Serendipitous knowledge discovery on the Web of Wisdom based on searching and explaining interesting relations in knowledge graphs

Eero Hyvönen[a,*,1]

[a]Aalto University, Department of Computer Science, and University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG), P. O. Box 15400, Espoo, 00076 AALTO, Finland

ARTICLE INFO

Keywords:
knowledge graphs
relational search
knowledge discovery
information retrieval
large language models
generative AI

ABSTRACT

This paper maintains that the Semantic Web is changing into a kind of Web of Wisdom (WoW) where AI-based problem solving, based on symbolic search and sub-symbolic methods, and Information Retrieval (IR) merge: IR is seen as a process for solving information-related problems of the end user with explanations, a form of knowledge discovery. As a case of example, relational search is concerned, i.e., solving problems of the type "How are $X_1 \dots X_n$ related to $Y_1 \dots Y_m$?". For example: how is *Pablo Picasso* related to *Barcelona*? The idea is to find explainable “interesting” or even serendipitous associations in Knowledge Graphs (KG) and textual web contents. It is argued that domain knowledge-based symbolic methods based of KGs are needed to complement domain-agnostic graph-based methods and Generative AI (GenAI) boosted by Large Language Models (LLM). By using domain specific knowledge, it is possible to find and explain meaningful reliable textual answers, answer quantitative questions, and use data analyses and visualizations for explaining and studying the relations.

1. Extending Search to Explainable Knowledge Discovery

Figure 1: Data-Information-Knowledge-Wisdom (DIKW) hierarchy of data science

According to the Data-Information-Knowledge-Wisdom (DIKW) hierarchy of data science [32], new value is created as ’data’ (know just the data, nothing about it) changes into ’information’ (know what the data is), then into ’knowledge’ (know how the information is used), and finally into ’wisdom’ (know why; explaining knowledge) (Cf. Fig. 1). This transition is happening on the Web that is gradually changing from a data/information publishing platform into a knowledge base, and finally into an intelligent question answering system [8], the Web of Wisdom (WoW).

In this paper one approach towards the WoW is considered: *relational search* (RS) where the notion of search is extended to finding “interesting” [35] or even serendipitous connections between the resources in a knowledge graph (KG) [21], such as persons, places, and events. For example: “How are German novelists of the 19th century related to France?” Such semantic connections can be based on various criteria: German people (or their family members) ware born or died in Paris, French topics were discussed in their novels, they wrote a novel or an article in French, their publisher was a French company, their portraits are in Louvre, they got a medal of honor in Lyon, and so on. In relational search, the document and entity search paradigms [4] are extended to finding explainable interesting semantic connections between entities.

This kind of functionality can be seen as a form of Knowledge Discovery (KD) [15] arguably with a touch of serendipity. Serendipity means ’happy accident’ or ’pleasant surprise’, even ’fortunate mistake’ in KD [3]. A paradigmatic example of serendipitous KD is discovery of penicillin in 1928, when Dr Alexander Fleming returned from a holiday to his laboratory to find mold growing on a Petri dish of Staphylococcus bacteria. He was surprised by the fact that mold seemed to be preventing the bacteria around it from growing. Serendipitous knowledge discovery is one of the grand promises and challenges of the Semantic Web. When linking local disjoint datasets based on rigorous semantics and aligned identifiers, richer and in many cases unexpected patterns of knowledge are likely to pop up.

In the following, the idea and previous research on what can be called *domain agnostic* relational search are first overviewed. After this a *knowledge-based approach* to RS is discussed, where the notion of heuristic search and problem

*Corresponding author
eero.hyvonen@aalto.fi (E. Hyvönen)
https://seco.cs.aalto.fi/u/eahyvone/ (E. Hyvönen)
ORCID(s): 0000-0003-1695-5840 (E. Hyvönen)
[1]The author is professor of computer science at the Aalto University and director of Helsinki Centre for Digital Humanities (HELDIG) at the University of Helsinki.

solving in AI is combined with search in Information Retrieval (IR). This approach is compared with domain agnostic approaches and using Generative AI (GenAI) and Large Language Models (LLM) for question answering, i.e., a kind of *linguistic relational search*. The comparison suggests that the symbolic and sub-symbolic approaches are complementary, leading to hybrid neuro-symbolic approaches [14, 6] in the future.

