Knowledge-based relational search in cultural heritage linked data

Eero Hyvönen
Semantic Computing Research Group (SeCo), Aalto University, Espoo, and University of Helsinki (HELDIG), Helsinki, Finland

Heikki Rantala
Semantic Computing Research Group (SeCo), Aalto University, Espoo, Finland

Abstract

This article presents a new knowledge-based approach for finding interesting semantic relations between resources in a knowledge graph (KG). The idea is to characterize the notion of 'interesting connection' in terms of generic ontological explanation patterns that are applied to an underlying linked data repository to instantiate connections. In this way, (1) semantically uninteresting connections can be ruled out effectively and (2) natural language explanations about the connections can be created for the end-user. The idea has been implemented and tested based on a KG of biographical data extracted from the life stories of 13,144 prominent historical persons in Finland, enriched by data linking to collection databases of museums, libraries, and archives. The demonstrator is in use as part of the BiographySampo portal of interlinked biographies.

Correspondence:
Eero Hyvönen, Aalto University, Espoo, and University of Helsinki (HELDIG), Helsinki, Finland.
E-mail:
eero.hyvonen@aalto.fi

1 Relational Search as Knowledge Discovery

Knowledge discovery (Baker and Cheung, 2007) is one of the grand promises and challenges of the Semantic Web and its applications in Digital Humanities (Hyvönen, 2020). This article concerns the problem of discovering relations (a.k.a. connections, associations) in semantically rich, linked Cultural Heritage (CH) data (Hyvönen, 2012), i.e. 'knowledge graphs' (KG). In particular, we focus on the 'relational search' problem of finding 'interesting' connections between the resources in a KG, such as persons, places, and events. For example, how are American novelists of the twentieth century related to France? Such semantic connections can be based on various criteria: a person (or her/his family member) was born or died in Paris, French topics were discussed in her/his novels, (s)he wrote a novel or an article in French, her/his publisher was a French company, her/his portrait is in Louvre, (s)he got a medal of honour in Lyon, and so on.

What is considered 'interesting' is a tricky question (Silberschatz and Tuzhilin, 1995). The answer also depends on the user. The more knowledgeable (s)he is, the less interesting new connections (s)he is likely to found as (s)he already knows a lot. There are also different degrees of interestingness. In this article, the anticipated users of the application to be presented

doi:10.1093/llc/fqab042

are school children, students, and ordinary people interested in learning history, not professional historians solving focused academic problems. In the following, a connection is considered interesting, if it is deemed new, surprising, and valuable to the anticipated user.

Given the richness of possible semantic connections, solving relational search problems can be seen as an instance of computational creativity (Boden, 2009), an example of the subtype 'exploratory creativity', where creativity refers to search within a predefined search space under given constraints for the solutions. In the following, relational search methods are first discussed. Two major challenges are identified: (1) filtering out interesting connections from not interesting ones and (2) creating explanations for the interesting connections, a challenge of explainable Artificial Intelligence (Došilovíc *et al.*, 2018). As a remedy, a new knowledge-based approach is presented. To test and evaluate the method, a case study of applying this approach is then presented in the CH domain by using a large KG of biographical data. In conclusion, lessons learned are discussed, and further research suggested. This article is an extended version of the papers/abstracts presented at the Digital Humanities in Nordic Countries Conference (DHN 2019) in Copenhagen and Digital Humanities 2019 conference (DH 2019) in Utrecht.

2 Approaches to Relational Search

In relational search, the query consists of two or more resources, and the task is to find semantic relations, i.e. the query results, between them that are of interest to the user. This problem has been addressed before in different domains. The approaches reported in the literature (Cheng *et al.*, 2017) differ in terms of the query formulation, underlying KG, methods for finding connections, and representation of the results. Some sources of inspiration for our own work are shortly reviewed below.

