

## Chapter 6

# Exploring Fuzzy Association Rules in Semantic Network Enrichment Improvement of the Semantic Indexing Process

**Souheyl Mallat**

*Faculty of Science of Monastir, Tunisia*

**Emna Hkiri**

*LATICE Laboratory, Tunisia*

**Mounir Zrigui**

*Faculty of Science of Monastir, Tunisia*

### ABSTRACT

*In the aim of natural language processing applications improvement, we focus on statistical approach to semantic indexing for multilingual text documents based on conceptual network formalism. We propose to use this formalism as an indexing language to represent the descriptive concepts and their weighting. Our contribution is based on two steps. In the first step, we propose the extraction of index terms using the multilingual lexical resource EuroWordNet (EWN). In the second step, we pass from the representation of index terms to the representation of index concepts through conceptual network formalism. This network is generated using the EWN resource and pass by a classification step based on association rules model. Our proposed indexing approach can be applied to text documents in various languages. Next, we apply the same statistical process regardless of the language in order to extract the significant concepts and their associated weights. We prove that the proposed indexing approach provides encouraging results.*

## 1. INTRODUCTION

It is known that ambiguities of natural language, have a detrimental effect on the results of query terms translation in the context of information retrieval by crossing languages. However, research efforts to integrate sense disambiguation techniques in machine translation (MT) have not been successful and get unconvincing results. In addition, our automatic translation system (ATS) (Jianfeng et al., 2001) requires a high precision of disambiguation to achieve an effect on the selection of the best translation in the target language of ambiguous words.

The semantic disambiguation process of the query in the target language is based on a similar document language as the query. This document is a list of relevant sentences (most similar to a user query); these sentences noted List\_S are satisfying the query, and they are classified according to their degree of linguistic relevance (semantic, morphological). Building this List\_S of words is presented in the work (Mallat et al., 2013) (Mallat et al., 2014). The same lists (of French and English sentences are the result of the multilingual parallel corpus alignment. Both versions of the lists are used as resources for the disambiguation process in the queries translation (Arabic-French) and (Arabic-English). The process is to match the query and the List\_S content to find the words of the query in the target language that best fits this List\_S. A key feature of the method of disambiguation is that the degree of matching of each translation of an ambiguous word and List\_S depends on the highest weight.

Note that this List\_S is expressed by singular characteristics of specific themes such as semantic and morphological wealth that is supposed to represent the best the relevant answers to a given query. Indeed, the disambiguation process improvement requires providing an effective method for representing and better analyzing the contents of this list.

In this paper, we focus on the extraction of the concepts (descriptors) or index concepts in order to associate for each document (List\_S) a representation of its contents by concepts and their associated weights.

To do this, we focus our work to propose a statistical approach to semantic indexing of multilingual documents (French or English) that are taken only on calculations of the frequency of words. Also, we focus on exploiting taxonomic and non taxonomic relations (contextual) between terms. The proposed indexing approach consists of:

1. Extracting the significant words or index terms associated with the concepts of a document in English or French, based on the two external lexical resource (multilingual thesaurus) EuroWordNet (EWN French and English EWN) (Gonzalo et al., 1998) (Vossen et al., 1997). As we consider that a EWN is composed of a set of lexicons and a set of relations between them designated by concepts.
2. The construction and exploitation of conceptual network formalism of a document that requires the extraction of concept nodes and relations between them extracted from the previous step. In the extraction of relations, we rely on the EWN resource to identify the taxonomic relations, and we add the fuzzy association rules model to identify non taxonomic relations (contextual) between concepts. This model represents an inference mechanism to discover these latent relations, buried in List\_S and carried by the semantic context. The goal of this model is to better represent the semantic content of the document. Thus, the novelty of this model involves two aspects: (1) co-occurrence of terms is taken into account during indexing of the List\_S. The model's descriptors are no longer words but sets of index terms (term-sets). The term sets capture the intuition that

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/exploring-fuzzy-association-rules-in-semantic-network-enrichment-improvement-of-the-semantic-indexing-process/196438](http://www.igi-global.com/chapter/exploring-fuzzy-association-rules-in-semantic-network-enrichment-improvement-of-the-semantic-indexing-process/196438)

## Related Content

---

### The Graph Signature: A Scalable Query Optimization Index for RDF Graph Databases Using Bisimulation and Trace Equivalence Summarization

Mustafa Jarrar and Anton Deik (2015). *International Journal on Semantic Web and Information Systems* (pp. 36-65).

[www.irma-international.org/article/the-graph-signature/136833](http://www.irma-international.org/article/the-graph-signature/136833)

### ECHO: A Layered Model for the Design of a Context-Aware Learning Experience

Hadas Weinberger (2010). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications* (pp. 539-555).

[www.irma-international.org/chapter/echo-layered-model-design-context/39190](http://www.irma-international.org/chapter/echo-layered-model-design-context/39190)

### A Model-Driven Engineering Approach for Defining Rich Internet Applications: A Web 2.0 Case Study

Francisco Valverde, Oscar Pastor, Pedro Valderas and Vicente Pelechano (2010). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications* (pp. 40-58).

[www.irma-international.org/chapter/model-driven-engineering-approach-defining/39163](http://www.irma-international.org/chapter/model-driven-engineering-approach-defining/39163)

### Nesting Strategies for Enabling Nimble MapReduce Dataflows for Large RDF Data

Padmashree Ravindra and Kemafor Anyanwu (2014). *International Journal on Semantic Web and Information Systems* (pp. 1-26).

[www.irma-international.org/article/nesting-strategies-for-enabling-nimble-mapreduce-dataflows-for-large-rdf-data/113712](http://www.irma-international.org/article/nesting-strategies-for-enabling-nimble-mapreduce-dataflows-for-large-rdf-data/113712)

### A Service Oriented Ontological Framework for the Semantic Validation of Web Accessibility

Rui Lopes, Konstantinos Votis, Luís Carriço and Spiridon Likothanassis (2009). *Handbook of Research on Social Dimensions of Semantic Technologies and Web Services* (pp. 49-67).

[www.irma-international.org/chapter/service-oriented-ontological-framework-semantic/35722](http://www.irma-international.org/chapter/service-oriented-ontological-framework-semantic/35722)