Knowledge Representation Issues and Implementation of Lexical Data Bases

F. Sáenz and A. Vaquero

Departamento de Sistemas Informáticos y Programación, Universidad Complutense de Madrid,
E-28040 Madrid, Spain
Tel: +34913947622
Fax: +34913947529
{fernan,vaquero}@sip.ucm.es

Abstract. We propose to apply classical development methodologies to the design and implementation of Lexical Databases(LDB), which embody conceptual and linguistic knowledge. We represent the conceptual knowledge as an ontology, and the linguistic knowledge, which depends on each language, in lexicons. Our approach is based on a single language-independent ontology. Besides, we study some conceptual and linguistic requirements; in particular, meaning classifications in the ontology, focusing on taxonomies. We have followed a classical software development methodology for implementing lexical information systems in order to reach robust, maintainable, and integrateable relational databases (RDB) for storing the conceptual and linguistic knowledge.

1   Introduction

Due to the immaturity of the knowledge representation topic, lack of standardization is broadly felt as a very undesirable state into the community around language resources [LREC 02]. For instance, standard terminology for a common reference ontology is yet a goal to be reached. There is no doubt about what lexicon means, but ontology is differently understood in the computational linguistic literature. For instance, WordNet is mentioned as an ontology [USC 96], CYC is provided with a formal ontology [PRI 01], etc. Here, ontology, in a LDB, is the set of concepts in the domain of the base and the relationships that hold among them, without including linguistic knowledge, and common to all of the languages supported in the base.

Weak attention has been paid on topics about development methodologies for building the software systems which manage LDB, and dictionaries in particular. We claim that the software engineering methodology subject is necessary in order to develop, reuse and integrate the diverse available linguistic information resources. Really, a more or less automated incorporation of different lexical databases into a common information system, perhaps distributed, requires compatible software architectures and sound data management from the different databases to be integrated. The database subject have already done a long way reaching a strong standardization, and supplying models and methods suitable to develop robust information systems. We apply RDB design methodologies to develop LDB consisting of ontologies and