Using Large Language Models for searching explainable relations in a cloud of Cultural Heritage knowledge graphs: SampoSampo as a neuro-symbolic system

Annastiina Ahola[1,*], Petri Leskinen[1], Heikki Rantala[1], Jouni Tuominen[2,3,1] and Eero Hyvönen[1,3]

[1]Aalto University, Department of Computer Science
[2]University of Helsinki, Helsinki Institute for Humanities and Social Sciences (HSSH)
[3]University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG)

Abstract

Knowledge discovery of "interesting" or even serendipitous relations in data, often called relational search, provides a novel Artificial Intelligence-based approach in Digital Humanities for studying Cultural Heritage. Relational search methods are traditionally symbolic, based on searching connections in knowledge graphs. In contrast, this paper presents a novel neuro-symbolic approach to relational search based on combining Large Language Models (LLM) with knowledge graphs (KG). It is argued that by using curated KG data and data models with Retrieval-Augmented Generation, hallucinations of LLMs can be mitigated and relational search extended also to web resources external to the underlying cloud of KGs. As a practical use case, first results of using the method for knowledge discovery as part of the new web service *SampoSampo – Connecting Everything to Everything Else* are presented.

Keywords

linked data, knowledge graphs, cultural heritage, relational search, relation extraction

1. Knowledge discovery of relations in a cloud of knowledge graphs

In Cultural Heritage (CH) data, such as person biographies, we are often not only interested in the people and other entities being described but also in how all these entities are interlinked between each other. That includes, for example, being able to answer questions such as *'How are Person X and Person Y related to each other?'* or *'How is Person X related to Place Y?'* regarding entities X and Y present in the data. Unfortunately, these relations might only be included in the data implicitly in the form of free-form text, based on contextual external data not present in the dataset at hand, or otherwise not modeled in a machine-readable way. This limits how well the relations can be searched and found.

Relation Extraction (RE) (Nasar, Jaffry, and Malik 2021) in Knowledge Discovery (KD) (Maimon and Rokach 2005) deals with the task of extracting and classifying relations from source data, such as texts and other documents, to a structured format, e.g., RDF. These extracted relations can then be used in Relational Search (RS) (Heim, Lohmann, and Stegemann 2010; Hyvönen and Rantala 2021; Hyvönen 2024) applications to provide the user with the tools to not only explore all the found relations, but also filter them on factors such as relation type, e.g., relations describing only student–teacher relationships between entities or a person's career being related to a specific place.

Digital Humanities in the Nordic and Baltic Countries, 10th Conference: Lost in Abundance – Encounters with the Non-Canonical, March 9-13, 2026, Aarhus Denmark
*Corresponding author.
annastiina.ahola@aalto.fi (A. Ahola); petri.leskinen@aalto.fi (P. Leskinen); heikki.rantala@aalto.fi (H. Rantala); jouni.tuominen@helsinki.fi (J. Tuominen); eero.hyvonen@aalto.fi (E. Hyvönen)
https://www.aalto.fi/en/people/annastiina-ahola (A. Ahola); https://www.aalto.fi/en/people/petri-leskinen (P. Leskinen); https://seco.cs.aalto.fi/u/rantalh3/ (H. Rantala); https://seco.cs.aalto.fi/u/jwtuomin/ (J. Tuominen); https://seco.cs.aalto.fi/u/eahyvone/ (E. Hyvönen)
0009-0008-6369-4712 (A. Ahola); 0000-0003-2327-6942 (P. Leskinen); 0000-0002-4716-6564 (H. Rantala); 0000-0003-4789-5676 (J. Tuominen); 0000-0003-1695-5840 (E. Hyvönen)

Digital Humanities in the Nordic and Baltic Countries Publications – ISSN: 2704-1441

This problem has been approached before by various methods based on searching connecting paths in knowledge graphs (KG). This paper shows a method for using Large Language Models (LLM) for the task in combination with data from knowledge graphs, leading to a new kind of hybrid neuro-symbolic system (Hitzler and Sarker 2022; Bhuyan and Tomar 2024). As a case study, the method is being implemented as part of the *"SampoSampo – Connecting Everything to Everything Else"* [1] (Hyvönen, Ahola, Leskinen, Rantala, et al. 2025; Hyvönen, Ahola, Leskinen, and Tuominen 2025) data service and semantic portal.

