



ΑΤΕΙ ΗΠΕΙΡΟΥ

ΤΜΗΜΑ ΤΗΛΕΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

ΣΥΓΚΡΙΣΗ ΛΟΓΙΣΜΙΚΩΝ ΠΑΚΕΤΩΝ ΓΙΑ ΑΝΑΛΥΣΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ LOG FILES



ΥΠΕΥΘΥΝΟΣ ΣΠΟΥΔΑΣΤΗΣ : ΔΡΑΚΟΣ ΙΩΑΝΝΗΣ
ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ : ΣΤΥΛΙΟΣ ΓΕΩΡΓΙΟΣ

ΑΡΤΑ 2004

ΣΥΓΚΡΙΣΗ ΛΟΓΙΣΜΙΚΩΝ ΠΑΚΕΤΩΝ ΓΙΑ ΑΝΑΛΥΣΗ
ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ LOG FILES

ΔΡΑΚΟΣ ΙΩΑΝΝΗΣ

ΣΕΠΤΕΜΒΡΙΟΣ 2004

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΜΕΡΟΣ ΤΩΝ ΑΠΑΙΤΗΣΕΩΝ ΤΟΥ
ΤΜΗΜΑΤΟΣ ΤΗΛΕΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

Abstract (σύντομη περιγραφή διπλωματικής)

Η παρακάτω διπλωματική εργασία έχει σκοπό την αξιολόγηση λογισμικών πακέτων που χρησιμοποιούνται για ανάλυση και επεξεργασία log files που αποθηκεύονται στο web server , δηλαδή πρόκειται για log analyzer , από αυτό συνήθως εξυπηρετούνται οι διαχειριστές των ιστοσελίδων , διότι τους βοηθάνε στην καλύτερη ανάλυση των log files , έτσι ώστε να προβούν στην καλύτερη εξόρυξη δεδομένων από το διαδίκτυο .

Η μέθοδος που ακολουθήθηκε κατά στην διάρκεια της διπλωματικής είναι η εξής στην αρχή να γίνει μια αναφορά σε κάποια πράγματα γενικά σε εξόρυξη δεδομένων, στην συνέχεια να γίνει μια αναφορά για τα log files , όπου θα υπάρχουνε κάποια παραδείγματα ,και κάποιες εξηγήσεις από log files. Ύστερα θα υπάρχει μια παρουσίαση για τα προγράμματα που θα αξιολογήσουμε ,όπου θα γίνει μια αξιολόγηση των συγκεκριμένων προγραμμάτων, στη συνέχεια θα εξάγουμε τα αποτελέσματα του κάθε προγράμματος αφού έχουμε τρέξει στο καθένα software τα log files, τέλος θα επισημάνουμε τα συμπεράσματα για κάθε λογισμικό πακέτο.

Περιεχόμενα	σελίδα
κεφαλαίο 1 εισαγωγή	1
1.1 Σκοπός και κίνητρο μελέτης διπλωματικής	1
1.2 Στόχος διπλωματικής	2
1.3 Δομή διπλωματικής	4
1.4 Προβλήματα διπλωματικής	4
κεφαλαίο 2 εξόρυξη δεδομένων	6
2.1 Ορισμός	6
2.1.1 Τρόποι και Διαδικασία εξόρυξης δεδομένων	6
2.2 Κατηγορίες των τεχνικών εξόρυξης δεδομένων	7
2.3 Web usage mining	9
2.3.1 Προεπεξεργασία δεδομένων	10
2.3.2 Καθαρισμός δεδομένων	11
2.3.3 Αναγνώριση χρηστή	11
2.3.4 Ανακάλυψη σχεδίων	12
2.3.5 Ανάλυση σχεδίων	13
κεφαλαίο 3 log files	14
3.1 Ορισμός και περιεχόμενα log files	14
3.2 Προβλήματα κατά την ανάλυση log files	17
3.2.1 Ιδανικό log file	18
3.3 Προεπεξεργασία log files	19
3.3.1 Καθαρισμός δεδομένων	19
3.3.2 Αναγνώριση χρηστή	20
κεφαλαίο 4 Ανάλυση μεθοδολογίας	22
4.1 Εύρεση log file	22
4.2 Επιλογή λογισμικών πακέτων	23
4.3 Βασικοί τρόποι αξιολόγησης	25
4.4 Προβλήματα (διαδικαστικά)	28
4.5 Σύντομη περιγραφή software	28
κεφάλαιο 5 Αξιολόγηση λογισμικών πακέτων	32
5.1 Παρουσίαση σελίδας	32
5.2 Αξιολόγηση των λογισμικών πακέτων	33
5.2.1 123loganalyzer version 3.10	33
5.2.2 Absolute Log Analyzer Demo v 2.33	34

5.2.3 Absolute Log Analyzer Pro Demo v 2.33	36
5.2.4 Advanced Log Analyzer v 1.5	37
5.2.5 AlterWind LogAnalyzer Professional v2.1	38
5.2.6 Azure Web Log 1.4	39
5.2.7 FastStats Analyzer v 2.81	41
5.2.8 OpenWebScope Web Statistics v 1.01	42
5.2.9 WebLog Expert Lite v 3.0	43
5.3 Εκτέλεση log files στα λογισμικά πακέτα	44
5.4 Σύγκριση log analyzer	52
Κεφάλαιο 6 Συμπεράσματα	58
Αναφορές-Βιβλιογραφία	59
Παράρτημα-log file	61

Εικόνες

εικόνα 1 ^η	2
εικόνα 2 ^η	3
εικόνα 3 ^η	12
εικόνα 4 ^η	16
εικόνα 5 ^η	20
εικόνα 6 ^η	21
εικόνα 7 ^η	25
εικόνα 8 ^η	33
εικόνα 9 ^η	45
εικόνα 10 ^η	46
εικόνα 11 ^η	46
εικόνα 12 ^η	47
εικόνα 13 ^η	48
εικόνα 14 ^η	49
εικόνα 15 ^η	50
εικόνα 16 ^η	51
εικόνα 17 ^η	52

ΔΗΛΩΣΗ ΠΕΡΙ ΛΟΓΟΚΛΟΠΗΣ

όλες οι προτάσεις οι οποίες παρουσιάζονται σ'αυτο το κείμενο και οι οποίες ανήκουν σε άλλους αναγνωρίζονται από τα εισαγωγικά και υπάρχει η σαφής δήλωση του συγγραφέα. Τα υπόλοιπα γραφομενα είναι επινόηση του γράφοντος ο οποίος φέρει και την καθολική ευθύνη γι αυτό το κείμενο και δηλώνω υπεύθυνα ότι δεν υπάρχει λογοκλοπή γι αυτό το κείμενο.

Όνοματεπώνυμο.....

Υπογραφή.....

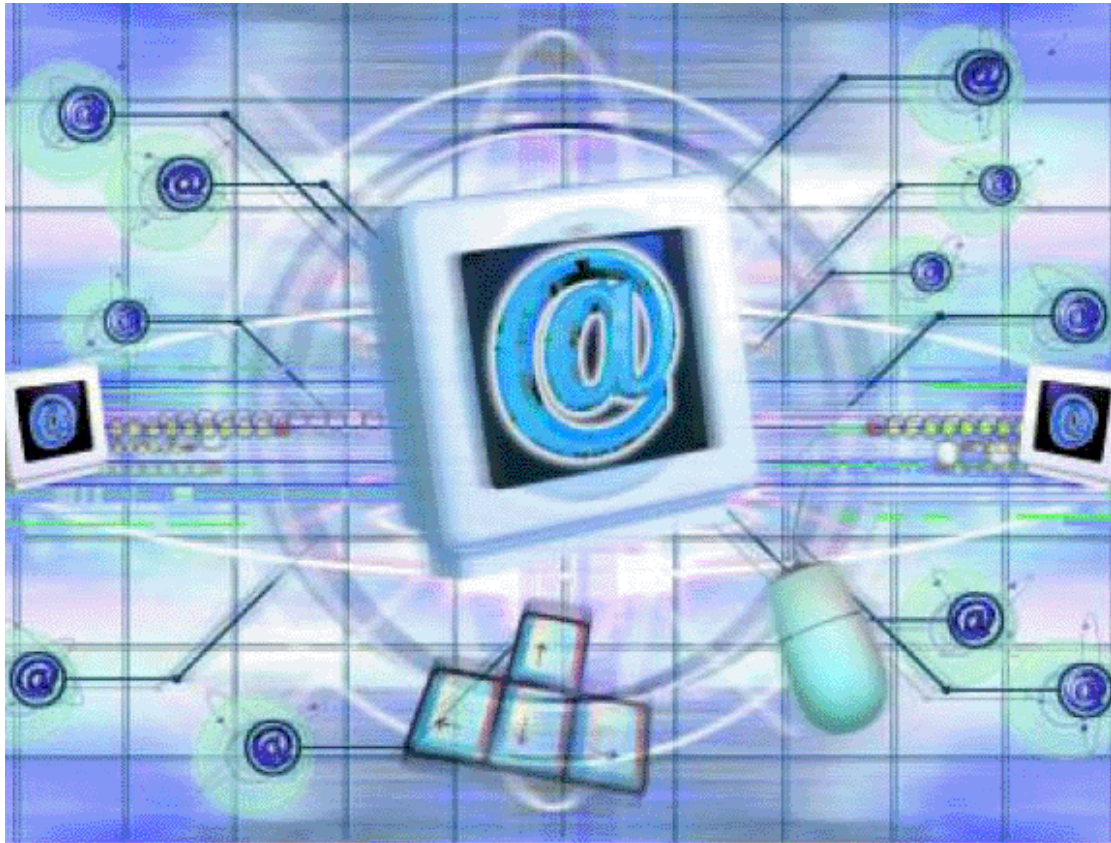
Ημερομηνία.....

κεφαλαίο 1 εισαγωγή

Σε αυτό το κεφαλαίο περιγράφουμε με λίγα λόγια τι θα πούμε και τι θα αναφέρουμε στην διπλωματική εργασία, πρόκειται για ένα καθαρό εισαγωγικό κεφαλαίο όπου δίνεται η δυνατότητα στον αναγνώστη διαβάζοντας αυτό το κεφαλαίο να καταλάβει με τι έχει να κάνει η διπλωματική εργασία, να πάρει μια γεύση από τι θα ακολουθήσει στη συνέχεια, αλλά ταυτόχρονα του δίνεται η δυνατότητα να καταλάβει το σκοπό της διπλωματικής εργασίας.

1.1 Σκοπός και κίνητρο μελέτης διπλωματικής

Πρώτα από όλα πρέπει να αναφέρουμε ότι ο σκοπός και το κίνητρο που δίνεται σε κάποιον που θέλει να μελετήσει την παρακάτω διπλωματική είναι για να αντλήσει κάποιες σημαντικές πληροφορίες που ίσως να του είναι χρήσιμες για αυτόν ή για κάτι άλλο. Ένα ακόμα κίνητρο για την μελέτη της παρακάτω διπλωματικής είναι να παρέχουμε πληροφορίες σε αυτόν που θέλει να ασχοληθεί με εξόρυξη δεδομένων ή με log files, έτσι ώστε να παρέχονται πληροφορίες σε άτομα που ασχολούνται με το διαδίκτυο ως διαχειριστές κάποιων ιστοχώρων ή είναι κατασκευαστές ιστοσελίδων, συνήθως αυτοί είναι, που ενδιαφέρονται για τέτοιου είδους πληροφορίες, και δείχνουν όλο και περισσότερο ενδιαφέρον σε τέτοιου είδους διπλωματικές εργασίες.

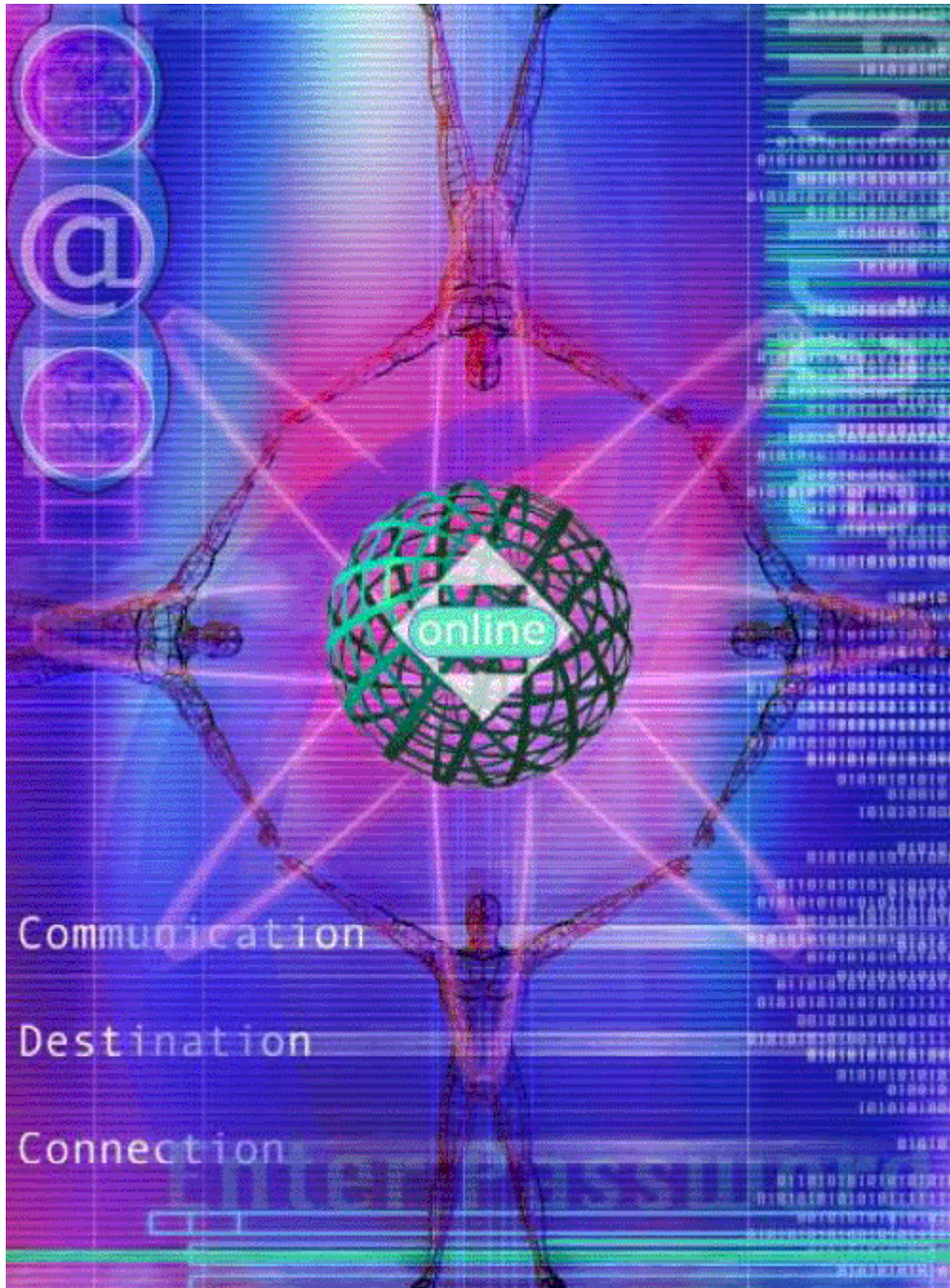


εικόνα 1^η Συλλογή πληροφορίες από web

1.2 Στόχος διπλωματικής

Ο στόχος της εργασίας μας είναι να γίνει μια σύγκριση σε κάποια συγκεκριμένα λογισμικά πακέτα που χρησιμοποιούνται στο εμπόριο για ανάλυση log αρχείων, με κάποια κοινά κριτήρια αξιολόγησης αυτών των λογισμικών. Αφού βέβαια πρώτα έχει γίνει κάποια αναφορά, για το τι είναι η εξόρυξη δεδομένων, τι είναι ειδικότερα η εξόρυξη δεδομένων στο διαδίκτυο και πως αυτή πραγματοποιείται, ακόμα τι είναι τα log files, για το τι πρόκειται, πως μπορεί να τα διαχωρίστουμε και πως μπορούν να τα επεξεργαστούμε. Επίσης γίνεται μια αναφορά στην ιστοσελίδα που πήραμε τα δεδομένα μας, δηλαδή η συγκεκριμένη ιστοσελίδα που αναφέραμε, με τη έχει να κάνει και ποιο σκοπό εξυπηρετεί αυτή η σελίδα. Μια βασική επιδίωξη μας στην εργασία μας είναι να δώσουμε και κάποια αποτελέσματα από το κάθε λογισμικό πακέτο έτσι ώστε να είναι σε θέση πολύ εύκολα οι διαχειριστές ιστοσελίδων

να αποφασίσουν ποιο είναι το καλύτερο πακέτο για την δικιά τους ιστοσελίδα.



εικόνα 2^η ο άνθρωπος ως διαχειριστής του web

1.3 Δομή διπλωματικής

Στη συνέχεια αναφέρεται η δομή της εργασίας. Πρώτα από όλα γίνεται μια αναφορά σε κάποιους ορισμούς για την εξόρυξη δεδομένων και ειδικά για την εξόρυξη δεδομένων στο διαδίκτυο, παράλληλα διαχωρίζετε η εξόρυξη δεδομένων σε κάποιες κατηγορίες, και ταυτόχρονα αναφέρονται τα στάδια επεξεργασίας της εξόρυξης δεδομένων. Ύστερα γίνεται μια εκτενέστερη αναφορά στα log files, δηλαδή δίνεται ο ορισμός για τα log file, εξηγούμε τι είναι τα log files, και παρέχουμε κάποιες κατηγορίες των log files, ταυτόχρονα αναφέρονται στην προεπεξεργασία των συγκεκριμένων αρχείων ένα βήμα πριν την τελική επεξεργασία. Στη συνέχεια αναφερόμαστε για την μεθοδολογία που θα ακολουθηθεί κατά την διάρκεια της επεξεργασίας και της ανάλυσης των log files, και αμέσως μετά θα υπάρχει μια αναφορά σε μια κλίμακα σύγκρισης έτσι ώστε να αξιολογηθούν καλύτερα τα συγκεκριμένα αρχεία. Η συγκεκριμένη κλίμακα σύγκρισης που χρησιμοποιήθηκε για την σύγκριση των λογισμικών πακέτων έγινε με κριτήρια που φτιάξαμε μόνοι μας αλλά και από κριτήρια που υπήρχανε στο διαδίκτυο για σύγκριση οποιοδήποτε λογισμικού πακέτου. Τέλος θα αξιολογείται το κάθε λογισμικό ξεχωριστά και θα υπάρχει μια αναφορά στα υπέρ και τα κατά των συγκεκριμένων λογισμικών.

