

**ΑΛΕΞΑΝΔΡΕΙΟ ΤΕΧΝΟΛΟΓΙΚΟ
ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ**

**ΤΜΗΜΑ ΒΙΒΛΙΟΘΗΚΟΝΟΜΙΑΣ ΚΑΙ
ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΗΣΗΣ**

ΘΕΜΑ: ΤΟ ΣΗΜΑΣΙΟΛΟΓΙΚΟ WEB'

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Ρουσσίδα Χρήστου
Α.Μ. 020/04

**Επιβλέπων καθηγητής
Κεφαλάς Μιχάλης**

Θεσσαλονίκη 2010

ΠΡΟΛΟΓΟΣ

Η παρούσα πτυχιακή εργασία εκπονήθηκε από τον φοιτητή Ρουσσίδα Χρήστο στο ΠΤ' εξάμηνο του τμήματος Βιβλιοθηκονομίας και Συστημάτων πληροφόρησης του Αλεξάνδρειου Τεχνολογικού Εκπαιδευτικού Ιδρύματος Θεσσαλονίκης.

Αναλύει τις βασικές αρχές και την προοπτική του σημασιολογικού ιστού που αποτελεί το μέλλον στα δικτυακά και πληροφοριακά τεκταινόμενα.

Αποτελείται από 5 κεφάλαια τα οποία περιλαμβάνουν:

- Εισαγωγή – Ιστορική αναδρομή – Τρέχουσα κατάσταση του Διαδικτύου
- Κύριο Μέρος – Παρουσίαση των δυνατοτήτων και τεχνολογιών του σημασιολογικού web
- Εργαλεία του Σημασιολογικού Ιστού
- Επίλογος – Σύνοψη
- Βιβλιογραφία - Γλωσσάριο

PROLOGUE

The present diplomatic dissertation was elaborated by the student Roussidis Christos at the final course studying year of the department of Library Science and Information Systems of Alexandreio Technological Institution of Thessaloniki.

The paper analyses the basic principles and the prospect of the semantic web, which represents the future in network and computer science.

It consists of 5 chapters which include:

- Introduction – Historic Throwback – Current Internet Status
- Main Part – Presentation of the abilities of the new web
- Semantic web construction tools
- Conclusion – Synopsis
- Bibliography - Vocabulary

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω αρχικά την οικογένεια μου και ιδιαίτερα την αγαπημένη μου μητέρα που ήταν δίπλα μου σε αυτή μου την προσπάθεια. Επίσης, ένα ευχαριστώ στους φίλους μου και φυσικά στον επόπτη καθηγητή μου κ. Μιχάλη Κεφαλά για τις πολύτιμες συμβουλές του, την καθοδήγηση του και την υπομονή του.

ΠΕΡΙΕΧΟΜΕΝΑ

Κεφάλαιο 1

Εισαγωγή – Ιστορική αναδρομή – Τρέχουσα κατάσταση του Διαδικτύου

- 1.1 – Ιστορική αναδρομή στο Internet και η δημιουργία του
- 1.2 – Το σημερινό web, οι παροχές, οι υπηρεσίες και οι αδυναμίες του
- 1.3 – Το όραμα του σημασιολογικού web

Κεφάλαιο 2

Κύριο Μέρος – Παρουσίαση των δυνατοτήτων και των τεχνολογιών του
Σημασιολογικού Ιστού

- 2.1 – Βασική παρουσίαση των δυνατοτήτων του σημασιολογικού web
 - 2.1.1 - Διαχείριση της γνώσης
 - 2.1.2 - Το ηλεκτρονικό εμπόριο μεταξύ επιχειρήσεων και χρηστών-καταναλωτών του Ιστού
 - 2.1.3 - Το ηλεκτρονικό εμπόριο μεταξύ επιχειρήσεων
 - 2.1.4 Τα συστήματα wiki
- 2.2 – Βασική παρουσίαση των τεχνολογιών του Σημασιολογικού Ιστού
 - 2.2.1 – Ρητά μεταδεδομένα
 - 2.2.2 – Οντολογίες
 - 2.2.3 – Λογική
 - 2.2.4 – Πράκτορες

Κεφάλαιο 3

Εργαλεία του Σημασιολογικού Ιστού

- 3.1.1 – Δεδομένα έγγραφα ιστού: XML
- 3.1.2 – Περιγραφή πόρων ιστού: RDF
- 3.1.3 – Γλώσσα οντολογιών ιστού: OWL

Κεφάλαιο 4

Επίλογος – Σύνοψη

Κεφάλαιο 5

Βιβλιογραφία – Γλωσσάριο

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

Ιστορική αναδρομή

Τρέχουσα κατάσταση του Διαδικτύου

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή – Ιστορική αναδρομή – Τρέχουσα κατάσταση του Διαδικτύου

1.1 Ιστορική αναδρομή στο Internet και η δημιουργία του

Το διαδίκτυο όπως το γνωρίζουμε σήμερα πέρασε από πολλές φάσεις δημιουργίας πριν καταλήξει στη σημερινή του μορφή και τις τρέχουσες δυνατότητες του προς τους χρήστες. Η ονομασία διαδίκτυο (Internet) σήμερα θεωρείται συνώνυμη του Παγκόσμιου Ιστού¹ (World Wide Web-WWW), πράγμα το οποίο όμως είναι λανθασμένο, καθώς ο Ιστός δεν είναι τίποτα άλλο παρά μια εφαρμογή του διαδικτύου, όπως για παράδειγμα είναι και το WWW2. Η βασική διαφορά μεταξύ τους είναι ότι το διαδίκτυο έχει και υλική υπόσταση, αντίθετα με τον ιστό, ο οποίος είναι στην ουσία πακέτα πληροφοριών.

Η αρχή του διαδικτύου έγινε το 1969 με την ονομασία ARPANET, δημιουργημένο από το Αμερικάνικο Υπουργείο Άμυνας, με κύριο στόχο στρατιωτικής μελέτης για την αξιόπιστη δημιουργία των δικτύων και της εσωτερικής ασφάλειας της χώρας. Η βασική αρχή ήταν η ανταλλαγή δεδομένων μεταξύ των ερευνητών και εκεί ήταν και η πρώτη δημιουργία του ηλεκτρονικού ταχυδρομείου.

Τέσσερα χρόνια μετά, και συγκεκριμένα το 1973, ένα νέο ερευνητικό πρόγραμμα ξεκινά με την ονομασία Internetworking Project, το οποίο στοχεύει στη διασύνδεση διάφορων δικτύων και όχι μόνο στρατιωτικής φύσης, για την ανταλλαγή δεδομένων μεταξύ των συμμετεχόντων δικτύων. Αποτέλεσμα αυτής της έρευνας ήταν η δημιουργία μιας νέας τεχνικής με την ονομασία Internet Protocol (IP), από το οποίο στη συνέχεια παίρνει το όνομα του το σημερινό μας internet. Οι υπολογιστές από διαφορετικά δίκτυα που χρησιμοποιούσαν το πρωτόκολλο IP, μπορούσαν να συνδέονται και να αποτελούν ένα διαδίκτυο. Σε ένα δίκτυο IP όλοι οι υπολογιστές είναι ισοδύναμοι, και έτσι δεν υπάρχουν περιορισμοί για το ποιος υπολογιστής θα

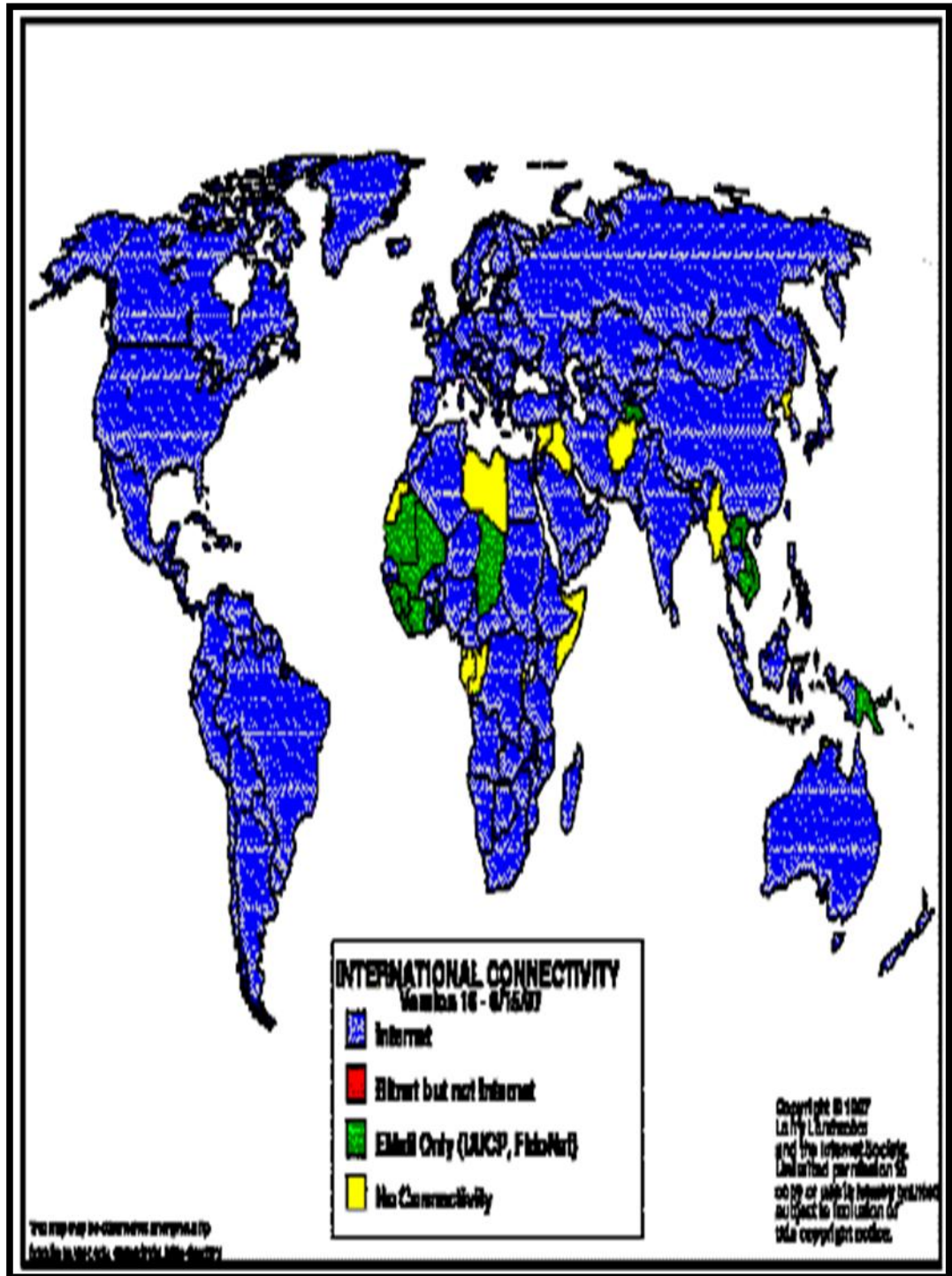
¹ Tim Burners-Lee. September 1998

Το σημασιολογικό web

επικοινωνεί με ποιον. Την ίδια χρονιά επίσης, σχεδιάστηκε ακόμα μια τεχνική για τον έλεγχο μετάδοσης δεδομένων η οποία χρησιμοποιείται και σήμερα, το Transmission Control Protocol (TCP), το οποίο ορίζει τις προδιαγραφές για τη μεταφορά αρχείων μεταξύ υπολογιστών.

Περνώντας στη δεκαετία του 80', το Υπουργείο Άμυνας των ΗΠΑ αναγνωρίζει πια το πρωτόκολλο με την ενιαία ονομασία TCP/IP και εκδίδει το λειτουργικό Berkeley UNIX, το οποίο περιλαμβάνει το TCP/IP και συμβάλλει στη ραγδαία εξάπλωση του ARPANET στα πανεπιστήμια της εποχής. Η τεράστια απήχηση του είχε ως αποτέλεσμα την επιβάρυνση του δικτύου, το οποίο αναγκαστικά χωρίστηκε σε δύο μέρη για να υποστηρίξει τη ζήτηση του. Έτσι λοιπόν γεννιούνται το MILNET για καθαρά στρατιωτικούς σκοπούς και το ARPANET για χρήση της πανεπιστημιακής κοινότητας. Από εκεί και έπειτα όλο και περισσότερα πανεπιστήμια δημιουργούν τα δικά τους δίκτυα και συνδέονται στο παγκόσμιο αυτό δίκτυο που αρχίζει να γίνεται γνωστό ως Internet, μέχρι που στο 1990 η ονομασία ARPANET καταργείται.

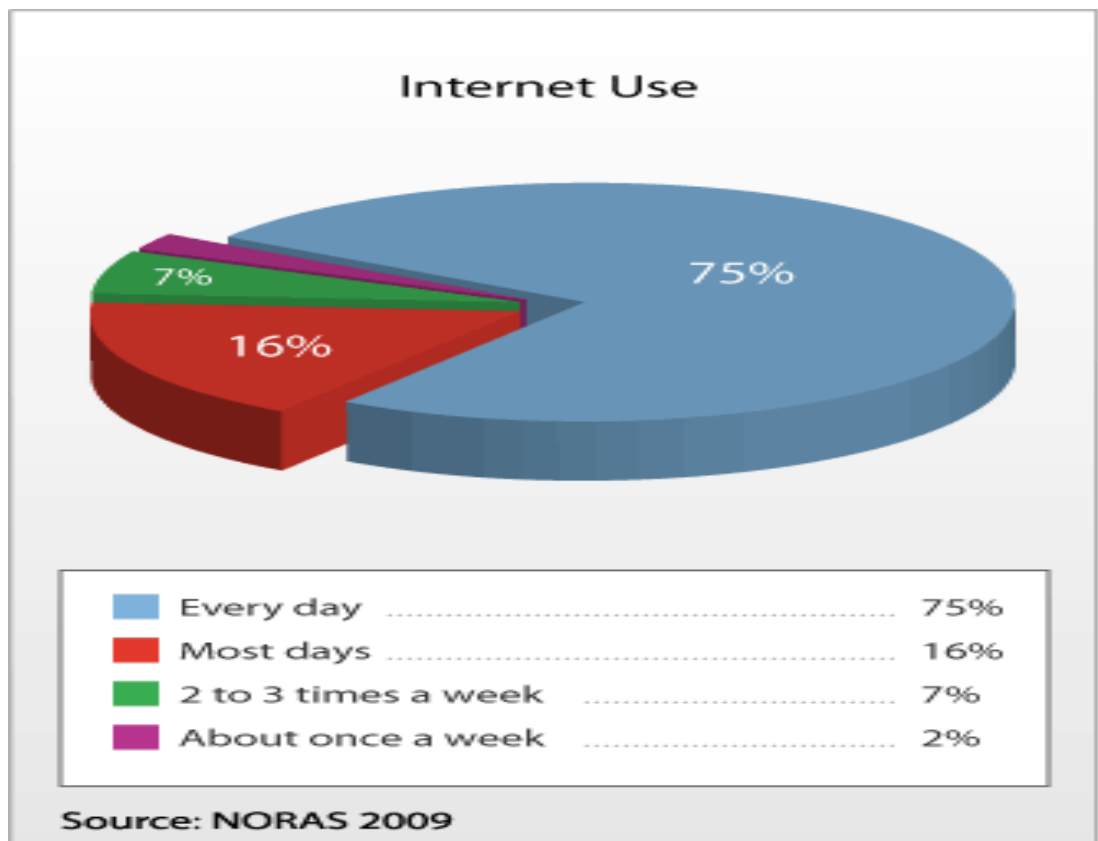
Με την είσοδο της νέας δεκαετίας, ο τότε προγραμματιστής στο Ευρωπαϊκό Κέντρο Φυσικής (CERN), Tim Burners-Lee, δημιουργεί το πρωτόκολλο http (hypertext transfer protocol), το οποίο είναι στην ουσία η γλώσσα επικοινωνίας μεταξύ των υπολογιστών, καθώς επίσης και το URI (Universal Resource Identifier), το οποίο είναι ένας τρόπος αναγνώρισης κάθε στοιχείου μέσα στο διαδίκτυο. Ο συνδυασμός των παραπάνω απέδωσε το σημερινό URL (Uniform Resource Locator), το οποίο δίνει τη μοναδικότητα κάθε συμμετέχοντα στο internet. Αφού έκανε όλα αυτά δημιούργησε και έναν περιηγητή για το δημιούργημα του και το ανέβασε δωρεάν στους servers της CERN. Η απήχηση ήταν τεράστια και από απλούς χρήστες αλλά και από κυβερνητικές οργανώσεις, στο σημείο που χρειάστηκε η δημιουργία νέου περιηγητή με την ονομασία Netscape Navigator και σε συνδυασμό με την αυξανόμενη απήχηση των windows και του γραφικού τους περιβάλλοντος έχουμε την ανάπτυξη του ιστού όπως τον γνωρίζουμε σήμερα.



