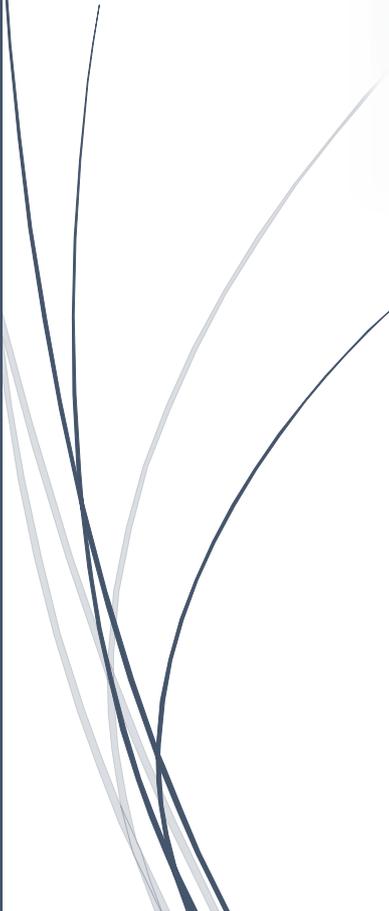


Διαδικτυακά εργαλεία και τεχνικές διαχείρισης Πληροφοριών



Η σελίδα είναι σκόπιμα λευκή



ΑΝΩΤΑΤΟ ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΗΠΕΙΡΟΥ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ Τ.Ε

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Διαδικτυακά εργαλεία και τεχνικές διαχείρισης Πληροφοριών

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:
ΣΤΥΛΙΟΣ ΧΡΥΣΟΣΤΟΜΟΣ

ΑΜ:10656

ΑΡΤΑ, 2013

Περιεχόμενα

Περιεχόμενα	3
Περίληψη	5
Abstract	6
1.0 Διαδικτυακά εργαλεία και τεχνικές διαχείρισης Πληροφοριών	7
1.2 Εργαλεία και τεχνικές για τη διαχείριση των πληροφοριών.....	8
1.3 Πληροφορίες Τεχνολογίας.....	8
1.4 Αυτοματοποιημένη διαδικασία εξόρυξης πληροφορίας από τον Παγκόσμιο Ιστό	9
2.0 Web crawlers	11
2.1 Αρχιτεκτονική των Web Crawlers.....	14
2.2 Crawler αναγνώρισης.....	15
2.3 Το crawling στο «βαθύ» ιστό	16
2.3 Ιχνηλάτες Ανοιχτού Κώδικα.....	16
2.4 Δυναμικές Ιστοσελίδες και Αλληλεπίδραση Χρηστών	17
3.0 Προγραμματιστικό Περιβάλλον- Πλατφόρμα Ανάπτυξης	19
3.1 Το Visual Studio .NET ως πλατφόρμα ανάπτυξης	21
3.2 Αρχιτεκτονική του Visual Studio .NET	22
3.3 Internet Information Server ως Web Server	26
α)Τι είναι ένας Web Server	26
3.3.1 Ο υπολογιστής Web Server ως υλικό.....	26
3.3.2 Ποιος είναι ο ρόλος του web server στο Internet	27
3.4 Apache HTTP εξυπηρετητής.....	29
3.5 IIS(Internet Information Server)	30
3.5.1Πρωτόκολλο TCP/IP	32
4.0 Βασικές Υπηρεσίες IIS.....	34
4.1 Η υπηρεσία WWW	35
4.2 Active Server Pages (ASP).....	37
4.3 Τρόπος Λειτουργίας ASP.NET αρχείων	38
4.4 Η γλώσσα προγραμματισμού C#	40
4.5 Γενικά Χαρακτηριστικά C#	41
5.0 Microsoft SQL Server	45
5.1 Χαρακτηριστικά	46
5.2 Αποθήκευση Δεδομένων	47

5.3 Ενδιάμεση μνήμη (Cache)	49
5.4 Από το Web 1.0 στο Web 2.0	54
5.4.1 Η μετακίνηση προς το Web 2.0	55
5.4.2 Από το Web 2.0 στο Web 3.0	57
5.5 Asynchronous JavaScript and XML (AJAX)	61
6.0 Τεχνολογίες	63
6.1 Σχεδιασμός Βάσης Δεδομένων	67
6.2 Μηχανές Αναζήτησης στο Παγκόσμιο Ιστό (λειτουργία)	69
6.3 Crawler-based μηχανές αναζήτησης.....	70
6.4 Θεματικοί κατάλογοι	71
6.5 Μετα-μηχανές αναζήτησης.....	72
7.0 Μετρικές Απόδοσης Αλγορίθμων Μηχανών Αναζήτησης.....	73
7.1.1 Ορισμός (πληροφορίες πλαίσιο ανάκτησης)	74
7.2 Μέθοδος υπολογισμού της σπουδαιότητας των όρων ενός κειμένου: η μετρική TF-IDF	75
7.3 Αρχιτεκτονική τριών επιπέδων (3-Tier Architecture)	78
7.4 Components	82
7.5 Αρχιτεκτονική Πληροφοριών (Information Architecture).....	84
7.6 Σχεδιασμός Σελίδων Ιστού (Web Design)	85
8.0 Σχεδιασμός Διεπαφής Χρήστη (User Interface Design)	88
8.1 Γενικότερα Θέματα Σχεδιασμού Ιστοχώρων	90
8.2 Κανόνες Design του Ιστοχώρου	92
8.3 Συμπεράσματα	94
9.0 Βιβλιογραφία.....	95

Περίληψη

Η πληροφορία στην κοινή γλώσσα είναι γεγονότα και απόψεις που προσφέρονται και λαμβάνονται από έμβια όντα, μέσα μαζικής επικοινωνίας, ηλεκτρονικούς υπολογιστές (κυρίως μέσω διαδικτύου) και από πάσης φύσεως φαινόμενα στο περιβάλλον. Στην επιστήμη της Πληροφορικής, όταν η πληροφορία εξετάζεται στατιστικά, απεκδύεται κάθε έννοιας νοήματος και αποσυνδέεται από κάθε «σημασία» με την έννοια που την αντιλαμβάνονται οι άνθρωποι. Τέλος όσον αφορά το διαδίκτυο συγκεντρώνει μεγάλες ποσότητες πληροφοριών, οι οποίες λόγω του όγκου τους πολλές φορές μένουν αναξιοποίητες. Οι τεχνικές εξόρυξης γνώσης μπορούν να αυτοματοποιήσουν τη διαδικασία της ανάκτησης χρήσιμων πληροφοριών από τον ιστό και να συνδυαστούν με συστήματα που μπορούν να αξιοποιήσουν τη γνώση αυτή.

Λέξεις κλειδιά: πληροφορία, τεχνικές εξόρυξης, ανάκτησης, Πληροφορική, διαδίκτυο

Abstract

The information in common language are facts and opinions that are offered and derived from living organisms , media, e-mail computers (mostly online) and all kinds of phenomena - me environment . In Computer Science, when the information considered statistically disclaims any notion of meaning and disconnected from any "meaning" in the sense that we perceive people. Finally, as regards the internet gathers large amounts of information, which due to their size often remain untapped. The mining techniques can automate the process of retrieving useful information from the web and combine them with systems that can exploit this knowledge.

Keywords: information extraction techniques, recovery, IT, internet

1.0 Διαδικτυακά εργαλεία και τεχνικές διαχείρισης Πληροφοριών

1.1 Τι είναι πληροφορία

Η πληροφορία είναι μια αλληλουχία συμβόλων, που είτε καταγράφονται είτε μεταδίδονται, η οποία μπορεί να ερμηνευτεί ως μήνυμα και μπορεί να επηρεάσει ένα δυναμικό σύστημα το οποίο είναι σε θέση να την επεξεργαστεί. Αντίστροφα, ερμηνεύεται ως στοιχείο γνώσης η νοητική οργάνωση μιας συλλογής από απλά πράγματα που αποκτά σημασία.

Η πληροφορία στην κοινή γλώσσα είναι γεγονότα και απόψεις που προσφέρονται και λαμβάνονται από έμβια όντα, μέσα μαζικής επικοινωνίας, ηλεκτρονικούς υπολογιστές (κυρίως μέσω διαδικτύου) και από πάσης φύσεως φαινόμενα στο περιβάλλον. Στην επιστήμη της Πληροφορικής, όταν η πληροφορία εξετάζεται στατιστικά, απεκδύεται κάθε έννοιας νοήματος και αποσυνδέεται από κάθε «σημασία» με την έννοια που την αντιλαμβάνονται οι άνθρωποι.

Άλλος ένας ορισμός της πληροφορίας είναι ο εξής είναι μερικά νόημα μήνυμα που καταγράφονται σε συμβατικά ή μη συμβατικά μέσα ,αποθηκεύονται και να υποβάλλονται σε επεξεργασία από τα συστήματα και τις υπηρεσίες με σκοπό την παροχή μιας περισσότερο ή λιγότερο μόνιμη μνήμη του μηνύματος και τη διάδοσή τους στους χρήστες.

Το British Standards¹ ορίζει την πληροφορία ως «την έννοια που ο άνθρωπος χρησιμοποιεί σημάδια λόγου χάρη ομιλία, χειρονομίες για να επικοινωνήσει ». Μπορεί να θεωρηθεί ως επεξεργασία δεδομένων, έτσι, οι πληροφορίες είναι δεδομένα με την έννοια, ότι είναι ένα αποτέλεσμα της ανθρώπινης δράσης.

Ένας αντισυμβατικός ορισμός ορίζει την πληροφορία ως ένα τρόπο εξωτερικό για να τροφοδοτήσει τις γνώσεις μας, έτσι ώστε να μπορεί να λειτουργήσει με επιτυχία και να επιτύχουμε τους στόχους μας.

¹ BS 3527, Part 1, 1976

1.2 Εργαλεία και τεχνικές για τη διαχείριση των πληροφοριών

Η ανάγκη για τη διαχείριση των πληροφοριών μέσα σε ένα θεσμικό όργανο δεν είναι κάτι καινούργιο, δύο μεγάλες τάσεις, ωστόσο, έχουν έρθει μαζί για να δημιουργήσουν ένα νέα σημασία των πληροφοριών. Πρώτον, η αυξανόμενη πολυπλοκότητα των οργανωτικών δομών και λειτουργιών, και, δεύτερον, ο αντίκτυπος της πληροφορικής και των επικοινωνιών για τις λειτουργίες εργασίας και τα πρότυπα. Κατά συνέπεια, είναι η επεξεργασία των πληροφοριών και των πόρων διαχείρισης των πληροφοριών εντός του οργανωτικού πλαισίου (Lewis, 1985).

1.3 Πληροφορίες Τεχνολογίας

Το πεδίο της επιστήμης βιβλιοθηκονομίας και των πληροφοριών έχει υποστεί μια ριζική αλλαγή κατά τη διάρκεια των τελευταίων τριών δεκαετιών, λόγω των επιπτώσεων της τεχνολογίας των πληροφοριών σχετικά με την παραγωγή, επεξεργασία, αποθήκευση, ανάκτηση και διάδοση / μετάδοση των πληροφοριών. Στην παρούσα συγκυρία, μπορεί να είναι χρήσιμο να θυμόμαστε ότι οι τέσσερις αλληλένδετες τάσεις της πληροφορικής: πολλαπλές πληροφορίες τεχνολογίας, διασπορά της πληροφορίας τεχνολογίας, Επιτάχυνση της πληροφορίας τεχνολογίας, και τις Διάχυτες πληροφορίες τεχνολογίας

Τα πλεονεκτήματα της πληροφορίας τεχνολογίας περιλαμβάνουν: διαμονή του αυξημένου φόρτου εργασίας, η επίτευξη μεγαλύτερης αποτελεσματικότητας, η ικανότητα για την παραγωγή των νέων υπηρεσιών, τη διευκόλυνση της συνεργασίας κλπ. μπορεί να σταθεί σε καλή θέση στην αναζήτηση της ποιότητας και της παραγωγικότητας και των υπηρεσιών. Αυτή η νέα τεχνολογία βασίζεται κυρίως σε ηλεκτρονικά μέσα φαίνεται να αντικαθιστούν σταδιακά τα συμβατικά εργαλεία σε δραστηριότητες ενημέρωσης, όπως είναι δεκτικά για την ένταξη σε αντίθεση με το τελευταίο.

Παρ' όλα αυτά, η ενσωμάτωση δεν έχει επιτευχθεί σε πολλές περιπτώσεις λόγω ασυμβατότητας στο υλικό του συστήματος και το σχεδιασμό του λογισμικού και τα σχετικά πρότυπα.

1.4 Αυτοματοποιημένη διαδικασία εξόρυξης πληροφορίας από τον Παγκόσμιο Ιστό

Μία από τις μεγαλύτερες, αν όχι η μεγαλύτερη, πηγές πληροφοριών την σημερινή εποχή αποτελεί ο Παγκόσμιος Ιστός ή με άλλα λόγια το διαδίκτυο. Στο διαδίκτυο πέραν των άλλων παρέχονται πληροφορίες για τεχνολογικές εξελίξεις, ερευνητικά αποτελέσματα ή καινοτομίες που εξελίσσονται.

Παγκόσμιος ιστός είναι το δίκτυο των συνδεδεμένων υπολογιστών και δικτύων σε παγκόσμια κλίμακα, το οποίο χρησιμοποιεί συγκεκριμένη ομάδα πρωτοκόλλων επικοινωνίας, γνωστή ως «http». Κάθε δίκτυο-δομική μονάδα του διαδικτύου αποτελείται από συνδεδεμένους υπολογιστές σε τοπικό επίπεδο, για παράδειγμα το δίκτυο υπολογιστών των κεντρικών γραφείων μιας εταιρίας. Αυτά τα δίκτυα με τη σειρά τους συνδέονται σε ευρύτερα δίκτυα, όπως εθνικά και υπερεθνικά. Το ευρύτερο δίκτυο στον κόσμο λέγεται παγκόσμιος ιστός το οποίο είναι μοναδικό (δηλαδή δεν υπάρχουν παραπάνω από ένα δίκτυα υπολογιστών παγκόσμιας κλίμακας), και συμπεριλαμβάνεται τόσο τα γήινα δίκτυα, όσο και τα δίκτυα των δορυφόρων της και άλλων διαστημικών συσκευών που είναι συνδεδεμένα σε αυτό.

Η τεχνολογία του ιστού καθιστά δυνατή την δημιουργία «υπερκειμένων», μία διασύνδεση δηλαδή πάρα πολλών μη ιεραρχημένων στοιχείων που παλαιότερα ήταν απομονωμένα. Τα στοιχεία αυτά μπορούν να πάρουν και άλλες μορφές πέραν της μορφής του γραπτού κειμένου, όπως εικόνας και ήχου.

Η τεχνολογία του ιστού δημιουργήθηκε το 1989 από τον Βρετανό Τιμ Μπέρνερς Λη, που εκείνη την εποχή εργαζόταν στον Ευρωπαϊκό Οργανισμό Πυρηνικών Ερευνών (CERN) στην Γενεύη της Ελβετίας. Το όνομα που έδωσε στην εφεύρεσή του ο ίδιος ο Lee είναι World Wide Web, όρος γνωστός στους περισσότερους από το «www». Αυτό που οδήγησε τον Lee στην εφεύρεση του Παγκόσμιου ιστού ήταν το όραμά του για ένα κόσμο όπου ο καθένας θα μπορούσε να ανταλλάσσει πληροφορίες και ιδέες άμεσα προσβάσιμες από τους υπολοίπους. Το σημείο στο οποίο έδωσε ιδιαίτερο βάρος ήταν η μη ιεράρχηση των διασυνδεδεμένων στοιχείων. Οραματίστηκε κάθε στοιχείο, κάθε κόμβο του ιστού ίσο

ως προς την προσβασιμότητα με τα υπόλοιπα. Αν σκεφτεί, όμως, κανείς τον βαθμό ιεράρχησης με τον οποίο λειτουργούν οι μηχανές αναζήτησης του ιστού, όπως για παράδειγμα το google, γίνεται εύκολα κατανοητό ότι στην πράξη κάτι τέτοιο δεν συμβαίνει, τουλάχιστον στον βαθμό που το είχε οραματιστεί ο Lee.

Επίσης είναι γνωστό ότι οι πληροφορίες από τα αποτελέσματα ερευνητικών και αναπτυξιακών αποτελεσμάτων, τις καινοτομίες και τις τεχνολογίες που αναπτύχθηκαν σε εθνικό επίπεδο είναι διαθέσιμα στο διαδίκτυο. Οι πληροφορίες αυτές παρέχονται από τους αντίστοιχους δικτυακούς τόπους που αναπτύσσονταν στα πλαίσια υλοποίησης ερευνητικών ή αναπτυξιακών προγραμμάτων ή σε δικτυακούς τόπους που περιέχουν και συντηρούν αντίστοιχες συλλογές δεδομένων.

Βάση των ανωτέρω θα έπρεπε να αναζητηθούν μέθοδοι αυτόματης εξόρυξης της ήδη υπάρχουσας πληροφορίας από το διαδίκτυο δίνοντας ιδιαίτερη έμφαση τόσο στην ποιότητα αυτής όσο και της σημασιολογικής της συνέπειας με του στόχους του έργου.

Η μέθοδος που είναι η πιο ευρέως διαδεδομένη και χρησιμοποιείται σε αντίστοιχες περιπτώσεις είναι οι web crawlers.

2.0 Web crawlers

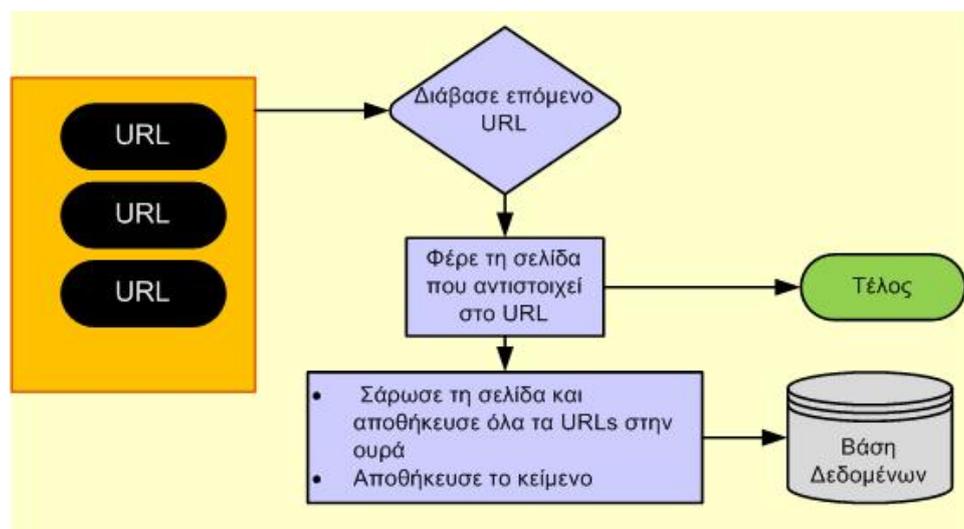
Ένας ιχνηλάτης του ιστοχώρου (Web crawler) είναι ένα πρόγραμμα το οποίο «φυλλομετρά» το web με ένα αυτόματο και συστηματικό τρόπο. Η διαδικασία αυτή αναφέρεται και ως crawling. Άλλοι συνώνυμοι όροι που χρησιμοποιούνται στη βιβλιογραφία είναι bots, Web spiders, Web robots. Ο ιχνηλάτης αποτελεί βασικό στοιχείο των μηχανών αναζήτησης, και πολλών άλλων web εφαρμογών, καθόσον από αυτόν εξαρτάται η συγκέντρωση, επικαιροποίηση αλλά και η ποιότητα των δεδομένων. Επιπλέον οι ιχνηλάτες μπορεί να χρησιμοποιηθούν για την αυτόματη συντήρηση ενός ιστότοπου (Web site), όπως για παράδειγμα για τον έλεγχο των συνδέσμων ή για την επιβεβαίωση HTML κώδικα, καθώς επίσης μπορούν να χρησιμοποιηθούν για την συλλογή συγκεκριμένου τύπου πληροφορίας από το web, όπως για παράδειγμα την συλλογή e-mail διευθύνσεων (spamming, spambot).

Ένας web crawler είναι ένα είδος πράκτορα λογισμικού (bot). Γενικά ο crawler ξεκινά από μια λίστα με URL (Uniform Resource Locator) για να επισκεφτεί τα οποία καλούνται σπόροι (seeds). Καθώς ο ιχνηλάτης επισκέπτεται αυτά τα URL, εντοπίζει όλους τους υπερσυνδέσμους (hyperlinks) σε μια σελίδα και τους προσθέτει στη λίστα των URL για να τα επισκεφτεί στη συνέχεια με μια συγκεκριμένη πολιτική. Επομένως οι ιχνηλάτες είναι ένα πρόγραμμα το οποίο διασχίζει το γράφο του web ακολουθώντας τους συνδέσμους και αυτόματα κατεβάζει τοπικά το περιεχόμενο των σελίδων του web. Η βασική δομή ενός ιχνηλάτη δίνεται στην εικόνα. Όπως βλέπουμε ο ιχνηλάτης εφοδιάζεται από μια λίστα (ουρά) από URLs. Επισκέπτεται το πρώτο URL κατεβάζει το περιεχόμενο της σελίδας και τυχόν νέους συνδέσμους που υπάρχουν στη σελίδα οι οποίοι και προστίθενται στην αρχική ουρά των URL. Η διαδικασία αυτή επαναλαμβάνεται έως ότου εξαντληθεί η λίστα των URLs .

Ένα πρόγραμμα ανίχνευσης Web Crawler αρχίζει με μια λίστα διευθύνσεων URL για να επισκεφθεί , και να καλέσει πόρους . Καθώς ο ανιχνευτής επισκέπτεται αυτές τις διευθύνσεις URL , προσδιορίζει όλες τις υπερ-συνδέσεις στη σελίδα και τους προσθέτει στον κατάλογο των URLs. Ο μεγάλος όγκος δεδομένων σημαίνει ότι ο ανιχνευτής μπορεί να κατεβάσει μόνο έναν περιορισμένο αριθμό

των ιστοσελίδων σε μια δεδομένη στιγμή , για αυτό πρέπει να δοθεί προτεραιότητα στις λήψεις του . Το υψηλό ποσοστό της αλλαγής συνεπάγεται ότι οι σελίδες να έχουν ήδη ενημερωθεί ή και να διαγραφεί .

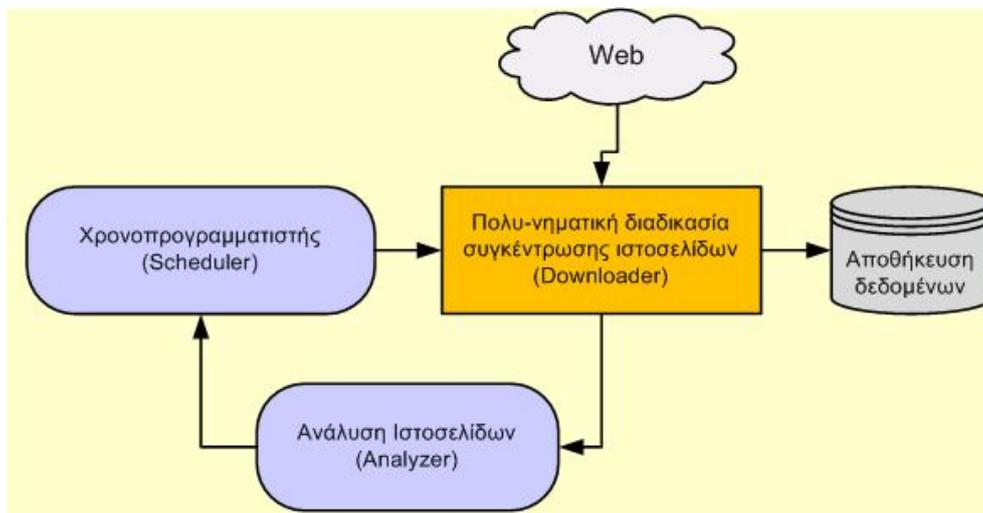
Για να κατανοήσουμε καλύτερα το νόημα του Web Crawler θα χρησιμοποιήσουμε ένα παράδειγμα έχουμε ένα μια online συλλογή φωτογραφιών που μπορεί να προσφέρει τρεις επιλογές για τους χρήστες , όπως ορίζεται μέσω των παραμέτρων HTTP GET στη διεύθυνση URL . Αν υπάρχουν τέσσερις τρόποι για να ταξινομήσουμε τις εικόνες , τρεις επιλογές μεγέθους thumbnail , δύο μορφές αρχείων , καθώς και μια επιλογή για να απενεργοποιήσουμε το χρήστη περιεχόμενο που παρέχεται , τότε το ίδιο σύνολο του περιεχομένου μπορεί να προσεγγιστεί με 48 διαφορετικές διευθύνσεις URL , τα οποία μπορούν να συνδεθούν με το site . Αυτός ο μαθηματικός συνδυασμός δημιουργεί ένα πρόβλημα για τους crawlers, καθώς πρέπει να ταξινομούν μέσω ατέλειωτων συνδυασμούς σχετικά μικρές κρυπτογραφημένες αλλαγές, προκειμένου να ανακτήσουν το μοναδικό περιεχόμενο .



Εικόνα 1.0: Η Δομή ενός φυλλομετρητή

Όπως βλέπουμε από το σχήμα ο ιχνηλάτης αποτελείται από τέσσερα βασικά στοιχεία: τη μονάδα ανάκτησης ιστοσελίδων μέσω του πρωτοκόλλου HTTP, (downloader), τη μονάδα ανάλυσης (parsing) των ιστοσελίδων και εξαγωγή των συνδέσμων που περιέχει, (analyzer), τον χρονοπρογραμματιστή (scheduler),

μονάδα διαχείρισης των URLs, και τέλος την αποθήκη που συγκεντρώνεται το περιεχόμενο των ιστοσελίδων.



Εικόνα 2: Αρχιτεκτονική web φυλλομετρητή

Ο downloader εκτός από την ανάκτηση των ιστοσελίδων δημιουργεί και ένα ευρετήριο με τα URLs των σελίδων. Η επιλογή της σειράς με την οποία επισκέπτονται οι κόμβοι του web καθορίζεται από τη μέθοδο Dequeue του χρονοπρογραμματιστή. Συνήθως η επίσκεψη γίνεται με διάσχιση κατά πλάτος-πρώτα (Breadth First Search). Συνήθως ο ιχνηλάτης χρησιμοποιεί πολλές διαδικασίες κατεβάσματος ιστοσελίδων ταυτόχρονα οι οποίες εκτελούνται σε διαφορετικούς επεξεργαστές. Λεπτομέρειες σχετικά με ιχνηλάτες δεν δημοσιεύονται (αποτελούν μυστικά των εταιρειών) ώστε να περιορίζονται οι κακόβουλοι χρήστες (spammers). Ένας ιχνηλάτης μπορεί να είναι εξειδικευμένος (focused) να συγκεντρώνει σελίδες που είναι σχετικές με ένα προ-καθορισμένο θεματικό πεδίο. Για παράδειγμα το CiteSeerX, μια μηχανή αναζήτησης και ηλεκτρονική βιβλιοθήκη για άρθρα στην επιστήμη των υπολογιστών χρησιμοποιεί εξειδικευμένο crawling για την συγκέντρωση των επιστημονικών άρθρων. Τέλος ένας ιχνηλάτης μπορεί να χρησιμοποιηθεί ως φίλτρο για την συγκέντρωση σελίδων κάποιου συγκεκριμένου τύπου, για παράδειγμα μόνο για HTML σελίδες. Αυτό μπορεί να γίνει όταν ο ιχνηλάτης εξετάζει το URL και μόνο όταν το URL έχει κατάληξη, για παράδειγμα, .html, .htm, να επιτρέπει να κατέβει το αρχείο. Η επίδοση ενός ιχνηλάτη αξιολογείται από τον ρυθμό ανάκτησης σελίδων από το web. Προφανώς ο ρυθμός αυτός έχει περιορισμούς, όπως για παράδειγμα, είναι το εύρος

των δικτύων. Συνήθως η διαδικασία αυτή γίνεται παράλληλα από πολλούς εξυπηρετητές ώστε να μεγιστοποιηθεί ο ρυθμός της συγκέντρωσης του περιεχομένου. Η φύση και το μέγεθος του web κάνουν το έργο του ιχνηλάτη, πολύπλοκο, πολύ δύσκολο και όπως θα δούμε, και ο σχεδιασμός ενός συστήματος ιχνηλάτησης του web αποτελεί μια πρόκληση για πολλά προβλήματα.

Η πολιτική που χρησιμοποιούν οι crawlers²

Η συμπεριφορά ενός Web crawler είναι το αποτέλεσμα ενός συνδυασμού των πολιτικών που περιλαμβάνουν:

- μια πολιτική επιλογή που αναφέρει ποιες σελίδες μπορούν να κατεβάσουν,
- μια εκ νέου επίσκεψη πολιτική που δηλώνει πότε πρέπει να ελέγξουμε για τις αλλαγές στις σελίδες
- μια πολιτική που δηλώνει το πώς να αποφευχθεί η υπερφόρτωση των ιστοσελίδων
- μια πολιτική που δηλώνει πώς να συντονίσει τους καταναμημένους crawlers Ιστού.

