

Joni Hänninen

**BING.COM -
PÄÄTÖKSENTEKKOKONE
ja hakukoneiden historia**

Opinnäytetyö
Tietotekniikka


Toukokuu 2010




MIKKELIN AMMATTIKORKEAKOULU

Mikkeli University of Applied Sciences

KUVAILULEHTI

 MIKKELIN AMMATTIKORKEAKOULU Mikkeli University of Applied Sciences	Opinnäytetyön päivämäärä 	
Tekijä(t) Joni Hänninen	Koulutusohjelma ja suuntautuminen Tietotekniikan koulutusohjelma	
Nimeke BING.com- päätöksentekokone ja hakukoneiden historia.		
Tiivistelmä <p>Opinnäytetyöni käsittelee hakukoneiden historiaa, sekä lähemmin Microsoftin viime kesänä julkaisemaa Bing nimistä hakukonetta, joka yrittää saada osan Googlen dominoimasta hakumarkkinoista itselleen. Microsoft tarjoaa hakukoneellaan erilaisia ominaisuuksia, jotka auttavat ihmisiä päätöksenteossa oli kyseessä sitten lomamatkan tai häiden suunnittelu.</p> <p>Teoriaosassa esittelen Internetin ja World Wide Webin syntyä, ja miten hakukoneet alkoivat yleistyä 1990 luvulta lähtien. Teoriaosuuden päättää Bing- hakukoneen esittely.</p>		
Asiasanat (avainsanat) Hakukone, hakuohjelmat, haku, Internet		
Sivumäärä 43 s.	Kieli Suomi	URN URN:NBN:fi:amk-2010053111044
Huomautus (huomautukset liitteistä) 5 s.		
Ohjaavan opettajan nimi Timo Mynttinen	Opinnäytetyön toimeksiantaja 	

DESCRIPTION

 MIKKELIN AMMATTIKORKEAKOULU Mikkeli University of Applied Sciences		Date of the bachelor's thesis	
Author(s) Joni Hänninen		Degree programme and option Information Technology	
Name of the bachelor's thesis BING.COM – The decision making engine and the history of search engines.			
Abstract My thesis deals with the history of search engines, and more specific about a search engine called Bing that Microsoft launched last summer to gain some market share of searches from the dominant Google. Microsoft offers variety of features with its search engine that will guide people in decision making whether they are planning a vacation or weddings. In the theory part of the thesis I talk about the birth of the Internet and World Wide Web, and how search engines became popular since the 90's. An introduction to the Bing search engine will end the theory part.			
Subject headings, (keywords) search engine, search applications, searching, Internet			
Pages 43 p.	Language Finnish	URN URN:NBN:fi:amk-2010053111044	
Remarks, notes on appendices 5 p.			
Tutor Timo Mynttinen		Bachelor's thesis assigned by	

SISÄLTÖ

1	JOHDANTO	1
2	INTERNETIN LYHYT HISTORIA	2
2.1	Ajatus verkosta	2
2.2	ARPANET	2
2.3	ARPANETistä Internetiin.....	4
2.4	Internet 1975-1988.....	4
2.5	Word Wide Web	6
3	HAKUKONE	7
3.1	Hakukoneiden historia	7
3.1.1	Ennen World Wide Webiä.....	7
3.1.1.1	Archie	8
3.1.1.2	Gopher.....	9
3.1.1.3	WAIS.....	10
3.1.2	Mosaic- selain	11
3.1.3	Hakukoneet World Wide Webin jälkeen	12
3.1.3.1	ALIWEB	13
3.1.3.2	Galaxy	13
3.1.3.3	Yahoo!.....	14
3.1.3.4	WebCrawler	15
3.1.3.5	Lycos	16
3.1.3.6	Excite.....	17
3.1.3.7	AltaVista	18
3.1.3.8	HotBot	19
3.1.3.9	Ask Jeeves	20
3.1.3.10	Northern Light.....	21
3.1.3.11	Google	22
3.2	Hakukoneen toimintaperiaate	23
3.2.1	Hakurobotti	25
3.2.1.1	Robots.txt	27
3.2.2	Tietokanta	29
3.2.3	Indeksointiohjelma.....	29
3.2.4	Palautusjärjestelmä	30

3.2.5	Graafinen liittymä	31
3.3	Hakukoneen käyttötarkoitus	31
3.4	Hakukoneen hyödyt ja haitat	32
3.5	Hakukoneiden tulevaisuus	32
4	BING - PÄÄTÖKSENTEKOKONE	33
4.1	Historia	33
4.2	Toiminta-ajatus	35
4.3	Ominaisuudet	36
4.3.1	Etusivu	36
4.3.2	Hakutulossivu	37
4.4	Markkinaosuus	41
4.5	Tulevaisuus	42
5	YHTEENVETO	42
6	LÄHTEET	44
7	LIITTEET	52

1 JOHDANTO

Ihmiset ovat siirtäneet jokapäiväiset asiat kuten pankissa käynnin, ostosten teon, postin lähettämisen ja tiedon etsimisen Internetiin. Harva nykyaikainen ihminen lähtee jonottamaan ruuhka-aikana pankkiin maksukseen laskujaan, kun sen voi tehdä työpaikalla verkkopankin kautta. Ostokset tehdään verkkokaupoissa, ja postista sitten haetaan se H&M- paketti tai jättitelevisio tuodaan kotiinkuljetuksena. Kirjeen kirjoittaminenkin on helpottunut, kun ei enää tarvitse etsiä oikeanlaista kirjepaperia, kirjekuorta tai postimerkkiä, saati sitten kiirehtiä ajoissa viemään se neljältä tyhjentyvään postilaatikkoon. Sähköpostilla voi lähettää pitkänkin novellin ystävälleen, ja liitteeksi voi laittaa hienon Powerpoint- esityksen, jossa on kauniita kuvia kissoista hattuineen kaikkineen. Facebookissa puolestaan kerrotaan puolittuntemattomille ”ystävilleskin” viikonlopun toilailuista, ja Twitterin avulla selviää, milloin julkkis on käymässä lähipubissa. Nuoria ihmisiä ei enää näy kirjastossa etsimässä tietoa Kolumbuksen löytöretkistä koulun jälkeen, kun tieto saadaan selville googlettamalla kotikoneelta - tietenkin sen jälkeen kun ensin on päivitetty status Facebookissa, käyty viljelemässä peltoa Farmvillessä sekä katsottu uudet Youtube- videot. Moderni yhteiskunta on siis siirtynyt elämään verkossa.

Verkko on muutakin kuin pelkkiä verkkokauppoja ja sosiaalisia verkostoja. Verkko pitää sisällään triljoonia niin hyödyllisiä kuin vaarallisiakin sivuja, mistä löytää tietoa joka lähtöön. Hakukoneet ovat kehitetty verkon käyttäjien hyväksi, jotta he löytäisivät etsimäänsä kysymykseen nopeasti vastauksen. Hakukoneelle vaan kerrotaan mitä etsit, niin se näyttää sinulle vaihtoehtosivuja, mistä kannattaa lähteä sitä tietoa etsimään.

Opinnäytetyössä puhun sekä Internetistä että verkosta, ja on varmaan hyvä käydä tässä vaiheessa läpi niiden erot. Miljoonat tietokoneet ympäri maailmaa muodostavat keskenään verkostoja, ja sitten verkostoja liitetään yhteen, jotta ne voisivat kommunikoida keskenään. Tämä verkostojen verkosto on nimeltään Internet. Verkko (World Wide Web) on puolestaan rakennettu Internetin päälle, eli verkon avulla pääsemme Internetin tietoihin käsiksi http- protokollan avulla. Verkkoselaimet, kuten Mozilla Firefox tai Google Chrome, päästävät meidät näkemään Internetissä olevia tietoja linkkien ja osoitteiden kautta, ja näemme palvelimille tallennetut html- sivut graafisen liittymän avulla selkokielellä.

2 INTERNETIN LYHYT HISTORIA

2.1 Ajatus verkosta

Leonard Kleinrock (1961) kehitti matemaattisen teorian tietoverkoista Kalifornian yliopistolle, UCLA:lle, tekemässään opinnäytetyössään vuonna 1961. Opinnäytetyön tarkoituksena oli tutkia tiedonkulun ongelmia suurissa viestintäverkoissa. Mahdolliset verkostot koostuisivat, solmuista, jotka ovat yhteydessä toisiinsa linkeillä. Solmut puolestaan vastaanottavat, lajittelevat, tallentavat ja lähettävät viestejä, jotka lähtevät tai tulevat solmuihin linkkien kautta. Kleinrock loi näin ollen perustan Internetille.

Lokakuussa 1965 käytiin maailman ensimmäinen tietokoneiden välinen digitaalisen verkoston kommunikointi kahden Massachusettsin teknisen yliopiston tutkijan, Lawrence Robertsin ja Thomas Marillin, tietokoneen välillä. Tämä oli myös ensimmäinen kerta, kun pakettitiedonsiirtoa käytettiin viestinnässä. (Roberts 1997.)

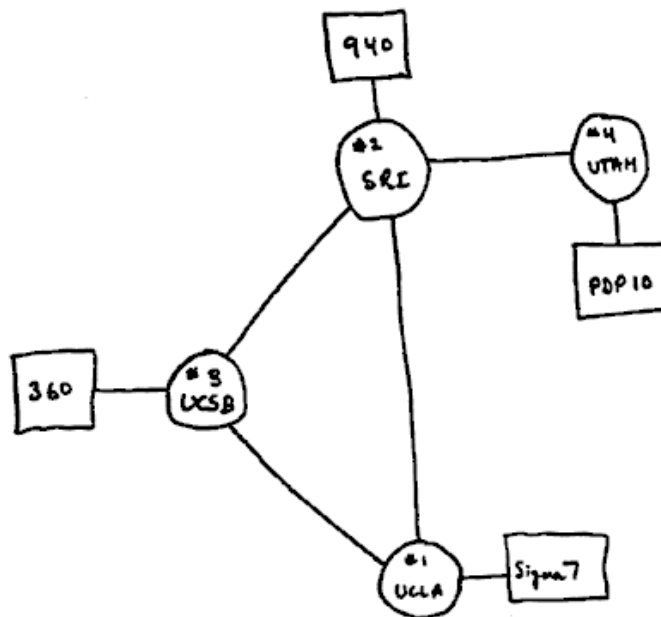
2.2 ARPANET

Lawrence Roberts (1997) palkattiin vuonna 1966 tutkijaksi ARPAan, Advanced Research Projects Agency, jota Yhdysvaltojen puolustusministeriö hallinnoi. Robertsin tehtävänä oli suunnitella ARPANET- verkosto, johon saataisiin luotettavasti kytkettyä monia tietokoneita yhteen resurssien jakamiseksi (Roberts 1999). Vuonna 1968 Kalifornian yliopiston verkon mittauskeskus NMC, Network Measurements Center, valittiin ensimmäiseksi solmukohdaksi ARPANETtiin, ja he saivat seuraavana vuonna käyttöönsä BBN – yrityksen valmistaman kytkimen. Tämä IMP, Interface Message Processor, oli pakettikytkentäinen solmukohta, mikä liittäisi verkostoja ARPANETtiin. (Internet Society 2010.) NMC:n tutkimuksia johti Leonard Kleinrock, ja hänen tehtävänään oli kytkeä ensimmäinen host, eli isäntäkone, IMP- kytkimeen. Kleinrockin (2005) mielestä tehtävänanto oli haastava, sillä sellaista yhteyttä ei ollut koskaan aiemmin yritetty.

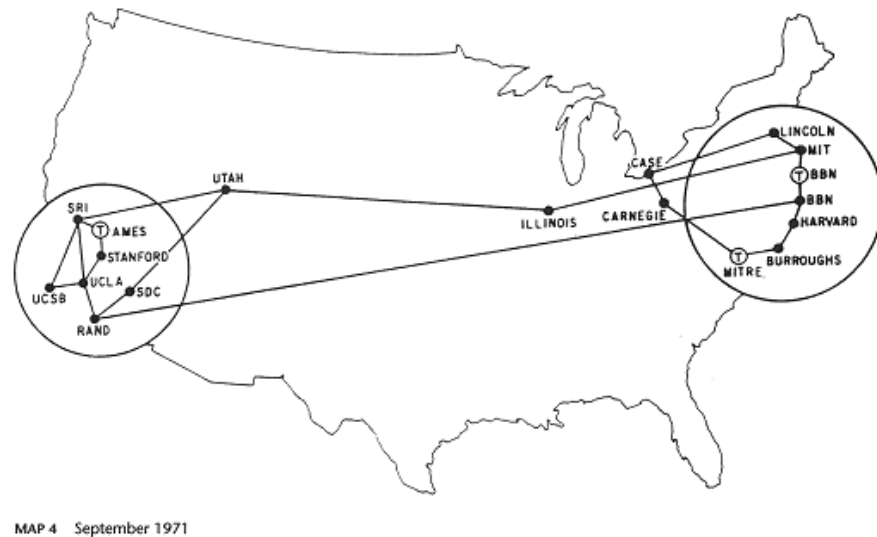
Kleinrockin (2005) tiimi onnistui lähettämään bittejä yliopiston koneelta IMP- kytkimeen tiistaina 2.9.1969, ja seuraavana päivänä jo viestejä. Kuukautta myöhemmin toinen solmu lisättiin Stanfordin yliopiston tutkimuslaitokseen, SRI, ja tarkoituksena

oli, että Kalifornian yliopiston opiskelijat kirjautuisivat Stanfordin tietokoneeseen, selaisivat sen tietokantoja ja yrittäisivät lähettää tietoa(Griffiths 2002a).

Esimmäinen isäntäkoneiden välinen viesti lähetettiin lokakuun 29. päivänä 1969 kello 22.30 (Kleinrock 2004). Kokeilu suoritettiin siten, että molemmalla koneen käyttäjällä olivat puhelinkuulokkeet, jolla varmistettiin viestin saapuminen perille. Kleinrock ja yksi hänen ohjelmoijistaan kirjautui SRI isäntäkoneeseen UCLA isäntäkoneesta, ja tarkoituksena oli kirjoittaa sana login siten, että UCLA koneelta kirjoitettaisiin kirjaimet log ja SRI kone täydentäisi sanan kirjaimilla in. Ensimmäiset kaksi kirjainta, eli lo, näkyivät SRI isäntäkoneella, mutta kolmas kirjain kaatoi järjestelmän. Seuraava yritys kuitenkin onnistui, ja vuoden 1969 lopussa verkostoon lisättiin vielä kaksi yliopistoa, Santa Barbara ja Utah (ks. KUVA 1). (Kleinrock 2002.) Kaksi vuotta myöhemmin, 1971, ARPANETissä oli 23 isäntäkoneita linkittyneenä toisiinsa (ks. KUVA 2) (Griffiths 2002a).



KUVA 1. Kleinrockin luonnos neljän yliopiston muodostamasta verkostosta 1969



KUVA 2. ARPANET vuonna 1971

2.3 ARPANETistä Internetiin

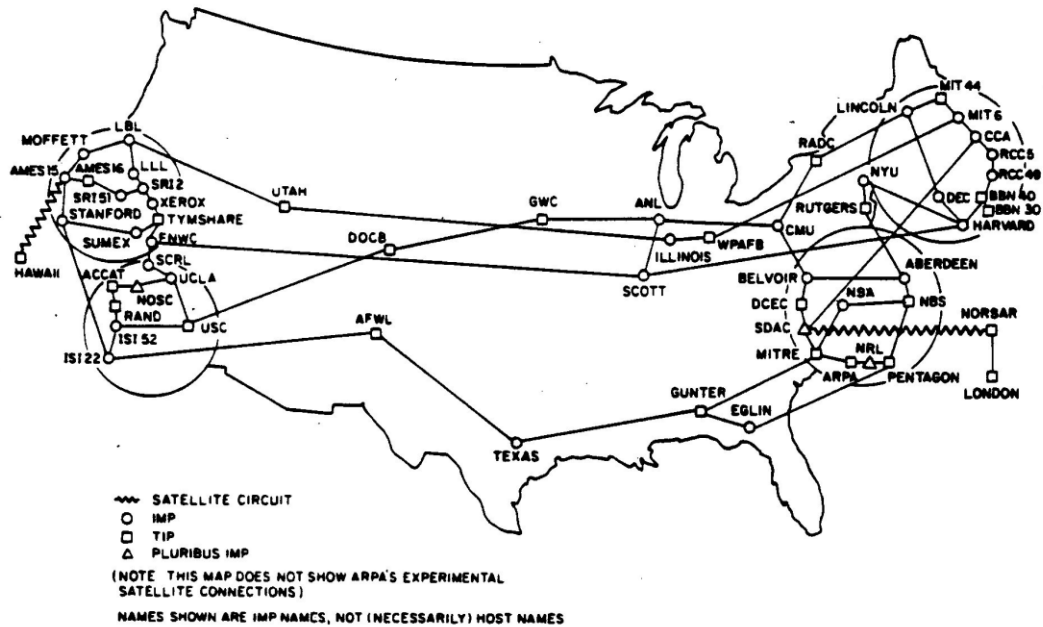
ARPANET julkistettiin yleiseen käyttöön lokakuussa 1972, ja järjestelmän toimintaa esitettiin linkittämällä 40 eri paikoissa olevaa tietokonetta yhteen. Pian aloitettiin muiden kuin ARPANET verkostojen kehittäminen. (Griffiths 2002a.) ARPANET- verkostoon lisätään ensimmäiset kansainväliset yhteydet, Lontoon yliopisto Englannissa ja Norjan kuninkaallinen tutkainstituutti, vuonna 1973 (Boswell 2010).

Vuonna 1974 Vinton Cerf ja Robert Kahn julkaisivat tutkimuksen, jossa kuvailtiin ensimmäistä kertaa TCP/IP protokollat, joiden tarkoituksena oli yhdistää erilaiset verkostot toisiinsa (Aboba 1993). Internet sanan käyttö terminä kuvatakseen yksittäistä maailmanlaajuista TCP/IP verkostoa, otettiin käyttöön puolta vuotta myöhemmin RFC 675 julkaisussa, jonka Vinton Cerf, Yogen Dalal ja Carl Sunshine kirjoittivat Stanfordin yliopistossa. Seuraavan 9 vuoden aikana protokollia hiottiin ja niitä toteutettiin useilla eri käyttöjärjestelmillä. (Wikipedia 2010a.)

2.4 Internet 1975-1988

1975 ensimmäinen kaupallinen pakettikytkentäinen verkosto, TELENET, Leonard Roberts'in toimesta.

