



OULUN YLIOPISTO
UNIVERSITY of OULU

WWW-sivun hakukoneoptimointi: tapaus Google

Oulun yliopisto
Tietojenkäsittelytieteiden laitos
Kandidaatin tutkielma
Heikki Heiskanen
12.4.2016

Tiivistelmä

Nykypäivänä internet toimii tiedon valtatieenä, jonka käyttäjinä toimivat niin yksilöt, yritykset kuin suuret organisaatiotkin. Nämä kaikki internetin käyttäjät pystyvät etsimään, muokkaamaan ja lisäämään tietoa internetiin. Tiedon lisääminen internetsivujen muodossa ei kuitenkaan takaa sitä, että muut käyttäjät löytävät sen internetistä.

Tässä työssä perehdytään hakukoneoptimoinnin käsitteen määrittelemiseen, hakukoneoptimoinnin vaikutusten tarkastelemiseen ja hakukoneoptimoinnin teknisen toteutuksen läpikäymiseen.

Työssä tarkastellaan lähemmin Googlen tapaan käsitellä hakukoneoptimointia ja osoittaa mitä sillä voidaan tehdä ja mitä ei, sekä näyttää esimerkein kuinka hakukoneoptimointia tehdään HTML-kielellä. Tässä työssä tarkastellaan myös hakukoneoptimoinnin tarpeellisuutta ja sen vaikutusta hakutuloksiin edellisten tutkimusten ja aiheeseen liittyvän tieteellisen kirjallisuuden pohjalta. Työn tarkoituksena on myös tuoda esille hakukoneoptimoinnissa parhaaksi koettuja käytänteitä.

Luvussa yksi käsitellään hakukoneita yleisesti, sekä tarkastellaan hakukoneiden markkinaosuuksia ja käyttäjien hakukonekäyttäytymistä. Luvussa kaksi tarkastellaan hakukoneoptimoinnissa käytettäviä käsitteitä, hakukoneoptimoinnin hyötyjä, hakukoneen yleistä toimintaa ja miten hakukone indeksoi internetsivuja. Luvussa kolme näytetään konkreettisesti miten hakukoneoptimointia tehdään eli miten www-sivun näkyvyyttä voidaan parantaa manipuloimalla HTML-lähdekoodia.

Avainsanat

hakukoneoptimointi, web-ohjelmointi, hakukoneet, internet, Google

Ohjaaja

Seppo Pahnila

Lyhenteet

HTML, HyperText Markup Language

CSS, Cascading Style Sheets

JS, JavaScript

PHP, Hypertext Preprocessor

SEO, Search Engine Optimization, hakukoneoptimointi

SEM, Search Engine Marketing, hakukonemarkkinointi

SMO, Social Media Optimization, sosiaalisen median optimointi

SERP, Search Engine Results Page, hakukoneen tulossivu

URL, web-osoite

Sisällys

Tiivistelmä.....	2
Lyhenteet	3
Sisällys	4
1. Johdanto	5
2. Hakukoneoptimoinnin taustaa.....	7
2.1. Hakukoneoptimoinnin käsitteitä	7
2.2. Hakukoneoptimoinnin hyödyt	9
2.3. Hakukoneen yleinen toiminta	9
2.4. Hämähäkit ja www-sivujen ryömiminen	10
3. Hakukoneoptimointi käytännössä	11
3.1. Otsikot ja alaotsikot	11
3.2. Metadata.....	13
3.3. Sivun sisältö ja runko-merkki	13
3.4. Sitemap	13
3.5. Robots-tekstitiedosto.....	14
3.6. Avainsanat	14
3.7. Responsiivisuus.....	15
3.8. Kuvat.....	16
Pohdinta ja yhteenveto	17
Lähteet.....	19

1. Johdanto

Tässä työssä tarkastellaan Googlen hakukonetta ja sen toimintaa. Tässä työssä ei käsitellä muita hakukoneita ja niiden hakualgoritmeja, eikä myöskään Googlen käyttämää PageRank-algoritmia matemaattisella tasolla, koska se on jo itsessään laaja kokonaisuus. Tässä työssä ei käsitellä internetissä tapahtuvaa mainostamista ja markkinointia.

Työ noudattaa rakenteeltaan Harold Davisin (2006) kirjassa mainittuja kolmea käytännön ohjetta. Ensimmäiseksi täytyy ymmärtää miten hakukoneet käsittelevät internetsivuja, toiseksi pitää pystyä tekemään järkeviä muutoksia internetsivuun hakukoneiden näkökulmaa hyödyntäen ja kolmanneksi täytyy pyrkiä välttämään huonoja optimointikäytänteitä. (Davis, 2006)

Yllä mainittua rakennetta soveltaen tämän työn luvussa kaksi käydään läpi miten hakukoneet toimivat ja miten hakukoneiden hämähäkit (crawlers) näkevät internetsivun. Hämähäkeillä tarkoitetaan tietokoneohjelmia, jotka käyvät läpi internetsivuja. Hämähäkeistä kerrotaan enemmän luvussa kaksi. Näillä tiedoilla pyritään antamaan konteksti hakukoneoptimoinnille. Luvussa kolme käsitellään miten internetsivun omistaja voi käyttää hakukoneita hyödykseen ja millaisia erilaisia tapoja on tehdä hakukoneoptimointia. (Davis, 2006)

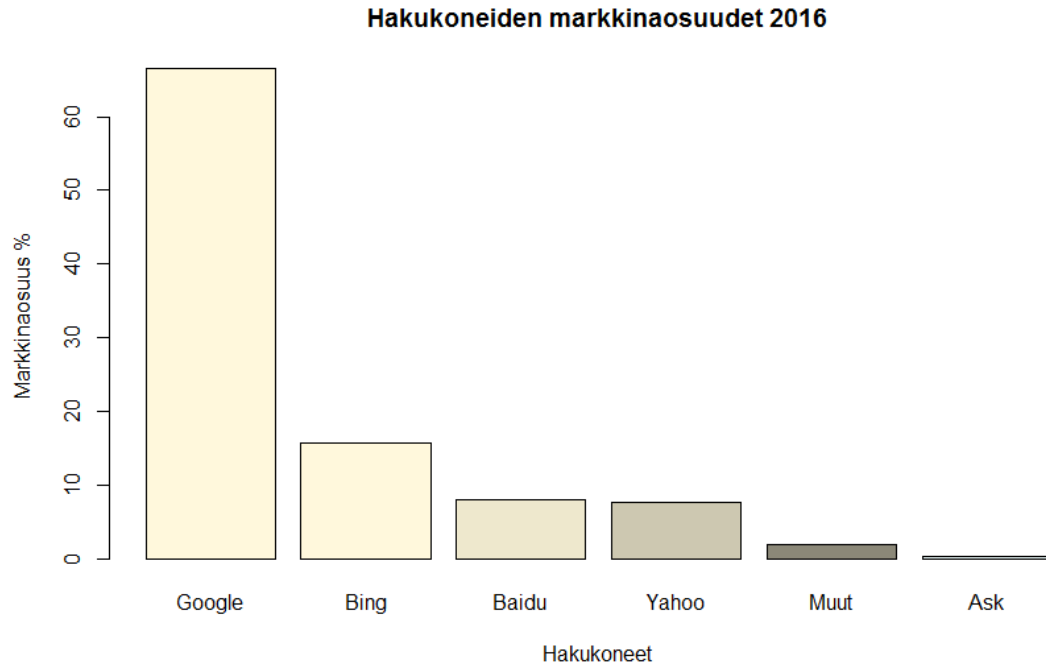
Hakukoneet mahdollistavat nopean tiedonhaun internetistä, joten käyttäjät ovat ottaneet hakukoneet jokapäiväiseen käyttöönsä. Hakukoneet vastaavat useisiin miljardeihin kyselyihin päivässä, joten hyvin hakukoneoptimoidut sivut voivat saada enemmän luontaista näkyvyyttä hakukoneilta. Tähän internetsivujen omistajat juuri pyrkivät hakukoneoptimoinnilla (Su, Hu, Kuzmanovic & Koh, 2014), sillä internetsivun näkyvyys hakukoneilla on suoraan verrannollinen kävijöiden määrään, mutta ei niiden laatuun (Davis, 2006).