2.1 Αρχιτεκτονική των Web Crawlers

Ένας Crawler δεν πρέπει μόνο να έχει μια καλή στρατηγική ανίχνευσης, αλλά θα πρέπει επίσης να έχει μια ιδιαίτερα βελτιστοποιημένη αρχιτεκτονική. Ο Shkapenyuk και Suel³ επισημαίνουν πως αν και είναι αρκετά εύκολο να οικοδομήσουμε έναν αργό crawler που κατεβάζει μερικές σελίδες ανά δευτερόλεπτο για ένα σύντομο χρονικό διάστημα, η δημιουργία ενός υψηλού απόδοσης συστήματος που μπορεί να κατεβάσει εκατοντάδες εκατομμύρια σελίδες κατά τη διάρκεια αρκετών εβδομάδων αλλά παρουσιάζει μια σειρά από προκλήσεις στο σχεδιασμό του συστήματος, I/O και το δίκτυο της αποτελεσματικότητας, και την ευρωστία και τη διαχειρισσιμότητα.

² Castillo, Carlos (2004). Effective Web Crawling (Ph.D. thesis). University of Chile.

³ Shkapenyuk, V. and Suel, T. (2002). Design and implementation of a high performance distributed web crawler

Οι Web Crawlers είναι ένα κεντρικό τμήμα των μηχανών αναζήτησης, οι λεπτομέρειες σχετικά με αλγορίθμους και την αρχιτεκτονική τους, διατηρούνται ως επιχειρηματικά απόρρητα. Όταν δημοσιεύονται σχέδια ανίχνευσης, υπάρχει συχνά μια σημαντική έλλειψη λεπτομέρειας που εμποδίζει τους άλλους να αναπαράγουν το έργο. Υπάρχουν επίσης ανησυχίες που εγείρονται σχετικά με το spamming στη μηχανή αναζήτησης, που εμποδίζουν μεγάλες μηχανές αναζήτησης από τη δημοσίευση στην ιεράρχηση των αλγορίθμων τους.

2.2 Crawler αναγνώρισης

Οι Web Crawlers αυτοπροσδιορίζονται συνήθως σε ένα διακομιστή Web με τη χρήση του User agent πεδίο αίτησης HTTP. Οι Διαχειριστές των ιστοσελίδων Web συνήθως εξετάζουν τους Web servers για να συνδεθούν και να χρησιμοποιήσουν οι users agents να καθορίσουν ποιοι crawlers έχουν επισκεφθεί το web server και πόσο συχνά. Το πεδίο user agent μπορεί να περιλαμβάνει μια διεύθυνση URL όπου ο διαχειριστής ιστοσελίδας μπορεί να βρει περισσότερες πληροφορίες σχετικά με τον crawler.

Εξετάζοντας τη καταγραφή του διακομιστή Web είναι επίπονη εργασία, κάποιιοι διαχειριστές χρησιμοποιούν εργαλεία όπως CrawlTrack⁴ ή SEO Crawllytics για τον εντοπισμό, την παρακολούθηση και την επαλήθευση των crawlers Web. Η Κακόβουλη χρήση και άλλα κακόβουλα προγράμματα ανίχνευσης Web είναι απίθανο να τοποθετήσουν τον προσδιορισμό των πληροφοριών στον τομέα user agent, ή μπορεί να αποκρύψει την ταυτότητά τους, όπως ένα πρόγραμμα περιήγησης.

Είναι σημαντικό για τους Web Crawlers Ιστού να αυτοπροσδιορίζονται έτσι ώστε οι διαχειριστές ιστοσελίδα μπορούν να επικοινωνήσουν με τον ιδιοκτήτη, αν χρειαστεί. Σε ορισμένες περιπτώσεις, οι crawlers μπορεί κατά λάθος να παγιδευτεί σε μια παγίδα ή μπορούν να υπερφορτώσουν ένα διακομιστή Web με τα αιτήματα. Η ταυτοποίηση είναι επίσης χρήσιμο για τους διαχειριστές που ενδιαφέρονται να γνωρίζουν πότε μπορούν να αναμένουν τις ιστοσελίδες τους και να αναπροσαρμόζονται από μια συγκεκριμένη μηχανή αναζήτησης.

⁴ <http://www.crawltrack.net/>

2.3 Το crawling στο «βαθύ» ιστό

Ένα τεράστιο ποσό των ιστοσελίδων που βρίσκονται στο «βαθύ ή αόρατο ιστό»⁵. Αυτές οι σελίδες είναι συνήθως προσβάσιμες μόνο με την υποβολή ερωτημάτων σε μια βάση δεδομένων, και την τακτική που ακολουθούν οι crawlers είναι σε θέση να βρίσκουν αυτές τις σελίδες, αν δεν υπάρχουν σύνδεσμοι που οδηγούν σε αυτά. Το πρωτόκολλο που χρησιμοποιούν τα Sitemaps της Google και το mod_oai⁶ (μοντέλο apache που επιτρέπει το Web Crawler να ανακαλύψουν αποτελεσματικά νέους, τροποποιημένους, πόρους από ένα web server χρησιμοποιώντας το OAI-PMH).

Το Deep web crawling πολλαπλασιάζει επίσης τον αριθμό των συνδέσεων στο διαδίκτυο για να ανιχνεύσει. Μερικοί crawlers παίρνουν μόνο μερικά από τα σε σχήμα URLs. Σε ορισμένες περιπτώσεις, όπως το Googlebot, το Web Crawler γίνεται σε όλο το κείμενο που περιέχεται μέσα στο περιεχόμενο υπερκειμένου ετικέτες, ή κείμενο.

2.3 Ιχνηλάτες Ανοιχτού Κώδικα

- crawler4j, Είναι γραμμένος σε Java και διατίθεται με Apache License.
- Heritrix, Είναι ένας crawler των Internet Archive γραμμένος σε java.
- Nutch, Είναι γραμμένος σε java και διατίθεται με Apache License. Μπορεί να χρησιμοποιηθεί με το πακέτο κατασκευής ευρετηρίων του lucene.
- WebSPHINX, (Website-Specific Processors for HTML INformation eXtraction) είναι μια βιβλιοθήκη κλάσεων της Java και ένα διαδραστικό περιβάλλον ανάπτυξης για ιχνηλάτες του ιστοχώρου.

⁵ Shestakov, Denis (2008). Search Interfaces on the Web: Querying and Characterizing. TUCS Doctoral Dissertations 104, University of Turku

⁶ Michael L Nelson; Herbert Van de Sompel; Xiaoming Liu; Terry L Harrison; Nathan McFarland (2005-03-24). "mod_oai: An Apache Module for Metadata Harvesting"

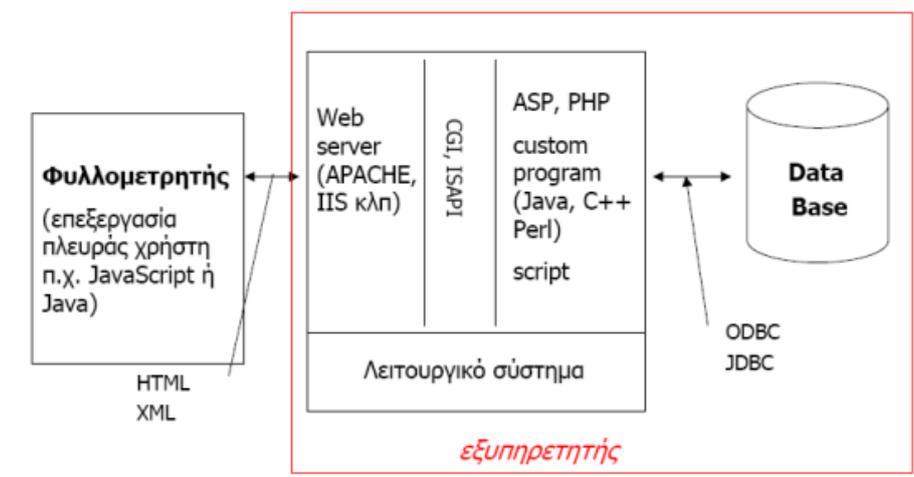
2.4 Δυναμικές Ιστοσελίδες και Αλληλεπίδραση Χρηστών

Είναι δύσκολο να πει κάποιος με ακρίβεια πότε οι «δυναμικές ιστοσελίδες» ξεκίνησαν αλλά η γενικότερη ιδέα ξεκίνησε μετά την εξάπλωση και ανάπτυξη των ιστοσελίδων. Το πρωτόκολλο HTTP άρχισε να χρησιμοποιείται από το 1990 και η γλώσσα HTML ξεκίνησε να χρησιμοποιείται για τον σχεδιασμό ιστοσελίδων από το 1996. Η εξάπλωση χρήσης των φυλλομετρητών ξεκίνησε το 1993 με τον φυλλομετρητή Mosaic. Η ιδέα της δημιουργίας δυναμικών ιστοσελίδων υπήρχε και πριν την ανάπτυξη των ιστοσελίδων. Για παράδειγμα πριν το 1990 υπήρχαν «Βάσεις Δεδομένων » (Bulletin Board Systems) στα οποία οι χρήστες συνδέονταν μέσω μόντεμ και το περιεχόμενο που έβλεπαν δημιουργούνταν δυναμικά.

Δυναμική ιστοσελίδα είναι μια ιστοσελίδα η οποία δημιουργείται δυναμικά την στιγμή της πρόσβασης σε αυτή ή την στιγμή που ο χρήστης αλληλεπιδρά με τον εξυπηρετητή ιστοσελίδων. Οι δυναμικές ιστοσελίδες θεωρούνται δομικό στοιχείο της νέας γενιάς του παγκόσμιου ιστού (Web 2.0) όπου η πληροφορία διαμοιράζεται σε πολλαπλές ιστοσελίδες. Η δυναμική ιστοσελίδα μπορεί να δημιουργείται δυναμικά από ένα σενάριο εντολών, το οποίο εκτελείται τοπικά στο πελάτη ή στον εξυπηρετητή ή και στον πελάτη και στον εξυπηρετητή.

Με άλλα λόγια αλληλοεπιδρά με τους επισκέπτες της ιστοσελίδας δίνοντας τους τη δυνατότητα για παράδειγμα να σχολιάσουν , να ψηφίσουν, να συμπληρώσουν φόρμες και να λαμβάνουν ενημερωτικά email (νέα της ιστοσελίδας ή προφορές). Η αλληλεπίδραση της ιστοσελίδας με τους επισκέπτες, τους κάνει να ξανάρχονται στην ιστοσελίδα σας. Τέλος γίνεται λοιπόν εμφανές το πρόβλημα των στατικών σελίδων έγκειται στο ότι αυτές δεν δίνουν καμία δυνατότητα αλληλεπίδρασης μεταξύ χρήστη και server πάρα παρέχουν στείρα πληροφορία.

Η αρχιτεκτονική λοιπόν που θα βασίζεται μία δυναμική σελίδα που θα αλληλοεπιδρά με τον χρήστη είναι η εξής:



Εικόνα 3: Αρχιτεκτονική Δυναμικών Ιστοσελίδων

3.0 Προγραμματιστικό Περιβάλλον- Πλατφόρμα Ανάπτυξης

α) Η Πλατφόρμα Microsoft .Net

Η Microsoft ξεκίνησε την ανάπτυξη του .NET Framework στα τέλη της δεκαετίας του 1990, αρχικά με το όνομα της Next Generation Windows Services (NGWS). Μέχρι τα τέλη του 2000, οι πρώτες εκδόσεις beta του .NET 1.0 κυκλοφόρησαν.

Η έκδοση 3.0 του .NET Framework περιλαμβάνεται στα Windows Server 2008 και Windows Vista. Η έκδοση 3.5 περιλαμβάνεται στα Windows 7 και Windows Server 2008 R2, και μπορεί επίσης να εγκατασταθεί σε Windows XP και Windows Server 2003⁷. Στις 12 Απριλίου 2010, το .NET Framework 4 κυκλοφόρησε μαζί με το Visual Studio 2010.

Το .NET Framework οικογένεια περιλαμβάνει επίσης δύο εκδόσεις για κινητά ή ενσωματωμένης χρήσης της συσκευής. Μια μειωμένη έκδοση του πλαισίου, το .NET Compact Framework, είναι διαθέσιμο για πλατφόρμες Windows CE, συμπεριλαμβανομένων των Windows Mobile συσκευές, όπως smartphones. Τέλος το .NET Framework Micro απευθύνεται σε αυστηρά περιορισμένους πόρους συσκευές.

Το .NET Framework (προφέρεται dot net) περιλαμβάνει μια μεγάλη βιβλιοθήκη και παρέχει διαλειτουργικότητα γλώσσας (κάθε γλώσσα μπορεί να χρησιμοποιήσει κώδικα γραμμένο σε άλλες γλώσσες) σε διάφορες γλώσσες προγραμματισμού . Προγράμματα που έχουν γραφτεί για το .NET Framework εκτελούνται σε ένα περιβάλλον λογισμικού (σε αντίθεση με το περιβάλλον υλικό) , που είναι γνωστή ως το Common Language Runtime (CLR) , μια εφαρμογή εικονικής μηχανής που παρέχει υπηρεσίες , όπως η ασφάλεια , η διαχείριση μνήμης , και το χειρισμό εξαιρεση. Η βιβλιοθήκη τάξη και η CLR μαζί αποτελούν το .NET Framework .

Η βασική βιβλιοθήκη του .NET Framework είναι Class παρέχει διεπαφή χρήστη, πρόσβαση σε δεδομένα , σύνδεσης με βάσεις δεδομένων , κρυπτογραφία ,

⁷ <http://msdn.microsoft.com/en-US/library/cc160717%28v=vs.90%29.aspx>

ανάπτυξη web εφαρμογών, αριθμητική αλγόριθμους , και στο δίκτυο επικοινωνιών . Οι προγραμματιστές παράγουν λογισμικό συνδυάζοντας το δικό τους κώδικα τους μέσα στα πλαίσια του .NET και άλλες βιβλιοθήκες.

Το . NET Framework προορίζεται να χρησιμοποιείται από τις περισσότερες νέες εφαρμογές που δημιουργούνται για την πλατφόρμα των Windows . Η Microsoft παράγει επίσης ένα ολοκληρωμένο περιβάλλον ανάπτυξης σε μεγάλο βαθμό για . NET λογισμικό που ονομάζεται Visual Studio .

3.1 Το Visual Studio .NET ως πλατφόρμα ανάπτυξης

Το Microsoft Visual Studio είναι ένα ολοκληρωμένο περιβάλλον ανάπτυξης (IDE) από τη Microsoft. Χρησιμοποιούνται για την ανάπτυξη της κονσόλας και γραφικές εφαρμογές διεπαφή χρήστη , μαζί με τα Windows Forms και WPF εφαρμογές, ιστοσελίδες , εφαρμογές web και των υπηρεσιών web , τόσο εγγενή κώδικα μαζί με διαχειριζόμενο κώδικα για όλες τις πλατφόρμες που υποστηρίζονται από τα Microsoft Windows , Windows Mobile , Windows CE , . NET Framework , . NET Compact Framework και το Microsoft Silverlight .

Το Visual Studio περιλαμβάνει έναν επεξεργαστή κώδικα υποστήριξη IntelliSense καθώς και ανακατασκευασμένο κώδικα. Το ολοκληρωμένο πρόγραμμα εντοπισμού σφαλμάτων λειτουργεί τόσο ως ένα πρόγραμμα εντοπισμού σφαλμάτων πηγής επίπεδο και σαν ένα μηχανήμα σε επίπεδο εντοπισμού σφαλμάτων . Άλλα ενσωματωμένα εργαλεία περιλαμβάνουν έναν «σχεδιαστή» για τη δημιουργία GUI εφαρμογών , web designer και τη βάση δεδομένων σχεδιαστή σχήματος .

Με το Visual Studio .NET, μπορεί να μετατραπεί εύκολα μια υπάρχουσα επιχειρηματική λογική σε υπηρεσίες Web της XML πολλαπλής χρήσης, συγκεντρώνοντας διαδικασίες και καθιστώντας τις διαθέσιμες σε οποιαδήποτε πλατφόρμα. Μπορούν να ενσωματωθούν εύκολα οποιουσδήποτε αριθμός υπηρεσιών Web οι οποίες περιλαμβάνονται και είναι διαθέσιμες σε πολλούς ανεξάρτητους καταλόγους UDDI, παρέχοντας στις εφαρμογές τους μια ισχυρή βάση όσον αφορά τις υπηρεσίες και την επιχειρηματική λογική.

Κάνοντας χρήση της XML, μιας τεχνολογίας βιομηχανικού προτύπου για την περιγραφή δεδομένων, μπορούν να δημιουργηθούν εφαρμογές υψηλών επιδόσεων που βασίζονται σε δεδομένα. Μπορούν να χρησιμοποιηθούν ενσωματωμένα εργαλεία ADO.NET τα οποία προορίζονται για μια ποικιλία βάσεων δεδομένων, συμπεριλαμβανομένης της βάσης δεδομένων SQL Server, της βάσης δεδομένων Oracle ή οποιασδήποτε άλλης προέλευσης XML. Με την εσωτερική υποστήριξη για XML, το εργαλείο ADO.NET επιτρέπει στους προγραμματιστές να κάνουν κοινή χρήση δεδομένων σε διαφορετικές πλατφόρμες υπολογιστών. Επιπλέον, το Visual Studio .NET περιλαμβάνει το μηχανισμό Microsoft Data

Engine (MSDE), μια βάση δεδομένων 100% συμβατή με τον SQL Server η οποία προσφέρει στους προγραμματιστές μια χρησιμοποιήσιμη βάση δεδομένων προγραμματισμού και υποστηρίζει τοπικά XML για μέγιστη διαλειτουργικότητα.

Από την άλλη μεριά Το Visual Basic .NET περιλαμβάνει τη γνωστή στους προγραμματιστές σύνταξη Visual Basic συν προαιρετικές δυνατότητες προγραμματισμού προσανατολισμένες στα αντικείμενα, συμπεριλαμβανομένης της μεταβίβασης και άλλων προαιρετικών δυνατοτήτων όπως το Structured Exception Handling και το Free-Threading.

Η Visual C# .NET, η νέα γλώσσα της Microsoft, προσφέρει αυξημένη παραγωγικότητα στον προγραμματιστή της C και της C++. Η Visual C# .NET παρέχει υποστήριξη υψηλού επιπέδου για στοιχεία με ιδιότητες, μεθόδους, ευρετήρια, χαρακτηριστικά, ιστορικό εκδόσεων και συμβάντα, ενώ παρέχει ισχυρή και αποτελεσματική υποστήριξη για την πλατφόρμα Microsoft .NET.

3.2 Αρχιτεκτονική του Visual Studio .NET

ΤΟ Visual Studio δεν υποστηρίζει οποιαδήποτε γλώσσα προγραμματισμού, αντί να επιτρέπει τη σύνδεση των λειτουργιών κωδικοποιείται ως VSPackage . Όταν εγκατασταθεί, η λειτουργία είναι διαθέσιμη ως υπηρεσία .

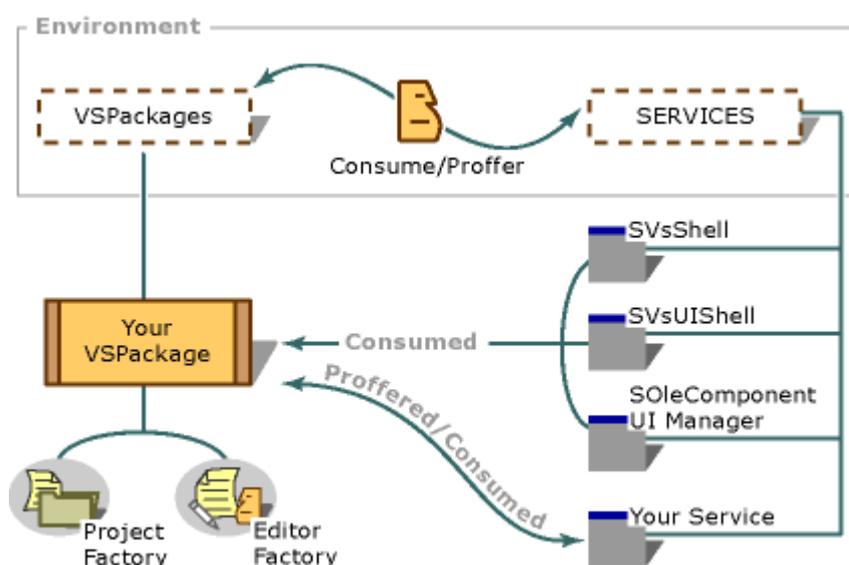
Το IDE παρέχει τρεις υπηρεσίες :

- SVsSolution , η οποία παρέχει τη δυνατότητα να απαριθμήσει τα έργα και τις λύσεις
- SVsUIShell , η οποία παρέχει παραθύρων και τη λειτουργικότητα UI (συμπεριλαμβανομένων των καρτελών , γραμμές εργαλείων και εργαλείων Windows)
- SVsShell , η οποία ασχολείται με την καταγραφή των VSPackages.

Επιπλέον , η IDE είναι επίσης υπεύθυνη για τον συντονισμό και επιτρέπει την επικοινωνία μεταξύ των υπηρεσιών⁸. Όλα οι συντάκτες , σχεδιαστές , τύποι

⁸ [http://msdn.microsoft.com/en-us/library/bb165114\(VS.80\).aspx](http://msdn.microsoft.com/en-us/library/bb165114(VS.80).aspx)

έργων και άλλα εργαλεία εφαρμόζονται ως VSPackages . Το Visual Studio χρησιμοποιεί COM για να μεταβεί στα VSPackages . Το Visual Studio SDK περιλαμβάνει επίσης το Managed Package Framework(MPF) , το οποίο είναι ένα σύνολο των υπό διαχειρίσεων γύρω από τις διεπαφές COM- που επιτρέπουν τα πακέτα να γραφτούν σε οποιαδήποτε γλώσσα με συμβατό CLI ⁹. Ωστόσο , το MPF δεν παρέχει όλη τη λειτουργικότητα που εκτίθενται από τις οπτικές διασυνδέσεις COM Studio . Οι υπηρεσίες μπορούν στη συνέχεια να καταναλωθούν για τη δημιουργία άλλων πακέτων, τα οποία προσθέτουν λειτουργικότητα στο Visual Studio IDE .



Εικόνα 4: Αποψη της αρχιτεκτονικής του Visual Studio

Υποστηρίζει γλώσσες προγραμματισμού, προστίθεται με τη χρήση συγκεκριμένου πακέτου VSPackage που ονομάζεται Υπηρεσία Γλώσσας. Μια υπηρεσία γλώσσας ορίζει διάφορες διεπαφές που η εφαρμογή VSPackage προσθέτει την υποστήριξη για διάφορες λειτουργίες¹⁰.

Οι λειτουργίες που μπορούν να προστεθούν με αυτόν τον τρόπο περιλαμβάνουν τη σύνταξη, αγκύλες, επεξηγήσεις πληροφοριών παραμέτρων, λίστες μελών και

⁹ http://www.codeguru.com/csharp/.net/net_general/visualstudionetadd-ins/article.php/c11835/Extending-Visual-Studio-2005.htm

¹⁰ [http://msdn.microsoft.com/en-us/library/bb166391\(VS.80\).aspx](http://msdn.microsoft.com/en-us/library/bb166391(VS.80).aspx)

δείκτες σφάλματος για το φόντο . Αν η διεπαφή υλοποιείται, η λειτουργία θα είναι διαθέσιμη για τη γλώσσα. Οι εφαρμογές μπορούν να επαναχρησιμοποιηθούν για τον κώδικα από το πρόγραμμα ανάλυσης ή του μεταγλωττιστή της γλώσσας. Οι υπηρεσίες της γλώσσας μπορούν να υλοποιηθούν είτε σε εγγενή κώδικα ή διαχειριζόμενο κώδικα. Για την εγγενή κώδικα, είτε οι εγγενείς διεπαφές COM ή το πλαίσιο Babel (μέρος του Visual Studio SDK) μπορεί να χρησιμοποιηθεί¹¹. Για διαχειριζόμενο κώδικα, η MPF περιλαμβάνει τη σύνταξη για διαχείριση γλωσσικών υπηρεσιών¹². Το Visual Studio δεν περιλαμβάνει καμία υποστήριξη ελέγχου προέλευσης που χτίστηκε, αλλά ορίζει δύο εναλλακτικούς τρόπους για συστήματα ελέγχου πηγή για να ενσωματωθούν με το IDE¹³. Ένα πακέτο VSPackage ελέγχου μπορεί να προσφέρει δικιά του προσαρμοσμένη διεπαφή του χρήστη .

Σε αντίθεση , ένα plugin για τον έλεγχο πηγή, με την MSSCCI παρέχει μια σειρά από λειτουργίες που χρησιμοποιούνται για την εφαρμογή των διαφόρων λειτουργιών ελέγχου προέλευσης , με ένα πρότυπο Visual Studio διεπαφή χρήστη . Το MSSCCI(Microsoft Source Code Control Interface) χρησιμοποιήθηκε για πρώτη φορά για την ενσωμάτωση του Visual SourceSafe με το Visual Studio 6.0 , αλλά αργότερα άνοιξε μέσω του Visual Studio SDK . Το Visual Studio . NET 2002 που χρησιμοποιεί το MSSCCI 1.1 και το Visual Studio . NET 2003 που χρησιμοποιεί το MSSCCI 1.2 . Το Visual Studio 2005 , 2008 και 2010, η χρήση του MSSCCI Version 1.3 , που προσθέτει υποστήριξη για να μετονομάσουμε και να διαγράψουμε ένα ασύγχρονο άνοιγμα .

Το Visual Studio εκτελεί πολλαπλές παρουσίες του περιβάλλοντος (το καθένα με το δικό του σύνολο των VSPackages). Οι περιπτώσεις που χρησιμοποιούν διαφορετικές ομάδες μητρώου για να αποθηκεύσουμε τη κατάσταση ρύθμισης και διαφοροποιούνται από AppID τους (ID Εφαρμογής) . Το Visual Studio Express edition εγκαθίσταται με τη δική τους AppIds , αλλά οι Standard,

¹¹ [http://msdn.microsoft.com/en-us/library/bb165670\(VS.80\).aspx](http://msdn.microsoft.com/en-us/library/bb165670(VS.80).aspx)

¹² [http://msdn.microsoft.com/en-us/library/bb166360\(VS.80\).aspx](http://msdn.microsoft.com/en-us/library/bb166360(VS.80).aspx)

¹³ <http://msdn.microsoft.com/en-us/library/bb165370.aspx>

Professional και η ομάδα προϊόντων Suite μοιράζονται την ίδια AppID . Κατά συνέπεια , μπορεί κανείς να εγκαταστήσει τις εκδόσεις Express side- by-side με άλλες εκδόσεις , σε αντίθεση με τις άλλες εκδόσεις που ενημερώνουν την ίδια εγκατάσταση . Η επαγγελματική έκδοση περιλαμβάνει ένα υπερσύνολο των VSPackages στην βασική έκδοση και η σουίτα ομάδα περιλαμβάνει ένα υπερσύνολο των VSPackages στις δύο άλλες εκδόσεις.

3.3 Internet Information Server ως Web Server

α) Τι είναι ένας Web Server

Ένας web server μπορεί να σημαίνει δύο πράγματα - ένα υπολογιστή στον οποίο μια ιστοσελίδα φιλοξενείται και ένα πρόγραμμα που τρέχει σε έναν τέτοιο υπολογιστή. Έτσι, ο όρος του web server αναφέρεται σε υλικό και λογισμικό.

3.3.1 Ο υπολογιστής Web Server ως υλικό

Μια ιστοσελίδα είναι μια συλλογή από ιστοσελίδες που είναι ψηφιακά αρχεία συνήθως γραμμένο με HyperText Markup Language (HTML) . Για μια ιστοσελίδα για να είναι διαθέσιμη σε όλους στον κόσμο, ανά πάσα στιγμή , θα πρέπει να αποθηκεύονται ή φιλοξενείται σε έναν υπολογιστή που είναι συνδεδεμένος στο διαδίκτυο. Ένας τέτοιος υπολογιστής είναι γνωστός ως ένα διακομιστής Web.

Μπορούμε να φιλοξενήσουμε έναν ιστοχώρο στον υπολογιστή του σπιτιού μας, αλλά αυτό συνεπάγεται πολλή δουλειά και συνεχή παρακολούθηση. Είναι πιο εύκολο να «αγοράσουμε» ένα web hosting από μια εταιρεία, διότι υπάρχουν χιλιάδες που προσφέρουν αυτή την υπηρεσία σε λογικές τιμές .

Υπάρχουν πολλές απαιτήσεις για έναν υπολογιστή Server - πρέπει να είναι γρήγορος, έχει μεγάλη χωρητικότητα αποθήκευσης σκληρό δίσκο και πολλή RAM . Αλλά το πιο σημαντικό είναι να έχει μια μόνιμη διεύθυνση στο διαδίκτυο , επίσης γνωστή ως μια IP (Internet Protocol) . Εάν η IP αλλάζει διεύθυνση , η ιστοσελίδα δεν θα βρεθεί και θα εμφανίζετα offline.