Vuonna 1977 tehtiin ensimmäinen kolmen verkoston sisäinen yhteisliitännä linkittämällä SATNET, PRNET ja ARPANET yhteen reitillä, joka kulki Kalifornian Menlo Park nimisestä kaupungista University College London-yliopistoon Englantiin ja takaisin University of Southern California- yliopistoon Marina del Rey:hin Kaliforniaan (ks. KUVA 3).



KUVA 3. ARPANET 1977

Samaisena vuonna 1977 keksittiin PC modeemi, mikä mahdollisti yksityisten ihmisten pääsyn Internetiin kotikoneilta. Ensimmäinen menestynyt kotitietokone Apple II näki myös päivänvalon vuonna 1977. Ja Internetin suosio kasvoi.

1978 Ensimmäinen ilmoitustaulu Internetissä auttoi tietokoneen käyttäjiä keskustelemaan toistensa kanssa ja vaihtamaan tietoja ja ohjelmistoja modeemin välityksellä.

1978 Ensimmäinen SPAM.

1979 kehitettiin Usenet. Ilmoitustaulu- konseptin laajentaminen kehittää Usenet- verkoston modeemien käyttäjistä. Ilmoitukset järjestettiin hierarkkisesti kategorioihin, oli kyseessä sitten uutisryhmä, tieteentien julkaisu tai yleinen keskustelu.

1980 englantilainen fyysikko Tim Berners-Lee aloittaa projektin nimeltä Enquire eurooppalaiselle hiukkasvoiman tutkimuskeskukselle Cern:lle. Enquire käytti hypertekstiä, eli linkkejä, ja sen tarkoituksena oli tiedottaa Cernin omia työntekijöitä projekteista, ohjelmistoista ja muista työntekijöistä. Enquire oli pohjana World Wide Webille.

1980 CSNET kasvatti Internetin suosiota liittämällä yliopistoja, jotka eivät kuuluneet ARPANETiin, verkkoon.

1984 Domain

1985 .com, .edu, .gov, .mil, .org

1986 protokollien sota: IP vs. OSI, isäntäkoneita 6 000

1987 isäntäkoneita 30 000

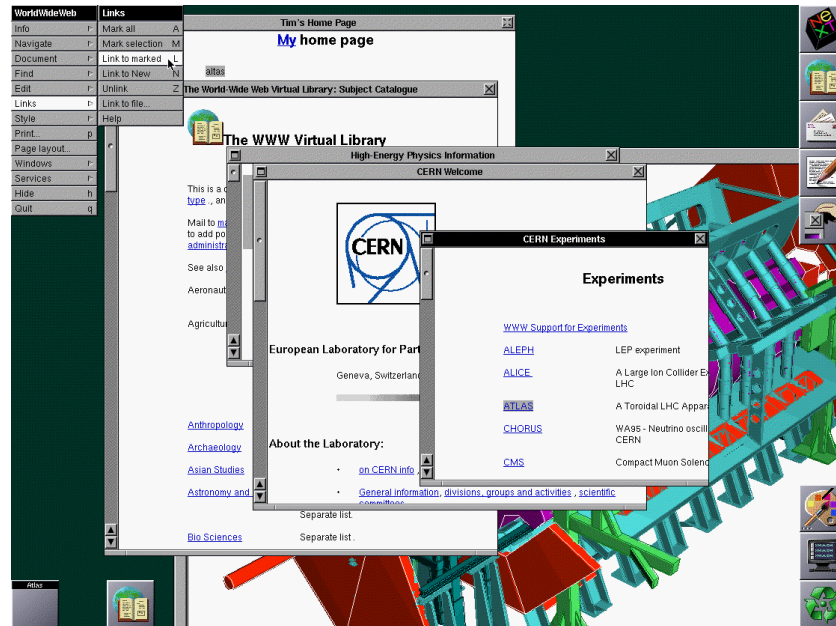
1988 Internet mato, IRC, NSFnetin päivitys → Internet Suomeen

2.5 Word Wide Web

1989 Tim Berners-Lee kirjoitti ehdotuksen työnantajalleen CERN:lle, jossa Berners-Lee oli huolissaan tiedon häviämisestä jatkuvasti päivittyvissä järjestelmissä. Vastauksena ongelmaan olisi idea linkitetyistä tietojärjestelmistä, jolloin tietokannoista tulisi näkyviin hypertekstin näkymä erilliseen porttiohjelmaan (nykyisin tunnetaan nimellä selain), josta käyttäjä löytäisi haluamansa tiedon. (Berners-Lee 1989.)

Marraskuun 12. päivänä 1990 Berners-Lee julkaisi yhdessä työkaverinsa Robert Cailliau'n kanssa kattavamman ehdotuksen rakentaakseen hypertekstiprojektin nimeltä WorldWideWeb (Berners-Lee & Cailliau 1990). Joulukuun 1990 mennessä Berners-Lee (2004) oli rakentanut tarvittavat työkalut toimivaa verkkoa varten: ensimmäisen verkkoselaimen WorldWideWeb (ks. KUVA 4), joka oli samalla interneteditori, ensimmäisen verkkopalvelimen, joka oli Berners-Leen oma tietokone sekä ensimmäisen verkkosivuston osoitteessa <http://nxoc01.cern.ch/hypertext/WWW/TheProject.html>, mikä kuvaili itse projektia. Kyseinen sivusto lakkautettiin samana vuonna, mutta ko-

pio siitä löytyy osoitteesta <http://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html>. (Berners-Lee 2007.)



KUVA 4. WordWideWeb- selain vuodelta 1993. Alkuperäinen selain vuodelta 1990 oli muuten samanlainen paitsi kokonaan harmaan sävyinen, ja logot kuten CERN, tai kirja näkyvät toisessa ikkunassa, sillä selain ei hyväksynyt aluksi tekstiin liitettäviä kuvia (Berners-Lee 2004).

3 HAKUKONE

3.1 Hakukoneiden historia

3.1.1 Ennen World Wide Webiä

World Wide Webin perustajan Tim Berners-Leen mukaan

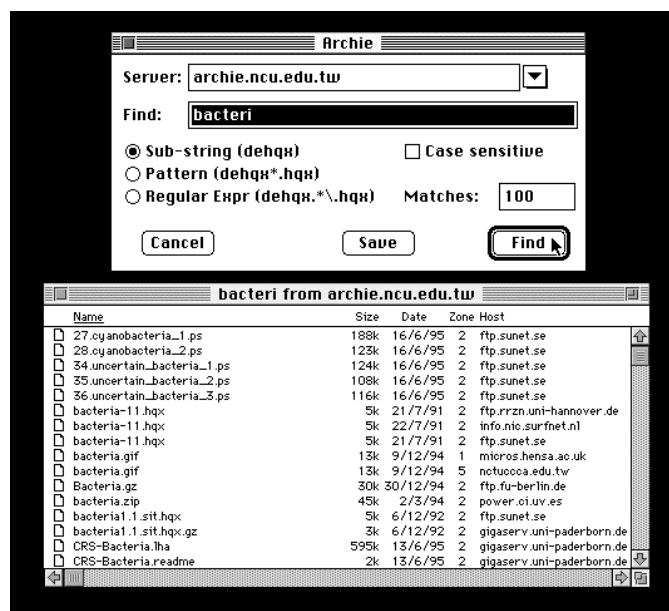
”Internet on verkostojen verkosto. Periaatteessa se on tehty tietokoneista ja kaapeleista... Verkko on informaation abstrakti, kuvitteellinen paikka. Internetistä löytää tietokoneita – Internetistä löytää tiedostoja, ääniä, videoita,... tietoa. Internetissä yhteydet ovat tietokoneiden välillä kulkevia kaapeleita, verkossa yhteydet ovat hypertekstilinkkejä. Verkko on olemassa, koska on olemassa ohjelmia, jotka kommunikoivat tietokoneiden välillä Internetissä. Verkkoa ei voi olla olemassa ilman Internetiä. Verkko teki

Internetistä hyödyllisen, sillä ihmiset ovat hyvin kiinnostuneita informaatiosta ja he eivät oikeastaan tahdo tietää tietokoneista ja kaapeleista. (Griffiths 2002a.)”

Internetin ollessa niin laaja kaikkine tietoineen, niin tuli tarve kehitellä keino etsiä haluttuja tietoja valtavasta verkostosta. Syntyi hakukone.

3.1.1.1 Archie

Ensimmäisen hakukoneen (mm. Järvinen 2003), Archien (ks. KUVA 5), kehitti Alan Emtage vuonna 1990 McGillin yliopistossa Montrealissa (Wikipedia 2010b). Archien tarkoituksena oli listata anonymien ftp- arkistosivujen sisältöä (Underwood 2004). Ohjelma haki öisin noin kerran kuukaudessa arkistoista päivitystietoja, ja kopioi tiedot hakutietokantaan (Deutsch 1993). Haku toteutettiin käyttämällä Unix grep komentoa (Wikipedia 2010b). Esimerkiksi komento `grep pine woods.txt`, näytti käyttäjälle kaikki rivit, joissa ilmeni sana pine tiedostossa woods.txt. Hakusanaa ei ollut rajoitettu, joten sanat kuten pineapple tai pines oli tuloksissa mukana. (Wikipedia 2010c.) Käyttäjälle palautettiin lista, mistä osoitteista haettua tiedostoa kannatti lähteä etsimään (Järvinen 2003). Palvelimille, joissa haettu tieto löytyy, päästiin monella eri tavalla; käyttäen paikallista clienttia (archie, anarchie), Telnetin avulla suoraan palvelimeen, tilaamalla kyseisen tiedoston sähköpostiin, ja myöhemmin World Wide Web käyttöliittymien avulla. (Wikipedia 2010b.)

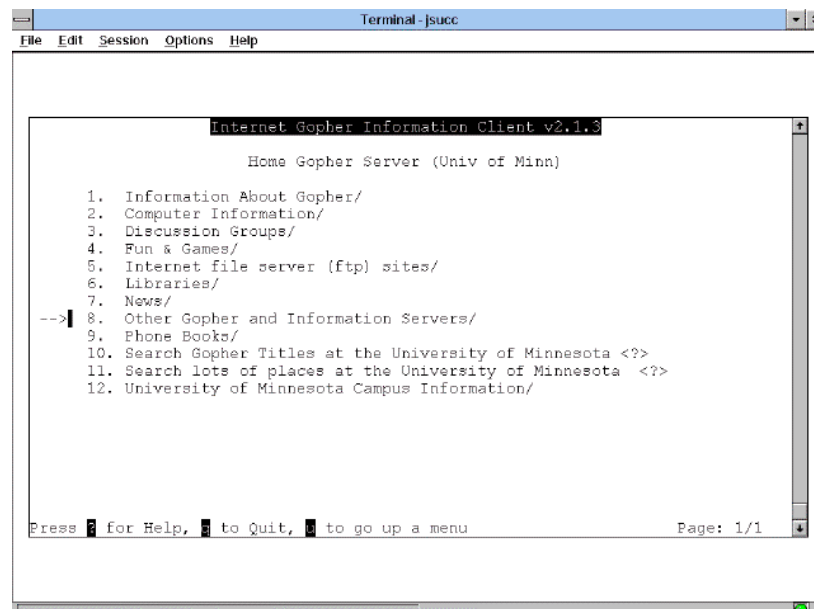


KUVA 5. Archie- hakukysely 1990

WWW teki tiedon hakemisesta paljon helpompaa ja enää ei ole kovin monta Archie-palvelinta maailmassa. Yhteen palvelimeen pääsee ainakin osoitteesta http://archie.icm.edu.pl/archie-adv_eng.html.

3.1.1.2 Gopher

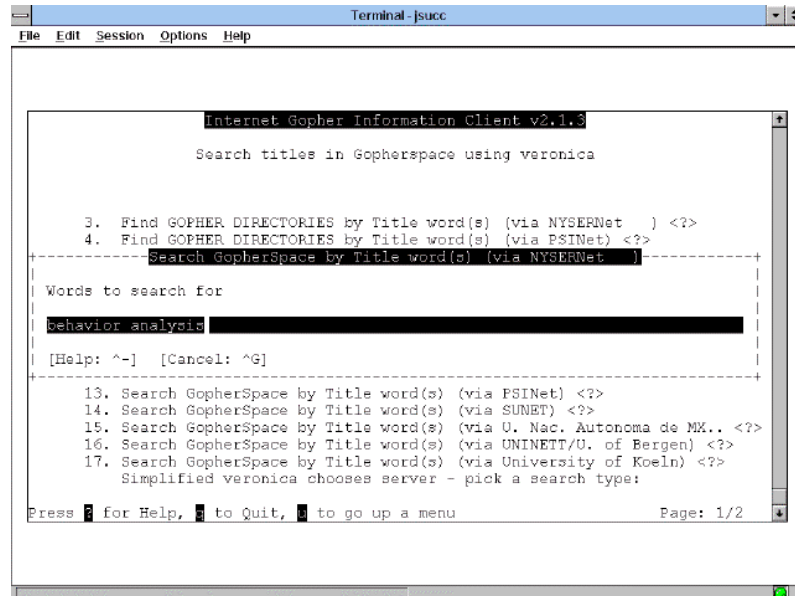
Archie etsi kuitenkin pelkkiä tiedostoja koneilta (Griffiths 2002b), ja niinpä käyttäjät eivät olleet varmoja mitä itse tiedostot pitivät sisällään. Vuonna 1991 Mark McCahill kehitti Minnesotan yliopistossa Gopher- hakuohjelman (ks. KUVA 6), mikä haki tiedostoissa olevan tekstin, ja listasi sen käyttäjälle (Underwood 2004). Gopher antoi listan linkeistä, johon pääsi enteriä painamalla suoraan lukemaan tiedostoa, eikä Archien käyttötavan mukaan tarvinnut enää etsiä tiedostoa itse sivustosta. Samalla pystyi kyseisen tekstin lähettämään itselleen sähköpostilla. Gopherista tulikin hyvin suosittu, juuri tämän tallennustavan takia. Nykyaikaisen Gopher palvelimen löytää täältä <http://gopher.quux.org:70/>. (Griffiths 2002b.)



KUVA 6. Gopher- etusivu 1991

Gopherin jalanjäljillä kulki ohjelma nimeltä Veronica (Very Easy Rodent-Oriented Network Index to Computerised Archives), joka haki tietoja suoraan Gopherin indeksoimista järjestelmistä (ks. KUVA 7). Veronica tarjosi tietokannan useimmista

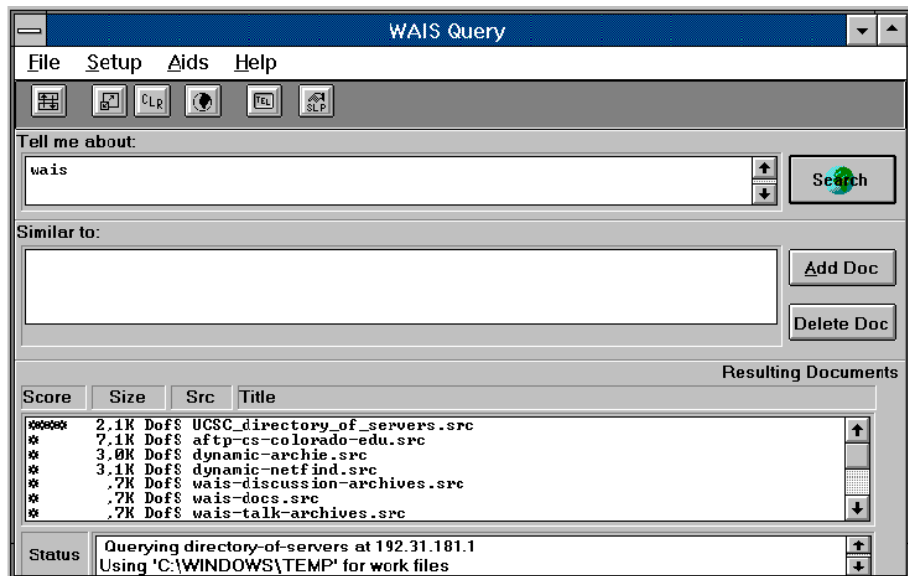
Gopherin valikoihin sisältyvistä hakusanoista.(Underwood 2004.) Veronicassa onnistui myös eri hakemistoista hakeminen (Griffiths 2002b).



KUVA 7. Hakusanan laittaminen Veronicassa

3.1.1.3 WAIS

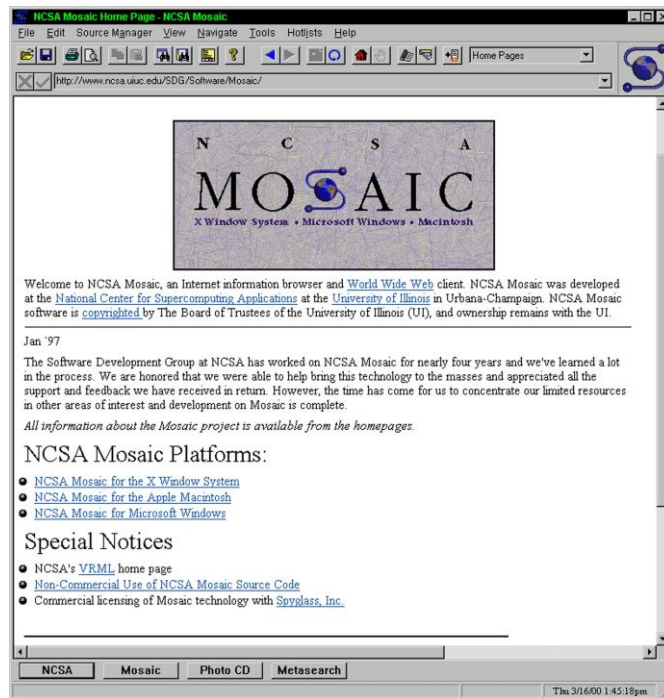
Wide Area Information Servers, eli WAIS, kehitettiin myös vuonna 1991 Thinking Machines Corporation toimesta (Wikipedia 2010d). WAIS on Internet-järjestelmä, jossa erikseen määritetyt aihehakemistot on perustettu monissa palvelinosoiteissa. Näitä palvelimia hallinnoi palvelinhakemisto, Directory of Servers, ja johon käyttäjät pääsivät käsiksi hakemaan tietoa WAIS client ohjelmilla. (TechTarget 2005.) WAIS hajotti tietokantansa erillisiin aiheisiin, ja käyttäjä pystyi näin rajaamaan hakuaan tiettyyn aihehakemistoon. WAIS järjesti hakutulokset sitä mukaan, miten useasti hakusana ilmeni tietokannoissa (ks. KUVA 8). (Griffiths 2002b.) Käyttäjälle näytettiin lista kaikista niistä palvelimista, joissa hakusanan sisältänyt tietokanta löytyi (TechTarget 2005).



