



TEKNILLINEN KORKEAKOULU  
Tietotekniikan osasto  
Informaatioverkostojen koulutusohjelma

Thomas Häggström

**Toimintakeskeisen semanttisen moninäkömäähaun toteutus ja  
evaluointi kulttuurialan portaalisovelluksessa**

Diplomityö, joka on jätetty opinnäytteenä tarkastettavaksi  
diplomi-insinöörin tutkintoa varten

Espoon Otaniemessä 7.12.2007

Valvoja      Professori Eero Hyvönen

Ohjaaja      Stina Westman DI

TEKNILLINEN KORKEAKOULU		DIPLOMITYÖN TIIVISTELMÄ	
Tietotekniikan osasto			
Informaatioverkostojen koulutusohjelma			
Tekijä		Päiväys	
Thomas Häggström		7.12.2007	
		Sivumäärä	
		104 + 13	
Työn nimi			
Toimintakeskeisen semanttisen moninäkömääräyksen toteutus ja evaluointi kulttuurialan portaalisovelluksessa			
Professuuri		Koodi	
Viestintäteknikka		AS-75	
Työn valvoja			
Professori Eero Hyvönen			
Työn ohjaaja			
Stina Westman DI			
<p>Diplomityön tavoitteena on tutkia semanttisen webin tarjoamia mahdollisuuksia tiedonhaussa. Tähän tavoitteeseen pyrin toteuttamalla tiedonhakujärjestelmän, jossa lukuisista eri museoista peräisin olevat heterogeeniset aineistot ovat haettavissa yhdellä käyttöliittymällä.</p> <p>Työni kirjallisuudessa käsittelen tiedonhaun teoriaa, metadatan käyttöä ja semanttisen webin toimintaperiaatteita. Kirjallisuudessa yhteydessä paneudun tutkimusryhmässä aikaisemmin tehtyyn tutkimukseen erityisesti sisällönkuvailun ja moninäkömääräyksen osalta.</p> <p>Työssäni kehitin heterogeenisen aineiston yhdistämisen mahdollistavaa toimintakeskeistä sisällönkuvailun menetelmää ja tietomallia. Mallin varaan toteutin moninäkömääräyksen periaatteella toimivan tiedonhakujärjestelmän, jolla toimintakeskeisesti kuvailtua aineistoa voidaan hakea. Saadakseni tarkkaa tietoa tiedonhakujärjestelmän toimivuudesta ja soveltuvuudesta suunniteltuun käyttötarkoitukseen, evaluoin valmista tiedonhakujärjestelmää käyttäjäkeskeisen evaluointimallin mukaisesti. Evaluointia varten suunnittelin kokeen, johon kuuluivat hakutehtävät ja tiedon keruun menetelmät, kuten kyselylomakkeet, transaktioloki sekä videokuvan ja äänen kaappaus.</p> <p>Toimintakeskeinen tietomalli osoittautui tiedonhakujärjestelmän sovelluskehityksen aikana toimivaksi ja heterogeenisiä aineistoja yhdistäväksi tietomalliksi. Työssä kehittämäni toimintakeskeinen semanttinen moninäkömääräyksen sekä aineiston kartta- ja aikajanaprojisointi toimivat täysin toimintakuvausten ja temaattisten roolien varassa. Käyttäjät testit puolestaan todistivat tiedonhakujärjestelmän toimivaksi. Hakujärjestelmä tuki jokaista tehtävätyyppiä ja käyttäjät pitivät arvioissaan järjestelmää uudentyyppisenä, tehokkaana, toimivana ja hyödyllisenä. Vaikka käyttöliittymä tarjosi perinteisen vapaatekstihaun, käytettiin toimintakeskeistä semanttista moninäkömääräyksen jokaisen tehtävätyypin suorituksen yhteydessä. Käyttäjät olivat melko tyytyväisiä hakujärjestelmän palauttamien dokumenttien relevanssiin. Kartta- ja aikajanaprojektioita pidettiin innovatiivisina ja toimivina lisäominaisuuksina.</p> <p>Käyttäjät testien aikana kritisoitiin moninäkömääräyksen käytön vaikeutta ja näkömien sisältöä pidettiin vaikeasti ymmärrettävänä. On selvää, että moninäkömääräyksen on edelleen kehitettävä. Evaluoinnin tulosten perusteella näyttää siltä, että moninäkömääräyksen ja vapaatekstihaun soveltuvat erityyppisten tehtävien suorittamiseen ja ovat näin ollen toisiaan täydentäviä hakutapoja. Moninäkömääräyksen hyödynnettiin tehtävissä, joissa lähtötiedot eivät olleet tarkkoja. Tehtävissä, joissa tehtävänannossa annettiin tarkkaa metatietoa, oli tekstihaun hyödyllisempi.</p>			
Avainsanat			
semanttinen web, tiedonhaku, ontologiat, moninäkömääräyksen, sisällönkuvailu			

HELSINKI UNIVERSITY OF TECHNOLOGY Department of Computer Science and Engineering Information Networks Degree Program		ABSTRACT OF MASTER'S THESIS	
Author  Thomas Häggström		Date	7.12.2007
		Pages	104 + 13
Title of thesis  Implementation and Evaluation of Activity-Centered Semantic View-Based Search in a Portal Application for the Museum Domain			
Professorship  Media Technology		Professorship Code  AS-75	
Supervisor  Prof. Eero Hyvönen			
Instructor  Stina Westman, MSc.			
<p>The research objective for this Master's thesis was to study the possibilities of semantic web technologies in information retrieval. The goal was to develop a method for describing heterogeneous documents, implement an information retrieval system for retrieving such documents and eventually evaluate both the method and the system.</p> <p>The literature review concentrated on the field of information retrieval, the use of metadata for describing documents and the principals of the semantic web. Previous research in our research group provided a strong basis for this work and was thus reviewed thoroughly. Especially previous work on event based document descriptions and view-based retrieval.</p> <p>At the core of the method developed for describing heterogeneous documents is a data model centered on activity. This data model can be used to describe events relating to any document in a uniform way. A large database from varied museum collections was gathered and described using the method. In order to evaluate the applicability of the method an information retrieval system was developed and evaluated following the user-centered approach of information retrieval evaluation. The experiment conducted consisted of simulated work task situations and data collection methods such as questionnaires, a transaction log and video capture.</p> <p>During application development the activity centered model proved to be a viable model for integration of heterogeneous documents. The whole user interface was built on top of the model and the experiment proved the system functional. It supported all given types of tasks and the test subjects found the application novel, powerful, functional and useful. Although the user interface provided traditional text retrieval, the activity-centered view-based retrieval features were used in conjunction with each different type of task. The test subjects also found the relevance of the search results good. During the evaluation several usability problems were encountered. It is clear that semantic view-based retrieval needs to be further developed. Based on the results of the evaluation it seems that text retrieval and view-based retrieval are suitable for different types of tasks and thus complement each other. View-based retrieval was used in task where the initial information was vague. Tasks that provided precise metadata were solved using text retrieval.</p>			
<p>Keywords semantic web, information retrieval, ontology, view-based search, document description</p>			

## Alkusanat

Tämä diplomityö on tehty Teknillisen korkeakoulun viestintäteknikan laboratoriossa semanttisen laskennan tutkimusryhmässä osana FinnONTO-projektia.

Haluan kiittää professori Eero Hyvöstä inspiroivasta aiheesta. Kiitokset työn ohjaajalle Stina Westmanille asiantuntevasta ohjauksesta, palautteesta ja tuesta työn eri vaiheissa.

Haluan myös kiittää koko SeCo-tutkimusryhmää, erityisesti työhuonetovereitani Tuukka Ruotsaloa ja Tomi Kauppista iloisesta työilmapiiristä.

Suuri kiitos Milkalle kannustuksesta ja myötäelämisestä.

Helsingissä 6.12.2007

Thomas Häggström

# Sisällysluettelo

1 Johdanto .....	1
1.1 Tutkimuksen rajaus ja tutkimusongelma .....	2
1.2 Työn rakenne .....	4
2 Tiedonhaku .....	6
2.1 Tiedontarve .....	7
2.2 Tiedonhakupöytäkirja .....	9
2.2.1 Tietomallit .....	10
2.2.2 Kyselyt .....	10
2.3 Tiedonhakupöytäkirjojen evaluointi .....	11
2.3.1 Laboratoriomalli .....	12
2.3.2 Käyttäjakeskeinen malli .....	13
2.3.3 Yhteisöllinen benchmarking .....	14
2.3.4 Käytettävyys .....	15
2.4 Yhteenveto tiedonhausta .....	17
3 Metadata ja sisällönkuvailu muistiorganisaatioissa .....	18
3.1 Metatiedon esittämismuodot .....	19
3.2 Toimintakeskeinen sisällönkuvailu .....	21
4 Semanttinen web .....	24
4.1 Pohjapalvelut: Internet- ja XML-tasot .....	25
4.2 Sisällönkuvailu: RDF-taso .....	25
4.3 Yhteiset käsitteet: ontologia-taso .....	27
4.4 Sovellukset: logiikka ja päättely .....	30
4.5 Totuuden jäljillä: luottamus .....	30
4.6 Yhteenveto semanttisesta webistä .....	31
5 Semanttinen moninäkökulma .....	32
6 KulttuuriSampo-portaali .....	36
7 KulttuuriSammon toteutus .....	40
7.1 Tietomalli ja aineiston käsittely .....	40
7.2 KulttuuriSammon käyttöliittymä .....	47
7.2.1 Toimintakeskeinen semanttinen moninäkökulma .....	47
7.2.2 Hakujen tekeminen .....	55
7.2.3 Hakutulosten esitystavat .....	57
7.3 KulttuuriSammon sovellusarkkitehtuuri .....	62
7.4 KulttuuriSammon toteutuksen yhteenveto .....	64
8 Järjestelmän evaluointi .....	65
8.1 Kokeen rakenne ja koejärjestelyt .....	65
8.1.1 Koehenkilöt .....	65

8.1.2 Kerätty aineisto .....	66
8.1.3 Kokeen kulku .....	67
8.2 Tehtävät .....	68
9 Tulokset .....	72
9.1 Hakutehtävien suoritus .....	72
9.1.1 Hakumäärät .....	72
9.1.2 Tehtäväsuoritusten kestot .....	73
9.1.3 Hakutavat .....	74
9.1.4 Hakusanat .....	77
9.1.5 Hakukategoriat ja näkymien käyttö .....	78
9.1.6 Hakujen onnistuminen ja tuloskuvat .....	81
9.2 Tehtävien arviointi .....	83
9.3 Tiedonhakujärjestelmän arviointi .....	87
9.4 Tulosten tarkastelu .....	89
9.4.1 Vastaukset tutkimuskysymyksiin .....	91
9.4.2 Tulosten arviointi ja tutkimuksen luotettavuus .....	92
10 Johtopäätökset .....	95
Lähteet .....	98
Liite A Tehtävien suoritusjärjestys koehenkilöittäin	
Liite B Alkukyselyn kyselylomake	
Liite C Tehtäväkohtainen kyselylomake	
Liite D Loppukyselyn kyselylomake	
Liite E Transaktiolokin tapahtumatyyppit	
Liite F Transaktiolokin esimerkki	

## Luettelo kuvista

2.1 Tiedonhallintajärjestelmän kaavakuva (Järvelin 1995) .....	9
3.1 Salmisen (2006) sisällönkuvailun malli .....	23
4.1 Semanttisen webin tasot eli ns. teknologiakakku (Berners-Lee 2000) .....	25
4.2 RDF:n lause-perusrakennneosan muodostavat subjekti, predikaatti ja objekti .....	26
4.3 Ontologioiden luokittelu käsitteellistämisen kohteen suhteen ja luokkien merkitys yleistettävyyteen ja käytettävyyteen (Gómez-Pérez et al. 2004) .....	29
5.1 Esimerkki semanttisen moninäkömahaun näkömäprojektioista (Mäkelä 2006) .....	34
5.2 MuseoSuomi - semanttinen moninäkömähakukäyttöliittymä .....	35
7.1 Tietomalli käsitteellisellä tasolla .....	43
7.2 Esimerkki maalauksen metatietojen ja biografian yhdistämisestä .....	44
7.3 Semanttisen rikastamisen prosessi .....	46
7.4 Esimerkki käyttöliittymän moninäkömahaun näkömästä .....	49
7.5 KulttuuriSammon hakusivu .....	56
7.6 KulttuuriSammon kohdesivu, jossa kohteen kuva, metatiedot, asiasanat ja toimintakuvaukset .....	57
7.7 Hakutulokset ryhmiteltynä kategorian <i>Lappi</i> mukaan .....	58
7.8 Hakutulokset kartalle projisoituna .....	59
7.9 Hakutulokset aikajalalle projisoituna .....	61
7.10 KulttuuriSammon arkkitehtuuri .....	63
9.1 Tehtäväkohtaiset tehtyjen hakujen keskiarvot, maksimit ja minimi .....	73
9.2 Tehtäväkohtaiset suoritusaikojen keskiarvot, maksimit ja minimi .....	73
9.3 Hakutapojen suhteellinen käyttö tehtäväkohtaisesti .....	74
9.4 Hakutapojen arviointi tehtäväkohtaisesti .....	75
9.5 Näkömien käyttö tehtäväkohtaisesti .....	80
9.6 Näkömien arviointi ja arvioiden keskihajonta .....	81
9.7 Hakujärjestelmän palauttamien dokumenttien relevanssi tehtäväkohtaisesti .....	82
9.8 Tehtävien arviointi .....	84
9.9 Tiesin tarkalleen minkälaista tietoa etsiä .....	85
9.10 Koehenkilöiden omat arviot suoriutumisestaan .....	86
9.11 Hakuprosessin stressaavuus tehtäväkohtaisesti .....	86
9.12 Hakujärjestelmän arviointi .....	87
9.13 Järjestelmällä hakeminen .....	88

## Luettelo taulukoista

3.1 Matriisirepresentaatio neljästä temaattisesta perusroolista .....	22
3.2 Tarkennetut temaattiset roolit tilanteen luonteen mukaan .....	23
7.1 KulttuuriSammossa käytettävät temaattiset roolit .....	42
8.1 Tehtävääsettelu .....	69
9.1 Tehtävissä suoritettujen hakujen määrät .....	72
9.2 Tehtävissä käytetyt hakusanat, joita käytettiin useammin kuin kerran, ja erilaisten hakusanojen kokonaismäärä .....	78
9.3 Tehtävissä käytetyt hakukategoriat .....	79
A.1 Tehtävien suoritusjärjestys .....	105



## Luettelo esimerkeistä

7.1 rdf_reachable-apupredikaatti .....	49
7.2 Aineistotyyppinäkömän kategoriapuun säännöt .....	50
7.3 Aineistotyyppinäkömän ripustussäännöt .....	50
7.4 Toimijat-näkömän kategoriapuun säännöt .....	50
7.5 Toimijat-näkömän ripustussäännöt .....	51
7.6 Toimintänäkömän kategoriapuun säännöt .....	51
7.7 Toimintänäkömän ripustussäännöt .....	51
7.8 Toiminnan väline -näkömän kategoriapuun säännöt .....	52
7.9 Toiminnan väline -näkömän ripustussäännöt .....	52
7.10 Paikkanäkömän kategoriapuun säännöt .....	53
7.11 Paikkanäkömän ripustussäännöt .....	53
7.12 Aikanäkömän kategoriapuun säännöt .....	53
7.13 Aikanäkömän ripustussäännöt .....	54
7.14 Kokoelmanäkömän kategoriapuun säännöt .....	54
7.15 Kokoelmanäkömän ripustussäännöt .....	55
7.16 Karttaprojektion toimintaperiaate SPARQL-kyselynä .....	60
7.17 Aikajanaprojektion toimintaperiaate SPARQL-kyselynä .....	61
9.1 E-tehtävän tuloskuvasarjoja .....	83

## Sanasto

Ajax	Asynchronous JavaScript And XML on tekniikka vuorovaikutteisten verkkosovellusten luomiseen. Ajax-tekniikassa selainohjelma vaihtaa pieniä määriä dataa palvelimen kanssa taustalla niin, ettei koko verkkosivua tarvitse ladata uudelleen joka kerta käyttäjän tehdessä muutoksen.
DC	Dublin Core on metatietosanastostandardi informaatio-resurssien kuvaamiseen.
Iconclass	Ikonografinen luokittelusysteemi. Taiteen tulkintametsodi, jossa kuvaa tarkastellaan sen kulttuurihistoriallisessa yhteydessä.
MAO	Museoalan Ontologia. Museoalan asiasanaston (MASA) ontologisoitu versio.
Metadataskeema	Metadataskeemalla tarkoitetaan sitä rakennetta, jolla metadata esitetään.
MIR	Multimedia Information Retrieval. Tiedonhakua useasta mediatyypistä koostuvasta aineistosta.
MuseoSuomi	MuseoSuomi on prototyyppi semanttisesta yhteisportaalista ja kansallisesta julkaisukanavasta, jonka avulla Suomen eri museot ja muut kulttuurisällön tuottajat voivat julkaista yhdessä kokoelmiaan ja muita aineistojaan (semanttisessa) webissä.
OWL	Web Ontology Language on W3C:n suositteleni kieli webissä käytettävien ontologioiden määrittelemistä varten.
Portaali	Portaali on kokoava verkkopalvelu, joka tarjoaa yhtenäisen käyttöliittymän aikaisemmin erillisiin palveluihin tai tietojärjestelmiin.
RDF	Resource Description Framework on W3C:n standardoima tiedonesityskieli, joka tarjoaa aihealueriippumattoman tietomallin.
Saanti	Tiedonhaussa käytettävä käsite, joka kertoo, kuinka suuren osan kaikista relevanteista dokumenteista haku palauttaa.

SeCo	Semantic Computing Research Group on Teknillisen korkeakoulun viestintätekniiikan laboratoriossa ja Helsingin yliopiston tietojenkäsittelytieteen laitoksella toimiva Semanttisen laskennan tutkimusryhmä.
Semanttinen samoilu	Tiedonhakujärjestelmän aineistoon tutustumista assosiativisten suosittelulinkkien avulla. Vastaa hyvin pitkälle WWW:ssä "surffaamista" eli selailua hyperlinkkejä seuraamalla ilman tarkkaa päämäärää.
Tarkkuus	Tiedonhaussa käytettävä käsite, joka kertoo, kuinka suuri osa tiedonhakujärjestelmän palauttamista dokumenteista on relevantteja hakuun nähden.
TREC	Text Retrieval Conference. Tiedonhaun tutkimusta edistävä vuosittainen konferenssi.
W3C	World Wide Web Consortium on kansainvälinen yritysten ja yhteisöjen yhteenliittymä, joka ylläpitää ja kehittää WWW:n standardeja tai suosituksia.
WWW	World Wide Web on Internetissä toimiva hajautettu hypertekstijärjestelmä.
XML	eXtensible Markup Language on W3C:n standardisoima yleiskäyttöinen merkkäuskieli, joka sallii käyttäjien määrittellä omat taginsa. XML:n helpottaa rakenteellisen datan välittämisen tietojärjestelmien välillä.
YSO	Yleinen Suomalainen Ontologia. Yleisen Suomalaisen Asiasanaston (YSA) ontologisoitu versio.

# 1 Johdanto

Museot ovat kansakunnan muistiorganisaatioita yhdessä arkistojen ja kirjastojen kanssa. Museoiden tehtävänä on kuvata kaikkea inhimillistä elämää; tallentaa ja hoitaa siitä kertovaa aineistoa kuten esineitä, näytteitä, taideteoksia, rakennuksia, muinaisjäännöksiä ja muistomerkkejä sekä niihin liittyviä tarinoita. Tallentamisen ja hoitamisen lisäksi museot välittävät tietoa: museot järjestävät näyttelyitä ja opetusta, julkaisevat tutkimuksiaan ja neuvovat omaan alaansa liittyvissä kysymyksissä. (Anon 2006a)

Kansainvälinen museoneuvosto ICOM (International Council of Museums) määrittelee museon seuraavasti: *Museo on pysyvä, taloudellista hyötyä tavoittelematon, yhteiskuntaa ja sen kehitystä palveleva laitos, joka on avoinna yleisölle ja joka tutkimusta ja opetusta edistääkseen ja mielihyvää tuottaakseen hankkii, säilyttää, tutkii, käyttää tiedonvälitykseen ja pitää näytteillä aineellisia todisteita ihmisestä ja hänen ympäristöstään* (Anon 2001).

Museoiden tehtävän oleellinen osa on siis tuoda tallennettu tieto ihmisten ulottuville mahdollisimman helpolla tavalla. Museot innostuivat heti World Wide Webin synnyttyä sen tarjoamista mahdollisuuksista museotoiminnan harjoittamiseen. WWW:n välityksellä museoiden aineistot eivät enää olisi museorakennukseen sidottuja, vaan niitä voisi hyödyntää paljon laajemmin, ajasta ja paikasta riippumatta. WWW-sivujen ja verkkonäyttelyiden avulla museotoiminnan laatua voitaisiin parantaa erityisesti tutkimuksen, opetuksen edistämisen ja mielihyvän tuottamisen kannalta. Näiden visioiden valossa pidettiin jo vuonna 1997 ensimmäinen Museums and the Web -konferenssi<sup>1</sup> Los Angelesissa Yhdysvalloissa, jossa museoalan edustajat kokoontuivat pohtimaan WWW:n tarjoamia mahdollisuuksia ja esittelemään uusimpia sovelluksia. Bearman ja Trant (1997) huomioivat silloin konferenssijulkaisun johdannossa: *Oikeasti elävät WWW-sivustot tulevat heijastamaan ymmärrystä ihmisten tavoista hyödyntää museoaineistoa. Seuraavan sukupolven WWW-sivustojen on luotava tiloja, jotka tukevat toimintoja kuten vertailua ja analyysia, ja tarjottava tapa integroida usean eri instituution tarjoamaa tietoa museovieraiden määrittelemiin pakkauksiin.*

Lähes kymmenen vuotta myöhemmin museoiden WWW-sivut eivät vielä täytä Bearmanin ja Trantin esittämää seuraavan sukupolven museoverkkosivustojen määritelmää. Aineistojen vertailu ja analyysi ei ole huomattavasti helpottunut, eikä museotietokantojen yhdistämisessä ole päästy kovinkaan pitkälle. Hyvin usein museoiden WWW-sovellusten käyttöliittymä on pelkkä hakukenttä ja hakua tarkentavien hakuehtojen sokkelo, jossa tuloksina esitetään kuva kohteesta ja sen metatietokentät. Museoiden aineistot ja kulttuuriperintö ovat kuitenkin kiinnostavampia ja ansaitsevat parempaa. On otettava huomioon loppukäyttäjää kiinnostavat aspektit, kuten esineiden historia ja

---

<sup>1</sup><http://www.archimuse.com/conferences/mw.html>

aineistojen punoutuminen toisiinsa. Esineiden yhtymäkohdat historiallisiin tapahtumiin ja jokapäiväiseen elämään luovat esineille elävämmän kontekstin ja syventävät kokemusta.

Semanttinen web tulee visioiden mukaan mullistamaan verkkopalvelut mahdollistamalla koneellisesti ymmärrettävän sisällönkuvailun ja sen myötä tarkemman tiedonhaun ja verkottuneemman rakenteen WWW-ympäristössä (Berners-Lee et al. 2001). Vaikka semanttista webiä tekniseltä puolelta on tutkittu paljon, on sen tarjoamien mahdollisuuksien osoittaminen suurelle yleisölle vielä tekemättä. Semanttisen webin uskottavuuden lisäämiseksi on suuri tarve tuotantokelpoisille ja hyödyllisille semanttisen webin sovelluksille, jotka toteuttaisivat visiot ja joista tulisi yleisesti käytettyjä ja hyväksytyjä.

Käytettävyyssiantuntija Kristine Faulkner (2000, s.6) väittää, että uusi teknologia on sen varhaisissa vaiheissa kykenemätön ottamaan käyttäjää huomioon ja että käytettävyyteen liittyvien asioiden tuleminen osaksi tutkimusta osoittaa teknologian kypsyttä. Semanttinen web on mielestäni nyt tässä murrosvaiheessa. Teknologiaa on kehitetty ja testattu pienissä piireissä pienin demosovelluksin. Nyt olisi kuitenkin aika pyrkiä saamaan kriittistä massaa tekniikan taakse, jotta käyttökelpoista aineistoa olisi saatavilla ja sen myötä hyödyllisiä sovelluksia olisi mahdollista toteuttaa. Seuraavaksi on mietittävä mitä semanttinen web voi tarjota loppukäyttäjille ja pyrkiä toteuttamaan näitä visioita.

Museoala ja nimenomaan kokoelmien julkaiseminen verkossa kiinnostavalla tavalla on alue, jossa olemassa olevat haasteet ja semanttisen webin visio kohtaavat. Kulttuurisisällöt ovat monimuotoisia, hajautuneita ja niiden kuvaukset keskenään yhteen sopimattomia. Erilaisten mediasisältöjen tallentamisessa on käytössä satoja erilaisia tietojärjestelmiä ja tiedonesitystapoja. Tästä huolimatta sisällöissä on lukemattomia kiinnostavia yhtymäkohtia, joita semanttinen web voisi tuoda esiin.

## **1.1 Tutkimuksen rajaus ja tutkimusongelma**

Tämä tutkimus on osa Suomalaiset semanttisen webin ontologiat (FinnONTO) 2003-2007 -projektia. FinnONTO-projekti on laaja kansallinen hanke, jonka visiona on kehittää avoin perusta suomalaisen semanttisen webin infrastruktuurille (Hyvönen et al. 2007). Projektilla on neljä tuloksellista tavoitetta: kehittää keskeisiä kansallisia ontologioita, perustaa palveluita ontologioiden kehittämiseen ja käyttöön, kehittää työkaluja semanttisen webin parissa työskentelyyn ja pilotoida kehitettyjä ontologioita, palveluja ja työkaluja portaalihankkeissa. Osana portaalihankkeita kehitetään sisällönkuvailumenetelmiä sekä näihin liittyviä haku-, samoilu- ja visualisointimenetelmiä, joissa sisällöt ja menetelmät kuvataan suhteessa ontologioihin.

Tässä työssä kuvaan semanttisen webin tekniikoille pohjautuvan museoalan portaalijärjestelmän, KulttuuriSammon, toteutuksen ja arvioin toteutetun järjestelmän

hyödyllisyyttä käyttäjätestein. KulttuuriSampo-portaalin tavoitteena on tarjota kulttuuriperinnöstä kiinnostuneille mahdollisuus hahmottaa kulttuurisisältöjen suhteita aikaisempaa laajemmassa perspektiivissä ja olla tehokas työvälinen tiedonhaussa ja tiedon jalostuksessa. Ajatuksena on, että eri lähteistä peräisin olevia kulttuurisisältöjä voidaan yhdistää ja julkaista semanttisen webin tekniikoita hyödyntäen niin, että tiedon haku on järjestelmän käyttäjille tuloksellista, helppoa ja hyödyllistä.

Työni pohjautuu Semanttisen laskennan tutkimusryhmässä (SeCo) FinnONTO-projektin ja sitä edeltäneen Semantic Web: älykkäät hakemistot -projektin aikana tehtyyn työhön ja projekteissa saavutettuihin tutkimustuloksiin. Hakujärjestelmän kehitystyössä perustana ovat olleet erityisesti kansainvälisesti palkittu MuseoSuomi-järjestelmä, semanttinen moninäkömähaku sekä toimintakuvauksiin perustuvan sisällönkuvailun tutkimuksen tulokset (Hyvönen et al. 2002; Hyvönen et al. 2004a; Hyvönen et al. 2004b; Hyvönen et al. 2004c; Hyvönen et al. 2004d; Hyvönen et al. 2004e; Mäkelä et al. 2004; Saarela 2004; Hyvönen et al. 2005; Mäkelä et al. 2005; Mäkelä 2006; Mäkelä et al. 2006; Junnila 2006; Junnila et al. 2006; Salminen 2006). Näihin aiheisiin perehdyn tarkemmin työni teoriaosuudessa luvuissa 3.2 ja 5.

Tutkimuksessani pyrin selvittämään, mahdollistaako toimintakuvauksiin perustuva sisällönkuvailu heterogeenisen museoaineiston semanttisen yhdistämisen ja hyödyllisten haku- ja visualisointijärjestelmien rakentamisen sisällönkuvailun varaan? Tästä tutkimusongelmasta nousee kaksi tutkimuskysymystä, joista toinen liittyy konstruktion rakentamiseen ja toinen sen käytöstä saatavaan hyötyyn. Ensimmäinen konstruktion rakentamiseen liittyvä tutkimuskysymys on: *Mahdollistaako toimintakuvauksiin perustuvan sisällönkuvailun tietomalli heterogeenisen aineiston yhdistämisen?* Mikäli yhdistäminen osoittautuu mallin avulla mahdolliseksi, haluan selvittää, koetaanko se myös lopullisissa käyttökohteissa hyödylliseksi. Toinen tutkimuskysymys on näin ollen: *Voiko toimintapohjaisen tietomallin päälle rakentaa käyttäjälle hyödyllisiä haku- ja visualisointijärjestelmiä?*

Lähestymistapani tutkimusongelmaan on konstruktivisen tutkimusotteen mukainen. Lukan (2000) mukaan konstruktivisella tutkimusotteella on kuusi tärkeää ominaisuutta:

- Tutkimus keskittyy sellaisiin ongelmiin, joiden ratkaisemisella on käytännön hyötyä.
- Tutkimus tuottaa innovatiivisen konstruktion, jonka on tarkoitus ratkaista ongelma.
- Tutkimus sisältää tavoitteen toteuttaa konstruktion ja arvioida sen käytännöllistä soveltuvuutta.
- Tutkimuksessa tapahtuu kokemusperäistä oppimista tutkijoiden ja kohdehenkilöiden läheisen yhteistyön ja osallistumisen ansiosta.

- Tutkimus on eksplisiittisesti linkitetty aikaisempaan teoreettiseen tietämykseen.
- Tutkimus keskittyy erityisesti empiiristen havaintojen heijastamiseen takaisin teoriaan.

Keskeinen käsite konstruktivisessa tutkimusotteessa on konstruktio, joka voi tilanteesta riippuen olla toteutukseltaan hyvinkin erilainen. Matemaattiset algoritmit ovat esimerkkejä teoreettisista konstruktioista ja lääketieteessä uuden lääkkeen kehitys tai uuden hoitomuodon kehittäminen noudattavat konstruktivista lähestymistapaa (Lukka 2001). Tässä tutkimuksessa konstruktio on KulttuuriSampo-portaali ja sen toteutuksessa tekemäni tekniset ratkaisut, joiden tarkoituksena on ratkaista heterogeenisten museoaineistojen yhdistämisen ongelma ja sillä tavoin tehdä kontribuutio tiedonhaun ja semanttisen webin tutkimuksen sekä museoaineistojen käsittelyn teoriaan.

Konstruktivinen tutkimus on kokeellista. Konstruktiota tulisi pitää testityökaluna yrityksessä havainnollistaa, testata tai jatkokehittää teoriaa tai kehitettäessä täysin uutta teoriaa. Konstruktivisessa tutkimuksessa uskotaan vahvasti, että kattavalla analyysillä siitä, mikä toimii tai ei toimi käytännössä, voi olla suurta merkitystä teoriaan. Ideaalitulos on, että ongelma ratkeaa konstruktioilla. On kuitenkin huomioitava, että myös konstruktiot, jotka eivät onnistu ratkaisemaan ongelmaa, voivat olla hyödyllisiä teorian osalta. (Lukka 2001)

Pyrin siis vastaamaan tutkimuskysymyksiin toteuttamalla toimintakuvauksiin perustuvan sisällönkuvailun päälle haku- ja selailujärjestelmän. Käytännöllinen kontribuutio tulee arvioimalla konstruktiota sovelluskehittäjien näkökulmasta sekä loppukäyttäjien näkökulmasta. Sovelluskehityksen osalta analysoin portaalin kehityksen aikana huomattuja hyötyjä ja vastaan tulleita haasteita. Konstruktiohyödyllisyyttä loppukäyttäjien näkökulmasta selvitän käyttäjätestein. Teoreettista kontribuutiota on toimintaperustaisen tietomallin arvioinnin lisäksi tutkimuksessa kehitettävä testipatteristo semanttisten portaalien arviointiin.

Tutkimukseni kuuluu informaatiotutkimuksen alaan ja tiedonhaun tutkimusalueeseen, jossa tutkitaan informaation kuvailuun, järjestämiseen ja hakemiseen liittyviä asioita. Näin ollen tiedonhaulla tutkimusalueena on iso rooli teoriaosuudessa. Koska osana tutkimustani arvioin kehittämäni järjestelmää, perehdyn myös tiedonhakujärjestelmien evaluointiin ja käytettävyyden kysymyksiin.

## 1.2 Työn rakenne

Aluksi luvussa 2 käsittelem tiedonhakua tieteellisen tutkimuksen alueena. Pohdin minkälaiset asiat tiedonhaussa yleisesti ovat ongelmia, ja miten erilaisia tiedonhaun järjestelmiä voidaan arvioida. Luvussa 3 määrittelen mitä metadata on ja käyn läpi työn kannalta keskeisiä metadatastandardeja. Metadatan yleiskatsauksen jälkeen syvennyn sisällön kuvailuun metadatan avulla, ja käyn läpi erilaisia menetelmiä ja tapoja kuvata sisältöä

formaalisti. Luvussa 4 perehdyn semanttisen webin visioon, periaatteisiin ja keskeisiin käsitteisiin. Ontologian käsitteeseen syvennyn erityisesti ontologioiden keskeisen aseman takia semanttisessa webissä ja tässä työssä. Luvussa 5 tarkastelen millaisia käyttöliittymiä ja tiedonhakujärjestelmiä metadatatalla ontologioihin sidotun tiedon hakuun ja visualisointiin on olemassa. Näistä haen ideoita KulttuuriSammon toteutusta varten.

Tämän jälkeen siirryn tarkastelemaan portaalin ja sen arvioinnin toteutuspuolta. Ensiksi esittelen tutkimusta varten toteutettuun portaaliohjelmiston aina visiosta käyttöliittymän komponentteihin asti luvuissa 6 ja 7. Käyn läpi järjestelmän toimintaperiaatteet ja suunnittelussa tehdyt valinnat. Luvussa 8 esittelen järjestelmän evaluoinnin toteutuksen. Sen jälkeen luvussa 9 esitän evaluoinnin tuottamat tulokset, analysoin tarkemmin tulosten merkitystä tutkimukseni kannalta ja arvioin tulosten luotettavuutta. Viimeisessä luvussa nostan esiin tutkimuksen keskeisimmät johtopäätökset ja esittelen mahdollisia jatkotutkimuksen aiheita.



## 2 Tiedonhaku

Tiedonhaku on yksi informaatiotutkimuksen keskeinen tutkimusalue. Informaatiotutkimuksessa tutkitaan tiedon välittymistä tiedon tuottajien ja tiedon käyttäjien välillä. Tämä tapahtuu tutkimalla yhteisöjen ja yksilöiden tiedonkäytön ympäristöjä, tiedontarpeita ja hankintatapoja sekä tiedon organisointia. Informaatiotutkimuksessa kehitetään käsitteitä, menetelmiä ja järjestelmiä tiedon välittymisen helpottamiseksi. Päämääränä on saada tarpeellinen tieto mahdollisimman vaivattomasti kenen tahansa sitä tarvitsevan käyttöön mahdollisimman hyödyllisessä ja helposti omaksuttavassa muodossa. (Järvelin & Kekäläinen 2002)

Ingwersen (1992, s. 49) on määritellyt tiedonhaun tutkimuksen seuraavalla tavalla: "tiedonhaun tutkimus tarkastelee tiedon esittämisen, tallettamisen ja etsimisen prosesseja, joita tarvitaan, jotta haluttu tieto saadaan hakutehtävän esittäneen henkilön käyttöön". Järvelinin ja Kekäläisen (2002) mukaan tiedonhaussa pyritään löytämään tiedontarpeiden tyydyttämistä mahdollisimman hyvin palveleva dokumentti tai dokumenttijoukko. Löydettyjen dokumenttien tulee olla rakenteensa, sisältönsä ja ulkoasunsa puolesta tiedontarvitsijoille sopivia ja hyödyllisiä (mt.).

Järvelinin (1995) mukaan tiedonhaun tutkimuksen peruskysymyksiä ovat:

- Kuinka kuvailla tallennettua informaatiota tai dokumentteja siten, että se olisi tai ne olisivat helposti löydettävissä?
- Kuinka järjestää informaatiota siten, että se olisi helposti järjestettävissä?
- Miten kehittää tiedonhakumekanismia siten, että ne auttaisivat tiedon löytymistä mahdollisimman tehokkaasti?

Semanttinen web liittyy keskeisellä tavalla tiedonhaun tutkimukseen. Semanttisen webin ideana on tarjota sellaiset rakenteet dokumenttien ja informaation kuvailuun, että WWW-ympäristössä sijaitseva informaatio olisi helpommin haettavissa ja hyödynnettävissä (Berners-Lee et al. 2001). Tässä työssä kehitän tiedonhakujärjestelmän suomalaisten museoiden sähköisten aineistokokoelmien hyödyntämistä varten. Tarkoitukseni on osoittaa semanttisen webin käyttökelpoisuus tiedonhakua parantavana tekniikkana. Työssäni tulen ottamaan kantaa Järvelinin esittämiin peruskysymyksiin informaation kuvailusta ja tiedonhakumekanismeista, sekä hyödyntämään tiedonhaun tutkimusperinteitä arvioidessani ratkaisujeni toimivuutta. Tässä luvussa kerron mitä tiedonhaku on ja minkälaisia asioita tiedonhaun tutkimukseen liittyy. Erityisesti keskityn tiedonhakujärjestelmiin ja menetelmiin, joilla tiedonhakujärjestelmien paremmuutta suhteessa toisiinsa voidaan arvioida. Keskityn näihin, koska haluan työssäni saada mahdollisimman realistisen kuvan kehittämäni järjestelmän toimivuudesta.

Tiedonhakuun on olemassa erilaisia näkökulmia. Teknisestä prosessinäkökulmasta tiedonhaku on vaiheita ja välineitä ja niiden toteutuksia ja toimintaa. Kognitiivisessa näkökulmassa huomio kiinnitetään haun yhteydessä esiintyviin ajattelu- ja tiedonkäsittelyprosesseihin. Evaluoivassa näkökulmassa huomio kiinnitetään tiedonhaun tuloksellisuuteen ja kustannuksiin. (Järvelin & Kekäläinen 2002)

Tiedonhaun ja tiedon tallennuksen tutkimukseen vaikuttavat useat muut tieteen alat. Näitä ovat esimerkiksi joukko-oppi ja logiikka, tietojenkäsittelyoppi ja tilastotiede, kielitiede, tiedon käytön, tarpeiden ja hankinnan tutkimus, kognitiotiede sekä psykologia, antropologia ja sosiologia (Järvelin & Kekäläinen 2002). Itse lisäisin listaan vielä käytettävyystudkimuksen. Laaja-alaisuudesta johtuen tiedonhaku on kokonaisuutena usein heikosti ymmärretty tai sitä lähestytään vain yhdestä näkökulmasta (Ingwersen 1992).

Perinteisesti tiedonhaun tutkimus on keskittynyt tekstihakuun, jossa suuresta tekstikorpuksesta pyritään löytämään tiedontarpeen tyydyttävät tekstinpätkät. Tietoa esiintyy kuitenkin kasvavassa määrin muissakin muodoissa, kuten digitaalisena äänenä ja videona. Kiinnostus muihin muotoihin kohdistuvaa tiedonhakua kohtaan on näin ollen kovassa kasvussa. Tätä tiedonhaun haaraa kutsutaan multimediatiedonhauksi. (Bertino et al. 1999, s. 325)

Multimediatiedonhaku on useasta mediatyypistä, kuten videoista, valokuvista ja tekstistä, koostuvan aineiston hakua. Multimediatiedonhaussa tutkimus jakautuu sisällön analysointiin perustuviin menetelmin ja sisällön kuvailuun perustuviin menetelmiin (Järvelin & Kekäläinen 2002). Sisällön analysointiin perustuvissa menetelmissä tarvittava tietämys pyritään keräämään lähdeaineistosta esimerkiksi analysoimalla videokuvaa koneellisesti (Lew et al. 2006). Hahmontunnistus on esimerkki sisällön analysointiin perustuvasta menetelmästä. Nykytekniikoilla esimerkiksi luonnollisista valokuvista irti saatava tieto on kuitenkin melko vähäistä (Järvelin & Kekäläinen 2002). Sisällön kuvailuun perustuvissa menetelmissä dokumentin sisältöä kuvataan käyttäen dokumentaatiokieltä (Alaterä & Halttunen 2002). Kuvailun perusteella dokumentit ovat helpommin haettavissa ja hyödynnettävissä. Tässä työssä keskityn sisällön kuvailuun perustuviin menetelmiin multimediatiedonhaussa.

## 2.1 Tiedontarve

Tiedontarpeella tarkoitetaan yksilön kognitiivista tilaa, jossa yksilö huomaa eron itsellään olevan tiedon ja tarvitsemansa tiedon välillä. Tiedontarpeen tyydyttämiseksi hän saattaa suorittaa hakuja tiedonhakujärjestelmiin tai muilla tavoin pyrkiä hankkimaan puuttuvan tiedon, esimerkiksi etsimällä sopivan tiedonvälittäjän. Aina tiedontarve ei kuitenkaan saa aikaiseksi toimintaa. Yksilö voi myös hyväksyä tiedontarpeen täyttymättömyyden ja tiedon puuttumisen. (Anon 2006b)

Tiedontarpeen käsite voidaan Taylorin (1968) mukaan jäsentää jatkumona, jossa tarpeen laatu muuttuu sitä mukaa kuinka jäsenneily tarve on. Ydintarve on taustalla oleva tarve, joka voi esiintyä esimerkiksi tyytymättömyyden tunteena. Se voi olla osaksi tiedostamaton ja näin ollen ei vielä kielellisesti ilmaistavissa. Tietoinen tarve on tiedostettu mutta ehkä vielä epämääräinen mielikuva. Tiedontarvitsija saattaa tässä vaiheessa keskustella muiden kanssa selventääkseen tarvettaan. Kun hän kykenee ilmaisemaan tarpeensa selvästi kielellisenä ilmaisuna, on kyseessä muotoiltu tarve. Sopeutettu tarve on tiedonhakupalvelua tarjoavalle henkilölle tai järjestelmälle esitetty hakutehtävä, joka on jo muokattu sen mukaan, mitä kysyjä odottaa voivansa saada palvelulta.

Hakujärjestelmälle esitetty kysely ei siis välttämättä kerro, mikä tiedonhakijan alkuperäinen ja todellinen tiedontarve on. Ilmaistut tarpeet heijastavat sitä, mitä käyttäjät uskovat järjestelmän pystyvän löytämään ja millaisia pyyntöjä se pystyy käsittelemään (Eakins & Graham 1999). Kun tiedontarve ulkoistetaan kirjoittamalla se hauksi tai kuvailemalla se, tarvitsija luo tarpeesta esityksen, joka eroaa hänen sisäisestä ymmärryksestään ja tarpeen kontekstista (Hertzum 2003). Pyynnöt eivät siis ole täydellisiä kuvauksia tiedontarpeesta.

Ingwersenin ja Järvelinin (2005, s. 290-293) mukaan tiedontarve voidaan luokitella kolmen dimension mukaan. Dimensiot ovat tiedonhakijan tarkoituksellisuus tai tavoite, tiedonhakijan nykyisen tiedon tyyppi ja tiedonhakijan nykyisen tiedon laatu. Näitä dimensioita yhdistelemällä tiedonhakijalla saattaa tiedonhaun alkaessa olla tiedontarve, joka kuuluu johonkin seuraavista kahdeksasta ryhmästä:

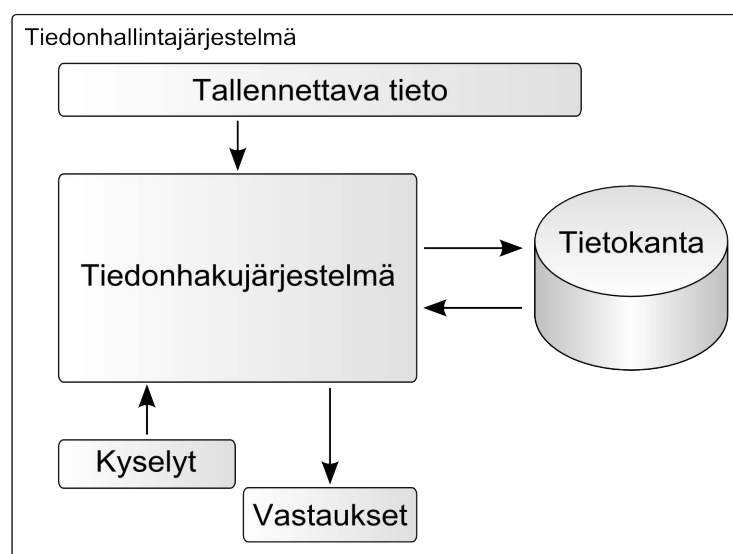
1. *Tunnetun kohteen haussa (Known item searching)* tunnetulla tiedolla pyritään löytämään dokumentti johon tieto liittyy. Tunnetun kohteen hausta on kyse, kun kirjaston tietokannasta etsitään kirjaa kirjailijan nimen perusteella.
2. *Tunnetun tiedon haussa (Known data element searching)* tunnetulla metadatalalla halutaan löytää tiedon palanen, kuten muita metatietoja. Esimerkiksi henkilön puhelinnumeron etsiminen nimen perustella kuuluu tähän luokkaan.
3. *Tunnetun aiheen tai sisällön haussa (Known topic or contents searching)* tunnetuilla sisällöllisillä attribuuteilla pyritään löytämään dokumentteja tai dokumenttijoukkoja. Tiedonhakija voi esimerkiksi olla kiinnostunut löytämään kaikki koira-aiheiset valokuvat valokuvatietokannasta.
4. *Faktatiedonhaussa (Factual data searching)* tunnetuilla sisällöllisillä attribuuteilla pyritään löytämään faktatietoa. Esimerkiksi kuka, missä ja milloin on maalannut kolmiosaisen Kalevala-aiheisen taulun?

5. *Epätarkan kohteen haku (Muddled item searching)* vastaa tunnetun kohteen hakua sillä erotuksella, että lähtötieto on puutteellista tai väärää. Tietyn kirjan hakija on esimerkiksi saattanut kuulla kirjailijan nimen väärin.
6. *Epätarkan tiedon haku (Muddled data element searching)* vastaa samoin tunnetun tiedon hakua puutteellisella tai väärällä lähtötiedolla.
7. *Epätarkan aiheen tai sisällön haku (Muddled topic or contents searching)* vastaa tunnetun aiheen tai sisällön hakua puutteellisella tai väärällä lähtötiedolla.
8. *Epätarkka faktatiedonhaku (Muddled factual data searching)* vastaa faktatiedonhakua puutteellisella tai väärällä lähtötiedolla.

Erilaiset tiedonhakujärjestelmät tukevat näiden ryhmien tiedontarpeita eri tavoin. Tunnetun tiedon hakemiseen tarvitaan erilaisia hakumenetelmiä ja kyselyjä kuin epätarkan tiedon hakuun. Edellä mainittuja tiedontarpeen luokkia hyödynnän luvussa 8 muodostaessani hakutehtäviä tiedonhakujärjestelmän evaluointia varten.

