

Sotasampo - talvi- ja jatkosota semanttisessa webissä

Eero Hyvönen

Semanttisen laskennan tutkimusryhmä (SeCo)

Aalto-yliopisto, tietotekniikan laitos ja

Helsingin yliopisto, Digitaalisten ihmistieteiden keskus HELDIG

<http://seco.cs.aalto.fi>, <http://heldig.fi>

G. W. F. Hegelin (1770–1831) mukaan opimme historiasta vain sen, että emme opi historiasta mitään. Sotasampo-järjestelmän tavoitteena on varmistaa, ettei näin käy Suomen historian suurimman tragedian, talvi- ja jatkosodan osalta. Mitä enemmän tiedämme sodan uhreista, sitä vähemmän sotia toivottavasti jatkossa syttyy. Sotasampo rakentaa rauhaa julkaisemalla toisen maailmansodan kansallisia tietoaaineistoja verkossa linkitetyn avoimen datan palveluna, jonka rajanpintojen varaan voidaan kehittää älykkäitä verkkosovelluksia. Järjestelmä on ensimmäinen laatuaan maailmassa ja sai vuonna 2017 arvostetun kansainvälisen LODLAM Challenge Open Data Prize -palkinnon.

Sotasampo-järjestelmä koostuu 1) linkitetyn avoimen datan (Linked Open Data) palvelusta, joka on tarkoitettu sovellusten kehittäjien käytettäväksi sekä 2) useita sovelluksia sisältävästä semanttisesta portaalista Sotasampo.fi, joka perustuu em. datapalveluun. Sotasampo julkistettiin 2015 ja se sai kolmessa päivässä yli 10 000 käyttäjää. Uusin sovellus ”Sankarihautausmaat” julkaistiin Suomen 100. itsenäisyyspäivän kunniaksi marraskuussa 2017 ja sai heti 57 000 käyttäjää.

Linkitetyn avoimen datan palvelu

| Datan luovuttaja | Aineisto (25.11.2015) | Määrä | Datan lähteet ja tekijänoikeus | Lisenssi linkitettyyn dataan |
|--|---|------------------|--|-------------------------------|
| 1 Puolustusvoimien SA- | TK-kuvat ja videot | 160 000 | Puolustusvoimat & Aalto/SeCo | CC BY 4.0 |
| 2 Kansallisarkisto | Menehtyneet 39-45 | 95 000 | Kansallisarkisto & Aalto/SeCo | CC BY 4.0 |
| 3 Kansallisarkisto | Organisaatiokortisto | 500 | Kansallisarkisto & Aalto/SeCo | CC BY 4.0 |
| 4 Kansallisarkisto | Sotapäiväkirjat | 26 500 | Kansallisarkisto & Aalto/SeCo | CC BY 4.0 |
| 5 Kansallisarkisto | Senaatin kartat | 404 | Kansallisarkisto & Aalto/SeCo | CC BY 4.0 |
| 6 Suomen Sotahistoriallinen Seura Bonnier Publications Oy | Kansa taisteli -lehtien artikkelit 1957-1986 | 3 360 | Metadata: Timo Hakala & Aalto/SeCo. Artikkelit © Suomen Sotahistoriallinen Seura ja Bonnier, http://sshs.fi . | CC BY 4.0 (metadata) |
| 7 Maanmittauslaitos | Paikannimirekisteri | 800 000 | Maanmittauslaitos (MML) & Aalto/SeCo | CC BY 4.0 |
| 8 Maanmittauslaitos | Karjalan topograf. kartat | 47 | MML & Aalto/SeCo | Ei-kaupallisiin tarkoituksiin |
| 9 Jyrki Tiittanen | Karjalan paikat geokood. | 35 000 | Jyrki Tiittanen & Aalto/SeCo | CC BY 4.0 |
| 10 Aalto-yliopisto | Sotien tapahtumat | 11 275 | Aalto/SeCo. Lähteet: Talvi- ja jatkosodan pikkujättiläiset, organisaatiokortisto ym. | CC BY 4.0 |
| Aalto-yliopisto | Henkilöt (Menehtyneet 39- 45 datan lisäksi) | 1 050 | Aalto/SeCo. Lähteet: Talvisodan historia, sotapäiväkirjat, organisaatiokortisto, Suomen Rintamamiehet 1939-1945, Maanpuolustuskorkeakoulu/Ohto Manninen, Wikipedia ym. | CC BY 4.0 |
| | Biografiat | 6 300 | Suomalaisen Kirjallisuuden Seura / Kansallisbiografia & Aalto/Seco | CC BY 4.0 |
| | Joukko-osastot | 3 180 | Aalto/SeCo. Talvisodan historia, Talvisodan sotapäiväkirjat, organisaatiokortisto | CC BY 4.0 |
| YHTEENSÄ | | 1 142 616 | | |

Kuva 1. Sotasammon tietoaaineistoja.

