

ΠΕΡΙΛΗΨΗ

Οι *Agents* είναι θεωρητικά «έξυπνα» προγράμματα, που βοηθούν το χρήστη να ολοκληρώσει κάποιες ιδιαίτερα χρονοβόρες διεργασίες. Οι δουλειές αυτές των Agents έχουν να κάνουν κυρίως με αναζήτηση, εύρεση και φιλτράρισμα πληροφοριών. Οι ρόλοι που παίζουν οι Agents και οι ιδέες, οι οποίες έχουν αντληθεί από τα θέματα των αυτόνομων πρακτόρων και της Τεχνητής νοημοσύνης, συνδυάζονται και δημιουργούν ένα περιβάλλον αλληλεπίδρασης, άλλοτε συνεργασίας και άλλοτε ανταγωνισμού.

Επίσης παρουσιάζονται τα διαφορετικά είδη Agents που συνθέτουν το μοντέλο των διαφορετικών αναγκών των χρηστών.

Τέλος, παρουσιάζονται αρκετά από τα υπάρχοντα προϊόντα που μπορεί εύκολα κανείς να βρει στο Internet.

Abstract:

Agents are semi-intelligent programs that assist the user in performing repetitive and time-consuming tasks. Information discovery and information filtering are a suitable domain for applying agent technology. Ideas drawn from the field of autonomous agents and artificial life are combined in the creation of an evolving ecosystem composed of competing and cooperating agents. A co-evolution model of information filtering agents that adapt to the various user's interests and information discovery agents that monitor and adapt to the various on-line information sources is presented.

Products found over the Internet are presented and discussed.

Keywords: Agents, SearchBots, World-Wide-Web

ΚΕΦΑΛΑΙΟ 1, ΤΙ ΕΙΝΑΙ AGENTS

ΕΙΣΑΓΩΓΙΚΑ ΣΤΟΙΧΕΙΑ

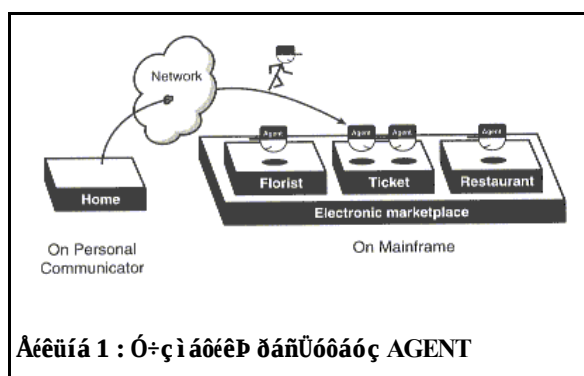
Τα εντυπωσιακά χαρακτηριστικά των Agents είναι για πολλά χρόνια αντικείμενο συζητήσεων στους κύκλους των ειδικών. Ποτέ άλλοτε ένα πεδίο έρευνας είναι τόσο πλούσιο και ποικίλο, ακόμα όμως ατελές, που το κύριο κεφάλαιο της έρευνας αυτής είναι το ερώτημα *τι ακριβώς είναι ένας Agent* ;

Αν το θέσουμε απλά, θα ορίσουμε τους Agents ως προσωπικούς βοηθούς σε επίπεδο software, με «εξουσιοδότηση» από τους χρήστες τους. Παλιοί οραματιστές όπως ο Nicholas Negreponte (1970, 1989) και Alan Kay (1984) ήταν από τους πρώτους που αναγνώρισαν την αξία των προσωπικών βοηθών σε επίπεδο software. Είχαν αναπτύξει την ιδέα της ανάθεσης Agents στο περιβάλλον επικοινωνίας (interface) συγκεκριμένων computer-based εργασιών. Πρόσφατα, αρκετοί κατασκευαστές υπολογιστών έχουν υιοθετήσει αυτήν την ιδέα σαν την ιδανική περιγραφή του περιβάλλοντος εργασίας του μέλλοντος · παράδειγμα, τα videos της Apple (1988). Επίσης , σύμφωνα με τα λόγια του Ted Selker (1994) από το κέντρο ερευνών της IBM στο Almaden, «*οι Agents είναι προγράμματα υπολογιστών που προσομοιώνουν μια ανθρώπινη σχέση, κάνοντας κάτι που διαφορετικά ο άλλος θα έκανε για μας.*» Παρακάτω θα δοθούν αρκετοί ορισμοί για τον όρο Agents, που βρέθηκαν σε διάφορες πηγές.

Η General Magic , μια από τις πρώτες εταιρείες της Silicon Valley, έχει αναπτύξει την απαιτούμενη τεχνολογία της γλώσσας προγραμματισμού των Agents, που την ονομάζει Telescript, υποστηρίζει τη χρήση software Agents σαν «προσωπικούς πράκτορες» μέσα στο δίκτυο. Ορίζει δηλαδή τον όρο αυτό ως κομμάτια κώδικα γραμμένου σε Telescript που στέλνονται σε όλα τα μήκη και πλάτη του δικτύου (White 1994). Ένα πρόγραμμα σε Telescript υπακούει στις οδηγίες του χρήστη για να επιτελέσει τις διάφορες εργασίες που ζητούνται, κινούμενο σε ηλεκτρονικούς δρόμους , καταλήγοντας σε διάφορα «μέρη». Παραδείγματα για τέτοια μέρη είναι τα διάφορα «ηλεκτρονικά γραμματοκιβώτια», ημερολόγια, ηλεκτρονικά μαγαζιά,

σημεία συνάντησης στο δίκτυο, κλπ.

Όπως φαίνεται στο παρακάτω σχήμα, χρήστες που δημιουργούν Telescript Agents, φαντάζονται τα προγράμματα αυτά ως ηλεκτρονικές προεκτάσεις του εαυτού τους, ικανές να μαζεύουν επιλεκτικά πληροφορίες, να κάνουν διαπραγματεύσεις, καθώς και να εκτελούν οικονομικές συνδιαλλαγές εκ μέρους τους.



Προφανώς τα προγράμματα αυτά μπορούν να παραμετροποιηθούν καταλλήλως. Ταυτόχρονα έχουν και μια δόση «ευφυΐας», με την έννοια ότι είναι δυνατό να προσαρμόζονται στις συνθήκες του περιβάλλοντος, και να ενεργούν ανάλογα με τα «μέρη» στα οποία βρίσκονται · π.χ. αλλάζοντας από ένα κύριο πλάνο δράσης στο εναλλακτικό του.

ΟΡΙΣΜΟΣ 1ος

Ερμηνείες λεξικών (Webster's New World Dictionary) θέλουν το λήμμα Agent ως: « Ένα άτομο ή μια οντότητα, η οποία πράττει ή είναι ικανή να πράξει ή είναι εξουσιοδοτημένη να πράξει εκ μέρους κάποιου άλλου.»¹

Σ'αυτή την ερμηνεία τονίζονται τρία πράγματα: το ότι ένας Agent αποτελεί μια ξεχωριστή οντότητα, με αρκετή έως απόλυτη αυτονομία, το ότι κάνει κάποιες ενέργειες και το ότι ένας Agent ενεργεί εκ' μέρους κάποιου άλλου.

ΟΡΙΣΜΟΣ 2ος

Όσον αφορά στον τομέα της πληροφορικής, βρέθηκε ο ορισμός της

¹ A person or thing that acts or is capable of acting or is empowered to act, for another.

λέξης Software Agent:

«Μια υπολογιστική μονάδα (κομμάτι ή λογισμικό), η οποία αναλαμβάνει κάποια εξειδικευμένα καθήκοντα αυτόνομα.»²

Ο N. Negroponte, διευθυντής του AI Lab του MIT έχει πει αναφορικά με το συγκεκριμένο θέμα:

«When I want to go out to the movies, rather than read reviews, I ask my mother-in-law. We all have an equivalent who is both an expert on movies and an expert on us. What we need to built is a digital mother-in-law»

Στο παραπάνω απόσπασμα γίνεται αναφορά σε έναν ηλεκτρονικό βοηθό, ο οποίος γνωρίζει πολύ καλά ένα συγκεκριμένο αντικείμενο, γνωρίζει όμως πολύ καλά και μας.

Η «ηλεκτρονική πεθερά» του Negroponte αποτελεί ένα από τα πιο επίκαιρα αντικείμενα επιστημονικής μελέτης σήμερα.

ΟΡΙΣΜΟΣ 3ος

Σύμφωνα με το βιβλίο «Internet Agents, Spiders, Wanderers, Brokers, and Bots» (σ.5) του Fah-Chun Cheong, το λήμμα Agent δηλώνει:

«Λογισμικό που λειτουργεί ως προσωπικός βοηθός με ξεχωριστές αρμοδιότητες από τον χρήστη του.»³

ΟΡΙΣΜΟΣ 4ος

Στην ηλεκτρονική διεύθυνση www.whatis.com , όσον αφορά το λήμμα Agent σε σχέση με το WWW βρέθηκε ο ορισμός :

«Ένας Agent είναι ένα πρόγραμμα το οποίο συγκεντρώνει πληροφορίες ή εκτελεί κάποιες άλλες λειτουργίες χωρίς την άμεση επίβλεψη του χρήστη.»⁴

ΣΥΝΟΨΗ:

Βλέπουμε ότι ο ορισμός 4, θα μπορούσε να χρησιμοποιηθεί από πολλούς προγραμματιστές, οι οποίοι θα ισχυριστούν ότι δημιούργησαν έναν

² A computing entity (piece of software) that performs user delegated tasks autonomously.

³ Agents can be considered personal software assistants with authority delegated from their users.

⁴ On the Internet, an agent is a program that gathers information or performs some other service without your immediate presence and on some regular schedule.

Software Agent.

Υπάρχουν όμως σημαντικές διαφορές ενός Agent με άλλα συμβατικά προγράμματα .

ΒΑΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ SOFTWARE AGENTS

Ένας βασικός Software Agent στηρίζεται σε τρία χαρακτηριστικά:

1. την αυτονομία (autonomy).
2. τη δυνατότητα αντίδρασης και προσαρμοστικότητας (reactive and adaptive).
3. τη δυνατότητα επικοινωνίας (communication ability).

Η έννοια της αυτονομίας δηλώνει ότι ο Agent λειτουργεί ανεξάρτητα από τον χρήστη και ευθύνεται ο ίδιος για τον έλεγχο των πράξεών του. Η δυνατότητα αντίδρασης και προσαρμοστικότητας δείχνουν ότι αισθάνεται συνεχώς το εξωτερικό του περιβάλλον και ότι προσαρμόζει συνεχώς τη συμπεριφορά του σύμφωνα με τις αλλαγές που γίνονται σ' αυτό. Τέλος, η δυνατότητα επικοινωνίας δείχνει ότι μπορεί να συνεργάζεται με άλλες οντότητες, είτε αυτές είναι κάποιοι χρήστες, είτε είναι άλλοι Agents, είτε κάποια αντικείμενα (objects), βάσεις δεδομένων, κλπ.

Αν προσθέσουμε στα παραπάνω χαρακτηριστικά του Software Agent την ικανότητα να θέτει στόχους, να διατηρεί μοντέλα πεποιθήσεων, να επεξεργάζεται και να κρίνει τις πράξεις του καθώς και τις πράξεις άλλων Agents (συμπεριλαμβανομένων και των χρηστών), και τέλος να μαθαίνει από τη συμπεριφορά του και να βελτιώνεται, τότε έχουμε έναν Intelligent Agent. Στην επόμενη ενότητα θα δούμε αναλυτικότερα τα χαρακτηριστικά ενός Intelligent Agent.

Σε γενικές γραμμές, πολλοί Intelligent Agents βασίζονται στην ιδέα ότι ο χρήστης πρέπει να ορίζει απλώς έναν «ανώτερο» στόχο δίχως να ασχολείται με επιμέρους εντολές, και ο Intelligent Agent χρησιμοποιώντας ρουτίνες που του χαρίζουν «νοημοσύνη» να αναλαμβάνει την ολοκληρωμένη υλοποίησή του χειρίζοντας το «πώς» και το «πότε» θα πραγματοποιήσει αυτό το στόχο.

Η τεχνολογία των Intelligent Agents, όπως γίνεται αντιληπτό, βασίζεται και πηγάζει από τη θεωρία της Τεχνητής Νοημοσύνης. Δεν είναι τυχαίο ότι Πανεπιστημιακή έρευνα πάνω στον τομέα αυτό γίνεται κυρίως σε εργαστήρια Τεχνητής Νοημοσύνης. Η επιστήμη των Agents μπορεί να θεωρηθεί από ακραίους ακόμα και ως η συνέχεια και το μέλλον της Τεχνητής Νοημοσύνης, ιδιαίτερα σε περιπτώσεις όπου χρησιμοποιούνται δίκτυα και επικοινωνίες.

ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΕΝΟΣ «ΕΞΥΠΝΟΥ» AGENT

Τα παρακάτω χαρακτηριστικά συνθέτουν την «προσωπικότητα» ενός Intelligent Agent. Βεβαίως, υπάρχουν Intelligent Agents, οι οποίοι δεν διαθέτουν όλα τα παρακάτω, έχουν όμως την ίδια ιδεολογία στον σχεδιασμό και τους στόχους τους με άλλους Intelligent Agents, οπότε και ανήκουν στην ίδια κατηγορία προγραμμάτων.

Αυτονομία (Autonomy):

Ο Intelligent Agent λειτουργεί αυτόνομα από τον χρήστη, στο βαθμό που καθορίζει ο ίδιος ο χρήστης. Η αυτονομία δηλαδή μπορεί να διαφέρει από Agent σε Agent. Για παράδειγμα υπάρχουν Agents οι οποίοι μπορούν αυτόνομα να δημιουργήσουν ένα απλό βραδινό backup και Agents οι οποίοι διαπραγματεύονται μόνοι τους τις τιμές αγοράς προϊόντων από on-line κέντρα πωλήσεων.

Αφοσίωση στο χρήστη (Delegation):

Η έννοια της αυτονομίας μπορεί να δίνει στο πρόγραμμα την ευχέρεια να λαμβάνει μόνο του τις περισσότερες αποφάσεις, απαλλάσσοντας έτσι τον χρήστη από βαρετές, καθημερινές και επαναλαμβανόμενες δουλειές. Δεν πρέπει να ξεχνάμε όμως ότι η εξουσία του Intelligent Agent περιορίζεται τελικά από τον χρήστη του γιατί αυτός θέτει πάντα τους τελικούς στόχους. Ένας Agent λειτουργεί πάντα για λογαριασμό του χρήστη του (ή ενός άλλου Agent, αν αυτός τον χρησιμοποιεί) και οι ενέργειές του θα πρέπει να

συμφωνούν με τις ενέργειες που θα έκανε και ο ίδιος ο χρήστης (ή ο άλλος Agent).

Δυνατότητες επικοινωνίας (Communication skills):

Ο Agent πρέπει να αλληλεπιδρά με τον χρήστη για να παίρνει οδηγίες από αυτόν καθώς και να τον ενημερώνει για την πορεία των εργασιών του. Επίσης ένας Agent μπορεί να αλληλεπιδρά με άλλους Agents για την επίτευξη ενός κοινού στόχου. Παρακάτω θα δούμε πώς μπορούν να συνεργαστούν πολλοί Agents για την επίτευξη ενός κοινού στόχου.

Παρακολούθηση (Monitoring):

Ένας Agent χρειάζεται να μπορεί να παρακολουθεί το περιβάλλον του για να λειτουργήσει αυτόνομα.

Προσαρμοστικότητα (Adaptive):

Οι Intelligent Agents προσαρμόζονται αυτόματα στις προτιμήσεις των χρηστών τους λαμβάνοντας υπ' όψη προηγούμενες εμπειρίες. Επίσης μπορούν και προσαρμόζονται αυτόματα στο περιβάλλον τους (από τη στιγμή που έχουν τη δυνατότητα της συνεχούς παρακολούθησής του).

Νοημοσύνη (Intelligence):

Ένας Intelligent Agent πρέπει να «καταλαβαίνει» τα συμβάντα που παρακολουθεί και να παίρνει αποφάσεις για να λειτουργήσει αυτόνομα. Η θεωρία της Τεχνητής Νοημοσύνης, των Νευρωνικών Δικτύων και της Ασαφούς Λογικής (Fussy Logic) συναντάται συνεχώς και στο σύνολο των Intelligent Agents.

Δυναμικότητα (Temporal continuity):

Ένας Agent χρειάζεται να επεξεργάζεται συνεχώς δεδομένα σε αντιδιαστολή με τα απλά προγράμματα. Ένα απλό πρόγραμμα κατά την εφαρμογή του συγκεντρώνει κάποια Inputs (στοιχεία προς επεξεργασία), τα επεξεργάζεται και εμφανίζει στον χρήστη τα αποτελέσματα (Outputs).

«Τρέχει» δηλαδή μια φορά και δίνει «στατικά» αποτελέσματα . Ένας Agent που παρακολουθεί για παράδειγμα τις τιμές μιας μετοχής στο χρηματιστήριο και ερευνά το αν θα πρέπει να επενδύσουμε στην μετοχή αυτή ή όχι, χρειάζεται συνεχή ενημέρωση από το περιβάλλον του. Δηλαδή «τρέχει» συνεχώς και δίνει διαφορετικά αποτελέσματα σε διαφορετικές χρονικές στιγμές.

Χαρακτήρας (Character):

Μερικοί Intelligent Agents έχουν μια καλά-καθορισμένη «προσωπικότητα» και ψυχολογική κατάσταση.

Δυνατότητα μεταφοράς (Mobile):

Μερικοί Agents έχουν τη δυνατότητα να μεταφέρονται από ένα υπολογιστικό σύστημα σε ένα άλλο χρησιμοποιώντας στο μέγιστο τις δυνατότητες των δικτύων και του Internet. Αυτό το χαρακτηριστικό το συναντάμε πολύ συχνά στους λεγόμενους «Internet Agents», «Web Agents», «Spiders», «WebCrawlers», ή «Worms».

ΠΛΕΟΝΕΚΤΗΜΑΤΑ

Τα παραπάνω χαρακτηριστικά δίνουν στους Intelligent Agents μοναδικές δυνατότητες σε σχέση με τα άλλα προγράμματα. Παρακάτω αναλύονται τα πλεονεκτήματα της χρήσης τους:

1.Πολυπλοκότητα της συναλλαγής.

Ένας Intelligent Agent ταιριάζει περισσότερο στο χειρισμό σύνθετων συναλλαγών. Μια απλή συναλλαγή δεν χρειάζεται «νοημοσύνη». Για παράδειγμα για την αγορά ενός μουσικού CD από ένα on-line κατάστημα του Internet, η μόνη μεταβλητή που θα πρέπει να υπολογιστεί από τον Agent θα είναι η τιμή του CD. Όταν όμως θα πρέπει να λάβουμε υπ'όψη μας και άλλους παράγοντες, τότε η «νοημοσύνη» είναι απαραίτητη για να υλοποιήσουμε σωστά τις συναλλαγές μας.

2.Μονοδιάστατα και Πολυδιάστατα προβλήματα.

Τα πολυδιάστατα προβλήματα είναι πιο δύσκολα στην επίλυσή τους από τον άνθρωπο. Όσο αυξάνει ο αριθμός των μεταβλητών, τόσο πιο σύνθετο γίνεται και το πρόβλημα. Τα συστήματα στήριξης αποφάσεων είναι σχεδιασμένα για να βοηθούν τους ανθρώπους στις αποφάσεις τους. Ένας Intelligent Agent όμως θα δώσει καλύτερες λύσεις γιατί λαμβάνει υπ' όψη του τις προσωπικές συνήθειες του χρήστη. Για παράδειγμα, ένας Agent υπεύθυνος για αυτόματες κρατήσεις δωματίων σε ξενοδοχεία, θα υπολογίσει τις ιδιαιτερότητες του χρήστη του για δωμάτιο που βρίσκεται κοντά στο αεροδρόμιο, κοντά στην έκθεση ή την εταιρία που θα επισκεφτεί ο χρήστης, Α κατηγορίας, με χρέωση για πρωινό μόνο αν ταιριάζουν οι ώρες της επίσκεψης του χρήστη στην πόλη, κλπ.

3.Κόστος της αναζήτησης.

Ένα σημαντικό πλεονέκτημα της χρήσης ενός Intelligent Agent είναι η μείωση των εξόδων για αναζήτηση πληροφοριών. Αν οι πληροφορίες που χρειαζόμαστε είναι δυσεύρετες και απαιτούν πολύ χρόνο για τη συγκέντρωση και την επεξεργασία τους, τότε η χρήση των Intelligent Agents είναι αναγκαία. Μια απλή αναζήτηση σε κάποια βάση δεδομένων δεν θεωρούμε ότι λειτουργεί με «νοημοσύνη». Ένας Intelligent Agent όμως αναζητά πληροφορίες σε πολλαπλές βάσεις δεδομένων και τις συνδυάζει κατάλληλα για να τις επεξεργαστεί και να μας δώσει το σωστό αποτέλεσμα.

4.Αυτόματη εκμάθηση και τυποποιημένη γνώση.

Ένα απλό πρόγραμμα αναζήτησης χρησιμοποιεί τυποποιημένες «Βάσεις Γνώσεων» (Knowledge Bases) για να αναπαραστήσει τις πληροφορίες. Αντίθετα, ένας Intelligent Agent χρησιμοποιεί ρουτίνες Τεχνητής Νοημοσύνης για να «μαθαίνει», ανάλογα με την προηγούμενη συμπεριφορά του χρήστη.

5.Αριθμός των συμβαλλόμενων (parties).

Ο αριθμός των συμβαλλόμενων στοιχείων ενός προβλήματος σχετίζεται άμεσα με την πολυπλοκότητα του προβλήματος αυτού. Ένα Infobot (απλό πρόγραμμα χωρίς νοημοσύνη που συλλέγει πληροφορίες από το Internet) μπορεί να συναλλάσσεται με διαφορετικούς φορείς αλλά θα πρέπει όλοι αυτοί να «μιλάνε» την ίδια γλώσσα. Ένας Intelligent Agent μπορεί να επικοινωνεί με τελείως διαφορετικούς φορείς (π.χ ξενοδοχεία, αεροπορικές εταιρίες, θέατρα, κλπ) για να πετύχει τον στόχο του.

6.Το στοιχείο του Χρόνου.

Ένας Intelligent Agent σε αντίθεση με ένα απλό πρόγραμμα, πρέπει να παρακολουθεί συνεχώς το περιβάλλον του και να ανταποκρίνεται στις αλλαγές. Για παράδειγμα ένας Intelligent Agent που ενημερώνει τον χρήστη του για τις καθημερινές ειδήσεις από πηγές του Internet, πρέπει να επεξεργάζεται τα δεδομένα του κάθε μέρα.

7.Τοπικό και Μακρινό.

Ένα απλό πρόγραμμα λειτουργεί συνήθως «τοπικά», στον υπολογιστή του κάθε χρήστη. Ένας Intelligent Agent μπορεί να χρησιμοποιήσει τεχνολογίες δικτύων (όπως το Internet) και να βρίσκεται μακριά από τον χρήστη του, απαλλάσσοντάς τον από χρόνο και υπολογιστική ισχύ.

ΜΕΙΟΝΕΚΤΗΜΑΤΑ

- Οι Intelligent Agents είναι προγράμματα σχεδιασμένα από ανθρώπους, όπως και όλα τα άλλα κατασκευάσματα της Τεχνητής Νοημοσύνης. Είναι πιθανό λοιπόν, να υπάρχουν λάθη στο σχεδιασμό και στην επιλογή των παραμέτρων τους. Έτσι, μπορούν να παρουσιάσουν προβλήματα στην επεξεργασία και γενικότερα διαχείριση των «γνώσεων» ενός Intelligent Agent, με αποτέλεσμα να δώσει λανθασμένα αποτελέσματα και να μη συμβουλέψει σωστά τον χρήστη.
- Η χρήση των Intelligent Agents προϋποθέτει μεγάλη υπολογιστική ισχύ

για να έχουμε τα επιθυμητά αποτελέσματα στο χρόνο τους. Η τοπική χρήση προηγμένων Intelligent Agents είναι σχεδόν απαγορευμένη σε προσωπικούς υπολογιστές για το λόγο αυτό.

- Υπάρχουν Intelligent Agents, οι οποίοι δεν δίνουν την επιλογή στον χρήστη να παρεμβαίνει ριζικά στις αποφάσεις τους. Διαθέτουν δηλαδή αυξημένη αυτονομία από τον χρήστη και έτσι μπορεί να οδηγηθούν σε υπερβολές.
- Η τεχνολογία των Intelligent Agents δεν είναι διαδεδομένη, παρά μόνο σε στρατιωτικούς οργανισμούς και σε πανεπιστημιακά εργαστήρια. Στον επιχειρηματικό τομέα, πέρα από τον κλάδο των ηλεκτρονικών υπολογιστών, δύσκολα συναντάμε Intelligent Agents (τουλάχιστον σήμερα).
- Ένα ακόμη μειονέκτημα των Intelligent Agents, ιδιαίτερα των κινητών (mobile), είναι η ασφάλεια. Μερικοί Web Administrators δεν θέλουν Intelligent Agents να βρίσκονται στο σύστημά τους και να έχουν διάφορες εξουσίες για λόγους ασφαλείας.
- Μερικοί Intelligent Agents που χρησιμοποιούν το Internet ως πηγή γνώσης τους μπορεί να έχουν περιορισμένη πρόσβαση σε συγκεκριμένες πηγές. Για παράδειγμα ένας μικρός λιανοπωλητής που διατηρεί on-line βιβλιοπωλείο στο Internet μπορεί να μη θέλει να «εκθέσει» τις τιμές των βιβλίων του σε έναν Intelligent Agent για να αποφύγει τον ανταγωνισμό από τα μεγαλύτερα βιβλιοπωλεία, τα οποία έχουν καλύτερες τιμές και θα προτιμηθούν από τον Agent.
- Άλλος ένας λόγος για τον οποίο η χρήση των Intelligent Agents στο Internet είναι αμφιλεγόμενη είναι οι υψηλές απαιτήσεις τους σε ταχύτητα και συνεπώς σε εύρος δικτύου (bandwidth).

