



## Biografiasammon tekoäly yhdistää ja rikastaa suomalaiset elämäkerrat semanttisessa webissä

Eero Hyvönen, Petri Leskinen, Minna Tamper, Heikki Rantala,  
Esko Ikkala, Jouni Tuominen ja Kirsi Keravuori

*Biografiasampo-järjestelmä käynnistää uuden aikakauden elämäkertakokoelmien julkaisemisessa ja käyttämisessä verkossa. Järjestelmän ydinaineistona on Kansallisbiografia ja muut Suomalaisen Kirjallisuuden Seuran (SKS) ja tieteellisten seurojen toimittamat pieniselämäkerrat, yhteensä 13 100 elämäntarinaa, joita on kirjoittanut 900 suomalaista tutkijaa. Biografiasammon innovaationa on luoda kieliteknologian, tekoälyn ja semanttisen webin teknologioiden avulla elämäkertojen teksteistä ja niihin eri lähteissä liittyvistä tiedoista tietämysverkko (knowledge graph) ja kansallinen tietoinfrastruktuuri, joka koostuu miljoonista tietojen välisistä yhteyksistä. Tietämysverkko on julkaistu linkitetyn datan palvelussa, jonka varaan on toteutettu seitsemästä sovellusnäköymästä koostuva älykäs, kaikille avoin ja maksuton verkkopalvelu **biografiasampo.fi** kansalaisten ja digitaalisten ihmistieteiden tutkijoiden käytettäväksi.*

Suomalaisen Kirjallisuuden Seuran (SKS) Biografiakeskus<sup>1</sup> julkaisee Suomen historian keskeisten henkilöiden pieniselämäkertoja kirjoina ja verkossa. Esimerkiksi yhteistyössä Suomen Historiallisen Seuran kanssa on koottu Kansallisbiografia<sup>2</sup>, johon kuuluu lähes 6500 elämäkerta. Nämä julkaistiin 10-osaisena, 9500-sivuisena suurteossarjana vuosina 2003-2008. Lisäksi SKS:n kokoelmiin kuuluvat erilliset tietokannat mm. tunnetuista sotilashenkilöistä, kirkonmiehistä ja talousvaikuttajista, ja aineistoja laajennetaan uusilla henkilöryhmillä yhteistyössä tieteellisten seurojen kanssa (ks. taulukko 1). Aineistot ovat olleet luettavissa verkossa osin maksullisen palvelun kautta.

	Elämäkertakokoelma	Koko
1	Kansallisbiografia	6478
2	Kenraalit ja amiraalit 1809-1917	481
3	Suomen papisto 1800-1920	1234
4	Talouselämän vaikuttajat	2235
5	Turun hiippakunnan paimenmuistio 1554-1721	2716
	<b>Yhteensä elämäkertoja</b>	<b>13144</b>

**Taulukko 1.** SKS:n Biografiakeskuksen pieniselämäkertojen tietokannat, jotka muodostavat Biografiasammon ytimen.

Vastaavia, mutta vielä laajempia kirja- ja verkkopalveluhankkeita on muissakin maissa, esimerkkeinä Britannian Oxford Dictionary of National Biography (ODNB)<sup>3</sup>, USA:n American National Biography<sup>4</sup>, Saksan

<sup>1</sup> <http://biografiakeskus.fi>

<sup>2</sup> <https://kansallisbiografia.fi/kansallisbiografia>

<sup>3</sup> <http://global.oup.com/oxforddnb/info>

<sup>4</sup> <http://www.anb.org/aboutanb.html>

Neue Deutsche Biographie, Ranskan Nouvelle Biographie Generale<sup>5</sup>, Hollannin Biography Portal of the Netherlands<sup>6</sup> ja BiographyNet ja Ruotsin Dictionary of Swedish National Biography<sup>7</sup>. Alan ehkä tunnetuin kokoelma ODNB sisältää n. 63 000 elämäkertaa ja Biography portal of the Netherlands lähes 150 000. Elämäkertojen pituus eri kokoelmissa vaihtelee alkaen matrikelimaisista yhteenvedoista seikkaperäisiin artikkeleihin.