2. Approaches to Relational Search

In relational search [9] based on a KG, the question or query contains two or more entities whose relations are searched for, and the task is to find explainable semantic relations (query results) between the entities for the end user. Relational search involves two major challenges: 1) *Finding and filtering out interesting connections from not interesting ones*. For example, the fact that two entities are connected as instances of a class Person is not interesting, but if the entities went to the same school at the same time, the relation is likely to be of interest. The key challenge here is how to formulate the notion of "interesting" for the computers. 2) *Explaining the relations*. That two (or more) entities are connected is not interesting without an explanation on how or why they are related. This is a challenge of explainable AI [2] or even creative AI [41].

Finding explanations for relations between entities is important not only for knowledge discovery but also, for example, in recommender systems and semantic browsing [27, 20]. Finding serendipitous relations makes it also possible to avoid over-personalization in recommender systems [25]. The semantic richness and number of possible connections in KGs suggest that relational search can be seen as an instance of computational creativity [7], an example of the subtype "exploratory creativity", where creativity refers to search within a predefined search space under given constraints for the solutions.

2.1. Domain Agnostic Relational Search

In [34] the idea of searching relations is applied to association finding in national security domain. Within the CH domain, CultureSampo[2] [16, 26] 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 ULAN[3] knowledge graph of historical persons. In RelFinder[4] [24, 13, 12], based on the earlier "DBpedia Relationship Finder" [22], 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 researchers, e.g., worked at the Princeton University. In WiSP [37], 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. WoolNet [38] is a recent system with an online ptototype[6] along the same line of research. Some applications, such as RelFinder and Explass [10], 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 most of which 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.

These approaches can be characterized as *Domain Agnostic Relational Search* as they are mostly based on generic graph algorithms. A benefit of this is that the same methods can be re-used in different application domains. However, this position paper argues that generic criteria are not enough to capture the semantics "interestingness" or "serendipity" but in many cases domain knowledge is needed, too.

As usual in AI, knowledge can be incorporated in systems in two major ways: 1) using symbolic knowledge representations and reasoning or 2) by sub-symbolic methods based typically on machine learning and neural networks. Also neuro-symbolic approaches combining benefits and mitigating challenges of these approaches can be used [14, 6]. To support and evaluate the argument for the need of knowledge-based relational search, some experiments are discussed next in the application domain of Cultural Heritage and Digital Humanities [11]. However, the ideas discussed are more general and methods of relational search are applicable to other domains, too (cf., e.g., [34]).

2.2. Knowledge-based Relational Search

In symbolic AI, (heuristic) search [28] is based on knowledge-based strategies and is used for intelligent problem solving and (logical) reasoning, while in IR search means finding objects based on indexing them in databases, knowledge graphs, and on the Web. In [18] and [30], the notion of *Knowledge-based Relational Search* was introduced combining the two notions of search. The idea here is to first search in the KG in the AI sense for interesting relations, i.e., paths between entities, using knowledge-based heuristics that give constraints for interestingness. For example, based on a genealogical family tree KG and knowledge, it is possible to see that two persons are related as cousins, if their parents are siblings, i.e., have the same parents.

This phase could be done dynamically as graph search like in the knowledge agnostic approaches, but there is also the option of doing this in advance in a pre-processing phase. In this case the original KG is transformed, using, e.g.,

[2] CultureSampo : http://www.kulttuurisampo.fi

[3] Getty ULAN KG: http://www.getty.edu/research/tools/vocabularies/ulan/

[4] RelFinder: http://www.visualdataweb.org/relfinder.php

[5] Wikidata KG: http://wikidata.org

[6] WoolNet system online: https://woolnet.dcc.uchile.cl/

SPARQL CONSTRUCT queries for finding graph paths, into another KG that contains instances of semantically interesting relations of different types with natural language explanations attached. The properties of relations include references to related entities and the explanation. This idea makes it possible to use search methods in the IR sense for finding explained relations [18, 30]: the new KG can now 1) be queried flexibly by using semantic faceted search [39] and 2) the results be visualized and analyzed using seamlessly integrated tools as suggested in the Sampo model [17] and Sampo-UI framework [19, 29]. On the user interface, a query is formulated by 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.