Εικ.1 Χαρακτηριστική εξάπλωση του Web το 1997.
Με μπλε χρώμα απεικονίζονται οι περιοχές με πλήρη πρόσβαση στον ιστό ενώ με κίτρινο
αυτές που δεν έχουν καθόλου.

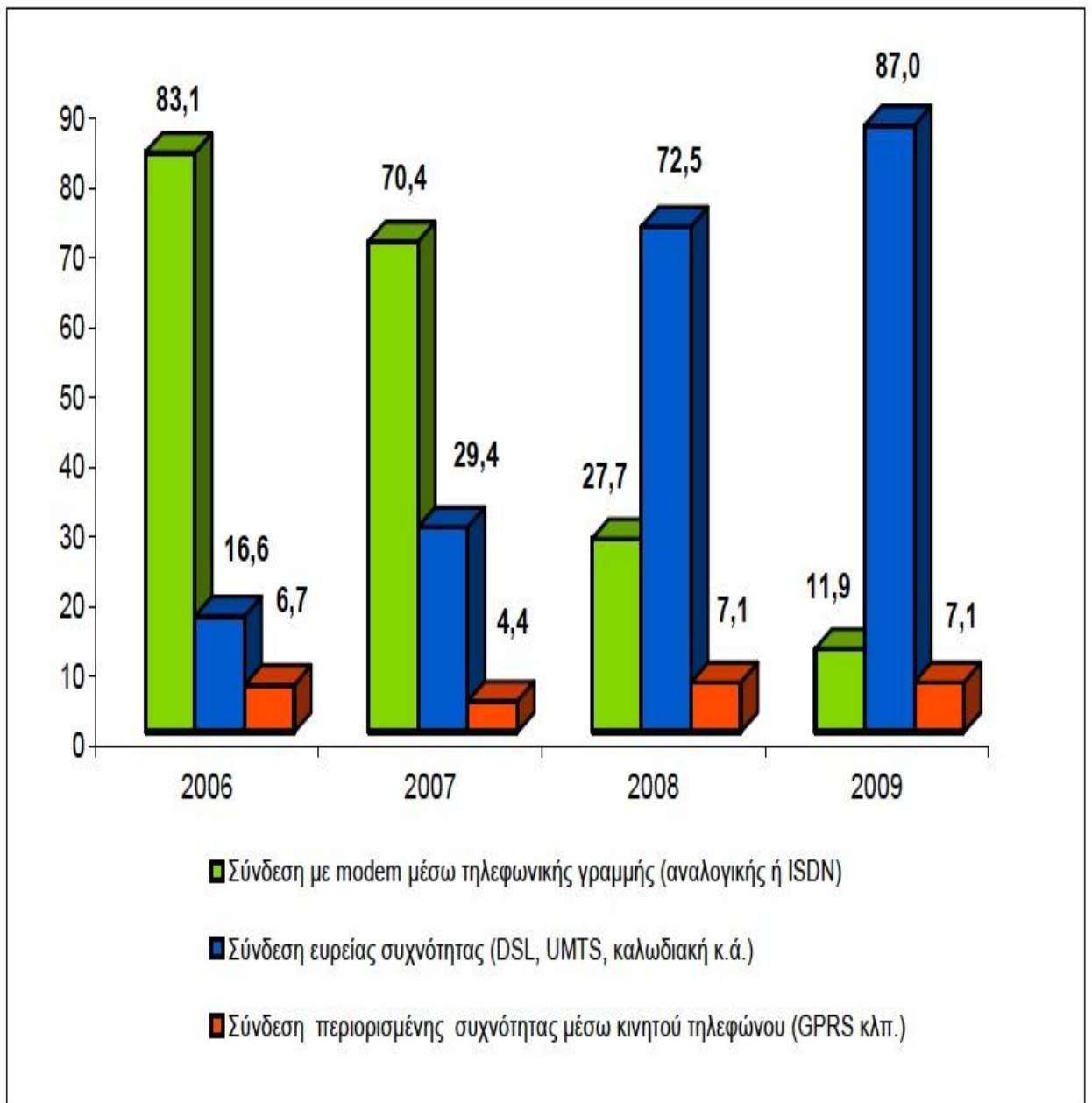
1.2 Το σημερινό web, οι παροχές, οι υπηρεσίες και οι αδυναμίες του

Φτάνοντας στη σημερινή εποχή, ο καθένας μας μπορεί να διαπιστώσει την τεράστια επίδραση του παγκόσμιου ιστού στην καθημερινότητα μας. Η χρήση του είναι καθημερινή και στις περισσότερες περιπτώσεις πολύωρη για διάφορους λόγους και με διάφορες χρήσεις. Ο παγκόσμιος ιστός σήμερα μας παρέχει επικοινωνία, κοινωνικότητα, διαπροσωπικές σχέσεις, επιχειρηματικές συναλλαγές, γνώση, πληροφόρηση, διασκέδαση και οτιδήποτε άλλο μπορεί να σκεφτεί κάποιος. Υπάρχει ατελείωτος όγκος πληροφοριών μέσα στο web είτε χρήσιμων είτε άχρηστων και λανθασμένων.



Εικ.2 Χρήση του broadband internet από δημοσκόπηση στην Μ. Βρετανία. Το 75% δηλώνει καθημερινή χρήση

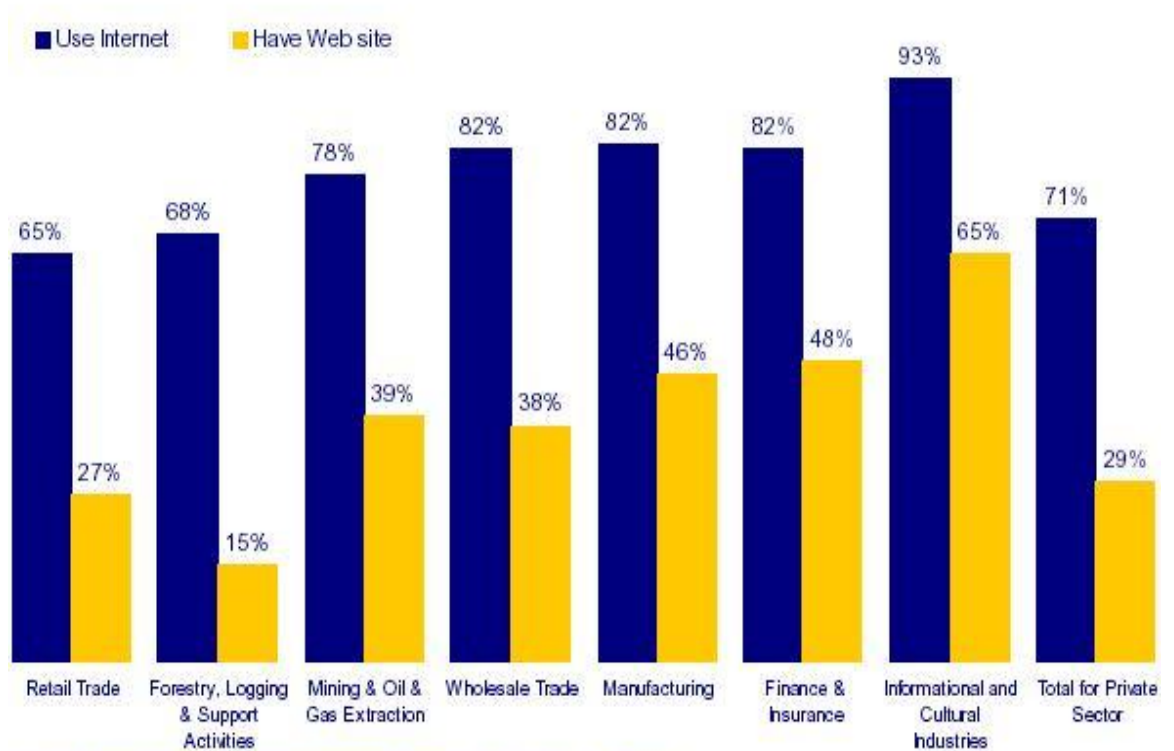
Ευρυζωνική σύνδεση στην κατοικία (2005 – 2009)



Εικ.3 Η διείσδυση των υπηρεσιών broadband internet στην Ελλάδα, η οποία αν και είναι τελευταία στην ευρωπαϊκή ζώνη δείχνει σημάδια βελτίωσης τα τελευταία χρόνια.

Internet Usage by Canadian Businesses Varies by Industry

Business Technology Use In The Canadian Private Sector by Industry, 2001



Source: Electronic Commerce and Technology Survey; Statistics Canada, April 2002.

Εικ.4 Χρήση του internet σε διάφορα πεδία βιομηχανιών στον Καναδά. Τεράστια νούμερα και διείσδυση

Οι υπηρεσίες του web στην ουσία δεν μπορούν να κατηγοριοποιηθούν, καθώς είναι ουσιαστικά μικρογραφία της πραγματικής ζωής. Ένας χρήστης του διαδικτύου μπορεί να αγοράσει πράγματα, να μιλήσει με συνανθρώπους του, να πληρώσει μέχρι και λογαριασμούς του. Κάποιες από τις υπηρεσίες του βέβαια είναι επί πληρωμή και κάποιες άλλες ελεύθερες προς όλους. Οι τυπικές χρήσεις βέβαια του σημερινού χρήστη του διαδικτύου είναι η αναζήτηση πληροφοριών, η αναζήτηση και επαφή με άλλους ανθρώπους και η παραγγελία προϊόντων μέσω του web. Οι παραπάνω δραστηριότητες υποστηρίζονται κυρίως από μηχανές αναζήτησης (search engines),

καθώς επίσης και από συνδέσμους (links) μέσω εγγραφών και ιστοσελίδων. Στο σημείο αυτό όμως εμφανίζεται και το κυριότερο μειονέκτημα και αδυναμία του σημερινού web: όλη η αρχιτεκτονική και ο σχεδιασμός του είναι ανθρωποκεντρική, δηλαδή τα πάντα είναι βασισμένα στο πως ο άνθρωπος αντιλαμβάνεται και χρησιμοποιεί τον ιστό και ο υπολογιστής εκτελεί τις εντολές του χρήστη χωρίς να κατανοεί τι του ζητείται. Έτσι λοιπόν παρατηρούνται οι παρακάτω αδυναμίες του σημερινού web:²

- Υψηλή ανάκληση – χαμηλή ακρίβεια: ακόμα και αν ανακτηθούν οι βασικές σελίδες με τα αποτελέσματα που αναζητά ο χρήστης, θα ανακτηθεί μαζί τους ένας τεράστιος αριθμός από λίγο σχετικά έως εντελώς άσχετα δεδομένα. Έτσι ο χρήστης βρίσκεται σε σύγχυση καθώς ο υπερβολικά μεγάλος αριθμός αποτελεσμάτων τον κάνει να σκέφτεται αν όντως διάλεξε τις σωστές πληροφορίες από τον τεράστιο σωρό.
- Χαμηλή ή καθόλου ανάκληση: είναι συχνό φαινόμενο να μην παίρνουμε κάποια σχετικά αποτελέσματα με την αναζήτηση μας ή να παίρνουμε άσχετες ιστοσελίδες. Παρόλο που είναι σπανιότερο το φαινόμενο, είναι ωστόσο υπαρκτό.
- Τα αποτελέσματα είναι ιδιαίτερα ευαίσθητα στο λεξιλόγιο που χρησιμοποιούμε: συχνά, οι αρχικές λέξεις-κλειδιά που χρησιμοποιούμε δεν επιστρέφουν τα επαρκή αποτελέσματα και σε αυτή την περίπτωση, η μηχανή αναζήτησης αλλοιώνει ελαφρώς τους όρους της αναζήτησης μας και μας επιστρέφει αποτελέσματα τα οποία όμως δεν έχουν συνάφεια με την αρχική μας αναζήτηση.
- Τα αποτελέσματα είναι μεμονωμένες ιστοσελίδες: αν χρειαζόμαστε για παράδειγμα αποτελέσματα από διάφορες ιστοσελίδες θα πρέπει να κάνουμε

² T. Burners-Lee, Hendler, J. and Lassila, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. 2001

Το σημασιολογικό web

πολλαπλές αναζητήσεις για να συλλέξουμε τις πληροφορίες και στη συνέχεια να ελέγξουμε τα αποτελέσματα μεμονωμένα και με μη αυτόματο τρόπο για να συνθέσουμε το τελικό μας αποτέλεσμα.

Από τα παραπάνω παρατηρούμε ότι παρά την βελτίωση των μηχανών αναζήτησης τα τελευταία χρόνια, τα προβλήματα παραμένουν στην ουσία τα ίδια. Ακόμα και αν μια αναζήτηση είναι επιτυχής, ο χρήστης είναι αυτός που θα πρέπει να ελέγξει τα επιλεγμένα αποτελέσματα για να εξάγει τις πληροφορίες που ψάχνει. Έτσι συμπεραίνουμε ότι δεν υπάρχει ιδιαίτερη υποστήριξη για την ανάκτηση πληροφοριών, η οποία είναι μια εξαιρετικά χρονοβόρα δραστηριότητα. Το βασικό εμπόδιο για την παροχή καλύτερης υποστήριξης προς τους χρήστες του ιστού είναι ότι το νόημα του ανακτώμενου περιεχομένου δεν είναι προς το παρόν προσπελάσιμο από τους υπολογιστές (machine accessible). Φυσικά υπάρχουν εργαλεία που βοηθούν στην ορθογραφία, στον διαχωρισμό των κειμένων σε τμήματα ή ακόμα και στην καταμέτρηση λέξεων, αλλά η διαδικασία που ακολουθούν είναι εντελώς αυτοματοποιημένη και χρειάζεται η επέμβαση του ανθρώπου για να γίνουν κατανοητά αυτά που αναγράφονται στην οθόνη.

Η παρούσα κατάσταση μπορεί να βελτιωθεί και υπάρχουν αυτή τη στιγμή δύο τρόποι. Ο πρώτος συνιστά την χρήση πολύπλοκων μεθόδων που βασίζονται στην Τεχνητή Νοημοσύνη (Artificial Intelligence) και την υπολογιστική γλωσσολογία (computational linguistics). Η μέθοδος αυτή ακολουθείται εδώ και αρκετό καιρό και παρόλο που υπάρχει κάποια πρόοδος, το όλο εγχείρημα μοιάζει δύσκολο να εξελιχθεί. Ο άλλος τρόπος είναι και η ανάπτυξη της επόμενης μορφής web με την ονομασία σημασιολογικός ιστός.

