

# **Προς μια Μεθοδολογία Καταγραφής και Ανάλυσης Επισκεψιμότητας Ιστοχώρων σε Εφαρμογές Διαδικτύου**

Μια διατριβή που παρουσιάστηκε στο Τμήμα  
Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών  
του Πολυτεχνείου Κρήτης  
σε εκπλήρωση των απαιτήσεων για την απόκτηση του  
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Μηχανική Υπολογιστών

Στέφανος Καρασαββίδης

Χανιά, Φεβρουάριος 2004

Δηλώνω ότι είμαι ο μοναδικός συγγραφέας αυτής της διατριβής. Εξουσιοδοτώ το Πολυτεχνείο Κρήτης να δανείζει αυτή τη διατριβή σε άλλους οργανισμούς, ή μεμονωμένα άτομα, για ερευνητικούς σκοπούς.

Επιπλέον, εξουσιοδοτώ το Πολυτεχνείο Κρήτης να αναπαράγει αυτή την διατριβή με φωτοτυπίες ή άλλους τρόπους, ολόκληρη ή μέρος της, μετά από αίτηση άλλων οργανισμών ή μεμονωμένων ατόμων για ερευνητικούς σκοπούς

Το λογισμικό που παρήχθη στα πλαίσια της διατριβής είναι διαθέσιμο ως ελεύθερο λογισμικό (free software) σύμφωνα με τις ρήτρες της άδειας χρήσης GNU General Public License έκδοση 2.0 (GPL 2.0) μέσω της διεύθυνσης <http://www.music.tuc.gr/webanalysis>.

Το Πολυτεχνείο Κρήτης ζητά τις υπογραφές όλων των προσώπων που χρησιμοποιούν ή αντιγράφουν αυτή τη διατριβή.

Παρακαλείστε να υπογράψετε παρακάτω και να αναγράψετε διεύθυνση και ημερομηνία.

# Περίληψη

Η ανάλυση επισκεψιμότητας ιστοχώρων (web usage analysis, web usage mining) είναι μία νέα και ενδιαφέρουσα περιοχή έρευνας με πολύ μεγάλο βιομηχανικό ενδιαφέρον, που μπορεί να χρησιμοποιηθεί παράλληλα με άλλες τεχνικές σχεδιασμού εφαρμογών όπως η ανάλυση χρησιμοποίησιμότητας (usability evaluation) σε εφαρμογές διαδικτύου. Σε αυτή την περιοχή έχουν προταθεί διάφοροι αλγόριθμοι στη διεθνή βιβλιογραφία που χαρακτηρίζονται από μεγάλο βαθμό πολυπλοκότητας. Όμως για τη μεγάλη πλειονότητα των επιχειρήσεων και οργανισμών, και ιδιαίτερα για μικρομεσαίες επιχειρήσεις (MME, SME), δεν είναι κατανοητό τι μπορεί να προσφέρει η ανάλυση επισκεψιμότητας ιστοχώρων στη δική τους κατηγορία εφαρμογών, και πως μπορεί να εφαρμόσει τους αλγόριθμους στους δικούς της ιστοχώρους. Εργαλεία που έχουν αναπτυχθεί από τη βιομηχανία για ανάλυση χρήσης ιστοχώρων είναι εξαιρετικά ακριβά και πρέπει επίσης να προσαρμοστούν στις ανάγκες των επιχειρήσεων με επιπρόσθετα μεγάλα κόστη συμβούλων. Έτσι μόνο μεγάλες επιχειρήσεις μπορούν να εγκαταστήσουν τέτοια εργαλεία και υπηρεσίες για ανάλυση χρήσης ιστοχώρων, ενώ επιχειρήσεις μικρότερης κλίμακας συναντούν ανυπέρβλητα προβλήματα.

Για να υποστηριχθεί μεγαλύτερο εύρος επιχειρήσεων και οργανισμών που δραστηριοποιούνται στο Διαδίκτυο χρειάζεται να αναπτυχθούν γενικής φύσης εργαλεία και μεθοδολογίες που να είναι βασισμένα σε φτηνό και διαδεδομένο λογισμικό και θα είναι προσαρμοσμένα σε συγκεκριμένες κατηγορίες εφαρμογών, ώστε να είναι εύκολα κατανοητά από όλες τις MME στην κατηγορία αυτή, και να προσαρμόζονται εύκολα στις ιδιαιτερότητες κάθε μίας επιχείρησης σε αυτή την κατηγορία. Αυτή η διατριβή συνεισφέρει στην ανάπτυξη συστηματικών μεθοδολογιών και λογισμικού για ανάλυση χρήσης και χρησιμοποίησιμότητας ιστοχώρων για συγκεκριμένες κατηγορίες εφαρμογών που περιλαμβάνουν πολλές μικρομεσαίες επιχειρήσεις.

Προτείνουμε μία συστηματική μεθοδολογία και σύστημα βάσης δεδομένων για την καταγραφή δεδομένων χρήσης ιστοχώρων που απευθύνεται σε ιστοχώρους τουριστικού προορισμού του Διαδικτύου. Προτείνουμε επίσης συστηματικές μεθοδολογίες ανάλυσης σεναρίων χρήσης τουριστικών ιστοχώρων. Παρουσιάζουμε μία σειρά αναλύσεων σεναρίων χρήσης και συμπεριφοράς επισκεπτών ενός συγκεκριμένου ιστοχώρου τουρισμού για να δείξουμε τα πλεονεκτήματα της προτεινόμενης μεθοδολογίας και εργαλείων λογισμικού για ιστοχώρους στην ίδια κατηγορία. Δείχνουμε ότι οι προτεινόμενες μεθοδολογίες και λογισμικό μπορούν να χρησιμοποιηθούν και για την ανακάλυψη προβλημάτων χρησιμοποίησιμότητας (usability) παράλληλα ή συμπληρωματικά με άλλες πιο γνωστές μεθοδολογίες, αλλά το εύρος των συμπερασμάτων που μπορούν να εξαχθούν με τις προτεινόμενες μεθοδολογίες ανάλυσης είναι

πολύ μεγαλύτερο και περιλαμβάνει και μη προβλεπόμενα σενάρια χρήσης.

Τα αποτελέσματα αυτής της διατριβής πιστεύουμε ότι είναι χρήσιμα σε ένα μεγάλο εύρος επιχειρήσεων που δραστηριοποιούνται στο Διαδίκτυο με ιστοχώρους τουριστικού περιεχομένου.

Το λογισμικό που υποστηρίζει τη μεθοδολογία για ανάλυση επισκεψιμότητας που προτείνουμε είναι διαθέσιμο για οποιαδήποτε επιχείρηση ή οργανισμό ως ελεύθερο λογισμικό (free software) μέσω της διεύθυνσης <http://www.music.tuc.gr/webanalysis>.

# Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Σταύρο Χριστοδουλάκη για την επίβλεψη της διατριβής και τη μοναδική του ικανότητα να διακρίνει το σημαντικό ανάμεσα στα ασήμαντα. Θα ήθελα επίσης να ευχαριστήσω τα άλλα δύο μέλη της τριμελούς επιτροπής αξιολόγησης της διατριβής, τους αναπληρωτή καθηγητή κ. Μανόλη Κουμπάρακη και αναπληρωτή καθηγητή Ευριπίδη Πετράκη, για την ανάγνωσή της και τις παρατηρήσεις τους.

Στα πλαίσια της εργασίας διεξήχθησαν και ορισμένα πειράματα, στα οποία συμμετείχαν άτομα χωρίς την βοήθεια των οποίων δε θα ήταν δυνατή η ολοκλήρωση της εργασίας. Σε αυτά τα άτομα απευθύνω τις θερμές μου ευχαριστίες για το χρόνο που αφιέρωσαν και τη βοήθεια που μου προσέφεραν. Ήταν οι Απόστολος Πάγκος, Επιμενίδης Βουτσάκης, Ιωάννης Βαρελάς, Παναγιώτης Καραγεωργάκης, Παναγιώτης Πατεράκης, Beate Bässler και Ute Birghof. Στα ίδια πλαίσια θέλω να απευθύνω ένα ιδιαίτερο ευχαριστώ στους Νεκτάριο Γιολδάση και Γιώργο Ανέστη.

Θέλω να ευχαριστήσω επίσης τη Γεωργία Κυριακάκη και το Σωτήρη Μπούρο για την πολύτιμη βοήθεια κατά την προετοιμασία της παρουσίασης της διατριβής καθώς και το Νεκτάριο Μουμουτζή για την βοήθεια που προσέφερε σε καθοριστικά σημεία της εργασίας και την προθυμία του να αφιερώσει τον περιορισμένο του χρόνο. Επίσης θέλω να ευχαριστήσω τον Αναστάσιο Ζαφειρόπουλο που βοήθησε στη δημιουργία παραστατικών διαγραμμάτων από τους πολυπληθής και ασήμαντους αριθμούς. Τέλος θέλω να ευχαριστήσω το Θεόδωρο Μαργαζά για την υποστήριξη και ενθάρρυνση που μου παρείχε καθ' όλη τη διάρκεια εκπόνησης της διατριβής.

Χανιά, Φεβρουάριος 2004

Στέφανος Καρασαββίδης

αφιερώνεται  
σ' αυτούς που νιώθουν  
ότι οι αγωνίες  
οι κόποι και οι θυσίες τους  
βρίσκουν ανταπόκριση  
με τις δικές μας επιτυχίες

# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή.....</b>	<b>19</b>
1.1	Σκοπός.....	20
1.2	Αναγκαιότητα.....	21
1.3	Δομή της Εργασίας.....	22
<b>2</b>	<b>Σχετική Έρευνα.....</b>	<b>25</b>
2.1	Αλληλεπίδραση Ανθρώπου Υπολογιστών (HCI).....	25
2.1.1	Μορφές Αλληλεπίδρασης.....	27
2.2	Χρησιμοποιησιμότητα (Usability).....	30
2.2.1	Ορισμός Usability.....	30
2.2.2	Ιδιαιτερότητες Εφαρμογών στον Παγκόσμιο Ιστό.....	32
2.2.3	Αξιολόγηση Χρησιμοποιησιμότητας.....	36
2.2.4	Προσέγγιση Web Usage Analysis.....	45
2.3	Web Usage Analysis/Mining.....	46
2.3.1	Τι είναι Web Mining.....	47
2.3.2	Ανάλυση και Αξιολόγηση Επισκεψιμότητας.....	48
2.3.3	Συλλογή Δεδομένων – Πρωτογενή Δεδομένα.....	54
2.3.4	Μετρικές και Μεγέθη για την Ανάλυση Επισκεψιμότητας.....	64
2.3.5	Προεπεξεργασία Δεδομένων (Data Preprocessing).....	76
2.3.6	Ανακάλυψη Γνώσης (Knowledge Discovery).....	93
2.3.7	Ανάλυση.....	100
2.4	Ιστοχώροι Τουριστικού Προορισμού.....	107
2.5	Συνεισφορά.....	109
2.6	Ανακεφαλαίωση.....	111
<b>3</b>	<b>Μεθοδολογία και Σύστημα Βάσης Δεδομένων για την Καταγραφή Δεδομένων Χρήσης Ιστοχώρων.....</b>	<b>112</b>
3.1	Σύστημα Βάσης Δεδομένων για την Καταγραφή Δεδομένων Χρήσης Ιστοχώρων.....	112
3.1.1	Βασικές Αρχές.....	113
3.1.2	Δομές Σελιδοπροβολής.....	114
3.1.3	Υπηρεσία Ενημέρωσης.....	117
3.1.4	Βάση Χρήσης.....	118
3.1.5	Αρχιτεκτονική Συστήματος.....	123
3.2	Μεθοδολογία Επιλογής Στοιχείων Καταγραφής.....	126
3.2.1	Μεθοδολογία Δημιουργίας Δομών Σελιδοπροβολής.....	127
3.2.2	Εφαρμογή της Μεθοδολογίας Καταγραφής στον Ιστοχώρο Τουριστικού Προορισμού....	129
3.3	Ανακεφαλαίωση.....	135
<b>4</b>	<b>Μεθοδολογία Ανάλυσης από το Σύστημα Καταγραφής Δεδομένων Χρήσης Ιστοχώρων.....</b>	<b>136</b>
4.1	Χαρακτηριστικά Επισκέψεων.....	137
4.1.1	Παράμετροι Επίσκεψης.....	137
4.1.2	Παράμετροι Εντός Ιστοχώρου.....	145
4.2	Προσωποποίηση Ιστοχώρου.....	148
4.3	Χρησιμοποιησιμότητα.....	152
4.4	Σενάρια Χρήσης.....	154
4.4.1	Εντοπισμός Σεναρίων.....	156
4.4.2	Συμπεράσματα από την Ανάλυση Χρήσης Σεναρίων.....	158
4.4.3	Γενική Μορφή Ερωτήσεων.....	158
4.4.4	Έκφραση Ερωτήσεων για Σενάρια Χρήσης.....	159
4.4.5	Σύνθετο Παράδειγμα Σεναρίου Χρήσης σε SQL92 και SQL99.....	167
4.4.6	Δυναμικά Σενάρια Χρήσης.....	172



4.5 Ανακεφαλαίωση.....	173
<b>5 Μελέτες Χρησιμοποίησιμότητας Ιστοχώρων με Παραδοσιακές Μεθόδους και με τη Μεθοδολογία Καταγραφής και Ανάλυσης Δεδομένων Χρήσης Ιστοχώρων.....</b>	<b>174</b>
5.1 Σενάρια Χρήσης Εφαρμογής Ιστοχώρου.....	174
5.1.1 Πρωταρχικά Σενάρια Χρήσης.....	176
5.1.2 Δευτερεύοντα Σενάρια Χρήσης.....	178
5.1.3 Συνδυασμός Σεναρίων Χρήσης Ιστοχώρου.....	179
5.2 Μελέτη Χρησιμοποίησιμότητας με Παραδοσιακές Μεθόδους.....	180
5.2.1 Ευριστική Αξιολόγηση.....	181
5.2.2 Πειράματα Χρηστών (User Tests).....	185
5.2.3 Συμπεράσματα Μελέτης Ιστοχώρου Τουρισμού με Παραδοσιακές Μεθόδους.....	194
<b>5.3 Μελέτη με τη Μεθοδολογία Καταγραφής και Ανάλυσης Δεδομένων Χρήσης Ιστοχώρων.....</b>	<b>195</b>
5.3.1 Γενικά Χαρακτηριστικά Ιστοχώρου.....	197
5.3.2 Προσωποποίηση Ιστοχώρου.....	221
5.3.3 Σύγκριση Εναλλακτικών.....	224
5.3.4 Σενάρια Χρήσης.....	234
5.3.5 Συμπεράσματα από τη Μελέτη με τη Μεθοδολογία Καταγραφής και Ανάλυσης Δεδομένων Χρήσης Ιστοχώρων.....	242
5.4 Σύγκριση Μεθόδων.....	244
<b>6 Προστασία Προσωπικών Δεδομένων κατά την Καταγραφή Χρήσης στο Web.....</b>	<b>247</b>
<b>7 Ανακεφαλαίωση - Συνεισφορά.....</b>	<b>251</b>
7.1 Επισκόπηση Περιεχομένων.....	251
7.2 Ερευνητική Συνεισφορά.....	252
<b>8 Μελλοντικές Επεκτάσεις.....</b>	<b>255</b>
<b>9 Αναφορές – Βιβλιογραφία.....</b>	<b>256</b>
<b>10 Παράρτημα Α – Ιστοχώροι Τουριστικού Προορισμού.....</b>	<b>262</b>
10.1 Χαρακτηριστικά Ιστοχώρων Τουριστικού Προορισμού.....	263
10.1.1 Ταξινόμηση Δεδομένων.....	263
10.1.2 Πολυγλωσσικότητα.....	263
10.1.3 Υπερκείμενο.....	263
10.1.4 Αναζήτηση.....	264
10.1.5 Κρατήσεις και Πληρωμές.....	264
10.1.6 Ψηφιακοί Χάρτες.....	264
10.1.7 Εποχικότητα.....	264
10.2 Ιστοχώροι Τουριστικού Προορισμού.....	265
10.2.1 www.austria-tourism.at.....	265
10.2.2 www.visit-sweden.com.....	265
10.2.3 www.tourspain.es.....	266
10.2.4 www.reiseland-brandenburg.de.....	266
10.2.5 www.hamburg-tourism.de.....	267
10.2.6 www.travelscotland.co.uk.....	267
10.2.7 www.visitsouthwest.co.uk.....	267
10.2.8 agn.hol.gr.....	268
10.3 Πιλοτική Ανάπτυξη Ιστοχώρου Τουριστικού Προορισμού.....	268
10.3.1 Ηλεκτρονικός Οδηγός.....	269
10.3.2 Περιοχές Ενδιαφέροντος και Αξιοθέατα.....	270
10.3.3 Λιμάνια και Αγκυροβόλια.....	271
10.3.4 Επιχειρήσεις Τουρισμού.....	272
10.3.5 Εκδηλώσεις.....	273
10.3.6 Γλωσσικός Οδηγός.....	273
10.3.7 Αναζήτηση Πληροφορίας.....	273

<b>11</b>	<b>Παράρτημα Β – Στιγμιότυπα από τον Ιστοχώρο Τουριστικού Προορισμού.....</b>	<b>274</b>
11.1	Αρχικές Εικόνες.....	274
11.2	Πilotική Λειτουργία.....	275
11.3	Αναμορφωμένη Έκδοση.....	277
<b>12</b>	<b>Παράρτημα Γ – Οδηγίες προς Αξιολογητές.....</b>	<b>286</b>
12.1	Purpose of Evaluation.....	286
12.2	The Tourism Site.....	286
12.2.1	Crete Guide.....	286
12.2.2	Areas & Sites.....	286
12.2.3	Harbors & Anchorages.....	287
12.2.4	Tourism Companies.....	287
12.3	Usage Scenarios.....	287
12.4	Evaluation Method.....	288
<b>13</b>	<b>Παράρτημα Δ – Σχέδιο Πειράματος Χρηστών.....</b>	<b>290</b>
13.1	Μελέτη Χρησιμοποιησιμότητας για τον Ιστοχώρο Τουριστικού Προορισμού.....	290
13.1.1	Σκοπός.....	290
13.1.2	Πρόβλημα – Αντικείμενο Μελέτης.....	290
13.1.3	Χρήστες.....	291
13.1.4	Μεθοδολογία Πειράματος.....	291
13.1.5	Λίστα Εργασιών.....	292
13.1.6	Περιβάλλον Πειράματος.....	299
13.1.7	Ο Ρόλος του Ειδικού.....	300
13.1.8	Μετρικές Αξιολόγησης.....	300
13.1.9	Συμφωνητικό.....	300
13.1.10	Ερωτηματολόγια.....	302

## Πίνακας Πινάκων

Πίνακας 2.1 Γνωστές και καθιερωμένες μορφές αλληλεπίδρασης	29
Πίνακας 2.2 Χαρακτηριστικές κινήσεις για πλοήγηση με την εφαρμογή Opera [49]	30
Πίνακας 2.3 Η γενική μορφή αιτήσεων και αποκρίσεων του HTTP πρωτοκόλλου	55
Πίνακας 2.4 Παράδειγμα αίτησης και απόκρισης HTTP πρωτοκόλλου	56
Πίνακας 2.5 Περιγραφές των κυριότερων διαθέσιμων πρωτογενών δεδομένων για την ανάλυση επισκεψιμότητας και οι πηγές τους	59
Πίνακας 2.6 Το Combined Log Format	60
Πίνακας 2.7 Εγγραφές σε Combined Log Format (CLF) από την εμφάνιση της αρχικής σελίδας του <a href="http://www.crete.tournet.gr">http://www.crete.tournet.gr</a> . Οι εναλλαγές των πεδίων του CLF εμφανίζονται με εναλλαγές στο φόντο των τιμών.	78
Πίνακας 2.8 Κανόνες για αναγνώριση αιτήσεων ως σελιδοπροβολές	80
Πίνακας 2.9 Τρία διαφορετικά σενάρια αιτήσεων ενός επισκέπτη για σελίδες στο [43]. Στην περίπτωση Α τα ονόματα των αιτήσεων επαρκούν για τον προσδιορισμό του περιεχομένου της κάθε σελιδοπροβολής, ενώ στην περίπτωση Β και Γ πρέπει να συμπεριληφθούν και οι παράμετροι του URI	81
Πίνακας 2.10 Εικονικό απόσπασμα των πεδίων IP, Request και User-Agent ενός CLF αρχείου καταγραφής. Διακρίνονται 2 διαφορετικές διευθύνσεις IP και 3 διαφορετικά User-Agent. Η ομαδοποίηση των αιτήσεων ανά επίσκεψη δεν είναι προφανής.	82
Πίνακας 2.11 Παραδείγματα τιμών του πεδίου User-Agent	83
Πίνακας 2.12 Ενδεικτικά ποσοστά επισκέψεων ανά όνομα περιοχής για το [43]	91
Πίνακας 2.13 Support & Confidence σε Association Rules από το [3]	95
Πίνακας 2.14 Η "κάθετη" δομή για τον υπολογισμό association rules	96
Πίνακας 2.15 Παραδείγματα κανόνων από Sequential Pattern Analysis	97
Πίνακας 2.16 Ορισμός των contact efficiency, relative contact efficiency και conversion efficiency από το [61]	104
Πίνακας 2.17 Ευριστικός αλγόριθμος για την εύρεση σελίδων που χρήζουν βελτίωσης (EVAL_conversion_efficiency [61])	105
Πίνακας 2.18 Πράξεις OLAP	106
Πίνακας 2.19 Διαστάσεις OLAP του [25] για την ανάλυση επισκεψιμότητας	107
Πίνακας 3.1 Περιεχόμενα του πακέτου σελιδοπροβολής	115
Πίνακας 3.2 Περιεχόμενα δομής σελιδοπροβολής	127
Πίνακας 3.3 Βασικές οντότητες καταγραφής για τον ιστοχώρο τουριστικού προορισμού με βάση τη μεθοδολογία επιλογής στοιχείων καταγραφής. Στην πρώτη στήλη φαίνονται οι Γενικές Σελίδες που καταγράφηκαν. Στη δεύτερη στήλη σημειώνεται εάν πρόκειται για Απλή ή Σύνθετη Γενική Σελίδα, ενώ στην τρίτη στήλη δίνεται μία σύντομη περιγραφή για το περιεχόμενο που παρουσιάζεται. Στην τέταρτη και τελευταία στήλη φαίνονται οι πιθανές ενέργειες της κάθε σελίδας.	134
Πίνακας 4.1 Παράμετροι που μπορεί να υπολογιστούν και να αναλυθούν όσον αφορά τις επισκέψεις ενός ιστοχώρου	138
Πίνακας 5.1 κλίμακα αξιολόγησης προβλημάτων χρησιμοποίησιμότητας	181
Πίνακας 5.2 Δημογραφικά στοιχεία χρηστών από το προκαταρκτικό ερωτηματολόγιο για κάθε συμμετοχή στα πειράματα	186
Πίνακας 5.3 Συγκεντρωτικά στοιχεία εκτέλεσης εργασιών στο πείραμα χρηστών	188
Πίνακας 5.4 Στοιχεία εκτέλεσης εργασιών στο πείραμα χρηστών	189
Πίνακας 5.5 Απαντήσεις των ερωτήσεων 1 έως 3 του ερωτηματολογίου περάτωσης	192
Πίνακας 5.6 Τα σημαντικότερα προβλήματα που εντοπίστηκαν ανά είδος μελέτης	195
Πίνακας 5.7 Ποσοστά επισκέψεων ανά Referring Host Name (Ιούνιος 2003)	203
Πίνακας 5.8 Κατανομή σελιδοπροβολών στις διάφορες σελίδες του ιστοχώρου (10/04/2003-30/06/2003). Τα γραμμοσκιασμένα ονόματα αναφέρονται σε σύνθετες σελίδες, παρουσιάζουν δηλαδή διαφορετικό περιεχόμενο της ίδιας κατηγορίας ανάλογα με τις παραμέτρους κλήσης του URL.	212
Πίνακας 5.9 Αριθμός σελιδοπροβολών ανά περιοχή ενδιαφέροντος και ανά γλώσσα εμφάνισης.	

Εμφανίζονται οι περιοχές που βρίσκονται στις 23 πρώτες προτιμήσεις σε κάποια γλώσσα προβολής (Ιούνιος 2003)	218
Πίνακας 5.10Επισκέψεις και σελιδοπροβολές για τις περιοχές ενδιαφέροντος και σελίδες σχετικής πληροφορίας για τις επισκέψεις με και χωρίς ειδικούς υπερσυνδέσμους	234
Πίνακας 5.11Γενικά χαρακτηριστικά για τις επισκέψεις πάνω στις οποίες μελετήθηκε το σενάριο χρήσης "Αναζήτηση Διαμονής"	235
Πίνακας 5.12Μήκος μονοπατιού εντός σελίδων του βέλτιστου μονοπατιού	236
Πίνακας 5.13Γενικές κατηγοριοποιήσεις στο σενάριο "αναζήτηση Διαμονής" με κριτήριο έναρξης την ακολουθία {"αρχική σελίδα", "φόρμα αναζήτησης διαμονής"}	238
Πίνακας 5.14Σελίδες που προηγούνται της εμφάνισης διαμονής, όταν υπάρχει αρχικά αποτυχία του σεναρίου με ακολουθία έναρξης {"αρχική σελίδα", "φόρμα αναζήτησης διαμονής"}	240
Πίνακας 5.15Είδος αναζήτησης όταν η προηγούμενη σελιδοπροβολή ήταν αποτελέσματα αναζήτησης	240
Πίνακας 10.1Πλήθος Περιοχών Ενδιαφέροντος ανά Νομό της Κρήτης	271
Πίνακας 10.2Πλήθος αξιοθεάτων ανά κατηγορία πρώτου επιπέδου	271
Πίνακας 10.3Πλήθος επιχειρήσεων τουρισμού ανά κατηγορία και είδος παρουσίασης όπως παρουσιάζονταν στον ιστοχώρο στο τέλος του προγράμματος	272

## Πίνακας Σχημάτων

Σχήμα 2.1 Επικοινωνία Ανθρώπου Υπολογιστή από το [65]	26
Σχήμα 2.2 Χρονική πορεία τεχνολογιών HCI και αντιστοιχία μετάβαση από ερευνητικό σε εμπορικό επίπεδο [66]	28
Σχήμα 2.3 Ένα μοντέλο για τα χαρακτηριστικά της αποδοχής ενός συστήματος από το [67] σελ. 25. Διακρίνονται και οι διαστάσεις του usability.	32
Σχήμα 2.4 Εκμάθηση εφαρμογής για αρχάριους και έμπειρους χρήστες [67] σελ. 28	38
Σχήμα 2.5 Σχέση αναλογίας κερδών/κόστους με αριθμό χρηστών σε πειράματα χρηστών [67] σελ. 174	42
Σχήμα 2.6 Σχέση ανάμεσα σε αριθμό αξιολογητών και ποσοστού προβλημάτων χρησιμοποιησιμότητας που εντοπίζονται [67] σελ. 156	44
Σχήμα 2.7 Διαδικασία Ανάλυσης Επισκεψιμότητας από το [8]	49
Σχήμα 2.8 Διαδικασία Ανάλυσης Επισκεψιμότητας από το [7]. Διακρίνονται τα σημεία ανάδρασης μετά από τα αποτελέσματα της ανάλυσης	50
Σχήμα 2.9 Διαδικασία Ανάλυσης Επισκεψιμότητας από το [22].	51
Σχήμα 2.10 Γενικό περίγραμμα της διαδικασίας ανάλυσης επισκεψιμότητας. Διακρίνονται τα στάδια της συλλογής δεδομένων, προεπεξεργασίας δεδομένων, ανακάλυψη γνώσεως και ανάλυσης αποτελεσμάτων. Η όλη διαδικασία γίνεται υπό την επίβλεψη εξειδικευμένου προσωπικού με ειδικές γνώσεις για τη δομή του ιστοχώρου, της ανάλυση και επεξεργασία των δεδομένων και για την περιοχή για την οποία προσφέρεται πληροφορία.	52
Σχήμα 2.11 Αλυσίδες επικοινωνίας HTTP πρωτοκόλλου	57
Σχήμα 2.12 Σχέση ανάμεσα αιτήσεων HTTP με σελιδοπροβολές, επισκέψεις και χρήστες	66
Σχήμα 2.13 Υπολογισμός Maximal Forward References από το [41].	90
Σχήμα 2.14 Παράδειγμα μονοπατιών για τον υπολογισμό του $conveff(P,T,*)$	105
Σχήμα 2.15 Οργάνωση δεδομένων OLAP από [63]	106
Σχήμα 3.1 Η φιλοσοφία του συστήματος καταγραφής δεδομένων χρήσης. Διακρίνονται οι δομές σελιδοπροβολής που στέλνονται με κάθε σελιδοπροβολή από διαφορετικές εφαρμογές ιστοχώρου στην υπηρεσία ενημέρωσης. Η υπηρεσία ενημέρωσης ενημερώνει με τη σειρά της το κεντρικό αποθηκευτικό σύστημα πάνω στο οποίο διενεργείται η ανάλυση	114
Σχήμα 3.2 Απλοποιημένο ER Διάγραμμα της Βάσης Χρήσης	119
Σχήμα 3.3 Φυσικό μοντέλο της Βάσης Χρήσης σε σχεσιακή βάση δεδομένων που υλοποιήθηκε για την αποθήκευση των δεδομένων χρήσης από εφαρμογές διαδικτύου	120
Σχήμα 5.1 Πιθανά μονοπάτια σελιδοπροβολών για την επιλογή διαμονής. Η εκκίνηση γίνεται είτε από την αρχική σελίδα, είτε από την παρουσίαση μιας περιοχής ενδιαφέροντος. Ο τερματισμός γίνεται με την παρουσίαση καταλυμάτων	177
Σχήμα 5.2 Όμοια με το Σχήμα 5.1 στην επιλογή ενοικίασης οχήματος ακολουθούνται τα ίδια μονοπάτια, με εξαίρεση την αναζήτηση στον τίτλο από την αρχική σελίδα	178
Σχήμα 5.3 Πιθανές ακολουθίες σελιδοπροβολών για την κατάληξη σε παρουσίαση περιοχής ενδιαφέροντος	179

## Πίνακας Εικόνων

Εικόνα 2.1 Απλή σελίδα HTML. Η αίτηση σε εξυπηρετητή HTTP για αυτή τη σελίδα δημιουργεί συνολικά 28 διαφορετικές αιτήσεις για τις διάφορες εικόνες	65
Εικόνα 2.2 Αρχική σελίδα του <a href="http://www.crete.tournet.gr">http://www.crete.tournet.gr</a> [43]	77
Εικόνα 5.1 Η λειτουργικότητα για "επιλεγμένες σελίδες" στην παλαιά έκδοση του ιστοχώρου	228
Εικόνα 5.2 Η επιλογή των επιλεγμένων σελίδων όπως ήταν αρχικά στη νέα έκδοση	229
Εικόνα 5.3 Η νέα ονομασία (e-φυλλάδιο) και η νέα τοποθεσία του σχετικού εικονιδίου	230
Εικόνα 5.4 Η παρουσίαση σχετικής πληροφορίας για επιχειρήσεις τουρισμού που βρίσκεται στο κάτω μέρος της κάθε παρουσίασης περιοχής ενδιαφέροντος	232
Εικόνα 5.5 Ειδικοί επιπλέον σύνδεσμοι για να μελετηθεί η αντίδραση των επισκεπτών σε διαφορετικές τοποθεσίες των συνδέσμων	233
Εικόνα 10.1 Χαρακτηριστική σελίδα του Ηλεκτρονικού Οδηγού του Crete TOURnet όπως αυτή φαινόταν στο τέλος του προγράμματος. Απόσπασμα από την ενότητα "Χλωρίδα"	270
Εικόνα 11.1 Εικόνα από τις πρώτες ημέρες λειτουργίας του ιστοχώρου	274
Εικόνα 11.2 Χάρτης περιοχών ενδιαφέροντος	275
Εικόνα 11.3 Παρουσίαση περιοχών ενδιαφέροντος	275
Εικόνα 11.4 Χάρτης λιμανιών και αγκυροβολίων	276
Εικόνα 11.5 Αναζήτηση διαμονής	276
Εικόνα 11.6 Αρχική σελίδα. Διακρίνονται στη δεξιά στήλη οι συντομεύσεις και η γενική αναζήτηση κειμένου. Στην αριστερή στήλη υπάρχει το πεδίο για αναζήτηση τίτλου διαμονής.	277
Εικόνα 11.7 Η σελίδα αναζήτησης καταλυμάτων. Διακρίνονται οι βασικές επιλογές για τις διευκολύνσεις, την κατηγορία και το είδος των καταλυμάτων, καθώς και η επιλογή για επιλογή περιοχής	278
Εικόνα 11.8 Το κάτω μέρος της παρουσίασης περιοχής ενδιαφέροντος. Διακρίνονται οι σύνδεσμοι για σχετική πληροφορία, όπως διαμονή κλπ.	279
Εικόνα 11.9 Αρχική σελίδα παρουσίασης διαμονής	280
Εικόνα 11.10 Παράθυρο επιλογής περιοχής που χρησιμοποιείται από τις φόρμες αναζήτησης διαμονής, ενοικίασης οχημάτων κλπ. Στο συγκεκριμένο στιγμιότυπο έχει επιλεγθεί στο χάρτη η περιοχή Χανίων και στην κάτω αριστερή περιοχή εμφανίζονται τα ονόματα των σχετικών περιοχών. Από αυτά ο χρήστης έχει τελικά επιλέξει τα Χανιά και τις Μουρνιές	281
Εικόνα 11.11 Το παράθυρο επιλογής περιοχής μέσω των επιλεγμένων περιοχών από το φυλλάδιο	282
Εικόνα 11.12 Εικόνα από τη σελίδα "Το Φυλλάδιο Μου". Διακρίνονται οι επιλογές ανά κατηγορία των καταλυμάτων, των αξιοθεάτων και των περιοχών ενδιαφέροντος	283
Εικόνα 11.13 Απόσπασμα από τη σελίδα αποτελεσμάτων αναζήτησης στο Crete TOURnet	284
Εικόνα 11.14 Ενότητα "Επικοινωνία & Τοποθεσία" στην παρουσίαση επιχειρήσεων τουρισμού	285

## Πίνακας Διαγραμμάτων

Διάγραμμα 3.1 Αρχιτεκτονική του Συστηματος.	124
Διάγραμμα 4.1 Απλοποιημένο σχήμα της σχεσιακής βάσης δεδομένων όπου φαίνονται οι σχέσεις για τις επισκέψεις (session), τις σελιδοπροβολές (page_impression) και τις ενέργειες σελίδας (action). Σημαντικό ρόλο παίζει το attribute sequence_order στη σχέση page_impression όπου αποθηκεύεται η σειρά της κάθε σελιδοπροβολής στην κάθε επίσκεψη.	161
Διάγραμμα 5.1 Αριθμός επισκέψεων στη διάρκεια του χρόνου και ανά έτος στο Crete TOURnet. Ο άξονας του χρόνου είναι ανά εβδομάδα. Η απότομη πτώση των επισκέψεων το καλοκαίρι του 2002 οφείλεται στην αποσύνδεση του ιστοχώρου λόγω τεχνικών προβλημάτων. Λόγω της αλλαγής της μορφής του ιστοχώρου παρουσιάζεται παρόμοια συμπεριφορά για λίγες ημέρες τον Απρίλιο του 2003 (1999-2003)	198
Διάγραμμα 5.2 Μέσος όρος αριθμού σελιδοπροβολών ανά επίσκεψη σε εβδομαδιαία βάση	199
Διάγραμμα 5.3 Κατανομή ποσοστών επισκέψεων στο χρόνο ανάλογα με τον αριθμό των σελιδοπροβολών. Οι επισκέψεις έχουν ομαδοποιηθεί ανάλογα με τον αριθμό σελιδοπροβολών. Στο διάγραμμα κάθε χρώμα αντιστοιχεί στο ποσοστό που έχει κάθε συγκεκριμένη ομάδα στο σύνολο των επισκέψεων (ποσοστό επισκέψεων με 1 σελιδοπροβολή, με 2-4 σελιδοπροβολές, 5-10, 11-30 και άνω των 30). Το άθροισμα των ποσοστών καλύπτει το σύνολο των επισκέψεων και είναι 100%.	201
Διάγραμμα 5.4 Ιστοχώροι προέλευσης σε ποσοστά επί του συνόλου των επισκέψεων (Ιούνιος 2003). Ομαδοποίηση ανά οργανισμό	204
Διάγραμμα 5.5 Εξέλιξη ιστοχώρων προέλευσης όπου υπήρχε σε εβδομαδιαία βάση ποσοστό άνω του 5%. (κανονικοποιημένα). Στο διάγραμμα κάθε χρώμα αντιστοιχεί σε συγκεκριμένο ιστοχώρο προέλευσης, και δείχνει για κάθε χρονική στιγμή το ποσοστό των επισκεπτών που προέρχονται από τον κάθε ιστοχώρο. Συμπεριλήφθηκαν μόνο οι προελεύσεις που κάποια χρονική στιγμή είχαν ποσοστό 5% των επισκεπτών. Έγινε κανονικοποίηση στο άθροισμα των επισκεπτών από αυτές και μόνο τις προελεύσεις.	206
Διάγραμμα 5.6 Επισκέψεις από τον ιστοχώρο abcnews.go.com. Παρατηρούμε ένα μικρό χρονικό διάστημα με πολλές επισκέψεις και στη συνέχεια μία απότομη και συνεχή πτώση. Τα δείγματα είναι ανά εβδομάδα	207
Διάγραμμα 5.7 Χώρες προέλευσης από τις καταλήξεις των ονομάτων περιοχής (domain names) των επισκεπτών (Ιούνιος 2003)	208
Διάγραμμα 5.8 Χώρες προέλευσης επισκεπτών για τον Ιούνιο 2003 με πηγή τη βάση GeoIP	209
Διάγραμμα 5.9 Κατανομή επισκεπτών ανά χώρα προέλευσης στη διάρκεια του 2003 ανά μήνα	210
Διάγραμμα 5.10 Κατανομή σελιδοπροβολών στις διάφορες σελίδες του ιστοχώρου (10/04/2003-30/06/2003)	212
Διάγραμμα 5.11 Κατανομή σελιδοπροβολών διαμονής ανά κατηγορία διαμονής (10/04/2003-30/06/2003)	213
Διάγραμμα 5.12 Κατανομή σελιδοπροβολών στις διάφορες σελίδες του ιστοχώρου (1/1/2002-31/12/2002)	214
Διάγραμμα 5.13 Σχετική κατανομή σελιδοπροβολών στη διάρκεια του χρόνου. Φαίνεται το ποσοστό επί του συνόλου των σελιδοπροβολών επιχειρήσεων τουρισμού, περιοχών ενδιαφέροντος (μαζί με αξιοθέατα και ενότητες εισαγωγής) και των υπολοίπων σελίδων. Ο οριζόντιος άξονας του χρόνου είναι ανά εβδομάδα (1/1/2002-30/6/2003)	215
Διάγραμμα 5.14 Ρυθμός μεταβολής ποσοστών ανά κατηγορία σελίδων στον ιστοχώρο. Είναι εμφανής η υπεροχή των σελίδων για επιχειρήσεις τουρισμού τους καλοκαιρινούς μήνες, έναντι των περιοχών ενδιαφέροντος (1/1/2002-30/6/2003)	216
Διάγραμμα 5.15 Ποσοστά επισκεψιμότητας ανά γλώσσα εμφάνισης και ανά περιοχή. Χαρακτηριστικές διαφορές ανάμεσα στους έλληνες επισκέπτες και τους ξένους (Ιούνιος 2003)	219
Διάγραμμα 5.16 Χαρακτηριστικές περιπτώσεις διαφοροποίησης επισκεψιμότητας περιοχών ενδιαφέροντος ανά χώρα προέλευσης. Φαίνονται τα ποσοστά προβολών περιοχών ανά χώρα προέλευσης	220
Διάγραμμα 5.17 Εξέλιξη επισκέψεων από <a href="http://www.gmx.net/frauen.html">http://www.gmx.net/frauen.html</a> . Διακρίνεται ότι η	

πρώτη επίσκεψη ήταν στις 4 Ιουνίου 1999 με απότομη αύξηση την αμέσως επομένη. Οι επισκέψεις διατηρήθηκαν σε μεγάλο αριθμό 4 ημέρες, ώσπου να σταματήσουν εντελώς μετά το Νοέμβριο του 1999, όπου και πιθανότατα αφαιρέθηκε η σελίδα.	223
Διάγραμμα 5.18 Φράσεις αναζήτησης ως ποσοστά επί των επισκέψεων στις οποίες χρησιμοποιήθηκαν (Ιούνιος 2003)	224
Διάγραμμα 5.19 Απόλυτος αριθμός σελιδοπροβολών μία εβδομάδα πριν και μία μετά την αλλαγή του ιστοχώρου	226
Διάγραμμα 5.20 Χρήση της υπηρεσίας "Ηλεκτρονικό Φυλλάδιο". Στο γράφημα στο κάτω τμήμα (γαλάζιο) φαίνεται το ποσοστό των επισκέψεων στις οποίες υπήρχε τουλάχιστον μία επίσκεψη στη σελίδα του ηλεκτρονικού φυλλαδίου. Από τις τελευταίες επισκέψεις, στο πάνω μέρος του γραφήματος (ερυθρό) φαίνεται τι ποσοστό επαναεπισκέφτηκε τη σελίδα, είχε δηλαδή πάνω από μία επίσκεψη. Στο κάτω μέρος διακρίνονται οι τρεις διαφορετικές φάσεις στο σχεδιασμό των συνδέσμων πρόσβασης στην υπηρεσία από το διαφορετικό εύρος που κυμαίνεται η καμπύλη.	231
Διάγραμμα 5.21 Γενικές κατηγοριοποιήσεις στο σενάριο "αναζήτηση Διαμονής" με κριτήριο έναρξης την ακολουθία {"αρχική σελίδα", "φόρμα αναζήτησης διαμονής"}	239
Διάγραμμα 6.1 Ο χρήστης επισκέπτεται πρώτα τον ιστοχώρο Α και μετά τον Β. Και οι δύο ιστοχώροι χρησιμοποιούν διαφημιστικά μηνύματα από τον Γ. Ο Γ γνωρίζει ότι ο χρήστης είδε της σελίδα στον ιστοχώρο Α και στον Β	248



## Πίνακας Ερωτήσεων

Ερώτηση 4.1 Αριθμός επισκέψεων ανά χώρα	139
Ερώτηση 4.2 Αριθμός επισκέψεων στη διάρκεια του χρόνου	139
Ερώτηση 4.3 Αριθμός επισκέψεων στη διάρκεια του χρόνου ανά χώρα προέλευσης	140
Ερώτηση 4.4 Αριθμός επισκέψεων ανά μηχανή προέλευσης	141
Ερώτηση 4.5 Αριθμός επισκέψεων ανά φράση αναζήτησης από εξωτερικές μηχανές αναζήτησης	141
Ερώτηση 4.6 Αριθμός επισκέψεων ανά φράση αναζήτησης και εξωτερικές μηχανή αναζήτησης	142
Ερώτηση 4.7 Αριθμός επισκέψεων με μία σελιδοπροβολή	143
Ερώτηση 4.8 Αριθμός επισκέψεων ανά όνομα περιοχής επισκέψεων	144
Ερώτηση 4.9 Κατανομή εφαρμογών πλοήγησης σε επισκέψεις	145
Ερώτηση 4.10 Αριθμός σελιδοπροβολών ανά Γενική Σελίδα	146
Ερώτηση 4.11 Αριθμός σελιδοπροβολών ανά Γενική Σελίδα ανά μήνα	146
Ερώτηση 4.12 Αριθμός σελιδοπροβολών ανά Γενική Σελίδα και ανά γλώσσα εμφάνισης ανά μήνα	147
Ερώτηση 4.13 Αριθμός σελιδοπροβολών ανά Γενική Σελίδα και ανά γλώσσα εμφάνισης ανά μήνα	147
Ερώτηση 4.14 Αριθμός σελιδοπροβολών ανά Γενική Σελίδα και ανά γλώσσα εμφάνισης ανά μήνα	148
Ερώτηση 4.15 Απότομη μεταβολή υπαρχόντων προελεύσεων	151
Ερώτηση 4.16 Εντοπισμός νέων προελεύσεων ανά εβδομάδα	152
Ερώτηση 4.17 Αριθμός χρήσης εναλλακτικής επιλογής ανά εναλλακτική	154
Ερώτηση 4.18 Παράδειγμα της γλώσσας MINT [71] για ανακάλυψη σειριακών σχημάτων	160
Ερώτηση 5.1 ψευδοSQL κριτήριο έναρξης σεναρίου αναζήτησης διαμονής	235
Ερώτηση 5.2 ψευδοSQL	236
Ερώτηση 5.3 ψευδοSQL	237
Ερώτηση 5.4 ψευδοSQL	238
Ερώτηση 5.5 ψευδοSQL	239
Ερώτηση 5.6 ψευδοSQL	240

## Πίνακας Πλαισίων

Πλαίσιο 4.1 Ακολουθία δύο συνεχόμενων σελιδοπροβολών	162
Πλαίσιο 4.2 Ακολουθία n συνεχόμενων σελιδοπροβολών	162
Πλαίσιο 4.3 Ακολουθία με περιορισμός στα a και c	163
Πλαίσιο 4.4 Ακολουθία με περιορισμός στα a και c και εξαγωγή των συγκεκριμένων επισκέψεων	163
Πλαίσιο 4.5 Ακολουθίες με τη χρήση wildcard	163
Πλαίσιο 4.6 Ακολουθία με περιορισμό στα a και c και wildcard ανάμεσά τους	164
Πλαίσιο 4.7 ΛΑΘΟΣ υπολογισμός ακολουθίας $a b^* c$ με περιορισμό στο όνομα της b	164
Πλαίσιο 4.8 Υπολογισμός ακολουθιών $a b^* c$ που δεν καλύπτουν τους περιορισμούς για το wildcard	165
Πλαίσιο 4.9 Υπολογισμός ακολουθιών $a b^* c$ που δεν καλύπτουν τους περιορισμούς για το wildcard	166
Πλαίσιο 4.10 Υπολογισμός ακολουθιών $a b^* c$ που δεν καλύπτουν τους περιορισμούς για το wildcard	167
Πλαίσιο 4.11 Ερώτηση σε ψευδοSQL για σενάριο επιτυχούς αναζήτησης διαμονής	168
Πλαίσιο 4.12 Ερώτηση σε SQL92 για σενάριο επιτυχούς αναζήτησης διαμονής	170
Πλαίσιο 4.13 Ερώτηση σε SQL99 για σενάριο επιτυχούς αναζήτησης διαμονής	171
Πλαίσιο 4.14 Ερώτηση σε ψευδοSQL για σενάριο επιτυχούς αναζήτησης διαμονής	173

# 1 Εισαγωγή

Το Διαδίκτυο υποσχέθηκε να φέρει την επανάσταση στη διάδοση πληροφοριών και στον τρόπο του επιχειρήν και του μάρκετινγκ. Η ρυθμός εμφάνισης νέων ιστοχώρων, από την αρχή της δεκαετίας του 90 και μετά, χαρακτηρίζεται ως εκρηκτικός. Όλοι πλέον οι οργανισμοί και οι εταιρείες προσπαθούν να εκμεταλλευτούν το νέο μέσο είτε για τη βελτίωση των παρεχόμενων υπηρεσιών είτε για τη δημιουργία νέων πηγών εσόδων. Δε χρειάστηκε βέβαια να εισέλθει ο χώρος της Νέας Οικονομίας σε κρίση, με χιλιάδες πτωχεύσεις και απολύσεις προσωπικού, για να διαπιστωθεί ότι οι αρχικές προσδοκίες από το Διαδίκτυο ήταν από μη ρεαλιστικές ως και ουτοπικές.

Σύντομα σε ερευνητικό αλλά κυρίως, λόγω του εμπορικού ενδιαφέροντος, σε επιχειρηματικό επίπεδο, άρχισε η διερεύνηση μεθόδων για τη βελτίωση των ιστοχώρων. Ο σκοπός ήταν και παραμένει η προσαρμογή των ιστοχώρων ώστε να καλύπτονται σε μεγαλύτερο βαθμό οι ανάγκες των χρηστών. Στους εμπορικούς ιστοχώρους υπάρχει η επιθυμία οι χρήστες να οδηγούνται σε συγκεκριμένες συμπεριφορές (για παράδειγμα να προχωρούν σε αγορές, να ανταποκρίνονται σε διαφημιστικά μηνύματα κλπ.). Γενικότερα γίνεται προσπάθεια οι επισκέπτες να βρίσκουν με μεγαλύτερη ευκολία την πληροφορία ή τις υπηρεσίες που αναζητούν.

Ο στόχος της βελτίωσης των ιστοχώρων αντιμετωπίστηκε αρχικά στο επίπεδο επικοινωνίας-ανθρώπου υπολογιστή (Human Computer Interaction- HCI), μεταφέροντας τις μεθοδολογίες και εμπειρίες που υπάρχουν από παραδοσιακές εφαρμογές ηλεκτρονικών υπολογιστών στις εφαρμογές Διαδικτύου (web applications). Οι ιδιαιτερότητες των εφαρμογών Διαδικτύου σε σύγκριση με τις παραδοσιακές εφαρμογές οδήγησαν σε προσαρμογές στις μεθοδολογίες που χρησιμοποιούνται, ενώ σήμερα πλέον αυτού του είδους οι εφαρμογές αποτελούν ένα ειδικευμένο είδος έρευνας και μελέτης στο χώρο του HCI. Ο όρος που χρησιμοποιείται στο HCI για να εκφράσει το βαθμό στον οποίο καλύπτονται οι ανάγκες των χρηστών είναι usability (χρησιμοποιησιμότητα).

Ένα σημαντικό στοιχείο που μελετάται είναι η συμπεριφορά των επισκεπτών των ιστοχώρων. Η απλή λειτουργία ενός ιστοχώρου δε δίνει όμως πληροφορίες για τη συμπεριφορά των επισκεπτών του. Συνήθως υπάρχει άγνοια ακόμη και για στοιχειώδη χαρακτηριστικά των επισκεπτών, όπως για παράδειγμα με ποιον τρόπο έφτασαν στις συγκεκριμένες σελίδες ή για ποιο θέμα ενδιαφέρονται να πληροφορηθούν.

Στα μέσα της δεκαετίας του 90 εμφανίστηκαν διαφορετικού είδους έρευνες [15] όπου διαπιστώθηκε ότι μελετώντας τα δεδομένα που προκύπτουν από τις εντολές ελέγχου του HTTP πρωτοκόλλου, εμφανίζονται χρήσιμες πληροφορίες και μπορούν να εξαχθούν συμπεράσματα για

ένα πλήθος στοιχείων που χαρακτηρίζουν τους επισκέπτες των ιστοχώρων. Η ανάλυση αυτής της πληροφορίας μπορεί να οδηγήσει σε καλύτερη κατανόηση των αναγκών των επισκεπτών και τελικά σε συμπεράσματα για τον τρόπο συμπεριφοράς τους. Με βάση αυτή την ανάλυση γίνονται κατάλληλες προσαρμογές στην εφαρμογή Διαδικτύου με σκοπό τη βελτίωσή της. Ο όρος που φαίνεται να καθιερώνεται για αυτού του είδους τις αναλύσεις είναι Web Usage Analysis (ή Web Usage Mining όταν εμπλέκονται και τεχνολογίες Data Mining), ενώ στα ελληνικά χρησιμοποιείται συχνά ο όρος «ανάλυση επισκεψιμότητας», αν και συνήθως αναφέρεται σε συνδυασμό με ένα περιορισμένο αριθμό μετρικών που αναλύονται.

Κοινό χαρακτηριστικό των περισσότερων συστημάτων καταγραφής χρήσης σε ιστοχώρους είναι η χρήση των αρχείων καταγραφής από τους εξυπηρετητές παγκόσμιου ιστού. Τα δεδομένα που παράγονται με αυτό τον τρόπο είναι τεράστια σε μέγεθος και χρήζουν προεπεξεργασίας για να αντιμετωπιστούν στρεβλώσεις που υπεισέρχονται από διάφορους παράγοντες όπως οι προσβάσεις στους ιστοχώρους από αυτόματες μηχανές, η ύπαρξη ειδικής δομής στις διευθύνσεις στα αρχεία καταγραφής που δίνουν νόημα σε έννοιες όπως το περιεχόμενο που εμφανίζεται με κάθε πρόσβαση, η γλώσσα εμφάνισης του περιεχομένου κ.α.

## **1.1 Σκοπός**

Στην παρούσα διατριβή διερευνάται η πιθανή συνεισφορά των δεδομένων που προκύπτουν από τη χρήση ιστοχώρων στη βελτίωση της χρησιμοποιησιμότητας εφαρμογών στο Διαδίκτυο.

Γι' αυτό το σκοπό προτείνουμε και παρουσιάζουμε τις λεπτομέρειες παρουσιάζεται ενός συστήματος καταγραφής που εκμεταλλεύεται την ύπαρξη ειδικής γνώσης για τη δομή και το περιεχόμενο ενός ιστοχώρου και καταγράφει σε πραγματικό χρόνο τις προσβάσεις σε αυτόν. Για να μειώσουμε το εύρος των πιθανών εφαρμογών ώστε να γίνει η μεθοδολογία που προτείνουμε πιο κατανοητή και πιο εύκολα προσαρμόσιμη σε μια συγκεκριμένη εφαρμογή ιστοχώρου, επικεντρωνόμαστε σε μια μεγάλη κατηγορία εφαρμογών ιστοχώρου που έχει πολύ όμοιες απαιτήσεις και χαρακτηριστικά: τους ιστοχώρους τουριστικού προορισμού.

Το σύστημα χρησιμοποιείται για την αναλυτική καταγραφή των κινήσεων των επισκεπτών σε έναν πραγματικό ιστοχώρο τουριστικού προορισμού. Με βάση αυτά τα στοιχεία παρουσιάζουμε μία μεθοδολογία και διαφορετικές προσεγγίσεις για την εξαγωγή συμπερασμάτων για την χρησιμοποιησιμότητα του ιστοχώρου. Για να υπάρξει κάποιο μέτρο σύγκρισης, παρουσιάζεται ταυτόχρονα μία ολοκληρωμένη μελέτη χρησιμοποιησιμότητας με παραδοσιακές μεθόδους.

Τα αποτελέσματα των δύο προσεγγίσεων συγκρίνονται μεταξύ τους. Ο στόχος είναι να βρεθεί ο

βαθμός στον οποίο ένα τέτοιο σύστημα μπορεί να συνεισφέρει στον εντοπισμό προβλημάτων χρησιμοποίησιμότητας και εάν μπορεί να αντικαταστήσει μερικώς παραδοσιακές μεθόδους που χρησιμοποιούνται έως τώρα.

Ένας άλλος σκοπός είναι να παρουσιαστεί το εύρος των χρήσεων της μεθοδολογίας ανάλυσης ιστοχώρων για ιστοχώρους τουριστικού προορισμού έτσι ώστε να μπορούν να αποτελέσματα να χρησιμοποιηθούν από μικρομεσαίες επιχειρήσεις για να αξιολογήσουν τα οφέλη που μπορεί να έχουν με τη χρήση αυτών των μεθοδολογιών στη συγκεκριμένη κατηγορία εφαρμογών (ιστοχώρων τουρισμού).

Τέλος, ένας άλλος στόχος είναι να παρουσιαστεί η μεθοδολογία σε όλη της τη λεπτομέρεια ώστε να μπορεί εύκολα να γίνει κατανοητή από οποιαδήποτε επιχείρηση που δραστηριοποιείται τον τομέα ιστοχώρων τουρισμού για να μπορεί εύκολα η ίδια να προσαρμόσει τη μεθοδολογία και το λογισμικό που παρουσιάζεται σε αυτή τη διατριβή στις συγκεκριμένες ανάγκες της.

## **1.2 Αναγκαιότητα**

Παρόλο που τα τελευταία χρόνια γίνεται εντατική έρευνα σε θέματα καταγραφής και ανάλυσης των δεδομένων από επισκέψεις σε ιστοχώρους (ανάλυση επισκεψιμότητας) και τεχνικές που μπορούν να χρησιμοποιηθούν, ο χώρος βρίσκεται ακόμα σε εμβρυακό στάδιο. Παρόλο που οι περισσότεροι ερευνητές έχουν ως αφετηρία τα δεδομένα καταγραφής του HTTP πρωτοκόλλου και αναφέρονται στις ίδιες μετρικές που εξάγονται από αυτά (αν και πολλές φορές με διαφορετική ορολογία), δεν έχει αποκρυσταλλωθεί ακόμα κάποιος τυποποιημένος τρόπος για την εξαγωγή συμπερασμάτων.

Επειδή ο χώρος παρουσιάζει έντονο εμπορικό ενδιαφέρον οι σχετικές δημοσιεύσεις είναι μάλλον φειδωλές στην παρουσίαση λεπτομερειών για τις μεθόδους και τα μοντέλα που χρησιμοποιούν. Τα συστήματα που προσφέρονται στην αγορά δεν μπορούν εύκολα να συγκριθούν ενώ το κόστος απόκτησής των είναι ιδιαίτερα μεγάλο (για παράδειγμα το σύστημα ανάλυσης της Netgenesis [36] χρεώνεται με \$100.000 ανά CPU). Προσπαθώντας να καλύψουν όσο το δυνατόν μεγαλύτερο τμήμα της αγοράς, αναγκάζονται να χρησιμοποιούν ως πηγαία δεδομένα τα τυποποιημένα αρχεία καταγραφής των εξυπηρετητών HTTP και με αυτόν τον τρόπο συναντούν δυσκολίες σε θέματα αντιστοίχισης των αιτήσεων (requests) σε μοναδικές σελίδες που εμφανίζονται στον τελικό χρήστη. Ειδικά στους σύγχρονους ιστοχώρους, όπου το περιεχόμενο των σελίδων που εμφανίζονται παράγεται με δυναμικό τρόπο, οι μόνοι που τελικά έχουν ακριβή γνώση για το τι ακριβώς στέλνεται ως σύνθετο έγγραφο στον τελικό χρήστη στην οθόνη του, είναι οι άνθρωποι που κατασκευάζουν τον ιστοχώρο και αυτή η γνώση δεν εμφανίζεται στα αρχεία καταγραφής των εξυπηρετητών.

Ειδικά μικρές και μεσαίες επιχειρήσεις, τα παραπάνω έχουν ως αποτέλεσμα να είναι πολύ δύσκολο να χρησιμοποιήσουν αυτές τις καινούριες τεχνολογίες. Το κόστος αγοράς των αδειών χρήσης του λογισμικού είναι απαγορευτικό για αυτές τις επιχειρήσεις, και ιδιαίτερα εφόσον αυτό πρέπει να προσαρμοστεί με μεγάλο επιπλέον κόστος συμβούλων στην εφαρμογή της εκάστοτε μικρομεσαίας επιχείρησης. Ένας δεύτερος λόγος είναι ότι οι μικρομεσαίες επιχειρήσεις δεν μπορούν να δουν άμεσα τα οφέλη που μπορεί να έχουν από ένα γενικής χρήσεως πρόγραμμα ανάλυσης επισκεψιμότητας ιστοχώρων, που είναι ανεξάρτητο από τη δική τους εφαρμογή. Πολύ πιο εύκολα μπορούν να πεισθούν με συγκεκριμένα παραδείγματα χρήσεων που μπορεί να έχουν στο δικό τους εμπορικό ή επιχειρησιακό περιβάλλον. Έτσι απαιτείται η εξειδίκευση των γενικών προγραμμάτων ανάλυσης επισκεψιμότητας σε κατηγορίες εφαρμογών ιστοχώρου που είναι αρκετά συγκεκριμένες (όπως ιστοχώροι τουρισμού) αλλά που να περιλαμβάνουν επίσης πολύ μεγάλο αριθμό επιχειρήσεων στην κατηγορία αυτή.

Οι μελέτες χρησιμοποίησιμότητας από τη μεριά τους βασίζονται σε μεθοδολογίες από παραδοσιακές εφαρμογές, ενώ οι εφαρμογές που ολοένα χρησιμοποιούνται πιο συχνά και “τρέχουν” στο Διαδίκτυο, έχουν διαφορετικά χαρακτηριστικά. Σε αυτά περιλαμβάνονται η ευκολία με την οποία μπορεί να γίνουν αλλαγές στην εφαρμογή και να κυκλοφορήσουν άμεσα σε όλους τους χρήστες. Το αποτέλεσμα είναι η οργάνωση και διενέργεια παραδοσιακών μελετών χρησιμοποίησιμότητας, που απαιτούν σημαντικό χρόνο και κόστος, να έρχεται σε αντίθεση με την ανάγκη για γρήγορη αντιμετώπιση ενδεχόμενων προβλημάτων, αλλαγών χρήσεως ή δημιουργία νέων αγορών που προκύπτουν πολύ συχνά στο Διαδίκτυο. Τέτοιες αλλαγές στη χρήση είναι δυναμικές και δεν προέρχονται από κάποιο προκαθορισμένο σενάριο χρήσης (use case) που οι κατασκευαστές του ιστοχώρου είχαν προβλέψει.

Ο εκμετάλλευση της ευκολίας που παρέχεται από συστήματα καταγραφής χρήσης ιστοχώρων για την απόκτηση άμεσης γνώσης για τη συμπεριφορά των επισκεπτών μπορεί να δώσει σημαντικά πλεονεκτήματα έναντι παραδοσιακών μορφών μελέτης χρησιμοποίησιμότητας, αλλά και πιθανόν να μπορούσε να δώσει καινούρια γνώση για το γενικότερο επιχειρησιακό περιβάλλον στο οποίο ανήκει η εφαρμογή ιστοχώρου που να μπορεί να χρησιμοποιηθεί από την επιχείρηση. Στόχος μας είναι να δείξουμε τέτοια συγκεκριμένα παραδείγματα για μια γενική κατηγορία ιστοχώρων όπως οι ιστοχώροι τουριστικού προορισμού.

### **1.3 Δομή της Εργασίας**

Στο Κεφάλαιο 2 δίνεται μία εισαγωγή στο γνωστικό αντικείμενο της αλληλεπίδρασης του ανθρώπου με υπολογιστές και περιγράφεται η διάσταση της χρησιμοποίησιμότητας ως μέτρο της

ανταπόκρισης των εφαρμογών στις ανάγκες των χρηστών. Επισημαίνονται τα νέα δεδομένα που προκύπτουν από την ολοένα και μεγαλύτερη χρήση εφαρμογών που χρησιμοποιούν τεχνολογίες Διαδικτύου και ειδικότερα εφαρμογές πλοήγησης με βάση το HTTP πρωτόκολλο. Γίνεται μία σύντομη αναφορά σε καθιερωμένες πρακτικές και μεθόδους για τη βελτίωση της χρησιμοποιησιμότητας και δίνεται μία αναλυτική περιγραφή της σχετικής έρευνας σε θέματα ανάλυσης των δεδομένων που προκύπτουν από τη χρήση των ιστοχώρων με βάση το HTTP πρωτόκολλο.

Στο κεφάλαιο 3 παρουσιάζεται ένα σύστημα συλλογής και καταγραφής δεδομένων χρήσης ιστοχώρων που αναπτύχθηκε στα πλαίσια της διατριβής και μία μεθοδολογία για την επιλογή των στοιχείων εκείνων που δημιουργούν τη βάση ανάλυσης επισκεψιμότητας.

Στο κεφάλαιο 4 παρουσιάζεται μία συστηματική μεθοδολογία για την ανάλυση επισκεψιμότητας με αναφορά σε εφαρμογές ιστοχώρων τουριστικού προορισμού. Παρουσιάζεται επίσης ένα μοντέλο για την αναπαράσταση σεναρίων χρήσης σε έναν ιστοχώρο, στηριζόμενο στα μονοπάτια που ακολουθούνται κατά την επίσκεψη σε έναν ιστοχώρο και τα χαρακτηριστικά της κάθε σελίδας που εμφανίζεται.

Στο Κεφάλαιο 5 περιγράφεται η διενέργεια δύο ξεχωριστών μελετών χρησιμοποιησιμότητας που πραγματοποιήθηκαν στα πλαίσια της διατριβής πάνω στον ιστοχώρο τουριστικού προορισμού. Η πρώτη χρησιμοποιεί δύο καθιερωμένες μεθόδους για τη μελέτη χρησιμοποιησιμότητας. Στη δεύτερη μελέτη γίνεται προσπάθεια να χρησιμοποιηθούν τα δεδομένα από τη χρήση του ιστοχώρου για την εξαγωγή συμπερασμάτων.

Στο Κεφάλαιο 6 αναπτύσσεται ο προβληματισμός που υπάρχει κατά τη συλλογή δεδομένων για την παραβίαση στοιχειωδών δικαιωμάτων των χρηστών για προστασία προσωπικών δεδομένων.

Στο Κεφάλαιο 7 γίνεται μία ανακεφαλαίωση της διατριβής και παρουσιάζεται η ερευνητική συνεισφορά.

Στο Κεφάλαιο 8 παρουσιάζονται μελλοντικές επεκτάσεις.

Η διατριβή κλίνει με τις αναφορές και βιβλιογραφία που χρησιμοποιήθηκαν για τη συγγραφή της.

Η διατριβή περιλαμβάνει επίσης ένα σύνολο παραρτημάτων σαν συμπληρωματική πληροφορία για να γίνεται περισσότερο κατανοητό το κείμενο και οι αναφορές της διατριβής.

Στο Παράρτημα Α περιγράφονται τα χαρακτηριστικά ιστοχώρων τουριστικού προορισμού και ένας

ιστοχώρος που καλύπτει την περιοχή της Κρήτης και χρησιμοποιήθηκε για την ανάλυση επισκεψιμότητας.

Στο Παράρτημα Β δίνονται ορισμένα στιγμιότυπα εικόνων (screenshots) από τον ιστοχώρο τουριστικού προορισμού στα οποία γίνεται αναφορά στο κυρίως κείμενο.

Στο Παράρτημα Γ υπάρχουν οι οδηγίες που δόθηκαν στους αξιολογητές της ευριστικής αξιολόγησης του ιστοχώρου (μία από τις παραδοσιακές μεθόδους που χρησιμοποιήθηκε).

Στο Παράρτημα Δ υπάρχει το αναλυτικό σχέδιο πειράματος (test plan) στο οποίο βασίστηκαν τα πειράματα χρηστών (user tests).



## 2 Σχετική Έρευνα

Σε αυτό το κεφάλαιο θα δοθεί μία σύντομη περιγραφή για το γνωστικό πεδίο της Αλληλεπίδρασης Ανθρώπου Υπολογιστή, και ειδικότερα για το τμήμα που αναφέρεται στην έννοια και τις μεθόδους για την αξιολόγηση συστημάτων αλληλεπίδρασης. Στόχος δεν είναι η εξάντληση του θέματος, μιας και υπάρχουν μια πληθώρα σχετικών δημοσιεύσεων. Η έμφαση θα δοθεί στις ιδιαιτερότητες που υπάρχουν στις εφαρμογές ιστοχώρου με τις οποίες ασχολείται η παρούσα διατριβή.

### 2.1 Αλληλεπίδραση Ανθρώπου Υπολογιστών (HCI)

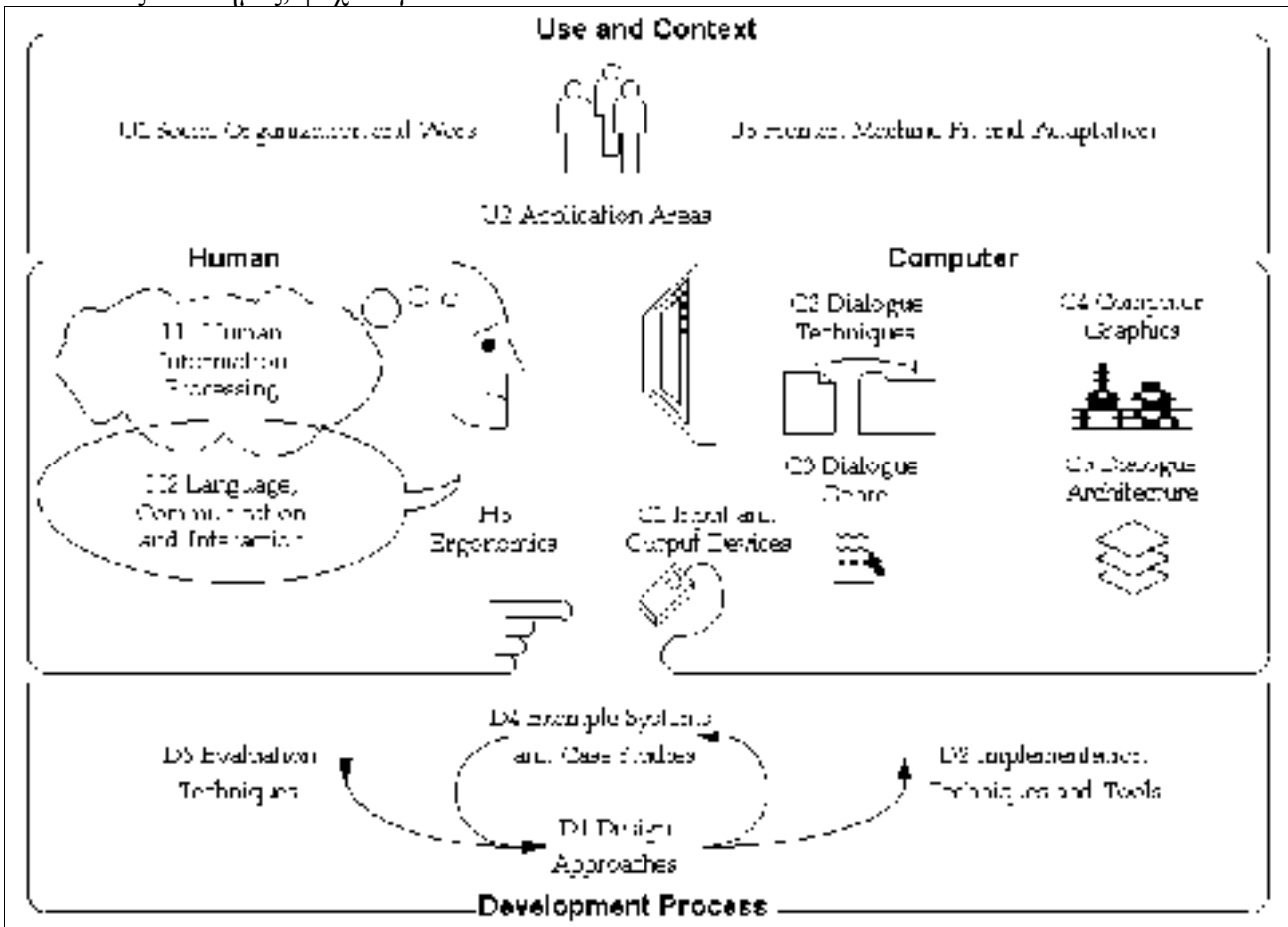
Η μελέτη της χρησιμοποίησιμότητας (Usability) μιας εφαρμογής εντάσσεται στο γνωστικό αντικείμενο της Αλληλεπίδρασης Ανθρώπου Υπολογιστή (Human Computer Interaction – HCI). Αν και δεν υπάρχει κάποιος ορισμός που να προσδιορίζει τους ακριβείς τομείς με τους οποίους ασχολείται η περιοχή, στο [65] χρησιμοποιείται για το HCI ο ορισμός

“Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them.”

Από την οπτική γωνία της Επιστήμης Υπολογιστών, η έμφαση δίνεται στην “αλληλεπίδραση” και πιο συγκεκριμένα στην αλληλεπίδραση ανάμεσα ενός ή περισσότερων ανθρώπων με μία ή περισσότερες υπολογιστικές μηχανές. Η αλληλεπίδραση γίνεται στο επίπεδο διεπικοινωνίας με το χρήστη (User Interface) που προσφέρει η εκάστοτε εφαρμογή.

Το πεδίο του HCI χαρακτηρίζεται στο [65] επιπλέον από το γεγονός ότι οι εργασίες γίνονται από κοινού από τον άνθρωπο και τη μηχανή. Επίσης το HCI ασχολείται με τη δομή της επικοινωνίας ανάμεσα στον άνθρωπο και τη μηχανή, τις ανθρώπινες ικανότητες να χρησιμοποιούν μηχανές (περιλαμβάνοντας την ευκολία εκμάθησης των Interface), τους αλγόριθμους και τον προγραμματισμό καθ' αυτών των interface, τα προβλήματα που εμφανίζονται στο σχεδιασμό και την ανάπτυξη των interface και τη διαδικασία της προδιαγραφής, του σχεδιασμού και της υλοποίησης των interface. Αναφέρεται χαρακτηριστικά ότι το HCI έχει πτυχές επιστήμης, μηχανικής και σχεδιασμού.

Ακριβώς επειδή στο HCI χειρίζονται θέματα επικοινωνίας των ανθρώπων και των μηχανών, απαιτείται η γνώση από επιστημονικά πεδία και από τις δύο πλευρές. Και αν από την πλευρά των υπολογιστικών μηχανών οι πτυχές της επιστήμης πληροφορικής που χρησιμοποιούνται (γραφική υπολογιστών, λειτουργικά συστήματα, γλώσσες προγραμματισμού κ.α.) φαντάζουν προσιτές, δεν ισχύει το ίδιο για την πλευρά του ανθρώπου όπου εμπλέκονται θεωρία επικοινωνίας, γλωσσολογία,



Σχήμα 2.1 Επικοινωνία Ανθρώπου Υπολογιστή από το [65]

Στο Σχήμα 2.1 διακρίνονται τέσσερις βασικές πτυχές του HCI έτσι όπως παρουσιάζονται στο [65]. Παρόλο που η δημοσίευση αναφέρεται στη σκιαγράφηση του HCI με σκοπό τη διαμόρφωση προτάσεων για τη διδασκαλία του πεδίου, εντούτοις δίνει μία πολύ καλή εικόνα για το εύρος των θεμάτων που άπτονται του HCI.

Πρόκειται για τη χρήση των υπολογιστών και το πλαίσιο στο οποίο γίνεται (U), τα ανθρώπινα χαρακτηριστικά (H), τα συστήματα υπολογιστών και η αρχιτεκτονική των interfaces (C) και, τέλος, τη διαδικασία ανάπτυξης (D).

Τα υπολογιστικά συστήματα συναντώνται μέσα σε ένα γενικότερο κοινωνικό και οργανικό περιβάλλον εργασίας (U1). Μέσα σε αυτό το πλαίσιο υπάρχουν εφαρμογές για τις οποίες υπάρχει η επιθυμία για υιοθέτηση υπολογιστικών συστημάτων (U2) αλλά απαιτείται οι πτυχές της εφαρμογής (ανθρώπινες, τεχνικές και εργασιακές) να συντονιστούν μεταξύ τους (U3).

Πρέπει επίσης να συνεκτιμηθούν ανθρώπινοι παράγοντες, όπως ο τρόπος επεξεργασίας

πληροφορίας από τον άνθρωπο (H1), η επικοινωνία (H2) αλλά και φυσικά χαρακτηριστικά του ανθρώπου (H3).

Στην πλευρά του υπολογιστή έχουν αναπτυχθεί ποικίλες τεχνολογίες που υποστηρίζουν την αλληλεπίδραση μέσω συσκευών εισόδου/εξόδου (C1), διαλόγων (C2), γραφικών (C4) και πιο σύνθετων λειτουργιών (C3, C5). Θα περιγραφούν πιο αναλυτικά παρακάτω κάποιες από αυτές τις τεχνολογίες.

Τέλος υπάρχει η διαδικασία της ανάπτυξης συστημάτων αλληλεπίδρασης, που εμπεριέχει το σχεδιασμό (D1), τεχνικές και εργαλεία για την υλοποίηση (D2) και τεχνικές και μεθόδους για την αξιολόγηση (D3).

Ειδικά το τελευταίο μέρος, δηλαδή η αξιολόγηση των interfaces θα αναπτυχθεί σε επόμενη ενότητα.

### **2.1.1 Μορφές Αλληλεπίδρασης**

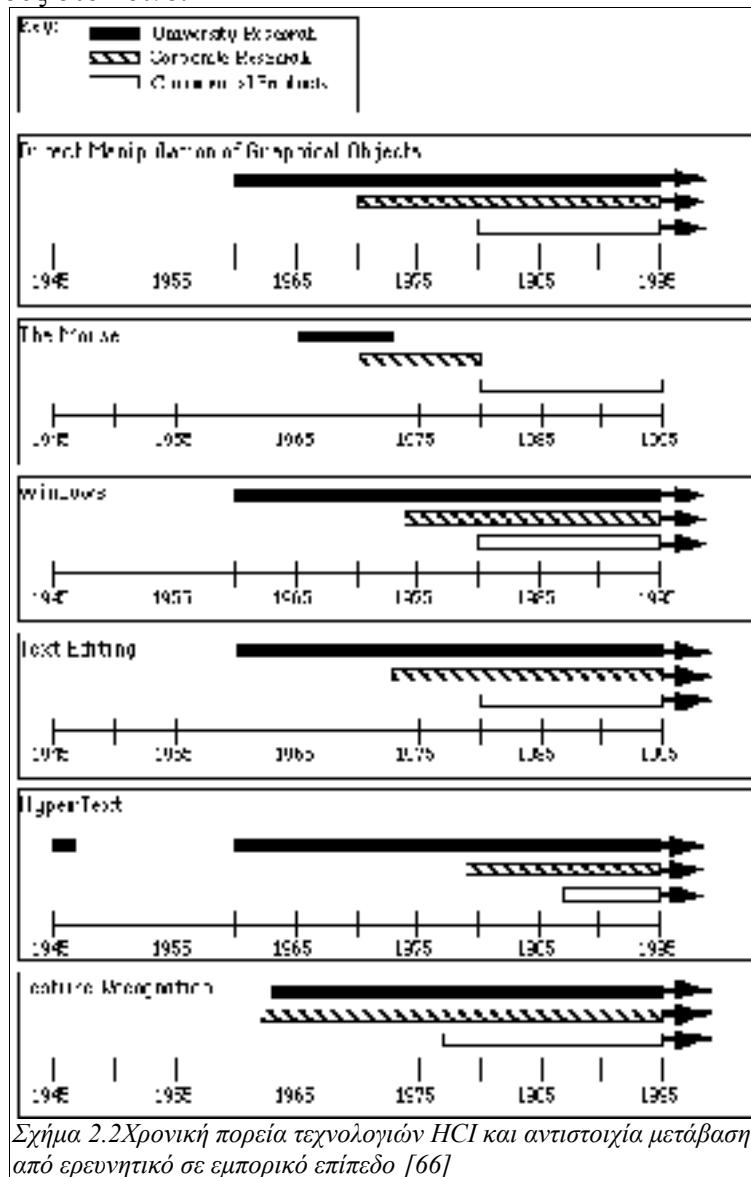
Όπως διακρίνεται στο Σχήμα 2.1, η αλληλεπίδραση του ανθρώπου με τον υπολογιστή γίνεται διαμέσω κατάλληλων συσκευών εισόδου εξόδου που υπάρχουν στον υπολογιστή. Οι συσκευές αυτές εκμεταλλεύονται τους διαύλους επικοινωνίας που διαθέτει ο άνθρωπος για να κάνουν εφικτή την επικοινωνία με το υπολογιστικό σύστημα. Τέτοιοι δίαυλοι είναι οι οπτικοί από την ανθρώπινη όραση, ακουστικοί, δίαυλοι αφής και η ανθρώπινη κίνηση.

Από τη μεριά του υπολογιστή, οι συσκευές εισόδου/εξόδου χωρίζονται συχνά σε συσκευές εισόδου κειμένου (πληκτρολόγια, συστήματα αναγνώρισης φωνής, συστήματα αναγνώρισης γραφής κ.α.), συσκευές εισόδου σημείων- pointing devices (ποντίκι, track ball, joystick, οθόνες αφής, συστήματα αναγνώρισης σημείων επικέντρωσης ματιού κ.α.) και συσκευές εξόδου όπως (οθόνες, εκτυπωτές, φωτεινές ενδείξεις κ.α.).

Οι συγκεκριμένες μορφές που παίρνει η αλληλεπίδραση του ανθρώπου με τα υπολογιστικά συστήματα εξελίχθηκε στο πέρασμα των χρόνων κυρίως λόγω της προόδου στις τεχνικές ικανότητες των υπολογιστικών συστημάτων. Συχνά καινοτόμες μορφές επικοινωνίας που μελετήθηκαν στα ερευνητικά εργαστήρια καθυστέρησαν πολλά χρόνια να εμφανιστούν στους υπολογιστές του ευρύ κοινού λόγω της περιορισμένης υπολογιστικής απόδοσης των υπολογιστών της εκάστοτε εποχής.

Όπως επισημαίνεται στο [66], σημαντικές καινοτομίες στον τρόπο επικοινωνίας ανθρώπων υπολογιστών που εμφανίζονται με εμπορικές εφαρμογές στο ευρύ κοινό, έχουν τις ρίζες τους σε

έρευνες στον τομέα του HCI από ερευνητικά εργαστήρια πανεπιστημίων που έγιναν πολλά χρόνια πριν την εμφάνισή τους στο κοινό.



Στο Σχήμα 2.2 φαίνεται η χρονική στιγμή που εμφανίστηκε μία νέα τεχνολογία στην αλληλεπίδραση ανθρώπων υπολογιστών σε ερευνητικά εργαστήρια και σε εμπορικά προϊόντα. Χαρακτηριστικά το ποντίκι “εφευρέθηκε” ήδη από το 1965 στο Stanford Research Laboratory, αλλά άρχισε να εμφανίζεται μετά το 1980 στα γραφεία των χρηστών.

<i>Τίτλος</i>	<i>Περιγραφή</i>
Γλώσσα εντολών (command language)	Από τις πρώτες μορφές αλληλεπίδρασης που χρησιμοποιήθηκαν και χρησιμοποιείται ακόμα ευρέως κυρίως από έμπειρους χρήστες. Ο χρήστης πληκτρολογεί στη “γραμμή εντολών” εντολές προς τον υπολογιστή χρησιμοποιώντας εξειδικευμένη γλώσσα.
Διάλογοι ερωτήσεων-απαντήσεων (question answer dialogs)	Ερωτήσεις και απαντήσεις εναλλάσσονται με συγκεκριμένη σειρά οδηγώντας στο επιθυμητό αποτέλεσμα.
Φόρμες εισαγωγής (fill in forms)	Χρησιμοποιούνται κυρίως για την εισαγωγή δεδομένων. Παρουσιάζεται μία φόρμα με κενά πεδία τα οποία καλείται να συμπληρώσει ο χρήστης. Συνήθως συνοδεύονται από έλεγχο λαθών, προκαθορισμένες τιμές, λίστες επιλογών κ.α.
Φυσική γλώσσα (natural language)	Ο χρήστης χρησιμοποιεί φυσική γλώσσα είτε “μιλώντας” στον υπολογιστή είτε γράφοντας σε χαρτί είτε στο πληκτρολόγιο. Ο υπολογιστής προσπαθεί να αναγνωρίσει συγκεκριμένες εντολές.
Κατάλογοι επιλογών (menus)	Ο χρήστης επιλέγει από ένα σύνολο δυνατών επιλογών που του παρουσιάζονται είτε με το ποντίκι είτε με το πληκτρολόγιο. Συνήθως η παρουσίαση των επιλογών ακολουθεί κάποια οργάνωση (ιεραρχία, αλφαβητική, εννοιολογική)
Απευθείας διαχείριση (direct manipulation)	Ο χρήστης ενεργεί απ' ευθείας πάνω σε ορατά αντικείμενα που συμβολίζουν συγκεκριμένες συσκευές ή έννοιες του υπολογιστή. Η αναπαράσταση των εννοιών γίνεται με ποικίλους τρόπους όπως εικονίδια, παράθυρα κ.α.

*Πίνακας 2.1 Γνωστές και καθιερωμένες μορφές αλληλεπίδρασης*






Στον Πίνακα 2.1 περιγράφονται συνοπτικά γνωστές και καθιερωμένες μορφές αλληλεπίδρασης με τους υπολογιστές. Ο αναγνώστης παραπέμπεται στην πληθώρα των σχετικών δημοσιεύσεων για πιο αναλυτική παρουσίαση των μορφών που μπορεί να πάρει η αλληλεπίδραση με τους υπολογιστές. Θα περιγραφεί πιο αναλυτικά, σαν παράδειγμα, μόνο η τεχνική των χειρονομιών ποντικιού (mouse gestures) όχι λόγω της καινοτομίας της αλλά επειδή τους τελευταίους μήνες φαίνεται να εισάγεται σε προϊόντα που διανέμονται μαζικά στο κοινό.

Οι χειρονομίες ποντικιού (mouse gestures) επιτρέπουν την αντιστοίχιση εντολών προς τον υπολογιστή σε κινήσεις που γίνονται με το ποντίκι. Εκεί που θα χρειαζόταν για παράδειγμα η επιλογή μιας εντολής από το menu της εφαρμογής, μπορεί η ίδια εντολή να δοθεί σχηματίζοντας ένα νοητό μονοπάτι με το ποντίκι. Το είδος των εντολών που δίνονται μπορεί να είναι οσοδήποτε πολύπλοκο. Τέτοια λειτουργικότητα υπάρχει ενσωματωμένη στην εφαρμογή πλοήγησης Διαδικτύου Opera [49] ενώ υπάρχουν διαθέσιμες επεκτάσεις και για πιο δημοφιλή εφαρμογές πλοήγησης όπως Mozilla ή Internet Explorer.

Στον Πίνακα 2.2 φαίνονται 5 χαρακτηριστικές κινήσεις για την πλοήγηση με την εφαρμογή Opera,

από την πληθώρα πιθανών κινήσεων που είναι διαθέσιμες.

Οι χειρονομίες ποντικιού υπάρχουν επίσης διαθέσιμες για γνωστά λειτουργικά συστήματα σαν προσθήκες, αλλά αναμένεται να ενσωματωθούν σε αυτά σε επόμενες εκδόσεις.

<i>Κίνηση</i>	<i>Περιγραφή</i>
	<b>Previous page in history.</b> Hold right button and move mouse left or hold right button and click left button
	<b>Next page in history.</b> Hold right button and move mouse right or hold left button then click right button
	<b>Go to home page.</b> Double-click in empty window
	<b>Open link in a new window in the foreground.</b> Move down then up
	<b>Minimize.</b> Move down then left

Πίνακας 2.2 Χαρακτηριστικές κινήσεις για πλοήγηση με την εφαρμογή Opera [49]

## 2.2 Χρησιμοποιησιμότητα (Usability)

### 2.2.1 Ορισμός Usability

Ως τώρα δόθηκε μία γενική εικόνα για τους τομείς με τους οποίους ασχολείται η Αλληλεπίδραση Ανθρώπων Υπολογιστών. Το προϊόν όλων αυτών των ενεργειών είναι το μέσον που δίνεται προς το χρήστη των εφαρμογών για να αξιοποιήσει τις υπηρεσίες που προσφέρονται από τον υπολογιστή. Το μέσον αυτό είναι ουσιαστικά το interface (επίπεδο διεπικοινωνίας) των εκάστοτε εφαρμογών και ο όρος που εκφράζει το πόσο “καλά” ένα interface ικανοποιεί το ρόλο του, ως ενδιάμεσο στη σχέση των χρηστών με τις εφαρμογές, είναι usability (χρησιμοποιησιμότητα). Ο όρος “user-friendly” που χρησιμοποιόταν παλαιότερα για να εκφράσει παρόμοια έννοια, έχει δεχτεί μεγάλη κριτική επειδή ερμηνευόντάς τον ετυμολογικά μπορεί να εξαχθεί το συμπέρασμα ότι ένα “φιλικό” interface είναι από μόνο του “καλό” interface. Ο Nielsen λέει χαρακτηριστικά στο [67] σελ. 23 ότι “...οι χρήστες δε χρειάζονται μηχανές που είναι φιλικές προς αυτούς, αλλά απλώς μηχανές που δεν τους

εμποδίζουν όταν θέλουν να κάνουν τη δουλειά τους”.

Αναζητώντας έναν αυστηρό ορισμό για το τι είναι usability διαπιστώνει κανείς ότι υπάρχουν τόσοι ορισμοί όσοι είναι οι συγγραφείς που ασχολούνται με το θέμα. Στο [68] συγκρίνονται ορισμοί που έχουν γίνει από διάφορους συγγραφείς. Αποσπάσματα από αυτή την εργασία παρατίθενται παρακάτω.

Σύμφωνα με το Merriam-Webster Dictionary “[Usability means]... *convenient and practical for use...*”. Αποτελεί έναν πολύ γενικό ορισμό με τον οποίο διαισθητικά θα συμφωνούσαν οι περισσότεροι.

Στο ISO-9241-11 (Guidance on Usability) usability ορίζεται ως “*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.*”. Αυτός ο ορισμός εισάγει αρκετούς περιορισμούς για το πλαίσιο στο οποίο πρέπει να γίνει η αξιολόγηση της χρησιμοποίησιμότητας. Αναφέρεται σε προσδιορισμένους χρήστες, που σημαίνει ότι πρέπει να είναι γνωστοί οι χρήστες του συστήματος και τα χαρακτηριστικά τους. Αναφέρεται επίσης σε προσδιορισμένους στόχους και σε προσδιορισμένο πλαίσιο χρήσης. Οι υπόλοιποι όροι δίνουν ήδη μία ένδειξη για τον τρόπο μέτρησης της χρησιμοποίησιμότητας.

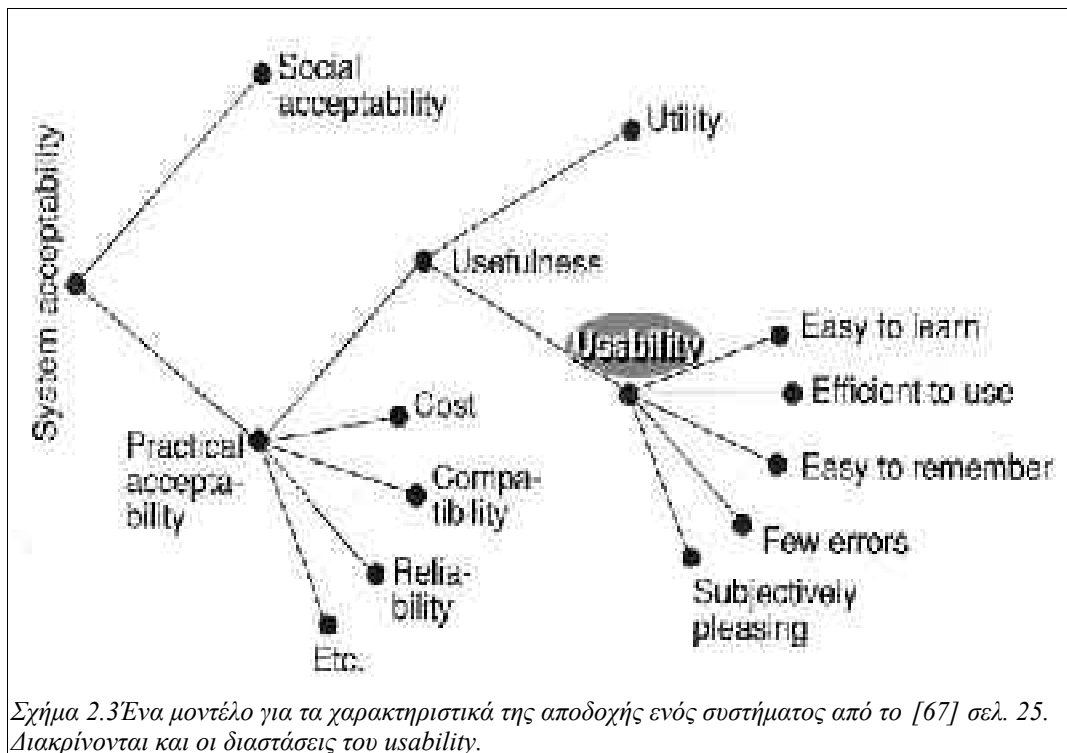
Ένας άλλος ορισμός των Allwood και Kalén, αναφέρει “*Usability [ . . . ] is an effect of the interaction between the user, the task, and the program, and it can be analysed into three parts: user-friendliness, user-acceptance and user-competence.*”. Σε αυτό τον ορισμό ενδιαφέρον παρουσιάζει το περί “user-competence” δηλαδή την επάρκεια ή ικανότητα του χρήστη να αντεπεξέλθει σε όλο το εύρος της λειτουργικότητας που προσφέρει η εφαρμογή. Θέτει δηλαδή ορισμένες απαιτήσεις και στο χρήστη πέραν της εφαρμογής και παραπέμπει έμμεσα και σε ζητήματα εκπαίδευσης των χρηστών.

Ο Löwgren από τη μεριά του ορίζει ότι “*Usability is a result of relevance, efficiency, learnability and attitude*”. Το relevance (σχετικότητα) αναφέρεται στο ότι το σύστημα υποστηρίζει τις “σωστές” λειτουργίες και έτσι εξυπηρετεί τις ανάγκες του χρήστη, ενώ η αποδοτικότητα ότι αυτές οι “σωστές” λειτουργίες γίνονται σε κάποιο επίπεδο απόδοσης (π.χ. σε σύντομο χρόνο). Επίσης το learnable σημαίνει ότι το σύστημα είναι κατ' αρχήν εύκολο στην εκμάθηση στην αρχική χρήση, αλλά ότι επίσης οι χρήστες μπορούν και θυμούνται τις λειτουργίες στο πέρασμα του χρόνου. Επίσης είναι σχετική η διάθεση που αποκτούν οι χρήστες με το σύστημα (θετική, αρνητική).

Όπως βέβαια επισημαίνει ο Booth “*Unfortunately [ . . . ] it appears unlikely that a generally agreed*

definition of usability will emerge within the next few years”.

Κλείνοντας παρατίθεται το σχήμα 2.3 από το [67] σελ. 25 για τη θέση του usability στην αποδοχή ενός συστήματος από τους χρήστες.



Χαρακτηριστικό στο παραπάνω σχήμα είναι ότι ο Nielsen διαχωρίζει την ικανότητα του συστήματος να διεκπεραιώσει αυτό που χρειάζεται (utility) από καθ' αυτό το usability που αναφέρεται στο πόσο καλά ο χρήστης μπορεί να εκμεταλλευτεί αυτές τις ικανότητες.

## 2.2.2 Ιδιαιτερότητες Εφαρμογών στον Παγκόσμιο Ιστό

Πριν γίνει αναφορά στην αξιολόγηση της χρησιμοποίησιμότητας των εφαρμογών στους ιστοχώρους (web applications) που εξετάζεται σε αυτή τη διατριβή, έχει ενδιαφέρον να εξεταστεί κατά πόσο αυτές οι εφαρμογές έχουν αξιοσημείωτες διαφορές από τις “παραδοσιακές” εφαρμογές των ηλεκτρονικών υπολογιστών. Το ζητούμενο είναι να εξακριβωθεί κατά πόσο οι μέθοδοι που έχουν αναπτυχθεί για την αξιολόγηση “παραδοσιακών” εφαρμογών μπορούν να εφαρμοσθούν με την ίδια αποτελεσματικότητα σε εφαρμογές στον παγκόσμιο ιστό.

Άμεσα προκύπτει η ανάγκη να οριστεί η έννοια της εφαρμογής παγκόσμιου ιστού και να προσδιοριστούν τα χαρακτηριστικά της. Δυστυχώς και σε αυτόν τον τομέα δεν υπάρχουν ορισμοί που να έχουν υιοθετηθεί ευρέως από την επιστημονική κοινότητα.



### **2.2.2.1 Εφαρμογές Ιστοχώρου**

Κατ' αρχήν, δεν υπάρχει διαφωνία στο ότι οι εφαρμογές του παγκόσμιου ιστού αναφέρονται πάντα σε εφαρμογές πλοήγησης που επικοινωνούν με το HTTP πρωτόκολλο με έναν ή περισσότερους HTTP εξυπηρετητές μέσω του Διαδικτύου. Από αυτή την άποψη η εφαρμογή βρίσκεται ουσιαστικά στην πλευρά των εξυπηρετητών και καλείται απομακρυσμένα από τις εφαρμογές πλοήγησης.

Οι απαντήσεις των εξυπηρετητών μπορεί να περιέχουν αποκλειστικά μη αλληλεπιδραστικά δεδομένα και η μόνη αλληλεπίδραση του χρήστη να είναι η επιλογή συνδέσμων που έχουν ως αποτέλεσμα την αποστολή νέων δεδομένων από τον εξυπηρετητή. Οι εφαρμογές πλοήγησης φροντίζουν για την εμφάνιση των δεδομένων που προέρχονται από τον εξυπηρετητή σύμφωνα με αναγνωρισμένα πρότυπα (π.χ. HTML, XML). Συνήθως οι απαντήσεις παρουσιάζονται σαν μία συνεχόμενη “σελίδα” στην εφαρμογή πλοήγησης και λέγεται ότι ο εξυπηρετητής στέλνει “σελίδες” στο χρήστη.

Η σελίδες μπορεί να στέλνονται χωρίς την εκτέλεση κάποιου προγράμματος ή μπορεί να είναι το αποτέλεσμα συνθέσεως δεδομένων από διάφορες πηγές και εκτέλεσης κάποιου προγράμματος. Στη δεύτερη περίπτωση το αποτέλεσμα μπορεί να εξαρτάται από το χρήστη που “ζήτησε” τα δεδομένα και μπορεί επίσης να εξαρτάται από τις προηγούμενες σελίδες που είδε ο χρήστης (η εφαρμογή δηλαδή έχει μία κατάσταση).

Οι σελίδες όμως μπορεί να περιέχουν και προς εκτέλεση πρόγραμμα, οπότε ένα μέρος του προγράμματος εκτελείται στην πλευρά της εφαρμογής πλοήγησης. Σε αυτή την περίπτωση η αλληλεπίδραση μπορεί να πάρει και πιο πολύπλοκες μορφές όπου στην ακραία περίπτωση έχουμε μία κλασσική εφαρμογή που εκτελείται αποκλειστικά ή εν μέρη στην πλευρά του χρήστη.

Σε σχετικές δημοσιεύσεις γίνεται συχνά ο διαχωρισμός ανάμεσα σε έναν “απλό” ιστοχώρο και σε μία “εφαρμογή” ιστοχώρου απαιτώντας για τον χαρακτηρισμό “εφαρμογή” να γνωρίζει το χρήστη που τη χρησιμοποιεί (με κάποιας μορφής αναγνώρισης χρήστη authentication) και να μπορεί ο χρήστης να αποθηκεύει μόνιμα δεδομένα της εφαρμογής (βλέπε π.χ. [69]). Αυτός ο ορισμός όμως είναι πολύ περιοριστικός και αποκλείει για παράδειγμα ιστοχώρους όπως μηχανές αναζήτησης. Για την παρούσα εργασία περιοριζόμαστε στη απαίτηση ο ιστοχώρος να ικανοποιεί κάποιον συγκεκριμένο στόχο (π.χ. την αναζήτηση πληροφορίας για ένα θέμα).

Συνοπτικά μπορούμε να πούμε ότι η Εφαρμογή Ιστοχώρου συνίσταται από ένα σύνολο *διασυνδεδεμένων σελίδων που εμφανίζονται σε εφαρμογή πλοήγησης που προέρχονται από έναν εξυπηρετητή HTTP πάνω από το Διαδίκτυο και εξυπηρετεί κάποιους συγκεκριμένους στόχους.*

Επειδή ο παραπάνω ορισμός είναι αρκετά ευρύς για το αντικείμενο της εργασίας εισάγονται και οι ακόλουθοι περιορισμοί:

- *οι εφαρμογές πλοήγησης που χρησιμοποιούνται δεν είναι εξειδικευμένες για ειδικές υπολογιστικές συσκευές. Ο περιορισμός αυτός χρειάζεται για να αποκλειστούν εφαρμογές που τρέχουν σε υπολογιστές παλάμης ή για παράδειγμα κινητά τηλέφωνα διότι θέτουν μεγάλους περιορισμούς στις τεχνολογίες που μπορεί να χρησιμοποιηθούν μέσα στις σελίδες, την πλοήγηση στον ιστοχώρο και κυρίως στο διαθέσιμο χώρο στην οθόνη του υπολογιστή. Η ανάπτυξη εφαρμογών προσαρμοσμένων σε τέτοιου είδους συσκευές γίνεται συνήθως έχοντας ξεχωριστές εκδόσεις των εφαρμογών.*
- *οι ιστοχώροι που εξετάζονται δε χρησιμοποιούν σε μεγάλο βαθμό άλλου είδους εφαρμογές ενσωματωμένες στις σελίδες του ιστοχώρου. Με αυτόν τον τρόπο αποκλείονται ιστοχώροι που ως κύριο μέσο αλληλεπίδρασης χρησιμοποιούν αυτόνομες τεχνολογίες ανάπτυξης εφαρμογών που απλώς είναι ενσωματωμένες σε σελίδες του ιστοχώρου. Τέτοιες τεχνολογίες είναι για παράδειγμα Java Applets, Macromedia Flash ή ActiveX Controls. Πέραν του ότι τέτοιες τεχνολογίες χρειάζονται ειδικές εγκαταστάσεις λογισμικού από το χρήστη για να μπορέσουν να χρησιμοποιηθούν, είναι επίσης κλειστές τεχνολογίες.*

#### **2.2.2.2 Ιδιαιτερότητες με παραδοσιακές εφαρμογές**

Το αρχικό ερώτημα ήταν εάν οι εφαρμογές ιστοχώρου είναι αρκετά διαφορετικές ώστε να δικαιολογούν τη διαφορετική αντιμετώπιση στην αξιολόγηση της χρησιμοποιησιμότητας σε σχέση με τις παραδοσιακές εφαρμογές. Οι απόψεις και σε αυτό το σημείο δυστυχώς δίστανται και γίνονται συζητήσεις πάνω στο θέμα [70].

Στο [26] καταλήγουν ότι δεν είναι ακόμα γνωστό ο τρόπος σχεδιασμού για αποτελεσματική ανεύρεση πληροφορίας ή για αποτελεσματική χρήση γραφικών, κειμένου, animation και αλληλεπίδρασης. Γι' αυτό, καταλήγουν, ο παγκόσμιος ιστός είναι διαφορετικός από το κοινό λογισμικό και έτσι δεν μπορεί ούτε να σχεδιαστεί αλλά ούτε και να αξιολογηθεί με τον ίδιο τρόπο [24].

Ως συχνά επιχειρήματα προβάλλεται το γεγονός ότι στον παγκόσμιο ιστό ο χρήστης έχει πλήρη ελευθερία πλοήγησης και δεν περιορίζεται στα στενά πλαίσια μίας εφαρμογής. Επίσης η εκάστοτε σελίδα θεωρείται ως μέρος ενός ευρύτερου χώρου που περιλαμβάνει όχι μόνο τις σελίδες του συγκεκριμένου ιστοχώρου, αλλά όλο τον Παγκόσμιο Ιστό [21]. Στον Παγκόσμιο Ιστό δεν υπάρχουν κανόνες σχεδιασμού (style guidelines) για το επίπεδο διεπικοινωνίας όπως υπάρχουν για όλα τα

λειτουργικά συστήματα, εξασφαλίζοντας ένα ενιαίο και συνεπή τρόπο αλληλεπίδρασης με τις εφαρμογές. Αντί αυτού ο κάθε ιστοχώρος προσφέρει ένα διαφορετικό τρόπο πλοήγησης και αλληλεπίδρασης στον οποίο ο χρήστης καλείται να αντεπεξέλθει.

Από τη φύση του ο Παγκόσμιος Ιστός είναι ένας ανοιχτός χώρος χωρίς περιορισμούς για το είδος του χρήστη που θα επισκεφθεί κάποιον ιστοχώρο με αποτέλεσμα να είναι πολύ πιο δύσκολο να σκιαγραφηθεί το είδος των χρηστών που θα τον επισκεφτούν.

Επίσης πολύ σημαντικός παράγοντας είναι οι διαφορετικές τεχνικές εγκαταστάσεις που διαθέτει ο κάθε χρήστης. Η διαφορά σε ταχύτητες πρόσβασης (56Kbit/sec απλού modem σε σχέση με 768Kbit/sec μιας γραμμής DSL) έχουν ως αποτέλεσμα οι χρήστες να βιώνουν με εντελώς διαφορετικό τρόπο την πλοήγηση στον ίδιο ιστοχώρο [20].

Ένα επίσης ιδιαίτερα σημαντικό χαρακτηριστικό είναι η ευκολία με την οποία μπορεί να προσφερθεί σε κάθε χρήστη μία προσαρμοσμένη σε αυτόν ειδική έκδοση του ίδιου ιστοχώρου (personalization). Ο κάθε χρήστης αντιλαμβάνεται έτσι με διαφορετικό τρόπο τον ίδιο ιστοχώρο επειδή έχει γίνει προσαρμογή του περιεχομένου και του σχεδιασμού ειδικά για αυτόν, είτε με επιλογές του ίδιου χρήστη, είτε με αυτόματες τεχνικές.

Αντίθετα με τα παραπάνω ο Butler στο [70] διαφωνεί και δηλώνει ότι οι ειδικοί στο HCI έχουν αντιμετωπίσει στο παρελθόν παρόμοιες αβεβαιότητες. Η διαμάχη που υπάρχει για το Web Usability απλώς επιβεβαιώνει ότι υπάρχει ακόμα πολύ δουλειά να γίνει στην αξιολόγηση τόσο του παγκόσμιου ιστού όσο και στις τεχνικές που χρησιμοποιούνται στο σχεδιασμό και στην αξιολόγηση [24]. Τα κύρια επιχειρήματα που συνηγορούν υπέρ της ομοιότητας του σχεδιασμού του επιπέδου αλληλεπίδρασης για εφαρμογές ιστοχώρου και παραδοσιακών εφαρμογών αναφέρονται σε σχετική δημοσίευση του Scott Butler [27] που απαντάει σε επιχειρήματα τις άλλης άποψης.

Έχοντας την εμπειρία από τις συνηθισμένες εφαρμογές που αναπτύσσονται για ηλεκτρονικούς υπολογιστές, φαντάζει παράλογο να πιστέψει κανείς ότι μπορεί να εφαρμοστούν οι ίδιες μέθοδοι στην αξιολόγηση ιστοχώρων, όταν δεν υπάρχει δυνατότητα για δημιουργία στοιχείων αλληλεπίδρασης όπως για τα πολύ συνηθισμένα menu. Η δημιουργία παρόμοιας λειτουργικότητας σε σελίδες ιστοχώρου γίνεται δυνατή μόλις τα τελευταία χρόνια με την ευρεία υποστήριξη ανοιχτών προτύπων από τη μεριά των εφαρμογών πλοήγησης (HTML, DOM, CSS, Event Handling) αλλά και πάλι αυτή πρέπει να προγραμματιστεί “με το χέρι” με αποτέλεσμα να αποθαρρύνονται οι προγραμματιστές στη χρήση.

Παρόλα αυτά ακόμα και το προηγούμενο παράδειγμα δεν εμποδίζει στην όμοια προσέγγιση της

χρησιμοποιησιμότητας με παραδοσιακές εφαρμογές. Έτσι στο [27] χρησιμοποιείται το παράδειγμα της Microsoft ως εταιρείας που αναπτύσσει λογισμικό για μεγάλη μερίδα του κοινού. Αναφέρει χαρακτηριστικά ότι τα προγράμματά της χρησιμοποιούνται από ένα πολύ ανομοιογενές σύνολο ανθρώπων και παραμένουν το ίδιο καλά (ή άσχημα) στη χρησιμοποιησιμότητα. Δεν στέκεται δηλαδή εμπόδιο στη χρησιμοποιησιμότητα το γεγονός ότι δεν μπορεί να προβλεφθεί η ακριβής κατηγοριοποίηση των χρηστών που θα χρησιμοποιήσει μία εφαρμογή, όπως συμβαίνει στον Παγκόσμιο Ιστό.

Το ζήτημα που δεν απαντάται όμως είναι εάν οι πόροι που απαιτούνται για τη βελτίωση της χρησιμοποιησιμότητας μπορεί να δαπανώνται με τον ίδιο τρόπο σε εφαρμογές διαδικτύου. Σε ένα απλοϊκό παράδειγμα θα επιχειρηθεί να απαντηθεί το ερώτημα:

Μία εταιρεία διαπιστώνει ότι υπάρχει πρόβλημα χρησιμοποιησιμότητας σε μία κλασική εφαρμογή της. Διορθώνει το πρόβλημα σε μικρό χρονικό διάστημα αλλά επειδή ο τρόπος διανομής της εφαρμογής είναι ιδιαίτερα χρονοβόρος (διανομή πακέτων σε μεταπωλητές ανά τον κόσμο, ειδοποίηση υπαρχόντων πελατών κλπ.) αποφασίζει να καθυστερήσει τη διανομή και να συμπληρώσει τη νέα έκδοση με νέες λειτουργίες. Στο σύνολο της νέας εφαρμογής έχει ήδη δαπανηθεί σημαντικός χρόνος και η εταιρεία πριν την τελική διανομή αποφασίζει να διενεργήσει μελέτη χρησιμοποιησιμότητας.

Μία άλλη εταιρεία διαθέτει εφαρμογή σε ιστοχώρο. Διαπιστώνει ότι υπάρχει πρόβλημα χρησιμοποιησιμότητας το οποίο διορθώνει σε μικρό χρονικό διάστημα. Βρίσκεται σε δίλημμα εάν πρέπει να “βγάλει στον αέρα” τη νέα έκδοση (ζήτημα δευτερολέπτων) ή να διενεργήσει μελέτη χρησιμοποιησιμότητας για τη μικρή αλλαγή που έκανε.

Από το παραπάνω γίνεται εμφανές ότι το δίλημμα που προκύπτει είναι στην ταχύτητα με την οποία μπορούν να διανεμηθούν αλλαγές στις εφαρμογές ιστοχώρου. Στις παραδοσιακές εφαρμογές δεν υπάρχει αυτό το δίλημμα, διότι ούτως ή άλλως οι αλλαγές γίνονται σε άλλες τάξεις μεγέθους του χρόνου. Αλλά ακριβώς επειδή ο παράγοντας χρόνος δεν έχει την ίδια σημασία στο Διαδίκτυο, οι εφαρμογές ιστοχώρου μπορεί να αλλάζουν πολύ συχνά. Και όντως αυτό συμβαίνει.

### **2.2.3 Αξιολόγηση Χρησιμοποιησιμότητας**

Ένα σημαντικό στοιχείο όταν γίνεται αναφορά στη χρησιμοποιησιμότητα είναι ότι αυτή είναι μετρήσιμη. Στο HCI τα πειράματα που μετρούν το πόσο χρησιμοποιήσιμο είναι ένα σύστημα, αποκαλούνται Usability Evaluations ή Usability Tests (θα χρησιμοποιηθεί ο ελληνικός όρος αξιολόγηση στη συνέχεια)..

Συχνά γίνεται σύγχυση με τον όρο Usability Engineering ο οποίος αναφέρεται σε μία μεθοδική προσέγγιση στην ανάπτυξη ενός User Interface. Είναι ένας πρακτικός και συστηματικός τρόπος για να παραδοθεί ένα προϊόν που “δουλεύει” για το χρήστη. Το Usability Engineering περιλαμβάνει πολλές μεθόδους που εφαρμόζονται σε διαφορετικά στάδια της ανάπτυξης και περιλαμβάνουν ανάμεσα σε άλλα τη συλλογή απαιτήσεων, την ανάπτυξη και δοκιμή πρωτοτύπων, την αξιολόγηση εναλλακτικών σεναρίων κ.α.

Η αξιολόγηση χρησιμοποίησιμότητας, από τη μεριά της, είναι μέρος της ευρύτερης έννοιας του Engineering και περιλαμβάνει μεθόδους για τη δοκιμή από πραγματικούς χρήστες ενός συστήματος και τη συστηματική καταγραφή και ανάλυση της συμπεριφοράς τους από ειδικούς του χώρου. Ο σκοπός είναι να εντοπιστούν προβλήματα στη χρησιμοποίησιμότητα και να προταθούν λύσεις.

### **2.2.3.1 Ο Παράγοντας Χρήστης**

Σε κάθε περίπτωση αξιολόγησης χρησιμοποίησιμότητας ο παράγοντας χρήστης είναι ίσως ο πιο σημαντικός. Οι διαφορές που υπάρχουν ανάμεσα στους χρήστες μπορεί να αποβούν καθοριστικές στα αποτελέσματα της μελέτης χρησιμοποίησιμότητας. Ο Nielsen χαρακτηρίζει στο [67] σελ. 43 τη γνώση γύρω από τους χρήστες έναν από τους σημαντικότερους παράγοντες στη μελέτη χρησιμοποίησιμότητας και αναφέρει χαρακτηριστικά ότι σε μελέτη για τη χρησιμοποίησιμότητα συστημάτων hypertext βρέθηκε ότι 4 από τα 10 πιο κρίσιμα προβλήματα των συστημάτων είχαν ως πηγή τις διαφορές ανάμεσα στους χρήστες. Αυτό βέβαια δε σημαίνει ότι δεν μπορούν να αναπτυχθούν συστήματα που να καλύπτουν τις ανάγκες μεγάλης γκάμας χρηστών.

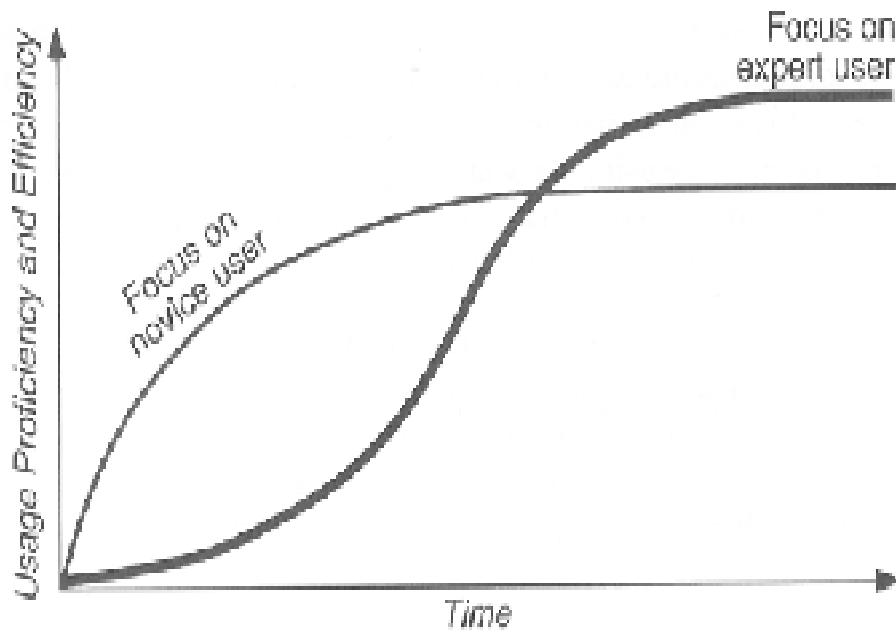
### **2.2.3.2 Μέτρηση Παραμέτρων**

Η μετρησιμότητα της χρησιμοποίησιμότητας, όπως αναφέρθηκε, είναι ουσιώδες στοιχείο στην όλη διαδικασία. Παρακάτω περιγράφονται προσεγγίσεις για τη μέτρηση συστατικών χρησιμοποίησιμότητας που φαίνονται στο Σχήμα 2.3.

#### **Learnability**

Η ικανότητα εκμάθησης ενός συστήματος είναι σημαντικό στοιχείο στη χρησιμοποίησιμότητα. Η πρώτη εμπειρία που αποκτάται από ένα σύστημα είναι στην προσπάθεια εκμάθησής του, και γι' αυτό πρέπει να είναι εύκολα στην εκμάθηση. Για συστήματα που πρόκειται να χρησιμοποιηθούν για μεγάλο χρονικό διάστημα ή και καθημερινά (π.χ. οι υπάλληλοι ενός ασφαλιστικού οργανισμού που περνούν στοιχεία ατυχημάτων σε σχετική εφαρμογή δουλεύουν καθημερινά με το σύστημα) μπορεί να προβλεφθεί σημαντικός χρόνος για την εκμάθηση τους. Ειδικά στο Διαδίκτυο όπου η πρόσβαση στις εφαρμογές γίνεται πολύ εύκολα (με την επιλογή ενός συνδέσμου) και επίσης εύκολα μπορεί να φύγουν από την εφαρμογή εάν για οποιοδήποτε λόγο δυσκολευτούν στη χρήση

τους. Παρόμοια χαρακτηριστικά παρουσιάζουν εφαρμογές που από τη φύση τους χρησιμοποιούνται μόνο μια φορά (π.χ. ένα πληροφοριακό σύστημα σε ένα μουσείο). Τέτοιου είδους συστήματα ο Nielsen στο [67] τα αποκαλεί “walk up and use systems” και πρέπει να έχουν ιδιαίτερα μικρό χρόνο εκμάθησης για να μην αποθαρρυνθούν οι χρήστες.



Σχήμα 2.4 Εκμάθηση εφαρμογής για αρχάριους και έμπειρους χρήστες [67] σελ. 28

Στο σχήμα Σχήμα 2.4 παρουσιάζονται δύο υποθετικές συναρτήσεις της ικανότητας και αποδοτικότητας χρήσης μίας εφαρμογής ανάλογα με το χρόνο. Στην πρώτη (που απευθύνεται σε αρχάριους χρήστες) σχετικά σύντομα οι χρήστες μπορούν να χρησιμοποιήσουν το σύστημα αλλά δε φτάνουν ποτέ σε πολύ ψηλά επίπεδα απόδοσης. Στη δεύτερη η έμφαση δίνεται σε έμπειρους χρήστες οι οποίοι θα αργήσουν να μάθουν την εφαρμογή (λόγω της πολυπλοκότητας και των προχωρημένων δυνατοτήτων) αλλά θα αποκτήσουν την ικανότητα για πολύ αποδοτική χρήση της.

Ένα απλός τρόπος μέτρησης της ικανότητας εκμάθησης είναι να καθοριστούν κάποιες λειτουργίες που είναι επιθυμητό να μπορούν να χρησιμοποιήσουν οι χρήστες και να μετρηθεί ο χρόνος που απαιτείται από χρήστες που δεν έχουν χρησιμοποιήσει στο παρελθόν την εφαρμογή, να φτάσουν σε αυτό το επίπεδο. Ένας άλλος τρόπος είναι να επαναλαμβάνουν οι χρήστες κάποια συγκεκριμένη εργασία και να διαπιστωθεί εάν βελτιώνεται ο χρόνος ολοκλήρωσης της εργασίας στις επαναλήψεις.

Έχοντας υπ' όψιν την εικόνα που παρουσιάζει το Σχήμα 2.4 θα μπορούσε να σκεφτεί ότι δεν είναι δυνατόν να αναπτυχθούν συστήματα που να απευθύνονται τόσο σε αρχάριους όσο και σε έμπειρους χρήστες. Στην πράξη όμως υπάρχουν αρκετές δυνατότητες για αντιστάθμιση των όποιων αρνητικών

στοιχείων της μίας ή της άλλης ομάδας χρηστών. Ένα γνωστό παράδειγμα είναι η χρήση “shortcuts” που προσφέρουν γρήγορη κλήση μιας λειτουργίας μέσω συνδυασμό πλήκτρων, αντί της πιο αργής αλλά και πιο προσιτής χρήσης ιεραρχίας menu.

### **Efficiency of Use**

Η αποδοτικότητα αναφέρεται σε έμπειρους χρήστες και μετριέται πάνω σε αυτούς. Οι χρήστες χαρακτηρίζονται έμπειροι όταν έχουν χρησιμοποιήσει για μεγάλο χρονικό διάστημα ένα σύστημα ή εάν το δηλώσουν από μόνοι τους. Ένας πιο τυπικός τρόπος χαρακτηρισμού τους είναι με βάση το Σχήμα 2.4 όταν δηλαδή παρατηρείται ότι η απόδοσή τους δεν αυξάνει πλέον.

Για τη μέτρηση απλώς επιλέγονται έμπειροι χρήστες και μετριέται ο χρόνος ολοκλήρωσης κάποιων προκαθορισμένων εργασιών.

### **Memorability**

Το σύστημα πρέπει να είναι σχεδιασμένο κατά τέτοιο τρόπο ώστε ο χρήστης να μπορεί να θυμάται τη χρήση του. Δηλαδή κατά πόσο ένας χρήστης που έχει χρησιμοποιήσει το σύστημα για αρκετό χρονικό διάστημα, είναι σε θέση να το χρησιμοποιήσει μετά από μία διακοπή χρήσης χωρίς να χρειαστεί να μάθει τη λειτουργία της εφαρμογής από την αρχή.

Η αναφορά γίνεται περισσότερο στους περιστασιακούς χρήστες, δηλαδή στους χρήστες που δεν είναι ούτε αρχάριοι αλλά ούτε έμπειροι, έχουν όμως χρησιμοποιήσει το σύστημα για αρκετό καιρό στο παρελθόν.