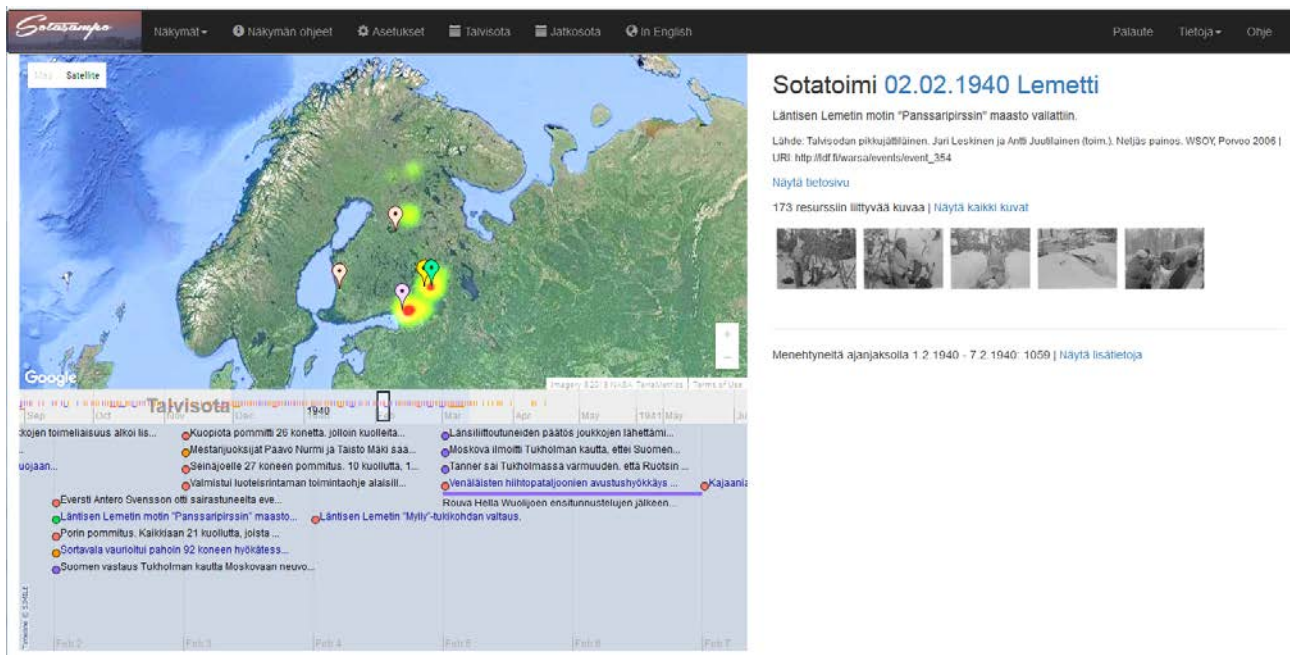
Sotasammon alkuperäiset tietoaaineistot lähteineen on lueteltu kuvassa 1. Heterogeeniset datajoukot harmonisoitiin keskenään yhteentoimiviksi muuntamalla ja linkittämällä ne semanttisen webin RDF-tietomallin mukaiseksi semanttiseksi tietämysverkoksi (engl. knowledge graph). Harmonisoinnin perustana käytettiin museoalan tapahtumapohjaista CIDOC CRM -tietomallia¹ joka on ISO-standardi vuodesta 2006. Sen ontologisena ideana on muuttaa tieto toisiinsa liittyviksi tapahtumiksi, mikä soveltuu hyvin historiallisten tapahtumien kuvaamiseen. Sotasammon tietämysverkossa on nykyään yli 10 miljoonaa

¹ <http://cidoc-crm.org>

tietojen välistä yhteyttä eli kolmikkoa, (engl. triplet), esimerkiksi se, että talvisota alkoi 30.11.1939. Data on avointa käytettävissä Linked Data Finland –palvelualustalla² SPARQL-palvelupisteen kautta.

Semanttisen portaalin sovellusnäkyvät

Sotasampo.fi-portaalin ideana on tarjota dataan erilaisia sovellusnäkyymiä, joita julkaistiin aluksi kuusi: Tapahtumat (taistelut, poliittiset tapahtumat ym. aikajanalla ja kartalla), Henkilöt (aineistoissa esiintyvät n. 100 000 henkilöä ja heille automaattisesti luodut ”kotisivut”), Joukko-osastot (linkitettyä tietoa tuhansista joukko-osastoista), Paikat (kymmeniä tuhansia luovutetun alueen paikkoja historiallisilla kartoilla), *Kansa Taisteli* -lehdet (tekstit linkitettyinä paikkoihin ja toimijoihin) ja Menehtyneet (kattavat tiedot viime sodissa menehtyneestä 95 000 Suomen armeijan sotilaasta).