1.3 Το όραμα του σημασιολογικού web

Η βασική ιδέα του σημασιολογικού web είναι ότι τα δεδομένα θα είναι προσβάσιμα, αναγνωρίσιμα και επεξεργάσιμα σε μορφή που θα είναι κατανοητή από τους υπολογιστές (machine processable). Είναι σημαντικό να τονιστεί πως ο σημασιολογικός ιστός δεν θα είναι κάτι παράλληλο με τον παγκόσμιο ιστό, αλλά θα εξελιχθεί μέσα από τον ήδη υπάρχοντα ιστό. Έτσι λοιπόν συμπεραίνουμε ότι θα υπάρχει συνεργασία ανθρώπου και υπολογιστή και ο τελευταίος θα διευκολύνει τον χρήστη, καθώς θα είναι σε θέση να συνδυάζει και να κατανοεί το αίτημα του χρήστη, σώζοντας πολύτιμο χρόνο και προσπάθεια του τελευταίου. Για να πραγματοποιηθούν τα παραπάνω βέβαια χρειάζεται ακόμα πολύς χρόνος και προσπάθεια από τον δημιουργό του, Tim Burners-Lee και την Κοινοπραξία Παγκόσμιου Ιστού (World Wide Web Consortium, W3C).

Τα βασικά εργαλεία του νέου ιστού είναι τα δεδομένα έγγραφα ιστού (XML), η οποία είναι μια γλώσσα με λεξιλόγιο που ορίζεται από τον χρήστη και είναι ιδιαίτερα κατάλληλη για την αποστολή εγγράφων στον ιστό, η περιγραφή πόρων ιστού (RDF), η οποία είναι μια γλώσσα και ένα βασικό μοντέλο δεδομένων για τη συγγραφή απών προτάσεων σχετικά με αντικείμενα του ιστού, οι οντολογίες και η γλώσσα οντολογιών ιστού (OWL), η οποία είναι η τρέχουσα πρόταση για μια γλώσσα οντολογιών ιστού, εργαλεία τα οποία θα αναλυθούν αναλυτικότερα στα ακόλουθα κεφάλαια.

Το όραμα της συνεργασίας ανθρώπου- υπολογιστή, και η κατανόηση των εγγράφων του ιστού από τον τελευταίο με στόχο τη διαχείριση και τη διευκόλυνση των χρηστών του διαδικτύου είναι σαφώς φιλόδοξη και μένει να αποδειχθεί αν θα είναι ρεαλιστική.

Ο ιδρυτής του σημασιολογικού web Tim Burners-Lee, στην προσπάθειά του να αποδώσει γραφικά τη δομή του νέου ιστού δημιούργησε το λεγόμενο «πύργο των επιπέδων» του semantic web, όπως φαίνεται στην εικόνα που ακολουθεί.

Ξεκινώντας μια βασική ανάλυση των επιπέδων του πύργου, έχουμε το πρώτο επίπεδο που περιλαμβάνει τα URI και Unicode τα οποία παρέχουν πρότυπους τρόπους για τον ορισμό αναφορών σε οντότητες και στην ανταλλαγή συμβόλων.

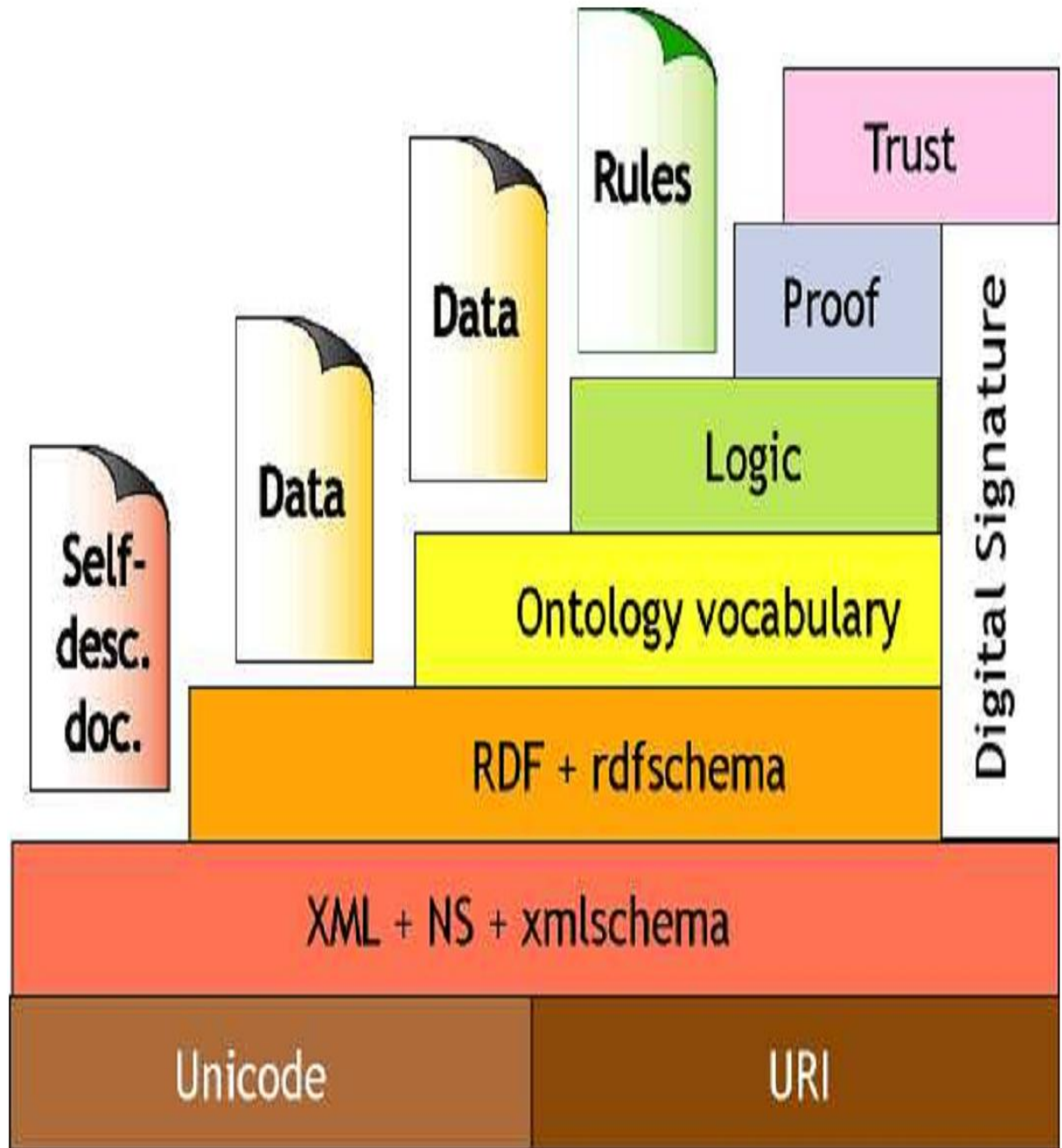
Στο δεύτερο επίπεδο έχουμε την XML. Η XML επιτρέπει τη συγγραφή δομημένων εγγράφων του ιστού με λεξιλόγιο ορισμένο από το χρήστη.

Το τρίτο επίπεδο του πύργου είναι το RDF και το RDF Schema το οποίο ορίζει ένα μοντέλο δεδομένων για την περιγραφή μηχανικά επεξεργασμένων σημασιολογικών δεδομένων. Επίσης, επιτρέπει τη συγγραφή απλών προτάσεων σχετικά με αντικείμενα του ιστού. Η RDF Schema παρέχει θεμελιώδη στοιχεία μοντελοποίησης για την οργάνωση των αντικειμένων του Ιστού σε ιεραρχίες.

Το τέταρτο επίπεδο περιέχει το λεξιλόγιο οντολογιών που είναι μία γλώσσα η οποία αναφέρεται σε ένα περιορισμένο υποσύνολο λογικής για την περιγραφή ορολογίας και αναπαράσταση πολυπλοκότερων σχέσεων μεταξύ των αντικειμένων του ιστού.

Στο πέμπτο επίπεδο συναντάμε την Λογική. Η λογική βρίσκεται πάνω από την οντολογία καθώς σε θεωρητικό επίπεδο θα πρέπει να μπορεί να προσφέρει μια πλουσιότερη γλώσσα λογικής. Τέλος χρησιμοποιείται για να ενισχύσει τη γλώσσα οντολογιών.

Τελειώνοντας, στα δυο τελευταία επίπεδα έχουμε την απόδειξη και την εμπιστοσύνη, τα οποία φαίνεται να είναι περισσότερο εφαρμογές παρά επίπεδα γλωσσών. Όπως και να χει οι υλοποιήσεις τους είναι ακόμα πολύ πρόωρες και μη ρεαλιστικές. Στην Απόδειξη θα περιλαμβάνεται η διαδικασία εξαγωγής συμπερασμάτων και η αναπαράσταση των αποδείξεων. Στην Εμπιστοσύνη θα χρησιμοποιούνται ψηφιακές υπογραφές από δικτυακούς πράκτορες για την επιβεβαίωση των υπηρεσιών της εκάστοτε εταιρείας προς το χρήστη.



Εικ.5 ο «πύργος των επιπέδων» του σημασιολογικού ιστού, από τον Tim Berners-Lee, ο οποίος περιγράφει τα κύρια επίπεδα της σχεδίασης και του οράματος του σημασιολογικού ιστού

ΚΕΦΑΛΑΙΟ 2

Κύριο μέρος

Παρουσίαση

των δυνατοτήτων και των τεχνολογιών του

Σημασιολογικού Ιστού

ΚΕΦΑΛΑΙΟ 2

Κόριο μέρος – Παρουσίαση των δυνατοτήτων και των τεχνολογιών του Σημασιολογικού Ιστού

2.1 Βασική παρουσίαση των δυνατοτήτων του Σημασιολογικού Web

Στο κεφάλαιο αυτό θα αναδείξουμε τη χρησιμότητα που μπορεί να έχει το νέο web στις υπάρχουσες υπηρεσίες του σημερινού ιστού. Θα εξετάσουμε αυτές τις δυνατότητες που επεκτείνονται στις περισσότερο χρησιμοποιούμενες μορφές του web 2.0, και τις πιθανές εξελίξεις και βελτιώσεις των μειονεκτημάτων που υπάρχουν σήμερα με απώτερο στόχο φυσικά την καλύτερη και ταχύτερη εξυπηρέτηση του χρήστη.

Εκατομύρια έγγραφα διακινούνται σήμερα στο διαδίκτυο (Internet) και σε εταιρικά ενδοδίκτυα (intranets), ενώ ο ρυθμός με τον οποίο αυτά πολλαπλασιάζονται κάθε μέρα αυξάνει. Με τέτοιες ποσότητες παραγόμενης πληροφορίας αποτελεί μεγάλο πρόβλημα η κατανάλωση, δηλαδή η ανακάλυψη, η ταξινόμηση και η χρήση αυτής. Για παράδειγμα, ο επισκέπτης του διαδικτύου που θα αναζητήσει μια πληροφορία δίνοντας λέξεις-κλειδιά σε μια μηχανή αναζήτησης, το πιο πιθανόν είναι ότι θα λάβει πάρα πολλές ιστοσελίδες που θα περιέχουν τις λέξεις αυτές, αλλά θα πρέπει να επωμιστεί ο ίδιος το χρονοβόρο ξεκαθάρισμα των περιττών ιστοσελίδων, προκειμένου να βρει ποιες από αυτές είναι συναφείς με το αντικείμενο που αναζητήσε. Από αυτήν την άποψη, ο άνθρωπος φαίνεται να λειτουργεί ως υπηρέτης της μηχανής. Το πρόβλημα της αναζήτησης της πληροφορίας περιπλέκεται ακόμα περισσότερο από το γεγονός ότι η αναζητούμενη πληροφορία πιθανώς να μην εντοπίζεται λόγω είτε της αποτύπωσής της σε διαφορετική μορφή είτε της συνωνυμίας των λέξεων.

Σε μια ιδανική περίπτωση, ο άνθρωπος με τη χρήση μηχανών θα έπρεπε να είχε πρόσβαση στο σύνολο των πληροφοριών, θέτοντας σύνθετες ερωτήσεις που θα απαιτούσαν εξειδικευμένη αναζήτηση σε πολλαπλές, κατανεμημένες και διαφορετικά δομημένες βάσεις πληροφοριών, με κατάλληλα κριτήρια για τον εντοπισμό

δεδομένων σύμφωνα με το νόημα των ερωτήσεων. Από την άλλη, έξυπνα προγράμματα-πράκτορες θα αναλάμβαναν το έργο της εξόρυξης γνώσης από διάσπαρτα συστήματα αποθήκευσης, αλλά και της μετάφρασης των διαφορετικών λεξιλογίων και αναπαράστασης αυτών σε μια κοινή μορφή. Προκύπτει, λοιπόν, η ανάγκη ώστε το διαδίκτυο να εμπλουτιστεί με τεχνικές που αφορούν στην αποθήκευση και αναπαράσταση της πληροφορίας, οι οποίες και θα επιτρέψουν στις μηχανές την ανάληψη ενός ουσιαστικού ρόλου διαχείρισης και ανεύρεσης της διαρκώς αυξανόμενης και σημασιολογικά σωστής πληροφορίας που διακινείται.

Στη συνέχεια θα παρουσιάσουμε πιο συγκεκριμένα τις βελτιώσεις που μπορεί να επιφέρει ο σημασιολογικός ιστός σε καθημερινές εφαρμογές και χρήσεις του web 2.0 και έτσι ώστε να προσπαθήσουν να καταρριφθούν τα προβλήματα που αναφέρθηκαν στις προηγούμενες παραγράφους.

2.1.1 Διαχείριση της γνώσης

Η διαχείριση της γνώσης (knowledge management)³ αναφέρεται στην απόκτηση, προσπέλαση και συντήρηση της γνώσης μέσα σε ένα οργανισμό, ή ακόμα και σε οποιονδήποτε χρήστη γενικότερα. Θεωρείται ως μια θεμελιώδης δραστηριότητα για όλους τους χρήστες του ιστού, από τον απλό ιδιώτη, έως τις μεγάλες επιχειρήσεις. Ειδικά οι επιχειρήσεις δίνουν τεράστια σημασία στη διαχείριση της γνώσης, καθώς μέσω αυτής μπορούν να αυξήσουν την ανταγωνιστικότητα τους, να δημιουργήσουν νέα γνώση και να αυξήσουν την παραγωγικότητα τους. Η εσωτερική γνώση μιας επιχείρησης θεωρείται πλέον ως ένα πνευματικό περιουσιακό στοιχείο και της δίνεται η ανάλογη προσοχή.

Με δεδομένα τα παραπάνω, ας δούμε ποια είναι τα ελαττώματα της σημερινής τεχνολογίας και τις αλλαγές που είναι διατεθειμένο να φέρει το νέο web.

³ Huang, C., 2009. Sharing knowledge in a supply chain using the semantic web.

