

Mirja Iivonen
Kai Halttunen

Web-tiedonhaun tarkastelua

Johdanto

Informaatioympäristössä 1990-luvulla tapahtuneet suurimmat muutokset liittyvät Internetiin, joka on kasvanut nopeasti ja tavoittaa tänään jo miljoonia käyttäjiä eri puolilla maailmaa. Internet on avoin tietoverkko, johon voidaan laskea kuuluvaksi fyysiset tiedonsiirtoverkot, eri pisteissä fyysisesti sijaitsevat tietokone-laitteet ja niissä käytettävät ohjelmistot sekä tietokoneiden muistiin tallennetut tiedostot ja niissä oleva informaatio. World Wide Web, jatkossa web, on Internetin kautta tapahtuvaan tiedonhaakuun kehitetty palvelu, jonka kautta voidaan siirtää dokumentteja. Tiedon saatavuuden kannalta web voidaan ymmärtää suurena, lähes rajattomana tietovarantona, koska suuri määrä potentiaalisesti hyödyllistä informaatiota on sen kautta saavutettavissa riippumatta siitä, mihin ko. informaatio on fyysisesti tallennettuna.

Vaikka vain pieni osa webin käytöstä on tarkoituk-sellista ja suunniteltua tiedonhakua, niin informaatio-tutkimuksen näkökulmasta katsottuna web-ympäris-tössä tapahtuvan tiedonhaun tarkastelu on tärkeää. Kiinnostavia kysymyksiä ovat mm. kuinka käyttäjät hakevat tietoa webistä ja miten heidän tiedonhakunsa webissä etenee, mitä webin tiedonhakupalveluita he käyttävät, kuinka käyttäjät arvoivat webistä löytämänsä informaatiota ja mihin tekijöihin perustuu heidän arvi-onsa webistä lyödetyn tiedon laadusta ja rele-vanttiudesta.

Webistä suoritettavaan tiedonhaakuun liittyy vielä paljon ongelmia. Tiedonhaku voi olla hyvin hämmen-tävää ja turhauttavaa, koska web poikkeaa huomatta-vasti traditionaalisesta informaatioympäristöstä. Se on suuri, hajanainen ja nopeasti muuttuva. Jotkut web-dokumentit voivat ilmestyä, kadota ja ilmestyä uudelleen. Vanha web-dokumentti voi korvautua uu-

della, samalla kun uusi web-dokumentti osoittautuukin jo jossakin olemassa olevaksi dokumentiksi. Sama web-dokumentti voi myös sijaita samanaikaisesti use-assa eri osoitteessa. (Hjerppe 1996.) Käyttäjä voi ha-kiessaan löytää helposti liikaa dokumentteja, ylikuormittua saamastaan informaatiosta ja kokea do-kumenttien valikoinnin vaikeaksi. Käyttäjä voi myös turhautua siitä, että hänellä on valittavanaan liian pal-jon vaihtoehtoja tiedonhaun jatkamiseksi. Käyttäjä voikin helposti tyytyä ensimmäisiin, mutta lähes epärelevantteihin dokumentteihin (Wang & Tenopir 1998).

Tiedonhaku webistä on vaikeaa edelleen myös sik-si, että web-dokumentit eivät sisällä riittävästi metadataa. Metadata on dataa datasta, vaikka siitä joskus yksinkertaistaen puhutaan metatietona, eli tie-tona tiedosta. Metadata tarkoituksena webissä on kuvailla webin sisältämiä dokumentteja. Tarkemmin se voidaan määritellä esim. seuraavasti: "Metadata is data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics." (Dempsey 1998.) Lisäksi webin hakupalveluiden katteesta, ajan-tasaisuudesta ja hakuominaisuuksista on vaikeaa ellei mahdotonta saada tietoa. Palvelut tyytyvät usein aino-astaan automaattiseen tietokantojen muodostamiseen, ilman tarpeellisia lisäarvopalveluita kuten esimerkiksi dokumenttien kuvailua ja arviointia. Tästä johtuen käyttäjällä voi web-tiedonhaun tuloksena olla valtava määrä epärelevantteja web-dokumentteja samaan ai-kaan kun häneltä on jäänyt löytymättä relevantteja web-dokumentteja.

Web-tiedonhausta jo tiedettyä

Webistä tapahtuvaan tiedonhaakuun ja -hankintaan liittyen on ilmestynyt jo melko paljon kirjallisuutta. Kar-keasti tämä kirjallisuus voidaan jakaa kahtia: opas-kirjoihin (guidebooks) ja tutkimuksiin. Useimmat opas-kirjat käsittelevät webistä tapahtuvaa tiedonhakua ja -hankintaa melko teknisesti. Ne kuvaavat, kuinka web tai laajemmin tarkasteltuna Internet toimivat ja minkä-laisia palveluita käyttäjille niiden kautta on saatavilla. Hakupalveluiden ja hakumahdollisuuksien lisäksi ne käsittelevät myös sähköpostia, postituslistoja, tiedostojen siirtoa, keskusteluryhmiä, reaa-liakommunikaatiota jne. (ks. esim. Comer 1994,

Wiggings 1994, Kriesel 1998, Kuvaja 1997, Keski-kiikonen 1997). Jotkut opaskirjat keskittyvät jonkun tietyn palvelun tai ohjelman tarkasteluun (ks. esim. Grimes & Kinkoph 1997, Irvine 1997). Web-tiedonhaakuun keskittyneitä oppaita on ilmestynyt kuitenkin vain muutamia (ks. esim. Eager ...et al. 1995, Korpi-mies 1995, Pfaffenberger 1996).

