

Ulla Lehtoranta

Moniulotteisen relevanssiarviointimenetelmän demonstrointi
Internet-aineistoissa tiedonhakuaiheina
Bernadette Soubirous, Heidi Hautala,
Mauri Kunnas ja Madonna Ritchie

Pro gradu -tutkielma

Tampereen yliopisto

Informaatiotutkimuksen laitos

TIIVISTELMÄ
TAMPEREEN YLIOPISTO
Informaatiotieteiden tiedekunta
Informaatiotutkimuksen laitos
LEHTORANTA, ULLA:

Moniulotteinen relevanssiarvointimenetelmän demonstrointi Internet-aineistoissa tiedonhakuaiheina Bernadette Soubirous, Heidi Hautala, Mauri Kunnas ja Madonna Ritchie

Pro gradu -tutkielma, 65 s., 9 liites.
Informaatiotutkimus
Marraskuu 2006

Tutkimuksessa sovelletaan moniulotteista relevanssiarviointia menetelmänä, sen soveltuvuutta erityisesti moni-ilmeisen Internet-aineiston relevanssiarviointiin. Moniulotteiseen relevanssiarviointimenetelmään liittyvät moniportaiset aihe relevanssiarviot sekä tutkittavien aineistojen aihe riippumattomat ominaisuudet. Moniulotteisen relevanssiarviointimenetelmän soveltamisessa ollaan kiinnostuneita dokumenttien sisällöllisistä eroista ja päällekkäisyyksistä. Aihe relevanssin määrittelemiseksi aineistosta tunnistetaan ensin dokumenttien sisällöllisistä eroista ja päällekkäisyyksistä. Aihe relevanssin määrittelemiseksi aineistosta tunnistetaan ensin sisällöllisiä teemoja esimerkiksi grounded theory -sisällönanalyysin tapaan eli aineistojen aihe relevanssin mukainen asiasisältö kategorisoidaan teemaulottuvuuksiin. Kunkin teeman osalta dokumentille voidaan antaa relevanssiarvo ja edelleen laskea niiden perusteella perinteinen, yksiulotteinen, yhdellä luvulla esitettävä relevanssiarvo.

Teemakohtainen relevanssiarvo voidaan määrittellä esimerkiksi moniportaisesti asteikolla 0, 1, 2 ja 3. Aihe riippumattomat ominaisuudet voivat olla mitä tahansa tiedonhakijalle tai relevanssiarvioijalle merkityksellisiä dokumentin piirteitä (laatu, tyyppi, kieli, tms.). Aihe riippumattomat ominaisuudet voivat olennaisesti määrittää aiheensa puolesta dokumenttien hyödyllisyyttä tietyn hakijan kannalta.

Moniulotteisen relevanssiarviointimenetelmän on arveltu aihe riippuvien ja aihe riippumattomien ominaisuuksiensa osalta soveltuvan hyvin Internet-aineiston relevanssiarviointiin. Myös tiedonhaun testikokoelmien yhteydessä aineistoja voitaisiin arvioida moniulotteisesti.

Tutkielman tarkoituksena on demonstroida henkilöihin liittyvän www-aineiston analysoinnissa. Testiaineistoksi valittiin neljää henkilöä koskevaa www-aineistoa, joihin moniulotteista relevanssiarviointimenetelmää sovellettiin relevanssikriteereiden havainnollistamiseksi. Aineistossa pyrittiin monipuoliseen kirjoon: historiallinen henkilö Bernadette Soubirous 1800-luvun Ranskassa, suomalainen nykypoliitikko Heidi Hautala, lastenkirjailija ja -kuvittaja Mauri Kunnas sekä poptaiteilija Madonna Ritchie.

Asiasanat: Moniulotteinen relevanssiarviointi, moniportainen relevanssi, aihe riippumattomat ominaisuudet, Internet-aineisto, www, web, tiedonhaku, tiedonhankinta

Sisällys

1 Johdanto	5
2 Peruskäsitteet ja tutkimussuuntaukset.....	7
2.1 Internetin määritelmiä.....	9
2.2 Informaatiolukutaidon osaamistavoitteet yliopisto-opiskelussa.....	11
2.3 Relevanssi informaatiotutkimuksen keskeisenä käsitteenä.....	12
2.4 Tiedonhaketutkimuksen relevanssikäsitteet.....	13
2.5 Cranfield-tyypin testikokoelmat	14
2.5.1 Vuorovaikutteinen TREC ja sen kritisointi IR-tutkimuksessa.....	16
2.5.2 Pia Borlundin testikokoelmakritiikki.....	18
2.6 Moniulotteiset ja dynaamiset relevanssikriteerit sekä Järvelinin ja Kekäläisen käyttöliittymäskenaariot.....	20
2.7 Testikokoelman rakentaminen.....	21
3 Moniulotteinen relevanssiarviointimenetelmä	23
3.1 Yksiulotteisista moniulotteisiin relevanssikuvauksiin.....	24
3.2 Relevanssikorpus hakutulosten eli dokumenttien ja niiden mukaisten teemojen kuvaajana.....	25
3.3 Moniulotteisten relevanssiarvioiden menettelytapa.....	27
3.4 Aiheriippumattomien ominaisuuksien merkitykset Internet-aineistossa.....	30
3.5 Aikaisemmat Internet-aineiston relevanssikriteeritutkimukset viitekehysinä relevanssiarvioinneille.....	30
4 Tutkimuksen toteutus.....	34
4.1 Aineiston valinta ja henkilötiedonhakutehtävät.....	36
4.2 Bernadette Soubirous.....	38
4.3 Heidi Hautala.....	43
4.4 Mauri Kunnas.....	47
4.5 Madonna Ritchie.....	52
5 Tulokset	56
6 Keskustelua ja johtopäätökset.....	58

LÄHDELUETTELO

I Julkaistut painetut lähteet.....	62
II Julkaistut elektroniset lähteet.....	64
III Painamattomat lähteet.....	65

LIITTEET

LIITE 1: Alkuperäinen tutkimusaineisto:

henkilötiedonhakutehtävien mukaiset URL-osoitteet Internetissä.....	66
<i>Hautala, Heidi.....</i>	<i>66</i>
<i>Kunnas, Mauri.....</i>	<i>67</i>
<i>Ritchie, Madonna.....</i>	<i>69</i>
<i>Soubirous, Bernadette.....</i>	<i>70</i>
LIITE 2: Tarina Bernadette Soubirouksesta ja Lourdesin Madonnasta.....	73

1 Johdanto

Merkittävä osa tiedonhakujärjestelmien tutkimuksesta tapahtuu ns. laboratorio-tutkimuksena, jossa hakujärjestelmin toimivuutta arvioidaan testikokoelman avulla. Testikokoelmia on käytetty jo 1950-luvun loppupuolelta alkaen. Ensimmäiset tiedonhaun testikokoelmat rakennettiin Cranfield-projektissa, jonka mukaan koko tutkimussuuntaus on saanut nimensä (Cranfield-paradigma) (Sormunen 1994). Testikokoelma koostuu dokumenttien muodostamasta tietokannasta, joukosta hakuaiheita sekä relevanssikorpuksesta, joka sisältää arvion siitä mitkä dokumentit kunkin hakuaiheen osalta tulisi tietokannasta löytää. Kansainvälisessä tiedonhaketutkimuksessa eniten käytetty TREC-kokoelma sisältää noin puoli miljoonaa dokumenttia, useita satoja määriteltyjä hakuaiheita ja näitä vastaavia relevanssikorpuksia. Relevanssiarviot ovat binäärisiä eli kukin dokumentti on joko relevantti (arvo 1) tai epärelevantti (arvo 0) sen mukaan käsittelee se hakuaihetta vai ei (Sormunen 1994). Binäärisiä relevanssiarvioita on arvosteltu epärealistisiksi mm. sen vuoksi, että ne eivät huomioi relevanssin asteen vaihtelua. Aihetta sivumennen käsittelevä dokumentti on relevanssiarvioissa täsmälleen yhtä arvokas kuin sitä laajasti käsittelevä dokumentti (Sormunen, 2002a).

Toinen binäärisen relevanssiarvioinnin ongelma on ns. riippumattomuusoletus, jonka mukaan jokainen relevantti dokumentti on relevantti muista relevanteista dokumenteista riippumatta (Sormunen 2002b). Jos hakujärjestelmä A hakee kaksi sisällöltään päällekkäistä ja hakujärjestelmä B kaksi sisällöltään toisiaan täydentävää relevanttia dokumenttia, riippumattomuusoletuksen seurauksena molempien hakujärjestelmien mitattu tehokkuus on yhtä suuri. Mittaustulos näyttää vääristyneeltä, sillä järjestelmä B tuotti käyttäjälle kaksinkertaisen määrän aiheeseen liittyvää informaatiota järjestelmään A verrattuna.

Sormunen (2002b) on esittänyt menetelmän moniulotteisten relevanssiarvioiden tuottamiseksi testikokoelmiin. Menetelmän tavoitteena on ratkaista yksiulotteisten relevanssiarvioiden rajoitteet dokumenttien sisällöllisen päällekkäisyyden tai erillisyyden tunnistamisessa. Menetelmä perustuu hakuaiheen jakamiseen teemoihin, joiden osalta dokumentin relevanssi arvioidaan erikseen eli arvioitavien dokumenttien teemaullottuvuuksille saadaan teemakohtaiset relevanssiarvot. Jos kaksi dokumenttia on

relevantteja saman teeman osalta, informaatioisällöt voivat olla päällekkäisiä. Jos vastaavasti kaksi dokumenttia ovat relevantteja eri teemojen osalta, voidaan olettaa informaatioisältöjen olevan toisiaan täydentäviä. Tässä tutkielmassa moniulotteisen relevanssiarvioinnin menetelmää sovelletaan www-aineistoon neljän esimerkkitehtävän puitteissa. WWW-aineisto on kiinnostava sovellusalue menetelmälle useasta syystä. Internetissä julkaistaan usein päällekkäistä aineistoa tai sama aineisto eri versioina. Internetin aineisto on laadullisesti hyvin heterogeenistä, joten käyttäjä joutuu tekemään valintoja monia kriteereitä soveltaen - menetelmää voi soveltaa www-pohjaisien testikokoelmien rakentamisen lisäksi tietyn aihepiirin www-aineiston rakenteen ja ominaisuuksien kartoittamiseen (Sormunen, 2006).

Moniulotteisessa relevanssiarviointimenetelmässä erotellaan sisällönanalyysin aiheiriippuvat ja aiheiriippumattomat relevanssipiirteet, jotka liittyvät esitettävään informaatioon tietyssä tiedonlähteessä tai tiedonlähteen erikseen määriteltävässä osassa tai osissa. Menetelmä on uusi teoreettinen näkökulma viitekehysineen moniulotteisiin ja joustavampiin relevanssikäsityksiin, arviointimenetelmään sekä sovelluksiin.

Tutkielman tarkoitus on demonstroida moniulotteisen relevanssiarviointimenetelmän soveltamista www-aineistoon neljään henkilöhuakeeseen liittyvän aineiston perusteella. Tarkastelun kohteena ovat erilaiset www-aineistojen sisällön ja ulkoisten ominaisuuksien arvioinnissa esiin nousevat tilanteet. Lähtökohtana on Sormusen (2002b) esittämä moniulotteinen relevanssiarviointimenetelmä mutta empiirisessä osuudessa pyritään huomioimaan myös eräiden ww-aineistojen relevanssiarviointia koskeneiden tutkimusten näkökulmana, mm. Rieh (2002), Tombros, Ruthven ja Jose (2005) sekä Crystal ja Greenberg (2006).

Tavoitteena on 1) demonstroida esitetyn menetelmän toimintaa ja käyttökelpoisuutta Internet-aineistossa tuottamalla valittuihin aiheisiin liittyvistä dokumenteista moniulotteinen relevanssikorpus ja 2) selvittää valittuihin henkilöihin liittyvien Internet-aineistojen rakennetta ja sisältöä. Tarkastelussa ovat sekä aiheeseen liittyvät teemat että temariippumattomat dokumenttien ominaisuudet, kuten tyyppi, laadukkuus tai kieli. Käyttäjä joutuu tyyppillisesti näkemään vaivaa arvioidessaan löytämäänsä aineistoa eri näkökulmista.

Moniulotteisen relevanssidatan tuottamisen merkitys liittyy tiedonhaun testikokoelmien kohtaamiin uusiin vaatimuksiin. Kyse on tutkimuksesta, jossa viitekehystenä toimivat aikaisemmat relevanssitutkimukset. Niissä on mm. selvitetty, millaisia relevanssikriteereitä käyttäjät ja tiedonhakijat pitävät tärkeinä arvioidessaan www-sivustoja (esim. Rieh, 2002). Aikaisempia tiedonhaun testikokoelmia ja tietokantoja ja niiden sisältämien dokumenttien sisältöjä on arvioitu pelkästään yksiulotteisesti joko binäärisellä tai moniportaisella asteikolla toteamalla kukin dokumentti joko relevantiksi tai sitten ei. Www-aineistot ovat heterogeenisiä ja niihin liittyy monenlaisia ominaisuuksia, jotka voidaan ja tulisi ottaa arvioinnin piiriin. Vaikka moniulotteisen relevanssiarvioinnin menetelmä on kehitetty testikokoelmien rakentamista varten, tässä tutkielmassa ei rakenneta www-testikokoelman relevanssikorpusta. Sen sijaan sovelletaan menetelmää vapaamuotoisella tavalla kerätäkseen vaikutelmia www-aineiston sisällöllisistä ja ulkoisista ominaisuuksista, joilla voi olla merkitystä tiedonhakijan relevanssiarviointiprosessissa.

Tässä tutkimuksessa itseäni kiinnostaa hakutehtävinä ranskalaiselle nuorelle, Bernadette Soubiroukselle, tapahtuneet Lourdesin Madonna-ilmiöt vuonna 1858 sekä ikinuori pop-kulttuurin vaikuttaja ja esikuva, laulaja, näyttelijä Madonna. Suomenkielisenä hakutehtävänä on lastenkirjailija ja -kuvittaja Mauri Kunnas MAUKU-tietokantaprojektin valmiissa viitetietokannassa sekä poliitikko, vihreiden kansanedustaja Heidi Hautala. Neljä tyypiltään hyvin erilaista henkilöahmoa antaa hyvän mahdollisuuden demonstroida moniulotteisen relevanssiarvioinnin menetelmää.

2 Peruskäsitteet ja tutkimussuuntaukset

Koko informaatiotutkimusta teorioineen ja käytännön tavoitteineen voidaan ajatella ”tiedonhankinnan tieteenä” (Järvelin & Vakkari, 1988). Tiedonhankintatutkimus selvittää tiedontarpeita ja niiden syntymistä ja ilmenemistä, ihmisten valitsemia tiedonhankinnan kanavia, kuinka usein ihmiset näitä kanavia käyttävät, ovatko he tyytyväisiä saamaansa tietoon ja miten he tietoa käyttävät. Informaatiotutkimuksen osa-alueilla pyritään ymmärtämään tiedonhankinnan erilaisia järjestelmiä sekä niiden käytäntöjä.

Tiedonhankinnan tutkimuksesta (IS) puhuttaessa joudutaan tarkastelemaan kolmea toisiinsa läheisesti liittyvää ilmiötä, joita ovat tiedontarve, tiedonhankinta ja tiedonkäyttö. Informaatiotutkimuksen käsitteistöön kuuluessaan tiedonhankintatutkimus rakentaa

tieteenalan käsitteistöön kuuluessaan tieteenalan käsitteellistä ja metodista perustaa. Tiedonhankinnan ja tiedonhaun merkityksen rajanveto on Savolaisen (1999, 73-109) mielestä vaikeaa, mutta käsitteet eivät kuitenkaan tarkoita samaa asiaa. Tiedonhankintatutkimuksen peruskäsite on tiedonkäyttö. Tiedonkäyttö on toimintaa, jossa yksilö hyödyntää, muokkaa ja soveltaa hankkimaansa informaatiota päämääräänsä ja tarkoituksiinsa nähden mielekkäällä tavalla.

Tiedonvälittäjät ovat (inhimillisiä) sekundääritiedonlähteitä: he eivät siis välttämättä itsessään omaa mitään alkuperäistä, tiedontarvetta varsinaisesti tyydyttävää informaatiota, mutta sen sijaan he osaavat kertoa tai näyttää, mistä yksilön tarvitsemaa informaatiota tulisi saada. Tiedonlähde on yksilön tietyssä tilanteessa tarvitseman tiedon varsinainen kantaja. Tiedonhankintakanaviin sisältyvät tiedonvälittäjät, tiedonlähteet ja erilaiset informaatioteknologiat. Informaatioteknologioita puolestaan on olemassa kolmenlaisia: 1) formaalit informaatioteknologiat, 2) sellaiset teknologiat, joilla on informatiivisia funktioita primäärien ei-informatiivisten funktioidensa lisäksi ja 3) muut ihmiset (informaalit tiedonhankintakanavat). Vaikka tarkkaan ottaen tulisikin puhua "informaationhankinnasta" tai sen "etsinnästä", käytetään yleensä ilmaisua "tiedonhankinta". Perusteena on, että ihminen ei yleensä keräile informaatiota, vaan käyttää sitä mm. ongelmanratkaisussa. Tämä päämäärän aspekti puoltaa ilmaisun "tiedonhankinta" käyttöä (Savolainen, 1998). Oman käsitykseni mukaan tiedonhankinta on juuri informaation keräilemistä.

Tiedontarve- ja tiedonhankintatutkimuksessa merkittävimmän tiedontarvetta jäsentävän mallin on esittänyt Brenda Dervin. Brenda Dervinin Sense making- teoria on lähestymistapa, metodi ja konsepti, prosessi ja kommunikatiivinen käyttäytyminen, jonka keskeisiä toimintoja ovat tiedon etsintä, muokkaus, luominen tai käyttö. Sense making -teorian lähestymistapa mallintaa tiedonhankintakäyttäytymistä kolmen perustekijän pohjalta. Nämä ovat 1) tiedonhankinnan tilannekonteksti (situation), 2) tarve hankkia tietoa (gap), 3) ne merkitykset, joita tiedon hankinnalla ja käyttämisellä on toimijalle (tai toivoo niillä olevan) (uses, helps). Luonteenomaisinta perusteelliselle sense making-teorialle on yksittäisen toimijan korostaminen, mutta nykyään on mukana enemmän yhteisöllisempi tarkastelu ja kielenkäyttö. Sense making -teorian perusolettamuksena on yhtäältä ihmillisen käyttäytymisen ainutlaatuisuuden ja monimuotoisuuden hyväksyminen, mutta toisaalta sen rajoittuneisuuden ymmärtäminen. Tältä pohjalta on helppo ymmärtää, miksi sense making -teorian mukaiseen toimintaan voi ottaa kriittisen kannan sen objektivismin

tai positivismin suhteen (Savolainen, 1998).

Tiedonhaku (information retrieval, IR) on se osa tiedonhankintaa, joka toteutetaan käyttämällä apuvälineenä tietokonetta ja tiedonhaku kohdistuu usein tietokantoihin. Tiedonhaussa saavutetaan tuloksia, jos osaa muotoilla oikean hakulausekkeen ja käyttää tietokannan edellyttämiä hakutekniikoita. Tiedonhaussa keskitytään hakulausekkeiden muotoiluun ja arvioidaan haun avulla löydettyjen dokumenttien relevanssia.

Tiedonhaun tavoitteena on löytää hyödyllinen informaatio tiedontarvitsijalle tai erityiseen tiedontarvetilanteeseen oikeassa muodossa. Relevantti dokumentti voi vaihdella aineistotyyppin ja sen sisällön, rakenteen ja ulkoasun suhteen. Relevanssiin vaikuttavat myös aineiston käytettävyys sekä saatavuus. Tiedonhaun perusongelmiin kuuluu usein hakupyynnön ja hakumekanismin epätäydellinen yhdisteleminen käyttöliittymässä tai hakumekanismin järjestelmä- tai käyttäjälähtöiset rajoittavat tekijät (Kekäläinen, 2002).

Alaterän ja Halttusen mukaan (2002) termi tiedonhaun tutkimus ei tarkoita vain haun tutkimusta, vaan myös tiedon tallennuksen, sen organisoinnin ja jäsentämisen tutkimusta. Tiedonhaun lähtökohta on informaation tarve, jossa yksilö toimii tiedon hankkijana / hakijana muotoilemalla hakukyselyn tietokantaan ja löytämällä hakua vastaavan tuloksen. Tietokantaan kuvailtu ja tallennettu tieto viittaa alkuperäisiin dokumentteihin, joiden alkuperäisinä ja vastuullisina tekijöinä ovat dokumenttien tuottajat. Tiedonhaussa on esitettävä kysymys, mihin informaation tarpeisiin tiedonhaku pyrkii vastaamaan (Sormunen, 2003b).

2.1 Internetin määritelmiä

Internet-aineistosta voisi käyttää termiä www-aineisto tai web-aineisto, mutta itse suosin termiä www-aineisto. Internet tai www eli world wide web, erityisesti web on tiedonhakuympäristö, johon voidaan soveltaa perinteisten tietojärjestelmien tiedonhaun ja tiedonhankinnan menetelmiä.

Hakala (2002) lainaa Internet-määritelmänsä Lynchin ja Berners-Leetä:

Internetiä ja erityisesti sen World Wide Webin nimellä tunnettua multimediodokumenttien

kokoelmaa, ei kirjastojen tavoin suunniteltu tukemaan organisoitua tiedon hakua ja tallennusta. Voidaan sanoa, että www:stä on tullut maailman ”elektronisten kirjapainojen” tuotteiden kaoottinen varasto. Tämä varasto ei sisällä ainoastaan kirjoja ja tutkimusraportteja, vaan myös kokouspöytäkirjoja, mainoksia, video- ja audiotallenteita ja tieteellistä raakadataa. Joutavanpäiväinen sekä pysyvästi arvokas aineisto esiintyy kaikkialla. Tähän asti World Wide Webiä on kehitetty pikemminkin ihmisille tarkoitetuksi dokumenttien välitysjärjestelmä ksi kuin informaation automaattiseksi käsittelyjärjestelmäksi. Rikastamalla Web-sivuja tietokoneille tarkoitettulla datalla ja lisäämällä verkkoon vain konelukuisiksi tarkoitettuja sivuja voimme muuttaa World Wide Webin semanttiseksi verkoksi.

Iivosen ja Halttusen (1999) mukaan Internet on avoin tietoverkko, johon voidaan laskea kuuluvaksi fyysiset tiedonsiirtoverkot, eri pisteissä fyysisesti sijaitsevat tietokonelaitteet ja niissä käytettävät ohjelmistot sekä tietokoneiden muistiin tallennetut tiedostot ja niissä oleva informaatio. World Wide Web, lyhennettynä web on Internetin kautta tapahtuvaan tiedonhakuun kehitetty palvelu, jonka kautta voidaan siirtää dokumentteja. Tiedon saatavuuden kannalta www voidaan ymmärtää suurena, lähes rajattomana tietovarantona, koska suuri määrä potentiaalisesti hyödyllistä informaatiota on sen kautta saavutettavissa riippumatta siitä, mihin kyseinen informaatio on fyysisesti tallennettuna (Halttunen & Iivonen, 1999).

Heikkisen, Aganderin, Ijäksen ja Laitisen (2005) mukaan suuri osa www-sivuista on käyttäjille avoimia, mutta etenkin moniin tietokantoihin ja elektronisten julkaisujen kokoelmiin on pääsy käyttäjätunnuksen tai salasanan takana, tai niihin pääsee vain paikallisen tietoverkon kautta, jonka käyttöön hallinnoiva organisaatio on ostanut käyttöoikeuden. Oma lukunsa ovat organisaatioiden sisäiseen viestintään tarkoitettut, muilta suljetut intranetit. Internetissä on hakupalveluita, joista suosituimpiin lukeutuu Google (www.google.fi) . Tiedonhakujärjestelmistä Google on tunnettu juuri hakukoneestaan ja Yahoo (www.yahoo.com) aihehakemistostaan. Kirjoittajat toteavat, että Internetissä on paljon materiaalia, jota hakukoneet eivät pysty indeksoimaan. Aikaisemmin todetusti sivuston tekijät ovat voineet estää sen. Hakukoneet eivät myöskään indeksoi sivuja, jotka luodaan esimerkiksi Internetissä olevissa tietokannoissa hetkellisesti tiedonhaun tuloksena. Googlella ei voi tarkistaa, löytyykö jonkun kirjaston kokoelmista haluamaasi kirjaa. Silloin on haettava ensin tiedontuottajan (korjaisiin tiedonvälittäjän) eli kirjaston kotisivu ja

etsittävä sitten ”teosta” kirjaston kokoelmätietokannan hakusivuilta (Heikkinen & Agander & Ijäs & Laitinen, 2005, 99-100, 113).

Tässä tutkielmassa keskitytään kuitenkin avoimen Internetin hakutehtävälähtöiseen relevanttien dokumenttien sisällönanalyysiin. Verkkotiedonhaku, josta tässäkin tutkielmassa on lopulta kyse, kuuluu tiedonhaku tutkimuksen uudempiin pääsuuntauksiin. Verkkotiedonhaussa voidaan tutkia joko hakumenetelmiä tai käyttäjiä. Verkkotiedonhaussa integroituu tiedonhaku- ja tiedonhankinta.

2.2 Informaatiolukutaidon osaamistavoitteet yliopisto-opiskelussa

Informaatiolukutaitoon viitataan usein tietoyhteiskuntakehityksen ja tiedonhaun yhteydessä. Informaatiolukutaito käsitteenä on monissa eri yhteyksissä kansalaisvalmius samalla tavalla kuin lukutaito on totuttu näkemään. Sormusen mukaan (2003) informaatiolukutaitona Google-tyyppisen logiikan sisäistäneille osaamisen laajentaminen voi olla vaikeaa. Tehokkaat hakustrategiat ovat usein välinekohtaisia. Informaation helppo saatavuus voi ruokkia älyllistä laiskuutta ja houkutus tuottaa leikkaa / liimaa -raportteja (Sormunen, 2003a).

