

Chapter III

OntoExtractor: A Tool for Semi-Automatic Generation and Maintenance of Taxonomies from Semi-Structured Documents

Marcello Leida
University of Milan, Italy

ABSTRACT

This chapter introduces OntoExtractor, a tool for the semi-automatic generation of the taxonomy from a set of documents or data sources. The tool generates the taxonomy in a bottom-up fashion. Starting from structural analysis of the documents, it produces a set of clusters, which can be refined by a further grouping created by content analysis. Metadata describing the content of each cluster is automatically generated and analysed by the tool for producing the final taxonomy. A simulation of a tool, based on an implicit and explicit voting mechanism, for the maintenance of the taxonomy is also described. The author depicts a system that can be used to generate the taxonomy from a heterogeneous source of information, using wrappers for converting the original format of the document to a structured one. This way, OntoExtractor can virtually generate the taxonomy from any source of information just adding the proper wrapper. Moreover, the trust mechanism allows a reliable method for maintaining the taxonomy and for overcoming the unavoidable generation of wrong classes in the taxonomy.

INTRODUCTION

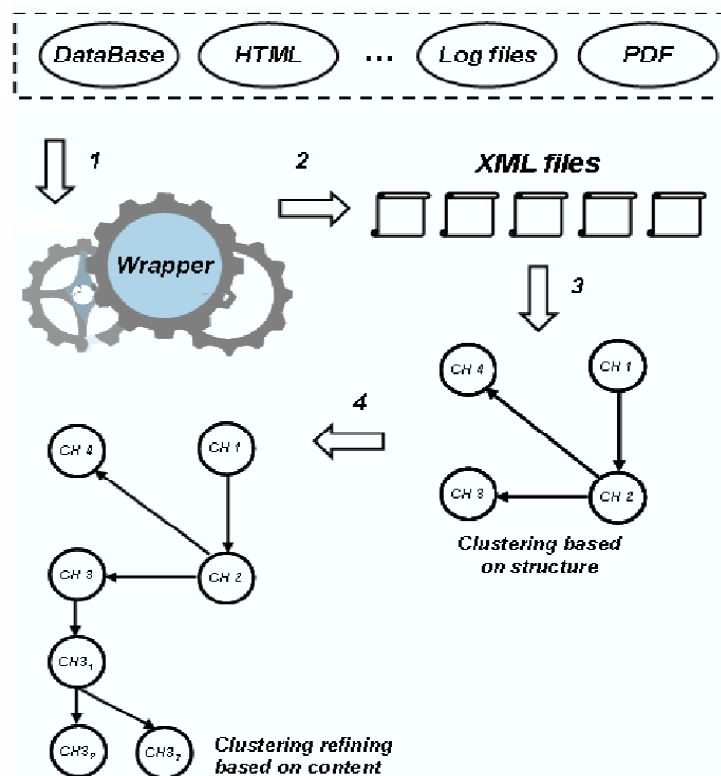
Nowadays, people can easily access virtually infinite sources of information. This unlimited supply of knowledge is potentially useful in a number of

application scenarios; on the other hand, information needs to be organized and structured to become a robust and trusted source of knowledge. *CoP (Communities of Practise)* (Lave et al., 1991) are forming spontaneously, for example, among

people sharing same interests, or according to an organization model, for instance among people working on the same project. Members of those communities interact and share information. The need for an organized and common structure that describes shared information is then evident. In the last few years, business ontology has been recognized as the most promising way to describe shared knowledge in a business environment. Shared information often can be redundant, incomplete, or subject to different interpretations. Therefore, we must be able to deal with different levels of uncertainty. Recent studies (Fagin, 1998; Fagin, 1999, & Fagin, 2002) propose using Soft Computer Techniques as a promising approach to handle uncertainty. *The Knowledge*

Management Group of the University of Milan, during the collaboration with BT Intelligent Systems Research Centre, has developed and, in part, implemented a fuzzy-based approach to extract metadata and description knowledge from heterogeneous information sources. Moreover, a fuzzy-based trust system for the maintenance of the generated hierarchy has been proposed. The following sections briefly present the theories behind this approach and the demo software developed on the basis of those theories. Section 2 is a general overview of the construction of the hierarchy; the original approach is intended to build classes of ontology from the knowledge base in a bottom-up fashion. The section is focused on presenting the techniques implemented in the *On-*

Figure 1. Overall structure of the process



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/ont extractor-tool-semi-automatic-generation/28811

Related Content

A Closer Look at Concept Map Collaborative Creation in Product Lifecycle Management

Daniela Oliveira, Mickael Gardoniand Kimiz Dalkir (2021). *Handbook of Research on Organizational Culture Strategies for Effective Knowledge Management and Performance* (pp. 260-282).

www.irma-international.org/chapter/a-closer-look-at-concept-map-collaborative-creation-in-product-lifecycle-management/286319

Using Corporate Universities to Failitate Knowledge Transfer and Achieve Competitive Advantage: An Exploratory Model Based on Media Richness and Type of Knowledge to be Transferred

M. Suzanne Clinton, Kimberly L. Merrittand Samantha R. Murray (2009). *International Journal of Knowledge Management* (pp. 43-59).

www.irma-international.org/article/using-corporate-universities-failitate-knowledge/37414

Social Software Support for Collaborative Innovation Development within Organizations

Michael Reinhardt, Martin Wiener, Marc René Frieß, Georg Grohand Michael Amberg (2012). *International Journal of Knowledge-Based Organizations* (pp. 56-76).

www.irma-international.org/article/social-software-support-collaborative-innovation/61428

Expert Knowledge in the University-Industry Cooperation: The Cases of Germany and Russia

Oxana Karnaukhovaand Oliver Hinkelbein (2018). *Social Media for Knowledge Management Applications in Modern Organizations* (pp. 218-236).

www.irma-international.org/chapter/expert-knowledge-in-the-university-industry-cooperation/185905

Knowledge Management in Hospitals

Kevin C. Desouza (2008). *Knowledge Management: Concepts, Methodologies, Tools, and Applications* (pp. 2191-2204).

www.irma-international.org/chapter/knowledge-management-hospitals/25252