Käyttäjien hakukonekäyttäytymistä tutkiessaan Broder (2002) jakoi haut kolmeen erilaiseen luokkaan, navigaatio-, vuorovaikutus- ja tietoa hakeviin hakuihin, perustuen haun tarkoitukseen eli mitä haulla pyrittiin etsimään. Tätä kyseistä jaottelua käyttivät myöhemmin Jansen, Booth ja Spink (2007) hyväksi tutkimuksessaan, jossa analysoitiin Dogpile nimisen hakukoneen hakulokia (search log).

Ensimmäiseen luokkaan kuuluvat navigaatioon liittyvät haut. Näillä tarkoitetaan hakuja, joiden tarkoituksena on löytää tietty internetsivu. Hakusanana voi toimia esimerkiksi internet osoitteen (URL) osa, kuten ”facebook.com”. Navigaation liittyviä hakuja Jansen kumppaneineen (2007) tekemän tutkimuksen mukaan oli noin 10 % mitatuista hauista. Toiseen luokkaan kuuluvat vuorovaikutukseen (transaction) liittyvät haut. Näillä tarkoitetaan hakuja, joilla pyritään löytämään muun muassa ladattavia ohjelmistoja, reseptejä ja multimediala. Vuorovaikutukseen liittyviä hakuja oli noin 9 % mitatuista hauista. Kolmanteen luokkaan kuuluu tietoa hakevat haut. Näitä oli verrattuna edellisiin kahteen luokkaan eniten, noin 80 % mitatuista hauista. Yleisimpiä hakusanoina tässä luokassa olivat luonnollisella kielellä muotoillut hakusanat, esimerkiksi ”how to...” alkuiset haut. (Jansen, Booth & Spink, 2007)

Hakukoneen markkinaosuudella tarkoitetaan hakukoneen käyttöastetta eli miten paljon hakukoneella on käyttäjiä suhteessa muihin hakukoneisiin. Kuten kuvasta (kuva 1) nähdään, niin Googlella on suurin markkinaosuus kaikista hakukoneista ja siksi se onkin tämän työn päähakukone. Seuraavaksi suurin on Bing hakukone, joka on Microsoftin oma hakukone. Lähes tasoissa markkinaosuuksiltaan ovat Baidu ja Yahoo. Baidu on kiinassa

toimiva hakukone, jota on yritetty myös laajentaa muualle maailmaan. Ask hakukoneella on alle prosentti kokonaisesta markkinaosuudesta ja on siksi aika tuntematon. Muihin hakukoneisiin kuuluu vähemmän tunnetut hakukoneet kuten AOL, Lycos ja DuckDuckGo. Näiden hakukoneiden markkinaosuus on yhteensä noin kaksi prosenttia.



Kuva 1. Hakukoneiden markkinaosuudet vuonna 2016 (Netmarketshare, 2016).

Googlen hakukoneella tehtiin vuonna 2014 noin kuusi miljardia hakua päivässä ja vuodessa yhteensä 2,1 biljoonaa hakua (Statisticbrain, 2015). Näin suureen määrään hakuja tarvitaan paljon laskentatehoa, joten Statisticbrain (2015) mukaan Googllella on yhteensä 920 000 palvelinta ympäri maailmaa, jotka sisältävät sata miljoonaa gigatavua dataa (Statisticbrain, 2015). Internetlivestats (2013) mukaan Google käsittelee yli 40 000 hakua joka sekunti, joka taas tarkoittaa yli 3.5 miljardia hakua päivässä (Internetlivestats, 2013). Näistä luvuista nähdään, että Googlen käyttöaste on kasvanut huomattavasti muutaman vuoden aikana (Internetlivestats, 2013).

Kuten aiemmin on todettu, niin hakukoneoptimointi nähdään ainoana tapana saada enemmän näkyvyyttä internetissä, joten siitä onkin muodostunut liiketoiminta-ala. Vaikka hakukoneoptimoinnista on tullut suurehko liiketoiminta-ala, niin alalla on silti paljon epäselvyyksiä oikean hakukoneoptimoinnin suhteen, koska hakukoneoptimoinnin metodeja ei voida täysin vahvistaa arkaluonteisen hakukoneen toiminnan takia (Su et al., 2014). Tästä syystä on olemassa paljon huonoja käytäntöjä ja suorastaan rikollisia tapoja tehdä hakukoneoptimointia ja näistä käytänteistä yleensä rangaistaan internetsivuja (Malaga, 2008).

Hakukoneiden valmistajat, kuten Google, eivät paljasta suoranaisesti omia algoritmejaan ja niiden painotuksia internetsivujen sijoitusten määräytymisessä. Tämä taas on ongelmallista hakukoneoptimointialan kannalta. Googlen hakukone käyttää yli 200 erilaista tekijää, jotka vaikuttavat sivun sijoittumiseen. Nämä tekijät ja niiden vaikutus sivun sijoitukseen ovat tuntemattomia. Tämän takia on vaikea sanoa varmuudella, mitkä kaikki asiat vaikuttavat hakukoneen antamiin tuloksiin. (Evans, 2007)

2. Hakukoneoptimoinnin taustaa

Tässä luvussa käsitellään hakukoneoptimoinnin käsitteistöä ja hakukoneiden yleistä toimintaa. Hakukoneoptimointiin liittyy erilaisia käsitteitä ja lyhenteitä, joiden merkitys voi jäädä lukijalle epäselväksi. Tämän luvun tarkoituksena on antaa lukijalle peruskäsitys siitä, miten Googlen hakukone toimii ja miten se löytää käyttäjän hakusanoilla relevantteja hakutuloksia tietokannastaan. Tässä luvussa käydään läpi miten internetsivut yleensäkin päätyvät Googlen tietokantaan. Lopuksi käydään läpi mitä hakukoneoptimoinnilla oikeastaan tarkoitetaan ja mitä sillä pyritään saamaan aikaiseksi.

2.1. Hakukoneoptimoinnin käsitteitä

Hakukoneoptimointi eli search engine optimization (SEO) tarkoittaa www-sivun HTML-lähekoodin (Hypertext Markup language) manipulointia niin, että internetsivu saa lisää näkyvyyttä internetissä (Killoran, 2013). Tätä kutsutaan internetsivun sisäiseksi hakukoneoptimoinniksi (on-page SEO). On olemassa myös internetsivun ulkopuolista hakukoneoptimointia (off-page SEO), jolla tarkoitetaan sisäänpäin tulevien linkkien määrää ja laatua (Stearns, 2013). On-page SEO:lla tarkoitetaan sivun omaa sisällön hyvyttä, oikein kohdistettuja avainsanoja ja sivun yleistä rakennetta. Off-page SEO:lla taas tarkoitetaan sivun ulkopuolisia asioita kuten linkkejä. Tarkemmin takaisin tulevia linkkejä (backlink). Sivun haltijalla on täysi kontrolli on-page hakukoneoptimoinnista, mutta off-page hakukoneoptimoinnista ei. (Stearns, 2013) Tästä kerrotaan enemmän luvussa kolme.

