

SATAKUNNAN AMMATTIKORKEAKOULU

Tarja Suojala

HAKUKONEIDEN VERTAILU JA INDEKSOINTI WWW:SSÄ

Liiketalous, matkailu, tietojenkäsittely ja viestintä

Tietojenkäsittelyn koulutusohjelma

2007

HAKUKONEIDEN VERTAILU JA INDEKSOINTI WWW:SSÄ

Suojala, Tarja Kristiina
Satakunnan ammattikorkeakoulu
Liiketalous, matkailu, tietojenkäsittely ja viestintä Pori
Tietojenkäsittelyn koulutusohjelma
Helmikuu 2007
Grönholm, Jukka
UDK: 004.738.5, 004.738.52, 025.4.03
Sivumäärä: 100

Asiasanat: hakupalvelut, Internet, metadata, tiedonhaku

Opinnäytetyö käsittelee Internetin hakukoneita ja niiden indeksointimenetelmiä. Tavoitteena oli selvittää, tukevatko tutkimukseen valitut viisi hakukonetta Boolean hakuja ja millä hakukoneella saadaan parhaat hakutulokset.

Työn teoriaosassa kerrotaan, mistä osista hakukone koostuu sekä miten hakua voidaan muotoilla ja rajata relevanttien hakutulosten saamiseksi. Yleinen tapa muotoilla hakulauseketta on käyttää Boolean operaattoreita AND, OR ja NOT, joiden avulla hakutulosta on mahdollista laajentaa tai supistaa.

Teoriataustassa syvennyttään myös hakukoneiden suorittamaan indeksointiin eli www-sivujen lisäämiseen hakukoneisiin. Indeksoinnissa hakukoneen hakurobotti hakee www-sivuja ja tallentaa niistä tiettyjä tietoja tietokantaan. Www-sivun sisällön kuvailemiseen ja muun sivuun liittyvän tiedon ilmoittamiseen voidaan käyttää metadataa, joka piilotetaan sivun HTML-koodiin. Yksi metadataformaateista on Dublin Core, joka kehitettiin alun perin erityisesti www-sivujen metadatan kuvailemiseen. Www-sivujen sijoittumiseen hakukoneiden hakutuloksissa voidaan pyrkiä vaikuttamaan sivujen rakenteella, sisällöllä ja linkityksillä. Tästä käytetään nimitystä hakukoneoptimointi.

Työn testausosassa on kaksi eri osiota. Ensimmäisessä osiossa on selvitetty kahdeksan hyvän hakukoneen kriteeriä ja arvioitu hakukoneita niiden perusteella. Jälkimmäisessä osiossa hakukoneita on testattu viidellä hakulauseella, jotka on muotoiltu niin, että kaikki Boolean operaattorit on saatu mukaan. Jokaisen hakukoneen antamista tulossivuista on tarkastettu 10 ensimmäistä, ja ne on pisteytetty asteikolla 1-3 siten, että mitä relevantimpi sivu on ollut käytettyjen hakusanojen kannalta, sitä enemmän se on saanut pisteitä. Lopuksi kunkin hakukoneen saamat pisteet on laskettu yhteen.

Tulosten perusteella tutkimuksessa mukana olleista hakukoneista paras oli markkinajohtaja Google, jonka etumatka muihin koneisiin oli selvä. Toiselle sijalle selvisi AltaVista ja kolmannelle Yahoo!. Microsoftin syksyllä 2006 julkaissama Live Search sijoittui neljänneksi ja juuri tutkimuksen teon aikoihin uusiutunut Www.fi viidenneksi.

COMPARISON OF SEARCH ENGINES AND INDEXING IN THE WWW

Suojala, Tarja Kristiina

Satakunta University of Applied Sciences

School of Business, Tourism, Business Information Systems, Media and Communication Pori

Degree Programme in Business Information Systems

February 2007

Grönholm, Jukka

UDK: 004.738.5, 004.738.52, 025.4.03

Pages: 100

Keywords: information search, Internet, metadata, search services

This thesis deals with Internet search engines and the indexing methods of them. The aim was to find out if the five search engines chosen for the study support Boolean searches and which search engine gives the best results.

The subtext reveals which parts a search engine consists of and how a search can be formulated and defined to get relevant search results. A common way to formulate a search phrase is to use Boolean operators AND, OR and NOT with which the search result can be expanded or limited.

The subtext also goes deeply into the search engine indexing, i.e. including websites into search engines. In indexing, the search robot of the search engine retrieves websites and saves certain information about them into the database. To describe the content of a website and to present additional information related to it, one can use metadata which is hidden in the HTML tag of the website. One of the metadata formats is Dublin Core which was originally developed especially for describing the metadata of websites. Website ranking can be influenced by the structure, the content and the linking of the site. This activity is called search engine optimization.

The empiric part consists of two sections. In the first one, there are eight criteria for a good search engine, and the search engines have been evaluated using them. In the latter one, the search engines have been tested using five test search phrases that include all Boolean operators. The first ten results returned by each search engine have been checked and they have been given points using scale 1-3. The more relevant the website is compared to the used search terms, the more it has been given points. Finally, the points of each search engine have been summed.

The results show that the best one of the search engines in this study was the market leader Google, the lead of which over other engines was clear. AltaVista came second and Yahoo! came third. Live Search, released by Microsoft in fall 2006, was fourth. Wwww.fi, which was renovated just before the study was made, was fifth.

SISÄLLYS

1 JOHDANTO	6
2 YLEISTÄ HAKUKONEISTA.....	8
2.1 Perinteiset hakukoneet	9
2.2 Metahakukoneet	13
3 HAUN MUOTOILU	14
3.1 Boolean logiikka	14
3.2 Läheisyysoperaattorit	18
3.3 Fraasihaku	19
3.4 Katkaisu	21
3.5 Hakukentät	22
3.6 Muita hakumahdollisuuksia	24
4 INDEKSOINTI	25
4.1 Hakurobotit	27
4.2 Metadata.....	33
4.2.1 Yleistä metadatatista.....	33
4.2.2 Dublin Core.....	35
5 HAKUKONEOPTIMOINTI JA -MARKKINOINTI.....	40
5.1 Hakukoneoptimointi.....	40
5.1.1 Avainsanat.....	41
5.1.2 Meta-komennot	43
5.1.3 Web Design.....	46
5.1.4 Linkkisuosion kasvattaminen.....	48
5.1.5 Teemat.....	50
5.2 Hakukonemarkkinointi.....	50
6 VERTAILTAVAT HAKUKONEET.....	52
6.1 Google	53

6.2 Yahoo!.....	56
6.3 AltaVista	58
6.4 Live Search.....	60
6.5 Www.fi.....	63
7 TUTKIMUKSESSA KÄYTETYT TUTKIMUSMENETELMÄT	65
8 KOEHAUT JA TULOKSET	69
9 YHTEENVETO	90
LÄHTEET	94
Kirjat	94
Www-sivut	95

1 JOHDANTO

Tässä tutkielmassa käsitellään Internetin hakukoneita ja niiden suorittamaa indeksointia eli sivujen lisäämistä hakutietokantaan sekä tehdään vertailu viiden hakukoneen välillä. Ensikosketukseni hakukoneisiin tapahtui 90-luvun puolessavälissä Internet-kurssilla, jolla pääsin kokeilemaan AltaVista-hakukonetta. Siitä lähtien olen käyttänyt hakukoneita aktiivisesti niin opinnoissani ja töissäni kuin harrastuksiini ja vapaa-aikaani liittyvän tiedon etsimisessä. Näiden kymmenen vuoden aikana Internet on kasvanut trendi-ilmiöstä osaksi arkea, ja sen sisältämien sivujen määrä on moninkertaistunut. Kasvaminen on asettanut omat haasteensa sekä tiedonhakijoille että hakukoneille. Samalla hakutermillä, jolla hakukone löysi kymmenen vuotta sitten parikymmentä dokumenttia, löytyy tänä päivänä helposti useita kymmeniä tuhansia sivuja. Tästä syystä hakukoneiden käyttäjien on täytyntä opetella keinoja, joilla hakua voidaan muotoilla ja rajata osuvampien hakutulosten saamiseksi. Myös hakukoneiden työ on lisääntynyt: uusia indeksoitavia sivuja ilmestyy Internetiin päivittäin tuhansia, minkä lisäksi uusia hakukoneita syntyy jatkuvasti jo olemassa olevien hakukoneiden puutteita korjaamaan. Tämä luonnollisesti lisää kilpailua ja pakottaa hakukoneet parantamaan indeksointiaan, sillä mikäli käyttäjä saa hakukoneella toistuvasti vääränlaisia tuloksia, hän voi helposti siirtyä jonkin toisen hakukoneen käyttäjäksi.

Tutkielmalla oli useita eri tavoitteita. Ensimmäinen halusin laatia tietopakettien keskeisimmistä hakukonehaun muotoilu- ja rajauskeinoista havainnollistavien esimerkkien kera. Tässä minua inspiroivat ne tuttavani, joilla on minua vähemmän kokemusta hakukoneiden käytöstä ja jotka pyytävät minulta usein apua "googletamisessa", niin kuin useimmat heistä Internet-tiedonhakua kutsuvat. Toiseksi tahdoin selvittää sekä lukijoille että itselleni, miten www-sivut päätyvät hakukoneiden tietokantoihin sekä miten sivun omistaja voi vaikuttaa sivunsa sijoittumiseen hakutuloksissa. Tästä tiedosta on varmasti hyötyä etenkin niille, joilla on omat www-sivut ja jotka haluavat sivunsa mahdollisimman monien nähtäville. Kolmas tavoite oli tutkia, onko Google - jonka käytöstä on todellakin tullut monille Internet-tiedonhaun synonyymi - ansaitusti maailman suosituin hakukone,

vai antaako jokin muu hakukone parempia ja osuvampia tuloksia. Googlelle valitsemisissä haastajissa on niin vanhoja hakukonekonkareita kuin uusia tulokkaitakin. Mielenkiinnosta otin ulkomaisten hakijoiden rinnalle myös yhden kotimaisen hakukoneen.

Mielestäni paras tapa mitata hakukoneiden paremmuutta ja tehokkuutta on koehakujen tekeminen, ja sitä tapaa käytetään tässäkin tutkimuksessa. Testaan hakukoneita koehakulauseilla ja annan löydetyille sivuille pisteitä sen mukaan, miten hyvin ne vastaavat hakulauseita. Tämän lisäksi arvioin hakukoneita kahdeksalla kriteerillä, joita on aiemminkin käytetty hakukoneita arvioitaessa. Koehakulauseet ja kriteerit esitellään tarkemmin luvussa 7. Se hakukone, jolla on lopussa eniten pisteitä, on tämän tutkimuksen paras.

Varsinainen tutkimusongelmani oli selvittää, tukevatko tutkimukseen valitut hakukoneet Boolean hakuja, ja jos tukevat, onko niiden välillä eroja sen suhteen, miten hyvin ne kyseisiä hakuja tukevat. Boolean logiikka tuli minulle tutuksi jo tietojenkäsittelymatematiikan opintojaksolla syksyllä 2003, mutta vasta paljon myöhemmin minulle selvisi, että Boolean operaattoreita voidaan hyödyntää hakukoneissa hakujen rajaamisessa. Tutkimuksen koehakulauseet on muotoiltu niin, että kaikki Boolean operaattorit on saatu käyttöön.

Aiheesta on olemassa aikaisempia tutkimuksia, muun muassa Hellsténin (2004) Savonia-ammattikorkeakoulussa tekemä. Kuten kuitenkin aiemmin totesin, Internet-tiedonhaku uudistuu ja kehittyy kaiken aikaa, minkä takia jo muutamaa kuukautta vanhemmat tutkimukset alkavat olla vanhentuneita. Tämän tutkimuksen hakukoneet ovat kahta lukuun ottamatta eri hakukoneita kuin Hellsténin tutkimuksessa.

2 YLEISTÄ HAKUKONEISTA

Internetissä olevan aineiston laajuus on suunnaton. Viimeisimpien arvioiden mukaan Internetissä on jopa 1 000 miljardia dokumenttia, joten on helppo uskoa, että tiedonhakijan olisi käytännössä mahdotonta löytää itseään kiinnostavia tietoja tästä sivustopaljoudesta ilman erityisiä apuvälineitä. Joskus on toki helppo päättellä, mistä tarvittava tieto löytyy - esimerkiksi Ylen kanavien tv-ohjelmätietoja kaipaava osaa todennäköisesti etsiä niitä osoitteesta <http://www.yle.fi> - mutta hyvin usein ongelma on nimenomaan se, ettei hakijalla ole aavistustakaan, mistä hän lähtisi tietoa etsimään. Ensimmäinen ratkaisu tähän saatiin vuonna 1994, kun ensimmäiset luettelot, Galaxy ja Yahoo! perustettiin. Nämä olivat kuitenkin vain luetteloita, joissa sivut oli jaoteltu eri ryhmiin aihepiirin mukaan, mikä helpotti tiedon löytämistä. Parempi ratkaisu Internet-kaaokseen saatiin, kun tiedonhaku automatisoitiin ja hakukoneet kehitettiin. (MTV3 Internet, viitattu 16.10.2006; Perkiömäki 1997; Uutistoimisto.com 2006).

Tämän päivän hektisessä maailmassa ihmiset arvostavat tiedonhaussakin nopeutta ja vaivattomuutta, ja yksi Internetin vahvuuksista on juuri sen nopeus etsittäessä tietoa. Ilman tehokasta hakukonetta tarvittavan tiedon löytäminen voisi olla verrattavissa kirjan etsimiseen suuresta kirjastosta, jossa ei ole kortistoa, jossa kirjat säilytetään täysin satunnaisessa järjestyksessä ja johon kuka tahansa meistä voisi itse lisätä omat kirjansa. Hyvä hakukone käyttää edistynyttä tekniikkaa osuvien hakutulosten saamiseen jokaiseen kyselyyn, ja tehokas palvelin vastaa asiakkaan kyselyyn laittaen www-sivustot tarkoituksenmukaiseen järjestykseen - tiedonhakijalla ei useinkaan ole aikaa selailla turhauttavia tuloksia, jotka eivät liity hakutermeihin millään tavoin. (Hirsjärvi, Remes & Sajavaara 2005, 84-85; MTV3 Internet, viitattu 16.10.2006; Nikkilä 2001, 176).

2.1 Perinteiset hakukoneet

Hakukoneet (*search engines*), joita kutsutaan toisinaan myös nimellä hakupalvelu, ovat kaikista Internetin tiedonhankintatavoista suosituimpia ja varmasti lähes kaikille Internetin käyttäjille tuttuja. Nimitys hakukone on itse asiassa hieman harhaanjohtava, sillä hakukone ei nimestään huolimatta ole kone. Tässä tutkielmassa sillä tarkoitetaan kokonaista järjestelmää, jossa on monta suhteellisen eri tyyppistä osaa. (Aniway 1999-2005; Haasio 2003, 14; Korhonen 2003, 49; Korpela 2006).

Hakukone muodostuu tyypillisesti seuraavista osista:

- Www-sivulla on **lomake**, joka muodostaa käyttöliittymän. Sen avulla hakija voi syöttää hakusanoja ja pyytää hakukonetta hakemaan tietokannastaan sivuja, joilta kyseiset sanat löytyvät. Lomake on luonnollisesti tärkeä - mutta myös yksinkertaisin - osa hakukonetta. Se voi olla rakenteeltaan hyvinkin pelkistetty.
- **Hakupalvelin** ottaa vastaan lomakkeen avulla annettuja hakupyynnöitä ja toteuttaa niitä. Suurissa hakupalveluissa on monta rinnakkaista hakupalvelinta niin, että erillinen järjestely pyrkii ohjaamaan pyynnöt eri palvelimille estäen niiden ylikuormittumista.
- **Tietokantaan** tallentuu dataa sivuilta. Kun hakupalvelin saa hakupyynnön, se ei etsi tietoa kaikilta Internetissä olevilta sivuilta, vaan tietokannastaan, joka on tehty mahdollisimman nopeaksi käsitellä. Jos hakusanoja vastaavia sivuja löytyy, lähettää hakupalvelin selaimelle - ja sen myötä hakijalle - tietoa löytämistään sivustoista, kuten niiden url-osoitteet, otsikot ja mahdollisesti jonkinlaiset tiivistelmät niiden sisällöstä. On mahdollista, että kun hakija yrittää seurata linkkiä löytyneeseen sivuun, ei sivua ole enää olemassa, koska se on poistunut www:stä sen jälkeen, kun se oli päätynyt tietokantaan.

Hakukone ei anna tuloksia pelkästään sivujen sisällön perusteella. Tämän tutkimuksen hakukoneista Googlen toiminta perustuu PageRank-ohjelmaan, joka käyttää hyväkseen Internetin linkkirakenteita. Ohjelma on nimetty Googlen perustajajäsenen Larry Pagen mukaan. Google tulkitsee sivulta A sivulle B johtavan linkin ikään kuin sivu A "äänestäisi" sivua B, eli mitä enemmän link-

kejä tiettyyn www-sivuun osoittaa, sitä arvokkaamman hakukone olettaa sen olevan. Hakutulosten kärjessä ovat ne sivut, joille osoittaa eniten linkkejä. Google analysoi myös sen sivun, joka antaa äänen: "tärkeiden" sivujen antamalla äänillä on enemmän arvoa, ja ne auttavat tekemään äänestämistään sivuista tärkeitä. Korkealaatuiset ja tärkeät www-sivut saavat korkeamman PageRank-arvon, jonka Google muistaa jokaisella hakukerralla.

PageRank-ohjelmaa voidaan valitettavasti huijata. Niin sanotut linkkifarmit pyrkivät keinotekoisesti nostamaan www-sivun arvoa korkeammaksi. Linkkifarmit ovat sivustoja, joissa on pelkästään linkkejä toisiin sivustoihin. Google pyrkii kuitenkin havaitsemaan ne ja saattaa jopa rangaista niitä nollaamalla niiden PageRank-äänimäärän.

- **Indeksointijärjestelmä** tarkoittaa automaattisesti toimivaa ohjelmistoa, joka käy läpi www-sivuja kartuttaen ja päivittäen hakukoneen tietokantaa. Indeksointijärjestelmä on tavallaan www:n käyttäjän vastine siinä mielessä, että se käy sivuilla, joilla se on aiemmin käynyt tai jotka sille erikseen kerrotaan, ja seuraa sivuilla olevia linkkejä. Lukemisen sijaan se kuitenkin käsittelee sivujen tekstuaalista sisältöä, huomioiden mahdollisesti merkkauksen (esimerkiksi painottaen otsikoita enemmän kuin normaalitekstiä) ja vie tietoa siitä tietokantaan.

(Google 2002; Koulutus- ja konsultointipalvelu KK Mediat 2000-2006; Korpela 2006; Setälä 2004).

Hakukoneet ovat yksi tärkeimpiä syitä tehdä www-sivuista "tekstiystävällisiä". Jos sivulla on esimerkiksi vain kuvagalleria ilman mitään selityksiä, pitää hakukone sitä sisällöttömänä sivuna. Mutta jos sivulla on otsikko, kuvatekstit sekä lyhyt selitys siitä, millaisista kuvista on kyse, lyhyesti sanottuna metadataa, on sillä jo indeksoitavaa. Tämän perusteella hakija, joka etsii tiettyyn aiheeseen liittyviä kuvia, voi löytää kyseisen kuvagallerian. (Korpela 2006).

Hakukoneella on tietty pisteytysjärjestelmä, jossa mitataan muun muassa sanan esiintymistäajuutta sivulla sekä sitä, löytyykö sana sivun otsikosta, url-osoitteesta tai alusta. Tehokas hakukone asettaa hakutuloksissa olevat www-sivut tarkoituksenmukaiseen järjestykseen: ne sivut, jotka hakukone on arvioinut parhaiksi, ovat listan kärjessä. (Hirsjärvi ym. 2004, 85; Korhonen 2003, 49-50).

Periaatteet, joilla hakukone määrittelee hakutulosten tärkeysjärjestyksen ja ne kriteerit, joilla www-sivut hakutulokseen päätyvät, riippuvat hakukoneesta. Hakukoneista ei tavallisesti saa kovin tarkkaa tietoa siitä, millä periaatteilla hakutulokset järjestetään. Tätä perustellaan sillä, että sisällöntuottajat käyttäisivät tietoa hyväkseen saadakseen sivunsa hakutulostilan kärkeen. (Yhdyskuntasuunnittelun tutkimus- ja koulutuskeskus 2005).

Pääsääntöisesti hakukoneet kuitenkin määrittävät tietokantoihinsa indeksoitujen sivujen relevanssin hakusanoihin nähden seuraavien periaatteiden mukaan:

- suhteellinen frekvenssi eli hakusanan esiintymistiheys sivulla
- hakusanan esiintyminen sivun alussa, otsikkokentässä, url-osoitteessa, META-kentissä, tyylilliset korostukset (fontti, tummennus jne.)
- hakusanat esiintyvät useammin kuin jollain toisella sivulla
- hakusanojen läheisyys toisiinsa nähden
- muita periaatteita (edellisiä täydentäen):
 - PageRank (erityisesti Google)
 - hakutulosten käytön määrä ja linkityssuhteet (esim. Alexa)
 - käsiteperustainen haku eli semanttisten suhteiden tunnistaminen
 - automaattinen fraasin tunnistaminen sekä hakusanojen syntaksin ja semantiikan tunnistaminen
 - esimerkinmukainen haku
 - tulosten ns. klusterointi (erityisesti Vivisimo)
 - ostettu relevanssi
 - kysymys-vastaus -parien tietämystietokanta (Ask)
 - suodattimet - hakutulosten suodattaminen etukäteen
 - linkin esiintyminen myös hakukoneen aihehakemistossa lisää relevanssia haussa

(Kirjastot.fi 1995 - 2006).

Jonkinlainen käsitys siitä, millaista tietoa Internetistä haetaan, saadaan silmäilemällä hakukoneiden suosituimpia hakusanoja. (Haasio 2002, 12). Vertasin maanantaina 18.09.2006 Eniro-hakukoneen (entinen Evreka) top 10 -listaa Haasion

(2002, 12-13) esittelemään saman hakukoneen top 10 -listaan päivältä 01.06.2002, ja tulos yllätti minut jossain määrin. Siinä missä 01.06.2002 suosituimpia hakusanoja olivat muun muassa "ilmaisia", "woman" ja "kuvat", oli lista 18.09.2006 klo 21 tämän näköinen:

1. herttoniemi
2. helsinki
3. levy
4. kokoelmat
5. fiat
6. eri
7. yhtiökokouksen
8. punto
9. midi
10. tdi

Vaikka tuohonkin joukkoon mahtuu pari hakusanaa, jotka voidaan tulkita viihteellisiksi, on lista hyvin eri näköinen kuin vuoden 2002 lista. Luonnollisesti lista muuttuu usein, ja todennäköisesti siihen vaikuttavat paljon mm. vuodenaika, viikonpäivä sekä kellonaika: katsoin Eniron top 10 -listaa uudelleen lauantaina 23.09.2006 klo 21, ja silloin se näytti jo aivan toisenlaiselta, joskin viihteellisten hakusanojen osuus oli silloinkin huomattavan pieni.

Hakusanalistan muutos kertonee omalta osaltaan Internetin arkipäiväistymisestä. Internet-liittymä on nykyisin jo useimmissa kodeissa: laajakaistaliittymiä oli vuonna 2004 Suomessa tuhatta asukasta kohti jopa 779 kappaletta, kun niiden määrä vuonna 2002 oli vain 237 (Tilastokeskus 2005). Internet ei ole enää erikoisuus, vaan sen käytöstä on tullut rutiininomaista samalla tavalla kuin vaikka matkapuhelimen käytöstä (Korpela 2005; Savolainen 2003). Tämä näkyy myös hakukoneiden käytössä: Internetistä haetaan nykyisin aivan arkipäiväisiä asioita juna-aikatauluista ruokaresepteihin ja huonekasvien hoito-ohjeista rengasliikkeiden yhteystietoihin.

Eniro-hakukoneen tuoreimman top 10 -listan voi käydä katsomassa osoitteessa http://www.eniro.fi/toplistat/?partnerid=s24suomi24/bubble_day.shtml.

Koska eri hakukoneet tuottavat usein samoilla hakusanoilla erilaisia tuloksia, kannattaa haku usein tehdä kahta tai useampaa hakukonetta käyttäen. Joskus vastaavan haun tekeminen toisella hakukoneella - tai kokonaan toista tietolähdettä käyttämällä - tuottaa paremman tuloksen. Yksi onnistuneen tiedonhaun strategia on tehdä monta peräkkäistä hakua samasta aiheesta tarkentaen hakulauseketta koko ajan. (Haasio 2003, 17-18; Haasio 2005b, 74, 80; Kirjastot.fi 1995-2006).

2.2 Metahakukoneet

Tavallisten hakukoneiden, joihin tämä tutkielma on keskittynyt, lisäksi on olemassa monihakuun perustuvia hakukoneita eli ns. **metahakukoneita**. Siinä missä perinteiset hakukoneet hakevat tietoa omista tietokannoistaan, metahakukoneet hyödyntävät muiden hakukoneitten keräämiä tietoja. Käytännössä ne tekevät haun monesta hakukoneesta yhden käyttöliittymän kautta ja valikoivat hakutulokseen eri hakukoneiden tuottamat parhaat osumat. Metahakukoneita ovat mm. Dogpile (<http://www.dogpile.com>), Ixquick (suomenkielinen versio löytyy osoitteesta <http://www.ixquick.com/suo/>), Mamma (<http://www.mamma.com>), Metasearch (<http://www.metasearch.com>), Search (<http://www.search.com>) sekä Vivisimo (<http://www.vivisimo.com>). (Haasio 2003, 70; Haasio 2005b, 80).

Metahakukoneet ovat hyödyllisiä, kun halutaan nopeasti selkeä ja rajattu tulosjoukko. Niiden avulla vältetään tuhansien sivujen saaminen hakutulokseksi. Esimerkiksi Vivisimo on klusteroiva metahakukone: se jaottelee hakutulokset aiheryhmittäin, jolloin hakutulosten läpikäyminen on vaivattomampaa. (Haasio 2005b, 81).

Kokeilumielessä etsin Vivisimolla tietoa yhdysvaltalaisesta feminististä ja kirjailijasta nimeltä Valerie Solanas, ja tulokset oli todellakin ryhmitelty kauniisti aiheen mukaan: löytyi elämäkertatietoja, valokuvia, Solanasin kuuluisia lausahduksia, tietoa hänen kirjoistaan ja paljon muuta.

Vaikka metahakukoneet ovat käteviä ja vaikka niitä kehitetään kaiken aikaa, ne eivät ole läheskään niin suosittuja kuin perinteiset hakukoneet. Yksi syy on se, etteivät metahakukoneet tavallisesti hyödynnä isoimpia hakukoneita, vaan hakevat tietoa pienemmistä hakukoneista. Metahakukoneita käytettäessä on myös muistettava, että niihin sisältyvissä yksittäisissä hakuohjelmissa on mahdollisesti erilaiset hakusäännöt, eikä metahaussa voida aina käyttää hyväksi yksittäisen hakuohjelman kaikkia ominaisuuksia. (Haasio 2003, 70-71; Kirjastot.fi 1995-2006).

3 HAUN MUOTOILU

Koska hakutulosten automaattisesta järjestämisperiaatteesta ei saa tarkkaa tietoa, on järjestykseen yritettävä vaikuttaa haun muotoilulla. Se, miten hakua voidaan muotoilla, riippuu paljolti käytettävästä hakukoneesta. Internetin hakukoneet ovat suuressa määrin samankaltaisia, joten niistä löytyy paljon samoja ominaisuuksia. (YTK 2005).

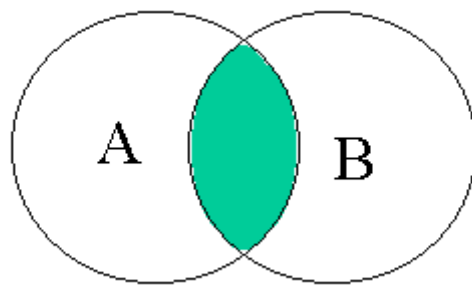
Hakukoneiden avaussivulla on oletuksena *perushaku* (*simple search*), jossa hakusana/-sanat kirjoitetaan hakulaatikkoon sellaisenaan ilman muita merkintöjä. Perushaussa haun muotoilumahdollisuudet ovat vähäiset. Haun rajaus tapahtuu tavallisesti hakukoneen *tarkennetussa haussa* (*advanced search, expert search, power search*), joka löytyy yleensä linkin takaa, tai kirjoittamalla rajauksen mahdollistava komento hakulaatikkoon. (Avoin yliopisto, viitattu 15.10.2006; Haasio 2005b, 74; YTK 2005).

