Knowledge-based Approach to Relational Search in Knowledge Graphs with Explanations: Case BiographySampo – Biographies on the Semantic Web

HEIKKI RANTALA, Aalto University, Semantic Computing Research Group (SeCo), Finland
EERO HYVÖNEN, University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG) and Aalto University (SeCo), Finland

This paper presents a new knowledge-based approach for finding serendipitous semantic relations between resources in a knowledge graph. 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 knowledge graph of biographical data extracted from the short biographies of 13 100 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 that has had some 50 000 users.

Additional Key Words and Phrases: linked data, knowledge discovery

ACM Reference Format:
Heikki Rantala and Eero Hyvönen. 2018. Knowledge-based Approach to Relational Search in Knowledge Graphs with Explanations: Case BiographySampo – Biographies on the Semantic Web. 1, 1 (October 2018), 16 pages. https://doi.org/10.1145/1122445.1122456

1 RELATIONAL SEARCH AS SERENDIPITOUS KNOWLEDGE DISCOVERY

Serendipitous knowledge discovery [3] is one of the grand promises and challenges of the Semantic Web and its applications in Digital Humanities [21]. Serendipity means 'happy accident' or 'pleasant surprise', even 'fortunate mistake'. According to the Merriam-Webster dictionary[1] serendipity is "the faculty or phenomenon of finding valuable or agreeable things not sought for".

This paper concerns the problem of discovering and explaining serendipitous relations (a.k.a connections, associations) in semantically rich, linked Cultural Heritage (CH) data [18], i.e., in *Knowledge Graphs* (KG). In particular, we focus on the *relational search* problem of finding "interesting" [38] connections between the resources in a KG, such as persons, places, and events. For example: how are American novelists of the 20th 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

[1] https://www.merriam-webster.com/dictionary/serendipity

Authors' addresses: Heikki Rantala, Aalto University, Semantic Computing Research Group (SeCo), Finland, heikki.rantala@aalto.fi; Eero Hyvönen, University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG) and Aalto University (SeCo), Finland.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2018 Association for Computing Machinery.
XXXX-XXXX/2018/10-ART $15.00
https://doi.org/10.1145/1122445.1122456

of honour in Lyon, and so on. Given the richness of possible semantic connections, solving relational search problems can be seen as an instance of computational creativity [6], 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 [10]. As a remedy, a new knowledge-based approach is presented. Previous solutions tend to focus on relations between two (or more) entities. To generalize this to search relations between groups of entities, such as occupational groups (e.g., artists) and larger geographical areas (e.g., Italy), we suggest a solution based on hierarchical ontologies and faceted search.

To test and evaluate the method, a case study of applying this approach is then presented in the Cultural Heritage domain by using a large KG of biographical data. In conclusion, lessons learned are discussed, and further research suggested. This paper is a substantially extended version of the short papers/abstracts presented at the Digital Humanities in Nordic Countries Conference (DHN 2019) in Copenhagen [23] and Digital Humanities 2019 conference (DH 2019) in Utrecht [24].

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 [8] 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 Related Works

In [37] the idea of searching relations is applied for association finding in national security domain. Within the CH domain, CultureSampo[2] [19, 32] contains an application perspective where connections between two persons were searched using a breadth-first algorithm, and the result was a list of chains of arcs (such as *student-of*, *patron-of*, etc.), connecting the persons based on the Getty ULAN[3] knowledge graph of historical persons. In RelFinder[4] [15, 16, 31], based on the earlier "DBpedia Relationship Finder" [29], 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 [40], several paths with a relevance measure between two resources in the WikiData KG[5] can be found, based on different weighed shortest path algorithms. The query results are represented as graph paths. One possible approach is ranking results based on how familiar the elements related to the information are to the user [1]. Some applications, such as RelFinder and Explass [9], allow filtering relations between two entities with facets.

From a methodological perspective, the main challenge in these systems 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

[2]http://www.kulttuurisampo.fi
[3]http://www.getty.edu/research/tools/vocabularies/ulan/
[4]http://www.visualdataweb.org/relfinder.php
[5]http://wikidata.org

page rank of nodes and commonness of arcs, can be used. The problem of ranking ranking the relations found is discussed, e.g., in [8], and learning how to rank in [4].

The notion of serendipitous knowledge discovery is also related to recommender systems [25], where the idea is to supplement direct search results with other indirect hits in exploratory information search [33]. The main approaches here are knowledge-based methods based on metadata and ontology structures [7, 34] and collaborative filtering techniques that utilize user data statistically. A challenge here is how to make the recommendations serendipitous [2] and break out from the "information bubble" ("echo chamber") of the user. For example, if a user is a football fan or a member of a political party, it would be refreshing to recommend also some related non-football content and views of other parties. This problem is relevant when creating personalized systems.

The problem of explaining associations between resources in knowledge graphs (KG) is addressed in the fields of explainable AI [11, 28]. In [5] two algorithms and a tool RECAP are presented in for this purpose: E4D based on explaining individual paths between given resources in a knowledge graph, and E4S where additional schema information and a target predicate are used for focusing on more interesting explanations. In contrast to these, our method is not based on the schema but on additional domain knowledge patters of interestingness, that are used both for finding the connecting paths in the first place, and for explaining them. Explanations have been studied also in the context of recommender systems [17]. The idea of proving 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 [27]. In our case, however, the focus is on formulating natural language explanations of answers to queries expressed as selections in faceted search [41]. In creative AI [44], novel serendipitous knowledge structures, such as music, texts, and images, are created by the machine. As there there are often lots of examples available, statistical learning techniques are nowadays commonly used here.