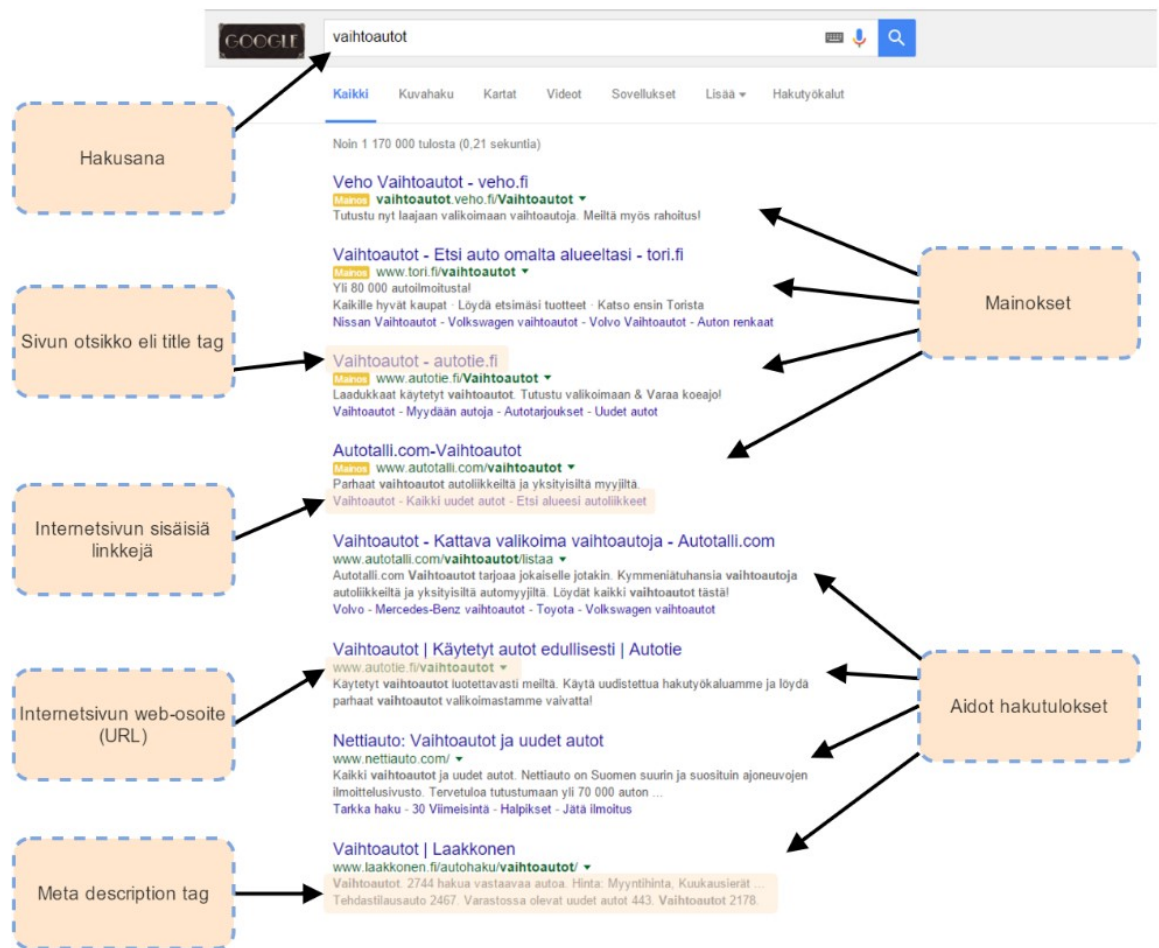
Hakukoneella haettaessa tietyillä hakusanoilla hyvin hakukoneoptimoitu internetsivu saa paremman sijoituksen hakutuloksissa. Näin hakukoneen käyttäjä löytää internetsivun helpommin useammalla eri hakusanalla ja täten palveluntuottajan kannalta internetsivu on näkyvämpi ja tuottavampi. On olemassa myös hakukonemarkkinointia eli search engine marketing (SEM), jota ei pidä sekoittaa hakukoneoptimointiin. Hakukonemarkkinointi tarkoittaa hakukoneoptimointia, johon on liitetty muita maksullisia markkinointipalveluita, kuten mainoskampanjoita (Killoran, 2013).

Hakukoneoptimointi painottuu hakukoneen luonnolliseen toimintaan eli hakukoneen algoritmeihin, eikä siihen siksi sisällytetä maksullisia palveluita, kuten hakukonemarkkinoinnissa (Killoran, 2013). Joskus luonnollisesta toiminnasta käytetään myös käsitettä ydin SEO (core SEO) ja maksullisesta toiminnasta non-core SEO (Davis, 2006). Näiden kahden käsitteen, hakukoneoptimoinnin ja hakukonemarkkinoinnin ero on se, että hakukoneoptimointi on kaikkien ulottuvilla ja, että se painottuu orgaanisiin hakutuloksiin eikä maksettuihin hakutuloksiin (Killoran, 2013).

Hakutulossivulla eli search engine result page:lla (SERP) on useita alueita (section) erilaisille hakutuloksille. Orgaanisiksi hakutuloksiksi sanotaan aluetta, johon ilmestyy hakukoneen luonnolliset hakutulokset (Höchstötter & Lewandowski, 2009). Nämä hakutulokset ovat pelkästään hakukoneen algoritmien etsimiä hakutuloksia eli niihin ei voi vaikuttaa rahalla (Killoran, 2013). Juuri näitä muutamaa ensimmäistä orgaanista hakutulosta tyypilliset käyttäjät klikkaavat ensimmäisenä (Höchstötter & Lewandowski, 2009). Toiseksi hakutulossivulta löytyy mainoksia, niin kutsuttuja sponsoroituja tuloksia (sponsored results), niin sivupalkista kuin myös orgaanisten hakutulosten päältä. Nämä sponsoroidut hakutulokset ovat siis maksettuja mainoksia (Killoran, 2013).

Hakutulossivulla organisten hakutulosten joukossa on yleensä ainakin yksi niin kutsuttu ”snippet”-hakutulos. ”Snippet”-hakutuloksella tarkoitetaan sellaista hakutulosta, jonka päälinkin alla on useampia alalinkkejä. Näitä alalinkkejä voivat olla esimerkiksi sivuston navigointilinkit. (Höchstötter & Lewandowski, 2009)

Hakutulossivu (kuva 2) koostuu mainoksista ja aidoista eli organisisista hakutuloksista. Kuvasta kaksi huomataan myös, että Google käyttää linkkeinä internetsivun otsikoita (title) ja internetsivun oikea web-osoite on vasta sen alapuolella. Web-osoitteen alapuolella oleva teksti on niin kutsuttu meta description eli sivun lyhyt kuvaus. Kuten kuvasta kaksi nähdään, joidenkin web-osoitteiden osia on lihavoitu. Nämä lihavoidut osat ovat suoraan otettu hakukenttään syötetystä hakusanasta, eli tässä tapauksessa hakusanasta ”vaihtoautot”.



Kuva 2. Kuvakaappaus Googlen hakutulossivusta (*google.fi*).

Hakukoneoptimointi voidaan jakaa kahteen alaluokkaan eettisyytensä mukaan: white hat- ja black hat-hakukoneoptimointiin. White hat-hakukoneoptimointi tarkoittaa sitä, että webmasterit, toisin sanoen internetsivuista vastaavat henkilöt, voivat käyttää hakukoneen omaa julkista dokumentaatiota ja julkisia parhaita käytäntöjä oman internetsivun mainostamisessa. Black hat-hakukoneoptimointi taas tarkoittaa sitä, että webmaster yrittää käyttää hakukoneelle ”laittomia” ja petollisia keinoja hyödykseen internetsivun markkinoinnissa. (Killoran, 2013)

Hakukoneoptimoinnissa yleensä puhutaan paljon avainsanoista (keywords). Näillä avainsanoilla tarkoitetaan siis käyttäjän kirjoittamaa hakulauseketta esimerkiksi Googlen hakukoneeseen, millä käyttäjä pyrkii löytämään haluamansa tiedon. (Stearns, 2013)

Avainsana voi olla siis esimerkiksi vaikka ”autokauppa” tai ”autokaupat Oulussa”. Avainsanoista kerrotaan enemmän luvussa kolme.

2.2. Hakukoneoptimoinnin hyödyt

Hakukoneet hallinnoivat noin 93 % koko internetin liikenteestä, joten tästä syystä on tärkeä ymmärtää miten kriittisiä hakukoneet ovat internetin käyttäjille (Egri & Bayrak, 2014).

Hakukoneet ovat mahdollistaneet internetissä olevaan valtavaan tietomäärään, sekä tutkijoille, että suuremmissa määrin julkiselle ja kaupalliselle puolelle. Tästä syystä hakukoneiden toimintaa alettiin arvioida ja tarkastelemaan lähemmin 1990-luvun lopulla. Haluttiin saada selville hakukoneiden algoritmien toimintaa ja haluttiin pyrkiä tunnistamaan tekijät, joilla on merkitystä hakukoneita ajatellen. (Zhang & Dimitroff, 2005)

Nykypäivänä suuret hakukoneet, kuten esimerkiksi Google, mahdollistavat ensisijaisesti sen, että potentiaaliset asiakkaat löytävät uusia palveluita. Hakukoneet ovat siis antaneet myyjille ja internetsivujen omistajille kannustimen optimoida sivujansa mahdollisimman hakukoneystävällisiksi (Wang, Savage & Voelker, 2011). Koska liiketoiminta on menossa enemmän internettiin, kuten verkkokauppatoimintaan, niin myyjien täytyy panostaa enemmän näkyvyyteen internetissä. Jotta he saisivat kilpailuedun yhä tiukentuvassa kilpailussa asiakkaista (Sagot, Fourgères, Ostrosi & Lacom, 2014). Palveluntuottajat pyrkivät saamaan internetsivun ensimmäiselle hakutulossivulle. iProspectin vuonna 2006 tehdyn tutkimuksen mukaan 62 % vastaajista klikkasi hakutulossivun ensimmäisellä sivulla olevia hakutuloksia (iProspect, 2006).

