SampoSampo: A Portal for Studying Enriched Data and Semantic Connections on a Cultural Heritage Linked Open Data Cloud

 $\begin{array}{c} \text{Eero Hyvönen}^{1,2[0000-0003-1695-5840]}, \text{Petri Leskinen}^{1[0009-0008-6369-4712]}, \\ \text{Annastiina Ahola}^{1[0009-0008-6369-4712]}, \text{Heikki Rantala}^{1[0000-0002-4716-6564]}, \\ \text{ and Jouni Tuominen}^{3,2,1[0000-0003-4789-5676]}, \end{array}$

 ¹ Semantic Computing Research Group (SeCo), Aalto University, Finland
 ² Helsinki Centre for Digital Humanities (HELDIG), University of Helsinki, Finland
 ³ Helsinki Institute for Social Sciences and Humanities (HSSH), University of Helsinki

Abstract. This paper presents use cases and demonstrates functionalities of the new SAMPOSAMPO portal based on a linked open data cloud (LOD) of related Cultural Heritage knowledge graphs of different application domains. The portal is used for searching, exploring, and analyzing entities in the LOD cloud for globally enriched data and for finding semantic "interesting" or even "serendipitous" connections (relations) between the entities with natural language explanations.

Keywords: linked data \cdot digital humanities \cdot entity alignment \cdot semantic portal \cdot data analysis \cdot knowledge discovery

1 Connecting Everything to Everything Else

Leonardo da Vinci (1452–1519) has said: "Learn how to see. Realize that everything connects to everything else". This wisdom is very true regarding Cultural Heritage (CH) data due to its rich linkedness. However, exposing and learning the connections (links) between resources in CH LOD is a challenge: The data are typically available in distributed data silos, the data are heterogeneous, and use different identifiers for the same entities in different data silos, which cuts off links and connections. As a remedy, the SAMPOSAMPO portal addresses the following challenges: 1) How to search and link data about entities in a LOD cloud for enriched descriptions about the entities. 2) How to search for connections (relations) between entities within and across KGs for finding out, e.g., how people are related to places or each other [4].

SAMPOSAMPO is based on the Sampo series⁴ of mutually related CH LOD services and semantic portals. The "Sampo model" [3]. was used to establish an entity alignment LOD service on top of which a meta-level Sampo, "SAMPOSAMPO", was created.

⁴ Sampo series of over 20 LOD services and CH portals: https://seco.cs.aalto.fi/ applications/sampo/

2 Authors Suppressed Due to Excessive Length

2 Using SampoSampo Portal

There are two main use cases for the SAMPOSAMPO portal. Firstly, it is often useful to search data about entities from different data sources as different datasets may contain complementary enriching information about the entities. For example, BiographySampo publishes general biographical data about 13 000 prominent Finns but there is more data available on them in other Sampos. For example, ParliamentSampo contains the speeches and activities in Parliament of Finland for 790 politicians presented in BiographySampo. Aggregated data also facilitates detecting mutual inconsistencies in the data sources. Secondly, searching for explained "interesting" implicit relations between entities within and across datasets is useful. For example: how are people related to places, say Finnish artists to France or Jean Sibelius to Berlin?

Based on the Sampo model [3] and the Sampo-UI framework [5,7] for UI design, the landing page of SAMPOSAMPO provides access to application perspectives where the instances of KG classes can be searched using semantic faceted search where the facets correspond to the properties of the class. After filtering results by making selections on the facets, the result set can displayed as a table or using a variety of data-analytic tools and visualizations, such as charts, maps, and timelines. By selecting an instance from the result set, aggregated linked data related to it can be displayed and data-analyses and visualizations pertaining to the entity instance can be shown.

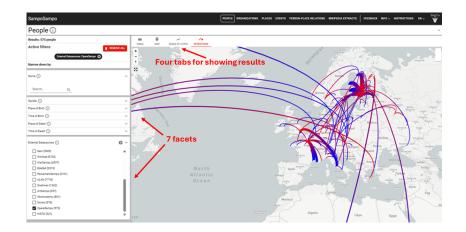


Fig. 1: Visualizing the life lines of people in OperaSampo from their place of birth (blue end of the line) to the place of death (red end).

