

ΕΥΦΥΕΙΣ ΠΡΑΚΤΟΡΕΣ ΓΙΑ ΤΟΝ ΠΑΓΚΟΣΜΙΟ ΙΣΤΟ:

Μια σύντομη αναφορά στο παρόν και στο μέλλον

Χρήστος Οροβας

Δρ. Πληροφορικής, Καθηγητής ΠΕ20

1^ο ΤΕΕ Κοζάνης, Βερμίου 26, 501 00 Κοζάνη

e-mail : orovas@sch.gr , URL : <http://users.sch.gr/orovas>

Περίληψη: Η ταχύτητα με την οποία αναπτύσσεται ο παγκόσμιος ιστός και ο όγκος της πληροφορίας που είναι διαθέσιμη είναι δυο παράγοντες που θα αποτελούσαν εμπόδιο στην αποτελεσματική χρήση και αξιοποίηση του. Οι ευφυείς πράκτορες του παγκοσμίου ιστού είναι προγράμματα τα οποία βοηθούν τον χρήστη αυτής της υπηρεσίας του Διαδικτύου να βρεί γρήγορα ιστοσελίδες σχετικές με το θέμα για το οποίο αναζητεί πληροφορίες. Στόχος τους είναι η εύρεση ιστοσελίδων οι οποίες σχετίζονται με τις λεγόμενες ως λέξεις κλειδιά της αναζήτησης. Η συνοπτική παρουσίαση της λειτουργίας τους και οι κατευθύνσεις στις οποίες κινείται η οργάνωση της πληροφορίας στον παγκόσμιο ιστό είναι το θέμα αυτής της ανακοίνωσης. Έμφαση δίνεται στο γεγονός ότι η μελλοντική εισαγωγή στοιχείων σχετικά με την δομή και το νόημα της πληροφορίας στον παγκόσμιο ιστό θα διευκολύνει τον αυτόματο χειρισμό της με χρήση ευφυών πρακτόρων λογισμικού και θα τους επιτρέψει να προσφέρουν ένα πλήθος από αναβαθμισμένες υπηρεσίες πέρα από την απλή αναζήτηση.

Λέξεις κλειδιά: Ευφυείς πράκτορες, Παγκόσμιος Ιστός, Οντολογίες.

Abstract: The speed by which the world wide web is being developed and the volume of the available information are two factors which could be a drawback in web's effective use. The intelligent agents for the world wide web are programs which help the user of this service of the internet find, within a short period of time, sites which are relevant with the subject under investigation. The aim of these programs is to locate web pages related with the keywords of the search. A concise presentation of their operation and the directions in which the organization of information in world wide web is moving into is the subject of this paper. Emphasis is given to the fact that the introduction of elements about the structure and the meaning of the information in the web will make its automatic manipulation using intelligent agents easier and will allow them to offer a number of advanced services beyond the level of simple search for keywords.

Keywords: Intelligent agents, World wide web, ontologies.

Εισαγωγή

Αποτελώντας ίσως το όνειρο κάθε μηχανικού μισό αιώνα νωρίτερα, το διαδίκτυο, αυτό το σύνολο από συνδεδεμένα δίκτυα υπολογιστών σε όλο τον Πλανήτη, προσφέρει την υποδομή και το πλαίσιο πάνω στο οποίο λειτουργούν οι υπηρεσίες του (ηλεκτρονικό ταχυδρομείο, απομακρυσμένη σύνδεση σε υπολογιστή, μεταφορά αρχείων, κανάλια ταυτόχρονης επικοινωνίας, παγκόσμιος ιστός) με σημαντικότερη ίσως και σίγουρα ταχύτερα εξελισσόμενη από αυτές αυτή του Παγκοσμίου Ιστού. Η ύπαρξη και η σημερινή μορφή του Παγκοσμίου Ιστού είναι κάτι που δύσκολα θα μπορούσε να φανταστεί κάποιος μερικές μόνο δεκαετίες νωρίτερα. Η επανάσταση της Πληροφορίας, στην εποχή της οποίας ζούμε και η ανθρωπότητα πέρασε με την εμφάνιση των πρώτων ηλεκτρονικών υπολογιστών, έχει επιφέρει τεράστιες αλλαγές στους περισσότερους, αν όχι σε όλους τους τομείς της ζωής μας. Αυτές οι αλλαγές δεν ήταν αρχικά εμφανείς στο ευρύ κοινό μέχρι την εποχή που πρωτοεμφανίστηκαν οι υπηρεσίες του διαδικτύου. Από τον τρόπο με τον οποίο εργαζόμαστε και ενημερωνόμαστε μέχρι και τον τρόπο με τον οποίο εκφραζόμαστε και επικοινωνούμε οι αλλαγές που υπάρχουν σε σχέση με ότι συνέβαινε στις αρχές π.χ. της δεκαετίας του 1990 είναι τεράστιες.