Στη συνέχεια θα εξάγονται τα αποτελέσματα από την επεξεργασία των δεδομένων μας, αφού προηγουμένως θα έχουμε τρέξει το κάθε πρόγραμμα ξεχωριστά και τέλος θα παρουσιάζονται τα τελικά αποτελέσματα μας για το κάθε λογισμικό.

1.4 Προβλήματα διπλωματικής

Βέβαια κάθε εργασία έχει και τις δυσκολίες της έτσι αναμενόμενο ήταν να υπάρχουνε και σε αυτήν την εργασία κάποιες δυσκολίες. Κατά την συγκεκριμένη εργασία μια δυσκολία ήτανε η απόφαση ποια software θα χρησιμοποιήσουμε, δηλαδή η κατάλληλη επιλογή των λογισμικών πακέτων, ένα άλλο πρόβλημα ήτανε ότι η συγκεκριμένη εργασία έγινε με

απομακρυσμένη επικοινωνία με τον επιβλέποντα καθηγητή, διότι η επικοινωνία εξ αποστάσεως ήταν αναγκαστική και δεν είναι πάντα ο καλύτερος τρόπος για να πραγματοποιηθεί μια διπλωματική εργασία. Ένα άλλο πρόβλημα κατά της διάρκειας της υλοποίησης αυτής της εργασίας είναι ότι οι περισσότερες πληροφορίες ήταν στην αγγλική γλώσσα, αφού η βιβλιογραφία που χρησιμοποιήσαμε ήταν στα αγγλικά γραμμένη και οι ιστοσελίδες που βρήκαμε υλικό για την διπλωματική ήταν επίσης στα αγγλικά, έτσι υπήρχε το πρόβλημα με την μετάφραση που έπρεπε να γίνει στην ελληνική γλώσσα. Τέλος ένα άλλο πρόβλημα που δημιουργήθηκε κατά την διάρκεια της υλοποίησης της εργασίας ήταν ότι μερικά λογισμικά πακέτα δεν μπορούσαμε να τα χρησιμοποιήσουμε καθόλου επειδή δεν μας επιτρεπόταν να γίνει η εγκατάσταση μας στον υπολογιστή που δουλεύαμε, έτσι ήταν αδύνατο να χρησιμοποιήσουμε αυτά τα προγράμματα.

κεφαλαίο 2 εξόρυξη δεδομένων

εισαγωγή

Σε αυτό το κεφαλαίο θα αναφερθούν κάποια στοιχεία που έχουν να κάνουν με εξόρυξη δεδομένων θα δώσουμε κάποιους ορισμούς περί εξόρυξης δεδομένων και του τρόπους που ακολουθείται. Στη συνέχεια θα γίνουμε πιο συγκεκριμένοι περί εξόρυξης δεδομένων στο διαδίκτυο, όπου εκεί θα αναφέρουμε κάποιους τρόπους πως γίνεται η εξόρυξη δεδομένων στο διαδίκτυο, και τη λογική σύνδεση με τα log files, όπου είναι ένας τρόπος εξόρυξης δεδομένων στο διαδίκτυο.

2.1 Ορισμός

Εξόρυξη δεδομένων είναι ένας τρόπος , μια μέθοδος που μας επιτρέπει να ανακαλύπτουμε πληροφορίες από κάποιες βάσεις δεδομένων, μας επιτρέπει να αντλούμε πληροφορίες από οποιαδήποτε βάση δεδομένων και στη συνέχεια για οποιοδήποτε λόγο εμείς θέλουμε να εκμεταλλευόμαστε τις συγκεκριμένες πληροφορίες.

2.1.1 Τρόποι και Διαδικασία εξόρυξης δεδομένων

Υπάρχουν 2 τρόποι εξόρυξης δεδομένων, ο πρώτος είναι να γίνεται πρώτα η ανασυρση της πληροφορίας και μετά η επεξεργασία της, ο δεύτερος τροπος είναι η απευθείας ανασυρση της πληροφορίας όπου θα είναι ήδη επεξεργασμένη μέσα στη βάση δεδομένων

ΔΙΑΔΙΚΑΣΙΑ

Η διαδικασία της εξόρυξης δεδομένων ακολουθεί κάποια βήματα για να φτάσουμε στο επιθυμητό αποτέλεσμα, τα βήματα είναι τα εξής

- 1) θέτουμε στόχους
- 2) επιλογή στοιχείων
- 3) προεπεξεργασία στοιχείων

- 4) ανάλυση
- 5) επικύρωση
- 6) προβλέψεις
- 7) παρουσίαση

Στην αρχή, είναι σημαντικό να γίνει κατανοητό ποιοι είναι οι στόχοι της ανάσυρσης δεδομένων.

Η επιλογή στοιχείων περιλαμβάνει την προσεκτική διαλογή των δεδομένων από μια ή περισσότερες βάσεις δεδομένων. Είναι καλύτερο να αρχίσει με λιγότερα στοιχεία για να έχουμε μια πιο καλύτερη εικόνα.

Μόλις τα μη επεξεργασμένα στοιχεία επιλεχτούν η προεπεξεργασία είναι το επόμενο στάδιο όπου συνήθως είναι απαραίτητη για να γίνει η διαπίστωση εάν λείπουν βασικά στοιχεία μας.

Το άμεσο επόμενο στάδιο είναι η ανάλυση των στοιχείων μας όπου γίνεται μέσα από κάποια εργαλεία που χρησιμοποιούνται στην εξόρυξη δεδομένων.

Στη συνέχεια η επικύρωση των αποτελεσμάτων από την ανάλυση είναι το επόμενο στάδιο όπου εκεί γίνεται ο έλεγχος και ταυτόχρονα η επιβεβαίωση για τα δεδομένα που αντλήσαμε.

Τέλος η παρουσίαση των δεδομένων είναι το τελευταίο στάδιο για την επεξεργασία των στοιχείων από την βάση δεδομένων, συνήθως η παρουσίαση γίνεται μέσα από σχεδιαγράμματα και στατιστικά στοιχεία που μας δίνουν τα εργαλεία που χρησιμοποιήσαμε παραπάνω,
(αναφορά IRM Data Mining. Copyright ©1999, all rights reserved)

2.2 Κατηγορίες των τεχνικών εξόρυξης δεδομένων

Οι τεχνικές της εξόρυξης δεδομένων έχουν αναπτυχθεί παρά πολύ τα τελευταία χρόνια έτσι ώστε μας επιτρέπουν να τις διαχωρίσουμε σε τρεις κατηγορίες:

- Στη κατηγορία ανάλογα με τη βάση δεδομένων.
- Στη κατηγορία ανάλογα με τα στοιχεία που περιέχονται στις βάσεις δεδομένων.
- Στη κατηγορία ανάλογα με τις τεχνικές που χρησιμοποιούμε στην εξόρυξη δεδομένων.

(αναφορά Data Mining: An Overview from Database Perspective
Ming-Syan Chen Elect. Eng. Department National Taiwan Univ.)

αναλόγως την βάση δεδομένων , δηλαδή η εξόρυξη δεδομένων έχει να κάνει με το είδος της βάσης δεδομένων μας , μια βάση δεδομένων μπορεί να ανήκει σε μια από τις παρακάτω κατηγορίες:σχεσιακές βάσεις δεδομένων, βάσεις δεδομένων συναλλαγής, αντικειμενοστραφείς βάσεις δεδομένων,

αναλόγως τα στοιχεία που περιέχονται στη βάση δεδομένων, δηλαδή αν μια συγκεκριμένη βάση δεδομένων περιέχει κάποια ίδιου τύπου στοιχεία ή τα στοιχεία της είναι διαφορετικά,

αναλόγως με τις τεχνικές που χρησιμοποιούμε για την εξόρυξη δεδομένων μπορούμε να τις χωρίσουμε σε κατηγορίες όπου έχουν να κάνουν με τον τρόπο που κάνουνε την όλη διαδικασία στην εξόρυξη δεδομένων,

(αναφορά Data Mining: An Overview from Database Perspective
Ming-Syan Chen Elect. Eng. Department National Taiwan Univ.
Taipei, Taiwan, ROC Jiawei Han □ School of Computing Sci. Simon Fraser
University B.C. V5A 1S6, Canada Philip S. Yu IBM T.J. Watson Res.
Ctr.P.O.Box 704 Yorktown, NY 10598, U.S.A.)

2.3 Web usage mining

Η εξόρυξη δεδομένων από το web είναι μια περίπτωση εξόρυξη δεδομένων που τα τελευταία χρόνια απασχολεί όλο ένα και περισσότερους διαχειριστές ιστοσελίδων. Ο λόγος είναι ότι το διαδίκτυο τα τελευταία χρόνια έχει αναπτυχθεί παρά πολύ και ο κάθε διαχειριστής θα ήθελε να ξέρει τι γίνεται με την ιστοσελίδα του σε σχέση με την επισκεψιμότητα του, ποιος την επισκέπτεται, ποιες σελίδες επισκέπτεται και ποτέ την επισκέπτεται. Ένας ακόμη λόγος που οι διαχειριστές μιας σελίδας ενδιαφέρονται για την ιστοσελίδα τους είναι αν και κατά πόσο η ιστοσελίδα τους ανταποκρίνεται στους όρους και στις προϋπόθεσης που οι ίδιοι θέσανε από την αρχή της κατασκευής της σελίδας τους, πόσο μάλλον αν πρόκειται για σελίδα που έχει άμεση σχέση με ηλεκτρονικό εμπόριο π.χ. ηλεκτρονικό βιβλιοπωλείο.

Επίσης ένας σημαντικός λόγος που πολλοί διαχειριστές θέλουν να γνωρίζουν την επισκεψιμότητα του ιστοχώρου τους είναι για να βελτιώσουν την ποιότητα των υπηρεσιών τους , και ακόμα για να χαράξουν ή να αλλάξουν τελείως την πολιτική που ακολουθούνε στην ιστοσελίδα τους.

Ένας διαχειριστής μπορεί να βγάλει συμπεράσματα μέσα από τον ISP(internet service provider)από τον παροχές υπηρεσιών διαδικτυου, όπου από εκεί αντλεί όλα τα στοιχεία που του είναι χρήσιμα για να βγάλει τα δικά του συμπεράσματα για τον ιστοχώρο, τα στοιχεία που ζητάει ένας διαχειριστής αποθηκεύονται στον ISP και πρόκειται για τα log files,

τα δεδομένα μπορούνε να αποθηκευτούνε σε πολλές πλευρές

- Μια πλευρά είναι στον υπολογιστή του πελάτη.
- Μια άλλη πλευρά είναι να αποθηκευτούνε στον υπολογιστή παροχές .
- Μια άλλη πλευρά είναι να αποθηκεύουνε σε proxy server.
- Και τέλος μια άλλη πλευρά είναι σε κάποια βάση δεδομένων που υπάρχει μόνο για αυτό το σκοπό.

τα δεδομένα μεταξύ τους διαφέρουν σε σχέση με τα στοιχεία που έχουν αλλά συνήθως μερικά-βασικά στοιχεία είναι τα ίδια επίσης εξαρτάται από την εφαρμογή που χρησιμοποιούνται τα δεδομένα μας

Η εξόρυξη δεδομένων στο διαδίκτυο χωρίζεται σε κάποιες κατηγορίες :

- Μια κατηγορία είναι αυτή που ψάχνει στο διαδίκτυο για αρχεία που δεν είναι μόνο σε μορφή κειμένου αλλά μπορεί να είναι και σε ήχο η και σε βίντεο,η εικόνα.
- Μια άλλη κατηγορία είναι αυτή που ψάχνει στο διαδίκτυο για να βρει με τον τρόπο που κρύβονται οι συνδέσεις μεταξύ των σελίδων και να βρει κάποιες ομοιότητες μεταξύ των διάφορων σελίδων.
- Τέλος μια άλλη κατηγορία είναι αυτή που προσπαθεί να κάνει εξόρυξη δεδομένων μεταξύ των αρχείων που κρατιούνται συνήθως στον διακομιστή(log files),αυτός είναι ο πλέον επικρατέστερος τροπος εξόρυξης από το διαδίκτυο για συγκεκριμένους ιστοχωρους.

2.3.1 Προεπεξεργασία δεδομένων

Μερικά από τα προβλήματα που συναντάμε όταν επιχειρούμαι να κάνουμε τέτοιου είδους εξόρυξης δεδομένων είναι ότι οι εξής

- Μπορεί ένας διακομιστής να έχει περισσότερες από μια IP διεύθυνση.
- Μπορεί μια IP διεύθυνση να αντιπροσωπεύει περισσότερους από ένα διακομιστή.
- Μπορεί ένας χρήστης να χρησιμοποιεί περισσότερες από μια IP διεύθυνση σε συνεχή χρονικά διαστήματα

- ο Αλλά προβλήματα κατά την διάρκεια της προεπεξεργασίας των δεδομένων είναι τα δεδομένα πρέπει να συλλεχθούν από τους διακομιστές.
- ο Τα περισσότερα δεδομένα αναφέρονται σε σελίδες, διακομιστές, υπερσυνδέσεις και αλλά τέτοια παρόμοια.
- ο Μια απλή σελίδα θα είναι καταγραμμένη σε πολλές τοποθεσίες στον διακομιστή και αυτό γιατί ο σημαντικός λόγος που διαδραματίζουν οι proxy servers μιας και από πίσω θα βρίσκονται κάποιοι υπολογιστές που χρησιμοποιούν την ίδια IP διεύθυνση.

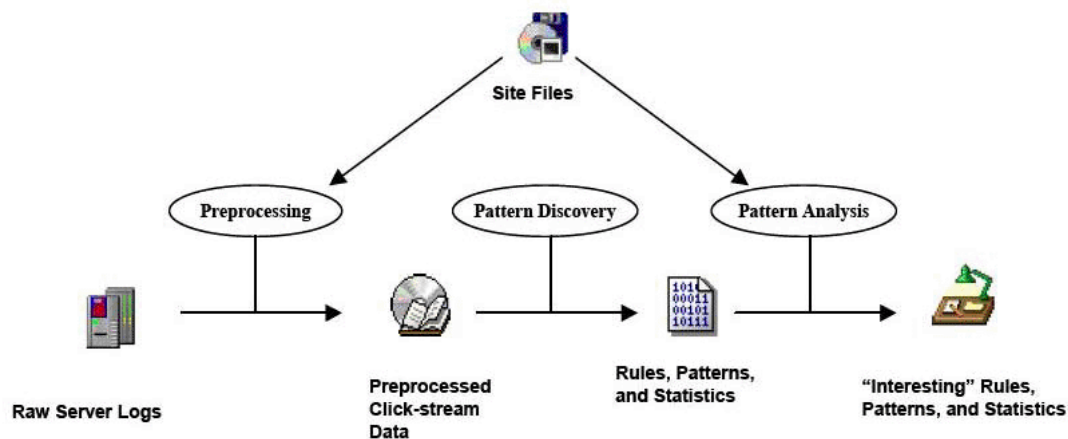
2.3.2 Καθαρισμός δεδομένων

Επίσης ένα άλλο πρόβλημα είναι ότι πρέπει κατά την διάρκεια της προεπεξεργασίας δεδομένων πρέπει να γίνει διαχωρισμός από τις σελίδες που περιέχουν καθαρά αρχεία .html και από τις σελίδες που περιέχουν κάποια αρχεία ήχου η βίντεο όπως τα .avi .gif .jpeg , έτσι ώστε τα αποτελέσματα μας να είναι πιο αντικειμενικά σε αυτό που θέλουμε να πετύχουμε , στην αντίθετη περίπτωση δηλαδή αν δεν γίνει κάποιος καθαρισμός δεδομένων τότε τα αποτελέσματα που θα έχουμε από την επεξεργασία των αρχείων δεν είναι αυτά που πραγματικά ζητάμε , έτσι θα έχουμε μια ψευδή εικόνα για την σελίδα μας και θα οδηγούμαστε σε λάθος συμπεράσματα .

2.3.3 Αναγνώριση χρηστή

Στη συνέχεια ένα άλλο πρόβλημα που έχει να κάνει με την όλη λειτουργία του διαδικτυου είναι ότι οι IP διευθύνσεις χωρίζονται σε στατικές και δυναμικές, αυτό σημαίνει ότι στην περίπτωση που οι διευθύνσεις είναι στατικές το πρόβλημα μας γίνεται πιο μικρό ενώ στην περίπτωση που οι διευθύνσεις είναι δυναμικές το πρόβλημα μας γίνεται όλο και πιο έντονο αφού δεν ξέρουμε αν κάποιος χρήστης εισέρχεται στο διαδικτυο πολλές φορές με διαφορετική IP

κάθε φορά, μιας και ο παροχές του δίνει διαφορετική διεύθυνση έτσι σε αυτή την περίπτωση δεν μπορούμε να έχουμε μια πλήρη εικόνα σε σχέση με τον ιστοχώρο μας.



εικόνα 3^η Προεπεξεργασία log files

2.3.4 Ανακάλυψη σχεδίων

Η ανακάλυψη σχεδίων αποτελείται από διάφορους μεθόδους και αλγόριθμους που χρησιμοποιούνται για γενική εξόρυξη δεδομένων και όχι μόνο στην εξόρυξη δεδομένων στο διαδίκτυο, επίσης χρησιμοποιούνται και για στατιστική χρήση, οι περισσότεροι από αυτούς είναι έτοιμοι ενώ οι υπόλοιποι τους φτιάχνουνε οι επιχειρήσεις για λογαριασμό δικό τους. Ένα χαρακτηριστικό παράδειγμα είναι στο marketing όπου οι αλγόριθμοι τους είναι πολλοί χρήσιμοι για να βγάλουνε συμπεράσματα πάνω στη αγορά.

