

Fire@uta.fi: Tiedonhaun tutkimusta Tampereen yliopistossa – osa 2

Kalervo Järvelin (toim.)

*(<http://www.info.uta.fi/research/fire.html>,
<http://www.info.uta.fi/tutkimus/labra.html>)*

Kalervo Järvelin (ed.), *FIRE@uta.fi* Information retrieval research at the University of Tampere – Part 2. Informaatiotutkimus 19(2): 51-57.

FIRE (the Finnish Information Retrieval Expert Group) is the research group on Information Retrieval active at the Department of Information Studies, University of Tampere, Finland. The first part of this article described the IR Laboratory of the department and the FIRE research activities in text retrieval. In the second part, multimedia retrieval, IR learning environments and IR in the information seeking context will be presented.

Editor's address: Department of Information Studies, University of Tampere, FIN-33014 TAMPERE, Finland. E-mail: kalervo.jarvelin@uta.fi

1. Johdanto

FIRE (Finnish Information Retrieval Expert Group) on Tampereen yliopiston Informaatiotutkimuksen laitoksella toimiva tiedonhaun tutkimusryhmä. Tutkimusryhmä edistää tiedonhaun tutkimusta Suomessa yhdessä ja jäsentensä kautta toteuttamalla tutkimusprojekteja, tuottamalla tutkimusjulkaisuja, ohjaamalla opinnäytteitä maisteri- ja tohtoritasolla, osallistumalla kansainväliseen tutkimusyhteistyöhön ja konferenssitoimintaan, antamalla opetusta ja pitämällä säännöllistä seminaaria. Koska FIRE toimii nimenomaan Tampereen yliopistossa, pääotsikko FIRE@uta.fi kertoo lyhyesti, mistä tässä artikkelissa on kyse: artikkelissa esitellään FIREn piirissä tapahtuvaa tutkimusta (sähköpostiosoitteena se ei kuitenkaan toimi). Ylläolevista verkko-osoitteista löytyy myöhemminkin ajankohtaista tietoa FIREn toiminnasta.

Tiedonhaun tutkimuksen ydinryhmä, tämän artikkelin kirjoittajat, koostuu seuraavista, pääasiassa tiedonhaun tutkimuslaboratorion yhteydessä toimivista tutkijoista: Riitta Alkula, Hannele Fabritius, Kai Halttunen, Turid Hedlund, Kalervo Järvelin, Jaana Kekäläinen, Heikki Keskustalo, Marjo Markkula, Ari Pirkola ja Eero Sormunen sekä Pertti Vakkari. He eivät kuitenkaan edusta kaikkea Tampereella

tehtävää tiedonhaun tutkimusta, sillä FIREn toimintaan osallistuu muitakin tutkijoita, joista lähemmin loppuluvussa.

Tämä katsaus jakautuu kahteen osaan: ensimmäisessä esittelimme Tiedonhaun laboratorion ja FIREn tekstitiedonhaun tutkimushankkeita; toisessa osassa esittelemme multimediatiedonhaun tutkimusta, tiedonhaun oppimisympäristöjen tutkimusta sekä tiedonhaun tutkimusta osana tiedonhankintaa. Artikkelin tutkimusesittelyt ovat tutkijoiden itsensä kirjoittamia. Kunkin osuuden kirjoittaja näkyy esittelyn yhteydessä.

2. Kuva- ja videotiedonhaku

Marjo Markkula (limamar@uta.fi) ja Eero Sormunen (lieeso@uta.fi)