2.3. Hakukoneen yleinen toiminta

Hakukone on järjestelmä, joka ensin etsii tietoa eli internetsivuja internetistä käyttämällä siihen tarkoitettuja tietynlaisia ohjelmia eli hämähäkkejä. Tämän jälkeen se organisoii ja järjestele löytämänsä tiedon sellaiseen järjestelmälliseen muotoon, että käyttäjät voivat etsiä tietoa erilaisilla hakusanoilla ja löytää niiden avulla relevanttia tietoa. (Cui & Hu, 2011)

Hakukoneet voidaan luokitella erilaisiin luokkiin niiden toiminnan mukaan. Näitä luokkia ovat muun muassa kokonaisen tekstin hakukoneet, jotka indeksoivat internetsivuja omaan tietokantaansa, metahakukoneita, jotka käyttävät hyväksi muiden hakukoneiden indeksejä ja kokoavat niistä oman hakutuloksen, sekä vertikaaliset hakukoneet, joiden hakualue on rajattu vain tiettyyn aihealueeseen. Esimerkiksi Googlen hakukone on kokotekstihakua tukeva hakukone (full text search engine). (Cui & Hu, 2011)

Googlen hakukone pyrkii vähentämään roskapostia (spam). Uusia roskapostisivustoja luodaan kokoajan. Esimerkiksi Googlen ”How search works”-internetsivulta nähdään millaista roskapostitusta tapahtuu jatkuvasti. Roskapostiksi luetaan sellaiset sivut, jotka harjoittavat toimintaa vastoin Googlen webmaster ohjesääntöä. Tällaista toimintaa on piilotetun tekstin tai hakusanojen liiallinen käyttäminen, käyttäjien luoma roskapostimainen sisältö, väliaikaiset tai huonoa sisältöä sisältävät sivustot, sisään tulevien ja ulospäin menevien linkkien huono laatu, huono isäntäpalvelu (host), erilaisen tiedon näyttäminen ihmisille ja hakukoneille (cloaking) ja tietomurrettu sivusto. (Google, 2016)

2.4. Hämähäkit ja sivujen ryömiminen

Hämähäkeillä tarkoitetaan tietokoneohjelmia, jotka pyrkivät käymään läpi kaikki internetissä olevat internetsivut yksi kerrallaan. Tätä kutsutaan ryömimiseksi (crawling). Jokaisesta internetsivusta kerätään dataa, jonka perusteella hämähäkit pisteyttävät internetsivun. Tähän dataan kuuluu muun muassa kaikki yksittäiset sanat, niiden sijainnit dokumentissa, fontin ja alkukirjaimen koot. Toisekseen hämähäkit käyvät läpi kaikki internetsivulta löytyvät linkit ja tallentavat datan niistä erilliseen tiedostoon. Tässä tiedostossa on tieto jokaisesta linkistä mistä ja mihin se osoittaa, sekä myös tieto linkin tekstistä (Brin & Page, 2012). Kun internetsivu on pisteytetty, se tallennetaan hakukoneen tietokantaan, jota kutsutaan indeksiksi (index). Tästä tietokannasta hakukone pystyy hakemaan internetsivuja hakusanoilla. (Stearns, 2013)

Hämähäkit eivät ryömi ja pisteytä kokonaisia internetsivustoja kerralla, vaan ne pisteyttävät yksittäisiä internetsivuja, jotka ovat kytkeytyneet toisiin internetsivuihin linkeillä. Hämähäkit seuraavat linkkejä, joten on tärkeää, että internetsivuston sisällä olevat useat yksittäiset internetsivut ovat kaikki linkitetty niin, että mikään sivu ei jää linkittämättä. Jos internetsivua ei ole linkitetty, niin hämähäkit eivät löydä sitä. (Stearns, 2013)

Brin & Page (2012) mukaan Google käyttää useaa erilaista yksikköä hämähäkissään ja jokaisella yksiköllä on omat tietyt tehtävänsä (Brin & Page, 1998). Tässä tutkielmassa hämähäkin sisäiseen toimintaan ei oteta kantaa. Tämä taas johtuu siitä, että tämän työn tarkoitus on antaa peruskäsitys hakukoneoptimoinnista ja sen taustasta, eikä hämähäkin sisäisestä toiminnasta. Hämähäkit eivät näe internetsivuja samalla tavalla kuin ihmiset. Jotta ymmärtäisi mitä hämähäkit näkevät, niin täytyy tietää mistä tavallinen internetsivu koostuu. Internetsivun koostumus on hyvä tietää siksi, että voisi ymmärtää miksi joitakin hakukoneoptimointeja tehdään. (Stearns, 2013)

Internetsivut koostuvat useista erilaisista tiedostoista: HTML-tiedostoista (internetsivun rakenne), CSS-tiedostoista (Cascading Style sheets eli suomeksi tyylitiedostoista), PHP-tiedostoista (Hypertext Preprocessor eli serverin toiminnallisuuksista vastaavista tiedostoista), JS-tiedostoista (JavaScript-tiedostoista, jotka vastaavat asiakasohjelman eli clientin toiminnallisuuksista) ja muista erilaisista tiedostoista, kuten esimerkiksi kuvatiedostoista.

Käyttäjät voivat nähdä internetsivuilla useaa erilaista sisältöä kuten videoita, kuvia, linkkejä ja tekstiä. Yleensä internetsivuilla on myös monenlaista toiminnallisuutta, kuten ponnahdusikkunoita ja kuvagallerioita. Hakukoneiden hämähäkit eivät kuitenkaan näe näistä mitään. Ne näkevät vain riisutun mallin internetsivusta, jossa ei ole muuta kuin tekstiä ja linkkejä muille sivuille, joko saman verkkotunnuksen (domain) sisällä oleville tai sen ulkopuolella oleville (Stearns, 2013). Tästä syystä hakukoneoptimoinnissa keskitytään vain HTML-elementteihin eikä esimerkiksi tekstin tai taustan väriin. Toisin sanoen internetsivun muotoilulla (design) ei ole hämähäkeille suurtakaan merkitystä (Killoran, 2013).

3. Hakukoneoptimointi käytännössä

Tässä luvussa esitellään useita erilaisia tapoja miten hakukoneoptimointia voidaan tehdä sivuston sisällä (on-page SEO). Näiden kappaleiden tarkoituksena on selittää taustaa, miksi valittiin vain nämä alaluvuissa olevat hakukoneoptimointiin käytetyt metodit. On kuitenkin huomioitava, että tässä luvussa esiteltyt asiat ovat todellisuudessa vain pieniä muutoksia itse HTML-koodissa. Tämä tarkoittaa sitä, että yksittäiset muutokset eivät välttämättä tuo parannusta sivun näkyvyyteen vaan, että monesta pienestä muutoksesta voi saada lopulta suuren hyödyn.

Moz sivuston tekemän kyselyn (Moz, 2015) mukaan, johon vastasi 150 markkinoinnin asiantuntijaa, Googlen hakukoneen algoritmissä internetsivun sijoitukseen vaikuttaa eniten verkkotunnus-tason tekijät. Näitä ovat esimerkiksi verkkotunnuksen nimen pituus ja onko sivustolla SSL-sertifikaatio vai ei. Näihin domain tason tekijöihin liittyy myös sisäänpäin tulevien linkkien määrä ja niiden laatu. (Moz, 2015)

Su et al. (2014) toteavat tutkimuksessaan, että on olemassa kolme tekijää, jotka vaikuttavat huomattavasti sivun sijoitukseen Googlen hakukoneella tehdyissä hauissa. Heidän tekemässä tutkimuksessa pyrittiin saamaan selville mitkä ovat internetsivujen sijoitukseen vaikuttavia tekijöitä ja vahvistamaan jo tiedossa olevien tekijöiden oikeellisuutta. Jotta pystyttäisiin löytämään ja vahvistamaan hakukoneiden käyttämiä tekijöitä Su et al. (2014) suunnittelivat ja implementoivat oman sijoitusjärjestelmän (ranking system).

Su et al. (2014) mukaan eräs tekijä on Googlen käyttämä PageRank-algoritmi. PageRank-algoritmi perustuu internetsivulla oleviin linkkeihin. Näitä linkkejä on kahdenlaisia sisäänpäin ja ulospäin meneviä. Jokaisella linkillä, tai jokaisella internetsivulla, on oma pisteytys, joka kertoo miten tärkeä linkki on tai miten paljon linkillä on painoarvoa. Kaikki linkit eivät siis ole samanarvoisia. Tästä johtuen pelkkien linkkien määrä ei takaa hyvää pisteytystä PageRank-algoritmilta (Page, Brin, Motwani & Winograd, 1999).

