

Chapter 33

Web Intelligence: A Fuzzy Knowledge–Based Framework for the Enhancement of Querying and Accessing Web Data

Jafreezal Jaafar

Universiti Teknologi PETRONAS, Malaysia

Kamaluddeen Usman Danyaro

Universiti Teknologi PETRONAS, Malaysia

M. S. Liew

Universiti Teknologi PETRONAS, Malaysia

ABSTRACT

This chapter discusses about the veracity of data. The veracity issue is the challenge of imprecision in big data due to influx of data from diverse sources. To overcome this problem, this chapter proposes a fuzzy knowledge-based framework that will enhance the accessibility of Web data and solve the inconsistency in data model. D2RQ, protégé, and fuzzy Web Ontology Language applications were used for configuration and performance. The chapter also provides the completeness fuzzy knowledge-based algorithm, which was used to determine the robustness and adaptability of the knowledge base. The result shows that the D2RQ is more scalable with respect to performance comparison. Finally, the conclusion and future lines of the research were provided.

INTRODUCTION

There is a growing number of hypes in today's world of data. The data that change the human interaction through leveraging power of accessibility. The pervasive applications and tools such as phones, computers and cars have built the knowledge base which needs large data stores and management. MapReduce, resource description framework (RDF) as well as simple protocol and RDF query language (SPARQL) are the current technologies in data science that enable Web users to access and query information in a suitable way.

DOI: 10.4018/978-1-4666-9840-6.ch033

Perhaps, due to plethora of data, the database structures need to be enhanced in a way that will ease the query for proper processing and big data exploration. This is a challenge in big data management in which the redundancy or/and unorganized information is of concerned. As schema is the backbone for every database but it is not sufficient when the data is unstructured. Similarly, the quality of data re-usability reduces as it increases everyday through Web by making it to be unstructured. This means that the presence of unstructured data reduces the integrity of data and making it to be difficult for reuse. Many distributed databases and database schemas are now connected with large information. More specifically, large amounts of information on the Web cause the problem of uncertainty and imprecision due to access and querying bunch of data. Uncertainties, imprecisions and inconsistencies in data model is one of the tactical challenges of big data (Jewell et al., 2014; Savas, Sagduyu, Deng, & Li, 2014). The uncertain or imprecise issue of data is called veracity which among the dimensions of big data. Significantly, big data tools such as MapReduce and Hadoop deal with both structured and unstructured data to simplify accessibility. Therefore, the main contribution of this chapter is to provide a fuzzy knowledge-based framework that will be used as channel of accessing and querying big data. The framework guides in providing the precise information to the user through querying multiple data sources. This process allows machine to reason intelligently. The reasoning approach of this work is specifically using fuzzy logic-based systems. The proposed work suggests the scalability of the fuzzy knowledge base (KB).

This chapter proceeds with the second section that discusses about the retrospective background knowledge on Web intelligence, big data and the uncertainty. Third section illustrates the overview about the proposed framework which includes: the general architecture of the system and fuzzy ontology. In section four, the implementation process has been presented. Subsequently, section five discusses the result of the proposed framework. Finally, section six provides the conclusion and future works.

WEB INTELLIGENCE

Web information has great impact in human life especially in the domain of world knowledge such as uncertainty and imprecision. The Web intelligence constitutes the usage of WWW as a phenomenon of retrieving information from the storage efficiently. It acts as an agent in which the machine can reason and conveys the message using Web tools. The agent goes round and integrates the resource which finally presents the information to user through a Web page. The resources or things depend on RDF that links the entire concept together through Uniform Resource Identifier (URI). Web intelligence is a well-known research area which converges subjects such as artificial intelligence, databases, Web science, Semantic Web, and information retrieval (Berners-Lee et al., 2006; Camacho et al., 2013; Shadbolt & Berners-Lee, 2008; Shroff, 2013; Williams et al., 2014). Therefore, reasoning in Web data is the first step in finding the solution of a problem in the knowledge-based system.

To work with intelligence, it is necessary to define the complex knowledge acquisition, knowledge inference, deduction and knowledge representation in order to make the conceptualization of a knowledge model suitable for reasoning (Camacho et al., 2013; G'abor, 2007; Russell et al., 2010; Zadeh, 2004). This may lead the ontology to be machine-processable and allows the precise interpretation of knowledge representation. The knowledge representation has major three flavors: concepts, relations and instances (C, R and I). Generally, the representations are being understood by humans. In this sense, if the knowledge representation is done through the concepts and relates it to humans using symbols on a particular group,

21 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/web-intelligence/150190

Related Content

Rough Set on Two Universal Sets Based on Multigranulation

D. P. Acharjya and Mary A. Geetha (2014). *Data Mining and Analysis in the Engineering Field* (pp. 159-179).

www.irma-international.org/chapter/rough-set-on-two-universal-sets-based-on-multigranulation/109981

Cybersecurity of Medical Data Based on Big Data and Privacy Protection Method

Jianhong Li, An Pan and Tongxing Zheng (2023). *International Journal of Data Warehousing and Mining* (pp. 1-16).

www.irma-international.org/article/cybersecurity-of-medical-data-based-on-big-data-and-privacy-protection-method/325222

Data Mining in Designing an Agent-Based DSS

Christian Bohm, Maria R. Galli and Omar Chiotti (2003). *Data Mining: Opportunities and Challenges* (pp. 421-436).

www.irma-international.org/chapter/data-mining-designing-agent-based/7612

A Hybrid Method for High-Utility Itemsets Mining in Large High-Dimensional Data

Guangzhu Guangzhu Yu, Shihuang Shao, Bin Luo and Xianhui Zeng (2009). *International Journal of Data Warehousing and Mining* (pp. 57-73).

www.irma-international.org/article/hybrid-method-high-utility-itemsets/1823

A Multi-Agent Neural Network System for Web Text Mining

Lean Yu, Shouyang Wang and Kin Keung Lai (2008). *Emerging Technologies of Text Mining: Techniques and Applications* (pp. 162-183).

www.irma-international.org/chapter/multi-agent-neural-network-system/10181