Ένας αλγόριθμος για γενική χρήση αλλά τόσο χρήσιμος είναι ο αλγόριθμος από το *Methods and Problems in Data Mining* Heikki Mannila, University of Helsinki όπου ονομάζεται *Knowledge Discovery in Databases*, ο συγκεκριμένος αλγόριθμος κάνει εξόρυξη μέσα από μια βάση δεδομένων που θέλουμε, χρησιμοποιεί ένα γενικό αλγόριθμο για να ανακαλύψει την γνώση και αυτό το λόγο μπορεί να χρησιμοποιηθεί σε πολλές βάσεις δεδομένων.

2.3.5 Ανάλυση σχεδίων

Η ανάλυση σχεδίων είναι το τελευταίο στάδιο στη γενική ανάλυση δεδομένων στο διαδίκτυο όπου σε αυτό το στάδιο περιλαμβάνει την διαλογή ανάμεσα από τους αλγόριθμους και τους μεθόδους, μιας και θέλουμε να χρησιμοποιήσουμε κάποιους συγκεκριμένους αλγόριθμους που μας εξυπηρετούνε καλύτερα. Συνεπώς αναλόγως την χρήση και την δουλεία που θέλουμε να κάνουμε επιλέγουμε και τον κατάλληλο αλγόριθμο που θα μας βοηθήσει κατάλληλα.

(αναφορά Communications of the Association for Information Systems (Volume 8, 2002) 267-296 267 Data Mining: A Conceptual Overview by J. Jackson DATA MINING: A CONCEPTUAL OVERVIEW Joyce Jackson Management Science Department University of South Carolina joyce.jackson@sc.edu)

κεφαλαίο 3 log files

εισαγωγή

Στο κεφαλαίο που ακολουθεί θα γίνει μια αναφορά για τα log files ,θα αναφέρουμε τι είναι τα log files ,που βρίσκονται ,τι σκοπό έχουνε και για ποιο ρόλο υπάρχουνε, επίσης θα αναφέρουμε τι είδους προεπεξεργασίας γίνεται, έτσι ώστε να είναι έτοιμα για την τελική επεξεργασία και ανάλυση τους μέσα από τα λογισμικά πακέτα.

Η παρακολούθηση της δραστηριότητας του server μπορεί να γίνει με την χρήση διαφόρων μεθόδων. Υπάρχει δυνατότητα ελέγχου της κατάστασης του server (servers status) σε πραγματικό χρόνο χρησιμοποιώντας το Πρωτόκολλο Μεταφοράς Υπερκειμένου (HTTP-HyperText Transfer Protocol) ή το Πρωτόκολλο διαχείρισης Απλού δικτύου (SNMP-Simple Network Management Protocol). Έτσι, ο έλεγχος του server πραγματοποιείται με την καταγραφή και την εξέταση των log αρχείων (log files) ή με την χρήση των εργαλείων ελέγχου απόδοσης, τα οποία παρέχονται από το λειτουργικό σύστημα.

3.1 Ορισμός και περιεχόμενα log files

Τα log αρχεία είναι πολύ δυναμικά εργαλεία. Κάθε φορά, που ένα αρχείο ζητείται από το web server τότε ο συγκεκριμένος διακομιστής κρατάει κάποια αρχεία για αυτήν την αίτηση, αυτά είναι τα γνωστά log files, ένα τέτοιο αρχείο καταγράφει όλες τις ημερήσιες επισκέψεις σε μια σελίδα. Επίσης μπορεί ένα log files να μην περιέχει μόνο αρχεία με επέκταση html αλλά και αρχεία με άλλες καταλήξεις όπως .php , .asp .js

Log files είναι δυναμικά αρχεία που αποθηκεύονται σε ένα συγκεκριμένο κατάλογο στον server και κρατάνε στοιχεία για τις επισκέψεις που δέχεται ο server.

(αναφορά LOG FILES ANALYSIS TO MEASURE THE UTILITY OF AN INTRANET

Stéphane Vokar, Céline Mariage, Jean Vanderdonckt

Université catholique de Louvain – Institut d'Administration et de Gestion

Place des Doyens, 1 - B-1348 Louvain-la-Neuve, Belgium

{vokar, mariage,vanderdonckt}@isys.ucl.ac.be –

<http://www.isys.ucl.ac.be/bchi>

)

Διαφορετικοί web servers καταγράφουν διαφορετικές πληροφορίες, όμως οι παρακάτω πληροφορίες υπάρχουν σχεδόν πάντα στα log αρχεία των web servers:

ο κάθε server κρατάει σε ένα συγκεκριμένο κατάλογο στο δίσκο του το log file όπου συνήθως αυτός ο κατάλογος δεν μπορεί να αλλάξει εύκολα . Στη συνέχεια παρουσιάζουμε μια γραμμή από το log file έτσι ώστε να αναλύσουμε τι είναι το κάθε στοιχείο από το αρχείο.

```
62.103.78.228 - - [28/Apr/2004:10:10:45 +0300] "GET /index.html HTTP/1.1"
304 - "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)"
```

«62.103.78.228» είναι η IP διεύθυνση του χρήστη .

«28/Apr/2004:10:10:45 +0300» είναι η ημερομηνία και η ώρα που έγινε η ζήτηση από τον χρήστη .

«GET /index.html HTTP/1.1» είναι η σελίδα που ζητάει ο χρήστης να δει στον φυλλομετρητή του.

«Mozilla/4.0» είναι ο φυλλομετρητής που διαθέτει ο διακομιστής.

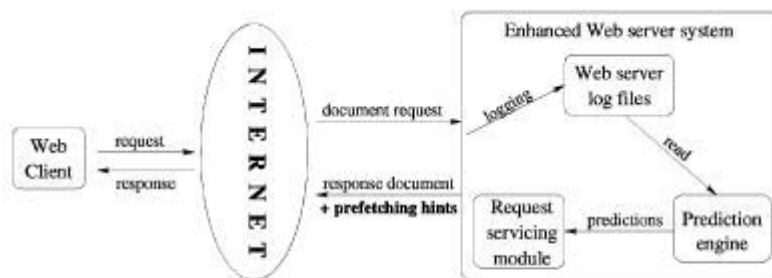
«; MSIE 6.0; Windows NT 5.1» είναι ο φυλλομετρητής του χρήστη και το λειτουργικό σύστημα που χρησιμοποιεί ο χρήστης καθώς και οι εκδόσεις τους.

- ο υπολογιστής που χρησιμοποιεί ο χρήστης (π.χ. n12.pa.isp.com ή στην μορφή της IP διεύθυνσης 127.192.192.1).

- η ημέρα και η ώρα της ζήτησης (request) π.χ 20/Apr/2004:10:10:45 +0300.
- το αρχείο, το οποίο ζητήθηκε από τον πελάτη (client) π.χ index.html.
- το μέγεθος του αρχείου π.χ. 50 KB.
- ένας κωδικός της HTTP κατάστασης (HTTP status code)(π.χ. γνωστό HTTP status code 404 file not found).

Επιπρόσθετα, κάποιοι web servers καταγράφουν (log) τις παρακάτω πληροφορίες:

- referring URL (την web σελίδα, που επισκέφτηκε ο χρήστης ακριβώς πριν μπει στον δικό μας web server)
π.χ. www.yahoo.gr και μετά www.infowood.gr
- το όνομα του πελάτη (client name)(το όνομα και την έκδοση του browser και του λειτουργικού συστήματος, κάτω από τα οποία τρέχει ο browser)
π.χ. MSIE 6.0; Windows NT 5.1



εικόνα 4^η Συναλλαγή client- server

Τα log αρχεία μπορούν να γίνουν πάρα πολύ μεγάλα. Ένα σχετικά μικρό web site, το οποίο έχει μερικές χιλιάδες επισκεπτών μηνιαία, θα έχει περίπου 25MB log αρχεία.

Το log file λαθών (error log file), το οποίο βρίσκεται στο `https-servername/logs/errors` στο root directory του server, δημιουργεί κατάλογο με όλα τα λάθη, που έχει καταμετρήσει ο server.

Το log file πρόσβασης (access log file), το οποίο βρίσκεται στο `https-servername/logs/access` στο root directory του server, καταγράφει πληροφορίες αναφορικά με ερωτήσεις προς τον server και απαντήσεις από τον server.

Συνήθως ερωτήσεις είναι οι αιτήσεις που ζητάνε οι χρήστες από τον διακομιστή ενώ απαντήσεις είναι οι πληροφορίες μέσω των σελίδων που δίνουν οι διακομιστές. Είναι εύκολο να διαμορφωθούν οι πληροφορίες, οι οποίες καταγράφονται στο access log file του server.

Συνήθως το αρχείο που χρησιμοποιείται πιο πολύ για επεξεργασία και ανάλυση είναι το δεύτερο (access log file) διότι περιέχει όλες εκείνες τις πληροφορίες που είναι απαραίτητες για να γίνει εξαγωγή με όλα τα αποτελέσματα που έχουν σχέση με την σελίδα μας.

(αναφορά Web Usage Mining on Proxy Servers:

A Case Study Jan Kerkhofs Prof. Dr. Koen Vanhoof Danny Pannemans
Limburg University Centre July 30, 2001)

3.2 Προβλήματα κατά την ανάλυση log files

Τα προβλήματα που παρουσιάζονται κατά την ανάλυση των log files είναι δυο το πρώτο είναι ότι τα log files περιέχουν στοιχεία ελλιπή για να επεξεργαστούνε στην συνέχεια το δεύτερο πρόβλημα είναι ότι τα log files μερικές φορές περιέχουν ξένα στοιχεία από εκείνα που πρέπει να υπάρχουνε μέσα στο server αυτά είναι δυο πιθανά προβλήματα που μπορεί να συμβαίνουν σε ένα αρχείο (access log file).

Εδώ και καιρό έχουμε αναπτυχθεί κάποια πακέτα με προγράμματα που κάνουνε ανάλυση και επεξεργασία των log files, αλλά τα περισσότερα από αυτά ,τα συμπεράσματα που εξάγουμε έχουμε άμεση σχέση με τον τομέα του marketing. Δηλαδή οι πληροφορίες που θα σου δώσουνε δεν είναι πάντα οι επιθυμητές και συνήθως έχουμε άμεση σχέση με την αγορά και το marketing π.χ. τα αποτελέσματα θα βοηθούσανε καλύτερα ένα υπεύθυνο marketing και όχι έναν διαχειριστή μιας σελίδας.

Οι περισσότεροι server κρατάνε διαφορετικά στοιχεία μέσα στα log files τους έτσι γίνεται όλο και πιο δύσκολη η επεξεργασία τους, πρόκειται για κάποια στοιχεία που συνήθως μας είναι άχρηστα.

Στη συνέχεια παρουσιάζονται πως θα έπρεπε να είναι τα ιδανικά log files, δηλαδή τι στοιχεία να περιέχουνε ,έτσι ώστε να μπορούμε να τα διαχειριζόμαστε καλύτερα και να εξάγουμε τα συμπεράσματα που εμείς ζητάμε και όχι αυτά που δεν μας ενδιαφέρουνε.

3.2.1 Ιδανικό log file

Το ιδανικό log file αποτελείται από τις παρακάτω προτάσεις:

- 1) Ποιος επισκέπτεται την σελίδα μας, έτσι ώστε να έχουμε μια εικόνα από τους χρηστές μας .
- 2) Ποιες σελίδες οι χρηστές "κατεβάζουνε στον υπολογιστή τους " και με ποιο τρόπο γίνεται η προσπέλαση δηλαδή μέσα από ποια link η από ποια icon.
- 3) Πόσο χρόνο κάθονται οι χρηστές μέσα στη σελίδα μας, έτσι ώστε να καταλαβαίνουμε αν η σελίδα μας είναι βαρετή η προκαλεί ενδιαφέρον στον επισκέπτη μας.

- 4) Που οι επισκέπτες αφήνουνε την σελίδα μας ,δηλαδή σε ποια υποσέλιδα οι επισκέπτες σταματάνε την επισκεψιμοτητα μας μέσα στο site, είναι ένας καλός τρόπος να σταματάμε εκεί το site μας.
- 5) Την επιτυχία που είχε η επίσκεψη ενός χρηστή στο site μας, δηλαδή πόσες σελίδες προσπέλασε, ποσά download εκανε,ποσα e-mail έστειλε στον διαχειριστή κ.τ.λ.

(αναφορά Web Usage Mining on Proxy Servers:

A Case Study Jan Kerkhofs Prof. Dr. Koen Vanhoof Danny Pannemans

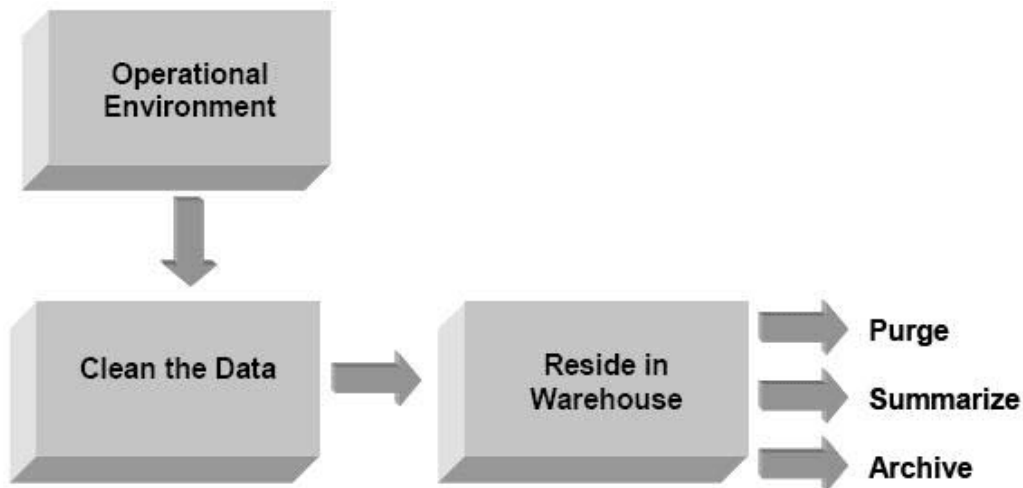
Limburg University Centre July 30, 2001)

3.3 Προεπεξεργασία log files

Στη συνέχεια γίνεται η απαραίτητη προεπεξεργασία των αρχείων log ,γίνεται σε δυο στάδια πρώτα γίνεται καθαρισμό δεδομένων και μετά η αναγνώριση χρηστή

3.3.1 Καθαρισμός δεδομένων

Πρώτα από όλα αφαιρούμε τα αρχεία που δεν θέλουμε να συμπεριλάβουμε στην προεπεξεργασία των log files όπως .wav , .gif, .jpeg και αλλα,στη συνέχεια κρατάμε εμείς ποια στοιχεία θέλουμε για να εξεργαστούμε δηλαδή μπορεί ο διακομιστής να μας παρέχει την δυνατότητα να κρατάμε όσες γίνεται περισσότερες πληροφορίες, εμείς όμως να μην χρειαζόμαστε κάποιες από αυτές, έτσι μπορούμε να τις αφαιρέσουμε από το αρχείο μας. π.χ. μπορεί κάποιος διαχειριστής να μην τον ενδιαφέρει ο browser που χρησιμοποιεί ο χρηστής ,έτσι μπορεί κάλλιστα να αφαιρέσει από το αρχείο ότι έχει σχέση με browser, όπως ο internet explorer ή ο Netscape navigator.



εικόνα 5^η Καθαρισμός δεδομένων

3.3.2 Αναγνώριση χρήστη

Το επόμενο σημαντικό βήμα στην προεπεξεργασία των αρχείων log είναι ότι δεν μπορούμε να έχουμε μια ακριβής εικόνα για τους χρήστες, διότι μπορεί ένας υπολογιστής να κατεβάσει μια σελίδα και μετά από λίγο να ξανακατεβάσει τη ίδια αλλά με διαφορετική IP διεύθυνση όταν έχουμε δυναμική IP, τότε ο διακομιστής τον αναγνωρίζει σαν διαφορετικό αλλά στην πραγματικότητα είναι ο ίδιος. Επίσης μπορεί ένας υπολογιστής να είναι στο διαδίκτυο για ώρες αλλά οι χρήστες του ανά τακτά χρονικά διαστήματα να είναι διαφορετικός, αυτό οφείλεται στην IP διεύθυνση όταν έχουμε στατική IP, γι' αυτό μας δεν μπορούμε να έχουμε μια πλήρης και ακριβής εικόνα.

Ένας διακομιστής μετράει τον επισκέπτη κάθε φορά που αυτός θέλει να μπει σε μια ιστοσελίδα, για κάθε διαφορετική σελίδα στον ίδιο χώρο μετράται σαν μια νέα αίτηση από τον διακομιστή, την στιγμή που κάνει αίτηση ο επισκέπτης καταγράφεται στο αρχείο log, ως είναι και σε υποσελίδα του ίδιου ιστοχώρου.

Μετά από όλα αυτές τις ενέργειες που έχουμε κάνει, μπορούμε να προχωρήσουμε στη ανάλυση και επεξεργασία των log files, αν και πρέπει να αναφερθεί ότι αυτός ο τρόπος εξόρυξης δεδομένων δεν ανήκει σε κάποιον αλγόριθμο ή σε κάποια συγκεκριμένη μέθοδο. Αλλά λόγω ότι είναι ένα πολύ

δυνατό εργαλείο για εξόρυξη δεδομένων στο διαδίκτυο χρησιμοποιείται ευρέως στην αγορά ,επίσης μπορούμε να πούμε ότι κατά κάποιο τρόπο υπάγεται στην γενική κατηγορία της εξόρυξης δεδομένων η ανάλυση αρχείων log αφού ουσιαστικά πρόκειται για εξόρυξη δεδομένων από το web,



εικόνα 6^η Προεπεξεργασία δεδομένων

(αναφορά Web Usage Mining on Proxy Servers:

A Case Study Jan Kerkhofs Prof. Dr. Koen Vanhoof Danny Pannemans

Limburg University Centre July 30, 2001)

κεφαλαίο 4 Ανάλυση μεθοδολογίας

εισαγωγή

Σε αυτό το κεφαλαίο θα είναι η ανάλυση της μεθοδολογίας που θα ακολουθήσουμε κατά την διάρκεια της διπλωματικής, δηλαδή να αναφέρουμε από που βρήκαμε τα log files, με ποιο τρόπο επεξεργαστήκαμε τα συγκεκριμένα λογισμικά πακέτα. Ύστερα θα πρέπει να αναφέρουμε τους τρόπους αξιολόγησης, δηλαδή την κατηγοριοποίηση, στη συνέχεια να επισημάνουμε τα προβλήματα που δημιουργήθηκαν κατά την συγκεκριμένη διαδικασία, και τέλος γίνει μια σύντομη περιγραφή του κάθε λογισμικού πακέτου.