ΚΕΦΑΛΑΙΟ 2, ΤΟ WEB & ΟΙ Agents

Τα τελευταία χρόνια το World Wide Web (WWW) έχει μεταμορφώσει το Internet. Η ύπαρξη του Web μετέτρεψε το Internet από ένα αποκλειστικό club μιας χώρας στο οποίο σύχναζαν μόνο λίγοι «άριστα συνδεδεμένοι», σε έναν τεράστιο δημόσιο χώρο, όπου τον επισκέπτονται καθημερινά άνθρωποι από όλα τα κοινωνικά και επαγγελματικά στρώματα .

Όλα αυτά έγιναν, εξαιτίας του γραφικού περιβάλλοντος εργασίας και πρόσβασης στις πληροφορίες κατά μήκους του Internet (GUI - Graphical User Interface). Το interface αυτό επιτρέπει τους χρήστες να δουν «αξιοθέατα», να ακούσουν ήχους, με έναν ιδιαίτερα καλαίσθητο τρόπο πλοήγησης.

Το World Wide Web λοιπόν, άνοιξε τις πόρτες του Internet στις μάζες . Από τη στιγμή που οι μάζες αυτές είναι διψασμένες για πληροφορία, συχνά θα χρησιμοποιούνται πλήθος Agents με στόχο την αυτοματοποίηση διαφόρων εργασιών. Για να κατανοηθεί ακριβώς αυτή η έντονη επίδραση του WWW στην ευρεία χρήση και εξάπλωση των Agents, είναι χρήσιμο να αναφερθούν κάποια θέματα κλειδιά του World Wide Web.

ΑΝΑΠΤΥΞΗ ΤΟΥ WEB

Ο προκάτοχος του WWW ήταν ένα μικρό σύστημα, σχεδιασμένο για τοπική και προσωπική χρήση, δημιουργημένο στην CERN, στο ευρωπαϊκό εργαστήριο Φυσικής της Geneva, με στόχο να κρατά τις τοπικές πληροφορίες ενός κατακευματισμένου project. Τα αποτελέσματα ήταν τόσο πολύ θετικά, που προέτρεψαν τους σχεδιαστές - προγραμματιστές των συστημάτων αυτών να ασχοληθούν και να αναπτύξουν την ιδέα αυτή, κάτι που ουσιαστικά οδήγησε και στο σημερινό WWW. Το Μάρτιο του 1989, ο Tim Berneslee στη CERN, άρχισε να κυκλοφορεί την πρότασή του για το χτίσιμο

ενός hypertext⁵ συστήματος, που θα παρείχε εύκολο διαμοιρασμό (sharing) της πληροφορίας μεταξύ γεωγραφικά χωρισμένων ομάδων ερευνητών.

Τον Οκτώβριο του 1990, άρχισε η ουσιαστική ανάπτυξη του Web και η ιδέα να παίρνει σάρκα και οστά . Μέχρι τα Χριστουγεννα του 1990, η πρόσβαση στα hypertext αρχεία καθώς και στις ομάδες ειδήσεων (newsgroups) του Internet γινόταν μέσω τα προγράμματα της NeXTStep σε περιβάλλον γραμμής εντολής ή γραφικών. Λίγο πριν το τέλος του 1991, η CERN ανακοίνωσε και παρουσίασε το Web στον κόσμο. Από εκεί και πέρα η αύξηση της χρήσης και της αξίας του Web ήταν αλματώδης.

ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΤΟΥ WEB

Το WWW οργανώνει, μεταδίδει και ανακτά πληροφορία κάθε τύπου, χρησιμοποιώντας ένα συνδυασμό τεχνολογιών από πλευράς hypertext, γραφικών και πολυμέσων (multimedia) · οι τεχνολογίες αυτές συνενωμένες σε ένα σύνολο που περιέχει συμβάσεις για τα ονόματα (name conventions), πρωτόκολλα επικοινωνίας των διαφορετικών δικτύων (network protocols) και συγκεκριμένες μορφοποιήσεις (formats) για τα κείμενα και τα διάφορα έγγραφα που κυκλοφορούν, εφαρμόζονται με τη χρήση μιας αρχιτεκτονικής πελάτη - υπηρέτη (client - server architecture).

Η αρχιτεκτονική του Web είναι έντονα ιεραρχικά - κλιμακωτά σχεδιασμένη. Το περιεχόμενό της είναι όλος ο κόσμος της δικτυακά προσβάσιμης πληροφορίας, την οποία οι δημιουργοί του Web έχουν ορίσει ως «η ενσάρκωση της ανθρώπινης γνώσης» .

Παράλληλα με το Web, υπάρχουν και άλλα πληροφοριακά συστήματα όπως το Gopher και το WAIS, που χρησιμοποιούν μια παρόμοια client - server αρχιτεκτονική. Αυτά τα συστήματα παίζουν ξεχωριστό ρόλο και έχουν διαφορετικούς στόχους · το Gopher, που είναι περίπου σαν το Web χωρίς την ύπαρξη πλήρους hypertext περιβάλλοντος, χρησιμοποιεί ένα δομημένο σύστημα επιλογών (menu) που επιτρέπει την ιεραρχική οργάνωση της πληροφορίας σε καταλόγους - φακέλους. Το WAIS δεν παρέχει

⁵ Hypertext = Ὁδᾱῖᾱῖᾱῖᾱῖᾱῖ, äç. ῆᾱῖᾱῖᾱῖᾱῖᾱῖ ἰᾱ ὁᾱῖᾱῖᾱῖᾱῖᾱῖ - ᾱᾱῖᾱῖᾱῖᾱῖᾱῖ (links) ᾱῖᾱῖ ἰᾱᾱῖᾱῖ ᾱᾱῖᾱῖᾱῖ ᾱᾱῖᾱῖᾱῖ ᾱᾱῖᾱῖᾱῖ ᾱᾱῖᾱῖᾱῖᾱῖᾱῖ, ῆᾱῖᾱῖᾱῖᾱῖᾱῖᾱῖ ὁᾱῖᾱῖᾱῖᾱῖᾱῖ.

δυνατότητες πλοήγησης, αλλά παρέχει καλά δομημένα ευρετήρια, αποκλειστικά για να μεταφέρει τους χρήστες στην επιθυμητή τοποθεσία του «πληροφοριακού χώρου». Κατ'αναλογία με τη δομή ενός κανονικού βιβλίου, στον πληροφοριακό αυτό χώρο το Gopher αποτελεί τον πίνακα των περιεχομένων, το WAIS τις σελίδες ευρετηρίων ενώ το Web αναλογεί στο υπερ-κείμενο του βιβλίου, όπου και βρίσκεται ο όγκος των πληροφοριών.

Τέλος ο κύριος όγκος του software αντιλαμβάνεται το Web σαν μια συμπαγής μορφή. Η αρχιτεκτονική του software λοιπόν, αποτελείται από τα ακόλουθα στοιχεία που αλληλεπιδρούν στο Internet :

- ♦ Πελάτες (Clients), που επιτρέπουν τους χρήστες να πλοηγούνται μέσα στο Web ή ακόμη και να αλληλεπιδρούν με τους Servers με έναν ιδιαίτερα ενδιαφέρον τρόπο.
- ♦ Εξυπηρετητές - Υπηρέτες (Servers), που επιτρέπουν τις διάφορες τοποθεσίες (sites) στο Internet να δημοσιεύουν και να εξάγουν πληροφορία.
- ♦ Proxies, που διευκολύνουν την επικοινωνία και παρέχουν έλεγχο της πρόσβασης σε Sites που βασίζονται σε άμεση επικοινωνία με το Internet (για παράδειγμα sites πίσω από firewall).

ΠΡΩΤΟΚΟΛΛΟ ΕΠΙΚΟΙΝΩΝΙΑΣ : HTTP

Το HTTP (HyperText Transfer Protocol) είναι ένας τρόπος πρόσβασης στους Web servers. Υιοθετεί ένα ανγνωστό στυλ κειμένου σχεδόν παρόμοιο με το αντίστοιχο του πρωτοκόλλου μεταφοράς αρχείων (FTP - File Transfer Protocol) και του πρωτοκόλλου μεταφοράς ειδήσεων (NNTP - Network News Transfer Protocol) , που έχουν χρησιμοποιηθεί με επιτυχία εδώ και χρόνια. Το HTTP δεν είναι τόσο ένας τρόπος μεταφοράς hypertext αρχείων, όπως ίσως υπονοούν τα αρχικά του, αλλά κυρίως ένας τρόπος ανάγνωσης και μεταφοράς πληροφορίας με βάση την αποτελεσματικότητα των hypertext διακλαδώσεων. Τα δεδομένα που μπορούν έτσι να μεταφερθούν είναι σχεδόν τα πάντα, για παράδειγμα απλό κείμενο, αρχεία εικόνων, αρχεία video και ήχου .

Επίσης το HTTP είναι ένα απλό πρωτόκολλο που κάνει αίτηση για κάτι και περιμένει ανταπόκριση. Ουσιαστικά υπάρχουν τέσσερα βήματα μιας συναλλαγής με τον τρόπο αυτό : Σύνδεση, αίτηση, απόκριση και κλείσιμο της επικοινωνίας.

HTML: Η ΓΛΩΣΣΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ

Η HTML (HyperText Markup Language) είναι μια απλή βασική γλώσσα προγραμματισμού hypertext αρχείων. Περιγράφει δηλαδή, την οργάνωση και τη δομή ενός κειμένου. Είναι ειδικά σχεδιασμένη με γνώμονα την απλούστευση της παραγωγής από τον άνθρωπο και κυρίως τον υπολογιστή.

Η ιδέα πίσω από την HTML είναι να μορφοποιείται η πληροφορία με τέτοιο τρόπο ώστε να είναι εφικτά η ηλεκτρονική ταξινόμηση, το ψάξιμο και η ανάκτηση των δεδομένων και της πληροφορίας. Πρέπει δηλαδή να μην υπάρχει εξάρτηση του τρόπου εμφάνισης της πληροφορίας με τον τρόπο εμφάνισης του κάθε διαφορετικού εγγράφου - κειμένου. Αυτό το αποτέλεσμα διευκολύνει προφανώς την παραγωγή hypertext αρχείων με τη χρήση HTML.

Αντίθετα, το format των αρχείων που παράγονται από έναν συμβατικό επεξεργαστή κειμένου, υπαγορεύει την εμφάνιση του κειμένου με ένα συγκεκριμένο τρόπο, που συνεπάγεται ασυμβατότητα με άλλα σχετικά προγράμματα. Η HTML δεν υπαγορεύει, αλλά από την άλλη μεριά προτείνει την καταλληλότερη παρουσίαση για ένα κείμενο. Επικεντρώνεται μόνο στη δημιουργία σωστής δομής, οπότε και δεν έχει σημασία ποιος φυλλομετρητής ιστοσελίδων (WEB browser) χρησιμοποιείται.

ΔΙΑΔΡΑΣΤΙΚΟΤΗΤΑ ΣΤΟ Web

Η πραγματική δύναμη του Web βρίσκεται στη δυνατότητα έκφρασης. Όπως είναι κατανοητό, από τη στιγμή που ολόκληρη η ποσότητα της πληροφορίας είναι κατανεμημένη με τη μορφή των hypertext κειμένων, θα πρέπει ο χρήστης να εκφράζει τις ανάγκες του κάνοντας ορισμένα μόνο κλικ με το mouse του Η/Υ του, σε ποικίλες επιλογές. Φόρμες, με διάφορα

αντικείμενα όπως πλήκτρα, πλαίσια ελέγχου, αναδυόμενα menu, τις οποίες ο χρήστης συμπληρώνει, εικόνες που λειτουργούν ως `imagemaps`, είναι μόνο κάποιες από τις τεχνικές βελτιώσεις που έχουν γίνει και έχουν ουσιαστικά πλουτίσει την «εκφραστικότητα» του Web, προσφέροντας επιπλέον δυνατότητα διαδραστικότητας.

