Learning to Rank Complex Semantic Relationships

Na Chen, Department of Computer Science, University of Southern California, Los Angeles, CA, USA
Viktor K. Prasanna, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA

ABSTRACT
This paper presents a novel ranking method for complex semantic relationship (semantic association) search based on user preferences. The authors’ method employs a learning-to-rank algorithm to capture each user’s preferences. Using this, it automatically constructs a personalized ranking function for the user. The ranking function is then used to sort the results of each subsequent query by the user. Query results that more closely match the user’s preferences gain higher ranks. Their method is evaluated using a real-world RDF knowledge base created from Freebase linked-open-data. The experimental results show that the authors’ method significantly improves the ranking quality in terms of capturing user preferences, compared with the state-of-the-art.

Keywords: Complex Semantic Relationship, Freebase, Learning to Rank, Semantic Association, Semantic Web, User Preferences

INTRODUCTION
Semantic relationships are at the heart of ontologies. They connect words, terms and entities through meaning, and thus enable a graph representation of knowledge with rich semantics. Complex semantic relationship, also known as semantic association (Aleman-Meza et al., 2003), is a sequence of consecutive properties that link two resource entities; in RDF graphs, it is a path consisting of labeled edges that connects two entity nodes. Semantic association mining is a critical step towards getting useful semantic information for better integration, search and decision-making. A number of search techniques and query languages have been developed for discovering semantic associations, such as \( \rho \)-Queries (Anyanwu & Sheth, 2003) and SPARQLeR (Kochut & Janik, 2007). A semantic association search query consists of a pair of entities, and the results contain all the semantic associations between them.

With the amount, scale and complexity of ontologies growing rapidly, the number of semantic associations between a pair of entities is becoming increasingly overwhelming. Thus, a semantic association search is very likely to return too many results for a user to digest. For example, we parsed the entire fictional_universe domain of Freebase linked-open-data (Google, 2011) into an RDF knowledge base containing 192K resources and 411K properties. We ob-
served that in such a knowledge base, even a simple query (e.g., between *Harry Potter* and *James Potter*) with a strict path length restriction (e.g., 10) returns thousands of semantic associations. Thus, an effective ranking technique is need for identifying the most relevant results.

A fundamental challenge in semantic association ranking is to understand user preferences. Different users can have different preferences in terms of personal interests and search intentions. For example, Figure 1 is a small fraction of the RDF knowledge base we created from Freebase data (Google, 2011). Given a query “finding semantic associations between *Harry Potter* and *James Potter*”, a few typical search results are listed in Table 1. Among these results, a user who is familiar with the fiction *Harry Potter* would find complicated relationships such as Result 4 more informative, while another user interested in the topic of superpower may want relationships like Result 5 to gain higher ranks. Given that such preferences are difficult to be explicitly expressed in current semantic association query languages (Anyanwu & Sheth, 2003; Kochut & Janik, 2007), it is the ranking method’s responsibility to cater for each individual user’s specific preferences.

Many ranking methods have been proposed towards personalized semantic association search, by allowing users to manually tune a predefined set of ranking metrics. For example, SemRank (Anyanwu *et al.*, 2005) provides a slider bar for users to vary the rank mode between *Conventional* and *Discovery* via a graphical interface, and thus allows them to view their results through different lenses; while the ranking method in Aleman-Meza *et al.* (2005) allows users to specify weights of certain metrics. However, manually adjusting the ranking metrics is a subtle task, which requires a considerable amount of time and effort. In addition, manual tuning can become extremely difficult when the volume and the complexity of search results are overwhelming, which is very common in large knowledge bases.

The goal of our work is to automatically capture user preferences and effectively leverage these preferences to personalize semantic association search results. Our method is mo-

---

Figure 1. A small fraction of the RDF knowledge base we created from Freebase data under the topic “fictional_universe”. The color of each instance node denotes its class.
Related Content

Family History Information Exchange Services Using HL7 Clinical Genomics Standard Specifications
[www.igi-global.com/article/family-history-information-exchange-services/2814?camid=4v1a](www.igi-global.com/article/family-history-information-exchange-services/2814?camid=4v1a)

Automatically Integrating Heterogeneous Ontologies from Structured Web Pages
[www.igi-global.com/article/automatically-integrating-heterogeneous-ontologies-structured/2836?camid=4v1a](www.igi-global.com/article/automatically-integrating-heterogeneous-ontologies-structured/2836?camid=4v1a)
Knowledge Management and Lifelong Learning in Archival Heritage: Digital Collections on a Semantic Scope for Educational Potential
www.igi-global.com/chapter/knowledge-management-lifelong-learning-archival/29292?camid=4v1a

A Theory of Social Agentivity and its Integration into the Descriptive Ontology for Linguistic and Cognitive Engineering
www.igi-global.com/article/theory-social-agentivity-its-integration/63645?camid=4v1a