SampoSampo is a new member in the Sampo series[2] of Linked Open Data (LOD) services and portals (Hyvönen 2022) in use in Finland. By combining information about the entities from different sources, the user will be able to get a more comprehensive picture about specific entities than they would from looking at it from a single source. In contrast to earlier Sampos, it is a "metasampo" based on a cloud of 11 other Sampos—such as BiographySampo, AcademySampo, LetterSampo Finland—and 8 related external data services on the Web, including Wikidata (Vrandečić and Krötzsch 2014), Geni.com, Getty ULAN[3], etc. The SampoSampo knowledge graph (KG) is a data linking service with resemblance especially to the viaf.org[4] service (Hickey and Toves 2014) provided by OCLC, but also to the works of ontology mapping, ontology services (Xia, Jiménez-Ruiz, and Cross 2015; Frosterus et al. 2015), Linked Open Vocabularies[5], and the proxy data model of Europeana (Isaac 2023).

Our work demonstrates how such a KG and LOD service can be used for the following practical purpose: how to discover new relations between entities over the KG cloud with explanations, following the philosophy of explainable AI (Došilović, Brčić, and Hlupić 2018)? This paper presents the first results of using SampoSampo in this use case in line with the vision for KD of "interesting" (Silberschatz and Tuzhilin 1995) or even serendipitous relations on the Web of Wisdom (Hyvönen 2024). The SampoSampo portal[6] and the underlying LOD service[7] were opened on the Web using the open MIT and CC BY 4.0 licenses on 9 January 2026[8].

2. Aggregating explained relations by complementary methods

An important aspect of the relational search methods in SampoSampo is to be able to integrate the relations created by different methods, based on different datasets, and across different sources in a cloud of interlinked KGs. Some of the sources allow these relations to be extracted from existing links between entities, such as extracting relations between people and places of education from AcademySampo or between people who have frequently exchanged letters with each other from LetterSampo Finland.

However, there is an untapped potential in the biographical texts present in BiographySampo. The relations present in the summaries at the end of the biographical texts have been extracted using a rule-based approach (Hyvönen et al. 2019) and turned into relations (Rantala, Hyvönen, and Tuominen 2023), but not from the free-form parts of the texts. With the advent of LLMs, automating the extraction of relation data from these texts is now more feasible.

This paper describes the work on extracting relations using LLMs and the biographical texts from BiographySampo as additional context in the extraction process. The initial focus of our work is on generating relations between person and place entities, with other relations such as person–person relations planned to be added later.

[1] SampoSampo project homepage: https://seco.cs.aalto.fi/projects/ss/
[2] Sampo systems online: https://seco.cs.aalto.fi/applications/sampo/
[3] ULAN: https://www.getty.edu/research/tools/vocabularies/ulan/
[4] Virtual International Authority Files: https://viaf.org
[5] Linked Open Vocabularies: https://lov.linkeddata.es/dataset/lov/
[6] SampoSampo portal: https://samposampo.fi/
[7] SampoSampo LOD service: https://ldf.fi/dataset/sampo/
[8] SampoSampo publication event: https://seco.cs.aalto.fi/events/2026/2026-01-09-samposampo/

3. Relation extraction process

The biographical texts from BiographySampo have already been run through a NER and NEL process (Tamper, Leskinen, and Hyvönen 2019). Therefore, the recognized and linked mentions of places in the texts can be used to determine person–place pairs that are more likely to produce documented relations as opposed to randomly trying to combine person and place entities.

As the biographical texts are in Finnish, the LLM chosen for the task was required to be able to handle Finnish in addition to English. For ease of use, it was preferred that the LLM could be accessed through some type of API as opposed to having to run it ourselves using a supercomputer. In addition to these two factors, we wanted the model to cite its sources for any claims it made, which would require it to be able to execute web searches in a feasible way. LLMs can notoriously be unreliable and hallucinate facts when going unchecked, so by asking it to include sources could at least in theory make it a bit easier for both us and the user to evaluate and verify the generated claims and sources used. In the best case scenario, these additional web sources could provide more details to relations described in the BiographySampo biographical texts or even provide new relations not explicitly mentioned in the texts. If unreliable sources are later spotted, the relations using these sources can be either removed or corrected as the source information is retained for all individual relations.