2.1 Domain knowledge agnostic approaches

In Sheth *et al.* (2005) the idea of searching relations is applied for association finding in a national security domain. Within the CH domain, CultureSampo[1] (Hyvönen *et al.*, 2009; Mäkelä *et al.*, 2012) contains an application perspective where connections between two persons were searched using a breath-first algorithm, and the result was a list of arcs (such as student-of, patron-of, etc.), connecting the persons based on the Getty Union List of Artist Names (ULAN)[2] KG of historical persons. In RelFinder[3] (Heim *et al.*, 2010, 2009; Lohmann *et al.*, 2010), based on the earlier 'DBpedia Relationship Finder' (Lehmann *et al.*, 2007), the user selects two or more resources, and the result is a minimal visualized graph showing how the query resources are related with each other. For example, Albert Einstein is related to Kurt Gödel in DBpedia/Wikipedia because both gentlemen, e.g. worked at the Princeton University. In WiSP (Tartari and Hogan, 2018), several paths with a relevance measure between two resources in the Wikidata KG[4] can be found, based on different weighed shortest path algorithms. The query results are represented as graph paths. Some applications, such as RelFinder and Explass (Cheng *et al.*, 2014), allow filtering relations between two entities with facets.

From a methodological perspective, the main challenge in the relational search systems discussed in the previous paragraph is how to select and rank the interesting paths, since there are exponentially many possible paths between the query resources in a KG that are not interesting. This problem can be approached by focusing only on 'simple paths' that do not repeat nodes, on only restricted node and arc types in the graph (e.g. social connections between persons), and by assuming that shorter, possibly weighted paths are more interesting than longer ones. For weighting paths, measures such as page rank of nodes and commonness of arcs, can be used.

The notion of serendipitous knowledge discovery is also related to recommender systems (Jannach *et al.*, 2011). The problem of explaining associations between resources in KG is addressed in the field of explainable Artificial Intelligence (AI) (Lecue, 2020). The idea of providing related information with explanations can also be seen in commercial search engines, such as Google, that provide the user with additional information about entities found, such as persons, e.g. schools (s)he studied in, books (s)he wrote, etc. In the related works above the notion of 'explanation' is a path or a subgraph connecting the target resources, such as persons and films or places. In contrast, our focus is on creating explanations written in natural language. Our work is also related to the field of question answering where answers to natural language questions are determined (Kolomiyets and Moens, 2011). In our case, however, the focus is on formulating natural language explanations of answers to queries expressed as selections in faceted search.

2.2 A knowledge-based approach

The graph-based methods above make use of generic graph traversal algorithms that are application domain agnostic. In contrast, this article suggests an alternative, a 'knowledge-based' approach to finding interesting connections in a KG. The idea is to formalize the notion of 'interestingness' (Silberschatz and Tuzhilin, 1995) in the application domain using general explanation patterns that can be instantiated in a KG using graph transforming rules. In this way, relational search of finding connection paths in a graph can be reduced into a search on explanation instances in a simpler search space created using knowledge-based rules.

The proposed method consists of the following steps:

(1) Identify and select entity types (e.g. persons and places) whose mutual relations are to be searched for.
(2) Organize the entities of these types into hierarchical facets (ontologies).
(3) Create knowledge-based graph transformation rules for creating instances of explanations whose properties include (a) the interestingly related entities and (b) a natural language explanation about their semantic connection.
(4) Solve relational search problems as faceted search problems (Tunkelang, 2009) in the new explanation instance search space. This means in practice that the user selects the end-point types or entity instances on the facets, after which the search results are the connections of interest between the selections, with explanations attached to them.

The argued benefits of this approach are: (1) nonsense relations between the query resources can be ruled out effectively by the knowledge-based rules and (2) the explanation patterns can be used for creating natural language explanations for the connections. The price to be paid is the need for crafting the transformation rules and their explanation patterns manually, based on application domain knowledge, as customary in knowledge-based system.

3 Finding Semantic Relations in a Biographical KG

To explain, test, and evaluate our knowledge-based approach in more detail we next consider its application in the semantic portal BiographySampo—Finnish Biographies on the Semantic Web[5] (Hyvönen *et al.*, 2019).

3.1 Knowledge graph

The KG underlying our system was created using the following interlinked datasets:

(1) The core dataset is the biographical data of BiographySampo extracted in RDF form[6] from 13,144 Finnish biographies, including, e.g. 51,937 family relations, 4,953 places, 3,101 occupational titles, and 2,938 companies. The data model used is an extension of CIDOC CRM.[7]
(2) HISTO ontology[8] of Finnish history including more than thousand historical events. Data for the events include, e.g. people and places related to the event and event type. The data were available in RDF format.
(3) The Fennica, National Bibliography of Finland,[9] is an open database of Finnish publications since 1488. The metadata includes, among other things, the author of the book and the subject matter of the book, which can include places. Also, these data were available in RDF form.
(4) BookSampo[10] data covering virtually all Finnish fiction literature in RDF format, maintained by the Finnish Public Libraries consortium Kirjastot.fi.
(5) The Finnish National Gallery[11] has published the metadata about the works of art in their collections. The metadata is described using Dublin Core standard and was available in

JSON and XML format that was transformed into RDF.