Toiseksi Su et al. (2014) mainitsee, että sivulle sopivan hakusanan käyttäminen web-soitteessa (URL) on erittäin isossa roolissa hakukoneoptimoinnin kannalta. Tämä tarkoittaa käytännössä siis sitä, että pyritään valitsemaan asianmukaiset nimet verkkosoitteelle ja tiedostojen nimille. Kolmanneksi huomattavaksi tekijäksi Su et al. (2014) havaitsi metakuvaus-merkin (description tag) käytön internetsivun pääelementissä (header).

On olemassa kuitenkin ristiriitaista tietoa meta-merkin käytöstä. Googlen virallisessa blogissa (2007) sanotaan, että täsmällisen metakuvaus-merkin käyttö voi lisätä kävijämääriä, mutta se ei vaikuta sivun sijoitukseen millään lailla (Google, 2007). Toisin sanoen Googlen hämähäkit eivät ota huomioon metakuvauksessa olevia hakusanoja. Toisekseen meta-merkeillä ei ole internetsivun sijoituksessa paljoakaan painoa, sillä ne nähdään mahdollisuutena helppoon hakukoneoptimointiin täyttämällä ne hakusanoilla (keyword stuffing). Tästä taas johtuu se, että jotkin hakukoneet pitävät meta-merkkejä lähes turhina. (Su et al., 2014)

3.1. Otsikot ja alaotsikot

Jokaisessa HTML-tiedostossa eli jokaisella internetsivulla tulisi olla otsikko eli otsikko-merkki (title tag). HTML-tiedostossa on kolme peruselementtiä, jotka tulisi olla jokaisella

sivulla. Ensimmäiseksi html-merkki, joka sanelee tiedoston käyttävän HTML-kieltä. Toisekseen esitellään pää-merkki (head tag), jonka sisällä on edellä mainittu otsikko-merkki ja muita metatietoja sivusta, kuten esimerkiksi sivun lyhyt esittely (description). Kolmanneksi ja viimeiseksi sivustolla täytyy olla sivun runko-merkki (body tag), jolla tarkoitetaan yksinkertaisesti internetsivun pääsisältöä. (Noruzi, 2007)

```
<html>
  <head>
    <title>Sivun avainsanat järkevänä lauseena.</title>
  </head>
  <body>
    ...
  </body>
</html>
```

Otsikko-merkki määrittelee internetsivun otsikon, joka toimii itseasiassa Googlen hakutulossivulla (SERP) linkkinä sivulle, sekä se näkyy käyttäjän selaimessa välilehden otsikkona. Koska otsikolla on niin suuri merkitys hakukoneoptimoinnissa, niin sen tulisi sisältää internetsivulle tärkeitä ja relevantteja haku- ja avainsanoja. Tärkeimmän hakusanan tulisi olla otsikon alussa. On myös hyvä huomata, että jokaisella saman verkkotunnuksen (domainin) alla olevalla internetsivulla tulisi olla uniikki otsikko, sekä sen maksimi pituus tulisi pitää alle 120 merkissä (Stearns, 2013). Googlen ohjesäännössä mainitaan myös, että otsikossa käytettävän hakusanan tulisi olla selkeä eikä epämääräinen, kuten esimerkiksi ”New Page” (Google, 2010).

```
<html>
  <head> ... </head>
  <body>
    <h1>Pääotsikko</h1>
    <h2>Alaotsikko</h2>
    <h3>...</h3>
  </body>
</html>
```

Alaotsikoilla tarkoitetaan HTML-kielessä h1, h2, h3, h4, h5 ja h6 merkkejä. Nämä merkit sijoitetaan html-dokumentin runko osaan (body). Näistä kuudesta tagista ensimmäinen eli h1-merkki on kaikkein tärkein ja sillä on eniten painoarvoa, niin hakukoneen kuin myös lukijan kannalta. Tämän takia se käyttää oletuksena suurinta fonttikokoa kaikista alaotsikko merkeistä. H1-merkin tarkoitus on olla sivulla ensimmäinen näkyvä otsikko lukijalle ja siksi sen tulisi sisältää sivun ensisijainen sanoma. Toisin sanoen se ilmaisee mikä sivun tarkoitus on. Tämän takia internetsivulla saa olla vain yksi h1-merkki, siksi otsikkotasolle h1 kannattaa lisätä internetsivulle soveliaita hakusanoja. Kuitenkin aina, kun puhutaan hakusanoista, niin täytyy olla tietoinen, että ei täytä kyseessä olevaa datakenttää useilla epämääräisillä hakusanoilla (keyword stuffing). Muilla alaotsikko merkeillä pyritään taas jaottelemaan sivun sisältö järkeviin lohkoihin niiden tärkeyden mukaan. (Stearns, 2013)

Kuten aiemmin on todettu, niin hämähäkit tallentavat internetsivulla olevien sanojen fonttien koot dataksi, jota käytetään internetsivun sijoituksen laskemisessa. Ei ole kuitenkaan varmaa kuinka suuri painoarvo fontin koolla on esimerkiksi avainsanoissa, sillä fontin kokoa voidaan muuttaa tyylitiedoston avulla (CSS). Kuitenkin itse otsikko-merkillä on suuri painoarvo, sillä se kuuluu HTML-kielen semantiikkaan.

3.2. Metadata

Metadatalle tarkoitetaan tässä kontekstissa taustatietoa internetsivusta, joka on tarkoitettu hakutulossivulla olevaksi tiedoksi (Moz, 2016). Tähän taustatietoon kuuluu esimerkiksi sivun esittely (description), hakusanat (keywords), sivulla käytettävä kieli ja mitä merkistöä käytetään (charset). (Stearns, 2013)

```
<html>
  <head>
    ...
    <meta charset="UTF-8">
    <meta name="description" content="Lyhyt ja ytimekäs
    esittely sivusta.">
    <meta name="keywords" content="hakusana1, hakusana2,
    hakusana3">
  </head>
  <body>
    ...
  </body>
</html>
```

HTML-kielessä metatiedot syötetään sivun pää-merkin (head) sisään, kuten yllä olevasta esimerkistä nähdään. Tässä esimerkissä on käytetty UTF-8 merkistöä, jolla mahdollistetaan skandinaavisten merkkien käyttö internetsivulla. Toisekseen esimerkiksi voidaan nähdä internetsivun esittely (description) ja käytetyt kolme avainsanaa (keyword).

Kuten aiemmin kuvasta 2. huomattiin, niin esittely (description) metatietoa käytetään hakutulossivulla olevien hakutulosten esittelytekstinä, ja vaikka Googlen blogikirjoituksen (2007) mukaan esittelyllä ei ole mitään vaikutusta internetsivun sijoitukseen, on se silti hyvä olla kirjoitettuna. Sillä se voi lisätä internetsivun kävijämääriä (Google, 2007).

3.3. Sivun sisältö ja runko-merkki

Internetsivun hyvä sisältö on yksi tärkeä osa hakukoneoptimointia. On erittäin tärkeä muistaa, että internetsivuja tehdään ihmisten eikä koneiden luettavaksi. Internetsivulla oleva sisältö voi vaikuttaa painoarvoltaan eniten internetsivun näkyvyyteen. Tämä johtuu siitä, että käyttäjät yleensä huomaavat hyvän sisällön ja tahtovat jakaa sitä muille. Tällainen kuulopuhe (word-of-mouth) on erittäin positiivista sivuston näkyvyyden kannalta. (Google, 2010) Tämä taas johtaa siihen, että internetsivulle saadaan sisäänpäin tulevia linkkejä, joka taas lisää PageRank-sijoitusta tiettyyn pisteeseen asti. (Page, Brin, Motwani & Winograd, 1999)

Oikea ja orgaaninen sisältö on myös huomattavissa, sekä käyttäjän, että hakukoneen näkökulmista. Hakukone huomaa hetkessä, jos internetsivulla on käytetty esimerkiksi samoja hakusanoja useita kertoja, vaikkapa title tagissa, h1 tagissa ja ensimmäisessä kappaleessa. Tästä se tietää, että sivua on yritetty hakukoneoptimoida tietyille hakusanalle tietoisesti. Tätä kutsutaan hakusanojen väärinkäytöksi (keyword stuffing/abuse), joka taas vähentää internetsivun sijoitusta huomattavasti (Stearns, 2013)

3.4. Sitemap

Sivustokartta (sitemap) tarkoittaa listaa sivustolla olevista internetsivuista. Sivustokartoja on kahdenlaisia: käyttäjille tarkoitettuja ja hakukoneille tarkoitettuja.