β) Πρόγραμμα του web server ως λογισμικό

Ένα πρόγραμμα web server είναι λογισμικό που τρέχει στο χώρο του web hosting Server στου υπολογιστή. Ο κύριος σκοπός του είναι να εξυπηρετεί τις σελίδες, αυτό σημαίνει ότι περιμένει αιτήσεις από προγράμματα περιήγησης στο Web και αποκρίνεται με την αποστολή των απαιτούμενων στοιχείων πίσω. Ο client-server αλληλεπίδραση είναι το σήμα κατατεθέν για το πώς λειτουργεί το διαδίκτυο.

Υπάρχουν πολλά προγράμματα του web server διαθέσιμο στο κατάλογο ελέγχου των web servers. Το πιο γνωστό και δημοφιλές από αυτά είναι όλα ο Apache έχει αναπτυχθεί από την Apache Foundation. Δεν είναι μόνο Apache δωρεάν, αλλά είναι επίσης διαθέσιμο για διάφορα λειτουργικά συστήματα όπως τα Windows, Macintosh και Linux/ Unix.

Η FYI (Η σειρά FYI των σημειώσεων έχει σχεδιαστεί για να παρέχει στους χρήστες του Διαδικτύου με ένα κεντρικό αποθετήριο πληροφοριών σχετικά με τυχόν θέματα που σχετίζονται με το Διαδίκτυο), διαφοροποιώντας το πρόγραμμα server (software) και ο υπολογιστής Server (hardware).

3.3.2 Ποιος είναι ο ρόλος του web server στο Internet

Ο Web Server- ο υπολογιστής ή το πρόγραμμα - έχουν σημαντικό ρόλο στο Internet . Ο server που έχει «εγκαταστήσει» την ιστοσελίδα στο σκληρό δίσκο του, ενώ το πρόγραμμα server βοηθά παραδώσει στις ιστοσελίδες και τα σχετικά αρχεία τους, όπως εικόνες , ταινίες σε πελάτες και άλλα.

Η διαδικασία της φόρτωσης μιας ιστοσελίδας / σελίδα σε ένα πρόγραμμα περιήγησης στο web ξεκινά με το χρήστη που εισέρχεται είτε τη διεύθυνση URL στη γραμμή διεύθυνσης ή κάνοντας κλικ σε ένα σύνδεσμο . Θα πρέπει να ξέρουμε ότι κάθε ιστοσελίδα έχει μια μοναδική διεύθυνση στο διαδίκτυο που σημαίνει ότι η ίδια η σελίδα δεν μπορεί να υπάρξει σε δύο μέρη.

Μόλις το αίτημα ξεκινήσει από τον χρήστη , ο browser στέλνει ένα αίτημα για την ιστοσελίδα . Η διεύθυνση URL της σελίδας που ζητήθηκε από το web έχει επιλυθεί σε μια διεύθυνση IP διεύθυνση , η οποία , στην αγγλική γλώσσα , μέσα , μετατρέπεται σε μια IP διεύθυνση - κάτι που καταλαβαίνουν οι υπολογιστές . Η IP διεύθυνση απευθύνει στη θέση της υποδοχής της ιστοσελίδας . Η αίτηση διαβιβάζεται στον υπολογιστή Server που πέρασε με το λογισμικό μέσω του διακομιστή.

Το λογισμικό διακομιστή πλέον «εργάζεται» και «κυνηγά» για τη ζητούμενη ιστοσελίδα στο σκληρό δίσκο . Με την εύρεση του αρχείου , στέλνει μια απάντηση στον browser που ακολούθησε το πραγματικό αρχείο ιστοσελίδα η οποία στη συνέχεια αρχίζει να εμφανίζει τη σελίδα .

Μια τυπική ιστοσελίδα δεν έχει μόνο κείμενο, αλλά επίσης ενσωματωμένο στοιχεία πολυμέσων , όπως εικόνες και κινούμενα σχέδια Flash . Αυτά τα «έξτρα» αρχεία είναι χωριστά από την πραγματική ιστοσελίδα και είναι δεν βρίσκονται στο πρόγραμμα περιήγησης.

Σημείωση: μόνο το πρόγραμμα περιήγησης στο web καθορίζει το πώς μια ιστοσελίδα εμφανίζεται. Ο web server δεν έχει κανέναν έλεγχο πάνω σε αυτό . Η δουλειά του web server τελειώνει τη στιγμή που οι αιτήσεις από το πρόγραμμα περιήγησης επεξεργάζεται και οι απαιτούμενες πληροφορίες αποστέλλονται .

Αν και μπορεί να φαίνεται ότι η διαδικασία αίτησης - και - απάντηση παίρνει πολύ χρόνο , ειδικά αν σκεφτεί κανείς ότι ο πελάτης υπολογιστές και διακομιστές μπορεί να είναι χιλιάδες μίλια μακριά , αυτό συμβαίνει στην πραγματικότητα πολύ γρήγορα . Αυτό συμβαίνει γιατί από πρωτόκολλο Hypertext Transfer Protocol (HTTP), το οποίο θέτει ένα σύνολο κανόνων για να διευκολύνει τη μεταφορά δεδομένων μέσω του διαδικτύου.

Για την επιλογή ενός server οι παράγοντες που πρέπει να εξεταστούν είναι:

- **Ταχύτητα**

Κατά κύριο λόγο η ταχύτητα ανταπόκρισης ενός server στις εισερχόμενες κλήσεις του εξαρτάται από το εύρος ζώνης που έχει η σύνδεση του με το internet και κατά δεύτερο λόγο από της δυνατότητας του hardware(ram, cpu, hard disk) και το είδος του λογισμικού που χρησιμοποιεί.

- **Ασφάλεια**

Αμέσως μετά την ταχύτητα η ασφάλεια θα πρέπει να είναι το μεγαλύτερο μέλημα. Ο server θα πρέπει να υποστηρίζει S-HTTP ή SSL και αυτό γιατί αυτά είναι τα πλησιέστερα που υπάρχουν σε πρότυπα ασφαλείας στο web.

- **Ευκολία εγκατάστασης και διαχείρισης**

Γενικά η εγκατάσταση ενός web-server θα πρέπει να είναι το ίδιο εύκολη με την εγκατάσταση οποιοδήποτε άλλου προγράμματος. Επίσης θα πρέπει να παρέχει ένα πρόγραμμα που θα συνδυάζει όλες τις δυνατότητες διαχείρισης σε ένα κοινό, εύχρηστο περιβάλλον από το οποίο ο administrator θα έχει γρήγορη πρόσβαση στις λειτουργίες του server όπως για παράδειγμα το ημερολόγιο πρόσβασης.

3.4 Apache HTTP εξυπηρετητής

Ο Apache HTTP γνωστός και απλά σαν Apache είναι ένας εξυπηρετητής του παγκόσμιου ιστού (web). Όποτε ένας χρήστης επισκέπτεται ένα ιστότοπο το πρόγραμμα πλοήγησης (browser) επικοινωνεί με έναν διακομιστή (server) μέσω του πρωτοκόλλου HTTP, ο οποίος παράγει τις ιστοσελίδες και τις αποστέλλει στο πρόγραμμα πλοήγησης. Ο Apache είναι ένας από τους δημοφιλέστερους εξυπηρετητές ιστού, εν μέρει γιατί λειτουργεί σε διάφορες πλατφόρμες όπως τα Windows, το Linux, το Unix και το Mac OS X. Κυκλοφόρησε υπό την άδεια λογισμικού Apache και είναι λογισμικό ανοιχτού κώδικα. Συντηρείται από μια κοινότητα ανοιχτού κώδικα με επιτήρηση από το Ίδρυμα Λογισμικού Apache (Apache Software Foundation). Ο Apache χρησιμοποιείται και σε τοπικά δίκτυα σαν διακομιστής συνεργαζόμενος με συστήματα διαχείρισης Βάσης Δεδομένων π.χ. Oracle, MySQL.

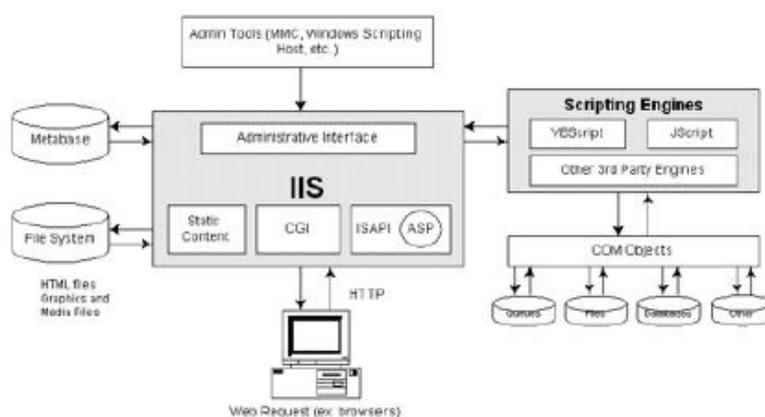
Η πρώτη του έκδοση, γνωστή ως NCSA HTTPd, δημιουργήθηκε από τον Robert McCool και κυκλοφόρησε το 1993. Θεωρείται ότι έπαιξε σημαντικό ρόλο στην αρχική επέκταση του παγκόσμιου ιστού. Ήταν η πρώτη βιώσιμη εναλλακτική επιλογή που παρουσιάστηκε απέναντι στον εξυπηρετητή http της εταιρείας Netscape και από τότε έχει εξελιχθεί στο σημείο να ανταγωνίζεται άλλους εξυπηρετητές βασισμένους στο Unix σε λειτουργικότητα και απόδοση. Από το 1996 ήταν από τους πιο δημοφιλείς όμως από τον Μάρτιο του 2006 έχει μειωθεί το ποσοστό της εγκατάστασής του κυρίως από τον Microsoft Internet Information Services και την πλατφόρμα .NET. Τον Οκτώβριο του 2007 το μερίδιό του ήταν

47.73% από όλους τους ιστοτόπους. Επιπλέον ο Apache χρησιμοποιείται σήμερα ευρέως γιατί α)είναι δωρεάν, β)δουλεύει σε unix και OS/2 και γ) είναι πάρα πολύ γρήγορος.

Τέλος το κύριο μειονέκτημα του είναι ότι ενώ η δημιουργία στατικών σελίδων είναι εύκολη, η δημιουργία δυναμικών σελίδων είναι αρκετά δύσκολη αφού ο server δεν παρέχει τεχνολογίες υποστήριξης.

3.5 IIS(Internet Information Server)

Ο πρώτος διακομιστής Microsoft στο web ήταν ένα ερευνητικό πρόγραμμα στο Ευρωπαϊκό Microsoft Windows NT Ακαδημαϊκό Κέντρο (EMWAC), τμήμα του Πανεπιστημίου του Εδιμβούργου στη Σκωτία, και διανεμήθηκε ως δωρεάν λογισμικό. Εντούτοις, δεδομένου ότι ο EMWAC διακομιστής δεν ήταν δυνατό να κλιμακώνεται και να χειριστεί τον όγκο της κυκλοφορίας που πηγαινει στη Microsoft.com, η Microsoft αναγκάστηκε να αναπτύξει το δικό του web server, IIS¹⁴.

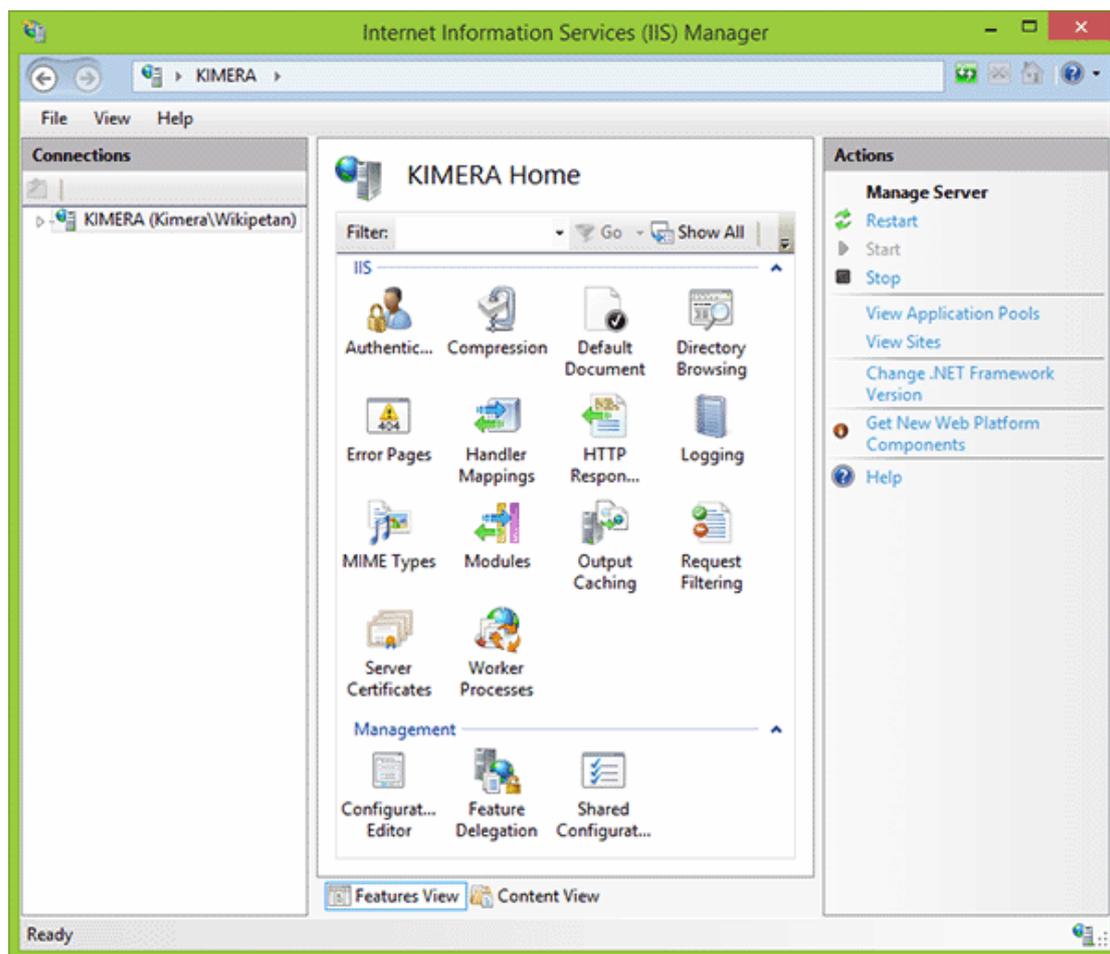


Εικόνα 5: IIS(Internet Information Server)

Ο Internet Information Services (IIS, πρώην Internet Information Server) είναι μια επέκταση του web server που δημιουργήθηκε από τη Microsoft για χρήση των Windows NT. Ο IIS υποστηρίζει HTTP, HTTPS, FTP, FTPS, SMTP και NNTP.

¹⁴ http://www.microsoft.com/misc/features/features_fishbk.htm

Αποτελεί αναπόσπαστο μέρος της οικογένειας των Windows NT και των Windows NT 4.0, αν και μπορεί να μην υπάρχει από κάποια έκδοση (π.χ. Windows XP Home Edition). Ο IIS δεν είναι ενεργοποιημένος από προεπιλογή όταν είναι εγκατεστημένα τα Windows. Ο διαχειριστής του IIS είναι προσβάσιμο μέσα από την κονσόλα διαχείρισης της Microsoft ή από τα εργαλεία διαχείρισης στον Πίνακα Ελέγχου.



Εικόνα 6: Κονσόλα διαχείρισης του IIS

Συνοψίζοντας ο IIS είναι ολοκληρωμένος με το λειτουργικό σύστημα Windows Server γεγονός που του επιτρέπει να χρησιμοποιεί των ίδιο κατάλογο χρηστών και έτσι παρέχει υψηλή ασφάλεια ελέγχοντας με αυτόν τον τρόπο τα δικαιώματα που έχει κάθε χρήστης σε εφαρμογές στο διαδίκτυο. Υποστηρίζει SSL και παρέχει μια σειρά από λειτουργίες που αυξάνουν την απόδοση του και τον καθιστούν

ικανό να ανταποκριθεί σε οποιαδήποτε απαίτηση για server. Το δυνατότερο σημείο του IIS είναι η υποστήριξη που παρέχει στην εύκολη ανάπτυξη δυναμικών σελίδων στο web αφού είναι ο server που υποστηρίζει το περιβάλλον ανάπτυξης εφαρμογών στο web : active server pages.

3.5.1 Πρωτόκολλο TCP/IP

Οι client εφαρμογές δικτύου επικοινωνούν με τις υπηρεσίες του IIS χρησιμοποιώντας το πρωτόκολλο TCP/IP. Μερικές φορές καλείται και Internet Protocol Suite.

Το πρωτόκολλο TCP/IP έχει τέσσερα επίπεδα:

- Επίπεδο σύνδεσης

Το επίπεδο σύνδεσης δημιουργεί το interface των Windows NT στο δίκτυο. Χειρίζεται όλες τις λεπτομέρειες σχετικά με τη σύνδεση μέσω interface του οδηγού συσκευής δικτύου με την κάρτα interface δικτύου.

- Επίπεδο δικτύου

Το επίπεδο δικτύου χειρίζεται την κίνηση των πακέτων στο δίκτυο. Το IP είναι σε αυτό το επίπεδο. Αποκρύπτει το βασικό φυσικό δίκτυο δημιουργώντας μια όψη εικονικού δικτύου. Το επίπεδο δικτύου είναι ένα αναξιόπιστο, best-effort connectionless πρωτόκολλο παράδοσης πακέτου. Μετακινεί πακέτα από την πηγή στον προορισμό αλλά δεν υπάρχει εγγύηση ότι αυτά θα φτάσουν εκεί.

- Επίπεδο μεταφοράς

Το επίπεδο μεταφοράς χειρίζεται τη ροή των δεδομένων μεταξύ της εφαρμογής client και του επιπέδου εφαρμογών. Το TCP (Transmission Control Protocol) είναι σε αυτό το επίπεδο. Το TCP παρέχει ανάκαμψη από σφάλματα και έλεγχο ροής των δεδομένων μεταξύ της client εφαρμογής δικτύου και των υπηρεσιών IIS.

- Επίπεδο εφαρμογών

Το επίπεδο εφαρμογών χειρίζεται τις συνδέσεις με τα επίπεδα δικτύου και μεταφοράς. Δεν σχετίζεται με το πώς τα δεδομένα κινούνται στο δίκτυο. Το Windows Sockets interface βρίσκεται στο επίπεδο εφαρμογών.

4.0 Βασικές Υπηρεσίες IIS

α)Υπηρεσία FTP

Η υπηρεσία FTP του IIS μπορεί να μεταφέρει οποιοδήποτε αρχείο μεταξύ του server κι ενός FTP client. Χειρίζεται τη σύγχρονη πρόσβαση από πολλαπλούς FTP clients. Κάθε FTP client εγκαθιστά μία σύνδεση socket στην υπηρεσία FTP του IIS και συνδέεται με αυτήν. Οι FTP clients χρησιμοποιούν ένα περιορισμένο σύνολο εντολών και έχουν περιορισμένη πρόσβαση αρχείων. Η σύνδεση socket στην υπηρεσία FTP διαρκεί έως ότου ο FTP client αποσυνδεθεί. Το FTP είναι ένα από τα πρώτα πρωτόκολλα TCP/IP. Οι Web browsers και άλλες γραφικές interface εφαρμογές έχουν αντικαταστήσει τις πρώιμες FTP client εφαρμογές. Οι περισσότερες υπηρεσίες FTP δεν παρέχουν περιγραφές αρχείων. Το browsing στους καταλόγους είναι μία αργή διαδικασία.

β)Υπηρεσία Gopher

Το Internet Gopher είναι ένα εργαλείο για browsing σε αρχεία και καταλόγους στο Internet. Ένας Gopher client εγκαθιστά μια socket σύνδεση στην υπηρεσία Gopher του IIS. Συνήθως δεν απαιτεί από τον client να συνδεθεί. Ο Gopher client εμφανίζει μια ιεραρχία από αντικείμενα και καταλόγους, όπως περίπου ένα σύστημα αρχείων σε ένα μενού από επιλογές. Μπορεί να είναι μια λίστα από αρχεία, υποκαταλόγους ή συνδυασμό των δύο. Ο Gopher client αντιγράφει ένα επιλεγμένο αρχείο στο δίκτυο και το εμφανίζει. Το μενού Gopher μπορεί να δείχνει σε άλλα αρχεία και καταλόγους που βρίσκονται σε άλλους Gopher servers στο Internet. Ήταν η πρώτη υπηρεσία Internet που προσέφερε αυτό το χαρακτηριστικό. Τέλος το Gopher έχει περιορισμένες δυνατότητες γραφικής αναπαράστασης. Δεν μπορεί να παρουσιάζει γραφικά και κείμενο ταυτόχρονα.

Το Gopher και το HTTP είναι παρόμοια πρωτόκολλα δικτύου. Έγιναν διαθέσιμα και τα δύο περίπου την ίδια περίοδο. Οι περισσότερες σελίδες στο Internet δεν προσφέρουν τις υπηρεσίες Gopher. Πολλές παλιότερες σελίδες έχουν σταματήσει να τις προσφέρουν. Έχουν μετατρέψει τα έγγραφα Gopher σε έγγραφα HTML επειδή το HTML μπορεί να παρουσιάζει γραφικά και κείμενο μαζί. Τα έγγραφα HTML βασίζονται στο πρωτόκολλο HTTP.

4.1 Η υπηρεσία WWW

Ο Web browser που επεξεργάζεται έγγραφα HTML χρησιμοποιεί το πρωτόκολλο HTTP ώστε να συναλλαγεί με την υπηρεσία WWW του IIS. Η υπηρεσία WWW γνωρίζει πώς να αποκριθεί σε μια αίτηση HTML αναλύοντας τα πεδία URL. Μπορεί να καθορίσει αν η αίτηση είναι για μια στατική HTML σελίδα, για μια εφαρμογή ISAPI, για μια εφαρμογή CGI, ή για ένα Internet σύνδεσμο βάσης δεδομένων (Internet Database Connector).

- Στατική HTML σελίδα

Η υπηρεσία WWW ψάχνει για ένα όνομα αρχείου ή ένα πεδίο μονοπατιού.

- Εφαρμογή ISAPI

Η υπηρεσία WWW ψάχνει για ένα όνομα αρχείου ή ένα πεδίο μονοπατιού με την κατάληξη DLL.

- Εφαρμογή CGI

Η υπηρεσία WWW ψάχνει για ένα όνομα αρχείου ή ένα πεδίο μονοπατιού με μια κατάληξη που έχει συνδεθεί με την εφαρμογή.

- Internet Database Connector

Η υπηρεσία WWW ψάχνει για ένα όνομα αρχείου ή ένα πεδίο μονοπατιού με την κατάληξη IDC.

Το Hypertext Transmission Protocol είναι ένα πρωτόκολλο χωρίς κατάσταση που σχεδιάστηκε για να επεξεργάζεται μια μοναδική συναλλαγή κατά τη διάρκεια τη σύνδεσης με έναν server. Βρίσκεται στα TCP και IP πρωτόκολλα.

Τα τέσσερα βήματα κατά τη διάρκεια μιας απλής συναλλαγής HTTP είναι:

- Σύνδεση

Ο client HTTP εγκαθιστά μια σύνδεση socket στον IIS

- Αίτηση

Ο client HTTP στέλνει μια αίτηση στην υπηρεσία WWW του IIS. Η αίτηση περιέχει τον τύπο της υπηρεσίας HTTP και άλλη πληροφορία

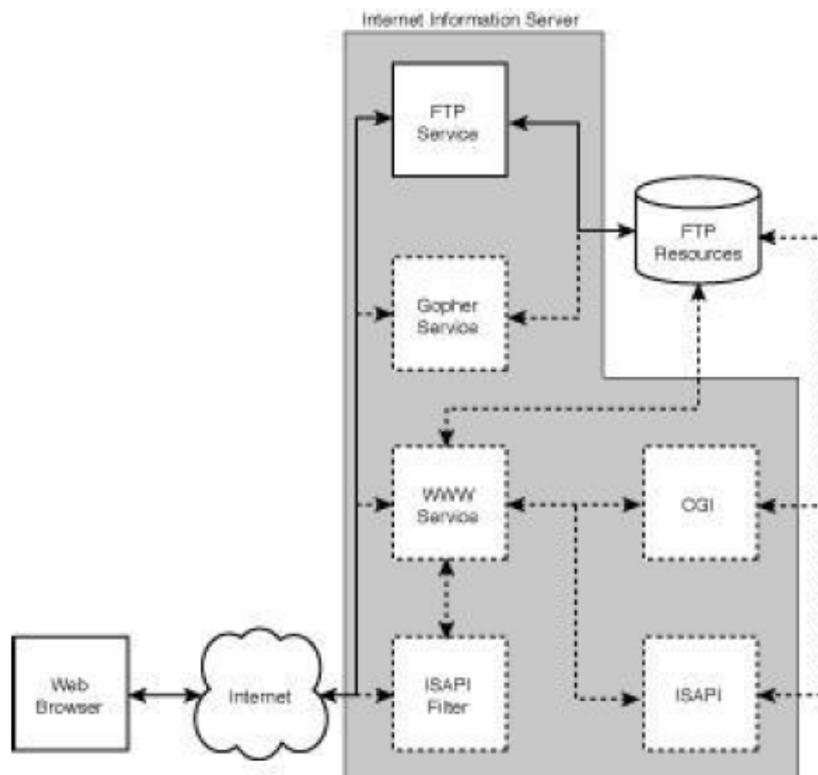
- Απόκριση

Η υπηρεσία WWW στέλνει μια απόκριση πίσω στον HTTP client. Η απόκριση περιέχει την κατάσταση της συναλλαγής και τα δεδομένα που ζητήθηκαν

- Αποσύνδεση

Η υπηρεσία WWW σηματοδοτεί τη λήξη της συναλλαγής κλείνοντας τη σύνδεση socket

Ο Web browser δημιουργεί μια σύνδεση με την υπηρεσία WWW για κάθε αρχείο που είναι μέρος του εγγράφου HTML. Δημιουργεί μια σύνδεση πρώτα για το κείμενο HTML. Ο IIS αποσυνδέεται αφού το στείλει. Ο Web browser αναλύει κάθε λέξη του αρχείου κειμένου ψάχνοντας για ονόματα αρχείων γραφικών εικόνων. Έπειτα δημιουργεί μια σύνδεση με τον IIS και ζητά ένα μοναδικό αρχείο γραφικής εικόνας. Ο IIS αποσυνδέεται αφού το στείλει. Αυτή η διεργασία επαναλαμβάνεται για κάθε αρχείο εικόνας.



Εικόνα 7:Υπηρεσία FTP,υπηρεσία Gopher, υπηρεσία WWW

4.2 Active Server Pages (ASP)

α) ASP.NET

Η ASP.NET (ή αλλιώς ASP+) είναι μια web πλατφόρμα η οποία παρέχει όλες τις λειτουργίες που απαιτούνται για τη δημιουργία δυναμικών διαδικτυακών εφαρμογών. Η πλατφόρμα αυτή είναι βασισμένη στο .NET Framework, επομένως όλα τα χαρακτηριστικά αυτού είναι διαθέσιμα στις ASP.NET εφαρμογές. Οι εφαρμογές αυτές μπορούν να προγραμματιστούν με οποιαδήποτε από τις υποστηριζόμενες, από το CLR (Common Language Runtime) γλώσσες με πιο διαδεδομένες την Visual Basic και τη C#.

β) Διαφορές ASP-ASP.NET

Η ASP.NET ήρθε να εγκαταστήσει την κλασσική ASP τεχνολογία η οποία ήταν η πρώτη προσπάθεια της Microsoft στην ανάπτυξη μιας πλατφόρμας που αφορά το διαδικτυακό προγραμματισμό. Μερικά από τα νέα χαρακτηριστικά της ASP.NET σε σχέση με την κλασσική ASP είναι τα εξής:

- Καθαρός διαχωρισμός μεταξύ εμφάνισης και κώδικα. Με την κλασσική ASP τεχνολογία ο προγραμματισμός του server γινόταν με αταξία μέσα στον HTML κώδικα με αποτέλεσμα αργότερα να γίνονται πολύ δύσκολα οι αλλαγές στη σελίδα.
- Η ASP.NET καθόρισε ένα μοντέλο ανάπτυξης εφαρμογών το οποίο είναι πολύ πιο κοντά σε αυτό της ανάπτυξης desktop εφαρμογών. Ως αποτέλεσμα, έγινε πολύ εύκολη η εναλλαγή μεταξύ desktop και web εφαρμογών για έναν προγραμματιστή.
- Ένα νέο περιβάλλον ανάπτυξης που ονομάστηκε Visual Studio ήρθε για να βοηθήσει τους προγραμματιστές να δημιουργήσουν και να προγραμματίσουν διαδικτυακές εφαρμογές με τη βοήθεια ενός οπτικού περιβάλλοντος προγραμματισμού.
- Με χρήση της ASP.NET τεχνολογίας δώθηκε στον προγραμματιστή η δυνατότητα να χρησιμοποιήσει ανάμεσα από έναν αριθμό αντικειμενοστραφών

γλωσσών προγραμματισμού με πιο διαδεδομένες τη Visual Basic .NET και τη C#

- Πρόσβαση σε ολόκληρο το .NET Framework, γεγονός που σημαίνει ότι για πρώτη φορά οι προγραμματιστές διαδικτυακών εφαρμογών απέκτησαν ένα μοναδικό και εύκολο τρόπο ώστε να μπορούν να έχουν πρόσβαση σε σημαντικά χαρακτηριστικά της πλατφόρμας για να εργάζονται με βάσεις δεδομένων, αρχεία, e-mails, εργαλεία δικτύων και άλλα πολλά.