3.1 Boolean logiikka

Enemmistö Internetin hakukoneista perustuu Boolean logiikan hyödyntämiseen. Vaikka minkään hakukoneen peruskäyttö ei edellytäkään sen hallintaa, on sen

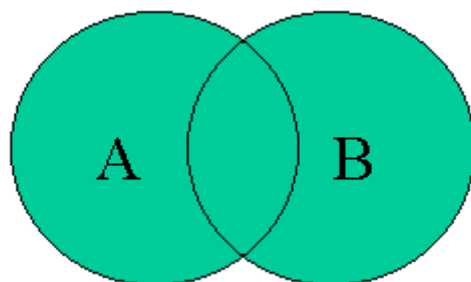
peruseriaatteet hyvä ymmärtää. Monipuolisten ja kehittyneempien hakujen onnistuminen vaatii Boolean logiikan sisäistämistä. (Haasio 2003, 25; Haasio 2005b, 78).

Boolean logiikka on algebrallinen järjestelmä, *täydellisen täsmäytyksen (exact match)* menetelmä, joka on saanut nimensä kehittäjältään, englantilaiselta matemaatikolta George Boolelta (1815 - 1864). Se pohjautuu kolmen *loogisen operaattorin* käyttöön hakua suoritettaessa: operaattori AND on *leikkaus (intersection)*, OR *yhdiste (union)* ja NOT *erotus (difference)*. Asiaa havainnollistetaan usein seuraavien nk. Vennin kaavioiden avulla. (Alaterä & Halttunen 2002, 38-39; Haasio 2005a, 14; Haasio 2005b, 78; Korhonen & Likitalo 2005, 82).



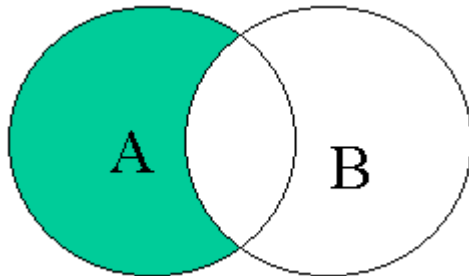
Kuva 1 Leikkaus (AND). (Alaterä & Halttunen 2001-2004).

Joukkojen A ja B leikkaukseen kuuluvat ne alkio, jotka kuuluvat *sekä* joukkoon A *että* joukkoon B. (Haasio 2005a, 15).



Kuva 2 Yhdiste (OR). (Alaterä & Halttunen 2001-2004).

Joukkojen A ja B yhdisteeseen kuuluvat kaikki ne alkiot, jotka kuuluvat *jompaan-kumpaan tai molempiin* joukkoihin. (Haasio 2005a, 15).



Kuva 3 Erotus (NOT). (Alaterä & Halttunen 2001-2004).

Joukkojen A ja B erotukseen kuuluvat ne alkiot joukosta A, jotka *eivät* kuulu joukkoon B. (Haasio 2005a, 15).

Käytännössä kyse on siitä, että jos on esimerkiksi kiinnostunut amerikanraudoista ja/tai moottoripyöristä, hakulause voidaan muotoilla kolmella eri tavalla. Jos haku muotoillaan

amerikanraudat AND moottoripyörät

tulokseksi saadaan dokumentit, joissa esiintyvät *molemmat* hakusanat. Tulosten määrä vähenee helposti, sillä kaikkien AND-operaattorilla yhdistettyjen sanojen edellytetään sisältyvän hakutuloksiin. (Haasio 2005b, 79). Äskeinen esimerkki voitaisiin kirjoittaa muillakin tavoilla:

- amerikanraudat moottoripyörät (useimmat hakukoneet lisäävät automaattisesti AND-sanan kyselyn sanojen väliin)
- +amerikanraudat +moottoripyörät
- amerikanraudat+moottoripyörät
- amerikanraudat & moottoripyörät

(Husu, Tarkoma & Vuorijärvi 2001, 21; Vanhala-Nurmi, viitattu 19.10.2006).

Jos vastaava haku muotoillaan käyttäen OR-operaattoria, se näyttää tältä:

amerikanraudat OR moottoripyörät (tai amerikanraudat | moottoripyörät)

Haun tulokseksi saadaan dokumentit, joissa esiintyy *ainakin toinen* hakusanoista. (Haasio 2005b, 79; Vanhala-Nurmi, viitattu 19.10.2006). Siinä missä AND-operaattori supistaa hakua, OR-operaattori laajentaa sitä.

NOT-operaattorin avulla voidaan sulkea tietyt hakusanat pois haun tuloksesta. Jos haluamme tietoa *ainoastaan* amerikanraudoista, emme moottoripyöristä, haku voidaan muotoilla näin:

amerikanraudat AND NOT moottoripyörät (tai amerikanraudat -moottoripyörät)

Tällä tavalla yksikään dokumentti, jossa esiintyy sana "moottoripyörät", ei esiinny haun tuloksissa. NOT-operaattorin käytössä kannattaa olla varovainen, koska sitä käyttämällä voi menettää hyviä viitteitä. (Haasio 2005b, 79; Heikkinen, Agander, Ijäs & Laitinen 2001, 31; Vanhala-Nurmi, viitattu 19.10.2006). NOT-operaattorin kanssa on syytä käyttää AND-sanaa, sillä hakukoneet eivät välttämättä ymmärrä NOT-sanaa, jos se esiintyy yksinään.

Jotkut hakukoneet eivät tunnista kaikkia loogisia operaattoreita: esimerkiksi Google ei tunnista NOT-operaattoria, mutta Yahoo! ja AltaVista tunnistavat. Jos siis AltaVistassa tehdään haku

opiskelija NOT tietojenkäsittely

tulee vastaava haku tehdä Googlessa muodossa

opiskelija -tietojenkäsittely (Hänninen 2005, 6).

3.2 Läheisyysoperaattorit

Useissa hakukoneissa voidaan käyttää myös *läheisyysoperaattoria* (*adjacency operator, proximity operator*) NEAR. Läheisyysoperaattorilla on mahdollista määritellä, miten kaukana toisistaan hakusanat saavat sijaita samalla www-sivulla. (Alaterä & Halttunen 2002, 43; Haasio 2003, 28; Haasio 2005a, 17)

NEAR-operaattorilla voidaan myös varmistaa, että haettavat termit liittyvät asiayhteydeltään toisiinsa. Esimerkiksi työt NEAR pojat -haku CompletePlanetissa varmistaa sen, että hakusanat ovat korkeintaan kymmenen sanan päässä toisistaan. Tällöin ne ovat tavallisesti samassa virkkeessä tai peräkkäisissä lauseissa. (Haasio 2003, 28).

NOT NEAR -operaattorin, joka on mahdollinen joissakin hakukoneissa, avulla voidaan määritellä, ettei jokin termi saa esiintyä toisen termin läheisyydessä. Esimerkiksi CompletePlanetissa NOT NEAR -komento tarkoittaa, että kahden hakusanan välisen etäisyyden täytyy olla yli kymmenen sanaa (Haasio 2005a, 17; Sarajärvi 2004, 6).

Jos hakulausekkeessa on useampia hakusanoja, jotka yhdistetään toisiinsa Boolean operaattoreilla sekä läheisyysoperaattoreilla, operaatiot suoritetaan tässä järjestyksessä: (sulut), NEAR, NOT, AND, OR. Jos tehdään esimerkiksi hakulauseke

joulu AND piparkakut OR joulutortut

hakutulokseksi saadaan dokumentit, joissa sanat joulu ja piparkakut esiintyvät samassa dokumentissa tai dokumentit, joissa esiintyy sana joulutortut. (Haasio 2005b, 79; Tampereen yliopiston kirjasto, viitattu 19.10.2006).

Jos taas tehdään hakulauseke

joulu AND (piparkakut OR joulutortut)

hakutulokseksi saadaan dokumentit, joissa esiintyvät sanat joulu ja jompikumpi sanoista piparkakut tai joulutortut. (Haasio 2005b, 79).

Eräissä tapauksissa voidaan käyttää myös operaattoreita AFTER ja BEFORE. BEFORE-operaattorin avulla haku voidaan muotoilla siten, että haluttu hakusana esiintyy korkeintaan kymmenen sanan päässä ennen toista hakusanaa. AFTER-operaattori mahdollistaa hakulausekkeen muotoilemisen niin, että toinen haluttuista hakusanoista esiintyy korkeintaan kymmenen sanan päässä toisen termin jälkeen. Näiden lisäksi eräissä hakukoneissa on mahdollista käyttää NOT BEFORE ja NOT AFTER -operaattoreita. Niillä voidaan kieltää jonkin sanan esiintyminen vaihtoehtoisesti vähintään kymmenen sanaa ennen tai jälkeen suhteessa toiseen hakusanaan. (Haasio 2005a, 18; Sarajärvi 2004, 6).

3.3 Fraasihaku

Yksi tehokkaan Internet-tiedonhaun keskeisimmistä asioista on *fraasihaun* (*exact phrase, phrase search*) osaaminen. Jos halutaan, että jokin tietty, monesta sanasta koostuva merkkijono esiintyy hakutuloksissa täsmälleen tietyssä muodossa, on fraasihaku hyvä olemassa. Sillä pakotetaan merkit ja välimerkit esiintymään peräkkäin ja juuri siinä järjestyksessä kuin ne on hakulausekkeeseen kirjoitettu. Fraasi muodostetaan useimmiten laittamalla haluttujen sanojen ympärille lainausmerkit, siis esimerkiksi "cavalier king charlesin spanieli" tai "life is like a box of chocolates". Fraasia ilmaiseva merkki vaihtelee hakukoneittain; joissakin hakukoneissa fraasi merkitään näin: 'life is like a box of chocolates'. (Haasio 2003, 29; Rautajoki 2005). Fraasihaku voidaan yhdistää muihin hakueroihin.

Tietyt sanat ovat hakukoneissa ns. *stop words* -sanoja, eli hyvin tavalliset sanat ja merkit (artikkelit a, an ja the, and, or yms.) jäävät hakulausekkeesta huomioimatta automaattisesti, koska niistä on yleensä enemmän haittaa kuin hyötyä: ne harvoin auttavat haun rajoituksessa, ja ne voivat hidastaa hakua huomattavasti. Jos halutaan sisällyttää tällainen sana hakuun, tulee sanan sisältyä fraasiin, ja hakulauseke merkitään fraasiksi. (Rautajoki 2005).

Selvennetään asiaa parilla esimerkillä:

- Etsittäessä jonkin henkilön nimeä se laitetaan lainausmerkkien sisään, esim. "Juha Veijonen". Nyt hakukone hakee vain ne dokumentit, joissa esiintyy fraasi Juha Veijonen. Mikäli nimi Juha Veijonen kirjoitettaisiin hakukoneeseen ilman lainausmerkkejä, kiinnostavien ja relevanttien sivustojen lisäksi olisi löytynyt myös paljon hakusanojen kannalta turhia sivustoja, esimerkiksi erään mainostoimiston, jossa ovat töissä muun muassa Juha Salaja ja Jimi Veijonen, kotisivut.
- Jos haetaan tietoa elektronisesta viittaamisesta, voidaan hakuterminä käyttää fraasia "elektroninen viittaaminen". Kun sanat elektroninen ja viittaaminen laitetaan lainausmerkkien sisään, hakukone etsii ainoastaan kyseistä fraasia, ei erillisiä elektroninen ja viittaaminen -sanoja, jotka eivät välttämättä ole asiayhteydessä toisiinsa.
(Haasio 2005a, 19-20).

Tässä on vielä esimerkkejä tilanteista, joissa fraasihausta on hyötyä:

- henkilönimien haku: "Sherilyn Fenn"
 - sanonnan tms. haku: "ei omena kauas puusta putoa"
 - lyriikat tms.: "sinä ja minä liikennevaloissa"
 - organisaatioiden nimet yms.: "Satakunnan ammattikorkeakoulun kehittäminen ja palvelukeskus"
- (Haasio 2005a, 19-20).

Huomattavaa on, että esimerkiksi fraasihaut "02 620 3370" sekä "02-620 3370" antavat Googlessa saman hakutuloksen. Tämä johtuu siitä, että Googlessa jotkut merkit toimivat lauseen liittiminä. Google tunnistaa mm. väliviivat, yhtäsuuruusmerkit sekä heittomerkit lauseen liittimiksi (Google 2002).

3.4 Katkaisu

Sanojen oikeaoppinen *katkaisu* ja merkkien korvaaminen (*masking*, *truncation*, *wildcard*) on yksi onnistuneen haun edellytyksistä. Hakukoneet etsivät tietokannoistaan merkkijonoja. Suomen kielessä on kuitenkin paljon sijamuotoja, ja jotta ne löydetäisiin, täytyy hakusana joskus katkaista. Katkaisun avulla on myös mahdollista välttää yhdyssanoihin ja sanan vartalossa tapahtuviin vokaalinvaihteluihin (esim. engl. tooth/teeth) liittyvät ongelmat. Katkaisumerkkejä voivat olla esimerkiksi ?, #, \$ tai *, joista viimeksi mainittu lienee käytetyin. Se, mikä katkaisumerkki missäkin järjestelmässä on käytössä, selviää hakukoneen käyttöohjeista. Kaikki hakukoneet eivät tue katkaisumerkkiä. Tästäkin kerrotaan hakukoneiden ohjeissa. (Alaterä & Halttunen 2002, 43; Haasio 2003, 28; Haasio 2005a, 18; Lappeenrannan kaupunginkirjasto 2005).

Sanan katkaisua kannattaa miettiä huolellisesti, ja tarpeetonta sanan katkaisua on syytä välttää, sillä joskus siitä voi olla enemmän haittaa kuin hyötyä. (Haasio 2003, 28; Haasio 2005a, 18). Valaistaan asiaa muutaman esimerkin avulla:

- Jos halutaan tietoa kaneista, termi kani voidaan katkaista sanan vartalosta (kan*) hakua varten. Näin hakutuloksiin saadaan mukaan kaikki taivutusmuodot, monikkomuoto jne. Toisaalta mukaan saadaan muutkin kan-alkuiset termit, kuten kansainvälisyys, kantelut, kansalainen ja kannanotto.
- Jos termi suklaa halutaan katkaista (sukla*), hakutulokseksi saadaan mm. termit suklaiden, suklaakahvi ja suklaalajeja.
- Katkaisumerkin avulla on joissakin hakukoneissa mahdollista korvata myös sanan keskellä oleva merkki tai merkit esimerkiksi näin: foot, feet = f*t.
(Alaterä & Halttunen 2002, 43; Haasio 2005a, 19).

Joidenkin sanojen vartalo on liian lyhyt hakutermiksi. Esimerkiksi lapset-sanana vartalo on la, joka ei ole riittävän pitkä hakutermiksi. Tällöin on vain kokeiltava hakua eri muodoilla, ellei sanan nominatiivimuoto anna tarpeeksi hakutuloksia. (Heikkinen ym. 2001, 40).

3.5 Hakukentät

Useimmissa hakukoneissa on erilaisia mahdollisuuksia tarkentaa hakua kohdistamalla haku tiettyyn kenttään www-sivuilla. Haun rajausmahdollisuudet laitetaan tavallisesti valikkoihin valmiiksi, mutta ne voidaan myös kirjoittaa suoraan hakulaatikkoon. (YTK 2005).

Hakukenttä	Selite	Esimerkki
ANCHOR	hakee sivuja, joilla haettu sana on teksti-linkkinä	anchor:ireland hakee sivuja, joiden hyperlinkissä on sana ireland
APPLET	hakee dokumentteja, joissa on Java-applet	applet:minesweeper hakee sivuja, joilla on minesweeper-niminen applet
DOMAIN	hakee dokumentteja, jotka ovat ko. domainilta	domain:.se hakee ruotsalaisia .se-päätteisiä sivuja
FILETYPE	hakee dokumentteja, jotka ovat tietyssä tiedostoformaattissa	filetype:pdf hakee pdf-tiedostoja
HOST	hakee dokumentteja vain tietyltä palvelimelta	host:utu.fi hakee Turun yliopiston palvelimen sivuja
IMAGE	hakee kuvia, joiden nimessä on haettu sana	image:heinola hakee dokumentteja, joissa esiintyvän kuvan nimessä on sana heinola
INBODY	hakee sivuja, joiden body-elementissä on haettu sana	inbody:puppy hakee dokumentteja, joiden body-elementissä esiintyy sana puppy
INTITLE	hakee sivuja, joiden otsikossa on haettu sana	intitle:edita hakee dokumentteja, joiden otsikossa esiintyy sana edita
INURL	hakee sivuja, joiden url-osoitteessa on haettu sana	inurl:music hakee dokumentteja, joiden url-osoitteessa esiintyy sana music
LANGUAGE	hakee tietyn kielisiä sivuja	language:ru hakee venäjänkielisiä dokumentteja
LINK	hakee dokumentteja, joista on viitattu haettavaan www-osoitteeseen (Google)	link:www.vr.fi hakee sivuja, joilla on linkki VR:n sivuille
LINKDOMAIN	hakee dokumentteja, joista on viitattu haettavaan www-osoitteeseen (Yahoo!)	linkdomain:www.samk.fi hakee sivuja, joilla on linkki SAMK:n sivuille
LINKFROMDOMAIN	hakee sivustoja, joille on linkki tietyltä sivustolta	linkfromdomain:samk.fi hakee dokumentteja, joille on linkki samk.fi-sivustolta
LOC/LOCATION	hakee tietyn maalaisia sivuja	location: ee tai loc:ee hakee virolaisia dokumentteja
SITE	hakee dokumentteja vain tietyltä palvelimelta	site:lamk.fi hakee Lahden ammattikorkeakoulun palvelimen sivuja
TEXT	hakee sivut, joissa hakusana esiintyy tekstissä	text:kennel etsii sivuja, joiden tekstissä on sana kennel
TITLE	hakee dokumentteja, joiden otsikossa eli selaimen otsikkopalkissa näkyvässä tekstissä esiintyy hakusana	title:edita hakee dokumentteja, joiden otsikossa mainitaan sana edita
URL	selvittää, onko haetun url-osoitteen omaava sivusto hakukoneen tietokannassa	url:epes.net hakee dokumentteja, joiden url-osoite on epes.net

Taulukko 1 Hakukentät (*field search*). (Fonecta 2006; Haasio 2005b, 74-75; Husu ym. 2001, 21; Notess 2006; Tampereen kaupunginkirjasto, viitattu 22.10.2006; Vanhala-Nurmi, viitattu 19.10.2006; YTK 2005).

3.6 Muita hakumahdollisuuksia

Hakukoneilla tehtävässä tiedonhaussa on mahdollista käyttää edellä mainittujen operaattorien lisäksi mm. *aika-* ja *kielirajausta*. Aikarajaus tehdään sen mukaan, milloin sivu on viimeksi päivitetty. (Haasio 2005b, 74; Vanhala-Nurmi, viitattu 19.10.2006). Kielirajausta (*search by language*) tehtäessä kannattaa muistaa, että vain hyvin pieni osa Internetin sisällöstä on suomenkielistä - ja monessa tapauksessa on haettava nimenomaan englanninkielistä aineistoa, jos haluaa ajantasaista tietoa. Lisäksi on hyvä muistaa, että keskeisimmät käsitteet kannattaa kirjoittaa hakulausekkeen alkuun. (Haasio 2005b, 74-75; Husu ym. 2001, 21; YTK 2005).

Hakukoneita käytettäessä ei pidä unohtaa, ettei hakukone - niin kuin ei tietokonekaan - ole viisas, vaan se on hyvin mekaaninen. Internetin perinteiset hakukoneet eivät hae asioita, vaan sivustoilla esiintyviä sanoja, tiettyjä merkkijonoja. Kannattaa siis miettiä, onko hakutermeillä synonyymejä (= yhtä asiaa voidaan kutsua monella eri sanalla) tai homonyymejä (= yksi sana tarkoittaa kahta eri asiaa, esimerkiksi hiiri (eläin) ja hiiri (tietokoneen osoitinlaite)). On hyvä yrittää asettua sisällöntuottajan asemaan ja miettiä, mitä sanoja hän on käyttänyt. Jos halutaan tietoa esimerkiksi kuvanlukijoista, tulee huomioida se, että kuvanlukijaa kutsutaan useimmiten nimellä skanneri. Vastaavasti, jos haetaan sanalla "kanalinnut", haku ei välttämättä tuo esiin sivuja, joilla mainitaan sen alakäsitteitä, kuten "fasaani" tai "metso". Kotimainen Webinfo on hakukoneiden joukossa poikkeus siinä suhteessa, että se ymmärtää www-sivuilla olevien suomenkielisten sanojen kaikki taivutusmuodot. Siksi esimerkiksi "kesämökki" ja "kesämökit" tuottavat samat hakutulokset. Webinfo etsii tietoa ensisijaisesti suomalaisten yritysten ja yhdistysten sekä julkishallinnon sivuilta, eikä kilpaile yleishakukoneiden kanssa, vaan keskittyy hyötypalveluiden ja tuotteiden löytämiseen. (Haasio 2003, 17-19; Haasio 2005b, 74; Heikkinen ym. 2001, 40; Info Center Finland ICF Oy 1997-2006; Karvonen 2006).

Tässä tutkimuksessa käsiteltävistä hakukoneista Googlen indeksissä on ns. latenttia semantiikkaa. *Latentissa semanttisessa indeksoinnissa (LSI)* on kyse puhtaasta matemaattisesta lineaarialgebraan perustuvasta menetelmästä, jossa ei pyritä ymmärtämään sanojen merkitystä, vaan hakusanan synonyymien esiintymistiheyksiä

dokumenteissa. Näin Googlen haku voi palauttaa hakutulostan alkuun viitteitä dokumentteihin, joilla ei edes mainita hakusanaa. Osittain tämä johtuu linkkien ankkuriteksteistä, mutta englanninkielisissä hauissa latenttisella semanttisella indeksoinnilla on voimakas vaikutus. Suomenkielisten sanojen synonyymejä Google ei ainakaan vielä toistaiseksi indeksoi. (Poutiainen 2005).

Tein testihaun termeillä ~cola ja cola. Ensiksi mainittu palautti Pepsin ja Coca-Colan hakutulostan kärkeen, kun taas jälkimmäinen antoi aivan erinäköisen tuloksen.

4 INDEKSOINTI

Aniway (1999 - 2005) määrittelee indeksoinnin www-sivun lisäämiseksi hakukoneen tietokantaan. Indeksi on eräänlainen kopio tietojen ja tiedostojen sisällystöstä, metadatasta yms. Indeksointi mahdollistaa hakutuloksen esittämisen sekunnin murto-osassa, sillä hakukone ei hae tiedostoja tiedonhakijalle suoraan Internetistä reaaliaikaisesti, vaan pelkästään indeksistä. (Systems Garden, viitattu 21.10.2006). Indeksi siis on tavallaan, niin kuin sen nimestäkin jo voi päätellä, hakukoneen tietokannan sisällysluettelo.

Hakukoneiden tavoitteena on indeksoida koko World Wide Web. Eri hakukoneiden www-sivuista muodostamien indeksien koko vaihtelee suuresti: esimerkiksi markkinajohtaja Googlen indeksi sisälsi syyskuussa 2005 noin kahdeksan miljardia sivua, ja kilpailija Yahoo!':n indeksi yhtiön oman ilmoituksen mukaan peräti 19,2 miljardia sivua (Karvonen 2005). Vaikka luvut kuulostavat suurilta, ei minkään hakukoneen indeksi käsitä kuin pienen osan kaikesta www:n materiaalista.

Jokaisella hakukoneella on oma tapansa suorittaa indeksointi, mikä on syy siihen, että eri hakukoneet tuottavat erilaisia hakutuloksia, vaikka niillä haettaisiin samaa

asiaa samoilla hakusanoilla. Yleisimmin sivuista indeksoidaan url-osoite, otsikkokentät, sisältö sekä sivuilla olevat linkit. Eri hakukoneet tulkitsevat kuitenkin sisällön indeksoimisen eri tavoin. Se saattaa tarkoittaa tekstinäytettä tärkeimmistä termeistä tai ensimmäistä 20 riviä. Hakukone voi indeksoida sivun koko tekstin, kaikkein pienimmätkin sanat (esim. AltaVista) tai jättää kaikkein yleisimmät sanat pois (esim. Lycos). Erittäin yleisiä sanoja ei indeksoida, vaan niitä varten on ns. sulkulista. Jäljelle jääneistä sanoista muodostetaan käännetty tietokanta, jossa jokainen termi viittaa sellaisiin www-sivuihin, jotka sisältävät kyseisen sanan. (Haasio 2003, 14; Hänninen 2005, 5; Tampereen kaupunginkirjasto, viitattu 22.10.2006).

Yleensä dokumenteista indeksoidaan vain relevanteimmat osat eli niin sanotut *avainsanat*. Tämä menetelmä lisää viitteiden tarkkuutta verrattuna siihen, että dokumenttien yleisimmätkin sanat huomioitaisiin. Mikäli sivuilla on ääni- tai kuvatiedostoja, suoritetaan niiden indeksointi välillisesti niiden liitännäistekstien eli niin sanottujen alt-tekstien sekä kuvien tiedostonimien avulla. (Hänninen 2005, 5; Kirjastot.fi 1995-2006; Tampereen kaupunginkirjasto, viitattu 22.10.2006).

Muutamit hakukoneet liittävät myös metadatan indeksoitaviin dokumentteihin, toisin sanoen dokumentista voidaan antaa jotakin ylimääräistä tietoa. Tämä metadata voi olla kyseisen dokumentin tekijän nimi tai muuta tietoa hänestä, lyhyt kuvaus dokumentista ja/tai luettelo käytetyistä avainsanoista. Dokumentin tekijä antaa itse metadatan sivunsa HEAD-osiossa erityisten META-koodien avulla. (Hänninen 2005, 5; Tampereen kaupunginkirjasto, viitattu 22.10.2006).

Keskeisin indeksoitava tyyppi on www-sivu, mutta myös tiedostoarkistoja indeksoidaan, ja useimmat hakukoneet indeksoivat myös uutisryhmiä. Tietokannoissa on paljon tutkimattomia viitteitä, koska tietoa tulee jatkuvasti lisää, ja hakuroboteilla on täysi työ pysyä perässä. (Tampereen kaupunginkirjasto, viitattu 22.10.2006). Suuresta työmäärästä kertoo se, että julkaistavat www-sivut ilmestyvät hakukoneiden tietokantaan yleensä muutaman viikon viiveellä.

Indeksoinnissa on omat ongelmansa. Hypertekstin indeksoiminen on hankalaa, ja kaiken Internetissä olevan tiedon indeksoiminen kokonaan on mahdotonta.

Www:ssä on valtava määrä erilaisia informaatiotyyppisiä, joiden yhdenmukainen indeksoiminen on erittäin haasteellista. Internet elää kaiken aikaa: uutta tietoa tulee jatkuvalla syötöllä samalla kun entistä tietoa häviää. Myös url-osoitteet muuttuvat usein. Jatkuva päivittäminen on tärkeää, mutta vaikeaa. Oman ongelmansa muodostaa myös se, että sisällöntuottajien osaamisen - ja sen myötä heidän tekemiensä www-sivujen - taso vaihtelee huomattavasti, ja jos HTML-kielellä tehdyissä sivuissa on runsaasti virheitä, vaikeuttaa se indeksointia ja sen myötä hakujen tekemistä. (Hänninen 2005, 5; Tampereen kaupunginkirjasto, viitattu 22.10.2006).

Ehkä merkittävimmät ongelmat liittyvät kuitenkin siihen, että erilaiset hakukoneet ja niiden käyttämät systeemit voivat kuormittaa verkkoa huomattavasti. Kustannus-hyötysuhde ei aina ole hyvä: Internetistä kerätään kokonaisia dokumentteja, joista indeksoidaan vain muutama elementti. Vaikka tiedostot indeksoidaisiin kokonaan, eivät hakutulokset välttämättä parantuisi, koska sanahauissa sanat irrotetaan asiayhteyksistään miettimättä niiden merkitystä. (Hänninen 2005, 5-6; Tampereen kaupunginkirjasto, viitattu 22.10.2006).

4.1 Hakurobotit

Hakukoneita kuullaan toisinaan kutsuttavan hakuroboteiksi, vaikka hakurobotti on itse asiassa vain yksi - joskin hyvin keskeinen - osa hakukonetta. Hakurobotit (*bots, crawlers, robots, spiders*) ovat ohjelmia, jotka automaattisesti tutkivat www:n hypertekstirakennetta yrittäen kerätä tietoja mahdollisimman monista eri www-sivuista hyödyntämällä verkossa olemassa yhteyskäytäntöjä eli protokollia. Hakurobotit ottavat yhteyden eri palvelimiin ja lataavat niiltä dokumentteja tutkittavaksi. Hakurobotit seuraavat www-sivuilta löytyviä linkkiviitteitä indeksoidakseen niiden osoittamat sivut - eli jos johonkin sivuun ei viitata, ei hakurobotti välttämättä sitä löydä, kun taas eniten viitatus sivut tulevat varmimmin indeksoiduiksi. (Hänninen 2005, 4; Tampereen kaupunginkirjasto, viitattu 22.10.2006).

