Three-Dimensional Representation of Conceptual Fuzzy Relations

Jose A. Olivas and Samuel Rios

Dept. of Information Technologies and Systems
University of Castilla-La Mancha
Paseo de la Universidad 4,
13071, Ciudad Real, Spain
Joseangel.olivas@uclm.es,
i92riros@uco.es

Abstract. In this work, T-DiCoR is presented (**T**hree **Di**mensional **Co**nceptual **R**epresentation) as a tool for representing the fuzzy relations among the most representative concepts of a domain. Using this tool in a Metasearcher, the user may observe what other concepts are related to the searched concept, and what the connection forces are (fuzzy relations between concepts). This knowledge can be useful for making new queries with words conceptually related in a specific domain with the original ones.

1 Introduction

Nowadays, the task of recovering information of big data sources, especially the Web, has great relevancy to many users of different areas and of the whole world. The search engines play a determinant role in this task, but increasingly, the users need help to centre the search on their aims.

In this work a tool is presented that can be useful to centre the search, since it allows showing the relations among the most relevant concepts of a certain domain. This allows the user to refine his queries, verifying what concepts are the most relevant in the domain and with which ones and how they are related.

The presented tool T-DiCoR (Three Dimensional Conceptual Representation) shows the user a three-dimensional form, a graph with the form of molecule, where the nodes are the most relevant concepts of a domain and the edges show the forces (fuzzy) that join them. These edges will be represented in different colours and thicknesses according to the intensity of the relation between the concepts. The user can change the view of the graph using the mouse.

It is very important to know how to get the input matrix that represents the fuzzy relations between the concepts. It is done with a fuzzy aggregation of different values from different sources (as is shown in point 2). In point 3 the T-DiCoR tool and the algorithms used are described. A complete example is explained in section 4, and the paper finishes with some conclusions and future trends.

H. Legind Larsen et al. (Eds.): FQAS 2006, LNAI 4027, pp. 681 – 690, 2006.

2 The Input Matrix

The input matrix contains the fuzzy relations among the 20 most important concepts of a specific domain. The weight values came from the aggregation of several different sources:

2.1 Contextual, Linguistic and Ontology Relations (FIS-CRM vectors)

FIS-CRM [1] is a model for representing the concepts contained in any kind of document. It can be considered an extension of the vector space model (VSM). Its main characteristic is that it is fed on the information stored in a fuzzy synonym dictionary and various fuzzy thematic ontologies. The dictionary stores the synonymy degree between every pair of recognized synonyms. Each ontology stores the generality degree between every word and its more general words. The way of calculating this value is the one proposed by Widyantoro & Yen in 2001 [2].

The key of this model is first to construct the base vectors of the documents considering the number of occurrences of the terms (what we call VSM vectors) and afterwards to readjust the vector weights in order to represent concept occurrences, using for this purpose the information stored in the dictionary and the ontologies.

The readjustment process involves sharing the occurrences of a concept among the synonyms which converge in the concept and give a weight to the words that represent a more general concept than the ones contained.

2.2 Causal Relations

To detect the causal relationships that exist in a collection of documents, a starting point could be to detect conditional phrases. Nevertheless, this is not an easy task. Descartes could not have possibly imagined that to propose his famous phrase “I think, therefore I am”, would have given birth to so many conjectures and interpretations for centuries after. In reality, what did he want to say, “First I think and after I am a person, or As I am capable of thought, I am a person”.

To sum up, even on this occasion the intention of Descartes seems clear when he expressed his maxim, it is not easy to interpret and format the information expressed in natural language, especially when it involves complex sentences with complicated turns.

With the aim of detecting conditional phrases, we have developed a basic system [3] of detecting structures and a classification of sentences that allows us to locate, in terms of basic components (verb tenses, adverbs, linguistic turns, etc.), certain causal forms.

To make the grammatical analysis, we have observed on the one hand, that we can separate certain causal relationships based on the verb form used, while on the other hand we can separate others based on the adverbs used in the sentences. Both analyses give rise to some causal rules that we will use afterwards to make an automatic extraction of knowledge. In the same way, every structure is subdivided into two structures which correspond to the antecedent and consequence of the causal relationship, and a parameter that measures the degree of certainty, conjecture, or compliance of