Conceptual Graphs and Ontologies for Information Retrieval

Catherine Comparot, Ollivier Haemmerlé, and Nathalie Hernandez

IRIT, Université de Toulouse le Mirail, Département de
Mathématiques-Informatique, 5 allées Antonio Machado, F-31058 Toulouse Cedex
{catherine.comparot,ollivier.haemmerle,nathalie.hernandez}@univ-tlse2.fr

Abstract. We propose a mechanism for annotating and querying document collections based on the semantic modeling of the context of a search. We model on the one hand the topics concerned by the content of the document and on the other hand the metadata associated with the documents, by means of two ontologies expressed in the conceptual graph model. The semantic annotating mechanism is done by automatically building conceptual graphs.

1 Introduction

Many works in Information Retrieval (IR) aim at enhancing the classic method of document indexation with keywords by an indexation based on semantic annotation. In accordance with those which add an annotation layer based on ontologies [1] to describe the meaning of the documents, we model documents with two ontologies, a topic one and a documentary one. We propose to represent this knowledge in the conceptual graph formalism [2]. This knowledge representation model is well suited for our application for several reasons: (i) the query algorithms have been widely studied and are well suited for IR; (ii) this graphical model allows non-computer scientists to express more sophisticated queries than a simple conjunction of keywords, without the complexity of usual query languages.

Our work takes place in the French WebContent project [3], which aims at creating a software platform to accommodate the tools necessary to exploit the Semantic Web. The input of our system are news releases which are provided by news agencies. They are composed of a set of metadata and a body expressed in free text. Our goal consists in annotating each of these releases in order to enable an Information Retrieval process on that corpus.

In this article, we present successively the ontologies and their representation and the annotation mechanism.

2 An Ontology Based Model

Topic and documentary ontologies. Our system is based on two ontologies, the topic ontology and the documentary ontology. In our case study, we consider a
Fig. 1. An example of documentary model

The documentary model. In our model, the documentary model organizes the metadata embedded in the corpus documents. It is represented by means of the conceptual graph $M_{Doc}$. It is represented in Fig. 1.

A set of conceptual graphs $M_{Doc}^*$ is associated with the documentary model. These conceptual graphs span the graph $M_{Doc}$. They represent the elementary pieces of information which can be returned during the semantic annotation step of the documents.

3 Semantic Annotation of the Documents

Each document is annotated by a conceptual graph which describes its content and its metadata. This conceptual graph is built by aggregating elementary conceptual graphs called motifs which are returned by the text analysis tools we