

Chapter IV

Search Engine: Approaches and Performance

Eliana Campi

University of Salento, Italy

Gianluca Lorenzo

University of Salento, Italy

ABSTRACT

This chapter presents technologies and approaches for information retrieval in a knowledge base. We intend to show that the use of ontology for domain representation and knowledge search offers a more efficient approach for knowledge management. This approach focuses on the meaning of the word, thus becoming an important element in the building of the Semantic Web. The search based on both keywords and ontology allows more effective information retrieval exploiting the Semantic of the information in a variety of data. We present a method for taxonomy building, annotating, and searching documents with taxonomy concepts. We also describe our experience in the creation of an informal taxonomy, the automatic classification, and the validation of search results with traditional measures, such as precision, recall and f-measure.

INTRODUCTION

The rapid increase of the information amount in an organization or in a distributed environment in general, makes difficult to find, organise, classify, access, and use this information. Therefore, as the number of documents used by an organiza-

tion increases, the need to classify them into an intuitive and meaningful hierarchical structure becomes a challenge in order to facilitate the retrieval and the use of information.

The main problem in information retrieval is to provide the user with a concise and significant set of documents as a result for a specific query

Search Engine

on a large set of information. The result depends on the terms the user specifies in the query and the quality of the result depends on the quality of the query.

Relevance of particular information is subjective to the individual, and the context of this information is the determining factor. If the search is bound to a particular context, it produces a poorer but more precise and effective set of results.

A survey of Delphy Group demonstrates that the categories and the hierarchical structure of information are able to narrow the search area and find relevant information faster (Delphy Group, 2002).

Therefore, a major challenge is how to create a hierarchy without examining each document, especially when there are thousands of documents to classify, and how to execute a semantic indexing of the information to achieve effectiveness in the information retrieval. We adopted an approach based on Semantic Web technology in the area of knowledge management.

This chapter describes the various methodologies to create taxonomy and fill it with documents, through indexing of information sources. We describe how to create “informal taxonomy,” classify documents into taxonomy categories, and evaluate performances of this classification. We show how taxonomy represents an efficient way to organize a specific domain of an organization. In particular, we present our methods related to KIWI (Knowledge-based Innovation for the Web Infrastructure) project and applied in the Virtual eBMS platform (Virtual platform of eBusiness Management school). Then, existing technologies for searching documents in a knowledge repository are considered and the results of these different technologies are evaluated through three measures of performance: precision, recall, F-measure.

KNOWLEDGE MANAGEMENT IN THE KIWI PROJECT

The KIWI project aims at representing and managing organizational knowledge using technologies and methodologies based on Semantic technologies. Taxonomy and/or more complex structures, like ontologies, are key elements to structure the knowledge base of an organization. The KIWI project develops a set of open source tools, which allow for a formal domain representation and enable creative cooperation among users and interaction with the knowledge base:

- OntoMaker, an ontology editor;
- OntoMeter, a tool to evaluate the quality of a formal representation;
- OntoAssistant, a tool for semiautomatic document annotations;
- Semantic Navigator, a tool to browse ontology and knowledge base.

We have proposed a methodology to provide a more effective automatic document classification process. In particular, the methodology comprises four steps:

1. Developing a conceptual structure (an informal taxonomy) that presents the knowledge domain of the eBMS-ISUFI;
2. Defining rules for each category through positive and negative training set;
3. Filling the hierarchical structure with documents through semi-automatic classification;
4. Testing the quality of the classification rules, through precision, recall and F-measure, refining of the classification rules and periodic taxonomy maintenance.

This methodology has been applied to the Virtual eBMS knowledge management system.

25 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/search-engine-approaches-performance/28812

Related Content

Benefits of the Transics Fleet Management System

Petre Iltchev, Anna Palczewska, Klaudia Pilichowska, Remigiusz Kozlowski and Micha Marczak (2017). *Handbook of Research on Information Management for Effective Logistics and Supply Chains* (pp. 287-302). www.irma-international.org/chapter/benefits-of-the-transics-fleet-management-system/166814

Strategic Knowledge Management in Food Crisis Communication: A Framework for Organizational Preparedness and Consumer Trust

Maria Kalyvaki and D. Q. Spencer (2025). *International Journal of Knowledge Management* (pp. 1-10). www.irma-international.org/article/strategic-knowledge-management-in-food-crisis-communication/380734

Social Networks and Organizational Performance: Exploring the Quality of Domain Knowledge Sources

Pamela Schmidt, Sharath Sasidharan and Ronald Freeze (2013). *International Journal of Knowledge Management* (pp. 47-64). www.irma-international.org/article/social-networks-and-organizational-performance/99643

Automation vs. Human Intervention: Is There any Room Left for the Analyst in the Data Mining Process?

Meryem Sevinc, Lawrence Locker and John D. Murray (2009). *Handbook of Research on Knowledge-Intensive Organizations* (pp. 414-424). www.irma-international.org/chapter/automation-human-intervention/20866

The Role of Expected Reciprocity in Knowledge Sharing

Megan L. Endres and Sanjib Chowdhury (2013). *International Journal of Knowledge Management* (pp. 1-19). www.irma-international.org/article/the-role-of-expected-reciprocity-in-knowledge-sharing/83609