Lähtökohta on kyky tunnistaa, rajata ja ilmaista tiedontarpeensa. Tämä on myös yliopisto- ja korkeakouluopetuksen informaatiolukutaidon ensimmäinen standardi, josta suositeltavampi sanamuoto on tosin osaamistavoite (Informaatiolukutaidon osaamistavoitteet yliopisto- ja korkeakouluopetuksessa, 2006). Toinen osaamistavoite on tehokas tiedonhaku, jonka indikaattorien tulokseksi voisi tiivistää, että informaatiolukutaitoinen korkeakouluopiskelija poimii, tallettaa ja käsittelee tietoa ja käytettyjä tiedonlähteitä. Kolmannen osaamistavoitteen mukainen tiedon käsitteleminen on tiedon arviointia, kriittistä tarkastelua sekä tiedon soveltamista. Neljäs osaamistavoite on tiedon käyttäminen tiettyyn tarkoitukseen. Osaamistavoiteisiin voi liittyä sen pohtiminen, mitä kuluja ja hyötyä tarvittavan tiedon hankkimiseen liittyy sekä suorituksen aikaansaamiseen tähtäävää kehitysprosessia. Viides osaamistavoite on oikean ja hyvän tavan mukainen tiedon käyttö, johon kuuluvat tekijänoikeuksiin liittyvät kysymykset, maksuttoman tai vapaan tiedon käyttäminen, kirjaston säännöt sekä tieteellinen viittaaminen. Kyky tunnistaa puutteita omassa toiminnassa luo edellytyksiä

itseohjautuvaan oppimiseen ja informaatiolukutaidon kehittämiseen (Informaatiolukutaidon osaamistavoitteet yliopisto- ja korkeakouluopetuksessa, 2006).

Kauton (2003) lisäys informaatiolukutaidon standardeihin lähetettynä Helsingin yliopiston ylläpitämälle tiedonhallinnan sähköpostilistalle on hänen mukaansa Maureen Nimomin päätelmä (2002). Päätelmä on, että omavaraisuus ei ole informaatiolukutaitoisen henkilön välttämätön ominaisuus. Informaatiolukutaitoinen henkilö kykenee arvioimaan, milloin hän tarvitsee asiantuntijan apua ratkaistaakseen informaationtarveongelmaansa.

Avonetin osalta informaation arviointivastuu on siirtynyt entistä enemmän käyttäjälle. Se on informaatiolukutaidon kannalta lisähaaste. Internetissä ei ole sisältökontrollia perinteisten julkaisukanavien tapaan. Web-aineistossa hakija joutuu arvioimaan entistä tarkemmin löytämiensä dokumenttien sisältöjä. Sisällön lisäksi dokumentin käyttökelpoisuutta ja luotettavuutta arvioidaan myös muiden dokumentin ominaisuuksien perusteella (ks. Rieh. 2002). Tässä tilanteessa on helppo nähdä verkkotiedonhaun tutkimuksen tarve kehittää monipuolisempia relevanssiarviointimenetelmiä. Dokumenttien eri ominaisuudet ja moni-ilmeiset erityispiirteet Internetissä pitäisi saada mukaan relevanssiarvioihin. Informaatiolukutaitoon liittyvä tiedonhaun tutkimus voisi hyötyä moniulotteisesta relevanssidatasta.

Ekholm (2002) toteaa tieteellisten kirjastojen olleen aina sisällöntuottajia, lähinnä palvelukokonaisuuksiin liittyvien sisältöjen, kuten tietokantojen ja www-sivustojen tai portaalien tuottajia. Kirjastot ovat kuitenkin hyvin niukasti toteuttaneet oppimateriaaleihin, sähköisiin kirjoihin ja aineistoihinsa liittyvää digitointia ja välitystä (Ekholm, 2002). Näissäkin palveluympäristöissä aineisto on moninaista, joten moniulotteisen relevanssidatan tarve on ilmeinen tämänkin alueen tiedonhaketutkimuksessa.

2.3 Relevanssi informaatiotutkimuksen keskeisenä käsitteenä

Relevanssi on tiedonhaun arvioinnin tai evaluoinnin keskeinen perusta. Relevanssi on informaatiotutkimuksen historiallinen käsite, jonka mittaamisen, määrittelemisen ja tarkastelunäkökulmien ongelmista on keskusteltu noin 50 vuotta. Relevanssi jakautuu klassisessa näkökulmassa käyttäjärelevanssiin ja aiherelevanssiin. Edellisen mukaan

tiedontarvitsija tarvitsee hyödyllisen dokumentin (useful document). Aiherelevanssin perusteella (topicality) dokumentti on tarkoituksenmukainen ja vastaa sisällöllisesti arvioijansa tiedontarvetta (aboutness).

Tuoreemmissa tulkinnoissa relevanssin lajeja ovat algoritminen relevanssi tai "järjestelmärelevanssi", aiherelevanssi, kognitiivinen relevanssi, tilannerelevanssi ja motivaatio- tai tunnerelevanssi. Relevanssin arvioiminen perustuu myös subjektiivisen tulkintaan tai merkityksenantoon todennäköisyysilmiönä. Relevanssin aspekteja voivat olla esim. hyöty, täsmäys, informatiivisuus, tyytyväisyys tai asianmukaisuus. Tarkastelun viitekehystenä voi olla tiedontarve, pyyntö, hakutehtävä tai näkökulma. Relevanssiarvioija voi olla kysyjä, välittäjä, asiantuntija, käyttäjä, henkilö tai automaattinen järjestelmä (Järvelin, Kekäläinen 1998).

Realismia korostavat tutkijat näkevät relevanssin suhteena sisällön, sen käyttäjän ja käyttötilanteen välillä. Tämä tarkoittaa käyttäjän arviota sisällön käyttökelpoisuudesta käyttötilanteessa käyttäjän tavoitteiden, arvojen ja odotusten kannalta. Nyrkkisääntönä ja ideaalina on, että käyttäjä määrittelee relevanssin. Relevantti informaatio sisältää tarvitun informaation ja relevanssin määrittelee tiedontarvitsija. Arvioijalle relevanssi on jatkuva muuttuja perustuen hänen ymmärrykseen hakuaiheesta. Relevanssia käytetään usein tiedonhakujärjestelmän tuloksellisuuden mittaamisen perustana (Kekäläinen, 2002).

Halttusen ja Iivosen (1999) mukaan hakutuloksen tärkein arviointikriteeri on laatu. Muina kriteereinä on pidetty hakutuloksen kokoa, saantia, tarkkuutta, uutuusarvoa, hyödyllisyyttä ja kustannuksia. Laadun arviointi on kuitenkin olennainen kysymys sekä tiedonhakijan että tiedonhakupalveluiden tuottajien kannalta.

2.4 Tiedonhaku tutkimuksen relevanssikäsitteet

Relevanssin käsitteen monet erilaiset "tasot" ovat tiedonhaku tutkimuksen ja tiedonhankinnan vaiheiden selittäviä tekijöitä. Tutkimuksessa tiedonhaun tuloksen arviointi eli evaluointi voidaan suorittaa varsinaisen tiedonhaku tapahtuman monista eri osa-alueista ja lähestymistavoista lähtien. Relevanssiarvioinnille ei ole olemassa vain yhtä lähtökohtaa (Kekäläinen, 1998).

Ingwersenin (1995) mukaan tiedonhaku tutkimuksen valtavirta on kiinnostunut hakualgoritmeista joka edustaa järjestelmälähtöistä näkökulmaa tiedonhakuun. Käyttäjäsuauntautunut lähestymistapa tiedonhakuun pohjautuu pääasiassa kognitiiviseen psykologiaan sekä sosiaalitieteiden metodeihin. Nämä lähestymistavat ovat tarjonneet merkitykselliset näkymät tiedonhakijoiden käyttäytymiseen ja heidän tiedonhankintansa ominaispiirteisiin sekä yksilöpohjalta että organisaationallisessa kontekstissa. Näin on mahdollista tutkia myös informaatioammattilaisen ja käyttäjän välistä vuorovaikutusta. Käyttäjille pitäisi suunnitella ja määritellä mukavakäyttöisiä järjestelmiä ja ratkaista muitakin IR-vuorovaikutuksen ongelmia. Tiedonhaun pitäisi olla ongelmanratkaisua ja päätöksentekoa tukeva prosessi.

Tiedonhaun evaluointi tai tulokset operationaalisten tiedonhaku järjestelmien tutkimuksissa perustuvat käyttäjän relevanssiarvioihin ja niiden perusteella laskettuun saantiin ja tarkkuuteen. Käyttäjäsuauntautuneessa, tehtäväsuuntautuneessa tai kognitiivisessa tutkimuksessa on paljon yhteisiä piirteitä, joissa korostuu tiedonhaku prosessina sekä tuloksiin pyrkivänä toimintana. Käyttäjäsuauntautuneen tutkimuksen kohteena voivat olla asiakkaan, informaation käyttäjän sekä välittäjän vuorovaikutus ja roolit sekä heidän hakukäyttäytymismallinsa ja -menetelmänsä. Tehtäväsuuntautuneessa tiedonhaku tutkimuksessa ollaan kiinnostuttu yksilön ongelmanratkaisuprosessista eli tehtävän ja tiedontarpeen vaikutuksesta hakukäyttäytymiseen konkreettisissa tilanteissa sekä tiedonhaun varsinaisesta tuloksellisuudesta.

Tiedonhakutehtävien luonteesta, tiedonhakijoista tai tiedon käytön tarkoituksista riippuen ei voi aina välittäjänä, tiedonhakijana tai -tarvitsijana ottaa kantaa haetun informaation totuusarvoon tai käytettävyys- ja saatavuustekijöihin. Informaatio voi olla fiktiota tai faktaa tai sen löytymiseen voivat vaikuttaa muut tilanne- tai taustasidonnaiset tekijät (Kekäläinen, 2002).

2.5 Cranfield-tyypin testikokeemat

Sormusen (2006) mukaan valtaosa nykyisistä testikokeelmista tiedonhaku järjestelmien evaluoinnissa perustuu malliin, joka kehitettiin Cranfield-projekteissa varhain 1960-luvulla ja jota sovelletaan edelleen kehitetyissä muodoissaan. TREC on tunnetuin Cranfield-

tyyppisistä testikokoelmista. Samankaltainen lähestymistapa on omaksuttu CLEF-konsortion kieltenvälisen tiedonhaun ja yksittäisten tutkijaryhmien testikokoelmissa. Perinteisiä testikokoelmia käytetään pääasiassa yksittäisten kyselyjen tai niiden muunnelmien eräajomuotoisessa evaluoinnissa, jossa käyttäjän ja järjestelmän sekä käyttäjän ja informaation välisellä vuorovaikutuksella on vähäinen rooli. Cranfieldin lähestymistapa yksinkertaistaa relevanssin aiheriippuvaan relevanssiin ja esittää sen kaksiportaisena, binäärisenä (binary values) arvona luonnehtien testikokoelmassa hakuaiheen ja yksittäisten dokumenttien staattista luonnetta. Viisitasoisista relevanssiskaalaa käytettiin kuitenkin Cranfield-projektissa (Cleverdon 1967, tässä Sormunen, 2006 mukaan).

Binäärisissä arvioissa julkaisu on relevantti arvioijana toimivalle hakuaiheen määrittelijälle arvolla 1 tai epärelevantti arvolla 0. Relevantin dokumentin voi määrittää arvolla 1 mikä tahansa relevanssin piirre esim. seuraavista: julkaisu on täydellinen vastaus, julkaisun relevanssin aste korkea, mutta jotain puuttuu, julkaisu on hyödyllinen ja sisältää yleistä taustaa tai julkaisu on vähäisesti kiinnostava mutta sisältää esimerkiksi historiallista tietoa relevantisti.

Cranfield ja sen jälkeiset moniulotteisemmat relevanssiarviointimenetelmät ovat paradigma tietojenkäsittelysuuntautuneessa tutkimuksessa. Sormusen kirjallisuuskatsauksen (2002a) mukaan (esim. Borlund, 2000) tarve ylläpitää kontrollia on johtanut ongelmiin realismin kustannuksella, johon kuuluvat muun muassa käyttäjien soveltamat moniulotteiset ja dynaamiset relevanssikriteerit. Järvelin ja Kekäläinen (2002) kutsuvat aiheenmukaisuuden (topicality) ylittäviä subjektiivisten relevanssikriteerien kokonaisuutta korkeamman mittaluokan relevanssiksi. Heidän mukaansa erittäin ratkaisevat dokumenttien neljä ominaisuutta ja niistä neljä merkittävintä ovat jääneet hunningolle aiheriippuvan relevanssin binäärisessä arvioinnissa.

Kriittisiä huomiotta jätettyjä tekijöitä ovat ensinnäkin relevanssin aste, joka määrittää kuinka perusteellisesti aihetta käsitellään dokumentissa. Toiseksi dokumenttien sisällölliset päällekkäisyydet tarkoittavat, että kaksi dokumenttia voi sisältää saman informaation sisällön. Informaation sisällöllä, jotka löydetään toiseen kertaan, ei ole käyttäjälle enää samanlaista arvoa. Kolmanneksi virheiden ja poikkeavuuksien osalta dokumentti voi sisältää virheitä tai olla ristiriidassa muiden dokumenttien sisältöjen kanssa. Erilainen tieto samasta asiasta on

tärkeää tiedontarvitsijalle, vaikka onkin jäänyt tarkastelematta aiheriippuvuuteen perustuvissa relevanssiarvioinneissa. Neljänneksi dokumenttien aiheriippumattomat ominaisuudet antavat paljon lisäinformaatiota ja ovat tärkeitä kriteereitä käyttäjille arvioitaessa dokumentin hyödyllisyyttä. Näitä ominaisuuksia voivat olla kieli, genre, tekijät, lähteet, ikä, konteksti, media, jne.

On tärkeää huomata, että Järvelinin ja Kekäläisen (2002) määrittelemät neljä tekijää ovat luonteeltaan staattisia ja relevanssiarvioija voi tulkita ne dokumentista tai sen metadatasta itsenäisestikin muista dokumenteista tai käyttötilanteesta riippumatta. Täten neljää mainittua dokumentin ominaisuutta voidaan soveltaa standarditestikokoelmien relevanssiarvioinnissa. Ongelmana on, kuinka kerätä ja esittää moniulotteista relevanssidataa ja soveltaa sitä testikokoelmissa tai tiedonhakukokeissa.

Perusmotivaatio moniulotteisten relevanssisuunnitelmien kehittämisessä testikokoelmiksi ei lähde siitä, että hakualgoritmien standarditestausta sitä vaatisi. Kekäläinen ja Järvelin (2002) relevanssin moniulotteisuuden merkityksen siinä, että se tarjoaa viitekehyksen kehittää tiedonhaun testikokoelmia kohtaamaan vaatimukset kehittyvissä vuorovaikutteisten hakuprosessien tutkimisena. Moniulotteinen relevanssiarviointimalli on aikaisemmista testikokoelmista kehitettynä suunnitelma moniulotteiselle relevanssitalle ja menettelytavalle kerätä dataa.

2.5.1 Vuorovaikutteinen TREC ja sen kritisointi IR-tutkimuksessa

Sormusen mukaan (2006) vuorovaikutteinen TREC (Interactive track of TREC) on selkeästi mallintanut standarditestikokoelmien käytön vuorovaikutteisten hakuprosessien opiskelussa. Vuorovaikutteinen TREC on hakutehtävä- ja testikokoelmapohjainen lähestymistapa vuorovaikutteisen tiedonhaun tutkimukseen. Relevassi perustuu näkökulma- tai esimerkkirelevanssiin (aspectual tai instance relevance) vailla muita riippumattomuusoletuksia. Binääriset tuloksellisuusmittarit ovat instanssisaanti ja - tarkkuus. TREC-testikokoelmien relevanssiarvioissa voidaan käyttää palkattuja arvioijia. Nämä ideoivat myös hakuaiheet. TREC 1:n ja TREC 2:n aikoina testikokoelmat suunniteltiin pääasiallisesti eräajokokeisiin. Samoja kokoelmia käytettiin myöskin tutkittaessa vuorovaikutteisia hakua. Tulokset eivät olleet rohkaisevia, koska kokeellinen

suunnittelu kärsi hakuaiheiden epärealistisesta luonteesta. Erillinen vuorovaikutteinen testi suunniteltiin TREC 3:a ja 4:ää varten, mutta kokoelmien sijasta hakuaiheissa sovellettiin niiden osajoukkoja. TREC 6:ssa, perinteiset relevanssiarvioinnit korvattiin näkökulmakohtaisilla (aspectual) arvioinneilla ja standardimuotoiset saanti- ja tarkkuusmittarit korvattiin näkökulmaperustaisilla tarkkuus- ja saantimittareilla. Hakutehtävänä oli: *löydä niin monta relevanttia dokumenttia kuin mahdollista 20 minuutissa niin, että ne yhdessä käsittelisivät mahdollisimman monta erilaista hakuaiheen näkökulmaa.* TREC-7:ssä näkökulmatermi (aspect), korvattiin termillä ”instance” ja haku aika rajoitettiin 15 minuuttiin, mutta palautettiin 20 minuuttiin TREC-8:ssa. Viimeisimmillä kierroksilla käytettiin kuutta aihetta, jotka räätälöitiin vuorovaikutteista tiedonhakua IIR:ää varten. Yhteenvedon mittoina käytettiin näkökulmatarkkuutta ja saantia sekä hakuun kulunutta aikaa.

Kokemukset vuorovaikutteisesta TRECistä ovat osoittaneet, että perinteiset testiasetelmat, jotka perustuvat binäärisiin relevanssiarviointeihin, eivät anna asianmukaista perustaa IIR:n eli vuorovaikutteisen tiedonhaun ilmiöiden opiskelua varten. Vuorovaikutteisen TRECin tärkein saavutus on ollut, että se on kiinnittänyt huomion informaation päällekkäisyyksien ongelmaan. Käyttäjälle tai järjestelmälle päällekkäisen tiedon löytämistä ei katsota ansioksi.

Sormusen (2002b) artikkeli TREC-7:n ja TREC-8:n dokumenttipoolien uudelleen arvioinnista moniportaisen relevanssiasteikon mukaan 38 aiheen (topic) osalta arviot tehtiin neliportaisesti poiketen normaalista binäärisestä arvioinnista. Tutkimuksen tarkoituksena oli luoda TRECin alakokoelma, jossa voitaisiin tutkia tiedonhakujärjestelmien kykyä keskittyä erittäin relevantteihin dokumentteihin, analysoida relevanttien dokumenttien kokoelmaa sekä oppia tasorelevanssiarvioiden tekoprosessista. Moniportaisten relevanssiarvioiden käyttäjärelevanssin käyttöä motivoi tarve kehittää tiedonhakujärjestelmiä, joilla on mahdollista löytää erittäin relevantteja dokumentteja. Tutkimusongelmana oli alkuperäisen binäärisen relevanssiarvion ja uudelleenarvioitujen relevanssikorpuksen erot.

Alkuperäisen TRECin ja Tampereen yliopiston (UTA) tekemien arvioiden eroavaisuudet olivat siinä, että alkuperäiset arvioijat olivat itse luoneet aihekategoriat, määritelleet kaksiportaiset relevanssikriteerit ja tuottaneet aihekuvaukset. UTATRECin projektissa

turvautettiin TRECin valmiisiin kuvauksiin ja relevanssikriteerit perustuivat samoihin oletuksiin, mutta niitä laajennettiin 4-portaisiksi. Toinen eroavaisuus tutkimusten välillä oli tutkimusaineisto. UTATRECin tutkimuksessa olivat mukana kaikki TRECin relevantit dokumentit, mutta vain 5 % epärelevanteista dokumenteista. Lähes 700 TRECin 5737 dokumentista, jotka on arvioitu TRECin tutkimuksessa relevanteiksi, osoittautuivat UTA:n arvioinnissa epärelevanteiksi. Lisäksi 1004 TRECin relevanteista dokumenteista arvioitiin olevan vain marginaalisesti relevantteja. UTAn projektissa arvioitsijoilla oli vaikeuksia tunnistaa marginaalisesti relevantteja dokumentteja. Tästä voidaan tehdä johtopäätös, jonka mukaan TRECin arvioinnissa dokumentin relevanssiaste voi olla alhainen saaden arvon relevantti. Artikkelin keskeiset johtopäätökset korostavat moniportaisten relevanssiarvioiden etuja binääriseen arviointiin nähden. Moniportaisesti luotujen testikokoelmien perusteella voidaan testata tiedonhakujärjestelmien kykyä löytää erittäin relevantteja dokumentteja.

Vuosien kuluessa päämäärä vuorovaikutteisten järjestelmien vertaamisessa (ihminen ja järjestelmä) automaattisiin järjestelmiin, on korvattu testeillä verrata vuorovaikutteisia järjestelmiä omanlaistensa joukossa. TREC-6:n jälkeen päämääräksi ovat sitoutuneet systemaattiset cross-site-vertailut. Erilaiset ryhmät noudattavat erilaista tutkimusaikataulua. Over (2001, tässä Sormusen 2002b mukaan) on tehnyt johtopäätöksen, että pienemmän skaalan yhteenliittymiä tarvitaan, että hypoteesien koordinoinnissa ja erityistehtävien ja -kokoelmien valikoinnissa olisi vähemmän ongelmia (Sormunen, 2006).

2.5.2 Pia Borlundin testikokoelmakritiikki

TRECin kokeellisesti suunnitellut järjestelyt ovat joutuneet Åbo Akademiassa informaatiotutkimuksesta väitelleen Pia Borlundin kritiikin kohteeksi seuraavasti (2000):

Hakupyynnöt ja hakuaiheet, joita käytetään, ovat staattisia, rajoittuneita ja epärealistisia. Tiedontarpeelle ei ole annettu kontekstia. Ennalta määritellyt hakuaiheet eivät salli kokeeseen osallistuvien hakijoiden kehittää henkilökohtaisia tulkintoja informaation tarpeista tai kehittää niitä istuntojen aikana. Hakupyynnöjen keksijät ja kokeelliset hakijat opastetaan arvioimaan relevanssia binäärisellä (kaksiportaisella) tavalla havaitun aiheenmukaisuuden mukaan. Todelliset käyttäjät tekevät relevanssiarvioita erilaisiin tarkoituksiin ja heidän täytyy kohdata erilaisia relevanssikriteereitä (moniulotteinen ja

dynaaminen relevanssi).

Pia Borlund on tehnyt johtopäätöksen, että interaktiivisen TRECin kokeellinen suunnittelu käsittää hakijan roolin niin, että vuorovaikutteista IR-järjestelmää on opiskeltu ei-vuorovaikutteisen IR-järjestelmän viitekehyksessä. TREC-mallin kritiikki perustuu siihen, että esimerkkitestikokoelmana kontrollia on korostettu todellisen realismin mukaisen relevanssiarvioinnin kustannuksella. Vuorovaikutteisuus on unohdettu testitiedonhauissa ja hakuaihe korostuu enemmän kuin hakijan tiedontarve.

Borlundin (2000) vaihtoehtoinen malli perustuu simuloituihin eli jäljitelyihin tehtävätilanteisiin. Arviointeihin tarvitaan uskottavat koehakijat. Relevanssin arviointien merkitys laajentuu perinteisten arvioiden rinnalla. Muodostuu uusi käsite: Relative Relevance - suhteellinen relevanssi. Mukaan tiedonhakuun voidaan ottaa tehokkuusmittareita. Mallin mukaista toimintaa voidaan kritisoida ja tarkastella tehtyjen relevanssiarvioiden mukaisesti samoin kuin työtehtävän simuloinnin onnistumisella.

Borlund (2000) teki ehdotuksia vuorovaikutteisten tiedonhakujärjestelmien arviointiin. Simuloitu työtehtävätilanne on tämän olennainen elementti. Se on pieni peitetarina (cover story), joka palvelee kahta tehtävää. 1) Se on käynnistävä tekijä ja kehittää kuvatun informaation tarpeen sallimalla käyttäjien tulkinnat tilanteesta johtaen kognitiivisesti yksilöä tulkintoihin, niin kuin ”todellisessa elämässä” ja 2) se on alusta, johon tilanteenmukaista relevanssia arvioidaan. ”Peitetarinalla” on tarkoitus varmistaa sekä realismi, että kokeen kontrollointi. Borlund (2000) osoittaa tarpeen kehittää realistisempia malleja vuorovaikutteisten hakujärjestelmien kontrolloituihin kokeellisiin arviointeihin. Hän painottaa myös moniulotteista sekä tiedonhakijapohjaista dynaamisuonteista relevanssia. Kuitenkaan hän ei tee yhtään ehdotusta, kuinka kehittää menetelmiä ja esittää ja tehdä objektiivisia relevanssiarvioita testikokoelmissa. Tämä vaikuttaa siihen, miten standarditestikokoelmia aiotaan saada työn alle IIR-järjestelmien opiskelemista varten (Sormunen, 2002b).

2.6 Moniulotteiset ja dynaamiset relevanssikriteerit sekä Järvelinin ja Kekäläisen käyttöliittymäskenaariot

Kritiikki Cranfieldin paradigmaa vastaan ja relevanssin yksinkertaistaminen relevanssiasteiden binäärisestä aihe riippuvuusarvosta on noussut esiin tiedonhankinnan ja käyttäjäsuuntautuneen tiedonhaun tutkimuksissa. Kuitenkin 2000-luvulle asti tiedonhakuvaluoinnin käytännössä binäärinen skaala on ollut normi. Todelliset tiedonhakijat tekevät kuitenkin ei-binäärisiä, subjektiivisia ja dynaamisia relevanssiarvioiteja tiedonhakuprosesseissa, mitkä vaikuttavat suoraan hakutilanteessa. Tässä tarkoituksessa relevanssi on moniulotteinen, eikä sitä voida johtaa mistään yksittäisestä relevanssikriteeristä. Relevanssi on monitasoinen ilmiö tarkoittaen, että jotkut dokumentit ovat relevantimpia kuin toiset käyttäjän tiedontarpeen näkökulmasta.