Ensimmäiset hakurobotit kulkivat syvyys ensin -periaatteella (*depth first*) eli tekivät kyselyjä yhdelle palvelimelle niin kauan, että saivat sen kaikki tiedostot kerättyä, ja siirtyivät vasta sitten seuraavalle palvelimelle. Tämä aiheutti sen, että palvelin kuormittui suunnattomasti, sen toiminta hidastui ja pahimmassa tapauksessa se saattoi kaatua, mistä luonnollisesti seurasi harmia käyttäjille. Hyvää tässä periaatteessa on se, että sillä voidaan koota kattavasti tietoa rajatusta palvelinjoukosta. Nykyisin käytössä on leveys ensin -periaate (*breadth first*) eli kyselyjä tehdään vuorotellen eri palvelimilta, eikä yhdellä palvelimella mennä kovin syvälle. Leveys ensin -periaatteella saadaan kerättyä näytteitä suuresta palvelinjoukosta, ja kyseinen menetelmä suosii sivuja, joihin viitataan oman palvelimen ulkopuolelta. (Hänninen 2005, 4-5; Tampereen kaupunginkirjasto, viitattu 22.10.2006). Huonoa tässä periaatteessa on se, että syvimmällä olevat dokumentit saattavat jäädä keräämättä.

Kesällä 1994 Internetin hakurobottien tekijät sopivat keskenään protokollasta, jolla web-sivustojen tekijät voisivat rajata sivustojensa näkyvyyttä hakuroboteille. Syntyi robotinsulkustandardi (*Robots Exclusion Standard*). "Standardi" on ainoastaan käytännön sopimus silloisten tekijöiden välillä: kukaan ei pakota hakurobotin ohjelmoijaa noudattamaan tai kunnioittamaan sitä. Käytännössä monet hakurobotit kuitenkin näin tekevät. (Heikniemi 2002; Hänninen 2005, 5; Tampereen kaupunginkirjasto, viitattu 22.10.2006).

Robotinsulkustandardi mahdollistaa sen, että palvelimien ylläpitäjät voivat halutessaan kieltää kaikkia tai joitakin hakurobotteja imuroimasta sen tiedostoja. Esto toteutetaan käytännössä niin, että hakurobotit pyytävät ensin palvelimelta tiedostoa robots.txt, johon palvelimen ylläpitäjä voi listata ne palvelimella olevat hakemistot ja tiedostot, joita he eivät tahdo indeksoida. robots.txt -tiedosto sijoitetaan www-palvelimen juureen, mikä rajoittaa tekniikan käytettävyyttä melkoisesti, sillä se edellyttää pääkäyttäjän tunnuksia. Suositeltavaa on, että palvelinten ylläpitäjät käyttäisivät robots.txt -tiedostoa, jotta hakurobotit eivät keräisi tietokantaansa sellaisia dokumentteja, jotka eivät ole yleisesti kiinnostavia. (Heikniemi 2002; Hänninen 2005, 5; Tampereen kaupunginkirjasto, viitattu 22.10.2006).

Tiedoston nimeämisessä tulee olla tarkkana: nimi on täsmälleen robots.txt, pienellä kirjoitettuna. Itse formaatti on yksinkertainen, ja tiedosto voidaankin luoda millä tahansa tekstieditorilla. Tiedosto robots.txt koostuu tietueista, joista jokainen kieltää jotain hakukonetta käsittelemästä jotain sivun osaa. Jokainen tietue koostuu User-agent- ja Disallow-määreistä, ja tietueiden välissä on aina vähintään yksi tyhjä rivi. Tiedostoon voi lisätä kommentteja liittämällä niiden eteen risuaitamerkin (#). (Heikniemi 2002).

User-agent-määre kertoo, mitä hakurobotteja tietueen sulkumäärietykset koskevat. Disallow-määre kertoo hakemistopolut, joista hakurobotti ei saa hakea tietoa. Molempia rivejä voi olla yhdessä tietueessa useampia. (Heikniemi 2002).

Havainnollistetaan asiaa muutamalla esimerkillä. Näin kielletään kaikkia hakurobotteja noutamasta dokumentteja:

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

Näin kielletään hakurobotteja vierailemasta temp- ja kuvat -hakemistoissa:

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

Näin taas kielletään Googlen hakurobottia indeksoimasta taulukot-hakemistoa:

```
User-agent: googlebot  
Disallow: /taulukot/
```

Jos halutaan sallia AltaVistan Scooter-hakurobotin vierailu, kirjoitetaan näin:

```
User-agent: scooter  
Disallow:
```

(Fonecta 2006; Heikniemi 2002; Robotstxt.org, viitattu 25.10.2006).

Disallow-otsake määrittelee suljettavan url-polun alkuosan. Jos siis suljettaisiin polku /kuvat, suljettaisiin myös /kuvat.HTML, mutta sulku /kuvat/ peittää ainoastaan kuvat-hakemiston alaiset tiedostot. Disallow-määreessä ei voi käyttää tähteä jokerimerkkinä, joten /temp/*HTML-tyyppiset rakenteet eivät toimi. (Heikniemi 2002).

Sivun HEAD-osaan sijoitettavalla ROBOTS-metatagilla voidaan myös ohjata hakurobottien toimintaa:

- INDEX kehottaa hakukonetta indeksoimaan sivun
- FOLLOW kehottaa hakukonetta seuraamaan sivulla olevia linkkejä
- ALL on INDEX- ja FOLLOW-tilojen yhdistelmä
- NOINDEX estää hakukonetta indeksoimasta yksittäistä sivua, mutta ei estä sivulla olevien linkkien seuraamista ja mahdollisesti niiden sisällön indeksoimista
- NOFOLLOW antaa indeksoida itse sivun, mutta estää seuraamista sillä olevia linkkejä
- NONE on NOINDEX- ja NOFOLLOW-tilojen yhdistelmä
(Heikniemi 2002; KK Mediat 2000-2006).

ROBOTS-metatagi upotetaan HTML-sivuun seuraavalla tavalla:

```
<HTML>
<HEAD>
<TITLE>Sivun otsikko</TITLE>
<META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">
</HEAD>
<BODY>
Sivun teksti ja muu sisältö
</BODY>
</HTML>
```

(Fonecta 2006; Heikniemi 2002).

ROBOTS-metatagilla ei määritelmänsä mukaan ole tarkoituksaan pystyä määrittelemään hakukoneiden käytöstä hakurobottikohtaisesti, kuten robots.txt -tiedoston tietueiden User-agent-kentällä. Tällainenkin ratkaisu on tosin nähty: esimerkiksi Google hyväksyy ROBOTS-tagin tilalla myös GOOGLBOT-nimisen metatagin. Lisäksi Google ymmärtää tagille myös NOARCHIVE-tilan, joka tarkoittaa sitä, ettei sivua tallenneta hakukoneen omaan välimuistiin. (Heikniemi 2002).

Hakurobottien toiminnassa on myös heikkouksia. Ne indeksoivat tavallaan sokeasti: ne eivät osaa tehdä valintoja eri dokumenttien välillä, koska ne eivät kykene päättämään, mikä on tärkeää tietoa ja mikä ei. Tämän takia niiden tietokantoja kuormittaa roska. Hakurobotit eivät myöskään välttämättä erota, mikä sivu on ns. pääsivu. Amerikkalainen aineisto on Internetissä ylikorostunutta, ja hakukoneetkin priorisoivat amerikkalaisia palveluja, erityisesti .com-päätteisiä kaupallisia osoitteita. Tätä voidaan pitää hakukoneiden kaupallistumisen huonona puoleena. (Haasio 2003, 19; Hänninen 2005, 5; Tampereen kaupunginkirjasto, viitattu 22.10.2006).

Hakurobottien tietokantojen päivitys on erittäin hidasta. Arviolta 15 prosenttia hakukoneiden tietokannoissa olevista linkeistä on niin sanotusti kuolleita, mikä tarkoittaa sitä, etteivät hakutuloksissa saadut viitteet johda mihinkään. Tietokannoissa on kerättynä myös paljon tutkimattomia ja indeksoimattomia dokumentteja sekä duplikaatteja eli moninkertaisia linkkejä samoille sivuille. Tämän lisäksi huonosti suunniteltu robotti kuormittaa verkkoa ja palvelimia huomattavasti esimerkiksi tekemällä kyselyjä palvelimelle liian usein tai ruuhkaisina aikoina. (Hänninen 2005, 5; Tampereen kaupunginkirjasto, viitattu 22.10.2006).

Ehkä suurin puute hakukoneissa on se, että suuri osa Internetin sisällöstä ei löydy niiden avulla lainkaan: on arvioitu, että jopa puolet kaikesta siitä olisi täydellisesti tai osittain hakukoneiden tavoittamattomissa. Tätä osaa Internetistä kutsutaan *näkymättömäksi*, *pimeäksi* tai *syväksi webiksi* (*deep web*, *invisible net*, *invisible web*). Syvän webin informaation määrää on arvioitu yli 500-kertaiseksi pinta-

Internetiin verrattuna. (Haasio 2005a, 9; Haasio 2005b, 93; Nikkilä 2001, 175; Sarajärvi 2004, 15; Tampereen ammattiopiston kirjasto- ja tietopalvelut, viitattu 23.10.2006).

Sivujen tavoittamista hakukoneilla hankaloittavat muun muassa seuraavat seikat:

- Sivuille ei ole linkkejä kyseisen kokonaisuuden ulkopuolelta.
 - Sivun tekijä on ehkä estänyt sivujensa keräämisen hakukoneisiin.
 - Sivulla käytetään sellaisia tiedostomuotoja, joita hakukoneet eivät osaa käsitellä.
 - Palvelun käyttö edellyttää rekisteröitymistä.
 - Sivusto on audiovisuaalisesti runsas, minkä takia haun kohteena toimivaa tekstuaalista informaatiota on vain vähän.
 - Informaatio päivittyy nopeasti.
 - Aineisto julkaistaan tietokantana. Hakukoneet löytävät osan tietokantojen sisältämästä informaatiosta, mutta eivät läheskään kaikkea. Suurin osa tietokantapohjaisista julkaisuista jää hakukoneiden tavoittamattomiin johtuen menetelmästä, jolla tietokanta on rakennettu.
 - Aineisto on dynaamisesti tuotettu, esimerkiksi personoitu palvelu.
- (Alaterä & Halttunen 2002, 112-113; Haasio 2003, 101; Haasio 2005a, 9; Haasio 2005b, 93).

Verkossa on kuitenkin sivustoja, jotka toimivat syvän webin lähteinä. Myös aihehakemistojen avulla voidaan kartoittaa syvään webiin kuuluvia lähteitä, kuten tietokantoja. CompletePlanet (<http://www.completeplanet.com>) on juuri tällainen aihehakemisto. Siitä löytyy aihealueittain luokiteltuna yli 70 000 tietokantaa ja erikoishakukonetta. CompletePlanetissa on myös hyvä hakukone, jonka avulla on mahdollista muotoilla hakulauseke sofistikoitusti sekä kohdentaa haku tarkasti tiettyyn aihepiiriin. CompletePlanet jakautuu eri aihealueisiin, kuten kirjallisuus, musiikki, politiikka jne. Haku voidaan kohdentaa joko kaikkiin aihealueisiin tai johonkin tiettyyn aihealueeseen. CompletePlanet tukee Boolean logiikan lisäksi läheisyysoperaattoreita. (Haasio 2005b, 94).

Pohjoismainen NordGuide (<http://www.otilib.fi/nordinfo/nordguide/>) on yksi tapa kartoittaa skandinaavisia tietokantoja. Palvelu on kattava, ja se tuntee hyvin muitakin kuin online-tietokantoja, esimerkiksi cd-romit. NordGuidea ei enää päivitetä, mutta käytännössä kaikki sen sisältämät linkit ovat yhä relevantteja, minkä takia sitä kannattaa edelleen käyttää. Lisäksi NordGuidessa on hyvä lisähaku-toiminto, jossa haku on mahdollista rajata muun muassa tietokannan tyyppin (esimerkiksi numeerinen tietokanta, kokotekstitietokanta, kuvatietokanta jne.), kielen, aiheen yms. seikkojen mukaan. Kotimaisia tietokantoja on koottu Kirjastot.fi -portaaliin (<http://www.kirjastot.fi>) sekä Suomen tieteellisten kirjastojen etusivulle (<http://www.lib.helsinki.fi/tilke/>). (Alaterä & Halttunen 2002, 113; Haasio 2005b, 95).

4.2 Metadata

4.2.1 Yleistä metadatatista

Metadata (kuvaustieto, liitännäistieto, metatieto) ei nimestään huolimatta ole dataa, vaan informaatiota - ja nimenomaan informaatiota informaatiosta. Termi "meta" tulee kreikan kielestä ja tarkoittaa muutosta. Metadata on kuvailevaa ja määrittelevää tietoa jostakin tallennetusta tietokokonaisuudesta. Web-sivujen yhteydessä puhutaan usein myös resursseista ja resurssien kuvailusta (*resources, resource description*). (Alaterä & Halttunen 2002, 59; Hakala & Stenvall 1998).

Metadataa on ollut olemassa jo paljon ennen Internetiä, mutta maailmanlaajuinen kiinnostus siihen heräsi digitaalisten julkaisujen ja Internetissä olevan laajan kuvailemattoman aineiston ansiosta. Meta-komentojen alkuperäinen tarkoitus oli auttaa hakukoneita luokittelemaan sivuja ja tuottamaan informaatiota grafiikkapainotteisille sivuille. Esimerkiksi Flashilla toteutetut sivut tai muuhun visuaaliseen ilmaisuun perustuvat verkkopalvelut hyötyivät metadatatista. Metadatan tärkein käyttötarkoitus lienee tiedon tallentamisen ja haun edistäminen, mutta sitä voidaan käyttää myös erilaisissa käyttöoikeuksien hallintaan ja kauppaan liittyvissä sovelluksissa sekä erilaisten versioiden hallinnassa. Eri tahoilla luodut meta-

datastandardit on suunniteltu ainakin osittain näitä tarkoituksia varten. (Alaterä & Halttunen 2002, 59; Hakala & Stenvall 1998; KK Mediat 2000-2006).

Metadata koostuu joukosta elementtejä ja attribuutteja, jotka ovat tarpeen tallenteen kuvailussa (Hakala & Stenvall 1998). Tyypillistä metadataa on esimerkiksi cd-levyyn liittyvä tieto (levyn nimi, esittäjän nimi, julkaisuvuosi jne.) tai tekstidokumentin liitännäistieto (omistaja, viimeisin tallennuspäivämäärä, version numero jne.). Haku voidaan kohdentaa metadatakuvailun eri kenttiin (Stenvall 2002).

Oleennaista metadataassa on dokumenttien rakenteinen kuvaileminen sovitun kieliopin eli syntaksin mukaan. Rakenteinen kuvailu tekee dokumenttien erottelun ja tunnistamisen mahdolliseksi, mikä taas mahdollistaa tarkennetun tiedonhaun. Metadataa on kahta eri tyyppiä:

- Sisäinen (*embedded*) metadata on upotettu kuvailtavaan dokumenttiin itseensä, esimerkiksi HTML-dokumentin HEAD-osaan, jossa se on hakurobotin poimittavissa. Verkkodokumenttien sisäisen metadatan tuottamiseen on kehitetty ainakin kahdenlaisia ratkaisuja: metadatan tuottaminen voidaan integroida www-editoreihin, eli ohjelmiin, joilla tehdään www-sivuja, tai metadatan tuottamisessa voidaan hyödyntää erilaisia verkossa toimivia tallennuspalveluita, joiden tuottama metadata liitetään www-sivun HEAD-osaan. Dublin Core -metadataformaatin suomenkielinen tallennuspalvelu on osoitteessa <http://www.lib.helsinki.fi/cgi-bin/dc.pl>.
- Ulkoisessa (*external*) metadataassa kuvailut on tallennettu muualle kuin kuvailtaviin dokumentteihin. Esimerkiksi kirjastojen tietokannoissa on ulkoista metadataa.
(Alaterä & Halttunen 2002, 59-60, 62-63).

Metadataa voidaan jäsentää rakenteisuutensa perusteella ainakin neljälle tasolle:

- Rikkaita metadataformaatteja ovat hyvin rakenteiset formaatit. Esimerkiksi Federal Geographic Data Committeen (<http://www.fgdc.gov/>) formaatti, jolla

kuvataan digitaalista maantieteellistä tietoa ja paikkatietoa, on rikas metadata-formaatti.

- Rakenteisia formaatteja ovat Dublin Core sekä IFA Templates. Näissä formaateissa esitetään noin 20 kenttää.
- Yksinkertaiset formaatit koostuvat muutamista kentistä. Tällaisina voidaan pitää esimerkiksi HTML-kielessä määriteltyjä muutamia kenttiä, kuten author, description, keywords ja robots.
- Automaattisesti poimittuja, "lennosta luotuja" kuvailuja ovat www:n hakukoneiden luomat tekstinäytteet hakutulossivuilleen. Ne koostuvat tyypillisesti sivun otsikosta, tekstinäytteestä sivun alusta, url-osoitteesta sekä tiedoston koon ja muutospäivämäärän kertovista tiedoista.
(Alaterä & Halttunen 2002, 60).

Metadatan merkitys sivuston hakutulossijoituksen parantajana on vähentynyt viime vuosina. Tällä hetkellä isoista hakukoneista vain MSN Search hyödyntää meta-komentoja. Muut hakukoneet, kuten Google tai Yahoo!, eivät huomioi niitä kuin poikkeustapauksissa. Tähän on syynä se, että hakukoneiden huijaajat eli spammaajat alkoivat käyttää metadataa väärin. Hakukoneiden vastaus tähän oli vähentää metadatan merkitystä asteittain ja keskittyä puhtaasti sisällön ja sen ilmaisuuden painotukseen perustuvaan arviointiin. Hyvin laaditusta metadatasta ei kuitenkaan ole haittaakaan. (KK Mediat 2000-2006; Net Site Story, viitattu 29.10.2006).

Hakukoneoptimoinnista sekä hakukonespamista kerrotaan tarkemmin luvussa 5.1 Hakukoneoptimointi.

4.2.2 Dublin Core

Dublin Core on metadataformaatti, joka on yksinkertainen mutta tehokas väline erilaisten verkkodokumenttien kuvailuun. Formaatti sai nimensä siitä, että sen kehitystyö aloitettiin vuonna 1995 Dublinissa Ohiossa. Tarkoituksena oli tehostaa verkkotallenteiden hakua luomalla väline niiden kuvailemiseen. Dublin Core -

standardi koostuu 15:sta elementistä (kentästä), joita valitsemassa on ollut kansainvälinen, eri alojen ammattilaisista koostuva yhteisö. Mukana on ollut asiantuntijoita muun muassa tietojenkäsittely- sekä kirjasto- ja tietopalvelualalta. (Alaterä & Halttunen 2002, 60; Stenvall 2002).

Taulukosta 2 nähdään, että osa Dublin Coren elementeistä liittyy kohteen sisältöön, osa tekijätietoihin ja osa kohteen tunnistamiseen (Hakala & Stenvall 1998).

Sisältö	Tekijätiedot	Tunnistaminen
Nimeke (Title)	Tekijä (Creator)	Identifiointitunnus (Identifier)
Aihe (Subject)	Muu tekijä (Contributor)	Formaatti (Format)
Kuvaus (Description)	Julkaisija (Publisher)	Aikamääre (Date)
Kate (Coverage)	Oikeudet (Rights)	Kieli (Language)
Laji (Type)		
Lähde (Source)		
Suhde (Relation)		

Taulukko 2 Dublin Core -elementit. (Stenvall 2002).

Dublin Coren elementtien lisäksi voidaan käyttää myös tarkenteita, joiden avulla kentän merkitystä pystytään tarkentamaan. Tarkenteet voidaan jakaa **kenttä-tarkenteisiin** ja **merkintäjärjestelmiin**. Tarkentamisessa on kysymys joko kentän merkityksen "jalostamisesta" kenttätarkenteiden avulla tai kentän arvon esittämisestä jonkin tietyn merkintäjärjestelmän mukaisesti. Esimerkiksi aikamääre-kentän yhteydessä voidaan käyttää kenttätarkennetta, joka ilmaisee, että kysymyksessä on nimenomaan luomispäivämäärä (*date created*) tai muokkauspäivämäärä (*date modified*). Merkintäjärjestelmän avulla voidaan kertoa, missä standardimuodossa aikamääre on annettu. Esimerkiksi "2006-06-09" voi viitata joko kesäkuun yhdeksänteen tai syyskuun kuudenteen päivään, ellei ilmoiteta, minkä standardin mukainen esitystapa on kyseessä. (Stenvall 2002).

Dublin Coren keskeiset tavoitteet ovat:

- Kuvailutietojen luomisen yksinkertaisuus ja ylläpito
Jotta kuka tahansa voisi tehdä Dublin Core -kuvailun omalle dokumentilleen, on Dublin Core -kentät pidetty yksinkertaisina ja lukumäärältään vähäisinä.

Kuvailemisen tulee olla helppoa ja edullista, mutta samalla sen on annettava tehokas väline tiedonhakuun.

- Yleisesti ymmärrettävä semantiikka

Tiedon löytyminen Internetin kaltaisesta laajasta tietoympäristöstä on usein rajoittunutta. Tähän ovat syynä erot terminologiassa ja eri organisaatioiden erilaiset kuvailukäytännöt. Dublin Core voi auttaa tiedonhakijaa löytämään informaationsa käyttämällä yleisiä elementtejä, joiden semantiikka on yleisesti tunnettua ja tuettua. Esimerkiksi tiedemies, joka tahtoo tietyn kirjoittajan artikkeleita ja tietyn taiteilijan taiteesta kiinnostunut tutkija ovat varmasti yhtä mieltä tekijä-kentän tarpeellisuudesta. Yhteisten kuvailuperiaatteiden käyttäminen parantaa resurssien löydettävyyttä ja saatavuutta.

- Kansainvälisyys

Dublin Core -yhteisöön kuuluu jäseniä useista eri maista. Tämä on varmistanut sen, että standardin kehityksessä on huomioitu elektronisen ympäristön monikielelliset ja -kulttuuriset ominaisuudet.

- Laajennettavuus

Dublin Core -standardin haaste on siinä, että se on pidettävä yhtä aikaa yksinkertaisena ja tehokkaana. Tästä syystä Dublin Coren kehittäjät ovat pitäneet tärkeänä tarjota mekanismia, jonka avulla Dublin Core -kenttiä on mahdollista laajentaa, jotta se palvelisi paremmin vaativia tietotarpeita. Tämä malli antaa eri tahoille mahdollisuuden käyttää Dublin Core -kuvailua tallenteiden "ytimen" kuvailuun, mutta samalla se tekee mahdolliseksi myös tarkemman kuvailun tiettyihin tarpeisiin.

(Stenvall 2002).

Tavoitteiksi voidaan lukea myös yhdenmukaisuus muihin olemassa oleviin ja kehitettyihin kuvailustandardeihin sekä yhteensopivuus olemassa oleviin kokoelmiin ja indeksointijärjestelmiin. (Stenvall 2002).

Dublin Coren kaikki elementit ovat vapaaehtoisia ja toistettavissa, eikä niiden järjestyksellä ole merkitystä. Kentän arvo voidaan tietyissä kentissä valita merkintäjärjestelmistä, joita ovat esimerkiksi aihe-kentässä käytetyt kontrolloidut sanastot, jotka ovat rajallinen kokoelma tarkkaan määriteltyjä ja johdonmukaisesti käytettyjä termejä. Näiden käyttö saattaa parantaa hakutuloksia huomattavasti,

koska kuten tässä tutkielmassa jo aiemmin todettiin, tietokoneet ovat hyviä vertailemaan sanoja merkki merkiltä, mutta huonoja ymmärtämään ihmisten käyttämiä synonyymejä eri käsitteille. Ilman kontrolloitua sanastoa esimerkiksi termejä "karamellit", "karkit" ja "makeiset" voidaan käyttää tarkoittamaan samaa käsitettä. Sanastojen käyttö voi myös vähentää kirjoitusvirheiden määrää metadatan tallentamisvaiheessa. (Stenvall 2002).

Tarkenteiden käyttö on vapaaehtoista. Dokumentteja voidaan hyvin kuvailla hyödyntämällä pelkästään ydinkenttiä ilman kenttätarkenteita tai merkintäjärjestelmiä. Tavallisesti puhutaankin yksinkertaisesta (*simple*) ja tarkennetusta (*qualified*) Dublin Coresta. Tiettyjen kenttien kohdalla tarkenteiden käyttö on kuitenkin suositeltavaa. (Stenvall 2002).

Tallennetta kuvaillaessa jokin tieto voi näyttää kuuluvan useampaan kuin yhteen kenttään. Vaikka normaalissa tilanteessa on selvä vaihtoehto jokaiselle tallennettavalle tiedolle, voi tietyissä tilanteissa esiintyä "semanttista" päällekkäisyyttä. Tällöin kuvailijan on itse päätettävä, mikä kenttä on sopivin. (Stenvall 2002).

Dublin Coren ydinkenttien lisäksi tarvitaan sopimuksia tai käytäntöjä siitä, miten nuo kentät esitetään. Sovelluksesta riippuen käytössä on erilaisia syntakseja, jotka määrittelevät oikean tavan kenttien, tarkenteiden ja kentän arvojen esittämiseksi. Tällä hetkellä yleisimmät syntaksimuodot ovat HTML (*Hyper Text Markup Language*) ja RDF/XML (*Resource Description Framework/Extensible Markup Language*). (Stenvall 2002).

Jotta metadataa voitaisiin hyödyntää tiedonhaussa, on indeksointiohjelmien kyettävä poimimaan kuvailutiedot HTML-dokumentista. Indeksointiohjelmalle voidaan kertoa, että metadata sijaitsee HEAD-osan META-kentissä, joten ne osaavat kerätä metadatan indeksiinsä. HEAD-osaan upotettu metadata ei näy dokumenttia luettaessa tai tulostettaessa, ellei katsota dokumentin lähdekoodia (Näytä - Lähdekoodi). (Stenvall 2002).

HTML-kielessä jokaisen Dublin Core -elementin määrittely alkaa "<META" -merkkauksella ja loppuu ">" -merkintään. Tämän META-kentän sisällä määritel-

lään kaksi attribuuttia: NAME ja CONTENT. Ensimmäinen kertoo, mistä elementistä on kysymys, ja toinen sen, minkä arvon se saa. (Stenvall 2002). Esimerkiksi rivi

```
<META NAME="DC.Creator" CONTENT="Suojala, Tarja">
```

kertoo, että kyseisen dokumentin tekijä on nimeltään Tarja Suojala.

HTML-kieleen on sovittu tapa, jolla metadatatiedot erotetaan ja tunnistetaan. Sopimukseen liittyy etuliitteen käyttö osoittamaan, että metadatatiedot kuuluvat Dublin Coreen tai johonkin muuhun metadataformaattiin. Dublin Coressa lukemisen helpottamiseksi käytetään etuliitettä "DC", joka kirjoitetaan isoin kirjaimin. Elementin nimi aloitetaan isolla kirjaimella. (Stenvall 2002).

HTML-kielessä suositellaan muiden merkkien tallentamista niiden ascii-muodossa, esimerkiksi näin:

```
<META NAME="DC.Title" CONTENT=" K&uuml;yynniss&uuml; olevat projektit ">
```

(Stenvall 2002).

RDF on World Wide Web Consortiumin (W3C) kehittämä perusta metadatan luomiselle, tallentamiselle ja käsittelylle. RDF on XML:n sovellus, joka kattaa tietoresurssien kuvailemisen. Myös Dublin Core -metadata voidaan helposti esittää RDF:n avulla. (Stenvall 2002).

RDF-syntaksissa käytettävä XML-kieli on ilmaisukykyinen ja mahdollistaa monipuolisen kuvailurakenteen. RDF:n mukainen metadata voidaan esittää myös muodossa, joka soveltuu liitettäväksi HTML-koodiin. Itse syntaksissa XML näkyy paitsi koodissa, myös alkumäärittelyssä: jokainen RDF:n mukainen kuvailu alkaa ilmaisemalla syntaksin taustalla oleva XML-versio, jonka jälkeen RDF-osa aloitetaan merkinnällä <rdf:RDF>. </rdf:RDF> päättää RDF-kuvailun. (Stenvall 1999, 1-2).