## 2.2 Tiedonhakujärjestelmä

Tiedonhakujärjestelmä on osa tiedonhallintajärjestelmää. Tiedonhallintajärjestelmä sisältää myös tiedon tallennusjärjestelmän. Kuva 2.1 esittää, miten talletettava tieto, tiedontarpeita kuvaavat kyselyt ja niihin saadut vastaukset ja hakutulokset liittyvät tiedonhallintajärjestelmän toimintaan. Uusien tietojen tallettamista tietokantaan ja siellä jo talletettujen tietojen ajan tasalla pitämistä kutsutaan päivittämiseksi. Tiedontarpeet muotoillaan kyselykielen mukaisiksi kyselyiksi, joita tiedonhallintajärjestelmä pystyy käsittelemään vastausten tuottamiseksi. (Järvelin 1995)



**Kuva 2.1. Tiedonhallintajärjestelmän kaavakuva (Järvelin 1995)**

### 2.2.1 Tietomallit

Rakenteen kannalta tarkasteltaessa tiedonhakujärjestelmien tietokantoja voidaan ryhmitellä niiden tietomallien mukaan. Tietomalli määrittelee, millaisia rakenteellisia alkioita tietokantaan voi sisältyä ja minkä tyyppisiä yhteyksiä niiden välillä voidaan muodostaa. Tietokanta koostuu aina jonkin tietomallin rakenneperiaatteiden mukaan jäsenneyistä tietoalkioista. Tietokantaa ylläpitävä tiedonhallintajärjestelmä määrää tietokannan tietomallin. Keskeisiä tietomalleja ovat esimerkiksi hierarkkinen tietomalli, relaatiotietomalli, verkkotietomalli, semanttiset tietomallit, ER-tietomalli ja oliotietomallit. (Järvelin & Kekäläinen 2002)

Semanttiset tietomallit syntyivät 1970 luvun lopulla paikkaamaan relaatiotietomallin puutteita. Semanttisten tietomallien ominaisuuksiin kuuluu, että tietokannan dataa tarkastellaan käsitteellisten olioiden kautta, oliot organisoidaan luokkiin, luokilla ja olioilla on ominaisuuksia ja suhteita kuvaavia attribuutteja, luokat ovat suhteessa toisiinsa rakenteellisten keskinäisliitäntöjen kautta ja tarjolla on valmiita perusprimitiivejä, joilla kuvata keskinäisliitäntöjä ja johdettuja attribuutteja. (Hammer & Mcleod 1981)

### 2.2.2 Kyselyt

Tiedonhakujärjestelmälle tiedontarve on muotoiltava tiedonhakujärjestelmän tarjoaman kyselykielen mukaiseksi kyselyksi (Järvelin & Kekäläinen 2002). Tässä prosessissa muotoillusta tiedontarpeesta tulee hakujärjestelmälle sopeutettu tiedontarve, joka ei välttämättä vastaa muotoiltua tiedontarvetta. Kyselykieli saattaa olla rajoittunut ja tiedontarve on tyydytettävä usealla kyselyllä tai tarvetta ei pysty ilmaisemaan lainkaan. Sopeutettua tarvetta kutsutaan myös kompromissitarpeeksi juuri kyselykielen asettamien rajoitteiden takia (Taylor 1968; Anon 2006b).

Kyselyssä käytettäviä tietoalkioiden arvoja kutsutaan hakuavaimiksi. Hakuavainten vertailua dokumenttien sisältöön kyselyä suoritettaessa kutsutaan täsmäytykseksi. (Järvelin & Kekäläinen 2002)

Tiedonhakujärjestelmiin kohdistuvat kyselyt ovat luonteeltaan joko täsmällisiä tai sumeita riippuen hakujärjestelmän palauttamien dokumenttien suhteesta kyselyyn. Kyselyt ovat täsmällisiä silloin, kun vastauksiksi haluttavat dokumentit voidaan kuvailla joidenkin ominaisuuksien mukaan yksikäsitteisellä tavalla, esimerkiksi antamalla tarkka arvo, luettelemalla joukko arvoja tai antamalla raja-arvoja (Järvelin & Kekäläinen 2002).

Täsmällisten kyselyjen vastaukset ovat täydellisiä sekä virheettömiä. Kaikki hakuavaimia vastaavat dokumentit löytyvät ja kaikki vastaukset ovat toivottuja. Esimerkiksi SQL-tietokannan relaatiosta sellaisten alkioiden hakeminen, joiden luontihetki on suurempi kuin vuosiluku 2007, palauttaa täydellisesti ja virheettömästi kantaan tallennetut

relaatiot. Täsmälliset kyselyt ovat siitä ongelmallisia, että ne karsivat armottomasti pois dokumentit, jotka melkein täsmäävät kyselyyn. Näissä dokumenteissa saattaa kuitenkin olla tiedontarpeen tyydyttämisen kannalta tärkeitä dokumentteja. Hakuavaimet ovat harvoin tarpeeksi ilmaisuvoimaisia tiedon tarpeen ilmaisemiseksi, joten niiden sumeuttaminen on hyvä keino laajentaa hakutulosta. Edellisessä esimerkissä hakutulokseen voitaisiin esimerkiksi hyväksyä myös vuoden 2007 joulukuussa luotuja alkioita.

### **2.3 Tiedonhakujärjestelmien evaluointi**

Jotta tiedonhakuvälineitä ja prosesseja voitaisiin kehittää, on oltava menetelmiä, joilla tiedonhakua voidaan evaluoida. Evaluoimalla pyritään todentamaan tietyn menetelmän toimivuus tai paremmuus suhteessa johonkin toiseen menetelmään. Tiedonhaun evaluointi voi kohdistua sekä hakujärjestelmiin, hakuihin, tietokantoihin että hakijoihin. (Järvelin & Kekäläinen 2002)

Edellä mainituista evaluoinnin osa-alueista kaikki ovat olennaisia tässä työssä kehittämäni hakujärjestelmän evaluoinnissa. Tutkimukseni kannalta pääpaino on kuitenkin hakujärjestelmässä ja tietokannassa eli tietomallissa. Niitä olisi pystyttävä arvioimaan sekä yhdessä että erillään toisistaan. Ristikkäiset vaikutukset pitäisi testijärjestelyin pystyä minimoimaan, koska evaluoitavaa tietomallia ei pidä tuomita huonon hakujärjestelmän vuoksi tai toisin päin. Käytännössä näiden osien erillään pitäminen on vaikeaa. Jotta näitä ristikkäisvaikutuksia saataisiin minimoitua ja evaluointi tehtyä mahdollisimman luotettavasti, käyn tässä alaluvussa läpi tiedonhaun evaluointimenetelmiä ja niiden soveltuvuutta erilaisiin tutkimustilanteisiin.

Evaluoinnin menetelmät jaetaan karkeasti ottaen kahteen ryhmään: niin sanottuun laboratoriomalliin ja käyttäjäkeskeiseen malliin. Molemmissa tavoite on testituloksen luotettavuus. Laboratoriomallissa evaluoinnin luotettavuus saavutetaan muuttujien kontrollilla ja testien toistettavuudella. Käyttäjäkeskeisessä mallissa sen sijaan pyritään realistisiin tuloksiin pohjaamalla evaluointi potentiaalisten käyttäjien oikeisiin tiedon tarpeisiin ja hakutapoihin. (Borlund 2003)

Hakujen evaluointikriteerit voivat koskea yksinkertaisesti hakutulosta, mutta myös vuorovaikutusta haun aikana. Vuorovaikutteinen hakujärjestelmä antaa käyttäjän muokata hakuetoja hakutulosten perusteella ja iteratiivisesti tarkentaa hakua. Vuorovaikutuksen evaluointi on tärkeää, koska se voi paljastaa seikkoja, jotka selittävät lopullisen hakutuloksen laatua (Järvelin & Kekäläinen 2002). Vuorovaikutteiset hakujärjestelmät ovat synnyttäneet käsitteen vuorovaikutteisesta tai interaktiivisesta tiedonhausta. Suurin osa nykyaikaisista hakujärjestelmistä voidaan luokitella vuorovaikutteiseksi (Saracevic 1995, s. 142).

Seuraavaksi käsittelen tarkemmin evaluoinnin laboratoriomallia ja käyttäjäkeskeistä mallia sekä niissä käytettäviä evaluointikriteereitä. Näiden evaluointimallien jälkeen kerron

benchmarking-yhteisöjen merkityksestä evaluoinnissa. Lopuksi avaan vielä käytettävyyden käsitettä ja käytettävyyden mittaamisen menetelmiä.

### 2.3.1 Laboratoriomalli

Laboratoriomallin soveltaminen on yleisin lähestymistapa tiedonhaun evaluointiin. Siinä arvioidaan järjestelmien välisiä eroja käyttäen mittareina saantia ja tarkkuutta. Saanti kertoo kuinka suuren osan kaikista relevanteista dokumenteista haku palauttaa. Tarkkuus puolestaan kertoo, kuinka suuri osa tiedonhakuprosessin palauttamista dokumenteista on relevantteja hakuun nähden. Saannin ja tarkkuuden laskemisen edellytyksenä on, että jokaisen dokumentin aihe relevanssi josta hakutehtävää kohden on arvioitu. (Baeza-Yates & Ribeiro-Neto 1999, s. 74)

Kun tämä tieto on käytössä, voidaan tiedonhakuprosessia evaluoida täsmäyttämällä hakupyyntöistä muodostettujen kyselyiden tulokset tietokannan dokumentteihin. Hakutuloksista voidaan laskea kullekin haulle sen tuloksellisuus esimerkiksi saanti ja tarkkuuskäyränä. Laskemalla keskiarvokäyrä koko testihakujoukolle saadaan selville tutkitun haku- tai indeksointimenetelmän suorituskyky. (Järvelin & Kekäläinen 2002)

Pienillä aineistomäärillä relevanssiarvioiden muodostaminen on vielä mahdollista, mutta luotettavat tulokset vaativat usein suuren määrän dokumentteja. Relevanssiarvioiden muodostaminen suuresta aineistomäärästä on kuitenkin työlästä ja siksi usein käytännössä mahdotonta. Relevanssiarviot sisältävä saantikanta muodostetaan usein puoliautomaattisesti mahdollisimman monia erilaisia menetelmiä käyttäen. Evaluoinnin tulokset perustuvat näin ollen suhteelliseen saantiin. (Järvelin & Kekäläinen 2002)

Saanti sopii parhaiten mittariksi evaluoitaessa aihehakuja, joissa tavoitteena on löytää sopiva määrä dokumentteja annetusta aiheesta. Tietokantojen käyttötavat eivät kuitenkaan rajoitu aihehakuun. Esimerkiksi organisaatioiden elektronisista arkistoista haetaan usein yhtä tiettyä dokumenttia, jonka tiedetään olevan olemassa. Haku onnistuu vain, jos dokumentti löytyy. Saannilla ei tässä tapauksessa ole merkitystä. (Järvelin & Kekäläinen 2002)

Laboratoriomallia on kritisoitu erityisesti realismin puutteesta (Ingwersen & Järvelin 2005, s. 2). Esimerkiksi vuorovaikutuksen tutkimus, jossa hakija muokkaa hakujaan saamiensa välitulosten perusteella, on ollut muuhun tutkimukseen nähden olematonta, vaikka käytännössä hakuprosessit ovat poikkeuksetta vuorovaikutteisia (Saracevic 1995, s. 142). Laadukkaan saantikannan muodostaminen on myös monessa hakuskenaariossa dokumenttien suuren määrän takia mahdotonta (Alaterä & Halttunen 2002, s. 128-129). Näiden ongelmien lisäksi relevanssi on subjektiivista ja tilanneriippuvaista, eivätkä asiantuntijajaneelin päättämät saantikannan relevanssiarviot näin ollen aina vastaa todellisen käyttäjän relevanssiarvioita (Borlund 2003).

### 2.3.2 Käyttäjakeskeinen malli

Suuri osa hakujärjestelmistä on vuorovaikutteisia hakujärjestelmiä. Vuorovaikutteisuudella tarkoitetaan hakujärjestelmän käyttäjän ja järjestelmän välistä vuoropuhelua. Koska laboratoriomalli perustuu staattisten hakujen suorittamiseen eikä vaadi testihenkilöiden käyttämistä hakutehtävien suorittamisessa, on mallin käyttökelpoisuus vuorovaikutteisen haun evaluointimenetelmänä kyseenalainen (Ingwersen & Järvelin 2005, s. 7).

Laboratoriomalli ei myöskään ota huomioon relevanssin subjektiivista, moniulotteista ja dynaamista luonnetta. Relevanssiarvio on laboratoriomallissa binäärinen aihekohtainen ominaisuus. Dokumentti joko on tai ei ole relevantti pysyvästi. Käyttäjakeskeisen mallin ideana on, että relevanssiarvion antaa koehenkilö suhteessa senhetkiseen tiedontarpeeseensa. Tämä on tärkeää, koska tiedontarve saattaa muuttua tehtävän suorituksen aikana. Jotta vuorovaikutteisen tiedonhakujärjestelmän evaluointi olisi mahdollisimman realistista, vaatii se ihmisen interaktiota, dynaamisia tiedon tarpeen tulkintoja sekä moniulotteista ja dynaamista relevanssin arviointia. (Borlund 2003)

Ingwersen ja Järvelin (2005 s. 4-10) argumentoivat, että laboratoriomalli ja käyttäjakeskeinen malli ovat molemmat käyttökelpoisia, kunhan niitä käytetään oikeassa kontekstissa. Laboratoriomalli on heidän mukaan hyvin toimiva evaluointimalli kehitettäessä tiedonhakujärjestelmissä käytettäviä algoritmeja. Tiedonhakujärjestelmä koostuu kuitenkin muistakin vähintään yhtä tärkeistä komponenteista, kuten käyttöliittymästä ja sen toimintaperiaatteesta. Näiden arvioimisessa käyttäjakeskeiset mallit toimivat paremmin.

Borlund (2003) on kehittänyt käyttäjakeskeisen tiedonhaun evaluointikehyksen, joka pitää sisällään tarvittavat komponentit realistiseen käyttäjakeskeiseen evaluointiin. Mallin kolme komponenttia ovat: potentiaalisten käyttäjien käyttäminen testihenkilöinä, subjektiivisten ja dynaamisten tiedon tarpeen tulkintojen luominen simuloituilla työtehtävätilanteilla ja moniulotteisten ja dynaamisten relevanssiarvioiden käyttö. Näiden komponenttien lisäksi kehys tarjoaa suosituksia simuloitujen työtehtävätilanteiden hyödyntämiseksi. Kokonaisuudessaan mallin on ajateltu tarjoavan käyttäjakeskeisen tiedonhaun evaluointiin paremmin soveltuvan mallin. (Borlund 2003)

Käyttäjakeskeisen evaluointikehyksen testipuitteet pitävät sisällään aineistotietokannan, hakujärjestelmän ja tilat. Tietokannan aiheellisen sisällön ja käyttäjien yhteisten ominaisuuksien perusteella voidaan luoda tilanteeseen sopivia työtehtävätilanteita. Tehtävän suorituksista kerätään tietoa erilaisilla tiedonkeruun menetelmillä. Nämä voivat olla esimerkiksi transaktioloki, kyselylomake, haastattelu, havainnointi testin aikana tai videon ja äänen tallennus testitilanteesta. Yhdessä nämä muodostavat kokeen rakenteen. Itse kokeessa on tärkeää, että kaikki koehenkilöt saavat samanlaisen käsittelyn. Kokeen esittely, hakujärjestelmän esittely ja koejärjestelyistä kertominen on hyvä olla mahdollisimman vakioitua, jotta testin tulokset ovat keskenään vertailukelpoisia. (Borlund 2003)



Vuorovaikutusta koskevia evaluointikriteerejä ovat muun muassa komentojen ja komentotyyppien määrä, hakusanojen ja hakusanatyyppien määrä sekä virheiden ja virhetyyppien määrä. Myös hakujaksojen ja kyselyjen muuntelun määrää seurataan. (Järvelin & Kekäläinen 2002)

### 2.3.3 Yhteisöllinen benchmarking

Yhteisöllinen benchmarking tarjoaa tutkimusyhteisöille ja järjestelmäkehittäjille tiedonhaun evaluointiin soveltuvia vakioituja aineistoja ja menetelmiä. Benchmarkkauksen suomenkielinen termi voisi olla vertailukehittäminen. Yhteisöllisen benchmarkkauksen idea on asettaa kaikki järjestelmät samalle viivalle tarjoamalla testeissä käytettävät aineistot ja määrittelemällä menetelmät, joilla järjestelmiä testataan. Vakioitujen aineistojen käytön hyöty on tulosten vertailukelpoisuus muiden samaa aineistoa käyttävien hakujärjestelmien välillä. Yhteisöllinen benchmarking nähdäänkin tärkeässä roolissa tiedonhakujärjestelmien evaluoinnissa ja on tuottanut paljon arvokkaita tutkimustuloksia. (Saracevic 1995, s. 142)

Aineistoja ja menetelmiä tarjoavia benchmarking-yhteisöjä ovat esimerkiksi TREC, TRECVID, Interactive TREC ja ImageCLEF. Testatessa järjestelmiä näillä menetelmillä tulokset ovat vertailukelpoisia kaikkien muiden tutkimusten kanssa, jotka on tehty samaa materiaalia käyttäen. TREC-konferenssi on julkistanut neljä tavoitetta (Voorhees 2004):

- Kannustaa tiedonhaun tutkimusta suurilla testiaineistoilla.
- Lisätä kommunikaatiota teollisuuden ja tutkimuksen välillä luomalla avoin foorumi tutkimusideoiden vaihtoa varten.
- Vauhdittaa teknologian siirtymistä tutkimuslaboratorioista kaupallisiin tuotteisiin osoittamalla merkittäviä parannuksia tosielämän ongelmiin kohdistuvissa tiedonhaun metodeissa.
- Lisätä tarkoitustenmukaisten evaluointimenetelmien saatavuutta tutkimuslaitosten ja yritysten käyttöön sekä kehittää uusia paremmin soveltuvia evaluointimenetelmiä.

TREC tarjoaa testikokoelmia, jotka ovat toimivan hakuympäristön abstraktioita. Testikokoelma koostuu kolmesta osasta: joukosta dokumentteja, joukosta tiedontarpeita (topics) ja relevanssiarvioita, jotka kertovat mitkä dokumentit tulisi palauttaa kuhunkin tiedontarpeeseen liittyen. Jotta testikokoelma vastaisi mahdollisimman hyvin oikeaa testitapausta, on dokumenttijoukon poikkeuksetta oltava iso. TREC-kokoelmat ovat yleensä kooltaan noin 2 gigatavua eli noin 500 000 - 1 000 000 dokumenttia. (Voorhees 2004)

Multimediatiedonhaun (MIR) evaluointi on ongelmallista monesta syystä. Äänen, kuvan ja videon sisältöpohjainen haku on monimutkaista, tekijänoikeuskysymykset monesti vaikeita ja lisäksi on usein vaikea päättää, mitkä olisivat relevantteja tuloksia haulle tietyn

henkilön mielestä (Lew et al. 2006). Yhteistä multimediabenchmarking-yhteisöä, joka tarjoisi materiaalit ja määrittäisi menetelmät multimedia-aineistojen hakuun, ei vielä ole olemassa. Tämä on selkeä puute ja hidastaa multimediatiiedonhakujärjestelmien kehitystä ja vertailukelpoisuutta (Jaimes et al. 2005; Hauptmann & Christel 2004).

### 2.3.4 Käytettävyys

Käyttäjakeskeisessä tuotekehityksessä pyritään tuottamaan tuotteita ja järjestelmiä, joita käyttäjät haluavat ja pystyvät käyttämään. Tämä tehdään määrittelemällä mittarit käytettävyydelle, asettamalla hyväksyttävät raja-arvot mittareille, tuomalla käyttäjiltä saatu palaute takaisin suunnitteluprosessiin ja iteroimalla edellisiä, kunnes käytettävyystavoitteet on saavutettu. (Faulkner 2000)

ISO määrittelee käytettävyyden ergonomiastandardissaan ISO DIS 9241-11 seuraavasti: *Käytettävyys on mittaa siitä, missä määrin tuloksellisuudessa, tehokkuudessa sekä käyttäjän subjektiivisessa tyytyväisyydessä mitattuna ennalta määrätyt käyttäjät voivat hyödyntää tuotetta ennalta määrättyjen tavoitteiden saavuttamiseksi tietyssä kontekstissa* (Faulkner 2000, s. 7). Käytettävyyden tutkimus on näin ollen oleellinen osa tiedonhaun tutkimusta kun tutkitaan tiedonhakujärjestelmän soveltuvuutta siihen tehtävään, johon järjestelmä on suunniteltu.

Tuloksellisuudella tarkoitetaan, että käyttäjä pystyy suorittamaan tehtävät. Esimerkiksi järjestelmä, jonka avulla käyttäjä pystyy suorittamaan neljä viidestä tehtävästä, on tuloksellisempi kuin järjestelmä, jolla käyttäjä suoriutuu kolmesta tehtävästä. Tuloksellisuudessa ei arvioida aikaa. Tehokkuutta mitattaessa tarkastellaan aikaa, ja mitä nopeammin käyttäjä suoriutuu tehtävistä, sen tehokkaampi järjestelmä on. Käyttäjän subjektiivinen tyytyväisyys on usean asian summa. Tyytyväisyyteen vaikuttaa esimerkiksi se, kuinka hyväksyttävä järjestelmä on, kuinka mukava sitä on käyttää ja pitääkö käyttäjä yhtä järjestelmää toista parempana? (Faulkner 2000, s. 7-8)

ISO ei määrittele, millä tavalla tuloksellisuutta, tehokkuutta tai käyttäjän subjektiivista tyytyväisyyttä tulisi arvioida. Jotta näistä muodostuvaa käytettävyyttä voitaisiin arvioida, on kehitettävä mitattavia käytettävyysattribuutteja eli metriikoita. (Faulkner 2000, s. 117-118)

Tuloksellisuus voidaan hajottaa seuraaviin mitattaviin osiin (Faulkner 2000, s. 118):

- *Tehtävän onnistuneen/epäonnistuneen loppuun suorittamisen suhdeluku.* Tehtävää voidaan tarkastella kokonaisuudessaan tai osien onnistumisessa. Koska tehtävillä on yleensä osatehtäviä, ja useimmiten on hyödyllistä arvioida suoriutumista näistä osatehtävistäkin.
- *Tiettyjen komentojen tai toimintojen käytön frekvenssi.* Järjestelmät koostuvat useimmiten useista osista, jotka tarjoavat erilaisia toiminnallisuuksia. Näiden käytön mittaaminen

kertoo järjestelmän eri osien hyödyllisyydestä. Tällä tavalla saadaan selville onko järjestelmissä turhia ominaisuuksia, miksi jotain tiettyä toimintoa ei käytetä ja onko olemassa vaihtoehtoinen tehokkaampi tapa suoriutua tehtävästä? Lisäksi se saattaa kertoa onko järjestelmässä harhaanjohtavia toimintoja, joita yritetään käyttää tiettyihin toimintoihin, vaikka ne eivät sovellu siihen?

- *Käyttäjien ongelmien määrä.* Faulkner ei selitä mitä hän tarkoittaa tällä kategoriolla, mutta mielestäni Jordan (1998) avaa ongelmien ja virheiden merkitystä kuvaavasti. Hänen mukaansa virheitä voidaan tarkastella usealla tasolla. Virheiden luokittelu niiden vakavuuden mukaan on tärkeää jatkokehityksen kannalta. Virheasteikkona voidaan käyttää esimerkiksi asteikkoa pienestä katastrofaaliseen. Tällä asteikolla pieni virhe on sellainen, joka on helposti huomattavissa ja korjattavissa, vaikka se saattaa aiheuttaa käyttäjälle ärtymystä. Suuren virheen korjaaminen vie käyttäjältä huomattavan paljon aikaa ja aiheuttaa suuren määrän ärtymystä. Tuhoisa virhe estää käyttäjää suoriutumasta tehtävästä. Katastrofaaliset virheet aiheuttavat ei-toivottuja sivuvaikutuksia sen lisäksi, että ne estävät käyttäjää suoriutumasta tehtävästä.
- *Lopputuloksen laatu.* Pelkkä onnistuminen tai epäonnistuminen ei ole riittävä mittari monessa tehtävässä. Tuloksen laadulla on monesti suuri painoarvo arvioitaessa tuloksellisuutta. Lopputuloksen laatu on mittari siitä, kuinka tuloksellista tehtävästä suoriutuminen itse asiassa oli.

Tehokkuutta puolestaan voidaan arvioida mittaamalla (Faulkner 2000, s. 118-119):

- Kuinka paljon aikaa kului tietyn tehtävän loppuun suorittamiseen.
- Erilaisten toimintojen määrää (työn määrää) tehtävän loppuun suorittamiselle.
- Aikaa, joka kului tiedon etsimiseen dokumentaatiosta.
- Aikaa, joka kului ohjeiden lukemiseen.
- Aikaa, joka kului virheiden kanssa kamppailuun.

Edellä luetellut mittarit ovat lopputuloksen laatua lukuun ottamatta määrällisiä, ja niitä koskeva tietojen keruu on helppo järjestää. Ne jättävät kuitenkin laadulliset kysymykset auki. Käyttäjän mielipiteitä järjestelmästä niillä ei saada selville. Käyttäjän tyytyväisyyttä voidaan mitata kyselyllä, jossa kysytään käyttäjän mielipiteitä järjestelmää kohtaan jollain asteikolla. Asteikko voi esimerkiksi olla Likertin asteikko yhdestä viiteen, jossa yksi tarkoittaa huonoa ja viisi erinomaista. Etukäteen tavoitteeksi kannattaa asettaa esimerkiksi, että 90 prosenttia vastauksista kyselyssä ovat arvoltaan neljä tai viisi. (Faulkner 2000).

## 2.4 Yhteenveto tiedonhausta

Tässä luvussa olen kertonut tiedonhausta tutkimusalueena. Jotta tiedonhakuprosessien paremmuutta voidaan verrata ja kehitystä seurata, on käytettävä luotettavia menetelmiä järjestelmien tuloksellisuuden ja toiminnan arviointiin.

Paljon käytetyn laboratoriomallin ongelmana on se, että se ei kuvaa järjestelmän toimintaa todellisessa käyttökontekstissa. Se soveltuukin paremmin kontekstista irrotettujen algoritmien evaluointiin. Laboratoriomallissa edellytyksenä on myös hyvälaatuinen ja suurikokoinen aineistokanta, jossa kaikille kannan dokumenteille on tehty etukäteen aihekohtainen relevanssiarvio. Tällaisen saantikannan muodostaminen on monesti mahdotonta. Käyttäjäkeskeinen malli pyrkii tulosten luotettavuuteen käyttämällä potentiaalisia käyttäjiä testihenkilöinä oikeita käyttötapauksia simuloivissa koetilanteissa. Relevanssiarviot pohjautuvat tällöin subjektiiviseen relevanssiin, joka kertoo enemmän järjestelmän hyödyllisyydestä, eikä vaadi aineistokannan dokumenttien relevanssin arviointia etukäteen. Yhteisöllisen benchmarkingin idea on vakioida aineistot ja menettelyt, joilla testejä tehdään, jotta tiedonhaun tutkimustulokset olisivat keskenään vertailukelpoisia.

Käytettävyyden tutkimuksessa muodostetaan mitattavia käytettävyyssattributteja, joilla pyritään takaamaan kehitettävän tuotteen laatu. Vaikka käytettävyys ei suoranaisesti kuulu tiedonhakuun, käsittelin sitä, koska palautettujen dokumenttien relevanssin lisäksi käyttäjien mielipiteisiin järjestelmästä vaikuttavat monet muutkin asiat, kuten käyttöliittymän toteutus. Näiden huomioiminen hakujärjestelmän kehityksessä ja evaluoinnissa on oleellista.

### 3 Metadata ja sisällönkuvailu muistiorganisaatioissa

Digitaalisen informaation määrän kasvaessa, on tiedon löytäminen, järjestäminen, hallinnoiminen ja uudelleenkäyttö entistä tärkeämpää. Hyvä metatieto auttaa näissä tehtävissä. Mitä tarkempaa ja kuvaavampaa metadata on, sen paremmin relevanttia sisältöä on löydettävissä. Tämän luvun tarkoituksena on selventää metadatan käsitettä ja perehtyä metadatan käyttötapoihin museoissa. Luvun lopussa kerron mihin SeCo:ssa tutkimuksen kohteena oleva toimintakeskeinen sisällönkuvailu perustuu.

Yleisin metadatan määritelmä on dataa datasta. Haase (2004, s. 204) avaa tätä metadatan määritelmää sanomalla, että metadata on mitä tahansa dataa, joka välittää esineeseen liittyvää tietoa ilman, että esinettä itseään olisi tarkasteltava. Metadata on oletusarvoisesti alkuperäistä tietoa helpommin käsiteltävässä muodossa.

Arkielämämme on täynnä tällaista auttavaa dataa. Kirjaston hakukoneet hakevat kirjaviitteet metatietojen perusteella. Museoesineiden selitetekstit antavat meille syvemmän ymmärryksen tarkasteltavasta esineestä. Metadata onkin hyvin laaja käsite, joka tarkoittaa kontekstista riippuen erilaisia asioita. Tämän työn yhteydessä tarkastelen elektronisessa muodossa olevan metadatan käyttöä tiedonhaun parantamisen yhteydessä.

Metadatan lisäämisen prosessia kutsutaan annotoinniksi. Englanninkielinen sana annotate tarkoittaa suoraan käännettynä huomautuksilla varustamista. Dokumenttiin siis yhdistetään sitä kuvaavia huomautuksia. Huomautukset voivat olla avainsanoja, tekstikuvauksia tai rakenteellisia huomautuksia (Salminen 2006, s. 6). Rakenteellisessa annotoinnissa tarvittava ja mahdollinen metadata määritellään metadataskeemoilla, jotka määrittelevät sallitut elementit ja niiden suhteet muihin elementteihin sekä joissain tapauksissa elementtien sallitut arvot (Sutton 2004, s. 144). Semanttisessa webissä rakenteellisista annotaatioista viitataan ontologioihin ja tällöin niitä voidaan kutsua semanttisiksi annotaatioiksi (Fensel 2004).

Metadataskeemoissa pyritään siis mallintamaan tallennettavan tiedon tarve ja kehittämään skeema, joka vastaa tätä tarvetta mahdollisimman tarkasti. Monessa yhteydessä tämä lähestymistapa on toimiva ja tuottaa hyvän tuloksen. Metadata ei kuitenkaan ole ongelmatonta. Samoin kuin tietokantasuunnittelussa, on metadataskeemat määriteltävä aineiston ominaisuuksien pohjalta. On määriteltävä minkälaisia käsitteitä ja näiden keskeisiä yhteyksiä on olemassa. Tämän määrittelytyön tuloksena syntyy kuvausmalli, joka monesti kuitenkin on kompromissi kuvausvoiman ja käytettävyyden välillä. Kompromisseja tehdään esimerkiksi kenttien tarkkuudessa, kuvausmallin tarkkuustasossa ja kuvausalueessa. Yleinen ongelma metadataskeemoissa on esimerkiksi se, että annotoija näkee tärkeäksi kuvata asioita, joiden kuvaaminen skeemassa ei ole mahdollista. Tämä johtaa tilanteisiin, joissa määriteltäviä kenttiä käytetään väärin. Esimerkiksi vuosilukukenttään syötetään usein kirjaimia kuvaamaan sumeita arvoja (n. 1820, 1700-luku) tai yhden arvon tekstikenttiin

lisätään pilkulla erotettuja listoja, kun annotoitava dokumentti yhtäkkiä kuuluukin useaan kategoriaan. Täydellistä metadataskeemaa on hyvin vaikea luoda. Nämä ongelmat eivät kuitenkaan vähennä hyvän metadatan hyödyllisyyttä

Gillilandin (2000, s. 1) mukaan millä tahansa oliolla on kolme ominaisuutta, joita metadatalta voidaan kuvata: sisältö, konteksti ja rakenne. Sisältö on se, mitä olio sisältää tai mistä se kertoo, ja on oliolle ominaista. Sisältöä on esimerkiksi se, mitä jokin valokuva esittää, tai tieto siitä, onko valokuva värillinen vai mustavalkoinen. Konteksti on olioon liittyvää, mutta oliolle ulkoista tietoa. Esimerkki tällaisesta metatiedosta on sanomalehden toimitusjärjestelmän artikkelien muutosloki, josta käy ilmi kuka, missä ja milloin on viimeksi käsitelty tekstiä. Rakenne kuvaa formaaleja suhteita olion sisällä tai suhteita muihin olioihin, ja voi näin ollen olla joko oliolle ominaista tai ulkoista. Esimerkki tällaisesta metadatalta on kuvista, tekstistä ja äänistä koostuvan multimediaesityksen esityskieli. Dushayn (2002, s. 246) luokittelu metadatalle nojaa samaan kolmijakoon. Metadatan voi hänen mukaan jakaa kuvaavaan, hallinnolliseen ja rakenteelliseen metadataan sen mukaan kuvaako se kuvattavan olion sisältöä, kontekstia vai rakennetta.

Museotoiminnassa kaikista kolmesta metadatan luokasta on hyötyä. Pääpaino on perinteisesti ollut hallinnollisessa metadatalta. Hallinnollinen metadata on tärkeää museoaineiston tallentamisessa ja hoitamisessa eli tehtävissä, joiden avuksi museoiden tietojärjestelmät alun perin on suunniteltu. Museojärjestelmissä hallinnollinen metadata kertoo missä aineisto sijaitsee, missä kunnossa se on ja minkälaisia toimenpiteitä sille on suunniteltu. Sisältöä kuvaavalle ja rakenteelliselle metadatalle on kuitenkin ilmaantunut selkeä tarve, kun museot ovat havahtuneet WWW:n tarjoamiin mahdollisuuksiin museotoiminnan harjoittamisessa. Aikaisemmin sisäiseen käyttöön tarkoitettujen tietokantojen julkaiseminen ja muuntaminen verkkonäyttelyiksi houkuttelee useimpia museoita. Museotietokantojen painopisteen siirtyessä sisäiseen käyttöön tarkoitetuista järjestelmistä julkisiin verkkonäyttelyihin, on sisältöä kuvaavaan metadataan panostettava entistä enemmän. Kun museotietokannoista luodaan virtuaalinäyttelyitä, ei hallinnollisella metadatalta ole paljonkaan merkitystä virtuaalinäyttelyn kiinnostavuuden kannalta. Sisältöä kuvaava metadata sen sijaan lisää kiinnostavuutta ja on edellytys toimivalle sisältöpohjaiselle tiedonhauille. Seuraavaksi kerron minkälaisia metadatan esittämismuotoja museoissa käytetään.

### **3.1 Metatiedon esittämismuodot**

Muistiorganisaatioiden aineistojen annotointiin on kehitetty lukuisia menetelmiä. Menetelmiin kuuluu niin metadataskeemoja kuin asiasanastoja ja ontologioitakin. Rajanvedot tyyppien välillä ovat vaikeita, sillä monesti yksi standardi saattaa määrittää sekä metadataskeeman että skeemassa käytettävät arvot, eli asiasanat tai ontologiset käsitteet. Keskeisimpiä menetelmiä ovat Dublin Core, Iconclass, ABC-ontology, Committee

for Documentation Conceptual Reference Model (CIDOC CRM), Visual Resources Association (VRA), Union List of Artist Names (ULAN), Art and Architecture Thesaurus (AAT), MPEG7 visuaalisen aineiston annotointiin sekä Suomen museot online (SMOL). Näistä metadataskeemoja ovat Dublin Core, VRA ja SMOL. Ne määrittelevät ainoastaan kuvailtavien aineistojen mahdolliset ominaisuudet ja jättävät arvot vapaasti annettavaksi.

Ehkä tunnetuin metadataskeema on Dublin Core Metadata Element Set (DC), jonka tarkoituksena on olla niin yleinen joukko metadatakenttiä, että sillä voi kuvata mitä tahansa resursseja ja näin saavuttaa yhteentoimivuutta (Gilliland 2000, s. 4). Toisaalta Dublin Corella kuvailu on hyvin yksinkertaisella tasolla, eikä se salli monimutkaisten rakenteiden käyttöä. Toisessa ääripäässä kehitellään hyvinkin tarkkoja metadataskeemoja tarkoin rajattuihin käyttötarkoituksiin. Esimerkiksi tässäkin työssä monen lähteen aineistot olivat SMOL-muodossa, joka on erityisesti suomalaisille museoille suunnattu metadataskeema.

Metadatan kannalta tärkeitä referenssisanastoja ja ontologioita, jotka tarjoavat annotaatioiden arvoina käytettäviä asiasanoja ja ontologioita käsitteitä, ovat ULAN, AAT ja Iconclass (Ossenbruggen et al. 2006). Esimerkiksi Ateneumin taidemuseossa on käytössä Iconclass-luokittelujärjestelmä. Iconclass on rakenteeltaan hierarkkinen aiheenmukainen luokitusjärjestelmä, jonka avulla on mahdollista ryhmitellä taideteosten ja valokuvien aiheet systemaattisesti. Viestintäteknikan laboratoriossa osana FinnONTO-projektia asiasanastojen pohjalta luodut Museoalan ontologia (MAO) ja Yleisestä suomalaisesta asiasanastosta kehitetty Yleinen suomalainen ontologia (YSO) voivat referenssiontoologioina olla jatkossa hyvin hyödyllisiä suomalaisille museoalan toimijoille. Tässä työssä käytän sekä YSO:a että MAO:a referenssiontoologioina aineiston annotoinnissa ja hakujärjestelmän toteutuksessa.

CIDOC CRM on formaali ontologia, jonka tarkoituksena on helpottaa heterogeenisen kulttuuriperintötiedon integrointia, välitystä ja vaihtoa. Se pyrkii tarjoamaan semanttiset määrittelyt ja selvennykset, joiden avulla paikallisista hajautetuista aineistosta saadaan koherentteja globaaleja resursseja. CIDOC CRM:n näkökulma ei ole sidottu mihinkään tiettyyn kontekstiin ja CRM:n rakenteet sekä yksityiskohtien taso määräytyvät tämän periaatteen mukaan. (Crofts et al. 2005)

CIDOC CRM:n on vahva kuvattaessa identifikaatitietoa, kuten nimeä ja tyyppiä, hankinta- ja omistustietoja, kohteen fyysisten siirtojen, sijoituspaikan tai uudelleensijoituspaikan tietoja, fyysisiä ominaisuuksia, kuten materiaalia, kokoa ja tunnusmerkkejä, ja tapahtumia, kuten esimerkiksi historiallisia tapahtumia tai mitä dokumentille on tapahtunut sen olemassaoloaikana. CIDOC CRM ei kuitenkaan sovellu kovin hyvin digitaalisten sisältöjen kuvaamiseen. Se ei tue dokumentin teknisiä, formaattia kuvaavia tietoja, digitaalisen median temporaalisia, spatiaalisia ja spatiotemporaalisia ominaisuuksia, visuaalisia ominaisuuksia,

kuten pintoja tekstuureja ja muotoja, eikä audiovisuaalisen dokumentin sisällön hierarkkista tai järjestyksellistä tietoa. (Hunter 2002)

FinnONTO-projektin aikana CIDOC CRM:ää tutkittiin sillä ajatuksella, että sitä voitaisiin mahdollisesti käyttää heterogeenisten museoaineistojen yhdistämisessä. Salminen (2006) tuli tutkimuksensa perusteella kuitenkin siihen tulokseen, että CIDOC CRM ei ole kovinkaan sopiva sisällön kuvailuun, vaan soveltuu enemmän hallinnollisen ja rakenteellisen tiedon tallentamiseen. Näin ollen tutkimusryhmässä päädyttiin kehittämään uutta sisällönkuvailun menetelmää, joka soveltuisi KulttuuriSammon käyttöön. Seuraavassa alaluvussa perehdyn tämän toimintakuvauksiin perustuvan sisällönkuvailun mallin teoreettiseen pohjaan.

### 3.2 Toimintakeskeinen sisällönkuvailu

Metadataskeemat saattavat olla keskenään hyvin erilaisia. Erilaisilla skeemoilla kuvatun tiedon yhdistäminen on vaikeaa, koska käsitteet on peilattava toisiinsa ennen kuin yhdistämistä voidaan tehdä. Yhtenäisen kuvailutavan hyödyt on tunnistettu ja joitain ehdotuksia yhteiseksi kuvailutavaksi muistiorganisaatioissa on tehty, kuten CIDOC CRM, Iconclass, MPEG7 ja ABC-ontologia. Usein ongelmana on kuitenkin mallien metadataskeeman tarkkaan määritely muoto jotain tiettyä käyttötapausta ajatellen. Kun erilaisia aineistotyyppjä halutaan vertailla keskenään, ei ratkaisu enää olekaan niin toimiva, koska tarkkaan määritely metadataskeema ei voi tarjota kaikkia kuvailuun tarvittavia rakenteita.

FinnONTO-projektissa tunnistettiin tarve uudelle ilmaisuvoimaiselle sisällönkuvailun menetelmälle, jolla olisi mahdollista kuvata heterogeenisiä museoaineistoja yhteismitallisella tavalla (Junnila et al. 2006, s. 1). Menetelmää ja mallia lähdettiin tutkimusryhmässä kehittämään MuseoSuomi-portaalin KulttuuriSampo 1 -jatkoprojektin aikana (Mt. s. 1). Mallia kehitettiin kahdessa pro gradu -työssä, ja sitä testattiin KulttuuriSammon ensimmäisessä versiossa (Salminen 2006, s. 25-26).

Ratkaisumalli sai pohjansa Sowan (2000) ajatuksista, jossa esimerkiksi kuvan voidaan ajatella esittävän osaa maailmasta, hetkellistä tietystä paikassa esiintyvää tilannetta. Tilannetta rajaa toiminta ja toiminnan puitteet kertovat, missä paikassa ja ajassa toiminta tapahtuu. Sowan temaattisia rooleja, joista kerron tarkemmin seuraavaksi, tutkimalla Salminen (2006, s. 25-30) päätyi viiteen tapahtumaa määrittelevään tekijään: toimintaan, toimijaan, välineeseen, kohteeseen ja tavoitteeseen.

Luonnollisessa kielessä lauseet rakentuvat verbin ympärille, joka on toimintaa kuvaava lauseenjäsen. Ilmaisuvoimaista sisällönkuvailun menetelmää kehitettäessä on loogista perustaa menetelmä luonnollisesta kielestä tuttuihin ja ilmaisuvoimaisiin rakenteisiin. Luonnollisissa kielissä toimintaan osallistuvien roolien erot tuodaan esiin kieliopillisten



merkkauksen avulla. Merkkaukset linkittävät toimintaa kuvaavat verbit osallisia kuvaaviin substantiiveihin. Kielitieteessä näitä merkkauksia kutsutaan temaattisiksi rooleiksi. Kielitieteilijät käyttävät temaattisia rooleja kuvailemaan semantiikan ja syntaksin välistä rajapintaa. Logiikassa vastaava rakenne voidaan kuvata predikaatein, jotka yhdistävät olioita toisiinsa merkityksellisellä tavalla. (Saeed 1997; Sowa 2000)

Sowa (2000) on kehittänyt tiedon kuvailun käsitteellisiä malleja temaattisten roolien ympärille. Hänen temaattisten roolien perusjakonsa perustuu neljän ominaisuuden leikkauksiin. Tämä jako on esitetty taulukossa 3.1 käyttäen Salmisen (2006) käännöksiä (Sowa 2000, s. 506). Tässä taulukossa ohjaaja määrää prosessin suunnan joko alussa toimijana tai lopussa tavoitteena. Osallistuja puolestaan on läsnä koko prosessin ajan, mutta ei aktiivisesti kontrolloi tapahtumia. Osallistuja voi näin ollen olla joko väline tai kohde. Lähteen on oltava läsnä prosessin alussa, mutta sen ei tarvitse osallistua prosessiin loppuun, kun taas tuloksen on oltava läsnä prosessin lopussa ilman osallistumista koko prosessiin. (Mt. s. 506)

### **Taulukko 3.1. Matriisirepresentaatio neljästä temaattisesta perusroolista**

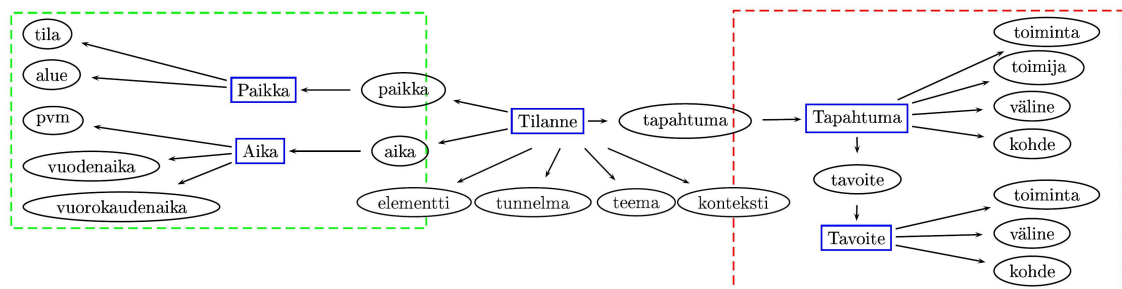
	<i>Lähde</i>	<i>Tulos</i>
<i>Ohjaaja</i>	Toimija	Tavoite
<i>Osallistuja</i>	Väline	Kohde

Näiden neljän perusroolin pohjalta Sowa on tarkentanut temaattisia rooleja tilanteen luonteeseen perustuen. Taulukossa 3.1 on avattu Sowan temaattiset roolit tilanteen luonteen mukaan. Esimerkiksi tilanteen luonteen ollessa siirtoa, on resurssi väline tai media ja tavoite kokija tai saaja. Toiminnasta puhuttaessa resurssit ovat välineitä ja tavoitteena on tulos tai vastaanottaja. Koska KulttuuriSammon sisällönkuvailun mallin luonteeksi valittiin toiminta, käytetään taulukon ensimmäistä riviä, eli temaattisia rooleja toimija, väline, tulos ja kohde. (Sowa 2000, s. 507-508)

**Taulukko 3.2. Tarkennetut temaattiset roolit tilanteen luonteen mukaan**

	Lähde		Tulos	
	Toimija	Resurssi	Tavoite	Kohde
<i>Toiminta</i>	Toimija, Vaikuttaja	Väline	Tulos, Vastaanottaja	Kohde, Teema
<i>Prosessi</i>	Toimija, Alkuperä	Aines	Tulos, Vastaanottaja	Kohde, Teema
<i>Siirto</i>	Toimija, Alkuperä	Väline, Media	Kokija, Saaja	Teema
<i>Tilallinen</i>	Alkuperä	Polku	Määränpää	Sijainti
<i>Ajallinen</i>	Alku	Kesto	Valmistuminen	Ajankohta
<i>Ympäröivä</i>	Alkuperä	Väline, Media	Tulos	Teema

Näitä toiminnan temaattisia rooleja Salminen hyödynsi tilanneskeemansa tapahtumaosiossa, joka on esitetty kuvan 3.1 oikeassa laidassa. Salminen erotteli tilanneskeemassaan tilanteeseen liittyvät puitteet tilanteen tapahtumista. Tämä ratkaisu on hänen mukaansa yleinen sisällönkuvailun tutkimuksessa sekä sisällönkuvailuontologioissa (Salminen 2006, s. 28). Tapahtuma ilmaisee tilanteessa tapahtuvan toiminnan, ja puitteet sisältävät ajan ja paikan ilmaisut, jotka rajaavat tilanteen aika-paikka-avaruudessa. Näiden lisäksi Salminen näki tarpeelliseksi sisällyttää tilanneskeemaansa tilannetta kuvaavia ominaisuuksia, kuten tunnelmaa ja teemaa. Näiden pohjalta Salmisen malli muotoutui kuvan 3.1 mukaiseksi.