Järvelinin ja Kekäläisen (2002) relevanssin käsitteiden analyysi on tuottanut tuloksia korkeamman asteen relevanssien identifioimisessa, joita ovat algoritmisen ja aihe relevanssin lisäksi kognitiivinen relevanssi ja tilanteenmukainen relevanssi. Dokumentin havaittu relevanssi perustuu hakijan subjektiiviseen arvioon. Hän voi käyttää monenlaisia enemmän tai vähemmän eksplisiittisiä (avoimia) kriteereitä arvioinnissa. Relevanssia pidetään näin ollen moniulotteisena kognitiivisena käsitteenä.

Moniulotteisen ja dynaamisen relevanssin haasteiden edessä (Järvelin & Kekäläinen, 2002) on olemassa epärealistisia oletuksia, mitä tulee dokumenttien itsenäisyyteen. Jotkut dokumentit voivat olla relevantteja vain, jos ne asetetaan rinnakkain. Dokumenttien sisällön samankaltaisuus ”relevanteissa” hakutuloksissa on toistetun sisällön uudelleen esiintymänä käyttäjän kannalta epärelevanttia.

Järvelinin ja Kekäläisen (2002) mukaan oletukset ovat kuitenkin välttämättömyys siitä lähtien, kun relevanssiarvointivaiheessa olevaa tiedonkäyttäjää ei ole informoitu mahdollisista dokumenttien yhdistelmistä, joita ei ole haettu kyselyn tai laajan tiedonhankintaprosessin yhteydessä. Kukaan ei ole pystynyt yli dokumenttien menevää arviointiprosessia. Sellainen jää kokonaan käyttäjän toiminta-alueelle.

Saanti ja tarkkuus ovat pääasiallisia tehokkuuden mittareita tiedonhaussa. Näiden mittarien

yhteydessä pitää kuitenkin muistaa, että yleensä ne perustuvat binäärisiin relevanssiarviointeihin. Ne heijastavat kuitenkin relevanssin luonnetta, jota käytetään aiheenmukaisissa tai siitä korkeampien moniulotteisten relevanssien arvioinneissa. Kekäläisen ja Järvelinin (2002) mukaan saanti ja tarkkuus voidaan yleistää käsittelemään ei-binäärisiä relevanssiarvioita.

Kekäläisen ja Järvelinin (2002) mukaan korkean mittaluokan relevanssin (high-order relevances) mukaan Järvelin ja Kekäläinen viittaavat relevanssiin, joka ei perustu puhtaasti aiheenmukaisuuteen vaan muihin kriteereihin. Samalla kun siirrytään rakenteettomien tekstidokumenttien rakenteisten dokumenttien kokoelmiin, mahdollisuudet soveltaa tilanteenmukaista relevanssia tulevat paremmiksi. Vaikka on helppo lisätä realismia korkean mittaluokan relevanssikriteereihin tarjoamalla niitä järjestelmäpiirteiksi, se on testikokoelmille haaste, sillä kriteerien vaihtelevuuden ymmärtämisessä on puutteita samoin kuin todisteiden yhdistämisessä kriteereihin.

Järvelin ja Kekäläinen ovat esittäneet tutkimuksessaan (2002) kaksi tiedonhaku-järjestelmien evaluointi- eli arviointiskenaariota. Ensimmäinen on laboratoriomallia laajempi lähestymistapa tiedonhakuun, kiinnittäen huomiota erilaisiin tiedonhakustrategioihin ja taktiikoihin sekä liittymiin, jotka tukevat niitä. Tämä on (IR Interfaces Scenario) tiedonhaun käyttöliittymäskenaario. Käyttöliittymäskenaario on järjestelmien kehittäjien näkökulma. Päämäärä on ensinnäkin kehittää tiedonhakualgoritmeja tukemaan informaatioavaruuksien tutkimista ja käsitteellistämistä (dokumentit ja käsitteet) auttamaan hakupyyntöjen muotoilua. Toinen tavoite on tunnistaa, identifioida ja pistää järjestykseen relevanteja dokumentteja vaihtelevia hakuaiheita käyttäen ja kolmanneksi esittämään sellaisia dokumentteja, joita on tarkoitus tukea arvioimalla niitä korkean relevanssijärjestyksen mukaisesti. Menestysmittoihin sisältyvät käytettävyys ja monipuolisuus, yhtä hyvin kuin saanti ja tarkkuus. Itsestäänselviä komponentteja voidaan arvioida saannin ja tarkkuuden mukaan, mutta koko algoritminen osatekijä tarvitsee laajempia mittoja.

2.7 Testikokoelmien rakentaminen

Tiedonhaku-järjestelmien evaluoinnin mahdollistavat testikokoelmat vaativat testiasetelman

rakentamista tutkimusongelman ratkaisemisen operationalisointina. Tutkimusongelma voi lähteä laboratoriomallista, aineistosta, käytännöstä, tutkimuksesta, teoriasta tai luovuudesta. Operationalisoitava tutkimusongelma lähtee tutkimuskysymyksistä, joihin halutaan saada vastaus. Niiden avulla yritetään selittää tai ymmärtää tutkittavia ilmiöitä. Tutkimusongelman operationalisointi tarkoittaa tutkimusongelman koeasetelmaa, sen sisältämiä aineistoja ja muuttujia suhteessa aineistoon (Kekäläinen & Sormunen, 2003).

Empiirinen tiedonhaketutkimus käsittää kenttätutkimukset, käyttäjätetit sekä laboratoriotetit. Empiirinen tiedonhaketutkimus on kenttätutkimuksena luonteeltaan ilmiötä kuvaileva. Tarkasteltavana ilmiönä tiedonhaku on työprosessin osa. Käyttäjätetit ovat vuorovaikutteista tiedonhaun tutkimusta kontrolloidussa ympäristössä. Tiedonhaketävät voivat olla annettuja tai omia. Laboratoriotesteistä esimerkkinä olivat siis Cranfield- ja TREC-tyypin testikokoelmat.

Aikaisempien teoreettisten tutkimusten ja huomattavienkin käytännön testikokeilujen ja -kokemusten perusteella perinteiset tiedonhaun testikokoelmat eivät ole sellaisenaan riittäviä nykyisen tiedonhaketutkimuksen ja tiedonhaun opetuksen tarpeisiin. Uudenlaiset interaktiivisen tiedonhaun käyttöliittymät ja järjestelmät sekä Internet-tiedonhaku ovat tuoneet tulevalle tiedonhaun tutkimukselle ja sen mukaiselle opetukselle haasteita. Järvelinin ja Kekäläisen (2002) esittämä tiedonhaun käyttöliittymäskenaarioksi kutsuttu evaluointimalli ja Borlundin (2000) esittämä evaluointimalli tiedonhaun tutkimusta suuntaavina esimerkkeinä näyttäisivät edellyttävän moniulotteista relevanssiarviointimenetelmää. Testikokoelmia käytetään hakutehokkuuden tutkimisen vakioituna testiympäristönä. Testikokoelman perusosat eli hakuaiheet, relevanssiarvioinnit ja dokumenttikokoelma ovat hyvin määritellyjä ja staattisia toimiakseen vakioituna testiympäristönä (Järvelin & Kekäläinen 2002).

Aitoihin käyttötilanteisiin tai hakijan valitsemiin aiheisiin perustuvassa vuorovaikutteisen tiedonhaun tutkimuksessa ei voida käyttää ennalta asetettuja saantikantoja. Saanti ja tarkkuus eivät ole tällöin hakutehokkuuden ainoat mittarit. Kun käytetään simuloituja tilanteita tai tehtäviä, saantikantoja voidaan soveltaa. Onnistunut tiedonhaun evaluointi lähtee tavoitteiden ja näkökulmien määrittelystä. Työ rajataan resurssien mukaan. Evaluoinnin kriteerien käyttö edellyttää mittaria kriteereille. Evaluointiohjelman suunnitteleminen on prosessi, johon kuuluu evaluoinnin varsinainen toteuttaminen ja

tulosten analysointi (Kekäläinen & Sormunen, 2003).

Vakkarin ja Hakalan (2000) tutkimuksen tavoitteena oli analysoida, kuinka muutokset graduopiskelijoiden omissa relevanssikriteereissä liittyivät eri tehtävävaiheisiin tutkimussuunnitelma- ja koko prosessin aikana. Tutkimuksessa seurattiin 11 opiskelijaa, jotka osallistuivat tutkimusseminaarissa pro gradun tutkimussuunnitelman muodostamiseen kevätlukukauden aikana 1999. Tarkoituksena oli saada selville, kuinka opiskelijoiden omassa tutkimustehtävässään eteneminen ja tehtävän ja prosessin vaiheet vaikuttivat heidän tapansa antaa relevanssikriteereitä haetuille dokumenttien viitteille ja itse dokumenteille. Tutkimuksen viitekehystenä oli Kulthaun esittämä tiedonhakuprosessin malli, johon on yhdistetty kognitiivista psykologiaa. Tulokset ja teoria osoittivat selvästi, että prosessiluonteisen tiedonhankinnan mukaiset relevanssiarvointikriteerit haetun tiedon ja informaatiotyyppien sekä dokumenttien hyödyllisyyden relevanssiarvioinnit riippuivat järjestelmällisesti tehtäväprosessin vaiheista sekä tiedonhakijan omista sisäisistä malleista (Vakkari & Hakala, 2000).

Oman tutkimukseni menetelmän soveltamisessa testitietokantana on lähtötilanteessa Internet avoimena verkkona ja tiedonhaku-ympäristönä. Hakutuloksien joukkoon voi päätyä täten Internetistä teoriassa ja käytännössä mikä tahansa dokumentti. Hakutulosten joukkoa eli tutkittavia otoksia rajoittavat kuitenkin monet sekä teoreettiset että käytännön tosiasiat.

3 Moniulotteinen relevanssiarvointimenetelmä

Moniulotteinen relevanssiarvointimenetelmä on Sormusen (2002a, 2006) kehittämä väline esittää moniulotteista relevanssidataa ja soveltaa tätä tiedonhaun ja tiedonhakutehtävien testikokoelmien rakentamiseen. Sormunen on täydentänyt moniulotteisen relevanssiarvointimenetelmien yksityiskohtia uudessa menetelmäartikkelissa (2006). Testikokoelmien yleinen tavoite on tukea interaktiivisten tiedonhakuprosessien tutkimusta ja tiedonhakukäyttäjien ymmärtämistä käyttäjälähtöisten relevanssikriteerien tunnistamisella kontrolloiduilla kokeilla. Testikokoelmatutkimusten avulla edistävän myös tehokkaiden tiedonhaun oppimisympäristöjen kehittämistä.

Moniulotteista relevanssiarvointimenetelmää on sovellettu perinteisiin dokumentteihin (Sormunen, 2006), mutta sen monimuotoiset tarkastelevat tutkimukselliset ominaisuudet on arvioitu tulevan paremmin esille www-aineistoissa. Tämä tutkimus antaa osaltaan tietoa menettelytavan soveltamisesta juuri Internet-aineistoissa.

Moniulotteisen relevanssiarvion lähtökohta on määritelty hakuaihe konteksteineen. Konteksti voi määrittää esimerkiksi minkätyyppisiä hakijoita hakuaiheen oletetaan kiinnostavan ja minkälaisia yleisiä ehtoja heillä on hakutuloksille. Nämä määräytyvät hakijan ajateltujen työtehtävien sekä informaatiokäyttämisen ja -vaatimusten mukaan. Tavoitteena on hakuaiheeseen liittyvän informaatioavaruuden kuvaus. Arvioitavana on (information space), joka tarkoittaa aiheperustaista informaatioavaruutta. Alustavilla hauilla tutustutaan hakuaiheeseen liittyviin dokumentteihin. Relevanttien hakutulosten perusteella kartoitetaan dokumenttien sisältöä aihe riippuvien teemojen ja aihe riippumattomien ominaisuuksien (non-topical attributes) valinnalla. Löydetyt ja valitut dokumentit, sivut, niiden kokonaisuudet tai niiden osat teemoina, jotka ovat muutettu teemoista sisältöä kuvaaviksi käsitteiksi, arvioidaan numeerisilla relevanssiluvuilla tai -asteilla (degree of theme relevance) moniportaisen asteikon mukaan eri tasoille niiden relevanssiasteen mukaisesti. Kiinnostuksen kohteena on dokumenttien joukko, tietokannan informaatio sisältö suhteessa hyvin määriteltyyn hakutehtävään. Aihepiirin sisällönanalyysi käsitellään tutkielmani hakuaiheissa kuitenkin tutkimusmotivaation, ajan ja -resurssien puitteissa (Sormunen, 2002a).

3.1 Yksiulotteisista moniulotteisiin relevanssikuvauksiin

Sormunen (2002b) esittää, että yksiulotteisista relevanssiarvioista halutaan siirtyä aineistolähtöisiin, dokumenttien sisältöjen ja ominaisuuksien kirjoa paremmin kuvaaviin relevanssikuvauksiin. Moniportaiset relevanssiarvot, joilla voidaan tarkemmin arvioida dokumenttien relevanssin astetta, kertovat informaatio sisällöstä enemmän kuin jyrkkä binäärinen arviointi, mutta nekään eivät ratkaise sisällöllisten päällekkäisyyksien (overlaps) ongelmaa. Moniulotteiseen relevanssiarvointiin kuuluvat sekä hakuaiheen eri osa-alueiden eli teemojen että dokumenttien aihe riippumattomien ominaisuuksien arviointi.

Moniulotteisuuden perusidea on se, että arvioinnissa otetaan huomioon useampi kuin yksi

sisällöllinen relevanssikriteeri tai dokumentin ominaisuus. Kun yksiulotteisessa relevanssiarvioinnissa dokumentin aiheenmukaista relevanssia (topicality) edustaa yksi luku, moniulotteisessa relevanssiarvioinnissa esitetään lukuja, joista kukin edustaa yhtä hakuaiheesta tunnistettua teemaa. Dokumentin ominaisuuksilla viitataan tässä dokumentin aiheiriippumattomiin piirteisiin, kuten tekstin kieleen, dokumentin tyyppiin tai laatuun, julkaisuvuoteen tms. Moniulotteinen relevanssidata voidaan esittää taulukkona, jossa kukin sarake edustaa yhtä relevanssikriteeriä ja kukin rivi yhtä dokumenttia. Moniulotteisesta relevanssidatasta voidaan tuottaa myös yksiulotteinen "kokonaisarvio" jokaista dokumenttia kohden. Kokonaisarvio voi olla esimerkiksi keskiarvo kaikkia sisällön eri teemoja kuvaavien relevanssilukujen relevanssiarvoista (Sormunen, 2006).

Aidon tiedonhakijan relevanssiarviot ja -kriteerit ovat dynaamisesti muuttuvia ja vahvasti subjektiivisia. Relevanssiarvioija pyrkii mahdollisimman suureen objektiivisuuteen mutta silti relevanssi on tehtävästä ja prosessista riippuva arvioijan konstruktio. Hakuaiheen taustalla oleva todellinen tai kuviteltu työ tms. tehtävä vaikuttaa dokumentin relevanssiarvioinnin kriteereihin. Mitä jäsentyneempi tehtävästä nouseva tiedontarve on, sitä paremmin relevanssiarvioija kykenee hahmottamaan dokumentin relevanssin. Nämä haasteet ja rajoitukset koskevat niin perinteistä kuin moniulotteista relevanssiarviointiakin.

3.2 Relevanssikorpus hakutulosten eli dokumenttien ja niiden mukaisten teemojen kuvaajana

Tämän tutkimuksen relevanssikorpuksella on erityinen esitysmuoto, joka sopii moniulotteisen relevanssidatan esittämiseen. Mitä rikkaampi relevanssikorpuksen (relevance corpus) tarjoama kuvaus on, sitä tarkempi kuva on aihepiirin informaatioavaruudesta eli kokonaisuudesta ja yksittäisen dokumentin roolista kokonaisuuden osana.

Tutkimuksen aineistoa kerätään hakutehtävän perusteella. Taustaoletuksena on, että jokainen saantikannan dokumentti sisältää jonkinlaista relevanttia informaatiota määriteltyyn hakutehtävään liittyen. Informaatioavaruudesta haettuja relevanttien dokumenttien joukkoa kutsutaan saantikannaksi (topical information base). Tämä staattinen relevantin informaation mahdollisuus arvioidaan ja esitellään

relevanssikorpuksessa, johon on valittu päällekkäisyyksien poistamiseksi. Nämä ovat relevanssiarviointimallin omia termejä, jotka kuvaavat mallin eri osatekijöitä ja käsitteitä moniulotteisessa relevanssiarviointiprosessissa (Sormunen 2002a).

Moniulotteiseen relevanssiarviointimenetelmään kuuluvat aiheriippuvien ulottuvuuksien lisäksi aiheriippumattomien ominaisuuksien määritelmät, jotka aiheriippumattomina määräytyvät olennaisesti tutkimuksen dokumenteista tai tietokantatyypistä. Www-kokoelman dokumentit voivat olla kirjoitettuna erilaisilla kielillä, tulla erilaisista lähteistä jne. Erityisesti Internet-ympäristö painottaa aiheriippumattomien relevanssikriteerien tärkeyttä (Sormunen 2002a).

Moniulotteisen relevanssiarviointimallin neljäs taustaoletus on, että relevanssikorpuksen täytyy tukea dokumenttien sisällöllisten päällekkäisyyksien (overlaps) analyysia. Ei riitä, että arvioidaan dokumentin potentiaalinen relevanssi kaikenkaikkiaan, vaan pitää myös ilmaista, kuinka dokumentti täyttää osuutensa aiheeseen perustuvassa informaatio-avaruudessa. Aiheen informaatioavaruus on käsite, joka korostaa, että kehittyneissä testikokoelmissa yksittäisen dokumentin käyttökelpoisuutta on arvioitava suhteessa muuhun saantikantaan kokoelmapohjaisessa kontekstissa (Sormunen, 2002a).

Moniulotteisen relevanssiarviointimenetelmän tutkittavia kysymyksiä ovat muun muassa teemojen ja ominaisuuksien valinnan logiikka eri aineistoissa. Hyvin olennaista tutkimuksen onnistumisen kannalta on arvioijalähtöinen teemojen valinta ja niiden relevanssiarviointi. Teemat kuvaavat informaatiotutkimuksen rakennetta eri aineistoissa. Tutkimuskysymyksissä voidaan päästä myös tuloksellisuusmittarien kehittämiseen. Esitetyssä moniulotteisen relevanssiarvioinnin menetelmässä voidaan arvioida aiheriippuvien ominaisuuksien lisäksi myös aiheriippumattomia ominaisuuksia, mikä on tärkeä piirre heterogeenistä web-aineistoa tarkasteltaessa.

Moniportaiselle asteikolle sijoittaminen täytyisi tapahtua relevanssiarviointimallin määriteltyjen ohjeiden mukaisesti. Tämän ohjekäytännön mukaan dokumentit haetaan, analysoidaan, teemoitellaan, tarkistetaan ja raportoidaan tietyin säännöin tietyssä järjestyksessä. Moniulotteinen relevanssiarviointimalli moniportaisine arvioineen on kuitenkin tämän tutkielman sovelluskäytössä arvioinnin osalta grounded theorya sovellettaessa riittämätön "tieteellisen tarkkoihin perusteluihin". Anselm Straussin

mukainen grounded theoryn käyttö ainakin suuntaa antavana menetelmänä on perusteltua informaatiotutkimuksenkin laadullisena menetelmänä, sillä teemat voidaan mielekkäästi valita vasta kun nähdään mitä dokumentit kokonaisuudessaan kertovat hakuaiheesta (Lowe 1996).

Grounded theory on laadullinen, aineistolähtöinen menetelmä, jota käytetään yhteiskuntatieteessä, lääketieteessä ja myös informaatiotutkimuksessa. Grounded teorian sanotaan muodostavan tietoa sekä induktiivisesti että deduktiivisesti. Tavoitteena on teoreettinen saturaatio, kertomuksen muodostamiseen ei tarvita enää uusia käsitteitä. Grounded teorian lähtökohta on aihetta koskevan ilmiön selittäminen niin monella käsitteellä kuin mahdollista (Lowe, 1996).

Tämän voisi tulkita moniulotteisen relevanssiarviointimallin indeksointikielen luomiseksi. Aiheperustaisia käsitteitä verrataan toisiinsa jatkuvasti. Moniportainen relevanssiarviointi tarkoittaa moniulotteista relevanssiarviointia, jossa hakutehtävän kontekstisidonnaisia aiheiriippuvia aiheulottuvuuksia kuvataan teemoilla. Teemat muodostavat hakuaiheulottuvuuden tietosisällön sisällysluettelon otsikot, jotka arvioidaan aiheen käsittelytapojen ja tekstisisällön laadullisen kattavuuden osalta hakutehtäväkohtaisen aineiston laajuudessa.

3.3 Moniulotteisen relevanssidatan menettelytapa

Tässä tutkielmassa moniulotteista relevanssiarviointimallia ja sen soveltuvuutta testikokoelmana toimivaan www-aineistoon arvioidaan ja käsitellään neljän hakutehtävän ja niiden henkilöhakuaihetehtävien perusteella. Mallin soveltuvuutta tarkastellaan siis tapauskohtaisesti, mutta tuloksena voi olla myös yleistettävissä olevia merkityksiä. Internet ja sen heterogeenisyys sekä monipuolisuus tuovat tarkasteluun kattavan joukon sisällöllisiä ja ulkoisia ominaisuuksia. Relevanssiarviointiprosessia suhteessa esimerkkietiedonhakutehtävään kuvaa lisäksi mm. termi sisällönanalyysi tai grounded theory lukuisine ala- ja yläkäsitteineen.

Moniulotteinen relevanssiarviointi Internet-aineistossa on samankaltaista kuin muissakin aineistotyypeissä. Internet sisältää moniulotteisen relevanssiarvioinnin kannalta monenlaista sisällöllistä tietoa sekä havainnollistavaa lisäinformaatiota, jonka

määrittelemiseen ja arviointiin on merkityksellistä löytää relevanssikriteereitä. Kirjallisuutta tai sanoma- tai aikakauslehtiä, niiden sisältöä ja yleisilmettä voitaisiin arvioida samalla menetelmällä. Seuraavassa on moniulotteisen relevanssiarviointimenetelmän mukainen kuvaus, miten junaonnettomuudesta kertovia sanomalehtiaineistoja tai tiedonhakukäyttäjyymistä koskevia tutkimuksia voitaisiin jakaa sisältöä kuvaaviin teemoihin:

Uutisaineistoissa teemat eli aiheriippuvat ominaisuudet (topical attributes) voivat liittyä eri vaiheisiin raportoidun uutisen tapahtumissa. Esimerkiksi teemat, jotka liittyivät aihekantaan "junaonnettomuus Jyväskylässä", voidaan jakaa teemoihin: todellinen onnettomuus (mitä tapahtui), seuraukset (kuolleet, loukkaantuneet), henkilökohtaiset kokemukset (haastattelut), onnettomuustutkinta, oikeudenkäynti, yleiset rautatieturvallisuusartikkelit, joita onnettomuus synnytti. Tutkimusjulkaisujen kokoelmassa asianmukaiset teemat voisivat olla erilaisia. Esimerkiksi aihe "hakukäyttäytyminen Webissä" voitaisiin jakaa teemoihin peruskäsitteet, mallit ja teoriat, tutkimusmenetelmät, empiiriset tulokset ja yleiset tutkimusympäristöt. Jokaisen teeman sisällä dokumentin relevanssi voidaan arvioida käyttämällä moniportaista asteikkoa. Korkein relevanssin taso voidaan varata dokumenteille, jotka sisältävät kaiken olennaisen informaation teemasta ja 0 dokumenteille, jotka eivät käsittele teemaa ollenkaan. Teemarelevansseja voidaan käyttää löydettyjen dokumenttien sisällöllisten päällekkäisyyksien vertailuun ja havaintojen tekemiseen siitä, kuinka käyttäjä onnistuu löytämään aihepiirin informaatioavaruuden eri osat (Sormunen, 2002a).

Testikokoelman saantikanta relevanssikorpuksena esitetään teemoissa. Niillä on sama rooli, kuin "instansseilla" interaktiivisessa TRECissä, mutta ne ymmärretään tässä yleisemmin. Teemoille voidaan antaa erilaisia alakohtaisia rooleja. Teemojen valinnan idea kannattaa ankkuroida tiettyyn hakutehtävään ja aiheenmukaisen informaatiokannan sisältöön. Toisaalta sellaiset aiheriippuvat näkökulmat, joista ei ole keskustelua dokumenteissa, voidaan jättää teemoista pois (Sormunen, 2006).

Jokaisen teeman sisällä dokumentin relevanssi voidaan arvioida käyttämällä moniportaista asteikkoa. Korkein relevanssin taso voidaan varata dokumenteille, jotka sisältävät kaiken olennaisen informaation teemasta esim. arvolla 3 ja 0 dokumenteille, jotka eivät käsittele teemaa ollenkaan. Teemarelevansseja voidaan käyttää löydettyjen dokumenttien

sisällöllisten päällekkäisyyksien vertailuun ja havaintojen tekemiseen siitä, kuinka käyttäjä onnistuu löytämään informaatiokannan eri osat. Teemakohtaista relevanssidataa voidaan soveltaa vuorovaikutteisen TRECin tapaan hakutehokkuuden mittarina. Teemarelevanssien perusteella voidaan laskea myös yleinen aiheenmukainen relevanssiarvo kullekin dokumentille saantikannassa ja suorittaa perinteisiä eräajoina tehtäviä tiedonhaketestejä (Sormunen, 2006).

Hakutehtävän mukainen aineistojen aiheriippuva arviointi perustuu aihealueen sisällönmukaisiin teemoihin eli sisällön aiheiden erittelemiseen sisällönanalyysinä. Hakutehtävän aiheenmukaiset teemat pyritään kartoittamaan kokonaisuudessaan valmiin testikokoelman laajuudelta, jolloin niiden esiintymät voidaan jokaisen hakutuloksen osalta havainnoida. Aiheriippuvat teemat on tarkoitus lajitella pää- ja sivuteemoihin. Relevanssi arvioidaan teemakohtaisesti. Arvioitavan dokumentin teema voi saada siis teemakohtaisesti lajiteltuna joko arvon 0 eli ei-relevantti tai relevanssiarvon 1, 2 tai 3 relevanssiasteen merkittävyyden perusteella.