Ένας τρόπος για τη μέτρηση είναι να αφηθεί ένας χρήστης, που έχει χρησιμοποιήσει το σύστημα για μακρύ χρονικό διάστημα, μακριά από αυτό για ένα διάστημα, και στη συνέχεια να κληθεί να πραγματοποιήσει μια σειρά από εργασίες όπου συγκρίνεται ο χρόνος ολοκλήρωσης πριν και μετά τη διακοπή χρήσης.

### **Few and Noncatastrophic Errors**

Ως λάθη θεωρούνται όλες οι ενέργειες που δεν έχουν το επιθυμητό (από το χρήστη) αποτέλεσμα. Ο σκοπός είναι να ελαχιστοποιηθούν αυτά τα λάθη. Η μέτρηση των λαθών μπορεί να γίνει στα πλαίσια άλλων πειραμάτων, συγκρίνοντας τον αριθμό λαθών που γίνονται σε συγκεκριμένες εργασίες. Τα λάθη μπορεί να χωριστούν σε δύο κατηγορίες: λάθη που γίνονται εμφανή με την πραγματοποίησή τους και μπορεί να διορθωθούν αμέσως, και λάθη που δε γίνονται αντιληπτά και οδηγούν σε λάθος αποτελέσματα ή λάθη που καταστρέφουν την εργασία του χρήστη. Ενώ η πρώτη κατηγορία λαθών μπορεί να θεωρηθεί ότι συμπεριλαμβάνεται στη μέτρηση της απόδοσης του συστήματος (χρόνος που απαιτείται για τη διόρθωση των λαθών), η δεύτερη κατηγορία είναι πιο

σημαντική και πρέπει να μετριέται ξεχωριστά.

### **Subjective Satisfaction**

Το σύστημα πρέπει να είναι σχεδιασμένο κατά τέτοιο τρόπο ώστε να είναι ευχάριστο στο χρήστη. Σε εφαρμογές όπου κύριο ή σημαντικό χαρακτηριστικό είναι η διασκέδαση του χρήστη (παιχνίδια, εκπαιδευτικές εφαρμογές για μικρά παιδιά κλπ..) αυτό το χαρακτηριστικό μπορεί να έχει μεγαλύτερη σημασία. Οι χρήστες τέτοιων συστημάτων ενδιαφέρονται κυρίως για την διασκέδαση παρά για την ταχύτητα με την οποία διεκπεραιώνονται οι διεργασίες. Για να μετρηθεί αυτό το χαρακτηριστικό το πιο απλό είναι οι χρήστες να δηλώσουν αν το σύστημα τούς ήταν ευχάριστο (π.χ. με μικρά ερωτηματολόγια μετά το πέρας των δοκιμών με χρήστες ή σε χρήστες από ήδη υφιστάμενες εγκαταστάσεις) και να υπολογιστεί ο μέσος όρος των απαντήσεων.

#### **2.2.3.3 Μέθοδοι αξιολόγησης**

Οι μέθοδοι και τα συστήματα για την αξιολόγηση της χρησιμοποίησιμότητας μπορεί να χωριστούν σε 4 διαφορετικές κατηγορίες. Στο [19] οι 4 κατηγορίες συνοψίζονται όπως φαίνεται παρακάτω. Για μία πιο αναλυτική εισαγωγή ο αναγνώστης παραπέμπεται στο [19] και [67] ενώ για την παρούσα εργασία θα γίνει πιο αναλυτική περιγραφή ορισμένων τεχνικών παρακάτω.

#### **Αυτόματη Αξιολόγηση**

Στην περίπτωση αυτή υπάρχει ένα σύστημα αξιολόγησης εφαρμογών, το οποίο εξάγει συμπεράσματα για την υπό εξέταση εφαρμογή. Το σύστημα αυτό μπορεί να είναι εντελώς αυτοματοποιημένο ή ημιαυτοποιημένο. Δεν υπάρχουν τέτοια γενικά συστήματα που να μπορούν να βγάλουν αξιόπιστα αποτελέσματα.

#### **Εμπειρική Αξιολόγηση**

Στην περίπτωση αυτή ένας προκαθορισμένος αριθμός χρηστών χρησιμοποιεί το σύστημα διεκπεραιώνοντας συγκεκριμένες λειτουργίες. Κατά τη διάρκεια του πειράματος λαμβάνονται μετρήσεις και γίνονται παρατηρήσεις από ειδικούς, και στη συνέχεια συντάσσονται αναφορές που καταγράφονται τα συμπεράσματα. Αυτές οι τεχνικές είναι ευρέως διαδεδομένες..

#### **Αναλυτική**

Χρησιμοποιούνται αναλυτικά ιδεατά μοντέλα της εφαρμογής και των χρηστών από τα οποία υπολογίζονται μετρικές που βοηθούν στην αξιολόγηση του συστήματος. Τα σύγχρονα αλληλεπιδραστικά συστήματα λόγω της πολυπλοκότητάς τους είναι σχεδόν αδύνατο να περιγραφούν από αναλυτικά μοντέλα, για αυτό το λόγο αυτές οι τεχνικές και δεν έχουν μεγάλη εφαρμογή.



## Ευριστική Αξιολόγηση

Βασίζεται στη χρησιμοποίηση γενικών ευριστικών κανόνων, γενικών οδηγιών, αλλά κυρίως στην ικανότητα, τη γνώση και την εμπειρία ειδικών οι οποίοι διενεργούν τον έλεγχο. Χρησιμοποιούνται ευρύτατα, εντοπίζουν λάθη που δεν εντοπίζονται από απλούς χρήστες και μειώνουν τον αριθμό των χρηστών που εμπλέκονται στην διαδικασία ελέγχου. Η επιτυχία αυτών των τεχνικών εξαρτάται σε μεγάλο βαθμό από την εμπειρία των ειδικών.

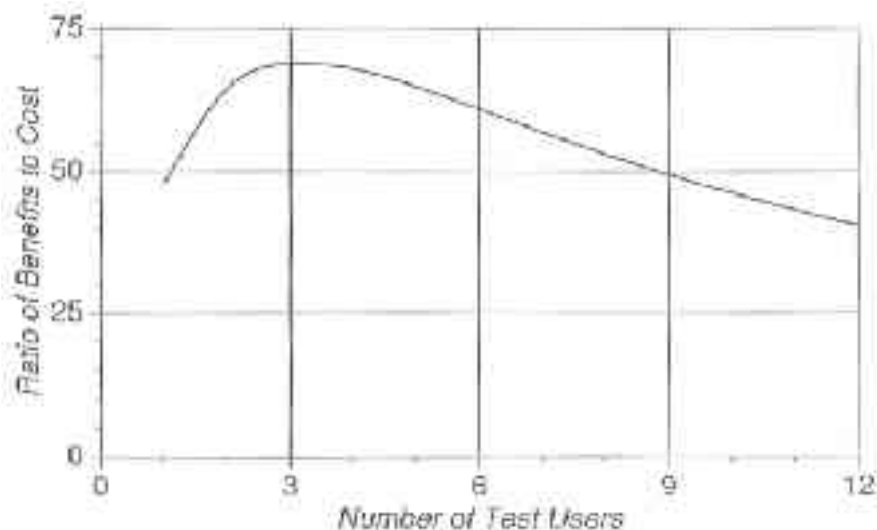
### 2.2.3.3.1 “Πειράματα” με Χρήστες

Ο τίτλος “πειράματα χρηστών” προέρχεται από τη μετάφραση του “User Testing”. Ακόμα και στην αγγλική του έκδοση η ονομασία μπορεί να παρερμηνευτεί διότι βεβαίως δε γίνεται κάποιου είδους έλεγχος του χρήστη αλλά εκτελείται μία διαδικασία με χρήστες.

Από πολλές απόψεις τα πειράματα χρηστών είναι θεμελιώδες στην αξιολόγηση χρησιμοποίησιμότητας και σε κάποιο βαθμό είναι αναντικατάστατα. Προσφέρει άμεση πληροφορία για το πώς ο κόσμος χρησιμοποιεί υπολογιστές και για τα ακριβή προβλήματα που συναντούν σε μία συγκεκριμένη εφαρμογή.

Προσοχή απαιτείται σε θέματα αξιοπιστίας αποτελεσμάτων (θα υπήρχαν τα ίδια αποτελέσματα εάν επαναλαμβανόταν το πείραμα;) και εγκυρότητας αυτών (τα αποτελέσματα όντως αντανακλούν τα ζητήματα χρησιμοποίησιμότητας που εξετάζονται;). Η αβεβαιότητα αυτή έγκειται στις μεγάλες διαφορές που υπάρχουν ανάμεσα στους χρήστες και στις μεγάλες αποκλίσεις στα μεγέθη των αποτελεσμάτων που παρατηρούνται. Έρευνες έχουν δείξει ότι η μέτρηση των λαθών έχει τις μεγαλύτερες διακυμάνσεις και απαιτεί έτσι μεγαλύτερο αριθμό χρηστών για την αξιοπιστία στα αποτελέσματα. Εάν τα αποτελέσματα που προκύπτουν από τα πειράματα είναι έγκυρα και στον πραγματικό κόσμο (πραγματικοί χρήστες, εκτός εργαστηρίου κλπ.) εξαρτάται πολύ από τη σύνθεση των χρηστών (είναι τα χαρακτηριστικά τους όμοια με τους πραγματικούς χρήστες;) αλλά και από τις συγκεκριμένες εργασίες πάνω στις οποίες γίνονται οι μετρήσεις ή/και παρατηρήσεις.

Η μέθοδος των πειραμάτων χρηστών περιλαμβάνει την ανεξάρτητη χρήση μιας εφαρμογής από ομάδα χρηστών. Οι χρήστες καλούνται να διεκπεραιώσουν ένα σύνολο προκαθορισμένων εργασιών. Κατά τη διάρκεια της χρήσης του συστήματος παρακολουθούνται διακριτικά από ειδικούς οι οποίοι κρατούν σημειώσεις ή κάνουν μετρήσεις. Τα αποτελέσματα των μετρήσεων ή παρατηρήσεων αναλύονται και εξάγονται συμπεράσματα.



Σχήμα 2.5 Σχέση αναλογίας κερδών/κόστους με αριθμό χρηστών σε πειράματα χρηστών [67] σελ. 174

Παραλλαγές υπάρχουν πολλές όπως “thinking aloud” όπου ο χρήστης κατά τη διάρκεια της χρήσης “σκέφτεται δυνατά” με αποτέλεσμα οι παρατηρητές να αποκτούν μία άποψη για τον τρόπο σκέψης των χρηστών. Επίσης μπορεί να συνδυαστεί η χρήση του συστήματος με την ακόλουθη συμπλήρωση ερωτηματολογίου (π.χ. για εκτίμηση της υποκειμενικής ικανοποίησης του χρήστη απέναντι στην εφαρμογή). Επίσης μπορεί να γίνει καταγραφή σε video του χρήστη με σκοπό την περαιτέρω ανάλυση κ.α.

Για μεγαλύτερη εμβάθυνση στη μέθοδο των πειραμάτων χρηστών ο αναγνώστης παραπέμπεται στο [67].

### 2.2.3.3.2 Ευριστικοί Κανόνες και Αξιολόγηση

Οι ευριστικοί κανόνες στηρίζονται βέβαια σε πραγματική και εμπειριστατωμένη έρευνα αλλά η εφαρμογή τους είναι περισσότερο θέμα εμπειρίας και συστηματικής ενασχόλησης με το αντικείμενο της χρησιμοποιησιμότητας. Παρόλα αυτά ακόμα και σχετικά άπειροι προγραμματιστές είναι σε θέση ακολουθώντας βασικούς κανόνες να αναπτύξουν “καλά” επίπεδα αλληλεπίδρασης των εφαρμογών.

Υπάρχουν επίσης κανόνες και συλλογές κανόνων που σχετίζονται με τα εκάστοτε λειτουργικά συστήματα (style guides) και έχουν ως σκοπό να παρουσιάσουν στο χρήστη όμοιες όψεις και λειτουργικότητες για τις εφαρμογές του ίδιου λειτουργικού συστήματος. Ένας τέτοιος κανόνας για παράδειγμα είναι ότι τα παράθυρα των εφαρμογών για το λειτουργικό σύστημα της Microsoft έχουν στην επάνω δεξιά γωνία ένα κουμπί ορισμένου σχήματος με το οποίο “κλείνει” το παράθυρο ή η εφαρμογή.

Άλλοι κανόνες έχουν σχέση με τις φυσικές ικανότητες του ανθρώπου και αναφέρονται για παράδειγμα στους συνδυασμούς χρωμάτων που είναι ή δεν είναι ενδεδειγμένοι για την εμφάνιση πληροφορίας στις οθόνες των υπολογιστών.

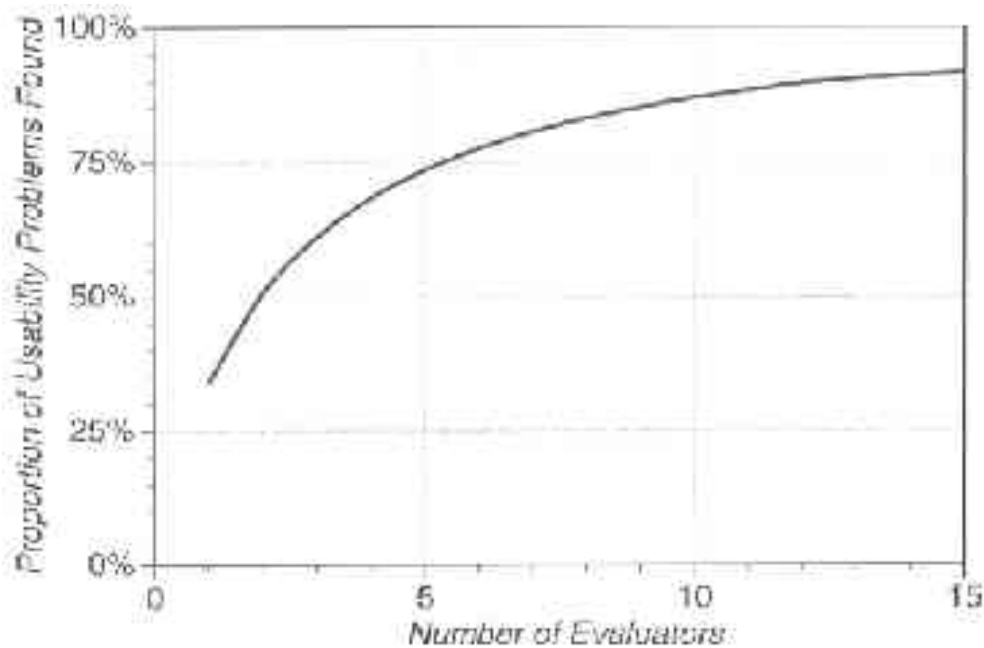
Ο Nielsen στο [67] σελ. 20 παρουσιάζει 10 σημαντικούς κανόνες που πρέπει να εφαρμόζονται από όλους τους σχεδιαστές εφαρμογών. Οι κανόνες περιλαμβάνουν θέματα όπως τη γλώσσα που χρησιμοποιείται για τα μηνύματα της εφαρμογής (να είναι γραμμένα με τη λογική που ακολουθεί ο χρήστης και όχι ο προγραμματιστής), θέματα συνέπειας σε μηνύματα ή τοποθεσία αλληλεπιδραστικών στοιχείων κ.α.

Συνήθως τέτοιου είδους ευριστική αξιολόγηση γίνεται έτσι κι αλλιώς από τους προγραμματιστές και σχεδιαστές μιας εφαρμογής. Δηλαδή κοιτώντας και χρησιμοποιώντας την εφαρμογή ακολουθούν κάποιους κανόνες από την εμπειρία τους.

Η ευριστική αξιολόγηση όπως νοείται στην παρούσα εργασία όμως ακολουθεί μία πιο συστηματική και αυστηρή μέθοδο. Σκοπός είναι να εντοπιστούν προβλήματα χρησιμοποίησιμότητας και να αντιμετωπιστούν σε μία επαναληπτική διαδικασία ανάπτυξης της εφαρμογής.

Στη μέθοδο αυτή μία σχετικά μικρή ομάδα αξιολογητών χρησιμοποιούν ανεξάρτητα μεταξύ τους την εφαρμογή και κρίνουν εάν είναι συμβατή με γνωστές σε αυτούς αρχές χρησιμοποίησιμότητας (τους ευριστικούς κανόνες δηλαδή). Προσπερνούν περισσότερο της μίας φορές όλο το επίπεδο διεπικοινωνίας της εφαρμογής.

Στο σχήμα Σχήμα 2.6 φαίνεται το ποσοστό των προβλημάτων που συνήθως ανακαλύπτονται ανάλογα με το πλήθος των αξιολογητών. Από το σχήμα συμπεραίνεται ότι 5 αξιολογητές είναι ικανός αριθμός για να εντοπιστούν τα πιο σημαντικά προβλήματα.



Σχήμα 2.6 Σχέση ανάμεσα σε αριθμό αξιολογητών και ποσοστού προβλημάτων χρησιμοποιησιμότητας που εντοπίζονται [67] σελ. 156

Η συλλογή των αποτελεσμάτων μπορεί να γίνει είτε με αναφορές από τους αξιολογητές είτε έχοντας κάποιον παρατηρητή την ώρα της αξιολόγησης. Οι αξιολογητές μπορεί μετά το πέρας της χρήσης της εφαρμογής να βρεθούν και να ανταλλάξουν τα συμπεράσματα στα οποία κατέληξαν.

Στην περίπτωση που υπάρχει παρατηρητής, η συμπεριφορά του πρέπει να είναι σαφώς πιο ανοιχτή σε προβλήματα των αξιολογητών, σε αντίθεση με τα “πειράματα” χρηστών όπου ο παρατηρητής μένει αμέτοχος. Σε αντίθεση επίσης με τα πειράματα χρηστών, ο παρατηρητής δεν έχει σκοπό να αξιολογήσει τη συμπεριφορά και τις ενέργειες των αξιολογητών διότι αυτή η αξιολόγηση έρχεται ήδη από τους ίδιους τους χρήστες.

Το είδος των αξιολογητών έχει βρεθεί να επηρεάζει σε μεγάλο βαθμό την ποιότητα και την ποσότητα των προβλημάτων που εντοπίζονται. Ειδικοί σε usability θα εντοπίσουν πολύ περισσότερα προβλήματα από αρχάριους χρήστες οπότε είναι προτιμότερο για τέτοιου είδους μεθόδους να επιλεγούν έμπειροι χρήστες.

### 2.2.3.3 Καταγραφή Πραγματικής Χρήσης

Σε αυτή τη μέθοδο καταγράφεται η πραγματική χρήση μιας εφαρμογής. Για παράδειγμα καταγράφονται στατιστικά στοιχεία με συχνότητες χρήσης των λειτουργιών της εφαρμογής ή λαθών που γίνονται. Βλέποντας για παράδειγμα ότι κάποια λειτουργία χρησιμοποιείται πολύ σπάνια, μπορεί να αναζητηθούν οι αιτίες στο οποίο οφείλεται (μπορεί να μην χρειάζεται η λειτουργία ή να μην είναι αρκετά εμφανής ο τρόπος ενεργοποίησής της).

Η καταγραφή μπορεί να γίνει είτε από τα μηνύματα χαμηλού επιπέδου του λειτουργικού

συστήματος είτε άμεσα από την ίδια την εφαρμογή. Η δεύτερη περίπτωση έχει το πλεονέκτημα ότι τα δεδομένα που καταγράφονται μπορεί να είναι υψηλού επιπέδου έννοιες που είναι πιο εύχρηστες στην ανάλυση.

#### **2.2.4 Προσέγγιση Web Usage Analysis**

Στην προηγούμενη ενότητα για τις μεθόδους ανάλυσης χρησιμοποισιμότητας αναφέρθηκε ήδη ότι μία μέθοδος βασίζεται στην καταγραφή της πραγματικής χρήσης μίας εφαρμογής και την ανάλυση των δεδομένων που προκύπτουν από αυτήν.

Στα πλαίσια της ανάλυσης χρησιμοποισιμότητας ιστοχώρων αυτή η μέθοδος αποκτάει νέα διάσταση επειδή προσφέρει μοναδικά πλεονεκτήματα που την ξεχωρίζουν από τις υπόλοιπες μεθόδους. Τα πλεονεκτήματα έχουν σχέση με τη φύση των ιστοχώρων που εισάγει αρκετές αβεβαιότητες για τις άλλες μεθόδους.

Αναφέρθηκε ήδη ότι ένα χαρακτηριστικό των εφαρμογών στον παγκόσμιο ιστό είναι ότι δεν υπάρχει σαφής άποψη για τα χαρακτηριστικά των χρηστών που τη χρησιμοποιούν. Ο παγκόσμιος ιστός προσφέρει απεριόριστη ευκολία για την επίσκεψη ενός ιστοχώρου που καθιστά αμφίβολη την ομοιογένεια, τις κατηγορίες των χρηστών και τους στόχους που αυτοί έχουν επισκεπτόμενοι έναν ιστοχώρο.

Ο παγκόσμιος ιστός είναι επίσης ευμετάβλητος στο χρόνο από πολλές απόψεις. Νέες εκδόσεις εφαρμογών πλοήγησης κάνουν εφικτή τη χρήση τεχνολογιών για την αλληλεπίδραση με το χρήστη που πριν φάνταζαν απίθανες ή πολύ δύσκολα υλοποιήσιμες. Οι αλλαγές στον ιστοχώρο με χρήση διαφορετικών τεχνικών στην αλληλεπίδραση μπορεί να είναι πολύ συχνές και χαρακτηρίζονται από στοιχεία πειραματισμού που καθιστά αναγκαία τη συνεχή παρακολούθηση της ανταπόκρισης των χρηστών στα νέα δεδομένα.

Επίσης οι σελίδες συχνά δημιουργούνται με δυναμικό τρόπο από βάσεις δεδομένων ή άλλα μέσα, με πληροφορία που μπορεί να αλλάζει συχνά. Ακριβώς επειδή υπάρχει η δυνατότητα σχετικά εύκολης αλλαγής του περιεχομένου αλλά και του τρόπου εμφάνισης ενός ιστοχώρου, θα ήταν επιθυμητό να ελεγχθούν πολλές εκδόσεις του ιστοχώρου ως προς την χρησιμοποισιμότητα.

Επίσης μπορούν μικρές αλλαγές να επιφέρουν ασύγκριτα διαφορετικά αποτελέσματα στους χρήστες ενός ιστοχώρου. Για παράδειγμα μπορεί ένας απλός σύνδεσμος από έναν ιστοχώρο, με μεγάλο αριθμό επισκέψεων, σε έναν άλλο ιστοχώρο να βγάλει το δεύτερο από την αφάνεια και να εκτοξεύσει τις επισκέψεις και τη χρήση του σε πολύ υψηλά επίπεδα. Αυτή η αλλαγή μπορεί να συνδυάζεται με εντελώς διαφορετικά χαρακτηριστικά των χρηστών για παράδειγμα σε όρους

πληροφορίας για την οποία ενδιαφέρονται.

Άλλο χαρακτηριστικό ορισμένων ιστοχώρων είναι ότι ο ίδιος ιστοχώρος μπορεί να επισκέπτεται περιοδικά από διαφορετικούς χρήστες στη διάρκεια του χρόνου. Για παράδειγμα διαφορετικούς επισκέπτες έχει ένας ιστοχώρος τουριστικού προορισμού για την Κρήτη πριν τους καλοκαιρινούς μήνες σε σχέση με την περίοδο του χειμώνα.

Τέλος, όπως αναφέρθηκε και σε προηγούμενη ενότητα, οι ιστοχώροι έχουν τη δυνατότητα να προσφέρουν στους χρήστες προσωποποιημένη όψη και υπηρεσίες, ανάλογα με τις προτιμήσεις του κάθε χρήστη.

Όλα τα παραπάνω χαρακτηριστικά καταστούν αναγκαία τη χρήση μίας μεθόδου για την αξιολόγηση χρησιμοποιησιμότητας η οποία θα είναι αρκετά ευέλικτη ώστε να μπορεί συνεχώς να εφαρμόζεται στον υπό μελέτη ιστοχώρο και να μπορεί να ανταποκρίνεται στην ποικιλομορφία που αυτός παρουσιάζει.

Οι “παραδοσιακές” μέθοδοι που χρησιμοποιούνται είναι αρκετά χρονοβόρες σε όρους προετοιμασίας και διενέργειας για να μπορέσουν να καλύψουν από μόνες τους αυτές τις ανάγκες. Σε αυτό το σημείο η ανάλυση με την καταγραφή της συμπεριφοράς των χρηστών μπορεί να καλύψει τις απαιτήσεις που υπάρχουν. Σε συνδυασμό με επιπλέον στοιχεία που είναι διαθέσιμα όταν η καταγραφή γίνεται σε ιστοχώρους αποκτάει ιδιαίτερο ενδιαφέρον για περαιτέρω μελέτη. Για παράδειγμα μπορεί να εξαχθούν στοιχεία για την προέλευση των χρηστών (χώρα, περιοχή) ή για το είδος της πληροφορίας που αναζητούν (από φράσεις αναζήτησης σε μηχανές αναζήτησης που οδήγησαν στον ιστοχώρο).

Η ονομασία που χρησιμοποιείται για αναλύσεις στον παγκόσμιο ιστό είναι Web Analysis ή Web Mining και ειδικά για τη συμπεριφορά των χρηστών Web Usage Analysis ή Web Usage Mining. Στη συνέχεια θα χρησιμοποιηθεί ο ελληνικός όρος Ανάλυση Επισκεψιμότητας.

Πρόκειται για ένα σχετικά νέο επιστημονικό πεδίο (όπως είναι εξάλλου και ο παγκόσμιος ιστός) με πρόσφατες δημοσιεύσεις και σχετικά λίγα παραδείγματα εφαρμογής των τεχνικών που αναπτύσσονται.

Ακριβώς επειδή η έρευνα σε θέματα Web Usage Analysis παρουσιάζει ξεχωριστό ενδιαφέρον, το θέμα έχει αναπτυχθεί με πιο εκτενή τρόπο στο Παράρτημα Β.

## **2.3 Web Usage Analysis/Mining**

Ο Παγκόσμιος Ιστός (World Wide Web) συνεχίζει την εκρηκτική επέκτασή του ως μέσον για την

διανομή πληροφορίας και τη διενέργεια συναλλαγών [16]. Η δυνατότητα εις βάθος ανάλυσης της χρήσης των ιστοχώρων παραμένει ένα ζητούμενο για τους διαχειριστές και σχεδιαστές των ιστοχώρων.

Η σχεδίαση ενός ιστοχώρου επικεντρώνεται στην οργάνωση της πληροφορίας σε κάθε σελίδα και τους υπερσυνδέσμους μεταξύ των σελίδων με σκοπό την εξυπηρέτηση των αναγκών των χρηστών. Για σχετικά μικρούς ιστοχώρους κάποιες απλοϊκές και άμεσες στατιστικές μπορεί να είναι αρκετές για την επαλήθευση της συμπεριφοράς των χρηστών. Αλλά όσο η πολυπλοκότητα και το μέγεθος αυξάνονται απαιτούνται πιο εξελιγμένες τεχνικές για την ανάλυση.

Στις επόμενες ενότητες θα περιγραφεί αρχικά ο όρος Web Mining για την ανάλυση ιστοχώρων και ειδικότερα την ανάλυση συμπεριφοράς των χρηστών. Θα ακολουθήσει μία περιγραφή για τα στάδια της ανάλυσης και για κάθε ένα από αυτό θα περιγραφούν αναλυτικά διάφορες τεχνικές που χρησιμοποιούνται για την εξαγωγή συμπερασμάτων.

### **2.3.1 Τι είναι Web Mining**

Το Web Usage Analysis ή Web Usage Mining είναι η εφαρμογή τεχνικών data mining σε μεγάλες αποθήκες δεδομένων για τον παγκόσμιο ιστό (Web data repositories). Προσφέρει αποτελεσματικές τεχνικές για την ανάλυση της συμπεριφοράς των χρηστών σε ιστοχώρους για την ανακάλυψη συχνά εμφανιζόμενων σχημάτων, κανόνων συσχετισμού (association rules), συνόλων ομοειδών σελίδων και χρηστών (page and user clusters) και για την ανάλυση μονοπατιών (path analysis).

Το Web Usage Mining ξεχωρίζει από το Web Content Mining στο ότι το δεύτερο ασχολείται με την πληροφορία (το περιεχόμενο) που παρέχεται στον παγκόσμιο ιστό και ξεχωρίζει από το Web Structure Mining στο ότι αυτό ασχολείται με τη δομή του παγκόσμιου ιστού.

Επειδή ο ίδιος ο παγκόσμιος ιστός είναι μία σχετικά νέα τεχνολογία, είναι φυσιολογικό η έρευνα σε θέματα ανάλυσης επισκεψιμότητας να βρίσκεται ακόμα σε εμβρυακό στάδιο. Η έρευνα σε θέματα Web Mining είναι ένα συγκλίνων επιστημονικό πεδίο από διάφορες επιστημονικές κοινότητες, όπως Βάσεις Δεδομένων, Information Retrieval, Artificial Intelligence και Machine Learning [2]. Στο [3], που είναι μία δημοσίευση του 2001, αναφέρεται ότι το “State of the Art σε θέματα Web Mining βρίσκεται ακόμα σε αρχέγονο στάδιο”. Οι περισσότερες δημοσιεύσεις έγιναν το δεύτερο μισό της δεκαετίας του 90, ενώ το μοναδικό βιβλίο που υπήρχε μέχρι πρόσφατα είναι το [40] από το 1999. Από τις πρώτες εργασίες που έδωσαν ένα survey με το διαχωρισμό σε Web Content και Web Usage Mining αναφέρεται το [15] του 1997, ενώ ο όρος Web Mining αναφέρεται πρώτη φορά σε εργασία του Etzioni το 1996 ([2] σελ. 9).

Επειδή το θέμα παρουσιάζει πολύ μεγάλο οικονομικό ενδιαφέρον, οι εταιρείες που ασχολούνται με το αντικείμενο είναι μάλλον φειδωλές στη δημοσίευση της ερευνητικής τους δουλειάς και περιορίζονται περισσότερο σε White Papers με γενικά χαρακτηριστικά για τα προϊόντα τους. Είναι επίσης χαρακτηριστικό ότι αποτελέσματα ερευνητικής δουλειάς, πολύ συχνά εμφανίζονται σύντομα ενσωματωμένα σαν προϊόντα νεοσύστατων εταιρειών.

Για μία ενημερωμένη συλλογή με σχετικές δημοσιεύσεις ο αναγνώστης παραπέμπεται στο

<http://www.cs.ualberta.ca/~wum/> .

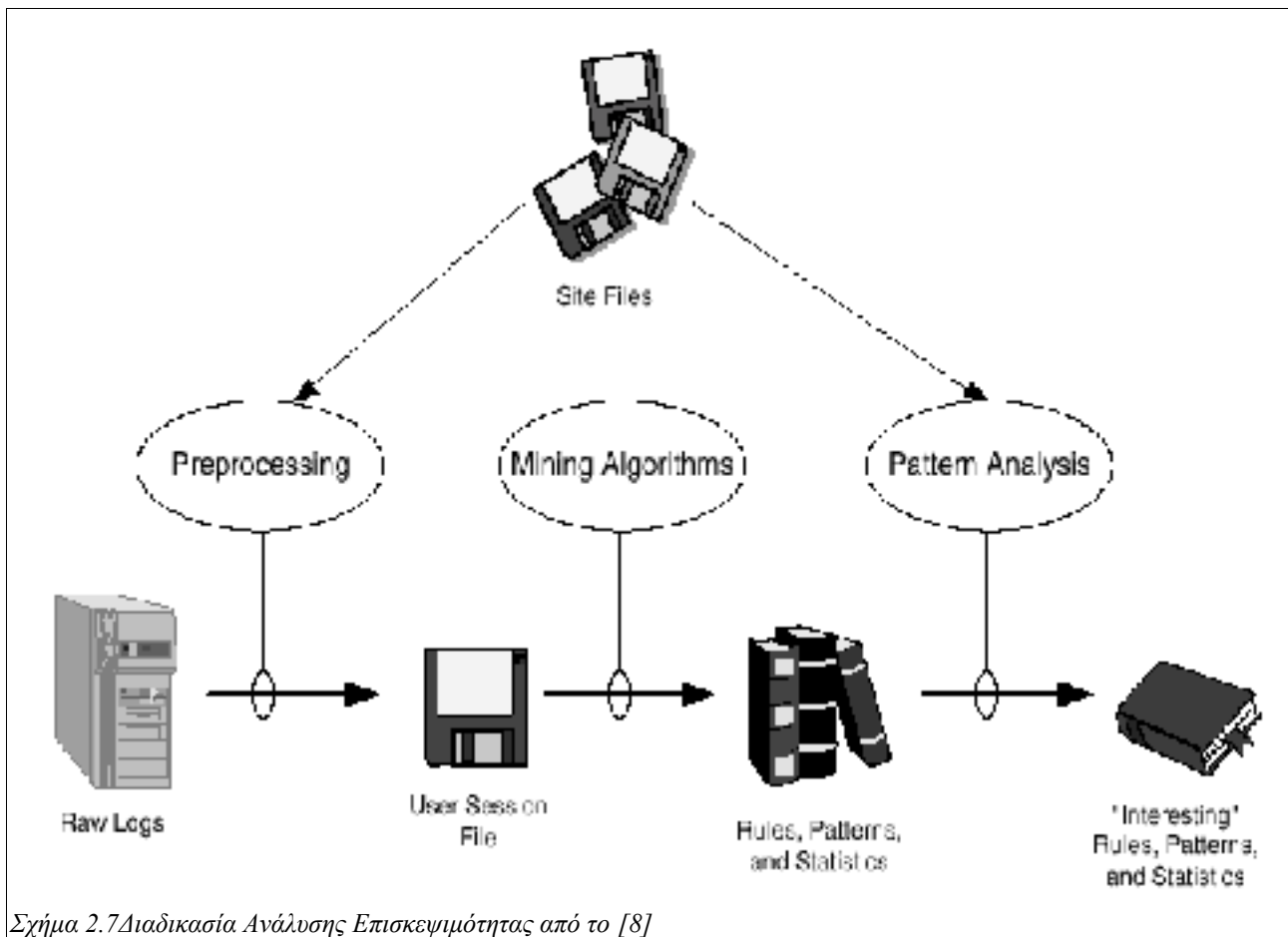
## **2.3.2 Ανάλυση και Αξιολόγηση Επισκεψιμότητας**

### **2.3.2.1 Στάδια Ανάλυσης**

Η ανάλυση επισκεψιμότητας δεν είναι μία μονοσήμαντη ή στερεότυπη διαδικασία. Παρόλο που τα αρχικά βήματα ακολουθούν συνήθως ένα κοινό σχήμα, στη συνέχεια χρησιμοποιούνται διαφορετικά μονοπάτια, ανάλογα με τους τομείς στους οποίους κάθε φορά επικεντρώνεται η ανάλυση. Πέραν των αρχικών δεδομένων από τις προσβάσεις στον εξυπηρετητή παγκόσμιου ιστού, η εκάστοτε μέθοδος μπορεί να εμπλουτιστεί και με άλλα δεδομένα που προέρχονται είτε άμεσα από τον ιστοχώρο (για παράδειγμα πληροφορίες για τη συσχέτιση των διαφορετικών σελίδων) είτε από τρίτες πηγές όπως δημογραφικές πληροφορίες για τους επισκέπτες.

Μία γενική κατηγοριοποίηση των εργασιών που γίνονται φαίνεται στο Σχήμα 2.7 που προέρχεται από το [8]. Μία αρχική φάση προεπεξεργασίας των πηγαίων δεδομένων ακολουθεί μία φάση όπου εφαρμόζονται τεχνικές Data Mining για την ανακάλυψη ενδιαφέροντων σχημάτων στα δεδομένα (Knowledge Discovery) τα οποία αναλύονται στο τελευταίο στάδιο της διαδικασίας.





Σχήμα 2.7 Διαδικασία Ανάλυσης Επισκεψιμότητας από το [8]

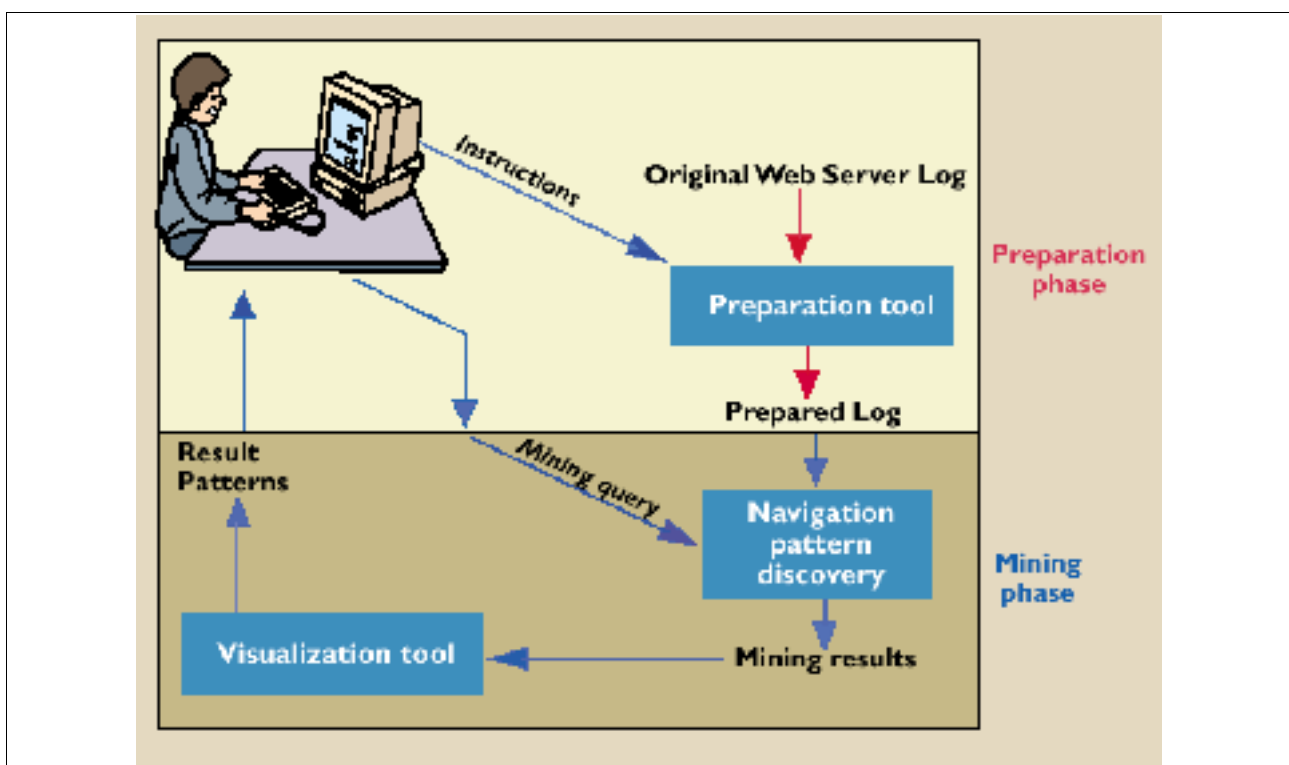
Συνήθως όμως υπάρχει μία ανάδραση από την αποκτούμενη γνώση η οποία μετασχηματίζει την παραπάνω σειριακή διαδικασία σε μορφή κυκλική. Στο [7], το οποίο επικεντρώνεται στην ανακάλυψη μονοπατιών πλοήγησης στον ιστοχώρο (Navigation Pattern Discovery), τα μονοπάτια συγκρίνονται με τα μονοπάτια που προέβλεψαν οι σχεδιαστές κατά την ανάπτυξη του ιστοχώρου ώστε να διαπιστωθεί εάν συμπίπτουν. Στην περίπτωση που δε συμπίπτουν επαναλαμβάνεται η ανάλυση αφού έχει γίνει η κατάλληλη επέμβαση στο σχεδιασμό του ιστοχώρου.

Όπως στο [8] έτσι και στο [7] υπάρχει μία φάση προεπεξεργασίας των πρωτογενών δεδομένων η οποία ακολουθείται από τη φάση εξαγωγής γνώσης. Η πληροφορία που αποκτήθηκε αναλύεται με εργαλεία γραφικής απεικόνισης δεδομένων και εξειδικευμένες γλώσσες ερωτήσεων Σχήμα 2.8.

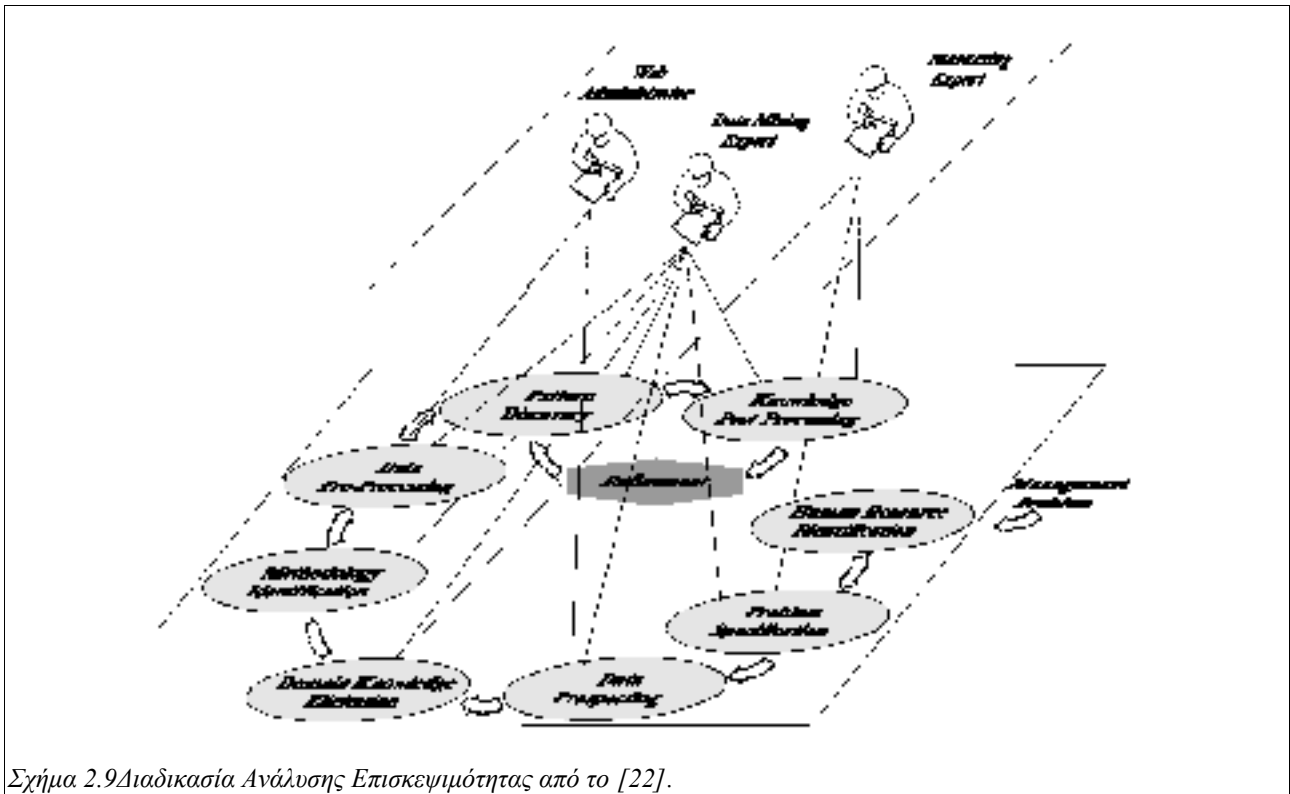
Μία προσπάθεια για γενικότερη προσέγγιση γίνεται στο [22] εισάγοντας την έννοια της γνώσης για την περιοχή της αγοράς που καλύπτει ένας ιστοχώρος. Η εργασία αναφέρεται σε εμπορικούς ιστοχώρους όπου υπάρχουν οι κλασσικές έννοιες της προσέλκυσης πελατών (customer attraction) και διατήρησης πελατών (customer retention) αλλά μπορεί να γενικευτεί για ιστοχώρους γενικής χρήσης όπου ο “πελάτης” είναι ο επισκέπτης και το προϊόν η πληροφορία και οι υπηρεσίες που προσφέρονται. Στην εργασία αναφέρονται τρεις διακριτές ομάδες ατόμων των οποίων η γνώση

χρησιμοποιείται στην ανάλυση επισκεψιμότητας. Πρόκειται κατ' αρχήν για το διαχειριστή του ιστοχώρου, ο οποίος έχει ακριβή γνώση για τη δομή του ιστοχώρου. Επίσης αναφέρεται ο ειδικός μάρκετινγκ για τη συγκεκριμένη περιοχή που καλύπτει ο ιστοχώρος, ο οποίος είναι σε θέση να επιβεβαιώσει τα αποτελέσματα της ανάλυσης. Τέλος είναι οι ειδικοί επί εργασιών Data Mining που θα οδηγήσουν τη διαδικασία ανακάλυψης της γνώσης από τα δεδομένα.

Πέραν των βημάτων που αναφέρθηκαν από τις προηγούμενες εργασίες (προεπεξεργασία πρωτογενών δεδομένων, ανάλυση δεδομένων κλπ.) στο [22] προβλέπεται και πλήθος επιπλέον βημάτων σε σχέση με την προσέγγιση του προβλήματος που σκοπεύεται να αντιμετωπιστεί και τη μεθοδολογία που θα ακολουθηθεί. Από αυτή την άποψη η εργασία παρέχει ένα ολοκληρωμένη διαδικασία για την ανάλυση επισκεψιμότητας. Η όλη διαδικασία απεικονίζεται στο Σχήμα 2.9.



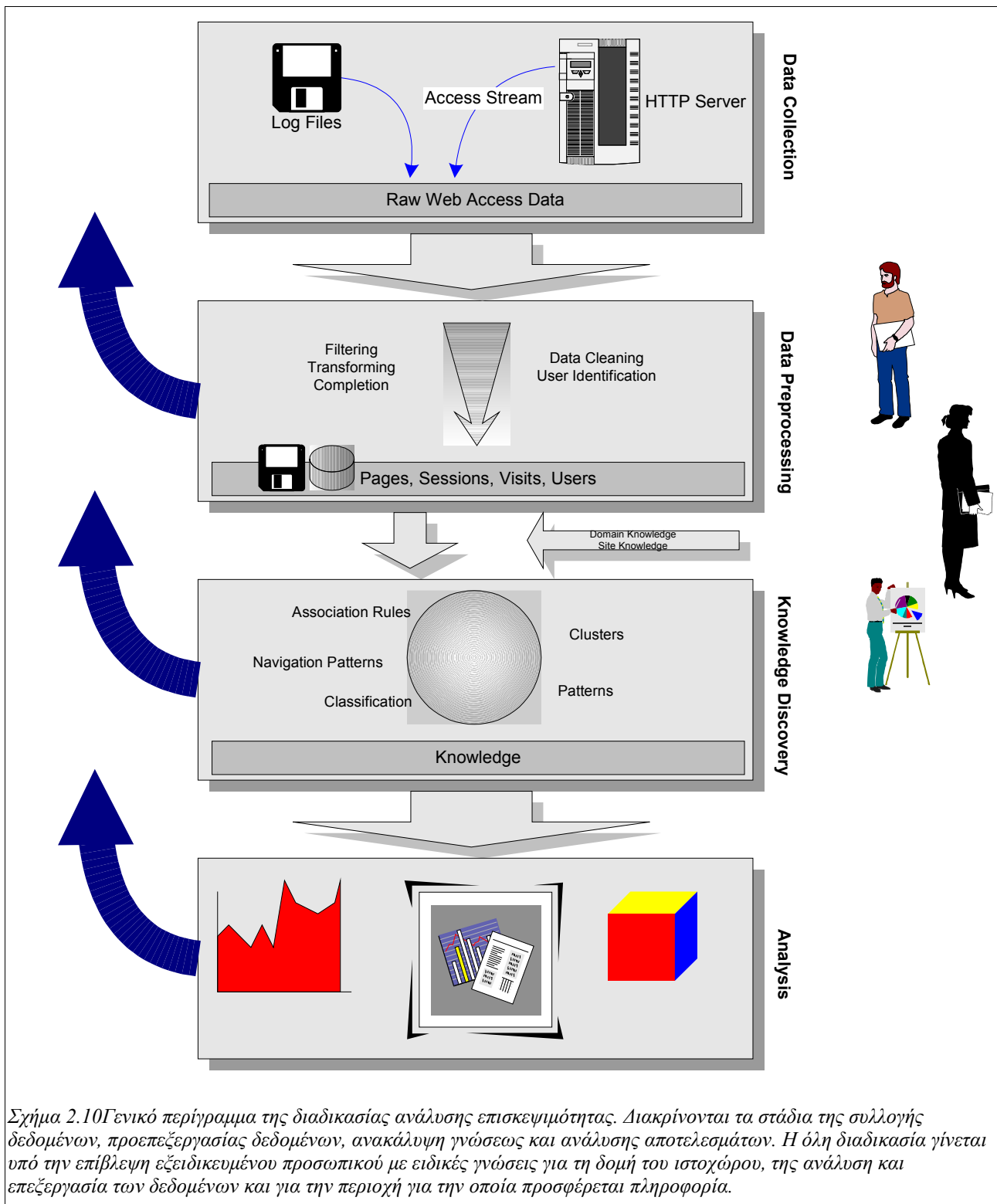
Σχήμα 2.8 Διαδικασία Ανάλυσης Επισκεψιμότητας από το [7]. Διακρίνονται τα σημεία ανάδρασης μετά από τα αποτελέσματα της ανάλυσης



Σχήμα 2.9 Διαδικασία Ανάλυσης Επισκευσιμότητας από το [22].

Σε κάθε περίπτωση η ακριβής διαδικασία που θα ακολουθηθεί εξαρτάται πάντα από τον τελικό στόχο που πρέπει να έχει τεθεί πριν την έναρξη της διαδικασίας. Ακόμα και για την περίπτωση όπου δεν έχουν οριοθετηθεί τα επακριβώς τα επιδιωκόμενα αποτελέσματα, μετά από μία αρχική ανάλυση θα πρέπει αυτά να προσδιοριστούν ώστε να προσαρμοστεί η διαδικασία για βέλτιστη απόδοση σε όρους ταχύτητας και ακρίβειας αποτελεσμάτων.

Ανακεφαλαιώνοντας και συνθέτοντας τις προηγούμενες μοντελοποιήσεις της διαδικασίας ανάλυσης επισκευσιμότητας, μπορούμε να καταλήξουμε σε ένα γενικό διαχωρισμό της διαδικασίας, όπως φαίνεται στο Σχήμα 2.10.



Αρχικά υπάρχει το στάδιο της συλλογής δεδομένων (data collection). Ακολουθεί το στάδιο της προεπεξεργασίας των δεδομένων (data preprocessing) και της ανακάλυψης γνώσης (knowledge discovery). Η διαδικασία τελειώνει με την ανάλυση των δεδομένων που έχουν προκύψει.

Τα δεδομένα από τις προσβάσεις στον ιστοχώρο μπορεί να εμπλουτιστούν με επιπλέον

πληροφορίες πριν το στάδιο της ανακάλυψης γνώσης (για παράδειγμα δημογραφικά δεδομένα).

Η όλη διαδικασία γίνεται υπό την επίβλεψη εξειδικευμένου προσωπικού με ειδικές γνώσεις για τη δομή του ιστοχώρου, της ανάλυση και επεξεργασία των δεδομένων και για την περιοχή για την οποία προσφέρεται πληροφορία.

Στη συνέχεια θα δοθεί μία γενική περιγραφή για τα τελευταία τρία στάδια της ανάλυσης, ενώ θα ακολουθήσει παρουσίαση σχετικής έρευνας και εργασιών σε κάθε έναν από τους τομείς.

### **2.3.2.1.1 Προεπεξεργασία Δεδομένων (Data Preprocessing)**

Σε αυτό το στάδιο προετοιμάζονται τα δεδομένα που αποκτήθηκαν από την κίνηση των εξυπηρετητών HTTP για την περαιτέρω ανάλυση. Συνηθισμένες εργασίες που εκτελούνται είναι

- καθαρισμός από μη επιθυμητή πληροφορία όπως αιτήσεις για πόρους που δεν υπάρχουν στον εξυπηρετητή
- φίλτρο που αφαιρεί αιτήσεις για πόρους που δεν παρουσιάζουν ενδιαφέρον (εικόνες και άλλα μέσα). Συνήθως διατηρούνται μόνο οι αιτήσεις για σελίδες HTML
- συμπληρώνονται προσβάσεις που λογικά θα πρέπει να έχουν γίνει, χωρίς αυτές να εμφανίζονται στα δεδομένα καταγραφής (path completion). Τέτοιες προσβάσεις προκύπτουν από μηχανισμούς caching στην αλυσίδα επικοινωνίας HTTP (Σχήμα 2.11)
- οι αιτήσεις ομαδοποιούνται ανά χρήστη και ανά επίσκεψη και υπολογίζονται έτσι τα μονοπάτια της επίσκεψης. Η ομαδοποίηση μπορεί να επεκταθεί και σε λογικούς συσχετισμούς ανά σελίδα (visit sessions, concept hierarchies, page taxonomies κλπ.).

Τα δεδομένα που προκύπτουν από αυτό το στάδιο φυλάσσονται συνήθως σε δομημένη μορφή σε σχεσιακές βάσεις δεδομένων ώστε να είναι εύκολη και ταχεία η πρόσβαση σε αυτά.

### **2.3.2.1.2 Ανακάλυψη Γνώσης (Knowledge Discovery)**

Το στάδιο της ανακάλυψης γνώσης είναι ίσως από τα πιο ενδιαφέροντα τμήματα στην ανάλυση επισκεψιμότητας. Σε αυτό το σημείο πρέπει να είναι ήδη γνωστό το σημείο στο οποίο πρέπει να επικεντρωθεί η ανάλυση, διότι πρέπει να επιλεγεί η μέθοδος που θα χρησιμοποιηθεί. Ανάλογα με τη μέθοδο πιθανώς να χρειαστεί αναμόρφωση των δεδομένων σε μορφή που να ταιριάζουν στη συγκεκριμένη μέθοδο.

Σε αυτό το σημείο τα δεδομένα μπορεί να εμπλουτιστούν με στοιχεία από τρίτες πηγές (ανεξάρτητα εάν αποκτήθηκαν από τον ίδιο ιστοχώρο ή από αλλού). Τέτοια στοιχεία μπορεί να είναι

δημογραφικά δεδομένα για τους επισκέπτες του ιστοχώρου (περιοχή διαμονής, εισόδημα κλπ).

Η πληροφορία που θα αποκτηθεί μπορεί να περιορίζεται σε απλά στατιστικά στοιχεία κίνησης του ιστοχώρου (αριθμός σελιδοπροβολών ανά επίσκεψη, ώρες αιχμής, σελίδες εισόδου και εξόδου κλπ.), αλλά η πραγματικά ενδιαφέρουσα πληροφορία προέρχεται από τη χρήση τεχνολογίας Εξόρυξης Δεδομένων (Data Mining), η οποία ανακαλύπτει συσχετισμούς που δεν είναι εμφανή με την πρώτη ματιά.

Οι κύριοι τομείς έρευνας είναι οι ακόλουθοι και προέρχονται από την κλασική εφαρμογή τεχνολογιών Data Mining:

- ανακάλυψη σχημάτων πλοήγησης (Navigation Pattern Discovery)
- κανόνες συσχετισμού (Association Rules)
- ακολουθιακά σχήματα (Sequential Patterns)
- ανακάλυψη ομοειδών (Clustering)
- κατάταξη (Classification)

Για κάθε μία από τις παραπάνω τεχνικές θα παρουσιαστούν σχετικές εργασίες

### **2.3.2.1.3 Ανάλυση**

Το στάδιο της ανάλυσης είναι το σημείο όπου κρίνεται η λειτουργία του ιστοχώρου ως προς τους στόχους για τους οποίους σχεδιάστηκε. Χρησιμοποιούνται συνήθως μέθοδοι οπτικοποίησης των δεδομένων από τα προηγούμενα στάδια, που περιλαμβάνουν από απλά διαγράμματα μέχρι και πολύπλοκα σχήματα από αναλύσεις OLAP (Online Analytical Processing).

### **2.3.3 Συλλογή Δεδομένων – Πρωτογενή Δεδομένα**

Σε αυτή την ενότητα θα γίνει αρχικά μία περιγραφή των συνήθως διαθέσιμων πρωτογενών δεδομένων και έπειτα θα παρουσιαστούν διαφορετικοί μέθοδοι που χρησιμοποιούνται για τη συλλογή τους. Θα ακολουθήσει μία σύγκριση των μεθόδων με σκοπό να αναλυθούν τα πλεονεκτήματα και μειονεκτήματα που παρουσιάζουν, και θα περιγραφούν οι τεχνικές δυσκολίες που εμφανίζονται και μειώνουν την αξιοπιστία των δεδομένων.

Η συλλογή των δεδομένων στα οποία θα βασιστεί η ανάλυση της συμπεριφοράς των επισκεπτών τοποθετείται στην αρχή μιας μακράς ακολουθίας ενεργειών που είναι απαραίτητη για την εξαγωγή συμπερασμάτων. Επειδή όλα τα βήματα που ακολουθούν τη συλλογή βασίζονται στα πρωτογενή

δεδομένα η αξιόπιστη και ακριβής καταγραφή τους είναι ιδιαίτερα σημαντική.

Μία συνηθισμένη παρανόηση είναι ότι όλα τα πρωτογενή δεδομένα για την ανάλυση επισκεψιμότητας συλλέγονται αυτόματα, και ότι η μόνη ενέργεια που απαιτείται είναι η αγορά ενός εμπορικού πακέτου ανάλυσης κίνησης του Web, το οποίο θα αναλάβει τα περαιτέρω [28]. Στην πραγματικότητα η συλλογή αξιόπιστων δεδομένων είναι μία μεγάλη πρόκληση και η ελαχιστοποίηση (και όχι η απάλειψη) των λαθών που μπορεί να γίνουν είναι ένα από τα βασικά αντικείμενα έρευνας στον τομέα της ανάλυσης επισκεψιμότητας. Η συλλογή και η επεξεργασία των πρωτογενών δεδομένων δυσχεραίνεται ακόμη περισσότερο λόγω του μεγάλου όγκου των δεδομένων που δημιουργούνται ακόμα και σε μικρούς ιστοχώρους.

### **2.3.3.1 Πρωτογενή Δεδομένα**

Σε αυτή την ενότητα θα παρουσιαστούν οι κυριότερες διαθέσιμες πηγές για τα πρωτογενή δεδομένα. Θα ακολουθήσει ένας πίνακας με την ονομασία και περιγραφή για κάθε ένα από αυτά.

#### **2.3.3.1.1 Το HTTP Πρωτόκολλο**

Η επικοινωνία με τους εξυπηρετητές παγκόσμιου ιστού (WEB Servers) γίνεται με το HTTP πρωτόκολλο (Hypertext Transfer Protocol) [29]. Η ιστορία του πρωτοκόλλου ξεκινάει ήδη από το 1990 με την έκδοση 0.9 ενώ σήμερα χρησιμοποιείται η έκδοση 1.1 όπου και έχει σταματήσει η επίσημη εξέλιξη από το World Wide Web Consortium (W3C) [30].

Το πρωτόκολλο τοποθετείται στο Application Layer του μοντέλου δικτύων OSI και χρησιμοποιεί συνήθως TCP/IP συνδέσεις για τη μεταφορά των δεδομένων. Πρόκειται για ένα πρωτόκολλο αίτησης/απόκρισης (request/response) όπου μία εφαρμογή πελάτη (user agent) στέλνει μία αίτηση σε εξυπηρετητή και λαμβάνει από αυτόν μία απάντηση. Η γενική μορφή των ερωτήσεων και αποκρίσεων φαίνεται στον πίνακα 2.3.

<i>Αίτηση</i>	<i>Απόκριση</i>
Request Method	Status Line
URI	Protocol Version
Protocol Version	Success/ Error Code
Request Modifiers	Server Information
Client Information	Entity Metainformation
Body Content	Entity Body- Content

*Πίνακας 2.3 Η γενική μορφή αιτήσεων και αποκρίσεων του HTTP πρωτοκόλλου*

Στην απλούστερη περίπτωση η αλυσίδα αιτήσεων και αποκρίσεων πραγματοποιείται πάνω από μία απλή σύνδεση. Σε πιο πολύπλοκες περιπτώσεις όμως μπορεί να παρεμβάλλονται στην αλυσίδα και

μεσάζοντες, με πιο γνωστούς τους εξυπηρετητές proxy, οι οποίοι είναι πράκτορες διαβίβασης (forwarding agents) των αιτήσεων από τους πελάτες προς τους εξυπηρετητές.

Κάθε τμήμα της αλυσίδας αιτήσεων και απαντήσεων του HTTP πρωτοκόλλου μπορεί να υλοποιεί και μηχανισμούς caching. Στο Σχήμα 2.11 φαίνονται επιλεγμένες περιπτώσεις αλυσίδας επικοινωνίας.

Ένα απλό παράδειγμα αίτησης και απάντησης με βάση το HTTP πρωτόκολλο φαίνεται στον πίνακα 2.4

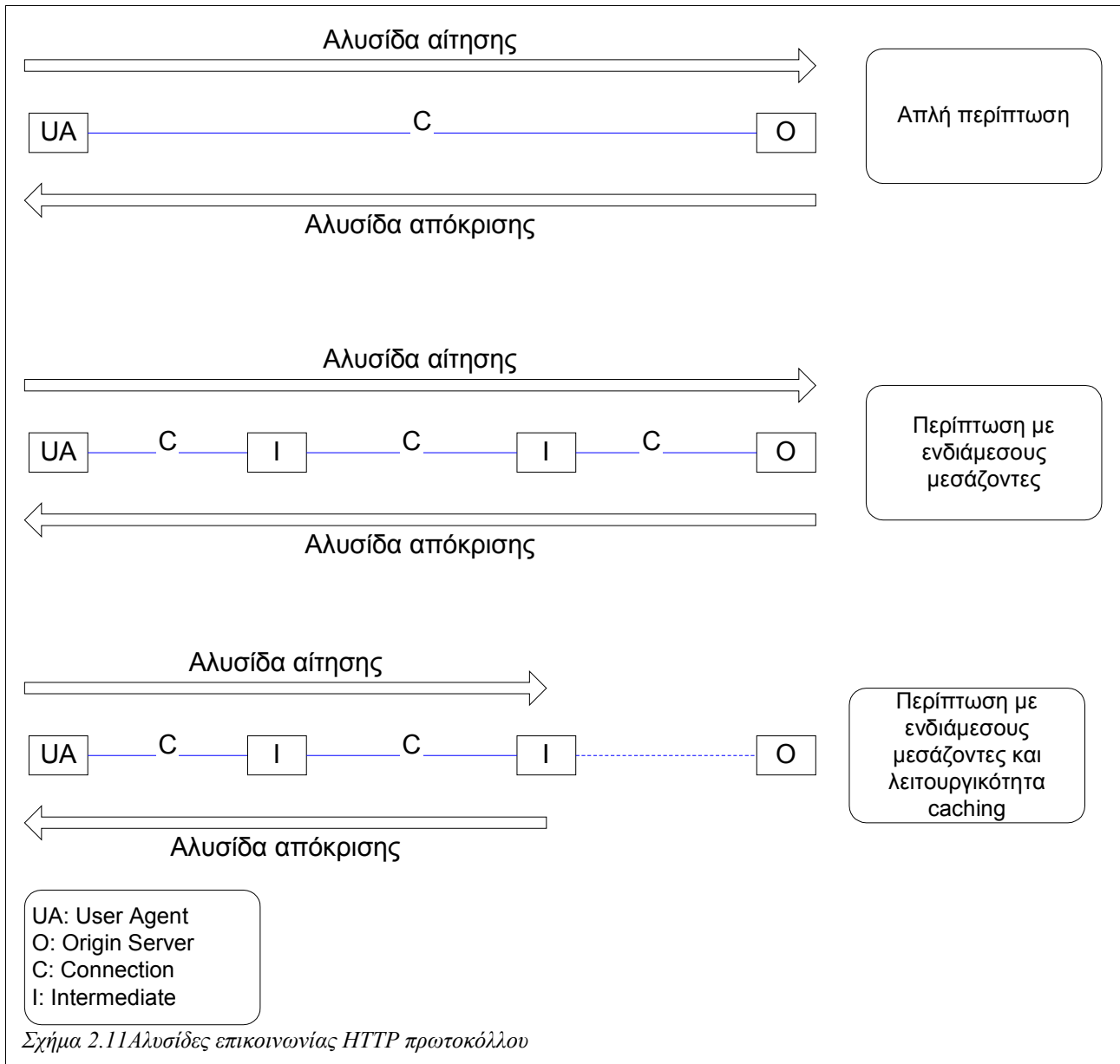
<i>A/</i> <i>A</i>	<i>HTTP Γραμμή</i>
1	GET /articles/index.shtml HTTP/1.0
2	User-Agent: Mozilla 4.0 (X; I; Linux-2.0.35i586)
3	Host: www.perlflect.com
4	Accept: image/gif, image/jpeg, */*
6	HTTP/1.0 200 OK
7	Date: Thus, 08 Oct 1998 16:17:52 GMT
8	Server: Apache/1.1.1
9	Content-type: text/html
10	Content-length: 1538
11	Last-modified: Mon, 05 Oct 1998 01:23:50 GMT
12	<i>CONTENT</i>

Πίνακας 2.4 Παράδειγμα αίτησης και απάντησης HTTP πρωτοκόλλου

Στη γραμμή 1 ξεκινάει η αίτηση από την εφαρμογή πλοήγησης. Ζητείται ο πόρος (σελίδα) με τη μέθοδο GET χρησιμοποιώντας την έκδοση 1.0 του πρωτοκόλλου. Η εφαρμογή που χρησιμοποιείται φαίνεται στη γραμμή 2 ενώ ο πόρος ζητείται από τον ιστοχώρο της γραμμής 3. Τέλος αναφέρονται και οι τύποι περιεχομένου MIME που δέχεται η εφαρμογή (γραμμή 4).

Η απάντηση του εξυπηρετητή δίνεται με την έκδοση 1.0 του πρωτοκόλλου και έχει κωδικό 200 (γραμμή 6). Αναφέρεται η ημερομηνία και ο τύπος του εξυπηρετητή (γραμμή 7 και 8) και επίσης η μορφή του περιεχομένου (γραμμή 9). Το περιεχόμενο έχει μέγεθος 1538 Bytes, έχει ενημερωθεί τελευταία φορά όπως φαίνεται στη γραμμή 11. Το κυρίως περιεχόμενο ακολουθεί μετά τη γραμμή 12.





### 2.3.3.1.2 Πρωτογενή Δεδομένα από το HTTP Πρωτόκολλο

Το HTTP πρωτόκολλο είναι ο σύνδεσμος των τελικών χρηστών με τα δεδομένα που διατίθενται από τους εξυπηρετητές WWW. Όλες οι κινήσεις των χρηστών για ανάκτηση των δεδομένων από τους εξυπηρετητές μεταφράζονται από την εφαρμογή πλοήγησης που χρησιμοποιούν σε πακέτα αιτήσεων σύμφωνα με αυτό το διεθνές πρότυπο. Από αυτή την άποψη η παρακολούθηση της επικοινωνίας των χρηστών με βάση το πρωτόκολλο μπορεί να δώσει πολύτιμες πληροφορίες για τις κινήσεις των χρηστών και επομένως για τη συμπεριφορά τους σε έναν συγκεκριμένο ιστοχώρο.

Βέβαια ο σκοπός του πρωτοκόλλου είναι πρωτίστως η αξιόπιστη επικοινωνία των εφαρμογών πλοήγησης με τους εξυπηρετητές και γι' αυτό το λόγο τα δεδομένα από το πρωτόκολλο δεν είναι σε εύχρηστη μορφή για την ανάλυση της συμπεριφοράς των χρηστών. Δίνουν όμως τα συστατικά από

τα οποία με την κατάλληλη επεξεργασία και μετατροπή η εξαγωγή πληροφορίας γίνεται πιο εύκολη.

### 2.3.3.1.3 Άλλες Πηγές Δεδομένων

Πέραν των δεδομένων που προκύπτουν από το HTTP πρωτόκολλο υπάρχουν και άλλες πηγές δεδομένων που χρησιμοποιούνται για την ανάλυση επισκεψιμότητας. Οι κυριότερες πηγές είναι οι ίδιοι οι εξυπηρετητές παγκόσμιου ιστού που διαχειρίζονται τόσο το HTTP πρωτόκολλο όσο και τις συνδέσεις δικτύου με τις εφαρμογές πλοήγησης. Επίσης σημαντική πηγή είναι ειδικές εφαρμογές που μπορεί να εκτελούνται στον υπολογιστή του τελικού χρήστη.

### 2.3.3.1.4 Πρωτογενή Δεδομένα

Στον πίνακα 2.5 αναφέρονται τα συνήθως χρησιμοποιούμενα πρωτογενή δεδομένα και οι πηγές από τις οποίες προέρχονται.

<i>Όνομα</i>	<i>Πηγή</i>	<i>Περιγραφή</i>
Host Ip	W	Η διεύθυνση IP του επόμενου σταθμού στην αλυσίδα αιτήσεων/αποκρίσεων από την πλευρά του εξυπηρετητή. Εάν πρόκειται για την απλή περίπτωση, τότε η διεύθυνση ανήκει στο μηχάνημα του τελικού χρήστη. Εάν όμως υπάρχουν και ενδιάμεσοι σταθμοί τότε ανήκει στον τελευταίο πριν τον εξυπηρετητή ενδιάμεσο σταθμό.
Host Name	WO	Το όνομα του μηχανήματος που προκύπτει από Reverse DNS Lookup του Host IP
User	W	Σε περιπτώσεις όπου χρησιμοποιείται Αναγνώριση Χρήστη (User Authentication) είναι το όνομα χρήστη
Time	WH	Η χρονική στιγμή που έγινε η αίτηση
Request URI, Method, Protocol	WH	Ο πόρος (resource) που ζητήθηκε με μία συγκεκριμένη αίτηση (ενδεχομένως με παραμέτρους), η έκδοση του πρωτοκόλλου HTTP και η μέθοδος που χρησιμοποιήθηκε (GET, PUT κ.α.)
Request Body	WH	Ενδεχόμενο μήνυμα ή δεδομένα που συνοδεύουν μία αίτηση
Response Size	WH	Μέγεθος της απάντησης που δίνει ο εξυπηρετητής
Referrer	WH	URI από το οποίο προήλθε η συγκεκριμένη αίτηση (π.χ. Πατώντας έναν σύνδεσμο από άλλη σελίδα)
User Agent	WH	Μία σειρά που περιγράφει την εφαρμογή πλοήγησης του τελικού χρήστη
Via Forwarded X-Forwarded-For	WH	Σε περίπτωση όπου υπάρχουν ενδιάμεσοι στη αλυσίδα αιτήσεων/αποκρίσεων περιέχει τις διευθύνσεις των ενδιάμεσων. Το πεδίο X-Forwarded-For δεν είναι μέρος του προτύπου HTTP αλλά επειδή χρησιμοποιείται από έναν πολύ διαδεδομένο proxy εξυπηρετητή χρησιμοποιείται συχνά σε αναλύσεις επισκεψιμότητας επειδή δίνει την πραγματική IP διεύθυνση του τελικού χρήστη.

<b>Όνομα</b>	<b>Πηγή</b>	<b>Περιγραφή</b>
Server response time	W	Ο χρόνος που χρειάστηκε ο εξυπηρετητής να επεξεργαστεί και να απαντήσει στην αίτηση (δεν πρέπει να συγχέεται με το χρόνο που απαιτήθηκε να φτάσει η απάντηση στον τελικό χρήστη)
Accept Language	WH	Πεδίο που στέλνουν οι εφαρμογές πλοήγησης και καθορίζουν σε ποιες γλώσσες προτιμάει ο τελικός χρήστης να βλέπει την πληροφορία που ζήτησε
Cookies	WH	Μικρές ποσότητες δεδομένων που μπορεί να ζητήσει ο εξυπηρετητής να αποθηκευτούν στην εφαρμογή πλοήγησης και να ξαναστέλνονται με κάθε νέα αίτηση στον ίδιο εξυπηρετητή. Τα δεδομένα αυτά μπορεί να καθοριστεί για πόσο χρονικό διάστημα πρέπει να φυλάσσονται. Είναι στην ευχέρεια των εφαρμογών πλοήγησης να δεχτούν αυτά τα δεδομένα.
User Agent Properties	U	Διάφορα στοιχεία που είναι διαθέσιμα από προγράμματα που τρέχουν στον υπολογιστή του τελικού χρήστη. Τέτοια στοιχεία μπορεί να είναι η ανάλυση της οθόνης του και ο αριθμός των χρωμάτων, εγκατεστημένα plugins μέχρι και στοιχεία για το ποιες άλλες σελίδες επισκέπτεται. Ο περιορισμός σε αυτά τα δεδομένα υπάρχει στο ότι είναι στην ευθύνη του τελικού χρήστη να επιτρέψει τη εκτέλεση τέτοιων προγραμμάτων.
<b>W:</b> WWW Server, <b>H:</b> HTTP Protocol, <b>U:</b> User, <b>O:</b> Other		

Πίνακας 2.5 Περιγραφές των κυριότερων διαθέσιμων πρωτογενών δεδομένων για την ανάλυση επισκεψιμότητας και οι πηγές τους

Παρόλο που τα παραπάνω δεδομένα καλύπτουν σχεδόν το σύνολο των διαθέσιμων στοιχείων που χρησιμοποιούνται για την ανάλυση επισκεψιμότητας, εντούτοις υπάρχουν και άλλα δεδομένα που μπορεί να χρησιμοποιηθούν αλλά εξαρτώνται σε μεγάλο βαθμό από τον τρόπο καταγραφής που επιλέγεται. Θα γίνει αναφορά σε επιλεγμένα από αυτά στις επόμενες ενότητες.

### **2.3.3.2 Μέθοδοι Συλλογής Δεδομένων**

Έχοντας περιγράψει τα κυριότερα πρωτογενή δεδομένα που χρησιμοποιούνται σαν βάση για την περαιτέρω ανάλυση της συμπεριφοράς χρηστών, απομένει να περιγραφεί ο τρόπος με τον οποίο αυτά τα δεδομένα συλλέγονται.

Η εταιρεία Accrue Software στο [31] περιγράφει 3 διαφορετικούς τρόπους που χρησιμοποιεί στα προϊόντα της οι οποίοι καλύπτουν σε μεγάλο βαθμό τους καθιερωμένους τρόπους συλλογής των πρωτογενών δεδομένων. Στις επόμενες ενότητες θα παρουσιαστούν αυτοί οι τρόποι καθώς και κάποιοι επιπλέον που χρησιμοποιούνται είτε για λιγότερο απαιτητικές αναλύσεις είτε επειδή απαιτούν εξειδικευμένο λογισμικό στον υπολογιστή του τελικού χρήστη.

#### **2.3.3.2.1 Συλλογή από Αρχεία Καταγραφής (Log Files)**

Όλοι οι εξυπηρετητές HTTP έχουν τη δυνατότητα να παράγουν αρχεία καταγραφής των

προσβάσεων σε αυτούς. Συνήθως υποστηρίζονται διάφορα είδη αρχείων καταγραφής που διαφέρουν στο πλήθος των στοιχείων που καταγράφονται και στο σχήμα ή τη διάταξη που χρησιμοποιούν, αλλά έχει καθιερωθεί ένα συγκεκριμένο είδος που υποστηρίζεται σχεδόν από όλους τους εξυπηρετητές και χρησιμοποιείται σχεδόν αποκλειστικά σε όλες τις εγκαταστάσεις εξυπηρετητών.

Πρόκειται για το Combined Log Format [32] που προκύπτει από το Common Log Format (CLF) προσαυξημένο με δύο επιπλέον πεδία. Αναλυτικά, το Combined Log Format αποθηκεύει τα στοιχεία του πίνακα 2.6.

<i>Όνομα</i>	<i>Περιγραφή</i>
host	client ip-address ή όνομα μηχανήματος
RFC1413	client identity (αποθαρρύνεται η χρήση)
user	όνομα χρήστη για προστατευόμενες σελίδες
date-time	μέρα και ώρα περάτωσης της αίτησης
request	request method, resource και πρωτόκολλο
status	response status
size	μέγεθος σε bytes της απόκρισης χωρίς τα HTTP headers
referrer	referring URL (από που ήρθε)
user-agent	browser που χρησιμοποιήθηκε

Πίνακας 2.6 Το Combined Log Format

Όλα τα στοιχεία του πίνακα, πλην του πεδίου RFC1413 περιέχονται στον πίνακα για τα πρωτογενή δεδομένα της προηγούμενης ενότητας (πίνακας 2.5). Το πεδίο RFC1413 δε χρησιμοποιείται πλέον και παραμένει μόνο για λόγους συμβατότητας.

Τα συστήματα ανάλυσης διαβάζουν περιοδικά τα αρχεία καταγραφής του εξυπηρετητή, τα φιλτράρουν και τα καθαρίζουν από μη επιθυμητή πληροφορία και ακολούθως είτε τα επεξεργάζονται απ' ευθείας είτε τα εισάγουν πρώτα σε κάποια βάση δεδομένων.

Να σημειωθεί ότι πέραν του Combined Log Format οι εξυπηρετητές είναι σε θέση να καταγράφουν και πολύ περισσότερα στοιχεία. Εναπόκειται στα συστήματα ανάλυσης των αρχείων να μπορούν να ερμηνεύσουν τη σειρά και το νόημα των επιπλέον πεδίων.

### **2.3.3.2 Συλλογή από την Παρακολούθηση του Τοπικού Δικτύου**

Σε αυτήν την περίπτωση ένα εξειδικευμένο υπολογιστικό σύστημα παρακολουθεί το τοπικό δίκτυο απ' όπου περνάνε οι αιτήσεις προς τον εξυπηρετητή και φιλτράρει τα πακέτα TCP/IP που αναφέρονται στην επικοινωνία με τον εξυπηρετητή HTTP (TCP/IP packet sniffing). Όπως και στην περίπτωση με τα αρχεία καταγραφής, τα δεδομένα που συλλέγονται συνήθως αποθηκεύονται σε κάποια βάση δεδομένων.

### **2.3.3.2.3 Συλλογή από Ενσωματωμένα Plugins**

Κάποιες εταιρείες έχουν αναπτύξει για τους σημαντικότερους εξυπηρετητές HTTP ειδικά modules/plug-ins που ενσωματώνονται στον εξυπηρετητή και αναλαμβάνουν την καταγραφή των προσβάσεων.

### **2.3.3.2.4 Συλλογή από Εξειδικευμένες Αιτήσεις**

Πρόκειται για μία μέθοδο που χρησιμοποιείται κυρίως από εταιρείες που προσφέρουν υπηρεσίες ανάλυσης επισκεψιμότητας σε τρίτους (παράδειγμα [33]). Η μέθοδος βασίζεται στην ενσωμάτωση μιας ειδικής εικόνας σε κάθε σελίδα του ιστοχώρου, η οποία προέρχεται όμως από ειδικό εξυπηρετητή της εταιρίας στατιστικής ανάλυσης. Έτσι με κάθε εμφάνιση μιας σελίδας του υπό εξέταση ιστοχώρου, η εφαρμογή πλοήγησης εκτελεί και μία αίτηση προς τον εξυπηρετητή ανάλυσης για την ειδική εικόνα. Έτσι ο εξυπηρετητής ανάλυσης έχει από αυτήν την αίτηση πλήρη πρόσβαση σε όλα τα στοιχεία του πελάτη που δεν αναφέρονται στο συγκεκριμένο ιστοχώρο από το HTTP πρωτόκολλο. Προσθέτοντας κατάλληλες παραμέτρους στο URL της εικόνας όπως είναι ένας κωδικός του ιστοχώρου ή το όνομα της σελίδας, ο εξυπηρετητής ανάλυσης έχει επίσης πρόσβαση σε στοιχεία του συγκεκριμένου ιστοχώρου. Σε κάθε περίπτωση όμως δεν υπάρχουν στοιχεία για το ειδικό περιεχόμενο της κάθε σελίδας που αυτό μπορεί να περιλαμβάνει τις διαφημίσεις που εμφανίζονται ή οι σύνδεσμοι που υπάρχουν.

Συνήθως το URL της ειδικής εικόνας εμπλουτίζεται και με επιπλέον στοιχεία, χρησιμοποιώντας κάποια scripting γλώσσα (JavaScript) που εκτελείται στην εφαρμογή πλοήγησης, και με αυτόν τον τρόπο μπορεί να αποκτήσει επιπλέον στοιχεία που είναι προσβάσιμα μέσω της γλώσσας. Τέτοια στοιχεία είναι για παράδειγμα η ανάλυση της οθόνης που χρησιμοποιεί ή τα Plugins που έχουν εγκατασταθεί και είναι διαθέσιμα στην εφαρμογή πλοήγησης του χρήστη. Ειδικά τα τελευταία στοιχεία δεν είναι διαθέσιμα με κανέναν άλλον τρόπο συλλογής στοιχείων που παρουσιάστηκαν παραπάνω επειδή είναι διαθέσιμα μόνο δια μέσω του υπολογιστή του επισκέπτη.

Συνήθως αυτή η ειδική εικόνα παρουσιάζει έναν μετρητή με τον αριθμό των επισκέψεων σαν υπηρεσία προς τον ιστοχώρο ή κάποια διαφήμιση της εταιρίας ανάλυσης (πέραν των αναλυτικών στατιστικών που προσφέρει η υπηρεσία σε ειδικές σελίδες), ή είναι αόρατη (1x1 transparent GIF image) στην περίπτωση που η υπηρεσία προσφέρεται επί πληρωμή.

Υπάρχουν όμως και περιπτώσεις όπου αυτή η μέθοδος χρησιμοποιείται για τη μέτρηση της επισκεψιμότητας ιστοχώρων με σκοπό την ουδέτερη αξιολόγηση με ενιαίο τρόπο από εταιρίες στο χώρο των διαφημίσεων. Ένα τέτοιο παράδειγμα είναι η [34] της οποίας το μηχανισμό χρησιμοποιούν γνωστοί ιστοχώροι της Γερμανίας όπως Spiegel Online ή Heise Online [35] και με

αυτόν τον τρόπο οι διαφημιζόμενοι έχουν μία ενιαία εικόνα για την τοποθέτηση των διαφημίσεών τους.

Οι ειδικές εικόνες, ειδικά όταν πρόκειται για αόρατες 1x1 εικόνες, αναφέρονται συχνά με τον όρο Web Bug

### **2.3.3.2.5 Συλλογή από Ειδικές Εφαρμογές στον Τελικό Χρήστη**

Μία επίσης διαδεδομένη μέθοδος για τη συλλογή δεδομένων για τις κινήσεις των χρηστών είναι η εγκατάσταση μίας ειδικής εφαρμογής στον υπολογιστή του τελικού χρήστη, η οποία παρακολουθεί όλες τις κινήσεις που κάνει ο χρήστης και στέλνει τα δεδομένα σε ειδικό εξυπηρετητή προς επεξεργασία.

### **2.3.3.3 Σύγκριση Μεθόδων Συλλογής Πρωτογενών Δεδομένων**

Όλες οι μέθοδοι που περιγράφηκαν χρησιμοποιούνται ευρέως από εταιρείες και ερευνητές για τη συλλογή των πρωτογενών δεδομένων από τις επισκέψεις σε ιστοχώρους. Το εύρος των δεδομένων που μπορεί να καταγραφούν δεν είναι όμως το ίδιο για όλες τις μεθόδους επειδή η κάθε μία ενσωματώνεται σε διαφορετικό σημείο στην αλυσίδα επικοινωνίας του HTTP πρωτοκόλλου. Τα κριτήρια με τα οποία μπορεί να αξιολογηθούν οι μέθοδοι μπορεί να χωριστούν σε κριτήρια για το εύρος των δεδομένων που μπορεί να καταγραφούν, τη δυνατότητα παρακολούθησης σε πραγματικό χρόνο της κίνησης καθώς και το βαθμό ευκολίας ενσωμάτωσης σε υπάρχουσες εγκαταστάσεις εξυπηρετητών. Για κάθε μία μέθοδος θα περιγραφεί η απόδοσή τους με βάση αυτά τα κριτήρια.

Η μέθοδος που εκμεταλλεύεται τα **αρχεία καταγραφής** είναι ίσως η πλέον διαδεδομένη διότι είναι εντελώς ανεξάρτητη από τη λειτουργία του εξυπηρετητή και δε χρειάζεται κανενός είδους παρέμβαση σε αυτόν. Όλοι οι εξυπηρετητές είναι σε θέση να παράγουν τέτοια αρχεία οπότε μπορεί να αναπτυχθούν γενικές εφαρμογές ανάλυσης που να καλύπτουν σχεδόν όλες τις υπάρχουσες εγκαταστάσεις εξυπηρετητών.

Το μειονέκτημά της είναι ότι ακριβώς επειδή η ανάγνωση των αρχείων γίνεται περιοδικά δεν επιτρέπεται η πρόσβαση στα δεδομένα σε πραγματικό χρόνο, οπότε αποκλείονται οι χρήσεις των αποτελεσμάτων της ανάλυσης για εφαρμογές ανάδρασης στη λειτουργία του ιστοχώρου σε πραγματικό χρόνο. Επίσης η επικέντρωση στην κάλυψη όσο το δυνατόν περισσότερων υφιστάμενων εγκαταστάσεων εξυπηρετητών περιορίζει τα πιθανά δεδομένα που μπορεί να καταγραφούν επειδή τα αρχεία καταγραφής πρέπει να ακολουθούν ένα συγκεκριμένο σχήμα. Για παράδειγμα στα Combined Log Format δεν υπάρχει καθόλου πληροφορία για τα cookies που όπως θα αναλυθεί αργότερα είναι ένα πολύ σημαντικό στοιχείο για την ανάλυση επισκεψιμότητας.

Επίσης δεν υπάρχει καμία πρόσβαση στα δεδομένα που μεταδίδονται ως κυρίως μέρος των αιτήσεων (body, PUT Method) που μπορεί να είναι για παράδειγμα τα στοιχεία που εισάγει ο χρήστης σε μία φόρμα αναζήτησης.

Η συλλογή δεδομένων από την **παρακολούθηση του τοπικού δικτύου** έχει σαφώς καλύτερα χαρακτηριστικά. Επειδή παρακολουθείται όλη η κίνηση από και προς τον εξυπηρετητή την ώρα που λειτουργεί, τα δεδομένα μπορεί να αποθηκεύονται και να επεξεργάζονται άμεσα, οπότε μπορεί να υπάρχει εικόνα της κίνησης στον εξυπηρετητή σε πραγματικό χρόνο. Τέτοιου είδους συστήματα όμως συνήθως απαιτούν εξειδικευμένα υπολογιστικά συστήματα και τοπολογίες στο δίκτυο (παράδειγμα ειδικό μηχάνημα που αναλαμβάνει αυτή την εργασία, ρύθμιση ενδεχομένου switch σε monitoring port ώστε να μεταδίδονται όλα τα δεδομένα προς το ειδικό μηχάνημα) που συνήθως αυξάνει το κόστος εγκατάστασης και λειτουργίας, ενώ καλύπτει μόνο τον εξυπηρετητή με τον οποίο είναι συνδεδεμένα.

Όσον αφορά τα δεδομένα που μπορεί να συλλεχθούν, περιορίζονται ουσιαστικά μόνο στα δεδομένα που μεταδίδονται κρυπτογραφημένα από την εφαρμογή του χρήστη στον εξυπηρετητή. Πέραν των δεδομένων που περιγράφηκαν παραπάνω, τέτοιου είδους εγκαταστάσεις μπορεί να εντοπίσουν και ενέργειες του χρήστη που αλλιώς δεν είναι εύκολα αναγνωρίσιμες, όπως είναι για παράδειγμα η διακοπή της μεταφοράς μιας εικόνας από το χρήστη.

Η **ενσωμάτωση** της λειτουργικότητας συλλογής των δεδομένων **στον εξυπηρετητή** έχει το μεγάλο πλεονέκτημα ότι υπάρχει πρόσβαση σε όλα τα δεδομένα της επικοινωνίας με τον εξυπηρετητή (ακόμη και της κρυπτογραφημένης πληροφορίας) και η επεξεργασία μπορεί να γίνει σε πραγματικό χρόνο. Επειδή όμως αυτή η ενσωμάτωση απαιτεί ειδική παρέμβαση στον εξυπηρετητή (είτε μέσω αλλαγής του πηγαίου κώδικα είτε μέσω ειδικών APIs που προσφέρουν οι εξυπηρετητές) η δυνατότητα χρησιμοποίησης τέτοιων λύσεων περιορίζεται στους εξυπηρετητές για τους οποίους έχουν αναπτυχθεί τέτοιες λύσεις, οπότε αποκλείονται συνήθως εξυπηρετητές με μικρή διάδοση.

Η συλλογή των πρωτογενών δεδομένων με την **ενσωμάτωση ειδικής εικόνας** στις σελίδες παρουσιάζει κατ' αρχήν πολλά πλεονεκτήματα. Επειδή η συλλογή των δεδομένων δε γίνεται με παρεμβάσεις σε συγκεκριμένο εξυπηρετητή, μπορεί με μία εγκατάσταση συλλογής να καλυφθούν πολλοί διαφορετικοί ιστοχώροι. Επίσης συλλέγονται δεδομένα για τους ίδιους χρήστες που επισκέπτονται διαφορετικούς ιστοχώρους εάν αυτοί οι ιστοχώροι χρησιμοποιούν το ίδιο σύστημα συλλογής δεδομένων. Η χρήση γλωσσών και προγραμμάτων που ενσωματώνονται στη σελίδα που προβάλλεται επιτρέπει την ενσωμάτωση δεδομένων που δεν είναι προσβάσιμες με τους προηγούμενους τρόπους, όπως για παράδειγμα τα δεδομένα για τα εγκατεστημένα Plugins της

εφαρμογής πλοήγησης.

Το μεγάλο μειονέκτημα είναι ότι η συλλογή βασίζεται στην πρόκληση αίτησης από την εφαρμογή πλοήγησης για την ειδική ενσωματωμένη εικόνα. Στο παρελθόν η μη εμφάνιση εικόνας συναντιόταν μόνο στις περιπτώσεις όπου χρησιμοποιούνταν εφαρμογές πλοήγησης χωρίς δυνατότητα εμφάνισης εικόνων, ή όταν ο χρήστης επέλεγε συνειδητά τη μη εμφάνιση καμίας εικόνας. Πρόσφατα όμως οι σύγχρονες εφαρμογές πλοήγησης έχουν ενσωματώσει ειδικές λειτουργίες που επιτρέπουν στο χρήστη να μη δέχεται εικόνες που προέρχονται από συγκεκριμένες διευθύνσεις. Με αυτόν τον τρόπο «κόβονται» οι διάφορες διαφημίσεις αλλά επίσης και αυτές οι ειδικές εικόνες που χρησιμοποιούνται για τη συλλογή δεδομένων. Επίσης υπάρχει έντονος προβληματισμός για παραβίαση των δικαιωμάτων των χρηστών όσον αφορά την επεξεργασία προσωπικών δεδομένων.

Οι εφαρμογές που εκτελούνται στον υπολογιστή του τελικού χρήστη παρουσιάζουν ιδιαίτερο ενδιαφέρον επειδή επιτρέπουν την παρακολούθηση κινήσεων του χρήστη που αλλιώς δεν είναι διαθέσιμες. Έχουν συνήθως πλήρη πρόσβαση στο μηχάνημα του χρήστη και με αυτόν τον τρόπο ουσιαστικά μπορούν να συλλέξουν οποιαδήποτε πληροφορία. Το μειονέκτημά τους έγκειται στο γεγονός ότι πρέπει πρώτα να πεισθεί ο χρήστης να εγκαταστήσει την εφαρμογή, πράγμα που δεν πρόκειται να κάνει εάν δε συνδυάζεται με κάποιο όφελος για αυτόν. Γι' αυτό συνήθως χρησιμοποιούνται με αντάλλαγμα κάποια υπηρεσία προς το χρήστη, ή όταν η ανάλυση γίνεται σε περιορισμένο και ελεγχόμενο σύνολο χρηστών.

Βεβαίως σε όλες τις περιπτώσεις αλλά και ειδικά στην τελευταία ανακύπτουν σοβαρά ζητήματα προστασίας προσωπικών (και μη) δεδομένων των χρηστών.

### **2.3.4 Μετρικές και Μεγέθη για την Ανάλυση Επισκεψιμότητας**

Στην προηγούμενη ενότητα περιγράφηκαν τα διαθέσιμα πρωτογενή δεδομένα που χρησιμοποιούνται για την ανάλυση επισκεψιμότητας σε ιστοχώρους.

Ο τελικός σκοπός στην ανάλυση του ιστοχώρου είναι η κατανόηση των χρηστών και η εξαγωγή συμπερασμάτων για τους λόγους που οδηγούν σε μία συγκεκριμένη συμπεριφορά. Όλοι οι τρόποι συλλογής των δεδομένων της κίνησης του ιστοχώρου έχουν ως αφετηρία τις μεμονωμένες αιτήσεις που γίνονται στον εξυπηρετητή HTTP και παρόλο που υπάρχει μία σαφής σχέση ανάμεσα στις αιτήσεις και στους χρήστες, οι αιτήσεις από μόνες τους έχουν μία πολύ μικρή αξία στην ανάλυση και εξαρτώνται περισσότερο από το σχεδιασμό του ιστοχώρου παρά από τη συμπεριφορά των χρηστών. Στην Εικόνα 2.1 φαίνεται ένα παράδειγμα μίας πραγματικής σελίδας HTML.





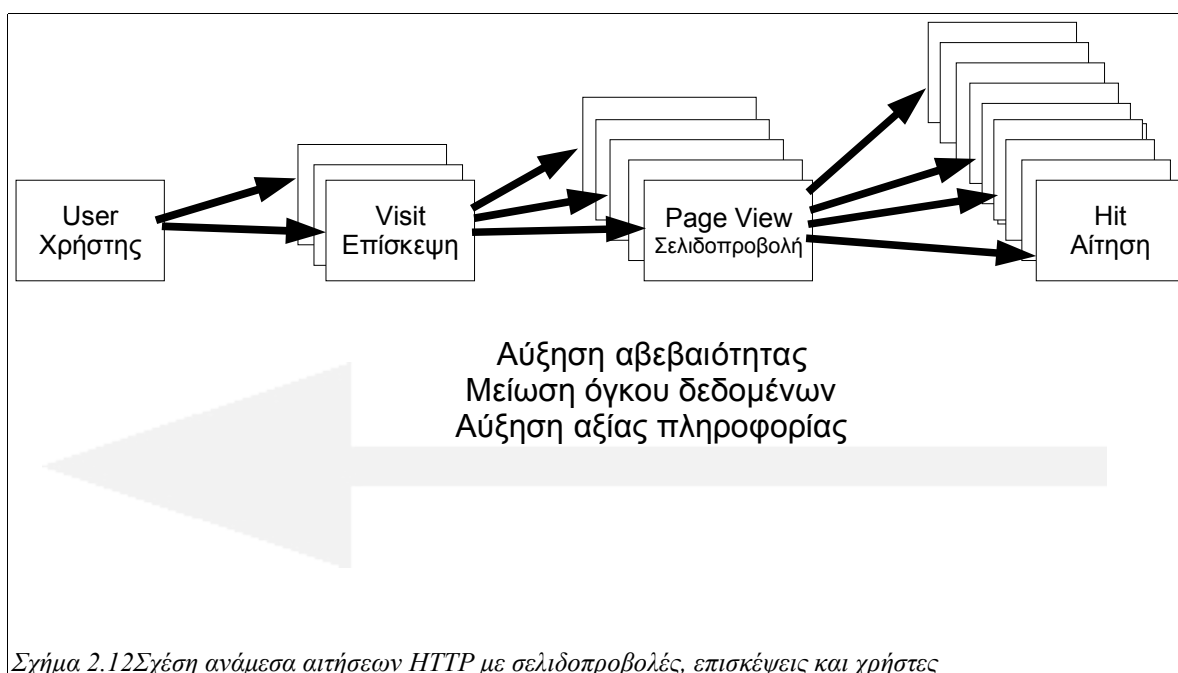
Εικόνα 2.1 Απλή σελίδα HTML. Η αίτηση σε εξυπηρετητή HTTP για αυτή τη σελίδα δημιουργεί συνολικά 28 διαφορετικές αιτήσεις για τις διάφορες εικόνες

Η σελίδα είναι προσβάσιμη από το URL <http://www.chania.gr>. Η εφαρμογή πλοήγησης, αφού μεταφέρει τον HTML κώδικα της σελίδας, μεταφέρει επίσης τις διάφορες εικόνες που αναφέρονται από τον HTML κώδικα. Για κάθε εικόνα δημιουργεί νέες αιτήσεις με αποτέλεσμα να δημιουργηθούν συνολικά 28 διαφορετικές αιτήσεις στον εξυπηρετητή. Με απλές αλλαγές του σχεδιασμού της σελίδας, μπορεί να αλλάξει δραματικά ο αριθμός των απαιτούμενων αιτήσεων. Για παράδειγμα εάν οι 5 σημαίες ενσωματωθούν σε ένα αρχείο εικόνας, μειώνονται οι αιτήσεις κατά 4. Επίσης εάν αφαιρεθούν οι “σφαίρες” στην ενότητα των νέων απαιτούνται 9 λιγότερες αιτήσεις. Για

τον τελικό χρήστη αυτές οι αλλαγές έχουν μικρή ή καμία σημασία επειδή αυτό που θα του εμφανιστεί στην οθόνη ουσιαστικά παραμένει το ίδιο. Στην πλευρά της συλλογής των δεδομένων όμως δημιουργείται μεγάλος όγκος δεδομένων με πολύ μικρή αξία για το σκοπό της ανάλυσης της επισκεψιμότητας του ιστοχώρου.

Το πρώτο βήμα στην ανάλυση των δεδομένων που συλλέγονται είναι η ομαδοποίηση των διαφόρων αιτήσεων σε σελιδοπροβολές, τις σελιδοπροβολές σε διακριτές επισκέψεις και τις επισκέψεις σε διακριτούς χρήστες (θα δοθούν στη συνέχεια συγκεκριμένοι ορισμοί). Αυτές οι ομαδοποιήσεις βασίζονται στα στοιχεία για τη διεύθυνση IP, το Referring URL και ενδεχόμενα cookies της κάθε αίτησης. Σε κάθε βήμα γίνονται ορισμένες παραδοχές (για παράδειγμα ο χρόνος που πρέπει να μεσολαβεί ανάμεσα σε διαδοχικές αιτήσεις ώστε να θεωρείται ότι ανήκουν σε μία επίσκεψη) ενώ λόγω τεχνικών περιορισμών στα παραπάνω στοιχεία, η όποια ανάλυση εμπεριέχει ένα σημαντικό περιθώριο σφάλματος. Για παράδειγμα θα μπορούσε να αναφερθεί ότι η αναγνώριση ενός χρήστη για τις αναλύσεις των ιστοχώρων είναι πολύ μικρότερης αξιοπιστίας σε σχέση με την αναγνώριση σε έναν ιστοχώρο τραπεζικών εργασιών μέσω Διαδικτύου, όπου λόγω της ευαισθησίας των δεδομένων μπορεί να επιβληθούν τεχνικοί περιορισμοί στους επισκέπτες.

Ξεκινώντας από τις αιτήσεις και προχωρώντας προς τους χρήστες, αυξάνεται με κάθε ανώτερο επίπεδο η αβεβαιότητα για την ακρίβεια της πληροφορίας που συλλέγεται, μειώνεται ο όγκος των δεδομένων που πρέπει να αποθηκευτούν και να επεξεργαστούν και αυξάνεται η σημαντικότητα και η αξία της πληροφορίας που μπορεί να αναλυθεί. Το Σχήμα 2.12 αναπαριστά την προηγούμενη πρόταση.



Οι σελιδοπροβολές, οι επισκέψεις και οι χρήστες θεωρούνται τα κύρια συστατικά μέρη στην ανάλυση των ιστοχώρων. Από μόνα τους δίνουν ήδη σημαντικές πληροφορίες για την επισκεψιμότητα του ιστοχώρου όπως αριθμός επισκέψεων στη διάρκεια του χρόνου, αριθμός σελιδοπροβολών ανά επίσκεψη, αριθμός χρηστών που χρησιμοποιούν τον ιστοχώρο, αριθμός επισκέψεων ανά χρήστη κλπ. Κάθε ένα από αυτά συνοδεύεται από ένα πλήθος χαρακτηριστικών που προσδίδουν μεγαλύτερη αξία στην πληροφορία. Για παράδειγμα μπορεί να είναι γνωστή η γεωγραφική τοποθεσία των χρηστών και να εξαχθούν στατιστικά συμπεράσματα για τα ενδιαφέροντα των χρηστών ανάλογα με την προέλευσή τους. Επίσης κάθε επίσκεψη χαρακτηρίζεται μεταξύ άλλων από μία συγκεκριμένη προέλευση (μηχανή αναζήτησης, φράσεις αναζήτησης) και από μία συγκεκριμένη διάρκεια που μπορεί να οδηγήσει σε συμπεράσματα για τα ενδιαφέροντα των χρηστών (από τις φράσεις αναζήτησης) ή για το εάν ο ιστοχώρος καλύπτει αυτά τα ενδιαφέροντα (από τις φράσεις αναζήτησης, από τη διάρκεια της επίσκεψης). Τέλος κάθε σελιδοπροβολή χαρακτηρίζεται μεταξύ άλλων από το περιεχόμενο που παρουσιάζεται στη σελίδα, τις φωτογραφίες και άλλα μέσα που συνοδεύουν το περιεχόμενο, τη θέση της σελιδοπροβολής στην ακολουθία των σελιδοπροβολών μιας επίσκεψης, τους συνδέσμους που επιλέγει ο χρήστης κλπ.

Ο συνδυασμός όλων αυτών των στοιχείων, στο βαθμό που αυτά θεωρούνται ακριβή, υποστηρίζει την ανάλυση της συμπεριφοράς των χρηστών του ιστοχώρου και αποτελεί έτσι μία πολύτιμη πηγή δεδομένων για την κατανόηση και αξιολόγηση του ιστοχώρου.

Στη συνέχεια θα οριστούν και θα περιγραφούν μετρικές και μεγέθη που χρησιμοποιούνται σε εμπορικά και ερευνητικά συστήματα ανάλυσης επισκεψιμότητας.

#### **2.3.4.1 Ορισμοί**

Η διαφημιστική αγορά ήταν η πρώτη που ενδιαφέρθηκε για το φορμαλισμό κοινώς αποδεκτών μετρικών για την περιγραφή των ιστοχώρων. Ενώ για παραδοσιακά μέσα διαφήμισης υπάρχει ένα σαφώς καθορισμένο πλαίσιο μέτρησης (οι εφημερίδες έχουν συγκεκριμένο αριθμό φύλων και αναγνωστικό κοινό, στην τηλεόραση υπάρχουν οι ζώνες μετάδοσης και οι εκπομπές που γίνονται και η απόδοση μετριέται με μετρήσεις τηλεθέασης), στο Διαδίκτυο υπήρχε (και σε ορισμένο βαθμό ακόμη υπάρχει) πλήρης σύγχυση για τους χρήστες που επισκέπτονται ιστοχώρους. Το 1996 έγινε με το [10] μία σημαντική προσπάθεια για το φορμαλισμό αποδεκτών μετρικών για το χώρο της διαφήμισης. Με την πάροδο του χρόνου η έρευνα επικεντρώθηκε και σε άλλους τομείς πέραν της διαφήμισης όπως μέτρηση ικανοποίησης πελατών σε online καταστήματα και αξιολόγηση της απόδοσης ενός ιστοχώρου στην πλοήγηση ή την ευκολία πρόσβασης στην πληροφορία που διατίθεται, οπότε και διευρύνθηκε περαιτέρω η ανάγκη καθορισμού συγκεκριμένης ορολογίας και

μετρικών. Αν και έχει διαμορφωθεί ένα σύνολο γενικά αποδεκτής ορολογίας ([10],[9],[34],[36]) στην πράξη μπορεί από τα ίδια πρωτογενή δεδομένα να προκύψουν διαφορετικές τιμές για τις μετρικές, και αυτό οφείλεται στα τεχνικά χαρακτηριστικά και τους αλγόριθμους που χρησιμοποιούνται από τα διάφορα συστήματα ανάλυσης.

Ακολουθεί μία απαρίθμηση χρησιμοποιούμενων μετρικών, μεγεθών και χαρακτηριστικών με τον ορισμό τους μαζί με μία περιγραφή για το πλαίσιο στο οποίο μπορεί να αναφέρονται.

#### **2.3.4.1.1 Hit – Αίτηση**

Ως Hit ορίζεται κάθε αίτηση προς τον εξυπηρετητή παγκόσμιου ιστού. Όπως εξηγήθηκε στην εισαγωγή του κεφαλαίου, οι αιτήσεις δεν παρουσιάζουν ιδιαίτερη αξία διότι εξαρτώνται σε μεγάλο βαθμό από το σχεδιασμό του ιστοχώρου. Σαν αυτόνομη μετρική παρουσιάζει ενδιαφέρον μόνο για εξυπηρετητές που χρησιμοποιούνται για εξειδικευμένες υπηρεσίες όπως εξυπηρετητές εικόνων διαφήμισης (πόσες φορές εμφανίστηκε κάποια διαφήμιση, πόσες φορές επιλέχτηκε) ή εξυπηρετητές video (πόσες φορές έγινε πρόσβαση σε κάποια ταινία).

#### **2.3.4.1.2 Pageview – Σελιδοπροβολή**

Μία σελίδα είναι κάθε σελίδα HTML, είτε στατική είτε δυναμικά παραγόμενη, που μπορεί να περιέχει κείμενο, εικόνες ή άλλα μέσα. Αποτελεί ένα από τα πιο σημαντικά στοιχεία μιας επίσκεψης, ειδικά σε συνδυασμό με στοιχεία για το περιεχόμενο κάθε σελίδας.

#### **2.3.4.1.3 Visit – Επίσκεψη**

Κάθε σειρά συνεχόμενων σελιδοπροβολών από έναν επισκέπτη ενός ιστοχώρου αποτελεί μία επίσκεψη. Όταν ένας επισκέπτης διακόψει τις σελιδοπροβολές για ορισμένο χρονικό διάστημα και ζητήσει μετά μία νέα σελιδοπροβολή, θεωρείται ότι εκτελεί μία νέα επίσκεψη. Συχνά χρησιμοποιείται και ο όρος session, αν και με sessions χαρακτηρίζονται επίσης ακολουθίες σελιδοπροβολών της ίδιας επίσκεψης αλλά με διαφορετική σημασιολογία στον ιστοχώρο. Στο [37] υπολογίστηκε σαν μέσος όρος δράσης ανά επίσκεψη τα 9,3 λεπτά και υπολόγισαν τα 25,5 λεπτά ως ικανός χρόνο αδράνειας για τη διακοπή της επίσκεψης. Από το τελευταίο προκύπτει ο χρόνος των 30 λεπτών που χρησιμοποιείται σχεδόν αποκλειστικά από όλα τα συστήματα σαν χρόνος που πρέπει να μεσολαβήσει χωρίς σελιδοπροβολές από έναν επισκέπτη ώστε να θεωρηθεί ότι διακόπηκε η επίσκεψη. Βέβαια ο μέσος χρόνος δράσης εξαρτάται σε πολύ μεγάλο βαθμό από τον ίδιο τον ιστοχώρο (είδους πληροφορίας που παρέχει, ταχύτητα μεταφοράς των δεδομένων κλπ.).

#### **2.3.4.1.4 Visitor – Επισκέπτης**

Ο όρος “Επισκέπτης” αναφέρεται σε ένα άτομο που επισκέπτεται κάποιον ιστοχώρο. Πρέπει να δοθεί ιδιαίτερη προσοχή ώστε να μην συμπερένεται ότι επισκέψεις από τον ίδιο υπολογιστή ή από

την ίδια εφαρμογή πλοήγησης αναφέρονται στο ίδιο άτομο (για παράδειγμα σε Internet Cafes μοιράζονται πολλά διαφορετικά άτομα τον ίδιο υπολογιστή). Στο [10] οι επισκέπτες χωρίζονται στις ακόλουθες κατηγορίες:

- **Unidentified Visitor:** γενικά ένα άτομο που επισκέπτεται έναν ιστοχώρο. Δεν υπάρχει καμία άλλη πληροφορία διαθέσιμη για αυτό.
- **Session Visitor:** υπάρχει διαθέσιμο ένα χαρακτηριστικό (π.χ. Cookie) ή μπορεί να βγει συμπέρασμα για ένα χαρακτηριστικό (π.χ. Συνδυασμός από διεύθυνση IP και τύπος εφαρμογής πλοήγησης) που αναγνωρίζει τις αποκρίσεις του επισκέπτη στην ίδια επίσκεψη
- **Tracked Visitor:** υπάρχει διαθέσιμο χαρακτηριστικό (π.χ. Cookie) που επιτρέπει την αναγνώριση του επισκέπτη σε περισσότερες της μίας επίσκεψης στον ιστοχώρο. Δεν υπάρχει άλλη πληροφορία διαθέσιμη εκτός από το χαρακτηριστικό
- **Identified Visitor:** όπως το tracked visitor αλλά με διαθέσιμα επιπλέον στοιχεία για τον επισκέπτη, όπως το όνομά του ή η περιοχή διαμονής του

Όπως θα γίνεται αντιληπτό η αναγνώριση του χρήστη είναι η πλέον πολύτιμη πληροφορία για τους εμπλεκόμενους στην ανάλυση επισκεψιμότητας. Η αναγνώριση του χρήστη είναι αναγκαία για την παροχή εξατομικευμένων υπηρεσιών και ο συνδυασμός της γνώσης περαιτέρω πληροφοριών για το πρόσωπο των επισκεπτών (π.χ. παρελθόντες παραγγελίες για online καταστήματα) δίνει τη δυνατότητα για ένα προς ένα μάρκετινγκ στο οποίο προσβλέπουν οι εμπορικές εταιρίες. Σε αυτό το σημείο δημιουργούνται οι προβληματισμοί και οι κίνδυνοι για τη χρήση προσωπικών δεδομένων των χρηστών.

#### **2.3.4.1.5 Duration Time**

Αναφέρεται στο χρόνο που μεσολαβεί ανάμεσα σε δύο γεγονότα, όπως ο χρόνος μέχρι την επόμενη σελιδοπροβολή ή την επόμενη επίσκεψη.

#### **2.3.4.1.6 Referrer – Προέλευση**

Το referrer είναι πεδίο που προβλέπεται από το HTTP πρωτόκολλο και αναφέρεται στο URL από το προήλθε μία συγκεκριμένη αίτηση. Ενδιαφέρον παρουσιάζει ειδικά όταν το URL βρίσκεται εκτός του εξεταζόμενου ιστοχώρου, οπότε δίνεται πληροφορία για το ποιοι άλλοι ιστοχώροι και σε ποιες σελίδες έχουν σύνδεσμο με τον εξεταζόμενο ιστοχώρο. Όταν ο referrer αναφέρεται σε μηχανή αναζήτησης υπάρχει πολλές φορές η πληροφορία για την αναζήτηση που έκανε ο επισκέπτης πριν φτάσει στον εξεταζόμενο ιστοχώρο (οπότε και πληροφορία για τα ενδιαφέροντα του επισκέπτη).

Ένα παράδειγμα εξωτερικού referrer από τη μηχανή αναζήτησης Google [39] είναι

<http://www.google.com/search?q=crete+hotel&hl=en&lr=&ie=UTF-8&start=10&sa=N>

Από το παραπάνω URL και με γνώση των παραμέτρων που χρησιμοποιεί η συγκεκριμένη μηχανή αναζήτησης προκύπτουν τα ακόλουθα συμπεράσματα:

- πρόκειται για έναν χρήστη που έκανε αναζήτηση στη μηχανή Google.
- Ο χρήστης έψαχνε για τη φράση “crete hotel” (q=crete+hotel),
- χρησιμοποίησε το αγγλικό interface της μηχανής (hl=en) και
- επέλεξε το σύνδεσμο από τη σελίδα που ξεκινάει από το 10ο αποτέλεσμα (start=10).

Τα referrers που είναι εντός του εξεταζόμενου ιστοχώρου βοηθούν στην ανακατασκευή της ακριβής ακολουθίας σελιδοπροβολών μίας επίσκεψης.

#### **2.3.4.1.7 Clickthrough**

Χρησιμοποιείται συνήθως για τη μέτρηση της απόδοσης διαφημιστικών μηνυμάτων. Εκφράζει το λόγο των αντιδράσεων σε ένα διαφημιστικό μήνυμα (click στη διαφήμιση) προς τον αριθμό εμφάνισης του μηνύματος. Σε περιπτώσεις όπου η διαφήμιση είναι ένα αρχείο (π.χ. μία εικόνα) με τον αριθμό των αιτήσεων προς τον εξυπηρετητή για την εικόνα και τον αριθμό των αιτήσεων για τη σελίδα στην οποία συνδέει η διαφήμιση.

#### **2.3.4.1.8 Client Domain**

Είναι το όνομα περιοχής (Domain Name) που αντιστοιχεί στη διεύθυνση IP του υπολογιστή του επισκέπτη. Μπορεί να δώσει μία ένδειξη για τη χώρα προέλευσης του επισκέπτη (κατάληξη του ονόματος π.χ. .gr, .de κλπ) καθώς και την ευρύτερη περιοχή διαμονής όταν πρόκειται για πρόσβαση από μη μόνιμη σύνδεση (οι εταιρείες παροχής υπηρεσιών διαδικτύου πολλές φορές δίνουν στο όνομα της IP διεύθυνσης στοιχεία για την πόλη ή περιοχή στην οποία υπάγεται). Το στοιχείο μπορεί να χρησιμοποιηθεί για την αναγνώριση επισκέψεων από μεγάλους οργανισμούς με πρόσβαση στο διαδίκτυο. Για παράδειγμα μία επίσκεψη από διεύθυνση με κατάληξη yperth.gr δίνει μία ισχυρή ένδειξη για επίσκεψη από υπολογιστή του Υπουργείου Παιδείας της Ελλάδας στο οποίο ανήκει αυτή η ονομασία περιοχής.

Το όνομα περιοχής μπορεί να είναι και εντελώς άχρηστο ή παραπλανητικό όπως για παράδειγμα για τα ονόματα με κατάληξη aol.com, που ανήκουν στη μεγάλη εταιρεία παροχής υπηρεσιών AOL, επειδή δίνεται αυτή η κατάληξη ανεξάρτητα από τη χώρα από την οποία γίνεται η σύνδεση.

#### **2.3.4.1.9 User Agent**

Επίσης ένα πεδίο που προβλέπεται από το HTTP πρωτόκολλο. Αναφέρεται στην εφαρμογή πλοήγησης που χρησιμοποιεί ο χρήστης. Δίνει μία ένδειξη για τις τεχνικές δυνατότητες που υποστηρίζονται από την εφαρμογή του χρήστη, οπότε μπορεί να προσαρμοστούν οι τεχνολογίες που χρησιμοποιούνται σε μία σελίδα ώστε να μπορεί να εμφανιστεί η σελίδα στο χρήστη. Είναι επίσης πολύ χρήσιμη πληροφορία για τις προσβάσεις από εφαρμογές αυτόματης επίσκεψης (εφαρμογές δεικτοδότησης από εφαρμογές αναζήτησης για παράδειγμα) ώστε να αποκλειστούν οι συγκεκριμένες πληροφορίες από την ανάλυση επειδή δεν αναφέρονται σε πραγματικούς χρήστες.

Επειδή αυτό το πεδίο δεν είναι αναγκαίο για την επικοινωνία των εφαρμογών με τους εξυπηρετητές υπάρχουν πολλές περιπτώσεις όπου είτε δε μεταφέρεται καθόλου, είτε δεν ανταποκρίνεται στην πραγματικότητα (παράδειγμα εφαρμογές αυτόματης μεταφοράς ολόκληρου ιστοχώρου).

#### **2.3.4.1.10 Cookies**

Όπως αναφέρθηκε είναι μικρές ποσότητες δεδομένων που αποθηκεύονται από την εφαρμογή πλοήγησης και στέλνονται στον εξυπηρετητή από τον οποίο προήλθαν με κάθε νέα αίτηση. Έχουν διάφορες χρήσεις αλλά για την ανάλυση επισκεψιμότητας χρησιμοποιούνται για την αναγνώριση επισκέψεων (session cookies) και επισκεπτών (visitor cookie).

Στην πρώτη περίπτωση αποθηκεύεται με την πρώτη πρόσβαση ένα μοναδικό χαρακτηριστικό για τη διάρκεια μιας επίσκεψης το οποίο σβήνεται με την πάροδο προκαθορισμένου χρόνου χωρίς αιτήσεις. Επειδή με κάθε αίτηση η εφαρμογή στέλνει και αυτό το χαρακτηριστικό το σύστημα ανάλυσης είναι σε θέση να διακρίνει με μεγάλη βεβαιότητα τις αιτήσεις που ανήκουν στην ίδια επίσκεψη.

Στη δεύτερη περίπτωση το χαρακτηριστικό αυτό διατηρείται και πέραν της επίσκεψης και ξαναστέλνεται σε μία επόμενη επίσκεψη, οπότε το σύστημα ανάλυσης είναι σε θέση να γνωρίζει ότι η επόμενη επίσκεψη έγινε από τον ίδιο υπολογιστή (ή λογαριασμό χρήστη) όπως και η πρώτη επίσκεψη. Ειδικά όταν χρησιμοποιείται η μέθοδος συλλογής με την ενσωμάτωση ειδικής εικόνας στις σελίδες, επειδή όλες αυτές οι εικόνες προέρχονται από τον ίδιο εξυπηρετητή και η πρόσβαση εκεί γίνεται μέσω διαφορετικών ιστοχώρων, το σύστημα ανάλυσης είναι σε θέση να παρακολουθεί επισκέψεις του χρήστη σε διαφορετικούς ιστοχώρους.

Η αυξανόμενη ευαισθησία των χρηστών για τη χρήση των δεδομένων που συλλέγονται με αυτόν τον τρόπο (ασχέτως εάν είναι μεγάλη ευκολία ειδικά για την παροχή καλύτερων υπηρεσιών στους επισκέπτες ακόμα και όταν δε χρησιμοποιούνται για την “παρακολούθησή” τους) έχουν αποφέρει

(δικαιολογημένα) στα cookies πολύ κακή φήμη και έτσι είναι πολλοί οι χρήστες που δεν επιτρέπουν τη χρήση τους. Ειδικά σε ιστοχώρους με θεματολογία που απευθύνονται σε επισκέπτες με τεχνικές γνώσεις παρατηρείται σε αυξημένο ποσοστό των επισκεπτών η μη αποδοχή των cookies [38].

#### **2.3.4.1.11 Language Header**

Οι εφαρμογές πλοήγησης μπορεί να στείλουν με κάθε αίτηση και στοιχεία για τη γλώσσα που είναι προτιμώμενη από τον επισκέπτη, οπότε υπάρχει ένδειξη τη γλώσσα που μιλάει ο χρήστης. Είναι πάντως ένα στοιχείο που καθορίζεται συνήθως αυτόματα κατά την εγκατάσταση της εφαρμογής πλοήγησης σύμφωνα με τις ρυθμίσεις του υπολογιστικού συστήματος (για παράδειγμα σε ένα σύστημα με Ελληνικά Windows θα ρυθμιστεί αυτόματα η γλώσσα Ελληνικά ως προτιμώμενη). Λίγοι είναι οι χρήστες που έχουν τη γνώση για το πως μπορούν να αλλάξουν αυτή τη ρύθμιση.

Για ιστοχώρους που παρέχουν πληροφορία για περισσότερες από μία γλώσσες αυτή η ρύθμιση είναι σημαντική για την αυτόματη προσαρμογή στις προτιμήσεις του επισκέπτη.

#### **2.3.4.1.12 Άλλα Μεγέθη**

Πέραν των μεγεθών που περιγράφηκαν ως τώρα, υπάρχει μία πληθώρα άλλων μεγεθών που μπορεί να καταγραφούν. Τα περισσότερα όμως, πέραν του ότι, όπως αναφέρθηκε, δεν είναι τυποποιημένα, εξαρτώνται σε μεγάλο βαθμό από τα χαρακτηριστικά του ιστοχώρου και δεν έχουν γενική εφαρμογή.

Σε ιστοχώρους όπου υπάρχει δυνατότητα αναζήτησης μπορεί να καταγραφούν οι φράσεις που χρησιμοποίησαν οι χρήστες για αναζήτηση, ο αριθμός των αποτελεσμάτων που επεστράφησαν, ο αριθμός των αποτελεσμάτων που επιλέχθηκαν από τον επισκέπτη κλπ.