**Kuva 3.1. Salmisen (2006) sisällönkuvailun malli**

Salminen ja Junnila rakensivat tilanneskeeman varaan KulttuuriSammon ensimmäisen version testatakseen skeeman käyttökelpoisuutta. Tätä versiota arvioitiin hyvin kevyesti vertailemalla toiminnallisuutta kuviteltuihin käyttötapauksiin. Tulokset kuitenkin osoittivat skeeman olevan käyttökelpoinen. Ongelmiksi osoittautuivat skeeman tarkkuustason ja helppokäyttöisyyden välillä tasapainoilu ja tulosten riippuvuus annotoijan tavasta käyttää skeemaa. (Salminen 2006, s. 51-56)

## 4 Semanttinen web

Suurin osa nykyisen WWW:n sisällöstä on ihmisten luettavaksi tarkoitettua aineistoa. Vaikka sisältö usein alun perin sijaitsee järjestettynä jossakin tietokannassa, se esitetään lopulta ihmislukijalle sopivasti muotoiltuna eikä muodossa jota koneet pystyisivät helposti hyödyntämään (Berners-Lee et al. 2001). Ihmiselle tarkoitettuja esitysmuotoja ovat esimerkiksi html, flash ja pdf.

Internetin käyttö painottuu kuitenkin yhä enemmän palveluihin ja tiedonhakuun. Tämänhetkiset WWW-sivut ja työkalut eivät ole kovinkaan tehokkaita näissä tehtävissä. Linkkien käyttö on yleensä liian vähäistä tiedon etsimiseen linkkejä seuraamalla. Jäljelle jäävät erilaiset sisältöä analysoivat hakukoneet kuten Google. Hakukoneiden ongelma on kuitenkin se, että ne eivät ymmärrä semanttisessa mielessä mitä ne etsivät, vaan palauttavat ainoastaan linkkejä sivuille, joissa jokin tietty kirjainyhdistelmä on joskus esiintynyt. (Antoniou & van Harmelen 2004, s. 1-2)

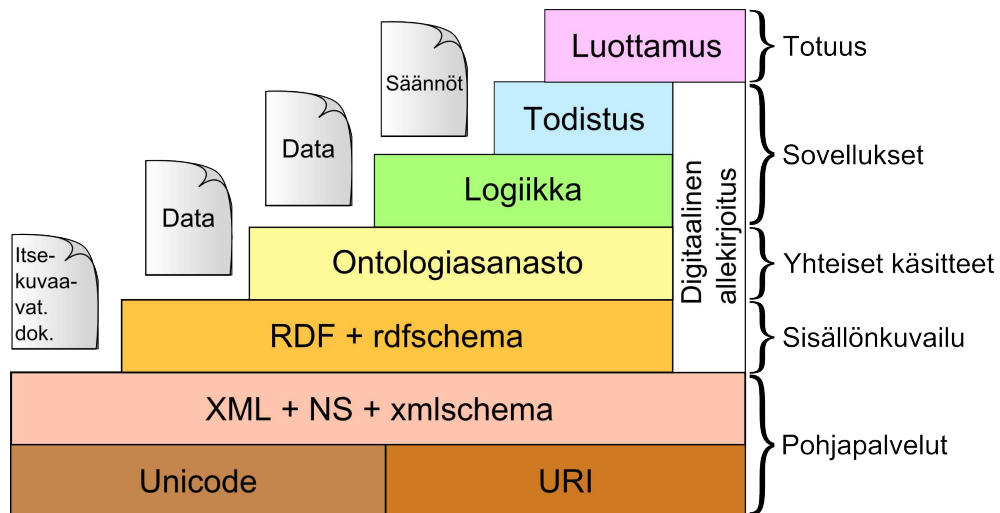
Semanttisen webin ydin on tiedon kuvaaminen tavalla, joka mahdollistaa tiedon koneellisen prosessoinnin ja hyödyntämisen. Semanttisen webin käyttötapaukset ovat tietämyksen hallinta, tietojen yhdistäminen ja tiedonhaku. W3C on määritellyt joukon tekniikoita, joilla WWW:ssä olevaa sisältöä voidaan esittää muodossa, jota koneet pystyvät analysoimaan. Nämä muodostavat semanttisen webin pohjan. Semanttinen web ei ole nykyisen WWW:n korvaaja, vaan laajennus, joka mahdollistaa koneiden ja ihmisten paremman yhteistoiminnan (Berners-Lee et al. 2001).

WWW:n kehittäjä Tim Berners-Lee on semanttisen webin keskeinen ideoija. Semanttinen web toteuttaisi hänen alkuperäisen visionsa WWW:stä suurena tietovarastona, jonne jokainen voi tuottaa merkityksellistä aineistoa. Kuten Berners-Lee toteaa: "Jos HTML ja web saivat online-dokumentit näyttämään yhdeltä valtavalta kirjalta, RDF(S) ja päättelykielet saavat maailman datan näyttämään yhdeltä valtavalta tietokannalta" (Berners-Lee & Fischetti 1999, s. 186). Tämän muutoksen myötä voi toteutua esimerkiksi älyllisten agenttien visio, jossa agentit hoitavat omatoimisesti asioita puolestamme (Mt.).

Yhtenä semanttisen webin kantavana ideana on alkuperäisessä WWW:ssä hyväksi todettu hajautettu rakenne. Hajautettu rakenne avaa mahdollisuuksia, joita voi olla vaikeaa tai jopa mahdotonta etukäteen kuvitella (Berners-Lee et al. 2001). WWW:n viimeaikaisten trendien eli niin sanotun Web 2.0:n kanssa semanttinen web jakaa ajatuksen avoimien rajapintojen ja hyödyllisen datan tarjoamisesta sillä perustelulla, että joku jossain voi hyvin keksiä niille hyödyllistä käyttöä. O'Reilly (2005) kutsuu tätä innovatiiviseksi yhteenliittämiseksi (O'Reilly 2005).

Semanttisen webin visio toteutuu usean erillään kehitetyn teknologian yhteenliittoksena. Berners-Lee (2000) on esittänyt semanttista webiä kuvaavan niin sanotun teknologiaakakun

(layer cake), jossa hän on kerännyt tasoina päällekkäin semanttisen webin toteutumiseen tarvittavat osakokonaisuudet. Malli on esitetty kuvassa 4.1. Tämän mallin tarkoituksena on myös havainnollistaa semanttisen webin kehitys kerroksittain ja tehdä selväksi käytettyjen tekniikoiden suhteet toisiinsa.



**Kuva 4.1. Semanttisen webin tasot eli ns. teknologiakakku (Berners-Lee 2000)**

Seuraavissa alaluvuissa käyn lyhyesti läpi eri tasojen sisältöä ja perehdyn niiden merkitykseen semanttisessa webissä.

#### 4.1 Pohjapalvelut: Internet- ja XML-tasot

Berners-Leen teknologiakakun alimmalla tasolla on niin sanottu Internet-taso, joka tarjoaa pohjapalveluita, kuten HTTP-protokollan ja URI-käytännön (Uniform Resource Identifier). Näiden päällä muut semanttisen webin teknologiat toimivat. Alimmalla tasolla on myös unicode-merkistö, jonka avulla merkistöongelmien pitäisi olla historiaa.

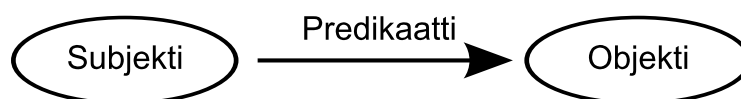
Internet-tason päällä on XML-taso. XML on merkkäuskieli strukturoitujen dokumenttien kirjoittamiseen käyttötapausta varten määriteltyjen sanastojen avulla. XML-dokumentteja on helppo käsitellä valmiiden rajapintojen ja työkalujen avulla, joita löytyy kaikille yleisimmille ohjelmointikielille.

#### 4.2 Sisällönkuvailu: RDF-taso

Teknologiakakun seuraava taso eli RDF-taso sisältää resurssien kuvaamiseen tarvittavan RDF (Resource Description Framework) -tiedonesityskielen sekä RDF-dokumenteissa käytettävien rakenteiden määrittelemiseen tarvittavan RDF Schema -kielen.

RDF on pohjimmiltaan aihealueriippumaton tietomalli (Antoniou & van Harmelen 2004, s. 62). Sen perusrakenneosa on subjekti-predikaatti-objekti kolmikosta (triple) rakentuva lause

(statement), jonka avulla resursseja (resource) voidaan kuvata. Kolmikko kuvaa subjektina olevaan resurssiin liittyvän ominaisuuden ja tämän ominaisuuden arvon.



#### **Kuva 4.2. RDF:n lause-perusrakenteen muodostavat subjekti, predikaatti ja objekti**

RDF-spesifikaatio määrittelee resurssin olevan mitä tahansa, jonka voi identifioida URI-viittauksella (URI reference) (Manola & Miller 2004). Näin ollen RDF:llä on mahdollista kuvata lähes mitä tahansa. Resurssi voi olla yhtä hyvin web-dokumentti, abstrakti käsite tai joku tietty henkilö. Resurssien ei tarvitse olla saavutettavissa webissä.

Resurssin voi ajatella olio-ohjelmointiparadigmasta tuttuna olion käsitteenä. Jokainen olio on uniikki, ja sillä on oltava uniikki tunniste. RDF:ssä tämä tunniste on URI. Jokaisella resurssilla on siis oma yksilöivä URI:nsa. URI:n tarkoitus on vähentää homonymiongelmia, jossa kaksi sanaa kirjoitetaan tai äännetään samalla tavalla, mutta niillä ei ole muuten mitään tekemistä toistensa kanssa. Viittaamalla resurssiin URI:n avulla voidaan varmuudella puhua samasta asiasta. URI-viittaus muodostuu alkuosasta ja vapaaehtoisesta #-merkillä erotetusta fragmentista (fragment identifier). Esimerkiksi URI-viittaus <http://www.esimerkki.org/index.html#luku2> koostuu kahdesta osasta, alkuosasta <http://www.esimerkki.org/index.html> ja fragmentista luku2.

Kaikki RDF-kolmikoiden osat voivat olla resursseja. Subjektien ja predikaattien on itse asiassa oltava resursseja. Objekti voi resurssin lisäksi olla myös literaali, kuten henkilön nimi. Literaalit löytyvät suunnattujen graafien muodostaman semanttisen verkon lehdistä. Joukosta kolmikoita muodostuu RDF-graafi, joka on jotain tiettyä aihealuetta kuvaava tietämuskanta. Tämä tietämuskanta voidaan nähdä suunnattuna verkkona, joka levittäytyy uri-viittausten kytkemänä internetin yli. (Manola & Miller 2004)

Abstrakti tietomalli tarvitsee kuitenkin konkreettisen konekäsiteltävän syntaksin, jotta se olisi käyttökelpoinen. Ihminen osaa tulkita graafeja, mutta koneelle tämä on vaikeampaa (Antoniou & van Harmelen 2004). RDF-tietomallin yleisin sarjallistus on XML-muodossa (Manola & Miller 2004). RDF:lle luotu XML-syntaksi mahdollistaa XML-tekniikoihin, kuten parsereihin ja prosessoreihin, perustuvan tehokkaan käsittelyn ja yhtenäisen DTD:hen sidotun esitystavan.

XML ei ole RDF:n ainoa esitystapa. RDF on tietomalli, eikä näin ollen sidottu mihinkään tiettyyn esitystapaan. RDF:n XML-syntaksi on ihmiselle vaikeaselkoinen, ja tiedon esittämiseen käytetään huomattavan paljon enemmän dataa, kuin optimoitu tapaus olisi tarvetta. RDF:lle on siksi kehitetty myös kompaktimpia ja helppolukuisempia notaatioita,

kuten N3<sup>1</sup> ja Turtle<sup>2</sup>. XML on kuitenkin vinyt voiton, koska sen käsittely on monelle entuudestaan tuttua, ja työkaluja XML-dokumenttien käsittelyyn on tarjolla suuria määriä. Tieteellisessä kirjoittamisessa kompaktimmat notaatiot ovat käytettympiä.

### 4.3 Yhteiset käsitteet: ontologia-taso

RDF:n tarjoaman perusrakenteen päällä on ontologiataso. Ontologia on eksplisiittinen ja formaali diskurssialueen kuvaus (Studer et al. 1998, s. 185). Se on tyypillisesti äärellinen lista termejä ja suhteita termien välillä, jossa termit kuvaavat tärkeitä diskurssialueen käsitteitä (Antoniou & van Harmelen 2004, s. 10). Sanan ontologia yleisin merkitys on filosofian ja tarkemmin metafysiikan osa-alue, eli oppi olevaisesta<sup>3</sup>, joka käsittelee todellisuuden yleistä luonnetta ja tutkii, millainen ympäröivä todellisuus on. Ontologian perimmäinen kysymys on "mitä on olemassa?".

Tietojenkäsittelytieteissä ja ohjelmistotuotannossa on jo pitkään mallinnettu reaali maailmaa tavoitteena luoda parempia tietojärjestelmiä. Tällaista mallinnusta ja abstrahointia tehdään niin olio-ohjelmoinnissa kuin tietokantasuunnittelussakin. Tarpeeksi pitkälle vietyinä mallinnuksissa voidaan nähdä piirteitä ontologiasta. Tästä syystä sana onkin kaapattu tietojenkäsittelytieteen tarkoituksiin. Tietojenkäsittelytieteen kontekstissa ontologialla tarkoitetaan formaalia ja eksplisiittistä kuvausta ongelma-alueen jaetusta käsitteellistyksestä. Formaali tarkoittaa, että ontologian tulisi olla koneymmärrettävä. Eksplisiittinen tarkoittaa, että käytetyt käsitteet ja niiden käytön rajoitukset on eksplisiittisesti määritelty. Jaettu tarkoittaa, että ontologian tietämys on jonkin laajemman ryhmän hyväksymää, ei yksittäisen henkilön yksityistä tietoa. (Studer et al. 1998, s. 185)

Perimmiltään ontologia on luettelo asioista, joiden oletetaan olevan olemassa alueessa D henkilön näkökulmasta, joka käyttää kieltä L puhuakseen D:stä (Sowa 2000, s. 492). Ontologioiden käyttö on lisääntynyt viimeisen kymmenen vuoden aikana. Ontologioita käytetään nykyään esimerkiksi tietämyksen hallintaan ja älykkäissä järjestelmäintegraatioissa (Gómez-Pérez et al. 2004, s. 8).

Ontologiat ovat keskenään hyvin erilaisia. Tämä johtuu siitä, että ontologia on laaja-alainen käsite ja ontologioilla on monia käyttötarkoituksia ja kohteita. Koska ontologiat muodostuvat aina käyttötarkoituksen mukaan, on tämän työn kannalta tärkeä ymmärtää ontologioiden luokittelu ja erilaisten ontologiatyyppien soveltuvuus eri tilanteisiin ja käyttötarkoituksiin.

Ontologioita on hyödyllistä luokitella kahden eri dimension suhteen: sisäisen struktuurin rikkauden suhteen ja käsitteellistämisen kohteen suhteen. Sisäisen struktuurin rikkaus

<sup>1</sup><http://www.w3.org/DesignIssues/Notation3.html>

<sup>2</sup><http://www.dajobe.org/2004/01/turtle/>

<sup>3</sup>Kreikan sanoista "oleminen" ja "oppi".

on mittari sille, kuinka formaali ja ilmaisuvoimainen ontologia on. Asiasanastot, joita jotkut saattavat pitää ontologioina, ovat tämän skaalan heikoimmassa päässä. Ne eivät ole formaaleja, eikä niissä yleensä ole mahdollista määritellä asioiden suhteita toisiinsa. Skaalan toisessa päässä ovat formaalit kehysontologiat, jotka ilmaisevat ominaisuuksien arvorajoitteita ja yleisiä loogisia rajoitteita. (Gómez-Pérez et al. 2004, s. 26-36)

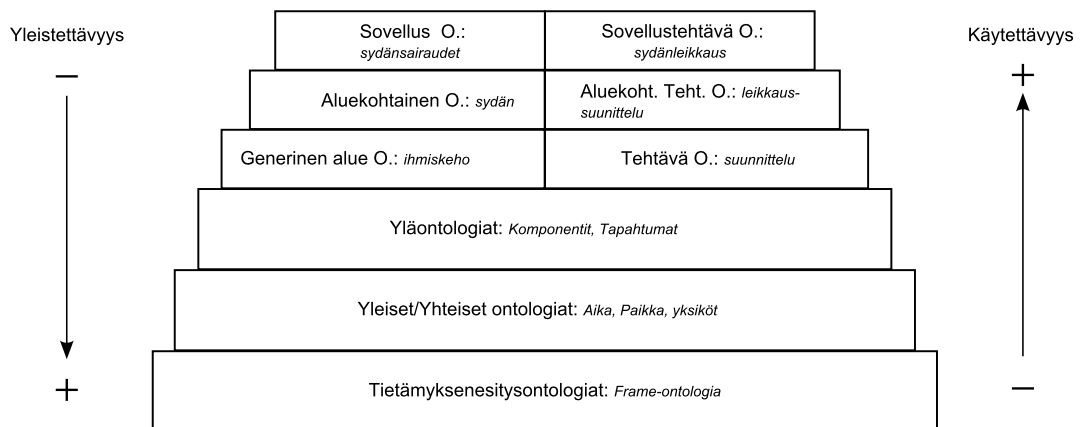
Jako sisäisen struktuurin rikkauden mukaan on selkeä, ja ontologiat on helppo luokitella tämän dimension mukaan. Käsitteellistämisen kohteen suhteen luokittelu on monimutkaisempaa, koska selkeiden rajojen vetäminen on vaikeaa. Ontologiat saattavat kuulua useaan luokkaan samaan aikaan. Gómez-Pérez et al. (2004, s. 29-35) ovat tehneet yhteenvedon erilaisista luokittelutavoista, ja koostaneet niiden perusteella seuraavan listan:

- *Tietämyksen esittämisen* (Knowledge Representation, KR) ontologiat tarjoavat primitiivit, joiden avulla tietämystä voi käsitteellistää tietyn KR-paradigman sisällä. Esimerkki tällaisesta on Gruberin (1993) kehysontologia (Frame Ontology), joka tarjoaa formaalit määritelmät kehyspohjaisissa järjestelmissä käytettävistä primitiiveistä, kuten luokka, aliluokka, attribuutti, arvo, relaatio ja aksioma. Kehysontologia mahdollistaa muiden ontologioiden rakentamisen kehyspohjaisten käytäntöjen avulla.
- *Yleiset ja yhteiset* ontologiat (General ontologies) edustavat maalaisjärkitietämystä, joka on uudelleenkäytettävissä eri käyttötarkoituksiin. Tällaisia ovat esimerkiksi sanastot, tapahtumat, aika, paikka, kausaalisuus, käyttäytyminen, toiminta ja mereologia.
- *Yläontologiat* (Upper-level ontologies) kuvaavat hyvin yleisiä käsitteitä, joihin kaikkien aluekohtaisten ontologioiden juuret tulisi liittää yleisen yhteentoimivuuden saavuttamiseksi. Tällä hetkellä yläontologioiksi on useita ehdotuksia. IEEE:n Standard Upper Ontology (SUO) työryhmä on kehittämässä standardoitua yläontologiaa. Tämän yläontologian nimi on Standard Upper Merged Ontology (SUMO), koska se on yhdistelmä vartenotettavimmista yläontologioista. (Niles & Pease 2001)
- *Aluekohtaiset ontologiat* (Domain ontologies) ovat uudelleenkäytettäviä aihealueen sisällä. Esimerkiksi museoalalle on FinnONTO-projektissa kehitetty oma museoalan ontologia (MAO), joka sisältää museoalan keskeiset käsitteet ja niiden suhteet. Yläontologioiden ja aluekohtaisten ontologioiden ero on selkeä. Aluekohtaiset ontologiat ovat yleensä yläontologioiden käsitteiden ja suhteiden tarkennuksia.
- *Tehtäväontologiat* (Task ontologies) kuvaavat sanastoa, jota käytetään jonkin geneerisen tehtävän yhteydessä. Tehtäviä voivat olla esimerkiksi myyminen, diagnosointi tai ajoittaminen. Tehtäväontologiat voivat ulottua usean eri aihealueen yli, jossa tehdään samantyyppisiä tehtäviä.

- *Aluekohtaiset tehtäväontologiat* (Domain-task ontologies) ovat edellisen ryhmän uudelleenkäytettävyydeltään rajoittuneempi kategoria ontologioita, joita voidaan uusiokäyttää saman aihealueen sisällä, mutta jotka eivät sovellu aihealueitten väliseen käyttöön.
- *Metodiontologiat* (Method ontologies) kuvaavat miten tehtäviä suoritetaan. Esimerkiksi Nokian RosettaNetin automatisoitua tarjouspyyntöprosessia voi kuvata metodiontologiaksi, koska se kuvaa miten ja missä järjestyksessä erilaisia tehtäviä tulee suorittaa.
- *Sovellusontologiat* (Application ontologies) tarjoavat määrittelyt tietyn sovelluksen tarvitseman tietämyksen mallintamiseen.

Tässä työssä kehitetään tietämyksen esittämisen ontologioita heterogeenisten aineistojen kuvailuun ja hyödynnetään suomalaisen semanttisen webin infrastruktuurin yleisiä sekä aluekohtaisia ontologioita käyttämällä FinnONTO-projektissa kehitettyjä YSO-, MAO- ja paikkaontologioita.

Kuva 4.3 havainnollistaa ontologioiden luokittelua käsitteellistämisen kohteen suhteen. Se kuvaa luokkien suhteita toisiinsa sekä luokkien yleistettävyyttä ja käytettävyyttä. Mitä tarkemmin ontologia on sidottu johonkin aihealueeseen, sitä parempi on sen käytettävyys. Yleistettävyyys puolestaan laskee ontologian tarkkuuden kasvaessa.



**Kuva 4.3. Ontologioiden luokittelu käsitteellistämisen kohteen suhteen ja luokkien merkitys yleistettävyyteen ja käytettävyyteen (Gómez-Pérez et al. 2004)**

Ontologioiden evaluointi on nousemassa tärkeäksi ontologioiden käytön yleistyessä. On oltava jokin tapa päätellä, mikä ontologia on hyvä ja mikä on huono, ja mitä kannattaisi tietystä yhteydestä käyttää. Ontologioiden sisällön evaluoinnin kannalta tärkeimmät



menetelmät ovat Methontology (Fernandez et al. 2006) ja OntoClean<sup>4</sup> (Gómez-Pérez et al. 2004). Nämä tarjoavat menetelmiä, joilla ontologian sisällöllisten rakenteiden oikeellisuuden voi varmistaa. Toisaalta monet ovat sitä mieltä, että ainoa kestävä tapa arvioida ontologioita on käyttää niitä sovelluksissa ja arvioida sovellusten suorituskykyä (Sure et al. 2004).

Web Ontology Language (OWL) on W3C:n suosittama kieli webissä käytettävien ontologioiden määrittämistä varten. Se perustuu OWL:ia edeltäneeseen DAML+OIL-kieleen. OWL on RDF-S:ää monipuolisempi kieli ja sisältää kehittyneempiä rakenteita ontologisten merkitysten ilmaisemiseksi. OWL-kielellä on mahdollista ilmaista monipuolisia luokkien välisiä suhteita, kuten kardinaliteettejä ja yhtäsuuruutta. Se tukee myös vahvempaa tyypitystä ja ominaisuuksien ominaisuuksia. Kielestä on määritelty kolme eri versiota, joiden ominaisuudet vaihtelevat ilmaisuvoiman ja laskennallisen vaativuuden suhteen. (McGuinness & van Harmelen 2004)

Laajin ontologiakieli on OWL Full. Siinä on mahdollista käyttää kaikkia OWL:n, RDF Scheman ja RDF:n rakenteita. OWL Fullia käyttäen voidaan kuitenkin laatia ontologioita, jotka eivät ole laskennallisesti ratkeavia koneellisen päättelyn kannalta. Tästä syystä täydestä versiosta on kehitetty karsittu versio OWL DL, jossa laskennallinen ratkeavuus on taattu. Kevein versioista on OWL Lite, joka on tarkoitettu esimerkiksi yksinkertaisten sanastojen määrittelyyn. Sitä voidaan käyttää sovelluksissa, joissa halutaan tarjota OWL-yhteensopivuus käyttämättä kuitenkaan monimutkaisempia OWL:n semanttisia rakenteita. (McGuinness & van Harmelen 2004)

#### **4.4 Sovellukset: logiikka ja päättely**

RDF-muotoista dataa ja ontologioita voidaan hyödyntää logiikan avulla ja päättely mahdollistaa uuden tietämyksen muodostamisen (Berners-Lee & Fischetti 1999, s. 185-186). Logiikkatasolla tulevat mukaan myös sovelluskohtaiset säännöt, jotka voivat olla käyttökelpoisia esimerkiksi verkkokaupan alalla (Antonioni & van Harmelen 2004, s. 12-18). Logiikka voi tarjota päättelyn sivutuotteena todistuksia ja selityksiä esitetyille asioille. Päättelyketjua voidaan seurata taaksepäin, ja päättelyaskelia seuraamalla vakuuttua esimerkiksi siitä, että lomamatkaa varaava älykäs agentti on tehnyt järkeviä päätöksiä (Berners-Lee & Fischetti 1999, s. 191-192).

#### **4.5 Totuuden jäljillä: luottamus**

Berners-Leen kerrosmallin (kuva 4.1) ylimmällä tasolla on luottamus. Semanttisessa webissä tarjottavaan tietoon ja palveluihin sekä agenttien tekemiin päätöksiin on pystyttävä luottamaan. Berners-Lee & Fischetti (1999 s.193) korostavat, että vasta luottamuksen myötä annamme koneiden tehdä asioita puolestamme. Luottamusluokituksia ja digitaalisia

---

<sup>4</sup><http://www.ontoclean.org/>

allekirjoituksia tullaan todennäköisesti tarjoamaan samalla tavalla kuin nykyisinkin luotettujen organisaatioiden kautta (Antoniou & van Harmelen 2004, s. 18). Kiinnostava filosofinen kysymys on, kenen totuus semanttisessa webissä on se oikea ja kenellä on valta valita se yksi totuus usean totuuden joukosta.

#### **4.6 Yhteenveto semanttisesta webistä**

Tässä luvussa kerroin semanttisen webin taustoista ja tekniikoista. Semanttisen webin ideana on tehdä WWW:n sisällöistä koneellisesti prosessoitavia älykkäämpien palveluiden mahdollistamiseksi. RDF-tietomalli ja OWL-muotoiset ontologiat ovat semanttisen webin tärkeimpiä rakenneosia. RDF:llä kuvataan resursseja ja OWL:n avulla kerrotaan minkälaisia asioita on olemassa. Ontologioita käsitellessäni kerroin myös, kuinka ontologioita voidaan kategorisoida käyttötarkoituksen mukaan ja mitä ontologiakategorioita tässä työssä hyödynnetään.

## 5 Semanttinen moninäkymähaku

Tiedonhaku on yksi WWW:n tärkeimpiä ja samalla haastavimpia sovelluksia. Tehokkaat ja helppokäyttöiset tiedonhakujärjestelmät ovat maailman menestyvimpien yritysten ydinalaa. Semanttinen web teoreettisena mallina tarjoaa kiinnostavia mahdollisuuksia tiedonhaun avuksi. Haaste on saada nämä mahdollisuudet hyödynnettyä ratkaisujen pysyessä helppokäyttöisinä.

Semanttisen webin yhtenä avainajatuksena on erottaa informaatio sen ihmiselle muotoillusta esitystavasta. Tätä informaatiota on kuitenkin vaikea käyttää ilman sitä tehokkaasti hyödyntäviä sovelluksia. Tässä työssä kehitän semanttisen webin sisällölle tiedonhakujärjestelmän, joka pyrkii piilottamaan alustan monimutkaisuuden ja tuomaan esiin sen vahvuudet helppokäyttöisessä muodossa. Uutta käyttöliittymää rakentaessa on tärkeää tiedostaa minkälaisia ratkaisuja aikaisemmin on tehty. Siksi tarkastelen tässä luvussa olemassa olevia tiedonhaun ja selaamisen menetelmiä, jotka jollain tavoin hyödyntävät semanttisen webin tekniikoita ja sisältöä.

Mäkelä (2006) on pro gradu -työssään tutkinut erilaisia hakutapoja ja niiden soveltuvuutta semanttisen tiedonhaun toteutustavaksi. Tutkimuksessa hän identifioi viisi erilaista lähestymistapaa semanttiseen tiedonhaakuun.

### 1. Tavanomaisen avainsanahaun vahvistaminen semanttisin tekniikoin

Tässä lähestymistavassa haettava aineisto ei ole semanttisessa muodossa, vaan lähestymistapaa voidaan käyttää esimerkiksi tiedonhaakuun WWW-sivuilta. Hakuavainten perusteella etsitään ontologisia vastaavuuksia ja pyritään löydetyn semanttisen rakenteen perusteella laajentamaan tai rajaamaan tekstihakua.

### 2. Yksinkertainen käsitteen paikantaminen

Tässä lähestymistavassa etsitään käsitettä esimerkiksi valitsemalla haluttu luokka luokkahierarkiasta ja rajaamalla sen attribuuttien arvot esimerkiksi vapaatekstihaulla tai päivämäärärajoituksella. Löydetyistä käsitteistä voidaan yleensä jatkaa samoilua muihin käsitteisiin viittaavien attribuuttien polkuja pitkin. Lähestymistavan kyselyt ovat muotoa "etsi kaikki julkaisut, joissa kirjoittajana on x ja projektina y".

### 3. Kompleksinen rajoitushaku

Monet monimutkaiset kyselyt voidaan ilmaista hakemalla tietyn tyyppisiä olioita, joita yhdistävät tietyt suhteet. Tällaisia rakenteita voidaan ilmaista graafihahmoina, joissa solmujen ja attribuuttien tyypit on kiinnitetty. Graafihahmoa voidaan käyttää kyselynä tietämuskantaan. Graafihahmot ovat siitä ongelmallisia, että niiden luominen on käyttäjille vaikeaa.

#### 4. Ongelmanratkaisu

Ongelmanratkaisu, jossa kuvaillaan ongelma ja päätellään ratkaisu ontologisen tietämyksen pojalta, on yksi yleisimmistä semanttiseen webiin yhdistetyistä käyttötapauksista. Ongelmanratkaisun toteutuksia on kuitenkin harvassa ja olemassa olevat ovat hyvin rajoittuneita.

#### 5. Yhdistävän polun löytäminen

Tämä lähestymistapa keskittyy löytämään kiinnostavia yhteyksiä resurssien välillä kiinnostavien resurssien sijaan.

Näiden lähestymistapojen erilaisista toteutuksista Mäkelä tunnisti joitakin yhteisiä menetelmäoppeja, jotka olivat RDF-polun kulkeminen, avainsanojen ja avainkäsitteiden peilaaminen toisiinsa, graafihahmojen käyttö, logiikan ja päättelyn hyödyntäminen sekä sumea logiikka. Näiden menetelmäoppien pohjalta Mäkelä päätyi nimeämään monimutkaiset graafihahmot hyödyllisyydessään ja laajennettavuudessaan semanttisen haun ydintekniikaksi.

Graafihahmojen luonti graafisella käyttöliittymällä on kuitenkin vaikeaa. Mäkelä näki ratkaisuna semanttisen moninäkömähaun (Hyvönen et al. 2004e), joka muodostaa graafihahmoja käyttäjän näkymävalintojen perusteella. Hyvönen et al. (2004e) esitteli semanttisen moninäkömähaun alun perin yliopiston promootiojuhlien valokuville toteutetussa hakujärjestelmässä.

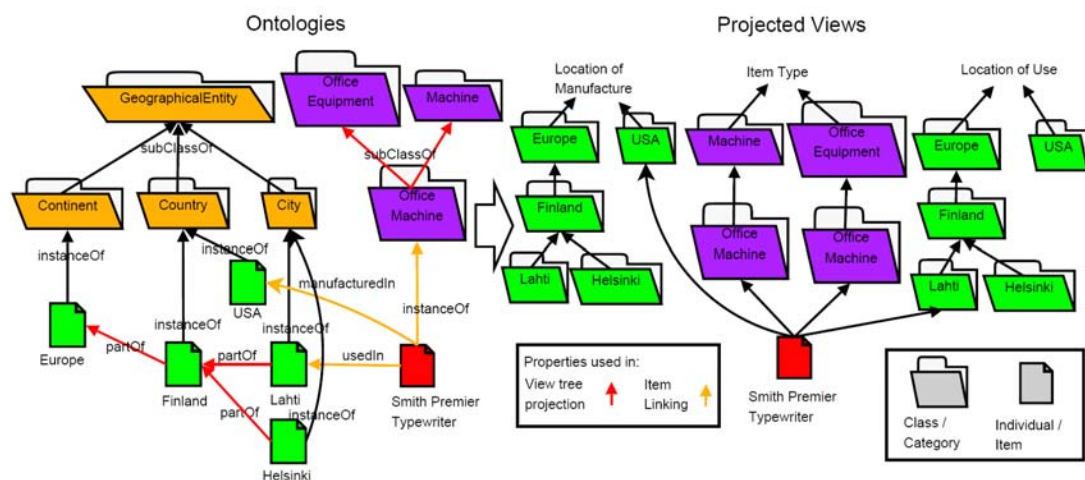
Moninäkömäinen luokittelutapa perustuu ajatukseen, että dokumenteilla ja olioilla on aina useita dimensioita. Taksonomia, jossa dokumentilla tai oliolla on tasan yksi paikka kiinteässä hierarkiassa, on näin ollen mahdottomuus. Moninäkömäajatuksen kehitti intialainen matemaatikko ja kirjastonhoitaja S.R. Ranganathan 1930-luvulla. (Rosenfield & Morville, 2002, s. 204-208)

Moninäkömähaussa luokittelun dimensioita kutsutaan näkymiksi tai faseteiksi, koska ne tarjoavat erilaisia tarkastelunäkökulmia luokittelun kohteena olevaan aineistokokoelmaan tai ikään kuin lohkaisevat tietyn pinnan näkyviin kokoelmasta. Itse näkymien esitystavat hakujärjestelmissä vaihtelevat. Joitain näkymiä on hyvä esittää hierarkioissa, toisia taas yksitasoisina "litteinä" listoina. Myös muita esitystapoja kannattaa harkita. Kartat ovat luonnollisin tapa maantieteellisen tiedon esittämiseen ja janat ajallisen tiedon esitystavaksi. (Rosenfield & Morville, 2002, s. 204-208)

Tiedonhaku moninäkömähaussa tapahtuu valitsemalla näkymistä kiinnostavia kategorioita. Jokainen valinta rajaa tulosjoukkoa niin, että tulosjoukko sisältää ne dokumentit, jotka kuuluvat valittuihin kategorioihin tai valittujen kategorioiden alikategorioihin (Hyvönen et al. 2004e).

Moninäkömähaku on tullut osaksi WWW:tä tiedonhakuovelluksissa, joista ensimmäinen kehitettiin suuren lääketieteellisten tekstien kokoelman hakuun ja oli nimeltään HiBrowse (Pollit 1998). HiBrowseen jälkeen moninäkömähäun ajatusta on hyödynnetty useissa järjestelmissä, kuten Flamenco (Yee et al. 2003) ja Relation Browser++ (Zhang & Marchionini 2005).

Semanttinen moninäkömähaku nojaa näkömähäun projisoimiseen ontologisista rakenteista. Prosessi on kaksiosainen: puunäkömähäun projisointi suunnatusta graafista ja dokumenttien ripustaminen projektion kategorioihin. Kuvan 5.1 esimerkki havainnollistaa prosessia. Hierarkkisen kategoriapuun projisointi tehdään seuraamalla graafia tiettyjen sääntöjen perusteella, poimimalla talteen vastaan tulevia oleellisia käsitteitä ja liittämällä ne kategoriapuuhun alla olevan tietämyskannan suhteiden mukaisesti. Useimmiten suhteet ovat hyponymioita, eli suhteita joissa käsitteen merkitys kokonaan sisältyy toisen käsitteen merkitykseen, tai meronymioita eli käsitteitä jotka merkitsevät tarkasteltavan käsitteen osaa. (Mäkelä 2006)



**Kuva 5.1. Esimerkki semanttisen moninäkömähäun näkömähäun projektioista (Mäkelä 2006)**

Toisessa vaiheessa dokumentit yhdistetään projisoituihin kategorioihin. Tämä tehdään yksinkertaisesti määrittelemällä ominaisuus (predikaatti), jolla dokumentin on viitattava kategoriana käytettyyn resurssiin. Tämä tehdään myös sääntöjen perusteella. Esimerkiksi kuvassa 5.1 Smith Premier Typewriter -kirjoituskone liittyy USA:han valmistuspaikkanäkömähäun, koska sen manufacturedIn-kaari viittaa USA-resurssiin, toimistokoneeseen esinetyypinäkymähäun koska sen instanceOf kaari viittaa Office Machine resurssiin sekä Lahteen käyttöpaikkanäkömähäun, koska usedIn property osoittaa Lahti-resurssiin.

Hakeminen semanttisessa moninäkömahaussa tapahtuu samalla tavalla kuin alkuperäisessäkin moninäkömahaussa: rajaamalla tulosjoukkoa kategoriavalintoja yhdistelemällä. Kuvassa 5.2 on esimerkki MuseoSuomen<sup>1</sup> käyttöliittymästä, jossa on valittu valmistuspaikkanäkömähästä kategoria Nokia ja materiaalinäkömähästä tekstiilimateriaalit. Hakutulos sisältää kenkiä ja vaatteita, mikä vastaa odotusarvoa erityisesti kumisaappaiden osalta.

The screenshot shows the MuseoSuomi website interface. At the top, there is a logo for the Helsinki Institute for Information Technology and the University of Helsinki. The main header features the MuseoSuomi logo and the tagline 'Suomen museot semanttisessa webissä'. Below the header, there is a search bar and navigation links. The search results are displayed in a grid format, showing various items such as shoes and clothing. Each item is accompanied by a small image and a text description including the category and a unique ID.

**Käsitteet:**    tarkenna hakua

**Hakuehdot**

**Kategoria:** Valmistuspaikka > Eurooppa > Suomi > Länsi-Suomen lääni > Pirkanmaa > Nokia (ryhmittele kahteet) (9/ista)

**Kategoria:** Materiaali > materiaalit > tekstiilimateriaalit (ryhmittele kahteet) (9/ista)

**Kohteet - ei ryhmittelyä**

**Kohteet 1-9/9**

			
Jalkineet, naisen kumisaappaat (LKM LHM LHM ES 97070 261)	Yöpuika, naisen trikooyöpuika (LKM LHM LHM ES 97070 296)	Yöpuika, naisen trikooyöpuika (LKM LHM LHM ES 97070 354)	Jalkineet, miehen lenkkikengät (LKM LHM LHM ES 2000020 94)
			
Jalkineet, lapsen kumisaappaat (LKM LHM LHM ES 87034 4)	Jalkineet, miehen samettikengät (LKM LHM LHM ES 93047 6)	Jalkineet, naisten kestäkengät, kangaskengät (LKM LHM LHM ES 97070 367)	Jalkineet, naisen kumijalkineet (LKM LHM LHM ES 85015 175)
	Jalkineet, naisen tenniskengät (LKM LHM LHM ES 97070 268)		

**Esinetyyppi** (koko luokittelu) (ryhmittele kahteet)  
jalkineet ja tekstiilit (9)

**Materiaali**: [kangas](#) > [materiaalit](#) > [tekstiilimateriaalit](#) (koko luokittelu) (ryhmittele kahteet)  
[kangas](#) (9), [nahka](#) (3), [pöytä](#) (1)

**Valmistaja** (koko luokittelu) (ryhmittele kahteet)  
[virvikset](#) (9)

**Valmistuspaikka**: [kangas](#) > [Eurooppa](#) > [Suomi](#) > [Länsi-Suomen lääni](#) > [Pirkanmaa](#) > [Nokia](#) (koko luokittelu) (ryhmittele kahteet)  
Ei valittavia kategorioita

**Valmistusajka** (koko luokittelu) (ryhmittele kahteet)  
[alkukaudet](#) (9), [vuosisadat](#) (9)

**Käyttäjät** (koko luokittelu) (ryhmittele kahteet)  
[henkilöt](#) (9)

**Käyttöpaikka** (koko luokittelu) (ryhmittele kahteet)  
[Eurooppa](#) (9)

**Käyttötilanne** (koko luokittelu) (ryhmittele kahteet)  
[valmistustekniikat](#) (2)

**Kokoelma** (koko luokittelu) (ryhmittele kahteet)  
[Lahden kaupungin museon kokoelmat](#) (9)

## Kuva 5.2. MuseoSuomi - semanttinen moninäkömähakukäyttöliittymä

KulttuuriSampo-projektissa ryhdyimme jatkokehittämään MuseoSuomi-portaalia tuomalla mukaan museoesineiden lisäksi muita aineistotyyppisiä. Toteutuksen pohjana oli näin ollen MuseoSuomi ja siinä käytetty semanttinen moninäkömähaku. Semanttinen moninäkömähaku haluttiin tutkimusryhmässä säilyttää myös KulttuuriSammon pääasiallisena hakukäyttöliittymänä. Luvussa 7.2.1 kerron miten toteutin semanttisen moninäkömähäun KulttuuriSampoon ja mitä ongelmia heterogeeniset aineistot aiheuttivat.

<sup>1</sup><http://www.museosuomi.fi/>

## 6 KulttuuriSampo-portaali

FinnONTO-projektilla on neljä tuloksellista tavoitetta: kehittää keskeisiä kansallisia ontologioita, perustaa palveluita ontologioiden kehittämiseen ja käyttöön, kehittää työkaluja semanttisen webin parissa työskentelyyn ja pilotoida kehitettyjä ontologioita, palveluja ja työkaluja portaalihankkeissa. Tässä diplomityössä kuvaan KulttuuriSampo-portaalin toteutuksen ja arvioin toteuttamaani portaalia käyttäjätestein. Seuraavaksi kuvaan KulttuuriSammon vision ja lähtökohdat hankkeeseen.

Kulttuurisesta näkökulmasta KulttuuriSampo-portaalin tavoitteena on olla kansallista kulttuuriamme dokumentoiva WWW-palvelu, joka tarjoaa erilaisia lähestymistapoja portaalin sisältöön. Visiona on tuoda yhteen heterogeenistä kulttuurisesti merkittävää ja kiinnostavaa aineistoa kansallisista muistiorganisaatioista ja tuottaa aineistosta yhtenäinen portaali, jossa aineistot punoutuvat yhteen semanttisten yhtymäkohtien kautta ja ovat haettavissa ja selailtavissa yhdellä käyttöliittymällä. Heterogeenisellä aineistolla tarkoitetaan eri metadataskeemojen mukaan kuvattua museokokoelmadataa, joka kuvaa kaikenlaisia kulttuuriartefakteja kuten valokuvia, videoita, maalauksia, esineitä, kansanperinnettä, kulttuurista perimätietoa, biografioita ja kulttuuripaikkoja. KulttuuriSammossa näitä artefakteja edustaisi mahdollisuuksien mukaan kuva, teksti, video tai esikatselukuva hyperlinkin päässä olevasta WWW-sivusta.

KulttuuriSampo-portaalilla on useita ajateltuja käyttötapauksia ja käyttäjäryhmiä, jotka ovat toimineet pohjana suunnittelutyössä. Portaalin yhtenä tavoitteena on palvella museoalan työntekijöitä, jotka voivat käyttää portaalia mukana olevien museoiden kokoelmien tarkasteluun ja yksittäisten kiinnostavien esineitten hakemiseen. KulttuuriSammon tulisi tarjota tehokas ja monikäyttöinen hakumenetelmä, jonka avulla käyttäjät saisivat tietämuskannasta louhittua tarvitsemansa tiedon. Toinen tärkeä käyttäjäryhmä on opiskelijat ja koululaiset. Heille portaalin tulisi tarjota elämyksiä, oppia ja apua esimerkiksi harjoitustöiden teossa. Opettajat voisivat hakea portaalin avulla tietoa jostain aikakaudesta ja siihen liittyvistä esineistä ja taideteoksista. Myös toimittajat ja taittajat voisivat käyttää palvelua kuvamateriaalin hakemiseen jostain tietyistä aiheesta. Antiikin keräilijät voisivat etsiä portaalin avulla tietynlaista esinettä ja saada tietoa samankaltaisista esineistä ja niiden historiasta. Viihdekäyttäjät saattaisivat innostua seuraamaan jotain tiettyä tarinaa ja siihen linkittyviä esineitä ja oppia asioita Suomen historiasta ja suomalaisista vaikuttajista. Myös semanttisen webin tutkijat ovat sovelluksen käyttäjäryhmä, onhan projektin tarkoitus tuottaa arvokkaita tutkimustuloksia ja esittelevän tekniikan sovelluskohteita. Jotta KulttuuriSammosta ei tulisi pelkkä tekniikkademonstraatio, on portaalin todellisten käyttäjien tavoitteet huomioitava alusta alkaen ja portaalia suunniteltava käyttäjien tarpeet mielessä.

KulttuuriSampo varten tunnistimme tutkimusryhmässä seuraavat käyttäjäryhmät:

- *Museoalan ammattilaiset* voivat käyttää palvelua sen tarjoamien kattavien hakuominaisuuksien ja portaaliluonteen takia. Heillä on yleensä paljon aikaa etsiä haluamaansa ja hakukriteerit ovat muita tarkempia.
- *Koululaiset, opiskelijat ja opettajat* voivat hyödyntää KulttuuriSampo opetus- ja oppimiskäytössä. Portaali on hyödyksi valmisteltaessa esitelmiä tai esseitä. Portaalin käyttö on vaihtoehto yleisten hakukoneiden käytölle tai tiedon etsimiselle kirjastosta. Koululainen tai opiskelija saa KulttuuriSammosta konkreettista sisältöä, kuten kuvia ja tekstiä. Kuvaamataidossa suomalaisten taiteilijoiden historia ja teokset voivat tarjota opettajille apumateriaalia.
- *Toimittajat ja taittajat* etsivät taustatietoja juttuihinsa useista lähteistä. KulttuuriSampo voi hyvin palvella toimittajia yhtenä lähteenä. Toimittajat ja taittajat saattavat tarvita historiallista kuvamateriaalia ja KulttuuriSampo voi olla hyvä väline kuvamateriaalin etsimiseen. Löydettyään sopivan kuvan, he voivat ottaa yhteyttä museoon ja pyytää kuvan käyttöoikeutta juttuunsa.
- *Viihdekäyttäjät* on kulttuurista ja historiasta jonkin verran kiinnostunut henkilö, joka etsii palvelusta elämyksiä. Hänellä ei välttämättä ole tarkkoja ajatuksia siitä, mitä hän etsii. Hakeminen on lähinnä selailua tai samoilua. Viihdekäyttäjän tavoitteena on viettää aikaa miellyttävällä tavalla ja mahdollisesti oppia ja kultivoitua siinä ohessa. Hän päätyy todennäköisesti portaaliin sattumalta eikä jaksa paneutua järjestelmän ominaisuuksiin yhtä paljon kuin ammattikäyttäjät. Mikäli viihdekäyttäjät kokee käyttöliittymän liian vaikeaksi, hän luovuttaa, eikä tule uudestaan takaisin.
- *Sisällöntuottaja* on henkilö, joka toimittaa sisältöä portaaliin. Hän työskentelee kulttuuriperinnön säilyttämistä tavoittelevassa organisaatiossa ja haluaa edistää halussansa olevan materiaalin hyötykäyttöä ja leviämistä. Sisällöntuottajan mielestä sisällöt on saatava esille, ja digitaalinen julkaiseminen on hänen mielestään erittäin hyvä tapa tehdä niin. Sisällöntuottaja on motivoitunut tehtävässään ja tekee mielellään työtä, joka edistää materiaalin leviämistä, kuten annotoi sitä tarvittavalla tavalla.

