Template Based Semantic Similarity for Security Applications

Boanerges Aleman-Meza, Christian Halaschek-Wiener[1], Satya Sanket Sahoo, Amit Sheth, I. Budak Arpinar

Large Scale Distributed Information Systems (LSDIS) Lab,
Department of Computer Science
University of Georgia, Athens, GA. 30602-7404
boanerg@cs.uga.edu, halasche@cs.umd.edu,
{sahoo, amit, budak}@cs.uga.edu
http://lsdis.cs.uga.edu/

Today's search technology delivers impressive results in finding relevant documents for given keywords. However many applications in various fields including genetics, pharmacy, social networks, etc. as well as national security need more than what traditional search can provide. Users need to query a very large knowledge base (KB) using semantic similarity, to discover its relevant subsets. One approach is to use templates that support semantic similarity-based discovery of suspicious activities, that can be exploited to support applications such as money laundering, insider threat and terrorist activities. Such discovery that relies on a semantic similarity notion will tolerate syntactic differences between templates and KB using ontologies. We address the problem of identifying known scenarios using a notion of template-based similarity performed as part of the SemDIS project [1, 3]. This approach is prototyped in a system named TRAKS (Terrorism Related Assessment using Knowledge Similarity) and tested using scenarios involving potential money laundering.

A *template* provides a means to represent a specific manner in which collection of entities are interconnected thus capturing a scenario or a set of circumstances of interest in security applications. The template is defined using classes and relationships of an ontology, forming a 'typed' directed graph. In terms of information retrieval, a template can be viewed as a query. Querying requires data to match the classes and the interconnections of the named relationships of the template. However, our approach exploits inheritance hierarchies in ontologies to detect similarities semantically. Computing similarity involves looking at syntactical, structural, and semantic properties of instance data with respect to the template.

Our work is aligned with the current semantic Web vision where ontologies play a central role. We used SWETO [2] as our dataset because it includes entities and relationships of relevance to security applications (e.g., banks, organizations, persons, watch lists). Known money laundering scenarios were described as templates to evaluate our approach. The results are ranked based on how close the types of entities or relationships are to those in the template. A graph-based visualization, based on TouchGraph, provides support better understanding of the results.

[1] Currently a Ph.D. student in the Computer Science Dept. at University of Maryland.

It is possible to test our prototype with available datasets and ontologies (using W3C's OWL recommendation). Figure 1 illustrates the architecture of TRAKS. Both the Web application and a technical report are available online[2].

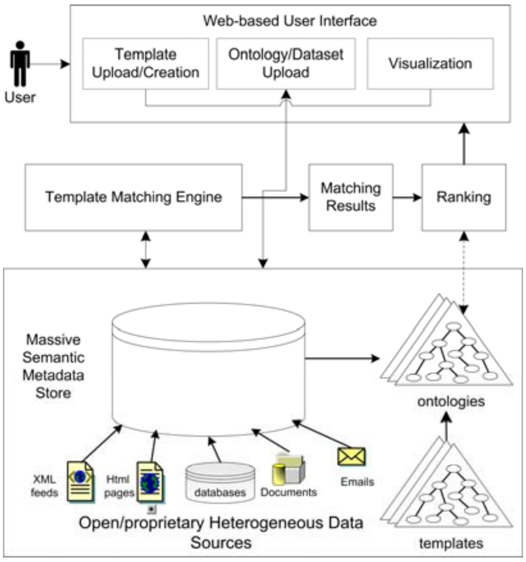

Fig. 1. TRAKS System Architecture

Acknowledgements. This work is funded by NSF-ITR-IDM Award#0325464 titled 'SemDIS: Discovering Complex Relationships in the Semantic Web' and NSF-ITR-IDM Award#0219649 titled 'Semantic Association Identification and Knowledge Discovery for National Security Applications.'

References

1. B. Aleman-Meza, C. Halaschek, I.B. Arpinar, A. Sheth, Context-Aware Semantic Association Ranking. Proceedings of Semantic Web and DB Workshop, Berlin, 2003, pp. 33-50
2. B. Aleman-Meza, C. Halaschek, A. Sheth, I.B. Arpinar, and G. Sannapareddy. SWETO: Large-Scale Semantic Web Test-bed. Proceedings of the 16th International Conference on Software Engineering and Knowledge Engineering (SEKE2004): Workshop on Ontology in Action, Banff, Canada, June 21-24, 2004, pp. 490-493
3. A. Sheth, B. Aleman-Meza, I.B. Arpinar, C. Halaschek, C. Ramakrishnan, C. Bertram, Y. Warke, D. Avant, F.S. Arpinar, K. Anyanwu, and K. Kochut. Semantic Association Identification and Knowledge Discovery for National Security Applications. Journal of Database Management, Jan-Mar 2005, 16 (1):33-53

[2] http://lsdis.cs.uga.edu/proj/traks/