In the end, OpenAI's GPT-5 mini model[9], was chosen for the task with web searches enabled through the API, as it fulfilled our requirements and offered a more cost-efficient alternative to the larger GPT-5 model. The model is tasked with generating relations in JSON format, specifying both the origin and target nodes for the relation, a relation type to categorize it, a short description of the relation, time period(s), sources for the claims made in the generated relation as well as a reasoning for including the relation in the first place. It is given a list of relation types that were used for the existing BiographySampo relations and tasked to prioritize their use whenever possible. However, if the model deems none of them appropriate, it is allowed to generate new ones. At the end of the prompt, the model also receives the biographical text in Finnish from BiographySampo alongside its page link, serving as an authoritative source that the model can use to generate relations. The complete template for the prompt is included in Appendix A.

If no documented and verifiable relations exist between two entities, the model is instructed to return an empty array. This could be possible in the cases of erroneous NER and NEL, where, for example, a mention of a person's surname or an adjective was instead linked to a place sharing the same name. For example, many of the biographical texts contain some inflection of the word *ruotsalainen* (engl. *Swedish*). In addition to functioning as an adjective describing things of or related to Sweden, it can also be a surname (*Ruotsalainen*) or refer to the island of *Ruotsalainen* located in Naantali, Southwest Finland. So, even though a place is recognized as mentioned in a text, it does not necessarily ensure that a relation should always exist between the person and the mentioned place. To avoid at least some unnecessary API calls, some of the most common erroneously recognized and linked places[10] were removed from the lists of possible person–place pairs.

4. Generated relation results

Table 1 includes an example response received from the model presented as a table. In the case of this relation, the model was asked to generate relations between Finnish cinematographer and actor Olavi Tuomi (1932–2006) and the municipality of Sodankylä in Lapland, which is a place mentioned in his biography. The model is able to generate a relation to describe Tuomi having received an award from a film festival that takes place in Sodankylä. In the original BiographySampo relations, the only relations related to Tuomi are his two relations to Helsinki, the capital of Finland, which is both his place of birth and death.

[9]OpenAI's GPT-5 mini model: https://platform.openai.com/docs/models/gpt-5-mini

[10]These were mostly common Finnish first and last names, which are also names of some more obscure Finnish places that are unlikely to be actually referenced in biographies.