Opaskirjojen tarkoitus on esitellä lukijoille Internet, sen tekniset ratkaisut ja/tai sen tarjoamat mahdollisuudet, neuvoa heitä Internet-palveluiden käytössä ja näin ollen viime kädessä auttaa lukijoita toimimaan onnistuneesti uudessa elektronisessa ympäristössä. Vaikka nämä opaskirjat ovat hyödyllisiä, ne eivät vielä kerro, mitä ihmiset todella tekevät, kun he käyttävät Internetiä tai kun he etsivät tietoa webistä. Opaskirjat kertovat ainoastaan sen, mitä ihmisten tulisi tehdä ja kuinka heidän tulisi käyttäytyä selviytyäkseen web/Internet-ympäristössä. Monesti opaskirjojen näkökulma onkin tietojenkäsittelyn näkökulma. Teknisen näkökulman korostuminen on haitallista silloin, jos ihmisiä pyritään yksioikoisesti esim. sopeuttamaan käyttäytymisensä uuden tekniikan mukanaan tuomiin muutoksiin, vaikka tekniikkaa pitäisi kehittää vastaamaan aiempaa paremmin ihmisten tapaa toimia. Tähän pääsemiseksi web-palveluiden kehittäjien pitäisi nykyistä paremmin tietää miten ihmiset käyttäytyvät hakiessaan webistä tietoa.

Internetiin ja webiin liittyen on julkaistu myös jo melko paljon tutkimuskirjallisuutta. Tutkimuskirjallisuudessa on kuitenkin keskitytty pääasiassa muihin kun tiedonhakuprosessia koskeviin kysymyksiin.

Tutkimuksissa on tarkasteltu seuraavia kysymyksiä:

1) Kuka käyttä Internetiä/webiä? Huomio on kiinnitetty käyttäjien demografisiin piirteisiin ja asemaan (ks. esim. Sproull, Kiesler and Siegel 1986, Tillotson, Cherry and Clinton 1995, Perry 1995, Kamner 1997, Senkevitch and Wolfram 1997).

2) Mitä Internetin palveluita tiedon hankintaan käytetään? Huomiota on kiinnitetty mm. sähköpostin, postituslistojen, keskusteluryhmien, tiedostopalvelimien, elektronisten lehtien ja erilaisten hakupalveluiden käyttöön. (ks. esim. Rice and Case, 1983, Carley and Wendt, 1991, Burton, 1994, Tillotson, Cherry and Clinton, 1995, Kamner, 1997, Merritt & Schechter, 1997).

3) Mihin Internetiä käytetään? On tarkasteltu mm. sitä, käytetäänkö Internetiä työtehtäviä, arkielämää vai harrastustoimintaa koskevan tiedon hankintaan (ks. esim. Sproull and Kiesler 1986, Tillotson, Cherry and Clinton 1995, Savolainen 1998).

4) Ovatko käyttäjät tyytyväisiä Internetiin? (ks. esim. Tillotson, Cherry and Clinton 1995, Covi and Kling 1996).

Itse tiedonhakuprosessia, ts. webissä suoritettavan tiedonhaun etenemistä ja sen aikana tehtäviä päätöksiä on toistaiseksi tutkimuksissa tarkasteltu melko vähän. Catledge ja Pitkow (1995, %: Hsieh-Yee 1998) tutkivat web-käyttäjien navigointistrategioita analysoimalla lokitiedostoja. He havaitsivat hakijoiden etenevän tiedonhaussa pääasiassa selailemalla linkkejä eteen- ja taaksepäin. He tunnistivat myös kolme erilaista selaajatyyppeä, satunnaisselaajat (screndipitous browser), yleisselaajat (general purpose browser) ja hakijat (searcher).

Hsieh-Yee (1998) tutki web-hakijoiden hakutaktiikoita simulaatioharjoitusten avulla. Hakijoita pyydettiin valitsemaan hakupalvelu ja muotoilemaan hakulauseke neljää eri hakutehtävää varten. Kahdessa hakutehtävässä piti löytää tietty web-dokumentti, kahdessa muussa edellisille samankaltaisia web-dokumentteja. Hakijoilta kysyttiin myös, miten he jatkaisivat hakua, jos he löytäisivät liikaa tai liian vähän dokumentteja. Hsieh-Yee havaitsi, että hakijat aloittaisivat hakunsa useimmiten sanahakupalvelulla. Hän havaitsi myös, että selailu oli suosituin hakutaktiikka silloin, jos hakutulos oli liian laaja ja hakutermin vaihtaminen silloin, jos alkuperäisellä hakulausekkeella löytyisi liian vähän dokumentteja. Hsieh-Yeen mukaan jotkut hakijoiden käyttämistä hakutaktiikoista olivat vastaavia, joita voi käyttää myös haettaessa onlineluetteloista, mutta taas toiset olivat ominaisia nimenomaan web-haulle. Tällaisiksi web-haulle ominaisiksi taktiikoiksi Hsieh-Yee mainitsi mm. oman isäntäorganisaation kotisivun tarkistamisen ja sanahakujärjestelmän vaihtamisen.