The demo implementation contains perspectives for searching People, Organizations, Places, Historical events, Connections between people and places, and Connections between entities based on links available in the Finnish Wikipedia and Wikidata. Figure 1 depicts the People perspective with seven facets on the left and search results on four alternative tabs on the right. In this case, OperaSampo was selected from that data source facet and a result set of 575 people there that are linked with at least some other data sources are shown on the right. The user has selected from the tabs the Migrations tab that shows the life lines of the people on a map. One can learn from the visualization, for example, that quite a few opera people have moved to Paris and died there.

Person-place relations ()					
Results: 37 relations		4 facets for discoverin	g connections.	Connections	with explanation
Active filters:	REMOVE ALL	Rovspergage 30 - 1-00 of 37 (< < > >1		Connoctions	
Title or occupation: kijalija 🙆 Piece	1000 0	Description ()	Person ()	Place (i)	Relation type ()
		Wika turbai on mekitu, tekijikai kirjaliseen teokseen Tõde Bestist, Lätut ja Leedust		Lista	Kiriotus kuras pakkaa
larrow down by:		Mineral and methods being an and a second seco		Lete	Kirjotus kuras pakkas
lerson ()	×	 An Watari on mekity, tekijikni kirjalisen technen Tessa apie Estis, Lancis i Lie 		100	Kriotus kuras pakkas
The or occupation (i)	e ^	Mika litektari on menistry teksisikai kicialiseen teokseen Totuus Virosta, Latvianta ja Li		Liena	Kirjohus kuvaa pakkaa
the protospenon ()	U ^	Mita Matari on metility telijika kitaliaren teokaren Tesa asir Estia Latria ir Lin		Lima	Kriptus kuras palitias
valtaneuvos (41)	ê	Mika Waitari on meninty tekliliksi kirjaliseen teokseen 75de Eestist, Likist ja Leedust		Ven	Kirjohys kuvaa pakkaa
🗸 kirjelija (37)		Mika Wahari on meninty tekijikai kirjaliseen teokseen Tõde Bestist, Liitist ja Leedust		Labia	Kirjotus kuras pakkas
kreivi (29)		 Mika Waltari on merkitty tekilikai kirjalliseen teokseen Totuus Virosta, Latviasta la Di 		Vire	Kirjohus kuras pelikas
vapeaherra (28)		 Mika Waltari on merkitty tekilikai kirjalliseen teokseen Tõde Eestist, Lätist ja Leedust 	Cicka kuvas pakkas Batta. Rainio, Leo	Lette	Kirjoitus kuraa pakkaa
Alakunhatia 24		 Mika Waltari on meninty tekilikai kirjaliseen teokseen Totuva Virosta, Latviesta kali 		Latria	Kirjohus kuvas palikaa
Urknown (22)		Mika Waltari on metritty tekilikai kirjaliseen teokseen Tiesa apie Estia. Latvia ir Lie	turki joka kursa pakkaa Vro. Rainio, Leo	Vire	Kirjoitus kuvaa palikaa
kerraaliluutnanti (20)		 Mika Waltari on metritry tekiliksi kirjaliseen teokseen Tiesa apie Estila. Latvia ir Liet 	tural ioka kunaa paikkaa Latvia. Rainin Leo	Labia	Kirjohya kuvas palikaa
everati (14) kansanedustala (14)	Ţ	 Henklön Viho Helenen usan tai opisieluun liittu paikassa Latvia tapahtuma: Helein Vinson, Latviaan vita 1920 	oin viooistosta. Oointomatioia Helander, Viho Veikko Päiviö .	. Labia	Ura tai opiskelu littyy pakkaan
pilspa (14)		 Henklön Viho Helenen ursan tal opiskeluin liitmi, pakassa Viro tapahtuma: Viron av shreeni 1922 - 1944 	ustamisen päätoimikunnan säätöin Helander, Viiho Veikko Päivid .	. Viro	Ura tai opiskelu littyy pakkaan
