

# ΤΕΧΝΟΛΟΓΙΕΣ ΔΙΑΔΙΚΤΥΟΥ ΚΑΙ ΣΧΕΔΙΑΣΗ ΣΤΟΝ ΙΣΤΟ II

Δρ Χριστιάνα Κολιούσκα

Δασολόγος – Περιβαλλοντολόγος ΑΠΘ

# ΜΗΧΑΝΕΣ ΑΝΑΖΗΤΗΣΗΣ



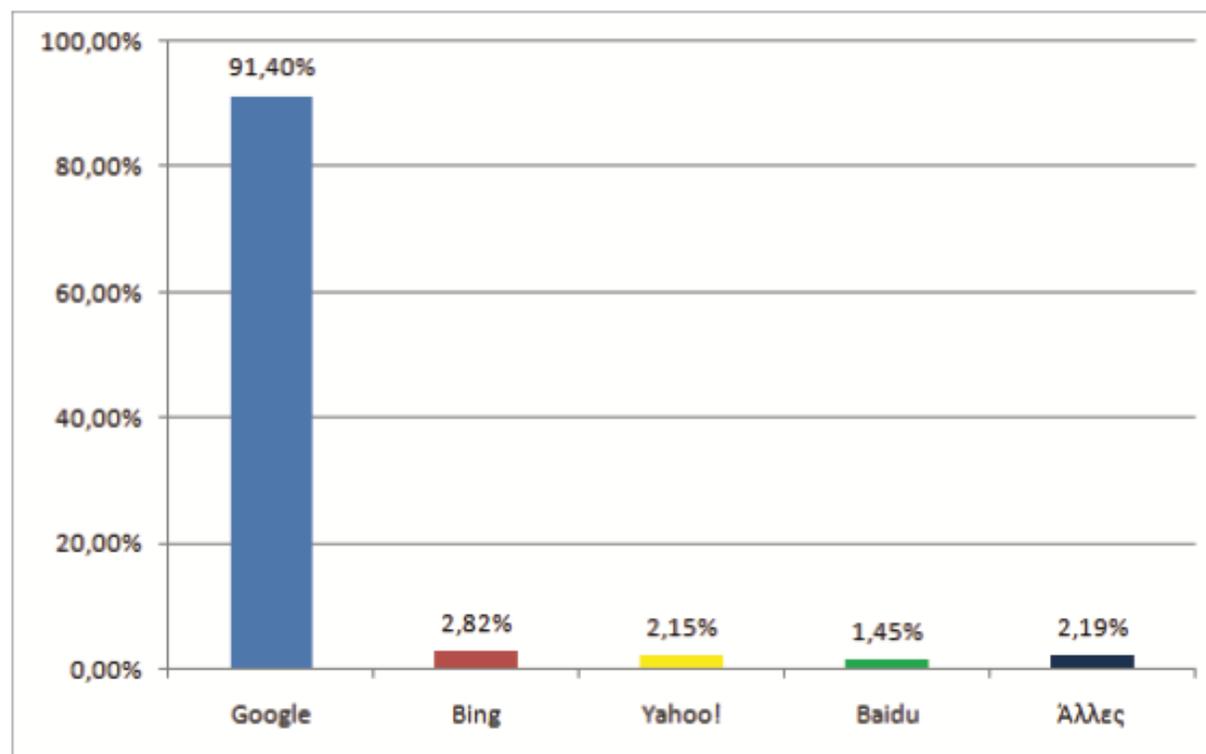
Οι μηχανές αναζήτησης έχουν γίνει σήμερα η αφετηρία σχεδόν κάθε δραστηριότητας στο διαδίκτυο, αποτελώντας τον «οδηγό» προς τις πληροφορίες που αναζητούνται, είτε πρόκειται για κάποια διαδικτυακή τοποθεσία, τις τελευταίες ειδήσεις κ.λπ. Αυτή η μορφή νοημοσύνης βασίζεται σε έναν αλγόριθμο, ένα σύνολο κανόνων οι οποίοι προσδιορίζουν τη σειρά και τον τύπο των αποτελεσμάτων που θα προκύψουν για κάθε εντολή αναζήτησης.

Μια μηχανή αναζήτησης είναι μια εφαρμογή που επιτρέπει την αναζήτηση κειμένων και αρχείων στο διαδίκτυο. Πρόκειται για ένα διαδικτυακό λογισμικό που επιτρέπει στους χρήστες να εντοπίσουν πληροφορίες στο διαδίκτυο. Δημιουργεί ευρετήρια με βάση το περιεχόμενο των σελίδων και συσχετίζει με αυτά ιστοσελίδες (ή εικόνες, video, κ.ά.) με λέξεις-κλειδιά που δόθηκαν από το χρήστη. Στη συνέχεια ταξινομεί τα σχετικά αποτελέσματα με κάποιον μάλλον πολύπλοκο τρόπο και τα προβάλλει στον χρήστη.

Ο τρόπος ταξινόμησης των αποτελεσμάτων αναζήτησης λαμβάνει υπόψη του πολλές παραμέτρους και δεδομένου ότι οι υψηλές θέσεις στην κατάταξη μιας αναζήτησης με σχετικές λέξεις-κλειδιά με τη σελίδα, οδηγεί σε υψηλή επισκεψιμότητα, αναπτύχθηκαν τεχνικές που επιδιώκουν την όσο το δυνατό υψηλότερη κατάταξη. Το σύνολο των τεχνικών ονομάζεται SEO, από τα αρχικά της φράσης Search Engine Optimization και σε ελεύθερη μετάφραση αποδίδεται ως «Βελτιστοποίηση Κατάταξης Ιστοσελίδων από τις Μηχανές Αναζήτησης».



