Tytuł

Autorzy

Afiliacje

Dane korespondencyjne

Abstrakt

Retrieval and Management of Scientific Information from Heterogeneous Sources*

Piotr Gawrysiak, Dominik Ryżko, Przemysław Więch, and Marek Kozłowski

Institute of Computer Science, Warsaw University of Technology,
Nowowiejska 15/19, 00-655 Warsaw, Poland
{p.gawrysiak,d.ryzko,pwiech,m.kozlowski}@ii.pw.edu.pl

Abstract. The paper describes the process of automated retrieval and management of scientific information from various sources including the Internet. Application of semantic methods in different phases of the process is described. The system envisaged in the project is a scientific digital library, with automated retrieval and hosting capabilities. An overall architecture for the system is proposed.

1 Introduction

Rapid advancements in computing and networking technology, that took place during the last two decades, transformed deaply the nature of scientific research worldwide. Nowadays, it is difficult to even imagine conducting a successful research project – both in humanities and in engineering – without exploiting vast knowledge resources provided by the global Internet, and without using the same network to disseminate research results.

The nature of contemporary Internet, used as a research tool, is however drastically different from what was envisioned in the 90-ties. The Internet is just a haphazard collection of non-coordinated knowledge sources. Most valuable repositories are not even centrally controlled. It is sometimes very difficult to evaluate the quality of data contained in non-professional sources, such as some Open Access journals [11]. The situation described above basically means that the concept of Semantic Web [9], promising the coordinated global network of information, failed to materialize. One of the primary reasons for this failure is the difficulty of creating and maintaining useful ontologies, describing all aspects of the world, that would drive exchange of information in the Semantic Web [6], and only such approach would fulfill the needs of a general purpose semangic search engine. The main reason for this is a state of ontology engineering, which is still mostly a manual process, very time-consuming, expensive and error prone. While some automated, or at least semi-automated, ontology building methods that are able to leverage the amount of information present in ever growing repositories of text

* This work is supported by the National Centre for Research and Development (NCBiR) under Grant No. SP/I/1/77065/10 by the strategic scientific research and experimental development program: „Interdisciplinary System for Interactive Scientific and Scientific-Technical Information".

Informacje bilbiograficzne

Wydawca

R. Bembenik et al. (Eds.): Intelligent Tools for Building a Scient. Info. Plat., SCI 390, pp. 37–48.
springerlink.com