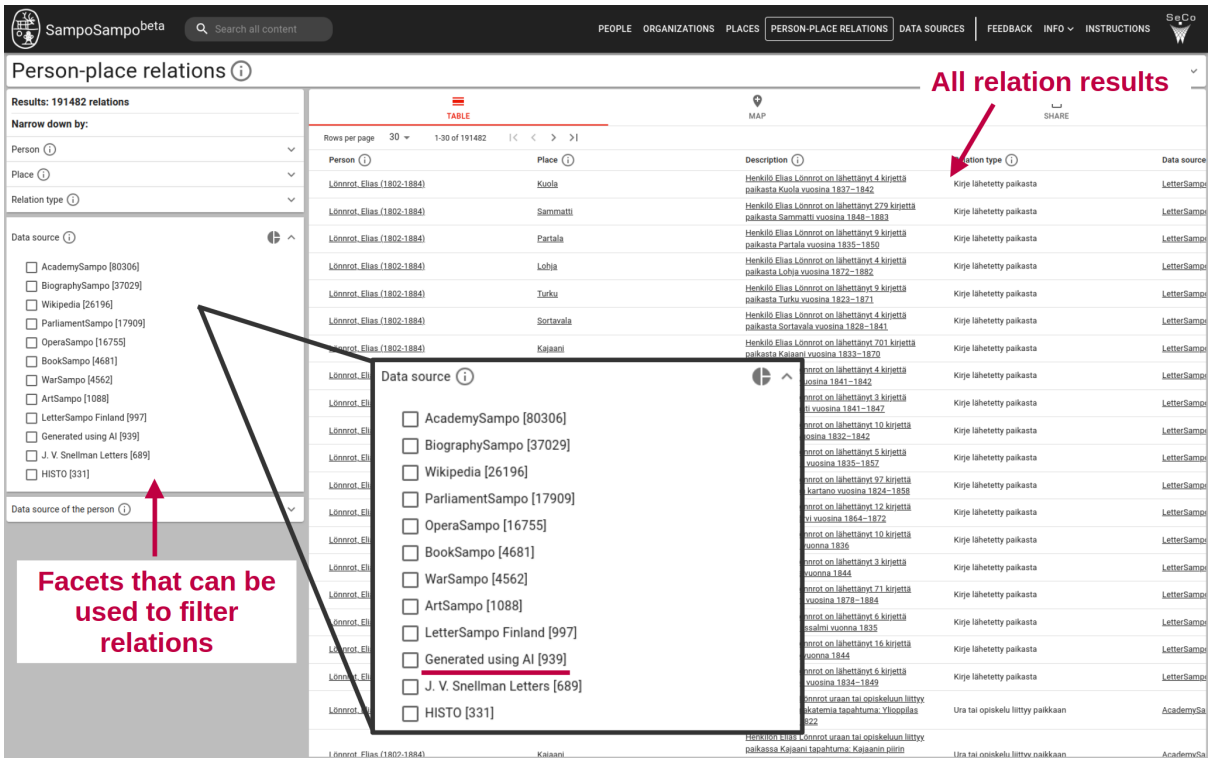

Figure 1: A screenshot showing the person–place relations faceted search view in SampoSampo portal with its data source facet open

The generated relations are transformed from the original response JSON format into RDF and added to SampoSampo KG, as well as made visible in the user interface (UI). The new relations primarily use the same properties as used for the original existing BiographySampo relations already present in the SampoSampo data with the addition of new properties for the web URL sources provided in the relation responses provided by the model. To ensure that the provenance of the relations and their potential unreliability due to hallucinations are clear to the user, the primary source of the generated relations is "Generated using AI". This also provides a way for the user to exclude these relations by only choosing non-AI sources in the UI if they wish (see Figure 4). The model was passed 316 requests and a total of 14.6M input tokens in those requests to produce the currently included 939 generated relations. This resulted in around 1960 executed web searches and 1.24M output tokens. In other words, one request for a person–place pair produced on average around three relations between the specified person and place.

5. Evaluation

Random samples of a total of 50 generated relations were selected to perform some evaluation of the initial results. Of these 50 relations, 20 were sampled from the very initial batch of 216 generated relations with no restrictions on the generated relation types. An additional 30 relations were sampled from an expanded set of generated relations (939 relations at the time of sampling), *excluding* relations with the relation types *Place of death* and *Place of birth*. This type restriction was intended to ensure the focus on the relations that are likely not already present in the KG from other sources and/or extraction methods.

These 50 generated relations included a total of 123 sources provided by the model—including both a link to a web source and a statement regarding what the model claimed the source contained that was relevant to the relation. To give some insight into what kind of relations ended up being included in the evaluation, 37 out of 50 generated relations used existing BiographySampo relation types (see

Table 1
An example relation generated by the model presented in table format

Attribute	Value
`origin_node`	Tuomi, Olavi (1932-2006)
`target_node`	Sodankylä
`relation_type`	award_received_from_place
`relation_description`	Olavi Tuomi was awarded the Sodankylä-palkinto (Sodankylä Award) by the Sodankylän elokuvajuhlat (Sodankylä Film Festival).
`time_period`	• 1989
`sources`	• *https://msfilmfestival.fi/historia/* The Midnight Sun Film Festival (Sodankylän elokuvajuhlat) history page lists recipients of the Sodankylä-palkinto and shows '1989: Olavi Tuomi, kuvaaja (cinematographer)'. • *https://msff.fi/index.php/fi-FI/historia* Festival history page (Finnish) records the Sodankylä-palkitut and names Olavi Tuomi as the 1989 recipient. • *https://yle.fi/a/3-8067750* Yle's article summarising Sodankylä-palkinto recipients includes the entry '1989 Olavi Tuomi, kuvaaja'. • *https://biografiasampo.fi/henkilo/p5788* Olavi Tuomi's BiographySampo page states that 'Olavi Tuomi sai merkittävästä työstään suomalaisen elokuvakulttuurin hyväksi Sodankylän elokuvajuhlien palkinnon 1989' (he received the Sodankylä Film Festival award in 1989).
`reasoning`	All cited sources (the official festival history page, a festival site mirror, a major Finnish broadcaster Yle, and Tuomi's BiographySampo entry) explicitly state that Olavi Tuomi was granted the Sodankylä-palkinto in 1989; this documents a verifiable directed relation from the person to the place as recipient of an award given by the Sodankylä film festival.