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 paper suggests an alternative, a *knowledge-based* approach to finding interesting connections in a KG. The idea is to formalize the notion of "interestingness" [38] in the application domain using general explanation patterns that can be instantiated in a KG using graph transforming rules. This approach is related to the idea of Ontology Design Patterns[6] (ODP) [12], where reusable solutions to recurring modeling problems are studied. In our approach, 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 1) the interestingly related entities and 2) a natural language explanation about their semantic connection.
(4) Solve relational search problems as faceted search problems [41] in the new explanation instance search space. This means in practise selecting the end point types or entity instances on facets, after which the search results are the connections of interest between the selections with explanations.

[6]http://ontalogydesignpatterns.org/

The argued benefits of this approach are: 1) Non-sense 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.

It is possible that the number of potentially interesting connection instances becomes very large and computationally challenging. For example, if n=1000 people are born in the same place P and one would search for connections of the type "persons X and Y are born in the same place P", then there would be $n * (n-1)/2 = 499500$ connections. However, if there are lots connections, then such common connections are likely not to be very interesting. Using a more refined connection type can reduce the number of connections. For example, one could search for connected persons born in a small village or born around the same time in a larger community.

3 FINDING SEMANTIC RELATIONS IN A BIOGRAPHICAL KNOWLEDGE GRAPH

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*[7] [22].

3.1 Knowledge Graph

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

(1) The core dataset is the biographical data of BiographySampo extracted in RDF form[8] from 13 144 Finnish biographies, including, e.g., 51 937 family relations, 4953 places, 3101 occupational titles, and 2938 organizations. The data model used is Bio CRM [42], an extension of CIDOC CRM[9] for biographical data.
(2) HISTO ontology[10] of Finnish history including more than thousand historical events. Data for the events include, e.g., related people, places, and times. The data was available in RDF format.
(3) The Fennica National Bibliography[11] is an open database of Finnish publications since 1488. The metadata includes, among other things, the author of the publication and its subject matter, which can include places. Also this data was available in RDF form.
(4) BookSampo[12] (meta)data covering virtually all Finnish fiction literature in RDF format, maintained by the Finnish Public Libraries consortium Kirjastot.fi.
(5) The Finnish National Gallery[13] 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[14] includes metadata and the texts written by J. V. Snellman, the national philosopher of Finland. The data includes, e.g., 1500 letters. We transformed the metadata 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.

[7] In use at http://biografiasampo.fi.
[8] https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/
[9] http://cidoc-crm.org
[10] https://seco.cs.aalto.fi/ontologies/histo/
[11] https://www.kansalliskirjasto.fi/en/services/conversion-and-transmission-services-of-metadata/open-data
[12] https://www.ldf.fi/dataset/kirjasampo/index.html
[13] https://www.kansallisgalleria.fi/en/avoin-data/
[14] http://snellman.kootutteokset.fi/

- (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[15] CONSTRUCT queries were used on top of the BiographySampo linked data service hosted by the Linked Data Finland platform[16] [20]. 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[17] [26] tool.

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 supercategory 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 in 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 the 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/> # Namespace for connection instances

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX skosxl: <http://www.w3.org/2008/05/skos-xl#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX gvp: <http://vocab.getty.edu/ontology#>
PREFIX schema: <http://schema.org/>
PREFIX rel: <http://ldf.fi/relse/>
PREFIX nbf: <http://ldf.fi/nbf/>

# Template for constructing connection instances
CONSTRUCT {
    ?uri a rel:Relation ;
        rel:relationType rel:deathPlace ;
        rel:personSubject ?person ;
        rel:placeObject ?place ;
        rel:date ?deathtime ;
        rel:source ?death ;
        rel:sourceName "Tapahtuma Semanttisessa kansallisbiografiassa" ;
        skos:prefLabel ?description .
}

# Matching the variables for constructing the connections above
WHERE {
# Person
```

[15] https://www.w3.org/TR/sparql11-overview/
[16] http://ldf.fi
[17] https://github.com/SemanticComputing/angular-semantic-faceted-search

```
    ?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: `xsd`, `skos`, `skosxl`, `foaf`, `crm`, `gvp`, and `schema` refer to well-known namespaces on the Web. `rel` contains, e.g., the schema of the application, and `nbf` is the namespace of BiographySampo. 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 knowledge graph. 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 form of created relation instances can be seen in the CONSTRUCT template of the above query: the class `Relation` has the following properties: type of the relation (`relationType`), the person of the relation (`personSubject`), the place of the relation (`placeObject`), the date of event (`date`), link to the underlying event (`source`), name of the underlying event source (`sourceName`), the explanation of the relation (`prefLabel`). An example of a connection instance telling that "Elin Danielson-Gambogi got the Florence City Art Award in 1899" is presented below as an example. Here the connection type is "person X received a honour related to place P".

```
a                  rel:Relation ;
rel:relationType   rel:honourAtPlace ;  # Connection type
rel:personSubject  nbf:p2264 ;          # Elin Danielson-Gambogi
rel:placeObject    rel:p5133 ;          # Florence
rel:date           "1899-01-01"^^xsd:date ;  # Date of the underlying event
rel:source         nbf:event28034 ;          # Event in BiographySampo
rel:sourceName     "Tapahtuma Semanttisessa kansallisbiografiassa" ;
skos:prefLabel     "Elin Danielson-Gambogi on vastaanottanut
                    kunnianosoituksen joka liittyy paikkaan Firenze:
                    'Firenzen kaupungin taidepalkinto 1899'." .
```

Table 1 contains related connection types "Painting depicts a place" and "Novel depicts a place". These connection types can be seen to represent a more general connection "Artwork depicts a place". Instances of both of these connection types could be created with a single SPARQL query corresponding to a more general artwork rule, but the resulting query would be more complex. We chose these connection types as a case study because these relationships were deemed interesting for the BiographySampo portal, and enough data was available in the material we had access to.