Παρά το γεγονός ότι προστέθηκαν πολλά νέα χαρακτηριστικά στην ASP.NET, αυξήθηκε κατά πολύ η πολυπλοκότητά της γεγονός που κάνει τη διαδικασία εκμάθησης της τεχνολογίας πιο δύσκολη για άπειρους ή νέους προγραμματιστές.

4.3 Τρόπος Λειτουργίας ASP.NET αρχείων

Το αρχείο που περιέχει μια ASP.NET σελίδα έχει την κατάληξη .aspx. Αυτά τα αρχεία για να “τρέξουν” απαιτούν την ύπαρξη ενός ειδικού λογισμικού στον server ο οποίος ονομάζεται web-server. Ο web-server που χρησιμοποιείται για να εκτελεστούν τα aspx αρχεία σε περιβάλλον Windows είναι ο Internet Information Server (IIS). Στον web-server υπάρχει ένα μικρό κομμάτι λογισμικού το οποίο αναλαμβάνει να εκτελέσει τον κώδικα σε asp.net.

Ένα aspx αρχείο μπορεί να περιέχει κώδικα σε HTML, CSS, Javascript, ASP αλλά και κώδικα στην γλώσσα .NET την οποία έχουμε επιλέξει για να προγραμματίσουμε την εφαρμογή στο κομμάτι που αφορά τον server. Γενικά μπορούμε να χωρίσουμε το αρχείο αυτό σε 2 τμήματα. Το πρώτο είναι αυτό το οποίο περιέχει τον κώδικα τον οποίο θα εκτελεστεί τοπικά στον περιηγητή. Το δεύτερο μέρος είναι ο κώδικας ο οποίος θα εκτελεστεί στον server.

Συνοπτικά η λειτουργία της ASP.NET τεχνολογίας είναι η εξής:

- Όταν ο περιηγητής ζητήσει ένα αρχείο HTML, ο server στέλνει το αρχείο αυτό.
- Όταν ο περιηγητής ζητήσει ένα ASP.NET αρχείο, ο IIS περνά την αίτηση στη μηχανή του server που επεξεργάζεται το ASP.NET κομμάτι.

- Αυτή η ASP.NET μηχανή διαβάζει το αρχείο, γραμμή προς γραμμή, και εκτελεί τα scripts που περιέχονται μέσα στο αρχείο.
- Τέλος, το ASP.NET αρχείο επιστρέφεται στον περιηγητή ως HTML αρχείο.

Επιδόσεις

Η ASP.NET στοχεύει σε καλύτερη επίδοση σε σχέση με τις άλλες scripting γλώσσες προγραμματισμού με το να μεταγλωττίζει τον κώδικα που εκτελείται στον server σε ένα ή περισσότερα dll αρχεία τα οποία αποθηκεύονται στον web-server. Η διαδικασία της μετάφρασης του serverside κώδικα γίνεται αυτόματα την πρώτη φορά που ζητείται μια σελίδα. Αυτό το χαρακτηριστικό παρέχει την ευκολία της ανάπτυξης εφαρμογών με scripting γλώσσες προγραμματισμού σε συνδυασμό με την αυξημένη επίδοση που παρέχει μια μεταγλωττισμένη βιβλιοθήκη.

Παρ' όλα αυτά θα υπάρξει μια μικρή καθυστέρηση εξαιτίας της μεταγλώττισης του κώδικα κατά την πρώτη φορά που θα ζητηθεί η σελίδα από τον web-server. Βέβαια αυτό θα προκύψει μόνο μια φορά, εκτός και εάν η σελίδα ανανεωθεί από τον προγραμματιστή οπότε θα χρειαστεί να ξαναγίνει η διαδικασία. Μια λύση στο πρόβλημα αυτό είναι ο προγραμματιστής να κάνει την μεταγλώττιση του κώδικα κατά τη διάρκεια της ανάπτυξης της εφαρμογής (η διαδικασία αυτή γίνεται στο Visual Studio) και στη συνέχεια να αποθηκεύσει στον εξυπηρετητή το .dll αρχείο που προέκυψε.

4.4 Η γλώσσα προγραμματισμού C#

Η C# είναι μια αντικειμενοστρεφής γλώσσα προγραμματισμού η οποία δημιουργήθηκε από τη Microsoft για το .NET Framework. Στη συνέχεια αναγνωρίστηκε σαν πρότυπο από τους οργανισμούς Ecma και ISO. Η C# ενθαρρύνει την αντικειμενοστρέφεια αλλά μπορεί να χρησιμοποιηθεί και για συναρτησιακό ή δομημένο προγραμματισμό. Στόχος της Microsoft ήταν να φτιάξει μια γλώσσα απλή, μοντέρνα, γενικού-σκοπού και αντικειμενοστρεφής.

α) Λόγοι Δημιουργίας

Καθώς η Microsoft εδώ και μια δεκαετία προσθέτει συνεχώς νέα χαρακτηριστικά στο λειτουργικό της, και στις βιβλιοθήκες των Windows, οι γλώσσες Visual Basic και C++ επεκτείνονται και αυτές σε μια παράλληλη τροχιά. Αυτό είχε ως αποτέλεσμα και οι δύο γλώσσες να καταλήξουν με πανίσχυρα χαρακτηριστικά αλλά και με προβλήματα που προέκυψαν από τον τρόπο με τον οποία εξελίχθηκαν. Στην περίπτωση της Visual Basic, η δύναμη της γλώσσας αντλούνταν από την απλότητα της χρήσης της. Σε βάρος της απλότητας, πολλά προγραμματιστικά προβλήματα που είχαν σχέση με το Windows API και το COM component δεν ήταν εύκολο να λυθούν. Αυτό συνέβαινε γιατί η Visual Basic δεν ήταν ποτέ, πραγματικά αντικειμενοστρεφής γλώσσα με αποτέλεσμα όσο μεγάλωνε το μέγεθος της εφαρμογής να μειώνεται η οργάνωση του κώδικα. Παράλληλα γινόταν πολύ δύσκολη η διατήρηση της εφαρμογής. Σε συνδυασμό με αυτά, η γλώσσα αυτή λόγω της καταγωγής της (δεν προοριζόταν για ανάπτυξη μεγάλων εμπορικών εφαρμογών, αλλά κυρίως για εκμάθηση) διατήρησε κάποια χαρακτηριστικά της τα οποία όμως δεν την έκαναν μια ολοκληρωμένη αντικειμενοστρεφή γλώσσα προγραμματισμού.

Από την άλλη πλευρά, η C++ έχει τις ρίζες της στην Ansi C++. Βέβαια δεν είναι ακριβώς ίδιες, καθώς η Microsoft έφτιαξε πρώτα τον μεταγλωττιστή για τη C++ και στη συνέχεια έγινε επίσημο το Ansi πρότυπο αλλά μοιάζουν σε πολλά στοιχεία. Δυστυχώς όμως αυτό οδήγησε σε δύο προβλήματα. Αρχικά, η Ansi C++ έχει τις ρίζες της σε παλιές τεχνολογίες, και αυτό υποδηλώνει έλλειψη υποστήριξης για ένα σύνολο νέων χαρακτηριστικών όπως είναι τα unicode strings και η

παραγωγή XML Documentation. Επίσης κάποια χαρακτηριστικά της γλώσσας είχαν σχεδιαστεί με βάση παλαιότερων μεταγλωττιστών. Κατά δεύτερον, η Microsoft προσπάθησε να εξελίξει την C++ σε μια γλώσσα προγραμματισμού η οποία θα χρησιμοποιούνταν για την ανάπτυξη, υψηλών προδιαγραφών και υψηλής απόδοσης, λογισμικού για την πλατφόρμα των Windows. Προκειμένου να γίνει αυτό η Microsoft αναγκάστηκε να προσθέσει ένα σύνολο βιβλιοθηκών και λέξεων κλειδίων με αποτέλεσμα η γλώσσα αυτή σε περιβάλλον Windows να είναι ένα «χάος». Ένα χαρακτηριστικό παράδειγμα αυτού του γεγονότος είναι ο μεγάλος αριθμός τύπων που υπάρχουν για τα αλφαριθμητικά (char, LPCTSTR, string, CString, wchar_t, OLECHAR κτλη).

Τέλος, με τον ερχομό του .NET Framework η Microsoft αναγκάστηκε να προσθέσει και άλλα χαρακτηριστικά στις δύο προαναφερόμενες γλώσσες. Συγκεκριμένα στην C++ προστέθηκαν νέες λέξεις κλειδιά και η Visual Basic άλλαξε ριζικά, προκύπτοντας η γλώσσα προγραμματισμού Visual Basic .NET.

Στη Visual Basic .NET διατηρήθηκε η σύνταξη της απλής Visual Basic αλλά ο σχεδιασμός του διέφερε εντελώς. Όλοι αυτοί οι λόγοι, συντέλεσαν στη δημιουργία μιας νέας γλώσσας από την Microsoft η οποία θα ήταν προορισμένη αποκλειστικά και μόνο για το .NET Framework και έτσι δημιουργήθηκε η C#.

4.5 Γενικά Χαρακτηριστικά C#

Η C# είναι μια γλώσσα απλή, μοντέρνα, αντικειμενοστρεφής και έχει στοιχεία από τις C, C++ και JAVA γλώσσες προγραμματισμού. Συντακτικά η C# μοιάζει πολύ στη C++ και στη JAVA καθώς πολλές λέξεις-κλειδιά είναι ίδιες. Επίσης μοιράζεται με αυτές, α) τη δομή των μπλοκ τα οποία ορίζονται με τα σύμβολα "{ " και "}" αλλά και β) την οριοθέτηση των εντολών με το σύμβολο ";" . Μπορεί με την πρώτη ματιά ενός κομματιού κώδικα σε C# να πούμε ότι είναι εμφανής η ομοιότητα με τις γλώσσες C++ και JAVA αλλά ή αλήθεια είναι ότι η C# είναι πιο εύκολη στην εκμάθηση από την πρώτη και παρόμοιας δυσκολίας με την δεύτερη. Ο σχεδιασμός της είναι συνυφασμένος με τα μοντέρνα προγραμματιστικά πρότυπα και παράλληλα έχει υιοθετήσει την απλότητα χρήσης της Visual

Basic αλλά και την υψηλή απόδοση της C++ σε θέματα χαμηλού επιπέδου διαχείρισης της μνήμης, σε περίπτωση που χρειαστεί.

Τα κύρια χαρακτηριστικά της C# είναι τα εξής:

- Πλήρης υποστήριξη για κλάσεις και αντικειμενοστραφή προγραμματισμό, η οποία περιέχει κληρονομικότητα διεπαφής και υλοποίησης, εικονικές συναρτήσεις και υπερφόρτωση τελεστών.
- Ένα συνεπές και καθορισμένο με σαφήνεια σύνολο βασικών τύπων μεταβλητών.
- Δεν υπάρχουν καθολικές μεταβλητές και μέθοδοι. Όλες οι μέθοδοι πρέπει να δηλώνονται μέσα στις κλάσεις. Οι δημόσιες (public) κλάσεις μπορούν να αντικαταστήσουν τις καθολικές μεταβλητές και τις καθολικές συναρτήσεις.
- Ενσωματωμένη δυνατότητα για αυτόματη παραγωγή XML documentation.
- Αυτόματο καθάρισμα της δυναμικά δεσμευμένης μνήμης.
- Δυνατότητα για μαρκάρισμα κλάσεων ή μεθόδων με κάποια συγκεκριμένα χαρακτηριστικά. Αυτό μπορεί να είναι χρήσιμο για την τεκμηρίωση της εφαρμογής αλλά και στην μεταγλώττιση της εφαρμογής. (για παράδειγμα μπορούμε να μαρκάρουμε κάποιες μεθόδους ώστε να μεταγλωττίζονται μόνο όταν η εφαρμογή είναι σε λειτουργία αποσφαλμάτωσης (debug mode)).
- Πλήρης πρόσβαση στη βασική βιβλιοθήκη του .NET Framework αλλά και εύκολη πρόσβαση στο Windows API (αν χρειαστεί).
- Προσπελαση της μνήμης με δείκτες ή απ' ευθείας (αν χρειαστεί). Γενικά η γλώσσα έχει σχεδιαστεί με τέτοιο τρόπο ώστε να μπορεί να λειτουργεί χωρίς αυτά τα χαρακτηριστικά στις περισσότερες των περιπτώσεων.
- Υποστήριξη για ιδιότητες (properties) και γεγονότα (events) στο στυλ της Visual Basic.
- Αλλάζοντας τις επιλογές του μεταγλωττιστή, δίνεται η δυνατότητα να μεταγλωττίσουμε το προγράμμα μας σε ένα εκτελέσιμο αρχείο ή μια βιβλιοθήκη η

οποία θα περιέχει .NET συστατικά τα οποία θα μπορούν να καλούνται από άλλα τμήματα κώδικα σαν ActiveX controls. (COM components)

- Η C# μπορεί να χρησιμοποιηθεί για συγγραφή α) δυναμικών ιστοσελίδων ASP.NET αλλά και β) XML Web Services

- Η C# υποστηρίζει έναν αυστηρό τύπο δεδομένων για boolean μεταβλητές, τον bool. Δηλώσεις οι οποίες έχουν κάποιες καταστάσεις όπως οι βρόγχοι while και if, απαιτούν μια έκφραση της οποίας ο προκύπτον τύπος θα είναι bool. Με απλά λόγια η έκφραση να μπορεί να είναι αληθής ή ψευδής. Ενώ η C++, υποστηρίζει boolean τύπους δεδομένων, αυτοί μπορούν εύκολα να μετατραπούν από/σε ακέραιους και εκφράσεις όπως οι if(a) απαιτούν απλά το a να μετατραπεί σε boolean τύπο, επιτρέποντας έτσι το a να είναι πχ. ακέραιος ή δείκτης. Η C# δεν επιτρέπει έναν ακέραιο να δείχνει αλήθεια ή ψεύδος οπότε αναγκάζει τους προγραμματιστές να χρησιμοποιούν εκφράσεις οι οποίες επιστρέφουν ακριβώς boolean τύπους δεδομένων. Ως αποτέλεσμα αυξάνεται η αποδοτικότητα και μειώνεται η πιθανότητα άσκοπων αλλά κοινών λαθών στον κώδικα.

Ένα τέτοιο πρόβλημα στη C++ περιγράφεται με το εξής παράδειγμα:

Όταν έχουμε μια έκφραση if(a=b) και χρησιμοποιούμε = αντί για ==.

- Σε αντίθεση με την δομή try...catch που χρησιμοποιούμε για χειρισμό λαθών, υπάρχει η δυνατότητα χρήσης της δομής try...finally η οποία εγγυάται εκτέλεση του κώδικα που βρίσκεται μέσα στο finally μπλοκ.

- Παρέχεται πλήρης δυνατότητα για reflection programming.

- Η C# (στην έκδοση 4.0) έχει 77 δεσμευμένες λέξεις. Τα περισσότερα από τα παραπάνω χαρακτηριστικά τα έχουν και οι γλώσσες προγραμματισμού Visual Basic .NET και C++. Το γεγονός όμως το ότι η C# σχεδιάστηκε για χρήση των δυνατοτήτων του .NET Framework την κάνει πιο πλήρη, και προσφέρεται στους προγραμματιστές με πιο απλή σύνταξη σε σχέση με τις άλλες δύο.

Μειονεκτήματα

Πέρα από τη δύναμη C#, υπάρχουν και κάποιοι περιορισμοί οι οποίοι την κάνουν ακατάλληλη γλώσσα προγραμματισμού για τη συγγραφή κάποιων εφαρμογών.

Το κύριο μειονέκτημά της είναι ότι δεν έχει σχεδιαστεί για τη σύγγραφη προγραμμάτων τα οποία έχουν σαν πρώτη προτεραιότητα τις ακραίες επιδόσεις. Αν λοιπόν ενδιαφέρει τον προγραμματιστή αν ένα βρογχος θα πάρει 1.050 κύκλους μηχανής αντί για 1.000, και αν κάθε δέκατο του δευτερολέπτου είναι σημαντικό για την ανάγκη που εξυπηρετεί μια εφαρμογή τότε η καλύτερη λύση μεταξύ των low-level γλωσσών παραμένει η C++. Παρ' όλα αυτά το σύνολο των εφαρμογών που ανήκουν σε αυτή την κατηγορία είναι πολύ μικρό.

Συνοψίζοντας η C# είναι μια μοντέρνα αντικειμενοστρεφής γλώσσα που δίνει την δυνατότητα στους προγραμματιστές να φτιάχνουν γρήγορα και εύκολα λογισμικό για την πλατφόρμα .NET. Η παρεχόμενη πλατφόρμα επιτρέπει την κατασκευή XML Web Services τα οποία είναι διαθέσιμα μέσω Internet από οποιαδήποτε εφαρμογή που τρέχει σε οποιαδήποτε πλατφόρμα. Η γλώσσα ενισχύει την παραγωγικότητα των υπευθύνων για την ανάπτυξη λογισμικού ενώ παράλληλα βοηθά τον προγραμματιστή στην μείωση των λαθών κατά τη φάση της υλοποίησης. Με αυτό το τρόπο αποτρέπει τις αυξανόμενες δαπάνες της ανάπτυξης λογισμικού. Τέλος, η C# βοηθά τη μετάβαση των C/C++ προγραμματιστών στην ανάπτυξη εφαρμογών για το παγκόσμιο ιστό ενώ παράλληλα διατηρεί τη δύναμη και την ευελιξία στην ανάπτυξη λογισμικού.

5.0 Microsoft SQL Server

Η Microsoft SQL Server είναι μια σχεσιακή βάση δεδομένων η οποία έχει αναπτυχθεί από τη Microsoft. Οι δύο πρωταρχικές γλώσσες ερωτημάτων που χρησιμοποιεί είναι οι T-SQL και ANSI-SQL.

Βάση της Microsoft SQL Server αποτελεί η Sybase SQL Server η οποία ήταν η λύση της Microsoft στην αγορά των προϊόντων βάσεων δεδομένων. Με αυτό το τρόπο η Microsoft έγινε ανταγωνίστρια εταιρία με τις Oracle, IBM και αργότερα με τη Sybase. Οι εταιρίες Sybase, Microsoft και Ashton-Tate συνεργάστηκαν προκειμένου να δημιουργήσουν και να προωθήσουν την πρώτη τους σχεσιακή βάση δεδομένων με όνομα SQL Server 1.0 για το λειτουργικό σύστημα OS/2 το 1989. Η έκδοση αυτή είναι ίδια με την SQL Server 3.0 που τρέχει σε λειτουργικά συστήματα UNIX και VMS. Η έκδοση 4.2 της SQL Server παρουσιάστηκε γύρω στο 1992 (ήταν διαθέσιμη για το λειτουργικό IBM OS/2 έκδοση 1.3). Στη συνέχεια η έκδοση 4.2 που βασιζόταν σε Windows παρουσιάστηκε μαζί με τη νέα έκδοση του λειτουργικού Windows NT 3.1. Η Microsoft SQL Server 6.0 είναι η πρώτη έκδοση της βάσης που σχεδιάστηκε αποκλειστικά για τα Windows NT και δεν ήταν σε συνεργασία με την εταιρία Sybase.

Μετά την παρουσίαση των Windows NT, η Sybase και η Microsoft χώρισαν τους δρόμους τους και η κάθε μια ανέπτυξε το δικό της προϊόν, και σχεδίασε με διαφορετικό τρόπο την προώθηση των βάσεων δεδομένων τους. Η Microsoft διαπραγματεύτηκε τα αποκλειστικά δικαιώματα της SQL Server για το λειτουργικό σύστημα Windows. Στη συνέχεια η Sybase μετέτρεψε το Server Enterprise για να αποφύγει περιπλοκές με τη βάση δεδομένων της Microsoft που διατήρησε το ίδιο όνομα. Έως το 1994, το προϊόν της Microsoft περιείχε τρία σημειώματα πνευματικών δικαιωμάτων της Sybase, γεγονός που δήλωνε την προέλευση του συγκεκριμένου προϊόντος. Αφού χωρίστηκαν οι δρόμοι των δύο εταιριών, πολλές ανεξάρτητες ανανεώσεις του προϊόντος πραγματοποιήθηκαν. Είναι γεγονός ότι η SQL Server 7.0 αποτελεί μια επανεγγραφή του κώδικα που είχε αναπτύξει η Sybase. Η πρώτη επιτυχία της Microsoft είναι η SQL Server 2000 για την οποία χρησιμοποιήθηκε η αρχιτεκτονική IA-64. Μέσα σε δέκα χρόνια από την εμφά-

νιση της έκδοσης 2000 της Microsoft SQL Server πραγματοποιήθηκαν ένα σύνολο από εξελίξεις και αναβαθμίσεις στο λογισμικό αυτό. Συγκεκριμένα δώθηκε μεγάλη σημασία στην απόδοση, στο εργαλείο IDE του πελάτη και σε ένα σύνολο άλλων βοηθητικών στοιχείων που έρχεται μαζί με το πακέτο SQL Server 2005. Σε αυτά τα στοιχεία περιέχονται τα εξής: ETL (Extract, transform, and load) εργαλεία για εξαγωγή δεδομένων, αλλαγή αυτών ώστε να ταιριάζουν στις ανάγκες του χρήστη, και φόρτωση αυτών στη βάση. Παράλληλα μέσα στο πακέτο λογισμικού του SQL Server 2005 περιέχεται ένας Server αναφορών, ένας OLAP (OnLine Analytical Processing) και data mining (εξόρυξη δεδομένων) Server, ένα σύνολο τεχνολογιών επικοινωνίας μεταξύ των διαφορών τμημάτων του λογισμικού. Συγκεκριμένα παραδείγματα τέτοιων υπο-προγραμμάτων είναι οι υπηρεσίες ανακοινώσεων και οι μεσίτες υπηρεσιών (Service Brokers).

5.1 Χαρακτηριστικά

α) Αρχιτεκτονική

Το layer protocol υλοποιεί το εξωτερικό επίπεδο του SQL Server. Όλες οι λειτουργίες οι οποίες μπορούν να πραγματοποιηθούν στον SQL Server επικοινωνούν μεταξύ τους μέσω ενός τύπου που έχει οριστεί από τη Microsoft και ονομάζεται Tabular Data Stream (TDS). Το TDS είναι μια εφαρμογή του layer protocol, το οποίο χρησιμοποιείται για να μεταφέρονται τα δεδομένα μεταξύ του εξυπηρετητή στον οποίο βρίσκεται η βάση δεδομένων, και του πελάτη. Αρχικά σχεδιάστηκε και αναπτύχθηκε από την εταιρία Sybase για τη σχεσιακή βάση δεδομένων Sybase SQL Server το 1984, και αργότερα από τη Microsoft για το αντίστοιχο προϊόν της με όνομα Microsoft SQL Server. Τα TDS πακέτα χρησιμοποιούνται και σε άλλες φυσικές μεταφορές όπως στο πρωτόκολλο TCP/IP, σε τεχνολογίες Named Pipes και διαμοιρασμού μνήμης. Ως αποτέλεσμα, η πρόσβαση στον SQL Server είναι εξίσου εύκολη πάνω από αυτά τα πρωτόκολλα. Τέλος, το SQL Server API είναι διαθέσιμο και μέσω υπηρεσιών διαδικτύου (Web Services).

5.2 Αποθήκευση Δεδομένων

Η κύρια μονάδα αποθήκευσης δεδομένων είναι η βάση δεδομένων. Αυτή αποτελείται από μια συλλογή πινάκων με γραμμές στις οποίες περιέχονται πληροφορίες συγκεκριμένων τύπων δεδομένων. Η SQL Server υποστηρίζει διάφορους τύπους δεδομένων. Σε αυτούς περιέχονται οι πρωταρχικοί τύποι όπως οι ακέραιοι, δεκαδικοί, αριθμοί κινητής υποδιαστολής, χαρακτήρες (και ακολουθίες χαρακτήρων), Varchar (σειρές χαρακτήρων μεταβλητού μήκους), δυαδικούς, κείμενο κ.α.

Η Microsoft SQL Server επιτρέπει καθορισμένους από το χρήστη τύπους δεδομένων (UDTs), δηλαδή τύπους οι οποίοι βασίζονται στους πρωταρχικούς τύπους αλλά μπορούν να τροποποιηθούν. Επίσης δημιουργεί στατιστικά τα οποία είναι διαθέσιμα σαν εικονικοί πίνακες (DMVs). Εκτός από πίνακες, μια βάση δεδομένων μπορεί να περιέχει και άλλα αντικείμενα όπως είναι οι προβολές (views), οι αποθηκευμένες συναρτήσεις, τα ευρετήρια (indexes), οι περιορισμοί (constraints), και ένα ιστορικό συναλλαγών του πελάτη με τη βάση. Μια SQL Server βάση δεδομένων μπορεί να περιέχει 231 αντικείμενα, και μπορεί να δημιουργήσει πολλαπλά αρχεία με μέγιστο μέγεθος τα 220 TB. Τα δεδομένα στη βάση δεδομένων αποθηκεύονται σε αρχεία με κατάληξη .mdf. Τα δευτερεύοντα αρχεία δεδομένων χαρακτηρίζονται από την κατάληξη .ndf και χρησιμοποιούνται για την αποθήκευση προαιρετικών μετά-πληροφοριών. Τα αρχεία αναφορών (log-files) έχουν την κατάληξη .ldf στο τέλος.

Ο χώρος αποθήκευσης που δεσμεύεται για μια βάση δεδομένων χωρίζεται σε αριθμημένες σελίδες οι οποίες έχουν 8KB μέγεθος η κάθε μία. Η σελίδα είναι η βασική μονάδα εισόδου-εξόδου (I/O) για τις λειτουργίες της SQL Server. Καθε σελίδα αποτελείται από μια επικεφαλίδα 96-bytes η οποία αποθηκεύει μετα-πληροφορίες γύρω από τη σελίδα. Τέτοιες πληροφορίες είναι ο αριθμός της σελίδας, ο τύπος της σελίδας, ο ελεύθερος χώρος που απομένει σε αυτή και το ID του αντικειμένου στο οποίο ανήκει. Ο τύπος της σελίδας ορίζει τα δεδομένα που

περιέχονται σ' αυτή, τα δεδομένα που βρίσκονται στη βάση δεδομένων, το ευρετήριο και τη δεσμευμένη περιοχή η οποία περιέχει πληροφορίες του τρόπου με τον οποίο οι σελίδες είναι δεσμευμένες σε πίνακες και ευρετήρια. Επίσης ορίζει τις αλλαγές οι οποίες γίνονται σε άλλες σελίδες από τη στιγμή του τελευταίου backup ή της τελευταίας σύνδεσης του χρήστη. Επιπρόσθετα, μπορεί να περιέχει μεγάλους τύπους δεδομένων όπως εικόνες ή κείμενο. Ενώ η σελίδα είναι η βασική μονάδα για να εκτελεστεί μια λειτουργία εισόδου-εξόδου, ο χώρος στην

πραγματικότητα διαχειρίζεται από μια επέκταση που αποτελείται από οκτώ σελίδες. Ένα αντικείμενο σε μια βάση δεδομένων μπορεί να επεκταθεί σε οκτώ σελίδες (uniform extent) ή να μοιραστεί τις σελίδες αυτές με επτά ή περισσότερα, διαφορετικά αντικείμενα (mixed extent). Μια γραμμή σε μια βάση δεδομένων δεν μπορεί να επεκταθεί σε περισσότερες από μία σελίδες, έτσι έχει περιορισμένο μέγεθος της τάξης των 8KB. Παρ' όλα αυτά εάν το μέγεθος ξεπεράσει τη χωρητικότητα αυτή και η γραμμή του πίνακα περιέχει Varchar ή Varbinary πληροφορία, τότε τα δεδομένα της γραμμής αυτής επεκτείνονται σε μια καινούρια σελίδα (ή σε μια ακολουθία σελίδων που καλείται Allocation Unit). Σε αυτή την περίπτωση, ένας δείκτης υποδεικνύει που βρίσκεται η πληροφορία.

Για φυσική αποθήκευση ενός πίνακα, οι γραμμές του διαχωρίζονται σε ένα σύνολο από κομμάτια τα οποία αριθμούνται από το ένα έως το n. Το μέγεθος του κάθε κομματιού καθορίζεται από τον χρήστη. Το προκαθορισμένο μέγεθος είναι όλες οι γραμμές να αποτελούν ένα κομμάτι. Ο διαχωρισμός των γραμμών σε κομμάτια γίνεται προκειμένου να επεκταθεί η βάση δεδομένων για χρήση σε cluster δηλαδή διαφορετικά κομμάτια να βρίσκονται σε διαφορετικούς servers. Οι γραμμές σε κάθε κομμάτι αποθηκεύονται σε δομή B-δέντρου ή σωρού. Εάν ο πίνακας έχει ένα ευρετήριο για γρήγορη ανάκτηση γραμμών, οι γραμμές αποθηκεύονται σύμφωνα με τις τιμές του ευρετηρίου το οποίο ορίζεται από το B-δέντρο. Τα δεδομένα βρίσκονται στα φύλλα του δέντρου, ενώ οι κόμβοι του δέντρου περιέχουν τις τιμές του ευρετηρίου γνωστοποιώντας έτσι σε ποιο φύλλο βρίσκονται τα δεδομένα που μπορούν να προσπελαστούν από αυτό τον κόμβο. Εάν το ευρετήριο δεν είναι clustered, οι γραμμές δεν είναι ταξινομημένες

σύμφωνα με το ευρετήριο. Μια προβολή ευρετηρίου (indexed view) έχει την ίδια δομή αποθήκευσης όπως ένας πίνακας ευρετηρίου. Ένας πίνακας δίχως ευρετήριο αποθηκεύεται σε μια δομή σωρού.