Table 2). 10 of them (originating from the initial batch of 20 sampled relations) were relations describing people's places of birth and death, which have also been extracted with rule-based methods from the summaries at the end of biography texts. The new generated relation types are sensible in regards to their respective relations, but there is some unwanted overlap with the existing relation type 'Career or studies related to place'. The relations with the type 'Educated at place', 'Held local government position in place' and 'Held local church office in place' should have all been under the type 'Career or studies related to place' instead based on the prompt instructions.

The vast majority of the links (121 out of 123) provided were valid web URLs and were accessible at the time of writing. The remaining two links did refer to existing sites but were missing parts of the links (e.g., `.html.stx#` missing at the end of an otherwise valid link). The provided links included sites such as wiki sites (e.g., Wikipedia in different languages), records in different library / museum catalogs as well as organizations' own websites documenting their history. There were a handful of cases where providing a link to a catalog's search page with filter might have been preferable (e.g., linking all documents written by a person) to the provided links that pointed to singular catalog records, but overall the sources provided were appropriate.

Table 3 shows the results for 1) the correctness of the relations and 2) the correctness of the statements accompanying the source web links. In this case, the correctness of the relations is evaluated purely on the basis of whether the nodes of the relation match the ones stated in the relation description and whether the relation description is a true statement. This evaluation does not attempt to evaluate, e.g., the semantic meaningfulness of the relations. For example, the relation that Mr. Olavi Tuomi was born in Finland would here be counted as a correct statement but not exactly an interesting relation in

Table 2
The relation types of evaluated relations

Relation type	Count
Existing relation types	*37*
– Place of death	6
– Place of birth	4
– Career or studies related to place	20
– Event took place at	4
– Received honor related to place	3
New relation types	*13*
– Received honor in place	2
– Educated at place	1
– Resided in place	2
– Membership in institution in place	1
– Family member worked in place	1
– Leisure activity in place	1
– Archives located in place	1
– Stage design depicts place	1
– Citizen of place	1
– Held local government position in place	1
– Held local church office in place	1

Table 3
The correctness of generated relations and source statements

Type	Total count	Correct	Partially correct	Incorrect
Relations	50	50	0	0
Statements	123	111	10	2

terms of relational search as we also have the more specific relation that he was born specifically in Helsinki, Finland. Another challenge not taken into account here is evaluating whether in some cases it is semantically correct for a relation to point to a place rather than an organization: For example, in the relation presented in Table 1, does the relation truly meaningfully link Mr. Olavi Tuomi to the place of Sodankylä or should this relation only be reserved for a relation between Mr. Olavi Tuomi and the organization of *Sodankylän elokuvajuhlat*?

The correctness of the statements was evaluated on the basis of whether the web source includes mentions of everything the statement claims it does. This means that a correct statement is fully backed by the source, but a statement with multiple claims and only some of them backed by the source, even if other sources confirm the additional claims, is evaluated as only partially correct. For example, the following statement "*The Kirpilä biography on the museum site lists posts including assistant physician at Kivelä Hospital (1956–1962).*" was marked as only partially correct, since the years listed in the statement (1956–1962) are present in the biography text itself, but not the web page provided as the source for this particular statement. The model would also sometimes have troubles with Finnish conjugation, leading to slightly incorrect names of things like places and events, e.g. *the battles of Talin* instead of *the battles of Tali*, which would also lead to the statement being evaluated as only partially correct if everything else was fine. The two statements evaluated as invalid stemmed from the model giving a source page that did not directly contain the information the model claims it did: One was a preview to a paper—the full paper of which supposedly contained the stated information—and another was a Finna.fi[11] catalog of a book that supposedly covered the information, but the Finna.fi page itself had no description for the contents of the book.