Παρόλο που οι περισσότερες πληροφορίες σήμερα παρέχονται με μια σχετικά δομημένη μορφή, όπως για παράδειγμα αρχεία κειμένου, εικόνες ή και βίντεο, από την πλευρά της διαχείρισης γνώσης παρουσιάζονται οι παρακάτω περιορισμοί⁴:

- Η αναζήτηση πληροφορίας: οι χρήστες και οι επιχειρήσεις εξαρτώνται από μηχανές αναζήτησης οι οποίες βασίζονται σε λέξεις-κλειδιά, τους περιορισμούς των οποίων αναλύσαμε σε προηγούμενο κεφάλαιο.
- Η εξαγωγή πληροφοριών: απαιτείται ανθρώπινη προσπάθεια και χρόνος για τον εντοπισμό των πληροφοριών από τα ανακτημένα αποτελέσματα. Οι υπάρχοντες πράκτορες του web δεν είναι ικανοί να εκτελέσουν τη συγκεκριμένη δουλειά με ικανοποιητικά αποτελέσματα, καθώς είναι τυποποιημένη και αυτοματοποιημένη η μέθοδος που ακολουθούν.
- Η συντήρηση πληροφοριών: σήμερα υπάρχουν προβλήματα στον διαχωρισμό των άσχετων και των απαρχαιωμένων πληροφοριών σχετικά με μια αναζήτηση.
- Η αποκάλυψη πληροφοριών: το σημείο αυτό αναφέρεται περισσότερο σε επιχειρήσεις, καθώς η νέα γνώση που υπάρχει μέσα σε βάσεις δεδομένων μιας εταιρείας, εξάγεται με τη χρήση εξόρυξης δεδομένων (data mining). Ωστόσο, αυτή η τακτική είναι δύσκολη για συλλογές εγγράφων που είναι δομημένα ανεπαρκή αν όχι εντελώς μη δομημένα.

Τελειώνοντας τώρα με την παρουσίαση των περιορισμών της σημερινής τεχνολογίας του ιστού, θα αναφέρουμε επιγραμματικά τους στόχους του σημασιολογικού ιστού για το μέλλον του web και για πολύ εξελιγμένα συστήματα διαχείρισης της γνώσης:

- Η γνώση θα είναι ομαδοποιημένη και κατανεμημένη σε εννοιολογικές περιοχές, ανάλογα με τη σημασία της.

⁴ F. van Harmelen and D. Fensel. Practical Knowledge Representation for the Web.

- Θα είναι εφικτός ο καθορισμός πρόσβασης σε συγκεκριμένα τμήματα πληροφοριών και εγγράφων.
- Η σημερινή αναζήτηση με λέξεις κλειδιά θα αντικατασταθεί με απαντήσεις σε ερωτήματα, και έτσι η παραγόμενη γνώση θα εξάγεται και θα παρουσιάζεται με φιλικό και κατανοητό τρόπο από τον χρήστη.
- Θα υποστηρίζεται η απάντηση ερωτημάτων σε περισσότερα του ενός έγγραφα, και έτσι θα ξεπεραστεί ο περιορισμός της μεμονωμένης αναζήτησης με λέξεις-κλειδιά.
- Θα υπάρχουν εργαλεία τα οποία θα λειτουργούν αυτοματοποιημένα και θα είναι υπεύθυνα για τη συντήρηση της γνώσης και θα ελέγχουν για ασυνέπειες δημιουργώντας νέα γνώση.

Συνοψίζοντας το μικρό αυτό υποκεφάλαιο παρατηρούμε ότι οι υποσχέσεις που πραγματοποιεί το επερχόμενο web, μόνο εύκολες δεν μπορούν να χαρακτηριστούν, καθώς ο τομέας της διαχείρισης γνώσης είναι εξαιρετικά σημαντικός και τα προβλήματα του σήμερα παραμένουν στάσιμα, παρόλη τη μικρή βελτίωση τους. Χρειάζεται χρόνος και ιδανική αξιοποίηση της νέας τεχνολογίας για να γίνουν τα παραπάνω πράξη.

2.1.2 Το ηλεκτρονικό εμπόριο μεταξύ επιχειρήσεων και χρηστών-καταναλωτών του Ιστού

Όπως ο καθένας μας ξέρει, το ηλεκτρονικό εμπόριο⁵ (e-commerce) ανάμεσα σε μια επιχείρηση και έναν καταναλωτή, είναι μια από τις κυρίαρχες εμπειρίες των χρηστών του ιστού. Ένας τυπικός χρήστης, δηλαδή, ο οποίος επισκέπτεται διάφορα on line καταστήματα και αφού δει και περιηγηθεί στις προσφορές και τις υπηρεσίες

⁵ Brian Blake, Simon Parsons and Terry R. Payne. The synergy of electronic commerce, agents and semantic web services.

του, κάνει την επιλογή του σύμφωνα με τα προσωπικά του κριτήρια και προχωρά στην παραγγελία των προϊόντων που επιθυμεί.

Η καλύτερη και προσδοκώμενη περίπτωση θα ήταν να μπορούσε ο χρήστης να συλλέξει πληροφορίες για αυτό που ψάχνει, όπως για τις τιμές, τους όρους, τις προϋποθέσεις και τη διαθεσιμότητα των προϊόντων σε πολλά δικτυακά καταστήματα, και στο τέλος θα προχωρούσε στην επιλογή της καλύτερης προσφοράς σύμφωνα με τα κριτήρια του. Εδώ όμως μπαίνει ο παράγοντας της μη αυτόματης φυλλομέτρησης των ιστοσελίδων του σημερινού ιστού, η οποία είναι εξαιρετικά χρονοβόρα για να πραγματοποιηθεί σε τέτοιο φάσμα αναζήτησης. Έτσι συνήθως ένας χρήστης καταλήγει να επισκέπτεται από ένα έως πολύ λίγα δικτυακά καταστήματα πριν κάνει την απόφασή του.

Η κατάσταση αυτή σήμερα έχει φτάσει σε σημείο να διευκολύνεται αρκετά από τα διαθέσιμα εργαλεία που υπάρχουν στον ιστό με τη μορφή ρομπότ αγορών (shopbots), τα οποία είναι στην ουσία πράκτορες λογισμικού που επισκέπτονται πολλά καταστήματα ταυτόχρονα, εξάγουν πληροφορίες για το χρήστη, όπως οι τιμές και τα διαθέσιμα προϊόντα, και στο τέλος δημιουργούν μια γενική επισκόπηση της αγοράς. Η λειτουργικότητα τους παρέχεται από προγράμματα τα οποία ονομάζονται προγράμματα περιτύλιξης (wrappers), τα οποία είναι προγράμματα λογισμικού που είναι ικανά να εξάγουν πληροφορίες από ένα δικτυακό κατάστημα. Για να λειτουργήσει η παραπάνω μέθοδος χρειάζεται να αναπτύσσεται ένα πρόγραμμα περιτύλιξης ανά κατάστημα. Η τεχνική αυτή όμως έχει μερικά μειονεκτήματα.

Οι πληροφορίες εξάγονται από την δικτυακή τοποθεσία του εκάστοτε καταστήματος με την μέθοδο της αναζήτησης με λέξεις-κλειδιά και άλλες παρεμφερείς μεθόδους ανάλυσης κειμένου. Η διαδικασία αυτή είναι επιρρεπής σε λάθη, καθώς λειτουργεί με τη χρήση υποθέσεων για την εγγύτητα των πληροφοριών, όπως για παράδειγμα η τιμή ενός προϊόντος δηλώνεται από τη λέξη 'τιμή' που ακολουθείται από το σύμβολο € και ένα θετικό αριθμό. Η παραπάνω μέθοδος ονομάζεται ευρετική (heuristic), και η λειτουργία της δεν είναι πάντα εγγυημένη. Αυτό γιατί συνήθως εξάγονται μόνο μεμονωμένες πληροφορίες και παραλείπονται άλλες εξίσου σημαντικές όπως, τα έξοδα αποστολής, οι ημερομηνίες παράδοσης, η διαθεσιμότητα του προϊόντος, οι περιορισμοί σχετικά με τις χώρες προορισμού, το επίπεδο ασφάλειας και οι πολιτικές προσωπικού απορρήτου (privacy policies).

Το σημασιολογικό web

Παρατηρούμε όμως ότι όλοι αυτοί οι παράγοντες μπορεί να παίζουν σημαντικό ρόλο στην τελική απόφαση του χρήστη.

Εδώ μπαίνει ο σημασιολογικός ιστός, ο οποίος αποβλέπει στην ανάπτυξη ειδικών και εξολοκλήρου νέων πρακτόρων λογισμικού, οι οποίοι θα είναι ικανοί να ερμηνεύουν τις πληροφορίες για τα προϊόντα και τους όρους παροχής υπηρεσιών:

- Οι πληροφορίες τιμολόγησης θα εξάγονται με το σωστό τρόπο και ολοκληρωμένα, και οι πολιτικές παράδοσης και προσωπικού απορρήτου θα ερμηνεύονται και θα συγκρίνονται με βάση τις απαιτήσεις του χρήστη.
- Πρόσθετες πληροφορίες σχετικά με την υπόληψη και την αξιοπιστία των δικτυακών καταστημάτων θα εξάγονται και θα αξιολογούνται από ανεξάρτητες πηγές, όπως σωματεία καταναλωτών, οργανισμούς αξιολόγησης η ακόμα και forums χρηστών του διαδικτύου.
- Ο προγραμματισμός χαμηλού επιπέδου προγραμμάτων περιτύλιξης θα είναι πια περιττός.
- Ακόμα περιπλοκότεροι πράκτορες λογισμικού θα είναι σε θέση να διεξάγουν αυτοματοποιημένες διαπραγματεύσεις με πράκτορες καταστημάτων για το όφελος του χρήστη.

Συνολικά παρατηρούμε πως οι αλλαγές του σημασιολογικού web θα είναι αν μη τι άλλο ριζικές και εάν καταφέρουν να υλοποιηθούν στο βαθμό που αναμένεται, η εμπειρία περιήγησης και εμπορικής χρήσης του ιστού θα αλλάξει εντελώς επίπεδο.

2.1.3 Το ηλεκτρονικό εμπόριο μεταξύ επιχειρήσεων

Οι περισσότεροι από μας συνδέουμε το εμπορικό κομμάτι του παγκόσμιου ιστού με το ηλεκτρονικό εμπόριο μεταξύ των επιχειρήσεων και των καταναλωτών, παρόλα αυτά όμως, η μεγαλύτερη προοπτική του νέου web βρίσκεται στο εμπόριο μεταξύ των επιχειρήσεων (business to business).

Μέχρι τις μέρες μας οι επιχειρήσεις ανταλλάζουν δεδομένα με τη χρήση της προσέγγισης της Ηλεκτρονικής Ανταλλαγής Δεδομένων (Electronic Data Interchange). Η τεχνολογία αυτή όμως είναι περίπλοκη και κατανοητή μόνο από ειδικούς. Είναι δύσκολη σε διάφορους τομείς, όπως της συντήρησης και του προγραμματισμού και είναι επιρρεπής σε σφάλματα. Κάθε επικοινωνία μεταξύ επιχειρήσεων απαιτεί να γίνεται κάθε φορά ξεχωριστός προγραμματισμός, πράγμα που είναι όπως αντιλαμβανόμαστε δαπανηρό. Πρόσθετα, τα δεδομένα που ανταλλάσσονται μέσω της Ηλεκτρονικής Ανταλλαγής Δεδομένων δεν μπορούν εύκολα να ενσωματωθούν σε άλλες επιχειρηματικές εφαρμογές.

Το διαδίκτυο φαίνεται να είναι η ιδανική υποδομή για την επικοινωνία μεταξύ των επιχειρήσεων. Οι επιχειρήσεις αναζητούν συνεχώς λύσεις βασισμένες στο διαδίκτυο και έχουν εμφανιστεί νέα επιχειρηματικά μοντέλα όπως οι Πύλες business to business. Ωστόσο, το ηλεκτρονικό εμπόριο μεταξύ επιχειρήσεων εμποδίζεται σήμερα, λόγω της έλλειψης προτύπων. Η HTML (Hypertext Markup Language) είναι πολύ ανίσχυρη σε αυτό τον τομέα, δηλαδή στο να υποστηρίζει αποτελεσματικά στις δραστηριότητες που χρειάζονται. Δεν παρέχει ούτε τη δομή, ούτε την απαιτούμενη σημασιολογία των πληροφοριών. Το νέο πρότυπο στον χώρο του σημασιολογικού ιστού ονομάζεται XML (Extensible Markup Language) και αποτελεί μια σημαντική βελτίωση στα προβλήματα που αντιμετωπίζει η HTML, αλλά με τον περιορισμό ότι μπορεί να υποστηρίξει επικοινωνίες μόνο στην περίπτωση ότι υπάρχει εκ των προτέρων συμφωνία για το λεξιλόγιο και το νόημα που θα χρησιμοποιηθεί.

Με την υλοποίηση του σημασιολογικού ιστού, θα επιτραπεί στις επιχειρήσεις να συνεταιριστούν και να συνεργάζονται χωρίς πολλές επιβαρύνσεις. Οι διαφορές στην ορολογία θα λύνονται με τη χρήση πρότυπων *αφηρημένων μοντέλων πεδίου* (abstract domain models), και τα δεδομένα θα είναι διαθέσιμα για ανταλλαγή με τη χρήση διάφορων υπηρεσιών μετάφρασης. Σχεδόν όλες οι διαπραγματεύσεις και οι συμφωνίες θα έρχονται σε πέρας αυτόματα, ή με μικρή ανθρώπινη παρέμβαση.

Η εν δυνάμει χρήση και διείσδυση του σημασιολογικού ιστού στο εμπόριο μεταξύ επιχειρήσεων μπορεί να αποδειχθεί καταλυτικός παράγοντας για την επιτυχία συμφωνιών ανάμεσα σε εταιρείες, και να διευκολύνει σε τεράστιο βαθμό τις σχέσεις ανάμεσα τους και ανάμεσα και στην ίδια την εταιρεία και τους εργαζομένους της.

2.1.4 Τα συστήματα wiki

Ο παγκόσμιος ιστός τώρα πια, διευρύνεται με εργαλεία τα οποία επιτρέπουν στο χρήστη να έχει ενεργή συμμετοχή στον ιστό. Η εξέλιξη αυτή θεωρήθηκε επαναστατική και το 2004 της δόθηκε και ονομασία: Ιστός 2.0 (web 2.0).

Η διεύρυνση αυτή επεκτείνεται σε πολλούς τομείς του σημερινού web 2.0, και ένα κομμάτι της είναι το αντικείμενο του κεφαλαίου μας, τα συστήματα wiki.⁶ Τα wiki λοιπόν, είναι συλλογές διαφόρων ιστοσελίδων που δίνουν τη δυνατότητα στους χρήστες τους να προσθέτουν το δικό τους περιεχόμενο, είτε με τη δημιουργία απευθείας κειμένου είτε με τη χρήση συνδέσμων (links), μέσω του browser που χρησιμοποιούν. Τα wiki επιτρέπουν τη συνεργατική δημιουργία γνώσης, καθώς παρέχουν τη δυνατότητα της πλήρους ελευθερίας προσθήκης και τροποποίησης πληροφοριών, χωρίς να τους περιορίζουν σε κανένα τομέα, όπως την κυριότητα του περιεχομένου (copyright), περιορισμούς πρόσβασης (π.χ. επί πληρωμή) και κάποιο περιορισμό στο χρόνο εργασίας του χρήστη. Χρησιμοποιούνται για διάφορους σκοπούς, όπως:

- Την ανάπτυξη νέας γνώσης, ως αποτέλεσμα συνολικής προσπάθειας κάποιας κοινότητας με συνεισφορά ενός πολύ μεγάλου αριθμού ανεξάρτητων χρηστών. Το πιο τρανταχτό παράδειγμα είναι η εγκυκλοπαίδεια γενικού περιεχομένου Wikipedia, η οποία πια έχει γιγαντωθεί.
- Η διαχείριση γνώσης μιας δραστηριότητας ή ενός έργου. Χαρακτηριστικά παραδείγματα είναι η ανταλλαγή πρακτικών συνεδρίων, ή η αναζήτηση και ανταλλαγή ιδεών μέσω μιας ιστοσελίδας κοινωνικής δικτύωσης.

