

# Ontology-Based Legal Information Retrieval to Improve the Information Access in e-Government

Asunción Gómez-Pérez

Facultad de Informática. Universidad  
Politécnica de Madrid  
Campus Montegancedo, s/n 28860  
Boadilla del Monte. Madrid. Spain  
34-913367439

asun@fi.upm.es

Fernando Ortiz-Rodríguez

Facultad de Informática. Universidad  
Politécnica de Madrid  
Campus Montegancedo, s/n 28860  
Boadilla del Monte. Madrid. Spain  
34-913363670

fortiz@fi.upm.es

Boris Villazón-Terrazas

Facultad de Informática. Universidad  
Politécnica de Madrid  
Campus Montegancedo, s/n 28860  
Boadilla del Monte. Madrid. Spain  
34-913363670

bvillazon@fi.upm.es

## ABSTRACT

In this paper, we present EgoIR, an approach for retrieving legal information based on ontologies; this approach has been developed with Legal Ontologies to be deployed within the e-government context.

## Categories and Subject Descriptors

H.3.3 Information Storage and Retrieval: Information Search and Retrieval – *query formulation, retrieval models, search process.*

## General Terms

Design, Experimentation

## Keywords

Ontology, Information Retrieval

## 1. INTRODUCTION

For more than two decades, the AI and Law community has been very active and productive. In the early 80's, research was focused on logic programming. Other approach adopted was the case-based reasoning. Knowledge Engineering was also of interest for the research community and the field most applied since it allowed developing and using the legal ontologies that underlie the growth of the Semantic Web.

The e-Gov has been strengthened with all these previous studies carried out by the research community and now its main concern is data representation and information management. By its nature, the e-Gov is supported by the legal domain.

Our contribution consists of an ontology based approach for legal information retrieval that we called EgoIR. This system has as a main goal to retrieve e-Gov documentation. EgoIR deals with Real-estate transaction documents, and gives an opportunity to the citizens, business and governments to integrate and recover documents. For this purpose EgoIR provides facilities for managing, searching and sharing e-Gov documentation.

## 2. EgoIR

EgoIR is an Ontology-Based Legal Information Retrieval System. This system is the result of integrating Ontological Workbench

WebODE<sup>1</sup>, and a text search engine library, Lucene<sup>2</sup>. In this section we describe the system architecture and the Legal Ontologies.

## 2.1 Architecture

The system integration of the EgoIR is built and composed by the Search Client, the Search Server and the Ontology Server modules, which are described in the next subsections. Figure 1 shows the general architecture of the system.

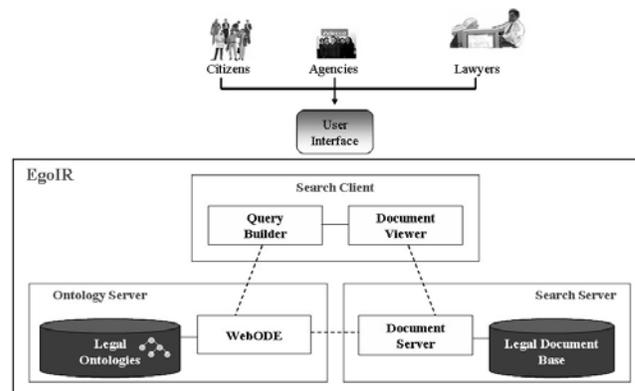


Figure 1. EgoIR System Architecture.

### 2.1.1 Ontology Server

This module defines how the knowledge is structured in the application domain. This module includes the Legal Ontologies within WebODE [2].

Within the Legal Ontologies, concept instances are associated with documents. Every time that a new concept instance is added the Ontology Server communicates with the Search Server to index its corresponding document.

### 2.1.2 Search Client

This module incorporates two sub-modules: a Query Builder and a Document Viewer.

Query Builder connects to the Ontology Server, in order to access Legal Ontologies, browse them and obtain concepts to build the query by using a graphical interface. This module sends the query to the Search Server.

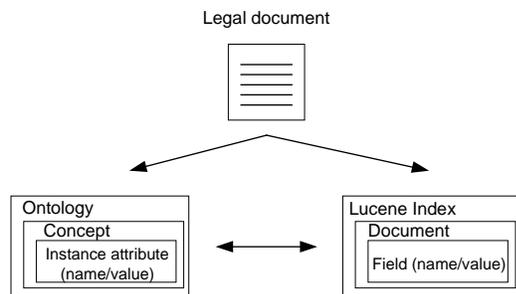
Demos and Posters of the 3rd European Semantic Web Conference (ESWC 2006), Budva, Montenegro, 11th - 14th June, 2006

<sup>1</sup> <http://webode.dia.fi.upm.es/>

<sup>2</sup> <http://jakarta.apache.org/lucene>

Document Viewer connects to the Search Server, in order to retrieve the legal documents satisfying the query, and to the Ontology Server to browse and display the documents.