Korkeinta relevanssitasoa 3 ei käytetä ensimmäisellä arviointikierröksellä. Tämä vaihe voidaan nähdä indeksointiprosessina eli tässä tapauksessa elektronisten dokumenttien sisällönkuvailuna. Voi osoittautua, että dokumentit sisältävät kiinnostavia teemoja, joita ei voi kattaa alkuperäisillä teemoilla, tai alkuperäisten teemojen määritelmiä täytyisi vaihtaa tai laajempi teema voi jakautua kahdeksi tarkemmin eriytyväksi teemaksi. Kaikki dokumentit täytyy uudelleenarvioida mahdollisten uusien tai muokattavien teemojen takia.

Kaikki dokumentit, jotka kuuluvat toiseksi korkeimmalle relevanssitasolle arvolla 2 tietyn teeman sisällä, vertaillaan tarkoituksena tarkastaa, voidaanko yhtään dokumenttia nostaa korkeimmalle relevanssitasolle 3, jonka voitaisiin nähdä sisältävän kaiken olennaisen informaation teemasta saantikannan kontekstissa (Sormunen, 2006).

Teemojen relevanssitasojen määrittely on seuraava:

- 3:** Erittäin relevantti: teemaa käsitellään kattavasti / perusteellisesti.
- 2:** Melko relevantti: joitakin teemaan liittyviä asioita käsitellään.
- 1:** Marginaalisesti relevantti, viittaa teemaan, mutta ei sisällä siitä olennaista
- 0:** ei-relevantti, teemaa ei käsitellä lainkaan.

3.4 Aiheriippumattomien ominaisuuksien merkitykset

Internet-aineistossa

Internet-aineistot ovat luonteeltaan hyvin heterogeenisiä, eivätkä juuri millään kriteerillä arvioiden yhtenäisiä, joten juuri tällä voi perustella sen, että Internet-sivustot ovat täynnä aiheriippumattomia ominaisuuksia.

Relevanssikorpuksen pitäisi sisältää informaatiota myös dokumenttien aiheriippumattomista ominaisuuksista, jotka voivat liittyä tiedonhakijalle hyödyllisiin tai vastakohtaisesti jopa merkityksettömiin relevanssikriteereihin. Oletuksena on, että aiheriippumattomat dokumentin ominaisuudet liittyvät tyypillisesti kriteereihin, joiden perusteella käyttäjät hylkäävät aihepiiriltään relevantteina pitamiään hakutuloksia. Käyttäjä ei esimerkiksi kykene käyttämään dokumentteja vieraan kielen takia, vaikka asiasisältö on relevantti. Hänen ei myöskään tarvitse arvostaa dokumenttia riittävästi voidakseen käyttää sitä tehtävänsä (Sormunen, 2002a).

Aiheriippumattomat ominaisuudet puolestaan viittaavat dokumenttien muihin ominaisuuksiin, jotka eivät tavallaan muuta dokumentin sanomaa. Aiheriippumattomia ominaisuuksia voivat olla esimerkiksi kieli, kirjoittaja, dokumentin ikä tai lähde. Aiheriippumattomien ominaisuuksien merkitystä tarkastellaan suhteessa aihe relevanssiin ja aiheriippumattomien laajuutta ja moni-ilmeisyyttä tarkastellaan tässä tutkielmassa varsinkin Internet-ympäristössä ja sen erityispiirteissä.

3.5 Aikaisemmat Internet-aineiston relevanssikriteeritutkimukset viitekehyksinä relevanssiarvioinneille

Aiemmalla käyttäjälähtöisellä Internet-tutkimuksella on tarkoitus selvittää, minkälaisia kriteereitä käyttäjillä on ollut, kun he ovat empiirisissä kokeiluissa arvioineet www-aineiston relevanssia hakutehtävien puitteissa. Näitä tutkimuksia on englanninkielisessä tiedemaailmassa tehty jo huomattavan paljon. Tässä pääluvussa esimerkkinä Internet-tutkimuksista on kolme käyttäjälähtöistä www-sivujen laatuun ja hyödyllisyyteen perustuvaa relevanssikriteereitä esittelevää tutkimusta. Näiden tutkimusten motivaationa on

ollut kerätä parempaa ymmärrystä niistä ominaispiirteistä, jotka tekevät www-dokumenteista hyödyllisiä tiedonhakijalle.

Rieh (2002) on tutkimuksessaan selvittellyt tekijöitä, jotka vaikuttavat arviointeihin www-aineiston informaation laadusta ja koetusta auktoriteetista ja näiden vaikutuksesta aineistonvalikointi-käyttäytymiseen. Lähtökohta Riehin tutkimukselle oli, ettei Internetissä ole laatukontrollia. Lisäksi on saatavilla useita vaihtoehtoisia resursseja ja käyttäjien on valittava informaatio-objekti usean vaihtoehdon joukosta. Kuitenkin hakijat käyttivät myös tietokantoja ja kirjastojärjestelmiä, jolloin käyttäjät olettivat, että heidän hakemansa tieto oli jo tiedontuottajien, tietokantojen ylläpitäjien ja kirjastonhoitajien valmiiksi arvioimaa. Riehin mukaan web on suhteellisen uusi resurssiympäristö, jossa on erilaiset säännöt ja kriteerit, kuin mitä käytetään painetussa kokoelmassa. Pitkäänkin perinteisillä informaatioaloilla työskennelleet voivat olla webin käytössä noviiseja.

Riehin (2002) vastausaineiston groundedteoreettinen ja induktiivinen tutkimusmateriaalin analyysi tähdensi kahta vaihtoehtoista arviointityyppiä, eli ennustavaa arviointia (predictive judgment) sekä evaluoivaa arviointia (evaluative judgment). Ennustava arviointi tapahtui siinä vaiheessa, kun henkilö siirtyi sivulta toiselle. Menetelmän oletuksena tutkimuksessa oli, että käyttäjät voivat tunnistaa ja keskustella informaatio-objektien luonteenomaisista piirteistä, jotka vaikuttavat heidän arviointeihinsa tiedon laadusta ja kognitiivisuudesta. On mahdollista koota relevanssikriteereitä kehittämällä luokitusta, joka pohjautuu tiedonhakijoiden omiin termeihin.

Tutkimuksen tuloksena oli, että web on vuorovaikutusympäristö, missä tutkija pystyi havainnoimaan ihmisten päätöksentekokäyttäytymistä ennustavasta vaiheesta evaluoivaan vaiheeseen jatkuvassa prosessissa. Ennustava arviointi johti uuden web-sivun aukaisemiseen, mahdollistaen henkilöitä tekemään evaluoivan arvioinnin, jossa he muodostivat web-sivun arvot ja mieltymyksensä niihin. Kun sivu ei vastannut heidän odotuksiaan, henkilöt päättivät etsiä uuden tai mennä edelliselle. Kun ennakoiva arviointi onnistui, henkilöt päättivät joko käyttää informaatiota tai jäädä sivulle. Tehtävästä, sen luonteesta eli aiheesta riippui, missä määrin aineiston valikoinnissa kiinnitettiin huomiota affektiivisiin tekijöihin, ennustaviin arvioihin tai tiedolliseen eli tietoiseen valikointiin.

Riehin (2002) tutkimuksen kaksi pääkäsitettä laadusta ja auktoriteetista ovat moniulotteisia

käsitteitä, jotka muodostuvat informanttien vastauksista, niiden erilaisista termeistä ja sanoista. Informaation laatua kuvailtiin sanoilla hyvä, tarkka, ajankohtainen, hyödyllinen, tärkeä, hyvä homma, huono, parempi, täydellinen, hieno, mukava, mahtava, paras, moitteeton, ihana, uskomaton, fantastinen, taiteellinen, hyvin ylläpidetty, hyvin kehitetty. Hyödyllisyysulottuvuuden arvioita edustivat termit hyödyllinen, käyttökelpoton, vaikea käyttää, informatiivinen, helpottava, ei auta, ei ole ymmärrettävä, ei ole paljon käytettävää tai käsiteltävää. Kognitiivinen pätevyys voidaan avainsanojen mukaan jakaa kuuteen fasettiin, jotka ovat luotettavuus, käyttövarmuus, tieteellisyys tai oppineisuus, uskottavuus, virallisuus ja auktoriteetti. Esimerkiksi oppineisuutta lausuttiin termeillä koulusivistynyt, akateeminen, ammatillinen tai biologinen.

Kun käyttäjät tekivät ennustavia arvioita ja päätöksiä, he ryhtyivät toimeen valitsemalla tietyn web-sivun, katsoen sen läpi ja siirtyivät tekemään evaluoivia arvioita. Mallit, jotka nousivat esille, osoittavat, että henkilöiden evaluoivat arvioinnit pohjautuivat enemmän informaatio-objektien ominaispiirteisiin ja lähteeseen mieluummin kuin henkilön omaan tietoon ja tilannetekijöihin. Informaatio-objektien ominaispiirteisiin he mainitsivat kuuluvaksi sisällön, grafiikan, sivun rakenteen tai organisaation ja informaation tyyppin suhteellisesti ottaen useammin kuin muut kriteerit. Yksi tutkimuksen löydöistä oli, että web-käyttäjät tekevät arvioita samanaikaisesti, kun tekevät hakuja.

Lähtötilanteesta Rieh (2002) pystyi laajentamaan tiedollisen auktoriteetin sekä laadukkuuden kahta pääkäsitettä ja relevanssikriteereitä kuvaavaa mallia seuraavilla näkökulmilla: laadun ja auktoriteetin fasetit monimuotoistuivat, webinformaation tunnusmerkit pystyttiin erottelamaan kahteen kategoriaan, joita ovat informaatio-objektien luonteenomaiset piirteet, kuten tyyppi, otsikko, sisältö, rakenne, esitystapa, grafiikka, toiminnallisuus sekä lähdekoneen ylläpitäjien tunnusmerkit kuten mm. URL-tunnus, tyyppi, maine, kirjoittaja- tai tekijäsuositukset. Lisäksi merkityksellistä on tietojärjestelmään ja tiedon substanssialueeseen liittyvän tiedon erottaminen sekä uusien vaikuttavien tekijöiden löytäminen laatuun ja auktoriteettiin, kuten tilanne, hakutulosten järjestys, yleinen oletus ja arviointifasettien kriteerien erilaisuuden painottaminen hakutehtävien luonteen mukaan.

Crystal & Greenberg (2006) arvioivat tutkimuksessa terveystiedon hakijoiden ja käyttäjien web-hauissa soveltamia relevanssikriteereitä. Hakujen kohteena olivat

kokoteksidokumentit, jotka käsittelivät ympäristön vaikutusta terveyteen. Käyttäjää pyydettiin hiirellä osoittamalla tunnistamaan käyttämänsä kriteerit relevanssiarvioinnissa. Avainkriteerit, jotka tulosten muodossa tunnistettiin niiden esiintymien frekvenssien perusteella, olivat tutkimus, hakuaihe, laajuus, tieto, vaikutus, tunteet, web-aineiston ominaisuudet ja auktoriteetti tai sivuja julkaiseva henkilö. Tutkimuksessa havaittiin eniten jakaantuvat dokumenttien relevanssikriteerit sekä todettiin, että tiedonhakujärjestelmien ja niiden hakukoneiden tulisi koota laajempaa kontekstisidonnaista aineistoa. Vuorovaikutus tiedonhakujärjestelmien hakutulosten kanssa on kaksivaiheinen prosessi. Ensin käyttäjät määrittelevät käyttöliittymäpiirteet, jotka käsittävät tietyt metadataelementit, kuten dokumentin otsikon, päivämäärän, jolloin dokumentti on päivitetty tai julkistettu, dokumentinkuvauksen, aiheen tai asiasanat. Tämä on Riehin (2002) tutkimuksen mukaisesti ennakoivaa arviointia, jossa yritetään ennustaa dokumentin hyödyllisyyttä. Evaluoivaa arviointia ovat käyttäjät kontekstin tilanteenmukaiselle tehtävälle relevantit dokumentit. Tiedonhaku tutkimuksen haaste on sellaisten tiedonhakujärjestelmien suunnitteleminen, joka tukee tehokkaasti käyttäjiä näissä molemmissa vaiheissa. Empiiriset tutkimukset voivat tarjota perustan tälle suunnittelun haasteelle vahvistamalla käyttäjämalleja ja käyttäjien mieltymyksiä.

Tombros, Ruthven ja Jose (2005) arvioivat tehtävälähtöisiä www-sivujen arviointikriteereistä. Tutkittavat viittasivat www-sivujen ominaisuuksiin, joita he pitivät olennaisina päättäessään sivujen hyödyllisyydestä simuloitujen tiedonhakutehtävien kontekstissa. Tutkimuksen tuloksena on asetelma käyttäjälähtöisiä kriteereitä, joita hakijat käyttivät arvioidessaan eri tyyppin tehtävien www-sivujen hyödyllisyyttä. Tutkijoiden peruslähtökohta oli tiedonhaku tutkimuksen relevanssikäsitys aihe relevanssina (topical-appropriateness or topicality) erityisesti käyttäjärelevanssin (user-utility) näkökulmasta. Tombrosin, Ruthvenin ja Josen tutkimuksessa painotettiin Barryn (1998) relevanssitekijöitä, joita ovat sisältö, henkilökohtaiset tekijät, jotka perustuvat lukijan ja informaation väliseen merkitykseen, sekä laatutekijät, jotka perustuvat dokumentin lähteeseen, esittämiseen, maineeseen sekä ajankohtaisuuteen. Kolme tutkijoiden tiedonhakutehtävää perustui erilaisiin tiedonhankinnallisiin oletuksiin. Nämä olivat taustatiedon hakeminen sekä toisena mahdolliseen päätökseen tiedonhakujen aikana johtava tiedonhaku. Kolmas tehtävä oli laatia lista kiinnostavista asioista, joita tiedonhakija haluaisi tehdä viettäessään viikonlopun Kiotossa.

Tehtävälähtöiset aiheen lisäksi sovelletut relevanssikriteerit pystyttiin lajittelemaan kolmeen fasettiin. Aiheenmukaiset relevanssikriteerit sisällön osalta olivat sen merkitys tekstinä ja numeroina. Aiheenmukaisia relevanssikriteereitä olivat lisäksi otsikot, hakutermit sekä aiheen käsittelyn laajuus. Kolmantena Tombros, Ruthven ja Jose esittelevät lisäksi laajan joukon aiheriippumattomia ominaisuuksia, jotka ovat merkittäviä tuloksia 24 tutkittavan hakuistunnoissa.

Näistä kolmesta tutkimuksesta ainoastaan tässä www-aineiston relevanssikriteeritutkimuksessa käsitellään hyvin tarkasti prosentteina ja frekvensseinä aiheriippumattomia fasettirakenteita. Fasetit ovat yläkäsitteitä niiden mukaisia asioita esitteleville alakäsitteille. Sama rakenteellisuus voidaan nähdä moniulotteisen relevanssiarvioinnin teemarakenteelle samoin kuin aiheriippumattomien ominaisuuksien kuvaamiselle.

Käyttäjälähtöiset aiheriippumattomat ominaisuudet (Tombros & Ruthven & Jose, 2005) ovat rakenne, laatu, kuvat, sekä fyysiset ominaispiirteet. Rakenneominaisuuksiin liittyvät taitto, linkit, linkkien laatu ja sivun tausta. Laatuominaisuuksia, jotka sivuavat samalla läheltä aihe relevanssia, ovat käsittelyn laajuus ja tyhjentyvyys sekä tekijä tai muu lähdetieto sekä kuvat. Aiheriippumattomia ominaisuuksia edelleen ovat ajankohtaisuus, yleinen ”heti näkyvä” laatu, sisällön luettavuus tai sivun virheellisyydet. Fyysisiä ominaisuuksia tai olomuotoja olivat seuraavat tapaukset: sivua ei löytynyt, sivun sijainti, sivu oli jo nähty, kieli, tiedostotyypit, tiedostokoko, yhteysnopeus ja julkinen tai rekisteröity palvelu.

Menetelmien rajoituksina voidaan pitää osallistujien relevanssikriteerien kuvaamisen subjektiivista luonnetta. Tombros, Ruthven ja Jose (2005) väittävät, että webin ominaispiirteet liittyvät lähes suoraan niiden hyödyllisyyteen. Aikaisemmin nähty tai koettu ominaisuus oli helppo tunnistaa ja sanoa päteväksi. Lisäksi hakutehtävät, joita käytettiin, eivät olleet todellisia, eivätkä heijastaneet osallistujien omia tiedontarpeita. Tämä on voinut vaikuttaa 24 tutkittavan tiedonhakijan www-sivujen arviointiin.

4 Tutkimuksen toteutus

Moniulotteisen relevanssiarviointimenetelmän tavoitteena tässä tutkimuksessa on esittää neljän eri henkilötiedonhakutehtävän hakutulosten mukaiset relevanssiarvioinnit, joissa on

tarkoitus perustella menetelmän käyttökelpoisuutta Internet-aineiston relevanssiin perustuvassa sisällönanalyysissä sekä aiheriippumattomien ominaisuuksien määrittelemisessä. Toinen tavoite on esittää dokumenttien aiheriippuvilla teemoilla relevanssikorpusten mukaista sisältöä selkeämmin relevanteimmiksi määritellyillä käsitteillä lyhyenä kuvauksena. On tarkoitus selvittää erityisesti www-aineiston erityispiirteitä aiheriippumattomina ominaisuuksina, jotka vaikuttavat sivujen korkean tai matalan kokonaisrelevanssin määritelmiin. Demonstroinnin ja aiheriippumattomien ominaisuuksien esittäminen lukijan kannalta ymmärrettävässä muodossa on tutkijan kannalta haastava tehtävä.

Tämän tutkielman moniulotteisen relevanssiarviointimenetelmän sovelluksena Internet-aineistoissa tutkitaan hakutehtäviä henkilöhakuihden perusteella. Hakutehtävät ovat Lourdesin Madonna -ilmiö Ranskassa vuonna 1858, poplegenda Madonna Ritchie 2000-luvulla, poliitikko Heidi Hautalan ura avoimessa Internetissä sekä lastenkirjailija ja -kuvittaja Mauri Kunnas avoimen Internetin valmiissa Mauku-viitetietokannassa informaatiotutkimuksen tietokantaprojektin vuoden 1997 harjoitustyössä.

Tutkimusongelmat määrittyvät henkilötiedonhakutehtävien mukaisesti sekä kolmen viitekehyksenä esiteltävän Internet-tiedonhakututkimuksen tulosten relevanssikriteerien perusteella. Kolme tämän tutkimuksen henkilötiedonhakutehtävistä suuntautuu avoimeen Internettiin. Mauri Kunnaksen tiedonhakutehtävä kohdistuu valmiiseen viitetietokantaan Internet-ympäristössä. Tarkoituksena on moniulotteisen relevanssimenetelmän muodossa harjoitella kolmella hakutehtävällä moniulotteisen relevanssidatan tuottamista ja neljännen tiedonhakutehtävän aiheenmukaisesti esittää kokonainen relevanssikorpus.

Viitekehyksenä Internetissä sijaitseviin aineistoihin voidaan käyttää aikaisempia tiedonhakijoiden ja -arvioiden tärkeinä pitäminä relevanssikriteereitä. Dokumenttien teemaperustaisia sisällön relevanssiarviointeja lähestytään groundedteoreettisesti relevanssiarvioijan ja tiedonkäyttäjän näkökulmista. Metodin sisällönanalyysissä yksi kerrallaan tarkasteltavien dokumenttien aiheriippuvien ja aiheriippumattomien ominaisuuksien kartoittaminen kuvataan hakutuloksia mallintavaan relevanssikorpukseen.

4.1 Aineiston valinta

Tiedonhakuaiheista Madonna Ritchie on ollut ensimmäinen mieleenjuolahtanut kiinnostava hakuaihe. Olen nuoresta lähtien pitänyt hänen musiikistaan ja kuunnellut hänen musiikkituotantoaan, sen kehittymistä, monipuolisuutta ja sanoitus- ja sävellyssuuntauksia aina ihailen. Olen ihailut hänen ja avustajaryhmiensä tanssillista esiintymistä näkemissäni konserteissa sekä niiden toteuttamisen tyylejä. Myös Madonnan yksityiselämässä ja elokuvissa näkyy hänen pettämätön tyylitajunsa. Arvostukseni on vain lisääntynyt Madonnan laajentaessa taiteellista tuotantoaan lastenkirjallisuuteen. Hänen neljä ilmestynyttä lastenkirjaansa ovat lumoavia. Madonna Ritchien oman tyttären nimi on Lourdes. Madonnalla sekä hänen nykyisellä aviomiehellään Guy Ritchiellä on poika Rocco. Madonna ja Guy Ritchie ovat juuri lokakuussa 2006 adoptoineet pojan Malawista. Tiedonhakutehtävänä Madonna Ritchien osalta on hänen elämänsä musiikin, elokuvien ja muuhun häneen liittyvän julkaisemisen osalta Internetissä 2000-luvulla.

Etsiessäni tietoa Madonna Ritchiestä Internetistä, löysin sivuja Lourdesin Madonnasta ja 14-vuotiaasta ranskalaisesta 1800-luvulla eläneestä tytöstä nimeltä Bernadette Soubirous. Hänen pienessä ranskalaisessa maalaiskylässä Lourdesissa kokemansa madonnailmiöt aiheuttivat minulle ahaa-elämyksen. Madonna-hakuaiheiden tulokset ovat pääasiassa englanninkielisiä. Tiedonhakulähtöisenä tehtävänä on selvittää Bernadette Soubirouksen elämää vuonna 1858 Lourdesissa sekä Lourdesin Madonna-ilmion käsittelyä Internet-ympäristössä.

Tiedonhakutehtävinä olevat Heidi Hautala ja Mauri Kunnas ovat tulleet mukaanotetuiksi tutkimukseeni vasta vuoden 2006 alussa. Nämä suomalaiset tunnetut henkilöt täydentävät relevanssikriteerien käsittelyä ja ovat suomalaisina hakuaiheina paremmin luettavissa pääasiassa suomenkielisten lähteidensä ansiosta. Heidi Hautalan osalta hakutehtävänä on hänen poliittinen ja pitkä, laaja-alainen uransa. Mauri Kunnaksen osalta määrittelen aiheiriippumattomia ominaisuuksia informaatiotutkimuksen opiskelijoiden (1997) MAUKU-tietokantaprojektissa.

Riehin (2002) mukaan merkityksellistä on tietojärjestelmään ja tiedon substanssialueeseen liittyvän tiedon erottaminen, sekä uusien vaikuttavien tekijöiden löytäminen laatuun ja

auktoriteettiin, kuten tilanne, hakutulosten järjestys ja aineistoon liittyvät ennako-odotukset. Yleiset ennako-odotukset Internetiä kohtaan voivat olla joko negatiivisia tai positiivisia. Tämä voi johtua joko tilanteesta tai tiedonhaussa hakutuloksien osalta onnistumisessa. Omiin sisäisiin malleihin huonosti sopiva hakutulos voi muuttaa laatukäsitteen hakutuloksen todellisesta relevanssityypistä muuksi kuin se todellisuudessa on.

Hakutehtävien mukaiset tiedonhauk Internetiin on tehnyt informaatiotutkimuksen laitoksen tutkija Leppänen tiedonhakijan roolissa informaatiotutkimuksen laitoksen palkattuna työntekijänä loppuvuodesta 2005. Alustavien hakutulosten mukaan jokaisen hakutehtävän aiheenmukaisia hakutuloksia www-sivujen muodossa on haettu yli 40 kpl, joista varsinaiseen relevanssikorpukseen ja relevanssiarviointiin valinnan teen etukäteen tutkimustehtävälähtökohtaisesti ja relevanssiarvioijalähtöisesti. Oma rohkeuteni tiedonhakuja kohtaan on viivästyttänyt tutkimuksen kirjoittamista. Aineistot on kerätty myös relevanssiarvioijan näkökulmasta. Leppänen on toiminut myös vertaisarvioijan roolissa tekemällä rinnakkaiset relevanssiarvioinnit relevanssikorpuksen muodossa. Oma relevanssiarvioijan osuuteni tulee mukaan tässä vaiheessa erillisenä, jolloin relevanssiarviot on tehty toistemme relevanssinäkemyksistä tietämättä.

Molemmat ovat tässä relevanssiarvioijan sekä tiedonhakijan rooleissa samalla myös tiedonkäyttäjää, Leppänen teki relevanssiarviot ammatilliseen työtehtäviinsä liittyen. Minulle tiedonkäyttäjänä näiden neljän hakutehtävän mukaisten relevanssiarviointien merkitys on myös ammatillinen, sillä niitä käytetään yliopistosta valmistumiseen ja filosofian maisterin tutkinnon edellyttämään pro gradu -tutkielmaan.

Olen valinnut www-aineistot tähän tutkimukseen Leppäsen (2005) tiedonhakujen linkkikokoelmista. Valintaperusteena Leppäsen tiedonhakujen ”lähes” koko Internetin informaatioavaruuteen eli avoimeen Internetiin kohdistuneiden hakujen joukosta olen aiheenmukaisten dokumenttien joukosta käyttänyt valintaperusteena tiedonhakutehtävien aiheenmukaisuutta sekä tutkimusasetelmallista lähtökohtaa. Leppäsen linkkikokoelmat olen tallentanut kirjanmerkeiksi omalle nettiselaimelleni. MAUKU-tietokannan aineiston käyttäminen testikokoelmana oli oma ajatukseni, mutta sen taustatiedonhakuun voi hyvin käyttää Leppäsen avoimesta Internetistä hakemia dokumentteja.

Vaikka relevanssiarvioinnin kohteena olevat sivustot ovat nyt saatavissa linkkien perusteella, aiheperustaisten teemojen tarkkaa jaottelua ei saa unohtaa sivustojen selailtavissa olevissa kokonaisarvioinneissa. Osa sivustoista voi olla poistunut käytöstä tai muuttunut alkuvuodesta 2006. Kyseessä ovat myös henkilöhakuaiheisiin perustuvat tiedonhankinnalliset tehtävät, joten relevanssiarvioinneissa kannattaa olla sen perusteella huolellinen. Tosin Leppäsen suorittamien tyhjentävien ja kattavien hakukyselyjen tuloksissa on valikoinnin varaa.