Wang ja Tenopir (1998, ks. myös Wang et al. 1998) tutkivat web-hakijoiden hakukäyttäytymistä ja heidän tiedonhakuun liittyviä tuntemuksiaan. 24 hakijaa pyydettiin hakemaan webistä vastaus kahteen faktakysymykseen. Heitä ei neuvottu eikä autettu haun aikana. Tutkimusaineisto kerättiin lokitiedostoina, äänen ajattelumenetelmällä, kyselyllä ja psykologisilla

testeillä. Lähes jokainen hakija löysi oikean vastauksen ensimmäiseen hakutehtävään, mutta vain puolet toiseen. Kuitenkin monet väärän vastauksen löytäneistä olivat varmoja siitä, että he olivat löytäneet oikean vastauksen. Wangin ja Tenopirin mukaan tiedonhaussa käytettiin pääasiassa kahta strategiaa. Hakijat joko lähtivät seuraamaan linkkejä tuntemiltaan web-sivuilta tai he käyttivät sanahakupalveluita ja muotoilivat hakulausekkeita. Tällöin he usein myös vaihtoivat sanahakupalvelua ja muuttivat hakulausekkeita hakujen aikana.

Spink, Bateman ja Jansen (1998) tutkivat webissä Excite-sanahakupalvelulla suoritettuja hakuja kyselyn avulla. He kartoittivat taustoittavien demografisten tietojen lisäksi Excite-hakupalvelun käytön useutta ja hakijoiden käyttämiä hakutermejä. Hakujen aiheet hajosivat useille eri aihealueille joskin eniten hakuja tehtiin aiheista, jotka liittyivät ihmisiin, tietokoneisiin tai liiketoimintaan. Spink, Bateman ja Jansen havaitsivat hakijoiden suorittavan melko yksinkertaisia ja nopeita hakuja. Hakijat käyttivät yhtä hakua kohden keskimäärin vain 3,3 hakutermiä. Tekijöiden mukaan hakijoiden kyky valita hyviä hakutermejä ja hyödyntää erilaisia hakustrategioita oli vähäinen.

Kuusivaara ja Tuomela (1999) tarkastelivat relevanssikäsityksen muodostumista web-tiedonhaun aikana relevanssia määrittävien dokumenttiattribuuttien ja informaatioattribuuttien avulla. Dokumenttiattribuuttien avulla he selvittivät, mistä web-tiedonhaussa löydetystä dokumenttiviitteen tiedoista käyttäjälle muodostuu käsitys viitteen takana olevan dokumentin relevanssista. Informaatioattribuuttien avulla he tarkastelivat käyttäjän arviota löydetyn dokumentin relevanssista. Tutkimuksen aineisto kerättiin syksyllä 1998 webbiin laitetun kyselylomakkeen avulla. Kyselylomakkeessa vastaajat raportoivat AltaVista sanahakupalvelulla ja BUBL LINK aihehakemistosta suorittamiaan tiedonhakuja. Dokumenttiattribuutteja tarkasteltaessa keskeisiksi attribuuteiksi osoittautuivat dokumentin nimeke ja aiheenkuvaustermit. Niiden perusteella käyttäjät arvioivat löytämäänsä viitettä tarkastellessaan, onko dokumentti, joka viitteen takaa löytyy, heille relevantti vai ei. Informaatioattribuutteja tarkasteltaessa tärkeimmäksi relevanssitekijäksi nousi tiedon paikkansapitävyys. Muita tärkeitä relevanssitekijöitä olivat dokumentin sisältö (vastasiko dokumentin sisältö haun aiheetta), dokumentin selkeys ja tiedon kattavuus.

Tiedonhaun konteksti

Fidel ja Soergel (1983) ehdottivat, että tiedonhaukua tutkittaessa seuraavat tekijät tulisi ottaa huomioon:

- 1) tiedonhakupähtymä (prosessi),
- 2) hakutulos,
- 3) hakupyynnö
- 4) hakijat (sekä välittäjät että asiakkaat),
- 6) tietovarannot (tietokanta ja tiedonhakujärjestelmä),
- 7) tiedonhaku ympäristö.

Fidelin ja Soergelin mainitsemat tekijät ovat oleellisia myös tutkittaessa tiedonhakuja web-ympäristössä. Se, että web tiedonhaku ympäristönä poikkeaa monessa suhteessa esim. traditionaalisesta online-ympäristöstä, tuo tarkasteluun uusia piirteitä. Seuraavassa pohdimme lyhyesti, mitä Fidelin ja Soergelin mainitsemien tekijöiden huomioon ottaminen web-tiedonhaun tutkimisessa voisi tarkoittaa.

Tiedonhakupähtymä

Siitä huolimatta, että tiedonhakupähtymää kuvataan usein nimeämällä eri vaiheita ja eri järjestyksessä, se kuitenkin yleensä hahmotetaan varsin laajaksi tapahtumasarjaksi, joka alkaa asiakkaan tiedontarpeen tunnistamisesta (on syy tiedonhauille) ja päättyy hakutuloksen arviointiin tai sen hylkäämiseen (mitään relevanttia ei löytynyt). Tiedonhakupähtymän aikana otetaan erilaisia "askeleita" (steps) ja tehdään sekä positiivisia että negatiivisia valintoja, ts. päätetään edetä tai olla etenemättä tiettyyn suuntaan. Tiedonhakupähtymää voidaan analysoida ja kuvata erilaisina hakustrategioina, -taktiikoina tai tiedonhaussa tehtyinä siirtoina (Bates 1979, Harter 1986, Bates 1987, livonen 1995).