Μέσα σε αυτή την κοινωνία της Πληροφορίας, όπως εύστοχα ονομάζεται, ο καθένας μπορεί να περιηγηθεί, περιπλανηθεί, αναζητήσει άλλα και ταυτόχρονα να προσθέσει κάτι στον ήδη τεράστιο όγκο πληροφορίας που είναι διαθέσιμος με πολύ μεγάλη ευκολία και διαμέσου μόνο του υπολογιστή μας. Ειδικότερα δε, στην εκπαιδευτική διαδικασία, ο πλούτος της προσφερόμενης γνώσης που είναι διαθέσιμη τόσο στον εκπαιδευτή όσο και στον εκπαιδευόμενο είναι ανεκτίμητος. Τόσο γιατί υπάρχει πρόσβαση σε τεράστιες ποσότητες πληροφορίας σχετικά με οποιοδήποτε θέμα όσο και γιατί υπάρχει πρόσβαση στις τελευταίες εξελίξεις και στις πιο «ενημερωμένες εκδόσεις» αυτής της πληροφορίας. Υπάρχουν όμως δυο ζητήματα που προκύπτουν: Πρώτον, πως μπορούμε να είμαστε ενημερωμένοι για το τι υπάρχει διαθέσιμο την τρέχουσα χρονική στιγμή στον τεράστιο και ταχύτατα εξελισσόμενο Παγκόσμιο Ιστό και δεύτερον, πως μπορούμε να είμαστε σίγουροι για την εγκυρότητα της διαθέσιμης πληροφορίας ; Η ανακοίνωση αυτή αναφέρεται με συντομία στο πρώτο ζήτημα. Μετά από μια σύντομη παρουσίαση του Παγκοσμίου Ιστού και του τρέχοντα τρόπου με τον οποίο είναι οργανωμένη η πληροφορία εκεί, εξετάζονται οι τρόποι που έχουμε στη διάθεση μας για αναζήτηση και αποτελεσματική «πλοήγηση». Το βάρος δίδεται στις αυτοματοποιημένες μεθόδους οι οποίες λειτουργούν με χρήση προγραμμάτων τα οποία ονομάζονται ευφυείς πράκτορες αναζήτησης. Μια σύντομη αρχική εικόνα των πραγμάτων

ΕΥΦΥΕΙΣ ΠΡΑΚΤΟΡΕΣ ΓΙΑ ΤΟΝ ΠΑΓΚΟΣΜΙΟ ΙΣΤΟ

που έρχονται υπάρχει στην τελευταία ενότητα της ανακοίνωσης όπου αναφέρεται το πλαίσιο πάνω στο οποίο μπορεί να δομηθεί η νέα μορφή του Παγκοσμίου Ιστού και το οποίο θα προσφέρει ένα περιβάλλον στο οποίο οι ευφυείς πράκτορες θα μπορούν να προσφέρουν ένα πλήθος από αναβαθμισμένες υπηρεσίες.

Παγκόσμιος Ιστός

Ο Παγκόσμιος Ιστός (Π.Ι) είναι μια υπηρεσία του διαδικτύου [α-διαδίκτυο]. Σαν ιδέα ξεκίνησε στις αρχές της δεκαετίας του 1990 με στόχο να διευκολύνει την οργάνωση, κοινή χρήση, διαθεσιμότητα και αναζήτηση πληροφοριών. Στηρίζεται στη μη γραμμική οργάνωση πληροφορίας, δηλ. στη δυνατότητα πρόσβασης σε κάποια στοιχεία χωρίς να είναι απαραίτητο να προσπελασθούν πρώτα τα προηγούμενα από αυτά στοιχεία. Αυτό επιτυγχάνεται με τη χρήση συνδέσμων που οδηγούν από ένα σημείο σε κάποιο άλλο. Οι πληροφορίες οργανώνονται με τη βοήθεια του υπερκειμένου (hypertext) που είναι ένα ηλεκτρονικό κείμενο το οποίο εκτός από κείμενο μπορεί να περιέχει και άλλες μορφές απεικόνισης της πληροφορίας (εικόνα, ήχο, κ.ά) και παρέχει την δυνατότητα χρήσης συνδέσμων που οδηγούν από ένα σημείο σε άλλα σημεία του ίδιου ή διαφορετικών υπερκειμένων.