In this approach "search" in the sense of IR is applied for finding the results of the classical AI "search". Knowledge discovery can be facilitated in two ways: 1) By finding explained connections between entities and 2) through data-analyses and visualizations of faceted search results. The argued benefits of this approach are: 1) Uninteresting relations between the query resources can be ruled out effectively by the knowledge-based constraints, and 2) the explanations for the relations can be created either in natural language or by using data-analyses and visualizations, such as networks connecting the entities.

However, there are two major challenges involved in this approach: Firstly, there is the knowledge acquisition problem of crafting the transformation rules and their explanation patterns manually, based on application domain knowledge, as customary in knowledge-based system. Secondly, the pre-compilation phase can result in exponential explosion in the number of relations generated. However, if the relations to be created are really interesting their number is probably not overwhelming.

2.3. Linguistics-based Relational Search

An approach to solve the knowledge acquisition problem is to use available linguistic texts and machine learning as a basis for providing explanations. This leads to *Linguistics-based Relational Search* approaches to solving relational search problems. A linguistic approach to explaining relationships between KG entities is presented in [40], based on mentions and links in textual documents. Here sentences about the related entities are first retrieved from a corpus (Wikipedia) using methods of sentence search. The sentences are then modified for better readability, and finally ranked on how well they describe the relationship embedded in them using machine learning. For example, when looking for an explanation on how Adolf von Becker is related to the Swiss village Vevey, the sentence "He died while vacationing in Vevey, Switzerland, aged 78." on his Wikipedia page provides a readable explanation for his relation to Vevey, after replacing 'he' with 'Adolf von Becker' in the sentence. This approach of building explanations from text corpora bears similarity with the idea of using LLMs for question answering: why not try to solve relational search problems by simply asking ChatGPT-like systems to explain relationships by asking questions, such as "How is Adolf von Becker related to Vevye?". A small informal study to be discussed in the next section was conducted for comparing ChatGPT with the knowledge-based relation search approach based on KGs.

To sum up, the approaches to relation search differ in four major ways: 1) *Query formulation*. The query is typically formed by fixing two (or more) entities between which relations are searched for. However, if GenAI is used for the task, the query can be a prompt in natural language and possibly some additional contextual data if methods of Retrieval-Augmented Generation (RAG) are used [23]. 2) *The underlying KG*. For example, DBpedia is used in RelFinder while in our own work relational instance KGs are used, transformed from different kind of KGs, such as biograpghical KGs and Getty ULAN. 3) *Methods for finding connections*. The methods can be symbolic, including domain agnostic methods on generic graph structures, knowledge-based methods for filtering out uninteresting relations, or linguistic methods based on texts. 4) *Representation of the results*. The associations found can be presented using visualizations, most notably as graph paths, but also using business graphics, maps, or timelines, or by using explanations in natural languages. When using visualizations, the task of knowledge discovery requires some human interpretation of the visual results.

3. Lessons Learned from the Knowledge-based approach

The knowledge-based approach has been evaluated with a few tests.

3.1. Relational search between people and places

In this case, the KG of BiographySampo [18], extracted from 13 000 short biographies of the Finnish National Biography and enriched by several other KGs, was used with the following facets: Person, Occupation, Place, and Connection type, including 10 subtypes, such as "Historical event in a place", "painting depicting a place", and "Accolate (award) related to a place". By making selections on the facets as a query, connections between persons and places are listed as a result with natural language explanations.

Fig. 2 illustrates the idea by a BiographySampo portal[7] screenshot. The facets are on the left and the relations found are explained on the right in a table where each row corresponds to a relation. A novelty here is the possibility to make generic prosopographical questions relating groups of people to areas with possible subareas, not only between particular entities, because the facets can be based on hierarchical ontologies. For example, by selecting from the facets Occupation=Painter and Place=Italy, connections of different types between Finnish painters and Italy can be found, such as "Elin Danielson-Gambogi received in 1899 the Florence City Art Award" and "Robert Ekman created in

[7]Portal: `https://biografiasampo.fi/`; project home: `https://seco.cs.aalto.fi/projects/biografiasampo/`