Jotkut traditionaalisessa tiedonhaku ympäristössä tunnistetut hakustrategiat ja taktiikat voivat toimia myös web-ympäristössä. Esimerkiksi pikahaku (haku yhdellä tai kahdella hakusanalla) käytetään usein haettaessa informaatiota webin tiedonhakupalveluiden avulla (vrt Spink, Bateman ja Jansen 1998) On kuitenkin oletettavaa, että kaikki traditionaalisessa online-ympäristössä tunnistetut hakustrategiat ja -taktiikat eivät toimi web-ympäristössä. Samoin on oletettavaa, että tarkasteltaessa web-ympäristössä toteutettavaa

tiedonhakua, voidaan löytää aivan uusia haku-strategioita ja taktiikoita, jotka ovat tyypillisiä nimenomaan web-tiedonhaulle (vrt. Hsieh-Yee 1998). Nämä uudenlaiset strategiat voivat olla yhä enemmän selailuun ja hypertekstissä navigointiin liittyviä strategioita. Koska web on tiedonhaku-ympäristönä nopeasti muuttuva, hajanainen ja suuri, on oletettavaa, että web-ympäristössä korostuvat tiedon valikointiin ja suodattamiseen liittyvät hakustrategiat ja -taktiikat (vrt. Green 1996). Web-tiedonhaussa noudatettavat hakustrategiat ja -taktiikat voidaan tunnistaa vain seuraamalla ja havainnoimalla ihmisten hakukäyttäytymistä.

Hakutulokset

Hakutulokset voidaan ymmärtää vastauksena tiedonhakuun. Hakutulokset on kuvattu aiemmissa tutkimuskirjallisuudessa viittaamalla mm. hakutulokseen kokoon, saantiin, tarkkuuteen, uutuusarvoon, hyödyllisyyteen ja kustannuksiin. Huomion kiinnittäminen näihin tekijöihin on tärkeää myös tutkittaessa web-ympäristössä tapahtuvaa tiedonhakua. Jotkut näistä tekijöistä voivat kuitenkin olla melko vaikeasti operationalisoitavissa. Esimerkiksi hakutulosten aiheuttamia todellisia kustannuksia voi olla web-ympäristössä vaikea selvittää ns. näkymättömien kustannusten vuoksi. Voidaan myös olettaa, että web-ympäristössä saavutettuja hakutuloksia tarkasteltaessa hakutuloksen tietyt kriteerit nousevat muita tärkeämmiksi. Tällaisia keskeisiä kriteereitä voisivat olla mm. hakutuloksen laatu ja relevanssi. Kuten Wang ja Tenopir (1998) totesivat, hakijat arvelevat löytäneensä oikean tiedon webistä silloinkin kun he eivät sitä löydä. Web on tiedonhaku-ympäristönä niin laaja, että sieltä on aina löydettävissä jotakin. Siksi nimenomaan hakutuloksen laadun arviointi on keskeinen kysymys sekä tiedonhakijan että tiedonhakupalveluiden tuottajien kannalta.

Hakupyynnöt

Hakupyynnöt on tiedontarpeen sanallinen kuvaus ja tiedonhaun käynnistäjä. Aiemmissa tutkimuksissa hakupyynnöt on tyypitelty monien eri tekijöiden mukaan. Huomiota on kiinnitetty mm. hakupyynnön alaan, niihin liittyviin ennako-oletuksiin ja rajauksiin sekä hakupyynnön spesifisyyteen, kompleksisuuteen ja selkeyteen (ks. tarkemmin Iivonen 1995). Aiempien

tutkimusten (Fidel 1985, Fidel 1987, Iivonen 1995) perusteella jo tiedetään, että hakijat käyttäytyvät eri-lailla käsitellessään erityyppisiä hakupyynnöitä ja että joihinkin hakupyynnöihin on helpompi löytää vastauksia kuin toisiin. Tämän voi olettaa pitävän paikkansa myös web-tiedonhaussa. Lisäksi voidaan olettaa, että web-tiedovarantona soveltuu paremmin tietäntyyppisten hakupyynnöiden käsittelyyn kun taas toisia hakupyynnöitä ei webistä kannata edes hakea. Vastaukset niihin joko löytyvät helpommin muualta tai jäävät kokonaan löytymättä.

Hakijat

Hakija on toimija, joka yrittää löytää hakupyynnön vastaukseksi tarvittavan informaation. Hän voi tehdä hakua itselleen (käyttäjä) tai toiselle henkilölle (välittäjä). Aiemmissä tutkimuksissa on kiinnitetty huomiota hakijoiden erilaisuuteen ja tarkasteltu mm. hakijoilla olevaa kokemusta, koulutusta, hakupyynnön aihealueen tuntemusta, persoonallisuutta ja hakutyyliä. Aiemmissä tutkimuksissa on huomattu, että erilaiset hakijat käyttäytyvät eri tavoin ja käyttävät erilaisia haku-strategioita. (Ks. tarkemmin Iivonen 1995). Lisäksi on huomattu, että hakijat ovat tavoilleen melko uskollisia ja siirtävät oppimansa ja hyväksi kokemansa strategiat myös uuteen ympäristöön ja uuteen tilanteeseen.