Το υπερκείμενο συντάσσεται με τη βοήθεια της γλώσσας περιγραφής υπερκειμένου HTML. Η HTML χρησιμοποιεί ένα σύνολο από ετικέτες για τη διαμόρφωση του τρόπου παρουσίασης του κειμένου. Ένα πολύ απλό υπερκείμενο που περιέχει κείμενο, την εικόνα που βρίσκεται σε κάποιο αρχείο εικόνας με όνομα niceday.gif και ένα σύνδεσμο στην ιστοσελίδα της Διεύθυνσης Δευτεροβάθμιας Εκπαίδευσης Φλώρινας φαίνεται στο σχήμα 1.

```
<HTML>
  <HEAD>
    <TITLE>Απλό Παράδειγμα</TITLE>
  </HEAD>
  <BODY>
    Αυτό το υπερκείμενο περιέχει μια εικόνα και ένα σύνδεσμο.

    <IMG SRC="niceday.gif">

    Πατήστε <A HREF="http://dide.flo.sch.gr"> εδώ </A> για να μεταβείτε
    στην ιστοσελίδα της Δ.Δ.Ε. Φλώρινας.

  </BODY>
</HTML>
```

Σχήμα 1. Απλό παράδειγμα υπερκειμένου

ΕΥΦΥΕΙΣ ΠΡΑΚΤΟΡΕΣ ΓΙΑ ΤΟΝ ΠΑΓΚΟΣΜΙΟ ΙΣΤΟ

Οι ετικέτες (γραμμένες με πλάγια γραμματοσειρά μέσα σε αγκύλες) καθορίζουν το τι θα εμφανιστεί και πώς. Εκτός από κάποιες ετικέτες οι οποίες αφορούν συγκεκριμένο τύπο πληροφορίας (π.χ <TITLE>), στη πλειονότητα τους αφορούν αποκλειστικά και μόνο στη μορφοποίηση του κειμένου χωρίς να προσδιορίζεται το αντικείμενο με το οποίο σχετίζεται η πληροφορία που βρίσκεται στο υπερκείμενο.

Υπερκείμενα τα οποία είναι διαθέσιμα στον Π.Ι ονομάζονται και ιστοσελίδες (websites ή sites). Τα προγράμματα τα οποία μας επιτρέπουν να δούμε ιστοσελίδες ονομάζονται φυλλομετρητές του Π.Ι (web browsers). Τα προγράμματα αυτά αποτελούν τα προγράμματα «πελάτες» (web-clients) του Π.Ι σε αντίθεση με τα προγράμματα εκείνα τα οποία κάνουν τις ιστοσελίδες διαθέσιμες και ονομάζονται «εξυπηρετητές» (web-servers) του Π.Ι. Μια ιστοσελίδα λοιπόν είναι αποθηκευμένη σε κάποιο υπολογιστή και μόνο μέσω του εξυπηρετητή ο οποίος εκτελείται στον ίδιο υπολογιστή γίνεται διαθέσιμη στον έξω κόσμο και το περιεχόμενο της μπορεί να σταλεί για απεικόνιση σε κάποιο πρόγραμμα πελάτη το οποίο εκτελείται σε κάποιο απομακρυσμένο, αλλά συνδεδεμένο με το διαδίκτυο, υπολογιστή.

Οι κανόνες επικοινωνίας μεταξύ των προγραμμάτων πελάτη και εξυπηρετητή του Π.Ι καθορίζονται από το πρωτόκολλο HTTP και κάθε σελίδα έχει μια μοναδική διεύθυνση που τη χαρακτηρίζει. Η διεύθυνση αυτή λέγεται URL (uniform resource locator) και περιέχει το όνομα του πρωτοκόλλου, το όνομα του μηχανήματος και το όνομα του αρχείου στο οποίο περιέχεται το υπερκείμενο (απουσία του ονόματος αρχείου χρησιμοποιείται ένα εξ'ορισμού αρχείο με όνομα index.html). Για παράδειγμα το URL της ιστοσελίδας της Δ.Δ.Ε. Φλώρινας είναι το <http://dide.flo.sch.gr>.

