

Chapter 1

Introduction

In the field of databases and information retrieval, there is an ever increasing demand for systems able to deal with flexible queries and answers. For instance with the limited capabilities of today's search engines, when one searches for *fast car*, the results are websites over the internet with texts which include the words *fast* and *car*. The search engines are simply not proficient enough to observe from one source that a particular car has a maximum speed of 350 Km/h, and thus it should be returned as an answer of the query. Moreover they do not possess the means of detecting not exact, but closely related answers. Here when we say related, we do not mean lexical similarity, but indeed meaningful semantic likelihood. So as another simple example we might state that when queried for *red car*, no search engine lists *orange cars* as a potential point of interest. What we would want from them is to tell us that this is not the exact answer, but still one which should be taken into consideration. Under this topic, dealing with similarity relations is a subject which requires utmost attention; where a similarity relation stand for the degree of affinity between two entities of the concerning domain.

Fuzzy Logic is used to represent vague information in the real world. In order to draw the similarity between objects in real world, we introduce similarity between predicates into Fuzzy Logic. Similarity itself is a fuzzy concept, which makes fuzzy logic suitable and appropriate for representing similarity.

There are different approaches which amalgamates *Logic Programming* with concepts coming from *Fuzzy Logic*. This type of works belong to logic programming classes such as *Fuzzy Logic Programming*, *Qualified Logic Programming* and *Similarity-Based Logic Programming*.

One intriguing element of the latter class is the *Bousi~ Prolog*. Briefly we may say that *Bousi~ Prolog* belongs to the *Similarity-Based Logic Programming* class, and it replaces the syntactic unification mechanism of classical SLD-resolution by a fuzzy unification algorithm based on fuzzy binary relations on a syntactic domain. This algorithm

provides a weak most general unifier as well as a numerical value, called the *approximation degree*. The approximation degree represents the truth degree associated with the computed instance(query). The result is an operational mechanism, called Weak SLD-resolution, which differs from other approaches in some aspects, based exclusively on similarity relations. Very crudely, the framework utilizes an algorithm which firstly annotates each formula in the rule set with a truth degree equal to 1. The rest of the formulas are annotated with their corresponding approximation degree. In case that several formulas of set generate the same approximate formula, with different approximations degrees, the one with the least degree is taken as annotation. The resulting set consist the core concepts of model and logical consequence of the particular similarity relation degree which is chosen at the beginning of the method.

On the other hand, a very promising representative of the first class is the *RFuzzy* framework, a Prolog-based tool for representing and reasoning with fuzzy information. In a nutshell the advantages of the framework in comparison to other tools in the same field of research are its easy, user-friendly syntax, and its expressivity through the availability of default values and types. [?]

The similarity concept has been recently investigated in the *Rfuzzy* framework. As of today, in *Rfuzzy* framework there are two existing methods for measuring the degree of similarity between two fuzzy predicates. Namely: *Interpretation Based* and *Structure Based* methods. Crudely, the former obtains the similarity between fuzzy predicates by the comparison between their interpretations. Whereas the latter method considers the structure of fuzzy predicates, in other words the way that fuzzy rules define fuzzy concepts.[?]

Even though these approaches are sound methodologies they are not without some flaws:

- As mentioned above, Interpretation Based Method(*IBM*) is a very naive method which only utilizes the commonality of the attributes of the predicates'. It can only work on a single layer of subconcepts thus does not have the means of tackling with complex definitions.
- Unlike *IBM*, Structure Based Method(*SBM*) makes use of the structural property by constructing the predicate tree via taking the head of a fuzzy rule as the root of the tree and by branching the tree regarding the body of the rule. Here one point of concern is that, the construction of the tree is recursive but not the evaluation of the similarity degree between the predicates. As a result of this the trees need to be structurally equivalent for the algorithm to be able to execute. This hole in the algorithm is patched by introducing identity predicates when one tree is lacking internal nodes or leaf nodes compared to the other one. But as Lu also confirms in [?], this causes a "distortion" effect on the final evaluation. Especially in cases such as where the branching factors differ by a big margin or when one tree is highly unbalanced, this effect would be clearly apparent.

With these shortcomings in mind, we believe a new approach which would handle the mentioned problems would carry significant importance. Very briefly the idea is firstly defining a generic entity matching problem(whether it being terms or predicates) mathematically as a graph problem and then introduce computational methods utilizing properties of graphs, to identify structural resemblance between parts of the graphs. That would be followed by fuzzyfying the process by introducing fuzzy logic terms on this generic framework. Lastly a methodology regarding computing the confidence value of the result of the process is constructed.

In this thesis, preliminaries are displayed in Chapter 2, where lattice theory, fuzzy set/relation, Fuzzy Logic and a brief introduction of RFuzzy framework are presented. In Chapter 3 we observe the *SBM* method [?] in detail and demonstrate its shortcomings. Chapter 4 depicts the new methodology we introduce in to the field. Since this methodology computes results which are not hundred percent valid in the cases of knowledge bases with incomplete information, in Chapter 5 we propose a method for computing the credibility values of fuzzy similarity relations. With the last chapter we wrap up the research and show on going research topics. The contribution in this work fixes the existing gaps in Lu's approach [?] and defines the roots for a much more ambitious project, as mentioned getting not only the results for *red car* but the ones for *orange car*, too.