Pyrin jokaisen hakutehtävän osalta samaan kokonaismäärään URL-osoitteiden osalta arvioitavassa relevanssiarvioinnissa. URL-osoitteiden ja niiden linkitettävyyden muille sivuille voi tehdä kokonaismäärällisestä kvantitatiivisesti laskettavissa olevista yhtenäisestä neljän hakutehtävän arvioinnista hieman hankalan. Kyse on kuitenkin lopulta neljästä eri hakutehtävästä ja linkitettävyydessä on kyse Internet-aineiston erityispiirteistä, jotka täytyy ottaa huomioon hakutehtäväkohtaisesti. Relevanssiarvioitavien sivustojen tai niiden käsitteiden kokonaismäärällinen laskeminen ei ole tutkimusmenetelmän kannalta olennaista. Hakutehtäväkohtaiset kontekstisidonnaisuuden piirteet on tärkeämpi huomioida relevanssiarvioinneissa, mutta menetelmän esitystavoissa on hyvä pyrkiä jonkinlaiseen yhtenäistämiseen.

4.2 Bernadette Soubirous

Ensimmäisenä tiedonhakutehtävänä on Bernadette Soubirous 1800-luvun Lourdesissa. Henkilötiedonhakutehtävässä painotetaan hänen elämäntarinaansa jättäen vähemmälle huomiolle ilmiön laajentuneet vaikutukset hänen kuolemansa jälkeen. Tuloksena esitetään aiheiriippuva sisällönanalyysi eli asiasisältö mahdollisimman lyhyenä kuvauksenomaisena kertomuksena, vähentämättä paljoakaan hakutehtävän mukaisia ulottuvuuksia. Teoreettinen saturaatio eli kaikkien relevanttien käsitteiden huomioonottaminen ei ole tutkijan kannalta mahdollista. Sivustoja ei ole saatavilla suomenkielisinä ja englanninkielen tai muiden vielä vieraampien kielten aiheeseen perustuva tulkinta on ongelmallista. Lähinnä pyrin hakuaiheen tiivistelmään otosaineiston rajoissa.

Bernadette Soubirouksen kertomuksen muodostamiseen arvioin ennakoivasti Leppäsen hakutuloksista (2006) ensin 22 sivua, jotka olivat organisaationallisilla tai kaupallisilla

sivustoilla. Nämä sivustot yli 40 sivuston joukosta valitsin siksi, että palvelin tai tuottajayhteisö käytti URL-osoitteessaan myös logoaan.

Tombros, Ruthven ja Jose 2005 määrittelivät käyttäjälähtöisiä aiheiriippumattomia ominaisuuksia, joita ovat rakenne, laatu, kuvat, sekä fyysiset ominaispiirteet. Rakenneominaisuuksista kiinnitin eniten huomiota taittoon, linkkien laatuun ja taustaan. Laatuominaisuuksia, jotka sivuavat samalla läheltä aihelevanssia, ovat käsittelyn laajuus ja tyhjentävyys, tekijä tai muu lähdetieto sekä kuvat. Kaikkien laatuominaisuuksien osalta jouduin kuitenkin tekemään rajauksen tiedonhakutehtävän määrittelemisessä, joka vaikuttaa relevanssiarvioinnin kohteeksi päätyvää aineiston määrää, mutta ei muulla tavoin jätä ulkopuolelle jäänyttä aineistoa periaatteessa merkityksettömäksi. Ajankohtaisuuden osalta tämä tarkoitti Bernadetten osalta 1800-luvun ajanjaksoa. Näin jäljelle jäi kolme korkeasti relevanttia sivua, joiden perusteella pyhän Bernadetten tarina tuli mahdollisimman tiivistetysti sanottua.

Aiheiriippumattomia ominaisuuksia ovat edelleen ajankohtaisuus, yleinen koettu laatu, sisällön luettavuus tai sivun virheellisyydet. Yleisen hetinäkyvän laadun pystyi tunnistamaan sivuilta niiden haku- ja selausvaiheessa sivujen luettavuudesta sekä URL-osoitteesta. Sivujen virheellisyydet näkyivät samoissa kohdissa sekä kaupallisten sivujen kyseenalaisessa asiayhteydessä. Fyysisiä ominaisuuksia tai olomuotoja olivat seuraavat tapaukset: sivua ei löytynyt aiemmin ilmoitetun sivun sijainnin perusteella, kieli oli joko liian vaikeasti ymmärrettävää tai uskonnollista tulkintaa vaativaa tai muuta kuin englantia. Tiedostotyyppit vaikuttivat silloin positiivisilta, kun edustavat kuvat oli sijoitettu asiayhteyteen. Tiedostokoolla ja muilla teknisillä tai fyysisillä ominaisuuksilla on ollut merkitystä vähintään relevanssitasolle 2 ennakoivalla arvioinnilla määrittelemilläni sivuilla. Jos relevanssitaso jäi arvolle 1 tai täysin epärelevantiksi, oli syy aiheiriippumattomissa ominaisuuksissa. Valinta julkisten, kaupallisten tai rekisteröityjen palvelujen välillä sai minut päättämään kertomuksessa vain kolmen organisationallisen sivuston käsittelyyn. Tombrosin, Ruthvenin ja Josenkin (2005) toteama Riehin (2002) kaltainen tutkimustulos ennustavasta arvioinnista, jossa ensisijainen mieltymys sivuihin muodostui, mahdollisti evaluoivan arvioinnin. Kun sivun ennakoiva arviointi ei sopinut omiin odotuksiini, päätin siirtyä seuraavalle tai palata edelliselle. Kun ennakoiva arviointi onnistui, päätin käyttää sivua tutkimuksessa yhtenä hakutuloksista. Bernadette Soubiroukseen liittyvän tehtävän luonteesta eli aiheesta riippui, missä määrin kiinnitin huomiota affektiivisiin tekijöihin,

ennustaviin arvioihin tai tiedolliseen eli tietoiseen valikointiin. Evaluoivan arvioinnin merkittävä laatutekijä oli sivun linkkien yhteys muille relevanteille sivuille.

Erittäin merkittävä tutkimusaineiston valikoinnin tulos oli, että www-käyttäjän ominaisuudessa tein arvioita samalla, kun tutkin Leppäsen (2006) hakutulosten mukaisia sivuja. Vaikka tehtävällä ei varsinaisesti ollut aikarajoitusta, mahdollisimman nopean tiedonhankinnan takia tein kuitenkin valinnat tutkijalähtöisesti niin, että ne voisivat olla missä järjestyksessä tahansa. Lopulta laitoin hakutulosvalintani tärkeysjärjestykseen sen mukaisesti, mikä oli tutkimusasetelman, -menetelmän, tai -tulosten kannalta oleellista. Valikointi tapahtui siis ensin ennakoivan arvioinnin perusteella, painottaen aiheriippumattomia ominaisuuksia sekä sen jälkeen evaluoivan arvioinnin mukaisesti, korostaen aiheriippuvia relevanssikriteereitä tiedonhakutehtävän kannalta sopivan kertomuksen muodostamisessa.

Ennakoivan arvioinnin piikkiin meni ensimmäinen yritys valita käsitteelliseksi lähtöoletukseksi sivu, joka esitteli aihepiirin tärkeitä käsitteitä selkeästi ensimmäisenä vastaantulevana kaikista käytössäni olevista hakutuloksista. Periaatteessa lähtösivuna voisi olla tulosten kannalta mikä tahansa käsittelyyn mukaan otettu sivu. Ensimmäinen vastaantullut käsitteellisesti jäsennetty sivu auttoi kaikkien sivustojen käsitteellistä kokonaisuuden ymmärtämistä ennakoivan arvioinnin aikana ja mahdollisti laajan aihepiiriin tutustumisen. Tässä tapauksessa tämä mallikäsitesivu oli <http://www.catholic-forum.com/saints/saintb06.htm> .

Tämän jälkeen havaitsin, että tutkimustulosten kannalta olisin ensin voinut lähteä liikkeelle käsitteellisesti tarkemmin jäsennetystä sivusta, joka oli Wikipedian http://en.wikipedia.org/wiki/Bernadette_Soubirous.

Grounded teorian kannalta sivujen käsittelyjärjestyksellä ei ole kuitenkaan merkitystä, sillä teorian kannalta kaikilta tutkimukseen liittyviltä sivuilta pitää valita kaikkiin teemaullottuvuuksiin sekä sisällönkuvailuun liittyvät teoreettiseen saturaatioon päätyvät teemat. Wikipedian sivu antoi relevantit käsitteet täsmällisempinä. Lähinnä nämä käsitteellistä informaatiota eniten antavat sivustot rajaavat sitä, mitä muita sivustoja kannattaa käsitellä tai jättää pois päällekkäisyyksien, virheellisyyksien, kielivaikeuksien tai muiden negatiivisten tai positiivisten relevanssikriteerien takia.

Aloitin Bernadette Soubirouksen elämäntarinan käsittelyn uudelleen käyttäen pohjana Wikipedian Bernadettesta kertovan artikkelin otsikkorakennetta, joka kuvasi ilmiötä 1800-luvulla. Laajemman ensivaikutelmakuvan useammilla eri käsitteellä kertova Wikipedian sivu antoi enemmän käsitteellisiä lähestymistapojen malleja muihin sivuihin, mutta karsi samalla hakutuloksia pois sillä perusteella, että oletuksena se asetti muille hyvin suurilukuiselle määrälle sivustoja korkeampia relevanssikriteerivaatimuksia. Oman kiinnostuksen mukaisesta aiheesta on vaikea karsia pois heti ennakoivalla tasolla, sillä tiedonhakujen mukaiset tulokset on vähintään käytävä läpi.

Sikäli lähtötilanne molemmilta mallisivuilta aloitettuna kuitenkin onnistui, sillä väljemmillä relevanssikriteereillä läpikäytyinä on moniulotteisen relevanssiarvioinnin kannalta helpompi päätellä useita toisistaan riippumattomia relevantteja dokumentteja, jotka eivät ehkä tule nousemaan korkeimmalle relevanssitasolle. Ne ovat kuitenkin merkittäviä olemassaolevia sivustoja lisätiedonhankinnan ja sisällön kaikkien puolien esiintymänä, mutta samalla syy rajaukseen. Hyvä puoli tiukemmillä teemaullottuvuuksilla liikkeellelähtemisessä on kuitenkin kokemuksen mukaan se, että teemakäsitteitä voi täten tiivistää enemmän ja varsinaisesta kertomuksesta tulee tarkempi tulosten ja johtopäätösten mukaan yhteenvedetty tiivistelmä.

Mielestäni Wikipedia-projekti ja sivusto kuvaa hyvin onnistuneena esimerkkinä sivustorakennelman asioiden kuvauksia, ketjuja ja yhteyksiä ja voisi toimia oletuksena mahdollisesti muidenkin korkeasti relevanttien aineistojen joukossa. Omassa tutkimuksessani Wikipedia lähtötilannesivuna voisi sopia avoimen Internetin henkilöhakutehtävistä parhaiten Madonna Ritchien Internet-aineistojen moniulotteiseen relevanssiarviointiin, sillä hänen taiteellinen uransa erityisesti 2000-luvun kontekstissa on jäsentymättömämpi ja puolueettomampi verrattuna esimerkiksi suomalaisen huippupoliitikkoon, Heidi Hautalaan, jonka puolueellisempi henkilötietosivu oletuksena sopisi paremmin kuvaan.

Käsitteellisesti paremmin jäsenetty Wikipedia sai siis tutkimaan relevanssiarvioinnin perustaa alusta uudestaan. Valitsin oletukseksi Wikipedian kertomuksen käsitteet 1800-luvun Bernadette Soubirouksesta. Aiheperustainen konteksti sai minut karsimaan relevanttien dokumenttien joukosta edelleen pois kaikki kaupalliset .com-päätteiset sivustot. Näin karsiutui ensimmäisenkin yrityksen oletussivu käsitteellisenä hakutuloksena

kokonaan, ja kertomuksen uudelleenmuotoilu alkaakin Wikipedian osoitteesta. Wikipediassa mainitun oman lähestymistapansa mukaan Bernadette Soubirouksen artikkeli on kirjoitettu neutraalista näkökulmasta. Kaikki kaupalliset sivut relevanssikriteerieni mukaan karsittuani jäljelle jäi ainoastaan kuusi relevanttia org-domainiin kuuluvaa sivua. Org-päätteisiä sivuja täytyi myös karsia siksi, että niitä ei joko löytynyt, ne sisälsivät liikaa päällekkäistä tietoa tai sisälsivät lisäksi sellaisia aiheriippumattomia kaupallisia tai muita ominaisuuksia, joiden merkitys Bernadette Soubirouksen asiayhteydessä jäi epäselväksi. Lopulta päädyin tutkimustehtävän karkeaan rajaukseen tulosten kannalta käsitellen lopulta vain Bernadettea hänen 35-vuotiaan elämänsä aikana. Hänen kokemansa ilmiöt käsittän syvinä teemaulottuvuuksina, mutta tiedonhakutehtävän painotus on Bernadette Soubirouksen henkilökohtaisessa elämässä ja sen ongelmassa. Hänelle tapahtuneet asiat tai kokemukset 1800-luvulla tai kuolemansa jälkeen ovat osa hänen elämäntarinaansa, mutta tiedonhakutehtävänä on Bernadette Soubirous itse.

Keräsin lyhyehkön kommentoidun linkkilistan kuvaamaan tutkimuksellista valikointiani. En halua luoda merkityksiä sivustojen tärkeysjärjestyksellä tai niiden valikoinnilla, vaan painotan kertomuksen koostumista sivustoilta täydentäen toisiaan, poislukien päällekkäisen tiedon määrää ja tämän tutkielman kannalta merkityksettömät teemaulottuvuudet. Ensimmäisen tehtävän sisällönanalyysi ei siis ole täydellinen moniulotteisen relevanssiarvointimallin sovellus, johon tässä tutkimuksessa Madonna Ritchien tiedonhakutehtävän osalta pyrin. Kyseessä on aiheriippuvien teemojen käsittelyyn perustuva groundedt eorian harjoitustehtävä aihepiirin luonteen, kielen ja käsitteellisen vaikeuden takia.

Kommentoitu linkkilista.

1) <<http://www.catholic-forum.com/saints/saintb06.htm>> .

Mallikäsitesivu ensimmäisenä lähtöoletuksena, joka jäi kaupallisen kontekstinsa takia kuitenkin käsittelyn ulkopuolelle. Kategorisoi käsitteellisesti suuren osan Bernadette Soubirouksen tarinasta teemojen yläkäsitteillä. Yläkäsitteinä toimivat linkit, joista lähdin hakutulosten mukaan rakentamaan tarinaa Bernadettesta, olivat mielestäni käsitteellisesti rajallisia mutta sisälsivät aihetta pääasiallisesti kuvaavia teemoja. Relevanssiarvo olisi tämän tutkimuksen kannalta ja aihepiirin määrittelevänä sivuna 3. Jää kuitenkin tämän tutkimuksen käsittelystä pois kaupallisena sivustona, jossa on lisäksi hieman käsitteellisen

tarkkuuden virheellisyyksiä.

2) <[http://en.wikipedia.org/wiki/Bernadette Soubirous](http://en.wikipedia.org/wiki/Bernadette_Soubirous)> .

Tyhjentävästi tietoa sisältävä sivu, jossa sisällöntuottajien tärkeinä pitämät termit on linkitetty erivärisillä kirjaimilla ja kaikilla muilla mahdollisilla tavoilla lisätietosivuille. Sisältää kaiken olennaisen informaation Bernadette Soubirouksesta ja on tutkimuksen tulosten kannalta esimerkillinen mallisivu. Relevanssiarvo 3.

3) <<http://198.62.75.1/www1/apparitions/><<http://pr00006.htm>> .

Kuvaus Bernadette Soubirouksen ja Lourdesin Madonnan kohtaamisesta ja kertomuksellista kuvausta heidän keskinäisistä kohtaamisistaan. Epäselvä, mutta syvää kiinnostusta herättävä asiayhteys. Bernadette Soubirousta käsitellään kuitenkin vain hänen kokemustensa kautta. Relevanssiarvo 2.

4) <<http://www.theworkofgod.org/Aparitns/Lourdes/Lourdes.htm>> .

Hyvin pelkistetty, mutta olennainen kuvaus ja olennaiset kuvat Lourdesin Madonnasta. Sopisi johdannoksi Bernadetten kokemuksille ja Lourdesin Madonna-ilmiöille sellaisenaan. Bernadette Soubirousta käsitellään kuitenkin vain hänen kokemustensa kautta. Relevanssiarvo 2.

Alkuperäisiä sivustoja tarkastaessani tulin siihen tulokseen, että sivujen 3) ja 4) osalta arvioitavat sivut olivat hyvin samankaltaisia. Tämän perusteella olisin voinut valikoida toisen tilalle eri organisaation sivun, joka olisi tuonut esittänyt aineistoa toisella tavalla esim. <http://www.catholicpilgrims.com/lourdes/ba_bernadette_intro.htm>. tai <http://www.wau.org/resources/saints/saintarchive/bernadette_soubirous.html>.

4.3 Heidi Hautala

Toinen tiedonhakutehtävä on Heidi Hautala suomalaisena poliitikkona 2000-luvulla. Lähestymistapa on laadullinen ja tehtävän käsittelyssä painotetaan Internet-aineiston aiheiriippuvien ominaisuuksien tutkimista. Tuloksena esitetään sivustojen korkeasti relevantteja ja relevantteja aiheiriippuvia ominaisuuksia. Tiedonhakutehtävän tavoitteena on kartoittaa Heidi Hautalan poliitikkouraa aiheiriippuvilla teema-aulottuvuksilla ja

teemakohtaisilla relevanssiarvoilla.

Relevanssiarvioitaviksi sivustoiksi valitsin Leppäsen (2006) tiedonhauista sellaiset, joiden URL-osoitteessa oli osassa tai toisessa Heidi Hautalan nimi joko kokonaisena tai osina. Teemarelevanssin kannalta sivuston muodolliset seikat ja layout ovat keskeisessä asemassa henkilöprofiilin muodostamisessa ja ne auttavat tutkimusraportoinnin huolellisuudessa tai selkeydessä. Jäljelle jäi kuusi sivustoa, joiden mukaan muodostan poliitikon henkilö- ja työprofiilin. Tämän ennakoivan ja evaluoivan sivustojen valikoinnin jälkeen tutkimuksen käsittelyä rajataan siis Heidi Hautalan poliitikkoprofiiliin ja -uraan 2000-luvulla, riippumatta siitä, mitä muita asioita sivustoilla esitetään. Tarkemman evaluoivan arvioinnin jälkeen päätin jättää käsittelystä päällekkäistä tietoa sisältävinä pois Heidi Hautalan presidentinvaalisivustot vuosilta 2000 ja 2006. Heidi Hautala oli Vihreän Liiton ehdokkaana molemmissa presidentinvaaleissa, joissa Tarja Halonen tuli valituksi Suomen presidentiksi ensimmäisenä naisena. Molemmat presidentinvaalilinkit ovat Heidi Hautalan Wikipedia-sivun linkkiluettelossa.

a) <<http://virtual.finland.fi/elections/president2000/english/hautala.html>> .

b) <<http://www.heidi2006.fi/>> .

Wikipedia olisi hyvä lähtökohta Heidi Hautalankin uran käsitteellistämiseksi, mutta poliittinen näkökulma voisi tuoda hänen toiminnastaan enemmän relevanssikriteereitä sekä teemaulottuvuuksia esille. Riehin (2002) relevanssikriteerien valossa voidaan lähteä selvittämään Heidi Hautalan 2000-luvun henkilökuvausta sekä teemojen kategorisointia aiheiriippuvien ominaisuuksien osalta aiheilevanteilta Heidi Hautalan omilta sivuilta. Oletusarvona lähtötilanteessa ennakoivana arviointina laitan kaikille teemaulottuvuuksille relevanssiarvon 2. Evaluoivan arvioinnin yhteydessä muutan käsiteltävän sivun valitsemieni teemojen relevanssiarvoa. Käsittelen sivuja seuraavassajärjestyksessä:

1) <<http://www.heidihautala.net/>> .

2) <<http://www.eduskunta.fi/faktatmp/hetekatmp/ed352e-su.htm>> .

3) <<http://www.livejournal.com/>> .

4) <http://fi.wikipedia.org/wiki/Heidi_Hautala> .

Alla oleva taulukko kuvaa eri aiheiriippuvien teemojen relevanssiarvot neljällä sivulla ja

oikeanpuoleinen sarake teemakohtaisten relevanssiarvojen summaa neljän sivun yli.

<u>AIHERIIPPUVAT TEEMAT:</u>	<u>RELEVANSSIARVO:</u>	<u>YHT:</u>
Yhteystiedot	2331	9
Esittely	3322	10
Ansioluettelo, koulutus ja ura	3322	10
Tiedotteet	3010	4
Kolumnit	3131	8
Muut kirjoitukset	2131	7
Puheet	2220	6
Puheenvuorot eduskunnassa	2230	7
Valtiopäivätoimet eduskunnassa: Aloitteet & kysymykset	3331	10
Kuvagalleria	3000	3
Linkkikokoelma	3333	12
Palaute	3130	7
Puheet Euroopan Parlamentin täysistunnossa	3010	4
Kysymykset		
Euroopan komissiolle ja neuvostolle	3010	4
Presidenttiehdokkaat ja presidentinvaalit	1012	4
Kansanedustaja	3332	11
Suomalaiset europarlamentaarikot	3332	11
Muut luottamustoimet	2232	9
Perhesuhteet	1101	3
Mielipiteet	3332	11
Maailman parantaminen	3333	12

Korkein mahdollinen sivustojen tuottama relevanssiarvo 3 yhteenlaskettuna neljältä sivulta voi olla siis 12. Näyttäisi siltä, että tämän osalta sivustoilla on huolehdittu eniten linkkikokoelman ylläpitämisestä sekä Heidi Hautalan toiminnan esittämisestä maailman parantamisena. Tätä korkeasti aiheriippuvaa relevanssia tukevat seuraavaavaksi eniten työkansanedustajana, europarlamentaarikkona ja mielipidevaikuttajana yhteenlasketulla

relevanssiarvoilla 11. Tähän näyttää vaikuttaneen omien relevanssikriteerieni perusteella Heidi Hautalan henkilöesittely, henkilökohtaiset ansiot, koulutus, ura, valtiopäivätoimet eduskunnassa sekä näiden lisäksi lähes yhtä suurella relevanssiarvolla omat yhteystiedot julkisena vaikuttajana sekä muut luottamustoimet. Yhteenlasketulle relevanssitasolle 8, 7 ja 6 ovat päässeet teema-aulottavuudet kolumneista, puheenvuoroista eduskunnassa ja muualla sekä palautteen antamisen mahdollistamisesta www-ympäristössä. Alemmille relevanssitasoille, jotka kaikki kokonaisuutena ja keskiarvona asettuisivat vielä kuitenkin relevanssitasolle 1, päätyivät tiedotteet, puheet Euroopan Parlamentin täysistunnossa, kysymykset Euroopan komissiolle ja neuvostolle, presidentinvaalit, presidenttiehdokkuus, perhesuhteet ja kuvagalleria.

Tämä tulos alimmille relevanssitasoille päätyneistä käsitteistä antaa minulle vaikutelman, että sivustoilta puuttuu perusteellinen asioiden kategoriallinen aiheriippuviin ominaisuuksiin perustuva järjestäminen, joka vaatisi tarkempaa käsitteellisyttä valtavan aineiston käsittävältä Heidi Hautalaa monilla eri tavoilla käsittelevältä sivustojen kokonaismäärältä. Tässä tarvittaisiin formaalia tiedonmuodostamista, joka näkyisi sivustojen uudenaikaisessa rakentelussa. Heidi Hautalan koulutus ja ura poliitikkona voisivat enemmän kohdistetusti esitettynä tuoda maailmanparantamisen vaikutusfoorumeita ja niiden tuloksellisuutta paremmin esille jopa käytännön tasolla. Toisaalta maailmanparantamisen imagolla rakennettu poliitikko näyttää alansa laajuudella olevan työtä pelkäämätön poliitikko, jolle näyttäisivät kelpaavan vähäpätöisemmätkin tehtävät.

Tämä näkemys herättää minussa ajatuksen, että Heidi Hautalan keskittyminen triviaalien tehtävien sijasta korkeampiin tuloksellisuutta vaativiin tehtäviin teoriassa ja käytännössä, avaisi työtä hänen osoittamillaan asioilla ja niiden mukaisilla tehtävillä monille muille samoista asioista kiinnostuneille, mutta kuka maksaisi kustannukset Vihreiden tai ympäristöasioiden ruohonjuuritason työstä? Korkeasti arvostetun Heidi Hautalan tehtävien kautta kuvattuna tämä nousee äärimmäisen tärkeäksi kyseisten asioiden käytännön toteuttamisen kannalta. Asioiden hajanaisuus voi vaikeuttaa sekä poliitikon omaa työtä että niiden asioiden Internet-sivuilla näkyvää kehittämistä, jonka mukaisesti kyseistä työtä arvostellaan, kommentoidaan, kiitetään tai tehdään. Kaikkien asioiden vähäisen relevanssitason julkinen käsitteleminen ei näyttäisi välttämättä tuovan lisäarvoa hyvien päämäärien mukaisten tavoitteiden toteutumiseksi.

