

Knowledge graphs within everyone's means

Maria Angela Pellegrino

Dipartimento di Informatica, Università degli Studi di Salerno, Fisciano (SA) ITALY

Abstract

A Knowledge Graph is a useful means for empowering users in actively exploring data of interest and manage acquired knowledge. However, its most common query language, SPARQL, proves to be too complex for lay users. Thus, the proposal of tools and interfaces to unlock their potentialities while masking underlying complexity in querying mechanisms is required. My Ph.D. research is situated exactly in this context and aims to scaffold end-users in taking advantage of Knowledge Graphs.

Keywords

Semantic Web, Knowledge management, Information retrieval, Data exploitation, Query builder

1. Introduction

The term Knowledge Graph (KG) has been recently used by the Semantic Web community to refer to any graph-based knowledge representation [1]. KGs, through a Semantic Web lens, allow for agile navigation of arbitrary entities thanks to defined links [1]. Thus, they are also referred to as Linked (Open) Data (LD or LOD)). Over the past decades, hundreds of datasets have been published using the Semantic Web standards covering any topical domain [2]. The LOD Cloud [3] (a KG that collects most of the published KGs) counted 12 datasets in 2007 and currently contains 1,239 datasets. Some of these KGs are proprietary, maintained internally by companies such as Google, Microsoft, Apple; while others, like DBpedia [4] and Wikidata [5], are openly available and maintained by dedicated communities. The central idea of LD is that data publishers support applications in discovering and integrating data by complying with a set of best practices in the areas of linking, vocabulary usage, and metadata provision [2]. Because of the extensive range of heterogeneous information stored in KGs, for their easy navigation, thanks to their quantitative and qualitative properties, they could behave as a critical resource for information retrieval (IR) and knowledge management (KM).

The exploitation of KG is mainly affected by i) required technical skills in query languages (e.g., SPARQL) and in understanding the semantics of the supported operators [6], too challenging for lay users, and ii) conceptualization issues to understand how data are modelled [6, 7].

These drawbacks have led to the development of tools and interfaces to support users in interacting with KGs by implicitly composing queries while hiding the underlying complexity.


My research is situated in this context and aims to propose approaches and prototyping tools to express users' needs or explore available data by a Natural Language (NL) interface to guide end-users (with different interests, types of background, age, and needs) to query KGs

CHIItaly 2021 Joint Proceedings of Interactive Experiences and Doctoral Consortium, July 11–13, 2021, Bolzano, Italy

✉ mapellegrino@unisa.it (M. A. Pellegrino)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

and take advantage of them without requiring technical skills in query languages. Instead of proposing a unified tool to address the heterogeneity of the target audience, I opt for proposing a unified approach to guide KG exploitation (described in Section 2) and instantiate it in different interfaces (detailed in Section 3) to fulfil specific requests of each target group.

2. From Data Querying to Data Exploitation Approach

The KM process includes 1) data retrieval, 2) data refinement, and 3) data exploitation (Fig. 1). It requires the contamination of IR, Information Visualization, and Human-Computer Interaction.

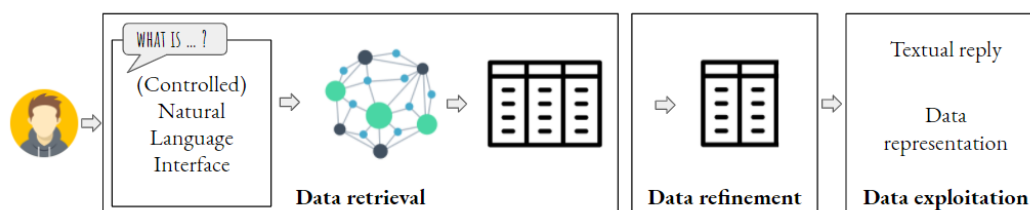


Figure 1: High-level workflow representation. End-users are guided in querying KGs by a (controlled) natural language interface, retrieved data are modeled as a tabular representation which can be refined (if needed), and, then, it can be exploited in a concrete (and reusable) artifact.

Data retrieval requires search activities that can be classified in lookup and exploratory search [8]. Lookup is a search task where users know what they are looking for and can formulate it as a direct question, as in question-answering (QA) applications, while an exploratory search task is an open-ended search that usually starts with vague information needs and requires iterative query formulation [9], facets or taxonomies [10], keyword search paradigm which includes auto-suggestions, instant results, partial matches [11, 12] to explore data.