The screenshot shows the Sotasampo.fi web application. At the top, there is a navigation bar with the Sotasampo logo and various menu items like 'Näkymät', 'Näkymän ohjeet', 'Asetukset', 'Talvisota', 'Jatkosota', and 'in English'. Below the navigation bar is a map of Finland with a heatmap overlay showing event density. A specific event is selected, showing a detailed view for 'Sotatoimi 02.02.1940 Lemetti'. This view includes the title, a source citation, a URI, a 'Näytä tietosivu' button, and a gallery of images. Below the gallery, there is a link to 'Menehtyneitä ajanjaksoilla 1 2 1940 - 7 2 1940 - 1059' and a 'Näytä lisätietoja' button. At the bottom of the map area, there is a timeline of events for the year 1940, with a list of events on either side of the timeline.

Kuva 2. Kuvankaappaus Sotasammon sovellusnäkyvästä Tapahtumat.

Kuvassa 2 on esimerkiksi näkyvä Tapahtumat-sovelluksesta. Erityyppiset sotilaalliset ja poliittiset tapahtumat näkyvät aikajanalla, johon voi zoomata janan päällä näkyvän aikaikkunan kautta. Kartta näyttää aikaikkunaan liittyvät tapahtumat kartalla ja visualisoi ajanjakson aikana menehtyneiden määrän lämpökartan avulla. Kuvassa käyttäjä on valinnut kartalta tapahtuman, joka on talvisodan sotatoimi Lemetissä 2.2.1940. Tapahtumaan linkittyvät tiedot, kuten valokuvat, Lemetti paikkana, tapahtuman tietosivu, ja sotatoimen päivä näkyvät oikealla olevassa tietoikkunassa, jonka linkkejä klikkaamalla saa lisätietoja. Tieto rikastuu automaattisesti eri datajoukkojen linkitysten kautta.

Alkuperäisiin sovellusnäkyymiin on sittemmin lisätty puolustusvoimien Sotamuseon 160 000 autenttisen historiallisen sotavalokuvan haku, selailu- ja suosittelusovellus, sekä yhteistyössä Kaatuneiden muistosäätiön ja Suomen Kameraseurojen Liiton kanssa Sankarihautausmaat-sovellus. Siinä kartoitettiin ja kuvattiin ensimmäistä kertaa maamme kaikki sankarihautausmaat, luotiin niille automaattisesti kotisivut, rikastettiin dataa tiedoilla kunkin hautausmaan sankarivainajista ja luotiin datan avulla graafisia visualisointeja oman alueen vainajien kuolinpaikoista historiallisilla kartoilla, kuolinajoista yms. Tekeillä on myös Sotavainajien muiston vaalimisyhdistyksen ja Kansallisarkiston kokoamaan sotavankitietokantaan (n. 5000 suomalaista sotavankia Neuvostoliitossa) perustuva uusi Sotavangit-sovellus.

² <http://ldf.fi>

Sotasammon avointa dataa ovat hyödyntäneet muutkin tahot kuin kehittäjäryhmä, kuten Sotapolku.fi-palvelu ja Suomi 100 -ohjelmaan kuuluva Kunniavartio-hanke, jossa maamme 100. itsenäisyyspäivänä järjestettiin sankarihaudoille kunniavartioita samanikäisten henkilöiden toimesta kuin hautausmaan vainajat. Sotasampo on tutkimuksellisesti uusi avaus ja innovaatio: siitä on kirjoitettu 14 kansainvälistä tutkimusartikkelia ja opinnäytetyötä, ja siihen liittyen on pidetty useita kansainvälisiä kutsuesitelmää.

Sotasammon taontaa ovat rahoittaneet opetus- ja kulttuuriministeriö, Suomen Kulttuurirahasto, Kaatuneiden Muistosäätiö, Sotavainajien muiston vaalimisyhdistys ja Suomen Akatemia. Työ on jatkoa Semanttisen laskennan tutkimusryhmän (SeCo) aiemmille ”sammoille” Kulttuurisampo.fi ja Kirjasampo.fi.

Lisätietoa

Lyhyt esittelyvideo: <https://vimeo.com/212249404>

Sotasampo portaali: <http://sotasampo.fi>

Sotasammon avoin datapalvelu: <http://www.ldf.fi/dataset/warsa>

Sotasampo hankkeen kotisivu ja julkaisut: <http://seco.cs.aalto.fi/projects/sotasampo/>

Artikkeli Sotasampo-visiosta ja -kokonaisuudesta:

Eero Hyvönen, Erkki Heino, Petri Leskinen, Esko Ikkala, Mikko Koho, Minna Tamper, Jouni Tuominen ja Eetu Mäkelä: WarSampo Data Service and Semantic Portal for Publishing Linked Open Data about the Second World War History. The Semantic Web – Latest Advances and New Domains (ESWC 2016) (Harald Sack, Eva Blomqvist, Mathieu d Aquin, Chiara Ghidini, Simone Paolo Ponzetto ja Christoph Lange (toim.)), pp. 758-773, Springer-Verlag, May, 2016. <http://seco.cs.aalto.fi/publications/2016/hyvonen-et-al-warsa.pdf>

Aiheesta suomeksi:

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