RDF:ssä käytetään nimiavaruuksia (*XML-namespaces*, *xmlns*), jotka määrittelevät kuvailussa käytettävien kenttien käsitteitä. Toisin kuin HTML:ssä, RDF-syntaksissa isoilla ja pienillä kirjaimilla on eroa - eli esimerkiksi "Title" ja "title" tulkittaisiin kahdeksi eri elementiksi. XML:ssä käytettävillä tageilla tulee olla myös lopetus, eli esimerkiksi <creator> merkkkaus on lopettava </creator> tagilla. RDF:ssä käytetään usein lyhennettyä merkkkausta: esimerkiksi <title> </title> -merkintä voidaan korvata lyhyemmällä <title /> -merkinnällä. Lyhennetty merkintä on hyödyllinen muun muassa käytettäessä tarkennettua Dublin Corea. (Stenvall 1999, 2).

HTML-osuudessa käytetty esimerkkirivi näyttäisi RDF:ssä tältä:

```
<dc:creator> Suojala, Tarja </dc:creator>.
```

5 HAKUKONEOPTIMOINTI JA -MARKKINOINTI

Hakukoneoptimoinnissa (*Search Engine Optimization*, *SEO*) sekä hakukone-markkinoinnissa (*Search Engine Marketing*, *SEM*) on kysymys siitä, miten www-sivustolle saataisiin mahdollisimman hyvä sijoitus hakukoneiden hakutuloksissa. Tarkoituksena on saada yrityksen sivusto jo ensimmäiselle hakutulossivulle - mieluiten vielä sen alkuun - ja sen myötä yritys ja sen tarjoamat tuotteet ja palvelut paremmin esille.

5.1 Hakukoneoptimointi

Hakukoneoptimointi on prosessi, jossa markkinoitavan sivuston näkyvyyttä hakukoneissa pyritään parantamaan keskittymällä sivuston ominaisuuksien muokkamiseen vallitsevien hakukonealgoritmien mukaisesti. Hakukoneilla suoritetaan maailmanlaajuisesti yli 300 miljoonaa hakua päivässä, ja jokainen niistä voi tuoda

mukanaan potentiaalisen asiakkaan, joten syy lisänäkyvyyden tavoittelemiseen on ilmeinen. Hakukoneoptimoinnin päämääränä on tavoittaa oikea kohderyhmä saattamalla juuri oikeata tietoa hakijan ulottuville mahdollisimman vaivattomasti. (Finkernet Markkinointi, viitattu 29.10.2006; KK Mediat 2000-2006; Net Site Story, viitattu 29.10.2006).

Hakukoneoptimoinnin imago on hieman kyseenalainen, koska iso osa epäeettisestä optimoinnista tapahtuu eri tyyppisille viihdesivustoille ja on johtanut siihen, että hakukoneiden tulokset ovat usein kaukana siitä, mitä hakija halusi. Asiasta tietämättömät puhuvat hakutulosten manipuloimisesta. Oikea hakukoneoptimointi ei kuitenkaan ole hakutulosten manipuloimista, vaan toiminnan tarkoitus on auttaa tiedonhakijaa löytämään informaatiopaljoudesta tarvitsemansa tieto. Jos taas päämääränä on parantaa sivun sijoittumista hakutuloksissa hakutermeillä, jotka eivät liity mitenkään sivun todelliseen sisältöön, voidaan puhua manipuloinnista. (KK Mediat 2000-2006; Net Site Story, viitattu 29.10.2006).

Hakukoneoptimointi eroaa muusta Internet-markkinoinnista siinä mielessä, että kyseessä on hyvin teknispainotteinen ja pitkäkestoinen prosessi. Lyhyen aikavälin tuloksia saadaan ideaalitulanteessa jo parissa viikossa, mutta pitempikestoisten tulosten muodostaminen voi vaatia vuosikausien työn. Yleisesti ottaen hakukoneoptimoinnissa pätevät samat periaatteet kuin muussakin markkinoinnissa: mitä enemmän aikaa, rahaa ja työtä on valmis sijoittamaan, sitä enemmän voi saada. Hakukoneoptimoinnin periaatteen voi kiteyttää yhteen lyhyeen lauseeseen: "Make pages for users, not for search engines". (KK Mediat 2000-2006).

5.1.1 Avainsanat

Avainsana (*keyword*) on yksittäinen sana tai useamman sanan ilmaisu, joka on tärkeä promotoitavalle palvelulle ja jolla käyttäjien halutaan tulevan sivulle. Laajempi määrittely huomioi myös odotetut sekä toteutuneet kävijämäärät ja keskittyy vain niihin avainsanoihin, jotka tuovat kokonaisvaltaisesti parhaan konversion. Avainsanojen karsinta ei siis perustu pelkkään arvioon, vaan tarkkaan tutki-

mukseen siitä, millaisia hakusanoja tiedonhakijat käyttävät hakiessaan tietoa hakukoneilla. Tätä kutsutaan nimellä avainsana-analyysi. (KK Mediat 2000-2006; Net Site Story, viitattu 29.10.2006).

Avainsanoja kannattaa ja voi sijoittaa kaikkialle dokumenttiin. Avainsanoja tulisi käyttää erityisesti seuraavissa elementeissä:

- title-otsikko
- meta-komennot (erityisesti keywords ja description)
- otsikot h1-h6
- sisältö
- hyperlinkit
- kuvien alt-tekstit

(KK Mediat 2000-2006).

Avainsanoja tulisi olla noin 10-20 kappaletta, ja yksittäistä avainsanaa suositellaan käytettäväksi 3-6 kertaa dokumentin sisällä. Suurin hyöty avainsanoista saadaan, kun käytetään ilmaisuasultaan rikasta kieltä: synonyymejä, taivutusmuotoja jne. Kohtuullinen avainsanojen korostaminen esimerkiksi lihavoinnilla nostaa usein sijoitusta hakutuloksissa. (KK Mediat 2000-2006; Net Site Story, viitattu 29.10.2006).

Avainsanoja valittaessa tulisi keskittyä mahdollisimman tarkasti rajatulle alueelle. Esimerkiksi Internet on liian yleinen hakusana liikkeenharjoittajille, mutta jos täsmennetään hakutermin muotoon "Internet palvelut Pori", se on jo huomattavasti parempi. Mitä yleisempi termi on kyseessä, sitä kovempi on kilpailu kärkisijoituksista. Jos esimerkiksi yksi sivuston avainsanoista on Internet, se pääsee todennäköisesti hakutulosten kärkeen ja saa sen myötä paljon kävijöitä, mutta se ei välttämättä tarjoa sitä, mitä kävijät hakevat. Jos taas käytetään tarkennettuja avainsanoja, kävijämäärät ovat pienempiä ja tarkemmin kohdennettuja. (KK Mediat 2000-2006).

Valitettavasti avainsanoja käytetään myös väärin. Vanhin ja yksinkertaisin tapa huijata hakukoneita on kirjoittaa pienellä fontilla ja/tai sivun taustan värisellä

tekstillä sivu täyteen avainsanoja, jotka eivät sille kuulu. Väärinkäyttöä on myös avainsanojen luetteleminen kommenttien sisällä tai piilottaminen CSS:n avulla. Edistyneempiä spammaamiskeinoja ovat nk. *cloaking* sekä *doorway pages*. Cloakingin ideana on, että spammaajat näyttävät kävijöille "oikean" sivunsa ja hakuroboteille toisen, runsaasti avainsanoja sisältävän sivun, jolloin sijoitukset hakukoneissa nousevat. Doorway pages -menetelmässä on kyse siitä, että spammaajat tekevät sivuja, joille on laitettu runsaasti avainsanoja. Nämä sivut ohjaavat kävijän spammaajan "oikealle" sivulle. Uudelleenohjaus on usein tehty esimerkiksi JavaScriptillä, jota hakurobotit eivät kykene seuraamaan. Hakukoneet pyrkivät tunnistamaan hakukonespamin ja rankaisevat siitä. (KK Mediat 2000-2006).

5.1.2 Meta-komennot

Meta-elementit sijoitetaan HTML- tai XHTML-dokumentin HEAD-osaan. XHTML:n mukainen syntaksi näyttää tältä:

```
<head >
<meta name="keywords" content="avainsana1, avainsana2, avainsana3" / >
<meta name="description" content="Tähän tulee kuvaus" / >
```

(KK Mediat 2000-2006).

Avainsanat määritellään attribuutilla **keywords**. Paremman sijoituksen saavuttamiseksi voidaan käyttää seuraavia keinoja:

- Avainsanat kannattaa sijoittaa tärkeysjärjestyksessä vasemmalta oikealle.
- Jokainen sivun otsikosta löytyvä avainsana tulisi löytyä sekä avainsanalistasta että kuvauksesta, minkä lisäksi näitä sanoja pitäisi olla joka puolella sivua.
- Avainsanan toistamista tulee välttää - se on spammaamista ja vähentää avainsanan arvoa.
- Oikeinkirjoituksen tarkistaminen on tärkeää.
- Pisteitä tai pilkkuja ei tule käyttää metan kuvauksessa.

- Avainsanoja, jotka eivät kuulu sivulle, ei pidä käyttää.
- Avainsanoista kannattaa muodostaa nk. avainlauseita.
(KK Mediat 2000-2006).

Jokaisella hakukoneella on hieman erilainen kyky käsitellä keywords-attribuutin sisältöä. Havaintojen mukaan maksimipituus sisällölle on keskimäärin 200 merkkiä. (KK Mediat 2000-2006).

Kuvaus määritellään attribuutilla **description**, johon pätevät hyvin pitkälti samat säännöt kuin keywords-attribuuttiinkin. Koska kuvaus voi joskus näkyä hakutuloksen yhteydessä, sen pitäisi olla mahdollisimman houkutteleva käyttäjiä ajatellen, mutta samalla tehokas hakukoneoptimointia ajatellen. Käytännössä tämä tapahtuu käyttämällä luonnollista tekstiä, johon on sijoitettu avainsanoja sopiviin kohtiin. (KK Mediat 2000-2006).

Useimmat hakukoneet näyttävät kuvauksesta ensimmäiset 170 merkkiä. Jos sisällön kuvaus on vaikea saada mahtumaan 170 merkin sisälle, kannattaa jakaa sivu alisivuihin. Tällöin se on myös helpompi indeksoida. Jos kuvaustekstiä ei ilmaista, käytetään sivun ensimmäisiä lauseita. (KK Mediat 2000-2006).

Muita meta-komentoja ovat:

language

```
<meta http-equiv="content-language" content="fi">
```

Attribuutti ilmaisee hakukoneelle ja hakurobotille sivun kielen. Kieli merkitään lyhenteellä, esimerkiksi fi= suomi, EN=englanti, FR=ranska jne.

robots

```
<meta name="robots" content="noindex, nofollow">
```

Attribuutti kertoo robotille, mitkä sivut tulisi indeksoida (tai jättää indeksoimatta).

author

```
<meta name="author" content="Suojala, Tarja">
```

Attribuuttia käytetään ilmoittamaan sivun tekijä tai ylläpitäjä.

copyright

```
<meta name="copyright" content="2K mediat 2000-2006">
```

Attribuutti ilmoittaa hakukoneille ja käyttäjille, että sivu on tekijänoikeudellisesti suojattu.

expires

```
<meta name="expires" content="Mon, 01 Jan 2007 18:00:00 GMT">
```

Attribuutti kertoo, että sivulla on vanhenemispäivämäärä. Tätä attribuuttia käytetään yleisesti esimerkiksi tietokantayhteyttä hyödyntävillä sivuilla.

refresh

```
<meta http-equiv="refresh" content=7;URL=http://www.2kmediat.com">
```

Attribuutti ohjaa käyttäjän annettuun url-osoitteeseen ilmoitetun viiveajan (esimerkissä seitsemän sekunnin) kuluttua. Hakukoneet pyrkivät hylkimään sivuja, jotka käyttävät tätä ominaisuutta, joten sitä kannattaa välttää.

abstract

```
<meta name="abstract" content="Kuvaus sivusta">
```

Tämä attribuutti on saman kaltainen kuin description. Sitä voidaan käyttää ilmoittamaan toinen kuvaus niihin hakukoneisiin, jotka sen tunnistavat.

distribution

```
<meta name="distribution" content="local">
```

Attribuutti kertoo sivun kohderyhmän. Arvot voivat olla

- global
- local
- IU (= *internal use*, sisäinen dokumentti eli intranet-käyttö)

revisit

```
<meta name="revisit-after" content="14 days">
```

Attribuutti ilmoittaa, miten usein robotin pitäisi vieraillla sivulla. Useimmat robotit eivät noudata tätä.

rating

```
<meta name="document-rating" content="14 years">
```

Attribuutti kertoo sivun sisällön tason. Mahdolliset arvot ovat

- Safe For Kids
- 14 Years
- General
- Restricted
- Mature

(Heikkinen 2004; KK Mediat 2000-2006).

5.1.3 Web Design

Hakukoneoptimojien yhteydessä puhutaan Web Designista usein hieman negatiiviseen sävyyn. Syy on se, että ulkoasultaan näyttävä www-sivu on yleensä koodirakenteeltaan sellainen, ettei sillä ole mahdollisuuksia sijoittua hakutuloksissa hyvin. Kun hakukoneoptimojen suositukset on tehty, sivusto on jouduttu rakentamaan uudelleen, minkä jälkeen ulkoasu on usein erilainen kuin alun perin on toivottu. (KK Mediat 2000-2006).

Sivua suunniteltaessa kannattaa yrittää asettua hakurobotin asemaan. Se ei näe grafiikkaa, hienoja skriptejä tai sovelluksia, ei ymmärrä asettelua eikä kuule ääniä. Se seuraa sivulla olevaa koodia rivi riviltä ja lukee sisällön sen perusteella. Sisällön merkitystä ei voi koskaan korostaa liikaa hakukoneoptimoinnissa. Laadittaessa sivustoa on hyvä pyrkiä selkeään ilmaisuun ja esittää asia mieluiten vähintään 250 - 300 sanalla sivua kohti. Toisaalta sivuston koko kannattaa pitää kohtuullisena, sillä se nopeuttaa indeksoimista. (KK Mediat 2000-2006; Net Site Story, viitattu 29.10.2006).

Useimmilla www-sivuilla on käytössä seuraavanlainen rakenne:

- sivun alussa on otsikko
- otsikkoa seuraa johdatus aiheeseen

- johdatusta seuraa varsinainen artikkelisisältö
- sivun lopussa on mahdollisesti yhteenveto (KK Mediat 2000-2006).

Hakukoneet ja hakurobotit ovat tietoisia rakenteesta ja antavat enemmän painoarvoa sivun yläosassa sijaitsevalle sisällölle, minkä vuoksi avainsanojakin kannattaa pyrkiä käyttämään paljon erityisesti sivun alkupuolella. (KK Mediat 2000-2006).

Kehysten käyttämistä suositellaan välttämään. Jos kuitenkin osaa välttää kehyksien käyttämiseen liittyvät ongelmakohdat, ei niiden pitäisi aiheuttaa vaikeuksia. Sivun alussa on ensinnäkin hyvä määrittää, että kyseessä on kehyssivu. Tämä tapahtuu käyttämällä HTML:n !doctype -komentoa. (KK Mediat 2000-2006; Net Site Story, viitattu 29.10.2006).

Elementillä **noframes** määritellään sisältö, jos selain ei tue kehyksiä. Tällöin näytetään noframes - komennon sisällä oleva sisältö. Noframes on jäänne Internetin alkuajoilta, jolloin selaimet eivät välttämättä ymmärtäneet kehyksiä. Vaikka selaimet ovat kehittyneet, ei noframes-tagia ole syytä unohtaa, sillä hakuroboteille siitä saattaa olla apua. Hakurobotit indeksoivat usein vain kehyssivun, joka sisältää ainoastaan kehysrakenteen. Ne eivät välttämättä osaa jatkaa matkaansa erillisille sisältösivuille, joilla on kaikki tarvittava tieto. (KK Mediat 2000-2006; Net Site Story, viitattu 29.10.2006).

Noframes-tagin sisään kannattaa laittaa lyhyt kuvaus sivustosta sekä sen jälkeen listata linkit yksittäisille sivuille. Hakukoneet seuraavat näitä linkkejä indeksoiden lopulta koko sivuston. Lisäksi kehyssivuun on hyvä sijoittaa vielä otsikko ja meta-komennot. Jos kehysten käyttäminen edelleen arveluttaa, vastaava kokonaisuus voidaan tehdä CSS-tyylimäärittelyjä hyödyntämällä. (KK Mediat 2000-2006; Net Site Story, viitattu 29.10.2006).

Linkeistä kannattaa rakentaa avainsanarikkaita sen sijaan, että käyttäisi "Klikkaa tästä" -tyyppisiä linkkitekstejä. Jos sivustolla on kuvia, niissä on hyvä käyttää alt-attribuuttia. Monipuolisten kuvaustekstien käyttäminen kaikilla sisältöelementeillä

on yleensäkin suositeltavaa, sillä kuten jo aiemmin todettiin, hakurobotit eivät indeksoi esimerkiksi kuva- tai äänitiedostoja niiden todellisen sisällön, vaan niiden tiedostonimien ja kuvaustekstien perusteella. (KK Mediat 2000-2006).

5.1.4 Linkkisuosion kasvattaminen

Linkkisuosion (*link popularity*) perusajatuksena on, että mitä useampi sivu linkittää kyseiseen sivuun, sitä tärkeämpää sen sisällön täytyy olla ja sitä korkeammalle sivun on sijoitettava hakutuloksissa. Linkkisuosio mitataan tarkkailemalla sivulle viittaavia linkkejä sekä niiden arvoa. (KK Mediat 2000-2006).

Helppimmin linkkejä saa esimerkiksi kirjoittelemalla www:ssä oleviin vieraskirjoihin, lisäämällä linkin FFA-linkkilistoille (*Free For All*) tai osallistumalla keskusteluun www:n keskustelupalstoilla. Hakukoneoptimoinnin alkuaikoina 1990-luvun lopulla kyseiset menetelmät olivat erittäin suosittuja, ja niitä käytetään vielä tänäkin päivänä. Nykyiset hakualgoritmit kuitenkin tunnistavat useimmat edellä mainitun tyyppiset resurssit, eivätkä anna niistä oleville linkeille paljoa arvoa. (KK Mediat 2000-2006).

Nykyään tavallisin ja samalla hyväksyttävoin tapa saada ulkopuolisia linkkejä ovat erilaiset hakemistolistaukset. Suurin, tunnetuin ja arvostetuin hakemisto on vapaaehtoistyöllä ylläpidettävä Open Directory Project eli ODP (<http://dmoz.org>), jonka listauksiin luottaa muun muassa Google. Sinne voi kuka tahansa ehdottaa laatusivustoa listattavaksi. DMOZ-toimittajat käyvät aktiivisesti läpi oman alansa sivuja etsien hyvää, ainutlaatuista ja merkittävää informaatiota. Toimittajille on laadittu tiukka eettinen koodisto, jolla pidetään huoli siitä, ettei toimittajan valtaa käytetä väärin esimerkiksi oman liiketoiminnan edistämiseksi. Huono puoli hakemistoissa on niiden jatkuva kaupallistuminen hakukoneoptimoinnin kustannuksella: yhä useammin hakemistolistauksia saa vain joko ostamalla tai vaihtokaupalla. (KK Mediat 2000-2006).

Toinen suosittu tapa kasvattaa linkkisuosiota on suora lähestyminen. Webmasterit viestivät keskenään esimerkiksi sähköpostitse ja pyytävät linkkejä muilta asiaan-kuvuluilta sivustoilta. Tämän metodin huono puoli ovat spammausviestit kymmenille tuhansille www-sivustojen ylläpitäjille. (KK Mediat 2000-2006).

Kolmas tapa ovat nk. vaihdot (*back links, reciprocals*), joissa www-sivustot vaihtavat linkin keskenään 1:1 -periaatteella. Sen vääristävä vaikutus hakutuloksiin on kuitenkin niin huomattava, että hakukoneet pyrkivät huomaamaan linkkivaihdot ja rankaisemaan niistä. Alalla on myös nk. link broker -yrityksiä, jotka tarjoavat mahdollisuutta joko ostaa tai vaihtaa suuria määriä linkkejä muiden webmasterien ja hakukoneoptimojien kanssa. Tämänkin toiminnan hakukoneet pyrkivät estämään. (KK Mediat 2000-2006).

Lähes kaikki edellä mainitut linkkisuosion kasvattamistavat pohjautuvat keino-tekoiseen suosion lisäämiseen, josta hakukoneet varoittavat ja useimmiten myös rankaisevat. Suositeltavin - ja yleensä myös edullisin - tapa on keskittyä laadukkaan sisällön ja sivuston tekemiseen. Käyttäjät ja muut webmasterit linkittävät sivustoon pyytämättäkin, jos he katsovat sen sisällön olevan tarpeeksi laadukasta. (KK Mediat 2000-2006).

Jos halutaan tietää, millä sivuilla on linkki esimerkiksi työministeriön sivustolle <http://www.mol.fi>, sen saa selville Googlessa haulla

link:www.mol.fi

Yahoo!-hakukoneessa vastaava haku tehdään näin:

linkdomain:www.mol.fi

Ero linkittävien sivujen määrässä eri hakukoneiden välillä on huomattava. Googlen kohdalla tulos on vain otos todellisesta tilanteesta, sillä Google näyttää ainoastaan mielestään tärkeimmät sivut. Yahoo!-n tulokseen taas sisältyvät myös sisäiset linkit. (KK Mediat 2000-2006).

Sekä Google että Yahoo! tarjoavat myös erillisen mittarin linkkisuosion seuraamiseen. Googlessa tämä mittari tunnetaan nimellä PageRank. Jokaiselle sivulle annetaan PR-arvo asteikolla 1-10. Suomessa vain harvat sivut saavat korkeampaa arvoa kuin 6/10. Yahoo!n vastaava mittari on nimeltään Yahoo!Rank. Molemmat mittarit on mahdollista saada käyttöön asentamalla kyseisen hakukoneen työkalurivin selaimeensa. (KK Mediat 2000-2006).

Linkkisuosiota voidaan tarkkailla myös www-sivuston omien lokitiedostojen avulla. Niistä saadaan selville, mitkä sivut tuovat kävijöitä ja siten myös linkittävät kyseiseen www-sivustoon. (KK Mediat 2000-2006).

5.1.5 Teemat

Hakukoneista ainakin Google hyödyntää hakualgoritmissään sisällön teemaa eli aihetta. Teemapohjainen ajattelu vaikuttaa erityisesti linkkisuosion laskemiseen. Jos esimerkiksi www-sivuja tekevä yrittäjä saa sivustolleen linkin puutarhatuotteita myyvältä asiakkaaltaan, tätä ei painoteta paljoa, koska sivustojen aihealueet ovat niin erilaiset. Teemapohjaisen ajattelun idea on vähentää linkkivaihtoa ja linkkien ympärille syntyneen kaupallisen toiminnan arvoa. (KK Mediat 2000-2006).

5.2 Hakukonemarkkinointi

Hakukonemarkkinointi on prosessi, jossa sivuston näkyvyyttä hakutuloksissa parannetaan keskittymällä nk. maksettuun mainontaan hakutulosten yhteydessä. Nämä mainokset näkyvät yleensä hakutuloksissa omana ryhmänään otsikon Sponsoritulokset, Sponsoroidut linkit tms. alla. Keskeisimmät hakukonemarkkinoinnin menetelmät ovat Googlen AdWords -hakukonemarkkinointijärjestelmä, Yahoo!n Sponsored Search, maksulliset hakemistolistaukset sekä muu maksettu Internet-mainonta. (KK Mediat 2000-2006).

Hakukonemarkkinoinnissa mainostaja tarjoaa tietyllä hakusanalla tai -lauseella sivulleen tulevasta kävijästä tietyn hinnan. Mitä korkeampi tarjottu hinta on, sitä näkyvämmälle paikalle mainostajan linkki sijoitetaan hakutuloksissa. Kysymyksessä on siis eräänlainen huutokauppa hakutulosten yhteydessä olevasta mainos-tilasta. Mainostaja maksaa tarjoamansa summan, kun kävijä klikkaa mainosta ja siirtyy mainostajan sivustolle. Klikkaushinta (*Pay Per Click, PPC*) muodostuu sen mukaan, miten suositusta hakusanasta on kysymys. Pienin veloitettava hinta on 5 senttiä per klikkaus, kun taas kalleimmista hakusanoista veloitetaan kymmeniä euroja. (Finkernet Markkinointi, viitattu 29.10.2006; KK Mediat 2000-2006; Net Site Story, viitattu 29.10.2006).

Sijainti korkeammalla hakutuloksissa tietää yleensä parempaa klikkausprosenttia: ensimmäisenä tuloksissa olevaa linkkiä klikataan yleensä enemmän kuin viidentenä tai kymmenentenä olevaa. Tämän vuoksi paremmasta sijainnista on myös maksettava korkeampi hinta. Mainostajat asettavat maksimihinnan, jonka haluavat maksaa hyvästä sijainnista. Järjestelmä pitää huolen siitä, että parhaalle paikalle pääsyt mainos maksaa yhden sentin enemmän kuin mikä on toiseksi parhaalle paikalle laitetun mainoksen maksimihinta. Toisena oleva mainos ei maksa omaa maksimihintaansa, vaan yhden sentin enemmän kuin mikä on kolmanneksi parhaalle paikalle päässeän mainoksen maksimihinta. (Net Site Story, viitattu 29.10.2006).

Markkinoilla on myös hakemistoja, joista voi ostaa parhaan paikan ja/tai hakusanaa vastaavan banneripaikan vuodeksi eteenpäin. Tällä menetelmällä ei voi saada varmuutta yksittäisen klikkauksen hinnalle, koska hinta on kiinteä eikä määräydy klikkausten mukaan. (Net Site Story, viitattu 29.10.2006).

Hakukonemarkkinointi on parhaimmillaan tarkkaan kohdennettu, nopea ja tehokas keino lisätä www-sivuston kautta tulevaa liikevaihtoa. Nopeimmillaan hakukonemarkkinointikampanja voidaan käynnistää alle puolessa tunnissa. Verrattuna hakukoneoptimointiin, jolla ei voi varmistaa korkeaa sijoitusta hakutuloksissa, hakukonemarkkinointi takaa mainostajalle hyvän paikan hakutuloksissa niin kauan kuin mainostaja on valmis maksamaan paikasta riittävästi. (Finkernet

Markkinointi, viitattu 29.10.2006; KK Mediat 2000-2006; Net Site Story, viitattu 29.10.2006).

6 VERTAILTAVAT HAKUKONEET

Tutkimuksessa on mukana viisi hakukonetta. Markkinajohtaja Google on itse-oikeutetusti mukana, koska sitä pidetään yleisesti parhaana ja koska halusin testata, onko se todella ylivoimainen muihin hakukoneisiin nähden. Yahoo! -hakukoneen otin mukaan siksi, että Yahoo! ilmoitti syksyllä 2005 oman hakuindeksinsä olevan Googlen hakuindeksiä suuremman - tarkoitukseni on tutkia, onko Yahoo! parempi myös laadullisesti, vai pelkästään määrällisesti.

Kysellessäni ihmisiltä, millä hakukoneella he hakivat tietoa Internetistä ennen Googlen tuloa, lähes kaikki vastaavat käyttäneensä AltaVistaa, jota itsekkin käytin paljon. Halusin AltaVistan mukaan tutkimukseeni nähdäkseni, onko se pysynyt kehityksen vauhdissa mukana.

Microsoftin syyskuussa 2006 julkaisema Live Search edustaa tutkimuksessa uusia hakukoneita. Ideana on, että se korvaa Microsoftin aiemman hakukoneen MSN Searchin, joka on pitänyt hakukoneiden suosituimmuskilpailussa kolmospaikkaa Googlen ja Yahoo!:n perässä. On mielenkiintoista nähdä, miten uusi tulokas pärjää vertailussa vanhoille konkareille.

Mielestäni vertailuun piti saada myös yksi puhtaasti kotimainen hakija. Ennen Googlen käyttäjäksi siirtymistäni oli oma suosikkihakukoneeni www.fi, johon olen turvautunut silloin tällöin myös viime vuosina. Sattumalta www.fi oli ajan-kohtainen juuri tutkimukseni teko aikaan, sillä se uudistui niin toiminnaltaan kuin ulkoasultaan lokakuussa 2006. Tämän vuoksi katsoin sen olevan hyvä valinta edustamaan suomalaisia hakukoneita vertailussa.

Tässä tutkielmassa keskitytään ensisijaisesti hakukoneiden suomenkielisiin käyttöliittymiin.