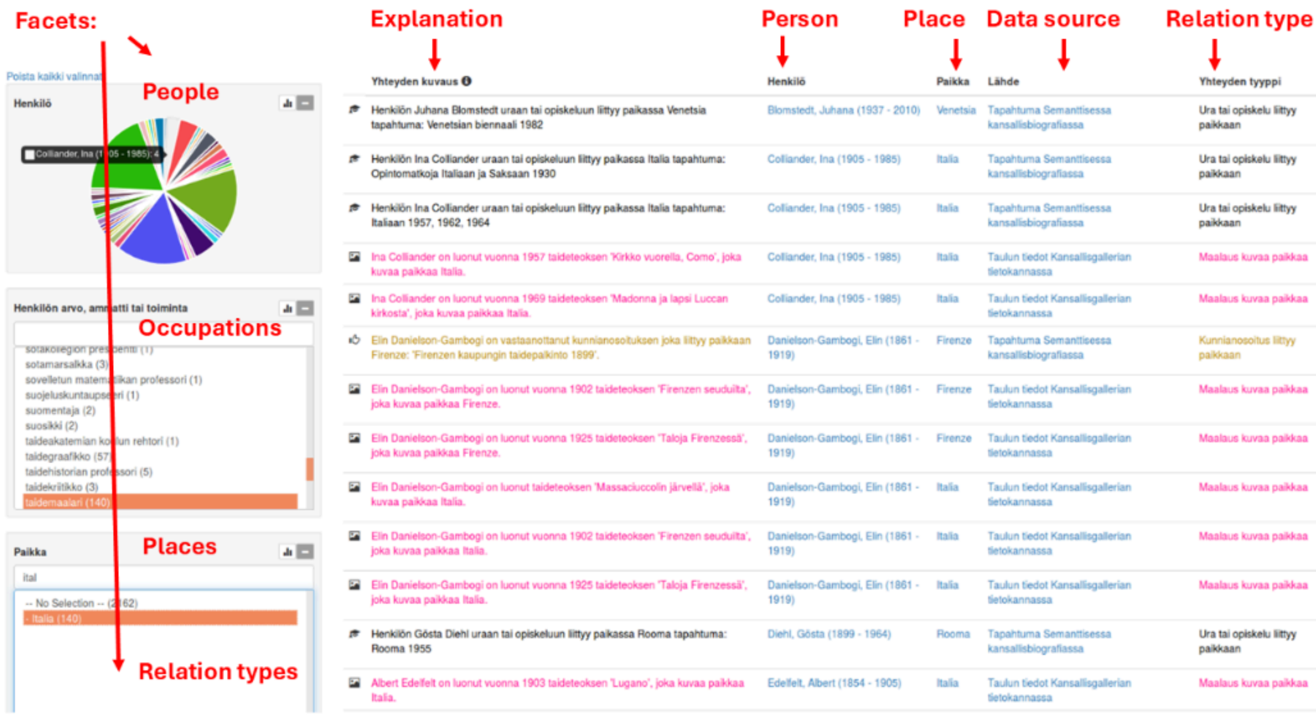

Figure 2: Relational search in BiographySampo. Search facets are on the left with the Person facet visualized as a pie chart. Explained results are listed on the right as a table.

1844 the painting 'Landscape in Subiaco' depicting a place in Italy" (in Finnish).

Another novelty of the approach is the ability to answer quantitative questions based on the hit counts of facet categories—we realized this possibility serendipitously only after the system was created. For example, the question "Who has got most awards in Germany" can be solved by selecting Connection type="Received an award in a place" on the connection type facet, and Place="Germany" that includes German cities and other places as subtypes. The hit count 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 given in Germany to Finns.

Using domain knowledge, the problem of combinatorial explosion of uninteresting relations can be addressed. For example, that two persons were born or lived at the same place during the same time can be interesting, but the place size matters. If the place is a country or a large city, lots of potential relations that are probably not interesting would be found. It seems that in this kind of semantically subtle situations domain-specific knowledge is necessary for constraining the search.

3.2. Relational search between people

In our next experiment, the focus was on searching relations between persons [31]. Here two facets for persons A and B were needed. A challenge here is to distinguish between the two persons and how to deal with directed, inverse, and symmetrical relations between them. For example, if Person A has been a teacher of Person B, then the relation that Person B was a student of A is redundant and not of interest. In addition, the non-directed relation that two persons A and B had the same teacher is of interest in, say art history, but if a teacher had n students, then listing all $n * (n-1)/2$ student pairs with the shared teacher leads to combinatorial explosion of relations. This approach was tested by using the Getty ULAN KG and the InTaVia KG [33] where national biographical KGs from the Netherlands, Austria, Slovenia, and Finland have been harmonized and aggregated into one. Again, knowledge-based tuning seems necessary to handle situations like this.