Σε αυτό το κεφαλαίο θα γίνει η ανάλυση την όλη μεθοδολογία που θα ακολουθήσουμε στην συνέχεια της διπλωματικής μας

4.1 Εύρεση log file

Πρώτα από όλα πρέπει να αναφέρουμε που βρήκαμε το log files, το αρχείο που θα χρησιμοποιήσουμε στην περαιτέρω πορεία της διπλωματικής εργασίας μας, το αρχείο που θα χρησιμοποιήσουμε το έχουμε πάρει από μια εταιρεία πληροφορικής που κατασκευάζει ένα σχεδιαστικό πρόγραμμα για εσωτερικό χώρο, όπου το αρχείο το πήραμε από τον παροχές υπηρεσιών διαδικτύου της συγκεκριμένης επιχείρησης όπου φιλοξενεί την ιστοσελίδα τους, πρόκειται για την σελίδα www.infowood.gr. Ο λόγος που διαλέξαμε τα αρχεία από την συγκεκριμένη σελίδα είναι ότι κατά το διάστημα που κατασκευάζαμε την πτυχιακή εργασία μπορούσαμε να προμηθευτούμε το συγκεκριμένο αρχείο από τον παροχέα, έτσι εκμεταλλευτήκαμε και το γεγονός ότι το εν λόγω αρχείο είναι από apache web server και επειδή ο συγκεκριμένος server χρησιμοποιείται για φιλοξενία ιστοσελίδων στους περισσότερους παροχής υπηρεσιών διαδικτύου λόγο του ότι είναι δωρεάν (open source).

4.2 Επιλογή λογισμικών πακέτων

Ο τρόπος επιλογής των λογισμικών πακέτων έγινε με βάση κάποια κριτήρια που αξιοποιήσαμε, έτσι ώστε να φτάσουμε στην επιλογή αυτών των πακέτων. Πρώτα από όλα τα λογισμικά πακέτα επιλέχθηκαν από το διαδίκτυο μέσα από διάφορους δικτυακούς τόπους, όπου οι περισσότεροι από αυτούς είχαν άμεση σχέση με log files ή ακόμα με log analyzer, δηλαδή πρόκειται για δικτυακούς τόπους, όπου παρουσιάζανε κάποια δικά τους προϊόντα, από συγκεκριμένες εταιρείες που κατασκευάζουν τέτοια λογισμικά πακέτα μιας και πρόκειται για δικά τους λογισμικά πακέτα, όπου είχαν να κάνουν με ανάλυση και επεξεργασία τέτοιου είδους αρχείου, βέβαια εδώ πρέπει να αναφέρουμε ότι η αναζήτηση και το κατέβασμα αυτών των λογισμικών πακέτων έγινε μέσα από κάποιες μηχανές αναζητήσεις όπως www.yahoo.com, www.google.com και www.download.com.

Εδώ πρέπει να αναφέρουμε ότι οποιοσδήποτε που επιθυμεί μπορεί κατεβάσει αυτά τα πακέτα, από το διαδίκτυο μιας και τα περισσότερα είναι ελεύθερα και δεν χρειάζεται να ξοδέψει κάποιος χρήματα για να τα αποκτήσει.

Βέβαια εδώ πρέπει να πούμε ότι οι οποιοσδήποτε κατηγορίες των διάφορων λογισμικών πακέτων χωρίζονται σε δυο μεγάλες κατηγορίες, η πρώτη κατηγορία έχει να κάνει με τα λογισμικά πακέτα που είναι ελεύθερα και δεν χρειάζεται κανένας περιορισμός για να τα κατεβάσεις, η άλλη μεγάλη κατηγορία έχει να κάνει με τα λογισμικά πακέτα όπου σου δίνουν το δικαίωμα να κατεβάσεις μια έκδοση ενός προγράμματος που λήγει σε κάποιο μικρό χρονικό διάστημα για να το δοκιμάσεις και άμα ταιριάζει σε αυτό που θες τότε μπορείς να αγοράσεις ολόκληρο το πακέτο πληρώνοντας κάποιο πόσο που αναλογεί στο κάθε λογισμικό πακέτο.

Τα διάφορα λογισμικά πακέτα που κατεβάσαμε από το διαδίκτυο ανήκουν τόσο στην πρώτη κατηγορία όσο και στην δεύτερη, προφανώς δεν προτιμήσαμε να μπούμε στην διαδικασία να αγοράσουμε κάποιο λογισμικό μιας και η δοκιμαστική έκδοση του προγράμματος μας είναι ότι πρέπει για την δουλεία που πρέπει να κάνουμε, και δεν ήτανε απαραίτητο, να γίνει η αγορά

του ολοκληρωμένου πακέτου του προγράμματος, αφού θα το χρησιμοποιήσουμε για λίγο χρονικό διάστημα.

Ύστερα η επιλογή των λογισμικών πακέτων έγινε μέσα από διάφορα site όπου το καθένα είχε και κάτι διαφορετικό να προβάλλει, συνήθως τα διαφορετικά site αποτελούνται από εταιρείες που κατασκευάζουν τέτοια λογισμικά πακέτα, ο λόγος ήταν ότι τα λογισμικά πακέτα έπρεπε να είναι από διαφορετικά site με διαφορετικό περιεχόμενο του κάθε ιστοχώρου, για να μην ανήκουν στην ίδια κατηγορία, και έτσι να έχουμε μια πιο πλήρη εικόνα για τα λογισμικά πακέτα,

έτσι ώστε οποιοσδήποτε που θέλει να χρησιμοποιήσει κάποιο η κάποια λογισμικά πακέτα, η επιλογή του να μην είναι τόσο δεσμευμένη, δηλαδή να υπάρχει μια πιο μεγάλη και ξεχωριστή ποικιλία από τέτοιου είδους λογισμικών πακέτων, για να μπορεί κάποιος να διαλέξει καλύτερα το λογισμικό πακέτο που του ταιριάζει.

Για παράδειγμα μερικά από τα software, είναι από κάποιες σελίδες από εταιρίες που ασχολούνται με κατασκευή λογισμικού, κάποια αλλά software είναι από κάποιους οργανισμούς δηλαδή τους χρησιμοποιούν κάποιοι οργανισμοί, όπως πανεπιστημιακά ιδρύματα, κολέγια, υπουργεία και άλλοι φορείς και τέλος κάποια αλλά software είναι από σελίδες που σκοπό έχουν να ανεβάζουν στο διαδίκτυο ότι ελεύθερο software υπάρχει, με άμεσο σκοπό να το χρησιμοποιούν οι χρηστές του διαδικτύου και να το δουλέψουν για δικό τους σκοπό.

Εν τέλει η επιλογή των συγκεκριμένων πακέτων δεν έγινε μέσα από ένα συγκεκριμένο δικτυακό τόπο για να υπάρχει μια ελεύθερη επιλογή των λογισμικών πακέτων, δηλαδή να γίνει η σύγκριση μέσα από πολλά τέτοια πακέτα.

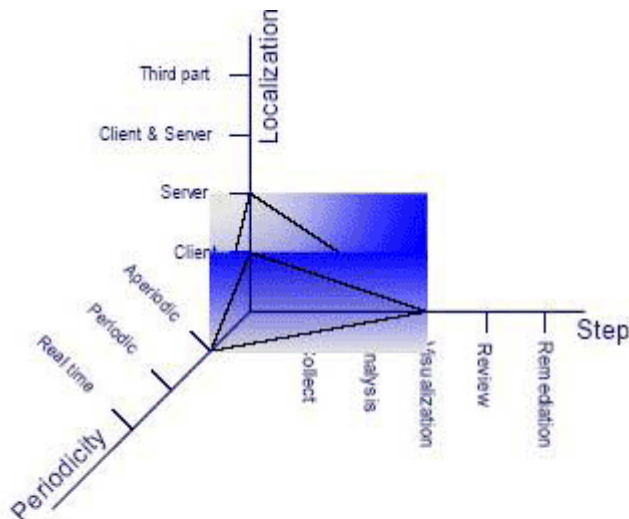


Figure 1. Classical log analyser

εικόνα 7^η Κλασικός τύπος log analyzer

Στη συνέχεια πρέπει να αναφερθούν κάποιοι τρόποι κατηγοριοποίησης και αξιολόγησης των λογισμικών πακέτων, συνήθως αυτοί οι τρόποι είναι κοινοί για όλα τα λογισμικά πακέτα, αλλά υπάρχουνε κάποιοι επιπλέον τρόποι αξιολόγησης για τέτοιους είδους software .

4.3 Βασικοί τρόποι αξιολόγησης

Η λογική με την οποία καθοριστήκανε τα κριτήρια αξιολόγησης ανήκει στην γενικότερη κατηγορία αξιολόγησης κριτηρίων και δεν αποτελεί μια ειδική περίπτωση για λογισμικά πακέτα που κάνουνε επεξεργασία log files , συνεπώς δεν διαφέρει από την γενικευμένη κατηγορία αξιολόγησης λογισμικών πακέτων, τα περισσότερα κριτήρια υπάρχουνε σε site εταιρειών που ασχολούνται με κατασκευή software όπως www.ahead.de , www.adobe.com , www.macromedia.com και άλλες

- ❖ Πρώτα από όλα ένα βασικό κριτήριο αξιολόγησης είναι οι ελάχιστες απαιτήσεις που απαιτεί ένα λογισμικό από τον υπολογιστή δηλαδή τι επεξεργαστή χρειάζεται για να δουλέψει, πόση μνήμη είναι απαραίτητη

για να λειτουργεί σωστά το πρόγραμμα, πόσο χώρο στο σκληρό δίσκο χρειάζεται, η κάρτα γραφικών τι ποιότητα να είναι και ποσά MB να υποστηρίζει, τέλος το cd rom τι ταχύτητα είναι, και σε ποιο είδους λειτουργικού συστήματος δουλεύει.

- ❖ Ένα άλλο βασικό κριτήριο που συσχετίζεται άμεσα με το προηγούμενο είναι αν και κατά πόσο εκμεταλεύεται σωστά η πλήρως το πρόγραμμα μας το hardware του υπολογιστή, δηλαδή αν χρησιμοποιεί όλη την μνήμη, αν η κάρτα γραφικών μπορεί να διαθέσει όλες τις λειτουργίες, μπορεί ο επεξεργαστής να κολλάει όταν δουλεύουμε το πρόγραμμα μας.
- ❖ Ένα άλλο σημαντικό κριτήριο είναι η τιμή του προγράμματος δηλαδή αν πρόκειται για κάποιο πρόγραμμα που είναι δωρεάν η απαιτείται να καταβάλουμε κάποιο χρηματικό πόσο , η αν το πρόγραμμα που είναι με χρηματικές δαπάνες υπάρχει κάποια έκδοση δοκιμαστική για κάποιο μικρό χρονικό διάστημα έτσι ώστε να βγάλουμε τα συμπεράσματα μας για το συγκεκριμένο πρόγραμμα ,εδώ πρέπει να αναφέρουμε ότι εμείς χρησιμοποιήσαμε προγράμματα που ήτανε δοκιμαστική έκδοση δηλαδή τα χρησιμοποιήσαμε μόνο για λίγες ημέρες ,αν θέλαμε για περισσότερο καιρό έπρεπε να γίνει η αγορά του πακέτου.
- ❖ Επίσης ένα άλλο κριτήριο για την αξιολόγηση των λογισμικών πακέτων είναι αν το πρόγραμμα διαθέτει κάποια βοήθεια που είναι να υπάρχει μέσα στο πρόγραμμα και αν υπάρχει τεχνική υποστήριξη για το συγκεκριμένο software μέσω τηλεφώνου ή μέσω ιστοσελίδας όπου να υπάρχει βοήθεια έτοιμη στην ιστοσελίδα η μέσω e-mail , η από την στιγμή της αγοράς και μετά δεν παρέχεται καμία βοήθεια , αυτό εξαρτάται από την κατασκευάστρια εταιρεία του προγράμματος.
- ❖ Ύστερα ένα άλλο κριτήριο για την αξιολόγηση του κάθε προγράμματος είναι οι επιδόσεις που έχει το κάθε πρόγραμμα δηλαδή αν κάποιο πρόγραμμα κάνει περισσότερη ώρα για να επεξεργαστεί κάποιο αρχείο τότε μπορεί να μην εκμεταλλευτεί από πολλούς διαχειριστές ιστοσελίδων, διότι να είναι χρονοβόρο στην λειτουργία του ,αυτό βέβαια

μπορούμε να το διαπιστώσουμε από την στιγμή που θα το χρησιμοποιήσουμε το πρόγραμμα και μετά , πιο πριν γίνεται μόνο αν υπάρχουν πληροφορίες στο site του προγράμματος ή σε άλλες πηγές πληροφορίας π.χ. άρθρα ,περιοδικά

- ❖ Στη συνέχεια ένα άλλο κριτήριο για την αξιολόγηση είναι οι δυνατότητες που μπορεί να μας προσφέρει ένα πρόγραμμα ,στη δικιά μας περίπτωση μπορεί ένα log analyzer να αποδίδει λιγότερες πληροφορίες από κάποιο άλλο στην έξοδο ,μετά αφού έχει γίνει η επεξεργασία των log files δηλαδή μπορεί ένα πρόγραμμα να βγάζει στην έξοδο μόνο γραφικές παραστάσεις ενώ κάποιο άλλο να δίνει τα αποτελέσματα και σε μορφή στατιστικών στοιχείων. Επίσης μπορεί κάποιο πρόγραμμα να δέχεται ένα συγκεκριμένο format από log files ενώ ένα άλλο μπορεί να δέχεται πιο πολλά format από log files τότε σίγουρα το δεύτερο έχει περισσότερες δυνατότητες από το πρώτο .
- ❖ Επίσης ένα άλλο κριτήριο έχει να κάνει με το περιβάλλον εργασίας εάν το πρόγραμμα μας είναι φιλικό προς το χρήστη η είναι πολύ δύσκολο στη λειτουργία του και πολύ περίπλοκο, επίσης αν πρόκειται για πρόγραμμα που να μπορεί να το χειριστεί κάποιος που δεν έχει μεγάλη σχέση με υπολογιστές.
- ❖ Ύστερα ένα άλλο βασικό κριτήριο αξιολόγησης είναι αν το πρόγραμμα μας δέχεται αναβάθμιση σε μια πιο καινούργια έκδοση, δηλαδή αν αγοράσουμε ένα λογισμικό πακέτο και μετά από ένα χρόνο βγει μια καινούργια, τότε θα μπορούμε να κάνουμε ενημερώσει στο ήδη υπάρχον πρόγραμμα η θα πρέπει να εγκαταστήσουμε από την αρχή τη νέα έκδοση μας,
- ❖ Τέλος ένα άλλο βασικό κριτήριο είναι η ασφάλεια που παρέχει το λογισμικό πακέτο μας, δηλαδή να μην υπάρχουνε προβλήματα στο πρόγραμμα μας αν κάποιος ιός μολύνει τον υπολογιστή μας ,να μην καταστρέψει το πρόγραμμα μας σε περίπτωση εξάπλωσης του ιού, π.χ. μπορεί κάποιος Hacker αν καταφέρει και εισβάλει στον

υπολογιστή να κλέψει στοιχεία που μπορεί να είναι πολύ χρήσιμα για την εταιρεία μας με άμεσο αποτέλεσμα οι πληροφορίες να διερευσουνε έξω από την εταιρεία, έτσι ώστε κάποιες ανταγωνιστικές εταιρείες να βρουν τα ξένα για αυτούς στοιχεία και να τα χρησιμοποιήσουνε εναντίον της θυγατρικής εταιρείας.

4.4 Προβλήματα (διαδικαστικά)

κατά τη διαδικασία της υλοποίησης της πτυχιακής παρουσιάστηκε κάποια προβλήματα όπου καταφέραμε να τα επιλύσουμε μετά από αρκετές προσπάθειες. Ένα πρόβλημα ήταν που τα περισσότερα λογισμικά πακέτα τα κατεβάζαμε από μια απλή γραμμή PSTN και με ένα modem 56 kbps, έτσι ο χρόνος κατά την διαδικασία του κατεβάσματος ήταν πολύ αργός σε σχέση με την τεχνολογία που υπάρχει αφού μπορούμε να κατεβάσουμε και με μια γραμμή ISDN με 64 kbps ή με μια ADSL, με 384 kbps, αυτό είχε ως αποτέλεσμα να γίνεται ιδιαίτερα κουραστικό και να απαιτείται πολύ μεγαλύτερο για τα προγράμματα που είναι μεγάλα σε χώρο, δηλαδή προγράμματα που ήταν αρκετά MBytes, επίσης ένα άλλο πρόβλημα που παρουσιάστηκε είναι ότι όλα τα αρχεία που βρήκαμε ήταν στην αγγλική γλώσσα και έτσι αναγκαζόμασταν να τα μεταφράζουμε στην ελληνική για να μπορούμε να καταλαβαίνουμε τι ακριβώς γίνεται, οτιδήποτε βιβλιογραφία είχαμε βρει στο διαδίκτυο ήταν στα αγγλικά γραμμένη με αποτέλεσμα να γίνεται πάντα μια προεπεξεργασία όπου συνήθως ήταν μια μετάφραση στα ελληνικά, και μετά η επεξεργασία του συγκεκριμένου υλικού ιδίως βιβλία σε ηλεκτρονική μορφή στο διαδίκτυο.

