

Sewelis: Exploring and Editing an RDF Base in an Expressive and Interactive Way

Sébastien Ferré¹ and Alice Hermann²

¹ IRISA/Université de Rennes 1, Campus de Beaulieu, 35042 Rennes cedex, France
`ferre@irisa.fr`

² IRISA/INSA de Rennes, Campus de Beaulieu, 35708 Rennes cedex 7, France
`alice.hermann@irisa.fr`

Abstract. Query-based Faceted Search (QFS), introduced in a research paper at ISWC'11, reconciles the expressiveness of querying languages (e.g., SPARQL), and the benefits of exploratory search found in faceted search. Because of the interactive nature of QFS, which is difficult to fully render in a research paper, we feel it is important to complement it with a demonstration of our QFS prototype, Sewelis (aka. Camelis 2). An important addition to the research paper is the extension of QFS to the guided edition of RDF bases, where suggestions are based on existing data. This paper motivates our approach, shortly presents Sewelis, and announces the program of the demonstration. Screencasts of the demonstration, as well as material (program and data) to reproduce it, are available at <http://www.irisa.fr/LIS/software/ sewelis>.

1 Motivation

A challenge of the Semantic Web is to enable the largest audience to explore and edit the knowledge expressed in the Semantic Web languages, in particular RDF(S) and OWL [6]. Because of the complexity of those languages, existing tools generally make a trade-off between expressiveness, i.e., the coverage of the chosen language, and usability. At one end of the spectrum, there are SPARQL endpoints and OWL editors (e.g., Protégé), which allow for full expressiveness but require advanced logical and technical knowledge. At the other end, there are Web applications that completely hide technical details but have limited functionalities (e.g., RSS). In between, there are graph visualization systems (e.g., Bramble [9]) and faceted search systems (e.g., SlashFacet [5], BrowseRDF [8], gFacet [3]) that offer a flexible and guided exploration of RDF graphs. However, those systems are far from covering the expressiveness of SPARQL (e.g., only paths of properties can be expressed). Concerning the edition of a RDF base, there exist tools that abstract away from the concrete syntax of formal languages, and guide users [7,2]. However, while their guiding generally takes into account language grammars and domain ontologies, it lacks of accuracy because it does not take into account the existing objects and their descriptions.

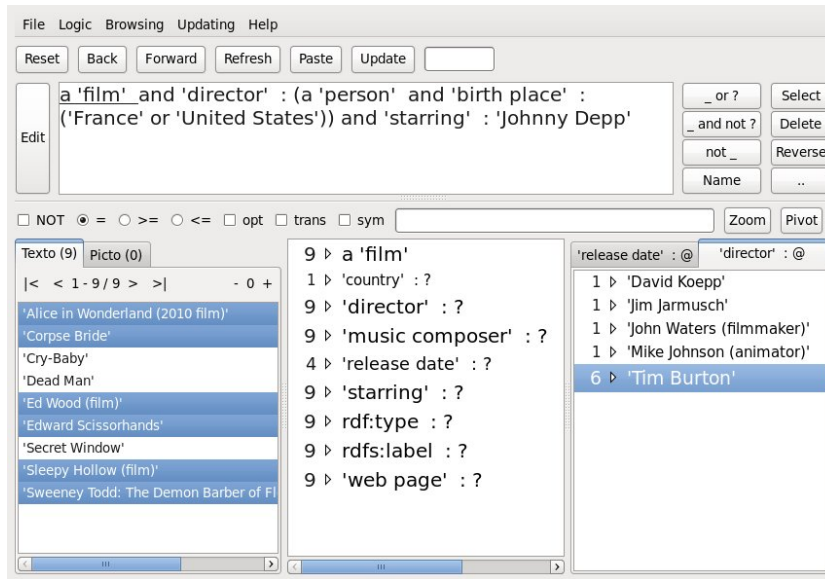


Fig. 1. Screenshot of Sewelis depicting the selection of films whose director was born in France or the United States, and starring Johnny Depp.

2 Sewelis

Sewelis combines in a same interface the exploration and the edition of a RDF base. It enables users to build in an interactive way, by guided navigation, complex queries and updates. The reachable queries cover most of the SPARQL language, and are displayed in a lighter syntax. The principle of guided navigation is to suggest to users lists of possible restrictions and other transformations for the current query or update. The set of restrictions provide rich feedback about the RDF base, and therefore support understanding-at-a-glance and exploratory search. In the case of a query, the suggested transformations are safe (no dead-end), and complete (every query that is not a dead-end is reachable). In the case of an update, the suggested transformations are derived from objects similar to the one being updated, in order to encourage consistency between object descriptions. Their computation is based on Query-based Faceted Search [1] for queries, and on UTILIS [4] for updates.