Πράκτορες του Παγκοσμίου Ιστού

Μέσα στο δαιδαλώδες πλαίσιο το οποίο έχει ήδη αναπτυχθεί, αλλά και βρίσκεται διαρκώς σε ολοένα και μεγαλύτερη ανάπτυξη στον Π.Ι με πληροφορίες σχετικά με ο,τιδήποτε μπορεί να φανταστεί κάποιος να βρίσκονται σε ιστοσελίδες οι οποίες είναι διάσπαρτες στους υπολογιστές που είναι συνδεδεμένοι με το διαδίκτυο, γίνεται πολύ εύκολα κατανοητό ότι η εύρεση μιας, ή ενός συνόλου ιστοσελίδων οι οποίες περιέχουν πληροφορίες σχετικά με κάποιο θέμα θα ήταν φοβερά δύσκολη έως αδύνατη.

Λύση σε αυτό το πρόβλημα δίνεται με δύο τρόπους. Ο πρώτος είναι η ύπαρξη ιστοσελίδων οι οποίες συντηρούν καταλόγους του Π.Ι με την έννοια του ότι είναι χωρισμένες σε θεματικές ενότητες (π.χ εκπαίδευση, διασκέδαση, τουρισμός, επιστήμη, κ.ά) στις οποίες υπάρχουν σύνδεσμοι προς ιστοσελίδες σχετικές με το συγκεκριμένο θέμα. Αν και οι

ΕΥΦΥΕΙΣ ΠΡΑΚΤΟΡΕΣ ΓΙΑ ΤΟΝ ΠΑΓΚΟΣΜΙΟ ΙΣΤΟ

ιστοσελίδες αυτές αποτελούν πολύ συχνά κάποια μορφή «πύλης εισόδου» στον Π.Ι και ίσως είναι από τις πρώτες ιστοσελίδες που επισκέπτεται κάποιος αναζητώντας πληροφορίες για κάποιο θέμα δεν είναι πάντα σίγουρο ότι θα βρούμε αυτό που ψάχνουμε, ιδιαίτερα όταν θέλουμε να κάνουμε μια εξειδικευμένη αναζήτηση.

Ο δεύτερος τρόπος είναι με την χρήση των μηχανών αναζήτησης. Με τον όρο αυτό χαρακτηρίζονται ιστοσελίδες οι οποίες παρέχουν την δυνατότητα εισαγωγής κάποιων λέξεων κλειδιών και μπορούν να μας απεικονίσουν ένα σύνολο από ιστοσελίδες μέσα στις οποίες αναφέρονται οι λέξεις αυτές. Στην πραγματικότητα, οι ιστοσελίδες των μηχανών αναζήτησης αποτελούν απλά την διεπαφή του χρήστη με κάποια σύνθετη εφαρμογή η οποία εκτελείται σε κάποιο απομακρυσμένο υπολογιστή και έχει ως στόχο την αναζήτηση μέσα σε μια βάση δεδομένων η οποία περιέχει εγγραφές της γενικής μορφής:

<λεξη κλειδί> <URL>

Αυτές οι βάσεις δεδομένων συντηρούνται με την βοήθεια των ευφυών πρακτόρων αναζήτησης του Π.Ι. [***v-2***]. Με τον όρο ευφυής πράκτορας εννοείται οτιδήποτε μπορεί να δεχθεί ερεθίσματα από το περιβάλλον του, να τα συνδυάσει με κάποια υπάρχουσα γνώση για αυτό και να προβεί σε μια σειρά πράξεων, δράσεων, προκειμένου να μεγιστοποιήσει μια συνάρτηση κέρδους ή απόδοσης [***AI***]. Όπως αναφέρθηκε και στην εισαγωγή, οι ευφυείς πράκτορες αναζήτησης του Π.Ι δεν είναι τίποτα άλλο παρά προγράμματα τα οποία επισκέπτονται συνεχώς ιστοσελίδες και καταγράφουν κάποια στοιχεία σχετικά με τα περιεχόμενα τους. Συνήθως γίνεται καταγραφή των λέξεων που υπάρχουν σε κάθε ιστοσελίδα, μπορούν όμως να χρησιμοποιηθούν και άλλες πληροφορίες οι οποίες είναι σχετικές με την ιστοσελίδα. Υπάρχουν πολλά ονόματα με τα οποία μπορούμε να αναφερθούμε στα προγράμματα αυτά. Συνήθως, εκτός από πράκτορες αναζήτησης είναι γνωστά και ως ρομπότ του Π.Ι, αράχνες του Π.Ι, κυνηγοί του Π.Ι, περιηγητές κ.ά (web robots, spiders, hunters, wanderers). Το περιβάλλον τους λοιπόν αποτελεί ο Π.Ι, τα ερεθίσματα προέρχονται από το περιεχόμενο των σελίδων και οι πιθανές δράσεις τους είναι οι μετακινήσεις στις σελίδες. Το δε ζητούμενο είναι η εύρεση κάποιας πληροφορίας.