As a data retrieval interface, I propose query builders enhanced with (controlled) NL interfaces to guide users in naturally posing questions by simulating, as much as possible, human interactions. If users have a clear objective, they can directly type or pronounce their requests. Vice versa, in exploratory search, users are guided in iteratively creating and refining questions.

As a query builder, NL queries can be translated to SPARQL to be run over a SPARQL endpoint (i.e., a way to publicly expose KG content). Among SPARQL constructs, SELECT query results can be naturally represented as tabular data. Thus, retrieved data are modelled as tables, which can be manually or automatically refined, and finally, used in data exploitation mechanisms.

It may result in textual replies or concrete artifacts, perhaps customizable and exportable, such as charts, data visualizations, data stories or virtual reality (VR) based data representations.

3. Proposed Prototypes and Explored Scenarios

To instantiate the general approach represented in Fig. 1, I considered target audience who might be interested in accessing KGs, and consequently, I designed, prototyped, and evaluated tools able to satisfy users' needs. I focus on Open Data (OD) experts and Public Administrations (PAs), education, and the Cultural Heritage (CH) community as target groups.

Open Data Experts and QueDI. During the latest years, my research lab, ISISLab, managed an European project, ROUTE-TO-PA, to support users in publishing high-quality OD and effectively exploiting them. It results in SPOD, a Social Platform for OD [13], a virtual place where users can co-create and exploit OD. SPOD is used by data producers or data enthusiastic (e.g., citizens, CH communities, PAs) with OD management skills, i.e., table manipulation and chart creation.

The problem I aimed to solve is how to make this target group able to access LOD and KGs without requiring explicit usage of SPARQL with the possibility to rely on their expertise in OD management. I proposed a *transitional approach* where OD experts are guided from LOD querying to their comfort zone. It resulted in *QueDI* (Query Data of Interest) which allows users to build queries step-by-step with an auto-complete mechanism and to exploit retrieved results by exportable and dynamic visualizations. QueDI scaffolds users in, first, creating a tabular representation of the dataset of interest by ELODIE, a query builder enhanced with a controlled NL interface. ELODIE (whose interface is visible in Fig. 2) realizes an exploratory search by organizing available data in facets and supporting users in automatically retrieving both user query results and data to go on with the query formulation by querying a configured SPARQL endpoint. Second, QueDI supports a manual dataset manipulation phase where users can exploit their skills in data refinement by aggregating, sorting, filtering, and cleaning data by interacting with a form-based interface that behaves as a SQL builder. Finally, it enables the creation of an exportable and reusable visualization. QueDI is freely available online¹. Besides its accuracy, expressivity and scalability features [14], it is *usable* according to the SUS score [14, 15].

Education. I proposed QueDI as a KM tool in the educational context to support future citizens in going beyond the passive inspection of results returned by a search engine, and in actively searching for the data that best answer their questions (as described in [16]).

Moreover, I investigated on the implicit exploitation of KGs in retrieving synonyms lookup in Novelette [17] (freely available on GitHub²), a digital storytelling environment where storytellers can create stories to graphically represent tales, data stories, or media stories. If users experience writer's block, Novelette has a suggestion provision mechanism. Users can type the word of interest and Novelette will automatically retrieve synonyms by querying BabelNet [18] and organizing retrieved results in (navigable) word clouds. It represents a keyword-based interface to implicitly explore KGs by navigating synonyms.