KUVA 8. WAIS hakukysely 1991

3.1.2 Mosaic- selain

Vuonna 1993 ilmestyi NCSA:lta, National Center for Supercomputing Applications, Internet-selain nimeltä Mosaic, joka sai paljon tunnustusta graafisen ja helppokäyttöisen käyttöliittymänsä takia (Wikipedia 2010e). Mosaic oli ensimmäinen selain, joka käytti ikoneita, kirjanmerkkejä ja näytti kuvia tekstin seassa (ks. KUVA 9). Pian yleisö löysikin tämän puoleensavetävän sovelluksen, joka helpotti niin sanotun tavallisen ihmisen Internetin käyttöä; yli 5000 kopiota ladattiin joka kuukausi, satoja tuhansia kyselyjä ohjelmasta tuli viikoittain ja vuotta myöhemmin Mosaicilla oli jo usean miljoonan käyttäjäryhmä ympäri maailmaa.



KUVA 9. Mosaic- selain 1993

Mosaicin kehittäminen ja tukeminen lopetettiin vuonna 1997, mutta sen voi vieläkin ladata NCSAn ftp-palvelimelta osoitteesta <ftp://ftp.ncsa.uiuc.edu/Web/Mosaic/>. (NCSA 2009.) Mosaic- selainta voi myös kokeilla osoitteessa <http://www.dejavu.org/1993win.htm>, Nykyiset Internet-selaimet kuten Mozilla ja Internet Explorer käyttävät samanlaisia ominaisuuksia ja toimintoja kuten Mosaic aikoi- naan, joten Mosaicin selaimella on ollut suuri vaikutus Internetin käytön hurjaan suosioon (Wikipedia 2010e).

World Wide Webin kehitys vuonna 1991 ja Mosaic- selaimen julkaisu 1993 mahdol- listavat Internetin valtavan laajentumisen, mutta ne mahdollistavat myös uusien ja helposti lähestyttävien hakukoneiden synnyn (Griffiths 2002b).

3.1.3 Hakukoneet World Wide Webin jälkeen

Hakukoneiden tiedonhakutekniikkaa paranneltiin merkittävästi 1990-luvulla, sillä Internetin yleistyessä sivujen määrä kasvoi eksponentiaalisesti, ja tiedon nopea hake- minen laajasta tietoverkostosta vaikeutui. 1993 Matthew Gray suunnitteli maailman ensimmäisen hakurobotin nimeltä World Wide Web Wanderer. Kyseessä oli ohjelma, joka tutki alun perin verkon kasvua laskemalla verkkopalvelimien määrän, mutta

myöhemmin keräsi talteen myös URL- merkkijonoja. Syntyi Wandex - verkon ensimmäinen tietokanta. (Underwood 2004.) Hakurobotit yleistyivät hakukoneiden käytössä 1990 luvulla, ja niiden toiminnasta kerron tarkemmin luvussa 3.2.1.

3.1.3.1 ALIWEB

Martijn Kosterin kehittämää ALIWEB- hakukonetta vuonna 1993 pidetään ensimmäisenä hakukoneena verkossa. ALIWEB, eli Archie- Like Indexing in the Web, oli rakenteeltaan Archie- hakukoneen tyylinen; se keräsi verkkopalvelimien indeksejä säännöllisesti kooten ne yhdeksi hakutietokannaksi. ALIWEBin toiminta perustui palvelimien ylläpitäjien tekemiin indeksitiedostoihin (ks. KUVA 10), joissa he kertovat palvelimella sijaitsevien tiedostojen sisällöstä. Kuvassa 10 on esimerkkinä osa NEXOR yrityksen verkkopalvelimen indeksitiedostosta. Ylläpitäjät lähettivät sitten tiedon indeksitiedoston sijainnista ALIWEB- tietokantaan verkkolomakkeella. (Koster 1994, 5-7.) ALIWEB- hakukoneen suosio jäi kuitenkin vähäiseksi, sillä hakukone vaati palvelinylläpitäjän tekemään ja päivittämään indeksitiedoston itse, mitä kaikki eivät osanneet tehdä. Myös Kosterin oma mielenkiinto jäi vähäiseksi uusien hakurobotiteknikan yleistyessä vuonna 1994. (Sherman 2002.)

Template-Type:	ORGANIZATION
Organization-Name:	NEXOR
URI:	/nexor/nexor.html
Description:	NEXOR is an OSI software company.
Keywords:	NEXOR, OSI
Template-Type:	DOCUMENT
Title:	Perl
URI:	/perl/perl.html
Description:	Information on the Perl Programming Language. Includes a local Hypertext Perl Manual, and the latest FAQ in Hypertext.
Keywords:	perl, perl-faq, language

KUVA 10. ALIWEBin haussa tarvitsema palvelimen indeksitiedosto

3.1.3.2 Galaxy

Galaxy oli vuonna 1994 ilmestyessään ensimmäinen verkkohakuluettelo, joka järjesti linkit hakemistoihin. Hakemistoja pystyi selailemaan suoraan tai hakusanalla etsimään

tietoa usealla eri vaihtoehdolla, kuten pelkästään tiedostojen otsikoista, linkeistä tai tekstistä (Szuprowicz 1997, 98). Galaxy käytti haussa verkkosivujen lisäksi Gopher- ja Telnet- tietokantoja. Hakukone ei käyttänyt hakurobotteja indeksoidakseen verkkosivuja, vaan verkkosivujen lisääminen tapahtui täyttämällä verkkolomakkeen, johon sivun perustietojen lisäksi pystyi valitsemaan oikean hakemiston sivulleen. (Sonnenreich 1997.)

3.1.3.3 Yahoo!

Neljä kuukautta Galaxyn julkaisemisen jälkeen huhtikuussa 1994 ilmestyi Jerry and David's Guide to the World Wide Web – verkkosivu, joka oli aluksi David Filon ja Jerry Yangin ylläpitämä linkkikokoelma heidän suosimistaan verkkosivuista (Yahoo! 2005). Kun linkkien määrä alkoi kasvaa, ne piti lajitella hakemistoihin (ks. KUVA 11) ja edelleen alihakemistoihin, ja verkkosivusta tuli haettava aihehakemisto. Verkkosivu vaihtoi samalla nimekseen Yahoo!, ja se erosi kilpailijoistaan, Wanderer- ja Galaxy-hakukoneista, siinä, että jokainen linkki sisälsi käsintehdyn kuvauksen linkistä johtavasta verkkosivusta. (Wall 2008.)



KUVA 11. Yahoo! kotisivu 1994

Yahoo! on kuitenkin enemmän verkkohakemisto kuin hakukone. Sen oma tietokanta on pieni verrattuna suuriin hakukoneisiin, mutta tietokanta on hyvin lajiteltu ja linkit helposti valittavissa hakemistoista. (Hock 2001, 183.) Yahoo! ulkoistikin hakuko-

neensa vuoteen aina vuoteen 2004 asti, jolloin yhtiö ymmärsi oman hakukoneensa tärkeyden (Wall 2008). Esimerkkinä tästä ulkoistumisesta on se, että Yahoo! teki Googlen kanssa sopimuksen kesäkuussa 2000, mikä johti siihen, että Yahoo! käytti sivustossaan Google- hakukoneen tuottamia vastauksia Yahoo! sivuilla tehtyihin hakuihin (Google Press Center 2000). Yahoo! yritti kilpailla Googlen kanssa parhaan hakukoneen tittelistä ostamalla mm. AltaVistan ja AllTheWeb- hakukoneet ja kehittämällä omaa hakukonettaan aina 29.7.2009 asti, jolloin Microsoftin Bing- hakukone ja Yahoo! yhdistyivät kymmenen vuoden sopimuksella haastamaan Googlen hakukonemarkkinoilla (MTV3- uutinen 2009).

3.1.3.4 WebCrawler

Huhtikuussa 1994 Brian Pinkerton julkaisi WebCrawler- hakukoneensa (ks. KUVA 12) verkossa, ja se oli ensimmäinen hakukone, joka indeksoi koko tekstin verkkosivulta. WebCrawler- hakukoneella oli aluksi 4000 verkkosivun tietokanta. Puoli vuotta myöhemmin hakukoneella tehtiin jo miljoonas haku. Kesäkuussa 1995 America Online osti WebCrawlerin. (Pinkerton, 2004.) WebCrawler- hakukoneesta tuli näin erittäin suosittu hakukone, jolla oli viikossa kaksi miljoonaa käyttäjää, jotka tekivät kolme miljoonaa hakua päivittäin. WebCrawler indeksoi tietokantaansa käymällä verkkosivulla muutaman kerran päivässä. Hakukone ilmoitti hakutulokset sen mukaan kuinka monta kertaa hakusana ilmeni verkkosivulla ja päättelämällä, miten ainutlaatuinen yksittäinen hakusana oli sivustoon nähden. (Szuprowicz 1997, 84–85.) Seuraavina vuosina ilmestyvistä hakukoneista ainakin Excite, AltaVista, HotBot ja Northern Light kopioivat WebCrawler- hakukoneen kehittämää hakualgoritmia (SEO Treo 2010).

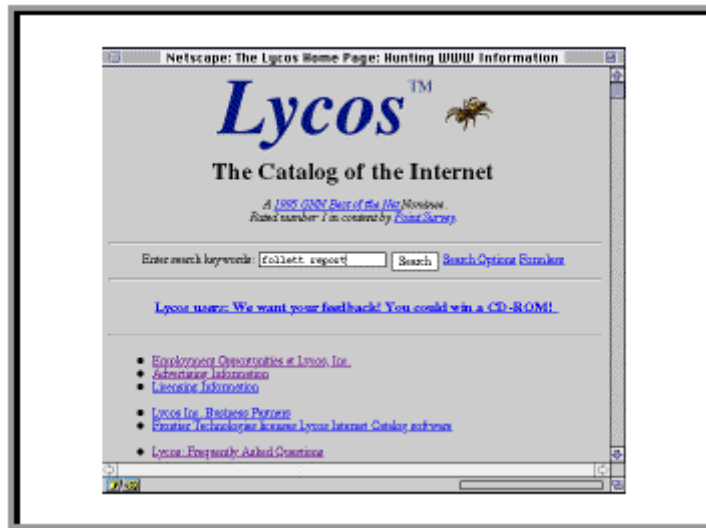


KUVA 12. WebCrawler kotisivu 1994

3.1.3.5 Lycos

Michael Mauldin (1997) julkaisi Lycos- hakukoneensa (ks. KUVA 13) heinäkuussa 1994, jonka tietokanta sisälsi ensimmäisenä päivänä 54 000 verkkosivua. Kuukautta myöhemmin tietokanta oli kasvanut 394 000 sivuun, puoli vuotta myöhemmin 1.5 miljoonaan ja vuoden 1996 loppuun mennessä tietokannassa oli jo yli 60 miljoonaa verkkosivua indeksoitu – enemmän kuin millään muulla hakukoneella.

Lycos- tietokanta indeksoi verkkosivulta sen nimen, otsikot, alaotsikot, 100 tärkeintä sanaa sivulta ja 20 ensimmäistä tekstiriviä. Indeksistä saa selville myös verkkosivun koon tavuissa ja sanojen määrän. (Szuprowicz 1997, 79.) Lycos tarjosi myös lyhyen kuvauksen hakutuloksissa linkin alapuolella, ja linkin vieressä oli prosenttiluku, joka kuvasi miten hyvin verkkosivun sisältö vastasi hakusanoja. Mauldin jätti Lycosin 1998, jonka jälkeen hakukone ”portalisoitui” keskittyen hakukoneena olemisen sijaan Chat- palveluun, henkilökohtaisiin kotisivuihin, horoskooppeihin ja muihin toimintoihin. (Sherman 2002.)



KUVA 13. Lycos kotisivu 1995

3.1.3.6 Excite

Excite aloitti toimintansa Architext- projektinimellä vuonna 1993. Kaksi vuotta myöhemmin projektilla oli rahoitusta 3 miljoonan dollarin edestä ja Excite- hakukone (ks. KUVA 14) julkaistiin lokakuussa 1995. (McIntyre 2007.) Excite- hakukone käyttää ICE, Intelligent Concept Extraction, - hakutekniikkaa indeksoimalla verkkosivuja ja päättämällä mitkä sanat liittyvät toisiin sanoihin. Uusien sanojen ilmaannuttua verkkoon, Excite opettelee sanojen välisiä suhteita jo olemassa oleviin sanoihin, ja auttaa näin tiedon hakemisessa. (Szuprowicz 1997, 74.)

Excite osti kilpailijansa WebCrawlerin 1997, saaden 44 prosentin markkinaosuuden hakukoneiden käyttäjistä. Excite- hakukoneen suosiota lisäsi myös sen sivistynyt yhteenvetotekniikka verkkosivuista tekemällä verkkosivuista tiivistelmiä tietokantaansa. (Szuprowicz 1997, 74–75.) Excite tarjoaa käyttäjilleen myös karttapalvelun, Keltaiset sivut, Uutiset ja MP3 hakupalvelun. Lisäksi käytössä on 12 erilaista kanavaa autoista urheiluun ja matkailuun sekä 14 eri kielen versiot hakukoneesta. (Hock 2001, 89–97.)

Ask (Jeeves)- hakukone osti Exciten vuonna 2005 (Richardson 2005).



KUVA 14. Excite kotisivu 1995

3.1.3.7 AltaVista

AltaVista (ks. KUVA 15) julkaistiin 16 miljoonan indeksoidun verkkosivun tietokannalla vuonna 1995. AltaVista oli alun perin projektityö, jolla esiteltiin Digital Equipment Corporation – yhtiön uutta ja nopeaa tietokonejärjestelmää, ja jolla pystyttiin tallentamaan jokaisen verkkosivun jokainen sana nopeasti haettavaksi indeksitietokannaksi. (The Web Search Workshop 2010.) AltaVista oli myös ensimmäinen hakukone, joka tarjosi monikielisen hakumahdollisuuden sekä erilaisia kehittyneitä hakumenetelmiä. Myöhemmin AltaVistaan tuli kuva-, ääni- ja videohaku sekä verkon ensimmäinen käännoistyökalu Babel Fish. (Hock 2001, 57.) Jo julkaisupäivänä yli 300 000 käyttäjää haki AltaVistalla tietoa, ja vuotta myöhemmin verkkosivuja oli indeksoitu lähelle 30 miljoonaa ja hakuja tehtiin päivittäin 19 miljoonaa kappaletta (Szuprowicz 1997, 70).

Vuonna 1999 AltaVistasta tuli verkkoportaali kilpaillakseen Yahoo!- hakukoneen kanssa markkinoista, ja keskittyi hakukoneen sijaan muihin ominaisuuksiin kuten ilmaiseen sähköpostiin sekä myymään tavaroita ja palveluita sivuillaan (Kopytoff 2000). Verkkoportaali ei kuitenkaan menestynyt, ja hakukonemarkkinoitakin hallitsi Google. Vuonna 2002 AltaVista lanseerattiin uudelleen keskittyen hakukoneena olemiseen (Glasner 2002). Se kannatti, sillä Yahoo! osti kilpailijansa vuoden päästä (Cullen 2003).



KUVA 15. AltaVista kotisivu

3.1.3.8 HotBot

HotBot- hakukone (ks. KUVA 16) ilmestyi Wired Magazine- lehden verkkopalveluna vuonna 1996, ja sen suosio perustui nopeaan hakurobottiin ”Slurp the Web Hound”, joka kävi läpi yli 10 miljoonaa verkkosivua päivittäin (Sonnenreich 1997). HotBot- hakukoneella oli siis päivitettyin hakutietokanta, joka sisälsi uusimmat verkkosivut ja nopeimmat päivitykset jo olemassa oleviin sivuihin. Slurp- robotti prosessoisi koko verkkosivun tekstin, teki siitä tiivistelmän ja lisäsi sen sopivaan tietokantaan. Robotti myös lisäsi tietokantaansa automaattisesti uudet verkkosivut, jos niihin oli viittauksia muista sivuista. (Szuprowicz 1997, 77–78.) HotBot järjesti hakutuloksensa muun muassa seuraavien tekijöiden avulla: hakusanan yleisyys itse verkkosivun tekstissä, hakusanojen yleisyys otsikoissa ja tekstin pituus (Hock 2001, 125).



KUVA 16. HotBot kotisivu 1996

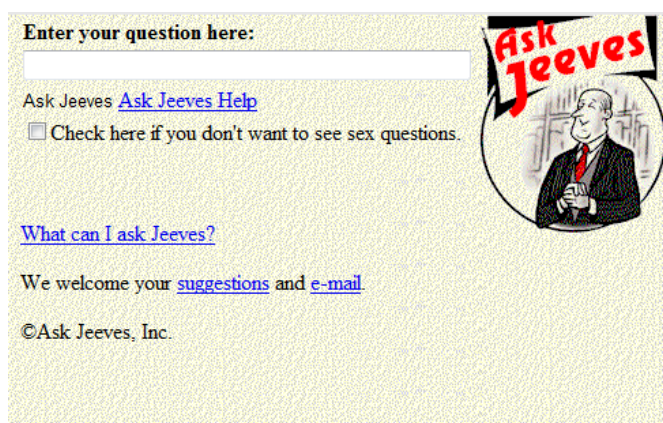
Lycos osti HotBot- hakukoneen 1998 (Lycos Acquires Wired Digital 1998) ja vuosien saatossa hakukoneen suosio romahtui kehittymättömyyteensä kilpailijoihin verrattuna. HotBot julkaistiin uudelleen 2002 metahakukoneena, joka siis etsii hakusanalla samanaikaisesti useasta eri hakukoneen tietokannasta. Hakusanan laitettuaan käyttäjä valitsi tietokannan, josta haku suoritetaan. Vuonna 2002 nämä hakukoneet olivat AllTheWeb, Inktomi, Teoma ja Google. (Search Journal 2003.) Vuonna 2010 vaihtoehtoina ovat Yahoo!, lyGO.com ja MSN(Bing).