Σε Online καταστήματα έχουν νόημα οι όροι Conversion, που αναφέρεται στο ποσοστό των επισκεπτών που εκπληρώνουν κάποιο προκαθορισμένο από το κατάστημα στόχο (μπορεί να είναι η εγγραφή τους στο κατάστημα ή η αγορά ενός προϊόντος), Retention που αναφέρεται σε ήδη πελάτες του καταστήματος που επανέρχονται με νέες παραγγελίες, και Loyalty που είναι πελάτες με συχνές επισκέψεις και αγορές από το κατάστημα (πιστοί οπαδοί του καταστήματος). Σε όλα αυτά τα στάδια υπάρχουν και μετρικές που χαρακτηρίζουν τους πελάτες που δεν επισκέπτονται πάλι τον ιστοχώρο για διάφορους λόγους [36].

#### **2.3.4.2 Αξιοπιστία Δεδομένων**

Ολόκληρος ο μηχανισμός για την καταγραφή των προσβάσεων των επισκεπτών βασίζεται στις αιτήσεις που γίνονται από τις εφαρμογές πλοήγησης στον εξυπηρετητή HTTP. Στην περίπτωση που για οποιοδήποτε λόγο η πληροφορία του ιστοχώρου εμφανιστεί στον επισκέπτη χωρίς να γίνει



αίτηση στον εξυπηρετητή, αποκρύπτεται από το σύστημα ανάλυσης η σχετική πληροφορία και δημιουργούνται στρεβλώσεις στα αποτελέσματα της ανάλυσης.

Περιπτώσεις όπου υπάρχει μία τέτοια συμπεριφορά υπάρχουν πολλές και οφείλονται κυρίως σε μηχανισμούς ενδιάμεσης αποθήκευσης των δεδομένων (caching), που άλλες φορές γίνεται για λόγους ταχύτητας και άλλες επειδή, λόγω τεχνικών περιορισμών, είναι ο μοναδικός τρόπος πρόσβασης στο δίκτυο.

Στις επόμενες υποενότητες παρουσιάζονται οι δύο κύριες περιπτώσεις που οδηγούν στην εμφάνιση πληροφοριών χωρίς τη διενέργεια σχετικής αίτησης στον εξυπηρετητή και οι μηχανισμοί που μπορούν να χρησιμοποιηθούν για την μερική αποφυγή.

#### **2.3.4.2.1 Bots και Crawlers**

Οι προσβάσεις σε έναν ιστοχώρο μπορούν να χωριστούν σε δύο κυρίως κατηγορίες. Η πρώτη και η σημαντικότερη αφορά τους απλούς επισκέπτες του ιστοχώρου και η δεύτερη αφορά τις προσβάσεις που γίνονται από λογισμικό αυτόματης πρόσβασης σε ιστοχώρους και δεν αφορά την πραγματική επίσκεψη από συγκεκριμένο άτομο. Τέτοιες προσβάσεις είναι κυρίως οι προσβάσεις από τις εφαρμογές δεικτοδότησης (indexing) (bots, crawlers) των μηχανών αναζήτησης, αλλά μπορεί να συμπεριληφθούν και οι εφαρμογές που μεταφέρουν και αποθηκεύουν με αυτόματο τρόπο τοπικά έναν ολόκληρο ιστοχώρο.

Ενώ η πρώτη κατηγορία των απλών επισκεπτών είναι αυτή που ενδιαφέρει κυρίως στην ανάλυση επισκεψιμότητας, η δεύτερη κατηγορία δεν αναφέρεται σε πραγματικές εμφανίσεις της πληροφορίας του ιστοχώρου και μπορεί να οδηγήσει με αυτόν τον τρόπο σε εσφαλμένα συμπεράσματα.

Ειδικά οι εφαρμογές δεικτοδότησης των μηχανών αναζήτησης μπορεί να δημιουργήσουν μεγάλη κίνηση σε έναν ιστοχώρο, και αυτό αναφέρεται τόσο στον όγκο δεδομένων που μεταφέρεται όσο και στις μετρήσεις σε επισκέψεις και σελιδοπροβολές που εμφανίζεται να γίνονται. Για αυτόν ακριβώς το λόγο οι προσβάσεις από τις εφαρμογές δεικτοδότησης των πιο διαδεδομένων μηχανών αναζήτησης αποφεύγουν να προβούν σε πολλές προσβάσεις σε μικρό χρονικό διάστημα ώστε να μη φορτώνεται ο εξυπηρετητής με μεγάλο υπολογιστικό φόρτο και να μη δημιουργείται μεγάλη σε όγκο κίνηση στο δίκτυο. Αυτή η συμπεριφορά όμως δεν είναι δεσμευτική για τις μηχανές αναζήτησης και υπάρχουν εφαρμογές δεικτοδότησης που δεν ακολουθούν αυτούς τους κανόνες.

Σε ιστοχώρους όπου περιλαμβάνεται μεγάλος όγκος πληροφορίας που κατανέμεται σε πολλές σελίδες, οι εφαρμογές δεικτοδότησης που ακολουθούν τους κανόνες για την αποφυγή μεγάλου

φόρτου φαίνονται σαν επισκέψεις που διαρκούν από πολλές ώρες μέχρι και πολλές ημέρες, ενώ οι εφαρμογές που δεν ακολουθούν αυτούς τους κανόνες δημιουργούν την εντύπωση μιας μεγάλης σε όγκο πληροφορίας (σελιδοπροβολές, όγκος δεδομένων) επίσκεψης από κάποιον χρήστη.

Σε κάθε περίπτωση κατά την ανάλυση των προσβάσεων είναι επιθυμητή η διάκριση ανάμεσα σε προσβάσεις από πραγματικούς επισκέπτες και μη. Η αναγνώριση των διαφόρων εφαρμογών που δεν αντιστοιχούν σε πραγματικούς χρήστες μπορεί να βασιστεί στα ακόλουθα στοιχεία:

- IP διεύθυνση: οι μηχανές αναζήτησης χρησιμοποιούν συγκεκριμένες διευθύνσεις IP για τη δεικτοδότηση των ιστοχώρων. Οι διευθύνσεις αυτές μπορεί να αλλάξουν με την πάροδο του χρόνου.
- Ονόματα DNS: τα ονόματα DNS από τα οποία γίνεται η δεικτοδότηση ακολουθούν ανά μηχανή αναζήτησης συγκεκριμένο σχήμα (για παράδειγμα το Google xxx.googlebot.com)
- User-Agent: στο HTTP πεδίο που αναφέρεται στην εφαρμογή πλοήγησης εμφανίζεται πολλές φορές το όνομα της εφαρμογής δεικτοδότησης

#### **2.3.4.2.2 Εξυπηρετητές Proxy**

Οι εξυπηρετητές proxy λειτουργούν ως μεσολαβητές στην επικοινωνία μέσω δικτύου ανάμεσα στις εφαρμογές πελάτη εξυπηρετητή. Οι πελάτες (δηλαδή οι εφαρμογές πλοήγησης) αντί να ανοίξουν απ' ευθείας σύνδεση με τον εξυπηρετητή που επιθυμούν, συνδέονται με τον proxy ο οποίος αναλαμβάνει να συνδεθεί με τον εξυπηρετητή. Τέτοιου είδους proxy υπάρχουν για διάφορα πρωτόκολλα επικοινωνίας όπως FTP, SSL, GOPHER ή και HTTP. Στην εξεταζόμενη περίπτωση ενδιαφερόμαστε για το πρωτόκολλο HTTP το οποίο χρησιμοποιείται για την πλοήγηση στον Παγκόσμιο Ιστό, αλλά ισχύουν όμοια πράγματα και για τα υπόλοιπα πρωτόκολλα.

Ο εξυπηρετητής HTTP proxy αφού έχει λάβει από την εφαρμογή πλοήγησης την αίτηση για έναν πόρο σε έναν εξυπηρετητή HTTP, αναλαμβάνει να προωθήσει την αίτηση στον HTTP εξυπηρετητή και να επιστρέψει στην εφαρμογή το όποιο αποτέλεσμα. Ο εξυπηρετητής HTTP καταγράφει έτσι μία αίτηση από τον proxy και όχι από τον πραγματικό επισκέπτη.

Οι proxy συνήθως λειτουργούν και ως cache δεδομένων για τα δεδομένα για τα οποία ήδη έχει γίνει αίτηση μέσω αυτών, οπότε στην περίπτωση που γίνει αίτηση για έναν πόρο που υπάρχει ήδη στον proxy δεν γίνεται καμία πρόσβαση στον HTTP εξυπηρετητή αλλά επιστρέφεται στην εφαρμογή πλοήγησης το αντίστοιχο περιεχόμενο της cache. Σε αυτήν την περίπτωση, παρόλο που στον τελικό χρήστη εμφανίζεται το περιεχόμενο μιας σελίδας από τον HTTP εξυπηρετητή, αυτός δεν

καταγράφει καμία εξωτερική πρόσβαση.

Οι εξυπηρετητές proxy είναι πολύ διαδεδομένοι επειδή προσφέρουν μία φαινομενική αύξηση της ταχύτητας πλοήγησης, επειδή πολλά δεδομένα είναι ήδη αποθηκευμένα στην cache τους, και δε χρειάζεται να «ταξιδέψουν» όλη την πραγματική διαδρομή από τον HTTP εξυπηρετητή στον πελάτη. Υπάρχουν όμως περιπτώσεις όπου αποτελούν το μόνο τρόπο σύνδεσης με το Διαδίκτυο, ειδικά όταν ένα εσωτερικό δίκτυο είναι αποκομμένο για λόγους ασφαλείας με το Διαδίκτυο και επιτρέπεται μόνο σε ένα μηχάνημα να συνδέεται προς τα έξω. Σε αυτό το μηχάνημα γίνεται εγκατάσταση του proxy και χρησιμοποιείται από όλα τα υπόλοιπα για την πρόσβαση στο Διαδίκτυο.

#### **2.3.4.2.3 Caching Εφαρμογών Πλοήγησης**

Έναν τοπικό μηχανισμό caching διατηρούν επίσης οι εφαρμογές πλοήγησης. Χρησιμοποιείται για την αποφυγή μεταφοράς δεδομένων στην περίπτωση που αυτά είχαν ήδη μεταφερθεί στο παρελθόν. Η cache είναι συνήθως δύο επιπέδων και αναφέρεται τόσο στη δευτερεύουσα όσο και στην κύρια μνήμη.

Το ίδιο πρόβλημα που παρουσιάζεται με την cache στους εξυπηρετητές proxy εμφανίζεται στην cache των εφαρμογών πλοήγησης. Δηλαδή στην περίπτωση που ο πόρος που ζητείται υπάρχει ήδη στην cache της εφαρμογής, δε γίνεται αίτηση στον εξυπηρετητή HTTP και έτσι δεν μπορεί να καταγραφεί η πρόσβαση.

#### **2.3.4.2.4 Αποφυγή Μηχανισμών Caching**

Ο έλεγχος του μηχανισμού cache τόσο στις εφαρμογές πλοήγησης όσο και στους εξυπηρετητές proxy μπορεί να γίνει με ειδικά header του HTTP πρωτοκόλλου που στέλνονται από τους εξυπηρετητές. Με αυτά τα headers μπορεί να καθοριστούν για κάθε URL στοιχεία όπως ημερομηνία λήξης ή ακόμα να καθοριστεί ότι δεν πρέπει να αποθηκευτεί καθόλου μία συγκεκριμένη απάντηση σε cache.

Μία πρώτη προσέγγιση θα ήταν να στέλνονται για όλες τις αιτήσεις προς τον HTTP εξυπηρετητή τα κατάλληλα header ώστε να μην αποθηκεύονται καθόλου οι απαντήσεις σε cache. Αυτό όμως θα ακύρωνε όλα τα οφέλη που προσφέρουν οι cache σε ταχύτητα πλοήγησης και θα είχε άλλες δυσμενείς επιπτώσεις προς τους επισκέπτες.

Μία δεύτερη προσέγγιση θα ήταν να στέλνονται αυτά τα headers μόνο για τις κύριες σελίδες του ιστοχώρου, ενώ για όλα τα υπόλοιπα αντικείμενα όπως οι διάφορες εικόνες, να μην στέλνονται. Και σε αυτήν την περίπτωση όμως δημιουργείται όμως επιπλέον δικτυακή κίνηση που μπορεί να έχει

δυσμενές επιπτώσεις στους επισκέπτες.

Η περίπτωση όπου για την στατιστική ανάλυση χρησιμοποιούνται ειδικές εικόνες ενσωματωμένες στο κυρίως περιεχόμενο επιτρέπει την καλύτερη προσέγγιση, επειδή αρκεί σε αυτήν την περίπτωση να μη γίνεται cache μόνο σε αυτήν την ειδική εικόνα. Εάν μάλιστα πρόκειται για μία συχνά χρησιμοποιούμενη αόρατη GIF εικόνα το μέγεθος της πληροφορίας που μεταφέρεται ελαχιστοποιείται. Αυτή η περίπτωση όμως πάσχει στην περίπτωση που ο χρήστης έχει επιλέξει να μη βλέπει εικόνες.

Παρόλο που με τις διάφορες μεθόδους που παρουσιάστηκαν παραπάνω μπορεί να διατηρηθεί σε μεγάλο μέρος ο έλεγχος επί των προσβάσεων στον εξυπηρετητή HTTP, δεν υπάρχει τεχνική λύση που να εγγυάται τον απόλυτο έλεγχο ως προς ποια πληροφορία τελικά βλέπουν οι επισκέπτες. Ακόμα και στην περίπτωση που με τα HTTP headers καλούνται οι εφαρμογές να μην αποθηκεύσουν τίποτα σε cache, δεν μπορεί τεχνικά να επιβληθεί στις εφαρμογές η τήρηση αυτών των headers.

Η αξιόπιστη ανακατασκευή σελιδοπροβολών που δεν καταγράφηκαν κατά τη συλλογή των πρωτογενών δεδομένων είναι ένα από τα πιο βασικά προβλήματα με τα οποία ασχολούνται οι σχετικές έρευνες για την ανάλυση επισκεψιμότητας. Παραδείγματα αντιμετώπισης τέτοιων προβλημάτων θα δοθούν στην επόμενη ενότητα

### **2.3.5 Προεπεξεργασία Δεδομένων (Data Preprocessing)**

Στην ενότητα 2.3.3 έγινε αναφορά για τους διάφορους τρόπους συλλογής των πρωτογενών δεδομένων από την κίνηση σε ιστοχώρους. Ο κοινός παρανομαστής σε όλες τις μεθόδους είναι το πρωτόκολλο HTTP που είναι η καρδιά της επικοινωνίας με τους εξυπηρετητές παγκόσμιου ιστού. Όλες οι μέθοδοι στην ουσία συλλέγουν τα στοιχεία που προκύπτουν από την επικοινωνία με τη χρήση αυτού του πρωτοκόλλου.

Στην ενότητα 2.3.4 αναφέρθηκαν χαρακτηριστικά μεγέθη που ενδιαφέρουν στην ανάλυση επισκεψιμότητας. Έγινε σαφές ότι η βασική μονάδα μέτρησης του HTTP πρωτοκόλλου, που είναι μία αίτηση στον εξυπηρετητή, έχει πολύ μικρή αξία στην κατανόηση της συμπεριφοράς των χρηστών, επειδή εξαρτάται περισσότερο από το σχεδιασμό του ιστοχώρου παρά από τις επιλογές που κάνει ο επισκέπτης στην πλοήγηση και αναζήτηση πληροφορίας.

Στην ανάλυση επισκεψιμότητας η βασική μονάδα μέτρησης είναι η εμφάνιση στην οθόνη του επισκέπτη μιας σελίδας υπερκειμένου, η οποία είναι το αποτέλεσμα τουλάχιστον μίας αίτησης προς τον εξυπηρετητή.

Ας πάρουμε ως παράδειγμα την αρχική σελίδα του ιστοχώρου τουρισμού <http://www.crete.tournet.gr> [43] που απεικονίζεται στην



Εικόνα 2.2 Αρχική σελίδα του <http://www.crete.tournet.gr> [43]

Η προβολή αυτής της σελίδας δημιουργεί στα αρχεία καταγραφής στον εξυπηρετητή τις εγγραφές σε μορφή Combined Log Format όπως φαίνονται στον πίνακα 2.7 (οι εναλλαγές των πεδίων του CLF εμφανίζονται με εναλλαγές στο φόντο των τιμών.)

<i>A/A</i>	<i>Combined Log Format</i>
	host user ident datetime request resp_code resp_size referrer useragent
1	sk.ced.tuc.gr - - [03/Oct/2002:01:29:35 +0300] "GET / HTTP/1.1" 200 10870 "-" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2b) Gecko/20020923"
2	sk.ced.tuc.gr - - [03/Oct/2002:01:29:35 +0300] "GET /ct-styles.css HTTP/1.1" 200 5351 "http://www.crete.tournet.gr/" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2b) Gecko/20020923"
3	sk.ced.tuc.gr - - [03/Oct/2002:01:29:35 +0300] "GET /script.js HTTP/1.1" 200 1023 "http://www.crete.tournet.gr/" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2b) Gecko/20020923"
4	sk.ced.tuc.gr - - [03/Oct/2002:01:29:35 +0300] "GET /dynimg/d_hotels-el.gif HTTP/1.1" 200 4094 "http://www.crete.tournet.gr/" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2b) Gecko/20020923"
5	sk.ced.tuc.gr - - [03/Oct/2002:01:29:36 +0300] "GET /dynimg/beach1.jpg HTTP/1.1" 200 11875 "http://www.crete.tournet.gr/" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2b) Gecko/20020923"
6	sk.ced.tuc.gr - - [03/Oct/2002:01:29:36 +0300] "GET /dynimg/minoan1.jpg HTTP/1.1" 200 14238 "http://www.crete.tournet.gr/" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2b) Gecko/20020923"
7	sk.ced.tuc.gr - - [03/Oct/2002:01:29:36 +0300] "GET /images/logosmall.gif HTTP/1.1" 200 6115 "http://www.crete.tournet.gr/" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2b) Gecko/20020923"
8	sk.ced.tuc.gr - - [03/Oct/2002:01:29:36 +0300] "GET /advs/chania.jpg HTTP/1.1" 200 12441 "http://www.crete.tournet.gr/" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2b) Gecko/20020923"
9	sk.ced.tuc.gr - - [03/Oct/2002:01:29:36 +0300] "GET /images/flags_small.gif HTTP/1.1" 200 438 "http://www.crete.tournet.gr/" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2b) Gecko/20020923"
10	sk.ced.tuc.gr - - [03/Oct/2002:01:29:36 +0300] "GET /images/transparent.gif HTTP/1.1" 200 258 "http://www.crete.tournet.gr/" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2b) Gecko/20020923"
11	sk.ced.tuc.gr - - [03/Oct/2002:01:29:36 +0300] "GET /advs/intourisme.jpg HTTP/1.1" 200 11302 "http://www.crete.tournet.gr/" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.2b) Gecko/20020923"

Πίνακας 2.7 Εγγραφές σε Combined Log Format (CLF) από την εμφάνιση της αρχικής σελίδας του <http://www.crete.tournet.gr>. Οι εναλλαγές των πεδίων του CLF εμφανίζονται με εναλλαγές στο φόντο των τιμών.

Η πρώτη εγγραφή αναφέρεται σε καθ' αυτή την σελίδα (request /). Η δεύτερη και τρίτη εγγραφή αναφέρονται στο Cascading Style Sheet και σε ένα JavaScript αρχείο που χρησιμοποιούνται στη σελίδα. Οι εγγραφές 4,5 και 6 αναφέρονται στην περιοχή 3 της εικόνας 2.2 που εναλλάσσονται. Η εγγραφή 7 είναι το λογότυπο της περιοχής 1, η εγγραφή 9 είναι οι σημαίες της περιοχής 2, οι εγγραφές 8 και 11 αναφέρονται στις εικόνες της περιοχής 4 και 5, ενώ η εγγραφή 10 αναφέρεται σε μία εικόνα αόρατη.

Κατ' αρχήν παρατηρούμε ότι σε κάθε εγγραφή, ο μεγαλύτερος όγκος στα δεδομένα που αποθηκεύονται αναφέρεται στο τελευταίο πεδίο του CLF που είναι η αναφορά της εφαρμογής

πλοήγησης. Επίσης παρατηρούμε ότι εκτός από την πρώτη εγγραφή, όλες οι άλλες έχουν στο πεδίο του referrer το URI από το οποίο προέρχονται οι αιτήσεις, και είναι το URI της σελίδας στην οποία αναφερόμαστε. Κάθε εγγραφή καταλαμβάνει περίπου 200 χαρακτήρες. Εάν μάλιστα περιγράψαμε μία σελίδα του ιστοχώρου που βρίσκεται σε μεγαλύτερο βάθος, το πεδίο του referrer θα ήταν πολύ μεγαλύτερο.

Από όλες αυτές τις εγγραφές, η μόνη που έχει πραγματική αξία είναι η πρώτη. Γνωρίζοντας ότι αναφερόμαστε στην αρχική σελίδα του συγκεκριμένου ιστοχώρου, όλες οι υπόλοιπες εγγραφές δε μας δίνουν περισσότερη πληροφορία εκτός του ότι “έγινε σελιδοπροβολή της αρχικής σελίδας”.

Σε απόλυτα μεγέθη όγκου δεδομένων, η προηγούμενη πρόταση περιγράφεται με τα 152 Byte της πρώτης εγγραφής, ενώ στο αρχείο καταγραφής αποθηκεύονται περίπου 2150 Bytes!

Οπότε το πρώτο βήμα στην προεπεξεργασία των δεδομένων αποτελείται από την απαλοιφή των αιτήσεων που δεν αναφέρονται σε αυτόνομες σελίδες. Παράλληλα απαλείφονται και άλλου είδους εγγραφές που δεν ανταποκρίνονται σε πραγματικές σελίδες του ιστοχώρου, όπως είναι αιτήσεις για σελίδες που δεν υφίστανται στον ιστοχώρο (για παράδειγμα όταν αναδιοργανώνεται ο ιστοχώρος και παραμένουν σύνδεσμοι σε διαγραφέντες σελίδες).

#### **2.3.5.1 Διάκριση Σελιδοπροβολών**

Η διάκριση ανάμεσα σε αιτήσεις που ανταποκρίνονται σε σελίδες και σε αυτές που δεν είναι σελίδες, συνήθως γίνεται με βάση την κατάληξη του πόρου της αίτησης που ζητείται. Έτσι αιτήσεις με κατάληξη .html ή .htm θεωρούνται σελίδες επειδή αυτή είναι η καθιερωμένη κατάληξη για στατικές σελίδες HTML. Όμως, όπως ορίζεται στο HTTP πρωτόκολλο [29], το πραγματικό περιεχόμενο μιας απάντησης από τον εξυπηρετητή καθορίζεται από το MIME Content Header της απάντησης (Πίνακας 2.4, γραμμή 9). Έτσι είναι δυνατόν πόροι ανεξαρτήτως κατάληξης να περιέχουν σελίδες που προβάλλονται στον επισκέπτη. Σε σύγχρονους ιστοχώρους, όπου το περιεχόμενο παράγεται δυναμικά τη στιγμή που γίνεται η αίτηση, είναι μάλιστα ο κανόνας οι πόροι να μην έχουν κατάληξη .html. Τέτοια παραδείγματα είναι ιστοχώροι που χρησιμοποιούν την τεχνολογία Active Server Pages (ASP) της Microsoft [44], Java Server Pages της Sun [45] ή τη γλώσσα PHP [46]. Για τέτοιους ιστοχώρους η απλή λύση είναι να προστεθούν και οι καταλήξεις αυτών των πόρων στις αιτήσεις σελίδων (asp, jsp, php αντίστοιχα).

Υπάρχουν όμως πόροι που δεν έχουν καμία κατάληξη, όπου εκεί πλέον είναι αδύνατον να υπάρξει ένας ενιαίος κανόνας για την αναγνώριση αιτήσεων για σελίδες. Για παράδειγμα στο [43] το URI

<http://www.crete.tournet.gr/servlet/msc.services.ct.maps.AreaOnMap>

παράγει HTML σελίδες ενώ ο πόρος αναφέρεται σε όνομα Java κλάσης. Το πρόβλημα περιπλέκεται ακόμα περισσότερο όταν πόροι δυναμικής παραγωγής περιεχομένου, όπως τα asp, jsp και php που αναφέρθηκαν, χρησιμοποιούνται για την παραγωγή μέσων που δεν είναι σελίδες. Για παράδειγμα τέτοιοι πόροι μπορεί να παράγουν κλασικές εικόνες. Οπότε η μέθοδος αναγνώρισης σελίδων που θα χρησιμοποιηθεί πρέπει να λαμβάνει υπ' όψιν τουλάχιστον τα στοιχεία που αναφέρονται στον πίνακα 2.8.

<i>Αναγνώριση πόρων με περιεχόμενο σελίδες</i>
Γνωστές καταλήξεις πόρων
Συγκεκριμένοι πόροι χωρίς κατάληξη
Εξαιρέσεις στους προηγούμενους κανόνες

*Πίνακας 2.8 Κανόνες για αναγνώριση αιτήσεων ως σελιδοπροβολές*

Ένα εντελώς διαφορετικό πρόβλημα συναντάται σε ιστοχώρους που χρησιμοποιούν HTML Frames ή IFrames. Η φιλοσοφία των HTML frames βασίζεται στην ιδέα ότι μία οθόνη στην εφαρμογή πλοήγησης μπορεί να αποτελείται από περισσότερες της μίας σελίδας HTML. Αν και η χρήση Frames κατέχει την πρώτη θέση του “Top 10 Mistakes in Web Design” του Jakob Nielsen ([47]), συναντώνται ακόμα συχνά σε ιστοχώρους. Το πρόβλημα έγκειται στο γεγονός ότι η εμφάνιση μιας σελίδας που αποτελείται από πολλά frames θα καταγραφεί ως πολλαπλές αιτήσεις για σελιδοπροβολές στον εξυπηρετητή, επειδή για κάθε σελίδα θα γίνει ξεχωριστή αίτηση.

Γίνεται κατανοητό ότι η χρήση μιας τέτοιας τεχνικής για την αναγνώριση σελίδων απαιτεί αναγκαστικά τη συμβολή των ανθρώπων που ανέπτυξαν τον ιστοχώρο, διότι μόνο αυτοί είναι σε θέση να γνωρίζουν τα ακριβή αποτελέσματα της κάθε αίτησης.

### **2.3.5.2 Χαρακτηρισμός Σελιδοπροβολών**

Απαλείφοντας λοιπόν όλες τις αιτήσεις που δεν αναφέρονται σε σελιδοπροβολές και παραμένοντας στον ιστοχώρο [43], μια πιθανή ακολουθία σελιδοπροβολών ενός επισκέπτη μπορεί να είναι αυτή του πίνακα 2.9A (φαίνονται μόνο το απόλυτο μονοπάτι των URI). Η περιγραφή αυτής της ακολουθίας μπορεί να αποτελείται από τη σειρά των απόλυτων μονοπατιών των σελίδων για κάθε αίτηση (Home, Contact, Harbours, Home, Intro, Home, About). Δηλαδή μπορεί να χρησιμοποιηθεί το όνομα της κάθε σελίδας για να περιγραφούν οι σελιδοπροβολές για το συγκεκριμένο επισκέπτη.

Στον πίνακα 2.9B έχουμε μία άλλη πιθανή ακολουθία σελιδοπροβολών. Ενώ στις 3 πρώτες αιτήσεις συμπίπτει με την περίπτωση A, από την τέταρτη αίτηση και μετά το απόλυτο μονοπάτι του URI παραμένει “Harbour.jsp” και διαφοροποιείται μόνο στο τμήμα του query του URI. Ανάλογα με το τμήμα του query προβάλλεται στον επισκέπτη διαφορετικό περιεχόμενο, οπότε δεν επαρκεί πλέον η



αναφορά του μονοπατιού για τον προσδιορισμό της σελιδοπροβολής. Μία λύση θα ήταν να αναφέρεται κάθε σελιδοπροβολή με το συνδυασμό του μονοπατιού και του τμήματος του URI. Σε περιπτώσεις όμως που ανεξάρτητα από το query το περιεχόμενο ανήκει σε μία συγκεκριμένη κατηγορία πληροφοριών που παρουσιάζεται (στη συγκεκριμένη περίπτωση πληροφορία για λιμάνι) θα ήταν πολύ χρήσιμο να διατηρηθεί αυτή η πληροφορία.

<i>Πιθανές Ακολουθίες Σελιδοπροβολών ενός Επισκέπτη</i>		
<b>A</b>	<b>B</b>	<b>Γ</b>
/Home.jsp	/Home.jsp	/Home.jsp?lang=en
/ContactUs.jsp	/ContactUs.jsp	/ContactUs.jsp?lang=en
/MainHarbour.jsp	/MainHarbour.jsp	/MainHarbour.jsp?lang=en
/Home.jsp	/Habour.jsp?ID=2	/Home.jsp?lang=en
/IntroTOC.jsp	/Habour.jsp?ID=25	/IntroTOC.jsp?lang=en
/Home.jsp	/Habour.jsp?ID=8	/Home.jsp?lang=en
/About.jsp	/Habour.jsp?ID=11	/About.jsp?lang=en

*Πίνακας 2.9 Τρία διαφορετικά σενάρια αιτήσεων ενός επισκέπτη για σελίδες στο [43]. Στην περίπτωση A τα ονόματα των αιτήσεων επαρκούν για τον προσδιορισμό του περιεχομένου της κάθε σελιδοπροβολής, ενώ στην περίπτωση B και Γ πρέπει να συμπεριληφθούν και οι παράμετροι του URI*

Μια τρίτη περίπτωση είναι αυτή που φαίνεται στην περίπτωση του πίνακα 2.9Γ. Σε αυτή την περίπτωση τα μονοπάτια είναι ίδια με την περίπτωση A αλλά διαφέρουν πάλι ως προς το τμήμα του query το οποίο προσδιορίζει τη γλώσσα εμφάνισης του περιεχομένου. Σε αντίθεση με την περίπτωση B δεν έχουμε παρουσίαση διαφορετικής πληροφορίας, αλλά έχουμε την ίδια πληροφορία να παρουσιάζεται σε διαφορετική γλώσσα. Παρόμοιο πρόβλημα υπάρχει όταν υπάρχουν στατικές σελίδες που παρουσιάζονται σε διαφορετικές γλώσσες, οπότε η διαφοροποίηση γίνεται είτε στο όνομα του πόρου είτε στο μονοπάτι που οδηγεί σε αυτό (για παράδειγμα “/en/index.html” και “/el/index.html”) οπότε και πάλι πρέπει κατά την ανάλυση να ληφθεί υπ' όψιν ότι παρόλο που πρόκειται για διαφορετικά μονοπάτια ουσιαστικά παρουσιάζεται το ίδιο περιεχόμενο. Τέτοιου είδους προβλήματα καλύπτονται με ιεραρχικές έννοιες που θα περιγραφεί σε επόμενη ενότητα.

### **2.3.5.3 Διάκριση Επισκέψεων**

Οι εξυπηρετητές παγκόσμιου ιστού διαχειρίζονται πολλές αιτήσεις ταυτόχρονα. Οι αιτήσεις μπορεί να προέρχονται είτε από έναν χρήστη είτε από περισσότερους. Όποια μέθοδος συλλογής δεδομένων και να χρησιμοποιηθεί, το αποτέλεσμα της πολλαπλής ταυτόχρονης διαχείρισης αιτήσεων είναι να υπάρχει ένα χρονικά διατεταγμένο σύνολο αιτήσεων όπου δεν είναι εμφανές η αντιστοίχιση των αιτήσεων σε μεμονωμένους επισκέπτες του ιστοχώρου. Δηλαδή δεν υπάρχει άμεση δυνατότητα να

ομαδοποιηθούν οι αιτήσεις ανά επίσκεψη στον ιστοχώρου.

Στον πίνακα 2.10 φαίνεται ένα πιθανό απόσπασμα ορισμένων πεδίων ενός αρχείου καταγραφής σε μορφή CLF. Διακρίνεται η ip διεύθυνση της εφαρμογής πελάτη, το μονοπάτι της αίτησης (όλες οι αιτήσεις θεωρούνται σελιδοπροβολές) και το πεδίο User-Agent που είναι αναγνωριστικό της εφαρμογής του τελικού χρήστη.

<i>A/A</i>	<i>IP</i>	<i>Request</i>	<i>User-Agent</i>
1	ip1	/page1	ua1
2	ip2	/page1	ua3
3	ip1	/page2	ua1
4	ip1	/page1	ua2
5	ip1	/page2	ua2
6	ip1	/page3	ua1
7	ip2	/page4	ua3
8	ip1	/page3	ua2
9	ip1	/page4	ua2
10	ip1	/page5	ua2
11	ip1	/page6	ua2
12	ip1	/page4	ua1
13	ip1	/page5	ua1
14	ip2	/page5	ua3
15	ip2	/page6	ua3
16	ip2	/page7	ua3

*Πίνακας 2.10 Εικονικό απόσπασμα των πεδίων IP, Request και User-Agent ενός CLF αρχείου καταγραφής. Διακρίνονται 2 διαφορετικές διευθύνσεις IP και 3 διαφορετικά User-Agent. Η ομαδοποίηση των αιτήσεων ανά επίσκεψη δεν είναι προφανής.*

Στον πίνακα υπάρχουν 2 διαφορετικές διευθύνσεις IP και 3 διαφορετικές τιμές για το User-Agent. Για την ομαδοποίηση των αιτήσεων ανά επίσκεψη, η πρώτη σκέψη είναι να ομαδοποιηθούν οι αιτήσεις με βάση τη διεύθυνση IP της κάθε αίτησης. Είναι μία πολύ συνηθισμένη τακτική που χρησιμοποιείται από απλές εφαρμογές απεικόνισης στατιστικών κίνησης όπως το [48], και βασίζεται στην ιδέα ότι οι διευθύνσεις IP που έχει κάθε μηχάνημα είναι μοναδικές. Για τον παραπάνω πίνακα θα υπήρχαν έτσι 2 διαφορετικές επισκέψεις, όπου η πρώτη θα ήταν αυτή με την IP διεύθυνση ip1 και η δεύτερη με ip2.

Η τελευταία παραδοχή όμως ακυρώνεται από τους μηχανισμούς παρεμβολής ενδιάμεσων σταθμών

στην αλυσίδα HTTP, όπου η τελική αίτηση στον εξυπηρετητή γίνεται όχι από το αρχικό μηχάνημα (δηλαδή τον τελικό χρήστη) αλλά από κάποιον ενδιάμεσο σταθμό (βλέπε Σχήμα 2.11 περίπτωση 3). Συνηθισμένες περιπτώσεις όπου χρησιμοποιούνται τέτοιοι μηχανισμοί είναι ενδιάμεσοι σταθμοί που λειτουργούν σαν proxy εξυπηρετητές για λόγους ταχύτητας ή ασφάλειας, ή gateways όταν δεν είναι διαθέσιμες αρκετές IP διευθύνσεις.

Όποιος και να είναι ο λόγος, το αποτέλεσμα είναι να εμφανίζονται στα πρωτογενή δεδομένα αιτήσεις με την ίδια διεύθυνση που ανήκουν σε διαφορετικές επισκέψεις. Μία ευρέως διαδεδομένη μέθοδος είναι να μη χρησιμοποιείται μόνο η IP διεύθυνση για το διαχωρισμό των επισκέψεων, αλλά ταυτόχρονα και στοιχεία από το πεδίο του User-Agent. Η μέθοδος βασίζεται στο γεγονός ότι σύμφωνα με το HTTP πρωτόκολλο ([29]) το πεδίο User-Agent πρέπει να είναι πάντα αυτό που δίνεται από την εφαρμογή που ξεκινάει την αίτηση (“user agent originating the request”). Η μορφή του πεδίου δεν είναι σαφώς καθορισμένη αλλά συνήθως περιέχει πληροφορίες για τον κατασκευαστή της εφαρμογής, την έκδοση της εφαρμογής και το λειτουργικό σύστημα. Παραδείγματα τιμών αυτού του πεδίου δίνονται στον πίνακα 2.11. Οπότε χρησιμοποιείται σαν αναγνωριστικό αιτήσεων που ανήκουν στην ίδια επίσκεψη ο συνδυασμός της διεύθυνσης IP, της έκδοσης του προγράμματος πλοήγησης και του λειτουργικού συστήματος. Οπότε στο παράδειγμα του πίνακα 2.10 προκύπτουν 3 διαφορετικές επισκέψεις (ip1-ua1, ip1-ua2, ip2-ua3).

<i>User-Agent</i>	<i>Σχόλια</i>
Mozilla/4.0 (compatible; MSIE 5.5; Windows 98)	Internet Explorer 5.5 Windows 98
Mozilla/4.75 [de] (Win98; U)	Netscape 4.75 Windows 98 German Edition
Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0)	Microsoft Internet Explorer 6.0 Windows 2000
Mozilla/4.0 (compatible; MSIE 5.0; Mac_PowerPC)	Microsoft Internet Explorer 5.0 Macintosh
Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.1b) Gecko/20020730	Mozilla 1.1beta Windows 2000
Openfind data gatherer, Openbot/3.0+(robot-response@openfind.com.tw;+http://www.openfind.com.tw/robot.html)	Crawler

Πίνακας 2.11 Παραδείγματα τιμών του πεδίου User-Agent

Η τελευταία μέθοδος είναι η πλέον διαδεδομένη όταν τα πρωτογενή δεδομένα συλλέγονται από τα αρχεία καταγραφής των εξυπηρετητών σε μορφή CLF. Δεν είναι όμως απαλλαγμένη από πιθανές περιπτώσεις λαθών. Τέτοιες περιπτώσεις ανακύπτουν από διάφορους παράγοντες που άλλες φορές οφείλονται σε συνθήκες της αγοράς του Διαδικτύου, ενώ άλλες από ιδιαιτερότητες του υπό ανάλυση ιστοχώρου. Για παράδειγμα είναι ευρέως διαδεδομένη η περίπτωση να χρησιμοποιείται άλλο User-Agent από το πραγματικό, επειδή πολλοί ιστοχώροι αρνούνται την εμφάνιση σελίδων όταν η πρόσβαση δε γίνεται από συγκεκριμένους τύπους εφαρμογών πλοήγησης. Έτσι εναλλακτικές εφαρμογές χρησιμοποιούν ως τιμή για το User-Agent κάποια που να είναι συμβατή με τους περισσότερους ιστοχώρους. Τέτοια περίπτωση είναι για παράδειγμα η εφαρμογή πλοήγησης Opera [49]. Άλλου είδους περίπτωση έχουμε όταν η πρόσβαση στον ιστοχώρο γίνεται από υπολογιστές και χρήστες που ακολουθούν κάποια συγκεκριμένη πολιτική του οργανισμού στον οποίο υπάγονται. Για παράδειγμα μπορεί να υπάρχει ένας ιστοχώρος στον οποίο έχουν πρόσβαση μόνο οι υπάλληλοι μιας εταιρείας, στην οποία έχει επιβληθεί η χρήση συγκεκριμένης εφαρμογής πλοήγησης. Εάν μάλιστα η πρόσβαση γίνεται δια μέσω ενιαίου proxy δημιουργείται μεγάλη σύγχυση στο διαχωρισμό των επισκέψεων.

Η πλέον ασφαλής μέθοδος για το διαχωρισμό των επισκέψεων βασίζεται στη χρήση session-cookies. Όπως αναφέρθηκε σε προηγούμενη ενότητα, πρόκειται για ένα μοναδικό χαρακτηριστικό το οποίο στέλνεται στην εφαρμογή πλοήγησης του χρήστη με την πρώτη αίτηση στον εξυπηρετητή και η εφαρμογή πλοήγησης την στέλνει πίσω στον εξυπηρετητή με κάθε νέα αίτηση που γίνεται εντός των ορίων μιας επίσκεψης. Τα session-cookies δεν αποθηκεύονται στα CLF αρχεία των εξυπηρετητών, αν και υπάρχει συνήθως η δυνατότητα να προστεθούν. Είναι σίγουρα διαθέσιμα στις υπόλοιπες μεθόδους συλλογής των πρωτογενών δεδομένων. Στην περίπτωση που χρησιμοποιούνται session-cookies για τη διάκριση των επισκέψεων, αναφερόμαστε στην περίπτωση των session-visitors της ενότητας 2.3.4.1 για του επισκέπτες.

Βέβαια, όπως αναφέρθηκε στην ενότητα 2.3.4.1 στην περιγραφή των cookies, είναι στην ευχέρεια του χρήστη να δεχτεί ή όχι τα cookies, οπότε και πάλι μπορεί να δημιουργηθούν στρεβλώσεις στα αποτελέσματα.

Επειδή ακριβώς δεν υπάρχει κάποιος αξιόπιστος τρόπος να ομαδοποιηθούν οι αιτήσεις σε επισκέψεις σε ανοιχτά προσβάσιμους ιστοχώρους, κρίνεται ως πλέον ενδεδειγμένος τρόπος ο συνδυασμός των παραπάνω μεθόδων. Δηλαδή αρχικά γίνεται προσπάθεια να ομαδοποιηθούν οι αιτήσεις με βάση την ύπαρξη session-cookie. Εάν διαπιστωθεί ότι δεν έγινε αποδοχή του cookie,

χρησιμοποιείται η μέθοδος συνδυασμού των IP και User-Agent πεδίων. Εάν και σε αυτή την περίπτωση παρατηρηθεί ότι δεν είναι αξιόπιστα τα δεδομένα (για παράδειγμα επειδή δε μεταφέρεται το User-Agent) τότε χρησιμοποιείται μόνο η διεύθυνση IP, ενώ μπορεί να ελεγχθεί και η περίπτωση να μην συμπεριληφθούν οι συγκεκριμένες αιτήσεις στην ανάλυση, αν και εφόσον αντιπροσωπεύουν μόνο μικρό μέρος των προσβάσεων στον ιστοχώρο.

Τέλος, πρέπει να επαναληφθεί από τον ορισμό της επίσκεψης στην ενότητα 2.3.4.1 ότι συνεχόμενες αιτήσεις προς τον ιστοχώρο από τον ίδιο χρήστη θεωρούνται ότι ανήκουν στην ίδια επίσκεψη εάν και μόνο εάν ο χρόνος που μεσολαβεί ανάμεσα σε δυο διαδοχικές αιτήσεις δεν υπερβαίνει ένα ανώτατο χρονικό όριο που έχει τεθεί. Με βάση την εργασία του [37] ο χρόνος που γενικά χρησιμοποιείται για αυτό το σκοπό είναι 30 λεπτά. Δεν είναι όμως λίγες οι περιπτώσεις όπου αυτός ο χρόνος μπορεί να προκαλέσει σημαντικότερες διαφοροποιήσεις στον αριθμό των επισκέψεων. Για παράδειγμα, όπως αναφέρεται στο [50], άλλαξε από την 01/01/2002 ο τρόπος με τον οποίο το [34] υπολογίζει τα στατιστικά για την κίνηση των ιστοχώρων που εξυπηρετεί. Μία από τις αλλαγές ήταν η αλλαγή του χρόνου νέας επίσκεψη από 4 λεπτά στο διεθνώς αναγνωρισμένο 30 λεπτά. Το αποτέλεσμα ήταν να υπολογίζονται έως και 90% λιγότερες επισκέψεις σε συγκεκριμένους ιστοχώρους. Οι ιστοχώροι αυτοί έχουν κάποια κοινά χαρακτηριστικά, όπως για παράδειγμα τη μεγάλη διάρκεια ανά σελιδοπροβολή (οι επισκέπτες δηλαδή έβλεπαν για πολύ ώρα μία σελίδα). Ορισμένοι ιστοχώροι για παράδειγμα πρόσφεραν στις σελίδες τους μέσα μεγάλου όγκου πληροφορίας (π.χ. Video) οπότε έως ότου μεταφερθεί η πληροφορία και ξαναγίνει νέα αίτηση περνούσαν πάνω από 4 λεπτά και θεωρούνταν ως νέα επίσκεψη.

#### **2.3.5.4 Διάκριση Επισκεπτών**

Όπως φάνηκε από την προηγούμενη παράγραφο, η διάκριση επισκέψεων από τα πρωτογενή δεδομένα έχει πολλά σημεία που μπορεί να οδηγήσουν σε λανθασμένα συμπεράσματα. Από αυτή την άποψη η διάκριση και αναγνώριση μεμονωμένων επισκεπτών είναι ένα πρόβλημα αρκετά πιο δύσκολο να επιλυθεί. Να σημειωθεί ότι αυτή η διάκριση μπορεί να οδηγήσει είτε σε tracked-visitor είτε σε identified-visitor της ενότητας 2.3.4.1. Επίσης πρέπει να τονιστεί ότι στην περίπτωση που υπάρχει μία και μοναδική επίσκεψη από έναν χρήστη στον ιστοχώρο, δεν έχει νόημα η έννοια του επισκέπτη επειδή συμπίπτει με αυτή τη μοναδική επίσκεψη. Η διάκριση επισκεπτών αποκτά ενδιαφέρον για επανεπισκέψεις του ίδιου χρήστη στον ίδιο ιστοχώρο.

Από την περιγραφή της διάκρισης επισκέψεων πρέπει να έχει γίνει κατανοητό ότι με μόνα δεδομένα τα στοιχεία από τα αρχεία καταγραφής CLF, στη γενική περίπτωση δεν είναι δυνατόν να

γίνει καμία είδους αναγνώριση των επισκεπτών. Ο τρόπος που μπορεί να χρησιμοποιηθεί είναι πάλι η χρήση cookies, με τη διαφορά ότι η λήξη τους δεν γίνεται αυτόματα με το τέλος μίας επίσκεψης αλλά μετά από πολύ μεγάλο χρονικό διάστημα. Είναι συνηθισμένο να χρησιμοποιούνται ημερομηνίες λήξης που βρίσκονται πολλά χρόνια στο μέλλον. Έτσι, ακόμα και αν ο χρήστης δεν επαναλάβει μία επίσκεψη για μεγάλο χρονικό διάστημα, τη στιγμή που θα γίνει η αίτηση στο ιστοχώρο θα μεταφερθεί το cookie που είναι ακόμα ενεργό και θα αναγνωριστεί.

Μεγαλύτερη ευκολία για την αναγνώριση των επισκεπτών έχουν ιστοχώροι που προσφέρουν τις υπηρεσίες τους μόνο μετά από ασφαλή αναγνώριση. Τέτοιοι ιστοχώροι μπορεί να είναι για παράδειγμα ιστοχώροι online τραπεζικής. Σε αυτές τις περιπτώσεις μάλιστα μπορεί να διακριθεί ο επισκέπτης ακόμα και αν για κάποιο λόγο δεν είναι διαθέσιμο το cookie (λόγω αλλαγής υπολογιστή για παράδειγμα), επειδή ο χρήστης θα αναγκαστεί να πιστοποιήσει την ταυτότητά του.

Πρέπει να τονιστεί ότι η ύπαρξη ενός visitor-cookie δεν σημαίνει αναγκαστικά ότι το ίδιο άτομο που επισκέφτηκε τον ιστοχώρο όταν πρωτοστάλθηκε το cookie είναι το ίδιο που θα το επισκέπτεται όταν διαπιστωθεί η ύπαρξη του cookie. Τα cookies συσχετίζονται με την εφαρμογή πλοήγησης ανά λογαριασμό στους υπολογιστές όπου είναι εγκατεστημένοι. Εάν είτε δεν υπάρχουν λογαριασμοί (οπότε όλοι οι χρήστες χρησιμοποιούν τον ίδιο λογαριασμό) είτε υπάρχουν αλλά δε χρησιμοποιούνται, είναι πολύ πιθανόν να “κρύβεται” στην πραγματικότητα άλλος επισκέπτης πίσω από αυτό το visitor-cookie.

### **2.3.5.5 Ανακατασκευή Μονοπατιών Επίσκεψης**

Στην ενότητα για τη διάκριση επισκέψεων περιγράφηκαν διάφορες μέθοδοι με τις οποίες ομαδοποιούνται οι αιτήσεις που καταγράφονται ανά μεμονωμένη επίσκεψη. Το αποτέλεσμα αυτής της εργασίας είναι ένα σύνολο, με χρονική σειρά, διατεταγμένων σελιδοπροβολών ανά επίσκεψη, το οποίο αναφέρεται ως μονοπάτι επίσκεψης. Σε συνδυασμό με την τοπολογία του ιστοχώρου, η ανάλυση των μονοπατιών που ακολουθούν οι χρήστες μπορεί να δώσει πολύτιμες πληροφορίες για τα ενδιαφέροντα των επισκεπτών, τον τρόπο αναζήτησης πληροφοριών στον ιστοχώρο και δυνατότητες βελτίωσης της τοπολογίας για καλύτερη ανταπόκριση στις επιθυμίες των επισκεπτών. Βασική προϋπόθεση όμως είναι τα δεδομένα για τα μονοπάτια που συλλέχθηκαν να είναι όσο το δυνατόν πιο ακριβή.

Δυστυχώς όμως υπάρχουν πολύ συνηθισμένες περιπτώσεις όπου λαμβάνει χώρα μία σελιδοπροβολή στην οθόνη του επισκέπτη χωρίς αυτή να μπορεί να καταγραφεί στην πλευρά του εξυπηρετητή. Τέτοιες περιπτώσεις περιγράφηκαν στην ενότητα “Αξιοπιστία Δεδομένων” και αναφέρονται γενικά στην ύπαρξη μηχανισμών cache των δεδομένων. Στην ίδια ενότητα

περιγράφηκαν κάποιες λύσεις για την αποφυγή αυτών των μηχανισμών, αλλά πέραν της μεθόδου συλλογής δεδομένων από ειδικά ενσωματωμένη εικόνα, δεν υπάρχει αποδεκτός τρόπος για την καταγραφή όλων των αιτήσεων προς τον εξυπηρετητή.

Αντί αυτού προκρίνεται η λύση της ανακατασκευής των τμημάτων του μονοπατιού που πιθανώς δεν καταγράφηκαν. Σημειώνεται ότι σε αυτή την περίπτωση προστίθενται δεδομένα αντί να αφαιρούνται όπως έγινε σε προηγούμενες φάσεις της συλλογής και προεπεξεργασίας δεδομένων. Ο όρος που χρησιμοποιείται για αυτή την εργασία είναι *path-completion*.

Ένας γενικός κανόνας που οδηγεί στην προσθήκη σελιδοπροβολών περιγράφεται στο [8] και κυρίως στο [14]. Βασίζεται στην ύπαρξη του πεδίου *referrer*, το οποίο δίνει το URI από το οποίο προήλθε μία σελιδοπροβολή, και την τοπολογία του ιστοχώρου που προσδιορίζει τις πιθανές μεταβάσεις από μία σελίδα σε άλλη. Έστω ότι κατά την ομαδοποίηση των σελιδοπροβολών ανά επίσκεψη συναντάται μία σελιδοπροβολή η οποία δεν είναι προσβάσιμη από την προηγούμενη σελιδοπροβολή που αναγνωρίστηκε. Εξετάζεται το πεδίο *referrer* για να εξακριβωθεί εάν πρόκειται για σελίδα που υπάρχει στο πρόσφατο ιστορικό της επίσκεψης (επισκέφτηκε την *referring* σελίδα ήδη). Σε αυτή την περίπτωση γίνεται η υπόθεση ότι ο χρήστης χρησιμοποίησε τη λειτουργία “back” του προγράμματος πλοήγησης μέχρι να φτάσει στη *referring* σελίδα, και από εκεί επέλεξε την εξεταζόμενη σελιδοπροβολή.

Εάν από το πεδίο *referrer* δεν είναι ξεκάθαρο τι έχει συμβεί, εξετάζεται η τοπολογία του ιστοχώρου και γίνεται παρόμοια υπόθεση. Εάν από την τοπολογία του ιστοχώρου βρεθούν περισσότερες της μίας σελίδας που συνδέουν στην εξεταζόμενη σελιδοπροβολή, γίνεται η υπόθεση ότι προήλθε από τη σελίδα που βρίσκεται πιο κοντά στην εξεταζόμενη σελιδοπροβολή.

### **2.3.5.6 Ιεραρχικές Έννοιες**

Στην περιγραφή των σελιδοπροβολών αναφέρθηκε ο προβληματισμός που προκύπτει όταν ουσιαστικά διαφορετικές σελιδοπροβολές μπορεί να ομαδοποιηθούν με βάση κοινά χαρακτηριστικά τους. Δόθηκε συγκεκριμένα το παράδειγμα όπου σελίδες με διαφορετικό URI παρουσιάζουν το ίδιο περιεχόμενο αλλά σε διαφορετική γλώσσα, ή παρουσιάζουν περιεχόμενο της ίδιας κατηγορίας. Στην πρώτη περίπτωση το στοιχείο που αλλάζει μία συγκεκριμένη σελίδα είναι η γλώσσα παρουσίασης ενώ στη δεύτερη καθ' αυτό το περιεχόμενο (για παράδειγμα σελίδες που όλες παρουσιάζουν περιοχές ενδιαφέροντος μιας περιφέρειας). Σε μία ψηφιακή βιβλιοθήκη θα μπορούσαν να ομαδοποιηθούν οι παρουσιάσεις βιβλίων ανάλογα με το αντικείμενό τους (επιστημονικό πεδίο, λογοτεχνική μορφή), τη γλώσσα συγγραφής ή είδος έκδοσης (βιβλίο, περιοδικό κλπ.).

Στην πράξη συναντώνται τέτοιες περιπτώσεις πολύ συχνά ειδικά όταν οι σελίδες παράγονται με δυναμικό τρόπο από βάσεις δεδομένων. Για την περιγραφή τέτοιων συσχετισμών χρησιμοποιούνται ιεραρχικές έννοιες (concept hierarchies) ή ταξινομίες (taxonomies). Η γενική ιδέα βασίζεται στην εισαγωγή ενός ή περισσότερων πιο αφηρημένων επιπέδων στα οποία γενικεύονται τα αντικείμενα που παρουσιάζονται στους ιστοχώρους. Μπορεί μάλιστα το ίδιο αντικείμενο (για παράδειγμα η παρουσίαση ενός βιβλίου) να αντιστοιχιστεί σε περισσότερες της μίας ταξινομίας.

Η δημιουργία τέτοιων κατηγοριοποιήσεων μπορεί να γίνει είτε αυτόματα είτε χειροκίνητα από τα άτομα που γνωρίζουν την περιοχή που καλύπτει ο ιστοχώρος. Στις αυτόματες κατηγοριοποιήσεις χρησιμοποιούνται ομοιότητες (similarity) στο κείμενο του περιεχομένου των σελίδων για την ομαδοποίησή τους, ή τις συνδέσεις (links) που υπάρχουν ανάμεσα στις σελίδες. Μία άλλη προσέγγιση εξετάζει τα μονοπάτια των επισκέψεων και ομαδοποιεί τις σελίδες που ανήκουν στα ίδια μονοπάτια. Γενικά αυτές οι ομαδοποιήσεις βασίζονται στο περιεχόμενο των σελίδων (content based, semantic taxonomies).

Στο [51], το οποίο εξετάζει τη συμπεριφορά στην πλοήγηση σε ιστοχώρους, χρησιμοποιείται η ομαδοποίηση με βάση τις υπηρεσίες του ιστοχώρου (service based conceptual hierarchy). Αναλύει ιστοχώρους που προσφέρουν υπηρεσίες αναζήτησης πληροφορίας, η οποία προσφέρεται στο χρήστη από δυναμικά παραγόμενες σελίδες. Τα αντικείμενα παρουσιάζονται με διαφορετικό βαθμό ανάλυσης (λίστα αποτελεσμάτων, αναλυτική παρουσίαση κλπ.). Έτσι εισάγονται οι έννοιες του Entity class που είναι αντικείμενα για τα οποία ο ιστοχώρος προσφέρει πληροφορία, Level of detail για το βαθμό ανάλυσης στην παρουσίαση των αντικειμένων και Specifiable properties για τα χαρακτηριστικά στα οποία μπορεί να βασιστεί μία ερώτηση.

Η έννοια των concept hierarchies χρησιμοποιείται και για την κατηγοριοποίηση άλλου είδους δεδομένων όπως δημογραφικές πληροφορίες (π.χ. χώρα προέλευσης επισκεπτών). Επίσης μπορεί να χρησιμοποιηθούν για ήδη υφιστάμενες συσχετίσεις όπως ταξινομίες προϊόντων εταιρειών [12].

Η κατασκευή των concept hierarchies είναι μία εργασία που ήδη αγγίζει το επόμενο βήμα της διαδικασίας ανάλυσης επισκεψιμότητας που είναι η ανακάλυψη γνώσης, επειδή ειδικά στις αυτόματες διαδικασίες ομαδοποίησης μπορεί να αποκαλύψει συνήθειες των επισκεπτών που δεν είναι εμφανής εξ' αρχής. Παρόλα αυτά εντάσσεται στο στάδιο της προεπεξεργασίας των δεδομένων επειδή δημιουργεί δομές που χρησιμοποιούνται περαιτέρω στα επόμενα στάδια της ανάλυσης.

### **2.3.5.7 Επεισόδια - Συναλλαγές**

Οι επισκέψεις, όπως περιγράφηκαν ως τώρα, θεωρούνται αυτόνομες έννοιες. Χαρακτηρίζονται από



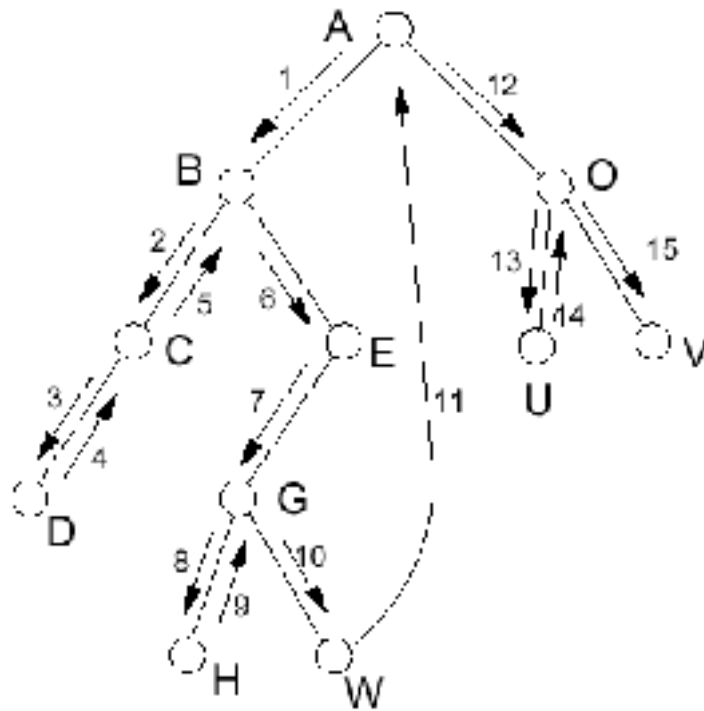
ένα σύνολο σελίδων πάνω σε ένα μονοπάτι που ακολούθησε ο επισκέπτης στη διάρκεια της επίσκεψης. Αυτή η ακολουθία σελιδοπροβολών μπορεί να σπάσει όμως σε μικρότερα τμήματα, είτε με βάση τη σειρά επίσκεψης των σελίδων είτε με βάση τη σημασιολογική συσχέτιση των σελίδων. Πολλές φορές οι νέες δομές που προκύπτουν δεν περιορίζονται μόνο σε σελιδοπροβολές αλλά προστίθενται γενικότερα οι επιλογές συνδέσμων του χρήστη (clicks – click stream).

Αυτό το υποσύνολο της ακολουθίας σελιδοπροβολών αναφέρεται με τον όρο επεισόδιο (episode) ή συναλλαγή (transaction) μιας επίσκεψης. Στο παράδειγμα του [43] θα περίμενε κανείς να ανακαλυφθούν επεισόδια που περιγράφονται με τους τίτλους “Αναζήτηση περιοχών επίσκεψης”, “Αναζήτηση μονάδων διαμονής”, “Αναζήτηση μεταφοράς” κλπ. τα οποία είναι αυτόνομες ενέργειες που γίνονται σε μία επίσκεψη στον ιστοχώρο. Παρόμοια επεισόδια μπορεί να ανακαλυφθούν σε ιστοχώρους ηλεκτρονικών καταστημάτων (αναζήτηση προϊόντων, επιλογή προϊόντων, διενέργεια παραγγελίας κλπ.)

Για τον εντοπισμό επεισοδίων στο [8] και [42] οι σελίδες του ιστοχώρου χωρίζονται πρώτα σε κατηγορίες ανάλογα με το περιεχόμενό τους (content page, navigation page etc.). Εισάγεται η έννοια του *Reference Length Module* όπου για κάθε είδους σελίδας επιτρέπεται ένας μέγιστος χρόνος προβολής και ξεκινάει μία νέα συναλλαγή όταν υπερβαίνεται αυτός ο χρόνος, και *Time Window Module* όπου υπάρχει ένα άνω χρονικό όριο για κάθε συναλλαγή.

Στο [41] θεωρείται σαν επεισόδιο η ακολουθία σελιδοπροβολών μέχρι την αναφορά σε προηγούμενη σελίδα (backward reference, όταν για παράδειγμα επιλέγεται η λειτουργία back στην εφαρμογή πλοήγησης). Ο αλγόριθμος που χρησιμοποιείται επικεντρώνεται έτσι στην εύρεση των μονοπατιών που ανήκουν στην επίσκεψη όπου δεν υπάρχει αναφορά προς τα πίσω. Αυτά τα μονοπάτια ονομάζονται *Maximal Forward References*. Έστω ότι το μονοπάτι μιας επίσκεψης αποτελείται από την ακολουθία {A,B,C,D,C,B,E,G,H,G,W,A,O,U,O,V}. Στο Σχήμα 2.13 φαίνεται το ίδιο μονοπάτι σε γράφο, όπου φαίνονται και οι επιστροφές (4,5,9,11,14). Από την ανάλυση της επίσκεψης προκύπτουν ως *Maximal Forward References* τα {ABCD, ABEGH, ABEGW, AOU, AOV}.

Όπως όμως επισημαίνεται στο [51] έχει βρεθεί ότι πάνω από το 50% των κινήσεων των χρηστών σε ιστοχώρους είναι κινήσεις προς τα πίσω, οπότε υπάρχει αμφιβολία για την αξία μιας τέτοιας ανάλυσης.



Σχήμα 2.13 Υπολογισμός Maximal Forward References από το [41].

### 2.3.5.8 Άλλα Στοιχεία

Ως τώρα παρουσιάστηκαν αρκετές εργασίες που γίνονται στο στάδιο της προεπεξεργασίας των πρωτογενών δεδομένων. Όλες κινούνται γύρω από τις σχέσεις σελιδοπροβολή/ επίσκεψη/ επισκέπτης και ασχολούνται με τον όσο το δυνατό πιο ακριβή διαχωρισμό των δεδομένων που είναι διαθέσιμα. Υπάρχουν όμως πολλά επιπλέον δεδομένα που αποκτούν αξία μετά τις παραπάνω εργασίες επειδή είναι άμεσα συσχετισμένες με τα δεδομένα που οργανώνονται στα προηγούμενα βήματα. Θα περιγραφούν κάποια από αυτά στη συνέχεια.

### IP Διεύθυνση

Η διεύθυνση IP του επισκέπτη, όταν είναι διαθέσιμη, μπορεί να δώσει πληροφορίες για την χώρα ή περιοχή απ' όπου προέρχεται η επίσκεψη. Η πληροφορία αυτή είναι διαθέσιμη (σε διαφορετικό βαθμό) από 2 μηχανισμούς: Reverse DNS Lookup και από βάσεις δεδομένων που αντιστοιχούν ip διευθύνσεις σε γεωγραφικές περιοχές.

Το Reverse DNS Lookup (ή name resolution) είναι μηχανισμός των DNS Server [52] που μπορεί να αντιστοιχεί μία IP διεύθυνση σε πλήρες όνομα υπολογιστή (όνομα υπολογιστή + όνομα περιοχής, Fully Qualified Domain Name ή FQDN). Οι καταλήξεις των ονομάτων περιοχής (top level domain - TLD) δίνουν ένδειξη για τη χώρα προέλευσης του επισκέπτη, όταν δεν αναφέρονται στα TLDs γενικού τύπου (com, net, org, gov, mil, int) αλλά στους κωδικούς των χωρών (gr για Ελλάδα, de

για Γερμανία κλπ.). Πέραν των χωρών προέλευσης, μεγάλη αξία έχουν και τα υπόλοιπα τμήματα των FQDN. Ειδικά όταν οι επισκέπτες έχουν πρόσβαση από οργανισμούς και δεν έχουν πρόσβαση από παροχέα Διαδικτύου, το όνομα δίνει μία ένδειξη για τον οργανισμό απ' όπου γίνεται η πρόσβαση. Στον πίνακα 2.12 φαίνονται ενδεικτικά ποσοστά επισκεπτών ανάλογα με το όνομα περιοχής τους από το [43]. Το όνομα περιοχής auth.gr που ανήκει στο Αριστοτέλειο Πανεπιστήμιο κατέχει το 0,11% των επισκέψεων ενώ το intel.com το 0,09% των επισκέψεων.

<b>Όνομα Περιοχής</b>	<b>Ποσοστό Επισκέψεων</b>
t-online.de	2,28%
cha.forthnet.gr	1,90%
tuc.gr	0,15%
auth.gr	0,11%
Intel.com	0,09%

Πίνακας 2.12 Ενδεικτικά ποσοστά επισκέψεων ανά όνομα περιοχής για το [43]

Ειδικά όταν η πρόσβαση γίνεται από παρόχους υπηρεσιών Διαδικτύου (internet Service Providers – ISP) πολλές φορές από το όνομα τη περιοχής μπορεί να εξαχθούν συμπεράσματα για τη συγκεκριμένη πόλη στην οποία βρίσκεται ο επισκέπτης. Για παράδειγμα ο ελληνικός ISP FORTHnet ακολουθεί στην ονοματολογία των ονομάτων περιοχής τον κανόνα να προσθέτει πριν την περιοχή forthnet.gr ένα ακρωνύμιο τριών χαρακτήρων που αντιστοιχεί στην πόλη απ' όπου συνδέθηκε ο επισκέπτης. Για παράδειγμα στο [43] το 1,9% των επισκεπτών έχει κατάληξη [cha.forthnet.gr](http://cha.forthnet.gr) που ανήκει στην πόλη Χανιά. Παρόμοιες τακτικές ακολουθούν σχεδόν όλοι οι ISP διεθνώς οπότε γνωρίζοντας τους κανόνες που ακολουθούν στην ονοματολογία δίνεται μία αρκετά ακριβής προσέγγιση για την περιοχή διαμονής των επισκεπτών. Υπάρχουν εξειδικευμένες υπηρεσίες που διαθέτουν αναλυτικές πληροφορίες για τέτοιου είδους αντιστοιχίσεις.

Ο δεύτερος τρόπος εξαγωγής πληροφορίας από τη διεύθυνση IP βασίζεται σε βάσεις δεδομένων που διατίθενται, στις οποίες υπάρχει αντιστοίχιση καθ' εαυτής της διεύθυνσης σε συγκεκριμένη περιοχή. Η λεπτομέρεια της αντιστοίχισης ποικίλει ανάλογα με τη χώρα. Έτσι για περιοχές των ΗΠΑ μπορεί να δοθεί πληροφορία μέχρι και σε επίπεδο Ταχυδρομικού Κώδικα [53]!

Βέβαια η προϋπόθεση για τα δεδομένα από τη διεύθυνση IP είναι να είναι αυτή διαθέσιμη, που όπως αναπτύχθηκε σε προηγούμενες ενότητες, κυρίως οι μηχανισμοί των εξυπηρετητών proxy αποκρύπτουν αυτή την πληροφορία. Ένδειξη ότι η IP διεύθυνση που φτάνει στον εξυπηρετητή WWW δεν ανήκει στην πραγματική διεύθυνση του επισκέπτη, αποτελεί η ύπαρξη του πεδίου VIA του HTTP πρωτοκόλλου, η οποία αναφέρει από ποιους ενδιάμεσους σταθμούς πέρασε η αίτηση.

Ορισμένοι, πολύ δημοφιλείς ([54]), εξυπηρετητές proxy προσθέτουν στα πεδία του HTTP πρωτοκόλλου το πεδίο X-Forwarded-For (πίνακας 2.5) στο οποίο αποθηκεύουν την πραγματική διεύθυνση του επισκέπτη. Αν και αυτό το πεδίο δεν ανήκει επίσημα στο HTTP πρωτόκολλο, η ύπαρξη αυτού του πεδίου εξασφαλίζει τη γνώση για τον πραγματικό επισκέπτη.

## Referrers

Το HTTP πεδίο `referrer` περιγράφηκε ήδη στις προηγούμενες ενότητες (Πίνακας 2.5, Πίνακας 2.6, Ενότητα 2.3.4.1, ανακατασκευή μονοπατιού). Δίνει την προέλευση από την οποία γίνεται η κάθε αίτηση στον ιστοχώρο. Στην ενότητα 2.3.4.1 περιγράφηκε η χρησιμότητά του όταν το `referrer` δείχνει σε ιστοχώρο που βρίσκεται εκτός του εξεταζόμενου ιστοχώρου. Από στατιστικής άποψης και μόνο παρουσιάζει μεγάλο ενδιαφέρον η γνώση για τους συνδέσμους που υπάρχουν σε άλλους ιστοχώρους ώστε να είναι δυνατή η εκτίμηση της “δημοφιλίας” του ιστοχώρου, πιθανώς σε σύγκριση με άλλους ανταγωνιστικούς. Μπορεί επίσης να αποκαλύψει ελλείψεις στην προβολή του ιστοχώρου όταν δεν εμφανίζονται επισκέψεις από ιστοχώρους απ' όπου θα περίμενε κανείς να έχει έναν αριθμό επισκέψεων. Για παράδειγμα ένας ιστοχώρος με κτηματομεσιτικές υπηρεσίες για μία περιοχή, θα περίμενε κανείς να έχει επισκέπτες από έναν ιστοχώρο που παρουσιάζει κτηματομεσίτες της ίδιας περιοχής.

Ιδιαίτερο ενδιαφέρον παρουσιάζουν οι μηχανές αναζήτησης επειδή συνήθως είναι διαθέσιμη πληροφορία για το περιεχόμενο της αναζήτησης που διενήργησε ο χρήστης πριν φτάσει στον εξεταζόμενο ιστοχώρο. Η οργάνωση των φράσεων αναζήτησης που χρησιμοποιήθηκαν για να φτάσουν οι επισκέπτες στον ιστοχώρο μπορεί να απαντήσει σε κρίσιμα ερωτήματα όπως

- τι ψάχνουν να βρουν οι επισκέπτες;
- καλύπτει ο ιστοχώρος το πεδίο για το οποίο ενδιαφέρονται οι επισκέπτες;
- βρήκαν οι επισκέπτες αυτό για το οποίο ενδιαφέρονταν (επισκέφτηκαν τις αντίστοιχες σελίδες);
- έρχονται επισκέπτες που ψάχνουν για θέματα άσχετα με το αντικείμενο του ιστοχώρου;
- καλύπτονται όλα τα θέματα του ιστοχώρου από αντίστοιχες επισκέψεις; Χρειάζεται μεγαλύτερη προβολή για κάποια θέματα;

Στη βιβλιογραφία δε δίνεται ιδιαίτερη έμφαση στα παραπάνω δεδομένα, παρόλο που όλες οι εργασίες αναγνωρίζουν την ύπαρξή τους και σε μεγάλο ή μικρότερο βαθμό τα συμπεριλαμβάνουν

στις έρευνές τους.

### **2.3.5.9 Δεδομένα από Εξωτερικές Πηγές**

Τα δεδομένα από εξωτερικές πηγές που αναφέρονται στη βιβλιογραφία περιορίζονται συνήθως σε στοιχεία για τα ενδιαφέροντα των πελατών ηλεκτρονικών καταστημάτων. Για παράδειγμα διατηρούνται εκτεταμένες πληροφορίες για την οικονομική ευρωστία των κατοίκων ανάλογα με την περιοχή διαμονής, και συσχετίζεται με τους επισκέπτες των ιστοχώρων όταν αυτοί εισάγουν για την εκτέλεση μιας παραγγελίας τον ταχυδρομικό κώδικα της διεύθυνσης αποστολής.

Στην πράξη μπορεί να χρησιμοποιηθούν κάθε είδους δεδομένων που εμπλουτίζουν τη διαθέσιμη πληροφορία είτε για τους επισκέπτες είτε για το περιεχόμενο του ιστοχώρου. Για παράδειγμα, όπως θα περιγραφεί αναλυτικά σε επόμενες ενότητες, παρατηρήθηκε σημαντική διαφορά στις προτιμήσεις των επισκεπτών του [43] σε τοποθεσίες ανάλογα με τη χώρα προέλευσης των επισκεπτών. Συνδέοντας τις τοποθεσίες με γεωγραφική πληροφορία ανακαλύφθηκε ότι οι τοποθεσίες ανήκουν σε συγκεκριμένες γεωγραφικές περιοχές.

### **2.3.5.10 Συμπεράσματα**

Σε αυτή την ενότητα παρουσιάστηκαν ορισμένες εργασίες που εκτελούνται στο στάδιο της προεπεξεργασίας των πρωτογενών δεδομένων. Η βιβλιογραφία που είναι διαθέσιμη ερευνά κυρίως τους τρόπους που θα οδηγήσουν σε αξιόπιστα δεδομένα όσον αφορά τις κινήσεις ανά επισκέπτη στους ιστοχώρους. Ένας έμμεσος στόχος είναι επίσης η μείωση των δεδομένων προς επεξεργασία.

Η ασάφεια που προκύπτει από τους τεχνικούς περιορισμούς στην επικοινωνία με τους εξυπηρετητές HTTP αναγκάζει τους μελετητές να χρησιμοποιήσουν ευριστικές μεθόδους για το διαχωρισμό των κινήσεων των επισκεπτών και για τον υπολογισμό των ακριβή μονοπατιών που χρησιμοποίησαν.

Ο όγκος και η πολυπλοκότητα της αρχικής επεξεργασίας των πρωτογενών δεδομένων πολλές φορές υποτιμάται. Τα δεδομένα που συλλέγονται με τον έναν ή τον άλλο τρόπο είναι τεράστια σε όγκο και η ακρίβεια των δεδομένων που θα προκύψουν μετά την προεπεξεργασία καθορίζουν και την αξιοπιστία των συμπερασμάτων που θα προκύψουν μετά την ανάλυση και επεξεργασία τους. Στο [22] αναφέρεται χαρακτηριστικά ότι “...*data preparation... which has proven the most complex and time-consuming task in almost all web-mining projects*”.

### **2.3.6 Ανακάλυψη Γνώσης (Knowledge Discovery)**

Από το στάδιο της προεπεξεργασίας των πρωτογενών δεδομένων έχουν προκύψει δομημένα δεδομένα για τις επισκέψεις στον υπό ανάλυση ιστοχώρο. Τα δεδομένα είναι απαλλαγμένα από

μεγάλο μέρος των στοιχείων που δημιουργούνται από την επικοινωνία με το πρωτόκολλο HTTP, αλλά διατηρούν την ουσιαστική πληροφορία για την πλοήγηση των επισκεπτών, ενώ έχουν οργανωθεί σε πιο αφηρημένες έννοιες όπως η έννοια της επίσκεψης και του επισκέπτη, και έχουν εμπλουτιστεί με επιπλέον πληροφορίες που τυχόν είναι διαθέσιμες για τους επισκέπτες.

Παρόλο που με την αφαίρεση των προσβάσεων για εικόνες και γενικά μη ουσιαστικών δεδομένων έχει μειωθεί πολύ ο όγκος των δεδομένων, τα εναπομείναντα δεδομένα συνεχίζουν να απαιτούν τεράστιους αποθηκευτικούς χώρους και η επεξεργασία τους μεγάλους υπολογιστικούς πόρους.

Αν και δεν αποκλείονται εξειδικευμένοι τρόποι αποθήκευσης των δεδομένων, συνήθως χρησιμοποιούνται σχεσιακές βάσεις δεδομένων οι οποίες έχουν καθιερωθεί στην αγορά και επιτρέπουν την υποβολή ερωτήσεων με καθιερωμένες γλώσσες ερωτήσεων ή τη χρήση γνωστών μεθόδων Data Mining [2]. Πολύ συχνά επίσης χρησιμοποιείται τεχνολογία OLAP (Online Analytical Processing) για τη γρήγορη πλοήγηση μέσα στα δεδομένα. Το OLAP θα περιγραφεί αναλυτικά στο στάδιο της ανάλυσης.

Στο στάδιο της ανακάλυψης γνώσης ο κύριος στόχος είναι να εντοπιστούν συσχετίσεις ανάμεσα στα δεδομένα που δεν είναι ορατές με μία απλή επισκόπηση ή με απλή στατιστική ανάλυση. Οι συσχετίσεις εκφράζονται για παράδειγμα με τη διατύπωση κανόνων ή την απαρίθμηση ενός συνόλου χαρακτηριστικών που οδηγούν σε ομαδοποιήσεις χρηστών. Χρησιμοποιούνται τεχνικές και μεθοδολογίες από πολλές επιστημονικές περιοχές, όπως στατιστική, data mining, machine learning και αναγνώριση προτύπων [14].

Ξεχωριστή βαρύτητα έχει η ανακάλυψη σχημάτων στα μονοπάτια επισκέψεων (Data Mining on User Navigation Patterns), όπου έχει κατευθυνθεί μεγάλο μέρος της έρευνας.

Οι τεχνικές που χρησιμοποιούνται για data mining είναι γνωστές από άλλες περιοχές, και έτσι σε αυτή την ενότητα θα δοθεί μία γενική περιγραφή για τις ενέργειες data mining που γίνονται στα δεδομένα από το χώρο του παγκόσμιου ιστού.

Είναι σημαντικό να τονιστεί ότι παρόλο που οι τεχνικές χρησιμοποιούνται και περιγράφονται παρακάτω (association rules, clustering κλπ.), δεν είναι άγνωστες, η εφαρμογή τους στα πρωτογενή δεδομένα από την κίνηση στους ιστοχώρους παρουσιάζει προβλήματα που χρήζουν περαιτέρω μελέτης. Όπως επισημαίνεται στο [59] *“the State of the Art in web mining is quite primitive”*. Εφαρμόζονται τεχνικές, που είναι προσαρμοσμένες σε αριθμητικά δεδομένα, για δεδομένα που δεν είναι αριθμητικά ενώ μέθοδοι που έχουν προταθεί βασίζονται σε ξεκάθαρα δεδομένα και σχέσεις ενώ για το web mining οι σχέσεις που εξετάζονται (π.χ. όμοιοι χρήστες) είναι μάλλον ασαφής.

Η οργάνωση ακολουθεί την παρουσίαση του [14] και [3] ενώ για μία πιο ολοκληρωμένη περιγραφή, ειδικά για την περιοχή ανάλυσης επισκεψιμότητας, ο αναγνώστης μπορεί να ανατρέξει στο [40].

### 2.3.6.1 Association Rules

Η ανακάλυψη association rules συσχετίζει σελίδες που επισκέπτονται μαζί στην ίδια επίσκεψη με συχνότητα εμφάνισης (support) που υπερβαίνει ένα ορισμένο κατώφλι. Αυτές οι σελίδες δεν πρέπει να έχουν σύνδεσμο μεταξύ τους, μιας και έτσι η συσχέτιση θα ήταν προφανής [14]. Στο [3] ο ορισμός γενικεύεται για συσχετίσεις αντικειμένων οποιασδήποτε μορφής, όπως η ενέργεια για αγορά προϊόντος. Στο [18] τα αντικείμενα που συσχετίζονται μπορεί να είναι και σύνολα πολλών αντικειμένων. Τα association rules συνοδεύονται συνήθως από δύο μεγέθη που εκφράζουν τη συχνότητα εμφάνισης του σχήματος και το βαθμό συσχέτισης των αντικειμένων (support και confidence αντίστοιχα). Μαθηματικά τα δύο μεγέθη περιγράφονται στον πίνακα 2.13.

Έστω A και B συναντώνται στο 1% των συναλλαγών, αλλά όποτε συναντάται το A υπάρχει 80% πιθανότητα να συναντηθεί και το B. Το support είναι το 1% της ταυτόχρονης εμφάνισης του A και B, ενώ 80% είναι το confidence του κανόνα.

<i>Support &amp; Confidence</i>	
Support s	$s(A \Rightarrow B) = P(A \cup B)$
Confidence c	$c(A \Rightarrow B) = P(B A) = \frac{P(A \cup B)}{P(A)}$

Πίνακας 2.13 Support & Confidence σε Association Rules από το [3]

Για τον υπολογισμό των κανόνων χρησιμοποιείται η έννοια του entity και του attribute, όπου ενδιαφέρει η ύπαρξη των attribute στο ίδιο entity. Στην περίπτωση της ανάλυσης επισκεψιμότητας ως entity μπορεί να θεωρηθεί η επίσκεψη ή ένα επεισόδιο και ως attribute οι σελιδοπροβολές που γίνονται ή για παράδειγμα η αγορά προϊόντων. Τα δεδομένα πρέπει να βρίσκονται είτε σε “οριζόντια” είτε σε “κάθετη” δομή όπου η κάθετη δομή βρίσκεται πιο κοντά στο σχεσιακό μοντέλο και είναι προτιμητέα [3]. Στην περίπτωση της κάθετης δομής τα δεδομένα συνδέουν το αναγνωριστικό της επίσκεψης με τις σελιδοπροβολές, χρησιμοποιώντας ένα tuple ανά σελιδοπροβολή (πίνακας 2.14). Στην περίπτωση που τα attributes έχουν συνεχόμενες τιμές, αυτές πρέπει να ενωθούν σε κλίμακες και να χρησιμοποιηθούν αυτές.

Στην περίπτωση που έχουν δημιουργηθεί ιεραρχικές έννοιες, αντί των σελιδοπροβολών μπορεί να συσχετιστούν οι έννοιες αυτές, δίνοντας μία πιο υψηλού επιπέδου άποψη για την επισκεψιμότητα, και μειώνοντας ταυτόχρονα τους δυνατούς συνδυασμούς και την πολυπλοκότητα στον υπολογισμό.

<i>SessionId/EpisodeId</i>	<i>PageId</i>
132	pageA
428	pageB
428	pageC
428	pageD
132	pageE

Πίνακας 2.14 Η "κάθετη" δομή για τον υπολογισμό association rules

Οι αλγόριθμοι που χρησιμοποιούνται είναι συνήθως παραλλαγές του Apriori [55]. Η αρχή λειτουργίας του Apriori βασίζεται στην αρχική δημιουργία κανόνων (candidate patterns) και στον μετέπειτα έλεγχο εάν αυτοί έχουν αρκετό support στη βάση δεδομένων [1]. Η λεπτομερής περιγραφή των αλγορίθμων είναι εκτός ενδιαφέροντος της παρούσας διατριβής και ο ενδιαφερόμενος αναγνώστης μπορεί να ανατρέξει για παραδείγματα και περιγραφές στα [3], [55], [56] ενώ για εφαρμογές με σχεσιακές βάσεις δεδομένων στο [57].

Παραδείγματα συσχετίσεων που μπορεί να βρεθούν είναι:

*30% των επισκεπτών που επισκέφθηκαν τη σελίδα /announcements/specialoffer.html έδωσαν παραγγελία στη σελίδα /products/product1*

Στην περίπτωση που έχουν δημιουργηθεί ιεραρχικές έννοιες για τις σελίδες, οι κανόνες μπορεί να αναφερθούν και σε αυτές. Για παράδειγμα

*40% των επισκεπτών που επισκέφθηκαν σελίδες μινωικών αξιοθεάτων έδωσαν παραγγελία για βιβλίο*

Επειδή τα δεδομένα προς επεξεργασία είναι μεγάλα σε όγκο, πολλές φορές γίνεται προσπάθεια για περιορισμό των συνδυασμών που πρέπει να εξεταστούν. Η γενίκευση των attributes σε ιεραρχίες είναι ένα σημαντικό βήμα προς αυτή την κατεύθυνση [3]. Παράδειγμα αυτοματοποιημένης γενίκευσης σελίδων παρουσιάζεται στο [23]. Επίσης μπορεί να γίνει εκμετάλλευση της γνώσης που υπάρχει για τη δομή του ιστοχώρου για τη μείωση των συνδυασμών που θα εξεταστούν. Έστω για παράδειγμα ότι το support για τη *A* είναι πολύ χαμηλό. Εάν οι σελίδες *B* και *C* είναι μόνο προσβάσιμες δια μέσω της *A* τότε μπορεί να αγνοηθούν οι σελίδες αυτές επειδή το support τους θα είναι πολύ μικρό [18].

### 2.3.6.2 Sequential Patterns

Η ανακάλυψη σειριακών σχημάτων αποτελεί ουσιαστικά την ανακάλυψη association rules εμπλουτισμένα με χρονικούς περιορισμούς. Δηλαδή ενδιαφέρουν γεγονότα της μορφής AB όπου



υπάρχει όμως συγκεκριμένη σειρά στα γεγονότα (το A προηγήθηκε του B). Δημιουργεί κανόνες με ορισμένο support και confidence από τους οποίους εξάγεται συμπέρασμα για την ύπαρξη ενός συνόλου ενεργειών δοθέντος ότι προηγήθηκε ένα άλλο σύνολο ενεργειών. Αυτός ο τομέας έχει άμεση εφαρμογή στην ανακάλυψη σχημάτων στην πλοήγηση των επισκεπτών στους ιστοχώρους και στην πρόβλεψη μελλοντικών επισκέψεων και συμπεριφορών των χρηστών.

Δύο παραδείγματα πιθανών κανόνων φαίνονται στον πίνακα 2.15. Ο βασικός περιορισμός συνίσταται στην αλληλουχία των γεγονότων, τα οποία καθ' εαυτά αποτελούν εκφράσεις σελιδοπροβολών. Έτσι στο πρώτο παράδειγμα, η ενοικίαση δωματίου είναι μία συγκεκριμένη σελιδοπροβολή (ορίζεται από το URL με τις παραμέτρους του) *A* όπως και η ενοικίαση αυτοκινήτου *B*. Ο κανόνας είναι *AB* με *support 30%*.

<i>Παραδείγματα Κανόνων από Sequential Pattern Analysis</i>
30% των επισκεπτών που κλείνουν δωμάτιο στην περιοχή A ( <b>A</b> ), έκλεισαν ενοικιαζόμενο αυτοκίνητο μέσα σε 30 μέρες μετά από το δωμάτιο ( <b>B</b> )
10% των επισκεπτών που παραγγέλνουν βιβλίο για θαλάσσιο αθλητισμό, επισκέφτηκαν τον ιστοχώρο την τελευταία εβδομάδα προερχόμενοι από το Google αναζητώντας με όρους w1 και w2

Πίνακας 2.15 Παραδείγματα κανόνων από Sequential Pattern Analysis

Η κάθε επίσκεψη μπορεί να μοντελοποιηθεί ως μονοπάτι σε ένα γράφο όπου κάθε κόμβος του γράφου αναπαριστά μία σελιδοπροβολή και οι ακμές συνδέσμους ανάμεσα σε σελίδες. Οι σύνδεσμοι δεν είναι αναγκαστικά σύνδεσμοι υπερκειμένου διότι ο επισκέπτης φτάνοντας σε μία σελίδα μπορεί να μεταβεί σε οποιαδήποτε άλλη (ανεξαρτήτως συνδέσμου υπερκειμένου) εισάγοντας το πλήρη URL τη επόμενης σελίδας<sup>1</sup>. Από αυτή την άποψη πρόκειται για έναν πλήρη συνδεδεμένο γράφο (fully connected graph). Ακόμα και εάν υποθέσουμε ότι οι επισκέπτες χρησιμοποιούν μόνο τους συνδέσμους υπερκειμένου που προσφέρονται στον ιστοχώρο, σε μεγάλους ιστοχώρους υπάρχει πολύ μεγάλη διασυνδεσιμότητα των κόμβων.

Με σκοπό την απλοποίηση του γράφου μπορεί να ομαδοποιηθούν οι σελιδοπροβολές που ανήκουν σε σημασιολογικά όμοιες ενέργειες και να μειωθεί έτσι η πολυπλοκότητα. Για παράδειγμα σε έναν ιστοχώρο τουριστικού προορισμού μπορεί όλες οι σελιδοπροβολές που σχετίζονται με την εύρεση διαμονής να ομαδοποιηθούν σε μία έννοια “εύρεση διαμονής”. Αυτές οι έννοιες αναφέρονται ως επεισόδια ή συναλλαγές (episodes- transactions). Η επίσκεψη τότε μοντελοποιείται σαν ακολουθία τέτοιων επεισοδίων.

Στο [18] για τον υπολογισμό σειριακών σχημάτων χρησιμοποιείται μία παραλλαγή του Apriori

<sup>1</sup> Στην πραγματικότητα υπάρχουν σελίδες που προσδιορίζονται από συνδυασμό URL και επιπλέον παραμέτρων με τη μέθοδο HTTP POST, οπότε δεν αρκεί το URL από μόνο του.

Αλγορίθμου [55]. Στο [58] δίνεται ο ορισμός του *g-sequence* (generalized sequence) όπου εισάγονται δύο νέοι δυνατοί περιορισμοί στην αναζήτηση σειρών: η δυνατότητα χρήσης wild-cards, όπου το wild-card συμβολίζει σελιδοπροβολές απροσδιόριστου αριθμού και είδους, και η δυνατότητα χρήσης στατιστικών περιορισμών στις σελιδοπροβολές, όπως πόσες φορές έχει προβληθεί μία συγκεκριμένη σελίδα. Οι δύο αυτές επιπλέον δυνατότητες επιτρέπουν την αναζήτηση σειρών της μορφής  $(A.I) * (B.I)$  όπου  $(A.I)$  συμβολίζει την πρώτη επίσκεψη στη σελίδα A και \* είναι wild-card. Επίσης ορίζεται η έννοια του *template* σαν δομή προσδιορισμού των σειρών που αναζητούνται. Και στο [58] χρησιμοποιείται ως βάση ο Apriori αλγόριθμος για τον υπολογισμό των *g-sequences*. Η γλώσσα για την υποβολή ερωτήσεων για την ανακάλυψη σχημάτων στα μονοπάτια περιγράφεται στο [71]. Μία επέκταση που επικεντρώνεται περισσότερο στην επιβεβαίωση υποθέσεων περιγράφεται στο [72].

Στο [1] προτείνεται ο αλγόριθμος *WAP-mine* που συγκρινόμενος με παραλλαγή του Apriori αλγόριθμου παρουσιάζεται ως μία τάξη μεγέθους ταχύτερος στην ανακάλυψη σειριακών σχημάτων, και χρησιμοποιεί WAP-trees (Web Access Pattern Trees). Στον αλγόριθμο απαιτείται μόνο 2 φορές η σάρωση των δεδομένων για την κατασκευή του δέντρου, και λειτουργεί μετέπειτα μόνο με τα δεδομένα της δεντρικής δομής.

### 2.3.6.3 Classification & Clustering

Το classification και clustering είναι τεχνικές που προσφέρουν αντιστοίχιση ανάμεσα σε αντικείμενα και κατηγορίες ή ομάδες ομοειδών.

Στο **classification** υπάρχουν ήδη δημιουργημένες κατηγορίες, που έχουν καθοριστεί από προηγούμενη γνώση. Εάν έχουν εξαχθεί χαρακτηριστικά επισκεψιμότητας για τις κατηγορίες που έχουν καθοριστεί, στο classification γίνεται προσπάθεια με βάση αυτά τα χαρακτηριστικά να γίνει αντιστοίχιση αντικειμένων στις κατηγορίες. Από αυτή την άποψη είναι ένα εργαλείο πρόβλεψης συμπεριφοράς των χρηστών επειδή κατατάσσοντας τους χρήστες σε συγκεκριμένες κατηγορίες αναμένεται ότι οι χρήστες θα συμπεριφερθούν όμοια με τους υπόλοιπους χρήστες της κατηγορίας. Στο [40] ο ορισμός που δίνεται για το classification είναι: “*Η διαδικασία χωρισμού συνόλου δεδομένων σε αμοιβαία αποκλειόμενες ομάδες έτσι ώστε τα μέλη των ομάδων να βρίσκονται όσο γίνεται πιο “κοντά” μεταξύ τους, και διαφορετικές ομάδες να βρίσκονται όσο γίνεται πιο “μακριά” μεταξύ τους, όπου η απόσταση μετριέται με βάση μία μεταβλητή που προσπαθούμε να προβλέψουμε*”. Η ανάγκη ύπαρξης προηγούμενων κλάσεων χαρακτηρίζει το classification ως “*supervised learning*”. Παραδείγματα κανόνων classification μπορεί να είναι

- clients who often access /faq/master\_vs\_phd tend to be from educational institutions

- 50% of clients who placed an online order in `/company/product2` were in the 20-25 age group and lived on the West Coast

Στα παραπάνω παραδείγματα οι κατηγορίες που έχουν δημιουργηθεί και στις οποίες έγινε η αντιστοίχιση είναι *“be from educational institution”*, *“20-25 age group”* και *“living on West Coast”*.

Σε αντίθεση με το classification, στο **clustering** δεν υπάρχει προηγούμενη γνώση για τα δεδομένα, και ο σκοπός είναι η καθ' αυτή δημιουργία ομοειδών ομάδων ή κατηγοριών. Εξετάζονται τα δεδομένα και με βάση κάποιων χαρακτηριστικών που επιλέγονται, ομαδοποιούνται τα αντικείμενα (για παράδειγμα οι επισκέπτες) με βάση την ομοιότητα στα χαρακτηριστικά. Σε αντίθεση επίσης με το classification το clustering χαρακτηρίζεται ως *“unsupervised learning”*.

Εάν το clustering γίνεται με βάση τα ονόματα των URL, τότε το clustering εκφυλίζεται στην ανακάλυψη association rules [18]. Ένας τρόπος για την ημιαυτόματη δημιουργία κατηγοριών για classification είναι η χρήση των κατηγοριών που βρέθηκαν από το clustering και ο εμπλουτισμός των κατηγοριών αυτών με δημογραφικά δεδομένα [18].

Σαν παράδειγμα για εφαρμογές clustering μπορεί να αναφερθεί το παρακάτω από το [23]:

*Ένας αριθμός επισκεπτών περνάει πολύ χρόνο σε σελίδες με θέμα “έπιπλα για μωρά”, “παιχνίδια για μωρά” και “πάνες”. Οι επισκέπτες ομαδοποιούνται σε κατηγορία στην οποία οι εξειδικευμένοι πάνω στον ιστοχώρο δίνουν το όνομα “αναμένοντες γονείς”. Σαν αποτέλεσμα της ανάλυσης στις τρεις θεματικές ενότητες προστίθενται σύνδεσμοι μεταξύ τους και επίσης όταν επισκέπτονται 2 από τις 3 κατηγορίες σελίδων δημιουργείται δυναμικά ειδικός σύνδεσμος στην τρίτη κατηγορία.*

Στο [23] γίνεται clustering των χρηστών με βάση τις σελίδες που επισκέπτονται ανά επίσκεψη και το χρόνο που δαπανούν ανά σελιδοπροβολή. Ως χρήστες θεωρούνται οι επισκέψεις οι οποίες προκύπτουν με κάθε νέα εμφάνιση IP διεύθυνσης και χρόνο αδράνειας 30 λεπτά. Για την αποφυγή μεγάλων διαστάσεων στα διανύσματα των δυνατών σελίδων, δημιουργείται μία ιεραρχία των σελίδων όπου τα φύλα θεωρούνται *απλές σελίδες* ενώ οι ενδιάμεσοι κόμβοι *γενικές σελίδες*. Οι σελιδοπροβολές απλών σελίδων αντικαθιστούνται με τις γενικές σελίδες της ιεραρχίας και ο χρόνος προβολής αθροίζεται για κάθε γενική σελίδα. Οι επισκέψεις που προκύπτουν ονομάζονται *generalized sessions* και πάνω σε αυτές εφαρμόζεται ο αλγόριθμος BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies). Αν και η δημιουργία των επισκέψεων είναι προβληματική για τους λόγους που παρουσιάστηκαν στην ενότητα για την προεπεξεργασία των δεδομένων, και η ιεραρχία των σελίδων είναι περιοριστική για σύγχρονους ιστοχώρους, η

συγκεκριμένη μέθοδος δίνει καλά αποτελέσματα σε καλούς χρόνους.

Στο [59] και τις σχετικές με αυτό δημοσιεύσεις από την ίδια ομάδα προτείνεται μία μετρική ομοιότητας ανάμεσα σε επισκέψεις (user session similarity measure) βασισμένη σε ασαφή λογική, η οποία περιλαμβάνει τα URL που επισκέφθηκαν οι χρήστες και τη δομή του ιστοχώρου.

Είναι σημαντικό να τονιστεί ότι επειδή στις τεχνικές για classification απαιτείται προηγούμενη γνώση για τη δημιουργία των κατηγοριών και η σύνδεσή τους με συγκεκριμένα χαρακτηριστικά, το αποτέλεσμα εξαρτάται άμεσα από την ποιότητα αυτής της γνώσης, και σε περίπτωση λανθασμένων εκτιμήσεων μπορεί να οδηγήσει σε λάθος συμπεράσματα.

### **2.3.7 Ανάλυση**

Το στάδιο της ανάλυσης είναι ίσως το πιο σημαντικό στη διαδικασία της ανάλυσης επισκεψιμότητας επειδή σε αυτό έχουν συγκεντρωθεί όλα τα στοιχεία για την κίνηση του ιστοχώρου από τα οποία θα κριθεί η περαιτέρω πορεία του.

Για κάποιον που ενδιαφέρεται για το εάν οι επισκέπτες προερχόμενοι από μηχανές αναζήτησης μπορεί να καλυφθούν από τις υπηρεσίες του ιστοχώρου, αρκεί μία ανάλυση των φράσεων που χρησιμοποιήθηκαν στις μηχανές αναζήτησης. Εάν στο μεγαλύτερό τους μέρος αναφέρονται σε θέματα που καλύπτονται από τον ιστοχώρο, τότε ο ιστοχώρος μπορεί να θεωρηθεί επιτυχημένος. Εάν οι φράσεις είναι άσχετες με τον ιστοχώρο, πρέπει να αναλυθούν οι λόγοι που οδηγούν σε αυτή την κατάσταση και να τροποποιηθούν αναλόγων οι σελίδες. Εάν πάλι οι φράσεις καλύπτουν μόνο μέρος των υπηρεσιών του ιστοχώρου πρέπει να τροποποιηθούν οι σελίδες ώστε να πλεονεκτούν περισσότερο στις επιπλέον φράσεις στις μηχανές αναζήτησης.

Εάν ενδιαφέρει η τοποθέτηση διαφημίσεων στον ιστοχώρο, αρχικά πρέπει να εκτιμηθεί πόσες διαφημίσεις (προβολές διαφήμισης) μπορεί να καλύψει ο ιστοχώρος. Ο αριθμός των σελιδοπροβολών ανά ημέρα ήδη μπορεί να δώσει αυτή την πληροφορία. Στην περίπτωση που η διαφήμιση επιδιώκεται να προσαρμοστεί περισσότερο στα ενδιαφέροντα των επισκεπτών, πρέπει να υπάρχει γνώση για τους επισκέπτες και να συμπεριληφθούν στην ανάλυση τα αποτελέσματα από την ανακάλυψη cluster χρηστών ή το classification.

Εάν πάλι ενδιαφέρει η καλύτερη εξυπηρέτηση των επισκεπτών σε όρους ταχύτητας ανεύρεσης πληροφορίας ή ταχύτητας εξυπηρέτησης, μπορεί να χρησιμοποιηθεί η γνώση από την ανακάλυψη association rules ώστε να βελτιστοποιηθεί ο ιστοχώρος ανάλογα με τις σελίδες ή υπηρεσίες που γίνεται προσπέλαση μαζί και να ελαχιστοποιηθεί ο αριθμός των επιλογών (σε κλικ) που πρέπει να κάνουν οι χρήστες.

Παραδείγματα όπως τα παραπάνω μπορεί να αναφερθούν πολλά και σε κάθε περίπτωση τα ακριβή βήματα που θα γίνουν και οι μέθοδοι που θα χρησιμοποιηθούν εξαρτώνται απόλυτα από τους στόχους που έχουν τεθεί και τις ιδιαιτερότητες του κάθε ιστοχώρου, και δεν υπάρχει κάποιο τυποποιημένο περίγραμμα ανάλυσης που μπορεί να ακολουθηθεί. Από αυτή την άποψη η ποιότητα και ακρίβεια των συμπερασμάτων εξαρτάται άμεσα από τους ανθρώπους που διενεργούν την ανάλυση και τους ρόλους που έχουν στο συγκεκριμένο ιστοχώρο. Επειδή η αναφορά γίνεται πάντα σε ένα άκρως τεχνικό θέμα, οι τεχνικοί του ιστοχώρου (διαχειριστές ιστοχώρου, δικτυακής εγκατάστασης, σχεδιαστές και κατασκευαστές ιστοχώρου) μπορεί να έχουν μεγάλη συμβολή στην ανάλυση. Μία απότομη πτώση των σελιδοπροβολών ανά επίσκεψη μπορεί να δικαιολογηθεί για παράδειγμα από προσωρινά προβλήματα ταχύτητας διασύνδεσης με το δίκτυο. Επίσης η αξιολόγηση των ανακαλυφθέντων association rules απαιτεί την εις βάθος γνώση της δομής και της διασύνδεσης των διαφόρων σελίδων. Ειδικοί σε μάρκετινγκ απαιτούνται στην αξιολόγηση τυχών διαφημιστικών προβολών του ιστοχώρου ενώ πολύ σημαντική είναι επίσης να συμμετέχουν στην ανάλυση άτομα που έχουν άριστη γνώση του αντικειμένου που πραγματεύεται ο ιστοχώρος (domain knowledge) ώστε να μπορούν να συνδυάσουν τη γνώση από μέσα προβολής του ίδιου περιεχομένου που δεν αναφέρονται στο Διαδίκτυο.

Ιδιαίτερη σημασία στη διαδικασία ανάλυσης έχει η ευχέρεια που παρέχεται στους αναλυτές να ανατρέχουν με ευκολία και ιδιαίτερα με μεγάλη ταχύτητα στα δεδομένα που έχουν συγκεντρωθεί. Να υπενθυμηθεί ότι αν και επεξεργασμένα και καθαρισμένα, τα δεδομένα είναι πολύ μεγάλα σε όγκο και οι συσχετισμοί που υπάρχουν μεταξύ τους πολύπλοκοι. Πολύ μεγάλη σημασία έχει επίσης η δυνατότητα οπτικοποίησης των αριθμητικών δεδομένων ώστε να μπορεί να φανούν οι μεταβολές των μετρικών ανάλογα με τις παραλλαγές στα χαρακτηριστικά που εξετάζονται. Τα δύο τελευταία σημεία καλύπτονται από εφαρμογές OLAP (Online Analytical Processing) [60] και είναι αρκετές οι δημοσιεύσεις για τη χρήση OLAP τεχνολογίας στην ανάλυση επισκεψιμότητας.

Σε αυτή την ενότητα θα δοθούν κάποια παραδείγματα για την αξιολόγηση των σχημάτων που μπορεί να βρεθούν από το Data Mining και τα συμπεράσματα που μπορεί να εξαχθούν από αυτά. Θα περιγραφεί πιο αναλυτικά η λειτουργικότητα της τεχνολογίας OLAP και θα παρουσιαστούν επιλεγμένες εργασίες που συνδυάζουν OLAP με την ανάλυση επισκεψιμότητας σε ιστοχώρους.

### **2.3.7.1 Αξιολόγηση Παρατηρήσεων**

Όπως αναφέρθηκε, ο σκοπός για τον οποίο διεξάγεται η ανάλυση καθορίζει απόλυτα τη μεθοδολογία που θα ακολουθηθεί για την ανάλυση και τις επιμέρους μετρικές που θα χρησιμοποιηθούν. Αυτό που επίσης εξαρτάται από τη συγκεκριμένη εφαρμογή είναι η αξιολόγηση

των παρατηρήσεων που θα γίνουν. Ανάλογα με το αντικείμενο και το είδος τους ιστοχώρου οι επιθυμητές τιμές που μπορεί να πάρουν οι διάφορες μετρικές ποικίλει και μεταβάλλεται. Στο [9] για παράδειγμα διατυπώνεται ο κανόνας ότι η χρονική διάρκεια μιας επίσκεψης προσδιορίζει την προσοχή που δίνει ο επισκέπτης στον ιστοχώρο (*measures the amount of attention spent*). Σε έναν ιστοχώρο συχνών και μικρών ειδησεογραφικών άρθρων θα περιμέναμε όμως ότι ο χρόνος επίσκεψης είναι πολύ πιο μικρός απ' ό,τι σε έναν ιστοχώρο τουριστικού προορισμού. Θα περιμέναμε επίσης ότι η συχνότητα επίσκεψης στον ειδησεογραφικό χώρο είναι πολύ μεγαλύτερη απ' ό,τι στον τουριστικό και να καλύπτει πολύ μεγάλη χρονική περίοδο.

Το αργότερο όταν φτάσει κανείς στο σημείο της αξιολόγησης των δεδομένων θα πρέπει να έχει προσδιορίσει την έννοια της “επιτυχίας” ή της απόδοσης του ιστοχώρου που εξετάζει. Κριτήρια για την αξιολόγηση της επιτυχίας θα μπορούσαν να είναι η ταχύτητα απάντησης του εξυπηρετητή, η ευκολία πλοήγησης στον ιστοχώρο, ο χρόνος παραμονής των επισκεπτών ή κλασικές μέθοδοι αξιολόγησης του επιπέδου διεπικοινωνίας (*user interface*) με αναφορά σε επιλεγμένο σύνολο χρηστών. Αλλά τέτοιου είδους κριτήρια αν και έχουν αυτόνομη αξία δεν έχουν μεγάλη σημασία εάν δεν εκπληρώνεται ο στόχος του ιστοχώρου. Ένα ηλεκτρονικό κατάστημα στο οποίο οι χρήστες παραμένουν για πολλές ώρες και πλοηγούν με ευκολία και ταχύτητα σε αυτό αλλά δεν προχωρούν σε αγορές δεν εκπληρώνει τους στόχους για τους οποίους κατασκευάστηκε (ηλεκτρονική αγορά προϊόντων) και μάλλον πρέπει να θεωρηθεί αποτυχημένο.

Απαιτείται η μέτρηση της επιτυχίας ενός ιστοχώρου με βάση τα αντικειμενικά κριτήρια που τίθενται από τους δικαιούχους του. Σε αυτό το πλαίσιο χρησιμοποιούνται για παράδειγμα η έννοιες του *contact efficiency* και του *conversion efficiency*. Το πρώτο επιστρέφει το ποσοστό των επισκεπτών που υπερβαίνει ένα κάτω χρονικό όριο παραμονής στον ιστοχώρο. Το δεύτερο επιστρέφει το ποσοστό που μετά από κάποια πλοήγηση προβαίνει στην επιθυμητή ενέργεια (π.χ. αγορά προϊόντος). Η επιτυχία του ιστοχώρου ορίζεται με το ποσοστό μετατροπής απλών επισκεπτών σε πελάτες του ιστοχώρου, και αυτό μπορεί να μετρηθεί χωρίς την εμπλοκή των ίδιων των χρηστών.

Οι τελευταίες δύο έννοιες επεκτείνονται στο [61] με σκοπό όχι μόνο τη μέτρηση της επιτυχίας του ιστοχώρου αλλά και τη βελτίωσή του. Κατ' αρχήν τονίζεται ότι ένας ιστοχώρος μπορεί να έχει πολλαπλούς σκοπούς και κάθε φορά πρέπει να είναι ξεκάθαρο ποιος από αυτούς ενδιαφέρει στην ανάλυση. Ο σκοπός που ενδιαφέρει κάθε φορά αναφέρεται ως “στόχος του ιστοχώρου” (*the goal of the site*). Στην πράξη οι πολλαπλοί στόχοι εμφανίζονται ως περισσότερες αυτόνομες υπηρεσίες. Για παράδειγμα μπορεί να υπάρχει ένας ιστοχώρος που προσφέρει πληροφορίες για προϊόντα και

υπηρεσία αναζήτησης προϊόντων. Οι δύο στόχοι του ιστοχώρου είναι α)βοήθεια ανεύρεσης προϊόντων και β)μάρκετινγκ προϊόντων.

Για τον ακριβή προσδιορισμό της επίτευξης του στόχου ενός ιστοχώρου, οι σελίδες του χωρίζονται σε “*action pages*” και “*target pages*”. Η αίτηση για μία *action page* σημαίνει ότι ο επισκέπτης επιδιώκει την εκπλήρωση του στόχου, ενώ η επίσκεψη σε *target page* ότι τον έχει εκπληρώσει. Θεωρείται ότι δεν μπορεί να φτάσει σε *target page* χωρίς να περάσει από *action page*. Σε μία ηλεκτρονική βιβλιοθήκη η επίσκεψη στη σελίδα αναζήτησης βιβλίων μπορεί να θεωρηθεί *action page* και η πρόσβαση στο κείμενο του βιβλίου το *target page*. Στη συγκεκριμένη εργασία η έννοια των σελίδων ξεχωρίζεται από το ακριβές περιεχόμενό τους και επιχειρείται η μοντελοποίηση των υπηρεσιών που προσφέρονται. Για παράδειγμα η υπηρεσία αναζήτησης μπορεί να προσφέρεται από όλες τις σελίδες του ιστοχώρου. Η μοντελοποίηση γίνεται με τη δημιουργία ιεραρχικών εννοιών με βάση το [51] των ίδιων συγγραφέων.

Στη συνέχεια ορίζεται ως “*active session*” η επίσκεψη στην οποία υπάρχει τουλάχιστον μία ενέργεια προς την εκπλήρωση του στόχου, δηλαδή υπάρχει επίσκεψη σε *action page*. Όλες οι υπόλοιπες επισκέψεις είναι *inactive*.

Ως “*customer session*” ορίζεται η επίσκεψη όπου έχει εκπληρωθεί ο στόχος. Επειδή ο στόχος εκπληρώνεται μόνο δια μέσου *action pages* όλα τα *customer sessions* είναι επίσης *active sessions*.

Το *contact efficiency* μίας συγκεκριμένης *action page*  $A$  ( $contacteff(A)$ ) ορίζεται ως ο λόγος των επισκέψεων που περιέχουν αυτή τη σελίδα προς το συνολικό αριθμό των επισκέψεων (Πίνακας 2.16). Υπολογίζοντας αυτό το μέγεθος για κάθε *action page* μπορεί να μετρηθεί η συμβολή κάθε σελίδας στην προσέλκυση επισκεπτών και μπορεί να ανακαλυφθούν σελίδες με μικρή συμβολή.

Για τη μέτρηση της σημαντικότητας μιας *action page* σε σχέση με τα υπόλοιπα *action pages* ορίζεται και το *relative contact efficiency of an action page* -  $Rcontacteff(A)$  ως ο λόγος των επισκέψεων που περιέχουν τη σελίδα  $A$  προς το σύνολο των *active sessions* ( $aSessions$ ) (Πίνακας 2.16).

Με όμοιο τρόπο ορίζεται το *conversion efficiency* μιας οποιασδήποτε σελίδας προς μία *target page*. Σε αυτή την περίπτωση όμως πρέπει να ληφθεί υπ' όψιν και το μονοπάτι που διανύθηκε στην πορεία προς την *target page*. Έτσι ορίζεται το “*conversion efficiency of a page  $P$  towards a target page  $T$  over a group of paths  $G$  from  $P$  to  $T$  -  $conveff(P,T,G)$ ” ως το λόγο του συνόλου των στοιχείων του  $G$  προς το λόγο του συνόλου όλων των *active sessions* που περιέχουν την  $P$  (Πίνακας 2.16).*

<b>Ορισμοί contact και conversion efficiency</b>	
contact efficiency	$contacteff(A) = \frac{card(\{\{s \in Sessions \mid A \in s\})}{card(Sessions)}$
relative contact efficiency	$Rcontacteff(A) = \frac{card(\{\{s \in Sessions \mid A \in s\})}{card(aSessions)}$
conversion efficiency	$conveff(P, T, G) = \frac{card(G)}{card(\{\{s \in aSessions \mid P \in s\})}$

**card()** cardinality

**{...}** multiset

**A** action page

**T** target page

**aSessions** active sessions

**P** arbitrary page

**G** group of paths from P to T

Πίνακας 2.16 Ορισμός των contact efficiency, relative contact efficiency και conversion efficiency από το [61]

Το conveff() προσδιορίζει την επιτυχία μίας σελίδας να βοηθήσει ή να οδηγήσει τον επισκέπτη στην εκπλήρωση του στόχου του ιστοχώρου. Τα μονοπάτια που μπορεί να ληφθούν υπ' όψιν βεβαίως μπορεί να είναι πάρα πολλά επειδή υπάρχει μεγάλη διασυνδεσιμότητα στις σελίδες ενός ιστοχώρου και έτσι είναι αντικείμενο data mining ο περιορισμός τους.

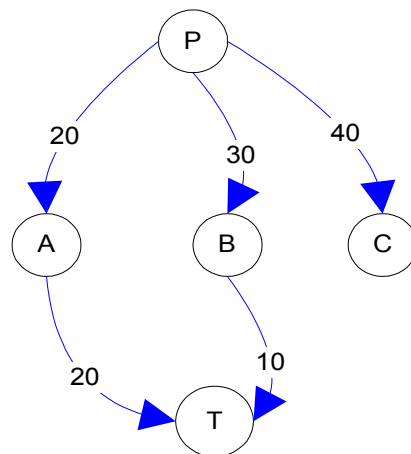
Επειδή όλα τα μονοπάτια που ανήκουν στο G περιέχουν και την P και την T θα περίμενε κανείς ο ονομαστής να είναι πολύ κοντά στον παρανομαστή. Στο Σχήμα 2.14 φαίνεται ένα παράδειγμα μονοπατιών. Ενδιαφέρει ο υπολογισμός του conveff(P,T,\*). Συνολικά υπάρχουν 100 επισκέψεις που περιέχουν την P. Από αυτές στις 20 επισκέπτεται την A και ακολούθως την T. Στις 30 η B και από αυτές τις 30 οι 10 ακολουθούν την T. Στις υπόλοιπες 40 ακολουθείται η C αλλά δεν πηγαίνουν στην T. Οπότε

$$conveff(P, T, *) = \frac{20+10}{100} = 0,3 \text{ το οποίο είναι μάλλον μικρό. Το ερώτημα που τίθεται είναι}$$

γιατί οι επισκέπτες που περνάνε από τη C δε συνεχίζουν προς την T.

Φάνηκε λοιπόν ένας τρόπος για την αξιολόγηση της επιτυχίας ενός ιστοχώρου βασιζόμενος στην συμβολή της κάθε σελίδας στην επίτευξη των στόχων του ιστοχώρου. Η εύρεση των μονοπατιών στα οποία θα εφαρμοσθεί η μέθοδος ακολουθεί τη μεθοδολογία που περιγράφηκε σε προηγούμενη ενότητα για την εύρεση των g-sequences και περιγράφεται πιο αναλυτικά στο [61] χρησιμοποιώντας όμοιες έννοιες με τις προηγούμενες. Στην ίδια εργασία προτείνεται και ένας μη αυτόματος αλγόριθμος για την αξιολόγηση και βελτίωση του ιστοχώρου βασιζόμενος στα μεγέθη που παρουσιάστηκαν. Ο αλγόριθμος φαίνεται στον πίνακα 2.17





Σχήμα 2.14 Παράδειγμα μονοπατιών για τον υπολογισμό του  $conveff(P, T, *)$  .

4. For each such pattern:

(a) Detection of pages that are themselves frequent but rarely lead to the target page

(b) If there are no such pages,

then the start page must be improved

else the detected pages must be improved

Πίνακας 2.17 Ευριστικός αλγόριθμος για την εύρεση σελίδων που χρήζουν βελτίωσης (*EVAL\_conversion\_efficiency* [61])

Το τελευταίο βήμα του αλγόριθμου ουσιαστικά εντοπίζουν εκείνες τις σελίδες των active sessions (είτε αρχικές είτε ενδιάμεσες) που παρουσιάζουν πρόβλημα επειδή δεν οδηγούν σε target page.

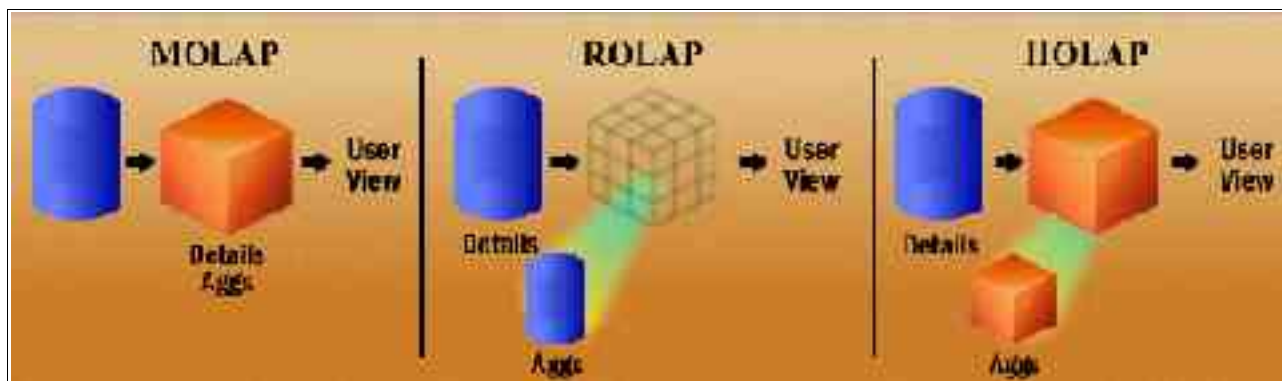
Το [61] κλίνει με μία ανάλυση ενός πραγματικού ιστοχώρου (SchulWeb) όπου μετά την αρχική ανάλυση και τη βελτίωση του ιστοχώρου το αποτέλεσμα είναι μία μετρήσιμη βελτίωση σύμφωνα με τους στόχους που είχαν τεθεί. Η αναλυτική περιγραφή της ανάλυσης είναι εκτός του σκοπού της παρούσας διατριβής και ο ενδιαφερόμενος αναγνώστης παραπέμπεται στη συγκεκριμένη εργασία.

### 2.3.7.2 OLAP & Data Mining

Το OLAP εμφανίστηκε το 1993 σε white paper [60] και περιείχε 12 κανόνες για το χαρακτηρισμό του OLAP. Αργότερα προστέθηκαν επιπλέον 6 αν και πλέον χρησιμοποιούνται συχνά οι 5 FASMI (Fast Analysis of Shared Multidimensional Information) κανόνες [62] για τον προσδιορισμό OLAP υπηρεσιών.

Στην πράξη OLAP τεχνικές προσφέρουν τη δυνατότητα πολυδιάστατης ανάλυσης δεδομένων και χρησιμοποιούνται κατά κόρον στα πλαίσια συστημάτων υποστήριξης αποφάσεων. Η έμφαση είναι

στις πολλές διαστάσεις και στην υποστήριξη ιεραρχιών και στην ταχύτητα απαντήσεων από το σύστημα.



Σχήμα 2.15 Οργάνωση δεδομένων OLAP από [63]

Η ανάλυση γίνεται πάνω σε αριθμητικές τιμές (measures) που προέρχονται από τα κυρίως δεδομένα (fact table) και υπολογίζονται για κάθε συνδυασμό διαστάσεων (dimensions). Τα δεδομένα (και οι μετρήσεις) αποθηκεύονται είτε σε εξειδικευμένες δομές (MOLAP – Multidimensional OLAP) είτε σε σχεσιακές βάσεις δεδομένων (ROLAP – Relational OLAP), ή ακόμα σε υβριδικές δομές που συνδυάζουν τους δύο τρόπους (HOLAP – Hybrid OLAP) (Σχήμα 2.15).

Η δομή που προκύπτει αναφέρεται ως κύβος OLAP (OLAP Cube) και πάνω σε αυτήν μπορεί να εκτελεστούν πράξεις που περιορίζουν ή επεκτείνουν κάθε φορά την άποψη στα δεδομένα (Πίνακας 2.18).

<i>Πράξεις OLAP</i>	
Drill-down	Η εξειδίκευση πάνω σε μία ιεραρχία. Για παράδειγμα από την άποψη των δεδομένων ανά μήνα πηγαίνουμε στην άποψη ανά ημέρα. Εξειδικεύουμε δηλαδή στην ιεραρχία του χρόνου.
Roll-up	Το αντίθετο από το drill-down
Slice	Περιορίζεται η άποψη σε υποκύβο με την επιλογή ενός χαρακτηριστικού σε μία διάσταση. Για παράδειγμα, επιλέγοντας .edu από το domain των επισκεπτών.
Dice	Περισσότερα του ενός slice σε περισσότερες διαστάσεις.

Πίνακας 2.18 Πράξεις OLAP

Στο SQL-99 επεκτείνεται η γλώσσα SQL για OLAP ερωτήσεις και αναμένεται ότι με την υποστήριξη του προτύπου από τους κατασκευαστές σχεσιακών βάσεων δεδομένων θα γίνει περισσότερο προσιτή η χρήση OLAP.