laamanni (15)		 Herklön Viho Heisnen usan tai opiskelun littu pakassa Talinna tapehtuma sisäi Talinnassa 1942 - 1944 	asiainministeriön toimistopäällikkö Helander, Vilho Veikko Päiviö .	Talima	Ura tai opiskelu littyy pakkaan
iace ()	B o	 Who Helanen on merkitty teküllesi kirjailiseen teokseen Suomalaiset Viron vagausso 	dessal joka kuvaa paikkaa Viro, Helander, Vilho Veikko Päiviö .	Vire	Kirjoitus kuvaa palikaa
0	e ^	 Herklön Viho Helenet usan tai opiskelun Tittin, baikassa Viro tabehtuma: Helsingi Latriaan vata 1920 	Header, had being range	. Vire	Ura tai opiskelu liittyy pakkaan
- Aasia (33)	Î	 Herklör Eik Honborg usan tei opiskelsun Tittu pakassa Virs tapahtume. Viron Vi kunnistuttos Tukholma 1935 	apaudemistin K. Filosofian Homborg, Erik	Vire	Ura tai opiskelu littyy pakkaan
Australia (2)		 Henklör Sighidus Aronus Forsius ursen tei opiskeluur liittur peikassa Viro tepehturt 	w maanmittari Virossa 1616 Forska, Sigfridus Aronius	Viro	Uta tai opiskelu liittyy pakkaan
Etelä-Amerikka (14)		 Herklör Softidus Aronus Forsius ursen tei poiskeluur littyo teikesse Tallinna teost suomelius nuotsalaisen seurakunnen papo, koulun opettaia 1591 	tuma: Talinnan tupmiskirkon Forsius. Sightidus Aronius	Talinna	Ura tai opiskelu littyy pakkaan
 Europpa (112) Batis (27) 		 Hecklön Aino Kallas ursan tei poiskeluut litter, pakassa Talinna tapahtuma. Talinna 	assa 1934 - 1944 Kallas, Airo	Talinna	Ura tai opiskelu liittyy pakkaan
Delgia (1)		 Ano Kalas on merkity teksikai kirjaliseen teokseen Nuori-Vira - muotokuria ja suor 	tavihoja' joka kunas paikkas Vro. Kallas, Aino	Viro	Kirjotus kuras paikkaa
 Espenja (7) 		 Aino Kallas on merkityu teki likai kirialliseen teokseen. Tukaaslane' joka kusaa pakka 	a Viro, Kallas, Airo	Viro	Kirjoitus kuraa palkkaa
Iso-Britannia (23)		 Airo Kallas on merkitu teksiksi kisaliseen teokseen Tähderlento: virolaisen turolli pakkaa Viro, 	jetteren Koldulan elämä joka kuvea Kallas, Aino	Vire	Kirjoitus kuraa palikaa
► [] Itala [22]		 Henklör Ains Kallas uzsan tai opiskeluur littur paikassa Viro tapahtuma: Viron kans 	alainen 1900 - 1953 Kallas, Airo	Vire	Ura tai opiskelu liittyy pakkaan
 ► ☐ Námeri (3) ► ☐ Námeta (12) 		 Henklich Ains Kalles unsen tei opiskelsun litter peikasse Viro tepehtume. Viron lähet 		Vire	Ura tai opiskelu liittyy pakkaan
	•	 Herklör Ivo Häkören ursen tel opiskelun littu pakassa Viru tepäturna. Pohjoint 1920-ja 1920-jaal 	nahin, Viroon ia Ehelä-Eurooppaati Härkönen, Ivo	Vire	Ura tai opiskelu liittyy pakkaan
elation type 💮	v	 Herhilds Sulo Mekho Pekhola usaan tai polskaluun littus pakaasa Viru tapahuma. P konskisäärisukkusen lohtala, vihuutaantii, siittevit Viroon, palannut Hankoon 1918 	brison subietudiuman Pekkola, Sulo Welkio	Vire	Ura tai opiskelu limyy pakkaan
		 Arturi Jinchorna on vestaanottanut kurnianosotuksen joka liittiv, salkkaan Latvia 1 pakinto 1914. 	Latrian Kolmen Tähden R. Berabornin Jänifuoma, Anturi	Labia	Kunnianosoitus liittyy paikkaan
		 Henklön Henry Parland ursen tel opisiellour littor palvasse Liettue tapahtuma. Shite Liettuesse 1929 - 	eri Ruotain konsulaatissa Kaunasissa Parland, Henry	Linu	Ura tei opiskelu littuv pekkaan

Fig. 2: Findig connections with explanations between people and places.