Ο τρόπος με τον οποίο οι πράκτορες αναζήτησης βρίσκουν τον δρόμο τους στον Π.Ι καθορίζεται από τον τρόπο αναζήτησης που χρησιμοποιούν. Χρησιμοποιώντας μεθόδους της τεχνητής νοημοσύνης [***AI***], μπορούν να επισκεφθούν ιστοσελίδες ακολουθώντας συνδέσμους από τις ιστοσελίδες κατά βάθος (εξέταση πρώτα στους συνδέσμους κάθε νέας σελίδας), κατά πλάτος (εξέταση όλων των συνδέσμων μιας σελίδας πριν εξετάσουμε τους συνδέσμους των νέων σελίδων), ευριστική (εξέταση σελίδων που πληρούν κάποια κριτήρια, π.χ με βάση το όνομα πεδίου που υπάρχει στο URL), κ.ά. Θα πρέπει εδώ να πούμε ότι οι

ΕΥΦΥΕΙΣ ΠΡΑΚΤΟΡΕΣ ΓΙΑ ΤΟΝ ΠΑΓΚΟΣΜΙΟ ΙΣΤΟ

πράκτορες αναζήτησης του Π.Ι λειτουργούν είτε σε κατάσταση αναζήτησης είτε σε κατάσταση συλλογής. Στην πρώτη περίπτωση επισκέπτονται τις ιστοσελίδες με στόχο την εύρεση σε αυτές κάποιων λέξεων κλειδιών. Στη δεύτερη περίπτωση η επίσκεψη γίνεται με στόχο την συλλογή πληροφοριών σχετικά με την ιστοσελίδα με σκοπό την δημιουργία και την συντήρηση της προαναφερθείσας βάσης δεδομένων η οποία χρησιμοποιείται ως λεξικό.

Η τρέχουσα μορφή του Π.Ι αποτελείται από ιστοσελίδες οι οποίες έχουν σχεδιασθεί έτσι ώστε να είναι αναγνώσιμες και κατανοήσιμες με ευκολία από τους ανθρώπους και όχι για να διευκολύνουν τον αυτόματο χειρισμό τους από προγράμματα [v-1]. Έτσι, δεν είναι λίγες οι φορές όπου τα αποτελέσματα της αναζήτησης δεν είναι ακριβώς τα αναμενόμενα, γεγονός που οφείλεται στην ύπαρξη ασάφειας τόσο κατά την υποβολή των ερωτημάτων προς τους πράκτορες αναζήτησης όσο και στις ίδιες τις σελίδες μιας και η χρήση της φυσικής γλώσσας δεν εξασφαλίζει την απόλυτη κατανόηση του περιεχομένου από ένα πρόγραμμα εκτελώντας μηχανικές διαδικασίες. Χαρακτηριστικά είναι τα προβλήματα που προκύπτουν από την ύπαρξη συνωνύμων ή πολλαπλών νοημάτων για την ίδια λέξη.

Αν και κατά τη δημιουργία ενός υπερκειμένου είναι δυνατή η χρήση κάποιων μετα-ετικετών (ετικέτες οι οποίες περιέχουν πληροφορίες σχετικά με τις πληροφορίες που περιέχονται στην ιστοσελίδα) η χρήση τους δεν αποτελεί πανάκεια [***μετα-ετικέτες***]. Οι μετα-ετικέτες τοποθετούνται στο τμήμα του υπερκειμένου που βρίσκεται ανάμεσα στις ετικέτες <HEAD> και </HEAD> και ένα παράδειγμα χρήσης τους φαίνεται στο σχήμα 2.

```
<HTML>
<HEAD>
  <TITLE>Απλό Παράδειγμα μετα-ετικετών</TITLE>
  <META name="description" content="Απλό παράδειγμα χρήσης των
μετα-ετικετών σε ιστοσελίδες." >
  <META name="keywords" content="meta, meta tags, μετα ετικέτες,
αναζήτηση, περιεχόμενο">
</HEAD>
```

Σχήμα 2. Παράδειγμα χρήσης μετα-ετικετών

Στο σχήμα 2 βλέπουμε τη χρήση δυο μετα-ετικετών. Η μία έχει ως στόχο να δώσει στοιχεία τα οποία περιγράφουν την ιστοσελίδα και η δεύτερη αναφέρει κάποιες λέξεις κλειδιά οι οποίες σχετίζονται με την ιστοσελίδα.

