

Historiallinen paikkatieto semanttisessa webissä – Biografiasampo.fi

Semanttinen web tarjoaa lupaavan lähestymistavan tiedon esittämiseen historiallisissa paikannimirekistereissä ja rekisterien julkaisemiseen verkkopalveluina sekä ihmisten että koneiden käyttöön. Esimerkki älykkäästä digitaalisesta kulttuurihistoriallisesta sovelluksesta on Suomen kansallisia elämäkertakokoelmia julkaiseva ja rikastava Biografiasampo.fi.

Historiallisen paikkatiedon käsittely on haasteellista: ajan kuluessa paikkojen nimet vaihtuvat, paikkojen sijainti, alue ja tyyppi voivat muuttua ja paikka voi kokonaan hävitä muinaisen Karthagon tapaan. Lisäksi tietomme historiallisista paikasta on epävarmaa (käytiinkö taistelu Vuoksen tuolla tai tällä puolella?), luonteeltaan sumeaa (mistä kohdasta alkaa Korvatunturi?) ja puutteellista (missä on antiikin maantieteilijän Pytheaan tarkoittama Ultima Thule?). Tilannetta ei helpota se, että paikoista on yleensä ollut käytössä monenlaisia ja erikielisiä nimiä.

Ontologiapalvelut edistävät yhteentoimivuutta

Semanttisessa webissä paikannimirekisterit (gazetteer) julkaistaan ns. ontologioina. Niissä jokaiselle paikalle on luotu yksilöivä HTTP URI -tunniste, johon liitetyt ominaisuudet määrittelevät paikan tiedot kuten eri nimet, tyyppi, koordinaatit, laajemman alueen ja paikan osat. Paikat ominaisuuksineen muodostavat verkkomaisen hierarkkisen tietämysgraafin (knowledge graph),

joka esitetään semanttisen webin perustana olevan RDF-tietomallin ja standardien avulla.

Olellainen idea on, että HTTP URI -tunniste toimii sekä yksilöivänä tunnisteena että URL-osoitteena, josta käyttäjä voi saada lisätietoa ko. paikasta Tim Berners-Leen linkitetyn datan periaatteiden mukaisesti. Se, mitä tietoa paikkaontologiapalvelusta milloinkin palautetaan, esimerkiksi paikan kuvaava RDF-muotoinen data selaimen JavaScript-sovellukselle tai tästä ihmisen tutkittavaksi muodostettu HTML-sivu, hoidetaan HTTP-protokollan mukaisen sisältöneuvottelun (content negotiation) avulla.