Στα πλαίσια της ανάλυσης επισκεψιμότητας οι τεχνικές OLAP από μόνες τους παρέχουν ένα πολύ εύχρηστο περιβάλλον, ειδικά για άτομα χωρίς τεχνικές γνώσεις, για την εύκολη και γρήγορη ανάλυση των δεδομένων. Δημιουργώντας διαστάσεις και ιεραρχίες για τα διάφορα χαρακτηριστικά από την κίνηση των ιστοχώρων δίνεται στους αναλυτές η δυνατότητα να εξερευνήσουν τη μεταβολή στατιστικών μεγεθών.

Στο [25] δημιουργείται έναν κύβος OLAP με διαστάσεις του πίνακα 2.19. Το μέγεθος που μπορεί να υπολογιστεί είναι για παράδειγμα ο αριθμός των σελιδοπροβολών ή των επισκεπτών.

<i>Διαστάσεις OLAP για ανάλυση επισκεψιμότητας</i>	
URL	Η σελίδα για την οποία έγινε αίτηση, σε ιεραρχία σύμφωνα με τη δομή καταλόγων του ιστοχώρου
Type of resource	Είδος αρχείου
Size	Μέγεθος αρχείου σε κατηγορίες μεγεθών
Request time	Η ώρα της αίτησης σε ιεραρχία χρόνου
Domain name	Περιοχή ονόματος επισκέπτη, σε ιεραρχία περιοχών
Agent	Ιεραρχία γνωστών προγραμμάτων πλοήγησης
User	Ιεραρχία χρηστών

*Πίνακας 2.19 Διαστάσεις OLAP του [25] για την ανάλυση επισκεψιμότητας*

Προσπάθεια για συνδυασμό OLAP και Data Mining γίνεται στο [64], [25] και [12].

## **2.4 Ιστοχώροι Τουριστικού Προορισμού**

Από τη στιγμή που ο δυνητικός ταξιδιώτης έχει επιλέξει τον προορισμό του, έρχεται αντιμέτωπος με ένα σύνολο πολύπλοκων προβλημάτων που πρέπει να επιλύσει. Τα προβλήματα αυτά συνοψίζονται στον τρόπο μετάβασης στον προορισμό, διαμονή, διατροφή, εξερεύνηση της περιοχής και γενικά η διασκέδαση. Βέβαια η βαρύτητα που δίνει ο καθένας σε κάθε μία από τις προηγούμενες παραμέτρους ποικίλλει ενώ η συγκεκριμένη λύση που θα επιλεγεί εξαρτάται από ζητήματα όπως διαθεσιμότητα ή από οικονομικά ζητήματα.

Η πολυπλοκότητα αυτή έχει οδηγήσει στην ανάγκη ύπαρξης ειδικών στο θέμα προγραμματισμού του ταξιδιού, που είναι τα γραφεία τουρισμού. Συνήθως προσφέρουν ολοκληρωμένες λύσεις υπό μορφή “πακέτων” αλλά σε αυτό ακριβώς το σημείο εμφανίζεται η δυσαρέσκεια των ταξιδιωτών, διότι στην προσπάθεια των γραφείων να εξυπηρετήσουν μία σχετικά μεγάλη ομάδα ανθρώπων το αποτέλεσμα καταλήγει να είναι αυτό που ονομάζεται “μαζικός τουρισμός” χωρίς να λαμβάνονται υπόψιν οι ιδιαιτερότητες του κάθε ταξιδιώτη.

Το Διαδίκτυο προσφέρει σε αυτό το σημείο μοναδικά πλεονεκτήματα σε σχέση με τα παραδοσιακά ταξιδιωτικά γραφεία ή ακόμα και άλλου είδους πηγών πληροφόρησης όπως βιβλία ή ψηφιακά offline μέσα. Η πληροφορία είναι εύκολο να ενημερώνεται με αποτέλεσμα να είναι πάντα επίκαιρη ενώ μπορεί να προσφέρει υπηρεσίες “στο σπίτι” του χρήστη όπως:

- άμεσος έλεγχος διαθεσιμότητας πτήσεων και διαμονής
- πληροφόρηση για καιρικές συνθήκες και για δρομολόγια μεταφορών
- ενήμερες συναλλαγματικές ισοτιμίες
- κράτηση εισιτηρίων για μεταφορές ή κράτηση δωματίων

Δεν ξαφνιάζει λοιπόν ότι στο Διαδίκτυο ο τουρισμός είναι ένας από τους κυρίαρχους τομείς ηλεκτρονικού επιχειρήν και κατέχει από τα μεγαλύτερα ποσοστά από άποψη όγκου συναλλαγών στο Διαδίκτυο.

Ως κατηγορία εφαρμογών στο Διαδίκτυο, οι ιστοχώροι τουριστικού προορισμού έχουν πολύ συγκεκριμένα και κοινά χαρακτηριστικά που συναντώνται σε όλους τους σχετικούς ιστοχώρους. Σε αυτά περιλαμβάνονται ιεραρχίες για τις περιοχές ενδιαφέροντος, κατηγοριοποιήσεις των επιχειρήσεων τουρισμού, γενική πληροφόρηση για τις περιοχές, μηχανισμοί αναζήτησης κ.α.

Παραδείγματα ιστοχώρων με αυτά τα χαρακτηριστικά είναι τα

- <http://www.austria-tourism.at/> που καλύπτει την Αυστρία ως περιοχή
- <http://www.visit-sweden.com> που καλύπτει τη Σουηδία ως περιοχή
- <http://www.tourspain.es/> που καλύπτει την Ισπανία ως περιοχή
- <http://www.reiseland-brandenburg.de/> που καλύπτει το κρατίδιο του Βραδενμβούργου της Γερμανίας
- <http://www.hamburg-tourism.de/> που καλύπτει την πόλη (και κρατίδιο της Γερμανίας) του Αμβούργου
- <http://www.travelscotland.co.uk/> που καλύπτει τη Σκωτία ως περιοχή
- <http://www.visitsouthwest.co.uk/> που καλύπτει τη νοτιοδυτική Αγγλία ως περιοχή
- <http://agn.hol.gr> που καλύπτει συνολικά την Ελλάδα

- <http://www.crete.tournet.gr/> που καλύπτει την Κρήτη ως περιοχή

Στο Παράρτημα Α περιγράφεται πιο αναλυτικά η κατηγορία αυτών των εφαρμογών και δίνεται μία περιγραφή για αντιπροσωπευτικούς ιστοχώρους. Περιγράφεται επίσης αναλυτικά ο ιστοχώρος τουριστικού προορισμού στον οποίο εφαρμόστηκαν οι μεθοδολογίες και τα πειράματα σε αυτή την εργασία.

## 2.5 Συνεισφορά

Η ανάλυση χρησιμοποίησιμότητας (usability) είναι σήμερα μία αποδεδειγμένα χρήσιμη μεθοδολογία η οποία ολοένα και καθιερώνεται ως αναπόσπαστο κομμάτι του σχεδιασμού και τη ανάπτυξης εφαρμογών λογισμικού, και χρησιμοποιείται παράλληλα με όλες τις άλλες μεθοδολογίες σχεδιασμού και ανάπτυξης όπως Use Cases, UML Diagrams, Rapid Prototyping κ.α.. Η ανάλυση χρήσης στον παγκόσμιο ιστό (Web Usage Analysis, Web Mining κλπ.) είναι μία σχετικά πρόσφατη περιοχή έρευνας με μεγάλο ερευνητικό και βιομηχανικό ενδιαφέρον. Και οι δύο μεθοδολογίες μπορούν να χρησιμοποιηθούν για τη βελτίωση της κατασκευής των εφαρμογών στο Διαδίκτυο με στόχο την προσέλκυση περισσότερων χρηστών στους ιστοχώρους, που παραμένουν περισσότερο χρόνο και γενικά χρησιμοποιούν τον ιστοχώρο με τρόπο αποδοτικό για την επιχείρηση ή τον οργανισμό που τον έχει αναπτύξει (σύμφωνα με κάποια σενάρια χρήσης που είχαν προδιαγραφεί ή που εμφανίστηκαν και έγιναν σημαντικοί μετά την κατασκευή).

Ο στόχος της διατριβής είναι να παρουσιάσει μια συστηματική μεθοδολογία καταγραφής χρήσης ιστοχώρων με σκοπό την ανάλυση της χρήσης. Μία τέτοια μελέτη παρουσιάζει μεγάλες δυσκολίες στη γενική της μορφή αν λάβει κανείς υπόψη το εύρος εφαρμογών στον παγκόσμιο ιστό, των σεναρίων χρήσης και των επιδιώξεων των επιχειρήσεων.

Είναι επίσης δύσκολο να τροποποιηθούν γενικής φύσεως σενάρια και λογισμικό για να αντικατοπτρίζουν τις ανάγκες μιας συγκεκριμένης εφαρμογής, καθώς και να εξηγηθεί η χρήση γενικής φύσεως μεθόδων και εργαλείων σε περιορισμένου μεγέθους επιχειρήσεις και οργανισμούς που δεν έχουν πολύ καταρτισμένο προσωπικό σε νέες τεχνολογίες. Έτσι προτείνουμε στα πλαίσια της διατριβής λογισμικό και μεθοδολογίες ανάλυσης επισκεψιμότητας ιστοχώρων για μια συγκεκριμένη κατηγορία εφαρμογών που όμως χρησιμοποιείτε και αναπτύσσεται από μεγάλο πλήθος επιχειρήσεων, όπως η κατηγορία εφαρμογών για ιστοχώρους τουριστικού προορισμού. Για να συγκεκριμενοποιήσουμε τη χρήση της μεθοδολογίας και του λογισμικού που προτείνουμε, τα εφαρμόζουμε σε μία συγκεκριμένη αλλά πολύ πλούσια σε δυνατότητες εφαρμογή τουρισμού στον παγκόσμιο ιστό, και δείχνουμε τα τα συμπεράσματα που μπορεί εξαχθούν, τα πλεονεκτήματα της μεθοδολογία και τρόπους εφαρμογής σε οποιοδήποτε ιστοχώρο με όμοια χαρακτηριστικά.

Για αυτή την κατηγορία εφαρμογών αναλύουμε τις βασικές λειτουργικότητες που παρέχονται και τα βασικά σενάρια χρήσης. Παρουσιάζουμε μια συστηματική μεθοδολογία καταγραφής και ανάλυσης χρησιμοποιησιμότητας ιστοχώρων βασισμένη σε μία βάση δεδομένων χρήσης ιστοχώρων και SQL ερωτήσεις πάνω σε αυτή. Η μεθοδολογία αυτή με τη χρήση βάσεων δεδομένων, επιτρέπει εκτός των άλλων και παρακολούθηση των επισκέψεων στον ιστοχώρο σε πραγματικό χρόνο για αντιμετώπιση έκτακτων καταστάσεων ή για παροχή υπηρεσιών σε τρίτους. Το λογισμικό που αναπτύξαμε για την καταγραφή και ανάλυση επισκεψιμότητας είναι διαθέσιμο σε μορφή ελεύθερου λογισμικού (free software).

Χρησιμοποιούμε τη μεθοδολογία καταγραφής και ανάλυσης χρήσης και παρουσιάζουμε ένα εύρος συμπερασμάτων που μπορούν να εξαχθούν με τη χρήση της σε έναν ιστοχώρο τουριστικού προορισμού. Συγκρίνουμε αυτή τη μεθοδολογία με πιο καθιερωμένες μεθοδολογίες ανάλυσης χρησιμοποιησιμότητας. Τα συμπεράσματα που προκύπτουν υποδεικνύουν ότι οι καθιερωμένες μεθοδολογίες είναι χρήσιμες για τον εντοπισμό προβλημάτων χρησιμοποιησιμότητας για προκαθορισμένα σενάρια χρήσης. Δείχνουμε όμως ότι προβλήματα χρησιμοποιησιμότητας μπορούν να εντοπιστούν και με τη μεθοδολογία ανάλυσης χρήσης ιστοχώρων, η οποία μάλιστα είναι ικανή να εφαρμοστεί για τον εντοπισμό σεναρίων χρήσης που δεν προβλέφθηκαν κατά την ανάπτυξη των ιστοχώρων. Τέλος δείχνουμε ότι η ανάλυση των δεδομένων χρήσης οδηγεί στην κατάταξη των χρηστών του ιστοχώρου σε κατηγορίες προσφέροντας παραπέρα δυνατότητες προσωποποίησης (personalization) της διεπαφής (user interface) και του περιεχομένου, ιδιαίτερα σε πολυγλωσσικά περιβάλλοντα. Οι δυνατότητες που προσφέρονται από ένα τέτοιο σύστημα είναι χρήσιμες όχι μόνο για τη βελτίωση των ιστοχώρων αλλά επίσης για οργανισμούς τοπικής αυτοδιοίκησης διαφόρων βαθμίδων (περιφέρεια, νομός κλπ.) και για οργανισμούς ή μεμονωμένες επιχειρήσεις τουρισμού για την ανάλυση των ενδιαφερόντων και προτιμήσεων των επισκεπτών του τουριστικού προορισμού.

Δείχνουμε μία μεγάλη σειρά από πειράματα που έγιναν με τη χρήση της μεθοδολογίας σε έναν ιστοχώρο τουριστικού προορισμού και συζητούμε τις παρατηρήσεις και τα συμπεράσματα που μπορούν να εξαχθούν. Η παρουσίαση των πειραμάτων και συμπερασμάτων δείχνουν τα συγκεκριμένα οφέλη που μπορούν να προκύψουν για επιχειρήσεις με ιστοχώρους τουρισμού με τη χρήση της μεθοδολογίας. Έτσι τις βοηθά προκειμένου να αποφασίσουν για μία ενδεχόμενη επένδυση στην τεχνολογία αυτή.

Για όλες αυτές τις χρήσεις των δεδομένων που συλλέγονται από τη χρήση του ιστοχώρου παρουσιάζουμε τυπικές SQL ερωτήσεις πάνω σε σχεσιακό σχήμα της Βάσης Χρήσης και

στατιστικές και αναλύσεις πιθανών συμπερασμάτων για την περίπτωση ιστοχώρων τουρισμού.

## **2.6 Ανακεφαλαίωση**

Σε αυτό το κεφάλαιο έγινε μία εισαγωγή στο γνωστικό αντικείμενο της αλληλεπίδρασης του ανθρώπου με υπολογιστές και περιγράφηκε η διάσταση της χρησιμοποιησιμότητας ως μέτρο της ανταπόκρισης των εφαρμογών στις ανάγκες των χρηστών.

Επισημάνθηκαν τα νέα δεδομένα που προκύπτουν από την ολοένα και μεγαλύτερη χρήση εφαρμογών που χρησιμοποιούν τεχνολογίες Διαδικτύου και ειδικότερα εφαρμογές πλοήγησης με βάση το HTTP πρωτόκολλο. Ιδιαίτερα τονίστηκε η διάσταση της ευκολίας με την οποία ενσωματώνονται αλλαγές σε εφαρμογές στο Διαδίκτυο, το οποίο δημιουργεί ανάγκες για ταχύτερη και αμεσότερη διερεύνηση της χρησιμοποιησιμότητας.

Έγινε μία σύντομη αναφορά σε καθιερωμένες πρακτικές και μεθόδους για τη βελτίωση της χρησιμοποιησιμότητας και δόθηκε μία αναλυτική περιγραφή της σχετικής έρευνας σε θέματα ανάλυσης των δεδομένων που προκύπτουν από τη χρήση των ιστοχώρων με βάση το HTTP πρωτόκολλο.

Τέλος περιγράφηκαν βασικά χαρακτηριστικά ιστοχώρων τουρισμού στους οποίους αναφέρεται η παρούσα εργασία και αναφέρθηκε εν συντομία η ερευνητική της συνεισφορά της διατριβής.

### **3 Μεθοδολογία και Σύστημα Βάσης Δεδομένων για την Καταγραφή Δεδομένων Χρήσης Ιστοχώρων**

Στο προηγούμενο κεφάλαιο αναπτύχθηκε η σχετική έρευνα στο χώρο της ανάλυσης επισκεψιμότητας ιστοχώρων (Web Usage Analysis). Η ανάλυση βασίζεται στα δεδομένα που καταγράφονται για τις κινήσεις των επισκεπτών στους ιστοχώρους και καθιερωμένη πρακτική είναι η καταγραφή να βασίζεται στα δεδομένα από τα αρχεία καταγραφής των εξυπηρετητών παγκόσμιου ιστού.

Σε αντίθεση με τη γενικότερη καταγραφή χρήσης εφαρμογών, οι εφαρμογές ιστοχώρου χαρακτηρίζονται από τη χρήση του HTTP πρωτοκόλλου ως μέσου επικοινωνίας, και ανήκουν όλες σε μία οικογένεια εφαρμογών με κάποιες βασικές κοινές μετρικές που πρέπει να ληφθούν υπόψιν.

Σε αυτό το κεφάλαιο θα περιγραφεί το σύστημα βάσης δεδομένων που αναπτύχθηκε στα πλαίσια της εργασίας με σκοπό την ανάλυση επισκεψιμότητας ιστοχώρων τουριστικού προορισμού. Θα περιγραφούν οι λόγοι που οδήγησαν στην ανάπτυξη ενός νέου συστήματος και θα δοθεί αναλυτική περιγραφή της αρχιτεκτονικής του.

Θα ακολουθήσει η παρουσίαση μίας μεθοδολογίας που αναπτύχθηκε με σκοπό τη συστηματική επιλογή των στοιχείων που κρίνονται κάθε φορά αναγκαία να καταγράφονται και τα οποία οδηγούν στη δημιουργία της δομημένης πληροφορίας στο σύστημα, η οποία αναλύεται.

Το κεφάλαιο θα κλείσει με μία ανακεφαλαίωση

#### **3.1 Σύστημα Βάσης Δεδομένων για την Καταγραφή Δεδομένων Χρήσης Ιστοχώρων**

Για τη συλλογή των δεδομένων χρήσης ιστοχώρων και την ανάλυση τους υπήρχαν διάφορες επιλογές. Η χρήση κάποιων εκ των ελεύθερων εφαρμογών καταγραφής χρήσης (για παράδειγμα AWStats ή Webalizer [16][48]) αποκλείστηκε λόγω των πολύ περιορισμένων δυνατοτήτων και της ελλιπής ή απύσας παραμετροποίησης που προσφέρουν. Εξειδικευμένες εφαρμογές που προσφέρονται από εξειδικευμένες επιχειρήσεις (π.χ. NetGenesis [36]) αποκλείστηκαν επίσης λόγω του πολύ μεγάλου κόστους (της τάξης των 100.000 \$ ανά CPU) ενώ εφαρμογές από ερευνητικές εργασίες (για παράδειγμα WUM [13]) αν και πολύ χρήσιμες, είναι πολύ εξειδικευμένες για τη γενικότερη ανάλυση που επιχειρείται στην παρούσα εργασία. Εξάλλου, όπως είναι παραδεκτό στο χώρο του Web Usage Mining, η συλλογή και ανάλυση των δεδομένων χρήσης ιστοχώρων είναι μία μη τετριμμένη εργασία και σε καμία περίπτωση δεν υπάρχουν “έτοιμες” λύσεις που μπορεί να



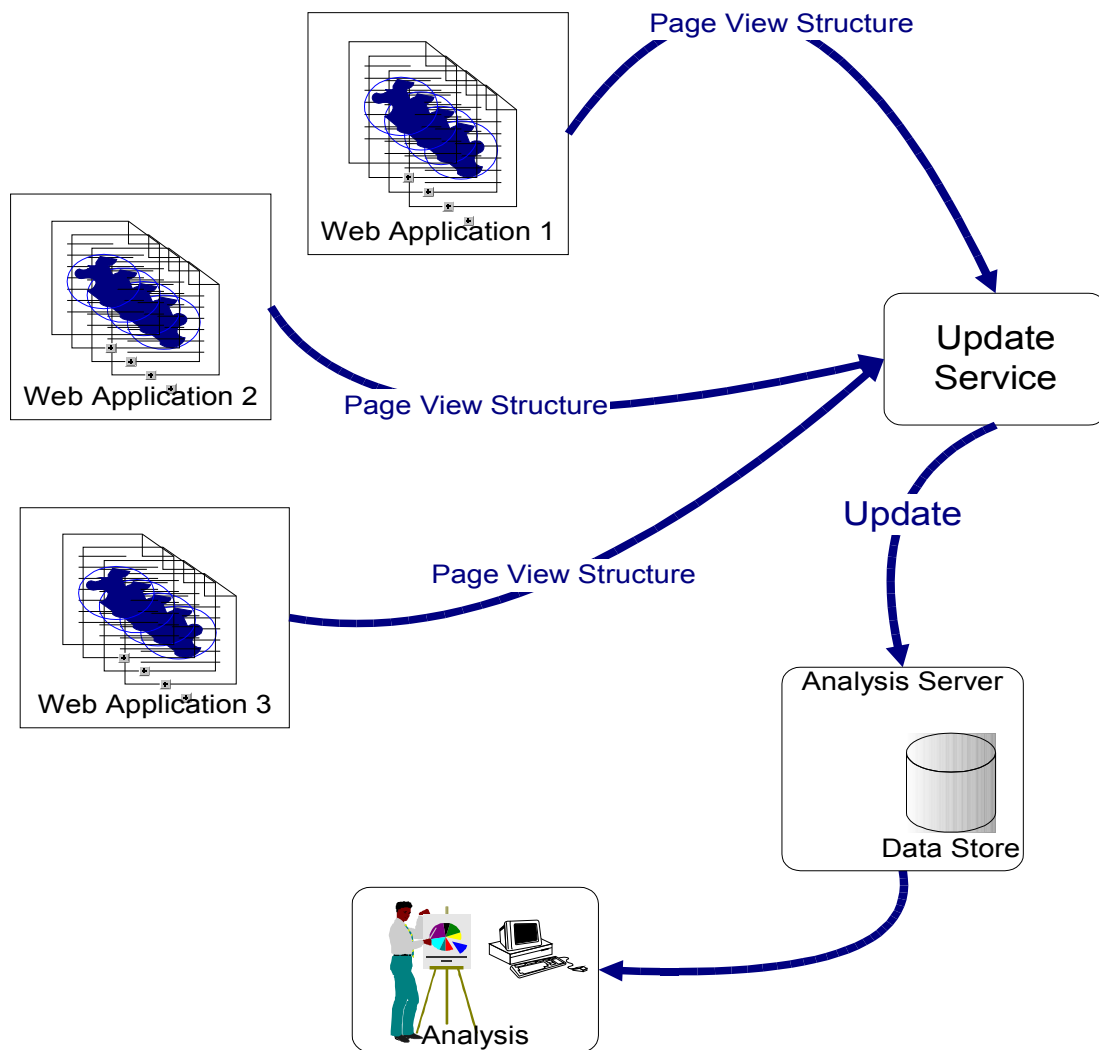
εφαρμοστούν γενικά.

Ειδικά για σύγχρονους ιστοχώρους τουριστικού προορισμού υπάρχει μία μεγάλη πρόκληση για το πως θα αξιοποιηθούν οι ιεραρχίες που συνήθως υπάρχουν στα δεδομένα που παρουσιάζονται στον ιστοχώρο (περιοχές, τοποθεσίες, αξιοθέατα, κατηγορίες) και είναι ήδη αποθηκευμένες με βάση ένα εξειδικευμένο σχήμα. Το σύστημα που αναπτύχθηκε εισάγει ένα επιπλέον επίπεδο περιγραφής της πληροφορίας που εμφανίζεται στις οθόνες των χρηστών, που αναφέρεται στο περιεχόμενο ανά σελιδοπροβολή, το οποίο προσφέρει αφενός επιπλέον δυνατότητες ανάλυσης όταν συνδυάζονται στην ίδια σελίδα πληροφορίες για περισσότερα και δυναμικά αντικείμενα, και αφετέρου επιτρέπει τη διασύνδεση με υπάρχουσες πηγές δεδομένων. Από αυτή την άποψη το σύστημα είναι αρκετά εύελκτο για να εφαρμοστεί και σε άλλους ιστοχώρους πέραν του ιστοχώρου για τον οποίο έγινε η ανάλυση στην παρούσα εργασία. Στην πραγματικότητα το ίδιο το σύστημα διαχειρίζεται ταυτόχρονα δεδομένα χρήσης από πολλές εφαρμογές.

Επιπλέον μεγάλο ενδιαφέρον παρουσίασε το γεγονός των πολλαπλών γλωσσών που υποστηρίζουν πολλοί ιστοχώροι και ειδικά αυτοί που απευθύνονται σε διεθνή κοινό. Η πολυγλωσσικότητα εισάγει ένα επιπλέον επίπεδο δυσκολίας. Εάν χρησιμοποιούταν ένα από τα υπάρχοντα συστήματα, η ίδια σελίδα παρουσίασης μιας περιοχής σε διαφορετικές γλώσσες θα φαινόταν ως διαφορετικές σελίδες. Ανάλογα με τον αριθμό των γλωσσών στον οποίο παρουσιάζεται η περιοχή, θα υπήρχαν ισάριθμες διευθύνσεις URL. Στην πραγματικότητα όμως πρόκειται για το ίδιο περιεχόμενο.

### **3.1.1 Βασικές Αρχές**

Σε αντίθεση με καθιερωμένες πρακτικές συλλογής δεδομένων χρήσης ιστοχώρων, στο σύστημα καταγραφής επισκεψιμότητας που αναπτύχθηκε τα δεδομένα δεν προέρχονται από τα αρχεία καταγραφής των εξυπηρετητών, αλλά από ειδικές δομές δεδομένων που στέλνονται με κάθε σελιδοπροβολή στο υποσύστημα καταγραφής. Οι πληροφορίες που περιέχονται σε αυτά καθορίζονται τη στιγμή της ανάπτυξης ή συντήρησης του ιστοχώρου, και αποθηκεύονται αμέσως. Αυτό έχει ως αποτέλεσμα την απάλειψη του σταδίου της προεπεξεργασίας δεδομένων που είναι ιδιαίτερα χρονοβόρο και δεν επιτρέπει έτσι την επεξεργασία των δεδομένων σε πραγματικό ή σχεδόν πραγματικό χρόνο (βλέπε και ενότητα σχετικής έρευνας στο κεφάλαιο 2). Έτσι είναι δυνατή η υποβολή ερώτησης στη βάση δεδομένων για πληροφορίες σε πραγματικό χρόνο (για παράδειγμα “πόσες ενεργές επισκέψεις υπάρχουν αυτή τη στιγμή”) ενώ δίνεται επίσης η δυνατότητα ανάδρασης προς την εφαρμογή για συγκεκριμένες παράγωγες πληροφορίες που εξάγονται από το σύστημα καταγραφής και προσφέρονται εν είδος υπηρεσίας προς τρίτους (για παράδειγμα “ποια ήταν η φράση αναζήτησης του χρήστη που μόλις επισκέφτηκε τον ιστοχώρο από το Google”).



Σχήμα 3.1 Η φιλοσοφία του συστήματος καταγραφής δεδομένων χρήσης. Διακρίνονται οι δομές σελιδοπροβολής που στέλνονται με κάθε σελιδοπροβολή από διαφορετικές εφαρμογές ιστοχώρου στην υπηρεσία ενημέρωσης. Η υπηρεσία ενημέρωσης ενημερώνει με τη σειρά της το κεντρικό αποθηκευτικό σύστημα πάνω στο οποίο διενεργείται η ανάλυση

Στο Σχήμα 3.1 φαίνεται μία γενική εικόνα του συστήματος. Διακρίνονται διαφορετικές εφαρμογές ιστοχώρου. Με κάθε σελιδοπροβολή σε κάθε μία από τις εφαρμογές δημιουργείται μία “Δομή Σελιδοπροβολής” η οποία περιέχει διάφορες πληροφορίες για τη σελίδα που εμφανίστηκε, καθώς και στοιχεία για τον χρήστη που τις ζήτησε.

Οι δομές σελιδοπροβολής στέλνονται στην “Υπηρεσία Ενημέρωσης” η οποία είναι υπεύθυνη για την επεξεργασία και αποθήκευση των δεδομένων.

Η αποθήκευση γίνεται σε σχεσιακή βάση δεδομένων στην οποία μπορεί να υποβληθούν ερωτήσεις για την ανάλυση των δεδομένων χρήσης.

### 3.1.2 Δομές Σελιδοπροβολής

Η σελιδοπροβολή είναι το βασικό δομικό στοιχείο με το οποίο οικοδομείται η εικόνα της

συμπεριφοράς των χρηστών στον ιστοχώρο. Οι αιτήσεις για σελίδες του ιστοχώρου που γίνονται από τους χρήστες μεταφράζονται σε ένα σύνολο υψηλού επιπέδου πληροφοριών τα οποία συναρμολογούνται σε Δομές Σελιδοπροβολής και στέλνονται στην υπηρεσία ενημέρωσης.

Κάθε δομή περιέχει τόσο στοιχεία από την επικοινωνία του HTTP πρωτοκόλλου όσο και δεδομένα σχετικά με τον ιστοχώρο που εξετάζεται και το συγκεκριμένο περιεχόμενο της σελίδας που ενεργοποίησε τη δημιουργία της δομής. Τα στοιχεία που συνθέτουν τη δομή φαίνονται στον Πίνακα 3.1.

Το ApplicationId αναφέρεται στην εφαρμογή ιστοχώρου όπου υπήρξε η σελιδοπροβολή. Το Client Ip αναφέρεται στη διεύθυνση IP του πελάτη σε όρους HTTP πρωτοκόλλου. Αυτή η διεύθυνση μπορεί να είναι διαφορετική από τη διεύθυνση του τελικού χρήστη σε περίπτωση που υπάρχουν ενδιάμεσοι στην αλυσίδα HTTP (proxies κλπ, βλέπε Σχήμα 2.11). Το Hostname αναφέρεται στο όνομα του μηχανήματος εφόσον αυτό είναι διαθέσιμο και δίνει μία ένδειξη για τη χώρα προέλευσης του χρήστη καθώς και για τον οργανισμό από τον οποίο προέρχεται.

<i>Δεδομένα του Πακέτου Σελιδοπροβολής</i>
Application Id
Client IP Address
Hostname
Referring URL
User Agent
Session Id
Date Time
Page Id
Visitor Id
Preferred Language
Page “Actions”
Post Impression Id
Forwarders (proxy)
Viewed Language

*Πίνακας 3.1 Περιεχόμενα του πακέτου σελιδοπροβολής*

Το Referring URL είναι η διεύθυνση από την οποία προήλθε η αίτηση για τη σελίδα, από την οποία μεταξύ άλλων μπορεί να εξαχθεί οι φράση αναζήτησης εάν προέρχεται από μηχανή αναζήτησης. Το User Agent είναι η εφαρμογή πλοήγησης που χρησιμοποιεί ο πελάτης. Το Session Id αναφέρεται στο cookie που έχει τεθεί για την τρέχουσα σύνοδο. Μεταδίδεται επίσης η ακριβής μέρα και ώρα

της αίτησης (Date Time) ενώ το Page Id είναι το όνομα της σελίδας που εμφανίστηκε. Το visitor id είναι το cookie που δε χάνεται στις διάφορες επισκέψεις από την ίδια εφαρμογή πλοήγησης. Το Preferred Language είναι η προτιμώμενη γλώσσα σύμφωνα με την εφαρμογή πλοήγησης ενώ το Viewed Language η γλώσσα στην οποία εμφανίζεται η σελίδα. Στο πεδίο forwarders αποθηκεύονται στοιχεία από τους ενδιάμεσους σταθμούς της αλυσίδας HTTP εφόσον υπάρχουν. Στο πεδίο page actions αποθηκεύονται τα περιεχόμενα της συγκεκριμένης σελίδας που σε συνδυασμό με το Page Id προσδιορίζουν το τι ακριβώς εμφανίζεται στο χρήστη. Ειδικό πεδίο είναι το Post Impression Id το οποίο αναφέρεται σε ενέργειες που ακολούθησαν μία σελιδοπροβολή αλλά συνδέονται με αυτή. Τέτοιες ενέργειες μπορεί να είναι για παράδειγμα η επιλογή ενός συνδέσμου.

Τα παραπάνω στοιχεία επαρκούν για την πλήρη περιγραφή όχι μόνο της ίδιας σελιδοπροβολής αλλά σχετικής με αυτήν πληροφορίας. Δίνει στοιχεία για την προέλευση του χρήστη, που μπορεί να είναι είτε από τον ίδιο τον ιστοχώρο είτε από εξωτερική διεύθυνση (για παράδειγμα μηχανή αναζήτησης). Επίσης περιγράφει το ακριβές περιεχόμενο της σελίδας που εμφανίστηκε στην οθόνη του χρήστη το οποίο παρουσιάζει μεγάλο ενδιαφέρον ειδικά σε δυναμικούς ιστοχώρους επειδή η εμφάνιση της κάθε σελίδας μπορεί να αλλάζει με το χρόνο ή ανάλογα με το χρήστη. Επίσης μπορεί να συνδυάσει μετέπειτα ενέργειες του χρήστη με τη συγκεκριμένη σελίδα. Τέτοια μπορεί να είναι η επιλογή ενός συνδέσμου ή κάποιας διαφήμισης κλπ.

Όπως αναφέρθηκε σε προηγούμενη παράγραφο, τα ακριβή στοιχεία που στέλνονται με κάθε σελιδοπροβολή καθορίζονται κατά την ανάπτυξη του ιστοχώρου από τους προγραμματιστές. Παρόλο που μπορεί να αυτοματοποιηθεί η διαδικασία για τις βασικές πληροφορίες που περιέχονται σε κάθε πακέτο, το πραγματικό ενδιαφέρον προέρχεται όταν συνδυάζεται η γνώση για τη συγκεκριμένη κατασκευή του ιστοχώρου.

Έστω για παράδειγμα ζητείται η δημιουργία μίας δομής σελιδοπροβολής για τα αποτελέσματα αναζήτησης. Με αυτόματο τρόπο θα μπορούσε να δημιουργηθεί η δομή που να περιέχει το όνομα της σελίδας (searchresults) και τα υπόλοιπα στοιχεία από το HTTP πρωτόκολλο όπως η προέλευση κλπ. Αυτό που όμως δεν μπορεί να καταγραφεί είναι για παράδειγμα πόσα αποτελέσματα βρέθηκαν και εμφανίστηκαν στο χρήστη, επειδή αυτό απαιτεί γνώση για τον ίδιο ιστοχώρο και το είδος της ερώτησης.

Αυτό που αναφέρεται στον πίνακα ως Page Actions είναι στην πραγματικότητα ένα σύνολο από δυάδες που αντιστοιχούν σε κάθε σελιδοπροβολή ένα αναγνωριστικό για το είδος της ενέργειας της σελίδας (για παράδειγμα εμφάνιση αποτελέσματος αναζήτησης) και ένα αναγνωριστικό για το αντικείμενο που σχετίζεται με την ενέργεια (για παράδειγμα το ID της περιοχής ενδιαφέροντος).

Έστω ότι στα αποτελέσματα αναζήτησης εμφανίζονται 4 περιοχές ενδιαφέροντος και 3 ξενοδοχεία. Η σελιδοπροβολή μπορεί να περιγραφεί με το όνομα της σελίδας (searchresults) και για κάθε αποτέλεσμα μία δυάδα με το ID της περιοχής ή του ξενοδοχείου και της ενέργειας “Εμφάνιση Αποτελέσματος”.

Παρόλο που η συγκεκριμένη προσέγγιση δημιουργεί αρκετά μεγάλο φόρτο στην ανάπτυξη του ιστοχώρου, προσφέρει από την άλλη μεριά μία μεγάλη ευελιξία στην περιγραφή των σελίδων που εμφανίζονται στο χρήστη.

Είναι σημαντικό επίσης να τονιστεί ότι οι δομές σελιδοπροβολής στη γενική περίπτωση είναι εντελώς ανεξάρτητες από τις προηγούμενες ενέργειες του χρήστη, δηλαδή δεν έχουν κατάσταση. Ο συσχετισμός των σελιδοπροβολών στην έννοια της επίσκεψης είναι ενέργειες που γίνονται από την υπηρεσία ενημέρωσης, η οποία είναι και υπεύθυνη για τη διασφάλιση της εγκυρότητας της πληροφορίας, όταν για παράδειγμα ο χρήστης δε δέχεται cookies.

### **3.1.3 Υπηρεσία Ενημέρωσης**

Η Υπηρεσία Ενημέρωσης αποτελεί το σημείο συλλογής των δομών σελιδοπροβολής απ’ τις εφαρμογές ιστοχώρου. Μετά από ελέγχους συνέπειας των δεδομένων ενημερώνεται η κεντρική βάση δεδομένων.

Πέραν της λειτουργίας συλλογής των δεδομένων, η πιο σημαντική ενέργεια που γίνεται είναι η δημιουργία της έννοιας της επίσκεψης ανά χρήστη. Πρέπει δηλαδή οι διάφορες σελιδοπροβολές να ομαδοποιηθούν ανά χρήστη για να δημιουργηθεί το μονοπάτι που ακολούθησε ο κάθε ένας από αυτούς. Για το σκοπό αυτό χρησιμοποιείται αρχικά η τιμή του SessionId που προέρχεται από ένα cookie. Στην περίπτωση που διαπιστωθεί ότι ο χρήστης δε δέχεται τα cookies χρησιμοποιούνται οι ευριστικές μέθοδοι που περιγράφονται στην ενότητα σχετικής έρευνας για Web Usage Analysis (βλέπε κεφάλαιο 2), για την αντιστοίχιση σε συγκεκριμένη επίσκεψη.

Στην υπηρεσία αυτή γίνονται και άλλες δευτερεύοντες λειτουργίες όπως η αποσυναρμολόγηση του ονόματος περιοχής του πελάτη σε χώρα και οργανισμό, ενώ για ειδικές περιπτώσεις έχει προβλεφθεί και περαιτέρω ανάλυση (για παράδειγμα σε Dialup χρήστες της Otenet και της Forthnet μπορεί από το όνομα του μηχανήματος που αντιστοιχεί στη σύνδεση να εξαχθεί συμπέρασμα για την περιοχή κατοικίας του χρήστη).

Εάν η σελιδοπροβολή προήλθε από γνωστή εξωτερική μηχανή αναζήτησης, εξάγεται από το Referring URL η φράση αναζήτησης που χρησιμοποίησε ο χρήστης και αποθηκεύεται και αυτή σαν χαρακτηριστικό της επίσκεψης. Η φράση αναζήτησης δίνει πολύτιμη πληροφορία για το πως

ψάχνουν οι χρήστες για σχετικές εφαρμογές και μπορεί να χρησιμοποιηθούν ώστε να βελτιωθούν οι σελίδες των ιστοχώρων ώστε να δεικτοδοτούνται πιο αποτελεσματικά από τις μηχανές αναζήτησης.

Όσον αφορά τη χώρα προέλευσης του χρήστη, όπως αναφέρθηκε χρησιμοποιείται η κατάληξη από το όνομα περιοχής (για παράδειγμα .gr ή .de). Πολλές φορές όμως αυτά τα στοιχεία δεν είναι διαθέσιμα επειδή δεν υπάρχει η απαραίτητη Reverse DNS ονομασία από τη διεύθυνση IP. Σαν δεύτερη πηγή για τη χώρα προέλευσης χρησιμοποιείται έτσι η βάση δεδομένων του GeoIP [11] η οποία διατηρεί την αντιστοίχιση διευθύνσεων σε χώρες προέλευσης. Ειδικά για περιπτώσεις κατάληξης όπου δεν υπάρχει αντιστοίχιση σε συγκεκριμένη χώρα (π.χ. .com ή .net) η βάση δεδομένων του GeoIP δίνει πολύ πιο ακριβή αποτελέσματα. Αυτή η βάση διατηρείται τοπικά οπότε δεν υπάρχει μεγάλο κόστος στην αντιστοίχιση διευθύνσεων IP σε χώρες, ενώ η ενημέρωση της βάσης γίνεται ανά τακτά χρονικά διαστήματα.

Επειδή η Υπηρεσία Ενημέρωσης είναι κεντρικό σημείο συλλογής των δεδομένων και διενεργεί ούτως ή άλλως ορισμένες εργασίες, προσφέρεται σαν σημείο υπηρεσίας για τις ίδιες τις εφαρμογές ιστοχώρου. Έτσι για παράδειγμα μπορεί να απαντήσει σε ερωτήσεις όπως “ποιος είναι ο αριθμός των ενεργών επισκέψεων” ή “ποια φράση αναζήτησης χρησιμοποίησε ο επισκέπτης από το Google” ή και “πότε ήταν η τελευταία φορά που επισκέφθηκε ο χρήστης τον ιστοχώρο” χωρίς την ανάγκη πρόσβασης απ' ευθείας στη βάση δεδομένων. Μπορεί δηλαδή να χρησιμοποιηθεί ως υπηρεσία για υψηλού επιπέδου πληροφορίες από τις εφαρμογές ιστοχώρου.

### **3.1.4 Βάση Χρήσης**

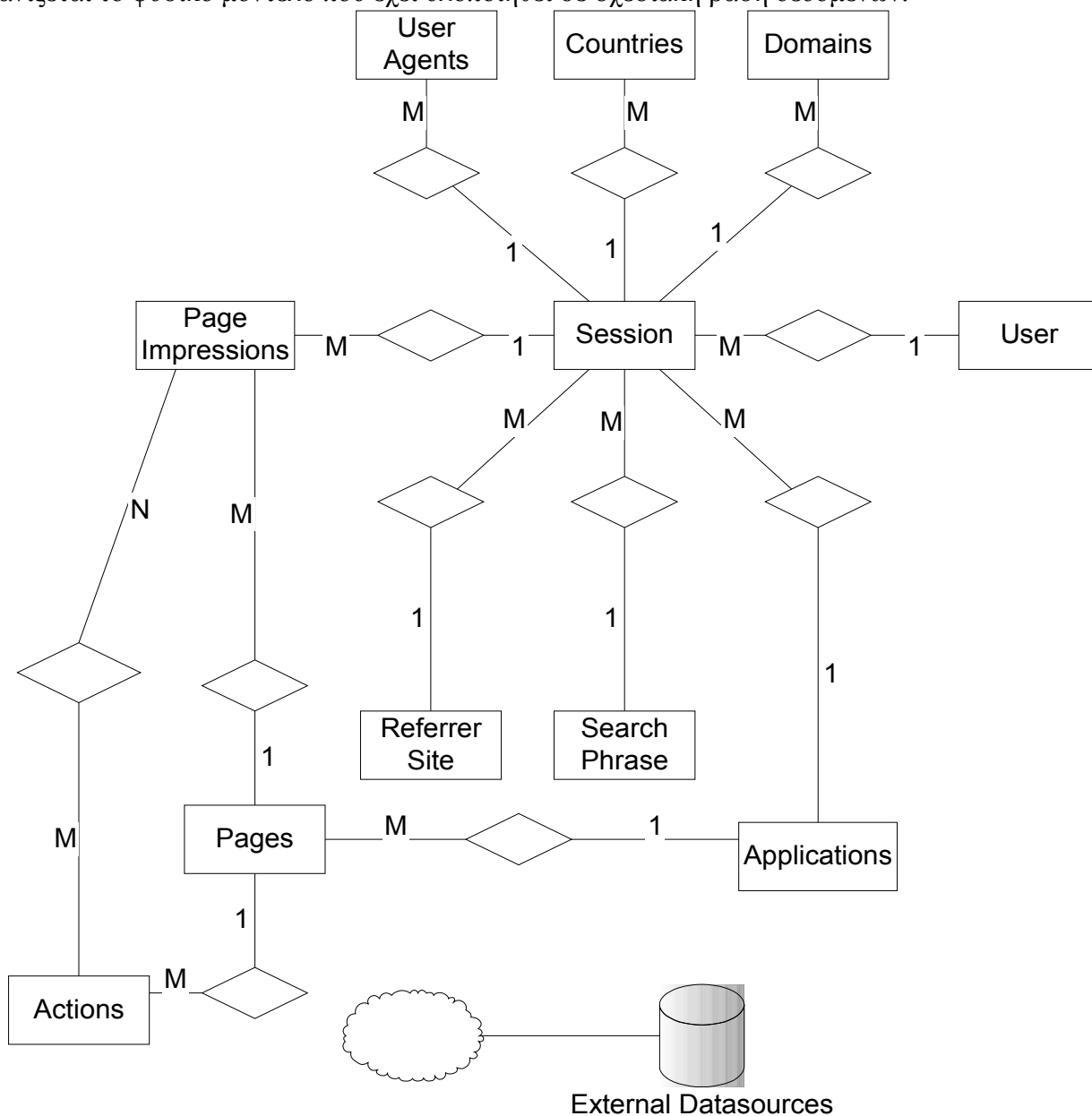
Στη βάση δεδομένων χρήσης ιστοχώρων είναι αποθηκευμένες οι πληροφορίες από τις κινήσεις των χρηστών και αποτελεί έτσι μία πολύτιμη πηγή για την περαιτέρω ανάλυση της συμπεριφοράς των χρηστών. Όπως αναφέρθηκε η βάση ενημερώνεται σε πραγματικό ή σχεδόν πραγματικό χρόνο από την υπηρεσία ενημέρωσης και η πληροφορία βρίσκεται άμεσα σε δομημένη μορφή, χωρίς να απαιτείται κανενός είδους προεπεξεργασία (βλέπε κεφάλαιο 2 σχετική έρευνα).

Το σχήμα που υλοποιήθηκε για την αποθήκευση των δεδομένων συμπεριλαμβάνει τις γνωστές έννοιες από το Web Usage Analysis (όπως η έννοιες του χρήστη, της επίσκεψης, μονοπάτια σελιδοπροβολής κλπ.) ενώ επιτρέπει ταυτόχρονα την αποθήκευση επιπλέον πληροφορίας για κάθε σελιδοπροβολή που αναφέρεται στο περιεχόμενο της κάθε σελίδας και τις ενέργειες που εκτελούνται από το χρήστη και δεν έχουν αντιστοίχιση σε σελιδοπροβολή.

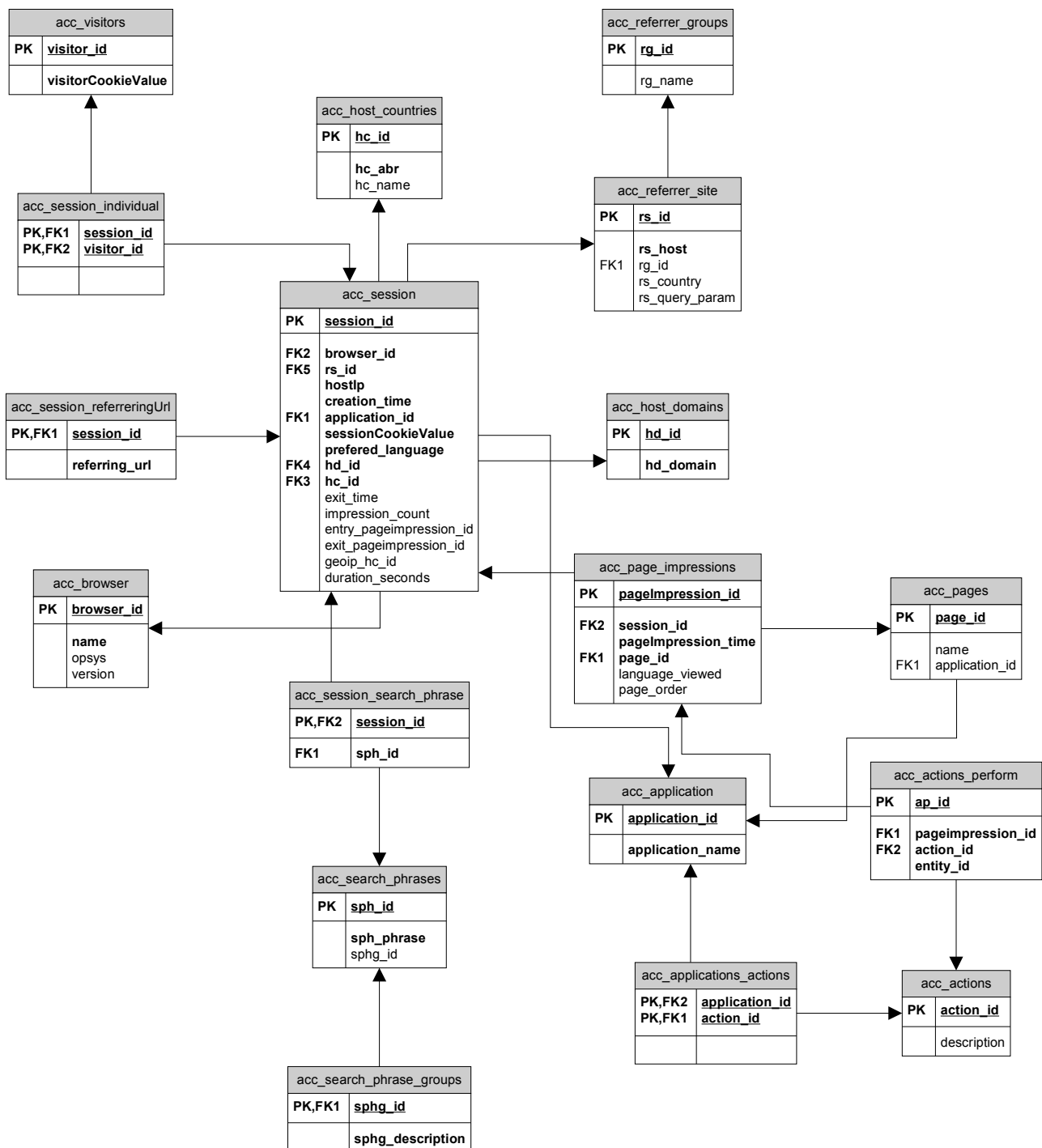
Το σχήμα λαμβάνει επίσης υπ' όψιν του την πολυγλωσσικότητα που μπορεί να έχουν οι ιστοχώροι, και τη διασύνδεση με άλλες πηγές δεδομένων ώστε να μπορεί να εκμεταλλεύεται ο αναλυτής τις

τυχών επιπλέον πληροφορίες που υπάρχουν σε αυτές.

Σε αυτή την ενότητα θα παρουσιαστεί το σχήμα της Βάσης Χρήσης. Στο Σχήμα 3.2 παρουσιάζεται ένα απλοποιημένο ER Διάγραμμα στο οποίο θα βασιστεί η περιγραφή, ενώ στο Σχήμα 3.3 εμφανίζεται το φυσικό μοντέλο που έχει υλοποιηθεί σε σχεσιακή βάση δεδομένων.



Σχήμα 3.2 Απλοποιημένο ER Διάγραμμα της Βάσης Χρήσης



Σχήμα 3.3 Φυσικό μοντέλο της Βάσης Χρήσης σε σχεσιακή βάση δεδομένων που υλοποιήθηκε για την αποθήκευση των δεδομένων χρήσης από εφαρμογές διαδικτύου

### 3.1.4.1 Εφαρμογή (Application)

Η οντότητα εφαρμογή αναπαριστά διαφορετικές εφαρμογές ιστοχώρου που μπορεί να διαχειρίζονται από την ίδια βάση χρήσης. Ανάλογα με την υλοποίηση μπορεί να συσχετιστούν οι χρήστες που επισκέπτονται διαφορετικές εφαρμογές ιστοχώρου. Κάθε εφαρμογή έχει ένα σύνολο σελίδων ενώ η πιο σημαντική συσχέτιση είναι οι επισκέψεις (sessions) που αντιστοιχούν στις εφαρμογές.



#### **3.1.4.2 Ιστοσελίδα (Pages)**

Οι σελίδες έχουν άμεση αντιστοίχιση στην έννοια των HTML σελίδων και αναπαριστούν την εικόνα που εμφανίζεται στις εφαρμογές πλοήγησης των χρηστών. Οι σελίδες είναι συστατικό των σελιδοπροβολών και αναπαριστούν μία υψηλού επιπέδου έννοια του ιστοχώρου, ανεξάρτητα από το συγκεκριμένο URL με το οποίο είναι προσβάσιμες. Στο Crete TOURnet για παράδειγμα υπάρχει η έννοια της “παρουσίασης περιοχής ενδιαφέροντος”, η οποία στη διάρκεια ανάπτυξης και εξέλιξης του ιστοχώρου άλλαξε τόσο στα URLs όσο και στο συγκεκριμένο περιεχόμενο της σελίδας. Το όνομα της σελίδας όπως αποθηκεύεται στη Βάση Χρήσης παρέμεινε το ίδιο. Κάθε σελίδα έχει επίσης ένα σύνολο ενεργειών (actions) με τα οποία περιγράφεται το περιεχόμενο της κάθε σελίδας καθώς και οι ενέργειες που γίνονται σε αυτήν (π.χ. ποιο σύνδεσμο επέλεξε ο χρήστης).

#### **3.1.4.3 Χρήστης (User)**

Ο χρήστης αναφέρεται στον τελικό χρήστη που επισκέπτεται έναν ιστοχώρο. Ο ίδιος χρήστης μπορεί να επισκέπτεται πολλές φορές τον ιστοχώρο, και είναι θεμιτό να αντιστοιχιστούν οι διάφορες επισκέψεις στο ίδιο πρόσωπο. Στην πράξη αυτό λόγω τεχνικών δυσκολιών δεν είναι εύκολο να επιτευχθεί. Στο συγκεκριμένο σχήμα ο χρήστης αντιστοιχεί στην έννοια του tracked visitor της παραγράφου 2.3.4.1.4.

#### **3.1.4.4 Επίσκεψη (Session)**

Η οντότητα session αναφέρεται σε διαδοχικές σελιδοπροβολές του ιστοχώρου από έναν χρήστη. Η επίσκεψη είναι ίσως η πιο σημαντική πληροφορία στην ανάλυση επισκεψιμότητας. Χαρακτηρίζεται από πλήθος στοιχείων, όπως η σελιδοπροβολή εισόδου και εξόδου, ώρα έναρξης, χώρα προέλευσης του χρήστη, προέλευση επίσκεψης (referrer) κ.α.

#### **3.1.4.5 User Agent**

Ο user agent αναφέρεται στην εφαρμογή πλοήγησης που χρησιμοποιεί ο χρήστης. Χρησιμοποιείται για την αντιστοίχιση σελιδοπροβολών σε επισκέψεις στην περίπτωση που δεν είναι διαθέσιμος άλλος τρόπος όπως cookies. Δίνει επίσης μία ένδειξη για τις τεχνικές δυνατότητες που έχουν οι εφαρμογές που χρησιμοποιούν οι χρήστες, και έτσι βοηθάει στην επιλογή των τεχνολογιών για την ανάπτυξη των ιστοχώρων.

#### **3.1.4.6 Χώρα (Country)**

Αναφέρεται στη χώρα προέλευσης των χρηστών. Η χώρα μπορεί να εξαχθεί είτε από την κατάληξη του ονόματος περιοχής του μηχανήματος που χρησιμοποιεί ο χρήστης, είτε από βάσεις δεδομένων που αντιστοιχούν απ' ευθείας τις IP διευθύνσεις σε χώρες. Για τη χώρα προέλευσης ισχύει η αίρεση της αβεβαιότητας από τους τυχών ενδιάμεσους σταθμούς στην αλυσίδα HTTP.

#### **3.1.4.7 Όνομα Περιοχής (Domain)**

Το Domain αναφέρεται στο όνομα περιοχής δεύτερου ή τρίτου επιπέδου του μηχανήματος του χρήστη. Γενικά μπορεί να χρησιμοποιηθεί για πληροφορία σχετικά με τον οργανισμό του χρήστη (π.χ. το όνομα περιοχής yperth.gr ανήκει στο Υπουργείο Εθνικής Παιδείας της Ελλάδας). Ειδικά για χρήστες Dialup συνδέσεων μπορεί να δώσει χωρίς τη χρήση τρίτων βάσεων δεδομένων πληροφορία για την περιοχή προέλευσης του χρήστη, και αυτό επειδή πολλοί πάροχοι τέτοιων υπηρεσιών αντιστοιχούν στο όνομα περιοχής και τον κόμβο απ' όπου έγινε η σύνδεση. Για παράδειγμα, χρήστες της Otenet που συνδέονται από το Ηράκλειο, έχουν όνομα περιοχής της μορφής heraxxx.otenet.gr, όπου το αρχικό hera υποδηλώνει το Ηράκλειο.

#### **3.1.4.8 ιστοχώρος Προέλευσης (Referrer Site)**

Ο ιστοχώρος προέλευσης αναφέρεται ακριβώς στη σελίδα απ' όπου προήλθε ο χρήστης. Πέραν της γενικότερης τοποθέτησης του ιστοχώρου στο Διαδίκτυο σε όρους διασύνδεσης και διεύθυνσης μπορεί να δώσει και πλήθος άλλων στοιχείων για τον ίδιο το χρήστη. Για παράδειγμα εάν είναι γνωστό ότι ο συγκεκριμένος ιστοχώρος προέλευσης ασχολείται με ένα συγκεκριμένο θέμα μπορεί βάσιμα να υποτεθεί ότι ο χρήστης στον υπό ανάλυση ιστοχώρο ενδιαφέρεται για το ίδιο θέμα. Ειδικά για μηχανές αναζήτησης μπορεί από το Referring URL να εξαχθεί η φράση αναζήτησης που χρησιμοποίησε ο χρήστης για να φτάσει στον υπό ανάλυση ιστοχώρο.

#### **3.1.4.9 Φράση Αναζήτησης (Search Phrase)**

Πρόκειται για τις φράσεις αναζήτησης από μηχανές αναζήτησης. Οι φράσεις μπορεί να “σπάσουν” σε πιο βασικές έννοιες συνδυασμός λέξεων (δε φαίνεται στο σχήμα) και να ομαδοποιηθούν σε κατηγορίες για τους χρήστες ανάλογα με τα ενδιαφέροντα τους. Για παράδειγμα μπορεί να χρησιμοποιηθεί η απλή φράση “Κρήτη” που δεν δίνει περισσότερη πληροφορία πέραν του ενδιαφέροντος για την Κρήτη. Εάν όμως υπάρχει η φράση “Ξενοδοχεία Κρήτη” υπάρχει ήδη γνώση για πιο συγκεκριμένα ενδιαφέροντα του επισκέπτη. Στη φράση “Αγροτουρισμός ξενοδοχείο Κρήτη” γίνεται ακόμα μεγαλύτερη εξειδίκευση του χρήστη.

#### **3.1.4.10 Σελιδοπροβολές (Page Impressions)**

Η έννοια των σελιδοπροβολών που αναφέρθηκε πολλές φορές σε προηγούμενες παραγράφους αποτελεί ουσιαστικά την κίνηση των χρηστών στον ιστοχώρο. Πολλές σελιδοπροβολές αντιστοιχούν σε μία επίσκεψη. Οι σελιδοπροβολές είναι διατεταγμένες ανάλογα με την ώρα που έγιναν και διαμορφώνουν το μονοπάτι που ακολούθησε ο χρήστης κατά την επίσκεψή του. Οι σελιδοπροβολές αναλύονται περαιτέρω στις ενέργειες σελιδοπροβολής (Page Actions).

### **3.1.4.11 Ενέργειες Σελίδας (Actions)**

Στην πιο απλή περίπτωση η σελιδοπροβολή περιγράφεται αποκλειστικά από το όνομα της σελίδας. Στην περίπτωση που υπάρχουν όμως δυναμικά παραγόμενες σελίδες, ή οι σελίδες αλλάζουν συχνά μορφή, οι σελιδοπροβολές μπορεί να αναλυθούν σε Ενέργειες Σελίδας. Οι ενέργειες προσδιορίζουν το ακριβές περιεχόμενο της κάθε σελιδοπροβολής και είναι στην απόφαση των αναλυτών και των κατασκευαστών του ιστοχώρου να καθορίσουν ποια ακριβώς στοιχεία επιθυμούν να καταγράφονται.

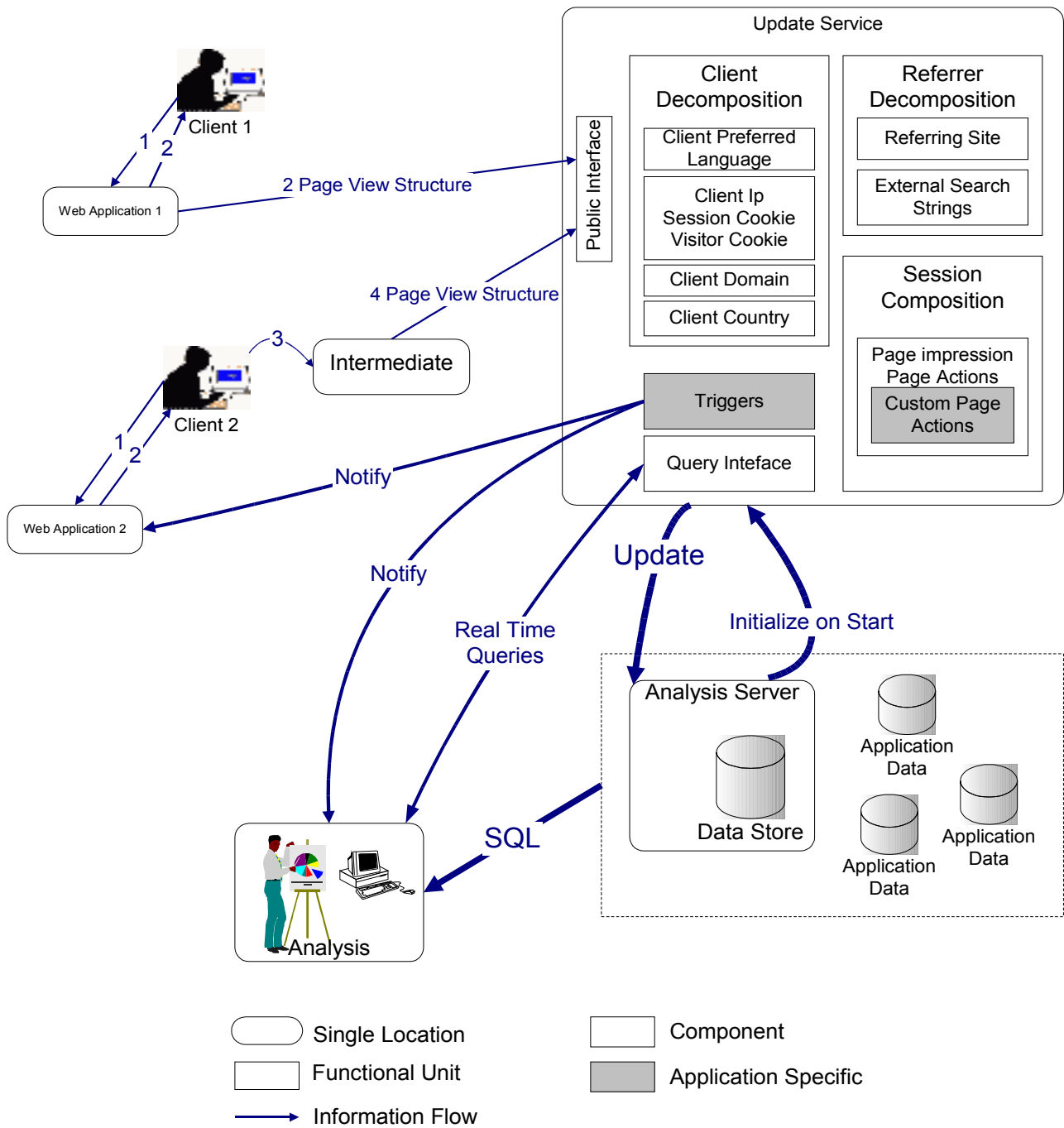
Οι ενέργειες μπορεί να χαρακτηριστούν με επιπλέον χαρακτηριστικά πέραν του αναγνωριστικού της ενέργειας (action identifier), τα οποία συνδέουν την ενέργεια με άλλες βάσεις δεδομένων. Για παράδειγμα σε έναν ιστοχώρο που παρουσιάζει περιοχές ή επιχειρήσεις χρησιμοποιούνται για να προσδιορίσουν ποια ακριβώς περιοχή ή επιχείρηση παρουσιάζεται (για παράδειγμα για την παρουσίαση της περιοχής Χανίων θα αποθηκευόταν κάτι της μορφής *Page area, action present area, entity Chania*).

### **3.1.4.12 Αναζήτηση εντός Ιστοχώρου**

Μία σημαντική έλλειψη του σχεσιακού σχήματος που παρουσιάστηκε για την αποθήκευση των δεδομένων χρήσης είναι η δομημένη αποθήκευση των κριτηρίων αναζήτησης από φόρμες HTML εντός κάποιου ιστοχώρου. Μπορεί να χρησιμοποιούνται τόσο πεδία ελεύθερου κειμένου όσο και στοιχεία δομημένα σε δυαδικές επιλογές ή λίστα πολλαπλών επιλογών. Το πρόβλημα προκύπτει από την ξεχωριστή αντιμετώπιση που έχει κάθε εφαρμογή με βάση το συγκεκριμένο σχήμα της πηγής ή των πηγών δεδομένων που χρησιμοποιεί και αντικατοπτρίζεται στις επιλογές που δίνονται στο χρήστη. Ανάλογα με τη συγκεκριμένη εφαρμογή μπορεί όμως τα δημιουργηθούν ειδικές δομές και να συνδεθούν με τη σχέση *acc\_actions\_perform* που συνδέει μία συγκεκριμένη σελιδοπροβολή με τις ενέργειες που γίνονται πάνω σε αυτήν (βλέπε Σχήμα 3.3 σελίδα 120).

### **3.1.5 Αρχιτεκτονική Συστήματος**

Στο διάγραμμα 3.1 φαίνεται η αρχιτεκτονική του συστήματος με τα διάφορα υποσυστήματα που το συνθέτουν, τη φορά της μετάδοσης δεδομένων και 2 εναλλακτικούς τρόπους για την αποστολή των δομών σελιδοπροβολής.



Διάγραμμα 3.1 Αρχιτεκτονική του Συστήματος.

Τα πλαίσια με τις καμπύλες αναφέρονται σε συστήματα που βρίσκονται σε ανεξάρτητες τοποθεσίες από τα υπόλοιπα. Αυτά είναι οι διάφορες εφαρμογές ιστοχώρου (Web Application 1 και 2), η υπηρεσία ενημέρωσης (Update Service), ο εξυπηρετητής ανάλυσης (Analysis Server), η υπηρεσία ενδιάμεσης αποστολής (Intermediate) και η τοποθεσία όπου γίνεται καθ' αυτή η ανάλυση (Analysis).

Η υπηρεσία ενημέρωσης δέχεται τις δομές σελιδοπροβολής που δημιουργούνται και αποστέλλονται

σε αυτήν με κάθε σελιδοπροβολής που λαμβάνει χώρα σε κάποια από τις συνεργαζόμενες εφαρμογές ιστοχώρου. Γι' αυτό το σκοπό προσφέρει ένα δημόσιο API, που για τη συγκεκριμένη υλοποίηση χρησιμοποιήθηκε Java Remote Method Invocation.

Στην υπηρεσία ενημέρωσης γίνεται η αποσύνθεση των δεδομένων για τα χαρακτηριστικά του πελάτη (Client Decomposition) και της προέλευσής του (Referrer Decomposition). Οι πληροφορίες που εξάγονται χρησιμοποιούνται στη σύνθεση της επίσκεψης του κάθε πελάτη σε όρους μονοπατιού και περιεχομένου των σελίδων που εμφανίστηκαν.

Από την υπηρεσία ενημέρωσης προσφέρεται επίσης μία λειτουργία για υποβολή ερωτήσεων για την τρέχουσα κατάσταση του συστήματος, όπως αριθμός ενεργών επισκέψεων, προέλευση τρεχούσων επισκεπτών κλπ. (Query Interface) Προσφέρει επίσης μία υπηρεσία που ενεργοποιείται και ενημερώνει τρίτους όταν εκπληρώνονται συνθήκες που καθορίζονται ανά εφαρμογή, όπως μεγάλος φόρτος, ασυνήθιστες αλλαγές κλπ. (Triggers).

Η υπηρεσία ενημέρωσης ενημερώνει τον εξυπηρετητή ανάλυσης (Analysis Server) με τα στοιχεία για κάθε σελιδοπροβολή και τις επισκέψεις στις οποίες ανήκουν. Κατά την εκκίνηση της υπηρεσίας ενημέρωσης, αυτή αρχικοποιείται με στοιχεία από τον εξυπηρετητή ανάλυσης που καλύπτουν θέματα όπως οι προδιαγραφές των URL προέλευσης για εξωτερικές μηχανές αναζήτησης.

Η ανάλυση των δεδομένων γίνεται σε διαφορετική τοποθεσία (Analysis) στην οποία χρησιμοποιούνται τόσο τα δεδομένα από τον εξυπηρετητή ανάλυσης, όσο και από άλλες πηγές δεδομένων που εξαρτώνται από την κάθε εφαρμογή. Για την ανάλυση χρησιμοποιείται η γλώσσα SQL για την υποβολή ερωτήσεων.

Στο διάγραμμα φαίνονται επίσης δύο εναλλακτικοί τρόποι για την αποστολή των δομών σελιδοπροβολής στην υπηρεσία ενημέρωσης. Στην πρώτη περίπτωση (Client 1, Web Application 1) ο πελάτης στέλνει στην εφαρμογή μία αίτηση για την προβολή μίας σελίδας (βήμα 1). Η εφαρμογή στέλνει στον πελάτη την απάντηση αλλά ταυτόχρονα συνθέτει και αποστέλλει στην υπηρεσία ενημέρωσης τη δομή σελιδοπροβολής με όλα τα στοιχεία για τον πελάτη και τη σελίδα που του στάλθηκε ως απάντηση (βήμα 2).

Στη δεύτερη περίπτωση (Client 2, Web Application 2) η αποστολή της δομής ενημέρωσης γίνεται από έναν ενδιάμεσο εξυπηρετητή. Αυτός ενεργοποιείται με αίτηση που γίνεται από τον πελάτη. Για να γίνει αυτή η αίτηση από τον πελάτη, η εφαρμογή στέλνει στον πελάτη μαζί με τη σελίδα ειδικό HTML κώδικα, ο οποίος ενεργοποιεί στην εφαρμογή πλοήγησης την αίτηση προς τον ενδιάμεσο (π.χ. μία εικόνα). Στην αίτηση που γίνεται προς τον ενδιάμεσο υπάρχει κωδικοποιημένη όλη η

πληροφορία για τη δημιουργία των δομών σελιδοπροβολής.

Με την παραπάνω αρχιτεκτονική γίνεται εμφανές ότι η ανάλυση ενός συγκεκριμένου ιστοχώρου δεν περιορίζεται από τα όρια του εξυπηρετητή που καταγράφει ή αποθηκεύει τα δεδομένα. Αντιθέτως μπορεί να γίνει σε εντελώς ανεξάρτητη τοποθεσία ή να ακολουθηθεί το μοντέλο του outsourcing για την προσφορά υπηρεσιών ανάλυσης για κάθε ιστοχώρο.

Η δυνατότητα που έχουν οι εφαρμογές ιστοχώρου να μεταθέτουν ουσιαστικά την αποστολή των δομών καταγραφής στον πελάτη (με τη βοήθεια του ενδιάμεσου εξυπηρετητή), γλιτώνει από τους κατασκευαστές της εφαρμογής από την προμήθεια και συντήρηση επιπλέον υλικού και πόρων δικτύου και τον επιπλέον φόρτο που θα υπήρχε.

### **3.2 Μεθοδολογία Επιλογής Στοιχείων Καταγραφής**

Στις προηγούμενες παραγράφους περιγράφηκε το σχήμα της Βάσης Χρήσης και η πληροφορία που μπορεί να αποθηκευτεί σε αυτήν. Το δομικό στοιχείο που χρησιμοποιείται για την ενημέρωση της βάσης χρήσης είναι δομές σελιδοπροβολής που περιγράφηκαν στην παράγραφο 3.1.2 (επαναλαμβάνεται η δομή στη συνέχεια).

Για κάθε σελιδοπροβολή που γίνεται στο ιστοχώρο, πρέπει να δημιουργηθούν οι κατάλληλες δομές. Για το μεγαλύτερο, αριθμητικά, πλήθος των στοιχείων που αποθηκεύονται στις δομές σελιδοπροβολής, υπάρχουν συγκεκριμένες και τυποποιημένες τιμές που επιλέγονται. Τέτοια στοιχεία είναι για παράδειγμα η διεύθυνση IP του επισκέπτη ή η διεύθυνση προέλευσης (referring URL). Τα στοιχεία όμως για το περιεχόμενο των σελίδων αλλά και καθ' αυτές τις σελίδες εξαρτώνται άμεσα από τον υπό ανάλυση ιστοχώρο. Για' αυτό το λόγο η δημιουργία των δομών σελιδοπροβολής απαιτεί εις βάθος γνώση του ιστοχώρου και ανάλογα με το σκοπό για τον οποίο γίνεται η ανάλυση μπορεί να επιλεγεί να δοθεί περισσότερο ή λιγότερο βάρος σε κάποιες σελίδες έναντι άλλων. Η επιλογή των στοιχείων που τελικά θα καταγραφούν είναι υπό την ευθύνη των ατόμων που αναπτύσσουν τον ιστοχώρο και απαιτεί τη συνεργασία με την ομάδα που θα κληθεί να τα αναλύσει .

Στη συνέχεια θα παρουσιαστεί μία μεθοδολογία που οδηγεί στη δημιουργία των δομών σελιδοπροβολής ενός ιστοχώρου. Ουσιαστικά η εφαρμογή της μεθοδολογίας οδηγεί στη δόμηση των στοιχείων που θα αποθηκευτούν στη Βάση Χρήσης.

Η συγκεκριμένη υλοποίηση για τον ιστοχώρο τουριστικού προορισμού θα παρουσιαστεί στο τέλος του κεφαλαίου.

Ο Πίνακας 3.2 απαριθμεί πάλι τα στοιχεία που περιέχονται σε κάθε δομή σελιδοπροβολής

<i>Δεδομένα της Δομής Σελιδοπροβολής</i>	
1	Application Id
2	Client IP Address
3	Hostname
4	Referring URL
5	User Agent
6	Session Id
7	Date Time
8	Visitor Id
9	Preferred Language
10	Forwarders (proxy)
11	Viewed Language
12	Page Id
13	Page actions {action_id, entity_id   null}+
14	Post Impression Id

Πίνακας 3.2 Περιεχόμενα δομής σελιδοπροβολής

### 3.2.1 Μεθοδολογία Δημιουργίας Δομών Σελιδοπροβολής

Για τη δημιουργία των δομών σελιδοπροβολής, υπάρχουν δύο έννοιες που πρέπει να προσδιοριστούν για τον κάθε ιστοχώρο. Η έννοια της σελίδας του ιστοχώρου και η έννοια του περιεχομένου σε ιστοχώρο. Για τη διευκόλυνση της περαιτέρω περιγραφής δίνονται οι ακόλουθοι ορισμοί:

**Περιεχόμενο ιστοχώρου:** ως *Περιεχόμενο* σε έναν ιστοχώρο ορίζουμε το σύνολο των δεδομένων που παρουσιάζονται σε εφαρμογές πλοήγησης παγκόσμιου ιστού και προέρχονται από τον εξυπηρετητή HTTP του ιστοχώρου

**Τμήμα Περιεχομένου:** το *Περιεχόμενο ενός Ιστοχώρου* μπορεί να χωριστεί σε διακριτά *Τμήματα Περιεχομένου* τα οποία μπορεί να είναι οργανωμένα σύμφωνα με ορισμένη *κατηγοριοποίηση*

**Σελίδα ιστοχώρου:** ως *Σελίδα Ιστοχώρου* ορίζεται το υποσύνολο των *Τμημάτων Περιεχομένου* που εμφανίζεται στα πλαίσια ενός παραθύρου σε εφαρμογή πλοήγησης παγκόσμιου ιστού

**Σημαντικό Περιεχόμενο:** τα *Τμήματα Περιεχομένου* του ιστοχώρου των οποίων τη χρήση θέλουμε να καταγράψουμε αποτελούν το *Σημαντικό Περιεχόμενο* του ιστοχώρου

**Γενική Σελίδα:** *Γενική Σελίδα* ονομάζεται η *Σελίδα Ιστοχώρου* που εμφανίζει *Σημαντικό Περιεχόμενο*

**Απλή Γενική Σελίδα:** ως *Απλή Γενική Σελίδα* ιστοχώρου ορίζεται η *Σελίδα* που εμφανίζει πάντα τα ίδια *Τμήματα Περιεχομένου* του *Σημαντικού Περιεχομένου*

**Σύνθετη Γενική Σελίδα:** ως *Σύνθετη Γενική Σελίδα* ορίζεται η *Σελίδα* που μπορεί να εμφανίζει διαφορετικά *Τμήματα Περιεχομένου* που ανήκουν στην ίδια *κατηγορία*, ακολουθώντας την ίδια μορφή παρουσίασης

**Ενέργεια Σελίδας:** *Ενέργεια Σελίδας* είναι ένα αναγνωριστικό που περιγράφει την εμφάνιση *Σημαντικού Τμήματος Πληροφορίας*

**Ακόλουθη Ενέργεια:** *Ακόλουθη Ενέργεια* είναι αναγνωριστικό που περιγράφει επιλογές του επισκέπτη που έπονται της εμφάνισης της σελίδας, αντιστοιχίζονται όμως σε αυτήν

Η δημιουργία των δομών σελιδοπροβολής βασίζεται στις *Γενικές Σελίδες* του ιστοχώρου και το *Σημαντικό Περιεχόμενο*. Ακολουθούνται τα εξής βήματα

## **1 Καταγραφή Γενικών Σελίδων**

Καταγράφονται όλες οι *Γενικές Σελίδες* του ιστοχώρου με βάση το *Σημαντικό Περιεχόμενο* που παρουσιάζουν.

## **2 Απόδοση Αναγνωριστικού**

Σε κάθε *Γενική Σελίδα* που καταγράφηκε δίνεται ένα μοναδικό αναγνωριστικό (identifier) με το οποίο διακρίνεται.

Σε κάθε *Τμήμα Πληροφορίας* του *Σημαντικού Περιεχομένου* δίνεται επίσης ένα μοναδικό αναγνωριστικό.

## **3 Ορισμός Ενεργειών Σελίδας**

Για κάθε *Σύνθετη Γενική Σελίδα* ορίζονται ενέργειες που περιγράφουν την εμφάνιση των *Τμημάτων Πληροφορίας* του *Σημαντικού Περιεχομένου*, και ενέργειες από τις επιλογές του χρήστη που έπονται της εμφάνισης της σελίδας (Ακόλουθες Ενέργειες).

Για τις *Απλές Γενικές Σελίδες* ορίζεται μία μοναδική ενέργεια που αντιστοιχεί στην εμφάνιση της σελίδας (default action).



## 4 Δημιουργία Δομών Σελιδοπροβολής

Με κάθε σελιδοπροβολή *Γενικής Σελίδας* που λαμβάνει χώρα στον ιστοχώρο δημιουργείται από τον κώδικα παραγωγής της σελίδας μία Δομή Σελιδοπροβολής με τον ακόλουθο τρόπο:

1. τα πεδία 1 ως 10 του πίνακα 3.2 συμπληρώνονται με τα στοιχεία του HTTP πρωτοκόλλου
2. το πεδίο 11 συμπληρώνεται με το αναγνωριστικό (identifier) της γλώσσας στην οποία εμφανίζεται το *Σημαντικό Περιεχόμενο*
3. το πεδίο 12 συμπληρώνεται με το αναγνωριστικό της *Γενικής Σελίδας*
4. για το πεδίο 13
  - εάν πρόκειται για *Σύνθετη Γενική Σελίδα*, δημιουργείται ένα σύνολο δυάδων που συσχετίζουν μία *Ενέργεια* με κάθε αναγνωριστικό του *Σημαντικού Τμήματος Πληροφορίας*
  - εάν πρόκειται για *Απλή Γενική Σελίδα*, τοποθετείται ως *Ενέργεια* το “default action” χωρίς αναγνωριστικό

## 5 Ακόλουθες Ενέργειες

Για κάθε Ακόλουθη Ενέργεια δημιουργείται από τον κώδικα παραγωγής της σελίδας μία δυάδα που συσχετίζει μία ενέργεια με ένα αναγνωριστικό και προστίθεται στο σύνολο των δυάδων του πεδίου 13. Για τη συσχέτιση της ακόλουθης ενέργειας με τη σελιδοπροβολή, χρησιμοποιείται το πεδίο 14.

### 3.2.2 Εφαρμογή της Μεθοδολογίας Καταγραφής στον Ιστοχώρο Τουριστικού Προορισμού

Η παραπάνω μεθοδολογία εφαρμόστηκε στον ιστοχώρο τουριστικού προορισμού με σκοπό την καταγραφή της χρήσης των υπηρεσιών του ιστοχώρου. Κύριο μέλημα ήταν να αντικατοπτρίζονται στις γενικές σελίδες οι διαφορετικές κατηγορίες περιεχομένου που παρουσιάζονται στον ιστοχώρο και να είναι δυνατή η εκμετάλλευση της δομής και ιεραρχίας των δεδομένων που ήδη υπάρχει στη βάση δεδομένων. Έτσι επιλέχθηκαν όπου είναι δυνατόν, ως αναγνωριστικά των τμημάτων πληροφορίας τα αναγνωριστικά της βάσης δεδομένων του ιστοχώρου.

Στο παράδειγμα των σελίδων παρουσίαση περιοχής ενδιαφέροντος, καθώς και στην αρχική σελίδα του ιστοχώρου τουριστικού προορισμού θα παρουσιαστεί η εφαρμογή της μεθοδολογίας, ενώ θα ακολουθήσουν όλες οι επιλογές που έγιναν συγκεντρωτικά.

Η αρχική σελίδα του ιστοχώρου αποτελεί ένα σημείο διασύνδεσης των επισκεπτών σε όλες τις υπηρεσίες του ιστοχώρου. Αν και παρουσιάζονται με κάθε προβολή διαφορετικές εικόνες-banner, αυτές δεν ενδιαφέρουν στη συγκεκριμένη ανάλυση. Το υπόλοιπο περιεχόμενο παραμένει το ίδιο. Οπότε πρόκειται για απλή γενική σελίδα. Δόθηκε το αναγνωριστικό “index” για τη σελίδα, ενώ συνδέεται με αυτή η ενέργεια “default action” που αναφέρεται στην καθ’ αυτή εμφάνιση της σελίδας.

Ο ιστοχώρος προσφέρει επίσης σελίδες παρουσίασης περιοχών ενδιαφέροντος. Για κάθε περιοχή υπάρχει μία χαρακτηριστική φωτογραφία, μία περιγραφή και σχετικοί με την κάθε περιοχή σύνδεσμοι.. Επίσης υπάρχει ένα εικονίδιο επιλέγοντας το οποίο προστίθεται η περιοχή στο “ηλεκτρονικό φυλλάδιο”. Ανάλογα με τις παραμέτρους του URL παρουσιάζεται διαφορετική περιοχή ενδιαφέροντος, αλλά πάντα με τη μορφή που περιγράφηκε. Οπότε πρόκειται για σύνθετη γενική σελίδα. Το αναγνωριστικό που δόθηκε στη σελίδα είναι “ar” και οι ενέργειες που συνδέονται με αυτή είναι οι “View Area” για την εμφάνιση καθ’ αυτή της περιοχής και “Add Brochure” ως ακόλουθη ενέργεια όταν προστίθεται στο “ηλεκτρονικό φυλλάδιο”. Τα αναγνωριστικά που συνδέονται με κάθε ενέργεια είναι το αναγνωριστικό της περιοχής που υπάρχει από τη βάση δεδομένων της εφαρμογής ιστοχώρου. Εκμεταλλευόμαστε έτσι την ιεραρχία περιοχών ενδιαφέροντος που υπάρχει ήδη στη βάση δεδομένων της εφαρμογής.

Με βάση την περιγραφή του ιστοχώρου στην παράγραφο 10.3 και τα βήματα της μεθοδολογίας, δημιουργήθηκε το σύνολο των σελίδων και οι ενέργειες που αντιστοιχίζονται σε κάθε μία από αυτές. Όλες οι σελίδες και οι ενέργειες παρουσιάζονται συγκεντρωτικά στον πίνακα 3.3.

Στην πρώτη στήλη φαίνονται οι Γενικές Σελίδες που καταγράφηκαν. Στη δεύτερη στήλη σημειώνεται εάν πρόκειται για Απλή ή Σύνθετη Γενική Σελίδα (Α και Σ αντίστοιχα), ενώ στην τρίτη στήλη δίνεται μία σύντομη περιγραφή για το περιεχόμενο που παρουσιάζεται. Στην τέταρτη και τελευταία στήλη φαίνονται οι πιθανές ενέργειες της κάθε σελίδας.

Με τις παραπάνω επιλογές επιτυγχάνεται:

- να υπάρχει για κάθε βασική ενότητα του ιστοχώρου μία Γενική Σελίδα που αναφέρεται στην παρουσίαση πληροφορίας της κάθε ενότητας. Έτσι υπάρχουν οι σελίδες ar για παρουσίαση περιοχής ενδιαφέροντος, si για παρουσίαση αξιοθεάτου, sm για παρουσίαση επιχείρησης τουρισμού, is για παρουσίαση ενότητας του ηλεκτρονικού οδηγού και ha για την παρουσίαση λιμανιού.
- τα entity\_ids στις ενέργειες σελίδας αντιστοιχούν σε αντικείμενα της βάσης δεδομένων του

ιστοχώρου τουριστικού προορισμού. Οι προηγούμενες Γενικές Σελίδες είναι όλες σύνθετες και συνοδεύονται από μία ενέργεια που αναγνωρίζει το συγκεκριμένο τμήμα πληροφορίας που παρουσιάζεται (ενέργειες View Area, View Site, View SME, view Intro Section και View Harbour). Μέσω της σύνδεσης με τη βάση δεδομένων που διαχειρίζεται τις πληροφορίες που παρουσιάζονται στον ιστοχώρο, είναι δυνατή η εξαγωγή πληροφοριών για τα χαρακτηριστικά της κάθε περιοχής/ αξιοθεάτου/ επιχείρησης κλπ. Για παράδειγμα στη βάση δεδομένων του ιστοχώρου τουριστικού προορισμού υπάρχει αποθηκευμένος ο τύπος της κάθε επιχείρησης σε πολλαπλά επίπεδα, και γνωρίζουμε έτσι για παράδειγμα ότι μία συγκεκριμένη παρουσίαση επιχείρησης αναφέρεται σε “Επιχείρηση Διαμονής” που είναι “Ξενοδοχείου Κλασικού Τύπου”.

- σε μία άλλη μελέτη θα μπορούσαν να καταγράφονται και οι εμφανίσεις διαφημιστικών μηνυμάτων και η αντίδραση των επισκεπτών σε αυτές. Επειδή τέτοια μηνύματα εμφανίζονται σε όλες τις σελίδες, όλες οι Γενικές Σελίδες που καταγράφηκαν θα μετατρέπονταν σε σύνθετες.

Ενώ οι Γενικές Σελίδες που καταγράφηκαν καλύπτουν πλήρως τον ιστοχώρο, οι ενέργειες που συνδέονται με κάθε σελίδα καλύπτουν τις ανάγκες που κρίθηκε ότι υπάρχουν για την καταγραφή κίνησης στον ιστοχώρο και μπορεί να μεταβληθούν ανάλογα με τα ενδιαφέροντα που υπάρχουν στη μελέτη συμπεριφοράς των επισκεπτών. Όπως θα περιγραφεί σε επόμενη ενότητα, κατά τη διάρκεια της μελέτης χρησιμοποίησιμότητας έγιναν ορισμένες τροποποιήσεις στο interface του ιστοχώρου για να μελετηθεί η αντίδραση των επισκεπτών. Για παράδειγμα στη σελίδα παρουσίασης περιοχών ενδιαφέροντος, τοποθετήθηκε δεύτερος σύνδεσμος για την εμφάνιση σχετικών επιχειρήσεων. Η προτίμηση των επισκεπτών του υπάρχοντος έναντι του δεύτερου συνδέσμου, μελετήθηκε με την προσθήκη ακόλουθης ενέργειας για τις δύο επιλογές. Έτσι ήταν δυνατόν να εξαχθούν δεδομένα για τους επισκέπτες που προτιμούν τον έναν ή τον άλλο τρόπο για την εμφάνιση σχετικών επιχειρήσεων.

Ένα άλλο πείραμα που διενεργήθηκε ήταν να εμφανίζεται ένας συγκεκριμένος σύνδεσμος με διαφορετική ψηφιακή εικόνα σε τμήμα των επισκεπτών. Η εμφάνιση της κάθε εικόνας καταγραφόταν με διαφορετική ενέργεια (εμφάνιση εικόνας A, εμφάνιση εικόνας B). Προστέθηκε επίσης ακόλουθη ενέργεια για του επισκέπτες που επέλεξαν το σύνδεσμο και έτσι ήταν δυνατόν να εξαχθούν συμπεράσματα για το βαθμό που επηρέαζε η κάθε μορφή της εικόνας τη συμπεριφορά των χρηστών.

<i>Αναγνωριστικό</i>	<i>Απλή/Σύνθετη<sup>1</sup></i>	<i>Περιγραφή</i>	<i>Ενέργειες<sup>2</sup></i>
contact	A	Σελίδα με πληροφορίες επικοινωνίας με τους συντελεστές	default action
contribute	A	Σελίδα με πληροφορίες για καταχώριση επιχείρησης	default action
link	A	Σελίδα με παραδείγματα εισαγωγής συνδέσμου από τρίτους ιστοχώρους	default action
terms	A	Όροι και προϋποθέσεις	default action
about	A	Ιστορικό του ιστοχώρου	default action
addinfo	A	Προσθήκη περιεχομένου πέραν της καταχώρισης επιχειρήσεων	default action
index	A	Αρχική σελίδα ιστοχώρου	default action
guide	A	Περιεχόμενα ηλεκτρονικού οδηγού	default action
hotels	A	Φόρμα αναζήτησης διαμονής	default action
sights	A	Φόρμα αναζήτησης περιοχών και αξιοθέατων	default action
travelagents	A	Φόρμα αναζήτησης γραφείων τουρισμού	default action
rentcar	A	Φόρμα αναζήτησης ενοικίασης οχημάτων	default action
harbours	A	Ενεργός χάρτης λιμανιών	default action
brochure	A	Ηλεκτρονικό φυλλάδιο	default action
ar	Σ	Παρουσίαση περιοχής ενδιαφέροντος	(View Area, Area Id) (Add Brochure, Area Id) <sup>2</sup>
si	Σ	Παρουσίαση αξιοθέατου	(View Site, Site Id) (Add Brochure, Site Id) <sup>2</sup>

<i>Αναγνωριστικό</i>	<i>Απλή/Σύνθετη<sup>1</sup></i>	<i>Περιγραφή</i>	<i>Ενέργειες<sup>2</sup></i>
im	Σ	Παρουσίαση φωτογραφίας/ φωτογραφιών	{(View Image, Image Id) (Add Brochure, Image Id) <sup>2</sup> }   (View Images, Area Id)   (View Images, Site Id)
sm	Σ	Παρουσίαση επιχείρησης	(View SME, SME Id) (Add Brochure, SME Id) <sup>2</sup>
ev	Σ	Παρουσίαση εκδήλωσης	(View Event, Event Id) (Add Brochure, Event Id) <sup>2</sup>
is	Σ	Παρουσίαση ενότητας ηλεκτρονικού οδηγού	(View Intro Section, Intro Section Id) (Add Brochure, Intro Section Id) <sup>2</sup>
ha	Σ	Παρουσίαση λιμανιού	(View Harbour, Harbour Id) (Add Brochure, Harbour Id) <sup>2</sup>
gc	Σ	Παρουσίαση ενότητας γλωσσαρίου	(View Glossary, Glossary Section Id)
map	Σ	Παρουσίαση χάρτη	(View Map, Map Id)
sendmail	Σ	Αποστολή ηλεκτρονικού μηνύματος σε επιχείρηση	(View email Form, SME ID) (Send Email, SME Id) <sup>2</sup>
areas_on_map	Σ	Παρουσίαση σημείου περιοχής σε χάρτη	(Area On Map, Area Id)   (SME On Map, SME Id)

<i>Αναγνωριστικό</i>	<i>Απλή/Σύνθετη<sup>1</sup></i>	<i>Περιγραφή</i>	<i>Ενέργειες<sup>2</sup></i>
searchresults	Σ	Αποτελέσματα αναζήτησης	{ (Search Type, Index Map)   (Search Type, Accommodation Form)   (Search Type, Accommodation Title Index)   (Search Type, Area or Sight Form)   (Search Type, Rental Form)   (Search Type, Travel Agent Form)   (Search Type, Area Related Accommodation Form)   (Search Type, General Side Form) } { (Search Result Area, Area Id)   (Search Result Site, Site Id)   (Search Result SME, SME Id)   (Search Result Harbour, Harbour Id)   (Search Result Intro Section, Intro Section Id)   (Search Result Image, Image Id) }*
areaselecttool	Σ	Επιλογή περιοχής μέσω χάρτη/ μέσω φυλλαδίου/ μέσω ονόματος	{(Select Area, Area Id)*}
1) Α: απλή Γενική Σελίδα, Σ: σύνθετη Γενική Σελίδα 2) Ακόλουθη ενέργεια			

Πίνακας 3.3 Βασικές οντότητες καταγραφής για τον ιστοχώρο τουριστικού προορισμού με βάση τη μεθοδολογία επιλογής στοιχείων καταγραφής. Στην πρώτη στήλη φαίνονται οι Γενικές Σελίδες που καταγράφηκαν. Στη δεύτερη στήλη σημειώνεται εάν πρόκειται για Απλή ή Σύνθετη Γενική Σελίδα, ενώ στην τρίτη στήλη δίνεται μία σύντομη περιγραφή για το περιεχόμενο που παρουσιάζεται. Στην τέταρτη και τελευταία στήλη φαίνονται οι πιθανές ενέργειες της κάθε σελίδας.

### **3.3 Ανακεφαλαίωση**

Σε αυτό το κεφάλαιο παρουσιάστηκε ένα σύστημα συλλογής και καταγραφής δεδομένων χρήσης ιστοχώρων που αναπτύχθηκε στα πλαίσια της διατριβής. Τα ιδιαίτερα χαρακτηριστικά του είναι ότι μπορεί και εκμεταλλεύεται την ειδική γνώση για κάθε ιστοχώρο και αντιμετωπίζει προβλήματα όπως την πολυγλωσσικότητα ιστοχώρων. Ειδικά για το θέμα της πολυγλωσσικότητας δε μας είναι γνωστό παρόμοιο σύστημα που να υποστηρίζει σε έναν ιστοχώρο περιεχόμενο που παρουσιάζεται σε διαφορετικές γλώσσες.

Για τη συλλογή δεδομένων προς ανάλυση προτάθηκε μία μεθοδολογία, που λαμβάνει υπ' όψιν τις ανάγκες ανάλυσης και ιδιαιτερότητες του κάθε ιστοχώρου, για να δημιουργηθούν οι κατάλληλες δομές καταγραφής. Με βάση αυτή τη μεθοδολογία, παρουσιάστηκε η συγκεκριμένη εφαρμογή στον ιστοχώρο τουριστικού προορισμού.

## 4 Μεθοδολογία Ανάλυσης από το Σύστημα Καταγραφής Δεδομένων Χρήσης Ιστοχώρων

Στο προηγούμενο κεφάλαιο παρουσιάστηκε ένα σύστημα που καταγράφει τη συμπεριφορά των επισκεπτών σε ιστοχώρους, και ειδικότερα σε ιστοχώρους τουριστικού προορισμού. Το αποτέλεσμα της καταγραφής είναι ένα δομημένο σύνολο δεδομένων όπου αντιστοιχίζεται η κάθε προβολή σελίδας, με τα συστατικά μέρη της, σε εγγραφές σε βάση δεδομένων. Το βάθος ή η λεπτομέρεια της πληροφορίας που καταγράφεται καθορίζεται από τους διαχειριστές του κάθε ιστοχώρου, ανάλογα με τις ανάγκες που κρίνεται ότι υπάρχουν κάθε φορά.

Σε αυτό το κεφάλαιο θα προταθεί μία μεθοδολογία για την ανάλυση των δεδομένων που καταγράφηκαν, στοχεύοντας κυρίως σε εφαρμογές που ασχολούνται με τον τουρισμό. Η μεθοδολογία αναφέρεται στις ακόλουθες απόψεις ανάλυσης ενός ιστοχώρου:

**Πληροφόρηση για τα χαρακτηριστικά των επισκέψεων:** εξάγονται στοιχεία για τα χαρακτηριστικά των επισκέψεων του ιστοχώρου με σκοπό την απόκτηση γνώσης για τις πραγματικές επισκέψεις που γίνονται στον ιστοχώρο και τα ενδιαφέροντά τους. Μπορεί να μελετηθούν μεταβολές στο χρόνο σε αυτά τα χαρακτηριστικά ή ακραίες τιμές με σκοπό να εντοπιστούν πιθανά σημεία βελτίωσης του ιστοχώρου.

**Προσωποποίηση ιστοχώρου (personalisation):** επιχειρείται να καταταχθούν οι επισκέψεις με όμοια χαρακτηριστικά ώστε να είναι δυνατή η στοχευμένη προσαρμογή του ιστοχώρου ανάλογα με τις τιμές των χαρακτηριστικών σε κάθε νέα επίσκεψη.

**Χρησιμοποιησιμότητα:** επιχειρείται η εξαγωγή συμπερασμάτων για τη βελτίωση της χρησιμοποιησιμότητας με τη σύγκριση εναλλακτικών μορφών του επιπέδου διεπαφής

**Στατικά σενάρια χρήσης:** μελετάται η απόδοση της εφαρμογής με βάση στατικά σενάρια χρήσης που είχαν προβλεφθεί κατά την ανάπτυξη του ιστοχώρου.

**Δυναμικά σενάρια χρήσης:** μελετώνται σενάρια χρήσης που δεν προβλέφθηκαν κατά την ανάπτυξη του ιστοχώρου και προκύπτουν κατά την πραγματική χρήση από τους επισκέπτες. Ειδικά ενδιαφέρει εάν τα σενάρια χρήσης προκύπτουν από την αδυναμία της εφαρμογής να καθοδηγήσει τους επισκέπτες στην κάλυψη των αναγκών που έχουν.

Για κάθε μία από τις παραπάνω περιοχές ανάλυσης θα παρουσιαστεί ο συγκεκριμένος τρόπος με τον οποίο καλύπτεται από το σύστημα καταγραφής. Στο επόμενο κεφάλαιο θα περιγραφεί η εφαρμογή της μεθοδολογίας σε έναν πρότυπο ιστοχώρο τουριστικού προορισμού και θα



παρουσιαστούν τα αποτελέσματα που προέκυψαν.

## 4.1 Χαρακτηριστικά Επισκέψεων

Η ρήση ότι “ο πελάτης είναι βασιλιάς”, στον παγκόσμιο ιστό μπορεί να παραφραστεί σε “ο επισκέπτης είναι ο βασιλιάς” και οι ανάγκες του είναι αυτές που τελικά θα καθορίσουν τις υπηρεσίες που θα του προσφερθούν. Στον ανώνυμο χώρο του παγκόσμιου ιστού η έλλειψη της πρόσωπο με πρόσωπο αντιμετώπισης του πελάτη και ο διεθνισμός που επικρατεί, μιας και οι υπηρεσίες προσφέρονται σε επισκέπτες από όλο τον κόσμο, καθιστά δύσκολη την απόκτηση γνώσης για τους επισκέπτες. Για να αντιμετωπιστεί αυτό το πρόβλημα μπορούμε αντί να επιδιώξουμε να μάθουμε στοιχεία απ' ευθείας για τον κάθε επισκέπτη, να στραφούμε σε χαρακτηριστικά των επισκέψεων ώστε έμμεσα να γνωρίσουμε και τους επισκέπτες.

Οπότε από τα πρώτα θέματα που ενδιαφέρουν κατά τη λειτουργία ενός ιστοχώρου είναι τα χαρακτηριστικά των επισκέψεων. Το πρώτο όμως πρόβλημα που συναντάται όταν λειτουργεί ένας ιστοχώρος είναι ότι οι ιδιοκτήτες του δεν μπορούν να απαντήσουν ακόμα και στις απλούστερες ερωτήσεις όσον αφορά τη χρήση του ιστοχώρου από τους πραγματικούς επισκέπτες. Η ερώτηση για παράδειγμα “πόσες επισκέψεις γίνονται καθημερινά στον ιστοχώρο” δεν μπορεί να απαντηθεί. Εφαρμογές στατιστικών κίνησης ιστοχώρων (π.χ. Webalizer [48]) που συνοδεύουν πολύ συχνά τα πακέτα φιλοξενίας των παρόχων υπηρεσιών φιλοξενίας (hosting service provider) εξάγουν μεν έναν αριθμό επισκέψεων από τα αρχεία καταγραφής, αλλά για παράδειγμα προσμετρούν σε αυτές και τις επισκέψεις εφαρμογών δεικτοδότησης των μηχανών αναζήτησης που βεβαίως δεν αντιστοιχούν σε πραγματικούς χρήστες. Εάν γίνουν λίγο ποιο σύνθετες οι ερωτήσεις, όπως “από ποιες χώρες προέρχονται οι επισκέπτες και ποιες μηχανές αναζήτησης προτιμούν ανάλογα με τη χώρα προέλευσης;”, οι απλές εφαρμογές στατιστικών δε δίνουν καμία πληροφορία.

Το σύστημα που αναπτύχθηκε στα πλαίσια της εργασίας μπορεί να δώσει απαντήσεις σε πολλές ερωτήσεις που αποκαλύπτουν τα χαρακτηριστικά των επισκεπτών και τις προτιμήσεις που έχουν σε κάθε ιστοχώρο. Η παρουσίαση θα χωριστεί σε χαρακτηριστικά που αναφέρονται σε καθ' αυτή την επίσκεψη (π.χ. χώρα προέλευσης, ώρα επίσκεψης) και σε χαρακτηριστικά που αναφέρονται στις κινήσεις των επισκεπτών μέσα στον ιστοχώρο (σελίδες που βλέπουν, επιμέρους περιοχές που προτιμούν κλπ.).

### 4.1.1 Παράμετροι Επίσκεψης

Στον πίνακα 4.1 φαίνονται οι παράμετροι που καταγράφονται για κάθε επίσκεψη. Για κάθε μία από αυτές μπορεί να υποβληθούν ερωτήσεις στο σχεσιακό σχήμα της βάσης δεδομένων. Μπορεί μάλιστα να συνδυαστούν σε μία ερώτηση πολλαπλές παράμετροι και να αποκαλυφθούν έτσι

αλληλοεξαρτήσεις ανάμεσά των.

Με αναφορά στο σχεσιακό σχήμα που φαίνεται στο Σχήμα 3.3 θα δοθούν συγκεκριμένα παραδείγματα ερωτήσεων SQL στη διάλεκτο του Microsoft SQL Server.

<i>A/A</i>	<i>Παράμετροι Επισκέψεων</i>
1	Χώρα προέλευσης
2	Χρονική στιγμή/ εποχή επίσκεψης
3	Ιστοχώρος προέλευσης
4	Φράσεις αναζήτησης από εξωτερική μηχανή αναζήτησης
5	Χρόνος παραμονής στον ιστοχώρο
6	Αριθμός σελιδοπροβολών
7	Σελιδοπροβολή εισόδου
8	Σελιδοπροβολή εξόδου
9	Όνομα περιοχής χρήστη
10	Εφαρμογή πλοήγησης
11	Προτίμηση γλώσσας

*Πίνακας 4.1 Παράμετροι που μπορεί να υπολογιστούν και να αναλυθούν όσον αφορά τις επισκέψεις ενός ιστοχώρου*

Η Ερώτηση 4.1 έχει ως αποτέλεσμα τον αριθμό των επισκέψεων ανά χώρα προέλευσης (σε φθίνουσα σειρά). Από την απάντηση αυτής της ερώτησης προκύπτει η κατανομή του κοινού ανάλογα με τη χώρα προέλευσης. Ιδιαίτερα σε ιστοχώρους που απευθύνονται στο διεθνή κοινό, όπως είναι ιστοχώροι τουριστικού προορισμού, αυτή η απάντηση είναι πολύτιμη διότι επιτρέπει να αξιολογηθούν τα ακόλουθα ζητήματα:

- υπάρχουν λιγότερες επισκέψεις από κάποια χώρα από όσες αναμενόταν λόγω των πραγματικών ταξιδιών που γίνονται στον τουριστικό προορισμό? Αν ναι, πρέπει να διερευνηθεί η δεικτοδότηση του ιστοχώρου σε μηχανές αναζήτησης που χρησιμοποιούνται σε αυτές τις χώρες.
- μήπως υπάρχουν αρκετές επισκέψεις από χώρες των οποίων η γλώσσα δεν υποστηρίζεται στον ιστοχώρο; Αν ναι, τότε πρέπει να διερευνηθεί εάν πρέπει να προσφερθεί επιπλέον μετάφραση του ιστοχώρου σε αυτή τη γλώσσα.
- υπάρχει ενδιαφέρον από χώρες με τις οποίες δεν υπάρχουν επαρκείς μεταφορικές συνδέσεις (αεροπορικές κλπ.);

```

SELECT      COUNT(session_id) AS sessioncount, geoup_hc_id as country
FROM        acc_session
GROUP BY    geoup_hc_id
order by    COUNT(session_id) desc

```

*Ερώτηση 4.1 Αριθμός επισκέψεων ανά χώρα*

Οι μεταβολές στα χαρακτηριστικά των επισκέψεων στη διάρκεια του χρόνου μπορεί να δώσει ιδιαίτερα χρήσιμα στοιχεία για τις προτιμήσεις των επισκεπτών σε έναν ιστοχώρο. Ιδιαίτερα σε ιστοχώρους τουρισμού, όπου συνήθως υπάρχει μία περίοδος μέσα στο χρόνο που αναφέρεται ως τουριστική περίοδος της περιοχής και συγκεντρώνει το μεγαλύτερο μέρος των πραγματικών επισκέψεων, αυτές οι μεταβολές στις επισκέψεις στον ιστοχώρο θα ανέμενε κανείς να αντικατοπτρίζονται στη συμπεριφορά των πραγματικών επισκεπτών ενός τουριστικού προορισμού.

Η Ερώτηση 4.2 χρησιμοποιεί το πεδίο 2 του πίνακα 4.1, που αναφέρεται στη χρονική στιγμή της επίσκεψης, για να εξάγει τον αριθμό των επισκέψεων ανά ημέρα του 2003.

Εάν απομονωθούν άλλοι παράγοντες που επηρεάζουν τον αριθμό των επισκέψεων σε έναν ιστοχώρο (όπως κατάταξη σε μηχανές αναζήτησης, άλλοι σύνδεσμοι σε τρίτους ιστοχώρους, διαφήμιση κλπ.) μπορεί ο αριθμός των επισκέψεων να θεωρηθεί ως μέτρο του ενδιαφέροντος του κόσμου στο περιεχόμενο ενός ιστοχώρου. Οπότε η Ερώτηση 4.2 ουσιαστικά μας απαντάει για το ενδιαφέρον του κόσμου στη διάρκεια του χρόνου. Σε έναν ιστοχώρο τουρισμού θα έπρεπε το ενδιαφέρον, άρα και οι επισκέψεις, να αυξάνονται πριν την κορύφωση της τουριστικής περιόδου και να μειώνεται μετά από αυτήν. Έτσι μπορεί να εξαχθούν συμπεράσματα για τη χρονική περίοδο που οι ενδιαφερόμενοι επισκέπτες ξεκινάνε να προγραμματίζουν ένα ενδεχόμενο ταξίδι τους, αφού τότε θα αρχίζει να αυξάνεται το ενδιαφέρον τους.

```

select datepart( year, creation_time ) as yearcolumn,
        datepart(month, creation_time )as monthcolumn,
        datepart(day, creation_time ) as daycolumn,
        count(session_id) as sessioncount
from acc_session
where
creation_time > '20030101'
and creation_time < '20040101'
group by datepart( year, creation_time ),datepart(month, creation_time ),
datepart(day, creation_time )
order by datepart( year, creation_time ),datepart(month, creation_time ),
datepart(day, creation_time )

```

*Ερώτηση 4.2 Αριθμός επισκέψεων στη διάρκεια του χρόνου*

Πραγματικό ενδιαφέρον ωστόσο παρουσιάζει να εντοπιστούν διαφοροποιήσεις στο μέγεθος του ενδιαφέροντος ανάλογα με άλλες παραμέτρους εκτός της χρονικής περιόδου. Στην Ερώτηση 4.3

εισάγεται μία επιπλέον ομαδοποίηση ανά χώρα προέλευσης. Εξάγονται πλέον ο αριθμός των επισκέψεων ανά ημέρα του χρόνου και ανά χώρα προέλευσης.

Τα δεδομένα από την ερώτηση μπορεί να χρησιμοποιηθούν για να επιβεβαιωθούν ή να αναθεωρηθούν γνωστές από άλλες μελέτες απόψεις όσον αφορά την έναρξη προγραμματισμού ταξιδιού στο συγκεκριμένο τουριστικό προορισμό ανάλογα με τη χώρα προέλευσης των επισκεπτών.

```
SELECT      DATEPART(year, creation_time) AS yearcolumn,
            DATEPART(month, creation_time) AS monthcolumn,
            DATEPART(day, creation_time) AS daycolumn,
            geoip_hc_id AS country,
            COUNT(session_id) AS sessioncount
FROM        acc_session
WHERE       (creation_time > '20030101') AND (creation_time < '20040101')
GROUP BY   DATEPART(year, creation_time),
            DATEPART(month, creation_time),
            DATEPART(day, creation_time), geoip_hc_id
ORDER BY   DATEPART(year, creation_time),
            DATEPART(month, creation_time),
            DATEPART(day, creation_time)
            COUNT(session_id) desc
```

*Ερώτηση 4.3 Αριθμός επισκέψεων στη διάρκεια του χρόνου ανά χώρα προέλευσης*

Τα πεδία 3 και 4 του πίνακα 4.1 αναφέρονται στις διευθύνσεις προέλευσης των επισκεπτών και στις φράσεις αναζήτησης που χρησιμοποιήσαν σε περίπτωση που πρόκειται για μηχανή αναζήτησης και είναι δυνατή η εξαγωγή της φράσης. Το πλήθος των συνδέσμων από απλούς ιστοχώρους (όχι μηχανές αναζήτησης ή καταλόγους) εκφράζει ως ένα βαθμό την άποψη άλλων ιστοχώρων για τον υπό εξέταση ιστοχώρο και δίνουν μία ένδειξη για την “άπλωση” (reach) του εξεταζόμενου ιστοχώρου στο Διαδίκτυο. Με το χρόνο είναι λογικό να αυξάνονται οι σύνδεσμοι επειδή γίνεται σιγά σιγά γνωστός ένας ιστοχώρος. Το πλήθος των επισκέψεων από μηχανές αναζήτησης σε συνδυασμό με τις φράσεις που χρησιμοποιήθηκαν δίνουν μία ένδειξη για την ποιότητα του περιεχομένου όσον αφορά την ευκολία δεικτοδότησης από τις μηχανές αναζήτησης.

Σε κάθε περίπτωση είναι θεμιτό να εξεταστούν οι προελεύσεις των επισκεπτών ώστε να εξασφαλιστεί μεγαλύτερος αριθμός επισκέψεων. Υπάρχουν κυρίως τέσσερα ζητήματα που πρέπει να εξεταστούν:

1. οι σύνδεσμοι από απλούς ιστοχώρους έχουν σχέση με το περιεχόμενο που καλύπτει ο ιστοχώρος;
2. οι φράσεις σε μηχανές αναζήτησης που χρησιμοποιήθηκαν έχουν σχέση με το περιεχόμενο του ιστοχώρου;

3. υπάρχουν ιστοχώροι που προσφέρουν συνδέσμους για θέματα που καλύπτει ο υπό ανάλυση ιστοχώρος, και δεν τον αναφέρουν ;
4. υπάρχουν φράσεις αναζήτησης που αναφέρονται σε θέματα που καλύπτει ο υπό ανάλυση ιστοχώρος και δεν εμφανίζονται στις προελεύσεις των επισκεπτών;

Για να εξεταστούν τα παραπάνω ζητήματα, απαιτείται η γνώση για τις προελεύσεις σε έναν ιστοχώρο. Αυτό επιτυγχάνεται με την Ερώτηση 4.4 η οποία εξάγει τον αριθμό των επισκέψεων ανά μηχανή προέλευσης (host name).

```
SELECT      COUNT(acc_session.session_id) AS sessioncount,
            acc_referrer_site.rs_host
FROM        acc_session INNER JOIN
            acc_referrer_site ON acc_session.rs_id = acc_referrer_site.rs_id
WHERE       (acc_session.creation_time > '20030101') AND
            (acc_session.creation_time < '20040101')
GROUP BY   acc_referrer_site.rs_host
ORDER BY   COUNT(acc_session.session_id) desc
```

*Ερώτηση 4.4 Αριθμός επισκέψεων ανά μηχανή προέλευσης*

Στην Ερώτηση 4.5 εξάγονται ο αριθμός των επισκέψεων ανά φράση αναζήτησης, ενώ η Ερώτηση 4.6 εξάγει τον αριθμό επισκέψεων ανά φράση αναζήτησης και μηχανή αναζήτησης.

```
SELECT      COUNT(acc_session.session_id) AS sessioncount,
            acc_search_phrases.sph_phrase
FROM        acc_session INNER JOIN
            acc_session_search_phrase ON acc_session.session_id =
            acc_session_search_phrase.session_id INNER JOIN
            acc_search_phrases ON acc_session_search_phrase.sph_id =
            acc_search_phrases.sph_id
WHERE       (acc_session.creation_time > '20030101') AND
            (acc_session.creation_time < '20040101')
GROUP BY   acc_search_phrases.sph_phrase
ORDER BY   COUNT(acc_session.session_id) DESC
```

*Ερώτηση 4.5 Αριθμός επισκέψεων ανά φράση αναζήτησης από εξωτερικές μηχανές αναζήτησης*

```

SELECT      COUNT(acc_session.session_id) AS sessioncount,
            acc_search_phrases.sph_phrase, acc_referrer_site.rs_host
FROM        acc_session INNER JOIN acc_session_search_phrase ON
            acc_session.session_id = acc_session_search_phrase.session_id
            INNER JOIN acc_search_phrases ON
            acc_session_search_phrase.sph_id = acc_search_phrases.sph_id
            INNER JOIN acc_referrer_site ON
            acc_session.rs_id = acc_referrer_site.rs_id
WHERE       (acc_session.creation_time > '20030101') AND
            (acc_session.creation_time < '20040101')
GROUP BY   acc_search_phrases.sph_phrase, acc_referrer_site.rs_host
ORDER BY   COUNT(acc_session.session_id), acc_referrer_site.rs_host DESC

```

*Ερώτηση 4.6 Αριθμός επισκέψεων ανά φράση αναζήτησης και εξωτερικές μηχανή αναζήτησης*

Οι φράσεις αναζήτησης μπορεί να συνδυαστούν και με άλλα στοιχεία, όπως η μεταβολή τους ανάλογα με την περίοδο του χρόνου ή και με την χώρα προέλευσης ώστε να εντοπιστούν διαφοροποιήσεις είτε σε σχέση με την τουριστική περίοδο είτε με τη χώρα προέλευσης των επισκεπτών και τις συνήθειες που υπάρχουν σε κάθε χώρα. Βεβαίως μπορεί να συνδυαστούν και οι τρεις παράγοντες μαζί.

Ανάλογα με τα αποτελέσματα των ερωτήσεων από τις προελεύσεις των επισκεπτών πρέπει να γίνει τροποποίηση του ιστοχώρου ώστε να καλύπτεται καλύτερα από μηχανές αναζήτησης (εξαρτάται από τον τρόπο δεικτοδότησης της κάθε μηχανής αναζήτησης), είτε να επιχειρηθεί συνεργασία με άλλους ιστοχώρους ώστε να γίνει προσθήκη συνδέσμου προς τον υπό εξέταση ιστοχώρο.

Τα πεδία 5 και 6 του πίνακα 4.1 αναφέρονται στο χρόνο παραμονής ενός επισκέπτη στον ιστοχώρο και στον αριθμό σελίδων που επισκέφτηκε. Υπάρχει σαφή εξάρτηση ανάμεσα σε αυτά τα δύο μεγέθη (όσες πιο πολλές σελίδες βλέπει ένας επισκέπτης, τόσο πιο πολύ ώρα θα χρειαστεί), αλλά εξαρτάται και από τη φύση του κάθε ιστοχώρου (σε έναν ιστοχώρο που προσφέρει ειδήσεις, θα ανέμενε κανείς ότι υπάρχουν πολλές επισκέψεις με μία μόνο σελιδοπροβολή και μικρής διάρκειας επειδή θα διαπίστωναν απλώς ότι δεν υπάρχουν νέες ειδήσεις.).

Και τα δύο μεγέθη δίνουν μία ένδειξη για το βαθμό που ο ιστοχώρος ενδιαφέρει τελικά τους επισκέπτες. Σε ιστοχώρους όπου από τη φύση τους απαιτούνται αρκετές σελιδοπροβολές για να εκτελεστεί μία εργασία, θα ανέμενε κανείς ότι οι επισκέψεις διαρκούν αρκετή ώρα. Εάν το ποσοστό των επισκέψεων στις οποίες γίνεται μία μόνο σελιδοπροβολή είναι μεγάλο, τότε πρέπει να διερευνηθούν οι λόγοι για τους οποίους αυτό συμβαίνει. Ένας τρόπος διερεύνησης είναι να εξεταστεί η φράση αναζήτησης που χρησιμοποιήθηκε από εξωτερικές μηχανές αναζήτησης και να εκτιμηθεί εάν το περιεχόμενο της φράσης καλύπτεται από τον υπό εξέταση ιστοχώρο. Εάν δεν καλύπτεται, τότε μπορεί η μία και μοναδική σελιδοπροβολή να θεωρηθεί φυσιολογική (ο επισκέπτης καταλαβαίνει ότι δεν μπορεί να βρει αυτό που ψάχνει και εγκαταλείπει). Εάν όμως ο

ιστοχώρος καλύπτει το θέμα της φράσης, πρέπει να γίνουν περαιτέρω έρευνες. Μπορεί να εξεταστεί η σελίδα εισόδου της επίσκεψης (πεδίο 7 του πίνακα 4.1) για να διαπιστωθεί εάν η πρώτη (και μοναδική) σελίδα της επίσκεψης έχει κάποιο πρόβλημα. Το πρόβλημα μπορεί να έγκειται για παράδειγμα σε πρόβλημα χρησιμοποίησιμότητας ή απλώς λανθασμένης λειτουργίας του λογισμικού. Μπορεί επίσης να αντιπαρατεθεί η φράση αναζήτησης που χρησιμοποιήθηκε από μηχανές αναζήτησης με το περιεχόμενο της πρώτης σελίδας. Εάν πρόκειται για άσχετη σελίδα σε σχέση με τη φράση, τότε πρέπει πάλι να τροποποιηθούν οι σελίδες ώστε να γίνει καλύτερη δεικτοδότηση από τις μηχανές αναζήτησης.

Στην Ερώτηση 4.7 υπολογίζεται ο αριθμός των επισκέψεων με μία σελιδοπροβολή. Με όμοιο τρόπο (αλλάζοντας τον περιορισμό για το πεδίο `impression_count`) υπολογίζονται και οι επισκέψεις με διαφορετικό αριθμό ή εύρος τιμών στις σελιδοπροβολές.

```
SELECT      COUNT(acc_session.session_id) AS sessioncount
FROM        acc_session
WHERE       (acc_session.creation_time > '20030101') AND
           (acc_session.creation_time < '20040101')
AND        impression_count = 1
```

*Ερώτηση 4.7 Αριθμός επισκέψεων με μία σελιδοπροβολή*

Το πεδίο 9 του πίνακα 4.1 αναφέρεται στο όνομα περιοχής του επισκέπτη (domain name) και προκύπτει από το Reverse DNS Lookup που προσφέρει το σύστημα DNS. Δηλαδή δίνεται μία διεύθυνση IP και επιστρέφεται το όνομα περιοχής που αντιστοιχεί σε αυτήν, εάν έχει γίνει σχετική ρύθμιση. Αυτό το πεδίο παρουσιάζει ειδικό ενδιαφέρον διότι επιτρέπει την εξαγωγή συμπερασμάτων για δύο στοιχεία:

1. οργανισμός στον οποίο εργάζεται/ βρίσκεται ο επισκέπτης εάν πρόκειται για όνομα περιοχής οργανισμού (π.χ. `tuc.gr` για το Πολυτεχνείο Κρήτης)
2. Γεωγραφική περιοχή που βρίσκεται ο επισκέπτης εάν πρόκειται για χρήστη DialUp και ο πάροχος κωδικοποιεί στο όνομα περιοχής την τοποθεσία του δικτύου του.

