

Tietotekniikan maisteriohjelma

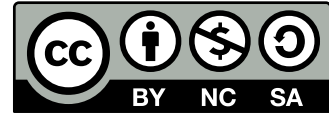
Promootiosampo

Järjestelmä suomalaisen promootioperinteen kuvaamiseen, julkaisemiseen
ja tutkimiseen semanttisessa webissä

Patrik Boman

© 2024.

Tämä teos on lisensoitu [Creative Commons](#) “Nimeä-EiKaupallinen-JaaSamoin 4.0 Kansainvälinen” -
käyttöluvalla.



Tekijä Patrik Boman

Työn nimi Promootiosampo — Järjestelmä suomalaisen promootioperinteen kuvaamiseen, julkaisemiseen ja tutkimiseen semanttisessa webissä

Koulutusohjelma Tietotekniikka

Pääaine Computer, Communication and Information Sciences

Työn valvoja Prof. Eero Hyvönen

Työn ohjaajat FM Heikki Rantala, DI Annastiina Ahola

Päivämäärä 20.5.2024

Sivumäärä 53+4

Kieli suomi

Tiivistelmä

Helsingin yliopiston filosofisen tiedekunnan 100. promootion juhluvuoden kunniaksi rakennettiin Promootiosampo sovellus, jolla voi tarkastella tietoja Helsingin yliopiston 380 vuotta jatkuneesta elävästä kulttuuriperinteestä ja siihen osallistuneista henkilöistä semanttisessa webissä. Promootiosampo on luotu Sampo-UI-järjestelmään perustuvan Akatemiasammon pohjalta, ja se hyödyntää sekä avointa linkitettyä dataa että käsin avoimista lähteistä sekä kirjallisuudesta louhittua tietoa promootioihin liittyvän tiedon kuvaamiseen. Näiden tietojen pohjalta oli mahdollista rakentaa ja julkaista monipuolista ja helposti analysoitavissa olevaa tietoa sisältävä palvelu suomalaisen promootioperinteen tutkimiseen.

Promootiosammon yhteen kerätty ja organisoitu data auttaa ymmärtämään perinteen kehitystä ja antaa mahdollisuuden tutkia osanottajia, heidän välisiä yhteyksiä sekä promootioperinteen historiaa uusilla tavoilla, joihin organisoimaton data ei kykene. Promootiosampo on helposti laajennettavissa oleva sovellus, johon on mahdollista lisätä dataa uusista promootioista sekä täydentää vanhoja historiatietoja tarpeen vaatiessa.

Avainsanat semanttinen web, sampo, fasettihaku, digitaaliset ihmistieteet, linkitetty data

Author Patrik Boman

Title Confermentsampo — An application for describing, publishing and researching Finnish conferment tradition in semantic web

Degree programme Computer Science

Major Computer, Communication and Information Sciences

Supervisor Prof. Eero Hyvönen

Advisors MSc (Tech.) Annastiina Ahola, MSc Heikki Rantala

Date 20 May 2024

Number of pages 53+4

Language Finnish

Abstract

ConfermentSampo was built in honor of the hundredth conferment ceremony of the Faculty of Philosophy of the University of Helsinki. It is a semantic web application that can be used to browse information about the history of this 380 year tradition and of the people who participated in it. ConfermentSampo was created based on the Sampo-UI software and the data of the AcademySampo service. ConfermentSampo uses open linked open data and handpicked information based on open data sources and literature to build a service that contains varied information relating to the Finnish conferment tradition in an easily analysable format.

The organised data of ConfermentSampo helps to understand the evolution of the tradition and yields an opportunity to examine the participating entities and their connections, and the history of the tradition itself in new ways that are impossible with unorganised data. ConfermentSampo is an easily expandable application to which information can be added regarding data of future conferments and newly discovered older historical knowledge as necessary.

Keywords semantic web, sampo, facet search, digital humanities, linked data

Esipuhe

Kiitokset diplomityöni valvojalle professori Eero Hyvöselle ja ohjaajille Annastiina Aholalle ja Heikki Rantalalle.

Kiitokset myös Helsingin yliopiston promootiotoimikunnalle yhteistyöstä Promootiosammon luonnissa, Aalto yliopiston tietotekniikan laitokselle työn rahoittamisesta sekä CSC Oy:lle Promootiosammon laskennallisen infrastruktuurin tarjoamisesta.

Haluan myös kiittää Semanttisen laskennan tutkimusryhmän jäseniä mainiosta työympäristöstä ja ideoista Promootiosammon kehittämiseksi.

Espoo, 20.5.2024

Patrik Boman

Sisällysluettelo

Tiivistelmä	3
Tiivistelmä (englanniksi)	4
Esipuhe	5
Sisällysluettelo	6
Lyhenteet	8
1 Johdanto	9
1.1 Työn tausta	9
1.2 Tutkimuskysymykset	10
1.3 Diplomityön rakenne	10
2 Digitaaliset ihmistieteet ja linkitetty data	11
2.1 Digitaaliset ihmistieteet	11
2.1.1 Aineeton kulttuuriperintö	11
2.1.2 Helsingin yliopiston filosofisen tiedekunnan promootioperinne	11
2.1.3 Linkitetty data ja kulttuuriperintö maailmalla	12
2.2 Semanttinen web	14
2.2.1 Resource Description Framework	14
2.2.2 Tietämysgraafit ja linkitetty data	15
2.2.3 Linkitetyn datan julkaiseminen	15
2.2.4 Semanttiset tietomallit ja ontologiat	19
2.3 Semanttisen webin käyttöliittymät	19
3 Taustajärjestelmät	23
3.1 Sampo-malli	23
3.1.1 Sampo-mallin kuusi periaatetta	23
3.1.2 Sampo-UI ja sen rakenne	24
3.2 Linkitetyn datan julkaisu	26
3.3 Akatemiasampo	27
3.4 Tulokset	28
4 Työn toteutus	29
4.1 Promootiosammon suunnitelma	29
4.2 Promootiosammon data ja sen keräys	30
4.3 Promootiosampo-portaali	34
4.3.1 Promootiosammon etusivu	34
4.3.2 Promootionäkymät	35
4.3.3 Henkilönäkymä	36
4.3.4 Promootiotapahtumat ja -roolit	36
4.4 SPARQL-palvelupiste	37

4.5	Promootiosammon tietomalli	38
5	Tulosten arviointi	41
5.1	Tulokset palvelun näkökulmasta	41
5.2	Promootiosammon tieteelliset ja kulttuurilliset hyödyt	41
6	Jatkokehitys	43
6.1	Tietokanta	43
6.2	Käyttöliittymä	44
6.3	Promootiosammon datan validointi	45
7	Yhteenveto	47
	Viitteet	48
A	Promootioiden tietomalli	54
B	Henkilöiden tietomalli	57

Lyhenteet

CC	Creative Commons
CIDOC CRM	CIDOC Conceptual Reference Model
CSV	Comma separated values
DC	Dublin Core
HTML	Hypertext Markup Language
JSON	Javascript Object Notation
LDF	Linked Data Finland
NER	Named Entity Recognition
scprs:	Lyhenne URI:lle http://ldf.fi/schema/promootiosampo/ Promootiosammon RDF-tiedostoissa
schema:	Lyhenne URI:lle http://schema.org/ RDF-tiedostoissa
SeCo	Semantic Computing Research Group – Semanttisen laskennan tutkimusryhmä
SPARQL	SPARQL Protocol and RDF Query Language
RDF	Resource Description Framework
regex	regular expression – säännöllinen lauseke
Turtle	Terse RDF triple language
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium

1 Johdanto

1.1 Työn tausta

Helsingin yliopiston filosofisen tiedekunnan 100. promootiota juhlistettiin vuonna 2023. Juhlavuoden kunniaksi haluttiin luoda Promootiosampo, jossa Semanttisen laskennan tutkimusryhmän¹ (SeCo) luoman Akatemiasammon²[34][33] kaltaisesti käyttäjät voisivat hakea tietoa Helsingin yliopiston filosofisen tiedekunnan promootioihin osallistuneista henkilöistä.

Helsingin yliopisto perustettiin vuonna 1640 Turun Kuninkaallisena Akatemiana [1] ja vuodesta 1643 alkaen yliopistolla on ollut perinteenä juhlia promootiojuhlia kolmen tai neljän vuoden välein muodostaen perinteen, joka on jatkunut lähes katkeamattomana 380 vuoden ajan. Promootioperinne sai oletettavasti alkunsa Italiassa Bolognan yliopistossa jo 1200-luvulla [20] ja levisi sieltä vähitellen ympäri Euroopan, rantautuen Suomeen Turun Kuninkaallisen Akatemian perustamisen myötä. Maistereiden promovointi tekee suomalaisista promootioista maailmanlaajuisesti ainutlaatuisia, sillä ulkomaisten yliopistojen promootioperinteissä ainoastaan tohtorien promovointi on säilynyt nykypäivään asti. Helsingin yliopisto on yksi maistereiden promootiope-
rinnettä harjoittavista yliopistoista ja perinne on tärkeä osa Suomen akateemista kulttuuria. Maistereiden promovointi on listattu Suomen Elävän perinnön kansalliseen luetteloon, mikä on osa Suomen UNESCO:n aineettoman kulttuuriperinnön suojelua koskevan yleissopimuksen täytäntöönpanoa [35]. Tältä listalta Maistereiden promovointi on ehkä mahdollista lisätä tulevaisuudessa UNESCO:n aineettoman kulttuuriperinnön listalle [32].

Promootiosammosta haluttiin luoda Akatemiasammon kaltainen avoimeen linkitettyyn dataan perustuva portaali, jonka kautta käyttäjä voi hakea tietoa promootioista ja promootioihin osallistuneista henkilöistä, sekä luoda näistä tiedoista erilaisia data-analyyssejä ja johtopäätöksiä. Promootiosammon data on peräisin aiheeseen liittyvästä kirjallisuudesta sekä julkisista lähteistä, kuten promootioihin liittyvistä digitoituista teksteistä sekä Akatemiasammon ja Ylioppilasmatrikkeleiden³ datasta.

Akatemiasammon Ylioppilasmatrikkeleihin perustuvalla tiedolla on suuri rooli promootiosammon henkilödatassa, koska suurin osa promovoiduista maistereista ja tohtoreista ovat aikanaan valmistuneet myös ylioppilaiksi. Alusta alkaen Promootiosampoa luotaessa on punnittu sen roolia Akatemiasammon rinnalla; olisiko se parempi toteuttaa joko Akatemiasammon lisäosana tai itsenäisenä palveluna. Promootiosammon rooli itsenäistyi projektin edetessä Akatemiasammosta erilliseksi palveluksi. Tämä puolestaan luo mahdollisuuden rikastaa Akatemiasammon dataa tulevaisuudessa Promootiosampoon kerätyllä tiedolla. Promootiosammon sisältämä jäsennetty data promootioista ja niihin liittyvistä henkilöistä on hyödyllistä esimerkiksi historian- sekä sukututkimuksessa. Promootiosampo auttaa myös ylläpitämään suomalaisen promootioperinteen aineetonta kulttuuriperintöä tekemällä promootioperinteeseen liittyvästä tiedosta sekä helposti löydettävää että helppoa selata ja täydentää.

¹<https://seco.cs.aalto.fi/>

²<https://akatemiasampo.fi/fi/>

³<https://ylioppilasmatrikkeli.fi/> ja <https://ylioppilasmatrikkeli.fi/1853-1899/>

1.2 Tutkimuskysymykset

Tämän diplomityön tutkimuskysymykset ovat:

- Miten promootiotapahtumista kertovaa dataa voidaan mallintaa?
- Miten tällaista dataa on mahdollista julkaista ihmisten ja verkkopalveluiden käytettäväksi?
- Miten Promootiosampo voisi edesauttaa promootioperinteen ja historiallisten henkilöiden tutkimusta?

1.3 Diplomityön rakenne

Diplomityössä perehdytään Promootiosammon rakentamisprosessiin ja sitä ympäröivään tietoon. Työ jakautuu seitsemään kappaleeseen, joista ensimmäisessä esiteltiin tutkimuskysymykset ja yleistä tietoa Promootiosammon taustamotivaatioista.

Toisessa kappaleessa käsitellään digitaalisia ihmistieteitä (digital humanities) ja kulttuuriperintöä aihealueina. Kappaleessa myös tutustutaan työssä hyödynnettyihin semanttisen webin teknologioihin sekä semanttisen webin käyttöliittymiin.

Kolmannessa kappaleessa tutustutaan Promootiosammon taustajärjestelmiin. Siinä käsitellään Sampo-UI-alustaa, linkitetyn datan julkaisua, sekä Promootiosammon edeltäjää, Akatemiasampo-palvelua.

Neljännessä kappaleessa käsitellään työn toteutusta, joka jakautui kahteen vaiheeseen, Promootiosammon datan keräykseen ja käsittelyyn, sekä Promootiosampo-palvelun luomiseen.

Viidennessä kappaleessa käsitellään työn tuloksia ja arvioidaan kuinka hyvin Promootiosampo onnistui palveluna sekä sen merkitystä tieteelle ja kulttuurille.

Kuudennessa kappaleessa pohditaan mahdollisuuksia Promootiosammon jatkokehitykseen ja parantamiseen eri osa-alueilla.

Seitsemännessä kappaleessa vedetään yhteen diplomityöhön ja Promootiosampoon liittyviä ajatuksia.

2 Digitaaliset ihmistieteet ja linkitetty data

Tässä kappaleessa perehdytään digitaalisten ihmistieteiden aineettoman sekä aineellisen kulttuuriperinnön osa-alueisiin. Tämän lisäksi kappaleessa tutustutaan semanttisen webin Promootiosammossa keskeisiin käsitteisiin sekä semanttisen webin käyttöliittymiin.

2.1 Digitaaliset ihmistieteet

Digitaaliset ihmistieteet tutkimusalueena sijaitsee kulttuurin ja teknologian risteyksessä [5]. Digitaalisia ihmistieteitä voi yleisesti kuvailla tietokoneiden laskentatehoa hyödyntäviksi ihmistieteiksi. Tutkimusalan rajat ovat häilyviä alan jatkuvan kehityksen sekä ristiriitaisten tulkintojen takia siitä, mitä digitaaliset ihmistieteet ylipäättänsä tarkoittavat. Yksi digitaalisten ihmistieteiden kasvava osa-alue on aineellisen ja aineettoman kulttuuriperinnön tutkiminen kokoelmatietojen avulla. Museoiden, kirjastojen ja muiden lähteiden avoimesti saatavilla oleva data tekee kulttuuriperinnöstä otollisen kohteen tutkimukselle, jossa juuri tietokoneiden laskentatehoa on mahdollista hyödyntää ihmistieteiden kysymysten ratkaisussa [25].

2.1.1 Aineeton kulttuuriperintö

Aineeton kulttuuriperintö terminä tarkoittaa elävän kulttuurin ilmenemismuotoja ja perinteitä, jotka määrittelevät tai ovat muuten tärkeä osa erilaisten yhteisöjen identiteettiä [36]. Aineettoman ja aineellisen kulttuuriperinnön erottaa toisistaan kulttuuriperinnön fyysisyys. Aineelliseen kulttuuriperintöön voi kuulua muun muassa taideteoksia, rakennuksia ja esineistöä. Aineeton kulttuuriperintö taas muodostuu juhlista ja tapahtumista, perinteistä ja toimintatavoista sekä mahdollisesti toimintaan liittyvistä esineistä, jotka siirtyvät sukupolvelta toiselle pääasiallisesti sosiaalisen kanssakäynnin kautta.

Maisteripromootion perinne on osa suomalaista aineetonta kulttuuriperintöä ja on listattu Opetusministeriön ylläpitämään Elävän perinnön kansalliseen luetteloon⁴. Perinteen paikka tällä listalla on merkittävä, koska promootioperinnettä on mahdollista ehdottaa lisättäväksi tältä luettelolta UNESCO:n aineettoman kulttuuriperinnön listalle, jolla on tunnustettu maailmanlaajuisesti eri kulttuurien huomattavimpia perinteitä [32].

2.1.2 Helsingin yliopiston filosofisen tiedekunnan promootioperinne

Suomen promootioperinteistä tekee ainutlaatuisen se, että Suomi on maailman ainoa maa, joka yhä jatkaa maistereiden promovoimista. Muissa ulkomaisissa yliopistoissa promootioperinne saattaa jatkua tohtori-promootioina, mutta suomalaisia yliopistoja

⁴https://wiki.aineetonkulttuuriperinto.fi/wiki/El%C3%A4v%C3%A4n_perinn%C3%B6n_kansallinen_luettelo/valitut/fi

lukuunottamatta muut yliopistot ovat luopuneet maisteripromootiosta vuosisatojen saatossa⁵.

Promootiosammossa mallinnetut Helsingin yliopiston filosofisen tiedekunnan promootiot ovat aikojen saatossa monipäiväisiksi juhliksi kehittynyt akateeminen perinne. Nykymuodossaan perinne alkaa toukokuun kolmantenatoista päivänä eli Floran päivänä ja jatkuu myöhemmin kolmipäiväisenä viikonlopun yli kestävästä tarkan protokollan säätelemänä juhlanä, jossa valmistuville maistereille ja tohtoreille annetaan heidän arvonsa tunnusmerkit.

Varhaiset maisteriksi promovoidut henkilöt asettuivat usein kirkollisvirkoihin kuten papeiksi, mutta myöhemmin Venäjän vallan aikana valmistuvat maisterit alkoivat suuntautua monipuolisemmin eri yhteiskunnan aloille kuten Suomen suuriruhtinaskunnan keskushallinnon virkoihin, jonne Venäjä tarvitsi niin paljon koulutettuja virkamiehiä, että se laati asetuksen mahdollistaen virkamiespätevyyden saamisen ilman filosofian maisterin tutkintoa. [20]

Promootiojuhlan jokainen päivä sisältää tarkkaan suunniteltuja ohjelmanumeroita tanssiaisista huviretkiin sekä juhlaillallisiin. Promootion sujuvasta toteutumisesta vastaa muun muassa promootiotoimikunta, joka koostuu promootion järjestelyä vapaaehtoistyönä tekevistä maisteri- ja tohtoripromovendeistä⁶.

Promootiotapahtumien toteutumista varten perinteeseen kuuluu myös monet roolit, joiden tehtäviin kuuluu promootion eri osa-alueiden tehtävien suorittaminen. Näistä rooleista ja niissä esiintyneistä ihmisistä on säilynyt hyvin tietoa koko perinteen historian ajalta ja yksi Promootiosammon tärkeimmistä datajoukoista taulukoi juuri Suomen akateemisten ihmisten rooleja promootiotapahtumissa monen vuosisadan ajalta.

2.1.3 Linkitetty data ja kulttuuriperintö maailmalla

Promootiosampo on harvinaislaatuinen tapa käyttää semanttista webiä hyväksi aineettoman kulttuuriperinnön taltioimisessa. Aalto- ja Helsingin yliopistoissa luotuja Sampoja[27] lukuunottamatta suomenkieliset linkitettyyn dataan perustuvat kulttuuriperintöä käsittelevät sovellukset ovat harvinaisia. Maailmanlaajuisesti on kuitenkin olemassa sovelluksia, joissa semanttisen webin teknologioita hyödynnetään aineettoman kulttuuriperinnön taltioimiseen ja visualisoimiseen.