6.1 Google

Syksyllä 2006 kahdeksan vuotta täyttäneen Googlen (<http://www.google.fi>) asemasta kertonee parhaiten se, että termi "googlettaa" (*to google*) on monelle tänä päivänä Internetin hakukoneilla tapahtuvan tiedonhaun synonyymi. Hakukoneen nimi tulee matemaattisesta termistä "googol", joka tarkoittaa lukua, jossa numeroa 1 seuraa sata nollaa. Termin käyttö Googlessa viittaa yrityksen tavoitteeseen järjestää www:n suunnatonta informaatiomäärää. HTML-tiedostojen lisäksi Google pystyy indeksoimaan muun muassa pdf-, asp-, doc-, xls-, ppt-, rtf-, swf- ja php-tyyppisiä tiedostoja. (Google 2006; Karvonen 2005).

Google on paljon enemmän kuin sanahakukone: siinä on kuvahaku, Usenet-keskusteluryhmien haku sekä Open Directory -aihehakemisto. Käyttäjä voi valita, kohdistetaanko haku kaikkiin sivuihin, vain suomenkielisiin sivuihin vai pelkästään suomalaisiin sivuihin. Englanninkielisen käyttöliittymän etusivulla on linkit myös muun muassa hakukoneen omiin video-, uutis-, kartta- sekä kirjahakuihin. Yksi Googlen erikoisuuksista on Google Suggest -palvelu, joka ehdottaa reaaliajassa vaihtoehtoisia hakusanoja. Kun hakulaatikkoon alkaa kirjoittaa esimerkiksi sana "identification", näyttää Google luetteloa jo kirjoitetuilla kirjaimilla alkavista hakusanoista (identity, identified jne.). Samalla hakijalle näytetään, kuinka monta viitettä milläkin hakusanalla löytyy Googlen tietokannasta. Seitsemän kirjaimen jälkeen ykkösvaihtoehdoksi kohoaa "identification". Tämän jälkeen tarvitsee vain painaa enteriä, ja Google suorittaa haun. Google Suggest -palvelu on vielä toistaiseksi beta-asteella, ja se löytyy osoitteesta <http://www.google.com/webhp?complete=1&hl=en>. (Google 2006; Haasio 2003, 41; Haataja 2005).

Hakutermin välissä ei tarvita AND- tai JA-sanaa eikä plusmerkkiä, koska AND on Googlen nk. *oletusoperaattori*, toisin sanoen Google lisää sen automaattisesti hakutermin väliin. Google suosii sivuja, joilla hakutermit sijaitsevat lähekkäin,

sillä on todennäköistä, että sivu on käyttäjän tiedontarpeen kannalta relevantti, jos hakutermit ovat lähellä toisiaan. Tällöin niillä on yleensä selkeä asiayhteys. Jos halutaan tehdä haku, jonka tuloksessa olevissa dokumenteissa ei mainita lainkaan jotakin tiettyä sanaa, voidaan sen esiintyminen kieltää laittamalla sen eteen miinusmerkki, joka vastaa Boolean NOT-operaattoria. Tuottaakseen mahdollisimman tarkkoja tuloksia Google hakee sanoja vain siinä muodossa kuin ne on kirjoitettu, mikä tarkoittaa käytännössä sitä, ettei se tue wildcard-hakuja. Käyttäjän on siis tarvittaessa toistettava haku esimerkiksi sanan eri taivutusmuodoilla. Tätä voidaan kuitenkin käyttää jokerimerkkinä fraasihaussa siten, että jokainen tähti korvaa yhden sanan. Esimerkiksi haulla "two * * three" voidaan saada tulokset "two two four three", "two out of three", "two or maybe three" jne. Jos taas halutaan hakea fraasilla "jokainen taaplaa tyylillään", eikä olla aivan varmoja keskimäisen sanan kirjoitusasusta, voidaan haku muotoilla "jokainen * tyylillään". Boolean operaattoreista Google hyväksyy vain OR-operaattorin, jonka täytyy olla kirjoitettu nimenomaan muodossa "OR", ei "or". Muutoin Google ei tee eroa isojen ja pienten kirjainten välillä, vaan kaikki kirjaimet käsitellään pieninä kirjaimina. Näin ollen ei ole merkitystä, kirjoitetaanko hakulaatikkoon Ruotsi, ruotsi vai ruOTsi. (Google 2006; Haasio 2003, 40; Notess 2006). Jos kuitenkin käytetään fraasihakua, Google etsii hakutermiä juuri siinä muodossa kuin se on kirjoitettu lainausmerkkien sisään, eli tällöin isoilla ja pienillä kirjaimilla on eroa.

Hakukentistä Googlessa toimivat muun muassa *intitle*, *inurl*, *link* ja *site*. Google jättää automaattisesti huomiotta stop words -sanat, kuten "http" ja ".com", sekä tietyt yksinumeroiset luvut ja yksittäiset kirjaimet, koska nämä auttavat harvoin haun rajoittamisessa ja voivat päinvastoin hidastaa hakua. Jos halutaan sisällyttää hakuun stop words -termi, tulee sen eteen laittaa plusmerkki. Plus- tai miinusmerkin jälkeen ei laiteta välilyöntiä, vaan ne tulevat välittömästi ennen hakutermiä. (Google 2006; Notess 2006).

Hakua voidaan rajata myös käyttämällä tarkennettua hakua. "Kaikilla sanoilla" -kenttä vastaa AND-operaattoria, "täsmällisellä ilmauksella" -kenttä fraasihakua, "millä tahansa sanoista" -kenttä OR-operaattoria ja "ilman sanoja" -kenttä NOT-operaattoria. Haun rajaus voidaan kohdistaa myös muun muassa tiedostomuotoon,

päivitysajankohtaan ja verkkotunnukseen sekä siihen, missä päin sivua (otsikossa, tekstissä, url-osoitteessa, sivuun osoittavissa linkeissä) hakuehto esiintyy.

Asetukset-linkkiä klikkaamalla päästään valitsemaan käyttöliittymän kieli, hakutulossivulla kerralla näytettävien tulosviitteiden määrä (joka on oletuksena kymmenen) sekä se, avataanko hakutulokset uudessa selainikkunassa. Käyttäjä voi määrittellä myös sen, minkä kielisiä sivuja haetaan. Kielivalinnat-sivulta löytyy muun muassa linkit muiden maiden Google-sivustoille. (Google 2006).

Hakutulostilan alussa näytetään vaaleansinisellä taustalla sponsoroidut linkit ja niiden jälkeen muut tulokset. Hakutuloksen ensimmäisellä rivillä näkyy löytyneen www-dokumentin otsikko, jonka alla on esimerkinomainen tekstipätkä kyseiseltä sivulta. Hakusanat näkyvät siinä lihavoituna. Nämä esimerkkitekstit näyttävät hakijalle, missä yhteydessä hakutermejä on sivulla käytetty jo ennen kuin hän siirtyy kyseiselle sivulle. Esimerkkitekstin alla on sivun url-osoite sekä sivun koko kilotavuissa. Kun Google löytää useita hakutuloksia samalta sivustolta, hakusanoja parhaiten vastaava sivu on listassa ensimmäisenä ja loput sivut on listattu sisennettynä sen alle. Hakutuloksessa olevaa Välimuistissa-linkkiä klikkaamalla käyttäjä voi nähdä sivun sisällön sellaisena kuin se oli sivun luetteloinnin aikaan. Jos sivun varsinainen osoite ei syystä tai toisesta toimi, voi käyttäjä silti nähdä välimuistissa olevan sivun ja mahdollisesti löytää hakemansa tiedon sieltä. Välimuistiin tallennetussa versiossa hakutermit näkyvät korostettuina. Samankaltaisia sivuja -linkin alta löytyy linkkisuosituksia sisällöltään vastaaville sivustoille. (Google 2006; Haasio 2003, 47).

Kokeilen onneani -nappi vie käyttäjän suoraan hakutulostilan kärjessä olevalle sivulle ilman, että hänelle näytetään muita tuloksia. (Google 2006). Esimerkiksi fraasihaku "Solar Films" ja Kokeilen onneani -napin klikkaus vievät suoraan Solar Films -tuotantoyhtiön sivuille.

Käyttäjä voi myös halutessaan asentaa selainikkunaan Googlen hakupalkin Google Toolbar, jonka avulla on mahdollista käyttää Googlea riippumatta siitä, millä www-sivulla on (Haasio 2003, 48). Hakupalkin uusimmassa versiossa olevaa hakukenttää on tehostettu. Kun käyttäjä kirjoittaa hakusanaa hakulaatikkoon,

hänelle näytetään lista hyödyllisiä ehdotuksia, aivan kuten Google Suggest - palvelussa, sekä oikeinkirjoitusehdotuksia, hakupalkin hakuhistoria ja kirjanmerkit. Hakulaatikon G-kuvaketta klikkaamalla päästään valitsemaan muun muassa se, kohdistetaanko haku pelkästään suomalaisiin sivuihin tai senhetkiseen sivuun, tai tehdäänkö kuva- tai Kokeilen onneani -haku. (Google 2006).

Kun url-osoite lähetetään Googlelle, menee lisäyksen ajoituksesta ja Googlen jonoista riippuen yhdestä neljään viikkoon, ennen kuin se on indeksoitu (Google 2006).

6.2 Yahoo!

Siinä missä Google on maailman suosituin hakukone, on Yahoo! (<http://fi.search.yahoo.com>) Alexa Internetin (01.11.2006) mukaan Internetin suosituin verkkosivusto. Yahoo!-n perustivat Stanford Universityn opiskelijat David Filo ja Jerry Yang tammikuussa 1994, ja sen pääkonttori on Sunnyvalessa Kaliforniassa. Nimensä Yahoo! sai sanoista Yet Another Hierarchical Official Oracle. (Yahoo! Inc. 2006).

Yahoo! on kasvanut olemassaolonsa aikana laajasta aihehakemistosta hakukoneen, aihehakemiston ja portaalin yhdistelmäksi. Yahoo!-hakukoneen etusivulla - joka muistuttaa yksinkertaisuudessaan ja selkeydessään huomattavasti Googlen etusivua - on sanahaun lisäksi kuva- sekä uutishaku. Englanninkielisen liittymän etusivulta löytyy myös muun muassa video- ja audiodokumentaaliset, aihehakemisto, työpaikkahaku sekä alueellinen haku, jossa haun voi rajata esimerkiksi postinumeron tai kaupungin mukaan. Asetuksista käyttäjä pääsee valitsemaan, mitkä välilehdet etusivulla näytetään. (Karvonen 2005; Notess 2006).

Yahoo!-n oletusoperaattori on sama kuin Googlessakin, eli AND. Näin ollen Yahoo! hakee vain sivuja, jotka sisältävät kaikki haussa käytetyt hakusanat. Poikkeuksena ovat hyvin tavalliset sanat kuten "it" ja "how". Sanoja voidaan sulkea pois hausta laittamalla miinusmerkki niiden eteen. Katkaisumerkkiä Yahoo! tukee ainoastaan fraasissa kokonaisen sanan korvaajana. (Yahoo! 2006). Esimerkiksi

haku "a * shade of pale" tuottaa hakutulokseksi viitteet, joissa mainitaan musiikkikappaleen 'A Whiter Shade of Pale' nimi.

Yahoo! tukee kaikkia Boolean operaattoreita. Jos halutaan etsiä dokumentteja, joissa mainitaan sana amerikanraudat, mutta ei sanaa moottoripyörät, se voidaan ilmaista kolmella eri tavalla:

amerikanraudat AND NOT moottoripyörät

amerikanraudat NOT moottoripyörät

amerikanraudat -moottoripyörät

Operaattorit tulee kirjoittaa isoilla kirjaimilla, mutta hakusanoissa Yahoo! ei tee eroa isojen ja pienten kirjaimien välillä. (Yahoo! 2006).

Asetukset-linkkiä klikkaamalla päästään määrittelemään, minkä kielisiä hakutulostuloksia halutaan sekä rajaamaan lapsilta kielletyn sisällön pois tuloksista. Käyttäjä saa myös halutessaan hakutulossivut avautumaan uudessa ikkunassa. Laajennetusta hausta löytyvät haun rajausvaihtoehdot "kaikki nämä sanat" (vastaa AND-operaattoria), "tarkassa lausekkeessa" (vastaa fraasihakua), "mikä tahansa näistä sanoista" (vastaa OR-operaattoria) sekä "ei mitään näistä sanoista" (vastaa NOT-operaattoria). Haku voidaan rajata myös päivitysajankohdan, tiedostomuodon, maan ja toimialueen mukaan. Hakukentistä Yahoo!ssa toimivat muun muassa *domain*, *intitle*, *inurl*, *link*, *linkdomain*, *site* ja *url* (Notess 2006).

Suomenkielisen Yahoo!:n hakutulossivu on hieman pelkistetty versio englanninkielisen Yahoo!:n hakutulossivusta, jossa on enemmän osia. Hakutulostulosten määrä tulossivulla on Yahoo!ssakin oletuksena kymmenen, mutta sitä voi halutessaan muuttaa Asetukset-kohdassa. Tulostulosten alla on sivusta pieni yhteenveto, jossa hakutermit näkyvät lihavoituina. Vaikka Yahoo! löytäisi käytetyn hakutermin saman sivuston useammalta sivulta, se näyttää hakutuloksessa kuitenkin vain yhden sivun kultakin sivustolta. Jos halutaan nähdä muutkin saman sivuston sivut, joilla hakutermin esiintyy, klikataan url-osoitteen perässä olevaa Lisää sivuja tältä sivustolta -linkkiä. Maksetut linkit ovat hakutuloksissa joko tulostulosten alussa

vaaleansinisellä taustalla tai sivun oikeassa laidassa Sponsoritulokset -otsikon alla. (Notess 2006; Yahoo! 2006).

Rekisteröitymällä Yahoo!-palvelun käyttäjäksi on mahdollista tilata erillinen hakupalkki, Yahoo! Toolbar, omalle koneelle. Samalla saa ilmaisen sähköpostiosoitteen. (Haasio 2003, 65).

Yahoo! ei lupaa tarkkaa aikaa, jonka kuluessa sille lähetetyt sivut lisätään hakukoneeseen, vaan se puhuu ainoastaan "useista viikoista" (Yahoo! 2006).

6.3 AltaVista

AltaVista (<http://fi.altavista.com>), jonka nimi tarkoittaa näkymää ylhäältä, on hakukoneiden edelläkävijöitä. Se julkistettiin jo vuonna 1995, ja pian siitä tuli yksi www:n käytetyimmistä ja kattavimmista hakukoneista. Se on ollut yksi keskeisimmistä Internet-tiedonhaun kehittäjistä. AltaVista esitteli muun muassa ensimmäisen konekääntäjäpalvelun www:ssä, ja se toi myös ensimmäisenä webiin audio-, kuva- ja videohaun mahdollisuuden. Lisäksi AltaVista oli ensimmäinen monikielinen hakukone. (Haasio 2003, 48-49; Overture Inc. 2006). Vuonna 2004 AltaVista myytiin Yahoo!:lle. AltaVista toimii Sunnyvalessa Kaliforniassa, ja sille on myönnetty 61 hakutoimintoihin liittyvää patenttia, mikä on enemmän kuin millekään muulle www-hakuihin erikoistuneelle yhtiölle (Overture 2006).

AltaVistan käyttäjä voi valita oman käyttöliittymänsä kielen monesta vaihtoehdosta. Valinnan pääsee tekemään klikkaamalla etusivun oikeassa yläkulmassa olevaa linkkiä, joka on merkitty alaspäin osoittavalla nuolella. Suomenkielisen AltaVistan etusivulla on sanahaun ohella kuva-, MP3- ja ääni- sekä videohaku. Klikkaamalla linkkiä "englanti, suomi" käyttäjä pääsee valitsemaan, minkä kieltä sivuja hän haluaa hakutulokseen. Oletuskielinä tässä ovat englanti ja suomi. Babel Fish -käännös -linkin takaa löytyy kääntäjä, jonka avulla voidaan kääntää joko enintään 150 sanan pituisia tekstejä tai kokonaisia verkkosivustoja 19 eri kieliparin välillä. Babel Fish kääntää enintään 10 kt yksittäisestä web-sivusta. Koska Babel Fish on automaattinen kääntäjä, se toimii parhaiten silloin, kun

käännettävä teksti on kieliopillisesti oikein. Slangi, väärin kirjoitetut sanat, huonosti sijoitetut välimerkit ja monimutkaiset tai pitkät lauseet vaikeuttavat sivun kääntämistä. Valitettavasti Babel Fish ei osaa kääntää tekstiä suomeksi tai suomesta vieraalle kielelle. (Haasio 2003, 147-148; Overture 2006).

AltaVista tukee fraasihakua, joten henkilönnimet, sanonnat yms. hakutermit (esim. "Satakunnan ammattikorkeakoulu", "lumi teki enkelin eteiseen"), joiden halutaan esiintyvän tietyssä muodossa, kannattaa laittaa lainausmerkkien sisään. Fraasihaku löytyy myös Lisähaku-linkin takaa, ja siellä se korvaa lainausmerkit. Yksi AltaVistan hyvistä puolista on, että se tukee sanojen katkaisua, mikä mahdollistaa muun muassa sanojen eri taivutusmuotojen huomioon. Katkaisumerkkinä toimii tähti, jota voidaan käyttää myös fraasihaussa termin korvaajana. AltaVista ei tee eroa sille, onko hakusana kirjoitettu isoilla vai pienillä kirjaimilla. Jos halutaan, että esimerkiksi iso alkukirjain esiintyy haussa ja pienellä alkukirjaimella alkavaa samaa sanaa (esimerkiksi Saksa/saksa) ei saa esiintyä, termi tulee kirjoittaa lainausmerkkeihin ("Saksa"). Jos hakusanaan lisätään korkomerkki, AltaVista etsii vain sanoja, joissa on kyseinen korkomerkki. Jos korkomerkkiä ei lisätä, AltaVista etsii sekä korkomerkillisiä että korkomerkittömiä sanoja. (Haasio 2003, 51; Overture 2006).

AltaVistan oletusoperaattori on AND. AND-operaattorin sijaan voidaan käyttää plusmerkkiä ja NOT-operaattorin sijaan miinusmerkkiä. Lisähaussa "kaikki nämä sanat" -vaihtoehto vastaa AND-operaattoria, "mikä tahansa näistä sanoista" OR-operaattoria ja "eikä yhtään näistä sanoista" NOT-operaattoria. NOT-operaattorin kanssa ei tarvitse käyttää AND-sanaa. Myös läheisyysoperaattorin NEAR pitäisi toimia AltaVistassa, vaikka sitä ei hakukoneen omassa ohjeessa jostain syystä mainitakaan. Haun rajaamisessa on mahdollista käyttää myös hakukenttiä *domain*, *host*, *link*, *title* ja *url*. Lisähaussa rajaus voidaan tehdä lisäksi muun muassa tulosten kielen, päivämäärän ja tiedostotyypin mukaan. (Haasio 2003, 51; Overture 2006).

Asetukset-linkkiä klikkaamalla päästään valitsemaan, mitä tietoja hakutulostulostalla näytetään. Sivun kuvauksen ja url-osoitteen lisäksi voidaan tuloksesta näyttää sivun koko ja kieli, ja hakutermin voidaan lihavoida hakutuloksissa. Oletuksena

AltaVista tulostaa hakutulossivulla kymmenen tulosta, mutta käyttäjä voi muuttaa tulosten määrää. Asetuksissa on myös perhesuodatin, jolla on mahdollista suodattaa pois epämiellyttävä materiaali hakutuloksista. Käyttäjä voi tallentaa tekemänsä asetukset sekä halutessaan myös palauttaa oletusasetukset. (Overture 2006).

AltaVistan hakutuloslistan alussa ovat maksetut linkit, jotka näytetään Sponsoroitujen tulokset -otsikon alla. Hakutuloksissa url-osoitteen alla olevan Lisää sivuja kohteesta -linkin klikkaaminen antaa lisää sivuja kyseiseltä sivustolta tai palvelimelta. Mikäli käytetään AltaVistan englanninkielistä käyttöliittymää, hakutulosten yhteydessä on Translate-linkki silloin, kun kone pystyy kääntämään kyseisen sivuston. (Overture 2006).

AltaVista olettaa, että "sivusto taitetaan kokoon" (*site collapse*). Toiminnon saa pois päältä lisähakusivulla ottamalla ruksin pois sitä koskevasta laatikosta. Sivuston kokoon taittamisella tarkoitetaan sitä, että AltaVista näyttää enintään kaksi hakutulosta sivustoa kohden. Kun sivuston kokoon taittaminen ei ole käytössä, AltaVista näyttää kaikki hakutulokset ja sisentää saman sivuston sivut ensimmäisen hakutuloksen alle relevanssijärjestyksessä. (Haasio 2003, 54).

Myös AltaVistalla on oma hakupalkkinsa, AltaVista Toolbar, jonka käyttäjä voi asentaa koneelleen.

Sivustojen lisäämisestä AltaVistaan vastaa Yahoo! Search Technology -teknologia, aivan kuten sen omistajan Yahoo!:nkin hakukoneessa (Overture 2006).

6.4 Live Search

Microsoftin virallinen hakukone Live Search (<http://www.live.com>) tunnetaan myös nimillä Windows Live Search tai lyhyemmin live.com. Microsoftin tavoitteena on, että Windows Vistaan integroitava Live Search olisi tulevaisuudessa

toiseksi suosituin hakukone Googlen jälkeen ja ohittaisi siis suosiossa Yahoo!n. (KotiMikro, viitattu 04.11.2006; NoframE Oy 2006; Notess 2006).

Hakukoneen etusivulla - joka on ulkoasultaan miellyttävän pelkistetty - on sana-haun ohella kuva- ja uutishaku. Englanninkielisessä käyttöliittymässä on lisäksi Windows Live Local -karttapalvelu sekä QnA-palvelu, joka on vielä beta-asteella. QnA on kysymys-vastaus-palvelu, jossa kuka tahansa voi kysyä mitä tahansa, ja mahdollisena vastaajana on joku tai jotkut miljoonista muista yhteisöön kuuluvista. (Microsoft 2006; NoframE 2006; Notess 2006).

Perushauissa hakusanojen kirjainkokoja ei oteta huomioon. Välimerkkejä tai hyvin yleisiä sanoja, kuten englannin kielen artikkelit *a*, *and* ja *the*, ei huomioida, elleivät ne ole lainausmerkkien sisällä. Hakulaatikkoon voidaan kirjoittaa enintään 150 merkkiä välilyönnit mukaan luettuna. (Microsoft 2006).

Live Search tukee kaikkia Boolean operaattoreita sekä fraasihakua. Oletus-operaattorin AND sijaan voidaan käyttää &-merkkiä ja OR-operaattorin sijaan |-merkkiä. Sekä NOT että AND NOT toimivat. Operaattorit tulee kirjoittaa isoilla kirjaimilla. NOT-operaattorin tilalla voidaan käyttää myös miinusmerkkiä, mutta sen ja hakutermin väliin ei saa jättää välilyöntiä. Laittamalla plusmerkki hyvin yleisen sanan, kuten sanan "the" eteen, voidaan sisällyttää tuo sana hakuun. Haku voidaan myös rajata eri hakukenttiin, kuten *filetype*, *link*, *site* ja *url*, mutta ei hakukenttään *title*. Katkaisua Live Search ei tue, eikä tähdellä voi myöskään korvata kokonaista sanaa fraasihaussa. (Notess 2006).

Hakutuloksissa löydetyistä viitteistä esitetään otsikko, lyhyt yhteenveto, jossa hakutermit ovat lihavoituina, url-osoite sekä linkki välimuistiin tallennetulle sivulle, jonka yhteydessä kerrotaan myös, milloin hakuohjelma viimeksi kävi kyseisellä sivulla. Oletuksena Live Search näyttää kymmenen viitettä kerrallaan, mutta määrää päästään muuttamaan klikkaamalla hakutulossivulla olevaa Asetukset-linkkiä. Vaihtoehdot ovat 10, 15, 30 tai 50 tulosta sivua kohti. Live Search ryhmittää hakutuloksissa saman sivuston tulokset niin, että parhaiten hakutermiä vastaava sivu näytetään ensin ja loput saman sivuston sivut sen alla sisennettynä. Oletuksena se näyttää vain kaksi tulosta sivustoa kohti. Tätäkin mää-

rää voidaan muuttaa Asetukset-sivulla, ja vaihtoehdot ovat 1, 2 tai 3 tulosta sivustoa kohti. (Notess 2006). Hakutuloksissa url-osoitteen alla olevan Näytä lisää tuloksia kohteesta -linkin klikkaaminen antaa lisää sivuja kyseiseltä sivustolta tai palvelimelta. Maksetut linkit esitetään heti hakutuloksien alussa harmaalla taustalla, Sponsoroidut sivustot -tekstin näkyessä sivun oikeassa laidassa.

Asetukset-sivulla on lisäksi mahdollista valita hakukoneen käyttöliittymän esityskieli sekä sopimattomien sivustojen suodatustapa, joka voi olla tiukka, normaali tai pois päältä. Oletusarvona on normaali. Hakua voidaan rajoittaa kielen mukaan, ja jos käyttäjä haluaa, linkit avataan aina uudessa selainikkunassa. (Microsoft 2006).

Lisäasetukset-linkki löytyy hakutulossivun yläosasta Asetukset-linkin vierestä, ei siis etusivulta, mikä on Live Searchin heikkous. Lisäasetuksilla voidaan haku rajata maan tai alueen (vastaa hakukenttää *loc*), kielen (vastaa hakukenttää *language*) sekä sivuston tai toimialueen (vastaa hakukenttää *site*) mukaan. Haku on lisäksi mahdollista rajoittaa sivuihin, joilta on linkki tietylle sivulle (vastaa hakukenttää *link*). Lisäasetuksista löytyvät myös rajausvaihtoehdot Kaikki nämä ehdot (AND), Mikä tahansa näistä ehdoista (OR), Tämä lause tässä muodossa (fraasihaku) sekä Ei mitään näistä ehdoista (NOT). (Microsoft 2006).

Hakutulokset lajitellaan relevanssin mukaan. Lisäasetusten Tulokset, joiden luokitus on -välilehdellä on kolme liukusäädintä, joilla käyttäjä voi painottaa hakutuloksien järjestykseen vaikuttavia tekijöitä. Näin samasta hausta voidaan saada eri tulokset. Liukusäädinten avulla voidaan valita seuraavia vaihtoehtoja:

- Päivitetty viime aikoina: Jos halutaan painottaa äskettäin hakuindeksiin lisätyjä sivustoja, siirretään vasemmanpuoleista liukusäädintä ylöspäin.
- Hyvin suosittu: Jos halutaan painottaa sivustoja sen mukaan, miten monessa muussa sivustossa on niihin linkki, siirretään keskimmäistä liukusäädintä ylöspäin.
- Vastaa suunnilleen: Jos halutaan painottaa eniten hakusanojen ja tulosten välistä vastaavuutta, siirretään oikeanpuoleista liukusäädintä alaspäin. Vastaa suunnilleen -valinta ohittaa kahden ensimmäisen liukusäätimen luokitukset.

(Microsoft 2006).

Myöskään Live Searchin sivuilla ei anneta lupaus siitä, miten nopeasti uudet sivut löytyvät hakukoneesta, vaan prosessin sanotaan kestävän useita viikkoja (Microsoft 2006).

6.5 Wwww.fi

"Hakukone joka ymmärtää suomalaista". Kyseisellä lauseella mainostaa itseään sekä televisiossa että www-sivuilla näkyvissä bannereissa Suomen vanhin, äskettäin täysin uudistettu hakukone www.fi (<http://www.fi>), joka on erikoistunut löytämään käyttäjän kannalta olennaiset suomalaiset www-sivut. Palvelu perustuu Fonectan TeliaSonera Finlandilta joulukuussa 2005 ostamaan hakukoneeseen. Uuden palvelun hakumoottori on peräisin norjalaiselta Fast-yhtiöltä, ja teknisenä yhteistyökumppanina on Tietoenator. (Fonecta 2006; Haakana 2006).

Automaattisen tiedonkeräämisen lisäksi www.fi:ssä täydennetään hakuvastauksia paikallistuntemukseen perustuen. Paikallinen sisältöymmärrys pohjautuu muun muassa palvelun ytimessä toimivaan sisältötyöryhmään, joka seuraa suomalaisille ajankohtaisia aiheita reaaliajassa ja kääntää kerätyn ymmärryksen suomalaisten kannalta paremmiksi hakutuloksiksi. Palvelun tietosisältöä päivitetään sekä automaattisesti että sisältötyöryhmän havaintojen perusteella jokaisena arkipäivänä. Hakutulosten kärkeen nostetaan sivuja seuraavin perustein:

- Hakusana löytyy sivun otsikosta.
 - Hakusana esiintyy sivulla monta kertaa. Toisaalta hyvin yleiset sanat eivät nosta sivun sijoitusta. Mitä harvinaisempi sana ja mitä useamman kerran se esiintyy yksittäisellä sivulla, sitä todennäköisemmin sivu listataan hakutulosten kärkipäähän.
 - Hakusana löytyy sivun url-osoitteesta.
- (Fonecta 2006).