⁶ Mukundan, S., 2004. Spinning the Semantic Web.

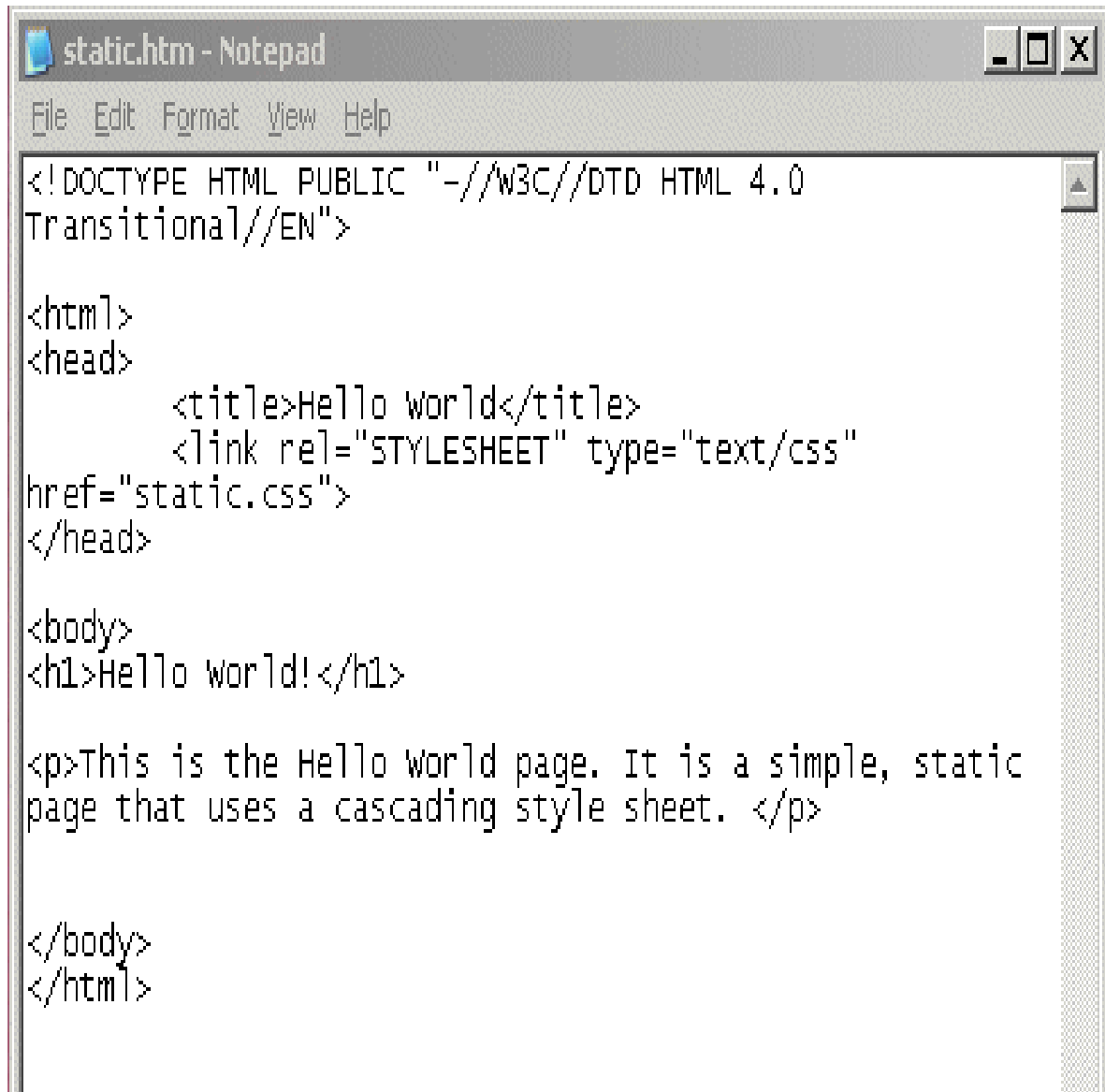
2.2 Βασική παρουσίαση των τεχνολογιών του Σημασιολογικού Ιστού

2.2.1 Ρητά μεταδεδομένα

Την δεδομένη χρονική στιγμή, το περιεχόμενο του ιστού είναι μορφοποιημένο για ανθρώπους αναγνώστες, παρά για προγράμματα και μηχανές. Η κυρίαρχη γλώσσα στην οποία είναι γραμμένες οι ιστοσελίδες είναι η HTML.

Οι πληροφορίες παρουσιάζονται με ικανοποιητική μορφή για τους ανθρώπους, αλλά οι υπολογιστές αντιμετωπίζουν προβλήματα. Οι αναζητήσεις γίνονται με λέξεις-κλειδιά που αναγνωρίζουν από τον κώδικα της ιστοσελίδας. Ένας πράκτορας λογισμικού μπορεί να αναγνωρίσει λεπτομέρειες ως ένα βαθμό. Θα έχει πρόβλημα όμως να ξεχωρίσει λεπτομέρειες που θα αφορούν στην αναζήτηση του χρήστη και εν τέλει θα αποπροσανατολιστεί.

Ένα παράδειγμα μιας τυπικής ιστοσελίδας σε HTML είναι το ακόλουθο:



```
static.htm - Notepad
File Edit Format View Help
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0
Transitional//EN">

<html>
<head>
    <title>Hello world</title>
    <link rel="stylesheet" type="text/css"
href="static.css">
</head>

<body>
<h1>Hello world!</h1>

<p>This is the Hello world page. It is a simple, static
page that uses a cascading style sheet. </p>

</body>
</html>
```

Εικ.6 Παράδειγμα κώδικα HTML: στο παράδειγμα αυτό φαίνεται ότι η διάταξη μιας σελίδας με HTML κώδικα είναι προσχεδιασμένη ως προς τη μορφή και τη δομή της, καθώς θα πρέπει να ξεκινάει πάντα με την ετικέτα `</html>` και να συνεχίζεται με την ετικέτα `</body>`. Τέλος όπως και οι αρχικές ετικέτες έτσι και οι ενδιάμεσες είναι προκαθορισμένης ονομασίας.

Η προσέγγιση του σημασιολογικού ιστού για την επίλυση των προβλημάτων κατανόησης, δεν είναι η δημιουργία ευφών πρακτόρων λογισμικού, αλλά το πρόβλημα προσεγγίζεται από την πλευρά των ιστοσελίδων. Αν η HTML αντικατασταθεί από άλλες καταλληλότερες γλώσσες, τότε οι ιστοσελίδες θα μπορούν οι ίδιες να μεταφέρουν το περιεχόμενο τους στο χρήστη. Εκτός από το να περιλαμβάνουν πληροφορίες για τους ανθρώπους αναγνώστες, θα περιλαμβάνουν και πληροφορίες σχετικά με το περιεχόμενό τους. Η αναπαράσταση αυτή είναι πολύ ευκολότερη για τους υπολογιστές. Η διαδικασία αυτή περιγράφεται από τον όρο μεταδεδομένα. Ο όρος μεταδεδομένα (metadata)⁷, αναφέρεται στα δεδομένα για τα δεδομένα. Τα μεταδεδομένα, αυτό που κάνουν στην ουσία είναι να συλλαμβάνουν το μέρος του νοήματος των δεδομένων, και έτσι προκύπτει και ο όρος *σημασιολογικός* στον σημασιολογικό ιστό.

Όπως συμβαίνει και σήμερα στην ανάπτυξη ιστοσελίδων, ο χρήστης δε θα είναι απαραίτητο να είναι ειδικός της επιστήμης των υπολογιστών για να δημιουργήσει μια ιστοσελίδα. Θα υπάρχουν εργαλεία διευκόλυνσης για αυτό το σκοπό. Υπάρχει βέβαια το ερώτημα γιατί οι χρήστες να εγκαταλείψουν την HTML από τη στιγμή που είναι ήδη γνωστή και καθιερωμένη. Η απάντηση βρίσκεται στην εξέλιξη της πληροφορικής, η οποία είναι πάντα ραγδαία και πρέπει συνεχώς να εγκαταλείπεται κάποιο πρότυπο όταν είναι παρωχημένο, για κάποιο νεότερο και πιο υποσχόμενο.

Τα παραπάνω βέβαια δεν συνιστούν ότι η XML είναι το απόλυτα σωστό μονοπάτι για το σημασιολογικό ιστό. Απλά αυτή τη στιγμή είναι το πρότυπο που είναι καθιερωμένο από τον οργανισμό W3C και αν δεν κριθεί επαρκής θα αλλάξει σε ένα άλλο.

Ακολουθεί ένα παράδειγμα της γλώσσας XML:

⁷ F. van Harmelen and D. Fensel. Practical Knowledge Representation for the Web.

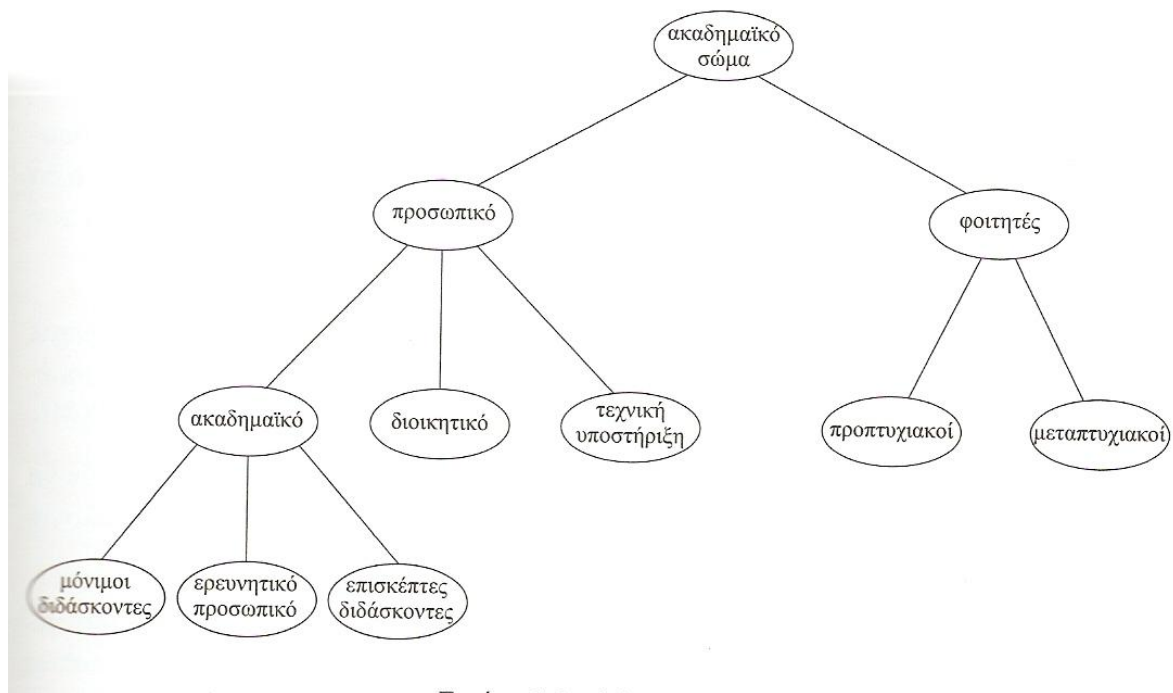
```
<?xml version="1.0"?>
<menu>
  <parent title="kirupaPicks">
    <child>
      <title>kirupa.com</title>
      <link>http://www.kirupa.com</link>
    </child>
    <child>
      <title>kirupaForum</title>
      <link>http://www.kirupa.com/forum/</link>
    </child>
    <child>
      <title>kirupa's Blog</title>
      <link>http://blog.kirupa.com</link>
    </child>
  </parent>
</menu>
```

Εικ.7 Παράδειγμα κώδικα XML. Παρατηρούμε ότι αντίθετα με την HTML η XML δεν περιορίζει τον χρήστη ως προς την ονομασία των ετικετών αλλά τον αφήνει ελεύθερο να ονομάσει τις ετικέτες σύμφωνα με το ζητούμενό του. Η μορφή φυσικά είναι παρόμοια για να μη ξενίσει και να κάνει τη μετάβαση ευκολότερη.

2.2.2 Οντολογίες

Η οντολογία⁸, στο χώρο του σημασιολογικού web εκφράζεται από τον ορισμό του R. Studer: «μια οντολογία είναι μια ρητή και τυπική προδιαγραφή μιας επίνουιας (conceptualization)».

Για να γίνουμε πιο κατανοητοί, μια οντολογία περιγράφει τυπικά ένα πεδίο ενδιαφέροντος. Συνήθως αποτελείται από μια λίστα όρων και τις σχέσεις μεταξύ των όρων αυτών. Οι όροι υποδηλώνουν σημαντικές έννοιες (κλάσεις) μεταξύ των αντικειμένων ενός πεδίου. Για παράδειγμα, ακολουθεί μια απλή οντολογία ενός πανεπιστημίου που θα περιλάμβανε τους όρους, καθηγητές, φοιτητές, μαθήματα, αμφιθέατρα και τους επιστημονικούς κλάδους.



Εικ.8 Παράδειγμα απλής οντολογίας, το οποίο περιλαμβάνει τους όρους και τις σχέσεις μεταξύ των πεδίων και των κλάσεων της οντολογίας. Στην αρχή βλέπουμε πως είναι το ακαδημαϊκό σώμα και στη συνέχεια απλώνεται ιεραρχικά η οντολογία.

⁸ J. Davies, D. Fansel and F. van Harmelen. Towards the semantic web: ontology-driven knowledge management.

Στο πλαίσιο του ιστού, οι οντολογίες παρέχουν μια *κοινή κατανόηση ενός πεδίου*. Για παράδειγμα με τη χρήση της οντολογίας μπορούν να ξεπεραστούν προβλήματα όρων με διαφορετική έννοια, όπως σε ένα πανεπιστήμιο ο όρος ‘course’ μπορεί να σημαίνει ένα ολόκληρο εξάμηνο, ενώ σε ένα άλλο να αναφέρεται απλά σε ένα μάθημα. “Οι οντολογίες δηλαδή υποστηρίζουν τη σημασιολογική διαλειτουργικότητα (semantic interoperability)” – Tim Burners-Lee.

Άλλα πλεονεκτήματα των οντολογιών είναι η οργάνωση και η πλοήγηση σε ιστότοπους, όπου οι ιστοσελίδες θα προβάλλουν την ιεραρχία τους και ο χρήστης θα μπορεί να επιλέγει ανάμεσα στις κατηγορίες και τις υποκατηγορίες.

Πρόσθετα, οι οντολογίες είναι χρήσιμες για την βελτίωση της ακρίβειας των αναζητήσεων στο web. Με τη χρήση των ανάλογων μηχανών αναζήτησης θα μπορούν να αναζητηθούν ιστοσελίδες που αναφέρονται στην ακριβή *έννοια* μιας οντολογίας, παρά να εμφανίσουν όλες τις σελίδες που περιέχουν διάφορες διεσπαρμένες λέξεις-κλειδιά.