Dive into intangible cultural heritage⁷ on UNESCO:n semanttisen webin teknologioita hyödyntävä aineettoman kulttuuriperinnön projekti. Sen tietämysgraafi sisältää yli 700 aineettoman kulttuuriperinnön instanssia⁸, jotka ovat linkitetty lähes tuhanteen käsitteeseen, jotka esiintyvät joko UNESCO Thesauruksessa tai käsitteitä täydentävässä sanastossa.

Dive into intangible cultural heritage projektin tietämysgraafia on visualisoitu projektisivulla viidellä eri tavalla, joista yksi on kuvassa 1 nähtävä Constellation

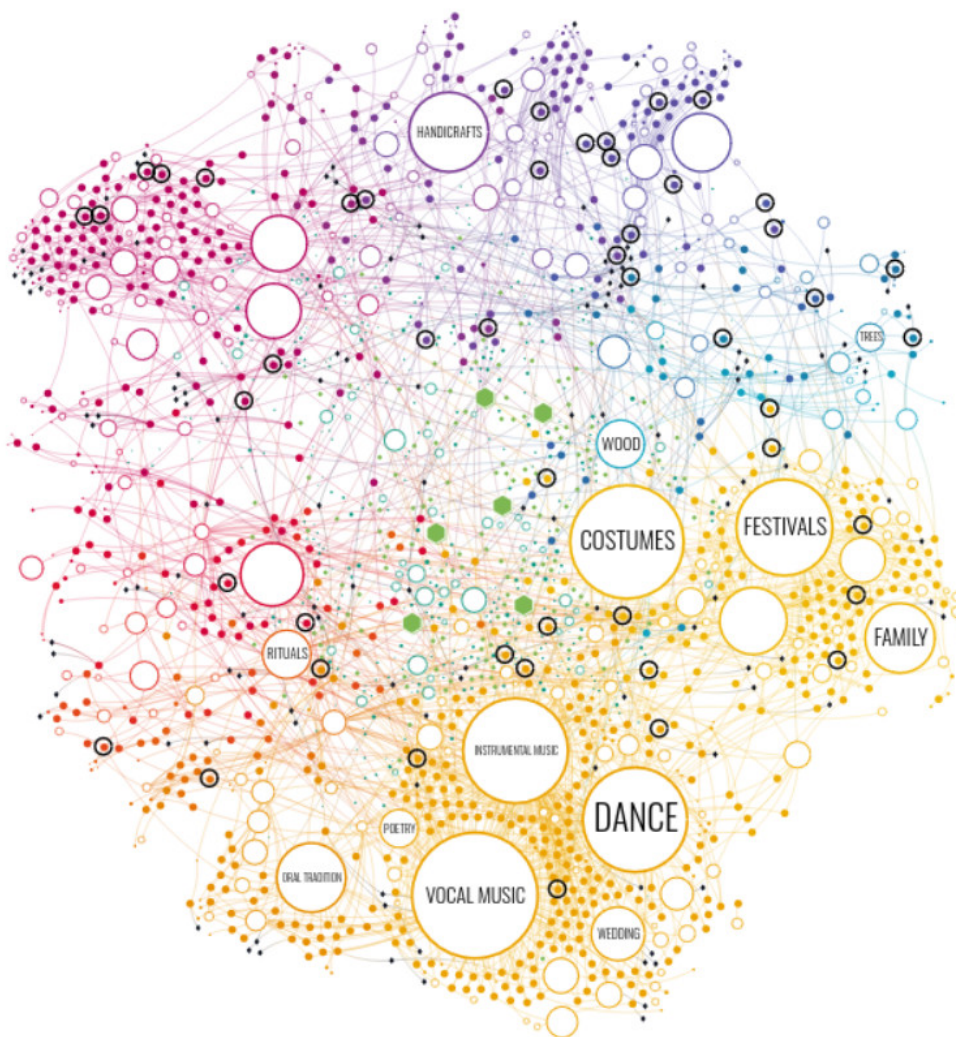
⁵https://wiki.aineetonkulttuuriperinto.fi/wiki/Maistereiden_promovoiminen

⁶<https://www.helsinki.fi/fi/projektit/promootion-riemuvuosi/promootion-historiaa/roolit-ja-rituaalit>

⁷<https://ich.unesco.org/en/dive>

⁸<https://ich.unesco.org/en/lists>

visualisaatio. Visualisaatiossa solmua klikkaamalla on mahdollista korostaa siihen käsitteiden kautta liittyvät muut solmut. Jos solmu on aineettoman kulttuuriperinnön kohde, voi solmun avata ponnahdusrutuun, jossa aineettoman kulttuuriperinnön kohde esitellään muun muassa lyhyiden artikkeleiden ja videotallenteiden muodossa.



Kuva 1: UNESCO:n 'Dive into intangible cultural heritage' projektin constellation visualisointi verkkosivulta <https://ich.unesco.org/en/dive>

Aineettoman kulttuuriperinnön sovellukset semanttisessa webissä ovat harvinaisia, mutta aineellisen kulttuuriperinnön sovellukset mukaan lukien on mahdollista havaita, että linkitettyä dataa voi hyödyntää monenlaisissa kulttuuriperintöä kartoittavissa ja taltioivissa sovelluksissa. Esimerkiksi Tommy Messaudi et al. [40] loivat ontologisen mallin tallentamaan historiallisen rakennuksen konservoinnin eri tiloja tarkoituksenaan luoda mahdollisimman kokonaisvaltainen malli, jossa yhdistetään analyttisen datan eri lähteitä kuten kolmiulotteista dataa ja valokuvia eri ajanjaksoihin siten, että niistä on mahdollista muodostaa esimerkiksi malli rakennuksen kunnan heikkenemisestä

ajan myötä.

Muita esimerkkejä semanttisen webin teknologioista kulttuuriperinnön taltionnissa on Amsterdamin museo⁹, jonka kokoelmien digitaalinen metadata ja tekijöiden tiedot on muunnettu viiden tähden linkitetyksi dataksi¹⁰. Tämä data on saatavilla osana Europeana Semantic Layer -alustaa. [11]

Europeana on Eurooppalainen digitaalinen kirjasto, johon on kerätty Eurooppalaiseen kulttuuriperintöön liittyvää metadataa lukuisista Euroopan kulttuuri-instituutioista. Tämä digitaalinen kirjasto hyödyntää muun muassa Dublin Core standardia, sekä CIDOC CRM -tietomallia sen alla olevissa eri kokoelmissa. [50]

2.2 Semanttinen web

Semanttinen web muodostuu linkitetystä datasta[22]. Eräitä sen peruskäsitteistä ovat tietämysgraafit, ontologiat sekä linkitetyn datan eri teknologiat. Monet näistä käsitteistä ovat vahvasti läsnä suuressa osassa semanttisen webin eri sovelluksia ja järjestelmiä Promootiosampo mukaan lukien.

2.2.1 Resource Description Framework

Resource Description Framework (RDF) on yleinen linkitetyn datan tietämysgraafissa käytetty malli. Se hyödyntää URI (Uniform Resource Identifier) merkkijonoja yksikäsitteisesti tunnistettavien resurssien kuvaamiseen [42]. URI-resurssit muodostavat tietämysgraafien solmut ja ne on mahdollista yhdistää toisiinsa linkeillä. Nämä linkit muodostuvat myös URI-merkkijoinoista ja kuvaavat sitä, millainen ominaisuus yhdistää kahta resurssia toisiinsa. RDF-data järjestyy näin kolmikoiksi, jotka sisältävät subjektin, predikaatin ja objektin.

Kolmikot ovat tehokas tapa kuvailla asiayhteyksiä eri resurssien välillä. Niiden avulla on mahdollista luoda tietämysgraafeja, jotka eivät ole ainoastaan ihmisten luettavissa, vaan myös tietokoneiden on mahdollista navigoida ja prosessoida niitä. URI:en lisäksi kolmikot voivat sisältää literaaliarvoja objekteina. Niitä on mahdollista käyttää resurssin ominaisuuksien kuvailemiseen. Näitä ominaisuuksia voivat olla esimerkiksi resurssin nimi, lukumäärä, päivämäärä tai sähköpostiosoite.

RDF-dokumentteja on mahdollista kirjoittaa eri serialisaatioformaateissa riippuen dokumentin käyttötarkoituksesta. Promootiosampo käyttää Turtle formaattia, joka on suunniteltu siten, että ihmisten on helppo lukea sekä kirjoittaa sitä. Kuvasa 2 on nähtävissä esimerkki Turtle tiedoston formaatista, jossa promootioentiteetti `f_promootio_1677` julistetaan kuuluvaksi luokkaan `scprs:Confement`. Entiteetille julistetaan myös muita ominaisuuksia kuten `nimi: 18. promootio (1677)` ja `tapahtumapäivä 27. heinäkuuta 1677`. Muita RDF syntakseja ovat muun muassa RDF/XML, RDF/JSON, N-Triples ja Notation 3 (N3) [17][59].

⁹<http://amsterdammuseum.nl>

¹⁰<https://5stardata.info/en/>


```

<http://ldf.fi/promootiosampo/promotions/f_promootio_1677> a scprs:Conferment ;
  scprs:ConfermentDescription "Vain kahdeksan viihättävää maisteria osallistui promootioon vuonna 1677" ;
  scprs:ConfermentDescriptionSource "Tero Halonen: Promootio - Elävän Yliopistoperinteen Historiaa" ;
  scprs:Image <https://finna.fi/Cover/Show?h=480&source=Solr&id=doria.10024_51658&index=0&size=large> ;
  scprs:ImageAttribution "Kansalliskirjasto Doria, via Finna.fi" ;
  scprs:ImageDescription "Väitöskirja vuodelta 1654, Moraalisesta hyveestä yleisesti, jossa kiitetään Andreas Petraeusta ntmeltä" ;
  scprs:ImageSource <https://finna.fi/Record/doria.10024_51658?sid=2990748186&imgid=1> ;
  scprs:conferrent <http://ldf.fi/promootiosampo/persons/anders_petraeus> ;
  scprs:magisterPrimus <http://ldf.fi/promootiosampo/persons/anders_hasselqvist> ;
  scprs:magisterUltimus <http://ldf.fi/promootiosampo/persons/simon_tolpo> ;
  scprs:promovend 8 ;
  scprs:tapahtumaPalva "1677-07-27" ;
  scprs:university <http://ldf.fi/promootiosampo/university/Kuninkaallinen_Turun_Akatemia> ;
  scprs:year 1677 ;
  scprs:yearEnd 1677 ;
  skos:prefLabel "10. promootio (1677)" .

```

Kuva 2: Turtle notaatiolla kirjoitettuja tietoja Helsingin yliopiston 18. promootiosta.

2.2.2 Tietämysgraafit ja linkitetty data

Tietämysgraafeille (eng. Knowledge Graph) ei ole olemassa yhtä yleisesti hyväksyttyä formaalia määritelmää. Sen sijaan lähteestä riippuen tietämysgraafeilla voidaan tarkoittaa monia semanttisen webin palveluiden keskeisiä komponentteja. Paulheimin [48] mukaan tietämysgraafin vähimmäisvaatimukseen kuuluvat muun muassa se, että se kuvaa olemassa olevia entiteettejä ja niiden välisiä yhteyksiä graafimaisessa muodossa. Färber et al. [12] määrittelevät tietämysgraafit subjektin, predikaatin ja objektin sisältävistä kolmikoista muodostuvina RDF-graafeina. Pujara et al. [53] määrittelevät tietämysgraafit internetistä louhituista faktoista luotuina graafeina.

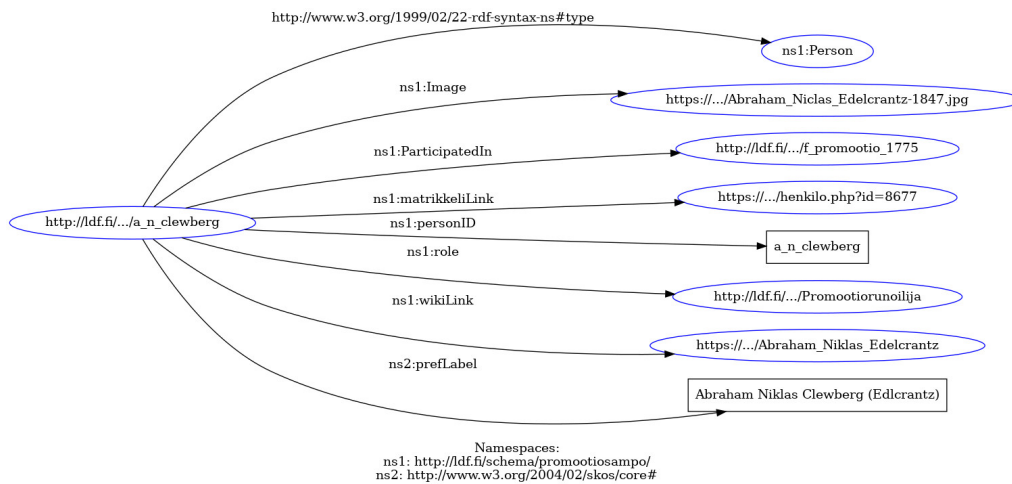
Yleisesti on mahdollista sanoa, että tietämysgraafit muodostuvat toisiinsa linkitetystä datasta. Visualisoidussa muodossa tietämysgraafeja voi kuvailla solmuista ja niitä yhdistävistä linkeistä muodostuvina verkkoina. Kuvassa 3 on Linked Data Finlandin RDF Grapher -palvelulla¹¹ tuotettu yksinkertaistettu graafi osasta Promootiosammon RDF-dataa, joka esittää Abraham Niklas Clewbergistä tunnettuja tietoja Promootiosammonsa. Kuvasta voi nähdä kuinka verkossa henkilöentiteetin solmu yhdistyy linkeillä henkilön ominaisuuksia esittäviin solmuihin ja literaaliarvoihin.

RDF-graafin solmut voivat kuvata joko URI-tunnisteella asiaa, tai sisältää literaaliarvon kuten numeroita, päivämääriä tai lauseita. Graafin linkit puolestaan kuvaavat URI-tunnisteella sitä, minkälainen ominaisuus yhdistää kahta solmua tai solmua ja literaalialia. Semanttisessa webissä tietämysgraafit yleensä noudattavat RDF-mallia, mutta tietämysgraafien monilukuisten määritelmien vuoksi myös muissa formaateissa olevia graafeja, kuten yritysten yksityisiä formaatteja hyödyntäviä graafeja, on mahdollista kutsua tietämysgraafeiksi.

2.2.3 Linkitetyn datan julkaiseminen

World Wide Web Consortium (W3C) suosittelee tiettyjä käytäntöjä linkitetyn datan julkaisemisessa. Yksi näistä käytännöistä on datan julkaiseminen koneiden käytettävissä olevassa muodossa esimerkiksi SPARQL-palvelupisteen kautta. W3C suosittelee myös linkitetyn datan julkaisijaa ilmoittamaan yleisölle uusien tietojoukkojen julkaisusta

¹¹ <https://www.ldf.fi/service/rdf-grapher>



Kuva 3: Yksinkertaistettu visualisointi osasta Promootiosammon dataa.

erilaisten julkaisu- ja tiedotuskanavien sekä sosiaalisten kanavien kautta. Yleisölle on myös hyvä tarjota kanava palautteen antamiselle, jotta esimerkiksi virheellinen data on helpompaa korjata sekä ilmoittaa lisenssi ja tekijänoikeudet, joiden alla data on julkaistu. [64]

Promootiosammon julkaisussa on huomioitu monet edellämaituista W3C:n suosituksista. Osa on integroitu palveluun itseensä, kuten linkki palautelomakkeeseen, mutta erinäiset tiedot esimerkiksi Promootiosammon Creative Commons lisenssistä löytyvät Linked Data Finland¹² (LDF) -palvelun kautta.

Linkitetty data julkaistaan usein verkossa kahdella eri tavalla. Ensimmäinen on linkitetyn datan tietokannan julkaisu RDF serialisoituna tiedostona esimerkiksi Turtle-formaatissa. Tämä on hyödyllinen muoto, mikäli käyttäjä haluaa selata tai käyttää tietoja omassa sovelluksessaan. Toinen tapa on luoda linkitetylle datalle SPARQL-palvelupiste.

SPARQL-protokollaa on mahdollista käyttää haluttujen tietojen löytämiseen RDF-tietokannasta. Protokolla mahdollistaa tämän lisäksi muun muassa kolmikoiden lisäämisen sekä poistamisen tietokannasta, sekä uusien RDF-tiedostojen luomisen alkuperäisen tietokannan tietojen pohjalta.

Yksinkertaisimmillaan SPARQL-hakuja voi käyttää tietämysgraafissa olevien rakenteiden tunnistamiseen [63] (pattern matching), mutta protokolla mahdollistaa myös monimutkaisempia hakuja. Esimerkiksi kuvassa 4 on kuvattu SPARQL-haku, joka etsii Promootiosammon datasta jokaista promootiota varten niissä esiintyneiden promovendien määriä. Haku myös laskee eri promootiovuosina filosofian maistereiksi promovoitujen henkilöiden lukumäärät Akatemiasammon datan avulla ja kuvassa 5 näkyy kyseisen haun antamia rinnakkaisia tuloksia taulukkonäkymässä. Tämänkaltaisten SPARQL-kyselyiden palauttamien tietojen pohjalta rakentuvat kaikki Promootiosammon esiintyvät visualisaatiot.

¹²<https://www.ldf.fi/>


```

SELECT ?label ?value ?count
WHERE {
  {
    SELECT ?label ?value
    WHERE {
      {
        ?id a scprs:Conferment .
        ?id scprs:promovend ?value .
        ?id scprs:year ?year .
        ?id skos:prefLabel ?label .
      } UNION {
        ?id a scprs:Conferment .
        ?id scprs:year ?year .
        ?id skos:prefLabel ?label .
        FILTER NOT EXISTS { ?id scprs:promovend [] . }
        BIND("null" as ?value)
      }
    }
  }
  ORDER BY ?year
} OPTIONAL {
  SELECT DISTINCT ?label (COUNT(?year) as ?precount)
  WHERE {
    {
      ?fm a :g3297450366287827456 ;
        ^(:has_event/:has_part) ?person__id ;
        skos:prefLabel ?abbr .
      BIND(REPLACE(REPLACE(REPLACE(STR(?abbr), "\\).*$", ""), "^.*\\.\"", ""), "FM \\(", "") AS ?year)
      FILTER regex(?year, "[0-9]{4}$")
      ?id a scprs:Conferment .
      ?id scprs:year ?pyear2 .
      ?id skos:prefLabel ?label .
      BIND(STR(?pyear2) AS ?pyear)
      FILTER(STR(?year) = STR(?pyear))
    }
  }
  GROUP BY ?value ?label
  ORDER BY ?year
}
BIND(COALESCE(?precount, "null") AS ?count)
}

```

Kuva 4: SPARQL-kysely, joka etsii muuttujaan ?value promootiosammon datasta tunnettuja promovendien määriä sekä laskee muuttujaan ?count promovendien osallistujalukumääriä Akatemiasammon henkilödatasta.

label	value	count
1. promootio (1643)	10	11
2. promootio (1647)	null	20
3. promootio (1650)	null	15
4. promootio (1653)	null	17
5. promootio (1656)	null	13
6. promootio (1661)	null	20
7. promootio (1664)	null	16
8. promootio (1667)	null	null
9. promootio (1672)	null	16
10. promootio (1677)	8	8
11. promootio (1679)	null	17
12. promootio (1682)	null	22
13. promootio (1685)	null	24
14. promootio (1688)	null	17
15. promootio (1691)	null	19
16. promootio (1694)	24	27
17. promootio (1697)	null	23
18. promootio (1700)	null	25
19. promootio (1703)	null	24
20. promootio (1707)	null	29

Kuva 5: Kuvan 4 SPARQL-kyselyn antamia tuloksia.