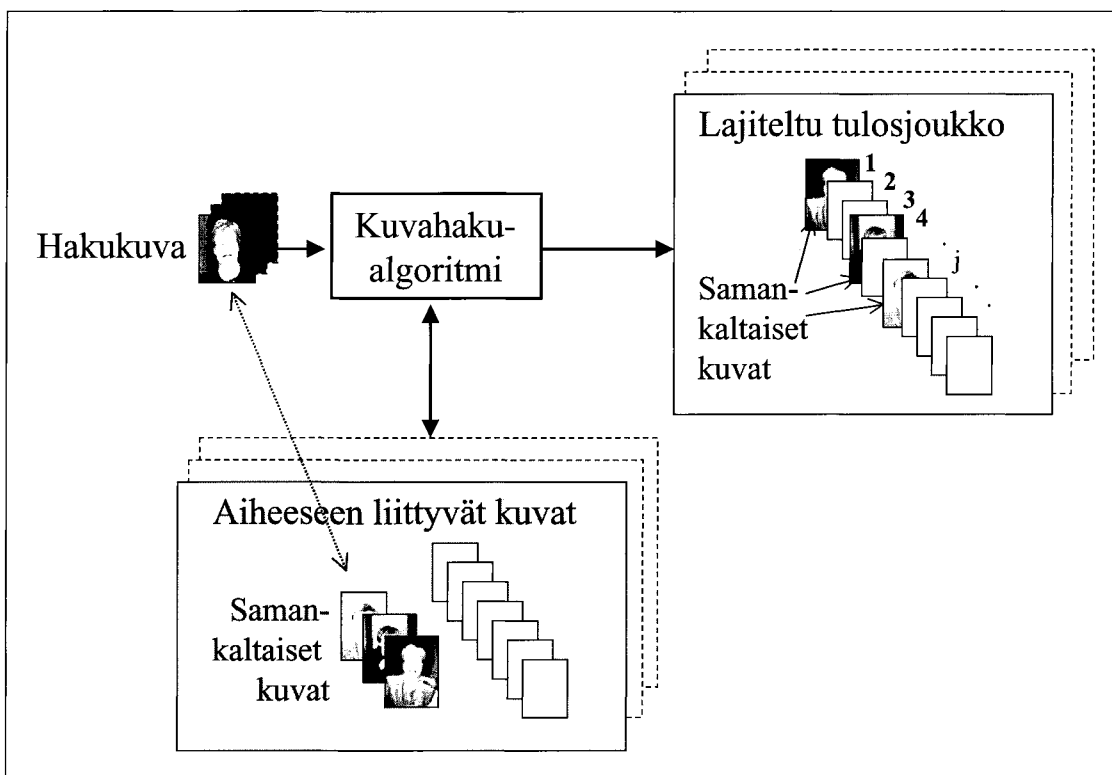
Digitaalisen median osahankkeessa on yhdistetty teknisen kehittämisen ja käyttäjätehtävien tutkimuksen näkökulmat tavoitteena kehittää uudenlaisia ratkaisuja digitaalisten kuva- ja videoarkistojen käyttäjien kokemuksiin hakuongelmiin. Projektin ensimmäisessä vaiheessa keskityttiin valokuvien hakumenetelmien kehittämiseen ja

evaluointiin. Projektissa tarkasteltiin yhtä kuvatie-
dönhakuympäristöä: sanomalehden valokuva-arkistoa.
Tutkimusta tullaan laajentamaan videotiedönhakuun ja
television uutistoimitukseen videoaineiston käyttö-
ympäristönä.

Kuvatiedonhaku jakautuu kahteen periaatteiltaan
erilaiseen menetelmään: kuviin liitettyihin sanallisiin
kuvauksiin perustuviin *tekstihakuihin* ja kuvan
automaattiseen visuaaliseen tulkintaan perustuviin
hahmohakuihin. Digitaalisten kuva-arkistojen pää-
asiallinen organisointitapa on tällä hetkellä sanalliseen
kuvailuun perustuva. Hahmopohjaiseen hakuun on
kehitetty automaattisia, digitaaliseen kuvankäsittelyyn
perustuvia menetelmiä. Hahmopohjainen kysely
palauttaa tietokannasta kuvia visuaalisen saman-
kaltaisuuden mukaan järjestettyinä. Kuvien visuaalisen
samankaltaisuuden laskeminen pohjautuu piirre-
vektoreihin, jotka mittaaavat mm. kuvan sisältämiä
värejä ja niiden sijoittumista kuvassa. Hahmopohjaisten
kuvahakumenetelmien käyttökelpoisuudesta operatio-
naalisissa käyttöympäristöissä tiedetään vielä vähän.
Menetelmien evaluointi on lähes koskematon alue
tiedonhaun tutkimuksessa.

Empiirinen kuvien hakua ja valintaa koskeva tutkimus
tehtiin Aamulehden toimituksessa (Markkula &
Sormunen, 1998; 2000). Hahmopohjaisten mene-
telmien osalta käyttäjätutkimuksen tulokset osoittivat,
että tekstiin perustuva kuvailu ja haku säilynee
sanomalehden kuva-arkistossa kuvahaun perus-
menetelmänä ainakin lähitulevaisuudessa. Hahmo-
pohjaisten menetelmien sovellusmahdollisuudet löytyvät
lähinnä hakuprosessin selausvaiheesta. Toimittajat
suosivat selaamista, mutta kokivat sen usein työlääksi
ja aikaavieväksi. Selattavana oli satoja, joskus jopa
tuhansia sanallisella kyselyllä saatuja sormenpääkuvia.
Automaattiset hahmopohjaiset menetelmät voisivatkin
tukea käyttäjää kuvien selaamisvaiheessa mahdol-
listamalla hakutuloksen järjestämisen. Kun yksi
halutunlainen kuva on löytynyt, muut tuloksessa olevat
visuaalisesti samankaltaiset kuvat voitaisiin ryhmittää
kuvajoukon kärkeen.