Για τη μεν πρώτη περίπτωση σε έναν ιστοχώρο τουριστικού προορισμού μπορεί να εξακριβωθεί κατά πόσο ενδιαφέρει μεγάλους οργανισμούς που κινούνται γενικότερα στο χώρο του τουρισμού ή στη συγκεκριμένη περιοχή που καλύπτει ο ιστοχώρος. Η δεύτερη παράμετρος επιτρέπει μία χονδρική εκτίμηση του ενδιαφέροντος των επισκεπτών ανάλογα με συγκεκριμένες γεωγραφικές τοποθεσίες.

Η Ερώτηση 4.8 δίνει τον αριθμό των επισκέψεων ανά όνομα περιοχής

```
SELECT      COUNT(acc_session.session_id) AS sessioncount,
            acc_host_domains.hd_domain
FROM        acc_session INNER JOIN
            acc_host_domains ON acc_session.hd_id = acc_host_domains.hd_id
WHERE       (acc_session.creation_time > '20030101') AND
            (acc_session.creation_time < '20040101')
GROUP BY   acc_host_domains.hd_domain
ORDER BY   COUNT(acc_session.session_id) DESC
```

*Ερώτηση 4.8 Αριθμός επισκέψεων ανά όνομα περιοχής επισκέψεων*

Το πεδίο 11 τέλος, αναφέρεται στην εφαρμογή πλοήγησης που χρησιμοποίησαν οι επισκέπτες κατά την πρόσβαση στον ιστοχώρο. Αν και τα τελευταία οι εφαρμογές πλοήγησης χρησιμοποιούν σε μεγαλύτερο βαθμό διεθνή πρότυπα για την περιγραφή των σελίδων, εντούτοις υπάρχουν ακόμα σημαντικά σημεία που ερμηνεύονται διαφορετικά ανάλογα με την εφαρμογή, με αποτέλεσμα είτε να μην είναι δυνατή η εμφάνιση των σελίδων ανάλογα με την εφαρμογή του επισκέπτη, είτε να λειτουργούν με λιγότερη λειτουργικότητα. Είναι οπότε χρήσιμο να γνωρίζουμε ποιες εφαρμογές πλοήγησης χρησιμοποιούν οι επισκέπτες ώστε να εξακριβωθεί εάν οι σελίδες του ιστοχώρου παρουσιάζονται ικανοποιητικά ανεξάρτητα με την εφαρμογή που χρησιμοποιείται. Εάν διαπιστωθεί ότι οι χρήστες χρησιμοποιούν σε μεγάλο ποσοστό μία εφαρμογή πλοήγησης η οποία δεν είναι ικανή να παρουσιάσει τις σελίδες ή μέρος των σελίδων, πρέπει να γίνει είτε σχετική τροποποίηση των σελίδων είτε να δοθούν σε αυτούς τους χρήστες εναλλακτικές σελίδες που να λειτουργούν ικανοποιητικά. Αυτό μπορεί εύκολα να γίνει ελέγχοντας με κάθε πρόσβαση στον ιστοχώρο την εφαρμογή πλοήγησης του χρήστη.

Συχνά σε συστήματα στατιστικών ιστοχώρων τα ποσοστά ανά εφαρμογή πλοήγησης δίνονται ως ποσοστά ως προς τις αιτήσεις προς τον ιστοχώρο. Επειδή όμως στην περίπτωση που μία εφαρμογή δεν μπορεί να χρησιμοποιηθεί σε έναν ιστοχώρο ο ανάλογος χρήστης θα τον εγκαταλείψει, οι αιτήσεις από μη συμβατές εφαρμογές φαίνονται με αυτόν τον τρόπο λίγες. Οπότε αυτό που πραγματικά ενδιαφέρει είναι πόσες επισκέψεις (και όχι πόσες αιτήσεις) γίνονται στον ιστοχώρο με προβληματικές εφαρμογές πλοήγησης.

Η ερώτηση που δίνει την κατανομή των εφαρμογών πλοήγησης φαίνεται στην Ερώτηση 4.9. Και σε αυτήν την περίπτωση είναι σημαντικό να συμπεριληφθεί ο παράγοντας του χρόνου για την εξαγωγή συμπερασμάτων, διότι οι χρήστες τείνουν να αλλάζουν εφαρμογές πλοήγησης με το χρόνο με αποτέλεσμα να μεταβάλλονται και οι δυνατότητες των εφαρμογών που χρησιμοποιούν.



```

SELECT      COUNT(acc_session.session_id) AS sessioncount, acc_browser.name,
acc_browser.opsys, acc_browser.version
FROM        acc_session INNER JOIN
acc_browser ON acc_session.browser_id = acc_browser.browser_id
WHERE
      (acc_session.creation_time > '20030101')
      AND (acc_session.creation_time < '20040101')
GROUP BY   acc_browser.name, acc_browser.opsys, acc_browser.version
ORDER BY   COUNT(acc_session.session_id) DESC

```

*Ερώτηση 4.9 Κατανομή εφαρμογών πλοήγησης σε επισκέψεις*

## 4.1.2 Παράμετροι Εντός Ιστοχώρου

Στην προηγούμενη ενότητα παρουσιάστηκαν παράμετροι που πρέπει να εξεταστούν για τους επισκέπτες χωρίς να συμπεριλαμβάνονται σε αυτές στοιχεία που προκύπτουν από την κίνηση των επισκεπτών μέσα στον υπό ανάλυση ιστοχώρο.

Μετά την είσοδο ενός επισκέπτη στον ιστοχώρο μπορεί να εξαχθούν συμπεράσματα κυρίως από τις σελίδες που επισκέπτεται. Γνωρίζοντας το περιεχόμενο των σελίδων (με τη σύνδεση σε πηγές δεδομένων από τις οποίες συνθέτουμε τις σελίδες και τη γνώση για τη δομή του ιστοχώρου) μπορούμε να εξάγουμε συμπεράσματα για τα ενδιαφέροντα των επισκεπτών.

Στην ανάλυση χρησιμοποιείται κυρίως η έννοια της Γενικής Σελίδας και της Σύνθετης Σελίδας που περιγράφηκε στο κεφάλαιο 3. Με αυτές περιγράφονται πρότυπα (templates) σελίδων που το περιεχόμενο της κάθε συγκεκριμένης σελιδοπροβολής περιγράφεται με δυνάδες από ενέργειες και συνδέσμους σε αντικείμενα εξωτερικών πηγών δεδομένων (βλέπε ενότητα 3.2 σελίδα 126).

Στη σχεσιακή βάση δεδομένων η παραπάνω έννοια αντικατοπτρίζεται στη σχέση `acc_page_impression` που αναφέρεται σε κάθε σελιδοπροβολή και που συνδέεται με τη σχέση `acc_action_perform` που συσχετίζει με τη σελιδοπροβολή ένα σύνολο ενεργειών και τα αντικείμενα στα οποία αναφέρονται (βλέπε Σχήμα 3.3 για το φυσικό μοντέλο της βάσης δεδομένων).

Στα πλαίσια της απόκτησης γνώσης γενικά για τους επισκέπτες που εξετάζεται σε αυτήν την ενότητα, ενδιαφέρει κυρίως να εντοπιστούν μεταβολές που υπάρχουν στις προτιμήσεις των επισκεπτών ανάλογα με άλλα χαρακτηριστικά τους. Για τις προτιμήσεις των επισκεπτών χρησιμοποιούνται οι Γενικές Σελίδες που εκφράζουν το τμήμα του ιστοχώρου που ενδιαφέρει τους επισκέπτες καθώς και οι κατηγορίες στις οποίες μπορεί να καταταχθούν τα αντικείμενα της κάθε σελιδοπροβολής ανάλογα με τον ιστοχώρο. Οι παράμετροι που εξετάζονται αναφέρονται κυρίως στις παράμετρους της επίσκεψης όπως παρουσιάστηκαν στην αμέσως προηγούμενη ενότητα (χώρα προέλευσης, φράσεις αναζήτησης κλπ.).

Σε έναν ιστοχώρο τουρισμού που καλύπτει τη θεματολογία όπως αυτή παρουσιάζεται στο

Παράρτημα Α, ενδιαφέρουν οι προτιμήσεις των επισκεπτών σε συγκεκριμένες περιοχές ενδιαφέροντος και κατηγορίες επιχειρήσεων όπως αυτές μεταβάλλονται στη διάρκεια του χρόνου ή ανάλογα με τη χώρα προέλευσης των επισκεπτών.

```
SELECT      COUNT(pageimpression_id) AS sessioncount, acc_pages.name
FROM        acc_session INNER JOIN acc_page_impressions ON
           acc_session.session_id = acc_page_impressions.session_id
           INNER JOIN acc_pages ON
           acc_page_impressions.page_id = acc_pages.page_id
           WHERE (acc_session.creation_time > '20030101') AND
           (acc_session.creation_time < '20040101')
GROUP BY acc_pages.name
ORDER BY COUNT(pageimpression_id) DESC
```

*Ερώτηση 4.10 Αριθμός σελιδοπροβολών ανά Γενική Σελίδα*

Η Ερώτηση 4.10 εξάγει τον αριθμό των σελιδοπροβολών ανά Γενική Σελίδα για μία χρονική περίοδο. Από τα αποτελέσματα αυτής της ερώτησης προκύπτει το ενδιαφέρον που δείχνουν οι επισκέπτες στα διάφορα τμήματα του ιστοχώρου. Δίνεται έτσι μία ένδειξη για ποια σημεία του ιστοχώρου πρέπει να επεκταθούν σε λειτουργικότητα ή περιεχόμενο.

Μπορεί όμως το ενδιαφέρον να μεταβάλλεται με το χρόνο. Σε έναν ιστοχώρο τουριστικού προορισμού για παράδειγμα το ενδιαφέρον για επιχειρήσεις τουρισμού ως ποσοστό επί του συνόλου των σελιδοπροβολών, θα είναι μεγαλύτερο την περίοδο πριν την τουριστική περίοδο επειδή τότε οι επισκέπτες προγραμματίζουν τις διακοπές τους. Η επόμενη ερώτηση (Ερώτηση 4.11) δίνει τον αριθμό των σελιδοπροβολών ανά σελίδα στη διάρκεια του χρόνου, από την οποία μπορεί να εξαχθούν οι μεταβολές στην κατανομή ανάλογα με τις σελίδες.

```
SELECT      DATEPART(year, acc_session.creation_time) AS yearcolumn,
           DATEPART(month, acc_session.creation_time) AS monthcolumn,
           COUNT(acc_session.session_id) AS sessioncount,
           acc_pages.name
FROM        acc_session INNER JOIN acc_page_impressions ON
           acc_session.session_id = acc_page_impressions.session_id
           INNER JOIN acc_pages ON
           acc_page_impressions.page_id = acc_pages.page_id
WHERE       (acc_session.creation_time > '20030101') AND
           (acc_session.creation_time < '20040101')
GROUP BY acc_pages.name
ORDER BY DATEPART(year, acc_session.creation_time), DATEPART(month,
acc_session.creation_time), COUNT(acc_session.session_id) DESC
```

*Ερώτηση 4.11 Αριθμός σελιδοπροβολών ανά Γενική Σελίδα ανά μήνα*

Η προηγούμενη ερώτηση μπορεί να επεκταθεί ώστε να λαμβάνεται υπ' όψιν είτε η γλώσσα εμφάνισης στον ιστοχώρο της κάθε σελίδας (σε πολυγλωσσικούς ιστοχώρους) είτε η χώρα προέλευσης των επισκεπτών. Η Ερώτηση 4.12 εξάγει τον αριθμό των σελιδοπροβολών ανά Γενική Σελίδα του ιστοχώρου και ανά γλώσσα εμφάνισης στη διάρκεια του χρόνου.

```

SELECT    DATEPART(year, acc_session.creation_time) AS yearcolumn,
          DATEPART(month, acc_session.creation_time) AS monthcolumn,
          COUNT(acc_session.session_id) AS sessioncount,
          acc_pages.name,
          acc_page_impressions.language_viewed
FROM      acc_session INNER JOIN acc_page_impressions ON
          acc_session.session_id = acc_page_impressions.session_id
          INNER JOIN acc_pages ON
          acc_page_impressions.page_id = acc_pages.page_id
WHERE     (acc_session.creation_time > '20030101') AND
          (acc_session.creation_time < '20040101')
GROUP BY acc_pages.name, acc_page_impressions.language_viewed
ORDER BY DATEPART(year, acc_session.creation_time), DATEPART(month,
acc_session.creation_time), COUNT(acc_session.session_id) DESC

```

*Ερώτηση 4.12 Αριθμός σελιδοπροβολών ανά Γενική Σελίδα και ανά γλώσσα εμφάνισης ανά μήνα*

Η έννοια της Γενικής Σελίδας ομαδοποιεί τις σελιδοπροβολές σε ένα πρότυπο σελίδας (template) αλλά δε μας δίνει τυχόν διαφοροποιήσεις που μπορεί να υπάρχουν ανάλογα με το περιεχόμενο της κάθε σελίδας. Σε έναν ιστοχώρο τουριστικού προορισμού είναι αναμενόμενο οι προτιμήσεις των επισκεπτών να διαφοροποιούνται ανάλογα με το βαθμό που μία περιοχή είναι τουριστικά αναπτυγμένη. Εάν πάρουμε ως παράδειγμα τη συλλογή δεδομένων για τον ιστοχώρο τουριστικού προορισμού όπως φαίνεται στον πίνακα 3.3 σελ. 134 τότε από το αναγνωριστικό της σελίδας για περιοχές ενδιαφέροντος (ar) και την ενέργεια που υποδηλώνει την παρουσίαση περιοχής ενδιαφέροντος (View Area), μπορούμε με την Ερώτηση 4.13 να εξάγουμε τον αριθμό των σελιδοπροβολών ανά περιοχή ενδιαφέροντος.

```

SELECT    COUNT(acc_session.session_id) AS sessioncount,
          acc_actions_perform.entity_id
FROM      acc_session INNER JOIN
          acc_page_impressions ON acc_session.session_id =
          acc_page_impressions.session_id INNER JOIN
          acc_actions_perform ON acc_page_impressions.pageImpression_id =
          acc_actions_perform.pageimpression_id INNER JOIN
          acc_actions ON acc_actions_perform.action_id =
          acc_actions.action_id INNER JOIN
          acc_pages ON acc_page_impressions.page_id = acc_pages.page_id
WHERE     (acc_session.creation_time > '20030101') AND
          (acc_session.creation_time < '20040101') AND
          (acc_pages.name = 'ar') AND
          (acc_actions.description = 'View Area')
GROUP BY acc_actions_perform.entity_id
ORDER BY COUNT(acc_session.session_id) DESC

```

*Ερώτηση 4.13 Αριθμός σελιδοπροβολών ανά Γενική Σελίδα και ανά γλώσσα εμφάνισης ανά μήνα*

Από μόνη της η παραπάνω ερώτηση δίνει μία γενική εικόνα για τις περιοχές για τις οποίες ενδιαφέρονται οι επισκέπτες, αλλά το πραγματικό ενδιαφέρον υπάρχει εάν μπορεί να εντοπιστούν περιοχές που έχουν μεγάλες διαφορές στις προτιμήσεις ανάλογα με άλλες παραμέτρους που χαρακτηρίζουν τις επισκέψεις. Τέτοιες μπορεί να είναι η χώρα προέλευσης, η γλώσσα εμφάνισης του περιεχόμενου και η χρονική περίοδος των επισκέψεων.

Η Ερώτηση 4.14 για παράδειγμα (με μικρή τροποποίηση της προηγούμενης) εξάγει τον αριθμό σελιδοπροβολών ανά γλώσσα εμφάνισης του περιεχομένου.

```
SELECT      COUNT(acc_session.session_id) AS sessioncount,
            language_viewed
FROM        acc_actions_perform.entity_id
FROM        acc_session INNER JOIN
            acc_page_impressions ON acc_session.session_id =
            acc_page_impressions.session_id INNER JOIN
            acc_actions_perform ON acc_page_impressions.pageImpression_id =
            acc_actions_perform.pageimpression_id INNER JOIN
            acc_actions ON acc_actions_perform.action_id =
            acc_actions.action_id INNER JOIN
            acc_pages ON acc_page_impressions.page_id = acc_pages.page_id
WHERE       (acc_session.creation_time > '20030101') AND
            (acc_session.creation_time < '20040101') AND
            (acc_pages.name = 'ar') AND
            (acc_actions.description = 'View Area')
GROUP BY   acc_actions_perform.entity_id, language_viewed
ORDER BY   COUNT(acc_session.session_id) DESC
```

*Ερώτηση 4.14 Αριθμός σελιδοπροβολών ανά Γενική Σελίδα και ανά γλώσσα εμφάνισης ανά μήνα*

## 4.2 Προσωποποίηση Ιστοχώρου

Η προσωποποίηση του ιστοχώρου αναφέρεται στην προσαρμογή του ιστοχώρου με τρόπο που να λαμβάνει υπόψιν την κατάταξη της κάθε επίσκεψης σε ομάδες επισκέψεων με όμοια χαρακτηριστικά. Γνωρίζοντας δηλαδή σε ποια ομάδα ή ομάδες επισκέψεων ανήκει μία συγκεκριμένη επίσκεψη, προσπαθούμε να τροποποιήσουμε τον ιστοχώρο ώστε να καλύπτει τις ανάγκες των επισκεπτών αυτής της ομάδας.

Εάν γνωρίζουμε για παράδειγμα ότι ο επισκέπτης που μόλις εισήλθε στον ιστοχώρο ενδιαφέρεται για πληροφορίες για συγκεκριμένη περιοχή ενδιαφέροντος, μπορεί να προσαρμοστεί ο ιστοχώρος ώστε να εμφανίζεται σε προβεβλημένη θέση απ' ευθείας σύνδεσμος προς τη σελίδα παρουσίασης της περιοχής, ώστε να μπορεί πιο γρήγορα να φτάσει ο επισκέπτης στον προορισμό του.

Η ομαδοποίηση των επισκέψεων μπορεί να γίνει με βάση τις παραμέτρους που παρουσιάστηκαν στην προηγούμενη ενότητα (χώρα προέλευσης, φράση αναζήτησης, χρονική περίοδος επίσκεψης κλπ.), αρκεί να εντοπιστούν σημαντικές διαφοροποιήσεις στα μεγέθη που παρουσιάζονται (π.χ. αριθμός επισκέψεων) ανάλογα με τις παραμέτρους. Τα μεγέθη είναι κυρίως ο αριθμός των επισκέψεων και ο αριθμός των σελιδοπροβολών.

Γενικά μπορεί να ακολουθηθεί η εξής μέθοδος:

1. Για κάθε μέγεθος (αριθμός σελιδοπροβολών και αριθμός επισκέψεων) εξάγονται οι τιμές του

μεγέθους σε συνάρτηση με τις τιμές των παραμέτρων της επίσκεψης.

2. Για τις περιπτώσεις όπου μία συγκεκριμένη τιμή μίας παραμέτρου ή εύρος τιμών μίας παραμέτρου ή συγκεκριμένος συνδυασμός τιμών περισσότερων παραμέτρων, έχει ως αποτέλεσμα σημαντική διαφοροποίηση στην τιμή του μεγέθους που εξετάζεται, δημιουργείται μία ομάδα επισκέψεων.
3. Για κάθε ομάδα επισκέψεων ερευνάται η πιθανή προσαρμογή του ιστοχώρου για νέες επισκέψεις που έχουν τις ίδιες τιμές στις παραμέτρους της επίσκεψης με αυτές της ομάδας.

Στην Ερώτηση 4.14 της προηγούμενης ενότητας για παράδειγμα υπολογίζεται ο αριθμός των σελιδοπροβολών ανά περιοχή ενδιαφέροντος σε ιστοχώρο τουρισμού ανάλογα με τη γλώσσα εμφάνισης του περιεχομένου. Έστω για ένα σύνολο περιοχών  $S$  παρατηρείται ότι οι σελιδοπροβολές σε αυτές είναι πολύ περισσότερες από τις σελιδοπροβολές στις υπόλοιπες περιοχές σε μία συγκεκριμένη γλώσσα **lang**. Ανάλογα με την πολιτική που ακολουθεί ένας ιστοχώρος μπορεί να προβεί στις ακόλουθες προσαρμογές:

Όταν γίνεται νέα επίσκεψη με γλώσσα **lang** τότε

- εάν είναι επιθυμητό η διευκόλυνση των επισκεπτών, εμφάνισε σε ορισμένο σημείο σε κάθε σελίδα του ιστοχώρου λίστα συνδέσμων από το σύνολο  $S$  ώστε οι επισκέπτες να έχουν άμεση πρόσβαση στην πληροφορία για αυτές τις περιοχές
- εάν είναι επιθυμητό να προβληθούν και άλλες περιοχές εκτός των  $S$ , εμφάνισε σε ορισμένο σημείο σε κάθε σελίδα του ιστοχώρου λίστα συνδέσμων που δεν ανήκουν στο  $S$  ώστε να διευρυνθούν οι περιοχές που επισκέπτονται οι χρήστες σε αυτή τη γλώσσα

Το προηγούμενο παράδειγμα έχει ιδιαίτερη εφαρμογή σε ιστοχώρους τουριστικού προορισμού επειδή συνήθως υπάρχουν περιοχές με μεγάλη τουριστική ανάπτυξη και είναι έτσι λογικό οι επισκέπτες να επιθυμούν να πληροφορηθούν για αυτές τις περιοχές. Αυτές οι περιοχές μπορεί να διαφοροποιούνται ανάλογα με τη χώρα προέλευσης των επισκεπτών. Εάν ο στόχος του ιστοχώρου είναι να διευκολύνει τους επισκέπτες, τότε θα δοθεί άμεση πρόσβαση σε πληροφορία για τις περιοχές. Εάν ο στόχος του ιστοχώρου όμως είναι να προωθηθούν και λιγότερο προβεβλημένες περιοχές τότε θα δώσει έμφαση σε αυτές τις περιοχές ώστε να κινήσει το ενδιαφέρον των τουριστών.

Πέραν των μεμονωμένων ερωτήσεων σε SQL στο σύστημα, μπορεί να χρησιμοποιηθούν και κύβου OLAP για τον εντοπισμό των ομάδων για τις οποίες θα γίνει προσαρμογή. Οι διαστάσεις των

κύβων θα είναι σε αυτή την περίπτωση οι παράμετροι των επισκέψεων, ενώ οι μετρικές τα μεγέθη αριθμός σελιδοπροβολών και επισκέψεων (βλέπε σχετικά ενότητα 2.3.7.2 σελίδα 105 περί ανάλυσης δεδομένων χρήσης με τεχνικές OLAP).

Στις προηγούμενες περιπτώσεις η ομαδοποίηση γίνεται με βάση τα παρελθόντα στοιχεία σε βάθος χρόνου. Μπορεί όμως να υπάρχουν και απότομες αλλαγές στα μεγέθη για τις οποίες είναι επιθυμητό να γίνει προσαρμογή σε σύντομο χρόνο. Για να εντοπιστούν αυτές οι αλλαγές απαιτείται η συνεχής παρακολούθηση των δεδομένων χρήσης ώστε να ενεργοποιούνται σχετικές λειτουργίες trigger που να ειδοποιούν για τις απότομες αλλαγές. Οι αναλυτές μελετώντας τις αλλαγές, είναι σε θέση να προβούν άμεσα σε προτάσεις για τροποποιήσεις στον ιστοχώρο ανάλογα με το είδος της αλλαγής.

Τέτοιου είδους απότομες αλλαγές προέρχονται κυρίως από αλλαγές στις προελεύσεις των επισκεπτών, με δύο κυρίως διαφοροποιήσεις:

1. μία μηχανή αναζήτησης μπορεί να πριμοδοτήσει τον ιστοχώρο για συγκεκριμένη φράση αναζήτησης με αποτέλεσμα να υπάρχουν απότομα πολλές επισκέψεις με αυτή τη φράση (για παράδειγμα επειδή μόλις πρόσθεσε τη σελίδα του ιστοχώρου στη βάση δεδομένων της, ή επειδή άλλαξε αλγόριθμο υπολογισμού της κατάταξης που χρησιμοποιεί)
2. ένας δημοφιλής τρίτος ιστοχώρος προσθέτει σύνδεσμο προς τον υπό ανάλυση ιστοχώρο οπότε υπάρχουν απότομα πολλές επισκέψεις από αυτόν. Όμοια συμπεριφορά υπάρχει εάν ένας τρίτος ιστοχώρος έχει για περιορισμένο χρονικό διάστημα πολλές επισκέψεις, και υπάρχει σε αυτόν σύνδεσμος προς τον υπό ανάλυση ιστοχώρο.

Στην πρώτη περίπτωση η απότομη αλλαγή αναφέρεται σε εξωτερικές φράσεις αναζήτησης, ενώ στη δεύτερη σε συγκεκριμένο URL προέλευσης. Το μέγεθος που αλλάζει είναι ο αριθμός των επισκέψεων.

Η περίπτωση των φράσεων αναζήτησης δεν παρουσιάζει ιδιαίτερο ενδιαφέρον, γιατί υποτίθεται ότι η σελίδα στην οποία οδηγούνται οι επισκέπτες είναι σχετική με τη φράση στην οποία έψαχναν.

Η δεύτερη περίπτωση όμως αποκτάει ενδιαφέρον επειδή ο εξωτερικός ιστοχώρος μπορεί να αναφέρεται σε πολύ συγκεκριμένη θεματολογία που καλύπτεται μεν από τον υπό ανάλυση ιστοχώρο αλλά μόνο ως τμήμα ευρύτερης πληροφόρησης που παρέχει. Οπότε είναι θεμιτό να εξεταστεί η κάθε τέτοια απότομη αλλαγή και να διερευνηθεί μήπως για τους επισκέπτες από αυτούς τους τρίτους ιστοχώρους πρέπει να γίνει ειδική προσαρμογή.

Η Ερώτηση 4.15 υπολογίζει εκείνες τις διευθύνσεις προέλευσης για τις οποίες υπάρχει αλλαγή μεγαλύτερη από 100% μεταξύ 2 ημερών και τουλάχιστον 10 επισκέψεις. Η ερώτηση λαμβάνει υπ' όψιν μόνο υπάρχουσες προελεύσεις.

```
select r2.thedate,
       r2.referring_url,
       r2.countbyurl,
       cast (cast (r2.countbyurl as float) / r1.countbyurl -1.0
            as decimal(5,3)) as metabolh
from
(
SELECT
    cast(cast(datepart(year, creation_time) AS char(4)) + '/'
    + cast(datepart(month, creation_time) AS varchar(2)) + '/'
    + cast(datepart(day, creation_time) AS varchar(2)) AS datetime)
    AS thedate,
    referring_url,
    COUNT(acc_session.session_id) AS countbyurl
FROM acc_session inner join acc_session_referrerringUrl on
acc_session.session_id = acc_session_referrerringUrl.session_id
WHERE rs_id not in (select rs_id from v_search_engine_hosts)
GROUP BY cast(cast(datepart(year, creation_time) AS char(4)) + '/'
+ cast(datepart(month, creation_time) AS varchar(2)) + '/' +
cast(datepart(day, creation_time) AS varchar(2)) AS datetime),
referring_url
) r1
right join
(
SELECT
    cast(cast(datepart(year, creation_time) AS char(4)) + '/'
    + cast(datepart(month, creation_time) AS varchar(2)) + '/'
    + cast(datepart(day, creation_time) AS varchar(2)) AS datetime)
    AS thedate,
    referring_url,
    COUNT(acc_session.session_id) AS countbyurl
FROM acc_session inner join acc_session_referrerringUrl on
acc_session.session_id = acc_session_referrerringUrl.session_id
WHERE rs_id not in (select rs_id from v_search_engine_hosts)
GROUP BY cast(cast(datepart(year, creation_time) AS char(4)) + '/'
+ cast(datepart(month, creation_time) AS varchar(2)) + '/' +
cast(datepart(day, creation_time) AS varchar(2)) AS datetime),
referring_url
) r2
on DATEADD ( dd , 1, r1.thedate ) = r2.thedate
and r1.referring_url = r2.referring_url

where r2.countbyurl > 10          -- second day more than 10
and cast (cast (r2.countbyurl as float) / r1.countbyurl -1.0 as decimal
(5,3)) > 1.0 -- change > 100%
order by r1.thedate
```

*Ερώτηση 4.15 Απότομη μεταβολή υπαρχόντων προελεύσεων*

Στην Ερώτηση 4.16 υπολογίζεται ο αριθμός επισκέψεων σε εβδομαδιαία βάση ανά προέλευση, αλλά εξάγονται μόνο οι επισκέψεις όταν πρωτοεμφανίζεται μία νέα προέλευση. Ελέγχοντας τα αποτελέσματα της ερώτησης μπορεί να εντοπιστούν έτσι νέες προελεύσεις και να απομονωθούν αυτές που προσθέτουν σημαντικό μερίδιο στο σύνολο των επισκέψεων. Για αυτές, ανάλογα με το περιεχόμενο της σελίδας προέλευσης μπορεί να εξεταστεί εάν πρέπει να δοθεί σε αυτούς τους

επισκέπτες προσαρμοσμένη μορφή του ιστοχώρου που εξετάζεται.

```
SELECT distinct r1.yearc, r1.weekc, r1.referring_url, r2.visitcount
from
(
SELECT      MIN(DATEPART(year, acc_session.creation_time)) AS yearc,
            MIN(DATEPART(week, acc_session.creation_time)) AS weekc,
            acc_session_referrerringUrl.referring_url
FROM        acc_session INNER JOIN
            acc_session_referrerringUrl ON acc_session.session_id =
            acc_session_referrerringUrl.session_id
WHERE       (acc_session.creation_time > '20030901')
GROUP BY   acc_session_referrerringUrl.referring_url,
            DATEPART(year, acc_session.creation_time),
            DATEPART(week, acc_session.creation_time)
) r1
inner join
(
SELECT      DATEPART(year, acc_session.creation_time) AS yearc,
            DATEPART(week, acc_session.creation_time) AS weekc,
            acc_session_referrerringUrl.referring_url,
            COUNT(acc_session.session_id) AS visitcount
FROM        acc_session INNER JOIN
            acc_session_referrerringUrl ON acc_session.session_id =
            acc_session_referrerringUrl.session_id
WHERE       (acc_session.creation_time > '20030901')
GROUP BY   acc_session_referrerringUrl.referring_url,
            DATEPART(year, acc_session.creation_time),
            DATEPART(week, acc_session.creation_time)
) r2
on r1.yearc=r2.yearc and r1.weekc=r2.weekc
order by r2.visitcount desc
```

*Ερώτηση 4.16 Εντοπισμός νέων προελεύσεων ανά εβδομάδα*

### 4.3 Χρησιμοποιησιμότητα

Οι προηγούμενες ενότητες περιέγραψαν το πως μπορούμε να γνωρίσουμε καλύτερα τους επισκέπτες ενός ιστοχώρου. Γνωρίζοντας τους επισκέπτες και τα ενδιαφέροντα που έχουν, έχουμε ήδη σημαντικά στοιχεία με βάση τα οποία μπορεί να αναπτυχθεί ένας ιστοχώρος που καλύπτει περισσότερο τις ανάγκες των επισκεπτών του όσον αφορά τη χρησιμοποιησιμότητα. Με την προσωποποίηση του ιστοχώρου στις ανάγκες ειδικών ομάδων επισκεπτών αυξάνεται επίσης η χρησιμοποιησιμότητα.

Σε αυτή την ενότητα θα παρουσιαστούν συγκεκριμένες μέθοδοι ανάλυσης των δεδομένων χρήσης που στοχεύουν απ' ευθείας στη βελτίωση της χρησιμοποιησιμότητας. Η μέθοδοι επικεντρώνονται στη σύγκριση εναλλακτικών μορφών του επιπέδου διεπαφής (interface) ενός ιστοχώρου. Τέτοιες συγκρίσεις είναι σύνηθες να γίνονται στα αρχικά στάδια ανάπτυξης μίας παραδοσιακής εφαρμογής,



αλλά αποκτούν μεγαλύτερη σημασία στο Διαδίκτυο όπου οι αλλαγές σε μία εφαρμογή μπορεί να γίνουν πολύ εύκολα και σύντομα, οπότε δεν υπάρχει αρκετός χρόνος για προηγούμενη μελέτη.

Στη γενική περίπτωση επιθυμούμε να συγκρίνουμε την αντίδραση των επισκεπτών σε διαφορετικούς τρόπους χρήσης των ίδιων υπηρεσιών του ιστοχώρου. Δηλαδή στον ιστοχώρο προσφέρεται τη μία περίπτωση ένας συγκεκριμένος τρόπος για την χρήση υπηρεσιών και σε μία άλλη περίπτωση ένας άλλος τρόπος. Επιθυμούμε να δούμε ποιος από τους 2 τρόπους είναι “καλύτερος”.

Ως μέτρο αξιολόγησης μπορούμε να θεωρήσουμε το ποσοστό των χρηστών που χρησιμοποιούν τη συγκεκριμένη υπηρεσία ανάλογα με την παραλλαγή που προφέρεται. Η χρήση της υπηρεσίας μεταφράζεται σε μία σελιδοπροβολή που προκλήθηκε από τις επιλογές του χρήστη.

Έστω για παράδειγμα θέλουμε να εξετάσουμε κατά πόσο ένα εικονίδιο γίνεται αντιληπτό από τους χρήστες. Η επιλογή (κλικ) του εικονιδίου οδηγεί σε συγκεκριμένη σελίδα, στην οποία φτάνει κανείς μόνο μέσω του εικονιδίου. Σχεδιάζουμε εναλλακτικά εικονίδια και εξετάζουμε το ποσοστό των επισκεπτών που επέλεξαν το κάθε εικονίδιο.

Εάν υπάρχουν διαφορετικοί τρόποι πρόσβασης στη συγκεκριμένη σελίδα, η παραπάνω μέθοδος δε θα δώσει ικανοποιητικά αποτελέσματα διότι δεν μπορούμε να γνωρίζουμε εάν οι επισκέπτες έφτασαν μέσω του εικονιδίου ή μέσω άλλου τρόπου στη σελίδα. Σε αυτή την περίπτωση προσφέρεται η έννοια της ακόλουθης ενέργειας σε μία σελιδοπροβολή. Πρόκειται για ενέργειες που αντιστοιχίζονται σε μία σελιδοπροβολή μετά την εμφάνισή της (βλέπε και ενότητα 3.2 σελ. 126). Οπότε αρκεί με την επιλογή του εικονιδίου να αντιστοιχιστεί μία ακόλουθη ενέργεια και να μετρηθεί απ' ευθείας πόσες φορές συναντάται η ενέργεια στη Βάση Χρήσης.

Εάν ακολουθήσουμε τη συλλογιστική, μπορούμε να φανταστούμε και εναλλακτικούς τρόπους χρήσης μίας υπηρεσίας στον ιστοχώρο που υπάρχουν ταυτόχρονα στις σελίδες. Έστω ότι προφέρεται σύνδεσμος προς μία άλλη σελίδα με δύο διαφορετικούς τρόπους ταυτόχρονα. Σε αυτή την περίπτωση μπορούμε να αντιστοιχίσουμε διαφορετικές ακόλουθες ενέργειες σε κάθε εικονίδιο ή σύνδεσμο, και να παρακολουθήσουμε πόσοι είναι οι επισκέπτες που κάνουν χρήση του κάθε εικονιδίου.

Έστω ότι υπάρχουν δύο εναλλακτικές μορφές της ίδιας επιλογής, και η πρώτη έχει ακόλουθη ενέργεια “xx” και η δεύτερη “yy”. Στην Ερώτηση 4.17 υπολογίζεται ο αριθμός των σελιδοπροβολών όπου επιλέχτηκε η επιλογή με τη μορφή “xx” και με τη μορφή “yy”. Εάν υπάρχουν σημαντικές διαφορές στη μία από τις δύο περιπτώσεις, τότε μπορούμε να συμπεράνουμε

ότι η μορφή που έχει η επιλογή με την μεγαλύτερη χρήση είναι πιο προσιτή στους επισκέπτες.

```
SELECT      COUNT(acc_actions_perform.pageimpression_id) AS usage,
            acc_actions.description AS action
FROM        acc_page_impressions INNER JOIN acc_actions_perform ON
            acc_page_impressions.pageImpression_id =
            acc_actions_perform.pageimpression_id INNER JOIN
            acc_actions ON acc_actions_perform.action_id =
            acc_actions.action_id
WHERE       (acc_actions.description = 'xx' OR
            acc_actions.description = 'yy')
            AND (acc_page_impressions.pageImpression_time > '20030601')
            AND (acc_page_impressions.pageImpression_time < '20030608')
GROUP BY   acc_actions.description
```

*Ερώτηση 4.17 Αριθμός χρήσης εναλλακτικής επιλογής ανά εναλλακτική*

## 4.4 Σενάρια Χρήσης

Τα σενάρια χρήσης αναφέρονται στο τρόπο χρήσης του ιστοχώρου όπως αυτό προβλέφθηκε κατά την ανάπτυξή του. Από αυτή την άποψη η μελέτη στα σενάρια χρήσης ενός ιστοχώρου είναι ίσως η πλέον σημαντική ενέργεια που πρέπει να γίνει διότι αποκαλύπτει πως τελικά οι επισκέπτες χρησιμοποιούν τον ιστοχώρο. Εάν όντως ο ιστοχώρος χρησιμοποιείται όπως προβλέφθηκε, τότε πρέπει να εξεταστεί εάν μπορεί να τροποποιηθεί ο ιστοχώρος ώστε να βελτιωθεί ακόμα περισσότερο ο τρόπος χρήσης του ιστοχώρου. Εάν από την άλλη διαπιστωθεί ότι οι χρήστες δε χρησιμοποιούν τον ιστοχώρο όπως αυτό προβλέφθηκε, τότε προκύπτουν πολλά ερωτήματα που πρέπει να απαντηθούν:

- μήπως οι επισκέπτες στον ιστοχώρο δεν ενδιαφέρονται για τις υπηρεσίες που προσφέρονται και μοντελοποιήθηκαν με τα σενάρια χρήσης; Αυτό μπορεί να οφείλεται σε προβληματική δεικτοδότηση από τις μηχανές αναζήτησης που “στέλνουν” έτσι διαφορετικούς χρήστες στον ιστοχώρο από αυτούς στους οποίους απευθύνεται. Απαιτείται έτσι αλλαγή του ιστοχώρου ώστε να γίνει καλύτερη δεικτοδότηση
- μήπως υπάρχει πρόβλημα χρησιμοποίησιμότητας στον ιστοχώρο που εμποδίζει τους επισκέπτες να εντοπίσουν εκείνες τις σελίδες που καλύπτουν τα σενάρια χρήσης; Απαιτείται σε αυτή την περίπτωση τροποποίηση του ιστοχώρου ώστε να γίνει περισσότερη εμφανή η λειτουργία του κάθε τμήματός του;
- μήπως οι επισκέπτες ξεκινούν εργασίες που προβλέφθηκαν από τα σενάρια χρήσης, εγκαταλείπουν όμως πρόωρα; Και σε αυτή την περίπτωση πρέπει να εξεταστεί μήπως υπάρχει πρόβλημα χρησιμοποίησιμότητας ή πρόβλημα στο περιεχόμενο του ιστοχώρου.

Έχοντας υπ' όψιν τα σενάρια χρήσης ενός ιστοχώρου είναι θεμιτό να γνωρίζουμε στοιχεία όπως το ποσοστό των χρηστών που ακολουθούν τα σενάρια ή το ποσοστό των χρηστών που ακολουθούν αρχικά ένα σενάριο αλλά τελικά δεν το ολοκληρώνουν. Από τους χρήστες που ακολουθούν και ολοκληρώνουν ένα σενάριο έχει ενδιαφέρον εάν η ολοκλήρωση επιτεύχθηκε με το βέλτιστο δυνατό μονοπάτι ή εάν χρειάστηκαν περισσότερες σελιδοπροβολές για την ολοκλήρωση. Από τους χρήστες που δεν ολοκληρώνουν έχει ενδιαφέρον να γνωρίζουμε σε ποια συγκεκριμένη σελιδοπροβολή διέκοψαν το σενάριο. Επίσης είναι θεμιτό να γνωρίζουμε τα χαρακτηριστικά των χρηστών κάθε κατηγορίας σε όρους διεύθυνσης προέλευσης (referring URL) ή χώρας προέλευσης.

Η μελέτη σεναρίων χρήσης είναι αρκετά σύνθετη διότι περιλαμβάνει πολλές παραμέτρους που πρέπει να εξεταστούν. Συνδυάζει τόσο ζητήματα λειτουργικότητας του ιστοχώρου, όσο θέματα χρησιμοποίησιμότητας και περιεχομένου. Για να μελετηθούν οι λόγοι που οδηγούν σε μία συγκεκριμένη συμπεριφορά απαιτείται γνώση για το συγκεκριμένο περιεχόμενο που εμφανίστηκε στους επισκέπτες σε κάθε σελιδοπροβολή.

Όπως αναφέρθηκε και στο κεφάλαιο της σχετικής έρευνας (βλέπε ενότητες 2.3.5.7, 2.3.6.1, 2.3.6.2 και σχετικές εργασίες [14], [3], [18], [55], [56], [23], [1], [71] και [58]), η κάθε επίσκεψη μπορεί να μοντελοποιηθεί ως μονοπάτι σε ένα γράφο όπου κάθε κόμβος του γράφου αναπαριστά μία σελιδοπροβολή και οι ακμές συνδέσμους ανάμεσα σε σελίδες. Τα σενάρια χρήσης μπορεί να μοντελοποιηθούν ως μονοπάτια πάνω σε αυτό το γράφο, και η μελέτη μπορεί να γίνει υποβάλλοντας ερωτήσεις πάνω σε αυτό το γράφο. Κάποιες τέτοιες ερωτήσεις είναι

- οι χρήστες που ξεκινούν από την αρχική σελίδα και πηγαίνουν μετά στη φόρμα αναζήτησης διαμονής, φτάνουν σε σελίδες παρουσίασης διαμονής; Εάν ναι, μετά από πόσες σελιδοπροβολές. Εάν όχι, σε ποιο σημείο διέκοψαν; Αυτοί που διέκοψαν, για πιο λόγο διέκοψαν; Ποια τα χαρακτηριστικά αυτών που διέκοψαν σε όρους εξωτερικών φράσεων αναζήτησης; Ποια τα χαρακτηριστικά των σελίδων στις οποίες διέκοψαν;
- οι χρήστες που έρχονται στον ιστοχώρο με συγκεκριμένη φράση αναζήτησης, βλέπουν τελικά σελίδες που να καλύπτουν αυτή τη φράση αναζήτησης;

Ένα βασικό πρόβλημα που υπάρχει στον παγκόσμιο ιστό, είναι ότι δεν γνωρίζουμε τις πραγματικές προθέσεις ενός επισκέπτη. Σε αντίθεση με τα πειράματα χρηστών σε παραδοσιακές μελέτες χρησιμοποίησιμότητας, όπου δίνονται στο χρήστη συγκεκριμένες εργασίες που καλείται να διεκπεραιώσει, για το χρήστη σε μία εφαρμογή ιστοχώρου δε γνωρίζουμε ποιες ακριβώς είναι οι πληροφορίες ή υπηρεσίες που αναζητάει ή ο ακριβής σκοπός της επίσκεψης.

Παρόλα αυτά, μπορεί από τα στοιχεία που συλλέγονται και καταγράφονται από ένα σύστημα καταγραφής, να εξαχθούν συμπεράσματα για το σκοπό του κάθε επισκέπτη. Εάν για παράδειγμα ο επισκέπτης προέρχεται από εξωτερική μηχανή αναζήτησης και είναι γνωστή η φράση που χρησιμοποίησε, μπορεί εύλογα να υποτεθεί ότι ενδιαφέρεται για θέματα που περιγράφονται από αυτή τη φράση.

Στη συνέχεια θα περιγραφούν τρόποι για τον εντοπισμό των προθέσεων των επισκεπτών σε έναν ιστοχώρο (και έτσι την κατάταξή τους σε συγκεκριμένο σενάριο χρήσης). Θα περιγραφούν επίσης τα γενικότερα συμπεράσματα που μπορεί να εξαχθούν μελετώντας σενάρια χρήσης. Θα περιγραφεί τέλος μία μέθοδος για την έκφραση ερωτήσεων σε για ακολουθίες που αναπαριστούν σενάρια χρήσης και θα παρουσιαστεί μία συγκεκριμένη μεθοδολογία για τη μελέτη πλήρως προκαθορισμένων σεναρίων (στατικά σενάρια) και δυναμικά σενάρια που προκύπτουν κατά τη μελέτη.

#### **4.4.1 Εντοπισμός Σεναρίων**

Όπως αναφέρθηκε, στον παγκόσμιο ιστό δεν υπάρχει ακριβή γνώση για τις προθέσεις ενός επισκέπτη όταν εισέρχεται σε έναν ιστοχώρο. Για να μελετηθεί όμως η συμπεριφορά των επισκεπτών πάνω σε συγκεκριμένο σενάριο χρήσης, πρέπει να υπάρχει σχετική βεβαιότητα ότι σε μία συγκεκριμένη επίσκεψη ο χρήστης προσπάθησε ή είχε σκοπό να προσπαθήσει να χρησιμοποιήσει υπηρεσία που καλύπτεται από το υπό μελέτη σενάριο χρήσης.

Χρειάζεται επίσης μία μέθοδος για να εντοπίζεται πότε ένα σενάριο έληξε (είτε με ολοκλήρωση είτε με εγκατάλειψη της προσπάθειας).

Για το σκοπό ενός ιστοχώρου τουριστικού προορισμού μπορεί να χρησιμοποιηθούν οι μέθοδοι που περιγράφονται στη συνέχεια για τον εντοπισμό της έναρξης ενός σεναρίου και της λήξης:

##### **4.4.1.1 Έναρξη από Σημασιολογία Προέλευσης**

Όταν εισέρχεται ένας επισκέπτης στον ιστοχώρο προερχόμενος από συγκεκριμένο URL μπορεί να εξαχθούν συμπεράσματα για το σκοπό της επίσκεψης από τα ακόλουθα στοιχεία:

- εάν πρόκειται για URL από μηχανή αναζήτησης και είναι γνωστή η φράση που χρησιμοποιήθηκε, τότε χρησιμοποιείται η φράση για να καταταχθεί ο επισκέπτης σε συγκεκριμένο σενάριο χρήσης. **Παράδειγμα:** υπάρχει επίσκεψη προερχόμενη από τη μηχανή αναζήτησης Google με φράση “Hotels in Crete”. Η επίσκεψη θεωρείται ότι καλύπτει το σενάριο “Αναζήτηση Διαμονής”

- εάν το URL προέλευσης καλύπτει συγκεκριμένη θεματολογία και προσφέρει σύνδεσμο προς τον ιστοχώρο τουριστικού προορισμού για περαιτέρω έρευνα στη θεματολογία αυτή, τότε η επίσκεψη κατατάσσεται στο σενάριο αυτής της θεματολογίας. **Παράδειγμα:** ένας εξωτερικός ιστοχώρος αναφέρεται σε συνέδριο που θα διεξαχθεί στην Κρήτη. Παρέχει σελίδα με προτάσεις διαμονής στην περιοχή του συνεδρίου και σύνδεσμο προς τον ιστοχώρο τουριστικού προορισμού για περαιτέρω έρευνα για αναζήτηση διαμονής. Η επίσκεψη θεωρείται ότι καλύπτει το σενάριο “Αναζήτηση Διαμονής”

#### **4.4.1.2 Έναρξη από Φράσεις Αναζήτησης εντός Ιστοχώρου**

Ο ιστοχώρος τουριστικού προορισμού παρέχει μία υπηρεσία γενικής αναζήτησης στον ιστοχώρο. Ανάλογα με τις φράσεις που χρησιμοποιούνται από τους επισκέπτες, μπορεί να καταταχθεί η επίσκεψη σε συγκεκριμένο σενάριο. **Παράδειγμα:** εισέρχεται ένας επισκέπτης στον ιστοχώρο και εισάγει στο πεδίο αναζήτησης “Hotels in Chania”. Η επίσκεψη κατατάσσεται στο σενάριο “Αναζήτηση Διαμονής”.

#### **4.4.1.3 Έναρξη από Μονοπάτι Επίσκεψης**

Μονοπάτια που ακολουθούνται κατά την επίσκεψη και συμπίπτουν με τα πρώτα τμήματα από μονοπάτια που προβλέφθηκαν από τα σενάρια χρήσης μπορεί να χρησιμοποιηθούν για την κατάταξη της επίσκεψης στο σενάριο με το οποίο συμπίπτει το μονοπάτι. **Παράδειγμα:** σε μία επίσκεψη εντοπίζεται το μονοπάτι “Αρχική Σελίδα” -> “Σελίδα Αναζήτησης Διαμονής”. Η επίσκεψη κατατάσσεται στο σενάριο “Αναζήτηση Διαμονής”.

#### **4.4.1.4 Λήξη Σεναρίου**

Η λήξη ενός σεναρίου χρήσης στην περίπτωση της επιτυχούς ολοκλήρωσης μπορεί να θεωρηθεί όταν γίνεται η πρώτη σελιδοπροβολή σε σελίδα τερματισμού του σεναρίου (για τον ιστοχώρο τουριστικού προορισμού οι σελίδες τερματισμού παρουσιάζονται στην ενότητα 5.1). Στην αναζήτηση διαμονής δηλαδή αυτό θα σήμαινε ότι όταν μετά την έναρξη του σεναρίου γίνεται μία σελιδοπροβολή παρουσίασης επιχείρισης διαμονής, τότε θεωρείται ότι έγινε επιτυχής ολοκλήρωση του σεναρίου χρήσης. Στην πραγματικότητα βέβαια μπορεί να ακολουθήσουν και άλλες τέτοιες σελιδοπροβολές που ανήκουν ακόμα στο ίδιο σενάριο χρήσης, αλλά το σημείο που ενδιαφέρει είναι ότι έγινε επιτυχής ολοκλήρωση του σεναρίου.

Στην περίπτωση της εγκατάλειψης ενός σεναρίου χρήσης, ο εντοπισμός μπορεί να γίνει εάν

- γίνεται έναρξη νέου σεναρίου χωρίς την ολοκλήρωση του προηγούμενου
- εγκαταλειφθεί ο ιστοχώρος χωρίς την ολοκλήρωση του σεναρίου

#### **4.4.2 Συμπεράσματα από την Ανάλυση Χρήσης Σεναρίων**

Τόσο στις περιπτώσεις όπου υπάρχει επιτυχής ολοκλήρωση των σεναρίων χρήσης όσο και στις περιπτώσεις εγκατάλειψης της προσπάθειας, μπορεί να εξαχθούν συμπεράσματα όσον αφορά τη χρησιμοποιησιμότητα ή να εντοπιστούν προβληματικά σημεία του ιστοχώρου.

Πρώτα απ' όλα ενδιαφέρει πόσοι είναι οι χρήστες που κατάφεραν να ολοκληρώσουν τα σενάρια χρήσης και πόσοι είναι αυτοί που τα εγκατέλειψαν.

Στην περίπτωση της επιτυχούς ολοκλήρωσης, ενδιαφέρει κυρίως ο αριθμός των σελιδοπροβολών που απαιτήθηκαν για την ολοκλήρωση, για να εξακριβωθεί εάν το επίπεδο διεπαφής του ιστοχώρου είναι αρκετά διαισθητικό ώστε να οδηγεί το χρήστη στο επιθυμητό αποτέλεσμα. Εάν αντιθέτως ανακαλυφθεί ότι οι χρήστες “δαπανούν” περισσότερες σελιδοπροβολές από τις απαιτούμενες για να ολοκληρώσουν την εργασία τους, τότε πρέπει να αναζητηθούν τα σημεία εκείνα που είναι προβληματικά.

Η περίπτωση της μη ολοκλήρωσης ενός σεναρίου παρουσιάζει σαφώς μεγαλύτερο ενδιαφέρον διότι εκεί υπάρχει σίγουρα κάποιο πρόβλημα στον ιστοχώρο που δεν επιτρέπει τους επισκέπτες να ολοκληρώσουν την εργασία που επιθυμούν. Η συγκεκριμένη σελιδοπροβολή στην οποία έγινε εγκατάλειψη της προσπάθειας από τον επισκέπτη καθώς και οι προηγούμενες ενέργειές του είναι ιδιαίτερα σημαντικές για την εξαγωγή συμπερασμάτων. Εάν η σελιδοπροβολή εγκατάλειψης αναφέρεται σε αποτελέσματα κάποιας αναζήτησης, μπορεί να γίνουν ποικίλες υποθέσεις που όμως συγκεκριμενοποιούνται από την παρουσίαση της συγκεκριμένης σελίδας. Για παράδειγμα η έλλειψη αποτελεσμάτων στη σελίδα αναζήτησης μπορεί να οφείλεται στο γεγονός ότι δεν υπάρχουν πληροφορίες στον ιστοχώρο για την αναζήτηση που διενήργησε ο χρήστης, αλλά μπορεί επίσης να οφείλεται σε πρόβλημα στο μηχανισμό αναζήτησης της εφαρμογής ιστοχώρου.

Η συγκέντρωση μεγάλου ποσοστού αποτυχημένων προσπαθειών που έχουν τα ίδια χαρακτηριστικά, εντοπίζει το προβληματικό σημείο αλλά δε δίνει λύση στο πρόβλημα. Εδώ είναι που παρεμβαίνει ο ειδικός σε θέματα χρησιμοποιησιμότητας σε συνεργασία με γνώστες του περιεχομένου του ιστοχώρου για να εντοπίσουν το ακριβές σημείο που δυσχεραίνει τους χρήστες στην χρήση του ιστοχώρου.

#### **4.4.3 Γενική Μορφή Ερωτήσεων**

Οι ανάγκες για ερωτήσεις τους είδους που περιγράφηκαν στις προηγούμενες παραγράφους μπορεί να περιγραφούν από ερωτήσεις για την ανακάλυψη σειριακών σχημάτων (sequential patterns) με την προσθήκη περιορισμών στο είδος των ακολουθιών που εξετάζονται και στα χαρακτηριστικά

των επισκέψεων και των σελιδοπροβολών. Το αποτέλεσμα της ερώτησης είναι τα αθροιστικά (aggregated) χαρακτηριστικά για τα σειριακά σχήματα που εξετάζονται.

Στη γενική περίπτωση οι ερωτήσεις αποτελούνται από τα εξής τμήματα:

1. ένα σύνολο κριτηρίων έναρξης, για τον εντοπισμό της έναρξης σεναρίων
2. ένα σύνολο κριτηρίων λήξης, για τον εντοπισμό της λήξης σεναρίου
3. ένα σύνολο κριτηρίων περιορισμού, για τον περιορισμό των εξεταζόμενων μονοπατιών
4. ένα σύνολο χαρακτηριστικών εξόδου ερώτησης, για τον προσδιορισμό της εξόδου της ερώτησης

Έστω ότι ένα σενάριο προβλέπει την αναζήτηση διαμονής σε ιστοχώρο τουριστικού προορισμού. Το σενάριο προβλέπει ότι ο χρήστης επιλέγει από την αρχική σελίδα σύνδεσμο που οδηγεί σε φόρμα αναζήτησης διαμονής. Σε αυτή τη φόρμα ίσως ο επισκέπτης επιλέξει ορισμένες περιοχές στις οποίες θέλει να περιορίσει τα αποτελέσματα. Ακολουθεί η προβολή αποτελεσμάτων αναζήτησης από τα οποία επιλέγει κάποιο και προβάλλεται μία σελίδα παρουσίασης διαμονής οπότε και ολοκληρώνεται το σενάριο.

Η ερώτηση που μπορεί να υποβληθεί είναι

Δώσε τις επισκέψεις και το μήκος των μονοπατιών των επισκέψεων όπου στο μονοπάτι συναντάται ακολουθία σελιδοπροβολών με τους ακόλουθους περιορισμούς: υπάρχει προβολή της σελίδας “index” ακολουθούμενη από προβολή της σελίδας “hotels form”. Έπειτα απροσδιόριστος αριθμός προβολών σε σελίδες ανάμεσα από “searchresults” ή “select area” ή “hotels form”. Έπειτα υπάρχει προβολή της σελίδας “view sme” με action “present sme” και entity το αναγνωριστικό επιχειρήσεων διαμονής.

#### 4.4.4 Έκφραση Ερωτήσεων για Σενάρια Χρήσης

Στο [71] παρουσιάζεται μία γλώσσα όμοια της SQL (MINT) για την υποβολή ερωτήσεων για την ανακάλυψη σειριακών σχημάτων και χρησιμοποιείται στο σύστημα Web Utilization Miner (WUM) [13]. Σε αυτή τη γλώσσα ορίζονται templates που πρέπει να ακολουθούν τα προς εύρεση μονοπάτια και επιτρέπονται περιορισμοί στον αριθμό των εμφανίσεων κάθε σελίδας στο μονοπάτι και τη συχνότητα εμφάνισης στο σύνολο των μονοπατιών (support). Ένα παράδειγμα φαίνεται στην Ερώτηση 4.18.

Σε αυτό το παράδειγμα τα  $x$  και  $y$  είναι μεταβλητές που αναπαριστούν κόμβους στο γράφο

```
SELECT t
FROM NODE AS xy, TEMPLATE x*y* AS t
WHERE x.url != "/balk.html"
AND x.support > 40
AND y.url = "/contact.html"
```

*Ερώτηση 4.18 Παράδειγμα της γλώσσας MINT [71] για ανακάλυψη σειριακών σχημάτων*

μονοπατιών. Η έκφραση  $x * y *$  περιορίζει το είδος μονοπατιού όπου ο αστερίσκος αναπαριστά απροσδιόριστο αριθμό και είδος κόμβων στο γράφο. Το  $x$  και  $y$  είναι μεταβλητές του κόμβου. Υπάρχουν περιορισμοί για το url των  $x$  και  $y$  ενώ απαιτείται επίσης το support (αριθμός εμφάνισης) του κόμβου  $x$  να είναι 40. Η ερώτηση θα λάβει υπ' όψιν όλα τα μονοπάτια που ξεκινούν από τη σελίδα  $x$  ακολουθούνται από απροσδιόριστο αριθμό σελίδων, περνάνε από τη σελίδα  $y$  και ακολουθούνται από απροσδιόριστο αριθμό σελίδων.

Η έξοδος των ερωτήσεων της MINT είναι ένας γράφος χωρίς κύκλους με τα μονοπάτια που καλύπτουν τα κριτήρια της ερώτησης. Η γλώσσα MINT μαζί με το σύστημα WUM (βλέπε [13]) στο οποίο εφαρμόζεται είναι ισχυρή στο να εντοπίζει μονοπάτια που ακολουθούν ένα συγκεκριμένο σχήμα και περιορίζονται από κάποια στατιστικά στοιχεία. Χωλαίνει όμως σε σύγχρονους ιστοχώρους όπου το περιεχόμενο είναι αποτέλεσμα ερωτήσεων σε μία ή περισσότερες πηγές δεδομένων και το URL μίας σελιδοπροβολής δεν επαρκεί για την περαιτέρω ανάλυση.

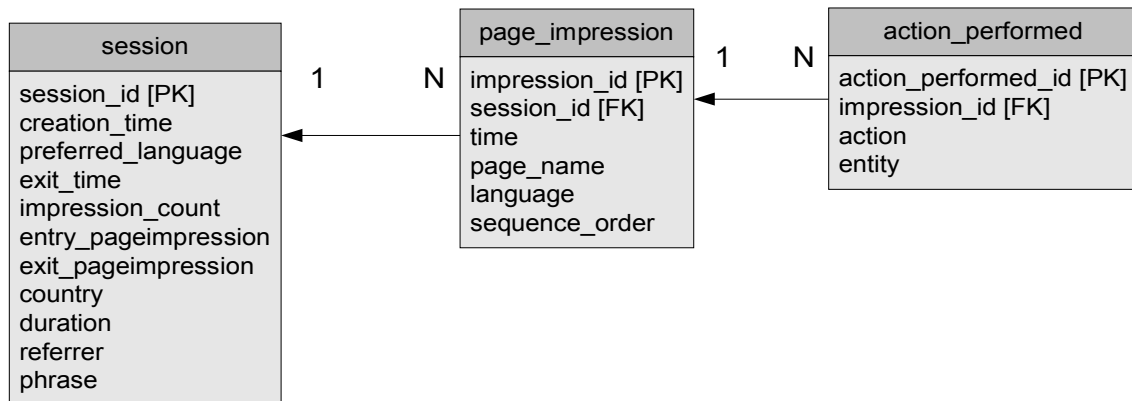
Για τη μελέτη των σεναρίων χρήσης, ο μηχανισμός των templates και τα wild cards της MINT καλύπτει μεγάλο μέρος των αναγκών, αλλά απαιτούνται επιπλέον δυνατότητες που έχουν σχέση με περιορισμό στα στοιχεία από την **επίσκεψη** (για παράδειγμα φράσεις από διευθύνσεις προέλευσης) ή από καθ' αυτές τις σελιδοπροβολές (είδος επιχείρισης που παρουσιάζεται). Επίσης απαιτείται ο περιορισμός στα χαρακτηριστικά των σελίδων που προσδιορίζονται με wild cards.

Στη συνέχεια θα παρουσιαστεί μία συστηματική μέθοδος για να υποβληθούν ερωτήσεις στη Βάση Χρήσης που αναφέρονται σε σενάρια χρήσης ως ακολουθία σελιδοπροβολών. Θα περιγραφεί επίσης πως αυτές οι ερωτήσεις μπορεί να εμπλουτιστούν με περιορισμούς στις παραμέτρους της επίσκεψης.

Για τη διευκόλυνση της περιγραφής θα χρησιμοποιηθεί ένα απλοποιημένο σχήμα της Βάσης Δεδομένων που φαίνεται στο Διάγραμμα 4.1. Φαίνεται η σχέση session που αναπαριστά τις επισκέψεις στον ιστοχώρο, η σχέση page\_impression που αναπαριστά τις σελιδοπροβολές και η σχέση action\_performed που αναπαριστά τις ενέργειες σε κάθε σελιδοπροβολή. Έχουν παραληφθεί οι σχέσεις για το όνομα της ενέργεια και της κάθε σελίδας (που απαιτούν ένα απλό join με τις αντίστοιχες σχέσεις) και περιλαμβάνονται στο σχήμα ως action και page\_name.



Ιδιαίτερη σημασία έχει το attribute `sequence_order` στη σχέση `page_impression`, όπου σε αυτό αποθηκεύεται η σειρά εμφάνισης της κάθε σελιδοπροβολής ανά επίσκεψη.



Διάγραμμα 4.1 Απλοποιημένο σχήμα της σχεσιακής βάσης δεδομένων όπου φαίνονται οι σχέσεις για τις επισκέψεις (*session*), τις σελιδοπροβολές (*page\_impression*) και τις ενέργειες σελίδας (*action*). Σημαντικό ρόλο παίζει το attribute `sequence_order` στη σχέση `page_impression` όπου αποθηκεύεται η σειρά της κάθε σελιδοπροβολής στην κάθε επίσκεψη.

Αρχικά θα υπολογιστούν ακολουθίες απλής μορφής του είδους “*a b c d*” όπου κάθε γράμμα αναπαριστά μία σελιδοπροβολή με ορισμένα χαρακτηριστικά ενώ πρόκειται για συνεχόμενες σελιδοπροβολές.

Στη συνέχεια θα εισαχθούν τα wild cards \* και + στην ακολουθία που συμβολίζουν απροσδιόριστο αριθμό σελιδοπροβολών ή τουλάχιστον μία σελιδοπροβολή αντίστοιχα. Έτσι στην ακολουθία “*a b w\* c d*” το *w\** συμβολίζει απροσδιόριστο αριθμό σελιδοπροβολών που παρεμβάλλονται ανάμεσα στις “*a b*” και “*c d*”.

Στη συνέχεια θα εισαχθούν επιπλέον περιορισμούς στο είδος των απροσδιόριστων σελιδοπροβολών. Δηλαδή θα περιορίσουμε στην προηγούμενη ακολουθία το *w\** να έχει συγκεκριμένα χαρακτηριστικά, όπως στο όνομα της σελίδας ή στις ενέργειες που συσχετίζονται με αυτή.

Στο τέλος θα εισαχθούν περιορισμοί στις παραμέτρους της επίσκεψης (URL προέλευσης, χώρα προέλευσης κλπ.) και θα δοθεί ένα παράδειγμα για σενάριο χρήσης.

#### 4.4.4.1 Ακολουθίες Απλής Μορφής

Στην απλή περίπτωση ενδιαφέρει να εντοπιστούν ακολουθίες με συνεχόμενες σελιδοπροβολές. Έστω ότι θέλουμε να βρούμε ακολουθίες της μορφής

$a_1 a_2 a_3 \dots a_n$

Στην ακολουθία αυτή τα  $a_1$  ως  $a_n$  αναπαριστούν  $n$  συνεχόμενες σελιδοπροβολές. Με βάση το σχήμα στο Διάγραμμα 4.1 η σχέση που συνδέει δύο συνεχόμενες σελιδοπροβολές είναι ότι το attribute `sequence_order` έχει διαφορά 1 και και οι δύο σελιδοπροβολές ανήκουν στην ίδια επίσκεψη. Οπότε για να βρεθούν όλες οι πιθανές συνεχόμενες δυάδες σελιδοπροβολών αρκεί ένα join της σχέσης `page_impression` με τον εαυτό της πάνω στα attributes `sequence_order` και `session_id`, όπως φαίνεται στο Πλαίσιο 4.1.

```
page_impression a1 JOIN page_impression a2 ON
a1.session_id = a2.session_id AND
a1.sequence_order = a2.sequence_order -1
```

*Πλαίσιο 4.1 Ακολουθία δύο συνεχόμενων σελιδοπροβολών*

Για ακολουθία μήκους  $n$  το join επεκτείνεται η ερώτηση αντίστοιχα  $n$  φορές όπως φαίνεται στο Πλαίσιο 4.2

```
page_impression a1 JOIN page_impression a2 ON
a1.session_id = a2.session_id AND
a1.sequence_order = a2.sequence_order -1
JOIN page_impression a3 ON
a2.session_id = a3.session_id AND
a2.sequence_order = a3.sequence_order -1
.
.
.
JOIN page_impression an ON
a(n-1).session_id = an.session_id AND
a(n-1).sequence_order = an.sequence_order -1
```

*Πλαίσιο 4.2 Ακολουθία  $n$  συνεχόμενων σελιδοπροβολών*

Με βάση την προηγούμενη βασική κατασκευή μπορούμε να εισάγουμε περιορισμούς στις παραμέτρους της κάθε σελιδοπροβολής. Έστω ότι ενδιαφέρει να βρεθούν οι τριάδες *abc* όπου το όνομα της σελίδας *a* είναι “index” και το όνομα της σελίδας *c* είναι “sme” και στην *c* υπάρχει στη σχέση `action_perform` η ενέργεια `action=’present sme’`. Η ερώτηση διατυπώνεται όπως φαίνεται στο Πλαίσιο 4.3. Το SELECT CLAUSE της ερώτησης εξάγει τις τριάδες ονομάτων των σελίδων *abc*.

Εάν ενδιαφέρει να εξαχθούν οι συγκεκριμένες επισκέψεις και το σημείο σε αυτές όπου υπάρχουν τα μονοπάτια της μορφής αυτής, τότε το SELECT CLAUSE επεκτείνεται όπως φαίνεται στο Πλαίσιο 4.4. Ουσιαστικά εξάγεται το αναγνωριστικό της επίσκεψης και το σημείο έναρξης της ακολουθίας (που είναι στην πρώτη σελιδοπροβολή *a*).

```

SELECT a.session_id, a.sequence_order, a.page_name, b.page_name, c.page_name
FROM
    page_impression a JOIN page_impression b ON
    a.session_id = b.session_id AND
    a.sequence_order = b.sequence_order -1
    JOIN page_impression c ON
    b.session_id = c.session_id AND
    b.sequence_order = c.sequence_order -1

    JOIN action_perform ap ON
    c.impression_id = ap.impression_id

WHERE
    a1.page_name = 'index' AND c.page_name = 'sme' AND
    ap.action = 'present sme'

```

*Πλαίσιο 4.4 Ακολουθία με περιορισμός στα a και c και εξαγωγή των συγκεκριμένων επισκέψεων*

Το support της παραπάνω ακολουθίας εξάγεται εύκολα με την έκφραση

```
count(concat (a.session_id, a.sequence_order))
```

#### **4.4.4.2 Ακολουθίες με Wildcards**

Ακολουθίες με wildcards έχουν τη μορφή που φαίνεται στο Πλαίσιο 4.5

```

a b c* d e
a b c+ d e

```

*Πλαίσιο 4.5 Ακολουθίες με τη χρήση wildcard*

Σε αυτές τις ακολουθίες το σύμβολο \* συμβολίζει απροσδιόριστο αριθμό σελιδοπροβολών, ενώ το σύμβολο + συμβολίζει τουλάχιστον μία σελιδοπροβολή.

Στο προηγούμενο παράδειγμα έχουμε ακολουθίες της μορφής a που ακολουθείται από b μετά από απροσδιόριστο αριθμό σελιδοπροβολών και καταλήγει με τη σελιδοπροβολή d και e.

Αντίστοιχα για το σύμβολο + θα είχαμε τουλάχιστον μία σελιδοπροβολή και όχι απροσδιόριστο αριθμό.

Ο υπολογισμός αυτών των ακολουθιών γίνεται εύκολα με μικρή μόνο αλλαγή των ερωτήσεων στα πλαίσια 4.1 έως 4.4. Στο σημείο όπου υπάρχει το wildcard, αντί να γίνει JOIN με περιορισμό η διαφορά ανάμεσα στις επόμενες σελιδοπροβολές να έχει διαφορά 1, στην περίπτωση του \* η διαφορά πρέπει να είναι μεγαλύτερη ή ίση του 1 και στην περίπτωση του + μεγαλύτερη του 1. Το πρόβλημα είναι ότι πλέον δεν μπορεί να εξαχθούν τα πλήρη μονοπάτια που ικανοποιούν την δοσμένη ακολουθία, διότι δε γνωρίζουμε το πλήρες μήκος της ακολουθίας.

Στο Πλαίσιο 4.6 υπολογίζεται η ακολουθία “a b\* c” με τους ίδιους περιορισμούς για το όνομα της

σελίδας a και c όπως στην τελευταία περίπτωση της προηγούμενης ενότητας

```
SELECT a.session_id, a.sequence_order
FROM
    page_impression a JOIN page_impression c ON
    a.session_id = c.session_id AND
    a.sequence_order <= c.sequence_order - 1

    JOIN action_perform ap ON
    c.impression_id = ap.impression_id
WHERE
    a1.page_name = 'index' AND c.page_name = 'sme' AND
    ap.action = 'present sme'
```

Πλαίσιο 4.6 Ακολουθία με περιορισμό στα a και c και wildcard ανάμεσά τους

#### 4.4.4.3 Ακολουθίες με Wildcards με Περιορισμούς

Στην αμέσως προηγούμενη ενότητα περιγράφηκε πως μπορεί να απαντηθούν με SQL ερωτήσεις όπου υπάρχει ένα wildcard στην ακολουθία σελιδοπροβολών που επιθυμούμε. Στην ερώτηση αυτή δεν υπήρχαν περιορισμοί στις σελίδες αυτές (π.χ. όνομα σελίδας) και η υλοποίηση δε διαφέρει ουσιαστικά από την περίπτωση ακολουθιών χωρίς wildcard. Εάν θελήσουμε να υπολογίσουμε την ερώτηση θέτοντας περιορισμούς στις σελίδες του wildcard, τότε η ερώτηση σε SQL γίνεται αρκετά πιο πολύπλοκη.

Έστω ότι επιθυμούμε την απάντηση της ερώτησης “ $a\ b^+c$ ”, όπου όμως θέλουμε να περιορίσουμε τις σελίδες b ώστε το όνομα των σελίδων να βρίσκεται ανάμεσα σε προκαθορισμένο σύνολο ονομάτων ( $n1, n2, n3$ ). Μία πρώτη προσέγγιση φαίνεται στο Πλαίσιο 4.7.

```
SELECT a.session_id, a.sequence_order
FROM
    page_impression a JOIN page_impression c ON
    a.session_id = c.session_id AND
    a.sequence_order < c.sequence_order - 1

    JOIN action_perform ap ON
    c.impression_id = ap.impression_id
WHERE
    b.page_name IN (n1, n2, n3)
```

Πλαίσιο 4.7 ΛΑΘΟΣ υπολογισμός ακολουθίας  $a\ b^+c$  με περιορισμό στο όνομα της b

Το πρόβλημα με την παραπάνω προσέγγιση είναι ότι αρκεί τουλάχιστον **μία** σελίδα ανάμεσα στην a και c με όνομα στο σύνολο ονομάτων που περιορίζουμε, για να επιστρέψει την ακολουθία, ενώ εμείς επιθυμούμε **όλα** τα ονόματα να ανήκουν στο σύνολο περιορισμού.

Παρατηρούμε όμως ότι μπορούμε εύκολα να υπολογίσουμε τα μονοπάτια που δεν καλύπτουν τους περιορισμούς για το wildcard, επειδή είναι όλα αυτά που έχουν τουλάχιστον μία σελίδα με όνομα διαφορετικά από το σύνολο περιορισμού. Στο Πλαίσιο 4.8 παρατηρούμε το WHERE CLAUSE όπου απαιτείται το b να **μη** ανήκει στο σύνολο ονομάτων.

```
SELECT a.session_id, a.sequence_order
FROM
    page_impression a JOIN page_impression c ON
    a.session_id = c.session_id AND
    a.sequence_order < c.sequence_order - 1

    JOIN action_perform ap ON
    c.impression_id = ap.impression_id
WHERE
    b.page_name NOT IN (n1, n2, n3)
```

*Πλαίσιο 4.8 Υπολογισμός ακολουθιών  $a b^* c$  που δεν καλύπτουν τους περιορισμούς για το wildcard*

Το πρόβλημα μπορεί να λυθεί ακολουθώντας τα επόμενα βήματα 1 έως 3

1. Υπολογίζουμε όλα τα μονοπάτια χωρίς περιορισμούς στο wildcard " $a b^* c$ ".
2. Υπολογίζουμε όλα τα μονοπάτια που **δεν** συμφωνούν με τον περιορισμό για το wildcard
3. Αφαιρούμε τα μονοπάτια του βήματος 2 από τα μονοπάτια του βήματος 1

Η συνολική ερώτηση για τον υπολογισμό της ακολουθίας " $a b^+ c$ " με  $b \in (n1, n2, n3)$  φαίνεται πλέον στο Πλαίσιο 4.9

Στη γραμμή 1 έως 9 υπολογίζονται οι ακολουθίες χωρίς τον περιορισμό στο wildcard. Στην γραμμή 13 ως 23 υπολογίζονται οι ακολουθίες που **δεν** καλύπτουν τον περιορισμό. Με το NOT IN της γραμμής 10 αφαιρούνται οι δεύτερες ακολουθίες από τις πρώτες, οπότε παραμένουν μόνο αυτές που καλύπτουν τους περιορισμούς. Για την αφαίρεση χρησιμοποιούμε το concat του session\_id με την έναρξη της ακολουθίας, το οποίο κάνει ένωση συμβολοσειράς..

```

1. SELECT a.session_id, a.sequence_order
2. FROM
3.     page_impression a JOIN page_impression c ON
4.     a.session_id = c.session_id AND
5.     a.sequence_order < c.sequence_order - 1
6.
7.     JOIN action_perform ap ON
8.     c.impression_id = ap.impression_id
9.
10. WHERE CONCAT(a.session_id, a.sequence_order) NOT IN
11.
12. (
13.     SELECT CONCAT(a.session_id, a.sequence_order)
14.     FROM
15.         page_impression a JOIN page_impression c ON
16.         a.session_id = c.session_id AND
17.         a.sequence_order < c.sequence_order - 1
18.
19.         JOIN action_perform ap ON
20.         c.impression_id = ap.impression_id
21.
22.     WHERE b.page_name NOT IN (n1, n2, n3)
23. )
24. )

```

*Πλαίσιο 4.9 Υπολογισμός ακολουθιών  $a b^* c$  που δεν καλύπτουν τους περιορισμούς για το wildcard*

#### **4.4.4.4 Ακολουθίες με Περιορισμούς στην Επίσκεψη**

Στις προηγούμενες ενότητες περιγράφηκαν περιπτώσεις για εντοπισμό ακολουθιών που έχουν περιορισμούς μόνο στις παραμέτρους των σελιδοπροβολών. Ο περιορισμός και σε παραμέτρους των επισκέψεων (όπως φράση αναζήτησης από εξωτερικές μηχανές αναζήτησης ή χώρα προέλευσης επισκέπτη) είναι πολύ εύκολο να υλοποιηθούν, μιας και το αποτέλεσμα των προηγούμενων ερωτήσεων περιλαμβάνει πάντα και το session\_id της επίσκεψης που ικανοποιεί τις συνθήκες της ζητούμενης ακολουθίας. Αρκεί δηλαδή ένα επιπλέον JOIN με τη σχέση session ώστε να γίνει περιορισμός και στις παραμέτρους της επίσκεψης.

Χρησιμοποιώντας το τελευταίο παράδειγμα για την εύρεση ακολουθιών της μορφής “ $a b^+ c$ ” με  $b \in (n1, n2, n3)$  θέτουμε τους ακόλουθους επιπλέον περιορισμούς:

1. το a να είναι η σελίδα με όνομα “index”
2. το c να είναι η σελίδα με όνομα “sm”
3. η χώρα προέλευσης των επισκεπτών να είναι η Γερμανία
4. οι επισκέπτες να προήλθαν από τη μηχανή αναζήτησης “[www.google.com](http://www.google.com)”

Οι ακολουθίες που ικανοποιούν αυτές τις συνθήκες βρίσκονται με την ερώτηση στο Πλαίσιο 4.10. Στη γραμμή 12 γίνεται ο περιορισμός για το όνομα της σελίδας a και c. Στη γραμμή 10 γίνεται το JOIN με τη σχέση session και στις γραμμές 13 και 14 γίνονται οι περιορισμοί για τη χώρα

προέλευσης και τη μηχανή προέλευσης.

```
1. SELECT a.session_id, a.sequence_order,
2. FROM
3.     page_impression a JOIN page_impression c ON
4.     a.session_id = c.session_id AND
5.     a.sequence_order < c.sequence_order - 1
6.
7.     JOIN action_perform ap ON
8.     c.impression_id = ap.impression_id
9.
10.    JOIN session on a.session_id = session.session_id
11. WHERE
12.     a.page_name = 'index' AND c.page_name='sm'
13.     AND session.country = 'Germany'
14.     AND session.referrer = 'www.google.com'
15.
16.     AND CONCAT(a.session_id, a.sequence_order) NOT IN
17.
18. (
19.     SELECT CONCAT(a.session_id, a.sequence_order)
20.     FROM
21.         page_impression a JOIN page_impression c ON
22.         a.session_id = c.session_id AND
23.         a.sequence_order < c.sequence_order - 1
24.
25.         JOIN action_perform ap ON
26.         c.impression_id = ap.impression_id
27.
28.     WHERE
29.         b NOT IN (n1, n2, n3)
30. )
```

*Πλαίσιο 4.10 Υπολογισμός ακολουθιών  $a b^* c$  που δεν καλύπτουν τους περιορισμούς για το wildcard*

#### 4.4.5 Σύνθετο Παράδειγμα Σεναρίου Χρήσης σε SQL92 και SQL99

Έστω ότι θέλουμε να απαντήσουμε στην ακόλουθη ερώτηση

Δώσε τις επισκέψεις και το μήκος των μονοπατιών των επισκέψεων όπου στο μονοπάτι συναντάται ακολουθία σελιδοπροβολών με τους ακόλουθους περιορισμούς:

- 1.υπάρχει προβολή της σελίδας “index” ακολουθούμενη από προβολή της σελίδας “hotels search form”.
- 2.Έπειτα απροσδιόριστος αριθμός προβολών σε σελίδες ανάμεσα από “searchresults” ή “select area” ή “hotels search form”.
- 3.Έπειτα υπάρχει προβολή της σελίδας “view sme” με action “present sme” και entity το αναγνωριστικό επιχειρήσεων διαμονής.

Η παραπάνω ερώτηση αναφέρεται στο σενάριο “αναζήτηση διαμονής” όπου γίνεται προβολή διαμονής (επιτυχής ολοκλήρωση). Μπορεί να αναπαρασταθεί με την ακολουθία “ $a b c+ d$ ” όπου η  $a$  είναι η σελίδα “index”, η  $b$  η σελίδα “hotels search form”, η  $c+$  τον απροσδιόριστο αριθμό σελιδοπροβολών που ανήκουν όμως σε συγκεκριμένο σύνολο ονομάτων, και η  $d$  αναπαριστά την εμφάνιση της σελίδα “view sme” με περιορισμούς στα action.

Χρησιμοποιώντας τη σημειογραφία για τις ακολουθίες των προηγούμενων ενοτήτων για τον προσδιορισμό του FROM CLAUSE της SQL, μπορούμε να εκφράσουμε την ερώτηση με τον τρόπο που φαίνεται στο Πλαίσιο 4.11. Θα αναφερόμαστε σε αυτή τη μορφή αναπαράστασης ως ψευδοSQL.

Η έκφραση  $d.action.entity\_id$  αναφέρεται στο attribute  $entity\_id$  που προκύπτει από το join της σχέσης  $page\_impression$  με τη σχέση  $action$ .

Στην ερώτηση στη γραμμή 8 περιορίζεται το  $entity\_id$  ώστε να ανήκει σε συγκεκριμένο σύνολο τιμών. Οι τιμές αυτές προκύπτουν από τρίτη πηγή δεδομένων.

```
1. select session_id, a.sequence_order, d.sequence_order - a.sequence_order
2. from a b c+ d
3. where
4. a.page_name='index'
5. and b.page_name='hotels form'
6. and c in ('searchresults', 'select area', 'hotels form')
7. and d.page_name='sme'
8. and d.action.entity_id in (valuelist)
9. and d.action.action_id='present sme'
```

*Πλαίσιο 4.11 Ερώτηση σε ψευδοSQL για σενάριο επιτυχούς αναζήτησης διαμονής*

Στη γραμμή 1, η έκφραση  $d.sequence\_order - a.sequence\_order$  εξάγει το μήκος του μονοπατιού, ενώ η έκφραση  $session\_id, a.sequence\_order$  προσδιορίζει την επίσκεψη και το σημείο έναρξης της ακολουθίας.

Στο Πλαίσιο 4.12 σελίδα 170 δίνεται η πλήρης υλοποίηση σε SQL 92, ακολουθώντας τη μεθοδολογία των προηγούμενων ενοτήτων.

1. αρχικά (γραμμές 1 έως 15) υπολογίζονται όλα τα μονοπάτια  $a b c+ d$ , με όλους τους περιορισμούς για τα  $a, b$  και  $d$  αλλά χωρίς περιορισμούς για το  $c$ .
2. ακολούθως (γραμμές 16 έως 32) υπολογίζονται όλα τα μονοπάτια που **δεν** καλύπτουν τη συνθήκη για το  $c$ . Αυτά αφαιρούνται από τα πρώτα. Η αφαίρεση γίνεται με βάση την ένωση της συμβολοσειράς  $session\_id$  με την  $a.sequence\_order$  (γραμμές 15 έως 17) ώστε να εξασφαλιστεί



ότι δε θα αφαιρεθούν μονοπάτια όπου υπάρχει ένα υπομονοπάτι που καλύπτει την ερώτηση και ένα άλλο υπομονοπάτι στην ίδια επίσκεψη που δεν καλύπτει την ερώτηση.

Σε πιο σύνθετες ερωτήσεις οι περιορισμοί σε σελίδες από wildcards, αυξάνουν σημαντικά την πολυπλοκότητα της ερώτησης. Το βασικό πρόβλημα είναι ότι είναι δύσκολο να εκφραστούν σε SQL92 οι ακολουθίες απροσδιόριστου μήκους, μιας και πρέπει να είναι γνωστό το ακριβές μήκος της κάθε ακολουθίας για να κατασκευαστεί με αντίστοιχο αριθμό JOIN μεταξύ της σχέσης page\_impression η ακολουθία.

Η SQL99 από την πλευρά της υποστηρίζει μηχανισμούς αναδρομικών ερωτήσεων με τη βοήθεια των οποίων επιτρέπεται η εύκολη έκφραση ακολουθιών. Επεκτείνοντας χάριν απλότητας τη σχέση page\_impression με attributes που δείχνουν στην προηγούμενη και επόμενη σελιδοπροβολή (prev\_page\_impressionid, next\_page\_impressionid) μπορούμε να εκφράσουμε την προηγούμενη ερώτηση για μονοπάτια αναζήτησης διαμονής σε SQL99 όπως φαίνεται στο Πλαίσιο 4.13 σελίδα 171.

Η βασική κατασκευή δημιουργεί στη γραμμή 1 ως 19 μία ακολουθία που ξεκινάει από το a και b (γραμμές 8 και 9) που ακολουθείται από σελίδα c με τους περιορισμούς που έχουν τεθεί (να είναι το όνομά της ανάμεσα σε σύνολο ονομάτων). Σε αυτή την βασική ακολουθία προστίθενται όλες οι επόμενες σελίδες που συνεχίζουν να ικανοποιούν τους περιορισμούς για τη c (γραμμές 12 έως 18).

Έχοντας αυτή την ακολουθία, αρκεί να βρεθούν οι επόμενες σελίδες που ικανοποιούν τους περιορισμούς για την d (γραμμές 23 έως 25).

## SQL92<sup>2</sup>

```
1. SELECT session.session_id, a.sequence_order as start_sequence_order,
2.       d.sequence_order as end_sequence_order,
3.       d.sequence_order - a.sequence_order as pathlength
4. FROM   session
5.       inner join page_impression as a on session.session_id=a.session_id
6.       inner join page_impression as b on a.session_id=b.session_id
7.       inner join page_impression as d on b.session_id=d.session_id
8.       inner join action_performed as act on d.page_impressionid=act.page_impressionid
9. WHERE  a.page_name='index'
10.      and b.page_name='hotels form'
11.      and d.page_name='sme'
12.      and act.actionid='present sme'
13.      and act.entityid in (valuelist)
14.      and a.sequence_order = b.sequence_order-1
15.      and b.sequence_order+1 < d.sequence_order
16.      and concat(a.session_id, a.sequence_order) not in
17. (
18. SELECT concat(a.session_id, a.sequence_order)
19. FROM   session
20.       inner join page_impression as a on session.session_id=a.session_id
21.       inner join page_impression as b on a.session_id=b.session_id
22.       inner join page_impression as notc on b.session_id=notc.session_id
23.       inner join page_impression as d on notc.session_id=d.session_id
24.       inner join action_performed as act on d.page_impressionid=act.page_impressionid
25. WHERE  a.page_name='index'
26.      and b.page_name='hotels form'
27.      and notc.page_name not in ('searchresults', 'select area', 'hotels form')
28.      and d.page_name='sme'
29.      and act.actionid='present sme'
30.      and act.entityid in (valuelist)
31.      and a.sequence_order = b.sequence_order-1
32.      and b.sequence_order+1 < d.sequence_order
33.      and notc.sequence_order > b.sequence_order
34.      and notc.sequence_order < d.sequence_order
35.)
```

*Πλαίσιο 4.12 Ερώτηση σε SQL92 για σενάριο επιτυχούς αναζήτησης διαμονής*

2 Να σημειωθεί ότι το παράδειγμα της SQL92 υπολογίζει και τα μονοπάτια της μορφής  $a \ b \ c + d \ (c + d) *$  το οποίο όμως δεν αποτελεί πρόβλημα διότι λύνεται εύκολα επιλέγοντας μόνο τις επισκέψεις με  $\text{MIN}(d.\text{page\_order})$ .

## SQL99

```
1. WITH RECURSIVE Sequence(from_page_impressionid, to_pageimpressionid) AS
2. (
3.     (
4.         SELECT cur.page_impressionid as from_page_impressionid, cur.next_page_impressionid as to_page_impressionid
5.         FROM     page_impression as prev
6.                inner join page_impression as cur  on prev.page_impressionid = cur.prev_page_impressionid
7.                inner join page_impression as next on cur.next_page_impressionid = next.page_impressionid
8.         WHERE   cur.page_name = 'hotels form'
9.                AND prev.page_name = 'index'
10.               AND next.page_name in ('searchresults', 'select area', 'hotels form')
11.     )
12.     UNION ALL
13.     (
14.         SELECT S.from_page_impressionid, P.page_impressionid as to_page_impressionid
15.         FROM Sequence S, page_impression P
16.         WHERE S.to_page_impressionid = P.prev_page_impressionid
17.                AND P.pageid in ('searchresults', 'select area', 'hotels form')
18.     )
19.)
20. SELECT S.session_id, d.sequence_order - S.sequence_order as pathlength
21. FROM   Sequence as S inner join page_impression as d on S.to_pageimpressionid=d.prev_page_impressionid
22.        inner join action_performed as act on d.page_impressionid=act.page_impressionid
23. WHERE  d.pageid = 'sme'
24.        AND act.entity_id in (valuelist)
25.        AND act.action_id = 'present sme'
```

*Πλαίσιο 4.13 Ερώτηση σε SQL99 για σενάριο επιτυχούς αναζήτησης διαμονής*

#### 4.4.6 Δυναμικά Σενάρια Χρήσης

Οι ακολουθίες για τις οποίες δόθηκε ο τρόπος εντοπισμού και εξαγωγής χαρακτηριστικών της επίσκεψης, αν και περιελάμβαναν wildcards, ήταν σε μεγάλο βαθμό προσδιορισμένες. Για τη μελέτη των σεναρίων χρήσης ενδιαφέρει όμως και η περίπτωση όπου το σενάριο δεν ολοκληρώνεται με τον τρόπο που έχει προβλεφθεί στην κατασκευή του ιστοχώρου. Και ενδιαφέρει να γνωρίζουμε τους λόγους που οδήγησαν σε αυτή την μη προβλεπόμενη συμπεριφορά.

Στο σενάριο αναζήτησης διαμονής που χρησιμοποιήθηκε στα προηγούμενα παραδείγματα, υποθέτουμε ότι οι επισκέπτες μετά την αρχική προβολή των αποτελεσμάτων αναζήτησης καταλήγουν σε παρουσίαση διαμονής όπου και ολοκληρώνεται το σενάριο. Τι γίνεται όμως στην περίπτωση που ποτέ δε φτάσουν σε παρουσίαση διαμονής; Το πραγματικό ερώτημα που πρέπει να απαντηθεί είναι “γιατί οι επισκέπτες δεν ολοκλήρωσαν το σενάριο;”.

Η απάντηση σε αυτό το ερώτημα δεν είναι μονοσήμαντη αλλά μπορεί να προσεγγιστεί μελετώντας τις σελιδοπροβολές πριν την εγκατάλειψη του σεναρίου και μετά. Από τις σελιδοπροβολές πριν τη εγκατάλειψη του σεναρίου και το ακριβές τους περιεχόμενο δίνονται ενδείξεις για τους λόγους που οδήγησαν τους επισκέπτες στην διακοπή του σεναρίου χρήσης. Από τις σελιδοπροβολές που ακολούθησαν μπορούμε να καταλάβουμε ποιές ήταν οι επόμενες κινήσεις των επισκεπτών και να προσεγγίσουμε έτσι τον τρόπο σκέψης τους. Ανάλογα με το περιεχόμενο του ιστοχώρου μπορεί να χρειαστεί να υποβληθούν αλληπάλλληλες ερωτήσεις που θα αναλύουν βήμα βήμα τις κινήσεις των επισκεπτών. Παράδειγμα τέτοιας ανάλυσης θα δοθεί στο επόμενο κεφάλαιο όπου θα μελετηθεί ένα σενάριο χρήσης σε έναν πραγματικό ιστοχώρο. Σ εαυτό το σημείο θα παρουσιαστεί μόνο ο τρόπος με τον οποίο μπορεί να εξαχθούν στοιχεία για τα χαρακτηριστικά προηγούμενων και επόμενων της εγκατάλειψης σελιδοπροβολών.

Α πάρουμε πάλι ως παράδειγμα την ακολουθία “ $a b c + d$ ” όπου η  $a$  είναι η σελίδα “index”, η  $b$  η σελίδα “hotels search form”, η  $c +$  τον απροσδιόριστο αριθμό σελιδοπροβολών που ανήκουν όμως σε συγκεκριμένο σύνολο ονομάτων  $S$ , και η  $d$  αναπαριστά την εμφάνιση της σελίδα “view sme” που ολοκληρώνει το σενάριο.

Στην περίπτωση που δεν ολοκληρώνεται το σενάριο, μας ενδιαφέρει να γνωρίζουμε την τελευταία σελίδα πριν την εγκατάλειψη του σεναρίου. Η ακολουθία που πρέπει να αναζητηθεί είναι η ακόλουθη: “ $a b c * d e$ ” όπου η  $e \notin S$ ,  $e \neq view\ sme$  και  $d \in S$ . Η  $e$  είναι οπότε η πρώτη σελίδα που είναι εκτός του “κανονικού” μονοπατιού. Οπότε η  $d$  είναι η τελευταία σελίδα μέσα στο κανονικό μονοπάτι. Με χρήση της μεθοδολογίας των προηγούμενων ενοτήτων μπορεί εύκολα να γίνει εξαγωγή των χαρακτηριστικών της  $d$ .

Στο Πλαίσιο 4.14 φαίνεται σε ψευδοSQL η ερώτηση που εξάγει το μήκος των μονοπατιών που οδηγούν σε εγκατάλειψη του σεναρίου, το όνομα της τελευταίας σελίδας εντός του κανονικού μονοπατιού, το περιεχόμενο της τελευταίας σελίδας μέσω της αναφοράς στο αναγνωριστικό της σελιδοπροβολής (pageimpression\_id), καθώς και το όνομα της επόμενης σελίδας.

```
1. select session_id, a.sequence_order,
2.         d.sequence_order - a.sequence_order as length,
3.         d.pageimpression_id
4.         d.page_name,
5.         e.page_name
6. from a b c* d e
7. where
8. a.page_name='index'
9. and b.page_name='hotels form'
10. and c in ('searchresults', 'select area', 'hotels form')
11. and d in ('searchresults', 'select area', 'hotels form')
12. and e not in ('searchresults', 'select area', 'hotels form', 'view sme')
```

*Πλαίσιο 4.14 Ερώτηση σε ψευδοSQL για σενάριο επιτυχούς αναζήτησης διαμονής*

## 4.5 Ανακεφαλαίωση

Σε αυτό το κεφάλαιο παρουσιάστηκε μία μεθοδολογία που οδηγεί στην λεπτομερή ανάλυση της επισκεψιμότητας σε έναν ιστοχώρο με βάση τα δεδομένα από το σύστημα καταγραφής της χρήσης ιστοχώρων που αναπτύχθηκε στα πλαίσια της εργασίας.

Η μεθοδολογία επικεντρώνεται σε πέντε διαφορετικές πλευρές που ενδιαφέρουν κατά τη λειτουργία ενός ιστοχώρου που είναι:

- Γενικά χαρακτηριστικά για την κίνηση στον ιστοχώρο και τις προτιμήσεις των επισκεπτών
- Ομαδοποίηση επισκεπτών με σκοπό της προσωποποίηση του ιστοχώρου
- Δυναμικά πειράματα για μελέτη χρησιμοποίησιμότητας
- Ανάλυση σεναρίων χρήσης

Σε κάθε ενότητα προτείνεται μία συστηματική προσέγγιση στο πρόβλημα που αναλύεται και παρουσιάζονται συγκεκριμένες ερωτήσεις σε γλώσσα SQL που εφαρμόζεται απ' ευθείας στο σχήμα της Βάσης Χρήσης.

Αποδεικνύεται ότι η εξαγωγή πληροφορίας για τη συμπεριφορά των επισκεπτών, ακόμα και αν πρόκειται για σύνθετες κινήσεις, είναι εφικτή με τυποποιημένες ερωτήσεις SQL, γεγονός που διευρύνει σημαντικά τους τομείς εφαρμογής της μεθοδολογίας.

## **5 Μελέτες Χρησιμοποιησιμότητας Ιστοχώρων με Παραδοσιακές Μεθόδους και με τη Μεθοδολογία Καταγραφής και Ανάλυσης Δεδομένων Χρήσης Ιστοχώρων**

Σε αυτό το κεφάλαιο θα παρουσιαστούν δύο μελέτες χρησιμοποιοησιμότητας (usability) εφαρμογών ιστοχώρου. Ο στόχος των μελετών αυτών είναι να γίνει σύγκριση των δυνατοτήτων και των αποτελεσμάτων που έχουν οι μεθοδολογίες που είναι βασισμένες στη χρησιμοποιοησιμότητα σε σχέση με τις μεθοδολογίες που είναι βασισμένες στην ανάλυση χρήσεως στο Web (Web Usage Analysis). Για λόγους συμβατότητας και δυνατότητας σύγκρισης των αποτελεσμάτων η ανάλυση έγινε στον ίδιο ιστοχώρο τουριστικού προορισμού Crete TOURnet [43]. Για λόγους πληρότητας ο ιστοχώρος περιγράφεται αναλυτικά στο Παράρτημα Α.

Η πρώτη μελέτη προσεγγίζει τον ιστοχώρο με παραδοσιακές μεθόδους για τη μελέτη χρησιμοποιοησιμότητας (usability) εφαρμογών και περιλαμβάνει ευριστική αξιολόγηση από ειδικούς στον τομέα (heuristic evaluation) και πειράματα χρηστών (user testing).

Η δεύτερη μέθοδος βασίζεται αποκλειστικά στα δεδομένα από πραγματική χρήση του ιστοχώρου, από τη Βάση Χρήσης που συλλέχθηκαν από το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της εργασίας. Εφαρμόζεται η μεθοδολογία που παρουσιάστηκε στο προηγούμενο κεφάλαιο για την εξαγωγή γενικότερων συμπερασμάτων για τον ιστοχώρο και ειδικότερα για ζητήματα χρησιμοποιοησιμότητας.

Αρχικά θα παρουσιαστούν οι στόχοι του ιστοχώρου και τα σενάρια χρήσης που προβλέφθηκαν κατά την ανάπτυξή του.

Θα ακολουθήσει η παρουσίαση της μελέτης με παραδοσιακές μεθόδους που χωρίζεται στην ευριστική ανάλυση και στα πειράματα χρηστών.

Στη συνέχεια θα παρουσιαστεί η μελέτη με τη μεθοδολογία ανάλυσης με το σύστημα καταγραφής.

Στο τέλος του κεφαλαίου θα επιχειρηθεί η αντιπαράθεση των αποτελεσμάτων από τις δύο προσεγγίσεις με σκοπό την αξιολόγηση της αποτελεσματικότητας της δεύτερης, και τη διερεύνηση της συμπληρωματικότητας και της συνεργατικότητάς των.

### **5.1 Σενάρια Χρήσης Εφαρμογής Ιστοχώρου**

Βασικός στόχος του προγράμματος Cretan TOURnet ήταν η υποστήριξη μικρομεσαίων

επιχειρήσεων τουρισμού της Περιφέρειας Κρήτης με την προβολή του τουριστικού προϊόντος που προσφέρουν στο Διαδίκτυο.

Σε αυτά τα πλαίσια, ο ιστοχώρος που αναπτύχθηκε επιχειρεί να διευκολύνει το δυνητικό επισκέπτη του νησιού στην εύρεση των υπηρεσιών τουριστικού χαρακτήρα που επιθυμεί. Στη συγκεκριμένη υλοποίηση του ιστοχώρου, αυτό περιλαμβάνει υπηρεσίες που προσφέρονται από 3 κυρίως κατηγορίες επιχειρήσεων:

- διανυκτέρευση
- ενοικίαση οχημάτων
- συνδυασμένες υπηρεσίες από γραφεία τουρισμού

Υποστηρικτικά με τις παραπάνω κατηγορίες επιχειρήσεων, έχουν ενσωματωθεί επίσης επιχειρήσεις με υπηρεσίες που είναι συνήθως απαραίτητες μετά την άφιξη του επισκέπτη στο νησί, όπως είναι καταστήματα με αναμνηστικά είδη ή παραδοσιακά προϊόντα, υπηρεσίες διατροφής, διασκέδαση κ.α.

Για την επιλογή της κατάλληλης επιχείρησης ο επισκέπτης του ιστοχώρου υποστηρίζεται με αναλυτικές πληροφορίες για της περιοχές ενδιαφέροντος και τα αξιοθέατα που υπάρχουν στο νησί, καθώς και με πληροφορίες γενικότερου ενδιαφέροντος όπως της ενότητας του ηλεκτρονικού οδηγού.

Από τα παραπάνω μπορεί να εξαχθούν πολλά σενάρια χρήσης του ιστοχώρου που αναφέρονται τόσο στο γενικότερο σκοπό του προγράμματος και χαρακτηρίζονται ως πρωταρχικά όσο και σε δευτερεύουσες ανάγκες (σε σχέση με τους στόχους του προγράμματος) των δυνητικών επισκεπτών του ιστοχώρου. Τα σενάρια αυτά μπορεί να ακολουθηθούν είτε αυτόνομα είτε σε συνδυασμό με άλλα σενάρια χρήσης.

Στη συνέχεια θα περιγραφούν ορισμένα από τα πιο σημαντικά σενάρια και οι παράμετροι που τα επηρεάζουν, καθώς και η ακολουθία των σελιδοπροβολών που μπορεί να ακολουθηθεί. Όπου δεν αναφέρεται διαφορετικά, η ακολουθία σελιδοπροβολών αναφέρεται στη δεύτερη έκδοση του ιστοχώρου (βλέπε Παράρτημα Α).

Ωστόσο είναι σημαντικό να τονιστεί ότι παρόλο που στα επόμενα διαγράμματα αναφέρονται σελίδες εκκίνησης των πιθανών ακολουθιών, από τη φύση του ο παγκόσμιος ιστός επιτρέπει την εκκίνηση από ενδιάμεσες σελίδες.

## 5.1.1 Πρωταρχικά Σενάρια Χρήσης

### 5.1.1.1 Επιλογή Διαμονής

Το σενάριο για την επιλογή διαμονής αναφέρεται στη διαδικασία εύρεσης κάποιου καταλύματος. Η διαδικασία αυτή μπορεί να έχει πολλές αφετηρίες, ανάλογα με τις επιλογές που έχει κάνει ο χρήστης πριν την επίσκεψη στον ιστοχώρο. Μπορεί να διακριθούν οι ακόλουθες γενικές περιπτώσεις:

1. ο χρήστης γνωρίζει ήδη το συγκεκριμένο κατάλυμα που θέλει να επιλέξει (για παράδειγμα από προηγούμενες επισκέψεις στο νησί ή από κάποιο έντυπο όπου αναφερόταν) και απλώς θέλει περισσότερες πληροφορίες ή στοιχεία επικοινωνίας
2. ο χρήστης γνωρίζει την περιοχή της Κρήτης όπου θέλει να διαμένει (για παράδειγμα επειδή θέλει να συμμετάσχει σε κάποιο συνέδριο που διεξάγεται στην περιοχή ή θέλει να επισκεφτεί αξιοθέατα της περιοχής) και επιθυμεί να επιλέξει ανάμεσα στα καταλύματα της περιοχής
3. ο χρήστης θέλει να διερευνήσει τις επιλογές που έχει σε διαμονή στην Κρήτη γενικότερα

Τα βασικά εργαλεία που διατίθενται στο χρήστη είναι

- ένα πεδίο για αναζήτηση κειμένου γενικά στον ιστοχώρο που είναι εμφανές από όλες τις σελίδες του ιστοχώρου (Εικόνα 11.6)
- ένα πεδίο αναζήτησης στον τίτλο επιχειρήσεων διαμονής από την αρχική σελίδα του ιστοχώρου (Εικόνα 11.6)
- Ένας “ενεργός” χάρτης της Κρήτης στην αρχική σελίδα, που με κλικ σε αυτόν εμφανίζει πληροφορίες για την περιοχή γύρω από το σημείο που επιλέχθηκε (Εικόνα 11.6)
- μία φόρμα αναζήτησης διαμονής προσβάσιμη από σύνδεσμο στην αρχική σελίδα ή από “συντομεύσεις” (Εικόνα 11.6) του ιστοχώρου προσβάσιμες από όλες τις σελίδες του ιστοχώρου. Η φόρμα αναζήτησης προσφέρει δυνατότητες για επιλογή περιοχής και βασικές διευκολύνσεις του καταλύματος (Εικόνα 11.7). Ειδικά η επιλογή περιοχής γίνεται με μία ειδική ενδιάμεση σελίδα σε μορφή ξεχωριστού παραθύρου που ανοίγει χωρίς να εγκαταλειφθεί η τρέχουσα σελίδα. Αυτό σημαίνει ότι ο χρήστης ενώ είναι στη σελίδα με τη φόρμα αναζήτησης διαμονής, πατώντας τη σχετική επιλογή του ανοίγεται ένα νέο παράθυρο περιορισμένων διαστάσεων, απ’ όπου μπορεί με διάφορους τρόπους να επιλέξει μία ή περισσότερες περιοχές. Η επιλογή γίνεται είτε μέσω χάρτη όπου εμφανίζονται όλες οι περιοχές σε συγκεκριμένη διάμετρο από το σημείο που επιλέχθηκε στο χάρτη, είτε αναζητώντας με βάση το όνομα της περιοχής, είτε από το

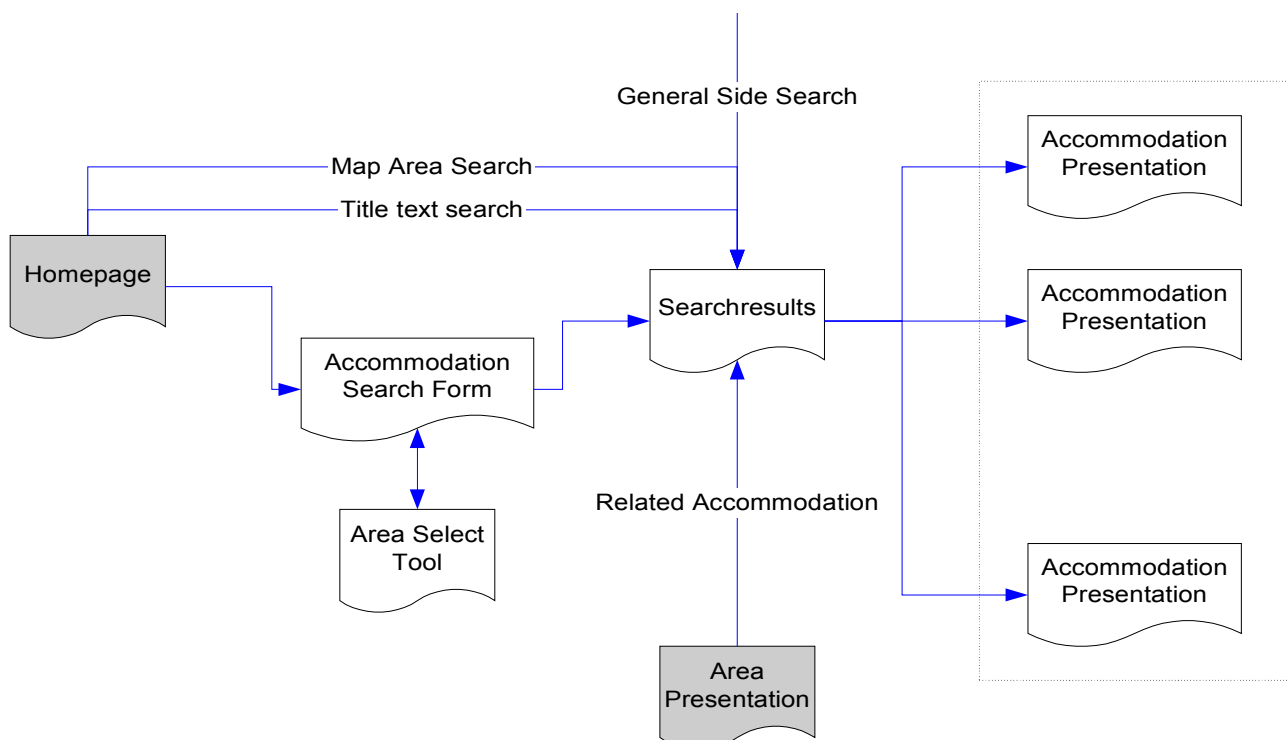


“ηλεκτρονικό φυλλάδιο” που θα περιγραφεί παρακάτω. Στιγμιότυπο φαίνεται στην Εικόνα 11.10

- από τις σελίδες παρουσίασης περιοχών ενδιαφέροντος, σύνδεσμος για τις επιχειρήσεις διαμονής της περιοχής (Εικόνα 11.8)

Η παρουσίαση των καταλυμάτων γίνεται σε πέντε διαφορετικές σελίδες. Στην πρώτη (Εικόνα 11.9) υπάρχει μία γενική περιγραφή κειμένου και μία φωτογραφία. Στη δεύτερη υπάρχουν πληροφορίες για την τοποθεσία καθώς και στοιχεία επικοινωνίας. Στην τρίτη και τέταρτη σελίδα υπάρχουν στοιχεία για διάφορες πτυχές της εγκατάστασης, όπως διευκολύνσεις δωματίων ή χαρακτηριστικά παραλίας. Στην τελευταία σελίδα υπάρχουν περισσότερες φωτογραφίες.

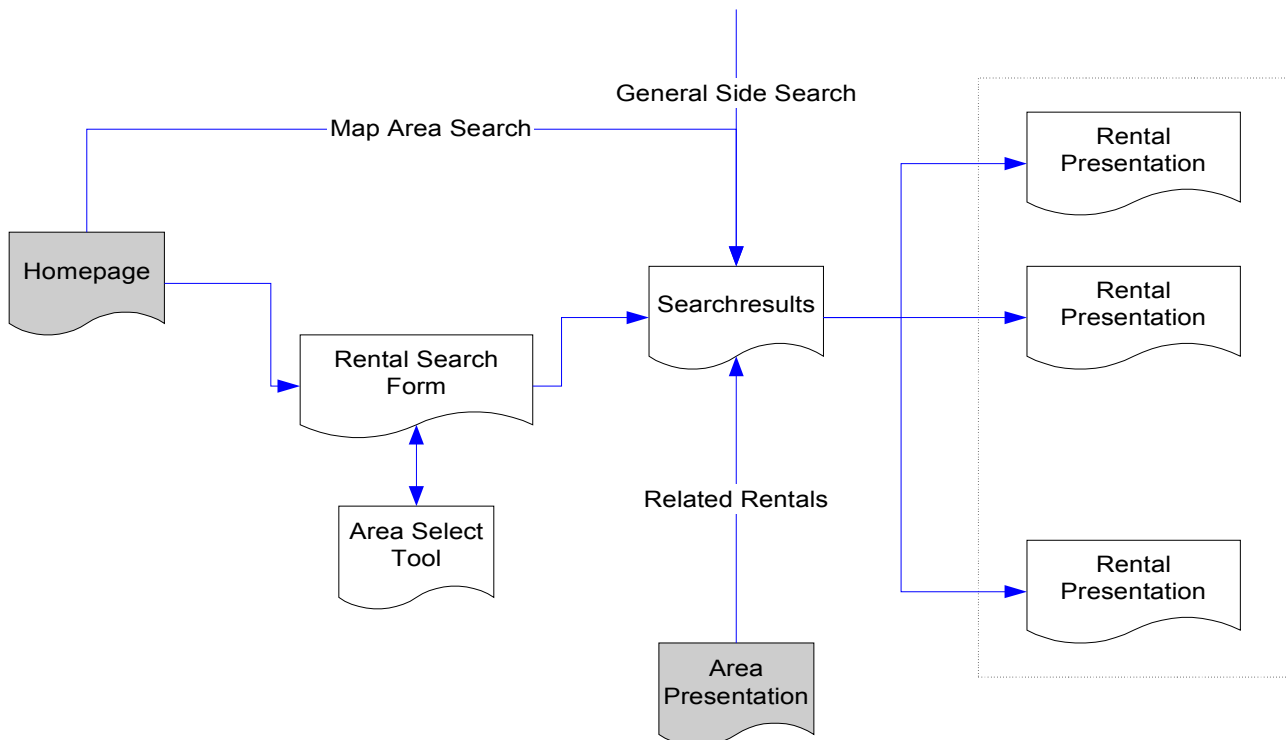
Από τα παραπάνω προκύπτουν τα μονοπάτια σελιδοπροβολών που φαίνονται στο Σχήμα 5.1.



Σχήμα 5.1 Πιθανά μονοπάτια σελιδοπροβολών για την επιλογή διαμονής. Η εκκίνηση γίνεται είτε από την αρχική σελίδα, είτε από την παρουσίαση μιας περιοχής ενδιαφέροντος. Ο τερματισμός γίνεται με την παρουσίαση καταλυμάτων

### 5.1.1.2 Επιλογή Ενοικίασης Οχήματος

Η επιλογή ενοικίασης οχημάτων έχει σχεδιαστεί με ακριβώς την ίδια φιλοσοφία όπως και η επιλογή διαμονής, με μόνη διαφορά το γεγονός ότι δεν υπάρχει στην αρχική σελίδα του ιστοχώρου το πεδίο για αναζήτηση στον τίτλο των επιχειρήσεων. Το σχετικό διάγραμμα φαίνεται στο Σχήμα 5.2.



Σχήμα 5.2 Όμοια με το Σχήμα 5.1 στην επιλογή ενοικίασης οχήματος ακολουθούνται τα ίδια μονοπάτια, με εξαίρεση την αναζήτηση στον τίτλο από την αρχική σελίδα

### 5.1.1.3 Επιλογή Γραφείου Τουρισμού

Όπως και η επιλογή ενοικίασης οχημάτων, έτσι και για τα γραφεία τουρισμού ισχύει η ακριβώς ίδια σχεδίαση. Οι ακολουθίες σελιδοπροβολών είναι οι ακριβώς ίδιες όπως και με τα ενοικιαζόμενα οχήματα και παραλείπονται.

### 5.1.2 Δευτερεύοντα Σενάρια Χρήσης

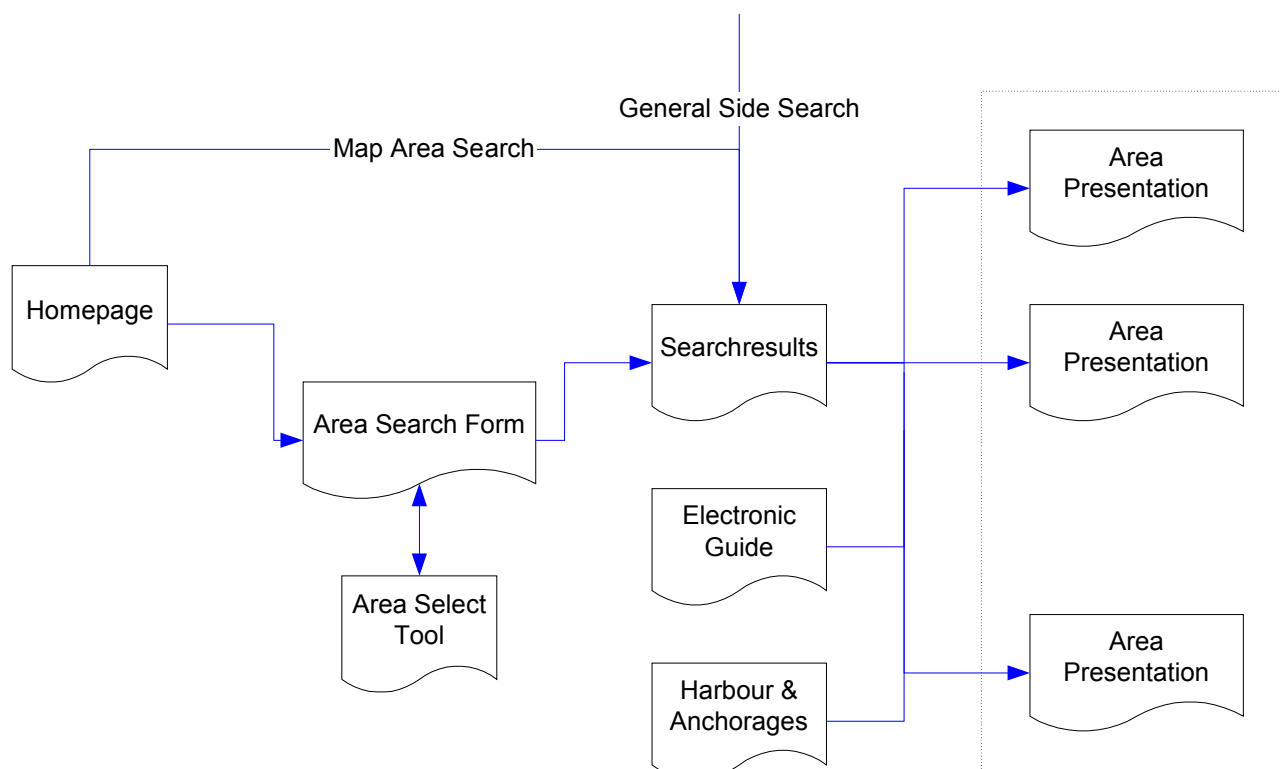
Όπως αναφέρθηκε σε προηγούμενη παράγραφο, ο ιστοχώρος Crete TOURnet περιέχει μεγάλο όγκο πληροφορίας και λειτουργικότητας που δεν αναφέρεται άμεσα στους βασικούς στόχους του προγράμματος που ήταν η προβολή των μικρομεσαίων επιχειρήσεων τουρισμού, αλλά απλώς συμπληρώνουν τον ιστοχώρο προς της κατεύθυνση των ιστοχώρων τουριστικού προορισμού.

Τέτοια πληροφορία και λειτουργικότητα είναι η ενότητα για τα λιμάνια και αγκυροβόλια της Κρήτης, ο εκτενής ηλεκτρονικός οδηγός και κυρίως η πληροφορία για τις περιοχές ενδιαφέροντος και τα αξιοθέατα της Κρήτης.

Ουσιαστικά όλη η πληροφορία που παρουσιάζεται στον ιστοχώρο επικεντρώνεται γύρω από την έννοια της περιοχής ενδιαφέροντος και όλες οι επιχειρήσεις αλλά και για παράδειγμα τα λιμάνια έχουν αντιστοιχία σε μία τέτοια περιοχή.

Στα διαγράμματα των σχημάτων 5.1 και 5.2 αυτό το γεγονός αντικατοπτρίζεται από το σημείο εκκίνησης της παρουσίασης περιοχής ενδιαφέροντος. Έχοντας φτάσει σε αυτό το σημείο, μπορεί να επιλεχτεί η εμφάνιση της πληροφορίας για επιχειρήσεις που έχουν αντιστοιχιστεί σε αυτή την περιοχή.

Η ακολουθία σελιδοπροβολών που οδηγεί στην παρουσίαση περιοχής ενδιαφέροντος δεν διαφέρει πολύ από τις ανάλογες ακολουθίες των προηγούμενων σεναρίων και παρουσιάζεται στο Σχήμα 5.3.



Σχήμα 5.3 Πιθανές ακολουθίες σελιδοπροβολών για την κατάληξη σε παρουσίαση περιοχής ενδιαφέροντος

### 5.1.3 Συνδυασμός Σεναρίων Χρήσης Ιστοχώρου

Πρέπει να έχει γίνει ήδη εμφανές από τις προηγούμενες παραγράφους ότι η πραγματική χρήση του ιστοχώρου δεν περιορίζεται στα στενά όρια που θέτουν τα σενάρια χρήσης που περιγράφηκαν. Η πραγματική αξία του ιστοχώρου προέρχεται από την ευελιξία που προσφέρει στην αναζήτηση πληροφοριών για τον προορισμό της Κρήτης.

Σε αυτά τα πλαίσια η ανάμιξη και ο συνδυασμός των προηγούμενων σεναρίων είναι το στοιχείο που συμπληρώνει για τον τελικό χρήστη την υπηρεσία για τον τουριστικό προορισμό.

Έτσι δεν είναι σπάνιο χρήστες που αναζητούν το κατάλυμα για την παραμονή τους στην Κρήτη, να

θελήσουν ταυτόχρονα να νοικιάσουν κάποιο όχημα για την περιήγησή τους στο νησί. Ένα επίσης πραγματικό σενάριο είναι οι χρήστες να θελήσουν να παραμείνουν σε διαφορετικές τοποθεσίες στο ίδιο ταξίδι.

Όπως αναφέρθηκε στα δευτερεύοντα σενάρια, η έννοια της περιοχής ενδιαφέροντος περικλείει τόσο τα αξιοθέατα μιας περιοχής όσο και τις επιχειρήσεις που δραστηριοποιούνται σε αυτήν. Οπότε προσφέρεται ως σημείο αναφοράς για την αναζήτηση υπηρεσιών από τους επισκέπτες του ιστοχώρου, το οποίο είναι εμφανές από τα προηγούμενα διαγράμματα όπου παρουσιάζεται ως σημείο εκκίνησης για την επιλογή διαμονής, ενοικίασης οχήματος κλπ.