Η χρήση των μετα-ετικετών αλλά και άλλων ετικετών που παρέχουν πληροφορίες για την ιστοσελίδα (π.χ. TITLE) είναι, εν δυνάμει, ζωτικής σημασίας για την αποτελεσματικότητα της αναζήτησης αλλά και για την βαθμολόγηση και ταξινόμηση των

ΕΥΦΥΕΙΣ ΠΡΑΚΤΟΡΕΣ ΓΙΑ ΤΟΝ ΠΑΓΚΟΣΜΙΟ ΙΣΤΟ

ιστοσελίδων αναφορικά με το πόσο κοντά βρίσκονται στο θέμα της αναζήτησης. Αν και είναι ένα βήμα προς τη σωστή κατεύθυνση, η προαιρετική φύση τους, τόσο όσον αφορά την ύπαρξη τους όσο και τη χρήση τους από τους πράκτορες αναζήτησης αλλά και τα παραδείγματα όχι σωστής χρήσης τους [***όχι-σωστα-μετα-ετικετ**] μας οδηγούν στο συμπέρασμα ότι θα πρέπει να υπάρχει ένα καλύτερα δομημένο πλαίσιο το οποίο θα παρέχει πληροφορίες σχετικά με την ιστοσελίδα χρησιμοποιώντας έναν ομοιόμορφο και κοινά αποδεκτό τρόπο. Στοιχεία αυτού του πλαισίου περιγραφής αναφέρονται στην επόμενη ενότητα.

Δομή και νόημα στις ιστοσελίδες

Σύμφωνα με τα παραπάνω γίνεται αντιληπτό ότι η αύξηση της αποτελεσματικότητας και των υπηρεσιών που παρέχονται από τους πράκτορες του Π.Ι καθώς και η μετεξέλιξη του Π.Ι σε περιβάλλον αυτόματης διαχείρισης, άντλησης πληροφοριών και εξαγωγής συμπερασμάτων προϋποθέτει την εφαρμογή ενός πλαισίου το οποίο θα εισάγει στις ιστοσελίδες πληροφορίες σχετικά με τη δομή και το νόημα των περιεχομένων τους. Ο σημαντολογικός Π.Ι (semantic web) [***v1,v3**], που προορίζεται να είναι η μετεξέλιξη του σημερινού Π.Ι, είναι ένα βήμα προς αυτή την κατεύθυνση και στηρίζεται πρωταρχικά στην χρήση οντολογιών και στους τρόπους με τους οποίους αυτές εφαρμόζονται στον Π.Ι.

Η χρήση του όρου οντολογία είναι δανεισμένη από την Φιλοσοφία όπου εκφράζει θεωρίες σχετικά με την φύση της ύπαρξης, δηλ. τι τύπου πράγματα, οντότητες υπάρχουν [***ΑΙ***]. Η χρήση της στην προκειμένη περίπτωση συνίσταται στην περιγραφή όρων, οντοτήτων και των μεταξύ τους σχέσεων. Η περιγραφή αυτή μπορεί να βρίσκεται αποθηκευμένη σε κάποιο αρχείο το οποίο με την σειρά του ονομάζεται οντολογία. Ουσιαστικά, μια οντολογία είναι μια ιεραρχημένη ταξινόμηση εννοιών, αντικειμένων, κτλ, οριζόμενη από το γενικότερο προς το ειδικότερο και αποτελεί ένα βασικό εργαλείο για την αναπαράσταση γνώσης. Όπως είναι φυσικό, οντότητες οι οποίες αποτελούν ειδικές περιπτώσεις μιας κατηγορίας κληρονομούν τα γενικά στοιχεία αυτής της κατηγορίας. Μπορούν δηλαδή να εκφραστούν κανόνες οι οποίοι περιγράφουν τις ιδιότητες μιας οντότητας και, εν γένει, καλύπτουν θέματα που αφορούν κατηγορίες αντικειμένων, ενέργειες, συμβάντα, πεποιθήσεις και εξαγωγή συμπερασμάτων.