5.3 Ενδιάμεση μνήμη (Cache)

Όλες οι σελίδες που αναφέρθηκαν παραπάνω αποθηκεύονται στην ενδιάμεση (buffer) μνήμη RAM η οποία είναι πιο γρήγορα προσπελάσιμη. Ως αποτέλεσμα ελαχιστοποιείται ο χρόνος που απαιτείται για την προσπέλαση και μεταφορά των πληροφοριών προς και από τον σκληρό δίσκο. Οποιαδήποτε σελίδα 8 KB μπορεί να είναι αποθηκευμένη στη μνήμη. Το σύνολο όλων των σελίδων που αποθηκεύονται σε μια χρονική περίοδο καλείται λανθάνουσα μνήμη (μνήμη cache). Ανάλογα με το μέγεθος της μνήμης που είναι διαθέσιμο στον εξυπηρετητή, η SQL Server αποφασίζει τον αριθμό των σελίδων που θα αποθηκευτούν στην λανθάνουσα μνήμη του υπολογιστή. Η διαχείριση της μνήμης αυτής γίνεται από διάφορους έξυπνους αλγόριθμους που έχουν σαν στόχο την βέλτιστη απόδοση. Ο κεντρικός διακομιστής στον οποίο βασίζεται η τρέχουσα εγκατάσταση του SQL Server είναι υπεύθυνος για να εξασφαλίσει ότι οποιαδήποτε αλλαγή στα στοιχεία πληρούν κάποιες προϋποθέσεις. Συγκεκριμένα χρησιμοποιεί τις συναλλαγές (transactions) μεταξύ των υποσυστημάτων για να εξασφαλίσει ότι οποιαδήποτε λειτουργία συμβεί, έχει δύο εν δυνάμει αποτελέσματα τα οποία είναι η αποτυχία ή η επιτυχία. Δηλαδή κάθε διαδικασία δεν πρόκειται να βρεθεί σε ενδιάμεσες καταστάσεις καθώς είτε θα ολοκληρωθεί συνολικά, είτε αποτυγχάνει συνολικά.

Με τη βοήθεια των συναλλαγών αυτών, ένα σύνολο από εντολές, μπορούν να αποτελέσουν μια ομάδα υπό την προϋπόθεση ότι είτε η εκτέλεση όλων των εντολών θα είναι επιτυχής, είτε η εκτέλεση όλων των εντολών θα είναι μη επιτυχής. Ο κεντρικός SQL εξυπηρετητής επιτρέπει την παράλληλη προσπέλαση της βάσης δεδομένων από πολλούς χρήστες. Για αυτό το λόγο θα πρέπει να ελέγχεται συνεχώς η ταυτόχρονη προσπέλαση στα κοινά δεδομένα, προκειμένου να εξασφαλιστεί η ακεραιότητα των στοιχείων. Για παράδειγμα όταν οι πελάτες ενημερώνουν το ίδιο στοιχείο, ή όταν ένας πελάτης προσπαθεί να αναγνώσει ένα στοιχείο το οποίο ενημερώνεται παράλληλα από έναν άλλο πελάτη.

α)Υπηρεσίες

Ο Microsoft SQL Server περιλαμβάνει ένα σύνολο από επιπρόσθετες υπηρεσίες. Αν και αυτές δεν είναι απαραίτητες για τη λειτουργία του συστήματος της βάσης δεδομένων, παρέχουν ένα σύνολο από ευκολίες οι οποίες αποτελούν το υψηλότερο επίπεδο της διαχείρισης του συστήματος της βάσης. Αυτές οι υπηρεσίες, τρέχουν είτε σαν μέρος του συστήματος SQL Server είτε ξεχωριστά σαν υπηρεσίες του λειτουργικού συστήματος που έχει ο server (στην προκειμένη περίπτωση Windows). Το σημαντικότερο χαρακτηριστικό όταν λειτουργούν ξεχωριστά είναι ότι είναι διαθέσιμο στους προγραμματιστές το API για τον έλεγχο τους αλλά και την αλληλεπίδραση μαζί τους. Παρακάτω περιγράφονται οι υπηρεσίες αυτές:

- Ο μεσίτης Υπηρεσιών (Service Broker) είναι ένα υποσύστημα στον 70 SQL Server ο οποίος είναι υπεύθυνος για την αποστολή και λήψη μηνυμάτων μεταξύ των υποσυστημάτων του κεντρικού ολοκληρωμένου συστήματος της βάσης δεδομένων. Επίσης η συγκεκριμένη υπηρεσία μπορεί να διαχειριστεί και θέματα αναμονής μηνυμάτων. Τέλος ο μεσίτης υπηρεσιών έχει ανώτερες δυνατότητες προγραμματισμού στο μέλλον (Scheduling).
- Οι υπηρεσίες συγχρονισμού (Replication) του SQL Server χρησιμοποιείται από το κεντρικό σύστημα προκειμένου να πραγματοποιείται ο συγχρονισμός των βάσεων δεδομένων, είτε ολοκληρωτικά, είτε μόνο ένα υποσύνολο αντικειμένων της βάσης. Ο συγχρονισμός μπορεί να επιτευχθεί με πέντε διαφορετικούς τρόπους: log shipping, mirroring, replication, back-ups και clustering. Αυτό είναι χρήσιμο σε περιπτώσεις που η κεντρική βάση δεδομένων καταστραφεί ολοσχερώς, καθώς δίνεται η δυνατότητα για αυτόματη λειτουργία της δευτερεύουσας συγχρονισμένης βάσης δεδομένων όταν συμβεί κάτι τέτοιο.
- Οι υπηρεσίες ανάλυσης (Analysis Services) του SQL Server κάνουν διαθέσιμες ικανότητες OLAP και περίπλοκης ανάκτησης δεδομένων για τις SQL βάσεις δε-

δομένων που έχουν περίπλοκη ανάλυση. Αυτό μπορεί να πραγματοποιηθεί δίχως τη σύνθεση δύσκολων ερωτημάτων προς τη βάση. Επίσης η μηχανή OLAP υποστηρίζει τα MOLAP, ROLAP, HOLAP εργαλεία ανάλυσης δεδομένων για την ορθή αποθήκευση των πληροφοριών στη βάση. Οι υπηρεσίες ανάλυσης υποστηρίζει XML τεχνολογίες για τα πρωτόκολλα επικοινωνίας που χρησιμοποιούνται. Τέλος, οι αλγόριθμοι που χρησιμοποιούνται στις υπηρεσίες ανάλυσης είναι οι εξής: δέντρα απόφασης, clustering, Bayes, time series analysis, sequence clustering, γραμμική και λογική ανάλυση και τέλος αλγόριθμοι νευρωνικών δικτύων για χρήση στην εξόρυξη δεδομένων.

- Ο SQL Server έχει υπηρεσίες Reporting για τη συγκέντρωση των στοιχείων που παρέχουν οι SQL βάσεις δεδομένων. Ως αποτέλεσμα, 71 υπάρχει η δυνατότητα παραγωγής μιας έκθεσης αναφοράς (Web Report) σχεδόν αυτόματα χωρίς να έχει προγραμματιστεί αυτή η λειτουργία σε HTML, CSS και javascript.

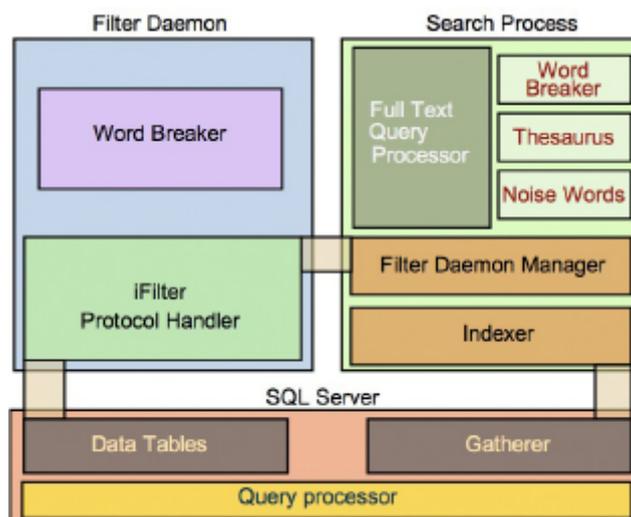
- Οι υπηρεσίες ανακοίνωσης (Notification Services) πρώτη φορά ενσωματώθηκαν στο λογισμικό του SQL Server στην έκδοση 2005. Οι υπηρεσίες ανακοίνωσης δίνουν τη δυνατότητα στους χρήστες να παράγουν e-mails απλά και γρήγορα. Παρ' όλα αυτά το χαρακτηριστικό αυτό από την έκδοση 2008 του SQL Server δεν βρίσκεται πια ενσωματωμένο στο σύστημα του SQL Server.

- Τα Integration Services του SQL Server χρησιμοποιούνται για να ενσωματώσουν τα δεδομένα από διαφορετικές πηγές δεδομένων. Χρησιμοποιούνται για τις ETL ικανότητες του SQL Server για τις ανάγκες αποθήκευσης δεδομένων (data warehousing needs). Οι ETL ικανότητες του SQL Server περιλαμβάνουν την εξαγωγή δεδομένων από εξωτερικές πηγές, την μετατροπή τους για να ταιριάξουν στις ανάγκες της λειτουργίας, και τέλος η αποθήκευσή τους σε ένα κοινό στόχο (συγκεκριμένα σε βάση δεδομένων είτε σε αποθήκες πληροφοριών ή αλλιώς data warehouses). Τα Integration Services περιλαμβάνουν εργαλεία GUI για να κατασκευαστούν τα workflows (σύνολο βημάτων που απαιτούνται) της εξαγωγής δεδομένων. Με αυτό τον τρόπο δίνεται έμφαση στην λειτουργικότητα για προβλήματα εξαγωγής δεδομένων από διαφορετικές πηγές, δημιουργία ερωτήσεων στα δεδομένα, μετατροπή δεδομένων συμπεριλαμβανομέ-

νου της αντιγραφής ή ένωσής τους και στη συνέχεια τη φόρτωση των αλλαγμένων δεδομένων σε άλλες αποθήκες πληροφοριών. Επίσης, υπάρχει η δυνατότητα αποστολής e-mails με λεπτομέρειες για την κατάσταση της κάθε λειτουργίας ανάλογα με τις επιλογές του χρήστη.

- Η υπηρεσία αναζήτησης πλήρους κειμένου (Full Text Search service) του SQL Server είναι μια υπηρεσία αναζήτησης πληροφορίας μέσα σε αδόμητο κείμενο (δηλ. όχι xml, html κτλπ). Η υπηρεσία αυτή μπορεί να αναζητήσει πληροφορία μέσα σε οποιαδήποτε γραμμή της βάσης δεδομένων που έχει δηλωθεί σαν char τύπος. Ενώ αυτή η λειτουργία μπορεί να γίνει με τη λέξη κλειδί LIKE της SQL γλώσσας, χρησιμοποιώντας την υπηρεσία αναζήτησης πλήρους κειμένου η διαδικασία αυτή απλοποιείται. Η υπηρεσία μπορεί να κάνει ταυτοποίηση κειμένου που στηρίζεται σε μια βαθμολογική σκαλά από το μηδέν έως το χίλια. Το κείμενο το οποίο έχει τον υψηλότερο βαθμό σημαίνει ότι είναι η πιο ακριβής ταυτοποίηση στο προς αναζήτηση κείμενο. Η μηχανή της αναζήτησης πλήρους κειμένου αποτελείται από δύο διεργασίες: την διεργασία Filter Daemon (msftefd.exe) και την διεργασία Search (msftesql.exe). Αυτές οι διεργασίες αλληλεπιδρούν με τον SQL Server. Η διεργασία της αναζήτησης περιλαμβάνει ένα μηχανισμό που φτιάχνει το ευρετήριο της διαδικασίας που αναφέρθηκε παραπάνω (indexer) και τον επεξεργαστή ερωτημάτων για το πλήρες, προς αναζήτηση, κείμενο. Ο indexer διαβάζει το κείμενο που βρίσκεται στη βάση δεδομένων. Επίσης υπάρχει η δυνατότητα αναζήτησης σε binary πληροφορία. Με τη χρήση των iFilters είναι δυνατή η εξόρυξη κειμένου από πχ. ένα αρχείο word το οποίο φυλάσσεται στη βάση δεδομένων σαν binary πληροφορία. Αφού πραγματοποιηθεί η εξόρυξη της πληροφορίας, η διεργασία Filter Daemon σπάει το κείμενο σε λέξεις δημιουργώντας έτσι το ευρετήριο. Στη συνέχεια διαγράφονται από αυτό λέξεις άχρηστης σημασίας όπως άρθρα, συνδέσμοι κτλπ. που δεν έχουν σημασία κατά την αναζήτηση πληροφορίας. Με τις σημαντικές λέξεις του ευρετηρίου δημιουργείται ένα ανεστραμμένο ευρετήριο στο οποίο δηλώνεται σε ποια γραμμή της βάσης βρίσκεται η κάθε λέξη. Σε περίπτωση αλλαγής κάποιας γραμμής της βάσης, είναι υπεύθυνος ο Συλλέκτης (Gatherer) ώστε να ενημερώσει το ευρετήριο. Όταν λαμβάνεται ένα ερώτημα για πλήρες κείμενο από τον επεξεργαστή ερωτημάτων του SQL Server, μεταβαίνει στον FTS επεξεργαστή ερωτημάτων που

βρίσκεται στη διεργασία αναζήτησης. Ο FTS επεξεργαστής ερωτημάτων “σπάει” το ερώτημα σε λέξεις, αφαιρώντας τα άχρηστα λήμματα και χρησιμοποιώντας ένα λεξικό όρων για να κατανοήσει διάφορους τύπους μιας λέξης (πχ ρήματα σε διάφορους χρόνους κτλπ.). Στη συνέχεια συγκρίνεται το ερώτημα με τις διάφορες λέξεις που βρίσκονται ήδη σε ανεστραμμένο ευρετήριο και προκύπτει η βαθμός ταυτοποίησης. Τα αποτελέσματα επιστρέφονται στον πελάτη μέσω της διεργασία του SQL Server. Συνοψίζοντας, είναι η υπηρεσία αυτή δίνει τη δυνατότητα αναζήτησης αδόμητης πληροφορίας σε κείμενα Word, pdf με έμφαση στην απόδοση. Παρακάτω φαίνεται η αρχιτεκτονική της υπηρεσίας αναζήτησης πλήρους κειμένου.



Εικόνα 8: Αρχιτεκτονική της υπηρεσίας αναζήτησης πλήρους κειμένου

5.4 Από το Web 1.0 στο Web 2.0

Μπορούμε να κατηγοριοποιήσουμε καθένα από τα Web x.x σαν μια διαφορετική κίνηση σύμφωνα με την χρήση του ίντερνετ. Το Web1.0 είναι η κίνηση που έλαβε χώρα όταν πρωτοξεκίνησε το ίντερνετ, αναφέρεται στην κατάσταση του World Wide Web και συμπεριλαμβάνει τους περισσότερους ιστοχώρους στην περίοδο μεταξύ 1994 και 2004. Είναι σημαντικό να σημειωθεί ότι το «Web1.0» έχει ονομαστεί αναδρομικά και μόνο μετά από την εισαγωγή του όρου «Web2.0». Έτσι, ανάμεσα στο Web1.0 και το Web2.0 υπάρχουν πολύ αόριστα καθορισμένα όρια. Αρχικά, η πρώτη χρήση του διαδικτύου είχε το μέσο της εκτύπωσης και του ηλεκτρονικού ταχυδρομείου. Το Web1.0 ξεκίνησε με τα βιβλία, τις ειδήσεις, τη μουσική και ότι άλλο μπορούσε να διακινηθεί σε ψηφιακή μορφή. Αυτή η μετακίνηση συνεχίζεται ακόμα και πιθανώς δεν θα σταματήσει ποτέ. Το συγκεκριμένο γεγονός συμβαίνει επειδή νέα στοιχεία διαρκώς διατίθενται, οπότε πρέπει να γίνονται απευθείας προσβάσιμα ηλεκτρονικά. Αυτό ακριβώς μας φέρνει στο Web2.0. Πολλοί σκέφτονται ότι αυτό είναι η τρέχουσα μετακίνηση του διαδικτύου, που κατά κάποιο τρόπο θεωρείται σωστό. Αφότου ταχυδρομήθηκαν ηλεκτρονικά όλα τα στοιχεία με τον Web1.0, η ηλεκτρονική κοινότητα άρχισε να ψάχνει τρόπους να μοιραστεί όλα αυτά τα στοιχεία. Η κύρια ερώτηση που οδήγησε αυτήν την μετακίνηση του Web2.0, είναι «πώς μπορώ να πάρω αυτό το στοιχείο και να το μοιραστώ με άλλους ανθρώπους;». Δεδομένης αυτής της ερώτησης, όλοι οι ιστότοποι που υπήρχαν εκείνη τη στιγμή στο Διαδίκτυο προσπάθησαν να απαντήσουν με διαφορετικές προσεγγίσεις κάθε φορά. Παρακάτω παρατίθενται κάποιες διαφορές μεταξύ Web1.0 και Web2.0:

- Το Web1.0 άρχισε ως ένας ηλεκτρονικός τόπος για τις επιχειρήσεις να μεταδίδουν τις πληροφορίες στους χρήστες/πελάτες. Τώρα, αρκετά χρόνια αργότερα, το μέρος αυτό έχει μετατραπεί σε μια θέση όπου και οι άνθρωποι μπορούν να προσφέρουν πληροφορίες.
- Στην αρχή, οι χρήστες διάβαζαν ότι οι άλλοι έγραφαν και ήταν συμβιβασμένοι με αυτή τη γνώση που αποκτούσαν. Αλλά, με την πάροδο του χρόνου, αυτοί

οι ίδιοι χρήστες αποφάσισαν ότι θέλησαν να δημιουργήσουν μέρος των πληροφοριών που διαμοιράζονται.

- Η κίνηση από το Web1.0 προς το Web2.0 έχει σχέση με τους ανθρώπους που ανταλλάσσουν πληροφορίες για να γίνουν μέρος μιας μεγάλης κοινότητας.
- Πλέον, οι χρήστες δεν είναι παθητικοί δέκτες απλά συλλέγοντας την πληροφορία που τους διατίθενται, αλλά παίρνουν μέρος σε blog, σε θέσεις φόρουμ και σχολιάζουν κάθε τι που γνωρίζουν.
- Στις εφαρμογές Web2.0, οι διεπαφές είναι πιο φιλικές, οι πληροφορίες είναι περισσότερες, οι άνθρωποι είναι περισσότερο συνδεδεμένοι και τα διάφορα δεδομένα αναζητούνται ευκολότερα σε σχέση με τις υπηρεσίες Web1.0.

5.4.1 Η μετακίνηση προς το Web 2.0

Το Web2.0 αναφέρεται σε μία υποτιθέμενη ή προτεινόμενη δεύτερη γενιά από web-based (βασισμένες στο διαδίκτυο) κοινότητες και υπηρεσίες που έχουν σκοπό να διευκολύνουν την δημιουργικότητα, τη συνεργασία και την ανταλλαγή ιδεών μεταξύ των χρηστών. Ο όρος εφευρέθηκε από την O'Reilly Media το 2004. Παραδείγματα τέτοιων υπηρεσιών είναι τα ακόλουθα: οι δικτυακοί τόποι κοινωνικής δικτύωσης (social networking sites), τα wikis, τα εργαλεία επικοινωνίας, και οι folksonomies¹. Αν και ο παραπάνω όρος προτείνει μια νέα έκδοση του Παγκόσμιου Ιστού Πληροφοριών (World Wide Web), δεν αναφέρεται στον εκσυγχρονισμό τεχνικών προδιαγραφών, αλλά σε αλλαγές στον τρόπο που οι προγραμματιστές λογισμικού και οι τελικοί χρήστες χρησιμοποιούν το διαδίκτυο. Δύο σημαντικοί συγγραφείς και ερευνητές δίνουν τους δικούς τους ορισμούς ο καθένας.

«Όπως είχε υποθεί από το Stephen Fry το Web2.0 είναι περισσότερο μία ιδέα στο μυαλό των ανθρώπων παρά μία πραγματικότητα. Είναι μία ιδέα στην οποία δίνεται έμφαση η αμοιβαιότητα μεταξύ του χρήστη και αυτού που υπηρεσιές.

Με άλλα λόγια, εννοούμε γνήσια αλληλεπίδραση, απλά επειδή δίνεται η δυνατότητα στους χρήστες τόσο να ανεβάζουν πληροφορίες στο διαδίκτυο όσο και να κατεβάζουν»¹⁵

Σύμφωνα με τον παραπάνω ορισμό η ιδέα του Web2.0 περιλαμβάνει την παραγωγή και την διάδοση περιεχομένου από τους χρήστες δίνοντας τους την δυνατότητα και την ελευθερία να το μοιραστούν και να το ξαναχρησιμοποιήσουν. Αυτό έχει σαν αποτέλεσμα μία αύξηση στην οικονομική αξία του διαδικτύου καθώς οι χρήστες μπορούν να κάνουν περισσότερα πράγματα online.

«Όπως είχε υποθεί από τον Tim O'Reilly το Web2.0 είναι μία επιχειρησιακή επανάσταση στη βιομηχανία των ηλεκτρονικών υπολογιστών που προκλήθηκε από την ιδέα της «θεώρησης» του διαδικτύου σαν πλατφόρμα και την προσπάθεια για κατανόηση των κανόνων για επιτυχία σε αυτή τη νέα πλατφόρμα »¹⁶

Κατά τον Tim O'Reilly, δηλαδή, το Web2.0 θεωρείται σαν μία επιχείρηση που δέχεται το διαδίκτυο σαν πλατφόρμα και χρησιμοποιεί τις δυνατότητές του (όπως για παράδειγμα το παγκόσμιο κοινό, δηλαδή την υπάρχουσα ομάδα των χρηστών του web).

Ο παραπάνω ορισμός αποτελεί και τον τελευταίο συγκεντρωτικό ορισμό του Web2.0 καθώς το ακριβές νόημά του παραμένει ανοιχτό προς αντιπαράθεση, και μερικοί ειδικοί, συμπεριλαμβανομένου και του Tim Berners Lee, έχουν αμφισβητήσει το γεγονός ότι ο όρος έχει κάποιο πραγματικό νόημα.

Στο πρώτο συνέδριο που έγινε για το Web 2.0 τον Οκτώβριο του 2004, ο Tim O'Reilly και ο John Battelle κατέληξαν στις βασικές αρχές του Web2.0. Τα συμπεράσματα τους είναι τα παρακάτω:

- Το διαδίκτυο (χρησιμοποιείται) σαν πλατφόρμα.
- Ανάπτυξη κοινωνικών δικτύων, με την συμμετοχή και την αλληλεπίδραση των χρηστών.
- Τερματισμός του λεγόμενου «perpetual beta».

¹⁵ <http://web.educastur.princast.es/eoi/eoimiere/myweb/blog/wp-content/uploads/2009/03/transcriptweb20.pdf>

¹⁶ <http://oreilly.com/web2/archive/what-is-web-20.html>

- Ευκολία βελτίωσης από νέους κοινωνούς .
- Καινοτομία στην διασύνδεση των συστημάτων και των ιστοσελίδων που συνθέτονται από συνεργαζόμενα χαρακτηριστικά ανεξάρτητων προγραμματιστών (ένα είδος «open source» ανάπτυξης).

5.4.2 Από το Web 2.0 στο Web 3.0

Τα πρότυπα στο διαδίκτυο έχουν αλλάξει σημαντικά κατά την τελευταία δεκαετία. Η τεχνολογία έχει μετατοπιστεί από το Web 1.0 στο Web 2.0 και προς την κατεύθυνση τώρα Web 3.0. Ενώ Web 1.0 ήταν πολύ βασικό στην προσέγγισή του, το Web 2.0 εργάστηκε σε μια ιδέα που ονομάζεται Folksonomy. Σήμερα, Web 3.0 λειτουργεί σε Me-onomy, δηλαδή αφορούν τη συγκεκριμένη ειδική την οργάνωση.

Οι διαφορές μεταξύ του Web 2.0 και Web 3.0 τα πρότυπα:

- Το Web 2.0 χρησιμοποιεί την ανάγνωση web εγγραφής, Ιστολόγια, web εφαρμογές, εμπλουτισμένων μέσων, ιογενή μέσα ενημέρωσης, την τοποθέτηση πινακίδων(tagging) ή Folksonomy ενώ μοιράζονται περιεχόμενο και εστιάζοντας στις τοπικές κοινωνίες.
- Το Web 3.0 Πρότυπο χρησιμοποιεί semantic web, drag n drop mash ups, widgets, τη συμπεριφορά των χρηστών ή Me-onomy, διαφήμιση, συμμετοχή των χρηστών, ενοποιεί το δυναμικό περιεχόμενο και επικεντρώνεται σε ιδιώτες.
- Επίσης το Web 3.0 χρησιμοποιεί την τεχνολογία του «Web Data» που χαρακτηρίζει τις δομές αρχεία δεδομένων που είναι δημοσιεύσιμα και επαναχρησιμοποιήσιμα στο διαδίκτυο μέσω των ερωτημάτων μπορεί να έχει μορφές όπως RDF, XML . Είναι το σκαλοπάτι για να ολοκληρωθεί ο σημασιολογικός ιστός (semantic web), η οποία επιτρέπει νέα επίπεδα λειτουργικότητας εφαρμογής, την ενοποίηση των δεδομένων και καθιστά τα δεδομένα ανοιχτά που συσχετίζονται και είναι προσβάσιμα με τη μορφή των ιστοσελίδων. Η πλήρης θεωρητική σκηνή web επεκτείνει το πεδίο εφαρμογής των δύο δομημένων και μη δομημένου περιεχομένου μέσω της OWL και RDF σημασιολογικών μορφών.

- Ακόμη το web 3.0 πρότυπο περιγράφει επίσης τις τελευταίες τάσεις για την τεχνητή νοημοσύνη. Το Web 3.0 έχει ως στόχο να αναδείξει την ευφυΐα σε ένα οργανικό τρόπο μέσα από την αλληλεπίδραση των ανθρώπων.
- Τέλος το web 3.0 συσχετίζεται με την σημασιολογική έννοια και τη χρήση του web με τη τεχνητής νοημοσύνης. Υπάρχει αφθονία στη διαρκή έρευνα για την ανάπτυξη λογισμικού, το οποίο χρησιμοποιεί λόγους-αιτίες με βάση ευφυείς πράκτορες και τη λογική περιγραφή. Αυτές οι εφαρμογές εκτελούν όλες τις λογικές λειτουργίες και τους λόγους-αιτίες χρησιμοποιώντας το σύνολο των κανόνων, που εκφράζει τη λογική σχέση μεταξύ των δεδομένων στο διαδίκτυο και τις ιδέες τους.

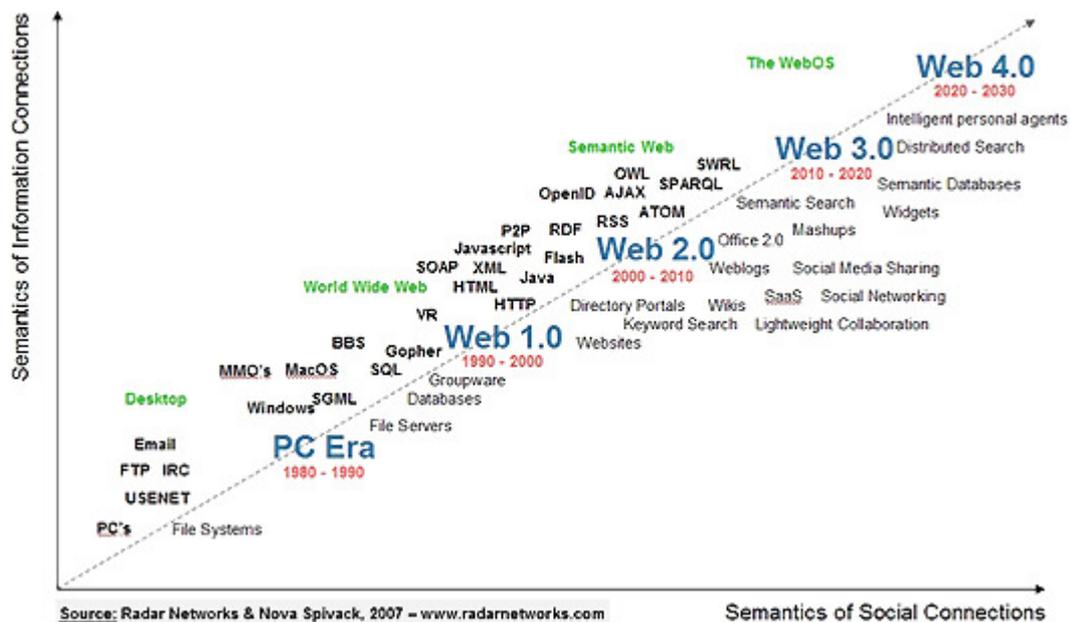
Η προτεινόμενη επέκταση της τεχνολογίας Web 3.0 περιλαμβάνει τα ακόλουθα:

- Συσκευές κινητού(smartphone), πανταχού συνδεσιμότητας, με κινητή πρόσβαση στο διαδίκτυο με την υιοθέτηση ευρυζωνικών υπηρεσιών.
- Το επιχειρηματικό μοντέλο SaaS με τα καταμεμημένα συστήματα πληροφορικής, cloud & grid computing (το σύννεφο και δίκτυο ηλεκτρονικών υπολογιστών) και web διαλειτουργικότητες υπηρεσιών.
- Ανοικτές τεχνολογίες(Open technologies), ανοικτές φόρμες δεδομένων και στοιχεία της άδειας(open data formats & data license), ανοικτές API(open API's) και πρωτόκολλα.
- Ανοιχτή ταυτότητα(Open identity), ανοιχτή φήμη(open reputation), δεδομένων προσωπικού χαρακτήρα .
- Οι ευφυείς τεχνολογίες web, σημασιολογικές τεχνικές όπως η OWL, RDF, SPARQL, SWRL και με βάση να αποθηκεύει δεδομένα με σημασιολογικό πλατφόρμες εφαρμογών.
- Ευφυής εφαρμογές, μηχανική μάθηση, αυτόνομοι παράγοντες, η συλλογιστική μηχανή και την επεξεργασία φυσικής γλώσσας.