(6) The collected works of the J. V. Snellman portal[12] include the texts written by J. V. Snellman, the national philosopher of Finland. The data include, e.g. 1,500 letters. We transformed the data into RDF.

3.2 Applying the Method

The four-step method of Section 2.2 was applied as follows:

- (Step 1) We decided to search for relations between people and places.
- (Step 2) Next, we used the person and place ontologies of BiographySampo as the basis of entity ontologies. The occupation ontology and place hierarchy of BiographySampo were used to allow faceted search based on properties of the entities. In addition, an ontology of relation types was created. In general, new ontologies could at any point be added and linked to the entity ontologies to allow faceting based on any property of the person, place, or the relation.
- (Step 3) As for the graph transformations rules, SPARQL[13] CONSTRUCT queries were used on top of the BiographySampo linked data service hosted by the Linked Data Finland platform[14] (Hyvönen *et al.*, 2014). The queries transformed (part of) the KG into a new KG of connection instances.
- (Step 4) Based on the transformed data, relational search queries can now be expressed in terms of selections on the facets and be solved efficiently using faceted search. In our case, the faceted search engine was implemented with the SPARQL Faceter[15] (Koho *et al.*, 2016) tool.

The person ontology we used from BiographySampo is based on manually coded metadata of the 13,144 biographees in the national biography collection of the Biography Centre at the Finnish Literature Society.[16] We limited this application to only dead people, to avoid possible data protection issues of personal data. The place ontology was based on the place name thesaurus YSO Paikat,[17] maintained by the National Library of Finland, but uses a more complete hierarchy created for BiographySampo. A key ontological resource in the KG and 'semantic glue' for determining the connections were CIDOC CRM events that were extracted from the semiformal textual parts of the biographies, as well as from the metadata of the collection items from memory organizations. For example, metadata about a painting was transformed into an event where the painting was created by the artist at the time and in the place mentioned in the metadata and linked with the underlying ontologies during knowledge extraction. The knowledge extraction and entity linking process of BiographySampo is explained in some more detail in (Tamper *et al.*, 2018; Hyvönen *et al.*, 2019). Although formal analyses of the data quality in BiographySampo have not been published, informal testing during the project of this article suggests that errors in the data are seldom encountered, and the overall data quality is good enough for practical applications.

A connection instance in the new search space has the following core properties: (1) a literal natural language expression that explains the connection in a human-readable form. (2) A set of properties that explicate the resources that are connected. Relation instances like this can be searched for in a natural way using faceted search, where the facets are based on the property values of the instances. By making selections on the facets the result set is filtered accordingly and hit counts in the facet categories are recalculated. Facet categories can be organized into hierarchies; selecting a super category then means that all subcategories are selected with one click. For example, selecting 'Finland' means that all places in Finland are automatically selected.

The focus of our demonstrator is on finding relations describing connections between people and places in Finnish cultural history. The relation instances listed in Table 1 were created using SPARQL CONSTRUCT queries whose application to the data generated connection instances with related natural language explanations. For example, the following query can be used to create connections between people and their death places and times:

```
# Namespace definitions
BASE <http://ldf.fi/relse/>
PREFIX nbf: http://ldf.fi/nbf/
```