Type of Connection	# 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	2528
Died in	7349
Painting depicts a place	1091
Novel depicts a place	290
Born in	7182
Career is related to a place	20536
In total	40901

Table 1. Connection types and instance counts

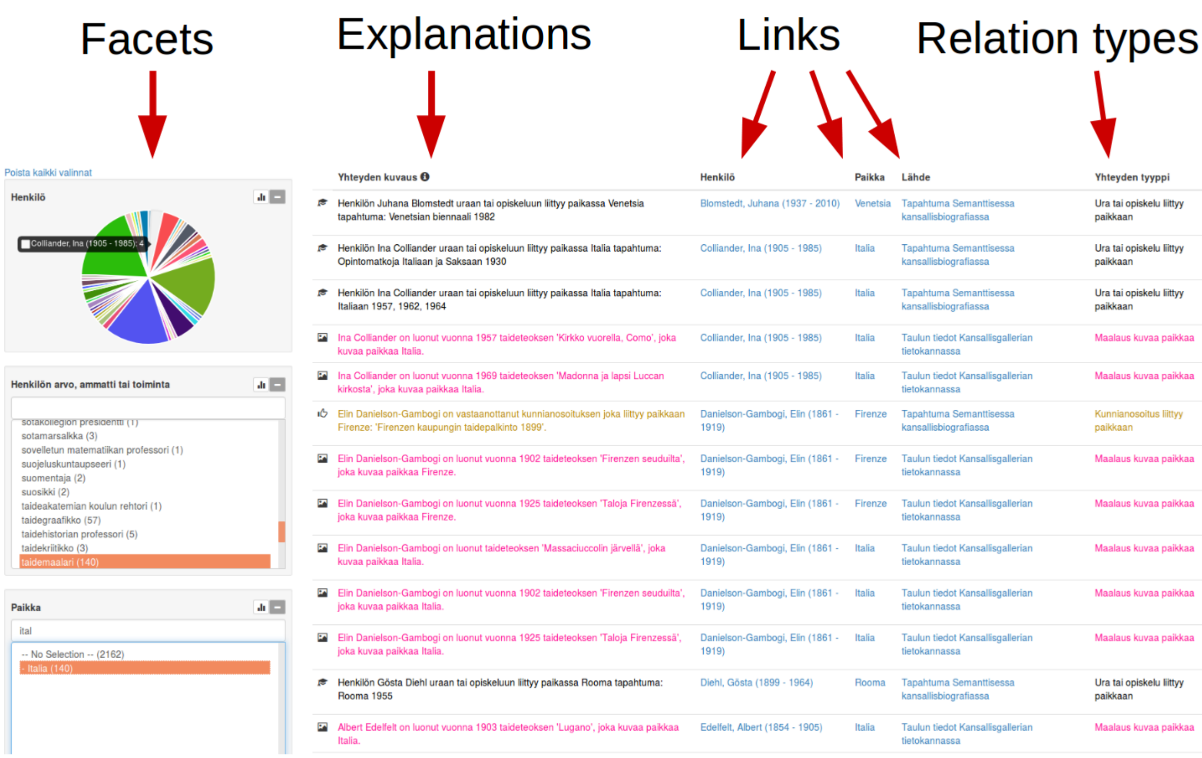

Fig. 1. View of the user interface

3.3 Data Service

The data underlying BiographySampo has been published on the Linked Data Finland platform[18] [20] according to the Linked Data publishing principles and other best practices of W3C [14], including, e.g., content negotiation

[18] https://ldf.fi

and provision of a SPARQL endpoint[19]. Most of the data of BiographySampo is open, but some parts, including the additional graphs of Faceted Relator are not publicly available due to certain copyright and data privacy issues. The portal, including the Faceter tool and various other JavaScript libraries used in its implementation, is based from a data perspective solely on querying the SPARQL endpoint from the client side using JavaScript. The portal is a demonstrator of the idea that versatile web applications can be implemented by separating the application logic and data services via SPARQL API, which facilitates developing new applications efficiently by re-using the same data.

In addition to the ready-to-use data application perspectives in the BiographySampo semantic portal, the underlying SPARQL endpoint can and is being applied to custom data analyses in Digital Humanities research using YASGUI[20] [36] and Python scripting in Google Colab[21] and Jupyter[22] notebooks. In our work, the ”FAIR guiding principles for scientific data management and stewardship” of publishing Findable, Accessible, Interoperable, and Re-usable data are used[23].

4 DEMONSTRATOR AT WORK

The section presents an implementation of the method, Faceted Relator, as part of the BiographySampo portal. The portal publishes biographical linked data of over 13 100 prominent Finnish people with seamlessly integrated Digital Humanities tools for biographical and prosopographical research, and has had some 40 000 users. In the following we first describe the ”Sampo” model on which BIographySampo and Faceted Relator are based. The tool is available online published as a separate application perspective of the BiographySampo portal[24].

4.1 Sampo Model

The Sampo model formulates the idea of aggregating and publishing distributed, heterogeneous local data sources in a global linked data service. In this way the data of all data providers can be enriched with each other's content, by reasoning based on Semantic Web standards, and the global data can be used easily across original local data silo boundaries. This arguably creates a sustainable ”business model” where every data provider wins through collaboration, and of course the end users, in particular, win by better access to more data in the service.