Figure 2 illustrates how a perspective for discovering and searching implicit semantic connections is used. The perspective is based on the class of connections whose instances represent connections between people and places with properties for them as well as an explanation about the connection. The facets on the left are: Person, Occupation or title of the person, Place, and Type of the connection, based on the events through which the connections have been established. Ten event types have been used, such as getting an honorary price in a place and creating a painting depicting a place or a book about it. The user has selected from the Occupation facet category "Author" and "Baltic countries" from the Place facet. On the right, 37 connections between people and places with their explanations are listed, such as "Relating to the studies or career of the person Aino Kallas, she was the spouse of the ambassador of Estonia in Finland in 1918 - 1922" (In Finnish: "Henkilön Aino Kallas uraan tai opiskeluun liittyy paikassa Viro tapahtuma: Viron lähettilään puoliso Helsingissä 1918 - 1922".)

SampoSampo aggregates data about 80 566 people entities from 13 Sampo systems and other data sources, 52 161 organizations, 20 676 place entities, and 1198 major historical events. For searching semantic connections data from our earlier relational search systems [8] was re-used.

4 Authors Suppressed Due to Excessive Length

3 Related Works and Contributions

SAMPOSAMPO portal is based on an entity alignment LOD service. Our work was inspired by related systems, such as the Linked Open Data Cloud⁵, VIAF⁶ [2], works of ontology mapping, ontology services [9,1], Linked Open Vocabularies⁷, and the proxy model of Europeana [6]. Entity alignment services have been used as *tools* for aggregating and enriching data in new applications. A major contribution of SAMPOSAMPO is to present novel use cases and possibilities of using an entity alignment service *itself* as a basis for applications for finding enriched entity data and implicit semantic connections, not only as a mapping tool. It can be used for research in Digital Humanities..

References

- Frosterus, M., Tuominen, J., Pessala, S., Hyvönen, E.: Linked open ontology cloud: managing a system of interlinked cross-domain light-weight ontologies. International Journal of Metadata, Semantics and Ontologies 10(3), 189–201 (2015). https://doi.org/10.1504/IJMSO.2015.073879
- Hickey, T.B., Toves, J.A.: Managing ambiguity in VIAF. DLib Magazine 20(7/8) (2014). https://doi.org/doi:10.1045/july2014-hickey
- Hyvönen, E.: Digital humanities on the Semantic Web: Sampo model and portal series. Semantic Web 14(4), 729–744 (2022). https://doi.org/10.3233/SW-223034
- Hyvönen, E.: Serendipitous knowledge discovery on the Web of Wisdom based on searching and explaining interesting relations in knowledge graphs. Journal of Web Semantics (2025). https://doi.org/DOI: 10.1016/j.websem.2024.100852
- Ikkala, E., Hyvönen, E., Rantala, H., Koho, M.: Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. Semantic Web 13(1), 69–84 (2022)
- 6. Isaac, A.: Europeana data model primer. Tech. rep., Europeana (2023), https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/ Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf
- Rantala, H., Ahola, A., Ikkala, E., Hyvönen, E.: How to create easily a data analytic semantic portal on top of a SPARQL endpoint: introducing the configurable Sampo-UI framework. In: VOILA! 2023 Visualization and Interaction for Ontologies, Linked Data and Knowledge Graphs 2023. vol. 3508. CEUR Workshop Proceedinbgs (2023), https://ceur-ws.org/Vol-3508/paper3.pdf
- Rantala, H., Leskinen, P., Peura, L., Hyvönen, E.: Representing and searching associations in cultural heritage knowledge graphs using faceted search. In: SE-MANTICS 2024, 20th International Conference on Semantic Systems, proceedings. IOS Press (September 2024), https://seco.cs.aalto.fi/publications/ 2024/rantala-et-al-searching-interesting-relations-2024.pdf, in press
- Xia, W., Jiménez-Ruiz, E., Cross, V.V.: Using BioPortal as a repository for mediating ontologies in ontology alignment. In: Workshop on Semantic Web Applications and Tools for Life Sciences (2015), https://api.semanticscholar.org/CorpusID: 37359417

⁵ Linked Open Data Cloud: https://lod-cloud.net/

⁶ Virtual International Authority File system: https://viaf.org

⁷ Linked Open Vocabularies: https://lov.linkeddata.es/dataset/lov/