Table 1. Connection types and instance counts

Type of connection	Number of connections
Historical event in a place	345
Letter sent from	575
Letter received from	124
Text describes a place	881
Received an award in a place	2,528
Died in	7,349
Painting depicts a place	1,091
Novel depicts a place	290
Born in	7,182
Career is related to a place	20,536
In total	40,901

```
PREFIX rel: <http://ldf.fi/relse/>
...
# Template for constructing connec-
tion instances
CONSTRUCT {
? uri a rel: Relation;
  rel: relationType rel: deathPlace;
  rel: personSubject? person;
  rel: placeObject? place;
  rel: date? deathtime;
  skos: prefLabel? description;
  rel: source? death.
}
# Matching the variables for con-
structing the connections above
WHERE {
# Person
   ? death crm: P100_was_death_of/
   ^foaf: focus? person.
  ? person skosxl: prefLabel/schema:
  familyName? familyName.
  ? person skosxl: prefLabel/schema:
  givenName? givenName.
# Place
  ? death nbf: place? place.
  ? place skos: prefLabel? placeName.
  FILTER(lang(? placeName) = 'fi').
# Time
   ? death nbf: time/gvp: estStart?
   deathtime.
  BIND (year(xsd: date(? deathtime))
  As? year)
# URI
  BIND(uri(encode_for_uri(concat
  (str(? person), str(? place),
"death_place", str(? death)))) as?
uri).
# Natural language explanation
  BIND(concat(str(? givenName), " ",
  str(? familyName),
   " on kuollut paikassa ", str(?
   placeName), " vuonna ",
  str(? year), ".") as? description).
}
```

The query consists of the following parts marked by comment lines beginning with '#': First, the prefixes for namespaces are introduced; only some of them are seen in the listing for brevity. Next, the CONSTRUCT template for generating connection instances is presented in terms of variables beginning with '?'. The value bindings for the variables are determined by matching the WHERE template in all possible ways with the underlying KG. The WHERE template matches first the person and then the place and time of death. After this, a URI identifier for the connection instance is concatenated from the matched variables using the concat function of SPARQL. Finally, the natural language explanation '? *givenName*? *familyName* has died in place? *placename* in the year? *year*' (in Finnish) of the connection instance is concatenated in the same way.

The CONSTRUCT queries are based on the unique identifiers in the KG, not on their literal labels, and therefore do not, e.g. introduce errors due to ambiguity of literal names. Literal names are only used for creating the natural language explanations.

4 Demonstrator at Work

The method was implemented as the tool FacetedRelator that was published as part of the BiographySampo portal and is in use online[18] as a separate application perspective. Figure 1 depicts the user interface of the application. The data and interface are in Finnish, but there is a Google Translate button in the right upper corner of the interface for foreign users available.