Käyttäjätutkimuksen tuloksiin tukeutuen kehitettiin
testausmenetelmä hahmopohjaisten kuvahaku-
algoritmien käyttökelpoisuuden arviointiin ja vertailuun
(Markkula, Sormunen & Järvelin, 1999). Evaluointimalli



Kuva 1. Testiympäristö hahmopohjaisten kuvahakualgoritmien evaluointiin

hyödyntää testikokoelman ja laboratoriotestauksen etuja perustuen kuitenkin todellisen käyttäjäyhteisön työtehtäviin (sanomalehtiartikkelin kuvittaminen), käyttäjiin (toimittajat) ja kuvatiedonhaku ympäristöön (sanomalehden kuva-arkisto).

Testikokoelma perustuu 10 toimittajan tekemiin kuvien samankaltaisuusarvioihin simuloitujen kuvitus-tehtävien yhteydessä. Kokoelma sisältää yhteensä 41 testitehtävää. Jokainen testitehtävä koostuu (1) kyselykuvasta (=toimittajan artikkelin kuvitukseksi valitsema kuva), (2) toimittajan tämän kanssa samankaltaisiksi arvioimista kuvista sekä (3) aiheeseen liittyvästä kuvajoukosta, joka saatiin sanallisella kyselyllä ja johon myös kyselykuva ja samankaltaisiksi valitut kuvat kuuluvat.

Uuden testikokoelman etuina voidaan mainita mm. seuraavat piirteet:

1. Testikokoelma on yleiskäyttöinen hahmopohjaisten kuvahakualgoritmien testausympäristö. Minkä tahansa kuvien visuaaliseen samankaltaisuuteen perustuvan algoritmin käyttökelpoisuutta voidaan arvioida ja verrata toisen algoritmin tai algoritmiversion käyttökelpoisuuteen.

2. Vertailussa käytetään mittarina uskottavia, todellisten käyttäjien samankaltaisuusarvioita ja lähteenä on ollut operationaalinen kuva-aineisto.

3. Testikokoelma on kompakti kokonaisuus. Vaikka aineisto simuloi tilannetta suurehkoissa kuvatietokannassa, testitehtävää kohden tarvitaan vain 100-200 kuvaa.

3. Tiedonhaun oppimisympäristöt

3.1. Tiedonhaun oppimisympäristön kehittäminen ja arviointi

Kai Halttunen (likaha@uta.fi)

Kai Halttunen tarkastelee tutkimuksessaan tiedonhaun oppimista sekä oppimisympäristöjen kehittämistä ja arviointia. Tietoverkkojen, erityisesti Internetin leviäminen tutkimusyhteisöistä kansalaisten jokapäiväiseksi informaatioympäristöksi - uudeksi joukkoviestintäkanavaksi, on tuonut tiedonhaun, tiedon tallennuksen ja tiedon hankinnan kysymykset yhä suuremman mielenkiinnon kohteiksi.

Tiedonhaun oppimisympäristöjen tutkiminen ja kehittäminen on tärkeää, koska yhä suurempi osa informaatioaineistosta tulee olemaan elektronisessa muodossa käyttäjien saatavilla. Niin joukkoviestintä-

yrityksillä kuin kansallisen kulttuurin säilyttäjäillä ja käyttöönsaattajilla on suuria aineistojen digitointiin ja verkkokäyttöön liittyviä hankkeita. Tämänkaltaisten elektronisten tiedonhaku ympäristöjen julkistus vaatii myös tiedontarvitsijoiden uudenlaista perehdyttämistä tiedonhaun maailmaan. Luonteva keino lisätä kansalaisten tietämystä tiedonhaun menetelmistä on verkkopohjaisten tiedonhaun oppimisympäristöjen rakentaminen sekä niiden tutkiminen ja evaluointi.

Digitalisoituvan ja verkottuvan informaatioympäristön tehokasta hyödyntämistä tai keinoja sekä välineitä, joiden avulla kansalaisten tiedonhakutaidot hajautetuista verkkotiedonlähteistä paranisivat, ei ole kuitenkaan vielä lainkaan tutkittu. Tutkimuksen ensimmäisessä vaiheessa evaluoidaan ja selvitetään Tampereen yliopiston informaatiotutkimuksen laitoksella kehitellyn interaktiivisen tiedonhakupelin käytettävyyttä tiedonhaun opetuksessa. Evaluointi perustuu tiedonhaun peruskurssilla kerättyihin eläytymismenetelmällä kirjoitettuihin kertomuksiin hyvistä ja huonoista oppimiskokemuksista tiedonhakupeliä käytettäessä. Aineistona on lisäksi käytetty videoita ja hakuistuntoja sekä havainnointia. Seuraavassa vaiheessa evaluointiin sisällytetään tiedonhakupeli tiedonhaun ohjaajien työvälineenä. Jatkossa tutkitaan tiedonhakupelin soveltuvuutta laajemman tiedonhaun oppimisympäristön osana sekä kehitetään konstruktivisen oppimiskäsityksen mukaista oppimisympäristöä, joka perustuu tiedonhaun osaamisen ja oppimistavoitteiden, oppimisympäristön elementtien ja niiden välisten suhteiden sekä verkkopohjaisen oppimisympäristön käyttöliittymän ja käytettävyyden analyysiin.