Το Web 2.0 χρησιμοποιεί την ευφυΐα μεταξύ των ανθρώπων που είναι παρόμοια με το πλήθος των γνώσεων για τη δημιουργία χρήσιμων γνώσεων. Αντίθετα, το web 3.0 στοχεύει προς τα εμπρός με την εφαρμογή , που σκέφτεται μόνη της και συνδέει μία εφαρμογή στην άλλη, παρέχοντας έτσι κάποια πολύ καλά αποτελέσματα.

Παραδείγματος χάρη:

Μια τυπική εφαρμογή web 2.0 μπορεί να χρησιμοποιήσει σε απευθείας σύνδεση τη κοινότητα της γνώσης, η οποία περιλαμβάνει την τοποθέτηση πινακίδων τους ανθρώπους του περιεχομένου και την κατηγοριοποίηση της. Από την άλλη πλευρά, web 3.0 εφαρμογή χρησιμοποιεί συστήματα διαχείρισης περιεχομένου, μαζί με τεχνητή νοημοσύνη. Αυτά τα συστήματα είναι σε θέση να απαντήσει στις ερωτήσεις μας, γιατί μπορεί να σκεφτεί και να βρει την πιο πιθανή απάντηση στο ερώτημα. Αυτό σημαίνει ότι ο web 3.0 μπορεί επίσης να χαρακτηριστεί ως πρότυπο «Αλληλεπίδραση ανθρώπου μηχανής».



Εικόνα 8: Ανάπτυξη Web

Το Web 3.0 εισάγει ένα ζευγάρι των νέων εννοιών που χρησιμοποιούνται στην αναζήτηση. Παρατίθενται παρακάτω:

- Εξειδικευμένη Αναζήτηση
- Εξατομικευμένη Αναζήτηση
- Κατάλληλα διαμορφωμένη αναζήτηση
- Παραγωγικός συλλογισμός
- Εξέλιξη του 3D Web

Οι πραγματικές δυνατότητες του web 3.0 δεν έχουν ξεδιπλωθεί ακόμη για να αποδείξει ότι είναι πολύ πιο έξυπνο από web 2.0. Ενσωματώνει πληροφορίες σε ολόκληρο το domain web. Αναπτύσσει ρομπότ Ιστού που είναι αρκετά έξυπνα για τη λήψη αποφάσεων στην απουσία παρεμβολής του χρήστη. Συμφωνεί σε ένα κοινό έδαφος για το διαδίκτυο για να εκπροσωπήσει κάθε είδους περιεχόμενο που είναι εύκολα κατανοητό από οποιοδήποτε άλλο πρόγραμμα υπολογιστή μέσω της κοινής σημασιολογικής έννοιας που εκφράζεται από την σελίδα.

Συμπερασματικά, το Διαδίκτυο θα γίνει ένα καλύτερο μέρος για να «ζήσει» ουσιαστικά, αφού τα πρότυπα του web 3 αναπτύξουν τον εαυτό τους και να γίνουν πλήρως προσαρμόσιμα σε μαζική κλίμακα από τους χρήστες του Διαδικτύου παγκοσμίως.

5.5 Asynchronous JavaScript and XML (AJAX)

α) Τι είναι Asynchronous JavaScript and XML

Ο όρος «AJAX» αρχικά δημιουργήθηκε από την συντομογραφία του «Asynchronous JavaScript and XML». Ουσιαστικά το AJAX δεν είναι μια νέα τεχνολογία, ούτε μια νέα γλώσσα προγραμματισμού, αλλά ένας νέος τρόπος χρήσης των υπάρχοντων τεχνολογιών. Είναι ένα γκρουπ από αλληλένδετες τεχνικές ανάπτυξης ιστοσελίδων, που χρησιμοποιούνται στην πλευρά του Πελάτη (client-side) για την δημιουργία διαδραστικών διαδικτυακών εφαρμογών.

Με το AJAX, οι διαδικτυακές εφαρμογές μπορούν να ανακτούν δεδομένα από τον Διακομιστή (server) ασύγχρονα, δηλαδή στο παρασκήνιο, χωρίς να παρεμβαίνουν στην εμφάνιση και την συμπεριφορά της υπάρχουσας σελίδας. Η χρήση των τεχνικών του AJAX έχει οδηγήσει σε αύξηση των διαδραστικών ή δυναμικών διεπαφών στις ιστοσελίδες. Συνήθως, τα δεδομένα ανακτώνται χρησιμοποιώντας το αντικείμενο XMLHttpRequest, το οποίο θα αναλυθεί παρακάτω.

Σε αυτό το σημείο, αξίζει να επισημανθεί το εξής. Παρά το όνομά του, το AJAX δεν απαιτεί την χρήση της XML για την λήψη των δεδομένων, καθώς επίσης δεν είναι απαραίτητο τα αιτήματα προς των Διακομιστή να είναι ασύγχρονα. Αυτό συμβαίνει γιατί από το 2005, που αρχικά παρουσιάστηκε το AJAX ως ιδέα, μέχρι σήμερα, οι προγραμματιστές διαδικτύου το έχουν εμπλουτίσει με διάφορους τρόπους και το έχουν επεκτείνει ενσωματώνοντας αρκετές νέες τεχνολογίες. Το όνομα όμως έχει παραμείνει το ίδιο, καθώς ήταν απαραίτητο ένα διακριτό αναγνωριστικό για την χρήση όλων αυτών των τεχνολογιών, κάτω από ένα κοινό όρο ώστε να μην υπάρχει σύγχυση.

Η φόρτωση ασύγχρονου περιεχομένου έγινε πραγματικότητα για πρώτη φορά το 1995, με την πρώτη έκδοση της γλώσσας προγραμματισμού JAVA. Τότε παρουσιάστηκαν για πρώτη φορά τα «Java Applets», τα οποία επέτρεπαν «μεταγλωττισμένο» κώδικα στην μεριά του Πελάτη (client-side) να φορτώσει δεδομένα ασύγχρονα από τον διαδικτυακό Διακομιστή (web server), μετά την φόρτωση μιας ιστοσελίδας. Την ίδια εφαρμογή είχε και το «IFRAME» στοιχείο της γλώσσας σήμανσης HTML, το οποίο παρουσίασε μια χρονιά αργότερα, το 1996,

ο Internet Explorer της Microsoft. Το 1999 έγινε το μεγάλο βήμα, με τον Internet Explorer 5 να ενσωματώνει στα ActiveX Controls του το «περιτύλιγμα» (wrapper) XMLHttpRequest, το οποίο μέχρι και σήμερα χρησιμοποιείται από σχεδόν όλους τους περιηγητές Διαδικτύου ως το έμφυτο αντικείμενο XMLHttpRequest. Όμως, παρά την κοινή αποδοχή και την αναγνώριση της σπουδαιότητας αυτής της τεχνολογίας, έπρεπε να περάσουν αρκετά χρόνια, ώσπου το 2004 πρώτη η Google, να κάνει εκτεταμένη χρήση του AJAX στο Gmail, μια υπηρεσία ανταλλαγής ηλεκτρονικής αλληλογραφίας. Μέχρι τότε η χρησιμότητα των HTTP αιτήσεων προς τον Διακομιστή στο παρασκήνιο, καθώς και γενικά οι ασύγχρονες τεχνολογίες στο διαδίκτυο, παρέμεναν αρκετά ασαφής και η εφαρμογή και υποστήριξη τους από τους προγραμματιστές ήταν πολύ περιορισμένη. Το 2005, η Google και πάλι, με το Google Maps, μια υπηρεσία περιήγησης χαρτών, έκανε ευρέως γνωστό το AJAX, και πολλοί ερευνητές και προγραμματιστές άρχισαν να ασχολούνται μαζί του, παρόλο που δεν υπήρχε καν σαν όρος ή ορισμός τότε.

Πρωτεργάτης του όρου «AJAX» είναι ο Jesse James Garrett πρόεδρος και ιδρυτής της εταιρίας Adaptive Path, ο οποίος πρώτος στον κόσμο, στο άρθρο του «AJAX: A New Approach to Web Applications» τον Φεβρουάριο του 2005, έδωσε ένα όνομα στην δυνατότητα που υπήρχε με τις τότε υπάρχουσες τεχνολογίες, να γεφυρωθεί το κενό μεταξύ Διακομιστή και Πελάτη με την ασύγχρονη επικοινωνία μεταξύ αυτών.

Στις 5 Απριλίου του 2006, η διεθνής κοινοπραξία για την κατοχύρωση προτύπων του παγκόσμιου προδιαγραφής για το αντικείμενο, σε μια προσπάθεια να δημιουργήσει ένα επίσημο διαδικτυακό πρότυπο.

6.0 Τεχνολογίες

Ο όρος «AJAX» ήρε να αναπαραστήσει ένα ευρύ γκρουπ από διαδικτυακές τεχνολογίες οι οποίες μπορούν να χρησιμοποιηθούν για να παρέχουν τα μέσα σε μια διαδικτυακή εφαρμογή να επικοινωνήσει με έναν Διακομιστή στο παρασκήνιο, χωρίς να παρεμβαίνουν με την τρέχουσα κατάσταση της σελίδας. Στο άρθρο του όπου επινόησε τον όρο «AJAX», ο James Garrett εξηγούσε ότι οι απαιτούμενες τεχνολογίες ήταν οι εξής:

- Την γλώσσα σήμανσης HTML ή την XHTML και την γλώσσα φύλλων στυλ CSS, για την παρουσίαση
- Το Μοντέλο Αντικειμένου Εγγράφου (Document Object Model) ή DOM, για την δυναμική έκθεση των δεδομένων καθώς και την αλληλεπίδραση με αυτά.
- Την γλώσσα σήμανσης XML και την γλώσσα μετασχηματισμού XSLT για την ανταλλαγή, την διαχείριση και προβολή των δεδομένων
- Το αντικείμενο XMLHttpRequest για την ασύγχρονη επικοινωνία
- Την γλώσσα προγραμματισμού JavaScript σαν ενωτικό κρίκο όλων των παραπάνω τεχνολογιών

Έκτοτε όμως, επήλθαν πολλές εξελίξεις στις τεχνολογίες που χρησιμοποιούνταν για την ανάπτυξη μιας εφαρμογής AJAX, καθώς και στον καθορισμό του όρου AJAX. Ειδικότερα, έχει διαπιστωθεί ότι:

- Η γλώσσα JavaScript δεν είναι η μόνη γλώσσα προγραμματισμού από την πλευρά του Πελάτη που μπορεί να χρησιμοποιηθεί για την ανάπτυξη μιας εφαρμογής AJAX. Άλλες γλώσσες όπως η VBScript είναι ικανές να παρέχουν την απαιτούμενη λειτουργικότητα. Παρόλα αυτά όμως, η JavaScript έχει παραμείνει η πιο δημοφιλής γλώσσα στον προγραμματισμό του AJAX κυρίως για την ενσωμάτωση και την συμβατότητα που έχει με την πλειοψηφία των σύγχρονων περιηγητών Ιστού .
- Η γλώσσα σήμανσης XML δεν είναι απαραίτητη για την ανταλλαγή δεδομένων και κατά συνέπεια και η γλώσσα μετασχηματισμού XSLU δεν είναι αναγκαία για την διαχείριση των δεδομένων. Συχνά χρησιμοποιείται το μορφότυπο JSON (JavaScript Object Notation) σαν ένα εναλλακτικό

πρότυπο για την ανταλλαγή δεδομένων, παρόλο που και άλλα πρότυπα, όπως η προ-διαμορφωμένη HTML ή ακόμα και το απλό κείμενο, μπορούν επίσης να χρησιμοποιηθούν.

α) Μειονεκτήματα και Αδυναμίες

Λόγω της δυναμικής φύσης τους, οι AJAX διεπαφές είναι συχνά δυσκολότερο να αναπτυχθούν σε σύγκριση με τις στατικές σελίδες. Οι εφαρμογές AJAX αναπόφευκτα περιλαμβάνουν εκτέλεση πολύπλοκων τμημάτων JavaScript κώδικα στην μεριά του Πελάτη. Το να γίνει ένας τέτοιος περίπλοκος κώδικας αποδοτικός και χωρίς σφάλματα, είναι ένας στόχος που πρέπει να ληφθεί σοβαρά υπόψη.

Επίσης, οι σελίδες που δημιουργήθηκαν δυναμικά με την χρήση διαδοχικών αιτημάτων AJAX δεν καταγράφουν αυτόματα τον εαυτό τους στην μηχανή ιστορικού του Περιηγητή. Αυτό έχει ως αποτέλεσμα το πάτημα του κουμπιού «Πίσω» του Περιηγητή να μην μεταφέρει τον χρήστη πίσω σε μια προγενέστερη κατάσταση της AJAX σελίδας, αλλά πιθανόν να τον μεταφέρει στην τελευταία ολόκληρη σελίδα που είχε επισκεφτεί πριν από αυτή. Κάποιες τεχνικές επίλυσης αυτού του προβλήματος περιλαμβάνουν την χρήση αόρατων «IFrames» για να προκαλέσουν αλλαγές στο ιστορικό του Περιηγητή και για να αλλάζουν το «μέρος άγκυρας» (anchor portion) της διεύθυνσης Url (ακολουθούμενο από ένα σύμβολο «#» ή Hash) όταν εκτελεστεί το AJAX, και παρακολουθώντας το στην συνέχεια για αλλαγές.

Οι αλληπάλληλες ενημερώσεις των δυναμικών σελίδων έχουν δημιουργήσει ένα επιπλέον πρόβλημα, που έγκειται στο ότι ο χρήστης δύσκολα μπορεί να προσθέσει στους σελιδοδείκτες του μια συγκεκριμένη κατάσταση της εφαρμογής. Λύσεις για αυτό το πρόβλημα υπάρχουν, και πολλές από αυτές βασίζονται στην χρήση του «αναγνωριστικού κομματιού» της διεύθυνσης URL (το κομμάτι ενός URL που βρίσκεται μετά το σύμβολο «#») για να παρακολουθούν, και επομένως να επιτρέπουν στον χρήστη να αποθηκεύει, μια δεδομένη κατάσταση της εφαρμογής.

Είναι γεγονός ότι οι περισσότεροι αυτόματοι καταχωρητές διαδικτύου (web crawlers) δεν εκτελούν κώδικα JavaScript. Επομένως, οι διαδικτυακές εφαρμογές που έχουν προοριστεί για δημόσια (ανοικτή προς όλους) καταχώρηση, πρέπει να παρέχουν έναν εναλλακτικό μέσο για την προσπέλαση του περιεχομένου τους το οποίο, υπό κανονικές συνθήκες, θα είχε ανακτηθεί με την χρήση του AJAX, έτσι ώστε να επιτρέπουν στις μηχανές αναζήτησης να το καταχωρήσουν στις βάσεις τους.

Ένα πολύ σημαντικό μειονέκτημα είναι αυτό της συμβατότητας με συσκευές που δεν υποστηρίζουν όλες τις τεχνολογίες που αξιοποιεί μια AJAX εφαρμογή. Κάθε χρήστης που ο Περιηγητής του δεν υποστηρίζει την χρήση της JavaScript ή του αντικειμένου XMLHttpRequest, ή απλώς έχει απενεργοποιημένες αυτές τις λειτουργίες, δεν θα είναι σε θέση να χρησιμοποιήσει κατάλληλα τις σελίδες που στηρίζονται στην χρήση του AJAX. Ομοίως, συσκευές όπως κινητά τηλέφωνα, προσωπικοί ηλεκτρονικοί βοηθοί (Personal Digital Assistants ή PDAS) και οθόνες ανάγνωσης (Screen Readers) ενδεχομένως να μην παρέχουν υποστήριξη για τις απαιτούμενες τεχνολογίες. Ακόμη και οι οθόνες ανάγνωσης που υποστηρίζουν AJAX, μπορεί να μην έχουν την δυνατότητα να «διαβάσουν» κατάλληλα το δυναμικά παραγόμενο περιεχόμενο. Ο μόνος τρόπος για να επιτραπεί στον χρήστη να χρησιμοποιήσει την εφαρμογή, είναι να γυρίσουμε πίσω σε μεθόδους που δεν χρησιμοποιούν καθόλου την JavaScript. Αυτό μπορεί να επιτευχθεί μόνο εφόσον επιβεβαιωθεί ότι όλες οι διασυνδέσεις και οι φόρμες της εφαρμογής μπορούν να αποδοθούν κατάλληλα, και δεν θα βασίζονται αποκλειστικά στην χρήση AJAX. Έπειτα, από την πλευρά της JavaScript, η υποβολή της φόρμας θα μπορούσε να σταματήσει με την εντολή «return false»

Ένας πολύ σημαντικός περιορισμός στην χρήση του AJAX έγκειται στην κοινή πολιτική καταγωγής. Η πολιτική αυτή είναι ένα σημαντικό σχέδιο ασφάλειας το οποίο ισχύει για ορισμένες γλώσσες προγραμματισμού που εκτελούνται στην πλευρά του Περιηγητή (δηλαδή του Πελάτη) όπως η JavaScript. Συνοπτικά, η πολιτική αυτή επιτρέπει στα τρέχοντα κομμάτια κώδικα (running Scripts) των σελίδων, τα οποία προέρχονται από τον ίδιο δικτυακό τόπο, να έχουν πρόσβαση στις μεταξύ τους μεθόδους και ιδιότητες χωρίς ειδικούς περιορισμούς, αλλά τους

αποτρέπει να έχουν πρόσβαση στο μεγαλύτερο μέρος των μεθόδων και των ιδιοτήτων κατά μήκος των σελίδων που βρίσκονται σε διαφορετικούς δικτυακούς τόπους. Με βάση τα παραπάνω λοιπόν, μερικές τεχνικές που εφαρμόζονται στο AJAX δεν επιτρέπεται να χρησιμοποιηθούν κατά μήκος των τομέων, αλλά θα πρέπει να περιοριστούν σε χρήση μόνο κάτω από τον ίδιο τομέα. Μια λύση έρχεται από την κοινοπραξία AJAX, με την διάθεση ενός προσχεδίου πάνω στο αντικείμενο XMLHttpRequest το οποίο προορίζεται να του επιτρέψει αυτήν την λειτουργία.

Τελειώνοντας, πρέπει να σημειωθεί πως όπως και άλλες τεχνολογίες του διαδικτύου, το A^X έχει το δικό του σύνολο από τρωτά σημεία, τα οποία οι προγραμματιστές καλούνται να αντιμετωπίσουν. Είναι γεγονός, ότι προγραμματιστές με εξοικείωση σε άλλες τεχνολογίες του διαδικτύου θα πρέπει να μελετήσουν πάνω σε νέες μεθόδους κωδικοποίησης και δοκιμών, για να είναι σε θέση να γράψουν ασφαλείς AJAX εφαρμογές. Ένα επιπλέον θέμα που προκύπτει αφορά το ότι οι διεπαφές που κάνουν χρήση του AJAX μπορούν να αυξήσουν δραματικά τον αριθμό των αιτήσεων που παράγουν οι χρήστες, προς τους διακομιστές διαδικτύου καθώς και στις λειτουργίες που κρύβουν πίσω τους, όπως οι Βάσεις Δεδομένων και άλλες. Αυτό μοιραία μπορεί να οδηγήσει σε μεγαλύτερους χρόνους απόκρισης ή/και πρόσθετες ανάγκες για επιπλέον ή καλύτερο υλικό.

6.1 Σχεδιασμός Βάσης Δεδομένων

Με τον όρο βάση δεδομένων (αγγλ.: database) εννοείται μία συλλογή από συστηματικά μορφοποιημένα σχετιζόμενα δεδομένα στα οποία είναι δυνατή η ανάκτηση δεδομένων μέσω αναζήτησης κατ' απαίτηση¹⁷. Ο Αμερικανός επιστήμονας υπολογιστών Τζιμ Γκρέϊ (Jim Gray) έχει γράψει για τις βάσεις δεδομένων: «Όταν οι άνθρωποι χρησιμοποιούν τις λέξεις βάση δεδομένων, διατυπώνουν στην ουσία ότι τα δεδομένα πρέπει να αυτοπροσδιορίζονται και να έχουν μια σχηματική δομή. Αυτό ακριβώς περιγράφουν οι λέξεις βάση δεδομένων»¹⁸.

Ειδικότερα, στην επιστήμη της πληροφορικής και στην καθημερινή χρήση των ηλεκτρονικών υπολογιστών, με τον όρο βάσεις δεδομένων αναφερόμαστε σε οργανωμένες, διακριτές συλλογές σχετιζόμενων δεδομένων ηλεκτρονικά και ψηφιακά αποθηκευμένων, στο λογισμικό που χειρίζεται τέτοιες συλλογές (Σύστημα Διαχείρισης Βάσεων Δεδομένων, ή DBMS) και στο γνωστικό πεδίο που το μελετά. Πέρα από την εγγενή της ικανότητα να αποθηκεύει δεδομένα, η βάση δεδομένων παρέχει μέσω του σχεδιασμού και του τρόπου ιεράρχησης των δεδομένων, τα αποκαλούμενα συστήματα διαχείρισης περιεχομένου, δηλαδή τη δυνατότητα γρήγορης άντλησης και ανανέωσης των δεδομένων.

Μία βάση δεδομένων είναι δομημένη σε οντότητες και γνωρίσματα, που αντιπροσωπεύουν πραγματικές οντότητες και τα ορίσματά τους καθώς επίσης και σε συσχετίσεις μεταξύ των οντοτήτων. Ο χαρακτηρισμός ενός αντικείμενου ως οντότητα ή γνώρισμα μπορεί να αποδειχτεί δύσκολος και εξαρτώνται πάντα από τις απαιτήσεις ή πάντα από τις απαιτήσεις της τελικής εφαρμογής.

Οι ορισμοί των δομικών στοιχείων είναι οι εξής:

- Οντότητες

Η Οντότητα είναι ένα αντικείμενο που αναπαρίσταται στην ΒΔ και μπορεί να διαχωριστεί πλήρως. Αντιπροσωπεύει ένα αντικείμενο ή μία έννοια του πραγμα-

¹⁷http://www.usg.edu/galileo/skills/unit04/primer04_01.phtml

¹⁸ Jim Gray (2007-01-11). A Transformed Scientific Method

τικού κόσμου. Μια οντότητα αντιπροσωπεύει πολλά πράγματα τα οποία μοιράζονται κοινά χαρακτηριστικά. Πχ δύο διαφορετικοί ερευνητές είναι δύο στιγμιότυπα της οντότητας Ερευνητής. Οι οντότητες επίσης περιέχουν περιγραφική πληροφορία για ένα φυσικό αντικείμενο.

- Γνώρισμα

Ένα γνώρισμα περιγράφει μία ιδιότητα κάποιας οντότητας. Κατά την καταγραφή των γνωρισμάτων μιας οντότητας πρέπει να ληφθούν υπόψη όλες οι πληροφορίες που θα χρειαστεί να αποθηκευτούν για κάθε οντότητα, έτσι ώστε να επιτελούνται οι λειτουργίες που απαιτούνται από τους τελικούς χρήστες. Τα γνώρισμα είτε περιγράφουν είτε ταυτοποιούν μία οντότητα.

- Συσχετίσεις

Μια συσχέτιση μεταξύ οντοτήτων αντιπροσωπεύει μια αλληλεπίδραση μεταξύ τους. Για παράδειγμα η σχέση «ανήκει σε» αποτελεί μια συσχέτιση μεταξύ των οντοτήτων Ερευνητικό και Φορέα. Τυπικά μια συσχέτιση αντιπροσωπεύει αυτό που στον πραγματικό κόσμο εκφράζεται με ρήμα.

6.2 Μηχανές Αναζήτησης στο Παγκόσμιο Ιστό (Λειτουργία)

α) Ιστορική αναδρομή

Οι μηχανές αναζήτησης αποτελούν μία από τις σημαντικότερες ανακαλύψεις στην ιστορία του παγκόσμιου ιστού. Πριν τη δημιουργία των μηχανών αναζήτησης για να αποκτήσει πρόσβαση κάποιος σε περιεχόμενο του διαδικτύου έπρεπε να γνωρίζει την ακριβή διεύθυνση τοποθεσίας του περιεχομένου. Με τη δημιουργία των μηχανών αναζήτησης η γνώση της ακριβούς διεύθυνσης τοποθεσίας δεν είναι αναγκαία. Ο χρήστης έχει τη δυνατότητα να αναζητήσει αυτό που επιθυμεί με βάση το περιεχόμενο.

Η ιστορία των μηχανών αναζήτησης ξεκινά το 1990 στο πανεπιστήμιο McGill του Μόντρεαλ, όπου δημιουργήθηκε μία εφαρμογή με το όνομα Archie. Το 1990 δεν υπήρχε ο παγκόσμιος ιστός, υπήρχαν όμως αρχεία σε διάφορα σημεία του διαδικτύου. Η εφαρμογή Archie συνέλεγε λίστες με αρχεία, τα κατηγοριοποιούσε και επέτρεπε την εύρεση ονομάτων αρχείων με βάση την προτίμηση του χρήστη. Η κύρια μέθοδος προσπέλασης των αρχείων ήταν το πρωτόκολλο File Transfer Protocol (FTP), το οποίο χρησιμοποιείται ευρέως και σήμερα για την ανταλλαγή αρχείων μεταξύ υπολογιστών.

Το 1991 στο πανεπιστήμιο της Μινεσότα δημιουργήθηκε μία συλλογή διαδικτυακών διευθύνσεων σε μορφή μενού, η οποία ονομάστηκε "gophers". Η gophers ήταν τοπική μηχανή αναζήτησης, παρέχοντας μόνο τίτλους αρχείων και προαιρετικά μία σύντομη περιγραφή. Με βάση τη συλλογή διευθύνσεων της gophers, το 1993 αναπτύχθηκε στο πανεπιστήμιο της Νεβάδα η μηχανή αναζήτησης Veronica. Λίγο αργότερα εμφανίστηκε η μηχανή αναζήτησης Jughead, η λειτουργία της οποίας ήταν παρόμοια με της Veronica. Η λειτουργία των μηχανών αναζήτησης με βάση τη gophers διήρκεσε μερικά χρόνια, έως τη ραγδαία ανάπτυξη του παγκόσμιου ιστού, ο οποίος επέτρεψε τους συνδέσμους μεταξύ ιστοσελίδων, την αναζήτηση πλήρους κειμένου, τους γραφικούς φυλλομετρητές και άλλες εύκολες στη χρήση τεχνολογίες.

πρώτη επιτυχημένη διαδικτυακή μηχανή αναζήτησης ήταν η WebCrawler, η οποία δημιουργήθηκε στο πανεπιστήμιο της Ουάσιγκτον, ξεκινώντας την επίσημη λειτουργία της το 1994. Μέσα σε ένα χρόνο έκαναν την εμφάνισή τους άλλες 3 μηχανές αναζήτησης, οι Lycos, Infoseek και OpenText.

Στα τέλη του 1995 εμφανίστηκαν οι AltaVista και Excite. Οι περισσότερες από τις τεχνολογίες αναζήτησης, οι οποίες χρησιμοποιούνται σήμερα υπήρχαν σε διάφορες μορφές στις πρώτες μηχανές αναζήτησης. Η μηχανή αναζήτησης HotBot εμφανίστηκε το 1996 και η Northern Light το 1997. Η HotBot παρείχε μία περίπλοκη αλλά εύκολη στη χρήση διεπαφή και μία αρκετά μεγάλη βάση δεδομένων, η οποία ήταν η μεγαλύτερη διαθέσιμη μέχρι το τέλος του 2007. Η Northern Light παρείχε αναζήτηση διαδικτύου παράλληλα με αναζήτηση ιδιοκτητών πληροφοριών.