Web-tiedonhakua tarkasteltaessa on huomattava, että webin avoimuudesta johtuen sen käyttäjien joukossa on aina enemmän ns. tavallisia käyttäjiä (itselleen hakua tekeviä henkilöitä) kuin välittäjiä. Web on siis korostetusti ei-ammattilaisten tiedonhaku-ympäristö. Lisäksi webin käyttäjien määrän nopea kasvu merkitsee sitä, että maailmanlaajuisesti ottaen web-tiedonhakijoiden joukossa on aina enemmän aloittaneita kuin kokeneita hakijoita. Vaikka joidenkin hakijoiden kokemus web-tiedonhausta koko ajan kasvaa, webin käyttäjien joukkoon tulee koko ajan lisää uusia käyttäjiä, joilta tämä kokemus puuttuu vielä kokonaan. Edellämainitut kaksi piirrettä, käyttäjien ei-ammattilaisuus ja noviisimaisuus merkinnevät sitä, että webissä suoritetaan aina melko yksinkertaisia hakuja.

Tietovarannot

Tietovarannot ovat informaatiokokoelmia ja dokumentteja, jotka ovat tiedonhaun kohteena pyrittäessä tyydyttämään ihmisten tiedontarpeita ja etsittäessä vastauksia erilaisiin hakupyynnöihin. Tietoliikenne-

yhteyksien avulla käytettävien tietovarantojen käyttö mahdollistui jo 1960-luvulla, kun ensimmäiset bibliografiset viitetietokannat tulivat julkiseen käyttöön. Ensimmäiset julkiset lähdetietokannat tulivat käyttöön 1970-luvulla (Williams 1977).

Internet ja erityisesti web ovat lisänneet tavoitettavien tietovarantojen määrää huomattavasti. Internetin kautta tavoitettavan aineiston määrää ja luonnetta on lähes mahdoton arvioida. Web-tekniikan käyttöönotto on myös helpottanut tietoresurssien käyttöä, koska se tarjoaa yleisesti käyttöönotetun teknisen ratkaisun niin tietokantojen kuin elektronisten julkaisujen ja asiointipalvelujenkin käyttöön Internet-verkossa. Vaikka tietoresurssien tekninen käyttö onkin helpottunut, on sisällöllinen saavutettavuus kuitenkin selkeästi jakanut. Webin kautta on tavoitettavana perinteisiä, laadukkaita ja maksullisia tietovarantoja (tietokantoja, elektronisia lehtiä jne) ja toisaalta yleisesti, vapaasti saavutettavia palveluita, joiden laadusta ja löytyvyydestä ei ole mitään takeita. Verkon käyttäjien tulisi kiinnittää huomiota verkosta löytyvän aineiston arviointiin. Julkisen verkon puolella olevissa palveluissa on merkittäviä ja hyödyllisiä tietoresursseja, mutta niiden löytäminen ja arviointi voi olla aikaavievä prosessi.

Aiempi tutkimus (ks. esim. Chatman 1992) osoittaa, että henkilö- ja muut informaalit lähteet ovat tiedonhankinnassa keskeisiä ja usein keskeisempiä kuin formaalit tiedonhankintakanavat. Internet tarjoaa tämänkaltaiseen tiedonhankintaan oivallisen mahdollisuuden. Sähköposti, postituslistat, keskusteluryhmät, irc ja chat-kanavat mahdollistavat globaalin informaalisen tiedonhankinnan. Myös web-tiedonhakua tarkasteltaessa em. viestintäpalvelut on otettava tutkimuksessa huomioon, koska postituslistojen ja keskusteluryhmien arkistoja julkaistaan tietokantoina webissä.

Web-tiedonhakuun voidaan käyttää erilaisia verkon tarjoamia tiedonhakupalveluita. Palvelut voidaan karkeasti jakaa kahteen ryhmään. Ensinnäkin on mahdollista käyttää erilaisia aihehakemistoja. Toiseksi käytävissä on joukko sanahakupalveluita (hakukoneita, hakurobotteja). Aihehakemistot ovat pääpiirteiltään samankaltaisia painettujen hakemistojen ja perinteisen kirjastoluetteloiden kanssa. Internet-resurssit on luokiteltu ja järjestetty aiheenmukaisiin ryhmiin. Haku tapahtuu hakemistoja selaamalla tai sanahakuna hakemistoon tallennettujen resurssien kuvauksista. Sanahakupalveluista aihehakemistot erottuvat juuri aineiston valinnan, arvioinnin ja järjestämisen kautta.

Hakemistoja tuottavat joko yksittäiset asiantuntijat tai toimituskunnat. Inhimillinen valinta ja arviointi takaavat myös kuvailtavien resurssien suhteellisen laadukkuuden. Esimerkkeinä yleisistä aihehakemistoista voidaan mainita eBlast : Encyclopaedia Britannica's Internet Guide, WWW Virtual Library, Looksmart, Infomine sekä Yahoo.