Οι προσπάθειες του εμπορικού και επιχειρηματικού κόσμου είναι επικεντρωμένες γύρω από τη μετατροπή του Web σε μια βάση για ηλεκτρονικό εμπόριο (e-commerce), με ταυτόχρονη επίτευξη υψηλών στάνταρ ασφάλειας των συναλλαγών και ηλεκτρονικής πληρωμής. Οι περισσότερες εφαρμογές e-commerce απαιτούν την απόδειξη γνησιότητας των δύο πλευρών που συναλλάσσονται καθώς και την ασφαλή ανταλλαγή κρίσιμων στοιχείων. Οι Agents στη σύγχρονη μορφή τους είναι αρκετά ικανοί να διεκπεραιώσουν αυτές τις δουλειές. Ταυτόχρονα, το Web έχει γίνει η ιδανική πλατφόρμα για τους λεγόμενους «εμπορικούς» Agents, οι οποίοι παρέχουν ενδιαφέρουσες υπηρεσίες όπως κυνήγι ευκαιριών, παρακολούθηση μετοχών, ανταλλαγές, μεσιτείες.

ΚΕΦΑΛΑΙΟ 3, ΤΑ ΔΙΑΦΟΡΕΤΙΚΑ ΕΙΔΗ Agents

ΜΗΧΑΝΕΣ ΑΝΑΖΗΤΗΣΗΣ

Κάθε έμπειρος χρήστης του Internet γνωρίζει λίγο πολύ τι σημαίνει μηχανή αναζήτησης. Αναφερόμαστε στις ηλεκτρονικές διευθύνσεις www.yahoo.com, www.webcrawler.com, www.lycos.com, www.hotbot.com, κλπ. Η ύπαρξη μιας τέτοιας υπηρεσίας αποσκοπεί στην εξυπηρέτηση του απλού χρήστη, ο οποίος ψάχνει για μια οποιαδήποτε πληροφορία, αλλά και κάποιου που κάνει εξεζητημένες αναζητήσεις άρθρων, ειδήσεων, κλπ. Οι σημαντικότερες μηχανές αναζήτησης βρίσκονται στον παρακάτω πίνακα.

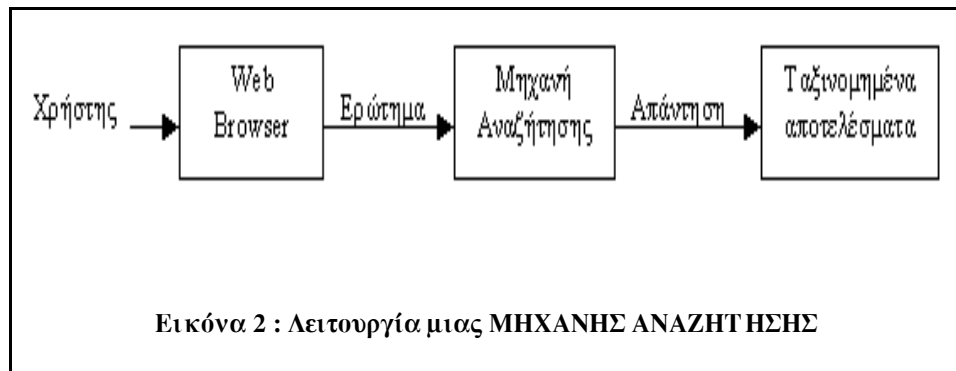
Η λειτουργία των μηχανών αναζήτησης είναι πολύ απλή. Κάθε μηχανή αναζήτησης περιέχει μια βάση δεδομένων με στοιχεία για τα περιεχόμενα κάποιων κόμβων στον Παγκόσμιο Ιστό (Internet). Τα στοιχεία αυτά είναι συνήθως τα παρακάτω:

- Τα URLs (Universal Resource Locator) των σελίδων που περιέχονται σε έναν κόμβο.
- Οι τίτλοι των σελίδων που περιέχονται σε έναν κόμβο.
- Η επικεφαλίδα της κάθε σελίδας.
- Μικρές περιγραφές των περιεχομένων του κόμβου.
- Αντιπροσωπευτικές λέξεις-κλειδιά για κάθε σελίδα.
- Συνήθως οι πρώτες 100 με 200 λέξεις της κάθε σελίδας.

Μερικές ισχυρές μηχανές αναζήτησης διατηρούν στη βάση δεδομένων τους όλο το περιεχόμενο των σελίδων ενός κόμβου (χωρίς βέβαια τις εικόνες). Η βάση αυτή ενημερώνεται και εμπλουτίζεται καθημερινά, με τη συνδρομή των διαχειριστών των εγγεγραμμένων κόμβων και με τη χρήση των λεγόμενων web robots (τα οποία θα μελετηθούν αργότερα).

Τι πρέπει να κάνει ο χρήστης για να βρει την πληροφορία που τον ενδιαφέρει; Αρχικά, πρέπει να συνδεθεί με μια μηχανή αναζήτησης, μέσω του φυλλομετρητή του (Web browser). Έπειτα, πρέπει να στείλει μια ή και περισσότερες λέξεις-κλειδιά που περιγράφουν την πληροφορία που ψάχνει

να βρει στη μηχανή αναζήτησης. Η μηχανή αυτή αναζητεί τις συγκεκριμένες λέξεις-κλειδιά στη βάση δεδομένων της και επιστρέφει στο χρήστη τα αποτελέσματα, τα URLs δηλαδή των διευθύνσεων που περιέχουν αυτές τις λέξεις-κλειδιά.



Γενικά, η ισχύς και η επιτυχία μιας μηχανής αναζήτησης κρίνεται συνήθως από δύο δείκτες. Τον δείκτη ακριβείας (precision) και τον δείκτη ανάκλησης (recall).

Ακρίβεια: Το κατά πόσο πέτυχε η αναζήτηση του χρήστη εξαρτάται από την ακρίβεια των αποτελεσμάτων, από το ποσοστό δηλαδή των σελίδων που ενδιέφεραν όντως τον χρήστη προς το σύνολο των σελίδων που επέστρεψε η μηχανή αναζήτησης.

Ανάκληση: Είναι το ποσοστό των σελίδων που επέστρεψε η μηχανή αναζήτησης προς το ποσοστό όλων των σχετικών σελίδων στον Ιστό. Η ανάκληση της μηχανής αναζήτησης συνδέεται άμεσα με την πληρότητά της (coverage), το ποσοστό δηλαδή των εγγεγραμμένων html σελίδων της βάσης δεδομένων της μηχανής αναζήτησης προς το ποσοστό όλων των html σελίδων που βρίσκονται στον Ιστό.

Σύμφωνα με στατιστική έρευνα του NEC Research Institute το 1997, το ερώτημα που δίνει ένας χρήστης σε μια μηχανή αναζήτησης, αποτελείται συνήθως από λίγες λέξεις-κλειδιά. Για παράδειγμα, στη μηχανή αναζήτησης NECI και με δείγμα 9,000 ερωτημάτων, βλέπουμε ότι τα 4,500 περίπου ερωτήματα περιείχαν μόλις μια λέξη-κλειδί.

Όπως καταλαβαίνουμε, τα αποτελέσματα που θα επιστρέψει η βάση δεδομένων στην περίπτωση που η αναζήτηση είναι απλοϊκή (πχ: μια λέξη-

κλειδί : University) θα είναι χιλιάδες. Θα πρέπει λοιπόν να ταξινομηθούν με τέτοιο τρόπο, έτσι ώστε να προτείνονται στο χρήστη τα URLs που περιέχουν πληροφορίες πλησιέστερες σ'αυτές που θα ήθελε. Η ταξινόμηση αυτή γίνεται με διάφορους αλγορίθμους και μπορεί να κρίνει την επιτυχία⁶ ενός κόμβου (site) στον Ιστό. Θα τη μελετήσουμε αναλυτικότερα στην επόμενη ενότητα.

Οι μηχανές αναζήτησης παρουσιάζουν συνήθως τα αποτελέσματά τους (URLs σχετικών html σελίδων) κατά δεκάδες. Αυτό γίνεται λόγω του μεγάλου όγκου των αποτελεσμάτων.

Επίσης, ταξινομούν τα αποτελέσματά τους με βάση τα κριτήρια αναζήτησης. Προτείνουν δηλαδή στο χρήστη τις html σελίδες που πιστεύουν ότι ανταποκρίνονται περισσότερο στην αναζήτηση του.

Σύμφωνα με στατιστικές μελέτες:

«αν ένα Web site δεν περιέχεται στις πρώτες τρεις σελίδες αποτελεσμάτων, στα πρώτα 30 δηλαδή προτεινόμενα URLs υπάρχει μόνο 28% πιθανότητα να επιλεγεί από έναν χρήστη. Αν δεν περιέχεται στα πρώτα 70 προτεινόμενα URLs, τότε το ποσοστό αυτό πέφτει στο 12%.»

Οι μηχανές αναζήτησης αποτελούν ένα από τα πιο διαδεδομένα μέσα αναζήτησης πληροφοριών στο Internet (μαζί με τους καταλόγους), και συνεπώς, ένα ισχυρό μέσο, το οποίο ενδέχεται να κρίνει την επιτυχία ή αποτυχία ενός site. Γίνεται κατανοητή συνεπώς η ανάγκη για μια καλύτερη θέση ενός site μέσα στα αποτελέσματα μιας μηχανής αναζήτησης.

Η θέση αυτή ονομάζεται ranking

Ιδιαιτερότητες συγκεκριμένων μηχανών αναζήτησης

Μερικές μηχανές αναζήτησης, όπως η WebCrawler, χρησιμοποιούν το λεγόμενο link popularity check για να προωθήσουν κάποιες σελίδες. Ελέγχουν δηλαδή, πόσες φορές παρουσιάζονται σε άλλες html σελίδες, links τα οποία στέλνουν στη σελίδα που πρέπει να βαθμολογήσει. Όσο πιο πολλά links βρει, τόσο πιο αξιοπρόσεκτη και αποδεκτή από την κοινότητα του

⁶ $\frac{\text{hits}_{\text{site}}}{\sum \text{hits}_{\text{all sites}}}$ όπου $\text{hits}_{\text{site}}$ είναι ο αριθμός των hits που λαμβάνει το site, και $\sum \text{hits}_{\text{all sites}}$ ο συνολικός αριθμός hits όλων των sites.

Internet θεωρεί πως είναι η σελίδα που μελετάει. Έτσι, αυξάνει το ranking της.

Η μηχανή αναζήτησης Hotbot ελέγχει και το μέγεθος της html σελίδας που μελετάει. Οι μικρές html σελίδες που έχουν τον ίδιο αριθμό λέξεων-κλειδιών με κάποιες μεγαλύτερες θα έχουν καλύτερο ranking.

Η μηχανή αναζήτησης Infoseek Ultra, χρησιμοποιεί βάρη σε μερικές λέξεις-κλειδιά. Οι ασυνήθιστες λέξεις, αυτές δηλαδή που παρουσιάζονται πιο σπάνια στη βάση δεδομένων της Infoseek, θα έχουν μεγαλύτερο βάρος από τις συνηθισμένες. Η λογική αυτή βοηθάει τη μηχανή αναζήτησης στην επίτευξη καλύτερου αποτελέσματος, όταν σαν κριτήριο αναζήτησης δίνονται περισσότερες από μία λέξεις-κλειδιά.

Πηγές:

<http://www.nswc.navy.mil/cosip/nov97/sugg1197-1.shtml>

WEB ROBOTS

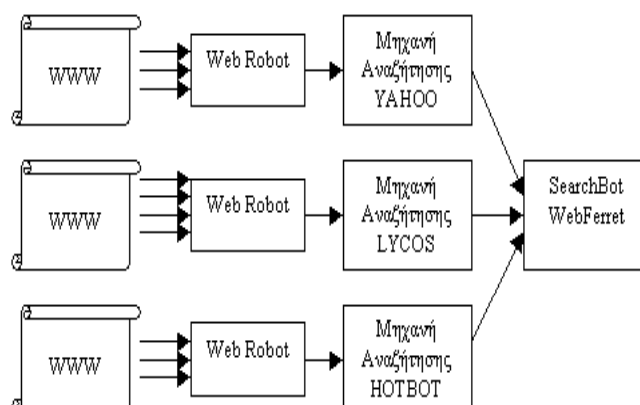
Μια μηχανή αναζήτησης αυτή καθ' αυτή δεν αποτελεί Search Agent. Η λειτουργία της βασίζεται στα λεγόμενα Web Robots (ή αλλιώς Spiders, Crawlers), τα οποία ανήκουν στην κατηγορία των Agents και η μελέτη τους παρουσιάζει μεγάλο ενδιαφέρον.