3.1.3.9 Ask Jeeves

Ilmestyessään vuonna 1997 Ask Jeeves- hakukoneen (ks. KUVA 17) tarkoituksena oli se, että ihmiset pystyisivät löytämään vastauksen etsimäänsä kysymyksen muodossa ja omalla äidinkielellään. Muutamia yksittäisiä hakusanoja ei enää tarvittaisi, vaan vastauksen etsimäänsä saisi suoraan kysymällä sitä hakukoneelta. (Buresh 1998.) Hakukoneesta tuli tämän takia hyvin suosittu ja vuotta myöhemmin hakuja tehtiin jo 300 000 päivittäin ja vuonna 1999 hakujen määrä oli jo yli 2 miljoonaa päivässä (Funding Universe 2005).

Ask Jeeves käyttää ExpertRank- hakualgoritmiä, joka ei pelkästään perustu etsittävän tiedon sisältävän verkkosivun linkityksen suosioon, eli siihen kuinka paljon linkkejä muilta sivuilta löytyy verkkosivulle, josta tieto löytyy. Toiminta perustuu myös aihekohtaiseen suosioon. (How does Ask search work 2010.) Hakukone etsii indeksitietokannastaan haulle vastaavat sanat, järjestää ne oikeaan muotoon ja etsii haulle vastauksen suodattamalla merkityksettömät sanat pois. Hakukone ei näytä kymmentä sinistä linkkiä vastauksena hakuun, vaan se näyttää parhaan vastauksen hakijan kysymyksen heti sivun ylälaudassa. (Web Search 2010.)

Vuonna 2006 Ask Jeeves- hakukone muutti nimensä pelkäksi Ask- hakukoneeksi (BBC- News 2006).



KUVA 17. Ask Jeeves kotisivu 1997

3.1.3.10 Northern Light

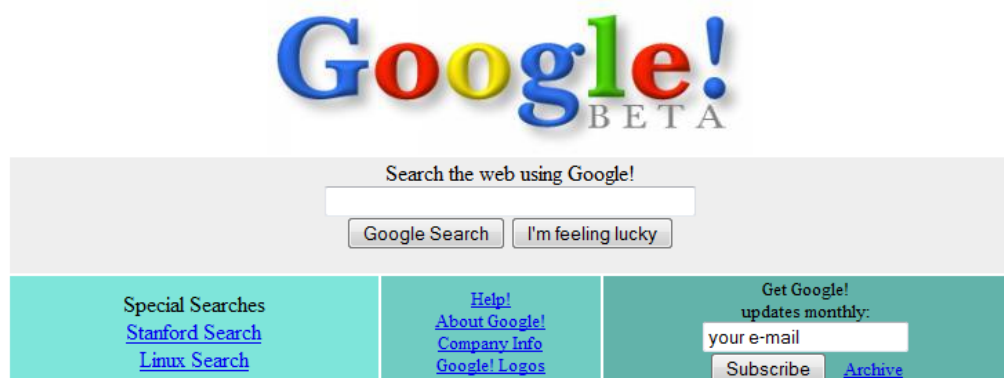
Northern Light- hakukone julkaistiin 1996, ja se tarjoaa vakavalle hakijalle apua tutkimustyöhönsä ja on tarkoitettu lähinnä yrityskäyttöön, ja se ei sisällä muita kuin valittujen yritysten julkaisuja (Northern Light Search 2010). Hakukoneen ainutlaatuisuus perustuu sen omaan laajaan julkaisutietokantaan kuin myös verkkohakuun, ja se myös järjestää hakutulokset erityisiin kansioihin. Northern Light tarjoaa oman julkaisukantansa artikkelien tiivistelmät ilmaiseksi ja maksua vastaan koko artikkelin tekstin. Maksamalla jäsenmaksun hakija saa käyttöönsä koko materiaalin, joka kattaa yli 7 000 yrityksen julkaisut muun muassa aikakauslehdet, kirjat, sanomalehdet, tiedotteet, uutiset ja tutkimukset. Northern Light tietokanta sisältää yli 360 miljoonaa verkkosivua ja yli 25 miljoonaa omaa artikkelia. Kun haku on suoritettu, hakukone järjes-

tää tulokset erityisiin kansioihin sitä mukaan, miten tulokset vastaavat toisiaan esimerkiksi aiheen ja tiedostotyyppin mukaan. (Hock 2001, 151–152; 172)

3.1.3.11 Google

Google (ks. KUVA 18) on ollut yksi menestyneimpiä, ellei menestynein, IT- alan yrityksiä viimeisen 15 vuoden aikana. On sanomattakin selvä, että Google on hallinnut hakukonemarkkinoita yli vuosikymmenen ajan, ja ihmiset ovat omaksuneet googlettamisen kuuluvan päivärutiineihin. Siitä kertovat huimat ideksointitiedot; 26 miljoonaa sivua aloittaessa, miljardin raja rikki 2000, ja 2008 sivuja oli indeksoitu huimat biljoona eli 1 000 000 000 000 (Alpert & Hajaj 2008). Googlen nopea suosio jo julkistamisvuonnaan 1998 perustui sen omaan PageRank- hakualgoritmiin, mikä antoi hakutermeille luotettavat tulokset nopeasti (Brandt 2002). Googlen PageRank toimii kuten äänestysjärjestelmä yksittäisille verkkosivuille. Kun verkkosivu on linkitetty toiseen, kumpikin sivu saa äänen. Google sitten laskee verkkosivun tärkeyden äänien lukumäärän ja laadun perusteella, mitkä toiset sivut ovat sille antaneet. (Rhodes 2010.)

John Battellen mukaan (2005) Google voitti hakukonemarkkinat AltaVistalta itselleen jo puolessa vuodessa sen takia, koska AltaVista portalisoitui ja unohti näin hakutoiminnon. Google saavutti paremmat hakutulokset PageRank- tekniikan avulla, ja Google on alusta asti pysynyt hakukoneena.



Copyright ©1998 Google Inc.

KUVA 18. Google kotisivu 1998

Microsoftin hakukoneen vaiheita tarkastellaan enemmän luvussa 4.1.

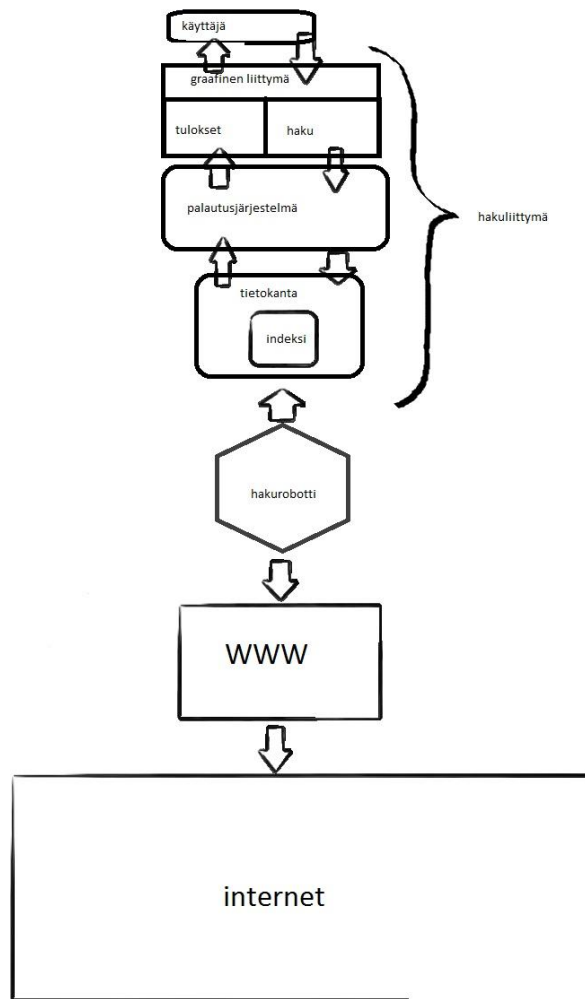
3.2 Hakukoneen toimintaperiaate

Internet- hakukoneet ovat verkkosivuja, jotka auttavat käyttäjiään löytämään muille sivuille tallennettua informaatiota. Eri hakukoneet etsivät tietoa eri lailla, mutta niillä kaikilla on samat periaatteet:

- ne etsivät tietoa Internetistä
- ne päivittävät luetteloaan (indeksi) löytämistään sanoista, ja siitä mistä sanat ovat löytyneet
- ne sallivat käyttäjän hakea etsimäänsä tästä luettelosta

Randolph Hock (2001) listaa viisi toiminnon osaa (ks. KUVA 19), joista nykyinen hakukone muodostuu:

- 1) **hakurobotti**, joka etsii verkkosivuja
- 2) **tietokanta**, sisältäen sekä hakurobottien että muiden lähteiden löytämät verkkosivut
- 3) **indeksointiohjelma**, mikä luetteloi tietokannan sisällön
- 4) **palautusjärjestelmä**, sisältäen algoritmin mikä järjestää hakutulokset sekä muut ohjelmat, jotka palauttavat hakusanoja vastaavat tiedot indeksistä
- 5) **graafinen liittymä**, HTML-sivu, joka kerää hakutiedot käyttäjältään palautusjärjestelmään, ja mikä näyttää lopulliset hakutulokset



KUVA 19. Hakukoneen toiminnon osat

Ymmärtääkseen miten nämä viisi eri toimintoa toimii keskenään, niin esittelen ne Google- hakukoneen toimintaperiaatteen avulla. Numeroidut kohdat viittaavat edellä mainittuihin toimintoihin eli 1) viittaa hakurobottiin, 2) tietokantaan ja niin edelleen.

Google tekee Internetistä omanlaisen version hakijoita varten, käyttämällä automaattisia ohjelmia nimeltä 1) Googlebotit. Ne etsivät Internetistä jatkuvasti muuttuvia tietoja, sellaisia sivuja, jotka päivittävät tietoaan usein. Esimerkiksi uutissivut skannataan muutaman minuutin välein, ja harvoin päivitettäviä sivustoja skannataan muutaman viikon välein. Googlebotit syöttävät dataa Googlen 2) keskusverkostoon. Data sisältää kaiken tarvittavan tiedon kyseisestä sivusta: URL-osoite, sivun koko tekstin, viitteet kuviin ja muihin sivun sisältämiin tiedostoihin, ja metadatan, jonka sivun omistaja sivusta tekee. Keskusverkostossa tiedot 3) indeksoidaan, eli muutetaan hakusanalueteloksi. Jokainen sana, jota voitaisiin käyttää hakusanana, on listattu Googlen keskus-

tietokoneisiin, jonka mukana on myös tieto miltä sivulta kyseisen sanan löytää. Jätti-läismäinen indeksiluettelo pilkotaan pienimmiksi osiksi, shardeiksi, jotka lähetetään 4) datakeskuksiin. (Layton ym. 2006.)

Googella on näitä datakeskuksia kymmeniä (Miller 2008) ympäri maailmaa. Datakeskus pitää sisällään suuria palvelinhuoneita, joissa on tuhansia koneita. Palvelimet ovat kahden yksikön, 2U, korkuisia ja ne vievät siis kahden hyllyn korkuisen tilan. Palvelinyksiköt on rakennettu tavallisten laivakonttien sisään, ja jokaisessa kontissa on 1160 palvelinta. (Mikrobitti 2009, 11.) Datakeskuksessa haku ajetaan indeksiluettelon läpi, ja hakua vastaavat termit lähetetään takaisin keskusverkkoon. Google päättää sitten ”salaisella reseptillään” mitkä verkkosivut ovat relevantteja hakutermejä kohtaan, ja näyttää käyttäjälle 5) selaimessa hakutulokset sisältäen verkkosivun tiivistelmän, eli snippetin. (Layton ym. 2006.)

Seuraavaksi kerron lisää hakukoneen jokaisesta toiminto-osasta.

3.2.1 Hakurobotti

Ennen kuin hakukone pystyy kertomaan käyttäjälleen mistä haettava tieto löytyy, sen on ensin löydettävä tiedon sisältävä verkkosivu tai tiedosto. Informaation löytämiseen valtavasta tiedon valtamerestä, hakukoneet käyttävät apunaan ohjelmia, jotka etsivät verkosta uusia sivuja lisätäkseen ne tietokantaansa sekä samalla päivittää tietokannassa jo olemassa olevia sivuja (Hock 2001, 6). Näitä ohjelmia kutsutaan hakuroboteiksi, englanniksi robots, spiders tai crawlers. Hakurobotin päätehtävänä on siis löytää verkkosivu palvelimelta, tehdä verkkosivusta tiivistelmä, ja tallentaa tiivistelmä hakukoneen tietokantaan. Toinen tehtävä on löytää muut sivut, joille pääsee jo löydettyltä sivulta linkkien avulla. (de Vrieze 2002, 10.)

Se, mitä hakurobotti verkkosivulta ottaa talteen, vaihtelee suuresti hakukoneesta riippuen. Toiset hakurobotit indeksoivat, eli luetteloivat, pelkästään sivun URL-osoitteen, otsikot tai alaotsikot, ja toiset taas koko verkkosivulla olevan tekstin. Esimerkiksi Googlebot indeksoi sivulta kaikki tärkeät sanat jättäen kuitenkin pois sanat ”a”, ”an” ja ”the”. AltaVista puolestaan indeksoi nämäkin sanat. Lycos indeksoi muun muassa otsikoiden sanat, 100 yleisintä sivulla käytettävää sanaa ja tekstin 20 ensimmäisen rivin sanat. (Frankilin 2000.)

Hakurobotti vierailee verkkosivulla pyytämällä palvelimelta sivun HTML- tiedostot (Szuprowicz 1997, 46). Se, miltä verkkosivu näyttää hakurobotin silmin, katso liitteet 2, 3 ja 4, joissa on kuvattu Mikkelin ammattikorkeakoulun Student- kotisivu osoitteessa student.mikkeli.amk.fi selaimen, lähdekoodin (HTML) sekä hakurobotin silmin.

Harold Davisin (2006) mukaan hakurobotille ei merkitse mitään verkkosivulla olevat kuvat. Ainoa tieto, mitä hakurobotti kuvista kerää, on kuvan HTML- kieleen alt- attribuuttiin lisätty kuvaus. Esimerkiksi HTML- koodi

``, lisää sivulle ”kukka” nimisen jpg- muodossa olevan kuvan, jossa on vaihtoehtoisena tekstinä sana ”ruusu”. Tämä vaihtoehtoinen teksti on tärkeä, sillä jos selain ei jostain syystä pysty lataamaan `kukka.jpg` -kuvaa, niin selain näyttää sivulla vaihtoehtoisen tekstin, eli tässä tapauksessa sanan ruusu. Jos kuvasta puuttuu tämä alt- attribuutti, niin kuvaa ei siinä tapauksessa indeksoida. Verkkosivulla olevat linkit tuottavat myös ongelmia hakuroboteille (mts. 54). Jos verkkosivulla on toimimaton linkki, eli mikä ei johda toiselle sivulle, niin tämä sivu jätetään indeksoimatta. Samoin sellaiset sivut, joihin tarvitaan kirjautua käyttäjätunnuksella ja salasalla, jäävät indeksoimatta.(Mt.)

Sivuja, joihin hakurobotti ei pääse käsiksi, kutsutaan ”syväksi verkoksi” (deep web). Näitä sivuja ovat muun muassa taloudelliset tiedot, lentoaikataulut, lääketieteelliset tutkimukset ja ne ovat verkkosivuja, jotka käyttävät omaa dynaamista tietokantaa. (Wright 2009.) Hakurobotti ei pääse käsiksi näihin sivuihin, koska ne eivät ole html- muodossa ja niihin ei ole linkkejä. Suomalaisia tällaisia sivustoja ovat esimerkiksi [eduskunnan](http://eduskunnan.fi) sivut sekä Valtion säädöstietopankki [Finlexin](http://finlexin.fi) sivut. (Järvinen 2003.) Bright Planetin (2000) tutkimusten mukaan tämä ”syvä verkko” on noin 500 kertaa suurempi kuin hakukoneiden löytämä ”pintaverkko”.

Hakurobotit käyvät useimmin suosituilla sivuilla, joilla käyttäjät käyvät yleisimmin ja jotka ovat eniten linkitettyjä, eli sivuille tulee toisilta sivuilta linkkejä. Hakurobotit voidaan ohjelmoida syvälliseen tai laajaan hakuun, tai molempiin. Syvällinen hakeminen ulottuu verkkosivun pääsivulta alisivuihin ja taas niiden alisivuihin ja niin edelleen asti. Laaja haku ottaa huomioon vain lähinnä verkkosivujen pääsivut, mutta ei välttämättä kaikkia alisivuja. (Hock 2001, 7.)

3.2.1.1 Robots.txt

Verkkosivuja on mahdollista myös suojata hakuroboteilta. Kotisivut voivat sisältää arkaluonteista ja yksityistä tietoa, kuten valokuvia, tai keskeneräisiä sivuja, mitä sivun omistaja ei halua hakukoneiden indeksoivan. Robots.txt on tekstitiedosto, mikä antaa sivun omistajalle mahdollisuuden hallita hakukoneiden käyttämiä hakurobotteja sivuihinsa nähden. The Robots Exclusion Protocol (REP), johon robots.txt toiminta perustuu, on hakurobottien suunnittelijoiden sekä verkkosivujen julkaisijoiden välinen kunnioitussopimus (Search Tools 2008).

