Sparnatural: a visual knowledge graph exploration tool

Thomas Francart[1][0000-0003-4648-0065]

[1] Sparna, 5 rue Georges Courteline 37540 Saint-Cyr-Sur-Loire, France.
`thomas.francart@sparna.fr`

Abstract. Knowledge Graphs are often relying on conceptual models with a high level of abstraction, making them hard to understand for nonexpert users, and are often hidden behind search forms, losing the flexibility to interact with the graph as a graph. In order to leverage knowledge graphs there is a need to propose end-user-oriented interactions demonstrating how they can allow to discover a knowledge domain. Sparnatural was designed and developed as a visual, configurable, easily deployable, open-source, client-side SPARQL query builder, allowing users to query any RDF knowledge graph. We describe here how Sparnatural works, how it can be configured, and we show how it was deployed in two demonstrators for the Archives Nationales de France and Bibliothèque Nationale de France. We show how Sparnatural has made it possible to give access to nonexpert users, with little effort, to these two large graphs. Workshops conducted with end-users showed they were enthusiastic about the possibility offered. They also revealed that this novel search paradigm is surprising and could benefit from user onboarding. The configuration of Sparnatural is key in providing efficient access to the graph and can be done through the specification of a search-oriented ontology.

Keywords: SPARQL, query builder, UI, Typescript, Knowledge Graph.

1   Introduction

The requirement for data interoperability has driven the development of conceptual models used for structuring knowledge graphs, like CIDOC-CRM [1] for cultural heritage, LRM [2] for libraries, or RIC-O [3] for archives. Knowledge graphs based on these ontologies are accessible either using a developer-oriented service (SPARQL endpoint), or sometimes they underly other search systems, typically multicriteria query forms or faceted search engines.

Data interoperability and data aggregation from silos in a homogeneous graph require a lot of effort, at the intellectual, business process and technical levels. These efforts are hard to demonstrate, because no good solution exists to make the graph tangible for managers and accessible for end-users: SPARQL endpoints are too technical, and search systems don't make the underlying graph tangible.

2   Sparnatural: an intuitive visual SPARQL query builder

Sparnatural [4] provides an innovative search paradigm to explore and discover these knowledge graphs. A presentation video [5] demonstrates its features. Technically speaking, Sparnatural is a visual SPARQL query builder, written in Typescript, operating purely on the client. Its only requirement is a SPARQL endpoint to which the queries can be sent. Sparnatural is open source, under a LGPL license.
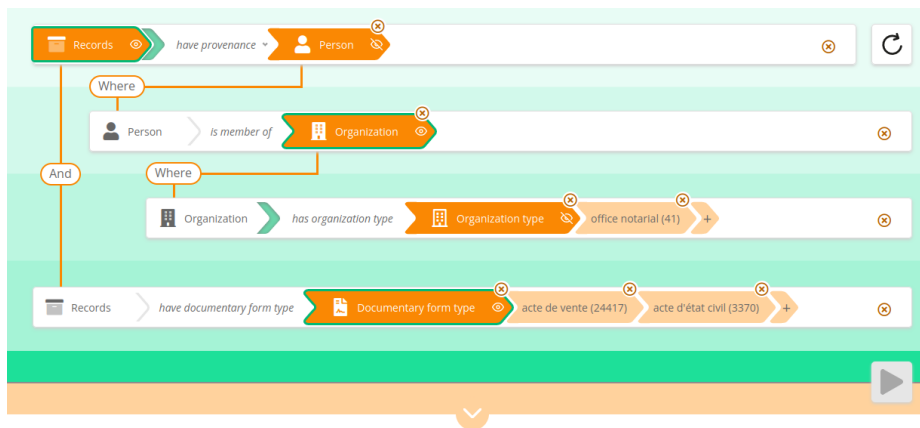

Fig. 1. A query edited in Sparnatural: "Archive records that have as provenance a Person who is a member of an organization of type 'Notarial Office', and that are of type 'bill of sale' or 'civil status document.'"

Each line in the query mimics the structure of an RDF triple pattern with a subject type, a predicate, an object type and, optionally, some values. At each step, dropdown lists enable the user to directly see the classes and predicates available.

Values of a criterion can be selected with different widgets (dropdown lists, autocompleted search fields, calendars, etc.) and the set of columns in the result set can be controlled by activating the "eye" icon on the arrows in the query pattern.

Sparnatural also offers the ability to load predefined queries, enabling data publishers to propose sample queries that can be loaded in one click. This can serve as query templates that the users can modify.

While OPTIONAL and FILTER NOT EXISTS are supported, Sparnatural does not have the objective of covering 100% of SPARQL keywords; currently it does not support UNION, FROM, or Aggregate functions like COUNT.

3   Demonstrators

The French National Archives (ANF) and the French National Library (BNF) conducted a project to add features to Sparnatural and to build two demonstrators to test

its deployment on their respective knowledge graphs. The project was successfully concluded in June 2022 by a presentation at the ANF [6].

The demonstrator of the ANF [7] allows exploring a dataset based on RIC-O [3], of about 50 million triples describing notarial archives. The demonstrator proposes two configurations of Sparnatural to navigate the same data: a generic one, that could be used for any RIC-O-based knowledge graph, and a more precise, showing specific criteria for notarial archives.

The demonstrator of the BNF [8] allows querying the data already accessible in the data.bnf.fr portal, of about 600 million triples, and based on an LRM-like [2] ontology. The performance of the SPARQL endpoint was sufficient to allow to deploy Sparnatural without further performance tuning.