4.5 Σύντομη περιγραφή software

Πριν αρχίσουμε να κάνουμε μια αναφορά στο κάθε λογισμικό ξεχωριστά πρέπει να επισημάνουμε ότι το κάθε πρόγραμμα το εγκαταστήσαμε και το τρέξαμε πάνω σε υπολογιστή με επεξεργαστή 1,7 GHz amd athlon xp, μνήμη

256 DDR,κάρτα γραφικών με τσιपाκι των 64MB,σκληρο δίσκο 40 GB και με λειτουργικό σύστημα microsoft XP,

βέβαια σε αυτό το σημείο πρέπει να επισημάνουμε ότι η αναφορά που γίνεται παρακάτω είναι μια μικρή και δεν αποτελεί την πλήρη ανάλυση και αξιολόγηση των software,η οποία θα γίνει στο επόμενο κεφαλαίο και θα μπορεί να δώσει στον αναγνώστη πλήρη εικόνα του καθενός λογισμικού πακέτου έτσι ωστε,να αποφασίσει πιο θα χρησιμοποιήσει για να κάνει την ανάλυση των log files που χρειάζεται, η οποία θα γίνει με τους τρόπους αξιολόγησης που αναφέραμε πιο πάνω,

τα προγράμματα που κατεβάσαμε και εγκαταστήσαμε στον υπολογιστή μας είναι τα ακόλουθα

- **123loganalyzer** πρόκειται για ένα πρόγραμμα που καταλαμβάνει μικρό χώρο στο δίσκο σχεδόν 2,3 MB,αυτη είναι μια έκδοση trial και με αρκετά εύκολο περιβάλλον εργασίας, μπορείτε να κατεβάσετε την έκδοση από την ιστοσελίδα τους www.loganalyzer.com η εγκατάσταση του προγράμματος είναι εύκολη, επίσης η συγκεκριμένη έκδοση του προγράμματος διαθέτει και βοήθεια για όσους την χρειάζονται.
- **absolute log analyzer** πρόκειται για ένα πρόγραμμα που καταλαμβάνει 14,50MB χώρο στο δίσκο ,πρόκειται για ένα πρόγραμμα που η έκδοση του είναι demo, παρέχει αρκετές δυνατότητες, κατά την διάρκεια της εξεργασίας των log files, επίσης η συγκεκριμένη έκδοση διαθέτει ένα πολύ εύκολο help,όσοι θέλουνε μπορούνε να το κατεβάσουνε από την www.birstrike.com/analyzer/.
- **absolute log analyzer pro** ,το συγκεκριμένο πρόγραμμα είναι το ίδιο με το προηγούμενο μόνο που η διάφορα τους είναι ότι το συγκεκριμένο πρόγραμμα η έκδοση του είναι professional δηλαδή παρέχει καποιες δυνατότητες παραπάνω από ότι το προηγούμενο, Πάρο όλα αυτά έχει

τα ίδια τεχνικά χαρακτηριστικά με την απλή demo έκδοση και οποίος επιθυμεί να μπορεί να το κατεβάσει από την ίδια σελίδα στο διαδίκτυο.

- **advanced log analyzer** πρόκειται για ένα μικρό πρόγραμμα με πολύ εύκολο περιβάλλον εργασίας, καταλαμβάνει γύρω στα 2,8 MH χώρο στο δίσκο, και με αρκετές δυνατότητες για ένα πρόγραμμα που είναι ελεύθερο στην αγορά ,χωρίς να χρειάζεται να το αγοράσει κανείς, όσοι ενδιαφέρονται μπορούν να το κατεβάζουνε,απο την σελίδα τους www.abacre.com.
- **alter wind log analyzer professional** πρόκειται για ένα λογισμικό πακέτο που είναι σε trial έκδοση και έχει πολύ εύκολο περιβάλλον εργασίας για professional έκδοση,ο χώρος που καταλαμβάνει στον δίσκο είναι μόνο 3,36MB ,αποτελεί μια απλή και αξιόπιστη λύση για επεξεργασία δεδομένων και μπόγε κάποιος να το κατεβάσει από την σελίδα www.alterwind.com.
- **azure web log** πρόκειται για ένα πρόγραμμα που αναλυει τα log files,καταλαμβάνει χώρο στο δίσκο μόνο 0,75MB πρόκειται για πρόγραμμα με αρκετές δυνατότητες, και η έκδοση του είναι ελεύθερη στο διαδίκτυο,μπορεί να κάποιος να την κατεβάσει από την σελίδα www.azuredesktop.com.
- **faststats analyzer** πρόκειται για ένα πρόγραμμα με πολύ απλό περιβάλλον εργασίας, η συγκεκριμένη έκδοση είναι demo, καταλαμβάνει μόνο 2,74 MH χώρο στο δίσκο και πρόκειται για μια απλή λύση που μπορεί να επεξεργαστεί αρκετά καλά,καποια log files.οποίος ενδιαφέρεται μπορεί να το κατεβεί από την συγκεκριμένη ιστοσελίδα www.mach5.com.
- **openwebscope web statistics** πρόκειται για μια έκδοση, η οποία είναι professional και παράλληλα είναι trial, η συγκεκριμένη έκδοση έχει πολύ απλό περιβάλλον και ευχάριστο, καταλαμβάνει μόλις 1,01MB

χώρο στο δίσκο και αποτελεί μια καλή λύση για επεξεργασία log files, οι ενδιαφερόμενοι μπορούν να την κατεβάσουν απτήν ιστοσελίδα www.openwebscope.com.

- **web log expert lite** πρόκειται ένα λογισμικό πακέτο που η συγκεκριμένη έκδοση είναι πολύ μικρή ,διαθέτει τις απαραίτητες δυνατότητες για να γίνει μια επεξεργασία log files, καταλαμβάνει 2,30MB χώρο στο δίσκο και το περιβάλλον εργασίας είναι πολύ απλό μιας και είναι τετοια η έκδοση του, όσοι ενδιαφέρονται μπορούν να το κατεβάζουν από την ιστοσελίδα www.weblogexpert.com.

Σε αυτό το σημείο πρέπει να αναφέρουμε ότι επιλέξαμε να συγκρίνουμε μόνο εννέα λογισμικά πακέτα γιατί ο συγκεκριμένος αριθμός δεν είναι ούτε πολύ λίγα ούτε και πάρα πολλά ,δηλαδή πιστεύω ότι ο αριθμός αυτός είναι αρκετά ικανοποιητικός για κάποιον που θέλει να επιλέξει ένα τέτοιο λογισμικό πακέτο.

κεφάλαιο 5 Αξιολόγηση λογισμικών πακέτων

εισαγωγή

Σε αυτό το κεφαλαίο θα παρουσιάσουμε την σελίδα την οποία θα δουλέψουμε τα log files, επίσης σε αυτό κεφαλαίο θα αναλύσουμε πλήρως τα λογισμικά πακέτα, που θα χρησιμοποιήσουμε σύμφωνα με τα κριτήρια αξιολόγησης που έχουμε αναφέρει στο προηγούμενο κεφαλαίο, επίσης θα αναφέρουμε τα πλεονεκτήματα και τα μειονεκτήματα του καθενός προγράμματος, στην συνέχεια θα τρέξουμε για κάθε πρόγραμμα με τα log files, που έχουμε στα χέρια μας και εκεί θα εξάγουμε τα αποτελέσματα από την ανάλυση των log files, όπου θα βγάλουμε και κάποια επιπλέον χρήσιμα συμπεράσματα για τα προγράμματα που αξιολογούμε.

5.1 Παρουσίαση σελίδας

Στη συνέχεια θα παρουσιάσουμε την σελίδα της οποίας χρησιμοποιούμε το αρχείο που θα γίνει η αξιολόγηση των λογισμικών πακέτων .Η συγκεκριμένη σελίδα πρόκειται για μια σελίδα όπου παρουσιάζει την όλη επιχείρηση και μια γενική παρουσίαση του συγκεκριμένου προγράμματος που εμπορεύεται. Πρόκειται για την σελίδα www.infowood.gr, σε αυτή την συγκεκριμένη σελίδα μπορούμε να βρούμε κάποια απλά παρουσιαστικά πράγματα που παρουσιάζουν το πρόγραμμα τους και την εταιρεία τους, επίσης μπορεί κάποιος να βρει κάποια πράγματα που βοηθάνε τους πελάτες της συγκεκριμένης εταιρείας, όπως λήψη αρχείων ,ομάδες συζητήσεων , συχνές ερωτήσεις. Επίσης μπορεί κάποιος να στείλει μέσω e-mail διευκρινιστικές ερωτήσεις πάνω στο πρόγραμμα και να ζητήσει τεχνική υποστήριξη μέσω της συγκεκριμένης ιστοσελίδας από κάποιες συχνές ερωτήσεις-απαντήσεις. Τέλος οποίος επιθυμεί μπορεί να βρει κάποια στοιχεία που θα τον βοηθήσουν και θα τον πείσουν να αγοράσει το πρόγραμμα τους, βέβαια για

περισσότερες πληροφορίες μπορούμε να επισκεφτούμε το δικτυακό τόπο στην παρακάτω σελίδα www.infowood.gr



εικόνα 8^η Κεντρική σελίδα www.infowood.gr

5.2 Αξιολόγηση των λογισμικών πακέτων

Στη συνέχεια θα αρχίσουμε να αξιολογούμε τα λογισμικά πακέτα σύμφωνα με τους τρόπους αξιολόγησης που έχουμε αναφέρει στο προηγούμενο κεφαλαίο,

5.2.1 123logalyzer version 3.10

- Το πρώτο software είναι το 123logalyzer version 3.10, προκειται για ένα απλό λογισμικό πακέτο με πολλές δυνατότητες στην επεξεργασία log files , το συγκεκριμένο πακέτο έχει πολλές μικρές απαιτήσεις από θέμα υπολογιστή, καταλαμβάνει πολύ μικρό χώρο στο δίσκο και το install του είναι πολύ μικρό σε MBytes .

- Εκμεταλλεύεται σε αρκετό μεγάλο βαθμό το hardware, αν και το συγκεκριμένο προϊόν δουλεύει ακόμα και σε αρκετά παλιούς υπολογιστές, συνεπώς δεν υπάρχει πρόβλημα από θέμα hardware.
- Από θέμα τιμής το συγκεκριμένο πακέτο έχει \$ 60 , η οποία τιμή είναι ικανοποιητική για όλα αυτά που προσφέρει το πρόγραμμα .
- Η συγκεκριμένη έκδοση παρόλο που είναι trial,προσφέρει ένα αρκετό ικανοποιητικό help του προγράμματος που μπορεί να βοηθήσει τον καθένα που έχει πρόβλημα.
- Το συγκεκριμένο πρόγραμμα έχει αρκετές δυνατότητες όπως μπορεί να αναλύει αρχεία τα οποία τα οποία σε μορφή .zip και μπορεί να σώζει τα αρχεία στο δίσκο με απλό τρόπο, επίσης μπορεί να αναλύουμε τα αρχεία και από command line,μόνο σε μεγάλα αρχεία χρειάζεται περισσότερη μνήμη.
- Το περιβάλλον εργασίας του προγράμματος μπορεί κάποιος να αναφέρει ότι είναι πολύ φιλικό με πολύ απλή διαχείριση των ενεργειών.
- Όσον αφορά την αναβάθμιση του προγράμματος το συγκεκριμένο πακέτο δέχεται αναβάθμιση μόνο που κοστίζει κάποια επιπλέον χρήματα αλλά σαφώς λιγότερα από ότι η κανονική έκδοση.
- Όσον αφορά την ασφάλεια το συγκεκριμένο πρόγραμμα έχει μια ασφάλεια αρκετή καλά, εφόσον υπάρχει και η αναμενόμενη από το λειτουργικό σύστημα πάνω σε αντιμετώπιση στους ιούς.

5.2.2 Absolute Log Analyzer Demo v 2.33

- Ένα άλλο λογισμικό πακέτο είναι το Absolute Log Analyzer Demo v 2.33 πρόκειται για μια έκδοση demo του προγράμματος, η οποία δεν

απαιτεί και παρά πολλές μεγάλες απαιτήσεις από θέμα hardware του υπολογιστή ,καταλαμβάνει αρκετά μεγάλο χώρο στο δίσκο .

- Το συγκεκριμένο πακέτο εκμεταλεύεται αρκετά καλά το hardware του υπολογιστή και σε βαθμό τέτοιο που να είναι αρκετός ένας υπολογιστής με μικρές δυνατότητες του hardware ,όπου περιέχονται και κάποιο εγχειρίδιο χρήσης για καλύτερη βοήθεια .
- Από θέμα τιμής το συγκεκριμένο πακέτο έχει \$ 65 η οποία είναι αρκετή ικανοποιητική για όλες τις δυνατότητες που μας προσφέρει.
- Η συγκεκριμένη έκδοση αποτελεί μια demo έκδοση και πάρο όλα αυτά έχει ένα ικανοποιητικό help που μπορεί να βοηθήσει αρκετά προβλήματα σε κάποιον που χρησιμοποιεί το πρόγραμμα.
- Οι δυνατότητες του προγράμματος είναι αρκετές, συγκεκριμένα με αυτό το πρόγραμμα μπορούμε να φιλτράρουμε τα περιττά στοιχεία των αρχείων, επίσης μπορούμε να χρησιμοποιήσουμε ένα περιττό στοιχείο στο αρχείο μας, αλλά μόνο ένα επειδή η έκδοση είναι trial , μπορούμε να τρέξουμε το αρχείο μας κατευθείαν από τον διακομιστή μας, μπορούμε να βρούμε συνδέσεις οι οποίες είναι σπασμένες, δηλαδή δεν έχουν ολοκληρωθεί, τέλος μπορούμε να εξαγάγουμε τα συμπεράσματα σε περισσότερα από μια μορφή όπως html αλλά και σε excel .
- Το περιβάλλον εργασίας είναι αρκετό απλό με πολλές δυνατότητες σε οπτικό περιβάλλον και μπορεί να χρησιμοποιηθούνε μπορεί εύκολα , με μια μπάρα εργασιών στο δεξιό μέρος του προγράμματος.
- Το συγκεκριμένο πρόγραμμα δέχεται αναβάθμιση και μπορεί να αναβαθμιστεί σχετικά εύκολα.
- Η ασφάλεια του συγκεκριμένου λογισμικού πακέτου είναι αρκετή καλή για μια έκδοση η οποία δεν είναι επαγγελματική πρόκειται δηλαδή για

απλή έκδοση του προγράμματος και πάρο όλα αυτά είναι και demo έκδοση.

5.2.3 Absolute Log Analyzer Pro Demo v 2.33

- Το επόμενο λογισμικό πακέτο είναι το Absolute Log Analyzer Pro Demo v 2.33 ,πρόκειται για ένα πακέτο που είναι όπως το προηγούμενο μόνο που πρόκειται για μια επαγγελματική έκδοση ,οι ελάχιστες απαιτήσεις είναι κάπως πιο αυξημένες από ότι ένα άλλο πρόγραμμα που δεν είναι επαγγελματικό .
- Μπορούμε να πούμε ότι εκμεταλευται σε μεγάλο βαθμό το υλικό μέρος του προγράμματος μιας και πρόκειται για μια έκδοση η οποία είναι professional .
- Από θέμα τιμής το συγκεκριμένο πακέτο έχει \$ 75 η τιμή είναι λογική για τις δυνατότητες και όλα αυτά που προσφέρει το πρόγραμμα.
- Η συγκεκριμένη έκδοση έχει ένα πλήρες help όπου μπορεί να σου προσφέρει βοήθεια από την αρχή μέχρι και το τέλος της ανάλυσης των αρχείων.
- Οι δυνατότητες του προγράμματος είναι οι εξής: μπορούμε να επεξεργαστούμε τα δεδομένα μας και από γραμμή εντολών, επίσης μπορούμε να χρησιμοποιήσουμε τα αρχεία κατευθείαν από τον διακομιστή ,χωρίς να τα μεταφέρουμε στο δικό μας υπολογιστή, επίσης μας επιτρέπεται το φιλτράρισμα των δεδομένων αυτών που δεν θέλουμε να χρησιμοποιήσουμε για επεξεργασία .
- Το περιβάλλον εργασίας είναι αρκετά απλό όπου στην δεξιά πλευρά του προγράμματος κάποιες συνδέσεις του προγράμματος και υπάρχουνε στην πάνω μπάρα του προγράμματος .

- Για το θέμα της αναβάθμισης το συγκεκριμένο πρόγραμμα αποτελεί μια καλή λύση στην περίπτωση που θέλουμε να αναβαθμίσουμε το ήδη υπάρχον πρόγραμμα μας .
- Όσον αφορά την ασφάλεια του προγράμματος η συγκεκριμένη έκδοση προσφέρει αρκετά καλή ασφάλεια μιας και οι απαιτήσεις του προγράμματος απαιτούνε καλή ασφάλεια για την καλή και αξιόπιστη λειτουργία του προγράμματος μας .

5.2.4 Advanced Log Analyzer v 1.5

- Το ακόλουθο λογισμικό είναι το Advanced Log Analyzer v 1.5 πρόκειται για ένα πρόγραμμα που μπορεί να χρησιμοποιηθεί με ελάχιστες απαιτήσεις υλικού και να μην αναγκάζει τον χρήστη να χρησιμοποιεί hardware πολύ εξειδικευμένο.
- Βέβαια το συγκεκριμένο λογισμικό πακέτο χρησιμοποιεί όλο το υλικό του υπολογιστή και μάλιστα αρκετά καλά για μια τέτοια έκδοση του προγράμματος.
- Από θέμα τιμής το συγκεκριμένο πακέτο έχει \$ 70 τιμή η οποία είναι αρκετά καλή για ένα τέτοιο πρόγραμμα .
- Η βοήθεια που προσφέρει το συγκεκριμένο πρόγραμμα είναι αρκετά καλή και προσφέρει όλο το φάσμα που μπορεί να προσφέρει ένα help ενός προγράμματος, με καλά δομημένο τρόπο και αρκετή ανάλυση.
- Οι δυνατότητες που μπορεί να προσφέρει το συγκεκριμένο πρόγραμμα είναι να εξάγει τα συμπεράσματα σε πολλές μορφές. Επίσης μπορεί να διαλέξει τα αρχεία από πολλούς διακόμισες ξεχωριστά και μπορεί να κάνει αναφορά σε κάθε link της σελίδας. Τελος υπάρχει ερωτηματολόγιο του προγράμματος που γίνεται on line.

