Chapter 67
Quality Measures for Semantic Web Application

Adiraju Prasanth Rao
Anurag Group of Institution, India

ABSTRACT

The Semantic Web is a standard of Common Data Formats on WWW with aim to convert the current web data of unstructured and semi-structured documents into a common framework that allows data to be shared and reused across applications, enterprises. The main purpose of the Semantic Web is driving the evolution of the current Web by enabling users to find, share, and combine information more easily. Humans are capable of using the Web to carry out tasks such as searching for the lowest price for a LAPTOP. However, machines cannot accomplish all of these tasks without human direction, because web pages are designed to be read by people, not machines. The semantic web is a vision of information that can be readily interpreted by machines, so machines can perform more of the tedious work involved in finding, combining, and acting upon information on the web. The chapter presents the architecture of semantic web, its challenging issues and also data quality principles. These principles provide a better decision making within organization and will maximize long term data integration and interoperability.

INTRODUCTION

This is the age of information which can be gathered from different sources. Information-oriented communication networks, services and applications in various domains have recently undergone rapid changes due to sudden increases in number of network enabled devices and sensors deployed in physical environments. Within the next ten years, it is expected that billions of devices will generate large volumes of real world data for many applications and services in various domains such as smart phones, health care, transport environmental monitoring system, mobile applications etc. Data generated by these devices is expected to be mostly multimodal in nature (like temperature, light, sound and video) and manifold in character. This implies that the quality of data can change with devices, location and time.

DOI: 10.4018/978-1-5225-5191-1.ch067

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

The amount of data available on the world-wide-web is already huge and is increasing at a rapid pace. About 2.5 billion bytes of data is generated each day, which includes sensory data and data from various other sources. The generated data from different sources can be analyzed to extract actionable information which enables better understanding about the physical world and value added products and services provided by the manufacturers.

Currently, every business transaction or decision is data based. Data has also become more and more important for social activities. Large amount of business related data is published on the web with the development of internet from "Web of Documents" to "Web of Data". Today's web consists of large libraries of web documents and interconnected documents that are transmitted by systems and are available to the public. Applications let people view, search and combine data like calendars, address books, playlist and spreadsheet. This technology has been developed from hypertext systems to which anyone can contribute. They also reveal that the quality of information or the consistency of documents cannot be constantly assured. The present web or traditional web can be categorized into the 2nd generation. The traditional web was an association among internet users, content providers and enterprises. Originally, data was posted on web sites and users simply read or downloaded the content. This has led to the concept of the internet as a huge distributed database with users performing three major operations - search, integration and web data mining. These operations are expanded in the following paragraphs.

- **Search:** The main goal of search is to identify and access the information or resources on the web.
- **Integration:** Integration is the process of combining and aggregating different resources to accomplish a specialized task. For example, searching for Indian food in the United States of America requires two resources - a restaurant and Indian food items. By integrating these two resources, we can look forward to a nice dinner.
- **Web Data Mining:** Web data mining is the nontrivial extraction of useful information from large data sets or databases. The author's D. Artz & Y. Gil. (2007 [5]) discussed Special data mining techniques are used to automatically discover and extract information dynamically from web documents and services. This involves three main tasks viz. resource finding which retrieves the intended web documents, information selection and preprocessing which automatically selects and preprocesses specific information retrieved from web resources and lastly a generalization which dynamically discovers the patterns in individual web sites as well as across multiple sites.

For all three categories of internet users, the internet is entirely meant for reading and is purely display oriented and is therefore based on keyword matching only. The main reason for this is that the current internet contains mostly unstructured data. The current web involves too much manual work and needs to involve more automation. In its current version, the web data mining search mechanism could be very expensive because applications are highly specialized and application oriented.

We thus identify three major problems faced by all three internet user categories and would look for solutions to make these activities more efficient. The expected web has following features:

1. The output data from search results should be relevant
2. More automation is needed in the case of integration searching
3. Web data mining needs to be less expensive.

9 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/quality-measures-for-semantic-web-application/198611

Related Content

A Systematic Study of Feature Selection Methods for Learning to Rank Algorithms
Mehrnoush Barani Shirzadand Mohammad Reza Keyvanpour (2018). *International Journal of Information Retrieval Research (pp. 46-67).*
www.irma-international.org/article/a-systematic-study-of-feature-selection-methods-for-learning-to-rank-algorithms/204466

SAR: An Algorithm for Selecting a Partition Attribute in Categorical-Valued Information System Using Soft Set Theory
Rabiei Mamat, Tutut Herawanand Mustafa Mat Deris (2011). *International Journal of Information Retrieval Research (pp. 38-52).*
www.irma-international.org/article/sar-algorithm-selecting-partition-attribute/68375

Visualizing Information Science Knowledge by Modelling Domain Ontology (OIS)
Ahlam F. Sawsaaand Joan Lu (2014). *International Journal of Information Retrieval Research (pp. 1-28).*
www.irma-international.org/article/visualizing-information-science-knowledge-by-modelling-domain-ontology-ois/113330

Solving the Cubic Cell Formation Problem Using Simulated Annealing
Hamida Bouazizand Ali Lemouari (2022). *International Journal of Information Retrieval Research (pp. 1-19).*
www.irma-international.org/article/solving-the-cubic-cell-formation-problem-using-simulated-annealing/290827

The Use of Text Mining Techniques in Electronic Discovery for Legal Matters
Michael W. Berry, Reed Esauand Bruce Kiefer (2012). *Next Generation Search Engines: Advanced Models for Information Retrieval (pp. 174-190).*
www.irma-international.org/chapter/use-text-mining-techniques-electronic/64425