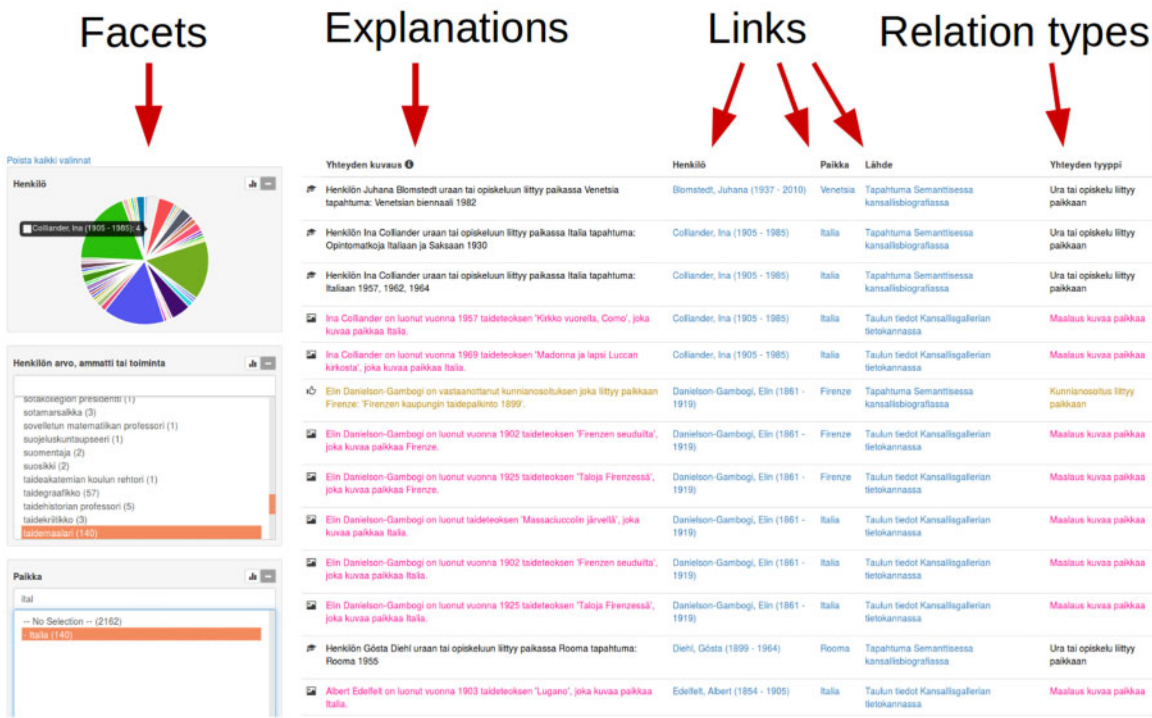

Fig. 1 View of the user interface

In this case study, FacetedRelator can be used for filtering relations with selections in four facets seen on the left: (1) person names, (2) occupations, (3) places, and (4) relation types. The idea of faceted search (Tunkelang, 2009), sometimes also called as 'view-based search' and 'dynamic ontologies', originates from the idea of faceted classification by S. R. Ranganathan in library science. Here the search items are projected on a set of orthogonal semantic classifications for filtering results, instead of on only one enumerative classification, such as the Dewey Decimal Classification. In our case, selections on the facets set constraints for the relation instances between people and places, i.e. results of search. The system shows a hit list of the relation instances that fit the selected filtering criteria in the facets. The user is not required to first input a person and a place but can limit the search at any time with a selection on any facet. Furthermore, the fact, that the facets use hierarchies defined in the ontologies, allows searching for relations between groups of people (on the occupations facet, e.g. 'film director') and larger areas ('e.g. South America') instead of individual persons or places. After each selection, the hit counts on the facet categories tell how many results there will be in the result set if a category is selected next. In this way, the user is guided towards filtering the solutions and never ends up in a 'no hits' situation. The hit counts can also be used for visualizing the distribution of the results along each facet dimension, which is useful in quantitative analyses.

Each connection instance is represented in a row in the hit list on the right. A row shows first the natural language explanation of the connection, then the related person, place, main data source, and finally the relation type (compare Table 1). Persons, places, and data sources are represented as links for further information. For example, the person link leads to the 'home page' of the person in BiographySampo that automatically reassembles and visualizes the life story of the person based on the various interlinked datasets of the system. Different types of relations are highlighted in different colors and have their own symbols to give the user a visual overview of different kind of relations found. At any point, the distribution of the hit counts in categories along each facet can be visualized using a pie chart—one of them can be seen in the left upper corner of Fig. 1.

For example, the question ‘How are Finnish painters related to Italy?’ is solved by selecting ‘Italy’ from the hierarchical place facet and ‘painter’ from the occupation facet. Any selection automatically includes its subcategories in the facet. For example, places such as Florence and Rome are in Italy, and Vatican further in Rome. The result set in this case contains 140 connections of different types whose distribution and hit counts can be seen on the connection type facet. In the same way, the person facet shows the hit count distribution along the person facet. Any facet could be used to filter the results further if needed. In this case, the 140 hits include, e.g. connection ‘Elin Danielson-Gambogi received in 1899 the Florence City Art Award’ and ‘Robert Ekman created in 1844 the painting “Landscape in Subiaco” depicting a place in Italy’.[19]

In faceted search, the hit counts of facet categories tell the quantitative distributions of the results along the facet categories. This feature is utilized in FacetedRelator by making it possible to study the distributions as pie charts by clicking on a button on a facet. This feature can be used in FacetedRelator for solving some quantitative research problems. For example, Fig. 2 illustrates how the question ‘Who has got most awards related to Germany’ can be solved by selecting the connection type ‘Received an award in a place’ (In Finnish: ‘Kunnianosoitus liittyy paikkaan’) on the connection type face on the bottom, and on the place facet above it ‘Germany’ (In Finnish: ‘Saksa’, including the cities and other places there listed as facet subtypes). By hitting a button on the people facet, the hit distribution and pie chart along the people facet shows immediately that general Carl Gustaf Mannerheim is the winner with eight awards out of the filtered 234 awards. When using the application, it is important to note that even in the best case the demonstrator is limited by the sources and data it uses. A relation can be missing for several reasons and this kind of relative numbers may not therefore reflect reality perfectly. However, they can be valuable for finding out interesting phenomena in the data for further close reading by the human expert.

5 Discussion: Lessons Learned

From a computational point view, the knowledge-based approach presented in this article was deemed feasible at least in this use case: the data transformations during the preprocessing phase could be performed efficiently in a few seconds per query. When querying the data in the portal, response times are nearly instant.