Οι οντολογίες αποτελούν ένα κρίσιμο στοιχείο του σημαντολογικού Π.Ι και μπορούν να εφαρμοστούν σε αυτόν με την χρήση των γλωσσών XML (Extended Markup Language), RDF (Resource Description Framework) και RDFS (RDF Schemas).

ΕΥΦΥΕΙΣ ΠΡΑΚΤΟΡΕΣ ΓΙΑ ΤΟΝ ΠΑΓΚΟΣΜΙΟ ΙΣΤΟ

Η γλώσσα XML διαφέρει από την HTML στο ότι χρησιμοποιεί ετικέτες οι οποίες έχουν άμεση σχέση με το περιεχόμενο της πληροφορίας που περικλείεται από αυτές και στο ότι οι ετικέτες αυτές μπορούν να καθορισθούν από τον χρήστη. Από την σύντομη επαφή μας με την HTML σε προηγούμενη ενότητα είδαμε ότι αυτό δεν συμβαίνει συνήθως εκεί αλλά στην HTML οι ετικέτες δίνουν απλά οδηγίες μορφοποίησης. Με τη χρήση της XML ο δημιουργός της σελίδας μπορεί να εισάγει δομή στα περιεχόμενα της. Το μόνο πρόβλημα όμως είναι ότι η δομή αυτή είναι κατά κάποιο τρόπο αυθαίρετη μιας και δεν είναι απαραίτητο να χρησιμοποιούνται οι ίδιες ετικέτες από όλους. Εκτός από δομή μπορούμε να εισάγουμε και κάποια σχέση νοήματος και περιεχομένου στα κείμενα που γράφονται με XML. Για να συμβεί πλήρως όμως αυτό θα πρέπει να γίνει χρήση της RDF όπου το νόημα εκφράζεται με προτάσεις της μορφής:

<αντικείμενο> <ιδιότητα> <τιμή>

Για παράδειγμα στην πρόταση: «Τα λόγια της πλώρης» δημιουργήθηκαν από τον Καρκαβίτσα, έχουμε ένα αντικείμενο («Τα λόγια της Πλώρης») να έχει μια ιδιότητα (έχει δημιουργηθεί από) η οποία έχει κάποια τιμή (τον Καρκαβίτσα). Αυτή η μάλλον μηχανιστική προσέγγιση περιγραφής δεδομένων μπορεί να βοηθήσει αρκετά στον αυτόματο χειρισμό τους. Με χρήση δε των URI (Universal Resource Identifier) τα οποία (σε αντιστοιχία με τα URL) ταυτοποιούν ένα αντικείμενο, μια ιδιότητα ή μια τιμή και είναι καθολικά διαθέσιμα και χρησιμοποιούμενα, είναι δυνατή μια ενιαία περιγραφή εννοιών και δεδομένων. Για την περιγραφή των κλάσεων, των υποκλάσεων και σχέσεων αυτής της μορφής υπάρχει η RDFS.

Ένα θετικό στοιχείο της όλης υπόθεσης είναι ότι όλα αυτά μπορούν να προστεθούν απλά στην σημερινή μορφή του Π.Ι χωρίς να είναι απαραίτητο να γίνουν ριζικές αλλαγές. Έτσι, με χρήση του συντακτικού της XML μπορούμε να εισάγουμε κοινά αποδεκτές και αναγνωρίσιμες (μέσω των URI) περιγραφές εννοιών και αντικειμένων μέσα σε ιστοσελίδες γραμμένες σε HTML. Η σταδιακή εξέλιξη του Π.Ι σε σημαντικόλογικό Π.Ι θα επιτρέψει στους πράκτορες του Π.Ι να προσφέρουν ένα πλήθος από αναβαθμισμένες υπηρεσίες όπως: ακριβέστερη αναζήτηση, συνδυασμός και εύρεση στοιχείων τα οποία δεν υπάρχουν κατά λέξη σε μια ιστοσελίδα αλλά ο εντοπισμός τους θα απαιτούσε ανάλυση του νοήματος και εξέταση του περιεχομένου μιας σελίδας από κάποιον χρήστη, διαδικασίες εξαγωγής συμπερασμάτων με βάση κάποιες αρχικές πληροφορίες, ικανότητα συνεργασίας και ανταλλαγής πληροφοριών μεταξύ των ευφυών πρακτόρων, ικανότητα «εύρεσης υπηρεσιών» στον Π.Ι όπου προγράμματα θα μπορούσαν να «ανακοινώνουν» τις λειτουργίες και υπηρεσίες που μπορούν να προσφέρουν, κ.ά.