Οι κυριότερες μηχανές αναζήτησης τα τελευταία χρόνια σε παγκόσμιο επίπεδο είναι τέσσερις, με μερίδιο αγοράς που έφτασε το 97.82% των συνολικών παγκόσμιων αναζητήσεων για το έτος 2018. Η Google κατείχε το 91.40% των αναζητήσεων της παγκόσμιας αγοράς, η Bing το 2.82%, η Yahoo! το 2.15% και η Baidu το 1.45%, η οποία δραστηριοποιείται στην Κινέζικη αγορά.



**Εικόνα 1: Μερίδιο Αγοράς Μηχανών Αναζήτησης (2018)**

*Πηγή: statcounter*

## *Google<sup>3</sup>*

Η Google ξεκίνησε τον Ιανουάριο του 1996 ως ερευνητικό πρόγραμμα από τους Larry Page και Sergey Brin, υποψήφιων διδασκτόρων της Επιστήμης των Υπολογιστών στο Stanford University της Καλιφόρνια. Ενώ οι συμβατικές μηχανές αναζήτησης κατέτασσαν τα αποτελέσματά τους μετρώντας πόσες φορές οι όροι αναζήτησης εμφανίζονταν στην σελίδα, οι δύο φοιτητές δημιούργησαν ένα καλύτερο σύστημα το οποίο ανέλυε τις σχέσεις μεταξύ των ιστοσελίδων. Ονόμασαν τη νέα τεχνολογία PageRank, η οποία καθόριζε τη συνάφεια της ιστοσελίδας με βάση τον αριθμό και τη σημαντικότητα των σελίδων που οδηγούν (link) πίσω σε αυτήν.



Η νέα μηχανή αναζήτησης αρχικά ονομάστηκε Back Rub και στη συνέχεια μετονομάστηκε σε Google, λέξη που προήλθε από αναγραμματισμό της λέξης «googol». Στην αρχή το Google έτρεχε στην ιστοσελίδα του Stanford University, με τα domain google.stanford.edu και z.stanford.edu. Το domain name της Google κατοχυρώθηκε στις 15 Σεπτεμβρίου του 1997, ενώ η εταιρεία ιδρύθηκε στις 4 Σεπτεμβρίου του 1998 ακολουθώντας γρήγορη εξέλιξη κυρίως όταν έγινε η επίσημη μηχανή αναζήτησης από τον Yahoo!. Στα μέσα του 1999 επεξεργαζόταν 500.000 αναζητήσεις ημερησίως και το 2004 έφτασε τις 200 εκατ. αναζητήσεις ημερησίως. Σύμφωνα με έρευνες αγοράς από την Millward Brown, τον Απρίλιο του 2007 η Google Inc. ανέλαβε την πρωτοκαθεδρία ως η διασημότερη εταιρεία παγκοσμίως, εκτοπίζοντας έτσι τη Microsoft Corp.

Η Google θεωρείται σήμερα η μεγαλύτερη μηχανή αναζήτησης στο διαδίκτυο. Η μηχανή χρησιμοποιεί σε μεγάλο βαθμό το κριτήριο της ανάλυσης υπερσυνδέσεων όσον αφορά την ιεράρχηση των αποτελεσμάτων Page Rank Technology. Ουσιαστικά μέσω αυτού του συστήματος η μηχανή δέχεται επαίνους για τα υψηλά ποσοστά συνάφειας των αποτελεσμάτων της σε σχέση με τους όρους της αναζήτησης.

Οι αναζητήσεις περιορίζονται σε μια συγκεκριμένη διαδικτυακή περιοχή ή σε μία μόνο γλώσσα, υποστηρίζοντας επίσης τελεστές boolean και αναζητήσεις φράσεων. Η Google αντιλαμβάνεται το κενό μεταξύ των λέξεων ως τον τελεστή AND, ενώ περαιτέρω μπορεί να γίνει αναζήτηση ακριβών εκφράσεων ή αποκλεισμός λέξεων από τα αποτελέσματα αναζήτησης.<sup>4</sup>

Παρέχει περισσότερες από 20 επιπλέον ειδικές λειτουργίες σε σχέση με την αρχική ικανότητα αναζήτησης για λέξη ή φράση. Σε αυτές περιλαμβάνονται συνώνυμα, προγνώσεις καιρού, χρονικές ζώνες, τιμές μετοχών, χάρτες, δεδομένα σεισμών, ώρες προβολής ταινιών και αποτελέσματα αγώνων. Προσφέρει ειδικές λειτουργίες για ημερομηνίες, τιμές, θερμοκρασίες, μετατροπές χρημάτων ή μονάδων, μαθηματικούς υπολογισμούς καθώς και μετάφραση των εμφανιζόμενων ιστοσελίδων σε διάφορες γλώσσες. Επίσης παρέχει τη δυνατότητα στους χρήστες να αναζητήσουν πληροφορίες στο διαδίκτυο μέσω φωνής (Google Voice Search) ή δίνοντας κάποια εικόνα (Search by Image).