This module is also a procedural mapping module that makes possible the interoperability between WebODE and Lucene. Figure 2 shows the relationships between ontology and Lucene index.



**Figure 2. Relationships between Ontology and Lucene Index**

### 2.1.3 Search Server

The Search Server module is based on Lucene and processes the Legal Document Base to create internally access structures. These structures (called indices) allow fast document location and are stored locally in the file system of the operating system.

The Legal Document Base consists of electronic documents that are stored in the file system. These electronic documents are: juridical term glossary, models of contracts, legal norms and jurisprudence. Currently the document's annotation process is manually done. When a concept instance is created, using WebODE interface, the values from its instance attributes are indexed using Lucene which includes the electronic document.

This module also performs a kind of mapping between ontology instances and indexed documents. We can see in Figure 3 that the query concepts with their syno-nyms are sent to the Search Server and it searches on the Lucene index, more specifically inside Legal Document that represents a Generic Document Concept. The Legal Document has inside the content field that represents an instance attribute that holds the document's content.

## 2.2 Legal Ontologies

Legal Ontologies [3] were built to represent the real-state transactions in the Spanish Government domain. These Legal Ontologies were developed with knowledge acquired by experts from academic and private sectors and built with the methodology METHONTOLOGY [2] and the workbench WebODE [2].

For the EgoIR system eleven ontologies have been developed: *person, civil personality, organization, location, tax, contract model, jurisprudence, Real-estate transaction verifications, Real-estate, legislation, and Real-estate transaction.*

## 3. RELATED WORK

There are many systems developed for managing legal information, but only a few deals with legal knowledge. In this section we describe briefly some legal IR systems.

In [5], CLIME (Computerized Legal Information Management and Explanation) aims at improving the access and understanding of large collections of legal information through the Internet. CLIME just combines conventional IR with artificial legal reasoning without ontologies. In [1], the authors describe the

Webocrat system whose goal is to provide new types of communication and service flows from public institutions toward citizens, thus improving the access of citizens to services and information of public administration. This system focuses on security issues.

Another work reported in [4] is the EULEGIS (European User Views to Legislative Information in Structured Form), whose main goal is to provide a consistent user interface for legal IR generated in different legal systems and at different legislative levels. This system focuses on user interfaces.

## 4. CONCLUSIONS AND FUTURE WORK

In this paper we present our first approach to an ontology-based legal IR, which aims to retrieve government documents in a timely and accurate way. This is an approach of an entirely new wave of legal knowledge systems. At this time we can mention that the utility of ontologies within an IR is twofold: On the one hand, as a social impact, ontologies are a good way to guide user to the legal terms, thus avoiding him/her to make mistakes at the query construction; and on the other hand, mostly technical, ontologies are a key to the development the Semantic Web and improving interoperability on the legal applications.

Finally, in the near future we will improve the performance of EgoIR and we will focus on further enhancement of the ontology-based retrieval mechanism by means of Natural Language Processing (NLP) techniques for an user friendlier environment; on the automatic semantic annotation of the documents to improve the search process; and on security issues by providing a summary of the retrieved documents.

## 5. ACKNOWLEDGMENTS

We would like to thanks to Rosario Plaza and Ángel López-Cima. This work is carried out within the ongoing Reimdoc Project which is supported by the Spanish Technology and Science Ministry (Project FIT-340100-2004-22).

This project is partially funded by a scholarship granted through the Program of Teaching Staff Improvement (PROMEP) at the Tamaulipas University.

## 6. REFERENCES

- [1] Dridi, F., Pernul, G.: The Webocracy project: Overview and Security Aspects. In Schnurr et al., *Professionelles Wissensmanagement: Erfahrungen und Visionen*. Shaker Verlag, Aachen, 2001.
- [2] Gómez-Pérez, A., Fernández-López, M., Corcho, O.: *Ontological Engineering*. Springer Verlag, 2003
- [3] Gómez-Pérez, A., Ortiz-Rodríguez, F., Villazón-Terrazas, B.: *Legal Ontologies for the Spanish e-Government*. In *Proceedings of the 11th Conference of the Spanish Association for Artificial Intelligence*. Santiago de Compostela, Spain, 2005.
- [4] Lyytikinen, V., Tiitinen, P., Salminen, A.: *Challenges for European Legal Information Retrieval*. *Proceedings of the IFIP 8.5 Working Conference on Advances in Electronic Government*. Zaragoza, Spain, 2000.
- [5] Winkels, R.G.F.: *CLIME: Legal Information Serving Put to the Test*. Pre-proceedings of the Second French-American Conference on AI and LAW. Available online at: <http://www.Iri.jur.uva.nl/~winkels/papers/AIL98-paper.html>