Collective Ontology Alignment

Jason B. Ellis, Oktie Hassanzadeh, Kavitha Srinivas, and Michael J. Ward

IBM T.J. Watson Research,
P.O. Box 704, Yorktown Heights, NY 10598
{jasone,hassanzadeh,ksrinivs,MichaelJWard}@us.ibm.com

1   Introduction

Enterprises are captivated by the promise of using big data to develop new products and services that provide new insights into their customers and businesses. However, there are significant challenges leveraging data across heterogeneous data stores, including making those data accessible and usable by non-experts. We designed a novel system called Helix which employs a combination of user-driven and automated techniques to bootstrap the building of a unified semantic model over virtualized data. Such a uniform semantic model allows users to query across data stores transparently, without needing to navigate a maze of data silos, data formats, and query languages.

In this poster, we discuss a specific aspect of Helix: the method by which it facilitates ontology alignment. Such alignments are very noisy and manually fixing the issues is a laborious process, especially when all the work must be done prior to putting the system into use. Instead, Helix proposes to engage users progressively in the process of ontology alignment through the course of their everyday use of the system. We do this by framing ontology alignment as a guided data exploration and integration task.

2   Guided Exploration, Linking, and Sharing

Helix provides a uniform user interface for heterogeneous data exploration and linking. This interface abstracts the underlying differences among data stores and data representations, with the goal of allowing the user to focus on their task rather than the technology. This interface engages users in the data alignment process through three key features:

1. Guided navigation - search and navigate to locate results of interest, assisted by suggestions based on semantic and schematic links
2. Saving results - save results of interest and share those results with others
3. Guided linking - users select two saved results and Helix guides them through ontology alignment. The resulting linked data can be saved, shared, and used in future links. (see Figure 1)

Through this process, users are creating ontology alignments by finding data of interest, aligning it, and saving/sharing what they find useful. In this way,
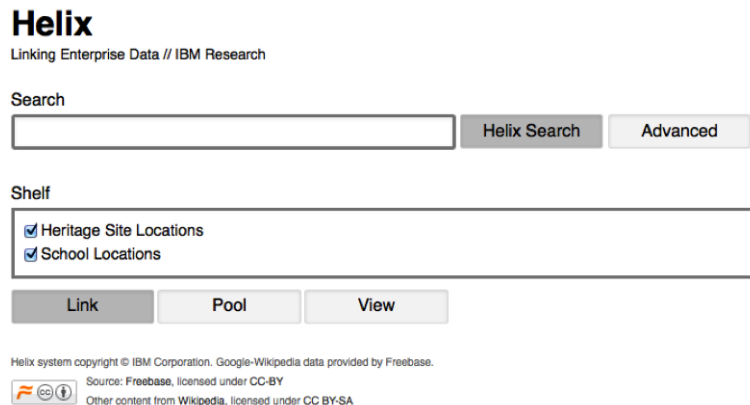
Fig. 1. Helix Front Page with two results selected for linking.

everyone who uses Helix is contributing to the alignment of the underlying data sources and can leverage the work of others.

Currently, sharing happens by one user explicitly sending saved results to another. However, we are building a recommender system that will automatically show relevant results to users through the course of their work, allowing them to more readily reuse ontology alignments.

Previous work has proposed systems that perform analysis and "pay-as-you-go" integration in specific domains using semantic technologies [2]. However, such systems typically leave the user out of the ontology mapping process. Helix explicitly engages users in linking the data they are interested in.

This work is also related to research on making it easier for non-expert users to query standard database management systems, particularly those that take an *exploratory search* approach [3]. Explorator offers a somewhat similar user experience, allowing users to explore RDF data through a process involving search, faceted navigation, and set operations [1]. By contrast, Helix allows users to navigate heterogeneous data and build complex queries through a process of progressively linking saved results. It also assists users through semantic & schematic guidance, linkage discovery, and (ultimately) recommendations.

References

1. de Araújo, S., Schwabe, D.: Explorator: a tool for exploring RDF data through direct manipulation. In: Linked Data on the Web WWW Workshop (2009)
2. Lopez, V., Kotoulas, S., Sbodio, M.L., Stephenson, M., Gkoulalas-Divanis, A., Mac Aonghusa, P.: QuerioCity: a linked data platform for urban information management. In: ISWC'12: Proceedings of the 11th international conference on The Semantic Web. Springer-Verlag (Nov 2012)
3. White, R.W., Drucker, S.M., Marchionini, G., Hearst, M., schraefel, m.c.: Exploratory search and HCI. In: CHI '07 extended abstracts. pp. 2877–2880. ACM Press, New York, New York, USA (2007)