3.2. Vuorovaikutteinen tiedonhakupeli

Eero Sormunen (lieeso@uta.fi)

Tiedonhakupeli on Web-sovellus, joka on kehitetty Informaatiotutkimuksen laitoksella kyselyjen analysoinnin välineeksi. Sovellus tarjoaa realistisen ympäristön demonstroida kyselyjen tuloksellisuutta eri hakutilanteissa. Tiedonhaun opetuksessa tiedonhakupeliä voidaan käyttää kyselyjen onnistumiseen vaikuttavien ilmiöiden havainnollistamisvälineenä. Tiedonhakupelin varaan voidaan rakentaa myös itsenäisesti tehtäviä verkkoharjoituksia, joiden suorittaja saa sovellukselta palautetta kyselyjen onnistumisesta. Tutkimusta tiedonhakupeli palvelee ideointivälineenä, jolla voidaan helposti selvittää, miten jokin kyselyn muotoilussa sovellettu tekniikka näkyy hakutuloksissa. Peliä on käytetty myös tiedonhaun käyttäjätesteissä, joka rekisteröi kaikki hakijan suorittamat kyselyt.

Tiedonhakupelin ensimmäinen ja nyt rutiinikäytössä oleva versio 3.0 valmistui vuoden 1998 lopussa. Sovellus on toteutettu tekniikalla, jossa HTML-dokumenttia käytetään WWW-käyttöliittymänä sitomaan yhteen tiedonhakupelin päätoiminnot ja niihin liittyvät CGI-ohjelmat. Järjestelmän pääkomponentit ovat (Sormunen & al., 1998):

1. Tiedonhaun testikokoelmat
 - a) TUTK-testikokoelma (54.000 suomalaista lehtiartikkelia, 35 hakutehtävää relevanssi-arvioineen)
 - b) ALMA97/98-tietokanta (tuoreempaa lehtiartikkeliaineistoa, muutamia hakutehtäviä)
 - c) TREC-testikokoelman otos (514.000 englanninkielistä tekstidokumenttia, 350 testitehtävää relevanssiarvioineen)
 - d) Lehtikuva98-kuvatietokanta (24.000 lehtivalokuvaa, muutamia hakutehtäviä relevanssiarvioineen)
2. Tiedonhakuohjelmat
 - a) InQuery, perustuu propabilistiseen tiedonhakumalliin
 - b) TRIP, perinteinen Boolean hakujärjestelmä
3. Käännössanakirjoja (suomi, ruotsi, englanti, saksa)
4. Morfologiset analyysiohjelmat (TWOL, Kstems)
5. Sovellusohjelmat kyselyn tulosten analysointiin ja visualisointiin.

Käyttäjätunnuksen haltijat voivat käyttää sovellusta verkon yli mistä tahansa työasemalta normaalia Web-selainta käyttäen, ks. lisätietoja <http://www.info.uta.fi/~lijula/pinball3/pinball3.html>. Pelin ominaisuuksia esittelevä opastustoiminto on avoin kaikille kiinnostuneille ilman salasanaa.

Tiedonhakupeliä on käytetty tiedonhaun luentojen havainnollistamisvälineenä ja ohjattujen harjoitusten välineenä kaikilla aineopintojen tiedonhaun opintojaksoilla (Halttunen & Sormunen, 2000). Sillä on myös rakenneltu itsenäisesti suoritettavia verkkoharjoituksia lähinnä teksti-, verkko- ja kuvatieonkurssilla, joissa opiskelijat tuntevat jo tiedonhaun perusteet. Toimittajien tiedonhankintakurssilla

välinettä käytettiin myös loppukäyttäjien tiedonhakuharjoituksissa, jolloin se toimi myös informaatiotutkimuksen opiskelijoiden opetusharjoittelun välineenä (Laitinen & al., 2000).