Henkilöhistoriallinen tieto on tärkeä komponentti historian tutkimuksessa, ja elämäkerrallisten tietojen saatavuus kätevästi yhdestä auktoritatiivisesta kokoelmasta helpottaa tutkijoiden työtä. Kansallisten biografiahankkeiden yhtenä pontimena on ollut myös kansallisvaltioiden identiteetin lujittaminen tarkastelemalla historiaa omien suurmiesten ja -naisten toiminnan kautta. Elämäkerrat kiinnostavat tutkijoiden ohella myös suurta yleisöä ja elämäkerrat ovat usein kirjakauppojen myyntitilastojen kärkisijoilla. Perinteistä näkökulmaa suurmiehistä ja -naisista on nykyisin laajennettu mm. vähemmistöihin ja ns. "tavallisiin" henkilöihin tasapainoisemman kuvan luomiseksi kansakunnan edustajista. Elämäkerrallista tietoa julkaistaan verkossa biografiakokoelmien ohella runsaasti muuallakin, kuten Wikipediassa, organisaatioiden ja ammattikuntien matrikkeleissa, kansainvälisissä auktoriteettitietokannoissa, sukututkimuksen piirissä, museoiden ja arkistojen kokoelmissa jne.

### Biografiasammon visio

Biografiakeskuksen suomalaiset elämäkerrat ja vastaavat kansainväliset kokoelmat ovat olleet lähes poikkeuksetta saatavilla vain tekstimuodossa ihmisen luettavaksi eikä semanttisena eli tietokoneen "ymmärtämänä" datana. Lisäksi tietokannoissa oleva metatieto rajoittuu yleensä biografisiin perustietoihin, ja valtaosa henkilötiedosta on olemassa vain rakenteettomana artikkelitekstinä, jota tietokone ei ymmärrä. Elämäkertatietojen analysoimiseen ja tutkimiseen ei juuri ole tarjolla työkaluja tai palveluita. Tämä olisi kuitenkin tarpeen henkilöhistorioiden data-analyttistä tutkimusta kuten verkostoanalyysiä varten ja työkalujen ja älykkäiden sovellusten kehittämiseksi digitaalisissa ihmistieteissä. SKS:n elämäkeroilla on ollut verkossa vuosittain n. 300 000 lukijaa, joten kysyntää kehittyneille palveluille olisi olemassa.

Biografiasammon uutuusarvo perustuu laskennallisten menetelmien uudelleen hyödyntämiseen sekä biografioiden julkaisussa että niiden tarjoamiseen loppukäyttäjien käytettäväksi helppokäyttöisinä työkaluina. Ideana on 1) parantaa aineistojen haku- ja selailuominaisuuksia ja mahdollistaa ihmisryhmien joustava muodostaminen prosopografista tutkimusta varten, 2) rikastaa elämäkerrallisia aineistoja toistensa avulla ja parantaa sekä monipuolistaa näin ihmisen lukukokemusta, 3) tarjota lukijalle välineet yksittäisen elämäkerran parempaan hahmottamiseen ja biografiseen tutkimiseen, 4) mahdollistaa ihmisryhmien ominaisuuksien prosopografista tutkimista ja vertailua, 5) muodostaa uutta tietämystä tekoälyn avulla sekä 6) avata mahdollisuus tutkia ja vertailla elämäkeroissa käytettyä kieltä.

### Esimerkkejä käytöstä

Biografiasampo on käytettävissä osoitteessa <http://biografiasampo.fi> ja sisältää kaikki SKS:n tietokantojen biografiat ilman maksumuuria. Seuraavassa havainnollistetaan Biografiasammon vision avaamia uusia mahdollisuuksia käytännön esimerkkien avulla.

#### 1) Hakuominaisuuksien parantaminen: henkilöt, ryhmät ja paikat

Aineistoista louhittu tietämysverkko mahdollistaa elämäkertojen ja ihmisryhmien haun joustavasti ns. fasettihaun avulla. Siinä elämäkertoja voi suodattaa perinteisen nimihaun lisäksi tekemällä valintoja ontologisten luokitusten avulla, joita ovat haluttu SKS:n tietokanta, linkitetty tietokanta, avainsanat, ajanjakso, sukunimi, biografia kirjoittaja, toimiala, yritys tai yhteisö, arvo, ammatti tai toiminta ja syntymäpaikka. Esimerkiksi Getty-säätiön historiallisten henkilöiden ULAN rekisteriin päässeet suomalaiset tai Helsingin yliopiston ylioppilasmatrikelissa 1853-1899 dokumentoidut henkilöt löytyvät yhdellä

<sup>5</sup> <https://fr.wikipedia.org/wiki/NouvelleBiographie%2C3%A9n%2C3%A9rale>