3.3. Relational search based on linked sentences

When constructing the relational KG it is possible to use different methods and KGs in extracting and aggregating the relations, as far as the same data model for representing relations is used. One possibility, suggested by the linguistic relational search approach, is to regard HTML links found in web publications as relations of interest, and the sentence including the link as the explanation. For example, in BiograpySampo the original biographies contained HTML links between biographees made by the human editors. These links could be used for creating ego-centric networks of the mutually linked biohgraphees in a data-driven fashion, with a potential for serendipitous knowledge discovery [36]. For example, *Mr Tapio Rautavaara*, a Finnish javelin thrower, actor, and singer, and *Mrs Aale Tynni*, a Finnish

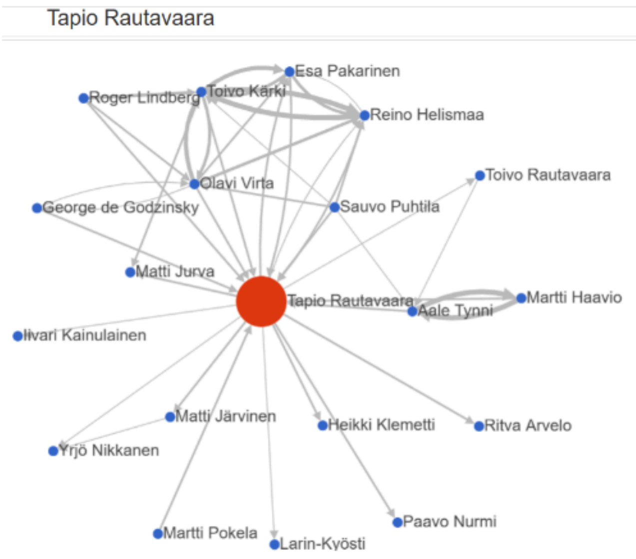

Figure 3: Ego-centric network of Tapio Rautavaara, a sportsman and popular singer, showing relations between persons in BiographySampo, including a serendipitous relation to Aale Tynni, an academician and author

academician and author, were connected in the eco-centric network of Tapio Rautavaara illustrated in Fig. 3, which was considered a potential bug in the data – or something serendipitous. The explanation given by the sentence with the connecting HTML link was that the both biographees got a gold medal in the Olympic Games in London 1948, Rautavaara in javelin and Tynni in poetry that was then a brand of Olympic sports. In the same vein, a connection in the egocentric networks between the painter Albert Edelfelt (1854–1905) and the Swedish Queen Bianca of the 14th century was explained by a painting of Edelfelt depicting the queen.

In [31] this idea was tested further by data extracted from the 14 407 English Wikipedia pages of the biographees in the InTaVia KG. Here some 180 000 sentences referring to 37 500 Wikipedia pages were found in the English Wikipedia/Wikidata. By extracting relations for different kind of HTML links, e.g., from person pages to place pages, between person pages, and from place pages to persons, new kind of relations could be added into a relational KG and be searched for with explanations given by their embedding sentences. In addition, data analyses about the link structures and network analyses with visualizations could be created for knowledge discovery.

3.4. Relational search based on LLMs

LLMs provide a new, alternative sub-symbolic approach to solving relational search problems worth studying further. In our case, initial tests have been made by simply asking ChatGPT 4.0 the same kind of questions as in above, e.g., how people and places are related. Although further and more formal research is needed, the following observations can be made. The answers of ChatGPT look very competent and interesting especially when regarding internationally well-known entities. However, when asking, for example, how Tapio Rautavaara and Aale Tynni, not so well-known, are related (cf. above), hallucinations start to emerge. In this case ChatGPT started to tell fully non-sense details about their marriage in 1954 that never even happened. The explanations provided by LLMs are qualitative texts without the possibilities of analyzing the underlying data quantitatively or visualizing the results. For example, asking or analyzing the awards given to Finns and related to Germany (cf. above) are impossible to ChatGTP to answer. Another issue in using LLMs is that the same prompt may result in different answers at different times. When using KGs, factual explanations based on symbolic structures can be given. Relations found by LLMs can be right but include errors, too. LLMs could be in principle be used to validate and explain their own answers, but this is of course tricky.