Haku ei palauta kaikkia löytyneitä viitteitä, vaan ainoastaan 4 000 ensimmäistä (Fonecta 2006).

Www.fi -hakukoneen etusivulla on sanahaun ohella uutishaku, josta löytyvät uusimmat uutiset kahden viikon ajalta, Picsearchin toteuttama kuvahaku sekä linkkirjasto, jonka avulla voidaan etsiä www-sivuja eri aihealueisiin luokitelluista sivuista. Yritystä voidaan hakea erillisellä haulla Fonecta Oy:n yritystiedoista. Käyttäjä voi myös valita, haetaanko sivuja kotimaasta vai ulkomailta. Ulkomaisia palveluja haettaessa hakutulokset tuottaa Gigablast-palvelu. (Fonecta 2006; Haakana 2006).

Mainospaikat myydään www.fi:ssä aikahinnoittelulla. Mainospaikan saa edullisimmillaan 360 eurolla, jolla yritys saa tekstimuotoisen, linkin sisältävän mainoksen valittuun hakutulossivuun kolmeksi kuukaudeksi. Www.fi -palvelussa asiakkaat eivät itse tee tai hallinnoi mainoksiaan, vaan niitä tekee ja myy Fonectan myynti. Mainospaikkoja tietyn haun tulossivulla on kolmesta viiteen kappaletta. Toisin kuin esimerkiksi Googlessa, mainospaikkojen myynti tapahtuu myyntijärjestyksessä: ensimmäiseksi paikan ostanut saa oman mainoksensa ensimmäiseksi näkyviin. (Haakana 2006). Maksetut linkit näkyvät heti hakutulostilan alussa vaaleansinisellä taustalla Sponsoroidut tulokset -otsikon kera.

Www.fi -hakukone tukee kaikkia Boolean operaattoreita, fraasihakua sekä katkaisua. Oletusoperaattori on tässäkin hakukoneessa AND, josta voidaan käyttää myös suomenkielistä versiota JA. OR-operaattorin sijaan TAI kelpaa aivan yhtä hyvin. NOT-operaattorin sijaan voidaan käyttää vaihtoehtoja AND NOT, EI, MUTTEI sekä miinusmerkkiä. Haku voidaan rajata myös hakukenttien *site* (voidaan käyttää myös suomenkielistä sanaa *palvelin*), *title* (voidaan käyttää myös suomenkielistä sanaa *otsikko*) ja *url* mukaan. Www.fi:ssä ei ole merkitystä, kirjoitetaanko hakusana isoilla vai pienillä kirjaimilla. (Fonecta 2006).

Yksi www.fi:n erikoisuuksista on ns. ennakoiva tekstin syöttö, joka ehdottaa käyttäjälle hakusanan kirjoittamisvaiheessa sanoja, joita palvelussa on käytetty (Fonecta 2006). Jos esimerkiksi aletaan kirjoittaa hakulaatikkoon sanaa "cockerspanieli", näyttää hakukone listaa jo kirjoitetuilla kirjaimilla alkavista haku-

sanoista (coca-cola, cocktail-lasi jne.), kunnes lopulta viiden kirjaimen jälkeen ykkösvaihtoehdoksi kohoaa cockerspanieli. Tämä toiminto vastaa Googlen Suggest -palvelua, ja se voidaan halutessa kytkeä pois päältä Omat asetukset -sivulla.

Hakutulossivulla esitetään sivun otsikon lisäksi lyhyt tiivistelmä sivusta, jossa hakusanat on lihavoitu, url-osoite sekä Lisää tuloksia sivusta -linkki, jota klikkaamalla saadaan tulokseksi lisää sivuja kyseiseltä sivustolta. Oletuksena hakukone näyttää kymmenen hakutulosviitettä sivua kohti, mutta tätä voidaan muuttaa Omat asetukset -sivulla. Kyseisellä sivulla on mahdollista myös rajata hakutuloksista pois lapsille sopimaton sisältö, määrittää tulosten kieli sekä valita, haluako hakutulosviitteiden avautuvan uuteen ikkunaan.

Tutkimuksen hakukoneista www.fi on ainut, joka antaa tarkkaa informaatiota siitä, minkä ajan kuluessa uudet sivut lisätään hakukoneeseen: sivusto luvataan lisätä palveluun niinkin pian kuin kahden viikon kuluessa lähettämisestä (Fonecta 2006). Tämä johtunee siitä, että suomalaisen hakukoneen jonot ovat huomattavasti ulkomaisten hakukoneiden jonoja lyhyemmät.

7 TUTKIMUKSESSA KÄYTETYT TUTKIMUSMENETELMÄT

Arvioin hakukoneita tekemällä koehakuja, sillä se on mielestäni ainut hyvä keino mitata hakukoneiden toiminnallisia ominaisuuksia. Koehauissa käytetään ainoastaan perushakua siksi, että suurin osa tiedonhakijoista käyttäneekin pelkästään sitä menemättä lisähaun puolelle.

Ennen varsinaisten koehakulauseiden käyttöä arvioin hakukoneita kriteereillä, joita myös Hellstén (2004) ja Perkiömäki (1997) ovat käyttäneet omissa hakukonevertailuissaan:

1. hakukone kertoo, montako sivustoa se sisältää ja luku löytyy avaussivulta
2. käyttäjälle kerrotaan haun kesto aika
3. käyttäjälle kerrotaan löydettyjen hakutulosten määrä
4. hakukone kertoo hakutuloksien yhteydessä, mitä hakusanaa käytettiin
5. hakukone kertoo, milloin löydetty sivu on viimeksi päivitetty
6. hakukone osaa korjata käyttäjän tekemiä kirjoitusvirheitä
7. hakukone tunnistaa skandit ja é-merkin
8. jos hakukone ei löydä mitään
 - a) käyttäjälle kerrotaan asiasta
 - b) käyttäjää neuvotaan tarkastamaan hakutermien kirjoitusasu
 - c) käyttäjää kehoitetaan muotoilemaan haku uudelleen
 - d) hakukone ehdottaa jotain muuta

Tulokset ovat taulukossa 3 numeroituna yhdestä kahdeksaan. Jos hakukoneessa on kyseinen ominaisuus, sen kohdalla on plusmerkki, mutta jos hakukoneesta puuttuu kyseinen ominaisuus, sen kohdalla on miinusmerkki.

Kohta 6 on jaettu kohtiin a ja b. Kohdassa a laitan hakusanaksi "kolestroli", koska monet kirjoittavat sanan "kolesteroli" sillä tavoin virheellisesti. Kohdassa b käytän hakusanaa "zyskovitz" odottaen, että hakukone osaisi korjata sen muotoon "zyskowicz".

Kohdassa 7 on alakohdat a, b, ja c. Kohdassa a hakusanana on "räikkönen", jotta nähtäisiin, ymmärtävätkö hakukoneet kirjaimia ä ja ö. Kohdassa b testataan Turun ruotsinkielistä nimeä "åbo" käyttämällä, käsittävätkö hakukoneet å-kirjaimen. Kohdassa c hakutermiä on "frangén", jotta testaukseen saadaan mukaan myös é-merkki.

Kohdassa 8 kirjoitan hakulaatikkoon "xwfrpx", koska se ei ole oikea sana, eikä mikään hakukone löydä sillä mitään.

Tämän jälkeen testaan hakukoneita viidellä hakulauseella. Koska tämän tutkimuksen ensisijainen tarkoitus on selvittää, miten hyvin hakukoneet tukevat Boolean hakuja, muotoilin hakulauseet sellaisiksi, että niissä käytetään kaikkia Boolean

operaattoreita sekä lisäksi sulkumerkkejä ja fraasihakua. Poikkeuksena on hakulause 1, jossa hakusanat kirjoitetaan sellaisinaan suoraan hakulaatikkoon ilman Boolean operaattoreita tai muita merkintöjä - tämä lienee se hakutapa, jota hakukoneiden käyttäjät tavallisimmin käyttävät. Kaikki käyttämäni hakulauseet ovat olleet minun tai tuttavieni todellisia tietotarpeita, eli ne eivät ole vain tätä tutkimusta varten väkisin keksittyjä "keinotekoisia" hakuja. Otin tarkoituksella tutkimukseen mukaan eri aihealueisiin liittyviä hakulauseita. Osa hakulauseista on englannin-, osa taas suomenkielisiä, riippuen siitä, kumpaa kieltä olen missäkin haussa pitänyt tarkoituksenmukaisena.

Hakulauseet ovat:

1. haight ashbury

Haetaan tietoa hippien 60-luvulla suosimasta kaupunginosasta.

2. suklaa AND reseptit

Haetaan ruokareseptejä, joissa käytetään suklaata.

3. nielsenin AND (lista OR heuristiikat)

Haetaan tietoa siitä, miten käytettävyydguru Jakob Nielsen määrittelee käytettävyyden.

4. goran visnjic AND (movies OR films) AND NOT er

Haetaan tietoa elokuvista, joissa näyttelijä Goran Visnjic on ollut mukana, mutta ei hänen roolistaan ER-televisiosarjassa.

5. "kultainen noutaja" AND kennel

Haetaan tietoa kenneleistä, joissa kasvatetaan kultaisia noutajia.

AND on oletusoperaattori kaikissa tutkimistani hakukoneista, joten käytännössä jätän sen hakulauseista pois tarpeettomana. AND NOT -operaattorin sijaan käytän joissakin hakukoneissa miinusmerkkiä.

Koska millä tahansa näistä hakulauseista saadaan tulokseksi tuhansia viitteitä, en voi käydä niitä kaikkia läpi, vaan tarkistan ainoastaan kymmenen ensimmäistä. Sponsoroidut linkit- tms. otsikon alla olevat maksetut linkit jätän pois. Löytyneet viitteet arvioin ja pisteytän relevanssiasteen mukaan osittain samalla tavalla kuin Hellstén (2004) tutkimuksessaan:

- 3 pistettä Löydetty sivu käsittelee nimenomaan haettua asiaa ja/tai on muutoin erittäin hyödyllinen haun kannalta.
- 2 pistettä Haettu asia esiintyy löydetyllä sivulla, mutta sitä käsitellään vain pintapuolisesti.
- 1 piste Haettu asia vain mainitaan sivulla.

Lisäksi käytän seuraavia merkintöjä:

- ? Haettua asiaa ei mainita sivulla lainkaan.
- Linkki ei johda mihinkään tai ei toimi.
- O Linkki on päällekkäinen jonkin toisen tuloslinkin kanssa.

Sivujen saamat pisteet lasketaan lopuksi yhteen.

Jos hakutulokset menevät päällekkäisissä linkeissä samalle sivustolle ja täsmälleen samalle sivulle, annan pisteitä vain sille sivulle, joka näistä on hakutuloslistalla ensimmäisenä, ja laitan sille merkinnän O. Muut jätän pisteittä ja laitan niille merkinnän O-O. Samalla tavalla kohtelen niitä dokumentteja, joiden sisällöt ovat keskenään samanlaiset, vaikka kyseessä olisikin eri sivusto tai sivu. Tämän perustelen sillä, ettei tiedonhakijalle ole hyötyä kahdesta eri dokumentista, jos niiden sisällöt ovat yhtäläiset.

Jos taas kyseessä on sama sivusto, mutta eri alisivu, lasken kaikki ensin löydetyn sivun pisteet mukaan ja laitan sille merkinnän O. Sen jälkeen tulevat samalle sivustolle johtavat viitteet saavat merkinnän O/2, ja niiden saamista pisteistä vähen-

netään puolet. Jos päällekkäisiä linkkejä löytyy useita, merkitsen ne pareittain: O ja O/2 vastaavat toisiaan, samoin O^2 ja $O/2^2$ jne.

Sivujen relevanssin määrittäminen on luonnollisesti melko subjektiivinen mittari. Uskon kuitenkin, että se antaa hakukoneiden käyttäjille hyödyllistä tietoa hakukoneiden ja niiden löytämien sivujen laadusta.

8 KOEHAUT JA TULOKSET

Arvioin hakukoneet ensin edellisessä luvussa esittelemäni kahdeksan kriteerin mukaan. Tulokset ovat taulukossa 3.

	AltaVista	Google	Live Search	www.fi	Yahoo!
1	-	-	-	-	-
2	-	+	-	-	+
3	+	+	+	+	+
4	-	+	+	+	+
5	-	-	-	-	-
6 a	-	+	-	+	-
b	-	+	-	-	-
7 a	+	+	+	+	+
b	+	+	+	+	+
c	+	+	+	+	+
8 a	+	+	+	+	+
b	+	+	+	+	+
c	+	+	+	+	+
d	+	-	+	+	+
Yht.	8/14	11/14	9/14	10/14	10/14

Taulukko 3 Kohtien 1-8 tulokset (10.11.2006).

Yksikään hakukoneista ei ilmoittanut avaussivullaan sisältämiensä dokumenttien lukumäärää, minkä havaitsin jo lukua 6 kirjoittaessani. Haun kestoajan kertoivat ainoastaan Google ja Yahoo!, joista ensin mainittu suoritti haun nopeammin. Löydettyjen hakutulosten määrän ilmoittivat kaikki hakukoneet. AltaVistaa lukuun ottamatta kaikki hakukoneet kertoivat hakutulossivulla, mitä hakusanaa/sanoja käytettiin - tosin käytetyt hakusanat näkyivät AltaVistan, niin kuin muidenkin hakukoneiden, tulosviitteiden kuvausteksteissä lihavoituina. En kuitenkaan antanut AltaVistalle plussia kohdassa 4. Mikään hakukoneista ei valitettavasti kertonut tulosviitteiden yhteydessä, milloin ne on viimeksi päivitetty.

Sanan "kolestroli" neuvoivat korjaamaan oikeaan muotoon "kolesteroli" vain Google ja www.fi, muille näytti kelpaavan väärin kirjoitettu muoto. "Zyskovitz" -sanana osasi korjata muotoon "zyskowicz" vain Google. Wwww.fi:kin ymmärsi, että sana on kirjoitettu väärin, mutta se ehdotti oikeaksi kirjoitusasuksi "zyskovicz", joten en antanut sille plussia kohdassa 6b. Ä-, ö- ja å-kirjaimet eivät tuottaneet hakukoneille ongelmia, niin kuin ei myöskään é-merkin sisältämä frangén-sana - tosin www.fi kysyi varmuuden vuoksi hakutulostilan yläreunassa: "Tarkoititko: frangin?".

Kirjainyhdistelmää "xwfrpx" ei löytänyt mikään hakukoneista, ja kaikki myös kertoivat sen hakijalle sen sijaan, että olisivat vain tulostaneet tyhjän hakutulostilan. Kaikki myös ehdottivat hakijalle haun uudelleenmuotoilua ja kirjoitusasun tarkistamista. Google oli ainut, joka ei ehdottanut näiden lisäksi mitään muuta. Yahoo! antoi linkin Yahoo! -haun tukikeskukseen, josta hakija voi saada hyödyllisiä hakuohjeita. AltaVista ehdotti maakohtaisen haun vaihtamista valintanapilla maailmanlaajuiseksi sekä hakusanan hakemista kuva-, video-, audio- ja uutishauista. Live Search antoi linkin sekä hakukoneen web-hakuohjeisiin että sivulle, jonka kautta käyttäjä voi lähettää hakukoneelle sellaisen url-osoitteen, jonka hän tietää olevan olemassa, mutta jota Live Search ei kuitenkaan löydä. Samanlaista mahdollisuutta tarjosi www.fi.

Näillä kriteereillä arvioituna Google otti niukan voiton muista hakukoneista. Toisen tilan jakavat Yahoo! ja www.fi, joiden jälkeen tulee Live Search yhtä pistettä viimeiseksi jätettyä AltaVistaa parempana. Edellä mainituilla kriteereillä ei kui-

tenkaan voida arvioida, mikä hakukoneista antaa käyttäjän kannalta relevantteimmat hakutulokset, joten seuraavaksi hakukoneita testataan aiemmin ilmoittamani hakulauseilla.

	AltaVista	Google	Live Search	www.fi	Yahoo!
1	http://www.sfgate.com/traveler/guide/sf/neighborhoods/haight.shtml	http://www.sfgate.com/traveler/guide/sf/neighborhoods/haight.shtml	http://www.hafci.org/	http://www.haight-ashbury-music.com/	http://www.sfgate.com/traveler/guide/sf/neighborhoods/haight.shtml
2	http://en.wikipedia.org/wiki/Haight-Ashbury	http://en.wikipedia.org/wiki/Haight-Ashbury	http://www.haight-ashbury-music.com/	http://www.hafci.org/	http://www.rockument.com/haimg.html
3	http://www.haight-ashbury-music.com/	http://www.sanfranciscobay.com/haightashbury/	http://en.wikipedia.org/wiki/Haight-Ashbury	http://www.rockument.com/haimg.html	http://www.haight-ashbury-music.com/
4	http://www.rockument.com/haimg.html	http://www.rockument.com/links.html	http://www.rockument.com/ha.html	http://landing.trafficz.com/index.php?domain=www.haight-ashbury.com	http://en.wikipedia.org/wiki/Haight-Ashbury
5	http://www.sanfranciscobay.com/haightashbury/	http://www.rockument.com/haimg.html	http://www.rockument.com/haimg.html	http://www.sfmuseum.org/hist1/habib.html	http://www.rockument.com/links.html
6	http://www.haightashbury.org/	http://www.hafci.org/	http://www.sanfranciscobay.com/haightashbury/	http://www.rockument.com/ha.html	http://landing.trafficz.com/index.php?domain=www.haight-ashbury.com
7	http://onlyinsanfrancisco.com/neighborhoods/description.asp?nid=28	http://www.haightstreetfair.org/	http://video.msn.com/v/us/v.htm?g=791006ec-7d68-468a-8a2d-752632305492&f=49	http://www.thefoodprogram.org/	http://www.sanfranciscobay.com/haightashbury/
8	http://www.rockument.com/links.html	http://www.haight-ashbury-music.com/	http://www.haightashbury.org/	http://www.sanfranciscobay.com/haightashbury/	http://www.hafci.org/
9	http://www.haightstreetfair.org/	http://www.haightashbury.org/	http://www.sfgate.com/traveler/guide/sf/neighborhoods/haight.shtml	http://www.ac.wwu.edu/~stephan/Maps/Map6.Haight.html	http://www.sfmuseum.org/hist1/habib.html
10	http://www.haightashburybeat.com/	http://www.answers.com/topic/haight-ashbury	http://www.redvic.com/haight.html	http://www.collectibles-shoppe.com/brian-baker-collection/haight-ashbury.html	http://www.ac.wwu.edu/~stephan/Maps/Map6.Haight.html

Taulukko 4 Haku 1: haight ashbury. Tulossivujen www-osoitteet. (10.11.2006).

	AltaVista	Google	Live Search	www.fi	Yahoo!
1	3	3	2	2	3
2	3	3	2	2	3 O
3	2	3	3	3 O	2
4	3 O	3 O	3 O	1	3
5	3	3 O/2	3 O/2	2	3 O/2
6	3	2	3	3 O/2	1
7	2	2	?	2	3
8	3 O/2	2	3	3	2
9	2	3	3	3	2
10	2	3	2	1	3
Yht.	24,5	25,5	22,5	20,5	23,5

Taulukko 5 Haku 1: haight ashbury. Tulossivuille annetut pisteet. (10.11.2006).

Eniten pisteitä sai Google, jonka jälkeen tulevat järjestyksessä AltaVista, Yahoo! ja Live Search. Www.fi jätettiin viimeiseksi.

Vaikka Google vastasi tiedonhakijan odotuksiin parhaiten, voidaan taulukosta 5 nähdä, ettei senkään pisteillä ole laskeva järjestys: pidän viitteitä 9 ja 10, joilla esitellään kattavasti Haight Ashburyn historiaa ja nykypäivää, tietotarpeeni kannalta parempina kuin viitteitä 6 (Haight Ashbury Free Clinicsin sivusto) ja 7 (vuosittaisen katutapahtuman, Haight Ashbury Street Fairin, sivusto).

Suurimmaksi osaksi hakukoneet löysivät samoja sivuja, mutta Yahoo!:ta lukuun ottamatta jokaisen kärkikymmenikössä oli silti ainakin yksi sivu, jota minkään muun hakukoneen kärkikymmenikössä ei ollut. Google, Yahoo! ja AltaVista rankkasivat parhaaksi saman sivun, jota voidaan kutsua varsinaiseksi löydöksi: San Francisco Chronicle -lehden sivulla kerrotaan hieman Haight Ashburyn historiasta sekä annetaan matkailijoille vinkkejä lukuisista nähtävyyksistä, ostospaikoista, ravintoloista yms., joihin siellä kannattaa tutustua. Jos tiedonhakija suunnittelisi matkaa Haight Ashburyyn, olisi tuo sivu hänelle epäilemättä hyvin arvokas. Kyseinen sivu oli myös Live Searchin tuloksissa, tosin vasta 9. sijalla, mitä ihmettelen.

Annoin kolme pistettä myös Wikipedian Haight Ashbury -sivulle, sillä vaikka monet - minä itse mukaan lukien - suhtautuvat Wikipediasta löytyvän tiedon todenperäisyyteen skeptisesti, sai sieltä loistavan tietopaketin Haight Ashburysta värikuvien kera. Wikipedian sivu oli Googlen ja AltaVistan tuloksissa toisena, Live Searchin tuloksissa kolmantena ja Yahoo!:n tuloksissa neljäntenä.

Oma suosikkini löydetyistä sivuista on viite, joka oli Googlessa sijalla 3 ja muissa hakukoneissa sijojen 5-8 välillä. Kyseessä on visuaalisesti todella näyttävä sivusto, josta löytyy tietoa Haight Ashburyn kaupoista ja nähtävyyksistä sekä paikallisista taiteilijoista yms., tapahtumakalenteri, valokuvia ja paljon muuta. Samalla voi nauttia hienosta grafiikasta, animaatioista ja 60-luvun hiteistä.

Lähes yhtä paljon pidin Rockument Channelin sivustosta, jolta olivat Googlen löytämät viitteet numero 4 ja 5. Tämä 60-luvun musiikin ystäville ehdottomasti tutustumisen arvoinen sivusto löytyy kaikkien muidenkin hakukoneiden kärkikymmeniköstä, tosin alasivut eivät välttämättä ole samoja. Sivusto tarjoaa laajan ja asiantuntevan katsauksen hippien aikakauden musiikkiin, kiinnostavia linkkejä sekä runsaasti aiheeseen liittyvää taidetta.

Yahoo!:n viite numero 6, joka oli myös www.fi:n viite numero 4, ei liity Haight Ashburyyn juuri mitenkään, vaikka sivun otsikko on www.haight-ashbury.com. Sivulla, joka on suomenkielinen, on vain linkkejä erinäisille kaupallisille sivuille. Yahoo!:n tuloslinkki numero 9 vei San Franciscon virtuaalisen museon sivustolle, jolla muun muassa kerrataan sanfranciscolaisen rockin historiaa kronologisessa järjestyksessä. Yksi Yahoo!:n parhaista löydöistä oli viite numero 10, joka oli myös www.fi:n viite numero 9: sivulta löytyy Haight Ashburyn kartta, johon on merkitty turistien kannalta kiinnostaviksi arvelut kohteet.

Live Searchin tulos numero 7 ihmetytti minua. Linkin klikkaus avaa MSN Video -ohjelman, joka näytti mainoksia ja huumorivideoita. Mikään sivulla ei liity Haight Ashburyyn.

Yleisesti ottaen olen positiivisesti yllätynyt siitä, miten hyviä dokumentteja hakukoneet löysivät Haight Ashburysta, mikä näkyy myös antamissani pisteissä.

Jokaisen hakukoneen kärkikymmenikössä oli useampi kolmen pisteen tulos, ja Googlen tuloksista niitä oli jopa reilusti yli puolet.

	AltaVista	Google	Live Search	www.fi	Yahoo!
1	http://www.myllynparas.fi/portal/suomi/reseptit/makeat_leivonnaiset/pullat/suklaa_kierrepullat/	http://www.pirkka.fi/ruoka/reseptit/suklaa-cookiepalat.aspx	http://www.oetker.fi/Reseptit/Pikkuleivat%20ja%20makeiset/Suklaa-kookostahdet/	http://www.ruokaohjereseptit.fi/reseptiohje/suklaa-banaanimuffinit/	http://perso.wanado.es/reseptit/suklaa_reseptit/suklaa_reseptit.html
2	http://www.santamaria.fi/servlet/se.daddy.sm.web.servlets.RecipeServlet?c=5&p=r5b&id=5466	http://www.nautiskelijat.net/wp/?cat=15	http://www.kustnatur.fi/reseptit.htm	http://elisa.net/ruokajaviini/reseptit/?id=R000174	http://www.santamaria.fi/servlet/se.daddy.sm.web.servlets.RecipeServlet?c=5&p=r5b&id=5466
3	http://elisa.net/ruokajaviini/reseptit/?id=R000174	http://www.herkut.net/resepti.php?id=876	http://www.herkut.net/luokkasivu.php?luokka=6	http://wanhat.rko.iki.fi/reseptit/suklaa.htm	http://elisa.net/ruokajaviini/reseptit/?id=R000174
4	http://www.myllynparas.fi/portal/suomi/reseptit/jalkiruokat_ja_valipalat/letut_ja_pannarit/suklaa_kauraletut	http://www.herkut.net/resepti.php?id=407	http://www.herkut.net/luokkasivu.php?luokka=7	http://www.pirkka.fi/ruoka/reseptit/kirsikka-suklaa-pikkuleivat.aspx	http://www.ruokaohjereseptit.fi/ruokaohjeet/suklaa/
5	http://www.ruokaohjereseptit.fi/ruokaohjeet/suklaa/	http://www.nettikeitto.fi/?id=1940	http://www.petsmobar.fi/fin/recepies/	http://www.rihmasto.fi/wiki/Suklaa	http://www.myllynparas.fi/portal/suomi/reseptit/makeat_leivonnaiset/pullat/suklaa_kierrepullat/
6	http://www.ruokaohjereseptit.fi/reseptiohje/appelsiininen-valkosuklaaterriini/	http://www.myllynparas.fi/portal/suomi/reseptit/makeat_leivonnaiset/taytekakut/suklaa_cookiesleivokset_tai_kakku/	http://www.maailmankaupat.fi/lehti/reseptit.html	http://www.santamaria.fi/servlet/se.daddy.sm.web.servlets.RecipeServlet?c=5&p=r5b&id=5466	http://www.ruokaohjereseptit.fi/reseptiohje/appelsiininen-valkosuklaaterriini/
7	http://elisa.net/ruokajaviini/reseptit/?id=R000570	http://www.myllynparas.fi/portal/suomi/reseptit/makeat_leivonnaiset/pikkuleivat,_keksit_ja_piparit/suklaa_cookies/	http://www.kuvakakuun.com/pages/reseptit.php	http://www.myllynparas.fi/portal/suomi/reseptit/makeat_leivonnaiset/kahvikakut/suklaa-ananaskakku/	http://elisa.net/ruokajaviini/reseptit/?id=R000570
8	http://www.pirkka.fi/ruoka/reseptit/suklaa-cappuccinomousse.aspx?sso_cookie=false	http://elisa.net/ruokajaviini/reseptit/?id=R000174	http://www.hirvelantalkkuna.fi/index.php?page=reseptit&res=suklaa	http://www.kotikokki.net/reseptit/nayta/975/Suklaa-appelsiinimousse	http://www.pirkka.fi/ruoka/reseptit/suklaa-mansikkatorrttu.aspx?sso_cookie=false
9	http://www.pirkka.fi/ruoka/reseptit/suklaa-cassissidan.aspx?sso_cookie=false	http://elisa.net/ruokajaviini/reseptit/?id=R000570	http://www.meira.fi/reseptit/kahvireseptit/kylmia_kahveja.html	http://esteri.iki.fi/~minna/reseptit/Suklaapekaanivuohenjuustotoast.html	http://www.pirkka.fi/ruoka/reseptit/suklaa-cassissidan.aspx?sso_cookie=false
10	http://www.paakari.net/reseptit.php?id=2038&cat=11	http://www.kotikokki.net/reseptit/nayta/379/suklaa-banaanipirtel%C3%B6	http://users.tkk.fi/%7Esmnummin/Suklaa/reseptit.html	http://tavaratalot.stockmann.fi/portal/suomeksi/herkku/reseptit/syysmetsan_makujuhlaa/suklaa-pahkinaleivos/	http://www.kenguru.fi/reseptit/viewrecipe.php?id=1105185514&category=

Taulukko 6 Haku 2: suklaa AND reseptit. Tulossivujen www-osoitteet. (11.11.2006).