<sup>6</sup> <http://www.biografischportaal.nl/en>

<sup>7</sup> <https://sok.riksarkivet.se/Sbl/Start.aspx?lang=en>

klikkauksella fasetista ”linkitetty tietokanta”. Professorit ja muiden ammattien edustajat taas suodattuvat esiin vastaavalla valinnalla ammattien ja arvojen fasetista. Fasettivalintojen kohdalla olevat automaattisesti päivittyvät lukumäärät kertovat, kuinka monta elämäkertaa löytyy, jos tekee kyseisen valinnan. Näin lukija ei koskaan päädy tilanteeseen, jossa haun tulos on tyhjä joukko, mikä on perinteisiin hakumenetelmiin liittyvä harmillinen ominaisuus. Kuvassa 1 käyttäjä on valinnut ammatti tai arvo -fasetista kategorian ”professori” ja sitten linkitettyjen tietokantojen fasetista ”eduskunta”, jolloin tuloksena on 11 professoria, jotka ovat toimineet myös kansanedustajana.

The screenshot shows the Biografiasampo search interface. The search criteria are set to 'Professori' (Professor) and 'Eduskunta' (Parliament). The search results display 11 profiles, each with a photo, name, and biographical information. The profiles include: Jörn Donner, Arvo Salo, Unto Miettinen, Edwin Linkomies, Paavo Korpiainen, Oskari Reinikainen, Yrjö Sirola, and Väinö Voionmaa. Each profile includes a small bio, a photo, and links to various sources like Wikipedia, Wikidata, Fennica, and Geni.com.

**Kuva 1.** Professori-kansanedustajien elämäkertojen haku fasettihaulla.

Henkilöiden ja ihmisryhmien haun lisäksi Biografiasampo tarjoaa koko maapallon kattavan karttaperustaisen hakunäkymän, jossa elämäkertoista louhitut kymmenet tuhannet tapahtumat on projisoitu kartalle n. 2600 eri paikkaan. Kartalle saa näkyviin myös historiallisia karttoja ja paikkoja esimerkiksi luovutetun Karjalan alueelta.

## 2) Elämäkertojen ja lukukokemuksen rikastaminen

Biografiasampo rikastaa käyttäjän lukukokemusta eri tavoin dataa automaattisesti eri lähteistä linkittämällä. Kone myös päättelee algoritmien avulla uutta tietoa kokonaisuutta analysoimalla. Henkilöistä tarjotaan elämäkerrallisia kuvauksia SKS:n aineistojen lisäksi myös muista tietolähteistä, kuten Wikipediasta, Kirjasammosta, sukututkijoiden aineistoista (Geni.com) ja Getty-säätiön rekistereistä (ks. taulukko 2). Ajatuksena on rekonstruoida henkilöiden elämätarinat mahdollisimman monen eri tietolähteen tietoja yhdistämällä ja luomalla jokaiselle henkilölle automaattisesti ”kotisivut”. Järjestelmän tekoäly ikäänkuin lukee elämäkerrat ja tutkii niihin liittyvät tietokannat ja tarjoaa tästä jäsenetyn yhteenvedon lukijalle edelleen tutkittavaksi. Ilman tietokoneen apua näin laajojen ja heterogeenisten aineistojen lukeminen ja jäsentäminen olisi erittäin työlästä tai mahdotonta.

	Tietolähde/palvelu	Linkit	Kuvaus/osoite
1	Wikipedia	5808	<a href="http://fi.wikipedia.org">http://fi.wikipedia.org</a>
2	Wikidata	6423	<a href="http://www.wikidata.org">http://www.wikidata.org</a>
3	Fennica	4007	Kansalliskirjaston kansallisbibliografia, linkitetyn datan palvelu
4	Sotasampo	287	<a href="http://sotasampo.fi">http://sotasampo.fi</a>
5	Vanhat Norssit	221	<a href="http://www.norssit.fi/semweb">http://www.norssit.fi/semweb</a>

6	Kirjasampo	715	<a href="http://kirjasampo.fi">http://kirjasampo.fi</a>
7	Biografiskt lexicon för Finland	1543	Ruotsinkieliset biografiat (SLS), <a href="http://www.blf.fi">http://www.blf.fi</a>
8	ULAN-ontologia	204	<a href="http://www.getty.edu/research/tools/vocabularies/ulan">http://www.getty.edu/research/tools/vocabularies/ulan</a>
9	VIAF	2696	Virtual International Authority Files, <a href="http://viaf.org">http://viaf.org</a>
10	Geni.com	5320	Kansainvälinen sukututkijoiden sukupuu-palvelu
11	Kotisivu	43	Henkilökohtaiset kotisivut
12	Eduskunta	637	Kansanedustajat 1917-2018
13	Ylioppilasmatrikkeli 1853-1899	379	Helsingin yliopiston tiedekuntien opiskelijat ja jäsenet
	<b>Yhteensä</b>	<b>26568</b>	