2.2.4 Semanttiset tietomallit ja ontologiat

RDF-tiedostot voivat muodostua yksinkertaisimmillaan hyvin yksinkertaisista rakenteista, mutta ajan myötä RDF:n luoman perustan päälle on rakennettu laajennuksia mahdollistamaan erilaisten luokkarakenteiden, rajoitteiden sekä sääntöjen luomisen. Eräitä näistä laajennuksista ovat ontologiat [15] ja tietomallit [49].

Borst [6] kannattaa määritelmää ontologioista jaettujen konseptualisatioiden muodollisina määritelmänä. Konseptualisaatiot ovat abstrakteja mallinnuksia erilaisista ilmiöistä, jotka on luotu tunnistamalla ilmiölle keskeiset elementit [58]. Jaettu konseptualisaatio korostaa, että konseptualisaation on oltava yleisesti hyväksytty, jotta siihen perustuvaa ontologiaa on mahdollista käyttää uudelleen eri tilanteissa. Käytännön termeissä ontologiat ovat tietorakenteita, jotka sisältävät mallin tietyn aihealueen käsitteistä sekä niiden välisistä suhteista.

Ontologiat eroavat monista tietomalleista siten, että ne ovat uudelleenkäytettävissä ja riippumattomia yksittäisestä sovelluksesta. Semanttiset tietomallit sen sijaan usein luotu yksittäistä sovellusta varten ilman odotuksia mallin uudelleenkäytöstä muissa tarkoituksissa. [57] Termien eroista huolimatta niitä käytetään joissain tilanteissa synonyymeinä.

Dublin Core¹³ (DC) ja CIDOC Conceptual Reference Model¹⁴ (CIDOC CRM) ovat tietomalleja, jotka ovat luotu helpottamaan eri verkkopalveluiden datasiilojen sisällön standardisointia [27]. CIDOC CRM on kulttuuriperinnön kuvailemiseen keskittyvä ontologia ja DC on metatietostandardi, joka auttaa kuvaamaan verkossa julkaistavia dokumentteja. Se tarjoaa metatietotermejä RDF-muodossa semanttisessa webissä käytettäväksi. DC on yksi käytetyimpiä ratkaisuja webin resurssien metadatan kuvailemiseen [8].

CIDOC CRM:n sisältämät entiteettityypit sekä niiden väliset suhteet on perustettu monien olemassa olevien kulttuuridatan datamallien luomalle pohjalle [46]. Sampoperheen palveluista CIDOC CRM -ontologiaa hyödyntää esimerkiksi Akatemiasampo, jossa CIDOC CRM:n tapahtumaperusteinen tiedon esittämistapa on keskeisessä osassa tietokannan rakennetta [29]. Käytännön hyötynä CIDOC CRM järjestelmästä on muun muassa mahdollisuus hakea tunnettuja CIDOC -luokkia tietokannasta, mutta haasteena on tehdä palvelun tietokanta CIDOC CRM:n sääntöjen kanssa yhteensopivaksi [38]. Akatemiasammosta poiketen CIDOC CRM -ontologia ei ole käytössä Promootiosammossa, mutta Promootiosampoa olisi mahdollista laajentaa siten, että sen promootiot esitetään CIDOC CRM -ontologian kanssa yhteensopivina tapahtumina. Promootiosammon oma yksinkertaistettu tietomalli on tällä hetkellä lähempänä DC:n dokumenttipohjaista rakennetta, joten muunnos CIDOC CRM:n tapahtumapohjaiseen rakenteeseen vaatisi palvelun tietokannan uudelleenmuotoilua.

2.3 Semanttisen webin käyttöliittymät

Linkitetyn datan ja semanttisen webin selaamiseen ja visualisoimiseen on kehitetty lukuisia erilaisia käyttöliittymiä. Bernasconi et al. [3] ovat kartoittaneet tutkimusartik-

¹³<https://www.dublincore.org/>

¹⁴<https://www.cidoc-crm.org/>

kelissaan useita eri käyttötarkoituksiin soveltuvia linkitettyä dataa hyödyntäviä työkaluja ja käyttöliittymiä. Artikkelissa järjestelmät ovat jaettu viiteen eri pääkategoriaan ja useisiin alakategorioihin.

Ensimmäinen työkalutyypin kategoria on semanttisen datan louhinnan työkalut. Tähän kategoriaan kuuluvia työkaluja käytetään semanttisen datan luomiseen järjestämättömästä datamassasta. Tämän kategorian työkaluja voidaan käyttää esimerkiksi tekstien täydentämiseen Wikipediaan johtavilla linkeillä.

AIDA (Accurate Online Disambiguation of Named Entities) on systeemi, joka hyödyntää YAGO2 tietokantaa nimettyjen entiteettien tunnistamiseen erilaisista teksteistä [65]. YAGO2 on automaattisesti koostettu tietokanta, jonka lähteinä ovat Wikipedia, GeoNames ja WordNet tietokannat [23]. Apache Stanbol [56] on avoimen lähdekoodin ohjelmisto, joka on luotu semanttisen sisällön hallintaa varten. Stanbol rakentuu erilaisia toimintoja tarjoavista komponenteista. DBPedia Spotlight [39] systeemi lisää automaattisesti DBPedia URI:a sille annettuihin tekstidokumentteihin. OpenCalais [7] verkkopalvelu tarjoaa ohjelmointirajapinnan, johon on mahdollista syöttää järjestämättömää tekstiä. Palvelu muuntaa tämän tekstin muun muassa koneoppimisalgoritmeja hyödyntäen RDF-formatoiduiksi entiteeteiksi ja tapahtumiksi. Semiosearh Wikifier [14] on sovellus, joka yhdistää nimettyjä entiteettejä sekä termejä niitä vastaaviin DBPedia entiteetteihin hyödyntäen monia tekniikoita heuristiikoista NER (Named Entity Recognition, nimettyjen entiteettien tunnistaminen) tekniikoihin. GLOBDEF System [44] on Apache Stanbolin kaltaisesti erilaisten dataa parantavien työkalujen hyödyntämiseen tähtäävä kehikko. Tämän systeemin tavoitteisiin kuuluvat muun muassa olemassa olevien dataa parantavien koneistojen integroinnin tukeminen, tuki lisättyjen merkintöjen päikkaansapitävyyden tarkistamiselle sekä valmius skaalautaa suurien datamassojen nopeaan käsittelyyn.

Toinen kategoria on perinteiset visuaalisen informaation etsimistyökalut, mitkä eivät käytä semanttisia teknologioita informaation etsinnässä. Tämän tyyppiset sovellukset hyödyntävät avainsanoja ja metadataa esimerkiksi kirjojen etsimiseen tietokannasta. Tähän kategoriaan kuuluvia palveluja voivat olla muun muassa kauppojen tai yritysten verkkosivut, joilta voi etsiä tuotteita eri avainsanojen avulla.

Kolmas kategoria on semanttisen datan visualisaatiojärjestelmät. Tämän kategorian sovelluksia voidaan käyttää RDF-tiedostoista ja SPARQL-kyselyistä saatujen tietämysgraafien visualisoimiseen. Visualisaatiot jakautuvat usein taulukko- tai graafimuotoisiin toteutuksiin, mutta myös muunlaisia toteutuksia on olemassa.

DBPedian [2] tavoitteena on muuntaa Wikipedian sisältö järjestetyksi dataksi. DBPedian laaja-alaisen kattavuuden takia se on mahdollista linkittää samanaikaisesti moniin massiivisia tietojoukkoja omaaviin palveluihin, joista esimerkkejä ovat US Census, Geonames ja WordNet. Aemoo [45] on semanttisen webin sovellus, jossa avainsanoja hakemalla on mahdollista tutkia eri entiteetteihin liittyvää useista lähteistä koottua tietoa. LODmilla [41] on useiden eri linkitetyn avoimen datan tietojoukkojen selaamista varten luotu semanttisen webin selain. Sen toimintoihin kuuluu muun muassa erilaisten resurssien visualisointi, resurssista ja graafeista tiedon etsiminen, dokumenttien manipulointi sekä selainäkymän tallentaminen ja jakaminen. Tarsier [62] työkalu on kehitetty RDF-graafien 3D visualisointiin. Graafivisualisoinnit on mahdollista jakaa erillisiin kerroksiin helpottaen datan ymmärtämistä ja analysointia.

SPARQLFilterFlow [18] on visuaalinen käyttöliittymä SPARQL-kyselyiden luomiseen, jonka tavoitteena on helpottaa monimutkaisten kyselyiden luomista myös niille henkilöille, joilla ei ole kokemusta SPARQL-kyselykielen käyttämisestä. QueryVOWL [19] on visuaalinen kyselykieli, jonka tavoitteena on mahdollistaa linkitetyn datan tutkiminen myös henkilöille, joilla ei ole kokemusta RDF graafeista tai SPARQL-kyselykielestä. RDF Explorer [61] on QueryVOWL:n tapaan visuaalinen kyselykieli, jonka tavoitteena on mahdollistaa maallikoille RDF-graafien tutkiminen ja kyselyiden luominen. Myös SPARQLBlocks [9] on visuaalinen kieli, jonka tarkoituksena on mahdollistaa käyttäjille SPARQL kyselyiden luominen ilman SPARQL syntaksin tunteesta. Sen tavoitteena on myös toimia mahdollisena väliaskelmana tekstisyntaksin itsensä oppimiselle. GraphBrain [13] on työkalu, jonka tarkoituksena on edesauttaa ja mahdollistaa tietämysgraafien yhteistyöllistä suunnittelua. Se tarjoaa työkaluja visualisoinnin lisäksi myös datan rikastamista ja analyysia varten. Ontodia [43] on verkkotyökalu, joka on luotu RDF tietojoukkojen sekä OWL-ontologioiden visualisoinnista, navigoinnista ja jakamista varten. Työkalu tukee sekä käyttäjän omien tiedostojen, että julkisten SPARQL palvelupisteiden käyttöä. TopBraid Composer on ohjelmistoperhe, jonka sovellukset ovat luotu pääasiallisesti ontologioiden muokkausta varten [47]. FenFire [21] on RDF-selain, joka tarjoaa sekä graafi-, että listanäkymän linkitetyn datan selaamiseen. H-BOLD [51] on työkalu linkitetyn datan selaamista varten ja se tarjoaa erilaisia näkymiä suurien linkitetyn avoimen datan tietojoukkojen visualisoinnista varten.

Neljäs kategoria on semanttiset tiedon täydennysjärjestelmät. Nämä sovellukset lisäävät semanttisia yhteyksiä dokumentteihin joko manuaalisesti, puoliautomaattisesti tai täysin automaattisesti olemassa olevien ontologioiden tai tietokantojen pohjalta.

OmekaS¹⁵ on käyttäjäystävällinen alusta, jota hyödynnetään digitalisoidun kulttuuriperinnön yhdistämiseen muihin online resursseihin[3]. OmekaS on esimerkiksi museoiden käytössä ja se toimii verkkojulkaisualustana, johon on mahdollista lisätä massoittain dataa niin kauan kuin se on formatoitu oikein [37]. AKTiveMedia [10] on systeemi, joka käyttää semanttisen webin teknologioita erilaisten dokumenttien rikastamiseen. Rikastaminen tapahtuu yhteisön voimin ja eri henkilöiden on mahdollista yhdistää dokumentteja tai niiden osia eri ontologioihin tai lisätä selityksiä tekstimuodossa. KIM [52] on alusta, jonka ominaisuuksiin kuuluvat automatisoitu semanttisten selitysten lisääminen dokumentteihin, sekä dokumenttien hallinta. KIM:a on mahdollista käyttää erilaisten selainpäätteiden kautta, joita on mahdollista kehittää KIM:n ohjelmointirajapinnan kautta. Nämä verkkopalvelut yhdistävät itsensä 'KIM Server'iin, joka ylläpitää alustan toimintoja. Thresher [24] on systeemi, jonka kautta käyttäjien on mahdollista käyttää verkkoselainta semanttisen datan erittelemiseen HTML-sivuilta. Työkalun algoritmi yrittää korostaa lisää samankaltaisia rakenteita käyttäjän toimien perusteella. Tämä mahdollistaa sen, että käyttäjät pystyvät erottelmaan verkkosivujen semanttisia rakenteita ilman verkkosivujen kehittäjien apua. Recogito [55] on avoimen lähdekoodin työkalu, jolla on mahdollista lisätä geograafiseen dataan liittyviä merkintöjä kuviin ja tekstidokumentteihin. Työkalu pystyy tunnistamaan paikan nimiä englanninkielisistä tekstidokumenteista itsenäisesti, mutta

¹⁵<https://omeka.org/>

tarvitsee ihmisen kirjoittamaan paikkoihin liittyvät merkinnät.

Viides kategoria on digitaalisten kirjastojen tutkimusjärjestelmät, joihin muun muassa Sampo-UI -sovellukset kuuluvat. Tämän kategorian sovellukset usein erikoistuvat tiettyjen ennalta määrättyjen tietokantojen tutkimiseen ja tarjoavat työkaluja juuri näiden digitaalisten kirjastojen tutkimiseen ja analysointiin.

Sampo-UI [31] on sovelluskehikko, joka helpottaa semanttisten portaalien luomista tarjoamalla käyttöliittymän ja helposti muokattavan rakenteen, missä monet portaalin vaatimat toiminnot ovat luotu kehittäjää varten etukäteen. Yewno Discover [16] on järjestelmä, joka sisältää akateemisista ja tieteellisistä lähteistä koottuja tietoja. Käyttäjä voi tutkia järjestelmän tietoja graafimaisen näkymän avulla ja muodostaa päätelmiä eri käsitteiden välisistä yhteyksistä. Talk to Books¹⁶ [54] on Googlen kehittämä palvelu, joka mahdollistaa kysymysten ja lausuntojen luomisen eri kirjoihin liittyen, minkä jälkeen algoritmi tuottaa vastauksen syötteeseen. Arca [4] on systeemi, jossa tekstikokoelmasta luotua tietämysgraafia on mahdollista tutkia Arcan käyttöliittymän kautta. Käyttöliittymä mahdollistaa muun muassa entiteettien etsimisen graafista sekä graafin itsensä navigoinnin.

Bernasconin et al. [3] tutkimusartikkeli esittelee yhteensä neljäkymmentäseitsemän eri semanttisen webin yhteisön kehittämää järjestelmää. Nämä käyttöliittymät ovat kategorisoitu viiteen pääkategoriaan niiden eri käyttötarkoitusten mukaan, joista erityisesti semanttisen datan visualisaatiojärjestelmät, semanttisen tiedon täydennysjärjestelmät ja digitaalisten kirjastojen tutkimusjärjestelmät käsittelevät semanttisen webin käyttöliittymiä.

Promootiosammon taustajärjestelmä Sampo-UI on yksi tutkimusartikkelin nimeämistä järjestelmistä. Ainutlaatuisen siitä tekee se, että Sampo-UI on ainoa artikkelissa mainittuja käyttöliittymiä, jonka keskeisenä ominaisuutena on uusien semanttisten portaalien luominen. Monet muut käyttöliittymät tarjoavat ominaisuuksia linkiteyn datan visualisointiin, tutkimiseen tai dokumenttien rikastamiseen, mutta ainoastaan Sampo-UI esitellään uusien palveluiden luomiseen soveltuvana kehikkona.

¹⁶<https://books.google.com/talktobooks/>

3 Taustajärjestelmät

Promootiosammon taustalta löytyy erilaisia järjestelmiä, joita ilman sampoa ei olisi ollut mahdollista rakentaa sen nykyisen kaltaisena toteutuksena. Näistä tärkeimmät ovat Sampo-mallia noudattava Sampo-UI kehikko, linkitetyn datan julkaisemiseen käytetty ja Promootiosampoon liittyvää metadatan sisältävä Linked Data Finland¹⁷ (LDF) -sivusto sekä Akatemiasampo, jonka pohjalle suuri osa Promootiosammon henkilödatasta perustuu.

3.1 Sampo-malli

Sampo-mallin keskeinen ajatus on tarjota käyttäjälle monenlaisia eri näkymiä, joiden kautta käyttäjä pystyy tutkimaan palvelun sisältöjä. Sampo-mallin mukaan rakennettu sovellus tarjoaa myös mahdollisuuden rajata näkymien tuloksia eri parametrien mukaan sekä tarjoaa data-analyysi työkaluja tiedon jatkojärjestelyä varten. [31]

Sampo-mallin mukaan julkaistun sovelluksen keskiössä on RDF-tietämysgraafi, joka sisältää sovelluksen keskeiset ontologiat, metadatan sekä linkityksiä ulkoisten palveluntarjoajien saatavilla oleviin palveluihin ja niiden metadataan. Tämä tietämysgraafi voidaan julkaista SPARQL-päätepisteenä muiden semanttisten palveluiden käytettäväksi linkitetyn datan periaatteiden mukaan. [26]

Sampo-mallin sovelluksessa on myös oltava käyttöliittymä, joka mahdollistaa datapalvelun tietojen tutkimisen monista eri perspektiivinäkymistä. Eri perspektiiveissä datapalvelun tiedot esitetään painottaen tietämysgraafin sisältämän datan eri osioita. Muun muassa Promootiosammossa perspektiiveinä toimivat promootiotapahtumat sekä promootioihin osallistuvat ihmiset. Useiden perspektiivien avulla on mahdollista luoda monipuolisia johtopäätöksiä, joita datan analysointi vain yhdestä perspektiivistä ei välttämättä mahdollista. Esimerkiksi promootioihin osallistuneiden kunniantohtoreiden määrää voi olla vaikea laskea promootioperspektiivistä, mutta osallistujaperspektiivin kautta tämä operaatio on suoritettavissa varsin helposti.

Tietojen analysointi on Sampo-malli sovelluksissa keskeinen toiminto, ja se toteutetaan kaksivaiheista filosofiaa käyttäen. Ensin mahdollistetaan sovelluksen tietojen seulonta fasettihakua hyödyntäen ja seuraavaksi mahdollistetaan tietojen analysointi erilaisten perspektiiveihin upotettujen data-analyttisten välilehtien avulla. Nämä välilehdet voivat sisältää esimerkiksi datan visualisointeja erilaisiin graafeihin, histogrammeihin, verkostoihin tai karttanäkymiin.