[11]Finna.fi search service: https://www.finna.fi/

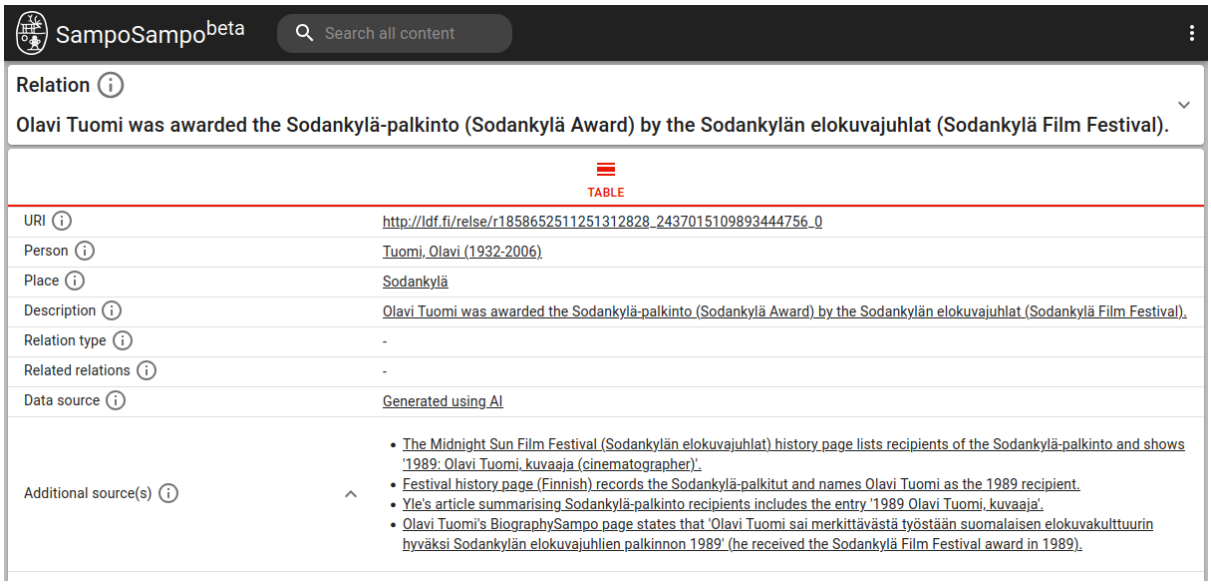

Figure 2: A screenshot showing the instance page of a relation in the SampoSampo portal

6. Using SampoSampo portal and data service

The SampoSampo portal offers multiple *application perspectives* to view the data, including a perspective for person–place relations as already seen in Figure 4. Selecting a perspective opens up its *faceted search view* with a facet menu and a result view (usually a table by default) as well as possible other visualizations included as different tabs for the result view.

In the person–place perspective, the user can view all the relations present in the data or filter a subset of data they want to look at by using the available facets. The user can limit the relations to those related to, e.g., a certain person or place, or the relations to be of a certain relation type(s) or from a specific data source or sources. For example, if the user is interested in how Finnish priest Petrus Bång's (1633–1696) career or studies are related to Uppsala Municipality based on AcademySampo data, they would select *Bång, Petrus (1633-1696)* from the Person facet, *Uppsalan kunta* (Engl. Uppsala Municipality) from the Place facet, *Ura tai opiskelu liittyy paikkaan* (Engl. Career or studies related to place) from the relation type facet and *AcademySampo* from the Data source facet. After making these selections, the results on the faceted search view page would automatically update to match these constraints.

Clicking on the label of an individual relation leads to that relation's own *instance page*, where the user can see its information in detail, including the sources for the AI-generated relations. Figure 6 shows how the instance page looks for the same AI-generated relation that was used as an example in Table 1. The additional sources listed also serve as links, leading to the web URL provided by the LLM as the source for the particular claim serving as the label for the link.

7. Conclusions and Future Work

Based on the initial result evaluation, extracting relations by providing the model with an authoritative information source, in this particular instance the BiographySampo KG, and prelimiting the person–place pairs based on that authoritative source produces sensible results, with all evaluated relations having factually correct descriptions.

However, even though the model was granted access to resources outside of its training data, such as web materials for verifying its facts, some hallucinations and minor errors still slipped in. For instance, the model ended up claiming that multiple facts are sourced from a web page, while that particular page only partially corroborates them, and would sometimes struggle with Finnish conjugation leading

to slightly misspelled entity names in its output. This highlights the importance of providing users enough information about the provenance of the data so that they can accurately assess its reliability and what potential biases it might be affected by. In the end, even with these hallucinations and errors, 90% of the statements evaluated were correct.

There are also other challenges, other than just generating factually correct relations. As briefly touched on in the evaluation section of this paper, not all relations are created equal in terms of their semantic meaningfulness. Eliminating "uninteresting" relations—especially duplicate relations entities describing the same connection between two entities just in different words—would make it easier for the user to come across potentially interesting relations in a serendipitous way.