**Taulukko 2.** Biografiasammon elämäkertoihin linkitetyt muut tietolähteet, joita voi käyttää fasettihaun kriteerinä.

**Analyysinäkömät (5)**

**Linkit tietolähteisiin (7)**

**Biografianäkymät (6)**

**Kansallisbiografian artikkeli**

**Suositus-linkit muihin biografioihin**

**Kuva 2.** Eliel Saarisen kotisivusto Biografiasammossa.

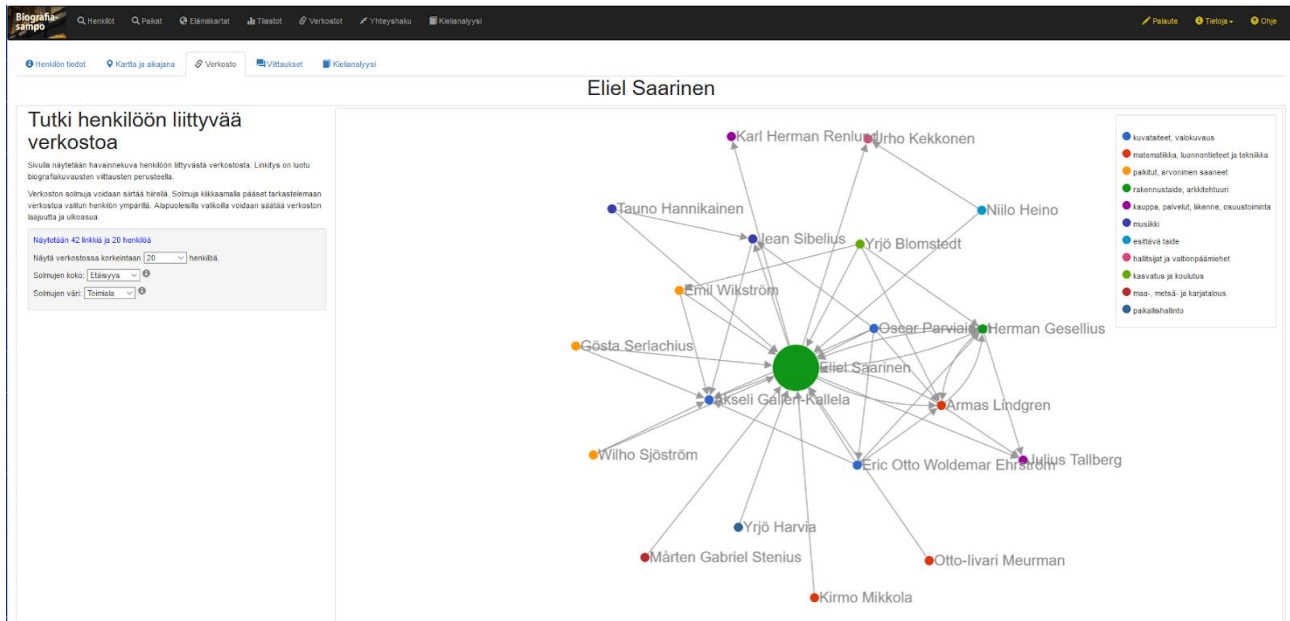
Kuvassa 2 näkyy esimerkkinä Eliel Saarisen automaattisesti muodostettu kotisivu. Ylärivissä on avattavissa viisi erilaista välilehteä, joiden kautta voi tarkastella Saarisen elämää eri tavoilla: tietosivuna (valittuna kuvassa), elämän tapahtumat kartalla ja aikajanalla, Saarisen egosentrinen verkosto perustuen elämäkertojen välisiin viittauksiin, viittausanalyysi muihin elämäkertoihin ja SKS:n elämäkerran kielianalyysi.

Saarisen eri lähteistä löydettyjen kuvien ja perustietojen alla on linkkejä erilaisiin tietolähteisiin, joissa Saarisenestä on tietoa, tässä tapauksessa seitsemään eri verkkopalveluun Suomessa ja ulkomailla. Hänestä löytyy Biografiasammossa kuusi erilaista biografista kuvausta, jotka ovat luettavissa eri välilehtien avulla: Biografiakeskuksen artikkeli alkuperäisenä ja uusilla linkeillä rikastettuna, tieto Linked Data Finland -datapalvelussa, Geni.com palvelun kotisivun sukupuunäkymä ja Wikipedian artikkeli. Näistä alkuperäinen Kansallisbiografian verkkopalvelussa oleva artikkeli on kuvassa avattu luettavaksi. Kuvan vasemmassa reunassa on vielä tietokoneen luomia erilaisia suosituslinkkejä Eliel Saariseen eri tavoin liittyviin muihin sivuihin, esimerkiksi toisiin sisällöltään samankaltaisiin elämäntarinoihin, jotka on tunnistettu automaattisesti niiden sanaston ja merkityksen perusteella.