KulttuuriSampo-tutkimusprojektiä varten saatiin kerättyä suuri määrä erilaista aineistoa keskeisiltä suomalaisilta kulttuuritoimijoilta ja muistiorganisaatioilta. Mukana olivat Valokuvataiteen museo, Suomen maatalousmuseo, Suomalaisen Kirjallisuuden Seura, Kansalliskirjasto, Espoon kaupunginmuseo, seitsemän kaupungin ja kolmentoista museon Antikvaria-ryhmä, Valtion taidemuseo, johon kuuluvat Ateneum ja Kiasma, Museovirasto, YLE ja Geologian tutkimuskeskus (GTK).

Edellä mainitut organisaatiot ovat antaneet KulttuuriSampo-projektin käyttöön kokoelmiensa sähköisessä muodossa olevat tietokannat. Kokoelmatyypit vaihtelevat esinekokoelmista grafiikka- ja piirustuskokoelmiin. Aineistotyypeistä puhuttaessa



aineistossa on esineitä, taideteoksia, kuten maalauksia, grafiikka ja piirustuksia, valokuvia (sekä taidevalokuvia että dokumentoivia valokuvia), videoaineistoa, ääntä, tarinoita, tekstejä ja biografioita sekä WWW-sisältöjä.

Suoraan organisaatioilta saatavien tietokantakokoelmien lisäksi KulttuuriSampo varten on tarkoitus tuottaa jonkin verran uutta aineistoa. Yhtenä tavoitteena on kuvata tarinat ja taidot semiformaalisti, jotta ne sitoisivat yhteen esineitä, taideteoksia ja audiovisuaalista materiaalia. Yksi suomalaisen taiteeseen vahvasti vaikuttanut tarina on Kalevala. Kalevala valittiinkin koko projektin keskeiseksi teemaksi, ja sen tapahtumia, tarinaa ja tulkintoja pyritään projektin yhteydessä formaalisti kuvaamaan. Tarinan vaiheiden yhteydessä KulttuuriSammon tavoite on tarjota linkkejä tarinaan liittyviin artefakteihin. Kalevalan kuvailuun ryhdyttiin kesällä 2006, mutta se ei valmistunut tätä työtä varten kehitettyyn versioon järjestelmästä.

KulttuuriSampo-projektissa on keskitytty tutkimaan eri aihealueiden välistä yhteentoimivuutta. Portaaliin tuotava aineisto on useasta aihealueesta museoalan sisällä, ja sen kuvailuun on käytetty erilaisia metadataskeemoja. Perinteisesti vastaavissa hankkeissa on ollut tapana peilata ja sovittaa kaikki aineistot yhteiseen tietokantarakenteeseen tai metadataskeemaan, kuten Dublin Coreen. Tätä lähestymistapaa kutsutaan syntaktiseksi yhdistämiseksi, koska siinä kaikki aineistot muunnetaan samaan muotorakenteeseen. (Hyvönen et al. 2006)

MuseoSuomi-hankkeessa sisällöt pyrittiin yhdistämään semanttisella tasolla. Semanttisen tason sisällön yhdistämisessä myös metadatan arvot ovat yhteisesti jaetut. Ne on järjestetty itsenäisiin ontologisiin rakenteisiin, joihin metadatat viitataan. Metadatan arvot, kuten kirjailijat, paikkakunnat ja aikakaudet viittaavat yhteisiin ontologioihin. Mikäli samat käsitteet esiintyvät useammassa ontologiassa, voidaan ontologiat peilata yhteneväsiltä osilta toisiinsa. Tällä tasolla tehtävä aineiston yhdistäminen mahdollistaa päättelyn aineiston sisällön semanttisista suhteista. Päättelyn avulla semanttinen haku, semanttinen samoilu, suosittelu, todistukset ja muut "älykkäät" palvelut tulevat mahdollisiksi. (Hyvönen et al. 2006)

Syntaktisessa yhdistämisessä ja MuseoSuomen kaltaisessa semanttisen tason sisällön yhdistämisessä on kuitenkin se ongelma, että mitä suurempi diversiteetti lähdeaineistossa on, sen vähemmän yhteisellä metadataskeemalla voi kuvata. Biografian kuvaamiseen tarvittavat käsitteet eroavat valokuvan kuvaamiseen tarvittavista käsitteistä. Kaikille artefakteille yhteiset käsitteet ovat harvassa ja korkealla abstraktiotasolla, eivätkä näin ollen enää ole kovinkaan kiinnostavia tietystä artefaktista kiinnostuneelle henkilölle. Koska KulttuuriSamossa on muun muassa valokuvia, videoita, maalauksia ja esineitä, ei MuseoSuomessa käytetty lähestymistapa ole toimiva. Tutkimusryhmässä ongelmaa lähdettiin ratkaisemaan siirtymällä sisällönkuvailussa toiminnan ja temaattisten roolien

tasolle (Junnila et al. 2006; Salminen 2006; Junnilla 2006). Tällä tasolla monimutkaisiakin asioita on mahdollista ilmaista keskenään yhteentoimivalla tavalla. Tähän tietomalliin perehdyin jo jonkin verran luvussa 3.2. Seuraavassa luvussa kertoessani tarkemmin KulttuuriSammon teknisestä toteutuksesta, kerron yksityiskohtaisemmin tietomallista, sekä siitä miten mallia muokattiin KulttuuriSampo varten ja millä tavalla aineistot muokattiin tietomallin mukaisiksi.

## 7 KulttuuriSammon toteutus

Edellisessä luvussa kuvasin KulttuuriSammon vision ja tavoitteet ja kerroin minkälaisia aineistoja sain museoilta portaalidemonstraatiota varten. Tätä työtä varten valmistunut versio KulttuuriSammosta käsittää vain osan portaalin kokonaisvisiosta. Tämä versio käsittää semanttisen moninäkömahaun ja aineiston peilaamisen erilaisille pinnoille. Vaikka nämä toiminnallisuudet ovat vain osa portaalin kokonaisvisiosta, ovat ne jo itsessään hyvin kiinnostavia. Tässä luvussa kerron teknisestä näkökulmasta toteuttamastani hakujärjestelmästä. Käyn läpi lopullisen tietomallin, kerron käyttöliittymän kehityksestä sekä ratkaisuksista ja avaam sovellusarkkitehtuuria, jonka tarvitsin portaalin toteuttamiseksi.

### 7.1 Tietomalli ja aineiston käsittely

KulttuuriSampoa varten luvussa 3 esittelemäni mallit eivät olleet suoraan hyödynnettävissä. Metadataskeemat olivat liian staattisia ja muut mallit enemmän teoreettisia kuin käytännöllisiä. KulttuuriSampoa varten tunnistettiin tarve ilmaisuvoimaiselle RDF-kielellä toteutetulle tietämyksenesityskielelle, joka sopisi erityisesti muistiorganisaatioiden heterogeenisten aineistojen kuvailuun, loogiseen päättelyyn ja semanttiseen julkaisemiseen.

Junnila (2006) loi prosessikuvauksiin kohdistuvalla työllään ensimmäisen version toimintaperustaisesta sisällönkuvailusta kehittämällä prosesseja kuvaavan tilanneontologian. Prosessikuvaukset koostuivat toisiaan seuraavista tapahtumista, joita määritteli tietty tekeminen ja siihen liittyvät toimijat. Mallissa oli myös otettu huomioon ympäristö, aika, tavoitteet ja tunnelma. Junnila identifioi toiminnan kuvailun tilanneontologian tärkeimmäksi osa-alueeksi. Hän arvioi mallia annotoimalla tarkasti pienen määrän aineistoa tilanneontologian mukaisesti ja rakentamalla prototyypinhakujärjestelmän. Tulokset olivat lupaavia.

Yhteistyössä Junnilan kanssa Salminen (2006) kehitti kuvien ja videoiden sisällönkuvailua. Salmisen malli perustuu pitkälti Sowan (2000) ja Saeedin (1997) ajatuksiin temaattisista rooleista. Hän päätyi aiempia sisällönkuvailuskeemoja tutkimalla ja tarpeisiin peilaamalla valitsemaan toimintaa määritteleviksi tekijöiksi toimijan, välineen, kohteen ja tavoitteen. Toiminnaksi hän määritteli kaikki luonnollisessa kielessä verbeillä ilmaistava aktiivinen tai passiivinen toiminta, kuten juokseminen ja leipominen tai istuminen ja makaaminen. (Salminen 2006)

Doerr (2003) huomauttaa ontologioiden rakenteen aina olevan kompromissi ilmaisuvoiman, ymmärrettävyyden ja helppokäyttöisyyden välillä. Junnilan ja Salmisen mallin ongelmaksi osoittautui juuri sen monimutkainen rakenne ja liian korkea tarkkuustaso. On kyseenalaista, voiko niin tarkkaa sisällönkuvailua edellyttää laajoilta aineistoilta ja onko se esimerkiksi

taloudellisesti mahdollista. On myös syytä kysyä, saavutetaanko tarkkuudella tarpeeksi lisähyötyä annotoinnin kasvaneeseen vaivaan nähden.

Sisällönkuvailun mallissa oli myös rakenteellisia ongelmia, jotka tulivat esiin pyrittäessä automatisoimaan annotointia ja rakentamaan mallin päälle hyödyllisiä palveluita. Mallin sidos logiikkaan ja RDF-malliin oli vielä liian hataralla pohjalla. Se oli toiminut KulttuuriSammon ensimmäisessä versiossa, jossa aineistona oli vain joitakin kohteita. KulttuuriSammon toisen version laajaan aineistoon ja ohjelmalliseen hyödyntämiseen malli ei ollut riittävä.

Ruotsalo ja Hyvönen (2007) identifioivat olemassa olevista metadataskeemoista ja tietämyksen esitystavoista seuraavat tarpeet sisällön kuvailun laatua parantaviin selvennyksiin (explication). Näiden selvennysten noudattaminen tekee tietämyksenesityskielestä ilmaisuvoimaisemman ja yksiselitteisemmän. Etuliitteellä dc viitataan Dublin Coren nimiavaruuteen<sup>1</sup> ja etuliitteellä ms MuseoSuomen nimiavaruuteen<sup>2</sup>.

1. *Suhteellisten roolien selventäminen tapahtumina.* Monien metadataskeemojen tarjoamat ominaisuudet, kuten dc:creator tai ms:placeOfManufacturing eivät ole luonteeltaan binäärisiä ominaisuuksia, vaan niihin kietoutuu useampia resursseja. Niiden merkitys määräytyy lisäksi käyttökontekstista. ms:placeOfManufacturing esimerkiksi viittaa paikkaan ja tiettyyn valmistamisen tapahtumaan, joka voi olla kontekstista riippuen maalaamista, rakentamista tai teollista valmistamista.
2. *Binääristen roolien selventäminen käsitteinä.* Metadataskeemojen binäärisillä ominaisuuksilla on niitä vastaavia käsitteitä ontologioiden hierarkioissa, jotka määrittelevät käsitteiden ylä- ja alakäsitteet. Nämä roolit tulisi selventää peilaamalla ne ontologioiden oikeisiin kohtiin.
3. *Puuttuvien temaattisten roolien selventäminen* tarkoittaa, että annotaatioissa on käsitteiden instansseja, joiden suhde annotoitavaan objektiin ei ole tiedossa. Esimerkiksi dc:subject sisältää usein joukon aihealueontologian käsitteitä, kuten yso:hevonen, yso:mies ja yso:ratsastaa. Ihminen pystyy päättelämään, että tällä tarkoitetaan miehen ratsastavan hevosella, eikä toisin päin. Automaattisessa päättelyssä Merkitys menee kuitenkin sekaisin, mikäli instanssien rooleja tapahtumassa ei selvennetä.

Tätä työtä varten kehitin yhdessä tutkija Tuukka Ruotsalon kanssa Junnilan ja Salmisen töihin pohjautuen KulttuuriSampo varten uuden yhtenäisen tietämyksen esitysmuodon, joka pohjautuu tapahtumille ja laajojen aihealuekohtaisten ontologioiden sekä yläontologioiden käyttöön. Menetelmä siirtää aineiston kuvailun tasolle, joka on lähempänä kielellistä

<sup>1</sup>dc: <http://purl.org/dc/elements/1.1/>

<sup>2</sup>ms: <http://www.cs.helsinki.fi/group/seco/ns/2004/03/18-esinekortti>

ilmaisua ja tapaa, jolla jäsentellemme ympäröivää maailmaamme. Menetelmässä otetaan huomioon edellä esitetyt kolme selvennystarvetta. Jokainen yksittäinen tapahtuma on instanssi toimintaa kuvaavasta ontologiakäsitteestä, mikä selventää tapahtuman luonteen. Tapahtumaa kuvaillaan sitomalla tapahtumaan muita instansseja temaattisilla rooleilla. Tämä ratkaisee sekä binääristen roolien että puuttuvien temaattisten roolien ongelmat.

Valitsimme tietämyksen esitysmuodollemme seitsemän temaattista roolia, jotka on listattu taulukossa 7.1. Nämä temaattiset roolit järjestimme hierarkkisesti osaksi YSO-ontologiaa niin, että kaikilla temaattisilla rooleilla on yhteinen yläluokka. Tämä helpottaa temaattisten roolien hyödyntämistä tapauksissa, joissa tarkalla roolilla ei ole merkitystä. Etuliite yso viittaa Yleisen suomalaisen ontologian (YSO) nimiavaruuteen<sup>3</sup>.

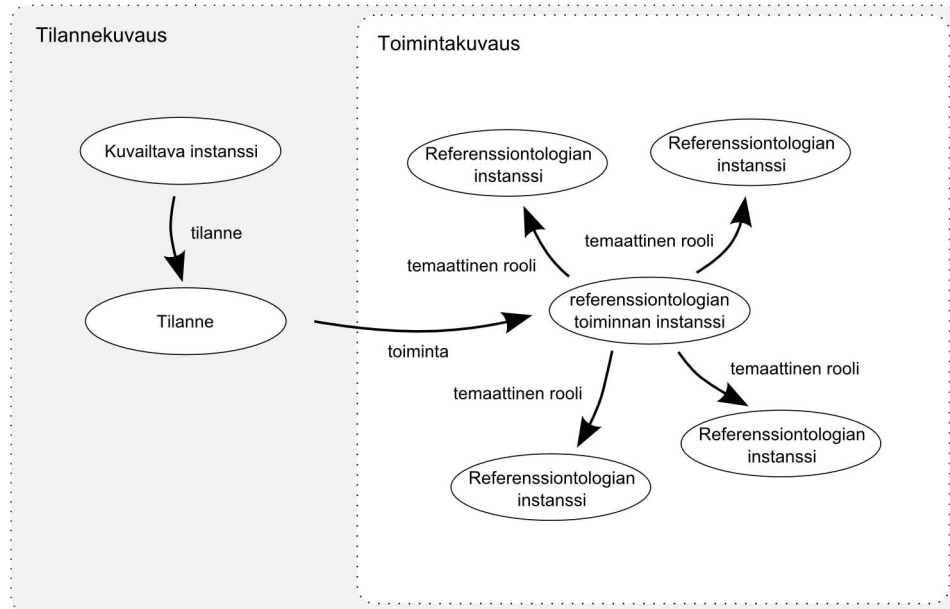
**Taulukko 7.1. KulttuuriSammossa käytettävät temaattiset roolit**

<i>Temaattinen rooli</i>	<i>Arvo</i>	<i>Merkitys</i>
yso:agent	eloisan käsitteen instanssi $x$	$x$ on toiminnan alullepanija
yso:patient	käsitteen instanssi $x$	$x$ muuttuu toiminnan seurauksena
yso:instrument	käsitteen instanssi $x$	$x$ :ää käytetään välineenä toiminnassa
yso:goal	käsitteen instanssi $x$	$x$ on kuvatus toiminnan tavoite
yso:participant	eloisan käsitteen instanssi $x$	$x$ on toiminnan osallinen
yso:place	tilallisen käsitteen instanssi $x$	$x$ on toiminnan tapahtumapaikka
yso:time	ajallisen käsitteen instanssi $x$	toiminta tapahtuu ajassa $x$

Nämä temaattiset roolit sitovat yhteen referenssiontologioiden käsitteitä tai niiden instansseja. Referenssiontologioina voidaan käyttää mitä tahansa ontologioita. KulttuuriSammossa käytettäviä referenssiontologioita ovat YSO, MAO, paikkaontologia ja aikaontologia. Kuvassa 7.1 on kaavakuva tietomallista. Kuvasta nähdään kuinka tietomalli koostuu tilannekuvauksesta ja siihen liittyvistä toimintakuvauksista. Kuvailtava instanssi, kuten taulu tai runo, viittaa tilanteisiin, joita taulusta tai runosta on pystytty eriyttämään. Taulussa tilanteita on todennäköisesti vain yksi, mutta runossa tilanteet saattavat seurata toisiaan. Tilanneinstanssi puolestaan viittaa toimintakuvauksiin. Toimintakuvauksen keskellä on jonkin toimintaa sisältävän ontologian sopivan toiminnan instanssi. Tämä instanssi voi olla instanssi esimerkiksi yso:kalastus- tai yso:nauraminen-käsitteistä. Tähän

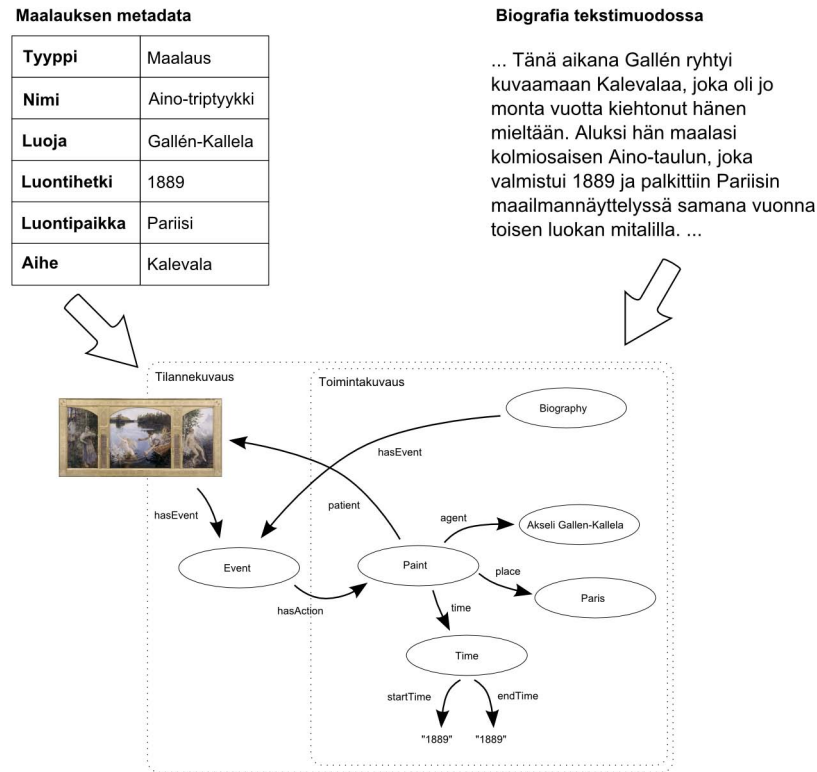
<sup>3</sup>yso: <http://yso.fi/YSO>

toimintaa kuvaavaan instanssiin sidotaan osallisia temaattisin roolein. yso:goal voi kalastuksen yhteydessä olla yso:kala ja yso:instrument puolestaan viitata mao:onkivapa-käsitteen instanssiin.



**Kuva 7.1. Tietomalli käsitteellisellä tasolla**

Tietämyskannan muodostava rikas verkko syntyy siitä, kun samoja instansseja tai saman luokan instansseja käytetään muissa toimintakuvauksissa samoissa tai muissa temaattisissa rooleissa. Esimerkki tällaisesta tapauksesta on kuvattu kuvassa 7.2. Kuvassa maalauksen metadatatista muodostettu tilannekuvaus ja taidemaalarin tekstimuotoisesta biografiasta muodostettu tilannekuvaus saavat saman muodon. Näin ollen kohdistettaessa tietämyskantaan haku, jossa sidotaan paikaksi Pariisi ja agentiksi Akseli Gallen-Kallela, tulee se palauttamaan viitteet sekä maalaukseen että biografiaan.



## Kuva 7.2. Esimerkki maalauksen metatietojen ja biografian yhdistämisestä

Rajaamalla temaattisten roolien määrää ja järjestämällä ne hierarkkisesti ontologiseen rakenteeseen saadaan malli, jota on hyvin helppo hyödyntää ohjelmallisesti. Kaikki tietämuskannan louhimiseen tarvittavat algoritmit voidaan rakentaa temaattisten roolien varaan. Erityistä tilannekohtaista käsittelyä ei enää tarvita, vaan monesti riittää geneerinen toteutus, jossa käytettyyn temaattiseen roolin ei oteta kantaa. Koska malli on rakenteeltaan luvussa 3.2 esittelemäni sisällönkuvailun skeemaa yksinkertaisempi, uskon että Salmisen (2006, s. 51-56) kokemat ongelmat annotoijien vaihtelevista tavoista käyttää kuvailuskeemaa eivät ole yhtä merkittävä ongelma tätä mallia käytettäessä.

Käytettyjen referenssiontologioiden laadulla on suora vaikutus tietämuskannan laatuun, koska ontologisten rakenteiden perusteella voidaan tehdä lisäpäätelyä. Esimerkiksi paikkaontologian tulisi kertoa missä maassa ja missä maanosassa Helsinki sijaitsee. Näiden tietojen perusteella Helsinkiä voidaan verrata muihin kaupunkeihin. Huonolaatuinen paikkaontologia saattaa sisältää vain kaupunkien nimiä ilman maantieteellistä tietoa, eikä vertailu olisi mahdollista. Ontologisen tiedon hyödyntäminen vaatii kuitenkin käytetyn ontologian tuntemusta ja ontologiakohtaisia algoritmeja.

Jotta aineistot saadaan edellä kuvaamani tietomallin mukaiseen muotoon, on ne yhteismitallistettava semanttisen rikastamisen prosessissa. Semanttisen rikastamisen prosessi alkaa käsittelemällä museoiden tietokannoista saatua museokohtaisesti muotoiltua

XML-muotoista dataa, joka sisältää tietokantakenttien nimet ja arvot. Prosessin ensimmäisessä vaiheessa nämä XML-tiedostot luetaan ja muunnetaan KulttuuriSampo varten määritellyn metadataskeeman mukaiseen RDF-muotoon projektia varten kehitetyllä konversiotyökalulla. Joissain tapauksissa museot ovat saaneet tuotettua kokoelmätietokannastaan yhteistä Suomen museot online (SMOL) XML-formaattia, joka helpottaa aineiston käsittelyä vähentämällä tarvittavien konversioiden määrää. Muissa tapauksissa konversio on tehtävä tapauskohtaisesti.

Konversiotyökalun laatu on kriittisin tekijä laadukkaasti aineiston saamiseksi ja luo koko perustaan onnistuneelle portaalille. Konversiotyökalun kehitys veikin huomattavan osan KulttuuriSampoon laitetusta työpanoksesta. Konversiotyökalu luotiin tutkimusryhmässä yhteisvoimin, ja sen luontiin osallistuivat itseni lisäksi tutkija Tuukka Ruotsalo sekä tutkimusapulaiset Mikko Virkkilä ja Olli Alm.

Konversion yhteydessä toimijoiden (kuten tekijä, lahjoittaja ja valmistaja) nimet pyritään tunnistamaan ja yhdistämään merkkijonoja vertailemalla tietämuskannassa jo oleviin toimijoihin. Tavoitteena on, että yhtä olemassa ollutta henkilöä vastaisi ainoastaan yksi uniikki resurssi. Tämä parantaa tietämuskannan laatua ja hakujen tarkkuutta. Sama toimenpide tehtiin myös paikoille, kuten käyttöpaikka ja valmistuspaikka. Paikat pyrittiin ripustamaan paikkojen nimiä vertailemalla paikkaontologiaan, joka sisältää hierarkkisen jaottelun lisäksi paikkojen koordinaattitiedot. Koordinaattitietojen avulla kohteet voidaan myöhemmin esimerkiksi näyttää kartalla.

Toimijoiden ja paikkojen nimien tunnistuksen lisäksi samankaltainen toimenpide tehtiin myös luettelointitiedoista löytyvien asiasanojen kohdalla. Artefaktien metatiedoista löytyvät sanat muunnettiin perusmuotoon ja näillä sanoilla haettiin ontologisia käsitteitä. Löydetyt sanat ripustettiin ontologioista löydettyihin käsitteisiin. KulttuuriSammossa käytetään ontologisoituja asiasanastoja, kuten Museoalan asiasanastoon pohjautuvaa Museoalan ontologiaa (MAO) ja Yleisen suomalaisen asiasanaston pohjalta koottua Yleistä suomalaista ontologiaa (YSO).

Tietojen tunnistaminen ja ripustaminen ontologisiin rakenteisiin ei itsessään tarjoa mitään, mitä tavanomaisilla relaatiotietokannoilla ei pystyisi tekemään. Tieto olisi tällä tavoin järjestettyä kaikissa maailman museotietokannoissa jo nyt, mikäli ne olisi suunniteltu hyvin. Tosiasia kuitenkin on, että tietokannat harvemmin suunnitellaan oikeaoppisesti niin, että tietokannassa olisi esimerkiksi henkilö-taulu, josta löytyisi ainoastaan yksi henkilöä vastaava relatio, johon kaikki hänen teoksensa vierasavaimella (foreign key) viittaisivat.

On myös mahdollista, että olemassa olleet relaatiot ovat hävinneet viettäessä tietokantoja XML-välitysmuotoon. Tätä prosessia emme päässeet valvomaan tai vaikuttamaan siihen. Merkkijonoja jälkikäteen vertailemalla ei päästä samaan tarkkuuteen, koska saman toimijan nimen kirjoitusasu vaihtelee tapauksesta toiseen ja samannimisiä henkilöitä on



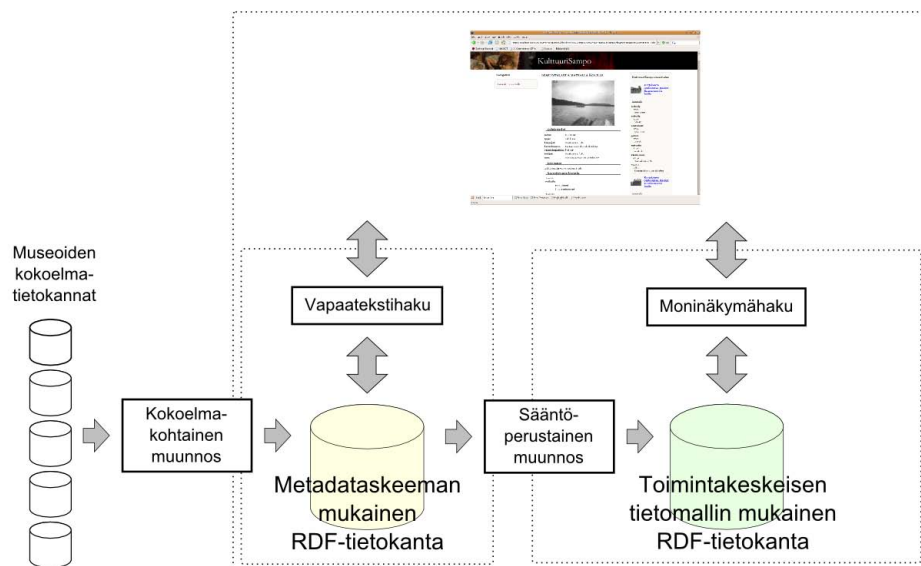
useita. Tällä menetelmällä esimerkiksi kaikki kaimat niputetaan saman nimen alle. On kuitenkin muistettava, että kyseessä on tutkimusprojekti ja tavoitteena on käyttökelpoisen testiaineiston luonti.

Edellä kuvattujen toimenpiteiden jälkeen käytössä on RDF-muodossa oleva ontologioihin ripustettu pohja-aineisto. Tämä pohja-aineisto on siis siinä semanttisen webin mukaisessa muodossa, jona Berners-Lee haluaisi kaiken WWW-sisällön olevan.

KulttuuriSammon julkaisuprosessissa aineistoa siis rikastetaan kolmella puoliautomaattisella toimenpiteellä, joista kaksi ensimmäistä kuvattiin edellä.

- Aineistosta löytyvien toimijoiden ja paikkojen yhdistäminen nimiä vertailemalla.
- Asiasanojen ripustaminen referenssiontologioihin, kuten YSO ja MAO.
- Tapahtumakehysten luonti toimintaan viittaavista asiasanoista.

Kuva 7.3 kuvaa rikastusprosessia ja lopputulosten päälle rakennettuja palveluja. Ensimmäinen tietokanta sisältää XML-konversiossa tuotettua aineistotyyppi-kohtaisten metadataskeemojen mukaista RDF-dataa. Tämän tietokannan päällä toimii perinteinen vapaatekstihaku, jossa etsitään asiasanoista ja artefaktien kuvailuista hakusanoihin täsmäviä sanoja. Toinen tietokanta sisältää ensimmäisestä tietokannasta sääntöpohjaisesti tuotettuja tapahtuma- ja toimintakuvauksia. Näiden kuvausten päällä toimii moninäkömähaku.



**Kuva 7.3. Semanttisen rikastamisen prosessi**

Edellä listatuista rikastustoimenpiteistä kolmas on ehdottomasti monimutkaisin ja kriittisin, mutta samalla eniten lisäarvoa tuottava. Siinä aineistotyyppi-kohtaisten metadataskeemojen mukaisia sisältöjä jatkokäsitellään sääntökielen avulla. Muunnosohjelman toteutti tutkija

Tuukka Ruotsalo. Käsittelyn tuloksena syntyy toimintakeskeisiä sisällönkuvailuja, eli tapahtuma- ja toimintakehyksiä. Muunnoksessa dokumenteista etsitään toimintaan viittaavia asiasanoja ja tiettyjen sääntöjen perusteella muodostetaan kehyksiä, joiden keskeinen toiminta on löydetyn toiminnan instanssi. Näiden kehysten avulla heterogeeniset sisällöt saadaan yhteismitallistettua, ja staattiset aineistot, kuten valokuvat, taulut ja esineet saadaan yhdistettyä moniulotteisempiin narratiiveihin ja prosesseihin, kuten elokuvaan, taitokuvauksiin ja runoihin.

## 7.2 KulttuuriSammon käyttöliittymä

KulttuuriSampo-projektissa halutaan laajentaa tapoja hakea ja selata kulttuuriaineistoa MuseoSuomeen verrattuna. Semanttisen moninäkömähäun rinnalle halutaan tuoda lisää aineiston haku- ja selailutapoja, jotka yhdessä antaisivat kokonaisvaltaisemman näkökulman aineistojen historiaan, käyttökohteisiin ja tapahtumiin. MuseoSuomen pohjalta tutkimusryhmässä ideoitiin useita uusia hakutapoja, kuten relaatiohaku, käsitehaku ja näyttelyhaku. MuseoSuomessa käytetty semanttinen moninäkömähaku haluttiin kuitenkin säilyttää yhtenä hakutapana, koska se oli osoittautunut toimivaksi MuseoSuomessa. Semanttista moninäkömähakua ei missään vaiheessa kuitenkaan oltu evaluoitu, joten perusteltujen muutosten tekeminen hakulogiikkaan tai käyttöliittymään oli vaikeaa. Koska MuseoSuomen ja KulttuuriSammon tietomallit erosivat olennaisesti toisistaan, heijastui se semanttisen moninäkömähäun toteutukseen. MuseoSuomen tietomalli oli museoesineiden toimialakohtainen tietomalli, jossa käytetyt luokat ja käsitteet olivat sidottuja museoesineisiin. KulttuuriSammossa toimitaan toimintakeskeisen yleisen tietomallin päällä, jossa sidokset ovat yleisemmällä tasolla. Tätä diplomityötä varten toteutin KulttuuriSampoon toimintakeskeisen semanttisen moninäkömähäun, jota kuvaan seuraavassa alaluvussa.

### 7.2.1 Toimintakeskeinen semanttinen moninäkömähaku

KulttuuriSampoon toteuttamani *toimintakeskeinen semanttinen moninäkömähaku* perustuu MuseoSuomessa kehitettyyn ajatukseen aineistokorpuksen ripustamisesta erilaisille ontologioista projisoituihin näkymiin. Koska MuseoSuomessa käsitellään ainoastaan museoesineitä, on hyödyllisten fasettien keksiminen MuseoSuomen moninäkömähakuun ollut melko yksinkertaista. Ne ovat pitkälti samat kuin käytetyn metadataskeeman museoesineille asettamat kentät: esinetyyppi, materiaali, valmistaja, valmistuspaikka, valmistusaika, käyttäjä, käyttöpaikka, käyttötilanne ja kokoelma.

Myös artefaktien ripustus näkymien kategorioihin on MuseoSuomessa suoraviivaista: näkömäpuut muodostetaan referenssiontologian suhteita seuraamalla ja tiettyyn näkömäpuun kategoriaan liittyvät ne esineet, joiden tietyn ominaisuuden arvo viittaa kategoriaa vastaavaan ontologiseen käsitteeseen. Näin esimerkiksi valitsemalla

YSO-ontologian tietystä haarasta muodostetusta käyttötilanne-näkymästä kategorian juustonvalmistus, palauttaa järjestelmä hakutuloksena juustomuotin, joka on asiasanoitettu YSO:sta löytyvään juustonvalmistuskäsitteeseen.

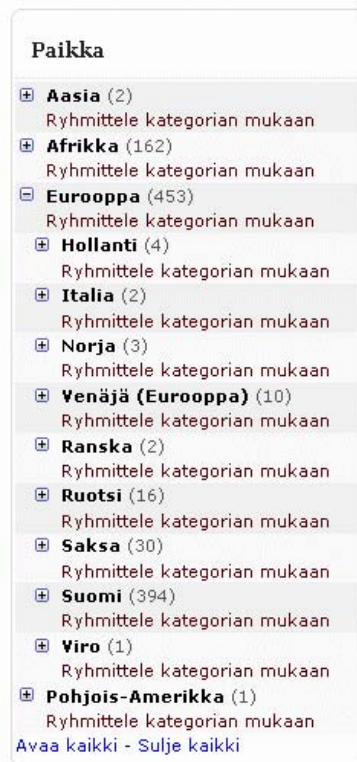
KulttuuriSammossa käytettävien näkymien valinta ei kuitenkaan ollut yhtä helppoa. Kysymys on pienimmästä yhteisestä nimittäjästä, eli mitkä artefaktien ominaisuudet löytyvät kaikilta portaalin artefakteilta? Vastaus on, että artefakteilla ei ole kovin montaa yhteistä ominaisuutta. Emme voi sanoa, että kaikilla artefakteilla olisi valmistuspaikka. Mikä on esimerkiksi WWW-sivun tai runon valmistuspaikka? Vaikka artefaktit voitaisiinkin pakottaa yhteiseen muottiin, määrittelemällä, että esimerkiksi WWW-sivun valmistuspaikka on sen palvelimen IP-osoitteen maantieteellinen sijainti, ei se olisi hakukoneen käyttäjälle intuitiivista. Tuloksena ei olisi kovin toimiva tiedonhakupöytä.

Toinen vaihtoehto on määritellä näkymät niin, että niissä on omat räätälöidyt näkymät jokaista aineistotyyppiä kohti. Tämä ei kuitenkaan palvele visiota aineistot saumattomasti yhdistävästä portaalista. Lisäksi näkymien määrä kasvaisi niin suureksi, että haun käytettävyys todennäköisesti kärsisi. Tutkimusryhmän MuseoSuomesta saatu palaute koskee useimmiten näkymien lukumäärästä sekavaksi muodostunutta käyttöliittymä.

Erilaiset artefaktit ovat kuitenkin löyhemmin kytköksissä toisiinsa. WWW-sivu saattaa kuvailla Espoota turistikohteena, kun taas Espoon kaupunginmuseon kokoelmista löytyviä nuken rattaita on työnnetty Tapiolan kaduilla. Tätä WWW-sivun ja rattaiden kytköstä ei kuvaa mikään muu kuin yhteinen paikka, ilman sen tarkempaa etuliitettä. Yhteistä niillä on se, että ne liittyvät johonkin toimintaan jossain temaattisessa roolissa.

MuseoSuomen näkymiä, KulttuuriSammon ensimmäisen prototyypin näkymiä sekä luotua tietomallia analysoimalla päädyin yhteensä seitsemään näkymään. Ne ovat aineistotyyppi, toimijat, toiminta, toiminnan väline, paikka, aika ja kokoelma. Kaikki näkymät aineistotyyppiä ja kokoelmaa lukuun ottamatta on sidottu toimintakeskeiseen tietomalliin. Seuraavaksi käyn tarkemmin läpi näkymät. Annan jokaisesta näkymästä sanallisen selityksen ja kerron säännöt, joilla kategoriapuut muodostetaan. Esitän myös näkymien ripustussäännöt, joiden mukaan päätetään liittyykö tietty kohde tiettyyn kategoriaan. Säännöt on kuvattu Prolog-kielellä SWI-Prolog<sup>4</sup> toteutuksen semantic web -apukirjaston avulla. Säännöissä hyödynnetään kirjaston RDF-predikaatteja, joille annetaan RDF-tietomallin mukaiset subjekti, predikaatti ja objekti. RDF-tietomallin tunteva osaa näin ollen lukea koodia, vaikka ei Prologia sen enempää tuntisikaan. Kuvassa 7.4 on esimerkki paikka-näkymästä käyttöliittymän esitysmuodossa, jossa näkymä esitetään avattavana ja suljettavana kategoriapuuna. Esimerkissä Eurooppa-kategorian solmu on avattu. Luvut kategorioitten nimien jälkeen kertovat, kuinka monta artefaktia kuhunkin kategoriaan liittyy.

<sup>4</sup><http://www.swi-prolog.org/>



**Kuva 7.4. Esimerkki käyttöliittymän moninäkömahaun näkymästä**

SWI-Prologin semantic web apukirjaston lisäksi säännöissä on käytetty joitain itse määriteltyjä RDFS-apupredikaatteja, joiden nimet pääosin kuvaavat niiden toimintaa. Esimerkiksi `rdfs_subClassOf/2` on tosi, mikäli ensimmäinen muuttuja on toisen muuttujan transitiivinen aliluokka, eli tulee jossain vaiheessa vastaan, kun seurataan RDFS:n `subClassOf`-kaarta. `rdf_reachable/3` on edellisen predikaatin yleistys, jolle voi antaa minkä tahansa kaaren ja selvittää ovatko muuttujat tavoitettavissa kaaren avulla. Tätä hyödynnetään temaattisten roolien hierarkiaa tutkittaessa. Esimerkissä 7.1 Property-muuttujan arvoksi sidotaan kaikki `yso:actualityInstance`-luokan aliluokat, eli mikä tahansa temaattinen rooli. Tämä predikaatti on käyttökelpoinen tilanteissa, joissa temaattisen roolin valinnalla ei ole merkitystä.

### Esimerkki 7.1. `rdf_reachable`-apupredikaatti

```
rdf_reachable(Property,
'http://www.w3.org/2000/01/rdf-schema#subPropertyOf',
'http://yso.fi/YSO#actualityInstance')
```

### Aineistotyyppi

Aineistotyyppinäkömä on näkymistä yksinkertaisin. Se jakaa portaalin sisällön aineistotyyppin mukaan kategorioihin sen mukaan onko kyseessä esine, valokuva, maalaus tai jokin muu KulttuuriSammon aineistotyypeistä. Kategoriapuun projektio lähtee liikkeelle `http://kulttuurisampo.fi/annotaatio#kulttuuriObjekti`-luokasta, joka on

kaikkien kokoelmakohteiden juuriluokka ja lisää kulttuuriObjektin aliluokat rekursiivisesti juurikategorian alisolmuiksi (esimerkki 7.2). Kohteiden aineistotyyppiä verrataan tämän jälkeen kategorian määrittävään aineistotyyppiin `rdf:type` ominaisuutta käyttäen (esimerkki 7.3). Aineistotyyppi-näkymä ei näin ollen ole sidottu toimintakeskeiseen malliin, vaan tarjoaa mallin ulkopuolisen mekanismin haun rajoittamiseksi.

### **Esimerkki 7.2. Aineistotyyppinäköymän kategoriapuun säännöt**

```
view_sub_category(URI, SubCatURI) :-
    rdfs_subClassOf(SubCatURI, URI).
```

### **Esimerkki 7.3. Aineistotyyppinäköymän ripustussäännöt**

```
view_leaf( ItemURI, CategoryURI) :-
    rdf( ItemURI, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type', CategoryURI ).
```

## **Toimijat**

Toimijat-näkymä sisältää kaikki ne resurssit, jotka on merkitty toimijoiksi johonkin toimintakehykseen. Kategoriapuun projektion juuri on näin ollen `http://kulttuurisampo.fi/annotaatio#toimijat`, jonka aliluokkina ovat henkilöt ja organisaatiot. Koska toimijat itsessään eivät ole haun alaisia kohteita, sisällytetään myös toimijainstanssit kategoriapuuhun. Eli kaikki henkilöinstanssit ovat listattuna ja valittavissa hakuavaimeksi henkilöt-solmun alta. Käytännössä toimintakehyksen toimijaksi voi asettaa minkä tahansa resurssin, kuten YSO:n tai MAO:n hevosen käsitteen. Prototyypin aineistossa tällaisia kuvauksia ei kuitenkaan ollut, joten toimijat-luokan aliluokat ja instanssit sisältävät kaikki annetut toimijat.

### **Esimerkki 7.4. Toimijat-näköymän kategoriapuun säännöt**

```
toimijat_sub_category(URI, SubCatURI) :-
    rdfs_subClassOf(SubCatURI, URI).

% Tässä näkymässä myös instanssit kuuluvat kategoriapuuhun.
% Näin voidaan hakea esim. kaikki Alvar Aaltoon liittyvät kohteet.
toimijat_sub_category(URI, SubCatURI) :-
    rdf(SubCatURI, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type', URI).
```

## Esimerkki 7.5. Toimijat-näkymän ripustussäännöt

```
% Kohde on mukana toiminnassa, jossa agentin roolissa on kategorian toimija
toimijat_leaf(ItemURI, CategoryURI) :-
    rdf(Toiminta, 'http://yso.fi/YSO#agentInstance', CategoryURI),
    rdf_reachable(Property, 'http://www.w3.org/2000/01/rdf-schema#subPropertyOf',
        'http://yso.fi/YSO#actualityInstance'),
    rdf(Toiminta, Property, ItemURI),
    rdf(ItemURI, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type', Tyyppi),
    rdfs_subclass_of(Tyyppi, 'http://kultuurisampo.fi/
annotaatio#kulttuuriObjekti').
```

```
% Kohteella on toimintakehys, jossa agenttina on kategorian toimija
toimijat_leaf(ItemURI, CategoryURI) :-
    rdf(Toiminta, 'http://yso.fi/YSO#agentInstance', CategoryURI),
    rdf(Scene, 'http://yso.fi/tilanne#toiminta', Toiminta),
    rdf(ItemURI, 'http://kultuurisampo.fi/annotaatio#hasScene', Scene).
```

## Toiminta

Toimintanäkymässä on koko inhimillisen toiminnan kirjo. Kategoriapuu projisoidaankin YSO:n <http://yso.fi/YSO#toiminta> käsitteestä lähtöisin, seuraten luokan aliluokkahierarkiaa. Kohteet liittyvät tiettyyn toimintanäkymän kategoriaan, mikäli ne ovat osallisena toimintakehyksessä, jossa keskeinen toiminta on annettu toiminta tai sen alitoiminta, tai omistavat sellaisen toimintakehyksen <http://kultuurisampo.fi/annotaatio#hasScene> kaaren perusteella.

## Esimerkki 7.6. Toimintanäkymän kategoriapuun säännöt

```
view_sub_category(URI, SubCatURI) :-
    rdfs_subClassOf(SubCatURI, URI).
```

## Esimerkki 7.7. Toimintanäkymän ripustussäännöt

```
toiminta_leaf( ItemURI, CategoryURI) :-
    rdf(Toiminta, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type', CategoryURI),

    rdf(Scene, 'http://yso.fi/tilanne#toiminta', Toiminta),
    rdf(ItemURI, 'http://kultuurisampo.fi/annotaatio#hasScene', Scene).
```

```
toiminta_leaf( ItemURI, CategoryURI) :-
    rdf(Toiminta, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type', CategoryURI),
    rdf_reachable(Property, 'http://www.w3.org/2000/01/rdf-schema#subPropertyOf',
        'http://yso.fi/YSO#actualityInstance'),
    rdf(Toiminta, Property, ItemURI),
    rdf(ItemURI, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type', Tyyppi),
    rdfs_subclass_of(Tyyppi, 'http://kultuurisampo.fi/
annotaatio#kulttuuriObjekti').
```

## Toiminnan väline

Toiminnan väline -näkyssä hakua voi rajoittaa sen mukaan, minkälaisia resursseja toiminnan yhteydessä on käytetty. Rakentamisessa käytetään esimerkiksi vasaraa tai poraa ja maalaamisessa pensseliä. Näitä toiminnan välineenä toimivia resursseja kuvataan toimintakehyksissä temaattisella roolilla yso:resourceInstance ja sen alirooleilla, kuten yso:materialInstance ja yso:instrumentInstance. Kategoriapuu projisoidaan käytettyjen referenssiontologioiden (YSO ja MAO) juurista. Kohde liittyy kategoriaan, mikäli sen toimintakuvauksessa resurssina on käytetty kategorialuokan instanssia.

