Creating the HISTO Ontology of Finnish History Events

Heikki Rantala[1], Eero Hyvönen[1,2], and Esko Ikkala[1]

[1] Aalto University, Semantic Computing Research Group (SeCo), Finland
[2] University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG), Finland
https://seco.cs.aalto.fi

1 An Ontology of History

This paper overviews work on creating an ontology of important events of Finnish history called HISTO[3]. HISTO is part of the larger Linked Open Data Infrastucture for Digital Humanities (LODI4DH) initiative[4] in Finland.

The ontology is based on a timeline that was originally created by the Agricola[5] network of Finnish historians. A limited early version of 220 events was released already in 2007 as part of the CultureSampo [5], and the idea was developed further in 2012 in a historical event ontology for the World War I [8]. More recent version of HISTO has been used as part of the BiographySampo [7,3], where 345 relations between person and a place were extracted using the event ontology. The ontology consists currently of over one thousand events. The events are categorized based on type, theme, and subject matter keywords. Currently there are 164 event types, 29 themes, and 1868 keywords used in the ontology. The keywords are taken from the KOKO ontology[6], but the types and themes are created specifically for the HISTO ontology. The event types constitute a classification of the types of historical events. Themes relate to what field of history the event relates to, for example, to cultural history or military history.

We are using Linked (Open) Data and RDF to represent the ontology. In Linked Data [2] every resource is represented with an universally unique identifier that allows it to be easily referenced even from outside sources. The events link to other events, places, people, and times. The events take place at some time and in some place, they have people as their participants in different roles, and events may consist of smaller subevents. In addition, events may cause other events and create causal chains. Our data model is based on the CIDOC CRM[7] ontology that is extended when necessary.

To represent people, we use an ontology based on the people extracted from the Finnish National Biography in the BiographySampo system. The Finnish

---

[3] https://seco.cs.aalto.fi/ontologies/histo/

[4] https://seco.cs.aalto.fi/projects/lodi4dh/

[5] Agricola portal (in Finnish): https://agricolaverkko.fi

[6] KOKO is a linked ontology cloud of Finnish core ontologies: https://finto.fi/koko/en

[7] http://cidoc-crm.org

National Biography collection used there includes 13 100 people with biographies and lots of references to related people. These people can be considered important in the Finnish history. For places we use an ontology that is based on the YSO Places vocabulary[8] maintained by the National Library of Finland, and extended and made more hierarchical in the BiographySampo system.

We are representing times using the CIDOC CRM model. Every event has it's own resource that represents the time-span of that event. Each time-span has the earliest and latest start and the earliest and latest end. This represents the length of the event, but can also represent uncertainty about the actual exact dates of the historical event.

Once published, the ontology of historical events can be used for data indexing and be linked from outside sources, and it is not limited to a specific place or person ontology. Events can be seen as semantic glue that connects people to other people and places in time. The HISTO ontology of events can also be used to link events in other ontologies. For example, the independence declaration of Finland might be referenced in multiple data sets. If all the references either use the same URI, or link to it, it is easy to automatically define that the data sets are referencing the same event even if they might use a slightly different name for it.

One central challenge in the project is applying CIDOC CRM model to historical events as Linked Data. CIDOC CRM is a somewhat general level model. Applying it as Linked Data is not straightforward and requires some adapting to fit the purpose of modeling historical events.

The modeling solutions developed in HISTO ontology can hopefully be useful in creating other ontologies of historical events. The ontologies for the types of historical events and the fields of history, that are created for HISTO, should also be possible to apply elsewhere.

2   Semantic Portal HistorySampo

To demonstrate the use of the ontology we are also creating a semantic portal based on the ontology for human users, not only a data services for the ontology. The portal will be based on the Sampo model [4] and it's user interface is created using the Sampo-UI[9] framework. The portal can hopefully be used for educational purposes to educate people of the Finnish history. The portal will have faceted search functionalities to search and browse events based on times, places, types, themes and keywords. The events will be visualized as a paginated table and on maps and timelines. Every event will have a homepage within the application with various links to related materials.