Regarding hallucinations and wrong answers, in relational search precision and reliability are more important than recall giving some advantage to symbolic methods. On the other hand, when exploring the space of possible connections, LLMs can be used in flexible ways and without semantically richly connected KGs and additional search interfaces available. However, scrutiny in fact-checking is needed. Interestingly, the latest version of ChatGPT released in autumn 2024, includes a new option for Web search[8]. This provides a novel approach combining ideas of question answering and searching web documents in the IR sense. The answers include links to web documents as a way for fact-checking and as references for finding explanations.

4. Conclusions

This paper maintained that the Web in changing into a Web of Wisdom addressing the information needs of end users directly by intelligent problem solving and question answering, based on a merger of symbolic and sub-symbolic methods of AI. This is a new possibility that the Semantic Web [5], "a new form of Web content that is meaningful to computers" was predicted to unleash, boosted by recent developments in LLMs based on deep learning and textual data. As a case study, the idea of using relational search in knowledge discovery was considered.

Based on the reported reviews above, relational search on the Web of Wisdom requires complementary symbolic and sub-symbolic methods. Domain specific knowledge and KGs are needed to find and explain meaningful reliable textual answers, answer quantitative questions, and use data analyses and visualizations for explaining and studying the relations.

As a topic for further research, a possibility for enhancing the capabilities of LLMs would be to finetune them for particular domain areas with additional contextual information using domain specific training materials. Such training material could be generated from new texts but

[8] ChatGPT search: `https://openai.com/index/introducing-chatgpt-search`

also from structured data sources, such as KGs. However, finetuning the very large LLMs is challenging in many ways, and using prompt engineering[9] techniques such as Retrieval-Augmented Generation (RAG) would provide an easier way to get better answers. Here prompting can be enhanced by using contextual data from KGs. There is also the possibility to apply LLMs and Generative AI to creating richer structured data in KGs [1]. A key issue here is whether precision or recall is more important in the application. For example, in explorative search applications recall is often preferred to miss less right results and lower precision is tolerated, but in data analyses errors are a more serious issue. More research is needed on how to combine the strengths and mitigate weaknesses of text-based LLMs and KGs in useful ways in different applications.

Acknowledgments Thanks to Heikki Rantala and Petri Leskinen for collaborations and discussions, and the Finnish Cultural Foundation for an Eminentia Grant for reflecting the work presented.

References

[1] Ahola, A., Peura, L., Leal, R., Rantala, H., Hyvönen, E., 2024. Using generative AI and LLMs to enrich art collection metadata for searching, browsing, and studying art history in digital humanities, in: Proceedings, 2nd International Conference on Data & Digital Humanities Generative Artificial Intelligence for Text and Multimodal Data, December 12–13, 2024, University of Minho, Braga, Portugal. URL: https://seco.cs.aalto.fi/publications/2024/ahola-et-al-genai-2024.pdf. accepted, forth-coming.

[2] Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J.M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., Herrera, F., 2023. Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. Information Fusion 99, 101805. doi:https://doi.org/10.1016/j.inffus.2023.101805.

[3] Baker, C., Cheung, K.H. (Eds.), 2007. Semantic Web—Revolutionizing Knowledge Discovery in the Life Sciences, Springer–Verlag. doi:10.1007/978-0-387-48438-9.

[4] Balog, K., 2018. Entity-Oriented Search. 1st ed., Springer–Verlag. doi:10.1007/978-3-319-93935-3.

[5] Berners-Lee, T., Hendler, J., Lassila, O., 2001. The Semantic Web. Scientific American 284, 34–43. URL: https://www.scientificamerican.com/issue/sa/2001/05-01/.

[6] Bhuyan, Bikram Pratim nsd Ramdane-Cherif, A., Tomar, R., 2024. Neuro-symbolic artificial intelligence: a survey. Neural Computing and Applications 36, 12809–12844. doi:10.1007/s00521-024-09960-z.

[7] Boden, M.A., 2009. Computer models of creativity. AI Magazine 30, 23–34. URL: http://www.aaai.org/ojs/index.php/aimagazine/article/view/2254.

[8] Bouziane, A., Bouchiha, D., Doumi, N., Malki, M., 2015. Question answering systems: Survey and trends. Procedia Computer Science 73, 366–375. doi:10.1016/j.procs.2015.12.005.