Tiedonhakupelistä on kehitetty versio 4.0, joka on testikäytössä. Uuden version tavoitteena on laajentaa tiedonhakupelin oppimisympäristöpiirteitä kehittämällä opettajan ja oppilaan työkaluja, sekä parantaa tiedonhakupelin laajennettavuutta ja ylläpidettävyyttä. Sovellus on toteutettu Java-sovelluksena asiakas-palvelin arkkitehtuurilla ja se sisältää neljä pääkomponenttia:

1. Käyttöliittymä on joko selaimen applettina tai erillisenä sovelluksena toimiva PC-ohjelma.
2. Palvelin on Unix-sovellus, joka huolehtii mm. käskyjen välittäjänä tiedonhakupelin sisäiseen SQL-tietokantaan ja eri hakujärjestelmien ajureille sekä toimii tiedostopalvelimenä tietorakenteiden tallentamista ja hakemista varten.
3. Tietokanpalvelin vastaa pelin sisäisestä SQL-tietokannasta, joka sisältää kaikki pelin käyttöä kuvaavat tiedot, käyttäjien salasanat sekä ulkopuolisten hakuohjelmistojen tarvitsemat indeksitaulukot.
4. Hakukanta-ajurit toteuttavat hakuohjelmisto- ja tietokantakohtaisen rajapinnan, joka on standardoitu palvelimeen päin. Ne muuttavat hakijan kyselyn kohdejärjestelmän syntaksiin ja palauttavat hakutulokset palvelimelle vakiomuodossa.

Tiedonhakupelin versio 4.0 edustaa kehittyneitä toteutustekniikkaa, joka antaa uusia kehittämismahdollisuuksia, mutta ei vielä sinällään ole versiota 3.0 laajempi sovellusalaltaan. Jatkokehityksessä on erityisen kiinnostuksen kohteena ns. tutkijaversioiden kehittäminen, joka mahdollistaisi monipuolisten koeasetelmien käytön vuorovaikutteiden tiedonhaku-tutkimuksen alueella. Toinen keskeinen laajentamistarve koskee verkkotiedonhakua, jossa verkkoaineiston erityisominaisuudet tuovat uusia haasteita ja mahdollisuuksia pelikonseptin kehittämiseen. Kolmas kehittämissuunta on tiedonhakupelin tuotteistaminen, jolloin se saataisiin myös laitoksen ulkopuolisten tutkimus- ja opetusyksiköiden ulottuville.

4. Tiedonhaku osana tiedonhankintaa

4.1. Tiedonhaku ongelmanratkaisuprosessin osana

Pertti Vakkari (lipeva@uta.fi)

Tutkimuksessa tarkastellaan sitä, miten henkilön tehtävän (ongelman) jäsenyisyyden aste on yhteydessä siihen, millaista tietoa hän hakee, millaisia taktiikoita ja termejä hän käyttää sekä millaiseksi hän arvioi löytyneiden viitteiden ja niiden perusteella hankittujen dokumenttien relevanssin. Tarkemmin tutkitaan sitä, miten 11 informaatiotutkimuksen opiskelijaa hakivat tietoa ja arvioivat löytyneiden viitteiden ja dokumenttien relevanssia laatiessaan pro gradunsa tutkimussuunnitelman lukukauden kestäväen seminaarin aikana keväällä 1999.

Työn viitekehystenä käytetään Kuhlthaun (1993) tiedonhankinnan prosessimallia yhdistettynä kognitiivisen psykologian ideoihin (esim. Gavin, 1998; Patel & Ramoni, 1997) havaitsemisen ja oppimisen riippuvuudesta henkilön aikaisemmasta tietämyksestä. (Vakkari, 1999.) Kuhlthaun mallia soveltaen opiskelijoiden tutkimussuunnitelman laadinta jaettiin kolmeen vaiheeseen. Ennen fokuksen muodostamista he valitsivat työnsä aihepiirin ja tutkiskelivat sitä. Fokuksen muotoiluvaiheessa he jäsenivät työnsä fokuksen. Muotoilun jälkeen ongelma oli jäsentynyt, jolloin he keräsivät tarkempaa tietoa ja ryhtyivät kirjoittamaan työtään.

Opiskelijoita pyydettiin tekemään tiedonhaku LISA-tietokannasta työtään varten seminaarin alussa keskivaiheilla ja lopussa. Näiden aikojen otaksuttiin heijastavan Kuhlthaun mallin mukaista kolmea vaihetta. Opiskelijoita pyydettiin ajattelemaan ääneen haun aikana. Hakulokit ja ääneenajattelu rekisteröitiin. Heitä haastateltiin ennen ja jälkeen jokaisen haun. Edellisissä mitattiin heidän ajatuksiaan, toimiaan ja tunteitaan tutkimussuunnitelman kirjoittamisesta käyttäen Kuhlthaun (1993) prosessisurveylomaketta (ml. käsittekartta) sekä teemahaastattelua. Jälkimmäisellä tutkittiin erityisesti heidän tietämystään työnsä aiheesta sekä sitä, millaista tietoa he olivat hakemassa ja miten he arvelivat sen auttavan työssä. Jälkimmäisissä opiskelijat arvioivat löytämiensä viitteiden relevanssia sekä sitä, mihin tulostetussa metainformaatioissa he arvioivat perustivat. Toisella ja kolmannella kierroksella he arvioivat myös aikaisemman haun perusteella hankkiemiensa dokumenttien relevanssia. Lopuksi opiskelijoita pyydettiin arvioimaan autoiko hakuistunto heitä jäsentämään tehtävänsä paremmin. Opiskelijat pitivät myös tutkimus- ja tiedonhankintapäiväkirjaa, jotka kerättiin kolme kertaa