---

[8] YSO places vocabulary: `https://finto.fi/yso-paikat/en/`
[9] Sampo-UI framework: `https://github.com/SemanticComputing/sampo-ui`

3   Related Work

In addition to CIDOC CRM, several other models for representing events has been proposed, such as SEM [1] and LODE [12]. BigraphySampo and WarSampo [6] are in-use applications where extensions of CIDOC CRM were developed for war history and historical biographical data, respectively. The idea of using semantic web technologies for representing history has been proposed in many works, such as [10,11]. An overview survey of using semantic web technologies in historical research is presented in [9]. In contrast to these approaches, the focus of our work on HISTO is to create a shared history ontology of events for data indexing and linking for the machines, and on top of it, a semantic portal of history for human consumption.

References

1. van Hage, W.R., Malaisé, V., Segers, R., Hollink, L., Schreiber, G.: Design and use of the simple event model (sem). J. Web Semant. **9**, 128–136 (2011)
2. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space (1st edition). Morgan & Claypool, Palo Alto, California (2011), `http://linkeddatabook.com/editions/1.0/`
3. Hyvönen, Rantala, H.: Knowledge-based relation discovery in cultural heritage knowledge graphs. In: DHN 2019 Digital Humanities in Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 4th Conference. pp. 230–239. CEUR Workshop Proceedings, Vol-2364 (March 2019), `http://www.ceur-ws.org/Vol-2364/`
4. Hyvönen, E.: Sampo model and semantic portals for digital humanities on the semantic web. In: Proc. of the Digital Humanities in the Nordic Countries (DHN). CEUR WS Proceedings (2020)
5. Hyvönen, E., Alm, O., Kuittinen, H.: Using an ontology of historical events in semantic portals for cultural heritage. In: Proc. of the Cultural Heritage on the Semantic Web Workshop at ISWC 2007 (2007)
6. Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., Tuominen, J., Mäkelä, E.: Warsampo data service and semantic portal for publishing linked open data about the second world war history. In: European Semantic Web Conference. pp. 758–773. Springer (2016)
7. Hyvönen, E., Leskinen, P., Tamper, M., Rantala, H., Ikkala, E., Tuominen, J., Keravuori, K.: BiographySampo – Publishing and enriching biographies on the Semantic Web for digital humanities research. In: Proceedings of the 16th Extended Semantic Web Conference (ESWC 2019). pp. 574–589. Springer–Verlag (2019)
8. Hyvönen, E., Lindquist, T., Törnroos, J., Mäkelä, E.: History on the semantic web as linked data – an event gazetteer and timeline for world war i. In: Proceedings of CIDOC 2012 - Enriching Cultural Heritage, Helsinki, Finland. CIDOC, http://www.cidoc2012.fi/en/cidoc2012/programme (2012), `https://seco.cs.aalto.fi/publications/2012/hyvonen-et-al-ww1-cidoc-2012.pdf`
9. Merono, A., Ashkpour, A., van Erp, M., Mandemakers, K., Breure, L., Scharnhorst, A., Schlobach, K., van Harmelen, F.: Semantic technologies for historical research: A survey. Semantic Web **6**(6), 539–564 (2015). https://doi.org/10.3233/SW-140158

10. Nagypal, G., Deswarte, R., Oosthoek, J.: Applying the semantic web: The VICODI experience in creating visual contextualization for history. Lit Linguist Computing **20**(3), 327–349 (2005), `http://dx.doi.org/10.1093/llc/fqi037`
11. Robertson, B.G.: Fawcett: A toolkit to begin an historical semantic web. Digital Studies / Le Champ Numerique **1**(2) (2009)
12. Shaw, R., Troncy, R., Hardman, L.: Lode: Linking open descriptions of events. In: Asian semantic web conference. pp. 153–167. Springer (2009)