Data alignment and linking in this approach is based on a shared global data model, and a set of shared domain ontologies (places, actors, etc.) that are used for describing the contents of the different data sources for semantic interoperability. In our use case, for example, the basic information about letters is fairly easy to model with a few metadata elements, such as author, sending/receiving places/times, subject matter, etc. The Getty vocabularies[25] TGN and ULAN can, for example, be used for harmonizing metadata field values, such as the sending place and sender of the letter. There are models for representing prosopografical information about historical people [42], and for representing places [39]. The problems for representing historical data like this, as well as aligning literal values in different languages used in the databases onto shared ontological structures for data harmonization are by no means solved yet in a global setting, but the technology stack of the Semantic Web provides us a promising standardized way to proceed [18].

The Sampo model also includes a kind of standardized model of semantic portals on top of a linked data service with the following characteristics. Firstly, the portal can provide the user with multiple “application perspectives”

[19]The homepage of the data service including, e.g., documentation of the data and pointers for linked data browsing and the SPARQL endpoint, is available at: https://www.ldf.fi/dataset/nbf
[20]https://yasgui.triply.cc
[21]https://colab.research.google.com/notebooks/intro.ipynb
[22]https://jupyter.org
[23]https://www.go-fair.org/fair-principles/
[24]http://biografiasampo.fi/yhteyshaku/
[25]https://www.getty.edu/research/tools/vocabularies/

for searching and exploring the underlying knowledge graph from different points of view. Secondly, each perspective provides the end user with a semantic faceted search engine [41], where the search results can be filtered and found flexibly by making selections using a set of orthogonal facets (e.g., letters, persons, places, times, etc.). Thirdly, after filtering down a target set of entities of interest in a perspective (say people, letters, or places), the result set can be studied and visualized using a variety of ready-to-use data-analytic tools. In our case, for example, map-based visualizations, timelines, and statistics are available. Fourthly, the data is provided back for the community to use as a linked data service, SPARQL endpoint, and application programming interfaces [14] based on W3C standards and best practices[26].

These ideas behind the Sampo model have been explored and developed before in different contexts. For example, the notion of collaborative content creation by data linking is a fundamental idea behind the Linked Open Data Cloud movement[27] and has been developed also in various other settings, e.g., in ResearchSpace[28]. The idea of providing multiple analyses and visualizations to a set of filtered search results has been used in other portals, such as the ePistolarium[29] [35] for epistolary data, and using multiple perspectives have been studied as an approach in decision making [30]. Faceted search [13, 41], also know as ”view-based search” and ”dynamic ontologies”, is a well-known paradigm for explorative search and browsing [33] in computer science and information retrieval, based on S. R. Ranaganathan’s original ideas of faceted classification in Libary Science. The two step usage model is used in prosopographical research [43] (without the faceted search component). The novelty of the Sampo model lies in combining several ideas and operationalizing them for developing applications in Digital Humanities.

The Sampo model outlined above has been developed and tested in several cultural heritage applications (2004–2020) that have had millions of users on the Semantic Web[30].

4.2 User Interface

Fig. 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.

In this case study, Faceted Relator 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 system shows a hit list of the relation instances that fit the selected filtering criteria in the facets. The user can limit the search at any time with a selection on any facet. Furthermore, the fact that the facets are hierarchical 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 it 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, name of the data source, and finally the connection type (cf. Table 1), based on the corresponding connection instance. Persons, places, and data sources are represented as links to 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 in order to give the user a visual overview of different kind of relations found. At any

[26]https://www.w3.org/standards/semanticweb/
[27]https://lod-cloud.net
[28]https://www.researchspace.org
[29]http://ckcc.huygens.knaw.nl
[30]See https://seco.cs.aalto.fi/applications/sampo/ for a list of Sampo portals and further information.

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”[31].

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 Faceted Relator 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 Faceted Relator for solving some quantitative research problems.

For example, Fig. 2 illustrates how the question ”Who has got most awards in 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 the demonstrator is limited by the sources and data it uses. A relation can be missing for a number of reasons and relative numbers may not therefore reflect reality perfectly. However, the tool can be valuable for finding out serendipitous phenomena in the data for further close reading by the human expert.

5 DISCUSSION: EVALUATION AND LESSONS LEARNED

In our work on implementing the demonstrator, creating the rules for extracting the relations was relatively easy in practice. A more difficult problem was linking the entity mentions in the data to the underlying ontologies. The rules (queries) for data transformation could be run in a few seconds. Of course, the size and data models used in other applications would lead to different times. We also tested whether creating more general rules would ease the work of designing the rules as less rules are then needed. It was indeed possible to create more general queries, but at least in our Finnish cultural heritage data, different kind of data models have been used, and it was easier to create several specialised queries for each data model than a more general rule for all of them. Furthermore, the execution times of the complicated general rules got longer.

5.1 Evaluating the System

To evaluate the quality of the relations and explanations given by the system we evaluated the results received with a small number of searches. We made a search to find relations starting from five different people and places.

The people selected were: (1) Elias Lönnrot (1802–1884), the creator of the Finnish national epic Kalevala, (2) Johan Ludvig Runeberg (1804–1877), the Finnish national poet, (3) Akseli Gallen-Kallela (1865–1931), one of the most prominent Finnish classical painters, (4) Ellen Thesleff (1869–1954), a female Finnish expressionist, and (5) Urho Kekkonen (1900–1986), the longest serving president of the Finland. After selecting a person in the person facet, Faceted Relator determined the related connections that were analysed manually.

[31]These explanations are in Finnish and are translated here in English for illustration.