Ennusteen mukaisesti Wikipedia täytti odotukset eikä kerro riittävän tyhjentävästi poliitikon urasta. Sivusto ei täytä minkään muun teeman kuin maailmanparantamisen teemaullottuvuuden osalta korkeinta relevanssiarvolukua, vaikka Heidi Hautalan käsittely artikkelissa painottuikin 2000-luvun asioihin. Riehin relevanssikäsitteiden mukaisesti arvioisin Wikipedian sivuston Heidi Hautalasta laadun ja auktoriteettikäsitteiden mukaisesti hyväksi, tärkeäksi, hyödylliseksi ja hyväksi hommaksi ajankohtaisuudeltaan. Toisaalta arvioin huonoksi tarkkuudeltaan, taiteellisuudeltaan, ylläpidoltaan sekä kehitykseltään. Hyödyllisyydeltään artikkeli saa kuitenkin negatiivisia aiheriippumattomia ominaisuuksia, kuten informatiivinen, ei ole paljon käytettävää tai käsiteltävää, ei ole helposti ymmärrettävissä, ei auta. Kognitiivisen pätevyyden voisi arvioida wikipedian osalta luotettavaksi, viralliseksi, ammatilliseksi. Sivun laatua heikentää kuvan tai kuvien puute.

4.4 Mauri Kunnas

Kolmas tiedonhakutehtävä on Mauri Kunnas kirjailijana ja kuvittajana. Tärkeimpänä lähestymistapana tehtävään on aiheriippumattomien ominaisuuksien painottaminen tietokannan Ammatit-hakusanalla tesauruksessa esiteltyjen asiasanojen osalta. Tiedonhakutehtävän rajauksena on sen kohdistaminen Informaatiotutkimuksen laitoksen vuoden 1997 kevätlukukauden opiskelijoiden MAUKU-tietokantaprojektin valmiiseen viitetietokantaan. MAUKU-tietokannan URL-osoite on

<<http://www.info.uta.fi/teaching/mauku/mauku.htm>>

Käsiteltävänä aineiston valitsen kvalitatiivisesti ja kvantitatiivisesti oman kiinnostukseni mukaisen tiedonhakutehtävän mukaisesti. Tuloksena esitetän tietokannan aiheriippumattomia ominaisuuksia. Tiedonhakutehtävän tarkoitus on tehdä tietokantalähtöisesti MAUKU-tietokannan aiheriippumattomien ominaisuuksien osalta moniulotteinen relevanssiarviointi tesauruksen ammatit-kuvailutermin osalta. Tiedonhakutehtävän rajauksena ja lähestymistapana on yhden laajemman termin käsitteellisen kuvauksen onnistuminen Mauri Kunnaksen painettuun tuotantoon perustuvassa viitetietokannassa.

Riehin (2002) tutkimuksessakin hakijat käyttivät myös tietokantoja ja kirjastojärjestelmiä, jolloin käyttäjät olettivat, että heidän hakemansa tieto oli jo tiedontuottajien, tietokantojen

ylläpitäjien ja kirjastonhoitajien valmiiksi arvioimaa. Riehin ja Belkinin mukaan web on suhteellisen uusi resurssiympäristö, jossa on erilaiset säännöt ja kriteerit, kuin mitä käytetään painetussa kokoelmassa. Pitkäänkin perinteisillä informaatioaloilla työskennelleet voivat olla webin käytössä noviiseja.

Tietokannan sisältämän aineiston valinneet ja kuvailleet tuottajat olivat informaatiotutkimuksen aineopintotason perusopiskelijoita. Itse olen relevanssiarvioijan roolissa syventävien opintojen pro gradu -opinnäytteen tekijä, joka arvioin tietokantaprojektissa valittua aineistoa tiedonhakijan, -hankkijan ja -käyttäjän näkökulmasta. En voinut sisällyttää analyysiin koko tietokantaa joten valikoin tietokannasta asiasanalla ”ammatit” ja sitä suppeammalla asiasanalla poimitut tietueet. Ammatit-asiasanan lähestymistapaa aineistoon perustelen käyttöliittymän ominaisuuksilla. Aakkosellisesti ammatit-termi sijaitsee käyttöliittymän ylimmissä osissa ja on täten helposti selailtavissa. Ammatit voivat hakuterminä ja asiasanana viitata myös moniin eriluonteisiin teoksiin alueensa mukaisesti. Näin voisi olla Mauri Kunnaksen Koiramäki-tuotantoa itsekin lukeneena myös laajemman eläimet-termin osalta, mutta ammattien kuvaus lastenkirjallisuuden kuvauksessa on mielenkiintoisempi ja harvinaisempi näinkin suuressa mittakaavassa yhden kirjailijan ja kuvittajan tuotannossa. Tontut yläkäsitteenä eli laajempina terminä viittaisivat oletettavasti vain jouluaiheisiin kirjoihin.

Näen tässä heti lastenkirjallisuuden ja kuvituksen osalta erään luovan käsitteellisen ongelman, jota tietokantaprojektin opiskelijat ovat kenties itsekin miettineet: eivätkö tontut kuulu ammatteihin? Perusteluni jatkuu kuitenkin tesauruksen käsitteellisen jäsentämisen kritisointina tämän osalta, että tontut laajempina terminä ovat rinnastettu RT:nä eli rinnakkaisen terminä poliitikkoihin, jotka sisältyvät ammatit-ryhmään, johon tonttuja ei ole merkitty. Eivätkö 10 eri tonttulajia ole ammatteja, jos ne RT:nä rinnastetaan poliitikkoihin? Tesaurusta selatessa tulee mieleen eivätkö aaveet, kummitukset ja peikot voisi olla tonttujen ja haltijoiden ammatteja, tai emännät, isännät, piijat tai rengit, joita ei niitäkään ole merkitty ammateiksi? Entä urheilijat ja toimittajien LT:nä toimitustyö? Entä käsitteet toimitustyö sekä yötyö työnteon suppeampana ST-muotona tai lehden tekeminen toimitustyön ST:nä? Käsitteellinen epäselkeys on myös urheilijoiden esittäminen KT:nä eli korvaavilla termeillä urheilujoukkueet ja kilpailijat. Urheilujoukkueiden osalta voin havaita varsinaisen virheellisyuden, koska kuitenkin urheilujoukkueiden osalta hakuterminä tesauruksessa esitetään urheilijoiden käyttämistä. Urheilijat eivät kuitenkaan alun perin kuulu MAUKU-

tietokannan ammattien joukkoon, mutta tesauruksen rakenteen käsittelyssä otan nämä kyseenalaistavat miellelyhtymät esiin. Eivätkö tontut ole ammatteja Mauri Kunnaksen tuotannossa, jos ne on suppeampina terminä jaoteltu kymmeneen alalajiin?

Luokkahaun mahdollistava käyttöliittymä jakaa viitetietoaineisto-osuuden seuraavilla hakutoimintoyläkäsitteillä: Tietoiheet, Möröt, Joulu, Kansanperinne, Poliitikka, Poppikulttuuri. Näiden yläkäsitteiden tai oikeastaan aineistoista riippumattomien otsikoiden muodostuminen on näppärä keksintö hakuominaisuutena, mutta ei aineistosta induktiivisesti kumpuava perusteellinen groundedteoreettinen sisällönanalyysi, vaan lähinnä toteutuksellinen niksi. Tämän perustelisin sillä, että möröt ja joulu voidaan mielestäni laskea kuuluvaksi kansanperinteen yhdeksi alakäsitteeksi. Poliitikka ja joulu voivat olla myös tietoiheita ja joulu kansanperinnettä sekä poppikulttuuri voi olla kansanperinnettä tai tietoihe jne. Nämä voitaisiin ymmärtää korkeina teemaullottuvuuksina, mutta niiden käsitteellisessä määrittelemisessä ei ole onnistuttu.

Hakiessani HAKU-toiminnolla kahdeksaa määriteltyä ammattia niiden LT:llä - ammatit-asiasanalla en pystynyt löytämään viitetietokannan tietoiheiden puolelta kuin LT:n ammatit mukaisesti indeksoituna muuta, kuin kirjan Yökirja: eli mitä kaikkea tapahtuu yöllä sekä kansanperinteen osalta teoksen Koiramäen lapset kaupungissa kaksi eri painosta, joissa ammateiksi on esitetty käsityöläiset, säätyläiset, perheet ja merimiehet, joista säätyläiset eivät kuulu tesauruksen alkuperäisiin ammattikäsitteisiin. Tämä tosin ei ole merkittävä virhe, sillä teokset löytyivät myös LT:n mukaisella ammatit-asiasanalla. Mörköjen, joulun, politiikan ja poppikulttuurin osalta ei löytynyt ammatteja sisältäviä teoksia.

Muuttamalla ammatit-asiasanan muotoista hakutoimintoa listaustoiminnon mukaiseksi, tietokanta ei tuottanut minkään kuuden hakutoiminnon osalta muuta kuin virheilmoituksia. Tältä osin olin täysin pettynyt tietokannan aineistoista riippumattomaan toimivuuteen ammatit-tiedonhakutehtävän valossa. Sanahauhallakaan LT:n eli ammatit -laajemman termin perusteella ei tuota kaikkien kahdeksan ammattitermin mukaisia tuloksia. Löytyneiden kahden teoksen osalta viitetietokanta antaa hyvin kirjasto- ja järjestelmälähtöisen sisällönkuvailumenetelmän mukaisen kuvauksen siitä, mitä sisällönkuvailu pahimmillaan voi käytännön näppäilytyönä olla liittämättä tähän lisäksi tietueiden kopiointi- tai ostopoliittisia kirjastomerkitä. Tietokannan omien kiinnostusten kohteiden mukaisena

kirjastolähtöisenä sisällönkuvausharjoituksena, vailla näiden viitetietueiden käyttöliittymäystävällisiä esittämismuotoja, tietokantaprojekti on varmasti ollut hyvä harjoitustyö käytännön työhön viitetietueiden muotoseikkojen ja aineistoon tutustumisen kannalta.

Ensivaikutelmana voisi ajatella, että tietokannan tesauksen sekä hakemiston, hahmojen, paikkojen sekä muiden nimien runsaslukuisesta luetteloinnista huolimatta nämä käsitteet eivät muodosta käsitteellistä ja yhtenäistä kokonaisuutta linkkirakelmana, hierarkiana tai näiden termien muodostamana selkeänä kuvauksena siitä, mitä viitetietokannalla halutaan sanoa tai tuoda esille. Tässä mielessä käyttöliittymää kokonaisuutena ei voi sanoa myöskään aiheen erityistietämyksen kannalta onnistuneena, sillä tietokannan yhteydessä ei kerrota, mihin asiaan tietokannan avulla on tarkoituksenmukaisinta saada vastaus. Asiasanat ja hakutermit ovat varmasti olemassaolevia oikeassa painetussa aineistossa, jonka käsitteelliseen listaukseen varsinaisina sunstantiiviasiasanoilla on pyritty ja muodostettu niiden perusteella tesauksen linkkihierarkia. Suuren määrän hahmojen, paikkojen ja muiden erisnimien listauksella on niiden esiintymänä varsinaisessa aineistossa varmasti iso merkitys, samoin kuin niiden listaamisen käsittämisenä äärimmäisen suuren luovan kirjallisen työn määrällisenäkin mittarina. Ne eivät vain aiheriippumattomina ominaisuuksina onnistu antamaan ennen tiedonhakuaiheeseen lähtemistä sellaisenaan linkitettyä kuvaa siitä, mitä tietokannan avulla on tarkoitus löytää ja mistä syystä sitä kannattaa etsiä.

Täytyy kuitenkin muistaa, että tietokannan rakentajat ovat käyttäneet MAUKU-tietokannan aiheriippumattomiksi ominaisuuksiksi käsitettävien formaalisten muotoseikkojen pohjana varsinaista aineistoa, jota tietokannan avulla ei pysty näkemään alkuperäisessä muodossaan. Silloin ei voi sanoa absoluuttista totuutta, miten asiat Mauri Kunnaksen satukirjanomaisessa tuotannossa on käsitelty ja onko tesaurus siinä mielessä rakenteellisesti johdonmukainen. Tällöin nämä formaaliset muotoseikat muodostuvat aineistoa näkemättömälle todellisiksi aiheriippumattomiksi ominaisuuksiksi moniulotteisen relevanssiarviointimenetelmän termeillä sanottuna. Kuitenkaan niiden totuudenmukaisuuteen ei voi ottaa kantaa aineistoa käsittelemättä, eikä tesaurus sellaisenaan anna siihen paljon johtolankojakaan. Tesauksen hakemistojen ja erisnimien listaus ei myöskään suoraan käyttöliittymässä johda varsinaisen tiedonhaun helpottamiseen luokkahakujen osalta.

Käyttöliittymänä MAUKU-tietokantaprojektin ensimmäinen ”näyteikkuna”, eli hakusivu on selkeä ja rauhoittava värisävyiltään. Aiheriippumattomien ominaisuuksien suureksi ansioksi pohjan osalta katson tesauruksen ja muiden hakutermien sijoittamisen sivun oikeaan laitaan, jolloin käyttäjä hakijana voi lähteä hakutoimintoihin tavallaan sivun vasemmasta laidasta kirjoittamalla tai hakumekanismin automaattisia ominaisuuksia käyttämällä. Väriykseltään pehmeillä ja vaaleilla lapsenomaisilla pastellisävyillä, vaaleanvihreällä, vaaleanpunaisella sekä vaaleansinisellä pohjalla on miellyttävä lukea käsitteellisiä hakutermejä ja erisnimiä sellaisenaan vailla yhtäkään kuvaa.

Jos kuvattomuus on tietoinen päätös Mauku-tietokannan yhteydessä, ilman sitä käyttöliittymän ilmeessä ilman käsitteellisiä merkityksiä on onnistuttu hyvin korkean relevanssiasteen mukaisesti relevanssiarvolla 3. Jos kuva täytyisi lisätä, sen valinnassa tulisi olla äärimmäisen huolellinen, ettei se yksinään laskisi yleisilmeen aiheriippumatonta relevanssia, sillä se antaisi ainakin hahmona, ellei jopa tietokannan käsitteellisen merkityksen yhteydessä käyttöliittymälle sellaisen piirteen, joka on näkyvissä koko ajan tiedonhaun käsitteistä riippumatta. Kuva, joka tässä tarkoituksessa voisi olla aiheriippumattomasti sopiva, olisi esimerkiksi Mauri Kunnaksen omakätinen nimikirjoitus tai kopio Ex Libriksestä, joka olisi Mauri Kunnaksen oma ja itse suunnittelema. Mauri Kunnaksen omasta mahdollisesta Ex Libriksestä tai sellaisen käyttämisestä minulla ei kuitenkaan ole mitään tietoa. Käyttöliittymäpiirteet varsinaisen viitetietokannan hakujen hakutulosten yhteydessä eivät ole yhtä houkuttelevia.

Crystal & Greenberg (2006) relevanssikriteerien tutkimuksessa käyttäjiä pyydettiin hiirellä osoittamalla tunnistamaan käyttämänsä kriteerit relevanssiarvioinnissa. Avainkriteerit, jotka tulosten muodossa tunnistettiin niiden esiintymien frekvenssien perusteella, olivat tutkimus, hakuaihe, laajuus, tieto, vaikutus, tunteet, web-aineiston ominaisuudet ja auktoriteetti tai sivuja julkaiseva henkilö. Nämä kaikki relevanssiikriteerit herättivät kiinnostukseni MAUKU-tietokantaa kohtaan ja tulivat käsitellyiksi moniulotteisen relevanssiarviointimenetelmän kannalta MAUKU-tietokannan aiheriippumattomaksi tekijöiksi käsittämieni asioiden osalta, jota kyseisessä tietokannassa oli koko tietokanta sekä kaikki sen osat, koska yhteyksiä ja linkkejä alkuperäiseen varsinaiseen aiheriippuvaan kokonaiseen aineistosisältöön ei ollut. MAUKU-tietokanta sellaisenaan oli positiivisessa merkityksessä erinomainen esimerkki siitä, minkälaista on aiheriippumattomien ominaisuuksien esittäminen Internet-aineistossa täydellä relevanssiarvolla 3.

Tutkimuksessa havaittiin, että tiedonhakujärjestelmien ja niiden hakukoneiden tulisi koota laajempaa kontekstisidonnaista aineistoa. Vuorovaikutus tiedonhakujärjestelmien hakutulosten kanssa on kaksivaiheinen prosessi. Ensin käyttäjät määrittelevät käyttöliittymäpiirteet, jotka käsittävät tietyt metadataelementit, kuten dokumentin otsikon, dokumentinkuvauksen, aiheen tai asiasanat. MAUKU-tietokantaprojektissa tätä oli tavoiteltu ansiokkaasti, mutta lopputulos on jäänyt tarkistamatta tai ajan ja resurssien puutteessa tulokseksi sellaisenaan. Kyseinen totuuden esittämisen mahdottomuus vailla selityksiä on myös tieteellinen argumentointi, kun muuta ei voida todeta. Evaluoivaa arviointia olivat käyttäjäkontekstin tilanteenmukaiselle ammatit-tehtävälle relevantit dokumentit ja niiden esittämisen traditionaalisuus yllätti viitetietokannan osalta. MAUKU-tietokannan käsittelyn neutraalius kuitenkin tekee tietokannasta käyttäjälähtöisen käyttäjän ikäryhmästä riippumatta.

Tiedonhaketutkimuksen haaste on edelleen Crystalin ja Greenbergin (2006) mukaan sellaisten tiedonhakujärjestelmien suunnittelemisen, joka tukee tehokkaasti käyttäjiä ennakoivissa ja evaluoivissa vaiheissa. Empiiriset tutkimukset voivat tarjota perustan tälle suunnittelun haasteelle vahvistamalla käyttäjämalleja ja käyttäjien mieltymyksiä.

4.5 Madonna Ritchie

Neljäs tiedonhakutehtävä on poplegenda Madonna Ritchie ja hänen brändinsä sekä elämä julkisuuden henkilönä 2000-luvulla. Tarkastelun pääpaino tutkimusmetodin kannalta on relevanssikorpuksen rakentamisessa. Valmis relevanssikorpus esitetään vain tämän tiedonhakutehtävän osalta. Madonna Ritchien 2000-luvun elämästä kertovien Internet-sivustojen tiedonhakutehtävänä on kartoittaa aineistoista teemaullottuvuuksien pää- ja sivuteemoja. Sen lisäksi tarkastellaan aineiston aiheiriippumattomia ominaisuuksia. Tavoitteena on moniulotteisen relevanssiarvioinnin mukainen relevanssikorpus Sormusen (2006) menetelmän mukaisesti. Madonna Ritchien tiedonhakutehtävän osalta näen grounded teoreettisen kertomuksen muodostamisen sisällöltään erilaisena kuin Bernadette Soubirouksen yhteydessä. Madonna Ritchien tiedonhakutehtävä kertomuksena on relevanssin ja relevanssikriteerien perusteleminen tai merkityksellistäminen sekä käyttäjä-että aihelähtöisesti.

Mukana olevat sivustot on valittu ensin www-palvelimensa perusteella. Sivuja ei ole kerätty aihehakemistoista eikä mainossivuilta, vaan (Leppäsen, 2006) yli 40 hakutuloksen joukosta. Palvelimet ainakin vaikuttavat olevan sellaisia, jotka käsittelevät sisällöntuotantona pelkästään Madonnan julkiseen elämään liittyviä asioita. Jokaisen palvelimen nimessä on Madonnan nimi URL-osoitteen osana. Sivustojen arvioinnissa lähdetään liikkeelle aina pääsivulta, josta tulisi näkyä sivu kuin kirjan kantena vähintään URL-osoitteen tasolla. Madonnan merkittävimmän kotisivun osalta tämä toteutuu ihan konkreettisesti, kun hänellä on uusimman lastenkirjansa kannesta kuva sivustonsa etusivuna, joka on lisätty kaikkien muiden sivujen etusivuksi lokakuun lopussa 2006. Etusivua on marraskuussa 2006 päivitetty.

Madonnan sivuista valikoitui arviointini kohteiksi kolme korkeasti relevanttia sivustoa, jotka eivät sisällä päällekkäisyyksiä, ja valikointini mukaan ne nostavat lähestymistapoina Madonnaan henkilönä esille erilaisia näkemyksiä.

Madonnan sivujen esittämisessä mainostamista ei voi sulkea pois 2000-luvun kontekstissa, sillä taiteilijan koko elämäntyö on ollut kaupallisten tuotteiden julkaisemista. Se ei kuitenkaan tee Madonnasta yksityis- tai julkisuuden henkilönä vähemmän kiinnostavaa, vaan aineiston perusteella yksityinen, julkinen ja kaupallinen maailma muodostavat Madonnasta suoraan sanottuna täydellisen ihmisen ikonin.

Madonnan sivujen arvioinnissa minulla ei ollut niistä ennakkotietoa tai ennako-oletuksia muiden kuin Madonnan kuuluisimman www.madonna.comin päivittäisestä pääsivulla vierailemisesta. Näin ennakoivat ja evaluoivat arvioinnit pohjautuvat niin tämän kuin muidenkin hakutehtävien osalta informaatio-objektien ominaispiirteisiin enemmän kuin omaan tietooni. Tilannetekijät ovat kuitenkin voineet vaikuttaa mahdollisimman hyviin arviointeihin käyttämieni korkealaatuisten teknisten laitteiden ansiosta sekä omista arvostuksistani tai merkityksenannoista johtuen.

Moniulotteisen relevanssidatan merkitys on kuitenkin tutkia relevanssikriteereitä, jotka tekevät valmiista sivuista relevantteja, tai sitten ei niiden käyttäjille. Viitekehysten jo tutkittujen relevanssikriteerien lisäksi olisi hyvä saada uusia. Madonnan osalta pystyn antamaan aiheriippumattomille ominaisuuksille sellaisenaan korkeasti relevantin relevanssikriteerin määrittelemällä ne ainakin valitsemani aineiston osalta hyvin korkeasti

relevanteiksi lisäarvotekijöiksi Internet-aineistojen yhteydessä. Riehin (2002) tutkimuksessa informaatiolähtöisiin relevanssikriteereihin kuuluvat sisältö, grafiikka, sivun rakenne, organisaatio ja informaation tyypit. Madonnan sivujen yhteydessä näkisin sivun rakenteen aiheriippumattomana ominaisuutena, mutta muut kriteerit aiheriippuvina.

Madonna Ritchie-hakuaiheessa Tombrosin, Ruthvenin ja Josen (2005) tehtävälähtöinen käyttäjätutkimus sopii parhaiten sivujen tarkastelun apuvälineeksi (ks. alaluku 3.6). Aiheriippumattomien ominaisuuksien esittämistapa alla olevien sivujen mukaisessa järjestyksessä ei välttämättä ole malliesimerkki siitä, miten aiheriippumattomia ominaisuuksia tulisi taulukoida. Aiheriippumattomat ominaisuudet ovat kuitenkin sivustojen lukemisen ja relevanssiarvioinnin perusteella yhteenvetoja sivustojen sellaisista ominaisuuksista, jotka vaikuttavat relevanssiarviointiin selaamisen ja varsinaisen relevanssiarvioinnin aikana. Aiheriippumattomat ominaisuudet ovat Internet-aineistoissa usein jo URL-osoitteen tasolla niitä, jotka vaikuttavat eniten Internet-aineiston havaitsemiseen ja arviointiin tietokoneen ruudulla.

- 1) <<http://www.madonna.com/>
- 2) <[http://en.wikipedia.org/wiki/Madonna_\(entertainer\)](http://en.wikipedia.org/wiki/Madonna_(entertainer))
- 3) <<http://www.beautifulmadonna.com/eng/>

<u>AIHERIIPPUVAT TEEMAT</u>	<u>RELEVANSSIARVO</u>	<u>YHT.</u>
Kuvaus Madonnasta taiteilijana	333	9
Kuvat	333	9
Madonnan taiteilijaelämä 2000-luvulla	333	9
Kaupallinen menestys	332	8
Näytteleminen ja elokuvaaura	331	7
Madonnan elämästä tehdyt muut dokumentit	333	9
Aikuiselämän menestyminen	333	9
Madonnan yksityiselämä	331	7
<i>Madonnan historia</i>	331	7
<i>Madonnan perhe</i>	333	9
<i>David Bandan adoptio lokakuussa 2006</i>	330	6
<i>Kabbalah Centre</i>	300	3
Madonnan oman tyylin kuvaus	333	9
Uusin tieto, tiedotteet, uutiset	333	9

Sivustolle kirjautumisen mahdollisuudet	333	9
Linkit muille relevanteille sivuille	333	9

Aiheriippumattomat ominaisuudet sivustojen mukaan järjestyksessä

1) Madonnan omilla sivuilla kaikkia aiheriippumattomia ominaisuuksia voi sanoa lisäarvotekijöiksi. Sivustot ovat itsessään Madonnan elämästä tehtyjä muita dokumentteja. Näin suurta oman tuotannon esittämistä Internetissä tai muissa tietomuodoissa ei voi erottaa yksityiselämästä. Kaupallisena sivustona Madonnan tuotannon käsittely ei ole täysin johdonmukaista, mutta kaupallisuuden osalta puolustaa aihelevanssia. Käyttäjystävällisyys näkyy sivustorakenteessa lähes koko Madonnan tuotannon esittelemisenä kiinnostuneelle ja tarkkaavaiselle selaajalle. Kuvien suuri määrä ja tekstin niukkuus lähinnä otsikkotason linkkeinä lisätietosivuilla tekee asian käsittelystä hieman epätarkan, mutta se johdattelee kiinnostuneen käyttäjän niille sivuston osille, josta käyttäjä on oikeasti kiinnostunut. Musiikin yhdistäminen sivuston ominaisuudeksi vailla ennakkovaroitusta on tiedonhakijan ympäristön kannalta hieman epäystävällistä, mutta tottuneille Internet-käyttäjille, monien selausikkunoiden käyttäjille sekä asian harrastajille varsinainen relevanssikriteeri.

2) Madonnan Internet-dokumentoinnin laajuus, kuvat, laatu, linkkiyhteydet, linkkien takuuarma toimivuus, käyttäjä- ja tuottajalähtöisyys ovat kaikki korkeasti relevantteja aiheriippumattomia ominaisuuksia. On kohteliasta ilmoittaa, että henkilökohtaisen elämän kuvaus on työn alla ja kehitysvaiheessa. Kuitenkin tiedon vapaa käytettävyys ja tulkinnan mahdollisuudet tekevät Wikipedian sivusta tässäkin tapauksessa todellisen aarteen. Tarkempaa tai laajempaa kuvausta Madonnan elämästä on todella vaikea löytää Internetistä muualta.