Ωστόσο αυτό που έκανε την Google σπουδαία είναι η ακρίβεια με την οποία η μηχανή αναζήτησης μπορεί να επιστρέψει αποτελέσματα αναζήτησης, η οποία αναπτύχθηκε όταν οι σχεδιαστές της Google συνδύασαν τις λέξεις-κλειδιά με διάσημους συνδέσμους. Ο συνδυασμός αυτός προσέδωσε μία μεγαλύτερη ακρίβεια στην κατάταξη της σελίδας συγκριτικά με τις λέξεις κλειδιά από μόνες τους. Στο πλαίσιο αυτό είναι σημαντικό όμως να κατανοηθεί ότι οι λέξεις-κλειδιά και οι σύνδεσμοι αποτελούν δύο μόνο από τις εκατοντάδες των διαφόρων κριτηρίων που μπορεί να χρησιμοποιήσει μία μηχανή αναζήτησης για την κατάταξη των ιστοσελίδων μέσω της σύγκρισης μεταξύ τους.<sup>5</sup>



## *Bing<sup>6</sup>*

Η Bing (γνωστή τα προηγούμενα χρόνια ως MSN Search, Windows Live Search και Live Search) είναι η μηχανή αναζήτησης της Microsoft (διαφημίζεται και ως «μηχανή λήψης αποφάσεων»). Το 1998 ξεκίνησε η λειτουργία της MSN Search από την Microsoft, η οποία έδειχνε τα αποτελέσματα αναζήτησης της Inktomi. Στις αρχές του 1999 παρουσίαζε αποτελέσματα από την Inktomi σε συνδυασμό με αποτελέσματα από την Looksmart, ενώ για ένα σύντομο χρονικό διάστημα μέσα στο 1999 παρουσίαζε αποτελέσματα από την AltaVista. Από το 2004 και μετά ξεκίνησε να χρησιμοποιεί τη δική της τεχνολογία αναζήτησης. Τα αποτελέσματα για αναζήτηση εικόνων (image search) παρέχονταν από την Picsearch. Το 2006 μετονομάστηκε σε Windows Live Search ξεκινώντας τη λειτουργία των δικών της αλγόριθμων αναζήτησης εικόνων, ενώ αργότερα το 2007 μετονομάστηκε σε Live Search. Η Bing παρουσιάστηκε για πρώτη φορά το Μάιο του 2009 από το CEO της Microsoft Steve Ballmer σε ένα συνέδριο στο San Diego της California. Εμφανείς διαφοροποιήσεις της ήταν οι λίστες με προτάσεις προς αναζήτηση, καθώς και λίστες με παρόμοιες αναζητήσεις που πραγματοποιήθηκαν, βασισμένες στην τεχνολογία semantic της Powerset, η οποία αγοράστηκε από την Microsoft το 2008.



Στις 29 Ιουλίου του 2009, η Microsoft και η Yahoo ανακοίνωσαν μια συμφωνία κατά την οποία η Bing θα τροφοδοτούσε το Yahoo!Search. Τον Οκτώβριο του 2011, η Microsoft δήλωσε ότι ετοιμάζε μια νέα έκδοση κώδικα για τη μηχανή αναζήτησης, η οποία αποσκοπούσε στην γρηγορότερη και με μεγαλύτερη σαφήνεια παράδοση αποτελεσμάτων στους χρήστες. Η νέα αναβάθμιση (γνωστή και ως Tiger) συγχωνεύτηκε στον κώδικα της Bing τον Αύγουστο του 2011. Το Μάιο του 2012 η Microsoft ανακοίνωσε άλλη μια αναβάθμιση της μηχανής της, η οποία θα περιλάμβανε το Sidebar, μια υπηρεσία που θα έψαχνε για πληροφορίες σχετικές με τα ερωτήματα στα κοινωνικά δίκτυα των χρηστών.

Η Bing, όπως και η Google, παρέχει και άλλες ειδικές λειτουργίες στους χρήστες οι οποίες περιλαμβάνουν, μεταξύ άλλων, άμεσα αποτελέσματα για αθλητικούς αγώνες, πληροφορίες για μετοχές, μετατροπές μονάδων, μαθηματικούς υπολογισμούς και τοπική αναζήτηση.



## *Yahoo!*<sup>7</sup>

Η Yahoo! είναι εταιρεία διαδικτυακών υπηρεσιών και ένας από τους πιο παλιούς καταλόγους ιστοσελίδων (web directory). Ξεκίνησε ως θεματικός κατάλογος που αργότερα όμως εξελίχτηκε και σε μία μηχανή αναζήτησης. Ιδρύθηκε τον Ιανουάριο του 1994 από τους τότε τελειόφοιτους του Stanford University, David Filo και Jerry Yang, οι οποίοι ήθελαν να ομαδοποιήσουν και να καταγράψουν ιστοσελίδες μείζονος ενδιαφέροντος και να τις ταξινομήσουν σε θεματικές ενότητες. Η ιστοσελίδα που δημιούργησαν ονομάστηκε αρχικά «Jerry and David's guide to the World Wide Web» και στη συνέχεια τον Μάρτιο του 1994 μετονομάστηκε σε Yahoo!. Στις 18 Ιανουαρίου 1995 δημιουργήθηκε και το domain «yahoo.com».

# YAHOO!

Αρχικά η μηχανή αναζήτησης της Yahoo βρισκόταν σε μια μηχανή που έστελνε ερωτήματα σε έναν κατάλογο στον οποίο ήταν τοποθετημένες όλες οι δηλωμένες στη Yahoo ιστοσελίδες. Τα αποτελέσματα εμφανίζονταν στο χρήστη υπό το brand της Yahoo, αλλά στην πραγματικότητα η αναζήτηση αυτή δε γινόταν ποτέ από τη Yahoo. Μέχρι το 2001 οι αναζητήσεις γίνονταν μέσω της Inktomi και αργότερα μέσω της Google μέχρι το 2004, όπου η Yahoo έγινε ανεξάρτητη αναπτύσσοντας το δικό της Web Crawler. Τον Ιούλιο του 2009 η Yahoo και η Microsoft ανακοίνωσαν μια συμφωνία τους κατά την οποία η Bing θα τροφοδοτούσε πλέον τα αποτελέσματα αναζήτησης της Yahoo.