- Το περιβάλλον εργασίας του προγράμματος είναι αρκετά δομημένο από την δεξιά πλευρά του προγράμματος να υπάρχουν κάποιες συνδέσεις που καθοδηγούνε τις κινήσεις κατά την επεξεργασία των αρχείων.
- Το πρόγραμμα μπορεί να υποστηρίξει και αναβάθμιση , αφού πρώτα έχει γίνει η αγορά του προγράμματος μας μιας και η έκδοση είναι trial .
- Όσον αφορά την ασφάλεια αυτή εξαρτάται κατά ένα μεγάλο ποσοστό στην ήδη υπάρχουσα ασφάλεια του λειτουργικού συστήματος.

5.2.5 AlterWind LogAnalyzer Professional v2.1

- Στη συνέχεια το επόμενο πρόγραμμα είναι το AlterWind LogAnalyzer Professional v2.1, προκειται για ένα πολύ απλό και πολύ αποτελεσματικό πρόγραμμα κατά την διάρκεια της επεξεργασίας τέτοιων αρχείων που ουσιαστικά χρησιμοποιεί πολύ μικρές απαιτήσεις από τον υπολογιστή μας και δεν υπάρχει πρόβλημα από θέμα hardware του υπολογιστή μας.
- Το συγκεκριμένο πρόγραμμα μπορούμε να πούμε ότι εκμεταλευται αρκετά το υλικό μέρος του υπολογιστή αν σκεφτούμε ότι πρόκειται για ένα απλό πρόγραμμα και δεν είναι απαραίτητος υπολογιστής με μεγάλες δυνατότητες.
- Από θέμα τιμής το συγκεκριμένο πακέτο έχει \$ 79 η τιμή του οποίου είναι αρκετά ικανοποιητική με βάση αυτά που μπορεί να μας προσφέρει .

- Από θέμα βοήθειας το συγκεκριμένο πακέτο εκτός από την καθιερωμένη βοήθεια που προσφέρει μπορεί να προσφέρει και βοήθεια από την ιστοσελίδα τους μέσω web η ακόμα και μέσω e-mail.
- Οι δυνατότητες που μπορεί να προσφέρει το συγκεκριμένο πρόγραμμα είναι κοινές με άλλα προγράμματα όπως ,να φιλτράρει τα στοιχεία των αρχείων, να κάνει εξόρυξη από διαφορετικούς διακόμιστες με διαφορετικά μορφής αρχεία, να επεξεργάζεται αρχεία με κατάληξη .zip .gz .
- Το περιβάλλον εργασίας είναι πολύ απλό και πολύ ξεκούραστο αφού δεν περιέχει παρά μόνο τα απαραίτητα στοιχεία για να γίνει μια ανάλυση δεδομένων και όχι περιττά στοιχεία.
- Η συγκεκριμένη έκδοση είναι trial και δέχεται αναβάθμιση αν και είναι πρόσφατη από εμάς ημερομηνίας έκδοσης της.
- Όσο για την ασφάλεια του προγράμματος είναι αρκετά ικανοποιητική αφού μπορεί να εξασφαλίσει κάποια καλά αποτελέσματα από την δουλεία μας χωρίς να υπάρχει πρόβλημα από ιούς που μπορεί να υπάρχουνε στον υπολογιστή μας και βεβαίως εξαρτάται και από την ασφάλεια του λειτουργικού συστήματος μας.

5.2.6 Azure Web Log 1.4

- Το επόμενο λογισμικό πακέτο είναι το Azure Web Log 1.4,προκειται για ένα πρόγραμμα που μπορεί να το δουλέψουμε και σε ελάχιστες απαιτήσεις και δεν χρειάζεται και σε ιδιαίτερο μεγάλο υπολογιστή με μεγάλες δυνατότητες για να δουλέψει.
- Όσον αφορά για το hardware μπορεί να το εκμεταλευτεί κατάλληλα και να καταφέρει να αντεπεξέλθει στην ανάλυση και επεξεργασία των

συγκεκριμένων αρχείων χωρίς να υπάρχουνε προβλήματα σε σχέση με το υλικό μέρος του υπολογιστή.

- Συμπεράσματα του για το hardware στο βασικό πρόγραμμα σε ένα σημείο του προγράμματος,
- Από θέμα τιμής το συγκεκριμένο πακέτο έχει \$ 65 η τιμή είναι ικανοποιητική για όλα αυτά που μας προσφέρει το πρόγραμμα.
- Η βοήθεια που περιλαμβάνει το πακέτο έχει και βοήθεια μέσω διαδικτυου και μέσω e-mail,πέρα από την συνηθισμένη βοήθεια που μπορεί να προσφέρει ένα πρόγραμμα.
- Οι δυνατότητες που προσφέρει το συγκεκριμένο πρόγραμμα είναι αρκετές για να αποτελέσει μια καλή επιλογή από κάποιον ενδιαφερόμενο, μπορεί κάποιος να δει όλα τα αποτελέσματα μέσα από το πρόγραμμα χωρίς να χρειάζεται να γίνει εξαγωγή σε άλλες μορφές όπως html και excel .
- Το περιβάλλον εργασίας είναι πολύ λιτό και πολύ απλό αφού δεν υπάρχει κανένα πρόβλημα δυσκολίας στο χειρισμό του συγκεκριμένου προγράμματος.
- Στο συγκεκριμένο πρόγραμμα προφανώς θα υπάρχει η δυνατότητα αναβάθμιση, μιας και πρόκειται για ένα πρόγραμμα που η έκδοση του είναι demo και δεν αποτελεί μια πλήρης έκδοση.
- Όσον αφορά την ασφάλεια του προγράμματος σε ενδεχόμενους ιούς, προφέρει μια αρκετή καλή ασφάλεια μιας και είναι για demo έκδοση το συγκεκριμένο πρόγραμμα.

5.2.7 FastStats Analyzer v 2.81

- Το επόμενο software είναι το FastStats Analyzer v 2.81, το οποίο πρόκειται για ένα απλό λογισμικό πακέτο που δεν απαιτεί και παρά πολύ μεγάλες απαιτήσεις από έναν υπολογιστή .
- Όσον αφορά το hardware η εκμετάλλευση από το πρόγραμμα γίνεται σε μεγάλο βαθμό σε σχέση με τις δυνατότητες του συγκεκριμένου πακέτου.
- Από θέμα τιμής το συγκεκριμένο πακέτο έχει 85 \$ τιμή η οποία ίσως είναι όχι και τόσο ικανοποιητική για όλα αυτά που μας προσφέρει .
- Από θέμα βοήθειας το συγκεκριμένο πακέτο μπορεί να προσφέρει μια βασική βοήθεια για ένα τέτοιο απλό πρόγραμμα ,μιας και η χρήση του είναι πολύ απλή και πολύ εύκολη και δεν σου δίνει την δυνατότητα για βοήθεια μέσω internet ή μέσω e-mail .
- Οι δυνατότητες του προγράμματος είναι εκτός από μια απλή εξεργασία αρχείων ,μπορεί ταυτόχρονα να επεξεργαστεί και αλλά ήδη format αρχείων, και με πολλές επιλογές κατά την έξοδο της ανάλυσης των αρχείων.
- Το περιβάλλον εργασίας ουσιαστικά ανήκει σε δυο παράθυρα τα οποία, το πρώτο ανήκει στην εισαγωγή των δεδομένων και στο δεύτερο γίνεται η επεξεργασία των δεδομένων μας, αλλά παρόλο αυτά μπορούμε να πούμε ότι δεν είναι ένα δύσκολο περιβάλλον εργασίας, μιας και ο χρήστης μπορεί εύκολα να κάνει την επεξεργασία χωρίς ιδιαίτερα προβλήματα .
- Όσον αφορά την αναβάθμιση του προγράμματος αυτή είναι μια περίπτωση που γίνεται αρκετά εύκολα μιας και υπάρχει σχετική ευκολία από το πρόγραμμα .

- Τέλος για την ασφάλεια αυτή είναι ικανοποιητική από θέμα ότι πρόκειται για μια έκδοση demo και δε έχουμε και πολλές απαιτήσεις από ένα τέτοιο πρόγραμμα.

5.2.8 OpenWebScope Web Statistics v 1.01

- Στη συνέχεια το επόμενο software είναι το OpenWebScope Web Statistics v 1.01 πρόκειται για ένα λογισμικό πακέτο που στην πραγματικότητα οι απαιτήσεις που έχει από ένα υπολογιστή είναι πολύ μικρές ,έτσι δεν υπάρχει πρόβλημα από θέμα hardware του υπολογιστή .
- Για το θέμα του hardware μπορούμε να πούμε ότι η εκμετάλλευση του είναι ικανοποιητική σε μεγάλο βαθμό, και δεν έχουμε και μεγάλες απαιτήσεις από το υλικό του υπολογιστή μιας και δεν χρειάζεται κάτι τέτοιο από το ίδιο το πρόγραμμα.
- Από θέμα τιμής το συγκεκριμένο πακέτο έχει \$ 99 τιμή η οποία δεν απεικονίζεται στις δυνατότητες του προγράμματος.
- Από θέμα βοήθειας το συγκεκριμένο λογισμικό πακέτο διαθέτει online βοήθεια εκτός από την κλασική βοήθεια που έχουν τα περισσότερα λογισμικά πακέτα δηλαδή μέσα από μια μπάρα ή από ένα κουμπί που θα έχει σύνδεση για την βοήθεια του προγράμματος .
- Οι δυνατότητες του προγράμματος είναι συνήθως οι ίδιες με τα περισσότερα προγράμματα ανάλυσης log files,το συγκεκριμένο πακέτο προσφέρει αρκετά πλεονεκτήματα και ευκολίες στην επεξεργασία τέτοιων αρχείων, μπορεί να εξάγει τα συμπεράσματα σε μορφή html .

- Το περιβάλλον εργασίας είναι ένα εύκολο σε χρήση από οποιονδήποτε που δεν διαθέτει κάποιες ιδιαίτερες γνώσεις πάνω σε υπολογιστές.
- Όσον αφορά το θέμα της αναβάθμισης ένα επιπλέον στοιχείο που έχει αυτό το πρόγραμμα είναι ότι μπορεί να κάνει ενημέρωση του προγράμματος και μέσω διαδικτυου, και αυτοματα από το πρόγραμμα χωρίς κάποιος να μπορεί να μπει σε ιδιαίτερες λεπτομέρειες και μεγάλες διαδικασίες.
- Όσο για την ασφάλεια αυτή μπορούμε ότι είναι σε αρκετό καλό επίπεδο αν σκεφτεί κανείς ότι μπορεί να κάνει ενημερώσει το πρόγραμμα μέσω διαδικτυου με πολύ απλοποιημένο τρόπο .

5.2.9 WebLog Expert Lite v 3.0

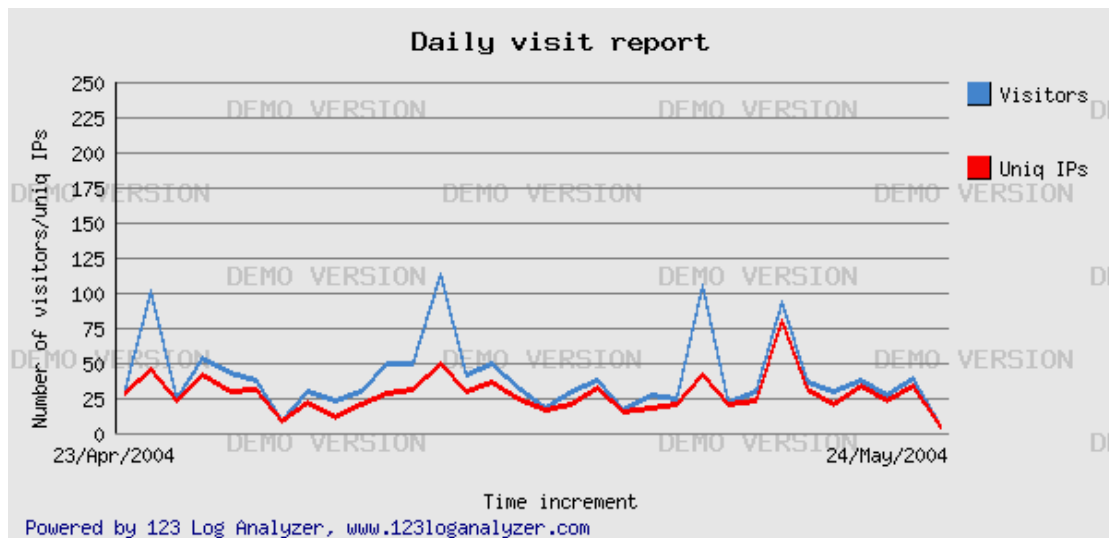
- Το αμέσως επόμενο πρόγραμμα και τελευταίο είναι το WebLog Expert Lite v 3.0 ,είναι ένα πρόγραμμα που η μορφή του είναι τόσο απλή έτσι ώστε οι απαιτήσεις να είναι πολύ ελάχιστες σε σχέση με κάποια αλλά που κάνουνε ακριβώς την ίδια δουλεία με το συγκεκριμένο πρόγραμμα.
- Η εκμετάλλευση του hardware γίνεται με παρά πολύ καλό τρόπο ,έτσι ώστε να μην υπάρχουνε καμία μέρη του υπολογιστή που δεν χρησιμοποιούνται σε αυτό το πρόγραμμα.
- Από θέμα τιμής το συγκεκριμένο πακέτο έχει \$ 75 τιμή ικανοποιητική για ένα τέτοιο πρόγραμμα.
- Η βοήθεια που προσφέρει το πρόγραμμα είναι σχετικά λιτή μιας και η έκδοση είναι λιτή άρα δεν έχουμε και παρά πολλές απαιτήσεις από την συγκεκριμένη βοήθεια του προγράμματος.

- Οι δυνατότητες που έχει το συγκεκριμένο πακέτο δεν είναι τίποτα άλλο από κάποιες συνηθισμένες, όπως να επεξεργάζεται αρχεία από διαφορετικούς διακομιστές ,να αναλύει αρχεία από με διαφορετική επέκταση όπως .zip και .gz .
- Το περιβάλλον εργασίας είναι ένα τελείως απλό περιβάλλον software με τα απολύτως απαραίτητα στοιχεία για επεξεργασία log files, ίσως από τα πιο απλά περιβάλλον εργασίας.
- Το πρόγραμμα δέχεται αναβάθμιση μιας και πρόκειται για μια έκδοση που είναι λιτή αυτό σημαίνει ότι η απλή έκδοση θα έχει κάποια επιπλέον στοιχεία.
- Όσον αφορά το θέμα της ασφάλειας το πακέτο αυτό στηρίζει πολύ την ασφάλεια στο ίδιο το λειτουργικό σύστημα.

5.3 Εκτέλεση log files στα λογισμικά πακέτα

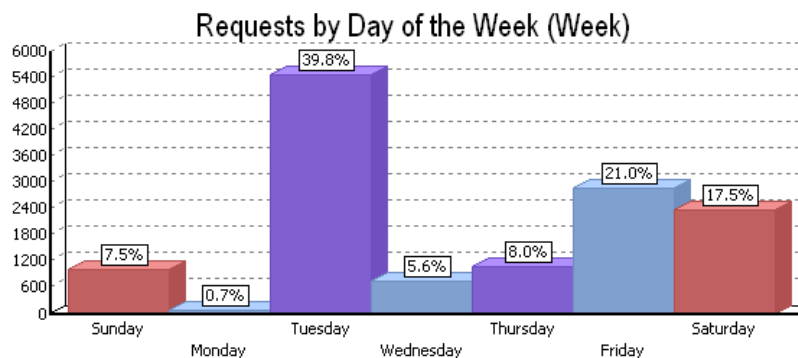
Στη συνέχεια αφού έχουμε περιγράψει το κάθε πρόγραμμα αναλυτικά και έχουμε αναφερθεί για τα πιο σημαντικά πράγματα για το κάθε πρόγραμμα , θα εκτελεί το κάθε πρόγραμμα με το log files που έχουμε στα χέρια μας και έχουμε αναφερθεί σε προηγούμενο κεφαλαίο μας.

Το πρώτο λογισμικό πακέτο που τρέξαμε το log file είναι το [123loganalyzer version 3.10](#) το συγκεκριμένο πακέτο εξάγει τα συμπεράσματα σε html μορφή αλλά και σε excel μορφή, η αναφορά που έχουμε από το συγκεκριμένο πακέτο μας δίνει μια πλήρη αναφορά για το ιστοχώρο μας και μας παρέχει αρκετές πληροφορίες για να εξάγουμε τα δικά μας συμπεράσματα μας, επίσης οι πληροφορίες μας παρέχονται σε πίνακα αλλά και σε γραφικές παραστάσεις που περιέχουν τα log files,



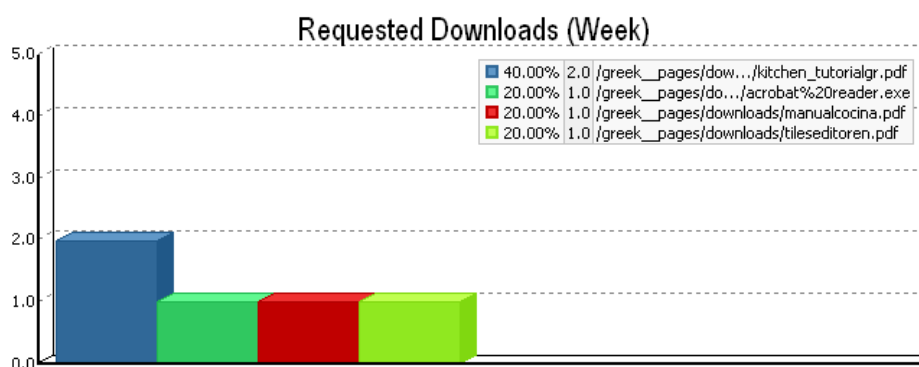
εικόνα 9^η εδώ απεικονίζεται η καθημερινή επισκεψιμότητα της ιστοσελίδας ,από αριστερά υπάρχει ο αριθμός των επισκεπτών όπου παρουσιάζεται με τη μπλε γραμμή ενώ με τη κόκκινη γραμμή παρουσιάζεται ο αριθμός των μοναδικών IP διευθύνσεων,

Το επόμενο λογισμικό πακέτο που τρέξαμε το Absolute Log Analyzer Demo v 2.33, το συγκεκριμένο πρόγραμμα εξάγει τα συμπεράσματα του σε ένα δικό του format και κρατάει όλα τα στοιχεία μέσα σε ένα συγκεκριμένο αρχείο που για να το διαβάσουμε πρέπει να το ανοίξουμε μέσα από το συγκεκριμένο πακέτο, αλλά ταυτόχρονα μας δίνει την δυνατότητα να κάνουμε εξαγωγή μέσα από το πρόγραμμα μας σε μορφή html που μας δίνει πολλά πλεονεκτήματα για την επεξεργασία των αποτελεσμάτων από την όλη διαδικασία όπως μειώνεται η πιθανότητα σφάλματος από το ίδιο το πρόγραμμα.



