Chapter XIII

Using Dynamically Acquired Background Knowledge for Information Extraction and Intelligent Search

Samhaa R. El-Beltagy,
Ministry of Agriculture and Land Reclamation, Egypt

Ahmed Rafea,
Ministry of Agriculture and Land Reclamation, Egypt

Yasser Abdelhamid,
Ministry of Agriculture and Land Reclamation, Egypt

ABSTRACT

This chapter presents a simple framework for extracting information found in publications or documents that are issued in large volumes and which cover similar concepts or issues within a given domain. The general aim of the work described is to present a model for automatically augmenting segments of these documents with metadata, using dynamically acquired background domain knowledge to help users easily locate information within these documents through a structured front end. To
realize this goal, both document structure and dynamically acquired background knowledge are utilized. A real life example where these ideas have been applied is also presented.

INTRODUCTION

This work is motivated by the fact that enterprises and organizations often contain information rich texts, but they rarely have the means by which these resources can be intelligently searched. In many cases, the search interface that is adopted is based on keywords and, though the indexing/matching techniques employed by those search engines may be very sophisticated, this approach suffers from the same limitations associated with the existing web search model (see El-Beltagy, 2000; Han & Chang, 2002).

This chapter addresses the particular problem of trying to extract information from organizational publications that are issued in large volumes and which cover similar concepts or issues and from which information cannot be extracted through the use of the structure of a document alone. The end goal is to enable individual sections of those documents to be automatically augmented with metadata so that users can perform structured searches using a predefined set of categories or classifications and obtain, as a result, only segments or sections of documents that fit their search criteria. The class of documents targeted by this work is, thus, that of resources that contain a set of information entities, most of which fall under known categories, but which contain no special markup to differentiate them from other information entities. The approach adopted toward this problem is to attempt to make use of background knowledge about those categories and to employ that background knowledge for an intelligent search. Rather than forcing predefined static background knowledge, the work presented allows for the dynamic acquisition of this knowledge as the system evolves.

Our goal is, thus, twofold: first, to provide the tools that can assist in ontology building and to utilize the background ontology for document indexing; and second, to provide an intelligent interface to allow for the retrieval of the stored information.

BACKGROUND

Information is a vital resource to individuals and organizations; its timely location can influence key decisions that affect both. It is, thus, no wonder that massive research efforts have been undertaken in recent years with the aim of
Related Content

Automatic Schema-Independent Linked Data Instance Matching System
www.igi-global.com/chapter/automatic-schema-independent-linked-data-instance-matching-system/198608?camid=4v1a

Foundation of Keyword Search in XML
www.igi-global.com/chapter/foundation-keyword-search-xml/69124?camid=4v1a

Agent Negotiation in Water Policy Planning
www.igi-global.com/chapter/agent-negotiation-water-policy-planning/73772?camid=4v1a
Interactive IR in Web Search Engine Environments
www.igi-global.com/chapter/interactive-web-search-engine-environments/24525?camid=4v1a