[9] Cheng, G., Shao, F., Qu, Y., 2017. An empirical evaluation of techniques for ranking semantic associations. IEEE Transactions on Knowledge and Data Engineering 29, 1. doi:10.1109/TKDE.2017.2735970.

[10] Cheng, G., Zhang, Y., Qu, Y., 2014. Explass: exploring associations between entities via top-k ontological patterns and facets, in: International Semantic Web Conference (ISWC), Springer–Verlag. pp. 422–437. doi:10.1007/978-3-319-11915-1_270.

[9] Prompt engineering guide: https://www.promptingguide.ai/

[11] Gardiner, E., Musto, R.G., 2015. The Digital Humanities: A Primer for Students and Scholars. Cambridge University Press, New York. doi:10.1017/CBO9781139003865.

[12] Heim, P., Hellmann, S., Lehmann, J., Lohmann, S., 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–Verlag. pp. 182–187. URL: http://dx.doi.org/10.1007/978-3-642-10543-2_21.

[13] Heim, P., Lohmann, S., Stegemann, T., 2010. Interactive relationship discovery via the semantic web, in: Proceedings of the 7th Extended Semantic Web Conference (ESWC 2010), Springer–Verlag, Berlin/Heidelberg. pp. 303–317. URL: http://dx.doi.org/10.1007/978-3-642-13486-9_21.

[14] Hitzler, P., Sarker, M.K. (Eds.), 2022. Neuro-Symbolic Artificial Intelligence: The State of the Art. IOS Press. doi:10.3233/FAIA210348.

[15] Ho, T.B., 2016. Knowledge discovery. Knowledge Science , 70–93.

[16] Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppälä, K., Takala, J., Puputti, K., Kuittinen, H., Viljanen, K., Tuominen, J., Palonen, T., Frosterus, M., Sinkkilä, R., Paakkarinen, P., Laitio, J., Nyberg, K., 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. URL: https://seco.cs.aalto.fi/publications/2009/hyvonen-et-al-culsa-mw-2009.pdf.

[17] Hyvönen, E., 2022. Digital humanities on the Semantic Web: Sampo model and portal series. Semantic Web – Interoperability, Usability, Applicability 14, 729–744. doi:10.3233/SW-223034.

[18] Hyvönen, E., Rantala, H., 2021. Knowledge-based relational search in cultural heritage linked data. Digital Scholarship in the Humanities (DSH) 36, 155–164. doi:https://doi.org/10.1093/llc/fqab042.

[19] Ikkala, E., Hyvönen, E., Rantala, H., Koho, M., 2022. Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. Semantic Web – Interoperability, Usability, Applicability 13, 69–84. doi:10.3233/SW-210428.

[20] Jannach, D., Zanker, M., Felfernig, A., Friedrich, G., 2011. Recommender Systems. An introduction. Cambridge University Press, Cambridge, UK. doi:10.1017/CBO9780511763113.

[21] Ji, S., Pan, S., Cambria, E., Marttinen, P., Yu, P.S., 2022. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems 33, 494–514. doi:10.1109/TNNLS.2021.3070843.

[22] Lehmann, J., Schüppel, J., Auer, S., 2007. Discovering unknown connections—the DBpedia relationship finder, in: Proc. of the 1st Conference on Social Semantic Web (CSSW 2007), GI. pp. 99–110. URL: http://subs.emis.de/LNI/Proceedings/Proceedings113/gi-proc-113-010.pdf.

[23] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.t., Rocktäschel, T., Riedel, S., Kiela, D., 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks, in: Proceedings of the 34th International Conference on Neural Information Processing Systems, Curran Associates Inc.. pp. 9459–9474.

[24] Lohmann, S., Heim, P., Stegemann, T., 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. pp. 421–422. URL: http://doi.acm.org/10.1145/1719970.1720052.

[25] Maccatrozzo, V., 2012. Burst the filter bubble: Using semantic web to enable serendipity, in: The Semantic Web – ISWC 2012, Springer–Verlag. pp. 391–398. doi:10.1007/978-3-642-35173-0_28.

[26] Mäkelä, E., Ruotsalo, T., Hyvönen, E., 2012. How to deal with massively heterogeneous cultural heritage data—lessons learned in CultureSampo. Semantic Web – Interoperability, Usability, Applicability 3, 85–109. doi:10.3233/SW-2012-0049.