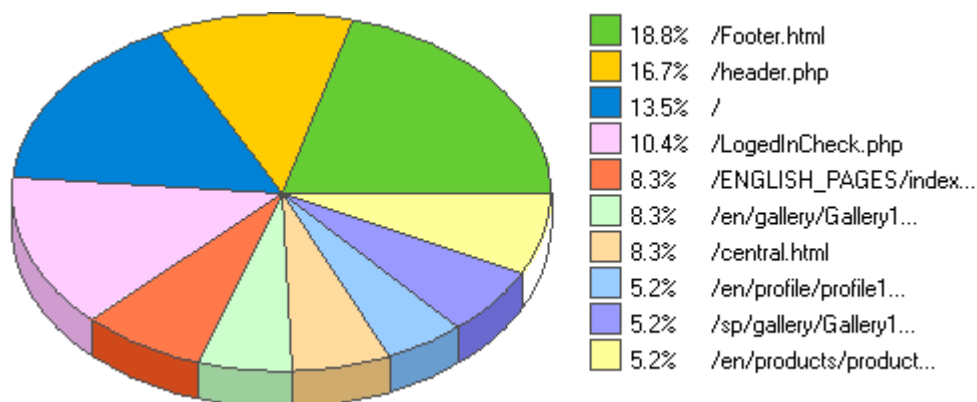
εικόνα 10^η εδώ απεικονίζεται η ημερήσια ζήτηση που έχει η ιστοσελίδα μας σε εβδομαδιαίο επίπεδο σε σχέση με τις 7 ημέρες της εβδομάδας ενώ από την κάθε μέρα μπορούμε να διακρίνουμε και το ποσοστό που αντιστοιχεί στην κάθε μέρα τις εκατό

Το αμέσως επόμενο λογισμικό πακέτο είναι το Absolute Log Analyzer Pro Demo v 2.33 το οποίο μπορεί να εξάγει τα συμπεράσματα του όπως το προηγούμενο πρόγραμμα δηλαδή σε δικιά του μορφή αλλά και σε html μορφή, μιας και πρόκειται για το ίδιο πρόγραμμα μόνο που πρόκειται για έκδοση που είναι επαγγελματική και αυτό συνεπάγεται κάποια επιπλέον πλεονεκτήματα για την όλη διαδικασία.



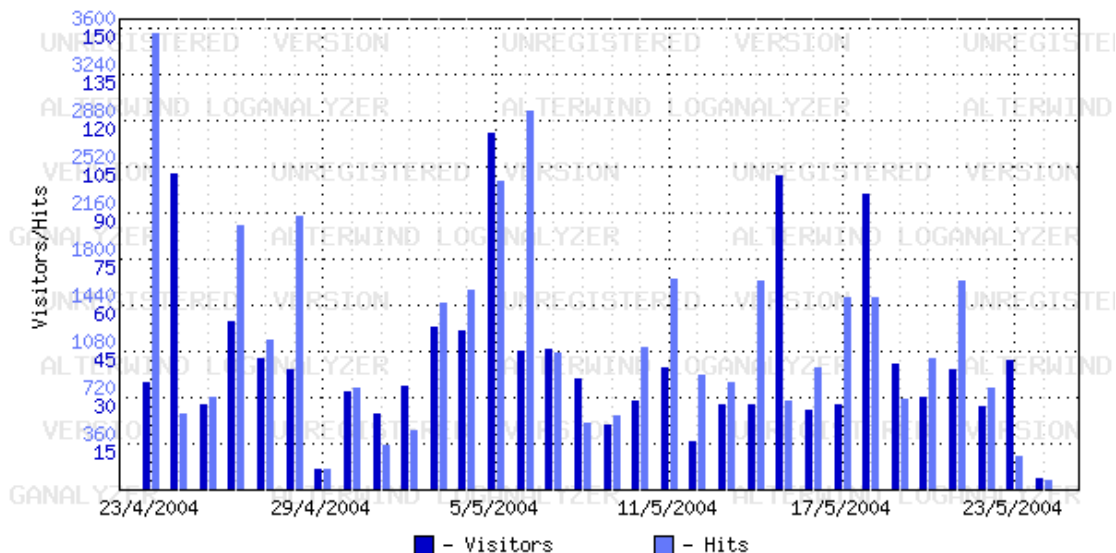
εικόνα 11^η εδώ απεικονίζεται η ζήτηση των αρχείων που έκαναν οι χρηστές download , το κάθε διαφορετικό χρώμα αντιστοιχεί και σε διαφορετικό αρχείο ενώ πάνω δεξιά στην εικόνα μπορούμε να δούμε και το ποσοστό του κάθε αρχείου

Το επόμενο λογισμικό πακέτο που τρέξαμε το log files είναι το Advanced Log Analyzer v 1.5 το οποίο μπορεί και επεξεργάζεται και αρχεία από iis server αλλά και από apache web server, επίσης μας δίνει την δυνατότητα να εξάγουμε τα συμπεράσματα ξεχωριστά για πιο server θέλουμε. Τέλος μας δίνεται η δυνατότητα να βλέπουμε τα συμπεράσματα μας και σε μορφή html όπου μπορούμε να δούμε τα αποτελέσματα μας σε πίνακες και σε γραφικές παραστάσεις, το συγκεκριμένο πρόγραμμα αναλύει προτού εκτελέσει την επεξεργασία του log file και μας δίνει πληροφορίες από ποιο server είναι το log file,



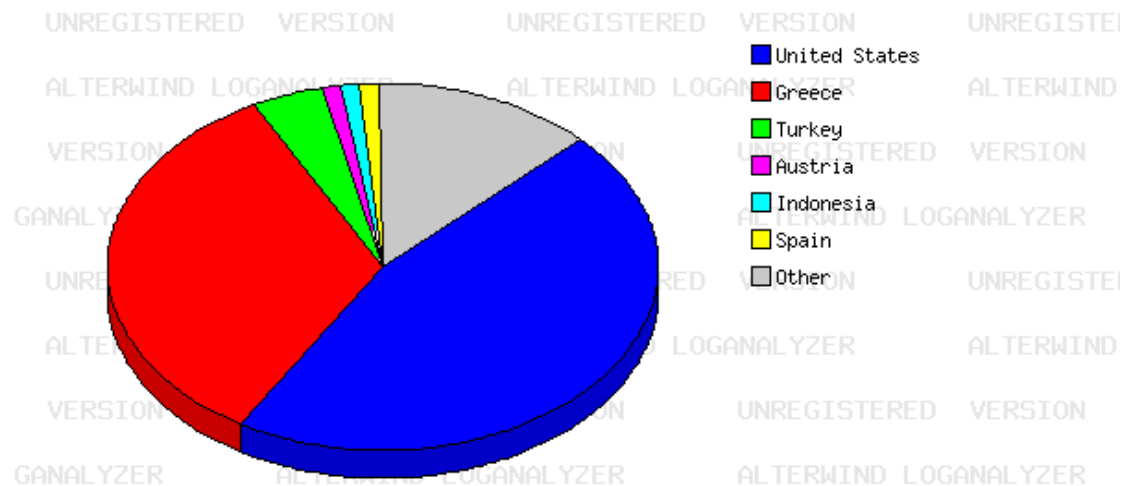
εικόνα 12^η εδώ απεικονίζεται η επισκεψιμότητα των κάθε link από τον ιστοχώρων μας ,το link ξεχωρίζει από τα άλλα αφού έχουν διαφορετικό χρώμα , εδώ μπορούμε να το δούμε σε μορφή ποσοστού και σε διάγραμμα,

Το ακολούθως επόμενο λογισμικό είναι το AlterWind LogAnalyzer Professional v 2.1 όπου πρόκειται για ένα απλό προγραμμα που επεξεργάζεται τα αρχεία log και τα εξάγει αυτόματα κατά την διάρκεια της επεξεργασίας σε html μορφή όπου μπορούμε να δούμε τα αποτελέσματα μας σε πίνακες αλλά και σε γραφικές παραστάσεις που μας βοηθάνε να εξάγουμε τα συμπεράσματα μας και σε στατιστικά στοιχεία.



εικόνα 13^η εδώ απεικονίζεται η επισκεψιμότητα ανά εβδομάδα με το σκούρο μπλε είναι οι επισκέπτες που είχαν επισκεπτεί την σελίδα μας ενώ με το ανοιχτό μπλε είναι οι επισκέψεις που δέχτηκε η σελίδα μας

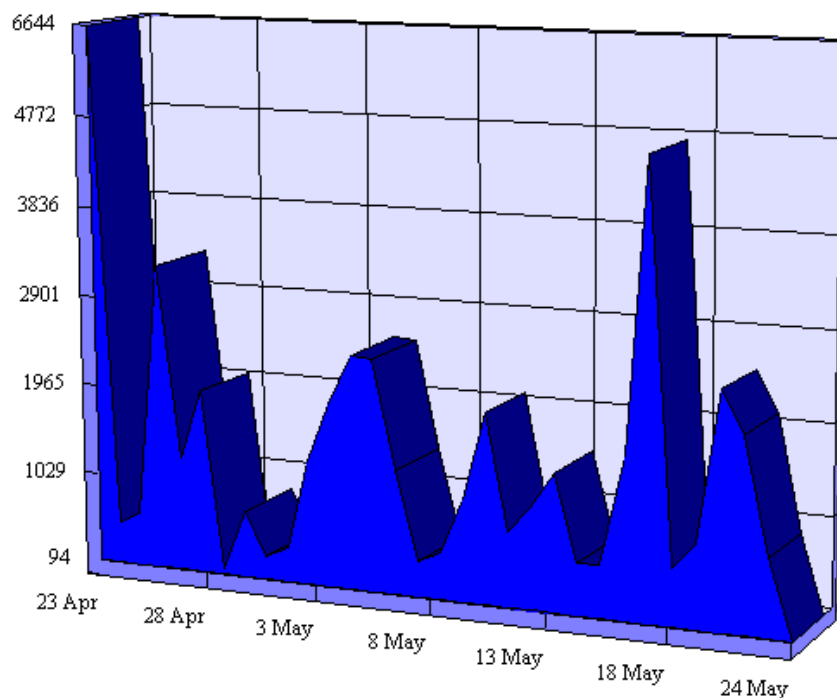
Το επόμενο λογισμικό πακέτο είναι Azure Web Log 1.4 στο οποίο τρέξαμε το αρχείο που είχαμε και τα συμπεράσματα που βγάλαμε ήτανε ότι εξάγει τα αποτελέσματα μας σε δικιά του μορφή και όχι σε html , excel η σε κάποια άλλη μορφή και αυτό έχει ως άμεση συνέπεια για να δούμε τα αποτελέσματα πρέπει αναγκαστικά να ανοίξουμε το πρόγραμμα και να τα διαβάσουμε μόνο μέσα από εκεί. Επίσης επειδή η έκδοση είναι demo δεν μας επιτρέπει να περιλάβουμε μέσα στο αρχείο μας περισσότερες από 10 ενεργές σελίδες όπως .php .asp και άλλες τέτοιες.



εικόνα 14^η εδώ απεικονίζεται η επισκεψιμότητα που είχε η σελίδα μας από κάθε χώρα για το συγκεκριμένο χρονικό διάστημα που είχαμε αναλύσει ,μπορούμε να ξεχωρίσουμε την κάθε χώρα από το διαφορετικό χρώμα που έχει στο διάγραμμα

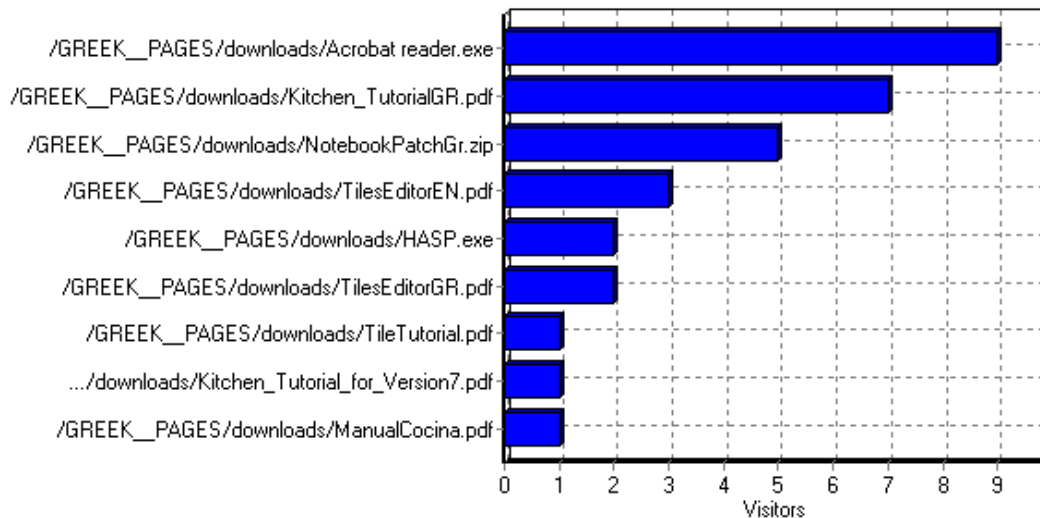
Το ακολούθως επόμενο λογισμικό πακέτο είναι το FastStats Analyzer v 2.81 πρόκειται για ένα αρκετά γρήγορο πρόγραμμα που είναι εύκολο στην εφαρμογή του μιας και εξάγει τα συμπεράσματα του σε διάφορες μορφές όπως .html excel και βεβαίως στη δικιά του μορφή όπου φαίνεται μέσα από το ίδιο το πρόγραμμα τους.

Hits per Day



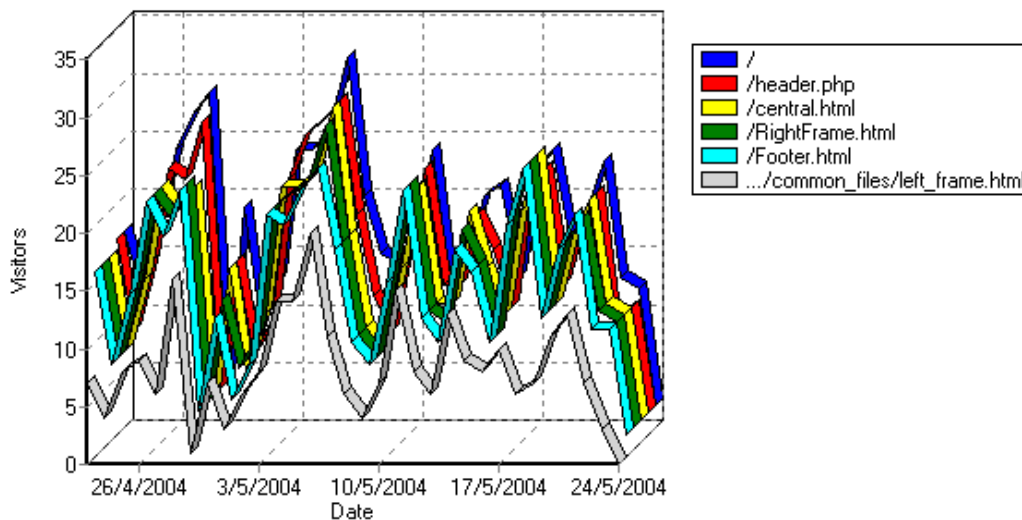
εικόνα 15^η εδώ απεικονίζεται η επισκεψιμότητα σε καθημερινή βάση αλλά μπορούμε και βγάλουμε συμπεράσματα και από εβδομαδιαία βάση αφού χωρίζονται σε εβδομάδες και στον κάθετο άξονα υπάρχει δομημένη η αρίθμηση για τις επισκέψεις ανά ημέρα

Το επόμενο λογισμικό πακέτο που αναλύσαμε είναι το OpenWebScope Web Statistics v 1.01 πρόκειται για demo έκδοση που μπορεί να εξάγει συμπεράσματα μόνο στην περίπτωση που είμαστε συνδεδεμένοι στο internet και με απομακρυσμένη σύνδεση με την σελίδα της οποίας θέλουμε να εξάγουμε τα συμπεράσματα μας, η εξαγωγή γίνεται σε html μορφή ,



εικόνα 16^η εδώ απεικονίζεται η ζήτηση των downloads , από αριστερά υπάρχει το αρχείο που έκαναν download οι χρήστες ενώ στον οριζόντιο άξονα υπάρχει το πόσες φορές το ζήτησαν οι χρήστες το συγκεκριμένο αρχείο

Τέλος το τελευταίο λογισμικό πακέτο όπου αναλύσαμε το αρχείο μας είναι το WebLog Expert Lite v 3.0 το οποίο μας επιτρέπει να εξαγάμε τα συμπεράσματα μας σε μορφή html και επειδή η έκδοση μας είναι demo δεν μας επιτρέπει να κάνουμε και παρά πολλά όπως να κάνουμε upload τα αρχεία μας σε κάποιο ftp server, τέλος το συγκεκριμένο πακέτο μας επιτρέπει να εξαγάμε και αρκετά συμπεράσματα μέσα από συγκεντρωτικούς πίνακες αλλά και μέσα από διάφορες γραφικές παραστάσεις.