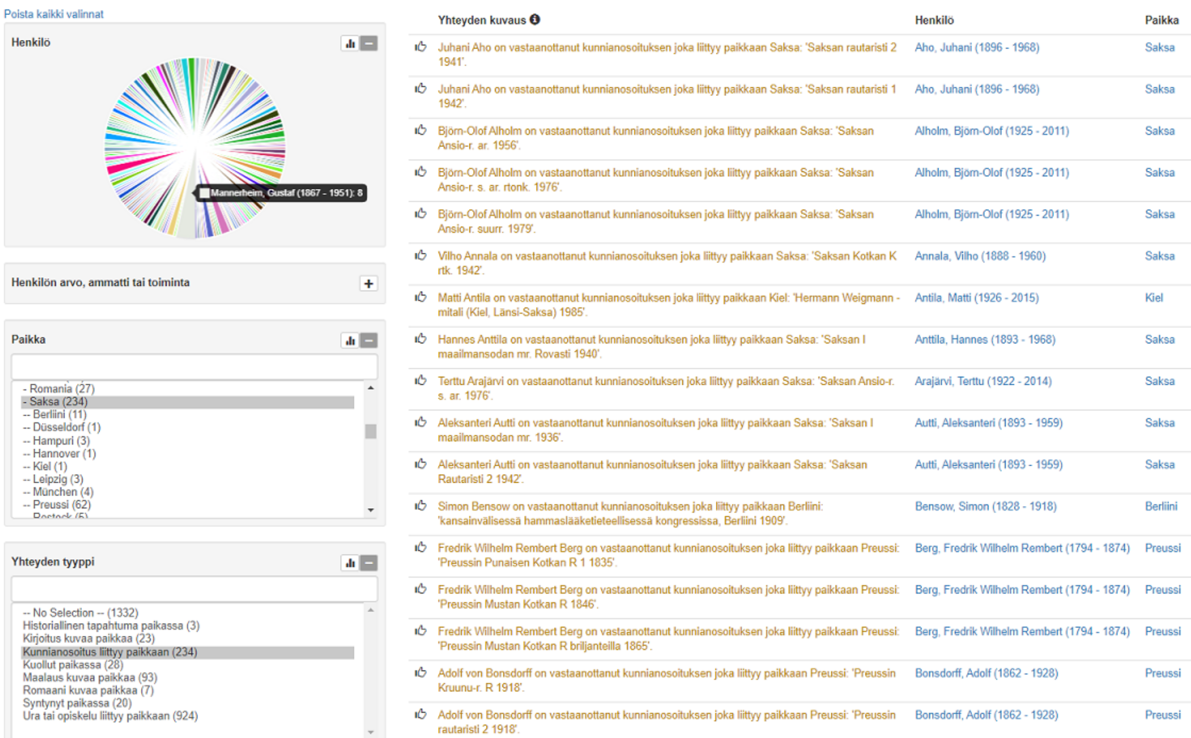

Fig. 2. Solving the problem: who has got most awards in Germany?

These people were selected for their significance to the Finnish history representing different fields and times. They do not represent the average people in the data but were expected to have many relations of diiferent kinds to places for evaluation in the data. The searches with the selected people yielded from 18 to 44 relations to places for each.

(1) Elias Lönnrot has 42 relations, including, for example, some places related to his career as a doctor of medicine in Oulu and Kajaani. He has received the Prussian Pour le Merite award. There are many letters, including some that he has sent from Estonia, relating him to places where he wrote or received letters. He is also the author of a few books concerning certain Finnish places. All the relations found seem to represent Lönnrot's life quite well, and the natural language explanations were good, too. A few of his books are mentioned multiple times, because they were published in multiple languages.
(2) Johan Ludvig Runeberg has 18 relations of multiple types. For example he has received Danish and Swedish honorary medals. Most of the explanation seem good, but one relation concerning his career is perhaps misleading. According to the system, Runeberg's career is related to Greece, because he was a teacher of classical Greek language at one point. This isn't entirely wrong, but might be seen as misleading.
(3) Akseli Gallen-Kallela has 44 relations. These concern only his birth, death and, his paintings, and certain books about his painting. Interestingly twenty of these are related to Africa, far more than, for example, to Paris. This may overstate the meaning of Africa for Gallen-Kallela's life, but it does reflect the fact that did spend almost two years in Africa. This can also be surprising information to someone with only passing information about Gallen-Kallela and inspire the user to learn more about him. It is notable that the system

doesn't show any career events for Gallen-Kallela. This seems to be because the biography of Gallen-Kallela is structured in such a way that no career events were picked to the Biographysampo knowledge graph, and therefore no relations can be generated based on them.

(4) Ellen Thesleff has 40 relations. These only concern her birth, death, and her paintings. Many of the painting are related to Italy, which does reflect the importance of Italy for her work. It is notable that both Thesleff and Gallen-Kallela lack relations concerning their career and awards in the system. Both of their biographies certainly include many interesting career events and awards that could, and ideally should, be included. These are lacking because the these events are collected to the Biographysampo knowledge graph from certain sections of the biographies, that are lacking with Gallen-Kallela and Thesleff.

(5) Urho Kekkonen has 29 relations. These include some historical events, incuding his presidential election, and notably many honours he has received in the form of honorary doctorates from around Finland and the World. Most of these have good natural language explanations, but few have somewhat mysterious looking explanations like ”Urho Kekkonen received an honor related to Varsova: Varsova 1964”. This refects the fact that Kekkonen received an honorary doctorate from Varsova in 1964, and therefore it is not wrong but the explanation is not good. This happens because in the biography certain types of honorary doctorates are given as a list. It might be possible to eliminate these kind of vague explanations when creating the relation entities, for example by automatically excluding all awards with too short explanations. However then these potentially interesting connections would not be shown. Notably the birth place of Kekkonen is missing. This is likely due to an omission in the mapping of place ontologies.

We also searched connections by starting from five places. The selected places are (1) Utsjoki, the most northernmost town of Finland to represent a smaller place, (2) Helsinki and (3) Turku, the two most important cities of the Finnish history, and (4) London and (5) Paris, representing important cities outside Finland.