### 3) Biografisen tutkimus

Käyttäjälle tarjotaan elämäkertatekstien ohella joustava ja helppokäyttöinen data-analyttinen välineistö (kuvan 1 analyysinäkömät), jolla voi tutkia ja visualisoida yksittäisen henkilön elämäntarinaa. Esimerkiksi kuvassa 3 näkyy Eliel Saarisen ns. egosentrinen verkko, josta näkyvät hänen yhteytensä muihin

elämäkertojen päähenkilöihin artikkeleissa tehtyjen viittausten perusteella, kuten Hvitträskin toimiston arkkitehtikollegoihin Herman Gesellius ja Armas Lindgren. Valitsemalla toisen välilehden, visualisoituvat hänen kansainvälisen elämänsä tapahtumat tyypeittäin jaoteltuna kartalla ja aikajanalla. Biografiasammon kieliteknologinen tekoäly louhii tapahtumat automaattisesti elämäkerta lukemalla ja muista tiedoista.

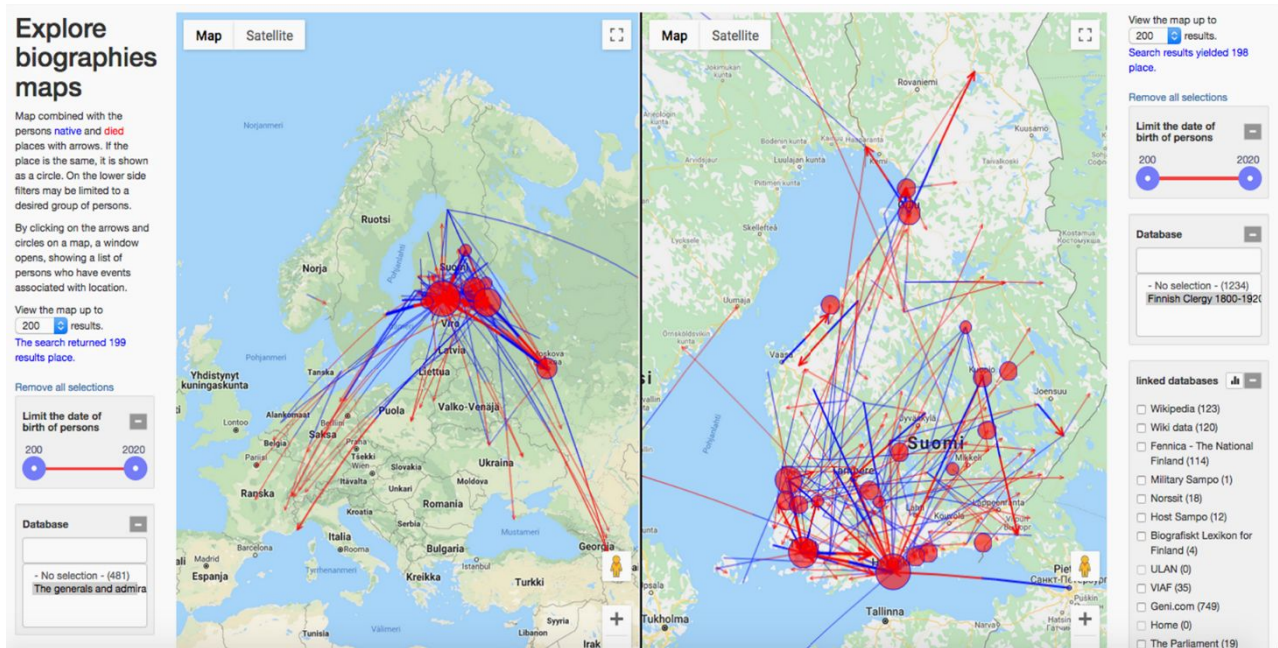


**Kuva 3.** Eliel Saarisen egosentrinen verkko, joka on tässä rajoitettu 20 henkilöön. Henkilöiden toimialat on merkitty eri väreillä.

