Demonstrating BiographySampo in Solving Digital Humanities Research Problems in Biography and Prosopography

Eero Hyvönen[1,2], Petri Leskinen[1], Minna Tamper[1], Heikki Rantala[1], Esko Ikkala[1], Jouni Tuominen[1,2], and Kirsi Keravuori[3]

[1] Semantic Computing Research Group (SeCo), Aalto University, Finland and
[2] HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
`http://seco.cs.aalto.fi, http://heldig.fi`
`firstname.lastname@aalto.fi`
[3] Finnish Literature Society (SKS)
`firstname.lastname@finlit.fi`

Abstract. This demonstration paper introduces the semantic portal "Biography-Sampo – Finnish National Biographies on the Semantic Web"[4] released on September 27th 2018 for public use[5]. BiographySampo *aims to make a paradigm shift in publishing biographical dictionaries on the web*. Firstly, the system provides the end user with an enhanced reading experience of biographies by enriching them with data linking and reasoning. Secondly, the BiographySampo includes versatile tooling for 1) biographical research of individual persons as well as 2) prosopographical research on groups of people.

BiographySampo generates for each person and place in the system a global "home page" for enhanced reading experience by enriching data from various interlinked data sources and by reasoning. The system is based on a Linked Data service on top of which seven application perspectives with tooling are provided.

1. *Persons*. Semantic faceted search view for filtering and finding protagonists of the biographies in flexible ways.
2. *Places*. Searching biographical events projected on interactive maps.
3. *Life maps*. Lives from birth to death of filtered person groups visualized as events or trajectories on maps.
4. *Statistics*. Various histogram and pie chart statistics of filtered person groups.
5. *Networks*. Analyzing networks of filtered people.
6. *Relations*. Finding serendipitous connections between persons and places.
7. *Language*. Tools for analyzing the language used in biographies.

The data was created by extracting knowledge from the underlying biographical texts, some 13 100 short biographies published by the Finnish Literature Society, using

---

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

[5] http://biografiasampo.fi

natural language technologies. After this, the data was enriched by linking it to 13 external biographical databases, and to some additional collection databases of memory organizations and semantic web data services. BiographySampo is the next step in our series of semantic "Sampo" portals [1] based on Linked Data, including CultureSampo (2009), TravelSampo (2011), BookSampo (2011), and WarSampo (2015). These applications have been used widely on the web, most notably BookSampo (up to 2 million users in 2018) and WarSampo (227 000 users in 2018).

BiographySampo builds upon our earlier prototype system, Semantic National Biography of Finland [3,5,4][6], with the following new contributions:

1. *Network analysis*. The system includes two tools for doing network analysis based on the mutual references of the underlying biographies, including automatic generation of egocentric networks and prosopographical group networks.
2. *Linguistic analysis of biographies*. In BiographySampo, the biographies can be analyzed by using linguistic analysis. For example, it turns out that the biographies of female members of the Finnish Parliament frequently contain words "family" and "child", but these words are seldom used in the biographies of male members.
3. *Relational search*. A new application perspective for discovery of connections in the biographical knowledge graph was created. This application is at the moment used to find out in what ways (groups of) people are related to places and areas, and it creates natural language explanations for such connections. For example: "How are Finnish artists related to Florence?" or "Who has got most awards in Germany?". The queries are formulated and the problems are solved in a novel way using faceted search.
4. The system also makes use of more datasets than before, and is linked to more external datasets for enriching the metadata.
5. The system is in use and the biographies can be read freely on the web without copyright concerns.

BiographySampo is explained in more detail in [2], including examples of using it, a discussion of its use in Digital Humanities research, and a comparison with related works.

Acknowledgements

---

[6] In these papers, related works are also discussed, unlike in this abstract.

[7] http://seco.cs.aalto.fi/projects/severi

[8] https://openscience.fi

References

1. Hyvönen, E.: Cultural heritage linked data on the Semantic Web: Three case studies using the Sampo model (2016), https://seco.cs.aalto.fi/publications/submitted/hyvonen-vitoria-2017.pdf, Invited talk: VIII Encounter of Documentation Centres of Contemporary Art: Open Linked Data and Integral Management of Information in Cultural Centres Artium, Vitoria-Gasteiz, Spain, October 19–20
2. 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 (2019), https://seco.cs.aalto.fi/publications/, submitted for review
3. Hyvönen, E., Leskinen, P., Tamper, M., Tuominen, J., Keravuori, K.: Semantic national biography of Finland. In: Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018). pp. 372–385. CEUR Workshop Proceedings, Vol-2084 (March 2018), http://www.ceur-ws.org/Vol-2084/short12.pdf
4. Leskinen, P., Miyakita, G., Koho, M., Hyvönen, E.: Combining faceted search with data-analytic visualizations on top of a SPARQL endpoint. In: Proceedings of VOILA 2018, Monterey, California. vol. 2187. CEUR Workshop Proceedings (October 2018)
5. Tamper, M., Leskinen, P., Apajalahti, K., Hyvönen, E.: Using biographical texts as linked data for prosopographical research and applications. In: 7th International Conference, EuroMed 2018, Nicosia, Cyprus. Springer-Verlag (November 2018)