seminaarin aikana. Edelliseen he kirjasivat ajatuksiaan suunnitelman laadinnasta. Jälkimmäiseen kirjattiin se, millaista tietoa he olivat hakemassa, mistä löysivät haetun dokumentin, oliko dokumentti relevantti ja mitä se antoi heidän työnsä.

Tulosten mukaan opiskelijat etenivät pääpiirteissään Kuhlthaun (1993) mallin mukaisesti tutkielmaa laatiessaan. Tulokset myös viittaavat vahvasti siihen, että opiskelijain tehtävänsuorituksen vaihe ja siten tehtävän jäsenyisyyden aste, on yhteydessä siihen millaisia hakutermejä ja –taktiikoita käytetään (Pennanen & Vakkari, 2000; Vakkari, 2000a), millaiseksi viitteiden ja dokumenttien relevanssi arvioidaan ja millaisia relevanssikriteereitä sovelletaan (Hakala & Vakkari, 1999; Vakkari & Hakala, 2000) sekä minkä tyyppinen informaatio koetaan hyödylliseksi (Vakkari, 2000b). Hankkeen tulokset kertovat myös sekä opiskelijoiden tehtäväprosessin vaiheen että heidän kognitiivisen tyyliinsä vaikuttavan hakutermin ja taktiikoiden valintaan.

Tutkimushanketta jatketaan kehittämällä käsitteellisesti mallia tehtäväprosessin vaiheen ja tehtävän jäsenyisyyden sekä tiedonhaun elementtien välisestä yhteydestä sekä keräämällä kattavampia empiirisiä aineistoja mallin validioimiseksi.

4.2. Tiedon ja muun aineiston hankinta ja käyttö journalistisessä juttuprosessissa

Hannele Fabritius (lihafa@uta.fi)

Hannele Fabritiuksen väitöskirjatutkimuksen tarkoituksena on selvittää journalistisen aineiston hankintaa ja käyttöä lehtijuttujen teon prosessissa. Tutkimuksessa kiinnitetään erityistä huomiota digitaaliseen aineistoon. Digitaaliset lähteet eivät muodosta yhtenäistä ryhmää vaan ne koostuvat Internetin avulla käyttöön saatavista aineistoista: perinteisistä, julkisista tietokannoista sekä lehtitalon sisäisistä tietokannoista, joista viime mainittua nimitetään myös sähköiseksi arkistoksi.

Tutkimusaiheen juuret löytyvät sekä tiedotusopin että informaatiotutkimuksesta. Tiedotusopin tutkimusperinteessä se paikallistuu journalistiikkaan lukeutuvaan lähdekäytäntöjen tutkimukseen. Informaatiotutkimuksen näkökulmasta tutkimuksessa integroituvat tiedonhankinnan ja -haun tutkimus.

Keskeisimmät tutkimuskysymykset ovat:

- Mitä lähteitä eri tyyppisissä juttuprosesseissa käytetään?

- Miten lähdeaineistoa käytetään jutunteossa?
- Mikä merkitys ja tehtävä erityisesti digitaalisista lähteistä haetulla aineistolla on jutunteossa?
- Miten ja millä perusteella jutun lähdeaineiston tärkeyttä arvioidaan?
- Minkälaisia esteitä ja vaikeuksia juttujen lähdeaineiston hankinnassa, haussa ja käytössä on ilmennyt?

Tutkimusaineisto on kerätty Helsingin Sanomissa vuosina 1996 - 1998. Triangulaation periaatteiden mukaisesti tutkimusasetelma ja tutkimusaineiston keruu ovat olleet vuorovaikutuksessa keskenään. Varsinkin tutkimuksen alkuvaiheessa käytn etnografisia kenttätyömenetelmiä eli osallistuvaa havainnointia, tutkimuspäiväkirjan pitoa ja valokuvausta sekä keräsin toimituksissa laadittuja dokumentteja. Toimitusosastot olivat kotimaa, talous, ulkomaat, sunnuntaisivut ja tutkiva ryhmä. Yhteensä haastateltiin kolmeakymmentä yleis- tai erikoistoimittajaa. Heistä kahdeksantoista suostui täyttämään juttupäiväkirjoja, joita kertyi lähes 300.