Ένα επιπλέον πρόβλημα που όμως προκύπτει είναι η λειτουργικότητα που προσφέρεται στο χρήστη να επιλέξει από ένα σύνολο προεπιλεγμένων επιχειρήσεων ή περιοχών. Για παράδειγμα, στην αναζήτηση διαμονής ο χρήστης μπορεί να έχει επιλέξει έναν περιορισμένο αριθμό ξενοδοχείων ως πιθανά καταλύματα. Το πρόβλημα εντοπίζεται στη συγκεντρωμένη παρουσίαση αυτών. Αυτή η ανάγκη καλύπτεται με την έννοια του “Ηλεκτρονικού Φυλλαδίου” που είναι ένας νοητός χώρος όπου ο χρήστης περιηγούμενος στον ιστοχώρο προσθέτει τις σελίδες που τον ενδιαφέρουν. Όταν επισκεφτεί τη σελίδα του φυλλαδίου, παρουσιάζονται όλες οι σελίδες που έχει επιλέξει με αυτόν τον τρόπο ανά κατηγορία.

Στην εφαρμογή η λειτουργία του φυλλαδίου έχει υλοποιηθεί προσθέτοντας για κάθε αντικείμενο που παρουσιάζεται (ξενοδοχείο, περιοχή κλπ.) ένα ειδικό εικονίδιο, πιέζοντας το οποίο προσθέτει το αντικείμενο στο φυλλάδιο. Στην Εικόνα 11.9 σελίδα 280 το σχετικό εικονίδιο για το κατάλυμα φαίνεται δίπλα στον τίτλο του καταλύματος. Στην ίδια τοποθεσία εμφανίζεται το εικονίδιο για όλες τις κατηγορίες σελίδων, όπως περιοχές ενδιαφέροντος, γραφεία τουρισμού, ενοικιάσεις οχημάτων κλπ.

Από κάθε σελίδα του ιστοχώρου, υπάρχει σύνδεσμος στο ηλεκτρονικό φυλλάδιο όπου εμφανίζονται ανά κατηγορία όλα τα επιλεγμένα αντικείμενα. Σχετική εικόνα εμφανίζεται στην Εικόνα 11.12 σελίδα 283. Οι επιλογές που γίνονται με αυτόν τον τρόπο προσφέρονται στο χρήστη και σαν επιλογή όταν αναζητεί με βάση την περιοχή ενδιαφέροντος. Στην Εικόνα 11.11 σελίδα 282 φαίνεται ένα στιγμιότυπο από τη σχετική σελίδα, όπου εμφανίζονται οι επιλογές ενδιαφέροντος, είτε αυτές έχουν προστεθεί αυτόνομα στο φυλλάδιο, είτε αναφέρονται στις περιοχές άλλων αντικειμένων.

## **5.2 Μελέτη Χρησιμοποίησιμότητας με Παραδοσιακές Μεθόδους**

Για τη μελέτη χρησιμοποίησιμότητας με παραδοσιακές μεθόδους επιλέχθηκε η χρήση των μεθόδων

της ευριστικής αξιολόγησης (heuristic evaluation) από ειδικούς στην ανάλυση χρησιμοποίησιμότητας σε συνδυασμό με πειράματα χρηστών (user tests). Ο λόγος είναι ότι αυτές οι δύο μέθοδοι θεωρούνται από τις πιο αποτελεσματικές στην ανακάλυψη προβλημάτων χρησιμοποίησιμότητας ενώ συνδυάζουν την εμπειρία ειδικών με τις ανάγκες που προκύπτουν από την πραγματική χρήση.

### 5.2.1 Ευριστική Αξιολόγηση

Η ευριστική αξιολόγηση αποσκοπεί στον εντοπισμό προβλημάτων χρησιμοποίησιμότητας εφαρμογών, από ειδικούς σε τέτοια θέματα επαγγελματίες ή επιστήμονες. Στην ευριστική αξιολόγηση ο στόχος δεν είναι η πρόταση λύσεων για τα προβλήματα που εντοπίζονται, αλλά ο κύριος σκοπός βρίσκεται καθαρά στον εντοπισμό αυτών.

Στην ευριστική αξιολόγηση, όπως περιγράφεται και στην παράγραφο 2.2.3.3.2, οι αξιολογητές έχοντας υπ' όψιν καθιερωμένους κανόνες χρησιμοποίησιμότητας (style guides, usability guides) ελέγχουν το interface της εφαρμογής που εξετάζεται για να εντοπίσουν στοιχεία που δεν ανταποκρίνονται σε αυτούς τους κανόνες.

Για την εφαρμογή ιστοχώρου διενεργήθηκαν 3 ανεξάρτητες αξιολογήσεις από δύο ειδικούς στο χώρο και το γράφοντα. Η αξιολόγηση έγινε στην έκδοση του ιστοχώρου όπως αυτός διαμορφώθηκε μετά την αναμόρφωσή του Απριλίου του 2003 (βλέπε Παράρτημα Α για τα στάδια ανάπτυξης της εφαρμογής). Στο Παράρτημα Γ υπάρχει παράθεση των γραπτών οδηγιών που δόθηκαν στους 2 πρώτους αξιολογητές. Τα αποτελέσματα της αξιολόγησης παραδόθηκαν υπό μορφής λίστας εντοπισμένων προβλημάτων, ενώ ακολούθησε και σχετική συζήτηση μεταξύ του γράφοντα και τους 2 πρώτους.

Για κάθε πρόβλημα που εντοπίστηκε έγινε και μία αξιολόγηση για τη σοβαρότητά του σε μία κλίμακα από 0 έως 4 που ακολουθεί την πρόταση από το [4]. Η σημασία της κάθε βαθμολογίας δίνεται στον πίνακα 5.1.

<b>Βαθμός</b>	<b>Σημασία</b>
0	Δεν πρόκειται για πρόβλημα χρησιμοποίησιμότητας
1	Αισθητικό πρόβλημα: δεν απαιτείται αντιμετώπιση εκτός και αν υπάρχει διαθέσιμος χρόνος
2	Μικρό πρόβλημα: πρέπει να αντιμετωπιστεί αλλά με χαμηλή προτεραιότητα
3	Μείζων πρόβλημα: είναι σημαντικό να αντιμετωπιστεί με μεγάλη προτεραιότητα
4	Καταστροφικό πρόβλημα: πρέπει οπωσδήποτε να αντιμετωπιστεί

Πίνακας 5.1 κλίμακα αξιολόγησης προβλημάτων χρησιμοποίησιμότητας

Από τις ευριστικές αξιολογήσεις εντοπίστηκαν συνολικά 12 επιμέρους προβλήματα στον ιστοχώρο με εύρος βαθμολογίας από 1 έως 3. Δεν υπήρξε κανένα με βαθμολογία 4.

Από τα συνολικά 12 προβλήματα τα 6 κάλυπταν τις ίδιες περιοχές του ιστοχώρου ενώ βαθμολογήθηκαν με 2 και 3.

Τα ζητήματα με βαθμολογία 1 περιλαμβάνουν θέματα όπως

- καλύτερη τοποθεσία σχετικών επιχειρήσεων τουρισμού στις σελίδες παρουσίασης περιοχών (Εικόνα 11.8)
- προσθήκη συνδέσμου στο εικονίδιο ένδειξης ύπαρξης ξεχωριστού ιστοχώρου για τις επιχειρήσεις που διαθέτουν
- θέματα συνέπειας στις ονομασίες των ενοτήτων (π.χ. “Διαμονή” έναντι “Ξενοδοχεία & Διαμερίσματα”)
- τοποθεσία συνδέσμου στα εικονίδια ενοτήτων της αρχικής σελίδας

Στα υπόλοιπα προβλήματα μπορεί να διακριθούν δύο κυρίως πόλοι γύρω από τους οποίους εντοπίστηκαν: Α) τη σελίδα επιλογής περιοχής στις φόρμες αναζήτησης (Εικόνα 11.10, Εικόνα 11.11) και Β) τη λειτουργία αναζήτησης καθ' αυτή.

Για κάθε ένα από τα προβλήματα με βαθμολογία 2 ή 3 θα δοθεί μία πιο αναλυτική περιγραφή στη συνέχεια.

### **5.2.1.1 Πρόβλημα Χρησιμοποιησιμότητας Αποτελεσμάτων Αναζήτησης**

Η αναζήτηση πληροφορίας για τον δυνητικό επισκέπτη της Κρήτης είναι από τις κύριες υπηρεσίες που προσφέρει ο ιστοχώρος. Από αυτή την άποψη ένα καλά χρησιμοποιήσιμο περιβάλλον διεπαφής (interface) για αυτή την υπηρεσία έχει ιδιαίτερα μεγάλη σημασία. Παρόλα αυτά ακριβώς σε αυτό το σημείο εντοπίστηκαν τα πιο σοβαρά προβλήματα χρησιμοποιοιμότητας από την ευριστική αξιολόγηση.

Το πρώτο πρόβλημα εντοπίζεται στο ότι στα αποτελέσματα αναζήτησης δεν είναι πουθενά εμφανές τα κριτήρια στα οποία αναφέρονται τα αποτελέσματα. Δηλαδή ο χρήστης ουσιαστικά δε βλέπει για τι είδους πληροφορία έψαξε, όπως αυτό είναι ευρέως διαδεδομένο σε μηχανές αναζήτησης. Στην Εικόνα 11.13 φαίνεται χαρακτηριστικό απόσπασμα μίας τέτοιας σελίδας. Στο χρήστη παρουσιάζεται ο αριθμός των σελίδων που καλύπτουν τα κριτήρια αναζήτησης, αλλά δε φαίνεται τι έψαξε.

Επίσης δεν μπορεί ο χρήστης να συνεχίσει την αναζήτηση από τη σελίδα αναζήτησης, τροποποιώντας ενδεχομένως τα κριτήρια αναζήτησης. Και αυτή η δυνατότητα είναι ευρέως διαδεδομένη σε άλλες μηχανές. Ακόμα και αν δε δοθεί η δυνατότητα επόμενης αναζήτησης άμεσα από τα αποτελέσματα αναζήτησης, παρατηρείται ότι όταν ο χρήστης επιλέγει τη λειτουργία “back” της εφαρμογής πλοήγησης, τα κριτήρια που είχε θέσει στις φόρμες χάνονται εν μέρη. Χαρακτηριστικά έγινε η παρατήρηση ότι ανάλογα με την εφαρμογή πλοήγησης (Microsoft Internet Explorer, Mozilla, Opera) είτε χάνονται όλα τα κριτήρια που είχαν επιλεγεί, είτε χάνονται για παράδειγμα μόνο τα κριτήρια για την περιοχή αναζήτησης.

Τα παραπάνω προβλήματα δε γίνονται αντιληπτά εάν ο χρήστης του ιστοχώρου βρει με την πρώτη προσπάθεια την πληροφορία που αναζητεί. Στην περίπτωση όμως που ο επισκέπτης θέλει πραγματικά να ερευνησει στην πληροφορία που παρέχεται, οι λειτουργίες και το περιβάλλον διεπικοινωνίας είναι ιδιαίτερα εχθρικό προς αυτόν.

#### **5.2.1.2 Πρόβλημα Χρησιμοποίησιμότητας Επιλογής Περιοχής Αναζήτησης**

Η επιλογή της περιοχής στην οποία γίνεται αναζήτηση γίνεται με εξειδικευμένη σελίδα που εμφανίζεται όταν ο χρήστης επιλέγει την κατάλληλη λειτουργία από τις φόρμες αναζήτησης (χάρτης στην Εικόνα 11.7). Στο παράθυρο που εμφανίζεται δίνεται η δυνατότητα επιλογής περιοχών είτε μέσω χάρτη (Εικόνα 11.10) είτε από περιοχές στο “Φυλλάδιο μου” (Εικόνα 11.11) είτε με αναζήτηση ονόματος περιοχής.

Στην περίπτωση του χάρτη και της αναζήτησης με βάση το όνομα, ο χρήστης πρέπει να επιλέξει από τα αποτελέσματα της σχετικής αναζήτησης τις περιοχές που τελικά θέλει να προσθέσει ως κριτήριο.

Όλοι οι αξιολογητές συμφωνούν ότι το σύνολο του interface στο παράθυρο επιλογής περιοχής είναι ιδιαίτερα δυσνόητο και δύσχρηστο. Δεν παρέχεται στο χρήστη κανενός είδους οδηγίες για τον τρόπο με τον οποίο μπορεί να λειτουργήσει η σελίδα.

Ο χάρτης παρόλο που παρέχει παρόμοια λειτουργικότητα με το χάρτη της αρχικής σελίδας (εμφανίζει δηλαδή περιοχές που βρίσκονται σε ορισμένη απόσταση από το σημείο που κάνει κλικ στο χάρτη) δεν παρουσιάζει την ίδια εικόνα. Συγκεκριμένα, όταν κυλάει ο δείκτης του ποντικιού πάνω στο χάρτη, δεν εμφανίζεται ο αντίστοιχος κύκλος με την περιοχή που θα κάλυπτε η ενδεχόμενη αναζήτηση.

Αφού έχουν εμφανιστεί τα αποτελέσματα αναζήτησης από το χάρτη ή από την αναζήτηση ονόματος, ο χρήστης μένει με την εντύπωση ότι οι περιοχές που εμφανίστηκαν είναι αυτές που

τελικά έχουν επιλεγεί, αλλά στην πραγματικότητα πρέπει να προσθέσει τις περιοχές που επιθυμεί σε ειδική λίστα που τελικά χρησιμοποιείται στην αναζήτηση.

Τέλος, έγινε η παρατήρηση ότι επειδή η επιλογή περιοχής αναζήτησης είναι σημαντική στην αναζήτηση πληροφορίας των δυνητικών επισκεπτών, αυτή η λειτουργία δε θα έπρεπε να είναι προσβάσιμη σε ξεχωριστό παράθυρο αλλά άμεσα από τις φόρμες αναζήτησης.

### **5.2.1.3 Πρόβλημα Χρησιμοποιησιμότητας Περιοχής σε Χάρτη**

Το Crete TOURnet περιέχει αναλυτικούς χάρτες για όλη την Κρήτη. Όλα τα αντικείμενα που παρουσιάζονται και μπορεί να αντιστοιχιστούν σε περιοχές ενδιαφέροντος μπορεί να εμφανιστούν σε χάρτη. Για αυτό το σκοπό υπάρχει ειδικό εικονίδιο που εμφανίζει την Κρήτη με ένα κόκκινο σημείο, μέσω του οποίου ο χρήστης βλέπει έναν απλοποιημένο χάρτη με την τοποθεσία του αντικειμένου. Παράδειγμα του εικονιδίου φαίνεται στα αποτελέσματα αναζήτησης στην Εικόνα 11.13.

Ο ίδιος χάρτης φαίνεται επίσης όταν ο επισκέπτης επισκεφθεί την ενότητα “Τοποθεσία και Επικοινωνία” των επιχειρήσεων τουρισμού (Εικόνα 11.14). Το σημείο στο χάρτη ανήκει στην περιοχή ενδιαφέροντος του αντικειμένου που παρουσιάζεται και είναι ενεργό. Επιλέγοντας το σημείο ο χρήστης μεταφέρεται στη σελίδα παρουσίασης της αντίστοιχης περιοχής.

Επίσης ο υπόλοιπος χάρτης είναι ενεργός, που οδηγεί στην παρουσίαση του γεωγραφικού χάρτη της περιοχής.

Το πρόβλημα είναι ότι ανάλογα με τον τρόπο με τον οποίο ο χρήστης επιλέγει να δει την τοποθεσία στο χάρτη, δημιουργούνται διαφορετικές ανάγκες που πρέπει να καλυφθούν. Για παράδειγμα εάν ο χρήστης επιλέξει να δει την τοποθεσία μιας περιοχής στο χάρτη, είναι θεμιτό η επιλογή του ενεργού σημείου στο χάρτη να οδηγεί πίσω στην παρουσίαση της περιοχής. Εάν όμως το σημείο εκκίνησης ήταν μία επιχείρηση, ο χρήστης θα ανέμενε ότι επιλέγοντας το ενεργό σημείο θα οδηγηθεί στην παρουσίαση της επιχείρησης (επειδή έβλεπε την τοποθεσία της επιχείρησης).

Επίσης δεν υπάρχει πουθενά κάποια ένδειξη ότι πιέζοντας στο υπόλοιπο τμήμα του χάρτη (εκτός του ενεργού σημείου) θα μεταφερθεί στην παρουσίαση αναλυτικού χάρτη στο σημείο που επέλεξε. Ιδιαίτερα για τις επιχειρήσεις τουρισμού, αλλά και για τα υπόλοιπα τμήματα του ιστοχώρου, η εμφάνιση χάρτη της περιοχής είναι ιδιαίτερα σημαντικό στοιχείο.



#### **5.2.1.4 Πρόβλημα Χρησιμοποίησιμότητας Ηλεκτρονικού Φυλλαδίου**

Για το τμήμα του ιστοχώρου που άπτεται του ηλεκτρονικού φυλλαδίου έγιναν δύο κυρίως παρατηρήσεις:

α) η μεγαλύτερη προβολή της προσφερόμενης υπηρεσίας ώστε να γίνεται πιο εύκολα αντιληπτή από τους επισκέπτες

β) η παρουσίαση περισσότερης πληροφορίας για κάθε αντικείμενο που έχει προστεθεί στο φυλλάδιο (Εικόνα 11.12)

Ιδιαίτερα για το δεύτερο σημείο είναι σημαντικό να παρουσιάζεται πληροφορία για την τοποθεσία του αντικειμένου (όνομα περιοχής, τμήμα χάρτη) ώστε ο επισκέπτης να μπορεί να προσανατολίζεται με μεγαλύτερη ευκολία ανάμεσα στις επιλογές που έχει κάνει.

#### **5.2.2 Πειράματα Χρηστών (User Tests)**

Για τα πειράματα χρηστών χρησιμοποιήθηκε η μέθοδος “Thinking Aloud” όπου οι χρήστες καλούνται να εκφράζουν δυνατά τις σκέψεις τους κατά τη διάρκεια διενέργειας των εργασιών που τους δίνονται. Συμμετείχαν συνολικά 7 χρήστες και εκτέλεσαν ο καθένας από 10 εργασίες που τους δόθηκαν στον αναμορφωμένο ιστοχώρο. Το αναλυτικό Σχέδιο Πειράματος Χρηστών (User Test Plan) παρατίθεται στο Παράρτημα Δ.

Παρακάτω θα παρουσιαστούν τα αποτελέσματα του πειράματος και στη συνέχεια θα γίνει σύνοψη των συμπερασμάτων και προτάσεις για αντιμετώπιση των προβλημάτων.

##### **5.2.2.1 Αποτελέσματα**

###### **5.2.2.1.1 Δημογραφικά Συμμετεχόντων**

Στον πίνακα 5.2 παρουσιάζονται οι απαντήσεις των συμμετεχόντων στις ερωτήσεις του προκαταρκτικού ερωτηματολογίου (βλέπε παράρτημα Δ).

<i>Ερώτηση</i>	<i>Σ1</i>	<i>Σ2</i>	<i>Σ3</i>	<i>Σ4</i>	<i>Σ5</i>	<i>Σ6</i>	<i>Σ7</i>
Ηλικία	23-28	23-28	23-28	23-28	18-22	18-22	18-22
Φύλλο	A	A	A	A	A	Θ	Θ
Υπηκοότητα	Ελληνική	Ελληνική	Ελληνική	Ελληνική	Ελληνική	Γερμανική	Γερμανική
Επάγγελμα	Μεταπτυχιακός Φοιτητής	Μεταπτυχιακός Φοιτητής	Προπτυχιακός Φοιτητής	Προπτυχιακός Φοιτητής	Προπτυχιακός Φοιτητής	Προπτυχιακός Φοιτητής	Προπτυχιακός Φοιτητής
Ομιλούμενες γλώσσες	Ελληνικά μητρική Αγγλικά πολύ καλά	Ελληνικά μητρική Αγγλικά άριστα	Ελληνικά μητρική Αγγλικά πολύ μέτρια	Γαλλικά μέτρια	Ελληνικά μητρική Αγγλικά πολύ καλά Γαλλικά μέτρια	Γερμανικά μητρική Αγγλικά πολύ καλά	Γερμανικά μητρική Αγγλικά πολύ καλά
Εμπειρία στο Διαδίκτυο	> 5	> 5	3 έως 5	3 έως 5	> 5	1 έως 3	3 έως 5
Εβδομαδιαίες ώρες στο Διαδίκτυο	5 έως 10	> 20	> 20	10 έως 20	> 20	< 5	5 έως 10
Αγαπημένη ασχολία στο Διαδίκτυο	Έρευνα	Διασκέδαση Εργασία	Έρευνα	Εργασία Έρευνα	Εργασία Έρευνα Αγορές	Διασκέδαση	Διασκέδαση Κοινωνικά θέματα
Διαδίκτυο για διακοπές στο παρελθόν	Ναι	Ναι	Όχι	Όχι	Ναι	Ναι	Ναι
Γνώση για τοποθεσίες Κρήτης	Στοιχειώδη	Σημαντικές	Σημαντικές	Σημαντικές	Σημαντικές	Καμία ιδιαίτερη	Στοιχειώδη

Πίνακας 5.2 Δημογραφικά στοιχεία χρηστών από το προκαταρκτικό ερωτηματολόγιο για κάθε συμμετοχή στα πειράματα

### **5.2.2.1.2 Αποτελέσματα Ανάλυσης Εργασιών από Χρήστες**

Στον πίνακα 5.3 παρουσιάζονται συνοπτικά στοιχεία για την εκτέλεση των εργασιών για όλες τις εργασίες. Για κάθε εργασία φαίνεται

- το ποσοστό των χρηστών που ολοκλήρωσε επιτυχώς την εργασία
- το ποσοστό των χρηστών που ολοκλήρωσε επιτυχώς την εργασία εντός του χρόνου αναφοράς
- το ποσοστό των χρηστών που ολοκλήρωσε επιτυχώς την εργασία με το βέλτιστο μονοπάτι
- ο μέσος όρος του χρόνου
- τυπική απόκλιση του χρόνου
- ο μέσος όρος του αριθμού σελιδοπροβολών
- τυπική απόκλιση του αριθμού σελιδοπροβολών

Στον πίνακα 5.4 παρουσιάζονται αναλυτικά για κάθε συμμετέχοντα χρήστη (Σ) και για κάθε εργασία (Ε) οι χρόνοι εκτέλεσης και ο αριθμός των σελιδοπροβολών στο μονοπάτι που ακολουθήθηκε.

<i>Εργασία</i>	<i>% επιτυχούς ολοκλήρωσης</i>	<i>% επιτυχούς ολοκλήρωσης εντός χρονικού ορίου</i>	<i>% επιτυχούς ολοκλήρωσης με βέλτιστο μονοπάτι</i>	<i>Μέσος όρος χρόνου (δευτερόλεπτα)</i>	<i>Τυπική απόκλιση χρόνου (δευτερόλεπτα)</i>	<i>Μέσος όρος αριθμού σελιδοπροβολών</i>	<i>Τυπική απόκλιση αριθμού σελιδοπροβολών (σελιδοπροβολές)</i>
<b>E1.</b>	-	-	-	-	-	-	-
<b>E2.</b>	100,00%	85,70%	85,70%	34,14	25,72	2,14	0,38
<b>E3.</b>	100,00%	85,70%	71,40%	81,43	101,17	3,86	1,86
<b>E4.</b>	100,00%	100,00%	100,00%	53	21,64	3	0
<b>E5.</b>	85,70%	0,00%	0,00%	182,33	80,37	7,67	1,75
<b>E6.</b>	100,00%	71,40%	85,70%	65,14	38,43	2,29	0,76
<b>E7.</b>	100,00%	57,10%	100,00%	125,71	78,74	3	0
<b>E8.</b>	100,00%	100,00%	71,40%	71	44,38	5,86	1,86
<b>E9.</b>	100,00%	85,70%	100,00%	31,29	33,16	1	0
<b>E10.</b>	100,00%	28,60%	28,60%	217	64,71	8	2,24

Πίνακας 5.3 Συγκεντρωτικά στοιχεία εκτέλεσης εργασιών στο πείραμα χρηστών

<i>Εργασία</i>	<i>Χρόνος (sec)</i>								<i>Σελιδοπροβολές</i>							
	<i>Χρόνος αναφοράς</i>	<i>Σ1</i>	<i>Σ2</i>	<i>Σ3</i>	<i>Σ4</i>	<i>Σ5</i>	<i>Σ6</i>	<i>Σ7</i>	<i>Βέλτιστο</i>	<i>Σ1</i>	<i>Σ2</i>	<i>Σ3</i>	<i>Σ4</i>	<i>Σ5</i>	<i>Σ6</i>	<i>Σ7</i>
<i>E1.</i>	-	250	265	180	300	180	190	230	-	-	-	-	-	-	-	-
<i>E2.</i>	<b>80</b>	15	38	20	24	90	22	30	<b>2</b>	2	2	2	2	3	2	2
<i>E3.</i>	<b>110</b>	60	35	45	310	40	46	34	<b>3</b>	4	3	3	8	3	3	3
<i>E4.</i>	<b>110</b>	40	75	27	87	53	35	54	<b>3</b>	3	3	3	3	3	3	3
<i>E5.</i>	<b>140</b>	240	55	110	229	-	210	250	<b>4</b>	9	6	5	9	-	8	9
<i>E6.</i>	<b>80</b>	50	120	120	55	30	35	46	<b>2</b>	2	2	4	2	2	2	2
<i>E7.</i>	<b>110</b>	300	90	110	75	116	76	113	<b>3</b>	3	3	3	3	3	3	3
<i>E8.</i>	<b>170</b>	30	85	120	140	38	44	40	<b>5</b>	5	5	6	10	5	5	5
<i>E9.</i>	<b>50</b>	20	105	30	15	10	15	24	<b>1</b>	1	1	1	1	1	1	1
<i>E10.</i>	<b>200</b>	300	180	285	120	249	210	175	<b>6</b>	7	10	6	7	12	8	6

Πίνακας 5.4 Στοιχεία εκτέλεσης εργασιών στο πείραμα χρηστών

Από τα στοιχεία των προηγούμενων πινάκων ξεχωρίζουν οι περιπτώσεις των εργασιών 5 , 7 και 10.

Η **εργασία 5** αναφέρεται στην εύρεση αναλυτικού χάρτη μιας περιοχής. Είναι κατ' αρχήν η μόνη εργασία όπου υπήρξε αποτυχία από έναν χρήστη στην ολοκλήρωσή της. Επίσης κανένας δεν κατάφερε να ολοκληρώσει την εργασία εντός του χρόνου αναφοράς αλλά ούτε χρησιμοποιώντας το βέλτιστο μονοπάτι.

Όλοι οι χρήστες κατάφεραν να φτάσουν στο σημείο όπου εμφανίζεται η τοποθεσία της περιοχής σε απλό χάρτη (είτε δια μέσω της σελίδας “Επικοινωνία & Τοποθεσία” στην παρουσίαση της επιχείρησης, είτε μέσω του εικονιδίου για το χάρτη της Κρήτης στα αποτελέσματα αναζήτησης). Σε εκείνο το σημείο όμως είτε επέλεξαν το σημείο στο χάρτη το οποίο οδηγεί στην παρουσίαση της περιοχής, είτε ξεκινούσαν από την αρχική σελίδα την εύρεση του χάρτη από την ενότητα “Χάρτες Κρήτης”.

Η **εργασία 7** αναφέρεται στην αναζήτηση καταλύματος με συγκεκριμένα κριτήρια, ένα από τα οποία είναι η επιλογή συγκεκριμένης περιοχής της Κρήτης. Ενώ όλοι επέτυχαν το βέλτιστο μονοπάτι που οδηγεί στην ολοκλήρωση της εργασίας, μόνο το 57% κατάφερε να ολοκληρώσει την εργασία εντός του χρόνου αναφοράς. Ανεξαιρέτως όλοι οι χρήστες συνάντησαν δυσκολίες στο παράθυρο “επιλογής περιοχής αναζήτησης” στο να κατανοήσουν τη λειτουργία του. Στο παράθυρο μπορεί να εκτελεστούν διάφορες λειτουργίες αλλά επειδή δεν αλλάζει η σελίδα κατά τη διάρκεια αυτών των λειτουργιών μετρήθηκε ως μία σελιδοπροβολή. Γι' αυτό παρατηρείται το φαινόμενο ο αριθμός των σελιδοπροβολών να είναι ο βέλτιστος για όλους τους χρήστες.

Ακόμα και στην περίπτωση όπου εμφανιζόταν η λίστα των περιοχών για τις οποίες ενδιαφέρονταν, τους ήταν ιδιαίτερα δύσκολο να κατανοήσουν ότι για την τελική επιλογή των περιοχών έπρεπε να επιλέξουν είτε μεμονωμένα, είτε μέσω της ειδικής επιλογής όλα μαζί. Όταν επέλεξαν κάποια περιοχή και η λίστα των επιλεγμένων περιοχών ήταν μεγάλη, δεν ήταν εμφανές ότι είχε προστεθεί η περιοχή με αποτέλεσμα να προσπαθούν να την ξαναπροσθέσουν.

Η **εργασία 10** αναφέρεται στη χρησιμοποίηση της λειτουργίας “Το φυλλάδιο μου” για την επιλογή ορισμένων ξενοδοχείων. Τόσο ο χρόνος εκτέλεσης όσο ο αριθμός των σελιδοπροβολών βρίσκονται στο μεγαλύτερο βαθμό εκτός των ορίων. Επειδή η συγκεκριμένη εργασία είναι η πλέον σύνθετη, ο αριθμός των σελιδοπροβολών δεν ξαφνιάζει στον ίδιο βαθμό όσο ο χρόνος εκτέλεσης.

Κατά τη διάρκεια του πειράματος παρατηρήθηκε το φαινόμενο οι χρήστες να ακολουθούν μεν τα σωστά βήματα για την εκτέλεση της εργασίας (επίσκεψη στη σελίδα του φυλλαδίου όπου από τις οδηγίες αποκτούν μία αντίληψη για τον τρόπο λειτουργίας του) αλλά σκοντάφτουν στην εμφάνιση

του εικονιδίου για προσθαφαίρεση αντικειμένων στο φυλλάδιο.

Συγκεκριμένα, για τεχνικούς λόγους που έχουν σχέση με την υλοποίηση της εμφάνισης του εικονιδίου ανάλογα εάν έχει προστεθεί ή όχι το αντικείμενο στο φυλλάδιο, το εικονίδιο αυτό εμφανίζεται κατόπιν της πλήρους φόρτωσης της HTML σελίδας (πλην των εικόνων). Ιδιαίτερα σε μεγάλες σε μέγεθος σελίδες, όπως είναι τα αποτελέσματα αναζήτησης, η μεταφορά ολόκληρης της σελίδας καθυστερεί και έτσι παρόλο που αρχίζουν να εμφανίζονται τα περιεχόμενα της κάθε σελίδας, το εικονίδιο για τη λειτουργία του φυλλαδίου δεν εμφανίζεται. Οι χρήστες που έχουν ενημερωθεί από την επίσκεψη στη σελίδα του φυλλαδίου για την ύπαρξη του εικονιδίου, δεν το εντοπίζουν με αποτέλεσμα την αναζήτησή του επιλέγοντας διάφορους συνδέσμους.

Αν και στις υπόλοιπες εργασίες υπήρχαν σημεία όπου είτε από τα σχόλια είτε από τη συμπεριφορά των χρηστών εντοπίζονται προβλήματα, γενικώς η ανταπόκριση της εφαρμογής στις ανάγκες των χρηστών κρίνεται ικανοποιητική, όπως εξάλλου φαίνεται από τα στοιχεία του πίνακα 5.3. Ορισμένα από τα σχόλια των χρηστών θα παρουσιαστούν σε επόμενη παράγραφο.

#### **5.2.2.1.3 Ερωτηματολόγιο Περάτωσης**

Στον πίνακα 5.3 φαίνονται οι απαντήσεις των ερωτήσεων 1 έως 3 του ερωτηματολογίου περάτωσης. Οι απαντήσεις αυτών των ερωτήσεων είναι γενικά σε πολύ θετικές περιοχές με εξαίρεση το χρήστη 4 όπου στη δυσκολία πλοήγησης έδωσε τη δεύτερη θετική γνώμη, και το χρήστη 5 που έδωσε για την οργάνωση πληροφορίας τη δεύτερη θετική γνώμη.

<i>Ερώτηση</i>	<i>Σ1</i>	<i>Σ2</i>	<i>Σ3</i>	<i>Σ4</i>	<i>Σ5</i>	<i>Σ6</i>	<i>Σ7</i>
1 (οργάνωση πληροφορίας)	Πολύ εξυπηρετική	Πολύ εξυπηρετική	Πολύ εξυπηρετική	Πολύ εξυπηρετική	Εξυπηρετική	Πολύ εξυπηρετική	Πολύ εξυπηρετική
2 (δυσκολία πλοήγησης)	Εύκολη	Εύκολη	Εύκολη	Κανονική	Εύκολη	Εύκολη	Εύκολη
3 (συμφωνία με σχεδιασμό)	Ναι	Ναι	Ναι	Ναι	Ναι	Ναι	Ναι

*Πίνακας 5.5 Απαντήσεις των ερωτήσεων 1 έως 3 του ερωτηματολογίου περάτωσης*



Οι παρατηρήσεις που έγιναν στις ερωτήσεις 4 και 5 (περισσότερο προφορικές) επικεντρώθηκαν στη δυσκολία χρήσης του παραθύρου επιλογής περιοχής. Κοινή ήταν η άποψη ότι είναι ιδιαίτερα δύσχρηστο και δυσνόητο.

Έγιναν επίσης προτάσεις για βελτίωση άλλων επιμέρους τμημάτων του ιστοχώρου, χωρίς όμως να αναφέρεται σε πρόβλημα χρησιμοποίησιμότητας παρά περισσότερο σε αισθητικά θέματα. Οι προτάσεις περιλάμβαναν θέματα όπως το χρώμα του φόντου των σελίδων ή την υποστήριξη συμβόλων υποκατάστασης (wild card) στον τίτλο για την αναζήτηση επιχειρήσεων.

Τέλος σημειώθηκε από τρεις διαφορετικούς χρήστες (1, 6 και 7) ως σημαντική έλλειψη του ιστοχώρου η απουσία τιμών για τα ξενοδοχεία.

#### **5.2.2.1.4 Σχόλια Συμμετεχόντων**

Κατά τη διάρκεια της εκτέλεσης των εργασιών έγιναν διάφορα σχόλια από τους χρήστες, από τα οποία τα σημαντικότερα συνοψίζονται παρακάτω:

- στην εργασία 7 σχολιάστηκε από το χρήστη 3 και 5 ότι ανέμεναν η λίστα των επιλεγμένων περιοχών να εμφανιζόταν με αλφαβητική σειρά
- στην εργασία 10 παρατηρήθηκε από το χρήστη 9 και 5 ότι θα ήταν χρήσιμο να εμφανίζεται πληροφορία για την τοποθεσία του κάθε αντικειμένου του φυλλαδίου
- το εικονίδιο που οδηγεί στον χάρτη τοποθεσίας αντικειμένων (αποτελέσματα αναζήτησης, παρουσίαση επιχειρήσεων και περιοχών κλπ.) θα μπορούσε να δείχνει απ' ευθείας την τοποθεσία και δε χρειάζεται το ενδιάμεσο στάδιο με τον απλό χάρτη. Θα έπρεπε πιέζοντας το εικονίδιο να εμφανίζεται απ' ευθείας ο αναλυτικός χάρτης.
- τα εικονίδια της κάθε ενότητας στην αρχική σελίδα θα έπρεπε να έχουν σύνδεσμο όπως και το κείμενο

#### **5.2.2.2 Συμπεράσματα**

Τα πειράματα χρηστών επικεντρώθηκαν στις λειτουργίες αναζήτησης πληροφοριών για επιχειρήσεις τουρισμού στον ιστοχώρο Crete TOURnet. Έφεραν στο φως σημαντικά προβλήματα χρησιμοποίησιμότητας του ιστοχώρου που επικεντρώνονται κυρίως

- στην εύρεση αναλυτικού χάρτη για κάποια περιοχή
- στην επιλογή περιοχή αναζήτησης

Σημειώνεται ότι και οι δύο λειτουργίες θεωρούνται σημαντικές όσον αφορά έναν ιστοχώρο τουριστικού προορισμού και ως τέτοιες τα προβλήματα που παρουσιάζουν χρήζουν άμεσης αντιμετώπισης.

Σε αυτές τις λειτουργίες παρουσιάστηκαν προβλήματα σε ανεξαιρέτως όλους τους χρήστες και η αποτυχία κρίνεται καθολική. Ειδικά η εργασία 5 (εύρεση αναλυτικού χάρτη) ήταν η μοναδική όπου υπήρξε διακοπή και μη ολοκλήρωση της εργασίας παρ όλη την προσπάθεια που ο χρήστης είχε καταβάλει. Εικάζεται ότι σε πραγματικές συνθήκες τα ποσοστά διακοπής θα ήταν μεγαλύτερα, διότι σε πειράματα τέτοιου είδους οι χρήστες συνήθως καταβάλλουν περισσότερο κόπο και δείχνουν περισσότερη υπομονή στην εκτέλεση των εργασιών.

Το τρίτο σημαντικό πρόβλημα που παρουσιάστηκε στη λειτουργία του ηλεκτρονικού φυλλαδίου είναι διαφορετικής υφής από τα προηγούμενα, μιας και δεν εντοπίζεται στο σχεδιασμό ή στην εμφάνιση του ιστοχώρου αλλά σε καθαρά τεχνικό θέμα. Η αναφορά γίνεται στην καθυστέρηση εμφάνισης του εικονιδίου για προσθαφαίρεση αντικειμένων στο φυλλάδιο. Από τη στιγμή που οι χρήστες ανακάλυπταν την καθυστέρηση στην εμφάνιση, ήταν σε θέση να προχωρήσουν με σύντομους ρυθμούς στην ολοκλήρωση της εργασίας.

Παρ όλες τις δυσκολίες που συνάντησαν οι χρήστες στην εκτέλεση των συγκεκριμένων προβληματικών εργασιών, ο ιστοχώρος (τουλάχιστον ως προς το τμήμα που εξετάστηκε) μπορεί να θεωρηθεί επιτυχημένος. Σε όλες τις υπόλοιπες εργασίες το ποσοστό επιτυχίας ήταν 100% και τόσο οι χρόνοι εκτέλεσης όσο και τα μονοπάτια που ακολουθήθηκαν δεν απείχαν σημαντικά από το χρόνο αναφοράς και το βέλτιστο μονοπάτι.

### **5.2.3 Συμπεράσματα Μελέτης Ιστοχώρου Τουρισμού με Παραδοσιακές Μεθόδους**

Στα πλαίσια της παρούσας εργασίας διενεργήθηκε μία μελέτη χρησιμοποιοσιμότητας ευριστικής αξιολόγησης και μία αντίστοιχη μελέτη με πειράματα χρηστών για τον ιστοχώρο τουριστικού προορισμού Crete TOURnet.

Τόσο από τις οδηγίες προς του αξιολογητές της ευριστικής αξιολόγησης όσο και από την επιλογή των εργασιών στα πειράματα χρηστών, το βάρος δόθηκε στην εύρεση πληροφορίας για επιχειρήσεις τουρισμού.

Από τις δύο μελέτες εντοπίστηκαν σημαντικά προβλήματα χρησιμοποιοσιμότητας που χρήζουν αντιμετώπισης για την καλύτερη ανταπόκριση του ιστοχώρου στις ανάγκες των χρηστών. Στον πίνακα 5.6 συνοψίζονται οι λειτουργίες του ιστοχώρου για τις οποίες εντοπίστηκαν τα πιο

σημαντικά προβλήματα σε κάθε μία μελέτη.

<i>Λειτουργία</i>	<i>Ευριστική</i>	<i>Πειράματα</i>
Επιλογή περιοχής αναζήτησης	×	×
Εύρεση αναλυτικού χάρτη περιοχής	×	×
Εμφάνιση εικονιδίων προσθαφαίρεσης φυλλαδίου		×
Εμφάνιση κριτηρίων αναζήτησης στα αποτελέσματα	×	
Περισσότερη πληροφορία στα αντικείμενα του φυλλαδίου	×	×

Πίνακας 5.6 Τα σημαντικότερα προβλήματα που εντοπίστηκαν ανά είδος μελέτης

Από τη σχετική σύγκριση προκύπτει ότι από τα συνολικά 5 σημαντικά προβλήματα που εντοπίστηκαν τα 3 ήταν κοινά ενώ τα υπόλοιπα 2 μοιράζονται στις διαφορετικές μελέτες. Θα πρέπει να τονιστεί ότι οι διαφορετικές μελέτες δεν ήταν εξαντλητικές και ότι είναι πολύ πιθανό να είχαν εντοπιστεί όλα τα προβλήματα εξίσου στις δύο μελέτες, ή και να είχαν εντοπιστεί και επιπλέον σημαντικά προβλήματα. Επίσης πρέπει να σημειωθεί ότι σύμφωνα με τις σχετικές μελέτες για ευριστική αξιολόγηση (βλέπε παράγραφο 2.2.3.3.2) ο αριθμός των αξιολογητών στη συγκεκριμένη μελέτη ήταν μάλλον μικρός.

### **5.3 Μελέτη με τη Μεθοδολογία Καταγραφής και Ανάλυσης Δεδομένων Χρήσης Ιστοχώρων**

Το Σύστημα Καταγραφής Χρήσης ιστοχώρων που αναπτύχθηκε στα πλαίσια της παρούσας διατριβής δημιουργεί από τις προσβάσεις στον ιστοχώρο τη Βάση Χρήσης του ιστοχώρου σύμφωνα με το Σχήμα 3.2. Στη Βάση Χρήσης μπορεί να αποθηκευτούν ένα πλήθος στοιχείων για την κάθε πρόσβαση στον ιστοχώρο. Αν και για τις βασικές έννοιες των επισκέψεων υπάρχουν προκαθορισμένα δεδομένα που αποθηκεύονται (όπως έναρξη επίσκεψης, όνομα σελίδων που προβάλλονται, προέλευση του επισκέπτη κλπ.) για ένα μεγάλο πλήθος δεδομένων απαιτείται ειδική μέριμνα από τους τεχνικούς που αναπτύσσουν τον ιστοχώρο ώστε να είναι δυνατή η καταγραφή των.

Τέτοιου είδους δεδομένα είναι κυρίως αυτά που αναφέρονται στο συγκεκριμένο περιεχόμενο της κάθε σελίδας, πράγμα που δεν είναι δυνατόν να εξαχθεί με απολύτως αυτόματο τρόπο, διότι απαιτείται βαθιά γνώση για τη δομή και το περιεχόμενο του ιστοχώρου. Η σχετικές βασικές επιλογές που έγιναν για τον ιστοχώρο τουριστικού προορισμού περιγράφηκαν στην ενότητα 3.2 στη σελίδα 126 και στον πίνακα 3.3.

Ένας άλλος τομέας όπου απαιτείται χειροκίνητη επέμβαση από τους τεχνικούς, είναι η εξαγωγή των φράσεων αναζήτησης όταν η προέλευση των επισκεπτών είναι από μηχανές αναζήτησης, όπου

η φράση αποθηκεύεται ως μέρος του URL. Επειδή η κάθε μηχανή χρησιμοποιεί διαφορετικούς τρόπους αναπαράστασης των φράσεων αναζήτησης στις παραμέτρους του URL, πρέπει για κάθε μία από αυτές να προσαρμοστεί ο σχετικός κώδικας που εξάγει τις φράσεις. Για τον υπό εξέταση ιστοχώρο έγινε η σχετική εργασία για το μεγαλύτερο μέρος των μηχανών αναζήτησης που εμφανίζονται ως προέλευση.

Το σύστημα άρχισε να καταγράφει τις πρώτες προσβάσεις στις αρχές του 1999, όταν ο ιστοχώρος είχε ακόμα την μορφή από την πιλοτική λειτουργία. Από τα μέσα του 2002 άρχισε η καταγραφή των δεδομένων από την τελική μορφή που έλαβε το σύστημα, και από τότε τα δεδομένα παρουσιάζουν τη μεγαλύτερη αξιοπιστία.

Το αποτέλεσμα είναι να υπάρχει ένας μεγάλος όγκος δεδομένων από τη χρήση του ιστοχώρου, σε μία δομημένη μορφή, ο οποίος μπορεί να αξιοποιηθεί για πολλούς σκοπούς, μερικοί εκ των οποίων αναλύονται στην ενότητα σχετικής έρευνας για Web Usage Analysis (Κεφάλαιο 2). Από την έναρξη της καταγραφής, και σε διάστημα περίπου 36 μηνών, έχουν αποθηκευτεί στη βάση χρήσης τα δεδομένα για περίπου 1.100.000 διακριτές επισκέψεις στον ιστοχώρο.

Στις επόμενες ενότητες θα παρουσιαστούν τα αποτελέσματα που προκύπτουν ακολουθώντας τη μεθοδολογία ανάλυσης από το σύστημα καταγραφής που προτάθηκε στο κεφάλαιο 4. Για τη διευκόλυνση του αναγνώστη θα ακολουθηθεί η ίδια δομή παρουσίασης με την παρουσίαση της μεθοδολογίας, η οποία είναι η ακόλουθη:

- Γενικά Χαρακτηριστικά Επισκέψεων
- Προσωποποίηση Ιστοχώρου
- Χρησιμοποιησιμότητα
- Σενάρια Χρήσης

Ο αναγνώστης πρέπει να έχει υπ' όψιν τις ακόλουθες σημαντικές παραμέτρους κατά την ερμηνεία των διαγραμμάτων:

- Όπου τα δεδομένα αναφέρονται σε χρονικές περιόδους ή άπτονται χρονικών παραμέτρων, αναφέρεται το χρονικό διάστημα στο οποίο έγιναν οι μετρήσεις.
- Ως σημαντική χρονική καμπή μπορεί να θεωρηθεί η 10η Απριλίου, διότι από τότε αντικαταστήθηκε η έκδοση της πιλοτικής λειτουργίας του ιστοχώρου με νέα έκδοση με σημαντικές αλλαγές στη δομή, στο επίπεδο αλληλεπίδρασης και στην παρουσίαση του

ιστοχώρου (βλέπε Παράτημα Α για αναλυτική περιγραφή των αλλαγών).

- Εάν δεν αναφέρεται διαφορετικά, τα δεδομένα αναφέρονται σε σελιδοπροβολές από πραγματικούς χρήστες και έχουν αφαιρεθεί (στο μέτρο του δυνατού) οι προσβάσεις στον ιστοχώρο από κάθε είδους αυτόματες μηχανές. Τέτοιες μηχανές είναι κυρίως οι εφαρμογές δεικτοδότησης των ιστοχώρων αναζήτησης που μπορεί να διαστρεβλώσουν σημαντικά τα αποτελέσματα, ειδικά σε ιστοχώρους με “μεγάλο” αριθμό URL και “μικρό” αριθμό επισκεπτών. Χαρακτηριστικά αναφέρεται ότι για τον Ιούνιο του 2003, το 20% των σελιδοπροβολών αντιστοιχεί σε προσβάσεις από εφαρμογές δεικτοδότησης, από το οποίο το 14% είναι του Google. Για τον εντοπισμό τέτοιων μηχανών χρησιμοποιήθηκε κυρίως το HTTP πεδίο για τον user-agent και γνωστές διευθύνσεις IP των μηχανών αναζήτησης. Βλέπε επίσης παράγραφο 2.3.4.2.1 σελίδα 73 του κεφαλαίου σχετικής έρευνας.

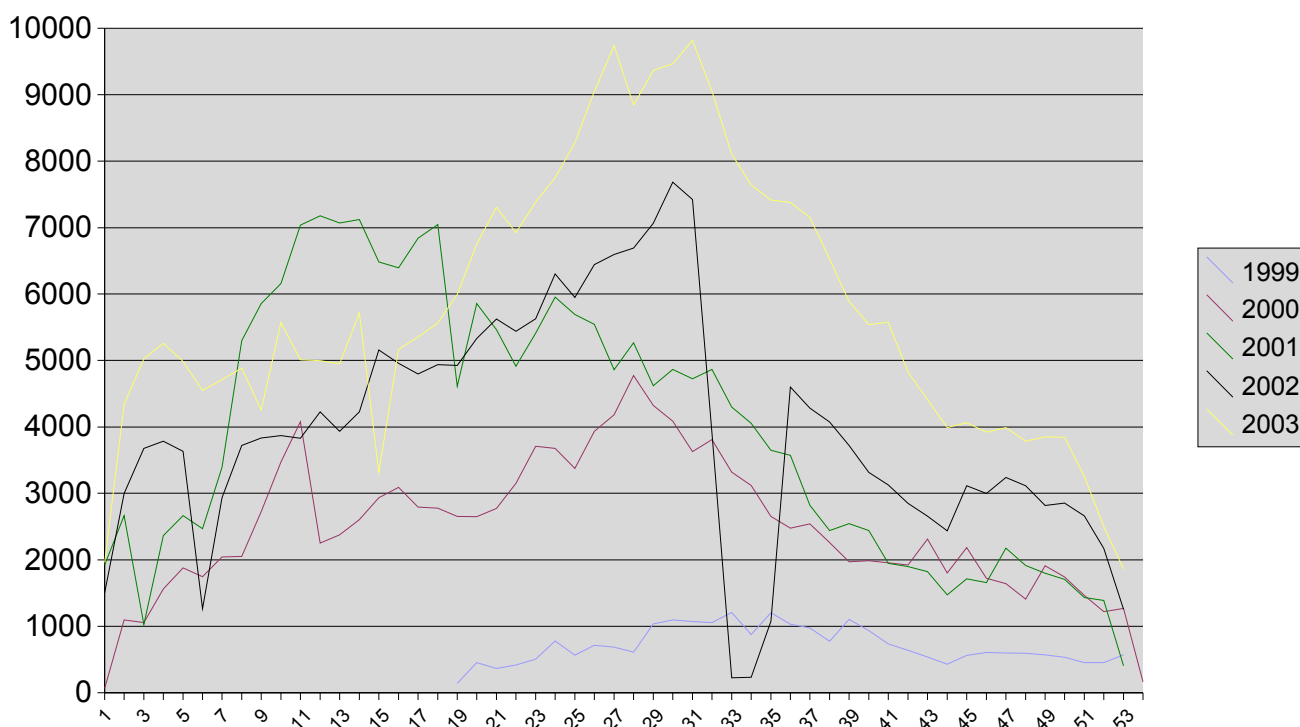
### **5.3.1 Γενικά Χαρακτηριστικά Ιστοχώρου**

Σε αυτή την ενότητα θα παρουσιαστεί ένα μεγάλο πλήθος στατιστικών δεδομένων για τη χρήση του ιστοχώρου τουριστικού προορισμού. Το μεγαλύτερο μέρος θα παρουσιαστεί υπό μορφή διαγραμμάτων με σχετικές επεξηγήσεις.

#### **5.3.1.1 Επισκέψεις Ιστοχώρου**

Πληροφορίες για τις επισκέψεις σε έναν ιστοχώρο έχουν τη μεγαλύτερη αξία σε όρους ανάλυσης, όπως περιγράφεται στο Κεφάλαιο 2.

Σε απόλυτους αριθμούς ανά εβδομάδα ο αριθμός των επισκέψεων στο Crete TOURnet στη διάρκεια του χρόνου παρουσιάζει την εικόνα που φαίνεται στο Διάγραμμα 5.1. Οι καμπύλες παρουσιάζουν καταρχήν μία σταθερή αύξηση στον αριθμό των επισκέψεων ανά έτος και μία πολύ χαρακτηριστική περιοδικότητα όσον αφορά την κατανομή τους στη διάρκεια ενός χρόνου. Οι επισκέψεις αρχίζουν να αυξάνονται μετά το Δεκέμβριο και παρουσιάζουν την κορύφωση στους καλοκαιρινούς μήνες και συγκεκριμένα στο τέλος Ιουλίου με αρχές Αυγούστου. Από εκεί και πέρα υπάρχει μία σταδιακή μείωση μέχρι και το Δεκέμβριο. Αυτή η εικόνα συμβαδίζει απόλυτα με την εποχικότητα του τουρισμού στην Κρήτη που κορυφώνεται τους μήνες Ιούλιο-Αύγουστο.



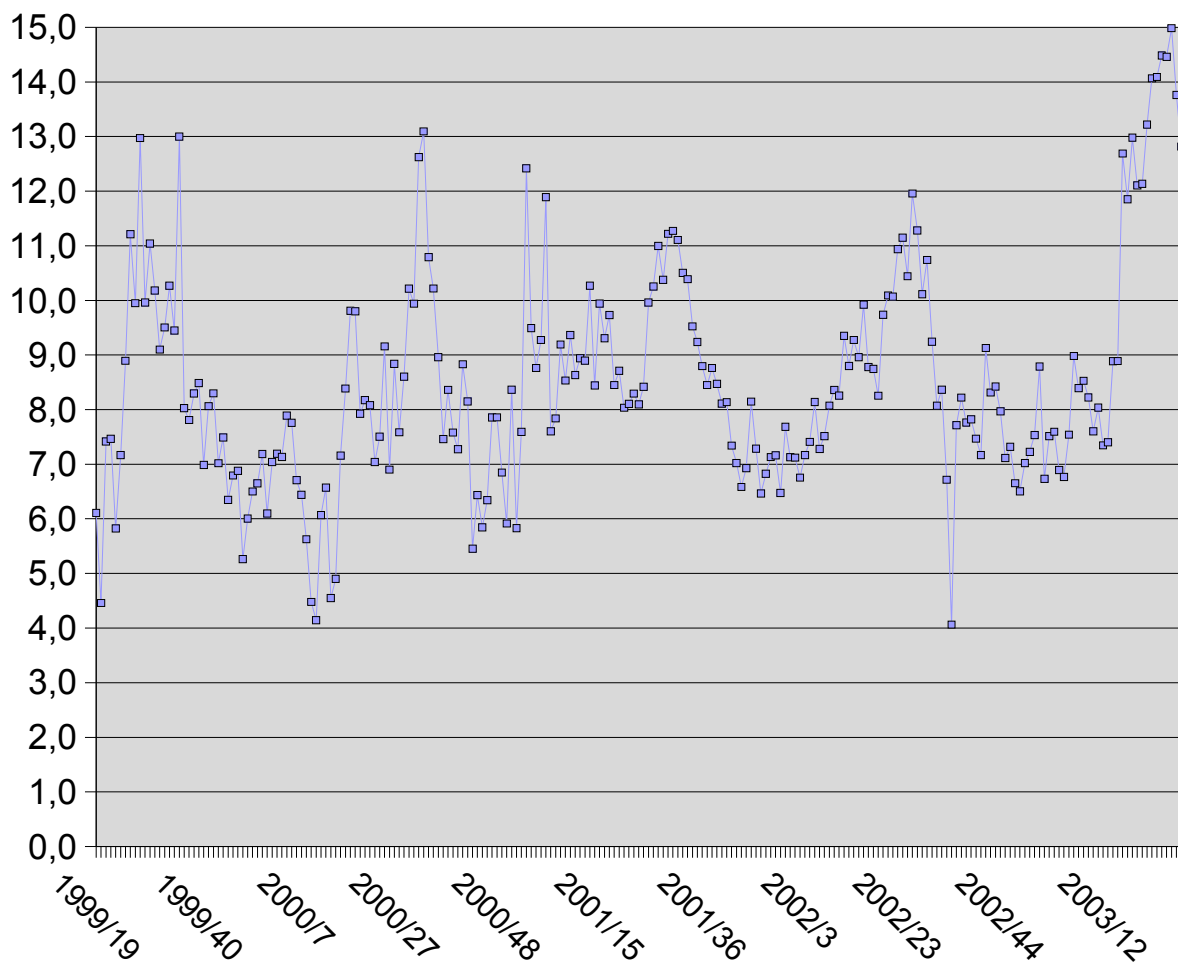
Διάγραμμα 5.1 Αριθμός επισκέψεων στη διάρκεια του χρόνου και ανά έτος στο Crete TOURnet. Ο άξονας του χρόνου είναι ανά εβδομάδα. Η απότομη πτώση των επισκέψεων το καλοκαίρι του 2002 οφείλεται στην αποσύνδεση του ιστοχώρου λόγω τεχνικών προβλημάτων. Λόγω της αλλαγής της μορφής του ιστοχώρου παρουσιάζεται παρόμοια συμπεριφορά για λίγες ημέρες τον Απρίλιο του 2003 (1999-2003)

Στην ερμηνεία του αριθμού των επισκέψεων σε έναν ιστοχώρο πρέπει να ληφθούν υπ' όψιν πολλοί παράγοντες. Έχοντας υπ' όψιν τον ορισμό της επίσκεψης, ως επίσκεψη θεωρείται το σύνολο των αιτήσεων προς τον ιστοχώρο από συγκεκριμένο χρήστη, των οποίων η χρονική στιγμή της αίτησης δεν απέχει περισσότερο από 30 λεπτά από την προηγούμενη. Στην περίπτωση που πρόκειται για την πρώτη αίτηση στο παράθυρο των 30 λεπτών, θεωρείται ως έναρξη νέας επίσκεψης, ενώ οι υπόλοιπες κατατάσσονται στην ίδια επίσκεψη με την πρώτη.

Αυτό σημαίνει ότι ακόμα και μία μεμονωμένη αίτηση στον εξυπηρετητή από κάποιον χρήστη, η οποία θα μπορούσε να γίνει ακόμα και “κατά λάθος”, μετρείται ως επίσκεψη. Οι πραγματικά ενδιαφέρουσες επισκέψεις για το συγκεκριμένο ιστοχώρο είναι αυτές που περιλαμβάνουν περισσότερες της μίας αίτησης οπότε και σελιδοπροβολής. Οι επισκέψεις που περιλαμβάνουν μόνο μία σελιδοπροβολή είναι υποψήφιες προς περαιτέρω ανάλυση για να εξακριβωθεί ο λόγος για τον οποίο δεν οδήγησαν σε παραπέρα “εξερεύνηση” από το χρήστη.

Στο Διάγραμμα 5.2 φαίνεται ο μέσος όρος του αριθμού σελιδοπροβολών ανά επίσκεψη με δείγματα ανά εβδομάδα. Σημαντική παρατήρηση που μπορεί να γίνει από αυτό το διάγραμμα είναι η

απότομη αύξηση των σελιδοπροβολών αμέσως μετά την ενεργοποίηση της νέας έκδοσης του ιστοχώρου στις αρχές του Απριλίου 2003. Επίσης είναι σημαντικές οι “ανωμαλίες” που παρουσιάζονται μέχρι τις αρχές του 2001 οι οποίες οφείλονται στην ανωριμότητα του συστήματος καταγραφής στη διάκριση πραγματικών επισκέψεων από επισκέψεις από αυτόματες μηχανές, οι οποίες βέβαια δεν αντιστοιχούν σε πραγματικούς χρήστες.



Διάγραμμα 5.2 Μέσος όρος αριθμού σελιδοπροβολών ανά επίσκεψη σε εβδομαδιαία βάση

Όπως αναφέρθηκε παραπάνω, οι επισκέψεις με μόνο μία σελιδοπροβολή, ή εάν γενικευτεί, οι επισκέψεις με “λίγες” σελιδοπροβολές, αναφέρονται σε χρήστες οι οποίοι για κάποιο λόγο δεν ικανοποιήθηκαν από τον ιστοχώρο και έτσι επέλεξαν να τον εγκαταλείψουν. Ο αριθμός των σελιδοπροβολών που θεωρούνται “λίγες” εξαρτάται σε μεγάλο βαθμό από τον ιστοχώρο στον οποίο αναφέρονται. Για παράδειγμα, ένας ιστοχώρος με νέα της ημέρας είναι λογικό να έχει πολλές επισκέψεις με μία μόνο σελιδοπροβολή, διότι πολλοί χρήστες θα τον επισκέπτονται πολλές φορές την ημέρα μόνο για να διαπιστώσουν ότι δεν έχει προστεθεί κάποιο νέο άρθρο.

Σε κάθε περίπτωση παρουσιάζει ενδιαφέρον να φανεί η κατανομή των επισκέψεων ανά αριθμό σελιδοπροβολών. Στο Διάγραμμα 5.3 (το οποίο εξάγεται με ερωτήσεις όπως φαίνεται στην Ερώτηση 4.7 σελίδα 143) εμφανίζεται η κατανομή των επισκέψεων σε ποσοστά στη διάρκεια του χρόνου με βάση τον αριθμό των σελιδοπροβολών. Έχει γίνει χωρισμός σε επισκέψεις με μία μόνο σελιδοπροβολή, με 2 με 4 σελιδοπροβολές, 5 με 10, 11 με 30 και άνω των 30 σελιδοπροβολών.

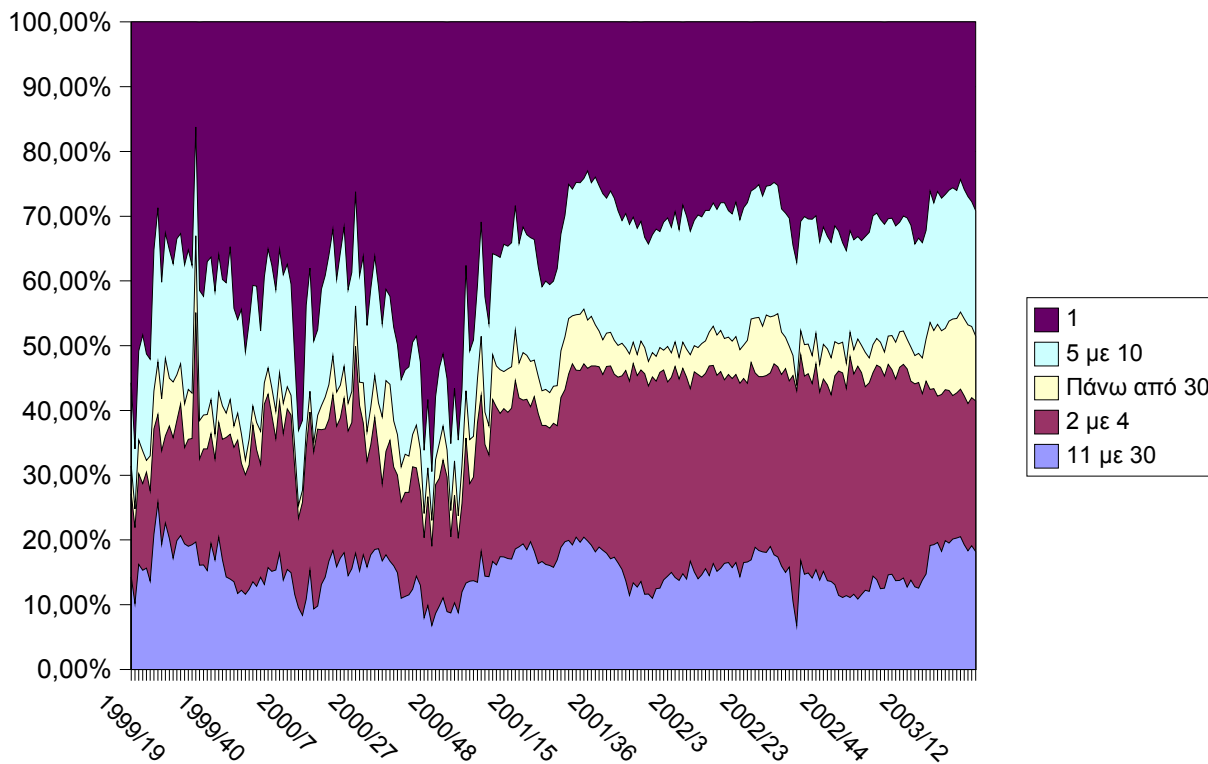
Το ποσοστό των επισκέψεων με μία μόνο σελιδοπροβολή φαίνεται να είναι πολύ μεγάλο, αν και μετά την ανανέωση του ιστοχώρου (Απρίλιος 2003) υπάρχει τάση να περιοριστεί προς όφελος των επισκέψεων άνω των 10 σελιδοπροβολών. Οι συσχετισμοί πριν τις αρχές του 2001 είναι επισφαλείς λόγω της ανωριμότητας του συστήματος καταγραφής που ήδη αναφέρθηκε.

Οι λόγοι που μπορεί να οδηγήσουν σε σύντομη αποχώρηση από τον ιστοχώρο μπορεί να είναι (χωρίς να περιορίζονται σε αυτούς):

1. ο ιστοχώρος είναι πολύ αργός
2. ο ιστοχώρος με την πρώτη ματιά φαίνεται να μην ανταποκρίνεται στο θέμα ενδιαφέροντος του χρήστη
3. ο ιστοχώρος παρουσιάζει σημαντικά προβλήματα χρησιμοποίησιμότητας

Οι επισκέψεις με μία μόνο σελιδοπροβολή μάλλον εμπίπτουν στην περίπτωση 1 και 2, ενώ οι επισκέψεις με πολύ λίγες (2-4) μάλλον στην περίπτωση 3.





Διάγραμμα 5.3 Κατανομή ποσοστών επισκέψεων στο χρόνο ανάλογα με τον αριθμό των σελιδοπροβολών. Οι επισκέψεις έχουν ομαδοποιηθεί ανάλογα με τον αριθμό σελιδοπροβολών. Στο διάγραμμα κάθε χρώμα αντιστοιχεί στο ποσοστό που έχει κάθε συγκεκριμένη ομάδα στο σύνολο των επισκέψεων (ποσοστό επισκέψεων με 1 σελιδοπροβολή, με 2-4 σελιδοπροβολές, 5-10, 11-30 και άνω των 30). Το άθροισμα των ποσοστών καλύπτει το σύνολο των επισκέψεων και είναι 100%.

### 5.3.1.2 Προέλευση Επισκεπτών

Ένα άλλο σημαντικό στοιχείο που επηρεάζει τον αριθμό των επισκέψεων είναι ο τρόπος με τον οποίο οι χρήστες έφτασαν στον ιστοχώρο. Διαφημιστικές καταχωρήσεις μπορεί να αυξήσουν σημαντικά τον αριθμό των επισκέψεων ενώ σύνηθες είναι μεγάλος αριθμός επισκέψεων να προέρχεται από μηχανές αναζήτησης. Δεδομένου ότι το Crete TOURnet δεν είχε ποτέ οργανωμένες διαφημιστικές καταχωρήσεις σε άλλους ιστοχώρους, είναι αναμενόμενο οι επισκέψεις να προέρχονται κυρίως από μηχανές αναζήτησης και από απλούς συνδέσμους άλλων ιστοχώρων που δεν έχουν τη μορφή διαφημίσεων.

Στη Βάση Χρήσης έχουν καταχωρηθεί από την έναρξη λειτουργίας του ιστοχώρου 186383 διαφορετικές διευθύνσεις προέλευσης (referring URLs) οι οποίες αντιστοιχούν σε 4293 διαφορετικά ονόματα μηχανημάτων (host names). Η μεγάλη διαφορά προέρχεται από το γεγονός ότι στις μηχανές αναζήτησης η διεύθυνση προέλευσης περιλαμβάνει πολλές παραμέτρους που διαφοροποιούν τα URL. Για παράδειγμα οι διευθύνσεις προέλευσης από τον ιστοχώρο [www.google.com](http://www.google.com) (και μόνο από αυτόν και όχι από τις παραλλαγές που υπάρχουν ανά χώρα όπως [www.google.com.gr](http://www.google.com.gr)) είναι 40479.

Ο Πίνακας 5.7 δίδει τα ποσοστά επισκέψεων ανά referring hostname για τον Ιούνιο του 2003 για τα πρώτα 40 μηχανήματα (από τα συνολικά 828 που υπήρχαν σε αυτό το μήνα). Το μεγαλύτερο ποσοστό φαίνεται να προέρχεται από άμεση πληκτρολόγηση της διεύθυνσης του ιστοχώρου στην εφαρμογή πλοήγησης (direct request), ακολουθούμενη από τη μηχανή αναζήτησης Google. Στην πραγματικότητα όμως η μηχανή αναζήτησης Google χρησιμοποιεί διαφορετικά ονόματα μηχανημάτων ή ακόμα και παραλλαγές ανάλογα με τη γλώσσα στην οποία εμφανίζεται το Google (βλέπε π.χ. A/A 2, 5, 6, 8, 12 κλπ.). Όμοια και άλλες μηχανές αναζήτησης ακολουθούν την ίδια τακτική, ενώ υπάρχουν και περιπτώσεις όπου ένας ιστοχώρος χρησιμοποιεί τα αποτελέσματα αναζήτησης από άλλες μηχανές για τις λειτουργίες αναζήτησης που προσφέρει (για παράδειγμα τα αποτελέσματα αναζήτησης από τη μηχανή αναζήτησης Netscape προέρχονται από το Google).

Υπάρχουν επίσης ιστοχώροι που κάτω από διαφορετικό όνομα περιοχής παρουσιάζουν ακριβώς το ίδιο περιεχόμενο, όπως για παράδειγμα ο Ελληνικός Οργανισμός Τουρισμού που είναι προσβάσιμος είτε ως [www.gnto.gr](http://www.gnto.gr) είτε ως [www.eot.gr](http://www.eot.gr).

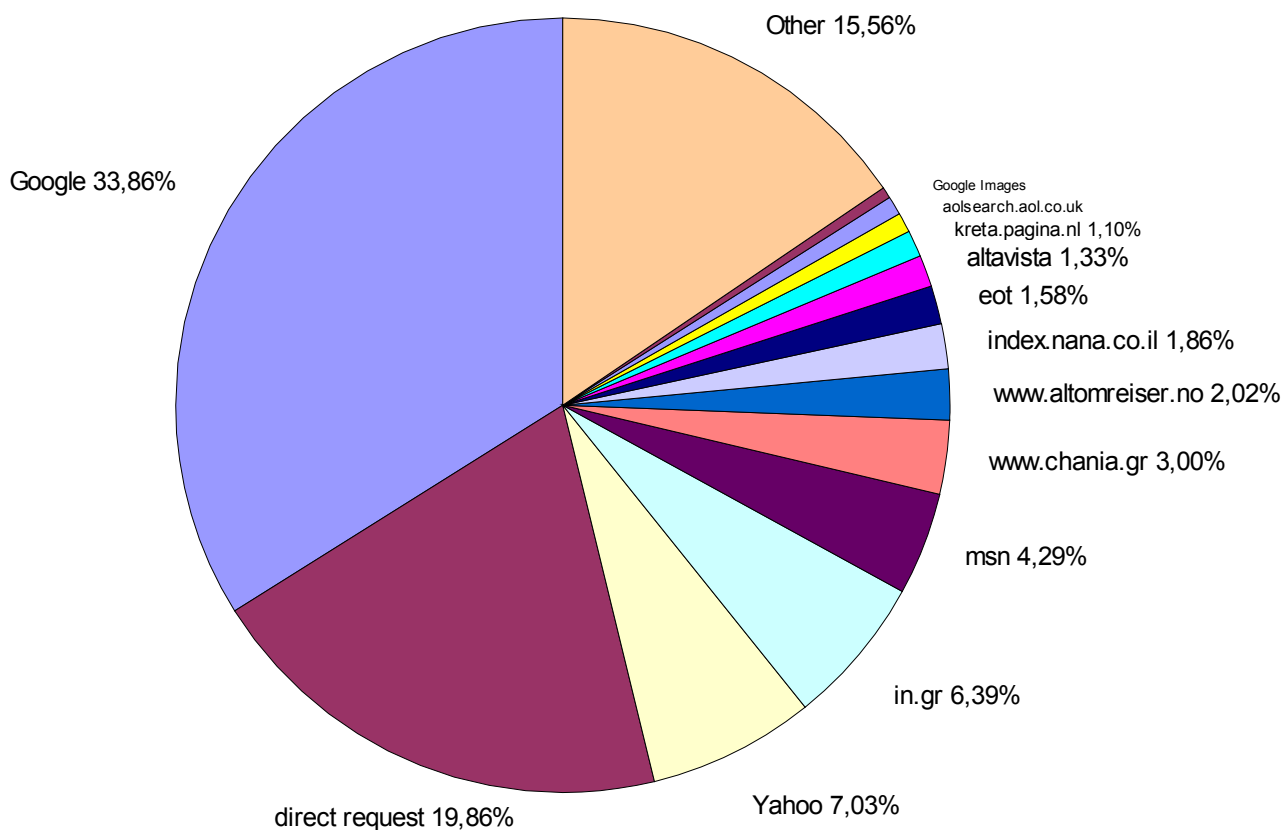
Επίσης είναι χαρακτηριστικό ότι οι ιστοχώροι που είναι γνωστοί ως μηχανές αναζήτησης ή κατάλογοι προσφέρουν πλέον μεγάλη γκάμα υπηρεσιών πέραν της κλασσικής απλής αναζήτησης. Έτσι για παράδειγμα το yahoo προσφέρει ηλεκτρονικό ταχυδρομείο (που φαίνεται από τη διεύθυνση προέλευσης ως μηχανήμα του είδους mailxx.yahho.xx) ή το Google προσφέρει αναζήτηση εικόνων που φαίνεται ως [images.google.xxxxx](http://images.google.xxxxx).

<i>A/A</i>	<i>Referring Host Name</i>	<i>Ποσοστό επί του συνόλου</i>
1	direct request	19,86%
2	www.google.com	13,22%
3	find.in.gr	5,40%
4	search.yahoo.com	4,50%
5	www.google.co.uk	4,33%
6	www.google.de	3,93%
7	www.chania.gr	3,00%
8	www.google.com.gr	2,05%
9	www.altomreiser.no	2,02%
10	search.msn.com	1,92%
11	index.nana.co.il	1,86%
12	www.google.fr	1,77%
13	search.msn.de	1,37%
14	www.google.it	1,30%
15	uk.search.yahoo.com	1,29%
16	kreta.pagina.nl	1,10%
17	www.google.be	1,02%
18	www.in.gr	0,99%
19	aolsearch.aol.co.uk	0,85%
20	www.google.at	0,83%
21	www.google.nl	0,83%
22	www.altavista.com	0,70%
23	www.google.ie	0,67%
24	www.gnto.gr	0,66%
25	www.google.ca	0,59%
26	www.google.ch	0,55%
27	odysseas.gnto.gr	0,52%
28	www.google.fi	0,49%
29	www.kretaforum.dk	0,47%
30	www.restplatzboerse.at	0,45%
31	www.google.pl	0,41%
32	www.eot.gr	0,40%
33	www.reiseguru.no	0,40%
34	www.google.co.il	0,39%
35	www.ifind.freemove.com	0,38%

<i>A/A</i>	<i>Referring Host Name</i>	<i>Ποσοστό επί του συνόλου</i>
36	www.google.com.au	0,38%
37	strony.wp.pl	0,34%
38	de.search.yahoo.com	0,33%
39	www.west-crete.com	0,32%
40	grieksetaal.pagina.nl	0,29%

Πίνακας 5.7 Ποσοστά επισκέψεων ανά Referring Host Name (Ιούνιος 2003)

Εάν ομαδοποιηθούν τα ονόματα του πίνακα 5.7 ανά οργανισμό προέλευσης παρουσιάζεται η εικόνα που φαίνεται στο Διάγραμμα 5.4. Οι επισκέψεις με προέλευση το Google καταλαμβάνουν την πρώτη θέση, ακολουθούμενοι από το Yahoo και το In.gr, ενώ υπάρχει και ένα ποσοστό 0,79% από images.google.xx που αναφέρεται σε αναζήτηση εικόνων.



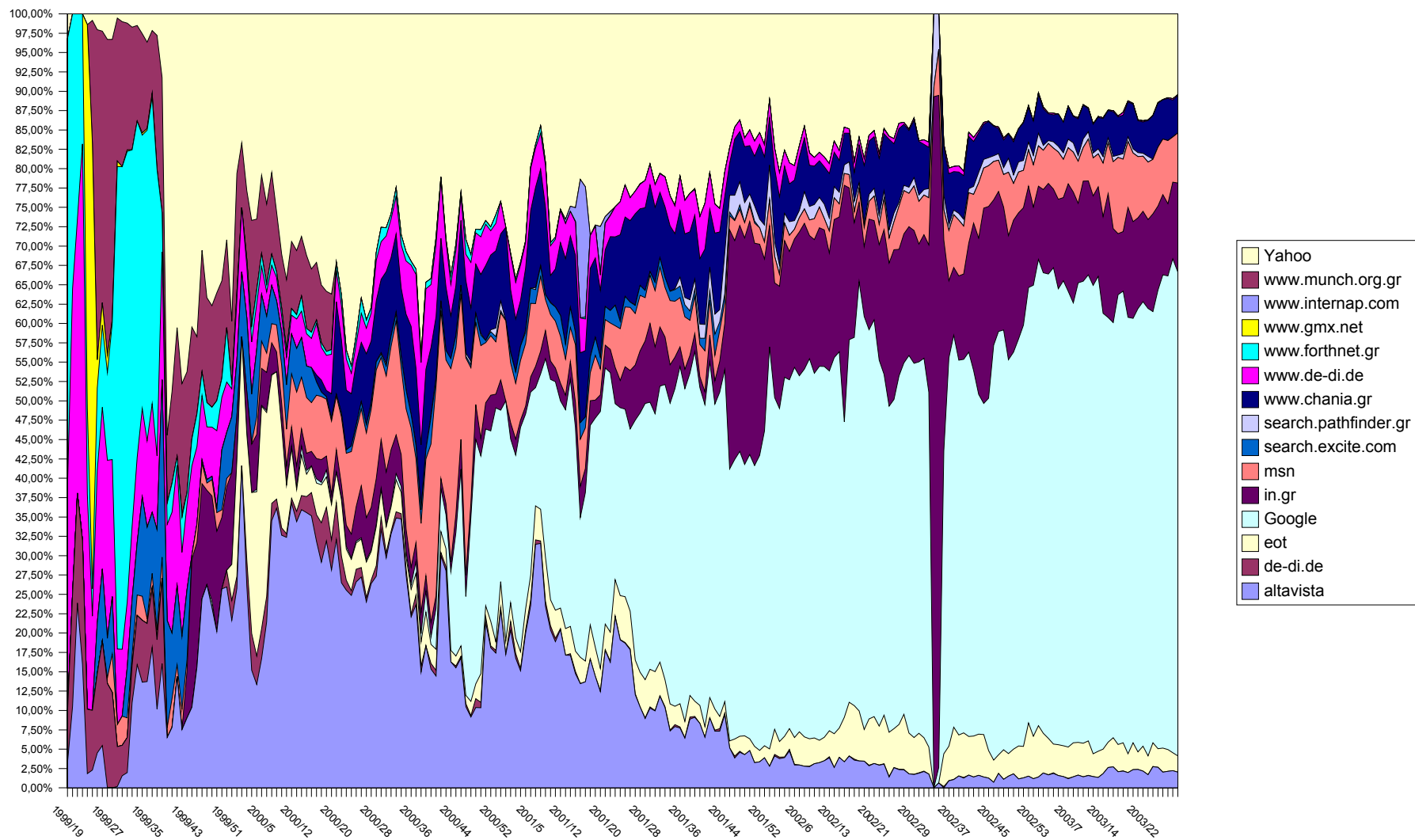
Διάγραμμα 5.4 Ιστοχώροι προέλευσης σε ποσοστά επί του συνόλου των επισκέψεων (Ιούνιος 2003). Ομαδοποίηση ανά οργανισμό

Βέβαια οι ιστοχώροι προέλευσης και τα ποσοστά που καταλαμβάνουν δεν είναι τα ίδια στη διάρκεια του χρόνου. Σε έναν απλό ιστοχώρο αρκεί ένας υπερσύνδεσμος για να αρχίσει να

εμφανίζεται στις προελεύσεις του Crete TOURnet και σε συνδυασμό με την επισκεψιμότητα του ιστοχώρου και την προβολή του συνδέσμου αυξομειώνονται οι επισκέψεις που προέρχονται από αυτόν.

Σε μία μηχανή αναζήτησης κυρίαρχο ρόλο παίζει η κατάταξη του Crete TOURnet στα αποτελέσματα αναζήτησης, ανάλογα πάντα με τη φράση αναζήτησης που έγινε στη μηχανή.

Στο Διάγραμμα 5.5 φαίνονται τα ποσοστά των επισκέψεων από εξωτερικούς συνδέσμους στη διάρκεια του χρόνου. Έχουν περιληφθεί όλα τα ονόματα των ιστοχώρων που σε εβδομαδιαία βάση κάποια στιγμή αντιστοιχούσαν σε τουλάχιστον 5% των επισκέψεων με εξωτερική προέλευση. Το καλοκαίρι του 1999 υπήρχε στην αρχική σελίδα της εταιρίας FORTHnet ειδικός διαφημιστικός σύνδεσμος στα πλαίσια του ερευνητικού προγράμματος. Οι σχετικές επισκέψεις ήταν πολλές και σταμάτησαν αμέσως μετά το πέρας της προβολής. Μεγάλο ποσοστό καταλάμβανε για μεγάλο χρονικό διάστημα η μηχανή αναζήτησης altavista το οποίο όμως μειώθηκε και παραμένουν κυρίαρχες οι επισκέψεις από Google, Yahoo και MSN. Την άνοιξη του 2000 ο Δήμος Χανίων άλλαξε διεύθυνση στον ιστοχώρο του (από [munch.org.gr](http://munch.org.gr) σε [chania.gr](http://chania.gr)) το οποίο αντικατοπτρίζεται σε διακοπή των επισκέψεων από την παλαιά διεύθυνση και την έναρξη επισκέψεων από τη νέα. Χαρακτηριστική είναι επίσης οι ανωμαλίες των αρχικών μηνών λειτουργίας όπου άρχισε να “γίνεται γνωστός” ο ιστοχώρος.



Διάγραμμα 5.5 Εξέλιξη ιστοχώρων προέλευσης όπου υπήρχε σε εβδομαδιαία βάση ποσοστό άνω του 5%. (κανονικοποιημένα). Στο διάγραμμα κάθε χρώμα αντιστοιχεί σε συγκεκριμένο ιστοχώρο προέλευσης, και δείχνει για κάθε χρονική στιγμή το ποσοστό των επισκεπτών που προέρχονται από τον κάθε ιστοχώρο. Συμπεριλήφθηκαν μόνο οι προελεύσεις που κάποια χρονική στιγμή είχαν ποσοστό 5% των επισκεπτών. Έγινε κανονικοποίηση στο άθροισμα των επισκεπτών από αυτές και μόνο τις προελεύσεις.

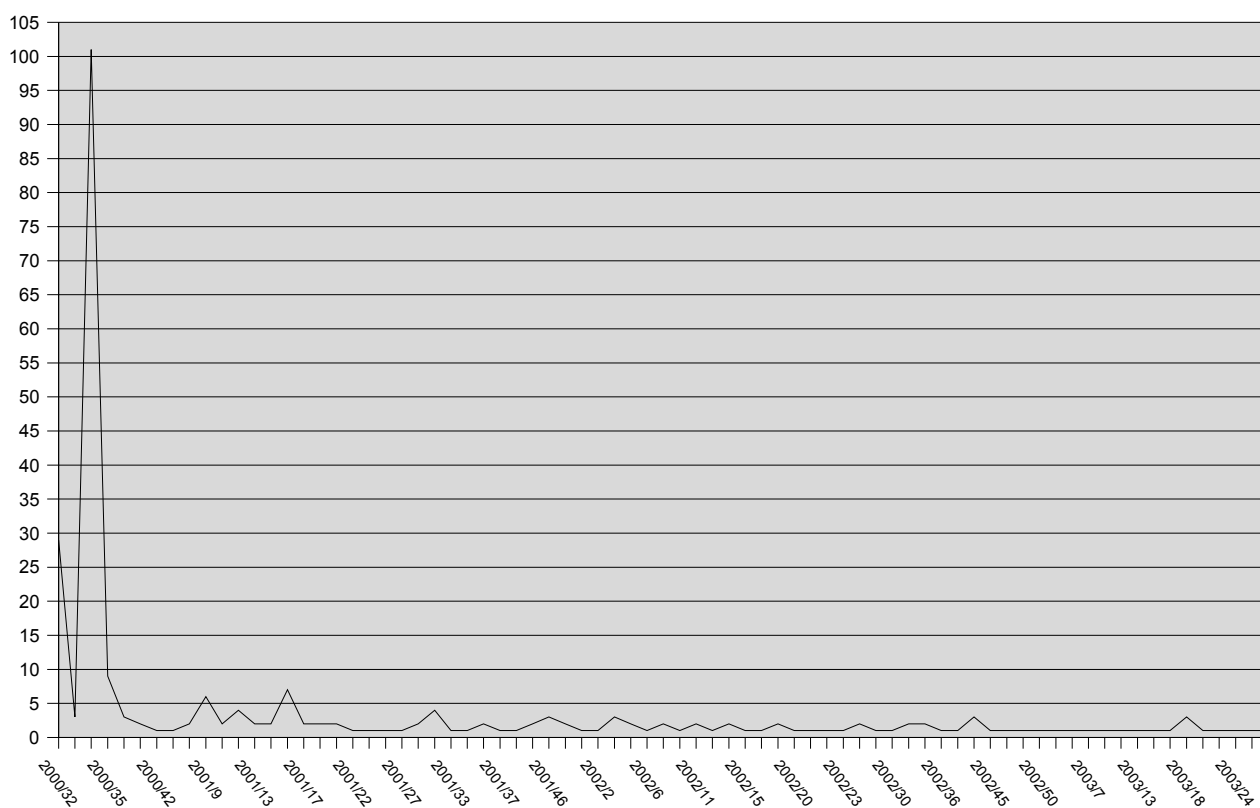
Ενδιαφέρον έχει επίσης όταν παρουσιάζεται για μικρό χρονικό διάστημα μεγάλος αριθμός επισκέψεων από συγκεκριμένο ιστοχώρο. Στο Crete TOURnet υπάρχουν 11 ιστοχώροι ως προέλευση από τους οποίους υπήρχαν τουλάχιστον 30 φορές περισσότερες εβδομάδες με περισσότερες από 60 επισκέψεις από εβδομάδες με λιγότερες από 10 επισκέψεις. Στο Διάγραμμα 5.6 φαίνεται χαρακτηριστικά το παράδειγμα του ιστοχώρου abcnews.go.com. Παρατηρούμε 2 κορυφές στην αρχή του διαγράμματος. Η πρώτη αντιστοιχεί στη διεύθυνση προέλευσης

<http://abcnews.go.com/sections/travel/AA/aa000802.html>

και η δεύτερη και υψηλότερη στη

<http://abcnews.go.com/sections/travel/DailyNews/Crete000815.html>

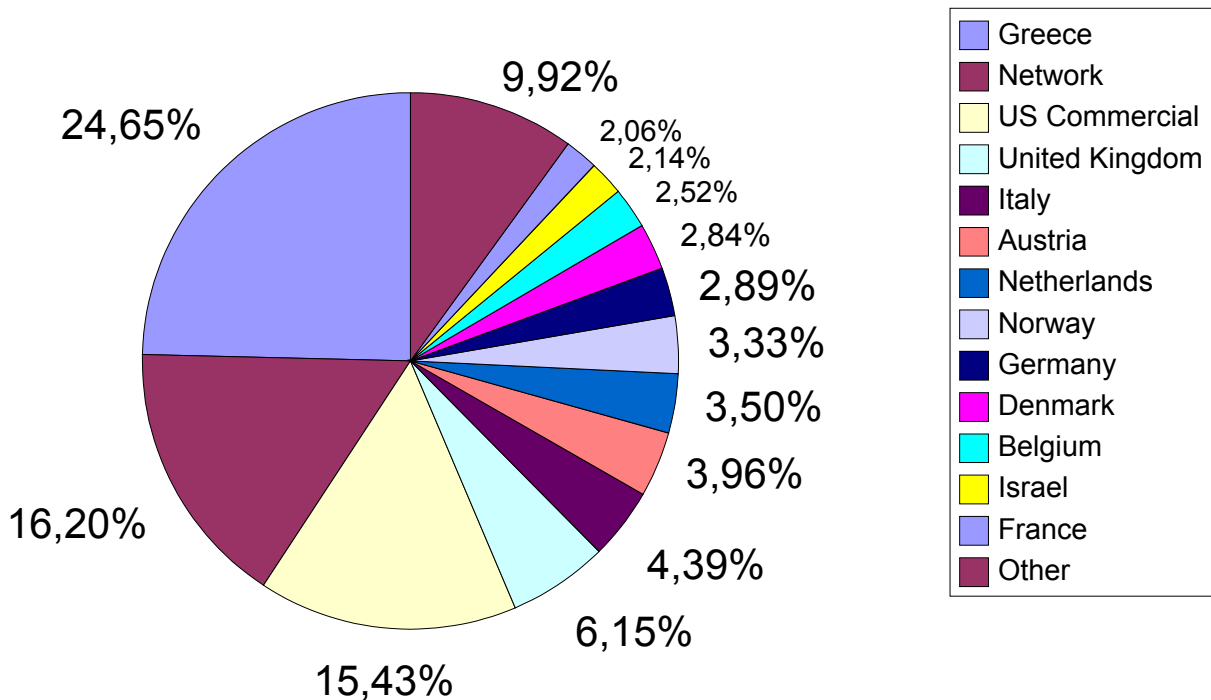
Στη συνέχεια υπάρχουν σποραδικές επισκέψεις κυρίως από τη δεύτερη διεύθυνση. Αυτή τη χρονική στιγμή μόνο εικασίες μπορούν να γίνουν για την ερμηνεία του φαινομένου, αλλά το πιο πιθανό είναι τα δύο άρθρα που αντιστοιχούν σε αυτές τις διευθύνσεις να υπήρχαν σε ιδιαίτερα προβεβλημένη θέση στο συγκεκριμένο ιστοχώρο με αποτέλεσμα οι επισκέπτες του ιστοχώρου να ακολουθούσαν και τους σχετικούς συνδέσμους. Αργότερα οι σελίδες τοποθετήθηκαν στο αρχείο του abcnews απ' όπου είναι ακόμα προσβάσιμες, αλλά με πολύ λιγότερους επισκέπτες.



Διάγραμμα 5.6 Επισκέψεις από τον ιστοχώρο abcnews.go.com. Παρατηρούμε ένα μικρό χρονικό διάστημα με πολλές επισκέψεις και στη συνέχεια μία απότομη και συνεχή πτώση. Τα δείγματα είναι ανά εβδομάδα

Όσον αφορά τις χώρες προέλευσης των επισκεπτών στο Διάγραμμα 5.7 φαίνονται οι χώρες προέλευσης των επισκεπτών για τον Ιούνιο 2003 όπως μπορεί να εκτιμηθούν από τις καταλήξεις των ονομάτων περιοχής (domain names) με βάση τη διεύθυνση IP των επισκεπτών (Ιούνιος 2003).

## Καταλήξεις Ονομάτων Περιοχής Επισκεπτών



Διάγραμμα 5.7 Χώρες προέλευσης από τις καταλήξεις των ονομάτων περιοχής (domain names) των επισκεπτών (Ιούνιος 2003)

Όπως εξηγήθηκε όμως στην ενότητα εισαγωγής, οι καταλήξεις ονομάτων περιοχής δε δίνουν σαφή ένδειξη για τις χώρες προέλευσης, ειδικά για τις καταλήξεις .com και .net που χρησιμοποιούνται από παρόχους υπηρεσιών διαδικτύου χωρίς εθνική ταυτότητα.

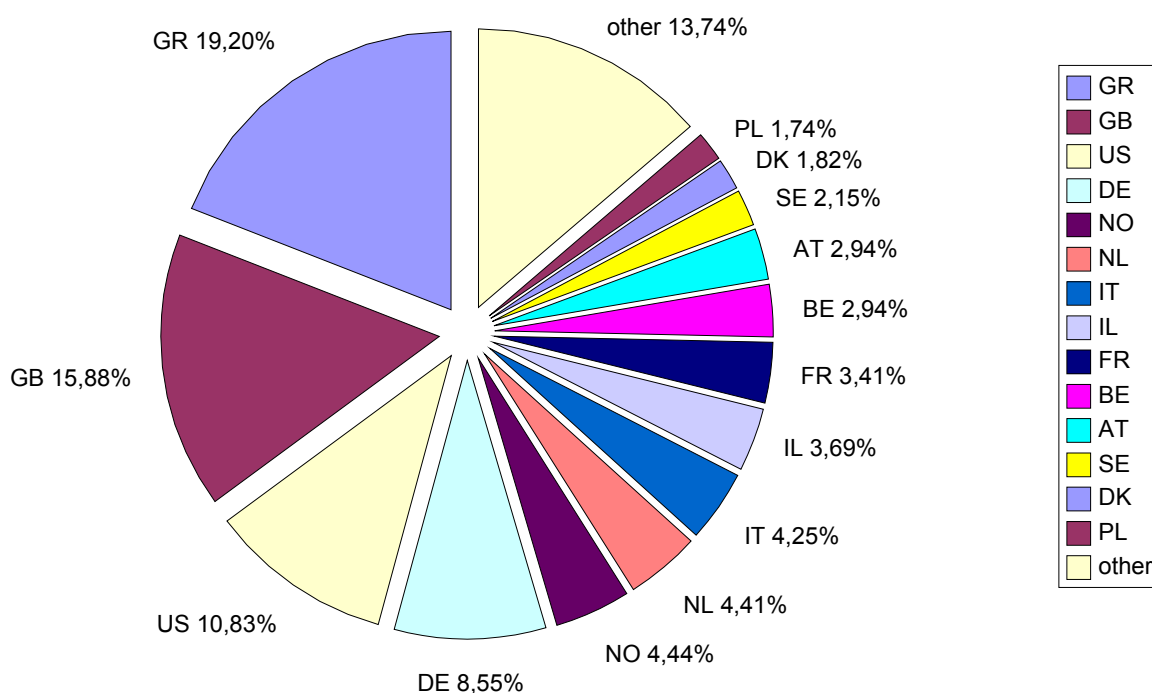
Γι' αυτό στο Διάγραμμα 5.8 χρησιμοποιείται η βάση δεδομένων της GeoIP για τις χώρες προέλευσης που διαφοροποιεί αρκετά τα ποσοστά. Για παράδειγμα οι επισκέπτες από Γερμανία που από τις καταλήξεις ονομάτων περιοχής φαίνεται να καταλαμβάνουν το 2,89%, με τα στοιχεία από την GeoIP φτάνουν στο 8,55% ενώ στη δεύτερη θέση έρχονται οι επισκέπτες από τη Μεγάλη Βρετανία με ποσοστό 15,88%!

Έχοντας υπόψιν ότι η πληροφορία του ιστοχώρου προσφέρεται στις γλώσσες ελληνικά, αγγλικά και γερμανικά, διαπιστώνεται ότι καλύπτει το 57% των επισκεπτών (Ελλάδα, Μεγάλη Βρετανία,



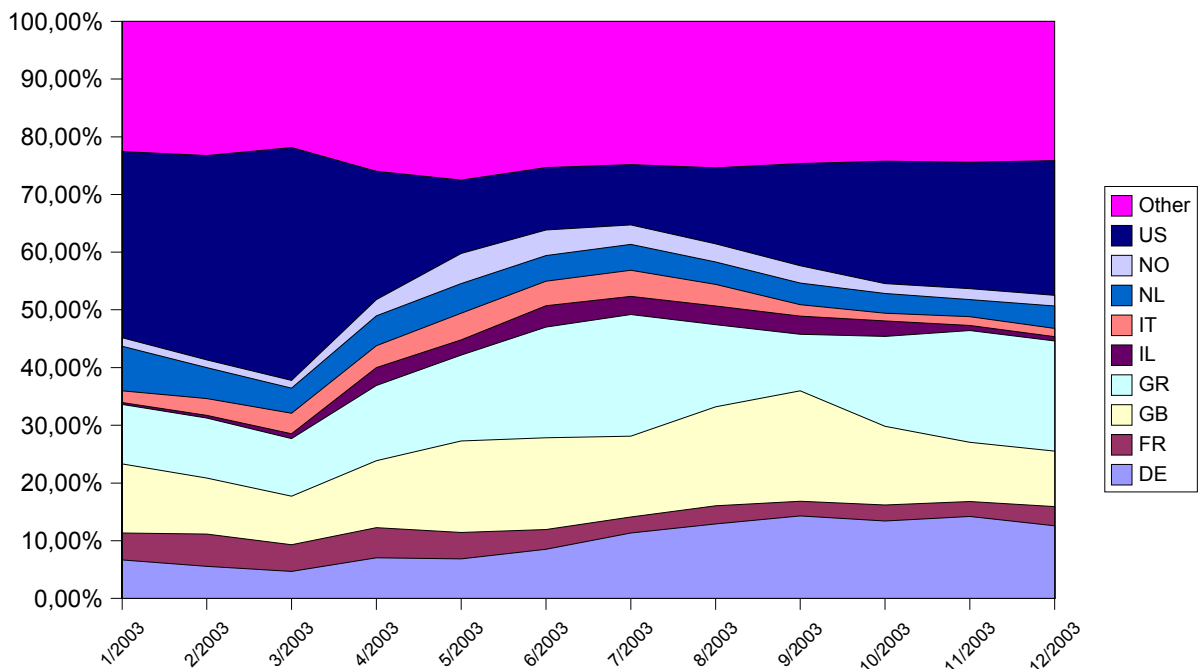
Γερμανία, Αυστρία) με τη γλώσσα της χώρας στην οποία ζούνε. Βέβαια είναι δεδομένο ότι εάν προσφερόταν ο ιστοχώρος σε τέταρτη γλώσσα θα υπήρχαν και περισσότεροι επισκέπτες από την αντίστοιχη χώρα διότι θα γινόταν δεικτοδότηση στις μηχανές αναζήτησης οι οποίες θα “έστελναν” τους επισκέπτες που αναζητούν στη μητρική τους γλώσσα σε μεγαλύτερο βαθμό στον ιστοχώρο.

Από αυτή την άποψη πρέπει να εξεταστεί μήπως υπάρχει στον υπάρχων ιστοχώρο σημαντικό ποσοστό επισκεπτών από χώρες που δεν καλύπτεται η αντίστοιχη γλώσσα. Σύμφωνα με το διάγραμμα το αμέσως μεγαλύτερο ποσοστό καταλαμβάνουν οι επισκέπτες από Νορβηγία (4,44%), Ολλανδία (4,41%), Ιταλία (4,25%), Ισραήλ (3,69%) και Γαλλία (3,41%). Τα ποσοστά επισκεπτών από αυτές τις χώρες είναι αρκετά μεγάλα και πρέπει να εξεταστεί το ενδεχόμενο μετάφρασης του ιστοχώρου σε αυτές τις γλώσσες, ειδικά εάν διαπιστωθεί ότι οι πραγματικοί τουρίστες επισκέπτες της περιοχής που καλύπτει ο ιστοχώρος προέρχονται όντως από αυτές τις χώρες.



Διάγραμμα 5.8 Χώρες προέλευσης επισκεπτών για τον Ιούνιο 2003 με πηγή τη βάση GeoIP

Ενδιαφέρον παρουσιάζει όμως και η κατανομή στις χώρες προέλευσης στη διάρκεια του χρόνου. Το σχετικό διάγραμμα φαίνεται στο Διάγραμμα 5.9. Το χαρακτηριστικό σε αυτές τις καμπύλες είναι ότι το ποσοστό των επισκεπτών από τις ΗΠΑ μειώνεται δραστικά την τουριστική περίοδο. Εικάζεται ότι οι επισκέπτες από τις άλλες χώρες (που είναι όλες Ευρωπαϊκές) ενδιαφέρονται περισσότερο για θέματα τουρισμού παρά για την υπόλοιπη πληροφορία που παρέχεται από τον ιστοχώρο.



Διάγραμμα 5.9 Κατανομή επισκεπτών ανά χώρα προέλευσης στη διάρκεια του 2003 ανά μήνα

### 5.3.1.3 Επισκεψιμότητα Σελίδων Ιστοχώρου

Στον πίνακα 5.8 και στο σχετικό με αυτόν Διάγραμμα 5.10 φαίνονται οι 33 διαφορετικές σελίδες που έχει ο ιστοχώρος και η κατανομή των σελιδοπροβολών ως ποσοστό επί του συνολικού αριθμού σελιδοπροβολών που έγιναν στον ιστοχώρο στο διάστημα από 10 Απριλίου έως 30 Ιουνίου 2003 (αναμορφωμένη έκδοση ιστοχώρου). Από τον πίνακα διακρίνεται ότι οι παρουσιάσεις επιχειρήσεων τουρισμού καταλαμβάνουν το μεγαλύτερο ποσοστό των σελιδοπροβολών, ακολουθούμενες από τα αποτελέσματα αναζητήσεων και τις παρουσιάσεις περιοχών ενδιαφέροντος.

Τα ποσοστά αυτά όμως δεν είναι απόλυτα ενδεικτικά για την πραγματική κατανομή των σελιδοπροβολών, διότι το περιεχόμενο των σελίδων με τα γραμμοσκιαμένα ονόματα μεταβάλλεται ανάλογα με τις παραμέτρους του URL. Έτσι οι σελίδες ar που αναφέρονται στην παρουσίαση περιοχών ενδιαφέροντος εμφανίζουν διαφορετική περιοχή ανάλογα με την παράμετρο για τον κωδικό της περιοχής.

<i>A/A</i>	<i>Όνομα Σελίδας</i>	<i>Ποσοστό επί Συνόλου</i>	<i>Περιγραφή</i>
1	sm	19,50%	Παρουσίαση επιχείρισης τουρισμού
2	searchresults	13,74%	Αποτελέσματα αναζήτησης
3	ar	9,95%	Παρουσίαση περιοχής ενδιαφέροντος
4	index	9,18%	Αρχική σελίδα
5	im	7,49%	Φωτογραφία/ φωτογραφίες
6	is	7,17%	Παρουσίαση ενότητας οδηγού
7	si	6,29%	Παρουσίαση αξιοθεάτου
8	maps3	4,86%	Χάρτης 3ου επιπέδου
9	maps2	3,47%	Χάρτης 2ου επιπέδου
10	hotels	2,98%	Φόρμας αναζήτησης διαμονής
11	maps1	2,19%	Χάρτης 1ου επιπέδου
12	guide	1,88%	Περιεχόμενα οδηγού
13	areaselecttool	1,85%	Εργαλείο επιλογής περιοχής από φόρμες αναζήτησης
14	show_all_related	1,84%	Εμφάνιση σχετικών επιχειρήσεων
15	sights	1,62%	Φόρμα αναζήτησης περιοχών και αξιοθεάτων
16	gc	1,56%	Γλωσσάρι
17	ha	0,91%	Παρουσίαση λιμανιού
18	rentcar	0,65%	Φόρμα αναζήτησης ενοικίασης οχήματος
19	areas_on_map	0,53%	Θέση περιοχής σε χάρτη
20	citymap	0,51%	Πολεοδομικός χάρτης
21	harbours	0,43%	Χάρτης λιμανιών
22	ev	0,43%	Παρουσίαση εκδήλωσης
23	brochure <sup>3</sup>	0,27%	Σελίδα “το φυλλάδιό μου”
24	travelagents	0,25%	Φόρμα αναζήτησης γραφείων τουρισμού
25	sendmail	0,19%	Αποστολή ηλεκτρονικού μηνύματος
26	advertise	0,09%	Σελίδα διαφήμισης
27	dm <sup>4</sup>	0,08%	Βυθομετρικός χάρτης λιμανιού
28	link	0,04%	Σελίδα για σύνδεσμο στον ιστοχώρο
29	contact	0,02%	Σελίδα επικοινωνίας
30	addinfo <sup>5</sup>	0,01%	Σελίδα για εισαγωγή επιχειρήσεων
31	about <sup>6</sup>	0,01%	Περί του ιστοχώρου

3 Ο σύνδεσμος για την πρόσβαση στο φυλλάδιο άλλαξε τοποθεσία και μορφή στις 18 Μαΐου 2003 με σημαντική αλλαγή στον αριθμό σελιδοπροβολών

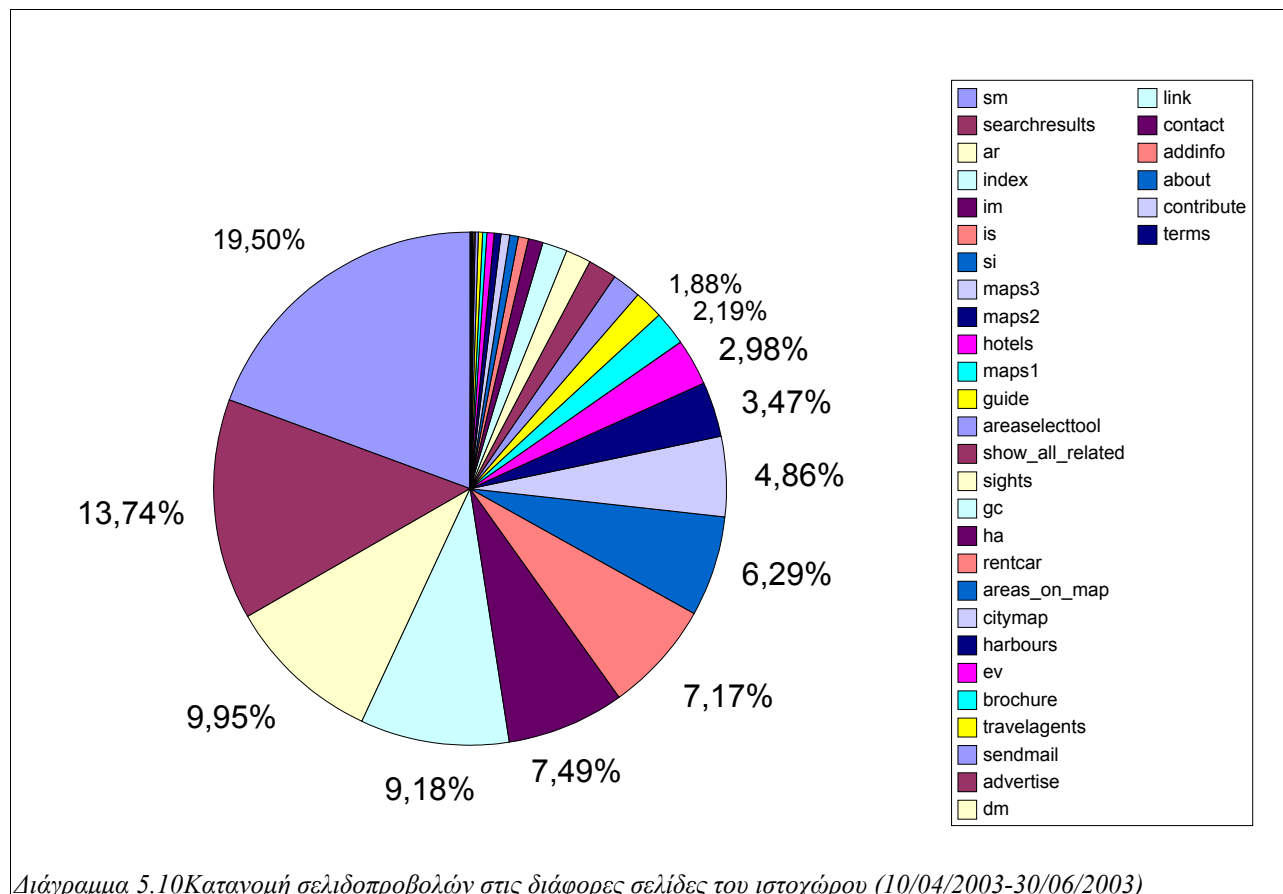
4 Η σελίδα καταργήθηκε στα τέλη Απριλίου

5 Μόνο στην ελληνική έκδοση

6 Μόνο στην ελληνική έκδοση

A/A	Όνομα Σελίδας	Ποσοστό επί Συνόλου	Περιγραφή
32	contribute <sup>7</sup>	0,00%	Σελίδα για συμβολή περιεχομένου
33	terms <sup>8</sup>	0,00%	Όροι χρήσης

Πίνακας 5.8 Κατανομή σελιδοπροβολών στις διάφορες σελίδες του ιστοχώρου (10/04/2003-30/06/2003). Τα γραμμοσκιασμένα ονόματα αναφέρονται σε σύνθετες σελίδες, παρουσιάζουν δηλαδή διαφορετικό περιεχόμενο της ίδιας κατηγορίας ανάλογα με τις παραμέτρους κλήσης του URL.

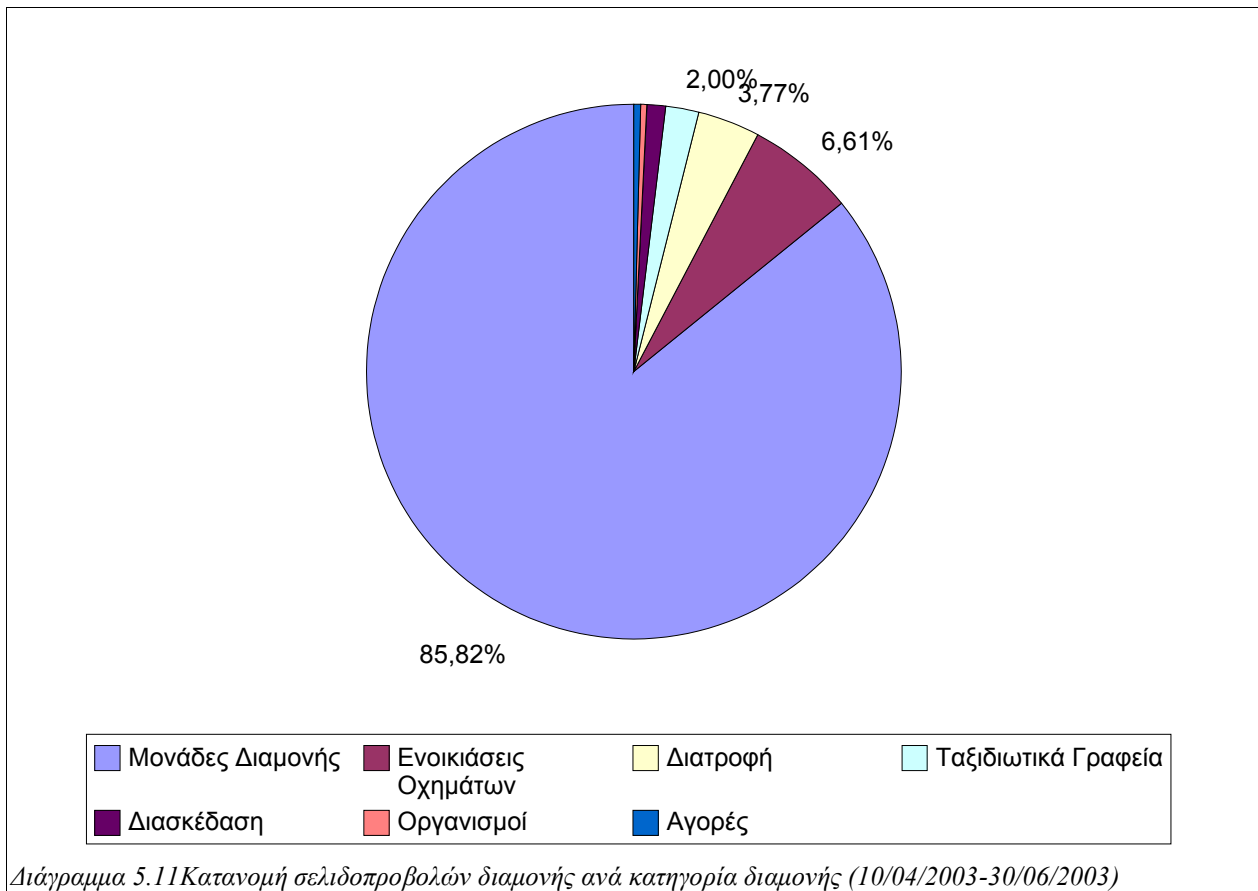


Ειδικά για τη σελίδα με το μεγαλύτερο ποσοστό σελιδοπροβολών, τις επιχειρήσεις τουρισμού δηλαδή, υπάρχει ενδιαφέρον να φανεί σε τι κατηγορίες επιχειρήσεων μοιράζονται οι σελιδοπροβολές. Αυτό είναι δυνατόν να βρεθεί επεκτείνοντας την ερώτηση στη βάση δεδομένων του Crete TOURnet (εκτός δηλαδή της Βάσης Χρήσης). Έτσι, όπως φαίνεται στο Διάγραμμα 5.11 οι επιχειρήσεις διαμονής συγκεντρώνουν με ποσοστό 85,82% το μεγαλύτερο μέρος των σελιδοπροβολών, και με αυτό τις προτιμήσεις των επισκεπτών του ιστοχώρου. Το μεγάλο ποσοστό της αρχικής σελίδας εξηγείται από το γεγονός ότι παρά τη μεγάλη διασύνδεση μεταξύ των σελίδων του ιστοχώρου, η αρχική σελίδα αποτελεί το σημείο εκκίνησης για πολλές ενέργειες του ιστοχώρου. Η σελίδα για τα αποτελέσματα αναζήτησης εμφανίζεται ως ενδιάμεση σε όλες τις

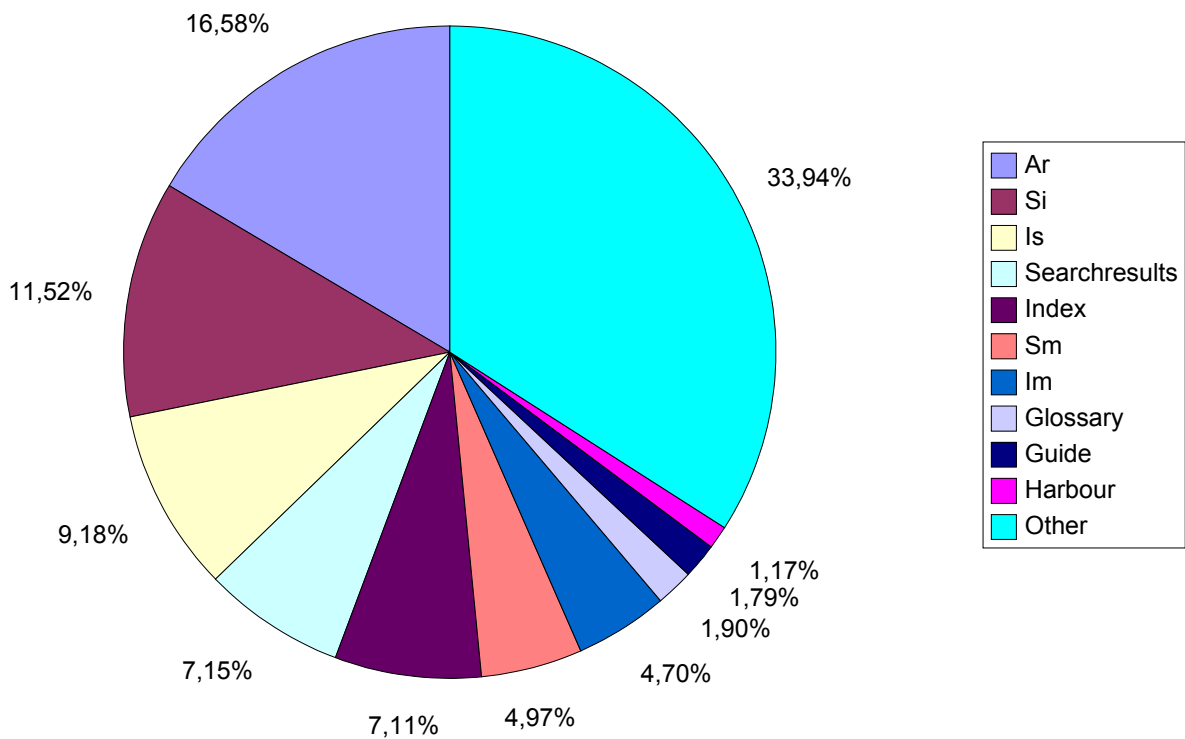
7 Μόνο στην ελληνική έκδοση

8 Μόνο στην ελληνική έκδοση

αναζητήσεις που γίνονται, και γι' αυτό καταλαμβάνει τη δεύτερη θέση σε ποσοστό σελιδοπροβολών.

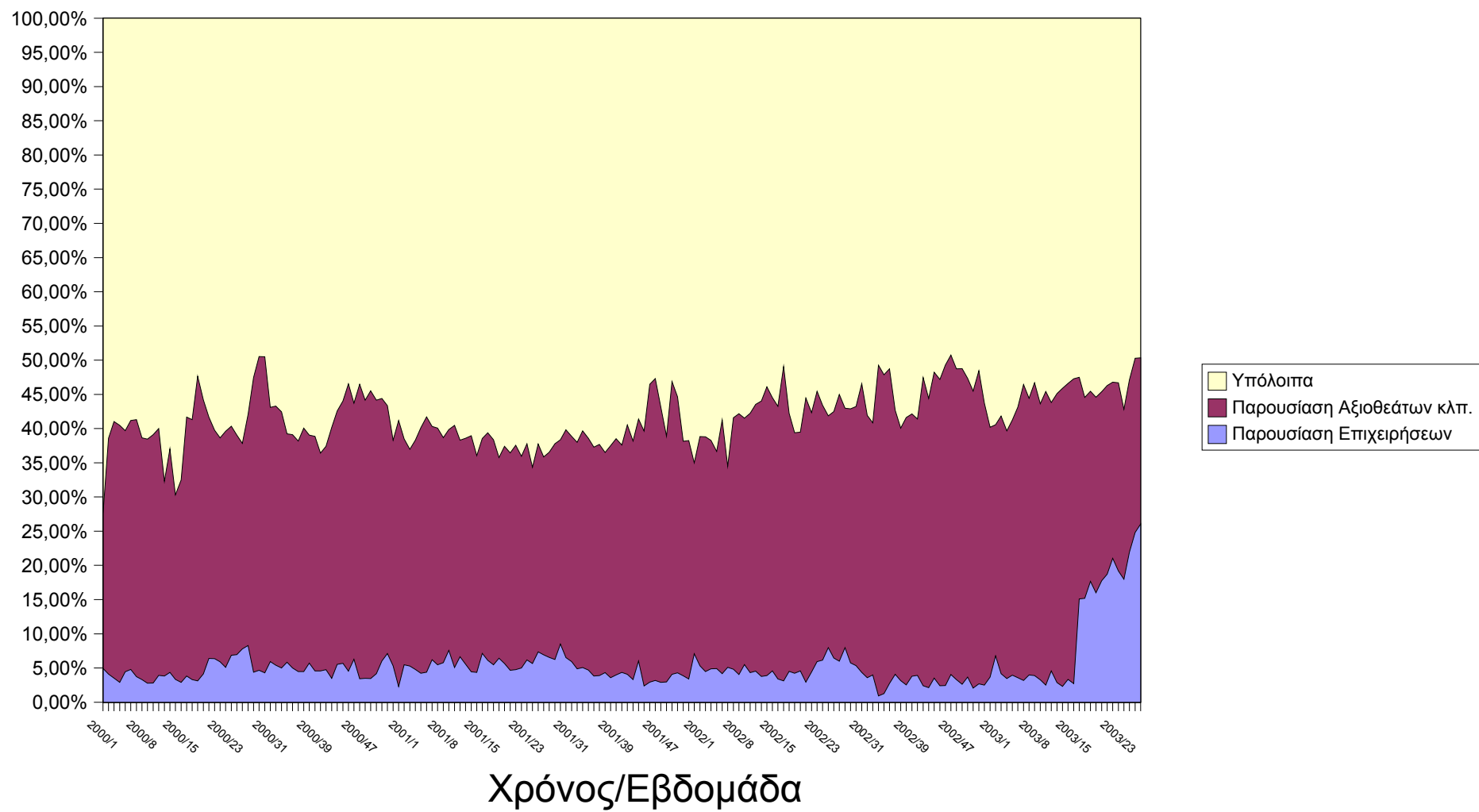


Όπως αναφέρθηκε, τα παραπάνω ποσοστά είναι από το διάστημα της ανανεωμένης λειτουργίας του ιστοχώρου. Το αντίστοιχο διάγραμμα για την προ της ανανέωσης λειτουργίας δίνει μία εντελώς διαφορετική εικόνα για τις προτιμήσεις των επισκεπτών. Στο Διάγραμμα 5.12 φαίνεται η αντίστοιχη κατανομή σελίδων για το διάστημα στη διάρκεια του 2002 (τα ονόματα παρόλο που ήταν διαφορετικά, έχουν προσαρμοστεί για ευκολότερη σύγκριση). Οι παρουσιάσεις επιχειρήσεων τουρισμού καταλαμβάνουν μόνο το 4,97% (sm) του συνόλου των σελιδοπροβολών ενώ οι σελιδοπροβολές για πληροφορίες γενικά με την Κρήτη (περιοχές, αξιοθέατα, οδηγός) καταλαμβάνουν συνολικά το 39,07% (ar, si, is, guide)). Αυτή η μεγάλη διαφορά που εμφανίστηκε αμέσως μετά την ενεργοποίηση της νέας σχεδίασης του ιστοχώρου (οπότε δεν οφείλεται σε παράγοντες όπως ταχύτητα πρόσβασης ή ενεργοποίησης νέου συνδέσμου προέλευσης από άλλον ιστοχώρο) θα ερμηνευτεί στην ανάλυση χρησιμοποίησιμότητας.