Τέλος, οι αναζητήσεις στον ιστό θα μπορούν να αξιοποιήσουν πληροφορίες γενίκευσης/ειδίκευσης. Θα μπορεί δηλαδή η μηχανή αναζήτησης να καταλάβει εάν ένα ερώτημα αποτύχει να εντοπίσει σχετικά αποτελέσματα και θα προτείνει στο χρήστη ένα πιο γενικό ερώτημα, ή θα είναι ικανή να εκτελέσει αυτή την ενέργεια εκ των προτέρων. Οι σημαντικότερες γλώσσες οντολογιών που θα αναλυθούν παρακάτω είναι οι εξής:

- RDF
- RDF Schema
- OWL

2.2.3 Λογική

Η λογική (logic) είναι το επιστημονικό πεδίο που μελετά τις αρχές της συλλογιστικής (reasoning). Ειδικότερα, η λογική προσφέρει τρεις βασικές ιδιότητες. Πρώτον, προσφέρει τυπικές γλώσσες για την έκφραση της γνώσης. Δεύτερον, παρέχει μια κατανοητή τυπική σημασιολογία (formal semantics). Τρίτον, τα προγράμματα αυτοματοποιημένης συλλογιστικής (automated reasoners) μπορούν να εξάγουν μπορούν να εξάγουν συμπεράσματα από την τρέχουσα γνώση, καθιστώντας έτσι ρητή την υπονοούμενη γνώση. Προγράμματα σαν και αυτά έχουν μελετηθεί εκτενώς στον τομέα της Τεχνίτης Νοημοσύνης.

Ένα σημαντικό πλεονέκτημα της λογικής, είναι ότι μπορεί να παρέχει αιτιολογήσεις συμπερασμάτων. Επιπλέον, «οι ερευνητές της Τεχνητής Νοημοσύνης έχουν αναπτύξει μεθόδους παρουσίασης μιας αιτιολόγησης με φιλικό προς τον άνθρωπο τρόπο, οργανώνοντας μια απόδειξη (proof) ως φυσική επαγωγή (natural deduction) και ομαδοποιώντας έναν αριθμό βημάτων συμπερασμού χαμηλού επιπέδου σε μετα-βήματα, τα οποία ο άνθρωπος θα θεωρούσε τυπικά ένα μόνο βήμα απόδειξης» - Aaron Stanton. Πιο απλά, οι αιτιολογήσεις είναι σημαντικές, καθώς θα αυξάνουν την εμπιστοσύνη των χρηστών απέναντι στους πράκτορες λογισμικού. Οι αιτιολογήσεις θα είναι επίσης απαραίτητες για τις δραστηριότητες μεταξύ αυτών καθ'αυτών των πρακτόρων λογισμικού.

Για να είναι χρήσιμη η λογική στον ιστό, θα πρέπει να μπορεί να είναι χρησιμοποιήσιμη μαζί με άλλα δεδομένα και να είναι και επεξεργάσιμη από τους υπολογιστές. Για αυτό τον λόγο, γίνεται συνεχείς προσπάθειες να αποτυπωθούν οι αποδείξεις και η λογική γνώση στις γλώσσες του σημασιολογικού ιστού.

2.2.4 Πράκτορες

Οι πράκτορες⁹ είναι προγράμματα λογισμικού που ενεργούν αυτόματα. Η δουλειά ενός προσωπικού πράκτορα στο σημασιολογικό ιστό, θα είναι να δέχεται κάποιες εργασίες και προτιμήσεις από το χρήστη, να αναζητεί πληροφορίες από τον ιστό, να επικοινωνεί με άλλους πράκτορες λογισμικού, να συγκρίνει πληροφορίες για τις απαιτήσεις και τις προτιμήσεις του χρήστη, να διαλέγει συγκεκριμένες επιλογές και να επιστρέφει τις απαντήσεις και τα αποτελέσματα στον χρήστη.

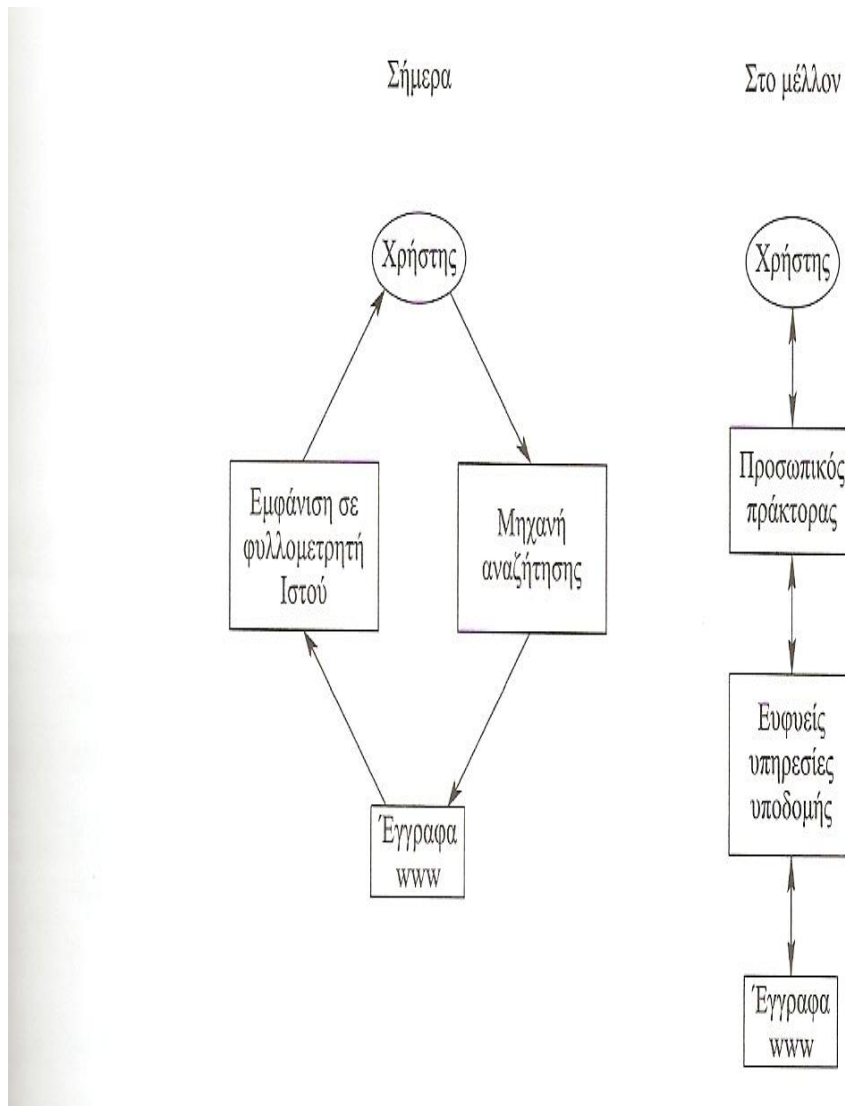
Πρέπει να επισημανθεί σε αυτό το σημείο ότι, οι πράκτορες λογισμικού, δεν θα αντικαταστήσουν τους ανθρώπους ως χρήστες στο σημασιολογικό ιστό, ούτε θα λαμβάνουν απαραίτητα αποφάσεις. Στις περισσότερες περιπτώσεις, ο ρόλος τους είναι καθαρά δευτερεύων, με κύριες απασχολήσεις την συλλογή και την οργάνωση πληροφοριών και την παρουσίαση των επιλογών, από τις οποίες οι χρήστες θα επιλέγουν αυτά που αναζητούν. Οι πράκτορες λογισμικού θα χρησιμοποιούν όλες τις τεχνολογίες που έχουμε περιγράψει, δηλαδή¹⁰:

- Τα μεταδεδομένα θα χρησιμοποιούνται για τον προσδιορισμό και την εξαγωγή πληροφοριών από πηγές του ιστού
- Οι οντολογίες θα χρησιμοποιούνται για την υποστήριξη των διαδικτυακών αναζητήσεων, την ερμηνεία των πληροφοριών που ανακτώνται και την επικοινωνία με άλλους πράκτορες λογισμικού
- Η λογική θα χρησιμοποιείται για την επεξεργασία πληροφοριών που ανακτώνται και την εξαγωγή συμπερασμάτων

Για να λειτουργήσουν βέβαια τα παραπάνω χρειάζονται ακόμη επιπλέον τεχνολογίες, όπως είναι οι γλώσσες επικοινωνίας πρακτόρων λογισμικού, οι οποίες ακόμα είναι σε βρεφικό στάδιο. Επίσης η αναπαράσταση των επιθυμιών, των πεποιθήσεων και των προθέσεων των πρακτόρων θα είναι χρήσιμες για τις προηγμένες εφαρμογές.

⁹ Alkhateeb, F. Web Semantics: Science, Services and Agents on the World Wide Web, 2009.

¹⁰ Hendler, James. Agents and the semantic web.



Εικ.9 Παράδειγμα χρήσης πρακτόρων λογισμικού. Η εικόνα αυτή μας δείχνει στα αριστερά τη σημερινή κατάσταση χρήσης του ιστού, όπου ο χρήστης μέσω της μηχανής αναζήτησης εισάγει τις λέξεις κλειδιά που χρειάζεται, η μηχανή αναζήτησης του επιστρέφει αποτελέσματα και ο χρήστης διαλέγει αυτά που θέλει και ο φυλλομετρητής τα εμφανίζει. Στο μέλλον όπως δείχνει η δεξιά εικόνα, ο χρήστης θα έχει ελάχιστη συμμετοχή στην αναζήτηση πληροφοριών, καθώς ο προσωπικός του πράκτορας σε συνεργασία με τις υπηρεσίες υποδομής άλλων ιστοσελίδων θα επικοινωνεί με αυτές και θα δίνει στον χρήστη πολλαπλά αποτελέσματα πληροφοριών.

ΚΕΦΑΛΑΙΟ 3

Εργαλεία

ΤΟΥ

Σημασιολογικού Ιστού

ΚΕΦΑΛΑΙΟ 3

Εργαλεία του Σημασιολογικού Ιστού

Στο κεφάλαιο αυτό θα κάνουμε μια βασική προσέγγιση στα τρία βασικά εργαλεία του σημασιολογικού ιστού αυτή τη στιγμή: την XML, την RDF και την OWL.



Εικ.10 χαρακτηριστικός κύβος με τα εργαλεία του Σημασιολογικού Ιστού

3.1.1 Δεδομένα έγγραφα ιστού: XML

Η XML¹¹ είναι μια εφαρμογή της SGML, δηλαδή του διεθνούς προτύπου για αναπαράσταση πληροφοριών από ανθρώπους σε υπολογιστές. Η HTML είναι και αυτή μια εφαρμογή της SGML. Η ιδέα για τη δημιουργία της XML προήρθε από τις ελλείψεις της HTML και το όραμα ενός καλύτερου web.

Όπως η HTML, έτσι και η XML είναι γλώσσα σήμανσης, δηλαδή δίνει το περιεχόμενο στον χρήστη της να διαμορφώσει το περιεχόμενο του και να παρέχει πληροφορίες σχετικά με το ρόλο του συγκεκριμένου περιεχομένου. Επίσης βασίζεται και αυτή στη χρήση ετικετών (tags), όπως για παράδειγμα, `</main>`. Όλες οι ετικέτες πρέπει να κλείνουν, και το περιεχόμενό τους ονομάζεται στοιχείο (element).

Μια παρατήρηση αρκετά προφανής είναι ότι οι άνθρωποι μπορούν να διαβάσουν και να κατανοήσουν και τις δύο κυρίαρχες μορφές κατασκευής πληροφοριών στον ιστό, δηλαδή τις HTML και την XML. Τι γίνεται όμως με τους υπολογιστές; Ο υπολογιστής δεν μπορεί να αναγνωρίσει το περιεχόμενο μιας πληροφορίας δημιουργημένης σε HTML, καθώς η τελευταία δεν περιέχει δομικές πληροφορίες για ένα έγγραφο, δηλαδή πληροφορίες για τμήματα του εγγράφου και τις σχέσεις μεταξύ τους. Από την άλλη πλευρά, το έγγραφο XML, είναι πολύ ευκολότερα προσπελάσιμο από υπολογιστές, καθώς περιέχει περιγραφές για όλες τις πληροφορίες.

Ένα άλλο χαρακτηριστικό της XML, είναι ότι, σε αντίθεση με την HTML, δεν περιγράφει τη μορφοποίηση του εγγράφου, αλλά ξεχωρίζει το περιεχόμενο από την μορφοποίηση. Αυτό μπορεί να το χαρακτηρίσει κάποιος ως μειονέκτημα, αλλά στην πράξη δεν είναι, καθώς η HTML αναγκάζει τον χρήστη να δημιουργήσει και τη μορφοποίηση, ενώ η XML τον αφήνει πιο ελεύθερο στον τομέα αυτό.

Στη συνέχεια θα δείξουμε κάποια πρότυπα σελίδων με HTML και με XML για να παρατηρήσουμε κάποιες μορφολογικές διαφορές.

¹¹ Shadbolt, N., 2006. The Semantic Web Revisited.

```
<html>
<head>
<title> My First Webpage</title>
</head>
<body>
<h1>This is a big heading for my page</h1>
<h3>This is a smaller heading for my page</h3>
The title in the head of the document above is
going to show on the top of my Browser. See how
HTML tags start and are then told to stop with
the forward slash.
<p>This is new paragraph on my page. The HTML
tag tells the little coding machine in your
browser to expect HTML code. Better end the
paragraph now with a slash in the tag</p>
<p><strong>This text is bold and in a new
paragraph and below is a horizontal rule<strong> </p>
<hr width="80%"> </hr>
</body>
</html>
```

Εικ.11 Παράδειγμα html κώδικα

```
<Objects>
  <Object name="Client">
    <Property name="Q3ClientId" type="Int" />
    <Property name="Name" type="String" />
    <CreateEvent name="NewClient" />
    <Event name="ClientFired">
      <Parameter name="Timestamp" type="AutoTimeStamp" />
      <Parameter name="LevelTime" type="Int" />
      <Parameter name="WeaponId" type="Int" />
      <Parameter name="ClientLocationX" type="Float" />
      <Parameter name="ClientLocationY" type="Float" />
      <Parameter name="ClientLocationZ" type="Float" />
      <Parameter name="Declination" type="Float" />
      <Parameter name="Ascension" type="Float" />
    </Event>
  </Object>
</Objects>
```

Εικ.12 Παράδειγμα XML κώδικα

Αυτό που παρατηρούμε από τα δύο αυτά παραδείγματα είναι και η μεγαλύτερη δυναμικότητα της XML έναντι της HTML. Επειδή η HTML έχει δημιουργηθεί με στόχο την παρουσίαση πληροφοριών, το σύνολο των ετικετών της είναι σταθερό: λίστες, χρώματα κλπ. Στην XML, από την άλλη, μπορούμε όπως προείπαμε να χρησιμοποιήσουμε τις πληροφορίες με διάφορους τρόπους, και έτσι ο ορισμός του λεξιλογίου εξαρτάται αποκλειστικά από το χρήστη, δίνοντας του έτσι απεριόριστη ελευθερία και δημιουργικότητα.