Αναλυτικότερα:

Spider ή Crawler λέγεται το πρόγραμμα που επισκέπτεται Web sites και διαβάζει τις σελίδες τους και άλλες πληροφορίες για να δημιουργήσει ή να ενημερώσει εγγραφές στον κατάλογο κάποιας μηχανής αναζήτησης. Υπάρχουν πολλοί τρόποι για να «συρθεί» ένας Spider μέσα σε ένα site. Για παράδειγμα, θα μπορούσε να ακολουθήσει τους συνδέσμους του hypertext της κεντρικής σελίδας και να συνεχίσει με τις επόμενες μέχρι να διαβάσει όλες τις διαθέσιμες σελίδες που περιέχει το site ή μέχρι να εκπληρωθεί ένας στόχος. Οι μεγαλύτερες μηχανές αναζήτησης στο WWW έχουν δικούς τους spiders (ο spider του AlitaVista λέγεται Scooter). Τα προγράμματα αυτά μπορούν να δραστηριοποιηθούν σε ολόκληρο το WWW, αλλά συνήθως

κάνουν επιλεκτικές επισκέψεις σε κόμβους, οι οποίοι έχουν δηλωθεί ως «νέοι» ή «προσφάτως ανακατασκευασμένοι» από τους διαχειριστές τους.

Στο παρακάτω σχήμα φαίνεται η αλληλεπίδραση των Search Engines με τον χρήστη, το WWW και τα Web Robots.



Εικόνα 3 : WEB ROBOTS & SEARCH ENGINES

Μια μηχανή αναζήτησης αποτελείται από τρία μέρη: τον server ερωτημάτων, τη βάση δεδομένων και το Web Robot. Ο server ερωτημάτων είναι υπεύθυνος για την αλληλεπίδραση με το χρήστη, δέχεται δηλαδή από αυτόν τις λέξεις-κλειδιά με τη μορφή κάποιου ερωτήματος (query) και στέλνει πίσω τα αποτελέσματα. Στη βάση δεδομένων βρίσκονται αποθηκευμένες όλες οι πληροφορίες που ζητάει ο χρήστης, για το που βρίσκονται οι html σελίδες που τον ενδιαφέρουν. Τέλος το Web robot είναι υπεύθυνο για την ενημέρωση της βάσης δεδομένων και την αλληλεπίδραση με το WWW.

Τα γνωστότερα Web Robots που συναντάμε στο Internet ανήκουν στις γνωστότερες μηχανές αναζήτησης. Παρακάτω βλέπουμε έναν πίνακα μ'αυτά.

<u>Μηχανή Αναζήτησης</u>	<u>Web Robot</u>
Alta Vista	Scooter
Excite	Crawlx.atext
HotBot	Slurp the Web
Hound	

InfoSeek	Sidewinder
Lycos	T-Rex
Northern Light	Gulliver
Web Crawler	Spidey

Πίνακας 1: ΓΝΩΣΤΑ WEB ROBOTS

Μια ολοκληρωμένη λίστα με Web robots διατηρείται στην παρακάτω σελίδα:

<http://info.webcrawler.com/mak/projects/robots/active/html/type.html>

Πηγή: Crawling the Web

<http://www.zdnet.com/pcmag/issues/1513/pcmg0045.htm>

Αρνητικά Σημεία των WebRobots

Τα Web Robots ζουν και εργάζονται ασταμάτητα στο WWW από το 1994 μέχρι σήμερα. Σε αυτό το χρονικό διάστημα, προσφέρουν υπηρεσίες χρήσιμες για την εύρεση των πληροφοριών του WWW και την ταξινόμησή τους, καθώς και χρήσιμα στατιστικά κάθε είδους για το Internet. Υπάρχουν όμως πολλοί, οι οποίοι τονίζουν τα αρνητικά στοιχεία που προσφέρει ένα Web Robot στην κοινότητα του Internet. Το κυριότερο από αυτά είναι οι υψηλές απαιτήσεις που έχει σε ταχύτητα και σε εύρος δικτύου (bandwidth).

Όπως ξέρουμε, η ικανότητα ενός Server να εξυπηρετεί εξαρτάται από τρεις παράγοντες:

- Το εύρος του δικτύου (bandwidth).
- Την ταχύτητα του Web Server.
- Την ποσότητα των δεδομένων που μεταφέρονται σε μια χρονική στιγμή, η οποία εξαρτάται άμεσα από τον αριθμό των απαιτήσεων (Requests) των χρηστών.

Βλέπουμε λοιπόν ότι η χρήση των Web robots (Spiders) και η συνεχόμενη σφυρηλάτηση ενός Web Server με Requests θα μπορούσε να μειώσει αισθητά την απόδοσή του και να τον οδηγήσει ακόμα και σε κατάρρευση. Έτσι η λογική των Spiders αποτελεί αμφιλεγόμενο θέμα ανάμεσα στους Web Administrators, κυρίως επειδή ανεύθυνοι

προγραμματιστές έχουν δημιουργήσει Spiders που τρέχουν πολύ γρήγορα, κάνοντας πολλαπλά Requests σε πολύ μικρό χρονικό διάστημα (αυτό ονομάζεται rapid fire).

Αν αναλογιστούμε και την αυξανόμενη χρήση των φορμών και των CGI scripts στις σελίδες του Internet (τα οποία ένας Spider δεν μπορεί να επεξεργαστεί), βλέπουμε την ανάγκη δημιουργίας ενός προτύπου, το οποίο θα βοηθάει τον Web Administrator να ελέγχει την είσοδο των διαφόρων Spiders στον server του. Το πρότυπο αυτό βρίσκεται ακόμη υπό κατασκευή.

Περισσότερες πληροφορίες μπορούν να βρεθούν στο άρθρο :

Robots in the Web: threat or treat?

<http://info.webcrawler.com/mak/projects/robots/threat-or-treat.html>

DELIVERY AGENTS

Ένα πρόβλημα που αντιμετωπίζει ο μέσος χρήστης του Internet είναι η ταχύτητα μετάδοσης των δεδομένων. Αν υποθέσουμε ότι το μεγαλύτερο μέρος των χρηστών συνδέονται με το Internet μέσω μιας Dial-up σύνδεσης με ταχύτητες που φτάνουν δύσκολα τα 56 Kbps (Kbytes per second), καταλαβαίνουμε ότι η μετάδοση της πληροφορίας (ενός αρχείου ουσιαστικά) μπορεί να γίνει μια πολύ χρονοβόρα και επίπονη διαδικασία, ιδιαίτερα όταν μιλάμε για μεγάλα αρχεία. Μια λύση για το πρόβλημα αυτό προσπαθούν να δώσουν οι λεγόμενοι Delivery Agents.

Background Ftp Της Forthnet

Η Forthnet αποτελεί έναν από τους μεγαλύτερους ISPs (Internet Service Provider = Φορέας Παροχής Υπηρεσιών του Internet) στην Ελλάδα. Μια από τις υπηρεσίες που προσφέρει είναι και η υπηρεσία Background FTP. Ο χρήστης, αφού επιβεβαιώσει ότι είναι συνδρομητής στη Forthnet, ορίζει την ηλεκτρονική διεύθυνση στην οποία βρίσκεται ένα αρχείο που τον ενδιαφέρει. Το background FTP αναλαμβάνει τότε να κατεβάσει (download) αυτό το αρχείο στον server της Forthnet, από τον οποίο μπορεί και ο χρήστης να το κατεβάσει στον Η/Υ του. Το κέρδος για τον χρήστη: η ταχύτητα

μετάδοσης του αρχείου από τον server της Forthnet στον υπολογιστή του είναι ενδεχομένως πολύ μεγαλύτερη από αυτή που θα είχε επιτευχθεί με μια απ'ευθείας σύνδεση με τον server στον οποίο ανήκε αρχικά το αρχείο (ο χρήστης από τη στιγμή που είναι συνδρομητής στη Forthnet έχει καλύτερη ταχύτητα στη σύνδεσή του με τον κόμβο της Forthnet). Το Background FTP της Forthnet δεν αποτελεί Agent. Σκεφτείτε όμως ένα πρόγραμμα υπεύθυνο για την αντιγραφή πολλών αρχείων από έναν server του Internet σε έναν άλλων ή ένα πρόγραμμα υπεύθυνο για την αντιγραφή των περιεχομένων ενός ολόκληρου server σε έναν άλλον. Ένα τέτοιο πρόγραμμα ονομάζεται Mirroring Program και θα το αναλύσουμε αμέσως παρακάτω.

Mirroring Programs

Πολλές φορές συναντάμε στο Internet τον όρο Site Mirror (mirror = καθρέφτης). Αυτό δηλώνει ότι τα περιεχόμενα ενός site «καθρεφτίζονται» σε ένα άλλο. Για παράδειγμα, τα περιεχόμενα του site του Μακεδονικού Πρακτορείου Ειδήσεων (www.mpa.gr) μπορεί να τα βρει κανείς και σε έναν κόμβο στην Αμερική, έτσι ώστε να εξυπηρετούν του ομογενείς που θέλουν να διαβάσουν την Ελληνική ειδησεογραφία. Οι ομογενείς στην Αμερική μπορούν να έχουν πρόσβαση και στο κεντρικό site του Μακεδονικού Πρακτορείου Ειδήσεων (το οποίο βρίσκεται στην Ελλάδα), αλλά η ταχύτητα πρόσβασής τους σ'αυτό θα είναι μικρότερη απ'αυτή που θα έχουν αν επισκεφθούν το Site Mirror.

Αν λάβουμε υπ'όψη το γεγονός ότι η Ελληνική ειδησεογραφία και προφανώς και το site του Μακεδονικού Πρακτορείου Ειδήσεων ανανεώνεται καθημερινά, θα διαπιστώσουμε πόσο απαραίτητη είναι η χρήση ενός Mirroring Program. Το πρόγραμμα αυτό θα έχει ως στόχο την καθημερινή αντιγραφή του site του Μακεδονικού Πρακτορείου Ειδήσεων στο Site Mirror της Αμερικής. Θα πρέπει να τρέχει μια φορά την ημέρα και να έχει ρουτίνες παρόμοιες με αυτές των Web Robots ή Spiders (που αναλύσαμε παραπάνω) για να καταλαβαίνει τη δομή του site και να αντιγράφει όλες τις πληροφορίες που περιέχονται σ'αυτό.

Μια άλλη σύγχρονη εφαρμογή των Agents, είναι οι λεγόμενοι Notification Agents. Πρόκειται συνήθως για προγράμματα που ειδοποιούν το χρήστη (ή άλλους Agents), όταν συμβαίνουν περιστατικά τα οποία τον ενδιαφέρουν. Στηρίζουν συνεπώς τη λειτουργία τους στην ανάγκη ενός Agent για αυξημένη προσαρμοστικότητα στις αλλαγές του περιβάλλοντός του. Θα μπορούσαμε να πούμε πως οι Notification Agents αποτελούν τους δράστες που ευθύνονται για την ισορροπία του ατόμου, στην περίπτωση που αυτό σκοντάψει. Οι δράστες αυτοί λειτουργούν συνήθως σε συνεργασία με άλλους δράστες, δηλαδή οι Notification Agents αλληλεπιδρούν συνήθως με άλλους Agents ή απευθείας με το χρήστη για να βοηθήσουν στην επίτευξη ενός συλλογικού στόχου. Ο στόχος αυτός μπορεί να αναφέρεται σε οποιαδήποτε υπολογιστική λειτουργία: από τη δημιουργία ενός απλού backup, όταν το περιστατικό που συμβαίνει είναι ένας απλός χτύπος του ρολογιού του H/Y, στην on-line αγοροπωλησία μετοχών, όταν η τιμή τους φτάσει σε ένα συγκεκριμένο κατώτατο ή ανώτατο επίπεδο.

WWW Monitoring Programs

Στη συγκεκριμένη βέβαια περίπτωση τα πράγματα είναι πολύ πιο απλά. Και αυτό γιατί στην ενότητα αυτή θα ασχοληθούμε μόνο με τους λεγόμενους Internet Notification Agents, μικρά προγράμματα δηλαδή, υπεύθυνα για την προσαρμογή χρηστών ή άλλων Agents σε περιστατικά τα οποία αφορούν την κοινότητα του Internet. Όταν λέμε περιστατικά, εννοούμε απλά την αλλαγή των πληροφοριών μιας html σελίδας (η οποία ενδιαφέρει το χρήστη) στο WWW.

Μερικές σελίδες στο WWW δίνουν τη δυνατότητα στο χρήστη να επιλέξει εάν θέλει να ειδοποιηθεί με e-mail σε περίπτωση που τα περιεχόμενα της σελίδας αυτής αλλάξουν. Ο χρήστης δίνει τον e-mail λογαριασμό του και μπαίνει σε μια λίστα με άλλους ενδιαφερόμενους χρήστες. Αν αλλάξουν τα περιεχόμενα της σελίδας αυτής, ο υπεύθυνος (webmaster) θα τρέξει έναν Notification Agent για να στείλει με email ένα τυποποιημένο μήνυμα μαζικά

στα μέλη της λίστας. Το μήνυμα θα έχει την εξής περίπου μορφή: «Σας ενημερώνουμε ότι τα περιεχόμενα της σελίδας www.site.com/page.html έχουν αλλάξει».