Sanahakupalveluiden tietokannat ovat laajoja kokoluokkia Internet-resursseista. Nämä palvelut koostavat tietokantansa web-sivuilla esiintyvistä merkkijonoista keräilyohjelmien (robots, harvesting programs, spiders) avulla. Keräilyohjelmat seuraavat web sivujen linkkirakenteita ja poimivat kohtaamiltan sivuilta lähes kaikki merkkijonot tietokantoihinsa. Vaikka nämä palvelut kattavat verkkoresursseista huomattavasti enemmän kuin aihehakemistot ovat sanahakupalveluiden tietokannat kuitenkin hyvin rajallisia. On arvioitu, että suurimmat hakupalvelut kattavat vain alle kolmanneksen internetin web-dokumenteista. Tietokantajulkaiseminen, jossa web-sivut luodaan ohjelmallisesti tietokannasta käyttäjän nähtäväksi, sekä muiden kuin html-muotoisten tiedostojen lisääntyvä käyttö vaikeuttaa sanahakupalveluiden käyttöä tiedonhaussa, koska näin tuotettuja dokumentteja ei ole sanahakupalveluiden avulla lainkaan löydettävissä. Sanahakupalveluiden käyttöä voidaan verrata avainsanahakuun perinteisissä tietokannoissa. Toisaalta avainsanahaku on käyttäjälle helppoa, mutta toisaalta tietokantojen sisältäessä web-sivujen kokotekstit yksittäisten hakutermien yhdistely tuottaa aivan liian suuria tulosjoukkoja. Sama haku eri sanahakupalveluissa tuottaa myös erilaisia tuloksia johtuen hakujärjestelmien erilaisesta katteesta. Tunnetuimpia sanahakupalveluita ovat Alta Vista, InfoSeek, Lycos, Excite, Northern Light Search ja HotBot.

Tiedonhakuympäristö

Tiedonhakuympäristöön kuuluvat sekä organisaatioympäristö (organisaation tehtävät, yleishyödyllisyys - kaupallisuus, aihealue ja tavoitteet), käytävissä oleva teknologia että asiakkaat, mikäli tiedonhakijat toimivat välittäjinä. Aikaisempi tutkimus osoittaa, että tiedonhakuympäristö vaikuttaa hakuprosessiin (ks. esim. Wanger, McDonald and Berger 1980, Harter 1984, Fidel 1991a, Fidel 1991b, Iivonen 1995).

Internet oli pitkään akateemisten tutkimus- ja opetusorganisaatioiden tietoverkko, mutta se levisi webin

kehityksen myötä nopeasti julkisen hallinnon ja sittemmin kaupallisten yritysten ja yksityisten kansalaisten käyttämäksi informaatiopalveluksi. Kun web-tiedonhakua tutkitaan osana organisaatioiden toimintaa, täyttyy kiinnittää huomiota tiedonhaku-ympäristöön ja sen mahdolliseen vaikutukseen hakuprosessiin. Merkille pantavaa on, että suuri osa tiedonhausta webissä tapahtuu täysin yksityisesti. Web on tuonut tiedonhaun kysymykset laajemman yleisön tietoon - tiedonhaku siirtyy yhä selkeämmin ammattimaisten välittäjien osaamisalueelta itsenäiskäyttöön. Kiinnostavaa on juuri se kuinka kansalaiset pystyvät hyödyntämään webin tarjoamia tiedonhakumahdollisuuksia elämässään. Koska Internet on niin heterogeeninen, nopeasti laajeneva, vapaasti leviävä media, jonka organisaatiot voivat ottaa käyttöön ajasta, paikasta ja organisaation koosta riippumatta, avaa se organisaatioille aivan uudenlaisia toimintamahdollisuuksia. Myös organisaatioiden toimintaan ja mahdollisuuksiin vaikuttavat muutokset tulisi ottaa huomioon tutkittaessa tiedonhakua webissä.

Organisaatioiden ja yksilöiden näkökulmasta webin tiedonhakupalvelut alkavat muistuttaa yhä enemmän joukkoviestintävälineitä. Tiedonhakupalvelut ovat mahdollisesti tulevaisuudessa erityyppisiä viestimiä ja palveluja integroivia ratkaisuja (esim. sanomalehti, radio ja televisio yhdistettynä tiedonhaku-, asiointi- ja ostospalveluihin). Monet webin tiedonhakupalvelut pyrkivät internetin etusivuksi, lähtöpisteeksi, täyden palvelun mediaaloiaksi, jotka tarjoavat niin uutisia, käyttötietoa, asiointi- ja ostosmahdollisuuksia kuin tiedonhakumahdollisuuksiakin. Vastaava kehitys on käynnissä myös joukkoviestintäpalveluiden ja vaikkapa verkkokaupan tarjoamissa palveluissa, joihin integroidaan puolestaan tiedonhakupalveluita. Näitä verkon palvelukeskuksia, etusivuja on alettu kutsua portaaleiksi (portals).

Lopuksi

Tiedonhaun laajeneminen web-ympäristöön pakottaa hahmottaan myös tiedonhakua uudella tavalla. Ideatasolla webin kautta on mahdollisuus saada laajat tietovarannot käyttöön nopeasti ja vaivattomasti. Käyttäjien ei tarvitse enää siirtyä fyysisesti paikasta toiseen saadakseen haluamansa tiedon. Käytännössä tarvittavan tiedon löytäminen webistä on kuitenkin monesti vaikeaa ja oikean tiedon perässä "surffailu" turhauttavaa, jopa ahdistavaa. Jotta web todella palve-

lisi tiedonhakua, webin hakupalvelujen kehittäminen vaatii vielä paljon työtä.