To see if the system is fit for its purpose, i.e. whether the connections returned make sense to the user and are valuable, an evaluation test was made. For this purpose, a small set of persons were selected and their connections to all places, found by the system, were analyzed manually. The test was then repeated reversely, starting from a set of selected places. In these tests, the system worked correctly in terms of precision returning only feasible explanations in all test cases.[20] This was not a big surprise, as the connections in our method are determined by explicit logical rules. As for recall, evaluation of the results is challenging, as there is no gold standard available, and failing to find a connection may be, e.g. due to sparsity of the data, not the method. In any case, as Table 1 shows, the system was able to find lots of interesting relations for further close reading in the data, and the approach looks promising.

According to Boden (2009), a system can be considered creative if it can create ‘new’, ‘surprising’, and ‘valuable’ ideas. At least from a layman perspective, this seems to be the case in FacetedRelator although measuring creativity is not easy. Given the large, semantically rich KG we believe that the system can provide insightful results even for an expert historian. However, more testing is needed to find out how interesting and surprising the results are for an expert of CH and how a system like this can be used for DH research.

The knowledge acquisition task of formulating a set of useful explanation patters and graph transformation rules in the demonstrator could be finished in a reasonably short time. At least in this case the number of connections found was not overwhelmingly large from a computational point of view, as shown in

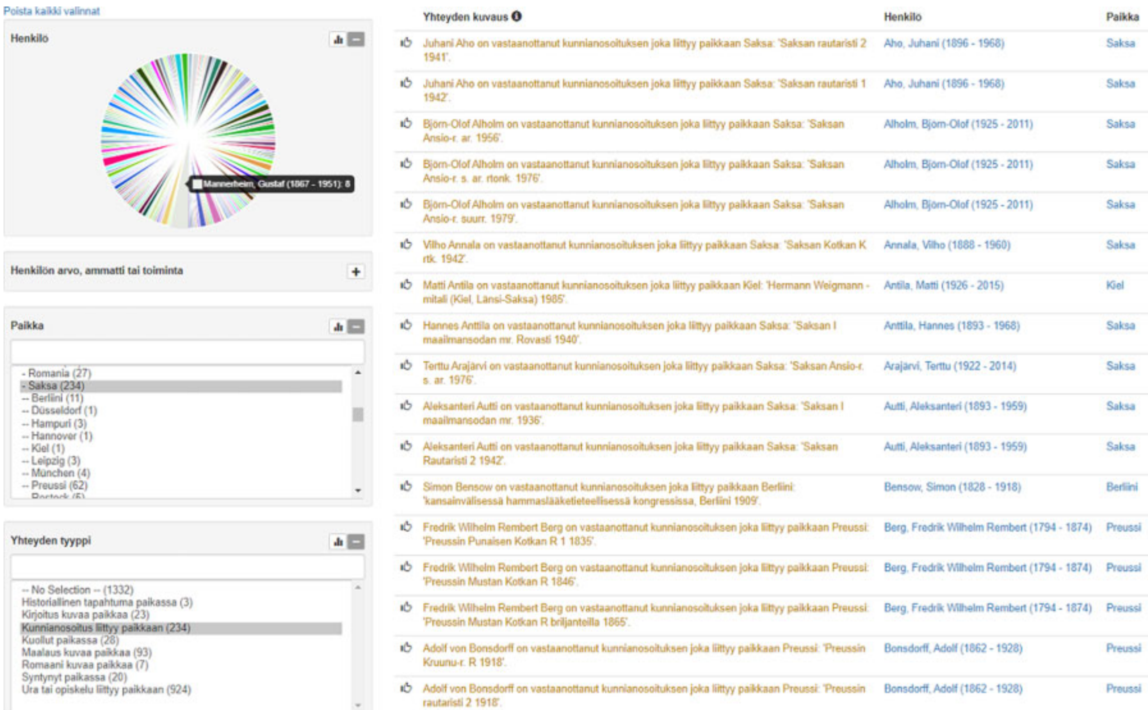

Fig. 2 Solving the problem: 'who has got most awards in Germany?'

Table 1, and could be generated quickly. From a human end-user perspective, the result set (40,901 connections) is still large enough to provide many non-trivial results and explanations. So, the suggested knowledge-based approach was deemed feasible at least in cases where the potentially interesting connection can be characterized logically, and their number is not very large. This seems to be the case in the biographical datasets of BiographySampo.