Datan julkaisusysteemi, käyttöliittymän perspektiivinäkymät ja kaksivaiheinen filtteröinti-analysointi sykli ovat kolme Sampo-mallin sovellusten keskeistä komponenttia, jotka ovat läsnä SeCo:n rakentamissa Sampo sarjan semanttisissa portaaleissa.

3.1.1 Sampo-mallin kuusi periaatetta

Edellisessä kappaleessa mainitut kolme komponenttia on mahdollista määrittellä laajemmin kuutena Sampo-mallin periaatteena[27]. Nämä periaatteet P1-P6 ovat:

¹⁷<https://www.ldf.fi/>

- P1: Tue yhteistyöllistä datan luomista ja julkaisua
- P2: Käytä jaettava avointa ontologiainfrastruktuuria
- P3: Jaa linkitetyn avoimen datan palvelut ja käyttöliittymä selvästi erilleen
- P4: Mahdollista saman datan tarkastelu useista eri perspektiiveistä
- P5: Standardisoi hakuportaalien käyttö filteröi-analysoi kaksiaskeleisella syklillä
- P6: Tue data-analyysiä ja tiedon löytämistä datan tutkimisen lisäksi

Ensimmäiset kolme periaatetta P1-P3 koskevat datapalvelujen kehitystä ja periaatteet P4-P6 koskevat semanttisten portaalien kehitystä.

Periaate P1 perustuu ideaan, että itsenäisistä datasiiloista voidaan aggregoida dataa ja julkaista se globaalisti. Jakamalla ontologiat ja metatietomallit globaalia palvelua on mahdollista hyödyntää edelleen, ja samalla itsenäisten palvelujen data saadaan helposti käytettäväksi kaikille, mikä hyödyttää kaikkia osapuolia.

Periaate P2:n keskeinen teema on käyttää uudelleen olemassa olevien palveluiden infrastruktuuria uusien palveluiden kehityksessä. Tämä helpottaa palveluiden jatkokehitystä sekä toisiinsa yhdistämistä.

Periaate P3 ehdottaa, että linkitetyn datan palvelu ja sen käyttöliittymä tulisi olla erillisiä kokonaisuuksia, joiden yhdistävänä siltana toimii SPARQL-rajapinta. Tämä mahdollistaa muun muassa datapalvelun käytön tutkimuksessa ja ohjelmistokehityksessä.

Periaate P4 kannustaa rakentamaan semanttisen portaalien käyttöliittymiä siten, että sen pohjalla piilevän datapalvelun tietoja on mahdollista käyttää uudelleen erilaisissa palveluissa ilman datapalvelun itsensä muokkaamista.

Periaate P5 pyrkii helpottamaan palvelun käyttöä edistämällä ideaa siitä, että ensin palvelun tarjoamat tiedot rajataan käyttäjää kiinnostavalle alueelle, jonka jälkeen palvelu tarjoaa rajatusta tiedosta visualisointeja sekä analyysejä valmiiksi rakennetuilla työkaluilla.

Periaate P6 kuvastaa tavoitetta kehittää semanttisia portaaleja tarjoamaan yhä edistyneempiä ominaisuuksia. Datapalvelun sisältämän datan tutkimisen lisäksi pyritään tarjoamaan työkaluja, jotka helpottavat datan analysointia ja ongelmien ratkaisua.

3.1.2 Sampo-UI ja sen rakenne

Sampo-UI on SeCo-tutkimusryhmän kehittämä alusta, joka on suunniteltu semanttisten portaalien luomista varten. Sovellus on JavaScript ohjelmisto ja vuodesta 2019 lähtien [31] se on toiminut pohjana uusien Sampo-palveluiden kehityksessä. Se mahdollistaa monipuolisesti muokattavien palveluiden luomisen, jotka pystyvät hyödyntämään SPARQL-päätepisteitä ja erilaisia JavaScript kirjastoja datan visualisointiin ja analysointiin. [31]

Taulukossa 2 on listattu SeCo:n aikaisemmin julkaisemia Sampoja.¹⁸ Vieraillemalla eri sammoissa on helppoa huomata Sampo-UI:n käyttöönoton vaikutus sampojen

¹⁸<https://seco.cs.aalto.fi/applications/sampo/>

visuaalisen ilmeen ja toteutuksen standardisoinnissa. Tämän lisäksi on mahdollista huomata Sampo-UI:n alustan luomisen jälkeen.

Sampo	Aihe	Julkaisuvuosi	Sampo-UI
MuseoSuomi	Museokokoelmat	2004	Ei
Kulttuurisampo	Kulttuuriperintö	2008	Ei
TerveSuomi	Terveystieto	2008	Ei
MatkailuSampo	Kulttuuri ja matkustaminen	2011	Ei
SotaSampo	Sotahistoria	2015	Ei
Norssi alumnit semanttisessa webissä	Historialliset henkilörekisterit	2017	Ei
U.S. Congress Protopographer	Historialliset henkilörekisterit	2017	Ei
BiografiaSampo	Biografiat	2018	Ei
NimiSampo	Paikannimet	2019	Kyllä
SotasurmaSampo	Sotahistoria	2019	Kyllä
Mapping Manuscript Migrations	Esimodernit manuskriptit	2020	Kyllä
Akatemiasampo	Suomen akateemiset henkilöt 1640-1899	2021	Kyllä
Löytösampo	Arkeologia	2021	Kyllä
Sotamuistosampo	Toisen maailmansodan veteraanien videoidut muistelmat	2021	Kyllä
Parlamenttisampo	Parlamentaarinen data	2023	Kyllä
Lakisampo	Lainsäädäntö ja oikeuskäytäntö	2023	Kyllä

Taulukko 2: SeCo:n luomia semanttisia portaaleita.

Sampo-UI:n perustalle rakennetut palvelut sisältävät aloitussivun (en. landing page) ja erilaisia perspektiivisivuja, joilta käyttäjä voi etsiä dataa, käyttää fasettihakumootoria datan seulomiseen, sekä käyttää sampoon sisäänrakennettuja data-analyysitoimintoja tietojen visualisoimiseen. Perspektiivisivuilla listatut dataentiteetit voivat sisältää myös omat instanssisivunsa, joille on mahdollisesti listattu entiteetistä yksityiskohtaista dataa ja muita lisätietoja, joita ei välttämättä ole näytetty itse perspektiivisivujen näkymissä.

Sampo-UI-sovelluksen käyttö alkaa sen aloitussivulta. Sivun sisältää vaihtelevan määrän kortteja, joiden kautta on mahdollista navigoida sovelluksen eri perspektiivisivuille. Sovellus sisältää myös yläpalkin, jonka kautta on mahdollista vaihtaa perspektiivinäkymiä, käyttää muita toimintoja kuten aineiston kattavaa sanahakua, löytää sammon käyttöohjeet ja palautelinkki sekä mahdolliset kielivalinnat.

Perspektiivinäkymät voivat koostua monista erilaisista sammon dataa visualisoivista näkymistä. Näistä yleisin on taulukkonäkymä, jossa perspektiivinäkymän kohdeinstanssit ja niihin liittyvä data on visualisoitu riveihin ja sarakkeisiin siten, että jokainen rivi edustaa yhtä instanssia, ja sarakkeet siihen liittyviä datapisteitä. Instanssia klikkaamalla UI ohjaa käyttäjän instanssisivulle, jossa on kerrottu yksityiskohtaisemmin entiteetin tietoja, joita ei välttämättä ole näytetty perspektiivisivun taulukkonäkymässä. Perspektiivinäkymien visualisoinnit hyödyntävät usein ulkoisia JavaScript-kirjastoja muun muassa erilaisten taulukkojen, graafien, verkkojen ja karttanäkymien luomiseen. Perspektiivisivun luomista varten tarvitaan kaksi tiedostoa: JSON tiedosto, jossa määritellään perspektiivin asetukset ja komponentit, sekä SPARQL-tiedosto, josta perspektiivin komponentit kutsuvat tarvitsemiaan SPARQL-kyselyitä.

Fasettihaku on perspektiivinäkymiin liittyvä sampojärjestelmän ominaisuus, jonka avulla perspektiivinäkymän tuloksia on mahdollista rajata halutuun kriteeriin. Fasettihaku perustuu fasettiluokittelun periaatteeseen, minkä juuret on mahdollista kytkeä jo Aristoteleen taksonomioihin. Ensimmäiset fasettianalyysiin perustuvat luokittelevat kehittyivät vasta 1900-luvulla Shiyali Ramamrita Ranganathan kehittäessä kaksoispistelukuokitusjärjestelmän kirjastojen aineistojen järjestämistä varten. [60] Sampo-UI:ssa fasettihaku toimii SPARQL-sovellusrajapintaa hyödyntäen. SPARQL-kyselyillä on mahdollista hakea sovelluksen tietokannasta tarkoin rajattua tietoa, ja fasettihaun erilaiset säätimet auttavat käyttäjää muokkaamaan näiden SPARQL hakujen parametreja.

Perspektiivinäkymien kaltaisesti instanssinäkymät voivat koostua erilaisista datan visualisoinneista. Näistä visualisoinneista tärkein on taulukkonäkymä, jossa kaikki instanssiin liitetyt tiedot ovat listattu. Taulukkonäkymän lisäksi instanssinäkymissä voi olla esimerkiksi karttanäkymiä, jotka voivat näyttää esimerkiksi paikkainstanssin sijainnin, tai verkkonäkymiä, joissa toistensa kanssa olleet tekemisissä olleet henkilöt visualisoidaan henkilöverkostoksi.

3.2 Linkitetyn datan julkaisu

Sampo-järjestelmien datapalveluita ylläpidetään Linked Data Finland (LDF) -verkkosivuilla. LDF on linkitetyn datan julkaisualusta, jonka kautta ylläpidetään suomalaisia semanttisen webin tietojoukkoja [30].

LDF on kehitetty tarjoamaan palveluita kahta pääasiallista ryhmää varten, jotka ovat datan julkaisijat ja semanttisen webin sovellusten kehittäjät. Datan julkaisussa LDF automatisoi muun muassa SPARQL-päätepisteen luomisen, sekä generoi tietojoukkoa kuvaavan verkkosivun, joka sisältää tietoa esimerkiksi sen tietojoukon skeemasta. Palvelujen kehitystä varten LDF tarjoaa myös palveluita linkitetyn datan tutkimiseen ihmisille luettavassa muodossa. [28]

LDF:ssä Tim Berners-Leen viiden tähden datamallia¹⁹ laajennetaan seitsemän tähden malliksi [30]. Tähtimallia käytetään kuvaamaan julkaistun datan laatua ja käytettävyyttä semanttisen webin sovelluksissa. Erityisesti LDF:n kuudes ja seitsemäs tähti ovat kehitetty edesauttamaan korkeampilaatuisen avoimen linkitetyn datan

¹⁹<https://www.w3.org/DesignIssues/LinkedData.html>

julkaisua, sillä alkuperäinen viiden tähden asteikko ei aseta korkeita vaatimuksia avoimen linkitetyn datan sisäiselle laadulle ja luotettavuudelle. Vaatimukset eri tähtien tasojen saavuttamiselle ovat luettavissa taulukosta 3.

Tähteä	Periaate	Luoja
1 tähti	Data julkaistu ja avoimesti saatavilla verkossa	Tim Berners-Lee
2 tähteä	Avoin data järjestettyä ja tietokoneen luettavissa	Tim Berners-Lee
3 tähteä	Saatavilla oleva data järjestetty avoimesti käytettävissä olevaan formaattiin	Tim Berners-Lee
4 tähteä	Käytä URI:a, jotta dataasi kohden voidaan luoda linkkejä	Tim Berners-Lee
5 tähteä	Linkitä data muiden dataan luoden kontekstia datallesi	Tim Berners-Lee
6 tähteä	Dokumentoi 5 tähden datassasi käytetyt skeemat	LDF.fi, SeCo
7 tähteä	Verifioi, että datasi noudattaa dokumentoitua skeemaa	LDF.fi, SeCo

Taulukko 3: Tähtimallin periaatteet

3.3 Akatemiasampo

Akatemiasampo on SeCo:n vuonna 2021 luoma linkitettyyn avoimeen dataan perustuva semanttinen portaali. Portaalin data perustuu suurelta osin saatavilla oleviin Ylioppilasmatrikkeleiden²⁰ vuosien 1640-1899 tietoihin.

Ylioppilasmatrikkeli-sivustot sisältävät tietoja noin 28 000 suomalaisesta ylioppilaasta, joista noin 18 000 esiintyy vuosien 1640–1852 ylioppilasmatrikkelissa ja 9 500 esiintyy vuosien 1853–1899 ylioppilasmatrikkelissa²¹. Matrikkelisivut päivittävät Vilhelm Laguksen rekonstruoimia Turun palossa tuhoutuneita matrikkeleita perustuen sukututkimuslähteisiin ja aikakaudelta säilyneisiin teksteihin kuten kirkonkirjoihin ja yliopiston arkistomateriaaleihin²².

Akatemiasammossa Ylioppilasmatrikkelit ovat muunnettu linkitetyksi dataksi, joka sisältää hyvin yksityiskohtaista tietoa ylioppilaiden elämästä. Keskeisiin henkilötietoihin kuuluvat muun muassa matrikkeleissa esiintyvien henkilöiden koulutus, elinajat ja elämänvaiheet kuten synnyin- ja kuolinpaikat. Lisäksi Akatemiasammossa on louhittu henkilöiden sukulaissuhteita ja ammatteja.

Vanhojen sampojen uudelleenkäyttö uusien Sampojen rakentamisessa on osoittautunut tehokkaaksi tavaksi kehittää sampoja. Akatemiasammon sisältämät biografiatiedot Suomen ylioppilaista monen vuosisadan ajalta antavat laajan pohjan

²⁰<https://ylioppilasmatrikkeli.fi/>

²¹<https://ylioppilasmatrikkeli.fi/1853-1899/>

²²<https://ylioppilasmatrikkeli.fi/esipuhe/aluksi.php>

Promootiosammolle, koska maisterin ja tohtorin oppiarvon ansainneet promovendit ovat usein suorittaneet myös suomalaisen ylioppilastutkinnon ja opiskelleet Suomessa, minkä seurauksena nämä sammut sisältävät suuren määrän samoja henkilöitä.

3.4 Tulokset

Tässä luvussa esitellään empiirisen tai taiteellisen työn tulokset ja vastataan tutkimuskysymyksiin, jotka on esitetty aiemmin, esimerkiksi johdannossa. Työn tieteellinen arvo mitataan saaduilla tuloksilla ja sillä, miten hyvin argumentointi tukee tutkimuskysymyksiin vastaamista. Kirjallisuustutkielmissa ei tyypillisesti ole erillistä päälukua tuloksille, vaan päätelmät, niiden arviointi ja merkitys kootaan yhteenvetolukuun.

Omien tulostensa merkitsevyyteen tulee suhtautua kriittisesti. Tuloksiaan ja omaa tulkintaansa niistä voi tarkastella kriittisesti joko tässä luvussa tai sitten myöhemmässä, tarkastelun tai johtopäätökset sisältävässä luvussa.

Tässä osassa tulisi käsitellä tutkimuksessa käytetyn aineiston luotettavuutta. Johtopäätöksien luotettavuutta voi käsitellä joko tässä luvussa tai myöhemmässä tarkasteluosassa. Asiaa voi käsitellä omassa luvussaan erillään yhteenvedosta tai johtopäätöksistä.

4 Työn toteutus

Promootiosampo sovelluksen luominen alkoi helmikuussa 2023 ja palvelu julkaistiin saman vuoden syksyn aikana. Promootiosammon luontiin kuului monia työvaiheita, joista ensimmäinen oli alustavan promootiodatan kerääminen internetistä HTML-muodossa ja sen muuntaminen RDF-muotoon. Tämä vaihe kesti noin kaksi kuukautta helmikuusta maaliskuun loppuun.

Huhtikuusta eteenpäin Promootiosammon luonnissa siirryttiin seuraavaan vaiheeseen, jossa Akatemiasammosta²³ monistettiin uusi arkisto GitLab²⁴ alustalle. Tästä arkistosta muotoiltiin Promootiosammon ensimmäinen prototyyppi, jota pystyttiin käyttämään paikallisesti tietokoneella ilman internetiä. Prototyyppiä kehitettiin sekä rikastamalla RDF-dataa entisestään että muokkaamalla käyttöliittymää itseään. Sampo-UI tarjoaa valmiiksi monia tapoja havainnollistaa dataa, mutta Promootiosammon näkymiä varten todettiin hyödylliseksi luoda uusia muokattuja komponentteja sammon tarpeita varten.

Promootiosammon luonnin viimeinen vaihe alkoi loppukesän aikana, jossa sammon ulkomuoto ja datasisältö oli varmennettu ja projektin huomio käännettiin palvelun viimeistelemiseen julkaisua varten. Viimeistelyyn kuului muun muassa koodin virheiden ja palvelun odottamattomien käyttäytymistilanteiden korjaaminen, sekä käyttöliittymän suomen- ja englanninkielisten versioiden oikoluku. Viimeistelyyn kuului myös palvelun pystyttäminen omalle internet-sivustollensa <https://promootiosampo.ldf.fi/fi/>. Promootiosampo julkaistiin 3. marraskuuta 2023.

4.1 Promootiosammon suunnitelma

Promootiosammon toteuttamista varten alustava suunnitelma oli kerätä internetistä promootioihin liittyvää keskeistä tietoa Helsingin yliopiston promootion historiaan ja rooleihin liittyviltä verkkosivuilta²⁵ sekä Helsingin yliopiston filosofisen tiedekunnan promootioita koskevilta Wikipedia-sivuilta²⁶. Promootiosammon rakentamisen aikana lista tietojen keräämistä varten hyödynnettävistä sivustoista laajeni sisältämään myös Finna²⁷- ja Kansalliskirjaston²⁸ palvelut.

Datan kokoamisen jälkeen se oli tarkoitus muuntaa linkitetyksi dataksi ja käyttää sitä Akatemiasammon kaltaisen palvelun luomiseen. Tässä palvelussa moni komponentti, etenkin henkilöihin liittyvät komponentit olisivat toteutettu Akatemiasammon vastaavien komponenttien pohjalta. Promootiosammossa olisi myös näkymä promootioille, ja promootiotapahtuman osanumeroille olisi omat sivunsa, joilla selitettäisiin osatapahtumien yksityiskohtia. Näille sivuille aggregoitaisiin myös muuta tietoa ja aineistoja kyseisiin tapahtumiin liittyen.

²³<https://akatemiasampo.fi/fi/>

²⁴<https://about.gitlab.com/>

²⁵<https://www.helsinki.fi/fi/projektit/promootion-riemuvuosi/promootion-historiaa/roolit-ja-rituaalit>

²⁶https://fi.wikipedia.org/wiki/Filosofisen_tiedekunnan_promootio

²⁷<https://finna.fi/>

²⁸<https://www.kansalliskirjasto.fi/fi>

4.2 Promootiosammon data ja sen keräys

Promootiosammon luonti alkoi tehtävän kuvauksella. Projektin tarkoitus oli rakentaa uusi Sampo-portaali Helsingin yliopiston 100. promootion juhluvuoden kunniaksi. Suunnitelmana oli muuntaa Wikipediassa ja Helsingin yliopiston verkkosivuilta löytyvät listat promootioiden osanottajista ensin CSV-tiedostomuotoon, ja sitten RDF-muotoon Sampo-UI:ssa käyttöä varten.