Data rikastuu linkittämällä ja päättelemällä

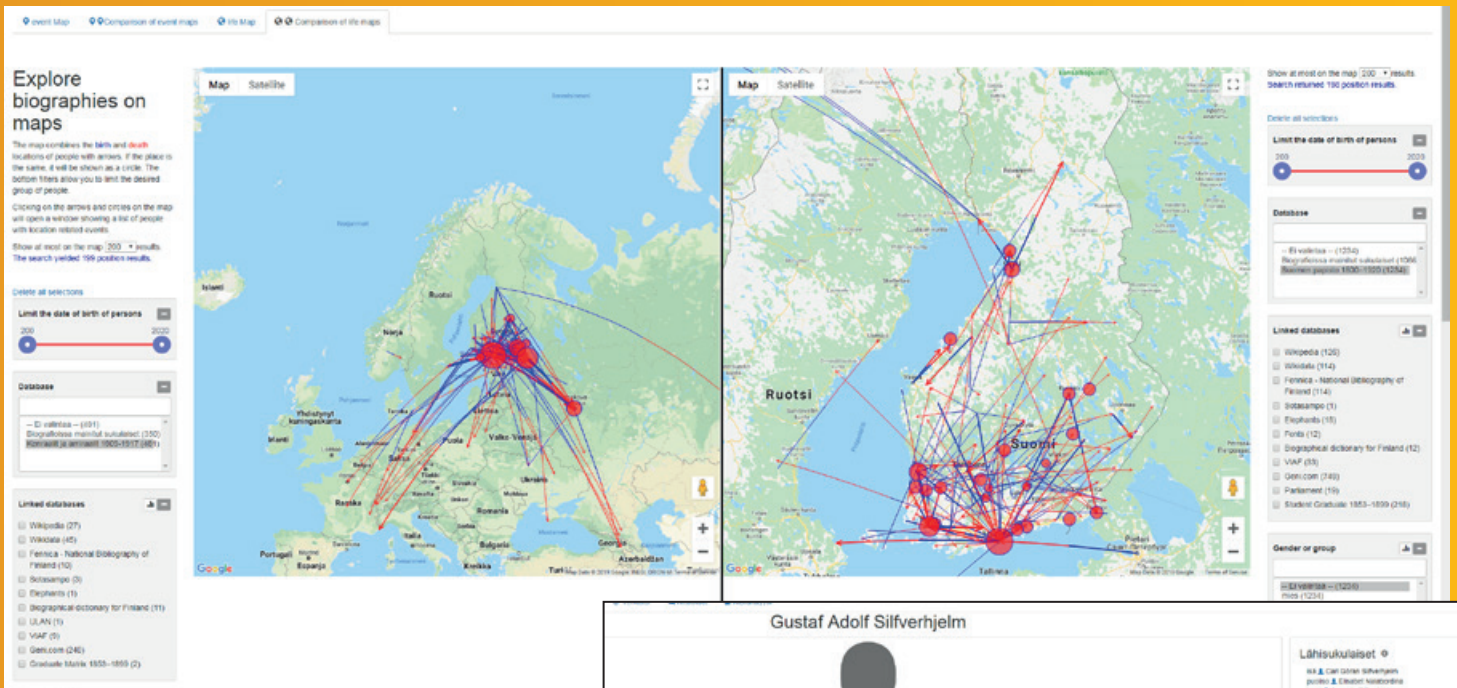
Ontologiapalveluiden keskeisenä tavoitteena on edistää paikkatietoaineistojen yhteentoimivuutta. Eri datalähteistä saatava paikkatieto saadaan linkittymään toisiinsa käyttämällä ontologioissa yhteisesti sovittuja tai niihin sillattuja (map, align) tunnisteita.

Esimerkiksi Wikidatan ja Wikipedioista louhitun DBpedian linkitetyn datan palvelut sisältävät valtavan määrän paikkoja, jotka on sillattu muun muassa eri maiden paikannimirekisteristä tuotettuun GeoNames-rekisteriin ja eri sovellusalojen paikkatietoaineistoihin maailmanlaajuisessa Linked Open Data -pilvessä. Ontologioiden toinen tavoite on datan sisällön rikastaminen datajoukkojen keskinäisellä linkityksellä sekä päättelyllä, joka perustuu semanttisen webin standardien taustalla olevaan formaaliin predikaattilogiikkaan.

Kansallisbiografia Biografiasammon ytimessä

Syksyllä 2018 julkistetussa semanttisen webin Biografiasampo.fi-palvelussa ydinaineistona on Suomalaisen Kirjallisuuden Seuran (SKS) toimittama Kansallisbiografia ja muut 13 100 elämä-

EERO HYVÖNEN ON HELSINGIN YLIOPISTON DIGITAALISTEN IHMITIETEIDEN KESKUKSEN JOHTAJA JA AALTO-YLIOPISTON TIETOTEKNIIKAN LAITOKSEN PROFESSORI. HÄN ON JULKAISSUT YLI 400 ARTIKKELIA JA KIRJAA TEKOÄLYYN JA SEMANTTISEEN WEBIIN LIITTYEN. UUSIMPANA SUOMEKSI "SEMANTTINEN WEB: AVOIMEN LINKITETYN DATAN KÄSIKIRJA" (GAUDEAMUS, 2018). S-POSTI: ETUNIMI.SUKUNIMI@AALTO.FI




Venäjänsotavoimissa palvelleiden kenraalien ja amiraalien (vasemmalla) ja papiston (oikealla) elämäntietojen prosopografinen vertailu Biografiasampo.fi-palvelussa.

kertaa, joista on louhittu tekoälyn kielitekniikan ja tietämyksen irrotusmenetelmien (knowledge extraction) avulla 120 miljoonaa yhteyttä sisältävä linkitetyn datan tietämysgraafi.