Τι συμβαίνει όμως όταν ο webmaster της συγκεκριμένης σελίδας δεν έχει προβλέψει για την ενημέρωση των χρηστών που την παρακολουθούν;

Υπάρχουν προγράμματα τα οποία κάνουν ακριβώς αυτή τη δουλειά για οποιαδήποτε html σελίδα. Επισκέπτονται δηλαδή συγκεκριμένες html σελίδες στο WWW και ελέγχουν αν έχουν αλλάξει τα περιεχόμενά τους. Τα προγράμματα αυτά θα μπορούσαν να τρέχουν τοπικά, από το Desktop ενός απλού χρήστη. Επειδή όμως κάτι τέτοιο θα σπαταλούσε τον πολύτιμο on-line χρόνο του σε συνεχείς ελέγχους html σελίδων, έχουν δημιουργηθεί sites που αναλαμβάνουν τη δουλειά αυτή εκ μέρους του. Η χρησιμότητα των υπηρεσιών αυτών διαφέρει από χρήστη σε χρήστη. Υπάρχουν Monitoring Agents που βοηθούν στην καταγραφή Ένα από αυτά είναι και το NETMind, το οποίο θα αναλυθεί παρακάτω.

Link Validation Programs

Παρόμοια με τα WWW Monitoring Programs θεωρούνται και τα λεγόμενα Link Validation Programs. Πρόκειται για εφαρμογές, οι οποίες ελέγχουν μια ή και περισσότερες html σελίδες για «σπασμένα» links. Έτσι μπορούν να διατρέξουν ένα ολόκληρο site και να μας ενημερώσουν για το ποια links δεν λειτουργούν πλέον σωστά, ποιες διευθύνσεις δηλαδή που υπάρχουν στο site που μελετάμε δεν στέλνουν σε έγκυρες html σελίδες (είτε γιατί αυτές δεν υπάρχουν πλέον, είτε γιατί το link δεν ήταν ποτέ έγκυρο). Τα Link Validation Programs χρησιμοποιούνται κυρίως από Web Administrators, για τον έλεγχο των html σελίδων τους και από μηχανές αναζήτησης, για να ενημερώνουν κατάλληλα τη βάση δεδομένων τους (σε περίπτωση που ένα site πάψει να υπάρχει).

OFF - LINE BROWSERS

Τα προγράμματα αυτά μοιάζουν πολύ με τα Mirroring Programs με τη διαφορά ότι έχουν ως στόχο την αντιγραφή ολόκληρων κόμβων (ή ενός

μέρους τους) τοπικά, στο σκληρό δίσκο του χρήστη και όχι σε άλλον κόμβο του Internet. Έτσι μπορεί ο χρήστης να επισκεφθεί έναν κόμβο που τον ενδιαφέρει off-line (χωρίς να είναι συνδεδεμένος) αργότερα.

Οι ρουτίνες που χρησιμοποιεί ένας Off-Line Browser είναι παρόμοιες με αυτές ενός Mirroring Program ή ενός Spider. Βασίζονται δηλαδή στο φιλτράρισμα των html σελίδων. Ο χρήστης δίνει συνήθως την κεντρική σελίδα ενός Website. Το πρόγραμμα την διαβάζει, την αντιγράφει στο σκληρό δίσκο και βρίσκει τα links τα οποία υπάρχουν στη σελίδα αυτή. Μετά διαβάζει ένα ένα τα links, τα αντιγράφει και βρίσκει τα δικά τους links. Η λειτουργία αυτή επαναλαμβάνεται μέχρι να εξαντληθούν όλες οι σελίδες του Website ή να εκπληρωθεί ένας άλλος στόχος του προγράμματος.

WEBSNAKE OFF-LINE BROWSER

Αποτελεί ένα από τα καλύτερα εμπορικά προγράμματα για Off-Line Web Browsing και χρησιμοποιεί και έναν μάγο (Wizard) για να διευκολύνει τους απλούς χρήστες.

Η ηλεκτρονική διεύθυνση: <http://www.anawave.com/websnake/>

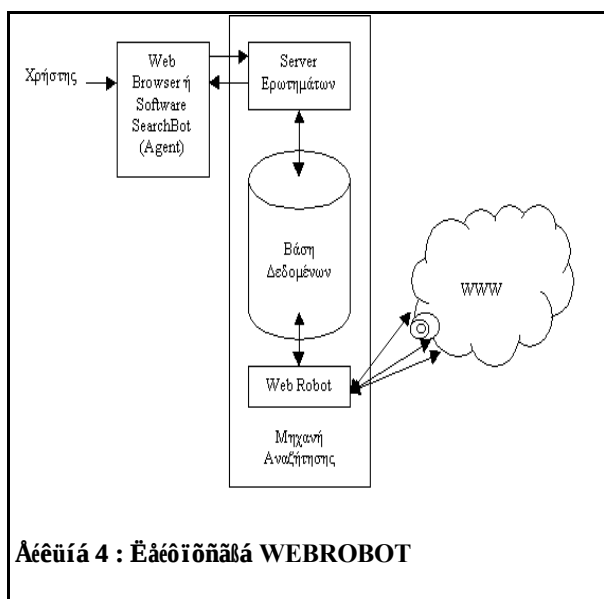
Όπως φαίνεται και από τον Μάγο (Wizard) που εμφανίζεται στην αρχή της εκτέλεσης της εφαρμογής, το WebSnake μπορεί να κάνει τις εξής λειτουργίες:

- ◆ Να αντιγράψει τα περιεχόμενα ενός ολόκληρου Website στο σκληρό δίσκο του χρήστη.
- ◆ Να αντιγράψει τα περιεχόμενα ενός ολόκληρου Website σε έναν άλλο κόμβο στο Internet (Mirroring).
- ◆ Να ψάξει όλες τις διευθύνσεις email που αναφέρονται σε ένα ολόκληρο site.
- ◆ Να δημιουργήσει σχηματικά τη δομή ενός Website (ποια html σελίδα συνδέεται με ποια, κλπ).
- ◆ Να ψάξει ένα Website για συγκεκριμένα αρχεία και να τα αντιγράψει τοπικά.

SEARCHBOTS

Σε προηγούμενο κεφάλαιο αναλύσαμε τη λειτουργία των μηχανών αναζήτησης. Είδαμε την αλληλεπίδραση των Web Robots με το WWW καθώς και την αλληλεπίδραση των μηχανών αναζήτησης με τον Web Browser του χρήστη. Φανταστείτε τώρα, την ύπαρξη κάποιων προγραμμάτων, τα οποία αναλαμβάνουν την αυτοματοποίηση της επικοινωνίας αυτής του Web Browser με τη μηχανή αναζήτησης. Αυτά τα προγράμματα ονομάζονται **Software Searchbots**.

Στο σχήμα της παρακάτω εικόνας, φαίνεται η σχέση ενός Software SearchBot (Software Search Robot) με τις μηχανές αναζήτησης, τα Web Robots και το WWW.



Η διαφορά ενός Web Browser με ένα SearchBot είναι ότι το SearchBot μπορεί να ρωτάει πολλές μηχανές αναζήτησης ταυτόχρονα.

Αναλυτικότερα:

Τα SearchBots είναι μικρά προγράμματα, που εγκαθίστανται και τρέχουν τοπικά, στο σκληρό δίσκο του χρήστη, έχουν δικό τους user interface και μπορούν να

λειτουργήσουν αυτόνομα αλλά και σε συνεργασία με άλλες εφαρμογές (πχ. Μπορούν να συνεργαστούν με τον default Web Browser). Κάνουν παράλληλες αναζητήσεις (requests) σε διάφορες μηχανές αναζήτησης και εμφανίζουν τα αποτελέσματα συγκεντρωμένα και ταξινομημένα. Τα πλεονεκτήματα που παρουσιάζουν σε σχέση με τον συμβατικό τρόπο αναζήτησης (πλοήγηση στο Website των μηχανών αναζήτησης) είναι η ταχύτητα και η απλότητα. Ένας νέος χρήστης μπορεί πιο εύκολα να διατυπώσει το ερώτημά του σε ένα SearchBot από το να ανοίξει τον Web

Browser του, να συνδεθεί με τη σελίδα της κάθε μηχανής αναζήτησης και να προσπαθήσει να αναζητήσει από εκεί τα αποτελέσματα της.

ΚΕΦΑΛΑΙΟ 4, ΥΠΑΡΧΟΝΤΑ ΠΡΟΪΟΝΤΑ

NETMIND WEB SITE

Το site αυτό αναλαμβάνει να ελέγχει αυτόματα μεμονωμένες και προεπιλεγμένες από τους χρήστες του html σελίδες. Αν τα περιεχόμενα μιας σελίδας αλλάξουν, τότε ο χρήστης που δήλωσε πως ενδιαφέρεται για τη σελίδα αυτή θα ειδοποιηθεί μέσω ηλεκτρονικού ταχυδρομείου. Το Netmind μάλιστα ειδοποιεί τους χρήστες ακόμα και σε περίπτωση «μετακίνησης» της σελίδας σε άλλο κόμβο. Χρησιμοποιεί Spiders και άλλα παρόμοια εργαλεία για να ελέγχει όλες τις δηλωμένες σελίδες, τις οποίες φιλτράρει για να βρει τη νέα τους διεύθυνση σε περίπτωση που αυτές μετακινήθηκαν. Η ηλεκτρονική διεύθυνση του NetMind είναι η: <http://mindit.netmind.com/>.

Προγράμματα όπως το NetMind είναι εξαιρετικά χρήσιμα σε εταιρίες, οι οποίες θέλουν να παρακολουθούν τις ανταγωνιστικές τους εταιρίες στο Internet, όπως επίσης και σε οποιονδήποτε Web Administrator, ο οποίος θέλει να ελέγχει αν οι html σελίδες του δέχτηκαν καμία επίθεση από hackers και άλλους.

ALTAVISTA

Η μηχανή Altavista αποτελεί μια από τις εγκυρότερες μηχανές αναζήτησης στο Internet. Ξεκίνησε ως ένα reasearch project στα Digital Equipment Corporations Research Laboratories στο Palo Alto της California (USA) και η βάση δεδομένων της περιλαμβάνει σήμερα πάνω από 150 εκτομμύρια σελίδες. Το Web robot της Altavista ονομάζεται Scooter. Για να τεθεί σε λειτουργία από έναν webmaster, θα πρέπει να δηλωθούν κάποια γενικά στοιχεία στην AddUrl σελίδα του site της Altavista. Ο Scooter τότε θα αρχίσει την «περιπλάνησή» του στις σελίδες του νεοεισερχόμενου site. Μπορεί να χρειαστεί μέχρι και τρεις μήνες για να περιπλανηθεί σε ολόκληρο το site και να αντιγράψει τα περιεχόμενά του στη βάση δεδομένων της Altavista (συνήθως ελέγχει δύο σελίδες από ένα site την εβδομάδα, λόγω του

μεγάλου όγκου των νέων sites). Αν ο Scooter αντιληφθεί προσπάθεια για spamming, τότε το URL του site που μελετάει θα δοθεί με e-mail σε έναν χειριστή για να το ελέγξει μόνος του.

Η Altavista δίνει μεγαλύτερη προτεραιότητα σε λέξεις-κλειδιά που εμφανίζονται στα Meta Tags μιας σελίδας και σ'αυτές που εμφανίζονται κοντά στο επάνω μέρος της. Αν δεν υπάρχει Meta Tag για την περιγραφή, τότε η μηχανή αναζήτησης θεωρεί ως περιγραφή τις πρώτες 30-40 λέξεις της σελίδας.

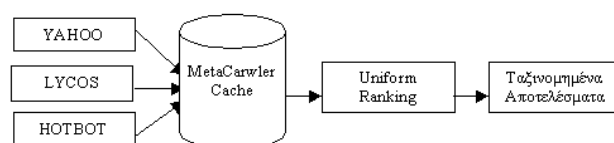
Πληροφορίες για τον Scooter υπάρχουν στη διεύθυνση:

http://www.altavista.com/av/content/about_our_story_2.htm#scooter

METACRAWLER

Η MetaCrawler είναι μια μηχανή αναζήτησης, η οποία δημιουργήθηκε το 1994 στο Πανεπιστήμιο της Washington από τους Erik Selberg, Oren Etzioni και Greg Lauckhart. Στεγάζεται στο Seattle, στον server της go2net Inc.

