

Chapter 2.14

A Semantic Web–Based Approach for Building Personalized News Services

Flavius Frasincar

Erasmus University Rotterdam, The Netherlands

Jethro Borsje

Erasmus University Rotterdam, The Netherlands

Leonard Levering

Erasmus University Rotterdam, The Netherlands

ABSTRACT

This article proposes Hermes, a Semantic Web-based framework for building personalized news services. It makes use of ontologies for knowledge representation, natural language processing techniques for semantic text analysis, and semantic query languages for specifying wanted information. Hermes is supported by an implementation of the framework, the Hermes News Portal, a tool which allows users to have a personalized online access to news items. The Hermes framework and its associated implementation aim at advancing the state-of-the-art of semantic approaches for personalized news services by employing Semantic Web standards, exploiting domain information, using a word sense disambiguation procedure, and being able to express temporal constraints for the desired news items.

INTRODUCTION

The simplicity, availability, reachability, and reduced exploitation costs have made the Web one of the most common platforms for information publishing and dissemination. This is particularly true for news agencies that use Web technologies to present emerging news regarding different types of events as for example business, cultural, sport, and weather events. Most of this information is published as unstructured text that is made available to a general audience by means of Web pages.

The heterogeneity of the Web audience and the diversity of the published information asks for more refined ways of delivering information that would enable users to access news items that interest them. For this purpose the Really Simple Syndication (RSS) (Winer, 2003) standard

was developed that publishes information in a semi-structured format that supports machine processing. This format is based on metadata that (1) associates news items with channels (feeds) that have properties like categories (e.g., business, sport, politics, etc.), title, publication date, etc., and (2) describes news items by means of their properties as categories (e.g., online business, business system, Internet marketing, etc.), release time, title, abstract, link to the original published information, etc.

Most of the annotations supported by the RSS feeds are coarse-grained in the sense that they fail to identify the different topics addressed in a certain news item. Also, the current annotations are only partially processable by machines as the tags do not have unique semantic meaning associated to them and thus have different interpretations. Being able to understand the semantic content of a news item would enable a fine-grained categorization of this information, thus better supporting the users (casual users, media analysts, stock brokers, etc.) information needs.

In order to make the Web data not only machine readable but also machine understandable the World Wide Web Consortium proposes the Semantic Web (Berners-Lee, Hendler, & Lassila, 2001), a sequence of technologies that allow for self-describing content. On the Semantic Web metadata is defined using semantic information usually captured in ontologies. Some of the most popular formats to describe ontologies on the Semantic Web are RDF(S) (Klyne & Carroll, 2004) (Brickley & Guha, 2004) and OWL (Bechhofer et al., 2004).

A special class of users who make daily use of (emerging) news is that of stock brokers. Because news messages may have a strong impact on stock prices, stock brokers need to monitor these messages carefully. Due to the large amounts of news information published on a daily basis, the manual task of retrieving the most interesting news items with respect to a given portfolio is a challenging one. Existing approaches such as Google Finance

or Yahoo! Finance are developed to meet these personalization needs by supporting automatic news filtering on the Web.

Current approaches to news filtering are able to retrieve only the news that explicitly mention the companies involved, failing to deliver indirect information which is also deemed relevant for the considered portfolio. For example, for a portfolio based on Google shares, such systems fail to deliver news items related to competitors of Google, such as Yahoo! or Microsoft, which might have an indirect influence on the share price of Google. Exploiting the semantic contextual information related to companies such as its competitors, CEO's, alliances, products, etc., enables a more comprehensive overview of relevant news with respect to a certain portfolio.

Another limitation of existing news filtering systems is their inability to cope with delivering news items satisfying temporal constraints. The time aspect is of utmost importance when, for example, one considers the fact that news items usually have an immediate impact on stock prices, or when one desires to do a historical analysis of past news and stock price evolutions. Being able to exploit the timestamps associated to news items enables retrieving only news that obey user-determined time-related constraints.

In this article we propose the Hermes framework, a semantic-based approach for retrieving news items related, directly or indirectly, to the concepts of interests from a domain ontology. In addition these news items might need to satisfy temporal constraints. For illustration purposes we focus here on the NASDAQ stock market domain (Kandel & Marx, 1997), but the genericity of our approach makes it applicable also to other domains, as, e.g., tourism or scientific domains. The Hermes News Portal (HNP) is an implementation of the Hermes framework, which allows the user to specify queries for the concepts of interest and temporal constraints, and retrieve the corresponding news items.

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-web-based-approach-building/37649

Related Content

Binary Self-Adaptive Salp Swarm Optimization-Based Dynamic Load Balancing in Cloud Computing

Bivasa Ranjan Parida, Amiya Kumar Rathand Hitesh Mohapatra (2022). *International Journal of Information Technology and Web Engineering* (pp. 1-25).

www.irma-international.org/article/binary-self-adaptive-salp-swarm-optimization-based-dynamic-load-balancing-in-cloud-computing/295964

Open Source in Web-Based Applications: A Case Study on Single Sign-On

Claudio Agostino Ardagna, Fulvio Fratiand Gabriele Gianini (2006). *International Journal of Information Technology and Web Engineering* (pp. 81-94).

www.irma-international.org/article/open-source-web-based-applications/2614

A Graphical User Interface (GUI) Testing Methodology

Zafar Singhera, Ellis Horowitzand Abad Shah (2008). *International Journal of Information Technology and Web Engineering* (pp. 1-18).

www.irma-international.org/article/graphical-user-interface-gui-testing/2643

The Impact of On-Line Consumer Reviews on Value Perception: The Dual-Process Theory and Uncertainty Reduction

Hsin Hsin Chang, Po Wen Fangand Chien Hao Huang (2016). *Web Design and Development: Concepts, Methodologies, Tools, and Applications* (pp. 1498-1524).

www.irma-international.org/chapter/the-impact-of-on-line-consumer-reviews-on-value-perception/137409

Spectral Graph and Minimal Spanning Tree for 3D Polygonal Meshes Fingerprinting

Emad E. Abdallah, Ibrahim Al-Oqily, Alaa E. Abdallah, Ahmed F. Otoomand Ayoub Alsarhan (2014). *International Journal of Information Technology and Web Engineering* (pp. 40-53).

www.irma-international.org/article/spectral-graph-and-minimal-spanning-tree-for-3d-polygonal-meshes-fingerprinting/124028