

An Ontology-Driven Approach for Semantic Information Retrieval on the Web

ANTONIO M. RINALDI
University of Napoli Federico II

The concept of relevance is a hot topic in the information retrieval process. In recent years the extreme growth of digital documents brought to light the need for novel approaches and more efficient techniques to improve the accuracy of IR systems to take into account real users' information needs. In this article we propose a novel metric to measure the semantic relatedness between words. Our approach is based on ontologies represented using a general knowledge base for dynamically building a semantic network. This network is based on linguistic properties and it is combined with our metric to create a measure of semantic relatedness. In this way we obtain an efficient strategy to rank digital documents from the Internet according to the user's interest domain. The proposed methods, metrics, and techniques are implemented in a system for information retrieval on the Web. Experiments are performed on a test set built using a directory service having information about analyzed documents. The obtained results compared to other similar systems show an effective improvement.

Categories and Subject Descriptors: H.3.1 **[Information Storage and Retrieval]**: Content Analysis and Indexing—*Dictionaries, linguistic processing, thesauruses*; H.3.3 **[Information Storage and Retrieval]**: Information Search and Retrieval—*Information filtering, query formulation, retrieval models, search process, selection process*; I.7.5 **[Document and Text Processing]**: Document Capture—*Document analysis*

General Terms: Algorithms, Design, Experimentation, Performance

Additional Key Words and Phrases: Ontologies, semantic relatedness metrics, WordNet

ACM Reference Format:

Rinaldi, A. M. 2009. An ontology-driven approach for semantic information retrieval on the Web. *ACM Trans. Internet Technol.* 9, 3, Article 10 (July 2009), 24 pages.
DOI = 10.1145/1552291.1552293 <http://doi.acm.org/10.1145/1552291.1552293>

1. INTRODUCTION

The production of digital contents is currently one of the most rapidly growing processes in the information age. This implies the creation of a plethora of information with related problems in organizing, managing, and searching in

Author's address: DIS-Dipartimento di Informatica e Sistemistica, Via Claudio, 21-80125, Napoli, IT; email: amrinald@unina.it.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
© 2009 ACM 1533-5399/2009/07-ART10 \$10.00

DOI 10.1145/1552291.1552293 <http://doi.acm.org/10.1145/1552291.1552293>