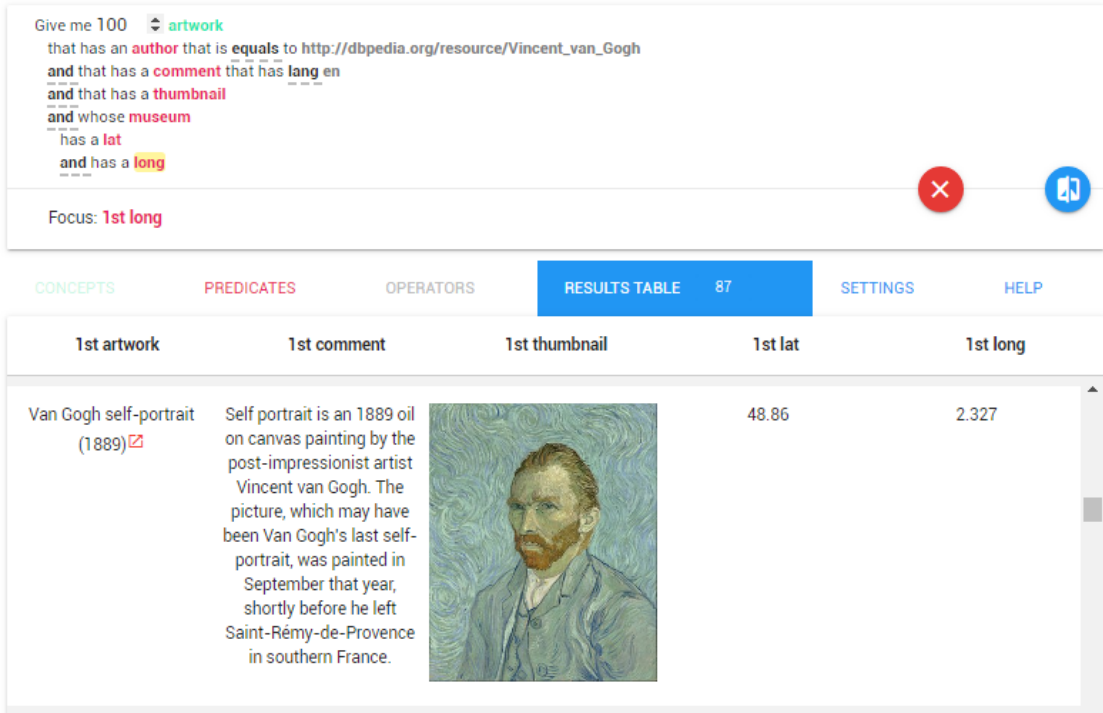
Cultural Heritage Community. In the last year, virtual exhibitions have been widely adopted to enhance physical tours. I propose to take advantage of CH KGs in an authoring platform for VR-based virtual exhibitions by combining ELODIE and an automatic mechanism to create VR-based solutions [19]. Assuming I am a Van Gogh's lover, I wish to visit all the museums that contain at least one painting by him. I can query DBpedia by ELODIE and collect all the artworks painted by Van Gogh by also retrieving information related to their geographical location. The corresponding query formulated in ELODIE and the resulting set of replies is visible in Fig. 2. Once I am satisfied with the retrieved results, instead of visualizing results by a chart as in QueDI, the generator guides me in creating VR-based virtual exhibitions³.

¹QueDI on GitHub: <https://github.com/routetopa/deep2-components/tree/master/controllets/splod-controllet>
Demo: <https://deep.routetopa.eu/deep2/COMPONENTS/controllets/splod-visualization-controllet/demo.html>

²Novelette links: <https://github.com/routetopa/storylet>, <http://www.isislab.it:19984/en/home-page-2/>

³Use case - dataset creation: https://youtu.be/63SmstO_x78.

Virtual exhibition tour and download: https://youtu.be/9LNdFY_2OJw; <https://www.isislab.it/en/virtual-museum/>.



Give me 100 artwork that has an author that is equals to http://dbpedia.org/resource/Vincent_van_Gogh and that has a comment that has lang en and that has a thumbnail and whose museum has a lat and has a long

Focus: 1st long

CONCEPTS PREDICATES OPERATORS RESULTS TABLE 87 SETTINGS HELP



1st artwork	1st comment	1st thumbnail	1st lat	1st long
Van Gogh self-portrait (1889) 	Self portrait is an 1889 oil on canvas painting by the post-impressionist artist Vincent van Gogh. The picture, which may have been Van Gogh's last self-portrait, was painted in September that year, shortly before he left Saint-Rémy-de-Provence in southern France.		48.86	2.327

Figure 2: ELODIE interface in the Van Gogh’s experience use case.

Moreover, I investigated how to make Virtual Assistants (VAs) compatible with KGs. It resulted in a community shared software framework (a.k.a. generator) that enables lay-users to create ready-to-use custom extensions for performing QA over KGs [20, 21]. This proposal represents a step forward in enabling direct search and lookup over KGs.