Työ alkoi lataamalla valitut sivut HTML-tiedostoina, ja poistamalla tiedostoista epäolennaiset osuudet kuten suuren osa HTML-koodin rakenteista. Prosessissa pyrittiin karsimaan mahdollisimman paljon datasta, joka ei suoraan liittynyt promootion henkilöihin. Tähän prosessiin käytettiin sekä manuaalista tiedoston muokkausta, että tiedostoa varten kirjoitettua Python-komentosarjaa, joka säännöllisiä lausekkeita käyttäen (regex) poisti mahdollisimman paljon HTML koodia myös vaikeasti editoitavista kohteista kuten URL-linkeistä puolijärjestettyjen datalistojen sisällä. Iteroivassa prosessissa Python-komentosarjaa parannettiin useita kertoja, kunnes osittain järjestetystä HTML-datasta saatiin muunnettua tyydyttävä määrä informaatiota CSV-muotoon.

Automatisoidun muuntamisen jälkeen CSV-taulukoita täydennettiin manuaalisesti lisäämällä komentosarjan puutteellisesti muuntamia tietoja, sekä korjaamalla väärin muutettuja tietoja, kuten puutteellisia henkilön nimiä. Tässä vaiheessa korjattiin myös henkilöiden nimiä yhdenmukaistettuun kirjoitusmuotoon, mikäli samasta henkilöstä käytettiin eri kirjoitusmuotoja eri lähteiden välillä. Joissain tapauksissa henkilön nimen kirjoitusmuoto saattoi vaihdella jopa saman lähteen sisällä vaatien tarkkaa tietojen eheyden tarkastelua.

Wikipedian ja Helsingin yliopiston verkkosivujen pohjalta laaditun taulukon avulla luotiin ensimmäinen Promootiosammon kolmesta tärkeästä taulukosta. Tämä taulukko listaa rivi riviltä jokaisen promootion ja sarakkeissa se sisältää promootioittain tietoja muun muassa promootion ajankohdasta, yliopiston aikakaudesta ja tärkeimpänä jokaiseen promootioon osallistuneiden henkilöiden nimet järjestettynä rooleittain sarakkeisiin. Esimerkki ensimmäisen promootiotietoja sisältävän taulukon rakenteesta on nähtävissä taulukossa 4.

Promootio	Vuosi	Promoottori	Primusmaisteri
1. Promootio	1643	simon_kexlerus	samuel_hartman
2. Promootio	1647	michael_wexionius	michael_jurvelius
3. Promootio	1650	nicolaus_nycopensis	jonas_betulander
4. Promootio	1653	michael_wexionius	henrik_carstenius

Taulukko 4: Yksinkertaistettu esimerkki muunnetusta datasta muodostetun CSV-taulukon rakenteesta

Henkilöiden dataa muunnettaessa keskeinen kysymys ilmeni pohdittaessa, kuinka dataa tallennettaessa voidaan varmistaa jokaisen henkilön yksikäsitteinen identiteetti. Yksi mahdollisuus olisi ollut luoda jokaiselle henkilölle tunnisteeksi yksilöllinen numerosarja Ylioppilasmatrikkeli-sivustojen ja Akatemiasammon tapaan. Promootiosammonsa päädyttiin lopulta identifioimaan henkilöt mallilla `etunimi_sukunimi` promootioissa esiintyneiden henkilöiden pienen lukumäärän vuoksi. Vajaasta 1200

promootioihin tiedettävästi osallistuneista henkilöistä vain kaksi on ollut täydellisiä nimikaimoja, jotka tarvitsivat numerosarjan lisäystä tarkennetuksi identifikaattoriksi.

Henkilöiden nimien ja roolien muuntamisen jälkeen seuraava askel datan rikastamisessa oli yhdistää promootioiden henkilöihin heihin liittyviä ulkoisia linkkejä ja lähteitä Wikipediasta kerätyn datan perusteella. Tätä tarkoitusta varten Promootiosampoa varten luotiin toinen CSV-taulukko, joka sisältää jokaiseen promootioihin tiedettävästi osallistuneeseen henkilöön liittyviä tietoja. Näitä tietoja ovat muun muassa henkilöihin liittyvät ulkoiset linkit muissa palveluissa kuten Ylioppilasmatrikkeleissa ja Wikipediassa. Muita oleellisia promootioiden henkilöihin liittyviä tietoja tässä uudessa CSV-taulukossa ovat tiedot henkilöiden käyttämistä nimistä, yliopiston osakunnista, joihin henkilöt kuuluivat, tiedot henkilöiden ammasteista, sekä linkit henkilöihin liittyviin kuviin, kuvateksteihin ja kuvien mahdollisiin tekijänoikeuslisensseihin. Esimerkki henkilötietoja sisältävästä taulukosta on nähtävissä taulukossa 5.

NameID	Name	YO-MatrikkeliLink	WikiLink
simon_kexlerus	Simon Kexlerus	https://ylioppilas.matrikkeli.fi/henkilo.php?id=U7	/wiki/Simon_Kexlerus
samuel_hartman	Samuel Hartman	https://ylioppilas.matrikkeli.fi/henkilo.php?id=321	/wiki/Samuel_Hartman
petter_torpensis	Petter Torpensis	https://ylioppilas.matrikkeli.fi/henkilo.php?id=317	
michael_wexionius	Michael Wexionius (Gylden stolpe)	https://ylioppilas.matrikkeli.fi/henkilo.php?id=U8	/wiki/Mikael_Wexionius

Taulukko 5: Yksinkertaistettu esimerkki Promootiosammon henkilöitä koskevan CSV-taulukon rakenteesta

Promootiosammon datan rikastamisessa tärkeä tekijä on ollut löytää promovendeihin ja muihin promootioihin osallistuneisiin henkilöihin liittyviä kuvia. Kuvien rikastamisessa Wikipedia toimi ensimmäisenä merkittävänä kuvien lähteenä ja promootioiden henkilöistä tiedettyjen Wikipedia linkkien avulla oli helppo löytää näihin henkilöihin liittyviä Wikimedia kuvia. Wikipediaa tärkeämpi kuvien lähde oli kuitenkin Kansallisarkiston ylläpitämä Finna-palvelu²⁹.

Finnan kautta käyttäjän on mahdollista hakea tietoa ja teoksia useiden suomalaisten museoiden kokoelmista. Hakemisen lisäksi käyttäjä voi tutkia teoksien digitoituja versioita, mikäli niitä on saatavilla. Laajojen digitaalisten kokoelmien lisäksi Finna-palvelu omaa vahvat hakutulosten rajaustoiminnot, jotka helpottavat aineistojen siivilöintiä. Esimerkiksi haettaessa yleisen nimen omaavaan henkilöön liittyviä teoksia Finnassa

²⁹<https://finna.fi/>

on mahdollista rajata hakutuloksia luomisajankohdan, teoksen tyyppin (kuva, kirja, esine, opinnäytetyö ja niin edelleen.), tekijänoikeuslupien, luojan tai otsikon mukaan.

Vahvojen rajaustoimintojen ja Ylioppilasmatrikkeleissa saatavilla olevien biografiatietojen avulla oli mahdollista etsiä tarkasti kuvia etenkin 1800-luvun promootioihin osallistuneista henkilöistä tai heidän käyttämästään esineistöstä. Valokuvakameroiden kehittämistä edeltäviltä aikakausilta 1600- ja 1700-luvuilla oli paljon vaikeampaa löytää henkilöihin liittyvää aineistoa kuten muotokuvamaalauksia tai esineistöä. Tämän sijaan Finnasta löytyi Kansallisarkiston digitoimia varhaisiin promootioihin osallistuneiden henkilöiden latinankielisiä opinnäytetöitä sekä muuta heidän kirjoittamaansa kirjallisuutta. Vaikka opinnäytetyöt eivät suoraan kerro paljoa yksittäisen henkilön ulkonäöstä, on varsin merkittävää huomata, että varhaisen Suomalaisen akateemisen kulttuurin henkilöistä on säilynyt varsin kattava määrä fyysistä ja digitoitua aineistoa.

Finna-palvelun aineistot ovat toimineet suuressa roolissa Promootiosammon datan rikastamisessa. Huolimatta siitä, että kaikki Finnasta kerätty data käsiteltiin ilman automaatiota, on sen arvo ja vaikutus Projektin sisältöön ollut huomattava. Kuvien lisäksi kuviin liittyvistä kuvauksista ja metatiedosta on ollut hyötyä Promootiosammossa, sillä näistä on monien henkilöiden kohdalla ollut mahdollista muodostaa niukkoja biografiaa kuvauksia tilanteissa, joissa henkilöstä ei tiedetä nimen ja promootioajankohdan lisäksi muuta informaatiota.

Promootiosammon datan rikastamiseen kuvien avulla liittyi myös haasteita. Suurin käytännön ongelma kuvien kanssa oli käyttöoikeuslupien varmistaminen. Monet valokuvat ja muotokuvat ovat vapaasti käytettävissä ilman tarvetta huolehtia mahdollisista tekijänoikeusloukkauksista, koska kuvan alkuperäisen tekijän kuolemasta on voinut kulua monta vuosisataa. Suuri osa kuvista on kuitenkin lisensoitu erilaisten Creative Commons³⁰ (CC) lisenssien alle, ja kuvien käytölle voi olla monenlaisia erilaisia rajoituksia. Tärkeimmät rajoitukset CC-lisensseissä ovat usein vaatimukset mahdollisimman kattavasta maininnasta kuvan lisenssiin ja alkuperäiseen tekijään tai kuvan omistajaan, vaatimukset ei-kaupallisesta käytöstä ja vaatimukset olla muokkaamatta alkuperäistä kuvaa. Poikkeuksen CC-lisensseihin muodostavat CC0- ja Public Domain-merkinnät, jotka antavat luvan täysin vapaaseen käyttöön.

Vaihtelevat käyttöilupien tekijänoikeuslisenssit muodostivat ajoittain esteitä yksittäisten kuvien käytölle. Ajoittain myös digitaalisten aineistojen käyttöilupia ei ollut merkitty aineistojen sivuille estäen kyseisten kuvien käytön. Tästä huolimatta suurimmalle osalle Promootiosammossa esiintyvistä henkilöistä oli mahdollista löytää korvaava kuva yhdestä tai toisesta lähteestä. Lopullisessa Promootiosammossa 685 henkilöllä 1181:sta oli heihin liitetty kuva. Prosentuaalisesti tämä tarkoittaa, että 58% henkilöistä omasi kuvan. Tämä on merkittävän suuri prosentti, kun otetaan huomioon promootiotapahtuman kasvu 1900-luvulta alkaen. Moni 1900- ja 2000-luvulla promootioihin osallistunut henkilö on vielä elossa ja heistä mahdollisesti otetut kuvat ovat tekijänoikeussyistä ja yksityisyydensuojan takia paljon harvinaisempia kuin 1600-1800-luvuilla eläneistä henkilöistä saatavilla olevat aineistot. Näin on merkittävää huomata, että ennen 1900-lukua promootioihin osallistuneista 624 henkilöstä 493:sta on saatavilla Promootiosammossa kuva joko heistä itsestään, heidän muotokuvasta,

³⁰<https://creativecommons.org/>

opinnäytetyöstä tai heidän aikoinaan omistamasta esineistöstä. Prosentuaalisesti henkilöistä tunnettujen kuvien määrä nousee tällöin 79 prosenttiin, mikä voidaan katsoa yllättävän suureksi määräksi.

Promootioiden henkilöihin liittyvien kuvien lisäksi Promootiosampoa rikastettiin myös Helsingin yliopiston filosofisen tiedekunnan promootioihin itseensä liittyvillä kuvilla. Finna-palvelusta etsittiin promootioihin liittyviä kuvia ja esineistöä. Mikäli promootioon itseensä liittyvää kuvaa ei ollut saatavilla, valittiin Promootiosammon promootionäkymässä promootiota edustamaan yksi siihen osallistuneista henkilöistä. Näitä promootioihin liittyviä kuvia ja niiden lisäinformaatiota kuten tekijänoikeuslisenasseja ja kuvatekstejä varten luotiin Promootiosammon kolmas CSV-taulukko. Tähän taulukkoon lisättiin myös sarakkeena kuvaukset promootioista itsestään. Kuvauksissa kerrotaan lyhyesti kyseiseen promootioon mahdollisesti liittyneistä tapahtumista, mikäli niistä löydettiin tietoa internetin tai kirjallisten lähteiden kautta.

Promootioiden kuvausten kautta nousee esille tämän akateemisen perinteen värikkäs historia. Esimerkki varhaisten promootioiden skandaaleista on syytökset mustan magian harjoittamisesta oppilaiden keskuudessa. Kuvauksista on myös mahdollista löytää promootioiden merkityksellisyys muun muassa Suomen kansallishengen kehityksessä ja itsenäisyyspyrkimysten näyttämönä.

Promootiosampoa varten kerätyn CSV-taulukoidun datan luomisen jälkeen viimeinen askel ennen käyttöönottoa Promootiosammossa oli muuntaa CSV-taulukot RDF-muotoon. Tätä varten luotiin Python-komentosarja, joka lukee kyseiset tiedostot ja luo niiden sisältämän tiedon pohjalta kaksi Turtle-notaatioon perustuvaa tiedostoa.

Ensimmäinen Python-komentosarjan luoma Turtle tiedosto sisältää promootioihin liittyvät RDF-kolmikot, jotka sisältävät tietoa muun muassa promootion ajankohdasta, osallistujista, promootioihin liittyvistä kuvista ja muuta mahdollista lisätietoa.

Toinen Python-komentosarjan luoma Turtle tiedosto taas sisältää promootioihin osallistuneiden henkilöiden tietoja. Promootioiden henkilöiden tietoihin kuuluu muun muassa tiedot siitä, mihin promootioihin he ovat osallistuneet, missä rooleissa he ovat osallistuneet promootioihin, henkilöön liittyvän kuvan tiedot ja kuvaus, tunnetut ammatit ja linkit henkilöön toisissa internet-palveluissa.

CSV-RDF muunnos-komentosarjan kirjoittamisessa hyödynnettiin rdflib³¹, pandas³² ja re³³ kirjastoja. Rdflib kirjastoa käytettiin RDF graafin luomiseen sekä sen populoimiseen RDF literaaleilla ja URI:illa. Pandas kirjastoa käytettiin CSV-taulukoiden lukemiseen ja kevyeen data-analyysiin suodattamalla pois tyhjiä soluja CSV-taulukoista. Re-kirjasto on Pythonin regex-kirjasto ja sitä käytettiin muokkaamaan CSV-taulukoiden sisältämiä tekstijonoja lopulliseen ihmisten lukemista varten tarkoitettuun muotoon. Regexiä tarvittiin muun muassa vaihtamaan taulukoiden tekstikuvauksista CSV-formaatin käyttämien merkkien tilalle vaihdetut tilapäismerkit niiden alkuperäiseen muotoon sen jälkeen, kun ne eivät enää vaaranna CSV-tiedoston eheyttä, sekä jakamaan yhdessä CSV-solussa sijaitsevat datainstanssit erilleen.

³¹<https://rdflib.readthedocs.io/en/stable/>

³²<https://pandas.pydata.org/>

³³<https://docs.python.org/3/library/re.html>

4.3 Promootiosampo-portaali

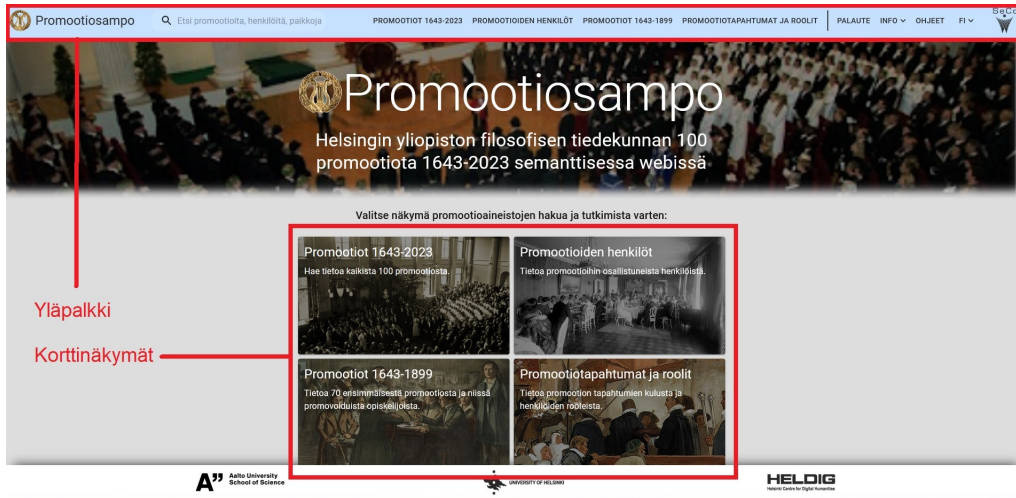
Promootiosampopalvelun rakentaminen alkoi alustavan datan keräyksen jälkeen luomalla kopio Akatemiasammon koodista uuteen GitLab-arkistoon. Tästä kopiosta ryhdyttiin muokkaamaan Promootiosampoä itsenäisenä, mutta vahvasti Akatemiasammon henkilödataan nojaavana palveluna. Promootiosampoä varten luotu CSV-data muunnettiin RDF-muotoon ja lisättiin sen tietokantaan.

Sammosta poistettiin Akatemiasammon perspektiivi- ja fasettikomponentit, kuten sen vanhat henkilösivut, ja tilalle luotiin uudet perspektiivinäkymät promootioita ja promootioissa esiintyviä henkilöitä varten. Perspektiivinäkymiä varten kirjoitettiin SPARQL-kysely komentosarjoja, joiden avulla Promootiosammon tietokannasta on mahdollista hakea ja järjestää perspektiivi- ja instanssinäkymissä käytettävää dataa.

Promootiosammon Sampo-UI-sivusto käyttää hyväkseen Promootiosammon SPARQL-palvelupistettä datan visualisoimiseen. Verkkosivuohjelmisto jäsentää ja lähettää SPARQL kyselyitä Sampo-UI ohjelmistoon koodattujen valmiiden kyselyiden sekä käyttäjän fasettihaussa tekemien valintojen perusteella. Parsituista kyselyistä saatujen vastausten perusteella Sampo-UI pystyy muuntamaan informaation visualisoitavaksi sivuston eri perspektiiveissä ja komponenteissa. Esimerkkejä fasettihaun tuloksista on nähtävissä kuvissa [7](#) ja [8](#).