#### 4) Prosopografisen tutkimus ja tilastot

Samalla tavalla voidaan Biografiasammonsa tutkia ja verrata toisiinsa henkilöryhmien toimintaa prosopografian keinoin. Voidaan esimerkiksi luoda ja visualisoida Biografiasampon linkitettyjen Norssin oppilaiden 1867-1992 myöhemmin muodostamia verkostoja SKS:n aineistojen sisäisten viittausten perusteella. Esiin nousee silloin mm. erillinen piiri filosofi Eino Kailan ympärillä sekä luokkatoverien Matti Klinge, Pentti Saarikoski ja Anto Leikola verkosto. Näin Biografiasampo itseasiassa automaattisesti rikastaa v. 2017 julkaistun Vanhat Norssit semanttisessa webissä –palvelun<sup>8</sup> toimintoja pelkästään palveluiden henkilöiden välisen siltauksen avulla. Kuvassa 4 käyttäjä vertaa toisiinsa kahden ihmisryhmän elämänlankoja, Suomen suuriruhtinaskunnan 1809-1917 amiraaleja ja kenraaleja (vasemmalla) ja papistoa 1800-1920 (oikealla). Ryhmät on muodostettu kahdella rinnakkaisella fasettihaulla Biografiasammon elämäkartojen vertailunäkymässä, jossa elämä kuvataan sinipunaisena nuolena syntymäpaikasta (sininen pää) kuolinpaikkaan (punainen pää). Yhdellä vilkaisulla selviää, että sotilaat liikkuvat pappeja kansainvälisemmin ja kohti etelää. Yhtä kaarta kartalla klikkaamalla pääsee käsiksi kaareen liittyviin elämäkertoihin tarkempaa tutkimusta varten. Esimerkiksi vasemmalla näkyvä poikkeava kaari Oulusta Länsi-Siperiaan osoittautuu Siperiaan maanmittaustöiden johtajaksi nimitetyn Gustav Adolf Silverhjelmin aiheuttamaksi.

<sup>8</sup> <http://www.norssit.fi/semweb>



**Kuva 4.** Suomen suuriruhtinaskunnan kenraalien ja amiraalin (vasemmalla) ja papiston (oikealla) elämänkaarien prosopografinen vertailu Biografiasammossa.

Biografiasampoon sisältyy lisäksi koko joukko tilastollisia näkymiä ihmisryhmien analyysejä varten. Pylväsdiagrammi-näkymän avulla selviää mm. faseteilla valitun henkilöryhmän ikä-jakautuma, lasten ja puolisoitten lukumäärät sekä elinajat ja iät. Piirakkadiagrammien avulla taas saa kuvan henkilöryhmän jakautumisesta sukupuolen, ammatin, toimialan, ammatin tai arvon sekä yrityksen tai yhteisön mukaan. Myös kuvan 3 kaltaiset eri ryhmien väliset tilastolliset vertailut ovat mahdollisia omilla välilehdillä.

### 5) Tietämyksen muodostus tekoälyllä

Semanttisen webin ja tekoälyn suuria lupauksia on uuden, yllätyksellisen tietämyksen automaattinen löytäminen linkitetyistä aineistoista, ns. serendipiteetti. Linkitetyn datan kokonaisuus on usein jo sellaisenaan enemmän kuin osiensa summa. Tarkasti määritelty tietämysverkon looginen semantiikka mahdollistaa lisäksi uusien tietojen välisten yhteyksien päättelyn ja lisäämisen verkkoon. Biografiasammossa tätä on hyödynnetty mm. fasettihaun ontologioissa, jossa vaikkapa ”Saksassa” syntyneitä etsittäessä löytyy Friedenauissa syntynyt sodanaikaisen Päämajan saksalainen kenraali Waldemar Erfurth, koska Friedenau on osa Berliiniä, joka puolestaan on osa Saksaa. Järjestelmään sisältyvässä erillisessä Yhteyshaku-sovelluksessa puolestaan tavoitteena on päätellä ja löytää henkilöiden ja paikkojen välisiä yhteyksiä ja muodostaa niille suomenkieliset selitykset. Esimerkiksi fasettivalinnoilla ”Italia” ja ”taidemaalari” löytyvät suomalaisten taidemaalarien erilaiset semanttiset yhteydet Italiaan, kuten että Elin Danielson-Gambogi vastaanotti Firenzen kaupungin taidepalkinnon vuonna 1899 ja että ”Robert Wilhelm Ekman on luonut vuonna 1844 taideteoksen 'Maisema Subiacosta', joka kuvaa paikkaa Italia”. Jälkimmäisessä tapauksessa yhteys on muodostunut Biografiasampoon yhdistettyjen Ateneumin kokoelmatietojen kautta. Ateneumin lisäksi yhteyshaussa hyödynnetään lisäksi kokoelmadataa kansallisbibliografia Fennicasta<sup>9</sup>, Kirjasammosta<sup>10</sup>, historiallisia tapahtumia kuvaavasta HISTO-ontologiasta<sup>11</sup> sekä J. V. Snellmanin kootuista teoksista<sup>12</sup>, joista luotiin linkitetyn datan versio.

