Chapter 5
Access to terminological ontologies

5.1 Introduction

Typical information infrastructures manage different terminological ontologies with a very different levels of specificity, language coverage integrated in their systems (e.g., from monolingual lists of terms to multilingual thesauri covering more than 20 languages), formalization, (e.g., from simple glossaries to well-structured thesauri), or size, (e.g., AGROVOC thesaurus [126] contains more than 16,000 concepts).

Each of these different ontology models have been traditionally stored, managed, used and updated independently from the rest of them. This lack of coordinated management makes very difficult to determine which models are used in which service and with which purpose. Additionally, this lack of coordination leads to the replication of the terminological ontologies along the infrastructure. Ensuring that all the copies of each ontology are identical requires high maintenance effort. Moreover, the lack of coordination makes the upgrade process very prone to errors. For example, it is very easy to forget updating one of the copies; and if this happen, the service using the obsolete terminology would produce deficient results. Additionally, the lack of knowledge of the used ontologies leads to a dispersion in their use. For instance, different but similar terminological ontologies are used for situations where a single one would be a better selection.

When ontologies are provided to the public by their creators, they are distributed through ad-hoc services developed by the institution providing each ontology. However, in systems which use a great amount of them, it is not viable to access a large amount of different incompatible services to retrieve the needed vocabularies. It is needed to collect all the required models obtained from external and internal sources and provide them homogeneously, using a single inner model, storage system, and access procedures.

Section 5.2 reviews the existent works in this area to determine if the proposed solutions cover all the required management needs. This study has shown that existent solutions tackle specific management problems independently of the rest. In this context, with the objective of managing terminological ontologies in a homoge-
neous way, section 5.3 describes the general architecture that has been proposed for the management of terminological ontologies. Each one of the following sections describes in detail a specific component of this architecture. Firstly, section 5.3.2 focuses on the need of a common repository to store all the required terminological ontologies and it makes a proposal for such a repository. Secondly, at management level, section 5.4 describes the management needs for terminological models and it proposes a solution for an efficient management. Finally, section 5.5 focuses on access issues and it describes the requirements and structure of a Web service to facilitate the distribution of the most suitable terminological models to the components of an information infrastructure.

The components described in this management architecture have been implemented and tested to verify their suitability for working with terminological ontologies. Section 5.6 shows the experiments performed and the results obtained.

5.2 Terminological ontology management

Because of the variety and large number of terminological models used in an information infrastructure, the harmonization of their management is a priority. The interest in creating terminological ontologies in the digital libraries field and other related disciplines has led to an increasing number of software packages for the construction of different types of terminological models. The web site of Willpower Information\(^1\) offers a detailed analysis of more than 40 tools, most of them designed for thesauri edition. Some tools are only available as a module of a complete information storage and retrieval system, but others also allow the possibility of working independently of any other software. Among these creation tools, one may highlight the following products:

- **BiblioTech\(^2\)**. This is a multi-platform tool that forms part of BiblioTech PRO Integrated Library System and can be used to build an ANSI/NISO standard thesaurus \([9]\).
- **Lexico\(^3\)**. This is a Java-based tool that can be accessed and/or manipulated over the Internet. Thesauri are saved in a text-based format. It has been used by the U.S. Library of Congress to manage vocabularies and thesauri such as the “Thesaurus for Graphic Materials”, the “Global Legal Information Network Thesaurus”, the “Legislative Indexing Vocabulary”, and the “Symbols of American Libraries Listing”.
- **MultiTes\(^4\)**. This is a windows based tool that provides support for ANSI/NISO relationships plus user defined relationships and comment fields for an unlimited number of thesauri (both monolingual and multilingual).

---

\(^1\) http://www.willpowerinfo.co.uk/thessoft.htm
\(^2\) http://www.inmagic.com/
\(^3\) http://www.pmei.com/lexico.html
\(^4\) http://www.multites.com/