Some duplicate relations are easier to automatically recognize (e.g., a person is only born once and only dies once), provided that the different relation types are used consistently between different data sources, which in itself is another challenge. With the original relation types used for the BiographySampo relations being limited to the ones that were relevant to what could be extracted back then, at least some level of harmonization and grouping of new generated relation types will have to be carried out in the future to ensure a more consistent experience for the user.

Acknowledgments: Our work is part of the national FIN-CLARIAH research infrastructure programme, funded by the Research Council of Finland and the European Union – NextGenerationEU instrument under grant number 346323. The last author was funded also by an Eminentia Grant of the Finnish Cultural Foundation for reflecting the research of the SeCo research group in 2001–2025. CSC – IT Center for Science has provided computational resources for our work.

References

Bhuyan, Amar, Bikram Pratim nsd Ramdane-Cherif, and Ravi Tomar. 2024. "Neuro-symbolic artificial intelligence: a survey." *Neural Computing and Applications* 36:12809–12844. https://doi.org/10.1007/s00521-024-09960-z.

Došilović, Filip Karlo, Mario Brčić, and Nikica Hlupić. 2018. "Explainable artificial intelligence: A survey." In *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)s.* IEEE.

Frosterus, Matias, Jouni Tuominen, Sini Pessala, and Eero Hyvönen. 2015. "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. https://doi.org/10.1504/IJMSO.2015.073879.

Heim, Philipp, Steffen Lohmann, and Timo Stegemann. 2010. "Interactive relationship discovery via the semantic web." In *Extended Semantic Web Conference,* 303–317. Springer.

Hickey, Thomas B., and Jenny A. Toves. 2014. "Managing Ambiguity In VIAF." *DLib Magazine* 20 (7/8). https://doi.org/doi:10.1045/july2014-hickey.

Hitzler, Pascal, and Md Kamruzzaman Sarker, eds. 2022. *Neuro-Symbolic Artificial Intelligence: The State of the Art.* IOS Press. https://doi.org/10.3233/FAIA210348.

Hyvönen, Eero. 2022. "Digital Humanities on the Semantic Web: Sampo Model and Portal Series." *Semantic Web Journal* 14 (4): 729–744. https://doi.org/10.3233/SW-223034.

Hyvönen, Eero. 2024. "Serendipitous knowledge discovery on the Web of Wisdom based on searching and explaining interesting relations in knowledge graphs." *Journal of Web Semantics* (December). https://doi.org/10.1016/j.websem.2024.100852.

Hyvönen, Eero, Annastiina Ahola, Petri Leskinen, Heikki Rantala, and Jouni Tuominen. 2025. "How to Create a Portal for Digital Humanities Research Using a Linked Open Data Cloud of Cultural Heritage Knowledge Graphs: Case SampoSampo." In *Proceedings of the Second International Workshop of Semantic Digital Humanities (SemDH 2025), co-located with the Extended Semantic Web Conference 2025 (ESWC 2025),* vol. 4009. CEUR Workshop Proceedings, June. https://ceur-ws.org/Vol-4009/paper_11.pdf.

Hyvönen, Eero, Annastiina Ahola, Petri Leskinen, and Jouni Tuominen. 2025. "SampoSampo: A Portal for Studying Enriched Data and Semantic Connections on a Cultural Heritage Linked Open Data Cloud." In *The Semantic Web: ESWC 2025 Satellite Events, Portoroz, Slovenia, June 1 - 5, 2025, Proceedings,* 67–74. Springer-Verlag. https://doi.org/10.1007/978-3-031-99554-5_13.

Hyvönen, Eero, Petri Leskinen, Minna Tamper, Heikki Rantala, Esko Ikkala, Jouni Tuominen, and Kirsi Keravuori. 2019. "BiographySampo - Publishing and Enriching Biographies on the Semantic Web for Digital Humanities Research." In *The Semantic Web. ESWC 2019,* edited by Pascal Hitzler, Miriam Fernández, Krzysztof Janowicz, Amrapali Zaveri, Alasdair J.G. Gray, Vanessa Lopez, Armin Haller, and Karl Hammar, 574–589. Springer-Verlag, June. https://doi.org/10.1007/978-3-030-21348-0_37. https://doi.org/10.1007/978-3-030-21348-0_37.