Käyttäjille tarkoitettu sivustokartta on yleensä vain internetsivulla oleva lista linkeistä, joita klikkaamalla käyttäjä pystyy navigoimaan sivustolla. (Google, 2010)

Hakukoneille tarkoitettu sivustokartta taas tarkoittaa xml-tiedostoa, joka on ainoastaan tarkoitettu hakukoneen hämähäkin luettavaksi. Tästä tiedostosta käy ilmi, mitä eri sivuja sivustolla on ja mistä sivut löytyvät. Koska voi käydä niin, että jotkin sivut eivät näy Googlen hämähäkeille esimerkiksi rikkoontuneen linkin takia, joten tämä sivustokartta antaa kuitenkin osoitteen, mistä kyseinen sivu löytyy. (Google, 2010) Sivustokartasta pystytään määrittelemään onko sivuston sivujen hierarkia kunnossa vai ei. Googlen (2010) mukaan parhaisiin käytäntöihin kuuluu se, että pyritään menemään yleisestä sisällöstä spesifisempään sisältöön ja vielä niin, että käyttäjän ei tarvitse käyttää kymmeniä klikkauksia eri sivujen välillä päästäkseen haluamalleen sivulle.

3.5. Robots tekstitiedosto

Robots tekstitiedosto on pelkkää tekstiä sisältävä tiedosto, joka sanelee Googlen hakukoneen hämähäkille, että mitkä sivut saa indeksoida ja mitkä ei. Yksityistä tai arkaluontoista tietoa sisältävät sivut voidaan siis jättää indeksoinnin ulkopuolelle. On kuitenkin huomioitava, että sivut silti löytyvät internetistä. (Google, 2010)

Robots tekstitiedostoa matkiva HTML-kielessä käytettävä linkkien erikoisattribuutti ”nofollow” on myös hyvä käytäntö, jolla sanellaan hakukoneen hämähäkille seurataanko linkkiä vai ei.

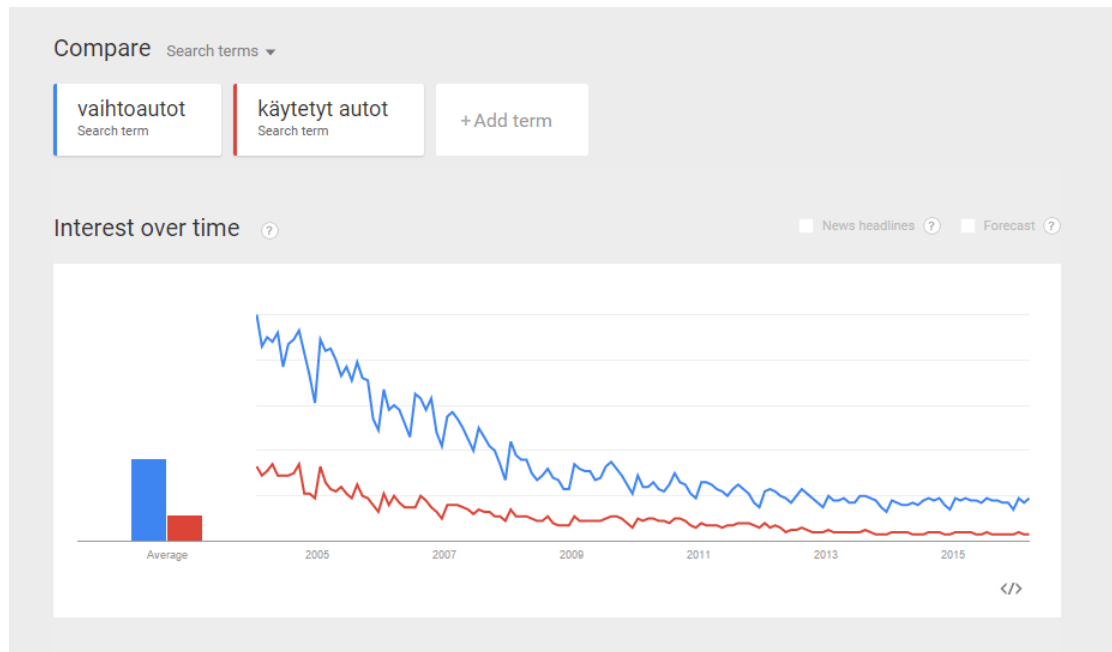
```
<a href="http://www.roskapostisivusto.net" rel="nofollow">Tämä on hyvä sivu!</a>
```

Erikoisattribuuttia ”nofollow” käytetään linkki-merkin (a) sisällä, kuten yllä on esitetty. Nofollow-erikoisattribuutti on käytännöllinen esimerkiksi silloin, kun internetsivulla on kommenttipalsta, johon käyttäjät voivat kommentoida ja liittää linkkejä. Jokaiseen linkkiin voidaan lisätä nofollow-erikoisattribuutti, jolloin hämähäkki ei ota huomioon linkkiä ollenkaan. Näin saadaan kitkettyä roskapostit pois sivulta. (Google, 2010)

3.6. Avainsanat

Avainsanalla (keyword), tarkoitetaan sanaa tai useita peräkkäisiä sanoja, joilla pyritään hakemaan tietoa hakukoneen avulla juuri tietyistä aiheista. Avainsanoja tulee käyttää orgaanisesti internetsivulla. Tällä tarkoitetaan sitä, että samaa avainsanaa ei toisteta useasti peräkkäin. Tämän takia onkin olemassa avainsanojen tiheyden mittaamiseen olevia ohjelmia, joilla voidaan löytää juuri oikea suhde avainsanoja ja muuta tekstiä sisällöstä. (Stearns, 2013) Tämä kuitenkin on ristiriidassa esimerkiksi Googlen hakukoneoptimoinnissa käytettäviä menetelmiä, sillä tekstin tulisi olla luontevaa eikä hienosäädettyä ja tarkalleen optimoitua (Google, 2010).

Yhtenä päätekijänä hakukoneoptimoinnissa on internetsivun kontekstiin sopivien avainsanojen käyttäminen internetsivun sisällössä. Internetsivun näkyvyyttä voidaan parantaa lisäämällä näiden avainsanojen tiheyttä internetsivun otsikossa ja sisällössä tiettyyn pisteeseen asti. Tästä pisteestä, joka ei ole täysin selvillä, alkaa internetsivun sijoitus kuitenkin tippua. (Zhang & Dimitroff, 2005)



Kuva 3. Kuvakaappaus Google Trends ohjelmasta (Google Trends, 2016).

Yllä olevasta kuvasta (kuva 3) nähdään kahden samaa tarkoittavan, mutta erilaisen hakusanan käyrät Google Trends web-aplikaatiossa. Tämä kuva kertoo sen, että ajan mittaan näiden hakusanojen käyttö on huomattavasti vähentynyt. Kuitenkin näiden kahden hakusanan välillä on selvä ero. Hakusanaa ”vaihtoaunut” käytetään huomattavasti enemmän Googlen hauissa kuin hakusanaa ”käytetyt autot”.

Tästä voidaan tulla siihen tulokseen, että hakusanalla ”vaihtoaunut” saadaan enemmän liikennettä internetsivulle kuin hakusanalla ”käytetyt autot”. On kuitenkin mahdollista, että hakusanaa ”vaihtoaunut” käytetään paljon jo olemassa olevien autokauppojen internetsivuilla hakusanana ja siksi sen valinta esimerkiksi uuden autoliikkeen internetsivun hakusanaksi voi olla väärä ratkaisu. Toisin sanoen hakusanalla ”käytetyt autot” uusi internetsivu voi saada enemmän näkyvyyttä Googlen hakukoneelta, koska sillä on vähemmän kilpailua vanhojen internetsivujen kanssa. Tätä hakusanojen etsimistä ja vertailu kutsutaan hakusana-analyysiksi (keyword research).