If the constraints on interestingness, i.e. the transformation rules, are loosened too much, there is the danger for combinatorial explosion of results. However, generating lots of connections probably means that they are not interesting, and should therefore not be done in the first place. For example, the connection that two persons are born in the same country would connect most of the people in our data and would not be interesting and worth generating. However, if the persons are born in a small village and about the same time, the connection would be much more interesting. We believe domain knowledge is useful and, in many cases, necessary in making such fine-grained distinctions of interestingness.

In our demonstrator, the connections are generated in a pre-processing phase, which may not be cost-efficient if there are lots of connections that will never be queried. In such situations, connection generation or part of it could be performed on query time to balance computational effort needed during pre-processing and querying. In our case, the relations are generated from multiple sources and endpoints with separate query for each type of relation. The queries were not in any way optimized and took time from few seconds to half an hour, which was computationally feasible.

When testing and evaluating the demonstrator, we also found out needs to improve the usability of the system. For example, the demonstrator now sorts results based on firstly the name of the person and secondly on the name of the place. The user should probably be offered the possibility to sort the relations freely along any facet. Faceted search does not corporate the notion of relevance of search results; the results can be clustered flexibly along various facet categories but there is no criterium for ordering them. However, especially if there are lots of results, ordering the results based on a relevance criterion, as in traditional search engines, such as Google, would be useful, too.

The knowledge-based approach presented in this article seems particularly suitable in applications, where knowledge and criteria for interestingness from the end user's perspective in available and can be formulated. The domain of CH applications seems promising from this perspective. From a technical perspective, the number of interesting connections should not be overwhelmingly large in the KG. However, if the transformation rules produce lots of such connections the criterion for interestingness is probably too loose, and the connections not so interesting.

Acknowledgements

The authors wish to acknowledge CSC—IT Center for Science, Finland, for computational resources.

Funding

Our research was supported by the Severi Project,[21] funded mainly by Business Finland.

References

Baker, C. and Cheung, K.-H. (eds) (2007). *Semantic Web—Revolutionizing Knowledge Discovery in the Life Sciences*. Berlin/Heidelberg: Springer.

Boden, M. A. (2009). Computer models of creativity. *AI Magazine*, **30**(3): 23–34.

Cheng, G., Shao, F., and Qu, Y. (2017). An empirical evaluation of techniques for ranking semantic associations. *IEEE Transactions on Knowledge and Data Engineering*, **29**(11): 1.

Cheng, G., Zhang, Y., and Qu, Y. (2014). Explass: exploring associations between entities via top-K ontological patterns and facets. In Mika, P., Tudorache, T., Bernstein, A., et al. (eds), *The Semantic Web – ISWC 2014*, Cham: Springer International Publishing, pp. 422–437.

Došilović, F. K., Brčić, M., and Hlupić, N. (2018). Explainable artificial intelligence: a survey. In 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)s, Rijeka, Croatia: IEEE

Heim, P., Hellmann, S., Lehmann, J., Lohmann, S., and Stegemann, T. (2009). RelFinder: revealing relationships in RDF knowledge bases. In *Proceedings of the 4th International Conference on Semantic and Digital Media Technologies (SAMT 2009)*, Springer.

Heim, P., Lohmann, S., and Stegemann, T. (2010). Interactive relationship discovery via the semantic web. In Lora A., Grigoris, A., Hyvönen, E., et al. (eds), *The Semantic Web: Research and Applications*, vol. **6088**, Berlin/Heidelberg: Springer, pp. 303–317.

Hyvönen, E. (2012). *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. Palo Alto, CA: Morgan & Claypool.

Hyvönen, E., Mäkelä, E., Kauppinen, T. et al. (2009). CultureSampo—Finnish culture on the Semantic Web 2.0. Thematic perspectives for the end-user. In *Proceedings of the Museums and the Web 2009*, Archives and Museum Informatics, Toronto.

Hyvönen, E., Tuominen, J., Alonen, M., and Mäkelä, E. (2014). Linked Data Finland: a 7-star model and platform for publishing and re-using linked datasets. In Presutti, V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., and Tordai, A., (eds), *The Semantic Web: ESWC 2014 Satellite Events*, Cham*:* Springer, pp. 226–30.