Hyvönen, Eero, and Heikki Rantala. 2021. "Knowledge-based relational search in cultural heritage linked data." *Digital Scholarship in the Humanities* 36 (Supplement 2): 155–164.

Isaac, Antoine. 2023. *Europeana Data Model Primer.* Technical report. Europeana. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf.

Maimon, Oded, and Lior Rokach. 2005. *Data mining and knowledge discovery handbook.* Vol. 2. 2005. Springer.

Nasar, Zara, Syed Waqar Jaffry, and Muhammad Kamran Malik. 2021. "Named entity recognition and relation extraction: State-of-the-art." *ACM Computing Surveys (CSUR)* 54 (1): 1–39.

Rantala, Heikki, Eero Hyvönen, and Jouni Tuominen. 2023. "Finding and explaining relations in a biographical knowledge graph based on life events: Case BiographySampo." In *CEUR Workshop Proceedings,* vol. 3443. RWTH Aachen University.

Silberschatz, Abraham, and Alexander Tuzhilin. 1995. "On Subjective Measures of Interestingness in Knowledge Discovery." In *KDD,* 95:275–281.

Tamper, Minna, Petri Leskinen, and Eero Hyvönen. 2019. "Visualizing and analyzing networks of named entities in biographical dictionaries for digital humanities research." In *International Conference on Computational Linguistics and Intelligent Text Processing,* 199–214. Springer.

Vrandečić, Denny, and Markus Krötzsch. 2014. "Wikidata: a free collaborative knowledgebase." *Communications of the ACM* 57 (10): 78–85.

Xia, Weiguo, Ernesto Jiménez-Ruiz, and Valerie V. Cross. 2015. "Using BioPortal as a Repository for Mediating Ontologies in Ontology Alignment." In *Workshop on Semantic Web Applications and Tools for Life Sciences.* https://api.semanticscholar.org/CorpusID:37359417.

A. Appendix: Prompt template

```
Context: You are a digital assistant with expert knowledge on person biographies and places.
Your task is to find documented and verifiable relations between people and places. You are
not allowed to make guesses or hallucinate information and all the sources you provide must
be accessible and trustworthy.

Your output should be an array in JSON format and you should not include any other
```

text outside of the JSON. Include all the reliable relations you can find; don't leave out relations if there are multiple different relations between the two entities. If there are no documented relations, you should return an empty array and no additional text. All relations should be directed.

The fields used should be the following:
1. "origin_node": [the origin of the directed relation]
2. "target_node": [the target of the directed relation]
3. "relation_type": [type of the relation]
4. "relation_description": [a brief textual description of the relationship, preferably one sentence in length, must include mentions of both the name of the person and the place in it]
5. "time_period": [an array of time period when the relation was applicable, in YYYY, YYYY-MM or YYYY-MM-DD format depending on known exactitude]
6. "sources": [an array of verifiable sources for the relation as web links (use attribute "url") as well as a description of the statements sourced from them in English (use attribute "statement")]
7. "reasoning": [the reasoning for including this relation in your answer]

For the field "relation_type", use the following types whenever possible:
1. "career_or_studies_related_to_place": The relation describes an activity related to the person's career or studies in the place.
2. "died_in_place": The relation describes the person's death in the place.
3. "born_in_place": The relation describes the person's birth in the place.
4. "honour_related_to_place": The relation describes the person having received an honorary award or title related to the place.
5. "painting_depicts_place": The relation describes the person's painting depicting the place.
6. "writing_depicts_place": The relation describes the person's writing depicting the place.
7. "letter_sent_from_place": The relation describes the person having sent a letter from the place.
8. "historical_event_in_place": The relation describes the person taking part in a historical event in the place.
9. "novel_depicts_place": The relation describes the person's novel depicting the place.
10. "letter_received_from_place": The relation describes the person having received a letter from the place.

If none of these are applicable, formulate a new relation type.

Question: How are *{LAST NAME, FIRST NAME (YEAR OF BIRTH - YEAR OF DEATH)}* and the place of *{NAME OF PLACE}* related to each other?

Here's additional information about *{LAST NAME, FIRST NAME (YEAR OF BIRTH - YEAR OF DEATH)}*'s life from their BiographySampo page (web link: *{URL OF PERSON'S BIOGRAPHYSAMPO PAGE}*):
"""
{PERSON'S BIOGRAPHICAL TEXT FROM THEIR PAGE}
"""