<sup>9</sup> <https://www.kansalliskirjasto.fi/fi/palvelut/metadatan-muunto-ja-valityspalvelut/avoin-data>

<sup>10</sup> <https://www.kirjasampo.fi/>

<sup>11</sup> <https://seco.cs.aalto.fi/ontologies/histo/>

<sup>12</sup> <http://snellman.kootutteokset.fi/>

## 6) Tekstien kielellinen analyysi

Biografioita voidaan tutkia Biografiasammossa paitsi henkilöiden, tapahtumien ja paikkojen myös teksteissä käytetyn kielen kautta. Aineistoista on muodostettu erillinen yli 100 miljoona yhteyttä sisältävä kielellinen tietämysverkko, jossa tekstien sanamuodot on perusmuotoistettu, analysoitu ja lauseet jäsennetty. Tämän datan avulla järjestelmään on toteutettu mm. työkalu, jolla voidaan tutkia ja verrata eri henkilöryhmien biografioissa käytettyjä sanoja. Esimerkiksi itsenäisyyden ajan nais- ja mieskansanedustajien elämäkertoja voidaan verrata vastaavatyypillisellä vertailunäkymällä kuin kuvassa 4. Näkymä luo tilaston sanojen käytöstä ja kertoo että esimerkiksi sanat ”perhe” ja ”lapsi” ovat naisilla erittäin yleisiä naisia kuvaavissa elämäkerroissa, mutta eivät juuri esiinny miesten kuvauksissa.

### Lähdekritiikkiä tarvitaan

Biografiasammon aineistot perustuvat asiantuntijoiden laatimiin kirjoituksiin, mutta tietojen rakenteistamisen, koostamisen, yhdistelyn, rikastamisen ja uuden tiedon muodostamisen on tehnyt etupäässä tietokone. Koska aineistot ovat laajoja, ei kokonaisuuden virheettömyyttä voida tarkistaa käsin kuin testimielessä sieltä täältä. Linkityksistä löytyy siksi enemmän virheitä, kuin jos se olisi tehty käsityönä, ja tilastollisten ja verkostanalyysien johtopäätösten kanssa on syytä olla tarkkana ja pitää mielessä, mihin dataan ja laskentamenetelmään ne perustuvat. Tämä on tyyppillistä digitaalisissa ihmistieteissä, jossa käsitellään usein niin laajoja aineistoja, ettei systemaattinen ihmistyö ole mahdollista aineistoja muodostettaessa. Laskennallisten tekniikoiden lisäarvo on kuitenkin erityisen suuri tällaista suurdataa (big data) käsiteltäessä ja puutteellinenkin tulos monasti parempi kuin ei mitään tulosta.

Järjestelmien käyttäjältä kuitenkin edellytetään erityisen tarkkaa lähdekriittistä ymmärrystä siitä, millaista dataa tietoaaineisto ja sovellus oikeastaan sisältävät, onko tieto missä määrin epätasällista tai puutteellista ja millaisia oletuksia järjestelmässä käytettävät ontologiat ja menetelmät kenties tekevät. Esimerkiksi faseteissa käytetyssä historiallisten paikkojen ontologiassa luovutetun Karjalan paikat eivät löydy Suomen alta, vaikka niiden avulla kuvatut tapahtumat yleensä liittyvä aikaan, jollain alue vielä oli osa Suomea.