Η διαφορά της μηχανής αυτής αλλά και όλων των Metasearch Engines από τις άλλες μηχανές αναζήτησης είναι ότι δεν έχουν δικές τους βάσεις δεδομένων αλλά χρησιμοποιούν τις άλλες search engines που υπάρχουν στο WWW. Όταν ο χρήστης κάνει ένα ερώτημα στην MetaCrawler, αυτή θα στείλει το ερώτημα αυτό στο Lycos, στο Infoseek, στο Excite, στο Altavista, στο Thunderstone και στο Yahoo. Έπειτα θα συλλέξει τα αποτελέσματα, θα τα αποθηκεύσει με ένα κοινό format, θα τα ταξινομήσει και θα τα παρουσιάσει στο χρήστη. Όπως καταλαβαίνει κανείς, η μηχανή MetaCrawler είναι λίγο πιο αργή από τις άλλες μηχανές αναζήτησης αλλά



τα αποτελέσματά της έχουν συνήθως μεγαλύτερη ακρίβεια και πληρότητα.

Τα κυριότερα χαρακτηριστικά της MetaCrawler μπορούμε να τα συνοψίσουμε παρακάτω :

ΠΛΗΡΗΣ ΚΑΤΑΓΡΑΦΗ ΤΟΥ WEB:

Εξαιτίας του μεγέθους και της συνεχούς ενημέρωσης του Web, μπορούμε να ισχυριστούμε ότι καμία συμβατική μηχανή αναζήτησης δεν μπορεί να προσφέρει στο χρήστη πλήρη καταγραφή των πληροφοριών του. Η MetaCrawler παρέχει σαφώς περισσότερες πληροφορίες επειδή συνεργάζεται με πολλές μηχανές αναζήτησης.

ΕΥΚΟΛΗ ΣΥΝΤΑΞΗ ΕΡΩΤΗΜΑΤΩΝ:

Κάθε μηχανή αναζήτησης ζητάει από το χρήστη διαφορετικό τρόπο σύνταξης των ερωτημάτων του (των λέξεων-κλειδιών δηλαδή). Ιδιαίτερα όταν έχουμε να κάνουμε με εξεζητημένες αναζητήσεις ο χρήστης για να πετύχει το σωστό αποτέλεσμα θα πρέπει να μάθει τις λεπτομέρειες χρήσης της κάθε μηχανής αναζήτησης ξεχωριστά. Η MetaCrawler προσφέρει ένα απλό τρόπο αναζήτησης, ο οποίος όμως «μεταφράζεται» κατάλληλα για να εξυπηρετεί τις απαιτήσεις των άλλων μηχανών αναζήτησης, με τις οποίες συνεργάζεται.

ΣΥΛΛΟΓΙΚΟ RATING:

Η MetaCrawler συνδυάζει και κανονικοποιεί τα scores (ratings) που δίνουν οι διάφορες μηχανές αναζήτησης για τα αποτελέσματά τους. Έτσι ο χρήστης έχει μια ολοκληρωμένη άποψη της θέσης κάθε URL συνολικά στο Web, και όχι μόνο στη βάση δεδομένων μια μηχανής αναζήτησης.

SAVVY SEARCH

Η μηχανή Savvy Search αποτελεί ένα Meta Search εργαλείο, το οποίο κάνει παράλληλες αναζητήσεις σε διάφορες μηχανές αναζήτησης. Εμφανίζει τα αποτελέσματά ταξινομημένα ανά μηχανή αναζήτησης και σύμφωνα με τις προτιμήσεις του χρήστη (χωρίς περιγραφές ή με περιγραφές

των URL). Τα URLs που επιστρέφει στο χρήστη δεν ταξινομούνται συλλογικά, σύμφωνα με εσωτερικά κριτήρια της Savvy Search, αλλά εμφανίζονται με τη σειρά που τα παίρνει η Savvy Search από τις άλλες μηχανές αναζήτησης.

Το μοναδικό προνόμιο της μηχανής αυτής σε σχέση με την Metacrawler είναι το γεγονός ότι το user-interface της είναι διαθέσιμο σε πολλές γλώσσες (Αγγλικά, Γαλλικά, Γερμανικά, Ελληνικά, ...).

Μπορείτε να την επισκεφθείτε στη διεύθυνση :

<http://www.savvysearch.com/>

BLACK WIDOW

Κατηγορία : OFF LINE BROWSERS

Η ηλεκτρονική διεύθυνση του προγράμματος είναι:

<http://www.softbytelabs.com/BlackWidow/>

Πρόκειται για ένα πρόγραμμα, υπεύθυνο για την μαζική αντιγραφή περιεχομένων του WWW στο σκληρό δίσκο του υπολογιστή όπου βρίσκεται. Ο χρήστης μπορεί να επιλέξει το είδος των αρχείων που χρειάζεται, καθώς και το βάθος της αναζήτησης. Η έκδοση που προσφέρεται δωρεάν στο δίκτυο είναι χρονικά περιορισμένη.

TELEPORT PRO

Κατηγορία : OFF LINE BROWSERS

Διεύθυνση στο Internet : <http://www.tenmax.com/pro.html>

Πρόκειται για ένα από τα καλύτερα προγράμματα του είδους. Μπορεί να χρησιμοποιηθεί και ως Mirroring Program, αν δοθούν οι κατάλληλες παράμετροι (user name και password του Web Administrator του κόμβου όπου θέλουμε να αντιγράψουμε τα περιεχόμενα άλλου κόμβου). Όπως φαίνεται στην εικόνα, μπορεί κάποιος να αντιγράψει τοπικά έναν ολόκληρο κόμβο, ή συγκεκριμένα αρχεία του και να κάνει αναζητήσεις για διευθύνσεις ηλεκτρονικού ταχυδρομείου ή ακόμα και για απλές λέξεις-κλειδιά. Το Teleport Pro κάνει παράλληλες αναζητήσεις και πολλαπλά requests

αυξάνοντας έτσι την ταχύτητα και την απόδοσή του. Η έκδοση που προσφέρεται είναι beta και ισχύει για έναν συγκεκριμένο αριθμό αναζητήσεων.

WEB MIRROR

Κατηγορία : OFF LINE BROWSERS

Διεύθυνση στο Internet: <http://www.maccasoft.com/>

Το Web Mirror διαφέρει από τα προηγούμενα προγράμματα σε μια μόνο ιδιότητα. Μπορεί κανείς να προγραμματίσει χρονικά τις λειτουργίες του. Για παράδειγμα, μπορεί κανείς να το αναθέσει να αντιγράφει τοπικά, σε καθημερινό επίπεδο τον κόμβο του Μακεδονικού Πρακτορείου Ειδήσεων. Επίσης, αν τρέξει κανείς την έκδοση του προγράμματος για Windows NT, μπορεί να το χρησιμοποιήσει και ως Mirroring Tool.

Παρακάτω, παρουσιάζονται κάποια από τα δημοφιλέστερα εμπορικά **SearchBots**.

WEB FERRET SEARCH AGENT

Σκοπός του Webferret είναι η αναζήτηση WWW σελίδων στο Internet με κριτήριο αναζήτησης μια ή περισσότερες λέξεις κλειδιά.. Δεν πρόκειται για μια σύνθετη και «βαριά» εφαρμογή. Το Webferret απλά χρησιμοποιεί τις ήδη υπάρχουσες μηχανές αναζήτησης του WWW (Excite, Webcrawler, Lycos, Yahoo, κλπ). Στέλνει δηλαδή σ'αυτές τα κριτήρια αναζήτησης (keywords) και ταξινομεί τα αποτελέσματα που παίρνει. Το πλεονέκτημα της εφαρμογής αυτής είναι η ταυτόχρονη (ασύγχρονη) αναζήτηση σε πολλαπλές search engines και η δυνατότητα να αποθηκεύει κανείς τις αναζητήσεις του. Κάνοντας διπλό κλικ πάνω στο όνομα ή στη διεύθυνση της WWW σελίδας που μας ενδιαφέρει, ανοίγει ο default Web Browser μας και συνδέεται με τη σελίδα αυτή.

Η εταιρία FerretSoft ασχολείται κυρίως με τη δημιουργία τέτοιων εφαρμογών και αν επισκεφθεί κανείς το site της (www.ferretsoft.com) μπορεί να βρει προγράμματα όπως τα παρακάτω :

AuctionFerretPRO. Ένας γρήγορος και ιδιαίτερα ισχυρός τρόπος αναζήτησης και εύρεσης δημοπρασιών (**auctions**) στο δίκτυο. Το πρόγραμμα αυτό υποβάλλει ερωτήματα (queries) σε γνωστά Site δημοπρασιών, οπότε και πολύ γρήγορα ενοπίζει αντικείμενα που ο χρήστης αναζητά.



Λέξιόν 6 : Η ιδιότητα AUTIONFERET

FileFerret Pro. Πρόγραμμα που ψάχνει στις γνωστές μηχανές αναζήτησης του WWW (Lycos, κλπ) για αρχεία σε ftp διευθύνσεις . Είναι ένας γρήγορος τρόπος για εύρεση και προσκόμιση αρχείων από το δίκτυο. Το FileFerretPRO ψάχνει και σε Web-based βάσεις δεδομένων αρχείων, σε διαμοιραζόμενες (shareware) βάσεις δεδομένων με Archie πρωτόκολλο, σε public domain software και άλλα αρχεία.

WebFerretPRO. Πρόγραμμα - utility για να βρίσκει ιστοσελίδες στο δίκτυο. Το WebFerretPRO υποβάλλει queries στις πιο γνωστές Web-based search engines, οπότε και οι σελίδες που αναζητά ο χρήστης βρίσκονται γρήγορα και αποτελεσματικά.

IRCFerretPRO. Είναι ένα utility που βρίσκει γρήγορα το κανάλι επικοινωνίας στο IRC (Internet Relay Chat) ενός ατόμου. Το IRCFerretPRO εντοπίζει εύκολα άλλους χρήστες στα IRC δίκτυα, κάνοντας αναζητήσεις με βάση λέξεις κλειδιά (keywords), ταιριάζοντάς τα με τα στοιχεία κάθε χρήστη σε επιλεγμένο IRC κανάλι.

NewsFerretPRO. Πρόγραμμα για τον εντοπισμό πολλαπλών ομάδων ειδήσεων (newsgroups), για τον εντοπισμό Usenet άρθρων. Ταυτόχρονα με τον εντοπισμό, αυτόματα συνδέονται τα μηνύματα και αποκωδικοποιούνται

τα δυαδικά attachments. Τα άρθρα αυτά μπορούν να ανακτηθούν και να προβληθούν ή να κρατηθούν για μετέπειτα επεξεργασία.

InfoFerret Pro. Πρόγραμμα που κάνει αναζητήσεις σε on-line εφημερίδες, περιοδικά, πρακτορεία για την εύρεση άρθρων.

MailFerret Pro. Πρόγραμμα που δέχεται ως κριτήριο αναζήτησης το ονοματεπώνυμο κάποιου και ψάχνει σε μηχανές αναζήτησης διευθύνσεων ηλεκτρονικού ταχυδρομείου για να βρει ποιο είναι το email του.

PhoneFerret Pro. Πρόγραμμα που κάνει αναζητήσεις σε on-line τηλεφωνικούς καταλόγους (Yellow Pages) με στόχο την εύρεση ενός τηλεφώνου.

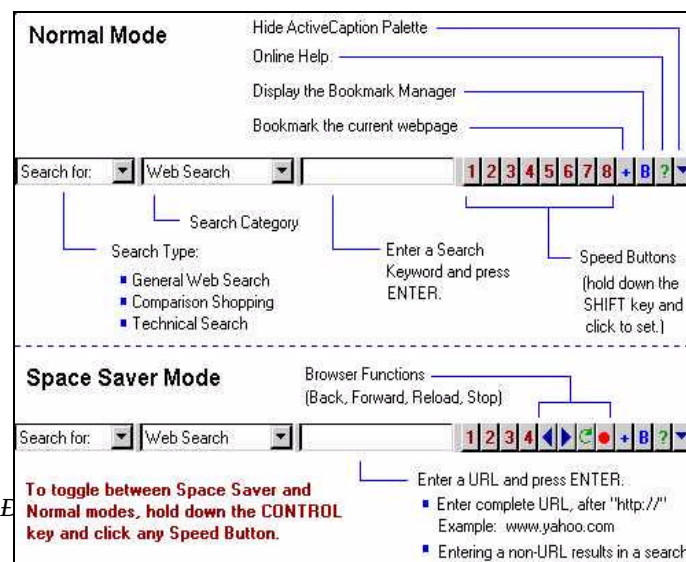
Όπως φαίνεται, όλα τα παραπάνω προγράμματα είναι μικρές εφαρμογές που βοηθούν το χρήστη στην εύρεση της Πληροφορίας μέσα στο WWW.