After selecting a place in the place facet, Faceted Relator determined the related connections that were analysed manually. Searching for people related to a certain place the user should first select a place from the place facet.

(1) Utsjoki has only 8 relations, so there is no need for narrowing the search any more as all the explanations can be easily read. There is a variety of different facts and this could well be used to find out about the local history of the town.

(2) There are more than 8000 connections for Helsinki, more than to any other place, which can be expected for the capital of Finland. When the user selects Helsinki, he is shown all the connections as a list ordered by name of the person in the connection. The number of connections is too large to go through and read them all. The lack of prioritization means that the user may not find interesting connections by just looking at the results. Here the user needs to explore the facets and narrow the search further to find interesting individual connections. Here the system is working as planned and invites the user to explore the data interactively, but some users might want a ranked selection of connections, so that they would be immediately offered most interesting results on top of the search results. To limit the results the user could further narrow the search to, for example, people of certain profession. However even without narrowing the search further, the user could compare the relative numbers of relations using the pie chart option and see that Helsinki has a relatively large number of connections to members of the Parliament and authors.

(3) Turku has over 3000 connections, and there is a need to narrow the search further as in the case of Helsinki. An interesting result are the relative numbers of the connections on the facets that can be visualized with a pie chart. Especially interesting might be a comparison with Helsinki. For example, Turku has a relatively larger number of connections to priests. The pie chart shows that the profession to which both Helsinki and Turku have most connections is Member of the Parliament.[32]

[32] A single person may have several connections to a place all of which are summed up here.

(4) London has 171 connections, and this might also be a too large number to go through and might require further narrowing the search to. For example, to find out how authors are related to London, an additional selection on the occupation/profession facet is needed. This would reveal, among other things, that the author Aale Tynni won a gold medal in poetry in the London Olympic Games of 1948. This is an example of serendipitous piece of information to those who do not know that poetry used be a competition in the Olympic Games.
(5) Paris has 446 connections in the system. Again there are a lot of connections, but the relative numbers might be interesting even without further narrowing of the search. These can be compared to other places such as London. The user could, for example, compare the profession distribution, and find out that Paris has more connections to painters than London. Also the fact that Paris has more connections altogether can be interesting. It hints that Paris has been more culturally significant for Finland than London.

The informal initial evaluation and testing of the demonstrator above, as well as some additional tests, showed that the method and system works as well in terms of precision. 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 golden standard available, and failing to find a connection may be due to sparsity of the data, not the method. In any case, as the Table 1 shows, the system was able to find lots interesting relations in the data and the approach looks promising. Theoretically it seems likely that this kind of approach will miss some truly serendipitous connections that represent some type of relation that could not be even thought of. This is because the nature of the method requires limiting the search to predetermined types of connections. It could be argued that this method gives preference to precision over recall.

According to [6], a system can be considered creative if it is able to create “new”, “surpising”, and “valuable” ideas. At least from a layman perspective, this seems to be the case in Faceted Relator although measuring creativity is not easy. Given the large, semantically rich knowledge graph 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.

5.2 Generalizability of the Knowledge-based Approach

In our example, we have searched for connections between prominent Finnish people and places. Generalizing the method to other application domains and datasets would in principle be straightforward, but in practice adaptation work is needed as the data about people and places in other datasets may be different, and it may also be represented using different data models and ontologies. If the data in several datasets is represented using standard data and vocabulary models, such as CIDOC CRM and SKOS, the same rules for instantiating connections can be re-used in different datasets.

More work would be needed to apply the method to different relations, such as relations between two people, pieces of art, or events. Different types of relations would require domain and dataset specific considerations. For example, when generating interesting connections between two people a connection “two prominent people were born in the same city” may generate a huge number of relations that would be both difficult to search efficiently and not generally particularly interesting if most people are born in the same few large cities. More interesting relation might be, for example, “two prominent people were born in the same small town or village”. This would keep the number of connections relatively low and the individual connections are more likely to be interesting. A weakness of knowledge-based methods like ours is the need to customize the method to fit to specific case and data—on the other hand fitting the method to particular applications is also a strength of the approach as non-interesting connections can be ruled out.

The number of connection instances needs to be limited to allow for efficient faceted search in interactive usage with a few seconds response times. A most computationally demanding part is counting of instances for all possible facet selections after each selection. According to our tests with Faceted Relator some 40 000 instances could be searched efficiently by faceted search using a database server corresponding to a normal personal computer. The needed efficiency is however dependent on the application. In our case, the application is mostly aimed for the general public and needs to work relatively fast. A system aimed for professional audience might allow for longer search times. The queries used in creating the connection instances are run in a separate preprocessing phase and do not make response times longer when using the portal but wise versa improves real time efficiency.

5.3 Lessons Learned

The knowledge acquisition task of formulating a set of useful explanation patters and graph transformation rules in the demonstrator was feasible. Furthermore, the number of connections found was not overwhelmingly large from a computational point of view, as shown in Table 1, and could be generated quickly. From a human end user perspective, the result set (40 901 connections) is still large enough to provided many non-trivial results and explanations. So, the suggested knowledge-based approach was deemed feasible at least in cases where the potentially interesting connections can be characterized logically and their number is not very large. This seems to be the case in the biographical datasets of BiographySampo.

A lesson learned from this project has been that the semantic approach is useful when searching relations between larger groups of entities, such as groups of people (e.g., painters) and larger geographical areas (e.g., Italy). This kind of relations would not be readily apparent in a search between individual entities; faceted search and ontologies allow for formulating this kind of generic search queries very naturally for, e.g., prosopographical analyses.