Paikkaontologian perustaksi otettiin Kansalliskirjaston YSA-sanastosta irrotetuista paikannimistä muodostettu ontologia YSO-paikat, jonka fragmentaarisia hierarkioita muokattiin, täydennettiin ja laajennettiin mm. Sotasampo.fi-palvelun luovutetun Karjalan paikoilla. Paikkaontologia yhdistää monenlaisia Biografiasammossa yhdistyviä, toisiaan rikastuttavia aineistoja. Hierarkkista paikkaontologiaa on hyödynnetty mm. älykkäässä fasettihaussa, jossa vaikkapa ”Saksassa” syntyneitä henkilöitä etsittäessä löytyy Friedenaussa syntynyt sodanaikaisen Päämajan saksalaisen yhdysesikunnan kenraali Waldemar Erfurth, koska Friedenau on osa Berliiniä, joka puolestaan on osa Saksaa.

Järjestelmään sisältyvässä Yhteyshaku-sovelluksessa tavoitteena on päätellä ja löytää henkilöiden ja paikkojen välisiä yhteyksiä ja muodostaa niille suomenkieliset selitykset. Esimerkiksi fasettilinnoilla ”Italia” ja ”taidemaalari” löytyy tieto, 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 taidemuseon kokoelmätietojen kautta. Ateneumin lisäksi yhteyshaussa hyödynnetään

Gustaf Adolf Silfverhjelm



1815–1899

Lähsukäsitteet

- & Carl Oskar Silfverhjelm
- & Elin Danielson-Gambogi
- & August Silfverhjelm
- & Nanna Silfverhjelm

Samanlaisia biografioita

- Carl Oskar Silfverhjelm
- Nanna Silfverhjelm
- August Silfverhjelm
- Elin Danielson-Gambogi
- Arthur Andersson
- Carl Axel Egeberg
- Olof Leonard von Bons

Gustaf Adolf Silfverhjelm tuli kersantiksi 1. Suomen jääkäriyhtymän Turun pataljoonan Hulttisten komppaniaan 13.7.1815. Hän opiskeli Haapaniemen suomalaisessa topografikoulussa Rantasalmella 17.2.1817–29.5.1819. Silfverhjelm ylennettiin vänrikiksi H. M. Keisarisen seurueen majoitusmestariosastoon (myöhemmin yleisesikunta) 24.2.1820. Hänet ylennettiin aliluutnantiksi 1823, luutnantiksi 1826, alikapteeniksi 26.4.1829 ja kapteeniksi 22.4.1832. Silfverhjelm määrättiin majoitusmestariksi 29. joulukuuta 1833 ja ylennettiin everstiluutnantiksi 4.5.1834. Hänet siirrettiin päämajajoukkojen Siperiaan olevaan jälkiväkijoukkoon 4.1.1835. Silfverhjelm ylennettiin everstiksi 6.4.1842 ja kenraalimajuriksi 18.12.1851. Silfverhjelm nimettiin 5.4.1859 maamittauslaitoksen johtajaksi Länsi-Siperiaan, missä tehtävässä hän toimi kuolemaansa saakka.

Elisabet Nalebordin. PV N. N. Nalebordin

eversti, vapaaherra Carl Oskar Silfverhjelm S 16.8.1759 Örebro, Ruotsi, K 28.11.1808 Oulu ja vapaaherra Ulrik Gustaf Lovén S 16.5.1774 Tukholma, K 17.8.1831 Pori.

Elisabet S 1641, K 1861, vapaaherra; Natalia S 1842, vapaaherra; Augusta, vapaaherra.

kokoelmadataa kansallisbibliografia Fennicasta, Kirjasammosta, historiallisia tapahtumia kuvaavasta HISTO-ontologiasta sekä **J. V. Snellmanin** kootuista teoksista, joista luotiin linkitetyn datan versio.

Visualisointeja paikkaontologialla

Paikkaontologian avulla tietoa voidaan myös visualisoida kartoilla. Mukaan on liitetty historialliset Karjalan kartat ja Senaatin kartasto. Esimerkiksi suomalaisia sotilaita ja pappeja kuvaavat ryhmät kartalla on muodostettu kahdella rinnakkaisella fasettihaulla Biografiasammossa elämäkertojen vertailunäkymässä, jossa elämä kuvataan sinipunaisten nuolien avulla syntymäpaikasta kuolinpaikkaan.