εικόνα 17^η εδώ απεικονίζεται η επισκεψιμότητα ανά εβδομάδα από κάποια συγκεκριμένα αρχεία που έχουμε επιλέξει για ανάλυση και επεξεργασία , τα αρχεία ξεχωρίζουν με διαφορετικό χρώμα ενώ αριστερά φαίνεται ο αριθμός των επισκεπτών στο κάθε αρχείο,

5.4 Σύγκριση log analyzer

Στον παρακάτω πίνακα μπορούμε να δούμε συνολικά τα χαρακτηριστικά όλων των λογισμικών πακέτων και έτσι ταυτόχρονα να βγάλουμε τα συμπεράσματα μας για τα πλεονεκτήματα και μειονεκτήματα του κάθε προγράμματος.

Χαρακτηριστικό	123 loganalyzer	AbsoluteLog Analyzer Demo	AbsoluteLog Analyzer Pro Demo	AdvancedLog Analyzer	AlterWind LogAnalyzer Professional	Azure Web Log	FastStats Analyzer	Open Web Scope Web Statistics	WebLog Expert Lite
Κόστος	\$ 60	\$ 65	\$ 75	\$ 70	\$ 79	\$ 65	\$ 85	\$ 99	\$ 75
Windows	All	All	All	All	All	All	All	All	All
Μέγεθος(Mbytes)	2,3	14,50	14,70	2,8	3,4	0,75	2,74	1,01	2,30
Χρόνος ανάλυσης αρχείου	≈ 3 sec	≈ 2.5 sec	≈ 2.7 sec	≈ 3.1 sec	≈ 2.8 sec	≈ 2.6 sec	≈ 2.2 sec	≈ 3 sec	≈ 3.1 sec
Τεχνική υποστήριξη		√	√	√	√			√	
Αναφορά σε HTML	√	√	√	√	√		√	√	√
Αποστολή σε e-mail		√	√					√	
Γραφική αναπαράσταση αναφορών	√	√	√	√	√	√	√		√
Υποστήριξη των περισσότερων log files formats	√			√	√				√

Επιγραμματικά μπορούμε να αναφέρουμε ότι για το κάθε λογισμικό πακέτο ισχύουν κάποια συγκεκριμένα χαρακτηριστικά , συγκεκριμένα μπορούμε να πούμε ότι :

- το 123 loganalyzer βγάζει αναφορά σε html και σε μορφή γραφικών παραστάσεων ενώ ταυτόχρονα υποστηρίζει τα περισσότερα format των log files , ενώ δεν σου δίνει την δυνατότητα για αποστολή μέσω e-mail και δεν παρέχει τεχνική υποστήριξη,
- το AbsoluteLog Analyzer Demo παρέχει τεχνική υποστήριξη για το πρόγραμμα τους , γίνεται αποστολή μέσω mail και η αναφορά για το log αρχείο μπορεί να είναι σε μορφή html και σε μορφή γραφικών παραστάσεων ενώ δεν είναι δυνατή η υποστήριξη των περισσότερων format,
- το AbsoluteLog Analyzer Pro Demo σου δίνει τις εξής δυνατότητες: παρέχει τεχνική υποστήριξη για το πρόγραμμα τους , γίνεται αποστολή μέσω mail και η αναφορά για το log αρχείο μπορεί να είναι σε μορφή html και σε μορφή γραφικών παραστάσεων ενώ δεν είναι δυνατή η υποστήριξη των περισσότερων format,
- το AdvancedLog Analyzer υποστηρίζει τεχνική υποστήριξη ενώ η αναφορά του είναι μόνο σε μορφή html χωρίς να σου δίνει την δυνατότητα αποστολή μέσω e-mail αυτής αλλά σου δίνει την δυνατότητα αναπαράσταση μέσω γραφικών παραστάσεων ,τέλος υποστηρίζει τα περισσότερα format από αρχεία
- το AlterWind LogAnalyzer Professional υποστηρίζει τεχνική υποστήριξη ενώ η αναφορά του είναι μόνο σε μορφή html χωρίς να σου δίνει την δυνατότητα αποστολή μέσω e-mail αυτής αλλά σου δίνει την

δυνατότητα αναπαράσταση μέσω γραφικών παραστάσεων ,τέλος υποστηρίζει τα περισσότερα format από αρχεία log

- το Azure Web Log είναι το μοναδικό πρόγραμμα που βγάζει τις αναφορές σε μορφή γραφικών παραστάσεων ενώ ταυτόχρονα δεν υποστηρίζει ούτε τεχνική υποστήριξη , ούτε αναφορά σε html ,ούτε αποστολή μέσω e-mail αυτής,
- το FastStats Analyzer μας δίνει τις αναφορές σε html αλλά και σε μορφή γραφικών παραστάσεων ενώ δεν υποστηρίζει τεχνική υποστήριξη και τα περισσότερα format των αρχείων log ,τέλος δεν μπορεί να γίνει αποστολή μέσω e-mail,
- το Open Web Scope Web Statistics υποστηρίζει τεχνική υποστήριξη , μπορεί η αναφορά να είναι σε μορφή html και να γίνει αποστολή αυτής μέσω e-mail ,ενώ δεν μπορεί να βγάλει την αναφορά σε γραφικές παραστάσεις αλλά υποστηρίζει μόνο μερικά συγκεκριμένα format από log files,
- το WebLog Expert Lite σου δίνει την δυνατότητα η αναφορά να είναι σε μορφή γραφικών παραστάσεων αλλά και σε html , επίσης υποστηρίζει αρκετά format τέτοιων αρχείων ενώ δεν σου παρέχει τεχνική υποστήριξη και αποστολή της αναφοράς μέσω e-mail.

Αναλύοντας το κάθε χαρακτηριστικό ξεχωριστά και συγκρίνοντας τα προγράμματα καταλήξαμε στα εξής συμπεράσματα:

- Για το κόστος το πιο ακριβό λογισμικό πακέτο είναι το Open Web Scope Web Statistics ενώ το πιο φθηνό είναι το 123 loganalyzer ενώ τα άλλα πακέτα οι τιμές περιορίζονται ανάμεσα σε αυτές τις δυο τιμές .
- Όσον αφορά το θέμα του λειτουργικού συστήματος όλα τα λογισμικά πακέτα μπορούνε να εγκατασταθούνε και να δουλέψουνε κάλλιστα σε όλα τα windows δηλαδή σε win95,win98,winME,win2000,winNT,winXP .
- Τα προγράμματα AbsoluteLog Analyzer Demo και AbsoluteLog Analyzer Pro Demo καταλαμβάνουνε τον περισσότερο χώρο στον σκληρό δίσκο του υπολογιστή που θα χρησιμοποιούνται 14,50 MB και 14,70 MB ενώ το Azure Web Log καταλαμβάνει μόλις 0,75 MB στον σκληρό δίσκο, ενώ τα άλλα προγράμματα χρησιμοποιούνε περίπου 3 MB στο σκληρό δίσκο.
- Όσον αφορά τον χρόνο επεξεργασίας των αρχείων όλα τα προγράμματα θέλουνε τον ίδιο χρόνο επεξεργασίας και ανάλυσης ενώ το FastStats Analyzer είναι το μοναδικό πρόγραμμα που θέλει το λιγότερο χρόνο επεξεργασίας
- Τεχνική υποστήριξη υποστηρίζουνε τα περισσότερα προγράμματα εκτός από τα εξής , 123 loganalyzer Azure Web Log , FastStats Analyzer και WebLog Expert Lite.
- Το Azure Web Log είναι το μοναδικό από τα προγράμματα που συγκρίναμε το οποίο δεν υποστηρίζει εξαγωγή των αποτελεσμάτων σε html .

- Τα προγράμματα AbsoluteLog Analyzer Demo , AbsoluteLog Analyzer Pro Demo και Open Web Scope Web Statistics είναι τα μοναδικά τα οποία υποστηρίζουνε την δυνατότητα αποστολής μέσω e-mail των αποτελεσμάτων από την επεξεργασία των αρχείων ενώ όλα τα άλλα δεν έχουνε αυτή την δυνατότητα .
- Όλα τα προγράμματα υποστηρίζουνε την δυνατότητα να απεικονίζεται η αναφορά σε μορφή γραφικών παραστάσεων εκτός του Open Web Scope Web Statistics .
- Τα προγράμματα 123 loganalyzer, AdvancedLog Analyzer , AlterWind LogAnalyzer Professional και WebLog Expert Lite υποστηρίζουνε τα περισσότερα format των log files ενώ τα υπόλοιπα προγράμματα δεν υποστηρίζουνε αυτή την δυνατότητα.

Κεφάλαιο 6 Συμπεράσματα

Σε αυτό το κεφαλαίο θα αναφερθούμε στα συμπεράσματα που έχουμε από την όλη διπλωματική μας εργασία ,

Πρώτα από όλα πρέπει να αναφέρουμε ότι πριν από όλα πρέπει να διαλέξουμε τον τρόπο που θέλουμε να κάνουμε την εξόρυξη δεδομένων μας, δηλαδή αν η εξόρυξη δεδομένων θα γίνει μέσα από μια συγκεκριμένη βάση δεδομένων ή θα γίνει μέσα από το web.

Στη συνέχεια αφού διαλέξουμε να κάνουμε την εξόρυξη δεδομένων μας μέσα από το web αυτό θα γίνει μέσα από κάποιο αρχείο (log files) βεβαίως πρέπει να κάνουμε μια βασική προεπεξεργασία πριν γίνει η τελική ανάλυση δηλαδή να γίνει ο καθαρισμός των δεδομένων μας , και να διαλέξουμε εμείς τα δεδομένα που θέλουμε να επεξεργαστούμε , με βάση τους στόχους που βάζουμε

Βέβαια τα κριτήρια για την επιλογή του προγράμματος κάθε φορά μπορεί να αλλάζουν αναλόγως το τι θέλουμε να κάνουμε , σε αυτή την διπλωματική όμως έχει γίνει μια αξιολόγηση αρκετών προγραμμάτων για τέτοιες περιπτώσεις, βεβαίως πρέπει να πούμε ότι αυτή τη δουλειά τις περισσότερες φορές την κάνουν οι διαχειριστές ιστοχώρων που θέλουν να ελέγξουν την λειτουργικότητα της ιστοσελίδας τους και να βγάλουν τα δικά τους συμπεράσματα τους .

Σε αυτό σημείο μπορεί να βοηθήσει η συγκεκριμένη διπλωματική αφού παραπάνω έχει γίνει μια αξιολόγηση και σύγκριση κάποιων λογισμικών πακέτων μέσα από ένα μεγάλο αριθμό που υπάρχει στην αγορά.

Εδώ πρέπει να αναφέρουμε ότι όπως γίνεται στα περισσότερα software έτσι γίνεται και στην δικιά μας περίπτωση τα λογισμικά πακέτα που η έκδοσή τους είναι professional έχουν περισσότερες δυνατότητες από ένα πρόγραμμα που η έκδοσή του δεν είναι professional , σαφώς και η professional έκδοση είναι πιο ακριβή από μια μη professional έκδοση,

Βέβαια τα περισσότερα πακέτα που η έκδοσή τους είναι trial οι λειτουργίες είναι περιορισμένες σε σχέση με αυτά που είναι ολοκληρωμένη η έκδοσή τους

Αλλά πάρο όλα αυτά κάνουνε την δουλειά τους εξίσου καλά σε σχέση με τα αγορασμένα προγράμματα. Αλλά το μοναδικό τους μειονέκτημα είναι ότι η συγκεκριμένη έκδοση αν είναι trial τότε θα λήγει σε μικρό χρονικό διάστημα. Πολλά trial δεν δίνουν όλες τις δυνατότητες των προγραμμάτων αλλά μερικές

Επίσης πρέπει να αναφέρουμε ότι η συμβατότητα του κάθε προγράμματος είναι κοινή για όλους τους υπολογιστές, δηλαδή τα προγράμματα που δουλέψαμε, δουλεύουνε σε όλους τους υπολογιστές αφού οι απαιτήσεις είναι πολύ μικρές σε σχέση με τους υπολογιστές που υπάρχουνε σήμερα και επίσης δουλεύουνε σε όλες τις εκδόσεις των windows μιας και είναι φτιαγμένα για τέτοιου είδους λειτουργικά συστήματα.

Τέλος πρέπει να γίνει μια επισήμανση μετά από την σύγκριση των όλων λογισμικών πακέτων που κάναμε ότι μερικά προγράμματα έχουνε πολλές κοινές δυνατότητες μεταξύ τους, αλλά υπάρχουνε και προγράμματα που οι δυνατότητες τους είναι ξεχωριστές σε σχέση με τις άλλες π.χ τα περισσότερα προγράμματα η αναφορά για ένα αρχείο μπορούσαμε να την δούμε σε μορφή html αλλά υπήρχε και πρόγραμμα που έβγαζε την αναφορά και σε μορφή για excel, ή ένα άλλο παράδειγμα αρκετά προγράμματα δεχότανε ως είσοδο το αρχείο και συμπιεσμένο όταν ήτανε αρκετά MB ενώ αλλά προγράμματα δεχότανε το αρχείο μόνο όπως ήτανε από τον διακομιστή

Αναφορές-Βιβλιογραφία

[1] IRM Data Mining. Copyright ©1999, all rights reserved(data%20mining.pdf)

[2] Data Mining: An Overview from Database Perspective

Ming-Syan Chen Elect. Eng. Department National Taiwan Univ. Taipei, Taiwan, ROC

Jiawei Han School of Computing Sci. Simon Fraser University

B.C. V5A 1S6, Canada

Philip S. Yu IBM T.J. Watson Res. Ctr.P.O.Box 704 Yorktown, NY 10598,
U.S.A. (han.pdf)

[3] DATA MINING: A CONCEPTUAL OVERVIEW

Joyce Jackson Management Science Department
University of South Carolina joyce.jackson@sc.edu (paper1a4.pdf)

[4] Web Usage Mining on Proxy Servers:

A Case Study Jan Kerkhofs Prof. Dr. Koen Vanhoof Danny Pannemans
Limburg University Centre July 30, 2001 (workshop_notes.pdf)

[5] LOG FILES ANALYSIS TO MEASURE THE UTILITY OF AN INTRANET

Stéphane Vokar, Céline Mariage, Jean Vanderdonck
Université catholique de Louvain – Institut d'Administration et de Gestion
Place des Doyens, 1 - B-1348 Louvain-la-Neuve, Belgium
{vokar, mariage,vanderdonck}@isys.ucl.ac.be –
<http://www.isys.ucl.ac.be/bchi>

[6] [http:// www.loganalyzer.com](http://www.loganalyzer.com)

[7] [http:// www.birstrike.com/analyzer/](http://www.birstrike.com/analyzer/)

[8] [http:// www.abacre.com](http://www.abacre.com)

[9] [http:// www.alterwind.com](http://www.alterwind.com)

[10] [http:// www.azuredesktop.com](http://www.azuredesktop.com)

[11] [http:// www.mach5.com](http://www.mach5.com)

[12] [http:// www.openwebscope.com](http://www.openwebscope.com)

[13] [http:// www.weblogexpert.com](http://www.weblogexpert.com)

Παράρτημα-log file

Στην συνέχεια στο παράρτημα που ακολουθεί υπάρχει ένα τμήμα από το log file που έχουμε επεξεργαστεί επειδή η έκταση του συγκεκριμένου αρχείου είναι πολύ μεγάλη , κρίθηκε αναγκαίο να μην εμφανιστεί όλο το αρχείο μιας και οι σελίδες του είναι περίπου 4000 .

Το αρχείο (access log file) καθώς και η διπλωματική εργασία υπάρχουνε στο διαδίκτυο στην παρακάτω ιστοσελίδα :

http://homepages.pathfinder.gr/giannis_drakos

όποιος επιθυμεί μπορεί να κατεβάσει το συγκεκριμένο αρχείο σε ηλεκτρονική μορφή αλλά και την εργασία σε μορφή .pdf .

```
62.103.78.228 - - [28/Apr/2004:11:48:19 +0300] "GET /index.html
HTTP/1.1" 304 - "-" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT
5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:19 +0300] "GET /header.php
HTTP/1.1" 200 1555 "http://www.infowood.gr/index.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:19 +0300] "GET /RightFrame.html
HTTP/1.1" 304 - "http://www.infowood.gr/index.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:19 +0300] "GET /central.html
HTTP/1.1" 304 - "http://www.infowood.gr/index.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /Footer.html
HTTP/1.1" 304 - "http://www.infowood.gr/index.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /imgs/logo.jpg
HTTP/1.1" 304 - "http://www.infowood.gr/header.php" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /SIMA1B.jpg
HTTP/1.1" 304 - "http://www.infowood.gr/header.php" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /pcmenu.js
HTTP/1.1" 304 - "http://www.infowood.gr/RightFrame.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /imgs/icon1K.jpg
HTTP/1.1" 304 - "http://www.infowood.gr/central.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /imgs/icon2K.jpg
HTTP/1.1" 304 - "http://www.infowood.gr/central.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /imgs/icon1B.jpg
HTTP/1.1" 304 - "http://www.infowood.gr/central.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET
/imgs/icon2Bath.jpg HTTP/1.1" 304 -
```

```
"http://www.infowood.gr/central.html" "Mozilla/4.0 (compatible; MSIE
6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /imgs/icon1P.jpg
HTTP/1.1" 304 - "http://www.infowood.gr/central.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET
/imgs/icon2Office.jpg HTTP/1.1" 304 -
"http://www.infowood.gr/central.html" "Mozilla/4.0 (compatible; MSIE
6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /imgs/icon4.gif
HTTP/1.1" 304 - "http://www.infowood.gr/central.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /imgs/newsBig.jpg
HTTP/1.1" 304 - "http://www.infowood.gr/central.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /closed.gif
HTTP/1.1" 304 - "http://www.infowood.gr/RightFrame.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
62.103.78.228 - - [28/Apr/2004:11:48:20 +0300] "GET /bar.jpg
HTTP/1.1" 304 - "http://www.infowood.gr/RightFrame.html" "Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)"
```