Robotstxt.org- sivusto (2007) on omistettu näille robots.txt- tiedostoille, ja sivustolta saa seuraavanlaista tietoa niistä. Kun hakurobotti haluaa käydä verkkosivulla esimerkiksi osoitteessa <http://www.mikkeli.amk.fi/opiskeluvaihtoehdot>, niin hakurobotti tarkistaa ensin robots.txt- tiedoston osoitteesta <http://www.mikkeli.amk.fi/robots.txt>. Näin hakurobotti saa tietää, onko sillä lupa indeksoida kyseiseltä sivustolta löytyvää tietoa. Robots.txt on siis pelkkä tekstitiedosto, mikä laitetaan sivuston juurihakemistoon, ja se on myös julkinen, eli jokainen voi selvittää mitä tietoja kyseinen sivu haluaa hakukoneelta estettävän laittamalla /robots.txt sivun URL- osoitteen loppuun yllä mainitulla tavalla. Robots.txt- tiedosto sisältää yksinkertaisuudessaan kaksi riviä tekstiä:

```
User-agent: *
Disallow: /
```

User-agent- syntaksi määrittelee hakurobotin, mikä torjutaan sivulla. Disallow- syntaksi määrittelee alueen hakemistopuussa, mikä on kielletty hakurobotilta. Yllä olevassa esimerkissä User-agent; * kuvaa sitä, että kyseeseen kuuluvat kaikki hakurobotit, ja Disallow: / kieltää hakurobotin käymästä millään sivulla. Kun halutaan kaikkien hakurobottien indeksoivan kaiken, niin robots.txt- tiedostoon laitetaan:

```
User-agent: *
Disallow:
```


Saman asian ajaa myös tyhjä robots.txt- tiedosto, tai jos sitä ei ole ollenkaan juurihakemistossa. Kun puolestaan halutaan hakurobottien jättävän tietyt sivustot rauhaan, laitetaan robots.txt- tiedostoon esimerkiksi:

```
User-agent: *  
Disallow: /kuvat/  
Disallow: /tärkeät/
```

Kun halutaan estää yksi hakurobotti, esimerkiksi Googlebot, niin robots.txt- tiedostoon laitetaan:

```
User-agent: Googlebot  
Disallow: /
```

Ja jos puolestaan halutaan vain yhden hakurobotin pääsy sivuille:

```
User-agent: Googlebot  
Disallow:
```

```
User-agent: *  
Disallow: /
```

Robots.txt- tiedosto rajoittaa poissulkemisen siten, että yhteen Disallow:- syntaksiin ei pysty laittamaan kuin yhden sivun tai tiedoston, mikä halutaan estettävän. Esimerkiksi Disallow: /kuvat/ /tärkeät/ /reseptit.html/ ei ole mahdollinen, vaan se pitää tehdä näin:

```
Disallow: /kuvat/  
Disallow: /tärkeät/  
Disallow: /reseptit.html/
```

Palatakseen vielä <http://www.mikkeli.amk.fi> sivulle, niin alapuolella on kyseisen sivun robots.txt- tiedoston sisältö:

```
User-agent: *  
Disallow: /admin/  
Disallow: /feedback/
```

Tästä voi päätellä sen, että sivulla sallitaan kaikki hakurobotit sekä kansioita admin ja feedback ei indeksoida hakukoneen tietokantaan.

3.2.2 Tietokanta

Hakukoneen tietokanta muodostuu kaikista yksittäisiltä verkkosivuilta tallennetuista tiedoista. Tietokanta sisältää niin hakurobottien löytämät sivut kuin sivut, jotka ovat tulleet tietokantaan muiden lähteiden ja tekniikoiden kautta. Suuri osa tietokantaan lisätyistä verkkosivuista tulee itse verkkosivun julkaisijan kautta. (Hock 2001, 7.) Hakukoneiden omilta sivuilta löytyy yleensä linkki, jonka kautta voi lisätä verkkosivunsa hakukoneen tietokantaan. Googlella tällainen sivu löytyy osoitteesta www.google.fi/addurl. Hakukoneen käyttäjän on syytä muistaa sellainen fakta, että hakukone ei etsi hakutermejä suoraan verkosta, vaan omasta tietokannastaan. Joten vaikka hakija tietää missä kyseinen tieto sijaitseekin verkossa, niin se ei välttämättä ole hakukoneen tiedossa sen tietokannassa.

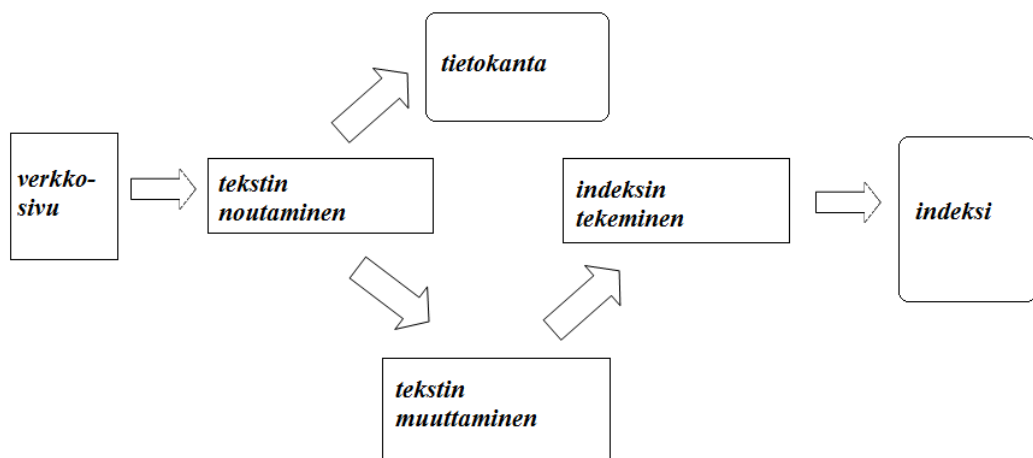
Kaikki hakukoneet eivät kuitenkaan tee omia tietokantoja, vaan ne käyttävät muiden tietokantoja hakutuloksissa lisäten niihin omia ominaisuuksia kuten lajittelualgoritmeja jne.

3.2.3 Indeksointiohjelma

Hakukoneen indeksointiohjelma tutkii hakukoneen tietokannassa olevan informaation, ja luo sopivat hakusanat indeksiin. Kun hakija suorittaa hakutoiminnon, niin se on tämä indeksi, mitä käytetään tunnistamaan vastaavat tulokset käytettyihin hakutermeihin. (Hock 2001, 8.) Hakukoneet indeksoivat eri tavalla informaatiota painottaakseen löytämiään tietoja tuloksissa. Tämän takia hakutulokset vaihtelevat samalla hakusalla eri hakukoneissa.

Indeksin tarkoituksena on löytää hakutermiä vastaavat tulokset nopeasti. Indeksien voi rakentaa muutamalla eri tavalla, mutta tehokkain keino on käyttää ns. hajautustaulu (hash table) - tekniikkaa. Tämän tekniikan avulla jokaiseen sanaan liitetään numeerinen arvo, käyttäen tiettyä kaavaa. Esimerkiksi suomen kielessä on kirjaimia, jolla useampi sana alkaa, ja puolestaan kirjaimia mitä käytetään harvemmin sanojen ensimmä-

mäisenä kirjaimena. Esimerkiksi kirjaimella ”N” alkaa enemmän sanoja kuin kirjaimella ”W”. Hakukoneen indeksoinnissa tämä tarkoittaa sitä, että sana mikä alkaa yleisemmällä kirjaimella kuin toinen, vie enemmän aikaa, jotta se löydetään indeksitietokannasta. Hajauttaminen vähentää tätä eroavaisuutta, ja vähentää hakutermiä vastaavan tiedon löytämiseen käytettävää aikaa. Hajautustaulu erottaa myös indeksin varsinaisesta hakusanasta. Hajautustaulu sisältää hajautusnumeron sekä osoittimen varsinaiseen tietoon, joka voidaan sitten tallentaa millä tahansa tehokkaalla tavalla. Tehokas indeksointi (ks. kuva) sekä tiedon varastointi edesauttavat tulosten nopean löytämisen, vaikka hakija tekisikin monimutkaisia hakuja. (Franklin 2000.)



KUVA 20. Indeksointiprosessi

3.2.4 Palautusjärjestelmä

Palautusjärjestelmä on ohjelma, joka vastaanottaa hakukyselyn ja etsii hakutermitte sopivat vastineet indeksistä palautusalgoritmin avulla ja lajittelee tulokset paremmuusjärjestykseen lajittelualgoritmia käyttäen (Hock 2001, 9-10; 25–27). Hakukoneiden välisen kilpailun takia nämä palautus- ja lajittelualgoritmit ovat tarkoin varjeltuja salaisuuksia, joten tietoja niistä ei juurikaan julkaista suurelle yleisölle. Hockin (2001, 27-29) mukaan muun muassa seuraavat asiat kuitenkin painottavat sanan tärkeyttä indeksointiprosessin aikana, ja kukin hakukone saattaa käyttää useampia tai vain muutamaa näistä:

- **sivun suosio**, eli sivulle johtaneiden sekä sivulta lähtevien linkkien määrä kertoo sivun olevan ihmisten suosiossa, ja sivulla olevien sanojen merkitys kasvaa.
- **sanan yleisyys**, mitä enemmän sama sana sisältyy sivun tekstiin, sitä enemmän sille annetaan painoarvoa.
- **hakusanojen sopivuus toisiinsa nähden**, eli jos hakukyselyssä on 4 hakusanaa ja ne kaikki löytyvät samasta hakutuloksesta, niin sille annetaan suurempi arvo lajittelussa.
- **sanojen harvinaisuus**, jos hakusanoina ovat 2 sanaa joista toinen on yleisempi sana hakukoneen tietokannassa ja toinen taas harvinaisempi, niin harvinaisen sanan sisältämä tulos saa enemmän tärkeispisteitä.
- **sana otsikossa**, antaa sanalle enemmän arvoa kuin sama sana myöhemmin itse tekstissä
- **sanojen läheisyys**, eli jos kaksi hakusanaa on peräkkäin tuloksessa, niin se saa lisäarvoa tulosten järjestämisessä.
- **päivämäärä**, eli tuoreimmat, päivitetty, tulokset saavat lisäpisteitä.

3.2.5 Graafinen liittymä

Graafinen liittymä on hakukoneen HTML- kielellä tehty liittymä, ts. hakukoneen kotisivu, jonka päätehtävänä on ottaa vastaa käyttäjän hakukysely, lähettää se hakukoneen prosessoitavaksi ja näyttää tulokset käyttäjälleen. Graafisen liittymän tarkoituksena on myös tarjota mainostajille mainospaikkoja, näyttää linkittämällä hakukoneen portaaliominaisuuksia (karttapalvelu, kääntäjä, uutiset yms.) sekä antaa informaatiota palveluntarjoajasta sen omilla tietosivuilla kuten Help- tai FAQ- osio.

3.3 Hakukoneen käyttötarkoitus

Tilastokeskuksen (2009) tutkimuksen mukaan suomalaisista 82 % 16–74- vuotiaista käytti Internetiä viimeisen kolmen kuukauden aikana. Internetin käyttäjistä puolestaan 82 % käytti Internetiä päivittäin tai lähes joka päivä. Suurin käyttäjäryhmä on 25- 34- vuotiaat 91 %:lla. Internet löytyy 78 prosentista suomalaisista kodoista. Sama tutkimus paljastaa myös suomalaisten Internetin käyttötarkoitukset. Internetin palveluista käytetyin on sähköposti, jota suomalaisista käyttää 91 prosenttia. Toiseksi suosituin on verkkopankit 87 % ja kolmantena hakukoneiden käyttö tavaroiden ja palveluiden et-

simiseen 86 %. Suomalaisista 68 prosenttia käyttää hakukonetta myös sairauksien, ravitsemuksen ja terveyteen liittyvissä kysymyksissä.

Suomalaisten suosituin hakukone on ollut jo vuosia Google, jonka käyttöprosentti huhtikuussa 2010 oli huimat 98,61 % (StatCounter... 2010a). Suomalaisten suosituimmat hakusanat vuonna 2009 olivat Googlen hakutilastojen (2010) mukaan facebook, youtube ja iltalehti (ks. KUVA 21).



KUVA 21. Suomalaisten suosituimmat Google- hakusanat 2009

3.4 Hakukoneen hyödyt ja haitat

Hakukoneen hyöty on siinä, että käyttäjä pääsee etsimään informaatiota helposti valtavasta tietoverkosta antamalla vain helpottavia avainsanoja hakukoneelle. Haittana voisi pitää sitä, että hakutulokset vaihtelevat eri hakukoneilla niiden eriävien algoritmien takia. Ympäristö myös kuormittuu hakukoneyhtiöiden valtavien datakeskusten takia ja lisääntyneiden Internet käyttäjien myötä, jotka käyttävät muun muassa paljon energiaa toimintaansa. Kirjastot ja kirjakaupat tekevät tappiota, kun kirjat digitalisoidaan.

3.5 Hakukoneiden tulevaisuus

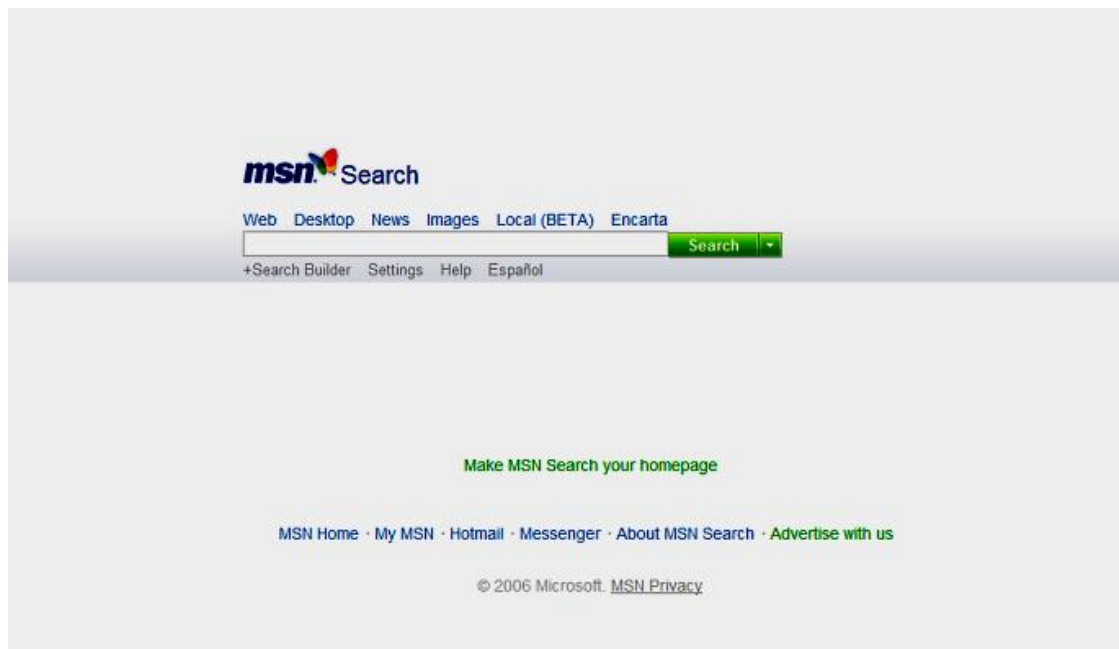
Google tulee olemaan vahvoilla ikuisuuden, ja pienemmät hakukoneet häviävät. Uusia toivottavasti tulee jatkossa vielä lisää Bingiä seuraten.

4 BING - PÄÄTÖKSENTEKOKONE

Tässä teoriaosion luvussa perehdytään Microsoftin uuteen hakukoneeseen nimeltä Bing. Bing on vielä suomenkielisenä versiona beta- testausvaiheessa, joten kaikki hienot Bingin tarjoamat ominaisuudet eivät ole täällä vielä käytössä. Siksi käytän opinnäytetyössäni Bing- hakukoneen Yhdysvallat- englanti- täysversiota, missä nämä ominaisuudet ilmenevät hyvin ja hakusanoina englanninkielisiä, sillä Bing- hakukoneen indeksi on suurempi englanninkielisenä kuin suomenkielellä.

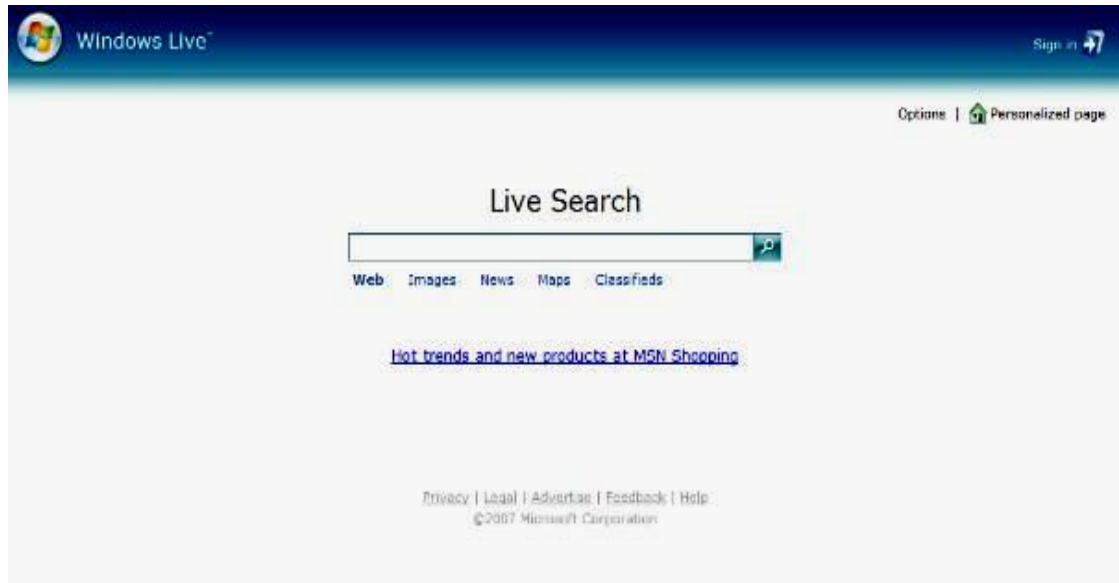
4.1 Historia

Microsoft julkaisi MSN Search (ks. KUVA 22)- palvelun vuonna 1998 käyttämällä Inktomin hakutietokantaa. Seuraavana vuonna MSN Search alkoi myös käyttää Looksmart- palveluntarjoajan tietokantaa haussa. Googlen menestymisen takia Microsoft alkoi kehittää omaa hakukonejärjestelmäänsä. Microsoftin oma MSN Search- hakupalvelu julkaistiin 2005 alussa, ja se alkoi myydä tietokantaansa muiden portaalien käyttöön kilpaillakseen tosissaan hakukonemarkkinoilla.



KUVA 22 .MSN Search kotisivu 2006

Windows Live Search (ks. KUVA 23) julkistettiin syyskuussa 2006 MSN Search tilalle. Uudessa hakukoneessa oli välilehdet haun rajaamisen helpottamiseksi kuten uutiset, kuvat ja musiikki. Kuvien etsimiseen oli kehitetty oma hakualgoritmi.