ACTIVE CAPTION

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://www.onesearch.com/ActiveCaption>

Το πρόγραμμα αυτό εισάγει μια μπάρα εργαλείων (tool bar) στο πάνω μέρος (caption) του default Web Browser του χρήστη. Οι επιλογές της μπάρας αυτής χρησιμοποιούνται στην αναζήτηση σε ορισμένες μηχανές αναζήτησης, και περιλαμβάνουν κάποια κουμπιά καθώς και ένα πλαίσιο κειμένου για να γράφει ο χρήστης τη λέξη-κλειδί που θέλει. Έτσι, μπορεί κάποιος να κάνει οποιαδήποτε αναζήτηση θέλει, χρησιμοποιώντας

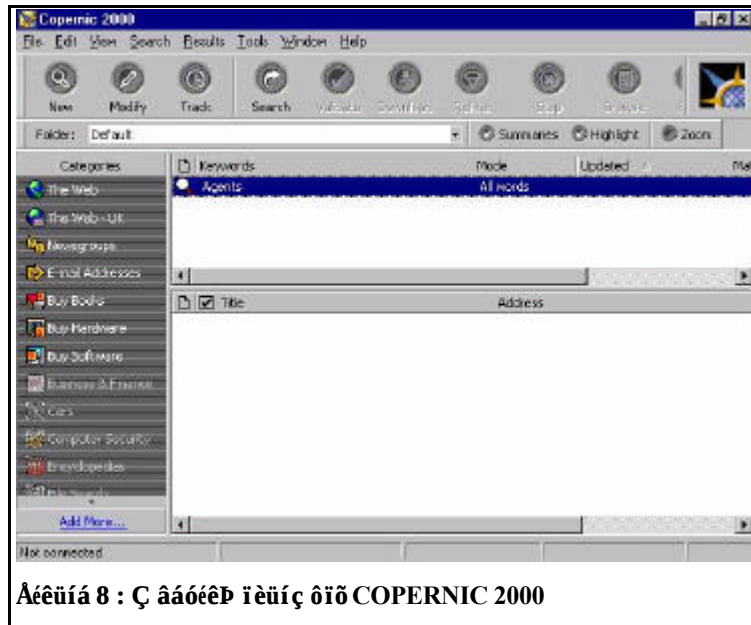


ουσιαστικά την μπάρα του Browser του, χωρίς να συνδέεται με την κεντρική σελίδα της κάθε μηχανής αναζήτησης.

COPERNIC 2000

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://www.copernic.com>



Αποτελεί ένα από τα πιο πλήρη SearchBots, παρόλο που η δωρεάν έκδοση από το δίκτυο είναι το demo του. Περιλαμβάνει επιλογές για αναζητήσεις html σελίδων, usenet νέων, διευθύνσεις ηλεκτρονικού ταχυδρομείου, βιβλίων, άρθρων, ταινιών, τρα-

γουδιών, μικρών αγγελιών κλπ. Μπορεί να ταξινομήσει τις αναζητήσεις του χρήστη με πολλούς τρόπους, καθώς και να τις αποθηκεύσει και να κρατήσει ένα ιστορικό των προτιμήσεων του.

WORLD SEARCH



Διεύθυνση κεντρικής σελίδας στο Internet:

<http://members.aol.com/neatt/index.htm>

Πρόκειται για ένα εύχρηστο SearchBot, το οποίο συνεργάζεται με τον προεπιλεγμένο Web Browser του χρήστη, για να κάνει αυτόνομες αναζητήσεις στις σημαντικότερες μηχανές αναζήτησης του Web.

CEASARS SEARCH BAR

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://www.geocities.com/SiliconValley/Horizon/8159/csb.htm>



Το πρόγραμμα αυτό, κατά την εγκατάστασή του μεταβάλλει τη λειτουργία της επιλογής «Search» του Microsoft Internet Explorer 4.01. Έτσι, τη θέση της μίας και μοναδικής μηχανής αναζήτησης του Internet Explorer, παίρνουν 43 διαφορετικές μηχανές αναζήτησης, τις οποίες ο χρήστης μπορεί να επισκεφθεί με ένα κλικ. Στη διπλανή εικόνα φαίνεται καλύτερα η Μπάρα Αναζήτησης (Search Bar) του προγράμματος.

7 SEARCH

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://www.meat.com>



Το πρόγραμμα αυτό προσφέρει ταχύτατη πρόσβαση και αναζητήσεις σε 90 περίπου μηχανές αναζήτησης. Η λειτουργία του είναι απλή: υπάρχει για κάθε μηχανή αναζήτησης, μια φόρμα, στην οποία ο χρήστης συμπληρώνει τα στοιχεία της αναζήτησης. Με το πάτημα ενός κουμπιού, ανοίγει ο προεπιλεγμένος (default) Web Browser και συνδέεται με τη μηχανή αναζήτησης, εμφανίζοντας τα αποτελέσματά της. Το πρόγραμμα μένει μόνιμα στο Task Bar των Windows και μπορεί να χρησιμοποιηθεί ανεξάρτητα από τον Web Browser.

SHETTY SEARCH

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://members.aol.com/satishetty>

Το Shetty Search είναι ένα εξαιρετικά γρήγορο εργαλείο, το οποίο συνεργάζεται με διάφορες μηχανές αναζήτησης και εμφανίζει τα αποτελέσματά τους σε μια απλή html σελίδα.

FTP WOLF

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://www.msw.com.au/>



Λέξη 12: Η διακίνηση των FTPWOLF

Πρόγραμμα από την εταιρία Wolf HQ, υπεύθυνο για την εύρεση αρχείων σε sites που χρησιμοποιούν το File Transfer Protocol (FTP). Το πρόγραμμα αυτό συνεργάζεται με 26 μηχανές αναζήτησης FTP και δημιουργεί reports σε html σελίδες για να παρουσιάσει τα αποτελέσματα στον χρήστη.

Η κατασκευάστρια εταιρία έχει δημιουργήσει και άλλα παρόμοια SearchBots, με το ίδιο User Interface, τα οποία θα μελετήσουμε παρακάτω.

IMAGE WOLF

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://www.msw.com.au/iwolf/index.html>

Πρόγραμμα από την ίδια εταιρία, το οποίο ψάχνει για εικόνες και αρχεία Video (*.avi).

WEB WOLF

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://www.msw.com.au/>

Πρόγραμμα της ίδιας εταιρίας, το οποίο λειτουργεί ως Desktop Spider. Χρησιμοποιεί δηλαδή παρόμοιες ρουτίνες με αυτές των Spiders (ή Web Robots) για να περιπλανηθεί μέσα σε ένα συγκεκριμένο site.

SEARCH WOLF

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://www.msw.com.au/>

Άλλο ένα πρόγραμμα από την Wolf HQ, υπεύθυνο αυτή τη φορά για παράλληλες αναζητήσεις στις γνωστότερες μηχανές αναζήτησης.

MATA HARI

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://thewebtools.com>

Το πρόγραμμα αυτό συνεργάζεται με 140 μηχανές αναζήτησης, κάνοντας τους διάφορα ερωτήματα. Συνδυάζει ένα πλήθος λειτουργιών για την αναζήτηση, το ranking, την εμφάνιση, το χειρισμό και την αποθήκευση των αποτελεσμάτων τους, δηλαδή των ζητούμενων πληροφοριών. Τις

πληροφορίες αυτές μπορεί να τις αντλήσει από το WWW, τα Newsgroups, Mailing lists, τοπικές βάσεις δεδομένων, bookmarks ή από σελίδες με Web links.

TOP 10 TRACKER

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://www.top-10.com>

Άλλο ένα παρόμοιο εργαλείο, αναζήτησης πληροφοριών στο Internet. Συνεργάζεται μόνο με τις σημαντικότερες μηχανές αναζήτησης.

MP3 LEECH

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://www.entropy.ezpublishing.com>

Το πρόγραμμα αυτό αναλαμβάνει την αναζήτηση αρχείων με επέκταση MP3. Τα αρχεία αυτά είναι συμπιεσμένα αρχεία ήχου και περιέχουν συνήθως γνωστά εμπορικά τραγούδια. Ο χρήστης δεν έχει παρά να δώσει μια μικρή περιγραφή του τραγουδιού και ο MP3 Leech θα το εντοπίσει.

IQ BOX

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://home1.swipnet.se/~w-10829/>



Αέεüιά 13 : Ç ιδύñά ðĩð IQBOX

Ένα ακόμη Searchbot, το οποίο τοποθετεί μια μπάρα επάνω στο Desktop του χρήστη. Όπως φαίνεται και από την εικόνα, ο χρήστης μπορεί να εισάγει μια ή περισσότερες λέξεις κλειδιά και να πατήσει το «Go!». Τότε θα ανοίξει ο default Web Browser του και η επιλεγμένη μηχανή αναζήτησης θα παρουσιάσει τα αποτελέσματα της. Το πρόγραμμα αυτό συνεργάζεται μόνο με τις σημαντικότερες

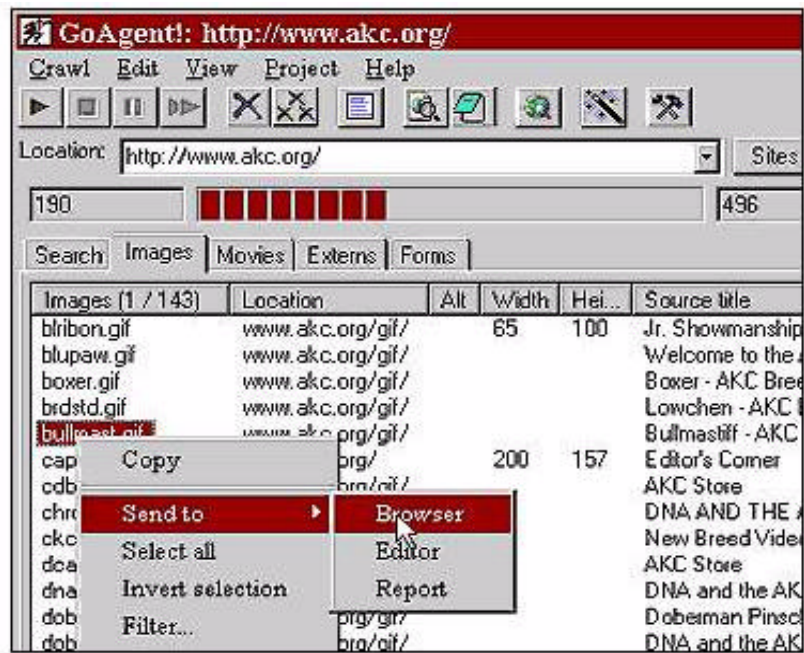
μηχανές αναζήτησης στο Internet.

GO AGENT

Διεύθυνση κεντρικής σελίδας στο Internet:

<http://www.goagent.com>

Το πρόγραμμα αυτό (όπως και όλα τα προηγούμενα) αναλαμβάνει την αυτόματη αναζήτηση σε πολλαπλές μηχανές αναζήτησης, με βάση συγκεκριμένες λέξεις-κλειδιά. Με τη βοήθεια ενός μάγου, ο χρήστης επιλέγει μαζί με άλλες παραμέτρους και το πόσο βαθιά θέλει να είναι η αναζήτησή του.



Αέεüíá 14 : Ôĩ ðãñéäÜëëíí ãñãáóðáò ôĩð GOAGENT

ΠΗΓΕΣ

- <http://filedudes.lvdi.net/win3xx/srchbots/>
- <http://home1.swipnet.se/~w-10829/>
- <http://info.webcrawler.com/mak/projects/robots/threat-or-treat.html>
Robots in the Web: threat or treat?
- <http://members.aol.com/neatt/index.htm>
- <http://members.aol.com/satishetty>
- <http://mindit.netmind.com/> : *NETMIND Notification Agent*
- <http://thewebtools.com>
- http://www.altavista.com/av/content/about_our_story_2.htm#scooter
- <http://www.anawave.com/websnake/> : *WEBSNAKE Off Line Browser*
- <http://www.copernic.com>
- <http://www.entropy.ezpublishing.com>
- <http://www.ferretsoft.com>
- <http://www.geocities.com/SiliconValley/Horizon/8159/csb.htm>
- <http://www.goagent.com>
- <http://www.it.uom.gr/Agents/Agentweb/index.htm>
Πρόγραμμα Πράκτορες στο Δίκτυο
- <http://www.maccasoft.com/>
- <http://www.meat.com>
- <http://www.metacrawler.com/>
- <http://www.msw.com.au/>
- <http://www.msw.com.au/iwolf/index.html>
- <http://www.nswc.navy.mil/cosip/nov97/sugg1197-1.shtml>
- <http://www.oneseek.com/ActiveCaption>
- <http://www.savvysearch.com/>

- <http://www.softbytelabs.com/BlackWidow/> : *BLACK WIDOW Off Line Browser*
- <http://www.tenmax.com/pro.html>
- <http://www.top-10.com>
- <http://www.w3.org> : *CERN*
- <http://www.whatis.com>
- <http://www.zdnet.com/pcmag/issues/1513/pcmg0045.htm> : *Crawling the Web*
- **Webster's New World Dictionary**
- **Internet Agents, Spiders, Wanderers, Brokers, and Bots, Fah-Chun Cheong** , New Riders Publishing , 1995.