Hyvönen, E. (2020). Using the Semantic Web in Digital Humanities: shift from data publishing to data-analysis and serendipitous knowledge discovery. *Semantic Web*, **11**(1):187–93.

Hyvönen, E., Leskinen P., Tamper M. et al. (2019). BiographySampo—publishing and enriching biographies on the semantic web for digital humanities research. In *Proceedings of the 16th Extendwed Semantic Web Conference (ESWC 2019)*, Springer.

Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2011). *Recommender Systems. An Introduction.* Cambridge: Cambridge University Press.

Koho, M., Heino, E., and Hyvönen, E. (2016). SPARQL Faceter—client-side faceted search based on SPARQL. In *Joint Proceedings of the 4th Interna- tional Workshop on Linked Media and the 3rd Developers Hackshop*. CEUR Workshop Proceedings, Heraklion, Crete, Greece, vol. **1615**.

Lecue, F.(2020). On the role of knowledge graphs in explainable AI. *Semantic Web*, **1**(11): 41–51.

Lehmann, J., Sch üppel, J., and Auer, S. (2007). Discovering unknown connections—the DBpedia relationship finder. In *Proceedings of the 1st Conference on Social Semantic Web (CSSW 2007)*, vol. **113** of LNI, Bonner Köllen, pp. 99–110.

Lohmann, S., Heim, P., Stegemann, T., and Ziegler, J. (2010). The RelFinder user interface: interactive exploration of relationships between objects of interest. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI 2010)*, ACM, Hong Kong, China, pp. 421–22.

Mäkelä, E., Ruotsalo, T., and Hyvönen, E. (2012). How to deal with massively heterogeneous cultural heritage data—-lessons learned in CultureSampo. *Semantic Web—Interoperability Usability, Applicability*, **3**(1): 85–109.

Kolomiyets, O. and Moens, M.-F. (2011). A survey on question answering technology from an information retrieval perspective. *Information Sciences*, **181**(24): 5412–34.

Sheth, A., Aleman-Meza, B., Arpinar, I. B. et al. (2005). Semantic association identification and knowledge discovery for national security applications. *Journal of Database Management on Database Technology*, **16**(1): 33–53.

Silberschatz, A. and Tuzhilin, A. (1995). On subjective measures on interestingness in knowledge discovery. In *Proceedings of KDD-1995*, AAAI Press, Montreal, Quebec, Canada.

Tartari, G. and Hogan, A. (2018). WiSP: weighted shortest paths for RDF graphs. In *Proceedings of VOILA 2018. CEUR Workshop Proceedings*, vol. **2187, Monterey, CA, USA**.

Tamper, M., Leskinen, P., Apajalahti, K., and Hyvönen, E. (2018). Using biographical texts as linked data for prosopographical research and applications. In *Digital Heritage. 7th International Conference on Progress in Cultural Heritage: Documentation, Preservation, and Protection, EuroMed 2018*, Springer, Nicosia, Cyprus.

Tunkelang, D. (2009). *Faceted Search. Synthesis Lectures on Information Concepts, Retrieval, and Services.* Palo Alto, Ca: Morgan & Claypool Publishers, pp. 1–80.

Notes

1 http://www.kulttuurisampo.fi.
2 http://www.getty.edu/research/tools/vocabularies/ulan/.
3 http://www.visualdataweb.org/relfinder.php.
4 http://wikidata.org.
5 In use at http://biografiasampo.fi.
6 https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/.
7 http://cidoc-crm.org.
8 https://seco.cs.aalto.fi/ontologies/histo/.
9 https://www.kansalliskirjasto.fi/en/services/conversion-and-transmission-services- of-metadata/open-data.
10 https://www.ldf.fi/dataset/kirjasampo/index.html.
11 https://www.kansallisgalleria.fi/en/avoin-data/.
12 http://snellman.kootutteokset.fi/.
13 https://www.w3.org/TR/sparql11-overview/.
14 https://ldf.fi.
15 https://github.com/SemanticComputing/angular-semantic-faceted-search.
16 https://kansallisbiografia.fi/.
17 https://finto.fi/yso-paikat/fi/
18 http://biografiasampo.fi/yhteyshaku/
19 These explanations are in Finnish and are translated here in English for illustration.
20 These tests will be elaborated in a forthcoming paper.