3.7. Responsiivisuus

Ihmiset käyttävät yhä enemmän mobiililaitteita internetin selaamiseen, joten on selvää, että internetsivuja täytyy myös pystyä käyttämään sujuvasti mobiililaitteilla (Mohorovičić, 2013). Siksi olisi suotavaa, että internetsivu olisi responsiivinen eli sivu muokkautuu laitteen näytön resoluutiolle sopivaksi niin, että käytettävyys ei kärsi. Responsiivisuuden yksi tärkeimmistä avainhyödyistä on laitteistoriippumattomuus (Mohorovičić, 2013). Toisin sanoen internetsivua voidaan käyttää täysin laitteesta riippumatta.

Metatieto-merkki sijoitetaan pää-merkin (head) sisään ja se kertoo käytettävälle selaimelle, että internetsivu on skaalautuva. Kun sovellusikkunan (viewport) meta-merkki on luotu, niin voidaan käyttää tiedusteluita (media query). Ne ovat yksinkertaisia jos-lauseita käytettävän laitteen näytön tai selaimen ikkunan leveydestä. Näin selain osaa ajaa vain tietyn koodin kerrallaan.

```
<meta name="viewport" content="width=device-width, initial-scale=1">
```

Responsiivisuuden takaamiseksi internetsivu voidaan rakentaa alusta asti käyttämällä kolmannen osapuolen valmistamia kehikkoja (framework), jotka sanelevat ruudukon (grid), jonka sarakkeiden leveys muuttuu laitteen näytön koon mukaan. Yleensä nämä ruudukot jakavat internetsivun kahteentoista samankokoiseen sarakkeeseen, mutta on olemassa myös esimerkiksi 16 sarakkeisia ruudukkoja. Tunnetuimpia kehikkoja ovat muun muassa foundation ja unsemantic. Nämä kehikot mahdollistavat mobiilisivujen luonnin ensimmäiseksi. Tätä kutsutaan mobiilisivut ensin-menetelmäksi (mobile-first).

Internetsivun responsiivisuus voidaan luoda myös käyttämällä CSS-kielen (cascading style sheets) tiedusteluita, mutta yleensä molempia sekä kehikkoja että tiedusteluita käytetään responsiivisuuden takaamiseksi internetsivuilla.

```
@media screen and (max-width: 600px) {
  .img {
    width: 50%;
  }
}
```

Yllä oleva koodi on tiedustelu, jossa tarkastetaan onko laitteen näytön leveys alle 600 pikseliä. Jos se on niin, muokataan img-nimistä luokkaa siten, että sen leveys puolitetaan. Tällä pyritään siihen, että kuva skaalautuu laitteen näytön mukaan.

3.8.Kuvat

Googlen SEO starter guiden (2010) mukaan jokaiselle internetsivulla olevalla kuvalla tulisi olla lisättyä "alt" attribuutti. Tällä tarkoitetaan yksinkertaisesti lyhyttä kuvausta kuvasta. Kuvaus näytetään kuvan sijasta käyttäjälle silloin kun kuva ei jostain syystä lataudu tai käyttäjä käyttää tekstilukijaa, joka ei tue kuvia (Google, 2010). Koska hakukoneiden hämähäkit eivät pysty käsittelemään kuvia, niin siksi kuvien turhaa käyttöä tulisi välttää. Tällä tarkoitetaan sitä, että esimerkiksi linkkien ja navigoinnin tulisi olla tekstiä.

```

```

Toisekseen Googlen (2010) mukaan kaikkien kuvien tulisi olla sijoitettuna palvelimella samassa kansiossa. Tämä on vain hyvä käytäntö, joka ei vaikuta internetsivun sijoitukseen. Kolmanneksi kuvien tulisi olla kaikkien käytetyimpien selainten tukemissa tiedostomuodoissa. Nämä tiedostomuodot ovat jpeg, gif, png ja bpm. (Google, 2010)

4. Pohdinta ja yhteenveto

Internet on erittäin laaja tiedonlähde, joten tarvittavan tiedon etsimiseen tarvitaan hakukoneita, jotka etsivät dataa hakusanoilla. Nämä hakukoneet ohjaavat 93 % koko internetin käyttäjävirrasta (Egri & Bayrak, 2014). Käyttäjät siis pääsääntöisesti kulkeutuvat internetsivuille hakukoneiden kautta. Tästä syystä hakukoneoptimointia ei pidä aliarvioida. Hakukoneoptimoinnilla pyritään parantamaan internetsivun näkyvyyttä erilaisin keinoin. Internetsivun näkyvyys auttaa internetsivua saamaan enemmän kävijöitä, joka taas tarkoittaa sitä, että internetsivulla on enemmän potentiaalisia asiakkaita. Hakukoneoptimointi voidaan jakaa kahteen erilliseen luokkaan: on-page ja off-page hakukoneoptimointiin (Stearns, 2013).

On-page hakukoneoptimoinnilla tarkoitetaan internetsivun lähdekoodin muokkausta hakukoneystävällisemmäksi hakukoneyritysten antamien ohjeiden mukaan. On-page hakukoneoptimointi on täysin internetsivun omistajan päätettävissä. Tässä työssä pyrittiin tuomaan esille on-page hakukoneoptimoinnin tapoja. Off-page hakukoneoptimoinnilla tarkoitetaan internetsivun ulkoista hakukoneoptimointia. Se keskittyy ulkoapäin tuleviin linkkeihin. Tällä tarkoitetaan sitä, että muut internetsivut linkittävät itsensä internetsivuumme. Off-page hakukoneoptimointi on siis täysin internetsivun omistajan päätättävällän ulkopuolella. Internetsivun omistaja ei voi tehdä off-page hakukoneoptimointia omalle sivulleen.

Google käyttää näitä ulkoapäin tulevia linkkejä laskiessaan PageRank-sijoitusta internetsivulle. PageRank-algoritmi on vain yksi osa, joka vaikuttaa internetsivun näkyvyyteen Googlen hakukoneella. PageRank-algoritmi toimii niin, että se pisteyttää kaikki ulkoapäin tulevat linkit niiden laadun mukaan. Esimerkiksi jos sivu, josta tulee sisäänpäin linkkejä meidän sivulle, on erittäin suosittu ja aito internetsivu, niin meidän internetsivu saa algoritmin mukaan paljon pisteitä.

Hakukoneoptimointi voidaan jakaa myös eettisyyden mukaan kahteen eri luokkaan (Killoran, 2013). White-hat hakukoneoptimoinnilla tarkoitetaan hakukoneoptimointia, joka noudattaa hakukoneiden yleisiä hyviä käytänteitä. Ne pyrkivät siihen, että ei käytettäisi black-hat hakukoneoptimointia. Black-hat hakukoneoptimoinnilla tarkoitetaan internetsivun näkyvyyden parantamista keinoja kaihtamatta. Siinä käytetään esimerkiksi maksettuja linkki-farmeja (Killoran, 2013). Nämä linkki-farmit ovat internetsivuja, joilla on tarkoituksena vain ulospäin lähtevien linkkien ylläpitäminen. Linkki-farmien toiminta perustuu Googlen PageRank-algoritmiin, joka pisteyttää internetsivut sisäänpäin tulevien linkkien määrän ja laadun mukaan (Page, Brin, Motwani & Winograd, 1999). Linkki-farmit voivat osoittaa internetsivuumme esimerkiksi 10 000 eri linkillä. Yleensä kuitenkin nämä linkki-farmit ovat huonolaatuisia, joten niistä lähtevät linkit saavat vain vähän painoarvoa Googlen PageRank-algoritmissa.

On olemassa myös tietoturvahyökkäyksiä, jotka kohdistuvat hyvän näkyvyyden omaaviin internetsivuihin. Näillä hyökkäyksillä pyritään tiputtamaan kohteena olevan internetsivun sijoitusta ja nostamaan jonku muun sijoitusta ylemmäksi. Tällainen hyökkäys toimii niin, että kohteena olevalle internetsivulle pyritään sijoittamaan salaa linkkejä erilaisille roskapostisivustoille. Hyökkääjä tietää, että nämä roskapostisivustoille johtavat linkit haittaavat huomattavasti kohteena olevan sivuston näkyvyyttä. (Dye, 2008)

Erilaisia hakukoneoptimoinnin tapoja on useita erilaisia. Tässä työssä mainittiin vain muutama näistä tavoista. Nämä mainitut tavat on pyritty valitsemaan niin, että niillä on

tutkitusti suurimmat vaikutukset internetsivun näkyvyyteen. Muita tapoja ovat HTML-kielen semanttisuus lähdekoodissa ja internetsivun nopea latautumisnopeus. Semanttisuudella tarkoitetaan HTML-kielessä sitä, että käytetään HTML 5 antamia merkkejä (tag) hyväksi. Esimerkiksi kun luodaan sivupaneelia (sidebar), niin voidaan käyttää aside-merkkiä. Sivupaneelin voi luoda myös epäsemanttisesti käyttämällä div-merkkiä, joka on yleinen sisällön erottelumerkki. Div-merkistä ei voida tietää suoraan mitä se pitää sisällään, kun taas aside-merkistä voi.