Washington Post uutisoi 24.1.1999 näkyvästi Direct Hit Technologies -yrityksen tuottaman uuden hakupalvelun www.directhit.com menestyksestä. Samalla lehti kertoi yrityksen liikeidean taustalla olevasta yrityksen perustajan Gary Cullissin epäonnistuneesta web-tiedonhausta. Culliss oli etsinyt hakusanalla Armagedon dokumentteja samannimisestä elokuvasta ja saanut tulokseksi tuhansia maailmanlopusta kertovia, itselleen täysin epärelevanttejä dokumentteja. Tiedonhaun asiantuntijan näkökulmasta Cullissin valitsemaa hakustrategiaa voidaan kritisoida, mutta se ei muuta sitä tosiasiaa, että Cullissin hakutulokset olivat hänen kannaltaan roskaa. Cullissin käyttämä hakustrategia oli myös Cullissin kannalta oikea, koska se oli hänen valitsemansa hakustrategia. Webin hakupalveluiden kehittämiseksi onkin tiedettävä, miten käyttäjät tosiasiasa webiä tiedonhakuun käyttävät. Tätä tietoa saadaan vain ottamalla selvää ihmisten tavasta toimia, ts. tutkimalla empiirisesti tiedonhakua web-ympäristössä.

Webin kautta tapahtuvaan tiedonhakuun tuo oman haasteellisuutensa myös se, että web toimii myös kanavana moniin sellaisiin tietovarantoihin, mm. isoihin tietokantoihin, joihin aiemmin on ollut pääsyä muuta kautta. Hyvän esimerkin tästä tarjoavat Kansallinen elektronisen kirjasto -hankkeen (FinELib - Finnish Electronic Library) kautta tavoitettavat palvelut kuten laajat elektronisten lehtien kokoelmat sekä kymmeniä viite- ja kokotekstietietokantoja tarjoavat hakupalvelut. Mahdollisuuksien määrä on jo nyt niin suuri, että niiden hallitseminen on lähes ylivoimainen tehtävä. Webin kautta tavoitettavien tietovarantojen hyödyntäminen vaatii näiden tietovarantojen jäsentämistä sekä käsitteellisesti että empiirisesti. Tämä jäsentäminen kuuluu eittämättä informaatiotutkimuksen tehtäviin. Mutta se kuuluu myös niille, jotka opettavat webin käyttöä muille. Esimerkiksi yliopistokirjastoissa annettavassa käyttäjäkoulutuksessa joudutaan jatkuvasti miettimään, miten webin käyttöä voidaan jäsentää parhaalla mahdollisella tavalla.

Kulthau (1991) on kuvannut tiedonhakua prosessiksi, joka etenee tunnetasolla epävarmuudesta aluksi optimismiin, sitten hämmennyksen, turhautumisen ja epäilyn kautta selkeyden ja varmuuden löytämiseen päätyen lopulta joko tyydytyksen ja helpotuksen tai vaihtoehdoisesti epäonnistumisen kokemiseen. Optimistisesti voisi uskoa, ettei web-tiedonhaunkaan tarvitse pysähtyä turhautumiseen ja ahdistamiseen webin kaoks-

sa vaan tiedonhaku webissäkin voi päättyä onnistuneesti. Jotta näin kävisi, web-tiedonhakupalveluita on kehitettävä vastaamaan paremmin ihmisten hakukäyttäytymistä. Tämä taas on mahdollista vain tuntemalla ihmisten tapa hakea tietoa netistä. Se taas edellyttää web-tiedonhaun tutkimista.

Hyväksytty julkaistavaksi 18.5.1999.