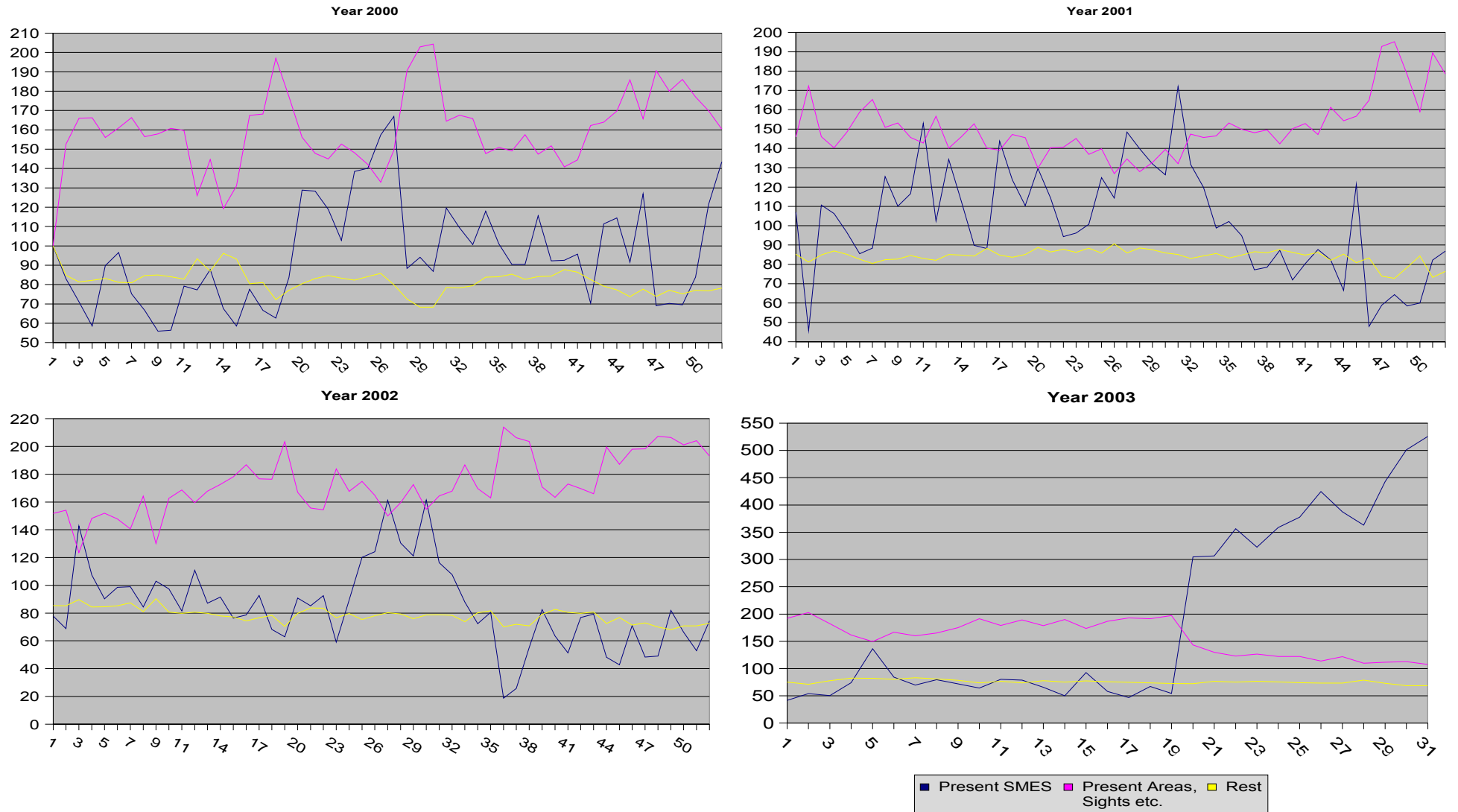


Διάγραμμα 5.12 Κατανομή σελιδοπροβολών στις διάφορες σελίδες του ιστοχώρου (1/1/2002-31/12/2002)

Ένα άλλο ενδιαφέρον ερώτημα είναι εάν παρουσιάζονται σημαντικές μεταβολές στην κατανομή σελιδοπροβολών ανά κατηγορία σελίδων. Ιδιαίτερα ενδιαφέρει η σχετική αναλογία ανάμεσα στις σελιδοπροβολές επιχειρήσεων τουρισμού και τις υπόλοιπες κατηγορίες περιεχομένου (με κύρια τις περιοχές ενδιαφέροντος, τα αξιοθέατα και τις ενότητες εισαγωγής). Στο Διάγραμμα 5.13 απεικονίζονται τα ποσοστά σελιδοπροβολών στη διάρκεια του χρόνου σε κατηγορίες επιχειρήσεων τουρισμού, πληροφορίες περιοχών ενδιαφέροντος κλπ. και των υπολοίπων σελίδων. Ειδικά μετά το 2001 όπου, όπως αναφέρθηκε, το σύστημα καταγραφής λειτουργεί στην πιο αξιόπιστη μορφή, διακρίνεται η αύξηση του ποσοστού των επιχειρήσεων τουρισμού τους καλοκαιρινούς μήνες. Στο Διάγραμμα 5.14 απεικονίζεται ο ρυθμός μεταβολής των ποσοστών όπου αυτό το φαινόμενο είναι πιο διακριτό.



Διάγραμμα 5.13 Σχετική κατανομή σελιδοπροβολών στη διάρκεια του χρόνου. Φαίνεται το ποσοστό επί του συνόλου των σελιδοπροβολών επιχειρήσεων τουρισμού, περιοχών ενδιαφέροντος (μαζί με αξιοθέατα και ενότητες εισαγωγής) και των υπολοίπων σελίδων. Ο οριζόντιος άξονας του χρόνου είναι ανά εβδομάδα (1/1/2002-30/6/2003)



Διάγραμμα 5.14 Ρυθμός μεταβολής ποσοστών ανά κατηγορία σελίδων στον ιστοχώρο. Είναι εμφανής η υπεροχή των σελίδων για επιχειρήσεις τουρισμού τους καλοκαιρινούς μήνες, έναντι των περιοχών ενδιαφέροντος (1/1/2002-30/6/2003)

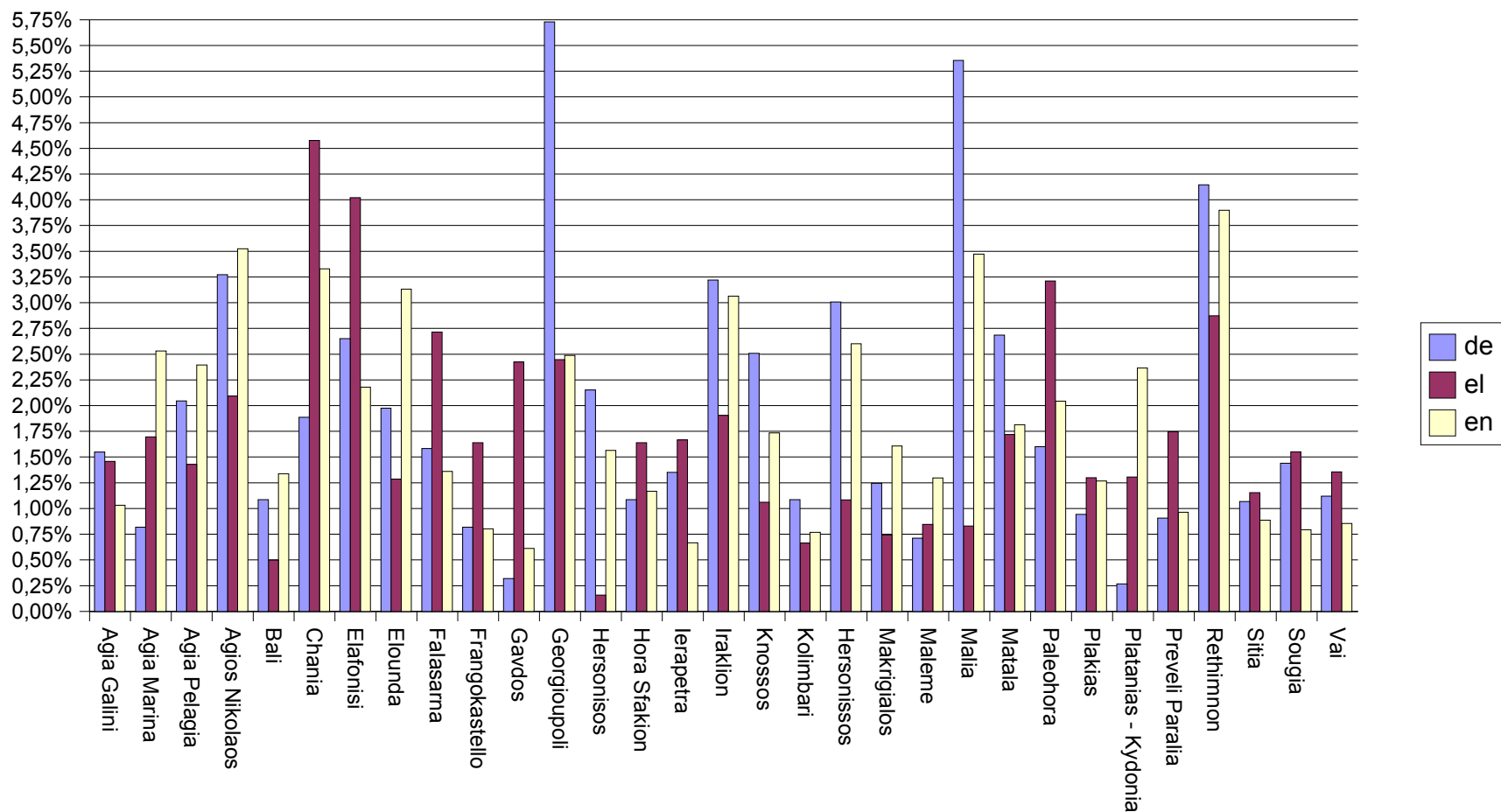


Από τη συχνότητα παρουσίασης ανά περιοχή ενδιαφέροντος μπορεί να εξαχθούν συμπεράσματα για τις δημοφιλές περιοχές των επισκεπτών του ιστοχώρου. Στον πίνακα 5.9 φαίνονται οι απόλυτοι αριθμοί των προβολών σελίδων περιοχών ενδιαφέροντος ανά περιοχή και ανά γλώσσα εμφάνισης (γερμανικά, ελληνικά και αγγλικά). Εάν υπολογιστούν να ποσοστά που καταλαμβάνει κάθε περιοχή ανά γλώσσα τότε προκύπτει η εικόνα στο Διάγραμμα 5.15.

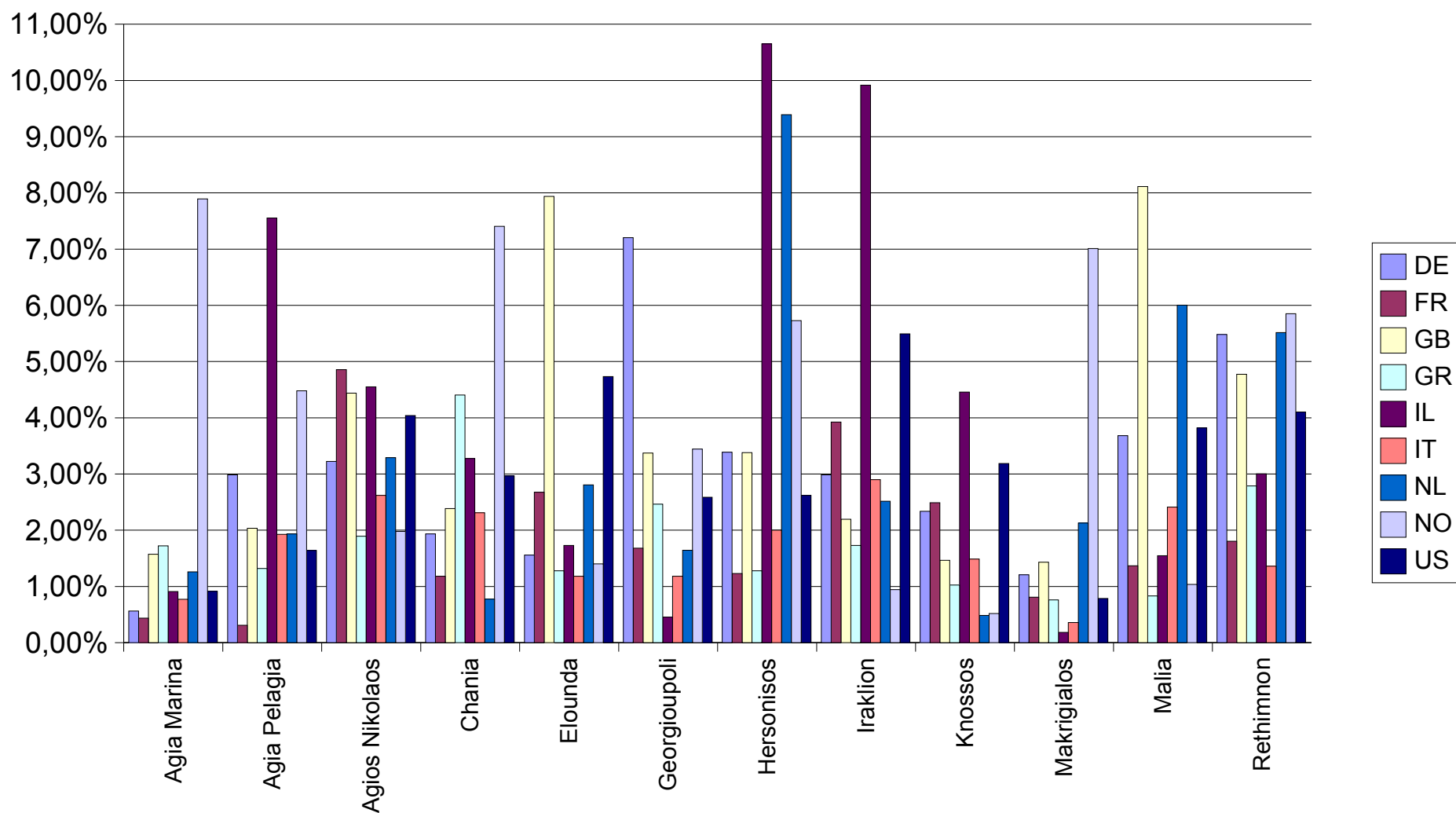
<i>Περιοχή</i>	<i>de</i>	<i>el</i>	<i>en</i>
Agia Galini	87	202	334
Agia Marina	46	235	820
Agia Pelagia	115	198	776
Agios Nikolaos	184	290	1142
Bali	61	69	433
Chania	106	634	1079
Elafonisi	149	557	706
Elounda	111	178	1015
Falasarna	89	376	441
Frangokastello	46	227	260
Gavdos	18	336	198
Georgiupoli	322	339	806
Hersonisos	121	22	507
Hora Sfakion	61	227	378
Ierapetra	76	231	216
Iraklion	181	264	993
Knossos	141	147	562
Kolimbari	61	92	249
Hersonissos	169	150	843
Makrigialos	70	103	521
Maleme	40	117	420
Malia	301	115	1125
Matala	151	238	588
Paleohora	90	445	662
Plakias	53	180	411
Platanias - Kydonia	15	181	766
Preveli Paralia	51	242	312
Rethimnon	233	398	1263
Sitia	60	160	287
Sougia	81	215	257
Vai	63	188	277

*Πίνακας 5.9.Αριθμός σελιδοπροβολών ανά περιοχή ενδιαφέροντος και ανά γλώσσα εμφάνισης. Εμφανίζονται οι περιοχές που βρίσκονται στις 23 πρώτες προτιμήσεις σε κάποια γλώσσα προβολής (Ιούνιος 2003)*

Ποσοστά Επισκεψιμότητας Περιοχών ανά Γλώσσα Προβολής



Διάγραμμα 5.15 Ποσοστά επισκεψιμότητας ανά γλώσσα εμφάνισης και ανά περιοχή. Χαρακτηριστικές διαφορές ανάμεσα στους έλληνες επισκέπτες και τους ξένους (Ιούνιος 2003)



Διάγραμμα 5.16 Χαρακτηριστικές περιπτώσεις διαφοροποίησης επισκεψιμότητας περιοχών ενδιαφέροντος ανά χώρα προέλευσης. Φαίνονται τα ποσοστά προβολών περιοχών ανά χώρα προέλευσης

Οι διαφορές ανά γλώσσα εμφάνισης στις προτιμήσεις των επισκεπτών είναι χαρακτηριστικές για τις διαφορές στις προτιμήσεις των ελλήνων επισκεπτών και των ξένων. Στους ξένους κυριαρχούν οι περιοχές με έντονο το φαινόμενο του μαζικού τουρισμού (Μάλια, Χερσόνησος, Γεωργιούπολη ιδιαίτερα για τους γερμανόφωνους,) ενώ οι έλληνες δείχνουν προτίμηση για περιοχές που χαρακτηρίζονται αντιθέτως για λιγότερο τουριστικά εκμεταλλευόμενες (Χανιά, Φαλάσαρνα, Γαύδος, Παλαιόχωρα, Σούγια κλπ.). Χαρακτηριστική είναι η διαφορά για τη Γαύδο όπου το ποσοστό των Ελλήνων επισκεπτών είναι πολλαπλάσιο του αντίστοιχου ποσοστού των ξενόγλωσσων επισκεπτών.

Εάν γίνει ανάλογος υπολογισμός ανά χώρα προέλευσης των επισκεπτών, προκύπτουν ακόμα πιο χαρακτηριστικές περιπτώσεις διαφοροποίησης. Στο Διάγραμμα 5.16 φαίνονται επιλεγμένες περιοχές όπου υπάρχει σημαντική διαφοροποίηση ανάλογα με τη χώρα προέλευσης. Για παράδειγμα οι επισκέπτες από Νορβηγία στο σύνολο των περιοχών ενδιαφέροντος που επισκέπτονται βλέπουν την περιοχή Αγία Μαρίνα στο 8% των περιπτώσεων, σε αντίθεση με τις άλλες χώρες όπου η περιοχή είναι σε χαμηλά ποσοστά. Οι τουριστικά πολύ ανεπτυγμένες περιοχές Χερσόνησος και Μάλια, που βρίσκονται πολύ κοντά γεωγραφικά, δείχνουν μία μεγάλη διαφοροποίηση ανάμεσα στους επισκέπτες από Ισραήλ και Μεγάλη Βρετανία. Οι επισκέπτες από τη Μεγάλη Βρετανία ενδιαφέρονται πολύ περισσότερο (περίπου 8%) για τα Μάλια αντί για τη Χερσόνησο (περίπου 3,4%). Σε αντίθεση με αυτούς οι επισκέπτες από το Ισραήλ ενδιαφέρονται για τη Χερσόνησο πολύ περισσότερο (πάνω από 11%) ενώ για τα Μάλια πολύ λιγότερο (λιγότερο από 2%).

### 5.3.2 Προσωποποίηση Ιστοχώρου

Στο τελευταίο διάγραμμα της προηγούμενης ενότητας (Διάγραμμα 5.15 σελίδα 219) φάνηκε ήδη μία σημαντική διαφοροποίηση των επισκεπτών ανάλογα με τη γλώσσα εμφάνισης που επιλέγουν στον ιστοχώρο. Με βάση τις αρχές που περιγράφηκαν στο κεφάλαιο της μεθοδολογίας ανάλυσης, τα αποτελέσματα που φαίνονται στο διάγραμμα προσφέρονται για την προσωποποίηση του ιστοχώρου ανάλογα με τη γλώσσα που επιλέγει ο κάθε επισκέπτης. Υπάρχουν δύο επιλογές που μπορεί να γίνουν, ανάλογα με την πολιτική που ακολουθεί ο ιστοχώρος:

- **προβολή λιγότερο προβεβλημένων περιοχών:** σε αυτή την περίπτωση ενδιαφέρει να προβληθούν περιοχές που δεν είναι τόσο γνωστές στην κατηγορία επισκεπτών. Αυτό σημαίνει ότι για παράδειγμα στην γερμανική και αγγλική έκδοση θα εμφανιστούν σε ειδικό σημείο στον ιστοχώρο λίστα με συνδέσμους για περιοχές που δεν ανήκουν στις τουριστικά ανεπτυγμένες περιοχές (Μάταλα, Γεωργιούπολη κλπ.).

- **πιο άμεση πρόσβαση στις δημοφιλείς περιοχές:** σε αυτή την περίπτωση επιδιώκουμε να προσφέρουμε στον επισκέπτη πιο άμεση πρόσβαση στις περιοχές που σύμφωνα με τη γλώσσα εμφάνισης επιθυμεί. Θα εμφανιστεί επομένως σε ειδικό σημείο στον ιστοχώρο λίστα με τις δημοφιλείς περιοχές όπως προκύπτει από το διάγραμμα. Έτσι στην ελληνική έκδοση θα συμπεριληφθούν για παράδειγμα οι περιοχές Χανιά, Ελφονήσι, Φαλάσαρνα κλπ.)

Ένα άλλο σημείο που περιγράφηκε στη μεθοδολογία ανάλυσης είναι η προσαρμογή του ιστοχώρου σε ασυνήθιστες αλλαγές.

### **5.3.2.1 Ασυνήθιστες Αλλαγές**

Από την περίπτωση των προσωρινά αυξημένων επισκέψεων με προέλευση το abcnews (Διάγραμμα 5.6 σελίδα 207) έχει γίνει ήδη εμφανές ότι τέτοιου είδους απότομες αλλαγές στους επισκέπτες του ιστοχώρου παρουσιάζουν ιδιαίτερο ενδιαφέρον εάν συνδυαστούν με τη γνώση για τα ενδιαφέροντά τους.

Για τις διευθύνσεις προέλευσης με συνθήκες

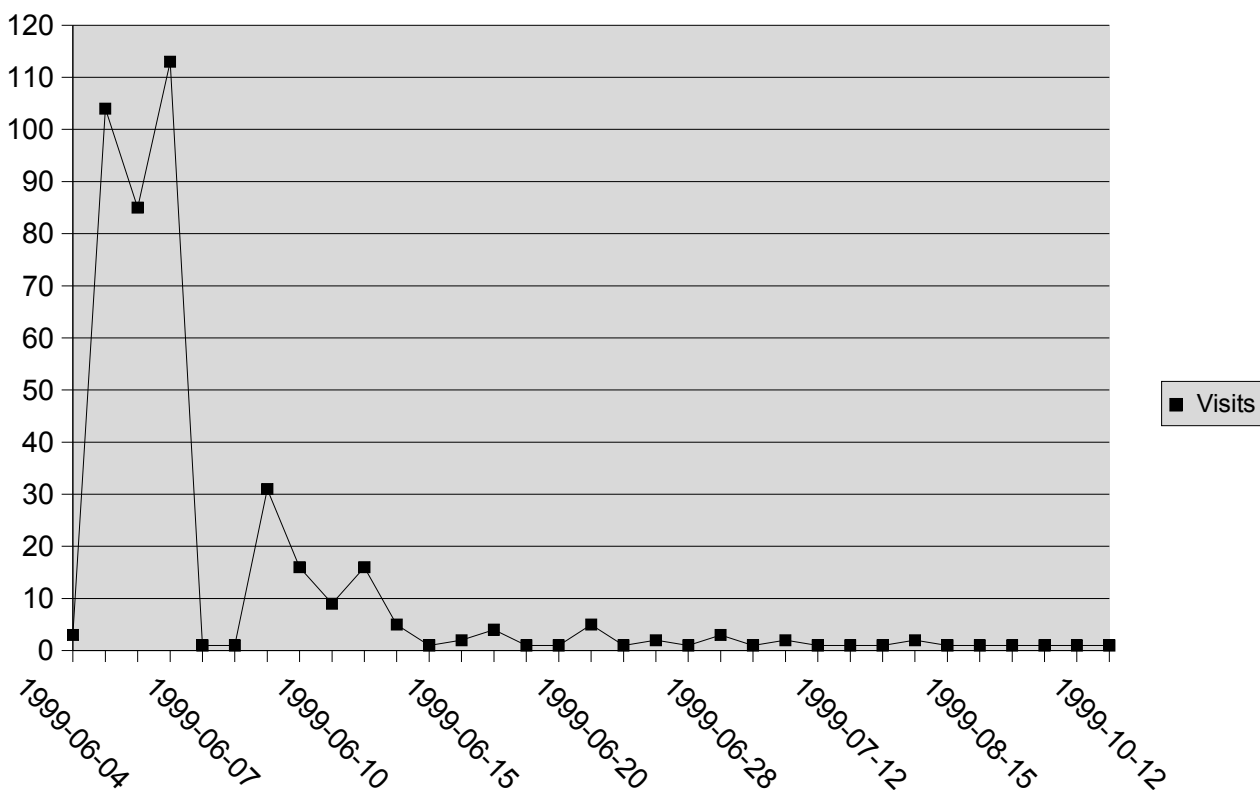
- η μεταβολή σε σχέση με την προηγούμενη ημέρα να είναι μεγαλύτερη από 100%
- το ποσοστό των συνολικών επισκέψεων από τη συγκεκριμένη διεύθυνση προέλευσης να είναι μεγαλύτερο του 1%

και τη συνολική χρονική περίοδο λειτουργίας του ιστοχώρου, προκύπτουν συνολικά 304 περιπτώσεις διευθύνσεων που θα ενεργοποιούσαν σχετική ειδοποίηση (βλέπε και Ερώτηση 4.14 σελίδα 148) . Αυτές ανήκουν σε 50 μοναδικές διευθύνσεις, το οποίο σημαίνει ότι υπάρχουν πολλές περιπτώσεις όπου οι παρατηρούμενες αυξομειώσεις πρέπει να θεωρηθούν φυσιολογικές. Μάλιστα, με περαιτέρω έρευνα, αναδεικνύεται ότι η διαφορά των 50 από τις 304 περιπτώσεις προέρχεται κυρίως από τον ιστοχώρο του Δήμου Χανίων, και σε ποσοστό 70% ανήκουν σε ημέρα Δευτέρα. Το τελευταίο εξηγείται με τη μεγάλη διακύμανση που έχει ο ιστοχώρος του Δήμου Χανίων ανάμεσα στο Σαββατοκύριακο και τις καθημερινές.

Χαρακτηριστικές περιπτώσεις από τις παραπάνω διευθύνσεις είναι η περίπτωση του abcnews που περιγράφηκε στο Διάγραμμα 5.6 σελίδα 207, καθώς και

- η διεύθυνση <http://www.gmx.net/frauen.html> (δεν είναι ενεργή πλέον) με 330% αύξηση στις 5 Ιουνίου 1999 Διάγραμμα 5.17
- η διεύθυνση <http://www.iasted.org/conferences/2002/greece/sppra.htm> με 240% αύξηση στις 16

Ιανουαρίου του 2002



Διάγραμμα 5.17 Εξέλιξη επισκέψεων από <http://www.gmx.net/frauen.html>. Διακρίνεται ότι η πρώτη επίσκεψη ήταν στις 4 Ιουνίου 1999 με απότομη αύξηση την αμέσως επομένη. Οι επισκέψεις διατηρήθηκαν σε μεγάλο αριθμό 4 ημέρες, ώσπου να σταματήσουν εντελώς μετά το Νοέμβριο του 1999, όπου και πιθανότατα αφαιρέθηκε η σελίδα.

Ενώ η πρώτη περίπτωση (<http://www.gmx.net/frauen.html>) δεν μπορεί να εξακριβωθεί πλέον επειδή δεν υπάρχει πλέον σαν σελίδα, η δεύτερη περίπτωση (<http://www.iasted.org/conferences/2002/greece/sppra.htm>) αναφέρεται σε συνέδριο στην Κρήτη όπου τοποθετήθηκε σύνδεσμος προς το Crete TOURnet ως ιστοχώρο με σχετική για την Κρήτη πληροφορία. Επειδή η σελίδα απευθύνεται κυρίως σε συμμετέχοντες στο συνέδριο, θα ήταν λογικό να υποθέσει κανείς ότι οι επισκέπτες του Crete TOURnet από τη συγκεκριμένη διεύθυνση θα θέλουν μεταξύ άλλων

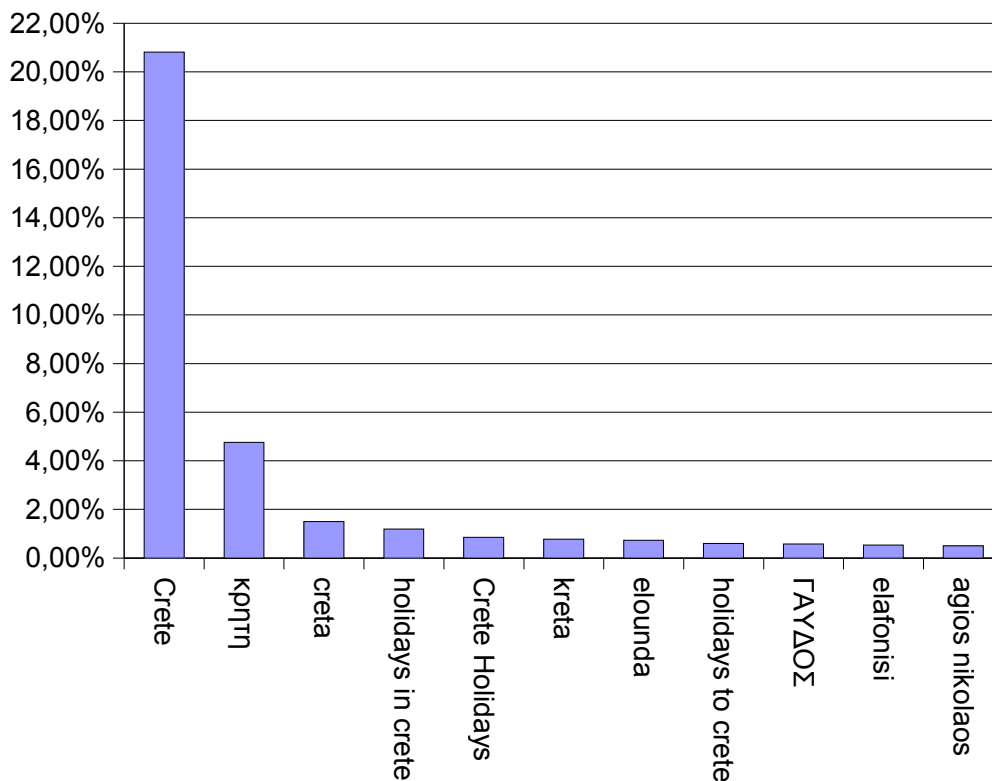
- να κλείσουν διαμονή στην περιοχή του συνεδρίου
- να ενημερωθούν για αξιοθέατα στην περιοχή του συνεδρίου

Η κατάλληλη προσαρμογή του ιστοχώρου ώστε στις επισκέψεις από αυτή τη διεύθυνση, και για την περίοδο πριν τη διεξαγωγή του συνεδρίου, ώστε να είναι πιο προσιτές οι παραπάνω πληροφορίες,

είναι μία εύκολη υπόθεση και εξασφαλίζει την μεγαλύτερη διευκόλυνση μέρους των επισκεπτών.

Η εικόνα που παρουσιάζεται με την παραπάνω μέθοδο στην περίπτωση που ως κριτήριο αντί της προέλευσης πάρουμε τις φράσεις αναζήτησης δεν είναι τόσο ξεκάθαρη. Όπως φαίνεται στο Διάγραμμα 5.18, που απεικονίζει τα ποσοστά των επισκέψεων που προήλθαν με συγκεκριμένες φράσεις αναζήτησης, υπάρχει μία μεγάλη διασπορά στις φράσεις αυτές λόγω της ποικιλομορφίας της πληροφορίας που παρουσιάζεται. Συγκεκριμένα από τις 5128 φράσεις που χρησιμοποιήθηκαν τον Ιούνιο του 2003 για πρόσβαση στο Crete TOURnet, μόνο οι 11 καταλαμβάνουν ποσοστό μεγαλύτερο του 0,5%, και μάλιστα από αυτές 4 έχουν ποσοστό μεγαλύτερο του 1%.

Οπότε υπάρχει μεγάλη δυσκολία στον εντοπισμό φράσεων που αυξάνουν προσωρινά σε σημαντικό βαθμό τις επισκέψεις στο Crete TOURnet.



Διάγραμμα 5.18 Φράσεις αναζήτησης ως ποσοστά επί των επισκέψεων στις οποίες χρησιμοποιήθηκαν (Ιούνιος 2003)

### 5.3.3 Σύγκριση Εναλλακτικών

Οι μελέτες χρησιμοποίησιμότητας πολλές φορές έχουν ως αντικειμενικό σκοπό τη σύγκριση εναλλακτικών μορφών αλληλεπίδρασης με το χρήστη. Εξετάζεται δηλαδή κατά πόσο ένας συγκεκριμένος σχεδιασμός του επιπέδου διεπικοινωνίας είναι περισσότερο αποδοτικός από κάποιον άλλο. Οι εξεταζόμενες παραλλαγές δε διαφέρουν κατ' ανάγκη ριζικά, αλλά μπορεί να αναφέρονται



σε μικρά τμήματα της όλης εφαρμογής. Στην ακραία περίπτωση οι συγκρινόμενες εναλλακτικές μορφές του επιπέδου διεπικοινωνίας μπορεί να διαφέρουν εντελώς και να αναφέρονται σε ανταγωνιστικές εφαρμογές.

Θα περιγραφούν τρεις χαρακτηριστικές περιπτώσεις σύγκρισης εναλλακτικών μορφών επιπέδων διεπικοινωνίας που έγιναν στα πλαίσια της εργασίας.

### **5.3.3.1 Ανάλυση Επισκεψιμότητας Επιχειρήσεων Τουρισμού**

Στην προηγούμενη ενότητα, όπου παρουσιάστηκαν, ορισμένα στατιστικά στοιχεία για τα μεγέθη του ιστοχώρου, έγινε ήδη εμφανές ότι η 10η Απριλίου όπου άλλαξε ριζικά η εμφάνιση του ιστοχώρου, μπορεί να θεωρηθεί ορόσημο στην εξέλιξη του Crete TOURnet.

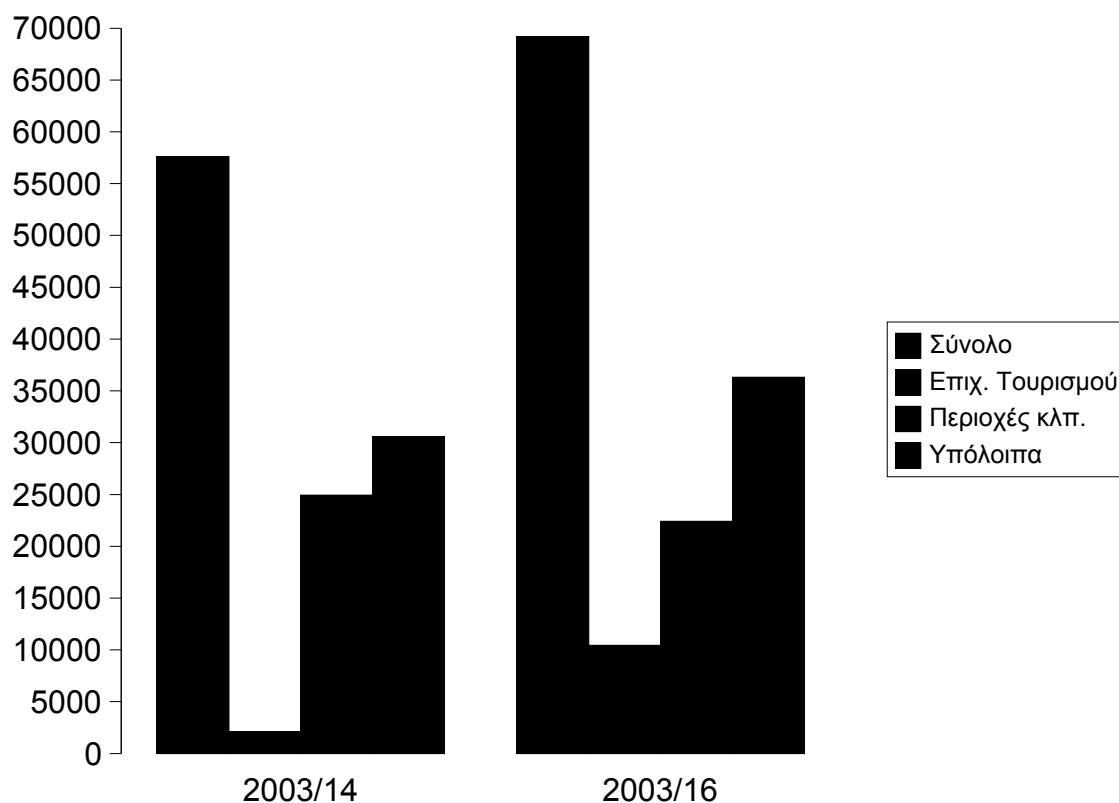
Κύρια χαρακτηριστικά των αλλαγών που επήλθαν είναι

1. σημαντική μείωση του ποσοστού επισκέψεων με λίγες σελιδοπροβολές ανά επίσκεψη (1-4) και ανάλογη αύξηση των επισκέψεων με φυσιολογικό ή μεγάλο αριθμό σελιδοπροβολών ανά επίσκεψη (Διάγραμμα 5.3 σελίδα 201)
2. μεγάλη αύξηση του μέσου όρου σελιδοπροβολών ανά επίσκεψη (Διάγραμμα 5.2 σελίδα 199)
3. μεγάλη αύξηση του ποσοστού σελίδων παρουσίασης επιχειρήσεων τουρισμού (Διάγραμμα 5.13 σελίδα 215)

Τα σημεία 2 και 3 έχουν άμεση σχέση μεταξύ τους, διότι η αύξηση των σελιδοπροβολών ανά επίσκεψη αντιστοιχεί σε μεγάλο βαθμό στις περισσότερες παρουσιάσεις επιχειρήσεων τουρισμού.

Στο Διάγραμμα 5.19 παρουσιάζεται ο απόλυτος αριθμός σελιδοπροβολών την εβδομάδα πριν και την εβδομάδα μετά την αλλαγή του ιστοχώρου. Αναφέρονται το σύνολο των σελιδοπροβολών, ο αριθμός των σελιδοπροβολών για επιχειρήσεις τουρισμού, ο αριθμός των σελιδοπροβολών για περιοχές ενδιαφέροντος, και αξιοθέατα και το υπόλοιπο.

Η αύξηση του συνόλου των σελιδοπροβολών οφείλεται στην αύξηση του μέσου όρου σελιδοπροβολών ανά επίσκεψη (Διάγραμμα 5.2 σελίδα 199) μιας και ο αριθμός των επισκέψεων δεν άλλαξε τόσο σημαντικά. Παρουσιάζεται μία μικρή μείωση στις παρουσιάσεις περιοχών, μία μικρή αύξηση στις υπόλοιπες σελίδες και μία μεγάλη αύξηση στις παρουσιάσεις επιχειρήσεων τουρισμού. Δηλαδή έχουμε μεγαλύτερη προβολή των επιχειρήσεων τουρισμού χωρίς αυτό να γίνεται σε βάρος της υπόλοιπης πληροφορίας που παρουσιάζεται στον ιστοχώρο.



Διάγραμμα 5.19 Απόλυτος αριθμός σελιδοπροβολών μία εβδομάδα πριν και μία μετά την αλλαγή του ιστοχώρου

Η αλλαγή αυτή θα μπορούσε να εξηγηθεί με αλλαγές στα ενδιαφέροντα των χρηστών που επισκέπτονται τον ιστοχώρο ή σημαντικές αλλαγές στις προελεύσεις από τις οποίες προέρχονται. Εάν για παράδειγμα υπήρχε ένας σύνδεσμος σε σημαντικό ιστοχώρο προς το Crete TOURnet ο οποίος αναφερόταν σε ξενοδοχεία, η αλλαγή θα μπορούσε να ερμηνευτεί διότι θα υπήρχαν πολλοί περισσότεροι επισκέπτες οι οποίοι ενδιαφέρονται για ξενοδοχεία. Αυτό όμως αποκλείεται διότι δεν υπήρξε σημαντική αύξηση του αριθμού επισκεπτών.

Δεν μπορεί επίσης να εξηγηθεί υποθέτοντας ότι με την αλλαγή του ιστοχώρου οι σελίδες των επιχειρήσεων τουρισμού έγιναν πιο “φιλικές” για τις μηχανές αναζήτησης, διότι αφενός δεν υπάρχει αύξηση επισκεπτών και αφετέρου (και ποιο σημαντικά) το χρονικό διάστημα στο οποίο παρατηρήθηκε η μεταβολή είναι πολύ μικρό για να ανταποκριθούν μηχανές αναζήτησης.

Το συμπέρασμα που συνάγεται είναι ότι η αναδιοργάνωση του ιστοχώρου έφερε πιο κοντά τις επιχειρήσεις στο χρήστη του ιστοχώρου.

### **5.3.3.2 Ανάλυση Χρήσης “Ηλεκτρονικού Φυλλαδίου”**

Κατά την πλοήγηση στον ιστοχώρο ο χρήστης βλέπει πολλές σελίδες για διάφορα θέματα, από τα οποία χρειάζεται στο τέλος να προβεί σε συγκεκριμένη επιλογή ανάλογα με τα ενδιαφέροντά του. Για παράδειγμα κατά την αναζήτηση κατάλληλης διαμονής, χρειάζεται να επιλέξει ανάμεσα σε ένα σύνολο καταλυμάτων που πληρούν τα κριτήρια του κάθε χρήστη. Έτσι κρίθηκε από νωρίς αναγκαίο να προσφερθεί λειτουργικότητα που να επιτρέπει στο χρήστη να επιλέγει σελίδες οι οποίες θα είναι προσβάσιμες οποιαδήποτε στιγμή κατά την πλοήγηση στον ιστοχώρο, χωρίς να χρειαστεί να προβεί εκ' νέου σε αναζήτηση (Εικόνα 11.12 όπως διαμορφώθηκε τελικά στη νέα έκδοση).



Η λειτουργικότητα φαινόταν στον ιστοχώρο με τον τίτλο “Selected” ή “Επιλεγμένα” στην παλαιά έκδοση και αρχικά και στη νέα. Στην παλαιά έκδοση η επιλογή για την εμφάνιση των επιλεγμένων σελίδων γινόταν με ειδικό κουμπί στο κάτω μέρος των φορμών αναζήτησης ή στην παρουσίαση επιχειρήσεων και αξιοθεάτων (Εικόνα 5.1). Στη νέα έκδοση αρχικά υπήρχε σχετική επιλογή στη δεξιά πλευρά του ιστοχώρου που είναι προσβάσιμη από όλες τις σελίδες. Όταν ο επισκέπτης επέλεγε κάποια σελίδα, το εικονίδιο αναβοέσβηνε (Εικόνα 5.2)

Αναζήτηση Διαμονής - Επίσημη, Crete, Kreta - Mozilla


Επι Επι Συν Συ Book marks Εξο Ψάξινα Έξο

Αναζήτηση Διαμονής - Επίσημη, Crete, Kreta Τελευταία επίσημη επίσημη - Crete TOULIST


Διεύθυνση: Λιμνοθάλασσα 70014 Χερσονήσος, Ηράκλειο  
Τηλέφωνο (+30): 28970 25575  
Fax (-30): 20370 20150

8. **Aldemar Bay View Village**  

Χαρακτήρες: Το ξενοδοχείο προσφέρει τις ιδανικές συνθήκες για διακοπές που εισάγουν τον επισκέπτη σε ένα νέο τρόπο ζωής και στη φιλοσοφία του σου ζην Πέλο ελληνικό, εκκεντωμένο, αρμόσιο, ομορφότερο και...  
Διεύθυνση: Λιμνοθάλασσα 70014 Χερσονήσος, Ηράκλειο  
Τηλέφωνο (+30): 28970 25705  
Fax (-30): 20370 21564

9. **Alexi's** 




Built in the most ideal location Alexi's Hotel offers a unique opportunity for the discerning tourist to enjoy magnificent surroundings in a quiet, comfortable...  
Διεύθυνση: Άνω Αρχαία Κελευθάρων 96, 73100 Νέα Κυδωνία, Νάνος  
Τηλέφωνο (+30): 28210 95703-4  
Fax (-30): 28210 90523

10. **Alfa Apartments** 

Πολύ μεγάλα διαμερίσματα που αποτελούν τον καλύτερο χώρο για να περάσετε τις διακοπές σας στην Κρήτη. Διαθέτουν μεγάλη βεράνδα και επιπλωμένα κήποι, που δίνουν την ιδανική ατμόσφαιρα για την απόλαυση.  
Διεύθυνση: Γερμανίας, 70200 Ηράκλειο, Νάνος  
Τηλέφωνο (+30): 20210 61112 - 20210 00301  
Fax (-30): 28210 01334

Σελίδες αποτελεσμάτων: 1 2 3 4 5 6 7 8 9 10 [Επόμενη >>]

---

Επιλογή για Αναζήτηση Επιστροφή

---

Ευρωπαϊκό Πρόγραμμα Crete Tourist, DIGSI, TEN-ISON, no 45549

Εικόνα 5.1Η λειτουργικότητα για "επιλεγμένες σελίδες" στην παλαιά έκδοση του ιστοχώρου

Η πηγή για πληροφορίες ταξιδιού για την Κρήτη



**Επίσημο Καζίνο**  
Επίσημος οδηγός για την Κρήτη που εστιάζει θέματα όπως Γεωγραφία, Μυθολογία, Ιστορία, Πολιτισμός, Γενικά και κληρονομιά του νησιού και πολλά άλλα. Πήρατε; κλικ εδώ

PORTO HENRI AND HOTEL EL LITHANIA OLD TOWN

**Εξυπηρέτηση & Διαμερίσματα**  
 Το καλύτερο ορατό  
 Περισσότερες πληροφορίες  
 κλικ εδώ

**Ενοίκιο οχημάτων**  
 Το καλύτερο ορατό  
 Περισσότερες πληροφορίες  
 κλικ εδώ

**Γλωσσάρι**  
 Λέξεις και φράσεις για τον ξένο ταξιδιώτη της Κρήτης και της Ελλάδας  
 κλικ εδώ

**Γρήγορο Ταξίδι**  
 Το καλύτερο ορατό  
 Περισσότερες πληροφορίες  
 κλικ εδώ

**Περιοχές & Αξιοθέατα**  
 Το καλύτερο ορατό  
 Περισσότερες πληροφορίες  
 κλικ εδώ

**Αεροπλάνο και Αγκυροβόληση**  
 Πληροφορίες για κινδύνους επιπέδου EFAS  
 κλικ εδώ

**Χάρτες Κρήτης**  
 Χάρτες της Κρήτης  
 κλικ εδώ

Εικόνα 5.2 Η επιλογή των επιλεγμένων σελίδων όπως ήταν αρχικά στη νέα έκδοση

Αργότερα (18 Μαΐου 2003) αποφασίστηκε η σχετική λειτουργία να προβληθεί περισσότερο και σχεδιάστηκε νέο εικονίδιο με νέο όνομα (Ηλεκτρονικό Φυλλάδιο).

Συνολικά υπάρχουν δηλαδή τρεις διαφορετικοί σχεδιασμοί στο επίπεδο διεπικοινωνίας όσον αφορά την πρόσβαση στην υπηρεσία του ηλεκτρονικού φυλλαδίου:

1. Παλαιά έκδοση ιστοχώρου (Εικόνα 5.1)
2. Νέα έκδοση ιστοχώρου με εικονίδιο στη δεξιά πλευρά (Εικόνα 5.2)
3. Νέα έκδοση ιστοχώρου με πιο προβλημένο εικονίδιο στη δεξιά πλευρά (Εικόνα 5.3)



Εικόνα 5.3 Η νέα ονομασία (e-φυλλάδιο) και η νέα τοποθεσία του σχετικού εικονιδίου

Η αίσθηση που υπήρχε ήταν ότι στην παλαιά έκδοση του ιστοχώρου η υπηρεσία δε θα χρησιμοποιόταν πολύ επειδή δεν ήταν εμφανής η σημασία του εικονιδίου. Αναμενόταν επίσης ότι η τρίτη έκδοση του σχεδιασμού θα είχε πολύ περισσότερες επιλογές από τη δεύτερη, ενώ ότι η δεύτερη θα είχε επίσης μεγαλύτερη χρήση από την πρώτη έκδοση επειδή το εικονίδιο ήταν πιο εύκολα προσβάσιμο και με μεγαλύτερη συνέπεια τοποθετημένο. Η πραγματική χρήση του ιστοχώρου διέψευσε όμως μέρος αυτών των προσδοκιών.

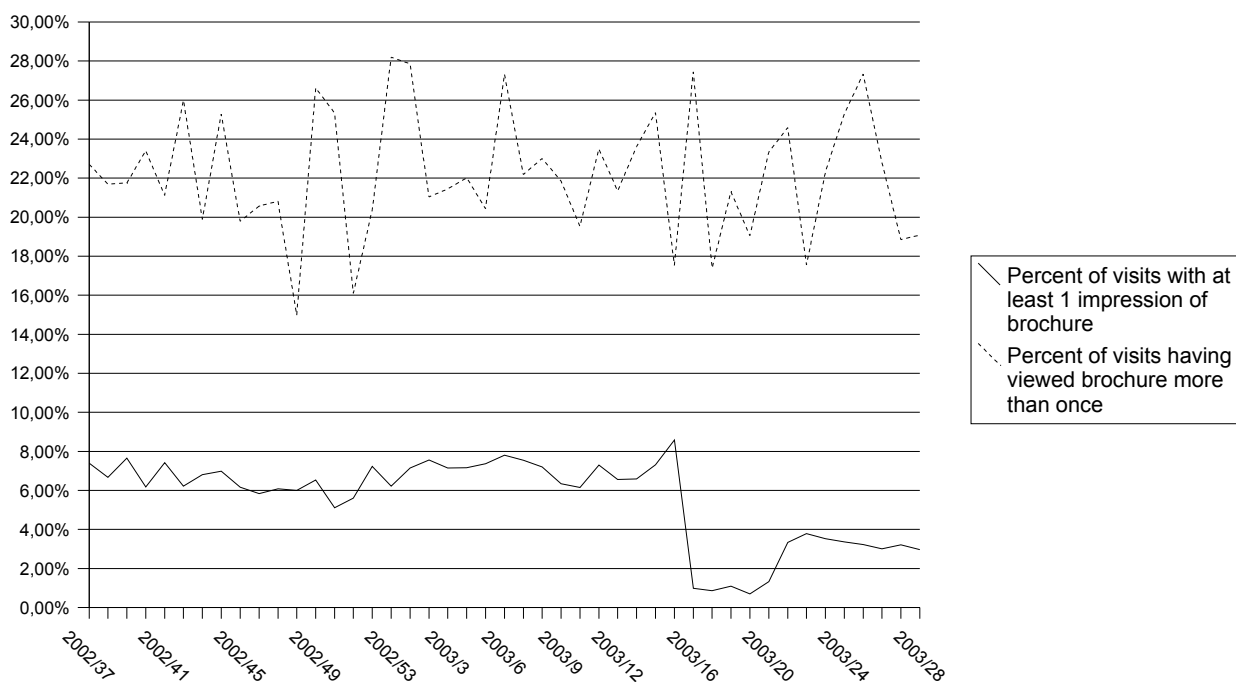
Στο Διάγραμμα 5.20 αποτυπώνεται η χρήση της υπηρεσίας στις διάφορες φάσεις σχεδιασμού των συνδέσμων πρόσβασης σε αυτήν. Στο κάτω μέρος η καμπύλη αναφέρεται στο ποσοστό από το σύνολο των επισκέψεων ανά εβδομάδα στις οποίες υπήρχε τουλάχιστον μία πρόσβαση στη σελίδα με την υπηρεσία.

Στο επάνω μέρος φαίνεται το ποσοστό, από τις επισκέψεις με τουλάχιστον μία πρόσβαση, που επισκέφθηκαν περισσότερες της μίας φορές την υπηρεσία.

Η κάτω καμπύλη μπορεί να θεωρηθεί ως βαθμός αναγνωρισιμότητας της υπηρεσίας, δηλαδή πόσο εμφανής είναι η πρόσβαση σε αυτήν. Η πάνω καμπύλη μπορεί να θεωρηθεί η χρησιμότητα που

φαίνεται να έχει η υπηρεσία για τους επισκέπτες.

Με αυτή την έννοια, η χρησιμότητα της υπηρεσίας παραμένει σε σταθερά επίπεδα και κυμαίνεται ανάμεσα σε 20% και 25%, ανεξάρτητα από τον τρόπο με τον οποίο έφτασαν στην υπηρεσία. Η αναγνωρισιμότητα όμως διαφέρει εντελώς στους τρεις διαφορετικούς σχεδιασμούς. Ενώ στην πρώτη περίπτωση υπήρχε ένα ποσοστό περίπου 6% των επισκεπτών που επέλεξαν την υπηρεσία, στη δεύτερη περίπτωση το ποσοστό έπεσε δραματικά. Στην τρίτη περίπτωση βελτιώθηκε όντως η αναγνωρισιμότητα, αλλά όχι στο βαθμό για να φτάσει τον αρχικό σχεδιασμό.



Διάγραμμα 5.20 Χρήση της υπηρεσίας "Ηλεκτρονικό Φυλλάδιο". Στο γράφημα στο κάτω τμήμα (γαλάζιο) φαίνεται το ποσοστό των επισκέψεων στις οποίες υπήρχε τουλάχιστον μία επίσκεψη στη σελίδα του ηλεκτρονικού φυλλαδίου. Από τις τελευταίες επισκέψεις, στο πάνω μέρος του γραφήματος (ερυθρό) φαίνεται τι ποσοστό επαναεπισκέφτηκε τη σελίδα, είχε δηλαδή πάνω από μία επίσκεψη. Στο κάτω μέρος διακρίνονται οι τρεις διαφορετικές φάσεις στο σχεδιασμό των συνδέσμων πρόσβασης στην υπηρεσία από το διαφορετικό εύρος που κυμαίνεται η καμπύλη.

### 5.3.3.3 Επηρεασμός από Τοποθεσία "Σχετικών Επιχειρήσεων"

Στο Crete TOURnet παρουσιάζεται ένα σύνολο περιοχών ενδιαφέροντος με πληροφορίες σε μορφή κειμένου και φωτογραφίες. Για κάθε περιοχή ενδιαφέροντος παρουσιάζονται επίσης επιχειρήσεις τουρισμού που βρίσκονται στην περιοχή. Η παρουσίαση γίνεται με σχετικούς υπερσυνδέσμους στο κάτω μέρος της κάθε σελίδας, όπου αναφέρονται ανά κατηγορία επιχείρησης οι 5 πρώτες με απ' ευθείας σύνδεσμο, ενώ εάν υπάρχουν περισσότερες από 5 επιχειρήσεις υπάρχει επίσης

υπερσύνδεσμος που οδηγεί σε σελίδα με όλες τις σχετικές επιχειρήσεις. Σχετική απεικόνιση υπάρχει στην Εικόνα 5.4.



Εικόνα 5.4Η παρουσίαση σχετικής πληροφορίας για επιχειρήσεις τουρισμού που βρίσκεται στο κάτω μέρος της κάθε παρουσίασης περιοχής ενδιαφέροντος

Στα πλαίσια της εργασίας εξετάστηκε σε ποιο βαθμό επηρεάζει την επισκεψιμότητα στις σελίδες των σχετικών επιχειρήσεων η τοποθεσία αυτών των συνδέσμων.

Για τη μελέτη της σχετικής συμπεριφοράς επιλέχθηκε η ακόλουθη τακτική:

- Για ορισμένες επισκέψεις τοποθετήθηκε σύνδεσμος για τις επιχειρήσεις διαμονής και ενοικίασης οχημάτων σε πιο προβλεπόμενη θέση της σελίδας, πριν το κείμενο περιγραφής της περιοχής (Εικόνα 5.5)
- Οι σύνδεσμοι εμφανίζονταν για το 50% των επισκέψεων για χρονικό διάστημα μίας εβδομάδας
- Οι σύνδεσμοι εμφανίζονταν για συγκεκριμένες δημοφιλείς περιοχές ενδιαφέροντος, που ήταν τα



Μάλια, Ελούντα, Ρέθυμνο, Χανιά και Γεωργιούπολη.

Στην Εικόνα 5.5 απεικονίζεται η σχετικά τροποποιημένη σελίδα.



Εικόνα 5.5 Ειδικό επιπλέον σύνδεσμοι για να μελετηθεί η αντίδραση των επισκεπτών σε διαφορετικές τοποθεσίες των συνδέσμων

Μετρήθηκε το ποσοστό των επισκεπτών που επέλεξε τους νέους συνδέσμους σε σχέση με το ποσοστό των επισκεπτών που επέλεξε τους κανονικούς συνδέσμους. Η περίπτωση της περιοχής Μάλια έχει ιδιαίτερο ενδιαφέρον διότι πρόκειται για δημοφιλή στους ξένους ταξιδιώτες περιοχή διακοπών, αλλά υπάρχει επίσης ο αρχαιολογικός χώρος Μάλια. Στον ιστοχώρο όμως υπάρχει πληροφορία μόνο για τον αρχαιολογικό χώρο αλλά παρόλα αυτά υπάρχουν πολλές επισκέψεις με σελίδα εισόδου αυτή την περιοχή. Υπάρχει η υποψία ότι πρόκειται για ταξιδιώτες που δεν ενδιαφέρονται κυρίως για τον αρχαιολογικό χώρο οπότε θα έχουν ιδιαίτερο ενδιαφέρον για επιχειρήσεις τουρισμού της περιοχής.

Το πείραμα έγινε από 23 έως 29 Ιουνίου 2003. Οι αριθμοί σε επισκέψεις και σελιδοπροβολές ανά

κατηγορία επίσκεψης φαίνονται στον πίνακα 5.10 (πάντα για τις επισκέψεις και τις προβολές για τις υπό εξέταση περιοχές).

	<i>Κανονικοί σύνδεσμοι</i>		<i>Σύνδεσμοι στο επάνω μέρος</i>	
<b>Αριθμός επισκέψεων</b>	695		660	
<b>Αριθμός προβολών περιοχών</b>	1416		1239	
<b>Προβολές ανά περιοχή</b>				
Μάλια	208		254	
Ελούντα	390		252	
Ρέθυμνο	247		229	
Χανιά	378		306	
Γεωργιούπολη	196		198	
<b>Επιλογή σχετικών επιχειρήσεων</b>				
Μάλια	23	11,06%	37	14,57%
Ελούντα	51	13,08%	30	11,90%
Ρέθυμνο	28	11,34%	30	13,10%
Χανιά	49	12,96%	63	20,59%
Γεωργιούπολη	16	8,16%	30	15,15%

*Πίνακας 5.10 Επισκέψεις και σελιδοπροβολές για τις περιοχές ενδιαφέροντος και σελίδες σχετικής πληροφορίας για τις επισκέψεις με και χωρίς ειδικούς υπερσυνδέσμους*

Παρατηρούμε ότι για την περίπτωση όπου υπήρχαν οι σύνδεσμοι για τις σχετικές επιχειρήσεις τουρισμού σε ειδικό προβεβλημένο σημείο, έχουμε μία αύξηση του ποσοστού των σελιδοπροβολών για σχετικές επιχειρήσεις, με εξαίρεση της περιοχής Ελούντα. Παρόλο που υπήρχε η υποψία ότι για την περιοχή Μάλια θα υπήρχε πολύ μεγαλύτερο ποσοστό, αυτό δεν επιβεβαιώνεται.

### **5.3.4 Σενάρια Χρήσης**

Στο κεφάλαιο 4 περιγράφηκε η μορφή που μπορεί να πάρουν ερωτήσεις σε SQL92 για να περιγραφούν σενάρια χρήσης που περιγράφονται από συνθήκες στο μονοπάτι μιας επίσκεψης και τα χαρακτηριστικά της κάθε σελιδοπροβολής και της επίσκεψης.

Με βάση αυτή τη μεθοδολογία έγινε μελέτη του σεναρίου χρήσης “Αναζήτηση Διαμονής” που περιγράφεται στο Σχήμα 5.1. Ενδιαφέρει κατά πόσο η εφαρμογή ιστοχώρου καλύπτει τους επισκέπτες που ενδιαφέρονται να αναζητήσουν διαμονή. Μελετήθηκε τόσο το στατικό σενάριο της ολοκλήρωσης του αναζήτησης διαμονής, όσο και η περίπτωση των επισκεπτών που διέκοψαν πρόωρα το σενάριο.

Θα παρουσιαστούν οι ερωτήσεις που υποβλήθηκαν σε κάθε βήμα χρησιμοποιώντας τη

σημειογραφία ψευδοSQL που παρουσιάστηκε στην ενότητα 4.4.4 σελίδα 159.

Στον πίνακα 5.11 απαριθμούνται μερικά γενικά χαρακτηριστικά των επισκέψεων για την χρονική περίοδο που έγινε η μελέτη τα οποία χρησιμοποιούνται στη συνέχεια.

<i>Περιγραφή</i>	
Χρονική Περίοδος	1 Ιουλίου 2003 έως και 7 Ιουλίου 2003
Αριθμός Επισκέψεων	11685
support ακολουθίας {"αρχική σελίδα","αναζήτηση διαμονής"} ανεξαρτήτως επίσκεψης	1417
support ακολουθίας {"αρχική σελίδα","αναζήτηση διαμονής"} τουλάχιστον μία φορά σε επίσκεψη	1158

*Πίνακας 5.11 Γενικά χαρακτηριστικά για τις επισκέψεις πάνω στις οποίες μελετήθηκε το σενάριο χρήσης "Αναζήτηση Διαμονής"*

Μελετήθηκε η περίπτωση με κριτήρια έναρξης σεναρίου την ακολουθία {"αρχική σελίδα","φόρμα αναζήτησης διαμονής"}. Θεωρήθηκε δηλαδή ότι επισκέπτες που επιλέγουν με αυτή τη σειρά σελίδες του ιστοχώρου (σε οποιοδήποτε σημείο της επίσκεψής τους) ενδιαφέρονται για καταλύματα για διαμονή. Η ακολουθία προσδιορίστηκε με βάση τη σημειογραφία (notation) της ενότητας 5.1 ως ακολούθως

```
select count(distinct session_id + a.sequence_order)
from a b
where a='index' and b='hotels form'
```

*Ερώτηση 5.1 ψευδοSQL κριτήριο έναρξης σεναρίου αναζήτηση διαμονής*

Αυτή η ακολουθία συναντάται 1417 φορές. σε 1158 επισκέψεις.

Το ιδανικό μονοπάτι για την εύρεση διαμονής είναι το

{"αρχική σελίδα", "φόρμα αναζήτησης διαμονής",["επιλογή περιοχής"],"αποτελέσματα αναζήτησης",["αποτελέσματα αναζήτησης"], "παρουσίασης επιχείρησης διαμονής"}

όπου ανάμεσα στα σύμβολα [] σημειώνονται προαιρετικές σελιδοπροβολές (επιλογή περιοχής και δεύτερη σελίδα αποτελεσμάτων για την εμφάνιση όλων των αποτελεσμάτων). Από αυτή την ακολουθία προκύπτει ότι το βέλτιστο μονοπάτι έχει μήκος από 4 έως 6 σελιδοπροβολές. Θεωρούμε ως ολοκλήρωση του σεναρίου την πρώτη εμφάνιση επιχείρησης διαμονής.

Η επόμενη ερώτηση υπολογίζει τις επισκέψεις και το μήκος του μονοπατιού όπου με αφετηρία το κριτήριο έναρξης του σεναρίου έχουμε επιτυχής ολοκλήρωση χωρίς την παρεμβολή τρίτων σελίδων (δηλαδή ακολουθούνται μόνο σελίδες μέσα από το βέλτιστο μονοπάτι). Η ερώτηση υπολογίζει το μήκος του μονοπατιού μόνο μέχρι την πρώτη σελιδοπροβολή επιχείρισης διαμονής.

```
select session_id, min(d.sequence_order) - a.sequence_order as pathlength
from a b c+ d
where
    a.page_name='index'
    and b.page_name='hotels form'
    and c in ('searchresults', 'select area', 'hotels form')
    and d.page_name='sme'
    and d.action.entity_id in (valuelist)
    and d.action.action_id='present sme'
group by session_id, a.sequence_order
```

*Ερώτηση 5.2 ψευδοSQL*

Από τα αποτελέσματα της ερώτησης προκύπτει η κατανομή στο μήκος του μονοπατιού που φαίνεται στον πίνακα 5.12. Όπως φαίνεται, σε συνολικά άνω του 88% των περιπτώσεων (4,5 και 6 σελιδοπροβολές) το μήκος του μονοπατιού δεν υπερβαίνει το βέλτιστο δυνατό μήκος για τα δοσμένα κριτήρια έναρξης.

<i><b>Μήκος Μονοπατιού</b></i>	<i><b>Ποσοστό</b></i>
4	47,76%
5	23,88%
6	16,42%
8	2,98%
9	4,48%
11, 13, 23	1,49%

*Πίνακας 5.12 Μήκος μονοπατιού εντός σελίδων του βέλτιστου μονοπατιού*

Το σημείο όμως που είναι ιδιαίτερα προβληματικό είναι ότι η εκτέλεση της ερώτησης δίνει μόνο 67 περιπτώσεις (4,72%) όπου όντως με τα δοσμένα κριτήρια υπάρχει επιτυχής ολοκλήρωση! Το ερώτημα που αμέσως προκύπτει είναι για ποιο λόγο, ενώ οι επισκέπτες ξεκινάνε την αναζήτηση διαμονής, δεν ολοκληρώνουν τελικά άμεσα το σχετικό σενάριο.

Πιθανές εξηγήσεις μπορεί να είναι οι ακόλουθες:

1. η υπόθεση για τα κριτήρια έναρξης σεναρίου (ότι δηλαδή ο επισκέπτης που βλέπει την αρχική σελίδα και μετά τη φόρμα αναζήτησης διαμονής ψάχνει για διαμονή) είναι λανθασμένη

2. ο ιστοχώρος δεν παρέχει την πληροφορία σύμφωνα με τις επιθυμίες (κριτήρια αναζήτησης) του επισκέπτη
3. ο ιστοχώρος παρουσιάζει πρόβλημα χρησιμοποίησιμότητας

Το σημείο 2 μπορεί να αναχθεί είτε σε έλλειψη δεδομένων του ιστοχώρου (δεν υπάρχουν καταχωρήσεις καταλυμάτων που καλύπτουν τα κριτήρια του επισκέπτη) είτε σε πρόβλημα χρησιμοποίησιμότητας.

Για να εξακριβωθεί το σημείο 1, εκτελείται η ακόλουθη ερώτηση:

```
select session_id
from a b c+ d e* f
where
    a.page_name='index'
    and b.page_name='hotels form'
    and c in ('searchresults', 'select area', 'hotels form')
    and d.page_name<>'sme'
    and f.page_name='sme'
    and f.action.entity_id in (valuelist)
    and f.action.action_id='present sme'
group by session_id, a.sequence_order
```

*Ερώτηση 5.3 ψευδοSQL*

Η ερώτηση θα έχει ως αποτέλεσμα τις επισκέψεις όπου υπάρχει μονοπάτι με το κριτήριο έναρξης για αναζήτηση διαμονής, δε θα εμφανιστεί διαμονή με ενδιάμεσες σελίδες τις σελίδες του βέλτιστου μονοπατιού (c+ d), αλλά θα εμφανιστούν σελίδες διαμονής (f) αργότερα (e\*) στην επίσκεψη. Η εκτέλεση της ερώτησης δίνει ως αποτέλεσμα ότι υπάρχουν 572 τέτοιες περιπτώσεις το οποίο αντιστοιχεί στο περίπου 50% των περιπτώσεων. Δηλαδή ενώ οι επισκέπτες δεν βρίσκουν με τη βέλτιστη διαδικασία επιχειρήσεις διαμονής, το πράττουν αυτό αργότερα με διαφορετικό τρόπο. Αυτό υποδεικνύει ότι το ενδιαφέρον για διαμονή είναι όντως υπαρκτό για τα δοσμένα κριτήρια έναρξης του σεναρίου.

Στην επόμενη ερώτηση επιχειρείται να μελετηθεί η περίπτωση όπου ξεκινάει μία ανεπιτυχής προσπάθεια εύρεσης διαμονής και δεν υπάρχει αργότερα στην επίσκεψη προβολή σελίδας διαμονής. Καλύπτει δηλαδή τις περιπτώσεις εκτός του προηγούμενου 50% (όπου τελικά γίνεται προβολή διαμονής). Συγκεκριμένα ενδιαφέρει να εξακριβωθεί εάν κατά τη διάρκεια αναζήτησης διαμονής μέσα από μονοπάτι που περιλαμβάνει τις βέλτιστες σελίδες, υπάρχει εγκατάλειψη του ιστοχώρου. Αυτή η ερώτηση μπορεί να σχηματιστεί με τον παρακάτω τρόπο.

Η ερώτηση δίνει όλες τις επισκέψεις όπου ξεκινάει αναζήτηση διαμονής με τα κριτήρια έναρξης, ακολουθείται το μονοπάτι αναζήτησης χωρίς προβολή διαμονής και γίνεται εγκατάλειψη του

```

select session_id
from a b c+ d
where
    a.page_name='index'
    and b.page_name='hotels form'
    and c.page_name in ('searchresults', 'select area', 'hotels form')
    and d.page_name in ('searchresults', 'select area', 'hotels form')
    and d.impression_id = session.exit_pageimpression
group by session_id, a.sequence_order

```

*Ερώτηση 5.4 ψευδοSQL*

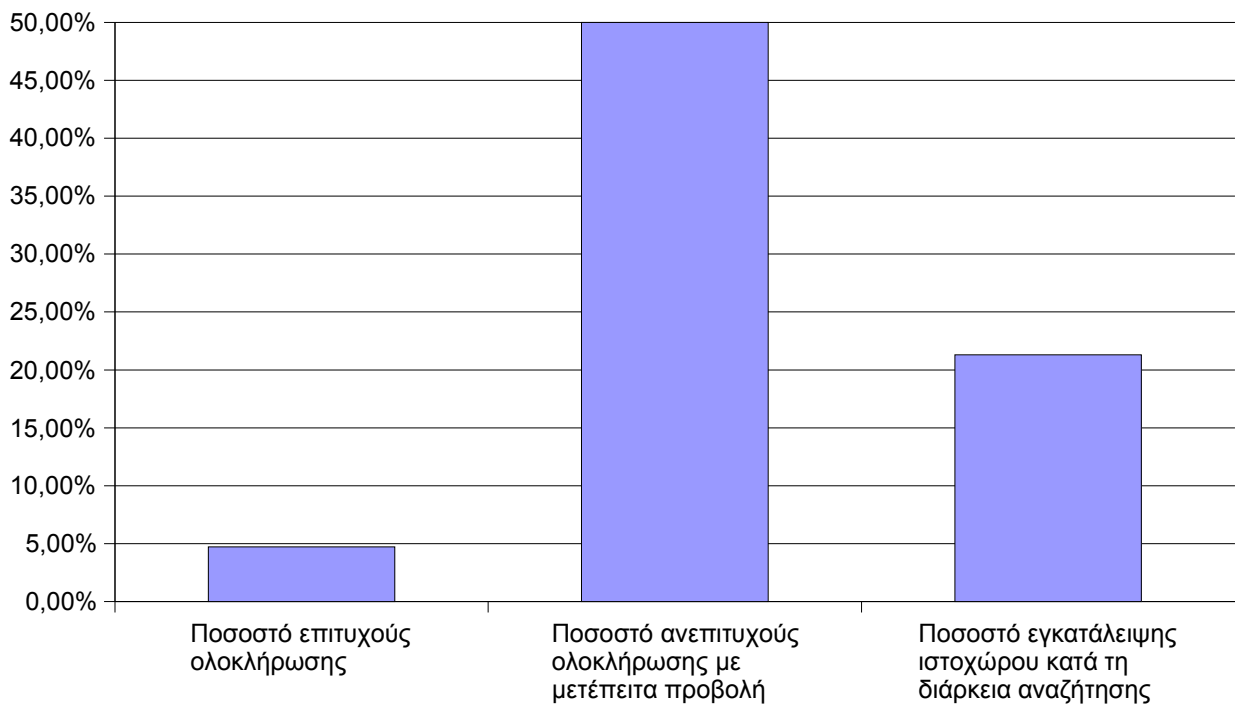
ιστοχώρου. Το αποτέλεσμα δίνει 270 τέτοιες περιπτώσεις (21,3%). Οπότε ένα σημαντικό μέρος των επισκεπτών που δεν καταφέρνουν να βρουν διαμονή από τη φόρμα αναζήτησης, εγκαταλείπουν τον ιστοχώρο.

Από την έως τώρα ανάλυση μπορεί να γίνει η κατηγοριοποίηση που φαίνεται στον πίνακα 5.13 και στο διάγραμμα 5.21.

<i>A/A</i>	<i>Περιγραφή</i>	<i>Ποσοστά</i>
<b>1</b>	Ποσοστό επιτυχούς ολοκλήρωσης	4,72%
<b>2</b>	Ποσοστό ανεπιτυχούς ολοκλήρωσης με μετέπειτα προβολή σελίδας διαμονής	50,00%
<b>3</b>	Ποσοστό εγκατάλειψης ιστοχώρου κατά τη διάρκεια αναζήτησης διαμονής	21,30%

*Πίνακας 5.13 Γενικές κατηγοριοποιήσεις στο σενάριο "αναζήτηση Διαμονής" με κριτήριο έναρξης την ακολουθία {"αρχική σελίδα", "φόρμα αναζήτησης διαμονής"}*

Όπως φαίνεται, το ποσοστό των επισκεπτών που τελικά χρησιμοποιούν επιτυχώς τη φόρμα αναζήτησης διαμονής είναι πολύ μικρό (περίπτωση 1 του πίνακα 5.13). Το ενδιαφέρον επικεντρώνεται στις άλλες δύο περιπτώσεις, και συγκεκριμένα:



Διάγραμμα 5.21 Γενικές κατηγοριοποιήσεις στο σενάριο "αναζήτηση Διαμονής" με κριτήριο έναρξης την ακολουθία {"αρχική σελίδα", "φόρμα αναζήτησης διαμονής"}

- για την περίπτωση 2 όπου τελικά, στη διάρκεια της επίσκεψης, γίνεται προβολή σελίδας διαμονής. Ενδιαφέρει με ποιον τρόπο έφτασαν στην προβολή διαμονής και επίσης γιατί εγκατέλειψαν την προσπάθεια από τη φόρμα αναζήτησης
- για την περίπτωση 3 όπου γίνεται εγκατάλειψη του ιστοχώρου ενδιαφέρει ο λόγος για τον οποίο εγκατέλειψαν τον ιστοχώρο

Η εκτέλεση της επόμενης ερώτησης εξάγει τις σελίδες που προηγούνται της εμφάνισης διαμονής (f.page\_name) για την περίπτωση 2 του πίνακα 5.13.

```
select f.page_name
from a b c+ d e* f g
where
    a.page_name='index'
    and b.page_name='hotels form'
    and c in ('searchresults', 'select area', 'hotels form')
    and d.page_name<>'sme'
    and g.page_name='sme'
    and g.action.entity_id in (valuelist)
    and g.action.action_id='present sme'
group by session_id, a.sequence_order
```

Ερώτηση 5.5 ψευδοSQL

Η εκτέλεση της ερώτησης έχει ως αποτέλεσμα την κατανομή του πίνακα 5.14

<i>Προηγούμενη Σελίδα</i>	<i>Κατανομή</i>
Προβολή Περιοχής Ενδιαφέροντος	8,60%
Αποτελέσματα Αναζήτησης	78,30%

*Πίνακας 5.14 Σελίδες που προηγούνται της εμφάνισης διαμονής, όταν υπάρχει αρχικά αποτυχία του σεναρίου με ακολουθία έναρξης {“αρχική σελίδα”, “φόρμα αναζήτησης διαμονής”}*

Από αυτά τα αποτελέσματα προκύπτει ότι ένα ποσοστό της τάξεως του 8,6% φθάνει τελικά στην προβολή διαμονής μέσω της περιοχής ενδιαφέροντος στην οποία ανήκει. Το μεγαλύτερο ποσοστό όμως (78,3%) προέρχεται πάλι από αποτελέσματα αναζήτησης, τα οποία προφανώς δεν αναφέρονται σε αναζήτηση από τη φόρμα αναζήτησης διαμονής.

Επεκτείνοντας την τελευταία ερώτηση (περιορισμός στο f να πρόκειται για αποτελέσματα αναζήτησης και προβολή της ενέργειας “Search Type”, βλέπε και πίνακα 3.3), προκύπτει για το είδος της αναζήτησης που έγινε, ότι πρόκειται στο μεγαλύτερο μέρος για αναζητήσεις μέσω του ενεργού χάρτη από την αρχική σελίδα. Συγκεκριμένα προκύπτουν τα ποσοστά του πίνακα 5.15

<i>Είδος Αναζήτησης</i>	<i>Κατανομή</i>
Αναζήτηση μέσω ενεργού χάρτη αρχικής σελίδας	88,90%
Σχετικές με περιοχή επιχειρήσεις	5,30%
Αναζήτηση από όνομα διαμονής από αρχική σελίδα	3,80%
Άλλο (events, food netertainment etc.)	2,00%

*Πίνακας 5.15 Είδος αναζήτησης όταν η προηγούμενη σελιδοπροβολή ήταν αποτελέσματα αναζήτησης*

Οπότε η μεγάλη πλειονότητα των χρηστών που δε βρίσκουν διαμονή μέσω της φόρμας αναζήτησης διαμονής, το πράττουν αυτό μέσω του ενεργού χάρτη που βρίσκεται στην αρχική σελίδα.

Συνεχίζοντας με την περίπτωση περίπτωση 2 του πίνακα 5.13, παραμένει ακόμα κενό στους λόγους που οδήγησαν στην εγκατάλειψη του σεναρίου χρήσης. Για αυτό το σκοπό εκτελείται η ακόλουθη ερώτηση

```
select c.page_name
from a b c+ d e* f
where
    a.page_name='index'
    and b.page_name='hotels form'
    and c.page_name in ('searchresults', 'select area', 'hotels form')
    and d.page_name <> 'sme'
    and f.page_name='sme'
    and f.action.entity_id in (valuelist)
    and f.action.action_id='present sme'
group by session_id, a.sequence_order
```

*Ερώτηση 5.6 ψευδοSQL*



Η ερώτηση επικεντρώνεται στις σελίδες c οι οποίες είναι οι σελίδες που βρίσκονται μέσα στο βέλτιστο μονοπάτι (αποτελέσματα, επιλογή περιοχής, φόρμα αναζήτησης) αλλά τελικά δεν οδηγούν σε προβολή διαμονής.

Από την εκτέλεση της ερώτησης προκύπτει μία ισοκατανομή ανάμεσα στη φόρμα αναζήτησης, τα αποτελέσματα αναζήτησης και τη σελίδα επιλογής περιοχής, με μία υπεροχή ελαφρώς για τη σελίδα των αποτελεσμάτων αναζήτησης. Δηλαδή ουσιαστικά οι επισκέπτες χρησιμοποιούν τη φόρμα αναζήτησης, τη σελίδα για επιλογή επιλογής και βλέπουν αποτελέσματα πολλές φορές, αλλά ποτέ δε φτάνουν σε σημείο να δουν κάποια συγκεκριμένη διαμονή.

Εξετάζοντας τις σελίδες των αποτελεσμάτων αναζήτησης προκύπτει ότι στα αποτελέσματα αναζήτησης υπάρχει πολύ μεγάλος αριθμός επιχειρήσεων. Συγκεκριμένα σε πάνω από 80% των περιπτώσεων τα αποτελέσματα αναζήτησης υπερβαίνουν τα 300. Συνδυάζοντας αυτό το γεγονός με την προσπάθεια που κάνουν οι επισκέπτες να χρησιμοποιήσουν τη σελίδα για επιλογή περιοχής, μπορεί να βγει το συμπέρασμα ότι οι χρήστες προσπαθούν να περιορίσουν τις ζητούμενες επιχειρήσεις σε συγκεκριμένη περιοχή, αλλά παρόλα αυτά δεν τα καταφέρνουν και τα αποτελέσματα είναι τόσα πολλά που δεν τους επιτρέπεται να εντοπίσουν την επιχείρηση που αναζητούν. Αυτό συνδυάζεται με τα αποτελέσματα στον πίνακα 5.15, όπου τελικά αυτοί οι χρήστες εγκαταλείπουν την προσπάθεια και τελικά βρίσκουν καταλύματα μέσω του ενεργού χάρτη της αρχικής σελίδας, δηλαδή με περιορισμό στη γεωγραφική περιοχή.

Εξετάζοντας με όμοιο τρόπο την περίπτωση 2 του πίνακα 5.13 (εγκατάλειψη ιστοχώρου), προκύπτει ότι στα αποτελέσματα αναζήτησης αυτής της κατηγορίας είτε εμφανίζονται πολλά αποτελέσματα, είτε δεν εμφανίζεται κανένα.

#### **5.3.4.1 Συμπεράσματα για την Ανάλυση Σεναρίου Χρήσης**

Από την ανάλυση της αμέσως προηγούμενης ενότητας για το σενάριο χρήσης “αναζήτηση διαμονής” με συνθήκη έναρξης την ακολουθία {“αρχική σελίδα”, “φόρμα αναζήτησης διαμονής”}, μπορεί να εξαχθούν τα ακόλουθα συμπεράσματα όσον αφορά τη χρησιμοποιησιμότητα του ιστοχώρου τουριστικού προορισμού:

- σε μικρό μόνο ποσοστό (4,97%) η φόρμα αναζήτησης διαμονής οδηγεί στην άμεση εύρεση διαμονής
- σε ποσοστό 21,3% οι επισκέπτες εγκαταλείπουν τον ιστοχώρο ενώ βρίσκονται στα ενδιάμεσα στάδια της αναζήτησης (αποτελέσματα, φόρμα, επιλογή περιοχής) με χαρακτηριστικό ότι στα αποτελέσματα είτε εμφανίζονται πάρα πολλά καταλύματα, είτε κανένα. **Κρίνεται έτσι ότι η**

**φόρμα αναζήτησης και και η σελίδα επιλογής περιοχής ανεπαρκείς στο ακριβή περιορισμό των αποτελεσμάτων σύμφωνα με τις επιλογές των επισκεπτών.**

- σε ποσοστό περίπου 50% οι επισκέπτες εγκαταλείπουν την αναζήτηση με βάση τη φόρμα αναζήτησης αλλά επιχειρούν με άλλους τρόπους που παρέχονται στον ιστοχώρο να βρουν καταλύματα. Κατά την εγκατάλειψη είναι χαρακτηριστικό ότι τα αποτελέσματα αναζήτησης είναι στη μεγάλη πλειονότητα πάρα πολλά, παρόλο που επιχειρείται με τη σελίδα επιλογής περιοχής να περιοριστεί η γεωγραφική περιοχή αναζήτησης. Οι επισκέπτες αυτής της κατηγορίας φαίνεται να καλύπτονται από αναζήτηση μέσω ενός ενεργού χάρτη που παρέχεται από την αρχική σελίδα, ο οποίος δίνει άμεσα αποτελέσματα από περιορισμένη γεωγραφική περιοχή. **Κρίνεται έτσι ότι η σελίδα επιλογής περιοχής είναι ανεπαρκής.**

Η επέμβαση τόσο στη φόρμα αναζήτησης όσο και στη σελίδα επιλογής περιοχής ώστε να επιτρέψει με μεγαλύτερη ακρίβεια τον περιορισμό των αποτελεσμάτων κρίνεται απολύτως αναγκαία.

### **5.3.5 Συμπεράσματα από τη Μελέτη με τη Μεθοδολογία Καταγραφής και Ανάλυσης Δεδομένων Χρήσης Ιστοχώρων**

Στην παρούσα ενότητα μελετήθηκε ο ιστοχώρος τουριστικού προορισμού με βάση τη μεθοδολογία ανάλυσης που προτείνεται στο Κεφάλαιο 4 και χρησιμοποιώντας τα δεδομένα, που καταγράφηκαν για τη συμπεριφορά των επισκεπτών, που συλλέχθηκαν από το σύστημα καταγραφής που παρουσιάζεται στο Κεφάλαιο 3.

Αν και οι ερωτήσεις που υποβλήθηκαν στο σύστημα καλύπτουν ένα πολύ μικρό μέρος των πιθανών ερωτήσεων που προτείνονται στη μεθοδολογία, μπορεί ήδη από αυτές να εξαχθούν χρήσιμα συμπεράσματα για το συγκεκριμένο ιστοχώρο ιστοχώρο και γενικότερα για τη μεθοδολογία.

Τα συμπεράσματα από τη μελέτη μπορεί να χωριστούν σε τρεις κατηγορίες:

- συμπεράσματα για τη βελτίωση χρησιμοποίησιμότητας και προσαρμογή του ιστοχώρου
- συμπεράσματα για την αγορά τουρισμού της γεωγραφικής περιοχής που καλύπτει ο ιστοχώρος
- συμπεράσματα για άλλους ιστοχώρους τουρισμού

Όσον αφορά καθ' αυτό τον ιστοχώρο και τα **προβλήματα χρησιμοποίησιμότητας** ιδιαίτερο ενδιαφέρον παρουσιάζει η δυνατότητα σύγκρισης εναλλακτικών μορφών αλληλεπίδρασης με το χρήστη, όπου μπορεί να αξιολογηθούν οι διάφορες ιδέες για το interface του ιστοχώρου και να επιλεγεί τελικά αυτό που φαίνεται να ανταποκρίνεται με καλύτερο τρόπο στους πραγματικούς

επισκέπτες. Τα ενδεικτικά πειράματα που έγιναν για την επισκεψιμότητα του “ηλεκτρονικού φυλλαδίου” και την επιλογή σχετικών επιχειρήσεων από τις περιοχές ενδιαφέροντος, ανάλογα με την τοποθεσία και μορφή των σχετικών συνδέσμων δείχνουν ότι μπορεί με τη μεθοδολογία και το σύστημα που αναπτύχθηκε να εντοπιστούν συγκεκριμένοι τρόποι βελτίωσης της χρησιμοποιησιμότητας ή να ελεγχθούν υποθέσεις για το επίπεδο διεπικοινωνίας του ιστοχώρου.

Το παράδειγμα με τις απότομες αλλαγές σε επισκέψεις από συγκεκριμένες προελεύσεις φανερώνει τους διαφορετικούς σκοπούς που μπορεί να αποκτήσει (έστω για περιορισμένο χρονικό διάστημα) ο ιστοχώρος. Η εκμετάλλευση τέτοιων αλλαγών είναι ιδιαίτερα εύκολη ακολουθώντας τη μεθοδολογία που προτάθηκε για την υποβολή ερωτήσεων για εντοπισμό ασυνήθιστων αλλαγών.

Επίσης ιδιαίτερο ενδιαφέρον παρουσιάζει η κατηγοριοποίηση των επισκεπτών με βάση τη γλώσσα που χρησιμοποιούν στον ιστοχώρο επειδή προκύπτουν σαφείς αλλαγές στις προτιμήσεις τους για περιοχές ενδιαφέροντος της γεωγραφικής περιοχής που καλύπτει ο ιστοχώρος. Και σε αυτή την περίπτωση προτάθηκαν συγκεκριμένοι τρόποι δυναμικής προσαρμογής του ιστοχώρου ανάλογα με την επιλογή της πολιτικής που είναι επιθυμητό να ακολουθηθεί για την προβολή περιοχών.

Πολύ σημαντικά κρίνονται τα αποτελέσματα από τη μελέτη του σεναρίου χρήσης για την αναζήτηση διαμονής, που καλύπτει έναν από τους βασικούς σκοπούς του ιστοχώρου, αποκάλυψε μεγάλα προβλήματα χρησιμοποιησιμότητας. Από τη μελέτη φάνηκε ότι η μεθοδολογία για μελέτη σεναρίων χρήσης είναι ικανή να χρησιμοποιηθεί για επιβεβαίωση (ή μη) των σεναρίων χρήσης που προβλέφθηκαν, αλλά και να αποκαλύψει τον πραγματικό τρόπο με τον οποίο χρησιμοποιείται ένας ιστοχώρος.

**Για την αγορά τουρισμού της περιοχής που καλύπτει ο ιστοχώρος, αν και οι επισκέπτες του ιστοχώρου δεν είναι αντιπροσωπευτικοί για τους πραγματικούς ταξιδιώτες του νησιού, προκύπτουν εξίσου σημαντικά συμπεράσματα. Το παράδειγμα με τις προτιμήσεις για συγκεκριμένες περιοχές ενδιαφέροντος ανάλογα με τη γλώσσα του ιστοχώρου, αποκαλύπτει και επιβεβαιώνει την προώθηση συγκεκριμένων περιοχών μαζικού τουρισμού στο εξωτερικό σε αντίθεση με πιο ποιοτικές επιλογές που γίνονται στην εγχώρια αγορά. Επίσης αποκαλύπτεται η περίοδος όπου αρχίζουν να προγραμματίζουν οι επισκέπτες το ταξίδι τους τόσο από την αύξηση σε επισκέψεις στον ιστοχώρο μετά την περίοδο των Χριστουγέννων, όσο και από την μετατόπιση του ενδιαφέροντος προς επιχειρήσεις τουρισμού (σε βάρος της γενικότερης πληροφόρησης) πριν την τουριστική περίοδο.**

**Συμπεράσματα εξάγονται όμως γενικότερα και για άλλους ιστοχώρους και ειδικότερα για**

ιστοχώρους τουρισμού. Από τα στοιχεία που μελετήθηκαν πιστεύουμε ότι υπάρχουν τουλάχιστον δύο σημεία που πρέπει να χρησιμοποιούνται για την δυναμική προσαρμογή ενός ιστοχώρου: το ένα αφορά τις προελεύσεις των επισκεπτών που παρουσιάζουν απότομες αλλαγές και δεν προέρχονται από μηχανές αναζήτησης. Σε αυτές τις περιπτώσεις ο σύνδεσμος που οδηγεί τους επισκέπτες μπορεί να αναφέρεται σε πολύ συγκεκριμένη θεματολογία (παράδειγμα ο σύνδεσμος για επιλογή διαμονής για τη διεξαγωγή συνεδρίου σε συγκεκριμένη περιοχή) και προσφέρεται έτσι για προσαρμογή του ιστοχώρου για καλύτερη κάλυψη των ενδιαφερόντων αυτών των επισκεπτών. Το δεύτερο σημείο αναφέρεται σε ιστοχώρους τουρισμού για γεωγραφικές περιοχές με διεθνή προσανατολισμό όπως είναι η Κρήτη. Σε αυτές τις περιπτώσεις φαίνεται ότι η συγκεκριμένες περιοχές που ενδιαφέρουν τους επισκέπτες εξαρτώνται άμεσα από τις χώρες που προέρχονται και προσφέρεται πάλι αυτό το στοιχείο για προσαρμογή του ιστοχώρου ανάλογα με την πολιτική που θέλει να ακολουθήσει.

Οι συγκεκριμένες ερωτήσεις που χρησιμοποιήθηκαν σε καμία περίπτωση δεν εξαντλούν τις δυνατότητες που παρέχονται από το Σύστημα Καταγραφής. Όπως φάνηκε από την ανάλυση στις προηγούμενες ενότητες η διαδικασία της μελέτης των δεδομένων μπορεί να γίνει ακολουθώντας πολλά διαφορετικά μονοπάτια και αυτή η ευελιξία είναι που φέρνει το σύστημα σε θέση να χρησιμοποιηθεί σε μεγάλο εύρος θεματολογίας ιστοχώρων.

## **5.4 Σύγκριση Μεθόδων**

Στις προηγούμενες ενότητες παρουσιάστηκαν δύο διαφορετικές προσεγγίσεις για την ανάλυση χρησιμοποιησιμότητας σε ιστοχώρους.

Η πρώτη βασίστηκε σε παραδοσιακές μεθόδους όπως είναι γνωστές από κλασικές εφαρμογές ηλεκτρονικών υπολογιστών και εφαρμόζονται σε εφαρμογές διαδικτύου (ευριστική ανάλυση και πειράματα χρηστών).

Η δεύτερη βασίστηκε αποκλειστικά στη μεθοδολογία καταγραφής και ανάλυσης δεδομένων χρήσης ιστοχώρων που αναπτύχθηκε και παρουσιάστηκε στα πλαίσια της παρούσας διατριβής.

Όπως ήταν εξάλλου αναμενόμενο για μία εφαρμογή για την οποία γίνεται πρώτη φορά ανάλυση χρησιμοποιησιμότητας, οι παραδοσιακές μέθοδοι έφεραν στο φως σημαντικά προβλήματα χρησιμοποιησιμότητας που υπάρχουν στον ιστοχώρο. Αλλά και από την εφαρμογή της μεθοδολογίας με βάση τη Βάση Χρήσης εντοπίστηκαν σημαντικά σημεία που χρήζουν επέμβασης ενώ ήταν δυνατόν να εξαχθούν και επιπλέον στοιχεία που διευκολύνουν στην κατανόηση των επισκεπτών και των αναγκών τους.

Από τις παραδοσιακές μεθόδους προέκυψαν 5 πολύ συγκεκριμένα τμήματα του ιστοχώρου που χρήζουν βελτίωσης. Περιλαμβάνουν τη λειτουργία για επιλογή περιοχής αναζήτησης, την εύρεση αναλυτικού χάρτη, το τεχνικό πρόβλημα στην εμφάνιση των εικονιδίων προσθαφαίρεσης στο φυλλάδιο, την εμφάνιση των κριτηρίων αναζήτησης και την εμφάνιση του φυλλαδίου.

Από την ανάλυση με εφαρμογή της μεθοδολογία προέκυψαν συγκεκριμένα αποτελέσματα από τη σύγκριση εναλλακτικών μορφών σε τμήματα του ιστοχώρου, και από τη μελέτη του σεναρίου χρήσης για την αναζήτηση διαμονής. Προέκυψε γνώση για τις συνθήκες που οδηγούν σε συγκεκριμένες συμπεριφορές ενώ δόθηκαν επίσης παραδείγματα ταχείας αντίδρασης σε αλλαγές στις κατηγορίες επισκεπτών με στόχο την καλύτερη προσαρμογή του ιστοχώρου στις ανάγκες αυτών των επισκεπτών. Τέλος, από τα αρχικά διαγράμματα που παρουσιάστηκαν από τη Βάση Χρήσης έγιναν εμφανείς οι δυνατότητες ανάλυσης που παρέχει για την συμπεριφορά των επισκεπτών.

Ιδιαίτερο ενδιαφέρον από τη Βάση Χρήσης παρουσιάζει το γεγονός των διαφορετικών προτιμήσεων σε περιοχές ανάλογα με τη χώρα προέλευσης ή τις συνολικά διαφορετικές προτιμήσεις σε επιχειρήσεις έναντι περιοχές ανάλογα με τη χρονική περίοδο. Τα στοιχεία αυτά και οι περαιτέρω δυνατότητες που παρέχονται, βοηθούν αποφασιστικά στην κατανόηση των αναγκών των επισκεπτών και καθιστούν εφικτή την προσαρμογή του ιστοχώρου για την καλύτερη εξυπηρέτησή τους.

Δεν έχει ξεκαθαριστεί εντελώς εάν με την εξάντληση της μεθοδολογίας σε όλες τις παραμέτρους που καταγράφονται θα προέκυπταν παρόμοια συμπεράσματα όπως στις παραδοσιακές μεθόδους εάν η ανάλυση επικεντρωνόταν ακριβώς στις ίδιες εργασίες που κλήθηκαν να εκτελέσουν οι χρήστες των πειραμάτων χρηστών. Το πρόβλημα της επιλογής περιοχής όμως που ανακαλύφθηκε τόσο από την παραδοσιακή μελέτη, όσο και από τη μελέτη με βάση τη Βάση Χρήσης, συνηγορεί στην άποψη ότι η μεθοδολογία είναι ικανή να δώσει σημαντικά περισσότερα χρήσιμα συμπεράσματα σε σχέση με τις παραδοσιακές μεθοδολογίες χρησιμοποίησιμότητας, που όμως εξαιτίας του πλήθους τους και του μεγέθους της ανάλυσης που χρειάζεται για να παραχθούν δεν ήταν δυνατόν να προκύψουν στα πλαίσια της παρούσας εργασίας.

Ένα σαφές πλεονέκτημα της μεθοδολογίας καταγραφής και ανάλυσης δεδομένων χρήσης είναι ότι πέραν της αρχικής εγκατάστασης του μηχανισμού καταγραφής χρήσης του ιστοχώρου, δεν απαιτεί περαιτέρω πόρους και μπορεί να χρησιμοποιηθεί για την παρακολούθηση σε πραγματικό χρόνο της ανταπόκρισης της εφαρμογής στους επισκέπτες. Επίσης μπορεί για τα βασικά στοιχεία καταγραφής (όνομα σελίδας, επίσκεψη κλπ.) να ενσωματωθεί πολύ εύκολα σε οποιοδήποτε ιστοχώρο.

Σε αντίθεση με το παραπάνω, τόσο η διενέργεια της ευριστικής αξιολόγησης, όσο και τα πειράματα χρηστών απαιτούν σημαντικούς πόρους κάθε φορά που επιχειρείται η διενέργεια σχετικής μελέτης. Οι ειδικοί για την ευριστική αξιολόγηση πρέπει να είναι διαθέσιμοι ενώ πρέπει να βρεθούν χρήστες για τα πειράματα χρηστών, πρέπει να υπάρχει ο ειδικός εξοπλισμός για τη διενέργεια των πειραμάτων και πρέπει να γίνει η προετοιμασία για τις εργασίες που θα κληθούν να εκτελέσουν. Όλα αυτά απαιτούν μεγάλους πόρους τόσο από οικονομική όσο και από χρονική άποψη.

Έχοντας υπ' όψιν τα χαρακτηριστικά των εφαρμογών στον Παγκόσμιο Ιστό, όπου αλλαγές γίνονται συχνά και χωρίς μεγάλο κόστος, η προσέγγιση της μεθοδολογίας καταγραφής και ανάλυσης δεδομένων χρήσης φαντάζει περισσότερο ευέλικτη και αποτελεσματική.

Εάν ληφθεί υπ' όψιν ότι το μεγαλύτερο μέρος των προβλημάτων από τις παραδοσιακές μεθόδους ανακαλύφθηκε από την ευριστική ανάλυση, μπορεί ως καλή λύση για εφαρμογές ιστοχώρου να θεωρηθεί ο συνδυασμός της ευριστικής ανάλυσης με την ανάλυση με τη μεθοδολογία καταγραφής και ανάλυσης δεδομένων χρήσης. Η πρώτη στηρίζεται στη γνώση των ειδικών σε θέματα χρησιμοποιησιμότητας και μπορεί να εφαρμοστεί στα αρχικά στάδια σχεδιασμού της εφαρμογής. Η δεύτερη βασίζεται στην πραγματική χρήση από τους πραγματικούς χρήστες και επιτρέπει τη δοκιμή νέων μορφών αλληλεπίδρασης χωρίς τις διαδικασίες που απαιτούνται για πειράματα χρηστών, την επιβεβαίωση των σεναρίων χρήσης για τα οποία σχεδιάστηκε η εφαρμογή και την παρακολούθηση και ταχεία αντίδραση σε νέες συνθήκες που προκύπτουν κατά τη λειτουργία της εφαρμογής.

## 6 Προστασία Προσωπικών Δεδομένων κατά την Καταγραφή Χρήσης στο Web

Προσωπικά δεδομένα θεωρούνται γενικώς όλα τα δεδομένα που μπορεί να οδηγήσουν στην αποκάλυψη της ταυτότητας του χρήστη. Σε αυτά περιλαμβάνονται για παράδειγμα το πραγματικό όνομα του χρήστη, η διεύθυνσή του ή το τηλέφωνό του. Δεδομένα που δεν οδηγούν **άμεσα** στην πραγματική ταυτότητα του χρήστη δεν περιλαμβάνονται σε αυτά (όπως στατιστικά στοιχεία, συχνά επισκεπτόμενες σελίδες κλπ.).

Υπάρχουν επίσης δεδομένα που εκ' πρώτης όψεως δεν αποκαλύπτουν την ταυτότητα του χρήστη, αλλά μπορούν με έμμεσο τρόπο να το πράξουν. Για παράδειγμα ο συνδυασμός της διεύθυνσης IP, της ώρας και και ημερομηνίας της πρόσβασης σε έναν ιστοχώρο μπορεί σε συνδυασμό με τα στοιχεία που καταγράφουν οι Πάροχοι Υπηρεσιών Διαδικτύου (Internet Service Provider - ISP) να αποκαλύψουν την ταυτότητα του χρήστη. Επειδή τα στοιχεία από τις προσβάσεις στον ιστοχώρο υπάρχουν σε διαφορετικό σημείο από τα στοιχεία χρήσης υπηρεσιών Διαδικτύου των ISP η καταγραφή των δεδομένων για τις διευθύνσεις IP εν' γένει δεν δημιουργεί προβλήματα προσωπικών δεδομένων.

Τα μη προσωπικά δεδομένα εν' γένει μπορεί να γίνουν αντικείμενο επεξεργασίας. Εάν, με οποιοδήποτε τρόπο, μπορεί να συνδεθούν διαφορετικές προσβάσεις στον ίδιο ή σε περισσότερους ιστοχώρους με τον ίδιο χρήστη (χωρίς τη γνώση της ταυτότητάς του), είναι δυνατόν να δημιουργηθούν αναλυτικά διαγράμματα (profiles) για το συγκεκριμένο χρήστη. Μπορεί δηλαδή να αποθηκευτεί για παράδειγμα το είδος της πληροφορίας που αναζητάει κάποιος χρήστης, τα προϊόντα για τα οποία ενδιαφέρεται ή αγοράζει και γενικά διάφορα στοιχεία που συνθέτουν ένα αναλυτικό προφίλ για τον χρήστη. Όσο αυτά τα στοιχεία παραμένουν απομονωμένα από δεδομένα που αποκαλύπτουν την ταυτότητα του χρήστη, τέτοιου είδους πρακτικές δεν παρουσιάζουν πρόβλημα.

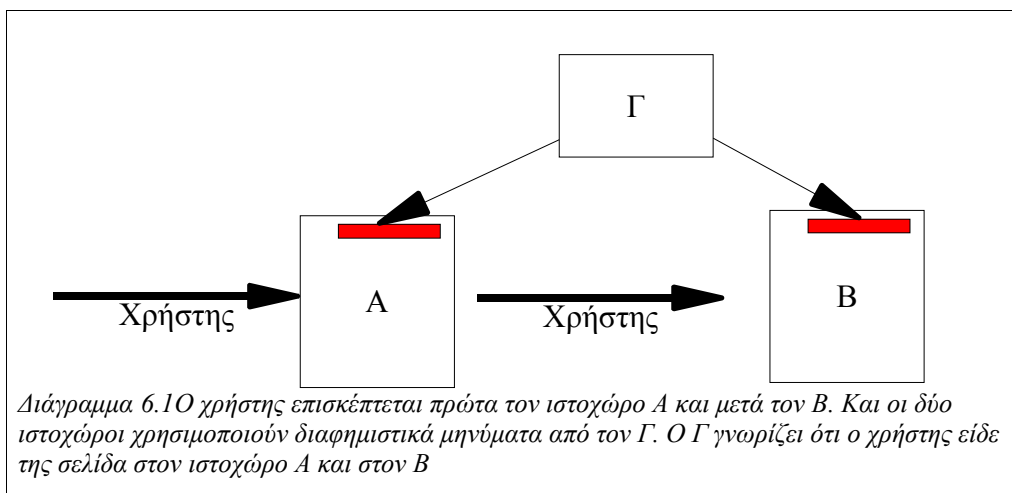
Υπάρχουν 3 κυρίως τρόποι για να γίνει αντιστοίχιση πολλών προσβάσεων στον ίδιο χρήστη (για αναλυτική περιγραφή βλέπε την ενότητα σχετικής έρευνας για Web Usage Analysis του Κεφαλαίου 2)

- χρησιμοποιείται session cookie και ο χρήστης το αποδέχεται. Αντιστοιχίζει προσβάσεις στον ιστοχώρο ο οποίος έθεσε το cookie για τη διάρκεια μίας επίσκεψης
- χρησιμοποιείται permanent cookie. Όμοια με το session cookie αλλά ακόμα και για διαφορετικές επισκέψεις

- εάν ο χρήστης χρησιμοποιεί μηχανήμα με σταθερή διεύθυνση IP, μπορούν να αντιστοιχιστούν οι προσβάσεις από τον ίδιο χρήστη

Παρόλο που υπάρχουν περιθώρια σφάλματος στα παραπάνω, εντούτοις αποτελούν έναν γενικά αξιόπιστο τρόπο στην αντιστοίχιση προσβάσεων από τον ίδιο χρήστη. Ειδικά η περίπτωση των cookies φαντάζει ασφαλής διότι αυτά στέλνονται μόνο στον εξυπηρετητή ο οποίος τα πρωτοέθεσε. Δηλαδή εάν τεθεί ένα τέτοιο cookie από τον ιστοχώρο Α, ο ιστοχώρος Β που ενδεχομένως θα επισκεφτεί ο χρήστης δεν γνωρίζει ότι επισκέφτηκε και τον Α.

Στην πράξη όμως υπάρχουν τρόποι όπου αποκτάται γνώση για τις προσβάσεις των χρηστών ακόμα και διαφορετικών ιστοχώρων. Αυτό γίνεται κυρίως από εταιρίες διαφημιστικών μηνυμάτων ή από εταιρίες που προσφέρουν υπηρεσίες καταγραφής κίνησης σε ιστοχώρους. Στο Διάγραμμα 6.1 φαίνεται ένα υποθετικό σχήμα για την επίσκεψη ιστοχώρων ενός χρήστη. Υπάρχει ο ιστοχώρος Α και ο Β. Και οι δύο έχουν ενσωματώσει ειδικό κώδικα ώστε να εξυπηρετούνται οι σελίδες τους με διαφημίσεις από την εταιρεία Γ (αντί διαφημίσεων θα μπορούσε να είναι καταγραφή κίνησης ιστοχώρων).



Ο χρήστης βλέπει σελίδες από τον Α και από τον Β. Επειδή σε κάθε σελίδα που βλέπει, εμφανίζεται το μήνυμα από τον Γ ο ιστοχώρος Γ μπορεί να θέσει cookie ο οποίος θα στέλνεται μόνο στον Γ. Αλλά αυτό του επιτρέπει να γνωρίζει ότι ο ίδιος χρήστης ενδιαφέρεται για το περιεχόμενο του Α και του Β.

Στην πραγματικότητα τέτοιου είδους εταιριών διαφημίσεων υπάρχουν πολλές οι οποίες έχουν δημιουργήσει πολύ μεγάλα δίκτυα συνεργαζόμενων ιστοχώρων. Είναι μάλλον απίθανο πλέον κάποιος χρήστης να θέλει να πλοηγηθεί στο Διαδίκτυο χωρίς να συναντήσει τέτοιου είδους διαφημίσεις. Επειδή τα δίκτυα είναι τεράστια, οι εταιρίες μπορούν και δημιουργούν πολύ αναλυτικά διαγράμματα για τους χρήστες.



Αρκεί κανείς να δει στο σκληρό δίσκο του υπολογιστή του τα permanent cookies που είναι αποθηκευμένα (στην περίπτωση που τα αποδέχεται) για να καταλάβει ότι από τους ιστοχώρους που επισκέπτεται κάποιος είναι δυνατόν να εξαχθούν πολλά συμπεράσματα για τον χρήστη (για παράδειγμα ποιες ηλεκτρονικές εφημερίδες διαβάζει).

Έως τώρα έγινε αναφορά σε μη προσωπικά δεδομένα που συλλέγονται με αυτόν τον τρόπο. Αλλά αρκεί κανείς να σκεφτεί ότι εάν σε ένα μόνο σημείο είτε ηθελημένα είτε άθελα αποκαλυφθεί η ταυτότητα του χρήστη, αυτός βρίσκεται πλέον εκτεθειμένος.

Ας υποθεθεί προς στιγμή ότι ο οργανισμός Γ στο παραπάνω παράδειγμα έχει σκοπό να βρει την πραγματική ταυτότητα ενός χρήστη. Μπορεί να ακολουθήσει διάφορους τρόπους (με διαφορετικές πιθανότητες επιτυχίας) για να το κατορθώσει, από τους οποίους θα περιγραφούν 3.

### **Τρόπος 1**

Η οντότητα Γ διατηρεί η ίδια ιστοχώρο όπου για τους σκοπούς του ιστοχώρου απαιτείται η εθελοντική για τους χρήστες υποβολή του ονόματός τους. Αυτόματα ο Γ μπορεί να συνδέσει τα στοιχεία που έχει καταγράψει από τις επισκέψεις σε άλλους ιστοχώρους. Παράδειγμα οργανισμών όπου αυτό είναι δυνατόν είναι ηλεκτρονικά καταστήματα τα οποία διαθέτουν και συνεργαζόμενους ιστοχώρους. Οι συνεργαζόμενοι ιστοχώροι εμφανίζουν διαφημιστικό μήνυμα του καταστήματος για τους χρήστες που θελήσουν να το επιλέξουν για κάποια αγορά. Το κατάστημα καθ' αυτό χρειάζεται τα στοιχεία του χρήστη για την αποστολή του προϊόντος. Παράδειγμα οργανισμού όπου αυτό θα ήταν δυνατόν είναι η διεθνής αλυσίδα Amazon.

### **Τρόπος 2**

Ο χρήστης χρησιμοποιεί στατική διεύθυνση. Ο Γ ερευνά δημόσιες λίστες συζητήσεων (mailing lists ή newsgroups όπου τα μηνύματα στέλνονται σε όλους τους εγγεγραμμένους χρήστες). Επειδή τα μηνύματα στέλνονται με email, εάν δεν έχει γίνει ειδική μέριμνα από τους διαχειριστές των λιστών, η διεύθυνση IP του αρχικού αποστολέα εμφανίζεται στο μήνυμα. Εάν εντοπιστεί από τον Γ ένα μήνυμα με IP αποστολής τη ζητούμενη IP διεύθυνση, γνωρίζει αυτόματα το email του χρήστη. Ακόμα και στην περίπτωση όπου η διεύθυνση δεν είναι στατική, εάν περιοριστεί η έρευνα σε μηνύματα που στάλθηκαν γύρω από τη χρονική στιγμή που έγινε η επίσκεψη στον ιστοχώρο, είναι εφικτή η εξακρίβωση της ταυτότητας του χρήστη.

### **Τρόπος 3**

Έστω ότι ο Α ή ο Β έχουν μία σελίδα που επεξεργάζεται τα στοιχεία μίας φόρμας όπου ο χρήστης

έδωσε το όνομά του. Ο χρήστης πιστεύει ότι δίνει τα στοιχεία στον Α ή το Β. Εάν (για οποιοδήποτε λόγο) τα στοιχεία του χρήστη μεταφέρονται υπό μορφή HTTP GET αίτησης (δηλαδή οι τιμές που έδωσε ο χρήστης μεταφέρονται στη διεύθυνση URL που πηγαίνει προς τον Α ή Β), τότε αυτή η διεύθυνση (μαζί με τα στοιχεία του χρήστη) θα εμφανιστεί στον Γ ως referrer.

Όλα τα παραπάνω παραδείγματα έγιναν με αναφορά εταιρίες διαφημίσεων, αλλά μπορεί το ίδιο καλά να γίνουν από υπηρεσίες που προσφέρουν ανάλυση επισκεψιμότητας. Τέτοιες υπηρεσίες υπάρχουν πολλές.

Τους τελευταίους μήνες, κυρίως λόγω σχετικών δημοσιευμάτων σε ειδικά περιοδικά, διαφαίνεται ένας προβληματισμός στους χρήστες για τους κινδύνους που εγκυμονεί η συλλογή τέτοιων στοιχείων. Δεν είναι σπάνιο να δίνονται συμβουλές από έγκυρα περιοδικά να απενεργοποιείται η εκτέλεση κώδικα JavaScript ή Cookies (είτε για λόγους ασφαλείας είτε για λόγους προστασίας προσωπικών δεδομένων).

Γι' αυτό το λόγο υπάρχουν πρωτοβουλίες τόσο από πλευράς της νομοθετικής εξουσίας όσο και από πλευράς της βιομηχανίας υπό μορφή αυτοελέγχου.

Από την πλευρά της βιομηχανίας αναπτύχθηκε υπό το W3C [30] το Platform for Privacy Preferences (P3P) [5] σαν πρότυπο για την αυτορύθμιση του χώρου. Συγκεκριμένα μπορεί οι ιστοχώροι να προσφέρουν μέσω ειδικού αρχείου στον εξυπηρετητή στοιχεία για την πολιτική που έχουν όσον αφορά την καταγραφή και επεξεργασία δεδομένων των χρηστών. Όταν ο χρήστης επισκέπτεται τον ιστοχώρο, η εφαρμογή πλοήγησης που χρησιμοποιεί (και είναι συμβατή με το πρότυπο) συγκρίνει τα στοιχεία του αρχείου με τις ρυθμίσεις που έχει κάνει ο χρήστης, και εάν δεν συμβαδίζουν ειδοποιεί σχετικά το χρήστη.

Νομοθετικά έχουν ρυθμιστεί τέτοια θέματα για παράδειγμα στη Γερμανία. Σύμφωνα με τον εκεί ισχύοντα νόμο, απαγορεύεται η κάθε είδους επεξεργασία ή καταγραφή προσωπικών δεδομένων, εφόσον ο χρήστης δεν έδωσε τη ρητή συγκατάθεσή του.

Από την πλευρά του χρήστη μπορούν επίσης να εφαρμοστούν μέθοδοι που προστατεύουν τα προσωπικά του δεδομένα. Για παράδειγμα μπορεί να ρυθμίσει για ποιους ιστοχώρους επιθυμεί να αποθηκεύονται cookies. Επίσης μπορεί να χρησιμοποιεί ειδικούς ενδιάμεσους σταθμούς (anonymizer) οι οποίοι φιλτράρουν τις αιτήσεις του ώστε να απομακρύνουν για παράδειγμα cookies, και καθιστούν αδύνατον στους ιστοχώρους να έχουν πρόσβαση στην πραγματική του διεύθυνση IP.

## 7 Ανακεφαλαίωση - Συνεισφορά

Στο Κεφάλαιο αυτό συνοψίζεται το υλικό που περιλαμβάνεται σε αυτή τη διατριβή και γίνεται επισήμανση της βασικής ερευνητικής συνεισφοράς.

### 7.1 Επισκόπηση Περιεχομένων

Στο Κεφάλαιο 1 έγινε μία σύντομη εισαγωγή στα κίνητρα και το σκοπό της διατριβής.

Στο Κεφάλαιο 2 έγινε μία εισαγωγή στο γνωστικό αντικείμενο της αλληλεπίδρασης του ανθρώπου με υπολογιστές και περιγράφηκε η διάσταση της χρησιμοποίησιμότητας ως μέτρο της ανταπόκρισης των εφαρμογών στις ανάγκες των χρηστών. Επισημάνθηκαν τα νέα δεδομένα που προκύπτουν από την ολοένα και μεγαλύτερη χρήση εφαρμογών που χρησιμοποιούν τεχνολογίες Διαδικτύου και ειδικότερα εφαρμογές πλοήγησης με βάση το HTTP πρωτόκολλο. Έγινε μία σύντομη αναφορά σε καθιερωμένες πρακτικές και μεθόδους για τη βελτίωση της χρησιμοποίησιμότητας και δόθηκε μία αναλυτική περιγραφή της σχετικής έρευνας σε θέματα ανάλυσης των δεδομένων που προκύπτουν από τη χρήση των ιστοχώρων με βάση το HTTP πρωτόκολλο.

Στο Κεφάλαιο 3 παρουσιάστηκε ένα σύστημα συλλογής και καταγραφής δεδομένων χρήσης ιστοχώρων που αναπτύχθηκε στα πλαίσια της διατριβής. Αναπτύχθηκε και παρουσιάστηκε μία μεθοδολογία η οποία οδηγεί στην επιλογή των δεδομένων που καταγράφονται.

Στο κεφάλαιο 4 παρουσιάστηκε μία μεθοδολογία για την ανάλυση επισκεψιμότητας ιστοχώρων βασισμένη στο σύστημα καταγραφής που αναπτύχθηκε.

Στο Κεφάλαιο 5 περιγράφηκε η διενέργεια δύο ξεχωριστών μελετών χρησιμοποίησιμότητας που πραγματοποιήθηκαν στα πλαίσια της διατριβής πάνω στον ιστοχώρο τουριστικού προορισμού. Η πρώτη χρησιμοποίησε ευριστική αξιολόγηση και πειράματα χρηστών που είναι καθιερωμένες μέθοδοι για τη μελέτη χρησιμοποίησιμότητας. Στη δεύτερη μελέτη έγινε προσπάθεια να χρησιμοποιηθούν τα δεδομένα από τη χρήση του ιστοχώρου και να εξαχθούν συμπεράσματα για τη χρησιμοποίησιμότητά του.

Στο Κεφάλαιο 6 αναπτύχθηκε ο προβληματισμός που υπάρχει κατά τη συλλογή δεδομένων για την παραβίαση στοιχειωδών δικαιωμάτων των χρηστών για προστασία προσωπικών δεδομένων.

## 7.2 Ερευνητική Συνεισφορά

Στην παρούσα διατριβή διερευνήθηκε η συνεισφορά των δεδομένων που προκύπτουν από τη χρήση ιστοχώρων στη βελτίωση της χρησιμοποίησιμότητας εφαρμογών στο Διαδίκτυο.

Η συλλογή και ανάλυση δεδομένων χρήσης ιστοχώρων, όπως και ο ίδιος ο Παγκόσμιος Ιστός, είναι μία σχετικά νέα ερευνητική περιοχή. Ως τέτοια έχει ακόμα μεγάλες ανάγκες για την αποκάλυψη ολόκληρου του εύρους των προβλημάτων που μπορεί να λύσει.

Η χρησιμοποίηση των δεδομένων χρήσης ιστοχώρων για την εξαγωγή συμπερασμάτων χρησιμοποίησιμότητας είναι ουσιαστικά μία επέκταση ήδη καθιερωμένων πρακτικών από παραδοσιακές μελέτες χρησιμοποίησιμότητας. Το νέο στοιχείο στις εφαρμογές ιστοχώρου είναι ότι έχουν ως κοινό παρανομαστή τη χρήση του HTTP πρωτοκόλλου, το οποίο επιτρέπει τη γενικευμένη καταγραφή της συμπεριφοράς των επισκεπτών σε τέτοιες εφαρμογές χωρίς την επιμέρους παρέμβαση σε κάθε εφαρμογή.

Η πολυπλοκότητα όμως των αλγορίθμων που παρουσιάζονται σε σχετικές ερευνητικές εργασίες μαζί με την αφαίρεση που αναγκαστικά γίνεται κάνει τη χρήση συστημάτων καταγραφής και ανάλυσης δεδομένων χρήσης ιδιαίτερα απρόσιτη και δυσνόητη στη μεγάλη πλειοψηφία των επιχειρήσεων και οργανισμών που θα θελήσουν να βελτιώσουν τους ιστοχώρους που αναπτύσσουν και συντηρούν. Για να γίνουν πιο προσιτά τέτοια συστήματα απαιτείται η τυποποίηση της ανάλυσης μέσω συστηματικών μεθοδολογιών και να προσαρμοστούν σε συγκεκριμένες κατηγορίες εφαρμογών στο Διαδίκτυο, ώστε να γίνουν εμφανή τα πλεονεκτήματα που έχουν.

Στα πλαίσια αυτής της διατριβής αναπτύχθηκε ένα τέτοιο σύστημα με τη χρήση ευρέως διαδεδομένων τεχνολογιών όπως σχεσιακών συστημάτων διαχείρισης βάσεων δεδομένων. Το χαρακτηριστικό του είναι ότι απαλείφεται το ιδιαίτερα χρονοβόρο στάδιο της προ επεξεργασίας των δεδομένων από τα αρχεία καταγραφής, ενώ γίνεται εκμετάλλευση των ειδικών γνώσεων που έχουν οι προγραμματιστές του ιστοχώρου για το είδος της πληροφορίας που παρουσιάζεται (περιεχόμενο, υπάρχουσες πηγές δεδομένων, γλώσσες εμφάνισης της ίδιας πληροφορίας κλπ.). Αναπτύχθηκε επίσης μία μεθοδολογία για την επιλογή των δεδομένων που θα καταγράφονται η οποία παρέχει μεγάλη ευελιξία στους αναλυτές να επικεντρώσουν στα σημεία εκείνα που επιθυμούν να μελετηθούν και αντιμετωπίζει προβλήματα που προκύπτουν από τη δυναμικότητα σύγχρονων ιστοχώρων που δεν μπορεί να λυθούν με παραδοσιακά συστήματα που βασίζονται στο URL της κάθε αίτησης για να προσδιορίσουν το περιεχόμενο του ιστοχώρου. Ειδικά η εγγενής υποστήριξη της έννοιας της πολυγλωσσίας του περιεχομένου, εξ' όσων γνωρίζουμε, δε συναντάται σε παρόμοιο προϊόν ή σχετική εργασία στην περιοχή.