Yhdellä vilkaisulla selviää, että sotilaat liikkuvat pappeja kansainvälisemmin ja vanhemmiten kohti etelää kuten eläkeläiset nykyään. Yhtä kaarta kartalla klikkaamalla pääsee käsiksi kaareen liittyviin elämäkertoihin. Esimerkiksi poikkeava elämänlanka Oulusta Länsi-Siperiaan osoittautuu Siperiaan ▶

KUVA: HELJÄ KORKIALA



Biografiasammon paikatieto-ontologian perustana toimii Kansalliskirjaston yleisen suomalaisen asiasanaston (YSA) paikannimistä muodostettu ontologia YSO-paikat.

maanmittaustöiden johtajaksi nimitetyn **Gustav Adolf Silverhjelmin** aiheuttamaksi.

Haasteena ajallisen muutoksen hallinta

Keskeinen ontologinen haaste historiallisten paikkojen esittämisessä on, että ajan kuluessa paikkojen ominaisuudet muuttuvat.

Jos haetaan vaikkapa Suomessa julkaistuja kirjoja, pitäisi Viipurissa painetut kirjat tulla hakutulokseen vain sotia edeltävältä ajalta.

Jo Suomen kuntien lähihistorian ajallinen paikkaontologia on varsin monimutkainen johtuen muutoksista kuntien laajentuessa, yhdistyessä, jakautuessa ja valtiorajojen muuttuessa.

Yhden hetkellisen ontologisen kuvauksen, kuten nykyisen Maanmittauslaitoksen Paikannimirekisterin (PNR) sijasta tarvitaan ontologioiden aikasarja, jossa on uusi paikkahierarkia periaatteessa jokaisen muutoskohdan jälkeen.

Aalto-yliopiston ja Geologian tutkimuskeskuksen (GTK) kehittämä Suomen ajallinen paikkaontologia SAPO kuntamuutoksista on yksi tapa lähestyä ongelmaa. Siinä luotiin uusi URI ja identiteetti jokaiselle ajalliselle erilaiselle kunnalle Suomessa 1800-luvulta lähtien aina, kun kunnan alue muuttui. Lisäksi luotiin oma tunniste koko kunnan käsitteelle eri aikoina. Käytettävissä ei ollut juurikaan tietoa kuntien polygoneista eri aikoina vaan ainoastaan kirjoissa dokumentoitua tietoa pinta-aloista ja muutoksista, kuten kuntaliitoksista. Tiedon perusteella muodostetusta ontologiasta voitiin kuitenkin päätellä alueellisia kattavuuksia tiedonhaku varten, kuten että vanhan Viipurin (ennen vuotta 1906) alueesta noin 12 prosenttia kuuluu edelleen Suomeen, nykyiseen Lappeenrantaan.

Suomalaista historiallisten paikkojen ontologiaa ja siihen liittyvää historiallisten karttojen palvelua

olla tutkimassa ja kehittämässä Aalto-yliopiston ja Helsingin yliopiston Digitaalisten ihmistieteiden keskuksen HELDIG yhteistyönä osana Linked Open Data Infrastructure for Digital Humanities in Finland (LODI4DH) kokonaisuutta.

Varteenotettava vaihtoehto ajallisen dimension tuomiseksi paikkaontologiaan on liittää sen ominaisuuksiin lisätietoa siitä, milloin kyseinen

Hetkellisen ontologisen kuvauksen, kuten Paikannimirekisterin, sijaan tarvitaan ontologioiden aikasarja.

ominaisuus on ollut voimassa, esimerkiksi milloin Pietari oli Venäjän pääkaupunki. Näin on menetelty mm. semanttisesti rikkaimmassa kansainvälisessä historiallisten paikkojen rekisterissä, Getty-säätiön Thesaurus of Geographical Names (TGN) -ontologiassa, joka nykyisin julkaistaan avoimena linkitettyinä datana verkossa. Suomessa kehitettävissä ontologioissa on hyvä huomioida kansainvälinen yhteentoimivuus paitsi W3C:n yleisiin semanttisen webin standardeihin myös TGN:n kaltaisiin sovelluslakohtaisiin de facto -standardeihin. ◀