	AltaVista	Google	Live Search	www.fi	Yahoo!
1	3 O	3	2	3	?
2	2	2	2	3	2
3	3 O ²	3 O	3 O	2	3 O
4	3 O/2	3 O/2	3 O/2	3	3 O ²
5	3 O ³	1	2	2	3
6	3 O/2 ³	3 O ²	2	2	3 O/2 ²
7	3 O/2 ²	3 O/2 ²	2	3	3 O/2
8	3 O ⁴	3 O ³	2	3	3 O ³
9	3 O/2 ⁴	3 O/2 ³	2	2	3 O/2 ³
10	-	3	3	2	2
Yht.	20	22,5	21,5	25	20,5

Taulukko 7 Haku 2: suklaa AND reseptit. Tulossivuille annetut pisteet. (11.11.2006).

Www.fi osoitti tässä haussa tehokkuutensa suomenkielisen aineiston hakijana: se päihitti toiseksi jääneen Googlen 2,5 pisteellä, mikä oli minusta yllättävän selkeä voitto. Kolmanneksi tuli Live Search yhden pisteen erolla Googleen. Yahoo! voitti viimeiseksi jääneen AltaVistan vain puolella pisteellä.

Www.fi:n voiton ratkaisi sama asia kuin AltaVistan heikon menestyksen: päällekkäiset linkit. Www.fi:ssä niitä ei tässä haussa ollut lainkaan, kun taas AltaVistan tuloksista kahta viitettä - joista toinen oli toimimaton linkki - lukuun ottamatta kaikki olivat päällekkäisiä toisen viitteen kanssa. Lisäksi www.fi:n tulokset olivat kaikki kolmen tai kahden pisteen arvoisia, kun taas esimerkiksi Googlen antamista tuloksista yksi oli vain yhden pisteen arvoinen. Yahoo!n parhaaksi rankaama tulos oli kuitenkin suurin yllätys. Toisin kuin url-osoitteen perusteella voisi odottaa, sivu on kokonaan espanjankielinen, joten siellä ei luonnollisesti mainita suomenkielisiä sanoja "suklaa" ja "reseptit".

Hakukoneiden tuloksissa oli paljon viitteitä sivujen

<http://www.pirkka.fi/ruoka/reseptit/>, <http://www.herkut.net>,

<http://www.myllynparas.fi/portal/suomi/reseptit>,

<http://elisa.net/ruokajaviini/reseptit/> ja <http://www.ruokaohjereseptit.fi> alasivuilta.

Olen antanut näiden sivustojen sivuille hyvät pisteet, koska niillä olevat reseptit vaikuttavat hyviltä ja käyttökelpoisilta. Kaikilla näillä sivustoilla on myös helppokäyttöinen reseptihaku, jolla voidaan etsiä lisää suklaa- tai muita reseptejä. Paras reseptihaku löytyy <http://www.kotikokki.net/reseptit> -sivun alisivuilta: siinä hakua voidaan rajata muun muassa valmistusajan ja reseptin työläyden mukaan sekä sen mukaan, mihin tilaisuuteen (illanistujaisiin, jouluksi, pääsiäiseksi, ylioppilajuhliin tms.) ruokaa ollaan tekemässä. Tämän vuoksi Googlen viite numero 10 ja www.fi:n viite numero 8 saivat minulta 3 pistettä.

Reseptihaku löytyy myös Yahoo!:n viitteestä numero 10, Kenguru.fi:n sivulta. Kyseisellä sivustolla on reseptien yhteydessä kerrottu niiden vaikeusaste, ja sivun alareunassa on linkki, jota klikkaamalla voi lähettää ystävilleen linkin resepteihin. Myös Santa Marian sivustolla, joka oli Yahoo!:n ja AltaVistan löytö numero 2 sekä www.fi:n löytö numero 6, on samanlainen vihjeenantomahdollisuus, samoin jo aiemmin mainitsemillani Myllyn Paras Oy:n, Herkut.netin ja Pirkan sivustoilla. Reseptivalikoiman pienuuden vuoksi sekä Kenguru.fi:n sivu että Santa Marian sivut ovat kuitenkin saaneet minulta vain 2 pistettä.

Live Searchin parhaaksi rankkaamalla sivulla on myös reseptihaku, mutta suklaaseen liittyviä reseptejä on vain muutama, minkä takia ihmettelen sen korkeaa sijoitusta. Olen kuitenkin pyrkinyt antamaan 2 pistettä niillekin sivuille, joilta löytyy vain yksi oikeanlainen resepti.

Live Searchin tuloslinkki numero 6 vei sivulle, jossa on useita suklaa-aiheisia reseptejä. Sivulla tunnutaan kuitenkin keskittymään pääasiassa reilun kaupan suklaaseen sekä yleensäkin reilun kaupan tuotteisiin, joten se ansaitsi minusta vain 2 pistettä.

Mielestäni tämän haun helmi oli Live Searchin löytö numero 10, jossa keskitytään määrän sijasta laatuun: reseptejä on vähän, mutta ne ovat sitäkin parempia ja houkuttelevampia. Itse ainakin uskon, että sacherkakun ja suklaamoussen kaltaisten klassikoiden reseptit kiinnostavat suklaan ystäviä. Sivustolla on lisäksi kerrottu suklaan historiasta ja terveysvaikutusta, kerätty suklaa-aiheisia aforismeja - sekä

listattu linkkejä sivuille, joilta löytyy lisää suklaaseen liittyviä reseptejä! Sivuston huono sijoitus kummastuttaa minua.

Www.fi:n linkki numero 5 vei sivulle, jossa on kotitekoisen suklaan resepti - ei siis reseptiä, jossa käytetään suklaata. Arvelin sen kuitenkin olevan suklaan suosijoiden mieleen, joten annoin sille 2 pistettä. Saman hakukoneen linkki numero 10 vei sivulle, jossa on runsas valikoima reseptejä sekä reseptihaku. Ongelmana oli vain se, ettei reseptihaku toiminut niin kuin sen olisi pitänyt: kun hain reseptejä hakusanalla "suklaa", se antoi tulokseksi paljon myös sellaisia reseptejä, joilla ei ollut mitään tekemistä suklaan kanssa. Tämän epämukavuuden takia vähensin viitteen pisteet kolmesta kahteen.

Hakutulosten perusteella ruuanlaitosta ja/tai leipomisesta pitävä suklaan ystävä löytäisi millä tahansa näistä viidestä hakukoneesta hyvältä kuulostavia ja kokeilemisen arvoisia reseptejä.

	AltaVista	Google	Live Search	www.fi	Yahoo!
1	http://www.cs.uta.fi/usabsem/versio2/heta_asiantuntijat_v2.doc	http://www.cs.uta.fi/usabsem/luvut/8_Korvenranta.pdf	http://www.atostek.com/kaytoliittymat.html	http://sokljoensuu.fi/aineistot/kasvatustiede/didaktiikka/valineet-ohjeet/parammetut_vee.htm	http://www.cs.uta.fi/usabsem/versio2/heta_asiantuntijat_v2.doc
2	http://www.cs.uta.fi/kurssit/usabsem/kommentit/Hetalle_opponeetit_v2.doc	http://www.atostek.com/kaytoliittymat.html	http://koti.mbnet.fi/kr7/tutkielma/index.php?PHPSESSID=2d98bc85ee559fe2a5d7d0a29f06c111&sivu=tutkielma_2&PHPSESSID=2d98bc85ee559fe2a5d7d0a29f06c111	http://matriisi.ee.tut.fi/hmopetus/hypmed04/demo/nielsen10/nielsen10.html	http://www.cs.uta.fi/kurssit/usabsem/kommentit/Hetalle_opponeetit_v2.doc
3	http://users.tkk.fi/~ptmannon/projects/wv/wkaytettavyys/heuristinen_arvio.html	http://www.student.oulu.fi/~Inissila/analyysi.htm	http://koti.mbnet.fi/kr7/tutkielma/index.php?PHPSESSID=2d98bc85ee559fe2a5d7d0a29f06c111&sivu=tutkielma_3&PHPSESSID=2d98bc85ee559fe2a5d7d0a29f06c111	http://www.cs.uta.fi/usabsem/luvut/8_Korvenranta.pdf	http://www.joensuu.fi/ajankohtaista/tiedotteet/2004/JoyDWNkaytettaavyystutkimus2.ppt
4	http://www.it.lut.fi/kurssit/03-04/010745002/lectures.html	http://www.soberit.hut.fi/T-76.115/03-04/palautukset/groups/Rajoitteiset/lu/Heuristinen_arviointi.pdf	http://www.it.lut.fi/kurssit/03-04/010745002/HTYOT/3/vaiheet.html	http://www.papunet.net/tiedostot/yleis/Kirsi_Kalenius.pdf	http://www2.uiah.fi/mediaudio/survey4/liitea1.html
5	http://www.it.lut.fi/kurssit/04-05/010745002/projects.html	http://www2.uiah.fi/mediaudio/survey4/liitea1.html	http://www.cs.uta.fi/~ov/usab/luennot/index.html	http://www.netlab.tkk.fi/opetus/s38310/04-05/Kalvot_04-05/Dty%F6_seminaari_eulenberger010305-1.pdf	http://www.it.lut.fi/kurssit/03-04/010745002/lectures.html
6	http://users.tkk.fi/~jhtikkal/dotcomrades/deliveries/i1/projektisuu nmitelma/laatukasikirja/kaytettavydenarviointi.html	http://www.cs.helsinki.fi/u/salaakso/kl2-2004/Kaytoliittymat2-Luento5-27.10.2004-Sari-A-Laakso.pdf	http://www.bensiinikauppiaat.fi/sivu/huoltamoala/	http://www4.hamk.fi/julkaisut/lataa.php?id=Mjg3fDIwMDYxMDA1&Julkaisuomintina_Session=21cf3f18b74f57a6b760e3efc56bd6d2	http://www.atostek.com/kaytoliittymat.html
7	http://www.student.oulu.fi/~Inissila/analyysi.htm	http://www.cs.helsinki.fi/u/lavuorin/nielsen/usabilityengineering.htm	http://www2.siba.fi/huiluseura/	http://kansalliskirjasto.fi/digiohjaus/20050829_Heuristinen_arviointi.pdf	http://www.student.oulu.fi/~kettumik/nnyrkki.html
8	http://www.soberit.hut.fi/T-121/T-121.2100/suomi/luentokalvot/T-121-2100_2006_6.pdf	http://www.joensuu.fi/ajankohtaista/tiedotteet/2004/JoyDWNkaytettavyystutkimus2.ppt	http://www.pinseri.com/2002/08/22/fi/	http://www.tekes.fi/julkaisut/Digi_TV_opas.pdf	http://www.cs.jyu.fi/~sakkinen/testaus/aineistot/6.5_Kaytettavyystestaus_v1.ppt
9	http://appro.mit.jyu.fi/essikurssi/testaus/t2/	http://www.cs.jyu.fi/~sakkinen/testaus/aineistot/6.5_Kaytettavyystestaus_v1.ppt	http://blogit.digitoday.fi/mobent/?p=36	http://viestintaministerio.fi/oliver/up1117-Arvid%201%202004.pdf	http://users.tkk.fi/~jparckin/nettikirjat.html
10	http://matriisi.ee.tut.fi/arvo/liitteet/tutkija101103_silius_tervakari.pdf	http://www.lib.helsinki.fi/digiohjaus/20050829_Heuristinen_arviointi.pdf	http://www.matkalla.org/linkit/linkit_hyoty.html	http://www.helia.fi/attachment/b56f00765cdb17b136e7ef855aa47df1/767e873862d8be731b639587b85c78eb/jaakko lahanna.pdf	http://ameba.lpt.fi/~toivjuss/Opinnaytetyoseminaari.ppt

Taulukko 8 Haku 3: nielsenin AND (lista OR heuristiikat). Tulossivujen www-osoitteet. (13.11.2006).

	AltaVista	Google	Live Search	www.fi	Yahoo!
1	2 O	2	2	?	2 O
2	2 O-O	2	3 O	3	2 O-O
3	3	3	3 O/2	2	2
4	3	3	3 O ²	3	3
5	3	3	3 O/2 ²	2	3
6	3	3 O	?	-	2
7	3	3 O/2	?	3	3
8	-	2	3	2 O	3
9	2	3	?	2 O-O	3
10	3	3	3	1	2
Yht.	22	25,5	17	16	23

Taulukko 9 Haku 3: nielsenin AND (lista OR heuristiikat). Tulossivuille annetut pisteet. (13.11.2006).

Tämän haun tulokset ovat hyvä käytännön esimerkki siitä, että Internetin hakukoneet - tai ainakin muutamat niistä - kelpaavat muuhunkin kuin viihteellisen aineiston hakemiseen. Erityisesti ykköseksi tulleella Googlella löytyi monta todella hyvää sivustoa. Kakkoseksi tuli Yahoo! yhden pisteen erolla AltaVistaan. Neljänneksi sijoittunut Live Search sai vain yhden pisteen enemmän kuin viimeiseksi jäänyt www.fi.

Yritin sivuja pisteyttäessäni asettua sellaisen tiedonhakijan asemaan, jonka täytyy arvioida jokin käyttöliittymä käyttäen Nielsenin heuristisen arvioinnin muistilistaa, mutta joka ei tiedä, millaisesta listasta on kyse. Hyviä pisteitä olen antanut sivuille, joista katsoin olevan tehtävän kannalta todellista hyötyä. Kolme pistettä saivat ne sivut, joilla luetellaan Nielsenin heuristiikat ja selitetään jollain tavalla, mitä ne tarkoittavat käytännössä, mutta myös ne sivut, joilta on suora linkki jollekin todella hyödylliselle sivulle. Sen sijaan ne sivut, joilla on vain listattu Nielsenin heuristiikat ilman selityksiä tai käytännön esimerkkejä, ovat saaneet minulta kaksi pistettä.

Haku antoi tulokseksi monia pdf-, doc- ja ppt-tiedostoja. Yleensä kyse oli joko opetusmateriaalista tai jonkun henkilön tekemästä tutkielmasta tms., ja sen tyyppiset dokumentit ovat verkossa useimmiten juuri edellä mainituissa muodoissa.

Googlen kärkikymmenikön kaikki viitteet olivat joko hyviä tai keskinkertaisia. Kaksi linkkiä olivat toistensa kanssa päällekkäiset, mutta yhtään toimimatonta linkkiä ei löytynyt, eikä myöskään sellaista sivustoa, joka ei olisi liittynyt Nielsenin heuristiikkoihin mitenkään. Myös Yahoo!:n kaikki viitteet saivat minulta 3 tai 2 pistettä, tosin sivut 1 ja 2 ovat käytännössä päällekkäiset: kyseessä on sama asiantuntija-arviointea käsittelevä tutkielma sillä erotuksella, että sivussa 2 on mukana opponenttien kommentit. Lähes sama dokumentti on Googlen tuloslistan numero 1, samoin www.fi:n numero 3.

AltaVistan kaikki toimivat linkit saivat minulta 3 tai 2 pistettä. Kaksi ensimmäistä linkkiä ovat samat kuin Yahoo!:n kaksi ensimmäistä linkkiä. Hyviä malleja heuristiikkojen soveltamisesta käytäntöön on Googlen viitteissä 3 (joka on myös AltaVistan viite 7) ja 4, Yahoo!:n viitteessä 7, AltaVistan viitteessä 3 sekä www.fi:n viitteissä 2 ja 4.

Tämän haun paras löytö oli mielestäni Googlen viite numero 5, joka oli myös Yahoo!:n nelonen. Kyseessä on web-käytettävyydestä tehty tutkielma, jossa on Nielsenin listasta sekä pitkä että lyhyt versio. Mukana on myös muun muassa yleistä teoriaa käytettävyydestä sekä hyödyllisiä linkkejä. Kokonaisuutta on helppo selaila sisällysluettelon ansiosta, minkä lisäksi tutkielma on mahdollista ladata omalle koneelle pdf-tiedostona. Toinen erittäin hyvä tutkielma käytettävyydestä on Live Searchin sivu 2 (ja 3). Googlen viitteessä 10 on heuristiikkojen lisäksi listattu Nielsenin iskulauseita selitysten kera. Olisin sijoittanut kyseisen sivun huomattavasti korkeammalle hakutuloksissa.

Live Searchin viitteet 6, 7 ja 9 eivät liity Nielsenin heuristiikkoihin millään tavalla, niin kuin ei myöskään www.fi:n parhaaksi rankkaama sivu. Googlen viite 10 ja www.fi:n viite 7 ovat sama dokumentti, vaikka url-osoite onkin eri. Www.fi taas tarjosi käytännössä samaa digi-tv:n palveluntekijän opasta sekä linkiksi 8 että linkiksi 9. Www.fi:n viitteestä 10 ei ole Nielsenin heuristiikoista tietoa hakeville

juurikaan hyötyä, sillä kyseessä on käytettävyystudkimuksen tiivistelmä, jossa vain mainitaan termit "heuristinen" ja "Nielsenin lista".

Tulosten perusteella uskon, että hakukoneiden löytämiin sivuihin tutustumalla sellainenkin käyttäjä, jolle Nielsenin heuristiikat ovat ennestään tuntemattomia, osaisi arvioida käyttöliittymän Nielsenin listan mukaan.

	AltaVista	Google	Live Search	www.fi	Yahoo!
1	http://movies.about.com/od/elektra/	http://movies.yahoo.com/movie/contributor/1800019268	http://2001.hiff.fi/elokuvat/elokuva.html?elokuva_id=24	http://www.123people.net/v/visnjic-goran/goran-visnjic-filmography.htm	http://movies.yahoo.com/shop?d=hc&id=1800019268&cf=ggen
2	http://movies.yahoo.com/shop?d=hc&cf=bioid=1800019268	http://www.bestprices.com/cgi-bin/vlink/actors/vhss-by-actor-Goran-Visnjic.html	http://www.aintitcool.com/talkback_display/19097	http://www.ropeofsilicon.com/profile.php?id=1964	http://www.hollywood.com/celebs/detail/id/192827
3	http://movies.about.com/cs/trailers/v/closetr042704.htm?iam=metaresults&terms=goran+visnjic	http://www.locafilm.com/acteur.php?id_artiste=7939	http://sg.movies.yahoo.com/Elektra/movie/11887/premiere-photos/	http://www.buy-movie.net/actors/Goran_Visnjic	http://movies.yahoo.com/shop?d=hc&cf=bioid=1800019268
4	http://shopping.yahoo.com/p:Goran%20Visnjic:1800019268;_ylt=A9ibyK_lbpFph4B6gdzXKMX;_ylu=X3oDMTBvdmM3bGlxBHBndANhdI93ZWJfcmVzdWx0BHNiYwNzcg--	http://www.allmovieportal.com/b/goranvisnjic.html	http://www.mrsiggl.es.com/hunks/categories.php?cat_id=33&sessionid=8b3ae1cd0de20de3434288de4228de70	http://www.precomania.com/search.php/contribid=1059798//	http://www.cinefil.com/cinefil2005/FichePersonnalite.cfm?ref=56620
5	http://us.imdb.com/name/nm0899681/photogallery	http://movies.google.com/goranvisnjic/f1059798	http://www.yee.ch/movies/E/EL/Elektra/Elektra.htm	http://www.popularq.com/articles/Actors/General/Goran-Visnjic/	http://www.hotflick.net/celebs/goran_visnjic.html
6	http://movies.yahoo.com/shop?d=hp&cf=prev&id=1808403529	http://movies2.nytimes.com/gst/movies/filmography.html?pid=226539	http://www.amazon.com/Spartacus-Complete-Miniserie-Robert-Dornhelm/dp/B0002PYTB4	http://www.star-run.de/Goran_Visnjic.htm	http://www.allocine.fr/personne/fichepersonne_gen_cpersonne=21637.html
7	http://us.imdb.com/gallery/granitz/1004/wi20010121_GoranVisnjic_Michelson_139099.jpg	http://entertainment.lycos.com/movies/celebrity.php?id=16856	http://www.amazon.com/Practical-Magic-Sandra-Bullock/dp/0790740060	http://www.dvdbooty.com/stars/goran-visnjic/	http://www.arsenick.net/Acteur-Goran-VISNJIC-1297.htm
8	http://www.themovieinsider.com/c/goran-visnjic/	http://www.allmoviephoto.com/c/GoranVisnjic.html	http://movies.yahoo.com/feature/elektra.html	http://www.reviewedby.com/galleries_v/Goran_Visnjic.html	http://www.cinebel.be/Servlets/Splash?lng=n1
9	http://www.rottentomatoes.com/click/movie-1140820/reviews.php?critic=columns&sortby=default&page=2&rid=1366521	http://uk.movies.yahoo.com/artists/v/Goran-Visnjic/index-326463.html	http://gvsite.com/landing.php?id=gvsite.com	http://nightdeveloping.com/rave-motion-picture-theatre.html	http://www.dvdempire.com/Exec/v4_list_cast.asp?userid=99364106946880&cast_id=33995
10	http://www.azreporter.com/entertainment/television/movies/spartacus.html	http://www.imdb.com/name/nm0899681/nowshowing	http://www.firstlookmedia.com/films/closeyoureyes/movie/index.html	http://www.matchflick.com/people/18988	http://us.imdb.com/gallery/granitz/1004/wi20010121_GoranVisnjic_Michelson_139099.jpg

Taulukko 10 Haku 4: goran visnjic AND (movies OR films) AND NOT er. Tulossivujen www-osoitteet. (15.11.2006).

	AltaVista	Google	Live Search	www.fi	Yahoo!
1	2 O	2	2	3	2 O
2	2 O ²	-	2	2	2
3	2 O/2	1	2	2	2 O/2
4	2	2	2	1	1
5	2 O ³	3	2	2	3
6	2 O/2 ²	2	2 O	1	1
7	2 O/2 ³	2	2 O/2	2	1
8	2	3	2	3	1
9	2	2	2	2	2
10	2	2	2	2	2
Yht.	17	19	19	20	16

Taulukko 11 Haku 4: goran visnjic AND (movies OR films) AND NOT er. Tulossivuille annetut pisteet. (15.11.2006).

Haussa numero 4 voiton vei ehkä hieman yllättäen kotimainen www.fi, jolle Google ja Live Search hävisivät yhdellä pisteellä. Toiseksi viimeinen oli AltaVista, yhden pisteen Yahoo!ta parempana.

Tulokset eivät olleet kehuttavia. Ainut hakukone, jonka antamiin viitteisiin olen todella tyytyväinen, on ykköseksi tullut www.fi, joka rankkasi parhaaksi viitteeksi erinomaisen sivuston. Live Searchin tarjoamissa linkeissä ei ollut yhtäkään todella hyvää löytöä, vaan kaikki viitteet olivat keskinkertaisia. Googlen kärkikymmenikössä oli yksi toimimaton linkki, ja AltaVista tarjosi paljon päällekkäisiä linkkejä. Yahoo!n viimeinen tila selittyi paljolti sillä, että sen tuloslistalla olevista sivuista peräti neljä oli ei-englanninkielisiä. Ainut asia, jonka näiltä sivuilta ymmärrän, on nimi Goran Visnjic. Siksi olen kohdellut näitä sivuja samalla tavalla kuin niitä sivuja, joilla hakusana vain mainitaan, eli olen antanut niille yhden pisteen.

Kolmen pisteen arvoisiksi olen katsonut sivut, joilta löytyy Goran Visnjicin koko filmografia, joilla on muuten laajasti tietoa hänen elokuvistaan tai joilla on hyödyllisiä linkkejä, mutta joilla ei mainita ER-televisiosarjaa. Kolmen pisteen sivuja

on hakutuloksista vain kymmenesosa, mihin on lähes kaikissa tapauksissa syynä se, että sivulla mainittiin selvästi kyseisen tv-sarjan nimi. Niille sivuille, joilla on mainittu ER, mutta jotka muuten ovat hyviä sivuja, olen antanut kaksi pistettä. Samoin ne sivut, joilla on vain osittainen filmografia eli joilla käsitellään vain yhtä tai muutamaa Goran Visnjicin elokuvaa, ovat saaneet pääsääntöisesti kaksi pistettä.

Yksi parhaista sivuista on Googlen viite numero 5, jolla on lueteltu Goran Visnjicin elokuvat uusimmasta vanhimpaan. Elokuvan nimeä klikkaamalla avautuu uusi sivu, jossa on kaikki olennainen tieto kyseisestä elokuvasta: muut näyttelijät, ohjaaja, käsikirjoittajat, genre, julkaisupäivämäärä, kestoaika jne. Nähtävänä on myös kuvia elokuvasta sekä kriitikoiden arvosteluja, minkä lisäksi käyttäjä saa antaa itse oman arvosanansa elokuvasta. Tämän tyyppisiä sivuja olisin halunnut nähdä hakutuloksissa enemmän.

AltaVistan parhaaksi rankkaamaa viitettä ihmettelen, sillä sen lisäksi, että siinä käsitellään lähinnä vain yhtä Goran Visnjicin elokuvaa, siinä on enemmän kyse eräästä toisesta näyttelijästä. Live Searchin ykkösviite taas yllätti siksi, että se on suomenkielinen - joskin siitä saa nähtäväksi myös englanninkielisen version.

Hyvin monella sivustolla (Googlen viitteissä 1, 7 ja 9, Yahoo!:n viitteissä 1-3, AltaVistan viitteissä 2, 4 ja 6, Live Searchin viitteissä 3 ja 8 sekä www.fi:n viitteessä 2) on täysin sama Goran Visnjicin biografia. Sivut eivät kuitenkaan ole muilta osin samanlaiset.

Oletin hakukoneilla löytyvän parempia sivuja Goran Visnjicistä ja elokuvista, joissa hän on esiintynyt. Kyseinen näyttelijä on tehnyt paljon muutakin kuin esittänyt lääkäriä ER-sarjassa, joten tulosten taso kummastutti minua. Kenties eri tavoin muotoillulla haulla saataisiin tarkoituksenmukaisempia tuloksia.

	AltaVista	Google	Live Search	www.fi	Yahoo!
1	http://personal.ineet.fi/private/windaldas/Index.htm	http://www.vaasanelainkeskus.fi/db/linkkilista.asp	http://www.telemail.fi/tosca-kasma/linkit.html	http://www.telemail.fi/tosca-kasma/linkit.html	http://personal.ineet.fi/private/windaldas/Index.htm
2	http://www.koirat.com/rotu.phtml?id=304	http://www.elisanet.fi/vappu/	http://www.raahekoirakerho.net/vkoirat.htm	http://www.juppsguard.net/kennel.htm	http://www.koirat.com/rotu.phtml?id=304
3	http://www.telemail.fi/tosca-kasma/linkit.html	http://www.tunturisuus.com/koiralinakit/kultsi.html	http://www.kpnetti.com/kokkola2005/	http://www.expag.e.com/13315	http://www.telemail.fi/tosca-kasma/linkit.html
4	http://www.tunturisuus.com/koiralinkit/kultsi.html	http://www.tunturisuus.com/koiralinkit/colliiepitka.html	http://www.sunpoint.net/~whesteam/about_us.htm	http://www.kiirramanna.net/leiri.htm	http://www.tunturisuus.com/koiralinkit/kultsi.html
5	http://www.ronsanas.com/	http://www.kpnetti.com/kokkola2005/	http://www.vaasanelainkeskus.fi/db/linkkilista.asp	http://freewebs.com/arany2/kosti.html	http://kotisivu.mtv3.fi/damiika/linkkeja.html
6	http://kotisivu.mtv3.fi/damiika/linkkeja.html	http://www.kpnetti.com/yli vieska2004/index.php?sivu=6&kieli=1	http://www.itisvet.net/dantaran/koirat.htm	http://www.dirlian.com/uutisia/uutisia2005.htm	http://www.elisanet.fi/vappu/
7	http://www.elisanet.fi/vappu/	http://www.raahekoirakerho.net/vkoirat.htm	http://www.opaasi.fi/p-k.kennelpiiri/metsoleiri.html	http://www.pijan.net/pijan.htm	http://www.freewebs.com/happy-sun_ripe/
8	http://www.kolumbus.fi/sari.makelal/	http://www.koululainen.fi/valkka/?mode=viewtopic&tid=29564&page=1	http://www.freewebs.com/kennelpiinkrouse/koirat.htm	http://kllk.vuodatus.net/page/Koirat	http://www.kpnetti.com/kokkola2005/
9	http://www.juppsguard.net/kennel.htm	http://www.saunalahti.fi/~lskp/elo/tulokset_su.html	http://www.pijan.net/pijan.htm	http://www.koirat.com/kennelsivu.phtml?id=1307	http://www.piners.com/Finska/mimmi.htm
10	http://www.amandiinas.net/	http://www.kennelliitto.fi/FI/Koiraamme/Vuoden+nayttelykoira/Kasvatatja+tulokset+2005.htm	http://www.koiranruokinta.fi/meinheim/kennel.html	http://personal.ineet.fi/koti/bernarossa.kennel/esittely.htm	http://www.freewebs.com/goldenrings/index.htm

Taulukko 12 Haku 5: "kultainen noutaja" AND kennel. Tulossivujen www-osoitteet. (19.11.2006).

