Chapter 10
A Web Knowledge Discovery Engine Based on Concept Algebra

Kai Hu
University of Calgary, Canada

Yingxu Wang
University of Calgary, Canada

Yousheng Tian
University of Calgary, Canada

ABSTRACT

Autonomous on-line knowledge discovery and acquisition play an important role in cognitive informatics, cognitive computing, knowledge engineering, and computational intelligence. On the basis of the latest advances in cognitive informatics and denotational mathematics, this paper develops a web knowledge discovery engine for web document restructuring and comprehension, which decodes on-line knowledge represented in informal documents into cognitive knowledge represented by concept algebra and concept networks. A visualized concept network explorer and a semantic analyzer are implemented to capture and refine queries based on concept algebra. A graphical interface is built using concept and semantic models to refine users' queries. To enable autonomous information restructuring by machines, a two-level knowledge base that mimics human lexical/syntactical and semantic cognition is introduced. The information restructuring model provides a foundation for automatic concept indexing and knowledge extraction from web documents. The web knowledge discovery engine extends machine learning capability from imperative and adaptive information processing to autonomous and cognitive knowledge processing with unstructured documents in natural languages.

INTRODUCTION

A central problem in web knowledge discovery, retrieval, and acquisition is how to formulate structured and effective queries on-line with a concept-oriented knowledge discovery tool. In the Internet environment, users often only submit short and incomplete queries that do not clearly express their actual needs (Spink et al., 2002). Therefore, an important issue in web knowledge mining is to improve search results by assisting users to express their information needs accurately and completely.

In order to achieve the above objectives, the following important issues must be dealt with for web-based knowledge searching engines: a) *Query Formulation:* An on-line search is preprocessed by a cognitive process to represent and formulate a query. In most information retrieval systems, this process is supposed to be an external activity and is not supported by the system. b) *Query Refinement*: When a primary query is formed with clearly identified domain, type, and attributes in an existing knowledge network, an accurate query refining process is needed to help users to efficiently formulate the query. c) *Query Expression:* There are a great variety of expression structures between the query initiator and the on-line information systems. Therefore, query expression is an important process in knowledge retrieval systems to transfer information between two heterogeneous information forms: the concept networks in the brain and the indexed databases in the web. It is the key for query expressing to effectively reduce the information leak in the transformation process from internal cognitive expressions to external formulated expressions.

A wide variety of techniques have been proposed to assist users to express a search request. Among them, an important method is *query expansion*, which adds relevant query terms to an initial query in order to improve retrieval results (Shaoira & Meirav, 2005; Na et al., 2005). *Query limitation* is another query-improvement strategy (Na et al., 2005) opposite to query expansion, where users are provided with options to limit their search in order to receive more focused results. These methods have not got satisfactory effectiveness due to uncompleted consideration of all crucial features in query formulations.

This paper presents a web knowledge discovery and acquisition engine on the basis of a denotational mathematics known as concept algebra (Wang, 2006b, 2008a, 2008c). A formal concept-driven methodology is adopted in information restructuring for web documents. Knowledge organizations and representations are modeled by concept algebra, which represents a two-level normalized semantic space that simulates the cognitive knowledge representation inside the brain. At the lower level, concepts are formalized by a 5-tuple in concept algebra with a set of algebraic concept manipulation rules. At the higher level, knowledge is formally modeled by concept networks with nine concept associations. The web knowledge discovery engine encompasses four coherent components known as the concept network explorer, the semantic analyzer, the conceptual query editor, and the XML query generator. The concept network explorer provides a visual thinking navigator for assisting users to locate, capture, and refine a query efficiently. A graphical interface of the knowledge query engine is developed to facilitate direct expression and refinement of queries. The computer-aided knowledge retrieval system generates refined queries that best fit not only users' requirements, but also rational knowledge structures of existing information systems based on concept algebra. An information restructuring model is designed to decode and map informal texts in web documents into structured concept network represented by a concept graph. Applying WorldNet, ConceptNet, and other domain ontology, a concept-based clustering method that considers semantic relations and dependencies are proposed to index the restructured information of on-line documents.

Related Content

Voice-Based Image Captioning System for Assisting Visually Impaired People Using Neural Networks

Nivedita M., AsnathVictyPhamila Y., Umashankar Kumaravelanand Karthikeyan N. (2023). *Principles and Applications of Socio-Cognitive and Affective Computing (pp. 177-199).*

www.irma-international.org/chapter/voice-based-image-captioning-system-for-assisting-visually-impaired-people-using-neural-networks/314320

Broad Autoencoder Features Learning for Classification Problem

Ting Wang, Wing W. Y. Ng, Wendi Liand Sam Kwong (2021). *International Journal of Cognitive Informatics and Natural Intelligence (pp. 1-15).*

www.irma-international.org/article/broad-autoencoder-features-learning-for-classification-problem/272860

Seeing for Knowing: The Thomas Effect and Computational Science

Jordi Vallverdú (2010). *Thinking Machines and the Philosophy of Computer Science: Concepts and Principles (pp. 280-293).*

www.irma-international.org/chapter/seeing-knowing-thomas-effect-computational/43703

Intelligent Medical Data Analytics Using Classifiers and Clusters in Machine Learning

Muthukumaran V., Satheesh Kumar S., Rose Bindu Joseph, Vinoth Kumar V.and Akshay K. Uday (2021). *Handbook of Research on Innovations and Applications of AI, IoT, and Cognitive Technologies (pp. 321-335).*

www.irma-international.org/chapter/intelligent-medical-data-analytics-using-classifiers-and-clusters-in-machine-learning/285697

Voice and Space: Agency of the Acousmêtre in Spatial Design

Richard Coyne (2009). *Exploration of Space, Technology, and Spatiality: Interdisciplinary Perspectives (pp. 102-112).*

www.irma-international.org/chapter/voice-space-agency-acousmêtre-spatial/18679