### Esimerkki 7.8. Toiminnan väline -näkyvän kategoriapuun säännöt

```
toiminnan_valine_sub_category(URI, SubCatURI) :-
    URI == 'http://www.yso.fi/kulttuurisampo/nakymat#toiminnan_valine',
    (
        view_sub_category('http://yso.fi/mao#mao-luokat', SubCatURI);
        view_sub_category('http://yso.fi/mao#uudet', SubCatURI);
        view_sub_category('http://yso.fi/YSO#fyysinen_objekti', SubCatURI)
    ).

toiminnan_valine_sub_category(URI, SubCatURI) :-
    view_sub_category(URI, SubCatURI).
```

### Esimerkki 7.9. Toiminnan väline -näkyvän ripustussäännöt

```
toiminnan_valine_leaf(ItemURI, CategoryURI) :-
    rdf(ResourceInstance, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type',
        CategoryURI),
    rdf_reachable(ResourceProperty, 'http://www.w3.org/2000/01/rdf-
schema#subPropertyOf', 'http://yso.fi/YSO#resourceInstance'),
    rdf(Toiminta, ResourceProperty, ResourceInstance),
    rdf(Scene, 'http://yso.fi/tilanne#toiminta', Toiminta),
    rdf(ItemURI, 'http://kulttuurisampo.fi/annotaatio#hasScene', Scene).

toiminnan_valine_leaf(ItemURI, CategoryURI) :-
    rdf(ResourceInstance, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type',
        CategoryURI),
    rdf_reachable(ResourceProperty, 'http://www.w3.org/2000/01/rdf-
schema#subPropertyOf', 'http://yso.fi/YSO#resourceInstance'),
    rdf(Toiminta, ResourceProperty, ResourceInstance),
    rdf_reachable(Property, 'http://www.w3.org/2000/01/rdf-schema#subPropertyOf',
        'http://yso.fi/YSO#actualityInstance'),
    rdf(Toiminta, Property, ItemURI),
    rdf(ItemURI, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type', Tyyppi),
    rdfs_subclass_of(Tyyppi, 'http://kulttuurisampo.fi/
annotaatio#kulttuuriObjekti').
```

## Paikka

Kaikki tekeminen tapahtuu aina jossain paikassa. Paikka voi olla maantieteellinen tai viitata tiloihin, kuten taloihin tai autoihin, joissa asioita voi tapahtua. Käytetyssä aineistossa on

vain maantieteellisiä paikkoja, joten näkymä sisältää vain kaikki maantieteelliset paikat. Näkymän juuri on paikkaontologian <http://www.yso.fi/kulttuuriSampo/paikat#maapallo>-käsite. Paikkaontologian käsittehierarkia hajaantuu kokonaisuus-rakenteella pienemmiksi osiksi. paikka:kokonaisuus-kaarta käytetään näin ollen kategoriapuun muodostamiseksi. Kohde liittyy paikkaan, mikäli toimintakuvauksessa on käytetty yso:placeInstance-roolin arvona kategorian paikkaa.

### Esimerkki 7.10. Paikkanäkymän kategoriapuun säännöt

```
paikka_sub_category(URI, SubCatURI) :-
  rdf( SubCatURI, 'http://www.yso.fi/kulttuuriSampo/paikat#kokonaisuus', URI).
```

### Esimerkki 7.11. Paikkanäkymän ripustussäännöt

```
paikka_leaf( ItemURI, CategoryURI) :-
  rdf(Toiminta, 'http://yso.fi/YSO#placeInstance', CategoryURI),
  rdf_reachable(Property, 'http://www.w3.org/2000/01/rdf-schema#subPropertyOf',
    'http://yso.fi/YSO#actualityInstance'),
  rdf(Toiminta, Property, ItemURI),
  rdf(ItemURI, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type', Tyyppi),
  rdfs_subclass_of(Tyyppi, 'http://kulttuurisampo.fi/
annotaatio#kulttuuriObjekti').
```

```
paikka_leaf( ItemURI, CategoryURI) :-
  rdf(Toiminta, 'http://yso.fi/YSO#placeInstance', CategoryURI),
  rdf(Scene, 'http://yso.fi/tilanne#toiminta', Toiminta),
  rdf(ItemURI, 'http://kulttuurisampo.fi/annotaatio#hasScene', Scene).
```

## Aika

Aikakaudet on kuvattu aikaontologian <http://www.yso.fi/kulttuuriSampo/ajat#timePeriod> luokan mukaisesti. Kategorian juuri on ajat:ajat. Tietämyskannan aikoja ei kuitenkaan ole sidottu projektiossa käytetyn ontologian aikoihin, vaan ne noudattavat omaa aikaskaemaansa. Näin ollen ripustuksen yhteydessä on peilattava tietämyskannan arvot aikaontologian aikakäsitteisiin. Tämä tehdään vertailemalla tietämyskannan ajankäsitteiden alku- ja loppuvuosia ontologian aikakäsitteiden alku- ja loppuvuosiin. Kohde ripustautuu kategoriaan, mikäli toimintakehyksen timeInstance-kaaren arvo viittaa aikaan, jonka alku- tai loppuvuosi on ontologian aikajakson sisällä, tai mikäli se kattaa koko ajanjakson.

### Esimerkki 7.12. Aikanäkymän kategoriapuun säännöt

```
% subclasses hierarchy
aika_sub_category(URI, SubCatURI) :-
  rdfs_subClassOf(SubCatURI, URI),
  not( SubCatURI == URI ),
  not( URI == 'http://www.yso.fi/kulttuuriSampo/ajat#timePeriod' ),
  not( SubCatURI == 'http://www.yso.fi/kulttuuriSampo/ajat#timePeriod' ).
```



### Esimerkki 7.13. Aikanäkymän ripustussäännöt

```
% Peilaa aikaoliot ajat.owl ontologian aikoihin
aika_leaf(ItemURI, CategoryURI) :-
  rdf(CategoryTime, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type',
    CategoryURI),
  rdf(CategoryTime, 'http://www.yso.fi/kulttuuriSampo/ajat#ajan_alkuvuosi',
    literal(Start)),
  rdf(CategoryTime, 'http://www.yso.fi/kulttuuriSampo/ajat#ajan_loppuvuosi',
    literal(End)),
  rdf(ItemURI, 'http://kulttuurisampo.fi/annotaatio#hasScene', Scene),
  rdf(Scene, 'http://yso.fi/tilanne#toiminta', Toiminta),
  rdf(Toiminta, 'http://yso.fi/YSO#timeInstance', Time),
  rdf(Time, 'http://kulttuurisampo.fi/annotaatio#alkuAika', literal(Start2)),
  rdf(Time, 'http://kulttuurisampo.fi/annotaatio#loppuAika', literal(End2)),
  nonvar(Start2),
  nonvar(End2),
  nonvar(Start),
  nonvar(End),
  catch(atom_to_integer(Start2, Start2Int), _, fail),
  catch(atom_to_integer(End2, End2Int), _, fail),
  catch(atom_to_integer(Start, StartInt), _, fail),
  catch(atom_to_integer(End, EndInt), _, fail),
  covers(Start2Int, End2Int, StartInt, EndInt).

% Alku ajanjaksossa
covers(Start1, End1, Start2, End2) :-
  Start1 > Start2,
  Start1 < End2.

% Loppu ajanjaksossa
covers(Start1, End1, Start2, End2) :-
  End1 > Start2,
  End1 < End2.

% Sisältyy täysin ajanjaksoon
covers(Start1, End1, Start2, End2) :-
  Start1 < Start2,
  End1 > End2.
```

### Kokoelma

Kokoelmanäkymä on aineistotyyppinäkymän tyylinen rajaava näkymä. Tämä näkymä on hyödyllinen, mikäli kiinnostukseen kohteet ovat jostain tietystä kokoelmasta. Kategoriapuun juuri on <http://kulttuurisampo.fi/annotaatio#kokoelma> ja kokoelman instanssit ovat juuren alisolmuja. Kohteet ripustetaan kategorioihin, mikäli <http://kulttuurisampo.fi/annotaatio#kokoelma> predikaatin arvo on kyseinen kokoelmainstanssi.

### Esimerkki 7.14. Kokoelmanäkymän kategoriapuun säännöt

```
kokoelma_sub_category(URI, SubCatURI) :-
  rdf(SubCatURI, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type', URI).
```

### **Esimerkki 7.15. Kokoelmanäkymän ripustussäännöt**

```
kokoelma_leaf(ItemURI, CategoryURI) :-
    rdf(ItemURI, 'http://kulttuurisampo.fi/annotaatio#kokoelmassa', CategoryURI).
```

Nämä edellä esittelemäni säännöt muodostavat näkymäpuut, joita käytetään moninäkymähaussa hakujen muodostamisessa. Seuraavassa alaluvussa kerron kuinka käyttöliittymällä tehdään hakuja.

#### 7.2.2 Hakujen tekeminen

Lopullinen käyttöliittymä koostuu hakusivusta ja kohdesivusta. Kuvassa 7.5 on ruudunkaappaus hakusivusta. Hakusivun vasemmassa ylälaudassa on vapaatekstihakukenttä, johon voi kirjoittaa yksittäisiä sanoja tai useampia sanoja yhdellä kertaa. Hakukentässä voi myös käyttää sekä boolean-hakuja että wildcard-hakuja.

Vapaatekstihaun alapuolella avautuvat kaikki KulttuuriSammon näkymät: aineistotyyppi, toimijat, toiminta, toiminnan väline, aika, paikka ja kokoelma. Jokainen näkymä on itsenäinen puu, jonka solmuja voi avata ja sulkea. Puun solmujen vieressä on numero, joka kertoo hakutulosten määrän, mikäli solmu valittaisiin. Solmu valitaan hakuavaimeksi klikkaamalla solmun nimeä. Niillä solmuilla, joilla on alasolmuja, on näkymäpuussa linkki "ryhmittele kategorian mukaan". Tämä valinta ryhmittelee hakutulokset alasolmujen mukaisiin ryhmiin.

[Uusi haku](#)

Vapaatekstihaku



NÄKYMÄT

Aineistotyyppi

**kokoelma** (69)

- teos** (11283)  
Ryhmittele kategorian mukaan
- Kuvataideteos** (3945)  
Ryhmittele kategorian mukaan
- Grafiikka** (369)
- Piirustus** (296)
- Veistos** (66)
- muséesine** (4909)
- Narratiivi** (36)  
Ryhmittele kategorian mukaan

[Avaa kaikki](#) - [Sulje kaikki](#)

Toimijat

**Henkilöt** (34)  
Ryhmittele kategorian mukaan

[Avaa kaikki](#) - [Sulje kaikki](#)

Toiminta

**kognitio** (37)  
Ryhmittele kategorian mukaan

**luominen** (37)  
Ryhmittele kategorian mukaan

**mentaalinen toiminta** (37)  
Ryhmittele kategorian mukaan

**olemassaolo** (37)  
Ryhmittele kategorian mukaan

**ruumiintoinnot ja hoito** (1)  
Ryhmittele kategorian mukaan

**vuorovaikutus** (37)  
Ryhmittele kategorian mukaan

[Avaa kaikki](#) - [Sulje kaikki](#)

Toiminnan väline

**aine** (17)  
Ryhmittele kategorian mukaan

**konkreettinen eloton objekti** (24)  
Ryhmittele kategorian mukaan

**ns\_30\_maoryhma:materiaalit\_ ja\_ain** (2)  
Ryhmittele kategorian mukaan

[Avaa kaikki](#) - [Sulje kaikki](#)

Aika

**masa:aikakaudet** (29)  
Ryhmittele kategorian mukaan

**vuosisadat** (29)  
Ryhmittele kategorian mukaan

[Avaa kaikki](#) - [Sulje kaikki](#)

Paikka

[Avaa kaikki](#) - [Sulje kaikki](#)

Kokoelma

**Ateneumin taidemuseo** (27)

**Kansatieteen kuvakokoelma** (8)

**Sinebrychoffin taidemuseo** (2)

[Avaa kaikki](#) - [Sulje kaikki](#)

Hakuehdot

[\[poista kaikki\]](#)

Kategoriat:

**Aineistotyyppi:**  
valokuva  OR Maalaus

Hakusana:










tyt\* 

37 osumaa

Ryhmitely Kartta Aikajana

ei ryhmittelyä

&lt;&lt; &lt; 1 2 3 4 5 &gt;&gt;

 tyttökoulun peruskivi lasketaan Kisamäellä, lyseon lähellä	 Vaimo ja tytär	 tytöt ruokivat kanoja
 Pitkälän tytöt kaloja perkaamassa	 Vihreähiuksinen tyttö	 Kaksi kangasta värjäävää tyttöä
 Tytön pää (Anna Slöör) ; Tyttö Kalelän interioörissä	 Ompeleva tyttö	 vanha nimi: taitelijan kaksi tyttäret

## Kuva 7.5. KulttuuriSammon hakusivu

Keskellä sivua on tilalaatikko, jossa näytetään valitut kategoriat ja hakusanat. Punaisesta X:stä kategorian voi poistaa hakuehdosta. Vaihtoehtoisesti sen voi poistaa näkymäpuusta klikkaamalla valittua solmua toistamiseen. Hakunäytön alapuolella on laskuri, joka kertoo osumien määrän.


Hakutuloksia voi tarkastella kolmella eri tavalla: ryhmiteltynä, kartalle projisoituna ja aikajanalle projisoituna. Ryhmitelty näkymä on sivutettu ja hakutuloksia voi selata vaihtamalla sivua nuolista tai valitsemalla tietyn sivunumeron.


Kaikista hakutuloksen esitystavoista pääse kohdesivulle klikkaamalla hakutuloksesta haluttua dokumenttia. Valinnasta käyttäjä siirtyy toiselle sivulle, jossa hän voi tarkastella kohteen metatietoja, asiasanoja ja kohteeseen liittyviä toimintakuvauksia. Kohdesivu näyttää aina kuvan kohteesta ja se tarjoaa myös pikanavigointinapit hakutuloksen seuraavaan ja edelliseen dokumenttiin. Esimerkki kohdesivusta on kuvassa 7.6.

**KULTTUURISAMPO**  
Hönnäkyshaku


**Navigointi**

Takaisin hakusivulle

Edellinen  
 **Pelkosen pariskunta**

Seuraava  
 **lehtikuusimetsää**

**SORTAVALAN SATAMA**



**Kohdetiedot**

mitat: 8 x 11 cm  
valmistusaika: 1909  
type: valokuva  
kuvaajat: Jääskeläinen, V.  
kokoelmassa: Kansatieteen kuvakokoelma  
valmistuspaikka: Lappajärvi  
tekijät: Jääskeläinen, V.  
nimi: Sortavalan satama  
id: 4878:592

**Asiasanat**

esiliinat satamat höyrylaivat purjelaivat pelastusveneet liput huivit

**Narratiivinen kuvailu**

Toiminta  
**Valokuvaus**  
tekijä Jääskeläinen, V.  
toimija Jääskeläinen, V.  
kohde **Sortavalan satama**

Toiminta  
**Sijaite**  
kohde **Sortavalan satama**  
paikka **Kansatieteen kuvakokoelma**

### **Kuva 7.6. KulttuuriSammon kohdesivu, jossa kohteen kuva, metatiedot, asiasanat ja toimintakuvaukset**

Seuraavaksi kerron miten toteutin kartta- ja aikajanaprojektiot toimintakeskeisessä ympäristössä.

#### **7.2.3 Hakutulosten esitystavat**

Oletusarvoisena tulosten tarkastelutapana on ryhmitelty näkymä, jossa hakutulokset listataan sivutettuun tuloslistaan. Hakutuloksia voi tässä näkymässä ryhmitellä näkymäkategorioiden perusteella. Jokaisen kategorian alla on *ryhmittele kategorian mukaan* -linkki, jota klikkaamalla hakutulokset jaetaan kategorian alakategorioiden mukaisiin ryhmiin. Tämä ryhmittely auttaa hakutulosten analysoinnissa ja lisähakujen tulosten ennakoinnissa. Kuvassa 7.7 hakutulos on ryhmitelty paikkanäkymän kategorian *Lappi* mukaan. Kuvassa nähdään miten tulokset on ryhmitelty Lapissa sijaitsevien paikkojen mukaisesti.

The screenshot shows a search interface with a sidebar on the left containing various filters. The main area displays search results for the category 'Lappi'. The filters include 'Käsitteet', 'Aineistotyyppi', 'Toimijat', 'Toiminta', 'Toiminnan väline', 'Aika', and 'Paikka'. The search results are displayed in a grid format, showing items like 'voiaastia', 'paahine', 'kasineet', 'lakki', 'mekko', and 'mekko'. The search results are filtered by the category 'Lappi'.

**Kuva 7.7. Hakutulokset ryhmiteltyinä kategorian *Lappi* mukaan**

Ontologioiden käyttö tarjoaa mahdollisuuden hyödyntää muualla kerättyä tietoa omissa sovelluksissa. KulttuuriSamossa metatietojen sijaintiviittaukset sidottiin paikkaontologiaan ja ajalliset hahmot aika-ontologiaan. Itsenäisesti kehitettyyn paikkaontologiaan tuotiin Geologian tutkimuskeskuksen paikkatiedot, ja näin paikkaontologian paikat saatiin sidottua koordinaatteihin. Tämä mahdollistaa sen, että kohteet voi helposti esittää karttapalvelussa.

KulttuuriSampo varten otimme käyttöön Googlen tarjoaman Google Maps -karttapalvelun<sup>5</sup>. Google Maps tarjoaa ohjelmointirajapinnan, jonka avulla Googlen tarjoamalle kartalle voi esimerkiksi asettaa pisteitä merkkamaan merkityksellisiä alueita. Otimme Google Mapsin käyttöön hakutulosten näyttämisen yhteydessä. KulttuuriSammon käyttäjä voi

<sup>5</sup><http://maps.google.com>

valita tulosjoukon tarkastelutavaksi kartan, ja hakutulos projisoidaan kartalle. Esimerkki hakutuloksen projisoinnista kartalle on kuvassa 7.8.



**Kuva 7.8. Hakutulos kartalle projisoituna**

Esimerkissä 7.16 on havainnollistettu SPARQL-kyselykielen avulla karttaprojektion toimintaperiaate. SPARQL on kyselykieli RDF-muotoiselle datalle, jolla voi muodostaa kyselyitä tietämuskantaan hyvin samantapaisesti kuin SQL-kyselykielellä relaatiotietokantoihin. SPARQL-kyselyissä voi asettaa pakollisia ja valinnaisia graafihahmoja sekä näiden konjunktioita ja disjunktioita. Graafihahmot muodostetaan tripleteistä, joissa joko subjekti, objekti tai predikaatti jätetään vapaaksi muuttujaksi. Muuttuja merkataan kyselyssä kysymysmerkillä ja sitä seuraavalla muuttujan nimellä. Kyselyn aikana muuttujiin sidotaan arvoja, joita kysely sitten voi palauttaa vastauksena, joko tulosjoukkoina tai RDF-graafeina. (Prud'hommeaux & Seaborne 2007)

Projisointi tapahtuu esimerkissä hyvin yksinkertaisesti etsimällä hakutuloksesta kohteita, joiden toimintakehyksissä on viittauksia paikkoihin, joilla on koordinaattitiedot. Esimerkissä

<#uri#> on hakutuloksen kohteen, kuten tietyn taulun URI-tunniste. Mikäli kohteeseen liittyy tapahtumia, ja tapahtumiin toimintaa, jossa temaattisella roolilla *paikka* on ilmaistu toiminnan tapahtumapaikka, tarkistetaan löytyykö paikkaontologiasta paikalle koordinaatteja. Mikäli tällaisia paikkoja löytyy, palauttaa kysely toiminnan, kohteen ja paikan tiedot koordinaatteineen. Kyselyiden paluuarvot järjestetään paikan mukaan ja niistä muodostetaan Google Maps ohjelmointirajapinnan avulla kartalle paikkakohtaisia pisteitä, jotka sisältävät kaikki kyseiseen paikkaan liittyvät hakutulokseen sisältyvät kohteet.

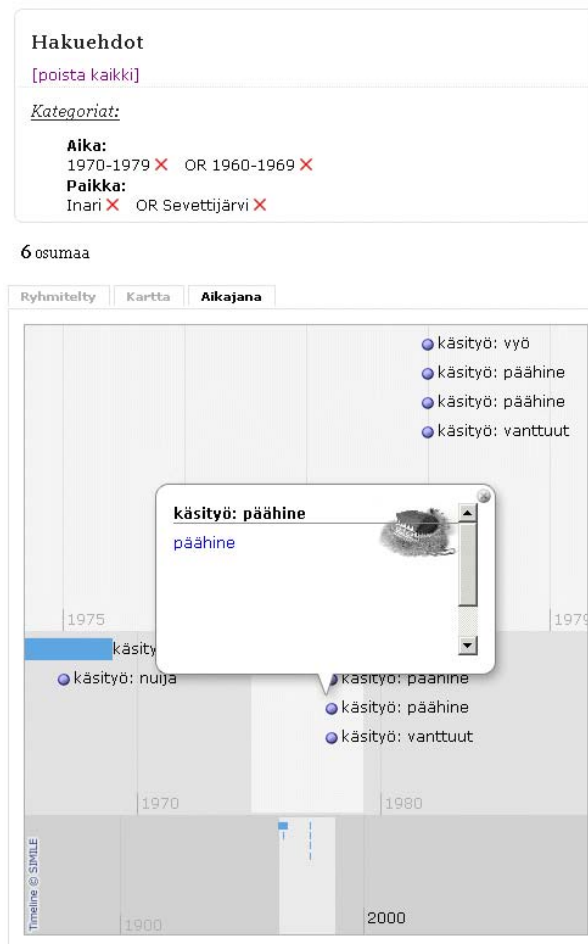
### **Esimerkki 7.16. Karttaprojektion toimintaperiaate SPARQL-kyselynä**

```
SELECT ?toiminta ?placeinstance ?valueLabel ?itakoordinaatti ?
pohjoiskoordinaatti ?itemuri
WHERE
<#uri#> <http://kulttuurisampo.fi/annotaatio#hasScene> ?y.
?itemuri <http://kulttuurisampo.fi/annotaatio#hasScene> ?y.
?y <http://yso.fi/tilanne#toiminta> ?toiminta.
?toiminta <http://yso.fi/YSO#placeInstance> ?placeinstance.
?placeinstance <http://www.yso.fi/kulttuuriSampo/paikat#itakoordinaattiWGS_84> ?
itakoordinaatti.
?placeinstance <http://www.yso.fi/kulttuuriSampo/
paikat#pohjoiskoordinaattiWGS_84> ?pohjoiskoordinaatti.
?placeinstance <http://www.w3.org/2000/01/rdf-schema#label> ?valueLabel.
```

Toinen kiinnostava näkökulma kulttuurikohteista puhuttaessa on ajallinen näkökulma, eli millä tavalla kulttuuriesineet levittäytyvät aikajanelle. SIMILE on kehittänyt Google Maps -tyylisen vapaasti käytettävän aikajanakomponentin<sup>6</sup>, jota hyödynsimme KulttuuriSamossa. Tulosjoukon tarkastelutavaksi voi käyttöliittymästä valita aikajanan, jolloin siirrytään tarkastelemaan kolmitasoista aikajanaa, jolle hakutulokset on projisoitu.

---

<sup>6</sup><http://simile.mit.edu/timeline/>



**Kuva 7.9. Hakutulos aikajanelle projisoituna**

Projisointi tapahtuu hyvin pitkälle samalla tavalla kuin karttaprojektiossa, navigoimalla toimintakeskeisen mallin mukaista RDF-verkkoa ja etsimällä viittauksia vuosilukuihin. Esimerkissä 7.17 on jälleen SPARQL:n avulla havainnollistettu toimintaperiaatetta. Ongelmia tuottavat yleisesti käytetyt "1800-luvun alku" -tyyliset sanalliset kuvaukset aikakaudesta. Ensimmäisessä versiossa näitä ei oteta huomioon, vaan tyydymme projisoimaan kartalle ainoastaan ne kohteet, joiden vuosiluvut ovat selkeästi annettu.

### **Esimerkki 7.17. Aikajanaprojektion toimintaperiaate SPARQL-kyselynä**

```
SELECT ?toiminta ?timeInstance ?alkuAika ?loppuAika ?itemuri
WHERE
<#uri#> <http://kulttuurisampo.fi/annotaatio#hasScene> ?y.
?itemuri <http://kulttuurisampo.fi/annotaatio#hasScene> ?y.
?y <http://yso.fi/tilanne#toiminta> ?toiminta.
?toiminta <http://yso.fi/YSO#timeInstance> ?timeInstance.
?timeInstance <http://kulttuurisampo.fi/annotaatio#alkuAika> ?alkuAika.
?timeInstance <http://kulttuurisampo.fi/annotaatio#loppuAika> ?loppuAika.
```



### 7.3 KulttuuriSammon sovellusarkkitehtuuri

KulttuuriSammon ensimmäinen versio oli toteutettu OntoViews-alustalla, jota tutkimusryhmässä oli kehitetty usean aikaisemman portaaliprojektin aikana (Mäkelä et al. 2004). Toivottujen toiminnallisuuksien valossa Apache Cocoonin<sup>7</sup> päälle toteutettu OntoViews ei mielestäni kuitenkaan ollut järkevä alusta jatkokehitykselle. Pahimmat puutteet näin käyttöliittymäkerroksessa. OntoViewsia käytettäessä käyttöliittymä olisi toteutettava täysin XSLT-muunnoksia käyttäen, muuntaen RDF:ää suoraan HTML:ksi. Tämä sitoo käyttöliittymäkerroksen niin tiukasti tietomallikerrokseen, etteivät mitkään osat aikaisemmissa projekteissa kehitetystä XSLT-koodista olisi olleet uudelleenkäytettäviä.

KulttuuriSammon toista versiota varten oli siis kehitettävä uusi ohjelmistoarkkitehtuuri, joka mahdollistaisi visioiden toteuttamisen. Suunnittelin yhdessä Eetu Mäkelän kanssa uuden arkkitehtuurin käytettäväksi kaikissa tutkimuslaboratorion samankaltaisissa semanttisen webin prototyypisovelluksissa. Arkkitehtuurin tuli tarjota helposti uudelleenkäytettävät rajapinnat yhteisiin taustasovelluksiin ja tarjota mahdollisimman selkeät ja löyhästi kytketyt kerrokset eritasoisten toiminnallisuuksien toteuttamiseksi, jotta kehitetyt palvelut ja komponentit hyödyttäisivät koko tutkimusryhmää jatkossakin. Semanttisen webin sovelluskehitystä varten oli lisäksi viime vuosina ilmaantunut lukuisia kiinnostavia apukirjastoja, joiden käyttöä halusimme projektin yhteydessä kokeilla.

Arkkitehtuuriratkaisumme on niin sanottu monikerroksinen arkkitehtuuri (n-tier architecture), jonka ideana on jakaa eri toiminnallisuutta tarjoavat osat selkeästi omiin kerroksiinsa ja kytkeä nämä kerrokset löyhästi yhteen. Näillä ratkaisuilla saavutetaan uudelleenkäytettävyyttä ja rajoitetaan muutosten vaikutusaluetta. Arkkitehtuurin kerrokset käyvät ilmi kuvasta 7.10.

Käyttöliittymä- ja sovelluskerrokseen, joka oli OntoViewsin pahin ongelmakohta, toimme Tapestry-sovelluskehityksen<sup>8</sup>. Tapestry toteuttaa Model View Controller (MVC) -mallin<sup>9</sup> WWW-ohjelmoinnissa ja tarjoaa mahdollisuuden luoda uudelleenkäytettäviä komponentteja ja käyttää valmiita komponenttitoteutuksia. Tapestrylle on tarjolla runsaasti valmiita käyttöliittymäkomponentteja ja omia on helppo tehdä. Tapestrya käyttämällä toivoimme saavamme aikaiseksi tutkimusryhmälle valmiiden komponenttien varaston, jota olisi mahdollista hyödyntää helposti myös muissa projekteissa.

KulttuuriSampo varten toteutin Tapestry-karttakomponentin Google Maps -palvelun päälle. Komponentin avulla hakutulokset voi projisoida kartalle. Toteutin myös SIMILE-tutkimusryhmän tarjoaman JavaScript-aikajanahan päälle aikajanakomponentin, joka

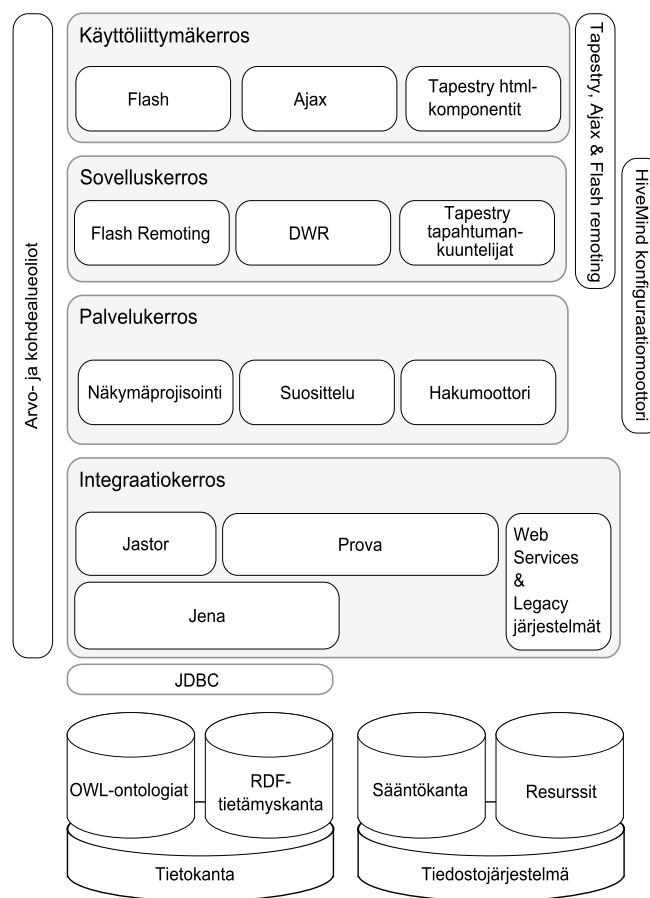
<sup>7</sup><http://cocoon.apache.org/>

<sup>8</sup><http://tapestry.apache.org/>

<sup>9</sup>Katso esimerkiksi (Gamma et al. 1995)

puolestaan projisoi hakutuloksen kohteet ajallisiin vuorovaikutteisiin janoihin. Molemmat komponentit hyödyntävät asynkronisia XML-RPC kutsuja, joita kutsutaan Ajax-tekniikaksi (Garrett 2005). Komponentit esittelin tarkemmin luvussa 7.2.3.

Sovelluserroksen alle toteutimme palveluserroksen, joka tarjoaa kaikissa SeCo:n portaaliprojekteissa tarvittavia yhteisiä palveluita, kuten moninäkömähaku-, projektio- ja suosittelumootorit. Nämä palvelut puolestaan on toteutettu integraatiokerroksen päälle, joka tarjoaa palveluille liityntärajapinnat tietokannoissa, sääntökannoissa ja tiedostojärjestelmässä sijaitseviin resursseihin. Integraatiokerros pitää sisällään erilaisia RDF:n ja ontologioiden hyödyntämiseen suunnattuja apukirjastoja, kuten Jena<sup>10</sup>, Jastor<sup>11</sup> ja Prova<sup>12</sup>.



### Kuva 7.10. KulttuuriSammon arkkitehtuuri

Tutkimusryhmässä päätimme jatkaa sovelluskehitystä Javalla. Halusimme pitää sovelluskehityksen mahdollisimman kevyenä, joten emme ottaneet käyttöön J2EE-

<sup>10</sup><http://jena.sourceforge.net/>

<sup>11</sup><http://jastor.sourceforge.net/>

<sup>12</sup><http://www.prova.ws/>

sovelluspalvelinta, vaan tyydyimme Apachen Tomcat Java-palvelimeen<sup>13</sup>, jonka voi pienikokoisuudesta johtuen helposti upottaa mukaan projektirunkoon.

KulttuuriSammon ohjelmistoarkkitehtuurista tuli tavoitteiden mukainen ja modulaarisuuden hyödyt nousivat esiin toistuvasti kehityksen aikana. Niin moninäkömähakumoottori kuin tietokantojen toteutustapa muuttui useaan otteeseen projektin aikana. Selkeiden moduulien ja rajapintojen ansiosta tukipalveluiden muutokset eivät kertaakaan heijastuneet käyttöliittymäkomponentteihin asti.

## 7.4 KulttuuriSammon toteutuksen yhteenveto

Tässä luvussa olen kuvannut mitä työvaiheita portaalin toteutus piti sisällään ja minkälaisilla teknisillä ratkaisulla hakujärjestelmä toteutettiin. Kerroin aluksi minkälaiseen tietomalliin päädyimme. Toimintakeskeinen tietomalli oli ratkaisumallimme heterogeenisen aineiston yhdistämisen ongelmaan. Museoilta saatujen aineistojen muuntaminen toimintakeskeiseen malliin oli monivaiheinen prosessi, jonka lopputuloksena oli tietämyskanta, jossa aineistot oli kuvattu toimintakeskeisesti.

Hakutapa toteuttamassani hakujärjestelmässä on toimintakeskeinen semanttinen moninäkömähaku, joka on semanttisen moninäkömähäun laajennus. Esitin Prolog-koodin, jolla näkymät muodostetaan osoittaakseni näkömuodostuksen yksinkertaisuuden. Kerroin myös millä tavalla tietämyskannan aineistot projisoidaan kartalle ja aikajanalle. Myös näiden komponenttien toteutuksissa toimintakeskeinen malli osoittautui helppokäyttöiseksi. Lopullinen käyttöliittymä muodostuu haku- ja kohdesivusta. Kerroin millä tavalla hakuja muodostetaan ja mitä tietoja artefakteista esitetään.

Lopuksi esittelin hakujärjestelmän sovellusarkkitehtuurin. Toteuttamaani portaalia varten jouduimme tutkimusryhmässä uusimaan sovellusarkkitehtuurin modulaarisemmaksi ja löyhemmin kytketyksi. Uuden arkkitehtuurin hyödyt nousivat esiin moneen otteeseen projektin aikana pohjapalveluita tuottavien komponenttien kehittyessä.

---

<sup>13</sup><http://tomcat.apache.org/>

## 8 Järjestelmän evaluointi

Semanttisen haun evaluointiin ja tulosten analysointiin ei ole olemassa yleisesti hyväksyttyä menetelmää (Rocha et al. 2004). Rocha et al. ovat myös sitä mieltä, että tiedonhaussa yleisesti käytetyt tarkkuus ja saanti eivät olisi semanttisessa haussa oleellisia mittareita. Kirjallisuuskatsauksen perusteella olen samaa mieltä evaluointimenetelmien puutteesta. Saanti ja tarkkuus puolestaan ovat mielestäni yhtä oleellisia semanttisessa haussa kuin perinteisessä tekstihaussa. Kuten luvussa 2.3.3 totesin, ei multimediatiedonhakuun tai semanttiseen hakuun kuitenkaan ole olemassa testimateriaalia, jossa relevantit dokumentit hakutapauksia kohden olisi määritelty.

Järjestelmää ei näin ollen voi evaluoida laboratoriomallin mukaisesti saannin ja tarkkuuden mittareilla. Käytettävissä oleva aineisto on niin laaja, että dokumenttien relevanssin arvioiminen hakutapauksia vasten olisi diplomityön puitteissa mahdoton tehtävä. Koska absoluuttista relevanssimittaria ei ole käytettävissä, on evaluointi tehtävä subjektiivisilla mittareilla käyttäjäkeskeisen evaluointimallin mukaisesti. Käyttäjiltä on hakutehtävien yhteydessä kysyttävä, ovatko tulostulosten dokumentit heidän mielestään relevantteja.

Tässä luvussa kerron minkälaisilla koejärjestelyillä evaluoin toteuttamaani hakujärjestelmää. Koejärjestelyt ja kokeen rakenne perustuivat Borlundin (2003) käyttäjäkeskeiseen evaluointimalliin, jossa simuloitujen työtehtävätilanteiden suorituksista kerätään määrällistä ja laadullista aineistoa.

### 8.1 Kokeen rakenne ja koejärjestelyt

Evaluoinnin kohteena oli luvussa 7 kuvaamani KulttuuriSampo-hakujärjestelmä. Koe suoritettiin käyttäjäkeskeisen evaluointimallin mukaisesti. Pyysin koehenkilöitä asettumaan aineiston ja hakujärjestelmän käyttötarkoituksiin pohjautuviin simuloituihin työtehtävätilanteisiin, ja keräsin tehtävien suorituksista sekä määrällistä että laadullista dataa.

#### 8.1.1 Koehenkilöt

Kokeet tein toukokuussa 2007 kahden viikon ajanjakson aikana. Koehenkilöt olivat miehiä ja naisia iältään 19-45 vuotta. Koehenkilöiden keski-ikä oli 26,3 vuotta. Yhteensä 12 henkilöä, joista 8 miehiä ja 4 naisia, osallistui kokeeseen. Koehenkilöt olivat opiskelijoita (9 kpl), tutkijoita (2 kpl) ja järjestelmäsuunnittelijoita (1 kpl). He eivät tunteneet KulttuuriSampo-järjestelmää entuudestaan, mutta osa tunsivat tutkimusryhmän tutkimuksen pääpiirteet. Yhden kokeen tekemiseen kuului keskimäärin lähes puolitoista tuntia, josta tehtäväsuorituksen osuus oli 28 minuuttia.

### 8.1.2 Kerätty aineisto

Koehenkilöiden suoriutumista tehtävistä seurasin useilla menetelmillä, joilla pyrin saamaan testitilanteesta mahdollisimman paljon tietoa irti. Tiedon keruuseen käytin kyselyjä, lokikirjoitusta, havainnointia, videokuvan kaappausta ja äänen tallennusta sekä vapaamuotoista haastattelua.

Tärkeimpänä tiedon keruun menetelmänä olivat kyselyt. Ennen tehtävien suorittamista koehenkilö täytti alkukyselyn, jossa kysyin perustietoja, kuten ikää, sukupuolta ja tietokoneenkäyttökokemusta. Jokaisen tehtävän jälkeen koehenkilö täytti tehtävää koskevan kyselylomakkeen, jossa kysyin tehtävään liittyviä kysymyksiä sekä järjestelmän toimintaan liittyviä kysymyksiä. Kaikkien tehtävien suorittamisen jälkeen koehenkilö vastasi vielä loppukyselyyn, jossa keskityin kokonaisvaikutelmiin koko koetilanteen ajalta. Kyselyillä keräsin sekä määrällistä että laadullista aineistoa. Keräämälläni aineistolla pyrin myös selvittämään testin mielekkyyttä ja testihenkilöiden mielipiteitä tehtävistä. Kyselylomakkeisiin otin mallia kysymyksenasettelun ja ulkoasun osalta Whiten (2004) tutkimuksesta. Kokeeni kyselylomakkeet ovat liitteissä B, C ja D.

Tehtäväkohtaisessa kyselylomakkeessa pyysin koehenkilöitä arvioimaan suoritettua simuloitua työtehtävätilannetta, hakuprosessia, tulosjoukkoa ja järjestelmää semanttisella differentiaaliasteikolla. Loppukyselyssä käytin semanttista differentiaaliasteikkoa järjestelmän arvioimiseen. Semanttista differentiaaliasteikkoa käytettäessä vastaajat merkitsevät oman arvionsa asteikolle, jonka päissä ovat vastakkaiset adjektiivit. Käyttämäni adjektiiviparit muokkasin kuvahakujärjestelmiä arvioineesta tutkimuksesta (Jose et al. 1998, s. 232-241).

Järjestin adjektiiviparit viisiportaisella Likertin asteikolle siten, että parissa esiintyvä positiivisen merkityksen perusjäsen, esimerkiksi "luotettava", on korkeamman arvon puolella, ja sen johdos, esimerkiksi "epäluotettava", pienemmän arvon puolella. Arviot ovat siis sitä parempia, mitä korkeampia niistä annetut pisteet ovat. Arvot 1 ja 5 merkitsevät voimakasta kallistumista kyseisen adjektiivin puoleen ja arvosanat 2 ja 4 lievempää kallistumista. Arvosana 3 on neutraali molempien adjektiivien suhteen. Sanallisesti eri arviot merkitsivät esimerkkiparin kohdalla seuraavaa: 1 = epäluotettava, 2 = melko epäluotettava, 3 = ei luotettava eikä epäluotettava, 4 = melko luotettava, 5 = luotettava.

Laadullista aineistoa keräsin kuvaruudunkaappausohjelmistolla<sup>1</sup>, joka tallensi koetilanteet videotiedostoiksi. Videotiedostoon tallentui myös ääni tietokoneen mikrofonista. Pyysin koehenkilöitä ajattelemaan koetilanteessa ääneen ja kertomaan mitä he tekevät ja minkä takia. Näitä videotiedostoja voidaan analysoida jälkikäteen. Ääniraidan kanssa ne paljastavat erityisesti konfliktit käyttäjän mallissa ja portaalin todellisessa toimintamallissa. Mallien

<sup>1</sup><http://camstudio.org/>

konfliktit ovat yleinen syy käytettävyysoongelmiin. Videotiedostoista voi myös jälkikäteen käydä tarkistamassa kyselyiden vastausten perusteella kiinnostavia tapahtumia.

Kyselyiden ja videokaappauksen lisäksi tallensin koetilanteesta niin sanottua jäljituspalkua. Ideana on kirjata lokitiedostoon kaikki käyttäjän interaktiot järjestelmän kanssa. KulttuuriSammon jäljituspalkun toteutin käyttöliittymän controller kerrokseen log4j-lokikirjoituskehyksellä<sup>2</sup>. Controller-kerros ottaa vastaan käyttäjän hiiren napin painalluksista syntyneet pyynnöt palvelimella. Pyyntöjen lisäksi lokiin kirjataan myös muita sovelluksesta saatavia tietoja, kuten sivujen latautumisaikoja. Latautumisaika saattaa joissain tapauksessa selittää hyvinkin paljon kyselyissä saatua huonoa palautetta.

Jokainen tapahtuma jäljituspalkussa on aikaan sidottu, joten se voidaan yhdistää videokaappaukseen tekstityksenä antamaan täydellisen kuvan käyttäjän interaktioista. Toisaalta jäljituspalkun tietoja voi käyttää määrällisenä tietona ja sen perusteella laskea esimerkiksi hakuiden solmujen avautumisen lukumäärä tehtävää kohti. Jäljituspalkun kaikki tapahtumatyyppit on kuvattu liitteessä E ja esimerkki tehtävän suorituksen aikana kerätystä jäljituspalkusta on liitteessä F.

### 8.1.3 Kokeen kulku

Borlund (2003) korostaa, että tulosten luotettavuuden varmistamiseksi koejärjestelyiden käytäntöjen on oltava mahdollisimman vakioituja. On hyvin tärkeää, että kaikki koehenkilöt saavat samanlaisen käsittelyn. Kokeen esittely, hakujärjestelmän esittely ja koejärjestelyistä kertominen on tehtävä samalla tavalla kaikille, jotta testin tulokset ovat keskenään vertailukelpoisia. Kokeessani kaikki koehenkilöt saivat samat ohjeet, opastukset ja tehtävänannot alla listatun ohjelman mukaan.

1. Koehenkilö toivotetaan tervetulleeksi ja hänelle kerrotaan yleisellä tasolla tutkimuksesta.
2. Koehenkilölle annetaan yksilöivä tunniste, joka merkitään kaikkiin tiedon keruuseen käytettäviin asiakirjoihin, lomakkeisiin ja lokeihin.
3. Koehenkilöä pyydetään vastaamaan taustatietokyselylomakkeeseen.
4. Koehenkilölle annetaan lyhyt opastus hakujärjestelmän käyttöön. Hänelle esitellään käyttöliittymä kohta kohdalta ennalta laaditun listan mukaisessa järjestyksessä. Näin varmistetaan, että jokainen koehenkilö saa samat lähtötiedot.
5. Koehenkilön annetaan harjoitella hakujärjestelmän käyttöä vapaasti korkeintaan 5 minuutin ajan.

---

<sup>2</sup><http://logging.apache.org/log4j/>

6. Koehenkilölle annetaan ensimmäinen tehtävänanto. Koehenkilön oppimisen vaikutuksen minimoimiseksi, tehtävien järjestys on jokaiselle koehenkilölle erilainen.
7. Koehenkilöä pyydetään vastaamaan tehtäväkohtaiseen kyselylomakkeeseen.
8. Loput tehtävänannot annetaan koehenkilölle yksi kerrallaan. Jokaisen tehtävän jälkeen koehenkilö vastaa tehtäväkohtaiseen kyselylomakkeeseen.
9. Lopuksi koehenkilö täyttää koko koetta koskevan loppukyselyn. Loppukyselyn jälkeen häntä haastatellaan vielä vapaamuotoisesti kokemuksista.

Minimoidakseni oppimisen vaikutuksen tuloksiin, tasapainotin tehtävien järjestyksen latin square -menetelmällä (MacKenzie 2002). Luomani koehenkilökohtainen tehtäväjärjestys on esitetty taulukkona liitteessä A.

Koejärjestelmän alustana käytin IBM ThinkPad T43 kannettavaa tietokonetta, jossa on 1,5 GB muistia, 1.86GHz Intel Pentium M -prosessori ja käyttöjärjestelmänä Microsoft Windows XP Professional Version 2002 SP2. Kannettavan käyttäminen helpotti koejärjestelyjä, koska pystyin kuljettamaan koko järjestelmän helposti mukani. Erityisesti videokuvan kaappaus söi kokeissa prosessoritehoja, joten sen vaikutus on otettava huomioon hakujärjestelmän vasteikatuloksia arvioitaessa. Todellinen suorituskyky tuotantokäyttöön rakennetulla sovelluspalvelimella ja erillisellä tietokantapalvelimella olisi huomattavasti parempi. Sujuvan videokuvankaappauksen mahdollistamiseksi laskin näytön resoluution 1024 pikselin tarkkuudelle. Myös tekstin koko näytöllä oli näin suurempi ja heikkonäköistenkin luettavissa.

## 8.2 Tehtävät

Tehtävänasettelun tavoitteena oli, että koehenkilöt käyttäisivät järjestelmää mahdollisimman realistisesti. Realistisuuteen päästäkseni kirjoitin ja laadin hakutehtävät simuloitujen työtehtävätilanteiden muotoon (Borlund 2003). Simuloidun työtehtävätilanteen tarkoitus on herättää koehenkilössä mahdollisimman aito tiedon tarve.

Evaluointia varten laadin kuusi simuloitua työtehtävätilannetta. Tehtävät kirjoitin niin, että ne heijastavat Ingwersenin ja Järvelinin (2005, s. 290-293) määrittelemiä, alaluvussa 2.1 kuvaamiani tiedon tarpeen kategorioita. Asemoimalla tehtävät tällä tavalla, toivoin saavani tarkempaa tietoa hakujärjestelmän soveltuvuudesta erilaisiin hakutilanteisiin. Kategorioista jätin pois epätarkat tehtävät, joissa lähtötiedot ovat jollakin tavalla virheelliset tai puutteelliset. En nähnyt tarpeelliseksi tarkoituksella harhaanjohtaa koehenkilöitä koetilanteessa. Näin ollen Ingwersenin ja Järvelinin jaottelun mukaisia tehtäviä oli neljä kappaletta.

Tiedontarpeen jaotteluun perustuvien tehtävien lisäksi laadin kaksi lisätehtävää. Ensimmäisen tarkoitus oli ohjata koehenkilöitä kohti käyttöliittymän paikka- ja aikajanakomponenttien käyttöä. Koska kaikki tehtävät on mahdollista suorittaa ilman näiden komponenttien käyttöä, saisin tämän tehtävän avulla suuremmalla todennäköisyydellä tuloksia myös niiden osalta. Toinen lisätehtävä oli vaikeampi tulkintatehtävä, jossa koehenkilön oli muodostettava perusteltu mielipide jostain asiasta portaalin tarjoamien työkalujen avulla. Tämän tehtävän asettamat raamit olivat löyhimmät, ja toivoinkin koehenkilön kokeilevan kaikkia eri tapoja hakea ja analysoida aineistoa ja muodostavan mielipiteensä mieluisimmasta tavasta käyttää portaalin erilaisia haku- ja analyysimenetelmiä.