On myös olemassa hakukoneoptimointiin rinnastettava tapa optimoida internetsivuja sosiaalisen median avulla. Tätä kutsutaan sosiaalisen median optimoinniksi (social media optimization). Sillä tarkoitetaan sitä, että pyritään jatkuvasti luomaan uutta ja mielenkiintoista sisältöä esimerkiksi Facebookiin tai Twitteriin. Tällä halutaan saada paljon tykkäyksiä ja uudelleen twiittauksia, koska ne jakavat sisällön tykkääjän koko sosiaaliselle verkostolle. (Bhargava, 2010)

Googlen hakukone uusiutui vuonna 2009. Silloin se otti käyttöön hakutulosten personoinnin kaikille käyttäjille. Hakutulosten personoinnilla tarkoitetaan sitä, että käyttäjälle näytetään Googlen mielestä oleellimmat hakutulokset käyttäjän hakuhistorian mukaan. Googlen käyttämä personointi toimii itsenäisenä Google-tilin ja yleisen selaushistorian rinnalla. (Google, 2009) Personoinnin takia Googlen hakukoneen käyttäjät saavat eri hakutuloksia perustuen heidän hakuhistoriaansa. Personoinnilla voidaan siis rajata käyttäjän tietämättä aihealue ”turvalliseen” alueeseen. Tällöin käyttäjä ei välttämättä löydä haluamaansa tietoa ollenkaan hakukoneella tai käyttäjä voi luulla kaiken merkittävän tiedon olevan esillä.

Tässä työssä käytiin läpi hakukoneita ja hakukoneoptimointia, sekä niihin liittyviä ilmiöitä. Tämän työn tarkoituksena oli tuoda esille hakukoneoptimoinnissa käytettäviä menetelmiä ja perustella niitä olemassa olevan artikkelien, tutkimusten ja kirjallisuuden avulla. Jatkotutkimuksen aiheita hakukoneoptimointiin liittyen on varmasti paljon. Tässä työssä käsiteltiin vain pinnallisesti personointia, haku- ja sijoitus-algoritmeja, sekä Googlen historiaa. Nämä ovat kaikki hyviä aiheita jatkotutkimuksen kannalta. Hakukoneoptimointi on aiheena erittäin laaja, joten tässä työssä ei myöskään selitetty tarkasti miten paljon sosiaalinen media vaikuttaa hakutuloksiin ja hakukoneiden toimintaan.

Tämän työn rajaukseen kuului, että tutkielmassa tarkastellaan vain Googlen hakukonetta ja sen toimintaa. Googlen hakukoneen toiminnasta rajattiin pois PageRank-algoritmin läpikäyminen matemaattisella tasolla. Yleisestä hakukoneoptimoinnin tarkastelusta jätettiin pois hakukoneiden merkitys markkinoinnin kannalta, vaikka se liittyy hakukoneoptimointiin. Markkinointi rajattiin pois, koska tämä tutkielma painottui tietojenkäsittelyyn ja hakukoneiden järjestelmiin.

Lähteet

- Bhargava, R. (2010). The 5 NEW Rules Of Social Media Optimization (SMO). Lainattu 21.3.2016, saatavilla: <http://www.rohitbhargava.com/2010/08/the-5-new-rules-of-social-media-optimization-smo.html>
- Brin, S., & Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 56(18), 3825.
- Broder, A. (2002). A taxonomy of web search. *SIGIR Forum*, 36(2), 3-10.
- Cui, M., & Hu, S. (2011). Search engine optimization research for website promotion. *Information Technology, Computer Engineering and Management Sciences (ICM), 2011 International Conference on*, 4. pp. 100-103.
- Davis, H. (2006). *Search engine optimization " O'Reilly Media, Inc."*.
- Dye, K. (2008). Website abuse for search engine optimisation. *Network Security*, 2008(3), 4-6.
- Egri, G., & Bayrak, C. (2014). The role of search engine optimization on keeping the user on the site. *Procedia Computer Science*, 36, 335-342.
- Evans, M. P. (2007). Analysing google rankings through search engine optimization data. *Internet Research*, 17(1), 21-37.
- Google (2016). How search works. Lainattu 21.3.2016, saatavilla: <http://www.google.com/insidesearch/howsearchworks/thestory/>
- Google (2010). Search engine optimization starter guide. Lainattu 21.3.2016, saatavilla: <http://static.googleusercontent.com/media/www.google.com/fi/webmasters/docs/search-engine-optimization-starter-guide.pdf>
- Google (2009). Personalized Search for everyone. Lainattu 21.3.2016, saatavilla: <https://googleblog.blogspot.fi/2009/12/personalized-search-for-everyone.html>
- Google (2007). Improve snippets with meta description. Lainattu 21.3.2016, saatavilla: <https://webmasters.googleblog.com/2007/09/improve-snippets-with-meta-description.html>
- Google Trends (2016). Lainattu 1.4.2016, saatavilla: <https://www.google.fi/trends/>
- Höchstötter, N., & Lewandowski, D. (2009). What users see – structures in search engine results pages. *Information Sciences*, 179(12), 1796-1812.
- Internetlivestats (2013). Google search statistics. Lainattu 21.3.2016, saatavilla: <http://www.internetlivestats.com/google-search-statistics/#trend>
- iProspect (2006). Search engine user behaviour study. Lainattu 21.3.2016, saatavilla: http://district4.extension.ifas.ufl.edu/Tech/TechPubs/WhitePaper_2006_SearchEngineUserBehavior.pdf

- Killoran, J. B. (2013). How to use search engine optimization techniques to increase website visibility. *Professional Communication, IEEE Transactions on*, 56(1), 50-66.
- Malaga, R. A. (2008). Worst practices in search engine optimization. *Communications of the ACM*, 51(12), 147-150.
- Moz (2016). Meta description. Lainattu 21.3.2016, saatavilla: <https://moz.com/learn/seo/meta-description>
- Moz (2015). Search ranking factors. Lainattu 21.3.2016, saatavilla: <https://moz.com/search-ranking-factors>
- Netmarketshare (2016). Desktop search engine market share. Lainattu 21.3.2016, saatavilla: <http://www.netmarketshare.com/>
- Noruzi, A. (2007). A study of HTML title tag creation behavior of academic web sites. *The Journal of Academic Librarianship*, 33(4), 501-506.
- Mohorovičić, S. (2013). Implementing responsive web design for enhanced web presence. *Information & Communication Technology Electronics & Microelectronics (MIPRO)*, 2013 36th International Convention on, pp. 1206-1210.
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. *Stanford InfoLab; Stanford InfoLab*.
- Sagot, S., Fougères, A., Ostrosi, E., & Lacom, P. (2014). Search engine optimization: From analysis based on an engineering meta-model towards integrative approaches. *Information Society (i-Society)*, 2014 International Conference on, pp. 274-281.
- Statisticbrain (2015). Google annual search statistics. Lainattu 21.3.2016, saatavilla: <http://www.statisticbrain.com/google-searches/>
- Stearns, J. (2013). Search engine optimization. Lainattu 21.3.2016, saatavilla: <https://web.archive.org/web/20130925162821/http://www.amprewave.com/guides/search-engine-optimization/>
- Su, A., Hu, Y. C., Kuzmanovic, A., & Koh, C. (2014). How to improve your search engine ranking: Myths and reality. *ACM Transactions on the Web (TWEB)*, 8(2), 8.
- Wang, D. Y., Savage, S., & Voelker, G. M. (2011). Cloak and dagger: Dynamics of web search cloaking. *Proceedings of the 18th ACM Conference on Computer and Communications Security*, Chicago, Illinois, USA. pp. 477-490.
- Zhang, J., & Dimitroff, A. (2005). The impact of metadata implementation on webpage visibility in search engine results (part II). *Information Processing & Management*, 41(3), 691-715.
- Zhang, J., & Dimitroff, A. (2005). The impact of webpage content characteristics on webpage visibility in search engine results (part I). *Information Processing & Management*, 41(3), 665-690.