Faceted search can be used to narrow the search for those relations that the user would consider most interesting. The incremental nature of faceted search makes it more likely to make serendipious discoveries, but it might be useful to augment the search with some sort of ranking of relations based on their presumed interestingness. This would be especially important in cases where the number of relation instances is very high.

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

5.4 Future Work

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. Developing the ontologies, such as the ontology of professions, might also improve usability of the system.

In the future, we would like to expand the system to include new connection types, such as relations between people.

Acknowledgements Our research was supported by the Severi project[33], funded mainly by Business Finland. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

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

REFERENCES

[1] Marwan Al-Tawil, Vania Dimitrova, and Dhavalkumar Thakker. 2020. Using knowledge anchors to facilitate user exploration of data graphs. *Semantic Web* 11, 2 (2020), 205–234. https://doi.org/10.3233/SW-190347

[2] R. S. Aylett, D. S. Bental, R. Stewart, J. Forth, and G.Wiggins. 2012. Supporting Serendipitous Discovery. In *Digital Futures (Third Annual Digital Economy Conference), 23-25 October, 2012, Aberdeen, UK*. http://www.serena.ac.uk/papers/#sthash.2aHjBNNz.dpuf

[3] Christopher Baker and Kei-Hoi Cheung (Eds.). 2007. *Semantic Web—Revolutionizing Knowledge Discovery in the Life Sciences*. Springer–Verlag.

[4] Federico Bianchi, Matteo Palmonari, Marco Cremaschi, and Elisabetta Fersini. 2017. Actively Learning to Rank Semantic Associations for Personalized Contextual Exploration of Knowledge Graphs. In *The Semantic Web*, Eva Blomqvist, Diana Maynard, Aldo Gangemi, Rinke Hoekstra, Pascal Hitzler, and Olaf Hartig (Eds.). Springer–Verlag, Cham, 120–135. https://doi.org/10.1007/978-3-319-58068-5_8

[5] Giuseppe Birró. 2020. Building relatedness explanations from knowledge graphs. *Semantic Web* 10, 6 (2020), 963–990.

[6] Margaret A. Boden. 2009. Computer Models of Creativity. *AI Magazine* 30, 3 (2009), 23–34. http://www.aaai.org/ojs/index.php/aimagazine/article/view/2254

[7] R. Burke. 2000. Knowledge-based recommender systems. In *Encyclopaedia of Library and Information Sciences*, A. Kent (Ed.). Marcel Dekker.

[8] Gong Cheng, Fei Shao, and Yuzhong Qu. 2017. An Empirical Evaluation of Techniques for Ranking Semantic Associations. *IEEE Transactions on Knowledge and Data Engineering* 29, 11 (2017), 1.

[9] Gong Cheng, Yanan Zhang, and Yuzhong Qu. 2014. Explass: exploring associations between entities via top-K ontological patterns and facets. In *International Semantic Web Conference (ISWC)*. Springer–Verlag, 422–437.

[10] Filip Karlo Došilović, 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.

[11] F. K. Došilović, M. Brčić, and N. Hlupić. 2018. Explainable artificial intelligence: A survey. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. Rijeka, Croatia, 210–215.

[12] A. Gangemi and V. Presutti. 2009. Ontology Design Patterns. In *Handbook on ontologies, 2nd. ed.*, S. Staab and R. Studer (Eds.). Springer–Verlag.

[13] Marti Hearst. 2006. Design recommendations for hierarchical faceted search interfaces. In *ACM SIGIR workshop on faceted search*. ACM, 1–5.

[14] T. Heath and C. Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space (1st edition)*. Morgan & Claypool, Palo Alto, California. http://linkeddatabook.com/editions/1.0/

[15] Philipp Heim, Sebastian Hellmann, Jens Lehmann, Steffen Lohmann, and Timo Stegemann. 2009. RelFinder: Revealing Relationships in RDF Knowledge Bases. In *Proceedings of the 4th International Conference on Semantic and Digital Media Technologies (SAMT 2009)*. Springer–Verlag, 182–187. http://dx.doi.org/10.1007/978-3-642-10543-2_21

[16] Philipp Heim, Steffen Lohmann, and Timo Stegemann. 2010. Interactive Relationship Discovery via the Semantic Web. In *Proceedings of the 7th Extended Semantic Web Conference (ESWC 2010)*, Vol. 6088. Springer–Verlag, Berlin/Heidelberg, 303–317. http://dx.doi.org/10.1007/978-3-642-13486-9_21

[17] J. H. Herlocker, J. A. Konstan, and J. Riedl. 2000. Explaining Collaborative Filtering Recommendations. In *Computer Supported Cooperative Work*. ACM, 241–250.

[18] Eero Hyvönen. 2012. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. Morgan & Claypool, Palo Alto, California.

[19] Eero Hyvönen, Eetu Mäkelä, Tomi Kauppinen, Olli Alm, Jussi Kurki, Tuukka Ruotsalo, Katri Seppälä, Joeli Takala, Kimmo Puputti, Heini Kuittinen, Kim Viljanen, Jouni Tuominen, Tuomas Palonen, Matias Frosterus, Reetta Sinkkilä, Panu Paakkarinen, Joonas Laitio, and Katariina Nyberg. 2009. CultureSampo – Finnish Culture on the Semantic Web 2.0. Thematic Perspectives for the End-user. In *Museums and the Web 2009, Proceedings*. Archives and Museum Informatics, Toronto.

[20] Eero Hyvönen, Jouni Tuominen, Miika Alonen, and Eetu Mäkelä. 2014. Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. In *The Semantic Web: ESWC 2014 Satellite Events*. Springer–Verlag, 226–230.