KUVA 23. Windows Live Search kotisivu 2007

2007 Microsoft ilmoitti, että se erottaa hakukoneen kehittelyn Windows Live- tuoteperheestä, nimeämällä hakukoneensa Live Search:ksi (ks. KUVA 24). Seuraavan kahden vuoden aikana Live Search tuotetta muutettiin monella eri tavalla muun muassa sulkemalla Live Search Books ja Live Search Academic- hakupalvelut, ja liittämällä kirjat ja akateemiset julkaisut saataviksi normaalin haun kautta. Muitakin samanlaisia palveluita lakkautettiin ja Microsoft sai näin kasvatettua yhtenäisen indeksitietokannan. Microsoftille tuottia harmia hakukoneensa Live- sana sillä uudistusten jälkeen Live tuoteperhettä ei ole olemassakaan, joten ettei nimi jäisi sitten pelkäksi Search:ksi, niin Microsoft vaihtoi Live Search hakukoneen nimeksi 3.6.2009 Bing- nimiseksi.



KUVA 24. Live Search kotisivu 2008

4.2 Toiminta-ajatus

Microsoft suunnitteli Bing- hakukoneen auttaakseen käyttäjiään käsittelemään hakujen ylikuormitusta, ja antaakseen nopeampia ja enemmän informoituja päätöksiä verkko-kohtaan yhteydessä. Microsoftin (2008) tekemän tutkimuksen mukaan 66 % ihmisistä kääntyy hakukoneen puoleen päättääkseen jostain asiasta. Microsoft ei ollut tyytyväinen omaan hakutapaansa, joten he kehittivät Bingistä päätöksentekokoneen tarjotakseen käyttäjälleen älykkäitä hakutyökaluja yksinkertaistaakseen hänen tehtäviään ja antaakseen käyttäjälleen enemmän informoituja vaihtoehtoja, oli kyseessä sitten nopeimmasta reitistä löytämisestä kotiin, tuotteen ostopäätökseen perustuvasta tutkimustyöstä tai matkan suunnittelusta.

Microsoftin tavoitteena on kaventaa hakijan ja tiedon välinen kuilu. Tarkoituksena on poistaa onni ja sattuman kauppa haun yhteydestä, jotta hakija löytäisi etsimänsä ensimmäisellä kerralla helposti ja vaivattomasti. Microsoftin tutkimuksen mukaan 35 % hakutuloksista ei tyydytä hakijaa, ja 24 % hylätään. Samaisen tutkimuksen mukaan 46 prosentilla menee yli puoli tuntia haun teossa pelkästään hakusanojen määrittelemiseen, ennen vastauksen löytymistä. Lähes puolet näistä kyselyistä palaa edelliseen kyselyyn esim. kirjoitusvirheen takia ja puolet tarvitsee lisää tarkennusta. Tutkimukseen osallistujien mukaan hyvin organisoitu hakukokemus auttaisi heitä löytämään

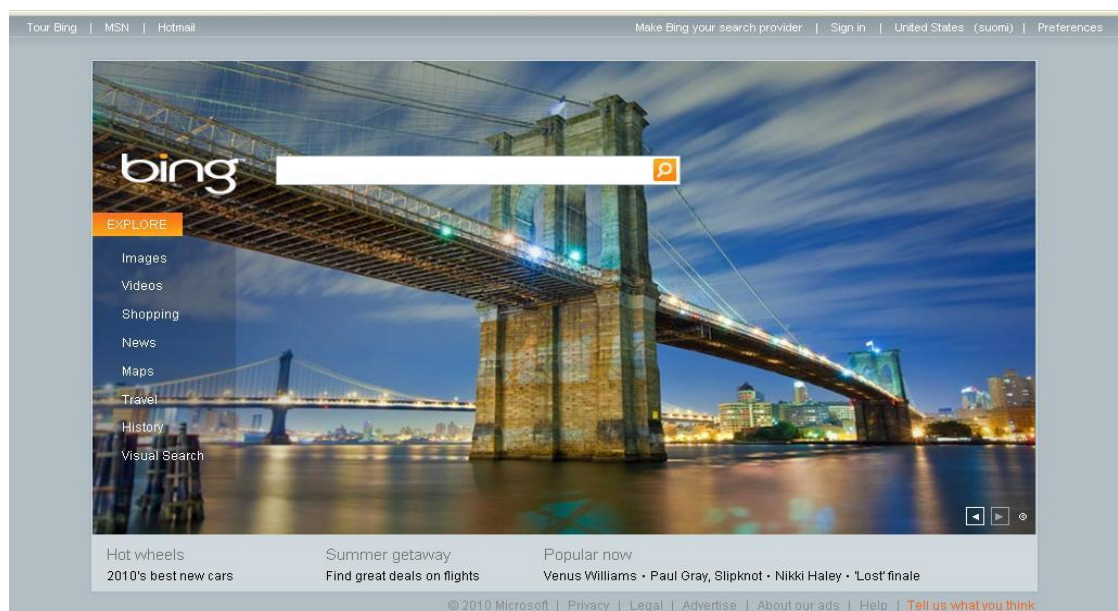
tiedon kaksi kertaa paremmin ja he saisivat toiminnot tehtyä nopeammin. Nykyaikaisen hakukoneen tulisi yksinkertaistaa toimintojaan ja tarjota sisäpiirin tietoa monista päätöksistä, joita käyttäjien täytyy tehdä. 75 % ihmisistä sanovat käyttävänsä hakukonetta varmistaakseen tuotteen ostopäätöksen, 62 % etsii paikallisen yrityksen tai kohteen, 45 % valitsee lennon tai hotellin, ja 43 % käyttää hakukonetta terveydellisiin tutkimustöihin. Tavalliset hakukoneet eivät ole suunniteltu tällaisiin toimintoihin, joten Bingillä on hyvä tilaisuus vallata markkinat tältä osa-alueelta.

Microsoft on paneutunut näihin ongelma alueisiin ja he ovat kehittäneet uuden päätöksentekokoneen. Bing tarjoaa mahtavat yleishakutulokset, merkityksellisyyden, suuren indeksitietokannan sekä nopeuden, eli kaikki tunnusmerkit hyvälle hakukoneelle tulee täytettyä.

4.3 Ominaisuudet

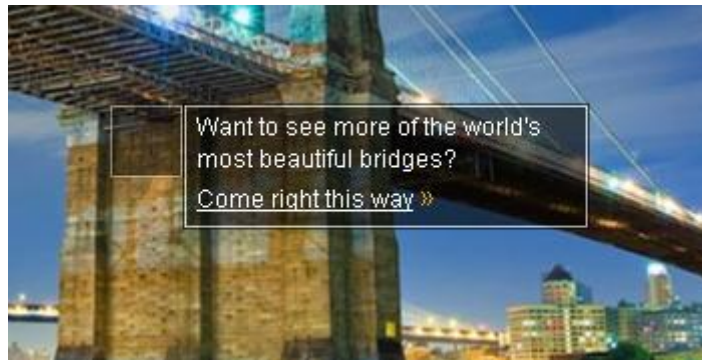
4.3.1 Etusivu

Bing.com kotisivulla (ks. KUVA 25) käyttäjän katse kohdistuu sivun taustakuvaan, joka vaihtuu päivittäin. Seuraavaksi huomataan kolikko valkea hakupalkki, mikä välillä pilaa taustakuvan sijainnillaan. Huomio kiinnittyy sen jälkeen vasemmalla olevaan Explore-palkkiin, missä haun voi rajata kuviin, videoihin, ostoksiin, uutisiin, karttoihin, matkailuun, historiaan tai visuaaliseen hakuun.



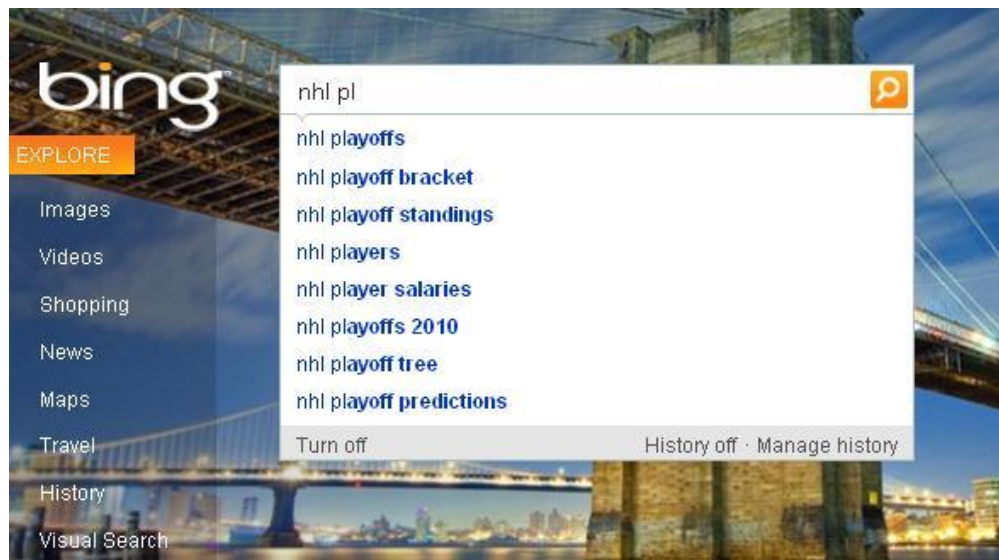
KUVA 25. Bing kotisivu

Tarkemmin kuin alkaa tutkia taustakuvaa hiirtä liikuttamalla edestakaisin sen päällä, niin taustakuvasta ponnahtaa esiin ns. hotspotteja (ks. KUVA 26), pieniä linkkejä, jotka liittyvät jotenkin päivän taustakuvaan antaen lisäinformaatiota.



KUVA 26. Hotspot

Haun aloittaessa huomaa, että Bing käyttää autosuggestiota hakukentässä (ks. KUVA 27), eli Bing antaa automaattisesti vastaavia hakuehdotuksia sitä mukaan kun kirjoitat kenttään.

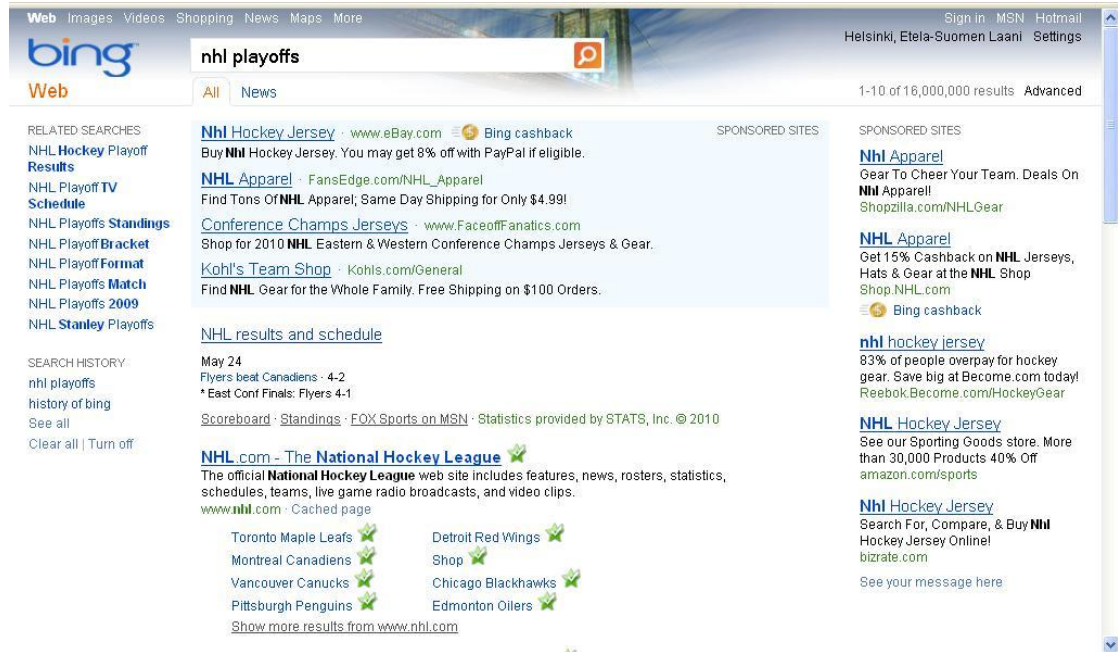


KUVA 27. Hakukentän autosuggestio

4.3.2 Hakutulossivu

Bing- hakukoneen tulossivu (ks. KUVA 28) näyttää aluksi olevan täynnä tekstiä, mutta tarkempi tutkimus osoittaa sen, että suurin osa tekstistä on hyödyllistä. Vasemmassa

laidassa, niin sanotussa Explore-paneelissa, on Related search- laatikko, mikä näyttää hakutermejä vastaavia vaihtoehtoisia sivuja. Esimerkiksi hakutermeille ”nhl playoffs” Bing tarjoaa linkit viimeisimpien otteluiden tuloksiin, TV- aikatauluun, sarjataulukon sekä viime vuoden playofftien tulokset. Tämän laatikon alapuolella on käyttäjän oma hakuhistoria kätevästi esillä, niin ei tarvitse kirjoittaa enää hakutermejä uudelleen jos pitää palata aikaisempiin hakukyselyihin. Tämän toiminnon saa samasta kohdasta nolattua tai poistettua kokonaan.



KUVA 28. Bing hakutulossivu

Hakukentän alapuolella ja sivun oikeassa laidassa on sponsorien linkkejä liittyen tässä tapauksessa NHL-tavaroiden myyjiin. Hakukentän alapuolella ennen linkkejä on kaksi välilehteä, joista voi valita koskeeko haku koko verkon aluetta vai sitten hakutermiä vastaavia uutisia. Tämä toinen (uutiset) välilehti vaihtelee hakutermin mukaan, ja se voi olla muun muassa matkailu, ostokset tai Wikipedia, ja niitä voi olla useampikin rinnakkain. Varsinaisten linkkien kohdalla on oikeassa laidassa kelluva palkki, ja jonka päälle mentäessä hiirellä aukeaa oikealle puolelle pieni ikkuna, missä kerrotaan lisää linkitetystä sivusta (ks. KUVA 29). Tässä tapauksessa käyttäjän on myös mahdollista hakea NHL.com-sivustolta suoraan, ettei tarvitse ensin mennä sen kotisivulle ja sitten vasta hakea sivun hakukentästä. Tämä nopeuttaa käyttäjän hakutoimintaa, jos linkistä ei pääse suoraan halutulle sivulle ja pitää vielä etsiä sivustolta haluamansa tiedon. Microsoftin tutkimuksen (2008) mukaan 24 % ihmisistä palaa takaisin hakutu-

loksiin, kun he huomaavat, että ehdotettu ja klikattu linkki ei annakaan vastausta. Näin on pystytty eliminoimaan turhat takaisin klikkailut.

Shipping for Only \$4.99!

General
Low-Priced **NHL** Gear. Shop Now.

on MSN · Statistics provided by STATS, Inc. © 2010

NHL League 🌟
This site includes features, news, rosters, statistics, adcasts, and video clips.

Detroit Red Wings 🌟
Shop 🌟
Chicago Blackhawks 🌟
Edmonton Oilers 🌟
nhl.com

hkin falls again... 🌟 11 hours ago
PN.com ... 1. The hits just keep on coming for Washington

nhl.com/5214763/five-things-learned-weekend · Cached

Yahoo! Sports 🌟
10 **Playoffs** news, results, features and more

Commerce Champions merchandise.
www.Esports.com

A MONSTER BOOST!

MORE ON THIS PAGE
Cammy's fortunate bounce The puck hits the end-boards, deflects to Cammalleri, and the red-hot forward makes no mistake.

Lapierre crushes Parent Max Lapierre catches Ryan Parent behind the net and introduces him to the boards...hard.

Gionta shows off his hands Brian Gionta takes a broken play and turns it into a 4-0 Montreal

More info ...

POPULAR LINKS
Colorado Avalanche
Gamecenter
Boston Bruins
Standings
Players

Search within nhl.com

[Go to the page](#)

KUVA 29. Pikasilmäys linkitettyyn sivuun

Bing tarjoaa myös niin sanottuja suoria vastauksia kysymyksiin. Tällaisia hakuja ovat perinteiset maantieteelliset kysymykset, kuten ”the capital of Norway”, valuutanmuutokset, ”50 euros in australian dollars (ks. KUVA 30)”, biologiset kysymykset, ”largest bird”, lentoaikataulut, ”lennonnumero”, ja niin edelleen. Vastaukset näkyvät tulossivun ensimmäisinä tuloksina korostettuna. Bing ottaa hauissa huomioon myös käyttäjän maantieteellisen sijainnin, joten jos haluat tietää mitä elokuvia oman kaupungin elokuvateattereissa (esimerkissä käyttäjä on newyorkilainen) on menossa, niin hakusanaksi riittää pelkkä ”movies” sen sijaan, että hakisit sanoilla ”New York movies”.

The screenshot shows a Bing search results page for the query "50 euros in australian dollars". The search bar at the top contains the query. Below the search bar, the results are categorized into "ALL RESULTS" and "Sponsored sites".

ALL RESULTS (1-10 of 16,900,000 results):

- Forex.com™** - www.Forex.com (Sponsored sites)
 - Free \$50,000 Practice Account With Real Time Charts, News & Research.
 - Convert Euros to Australian Dollars
 - 50 EUR = 72.39358 AUD
 - Source: MSN Money
 - Convert Euros to Australian Dollars (with star icon)
 - Sterling to Australian Dollar, Euro: Euro to US Dollar, Euro to Pound Sterling ... 50 = \$ 76.0400 € 100 = \$ 152.0800 € 500 = \$ 760.4000 € 1000 = \$ 1520.8000
 - www.advfn.com/currency-converter/euro-to-australian-dollar.html - Cached page
 - Exchange Rates Graph (Australian Dollar, Euro) (with star icon)
 - Currency Exchange Rate Graph for Australian Dollar vs Euro
 - www.x-rates.com/INR/AUD/graph1?fi.html - Cached page

Sponsored sites:

- 50 Dollars**
 - Explore 10,000+ Gift Baskets. Save On 50 Dollars!
 - Shopzilla.com
- Dollars To Euros**
 - Find Dollars To Euros at Great Prices
 - www.Pronto.com
 - See your message here

RELATED SEARCHES:

- Convert 50 Euros to Dollars
- Currency Converter
- Eurodollar Conversion
- Pounds in Dollars
- Euro Dollar Exchange Rate
- 100 Dollar in Euro
- 10 Euro in Dollar
- 500 Euros in Dollars

SEARCH HISTORY:

- 50 euros in australian

KUVA 30. Valuutanmuunnoksen suora vastaus

Bing auttaa käyttäjiään myös luokittelemalla hakutulokset ryhmiin (Web groups). Tietenkään ihan kaikkia hakukyselyjen tuloksia ei pysty luokittelemaan itse hakutermien takia, mutta jos tuloksia on paljon, niin silloin luokittelu on järkevää ja mahdollista. Luokitellut tulokset säästävät aikaa, kun tuloksia on paljon. Luokittelu näkyy hakijalle, kun hakutermiä vastaava luokittelu löytyy hakukoneesta. Näitä ovat esimerkiksi auto, matkustelu, ihmiset, urheilu ja terveyst. Esimerkiksi julkisuuden henkilön hakutulokset luokitellaan muun muassa seuraavien otsikoiden alle, elämäkertaa, valokuvat, videot ja blogit (ks. KUVA 31). Puolestaan tietty kaupunki hakusanana luokitellaan sään, hotellien, ravintoloiden, ja niin edelleen, alle.