4. Conclusions

KGs are crucial for KM and active IR, but their query languages are difficult to use, above all by lay users. Thus, I aim to unlock the potentialities of KGs by enabling a natural human-KG interaction. I proposed solutions to explore KGs (such as ELODIE) or to directly lookup data of interest (such as by VA devices). Furthermore, the proposed data exploitation mechanisms enable agile understanding and sharing of achieved knowledge to support discussions. For each target audience, I designed an interface that replies to specific and concrete needs and relies on users’ skills without requiring them to getting aware of data format and query languages.

References

- [1] H. Paulheim, Knowledge graph refinement: A survey of approaches and evaluation methods, *Semantic Web* 8 (2016) pp. 489–508.

- [2] M. Schmachtenberg, C. Bizer, H. Paulheim, Adoption of the linked data best practices in different topical domains, in: Proc. of ISWC, Springer, 2014, pp. 245–260.
- [3] J. P. McCrae, The LOD cloud, 2007. URL: <http://lod-cloud.net>, access 2021/04/21.
- [4] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. Van Kleef, S. Auer, et al., Dbpedia—a large-scale, multilingual knowledge base extracted from wikipedia, *Semantic web 6* (2015) 167–195.
- [5] D. Vrandečić, M. Krötzsch, Wikidata: a free collaborative knowledge base, *Communications of the ACM 57* (2014) 78–85.
- [6] H. Vargas, C. B. Aranda, A. Hogan, C. López, RDF explorer: A visual SPARQL query builder, in: Proc. of ISWC, volume 11778, Springer, 2019, pp. 647–663.
- [7] P. Bellini, P. Nesi, A. Venturi, Linked open graph: Browsing multiple sparql entry points to build your own lod views, *J. of Visual Languages & Computing 25* (2014) pp. 703 – 716.
- [8] G. Marchionini, Exploratory search: from finding to understanding, *Communications of the ACM 49* (2006) 41–46.
- [9] M. Hearst, Search user interfaces, Cambridge university press, 2009.
- [10] D. Tunkelang, Dynamic category sets: An approach for faceted search, in: ACM SIGIR, volume 6, Citeseer, 2006.
- [11] P. Morville, J. Callender, Search patterns: design for discovery, O’Reilly Media, Inc., 2010.
- [12] T. Russell-Rose, T. Tate, Designing the search experience: The information architecture of discovery, Newnes, 2012.
- [13] G. Cordasco, R. De Donato, D. Malandrino, G. Palmieri, A. Petta, D. Pirozzi, G. Santangelo, V. Scarano, L. Serra, C. Spagnuolo, L. Vicidomini, Engaging citizens with a social platform for open data, in: Proc. of Digital Government Research, ACM, 2017, p. 242–249.
- [14] R. De Donato, M. Garofalo, D. Malandrino, M. A. Pellegrino, A. Petta, V. Scarano, QueDI: From knowledge graph querying to data visualization, *Semantic Systems* (2020) 70.
- [15] R. D. Donato, M. Garofalo, D. Malandrino, M. A. Pellegrino, A. Petta, V. Scarano, Linked data queries by a trialogical learning approach, in: Proc. of Int. Conf. on Computer Supported Cooperative Work in Design, CSCWD, IEEE, 2019, pp. 117–122.
- [16] R. De Donato, M. Garofalo, D. Malandrino, M. A. Pellegrino, A. Petta, Education meets knowledge graphs for the knowledge management, in: Proc. of Methodologies and Intelligent Systems for Technology Enhanced Learning, Springer, 2020, pp. 272–280.
- [17] A. Addone, R. De Donato, G. Palmieri, M. A. Pellegrino, A. Petta, V. Scarano, L. Serra, Visual storytelling by novelette, in: Proc. of IV, IEEE, 2020, pp. 723–728.
- [18] R. Navigli, S. P. Ponzetto, Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network, *Artificial Intelligence 193* (2012) 217–250.
- [19] D. Monaco, M. A. Pellegrino, V. Scarano, L. Vicidomini, The role of linked open data in authoring virtual exhibitions, 2021. Submitted to *Journal of Cultural Heritage* in Dec. 2020.
- [20] M. A. Pellegrino, M. Santoro, V. Scarano, C. Spagnuolo, Automatic skill generation for knowledge graph question answering, in: Proc. of ESWC, 2021.
- [21] M. A. Pellegrino, V. Scarano, C. Spagnuolo, Move cultural heritage knowledge graphs in everyone’s pocket, 2021. Submitted to *Semantic Web Journal* in Mar. 2021.