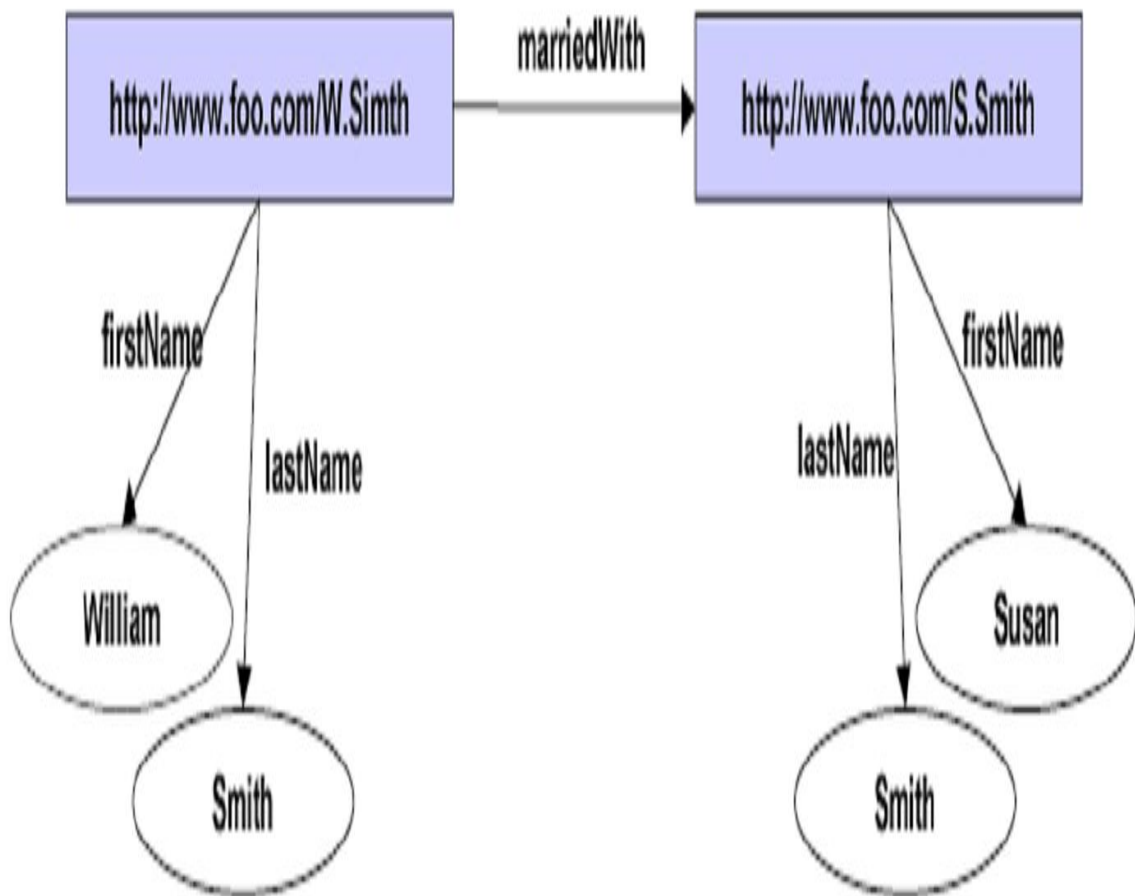
Τέλος, θα πρέπει να επισημάνουμε ότι η XML μπορεί να χρησιμεύσει ως μια ενιαία μορφή ανταλλαγής δεδομένων μεταξύ εφαρμογών. Αυτή η δυνατότητα μπορεί να εφαρμοστεί σε όλους τους τομείς του φάσματος των χρηστών του web, από τους απλούς χρήστες έως τις επιχειρήσεις.

3.1.2 Περιγραφή πόρων Ιστού: RDF

Η XML όπως αναφέρθηκε και προηγουμένως, είναι μια μετα-γλώσσα περιγραφής, σχεδιασμένη για να καλύψει το όραμα του νέου ιστού, αλλά δεν είναι ακόμα αναπτυγμένη σε επαρκές επίπεδο, με αποτέλεσμα να μην είναι πλήρης. Έτσι, παρόλο που παρέχει ένα ενιαίο πλαίσιο και ένα σύνολο εργαλείων, όπως οι agents και οι descriptors, για την ανταλλαγή δεδομένων και μεταδεδομένων μεταξύ εφαρμογών, δεν παρέχει κάποιον τρόπο καθορισμού της σημασιολογίας των δεδομένων.

Εδώ λοιπόν μπαίνει η RDF¹², η οποία παρόλο που συχνά αποκαλείται γλώσσα, είναι στην ουσία ένα μοντέλο δεδομένων, το οποίο σχετίζεται με γλώσσες προγραμματισμού- βλ.XML - και δανείζεται χαρακτηριστικά και πλεονεκτήματα τους. Η RDF εστιάζει στο πλαίσιο της πηγής (resource) και της ιδιότητας (property). Για να γίνει πιο κατανοητό το παραπάνω θεωρούμε ότι μια πηγή θεωρείται μια αναγνωρίσιμη οντότητα, π.χ. ένας άνθρωπος, ένα κτίριο ή μια ιστοσελίδα, και η ιδιότητα είναι η σχέση μεταξύ δύο πηγών ή μεταξύ μιας πηγής και μιας αφηρημένης έννοιας, όπως π.χ. η σχέση ενός ανθρώπου με το ονοματεπώνυμο του. Τον συνδυασμό και την περιγραφή των παραπάνω τον κάνει το μοντέλο δεδομένων RDF. Παρατίθεται και μια εικόνα που δείχνει και γραφικά αυτό που περιγράφω.

¹² Dos Santos, H.L., 2003. A proposal for management of RDF and RDF schema metadata in MOF.



Εικ.13 Απλό παράδειγμα της RDF. Στο παράδειγμα αυτό φαίνεται η χρήση της RDF , η οποία προσδίδει χαρακτηριστικά σε κάθε μέρος ενός πεδίου και τα περιγράφει έτσι ώστε να είναι διακριτές οι σχέσεις και η σημασία κάθε πεδίου για την αναζήτηση ενός χρήστη. Αναγνωρίζει δηλαδή, ότι το “William” είναι το μικρό όνομα αυτού που περιγράφεται, ενώ το “Smith” είναι το επίθετο. Επίσης το ίδιο γίνεται για τη γυναίκα του “Susan Smith”. Τέλος, η σχέση μεταξύ των δύο περιγράφεται με την έννοια “married with”, η οποία προσδίδει πως είναι αντρόγυνο.

Οι βασικές και θεμελιώδεις έννοιες της RDF είναι τρεις και παρόλο που ποικίλλουν σε ονομασίες στα διάφορα άρθρα που μπορεί να συναντήσει κάποιος, ο αριθμός τους μένει ίδιος. Έτσι καταλήγουμε στις εξής: πόροι (resources), ιδιότητες (properties) και προτάσεις (statements). Η τριπλέτα αυτή μπορεί να συναντηθεί σε διάφορες ονομασίες, όπως subject- predicate- object, αλλά στην ουσία αναφέρει τα ίδια πράγματα.

Ξεκινώντας με την ανάλυση της πρώτης έννοιας της RDF, έχουμε να δούμε ότι ως πόρος (resources) θεωρείται ένα αντικείμενο για το οποίο θέλουμε να μιλήσουμε. Όπως αναφέρθηκε και πριν, ένας πόρος μπορεί να είναι ένας άνθρωπος, ένα κτίσμα, μια τοποθεσία κλπ. Κάθε πόρος έχει μια διεύθυνση URI (Ενιαίο Αναγνωριστικό Πόρου). Η διεύθυνση URI μπορεί να είναι μια διεύθυνση URL ή κάποιο άλλο είδος μοναδικού χαρακτηριστικού.

Συνεχίζοντας με τις ιδιότητες (properties) παρατηρούμε ότι, είναι μια ειδική περίπτωση πόρων, καθώς περιγράφουν τις σχέσεις μεταξύ των πόρων, όπως για παράδειγμα, «γραμμένο από», «αριθμός σελίδων», «τίτλος» και διάφορα άλλα. Οι ιδιότητες της RDF καθορίζονται επίσης και από τα URI που αναφερθήκαμε προηγουμένως. Η ιδέα της χρήσης των URI και για τον καθορισμό των σχέσεων μεταξύ των δεδομένων είναι πολύ σημαντική, καθώς μας παρέχει το προνόμιο μιας μοναδικής παγκόσμιας ονομασίας του αντικειμένου που περιγράφεται, μειώνοντας δραστικά το πρόβλημα της ανωνυμίας που διαπρέπει το σημερινό web. Τελειώνοντας με το τρίτο χαρακτηριστικό του RDF, τις προτάσεις, βλέπουμε ότι, οι προτάσεις (statements) βεβαιώνουν τις ιδιότητες των πόρων. Η πρόταση είναι, ουσιαστικά, μια τριάδα αντικειμένου- ιδιότητας- τιμής, η οποία αποτελείται από έναν πόρο, μια ιδιότητα και μια τιμή.

Η RDF είναι μια ανοιχτή γλώσσα σχεδιασμού και δεν περιορίζει τους χρήστες με ειδικό λεξιλόγιο για την περιγραφή πόρων, όπως επίσης και δεν καθορίζει τη σημασιολογία κάποιου πεδίου. Αυτό όμως, μπορεί να γίνει από το χρήστη με το παρακλάδι της RDF τη γλώσσα RDF Schema.

Η RDF Schema είναι¹³, στην ουσία, μια γλώσσα μοντελοποίησης και κατηγοριοποίησης των ελεύθερων και χωρίς περιορισμό δεδομένων της RDF. Παρέχει ένα μηχανισμό για την περιγραφή συγκεκριμένων πεδίων, με τη χρήση στοιχείων μοντελοποίησης με σταθερό νόημα. Οι έννοιες-κλειδιά της RDF Schema είναι οι κλάσεις, οι σχέσεις υποκλάσης και ο προσδιορισμός των υποκατηγοριών σε μια ιδιότητα για τη διευκόλυνση της περιγραφής. Ας δούμε όμως ένα απλό παράδειγμα για την καλύτερη κατανόηση της RDF Schema. Έστω ότι έχουμε την φράση «Τα μαθηματικά διδάσκονται από τον Γιάννη Παπαδόπουλο». Στην πρόταση αυτή μπορούμε να ορίσουμε διάφορες κλάσεις και παραμέτρους με τη χρήση της RDFS όπως, ιδιότητα (property), υποιδιότητα (sub property of), κλάση (class), υποκλάση (subclass of) και άλλες. Έτσι, αν θέλουμε να διατυπώσουμε ότι η κλάση «λέκτορας» είναι υποκλάση της κλάσης «μέλος ακαδημαϊκού προσωπικού», μπορούμε να 1) ορίσουμε τους απαιτούμενους πόρους περιγραφής για το «λέκτορας», 2) να ορίσουμε τις ιδιότητες της περιγραφής και τέλος 3) να εφαρμόσουμε την τριάδα των βασικών ιδεών της RDF και να τελειώσουμε την περιγραφή μας. Υπάρχουν, βέβαια και προβλήματα της RDFS όπως ότι είναι αρκετά στοιχειώδης ως γλώσσα και δε βοηθά για μια ανεπτυγμένη περιγραφή, και επίσης ότι είναι δύσχρηστη και απαιτεί εξοικείωση από τη μεριά του χρήστη. Ακολουθεί και ένα γραφικό παράδειγμα της RDF Schema.

¹³ Palmisano, I. , 2006. A RDF-Based Framework for User Profile Creation and Management

3.1.3 Γλώσσα οντολογιών Ιστού: OWL

Οι γλώσσες οντολογιών¹⁴ επιτρέπουν στους χρήστες να γράφουν ρητές, τυπικές επίνσεις (conceptualizations) για μοντέλα πεδίων. Βασική απαίτηση αυτών των γλωσσών είναι η καλά ορισμένη σύνταξη, η αποδοτική υποστήριξη συλλογισμών, η τυπική σημασιολογία η επαρκής εκφραστική ισχύς και η ευκολία στην έκφραση.

Η γλώσσα OWL βασίζεται στην γλώσσα RDF και έχει το ίδιο είδος σύνταξης με αυτή. Επειδή και η OWL είναι μια γλώσσα προγραμματισμού, καταλαβαίνουμε πόσο αναγκαία είναι η καλά ορισμένη σύνταξη. Αν και η σύνταξη της RDF δεν θεωρείται ότι ευκολότερο για τον άνθρωπο, είναι το πρότυπο αυτή τη στιγμή και έτσι το μοιράζεται και η OWL.

Πρόσθετα, η OWL περιλαμβάνει δύο τύπους ιδιοτήτων για την περιγραφή των εγγράφων της. Ο πρώτος είναι η ιδιότητα αντικειμένου, η οποία συσχετίζει αντικείμενα με άλλα αντικείμενα. Για παράδειγμα, η ιδιότητα *is taught by* (διδάσκεται από) προσδίδει μια ιδιότητα στο αντικείμενο που περιγράφει. Ο δεύτερος είναι η ιδιότητα τύπου δεδομένων, οι οποίες συσχετίζουν αντικείμενα με τιμές ενός τύπου δεδομένων. Παράδειγμα αποτελούν τα *phone*, *title* ή *age*.

Συνοψίζοντας την OWL, είδαμε ότι είναι μια γλώσσα οντολογιών, η οποία βασίζεται στην RDF και υπό προϋποθέσεις θα μπορούσε να είναι και παρακλάδι της. Μας επιτρέπει να περιγράψουμε τη σημασιολογία της γνώσης με τρόπο κατανοητό από υπολογιστές.

¹⁴ Brian Blake, Simon Parsons and Terry R. Payne. The synergy of electronic commerce, agents and semantic web services.

ΚΕΦΑΛΑΙΟ 4

Επίλογος

Σύνοψη

ΚΕΦΑΛΑΙΟ 4

Επίλογος - Σύνοψη

Από όλα όσα ειπώθηκαν παραπάνω θα μπορούσαμε να κατηγοριοποιήσουμε σε δύο ερμηνείες τον σημασιολογικό ιστό. Η πρώτη θα μπορούσε να είναι ο σημασιολογικός ιστός ως ιστός δεδομένων και η δεύτερη ο σημασιολογικός ιστός ως εμπλουτισμός του υπάρχοντος ιστού.

Στην πρώτη ερμηνεία, ο βασικός στόχος του σημασιολογικού ιστού είναι η δυνατότητα ενοποίησης των δομημένων και των ελλιπώς δομημένων πηγών στο διαδίκτυο.

Στην δεύτερη ερμηνεία, ο στόχος του σημασιολογικού ιστού είναι να βελτιώσει τον υπάρχοντα παγκόσμιο ιστό. Οι τρόποι που θα μπορούσε να το κάνει αυτό είναι η βελτίωση των μηχανών αναζήτησης, ο σημασιολογικός εμπλουτισμός των υπαρχουσών ιστοσελίδων με μεταδεδομένα, την ταχύτητα πλοήγησης και ανάκτησης πληροφοριών και την δυνατότητα εξαγωγής πληροφοριών από πολλαπλές πηγές ταυτόχρονα. Όπως και να έχει, είναι ενδιαφέρον να δούμε μερικές πιθανές εφαρμογές του νέου ιστού.

Αρχικά μπορούμε να δούμε στον βιομηχανικό τομέα, και γενικότερα στον τομέα των επιχειρήσεων τις εξής προοπτικές:

- Διαχείριση γνώσης, κυρίως στα ενδοδίκτυα μεγάλων εταιρειών
- Ενοποίηση δεδομένων και αρχείων εταιρειών
- Ηλεκτρονικές επιστήμες και βελτίωση υπηρεσιών τους

Περνώντας τώρα σε πιο προσωπικά και ατομικά δεδομένα, ο σημασιολογικός ιστός θα μπορούσε να ωφελήσει πολύ σε διάφορες καθημερινές λειτουργίες όπως:

Το σημασιολογικό web

- Την εκπαίδευση, μέσω της χρήσης των μεταδεδομένων και της τηλεεκπαίδευσης
- Την εξατομίκευση του ιστού μέσω της χρήσης των προηγμένων τεχνολογιών αναζήτησης, ανάκτησης κλπ.
- Την καλύτερα αξιοποίηση των broadband υπηρεσιών διαδικτύου και ταχυτήτων μέσω των προηγμένων μηχανών αναζήτησης και των έξυπνων πρακτόρων λογισμικού

Ακόμα ίσως να είναι νωρίς να πούμε με βεβαιότητα εάν το όραμα του σημασιολογικού ιστού θα καταφέρει να γίνει πραγματικότητα, καθώς οι δυσκολίες αυτού του εγχειρήματος είναι πολύ μεγάλες, και για να δημιουργηθούν όλες αυτές οι οντολογίες και τα μεταδεδομένα που χρειάζονται θα χρειαστεί πολύς χρόνος και προσπάθεια από τους ερευνητές, και φυσικά από τους χρήστες, καθώς η μετάβαση από μια σταθερή κατάσταση σε μια άλλη δεν είναι ποτέ απλή.