## *Baidu*<sup>8</sup>

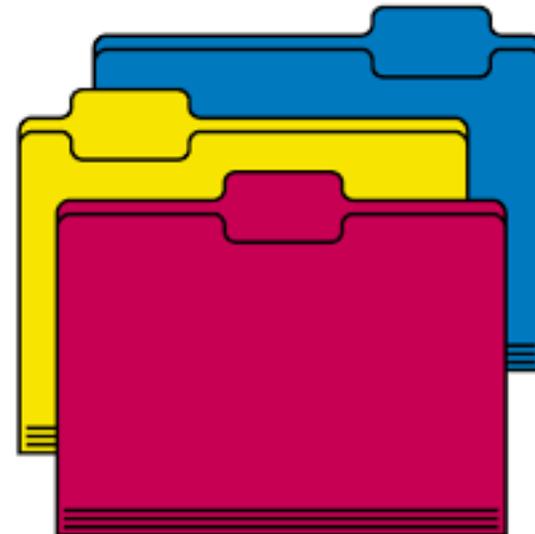
Η Baidu ιδρύθηκε το 2000 από τους Robin Li και Eric Xu. Η εταιρεία προσφέρει ένα μεγάλο φάσμα διαδικτυακών υπηρεσιών και κυρίως μια διαδικτυακή μηχανή αναζήτησης για εικόνες, βίντεο και ιστοσελίδες, χρησιμοποιώντας ως κύρια γλώσσα τα Κινέζικα. Η Baidu δραστηριοποιείται στην Κινέζικη αγορά, αποτελώντας την πιο διαδεδομένη μηχανή αναζήτησης στην Κίνα.

Για τη δημιουργία της χρησιμοποιήθηκε ένας αλγόριθμος βαθμολόγησης ιστοσελίδων για την κατάταξή τους από μηχανή αναζήτησης με την ονομασία RankDex, ο οποίος αναπτύχθηκε το 1996 από τον Robin Li. Η Baidu προσφέρει πολλές λειτουργίες για τον εντοπισμό πληροφοριών, προϊόντων και υπηρεσιών, χρησιμοποιώντας όρους αναζήτησης στην Κινεζική γλώσσα. Μερικές από αυτές είναι η φωνητική αναζήτηση, τιμές μετοχών, ειδήσεις, πληροφορίες για τον καιρό και δρομολόγια τρενών και αεροπλάνων.



## Ταξινόμηση μηχανών αναζήτησης

Σύμφωνα με τον Jerri Ledford (2009) στο βιβλίο του «Search Engine Optimization Bible», οι μηχανές αναζήτησης μπορούν να ταξινομηθούν σε τρεις διαφορετικούς τύπους: (α) πρωτοβάθμιες, (β) δευτεροβάθμιες και (γ) στοχευόμενες μηχανές. Η κατάταξη αυτή γίνεται βάσει του τρόπου εξαγωγής των αποτελεσμάτων αλλά και των ιδιοτήτων της κάθε μηχανής αναζήτησης.



# Πρωτοβάθμιες μηχανές αναζήτησης

Πρόκειται για τις πλέον γνωστές μηχανές αναζήτησης στους χρήστες του διαδικτύου. Συνήθως όταν γίνεται αναφορά σε πρώτου βαθμού μηχανές αναζήτησης εννοούνται η Google, η Yahoo!, το MSN. Οι μηχανές αυτές παράγουν το μεγαλύτερο μέρος της κυκλοφορίας σε μια ιστοσελίδα και γι' αυτό θα αποτελούν τον πρωταρχικό στόχο στην προσπάθεια εφαρμογής μιας στρατηγικής SEO. Όλες οι πρωτοβάθμιες μηχανές αναζήτησης διαφέρουν ελαφρώς μεταξύ τους όσον αφορά τα αποτελέσματα αναζήτησης (SERPs) τα οποία οφείλονται στους διαφορετικούς αλγορίθμους που χρησιμοποιούν.

## Πρωτοβάθμιες μηχανές αναζήτησης

Οι περισσότερες πρωτοβάθμιες μηχανές αναζήτησης δεν παρέχουν στο χρήστη μόνο την προκαθορισμένη αναζήτηση αλλά επιπρόσθετες δυνατότητες και χαρακτηριστικά όπως e-mail, τις ειδήσεις, τη χαρτογράφηση, καθώς και τα διάφορες εφαρμογές ψυχαγωγίας. Τα παραπάνω στοιχεία προστέθηκαν πολύ καιρό αργότερα ως ένα μέσο προώθησης των μηχανών με αποτέλεσμα την αύξηση της επισκεψιμότητάς τους. Παρά το γεγονός ότι αυτά τα χαρακτηριστικά δεν επηρεάζουν τον τρόπο αναζήτησης, θα μπορούσαν να επηρεάσουν τα κριτήρια των χρηστών για την επιλογή της μηχανής αναζήτησης.<sup>9</sup>

## Δευτεροβάθμιες μηχανές αναζήτησης

Αυτές οι μηχανές αναζήτησης απευθύνονται σε μικρότερο, πιο συγκεκριμένο κοινό στόχο, αν και το περιεχόμενό τους είναι ίδιο γενικά. Δεν συγκεντρώνουν τόσο μεγάλη κίνηση σε σχέση με τις πρωτοβάθμιες μηχανές, αλλά είναι χρήσιμες και εστιάζονται περισσότερο στις περιφερειακές αναζητήσεις. Παραδείγματα τέτοιων μηχανών είναι οι Ask.com, LookSmart, Miva και η Espotting.

