

# WebKB-2: Cooperatively-built Knowledge Bases

Philippe Martin

DSTC

Griffith University - PMB 50 GCMC, QLD 9726 Australia

philippe.martin@gu.edu.au

Peter Eklund

DSTC

Griffith University - PMB 50 GCMC, QLD 9726 Australia

p.eklund@gu.edu.au

## ABSTRACT

We describe some components of our large-scale multi-users knowledge server WebKB-2 (usable from <http://meganesia.int.gu.edu.au/~phmartin/WebKB/>). Details may found at this site or in [1] and [2].

## Keywords

Knowledge Representation/Retrieval, Ontology Server, Cooperation, RDF.

## 1. INTRODUCTION

Current Web search engines can retrieve documents that include keywords but they cannot retrieve precise information in answer to precise queries, e.g. "list the characteristics of American cars for sell under 20,000 AUD in Sydney?", "what are the 5 cheapest ways to travel from Sydney to Brisbane in March 2001?", and "which database systems can handle interactively modifiable schemas?". Answering such queries requires a large-scale cooperatively-built knowledge base (KB) because: (i) the task of producing the necessary information has to be distributed among many providers, (ii) the KB should be not be restricted to a limited domain (which KB should be searched or updated if the knowledge is scatered among many KBs; there is no reliable way to connect, merge or cross-check knowledge from separated KBs).

We developp a Web server (WebKB-2) for such large-scale cooperatively-built KBs. We expect it to be used for supporting "Yellow-Pages like catalogs" or "corporate memories". In this article, we briefly present a few of the elements that we designed to enable a large-scale cooperatively-built KB base and summarise the cooperation and search mechanisms adopted.

## 2. ELEMENTS FOR A LARGE-SCALE COOPERATIVELY-BUILT KB

A knowledge representation language is needed for representing knowledge. We designed two **notations** -- Frame-CG (FCG) and Formalized English (FE) -- to improve on the readability and expressivity of the Conceptual Graph linear notation. For instance, the sentence "According to Dr Foo, most cars have 4 wheels." may be represented by the FCG statement `[most cars, part: 4 wheels](Foo@bar.au)` or the FE statement ``most cars have for part 4 wheels'(Foo@bar.au)` -- the identifier of Dr Foo being `Foo@bar.au`. These notations encourage the user to be explicit and exact in his/her knowledge representations, and limit the number of ways an piece of information is expressed. This simplifies procedures for comparing/retrieving/cross-checking knowledge

For the same reason, we also developed **knowledge representation guidelines** for the WebKB users: lexical guidelines (e.g. use English singular nouns as category names), semantic guidelines (be precise, contextualize statements, re-use and complement existing knowledge), syntactic guidelines (e.g. how to represent various kinds of quantifiers, collections, intervals, contexts, 2nd order types/relations), ontological guidelines (e.g. how to represent states and processes, descriptions, indexations, characteristics, measures, numbers, collections, temporal/spatial/logical entities/relations).

**Each element of the KB** (categories, category names, links between categories, concept nodes and relation nodes) **is associated with an identifier of the user who created it**. This is required to support updates by multiple

users and permit each user to filter or focus on knowledge from certain users. **Each category** (concept/relation types, individuals) has a unique identifier (e.g. `wn#car`, "wn" being the creator identifier) but **may have several names** (e.g. the category `wn#car` may also be referred by `wn#car__auto__automobile__machine__motorcar`; this second way shows not only the "key name" but all the names). Conversely, a name may be shared by several categories if it has a variety of meanings. Within statements, categories may be referred via names instead of identifiers when there is no ambiguity about each referred category (either because a name refers to only one category or because the signature associated to the used relations can be exploited to reduce the possibilities). With our current KB initialized with the [WordNet](#) lexical database, the statement `[most cars, part: 4 wheels]` is ambiguous: there are 5 categories with name "car" and 6 categories with name "wheel". An unambiguous statement is `[most wn#car, part: 4 wn#wheel]` (or using abbreviations: `[most #car, part: 4 #wheel]`). When WebKB-2 parses an ambiguous statement, it rejects it but helps the user to refine it by displaying the various possible categories for each ambiguous name.

Knowledge retrieval, update and checking in WebKB is greatly supported, and guided by our **reuse of the WordNet ontology** (the 66,000 categories related to nouns) and **its insertion into an ontology of 100 top-level concept types and 140 basic relation types** signed on these top-level concept types. These relation types were designed to permit the direct representation of most natural language sentences but are also relevant for more model-oriented representations. We distinguished the WordNet *specialization* links into *subtype* links and *instance* links by manually isolating 2900 individuals. We also made a few other structural corrections. Finally, we added some categories from other ontologies, e.g. Ontolingua and the RDF basic standard schema. The categories and links are not stored in a fixed schema; they can be updated interactively. An object-oriented database called [FastDB](#) is exploited for knowledge storage.

Here is a summary of **the protocols for the cooperative edition of the KB**. Any user can reuse any category in links or statements (unless that induces a detected inconsistency), create new categories, links or statements, remove the ones s/he created and filter out the others. A user may not modify a statement that s/he has not created but s/he can connect it to another statement via a relation of type `pm#corrective_specialization`, `pm#corrective_generalization` or `pm#correction`. A user must also use these relations in order to add a statement that WebKB detect as a specialization or generalization of another statement. Thus, conflicts between users are made explicit, re-use is maximized and redundancy is prevented.

Statements can be **retrieved** via search for **specializations** of a query graph. Categories may be retrieved according to their names or connected links. Links may be **recursively explored**. In all cases, **filtering on creators** may also be applied.

### 3. REFERENCES

1. Ph. Martin. Conventions and Notations for Knowledge Representation and Retrieval. *ICCS 2000, LNAI 1867*, pp. 41-54
2. Ph. Martin., P. Eklund. Large-scale cooperatively-built KBs. *ICCS 2001 (9th Int. Conf. on Conceptual Structures)*; to appear