3) Sivusto on itsessään Madonnan elämästä ja sen tulkinnallisuudesta tehty dokumentti ja hänen luovan persoonansa ”moniulotteisuutta” uudella tavalla soveltava sivusto. Relevanssikriteereiltään se on takuuarma nuorisosivu. Näkökulma Madonna Ritchien tuotantoon ja sen käsittelemiseen tuo sille kohderyhmäkohtaisen ulottuvuuden, joka ei kuitenkaan aseta ikärajaa kyseisen sivuston käyttäjille.

5 Tulokset

Tutkimuksen yhteydessä on tärkeää muistaa kysyä, mikä on tiedonhaun tavoite? Tiedonhaun tavoitteena on löytää hyödyllinen informaatio tiedontarvitsijalle tai erityiseen tiedontarvetilanteeseen oikeassa muodossa. Tiedonhaun tutkimus on siis myös tiedon tallennuksen ja sen organisoinnin tutkimusta. Tiedonhaun vastauksena ovat relevantit dokumentit tai relevanssi. Koska relevanssin merkitys on hyvin keskeinen IR-tutkimuksessa, ovat jo suuret määrät käyttäjätutkimuksia määritelleet, miten relevanssiarviointeja tehdään myös avoimessa Internetissä tai kokoelma- tai hakukonelähtöisesti järjestetyissä Internetin osissa (Crystal & Greenberg, 2006).

Aiheriippuvien ja aiheriippumattomien ominaisuuksien erottaminen Internet-sivustoilla on hyvin helppoa. Oletuksia määrittelevät tutkimukset antavat tähän hyvin ymmärrettäviä suuntia antavia käsitteitä. Internet-aineistoissa on erityisen paljon aiheriippumattomia ominaisuuksia, jotka liittyvät aineiston esittämiseen Internetissä. Välittäjänä aineiston esittämiseksi on tietokone ja sen oheislaitteet ja Internet-aineiston kirjoittamiseen käytetty koodauskieli, esimerkiksi HTML ja muut ohjelmointikielet.

Hakutuloksen selaamisen yhteydessä totesin, että havaitsen Internet-sivuilta ensin aiheriippumattomia ominaisuuksia. Ne vaikuttivat ennakoivana arviointina siihen, mitä sivustoja lopulta valikoin tutkimuksen relevanssiarvioiden kohteeksi. Näitä ”heti näkyviä” aiheriippumattomia ominaisuuksia olivat URL-osoitteen nimen lisäksi aiheen otsikointi, sivuston layout, taitto ja siinä käytetyt kuvat ja värit ja näiden kaikkien sijainti tietokoneen ruudulla näkyvänä sijaintina. Aiheriippumattomia ominaisuuksia olivat myös mainokset sekä kuvien puute ja tekstin kieli, jonka tulkitsemisessa englanninkielinen sanakirja, joko painettuna tai suoraan netistä käytettynä oli usein tarpeen. Epämääräinen aloitussivu tai aiheriippuvan sivun esittäminen irrallisena asianhaarana varsinaiselta palvelimen selittämättä tätä yhteyttä heikensi aiheriippuvaa asiasisältöä, vaikka asia tekstinä olisikin hyvin esitetty.

Internet-ympäristössä, jossa useiden tekniikoiden ja samalla aiheriippumattomien ominaisuuksien esittäminen on mahdollista, aihe relevanssiin eivät vaikuttaneet pelkät tekstit. Arviointiin vaikuttavat enemmän ne piirteet, jotka tekevät Internetistä sellaisen

tallennusmuotofoorumin, joka se oikeasti on. Aiherelevanssiin voivat vaikuttaa aiheriippumattomina ominaisuuksina yhtä paljon värit ja kuvat, niinkuin painetussa aineistossa. Eroavaisuus on se, että Internet on digitaalinen tiedon tallennus- ja tiedonhakuympäristö. Painetun aineiston sijainti on riippuvaista sen fyysisestä olomuodosta ja varsinaisesta sijainnista, mutta digitaalinen Internet on olemassa ja käytettävissä teknisillä laitteilla riippumatta siitä, mistä niitä käytetään. Käytettävissä oleva tekninen laitteisto voi vaikuttaa Internetin sivujen relevanssiarvioihin digitaalisella yhteysnopeudellaan, teknisellä tasollaan ja käyttäjän teknisten laitteiden kokemuksen ja käsittämisen mukaan.

Kirjapainotaidonkin tulokset olivat ensin yllätys, ne aiheuttivat vuosisatoja kestäneen sensuurin tarpeen kirjastoissakin bibliografisena kontrollina, samoin kuin Internet-aineisto voidaan aiheriippumattomasti kokea epäluotettavana. Tästä huolimatta Internet-aineisto ei kuitenkaan ole ainoa tallennusmuoto, jossa voidaan tehdä moniulotteisia relevanssiarviointeja. Moniulotteinen relevanssiarviointimenetelmä sopii Internet-aineiston sisällönkuvailemiseen tai hieman suppeampaan kuvaamiseen erinomaisesti.

Grounded theory -menetelmän laajempi opiskelu ja menetelmällinen käsittely auttaisi relevanssiarvioinneissa ja aineistojen tulkinnoissa. Käsitteellinen sisällönymmärtäminen voi oikeilla rajauksilla lähteä myös kirjastoammatillisesta käsitteiden määrittelyssä, jota tapahtuu kirjastoissa tesauruksien ja aineistojen sisällönkuvaileuoppaiden pohjalta. Internet-aineistojen käsitteellinen tulkinta voi tesauruspohjalta antaa aihetta miettiä käsitteiden termien lisäämistä painettujenkin aineistoon tesauruksiin ja luettelointisääntöihin. Samalla tavalla tehdään tilaa uusille alusta asti lähtien suunnitelluille aineistontuotannoille, joiden käsittely ylittää yhden teoksen aiheriippuvuuden niin, ettei tietoa tarvitse etsiä turhan monesta paikasta relevantin ja tarpeellisen tiedon löytämiseen. Internet-aineistojen käsitteellinen läpikäyminen voi antaa aiheita päivittää myös voimassa olevia elektronisten aineistojen luettelointisääntöjä avain- ja asiasanoineen. Kokonaisia teoksia yhteyksineen ja suorine lisätietolinkkeineen voi tallentaa digitaalisesti siis myös Internetiin. Internet on tämän asian kannalta painettua aineistoa parempi paikka sisältäen suorat yhteydenpitofooromit tuottajasta, julkaisijasta ja tekijästä käyttäjiin, välittäjiin sekä arvioihin.

Relevanssiarvioinneissa olen pyrkinyt mahdollisimman objektiiviseen arviointiin, mutta

subjektiivisuutta ei voi mitenkään sulkea pois. Objektivisen arvioinnin kohteena on varmastikin varsinaisen relevanssiarvioinnin lisäksi vähintään yhtä paljon tutkimusmetodillinen lähestymistapa ja opinnäytteen tutkimusmenetelmällinen onnistuminen. Moniulotteisen relevanssiarviointimenetelmän osalta aineistoja tutkimuskohteiksi Internetissä sekä lähes kaikissa formaateissa on loputtomiin. Analyysitaitoja menetelmien osalta tarvitaan myös valmiin relevanssidatan tulkinnassa, päivittämisessä, muuttamisessa tai monissa Internet-kontekstin mukaisissa mahdollisissa jatkotutkimuksissa.

Informaatiolukutaitostandardien muistaminen muiden tekstien lukemisen sekä oman opiskelun tai kirjoittamisen yhteydessä olisi kuitenkin hyvä keino ylläpitää omien tiedonhaullisten tavoitteiden kontrollia. Tyhjänpäiväinen tietämys ei ole tarpeen hätätilanteissakaan. Informaatiolukutaidon standardimuotoinen tietämys dokumentaationa edistäisi paremmin koossa pysyvää tiedonhallintaa muodollisissa tehtävissä. Osaamistavoitteen sanamuodossa ilmaistuna informaatiolukutaitostandardeissa painottuisi kaikenikäisten, myös muiden kuin yliopisto-opiskelijoiden omien kriteereiden käyttäminen tiedonhakukiinnostuksen kohteiden ylläpitämisenä sekä relevanssikriteereinä. Välttämättä informaatiolukutaidon ei tarvitse aina johtaa kehitykseen ja kehittämiseen, vaan oppimiseen niillä välineillä, jotka ovat jo olemassa. Omien kiinnostuksen kohteiden mukainen seuraaminen tai harrastaminen ja ajan tasalla pysyminen ovat jo sinänsä hyviä perusteita tiedonhauille ja tiedonhankinnalle. Varsinaisten aineistojen ja palveluiden kehittäminen ei ole käyttäjän vaan sisällöntuottajien vastuulla. Tiedonhakijoiden on tarkoitus ottaa selvää kiinnostuksen kohteidensa mukaisista asioista omilla relevanssikriteereillään.

6 Keskustelua ja johtopäätöksiä

Halttusen ja Iivosen mukaan (1999) Web-tiedonhakupähtymää on perusteellisesti kuvattu suorituksena ja sen merkityksenä alusta loppuun. Tärkeimpiä johtopäätöksiä heidän artikkelistaan tämänkin tutkimuksen kohdalla on saada tietoa niistä menetelmistä ja relevanssikriteereiden mukaisista tavoista, joilla ihmiset käyttäytyvät hakiessaan webistä tietoa. Heidän mukaansa tekniikkaa täytyy muuttaa vastaamaan paremmin ihmisten tapaa toimia. Tutkimukseni aiheeseen liittyvä taustatyö oppimisympäristöjen rakentamisessa tai Internet-sivustojen suunnittelussa on työlästä aiheen parissa nyt ja tulevaisuudessa

työskenteleville. Moniulotteisen relevanssiarvioinnin kehittäminen on myös alkutekijöissään tämän tutkielman ollessa vain opinnäyte.

Kenenkään ei ole luultavasti mahdollista tietää, onko hakutehtävän aiheista löydettävissä kattavan sisällönanalyysin mahdollistava otos Internet-aineiston puitteissa tai tietoverkkojen ulkopuolisissa kokoelmissa, käytännössä tai teoriassa. Päällekkäisyydet tutkituissa aineistoissa tai mahdolliset virhetulkinnat sisällönanalyysissä ovat riippumattomia asioiden todellisuudesta, joka on olemassa sellaisenaan tutkimusmenetelmästä huolimatta. Itselleni niin kuin varmaankin gradunohjaajalle kyseessä on ollut subjektiivisen suhteellisuudentajun koetteleminen sen suhteen, milloin olen ollut valmis pro gradun loppuunkirjoittamiseen ja -tutkimiseen.

Tutkimuksen tuloksena ovat lisäkysymykset, mitä nyt? Uudet tutkimukset voivat olla ehdotuksia jatkotutkimukselle www-tiedonhaun kehittämiseen teoriassa, käytännössä tai tietopalvelu - ja kirjastoalan kenttätyössä. Yhteydenpito ja yhteistyö tiedontuottajien tai tiedonvälittäjien kesken voi aiheuttaa tuloksellisuusraporttien laatimisen tarvetta tai kehitysehdotuksia ja niiden käynnistämisen ja toimivuuden raportointia yksikkökohtaisesti erilaisissa tietopalveluyksiköissä. Tuloksena voi löytyä puutteellisuuksia heterogeenisen aineiston suunnittelussa, julkaisemisessa ja niiden haettavuudessa www-ympäristössä. Tiedontuottajilta on voinut jäädä ottamatta huomioon erilaiset tai eri-ikäiset käyttäjäryhmät tiedonjulkaisuvaiheessa. Jos lapset eivät pysty tai halua käyttää heille tarkoitettuja tietokantoja, niiden suunnittelussa voi olla puutteellisuuksia. Ne voivat olla liian monimutkaisia tai ne eivät ole tarpeeksi kiinnostavia.

Rieh (2002) on ehdottanut jatkotutkimukselle suuntaa toteuttaa web-järjestelmässä sellaisia käyttöliittymäpiirteitä ja toiminnallisuuksia, joita tässä tutkimuksessa tunnistettiin. Niiden perusteella pitäisi evaluoida kyseisen uuden systeemin tehokkuutta ja käytettävyyttä. Alan tutkimuksella on mahdollisuuksia laajentaa ymmärrystämme tiedon laadun ja pätevyuden arviointi- ja päätöksenteon prosesseista ja sen tulisi vähitellen johtaa sellaisten web-systeemien suunnitteluun, jotka tehokkaasti tukevat näitä prosesseja silloin, kun tietoa haetaan.

Tiedonhakujärjestelmien metadatan, jota käytetään hakukoneiden yhteydessä, tulisi kehittää ottamalla huomioon relevanssikriteerien todellinen hyötykäyttö. Sellaisen metadatan tulee

olla käyttäjakeskeinen perustuen käyttäjien omaan tietoon sekä kontekstisidonnainen sopien käyttäjien tilanteisiin ja tehtäviin.

Ovatko vaihtoehtoiset ja samoja asioita päällekkäisesti julkaisevat sivustot haitallisia esteitä rationaaliselle, laadulliseen tarkkuuteen perustuvalla tiedonhauulle? Merkityksellistävätkö ne sittenkin tiedonhankinnan lähtökohtien motivointia ja antavatko ne laajempia mahdollisuuksia tiedonlähteinä monien eri tuottajien näkökulmista, julkaisumuotojen ansioina tai riskitekijöinä? Onko Internet-aineiston julkaisukriteereitä syytä nostaa ja tallettaa sinne unohtuneet aineistot kansainvälisellä ohjelmalla, jos joku niitä jälkeinpäin vielä kyselee tai muistelee? Onko painetun aineiston kaltainen bibliografinen kontrolli otettava käyttöön kaikille palvelintunnusten hankkijoille lupana sisällönkuvailla aineistoja informaatiotutkimuksellisin dokumentointeina tai relevanssiarviointimenetelmin? Erilaisia teemaullottuvuuksia painottavilla tai sivuavilla aineistoilla voidaan esittää kuitenkin käsiteltävistä ilmiöistä sellaisia asioita, joita on tarkoitus tuoda esille. Voidaanko eri julkaisijayhteisöjen katsoa myös tukevan toisiaan silloin, kun toisten organisaatioiden resurssit eivät ole riittäviä viestintään ja tiedottamiseen. Eikö paikallisten kokoelmien ja verkostoyhteistyön sekä yhteistyön tavoitteena ole konsultointi, avun tai lisätiedon antaminen silloin, kun joku organisaatio on kehitys- tai muutostilassa tai tavoittamattomissa kirjastojen ja tietopalvelunkin osalta?

Internetissä bibliografinen kontrolli on alkutekijöissään. Wikipedia on pitkälle kehittynyt yritys Internetiä ja aihe- tai käsiterelevanssia yhteenlinkittävänä käyttöliittymänä, jossa tiedonkondrolloijina ovat Internetin vapaaehtoiset käyttäjät. Vaikka Wikipedia ei perustukaan kirjoittajalähtöisiin vastuullisiin tekijöihin, se perustuu yhteisölliseen kontrolliin järjestelmänä, jonka tarkoitus on esittää aihelevansseja sisällönkuvauksia Internetissä riippumatta painetuista aineistoista. Asioiden linkitettävyyys käsite käsitteeltä tai täydentävät lisätietoviitteet painettuun aineistoon tai linkit suoraan muihin Internet-osoitteisiin ovat selkeästi Internetin kondrolloimista. Harrastelijapohjalta toimiva järjestelmä tuottaa tietoa, mutta sen uskottavuutta ja luotettavuutta Internetissä lisäksi se, jos puolueettomuuden vaatimus voitaisiin osoittaa niin, että puolueelliset mielipiteet tai asiat esitettäisiin selkeästi näkökulmina käsiteltävään ilmiöön.

Kirjastojen oma bibliografinen kontrolli liittyy kirjastojen kokoelmapolitiikkaan ja kokoelmien mukaisten tietojen kuvaukseen ja esittämiseen kirjastokohtaisissa tai yhteisissä

tietojärjestelmissä ja hakutoiminnoissa sekä sijaintitietojen ja muiden viitetietojen mukaisissa tuloksissa. Tämä ei suoraan tarkoita sitä, että jo elektronisoitujen Internetissä sijaitsevien kirjastojärjestelmien painettujen aineistojen kokoelmiin kohdistuvat vastaukset olisivat ainoita ratkaisuja tiedontarve- ja tiedonhakukysymyksiin. Kirjastojärjestelmien saatavuus- ja sijaintitiedot voivat olla aiheriippuvaisia, mutta eivät välttämättä aiheriippuvuuden ylittäviä korkeasti relevantteja vastauksia tiedontarvetilanteeseen tai tiedonhakijan tehtävän mukaisen relevanssin vaatimuksien mukaan riittäviä. Täsmäyttävät tiedonhakumekanismit tavallisissa kirjastojärjestelmissä samoin kuin Internetissä kohdistuvat siihen kokoelmaan, josta tietoa etsitään. Tällöin eivät järjestelmistä riippumatta löydy sellaiset aineistot, jotka ovat poistettu kokoelmista, joita ei ole hankittu kokoelmiin tai niitä ei siellä muuten vielä ole. Tietojärjestelmäkohtaiset tiedonhaukset eivät pysty myöskään löytämään aineistoja sellaisista kokoelmista joihin kyseinen tietojärjestelmä ei yllä. Tietojärjestelmä ei pysty ilman ammattitaitoista henkilötyövoimaa myöskään välittämään tietoa sellaisille tiedontuottajille, joiden mukaista aineistoa aihe- tai käyttäjärelevanssin kannalta tarvitaan juuri tietyssä muodossa tai laajempaan yhteenliitettynä kokoelmana erilaisiin tiedontarve- tai työtehtävälanteisiin.

Kirjastokokoelmajärjestelmien linkitettävyyden muihin aihelevansseihin aineistoihin on kuitenkin samankaltaista kuin Internetissä. Tällöin tuloksellisuus kirjastoissa riippuu monista tekijöistä ja yhteydenpidosta tiedon tuotannosta tiedonvälittämiseen ja tiedon loppukäyttämiseen. Tuloksellisuus kirjastoissa ei lähde tällöin lainausluvuista tai asiakasmääristä vaan informaatiotutkimuksen liian helposti ohiluettavasta perustavoitteesta, joka tulee vastaan vasta informaatiotutkimuksen syventävien opintojen perusteella. Informaatiotutkimus tarkastelee tiedon hankintatapoja ja tiedonkäyttöä eri tilanteissa sekä tiedon organisointia, välittämistä ja säilyttämistä. Sen tavoitteena on kehittää käsitteitä, menetelmiä ja järjestelmiä, joiden avulla tieto saadaan vaivattomasti sitä tarvitsevien ulottuville helposti omaksuttavassa muodossa.

Tiedonhakututkimuksen osa-alueen opiskelun perimmäisenä tavoitteena on kehittää käsitteitä, menetelmiä ja järjestelmiä, joiden avulla dokumenttien kaikki sisältö, olipa se missä tahansa muodossa ja missä tahansa paikassa, saadaan vaivattomasti kenen tahansa sitä tarvitsevan ulottuville mahdollisimman hyödyllisessä ja helposti omaksuttavassa muodossa.

LÄHDELUETTELO

I Julkaistut painetut lähteet

Alaterä, A. & Halttunen, K. 2002. Tiedonhaun perusteet - osa lukutaitoa. Helsinki: BTJ Kirjastopalvelu.

Borlund, P. 2000. Evaluation of interactive information retrieval systems. Åbo Akademi University. Akademisk avhandling. Åbo: Åbo Akademi University Press.

Crystal, A. & Greenberg, J. 2006. Relevance criteria indentified by health information users during Web searches. Research Articles Source. Journal of the American Society for Information Science and Technology. Volume 57. (10). 1368 - 1382.

Ekholm, K. 2002. Tieteelliset kirjastot sisältöuottajina. Teoksessa A. Haasio ja J. Piukkula. (toim.) Tietoverkot & kirjastot. Helsinki: BTJ Kirjastopalvelu, 100.

Hakala, J. 2002. METADATA: käyttökohteet ja formaatit. Teoksessa A. Haasio ja J. Piukkula. (toim.) Tietoverkot & kirjastot. Helsinki: BTJ Kirjastopalvelu, 117-118.

Halttunen, K. & Iivonen, M. 1999. Web-tiedonhaun tarkastelua. Aiemman version korvaava eripainos. Informaatiotutkimus 18 (2).

Hirsjärvi, S. & Remes, P. & Sajavaara, P. 2005. Tutki ja kirjoita. 11. painos. Helsinki: Tammi.

Ingwersen, P. 1995. An Introduction to Algorithmic and Cognitive Approaches for Information Retrieval. Libri; 45 (3/4) Sep/Dec 95. pp.160-177.

Järvelin, K. & Kekäläinen, J. 1998. Tiedonhaun menetelmät. Luentomateriaali. Slk-1998. Informaatiotutkimuksen laitos. Tampereen yliopisto. (Aineistot tekijöiden hallussa.)

- Järvelin, K. & Kekäläinen, J. 2002. Evaluating Information Retrieval Systems Under the Challenges of Interaction and Multidimensional Dynamic Relevance. Proceedings of the 4th CoLIS Conference. Greenwood Village, CO: Libraries Unlimited, pp. 253-270.
- Kekäläinen, J. 2002. Tiedonhaun menetelmäluento. Kik-2002. Informaatiotutkimuksen laitos. Tampereen yliopisto. (Aineisto tekijän hallussa.)
- Kekäläinen, J. & Sormunen, E. 2003. Tiedonhakujärjestelmien evaluoinnin luennot. Kik-2003. Tampereen yliopisto. Informaatiotutkimuksen laitos. (Aineistot tekijöiden hallussa.)
- Lowe, A. 1996. An Explanation of Grounded Theory. European Union Post Doctoral Fellow. Swedish School of Economics and Business Administration. Helsinki, Finland. Working Paper. December 1996.
- Mäkinen, I. & Pennanen, M. & Tanni, M. 2004. Ohje kandidaatintutkielman tekemiseen. Päivitetty 1.9.2004. Tampereen yliopisto. Informaatiotutkimuksen laitos.
- Rieh, S. Y. 2002. Judgment of Information Quality and Cognitive Authority in the Web. Journal of the American Society for Information Science and technology, 53.(2).145-161.
- Savolainen, R. 1998. Tiedontarve- ja tiedonhankintatutkimuksen luennot. Kik-1998. Tampereen yliopisto. Informaatiotutkimuksen laitos. (Aineisto tekijän hallussa.)
- Savolainen, R. 1999. Tiedontarpeet ja tiedonhankinta. Teoksessa I. Mäkinen (toim.) Tiedon tie - johdatus informaatiotutkimukseen. Helsinki: BTJ Kirjastopalvelu, 73-109.
- Sormunen, E. 1994. Vapaatekstihaun tehokkuus ja siihen vaikuttavat tekijät sanoma lehtiaineistoa sisältävässä tekstikannassa. [The effectiveness of free-text searching in full-text databases containing newspaper articles and abstracts]. Espoo: Technical Research Centre of Finland, Research Publications 790. Lisensiaatintyö.

- Sormunen, E. 1998. Tiedonhaku avoimissa verkkoympäristöissä. Internet-tiedonhaun luento. Luento slk -98. Tampereen yliopisto. Informaatiotutkimuksen laitos. (Aineisto tekijän hallussa.)
- Sormunen, E. 2002b. Liberal Relevance Criteria of TREC - Counting on Negligible Documents? SIGIR '02. August 11-15, 2002. Tampere, Finland. University of Tampere
- Sormunen, E. 2003b. Tiedonhakututkimuksen luennot. Luennot klk-2003. Informaatiotutkimuksen laitos. (Aineisto tekijän hallussa.)
- Tombros, A. & Ruthven, I. & Jose, J. M. 2005. How Users Assess Web Pages for Information Seeking. *Journal of the American Society for Information Science and Technology*. 56(4):327-344.
- Vakkari, P. & Hakala, N. 2000. Changes in Relevance Criteria and Problem Stages in Task Performance. *Journal of Documentation*, vol. 56, no. 5, pp. 540-562.

II Julkaistut elektroniset lähteet

- Informaatiolukutaidon osaamistavoitteet yliopisto- ja korkeakouluopetuksessa. 2006. Information Literacy Competency Standards for Higher Education. 2000. Association of College Research Libraries. suom. HY. Kielikeskus (tark. 2001 Eija Suikkanen, 2006 Anne Lehto) Saatavilla:
<<http://www.helsinki.fi/opiskelijakirjasto/hankkeet/arkisto/infoluku/osaamistavoitteet.htm>>. Luettu 30.10.2006.
- Heikkinen, R. & Agander, A. & Ijäs, E. & Laitinen, M. 2005. Tiedonhakijan teho-opas. Docendo Finland. Elektroninen aineisto. E-kirja. Kirkes-kirjastot. Ellibs. Saatavilla [www-muodossa](http://www.muodossa): 24.10.2006. Luettu: 24.10.2006.
- Sormunen, E. 2003a. Informaatiolukutaito tiedonhaun tutkimuksen haasteena. Tiivistelmä informaatiotutkimuksen professori Eero Sormusen virkaanastujaisesityksestä Tampereen yliopistossa. Tampereen yliopisto tiedottaa. Tulostettu 22.2.2003.

Saatavilla: <<http://www.uta.fi/ajankohtaista/tiedotteet/2003/1803c.html>>.

Luettu 1.11.2006.

Kautto, V. 2003. Mitä vaaditaan informaatiolukutaitoiselta: lisäys. Email tiedon-
hallinta@helsinki.fi 25.4.2003. Tulostettu 25.4.2003.

III Painamattomat lähteet

Sormunen, E. 2002a. Multidimensional Relevance Assesments in Test Collections for
Interactive Information Retrieval. University of Tampere. Department of
Information Studies.

Sormunen, E. 2006. Multidimensional Relevance Assesments: a method for the
estimation of redundancy in document contents. University of Tampere.
Department of Information Studies.