Οι δευτεροβάθμιες μηχανές αναζήτησης μοιάζουν με τις πρωτοβάθμιες, απλά διαφέρουν ως προς τον τρόπο κατάταξης των αποτελεσμάτων. Μερικές από αυτές στηρίζονται σε μεγάλο βαθμό από τις λέξεις-κλειδιά, ενώ άλλες βασίζονται στους αμοιβαίους συνδέσμους των ιστοσελίδων. Κάποιες άλλες μπορεί να έχουν στηριχθεί σε κριτήρια όπως meta tags ή κάποια άλλα κριτήρια, τα οποία να προσαρμόζονται ανάλογα με την πολιτική της εταιρείας.<sup>10</sup>

## Δευτεροβάθμιες μηχανές αναζήτησης

Οι δευτεροβάθμιες μηχανές αναζήτησης θα πρέπει επίσης να συμπεριλαμβάνονται σε κάθε σχέδιο SEO. Παρά το γεγονός ότι δεν παρουσιάζουν τόσο μεγάλη κίνηση αναφορικά με την αναζήτηση των αποτελεσμάτων, αυτές οι μηχανές αναζήτησης σίγουρα δε θα πρέπει να αγνοηθούν καθώς ακόμα και αυτή η περιορισμένη κίνηση που παράγουν είναι πολύτιμη.

## Στοχευμένες μηχανές αναζήτησης

Οι στοχευόμενες μηχανές αναζήτησης (συχνά συναντώνται και ως επίκαιρες) περιέχουν συγκεκριμένη αναζήτηση, εστιάζοντας την προσοχή σε ειδικά θέματα, όπως είναι ο τομέας της πληροφορικής ή διάφορες πληροφορίες για φάρμακα ή διάφοροι κλάδοι της ιατρικής και της επιστήμης, τα ταξίδια, τον αθλητισμό, ή κάποιο άλλο γενικό θέμα. Παραδείγματα τέτοιων μηχανών είναι η Yahoo!, το CitySearch, Travel, Music Search. Όπως συμβαίνει και στους άλλους τύπους των μηχανών αναζήτησης, έτσι και εδώ τα κριτήρια κατάταξης των αποτελεσμάτων διαφέρουν από τη μια μηχανή στην άλλη.

Αναφορικά με το SEO, για αυτού του είδους τις μηχανές αναζήτησης πρέπει να ληφθεί υπόψη ότι πολλές από αυτές είναι πολύ πιο περιορισμένης εμβέλειας μηχανές αναζήτησης συγκριτικά με τις πρωτοβάθμιες και τις δευτεροβάθμιες, και συνήθως βασίζονται σε συγκεκριμένα θέματα. Στο πλαίσιο αυτό θα πρέπει να βρεθούν εκείνες οι στοχευόμενες μηχανές που είναι σχετικές με το συγκεκριμένο θέμα αναζήτησης.<sup>11</sup>

## Τύποι μηχανών αναζήτησης

Σύμφωνα με τον εκπρόσωπο της Google, Danny Sullivan, ο όρος «μηχανή αναζήτησης» χρησιμοποιείται γενικά για την περιγραφή των crawler-based search engines αλλά και των human-powered directories. Οι δύο αυτοί τύποι των μηχανών αναζήτησης συλλέγουν τις καταχωρήσεις τους με ριζικά διαφορετικούς τρόπους, που σχετίζονται με την συγκέντρωση των καταχωρήσεων. Οι crawler-based search engines χρησιμοποιούν μηχανισμούς που είναι αυτόματοι στην ανίχνευση ιστοσελίδων που βρίσκονται στον παγκόσμιο ιστό, ενώ στους human-powered directories η εργασία καταχώρησης πραγματοποιείται από άνθρωπο. Ωστόσο υπάρχουν και οι hybrid search engines που αποτελούν συνδυασμό των δύο προηγούμενων.

## Crawler-based search engines

Οι crawler-based search engines χρησιμοποιούν ένα κομμάτι λογισμικού για τη συλλογή πληροφοριών σχετικά με τις ιστοσελίδες του διαδικτύου. Τέτοιου είδους πληροφορίες είναι συνήθως οι λέξεις κλειδιά, το URL της σελίδας, οι σύνδεσμοι από και προς την ιστοσελίδα και ο κώδικας δημιουργίας της. Στη συνέχεια αυτές οι πληροφορίες αποθηκεύονται με τη σειρά τους σε μία βάση δεδομένων, η οποία αποτελεί το ευρετήριο μιας μηχανής αναζήτησης. Αυτό το κομμάτι του λογισμικού ονομάζεται αράχνη ή διαφορετικά ανιχνευτής (crawler) ή καμία φορά και robot.

## Crawler-based search engines

Οι μηχανές αυτές είναι πολύ αποδοτικές ειδικά όταν το ερώτημα που τους τίθεται είναι συγκεκριμένο. Ωστόσο όταν η αναζήτηση πραγματοποιηθεί με κάτι που δεν είναι ακριβές, η μηχανή αναζήτησης μπορεί να εμφανίσει ως αποτελέσματα πολλά που δεν έχουν σχέση με αυτό που ο χρήστης είχε στο μυαλό του.<sup>12</sup>

## Human-powered directories

Οι human-powered directories (θεματικοί κατάλογοι στα ελληνικά) επιτρέπουν τη καταχώρηση ιστοσελίδων από ανθρώπους. Μερικές από τις πιο δημοφιλείς μηχανές αυτού του τύπου είναι το Yahoo Directory, το Open Directory, Mahalo και Rollyo. Η εγγραφή στον κατάλογο γίνεται μετά την υποβολή μιας σύντομης περιγραφής του ενδιαφερόμενου ιδιώτη της ιστοσελίδας στον κατάλογο σχετικά με την ιστοσελίδα του, είτε διαφορετικά των συντακτών, μιας ομάδας εξειδικευμένων ανθρώπων που σχολιάζουν και αξιολογούν μια ιστοσελίδα. Συνήθως έχουν ιεραρχική δομή και κατατάσσουν τις πληροφορίες κατά θέμα, δηλαδή περιλαμβάνουν μεγάλες θεματικές περιοχές οι οποίες με τη σειρά τους χωρίζονται σε μικρότερες πιο ειδικές κατηγορίες.<sup>13</sup>