Kaikki tehtävänannot pohjautuivat haettavissa olevaan aineistoon, suunnitteluvaiheessa määriteltyihin käyttötapauksiin ja kohdekäyttäjryhmien oletettuihin tarpeisiin. Tehtävänannot eivät siis pohjautu koehenkilöiden todellisiin tarpeisiin ja kiinnostuksen kohteisiin, vaan pyrkivät asettamaan koehenkilöt kuviteltuihin rooleihin. Tämä helpotti myös koehenkilöiden rekrytoimista. Tehtävät noudattavat taulukon 8.1 jakoa tehtävätyyppeihin.

### Taulukko 8.1. Tehtäväasettelu

Tehtävän tunniste	Tehtävätyyppi
A	Aika- ja paikkahaku. Pyrkii ohjaamaan hakijaa kartan tai aikajanan käyttöön.
B	Tunnetun kohteen haku. Tunnetulla metadatala pyritään löytämään dokumentti, johon metadata liittyy. Tunnetun kohteen hausta on kyse, kun kirjaston tietokannasta etsitään kirjaa kirjailijan nimen perusteella.
C	Tulkintatehtävä. Tulkintatehtävässä hakijan on kerättävä tiedon palasia eri keinoja käyttäen ja tulkitsemalla keräämiänsä tietoja muodostettava vastaus esitettyyn kysymykseen.
D	Tunnetun tiedon haku. Tunnetulla metadatala halutaan löytää tiedon palanen, kuten muita metatietoja. Esimerkiksi henkilön puhelinnumeron etsiminen nimen perustella kuuluu tähän luokkaan.
E	Tunnetun aiheen tai sisällön haku. Tunnetuilla sisällöllisillä attribuuteilla pyritään löytämään dokumentteja tai dokumenttijoukkoja. Tiedonhakija voi esimerkiksi olla kiinnostunut löytämään kaikki koira-aiheiset valokuvat valokuvatietokannasta.
F	Faktatiedonhaku. Tunnetuilla sisällöllisillä attribuuteilla pyritään löytämään faktatietoa. Esimerkiksi kuka, missä ja milloin on maalannut kolmiosaisen Kalevala-aiheisen taulun?



Tehtävänantoja oli siis yhteensä kuusi kappaletta. Nämä tehtävät annettiin koehenkilöille paperilla koehenkilökohtaisen tehtäväjärjestyksen mukaisessa järjestyksessä. Luettelen seuraavaksi nämä tehtävät.

### Tehtävä A

Uusien sähkökanteleiden myötä olet innostunut perinteikkästä kanteleen soitosta. Haluaisit nyt tutkia tarkemmin soittimen kehitystä läpi vuosien.

Etsi KulttuuriSammosta vanhin kantele ja selvitä, miltä paikkakunnalta se on peräisin. Kirjaa tiedot muistiin. Ilmoita tutkijalle, kun olet valmis tai kun sinusta tuntuu, ettei haku enää edisty.

### Tehtävä B

Digikuvauksen innoittamana osallistut työväenopiston valokuvauskurssille. Yhtenä kurssin pakollisena harjoitustyönä on esitelmän pitäminen kiinnostavasta suomalaisesta valokuvaajasta. Olet kuullut, että valokuvaaja I.K. Inha on kuvannut 1800- ja 1900-lukujen taitteessa paljon suomalaisia maisemia ja elämää. Haluaisit nyt nähdä näitä valokuvia ja tutustua niihin tarkemmin, jotta voisit arvioida, onko I.K. Inha kiinnostava esitelmän aihe.

Etsi 5 kappaletta I.K. Inhan ottamia valokuvia, ja kirjaa kuvien id:t muistiin. Ilmoita tutkijalle, kun olet valmis tai kun sinusta tuntuu, ettei haku enää edisty.

### Tehtävä C

Olet menossa kaupunkilomalle Helsinkiin pariksi päiväksi. Kulturellina ihmisenä haluat tietysti käydä taidenäyttelyssä. Olet erityisen kiinnostunut maalauksista. Et kuitenkaan ole varma, mihin museoon mennä, koska sinulle on epäselvää, miten museoiden kokoelmat eroavat toisistaan.

Selvitä KulttuuriSammon avulla, millä tavoin Helsingissä sijaitsevien Ateneumin taidemuseon ja Nykytaiteen museon kokoelmat eroavat toisistaan. Kirjoita muistiin tärkeimmät löytämäsi erot. Ilmoita tutkijalle, kun olet valmis tai kun sinusta tuntuu, ettei haku enää edisty.

### Tehtävä D

Akseli Gallen-Kallela on maalannut useita tunnettuja Kalevala-aiheisia tauluja, kuten Aino-taru, Sammon taonta ja Lemminkäinen Tuonelassa. Haluat selvittää taideopintojesi ainetta "Kalevalan kuvittajat" varten missä järjestyksessä taulut on maalattu.

Selvitä, minä vuosina yllä mainitut taulut on maalattu, ja listaa ne kronologisessa järjestyksessä. Ilmoita tutkijalle, kun olet valmis tai kun sinusta tuntuu, ettei haku enää edisty.

### Tehtävä E

Olet tekemässä lehtijuttua luomuviljelystä, ja kaipaat kuvitukseksi kuvia vanhoista viljelymenetelmistä ja käytetyistä työkaluista Suomessa. Olet kuullut, että KulttuuriSampo sisältää paljon vanhoja valokuvia ja esineitä, joten päätät etsiä sopivaa kuvitusta KulttuuriSammosta.

Hae järjestelmästä 3 kappaletta aiheeseen sopivaa kuvaa, ja kirjaa jokaisen kohteen id muistiin. Ilmoita tutkijalle, kun olet valmis tai kun sinusta tuntuu, ettei haku enää edisty.

### Tehtävä F

Mummosi on muuttamassa omakotitalosta kerrostaloasuntoon. Muuton yhteydessä löysit mummosi omakotitalon vintiltä kiinnostavan esineen (kuvassa alla), jonka käyttötarkoitusta tai alkuperää et tunne. Olet kuullut, että KulttuuriSammossa on paljon tietoa museoesineistä. Päätät etsiä KulttuuriSammosta samankaltaisia esineitä, joiden avulla löytämäsi esineen voisi tunnistaa.

Selvitä KulttuuriSammon avulla, mikä esine on kyseessä, ja kirjaa tieto muistiin. Ilmoita tutkijalle, kun olet valmis tai kun sinusta tuntuu, ettei haku enää edisty.



## 9 Tulokset

Tässä luvussa raportoin edellisessä luvussa kuvatun kokeen tuloksia, analysoin tuloksia tarkemmin ja pohdin niiden merkitystä työni kannalta.

### 9.1 Hakutehtävien suoritus

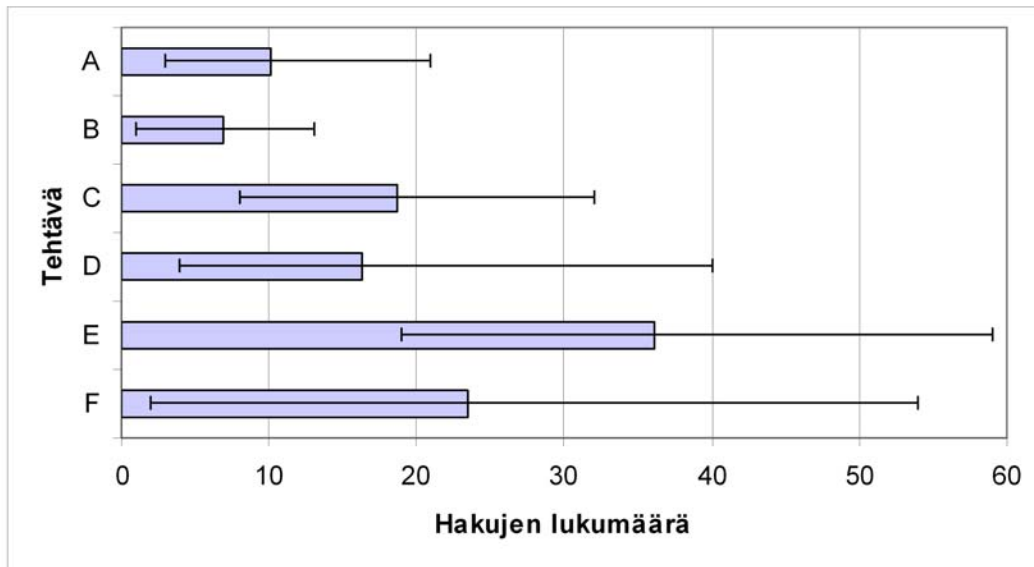
Aluksi pureudun tehtävien suorittamisen aikana keräämäni datan analysointiin.

#### 9.1.1 Hakumäärät

Taulukossa 9.1 on esitetty tehtäväkohtaiset lukumäärät hauille ja muita hakumääriä kuvaavia tunnuslukuja. Taulukosta nähdään, että tehtävässä E tehtiin keskimäärin selvästi eniten hakuja (36,1 hakua). Suurimmillaan tehtävän suoritus vaati jopa 59 hakua. Tämä oli hieman yllättävää, sillä alkujaan en pitänyt tehtävää vaikeimpana. Tehtävä vaati kuitenkin johdonmukaisesti eniten hakuja, mistä kertoo maltillinen keskihajonta. Tehtävän F keskihajonta puolestaan on selkeästi korkein ja ylittää muiden tehtävien arvot moninkertaisesti. Korkean keskihajonnan syy on tehtävänasettelussa, jossa annettiin ainoastaan kuva etsittävästä esineestä. Osa koehenkilöistä tiesi kuitenkin heti, mikä esine on kyseessä, kun taas toisten oli etsittävä esinettä sen kuvasta pääteltävien ominaisuuksien perusteella. Selkeästi vähiten hakuja tehtiin tehtävässä B, jossa etsittiin I.K. Inhan valokuvia ilman tarkempia sisällöllisiä kriteereitä. Tehtävän B helppous oli odotettua, ja hakujen määrä olisi ollut jopa nykyistä pienempi, mikäli pyydettyjen dokumenttien määrä olisi ollut pienempi kuin (turhan suuri) 5 kappaletta. Kuva 9.1 havainnollistaa hakujen lukumäärien eroja tehtäväkohtaisesti.

#### Taulukko 9.1. Tehtävissä suoritettujen hakujen määrät

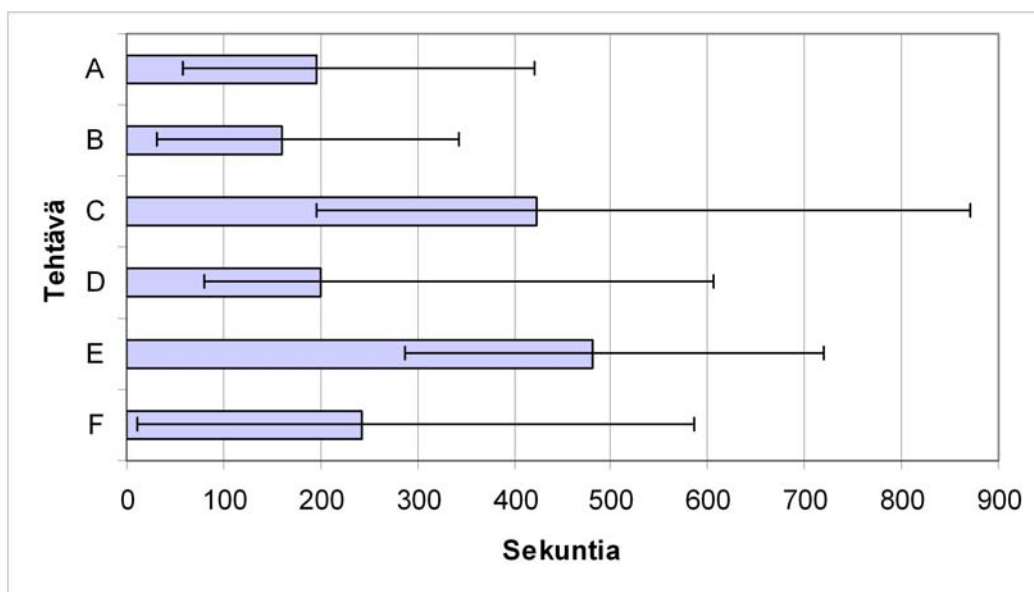
Tehtävä	A	B	C	D	E	F
Hakuja yhteensä	122	83	224	195	433	282
Keskiarvo / käyttäjä	10,2	6,9	18,7	16,3	36,1	23,5
Mediaani / käyttäjä	9	7	17,5	11,5	35,5	19
Keskihajonta	6,5	4,3	7,7	11,8	10,1	19,2
Maksimi	21	13	32	40	59	54
Minimi	3	1	8	4	19	2



**Kuva 9.1. Tehtäväkohtaiset tehtyjen hakujen keskiarvot, maksimit ja minimi**

### 9.1.2 Tehtäväsuoritusten kestot

E-tehtävän suorittaminen vei keskimäärin pisimmän ajan. Tehtävän suoritukseen käytettiin keskimäärin noin 8 minuuttia. Myös C-tehtävässä suoritusajat venyivät pitkiksi. Keskimäärin suoritus aika oli noin 7 minuuttia, mutta jotkut henkilöt käyttivät tähän tehtävään jopa 16 minuuttia. A-, B-, D- ja F-tehtävien kestoissa ei ollut suuria eroja. Tehtäväsuoritusten kestot ja hakujen lukumäärä korreloivat vahvasti keskenään (0,81), mikä nähdään myös vertailemalla kuvia 9.1 ja 9.2. C-tehtävässä tulkintatyö on vienyt paljon aikaa, eikä hakuja siitä syystä ole suhteessa tehty yhtä paljon.

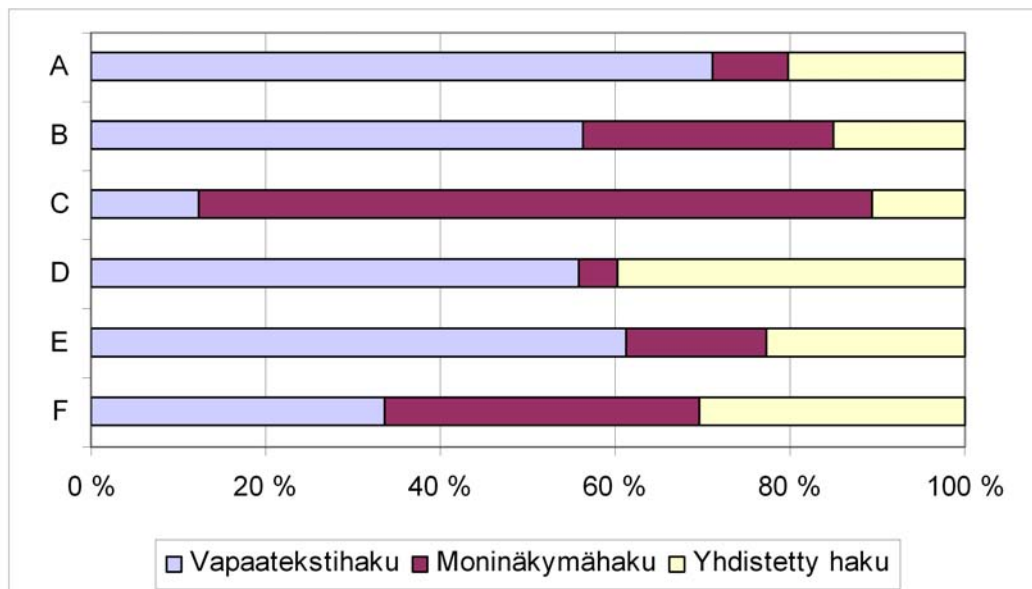


**Kuva 9.2. Tehtäväkohtaiset suoritusajojen keskiarvot, maksimit ja minimi**

### 9.1.3 Hakutavat

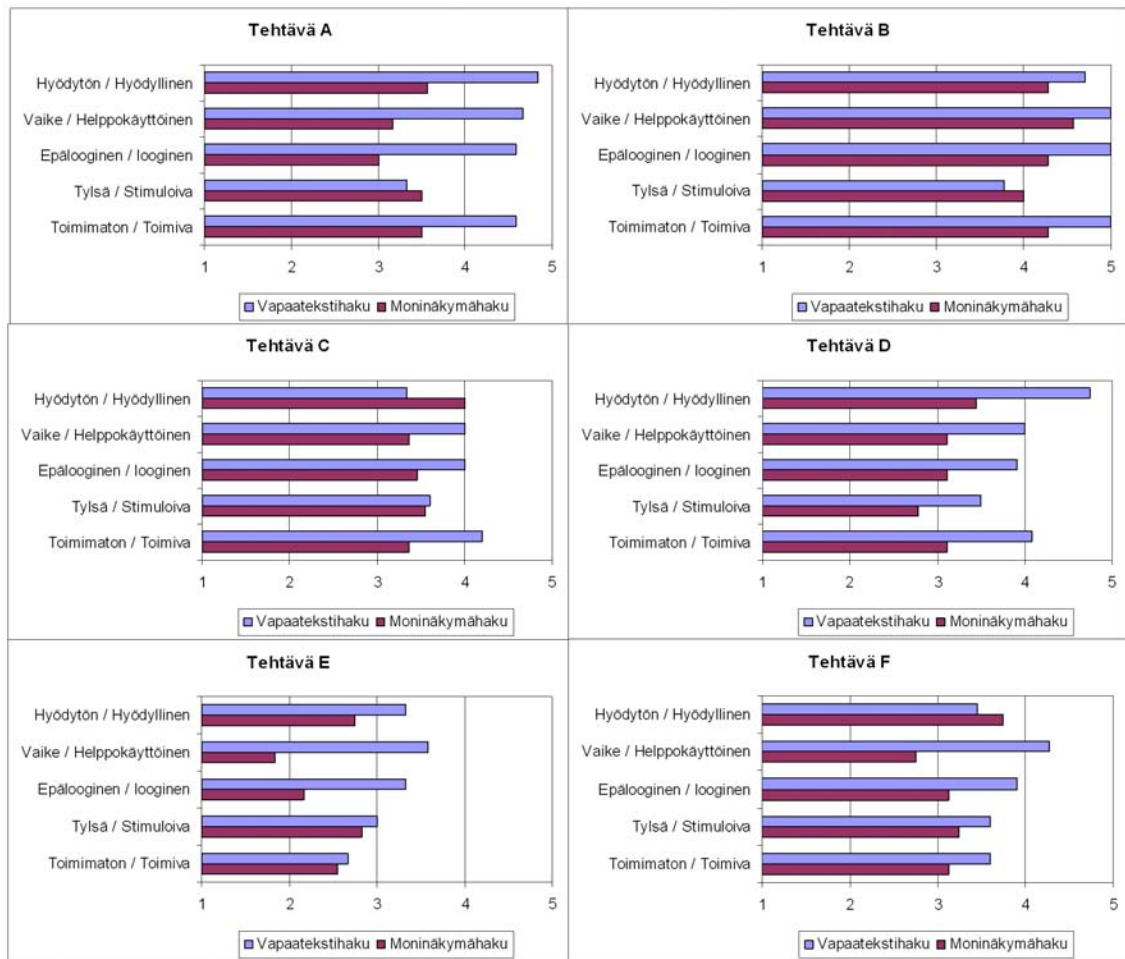
Kaiken kaikkiaan hauista 47 % oli puhtaita vapaatekstihakuja eli sellaisia, joissa käyttäjällä ei ollut valittuna yhtään kategoriala näkymäpuista tekstihakua tehdessä. Puhtaita moninäkömähakuja hakujen kokonaismäärästä oli 29 %. Loput 24 % olivat näiden kahden hakutapojen yhdistelmiä eli sellaisia, joissa käytettiin samaan aikaan sekä kategoriavalintoja että vapaatekstihakua.

Kuvassa 9.3 on esitetty hakutapojen suhteellinen käyttö tehtäväkohtaisesti. Kuvasta huomataan selkeitä tehtäväkohtaisia eroja hakutapojen käytössä. C-tehtävässä, joka oli tulkintatehtävä, käytettiin moninäkömähakua selvästi eniten. A ja E-tehtävissä sen sijaan luotettiin vapaatekstihaun voimaan. Näissä tehtävissä yli 60 % hauista oli vapaatekstihakuja. Yhdistelmähakua tehtiin kaikissa tehtävissä. D-tehtävässä, jossa tehtiin suhteellisesti vähiten puhtaita moninäkömähakuja, sai yhdistetty hakeminen suurimman painoarvon.



**Kuva 9.3. Hakutapojen suhteellinen käyttö tehtäväkohtaisesti**

Jokaisen tehtäväsuorituksen jälkeen kysyin testihenkilöiltä, mitä mieltä he olivat eri hakutavoista. Kuvassa 9.4 on yhteenveto kyselyjen tuloksista tehtäväkohtaisina keskiarvoina. Kuvasta huomataan, että vapaatekstihaku on arvioitu paremmaksi lähes kaikilla mittareilla. Ainoastaan hyödyllisyydessä moninäkömähaku on arvioitu tekstihakua paremmaksi tehtävissä C ja F ja stimuloivuudessa tehtävissä A ja B. Erot eivät kuitenkaan ole kovin suuria. Suurimmat erot hakutapojen välillä muodostuivat helppokäyttöisyyttä kysyttäessä. Laskettaessa keskiarvot kaikkien tehtävien yli, ovat molempien hakutapojen arviot kaikkien mittarien osalta plussan puolella eli  $> 3$ . Näin ollen molemmat hakutavat on kokeessa arvioitu hyödyllisiksi, helppokäyttöisiksi, loogisiksi, stimuloiviksi ja toimiviksi. Moninäkömähäun marginaali on kuitenkin selvästi pienempi.



**Kuva 9.4. Hakutapojen arviointi tehtäväkohtaisesti**

Käyttäjätestien aikana moninäkömähakua kritisoitiin. Näkymien sisältöä pidettiin vaikeasti ymmärrettävänä ja käyttöliittymää sekavana.

"Haku kategorioista selaamalla oli haastavaa, sillä aina kategorioiden nimistä ei pystynyt loogisesti päättämään, mitä kategoria sisältää."

"Moninäkömähaku ei tuonut lisäarvoa. 5 tehtävää kuudesta tein pelkällä tekstihaulla. Moninäkömähaku oli epälooginen ja sekava johtuen sanojen luokittelusta kuten "sjömanin kokoelma", "masa:aikakaudet" tai "kognitio"."

"Kategorioiden käyttöön pitäisi tutustua paremmin, hypähdykset ja viiveet haittasivat käyttöä. kategorioiden opasteet?"

"Näkömähaku oli hieman sekava ja sitä kautta vähän turha."

Järjestelmällä oli hakutuloksen oletustarkastelutavan lisäksi mahdollista projisoida hakutuloksia kartalle ja aikajanalle. Karttaprojektioa käytettiin selvästi vähiten. Vain 3 % hakutulosten esityksistä tapahtui kartalle projisoituna. Aikajanaprojektio oli hieman

suositumpi. Aikajanalla hakutuloksia tarkasteltiin 9 % osuudella. Tuloksissa on otettava huomioon, että ryhmitelty-esitystapa oli oletusesitystapa, ja näin ollen automaattisesti keräsi suurimman osuuden.

Hakutulosten projisointia aikajanalle ja kartalle keuhuttiin testien aikana.

"Karttanäkymä ja aikajananäkymä olivat mielenkiintoisia. Olisi tehnyt mieli tutkia kokoelmia tarkemmin niiden avulla."

"Karttasysteemi oli todella hieno ja innovatiivinen. Samoin aikajana."

Ominaisuuksia pidettiin hyvinä, vaikka niissä esiintyi ongelmia. Sekä kartta- että aikajanaprojektioita vaivasi se, että vain osa hakutuloksista pystyttiin näyttämään kartalla. Esimerkiksi paikkatiedot maalauksista puuttuivat kokonaan, jolloin projisointi oli mahdotonta.

"Yritin käyttää aikajanaa, mutta valokuvat eivät ilmestyneet aikajanalle, vaikka niistä osassa oli kuvausaika tiedossa. Siirryin tekstihakuun."

Käyttöliittymä ei millään tavalla kertonut, kuinka moni hakutuloksista pystyttiin projisoimaan kartalle tai aikajanalle. Tämän asian korjaaminen on helppoa, ja tällainen laskuri oli jo lokikirjoituksessa käytössä.

Toinen projektoiden käytettävyysoongelma oli se, että kartan ja aikajanan tila ei tallentunut, ja käyttäjän oli joka kerta etsittävä oikea paikka uudelleen toisella sivulla käymisen tai uuden hakuehdon lisäämisen yhteydessä. Tämä ärsytti monia testihenkilöitä, ja saattoi vaikuttaa hakuihin kuluneeseen aikaan sekä mielipiteisiin järjestelmästä.

"Aikajana voisi tallentaa edellisen tapahtuman, eikä aina kadottaa missä edellinen objekti oli."

Aikajanalle toivottiin muitakin lisäyksiä. Yksi toivottu lisäys oli se, että aikajana hakeutuisi automaattisesti ensimmäisen tai viimeisen tapahtumaan kohdalle aikajanalla, ja että näihin paikkoihin olisi myös pikalinkit. Toinen pieni lisäys olisi tiimalasin näyttäminen silloin kuin tapahtumia vielä etsitään. Nyt oli monesti epäselvää, onko haku päättynyt, löytyikö yhtään osumaa ja se, onko aikajana vain täysin väärässä ajassa. Yhdessä parannetun hakutulosten määrän esityksen kanssa, näillä melko yksinkertaisilla toimenpiteillä aikajanan käytettävyys paranisi huomattavasti.

Tehtävissä A ja D pyydettiin hakutuloksia jossain tietyssä järjestyksessä. Näiden tehtävien suoritusten yhteydessä nousi esiin KulttuuriSammon puutteellinen tuki hakutulosten

järjestämislle. Käyttöliittymän avulla pystyy kyllä ryhmittelemään hakutuloksia ja projisoimaan niitä erilaisille pinnoille, mutta selkeä nousevaan tai laskevaan järjestykseen laittaminen tietyn ominaisuuden arvon perusteella puuttui.

"Ryhmittely-näkymästä ei ollut apua, kun järjestämistapaan ei voinut vaikuttaa."

"Järjestely kohteen nimen mukaan olisi kiva..."

Esimerkiksi A-tehtävässä vanhimman kanteleen löytämien osoittautui työlääksi. Aikajanaa käyttäneet joutuivat varmuuden vuoksi vierittämään aikajanaa aikaan ennen Kristuksen syntymää aina kivikaudelle asti, jotta pystyivät varmistumaan siitä, että heidän löytämä kantele todella oli se vanhin kantele. Osa koehenkilöistä tarkisti asian Aika-näkymää käyttäen, ja näkivät sieltä vuosisadat-kategorian avattuaan, että 1700-luvulla näytti olevan vain yksi kantele, eikä sitä ennen yhtään. Myös ryhmittely-ominaisuutta käytettiin, ja kanteleet ryhmiteltiin vuosisatojen mukaan. Kaikki nämä tavat vaikuttivat kuitenkin vaivalloisilta yksinkertaisen nousevan tai laskevan numeron järjestykseen tai aakkojärjestykseen verrattuna.

#### 9.1.4 Hakusanat

Jokaista tehtävätyyppiä pyrittiin ratkomaan tekstihakua käyttäen. Tehtävissä A, D ja E jokainen koehenkilöistä käytti vapaatekstihakua tehtävän suorittamiseen. Museoiden kokoelmien vertailuun keskittyneessä C-tehtävässä puolestaan vain neljä henkilöä kahdestatoista käytti tekstihakua tehtävänsuorituksessa. B-tehtävässä vastaava suhde oli 9/12 ja F-tehtävässä 10/12. Tehtävänsuorituksessa useamman kerran käytetyt hakusanat on listattu tehtäväkohtaisesti taulukossa 9.2. Taulukosta käy myös ilmi erilaisten hakusanojen kokonaismäärä tehtävää kohden.

Tehtävissä A, B ja C erilaisten hakusanojen määrä pysyi pienenä. A-tehtävässä, jossa piti tutkia kanteleen kehitystä, kaikki koehenkilöt käyttivät sanaa kantele hakusanana, ja erilaisten hakusanojen määrä oli pienin. C-tehtävässä vapaatekstihakuja tehtiin ylipäättänsä vähän. Puhdas moninäkömahaku hakutapana dominoi tässä tehtävässä, kuten kuvasta 9.3 nähtiin. Hakutehtävässä vertailtavat museot löytyivät suoraan kokoelma-näkymästä, ja vertailua oli helpompi tehdä valitsemalla kategoriat suoraan. B-tehtävässä, jossa etsittiin I.K. Inhan valokuvia, hakusanat pyörivät pitkälti valokuvaajan nimen ja ammatin ympärillä. Kirjoitusasu vaihteli kuitenkin. Osa kirjoitti ainoastaan Inha ja toiset koko nimen I.K. Inha niin kuin se tehtävänannossa oli annettu.

Suurin hajonta käytetyissä hakusanoissa oli tehtävässä E. Tehtävänsuorituksissa kokeiltiin neljäkymmentäkolmea erilaista hakusanaa. Hakusana viljely oli selvästi suosituin, ja



sitä kokeilivat kaikki koehenkilöt yhtä lukuun ottamatta. Koska hakusana ei tuottanut tulosta, oli käyttäjien keksittävä viljelyyn jollain tavalla liittyviä hakusanoja, joilla sopivia kuvia olisi mahdollista löytää. Käyttäjien perehtyneisyys maanviljelyyn vaihteli selvästi, ja erilaisten hakusanojen määrä kasvoi myös siitä syystä. D-tehtävässä puolestaan hakusanojen suuri määrä johtui lähinnä haettavien kohteiden ja tehtävänannossa annettujen nimien suuresta lukumäärästä. Hakusanat pyörivät kuitenkin samoissa asioissa, mutta kirjoitusasu ja hakusanojen yhdistelmät vaihtelivat. Tehtävässä F noin puolet koehenkilöistä tiesi alusta alkaen esineen olevan keritsin. Henkilöt, jotka eivät tunnistaneet esinettä keritsimiksi, kokeilivat hyvin erilaisia hakusanoja (24 erilaista), joista saksit ja työkalu olivat käytetyimpiä.

**Taulukko 9.2. Tehtävissä käytetyt hakusanat, joita käytettiin useammin kuin kerran, ja erilaisten hakusanojen kokonaismäärä**

Tehtävä	Käyttäjien määrä: hakusana	Erilaisia kaikkiaan
A	12: kantele 2: kantele AND vanhin	4
B	7: Inha 3: I.K. Inha	7
C	2: Ateneum, nykytaiteen museo	8
D	7: Aino-taru 5: sammon taonta 4: gallen-kallela 3: Kalevala, aino 2: lemminkäinen tuonelassa, Akseli, lemminkäinen, sammon, Akseli Gallen-Kallela	26
E	11: viljely 7: maanviljely 4: työkalu 3: pelto, vilja 2: luomuviljely, aura, maan*, vilje*, käsityö, sadonkorjuu, maatalous, maatila	43
F	5: keritsin 4: keritsimet 3: lammas 2: saksit, työkalu	24

**9.1.5 Hakukategoriat ja näkymien käyttö**

Moninäkömähakua käytettiin myös jokaisen tehtävätyypin yhteydessä. C-tehtävässä moninäkömähakun käyttö oli suosituinta. Siinä kymmenen henkilöä kahdestatoista hyödynsivät moninäkömähakua tehtävän suorituksessa. Tehtävissä B, D, E ja F moninäkömähakua käyttäneiden määrät olivat seitsemän, kahdeksan, yhdeksän ja kahdeksan. A-tehtävässä moninäkömähakua käytti viisi henkilöä eli alle puolet koehenkilöistä.

Taulukossa 9.3 on eritelty kaikki käytetyt hakukategoriat. Tehtävissä E ja F kategorioiden valinnoissa oli eniten vaihtelua. E-tehtävän kohdalla tämä voi johtua siitä, että tulosten kriteerit olivat löyhät ja tekstihaku tuotti huonosti tuloksia. E-tehtävässä myös erilaisten

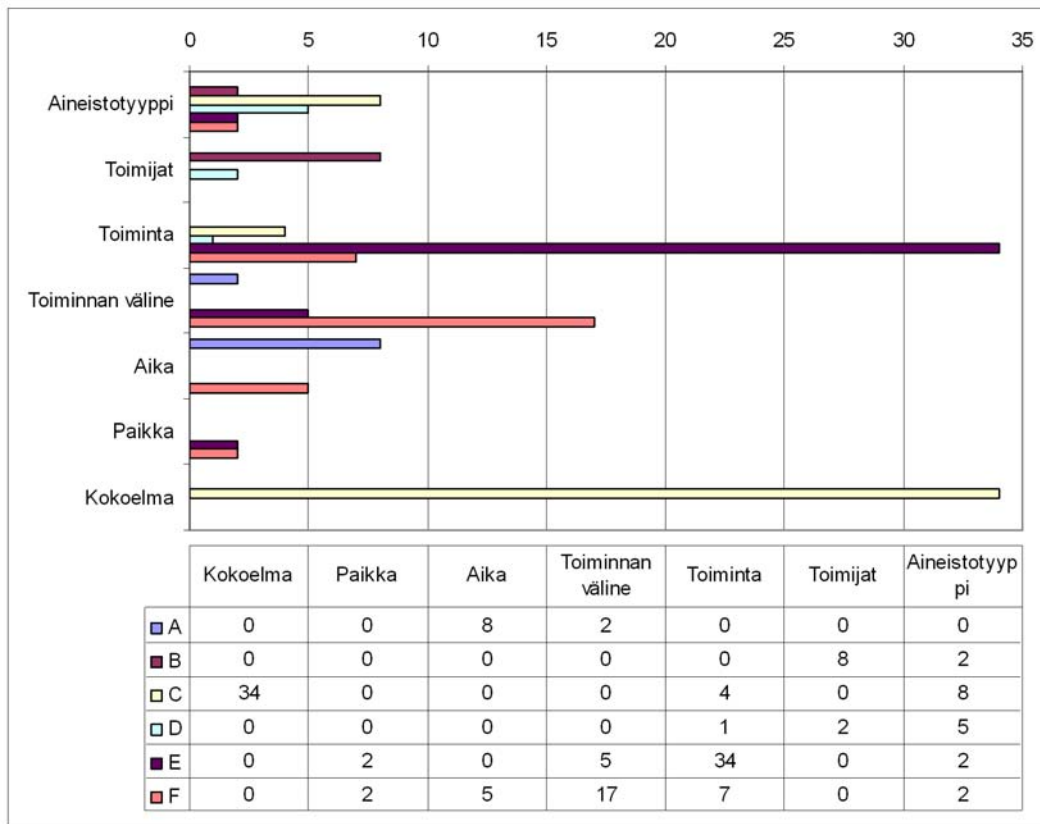
hakusanojen määrä oli suurin. Tehtävässä F annetut lähtötiedot olivat epämääräisimmät. Käyttäjille annettiin ainoastaan kuva tunnistettavasta esineestä. Tästä lähtöasetelmasta ratkaisua lähdettiin hakemaan erilaisin taktiikoin. Toiminnan väline -näköisestä etsittiin sopivia kategorioita materiaalien ja työvälineiden alta, kun taas toiminta-näköisestä valittiin työntekoon liittyviä kategorioita. Myös aikaa ja paikkaa rajaamalla pyrittiin pienentämään tulosjoukkoa, josta esinettä sitten selaamalla etsittiin.

### Taulukko 9.3. Tehtävissä käytetyt hakukategoriat

Tehtävä	Käyttäjien määrä: kategorian nimi
A	2 vuosisadat 2: masa:aikakaudet 1: konkreettinen eloton objekti, esineet
B	4: Inha, I. K., Henkilöt 2: valokuva
C	10: Ateneumin taidemuseo, Nykytaiteen museo Kiasma 3: Maalaus 1: hallinta, kognitio, mittaus, punnitus, Kuvataideteos, Veistos, valokuva
D	5: Maalaus 1: Akseli Gallen-Kallela, Kuvataideteos, Henkilöt
E	5: kylvä 4: viljely 3: luominen, istutus 2: kotieläintalous, maataloustuotanto, Suomi, maitotalous, työvälineet, esineet 1: VIIPURIN MUSEO:KÄSITYÖKALUT, alkutuotanto, kasvinviljely, Kuvataideteos, harvennus, kastelu, hallinta, kulutus ja käyttö, kasvatus, museoesine, ympäristö
F	3: työvälineet 2: Suomi, ns_30_maoryhmä:materiaali_metalli, museoesine, aine, konkreettinen eloton objekti, materiaalit 1: olemassaolo, 1800-luku, maitotalous, alkutuotanto, metalliseokset, omistus ja omistuksen siirtyminen, työ, lypsy, masa:aikakaudet, esineet, 1900-luku, fajanssi, metallit, teräs

Tarkasteltaessa näköisten käyttöä tehtävittäin (kuva 9.5) huomataan, että aineistotyyppi-näköistä käytettiin kaikista tasaisesti läpi tehtävien. Aineistotyyppiä rajattiin A-tehtävää lukuun ottamatta kaikissa tehtävissä. Myös toiminta-näköistä oli melko suosittu. Sitä hyödynnettiin neljässä tehtävässä. Toiminnan väline-näköistä puolestaan käytettiin

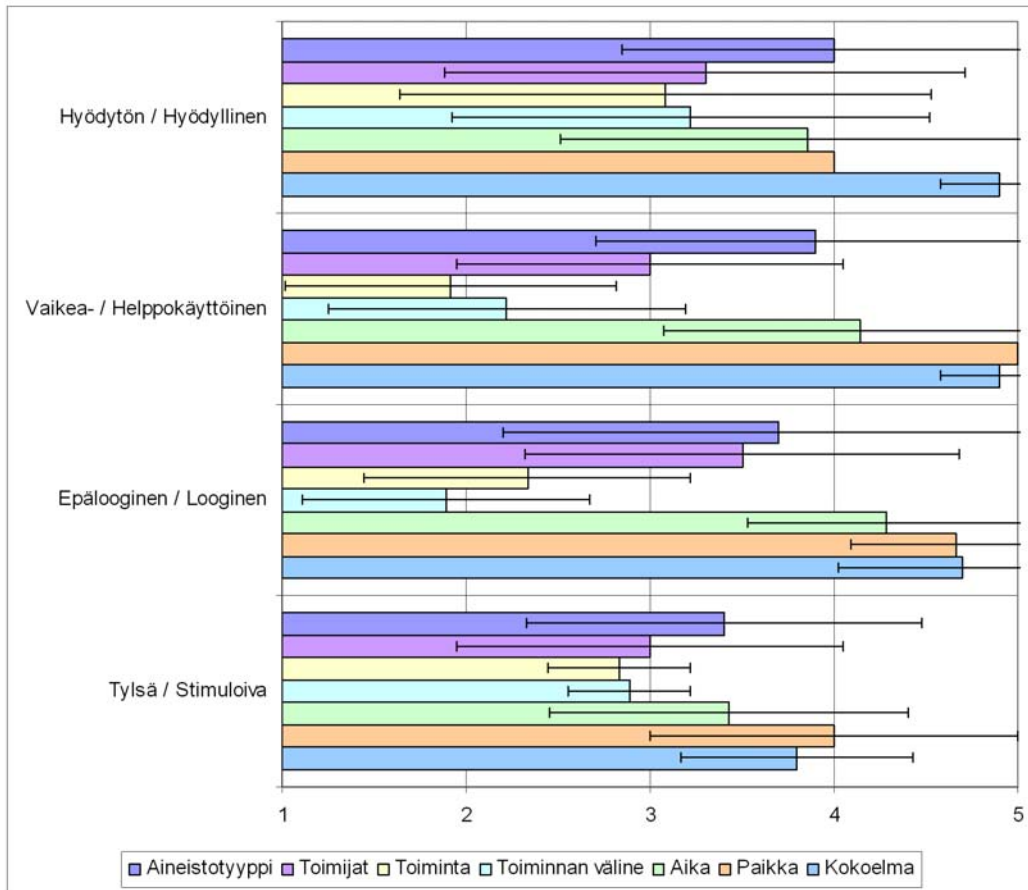
kolmessa tehtävässä. Toimijat-, paikka- ja aika-näkymiä käytettiin kutakin vain kahden tehtävän yhteydessä. Kokoelma-näkymää käytettiin ainoastaan C-tehtävässä, vaikka siinä sen kategorioita valittiin 34 kertaa.



**Kuva 9.5. Näkymien käyttö tehtäväkohtaisesti**

Loppukyselyssä testihenkilöiltä kysyttiin tarkemmin moninäkömahaun muodostavista komponenteista eli näkymistä. Jokaisesta näkymästä kysyttiin samat kysymykset, jotta saataisiin selville, miten erilaiset näkymät toimivat kokeen aikana. Kyselyn tulokset on annettu kuvassa 9.6.

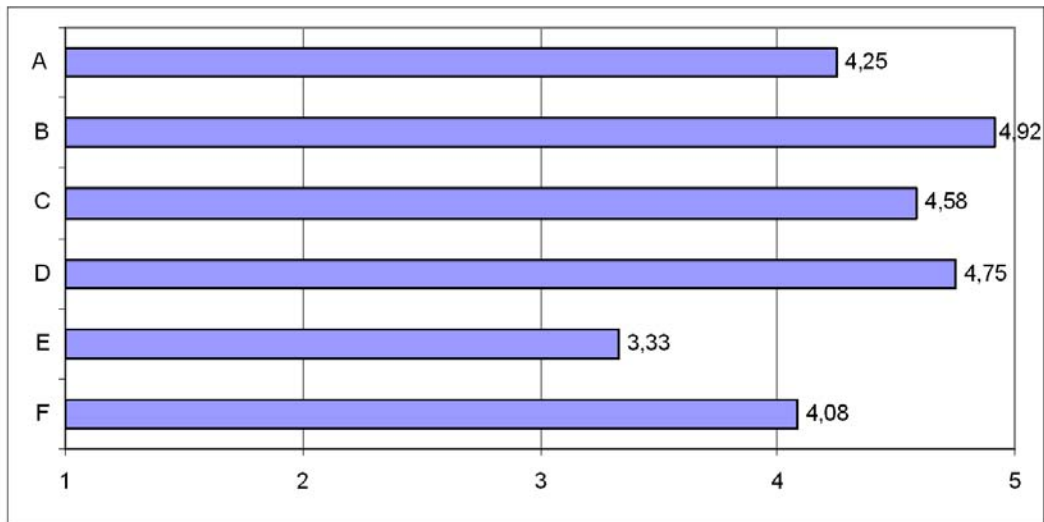
Ehdottomasti hyödyllisimmäksi näkymäksi testihenkilöt arvioivat C-tehtävässä käytetyn kokoelma-näkymän. Paikka-, aika- ja aineistotyyppinäkymät seurasivat kolmen ryhmässä perässä. Toimintaan suoraan viittaavat näkymät toimijat, toiminta ja toiminnan väline saivat huonoimmat arvosanat hyödyllisyydessä mitattuna. Helppokäyttöisyydessä paikka-näkymä sai keskiarvoksi viisi pistettä. Toisaalta vain kolme henkilöä kahdestatoista käytti ylipäättänsä paikka-näkymää kokeen aikana. Näkymiä toiminta ja toiminnan väline pidettiin melko vaikeakäyttöisenä. Samaan ryhmään näiden kanssa hyödyllisyyttä mitattaessa kuulunut toimijat-näkymä arvioitiin nyt selvästi kahta muuta näkymää helppokäyttöisemmäksi. Sama tilanne toistuu myös loogisuutta arvioitaessa. Tässä on kuitenkin huomioitava, että huonoimmat pisteet loogisuudessa on saanut toiminnan väline, joka hyödyllisyydessä ja helppokäyttöisyydessä mitattuna oli toiminta-näkymää edellä.



**Kuva 9.6. Näkymien arviointi ja arvioiden keskihajonta**

### 9.1.6 Hakujen onnistuminen ja tulokset

Jokaisen tehtäväsuorituksen jälkeen kysyin koehenkilöiltä, kuinka relevantteina he pitivät järjestelmän palauttamaa dokumentteja. Koehenkilöt pitivät E-tehtävää lukuun ottamatta hakutuloksia melko tai erittäin relevantteina. E-tehtävässä tulokset eivät kuitenkaan vastanneet koehenkilöiden odotuksia, ja relevanssi sai huomattavasti matalamman keskiarvon 3.33.



**Kuva 9.7. Hakujärjestelmän palauttamien dokumenttien relevanssi tehtäväkohtaisesti**

Kaikissa tehtävissä pyysin testihenkilöitä kirjaamaan paperille vastauksensa tehtävänannon kysymykseen. Vastaukset vaihtelivat esineiden tunnistenumeroista kokoelmien eroavaisuuksien analyysihin. Tehtävistä suoriutumisen mittaustapa eroaa siten tehtäväkohtaisesti. Tehtävistä suoriutumista pystyin mittaamaan sekä absoluuttisesti että annettujen vastausten perusteella. Absoluuttisesti mitattuna testihenkilöt joko antoivat vastauksen kysymykseen tai luovuttivat ilman vastauksen antamista.

Täydellisiä luovutuksia tapahtui vain tehtävissä E ja F. E-tehtävässä etsittiin luomuviljelystä kertovaan lehtijuttuun kuvitusta ja F-tehtävässä kuvan perusteella esineen nimeä. E-tehtävässä luovuttaneita oli kaksi, kun taas kuvan perusteella keritsimet löysi kaikki paitsi yksi henkilö.

Vastauksia tarkastelemalla myös muita tehtäväsuorituksia voisi pitää epäonnistuneena. Esimerkiksi tehtävissä A, F ja D on olemassa selkeät oikeat vastaukset kysymyksiin. Sen sijaan tehtävien B, C ja E kohdalla tehtävästä suoriutumista tulee tarkastella myös vastauksen sisällön perusteella. I. K. Inhan kuvia etsittäessä huomasin esimerkiksi, että monet valitsivat vain tekstihaun antamat viisi ensimmäistä kuvaa tarkistamatta, olivatko kuvat todellakin I. K. Inhan ottamia.

"Ehkä olisi vielä lopuksi voinut varmistaa, että Inha oli kuvaajana kaikissa, sillä Toimija-haku ilmeisesti osaisi hakea muitakin kuin tekijöitä."