Η μηχανή αναζήτησης Google εμφανίστηκε το 1998 και με τα αποτελέσματά της, τα οποία βασίζονται στη δημοτικότητα των ιστοσελίδων, αλλά και με την πολύ απλή διεπαφή της κατάφερε να αποκτήσει αμέσως δημοτικότητα μεταξύ των χρηστών μηχανών αναζήτησης. Τα επόμενα χρόνια όσες μηχανές αναζήτησης δε μπόρεσαν να ακολουθήσουν την εξέλιξη της διαδικτυακής τεχνολογίας εξαφανίστηκαν ή εξαγοράστηκαν από άλλες μηχανές αναζήτησης. Δημιουργήθηκαν και μερικές νέες μηχανές αναζήτησης χωρίς ωστόσο να αντέξουν τον ανταγωνισμό, με μόνη εξαίρεση τη Bing, η οποία είναι ιδιοκτησία της εταιρείας Microsoft Corporation. Η μηχανή αναζήτησης Bing δεν είναι στην ουσία μία νέα μηχανή αναζήτησης, αλλά αντικατέστησε τις προηγούμενες μηχανές αναζήτησης Live Search, Windows Live Search και MSN Search της ίδιας εταιρείας.

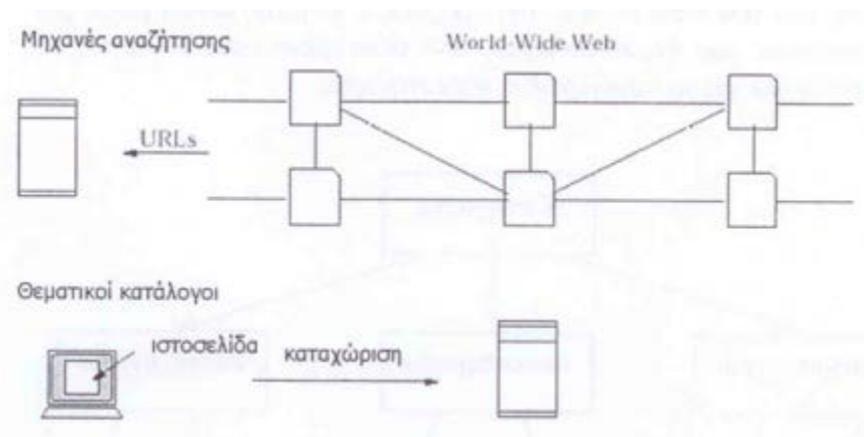
6.3 Crawler-based μηχανές αναζήτησης

Ο Jerri L. Ledford(2008) ορίζει τις crawler-based μηχανές αναζήτησης ως ένα κομμάτι λογισμικού, το οποίο χρησιμοποιεί εφαρμογές για να συλλέξει πληροφορίες σχετικά με τις ιστοσελίδες του παγκόσμιου ιστού. Οι πληροφορίες, οι οποίες συλλέγονται είναι συνήθως λέξεις-κλειδιά. Οι λέξεις-κλειδιά είναι πιθανοί δείκτες του περιεχομένου της σελίδας, η διεύθυνση της ιστοσελίδας, ο κώδικας, ο οποίος δημιουργεί τη σελίδα και σύνδεσμοι από και προς την ιστοσελίδα.

Οι πληροφορίες αποθηκεύονται σε βάσεις δεδομένων. Το μέρος των μηχανών αναζήτησης, το οποίο αλληλεπιδρά με τους χρήστες, είναι μία ιστοσελίδα. Στην ιστοσελίδα κάθε χρήστης μπορεί να εισάγει μία λέξη ή φράση, για την οποία αναζητεί πληροφορίες. Μόλις ο χρήστης «πατήσει» το κουμπί αναζήτησης, ένας αλγόριθμος εξετάζει τις αποθηκευμένες πληροφορίες και ανακτά συνδέσμους σε ιστοσελίδες, οι οποίες έχουν σχέση με τη λέξη ή φράση εισαγωγής του χρήστη.

6.4 Θεματικοί κατάλογοι

Οι θεματικοί κατάλογοι επιτρέπουν την καταχώρηση ιστοσελίδων από ανθρώπους. Διαθέτουν πολύ καλή οργάνωση αλλά αρκετά λιγότερο υλικό από τις μηχανές αναζήτησης. Ένας θεματικός κατάλογος έχει τη μορφή ενός ιεραρχικού γραφήματος με βασικές κατηγορίες στα υψηλά επίπεδα της ιεραρχίας και ειδικότερες κατηγορίες στα χαμηλότερα επίπεδα. Ο χρήστης κατεβαίνει στην ιεραρχία ακολουθώντας τις κατάλληλες υπερ-συνδέσεις. Στο τελευταίο επίπεδο της ιεραρχίας ο χρήστης καταλήγει σε μία λίστα εγγράφων, τα οποία έχουν επιλεγεί από ειδικευμένο προσωπικό ως τα πλέον κατάλληλα για τη συγκεκριμένη υποκατηγορία. (Ανδρέας Βεγλής, Ανδρέας Πομπόρτσας, Ευαγγελία Αβραάμ (2004))¹⁹



Εικόνα 9:Μηχανές αναζήτησης και θεματικοί κατάλογοι

¹⁹ Έρευνα και συλλογή πληροφοριών στο διαδίκτυο

6.5 Μετα-μηχανές αναζήτησης

Εκτός από τις μηχανές αναζήτησης και τους θεματικούς καταλόγους υπάρχουν και οι μετα-μηχανές αναζήτησης. Στην πραγματικότητα δεν είναι μηχανές αναζήτησης, δεν χρησιμοποιούν μία αράχνη για να συγκεντρώσουν τη δική τους βάση δεδομένων. Χρησιμοποιούν τα αποτελέσματα των μηχανών αναζήτησης και τα παρουσιάζουν στο χρήστη σε μία συγκριτική σελίδα. Ο μηχανισμός των μετα-μηχανών αναζήτησης ανακτά μόνο το 10% των αποτελεσμάτων από κάθε μηχανή αναζήτησης. Για να εμφανιστεί μία ιστοσελίδα σε μία μετα-μηχανή αναζήτησης, πρέπει να εμφανίζεται στο 10% κάποιας από τις μηχανές αναζήτησης, τις οποίες χρησιμοποιεί η συγκεκριμένη μετα-μηχανή αναζήτησης για να ανακτήσει αποτελέσματα.

Συνοψίζοντας τα ερωτήματα τα οποία κατά τη σύνταξή τους κάνουν χρήση λογικών ή άλλων τελεστών γίνονται αντικείμενο ιδιαίτερης επεξεργασίας. Οι τελεστές ομοιότητας (AND, +) υπαγορεύουν στους μηχανισμούς εντοπισμού πληροφορίας του συστήματος την αναγκαιότητα εμφάνισης των όρων αυτών μέσα στα δεικτοδοτημένα κείμενα, ενώ αντίθετα τελεστές απουσίας (NOT, -) υπαγορεύουν το αντίθετο. Σε περιπτώσεις όπου ένα ερώτημα διατυπώνεται με τη χρήση (εκ)φράσεων, οι αντίστοιχες φράσεις αναζητούνται τόσο στους όρους δεικτοδότησης, όσο και στα ίδια τα κείμενα. Συνοπτικά, οι μηχανισμοί ανίχνευσης της επιθυμητής πληροφορίας μέσα στο πλήθος των δεικτοδοτημένων ιστοσελίδων ποικίλλουν σε κάθε μηχανή αναζήτησης και μεταβάλλονται διαρκώς, για τη βελτίωση της απόδοσης των συστημάτων κατά την ανάκτηση πληροφορίας.

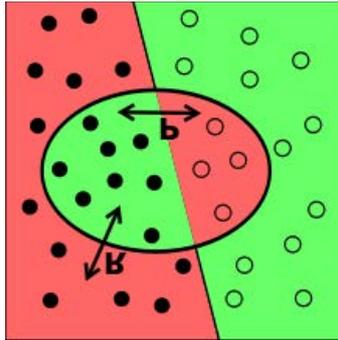
7.0 Μετρικές Απόδοσης Αλγορίθμων Μηχανών Αναζήτησης

α) Ακρίβεια και ανάκληση

Στην αναγνώριση προτύπων και την ανάκτηση πληροφοριών, την ακρίβεια (ονομάζεται επίσης θετική προγνωστική αξία) είναι το κλάσμα των ανακτημένων περιπτώσεων που σχετίζονται, ενώ η ανάκληση (επίσης γνωστή ως ευαισθησία) είναι το κλάσμα των σχετικών περιπτώσεων που ανακτώνται. Τόσο η ακρίβεια και η ανάκληση σχετίζονται με βάση την κατανόηση και το μέτρο της συνάφειας. Ας υποθέσουμε όταν μια μηχανή αναζήτησης επιστρέφει 30 σελίδες μόνο, από τις οποίες 20 ενώ αποτυγχάνει σχετικά να επιστρέψει 40 επιπλέον σχετικές σελίδες, για την ακρίβεια είναι $20/30 = 2/3$ ενώ η ανάκλησή της είναι $20/60 = 1/3$.

Όσον αφορά τις στατιστικές, αν η μηδενική υπόθεση είναι ότι όλα και μόνο τα σχετικά στοιχεία ανακτώνται, η απουσία του τύπου I και τύπου II λάθη αντιστοιχεί αντίστοιχα στη μέγιστη ακρίβεια (δεν υπάρχουν ψευδώς θετικά) και η μέγιστη ανάκληση (δεν υπάρχει ψευδώς αρνητικά). Το παραπάνω παράδειγμα αναγνώρισης προτύπων περιείχε $7 - 4 = 3$ τύπου I λάθη και $9 - 4 = 5$ λάθη τύπου II. Η ακρίβεια μπορεί να θεωρηθεί ως μέτρο ακρίβειας ή της ποιότητας, ενώ η ανάκληση είναι ένα μέτρο της πληρότητας ή της ποσότητας.

Με απλά λόγια, η υψηλή ανάκληση σημαίνει ότι ένας αλγόριθμος επέστρεψε τα περισσότερα από τα σχετικά αποτελέσματα, ενώ η υψηλή ακρίβειας σημαίνει ότι ένας αλγόριθμος επέστρεψε πολύ πιο σχετικά αποτελέσματα.



Εικόνα 10: Σε αυτό το σχήμα, τα σχετικά στοιχεία βρίσκονται στα αριστερά της ευθείας γραμμής, ενώ τα ανακτημένα στοιχεία στο οβάλ. Οι κόκκινες περιοχές αντιπροσωπεύουν λάθη. Στα αριστερά είναι τα σχετικά στοιχεία που δεν ανακτώνται (ψευδώς αρνητικά), ενώ στα δεξιά είναι τα ανακτηθέντα στοιχεία που δεν είναι σχετικά (ψευδώς θετικά).

7.1.1 Ορισμός (πληροφορίες πλαίσιο ανάκτησης)

Στο πλαίσιο ανάκτησης πληροφοριών, την ακρίβεια και την ανάκληση ορίζονται από την άποψη της ένα σύνολο των ανακτημένων εγγράφων (π.χ. τον κατάλογο των εγγράφων που παράγεται από μια μηχανή αναζήτησης στο διαδίκτυο για ένα ερώτημα) και μια σειρά από σχετικά έγγραφα (π.χ. ο κατάλογος όλων των εγγράφων στο διαδίκτυο που είναι σημαντικές για ένα συγκεκριμένο θέμα.

α) Ακρίβεια

Στον τομέα της ανάκτησης πληροφοριών, η ακρίβεια είναι το κλάσμα των ανακτημένων εγγράφων που σχετίζονται με την αναζήτηση:

$$\text{Ακρίβεια} = \frac{|\{\text{σχετικά έγγραφα}\} \cap \{\text{ανακτημένα έγγραφα}\}|}{|\{\text{ανακτημένα έγγραφα}\}|}$$

Η ακρίβεια λαμβάνει όλα τα ανακτημένα έγγραφα υπόψη, αλλά μπορεί επίσης να αξιολογηθούν σε μια δεδομένη βαθμίδα cut-off, λαμβάνοντας υπόψη μόνο τα πιο πάνω αποτελέσματα που επιστρέφονται από το σύστημα.

Για παράδειγμα, για μια αναζήτηση κειμένου σε ένα σύνολο εγγράφων η ακρίβεια είναι ο αριθμός των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό όλων των επιστρεφόμενων αποτελεσμάτων.

β) Ανάκληση

$$\text{Ανάκληση} = \frac{|\{\text{σχετικά έγγραφα}\} \cap \{\text{ανακτημένα έγγραφα}\}|}{|\{\text{ανακτημένα έγγραφα}\}|}$$

Για παράδειγμα, για την αναζήτηση κειμένου σε ένα σύνολο εγγράφων η ανάκληση είναι ο αριθμός των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό των αποτελεσμάτων που θα έχουν επιστραφεί.

Στη δυαδική ταξινόμηση, ανάκληση ονομάζεται ευαισθησία. Έτσι μπορεί να αντιμετωπίζεται ως η πιθανότητα από ένα σχετικό έγγραφο που ανακτάται από το ερώτημα.

Είναι ασήμαντο για να επιτύχει η ανάκληση 100% με την επιστροφή όλων των εγγράφων σε απάντηση σε οποιοδήποτε ερώτημα. Ως εκ τούτου, η ανάκληση από μόνο της δεν είναι αρκετή, αλλά κάποιος πρέπει να μετρήσει τον αριθμό των μη-σχετικών εγγράφων, για παράδειγμα με τον υπολογισμό της ακρίβειας.

7.2 Μέθοδος υπολογισμού της σπουδαιότητας των όρων ενός κειμένου: η μετρική TF-IDF

Η συγκρουόμενη φύση των μέτρων της ακρίβειας και της ανάκτησης επιβάλλουν την επιλογή της κατάλληλης μεθόδου για τον υπολογισμό των βαρών, με τέτοιο τρόπο ώστε να ικανοποιούνται και τα δύο μέτρα απόδοσης. Αυτό επιτυγχάνεται με τη χρήση πολύπλοκων παραγόντων στον υπολογισμό του βάρους, οι οποίοι περιέχουν συστατικά μέρη που ικανοποιούν τόσο το μέτρο της ακρίβειας, όσο και της ανάκτησης. Αυτοί οι παράγοντες είναι τρεις, και επηρεάζουν σημαντικά τη σύνδεση της μεθόδου με τα μέτρα απόδοσης: Ο παράγοντας συχνότητας όρων (term frequency factor), ο παράγοντας αντίστροφης συχνότητας εγγράφου (inverse document frequency factor) και ο παράγοντας κανονικοποίησης (normalization factor).

Αναλυτικά:

Παράγοντας συχνότητας όρων: Πρώτα από όλα, οι όροι οι οποίοι εμφανίζονται συχνά σε ένα έγγραφο είναι χρήσιμοι, σε ό,τι αφορά στην ικανοποίηση του παράγοντα της ανάκτησης. Μια λέξη που εμφανίζεται με μεγάλη συχνότητα στο έγγραφο, είναι πιθανό να σχετίζεται με το θέμα του εγγράφου και σαν όρος συντελεί στην ανάκτηση του, όταν γίνει ερώτηση για σχετικό θέμα. Επομένως,

είναι απαραίτητο στον υπολογισμό του βάρους να συμπεριληφθεί ο παράγοντας συχνότητας εμφάνισης όρου (term frequency-tf), ο οποίος ισούται με τον αριθμό των εμφανίσεων ενός όρου σε ένα έγγραφο. Ο παράγοντας της συχνότητας εμφάνισης όρου έχει χρησιμοποιηθεί από πολύ νωρίς στο σχεδιασμό ανάλογων συστημάτων αυτόματης ανάλυσης κειμένου. Ο παράγοντας tf εξασφαλίζει το ανώτατο όριο του νόμου του Zipf που αναφέρθηκε παραπάνω, αποκλείοντας τις λέξεις με χαμηλή συχνότητα εμφάνισης από την αντιπροσώπευση του εγγράφου.

Παράγοντας αντίστροφης συχνότητας εγγράφου: Ο παράγοντας της συχνότητας των όρων δεν εξασφαλίζει από μόνος του ικανοποιητική απόδοση στην ανάκτηση εγγράφων από το σύστημα. Οι όροι με υψηλές συχνότητες εμφάνισης τείνουν να απαντώνται σε πολλά έγγραφα μιας συλλογής, με αποτέλεσμα να μην αναγνωρίζεται ικανοποιητικά το θεματικό περιεχόμενο των κειμένων κάνοντας χρήση αυτών. Επιπλέον, στην περίπτωση ενός μόνο εγγράφου, η μεγάλη υψηλή εμφάνιση μιας λέξης δε διασφαλίζει εκ προοιμίου πως η υφιστάμενη λέξη σχετίζεται άμεσα με το θέμα του υπό εξέταση εγγράφου. Είναι πιθανόν η εν λόγω λέξη να είναι βοηθητική και να απαιτείται η χρήση κάποιου περισσότερο χαρακτηριστικού όρου του κειμένου για την αντιπροσώπευση του θέματος αυτού. Τούτο έχει σαν αποτέλεσμα την αδυναμία ικανοποίησης του μέτρου της ακρίβειας. Ο παράγοντας idf λαμβάνει υπόψη του την εμφάνιση ενός όρου σε πλήθος εγγράφων και αντισταθμίζει την υπερβολικά υψηλή συχνότητα εμφάνισης ενός όρου, γεγονός που μπορεί να σημαίνει ότι ο όρος είναι βοηθητικός και όχι χαρακτηριστικός για το θεματικό περιεχόμενο του εν λόγω κειμένου. Εξασφαλίζει, δηλαδή, το ανώτατο όριο στο νόμο του Zipf που αναφέρθηκε προηγουμένως.

Παράγοντας κανονικοποίησης: Σε συστήματα όπου δεν υπάρχει συγκεκριμένο μήκος για τα διανύσματα όρων, υπάρχει η τάση τα μικρού όγκου έγγραφα να αντιπροσωπεύονται από διανύσματα με μικρότερο αριθμό όρων, σε αντίθεση με τα εκτενή έγγραφα, των οποίων η αντιπροσώπευση γίνεται από διανύσματα με περισσότερους όρους. Κάτι τέτοιο βέβαια, σε μία αναζήτηση εγγράφου αδικεί τα σύντομα έγγραφα διότι η πιθανότητα ανάκτησής τους είναι μικρότερη από

ότι για τα εκτενή έγγραφα. Η ανισότητα αυτή πρέπει να εξαλειφθεί, καθώς στην αναζήτηση ενός εγγράφου με συγκεκριμένο θέμα πρέπει όλα τα έγγραφα να αντιμετωπίζονται με τον ίδιο τρόπο, ανεξάρτητα από το μέγεθός τους. Για την εξάλειψη της ανισότητας στην πιθανότητα ανάκτησης, συμπεριλαμβάνουμε στη μέθοδο υπολογισμού του βάρους των όρων έναν παράγοντα κανονικοποίησης. Ο παράγοντας κανονικοποίησης εξισορροπεί τον αριθμό των χαρακτηριστικών όρων, κανονικοποιώντας τα βάρη και εξασφαλίζοντας ότι όλα χαρακτηρίζονται από τιμές που κυμαίνονται μεταξύ του 0 και του 1. Όταν επιλέξου με λοιπόν τα όρια των βαρών, μέσα στα οποία περιέχονται οι λέξεις κλειδιά, θα έχουμε εξισορρόπηση του πλήθους τους για όλα τα έγγραφα.

Συνεπώς, ο υπολογισμός του βάρους ενός όρου k , σύμφωνα με τα παραπάνω, επιτυγχάνεται με την εφαρμογή του τύπου:

$$\text{Βάρος}_k = \frac{tf_k \times idf_k}{\sqrt{\sum_{i \in \text{vector}} (tf_i \times idf_i)}}$$

όπου tf_k είναι ο παράγοντας συχνότητας του όρου k , idf_k είναι ο παράγοντας αντίστροφης συχνότητας εγγράφου του όρου k , tf_i είναι ο παράγοντας συχνότητας ενός όρου i που ανήκει στο διάνυσμα όρων, και idf_i είναι ο παράγοντας αντίστροφης συχνότητας εγγράφου ενός όρου i . Το Ευκλείδειο μήκος του διανύσματος όρων, που είναι ο παράγοντας κανονικοποίησης, υπολογίζεται με τη χρήση των παραγόντων $tf \times idf$ όλων των όρων που ανήκουν στο διάνυσμα όρων. Η παραπάνω μέθοδος υπολογισμού των βαρών ονομάζεται μέθοδος TFIDF (Term Frequency-Inverse Document Frequency).

Ο τρόπος υπολογισμού των παραγόντων δεν είναι μοναδικός. Υπάρχουν πολλοί διαφορετικοί τρόποι να υπολογιστεί κάθε παράγοντας και οι συνδυασμοί αυτών ορίζουν διάφορες μορφές της μεθόδου TFIDF για τον υπολογισμό των βαρών²⁰.

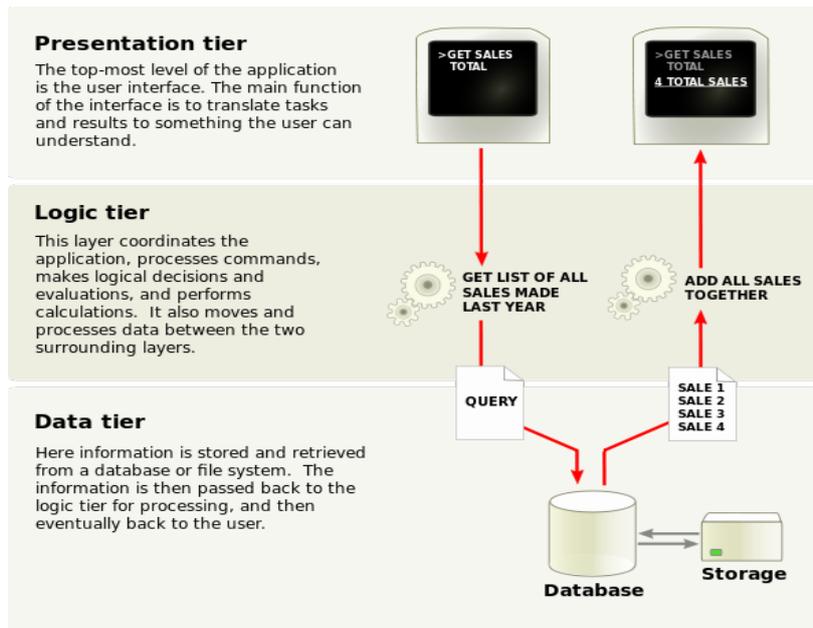
²⁰ Term-weighting approaches in automatic text retrieval (1988) Cached by Gerard Salton , Christopher Buckley

7.3 Αρχιτεκτονική τριών επιπέδων (3-Tier Architecture)

Η αρχιτεκτονική 3-tier αποτελείται από τρία επίπεδα (tiers). Τα επίπεδα αυτά αποτελούν λογικό διαχωρισμό και όχι φυσικό διαχωρισμό. Δεν είναι απαραίτητο να βρίσκονται σε διαφορετικούς ηλεκτρονικούς υπολογιστές. Ο κώδικας της εφαρμογής εμπεριέχεται σε αυτόνομα κομμάτια τα οποία αποκαλούνται components.

Τα τρία επίπεδα είναι τα εξής:

- Data Tier – Υπηρεσίες που έχουν σχέση με την ανάκτηση ή την αποθήκευση δεδομένων. Σε αυτό το επίπεδο υπάγονται οι βάσεις δεδομένων, παλιότερες μονολιθικές εφαρμογές που παρέχουν πρόσβαση σε δεδομένα και άλλες, εξωτερικές εφαρμογές που αποτελούν οι ίδιες πηγές δεδομένων.
- Business Tier – Η «λογική» των εφαρμογών. Οι υπηρεσίες που προσφέρει η εφαρμογή. Σε αυτό το επίπεδο υπάγονται τα κομμάτια της εφαρμογής που περιέχουν την λογική της εφαρμογής. Δηλαδή οι κανόνες και η λειτουργικότητα της εφαρμογής
- Presentation Tier – Αυτό που παρουσιάζεται στον χρήστη. Το User Interface. Σε αυτό το επίπεδο υπάγονται τα διάφορα User Interfaces που συναντούμε. Αυτά χωρίζονται σε thin clients ή rich clients. Τα thin clients είναι σχετικά απλά interfaces, όπως παραδείγματος χάριν μια σελίδα HTML την οποία βλέπουμε μέσω κάποιου browser. Τα rich clients από την άλλη είναι πλούσια interfaces, όπως παραδείγματος χάριν μια 32-bit η εφαρμογή Windows. Υπάρχουν φυσικά και τα ενδιάμεσα στάδια, όπως είναι η χρήση ActiveX components και DHTML σε μια ιστοσελίδα κλπ.



Εικόνα 11:Αρχιτεκτονική τριών επιπέδων (3-Tier Architecture)

Σημείωση: Ο όρος εφαρμογή «n-tier» είναι ένας διαφορετικός τρόπος αναφοράς σε κατανεμημένες εφαρμογές και αποτελεί έναν ειδικό τύπο 3-tier εφαρμογών, όπου ένα ή παραπάνω από τα επίπεδα της εφαρμογής χωρίζεται σε επιπρόσθετα επίπεδα. Οι αρχές που ισχύουν για 3-tier εφαρμογές ισχύουν και για n-tier εφαρμογές

α) Πλεονεκτήματα εφαρμογών 3-tier

Εφαρμογές που έχουν βασιστεί στην αρχιτεκτονική 3-tier αποκαλούνται συχνά server-centric, διότι επιτρέπουν στα κομμάτια της εφαρμογής (components) που ανήκουν στο business tier να τρέχουν σε ειδικούς servers που είναι εντελώς ανεξάρτητοι από το user interface ή την υλοποίηση των βάσεων δεδομένων που ανήκουν στο data tier. Αυτή η ανεξάρτηση της λογικής της εφαρμογής από το presentation tier και το data tier επιφέρει πολλά πλεονεκτήματα.

Μερικά από αυτά τα πλεονεκτήματα συνοψίζονται στον παρακάτω πίνακα.

Υποστήριξη πολλαπλών προγραμματιστικών γλωσσών	Τα components μιας εφαρμογής μπορούν να γραφτούν άνετα σε διαφορετικές προγραμματιστικές γλώσσες. Θα μπορούσε, παραδείγματος χάριν, η
--	---

	<p>λογική της εφαρμογής να υλοποιηθεί σε C++, ενώ το user interface θα μπορούσε να υλοποιηθεί σε Visual Basic.</p>
Κεντριοποιημένα components	<p>Τα components όλων των εφαρμογών μπορούν να τρέχουν σε κεντρικούς διακομιστές (servers), έτσι ώστε να γίνεται πολύ απλούστερη η διαδικασία συντήρησης ή αλλαγής τους.</p>
Load Balancing	<p>Τα components μιας εφαρμογής μπορούν να μοιραστούν σε πολλαπλούς servers στο δίκτυο, επιτρέποντας έτσι την εξισορρόπηση του φόρτου εργασίας του κάθε server.</p>
Καλύτερη πρόσβαση στα δεδομένα	<p>Περιορισμοί που πιθανόν να υπήρχαν στην σύνδεση πολλαπλών χρηστών στην βάση δεδομένων εκ μηδενίζονται, καθώς η βάση «μιλάει» πλέον μόνο με κάποια συγκεκριμένα components. Επίσης, οι οδηγοί και η σύνδεση για την βάση δεδομένων δεν χρειάζεται πλέον να βρίσκονται τοπικά στο μηχάνημα του χρήστη. Οι πόροι του συστήματος χρησιμοποιούνται πολύ καλύτερα , καθώς στην αρχιτεκτονική 3-tier οι συνδέσεις προς την βάση δεδομένων γίνονται μόνο όταν χρειάζονται και απελευθερώνονται αμέσως μόλις εκπληρώσουν τον σκοπό τους .</p>

<p>Βελτιωμένη ασφάλεια</p>	<p>Οι components που ανήκουν στο Business Tier μπορούν πλέον να διασφαλιστούν χρησιμοποιώντας κάποια κεντρικοποιημένη υποδομή.</p> <p>Η πρόσβαση στο κάθε component μπορεί να δωθεί ή να αρνηθεί ξεχωριστά ,διευκολύνοντας το ρόλο των Administrators.</p>
<p>Απλοποιημένη πρόσβαση σε εξωτερικούς πόρους</p>	<p>Η πρόσβαση σε εξωτερικές εφαρμογές όπως είναι εφαρμογές που τρέχουν πάνω σε mainframes που συλλέγουν δεδομένα απλοποιείται σημαντικά. Ένας gateway server π.χ., θεωρείται απλά άλλο ένα component που χρησιμοποιεί η εφαρμογή</p>

7.4 Components

Components αποκαλούνται ανεξάρτητες, αυτοτελείς μονάδες λογισμικού οι οποίες παρέχουν συγκεκριμένη λειτουργικότητα στα πλαίσια μιας εφαρμογής. Υπάρχουν πολλοί λόγοι για την χρήση components σε κάποιο application.

- Επαναχρησιμοποιούμενος κώδικας

Ένα component μπορεί εύκολα να χρησιμοποιηθεί από τον προγραμματιστή που το ανέπτυξε ή από κάποιον άλλο. Ένα component σχεδιασμένο για πρόσβαση σε μια βάση δεδομένων παραδείγματος χάρη, μπορεί να χρησιμοποιηθεί από πολλές διαφορετικές εφαρμογές, αν έχουν περίπου τις ίδιες ανάγκες πρόσβασης σε δεδομένα.

