



Chapter XX

Semantic Analytics in Intelligence: Applying Semantic Association Discovery to Determine Relevance of Heterogeneous Documents

Boanerges Aleman-Meza, University of Georgia, USA

Amit P. Sheth, University of Georgia, USA

Devanand Palaniswami, University of Georgia, USA

Matthew Eavenson, University of Georgia, USA

I. Budak Arpinar, University of Georgia, USA

ABSTRACT

We describe an ontological approach for determining the relevance of documents based on the underlying concept of exploiting complex semantic relationships among real-world entities. This research builds upon semantic metadata extraction and annotation, practical domain-specific ontology creation, main-memory query processing, and the notion of semantic association. A prototype application illustrates the approach by supporting the identification of insider threats for document access. In this scenario, we describe how investigative assignments performed by intelligence analysts are captured into a context of investigation by including concepts and

relationships from the ontology. A relevance measure for documents is computed using semantic analytics techniques. Additionally, a graph-based visualization component allows exploration of potential document access beyond the 'need to know'. We also discuss how a commercial product using Semantic Web technology, Semagix Freedom, is used for metadata extraction when designing and populating an ontology from heterogeneous sources.

INTRODUCTION

Creating applications that allow users to gain insightful and actionable information or mine for interesting patterns from vast amounts of heterogeneous information is one of the most exciting new areas of information systems research. This information may come from numerous sources spanning proprietary, trusted, or even open-source locations, including intranets, the deep Web and the open Web. The fast-emerging markets of business intelligence as well as national and homeland security are finding themselves in increasing need of a class of applications dealing with risk and compliance (Sheth, 2005). A representative example of this type of software is the *Insider Threat* application, which involves validating the legitimate access of sensitive documents. While physical security measures may help reduce malevolent or otherwise undesirable access to documents by employees within an organization, the development of new information-based security systems provides additional capabilities for defense against insider threat attacks. The intent of this application is to ensure that analysts who are assigned various investigative tasks access the information on a "need to know basis," and that the system identifies access to irrelevant information in an attempt to reduce the chances that confidential information is leaked or otherwise released inappropriately.

Research into techniques for searching documents was a critical component of the first generation of the Web and has since gone from academia to mainstream. A second generation "Semantic Web" will be built by adding semantic annotations to Web content that software can utilize and from which humans can benefit. Large-scale semantic annotation of data (domain-independent or domain-specific) is now possible because of numerous advances in the areas of entity identification, automatic classification, taxonomy and ontology development, and metadata extraction (Dill et al., 2003; Hammond, Sheth, & Kochut, 2002; Shah, Finin, Joshi, Cost, & Mayfield, 2002). Relationships are at the heart of semantics (Sheth, Arpinar, & Kashyap, 2003; Woods, 1975). The next frontier, which will fundamentally change the way we acquire and use knowledge, is to automatically identify complex relationships between entities in this semantically annotated data. Instead of a search engine that merely returns documents containing terms of interest, we propose an approach that supports semantic analysis of heterogeneous content to return actionable information that gives useful insight into the connection between documents and real-world entities, thus providing superior support for important decisions and actions. Demonstrating this approach is a prototype application that supports the task of ensuring that an intelligence analyst accesses documents on a "need to know" basis, which means that only those documents relevant to the analyst's investigative assignment should be viewed. However, this is only one of the many semantic applications needed as part of the advanced information technology pool to support homeland security.

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/semantic-analytics-intelligence/4402

Related Content

Design of a Data Model for Social Network Applications

Susanta Mitra, Aditya Bagchi and A.K. Bandyopadhyay (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 414-439). www.irma-international.org/chapter/design-data-model-social-network/7924

Application of Decision Tree as a Data Mining Tool in a Manufacturing System

S. A. Oke (2009). *Selected Readings on Database Technologies and Applications* (pp. 234-251). www.irma-international.org/chapter/application-decision-tree-data-mining/28555

From Chinese Philosophy to Knowledge Discovery in Databases A Case Study: Scientometric Analysis

Pei Liu and Eric Boutin (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 632-642). www.irma-international.org/chapter/chinese-philosophy-knowledge-discovery-databases/20749

Parallel Skyline Computation Exploiting the Lattice Structure

Markus Endres and Werner Kießling (2015). *Journal of Database Management* (pp. 18-43). www.irma-international.org/article/parallel-skyline-computation-exploiting-the-lattice-structure/153516

Using Ontology Languages for Conceptual Modeling

Palash Bera, Anna Krasnoperova and Yair Wand (2010). *Journal of Database Management* (pp. 1-28). www.irma-international.org/article/using-ontology-languages-conceptual-modeling/39114