Analyysin ytimen muodostavat toimittajien kirjoittamat jutut, niiden työprosesseja kuvaavat, toimittajien itse täyttämät juttupäiväkirjat ja näitä kumpiakin tarkentavat haastattelut.

Alustavat tulokset osoittavat, että:

- Helsingin Sanomien sisäisiä tekstikantoja eli sähköistä arkistoa käytettiin paljon, lähes puolessa kaikista jutuista. Tekstikannat kuitenkin arvioitiin vain jonkin verran tärkeiksi jutun teossa. Henkilölähteiltä saatua aineistoa käytettiin lähes 90 prosentissa jutuista, ja tällä tavoin saadun aineiston käyttö arvioitiin juttuprosesseissa erittäin tärkeäksi. Yleisin lehden sähköisen arkiston käyttötapa oli jutun tausta-aineiston löytäminen ja sen selvittäminen, milloin ja miten aihetta oli viimeksi käsitelty lehdessä. Merkittävänä käyttö-tarkoituksena olivat myös erilaiset tarkistustehtävät. Sähköiseen arkistoon lehteä on tallennettu vuodesta 1990. Sitä vanhempaa, paperimuotoista arkistoa tarvittiin harvoin.

- Internetin tarjoaman aineiston käyttö oli tutkimusaineiston keruun aikana vasta alkamassa. Vastoin odotuksia toimittajat käyttivät sanahakupalveluita tai hakemistoja verraten vähän. Niiden sijasta toimittajat suunnistivat suoraan eri organisaatioiden, yhtiöiden ja instituutioiden sivuille. Sivustojen

osoitteita he saivat tiedottajilta, eri alojen asiantuntijoilta tai kollegoilta. Internetistä haettu aineisto arvioitiin juttuprosessissa tärkeämmäksi kuin talon sisäisten kantojen sisältämä aineisto, mutta jutussa keskeisessä asemassa olevana aineistona se oli harvemmin kuin virkamiehä, asiantuntijoita tai tutkijoita haastatteleamalla juttuihin saatu aineisto.

- CD-ROM -kantoja toimittajat käyttivät erittäin vähän.

Fabritius jatkaa laadullista analyysia ja tulkintaa tarkastelemalla juttuprosesseissa esiintyvien journalististen lähdeaineistojen käyttöä, tärkeysarviointeja sekä toimittajien lähteiden käytössä kohtaamia vaikeuksia.

5. Lopuksi

FIREn tutkimustoiminta on määrältään laajaa, aihepiireiltään monipuolista ja koko lailla menestyksellistäkin, kuten lähdeluettelosta voidaan päätellä. Kymmenen vuoden kuluessa toiminta on laajentunut 4 tutkijan tekstitiedonhau tutkimuksesta ja laboratorion rakentamisesta eurooppalaisittain kohtuullisen suureksi tiedonhau tutkimusryhmäksi. Laboratorio on edelleen Pohjoismaiden ainoa tiedonhau tutkimuslaboratorio, mutta perustamishankkeita on vireillä Ruotsissa ja Norjassa. Oma kokemuksemme viittaa kuitenkin siihen, että laboratorion laitteiden, ohjelmistojen ja tietokantojen asentamisesta on vielä pitkä matka tulokselliseen tutkimustoimintaan. Onneksi kaikki tiedonhau tutkimus ei vaadi laboratoriota tuekseen. FIRE on nykyisin selvästi Pohjoismaiden suurin kansainvälisesti näkyvän tiedonhau tutkimuksen tuottaja.

FIREn toimintaan osallistuu muitakin tutkijoita, jotka eivät esiinny tämän artikkelin kirjoittajina, kuten mm. Preben Hansen, Tapio Lehtiniemi, Nils Pharo, Tuija Sonkkila, joiden työstä toivottavasti saamme kuulla myöhemmin tässä lehdessä. Preben Hansen (Kista, Sweden) tutkii tiedonhakua aitojen työtehtävien yhteydessä tavoitteenaan tiedonhakujärjestelmien käyttöliittymien kehittäminen. Nils Pharo (Oslo, Norja) tutkii verkkotiedonhakua aitojen tehtävien yhteydessä tarkoituksenaan analysoida tehtävän, hakijan, hakustrategian ja -prosessin vaikutusta verkkotiedonhakuprosessin muotoutumiseen. Tuija Sonkkila tutkii rakenteisten dokumenttien käyttöä yliopiston julkaisusarjojen julkaisukanavana. Kalervo Järvelin ja Timo Niemi tutkivat rakenteisten dokumenttien rakenteellista muokkausta, tietojen tilastollista aggregointia ja deduktiivista käsittelyä sekä näiden sovelluksia mm.

informetriikkaan (Järvelin & Niemi, 1999; Järvelin, Ingwersen & Niemi, 2000). Näiden tutkijoiden työt täydentävät ja rikastavat FIREn tutkimusprofiilia. FIREn tutkijajoukko on myös kasvamassa uusilla jatko-opiskelijoilla / tutkijoilla sitä mukaa kuin aiemmin aloittaneet väittelevät.