Αυτή τη στιγμή βρισκόμαστε στο σημείο που ονομάζεται web 2.0 (αν και μερικοί ήδη αναφέρονται στο web 2.5 ως πρόδρομο του web 3.0). όπως αναφέρθηκε το web 2.0 είναι βασισμένο και προσανατολισμένο για τη διευκόλυνση του χρήστη και δεν δίνει ιδιαίτερη σημασία στην πλευρά των υπολογιστών. Το νέο web θα είναι το web 3.0, δηλαδή το σημασιολογικό web, το οποίο αναπτύσσεται ήδη παράλληλα με το σημερινό web και θα είναι προσανατολισμένο στους υπολογιστές και θα τους βοηθά να κατανοούν μια πληροφορία παρά απλά να τη μεταφράζουν βοηθώντας έτσι στην εξέλιξη του ιστού.

Μένει να δούμε στην πράξη λοιπόν εάν θα μπορέσουν να πραγματοποιήσουν αυτά που υπόσχονται, και αν το καταφέρουν θα είναι μια πραγματική επανάσταση και το επόμενο βήμα στην επικοινωνία μεταξύ των ανθρώπων από όλες τις απόψεις και οπτικές γωνίες το δει κάποιος.

ΚΕΦΑΛΑΙΟ 5

Βιβλιογραφία

Γλωσσάριο

ΚΕΦΑΛΑΙΟ 5

Βιβλιογραφία

- 1) Tim Burners-Lee. Semantic web road map. September 1998
- 2) E. Dumbill. The semantic web: a primer. November 2000
- 3) F. van Harmelen and D. Fensel. Practical Knowledge Representation for the Web. 1999
- 4) J. Davies, D. Fensel and F. van Harmelen. Towards the semantic web: ontology-driven knowledge management. New York: Wiley, 2002
- 5) I. Horrocks B. Parsia, P. Patel-Schneider and J. Hendler. Semantic Web Architecture: Stack or Two Towers? Springer 2005, σ.37-41
- 6) Dos Santos, H.L., 2003. A proposal for management of RDF and RDF schema metadata in MOF. *Springer-Verlag Berlin Heidelberg*, [online]. σ. 1014-1031
- 7) Han, Y., 2005. A RDF-based digital library system. *Emerald Group Publishing Limited*, [online]. Vol. 24 No. 2, σσ. 234-240,
- 8) Palmisano, I. , 2006. A RDF-Based Framework for User Profile Creation and Management, *Springer-Verlag Berlin Heidelberg*, [online]. σσ. 606-613,
- 9) Bernon, C. , 2005. A Study of Some Multi-agent Meta-models, *Springer-Verlag Berlin Heidelberg*, [e-journal] . σσ. 62-77. Abstract only

- 10) Feigenbaum, L., 2007. Boca: an open - source RDF store for building Semantic Web applications. *Briefings in bioinformatics*, vol.8 (3) σσ. 195-200.
- 11) Βαρλάμης, Η. , *Επισκόπηση τεχνικών και συστημάτων ολοκλήρωσης ετερογενούς σημασιολογικής πληροφορίας και γνώσης*. [online] Αθήνα: Οικονομικό Πανεπιστήμιο Αθήνας.
- 12) Alkhateeb, F. , 2009. Extending SPARQL with regular expression patterns. *Web Semantics: Science, Services and Agents on the World Wide Web*, [e-journal], vol.7 σσ. 57-73. Abstract only.
- 13) van Harmelen, F. 2009. *Knowledge Engineering Rediscovered: Towards Reasoning Patterns for the Semantic Web*. [online] Amsterdam: Vrije Universiteit
- 14) Gergatsoulis, M., 2005. Multidimensional RDF. *Springer-Verlag Berlin Heidelberg*, [e-journal], σσ. 1188-1205. Abstract only.
- 15) Downes, S., 2005. Semantic networks and social networks. *The Learning Organization*, [e-journal], vol. 12 (5) σσ. 411-417. Abstract only.
- 16) Broekstra, J., 2002. Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema. *Springer-Verlag Berlin Heidelberg*, [e-journal], σσ. 54-68. Abstract only.
- 17) Huang, C., 2009. Sharing knowledge in a supply chain using the semantic web. *Expert systems with applications*, [e-journal], σσ. 3145-3161. Abstract only.
- 18) Mukundan, S., 2004. Spinning the Semantic Web. *On the Horizon*, [e-journal], vol 12(2) σσ. 74-78. Abstract only.

- 19) Shadbolt, N., 2006. The Semantic Web Revisited. *IEEE COMPUTER SOCIETY*, [online]
- 20) Brian Blake, Simon Parsons and Terry R. Payne. The synergy of electronic commerce, agents and semantic web services. 2004
- 21) Hendler, James. Agents and the semantic web. 2001
- 22) Uschold, Mike. A semantic continuum on the semantic web. 2002
- 23) T. Burners-Lee, Hendler, J. and Lassila, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. 2001
- 24) Mc Liraith, S., Cao Son, T. and Zeng, H. Semantic Web Services. 2001
- 25) Sure, Y. and Studer, R. Semantic Web technologies for digital libraries. 2005
- 26) Burke, M. The semantic web and the digital library. 2009
- 27) Όλο το υλικό από την επίσημη σελίδα του Semantic Web www.w3c.org

Γλωσσάριο

Artificial Intelligence - Τεχνητή νοημοσύνη: Ο όρος αναφέρεται στον κλάδο της επιστήμης υπολογιστών ο οποίος ασχολείται με τη σχεδίαση και την υλοποίηση υπολογιστικών συστημάτων που μιμούνται στοιχεία της ανθρώπινης συμπεριφοράς τα οποία υπονοούν έστω και στοιχειώδη ευφυΐα: μάθηση, προσαρμοστικότητα, εξαγωγή συμπερασμάτων, κατανόηση από συμφραζόμενα, επίλυση προβλημάτων κλπ.

Browser - φυλλομετρητής: είναι ένα πρόγραμμα που εμφανίζει τις ιστοσελίδες στην οθόνη του υπολογιστή και επιτρέπει την μετάβαση μέσω διάφορων συνδέσμων σε άλλες ιστοσελίδες του ίδιου ή και άλλου server (εξυπηρετητή) του Παγκόσμιου Ιστού.

Community – Κοινότητα: Πρόκειται για μια εικονική «κοινότητα», μια ομάδα χρηστών του Διαδικτύου, οι οποίοι ενδιαφέρονται για ένα κοινό θέμα και έτσι δημιουργούν ένα χώρο αφιερωμένο στο συγκεκριμένο ζήτημα.

Copyright - Πνευματική ιδιοκτησία: Παρέχει αποκλειστικό δικαίωμα στους δημιουργούς ενός έργου να διανέμουν, να αναπαράγουν και να διαθέτουν το έργο αυτό στο ευρύ κοινό. Επιτρέπει στους δημιουργούς να συντηρούνται από τα δημιουργήματά τους και να ελέγχουν τη χρήση του έργου τους.

Email - Ηλεκτρονικό ταχυδρομείο: Μέσω του Διαδικτύου έχουμε τη δυνατότητα να στείλουμε ηλεκτρονική αλληλογραφία σε άλλους χρήστες από τη διεύθυνση ηλεκτρονικού ταχυδρομείου που έχουμε λάβει μαζί με τον προσωπικό μας λογαριασμό (Account) στο Διαδίκτυο.

Forum: Υπηρεσία όπου σε μια περιοχή ενός ηλεκτρονικού online πίνακα οι χρήστες με ένα κοινό ενδιαφέρον μπορούν να τοποθετήσουν τις απόψεις/σημειώσεις τους πάνω σε ένα θέμα. Τα φόρουμ χρησιμοποιούνται συνήθως για να υποβάλουν οι χρήστες ερωτήσεις, να μοιράζονται πληροφορίες, ή να συζητούν τις ιδέες τους. Ένα φόρουμ μπορεί να διευθύνεται από ένα συντονιστή.

HTML - Hypertext Markup Language: Η HTML είναι η γλώσσα με την οποία δημιουργούνται σελίδες στο Διαδίκτυο, οι οποίες αναγνωρίζονται από κάθε υπολογιστή στον κόσμο.

Hypertext – Υπερκείμενο: Είναι οποιοδήποτε κείμενο περιέχει συνδέσμους οι οποίοι παραπέμπουν κάπου αλλού. Οι ιστοσελίδες γενικότερα είναι γεμάτες από υπερκείμενα και υπερσυνδέσμους.

Internet – Διαδίκτυο: Το Internet είναι ένα τεράστιο δίκτυο στο οποίο είναι συνδεδεμένα μεταξύ τους πολλά άλλα μικρά ή μεγάλα δίκτυα υπολογιστών. Πολλές φορές συγχέεται το Διαδίκτυο με τον παγκόσμιο ιστό το World Wide Web (WWW). Το WWW αποτελεί μόνο ένα μέρος του Διαδικτύου.

Library 2.0: Η υιοθέτηση του Web 2.0 στις διαδικτυακές υπηρεσίες της βιβλιοθήκης. Ανάμεσα στα χαρακτηριστικά της είναι η χρησιμοποίηση τεχνολογιών πολυμέσων που επιτρέπουν την αλληλεπίδραση και τη συμμετοχή των χρηστών, στις διαδικτυακές υπηρεσίες και συλλογές της βιβλιοθήκης. Είναι επικεντρωμένη στον χρήστη, προσφέρει τις υπηρεσίες της χρησιμοποιώντας πολυμέσα, είναι πλούσια σε κοινωνικές αλληλεπιδράσεις και επιτρέπει στους χρήστες να συμβάλλουν στις αλλαγές της βιβλιοθήκης

Link - Σύνδεση: είναι μια εικόνα ή κείμενο σε ένα έγγραφο του World Wide Web που ενεργεί ως διαδρομή σε μια άλλη ιστοσελίδα ή αρχείο στο Διαδίκτυο. Η αξία ενός site στο web μπορεί να καθοριστεί και από πόσο συχνά αναφέρεται ως πηγή, και συνεπώς τείνει να χαρακτηριστεί ως αυθεντία στο είδος του. Τα links στον κόσμο των blogs είναι ακόμα πιο σημαντικά καθώς οι bloggers δημιουργώντας links μπορούν να διανέμουν με συλλογικό τρόπο την αναγνωσιμότητα του κάθε blog.

Το σημασιολογικό web

υποδηλώνει τη σύνδεση μεταξύ διαφορετικών ιστοσελίδων. Σε μια ιστοσελίδα μπορούμε να βρούμε συνδέσμους είτε επάνω σε κείμενο (που είναι τότε μαρκαρισμένο με διαφορετικό χρώμα), είτε σε φωτογραφίες και γραφικά. Εάν κάνουμε κλικ με το ποντίκι μας επάνω σε ένα σύνδεσμο, τότε θα οδηγηθούμε στην ιστοσελίδα που κρύβεται πίσω από το σύνδεσμο αυτό.

OWL – Γλώσσα οντολογιών Ιστού: Η OWL είναι μια οικογένεια γλωσσών προγραμματισμού για την αναπαράσταση οντολογιών που έχει υιοθετηθεί από τον υπεύθυνο οργανισμό οντολογιών Web Ontology Consortium.

RDF - Resource Definition Framework: Είναι ένα σύνολο κανόνων (γλώσσα) για τη δημιουργία περιγραφών πληροφοριών, ειδικά για πληροφορίες που είναι διαθέσιμες στο World Wide Web.

Semantic Web - Σημασιολογικός ιστός: «Ο Σημασιολογικός Ιστός είναι μια επέκταση του σημερινού ιστού όπου η πληροφορία έχει καλά καθορισμένο νόημα, καθιστώντας τη συνεργασία μεταξύ ανθρώπων και υπολογιστών πιο αποτελεσματική» Tim Berners-Lee. Θα είναι το κύριο χαρακτηριστικό του Web 3.0. Αυτή τη στιγμή για να πραγματοποιήσει μια λειτουργία ο υπολογιστής θα πρέπει ένας άνθρωπος μέσα από κάποιες ιστοσελίδες να τον καθοδηγήσει. Στο Semantic Web ένας υπολογιστής δεν θα χρειάζεται έναν άνθρωπο για να πραγματοποιήσει λειτουργίες όπως αναζητήσεις, συγκρίσεις, sharing και mahups. Θα μπορεί από μόνος του να καταλάβει τις πληροφορίες και να τις επεξεργαστεί. Επί της ουσίας μιλάμε για ένα internet φτιαγμένο όχι μόνο για ανθρώπους αλλά και για μηχανές.

URL - Uniform Resource Locator: Με τον όρο περιγράφεται μια διαδικτυακή διεύθυνση, με την οποία μεταφέρεστε σε κάποιο πολύ συγκεκριμένο (μονοσήμαντο) ιστότοπο στο Διαδίκτυο. Κάθε URL υπάρχει μία και μόνο φορά στον κόσμο.

Web 2.0: Ο όρος χρησιμοποιείται για να υποδηλώσει όλες τις διαδικτυακές υπηρεσίες και εφαρμογές που προωθούν τη συνεργασία και την επικοινωνία μεταξύ των χρηστών του Διαδικτύου, όπως για παράδειγμα οι ιστοσελίδες κοινωνικής δικτύωσης, τα blogs, τα wikis κ.α. Στο Web 2.0, οι χρήστες έχουν χαρακτηριστικά μεγαλύτερη διάδραση με την σελίδα. Δημιουργούν δικά τους εξατομικευμένα προφίλ και αναλόγως με την θεματολογία του site μπορούν να παράγουν περιεχόμενο. Αυτή ακριβώς είναι η νοοτροπία του Web 2.0. Όλοι είναι δημιουργοί, όλοι είναι χρήστες, και όλοι είναι δικτυωμένοι. ο κοινό όλων αυτών των σελίδων είναι ότι δεν έχουν «δικό» τους περιεχόμενο, αντιθέτως τα πάντα είναι «user-generated».

Web 3.0: ενώ στο Web 2.0 η επικέντρωση βρισκόταν στη δημιουργία γνώσης και περιεχομένου από την πλευρά των χρηστών καθώς και η διαμοίρασή της με εργαλεία όπως τα Blogs και τα Wikis, στο Web 3.0 σημαντικό ρόλο παίζει η οργάνωση αυτής της γνώσης και η ενσωμάτωση αυτού του δυναμικού περιεχομένου (dynamic content) στον Παγκόσμιο Ιστό.

Wiki: Χαρακτηρίζονται οι ιστοσελίδες των οποίων το περιεχόμενο διαμορφώνεται συλλογικά από πολλούς χρήστες. Ο κάθε επισκέπτης έχει τη δυνατότητα να επεξεργαστεί τα ήδη υπάρχοντα δεδομένα, αλλά και να προσθέσει νέα.

XML - (eXtensible Markup Language): Μια γλώσσα που σχεδιάστηκε ειδικά για την διαχείριση των web αρχείων κειμένου, ανεξάρτητα από λογισμικό και hardware. Έχει ανάγκη για λογισμικό που να μπορεί να το λάβει, να το στείλει και να το παρουσιάσει.