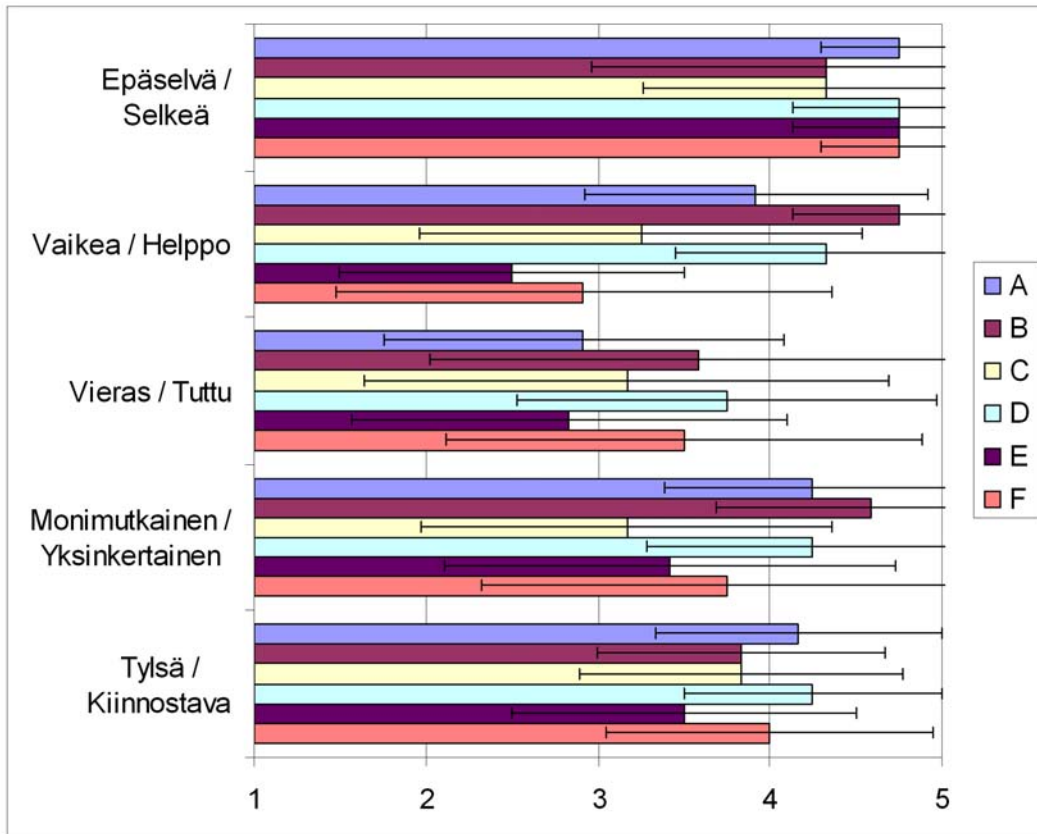
Luomuviljelykuvitusta lehtijuttuun hakevassa tehtävässä puolestaan monien kuvat olivat tehtävänannon rajauksen rajoilla. Ne saattoivat esimerkiksi vain olla vanhoja kuvia maatalan rakennuksesta. Alla esimerkissä 9.1 joitakin tehtävä E:n tulostuloksia, joista tuloksen laadun erot selvästi näkyvät. Tämän työn puitteissa en kuitenkaan ryhtynyt analysoimaan koehenkilöiden vastausten laatua tämän tarkemmalla tasolla.

## Esimerkki 9.1. E-tehtävän tulokuvasarjoja



## 9.2 Tehtävien arviointi

Tehtäviä A, D, E ja F pidettiin kaikista selkeimpinä, kuten kuvasta 9.8 näemme. B ja C-tehtäviäkin pidettiin melko selkeinä, joten suuria eroja tehtävääntojen selkeydessä ei ollut. Tehtävien vaikeutta arvioitaessa syntyivät tehtävääarvioiden suurimmat erot. E-tehtävää pidettiin selvästi vaikeimpana (2,5). Juuri tämä tehtävä oli se, jonka kaksi henkilöä jätti kesken. E oli myös se tehtävä, joka vei eniten aikaa, ja vaati suurimman määrän hakuja. Vaikeana tehtävänä pidettiin myös tehtävää F. Tehtävä olisi mahdollisesti arvioitu vieläkin vaikeammaksi, mikäli harvempi henkilö olisi tiennyt alusta alkaen, mikä esine kyseessä oli. C-tehtävä on kallistunut helpon puolelle, mutta kuitenkin arvioitu selvästi A, B ja D-tehtäviä vaikeammaksi. Ehdottomasti helpoimpana tehtävänä arvolla 4,75 pidettiin B-tehtävää. Tämä arvio on myös linjassa tehtävien vaatiman ajan ja hakumäärien kanssa, jotka B-tehtävälle olivat pienimmät.



**Kuva 9.8. Tehtävien arviointi**

Vieras/Tuttu-adjektiivipariin vastaaminen oli ilmeisen vaikeaa, sillä kaikkien tehtävien arvioiden keskiarvot jäivät kolmosen tuntumaan. Myös hajonta oli vastauksissa kysymyssarjan suurinta.

Museokokoelmien analysointiin rakennettua C-tehtävää pidettiin odotetusti monimutkaisimpana tehtävänä. Arvo jäi kuitenkin niukasti yksinkertaisen puolelle (3,17), joten kovin monimutkaisena tätäkään tehtävää ei pidetty. A, B ja D -tehtäviä pidettiin myös yksinkertaisina, ja helppouden sekä yksinkertaisuuden välillä oli selvä yhteys. Positiivista tehtävääarvioinneissa oli myös se, että tehtäviä pidettiin kauttaaltaan melko kiinnostavina.

"Tehtävät olivat mukavia ja sopivan helppoja, helpottuivat loppua kohden."

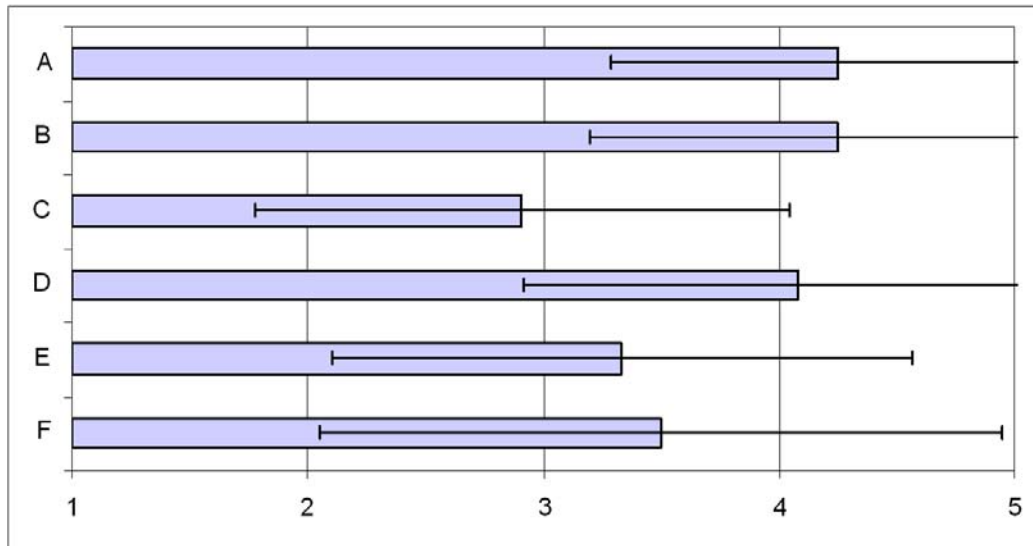
"Mukava koe, mielenkiintoinen."

Myös vapaamuotoisissa loppuhaastatteluissa kiiteltiin tehtävien, aineiston ja järjestelmän kiinnostavuutta sekä mielekkyyttä.

Kysymykseen, tiesivätkö testihenkilöt tarkalleen minkälaista tietoa heidän tulisi etsiä, ainoastaan tehtävässä C vastaukset kallistuivat epätietoisuuden puolelle. Tehtävissä A, B ja

D koehenkilöillä oli melko hyvä kuva siitä minkälaista tietoa he etsivät. Tehtävissä E ja F koehenkilöiden vastaukset kallistuivat vain niukasti väittämän puolelle.

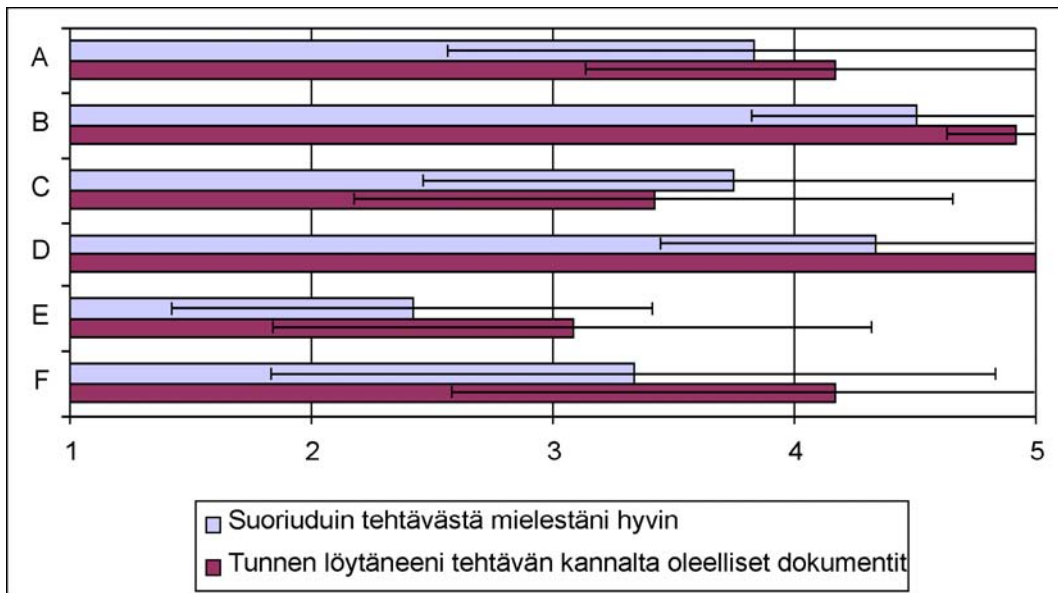
"Tiesin mitä etsiä, mutten keksinyt keinoa löytää sitä helposti."



**Kuva 9.9. Tiesin tarkalleen minkälaista tietoa etsiä**

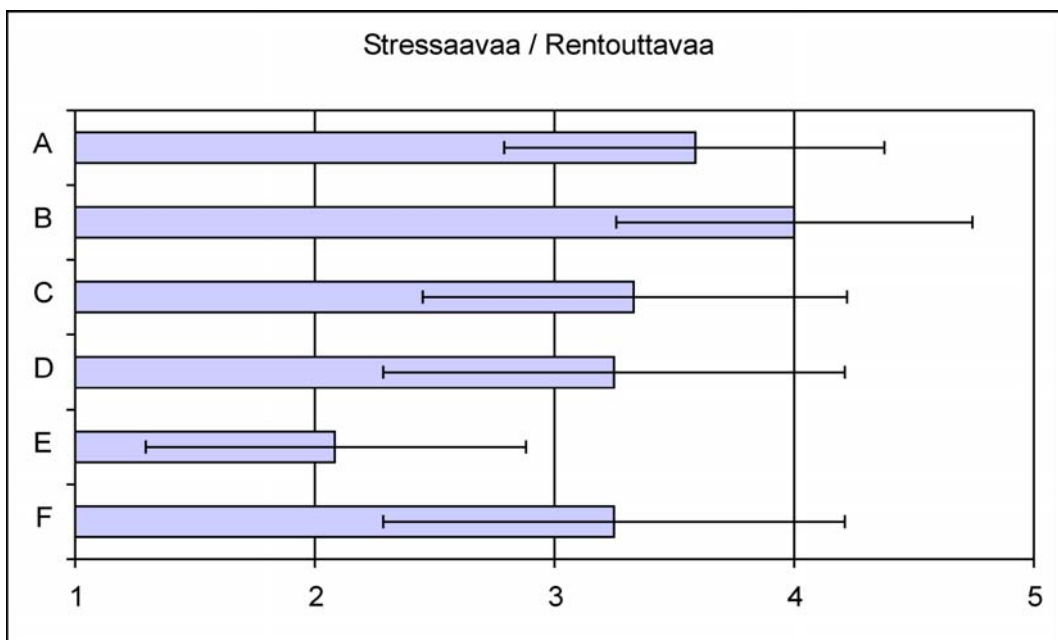
B- ja D-tehtävissä koehenkilöt olivat erittäin varmoja löytäneensä kaikki tehtävän kannalta oleelliset dokumentit. Tämä käy ilmi kuvasta 9.10. Näissä tehtävissä myös tyytyväisyys omaan suoritukseen oli korkeinta. C-tehtävä on ainoa, jossa arvio oleellisten dokumenttien löytämisestä on saanut huonommat pisteet kuin arvio omasta suoriutumisesta. Tämä johtuu todennäköisesti siitä, että tehtävässä ei pyydetty etsimään mitään tiettyä dokumenttia tai dokumenttityyppiä, vaan analysoimaan yleisemmin kahden museon eroja. Koehenkilöt olivat kaikista pettyneimpiä omaan suoritukseen E-tehtävässä. Tämä on linjassa kaikkien muiden mittareiden kanssa, jotka osoittavat E-tehtävän olleen vaikea.





**Kuva 9.10. Koehenkilöiden omat arviot suoriutumisestaan**

Lisäksi pyysin testihenkilöitä arvioimaan tehtävien/hakujen stressaavuutta ja rentouttavuutta. E-tehtävä, joka vaati eniten hakuja ja vei eniten aikaa pidettiin kaikista stressaavimpana. Eli mitä vaikeampi oli löytää etsimäänsä helposti ja nopeasti, sitä stressaavampana hakua pidettiin. Toisaalta C-tehtäväkin vei paljon aikaa, mutta sitä ei pidetty niin stressaavana. C-tehtävässä moninäkömähakua käytettiin suhteellisesti eniten.

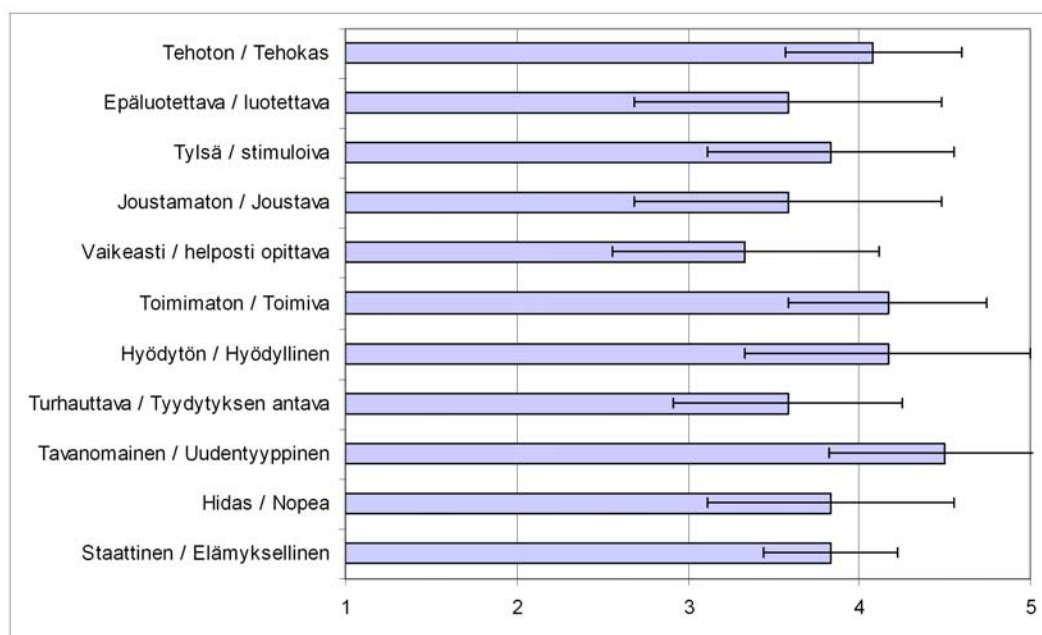


**Kuva 9.11. Hakuprosessin stressaavuus tehtäväkohtaisesti**

### 9.3 Tiedonhakujärjestelmän arviointi

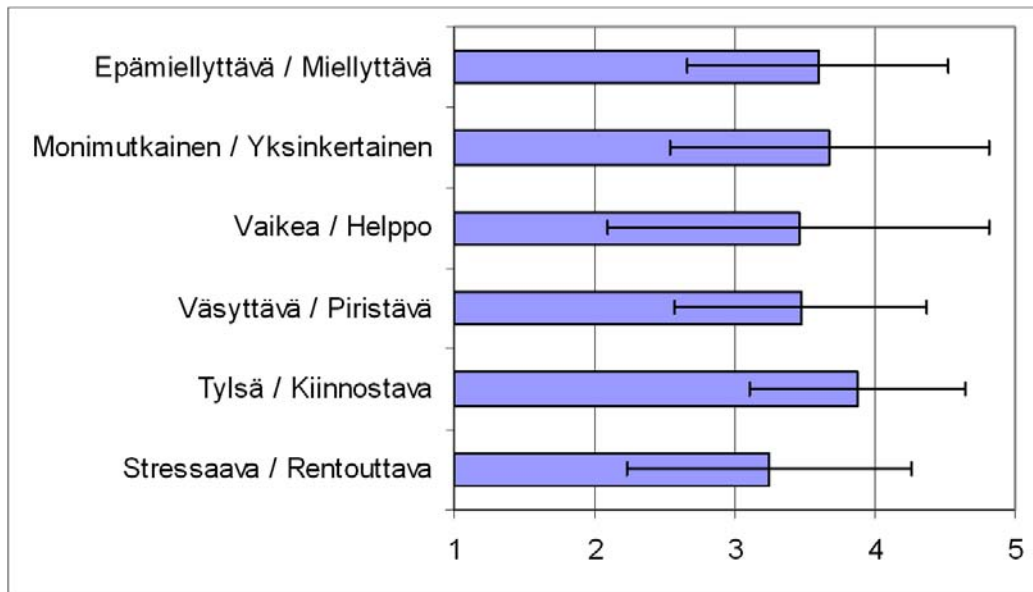
Loppukyselyssä järjestelmästä kokonaisuutena annettu palaute oli positiivista. Testihenkilöiden mielipiteitä järjestelmästä selvitettiin yhdentoista semanttisen differentiaalilin avulla. Kuvassa 9.12 ovat esillä testihenkilöiden arviot järjestelmästä.

Hakujärjestelmän suorituskyky oli selvästi kohdallaan. Järjestelmää pidettiin tehokkaana (4,08), toimivanan (4,17), hyödyllisenä (4,17) ja nopeana (3,83). Huonoimmat pisteet tulivat opittavuudesta (3,33). Moninäkömähakua pidettiin hankalasti opittavana, ja moni sanoikin kokeen aikana, että he tarvitsisivat enemmän aikaa tutustuakseen näkyymiin. Hakujärjestelmää pidettiin myös hyvin uudentyypisenä (4,5).



**Kuva 9.12. Hakujärjestelmän arviointi**

Edellä esitetyt arviot annettiin loppukyselyssä. Myös tehtävien jälkeen pyydettiin arvioimaan hakujärjestelmää tai tarkemmin hakuprosessia järjestelmällä. Yhteenveto näistä arvioista on esitetty kuvassa 9.13. Kuvasta näemme, että jotkut testihenkilöt pitivät järjestelmällä hakemista stressaavana ja vaikeana. Toisaalta näilläkin mittareilla kaikki keskiarvot nousivat yli kolmen.



**Kuva 9.13. Järjestelmällä hakeminen**

Tiedon esillepanoa pidettiin yleisesti selkeänä. Kohdesivu arvioitiin selkeäksi ja informatiiviseksi. Palautetta tuli siitä, että kohdesivulta poimitut tiedon palaset eivät olleet suoraan valittavissa hakuehdoiksi ja että kopioitaessa niitä ei aina löytynyt vapaatekstihauulla tai kategorioista, vaikka näin olisi olettanut.

"En löytänyt hakusanoilla 'maatalous', 'maanviljely', 'sadonkorjuu' yms yhtään kohdetta, vaikka nämä kyseiset asiasanat löytyivät valitsemani kolmen esineen asiasanoista."

"Kohdesivussa voisi olla esillä kategoria, johon esine kuuluu, niin olisi helpompi löytää muita samaan kategoriaan kuuluvia kohteita."

Erityisesti E-tehtävässä, jossa etsittiin luomuviljelyyn liittyviä kuvia, kaivattiin suoraa valintaa kohdesivuilta. Vapaatekstihakuun oli monella vaikea keksiä sopivia hakusanoja, ja toimintanäkymästä maanviljelyyn liittyviä kategorioita oli vaikea löytää. Löydettyään sopivan kuvan he olisivat halunneet jatkaa etsimistä sen perusteella kuvan metatiedoista löytyvillä avainsanoilla ja luokilla. Tätä mahdollisuutta käyttöliittymä ei kuitenkaan tarjonnut.

Itselleni hakukäyttöliittymän toteuttajana tuli yllätyksenä, kuinka moni yritti käyttää järjestelmää useaa selainikkunaa yhtä aikaa käyttäen. Tämä käyttötapa ei kuitenkaan toiminut, sillä järjestelmä tallensi haun tilan sessioon, joka oli selaimelle yhteinen. Näin ollen hakutilanne päivittyi jokaiseen avoimeen selaimen ikkunaan uuden haun jälkeen. Session tallennuspaikka tulisi siirtää url-tekstiin, jotta se tukisi tätä käyttötapaa. Session tallentamisesta ei voine täysin luopua, sillä siihen nojaavia ominaisuuksia toivottiin jopa

lisää. Session kuljettaminen urlissa tekisi kuitenkin usean samanaikaisen haun tekemisen mahdolliseksi ja tukisi näin olleen tätä käyttöliittymän tehokäyttötapaa.

Useat testihenkilöistä myös ikään kuin tallensivat välituloksia toiseen selainikkunaan. Toimintatapa oli kiinnostava huomioida, sillä en ollut tätäkään osannut alun perin ottaa huomioon. Tätä käyttötapaa järjestelmä kuitenkin tuki hyvin. Edellä esitettyjä usean selainikkunan käyttötapoja eräs koehenkilöistä luonnehti vertailevaksi hauksi. Vertailevan haun tukea tulisi kehittää järjestelmässä edelleen.

## 9.4 Tulosten tarkastelu

Tulosten perusteella näyttää siltä, että moninäkömähaku ja vapaatekstihaku soveltuvat erilaisten tehtävien suorittamiseen. Moninäkömähakua hyödynnettiin tehtävissä, joissa lähtötiedot eivät olleet tarkkoja. Tällaisia tehtäviä olivat F-tehtävä, jossa kuvan perusteella etsittiin esinettä, museoiden vertailuun perustuva C-tehtävä ja lehtijutun kuvituksen hakemiseen kohdistunut E-tehtävä. Tehtävä F osoitti, että mikäli lähtötiedot ovat kovin puutteelliset, voi moninäkömähaku tarjota tukea hakemiselle tarjoamalla valittavaksi vartenotettavia luokkia. Tehtävissä, joissa tehtävänannossa annettiin tarkkaa metatietoa, oli tekstihaku hyödyllisempi. Esimerkiksi D-tehtävässä, jossa lähtötiedot ovat hyvin tarkat, oli moninäkömähaku hyödytön. Vapaatekstihaulla etsittäessä oikeat taulut löytyivät nimien perusteella, eikä niitä tarvinnut lähteä etsimään näkymiä rajaamalla. Gallen-Kallelan taulut löytyivät D-tehtävässä helposti niiden nimillä, I.K. Inhan kuvat B-tehtävässä kuvaajan nimellä ja A-tehtävässä kanteleet yksinkertaisesti hakusanalla kantele. Kokeen tehtävät oli positioitu luvussa 2.1 esittelemäni Ingwersenin ja Järvelinin tiedontarpeen luokituksen mukaisesti. Tehtävistä B ja D oli asemoitu tiedonhakijan nykyisen tiedon tyyppin osalta tunnetun metadatan tehtäviksi, kun taas E ja F tunnettujen sisällöllisten attribuuttien tehtäviksi. Nykyisen tiedon tyyppi näyttää näin ollen vahvasti vaikuttavan hakutavan valintaan. Tekstihaku ja moninäkömähaku ovat näin ollen enemmän toisiaan täydentäviä hakutapoja kuin toisensa poissulkevia vaihtoehtoja.

Tarkempaa tietoa vaihtoehtoisten hakutapojen eroista suorituskyvyssä saataisiin suorittamalla sama koe vertailuryhmillä, jossa toinen ryhmä suorittaisi kokeen vapaatekstihaulla ja toinen moninäkömähauulla. Tässä kokeessa halusin kuitenkin evaluoida hakujärjestelmää kokonaisuutena, kaikki ominaisuudet mukaan luettuna.

Käyttäjätiestien aikana kritisoitiin erityisesti moninäkömähauksen käytön vaikeutta. Näkymien sisältöä pidettiin vaikeasti ymmärrettävänä, ja monet ilmaisivat testin loppuhaastattelussa, etteivät he vieläkaan ymmärtäneet toimintakeskeisten näkymien (toimija, toiminnan väline, toiminta) sisältöä. Näihin kommentteihin näen kaksi syytä. Ensimmäinen syy on puhtaasti käyttöliittymäongelma. Näkymien esittäminen avattavana ja suljettavana puuna ei välttämättä ole paras ratkaisu. Puun solmujen avaaminen ja sulkeminen vie aikaa, ja samalla

käyttäjältä saattaa kadota tieto missä kohdin puuta hän on menossa, mikäli puu muuttuu huomattavasti avausten ja sulkemisien johdosta. Valittavien kategorioiden etsimiseen tulisi näin ollen tarjota jokin parempi ratkaisu. Helposti toteutettava apu ongelmaan olisi näkymäkohtaisen kategoriahakukentän lisääminen, jolla pystyisi vapaatekstihaulla etsimään näkymästä kategorioita. Tällä tavalla voitaisiin kiertää puun käyttäminen solmuja avaamalla ja sulkemalla. Jotkut testihenkilöistä keksivätkin tämän tavan jo kokeen aikana. He avasivat koko näkymäpuun, ja käyttivät selaimen hakutoimintoa sopivien kategorioiden etsimiseen avatusta näkymästä.

Toinen näkymiin liittyvä ongelma johtui näkymien projisointitavasta, jonka esittelin luvussa 7.2.1. KulttuuriSammon näkymät on projisoitu ontologioista antamalla juurikäsité, ja seuraamalla juuresta jotain kaarta lehtiin. Näin ollen näkymäpuut muodostuvat sellaisiksi, että näkymän juurina ovat ontologioiden ylimmät käsitteet, jotka voivat olla hyvinkin yleisellä tai abstraktilla tasolla. Haun kannalta kiinnostavat kategoriat löytyvät näkymäpuista monesti vasta syvemmältä ontologian rakenteista. Abstraktit yläkäsitteet on kuitenkin vaikea yhdistää näihin haun kannalta oleellisiin käsitteisiin, ja näin ollen käyttökelpoiset kategoriat jäivät usein löytymättä. Toiminta-näkymässä ylimmän tason käsitteitä ovat esimerkiksi kognitio, olemassaolo ja mentaalinen toiminta. On vaikea kuvitella, mitä kaikkea näiden käsitteiden alta voisi löytyä. Käyttäjät pitivätkin toiminta- ja toiminnan väline -näkyviä vaikeakäyttöisinä ja epäloogisina.

Näkymiin tulisi myös lisätä mahdollisuus ryhmitellä tulokset näkymän juurien mukaan. Nyt esimerkiksi kokoelma-näkymän mukaan ryhmittely oli mahdotonta, mutta monet kaipaivat mahdollisuutta esimerkiksi C-tehtävän yhteydessä, jossa vertailtiin kahden kokoelman sisältöjä.

Vaikka moninäkymähaun toimintakeskeisyys ei kaikille käyttäjille avautunut, olivat toiminta ja toiminnan väline aineistotyypin jälkeen käytetyimmät näkymät. Tämä viittaisi mielestäni siihen, että hakeminen toiminnan kautta on ihmisille intuitiivista, ja nämä näkymät pitävät sisällään huomattava potentiaalin. Tässä tapauksessa käyttöliittymän toteutus ei vain ollut riittävän hyvä. Samalla kannattaa muistaa, että myös aika, paikka ja toimijat olivat toimintakeskeisiä, ja ne saivat huomattavasti parempia arvioita.

Useilla testihenkilöillä oli ongelmia hahmottaa vuorovaikutteisen hakujärjestelmän toimintaa. Loppuhaastattelun yhteydessä nousikin esiin, että käyttäjillä oli monesti kokeen aikana vaikea muistaa, että aikaisemmin tehdyt kategoriavalinnat ja tekstihaut jäivät voimaan. Käyttäjä saattoi esimerkiksi tehdä pitkän aikaa tekstihakuja niin, että hänellä oli tiukasti rajaavia kategorioita unohtunut valituksi. Tämä ei välttämättä ole vakava käytettävyysongelma, koska huomasin selvästi, että kokeen edetessä tämän tyyppiset ongelmat vähenivät oppimisen ansiosta. Kysymys on lähinnä uudentyyppisen järjestelmän käytön opettelusta. Jossain vaiheessa olin suunnitellut, että tekstihakukentän alapuolella

olisi valintaruutu, jolla pystyisi kontrolloimaan haun luonnetta kertahaun ja iteratiiviseen haun välillä. Tämä kuitenkin tekisi hakuprosessista entistä monimutkaisemman, ja siksi jätin ominaisuuden toteuttamatta.

Kysyttäessä testihenkilöiltä miksi he eivät olleet käyttäneet jotain tiettyä näkymää, käyttäjät kritisoivat käyttöliittymän komponenttien asettelua. Testihenkilöiden mielestä monet näkymät olivat niin pitkällä alhaalla sivua, ettei niitä edes huomannut, tai että alas asti skrollaaminen oli työlästä. Hakukomponenttien asettelua tulisi harkita uudestaan. Mikäli edellä esittämiini näkymien puuesitystavan ongelmiin löydetään ratkaisu, joka mahdollistaisi näkymien esittämisen pienemmässä tilassa, voisi näkymät järjestää näytön halki vaakasuorassa, nykyisen pystysuoran järjestyksen sijaan.

Karttaprojektiota käytettiin hyvin vähän kokeen aikana johtuen tehtävänasettelun epäonnistumisesta kartan käyttöön ohjauksessa. Paikkakuntaa ja paikkaa kyllä kysyttiin useassa tehtävässä, mutta aina niin, että oli helpompaa etsiä esine tai kohde jollain muulla tavalla ja katsoa paikkatiedot löydetyn kohteen metatiedoista. Näin ollen kartan käytettävyydestä en saanut testeissä kuvaa. Aikajanan käytössä esiintyneet ongelmat tilan säilyttämisestä pätevät kuitenkin melko varmasti myös karttaan. Vaikka ohjaus kartan ja aikajanan käyttöön ei täysin onnistunut, voi käyttömääristä päätellä, että erilaiset projektiotavat ovat enemmänkin hauskoja lisäominaisuuksia, joita käytetään harvemmin kuin perinteistä hakutulosten listausta.

Yleisesti ottaen järjestelmä osoittautui testien aikana olevan melko vakaa ja suorituskykyinen. Ainoastaan yhden kerran jouduin käynnistämään järjestelmää pyörittäneen sovelluspalvelimen uudestaan järjestelmän mentyä jumiin. Osa käyttäjistä onnistui aiheuttamaan virheilmoituksen, joka liittyi jollain tavalla tekstihakumoottoriin. Testien perusteella en päässyt kiinni virheilmoitusten syyhyn, sillä virhe näytti ilmaantuvan aina erilaisissa tilanteissa. Virheistä selvittiin kuitenkin aina nollaamalla sessio ja aloittamalla haku uudestaan. Tämä luonnollisesti vaikutti hieman testin tuloksiin hakujen määrän ja suoritusajan osalta. Ongelma esiintyi sen verran harvoin, ettei sillä mielestäni ollut suurta merkitystä tuloksiin. Sen sijaan arvioihin järjestelmän suorituskyvystä se olisi saattanut vaikuttaa, mutta kyselyssä saadut pisteet olivat sen verran hyviä, että en usko ajoittaisten virheilmoitusten häirinneen käyttäjiä kovinkaan paljon. Positiivisiin arvioihin varmaan vaikutti se, että käyttäjät tiesivät järjestelmän olevan vielä kehitysvaiheessa, ja antoivat pienet ongelmat anteeksi. Todellisuudessa en usko ihmisten palaavan hakukoneisiin, jotka kaatuilevat. Järjestelmän vakautta on siis parannettava ennen järjestelmän julkaisua WWW:ssä.

#### 9.4.1 Vastaukset tutkimuskysymyksiin

Ensimmäinen tutkimuskysymykseni oli, *mahdollistaako toimintakuvauksiin perustuvan sisällönkuvailun tietomalli heterogeenisen aineiston yhdistämisen?* Projektin aikana tutkija

Ruotsalon kanssa kehittämämme tietomalli osoittautui sovelluskehityksessä ja aineistojen rikastamisprosessissa kuvausvoimaiseksi malliksi, joka soveltuu kaikkien portaaliin mukaan otettujen aineistotyyppien kuvailuun. Käyttäjätiestien hakutehtävät muotoilin niin, että tehtävissä etsittiin vaihtelevasti valokuvia, maalauksia ja museoesineitä. Joissain tehtävissä haettavaa aineistotyyppiä ei oltu lyöty lukkoon ja käyttäjä sai valita vapaasti parhaiten sopivat hakutulokset. Hakujärjestelmä oli toimiva kaikkien aineistotyyppien kanssa, vaikkakin alkuperäisten metatietojen tasojen vaihtelusta johtuen aineistojen löydettävyyttä vaihteli. Toimintakeskeisiä hakutapoja käytettiin ja hakutuloksissa oli niin valokuvia, tauluja kuin myös esineitä. Näiden tulosten valossa voidaan sanoa, että tietomalli mahdollistaa heterogeenisen aineiston yhdistämisen.

Toinen tutkimuskysymykseni oli, *voiko toimintapohjaisen tietomallin päälle rakentaa käyttäjälle hyödyllisiä haku- ja visualisointijärjestelmiä?* Hakujärjestelmän toimintakeskeinen semanttinen moninäkömähaku on rakennettu täysin toimintakuvauksiin perustuvan sisällönkuvailun tietomallin päälle. Myös projektiot aikajanelle ja kartalle tehtiin tietomalliin pohjautuen. Käyttäjätesteissä käyttäjät pitivät järjestelmää toimivana ja hakutuloksia relevantteina hakuihin nähden. Hakeminen oli myös tuloksellista. Vain kolme tehtäväsuoritusta jäi kesken. Erityisesti kartta- ja aikajanaprojektioita kehuttiin. Vaikka moninäkömähakua kritisoitiin, käytettiin sitä tuloksellisesti monien tehtäväsuoritusten yhteydessä. Kritiikki kohdistui lähinnä käyttöliittymän käytettävyysoongelmiin ja näkymien sekavuuteen, jotka eivät johtuneet toimintakeskeisestä mallista. Näin ollen voin todeta, että käyttäjillensä hyödyllisten haku- ja visualisointijärjestelmien rakentaminen tietomallin päälle on mahdollista ja mielekäästä.

#### 9.4.2 Tulosten arviointi ja tutkimuksen luotettavuus

Konstruktiivinen tutkimusote sisältää tavoitteen arvioida konstruktion käytännöllistä soveltuvuutta toteuttamalla konstruktion ja heijastamalla empiirisiä havaintoja takaisin teoriaan (Lukka 2001). Diplomityössäni olen kertonut, kuinka kehitin heterogeenisten aineistojen yhdistämisen ongelman edessä ratkaisuksi uuden tietomallin ja tietomallia hyödyntävän hakukäyttöliittymän tietomallin mukaisesti kuvailun aineiston hakemiseksi. Tekemäni ratkaisut nojaavat vahvasti tutkimuslaboratoriomme aikaisempaan työhön semanttisten portaalien, moninäkömähäun ja sisällönkuvailun osalta ja linkittyvät teoriaan tietämyksen esittämisestä, logiikasta ja kielitieteistä. Empiirisiä havaintoja keräsin tutkimuskysymyksiäni valossa sekä portaalin ohjelmointi- ja kehitysvaiheessa että käyttäjätestein.

Konstruktiivinen tutkimus on kokeellista tutkimusta. Kokeellisen tutkimuksen perustana on saada aikaiseksi tapahtumia olosuhteissa, joita voidaan hallita ja joissa tapahtumista voidaan tehdä systemaattisia havaintoja (Järvenpää & Kosonen 1996, s. 15). Selittäviä muuttujia ovat ne, jonka vaikutuksia tutkitaan, ja selitettäviä muuttujia ne, joihin selittävän muuttujan

vaikutuksia tutkitaan (Mt., s. 15). Tärkeimpiä selittäviä muuttujia kokeessani olivat eri tavoin positioidut tehtävänannot.

Portaalin evaluoinnissa käytin tiedonhaun käyttäjäkeskeistä mallia. Käyttäjäkeskeisessä mallissa pyritään realistisiin tuloksiin pohjaamalla evaluointi potentiaalisten käyttäjien oikeisiin tiedon tarpeisiin ja hakutapoihin (Borlund 2003). Tiedon keruuseen käytin tietokoneelle reaaliaikaisesti tallentuvaa lokia eli niin sanottua jäljityspolkua, kyselylomakkeita, vapaamuotoisia haastatteluja, havainnointia testin aikana sekä videon ja äänen tallennusta testitilanteesta. Menetelmät toimivat hyvin ja sain kerättyä kattavan aineiston. Ensimmäinen koehenkilötesti oli pilottitesti, mutta koska testi sujui ongelmitta, eikä muutoksia koejärjestelyihin videotallennuksen codecin lisäksi jouduttu tekemään, otin pilottitestin osaksi varsinaista koedataa.

Määrällistä aineistoa keräsin erityisesti jäljityspolulla. Jäljityspolkuun tallentuvat tiedot ovat kiistattomia, koska ne sisältävät ainoastaan asioita, joita käyttäjä on tehnyt koetilanteen aikana. Jäljityspolun avulla kerätyn datan analysoinnissa oli kuitenkin ongelmallista päättää esimerkiksi se, missä vetää raja uudelle haulle. Onko näkymäpuun avaaminen uusi haku, vai ainoastaan toimenpide uuteen hakuun valmistauduttaessa, kuten kohdesivulla tarjottavan tekstimuotoisen tiedon lukeminen. Päädyin käsittelemään jokaista näkymäpuun interaktiota hakuna, koska pidän sitä olennaisena osana moninäkömähakua. Jäljityspolkuun tallentuvat myös kaikki turhat haut, joita käyttäjät tekevät vahingossa tai joiden tuloksiin he eivät edes tutustu. Lokitietojen tarkempi analyysi ja siivoaminen olisi kuitenkin tämän työn puitteissa vienyt liian paljon aikaa.

Kyselylomakkeet loin Microsoftin Infopath -ohjelmalla<sup>1</sup> ja kyselylomakkeet täytettiin tietokoneella. Sähköisten lomakkeiden käyttö vähentää virheitä, kun tietojen siirtovaihe tippuu pois ja kyselyjen tulokset ovat siirrettävissä suoraan analysoinnissa käytettävään ohjelmaan. Pääosin kyselylomakkeet toimivat hyvin ja koehenkilöt osasivat vastata niissä esitettyihin kysymyksiin. Loppukyselyn "Hakujärjestelmällä on mahdollista" -kysymykset olivat kuitenkin epäonnistuneita. Testin aikana huomasin, että en itsekään kyennyt vastaamaan niihin. Näin ollen on mahdotonta sanoa, mikä on oikea tai väärä vastaus. Jätin tämän osan pois analyysistä. Tehtäväkohtaisessa kyselylomakkeessa tulosjoukon esitystapoihin liittyvät kysymykset olivat hankalia erityisesti oletuksena olleen esitystavan osalta. Lisäksi sen nimeäminen ryhmitellyksi sai monet tulkitsemaan kysymyksen niin, että se edellytti ryhmittely-toiminnon käyttämistä ja vastasivat, että he eivät olleet käyttäneet esitystapaa laisinkaan.

Tehtävänasettelu oli mielestäni pääosin onnistunut. Tehtävät heijastivat aineistotietokannan sisältöä ja portaalin suunniteltuja käyttökohteita. Tehtäviä kommentoitiin positiiviseen

---

<sup>1</sup><http://office.microsoft.com/infopath/>



sävyyn loppukyselyssä. Tehtävää F voidaan kuitenkin pitää epäonnistuneena, koska koehenkilöiden esitiedot vaikuttivat niin suuresti lopputulokseen. Ne henkilöt, jotka tunnistivat esineen suoraan ja löysivät sen tekstihaulla, pitivät tehtävää erittäin helppona, kun taas ne, joilla ei ollut aavistustakaan esineen käyttötarkoituksesta, joutuivat tekemään hakuja kuvasta irtisaatavan tiedon perusteella. Tämä oli luonnollisesti erittäin paljon vaikeampaa ja osa jätti tehtävänsuorituksen kesken. Vaikka vastaava tilanne eritasoisista lähtötiedoista tosielämässä on jokapäiväinen, olisi testitilanteen suonut asettaa kaikki käyttäjät samalle linjalle. Myös B-tehtävän tuloksia tarkasteltaessa on huomioitava, että hakujen raportoitu määrä ja tehtäväsuorituksen kestot ovat suurempia kuin mitä niiden olisi tarvinnut olla. Syy tähän on se, että tehtävässä pyydettiin viiden kuvan hakuja. Osa tuskastui kuvien suuren määrän takia, eivätkä suorittaneet tehtävää tosissaan. Toiset taas jaksoivat etsiä vaihtelevia kuvia, eivätkä tyytyneet viiteen ensimmäiseen järjestelmän palauttamaan kuvaan. Pienempi määrä kuvia olisi tasoittanut tuloksia tehtävän osalta.

Yksi pahimmista virheistä järjestelmässä oli näkymäkategorioiden ennustavien hakutulosten näyttöjen arvojen virheet. Näiden lukujen tuli siis kertoa, kuinka monta hakutulosta tietyn kategorian valinta antaisi tulokseksi. Luvut eivät kuitenkaan päivittyneet oikealla tavalla vapaatekstihaakujen yhteydessä, ja monesti ne näyttivät muutenkin virheellisiä lukuja. Useasti huomasin esimerkiksi virheellisten 0-lukujen hämänneen käyttäjiä niin, että he eivät valinneet jotain oleellista kategoriaa hakuehtoonsa, koska olettivat sen johtavan tyhjiin hakutulokseen. Virheen syy oli se, että sain vapaatekstihaun liitettyä hakujärjestelmään hyvin myöhäisessä vaiheessa, enkä ehtinyt testata ominaisuutta riittävästi. On vaikea arvioida kuinka paljon tämä vaikutti käyttäjätietojen tuloksiin, mutta uskon että se heikensi käyttäjien arvioita moninäkömahausta ja näkymien käyttökelpoisuudesta.

Virhetekijöiden poistamiseksi kokeistani käytin vakiointimenettelyä. Pidin koetilanteen mahdollisimman samanlaisena kaikille koehenkilöille. Opastin käyttöliittymän käyttöön kohta kohdalta listatussa järjestyksessä. Annoin käyttäjien harjoitella itse ennen ensimmäistä tehtävää, ja kerroin heille samat asiat kokeesta ja sen taustoista. Kokeet suoritettiin kahdessa eri paikassa, mutta aina täsmälleen samalla laitteistolla ja samankaltaisissa häiriöttömissä tiloissa. Tehtävien suoritusjärjestyksen tasapainotin latin square -menetelmällä. Kaiken kaikkiaan uskon tulosten olevan luotettavia ja antavan tärkeää tietoa moninäkömahausta, toimintakeskeisestä tietomallista ja kokonaiskonstruktion toiminnasta.

## 10 Johtopäätökset

Tämän työn tavoitteena oli tutkia semanttisen webin tarjoamia mahdollisuuksia tiedonhaussa toteuttamalla kansallinen museoalan portaali, jossa eri museoista peräisin olevat heterogeeniset aineistot ovat haettavissa yhdellä käyttöliittymällä. Portaalien tarkoitus on pilotoida FinnONTO-projektissa kehitettyjä ontologioita, palveluja ja työkaluja. Työssäni kehitin sisällönkuvailumenetelmän ja siihen liittyviä haku- ja visualisointimenetelmiä. Valmista portaalia evaluoin käyttäjätestein saadakseni tarkempaa tietoa hakujärjestelmän toimivuudesta ja soveltuvuudesta käyttötarkoitukseen. Tässä luvussa esittelen tutkimuksen keskeisimmät johtopäätökset ja pohdin mahdollisia jatkotutkimuksen aiheita.

Toimintakeskeinen tietomalli osoittautui portaalien sovelluskehityksen aikana ja käyttäjätestien tulosten valossa toimivaksi tietomalliksi heterogeenisten aineistojen yhdistämisessä. Jalostettuani aineistoa sääntöpohjaisesti luvussa 7.1 selitetystä semanttisen rikastamisen prosessissa toimintakeskeiseen malliin, oli toimintakeskeisen mallin päällä toimiva hakujärjestelmä mahdollista toteuttaa hyvin generisesti. Luvun 7.2 kategoriaprojektiot ja dokumenttien ripustaminen kategorioihin sekä aineiston kartta- ja aikajanaprojisointi toimivat täysin toimintakuvausten ja temaattisten roolien varassa.

Käyttäjätetit puolestaan todistivat rakentamani hakujärjestelmän toimivaksi. Se tuki jokaista tehtävätyyppiä ja käyttäjät pitivät arvioissaan järjestelmää uudentyyppisenä, tehokkaana, toimivana ja hyödyllisenä. Vaikka käyttöliittymä tarjosi myös perinteisen vapaatekstihaun, käytettiin toimintakeskeistä semanttista moninäkömähakua jokaisen tehtävätyypin suorituksen yhteydessä. Käyttäjät olivat myös melko tyytyväisiä hakujärjestelmän palauttamiin dokumentteihin. Kartta- ja aikajanaprojektioita pidettiin hyvinä ja toimivina lisäominaisuuksina. Näin ollen tässä työssä konseptoinni toimintakeskeistä tietomallia voidaan pitää järjestelmätoteutuksen kannalta toimivana ratkaisuna heterogeenisten aineistojen yhdistämisessä ja saattamisessa haettavaksi.

Aineiston rikastamisprosessin puoliautomaattinen toimintakuvausten generointi osoittautui sen sijaan ongelmakohtaksi. Aineisto ei yksinkertaisesti ollut tarpeeksi rikasta tarjotakseen toivotun kaltaisen toimintakeskeisen semanttisen moninäkömähäun. Kiinnostavien kategorioiden määrä jäi monessa näkymässä alhaiseksi ja kohteet ripustuivat pitkälti samoihin kategorioihin, jolloin moninäkömähäun hakuperiaate ei päässyt oikeuksiinsa. Toisaalta tietomalli kykeni kuvaamaan aineistossa olleen tiedon yhtä hyvin kuin skeemaominaisuuksiin nojautuva järjestelmä.