Viiteluettelo

Gavin, H. (1998). *The Essence of Cognitive Psychology*. London: Prentice Hall Europe.

Hakala, N. & Vakkari, P. (1999). Relevanssikriteerien muutokset ja ongelman jäsentymisen tehtävä-prosessissa. *Informaatiotutkimus* 18(2), 39-50.

Halttunen, K. & Sormunen, E. (2000). *Learning Information Retrieval through an Educational Game - Is Gaming Sufficient for Learning?* (Käsikirjoitus.)

Järvelin, K., Ingwersen, P. & Niemi, T. (2000). Informetrics through advanced data management. *Journal of Documentation* 56(2) (hyväksytty julkais-tavaksi).

Järvelin, K. & Niemi, T. (1999). Integration of complex objects and transitive relationships for information retrieval. *Information Processing & Management* 35(5), 655-678.

Kuhlthau, C. (1993). *Seeking Meaning*. Norwood, NJ: Ablex.

Laitinen, H., Sormunen, E., Aarnio, A., Halttunen, K., Hurme, P., Lehtikoinen, T., Näränen, P., Vanhanen, H. & Ylönen, O. (2000). "Tartteeks meidän tietää tuo kaikki?" - kokemuksiä opetusteknologioiden ja ammatillisten roolien soveltamisesta toimittajan tiedonhankintakurssilla. <URL: <http://www.info.uta.fi/opymp/raportti>>

Markkula, M. & Sormunen, E. (1998). Searching for photos - Journalists' practices in pictorial IR. Teoksessa: Eakins, J.P., Harper D.J. & Jose, J. (toim.), *The Challenge of Image Retrieval. Proceedings. A Workshop and Symposium on Image Retrieval*. University of Northumbria at Newcastle, Newcastle upon Tyne, 5-6.2.1998. British Computer Society (BCS), Electronic Workshops in Computing. <URL: <http://www.ewic.org.uk/ewic/workshop/view.cfm/CIR-98>>

Markkula, M. & Sormunen, E. (2000). End-User Searching Challenges Indexing Practices in the Digital Photo Archive. *Information Retrieval* 1(4), 259-285.

Patel, V. & Ramoni, M. (1997). Cognitive Models of Directional Inference in Expert Medical Reasoning. Teoksessa: Feltoovich, P. & Ford, K. & Hoffman, R. (toim.), *Expertise in Context: Human and Machine*. Menlo Park (Calif.): AAAI Press, 67-99.

Pennanen, M. & Vakkari, P. (2000). Ongelman jäsentymisen yhteys tiedonhaun muutoksiin tehtävä-prosessin aikana. *Informaatiotutkimus* 19(1), 2000.

Sormunen, E., Laaksonen, J., Keskustalo, H., Kekäläinen, J., Kemppainen, H., Laitinen, H., Pirkola, A. & Järvelin, K. (1998). The IR Game - A Tool for Rapid Query Analysis in Cross-Language IR Experiments. Teoksessa: *PRICAI '98 Workshop on Cross Language Issues in Artificial Intelligence*. Singapore, Nov 22-24, 1998, 22-32.

Sormunen, E. & Markkula, M. & Järvelin K. (1999). The perceived similarity of Photos - Seeking a Solid Basis for the Evaluation of Content-based Image Retrieval Algorithms. Teoksessa: Draper S.W., Dunlop M.D., Ruthven I. & van Rijsbergen, C.J. (toim.), *Mira 99: Evaluating interactive information retrieval*. Glasgow, UK, 14-16 April 1999. British Computer Society (BCS), Electronic Workshops in Computing. <URL: <http://www.ewic.org.uk/ewic/workshop/view.cfm/MIRA-99>>

Vakkari, P. (1999). Task Complexity, Problem Structure and Information Actions. Integrating Studies on Information Seeking and Retrieval. *Information Processing & Management* 35(6), 819-837.

Vakkari, P. (2000a). Cognition and Changes of Search Terms and Tactics during Task Performance: A Longitudinal Study. *Proceedings of the RIAO'2000 Conference* (hyväksytty julkaista-vaksi).

Vakkari, P. (2000b). *Relevance and Contributing Information Types of Searched Documents in Task Performance*. In Proceedings of the 23rd Annual International ACM Sigir Conference on the Research and Development in Information Retrieval, Athens 24-28 July, 2000 (hyväksytty julkaistavaksi).

Vakkari, P. & Hakala, N. (2000). Changes in Relevance Criteria and Problem Stages in Task Performance. *Journal of Documentation* (hyväksytty julkaistavaksi).