## Human-powered directories

Οι χρήστες μπορούν να εισχωρήσει από μία μεγάλη θεματική κατηγορία σε μία πιο ειδική υποκατηγορία και τελικά να οδηγηθούν σε μία συγκεκριμένη ιστοσελίδα. Στην αρχική σελίδα εμφανίζεται μια λίστα με τις βασικές θεματικές κατηγορίες. Μετά την επιλογή οι χρήστες μεταφέρονται σε μια δεύτερη σελίδα που εμφανίζονται οι υποκατηγορίες της, ενώ δίπλα από κάθε υποκατηγορία εμφανίζεται ένας αριθμός που δηλώνει τον αριθμό των καταχωρήσεων που περιλαμβάνει αυτή η υποκατηγορία. Οι κατάλογοι αυτοί είναι χρήσιμοι όταν υπάρχει ενδιαφέρον για ένα ειδικό θέμα αναζήτησης. Στην περίπτωση αυτή μπορούν να συγκεκριμενοποιήσουν μία αναζήτηση, επιστρέφοντας πίσω πιο εξευγενισμένα και πιο ακριβή αποτελέσματα.<sup>14</sup>

## Hybrid search engines

Οι υβριδικές μηχανές αποτελούν συνδυασμό των crawled-based και των human-powered directories. Αυτό σημαίνει ότι υπάρχει ένας web crawler που ψάχνει στο διαδίκτυο για τις ιστοσελίδες που πρόκειται να συμπεριληφθούν, ενώ την ίδια στιγμή ο ανθρώπινος παράγοντας παίζει ρόλο για τη χειροκίνητη υποβολή των ιστοσελίδων τους στις μηχανές αναζήτησης.

## Τρόπος λειτουργίας μηχανών αναζήτησης

Οι μηχανές αναζήτησης αποτελούνται από τρία διαφορετικά είδη λογισμικού, η σύμπραξη των οποίων προσδίδει την πλήρη λειτουργία τους. Τα τμήματα αυτά είναι το spider software (αράχνες-ερευνητές των μηχανών αναζήτησης), το index software (λογισμικό περιεχομένου) και το query software (λογισμικό όρων αναζήτησης).<sup>15</sup>

## Τρόπος λειτουργίας μηχανών αναζήτησης

Το spider software είναι η λεγόμενη «αράχνη» που χρησιμοποιούν οι μηχανές αναζήτησης. Το πρόγραμμα αυτό, προσπελαύνει ολόκληρο το διαδίκτυο αναζητώντας νέες ιστοσελίδες για να τις προσθέσει στις βάσεις δεδομένων των μηχανών αναζήτησης. Ο τρόπος με τον οποίο λειτουργεί είναι παρόμοιος με οποιοδήποτε browser (π.χ. Google Chrome ή Mozilla Firefox). Η διαφορά τους είναι ότι τα spiders αγνοούν το design και τα γραφικά μιας ιστοσελίδας, και ασχολούνται αποκλειστικά με το κείμενο και το υπερκείμενο (URLs, links, κ.ά.) που μπορεί να περιέχει μια ιστοσελίδα. Θα μπορούσε να ειπωθεί ότι τα spiders είναι «κυνηγοί» κειμένου και υπερκειμένου, που εστιάζουν στον τρόπο που αυτά είναι τοποθετημένα σε μια σελίδα, καθώς και στον τρόπο με τον οποίο συνδέουν τις διάφορες σελίδες μεταξύ τους, από που προέρχονται και που οδηγούν.

## Τρόπος λειτουργίας μηχανών αναζήτησης

Το index software είναι το λογισμικό που διαθέτουν οι μηχανές αναζήτησης και αποτελεί τον αποδέκτη της έρευνας των spiders, καθώς δέχεται τα δεδομένα που αυτά έχουν συλλέξει. Το πρόγραμμα διαχωρίζει τον κύριο όγκο ενός κειμένου από τα links και τα URLs μέσω κατάλληλου αλγορίθμου που ταξινομεί τις λέξεις, τα ζευγάρια λέξεων και τις φράσεις με τέτοιο τρόπο που επιτρέπει στη μηχανή αναζήτησης να κατανοήσει το πόσο σημαντική είναι μια σελίδα σχετικά με θέμα αναζήτησης του χρήστη. Κατόπιν αποθηκεύει τα αποτελέσματα της κάθε έρευνας προκειμένου να μπορούν να χρησιμοποιηθούν ξανά.

## Τρόπος λειτουργίας μηχανών αναζήτησης

Το query software είναι αυτό που βλέπουν οι χρήστες όταν χρησιμοποιούν μια μηχανή αναζήτησης σε κάποιον browser. Πρόκειται δηλαδή για τη σελίδα με το πεδίο αναζήτησης, η οποία όταν συμπληρωθεί θα δώσει σε επόμενη σελίδα τα αποτελέσματα. Το βασικό χαρακτηριστικό αυτής της σελίδας είναι ακριβώς το πεδίο της αναζήτησης, δηλαδή το πλαίσιο στο οποίο οι χρήστες πληκτρολογούν τα θέματα προς έρευνα. Κάποιος πληκτρολογεί φράσεις σχετικές με το πεδίο αναζήτησής του και η μηχανή εμφανίζει τα καλύτερα δυνατά αποτελέσματα, έχοντας ερευνήσει ολόκληρο το διαδίκτυο. Βέβαια αυτό δεν είναι τελείως ακριβές, καθώς μια μηχανή αναζήτησης δεν «ψάχνει το διαδίκτυο» αλλά αυτό που πραγματικά κάνει είναι έλεγχο των καταλόγων της που δημιουργήθηκαν προηγουμένως από το index software, βάσει του υλικού που συλλέχθηκε από τα spiders.