The project included three workshops to confront end-users with the tool and gather feedback. Users were mostly enthusiastic, but some were lost by this new search paradigm. Some feedback was considered, such as the inclusion of the query execution button, or the "reset" button.

4   Ontology-driven configuration of Sparnatural

Sparnatural is configured by an OWL ontology, defining which classes and relationships are shown in the interface. The ontology defines the (multilingual) labels and tooltips, icons, or which value selection widget should be used for each predicate. The dropdown list and autocomplete widgets are also associated with a SPARQL query that populates the list or the autocomplete suggestions.

A key aspect of that configuration ontology is that it does not need to be the same as the ontology of the underlying knowledge graph. Classes or properties can be removed or labelled differently; shortcuts can be proposed: a single link in the search interface can correspond to a property path in the underlying graph; classes presented to users can be subsets of the (in general relatively abstract) classes of the underlying ontology.

The configuration ontology can be edited in the Protégé OWL editor, by importing two configuration ontologies [9] that provides base classes for the configuration.

The decoupling of the search ontology from the graph structure enables to show the same graph in different ways for different users. The configuration ontologies can also be shared, such as the generic configuration for the ANF demonstrator [9].

This has been a design choice that this configuration must be defined manually. We believe it is a key aspect of any knowledge graph publishing work. It does require additional work but allows to decouple the user search configuration from the actual graph structure. Work is ongoing to facilitate this process by generating a base configuration from the actual dataset, that can be further edited to produce the actual Sparnatural configuration.

5   Related work

Sparnatural builds on the visual paradigm of ResearchSpace semantic search component [10], but improves the user experience and is more expressive, with support for OPTIONAL and NOT EXISTS.

RDFExplorer [11] is another graph based SPARQL builder. Sparnatural is less expressive than RDFExplorer (it can generate only tree-shaped queries, not complete basic graph pattern), but we believe more end-user-friendly. The presentation of RDFExplorer [11] cites Bhowmick et al. [12] to summarize the challenges of visual graph querying paradigm: (1) development of graph queries requires a considerable cognitive effort; (2) users need to be able to express their goal in a systematic and correct manner, which is antagonistic with the goal of catering to lay users; (3) it is more intuitive to "draw" graph queries than to write them, which implies the need for intuitive visual interfaces. In the same family of tools, Visual SPARQL Builder [13] proposes a visual query building pane like RDFExplorer, while A-QuB [14] is more form-based. Sparnatural addresses the challenges of visual graph querying by providing the following set of features: UX is intuitive and "gamifies" the query experience; dynamic results can be provided on-the-fly; example queries can be loaded; some level of non-emptiness guarantee are provided.

6   Summary

Sparnatural offers an easy way to leverage knowledge graphs. Users with no technical knowledge of SPARQL and no a priori knowledge of the graph structure can query the data. Its client-side only deployment allows to easily include it in a webpage, replacing or complementing existing SPARQL endpoint forms.

Future work on Sparnatural includes addition of geographical and numerical search widgets; and on the configuration side, the ability to automatically derive a configuration ontology from the underlying graph structure, that can be manually tuned to be presented to end-users. SHACL-based configuration is also considered.

References

1. Le Bœuf, P., Doerr, M., Emil Ore, C., Stead, S.: CIDOC Conceptual Reference Model. https://cidoc-crm.org (last accessed 18/04/2023)
2. Riva, P., Le Boeuf, P., Žumer, M.: IFLA Library Reference Model: A Conceptual Model for Bibliographic Information. https://repository.ifla.org/handle/123456789/40 (2018).
3. Records In Contexts Ontology. https://www.ica.org/en/records-in-contexts-ontology (last accessed 18/04/2023)
4. Sparnatural https://sparnatural.eu (last accessed 18/04/2023)
5. Sparnatural presentation video. https://youtu.be/jcfldxjsBkk (last accessed 18/04/2023)
6. Sparnatural project presentation (French). https://sparnatural.eu#restitution-an-bnf (last accessed 18/04/2023)

7. Sparnatural ANF demonstrator. https://sparna-git.github.io/sparnatural-demonstrateur-an/ (last accessed 18/04/2023)

8. Sparnatural BNF demonstrator. https://data.bnf.fr/sparnatural (last accessed 18/04/2023)

9. Sparnatural configuration ontologies : http://data.sparna.fr/ontologies/sparnatural-config-core and data.sparna.fr/ontologies/sparnatural-config-datasources (last accessed 18/04/2023)

10. Sparnatural configuration for ANBF demonstrator A. https://github.com/sparna-git/sparnatural-demonstrateur-an/blob/main/sparnatural-config-A.ttl (last accessed 18/04/2023)

11. Oldman, D., Tanase, D.: Reshaping the knowledge graph by connecting researchers, data and practices in ResearchSpace. In: International Semantic Web Conference. pp. 325–340. Springer, Cham (2018).

12. Vargas, H., Buil-Aranda, C., Hogan, A., López, C.: RDF Explorer: A Visual SPARQL Query Builder. In: International Joint Conference on Aritificial Intelligence (IJCAI–PRICAI) (2020).

13. S. S. Bhowmick, B.C., Li, C.: Graph querying meets HCI: State of the art and future directions. In: ACM International Conference on Management of Data. pp. 1731–1736. ACM (2017).

14. Eipert, L.: Visual SPARQL Builder. https://leipert.github.io/vsb/# (last accessed 18/04/2023)

15. Assistive Query Building for Semantic Data. In: SEMANTICS Posters&Demos (2018).