[21] Eero Hyvönen. 2020. Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-analysis and Serendipitous Knowledge Discovery. *Semantic Web* 11, 1 (2020), 187–19.

[22] Eero Hyvönen, 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 *Proceedings of the 16th Extendwed Semantic Web Conference (ESWC 2019)*. Springer–Verlag.

[23] Eero Hyvönen and Heikki Rantala. 2019. 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* (Copenhagen, Denmark). CEUR Workshop Proceedings, Vol-2364, 230–239. http://www.ceur-ws.org/Vol-2364/

[24] Eero Hyvönen and Heikki Rantala. 2019. Relational Search in Cultural Heritage Linked Data: A Knowledge-based Approach. In *Digital Humanities 2019 Conference Papers, Book of Abstracts* (Utrecht, the Netherlands). University of Utrecht. https://dev.clariah.nl/files/

dh2019/boa/0445.html

[25] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2011. *Recommender Systems. An introduction.* Cambridge University Press, Cambridge, UK.

[26] Mikko Koho, Erkki Heino, and Eero Hyvönen. 2016. SPARQL Faceter – Client-side Faceted Search Based on SPARQL. In *Joint Proceedings of the 4th International Workshop on Linked Media and the 3rd Developers Hackshop*. CEUR Workshop Proceedings, 53–63. http://ceur-ws.org/Vol-2187/paper5.pdf

[27] Oleksandr Kolomiyets and Marie-Francine Moens. 2011. A survey on question answering technology from an information retrieval perspective. *Information Sciences* 181, 24 (2011), 5412–5434. https://doi.org/10.1016/j.ins.2011.07.047

[28] Freddy Lecue. 2020. On The Role of Knowledge Graphs in Explainable AI. *Semantic Web – Interoperability, Usability, Applicability* 11, 1 (2020), 41–51.

[29] Jens Lehmann, Jörg Schüppel, and Sören Auer. 2007. Discovering Unknown Connections—the DBpedia Relationship Finder. In *Proc. of the 1st Conference on Social Semantic Web (CSSW 2007) (LNI, Vol. 113)*. GI, 99–110. http://subs.emis.de/LNI/Proceedings/Proceedings113/gi-proc-113-010.pdf

[30] Harold A. Linstone. 1989. Multiple perspectives: Concept, applications, and user guidelines. *Systems practice* 2, 3 (1989), 307–331. https://doi.org/10.1007/BF01059977

[31] Steffen Lohmann, Philipp Heim, Timo Stegemann, and Jürgen Ziegler. 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, 421–422. http://doi.acm.org/10.1145/1719970.1720052

[32] Eetu Mäkelä, Tuukka Ruotsalo, and Hyvönen. 2012. How to deal with massively heterogeneous cultural heritage data—lessons learned in CultureSampo. *Semantic Web – Interoperability, Usability, Applicability* 3, 1 (2012), 85–109.

[33] Gary Marchionini. 2006. Exploratory Search: From Finding to Understanding. *Commun. of ACM* 49, 4 (April 2006), 41–46. https://doi.org/10.1145/1121949.1121979

[34] Stuart E. Middleton, David De Roure, and Nigel R. Shadbolt. 2009. Ontology-based Recommender Systems. In *Handbook on ontologies (2nd Edition)*, S. Staab and R. Studer (Eds.). Springer–Verlag, 779–796.

[35] Walter Ravenek, Charles van den Heuvel, and Guido Gerritsen. 2017. The ePistolarium: Origins and Techniques. In *CLARIN in the Low Countries*, Arjan van Hessen and Jan Odijk (Eds.). Ubiquity Press, 317–323. https://doi.org/10.5334/bbi

[36] Laurens Rietveld and Rinke Hoekstra. 2017. The YASGUI family of SPARQL clients. *Semantic Web – Interoperability, Usability, Applicability* 8, 3 (2017), 373–383. https://doi.org/10.3233/SW-150197

[37] A. Sheth, B. Aleman-Meza, I. B. Arpinar, C. Bertram, Y. Warke, C. Ramakrishnan, C. Halaschek, K. Anyanwu, D. Avant, F. S. Arpinar, and K. Kochut. 2005. Semantic Association Identification and Knowledge Discovery for National Security Applications. *Journal of Database Management on Database Technology* 16, 1 (2005), 33–53.

[38] Avi Silberschatz and Alexander Tuzhilin. 1995. On subjective measures on interestingness in knowledge discovery. In *Proceedings of KDD-1995*. AAAI Press, 275–281.

[39] H. Southall, R. Mostern, and M. L. Berman. 2011. On Historical Gazetteers. *International Journal of Humanities and Arts Computing* 5 (2011), 127–145.

[40] Gonzalo Tartari and Aidan Hogan. 2018. WiSP: Weighted Shortest Paths for RDF Graphs. In *Proceedings of VOILA 2018*. CEUR Workshop Proceedings, vol. 2187, 37–52.

[41] D. Tunkelang. 2009. Faceted search. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 1, 1 (2009), 1–80.

[42] Jouni Tuominen, Eero Hyvönen, and Petri Leskinen. 2018. Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research. In *BD2017 Biographical Data in a Digital World 2017, Proceedings* (Linz, Austria). CEUR Workshop Proceedings, 59–66. http://ceur-ws.org/Vol-2119/paper10.pdf.

[43] Koenraad Verboven, Myriam Carlier, and Jan Dumolyn. 2007. A short manual to the art of prosopography. In *Prosopography approaches and applications. A handbook*. Unit for Prosopographical Research (Linacre College), 35–70.

[44] Agnieszka Ławrynowicz. 2020. Creative AI: a New Avenue for Semantic Web? *Semantic Web – Interoperability, Usability, Applicability* 11, 1 (2020), 69–78.