## Τρόπος λειτουργίας μηχανών αναζήτησης

Τα τρία αυτά λογισμικά συνεργάζονται, ολοκληρώνοντας μαζί μια μηχανή αναζήτησης. Το spider software εξερευνά σε τακτική βάση ολόκληρο το διαδίκτυο και συλλέγει υλικό που αποτελείται από κείμενα, υπέρ-συνδέσμους και URLs, το index software ταξινομεί και «τακτοποιεί» το παραπάνω υλικό σε βάσεις δεδομένων, ενώ το query software δέχεται τα ερωτήματα από τους χρήστες και εξερευνά αυτές τις βάσεις δεδομένων για να προβάλει τα αποτελέσματα που προκύπτουν.

Οι μηχανές αναζήτησης λοιπόν έχουν ήδη κάνει την έρευνα και την κατάταξη-διαχωρισμό των πληροφοριών στο δίκτυο και διαθέτουν αυτή την πληροφορία σε όποιον πραγματοποιεί την κατάλληλη αναζήτηση στην ιστοσελίδα τους. Οι λέξεις που χρησιμοποιούν οι χρήστες κατά την αναζήτηση των αποτελεσμάτων που παρουσιάζονται ονομάζονται keywords (λέξεις κλειδιά).

## Τρόπος λειτουργίας μηχανών αναζήτησης

Η χρησιμότητα και συνεπώς η επιτυχία μιας μηχανής αναζήτησης εξαρτάται από τα αποτελέσματα που εμφανίζει. Ενώ μπορεί να υπάρχουν εκατομμύρια ιστοσελίδες που περιέχουν τη λέξη ή τη φράση που ψάχνει ο χρήστης, είναι δουλειά της μηχανής αναζήτησης να εμφανίσει τα καλύτερα αποτελέσματα ανάλογα με τη σχετικότητα, τη δημοφιλία και την ποιότητά τους. Οι μηχανές αναζήτησης χρησιμοποιούν διάφορες μεθόδους για να κατατάξουν τα αποτελέσματά τους και να εμφανίσουν τα καλύτερα από αυτά στις πρώτες θέσεις των Search Engine Result Pages (SERPs).

## Τρόπος λειτουργίας μηχανών αναζήτησης

Αυτές οι μέθοδοι είναι διαφορετικές από μηχανή σε μηχανή και αλλάζουν με τα χρόνια ανάλογα με τη χρήση του διαδικτύου και την ανάπτυξη νέων τεχνικών. Λεπτομέρειες γι' αυτές τις μεθόδους παραμένουν μυστικές. Ωστόσο είναι γνωστό ένα γενικό περίγραμμα, το οποίο αποτελείται από ένα συνδυασμό οκτώ (8) παραγόντων χωρισμένων σε τρεις (3) κατηγορίες:<sup>16</sup>

# Τρόπος λειτουργίας μηχανών αναζήτησης

- (1) Μέτρηση της σπουδαιότητας της σελίδας ανεξάρτητα από το ερώτημα
  - Με βάση τους συνδέσμους που «δείχνουν» στη σελίδα (PageRank).
  - Με βάση άλλες εκτιμήσεις (π.χ. ιστότοποι όπως η Wikipedia ή σελίδες με .edu TLD [Top-Level Domain] θεωρούνται υψηλής ποιότητας).

# Τρόπος λειτουργίας μηχανών αναζήτησης

(2) Μέτρηση της σχετικότητας της σελίδας ανάλογα με το ερώτημα

- Με βάση το περιεχόμενο της σελίδας.
- Με βάση τους συνδέσμους που «δείχνουν» στη σελίδα.
- Με βάση τα αρχεία καταγραφής ερωτημάτων (query logs): Τα περισσότερα ερωτήματα που δέχεται η μηχανή αναζήτησης τα έχει λάβει πολλές φορές στο παρελθόν. Τα query logs καταγράφουν τις σελίδες που οι χρήστες επέλεξαν να επισκεφτούν, σε κάποιο βαθμό, ακόμη και το χρόνο που πέρασαν σε αυτές.
- Εξατομικευμένη Αναζήτηση: Πολλές μηχανές αναζήτησης, όπως η Google και η Bing, εμφανίζουν προσαρμοσμένα αποτελέσματα ανάλογα με το ιστορικό δραστηριοτήτων του χρήστη (π.χ. τα ερωτήματα που έχει κάνει, τους συνδέσμους που έχει επισκεφτεί).
- Με βάση συγκεκριμένες απαντήσεις σε γνωστά ερωτήματα (π.χ. το πρώτο αποτέλεσμα για το ερώτημα «convert grams to ounces» είναι και η απάντησή του).

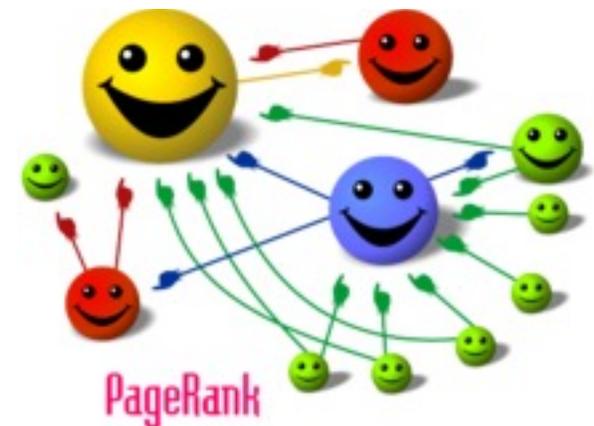
## Τρόπος λειτουργίας μηχανών αναζήτησης

### (3) Διασπορά του Θέματος

- Αν ένας όρος αναζήτησης έχει πολλές σημασίες, τότε οι σημαντικότερες εξ αυτών θα πρέπει να βρίσκονται μεταξύ των πρώτων αποτελεσμάτων, ανεξάρτητα από την ενυπάρχουσα σπουδαιότητα της σελίδας.