- Μειωμένη πολυπλοκότητα

Δημιουργώντας components, ένας προγραμματιστής 'κρύβει' την πολυπλοκότητα του κώδικά του. Η μόνη γνώση που χρειάζεται να έχει κάποιος άλλος προγραμματιστής για να χρησιμοποιήσει το component του είναι γνώση των μεθόδων (methods – functions) του component. Πρέπει δηλαδή να ξέρει μόνο τι πληροφορίες πρέπει να δώσει στο component και τι πληροφορίες να περιμένει από το component.

- Ευκολότερη συντήρηση

Χρησιμοποιώντας components, γίνεται πιο εύκολη η συντήρηση του κώδικα. Αν παραδείγματος χάρη κάποιος προγραμματιστής φτιάξει ένα component που υλοποιεί κάποιους επιχειρησιακούς κανόνες και οι κανόνες αυτοί κάποια στιγμή αλλάξουν, αρκεί η ενημέρωση αυτού του συγκεκριμένου component. Η υπόλοιπη εφαρμογή μένει άθικτη.

- Ευκολότερη ανάπτυξη

Διαχωρίζοντας τον κώδικα μιας εφαρμογής σε components, δίνει την δυνατότητα ανάπτυξης και ελέγχου (testing) μικρών, ανεξάρτητων κομματιών της εφαρμογής. Αυτό έχει σαν αποτέλεσμα και την διευκόλυνση των προγραμματιστών αλλά και την ευκολότερη συνεργασία μεταξύ τους.

- Απεξάρτηση από συγκεκριμένες γλώσσες προγραμματισμού

Σε οποιαδήποτε γλώσσα κι αν γραφτεί ένα component, διατηρεί την ικανότητα να επικοινωνεί με άλλα components, που ίσως είναι γραμμένα σε διαφορετική γλώσσα.

7.5 Αρχιτεκτονική Πληροφοριών (Information Architecture)

Το Ινστιτούτο Αρχιτεκτονικής Πληροφοριών (Information Architecture Institute) είναι ένας διεθνής οργανισμός που έχει σαν σκοπό να αναβαθμίζει και να προωθεί την πρακτική της Αρχιτεκτονικής των Πληροφοριών (Information Architecture).

Το Ινστιτούτο Αρχιτεκτονικής Πληροφοριών (Information Architecture Institute) ορίζει την Αρχιτεκτονική της Πληροφορίας σαν την τέχνη και επιστήμη της οργάνωσης και κατηγοριοποίησης των ιστοσελίδων, εσωτερικών διαδικτύων, απευθείας διασύνδεσης κοινοτήτων και λογισμικού για την ενίσχυση της χρηστικότητας.

Καθώς οι πληροφορίες πολλαπλασιάζονται εκθετικά με γεωμετρική πρόοδο η χρηστικότητα γίνεται ένας καθοριστικός παράγοντας επιτυχίας για τις ιστοσελίδες και εφαρμογές λογισμικού. Καλή αρχιτεκτονική της πληροφορίας τοποθετεί τα κατάλληλα θεμέλια για ένα πληροφοριακό σύστημα που είναι κατανοητό από τους χρήστες.

α) Γιατί χρησιμοποιούμε την Αρχιτεκτονική Πληροφοριών;

Σήμερα, κάθε επιχείρηση ή οργανισμός έχει ένα πρόβλημα στο πληροφοριακό της σύστημα. Είτε προβληματίζεται με το πως θα οργανώσει ένα τεράστιο αριθμό δεδομένων, είτε πως θα ταξινομήσει ένα νέο προϊόν, η αρχιτεκτονική πληροφοριών αποτελεί ένα σημαντικό μέρος της λύσης. Μερικές επιχειρήσεις έχουν ξεκάθαρα ορίσει θέσεις για Αρχιτέκτονες της Πληροφορίας (Information Architects), που έχουν ηγετικό ρόλο στην ανάπτυξη των διαδικασιών για τα πάντα από αρχειοθέτηση συστημάτων μέχρι την παραγωγή αρχιτεκτονικής. Άλλες επιχειρήσεις δεν έχουν επίσημες θέσεις, αλλά παρόλα αυτά υπολογίζουν τους Αρχιτέκτονες της Πληροφορίας σαν ένα κρίσιμο παράγοντα ανταγωνιστικότητας για πολλούς, όπως τους στρατηλάτες του διαδικτύου, σχεδιαστές διάδρασης (interaction designers) και αρχιτέκτονες της γνώσης. Και στους δυο τύπους επιχειρήσεων ή οργανισμών, οι καλύτερες πρακτικές στην αρχιτεκτονική

πληροφορίας υποστηρίζουν την ανάπτυξη διασυνδέσεων που βοηθούν την ροή χρήσιμων και σχετικών πληροφοριών για τον χρήστη.

Η επιχείρηση ή οργανισμός χρειάζεται αρχιτεκτονική της πληροφορίας όταν:

- Οι επιχειρησιακοί στόχοι υπαγορεύουν σχεδιασμό ή δηλώνουν ενδεικτικά επανασχεδιασμό μιας διεπιφάνειας χρήστη (user interface) ή ιστοσελίδας.
- Μη πρόσβαση πληροφοριών στους πελάτες και υπαλλήλους αυξάνει τα κόστη, συμπεριλαμβανομένου του τηλεφωνικού κέντρου και του γραφείου βοήθειας.
- Οι πρωτοβουλίες διαχείρισης των γνώσεων (knowledge management initiatives) μεταφέρονται από τις επιφάνειες εργασίας (desktop) σε έναν κεντρικό φάκελο του συστήματος ή σε ένα εσωτερικό δίκτυο.

7.6 Σχεδιασμός Σελίδων Ιστού (Web Design)

Ο σχεδιασμός στο Διαδίκτυο αφορά την ικανότητα δημιουργίας παρουσιάσεων περιεχομένου (συνήθως κειμένου ή πολυμέσων) οι οποίες φτάνουν στον τελικό-χρήστη μέσω του Παγκόσμιου Ιστού, με τη χρήση λογισμικού όπως ένας φυλλομετρητής (web browser) ή άλλου λογισμικού σχεδιασμένου για το διαδίκτυο όπως η τηλεόραση μέσω διαδικτύου, κινητών τηλεφώνων κλπ.

Η ιστοσελίδα είναι ένα ηλεκτρονικό αρχείο ή ένα σύνολο από ηλεκτρονικά αρχεία που υπάρχουν σε έναν ή και περισσότερους εξυπηρετητές (server/servers) και παρουσιάζει κείμενα και εφαρμογές πολυμέσων στον τελικό-χρήστη. Τέτοια στοιχεία όπως κείμενο, εικόνες (σύμφωνα με τα πρότυπα SVG, BMP, GIF, JPEG ή PNG) και φόρμες μπορούν να τοποθετηθούν στη σελίδα με τη χρήση γλωσσών σήμανσης υπερκειμένου όπως HTML/ΧHTML/XML. Η αναπαραγωγή πιο σύνθετων πολυμέσων (ανυσματικών γραφικών, βίντεο, ήχων, γραφικών με ενσωματωμένο ήχο και εικόνα) μπορεί να γίνει με πρόσθετα (plug-ins) όπως το Flash, το QuickTime, το περιβάλλον χρόνου εκτέλεσης Java, κ.α ή με τεχνολογίες όπως οι γλώσσες σήμανσης (X) HTML5, XML και MathML, και το πρότυπο φύλλων στυλ CSS 3.

Οι καινούριες εκδόσεις των προγραμμάτων περιήγησης (Internet Explorer 7, 8, Firefox 3.6, safari, chromium κλπ) που ακολουθούν τα W3C πρότυπα οδήγησαν σε μια ευρεία αποδοχή και χρήση των XHTML/XML σε συνδυασμό με τα CSS (Cascading Style Sheets) για την τοποθέτηση και διαχείριση των στοιχείων και αντικειμένων της ιστοσελίδας. Τα τελευταία πρότυπα στοχεύουν στο να αποκτήσουν τα προγράμματα περιήγησης την δυνατότητα να προσφέρουν μια ευρεία γκάμα επιλογής πολυμέσων και πρόσβασης στους πελάτες χωρίς τη χρήση των προσθέτων (plug-ins).

Γενικότερα οι ιστοσελίδες διαχωρίζονται σε στατικές και δυναμικές:

- Οι στατικές δεν αλλάζουν περιεχόμενο και διάταξη (layout) με οποιοδήποτε αίτημα εκτός και αν ο προγραμματιστής αναβαθμίσει (update) τη σελίδα. Μια απλή HTML σελίδα είναι παράδειγμα στατικού περιεχομένου.
- Οι δυναμικές προσαρμόζουν το περιεχόμενο και/ή την εμφάνισή τους σύμφωνα με την καταχώρηση/αλληλεπίδραση ή τις αλλαγές του τελικού χρήστη στο περιβάλλον προγραμματισμού (χρήστης, ώρα, τροποποιήσεις στη βάση δεδομένων, κτλ) Το περιεχόμενο μπορεί να αλλάζει στον υπολογιστή του τελικού-χρήστη με τη χρήση των γλωσσών προγραμματισμού που εκτελούνται στον υπολογιστή του χρήστη (JavaScript, VBScript, Actionscript, etc.). Το περιεχόμενο στις δυναμικές σελίδες συχνά μεταφράζεται στον εξυπηρετητή (server) μέσω γλωσσών προγραμματισμού που εκτελούνται στον εξυπηρετητή (Perl, PHP, ASP, JSP, ColdFusion, .NET κτλ).

Με την συνεχόμενη ειδίκευση στην τεχνολογία της Πληροφορικής δημιουργείται η ανάγκη διαχωρισμού της Σχεδιασμού Ιστοσελίδων από την Προγραμματιστή Ιστοσελίδων.

Για τη διαδικασία σχεδιασμού μιας ιστοσελίδας, μιας εφαρμογής ή ενός πολυμέσου για το διαδίκτυο μπορεί να συνδυάζονται πολλοί κλάδοι όπως animation, συγγραφή, επικοινωνιακός σχεδιασμός, εταιρική ταυτότητα, σχεδιασμός γραφι-

κών, αλληλεπίδραση ανθρώπου-υπολογιστή, αρχιτεκτονική υπολογιστών, σχεδιασμός αλληλεπίδρασης, marketing, φωτογραφία, βελτιστοποίηση μηχανών αναζήτησης και τυπογραφία.

Τα εργαλεία και οι τεχνολογίες που χρησιμοποιούνται περιλαμβάνουν μεταξύ άλλων:

- Γλώσσες σήμανσης: HTML, XHTML, XML
- Γλώσσες φύλλων φυλλομετρητή: CSS, XML
- Σενάρια στον πελάτη (Client-side scripting): JavaScript
- Σενάρια στον εξυπηρετητή (Server-side scripting): PHP, ASP, Perl, Ruby/Ruby on Rails
- Τεχνολογίες βάσεων δεδομένων: MySQL, PostgreSQL, Microsoft SQL Server
- Τεχνολογίες πολυμέσων: Flash, Silverlight

Οι ιστοσελίδες μπορούν να είναι στατικές ή δυναμικές. που αναπροσαρμόζουν αυτόματα το περιεχόμενό τους ή την εμφάνισή τους βασιζόμενες σε μια ποικιλία παραγόντων, όπως μια καταχώρηση από τον τελικό-χρήστη, μια καταχώρηση ή αλλαγή στο περιβάλλον προγραμματισμού από τον χειριστή της ιστοσελίδας (όπως εν παραδείγματι μια μορφοποίηση της βάσης δεδομένων).

Εξαιτίας της συνεχούς εξειδίκευσης στους τομείς της επικοινωνιακής Πληροφορικής, δημιουργείται μια σαφής τάση διαχωρισμού του σχεδιασμού στο Διαδίκτυο από την ανάπτυξη του διαδικτύου ως μέσου ροής της πληροφορίας και των αγαθών προς όλες τις διαδικτυακές υπηρεσίες.

8.0 Σχεδιασμός Διεπαφής Χρήστη (User Interface Design)

Ο σχεδιασμός διεπαφής χρήστη ή μηχανική διεπαφή χρήστη είναι ο σχεδιασμός των ιστοσελίδων, ηλεκτρονικών υπολογιστών, συσκευών, μηχανημάτων, συσκευές κινητών επικοινωνιών και εφαρμογών λογισμικού με την εστίαση στην εμπειρία και την αλληλεπίδραση του χρήστη.

Ο στόχος του σχεδιασμού της διεπαφής χρήστη είναι να κάνει την αλληλεπίδραση του χρήστη με απλό και αποτελεσματικό τρόπο, όσον αφορά την επίτευξη των στόχων των χρηστών, αυτό που συχνά αποκαλείται χρηστοκεντρική σχεδίαση. Ο καλός σχεδιασμός διεπαφής χρήστη διευκολύνει την ολοκλήρωση του έργου. Το Graphic design μπορεί να χρησιμοποιηθεί για να υποστηρίξει τη χρησιμότητά του. Η διαδικασία σχεδιασμού πρέπει να είναι ισοσκελισμένος τεχνική λειτουργικότητα και τα οπτικά στοιχεία (π.χ., νοητικό μοντέλο) για να δημιουργήσει ένα σύστημα που δεν είναι μόνο λειτουργική, αλλά μπορεί επίσης να χρησιμοποιηθεί και προσαρμόσιμα στις μεταβαλλόμενες ανάγκες των χρηστών.

Ο σχεδιασμό διεπαφής χρήστη απαιτεί μια καλή κατανόηση των αναγκών των χρηστών. Υπάρχουν αρκετές φάσεις και διεργασίες στο σχεδιασμό διεπαφής χρήστη, ανάλογα με το έργο²¹

Προϋπόθεση για τη σχεδίαση μιας επιτυχημένης διεπαφής είναι το να έχει γίνει επακριβώς κατανοητός ο στόχος που πρόκειται να εξυπηρετήσει το λογισμικό και οι απαιτήσεις χρηστικότητας του χρήστη από αυτό. Ο καλύτερος τρόπος για να απαντηθεί η ερώτηση «τι θέλει ο χρήστης να κάνει με αυτή την εφαρμογή;», είναι να μελετηθεί στην πράξη η εργασία του χρήστη. Η διαδικασία ανάπτυξης πολλαπλών σεναρίων, διαφορετικών μεταξύ τους, που περιγράφει το καθένα ξεχωριστούς τρόπους αλληλεπίδρασης και επικοινωνίας με το χρήστη για την υλοποίηση της εφαρμογής και στη συνέχεια η εξαγωγή συμπεράσματος για ποιο από αυτά θα προωθήσει καλύτερα τους σκοπούς της εργασίας, πάντοτε από

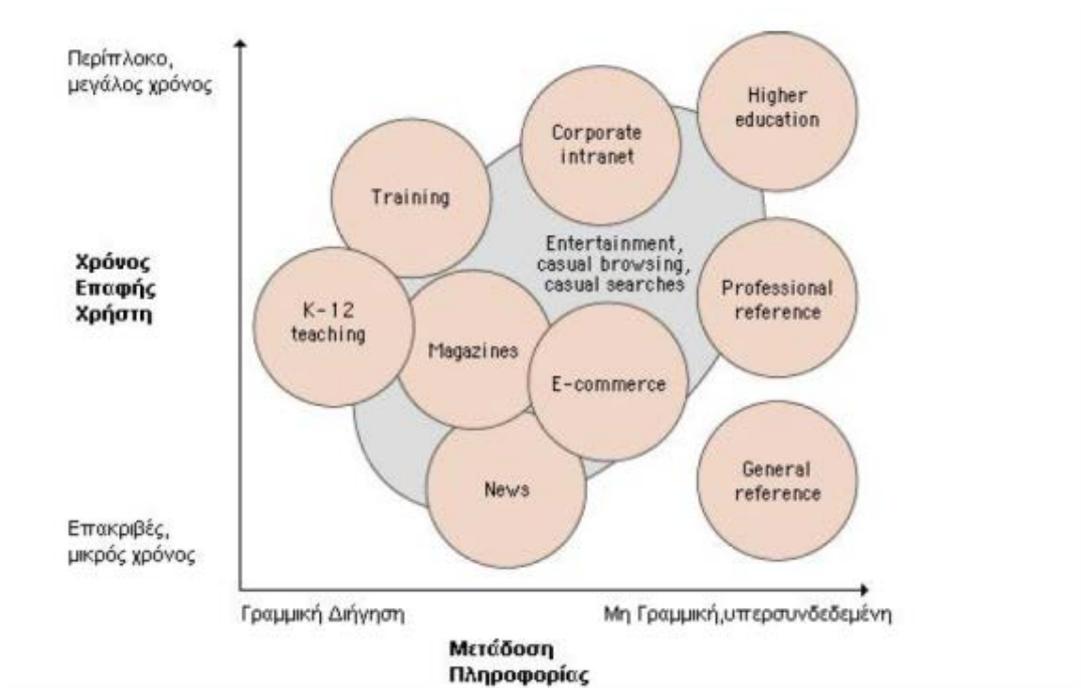
²¹ <http://web.inxpo.com/casting-calls/bid/105506/6-Tips-for-Designing-an-Optimal-User-Interface-for-Your-Digital-Event>

την πλευρά του χρήστη, είναι μια διαδικασία η οποία θα πρέπει να γίνει πριν από την έναρξη σχεδίασης της διεπαφής χρήστη και η οποία ονομάζεται ανάλυση εργασιών χρήστη (user task analysis).

Τέλος η επικοινωνία μιας εφαρμογής με το χρήστη είναι τόσο σημαντική, όσο και το ίδιο το περιεχόμενο της εφαρμογής. Χωρίς μια καλή διεπαφή χρήστη, πιθανά ο χρήστης δεν θα μπορέσει ποτέ να πάρει το πλήρες περιεχόμενο της εφαρμογής.

8.1 Γενικότερα Θέματα Σχεδιασμού Ιστοχώρων

Όλες οι παρουσιάσεις πληροφορίας κυβερνώνται από παραμέτρους ορισμένες από τους σκοπούς, την πρακτική λογική του επιλεγθέντος μέσου και το κοινό. Το παρακάτω σχήμα απεικονίζει τα επικρατέστερα θέματα μετάδοσης πληροφορίας σε σχέση με δύο βασικές μεταβλητές- τη γραμμικότητα της δομής της παρουσίασης και τον χρόνο επαφής με το μέσο χρήστη:



Προχωρώντας στο σχήμα αυτό διαγώνια προς τα πάνω δεξιά, μειώνεται η αναμενόμενη «αίσθηση» του Ιστοχώρου, δηλαδή τα στοιχεία εντυπωσιασμού κλπ και αυξάνεται η πληροφορία που περιέχεται και που απαιτεί ο χρήστης. Η εφαρμογή που παρουσιάζεται εμπίπτει στην κατηγορία "Reference"- αναφορά. Τα καλύτερα σχεδιασμένα sites αναφοράς επιτρέπουν στον χρήστη να μπει γρήγορα, να βρει αυτό που θέλει και ενδεχομένως να το κατεβάσει ή να το εκτυπώσει. Ουσιαστικά δεν υπάρχει κάποια «ιστορία» προς διήγηση, οπότε τα μοτίβα δεν είναι γραμμικά για την γενικότερη παρουσίαση του εργαλείου. Τα γραφικά είναι αφαιρετικά, ώστε να μην αποσπούν την προσοχή και το περιεχόμενο και η δομή των menu επιλογών προσεκτικά οργανωμένα ώστε να υποστηρίζουν γρήγορη αναζήτηση και ανάκληση δεδομένων. Η επιλογή της σύνθετης αναζήτησης ικανοποιεί τους πιο έμπειρους και σχετικούς με το αντικείμενο χρήστες,

ενώ μέσω της επιλογής σύνθετη αναζήτηση για την εισαγωγή παραμέτρων διατηρείται η επαφή του χρήστη με τον Ιστοχώρο χρονικά όσο «πρέπει», όπου ισχύει ο (εμπειρικός) κανόνας ότι όσο το λιγότερο τόσο το καλύτερο για εκείνον.

8.2 Κανόνες Design του Ιστοχώρου

- Το απλό είναι όμορφο

Η σελίδα που έχει πολλά πράγματα τα οποία δεν χρειάζονται δημιουργεί σύγχυση. Οι επισκέπτες είναι απογοητευμένοι όταν πρέπει να ανιχνεύσουν μέσω από πολλούς συνδέσμους και εικόνες για να βρουν αυτό που ψάχνουν. Διατηρώντας τις σελίδες απλές, η ιστοσελίδα θα είναι πιο εύκολη στη χρήση.

- Σχεδιασμός είναι «υψίστης σημασίας»

Για παράδειγμα όταν συναντάμε κάποιον για πρώτη φορά, θέλουμε να κάνουμε μια καλή πρώτη εντύπωση. Το ίδιο θα πρέπει να ισχύει και για την ιστοσελίδα. Η συνολική εμφάνιση και την αισθητική του site είναι το πρώτο πράγμα που οι επισκέπτες θα παρατηρήσουν.

- Η πλοήγηση θα πρέπει να είναι διαισθητική

Υπάρχουν μερικά πράγματα που είναι απογοητευτικά σε κάποιες σελίδες και οι περισσότεροι χρήστες αδυνατούν να τα ψάξουν. Οι σελίδες πρέπει να είναι καλά οργανωμένες με ένα σχεδιασμό top-down έτσι ώστε οι επισκέπτες μπορούν εύκολα να περιηγηθείτε μέσα από τις διάφορες ενότητες στην ιστοσελίδα.

- Η συνέπεια είναι το κλειδί

Οι επισκέπτες δεν θα πρέπει να αισθάνονται σαν να είναι η επίσκεψη σε ένα νέο δικτυακό τόπο κάθε φορά που ανοίγουν μια νέα σελίδα στο site. Συνοχή μεταξύ των σελίδων του site κάνει την πλοήγηση πολύ πιο εύκολο έργο.

- Τα χρώματα είναι ζωτικής σημασίας.

Κατά την επιλογή των χρωμάτων, χρησιμοποιούμε μια συνεπή παλέτα των χρωμάτων που δεν έρχονται σε «σύγκρουση» και να βεβαιωθούμε ότι υπάρχει μια έντονη αντίθεση μεταξύ του κειμένου και του φόντου.

- Ανάπτυξη για πολλαπλά προγράμματα περιήγησης.

Οι Browsers υποτίθεται ότι πρέπει να καθιστούν τις ιστοσελίδες με τον ίδιο τρόπο, αλλά δεν το κάνουν. Ως εκ τούτου, φροντίζουμε να ελέγχουμε την ιστο-

σελίδα σε πολλαπλά προγράμματα περιήγησης για να βεβαιωθούμε ότι τα πάντα φαίνεται σωστά. Το καλύτερο είναι να εντοπισμός των ενδεχόμενων προβλημάτων πριν από την ώρα, αντί να βασίζεται σε κατηγορίες του επισκέπτη.

- Έλεγχος της ιστοσελίδας για σφάλματα.

Όπως κάθε έμπειρος συντάκτης , ένα μεγάλο κομμάτι της εργασίας μπορεί να αμαυρωθεί από ένα μικρό απλό λάθος. Οι Webmasters θα πρέπει να ελέγχουν τακτικά τους δικτυακούς τόπους για τυπογραφικά λάθη, σπασμένα links, και εικόνες που δεν έχουν φορτώσει σωστά.

- Το περιεχόμενο.

Ακόμη και αν το site είναι όμορφα σχεδιασμένα, είναι ένα άδειο κέλυφος χωρίς περιεχόμενο. Μια καλή ιστοσελίδα έχει τόσο καλό σχεδιασμό και το μεγάλο περιεχόμενο. Ως εκ τούτου, βεβαιωνόμαστε ότι όλες οι σελίδες έχουν μοναδικό, πρωτότυπο περιεχόμενο, που τους κάνει να αξίζει να τις επισκεφθούμε.

8.3 Συμπεράσματα

Όποια και αν είναι τα Διαδικτυακά εργαλεία και τεχνικές διαχείρισης Πληροφοριών, ίσως , ασχολούνται με τη διάδοση και την ανάπτυξη της τεχνολογίας . Για την επίτευξη του στόχου πολλά εργαλεία και τεχνικές , τόσο συμβατικών και μη συμβατικών, έχουν αναπτυχθεί. Αυτές περιλαμβάνουν εργαλεία και τεχνικές , όπως , Πληροφορική , Πρότυπα , Ανάλυση Συστημάτων , Τεχνικές Ανάλυσης εργασίας, τεχνικές παρακολούθησης , τεχνικές αξιολόγησης , οι συμβατικές τεχνικές ταξινόμησης. Τα ο σύνολο της παρτίδας των τεχνικών διαχείρισης φαίνεται να ισχύει για τη διαχείριση των πληροφοριών . Δεδομένου ότι οι τεχνικές αυτές έχουν συζητηθεί ευρέως είναι καλά γνωστές στους επαγγελματίες της πληροφόρησης .

9.0 Βιβλιογραφία

1. *BS 3525*. s.l.:British Standards.
2. *Library Learning*. [Ηλεκτρονικό]
Available
at: http://www.usg.edu/galileo/skills/unit04/primer04_01.phtml
[Πρόσβαση 13 10 2013].
3. Anon., 2013. *CrawlTrack, your web statistics tool*. [Ηλεκτρονικό]
Available at: <http://www.crawltrack.net/>
[Πρόσβαση 15 10 2013].
4. Arampatzis A., v. d. W. T. v. B. P. K., 2000. *Linguistically-Motivated Information Retri*. New York: MarcelDekker.
5. Baeza-Yates, R. C. C. M. M. a. R. A., 2005. *Crawling a Country: Better Strategies than Breadth-First for Web Page Ordering*. In *Proceedings of the Industrial and Practical Experience track of the 14th conference on World Wide Web, pages*. s.l.:ACM Press.
6. Behlendorf, B., 2012. *What is the Apache HTTP Server* , s.l.: s.n.
7. Bharat, R. C. M. a. K., 1998. *SPHINX: A framework for creating personal, site-specific web crawlers*, s.l.: s.n.
8. Castillo, C., 2004. *Effective Web Crawling*. s.l.:s.n.
9. Ceri, S. P. F. A. B. M. B. S. C. a. M. M., 2003 . *Designing Data-Intensive Web Applications*. Morgan Kauffmann επιμ. Usa: s.n.
10. Chakrabarti, S., 2003. *Mining the Web*. s.l.:Morgan Kaufmann Publishers.
11. Fry, S., 2009. *Educastur*. [Ηλεκτρονικό]
Available at: <http://web.educastur.princast.es>
[Πρόσβαση 17 10 2013].
12. Gerard Salton, C. B., 1988. *Term-weighting approaches in automatic text retrieval* , Usa: s.n.
13. Gray, J., 2009. *A Transformed Scientific Method*. Mountain View, California: Hey, Tony.
14. Hersh W., T. A. P. S. C. B. K. D., n.d. s.l.:s.n.
15. Kobayashi, M. a. T. K., 2000. *Information retrieval on the web*. s.l.:ACM Press.
16. Kramer, D., 1999. *Flashback*. [Ηλεκτρονικό]
Available

- at: http://www.microsoft.com/misc/features/features_flsbk.htm
[Πρόσβαση 18 10 2013].
17. M, C. H. R., 1996. *Search Engines for the World Wide Web: A Comparative Study and Evaluation*. s.l., s.n.
 18. Mehta, V., 2006. *Codeguru*. [Ηλεκτρονικό]
Available
at: http://www.codeguru.com/csharp/.net/net_general/visualstudionet_add-ins/article.php/c11835/Extending-Visual-Studio-2005.htm
[Πρόσβαση 18 10 2013].
 19. Microsoft, 2005. *Microsoft*. [Ηλεκτρονικό]
Available at: <http://msdn.microsoft.com/en-us/library/bb165114%28VS.80%29.aspx>
[Πρόσβαση 18 10 2013].
 20. Microsoft, 2005. *Microsoft*. [Ηλεκτρονικό]
Available at: [http://msdn.microsoft.com/en-us/library/bb166391\(VS.80\).aspx](http://msdn.microsoft.com/en-us/library/bb166391(VS.80).aspx)
[Πρόσβαση 19 10 2013].
 21. Microsoft, 2010. *Microsoft*. [Ηλεκτρονικό]
Available at: <http://msdn.microsoft.com/en-US/library/cc160717%28v=vs.90%29.aspx>
[Πρόσβαση 17 10 2013].
 22. Nelson, M. L., 2005. *mod_oai: An Apache Module for Metadata Harvesting*. s.l.:s.n.
 23. O'Reilly, T., 2005. *O'Reilly*. [Ηλεκτρονικό]
Available at: <http://oreilly.com/web2/archive/what-is-web-20.html>
[Πρόσβαση 17 10 2013].
 24. Shestakov, D., 2008. *Search Interfaces on the Web*. s.l.:s.n.
 25. Suel, V. a., 2008. *Design and Implementation of a High-Performance*.
New York: s.n.
 26. Tanenbaum, A. S., 2002. *Computer networks*. Usa: Upper Saddle River.
 27. Tzoukermann., H. J. a. E., 1999. *Information retrieval based on context distance and*. s.l.:University of Berkeley.
 28. Ανδρέας, Β., 2004. *Έρευνα και συλλογή πληροφοριών στο διαδίκτυο*.
Αθήνα: Τζόλας.