Koska puoliautomaattinen toimintakuvausten generointi osoittautui niin vaikeaksi, tulisi siihen keskittyä jatkotutkimuksessa. Prosessia on edelleen kehitettävä, jotta tarkkuus paranee ja turhalta toistolta vältytään. Automaattista generointia voidaan mielestäni parantaa kahdella tavalla: parantamalla ontologisen tietämyksen hyödyntämistä sekä kehittämällä

tekstin analysoinnin tekniikoita. Erityisesti avainsanoihin perustuvaa toimintakuvausten generointia voisi parantaa hyödyntämällä ontologista tietämystä. Ontologisen tietämyksen pohjalta voisi olla mahdollista päätellä tai arvata toiminnassa osallisten temaattisia rooleja nykyistä tarkemmin. Roolituksen ollessa lopputuloksessa tarkemmalla tasolla, haut toimivat paremmin ja päättely antaa tarkempia tuloksia. Useimmiten artefaktien kiinnostavimmat annotaatiot on kuitenkin kuvattu tekstimuodossa. Pelkkien avainsanojen perusteella on lähes mahdoton rekonstruoida teksteissä esiintyviä tilanteita. Jotta toimintakeskeistä sisällönkuvailua voidaan tehdä biografiolle, runoilille ja muille kirjallisille teoksille, tarvitaan tapa irrottaa kehyksiä, tai edes kehusehdotuksia, luonnollisesta kielestä.

Tätä työtä varten valmistui KulttuuriSammon vision erilaisista hakutavoista toimintakeskeinen semanttinen moninäkömähaku sekä projektiot kartalle ja aikajanelle. Semanttisen moninäkömähäun käyttöliittymän toteutus vaatii edelleen kehitystyötä. Käyttäjättestit todistivat, että näkömien selaaminen ja sopivien kategorioiden etsiminen on tehtävä huomattavasti nykyistä helpommaksi. Tässä työssä olen esittänyt joitakin käyttäjättestien tuloksiin ja huomioihin perustuvia parannusehdotuksia. Näiden parannusten vaikutuksia käytettävyyteen olisi kiinnostava tutkia käyttämällä samaa testipatteristoa ja vertailemalla tuloksia tämän tutkimuksen tuloksiin. Jatkossa tulisi kehittää myös muita vaihtoehtoisia hakutapoja, kuten näyttelyjä, koostehakua ja teemahakua. Jatkotutkimuksessa tulisi myös tutkia videon ja audion kuvailua toimintakeskeisesti ja niiden kytkemistä hakujärjestelmään.

On hyvin todennäköistä, että pian ilmestyy multimediatiedonhakuun soveltuvia benchmarking-yhteisöjä. Jotta SeCo:ssa kehitettäviä semanttisia tiedonhakujärjestelmiä voidaan verrata muihin multimediatiedonhaun lähestymistapoihin, näen tärkeänä, että SeCo osallistuu benchmarking-yhteisöihin ja testaa kehitettyjä järjestelmiä yhteisöjen testiaineistoilla ja menetelmin. Tätä kautta saadaan luotettavia tutkimustuloksia hakujärjestelmien suorituskyvystä muihin hakujärjestelmiin verrattuna.

Työssäni tutkin semanttista webiä täysin suljetussa ympäristössä, yhden sovelluksen teknisenä alustana. Suljetussa ympäristössä on huomattavasti helpompi pilotoida kehitettyjä ontologioita, palveluja ja työkaluja. Jatkotutkimuksessa KulttuuriSampo tulisi kuitenkin viedä suuntaan, jossa se olisi hakujärjestelmän lisäksi myös WWW:ssä oleva julkinen tietokanta, joka tarjoaa kiinnostavaa aineistoa RDF-muodossa muiden järjestelmien hyödynnettäväksi. Semanttisen webin idea on kuitenkin kaikkea muuta kuin suljettu. Sen sijaan, että pohdimme museoissa ja tutkimusryhmissä, minkälaisia kiinnostavia loppukäyttöliittymiä voisimme tarjota loppukäyttäjille, voimme tarjota tekijöille aineistot ja katsoa mitä he keksivät tehdä niillä. Nämä niin sanotut mashupit, joissa hyödynnetään kahden tai useamman erillisen palvelun sisältöjä tai alustoja ja liitetään ne luovalla tavalla yhteen, ovat WWW:n kuumien sanojen tällä hetkellä. Alkuun tutkimuksessa pääsisi yksinkertaisesti antamalla RSS/RDF-tiedot hakutuloksista tarjolle.

Uuden teknologian parissa työskennellessä aika kuluu ongelmia ratkoessa. Niin tässäkin tapauksessa kului. Kehitystyö oli jatkuvaa taistelua työkalujen, prosessointitehon, muistinkäytön ja tilavaatimusten kanssa. Kohtaamani ongelmat johtuivat pääasiassa aineiston suuresta määrästä ja muodostetun tietämuskannan koosta. Olemassa olevat työkalut, kuten ontologiaeditorit, on suunniteltu yksinkertaisten, korkeintaan pari sataa dokumenttia sisältävien tietämuskantojen käsittelyyn. Työkalua, joka pystyisi käsittelemään KulttuuriSammon yli 10 000 dokumentin aineistoa ongelmitta, ei ole olemassa. Prosessointitehon rajoitteita puolestaan jouduin usein kiertämään tekemällä kattavaa esilaskentaa ja rakentamaan välimuistitoteutuksia. Todellisesta reaaliaikaisuudesta ollaan vielä kaukana.

Vaikka RDF:n ja OWL:n käsittely tuottaa nyt ongelmia, uskon menetelmien ja tietokoneiden kehittyvän niin, että nämä ongelmat voitetaan. Tämänkin työn aikana tehdyt suorituskykyharppaukset hakujärjestelmässä olivat huimia. Syksyllä 2006 yhden haun tekeminen kesti noin 10 sekuntia. Keväällä 2007 olimme jo päässeet siedettäviin puolen sekunnin aikoihin. Semanttisen webin suurimpana haasteena näen kuitenkin RDF-muodossa olevan aineiston tuottamisen työläyden. Nykyisellään sisällön tuottaminen KulttuuriSampon on vaivalloista ja vie paljon aikaa. Ryhdytäänkö työhön ilman vakuuttavia esimerkkejä työn hyödyistä kuvailujen tuottajalle? Ongelman ratkaisemiseksi sisällönkuvailuun voisi ottaa wiki-tyyliin loppukäyttäjät mukaan. Heille tulisi tarjota keino lisätä itse toimintakuvauksia ja näin yhteisöllisesti parantaa portaalin sisällön laatua. Kiinnostavien sisältöjen osalta tämä lähestymistapa saattaisi toimia.

## Lähteet

ALATERÄ, Anu ja HALTTUNEN, Kai. (2002) *Tiedonhaun perusteet - osa lukutaitoa*. 2002. 144 s. ISBN 951-692-527-8.

ANON. (2001) *ICOM Statutes* [online]. International Council of Museums, 2001 [viitattu 4.4.2006]. Article 2: definitions. Saatavilla WWW-muodossa: <<http://icom.museum/statutes.html#2>>.

ANON. (2006a) *Mikä on museo?* [online]. Suomen museoliitto, 2006a [viitattu 4.4.2006]. Saatavilla WWW-muodossa: <<http://www.museot.fi/>>.

ANON. (2006b) *Informaatiotutkimuksen perusteet oppimateriaali* [online]. Oulun yliopisto, suomen kielen, informaatiotutkimuksen ja logopedian laitos., 2006b [viitattu 16.1.2007]. Saatavilla WWW-muodossa: <<http://www oulu.fi/informaatiotutkimus/perusteet/materiaali/>>.

ANTONIOU, Grigoris ja VAN HARMELEN, Frank. (2004) *A Semantic Web primer*. The MIT Press, 2004. 259 s. ISBN 0-262-01210-3.

BAEZA-YATES ja RIBIERO-NETO. (1999) *Modern Information Retrieval*. 1999. 513 s. ISBN 0-201-39829-X.

BEARMAN, David ja TRANT, Jennifer. (1997) Museums and the Web: a conference, a community, and a challenge. Teoksessa *Museums and the Web, 1997: Selected Papers*. 1997.

BERNERS-LEE, Tim ja FISCHETTI, Mark. (1999) *Weaving the web*. San Fransisco: Harper, 1999.

BERNERS-LEE, Tim. (2000) *RDF and the Semantic Web*. 2000. Keynote-puheen kalvot. XML2000 3-8.12.2000. Saatavilla WWW-muodossa: <<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>>.

BERNERS-LEE, Tim, HENDLER, James, ja LASSILA, Ora. (2001) The Semantic Web. *Scientific American*. 2001, no. 284, s. 34-43.

BERTINO, Elisa, CATANIA, Barbara, ja FERRARI, Elena. (1999) Multimedia IR: Models and Languages. Teoksessa BAEZA-YATES ja RIBIERO-NETO. *Modern Information Retrieval*. 1999.

BORLUND, Pia. (2003) The IIR evaluation model: a framework for evaluation of interactive information retrieval systems . *Information Research*. 2003, vol. 8, no. 3.

BRICKLEY, Dan ja GUHA, R. (2004) *RDF Vocabulary Description Language 1.0: RDF Schema* [online]. 2004 [viitattu 13.6.2006]. Saatavilla WWW-muodossa: <<http://www.w3.org/TR/rdf-schema/>>.

CROFTS, Nick, et al. (2005) *Definition of the CIDOC Conceptual Reference Model. Version 4.2.* 2005. Saatavilla WWW-muodossa: <[http://cidoc.ics.forth.gr/docs/cidoc\\_crm\\_version\\_4.2.pdf](http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.pdf)>.

DOERR, Martin. (2003) The CIDOC CRM - an ontological approach to semantic interoperability of metadata. *AI Magazine*. 2003, vol. 24, no. 3.

DUSHAY, Naomi. (2002) Localizing experience of digital content via structural metadata. *JCDL '02: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*. 2002, s. 244-252.

EAKINS, J. ja GRAHAM, M.E. (1999) Content-based Image Retrieval. *A report to the JISC Technology Applications Programme: JTAP report*. 1999, no. 39.

FAULKNER, Kristine. (2000) *Usability Engineering*. Palgrave, 2000. 244 s. ISBN 0-333-77321-7.

FENSEL, Dieter. (2004) *Ontologies: Silver bullet for knowledge management and electronic commerce (2nd Edition)*. Springer-Verlag, 2004.

FERNANDEZ, Mariano, GOMEZ-PEREZ, Asuncion, ja JURISTO, Natalia. (1997) METHONTOLOGY: from Ontological Art towards Ontological Engineering. *Proceedings of the AAAI97 Spring Symposium Series on Ontological Engineering*. 1997, s. 33-40.

GAMMA, Erich, et al. (1995) *Design patterns: elements of reusable object-oriented software*. 1995. 395 s. ISBN 0-201-63361-2.

GARRETT, Jesse. (2005) *Ajax: A New Approach to Web Applications* [online]. 2005 [viitattu 1.12.2007]. Saatavilla WWW-muodossa: <<http://www.adaptivepath.com/ideas/essays/archives/000385.php>>.

GILLILAND, Anne. (2000) Setting the Stage. Teoksessa *Introduction to metadata: Pathways to Digital Information* [online]. Online Edition. 2000 [viitattu 29.4.2006]. Saatavilla WWW-muodossa: <[http://www.getty.edu/research/conducting\\_research/standards/intrometadata/setting.pdf](http://www.getty.edu/research/conducting_research/standards/intrometadata/setting.pdf)>.

GÓMEZ-PÉREZ, Asunción, FERNÁNDEZ-LÓPEZ, Mariano, ja CORCHO, Oscar. (2004) *Ontological Engineering*. 2004. 403 s. ISBN 1-85233-551-3.

- GRUBER, T. (1993) A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*. 1993, vol. 5, no. 2, s. 199-220.
- HAASE, Kenneth. (2004) Context for semantic metadata. *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*. 2004, s. 204-211.
- HAMMER, Michael ja MCLEOD, Dennis. (1981) Database Description with SDM: A Semantic Database Model. *ACM Transactions on Database Systems*. 1981, vol. 6, no. 3, s. 351-386.
- HAUPTMANN, Alexander ja CHRISTEL, Michael. (2004) Successful approaches in the TREC video retrieval evaluations. *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*. 2004, s. 668-675.
- HERTZUM, M. (2003) Requests for Information from a Film Archive: A Case Study of Multimedia Retrieval. *Journal of Documentation*. 2003, vol. 2, no. 59, s. 168-186.
- HUNTER, Jane. (2002) Combining the CIDOC CRM and MPEG-7 to Describe Multimedia in Museums. *Museums and the Web 2002: Selected Papers*. 2002.
- HYVÖNEN, Eero, et al. (2002) Semantic Interoperability on the Web: Case Finnish Museums Online. *Towards the semantic web and web services, Proceedings of XML Finland 2002 Conference*. 2002, s. 41-53.
- HYVÖNEN, Eero, et al. (2004a) A Semantic Portal for Publishing Museum Collections on the Web. *Proceedings of ECAI/PAIS 2004*. 2004a.
- HYVÖNEN, Eero, et al. (2004b) Finnish Museums on the Semantic Web. User's Perspective on MuseumFinland. *Proceedings of Museums and the Web 2004*. 2004b.
- HYVÖNEN, Eero, et al. (2004c) A Cultural Community Portal for Publishing Museum Collections on the Semantic Web. *Proceedings of 16th European Conference on Artificial Intelligence (ECAI2004): Workshop on Application of Semantic Web Technologies to Web Communities*. 2004c.
- HYVÖNEN, Eero, et al. (2004d) Publishing Museum Collections on the Semantic Web - the MuseumFinland Portal. *Proceedings of WWW2004, Alternate Track Papers and Posters*. 2004d.
- HYVÖNEN, Eero, SAARELA, Samppa, ja VILJANEN, Kim. (2004e) Application of Ontology Techniques to View-Based Semantic Search and Browsing. *The Semantic Web: Research and Applications. Proceedings of the First European Semantic Web Symposium (ESWS 2004)*. 2004e.

- HYVÖNEN, Eero, et al. (2005) MuseumFinland - Finnish Museums on the Semantic Web. *Journal of Web Semantics*. 2005, vol. 3, no. 2, s. 25.
- HYVÖNEN, Eero, et al. (2006) CultureSampo-Finnish Culture on the Semantic Web: The Vision and First Results . *Developments in Artificial Intelligence and the Semantic Web - Proceedings of the 12th Finnish AI Conference STeP 2006*. 2006.
- HYVÖNEN, Eero, et al. (2007) Elements of a National Semantic Web Infrastructure - Case Study Finland on the Semantic Web (Invited paper). *Proceedings of the First International Semantic Computing Conference (IEEE ICSC 2007), Irvine, California*. 2007.
- INGWERSEN, Peter ja JÄRVELIN, Kalervo. (2005) *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer, 2005. 448 s. ISBN 1-4020-3850-X.
- INGWERSEN, Peter. (1992) *Information retrieval interaction*. London: Taylor Graham, 1992.
- JAIMES, Alejandro, et al. (2005) Multimedia information retrieval: what is it, and why isn't anyone using it?. *MIR '05: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*. 2005, s. 3-8.
- JORDAN, Patrick. (1998) *An introduction to usability*. London: Taylor and Francis, 1998.
- JOSE, J.M., FURNER, J., ja HARPER, D.J. (1998) Spatial querying for image retrieval: a user-oriented evaluation. *In Proceedings of SIGIR'98*. 1998.
- JUNNILA, Miikka. (2006) *Tietosisältöjen semanttinen yhdistäminen toimintakuvausten avulla*. 2006.
- JUNNILA, Miikka, HYVÖNEN, Eero, ja SALMINEN, Mirva. (2006) Describing and Linking Cultural Semantic Content by Using Situations and Actions. *Developments in Artificial Intelligence and the Semantic Web - Proceedings of the 12th Finnish AI Conference STeP 2006*. 2006.
- JÄRVELIN, Kalervo ja KEKÄLÄINEN, Jaana. (2002) *Tiedonhaun menetelmät opintoaineisto* [online]. 2002 [viitattu 13.6.2006]. Saatavilla WWW-muodossa: <<http://www.internetix.fi/opinnot/opintojaksot/0viestinta/informaatiotutkimus/po4/>>.
- JÄRVELIN, Kalervo. (1995) *Tekstitiedonhaku tietokannoista: Johdatus periaatteisiin ja menetelmiin*. Espoo: Suomen ATK-kustannus, 1995.
- JÄRVENPÄÄ, Eila ja KOSONEN, Karri. (1996) *Johdatus tutkimusmenetelmiin ja tutkimuksen tekemiseen*. Espoo: Teknillinen korkeakoulu. Työpsykologian ja johtamisen laboratorio, 1996. 101 s.



LEW, Michael, et al. (2006) Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.* 2006, vol. 2, no. 1, s. 1-19.

LUKKA, Kari. (2000) The Key Issues of Applying the Constructive Approach to Field Research . Teoksessa *Management Expertise for the New Millenium: In Commemoration of the 50th Anniversary of the Turku School of Economics and Business Administration*. Turun kauppakorkeakoulun julkaisuja, 2000, s. 113-128.

LUKKA, Kari. (2001) *Konstrukttiivinen tutkimusote* [online]. 2001 [viitattu 11.6.2006]. Saatavilla WWW-muodossa: <<http://www.metodix.com>>.

MACKENZIE, Scott. (2002) *Research Note: Within-subjects vs. Between-subjects Designs: Which to Use?* [online]. 2002 [viitattu 10.4.2007]. Saatavilla WWW-muodossa: <<http://www.yorku.ca/mack/RN-Counterbalancing.html>>.

MANOLA, Frank ja MILLER, Eric. (2004) *RDF Primer: W3C Recommendation, helmikuu 2004* [online]. 2004 [viitattu 12.6.2006]. Saatavilla WWW-muodossa: <<http://www.w3.org/TR/rdf-primer/>>.

MCGUINNESS, Deborah ja VAN HARMELEN, Frank. (2004) *OWL Web Ontology Language Overview: W3C Recommendation 10 Feb 2004*. 2004.

MÄKELÄ, Eetu. (2006) *View-Based Search Interfaces for the Semantic Web*. 2006.

MÄKELÄ, Eetu, et al. (2004) *OntoViews - A Tool for Creating Semantic Web Portals. Proceedings of the 3rd International Semantic Web Conference (ISWC 2004)*. 2004.

MÄKELÄ, Eetu, HYVÖNEN, Eero, ja SIDOROFF, Teemu. (2005) *View-Based User Interfaces for Information Retrieval on the Semantic Web. Proceedings of the ISWC-2005 Workshop End User Semantic Web Interaction*. 2005.

MÄKELÄ, Eetu, HYVÖNEN, Eero, ja SAARELA, Samppa. (2006) *Ontogator - A Semantic View-based Search Engine Service for Web Applications. Proceedings of the 5th International Semantic Web Conference (ISWC 2006)*. 2006.

NILES, Ian ja PEASE, Adam. (2001) *Towards a Standard Upper Ontology. FOIS '01: Proceedings of the international conference on Formal Ontology in Information Systems*. 2001, s. 2-9.

O'REILLY, Tim. (2005) *What Is Web 2.0* [online]. 2005 [viitattu 24.9.2006]. Saatavilla WWW-muodossa: <<http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>>.

OSSENBRUGGEN, Jacco, et al. (2006) *Image Annotation on the Semantic Web: W3C Working Draft 22 March 2006* [online]. 2006 [viitattu 30.4.2006]. Saatavilla WWW-muodossa: <<http://www.w3.org/TR/swbp-image-annotation/>>.

POLLIT, Steven. (1998) The key role of classification and indexing in view-based searching. *International Cataloguing and Bibliographic Control*. 1998, vol. 2, no. 27, s. 37-40.

PRUD'HOMMEAUX, Eric ja SEABORNE, Andy. (2007) *SPARQL Query Language for RDF: W3C Proposed Recommendation 12 November 2007* [online]. 2007 [viitattu 1.12.2007]. Saatavilla WWW-muodossa: <<http://www.w3.org/TR/rdf-sparql-query/>>.

ROCHA, Crisiano, SCHWABE, Daniel, ja POGGI DE ARAGÃO, Marcus. (2004) A Hybrid Approach for Searching in the Semantic Web. *In Proceedings of the 13th International World Wide Web Conference*. 2004.

ROSENFELD, Louis ja MORVILLE, Peter. (2002) *Information architecture for the World Wide Web*. Second Edition. 2002. 461 s. ISBN 0-596-00035-9.

RUOTSALO, Tuukka ja HYVÖNEN, Eero. (2007) An Event-based Approach for Semantic Metadata Interoperability. *Proceedings of the 6th International Semantic Web Conference (ISWC 2007)*. 2007.

SAARELA, Samppa. (2004) *Näkymäpohjainen RDF-haku*. 2004. Saatavilla WWW-muodossa: <<http://www.seco.hut.fi/publications/2004/saarela-nakymapohjainen-rdf-haku-2004.pdf>>.

SAEED, John. (1997) *Semantics*. Oxford: Blackwell Publishers Ltd, 1997.

SALMINEN, Mirva. (2006) *Kuvien ja videoiden semanttinen sisällönkuvailu*. 2006. Saatavilla WWW-muodossa: <<http://www.seco.tkk.fi/publications/2006/salminen-sisallonkuvailu-2006.pdf>>.

SARACEVIC, Tefko. (1995) Evaluation of evaluation in information retrieval. *Proceedings of the 18th ACM Sigir Conference on Research and Development of Information Retrieval*. 1995, s. 138-146.

SOWA, John. (2000) *Knowledge representation: logical, philosophical and computational foundations*. Brooks/Cole, 2000. ISBN 0 534-94965-7.

STUDER, R, BENJAMINS, V, ja FENSEL, D. (1998) Knowledge engineering: Principles and methods. *IEEE Transactions on Data and Knowledge Engineering*. 1998, vol. 25, no. 1 & 2, s. 161–197.

SURE, York, et al. (2004) Why Evaluate Ontology Technologies? Because It Works!. *IEEE Intelligent Systems*. 2004, vol. 19, no. 4, s. 74-81.

SUTTON, Stuart. (2004) Digital Library Infrastructure: Metadata & the Education Domain. Teoksessa HILLMANN, Diane ja WESTBROOKS, Elaine. *Metadata in Practice*. American Library Association, 2004.

TAYLOR, Robert. (1968) Question-Negotiation and Information Seeking in Libraries. *College and Research Libraries*. 1968, s. 178-194.

WHITE, Ryen. (2004) *Implicit Feedback for Interactive Information Retrieval*. 2004.

VOORHEES, Ellen. (2004) Overview of TREC 2004. *Text Retrieval Conference*. 2004.

YEE, Ka-Ping, et al. (2003) Faceted metadata for image search and browsing. *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*. 2003, s. 401-408.

ZHANG, Junliang ja MARCHIONINI, Gary. (2005) Evaluation and evolution of a browse and search interface: relation browser++. *dg.o2005: Proceedings of the 2005 national conference on Digital government research*. 2005, s. 179-188.

## Liite A Tehtävien suoritusjärjestys koehenkilöittäin

Taulukko A.1. Tehtävien suoritusjärjestys

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
H1	1	2	3	4	5	6
H2	5	1	4	6	2	3
H3	3	6	1	5	4	2
H4	6	4	2	1	3	5
H5	2	5	6	3	1	4
H6	4	3	5	2	6	1
H7	6	5	4	3	2	1
H8	2	6	3	1	5	4
H9	4	1	6	2	3	5
H10	1	3	5	6	4	2
H11	5	2	1	4	6	3
H12	3	4	2	5	1	6

# Liite B Alkukyselyn kyselylomake

## Alkukysely

Tämän lomakkeen tiedot tarjoavat meille taustatietoa, joka on hyödyllistä analysoidessamme myöhemmissä vaiheissa antamiasi vastauksia.



ID

### Perustiedot

Sukupuoli:

Ikä:

Ammatti:

Koulutustaso:

### Kokemus tiedonhausta

Käyttätkö työssäsi tai vapaa-ajanasi tiedonhakujärjestelmiä?

Kyllä

En

Kuinka usein käytät tiedonhakujärjestelmiä?

jotain kertoja vuodessa

Löydän hakemani tiedon

en koskaan

Mitä seuraavista tiedonhakujärjestelmistä käytät

Internetin hakukoneita (Google, Yahoo jne.)

Kirjastojen hakukoneita

Wiki-tyyisiä online-tietosanakirjoja

Muita tiedonhakujärjestelmiä

Tiedonhakujärjestelmien käyttö on minulle yleisesti

vaikeaa

helppoa

miellyttävää

yksinkertaista

# Liite C Tehtäväkohtainen kyselylomake

## Tehtäväkohtainen kyselylomake

Järjestelmän arvioimiseksi pyydämme sinua vastaamaan juuri tekemääsi hakua ja käyttämäsi järjestelmää koskeviin kysymyksiin. Olemme kiinnostuneita mielipiteistäsi, joten vastaa vapaasti ja muista, että ei ole olemassa oikeita tai vääriä vastauksia. Muista että arvioimme järjestelmää, emme sinua.

ID   
Tehtävä

valkeakäyttöinen  C  C  C  C helppokäyttöinen  
 epälooginen  C  C  C  C looginen  
 tylsä  C  C  C  C stimuloiva  
 toimimaton  C  C  C  C toimiva

En käyttänyt vapaatekstihakua

## Osa 1: Kysymykset tehtävästä

### Annettu tehtävä oli

epäselvä  C  C  C selkeä  C  
 vaikea  C  C  C helppo  C  
 vieras  C  C  C tuttu  C  
 monimutkainen  C  C  C yksinkertainen  C  
 tylsä  C  C  C kiinnostava  C

### Tiesin tarkalleen minkälaista tietoa etsiä

täysin eri mieltä  C  C  C  C täysin samaa mieltä

### Suoriuduin tehtävästä mielestäni hyvin

täysin eri mieltä  C  C  C  C täysin samaa mieltä

### Tunnen löytäneeni tehtävän kannalta oleelliset dokumentit

täysin eri mieltä  C  C  C  C täysin samaa mieltä

### Vapaat kommentit tehtävästä



### Moninäkymähaku oli tämän tehtävän suorituksessa

hyödytön  C  C  C hyödyllinen  
 valkeakäyttöinen  C  C  C helppokäyttöinen  
 epälooginen  C  C  C looginen  
 tylsä  C  C  C stimuloiva  
 toimimaton  C  C  C toimiva

En käyttänyt kategoriahakua

### Vapaat kommentit tiedonhaun menetelmistä

## Osa 2: Tiedonhaun menetelmät



### Vapaatekstihaku oli tämän tehtävän suorituksessa

hyödytön  C  C  C hyödyllinen

## Osa 3: Tulostajako/löydyneet dokumentit

### Tulostajako (löydetty kohteet) oli tehtävän hakuja tehdessäsi

irrelevantti  C  C  C relevantti  
 epälooginen  C  C  C looginen



**Vapaat kommentit tulosjoukosta**

**Osa 4: Hakuprosessi / järjestelmällä hakeminen**

**Hakuprosessi oli**

stressaava	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	rentouttava
tylsä	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	kiinnostava
väsyttävä	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	piristävä
valkea	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	helppo
monimutkainen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	yksinkertainen
epämielelyttävä	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	mielelyttävä

**Vapaat kommentit hakuprosessista**



## Loppukysely

Kokeen tarkoituksena oli arvioida hakujärjestelmän ja sen osien hyödyllisyyttä hakutehtävien suorituksessa.

Vastaa seuraaviin kysymyksiin niin hyvin kuin mahdollista.

ID



### 1. Hakunäkymät

#### Toimijat

- Henkilöt** (3624)  
Ryhmittäile kategorian mukaa
  - organisaatiot** (446)  
Ryhmittäile kategorian mukaa
- Avaa kaikki - Sulje kaikki

#### Tehtävien suorittamisen yhteydessä

##### Toimijat-näkymä oli

hyödytön	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
vaikeakäyttöinen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
epälooginen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
tylsä	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
stimuloiva	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

En käyttänyt

#### Paikka

- Aasia** (2)  
Ryhmittäile kategorian mukaa
  - Afrikka** (164)  
Ryhmittäile kategorian mukaa
  - Eurooppa** (3691)  
Ryhmittäile kategorian mukaa
  - Pohjois-Amerikka** (1)  
Ryhmittäile kategorian mukaa
- Avaa kaikki - Sulje kaikki

#### Tehtävien suorittamisen yhteydessä

##### Paikka-näkymä oli

hyödytön	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
vaikeakäyttöinen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
epälooginen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
tylsä	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
stimuloiva	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

En käyttänyt

#### Toiminta

- havainnointi ja aistiminen**  
Ryhmittäile kategorian mukaa
  - kognitio** (7892)  
Ryhmittäile kategorian mukaa
  - kulutus ja käyttö** (39)  
Ryhmittäile kategorian mukaa
  - liikkuminen** (19)  
Ryhmittäile kategorian mukaa
  - luominen** (8193)  
Ryhmittäile kategorian mukaa
  - mentaalinen toiminta** (787)  
Ryhmittäile kategorian mukaa
  - muuttaminen** (94)  
Ryhmittäile kategorian mukaa
  - olemassaole** (7876)  
Ryhmittäile kategorian mukaa
  - omistus ja omistuksen siirtyminen** (2804)  
Ryhmittäile kategorian mukaa
  - ruumiintoinnnot ja hoito**  
Ryhmittäile kategorian mukaa
  - säälimiät** (2)  
Ryhmittäile kategorian mukaa
  - vuorovaikutus** (7890)  
Ryhmittäile kategorian mukaa
- Avaa kaikki - Sulje kaikki

#### Tehtävien suorittamisen yhteydessä

##### Toiminta-näkymä oli

hyödytön	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
vaikeakäyttöinen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
epälooginen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
tylsä	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
stimuloiva	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

En käyttänyt

#### Toiminnan väline

- aine** (1100)  
Ryhmittäile kategorian mukaa
  - konkreettinen eloton obje**  
(1145)  
Ryhmittäile kategorian mukaa
  - orgaaninen rakenne** (10)  
Ryhmittäile kategorian mukaa
- ns\_30\_maoryhma:arkisto\_Ja\_ (8)

### Tehtävien suorittamisen yhteydessä Toiminnan väline -näkymä oli

hyödytön	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
valkeakäyttöinen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
epälooginen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
tylsä	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
stimuloiva	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

En käyttänyt

### Tehtävien suorittamisen yhteydessä Aika-näkymä oli

hyödytön	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
valkeakäyttöinen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
epälooginen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
tylsä	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
stimuloiva	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

En käyttänyt

### Tehtävien suorittamisen yhteydessä Aineistotyyppi-näkymä oli

hyödytön	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
valkeakäyttöinen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
epälooginen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
tylsä	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
stimuloiva	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

En käyttänyt

#### Kokoelma

- Antellin kokoelmat (1483)
- Arkeologian esinekokoelma (
- Armifitit kokoelma (9)
- Aspelin-Haapyjän kokoelma
- Ateneumin taidemuseo (1549)
- Bergmanin kokoelma (80)
- Enckellin kokoelma (5)
- Finnen kokoelma (21)
- Historialliset kokoelmat (648)
- Idestamin kokoelma (16)
- Kansatieteelliset kokoelmat (
- Kansatieteen kuvakokoelma
- Kotkanien kokoelma (2)
- LAATOKAN-KARJALAN MUSEO
- NYKYTAITEEN MUSEO KIASMA (1
- POHJOIS-KARJALAN MUSEO (1
- Rakennushistorian kokoelma
- Seurasareen ulkomuseon
- kokoelmat (144)
- Sinebrychhoffin taidemuseo (3
- Sjomanin kokoelma (1)
- Sundströmin kokoelma (3)
- Suomalais-ugrilaiset kokoel
- (2679)
- Suomen merimuseon kokoelr
- (185)
- Topeliuksen kokoelma (2)
- Urajärven kokoelma (4)
- Urho Kerkkosen kokoelma (1)
- VALOKUVAKOKOELMA (1)
- VIIPURIN MUSEO:KIRKOLLISE
- (19)
- VIIPURIN MUSEO:KÄSTTYÖKAL
- (1)
- VIIPURIN MUSEO:MAALAUKSE
- (114)
- VIIPURIN MUSEO:VAATTEET...
- Yleisnouraritiset kokoelmat (

## Tehtävien suorittamisen yhteydessä Kokoeima-näkymä oli

hyödytön	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	hyödyllinen
vaikeakäyttöinen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	helppokäyttöinen
epälooginen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	looginen
tylsä	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	stimuloiva

En käyttänyt

## 2. Järjestelmä

### Hakujärjestelmä oli

tehoton	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	tehokas
epäluotettava	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	luotettava
tylsä	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	stimuloiva
joustamaton	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	joustava
vaikeasti opittava	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	helposti opittava
toimimaton	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	toimiva
hyödytön	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	hyödyllinen
turhauttava	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	tyytytyksen antava
tavanomainen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	uudentyyppinen
hidas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nopea
staattinen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	elämyksellinen

### Hakujärjestelmä pitää sisällään

Yli 10 000 hakukohdetta	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa
Kohteiden paikkatiedot ja koordinaatit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa
Valokuissa käytetyn kameran mallin ja objektiivin polttovälin	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa
Maalauksissa taidekritiikon arvion maalauksesta	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa

## Hakujärjestelmällä on mahdollista

Tarkastella hakutuloksia monin eri tavoin esitettynä	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa
Selata kulttuurihenkilöiden verkostoja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa
Muokata hakuvaimia kesken haun	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa
Hakea vapaatekstihaun ja näkymähaun yhdistelmällä	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa
Hakea näkymähaun kategorioita tekstihaulla	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa
Ryhmitellä hakutuloksia loogisiin kokonaisuuksiin	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa
Löytää sinuun itseesi liittyvät dokumentit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa
Rajata haettavaa paikkaa kartalla	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Kyllä
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ei
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	En osaa sanoa

## 5. Vapaat kommentit

### Mitkä asiat jäivät päällimmäiseksi mieleen hakujärjestelmästä?

### Mitkä asiat jäivät päällimmäiseksi mieleen kokeesta?

## Liite E Transaktiolokin tapahtumatyypit

```
#-----  
# Käyttäjätesteissä käytettävän jäljituspolut merkinnät ja formatointi  
# määräytyvät tämän tiedoston mukaan.  
#-----  
  
#-----  
# Yleiset interaktiotagit  
#-----  
#PAGE: Ladattiin sivu  
#PAGE ; sivun nimi  
PAGE=PAGE;%1$s  
  
#LOADTIME: Sivun lataamisen kesto  
#LOADTIME ; sivun nimi ; aika]  
LOADTIME=LOADTIME;%1$s;%2$s  
  
#-----  
# Hakutuloksen tagit  
#-----  
#RESDOC: palautettujen dokumenttien kokonaismäärä  
#RESDOC ; dokumenttien määrä  
RESDOC=RESDOC;%1$s  
  
#RESVIEW: palautettujen dokumenttien katselumoodi (ryhmitelty, kartta, aikajana)  
#RESVIEW ; moodi  
RESVIEW=RESVIEW;%1$s  
  
#SELECTDOC: valittiin objekti tarkemman tarkastelun kohteeksi  
#SELECTDOC ; kohteen uri  
SELECTDOC=SELECTDOC;%1$s  
  
#-----  
# Tulosjoukon aikajanakatselumoodin tagit  
#-----  
#TIMELINERESDOC: Aikajananäkymän dokumenttimäärä verrattuna kokonaismäärään  
#TIMELINERESDOC ; ajallisesti löydettyt dokumentit ; dokumenttien kokonaismäärä  
TIMELINERESDOC=TIMELINERESDOC;%1$s;%2$s  
  
#TIMELINELOADTIME: Aikajananäkymän latausaika  
#TIMELINELOADTIME ; aika; lopetettiin laskenta kesken (true/false)  
TIMELINELOADTIME=TIMELINELOADTIME;%1$s;%2$s  
  
#-----  
# Tulosjoukon karttakatselumoodin tagit  
#-----  
#MAPRESDOC: Karttatulosjoukon dokumenttimäärä verrattuna kokonaismäärään  
#MAPRESDOC ; kartalle löydettyt dokumentit ; dokumenttien kokonaismäärä  
MAPRESDOC=MAPRESDOC;%1$s;%2$s  
  
#MAPLOADTIME: Karttanäkymän latausaika
```

```
#MAPLOADTIME ; aika ; lopetettiin laskenta kesken (true/false)
MAPLOADTIME=MAPLOADTIME;%1$s;%2$s

#-----
# Tulosjoukon ryhmitely katselumoodin tagit
#-----
#NMBGROUPS: Ryhmien lukumäärä
#NMBGROUPS ; lukumäärä
NMBGROUPS=NMBGROUPS;%1$s

#GOTOPAGE: Sivutetun tulosjoukon sivun vaihto
#GOTOPAGE ; ryhmän tunniste ; ryhmän nimi ; sivunumero
GOTOPAGE=GOTOPAGE;%1s;%2s;%3s

#-----
# Puunäkymätagit
#-----
#EXPANDNODE: Käyttäjä avasi puun solmun
#EXPANDNODE ; avatun solmun id ; avatun solmun otsikko
EXPANDNODE=EXPANDNODE;%1$s;%2$s

#EXPANDALLNODES: Käyttäjä avasi näkymän kaikki solmut
#EXPANDALLNODES ; avatun näkymän otsikko
EXPANDALLNODES=EXPANDALLNODES;%1$s

#COLLAPSENODE: Käyttäjä sulki puun solmun
#COLLAPSENODE ; suljetun noodin id ; suljetun noodin otsikko
COLLAPSENODE=COLLAPSENODE;%1$s;%2$s

#COLLAPSEALLNODES: Käyttäjä sulki näkymän kaikki solmut
#COLLAPSEALLNODES ; suljetun näkymän otsikko
COLLAPSEALLNODES=COLLAPSEALLNODES;%1$s

#GROUPBY: Käyttäjä ryhmitteli tulokset solmun mukaan
#GROUPBY ; ryhmittelyyn käytetyn solmun id ; solmun label]
GROUPBY=GROUPBY;%1$s;%2$s

#UNGROUP: Käyttäjä poisti ryhmittelyn
#UNGROUP
UNGROUP=UNGROUP

#-----
# Kysely- ja kyselyn muokkaustagit
#-----
#CURQUERY: Hakuavaimet
#CURQUERY;Hakulause;Hakukategoriat (muodossa {x,y,z,..})
CURQUERY=CURQUERY;%1$s;%2$s

#NEWSEARCH: Käyttäjä aloitti alusta (aloita alusta, tai poista kaikki hakuehdot)
#NEWSEARCH
NEWSEARCH=NEWSEARCH
```

#SELECTCAT: Käyttäjä valitsi kategorian  
#SELECTCAT ; Kategorian id ; Kategorian otsikko  
SELECTCAT=SELECTCAT;%1\$s;%2\$s

#REMOVECAT: Käyttäjä poisti kategorian  
#REMOVECAT ; Kategorian id ; Kategorian otsikko  
REMOVECAT=REMOVECAT;%1\$s;%2\$s

#BROADEN: Kategoriaehtoa laajennettiin (valittiin valitun kategorian  
yläkategoria)  
#BROADEN ; Valittu kategoria ; alakategoria %; skipattujen tasojen määrä%  
BROADEN=BROADEN;%1\$s;%2\$s

#NARROW: Kategoriaehtoa supistettiin (valittiin valitun kategorian alikategoria)  
#NARROW ; Valittu kategoria ; yläkategoria %; skipattujen tasojen määrä%  
NARROW=NARROW;%1\$s;%2\$s

#KEYWORD: Käyttäjän syöte vapaatekstihakukenttään  
#KEYWORD ; hakusana  
KEYWORD=KEYWORD;%1\$s

#CLEARKEYWORD: Käyttäjä poisti hakusanan hausta  
#CLEARKEYWORD  
CLEARKEYWORD=CLEARKEYWORD

#-----  
# Suositteletagit  
#-----

#RECOMDOC: Suositellut dokumentit  
#RECOMDOC ; Pilkulla erotettu lista dokumenttiureista  
RECOMDOC=RECOM;%1\$s

#RECOMFOLLOW: Valittu suositus  
#RECOMFOLLOW ; Valittu dokumenttiuri  
RECOMFOLLOW=RECOMFOLLOW;%1\$s

## Liite F Transaktiokin esimerkki

124031 ; 2007-04-08 18:21:58,984 ; SELECTCAT ; KC6.15 ; Nykytaiteen museo Kiasma  
124047 ; 2007-04-08 18:21:59,000 ; PAGE ; Browsing  
124781 ; 2007-04-08 18:21:59,734 ; NMBGROUPS ; 1  
125078 ; 2007-04-08 18:22:00,031 ; LOADTIME ; Browsing ; 1031  
125078 ; 2007-04-08 18:22:00,031 ; RESVIEW ; grouped  
125093 ; 2007-04-08 18:22:00,046 ; RESDOC ; 153  
129625 ; 2007-04-08 18:22:04,578 ; SELECTCAT ; KC6.5 ; Ateneumin taidemuseo  
129640 ; 2007-04-08 18:22:04,593 ; PAGE ; Browsing  
130406 ; 2007-04-08 18:22:05,359 ; NMBGROUPS ; 1  
130687 ; 2007-04-08 18:22:05,640 ; LOADTIME ; Browsing ; 1047  
130687 ; 2007-04-08 18:22:05,640 ; RESVIEW ; grouped  
130687 ; 2007-04-08 18:22:05,640 ; RESDOC ; 1702  
144984 ; 2007-04-08 18:22:19,937 ; RESVIEW ; timeline  
144984 ; 2007-04-08 18:22:19,937 ; PAGE ; Browsing  
145078 ; 2007-04-08 18:22:20,031 ; LOADTIME ; Browsing ; 94  
145078 ; 2007-04-08 18:22:20,031 ; RESVIEW ; timeline  
145078 ; 2007-04-08 18:22:20,031 ; RESDOC ; 1702  
151156 ; 2007-04-08 18:22:26,109 ; TIMELINERESDOC ; 100 ; 1702  
151172 ; 2007-04-08 18:22:26,125 ; TIMELINELOADTIME ; 5485 ; true  
225718 ; 2007-04-08 18:23:40,671 ; SELECTDOC ; http://kulttuurisampo.fi/  
annotaatio#Instance-maalauk-at-A\_I\_30-1160408414574  
225718 ; 2007-04-08 18:23:40,671 ; PAGE ; ShowDetails  
231734 ; 2007-04-08 18:23:46,687 ; LOADTIME ; ShowDetails ; 6000  
239562 ; 2007-04-08 18:23:54,515 ; PAGE ; ShowDetails  
240609 ; 2007-04-08 18:23:55,562 ; LOADTIME ; ShowDetails ; 1047  
242797 ; 2007-04-08 18:23:57,750 ; PAGE ; ShowDetails  
243500 ; 2007-04-08 18:23:58,453 ; LOADTIME ; ShowDetails ; 703  
245297 ; 2007-04-08 18:24:00,250 ; PAGE ; ShowDetails  
246031 ; 2007-04-08 18:24:00,984 ; LOADTIME ; ShowDetails ; 734  
252078 ; 2007-04-08 18:24:07,031 ; PAGE ; Browsing  
252125 ; 2007-04-08 18:24:07,078 ; LOADTIME ; Browsing ; 47  
252125 ; 2007-04-08 18:24:07,078 ; RESVIEW ; timeline  
252125 ; 2007-04-08 18:24:07,078 ; RESDOC ; 1702  
257203 ; 2007-04-08 18:24:12,156 ; TIMELINERESDOC ; 100 ; 1702  
257203 ; 2007-04-08 18:24:12,156 ; TIMELINELOADTIME ; 4453 ; true  
276906 ; 2007-04-08 18:24:31,859 ; SELECTDOC ; http://kulttuurisampo.fi/  
annotaatio#Instance-maalauk-at-A\_II\_911-1160408389331  
276906 ; 2007-04-08 18:24:31,859 ; PAGE ; ShowDetails  
277593 ; 2007-04-08 18:24:32,546 ; LOADTIME ; ShowDetails ; 687  
290750 ; 2007-04-08 18:24:45,703 ; TIMELINERESDOC ; 100 ; 1702  
290750 ; 2007-04-08 18:24:45,703 ; TIMELINELOADTIME ; 5578 ; true  
294937 ; 2007-04-08 18:24:49,890 ; SELECTDOC ; http://kulttuurisampo.fi/  
annotaatio#Instance-maalauk-at-A\_IV\_4279-1160403633286  
294937 ; 2007-04-08 18:24:49,890 ; PAGE ; ShowDetails  
295531 ; 2007-04-08 18:24:50,484 ; LOADTIME ; ShowDetails ; 594  
298078 ; 2007-04-08 18:24:53,031 ; PAGE ; ShowDetails  
298547 ; 2007-04-08 18:24:53,500 ; LOADTIME ; ShowDetails ; 469  
299640 ; 2007-04-08 18:24:54,593 ; PAGE ; ShowDetails  
300406 ; 2007-04-08 18:24:55,359 ; LOADTIME ; ShowDetails ; 766  
305218 ; 2007-04-08 18:25:00,171 ; PAGE ; Browsing

305312 ; 2007-04-08 18:25:00,265 ; LOADTIME ; Browsing ; 94  
305312 ; 2007-04-08 18:25:00,265 ; RESVIEW ; timeline  
305328 ; 2007-04-08 18:25:00,281 ; RESDOC ; 1702  
310859 ; 2007-04-08 18:25:05,812 ; TIMELINERESDOC ; 100 ; 1702  
310859 ; 2007-04-08 18:25:05,812 ; TIMELINELOADTIME ; 4344 ; true  
311047 ; 2007-04-08 18:25:06,000 ; EXPANDNODE ; KC0.6 ; teos  
311047 ; 2007-04-08 18:25:06,000 ; PAGE ; Browsing  
311093 ; 2007-04-08 18:25:06,046 ; LOADTIME ; Browsing ; 46  
311093 ; 2007-04-08 18:25:06,046 ; RESVIEW ; timeline  
311093 ; 2007-04-08 18:25:06,046 ; RESDOC ; 1702  
312000 ; 2007-04-08 18:25:06,953 ; EXPANDNODE ; KC0.10 ; Kuvataideteos  
312000 ; 2007-04-08 18:25:06,953 ; PAGE ; Browsing  
312031 ; 2007-04-08 18:25:06,984 ; LOADTIME ; Browsing ; 31  
312062 ; 2007-04-08 18:25:07,015 ; RESVIEW ; timeline  
312062 ; 2007-04-08 18:25:07,015 ; RESDOC ; 1702  
319234 ; 2007-04-08 18:25:14,187 ; SELECTCAT ; KC0.12 ; Maalaus  
319281 ; 2007-04-08 18:25:14,234 ; PAGE ; Browsing  
319984 ; 2007-04-08 18:25:14,937 ; LOADTIME ; Browsing ; 703  
319984 ; 2007-04-08 18:25:14,937 ; RESVIEW ; timeline  
319984 ; 2007-04-08 18:25:14,937 ; RESDOC ; 1090  
322375 ; 2007-04-08 18:25:17,328 ; TIMELINERESDOC ; 100 ; 1090  
322390 ; 2007-04-08 18:25:17,343 ; TIMELINELOADTIME ; 1188 ; true  
337250 ; 2007-04-08 18:25:32,203 ; SELECTDOC ; [http://kulttuurisampo.fi/annotaatio#Instance-maalaus-at-A\\_III\\_2146-1160477082775](http://kulttuurisampo.fi/annotaatio#Instance-maalaus-at-A_III_2146-1160477082775)  
337250 ; 2007-04-08 18:25:32,203 ; PAGE ; ShowDetails  
337734 ; 2007-04-08 18:25:32,687 ; LOADTIME ; ShowDetails ; 484  
344281 ; 2007-04-08 18:25:39,234 ; PAGE ; ShowDetails  
344828 ; 2007-04-08 18:25:39,781 ; LOADTIME ; ShowDetails ; 547