	AltaVista	Google	Live Search	www.fi	Yahoo!
1	3	3	3	3	3
2	3	3	3	3	3
3	3	3 O	2	?	3
4	3	3 O/2	3	3	3
5	3	2 O ²	3	?	3
6	3	2 O/2 ²	3	3	3
7	3	3	3	3	?
8	?	2	?	?	2
9	3	1	3	3	3
10	3	3	1	1	?
Yht.	27	22,5	24	19	23

Taulukko 13 Haku 5: "kultainen noutaja" AND kennel. Tulossivuille annetut pisteet. (19.11.2006).

Viimeisen haun ykkönen oli AltaVista kolmen pisteen erolla Live Searchiin. Kolmanneksi sijoittui Yahoo! vain puolen pisteen erolla neljänneksi tulleeeseen Googleen. Vaikka "kultainen noutaja" on puhtaasti suomenkielinen hakutermin, ei kotimainen hakukone menestynyt: viimeiseksi jääneen www.fi:n ja voittaja AltaVistan välinen ero on niinkin iso kuin kahdeksan pistettä.

Jos edellisen haun viitteet olivat suurimmaksi osaksi keskinkertaisia tai huonoja ja tuottivat siksi hakijalle pettymyksen, näyttivät hakukoneet tässä haussa jälleen hyödyllisyytensä: yli kaksi kolmasosaa tuloksista oli kolmen pisteen arvoisia. Voittajahakukone AltaVistan kärkikymmenikössä oli peräti yhdeksän kolmen pisteen sivua - ainut sivu, jolle en antanut kolmea pistettä, ei liittynyt kultaisiin noutajiin mitenkään.

Yritin tässä haussa asettua sellaisen ihmisen asemaan, joka harkitsee kultaisen noutajan hankkimista ja haluaa tietoa kenneleistä, joissa kyseistä rotua kasvatetaan. Näin ollen olen antanut kolme pistettä kultaisia noutajia kasvattavien kenne-

lien sivuille sekä niille sivuille, joilla on hyödyllisiä linkkejä. Kaksi pistettä ovat saaneet esimerkiksi koiranäyttelyiden tulossivut, joilta saa usein hyödyllisiä vinkkejä siitä, minkä nimisissä kenneleissä tiettyä rotua kasvatetaan.

Tuloksissa oli jonkin verran myös niin sanottujen virtuaalikenneliien sivuja. Haun tarkoitus oli kuitenkin löytää oikeita kenneleitä, joissa kasvatetaan oikeita koiria. Tämän takia olen kohdellut virtuaalikenneliien sivuja samalla tavalla kuin niitä sivuja, jotka eivät liity hakuaiheeseen millään tavoin.

Kuten jo antamistani pisteistä on pääteltävissä, hakukoneet löysivät paljon hyviä sivuja. Yksi suosikeistani on Tunturisuden koiralinkit -sivusto, joka on Googlen viite 3 (ja 4) sekä Yahoo!:n ja AltaVistan viite 4. Sivusto kiinnostanee useimpia koiranomistajia tai koiran hankkimista suunnittelevia: tarjolla on runsaasti linkkejä muun muassa kasvattajien, eläinlääkäreiden, rotujärjestöjen ja eläintarvike-liikkeiden sivuille, ohjeita koiran hankkimista suunnitteleville, tietoa koiran kouluttamisesta, ravinnosta ja sairauksista sekä keskustelupalsta. Tekstin ohella sivustolla on myös paljon valokuvia koirista.

Toinen erittäin hyödyllinen sivu on Kennelliiton sivu, jolle pääsee Googlen linkistä 10. Siellä on informaatiota tulevista näyttelyistä, rotumääritelmät pdf-dokumentteina, kattava tietopaketti koiranpennun hankintaa suunnitteleville ja paljon muuta.

Kolmas laatulöytö oli Yahoo!:n ja AltaVistan kakkosena ollut koirat.com-sivusto, joka löytyy myös www.fi:n sijalta 9. Koirat.com-sivustolta on mahdollista hakea muun muassa kenneleitä nimen tai rodun perusteella, uusia tai suunnitteilla olevia pentueita, eri rotuisten koirien kuvia sekä tulevia näyttelyitä. Mukana on myös keskustelupalsta sekä linkkejä rotujärjestöjen sivuille. On hienoa, että Yahoo! ja AltaVista rankkasivat sivuston korkealle tuloksissaan - ja harmillista, ettei se löytynyt lainkaan Googlen ja Live Searchin kärkikymmeniköstä.

Niin ikään maininnan arvoinen on Yahoo!:n ja AltaVistan viite 3, joka oli myös Live Searchin ja www.fi:n parhaaksi rankkaama viite. Kyseinen sivusto keskittyy ensisijaisesti kultaisiin noutajiin, ja sillä on paljon aiheeseen liittyviä linkkejä ja

kuvia. Linkkejä löytyy paljon myös Yahoo!:n viitteestä 5, joka on AltaVistan viite 6.

Kennelien omat sivut, kuten Googlen viite 2 (joka oli myös Yahoo!:n viite 6 ja AltaVistan viite 7), Yahoo!:n viitteet 1 (joka oli myös AltaVistan viite 1) ja 9 sekä AltaVistan viitteet 5 ja 10, ovat todella hyviä. Kennelit esittelevät yleensä sivuillaan koiransa sukutaulujen ja valokuvien kera sekä ilmoittavat, milloin heille on tulossa uusia pentuja. Myös yhteystiedot ovat tavallisesti hyvin näkyvillä.

Tulosten perusteella uskallan luvata, että kultaisen noutajan pentua hankkimassa oleva löytää hakukoneilla tarpeeksi tietoa päätöksentekonsa tueksi. Löytyneistä sivuista on varmasti iloa muillekin koirankasvatuksesta kiinnostuneille.

9 YHTEENVETO

Taulukosta 14 nähdään hakukoneiden saamat yhteispisteet.

	AltaVista	Google	Live Search	www.fi	Yahoo!
Taulukko 3	8	11	9	10	10
Haku 1	24,5	25,5	22,5	20,5	23,5
Haku 2	20	22,5	21,5	25	20,5
Haku 3	22	25,5	17	16	23
Haku 4	17	19	19	20	16
Haku 5	27	22,5	24	19	23
Yhteensä	118,5	126	113	110,5	116

Taulukko 14 Hakukoneiden yhteispisteet.

Yhteispisteissä voiton vei - ehkä odotetustikin - Google. Toiseksi sijoittui AltaVista ja kolmanneksi Yahoo!. Neljäs oli Live Search, ja www.fi jätettiin viimeiseksi.

Googlen etumatka muihin hakukoneisiin on melko selkeä: toiseksi tullut AltaVista hävisi sille 7,5 pisteellä. Loppupisteissä sijoille 2-5 sijoittuneet hakukoneet ovat vain muutaman pisteen päässä toisistaan. Enemmän hajontaa olisi voitu saada käyttämällä esimerkiksi pisteasteikkoa 1-4 tai 1-5 tässä käytetyn 1-3-asteikon sijaan, tai käymällä tulosviitteistä läpi kymmenen ensimmäisen sijaan 15 tai 20 ensimmäistä. Kolmas mahdollisuus olisi ollut käyttää useampaa hakulausetta, vaikkapa kymmentä. Harkitsinkin tätä vaihtoehtoa tutkimusta suunnitelllessani, mutta päädyin kuitenkin viiteen, koska viisi hakulausetta riitti kaikkien Boolean operaattoreiden testaamiseen, mikä oli tämän tutkimuksen keskeisin tarkoitus.

Linkkejä, jotka eivät liittyneet hakuaiheeseen mitenkään, löysi eniten Live Search (5 kpl). Toisena oli www.fi (4 kpl), kolmantena Yahoo! (3 kpl) ja neljäntenä AltaVista (1 kpl). Googella näitä viitteitä ei ollut ainuttakaan.

Linkkejä, jotka eivät toimineet, palautti eniten AltaVista (2 kpl). Toisen sijan jakavat Google ja www.fi (1 kpl). Kaikki Yahoo!:n ja Live Searchin antamat linkit olivat toimivia.

Päällekkäisiä linkkejä oli eniten AltaVistalla (18 kpl). Toiseksi sijoittui Google (14 kpl), kolmanneksi Yahoo! (12 kpl) ja neljänneksi Live Search (10 kpl). Vähiten näitä linkkejä oli www.fi:llä (4 kpl).

Yksi kriteeri, jota hakukoneiden arvioinnissa ja pisteytyksessä olisi myös voinut olla käytössä, on nopeus, sillä se on hakutulosten relevanssin ohella toinen käytettävyyteen vaikuttava tekijä. Jos hakukoneen aloitussivu ja etenkin hakutuloslista latautuvat hitaasti, siirtyy tiedonhakija pian käyttämään toista hakukonetta.

Haun 4 tulossivut näyttivät, etteivät hakukoneet tue NOT-operaattoria yhtä hyvin kuin muita Boolean operaattoreita. Hyvin monessa hakukoneiden löytämässä si-

vussa mainittiin termi, jonka olin halunnut sulkea tuloksista pois NOT-operaattorilla. Vika saattoi toki olla myös hakulauseen puutteellisessa muotoilussa.

Googlen voiton ratkaisi sen tulosten tasaisuus: se oli taulukon 3 sekä hakujen 1 ja 3 tuloksissa paras, sekä lisäksi haussa 2 toiseksi paras ja haussa 4 jaetulla toisella sijalla, mutta ei minkään haun huonoin. Itse olen käyttänyt sitä vuodesta 2001 lähtien, ja se on täyttänyt tietotarpeeni niin hyvin, että vain hyvin harvoin minun on ollut turvauduttava muihin hakukoneisiin. Olen voinut luottaa siihen, olenpa ollut sitten etsimässä tietoa opiskelutehtävieni tueksi tai vapaa-ajan mielenkiinnon kohteisiin liittyen. Viimeksi Googlestä oli minulle suuri apu tämän tutkielman teoriaosan laadinnassa. Erityisesti pidän sen käyttöohjeesta, joka on mielestäni kattava ja selkeä. Miinuspuolena voidaan mainita päällekkäisten linkkien suurehko määrä hakutuloksissa. Tämän tutkimuksen tulosten perusteella Google on kuitenkin ansainnut markkinajohtajan asemansa.

Vaikka AltaVista ei aivan pysynytäkään Googlen vauhdissa mukana, ei se ainakaan tämän tutkimuksen koehakujen perusteella ole mikään eilispäivän hakukone. Se oli haussa 1 toiseksi paras ja haussa 5 selvä voittaja. Lähes kaikki sen tarjoamat linkit olivat joko kolmen tai kahden pisteen arvoisia, mutta koska sen hakutuloksissa oli vielä enemmän päällekkäisiä linkkejä kuin Googlen hakutuloksissa, ei se yltänyt voittoon saakka.

Yahoo! tuli minulle tutuksi jo 90-luvun puolella, mutta olen käyttänyt enemmän sen aihehakemistoa kuin sanahakukonetta. Se sijoittui taulukon 3 tuloksissa jaetulle toiselle sijalle sekä haussa 3 toiselle sijalle, mutta toisaalta haussa 4 se jätettiin viimeiseksi. Plussaa Yahoo! saa siitä, ettei sen hakutuloksissa ollut ainuttakaan sellaista linkkiä, joka ei olisi toiminut. Miinusta annan käyttöohjeesta, joka on ainakin minun mielestäni puutteellinen ja sekava.

Live Search, jolla Microsoft pyrki syrjäyttämään Googlen hakukoneiden kärki-paikalta, ei ainakaan vielä tässä tutkimuksessa vakuuttanut. Se ei ollut minkään haun heikoin, mutta se oli taulukon 3 tuloksissa sekä haussa 1 ja 3 toiseksi huonoin. Live Searchilla oli mahdollisuuksia pärjätä paremmin. Hyvin suuri osa sen

löytämistä viitteistä oli kolmen tai kahden pisteen arvoisia. Sen hakutuloksissa ei ollut lainkaan kuolleita linkkejä, ja päällekkäisiä linkkejä oli vain vähän. Live Searchin pisteitä vähensi kuitenkin se, että se tarjosi hakutuloksissaan enemmän hakuaiheeseen liittymättömiä linkkejä kuin muut hakukoneet. Uskon, että Microsoftilla on tarkoitus panostaa hakutoimintojen kehittämiseen vahvasti jatkossakin - sekä myös resursseja siihen - joten jos esimerkiksi vuoden kuluttua toteutettaisiin tämän tyyppinen tutkimus, Live Search todennäköisesti menestyisi jo huomattavasti paremmin.

Www.fi:ltä odotin enemmän erityisesti suomenkielisissä hauissa. Parhaimmillaan se olikin haussa 2, joka oli nimenomaan suomenkielinen: kaikki sen siinä löydämät viitteet olivat joko kahden tai kolmen pisteen arvoisia. Se oli paras myös haussa 4 sekä jaetulla toisella sijalla taulukon 3 tuloksissa. Muiden hakujen tuloksissa se jäi kuitenkin jälkeen muista hakukoneista. Www.fi tarjosi päällekkäisiä linkkejä vähemmän kuin muut hakukoneet, mutta sen pisteitä vähensivät monet hakuaiheeseen liittymättömät linkit. Itse kuitenkin pidän edelleen www.fi:stä ja uskon sillä olevan mahdollisuuksia kehittyä suosituksi hakukoneeksi. Sen yleisilme on mielestäni miellyttävä, käyttöliittymä on selkeä, ja ennakoiva tekstinsyöttö vähentää tiedonhakijan vaivaa.

Koehakujen tulokset eivät kerro, mikä hakukone on absoluuttisesti paras - ne kertovat ainoastaan, mitkä olivat näissä hauissa ja näillä kriteereillä mitattuna parhaita. Tiedonhakijoiden tietotarpeet vaihtelevat suuresti: jotkut etsivät puhtaasti viihdettä, jotkut taas tieteellistä tietoa, jotkut käyttävät hakukoneita yhden tietyn dokumentin aktiiviseen hakuun ja jotkut yleiseen selailemiseen. Näin ollen hakukoneita ei voi laittaa yleispätevään paremmuusjärjestykseen. Hakukoneiden toimintoja kehitetään kaiken aikaa, ja indeksointia sekä rankkausperiaatteita saataan muutella. Tämän tutkimuksen tulokset olivat voimassa vain tässä tutkimuksessa käytetyillä hakulausekkeilla vain sillä hetkellä, kun haut tehtiin. Ne antavat kuitenkin hyvän yleiskuvan siitä, miten hyvin hakukoneet ymmärtävät Boolean hakuja ja minkä tyyppisen tiedon hakijana mikin hakukone on parhaimmillaan.

LÄHTEET

Kirjat

Alaterä Anu, Halttunen Kai 2002: Tiedonhaun perusteet - osa lukutaitoa. Tampereen yliopiston täydennyskoulutuskeskus ja Otavan Opisto/Internetix. BTJ Kirjastopalvelu Oy. Helsinki.

Haasio Ari 2002: Internet kansalaisen tietovarantona. Teoksessa Haasio Ari, Piukkula Juha (toim.): Tietoverkot ja kirjastot. BTJ Kirjastopalvelu Oy. Helsinki. 7-21.

Haasio Ari 2003: Internet-tiedonhaun opas. BTJ Kirjastopalvelu Oy. Helsinki.

Haasio Ari 2005a: Internetin tietokannat. BTJ Kirjastopalvelu Oy. Helsinki.

Haasio Ari 2005b: Tiedon lähteillä 1. BTJ Kirjastopalvelu Oy. Helsinki.

Heikkinen Risto, Agander Anja, Ijäs Eva, Laitinen Mirja 2001: Tiedonhakijan teho-opas. 5., uudistettu painos. Docendo Finland Oy. Jyväskylä.

Hellstén Tanja 2004: Hakukoneiden vertailu. Savonia-ammattikorkeakoulu. Varkauden yksikkö. Tietojenkäsittelyn koulutusohjelma. Opinnäytetyö.

Hirsjärvi Sirkka, Remes Pirkko, Sajavaara Paula 2005: Tutki ja kirjoita. 11. painos. Gummerus Kirjapaino Oy. Jyväskylä.

Husu Milja, Tarkoma Elise, Vuorijärvi Aino 2001: Ammattisuomen käsikirja. 6., uudistettu painos. WS Bookwell Oy. Porvoo.

Korhonen Johanna 2003: Iloa internetistä. Gummerus Kirjapaino Oy. Jyväskylä.

Korhonen Mikko, Likitalo Heikki 2005: Matematiikkaa ja logiikkaa tietojenkäsittelijöille. Savonia-ammattikorkeakoulun julkaisusarja B4/2005. Savonia-ammattikorkeakoulu, Kehittämisen- ja palvelukeskus. Kuopio.

Nikkilä Timo 2001: Internet 2001 - Eilen, tänään ja huomenna. Docendo Finland Oy. Jyväskylä.

Www-sivut

Alaterä Anu, Halttunen Kai 2001-2004: Tiedonhaun perusteet - osa lukutaitoa? [online]. Tampereen yliopiston täydennyskoulutuskeskuksen, Tampereen yliopiston informaatiotutkimuksen laitoksen sekä Otavan Opiston / Internetixin yhteishankkeena tuotettu oppimateriaali. [Viitattu 19.10.2006]. Saatavissa <http://www.internetix.fi/opinnot/opintojaksot/0viestinta/informaatiotutkimus/po2/>.

Alexa Internet 2006: Top 500 Sites. [online]. [Viitattu 01.11.2006]. Saatavissa <http://www.alexa.com/site>.

Aniway 1999-2005: Hakukoneet ja hakemistot. [online]. [Viitattu 26.09.2006]. Saatavissa http://www.aniway.fi/hakukoneet_hakemistot_1.htm.

Avoin yliopisto, ei päiväystä: Tiedonhaku. [online]. [Viitattu 15.10.2006]. Saatavissa <http://www.avoinyliopisto.fi/kirjastot/tiedhaku.html>.

Finkernet Markkinointi, ei päiväystä: Hakukonemarkkinointi. [online]. [Viitattu 29.10.2006]. Saatavissa <http://www.finkernet.com/fi/hakukonemarkkinointi.htm>.

Fonecta 2006: Www.fi - hakukone joka ymmärtää suomalaista. [online]. [Viitattu 25.10.2006]. Saatavissa <http://www.fi/WebSearch>.

Google 2006: Kaikki Googlesta. [online]. [Viitattu 31.10.2006]. Saatavissa <http://www.google.fi/intl/fi/about.html>.

Haakana Kari 19.10.2006: Fonecta lanseeraa www.fi-hakukoneen uudelleen. Tietokone [verkkolehti]. [Viitattu 05.11.2006]. Saatavissa http://www.tietokone.fi/uutta/uutinen.asp?news_id=28463&tyyppi=1.

Haataja Juha 24.02.2005: Google ällistyyttää. Mikropc.net [verkkolehti]. [Viitattu 31.10.2006]. Saatavissa <http://mikropc.net/kolumni/index.jsp?id=j223>.

Hakala Juha, Stenvall Jani 1998: Dublin Core -formaatin käyttöopas. [online]. Helsingin yliopiston kirjasto. [Viitattu 25.10.2006]. Saatavissa <http://www.lib.helsinki.fi/meta/dc-opas.html>.

Heikkinen Mikko 30.07.2003, päivitetty 18.09.2004: Metatieto. [online]. [Viitattu 27.10.2006]. Saatavissa <http://www.biomi.org/internet/metadata.html>.

Heikniemi Jouni 13.04.2001, päivitetty 30.03.2002: Kuinka estää hakukonetta löytämästä minua? [online]. [Viitattu 25.10.2006]. Saatavissa <http://www.heikniemi.net/kirj/web/robots.html>.

Hänninen Harri 2005: Hakukoneet. [online]. Digitaalisen viestintätekniiikan seminaari. Lappeenrannan teknillinen yliopisto. Tietotekniikan osasto. Tietoliikennetekniikan laitos. [Viitattu 02.10.2006]. Saatavissa http://www.it.lut.fi/kurssit/04-05/010850000/Harri_hanninen.pdf.

Info Center Finland ICF Oy 1997-2006: Webinfo - OHJEET. [online]. [Viitattu 19.10.2006]. Saatavissa <http://www.webinfo.fi/static/hakuohjeet.html>.

Karvonen Tuomas 27.09.2005: Google nokittaa haettavissa web-sivuissa. Digitoday.fi [verkkolehti]. [Viitattu 31.10.2006]. Saatavissa http://www.digitoday.fi/page.php?page_id=11&news_id=200515621.

Karvonen Tuomas 24.02.2006: Hakukone taitaa suomen taivutusmuodot. ITviikko.fi [verkkolehti]. [Viitattu 19.10.2006]. Saatavissa http://www.itviikko.fi/page.php?page_id=15&news_id=2006998.

Kirjastot.fi 1995-2006: Verkkotiedonhaun perusteita. [online]. [Viitattu 22.10.2006]. Saatavissa <http://www.kirjastot.fi/tiedonhaku/perusteita/>.

Korpela Jukka K. 02.05.2005: Uskaltaako netissä kysyä? Mikropc.net [verkkolehti]. [Viitattu 18.10.2006]. Saatavissa <http://mikropc.net/kolumni/index.jsp?id=j232>.

Korpela Jukka K., päivitetty 12.10.2006: Web-julkaisemisen opas. [online]. [Viitattu 17.09.2006]. Saatavissa <http://www.cs.tut.fi/~jkorpela/webjulk/>.

KotiMikro, ei päiväystä: Windows Live Search livenä. KotiMikro [verkkolehti]. [Viitattu 04.11.2006]. Saatavissa <http://www.kotimikro.fi/Crosslink.jsp?d=426&a=6717>.

Koulutus- ja konsultointipalvelu KK Mediat 2000-2006: Internet markkinointi. [online]. [Viitattu 25.10.2006]. Saatavissa <http://www.2kmediat.com/nettimarkkinointi/>.

Lappeenrannan kaupunginkirjasto 2005: Opettajan tiedonhakuopas. [online]. [Viitattu 19.10.2006]. Saatavissa <http://www.lappeenranta.fi/?deptid=12326>.

Microsoft 2006: Microsoft Liven ohje. [online]. [Viitattu 05.11.2006]. Saatavissa <http://www.live.com>.

MTV3 Internet, ei päiväystä: Internet ja tiedonhaku. [online]. [Viitattu 16.10.2006]. Saatavissa <http://www.mtv3.fi/liittymat/tuki.shtml/62940?Internetopas>.

Net Site Story, ei päiväystä: Hakukonemarkkinointi ja optimointi. [online]. [Viitattu 29.10.2006]. Saatavissa <http://www.netsitestory.com/SEO/>.

NoframE Oy 2006: Microsoftin Live-hakupalvelu aloitti virallisesti. [online]. [Viitattu 05.11.2006]. Saatavissa <http://www.noframe.fi/hakukoneuutiset.htm>.

Notess Greg R. 1999-2006, päivitetty 10.10.2006: Search Engine Showdown: The Users' Guide to Web Searching. [online]. [Viitattu 01.11.2006]. Saatavissa <http://www.searchengineshowdown.com/>.

Overture Services, Inc. 2006: AltaVista Ohje. [online]. [Viitattu 02.11.2006]. Saatavissa <http://fi.altavista.com/help/>.

Perkiömäki Mika 1997: Tiedonhaku World-Wide Webistä. [online]. Tampereen yliopisto. Tietojenkäsittelyopin laitos. Julkaisussa Pieniä atk-alan tutkimuksia - syksy 1996, toim. Pertti Järvinen. [Viitattu 07.09.2006]. Saatavissa <http://mikap.iki.fi/papers/wwwhaku.html>.

Poutiainen Risto 04.02.2005: Latentti semanttinen indeksointi Googlen algoritmossa. [online]. [Viitattu 19.10.2006]. Saatavissa <http://www.hakukone.info/googleblog/2005/02/latentti-semanttinen-indeksointi.htm>.

Rautajoki Marja 2005: Haku Internetistä. [online]. VirtuaaliAMK:n Verkkotuottaminen-opintokokonaisuuden Internet, Intranet ja palvelut -opintjakson opiskelumateriaali. [Viitattu 10.09.2006]. Saatavissa <https://www.virtuaaliamek.fi/opintokokonaisuudet/55IudTyT/1100243557888/1100243689922/1100256087746/1100257979190.html.stx>.

Robotstxt.org, ei päiväystä: Scooter. [online]. [Viitattu 25.10.2006]. Saatavissa <http://www.robotstxt.org/wc/active/html/scooter.html>.

Sarajärvi Vesa 23.12.2003, päivitetty 06.04.2004: Tiedon lähteillä. [online]. Tiedonhakukurssi Ranuan lukiolle ja yläasteelle. [Viitattu 23.10.2006]. Saatavissa <http://www.ranua.fi/files/20040114084723.doc>.

Savolainen Reijo 2003: Internetistäkö kansalaisten tiedonhankinnan tärkein kanava? [online]. Julkaisussa Salonharju Inkeri (toim.): Kirja tietoverkkojen maailmassa. Helsingin yliopiston kirjasto. [Viitattu 25.10.2006]. Saatavissa

<http://www.lib.helsinki.fi/julkaisut/kirjatietoverkkojenmaailmassa/savolainen1.html>.

Setälä Heidi 23.04.2004: Hei me googlataan. Tutka [verkkolehti]. [Viitattu 18.10.2006]. Saatavissa <http://www.tutka.org/juttu?juttu=3603>.

Stenvall Jani 21.09.1998, päivitetty 03.01.2002: Dublin Core -formaatin käyttö-opas. [online]. Helsingin yliopiston kirjasto. [Viitattu 25.10.2006]. Saatavissa http://www.lib.helsinki.fi/dublin_core/dc-opas2002.html.

Stenvall Jani 18.08.1999: Dublin Core kuvailuformaatin RDF-syntaksi. [online]. Helsingin yliopiston kirjasto. [Viitattu 25.10.2006]. Saatavissa <http://www.lib.helsinki.fi/eva/dcsyntRDF.pdf>.

Systems Garden, ei päiväystä: Mitä tarkoittaa indeksointi? [online]. [Viitattu 21.10.2006]. Saatavissa <http://www.sguugle.com/asp/system/empty.asp?P=98&VID=default&SID=427843605556242&S=1&A=closeall&C=31954>.

Tampereen ammattiopiston kirjasto- ja tietopalvelut, ei päiväystä: Tiedonhaun opas. [online]. [Viitattu 23.10.2006]. Saatavissa <http://www.tao.tampere.fi/kirjasto/opas.htm>.

Tampereen kaupunginkirjasto, ei päiväystä: Sanahakupalvelujen toiminta ja hakuominaisuudet. [online]. [Viitattu 22.10.2006]. Saatavissa <http://www.tampere.fi/kirjasto/ansues.htm>.

Tampereen yliopiston kirjasto, ei päiväystä: Kirjaston käytön ja tiedonhankinnan perusteet. [online]. Yhteiskuntatieteellisen tiedekunnan valinnaisiin opintoihin kuuluvan Kirjaston käytön ja tiedonhankinnan perusteet -opintojakson www-sivusto. [Viitattu 19.10.2006]. Saatavissa <http://www.uta.fi/laitokset/kirjasto/opetus/kirjastokurssi/index.html>.

Tilastokeskus, päivitetty 20.12.2005: Tietoliikenne. [online]. [Viitattu 18.10.2006]. Saatavissa http://www.stat.fi/tup/suoluk/suoluk_tiede.html.

Uutistoimisto.com 1999-2006, päivitetty 12.09.2006: Lyhyt hakukoneopas. [online]. [Viitattu 16.10.2006]. Saatavissa http://www.uutistoimisto.com/lyhyt_hakukoneopas.php.

Vanhala-Nurmi Vuokko, ei päiväystä: Tiedon hakeminen Internetistä. [online]. Helsingin liiketalouden ammattikorkeakoulun Käytännön tietoliikenne -kurssin opiskelumateriaali. [Viitattu 19.10.2006]. Saatavissa <http://myy.helia.fi/~vanvu/tietoliikenne/thkehys.html>.

Yahoo! Inc. 2006: Company Information. [online]. [Viitattu 01.11.2006]. Saatavissa <http://docs.yahoo.com/info/>.

Yhdyskuntasuunnittelun tutkimus- ja koulutuskeskus, päivitetty 11.09.2005: Tiedonhaun perusteet. [online]. [Viitattu 19.10.2006]. Saatavissa <http://virtuaali.tkk.fi/yhdyskuntasuunnittelu/ytk-eri/Tiedonhaku/tiedonhakuindex.html>.