The screenshot in Figure 1 shows at the bottom left the list of films matching the query (at the top), i.e., films whose director is a person born in France or the United States, and starring Johnny Depp. The middle list is a facet hierarchy that shows the types and properties that apply to those films. The 9 selected films are all described with directors, music composers, actors, and web pages, but only 4 have a release date defined, and only 1 has a country defined. The bottom right part is a set of open facets, here 'release date' and 'director'. The visible facet

lists the directors of the selected films, and for each, the number of selected films they have directed (e.g., 6 for Tim Burton). A highlighting mechanism enables to visualize the individual relationships (i.e., triples) between elements in the left list (i.e., subjects) and elements in the right list (i.e., objects). Here, Tim Burton has been selected in the right list, and his 6 films have been highlighted accordingly in the left list. The demonstration shows how such a query can be built by successively selecting elements in the three lists (resources, classes, and properties), and pushing buttons at the top right (disjunction, negation, and co-references).

Sewelis strives for compatibility with Semantic Web standards and models, possibly extending them but in a consistent way. At the moment, it features full RDFS inference plus transitive and symmetric closures of properties from OWL. Its query and update language, LISQL, covers most of SPARQL (missing: filter primitives, named graphs). RDF/XML files can be imported, and other common formats such N-Triples and Turtle will be accepted in the near future. Sewelis supports the *follow-your-nose* pattern of Linked Data, i.e., for any resource in the base, its Linked Data description can be imported in one click. Labels and namespaces can be defined in a few clicks to allow for a concise representation of URIs, queries, and restrictions.

3 Demonstration

The demonstration illustrates the kind of exploration and edition in a RDF knowledge base that can be performed with Sewelis. The demonstration base is about films and people related to those films (directors, music composers, actors). It has been extracted from DPpedia³ as a RDF/XML file. Our base contains a selection of 120 films, 396 related people, 37 countries, and a total of about 3000 triples. Every film is possibly described by its title (as a label), production country, director(s), main actor(s), music composer, release date, and Wikipedia page. Every person is possibly described by his/her name (as a label), birth date (a year), birth place (a country), and Wikipedia page. Every country is described by its name (as a label).

The demonstration scenario splits into three phases:

1. import and untargeted exploration of the base, discovering classes and properties, the distribution of values, most frequent directors and actors, etc.;
2. targeted exploration, looking for films or people satisfying more or less complex combinations of properties. The following questions are successively answered along the principles of Query-based Faceted Search:
 - *Find the films directed by Tim Burton... and starring Johnny Depp... and starring Helena Bonham Carter* (standard faceted search),
 - *Find the films whose director was born in an English-speaking country... and released since 2000* (property paths, disjunction of values),

³ <http://www.dbpedia.org>

- Find the films that are related to France (either by the production country... or by the director’s birthplace... or by some actor’s birthplace)... or not (disjunction and negation of complex subqueries),
 - Find the people born in the US... and who are the director of a film... starring Johnny Depp... and released since 2000 (branching properties),
 - Find the people that directed a film... and played in the same film (“same as” co-references),
 - Find the films that were produced in some country... and whose director was born in another country (“other than” co-references);
3. creation or update of a film or a person in the base, using existing similar objects as models along the principles of UTILIS:
 - Create the film “Charlie and the Chocolate Factory” along with its full description,
 - Update the film “Finding Neverland” with additional actors.

Each step of this demonstration is available as a screencast on the Sewelis’ page at <http://www.irisa.fr/LIS/software/sewelis>. Moreover, a Linux binary of the prototype and the RDF/XML file of the film base can be downloaded for reproducing the demonstration and for free exploration and edition.

References

1. Ferré, S., Hermann, A.: Semantic search: Reconciling expressive querying and exploratory search. In: Aroyo, L., Welty, C. (eds.) *Int. Semantic Web Conf.* Springer (2011)
2. Haller, H.: QuiKey – an efficient semantic command line. In: *Knowledge Engineering and Management by the Masses (EKAW)*. pp. 473–482. Springer (2010)
3. Heim, P., Ertl, T., Ziegler, J.: Facet graphs: Complex semantic querying made easy. In: *et al., L.A. (ed.) Extended Semantic Web Conference*. pp. 288–302. LNCS 6088, Springer (2010)
4. Hermann, A., Ferré, S., Ducassé, M.: Guided creation and update of objects in rdf(s) bases. In: *Musen, M.A., Corcho, Ó. (eds.) Int. Conf. Knowledge Capture (K-CAP)*. pp. 189–190. ACM (2011)
5. Hildebrand, M., van Ossenbruggen, J., Hardman, L.: /facet: A browser for heterogeneous semantic web repositories. In: *et al., I.C. (ed.) Int. Semantic Web Conf.* pp. 272–285. LNCS 4273, Springer (2006)
6. Hitzler, P., Krötzsch, M., Rudolph, S.: *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC (2009)
7. Kuhn, T.: How controlled english can improve semantic wikis. In: *Semantic Wiki Work. (SemWiki) at the Eu. Semantic Web Conf. (ESWC)*. vol. 464. CEUR-WS (2009)
8. Oren, E., Delbru, R., Decker, S.: Extending faceted navigation to RDF data. In: *et al., I.C. (ed.) Int. Semantic Web Conf.* pp. 559–572. LNCS 4273, Springer (2006)
9. Schmitt, N., Niepert, M., Stuckenschmidt, H.: BRAMBLE: A web-based framework for interactive rdf-graph visualisation. Demo track at ISWC’10 (2010)