4.3.1 Promootiosammon etusivu

Kuvassa [6](#) nähtävissä oleva Promootiosammon etusivu koostuu korttinäkymästä, joista kolme ensimmäistä ohjaavat käyttäjän tutkimaan sammon promootioihin tai henkilöihin liittyviä perspektiivinäkymiä. Promootiosammon etusivun neljäs kortti ohjaa käyttäjän promootiotapahtumista ja sen rooleista kertovalle sivulle. Promootiosammon sivut sisältävät myös yläpalkin, jonka kautta on mahdollista navigoida Promootiosammon eri näkymien välillä, hakea avainsanoja, vaihtaa käyttöliittymän kieltä sekä avata ohje- ja yhteydenottosivuja.



Kuva 6: Promootiosammon etusivu

4.3.2 Promootionäkymät

Promootiosampo sisältää kaksi promootionäkymää: Promootiot 1643-2023 -näkyvässä on listattu kaikki Helsingin yliopiston filosofisen tiedekunnan kaikki 100 promootiota. Promootiot 1643-1899 -näkyvä on kuin edeltävä näkyvä, mutta se on rajattu ensimmäiseen 70 promootioon Akatemiasammossa ja Ylioppilasmatrikkeleissa esiintyvien henkilöiden aikakauden mukaan.

Promootionäkymissä on 24 roolia, joissa henkilö on saattanut ottaa osaa promootioihin ja 37 eri muuttujaa yhteensä, joihin kuuluu muun muassa valokuvia ja tietoja promootiossa esitetyistä runoista ja kantaateista. Osa muuttujista on visualisoitu perspektiivinäkymässä, mutta toiset näytetään ainoastaan promootioiden instanssinäkymissä. Esimerkiksi käytettyjen kuvien tekijänoikeustiedot ja -lisenssit löytyvät merkittynä vain instanssisivuilta, jotta perspektiivinäkymässä on enemmän tilaa ja näkyvyyttä merkittäville muuttujille.

Promootioperspektiiveillä on kuvassa 7 nähtävän taulukkonäkymän lisäksi kaksi tilastonäkymää, joista ensimmäinen on histogrammi, joka näyttää promootioon osallistuneiden henkilöiden määrät promootioittain. Toinen tilastonäkymä näyttää promootioiden promovendien kirjallisista lähteistä lasketut ja Akatemiasammon datasta lasketut lukumäärät histogrammeina.

The screenshot shows the 'Promootiosampo' application interface. At the top, there is a search bar and navigation links. The main content area is titled 'Promootiot 1643-2023' and 'Perspektiivinäkymän visualisaatio'. A sidebar on the left shows filters, including 'Aktiiviset suodattimet' and 'Rajoiita:'. The main table displays a list of promotions with columns for 'Promootio', 'Promootion kuvaus', 'Kuva', 'Kuvateksti', 'Tapahtumapäivä', 'Yliopisto', and 'Promootori'. The table is filtered to show only promotions where Zacharias Topelius is involved. The table has a 'Taulukkonäkymä' (Table view) label at the bottom right.

Promootio	Promootion kuvaus	Kuva	Kuvateksti	Tapahtumapäivä	Yliopisto	Promootori
56. promootio (1844)	Aikalaiskäsien mukaan promootio meni vanhan tavan mukaan, heidän koreettijoukkoille, puhuta latinaksi, Valtava kuusivuorokauden ja tanssi.		Zachris Topelius, suomenruotsalainen kirjailija, runoilija, toimittaja, historioitsija ja Keisarikunnan Aleksanterin Yliopiston rehtori	1844-01-21	Keisarikunnan Aleksanterin Yliopisto	Nathaniel Gerhard af Schulten
59. promootio (1853)	Vinainen promootio varhaisilla statustilla ennen vuoden 1852 statustien käyttöönottoa promootiossa. Samana vuonna julkaisiin myös huvilyhtymä Resan...		Samuel Werner von Trui oli suomalainen senaattori, pankinjohtaja, tilanomistaja ja vapaaherra. Hän oli myös Turun kaupunginvaltuuston ensimmäinen puheenjohtaja.	1853-05-30	Keisarikunnan Aleksanterin Yliopisto	Ernst Woldemar
60. promootio (1857)	Uuden Nikolai:n statustien myötä filosofien tiedekunta jaettiin kahtia historiallis-kulttuurilliseen ja fyysis-matematiikan tiedekuntaan, joihin kalleen samankalaiseen...		Carl Gustaf Estlander oli suomalainen kirjallisuuden tutkija, Helsingin yliopiston ensimmäinen ja uudemman kirjallisuuden professori ja valtiopäivämies	1857-05-29	Keisarikunnan Aleksanterin Yliopisto	Adolf Edward Arppe

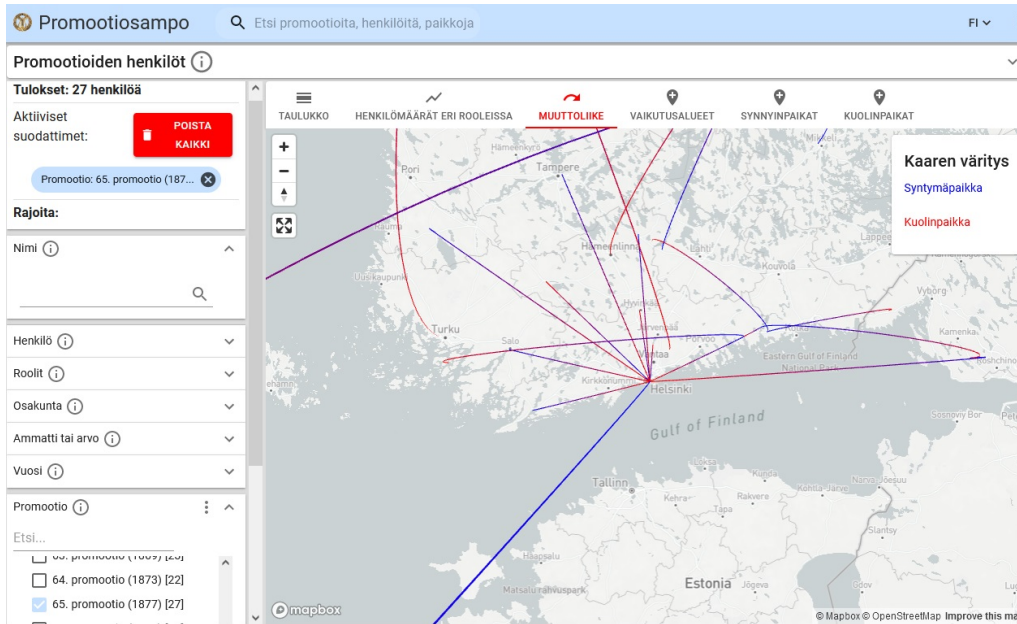
Kuva 7: Kuva Promootiosammon perspektiivinäkymästä, jossa tulokset ovat suodatettu näyttämään ne promootiot, joihin Zacharias Topelius on eri rooleissa.

4.3.3 Henkilönäkymä

Henkilönäkymässä on listattu kaikki tunnetut promootioissa eri rooleissa esiintyneet henkilöt. Promootiosammon henkilöperspektiivinäkymä sisältää kuusi muuttujaa ja henkilöiden instanssisivut 22 muuttujaa. Yhteensä henkilöitä on 1181 kappaletta. Henkilöperspektiivissä on kuusi eri näkymää, kun mukaan lasketaan henkilöstä tunnettua dataa kertova taulukkonäkymä. Yksi näkymä visualisoi henkilöiden määrää eri promootioissa kuten promootionäkymissäkin, mutta neljä muuta näkymää käyttävät hyödykseen Akatemiasammon pohjalta saatua dataa visualisoimaan kartalla muun muassa kuvassa 8 nähtävän henkilöiden muuttoliikkeen heidän syntymä- ja kuolinpaikkojen välillä. Loput näkymät visualisoivat henkilöiden synnyinpaikat, kuolinpaikat ja Ylioppilasmatrikkeleiden teksteissä mainitut vaikutusalueet kartalla numeromääräisinä pisteinä.

4.3.4 Promootiotapahtumat ja -roolit

Kuvassa 9 nähtävä Promootiosammon promootiotapahtumat ja roolit -sivu sisältää käsinkirjoitettuja tietoja modernin promootiojuhlan tapahtumankulusta ja tapahtumamateriaalista sekä selityksiä promootiossa esiintyvien roolien työnkuvasta ja vastuualueista. Tämä sivu Promootiosammon perspektiivi- ja instanssisivuista poiketen ei generoi sen sisältöä SPARQL-päätepisteen pohjalta, vaan se on kirjoitettu manuaalisesti React-kielen avulla. Promootiotapahtumat ja roolit -sivun pohja ei ole alkuperäinen osa Sampo-UI järjestelmää, vaan sivun komponentti kehitettiin Promootiosammon tarpeita varten.



Kuva 8: Kuva Promootiosammon henkilönäkymän muuttoliikkeen visualisaatiosta kartalla rajattuna 65. promootion henkiöihin. Kuvan kartan tekijänoikeudet kuuluvat MapBox ja OpenStreetMap palveluille.

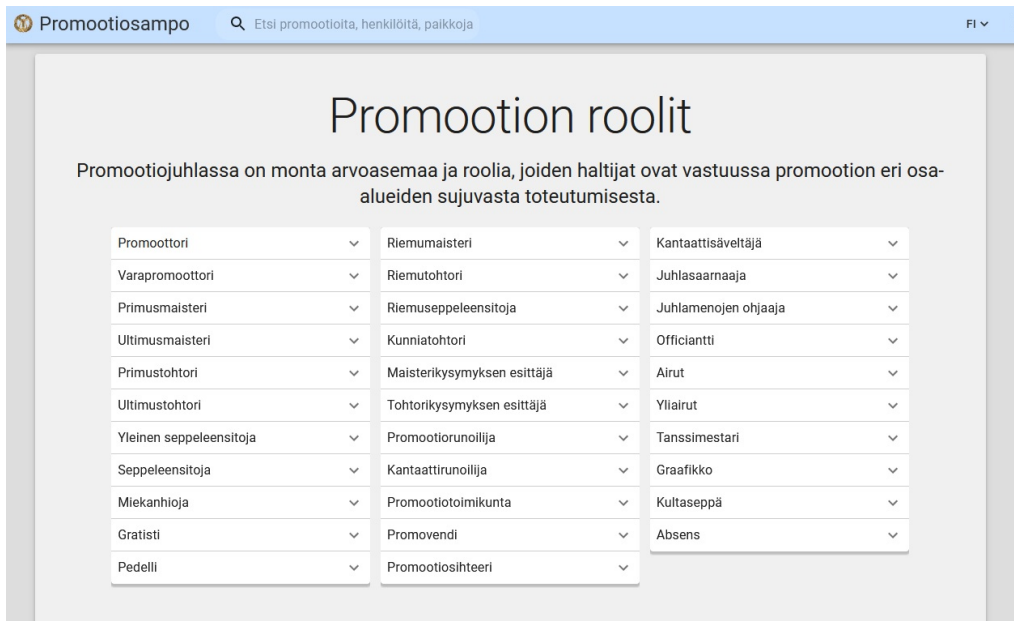
4.4 SPARQL-palvelupiste

Promootiosampon kuuluu visuaalisen käyttöliittymän lisäksi SPARQL-palvelupiste, jonka kautta käyttäjät voivat päästä käsiksi Promootiosammon dataan ja muodostaa siitä omia analyysejään. Esimerkkinä tästä on kuvan 10 histogrammi, joka on luotu syöttämällä kuvassa 11 nähtävä SPARQL-kysely Promootiosammon SPARQL-pääteeseen Yasgui³⁴-verkkopalvelussa. Päätepiesteen palauttamasta datasta pystytään laskemaan esimerkiksi lukumääriä eri sivustoille johtaville ulkoisille linkeille, ja tämä voidaan visualisoida suoraan palvelussa kuvan 10 mukaisesti.

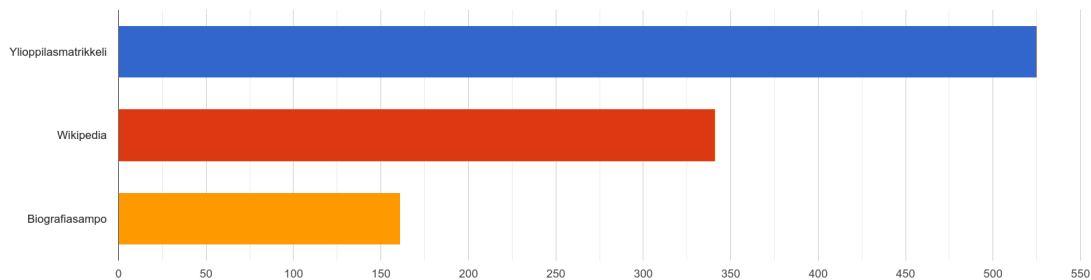
Tässä esimerkissä laskettiin viittauksia Wikipediaan, Ylioppilasmatrikkeleihin ja Biografiasampon³⁵ Promootiosammosta johtavien ulkoisten linkkien määrät. Tämän ohella SPARQL-kyselyitä on mahdollista käyttää myös muunlaisen tiedon hakemiseen Promootiosammosta. Kyselyillä on mahdollista hakea tilastollista tietoa muun muassa siitä, kuinka moniin promootioihin eri henkilöt ovat osallistuneet. On myös mahdollista rajata ja hakea tietoa promootioiden henkilöistä muun muassa sen mukaan, missä ammatissa he ovat työskennelleet. Promootiosammon Sampo-UI päätepieste luo monia tämänkaltaisia kyselyitä ja kysyy niitä Sammon SPARQL-palvelupisteeltä luodakseen sen perspektiivinäkömissä tietojen visualisointeja.

³⁴<https://yasgui.tripty.cc/>

³⁵<http://biografiasampo.fi/>



Kuva 9: Kuva Promootiosammon roolien ja tapahtumien sivusta.



Kuva 10: Kuvassa histogrammi, jossa lueteltuna kuinka monella Promootiosammon henkilöllä on linkki eri ulkoisiin palveluihin.

4.5 Promootiosammon tietomalli

Promootiosammon skeema oli viimeisiä projektissa työstettyjä osa-alueita. Skeema on RDF-graafi, joka kertoo Promootiosammon luokista ja ominaisuuksista. Skeeman luonti on tarpeellista, jotta Promootiosammon datan laatu voidaan korottaa kuuden tähden dataksi. Skeeman avulla on mahdollista varmentaa, että Promootiosammon tietokanta noudattaa sitä, mahdollistaen seitsemännen tähden datan luokituksen ansaitsemisen. Sisällöltään Promootiosammon skeema käsittelee promootioihin ja promootioiden henkilöihin liittyvän datan rakennetta.

Taulukko 6 kuvaa osaa Promootiosammon tietomallista. Promootiosampo sisältää kolme luokkaa; promootiot, Ylioppilasmatrikkeleihin perustuvat vuosien 1643-1899 promootiot ja promootioiden henkilöt. Liitteiden A ja B kautta, jotka kuvaavat edellä-

```

PREFIX scprs: <http://ldf.fi/schema/promootiosampo/>
PREFIX nbf: <http://ldf.fi/nbf/>

SELECT ?label (COUNT (?matrikkeli) AS ?matrikkeliCount) (COUNT (?wiki) AS ?wikiCount) (COUNT (DISTINCT ?yoma) AS ?bioCount)
WHERE {
  ?x a scprs:Person
  OPTIONAL{
    ?x scprs:matrikkeliLink ?matrikkeli
    BIND("Ylioppilasmatrikkeli" AS ?label)
  }
  OPTIONAL{
    ?x scprs:wikiLink ?wiki
    BIND("Wikipedia" AS ?label)
  }
  OPTIONAL {
    ?a nbf:yoma ?yoma .
    ?b scprs:matrikkeliLink ?yoma
    BIND ("Biografiasampo" AS ?label)
  }
}
GROUP BY ?label

```

Kuva 11: Kuvassa SPARQL-kysely, jolla kuvan 10 histogrammi on saatu aikaan.

mainitut luokat, on mahdollista tutustua laajempaan Promootiosammon tietomalliin.

Promootio-luokka ja Ylioppilasmatrikkeleihin perustuvien promootioiden luokka ovat identtiset, ja niiden funktio Promootiosammossa on jakaa promootiot kahteen helposti Sampo-UI:ssa käsiteltävään luokkaan Promootiosammon eri perspektiivinäkymiä varten. Promootioluokat sisältävät muun muassa perusluonteista tietoa jokaisesta promootioentiteetistä kuten ajankohdan, yliopiston nimen sekä promootioon liittyvän kuvan URI:t ja tekijänoikeusmerkinnät. Perustietojen lisäksi promootioluokka sisältää linkitykset kaikkiin tunnettuihin nimetyissä rooleissa kyseiseen promootioon osallistuneisiin henkilöihin. Henkilöluokka on promootioluokkaa paljon pienempi ja sisältää tietoja kuten henkilön identifikaatiotunnuksen, osakunnan, tunnettuja ammatteja tai arvoasemia sekä henkilöihin liitetyn kuvan URI:t ja tekijänoikeusmerkinnät. Tämän lisäksi henkilöluokka sisältää tunnetut ulkoiset linkit henkilöiden Wikipedia- ja Ylioppilasmatrikkelisivuille. Näitä ulkoisia linkkejä käytetään muodostamaan siltoja Akatemiasammon henkilödataan, mikä mahdollistaa Promootiosammon henkilöinstanssisivujen populoinnin Akatemiasammosta saatavissa olevalla linkitetyllä datalla.

Elementti	Kuvaus	Nimi
Promootio		
scprs:backupConferrer	Rooli	Varapromoottori
scprs:celebrationPreacher	Rooli	Juhlasaarnaaja
scprs:confermentCantata	Promootiossa esitetyn kantaatin nimi	Promootiokantaatti
scprs:confermentDescription	Lyhyt kuvaus promootioon liittyvistä merkittävistä tapahtumista	Promootion kuvaus
Henkilö		
scprs:matrikkeliLink	Linkki henkilöön liittyvälle Ylioppilasmatrikkelisivulle	Ylioppilasmatrikkeli linkki
scprs:Nation	Henkilön osakunta yliopistossa	Osakunta
scprs:personID	Henkilön uniikki ID	Henkilö ID
scprs:title	Henkilön ammatti tai arvo	Ammatti tai arvo
scprs:wikiLink	Linkki henkilöön liittyvälle Wikipediasivulle	Wikipedia linkki

Taulukko 6: Ote laajemmasta Promootiosammon tietomallista. Scprs: on Promootiosammon Turtle tiedostoissa käytetty lyhenne URI:lle <http://ldf.fi/schema/promootiosampo/>.

5 Tulosten arviointi

Tässä luvussa arvioidaan Promootiosammon onnistumista ja tuloksia sen valmistuksen jälkeen. Luvussa arvioidaan myös projektin suunnitelman toteutumista, sekä merkitystä laajemmassa mittakaavassa.

5.1 Tulokset palvelun näkökulmasta

Promootiosampo on palveluna toimiva eikä se sisällä palvelun toimintaa huomattavasti häiritseviä virheitä koodissa. Projektin lopputuloksen on mahdollista katsoa saavuttaneen projektin alussa asetetut tavoitteet ja Promootiosampo on osa laajenevaa Sampo-palveluiden verkostoa.