[See more results](#)

Learn more about **Jessica Alba** »



Category: American television and film actress

Date of birth: April 28, 1981

Profession: Actor

Awards won: MTV Movie Award for Sexiest Performance

Appears in: Awake, Bill, Fantastic Four, Honey, The Eye, ...

TV Series: Dark Angel, Flipper

Source: [Freebase](#) (with star icon)

Blogs for **Jessica Alba** »

Jessica Alba & Family Have Some Memorial Day Weekend ... (with star icon)

Jessica Alba and husband, Cash Warren, took daughter, Honor, out for the Memorial Day weekend. A town car came to pick up the...

by Taylor Blue 3 hours ago · [Knocked Up Celebs](#)

Dress by Number: Jessica Alba's White Dress and Gladiator... (with star icon)

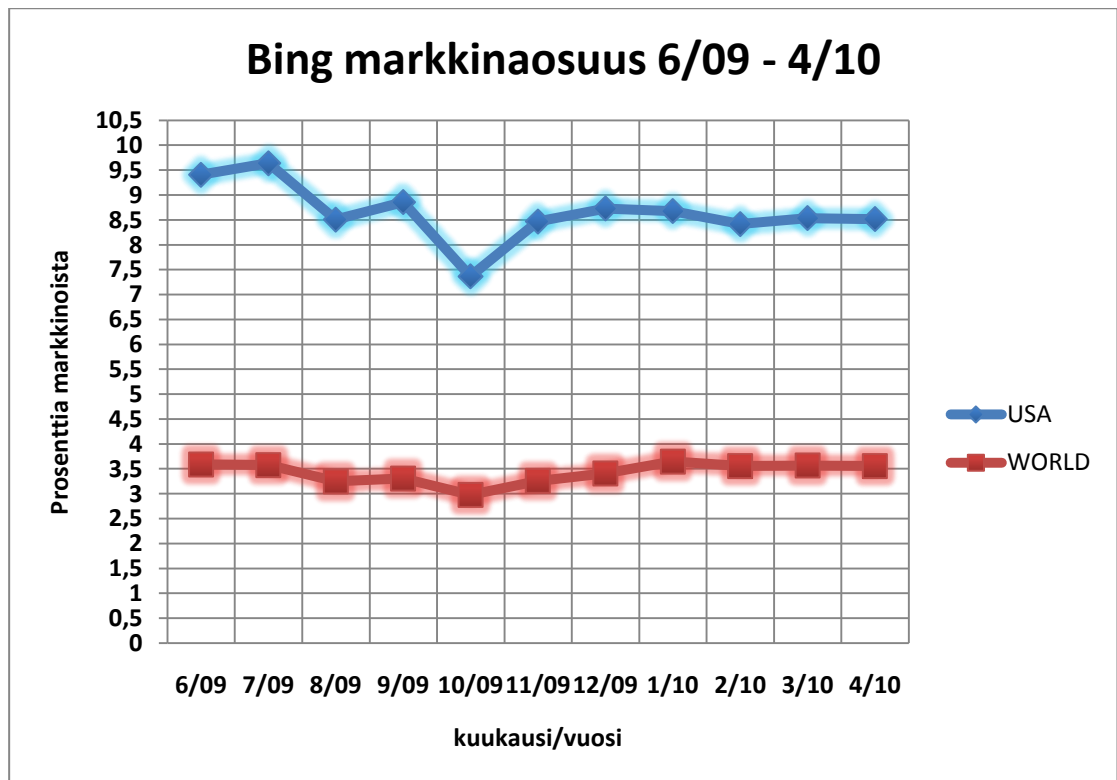
· by The Budget Babe 11 days ago

KUVA 31. Luokitellut hakutulokset

Kuvien ja videoiden selaaminen on tehty Bingillä helpoksi. Ensinnäkin hakutuloksia ei näytetä monella eri sivulla, vaan käytössä on yksi sivu, jota voi vierittää loputtomiin asti. Enää ei tarvitse odottaa seuraavan sivun latautumista. Lisäksi haun rajaaminen onnistuu kuvissa esimerkiksi koon, värin, tyylin tai jopa kasvojen(näkykö kuvassa pelkät kasvot vai kasvot ja hartiat) mukaan. Videon voi rajata pituuden, kuvasuhteen, resoluution tai lähteen mukaan. Videoita pystyy myös esikatselamaan 30 sekuntia siirtämällä hiiren kursorin videon päälle. Tämä on hyvä ominaisuus etenkin lyhyiden videoiden osalta, sillä videon voi katsoa jo suoraan hakutuloksissa menemättä itse isäntäosoitteeseen. Tällaisia videoita ovat esimerkiksi jääkiekon tai jalkapallon maali-koosteet, uutiset ja muun muassa hauskat kotivideo- tyyppiset videopätkät. Tämä ominaisuus sai hakukoneen julkaisun aikana paljon kritiikkiä, sillä pornoelokuvat näkyivät koneilla, missä oli suodattimet tätä vastaan (Linnake 2009). Reilun kahden viikon päästä julkaisusta Microsoft ilmoitti siirtävänsä kaiken pornograafisen materiaalin toiselle palvelimelle osoitteeseen explicit.bing.net. Näin Bingin on helpompi järjestellä pornograafiset materiaalit yhteen paikkaan, ja näin erotella ne tavallisesta hausta. Yritysten on myös helpompi rajata koneilleen hakutulokset, kun voivat jättää ”pornopalvelimen” pois niistä. (Gaudin 2009.)

4.4 Markkinaosuus

Hakumarkkinoiden lähteenä käytän StatCounter (2010b)- yrityksen tilastoja. Microsoft Live- hakukoneen osuus hakukonemarkkinoissa, ennen kesäkuun Bing julkaisua, oli maailman laajuisesti 1,87 % ja Yhdysvalloissa 4,95 %. Hakukoneista suurin on tietenkin vielä Google, joka tuolloin piti maailman laajuisesti paalupaikkaa lähes 89,8 % ja Yhdysvalloissa 78,5 % osuudella. Bing aloitti vahvasti Yhdysvalloissa, sillä 4.6.2009 vain 3 päivää julkaisun jälkeen, Bing ohitti Yahoo!- hakukoneen tullen toiseksi suurimmaksi hakukoneeksi Yhdysvalloissa 15,64 % osuudellaan. Tämä kuitenkin kesti vain muutaman päivän ja siitä lähtien sekä Bing (ks. KUVIO 1) että Yahoo! ovat kumpikin pysytelleet 7-10 prosentissa.



KUVIO 1. Bingin suosio 6/09-4/10 sekä Yhdysvalloissa että koko maailmassa

4.5 Tulevaisuus

Omasta mielestäni Bing tulee saamaan yhä enemmän käyttäjiä sivuilleen, kunhan se ensin saa suuremman indeksitietokannan ja kaikki ominaisuudet käytettäväksi muissakin maissa. Mutta ainakin Yhdysvalloissa uskon Bingin markkinaosuuden kasvavan jopa 20 % vuoteen 2015 mennessä.

5 YHTEENVETO

Internet on täynnä informaatiota kaikesta mahdollisesta aina koirien vaipoista pomminteko-ohjeisiin, ja näitä joillekin hyödyllisten tietojen löytämiseksi auttavat verkon lukuisat hakukoneet. Hakukone pyytää antamaan muutaman hakusanan, jolla se lähtee etsimään vastauksia indeksitietokannastaan. Monimutkaisten algoritmiensa avulla hakukone näyttää hakusanoja vastaavat tulokset paremmuusjärjestyksessä hakijalle. Hakijalle jää vastuuksi valita linkki, mistä luulee löytävänsä etsimänsä tiedon. Siinä puolestaan auttaa hakukoneen muodostama tiivistelmä linkin alapuolella, mistä saa pienen kuvan missä yhteydessä hakutermit kyseisellä sivulla on.

Microsoft on tuonut uuden, mielenkiintoisen hakukoneen maailmaan, jossa Google on ollut synonyyminä hakukoneelle. Bing tarjoaa mahtavia hakutuloksia, järjestyksellemmän käyttökokemuksen sekä nopeamman keinon päätöksentekoon. Bingillä on monta hienoa ominaisuutta, jotka auttavat käyttäjää tiedon haussa ja helpottamalla ylimääräisiä hakuja tarjoamalla automaattisesti vaihtoehtoisia hakuja, mistä hakija saattaisi olla kiinnostunut. Bing siis säästää aikaa päätöksenteossa, ja samalla se kertoo ylimääräistä tai toisin sanoen sellaista tietoa, mitä et osannut edes ajatella.

En usko, että Googlen maailman herruutta hakukoneissa tullaan koskaan viemään, ellei se tee jotain todella typerää hakukoneensa kanssa. Bing on kuitenkin raikas uutuus hakukonemarkkinoilla, josta toivon tulevan Googlen pahin vastus. Uskon, että lähiaikoina Googlekin alkaa parantaa hakutulosten näytettävyyttä Bingin mukaisesti. Itse aion jatkossakin kyllä käyttää Googlea hakukoneena, sillä se on helppokäyttöinen ja nopea. Kuitenkin, jos Bing saisi Suomen versionkin samanlaiseen kuosiin kuin Yhdysvalloissa, niin olen valmis ainakin kokeilemaan sitä. Paras hakukone olisikin sellainen, mikä koostuu Googlen nopeudesta ja indeksistä sekä Bingin tulossivun ominaisuuksista. Ehkäpä sellainenkin hakukone näkee päivänvalon lähitulevaisuudessa – miksei jopa Bing itse.

6 LÄHTEET

Aboba, Bernard 1993. Vinton Cerfin haastattelu. How the Internet Came to be. WWW-dokumentti.

<http://www.virtualschool.edu/mon/Internet/CerfHowInternetCame2B.html>. Päivitetty 6.5.2010. Luettu 6.5.2010.

Alpert, Jesse & Hajaj, Nissan 2008. We knew the web was big... Google Blog.

<http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>. Päivitetty 25.7.2008. Luettu 16.5.2010.

Battelle, John 2005. Search engines: your questions answered. Verkkolehti.

<http://www.guardian.co.uk/technology/2005/oct/05/news.searchengines>. Päivitetty 5.10.2005. Luettu 12.5.2010.

BBC-News 2006. Search site retires iconic Jeeves. WWW-dokumentti.

<http://websearch.about.com/b/2006/02/11/goodbye-jeeves-ask-jeeves-butler-gets-a-pink-slip.htm>. Päivitetty 11.2.2006. Luettu 12.5.2010.

Berners-Lee, Tim 1989. Information Management: A Proposal. WWW-dokumentti.

<http://www.w3.org/History/1989/proposal.html>. Päivitetty 20.6.2001. Luettu 9.5.2010.

Berners-Lee, Tim & Robert Cailliau 1990. WorldWideWeb: Proposal for a HyperText Project. WWW-dokumentti. <http://www.w3.org/Proposal.html>. Päivitetty 7.10.1991.

Luettu 9.5.2010.

Berners-Lee, Tim 2004. The WorldWideWeb Browser. WWW-dokumentti.

<http://www.w3.org/People/Berners-Lee/WorldWideWeb>. Päivitetty 2.11.2004. Luettu 9.5.2010.

Berners-Lee, Tim 2007. Examples of early WWW hypertext. WWW-dokumentti.

<http://www.w3.org/People/Berners-Lee/FAQ.html#Examples>. Päivitetty 20.11.2007. Luettu 9.5.2010.

Boswell, Wendy 2010. History of the Internet. WWW-dokumentti.

<http://websearch.about.com/od/whatistheinternet/a/historyinternet.htm>. Päivitetty 7.5.2010. Luettu 7.5.2010.

Brandt, Daniel 2002. PageRank: Google's Original Sin. WWW-dokumentti.

<http://www.google-watch.org/pagerank.html>. Päivitetty 23.9.2008. Luettu 18.5.2010.

Bright Planet 2000. The Deep Web: Surfacing Hidden Value. PDF- tiedosto.

<http://brightplanet.com/images/uploads/12550176481-deepwebwhitepaper.pdf>. Luettu 22.5.2010.

Buresh, Scott 1998. Ask.com Search Engine – A Brief History. WWW-dokumentti.

<http://www.searchengineguide.com/scott-buresh/askcom-search-engine-a-brief-history.php>. Päivitetty 13.5.2010. Luettu 13.5.2010.

Cullen, Drew 2003. Yahoo! buys! Overture! Verkkolehti.

http://www.theregister.co.uk/2003/07/14/yahoo_buys_overture/. Päivitetty 14.7.2003. Luettu 12.5.2010.

Davis, Harold 2006. Google. Kuinka varmistat verkkonäkyvyytesi. Suomentanut

Niemi, Marko 2007. Jyväskylä: Gummerus.

de Vrieze, Paul 2002. Improving search engine technology. Opinnäytetyö.

Deutsch, Peter 1993. What is Archie. WWW-dokumentti.

<http://www1.chapman.edu/gopher-data/archives/Internet%20Information/whatis.archie>. Päivitetty 7.5.1993. Luettu 8.5.2010.

Franklin, Curt 2000. How Internet Search Engines Work. WWW-dokumentti.

<http://computer.howstuffworks.com/internet/basics/search-engine1.htm>. Päivitetty 14.5.2010. Luettu 14.5.2010.

Funding Universe 2005. Ask Jeeves, Inc. Company history. WWW-dokumentti.
<http://www.fundinguniverse.com/company-histories/Ask-Jeeves-Inc-Company-History.html>. Päivitetty 22.3.2010. Luettu 11.5.2010.

Gaudin, Sharon 2009. Microsoft Bing Bonks Porn in Searches. Computerworld uutiset.
http://www.pcworld.com/article/166712/microsoft_bing_bonks_porn_in_searches.html. Päivitetty 17.6.2009. Luettu 28.5.2010.

Glasner, Joanna 2002. AltaVista Makeover: A Better View. Verkkolehti.
<http://www.wired.com/techbiz/media/news/2002/11/56335>. Päivitetty 13.11.2002.
Luettu 12.5.2010.

Google hakutilasto 2010. Suomen suosituimmat hakusanat 2009. WWW-dokumentti.
<http://www.google.com/insights/search/#geo=FI&date=1%2F2009%2012m&cmpt=q>.
Päivitetty 23.5.2010. Luettu 23.5.2010.

Google Press Center 2000. Yahoo! Selects Google as its Default Search Engine Provider. WWW-uutinen. <http://www.google.com/press/pressrel/pressrelease25.html>.
Päivitetty 13.5.2010. Luettu 15.5.2010.

Griffiths, Richard 2002a. Chapter Two: From ARPANET to World Wide Web. WWW-luento. <http://www.let.leidenuniv.nl/history/ivh/chap2.htm>. Päivitetty 15.12.2006. Luettu 14.5.2010.

Griffiths, Richard 2002b. Chapter Four: Search Engines. WWW-luento. <http://www.let.leidenuniv.nl/history/ivh/chap4.htm>. Päivitetty 15.12.2006. Luettu 14.5.2010.

Hock, Randolph 2001. The Extreme Searcher's Guide to Web Search Engines: A Handbook for the Serious Searchers, Second Edition. Medford: CyberAge Books.

How does Ask search work 2010. Yrityksen WWW-sivut.
http://asksupport.custhelp.com/app/answers/detail/a_id/2475/session/L3NpZC9fNFA2aWwqag==/sno/0. Päivitetty 13.5.2010. Luettu 13.5.2010.

Internet Society 2010. A Brief history of Internet. WWW-dokumentti.

<http://www.isoc.org/internet/history/brief.shtml>. Päivitetty 6.5.2010. Luettu 6.5.2010.

Linnake, Tuomas 2009. Bing aiheuttaa harmaita hiuksia pornopelkoisille. IT-Viikko uutiset. <http://www.itviikko.fi/ratkaisut/2009/06/04/bing-aiheuttaa-harmaita-hiuksia-pornopelkoisille/200913827/7>. . Päivitetty 4.6.2009. Luettu 28.5.2010.

Järvinen, Petteri 2003. Hakukone hakee ja löytää. Tietokone 4, 54.

Kleinrock 1961. WWW-dokumentti.

<http://www.lk.cs.ucla.edu/LK/Bib/REPORT/PhD/proposal.html>. Päivitetty 24.10.2004. Luettu 8.5.2010.

Kleinrock 2004. WWW-dokumentti.

<http://www.docstoc.com/docs/2187564/Microsoft-PowerPoint---Len-Kleinrocks-Brief-History-of-the/>. Päivitetty 8.5.2010. Luettu 8.5.2010.

Kleinrock 2005. WWW-dokumentti.

http://www.lk.cs.ucla.edu/personal_history.html. Päivitetty 25.11.2008. Luettu 8.5.2010.

Kopytoff, Verne 2000. Altavista Switches Web Portal Into High Gear. WWW-dokumentti. <http://www.sfgate.com/cgi-bin/article.cgi>. Päivitystietoja ei saatavilla. Luettu 15.5.2010.

Koster, Martijn 1994. ALIWEB – Archie- Like Indexing in the WEB. PDF-tiedosto.

http://www.iicm.tugraz.at/thesis/cguetl_diss/literatur/Kapitel05/References/Koster_1994/aliweb.pdf. Luettu 10.5.2010.

Layton, Steve, McCormick, John & Rust, Max 2006. How engine works. WWW-

uutinen. <http://www.chicagotribune.com/technology/chi-060921google-graphic,0,5529341.graphic>. Päivitetty 21.9.2006. Luettu 16.5.2010.

Lycos Acquires Wired Digital. 1998. Wired News Report. Verkkolehti.
<http://www.wired.com/techbiz/media/news/1998/10/15437>. Päivitetty 10.6.1998 .
Luettu 21.5.2010.

Mauldin, Michael 1997. Lycos: Design choices in an Internet search service. WWW-dokumentti. <http://robot-club.com/lti/pub/ieee97.html>. Päivitetty 16.8.2002. Luettu 12.5.2010.