Με βάση αυτό το σύστημα προτάθηκε μία συστηματική μεθοδολογία για την ανάλυση των δεδομένων που συλλέγονται, έχοντας υπόψη την κατηγορία εφαρμογών για τουριστικούς προορισμούς. Αναλύεται πώς, με τη συστηματική μελέτη των μεταβολών που υπάρχουν στον αριθμό των επισκέψεων και στον αριθμό των σελιδοπροβολών με βάση συγκεκριμένες παραμέτρους, μπορεί να εξαχθούν συμπεράσματα για τη συμπεριφορά των επισκεπτών. Σημαντικό εργαλείο για τη μελέτη της συμπεριφοράς των επισκεπτών δίνεται με την απεικόνιση των σεναρίων χρήσης σε σειριακά σχήματα και η δυνατότητα υποβολής ερωτήσεων που περιλαμβάνουν περιορισμούς σε όλα τα στοιχεία που χαρακτηρίζουν την κάθε επίσκεψη και την κάθε σελιδοπροβολή. Δόθηκαν συγκεκριμένες τεχνικές που αναλύουν όχι μόνο στατικά προκαθορισμένα σενάρια χρήσης, αλλά μπορούν επίσης να αναλύσουν με συγκεκριμένα βήματα σενάρια που δεν προβλέφθηκαν κατά τον σχεδιασμό του ιστοχώρου.

Για να φανούν τα πλεονεκτήματα της μεθοδολογίας, αυτή εφαρμόστηκε σε έναν πραγματικό ιστοχώρο τουριστικού προορισμού. Από τα αποτελέσματα της σχετικής μελέτης προέκυψαν σημαντικά συμπεράσματα που σχετίζονται τόσο με προβλήματα χρησιμοποίησιμότητας που ανακαλύφθηκαν όσο και γενικότερα συμπεράσματα που μπορεί να χρησιμοποιηθούν από επιχειρήσεις ή οργανισμούς που ενδιαφέρονται για τον τομέα του τουρισμού της περιοχής. Για παράδειγμα αποκαλύφθηκαν σημαντικές διαφοροποιήσεις στις προτιμήσεις των επισκεπτών για πληροφόρηση για περιοχές ενδιαφέροντος ανάλογα με τη χώρα προέλευσής τους και φάνηκε τα εντελώς διαφορετικά κριτήρια που έχουν οι εγχώριοι επισκέπτες σε σχέση με αυτούς άλλων χωρών για την επιλογή του τόπου διακοπών. Επίσης επιβεβαιώνεται η χρονική περίοδος έναρξης της αναζήτησης προορισμού διακοπών από τον αριθμό των επισκέψεων που αρχίζει αυξάνεται αρκετά πριν την τουριστική περίοδο. Τα πειράματα με τις εναλλακτικές μορφές αλληλεπίδρασης τόσο σε μορφή εικονιδίων όσο και στην τοποθεσία συνδέσμων επηρεάζουν σημαντικά τη χρήση τους από τους επισκέπτες και προσφέρεται έτσι ως κριτήριο για την αξιολόγηση του επιπέδου διεπικοινωνίας.

Όλες οι μέθοδοι που χρησιμοποιήθηκαν απαιτούν μόνο στοιχειώδη γνώση στατιστικής και γνώσης της περιοχής που αναφέρεται ο ιστοχώρος ενώ η χρήση της SQL ως γλώσσα ερωτήσεων στη Βάση Χρήσης . Ιδιαίτερα σημαντικό κρίνεται το γεγονός ότι δε χρησιμοποιήθηκαν προσωπικά δεδομένα των επισκεπτών.

Όσον αφορά τη συνιστώσα της χρησιμοποίησιμότητας (usability) προέκυψε ότι η ανάλυση επισκεψιμότητας είναι χρήσιμη ως συμπλήρωμα σε παραδοσιακές μεθόδους ανάλυσης χρησιμοποίησιμότητας και αναγκαία λόγω των ιδιοτήτων που παρουσιάζουν οι εφαρμογές

ιστοχώρου. Μπορεί να δώσει άμεσα απαντήσεις για τη σχετική επιτυχία ή αποτυχία νέων μορφών αλληλεπίδρασης χωρίς τη μεσολάβηση χρονοβόρων διαδικασιών προετοιμασίας και διεξαγωγής πειραμάτων χρηστών. Σε ορισμένο βαθμό μπορεί ο συνδυασμός ενός τέτοιου συστήματος με μία λιγότερο δαπανηρή παραδοσιακή μέθοδο όπως η ευριστική ανάλυση να συμβάλλει στη μείωση των απαιτούμενων πόρων (χρόνος, εξοπλισμός, κόστος) για τη βελτίωση χρησιμοποίησιμότητας και να κάνει το αντικείμενο πιο προσιτό σε μεγαλύτερη μερίδα ιστοχώραν.

## 8 Μελλοντικές Επεκτάσεις

Το σύστημα βάσης δεδομένων και η μεθοδολογία καταγραφής και ανάλυσης που αναπτύχθηκαν είχαν ως αφετηρία την κατηγορία εφαρμογών στο Διαδίκτυο με θεματολογία τον τουρισμό μίας γεωγραφικής περιοχής.

Από αυτή την άποψη θα ήταν χρήσιμο να εφαρμοστεί η μεθοδολογία και σε άλλους ιστοχώρους τουρισμού με σκοπό τη διερεύνηση διαδικασιών και μεθοδολογιών προσαρμογής του συστήματος σε αυτούς, ενώ πρέπει να εξεταστεί κατά πόσο το μοντέλο των δεδομένων που αποθηκεύονται απαιτεί τυχόν επεκτάσεις. Θεωρούμε όμως ότι είναι εξαιρετικά δύσκολο να επιτραπεί από τους ιδιοκτήτες παρόμοιων ιστοχώρων ο “πειραματισμός” με τα συστήματά τους μιας και επιδιώκουν την αδιάλειπτη λειτουργία, ενώ θα έπρεπε να αποκαλύψουν και τα μοντέλα δεδομένων που χρησιμοποιούν εσωτερικά.

Παραμένει επίσης ανοικτό το θέμα κατά πόσο μπορεί να γίνει εφαρμογή της μεθοδολογίας και σε ιστοχώρους που δεν έχουν ως θέμα τον τουρισμό. Θεωρούμε ότι θα ήταν χρήσιμο να επιχειρηθεί η εφαρμογή σε ιστοχώρους που έχουν πηγές δεδομένων με πλούσια κατηγοριοποίηση πληροφορίας ή ιστοχώρους με μεγάλο αριθμό επιχειρήσεων (π.χ. Εμποροβιομηχανικό Επιμελητήριο). Όμοιο ενδιαφέρον θα υπήρχε για εφαρμογή σε ιστοχώρους ηλεκτρονικού εμπορίου για “ανάλυση καλαθιού αγορών” (market basket analysis).

Όσον αφορά καθ' αυτό το σύστημα καταγραφής, παρουσιάζει ενδιαφέρον η τυποποίηση των APIs που προσφέρει για ερωτήσεις από τρίτους και η προσφορά τέτοιου API υπό μορφή Web Services για διάφορες πτυχές της συμπεριφοράς των επισκεπτών (π.χ. οι πιο συχνές φράσεις αναζήτησης, οι πιο συχνά επισκεπτόμενες περιοχές κλπ.).

## 9 Αναφορές – Βιβλιογραφία

- [1] Jian Pei, Jiawei Han, Behzad Mortazavi-asl, Hua Zhu, Mining Access Patterns Efficiently from Web Logs (2000), Pacific-Asia Conference on Knowledge Discovery and Data Mining
- [2] Raymond Kosala, Hendrik Blockeel, *Web Mining Research: A Survey*, SIGKDD: SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining, ACM, 2000
- [3] Catherine Bounsaythip, Esa Rinta-Runsala, *Overview of Data Mining for Customer Behavior Modeling*, Research Report TTE1-2001-18, VTT Information Technology, June 2001, Espoo, Finland
- [4] Jakob Nielsen, *Severity Ratings for Usability Problems*, <http://www.useit.com/papers/heuristic/severityrating.html>
- [5] Platform for Privacy Preferences (P3P) Project, <http://www.w3.org/P3P/>
- [6] Μανόλης Αναστασιάδης, *Tourism Information Systems, A State of the Art*, Internal Report Campiello Long Term Research Project 25572, Technical University of Crete
- [7] Myra Spiliopoulou, *Web usage mining for web site evaluation*. Communications of the ACM, 43(8):127--134, 2000.
- [8] R. Cooley, B. Mobasher, and J. Srivastava. *Data Preparation for Mining World Wide Web Browsing Patterns*. Journal of Knowledge and Information Systems, 1(1):5-32, February 1999.
- [9] James Pitkow Xerox Palo Alto Research Center. *In Search of Reliable Usage Data on the WWW*, In Proceedings of the Sixth International World Wide Web Conference (1997), pages 452-463
- [10] Thomas P. Novak and Donna L. Hoffman, *New Metrics for New Media: Toward the Development of Web Measurement Standards*, World Wide Web Journal, Winter, 2(1), 213-246 (1996)
- [11] Υπηρεσία GeoIP της Maxmind, <http://www.maxmind.com/>
- [12] Alex G. Büchner, Maurice D. Mulvanna, *Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining*, SIGMOD Record 27(4): 54-61(1998)



- [13] Web Utilization Miner (WUM), <http://wum.wiwi.hu-berlin.de/>
- [14] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan, *Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data*, SIGKDD Explorations, (1) 2, 2000
- [15] R. Cooley, B. Mobasher, J. Srivastava, *Web Mining: Information and Pattern Discovery on the World Wide Web*, In Proceedings of the 9<sup>th</sup> IEEE International Conference on Tools with Artificial Intelligence 1997
- [16] AWStats, <http://awstats.sourceforge.net/>
- [17] *Tourism and the European Union*, Enterprise Directorate-General of the European Commission, <http://europa.eu.int/comm/enterprise/services/tourism/tourismeu.htm>
- [18] Bamshad Mobasher, Namit Jain, Eui-Hong (Sam) Han, Jaideep Srivastava. *Web Mining: Pattern Discovery from World Wide Web Transactions*, Technical Report TR 96-050, University of Minnesota, Dept. of Computer Science, Minneapolis, 1996.
- [19] Θεόδωρος Μαργαζάς, *Μεθοδολογία και σύστημα για την αξιολόγηση οπτικών αντικειμένων συστημάτων αλληλεπίδρασης χρήστη υπολογιστή*, Διατριβή που παρουσιάστηκε στο Τμήμα Ηλεκτρονικής και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης (2000)
- [20] George Murray and Tania Costanzo, *Usability and the Web: An Overview*, Network Notes #61, ISSN 1201-4338, Information Technology Services, National Library of Canada, August 1999, <http://www.nlc-bnc.ca/9/1/p1-260-e.html>
- [21] Jakob Nielsen, *The Difference Between Web Design and GUI Design*, <http://www.useit.com/alertbox/9705a.html>
- [22] Alex G. Büchner, Maurice D. Mulvenna, Sarab S. Anand, John G. Hughes, *An Internet-enabled Knowledge Discovery Process*, 9th Int'l. Database Conf., Hong Kong, pp. 13-27, July 1999
- [23] Yongjian Fu, Kanwalpreet Sandhu, Ming-Yi Shih, *Clustering of Web Users Based on Access Patterns*, In Proceedings of the 1999 KDD Workshop on Web Mining, San Diego, CA. Springer-Verlag

- [24] Chetz Colwell & Helen Petrie, *Evaluation of Guidelines for Designing Accessible World Wide Web Pages*, <http://www.snv.jussieu.fr/inova/publi/ntevh/evaluation.htm#creation>
- [25] Osmar R. Zaiane, Man Xin, Jiawei Han, *Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs*, In Proc. Advances in Digital Libraries Conf. (ADL'98), Santa Babara, CA, April, 1998
- [26] Spool, Jared., Scanlon, Tara., Schroeder, Will., Snyder, Carolyn., DeAngelo, Terri *Web Site Usability: a Designer's Guide*, User Interface Engineering, North Andover, Massachusetts (1997)
- [27] Scott Butler, *Why Web Design is Similar to GUI Design*, από αναδημοσίευση στο <http://www.useit.com/alertbox/scottbutler.html>
- [28] <http://www.webusagemining.com/sys-tmpl/door/>
- [29] RFC2616, Hypertext Transfer Protocol – HTTP/1.1, <http://www.w3.org/Protocols/rfc2616/rfc2616.txt>
- [30] The World Wide Web Consortium (W3C), <http://www.w3.org>
- [31] Accrue Software, A Buyer's Guide to eBusiness Intelligence, <http://www.accrue.com>
- [32] Combined Log Format <http://httpd.apache.org/docs/logs.html#combined>
- [33] Ενδεικτικές εταιρείες εξωτερικής παροχής υπηρεσιών ανάλυσης <http://www.hitbox.com>, <http://www.counter.com>
- [34] Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern <http://www.ivw.de>
- [35] Spiegel Online <http://www.spiegel.de>, Heise Online <http://www.heise.de>
- [36] Matt Cutler, Jim Sterne NetGenesis Corp. <http://www.netgen.com>. *E-Metrics- Business Metrics for the New Economy*, White Paper 2000
- [37] Lara D. Catledge, James E. Pitkow, *Characterizing Browsing Strategies in the World-Wide Web*, In Computer Networks and ISDN Systems 27. Elsevier Science, 1995
- [38] heise online aktiviert neues IVW-Messverfahren, <http://www.heise.de/newsticker/data/mw-19.12.01-000/>

- [39] Google Search Engine, <http://www.google.com>
- [40] Jesus Mena, *Data Mining Your Website (1999)*
- [41] Ming-Syan Chen, Jong Soo Park, Philip S. Yu, *Efficient Data Mining for Path Traversal Patterns*, IEEE Trans. on Knowledge and Data Engineering, 10(2):209--221, March 1998
- [42] R. Cooley, B. Mobasher, J. Srivastava, *Grouping Web Page References into Transactions for Mining World Wide Web Browsing Patterns*, Technical Report TR 97-021, Dept. of Computer Science, Univ. of Minnesota, Minneapolis, USA, June 1997
- [43] Ευρωπαϊκό Πρόγραμμα Crete TOURnet, DGXIII, TEN-ISDN, No 45549, <http://www.crete.tournet.gr>
- [44] The Microsoft Corporation, <http://www.microsoft.com>
- [45] Sun Java Technology, <http://java.sun.com>
- [46] PHP.ORG, <http://www.php.org>
- [47] [Jakob Nielsen](#)'s Alertbox for May 1996, Top Ten Mistakes in Web Design, <http://www.useit.com/alertbox/9605.html>
- [48] The Webalizer, <http://www.webalizer.org>
- [49] Opera Software, <http://www.opera.com>
- [50] IVW stellt umstrittenes Messverfahren um, <http://www.onlinejournalismus.de/webwatch/ivw.shtml>
- [51] B. Berendt and M. Spiliopolou, *Analysis of navigation behaviour in web sites integrating multiple information systems*, VLDB Journal 9 (2000) 56-75
- [52] DNS Related RFCs, <http://www.dns.net/dnsrd/rfc/>
- [53] GeoIP, <http://www.maxmind.com/>
- [54] Squid Web Proxy Cache, <http://www.squid-cache.org/>
- [55] Rakesh Agrawal, Ramakrishnan Srikant, *Fast Algorithms for Mining Association Rules*, Proc. 20th Int. Conf. Very Large Data Bases, VLDB (1994)

- [56] Jochen Hipp, Ulrich Güntzer, Gholamreza Nakhaeizadeh, *Algorithms for Association Rule Mining A General Survey and Comparison*, IKGDD Explorations 2000
- [57] Sunita Sarawagi, *Integrating Association Rule Mining with Relational Database Systems: Alternatives and Implications*, In Proc. 1998 ACM-SIGMOD, pp 343--354
- [58] Myra Spiliopoulou. *The laborious way from data mining to web mining*. Int. Journal of Comp. Sys., Sci. & Eng., Special Issue on ``Semantics of the Web'', 14:113-126, Mar. 1999
- [59] Anupam Joshi, Raghu Krishnapuram, *Robust Fuzzy Clustering Methods to Support Web Mining*, Proc. ACM SIGMOD Workshop on Data Mining and Knowledge Discovery, Aug 98
- [60] Codd, E.F., *Providing OLAP (on-line analytical processing) to user-analysts: an IT mandate*, Technical Report , E.F. Codd and Associates, 1993
- [61] M. Spiliopoulou and C. Pohle. *Data mining for measuring and improving the success of web sites*. In R. Kohavi and F. Provost, editors, Journal of Data Mining and Knowledge Discovery, Special Issue on E-commerce, volume 5, pages 85-114. Kluwer Academic Publishers, Jan.-Apr. 2001.
- [62] Nigel Pendse, *Fast Analysis of Shared Multidimensional Information*,  
<http://www.olapreport.com/fasmi.htm>
- [63] Tom Conlon and Rob Zare, Program Managers SQL Server Business Intelligence Unit Microsoft Corp., *Introduction to Analysis Services and OLAP Technology*, Tech Edd 2002
- [64] Jiawei Han, *OLAP Mining: An Integration of OLAP with Data Mining*, In Proceedings of the 7th IFIP 2.6 Working Conference on Database Semantics (DS-7), pages 1--9, 1997
- [65] *ACM SIGCHI Curricula for Human-Computer Interaction*, Report of the ACM Special Interest Group on Computer-Human Interaction (SIGCHI) Curriculum Development Group,  
<http://www.acm.org/sigchi/cdg/index.html>
- [66] Brad A. Myers. "A Brief History of Human Computer Interaction Technology." ACM interactions. Vol. 5, no. 2, March, 1998. pp. 44-54.
- [67] Nielsen, J. (1993). *Usability Engineering*. AP Professional, Cambridge, MA.
- [68] Pär Carlshamre (1994). *A Field Study in Usability Engineering: Bringing in the Technical Communicators*

- [69] Bob Baxley, *What is a Web Application?*  
[http://www.boxesandarrows.com/archives/what\\_is\\_a\\_web\\_application.php](http://www.boxesandarrows.com/archives/what_is_a_web_application.php)
- [70] B. Shneiderman, J. Nielsen, S. Butler, M. Levi, and F. Con-rad., *Is the web really different from everything else?* In Proc. of CHI98, 1998
- [71] M. Spiliopoulou, L. C. Faulstich, *WUM - A tool for WWW Utilization Analysis*, Proceedings of EDBT Workshop WebDB'98, LNCS1590, pages 184-203. Springer Verlag, 1999
- [72] Bettina Berendt, *Web usage mining, site semantics, and the support of navigation*, WEBKDD 2000 Worskhop on Web Mining for E-Commerce
- [73] Berendt, B. & Spiliopoulou, M. (2000). *Analysis of navigation behaviour in web sites integrating multiple information systems*. The VLDB Journal, 9, 56-75.
- [74] Berendt, B. (2001). *Understanding web usage at different levels of abstraction: coarsening and visualising sequences*. In Proceedings of the Workshop "WEBKDD 2001 - Mining Log Data Across All Customer TouchPoints"

## 10 Παράρτημα Α – Ιστοχώροι Τουριστικού Προορισμού

Τα ταξίδια πάντα θεωρούνταν από τις αγαπημένες δραστηριότητες των ανθρώπων επειδή μεταξύ άλλων δίνουν την ευκαιρία για εξερεύνηση νέων τόπων, τη γνωριμία νέων ανθρώπων και διαφορετικών τρόπων ζωής. Ως εκ τούτου αποτελεί ευκαιρία για επιχειρηματική εκμετάλλευση.

Ο τουρισμός αποτελεί μία πολύ σημαντική συνιστώσα στα οικονομικά αλλά και κοινωνικά και πολιτισμικά δρώμενα πολλών χωρών. Στην Ευρωπαϊκή Ένωση υπολογίζεται ότι 8 εκατομμύρια εργαζόμενοι απασχολούνται άμεσα στον τουριστικό τομέα και αντιπροσωπεύουν το 5% του εργατικού δυναμικού και του ακαθάριστου ευρωπαϊκού προϊόντος. Το 2000 πέντε χώρες-μέλη της Ευρωπαϊκής Ένωσης συγκαταλέγονταν ανάμεσα στους 10 πρώτους τουριστικούς προορισμούς παγκοσμίως [17].

Από τη στιγμή που ο δυνητικός ταξιδιώτης έχει επιλέξει τον προορισμό του, έρχεται αντιμέτωπος με ένα σύνολο πολύπλοκων προβλημάτων που πρέπει να επιλύσει. Τα προβλήματα αυτά συνοψίζονται στον τρόπο μετάβασης στον προορισμό, διαμονή, διατροφή, εξερεύνηση της περιοχής και γενικά η διασκέδαση. Βέβαια η βαρύτητα που δίνει ο καθένας σε κάθε μία από τις προηγούμενες παραμέτρους ποικίλλει ενώ η συγκεκριμένη λύση που θα επιλεγεί εξαρτάται από ζητήματα όπως διαθεσιμότητα ή από οικονομικά ζητήματα.

Η πολυπλοκότητα αυτή έχει οδηγήσει στην ανάγκη ύπαρξης ειδικών στο θέμα προγραμματισμού του ταξιδιού, που είναι τα γραφεία τουρισμού. Συνήθως προσφέρουν ολοκληρωμένες λύσεις υπό μορφή “πακέτων” αλλά σε αυτό ακριβώς το σημείο εμφανίζεται η δυσανεμία των ταξιδιωτών, διότι στην προσπάθεια των γραφείων να εξυπηρετήσουν μία σχετικά μεγάλη ομάδα ανθρώπων το αποτέλεσμα καταλήγει να είναι αυτό που ονομάζεται “μαζικός τουρισμός” χωρίς να λαμβάνονται υπόψιν οι ιδιαιτερότητες του κάθε ταξιδιώτη.

Το Διαδίκτυο προσφέρει σε αυτό το σημείο μοναδικά πλεονεκτήματα σε σχέση με τα παραδοσιακά ταξιδιωτικά γραφεία ή ακόμα και άλλου είδους πηγών πληροφόρησης όπως βιβλία ή ψηφιακά offline μέσα. Η πληροφορία είναι εύκολο να ενημερώνεται με αποτέλεσμα να είναι πάντα επίκαιρη ενώ μπορεί να προσφέρει υπηρεσίες “στο σπίτι” του χρήστη όπως:

- άμεσος έλεγχος διαθεσιμότητας πτήσεων και διαμονής
- πληροφόρηση για καιρικές συνθήκες και για δρομολόγια μεταφορών
- ενήμερες συναλλαγματικές ισοτιμίες

- κράτηση εισιτηρίων για μεταφορές ή κράτηση δωματίων

Δεν ξαφνιάζει λοιπόν ότι στο Διαδίκτυο ο τουρισμός είναι ένας από τους κυρίαρχους τομείς ηλεκτρονικού επιχειρήν και κατέχει από τα μεγαλύτερα ποσοστά από άποψη όγκου συναλλαγών στο Διαδίκτυο.

Σε αυτό το παράρτημα θα αναπτυχθούν τα χαρακτηριστικά ιστοχώρων που καλύπτουν το τουριστικό προϊόν μιας γεωγραφικής περιοχής με βάση τη σχετική εργασία στο [6] και θα περιγραφούν χαρακτηριστικοί ιστοχώροι που λειτουργούν στο Διαδίκτυο. Τέλος θα περιγραφεί ο σχετικός ιστοχώρος για την Κρήτη που αναπτύχθηκε στα πλαίσια Ευρωπαϊκού ερευνητικού προγράμματος [43] ο οποίος καλύπτει τα τυπικά περιοχόμενα τέτοιων ιστοχώρων.

## **10.1 Χαρακτηριστικά Ιστοχώρων Τουριστικού Προορισμού**

Οι υπάρχοντες ιστοχώροι τουρισμού έχουν κοινά χαρακτηριστικά και προσφέρουν όμοιες λειτουργικότητας οι οποίοι περιγράφονται συνοπτικά παρακάτω.

### **10.1.1 Ταξινόμηση Δεδομένων**

Θα πρέπει να έχει γίνει κατανοητό ότι ιστοχώροι τουρισμού διαχειρίζονται μεγάλες ποσότητες δεδομένων, γεγονός που καθιστά αναγκαία την ταξινόμηση των δεδομένων.

Συνήθως χρησιμοποιούνται ιεραρχίες κατηγοριών που διευκολύνουν τόσο τους τελικούς χρήστες όσο και τους κατασκευαστές των ιστοχώρων.

### **10.1.2 Πολυγλωσσικότητα**

Τα περισσότερα συστήματα είτε απευθύνονται σε πληθυσμό που έχει τα αγγλικά ως μητρική γλώσσα, είτε απλώς θεωρούν ότι οι επισκέπτες τους μιλάνε αγγλικά. Στην πράξη όμως η υποστήριξη της μητρικής γλώσσας του κάθε επισκέπτη είναι θεμιτό ενώ παρατηρείται ότι κάτοικοι χωρών που έχουν συνηθίσει στη χρήση του Διαδικτύου στη μητρική τους γλώσσα δύσκολα εγκαταλείπουν αυτή τη συνήθεια.

### **10.1.3 Υπερκείμενο**

Τα έγγραφα υπερκείμενου είναι βασικό χαρακτηριστικό του Παγκόσμιου Ιστού και θεωρείται ότι οι χρήστες είναι εξοικειωμένοι με τη λειτουργία του.

#### **10.1.4 Αναζήτηση**

Υπηρεσίες αναζήτησης πληροφορίας θεωρούνται από τις πιο σημαντικές διότι οι χρήστες ποτέ δε θα έχουν το χρόνο να “ξεφυλλίσουν” όλο το περιεχόμενο του ιστοχώρου για να βρουν αυτό που θέλουν.

Η αναζήτηση συνήθως περιλαμβάνει δυαδικά κριτήρια, αλλά δεν είναι σπάνια η προσφορά λειτουργικότητας για αναζήτηση σχετικότητας (similarity search).

#### **10.1.5 Κρατήσεις και Πληρωμές**

Online κρατήσεις περιλαμβάνουν κρατήσεις για διαμονή, μεταφορές ή ακόμα και εισιτηρίων για εκδηλώσεις κλπ. Η κράτηση μπορεί να γίνει άμεσα από τον ιστοχώρο με την εξέταση διαθεσιμότητας, ή έμμεσα εάν ο ιστοχώρος προσφέρει τρόπους ηλεκτρονικής επικοινωνίας (για παράδειγμα ηλεκτρονικό ταχυδρομείο).

Η πληρωμή είναι ένα πολύ πιο πολύπλοκο και δύσκολο θέμα, διότι υπάρχουν περιορισμοί από τις υποδομές του κάθε συμβαλλόμενου (τράπεζες, ξενοδοχεία, χρήστες) ενώ οι χρήστες έχουν ακόμα μεγάλες επιφυλάξεις όσον αφορά την ασφάλεια της συναλλαγής. Από τη μεριά των επιχειρήσεων υπάρχει επίσης μεγάλη ανασφάλεια για την τελική πληρωμή των ηλεκτρονικά κρατηθέντων ποσών, που ενισχύεται από το γεγονός ότι για παράδειγμα σε πιστωτικές κάρτες οι περισσότερες απάτες γίνονται σε ηλεκτρονικές πληρωμές.

#### **10.1.6 Ψηφιακοί Χάρτες**

Για έναν άγνωστο προορισμό οι χάρτες αποτελούν ένα από τα βασικότερα εργαλεία προσανατολισμού και εύρεσης πληροφορίας. Ως τέτοιοι οι χάρτες αποτελούν μία από τις σημαντικότερες υπηρεσίες που προσφέρονται σε ιστοχώρους για συγκεκριμένους τουριστικούς προορισμούς.

Ψηφιακοί χάρτες είναι μία άμεση μεταφορά ενός οικείου για τους χρήστες μέσου στο Διαδίκτυο. Χάρτες χρησιμοποιούνται κατά κόρον στα ταξίδια των χρηστών και η προσφορά ανάλογης και ενισχυμένης λειτουργικότητας στον ιστοχώρο εξυπηρετεί από πολλές απόψεις. Ο ψηφιακός χάρτης όμως μπορεί να προσφέρει πολύ περισσότερα από μία στατική αναπαράσταση πληροφορίας και συνδυάζεται με λειτουργικότητα όπως ερωτήσεις εγγύτητας.

#### **10.1.7 Εποχικότητα**

Σε αντίθεση με άλλου ιστοχώρους, ιστοχώροι τουρισμού παρουσιάζουν μία έντονη εποχικότητα σε αναλογία με τις τουριστικές περιόδους της περιοχής που αναφέρονται.



Ο Έλληνας χρήστης που θα επισκεφτεί έναν ιστοχώρο τουρισμού για την Κρήτη το Νοέμβριο, όταν έχει λήξει η τουριστική περίοδος, είναι πολύ διαφορετικός από αυτόν που θα τον επισκεφτεί την πρώτη εβδομάδα του Αυγούστου όταν οι Έλληνες προετοιμάζονται για τις εγχώριες διακοπές τους.

## 10.2 Ιστοχώροι Τουριστικού Προορισμού

Ο τουρισμός είναι από τους κυρίαρχους τομείς στο Διαδίκτυο από άποψη όγκου συναλλαγών και έτσι το εύρος των επιχειρήσεων και οργανισμών που δραστηριοποιούνται στον τομέα είναι πολύ μεγάλο. Σε αυτό περιλαμβάνονται μεγάλοι διεθνείς ιστοχώροι με ποικιλία θεματικών περιοχών που περιλαμβάνουν και τμήματα τουρισμού (π.χ. Expedia ως τμήμα του δικτύου MSN της Microsoft <http://expedia.msn.com>, Yahoo Travel ως τμήμα του Yahoo <http://travel.yahoo.com/>) αλλά και πολλοί εθνικοί και τοπικοί ιστοχώροι που περιορίζονται σε συγκεκριμένες περιοχές.

Σε αυτή την ενότητα θα δοθούν σύντομες περιγραφές για ορισμένους αντιπροσωπευτικούς ιστοχώρους τουριστικού προορισμού.

### 10.2.1 [www.austria-tourism.at](http://www.austria-tourism.at)



Πρόκειται για τον επίσημο ιστοχώρο του ημιδημόσιου Εθνικού Γραφείου Τουρισμού της Αυστρίας με σκοπό την προώθηση της εθνικής βιομηχανίας τουρισμού στον Παγκόσμιο Ιστό. Προσφέρει ενότητες για “πόλεις και περιφέρειες” με περιοχές ενδιαφέροντος που ακολουθούν τη διοικητική ιεραρχία της Αυστρίας, πληροφορία για πολιτιστικά θέματα, διαμονή, τοπικά νέα, γενική πληροφόρηση για την Αυστρία ως προορισμό τουρισμού

καθώς και θεματικές ενότητες ανάλογα με την χρονική περίοδο ενός ταξιδιού όπως “ταξίδι το καλοκαίρι”, “ταξίδι το χειμώνα” κλπ.

Οι κύριες γλώσσες του περιεχομένου είναι τα γερμανικά και αγγλικά ενώ υπάρχουν εκδόσεις του ιστοχώρου που καλύπτουν συνολικά 22 γλώσσες (για παράδειγμα ρώσικα, γιαπωνέζικα και κινέζικα).

### 10.2.2 [www.visit-sweden.com](http://www.visit-sweden.com)

Ο ιστοχώρος αναφέρεται στη Σουηδία ως τουριστικό προορισμό και ανήκει στον ημιδημόσιο οργανισμό “Swedish Travel & Tourism Council”. Καλύπτει ολόκληρη τη Σουηδία η οποία χωρίζεται σε 5 γεωγραφικές περιοχές που περιέχουν περιοχές ενδιαφέροντος και 3 περιοχές ενδιαφέροντος που αναφέρονται σε μεγάλες πόλεις της Σουηδίας. Ο ιστοχώρος παρέχει μεγάλο



πλήθος πληροφορίας το οποίο είναι χωρισμένο σε ενότητες για πιθανές δραστηριότητες, εύρεση διαμονής, πληροφορία για περιοχές ενδιαφέροντος και μεταφορές, ενώ παρέχει και γενικότερη πληροφορία όπως ψηφιακοί χάρτες, στατιστικά στοιχεία της χώρας και οδηγό εκθέσεων.

Το περιεχόμενο υπάρχει σε τρεις κύριες γλώσσες που είναι τα αγγλικά, γερμανικά και σουηδικά ενώ υπάρχουν ειδικές ενότητες σε 16 γλώσσες ανάλογα με τη γεωγραφική περιοχή ή τη χώρα

προέλευσης των πιθανών επισκεπτών.

### 10.2.3 [www.tourspain.es](http://www.tourspain.es)



Ο ιστοχώρος ανήκει στο “Instituto de Turismo de España, TURESPAÑA” και καλύπτει την Ισπανία ως τουριστικό προορισμό. Οι κύριες θεματικές ενότητες καλύπτουν περιοχές ενδιαφέροντος που είναι οργανωμένες ως ευρύτερες περιοχές σημαντικών πόλεων και περιφερειών, εκδηλώσεις, διατροφή και διαδρομές στην Ισπανία, καθώς και υπηρεσίες εύρεσης και παρουσίασης διαμονής και μεταφορών. Παρέχει επίσης ειδική λειτουργία

για τη “σύνθεση” διαδρομών από τον επισκέπτη καθώς και γενικότερες πληροφορίες για τους επισκέπτες της χώρας. Η γλώσσες που υποστηρίζει ο ιστοχώρος είναι είναι ισπανικά και αγγλικά.

### 10.2.4 [www.reiseland-brandenburg.de](http://www.reiseland-brandenburg.de)



Ο ιστοχώρος καλύπτει το ομόσπονδο κρατίδιο του Βραδενμβούργου της Γερμανίας και λειτουργεί από το δημόσιο φορέα “TMB Tourismus-Marketing Brandenburg GmbH”. Καλύπτει πολύ μεγάλη θεματολογία γύρω από το ταξίδι στο κρατίδιο με όλες τις συνηθισμένες υπηρεσίες ενός ιστοχώρου τουριστικού προορισμού. Πέραν της κλασικής αναζήτησης σε δομημένη πληροφορία παρέχει επίσης και αναζήτηση

στο πλήρες κείμενο.

Η κύρια γλώσσα του ιστοχώρου είναι τα γερμανικά ενώ υπάρχει επίσης περιορισμένη σε εύρος πληροφορία στα αγγλικά. Ο ιστοχώρος βραβεύτηκε το 2002 ως καλύτερος ιστοχώρος τουρισμού στην κατηγορία “πλατφόρμα τουρισμού” με ιδιαίτερο χαρακτηριστικό τη δυνατότητα ηλεκτρονικής κράτησης. Η βράβευση έγινε από τη “Γερμανική Ένωση Τουρισμού”.

### 10.2.5 [www.hamburg-tourism.de](http://www.hamburg-tourism.de)

Ο ιστοχώρος συνδυάζει ένα πληροφοριακό σύστημα για μία πόλη με ένα σύστημα τουριστικού προορισμού για την πόλη του Αμβούργου (που ταυτόχρονα είναι και ομόσπονδο κρατίδιο της Γερμανίας) και λειτουργεί από το δημόσιο φορέα “Hamburg Tourismus GmbH”. Παρέχει τόσο



πληροφορίες για εκδηλώσεις και δραστηριότητες της πόλης όπως και πλήρες πληροφορίες σαν ιστοχώρος τουρισμού με υπηρεσίες για εύρεση διαμονής, αξιοθέατα, επαγγελματικά ταξίδια κλπ.

Οι κύριες γλώσσες του ιστοχώρου είναι τα γερμανικά και αγγλικά ενώ καλύπτει επίσης (σε μικρότερο αλλά εκτενή εύρος) άλλες 21 γλώσσες. Ο ιστοχώρος βραβεύτηκε το 2002 και το 2003 ως καλύτερος ιστοχώρος τουρισμού στην κατηγορία “μεγάλες πόλεις”. Η βράβευση έγινε από τη “Γερμανική Ένωση Τουρισμού”.

### 10.2.6 [www.travelscotland.co.uk](http://www.travelscotland.co.uk)



Πρόκειται για έναν ιδιωτικό ιστοχώρο που καλύπτει την περιοχή της Σκωτίας. Παρέχει έναν ηλεκτρονικό οδηγό, υπηρεσίες αναζήτησης όλων των ειδών επιχειρήσεων τουρισμού καθώς και ειδικές ενότητες για σημαντικές πόλεις της Σκωτίας με πληροφορίες για εκδηλώσεις και δραστηριότητες. Επιτρέπει την ηλεκτρονική κράτηση διαμονής. Το περιεχόμενο είναι μόνο στην αγγλική

γλώσσα.

### 10.2.7 [www.visitsouthwest.co.uk](http://www.visitsouthwest.co.uk)



Πρόκειται για επίσημο ιστοχώρο δημόσιου φορέα που καλύπτει τη νοτιοδυτική Αγγλία. Η κύρια κατηγορία επιχειρήσεων τουρισμού είναι διαμονή ενώ παρέχει επίσης πληροφόρηση γενικότερα για την περιοχή ενώ έχει ειδικές ενότητες για δραστηριότητες όπως για παράδειγμα

αναζήτηση απολιθωμάτων. Το περιεχόμενο παρέχεται μόνο στην αγγλική γλώσσα.

## 10.2.8agn.hol.gr



Είναι ένας ιστοχώρος που καλύπτει την Ελλάδα ως τουριστικό προορισμό και λειτουργεί από ιδιωτικό φορέα. Καλύπτει δύο κατηγορίες επιχειρήσεων τουρισμού που είναι διαμονή και ταξιδιωτικά γραφεία με δυνατότητες αναζήτησης ανά περιοχή ενδιαφέροντος, ενώ παρέχει επίσης απλούς ψηφιακούς χάρτες και

χρήσιμες πληροφορίες για τους επισκέπτες. Παρέχει επίσης δυνατότητα ηλεκτρονικών κρατήσεων. Ο ιστοχώρος παρέχει την πληροφορία μόνο στα αγγλικά.

## 10.3 Πιλοτική Ανάπτυξη Ιστοχώρου Τουριστικού Προορισμού

Ο ιστοχώρος Crete TOURnet ήταν το αποτέλεσμα του Ευρωπαϊκού ερευνητικού προγράμματος “Integrated and Complete Support for the Tourism-Related SMEs of the Region of Crete through Interactive Media and Networks- Cretan TOURnet” (TEN-ISDN αριθμός αναφοράς 45549) με κύριο συμβαλλόμενο την Περιφέρεια Κρήτης (Prime Contractor) ενώ ο σχεδιασμός και η ανάπτυξη έγινε από την εταιρία Multimedia Systems Center. Στο πρόγραμμα συμμετείχαν επίσης το Πολυτεχνείο Κρήτης, το Εμποροβιομηχανικό Επιμελητήριο Χανίων, ο Δήμος Χανίων και άλλοι ιδιωτικοί και δημόσιοι φορείς. Σκοπός του προγράμματος ήταν η υποστήριξη μικρομεσαίων επιχειρήσεων τουρισμού της Περιφέρειας Κρήτης και η προώθηση του τουριστικού προϊόντος δια μέσω του Διαδικτύου.

Το Crete TOURnet εντάσσονταν σε ένα ευρύτερο δίκτυο προγραμμάτων με περισσότερες από 40 περιφέρειες της Ευρώπης (IntouriSME) κατά την ανάπτυξη του οποίου λειτούργησε σαν πιλοτικό σύστημα για δοκιμές και βελτιώσεις πριν τη διανομή στα υπόλοιπα συστήματα του δικτύου.

Η εξέλιξη του ιστοχώρου που αναπτύχθηκε μπορεί να χωριστεί σε 5 στάδια:

1. **Αρχικός σχεδιασμός και συλλογή πληροφοριών:** στο στάδιο αυτό έγιναν τα πρώτα βήματα σε περιοχές όπως τον καθορισμό των ενοτήτων του ιστοχώρου, την κατηγοριοποίηση των επιχειρήσεων τουρισμού που θα παρουσιάζονται, την ανάπτυξη αναλυτικών φορμών προς συμπλήρωση από τις επιχειρήσεις, καθορισμό τεχνικών προδιαγραφών κλπ. Ο ιστοχώρος δεν έχει τεθεί ακόμα σε λειτουργία.
2. **Ανάπτυξη:** ολοκλήρωση του καθορισμού προδιαγραφών και έναρξη της ανάπτυξης του ιστοχώρου. Ο ιστοχώρος αρχίζει να εμπλουτίζεται διαρκώς με πληροφορίες αλλά βρίσκεται ακόμα σε πρώιμο στάδιο
3. **Ολοκλήρωση:** ολοκληρώνεται η λειτουργικότητα του ιστοχώρου και γίνεται εντατική εργασία

με σκοπό τη διασύνδεση με το δίκτυο αντίστοιχων συστημάτων στα πλαίσια του IntouriSME. Η συλλογή πληροφοριών για τις επιχειρήσεις τουρισμού συνεχίζεται διαρκώς. Η ανάπτυξη βρίσκεται πάντα στα πλαίσια της εκπλήρωσης των απαιτήσεων του προγράμματος. Το πρόγραμμα λήγει με επιτυχία στα τέλη του 2000.

4. **Πιλοτική λειτουργία:** ο ιστοχώρος συνεχίζει να λειτουργεί πιλοτικά. Αρχίζει η ανάπτυξη του συστήματος για την καταγραφή της χρήσης από τους επισκέπτες που θα περιγραφεί παρακάτω, η οποία ολοκληρώνεται στα τέλη του 2002.
5. **Επανασχεδιασμός και νέα υλοποίηση:** Ο ιστοχώρος δέχεται ριζική αλλαγή στο σχεδιασμό και εμφανίζεται με νέο “πρόσωπο” στις αρχές του 2003 για να ανταποκριθεί στις νέες εξελίξεις της τεχνολογίας και να αντιμετωπιστούν προβλήματα λειτουργικότητας και χρησιμοποίησιμότητας.

Ο ιστοχώρος του Crete TOURnet, με το πέρας του προγράμματος (2000) προσέφερε εκτενή πληροφόρηση για πιθανούς επισκέπτες του νησιού, τόσο από την άποψη των υπηρεσιών τουρισμού όσο και σχετικά με γενικότερα ενδιαφέροντα που μπορεί κανείς να έχει για την Κρήτη. Το μεγαλύτερο μέρος της πληροφορίας παρουσιαζόταν σε 6 γλώσσες (ελληνικά, αγγλικά, γερμανικά, ιταλικά, γαλλικά και νορβηγικά) ενώ οι πληροφορίες για τις επιχειρήσεις είχαν επί το πλείστον ελληνικές και αγγλικές περιγραφές. Οι τομείς που καλύπτονται από τον ιστοχώρο περιγράφονται εν συντομία παρακάτω. Στην αλληλεπίδραση με τον ιστοχώρο και τη διασύνδεση των σελίδων θα γίνει ξεχωριστή αναφορά στο κεφάλαιο για την ανάλυση χρησιμοποίησιμότητας.

Για περισσότερες και πιο αναλυτικές πληροφορίες για το Crete TOURnet ο αναγνώστης παραπέμπεται στο σχετικό τεχνικό δελτίο του προγράμματος και τις αναφορές προς την Ευρωπαϊκή Ένωση. Στο Παράρτημα Β – Στιγμιότυπα από τον Ιστοχώρο Τουριστικού Προορισμού παρατίθενται εικόνες- στιγμιότυπα (screen shots) του ιστοχώρου..

### **10.3.1 Ηλεκτρονικός Οδηγός**

Η ενότητα του ηλεκτρονικού οδηγού καλύπτει ένα ευρύ φάσμα θεμάτων γύρω από την Κρήτη και είναι οργανωμένη σε μορφή βιβλίου. Τα 14 συνολικά κεφάλαια καλύπτουν αναλυτικά ενότητες όπως τη Γεωγραφία της Κρήτης, τη Μυθολογία και Ιστορία και θέματα πολιτισμού ενώ υπάρχουν επίσης ενότητες όπως οι εκκλησίες και κατασκευές.

Όλες οι ενότητες περιέχουν αναλυτικό κείμενο και είναι πλαισιωμένες από χαρακτηριστικές για το κάθε θέμα φωτογραφίες.

Στην Εικόνα 10.1 φαίνεται μία χαρακτηριστική σελίδα από τον Ηλεκτρονικό Οδηγό



Εικόνα 10.1 Χαρακτηριστική σελίδα του Ηλεκτρονικού Οδηγού του Crete TOURnet όπως αυτή φαινόταν στο τέλος του προγράμματος. Απόσπασμα από την ενότητα "Χλωρίδα"

### 10.3.2 Περιοχές Ενδιαφέροντος και Αξιοθέατα

Σε αυτή την ενότητα παρουσιάζονται πληροφορίες για περιοχές ενδιαφέροντος και αξιοθέατα της Κρήτης. Η οργάνωση της πληροφορίας ακολουθεί ιεραρχική μορφή όπου κάθε περιοχή μπορεί να περιέχει αξιοθέατα ή άλλες περιοχές ενδιαφέροντος. Τα αξιοθέατα ακολουθούν μία αναλυτική κατηγοριοποίηση. Η έννοια της Περιοχής Ενδιαφέροντος είναι βασική για ολόκληρο τον ιστοχώρο, διότι σε αυτές αντιστοιχίζονται σχεδόν όλες οι υπόλοιπες πληροφορίες που παρουσιάζονται σε αυτόν (επιχειρήσεις, εκδηλώσεις κλπ.). Από τη μεριά του οι περιοχές ενδιαφέροντος είναι αντιστοιχισμένες στους οργανισμούς τοπικής αυτοδιοίκησης.

Συνολικά έχει περιληφθεί πληροφορία για περισσότερες από 350 περιοχές και 900 αξιοθέατα, ενώ κάθε αξιοθέατο πλαισιώνεται από ένα σύνολο φωτογραφιών, των οποίων το πλήθος ξεπερνάει τις 3000.

Στον Πίνακα 10.1 φαίνεται ο αριθμός των περιοχών ανά νομό της Κρήτης ενώ στον πίνακα ο αριθμός των αξιοθεάτων ανά κατηγορία πρώτου επιπέδου.

<i>Νομός</i>	<i>Πλήθος Περιοχών</i>
Χανίων	114
Ηρακλείου	105
Λασιθίου	63
Ρεθύμνης	89

*Πίνακας 10.1 Πλήθος Περιοχών Ενδιαφέροντος ανά Νομό της Κρήτης*

<i>Κατηγορία Αξιοθεάτου</i>	<i>Πλήθος Αξιοθεάτων</i>
Αρχαία Ελληνικά	41
Βυζαντινά	380
Γεωγραφικοί Σχηματισμοί	111
Μινωικά	165
Μουσεία	16
Φυσικά	30
Νεότερα	20
Ρωμαϊκά	27
Τούρκικα	27
Πόλεων	96
Ενετικά	108
Θέα	95

*Πίνακας 10.2 Πλήθος αξιοθεάτων ανά κατηγορία πρώτου επιπέδου*

Για χαρακτηριστικές εικόνες από τον ιστοχώρο βλέπε Εικόνα 11.3, Εικόνα 11.2

### **10.3.3 Λιμάνια και Αγκυροβόλια**

Η ενότητα για τα λιμάνια και αγκυροβόλια καλύπτει τις ανάγκες για αυτούς που επιθυμούν την προσέγγιση του νησιού από τη θάλασσα, αλλά επίσης για γενικότερου ενδιαφέροντος πληροφορίες για τα λιμάνια. Παρουσιάζονται συνολικά περίπου 30 λιμάνια και αγκυροβόλια με αναλυτικές περιγραφές, φωτογραφίες της περιοχής, βυθομετρικούς χάρτες κ.α.

Σχετικές εικόνες από τον ιστοχώρο Εικόνα 11.4

### 10.3.4 Επιχειρήσεις Τουρισμού

Όπως αναφέρθηκε, ο κύριος σκοπός του προγράμματος ήταν η υποστήριξη μικρομεσαίων επιχειρήσεων τουρισμού στο Διαδίκτυο. Για αυτό το σκοπό εισήχθησαν και παρουσιάζονταν στον ιστοχώρο ένα μεγάλο πλήθος επιχειρήσεων του τουριστικού τομέα της Κρήτης.

Παρόλο που το μεγαλύτερο αριθμητικά μέρος των επιχειρήσεων παρουσιάζονταν με την απλή αναφορά του τίτλου, στοιχείων επικοινωνίας και τοποθεσίας που προήλθαν από τη συνεργασία με το εμποροβιομηχανικό επιμελητήριο και την ένωση ενοικιαζόμενων διαμερισμάτων, στα πλαίσια του προγράμματος επιτεύχθηκε η συλλογή και παρουσίαση μεγάλου πλήθους επιχειρήσεων με αναλυτικές πληροφορίες και φωτογραφίες που προήλθαν από τη συμπλήρωση σχετικών ερωτηματολογίων από τις ίδιες τις επιχειρήσεις.

Για όσες επιχειρήσεις ήταν διαθέσιμος ηλεκτρονικός τρόπος επικοινωνίας (email, ιστοχώρος των ιδίων των επιχειρήσεων) παρέχονταν είτε σχετικός σύνδεσμος είτε ηλεκτρονική φόρμα επικοινωνίας.

Ο Πίνακας 10.3 παρουσιάζει το πλήθος των επιχειρήσεων που ήταν διαθέσιμα από τον ιστοχώρο ανά κατηγορία.

<i>Κατηγορία Επιχειρήσεων</i>	<i>Σύνολο Επιχειρήσεων</i>	<i>Αναλυτική παρουσίαση</i>
Ξενοδοχεία	1665	277
Άλλοι τύποι διαμονής	981	368
Γραφεία τουρισμού	467	42
Διατροφή	888	65
Ενοικιάσεις οχημάτων	275	40
Παραδοσιακά προϊόντα	351	58
Διασκέδαση	62	7
Οργανισμοί	366	51
Μεταφορές	36	0
<b>Σύνολα</b>	<b>5091</b>	<b>908</b>

*Πίνακας 10.3 Πλήθος επιχειρήσεων τουρισμού ανά κατηγορία και είδος παρουσίασης όπως παρουσιάζονταν στον ιστοχώρο στο τέλος του προγράμματος*



### **10.3.5 Εκδηλώσεις**

Η ενότητα των εκδηλώσεων επιτρέπει την παρουσίαση πληροφορίας για κάθε είδους εκδηλώσεων που διεξάγονται στην Κρήτη. Αυτές περιλαμβάνουν τόσο εκδηλώσεις που γίνονται άπαξ όσο και περιοδικές εκδηλώσεις που διεξάγονται είτε σε ορισμένη ημερομηνία (εορτές κλπ.) είτε σε ορισμένη χρονική περίοδο (πανηγύρια κλπ.).

### **10.3.6 Γλωσσικός Οδηγός**


Ο γλωσσικός οδηγός περιλαμβάνει κοινές φράσεις και προτάσεις που είναι χρήσιμες για τον ξενόγλωσσο επισκέπτη του νησιού αλλά και γενικότερα της Ελλάδας. Οι μεταφράσεις παρουσιάζονται για τις 6 γλώσσες που υποστηρίζει ο ιστοχώρος ενώ για κάθε φράση εμφανίζεται η μετάφραση στα ελληνικά μαζί με τη φωνητική αναπαράσταση. Για κάθε φράση παρέχεται επίσης αρχείο ψηφιακού ήχου για την προφορά της.

### **10.3.7 Αναζήτηση Πληροφορίας**

Η αναζήτηση πληροφορίας, όπως αναφέρθηκε στην προηγούμενη ενότητα περί ιστοχώρων τουριστικού προορισμού, είναι από τις κεντρικές υπηρεσίες που πρέπει να προσφέρουν τέτοιου είδους ιστοχώροι.

Η αναζήτηση στον ιστοχώρο του Crete TOURnet πριν τη ριζική αναμόρφωσή του βασιζόταν στη δομημένη πληροφορία της βάσης δεδομένων. Έτσι παρέχονταν η δυνατότητα μέσα από φόρμες HTML για αναζήτηση με βάση κάθε είδους χαρακτηριστικού των αντικειμένων που παρουσιάζονται. Αυτό περιλάμβανε για παράδειγμα τις κατηγορίες των αξιοθεάτων, τις κατηγορίες των μονάδων διαμονής, τις διευκολύνσεις που αυτά προσφέρουν κλπ.

Το μόνο πεδίο που επέτρεπε την ελεύθερη εισαγωγή κειμένου ήταν το πεδίο για το όνομα του αντικειμένου, το οποίο συνδυάζονταν με έναν ή περισσότερους δυαδικούς περιορισμούς (με check boxes) για τα υπόλοιπα χαρακτηριστικά.

Μία επίσης βασική παράμετρος στην αναζήτηση ήταν η επιλογή γεωγραφικής περιοχής με βάση του Δήμους όπως αυτοί προέκυψαν από το “Νόμο Καποδιστρια” που χώρισε την Κρήτη (και την υπόλοιπη Ελλάδα) σε ευρύτερες διοικητικές μονάδες με συγχωνεύσεις των παλαιών οργανισμών τοπικής αυτοδιοίκησης. Όπως αυτό θα φανεί στο κεφάλαιο για την ανάλυση χρησιμοποίησιμότητας, ειδικά για τον ξένο επισκέπτη, αλλά επίσης και τους Έλληνες, αυτού του είδους ο περιορισμός για την περιοχή αναζήτησης προσφέρει μικρή πραγματική βοήθεια για την εύρεση πληροφορίας για μια συγκεκριμένη περιοχή. Για χαρακτηριστικές εικόνες από τις σελίδες αναζήτησης και αποτελεσμάτων βλέπε για παράδειγμα  Εικόνα 11.5

# 11 Παράρτημα Β – Στιγμιότυπα από τον Ιστοχώρο Τουριστικού Προορισμού

Σε αυτό το παράρτημα παρουσιάζονται εικόνες-στιγμιότυπα (screenshots) από το Crete TOURnet [43] όπως ήταν διαθέσιμο κατά τη διάρκεια και μετά το πέρας του προγράμματος (τέλος 2000) καθώς και μετά την αλλαγή στο σχεδιασμό του ιστοχώρου στις αρχές του 2003.

## 11.1 Αρχικές Εικόνες



Εικόνα 11.1 Εικόνα από τις πρώτες ημέρες λειτουργίας του ιστοχώρου

## 11.2 Πιλοτική Λειτουργία



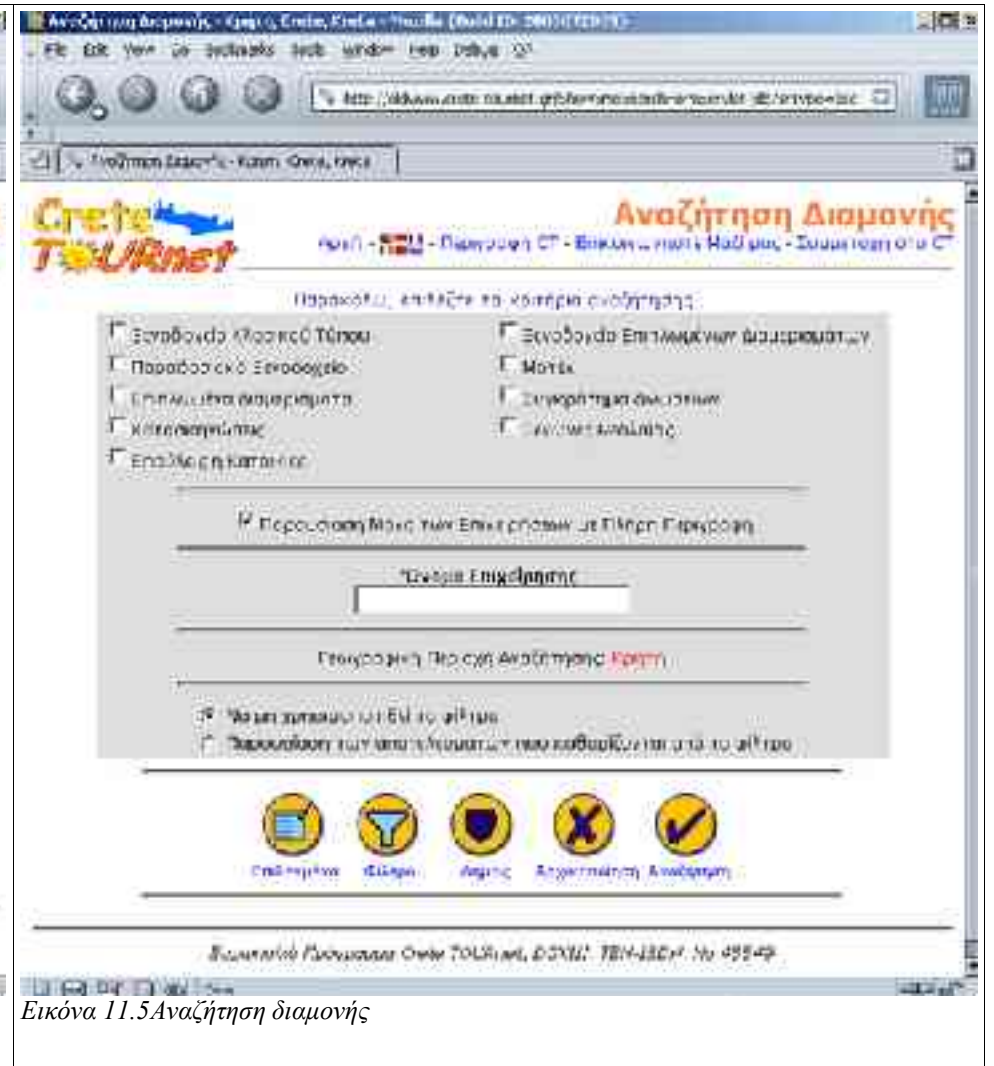
Εικόνα 11.2 Χάρτης περιοχών ενδιαφέροντος



Εικόνα 11.3 Παρουσίαση περιοχών ενδιαφέροντος

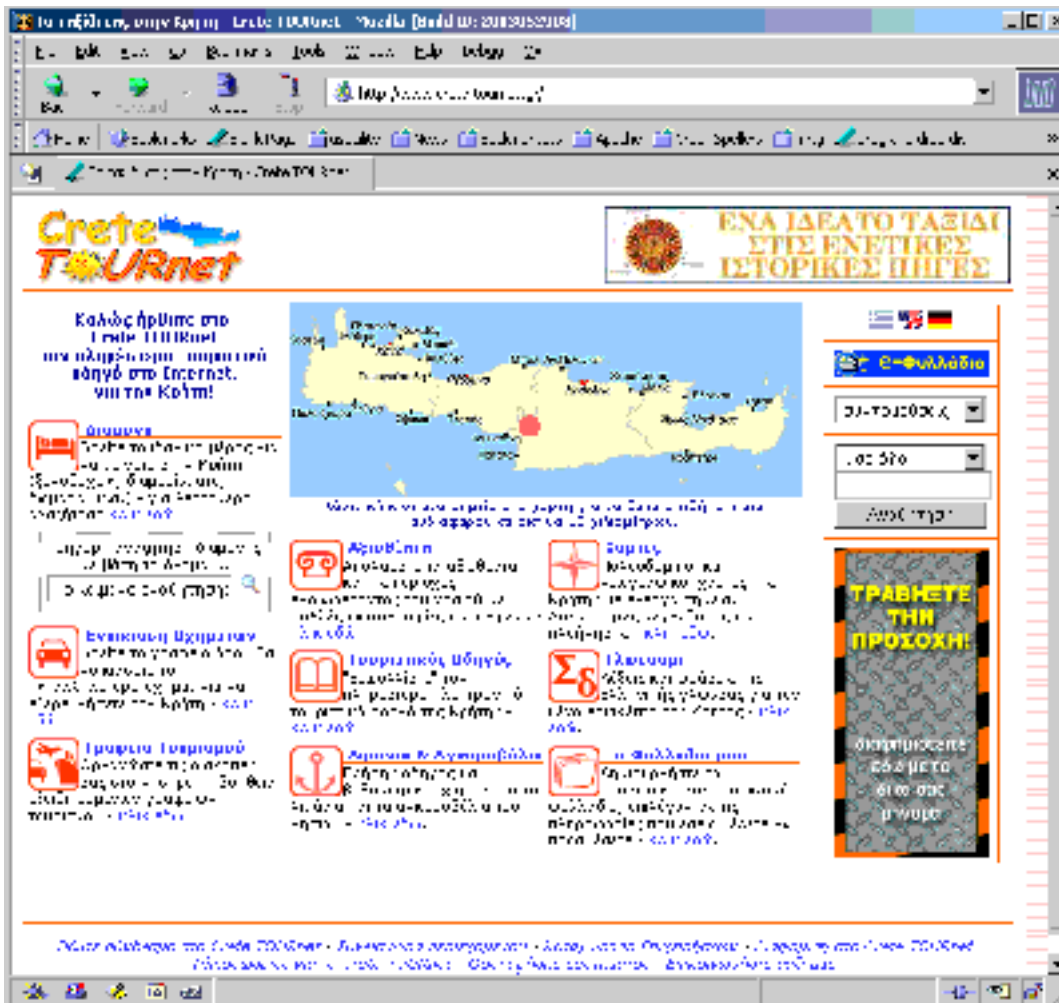


Εικόνα 11.4 Χάρτης λιμανιών και αγκυροβολιών



Εικόνα 11.5 Αναζήτηση διαμονής

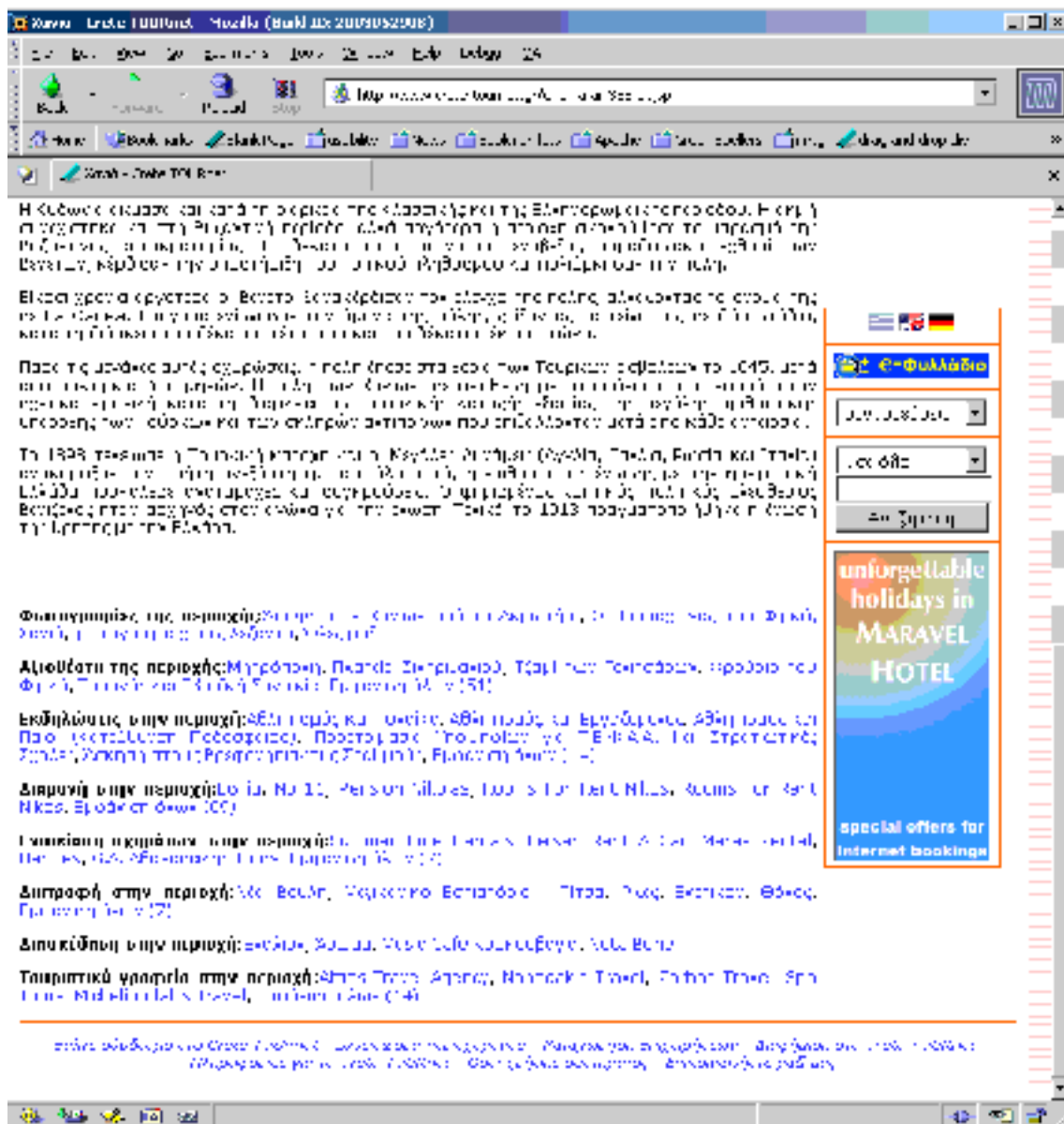
## 11.3 Αναμορφωμένη Έκδοση



Εικόνα 11.6 Αρχική σελίδα. Διακρίνονται στη δεξιά στήλη οι συντομεύσεις και η γενική αναζήτηση κειμένου. Στην αριστερή στήλη υπάρχει το πεδίο για αναζήτηση τίτλου διαμονής.



Εικόνα 11.7Η σελίδα αναζήτησης καταλυμάτων. Διακρίνονται οι βασικές επιλογές για τις διευκολύνσεις, την κατηγορία και το είδος των καταλυμάτων, καθώς και η επιλογή για επιλογή περιοχής



Εικόνα 11.8 Το κάτω μέρος της παρουσίασης περιοχής ενδιαφέροντος. Διακρίνονται οι σύνδεσμοι για σχετική πληροφορία, όπως διαμονή κλπ.

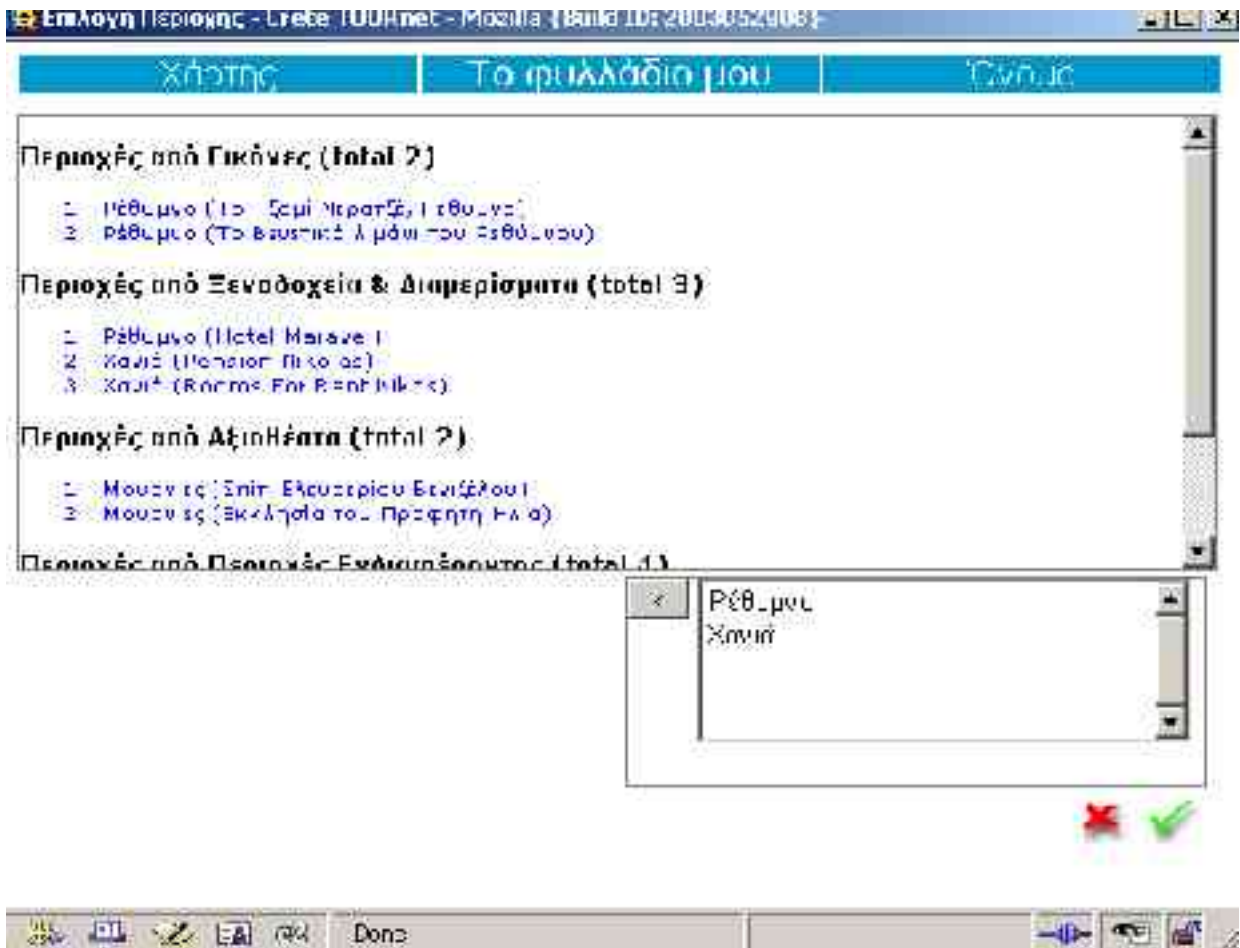


Εικόνα 11.9 Αρχική σελίδα παρουσίασης διαμονής

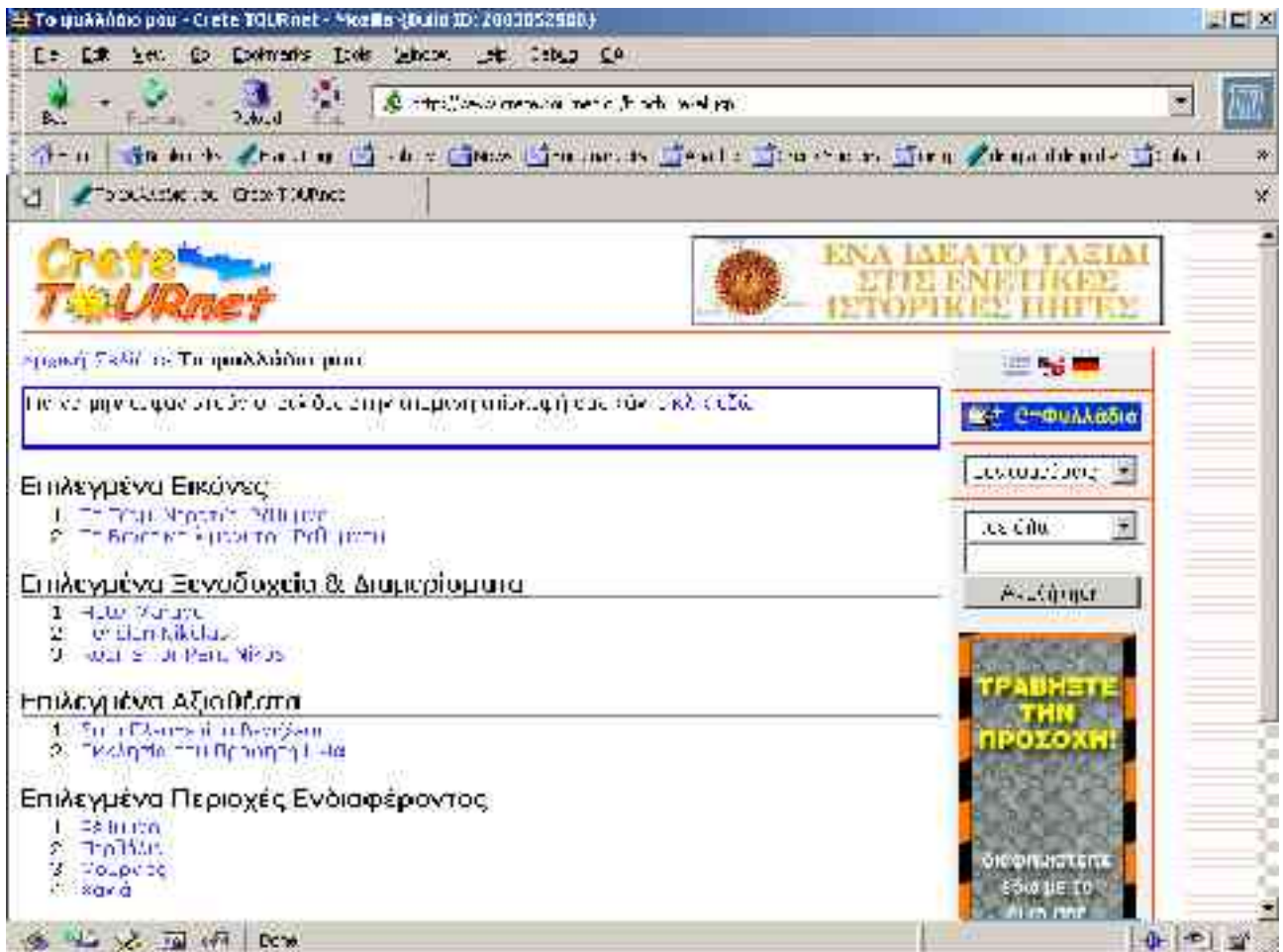




Εικόνα 11.10 Παράθυρο επιλογής περιοχής που χρησιμοποιείται από τις φόρμες αναζήτησης διαμονής, ενοικίασης οχημάτων κλπ. Στο συγκεκριμένο στιγμιότυπο έχει επιλεγθεί στο χάρτη η περιοχή Χανίων και στην κάτω αριστερή περιοχή εμφανίζονται τα ονόματα των σχετικών περιοχών. Από αυτά ο χρήστης έχει τελικά επιλέξει τα Χανιά και τις Μουριές



Εικόνα 11.11 Το παράθυρο επιλογής περιοχής μέσω των επιλεγμένων περιοχών από το φυλλάδιο



Εικόνα 11.12 Εικόνα από τη σελίδα "Το Φυλλάδιο Μου". Διακρίνονται οι επιλογές ανά κατηγορία των καταλυμάτων, των αξιοθεάτων και των περιοχών ενδιαφέροντος

Αποτελέσματα Αναζήτησης Crete TOURnet

Αρχική Σελίδα: Αποτελέσματα Αναζήτησης

**Βρέθηκαν συνολικά 220 σελίδες**

Περιοχές Ενδιαφέροντος (9)	Διασκέδαση (4)
Αξιοθέατα (16)	Επιδημιολογία (1)
Σταθμοί της Ο.Διαμερισμού (72)	Εστιατόρια (2)
Επιχειρήσεις (1)	Αεροπλάνα (1)
Ταξιδιωτικά Γραφεία (15)	Εκδόσεις (21)
Υπερήφανοι (1)	

**Αποτελέσματα για Περιοχές Ενδιαφέροντος (9)**

- Μονή Χριστοπηγής**

Θέση: Δήμος Σουδά, Χώρα Χανιά  
Βρίσκεται από τη θάλασσα και αποτελείται από τρία κτίρια που κρύβουν την ομορφιά του πέδου της πλευράς του βράχου, περίπου 2 χιλιόμετρα μετά το φαράγγι και τα αρχαία πυρρική κελιά (βλ. φωτογραφία) που κρύβονται στην κορυφή. Το κτίριο που βρίσκεται στην κορυφή είναι κεντρικό και κεντρικό. Αν κινηθείτε κυκλωρικά θα δείτε το Μοναστήρι της Κ...
- Σουδά**

Θέση: Δήμος Σουδά, Χώρα Χανιά  
Η Σουδά βρίσκεται 6 χιλιόμετρα από τα Χανιά αποτελεί ένα συνεχόμενο μυστικό με όμορφη θέα στην κατ'εξοχή περιοχή. Είναι η περιοχή της παλαιάς πόλης του 19ου αιώνα που παλαιότερα ήταν η πρωτεύουσα της Χανιάς. Η κεντρική της θέση είναι ένα από τα στρατηγικά σημαντικότερα σημεία της πόλης της Χανιάς.
- Καρυδιόφυτο**

Θέση: Δήμος Βαλτινίου, Βαλτινός, Χώρα Χανιά  
Το χωριό Καρυδιόφυτο βρίσκεται νότια των Χανίων, στα σύνορα της πόλης εκεί όπου προηγουμένως υπήρχε ο οικισμός των Καρυδιών.
- Χανιά**

Εικόνα 11.13 Απόσπασμα από τη σελίδα αποτελεσμάτων αναζήτησης στο Crete TOURnet



# 12 Παράρτημα Γ – Οδηγίες προς Αξιολογητές

Παράθεση των οδηγιών που δόθηκαν στους αξιολογητές της ευριστικής αξιολόγησης.

## Guide for the Heuristic Usability Evaluation of the Crete TOURnet Site

This document provides basic information for the evaluators of the Crete TOURnet Site. It includes a brief description of the site, some sample usage scenarios and information for the evaluation method to be used.

For any further information please contact Mr. Stefanos Karasavvidis, [stefos@ced.tuc.gr](mailto:stefos@ced.tuc.gr).

### 12.1 Purpose of Evaluation

This evaluation is performed in the frame of a master thesis at the Technical University of Crete in order to study the potential of web usage logging in finding usability problems on web sites.

For the purpose of comparing the findings of the developed method, a heuristic evaluation of a web site is conducted as part of several other “traditional” usability evaluation methods.

### 12.2 The Tourism Site

The Crete TOURnet Site (<http://www.crete.tournet.gr/>) resulted from a EU funded project aiming to promote Small & Medium Sized Enterprises in the tourism sector of Crete Island through the Internet. For this purpose the site provides information on several aspects of traveling to Crete Island.

The Site provides roughly information on the following categories:

#### 12.2.1 Crete Guide

A guide through various aspect of Crete Island, presented in the form of consecutive web pages and organized as Chapters and Sections. The subjects covered include Geography, History, Mythology, Minoan Civilization, Venetian Castles etc.

#### 12.2.2 Areas & Sites

Information for Areas of Interest on Crete Island and the Sites that can be found in those areas. Every Area can contain multiple sites. Every area and site is accompanied by images and text description and links to subjects of interest residing in the areas (hotels, car rentals etc)

### **12.2.3 Harbors & Anchorages**

This section contains information on the harbors and anchorages of Crete Island.

### **12.2.4 Tourism Companies**

The main categories of presented companies of the tourism sector are

- accommodation units
- car & bike rentals
- travel agents

Each company may be presented with a text description, photographs, and data for the provided facilities, as well as contact information.

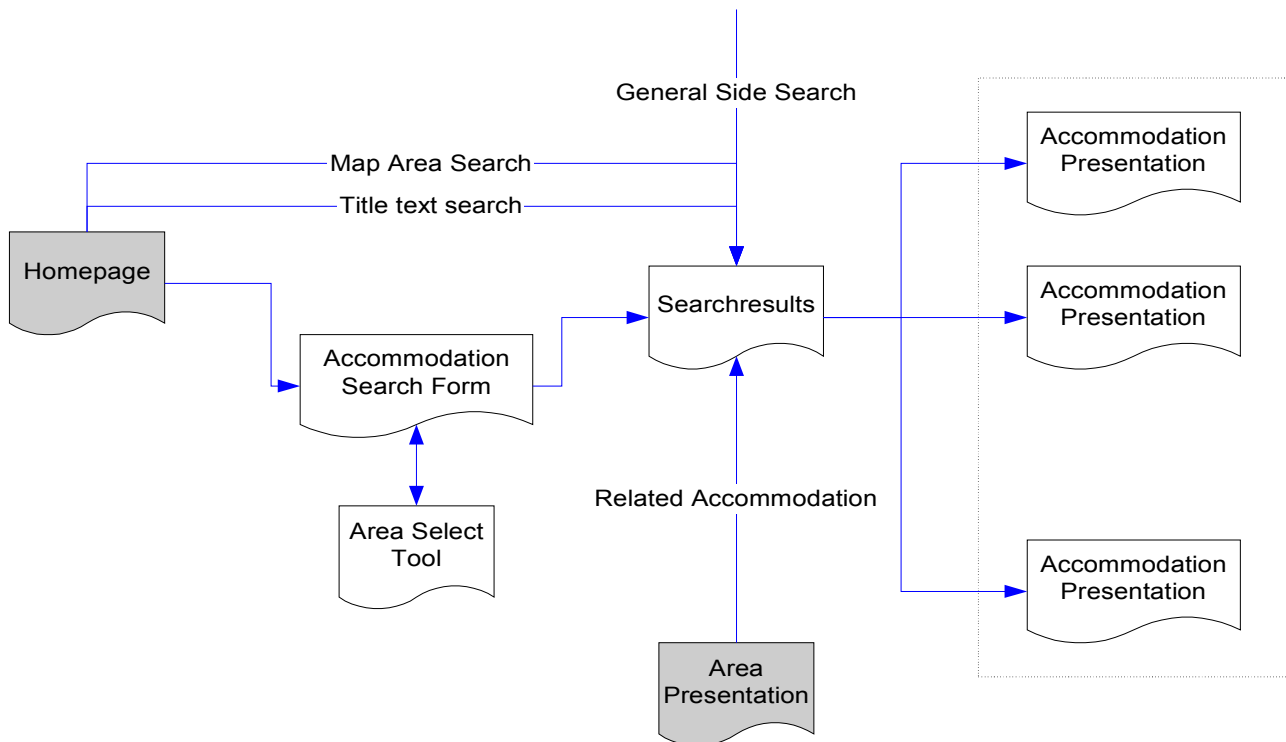
## **12.3 Usage Scenarios**

The site and the presented information are quite large in volume and complexity and can lead to a large number of possible usage scenarios. The purpose of the site focuses on the promotion of tourism companies and thus the presentation of tourism companies is considered as the final goal of the site.

All the scenarios considered as primary scenarios have a common goal to find a category of tourism company with a different starting point.

1. The user knows the name of the company for which he is trying to find information
2. The user knows the region (name, location on map) he would like to get tourism services
3. The user wants to find tourism services without having any limitations in the name or the area

For the example of finding a suitable accommodation unit the following diagram can be used as possible page sequences



Starting from the homepage or from the presentation of an area of interest, the user gets to the presentation of accommodation units, where he can find descriptions, photographs and contact information for further infos.

Similar sequences can be used for both car rentals and travel agents.

The site does also provide a “My Brochure” facility as a way to “remember” pages while browsing through the site.

As a secondary scenario, finding of information for a specific area is important, as this can be the starting point for investigating the companies within this area.

## 12.4 Evaluation Method

An informal heuristic evaluation of the site should be performed in order to identify possible usability problems. The evaluators should have in depth background on usability engineering in general and usability guidelines for the World Wide Web.

The evaluators should “work” through the site having in mind the above mentioned scenarios, but not limited to them.

The result of the evaluation should be a list of possible usability problems with short descriptions and a severity rating from 0 to 4 based on the suggestion of



<http://www.useit.com/papers/heuristic/severityrating.html>

The focus should be on identifying the usability problems and not to suggest possible solutions, although such comments will be appreciated.

As a help to evaluators not familiar with Crete Island at all, the following list of area names may be useful in conducting the evaluation: Chania, Rethimnon, Plakias, Matala, Knossos, Sougia, Paleohora, Ierapetra, Vai.

## **13 Παράρτημα Δ – Σχέδιο Πειράματος Χρηστών**

Σε αυτό το παράρτημα παρατίθεται το σχέδιο του πειράματος χρηστών (test plan) στο οποίο βασίστηκε η διενέργεια της μελέτης χρησιμοποίησιμότητας με πραγματικούς χρήστες του Crete TOURnet. Τα αποτελέσματα του πειράματος και τα σχετικά συμπεράσματα παρουσιάζονται στο κεφάλαιο 4.

### **13.1 Μελέτη Χρησιμοποίησιμότητας για τον Ιστοχώρο Τουριστικού Προορισμού**

#### **13.1.1 Σκοπός**

Αξιολόγηση χρησιμοποίησιμότητας του ιστοχώρου τουριστικού προορισμού Crete TOURnet με χρήση πραγματικών χρηστών.

#### **13.1.2 Πρόβλημα – Αντικείμενο Μελέτης**

Εξετάζεται η εφαρμογή ιστοχώρου Crete TOURnet ως προς τη χρησιμοποίησιμότητα για την εύρεση υπηρεσιών τουρισμού για τον προγραμματισμό ταξιδιού στην Κρήτη. Συγκεκριμένα επιχειρείται η διερεύνηση των ακόλουθων θεμάτων:

- είναι κατανοητά τα περιεχόμενα των ενοτήτων του ιστοχώρου όπως παρουσιάζονται στην αρχική σελίδα;
- είναι κατανοητή η λειτουργία του “φυλλαδίου μου”;
- είναι οι χρήστες ικανοί να βρουν πληροφορίες για διαμονή, ενοικίαση οχημάτων και ταξιδιωτικά γραφεία με βάση ορισμένα σημεία εκκίνησης;
- είναι οι χρήστες ικανοί να βρουν πληροφορίες για συγκεκριμένες τοποθεσίες με βάση ορισμένα σημεία εκκίνησης;
- ποια είναι η ταχύτητα εύρεσης πληροφοριών για επιχειρήσεις τουρισμού;
- είναι κατανοητή η λειτουργία των φορμών αναζήτησης για τους διάφορους τύπους επιχειρήσεων τουρισμού;
- είναι κατανοητός ο τρόπος επικοινωνίας με συγκεκριμένη επιχείρηση τουρισμού;
- ποιος είναι ο βαθμός ικανοποίησης των χρηστών από την εφαρμογή

### **13.1.3 Χρήστες**

Οι χρήστες έχουν εμπειρία στη χρήση του Παγκόσμιου Ιστού και έχουν χρησιμοποιήσει φόρμες αναζήτησης HTML. Έχουν μικρή έως καμία γνώση για τη γεωγραφική περιοχή που καλύπτει ο ιστοχώρος.

### **13.1.4 Μεθοδολογία Πειράματος**

#### **1. Χαιρετισμός και Ερωτηματολόγιο Χρηστών**

Κάθε χρήστης θα χαιρετιστεί προσωπικά από τον ειδικό πειράματος και θα επιδιωχθεί να νιώσει άνετα και χαλαρά. Θα ζητηθεί από τους χρήστες να συμπληρώσουν ένα σύντομο προκαταρκτικό ερωτηματολόγιο για το υπόβαθρό τους.

#### **2. Προσανατολισμός**

Θα γίνει μία έντυπη και προφορική εισαγωγή με σκοπό τον προσανατολισμό των χρηστών, όπου θα εξηγείται ο σκοπός και οι στόχοι του πειράματος. Ιδιαίτερα θα τους διαβεβαιωθεί ότι ο σκοπός είναι η αξιολόγηση του ιστοχώρου και όχι των ίδιων των χρηστών. Θα παρουσιαστεί στους χρήστες ο ηλεκτρονικός υπολογιστής και θα τους εξηγηθεί ότι θα καταγράφεται ο ήχος και η εικόνα της οθόνης του υπολογιστή. Οι χρήστες θα κληθούν να υπογράψουν τυπική σύμβαση αποδοχής εάν δεν το έχουν ήδη πράξει.

Ο χρήστης επιλέγει πριν την έναρξη των πειραμάτων την εφαρμογή πλοήγησης που θέλει να χρησιμοποιήσει (προσφέρονται Microsoft Internet Explorer 6, Mozilla Firebird 0.6 και Opera 7).

#### **3. Πείραμα Απόδοσης**

Το πείραμα απόδοσης αποτελείται από ένα σύνολο εργασιών που καλείται να εκτελέσει ο κάθε χρήστης ενώ αυτός παρακολουθείται. Οι εργασίες θα γίνουν με τον ακόλουθο τρόπο:

- Στους χρήστες δίνεται ένα κείμενο που περιγράφει αναλυτικά την εργασία, το οποίο και διαβάζουν. Ακολούθως προσπαθούν να εκτελέσουν την εργασία. Οι εργασίες δίνονται και εκτελούνται ξεχωριστά και αφού έχει ολοκληρωθεί η προηγούμενη. Ο χρήστης υπενθυμίζεται στις οδηγίες να “σκέφτεται δυνατά” (thinking aloud).
- Μετά την εκτέλεση των εργασιών οι χρήστες καλούνται να συμπληρώσουν ένα σύντομο ερωτηματολόγιο περάτωσης.

Κατά τη διάρκεια των εργασιών ο ειδικός σημειώνει τους χρόνους εκτέλεσης, τους αριθμούς σελιδοπροβολών και τα λάθη που γίνονται. Σημειώνει επίσης ειδικές ή ασυνήθιστες καταστάσεις

που συναντάει ο χρήστης.

#### **4. Ανασκόπηση**

Ο ειδικός κάνει ανασκόπηση με κάθε χρήστη μαγνητοφωνώντας τη συζήτηση. Οι ερωτήσεις επικεντρώνονται σε θέματα όπως:

- τη γενική αποδοχή του χρήστη με τον ιστοχώρο
- προτάσεις και σχόλια του χρήστη όσον αφορά το σχεδιασμό
- απαντήσεις του χρήστη σε ερωτήσεις από τον ειδικό για συγκεκριμένα προβλήματα και λάθη που παρατηρήθηκαν κατά τη διάρκεια του πειράματος

#### **13.1.5 Λίστα Εργασιών**

Ακολουθούν οι εργασίες που κλήθηκαν να εκτελέσουν οι χρήστες. Αναφέρεται

- ο τίτλος της εργασίας
- το σημείο εκκίνησης
- το κριτήριο επιτυχούς ολοκλήρωσης
- οι οδηγίες που δόθηκαν στους χρήστες
- ο χρόνος αναφοράς
- το βέλτιστο μονοπάτι και ο αριθμός σελιδοπροβολών (εκτός σημείου εκκίνησης)

Ο χρόνος αναφοράς υπολογίστηκε με βάση το βέλτιστο αριθμό σελιδοπροβολών. Για κάθε σελιδοπροβολή προσθέτονται 30 δευτερόλεπτα και στον τελικό αριθμό 20 δευτερόλεπτα. Ο χρόνος των 30 δευτερολέπτων υπολογίστηκε με βάση το φόρτο των σελίδων του ιστοχώρου σε μέγεθος πληροφορίας και μεμονωμένων αιτήσεων και της ταχύτητας της γραμμής σύνδεσης. Βλέπε και “περιβάλλον πειράματος”.

<i>Εργασία 1</i>	
<b>Τίτλος</b>	Αναγνώριση ενοτήτων αρχικής σελίδας
<b>Σημείο Εκκίνησης</b>	Αρχική σελίδα ιστοχώρου στην οθόνη και σε χαρτί. Κενό χαρτί για σημειώσεις
<b>Κριτήριο επιτυχούς ολοκλήρωσης</b>	Ο χρήστης σημειώνει όλες τις ενότητες στην εικόνα και περιγράφει κάθε ενότητα στο κενό χαρτί.
<b>Οδηγίες</b>	Βλέπετε την αρχική σελίδα του ιστοχώρου. Σημειώστε με κύκλους στο χαρτί τις ενότητες που μπορείτε να διακρίνετε, και για κάθε ενότητα σημειώστε το περιεχόμενο που πιστεύετε ότι υπάρχει σε κάθε μία από αυτές.
<b>Χρόνος αναφοράς</b>	Δεν εφαρμόζεται
<b>Βέλτιστος αριθμός σελιδοπροβολών</b>	Δεν εφαρμόζεται

<i>Εργασία 2</i>	
<b>Τίτλος</b>	Εύρεση διαμονής με βάση τον τίτλο
<b>Σημείο Εκκίνησης</b>	Αρχική σελίδα ιστοχώρου
<b>Κριτήριο επιτυχούς ολοκλήρωσης</b>	Ο χρήστης επιλέγει σύνδεσμο που εμφανίζει τη σελίδα παρουσίασης της ζητούμενης διαμονής
<b>Οδηγίες</b>	Ενδιαφέρεστε να βρείτε πληροφορίες για το ξενοδοχείο με όνομα “Maravel”. Έχετε βρεθεί στην αρχική σελίδα του ιστοχώρου. Προσπαθήστε να βρείτε το ξενοδοχείο.
<b>Χρόνος αναφοράς</b>	80 δεύτερα
<b>Βέλτιστος αριθμός σελιδοπροβολών</b>	2 (αναζήτηση μέσω τίτλου) αποτελέσματα αναζήτησης → παρουσίαση διαμονής

<i>Εργασία 3</i>	
<b>Τίτλος</b>	Εύρεση συγκεκριμένης πληροφορίας δωματίων για ένα ξενοδοχείο
<b>Σημείο Εκκίνησης</b>	Αρχική σελίδα ιστοχώρου
<b>Κριτήριο επιτυχούς ολοκλήρωσης</b>	Ο χρήστης αναγνωρίζει την ύπαρξη συγκεκριμένης διευκόλυνσης δωματίου ενός συγκεκριμένου ξενοδοχείου από την ενότητα “διευκολύνσεις δωματίων” που υπάρχει σε κάθε παρουσίαση
<b>Οδηγίες</b>	Έχετε κλείσει δωμάτιο στο ξενοδοχείο “Αίολος” από παραδοσιακό ταξιδιωτικό πράκτορα. Ενδιαφέρεστε να μάθετε εάν τα δωμάτια του ξενοδοχείου διαθέτουν κλιματισμό. Προσπαθήστε ξεκινώντας από την αρχική σελίδα να βρείτε πληροφορίες για το συγκεκριμένο ξενοδοχείο και να επιβεβαιώσετε εάν διαθέτουν τα δωμάτια κλιματισμό.
<b>Χρόνος αναφοράς</b>	170 δεύτερα
<b>Βέλτιστος αριθμός σελιδοπροβολών</b>	3 (αναζήτηση μέσω τίτλου) αποτελέσματα αναζήτησης → παρουσίαση διαμονής → διευκολύνσεις δωματίων

<i>Εργασία 4</i>	
<b>Τίτλος</b>	Εύρεση πληροφορίας επικοινωνίας για συγκεκριμένο ξενοδοχείο
<b>Σημείο Εκκίνησης</b>	Αρχική σελίδα ιστοχώρου
<b>Κριτήριο επιτυχούς ολοκλήρωσης</b>	Ο χρήστης φτάνει στην ενότητα επικοινωνία & τοποθεσία του συγκεκριμένου ξενοδοχείου
<b>Οδηγίες</b>	Έχετε κλείσει δωμάτιο στο ξενοδοχείο “Fodele Beach” από παραδοσιακό ταξιδιωτικό πράκτορα. Ενδιαφέρεστε να μάθετε την ακριβή απόσταση του ξενοδοχείου από κάποιο αεροδρόμιο του νησιού. Για αυτό το σκοπό θέλετε να έρθετε σε επικοινωνία με το ξενοδοχείο μέσω τηλεφώνου ή ηλεκτρονικού μηνύματος για να κάνετε σχετική ερώτηση. Ξεκινώντας από την αρχική σελίδα προσπαθήστε να βρείτε αυτή την πληροφορία.
<b>Χρόνος αναφοράς</b>	110 δεύτερα
<b>Βέλτιστος αριθμός σελιδοπροβολών</b>	3 (αναζήτηση μέσω τίτλου) αποτελέσματα αναζήτησης → παρουσίαση διαμονής → επικοινωνία & τοποθεσία

<i>Εργασία 5</i>	
<b>Τίτλος</b>	Εύρεση αναλυτικού γεωγραφικού χάρτη από την περιοχή συγκεκριμένου ξενοδοχείου
<b>Σημείο Εκκίνησης</b>	Αρχική σελίδα ιστοχώρου
<b>Κριτήριο επιτυχούς ολοκλήρωσης</b>	Ο χρήστης φτάνει στον αναλυτικό χάρτη της περιοχής του συγκεκριμένου ξενοδοχείου.
<b>Οδηγίες</b>	Θέλετε να βρείτε έναν αναλυτικό γεωγραφικό χάρτη της περιοχής του ξενοδοχείου “Sunning Dale”. Ξεκινώντας από την αρχική σελίδα, προσπαθήστε να βρείτε τον αναλυτικό χάρτη της περιοχής.
<b>Χρόνος αναφοράς</b>	140 δεύτερα
<b>Βέλτιστος αριθμός σελιδοπροβολών</b>	4 (αναζήτηση μέσω τίτλου) αποτελέσματα αναζήτησης → παρουσίαση διαμονής → επικοινωνία & τοποθεσία → (από χάρτη τοποθεσίας) χάρτης περιοχής

<i>Εργασία 6</i>	
<b>Τίτλος</b>	Εύρεση ξενοδοχείων με συγκεκριμένα κριτήρια
<b>Σημείο Εκκίνησης</b>	Αρχική σελίδα ιστοχώρου
<b>Κριτήριο επιτυχούς ολοκλήρωσης</b>	Ο χρήστης επιλέγει στη φόρμα αναζήτησης ξενοδοχείων τα σωστά κριτήρια και καταλήγει στα αποτελέσματα αναζήτησης
<b>Οδηγίες</b>	Ψάχνετε για ένα κατάλυμα για να περάσετε τις διακοπές σας. Θέλετε το κατάλυμα να είναι κοντά στη θάλασσα, να έχει κλιματισμό στα δωμάτια και να είναι τουλάχιστον Β κατηγορίας. Προσπαθήστε ξεκινώντας από την αρχική σελίδα του ιστοχώρου να βρείτε καταλύματα που καλύπτουν αυτά τα κριτήρια.
<b>Χρόνος αναφοράς</b>	80 δεύτερα
<b>Βέλτιστος αριθμός σελιδοπροβολών</b>	2 φόρμα αναζήτησης διαμονής → αποτελέσματα αναζήτησης

<i>Εργασία 7</i>	
<b>Τίτλος</b>	Εύρεση ξενοδοχείων με συγκεκριμένα κριτήρια σε συγκεκριμένη περιοχή
<b>Σημείο Εκκίνησης</b>	Αρχική σελίδα ιστοχώρου
<b>Κριτήριο επιτυχούς ολοκλήρωσης</b>	Ο χρήστης επιλέγει στη φόρμα αναζήτησης ξενοδοχείων τα σωστά κριτήρια και καταλήγει στα αποτελέσματα αναζήτησης.
<b>Οδηγίες</b>	Ψάχνετε για ένα κατάλυμα για να περάσετε τις διακοπές σας. Θέλετε το κατάλυμα να είναι κοντά στη θάλασσα, να έχει κλιματισμό στα δωμάτια και να είναι τουλάχιστον Β κατηγορίας. Επίσης θέλετε να βρίσκεται στην ευρύτερη περιοχή των Χανίων ή του Ακρωτηρίου. Προσπαθήστε ξεκινώντας από την αρχική σελίδα του ιστοχώρου να βρείτε καταλύματα που καλύπτουν αυτά τα κριτήρια.
<b>Χρόνος αναφοράς</b>	110 δεύτερα
<b>Βέλτιστος αριθμός σελιδοπροβολών</b>	3 φόρμα αναζήτησης διαμονής → παράθυρο επιλογής περιοχής → αποτελέσματα αναζήτησης

<i>Εργασία 8</i>	
<b>Τίτλος</b>	Εύρεση πληροφορίας από την περιοχή συγκεκριμένου ξενοδοχείου
<b>Σημείο Εκκίνησης</b>	Αρχική σελίδα ιστοχώρου
<b>Κριτήριο επιτυχούς ολοκλήρωσης</b>	Ο χρήστης βρίσκει το ξενοδοχείο και επιλέγει από την ενότητα Επικοινωνία/Τοποθεσία την περιοχή που βρίσκεται. Από εκεί επιλέγει τις σχετικές επιχειρήσεις ενοικίασης οχημάτων
<b>Οδηγίες</b>	Έχετε ήδη κράτηση στο κατάλυμα “Asterion”. Επιθυμείτε να βρείτε μία επιχείρηση για ενοικίαση οχημάτων στην περιοχή του ξενοδοχείου. Βρείτε το ξενοδοχείο και προσπαθήστε να πάρετε πληροφορίες για την περιοχή όπου βρίσκεται και να βρείτε ενοικιάσεις οχημάτων σε αυτήν.
<b>Χρόνος αναφοράς</b>	170 δεύτερα
<b>Βέλτιστος αριθμός σελιδοπροβολών</b>	5 (αναζήτηση μέσω τίτλου) αποτελέσματα αναζήτησης → παρουσίαση διαμονής → επικοινωνία & τοποθεσία → (από χάρτη τοποθεσίας ή όνομα περιοχής) παρουσίαση περιοχής → (σχετικές επιχειρήσεις) παρουσίαση ενοικίασης οχημάτων



<i>Εργασία 9</i>	
<b>Τίτλος</b>	Εύρεση επιχειρήσεων από τη σελίδα παρουσίασης περιοχής
<b>Σημείο Εκκίνησης</b>	Σελίδα παρουσίασης περιοχής
<b>Κριτήριο επιτυχούς ολοκλήρωσης</b>	Ο χρήστης επιλέγει “Τουριστικά γραφεία στην περιοχή
<b>Οδηγίες</b>	Ψάχνετε πληροφορίες για τα Χανιά και φτάνετε στη σελίδα παρουσίασης των Χανίων. Αποφασίζετε ότι θέλετε να περάσετε τις διακοπές σας στα Χανιά και ψάχνετε για τοπικό ταξιδιωτικό γραφείο. Προσπαθήστε ξεκινώντας από τη σελίδα παρουσίασης των Χανίων να βρείτε τοπικά ταξιδιωτικά γραφεία.
<b>Χρόνος αναφοράς</b>	50 δεύτερα
<b>Βέλτιστος αριθμός σελιδοπροβολών</b>	1 (σχετικές επιχειρήσεις) παρουσίαση ταξιδιωτικών γραφείων

<i>Εργασία 10</i>	
<b>Τίτλος</b>	Χρήση “το φυλλάδιο μου”
<b>Σημείο Εκκίνησης</b>	Αρχική σελίδα ιστοχώρου
<b>Κριτήριο επιτυχούς ολοκλήρωσης</b>	Ο χρήστης έχει επιλέξει στο φυλλάδιο τις περιοχές Χανίων και Ρεθύμνου και έναν αριθμό ξενοδοχείων. Βλέπει το φυλλάδιο με τις επιλεγμένες επιχειρήσεις
<b>Οδηγίες</b>	Θέλετε να επιλέξετε ανάμεσα σε διάφορες περιοχές και ανάμεσα σε διάφορα ξενοδοχεία ανά περιοχή τον τελικό τόπο διακοπών σας. Έχετε καταλήξει στην περιοχή Χανίων και Ρεθύμνου και σε ορισμένες επιχειρήσεις σε αυτές. Χρησιμοποιήστε το “φυλλάδιο μου” για να επιλέξετε ξενοδοχεία σε αυτές τις περιοχές. Για κάθε ξενοδοχείο που έχετε επιλέξει βρείτε τρόπους επικοινωνίας.
<b>Χρόνος αναφοράς</b>	200 δεύτερα
<b>Βέλτιστος αριθμός σελιδοπροβολών</b>	6 (πλαϊνό εικονίδιο ή ενότητα φυλλαδίου) φυλλάδιο → αρχική σελίδα → φόρμα αναζήτησης διαμονής → παράθυρο επιλογής περιοχής → αποτελέσματα αναζήτησης (επιλέγει μερικές επιχειρήσεις) → φυλλάδιο

### 13.1.5.1 Γερμανικές Μεταφράσεις των Οδηγιών

#### *Aufgabe 1*

**Anleitung** Sie sehen die Homepage der Site und eine auf Papier gedruckte Version. Welche unabhängige Sektionen können Sie erkennen? Kreisen Sie diese auf Papier ein und geben Sie für jede Stichwörter mit dem Inhalt der sich Ihrer Meinung nach darunter verbirgt.

#### *Aufgabe 2*

**Anleitung** Sie möchten Informationen zu dem Hotel „Maravel“ erhalten. Sie sind auf der Homepage der Site. Versuchen Sie eine Beschreibung des Hotels zu finden.

#### *Aufgabe 3*

**Anleitung** Sie haben bei einem traditionellen Reisebüro ein Zimmer im Hotel „Aiolos“ gebucht. Sie möchten nun wissen ob die Zimmer des Hotels über eine Klimaanlage verfügen.  
Versuchen Sie herauszufinden ob die Zimmer dieses Hotels über eine Klimaanlage verfügen.

#### *Aufgabe 4*

**Anleitung** Sie haben ein Zimmer im Hotel „Fodele Beach“ gebucht. Sie möchten nun die Entfernung des Hotels zu einem Flughafen der Insel wissen. Deshalb möchten Sie eine Telefonnummer oder eine Email des Hotels um eine entsprechende Anfrage zu machen.  
Versuchen Sie eine Telefonnummer oder eine Email des Hotels zu finden.

#### *Aufgabe 5*

**Anleitung** Sie möchten eine detaillierte Landkarte für die Umgebung des Hotels „Sunning Dale“.  
Versuchen diese Landkarte zu finden.

#### *Aufgabe 6*

**Anleitung** Sie suchen nach einem Hotel auf Kreta. Sie möchten dass das Hotel in der Nähe zu einem Strand ist, dass die Zimmer eine Klimaanlage haben und dass das Hotel mindestens von Kategorie B ist.  
Versuchen Sie Hotels mit diesen Kriterien zu finden.

### *Aufgabe 7*

**Anleitung** Sie suchen nach einem Hotel auf Kreta. Sie möchten dass das Hotel in der Nähe zu einem Strand ist, dass die Zimmer eine Klimaanlage haben und dass das Hotel mindestens von Kategorie B ist. Sie wollen dass sich das Hotel in der weiteren Umgebung von Chania befindet.

Versuchen Sie Hotels mit diesen Kriterien zu finden.

### *Aufgabe 8*

**Anleitung** Sie haben ein Zimmer im Hotel „Asterion“ gebucht. Sie möchten ein Auto von einer Autovermietung in der Umgebung des Hotels mieten.

Versuchen Sie Informationen zum Gebiet des Hotels und Autovermietungen zu finden.

### *Aufgabe 9*

**Anleitung** Sie suchen nach Informationen für Chania und sind auf die Chania Präsentationsseite gekommen. Sie entscheiden sich in Chania Ihren Urlaub zu verbringen und suchen nach lokalen Reisebüros.

Versuchen Sie von der Seite der Stadt Chania aus Reisebüros der Umgebung zu finden.

### *Aufgabe 10*

**Anleitung** Sie möchten sich für Ihren Urlaubsort und Ihrem Hotel entscheiden. Sie wissen schon vorher dass es entweder in der Umgebung von Chania oder Rethimnon sein wird.

Benutzen Sie „Meine Broschüre“ um einige Hotels in diesen Gebieten auszuwählen. Zu jedem Hotel versuchen Sie Kontaktinformationen zu bekommen.

## **13.1.6 Περιβάλλον Πειράματος**

Τα πειράματα γίνονται σε αίθουσα του Πολυτεχνείου Κρήτης όπου υπάρχει ένας ηλεκτρονικός υπολογιστής σε γραφείο, συσκευή video, τηλεόραση, μικρόφωνο και δύο καρέκλες παρατηρητών. Ο Η/Υ είναι συνδεδεμένος στο Διαδίκτυο με DialUp σύνδεση 33,6KBit/s. Πριν από κάθε χρήση

εκκενώνεται το ιστορικό της εφαρμογής πλοήγησης και η κρυφή μνήμη (cache). Η εφαρμογή πλοήγησης μπορεί να είναι είτε ο Microsoft Internet Explorer 6.0 ή Mozilla Firebird 0.6.

Οι ώρες διενέργειας των πειραμάτων αναφέρονται στις ώρες 9π.μ. με 1μ.μ. Όπου έχει διαπιστωθεί ότι ο φόρτος κίνησης του ιστοχώρου σε όρους φόρτου στους πόρους του υπολογιστικού συστήματος και φόρτου γραμμής μεταφοράς δε δημιουργεί επιπρόσθετες καθυστερήσεις.

### **13.1.7Ο Ρόλος του Ειδικού**

Ο ειδικός κάθεται μαζί με κάθε χρήστη στο δωμάτιο ενώ εκτελούνται τα πειράματα. Ο ειδικός δεν προσφέρει καμία βοήθεια ούτε απαντάει άμεσα σε ερωτήσεις των χρηστών εκτός και εάν διαπιστώσει ότι ο χρήστης έχει ξεφύγει εντελώς, και με ελάχιστες πιθανότητες επανόδου, από το μονοπάτι εκτέλεσης της εργασίας. Ο ειδικός βοηθάει σε περιπτώσεις τεχνικών προβλημάτων του εξοπλισμού.

### **13.1.8Μετρικές Αξιολόγησης**

- Επιτυχής ολοκλήρωση εργασίας
- Χρόνος αποπεράτωσης
- αριθμός σελιδοπροβολών
- Σχόλια και εκφράσεις χρηστών
- Αντίληψη χρηστών για τη χρησιμότητα της εφαρμογής

### **13.1.9Συμφωνητικό**

Κωδικός Χρήστη: # \_\_\_\_\_

Στα πλαίσια ερευνητικής διατριβής στο Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης, διεξάγονται πειράματα χρηστών εφαρμογής στο Διαδίκτυο. Ο σκοπός των πειραμάτων είναι να διερευνηθεί ο βαθμός χρησιμοποιησιμότητας (ευχρηστίας) της εφαρμογής. Σε καμία περίπτωση δεν κρίνονται οι χρήστες για την άνεσή τους ή την επιτυχία τους στη χρήση της εφαρμογής.

Πριν την έναρξη του πειράματος, παρακαλούμε να συμφωνήσετε στα εξής. Εκτός της τελευταίας επιλογής, όλες οι υπόλοιπες πρέπει να απαντηθούν με “ναι”.

όχι___	ναι___	Γνωρίζω ότι κατά τη διάρκεια του πειράματος ηχογραφείται ο ήχος και μαγνητοσκοπείται η οθόνη του υπολογιστή
όχι___	ναι___	Γνωρίζω ότι τα αποτελέσματα του πειράματος σε καμία περίπτωση δε θα συνδεθούν με το πρόσωπό μου
όχι___	ναι___	Συμφωνώ να επεξεργαστούν τα αποτελέσματα από το πείραμα για τους σκοπούς ερευνητικής διατριβής στο Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης
όχι___	ναι___	Συμφωνώ να αναφερθεί το όνομά μου ως συμμετέχοντα στο πείραμα

Όνομα και υπογραφή:

---

### **13.1.9.1 Γερμανική Μετάφραση Συμφωνητικού**

Tester Nr.: # \_\_\_\_\_

Im Rahmen einer wissenschaftlichen Arbeit im Department of Electronic & Computer Engineering der Technischen Universität Kreta, werden Benutzer Tests einer Internetapplikation durchgeführt. Ziel der Tests ist es den Grad der Benutzbarkeit (Usability) der Applikation zu bewerten. Es werden in keinem Fall die Benutzer die an den Tests teilnehmen bewertet.

Bevor dem Testbeginn bitten wir Sie zu folgenden Punkten zuzustimmen. Ausser den Letzten Punkt müssen Sie alle weiteren mit "Ja" beantworten.

Nein___	Ja___	Es ist mir bewusst dass während des Tests der Ton und das Bild des Computermonitors aufgezeichnet wird.
Nein___	Ja___	Es ist mir bewusst dass das Resultat des Tests nicht mit meiner Person in Bezug gebracht wird.
Nein___	Ja___	Ich stimme zu dass das Resultat des Tests zum Zweck der wissenschaftlichen Arbeit des Department of Electronic & Computer Engineering der Technischen Universität Kreta verwendet wird.
Nein___	Ja___	Ich stimme zu dass mein Name als Teilnehmer des Tests angegeben wird.

---

Nein ___	Ja ___	Es ist mir bewusst dass während des Tests der Ton und das Bild des Computermonitors aufgezeichnet wird.
----------	--------	---

Name und Unterschrift:

---

### **13.1.10 Ερωτηματολόγια**

#### **13.1.10.1 Προκαταρκτικό Ερωτηματολόγιο**

Πριν ξεκινήσουμε παρακαλούμε να συμπληρώσετε αυτό το σύντομο ερωτηματολόγιο. Οι απαντήσεις που θα δώσετε είναι αυστηρά εμπιστευτικές και δε θα συσχετιστούν με κανένα τρόπο με το πρόσωπό σας.

Κωδικός Χρήστη: # \_\_\_\_\_

**Ερώτηση 1:** σε ποια ηλικιακή ομάδα ανήκετε;

\_\_\_\_\_ 18-22

\_\_\_\_\_ 23-28

\_\_\_\_\_ 29-45

\_\_\_\_\_ 46-

**Ερώτηση 2:** παρακαλώ σημειώστε το φύλο σας

\_\_\_\_\_ άντρας                      \_\_\_\_\_ γυναίκα

**Ερώτηση 3:** παρακαλώ σημειώστε το επάγγελμά σας

\_\_\_\_\_

**Ερώτηση 4:** παρακαλώ σημειώστε τις γλώσσες που μιλάτε

\_\_\_\_\_ μητρική \_\_\_άριστα \_\_\_πολύ καλά \_\_\_μέτρια

\_\_\_\_\_ μητρική \_\_\_άριστα \_\_\_πολύ καλά \_\_\_μέτρια

**Ερώτηση 5:** ποια η εμπειρία σας στη χρήση του Διαδικτύου;

\_\_\_\_\_ λιγότερο από χρόνο

\_\_\_\_\_ ανάμεσα 1 και 3 χρόνια

\_\_\_\_\_ ανάμεσα 3 και 5 χρόνια

\_\_\_\_\_ περισσότερο από 5 χρόνια

**Ερώτηση 6:** πόσες ώρες την εβδομάδα «περνάτε» στο Διαδίκτυο;

\_\_\_\_\_ λιγότερο από 5 ώρες την εβδομάδα

\_\_\_\_\_ ανάμεσα 5 και 10 ώρες

\_\_\_\_\_ ανάμεσα 10 και 20 ώρες

\_\_\_\_\_ περισσότερο από 20 ώρες

**Ερώτηση 7:** ποια είναι η αγαπημένη σας ασχολία στο Διαδίκτυο;

\_\_\_\_\_ διασκέδαση

\_\_\_\_\_ εργασία

\_\_\_\_\_ κοινωνικά θέματα

\_\_\_\_\_ αγορές

\_\_\_\_\_ έρευνα

άλλο \_\_\_\_\_

**Ερώτηση 8:** έχετε ποτέ στο παρελθόν ερευνήσει στο Διαδίκτυο για την περιοχή που θέλετε να περάσετε τις διακοπές σας;

\_\_\_\_\_ ναι

\_\_\_\_\_ όχι

**Ερώτηση 9:** πως θα χαρακτηρίζατε τη γνώση σας για τις περιοχές και τοποθεσίες της Κρήτης;

\_\_\_\_\_ γνωρίζω άριστα όλες τις τοποθεσίες της Κρήτης

\_\_\_\_\_ γνωρίζω όλες τις σημαντικές τοποθεσίες της Κρήτης

\_\_\_\_\_ έχω στοιχειώδη γνώση για τις τοποθεσίες της Κρήτης

\_\_\_\_\_ δεν έχω ιδιαίτερη γνώση

#### **13.1.10.1 Γερμανική Μετάφραση Προκαταρκτικού Ερωτηματολογίου**

Bevor wir beginnen, bitten wir Sie den Folgenden Fragebogen auszufüllen. Die Antworten die Sie geben werden streng vertraulich behandelt und werden nicht mit Ihrer Person in Bezug gebracht.

Tester Nr.: # \_\_\_\_\_

**Frage 1:** in welcher Altersgruppe gehören Sie?

\_\_\_\_\_ 18-22

\_\_\_\_\_ 23-28

\_\_\_\_\_ 29-45

\_\_\_\_\_ 46-

**Frage 2:** bitte geben Sie Ihr Geschlecht an

\_\_\_\_\_ Mann

\_\_\_\_\_ Frau



**Frage 3:** bitte geben Sie Ihren Beruf an

\_\_\_\_\_

**Frage 4:** bitte geben Sie die Sprachen die Sie sprechen an

\_\_\_\_\_ \_\_\_Muttersprache \_\_\_perfekt \_\_\_sehr gut \_\_\_mittelmässig

\_\_\_\_\_ \_\_\_Muttersprache \_\_\_perfekt \_\_\_sehr gut \_\_\_mittelmässig

**Frage 5:** wie lange benutzen Sie das Internet?

\_\_\_\_\_ weniger als ein Jahr

\_\_\_\_\_ zwischen 1 und 3 Jahren

\_\_\_\_\_ zwischen 3 und 5 Jahren

\_\_\_\_\_ mehr als 5 Jahre

**Frage 6:** wie viele Stunden in der Woche "surfen" Sie im Internet?

\_\_\_\_\_ weniger als 5 Stunden

\_\_\_\_\_ zwischen 5 und 10 Stunden

\_\_\_\_\_ zwischen 10 und 20 Stunden

\_\_\_\_\_ mehr als 20 Stunden

**Frage 7:** mit was beschäftigen Sie sich hauptsächlich im Internet?

\_\_\_\_\_ Unterhaltung

\_\_\_\_\_ Arbeit

\_\_\_\_\_ Soziale Themen

\_\_\_\_\_ Shopping

\_\_\_\_\_ Informationssuche

anderes Thema \_\_\_\_\_

**Frage 8:** haben Sie in der Vergangenheit nach Informationen zu Ihrem geplanten Urlaubsort gesucht?

\_\_\_\_\_ Ja

\_\_\_\_\_ Nein

**Frage 9:** wie würden Sie Ihre Kenntnisse um die Gebiete und Sehenswürdigkeiten Kretas beschreiben?

\_\_\_\_\_ ich kenne sehr gut alle Orte Kretas

\_\_\_\_\_ ich kenne alle wichtigen Orte Kretas

\_\_\_\_\_ ich habe Basiswissen über die Orte Kretas

\_\_\_\_\_ ich habe keine speziellen Kenntnisse über Kreta

### **13.1.10.2 Ερωτηματολόγιο Περάτωσης**

Πριν το τέλος του πειράματος παρακαλούμε να συμπληρώσετε ένα τελικό ερωτηματολόγιο. Οι απαντήσεις που θα δώσετε είναι αυστηρά εμπιστευτικές και δε θα συσχετιστούν με κανένα τρόπο με το πρόσωπό σας.

Κωδικός Χρήστη: # \_\_\_\_\_

**Ερώτηση 1:** πόσο εξυπηρετική σας φάνηκε η οργάνωση της πληροφορίας στον ιστοχώρο όσον αφορά την εκτέλεση των εργασιών;

\_\_\_\_\_ πολύ εξυπηρετική

- εξυπηρετική
- μέτρια
- μη εξυπηρετική
- καθόλου εξυπηρετική

**Ερώτηση 2:** περιγράψτε τη δυσκολία της πλοήγησης στον ιστοχώρο κατά τη διάρκεια εκτέλεσης των εργασιών;

- εύκολη, δε χρειάστηκε ιδιαίτερος κόπος για την ολοκλήρωση των εργασιών
- χρειάστηκε “κανονικός” κόπος για την ολοκλήρωση των εργασιών
- χρειάστηκε πολύς κόπος για την ολοκλήρωση των εργασιών
- χρειάστηκε υπερβολικά μεγάλος κόπος για την ολοκλήρωση αυτών των εργασιών

**Ερώτηση 3:** συμφωνείτε γενικά με το σχεδιασμό του ιστοχώρου όσον αφορά την εύρεση πληροφορίας;

- ναι
- όχι

**Ερώτηση 4:** εάν όχι, τι θα προτείνατε για τη βελτίωση των σελίδων ώστε να γίνουν πιο εύχρηστες;

---

---

**Ερώτηση 5:** υπάρχει κάτι που θα θέλατε να γνωρίζουν οι σχεδιαστές του ιστοχώρου;

---

---

### **13.1.10.2.1 Γερμανική Μετάφραση Ερωτηματολογίου Περάτωσης**

Bevor wir die Tests beenden, bitten wir Sie folgenden Fragebogen auszufüllen. Die Antworten die Sie geben werden streng vertraulich behandelt und werden nicht mit Ihrer Person in Bezug gebracht.

Tester Nr.: # \_\_\_\_\_

**Frage 1:** wie hilfreich haben Sie die Organisation der Informationen der WWW-Seiten gefunden um die Aufgaben auszuführen?

\_\_\_\_\_ sehr hilfreich

\_\_\_\_\_ hilfreich

\_\_\_\_\_ mässig hilfreich

\_\_\_\_\_ weniger hilfreich

\_\_\_\_\_ nicht hilfreich

**Frage 2:** beschreiben Sie die Schwierigkeit die Sie hatten während der Ausführung der Aufgaben.

\_\_\_\_\_ leicht. Ich hatte keine nennenswerten Schwierigkeiten die Aufgaben auszuführen

\_\_\_\_\_ es war mässig schwer die Aufgaben auszuführen.

\_\_\_\_\_ es war schwer die Aufgaben auszuführen

\_\_\_\_\_ es war unheimlich schwer die Aufgaben auszuführen

**Frage 3:** wenn es auf die Informationsfindung ankommt, stimmen Sie mit dem Design und der Organisation der WWW- Seiten zu?

\_\_\_\_\_ Ja

\_\_\_\_\_ Nein

**Frage 4:** falls nein, haben Sie einen Vorschlag um die WWW- Seiten mehr benutzerfreundlich zu gestalten?

---

---

**Frage 5:** gibt es etwas das Sie den Programmierern der WWW- Seiten sagen möchten;

---

---