# PageRank

Το PageRank (μονάδα βαθμολόγησης της αξίας μιας ιστοσελίδας κατά τη Google) είναι μία αριθμητική τιμή η οποία αντιπροσωπεύει το πόσο σημαντική και σπουδαία είναι μία ιστοσελίδα στο διαδίκτυο, με κριτήρια που περιέχονται στον αλγόριθμο της Google μηχανής αναζήτησης. Το PageRank (PR) μαζί με κάποιους επιπλέον παράγοντες, επηρεάζει σημαντικά την κατάταξη στα αποτελέσματα αναζήτησης. Για τον προσδιορισμό του χρησιμοποιείται ένας πολύπλοκος αλγόριθμος που περιλαμβάνει εκατομμύρια μεταβλητές και όρους, ενώ ο υπολογισμός του γίνεται με αδιάβλητο αυτοματοποιημένο τρόπο.



# PageRank

Η βαθμολόγηση της ιστοσελίδας γίνεται με βάση τον αριθμό των links που δείχνουν προς την σελίδα και τη βαθμολογία αυτών των. Μαθηματικά ο τύπος που δίνει το PR μιας σελίδας είναι μια αναδρομική ακολουθία Markov η οποία:

- (1) υπολογίζει τον αριθμό των links που δείχνουν προς την σελίδα,
- (2) χρησιμοποιεί το PR των σελίδων στις οποίες βρίσκει τα link ως βάρη,
- (3) πολλαπλασιάζει το βάρος της κάθε σελίδας με έναν συντελεστή και
- (4) αθροίζει το PR που δίνει στην σελίδα κάθε ξεχωριστό link υπολογίζοντας το συνολικό PR της σελίδας.



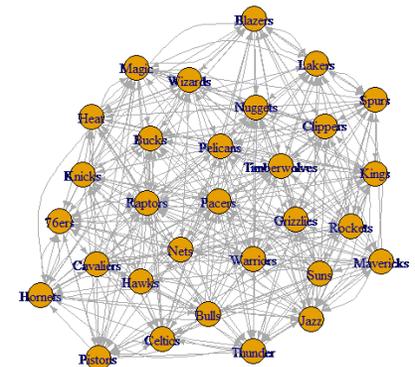
## PageRank

Η κλίμακα διαβάθμισης του PR είναι από το 1 έως το 10 (PR1-PR10). Οι αριθμοί αυτοί είναι φυσικά ενδεικτικοί και αντιστοιχούν σε μια εκθετική κλίμακα που ξεκινάει από το 1 link και φτάνει στα εκατομμύρια. Ο αλγόριθμος του Google ξεχωρίζει πολλές υποδιαίρεσεις μεταξύ των 10 αυτών ψηφίων, αλλά στο κοινό είναι διαθέσιμες αυτές οι βαθμολογίες που αποτελούν στρογγυλοποίηση. Το ακριβές PR της σελίδας δηλαδή είναι κρυφό και μπορεί κανείς να δει μόνο την βαθμολογία 0-10. Όσο μεγαλύτερο PR προσλαμβάνει μια ιστοσελίδα, τόσο υψηλότερη θέση έχει στην κατάταξη στα αποτελέσματα αναζήτησης (SERPs). Αυτός δεν είναι ο μόνος παράγοντας που χρησιμοποιεί η Google για την ταξινόμηση των σελίδων, αλλά είναι πολύ σημαντικός.



# PageRank

Άλλος σημαντικός παράγοντας αξιολόγησης είναι το περιεχόμενο των σελίδων. Η Google αναλύει όχι μόνο τα meta tags αλλά το σύνολο του περιεχομένου της ιστοσελίδας. Λαμβάνονται υπόψη πολλοί παράμετροι όπως οι τίτλοι, το μέγεθος και ο χρωματισμός των γραμματοσειρών, οι παράγραφοι, οι λέξεις και η συνάφειά τους με το κείμενο, καθώς και το υπόλοιπο περιεχόμενο των άλλων ιστοσελίδων. Όπως είχε αναφέρει ο Bill Gates το 1996 «Content is King», τονίζοντας τη μεγάλη σημασία του περιεχομένου αλλά και της ποιότητας του, στην αξιολόγηση μίας ιστοσελίδας από τις μηχανές αναζήτησης.



# PageRank

Ο δείκτης PR συχνά υπερτιμάται, καθώς στην πραγματικότητα υπάρχουν ιστοσελίδες με μηδενικό PR, οι οποίες εμφανίζονται υψηλότερα από ιστοσελίδες με πολύ υψηλό PR. Ο λόγος είναι ότι το PR αποτελεί έναν μόνο από τους παράγοντες που προσμετρούνται για την κατάταξη. Εξάλλου, από κάποιο χρονικό σημείο και μετά, υπήρξε η άνθιση ορισμένων εταιρειών που εξασφάλιζαν στους πελάτες τους υψηλό PR, πουλώντας απλά τους υπερσυνδέσμους από σελίδες με υψηλό PR. Όπως ήταν αναμενόμενο, ο αλγόριθμος του PR διαφοροποιήθηκε από τη Google και η σημασία του στις λίστες κατάταξης μειώθηκε.<sup>17</sup>