McIntyre, Ryan 2007. A Brief History of Excite. WWW-dokumentti.
<http://www.ryanmcintyre.com/wp/archives/2007/05/a-brief-history-of-excite.html>.
Päivitetty 21.5.2007. Luettu 11.5.2010.

Miller, Rich 2008. Google Data Center FAQ. WWW-dokumentti.
<http://www.datacenterknowledge.com/archives/2008/03/27/google-data-center-faq/>.
Päivitetty 27.3.2008. Luettu 15.5.2010.

MTV3- uutinen 2009. Bing haastaa Googlen. WWW-uutinen
<http://www.mtv3.fi/uutiset/talous.shtml/arkistot/talous/2009/07/922703>. Päivitetty
9.7.2009. Luettu 26.5.2010.

NCSA 2009. About NCSA Mosaic. WWW-dokumentti.
<http://www.ncsa.illinois.edu/Projects/mosaic.html>. Päivitetty 16.5.2010. Luettu
16.5.2010.

Northern Light Search 2010. General Questions. Yrityksen WWW-sivut.
http://nlsearch.com/faq_general.php#q01. Päivitetty 14.5.2010. Luettu 14.5.2010.

Pinkerton, Brian 2004. WebCrawler facts: WebCrawler Timeline. WWW-dokumentti.
<http://www.thinkpink.com/bp/WebCrawler/History.html> Päivitetty 28.6.2004. Luettu
10.5.2010

Rhodes, Matt 2010. Understanding How Google PageRank Works For Your Website.
WWW-dokumentti. <http://ezinearticles.com/?Understanding-How-Google-PageRank-Works-For-Your-Website&id=3700183>. Päivitetty 15.5.2010. Luettu 15.5.2010.

Richardson, Tim 2005. Ask Jeeves buys Excite Europe. WWW-uutinen.
http://www.theregister.co.uk/2005/05/20/excite_tiscali/. Päivitetty 20.5.2005. Luettu 14.5.2010.

Roberts 1997. Internet Chronology. WWW-dokumentti.
<http://www.ziplink.net/users/lroberts/InternetChronology.html>. Päivitetty 25.10.1999. Luettu 8.5.2010.

Roberts 1999. The First Theory of Packet Networks 1957-1964. WWW-dokumentti.
http://www.ziplink.net/users/lroberts/SIGCOMM99_files/v3_document.htm. Päivitetty 16.9.1999. Luettu 8.5.2010.

Robotstxt.org 2007. About/ robots.txt. WWW-dokumentti.
<http://www.robotstxt.org/robotstxt.html>. Päivitetty 20.5.2010. Luettu 20.5.2010.

Search Journal 2003. Lycos Heats Up HotBot With The Launch of HotBot's Quick-Search Deskbar. WWW-dokumentti.
http://www.searchjournal.com/content/hotbot_deskbar.html. Päivitetty 24.6.2003. Luettu 12.5.2010.

Search Tools 2008. The Robots Exclusion Protocol (REP). WWW-dokumentti.
<http://www.searchtools.com/robots/robots-exclusion-protocol.html>. Päivitetty 11.7.2008. Luettu 20.5.2010.

SEO Treo 2010. Search Engine History. WWW-dokumentti.
<http://www.seotreo.com/seo-blog/search-engine-history.html>. Päivitetty 19.3.2010. Luettu 11.5.2010

Sherman, Chris 2002. Happy Birthday, Aliweb. WWW-dokumentti.
<http://searchenginewatch.com/2161311>. Päivitetty 3.12.2002. Luettu 11.5.2010.

Sherman, Chris 2002. Happy Birthday, Lycos! WWW-dokumentti.
<http://searchenginewatch.com/2160551>. Päivitetty 14.8.2002. Luettu 11.5.2010.

Sonnenreich, Wes 1997. A history of Search Engines. WWW-dokumentti. <http://www.wiley.com/legacy/compbooks/sonnenreich/history.html#crawler>. Päivitetty 10.5.2010. Luettu 10.5.2010.

StatCounter Global Stats 2010a. Top Search Engines in Finland from Jul 08 to May 2010. WWW-dokumentti. http://gs.statcounter.com/#search_engine-FI-monthly-200807-201005. Päivitetty 23.5.2010. Luettu 23.5.2010.

StatCounter Global Stats 2010b. Yrityksen WWW-sivut. <http://gs.statcounter.com/>. Päivitetty 23.5.2010. Luettu 23.5.2010.

Szuporwicz, Bohdan O. 1997. Search Engine Technologies for the World Wide Web and Intranets. Charleston: Computer Technology Research Corp.

TechTarget 2005. Verkkolehti. http://whatis.techtarget.com/definition/0,,sid9_gci213333,00.html#. Päivitetty 17.5.2010. Luettu 17.5.2010.

The Web Search Workshop 2010. A brief history of the AltaVista search engine. WWW-dokumentti. <http://www.websearchworkshop.co.uk/archive/altavista-history.htm>. Päivitetty 14.5.2010. Luettu 14.5.2010.

Tilastokeskus 2009. Internetin käytön yleistymisen pysähtyi. WWW-dokumentti. http://www.stat.fi/til/sutivi/2009/sutivi_2009_2009-09-08_tie_001.html. Päivitetty 8.9.2009. Luettu 23.5.2010.

Underwood, Lee 2004. Search history. WWW-dokumentti. http://www.webreference.com/authoring/search_history/. Päivitetty 11.5.2010. Luettu 11.5.2010.

Wall, Aaron 2008. History of Search Engines: From 1945 to Google 2007. WWW-dokumentti. <http://www.searchenginehistory.com/#directories>. Päivitetty 14.5.2010. Luettu 14.5.2010.

Web search 2010. About. Yrityksen WWW-sivut.

<http://about.ask.com/en/docs/about/webmasters.shtml>. Päivitetty 12.5.2010. Luettu 12.5.2010.

Wikipedia 2010a. Internet. WWW-dokumentti. <http://en.wikipedia.org/wiki/Internet>. Päivitetty 4.5.2010. Luettu 5.5.2010.

Wikipedia 2010b. Archie search engine. WWW-dokumentti.

http://en.wikipedia.org/wiki/Archie_search_engine. Päivitetty 7.5.2010. Luettu 9.5.2010.

Wikipedia 2010c. Grep. WWW-dokumentti. <http://en.wikipedia.org/wiki/Grep> Päivitetty 8.5.2010. Luettu 9.5.2010.

Wikipedia 2010d. Wide Area Information Servers. WWW-dokumentti.

http://en.wikipedia.org/wiki/Wide_Area_Information_Servers. Päivitetty 9.5.2010. Luettu 9.5.2010.

Wikipedia 2010e. Mosaic browser. WWW-dokumentti.

http://en.wikipedia.org/wiki/Mosaic_browser. Päivitetty 17.5.2010. Luettu 20.5.2010.

Wright, Alex 2009. Exploring “Deep Web” That Google Can’t Crasp. Verkkolehti.

http://www.nytimes.com/2009/02/23/technology/internet/23search.html?_r=1. Päivitetty 23.2.2009. Luettu 20.5.2010.

Yahoo! 2005. The History of Yahoo! – How It All Started. WWW-dokumentti.

<http://docs.yahoo.com/info/misc/history.html>. Päivitetty 14.5.2010. Luettu 14.5.2010.

7 LIITTEET

LIITE 1(1). Listausta muutamasta hakurobotista

ABCdatos BotLink
Acme.Spider
Ahoy! The Homepage Finder
Alkaline
Anthill
Walhello appie
Arachnophilia
Arale
Araneo
AraybOt
ArchitextSpider
Aretha
ARIADNE
arks
AskJeeves
ASpider (Associative Spider)
ATN Worldwide
Atomz.com Search Robot
AURESYS
BackRub
Bay Spider
BBot
Big Brother
Bjaaland
BlackWidow
Die Blinde Kuh
Bloodhound
Borg-Bot
BoxSeaBot
bright.net caching robot
BSpider
CACTVS Chemistry Spider
Calif
Cassandra
Digimarc Marcs spider/CGI
Checkbot
ChristCrawler.com
churl
cIeNcIaFiCcIoN.nEt
CMC/0.01
Collective
Combine System
Conceptbot
ConfuzzledBot
CoolBot
Web Core / Roots

JATKU...

LIITE 1(2). Listaus muutamasta hakurobotista


XYLEME Robot
Internet Cruiser Robot
Cusco
CyberSpyder Link Test
CydralSpider
Desert Realm Spider
DeWeb(c) Katalog/Index
DienstSpider
Digger
Digital Integrity Robot
Direct Hit Grabber
DNAbot
DownLoad Express
DragonBot
DWCP (Dridus' Web Cataloging Project)
e-collector
EbiNess
EIT Link Verifier Robot
ELFINBOT
Emacs-w3 Search Engine
ananzi
esculapio
Esther
Evliya Celebi
FastCrawler
Fluid Dynamics Search Engine robot
Felix IDE
Wild Ferret Web Hopper #1, #2, #3
FetchRover
fido
Hämähäkki
KIT-Fireball
Fish search
Fouineur
Robot Francoroute
Freecrawl
FunnelWeb
gammaSpider, FocusedCrawler
gazz
GCreep
GetBot
GetURL
Golem
Googlebot
Grapnel/0.01 Experiment
Griffon
Gromit
Northern Light Gulliver
Gulper Bot
HamBot

LIITE 2. Studentin kotisivu selaimessa



MIKKELIN AMMATTIKORKEAKOULU
Mikkeli University of Applied Sciences

STUDENT



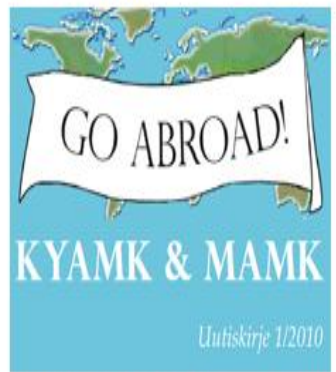
00:00 | 00:00

AJANKOHTAISTA
PERUSTIETOA
OPINNOISTA
TUKIPALVELUT
MAMK-INFO
YHTEYSTIEDOT
LAADUNVARMISTUS
TYÖVÄLINEET
OPINNÄYTETYÖ
OPISKELIJATOIMINTA
KESKUSTELU
PALAUTE

Oikopolut
Valitse...
Verkkosivut
Valitse...
Ruokalistat
Valitse...

Suosikit
Lukujärjestykset
Tuutorit
Terveystuulet
Ilmoitustaulu
Opinto-opas

MAMK-utiset



KYAMK & MAMK UUTISKIRJE 1/2010

Opastettu kierros Savonlinnassa Talvisalon
hautausmaalla 25.5. klo 17.00.
21.05.2010

Vanhan Moodle väistyy uuden tieltä lopullisesti
ti 1.6.
20.05.2010

Mikkeliin opiskeluterveydenhuollon toiminta
kesääjalla
19.05.2010

Future Science Groupin lääketieteen
lehtikokeimat ovat koekäytössä 31.12.2010
asti
18.05.2010

AMK-hallitus esittää muotoilujen ja
Savonlinnan liiketalouden lopettamista
17.05.2010

Yhden koulun opintotoimien sulkeutuminen

Tapahtumat

28.5.2010
Valmistumis- ja kevätkuuhla Mikkeliissä

28.5.2010
Valmistumis- ja kevätkuuhla Pieksämäellä

28.5.2010
Valmistumis- ja kevätkuuhla Savonlinnassa

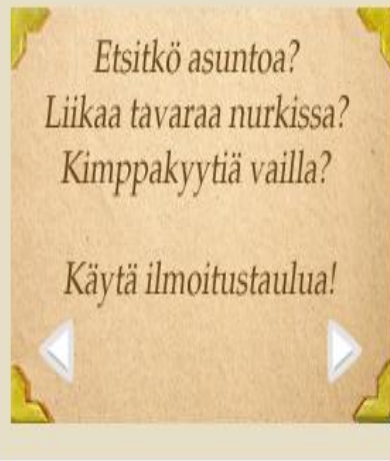
11.6.2010
Mikkeli soikoon - Rock on the Corner -kiertue

12.6.2010
Mikkeli soikoon - Lava X

29.7.2010-1.8.2010
MAMK mukana Farmari-messuilla

2.8.2010-6.8.2010
Kampusjopo Ride Hanko-Mikkeli

13.8.2010-14.8.2010
MAMKin tapahtumateltta Jurassic Rockissa




Pikauutiset

[In English](#)

Tekstikoko: [A](#) [A](#) [A](#)

Haku sisällöstä
Hae

 [RSS-Feed](#)

Gallup

Minkä arvosanan annat
itsellesi kuluneen
lukuvuoden työstä?

Kiitettävä
 Hyvä
 Tyydyttävä
 Väiltävä
 Heikko

Vastaa

KK2020

X-rakennus valmistuu
aikataulussa 2011
3000 opiskelijaa
Kasarmien kampuksella
2010. vuodelle

LIITE 3. Pätkä Studentin kotisivua HTML-kielillä

```

<html>
<head>

<title>STUDENT</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
<meta http-equiv="Pragma" content="no-cache">
<meta name="Description" content="">
<meta name="Keywords" content="">
<meta name="Copyright" content="">
<meta NAME="ROBOTS" CONTENT="NOARCHIVE">
<meta name="Robots" content="index,follow">
<script defer type="text/javascript" src="/gui/default/js/pngfix.js"></script>
<script type="text/javascript" src="/gui/default/js/prototype.js"></script>
<script type="text/javascript" src="/gui/default/js/effects.js"></script>
<script type="text/javascript" src="/gui/default/js/glider.js"></script>
<link rel="stylesheet" href="/gui/default/css/rotaattori.css" type="text/css" />
<link rel="stylesheet" href="/gui/default/css_student/ie_new.css" type="text/css" />
<link rel="stylesheet" href="/gui/default/css_student/fontit.css" type="text/css" />
<link rel="stylesheet" href="/gui/default/css_student/fr_styles.css" type="text/css">
<link rel="stylesheet" href="/gui/default/css/fontit.asp" type="text/css" media="screen,print" />
<link rel="stylesheet" href="/gui/default/css_student/print.css" type="text/css" media="print" />

<script language="JavaScript" type="text/javascript">ulm_save_doc=true; function adddownfavorite(link) { var agree=confirm("Tallennetaanko sivu omiin suosikkeihin?"); if(agree){ window.document.savefavorite.action=link; window.document.savefavorite.method="POST"; window.document.savefavorite.submit();} }</script>

<style>
#layout-taulukko {}
#topkuva {}
#valikko {}
#sisalto {}
#footer {}
.mainmenu {color: black}

/* #frameless_top {display: block; width: 200px; background-color: transparent; border: 0px; position: absolute;} */

#mamk_oikopolut {margin-left: 20px; width: 140px; border-top: 1px solid white; border-bottom: 1px solid white; padding-top: 10px; padding-bottom: 5px;}
#mamk_oikopolut select {width: 140px; margin-bottom: 7px;}
</style>

</head>
<body bgcolor=white marginheight="0" marginwidth="0" topmargin="0" leftmargin="0" onload="loadBody();">
<form name="savefavorite"><input type="hidden" name="sft" value="1"></form>

<form name="ws_form" action="" method="" target="">
    <input type=hidden name="" id="ws_searchparam" value="">
    <input type=hidden name="ws_id" id="ws_id" value="">
</form>

<!-- <meta http-equiv="Refresh" content="">-->

```

LIITE 4. Studentin kotisivu hakurobotin silmin

STUDENT FIN ENG SWE Mikkelin ammattikorkeakoulu Etusivulle Get the Flash
 Player to see this player AJANKOHTAISTA PERUSTIETOA OPINNOISTA
 TUKIPALVELUT MAMK INFO YHTEYSTIEDOT LAADUNVARMISTUS TYÖ
 ÄLINEET OPINNÄYTETYÖ OPISKELIJATOIMINTA KESKUSTELU PALAUTE
 Oikopolut Valitse Asio opiskelijatoiminnot Mikki Moodle Moodle uusi Netstorage
 Opinto opas ähköposti Theseus Webropol Wiki Verkkosivut Valitse CampusTV Jobs-
 tep Kampusfestari Kirjasto MAMK WWW sivut MAMOK Ravintola Talli Sinne on
 äästä ä Vanha Student VirtuaaliAMK Ruokalistat Valitse Florence Kasarmina Kult-
 tuurin ja nuorisotyö kampus Nikkarilan kampus Ravintola Talli Savonniemen kampus
 Suosikit Lukujärjestykset Tuutorit Terveystieteiden Ilmoitustaulu Opinto opas
 MAMK uutiset KYAMK MAMK UUTISKIRJE 2010 Opastettu kierros Savonlinnas-
 sa Talvisalon hautausmaalla 25 klo 17 00 21 05 2010 Vanhan Moodle äistyy uuden
 tieltä lopullisesti ti 20 05 2010 Mikkelin opiskeluterveydenhuollon toiminta kesäajalla
 19 05 2010 Future Science Groupin ääketieteen lehtikokoelmat ovat koekäytössä 31
 12 2010 asti 18 05 2010 AMK hallitus esittää muotoilujen ja Savonlinnan liiketalou-
 den lopettamista 17 05 2010 KN kampuksen opintotoimiston aukioloajat toukokuussa
 12 05 2010 Lisää Tapahtumat 28 2010 Valmistumis ja kevätjuhla Mikkeli 28 2010
 Valmistumis ja kevätjuhla Pieksä äellä 28 2010 Valmistumis ja kevätjuhla Savonlin-
 nassa 11 2010 Mikkeli soikoon Rock on the Corner kiertue 12 2010 Mikkeli soikoon
 Lava 29 2010 2010 MAMK mukana Farmari messuilla 2010 2010 Kampusjopo Ride
 Hanko Mikkeli 13 2010 14 2010 MAMKin tapahtumateltoa Jurassic Rockissa Pikauu-
 tiset MAMK Facebookissa In English Tekstikoko Haku sisällöstä Henkilöhaku RSS
 Feed Gallup Minkä arvosanan annat itsellesi kuluneen lukuvuoden työstä Kiitettä ä
 Hyvä Tyydytt ä älttä ä Heikko KK2020 rakennus valmistuu aikataulussa 2011 15 04
 2010 3000 opiskelijaa Kasarmin kampuksella 2010 syksyllä 18 02 2010 Avoimet työ-
 paikat Ei tiedostoja Student Selaaja Tulosta sivu Edellinen sivu Sivun ylö