Huomattava on myös esimerkiksi se, että Biografiasammon verkostanalyysien perustana ovat Biografiakeskuksen elämäkertojen väliset linkit, jotka ovat aineistojen kirjoittajien ja toimittajien luomia aineistoja kirjoitettaessa. Tällöin linkki ei tarkoita sitä, että yhdistetyt henkilöt esimerkiksi olisivat välttämättä edes tienneet toisistaan. Esimerkiksi Blanka Namurilaisen (Ruotsin kuningatar 1318-1363), Ruotsin ja Norjan kuninkaan Maunu Eerikinpojan ranskalaisen puolison verkosto kuvaa paitsi hänen lähipiiriään myös hänen jälkimainettaan. Oman aikansa ruotsalaisen ylimystön lisäksi kuningatar Blankan egosentrisen verkoston keskeisiä linkejä ovat viisisataa vuotta myöhemmin eläneet Zachris Topelius ja Albert Edelfelt. Topeliuksen Lukemisia lapsille ja Edelfeltin maalaus tekivät hänestä 1800-luvun Suomessa yhden tunnetuimmista ruotsalaisista kuningattarista. Yllättävä yhteys keihäänheitäjä Tapio Rautavaaran ja runoilija Aale Tynnin välille puolestaan syntyy siksi, että molemmat voittivat olympiakultaa Lontoossa, Rautavaara keihäänheitossa ja akateemikko Tynni lyriikassa, joka oli vielä tuolloin olympialaji. Tällaisten yhteyksien luonteen tarkempaa selvittämistä varten voidaan käyttää hyväksi Biografiasammon valmiiksi esiin louhimia lauseita, joissa linkit esiintyvät ja selittyvät.

Biografiasammon kansainvälisesti uudenlainen julkaisukonsepti ja -järjestelmä<sup>13</sup> on kehitetty Aalto-yliopiston tietotekniikan laitoksen ja Helsingin yliopiston Digitaalisten ihmistieteiden keskuksen HELDIG Semanttisen laskennan tutkimusryhmässä (SeCo) yhteistyössä Suomalaisen Kirjallisuuden Seuran kanssa. Työ on osa laajempaa kieliteknologian ja semanttisen laskennan tutkimusta LODI4DH, jonka tavoitteena on kansallisen tietoinfrastruktuurin luominen digitaalisten ihmistieteiden tutkimukseen<sup>14</sup>. Biografiasampo on jatkoa SeCo-ryhmän aiemmille kansainvälisesti ja kotimaassa palkituille

<sup>13</sup> <http://seco.cs.aalto.fi/projects/biografiasampo>

<sup>14</sup> <http://seco.cs.aalto.fi/projects/lodi4dh>

"sampo-järjestelmille" Kulttuurisampo (2008)<sup>15</sup>, Kirjasampo (2011)<sup>16</sup>, (2 miljoonaa käyttäjää 2017) Matkailusampo (2011) ja Sotasampo (2015)<sup>17</sup> (130 000 käyttäjää vuonna 2017).

### **Tutkimusta Biografiasammosta**

Artikkelit ovat avoimesti saatavilla osoitteessa <http://seco.scs.aalto.fi/prorojects/biografiasampo>.

Eero Hyvönen, Petri Leskinen, Minna Tamper, Jouni Tuominen and Kirsi Keravuori: Semantic National Biography of Finland. Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018), pp. 372-385, CEUR Workshop Proceedings, Vol-2084, Helsinki, Finland, March, 2018.

Eero Hyvönen and Heikki Rantala: Knowledge-based Relation Discovery in Cultural Heritage Knowledge Graphs. 2019. Submitted for review.

Petri Leskinen, Eero Hyvönen and Jouni Tuominen: Analyzing and Visualizing Prosopographical Linked Data Based on Biographies. Proceedings of the Second Conference on Biographical Data in a Digital World 2017 (BD2017), vol. 2119, pp. 39-44, CEUR Workshop Proceedings, Linz, Austria, 2018.

Petri Leskinen, Goki Miyakita, Mikko Koho and Eero Hyvönen: Combining Faceted Search with Data-analytic Visualizations on Top of a SPARQL Endpoint. Proceedings of VOILA 2018, Monterey, California, CEUR Workshop Proceedings, Vol. 2187, 2018.

Minna Tamper, Petri Leskinen, Kasper Apajalahti and Eero Hyvönen: Using Biographical Texts as Linked Data for Prosopographical Research and Applications. Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. 7th International Conference, EuroMed 2018, Nicosia, Cyprus, Springer-Verlag, 2018.

Jouni Tuominen, Eero Hyvönen and Petri Leskinen: Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research. Proceedings of the Second Conference on Biographical Data in a Digital World 2017 (BD2017), vol. 2119, pp. 59-66, CEUR Workshop Proceedings, Linz, Austria, 2018.

### **Lisätietoa suomeksi semanttisesta webistä ja linkitetystä datasta**

Eero Hyvönen: Semanttinen web. Linkitetyn avoimen datan käsikirja. Gaudeamus, 271 ss., 2018.

---

<sup>15</sup> <http://kulttuurisampo.fi>

<sup>16</sup> <http://kirjasampo.fi>

<sup>17</sup> <http://sotasampo.fi>