LIITTEET

Liite 1: Alkuperäinen tutkimusaineisto Henkilötiedonhakutehtävien mukaiset URL-osoitteet Internetissä:

Hautala, Heidi. Kaikki alla olevat osoitteet saatavilla ja luettu: 12.11.2006

<<http://www.heidihautala.net/>>.
<<http://www.heidi2006.fi/>>.
<http://www.livejournal.com/users/heidi_hautala/>.
<<http://www.yle.fi/vaalit/ehdokkaat/id20180.html>>.
<http://fi.wikipedia.org/wiki/Heidi_Hautala>.
<<http://www.city.fi/lehti/article.php?id=1728>>.
<<http://www.iltasanomat.fi/uutiset/erikoissivut/Presidentinvaalit+2006/1083426/108442>>.
<<http://www.vihrealiitto.fi/EU96/HeidiHautala.html>>.
<http://www.helsinginvihreat.fi/tiedotteet/heidi_hautala_14.html>.
<<http://koti.mbnet.fi/neptunia/politics/hautala.htm>>.
<<http://www.bioturva.org/kann/kann002.htm>>.
<<http://www.ristopakarinen.com/suomi/hautala.htm>>.
<http://www.prokarelia.net/xx/index.php?x=artikkeli&article_id=629&author=10>.
<http://www.helsinginvihreat.fi/tiedotteet/heidi_hautala_11.html>.
<<http://www.apu.fi/i ihmiset/article122987-1.html>>.
<<http://www.vihrealanka.fi/hautala.html>>.
<http://www.verkkouutiset.fi/arkisto/Arkisto_1999/17.joulukuu/heha5099.htm>.
<<http://www.kolumbus.fi/suomi-tshetszenia-seura/heidi.htm>>.
<<http://www.helsinginsanomat.fi/uutiset/tuoreet/artikkeli/1101981587556>>.
<http://www.aga.fi/international/web/lg/fi/like/gagafi.nsf/docbyalias/again_405_heidi_hautala>.
<http://www.kepa.fi/kumppani/arkisto/2002_3/2425>.
<<http://www.sumuvuori.net/hautala.html>>.
<<http://virtual.finland.fi/elections/president2000/english/hautala.html>>.
<http://www.fczb.de/projekte/wid_db/Interviews/Eu/Hautala.html>.
<<http://www.eduskunta.fi/faktatmp/hetekatmp/ed352e-su.htm>>.
<http://g05.netedit.info/en/g05.aspx?sortcode=2.11.19.19&id_article=227&starting=&ending=>>.
<<http://www.motherearth.org/hermaja/su/news16.php>>.

<<http://www.statewatch.org/news/2001/jul/03hh.htm>>.
<http://www.digitoday.fi/showPage.php?page_id=14&news_id=50398>.
<http://arkisto.sll.fi/vuotos/Hautala_tiedote.html>.
<http://www.digitoday.fi/showPage.php?page_id=14&news_id=48582>.
<<http://www.sr.se/cgi-bin/sisuradio/nyheter/artikel.asp?artikel=731371>>.
<<http://www.health.fi/index.php?page=ymparisto&i=112>>.
<<http://www.itviikko.fi/uutiset/uutisalue.asp?alue=paiva&UutisID=71110>>.
<<http://personal.fimnet.fi/luonto/vapaavuotos/hvalitus.html>>.
<<http://www.om.fi/14581.htm>>.
<<http://www.turunsanomat.fi/verkkolehti/?ts=1,4:8:0:0,4:8:1:1:2005-10-28,4:336224,1:0:0:0:0:>>>.
<http://www.kuluttajat-konsumenterna.fi/docs/heidi_hautala_eu_vastaus.html>.
<http://www.verkkouutiset.fi/arkisto/Arkisto_1998/15.toukokuu/HAUT1998.HTM>.

Leppänen, E. 2005. Heidi Hautala -linkit.

Email: ulla.lintukoto@pp.inet.fi 30.11.2005. Tulostettu: 1.12.2005.

Kunnas, Mauri. Kaikki alla olevat osoitteet saatavilla ja luettu: 12.11.2006

<<http://www.maurikunnas.net/index.cfm?flashcheck=ok>>.
<<http://virtual.finland.fi/netcomm/news/showarticle.asp?intNWSAID=26186>>.
<<http://www.suomalainen.com/sk/servlets/ProductServlet?action=getAuthorProducts&authorid=351447>>.
<<http://www-fi.valitutpalat.fi/lehti/lehti9912/artikkeli02.html>>.
<<http://www.suomenkuvalehti.fi/?id=4621>>.
<<http://www.lalli.fi/lalpahaku.asp?fmita=uutinen3&kohde=1231>>.
<fi.wikipedia.org/wiki/Mauri_Kunnas>.
<http://www.helsinki.fi/~lakoma/comics/nyrok_city.html>.
<http://www.peliko.fi/mauri_kunnas.php>.
<<http://www.kvaak.fi/keskustelu/index.php?threadid=3021>>.
<http://www.lambiek.net/artists/k/kunnas_mauri.htm>.
<<http://www.perunamaa.net/ankistit/kunnas.htm>>.
<<http://agricola.utu.fi/nyt/arvos/arvostelut.php?hakukohde=&jarjestys=&haku=&sivu=9&vps=100&arvostelu=932>>.

<<http://www.kyamk.fi/~vi4lohi/>>.
<<http://www.htk.fi/hyk/sinuhe/kunnasd.html>>.
<<http://www.yle.fi/a2/juttu.php?tunnus=1347>>.
<http://www.finpop.net/gallery/mauri_kunnas.php>.
<<http://www.helsinkiabc.fi/ABCaineisto.cgi?id=2&aid=379>>.
<http://www.otava.fi/kirjailijat/kotimaiset/kunnas_mauri/fi_FI/kunnas_mauri/>.
<<http://www.ylioppilaslehti.fi/1997/121297/pukki.html>>.
<<http://dbgw.finlit.fi/fili/bff/103/kunnas.html>>.
<<http://www.rovaniemi.fi/aurora/mlk/joulu/koti3harmaa.html>>.
<<http://www.herrahakkaraisentalo.net/englanti.html>>.
<<http://www.info.uta.fi/teaching/mauku/mauku.htm>>.
<<http://www3.lahti.fi/museot/images/taide/nayttelyt/kunnas/kunnas.html>>.
<<http://www.publiscan.fi/bl12e-7.htm>>.
<<http://www.operafestival.fi/en/operas/spooky.html>>.
<<http://oulu.ouka.fi/ppm/English/doghill.htm>>.
<<http://www.satakunnankansa.fi/teema/kirja-arvostelut/1039963.shtml>>.
<http://www.laakisarjat.fi/sarjis_alb5b.html>.
<<http://www.kirja-arvostelut.com/mauri%20kunnas>>.
<<http://www.finland.si/netcomm/news/showarticle.asp?intNWSAID=44137&intIGID=62&LAN=FI&Thread=44137,42075,44711,44547,44399,44380,44304,44137,42461,42075,40265,&intThreadPosition=0>>.
<<http://www.uta.fi/FAST/FIN/FFOLK/rk-mauri.html>>.
<<http://www.espoo.fi/kirjailijat/ruotsi/kunnas.html>>.
<<http://www.ijb.de/whiteravens/wr03/finland.htm>>.
<<http://www.aikakone.org/arkisto/arvostelut/k89avaruuskirja.htm>>.
<<http://www.uta.fi/festnews/fn97/touutis/kunnas.htm>>.
<<http://www.ssn.flinders.edu.au/scanlink/normotes/vol3/reviews/kalevala.html>>.
<<http://www.savonmedusa.fi/index.php?id=953>>.
<<http://ok.otava.fi/pagesingle.cfm?book=194>>.

Leppänen, E.2005. Kunnas-sivustot.

Email: ulla.lintukoto@pp.inet.fi 7.12.2005. Tulostettu: 8.12.2005.

Ritchie, Madonna. Kaikki alla olevat osoitteet saatavilla ja luettu: 12.11.2006

<<http://www.madonna.com/>>.
<<http://www.madonnaweb.com/>>.
<<http://www.imdb.com/name/nm0000187/>>.
<<http://www.madonna-online.ch/>>.
<[http://en.wikipedia.org/wiki/Madonna_\(entertainer\)](http://en.wikipedia.org/wiki/Madonna_(entertainer))>.
<<http://www.madonnacatalog.com/>>.
<<http://music.yahoo.com/ar-256352---Madonna>>.
<<http://www.drownedmadonna.com/>>.
<<http://www.madonavillage.com/forum/>>.
<<http://www.allaboutmadonna.com/>>.
<<http://www.madonna.nu/>>.
<<http://www.vh1.com/artists/az/madonna/artist.jhtml>>.
<<http://www.hellomagazine.com/profiles/madonna/>>.
<http://www.askmen.com/women/singer/5_madonna.html>.
<<http://www.azlyrics.com/m/madonna.html>>.
<<http://www.beautifulmadonna.com/eng/>>.
<<http://www.madonnacollectibles.com/>>.
<<http://www.discogs.com/artist/Madonna>>.
<<http://www.mad-eyes.net/>>.
<<http://www.celebrityzone.co.uk/madonna/>>.
<<http://amazingcele4.users1.50megs.com/madonnabio.htm>>.
<<http://www.kidzworld.com/site/p531.htm>>.
<<http://www.lyricsdownload.com/madonna-lyrics.html>>.
<<http://lyrics.duble.com/lyrics/M/madonna-lyrics/madonna-lyrics.htm>>.
<<http://www.madonnatribe.com/>>.
<<http://homepage.ntlworld.com/alan.stuart/music/madonna/biograph.html>>.
<<http://www.celebritycd.com/madonnaciccone/>>.
<<http://www.facingthechallenge.org/madonna.htm>>.
<<http://www.celebjunction.com/artists/m/madonna/index.shtml>>.
<<http://madonna.lyrics-songs.com/>>.
<<http://www.moono.com/html/madonna/madonna-pictures.cfm>>.
<<http://www.artistdirect.com/nad/music/artist/card/0,,414944,00.html#bio>>.

<<http://www.celebrityspider.com/madonna.html>>.
<<http://marriage.about.com/cs/celebritymarriages/p/madonnaguy.htm>>.
<<http://www.lyricsondemand.com/m/madonnalyrics/>>.
<<http://popwatch.ew.com/popwatch/madonna/index.html>>.
<<http://www.galleryofcelebrities.com/madonna.htm>>.
<<http://www.canehdian.com/non/madonna/biography.html>>.
<<http://www.rotten.com/library/bio/entertainers/music/madonna/>>.
<<http://www.eggcomm.com/madonna/>>.
<<http://www.themilkfactory.co.uk/music/madonna.htm>>.
<<http://www.geocities.com/~allmadonna/>>.
<<http://www.biogs.com/madonna/>>.
<<http://www.celebwelove.com/Madonna/>>.
<<http://www.atpictures.com/madonna/>>.
<<http://www.celebritypictures.com/MadonnaCiccone/madonna1.htm>>.
<http://www.celebritywizard.com/madonna_ciccone.html>.

Leppänen, E. 2005. Madonna-linkit.

Email: ulla.lintukoto@pp.inet.fi 23.11.2005. Tulostettu: 1.12.2005.

Soubirous, Bernadette. Kaikki alla olevat osoitteet saatavilla ja luettu: 12.11.2006

<http://www.catholicpilgrims.com/lourdes/ba_bernadette_intro.htm>.
<<http://www.catholic-forum.com/saints/saintb06.htm>>.
<http://www.catholic-pages.com/saints/st_bernadette.asp>.
<http://en.wikipedia.org/wiki/Bernadette_Soubirous>.
<http://www.heiligenlexikon.de/start.html?BiographienB/Bernadette_Soubirous_Marie_Bernard.htm>.
<http://womenshistory.about.com/cs/religion/p/p_bernadette.htm>.
<<http://www.saintpatrickdc.org/ss/0416.htm>>.
<<http://www.americancatholic.org/Features/SaintOfDay/default.asp?id=1355>>.
<<http://www.ewtn.com/library/MARY/BERNADET.htm>>.
<<http://www.christianmystics.com/traditional/women/bernadette1.shtml>>.
<http://www.wau.org/resources/saints/saintarchive/bernadette_soubirous.html>.
<<http://www.marypages.com/bernadetteEng.htm>>.

<http://www.deaconlaz.org/bernadette_soubirous.htm>.
<<http://www.ascension-research.org/bernadet.html>>.
<<http://www.nmia.com/~paulos/>>.
<<http://www.cin.org/saints/gotellpr.html>>.
<<http://home1.pacific.net.sg/~wallst/interest.htm>>.
<<http://www.dublinlourdes.net/bernadette.html>>.
<http://www.lourdes-france.org/index.php?goto_centre=ru&contexte=en&id=405>.
<http://www.reference.com/browse/wiki/Bernadette_Soubirous>.
<<http://www.sainte-bernadette-nevers.com/anglais/frsainte.htm>>.
<<http://bernadette-soubirous.biography.ms/>>.
<<http://www.who2.com/bernadettesoubirous.html>>.
<<http://www.saunalahti.fi/arnoldus/lourdes.html>>.
<http://www.catholic.org/saints/saint.php?saint_id=147>.
<<http://elvis.rowan.edu/~kilroy/JEK/04/16.html>>.
<<http://www.clairval.com/lettres/en/99/y31019920998.htm>>.
<<http://www.hwcdsb.edu.on.ca/stbernad/bio.htm>>.
<<http://catholic.net/RCC/Periodicals/Faith/1998-01-02/lourdes.html>>.
<<http://198.62.75.1/www1/apparitions/<http://pr00006.htm>>.
<<http://home.hiwaay.net/~oliver/bernadette.html>>.
<<http://members.aol.com/bjw1106/marian5.htm#TARGET9>>.
<<http://www.theturf.demon.co.uk/lourdes/lourdes1.htm>>.
<<http://www.ewtn.com/library/Prayer/ROSLOURD.HTM>>.
<<http://justus.anglican.org/resources/bio/137.html>>.
<<http://www.angelfire.com/al/srcs/bsb.html>>.
<<http://www.lourdes-store.com/us/bernadette.htm>>.
<<http://www.geocities.com/Heartland/Village/7355/churchhis.htm>>.
<<http://www.theworkofgod.org/Aparitns/Lourdes/Lourdes.htm>>.
<<http://web.singnet.com/~dcc1234/para0002.htm>>.
<<http://www.monksOfadoration.org/tabjnf94.html>>.
<<http://mike.friese.com/pilgrimage/paris/bernaut.html>>.
<<http://www.essaySample.com/essay/002854.html>>.
<<http://www.medjugorjeusa.org/lourdes.htm>>.
<<http://www.catholictradition.org/>>
<<http://www.familyplayhouse.com/SB.htm>>.

<<http://www.thesacredheart.com/sts/sabern.htm>>.
<<http://www.freewebs.com/mother/lourdes.htm>>.
<<http://www.anomalies-unlimited.com/Religion/Bernadette.html>>.
<<http://www.dailycatholic.org/issue/04Apr/apr5gab.htm>>.
<<http://www.catholicpilgrims.com/index.htm>>.
<http://www.catholicpilgrims.com/lourdes/ba_bernadette_intro.htm>.
<<http://www.catholic-pages.com/>>.
<http://www.catholic-pages.com/saints/st_bernadette.asp>.
<http://en.wikipedia.org/wiki/Main_Page>.
<<http://www.saintpatrickdc.org/>>.
<<http://www.marypages.com/mainframeEnglish.html>>.
<<http://www.deaconlaz.org/>>.
<<http://www.ascension-research.org/bernadet.html>>.
<<http://www.nmia.com/~paulos/stbernadette.html>>.
<<http://www.nmia.com/~paulos/berncor.html>>.
<<http://home1.pacific.net.sg/~wallst/>>.
<<http://www.dublinlourdes.net/>>.
<<http://www.lourdes-france.org/index.php>>.

Leppänen, E. 2005. 20 sivustoa. Email: ulla.lintukoto@pp.inet.fi. 8.11.2005.

Tulostettu: 1.12.2005

Leppänen, E. 2005. Lourdes-linkit. Email: ulla.lintukoto@pp.inet.fi 21.11.2005.

Tulostettu: 1.12.2005

Liite 2:

Tarina Bernadette Soubirouksesta ja Lourdesin Madonnasta

Bernadette Soubirouksen lapsuus

Pyhä Bernadette Soubirous (7.1.1844-16.4.1879) oli lammaspaimentyttö eteläisestä Ranskasta Lourdesin kaupungista. Helmikuusta heinäkuuhun 1858, hän raportoi 18 ”Lady’n” ilmestymistä. Huolimatta katolisen kirkon epäilystä, nämä väitteet julistettiin uskomisen arvoisiksi myöhemmin kanonisoinnin aikana, jolloin hänet julistettiin katolisen kirkon pyhimykseksi. Ilmiö teki kaupungista pyhiinvaeltajien tutun kohteen, joka houkuttelee miljoonia katolilaisia vuosittain.

Bernadette oli mylläri François Soubirouksen (1807-1871) tytär. Hänen äitinsä Louise Castérot (1825-1866) oli pyykkäri. Bernadette oli vanhin kuudesta lapsesta. Kovat ajat koskettivat ranskalaista maaseutua ja perhe eli äärimmäisessä köyhyudessa. Naapureiden mukaan perhe eli sopusointuisasti, ilmeisesti luottaen rakkauteensa ja toinen toistensa tukeen sekä uskonnolliseen hartauteensa. Kaikki perheenjäsenet hakivat työtä, mitä vain pystyivät saamaan. Bernadette teki maalaistalon töitä, paimensi lampaita ja auttoi tatiään hänen majatalossaan. Hän palasi Lourdesiin tammikuussa 1858 mennäkseen vapaaseen kouluun, jota hallinnoivat kristilliset hyväntekeväisyysseurat. Hän pystyi lopettamaan opiskelunsa vastaanottaakseen ensimmäisen pyhän ehtoollisensa. Hänen kouluvaikeutensa määriteltiin yksinkertaiseksi luonteeksi, jota käytettiin myöhemmin kuvaamaan hänen viattomuuttaan, mutta siitä lähtien kun kaikille opetettiin klassista ranskaa paikallisen gasconin sijasta, on hyvin todennäköistä, että Bernadette ei ollut ainoa, jolla oli oppimisvaikeuksia. Bernadette ei pitänyt huomiosta, jonka kohteena hän oli. Siksi hän liittyi kouluun, jossa oppi lukemaan ja kirjoittamaan.

Bernadette oli sairas lapsi. Hänellä oli lapsuudessaan kolera ja hän kärsi suurimman osan elämästään astmasta. Ne, jotka haastattelivat häntä hänen paljastamistaan näyistä, ajattelivat hänen olevan yksinkertainen. Huolimatta siitä, että katolisen kirkon ja Ranskan hallituksen virkamiehet kuulustelivat häntä ankarasti, hän pysyi johdonmukaisesti tarinassaan. Hänen käyttäytymisensä tuona ajanjaksona on testiasetelma niille kaikille, jotka väittävät näkyjen ja mystisten kokemusten sekä ihmeiden olevan nyt kirkon auktoriteettien arvioitavina.

Bernadetten aikuiselämä

Bernadette liittyi Sisters of Nevers hyväntekeväisyysnunnaluostariin, kun oli iältään 22. Hän vietti suurimman osan lyhyestä elämästään siellä, työskennellen apulaisena sairaiden osastolla ja myöhemmin suntiona luoden kauniita kirjailuja alttarivaatteisiin ja messupukuihin. Vakavan astmakohtauksen aikana hän pyysi vettä Lourdesin lähteestä, ja hänen oireensa vaimentuivat, eivätkä koskaan palanneet. Kun hänen oikeaan polveensa tuli tuberkuloosi, hän ei enää parantunut. Hän oli seurannut Lourdesin pyhiinvaeltajakehitystä, mutta ei ollut mukana basilikan vihkimisessä 1876. Ilmestykset julistettiin aidoiksi 1862 ja eteläisen Ranskan Lourdesista tuli nopeasti yksi maailman suurimmista pyhiinvaeltajien kohteista. Bernadettesta tuli yksi Sisters of Notre Dame Nevers -jäsenistä ja hän kuoli heidän luostarissaan pitkän ja kivuliaan sairauden jälkeen 35 vuoden iässä. Hänet julistettiin katolisen kirkon pyhimykseksi vasta 1933, mutta ei hänen kokemiensa Lourdesin Madonna-ilmioiden vuoksi, vaan yksinkertaiselle elämälle omistautumisensa ja palvelemisensa vuoksi. Hänen ruumistaan säilytetään häviämättömänä.

”Lady”, ”Our Lady of Lourdes”, The Immaculate Conception.

Ilmiöt Bernadetten kokemana: näyt ja ihmeet

Lady Lourdes ilmestyi 18 kertaa Bernadette Soubiroukselle, nuorelle, köyhälle ja sairaalle tytölle Massabiellen luolassa, lähellä Ranskan Lourdesia vuonna 1858. 14-vuotiaana Bernadette oli keräämässä takkapuita siskonsa ja kaverinsa kanssa Massabiellen luolassa Lourdesin ulkopuolella. Sillä aikaa kun Bernadetten sisko Marie ja hänen kaverinsa

ylittivät joen etsiäkseen puita toiselta puolelta, Bernadette pysyi kuivalla maalla. Puita etsiessään hän kuuli kovan äänen, kuin myrsky olisi tulossa läheltä Massabiellen luolaa. Luolan suulla oli ruusupensas, joka liikkui ikäänkuin olisi ollut tuulista, vaikka ei ollut. Luolan sisäpuolella sisäänkäynnin kohdalla hän näki kullanvärisen pilven ja pian sen jälkeen nuoren ja kauniin naisen, joka oli asettanut itsensä ruusupensaan yläpuolelle. Nainen hymyili Bernadettelle ja pyysi häntä etenemään. Bernadette ei tuntenut pelkoa, vaikka olisi voinut, ja putosi polvilleen rukoilemaan rukousnauhan kanssa. Sen jälkeen kun molemmat olivat suorittaneet nämä rukoukset, nainen vetäytyi hitaasti luolaan ja katosi. Bernadette väitti nähneensä ensimmäisen 18 näystä, joita hän kutsui pieneksi nuoreksi naiseksi, joka seisoo kalliosyvennyksessä. Muut tytöt totesivat, etteivät nähneet mitään. Bernadette oli kuullut "tuulen äänen" ja nähnyt sitten valon, joka nopeasti valaisi silhuetin nuoren tytön ympäri "joka oli yhtä nuori ja pieni kuin hän itse"; 1,40m ja 14 vuotta! Tyttö hymyili hänelle ja pyysi rukoilemaan. Bernadette koki Marian 18 vierailua kuuden kuukauden jakson aikana saaden tietää vasta viimeisellä kerralla, kuka nainen oli. Bernadette itse sanoi, että usko ja rukoilijat paransivat sairaita. Lupauksiin tuli muutos: "Tekisitkö minulle palveluksen tulemalla tänne kahdeksi viikoksi?" Bernadette sanoi kyllä. "Minä en lupaa tehdä sinua onnelliseksi tässä maailmassa, mutta toisessa". Toinen maailma on rakkauden maailma, jonka Bernadetten oli määrä kokea maan päällä kärsimyksistä huolimatta. Kahden viikon aikana toisen maailman tavat selvenivät hänelle kääntymyksenä sekä sydämessään.

Osa Bernadetten väitetyistä näyistä olivat yksinkertaisia ja kohdistuivat rukouksen ja rangaistuksen tarpeeseen. Bernadette kertoi perheelleen, että nainen oli sanonut: "Menethän pappien luokse ja kerrot että tänne pitää rakentaa kappeli. Antaa prosessin edetä." Yhdessä kahden tätinsä kanssa Bernadette vastuuntuntoisesti meni pitäjän papin, Dominique Peyramalen, luokse pyyntöineen. Älykäs, mutta usein kovapuheinen mies Peyramale, joka ei suuresti uskonut väitteisiin näyistä ja ihmeistä, pyysi Bernadettelta, että naisen täytyisi antaa tunnistaa itsensä. Bernadette toisti papin sanat naiselle seuraavalla vierailullaan, mutta nainen kumartui vähän, hymyili eikä sanonut mitään.

Bernadette kertoi myöhemmin perheelleen, kirkolle ja siviilitutkijoille kokemuksistaan. Bernadettea syytettiin valehtelemisesta ilmestyksistä ja häntä käskettiin ottamaan selvää naisesta, kuka hän oli, ja vaatimaan häntä esittämään ihmeen laittamalla ruusupensaan luolassa kukkimaan. "I am the Immaculate Conception" Näillä sanoilla nainen vahvisti sen, minkä paavi oli julistanut virallisena katolisen kirkon oppina neljä vuotta aikaisemmin. Miten koulututtamaton 14-vuotias voisi tietää Ranskan maaseudulla katolisen kirkon faktoja?

Bernadetten mukaan, hän jälleen kysyi naisen nimeä, mutta nainen vain hymyili takaisin. Bernadetten tiedusteluihin Bernadette vastasi eteläranskalaisella Occitanin kielellä samoin kuin ennenkin: "I am the Immaculate Conception." Neljä vuotta aikaisemmin Paavi Pius IX oli kuuluttanut dogmaattisen opin virheettömästä, tahrattomasta käsitteestä. Ainoa kaikista ihmisistä, jotka olivat koskaan eläneet, vain Maria, Jeesuksen äiti syntyi ilman perisyntiä. Tämä tiedettiin hyvin katolisen uskonnon keskuudessa, mutta se ei varmasti ollut ilmaisu, jonka yksinkertainen talonpoikaistyttö olisi voinut lukea. Hänen vanhempansa, opettajat ja papit todistivat kaikki myöhemmin, että Bernadette ei ollut koskaan aikaisemmin kuullut käsitettä Immaculate Conception heiltä.

Bernadette näki Marian jälleen pääsiäisen jälkeen ja huolimatta etäisyydestä hän pystyi sanomaan: "Näin hänet juuri sellaisena kuin minä olisin ollut luolassa. Se oli ensimmäinen kerta, kun joku puhui minulle niin kunnioittavaan sävyyn." Siitä lähtien Bernadette alkoi samaistua näkemäänsä Mariaan.