Promootiosammon data on rikasta ja etenkin palvelun promootio- ja henkilöinstansseihin liitettyjen kuvien määrä, sekä henkilöinstanssien yhdistäminen Akatemiasammon tarjoamaan dataan ovat olleet tärkeä tekijä palvelun onnistuneen sisällön ja ilmiäsuon luonnissa. Toinen onnistunut tekijä palvelun mielenkiintoisuuden lisäämisessä ovat olleet käsin kirjoitetut kuvaukset promootioista sekä kuviin lisätyt kuvatekstit. Tuloksena Promootiosampo sisältää kiinnostavaa tietoa sekä tutkimusta että vapaa-ajan selaustarkoituksia varten.

Promootiosampo täyttää viiden tähden datan vaatimukset; sen RDF-data on julkaistu ja linkitetty toisiin lähteisiin. Ennen Promootiosammon luontia internetissä promootioista oli julkaistu dataa van yhden tähden tasolla, eli digitoituna, mutta pääosin järjestämättömänä datana. Data oli myös pitkälti puutteellisesti ja jopa väärin linkitettyä. Esimerkiksi henkilöitä saattoi olla satunnaisesti linkitetty heihin liittyville Wikipedia- tai Ylioppilasmatrikkelisivuille, mutta Promootiosammon näitä linkejä on täydennetty sadoilla uusilla linkeillä järjestelmällisesti.

Promootiosampo täyttää myös LDF:n kuudennen tähden vaatimukset. Promootiosammon datarakenteesta on muodostettu tietomalli, joka on julkaistu LDF-sivustolla, mutta datan oikeellisuutta ei ole tarkistettu. Tämän vuoksi Promootiosammon data ei yllä tällä hetkellä seitsemännen tähden tasolle.

5.2 Promootiosammon tieteelliset ja kulttuurilliset hyödyt

Promootiosammon julkaiseminen on merkityksellinen tapahtuma suomalaisen aineettoman kulttuuriperinnön näkökulmasta. Promootiosampo on nimittäin ensimmäinen internetissä julkaistu promootioista kertova jäsenetty tietokanta. Promootiosampo edeltävät aineistot internetissä ovat olleet luettelomaisia listoja, joiden sisältämää dataa ei ole järjestetty yhdenmukaisesti rakenteisiin.

Järjestetty digitaalinen data tulee olemaan hyödyllistä, jos aiheesta tullaan tekemään tulevaisuudessa tutkimusta. Promootiosammon SPARQL-palvelupiste mahdollistaa monimutkaisten hakujen tekemisen promootiosammon dataan, sekä erilaisten siihen liittyvien tilastojen luomisen. Data on myös mahdollista ladata, muokata sekä käyttää uudelleen uusissa tarkoituksissa.

Kompaktien promootioihin liittyvien faktojen lisäksi Promootiosampo sisältää kuvauksia promootioista ja henkilöistä itsestään. Nämä kuvaukset antavat lisää valoa

itse tapahtumiin, niiden aikaiseen maailmaan, sekä promootioihin osallistuneiden henkilöiden elämänvaiheisiin. Historiantutkimuksen kannalta tällaiset kuvaukset voivat esimerkiksi auttaa yhdistämään henkilöitä eri historiallisten lähteiden välillä.

6 Jatkokehitys

Vaikka Promootiosampo rakennettiin ja julkaistiin onnistuneesti, sitä on mahdollista jatkokehittää vielä monin tavoin. Promootiosammon tietokantaa on mahdollista rikastaa monenlaisilla uusilla datapisteillä sekä verkkosivun luovaa Sampo-UI koodia on mahdollista parannella monin tavoin. Tämän lisäksi Promootiosammon data olisi mahdollista validoida SeCo:n käyttämän seitsemän tähden järjestelmän korkeimmalle tasolle.

6.1 Tietokanta

Promootiosampo sisältää potentiaalia jatkokehitystä varten. Etenkin Promootiosammon henkilöihin liittyvää tietoa on mahdollista rikastaa monin tavoin. Sammon datassa on selkeä jakoraja Ylioppilasmatrikkelin ja Akatemiasammon vuosilta 1643-1899 saatavilla olevan datan, sekä myöhemmän, pääosin Wikipediasta kerätyn datan välillä. Vanha Ylioppilasmatrikkeleihin pohjautuva tieto on paikoitellen hyvin yksityiskohtaista, mutta uudempien, 1900-luvulta alkaen vaikuttaneiden henkilöiden tiedot, jotka eivät nojaa Akatemiasammon tarjoamaan dataan eivät ole yhtä kattavia. Promootiosammon dataa olisi mahdollista rikastaa muun muassa Wikipedian kautta saatavissa olevilla tiedoilla esimerkiksi henkilöiden synnyin- ja kuolinpaikoista, ammasteista, sekä vaikutusalueista. Tätä tarkoitusta varten voisi olla mahdollista hyödyntää esimerkiksi Wikidataa³⁶ tai DBpediaa³⁷. Myös manuaalinen tiedon louhinta henkilöiden Wikipediasivuilta on mahdollisuus.

Wikidata ja DBpedia ovat eräitä maailman suurimmista avoimista tietämysgraafeista, joten Promootiosammon liittäminen niihin voisi tuoda monia etuja linkitetyn datan saralla.

Monet Promootiosammon tietokannan literaaliarvot ovat suomenkielisiä sanoja ja lauseita. Nämä tekstit olisi mahdollista kääntää myös englanniksi, jotta myös kansainvälinen yhteisö voisi tutkia ja hyötyä Promootiosammon sisällöstä. Jotta englanninkielistä linkitettyä dataa olisi mahdollista käyttää yhdessä suomenkielisen datan kanssa, Promootiosampo olisi tarpeen muuntaa itse Sampo-UI:n järjestelmätasolla. Sampo-UI ei Promootiosammon luomisajankohtana tukenut käyttäjän komennosta vaihtuvaa kieltä sen tietokannasta antamissa hakutuloksissa ja tämä ominaisuus tulisi imitoida esimerkiksi FINEST-Lakisammosta³⁸, jossa monikielisyys on yksi sammon keskeisiä ominaisuuksia.

Promootiosammon dataa olisi mahdollista laajentaa erityisesti lisäämällä tietoja muiden kuin Helsingin yliopiston filosofisen tiedekunnan promootioista. Tällä hetkellä Promootiosampo sisältää tietoja vain yhden tiedekunnan promootioista, mutta suomalaisilla yliopistolla on monia promootioita järjestäviä tiedekuntia. Esimerkiksi Helsingin yliopistolla on teologinen, oikeustieteellinen, lääketieteellinen, valtiotieteellinen, maatalous-metsätieteellinen sekä eläinlääketieteellinen tiedekunta, jotka

³⁶https://www.wikidata.org/wiki/Wikidata:Main_Page/Content/fi

³⁷<https://www.dbpedia.org/about/>

³⁸<https://finestlaw.demo.seco.cs.aalto.fi/fi>

ovat järjestäneet omia promootiojuhliansa. Näiden tiedekuntien promootioseremonioiden lukumäärä on vähäisempi filosofiseen tiedekuntaan verrattuna, mutta useilla tiedekunnista on tästä huolimatta 1600- ja 1700-luvuille yltävät perinteet filosofisen tiedekunnan rinnalla.

Promootiosammon tietokannassa olisi parannettavaa myös ulkoisten linkkien saralla. Tällä hetkellä Promootiosampo luo ulkoisia linkkejä Biografiasampoon lukemalla Biografiasammon SPARQL-palvelupisteen pohjalta luotua RDF-tiedostoa, joka listaa Biografiasammon henkilöinstanssien URI:t Ylioppilasmatrikkelilinkkien kanssa. Näistä kolmikoista Promootiosammossa päätellään SPARQL-hakujen avulla mitkä promootiohenkilöiden instanssit ovat samoja Biografiasammossa esiintyneitä henkilöinstansseja. Ongelma ilmenee siinä, että Ylioppilasmatrikkeleita siltana käytettäessä Promootiosammon linkitykset Biografiasampoon rajautuvat Ylioppilasmatrikkeleissa esiintyneiden henkilöiden lailla vuosiin 1643-1899, ja myöhemmille henkilöille, esimerkiksi presidentti Sauli Niinistölle, ei tule luotua ulkoista linkkiä Biografiasampoon siitä huolimatta, että hänellä on omat instanssisivunsa molemmissa sammoissa. Tähän epäkohtaan ratkaisuna olisi lisätä manuaalisesti Promootiosammon tietokantaan Biografiasampoon osoittavia ulkoisia linkkejä sen sijaan, että linkitysten luonnissa luotetaan SPARQL-kyselyiden tarjoamaan automaatioon.

6.2 Käyttöliittymä

Sampo-UI:ta itseään on myös mahdollista jatkokehittää. Osa sen komponenteista on suotavaa päivittää ajantasaisiin versioihin ja Sampo-UI:n dokumentaatiota on mahdollista kehittää kertomaan paremmin sovelluksen komponenttien käytöstä sekä dokumentoimaan, kuinka kolmannet tahot voivat ottaa palvelun käyttöön.

Palvelun komponentteja olisi mahdollista kehittää edelleen, jotta monikielisen datan käyttö olisi mahdollista. Tämä olisi erityisen suotavaa ottaen huomioon sen, että käyttöliittymä tukee jo monikielisyyttä käyttäjäpäätteen tasolla.

Muita mahdollisia kehityksen kohteita olisi kehittää fasettihaun komponenttia. Promootiosammon erikoisia käyttäytymisen ominaisuuksia fasettihaussa nousee esille etenkin henkilöperspektiivin ammattien fasettihakukentässä, jossa ammattidataa syötetään yhteen fasettiin kahdesta eri lähteestä. Ensimmäinen lähde on Akatemiasammon sisältämä henkilöiden arvoihin ja ammatteihin liittyvä data ja toinen lähde on Promootiosammon oma ammattidata, joka on louhittu Helsingin yliopiston promootioita koskevilta verkkosivuilta. Kahden lähteen ongelma nousee esille erityisesti silloin, kun kumpikin lähde omaa tietoja samasta ammatista. Akatemiasammon data voidaan hakea muodostamalla RDF-kolmikko `"?id scprs:matrikkeliLink/^schema:sameAs/:has_title ?title"` ja promootiosammon data haetaan kolmikolla `"?id scprs:title ?title"`. Lopputuloksena on, että fasettihaun näyttämässä listassa on kaksi ammattia, joilla on identtinen `"skos:prefLabel"`, ihmisluettava nimi, esimerkiksi `"professori"`, mutta entiteettinä kyse on kahdesta eri resurssista. Tämän seurauksena fasettihaun lista ei pysty yhdistämään näitä samoja ammatteja. Fasettihaussa olisi hyvä olla tämänkaltaisia tilanteita varten mahdollisuus sulauttaa yhteen ne listan kohteet, joiden ihmisluettavat nimet ovat identtisiä joidenkin muiden listan kohteiden kanssa.

Promootiosampoon jäi kehitettävää myös promootionäkymiin luotujen uusien histogrammivisualisaatioiden kanssa. Histogrammi, jota käytetään visualisoimaan promovendien määrää promootioissa kirjallisten lähteiden ja Akatemiasammosta laskeutujen henkilöiden perusteella ei rajaudu fasettihaun rajaamien tulosten mukaan. Kyseinen komponentti on hyvin samankaltainen sen henkilömääriä eri rooleissa laskevan sisäkomponentin kanssa, joka reagoi oikein fasettihaun suodattamiin tuloksiin, joten kyseessä on todennäköisesti virhe lähdekoodissa tai komponentin kutsunnassa.

Promootiosampo käyttää yhdistettyjä hakuja (federated query) ottamaan yhteyden Akatemiasammon SPARQL-palvelupisteeseen. Tämä hidastaa hakuja merkittävästi kirjoittamishetken toteutuksessa. Haut olisi mahdollista optimoida lataamalla päivitetty Akatemiasammon RDF-data Promootiosammon tietokantaan ja muokkaamalla Promootiosammon käyttämät SPARQL-kyselyt käyttämään sisäistä tietokantaa yhdistettyjen hakujen sijaan.

Promootiotapahtumat ja roolit -sivun data on osa Promootiosammon Sampo-UI osuutta RDF-datan sijaan, mikä tarkoittaa, että perspektiivi ei ole rakennettu linkitetyn datan pohjalle. Ratkaisun hyöty on, että informaatio tapahtumista ja rooleista on helppo kääntää suomeksi ja englanniksi, mutta haittapuolena on, että data ei ole saatavilla SPARQL-palvelupisteen kautta. Mahdollinen kehityssuunta olisi muuntaa tieto tapahtumista ja rooleista RDF-dataksi ja luoda jokaiselle roolille ja tapahtumalle omat instanssisivunsa Sampo-UI-järjestelmässä. Näin tieto olisi saatavilla myös SPARQL-palvelupisteen kautta. Yksilöllisissä instanssisivuissa olisi myös hyötyä uuden datan lisäyksessä. Esimerkiksi rooleihin liittyviä kuvia asuista ja esineistä olisi mahdollista koostaa instanssisivuille paljon helpommin nykyiseen toteutukseen verrattuna. Myös tapahtumien tietoja olisi mahdollista rikastaa monin tavoin. Esimerkiksi tapahtuman sijainti tai kulkueen kulkureitti olisi mahdollista visualisoida instanssisivulle luodulla kartalla.

Henkilöiden ja promootioiden instanssinäkymiä olisi mahdollista rikastaa luomalla automaattisesti instanssiin liittyviä hakulinkkejä Finna-palveluun. Finnaa on käytetty monien kuvien ja opinnäytetöiden löytämiseen etsittäessä kuvia instansseille, mutta monesta instanssista on olemassa useampi kuva kuin Promootiosampoon on valittu. Näiden tulosten selaamista varten Finna-hakulinkki esimerkiksi henkilön nimen perusteella rajattuna samaisen henkilön tunnettuun elinaikaan voi tuoda potentiaalisia tuloksia. Automatisoitujen hakujen haittapuolena on se, että joidenkin henkilöiden nimet voivat erota huomattavasti Promootiosammon ja Finnan välillä, joten haettavien henkilöiden identiteetti ei välttämättä ole yksintertaisesti määriteltävissä.

6.3 Promootiosammon datan validointi

Promootiosammon data olisi mahdollista varmentaa seitsemän tähden dataksi. Tällä hetkellä Promootiosampo kattaa viiden tähden datan edellytykset Tim Berners-Leen laatiman viiden tähden datan mallissa ja SeCo:n ehdottaman seitsemän tähden mallissa Promootiosammon datalla on edellytykset kuudennen tähden ansaitsemiselle. Seitsemättä tähteä varten Promootiosammon skeema tulisi testata esimerkiksi SHACL

Shapes Constraint Languagea³⁹ käyttäen, jotta voidaan olla varmoja, että Promootio-
sampo noudattaa kuudennen tähden antavaa skeemaa.

³⁹<https://www.w3.org/TR/shacl/>

7 Yhteenveto

Idea Promootiosammon kehittämiseen syntyi Suomen promootioperinteen juhla-vuodesta 2023, jolloin Helsingin yliopiston filosofinen tiedekunta vietti sadannetta promootiojuhlaansa 380 vuotta ensimmäisen promootion jälkeen. Tämän juhlan kunniaksi haluttiin luoda Promootiosampo, joka sisältää promootioihin liittyvää tietoa linkitettyinä datana semanttisessa webissä.

Henkilökohtaisella tasolla Promootiosammon luonti ja tämän opinnäytetyön kirjoittaminen ovat olleet opettavaisia kokemuksia. Semanttiseen webiin liittyvät aiheet ovat tulleet minulle aiempaa tutummiksi Promootiosammon kanssa työskennellessä. Muun muassa RDF, Turtle ja SPARQL ovat olleet keskeisiä Promootiosammon kehittämiseen liittyviä standardeja. Projektin tasolla Promootiosampo on toimiva osa laajenevaa Sampo- ja tavoite promootiodataa linkitetyksi dataksi jäsentävän palvelun luomisesta voidaan katsoa saavutetuksi. Henkilökohtaisena saavutuksena pidän Promootiosammon henkilödatan rikastamista kattavasti, erityisesti 1600-1900-luvuilla esiintyneiden henkilöiden kohdalta.

Promootiosampoa luotaessa sen sisältöä varten louhittiin paljon dataa käsityönä erityisesti promootioihin osallistuneisiin henkilöihin liittyviin kuviin ja henkilöitä kuvaileviin kuvateksteihin liittyen. Myös promootioista itsestään louhittiin paljon niihin liittyvää historiallista tietoa. Tämä tieto Promootiosampoon valjastettuna auttaa luomaan mielikuvia eri promootioiden aikakausista ja maailmantilanteista, joissa ne järjestettiin. Esimerkiksi Suomen kansallismielisyyden ja itsenäistymispyrkimysten heräämisen vuodet ovat vahvasti olleet vaikuttamassa promootioperinteeseen 1800-luvun lopun ja 1900-luvun alun aikana.

Itse Promootiosampo-sovelluksen kehitys on ollut vaiheikasta. Promootiosampoa varten Sampo-UI:hin luotiin uusia, sekä muokattiin vanhoja komponentteja. Tässä prosessissa ilmeni sekä Sampo-UI:hin liittyviä vahvuuksia että heikkouksia. Sampo-UI on monimutkainen kokonaisuus, joka tarvitsee lisää dokumentaatiota, jotta sen käyttöön ottaminen olisi helpompaa. Jyrkästä oppimiskäyrästä huolimatta se monipuolinen alusta, jonka pohjalta on mahdollista kehittää semanttista webiä monipuolisesti hyödyntäviä palveluita.

Viitteet

- [1] Artukka, Topi ja Ratia, Mari. The History of Ceremonial Conferments in Turku. *AURAICA. Scripta a Societate Porthan edita*, 13 (2022), 42–49.
- [2] Auer, Sören ja Bizer, Christian ja Kobilarov, Georgi ja Lehmann, Jens ja Cyganiak, Richard ja Ives, Zachary. DBpedia: A nucleus for a web of open data. *International Semantic Web Conference (2007)*, Springer, 722–735.
- [3] Bernasconi, Eleonora ja Ceriani, Miguel ja Di Pierro, Davide ja Ferilli, Stefano ja Redavid, Domenico. Linked data interfaces: a survey. *Information* 14, 9 (2023), 483.
- [4] Bernasconi, Eleonora ja Ceriani, Miguel ja Mecella, Massimo ja Catarci, Tiziana. Design, realization, and user evaluation of the ARCA system for exploring a digital library. *International Journal on Digital Libraries*.24, 1 (2023), 1–22.
- [5] Berry, David M. ja Anders Fagerjord. *Digital humanities: Knowledge and critique in a digital age*. John Wiley & Sons, 2017.
- [6] Borst, Willem Nico. *Construction of engineering ontologies for knowledge sharing and reuse*, 1997.
- [7] Butuc, Marius-Gabriel. *Semantically Enriching Content Using OpenCalais*. *Editia* 9 (2009), 77–88.
- [8] Carrasco, Laís ja Vidotti, Silvana A Borsetti Gregorio. *Proceedings of the International Conference on Dublin Core and Metadata Applications (2015)*, Dublin Core Metadata Initiative.
- [9] Ceriani, Miguel ja Bottoni, Paolo. SparqlBlocks: Using Blocks to Design Structured Linked Data Queries. *Language (XSD)* 1, 11 (2017).
- [10] Chakravarthy, Mr Ajay ja Ciravegna, Prof Fabio ja Lanfranchi, Ms Vitakeska. *AKTiveMedia: Cross-media document annotation and enrichment*, 2006.
- [11] De Boer, Victor ja Wielemaker, Jan ja Van Gent, Judith ja Hildebrand, Michiel ja Isaac, Antoine ja Van Ossenbruggen, Jacco ja Schreiber, Guus. Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Case Study. In *The Semantic Web: Research and Applications: 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27-31, 2012*. *Proceedings* 9 (2012), Springer, pp. 733–747.
- [12] Färber, Michael and Bartscherer, Frederic and Menne, Carsten and Rettinger, Achim. Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web* 9, 1 (2018), 77–12.