Lähteet

- Bates, Marcia (1979). Information search tactics. *Journal of the American Society for Information Science* 30 (4): 205-214.
- Bates, Marcia (1987). How to use information search tactics online. Online (May): 47-54.
- Burton, Paul F. (1994). Electronic mail as an academic discussion forum. *Journal of Documentation* 50 (2): 99-110.
- Carley, Kathleen, Wendt, Kira (1991). Electronic mail and scientific communication. *Knowledge: Creation, Diffusion, Utilization* 12 (4): 406-440.
- Chatman, Elfreda (1992). The information world of retired women. Westport, CT: Greenwood Press.
- Comer, Douglas E. (1994). The Internet book: everything you need to know about computer networking and how the Internet works. Englewood Cliffs, NJ, Prentice Hall.
- Covi, Lisa, Kling, Rob (1996). Organizational dimensions of effective library use: Closed rational and open natural systems models. *Journal of the American Society of Information Science* 47 (9): 672-689.
- Dempsey, Lorcan, Heere, Rachel: Metadata: A current view of practice and issues. *Journal of Documentation* 54 (2): 142-172, 1998
- Eager, William ...et al. (1995). Net. Search: quickly find anything you need on the internet. Indianapolis: Que.
- Fidel, Raya (1991a). Searchers' selection of search keys: II. Controlled vocabulary or free-text searching. *Journal of the American Society for Information Science* 42 (7): 501-514.
- Fidel, Raya (1991b). Searchers' selection of search keys: III. Searching styles. *Journal of the American Society for Information Science* 42 (7): 515-527.
- Fidel, Raya, Soergel, Dagobert (1983). Factors affecting online bibliographic retrieval: A conceptual framework for research. *Journal of the American Society for Information Science* 34 (3): 163-180.
- Fidel, Raya (1985). Individual variability in online searching behavior. in: *ASIS '85: Proceedings of the American Society for Information Science 48th annual meeting*. Vol. 22. Ed. by Carol A. Parkhurst, p. 69-72. White Plains, NY: Knowledge Industry Publications.
- Fidel, Raya (1987). What is missing in research about online searching behavior. *Canadian Journal of Information Science* 12 (3/4): 54-61.
- Green, Rebecca (1996). Description in the Electronic Environment. In: *Knowledge Organization and Change*. Ed. by Rebecca Green, p. 10-25. Indeks Verlag.
- Grimes, Galen, Kinkoph, Sherry (1997). 10 minute guide to Netscape Communicator 4. Indianapolis, Que, Macmillan.
- Harter, Stephen P. (1984). Online searching styles: An exploratory study. *College and Research Libraries* 45 (4): 249-258.
- Harter, Stephen P. (1986). Online information retrieval. Orlando: Academic Press.
- Hsieh-Yee, Ingrid (1998). Search tactics of Web users in searching for texts, graphics, known items and subjects: A search simulation study. *Reference Librarian* 60: 61-85.
- Hjerppe, Ronald (1996). Go with the flow, or abide by the side, or watch the waves? Challenges of Change for Knowledge Organization. In: *Knowledge Organization and Change*. Ed. by Rebecca Green, p. 242-248. Indeks Verlag.
- Iivonen, Mirja (1995). Hakulausekkeiden muotoilun yhdenmukaisuus onlineviitehaussa. *Acta Universitatis Tamperensis*. Ser. A. Vol. 443. University of Tampere, Tampere, 1995.
- Irvine, Martin (1997). Web works. New York: W.W. Norton & Company.
- Kaminer, Noam (1997). Scholars and the use of the Internet. *Library & Information Science Research* 19 (4): 329-345.
- Keskikiikonen, Mika (1997). Opi yhdellä silmäyksellä Internet: Visuaalinen opas. Espoo : Suomen atk-kustannus.
- Kriesel, Ronald W. (1998). Suggested internet research strategies [online], päivitetty 4.6.1998 [viitattu 10.3.1999]. Saatavilla www.concentric.net/~Rkriesel/Search/Strategies.shtml
- Korpimies, Kai (1995). WWW-navigointi. Espoo: Suomen atk-kustannus.
- Kuhlthau, Carol C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*. 42 (5): 361-371.
- Kuusivaara, Tuomas ja Tuomela, Sanna (1999). Relevanssikäsityksen muodostuminen webin tiedonhaussa. Pro gradu-tutkielma. Oulun yliopisto. Informaatiotutkimuksen laitos.
- Kuvaja, Arto (1997). WWW-käyttäjän opas. Espoo : Suomen atk-kustannus.
- Merritt, J. Greg, Schechter, Michael G. (1997). When navigating the world, don't just give them a few wonderful sites: Reflections on teaching, learning, the World Wide Web and the international studies curriculum. *International Studies Notes* 2 (2): 23-28.
- Pfaffenberger, Bryan (1996). Web search strategies. NY: Mis Press.

- Perry, Clifford (1995). Travelers on the Internet: A survey of Internet users. *Online* 19 (2): 29-34.
- Savolainen, Reijo (1998). Tietoverkot kansalaisten käytössä: Internet ja suomalaisen tietoyhteiskunnan arki. Tampere : Tampereen yliopisto, informaatiotutkimuksen laitos.
- Senkevitch, Judith J., Wolfram, Dietmar (1997). Internetworking an urban community: A longitudinal study of approaches to including adult new users to electronic information resources. *Library & Information Science Research* 19 (3): 249-264.
- Sproull, Lee, Kiesler, Sara, Siegel, Jane (1986). Reducing social context cues: Electronic mail in organizational communication. *Management Science* 32 (11): 1492-1512.
- Tillotson, Joy, Cherry, Joan and Clinton, Marchall (1995). Internet use through the University of Toronto Library: Demographics, destinations, and users' reactions. *Information Technology and Libraries* 1, September: 190-198.
- Wang, Peiling, Tenopir, Carol (1998). An exploratory study of users' interaction with World Wide Web resources: Information skills, cognitive styles, affective states, and searching behavior. Contributed paper to Annual National Online Meeting 1998.
- Wang, Peiling, Tenopir, Carol, Layman, Elizabeth, Penniman, David and Collins, Shawn (1998). An exploratory study of user searching of the World Wide Web: A holistic approach. Contributed paper to ASIS Annual Meeting 1998.
- Wanger, Judith, McDonald, Dennis, Berger, Mary C. (1980). Evaluation of the online search process. Santa Monica CA.: Caudra Associates.
- Williams, Martha (1977). Data bases: A history of development and trends. *Journal of American Society of Information Science* 28 (2): 71-78.
- Wiggins, Richard W. (1994). *The Internet for everyone: a guide for users and providers*. New York: McGraw-Hill.