Συμπεράσματα

Ο Παγκόσμιος Ιστός αποτελεί ένα περιβάλλον όπου μπορεί κανείς να συναντήσει πληροφορίες για ο,τιδήποτε. Έχοντας επιφέρει ήδη τεράστιες αλλαγές σε πολλούς τομείς της ζωής μας αποτελεί ταυτόχρονα ένα μοναδικό εργαλείο όταν θέλουμε να είμαστε ενημερωμένοι όχι μόνο για ο,τιδήποτε αφορά αλλά και για τις τελευταίες εξελίξεις σε κάποιο πεδίο. Στον τομέα της εκπαίδευσης η χρήση του συντελεί τα μέγιστα, βοηθώντας τόσο τον εκπαιδευτή όσο και τον εκπαιδευόμενο.

Μέσα σε αυτόν τον τεράστιο όγκο πληροφορίας που υπάρχει θα ήταν δύσκολο να κινηθούμε και να κάνουμε αποδοτική χρήση αυτής αν δεν υπήρχαν οι ιστοσελίδες κατάλογοι του Π.Ι και οι μηχανές αναζήτησης οι οποίες κάνουν χρήση των προγραμμάτων αναζήτησης του Π.Ι που ονομάζονται ευφυείς πράκτορες του Π.Ι. Με τη χρήση τους είναι δυνατή η ταχύτατη εύρεση ιστοσελίδων σχετικά με κάποιο θέμα δίνοντας απλά κάποιες λέξεις κλειδιά. Όμως, όντας κάποιος χώρος όπου η πληροφορία είναι οργανωμένη και τοποθετημένη έτσι ώστε να είναι εύκολα αναγνώσιμη και κατανοήσιμη από τους ανθρώπους, ο αυτόματος χειρισμός αυτής της πληροφορίας μπορεί να γίνει προβληματικός λόγω ασάφειας και έλλειψης ενός προτύπου πλαισίου το οποίο θα βοηθάει την εισαγωγή δομής και νοήματος στις ιστοσελίδες.

Ο σημαντολογικός Π.Ι που θα αποτελέσει την μετεξέλιξη του Π.Ι κάνει χρήση οντολογιών οι οποίες περιγράφουν τα περιεχόμενα των ιστοσελίδων όσον αφορά την δομή και το νόημα τους. Η χρήση των γλωσσών περιγραφής XML, RDF και RDFS είναι καθοριστικής σημασίας και το θετικό σημείο είναι ότι οι μετα-πληροφορίες που εισάγονται για να περιγράψουν τις πληροφορίες που υπάρχουν σε μια ιστοσελίδα μπορούν να συνυπάρξουν με την απλή HTML.

Χρησιμοποιώντας αυτό το νέο πλαίσιο ως περιβάλλον δράσης των ευφύων πρακτόρων θα μπορούμε να έχουμε στη διάθεση μας αναβαθμισμένες υπηρεσίες, ουσιαστικά θα είναι μια προσέγγιση του να έχουμε ζητήσει από κάποιον χρήστη να επιλύσει για μας ένα αριθμό από ζητήματα τα οποία εκτός από απλή αναζήτηση προϋποθέτουν κατανόηση του περιεχομένου, χρήση κανόνων εξαγωγής συμπερασμάτων και σχέσεων ανάμεσα στις οντότητες που περιγράφονται σε μια σελίδα. Αυτό το νέο πλαίσιο ανάπτυξης, σε συνεργασία με τις νέες τεχνολογίες ανάπτυξης του διαδικτύου (π.χ. IPv6 [***IPv6]) θα έρθει με την σειρά του να αποτελέσει ένα ακόμα εργαλείο στην διάθεση μας με στόχο την αποτελεσματική διαχείριση γνώσης.

Αναφορές

[α-διαδίκτυο]

[π.ι-ρομποτ] Fah-Chun Cheong, «Internet Agents: Spiders, Wanderers, Brokers, and Bots»,
New Riders, ISBN 1-56205-463-5.

[v-2] Web Hunting...