- [13] Ferilli, Stefano ja Redavid, Domenico. The GraphBRAIN system for knowledge graph management and advanced fruition. *Foundations of Intelligent Systems: 25th International Symposium, ISMIS 2020, Graz, Austria, September 23–25, 2020, Proceedings (2020)*, Springer, 308—317.
- [14] Gangemi, Aldo. A Comparison of Knowledge Extraction Tools for the Semantic Web. *The Semantic Web: Semantics and Big Data: 10th International Conference, ESWC 2013, Montpellier, France, May 26-30, 2013. Proceedings 10 (2013)*, Springer, 351—366.
- [15] Gómez-Pérez, Asunción ja Corcho, Oscar. Ontology languages for the semantic web. *IEEE Intelligent systems* 17, 1 (2002), 54—60.
- [16] Gramatica, Ruggero ja Pickering, Ruth. Start-up story Yewno: an AI- driven path to a knowledge-based future. *Insights: The UKSG Journal* 30, 2 (2017), 107–112.
- [17] Grlický, Vladimír. An overview of RDF model representation formats. In *IIT. SRC 2005: Student Research Conference (2005)*, Citeseer, s. 242
- [18] Haag, Florian and Lohmann, Steffen ja Ertl, Thomas. SparqlFilterFlow: SPARQL query composition for everyone. In *The Semantic Web: ESWC 2014 Satellite Events: ESWC 2014 Satellite Events, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers 11 (2014)*, Springer, 362—367.
- [19] Haag, Florian ja Lohmann, Steffen ja Siek, Stephan ja Ertl, Thomas. . QueryVOWL: A visual query notation for linked data. In *The Semantic Web: ESWC 2015 Satellite Events: ESWC 2015 Satellite Events, Portorož, Slovenia, May 31–June 4, 2015, Revised Selected Papers 12 (2015)*, Springer, 387—402.
- [20] Halonen, Tero. *Promootio : Elävän yliopistoperinteen historiaa, 2023. Kustannusosakeyhtiö Otava.*
- [21] Hastrup, Tuukka ja Cyganiak, Richard ja Bojars, Uldis. *Browsing Linked Data with Fenfire, 2008.*
- [22] Heath, Tom ja Bizer, Christian. *Linked data: Evolving the web into a global data space, vol. 1. Morgan & Claypool Publishers, 2011.*
- [23] Hoffart, Johannes ja Suchanek, Fabian M ja Berberich, Klaus ja Lewis-Kelham, Edwin ja De Melo, Gerard ja Weikum, Gerhard. YAGO2: Exploring and Querying World Knowledge in Time, Space, Context, and Many Languages. In *Proceedings of the 20th international conference companion on World wide web (2011)*, 229—232.
- [24] Hogue, Andrew ja Karger, David. Thresher: automating the unwrapping of semantic content from the World Wide Web. In *Proceedings of the 14th international conference on World Wide Web (2005)*, 86—95.

- [25] Hyvönen, Eero. Using the Semantic Web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery. *Semantic Web* 11, 1 (2020), 187—193.
- [26] Hyvönen, Eero. “Sampo” model and semantic portals for digital humanities on the Semantic Web. In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020)* (2020), CEUR-WS. org.
- [27] Hyvönen, Eero. Digital humanities on the Semantic web: Sampo model and portal series. *Semantic Web* 14, 4 (2023), 729–744.
- [28] Hyvönen, Eero ja Alonen, Miika ja Tuominen, Jouni ja Mäkelä, Eetu. Linked Data Finland: Towards a 7-star service platform for linked datasets. In *The First Annual KnowEscape Conference-KnowEscape* (2013), pp. 33—40.
- [29] Hyvönen, Eero ja Leskinen, Petri ja Rantala, Heikki ja Ikkala, Esko ja Tuominen, Jouni. Akatemiesampo-portaali ja-datapalvelu henkilöiden ja henkilöryhmien historialliseen tutkimukseen (Academysampo Portal and Data Service for Biographical and Prosopographical Research). *Infor- maatiotutkimus* 40, 2 (2021), 28—56.
- [30] Hyvönen, Eero ja Tuominen, Jouni ja Alonen, Miika ja Mäkelä, Eetu. Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets. In *The Semantic Web: ESWC 2014 Satellite Events: ESWC 2014 Satellite Events*, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers 11 (2014), Springer, 226—230.
- [31] Ikkala, Esko ja Hyvönen, Eero ja Rantala, Heikki ja Koho, Mikko. Sampo- UI: A full stack Javascript framework for developing semantic portal user interfaces. *Semantic Web* 13, 1 (2022), 69—84.
- [32] Kivilaakso, Aura ja Marsio, Leena. Safeguarding living heritage with participatory Wikiinventorying in Finland. *Memoriamedia Review* 1 (2017).
- [33] Leskinen, Petri ja Hyvönen, Eero. Reconciling and using historical person registers as linked open data in the AcademySampo portal and data service. In *International Semantic Web Conference* (2021), Springer, pp. 714—730.
- [34] Leskinen, Petri ja Hyvönen, Eero ja Rantala, Heikki. Analyzing the Lives of Finnish Academic People 1640-1899 in Nordic and Baltic Countries: AcademySampo Data Service and Portal. In *CEUR Workshop Proceedings* (2022), vol. 3232, RWTH Aachen University, pp. 94—10.
- [35] Mäkelä, Heidi Henriikka. Digesting the Finnish Nature and Past: Food, Pastness, and the Naturalness of the National in the Wiki-Inventory for Living Heritage. *Journal of ethnology and folkloristics* 15, 2 (2021), 89—111.

- [36] Marilena Alivizatou-Barakou, Alexandros Kitsikidis, Filareti Tsalakanidou, Kosmas Dimitropoulos, Chantas Giannis, et al. *Intangible Cultural Heritage and New Technologies: Challenges and Opportunities for Cultural Preservation and Development. Mixed Reality and Gamification for Cultural Heritage*, Springer International Publishing, 2017.
- [37] Massari, Arcangelo and Peroni, Silvio. *HERITRACE: Tracing Evolution and Bridging Data for Streamlined Curatorial Work in the GLAM Domain*. arXiv preprint arXiv:2402.00477 (2024).
- [38] Mazurek, Cezary ja Sielski, Krzysztof ja Walkowska, Justyna ja Werla, Marcin. *Applicability of CIDOC CRM in digital libraries*. In *Proceedings of CIDOC 2011 Knowledge Management and Museums Conference, Sibiu, Romania* (2011).
- [39] Mendes, Pablo N and Jakob, Max and García-Silva, Andrés and Bizer, Christian. *DBpedia spotlight: shedding light on the web of documents*. In *Proceedings of the 7th international conference on semantic systems* (2011), 1—8.
- [40] Messaoudi, T., Véron, P., Halin, G., & De Luca, L. *An onto- logical model for the reality-based 3D annotation of heritage building conservation state*. *Journal of Cultural Heritage* (2018).
- [41] Micsik, András ja Tóth, Zoltán ja Turbucz, Sándor. *Lodmilla: Shared visualization of linked open data*. In *Theory and Practice of Digital Libraries—TPDL 2013 Selected Workshops: LCPD 2013, SUEDL 2013, DataCur 2013, Held in Valletta, Malta, September 22-26, 2013. Revised Selected Papers 3* (2014), Springer, 89—100.
- [42] Miller, Eric. *An Introduction to the Resource Description Framework*, 1998.
- [43] Mouromtsev, Dmitry ja Pavlov, Dmitry S ja Emelyanov, Yury ja Morozov, Alexey V ja Razdyakonov, Daniil S ja Galkin, Mikhail. *The Simple Web-based Tool for Visualization and Sharing of Semantic Data and Ontologies*. *ISWC (Posters & Demos)* (2015)
- [44] Nisheva-Pavlova, Maria ja Alexandrov, Asen. *GLOBDEF: a framework for dynamic pipelines of semantic data enrichment tools*. In *Metadata and Semantic Research: 12th International Conference, MTSR 2018, Limassol, Cyprus, October 23-26, 2018, Revised Selected Papers 12* (2019), Springer, 159—168.
- [45] Nuzzolese, Andrea Giovanni ja Presutti, Valentina ja Gangemi, Aldo ja Musetti, Alberto ja Ciancarini, Paolo. *Aemoo: exploring knowledge on the web*. *Proceedings of the 5th Annual ACM Web Science Conference* (2013), 272—275.
- [46] Oldman, Dominic ja CRM Labs. *The CIDOC conceptual reference model (CIDOC-CRM): primer*. *CRM Labs 5* (2014).

- [47] Palonen, Tuomas ja Suominen, Osma ja Pessala, Sini ja Lappalainen, Mikko ja Nykyri, Susanna ja Frosterus, Matia. ONKI-projektin ontologiaeditorien vertailuraportti, 2013.
- [48] Paulheim, Heiko. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web* 8, 3 (2017), 489—508.
- [49] Peckham, Joan ja Maryanski, Fred. Semantic data models. *ACM Computing Surveys (CSUR)* 20, 3 (1988), 153—189.
- [50] Peroni, Silvio ja Tomasi, Francesca ja Vitali, Fabio. Reflecting on the Europeana data model. In *Digital Libraries and Archives: 8th Italian Research Conference, IRCDL 2012, Bari, Italy, February 9-10, 2012, Revised Selected Papers 8* (2013), Springer, 228—240.
- [51] Po, Laura ja Malvezzi, Davide ja muut. High-level visualization over big linked data. In *CEUR WORKSHOP PROCEEDINGS* (2018), vol. 2180, CEUR-WS.
- [52] Popov, Borislav ja Kiryakov, Atanas ja Ognyanoff, Damyan ja Manov, Dimitar ja Kirilov, Angel. KIM—a semantic platform for information extraction and retrieval. *Natural language engineering* 10, 3-4 (2004), 375—392.
- [53] Pujara, Jay ja Miao, Hui ja Getoor, Lise ja Cohen, William. Knowledge graph identification. In *The Semantic Web—ISWC 2013: 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I* 12 (2013), Springer, 542—557.
- [54] Ragini, Ch Suvarna. Self-selected reading for pleasure through “Talk to Books” app to improve reading skills among tertiary level learners. *Journal of English Language Teaching* 61, 3 (2019), 12—18.
- [55] Simon, Rainer ja Barker, Elton ja Isaksen, Leif ja de Soto Cañamares, Pau. Linking early geospatial documents, one place at a time: annotation of geographic documents with Recogito. *e-Perimtron* 10, 2 (2015), 49—59.
- [56] Sinaci, Ali Anil ja Gonul, Suat. Semantic Content Management with Apache Stanbol. In *The Semantic Web: ESWC 2012 Satellite Events: ESWC 2012 Satellite Events, Heraklion, Crete, Greece, May 27-31, 2012. Revised Selected Papers 9* (2015), Springer, pp. 371—375.
- [57] Spyns, Peter ja Meersman, Robert ja Jarrar, Mustafa. Data modelling versus ontology engineering. *ACM SIGMod Record* 31, 4 (2002), 12—17.
- [58] Studer, Rudi ja Benjamins, V Richard ja Fensel, Dieter. Knowledge engineering: Principles and methods. *Data & knowledge engineering* 25, 1-2 (1998), 161—197.
- [59] Tomaszuk, Dominik. Named graphs in RDF/JSON serialization. *Zeszyty Naukowe Politechniki Gdańskiej* (2011), 273.

- [60] Daniel Tunkelang. *Faceted Search*. Springer Cham, 2009.
- [61] Vargas, Hernán ja Buil-Aranda, Carlos ja Hogan, Aidan ja López, Claudia. RDF explorer: A visual SPARQL query builder. In *The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part I 18 (2019)*, Springer, 647—663.
- [62] Viola, Fabio and Roffia, Luca and Antoniazzi, Francesco and D’Elia, Alfredo and Aguzzi, Cristiano and Salmon Cinotti, Tullio. Interactive 3D exploration of RDF graphs through semantic planes. *Future Internet* 10, 8 (2018), 81.
- [63] World Wide Web Consortium et al. *SPARQL 1.1 overview*, 2013.
- [64] World Wide Web Consortium et al. *Best practices for publishing linked data*, 2014.
- [65] Yosef, Mohamed Amir and Hoffart, Johannes and Bordino, Ilaria and Spaniol, Marc and Weikum, Gerhard. AIDA: An Online Tool for Accurate Disambiguation of Named Entities in Text and Tables. *Proceedings of the VLDB Endowment* 4, 12 (2011).

A Promootioiden tietomalli

Promootiosampo sisältää kaksi luokkaa, jotka käyttävät identtistä promootioiden tietomallia. Ensimmäinen luokka kattaa kaikki Helsingin yliopiston filosofisen tiedekunnan promootiot ja toinen luokka kattaa vuosien 1640-1899 aikana pidetyt filosofisen tiedekunnan promootiot. Tämä jako on seurausta tarpeesta luoda nopeita SPARQL-kyselyitä, jotka pytyvät erottelemaan promootioiden Ylioppilasmatrikkeleiden tietoja sisältävän osajoukon.

Promootioiden tietomalli sisältää rakenteen promootioluokille jossa ominaisuudet kuvaavat muun muassa promootioinstanssiin eri rooleissa osallistuneet henkilöt, promootioon liittyvät kuvat sekä tiedot promootion tapahtumapäivämäärästä ja -vuodesta.

Ominaisuus	Kuvaus	Nimi
Promootio		
scprs:backupConferrer	Rooli	Varapromoottori
scprs:cantataPoet	Rooli	Kantaattirunoilija
scprs:cantataComposer	Rooli	Kantaattisäveltäjä
scprs:celebrationPreacher	Rooli	Juhlasaarnaaja
scprs:confermentCantata	Promootiossa esitetyn kantaatin nimi	Promootiokantaatti
scprs:confermentDescription	Lyhyt kuvaus promootioon liittyvistä merkittävistä tapahtumista	Promootion kuvaus
scprs:conferment-DescriptionSource	Lähde josta kuvauksen tiedot löydetty	Lähde promootion kuvaukselle
scprs:confermentPoem	Promootiossa esitetyn runon nimi	Promootioruno
scprs:confermentPoet	Rooli	Promootiorunoilija
scprs:conferrer	Rooli	Promoottori
scprs:danceMaster	Rooli	Tanssimestari
scprs:doctorPrimus	Rooli	Primustohtori
scprs:doctorUltimus	Rooli	Ultimustohtori

Ominaisuus	Kuvaus	Nimi
Promootio		
scprs:doctorQuestioneer	Rooli	Tohtorikysymyksen esittäjä
scprs:externalLinks	Promootiohin liittyviä ulkoisia linkkejä	Ulkoiset linkit
scprs:graphicDesigner	Rooli	Graafikko
scprs:goldsmith	Rooli	Koruseppä
scprs:gratisti	Rooli	Gratisti
scprs:headMarshal	Rooli	Yliairut
scprs:honoraryDoctor	Rooli	Kunniatohtori
scprs:Image	Henkilöön liittyvän kuvan URI	Kuva
scprs:ImageAttribution	Tietoja kuvan tekijänoikeuslissensistä	Valokuvan tekijänoikeusviittaukset
scprs:ImageDescription		Kuvaan liittyvä kuvaus
scprs:ImageSource	URI, joka johtaa sivulle, jolta henkilöön liittyvä kuva on löydetty	Valokuvan lähdesivu
scprs:jubileeDoctor	Rooli	Riemutohtori
scprs:jubileeGratisti	Rooli	Riemugratisti
scprs:jubileeMagister	Rooli	Riemumaisteri
scprs:jubileeWreathWeaver	Rooli	Riemuseppeleensitojatar
scprs:magisterPrimus	Rooli	Primusmaisteri

Ominaisuus	Kuvaus	Nimi
Promootio		
scprs:magisterUltimus	Rooli	Ultimusmaisteri
scprs:magister-Questioneer	Rooli	Maisterikysymyksen esittäjä
scprs:masterOf-Ceremonies	Rooli	Juhlamenojen ohjaaja
scprs:promovend	Luku	Promovendien lukumäärä
scprs:tapahtumaPaiva	Päivämäärä, jolloin promootio pidettiin	Promootion päivämäärä
scprs:university	Yliopiston nimi, jossa henkilö osallistui promootioon	Yliopisto
scprs:wreathWeaver	Rooli	Seppelensitoja
scprs:year	Vuosi, jolloin promootio pidettiin	Promootion vuosi
scprs:yearEnd	Vuosi, jolloin promootio pidettiin	Promootion vuosi

Taulukko A1: Promootiosammon Promootioiden tietomalli

B Henkilöiden tietomalli

Henkilöiden tietomallin ominaisuudet sisältävät tietoa muun muassa promootioihin osallistuneiden henkilöiden kuvista, rooleista ja ammateista. Ominaisuudet sisältävät myös tietoa henkilöihin liittyvistä ulkoisista linkeistä ja Promootiosampo käyttää näitä linkkejä muodostamaan siltoja esimerkiksi Akatemiasammon dataan.

Ominaisuus	Kuvaus	Nimi
Henkilö		
scprs:Image	Henkilöön liittyvän kuvan URI	Kuva
scprs:ImageAttribution	Tietoja kuvan tekijänoikeuslissenssistä	Valokuvan tekijänoikeusviittaukset
scprs:ImageDescription	Kuvaan liittyvä kuvaus	
scprs:ImageSource	URI sivulle, jolta henkilöön liittyvä kuva on löydetty	Valokuvan lähdesivu
scprs:matrikkeliLink	Linkki henkilön Ylioppilasmatrikkelisivulle	Ylioppilasmatrikkeli linkki
scprs:Nation	Henkilön osakunta yliopistossa	Osakunta
scprs:ParticipatedIn	Promootiot, joihin henkilö on osallistunut	
scprs:personID	Henkilön ainutlaatuinen identifikaatiotunniste	Henkilö ID
scprs:prefLabel	Henkilön nimi	Henkilön nimi
scprs:role	Henkilön roolit	Rooli
scprs:title	Henkilön ammatti tai arvo	Ammatti tai arvo
scprs:wikiLink	Linkki henkilön Wikipediasivulle	Wikipedia linkki

Taulukko B1: Promootiosammon henkilöiden tietomalli