[27] Middleton, S.E., Roure, D.D., Shadbolt, N.R., 2009. Ontology-based recommender systems, in: Staab, S., Studer, R. (Eds.), Handbook on ontologies (2nd Edition). Springer–Verlag, pp. 779–796. doi:10.1007/978-3-540-92673-3_35.

[28] Pearl, J., 1984. Heuristics: Intelligent Search Strategies for Computer Problem Solving. Addison-Wesley.

[29] Rantala, H., Ahola, A., Ikkala, E., Hyvönen, E., 2023a. How to create easily a data analytic semantic portal on top of a SPARQL endpoint: introducing the configurable Sampo-UI framework, in: Proceedings of 8th International Workshop on the Visualization and Interaction for Ontologies and Linked Data co-located with the 22nd International Semantic Web Conference (ISWC 2023) in Athens, Greece, CEUR Workshop Proceedings, Vol. 3508. URL: https://ceur-ws.org/Vol-3508/paper3.pdf.

[30] Rantala, H., Hyvönen, E., Leskinen, P., 2023b. Finding and explaining relations in a biographical knowledge graph based on life events: Case biographysampo, in: ESWC 2023 Workshops and tutorials joint proceedings, CEUR Workshop Proceedings. URL: https://ceur-ws.org/Vol-3443/ESWC_2023_SEMMES_relations.pdf.

[31] Rantala, H., Leskinen, P., Peura, L., Hyvönen, E., 2024. Representing and searching associations in cultural heritage knowledge graphs using faceted search, in: Knowledge Graphs in the Age of Language Models and Neuro-Symbolic AI. Proceedings of the 20th International Conference on Semantic Systems, 17–19 September 2024, Amsterdam, The Netherlands, IOS Press. pp. 420–435. doi:10.3233/SSW240033.

[32] Rowley, J.E., 2007. The wisdom hierarchy: representations of the dikw hierarchy. Journal of Information Science 33, 163—180. URL: https://api.semanticscholar.org/CorpusID:17000089.

[33] Schlögl, M., Tuominen, J., Kesäniemi, J., Leskinen, P., de Boer, V., Sugimoto, G., Dokler, J., 2023. The InTaVia knowledge graph - publishing European national biographical and cultural heritage object data. URL: https://seco.cs.aalto.fi/publications/2023/schlogl-et-al-intavia-kg-2023b.pdf.

[34] Sheth, A., Aleman-Meza, B., Arpinar, I.B., Bertram, C., Warke, Y., Ramakrishnan, C., Halaschek, C., Anyanwu, K., Avant, D., Arpinar, F.S., Kochut, K., 2005. Semantic association identification and knowledge discovery for national security applications. Journal of Database Management on Database Technology 16, 33–53.

[35] Silberschatz, A., Tuzhilin, A., 1995. On subjective measures on interestingness in knowledge discovery, in: Proceedings of KDD-1995, AAAI Press. pp. 275–281.

[36] Tamper, M., Leskinen, P., Hyvönen, E., Valjus, R., Keravuori, K., 2023. Analyzing biography collection historiographically as linked data: Case National Biography of Finland. Semantic Web – Interoperability, Usability, Applicability 14, 385–419. doi:10.3233/SW-222887.

[37] Tartari, G., Hogan, A., 2018. WiSP: Weighted shortest paths for RDF graphs, in: Proceedings of VOILA 2018, CEUR Workshop Proceedings, vol. 2187. pp. 37–52. URL: https://ceur-ws.org/Vol-2187/.

[38] Torres, C., Hogan, A., 2024. WoolNet: Finding and visualising paths in knowledge graphs, in: Proceedings of the Alberto Mendelzon Workshop (AMW), Mexico City, Mexico, Sept 30 – Oct 4. URL: https://aidanhogan.com/docs/woolnet_paths_knowledge_graphs.pdf.

[39] Tunkelang, D., 2009. Faceted search. Synthesis Lectures on Information Concepts, Retrieval, and Services 1, 1–80. doi:10.1007/978-3-031-02262-30.

[40] Voskarides, N., Meij, E., Tsagkias, M., De Rijke, M., Weerkamp, W., 2015. Learning to explain entity relationships in knowledge graphs, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 564–574.

[41] Ławrynowicz, A., 2020. Creative AI: a new avenue for semantic web? Semantic Web – Interoperability, Usability, Applicability 11, 69–78. doi:10.3233/SW-190377.