



Cairo University

Faculty of Computers and Information

Information Systems Department



# **Enrichment Web Content Using Ontology Engineering and Annotation Methods**

Submitted By:

**Hanaa Ghareib Mohamed Hendi**

Demonstrator, Information Systems Department

Faculty of Computers and Information, Fayoum University

A Thesis Submitted to the Faculty of Computers and  
Information, Cairo University

In Partial Fulfilment of the Requirements for the Degree of  
Master of Information Systems

Supervised by:

**Assoc.Prof.Dr. Ehab Ezzat Hassanein**

Assoc. Professor, Information Systems Department,  
Faculty of Computers and Information – Cairo University

**Assoc.Prof.Dr. Haytham Tawfeek Al-Feel**

Assoc. Professor, Information Systems Department,  
Faculty of Computers and Information – Fayoum University

Cairo, 2018

## ABSTRACT

Automatic document classification is one of the important tasks in web 3.0 due to the rapid growth of the number of electronic documents on the web available nowadays. Classification aims to assign the document to a predefined category automatically based on its contents. In general, text classification plays an important role in information extraction and summarization, text retrieval, question answering and web page content filtering. The Semantic Web aims to describe the meaning of information published on the Web to enable retrieval of information based on the accurate understanding of information's semantics. Semantic Web plays a major role in enriching the web content where it allows user's data to be linked from a source to another to be later understood by machines.

Semantic annotation can be described as the tagging of an ontology class instance and mapping it to the corresponding ontology class so that data becomes meaningful. This will lead to classify objects based on their content. In addition to that, increment the number of metadata which describes the web content in an efficient manner, facilitates the searching via the web between machines on the web.

Creating metadata is one of the major techniques that helps machines to understand and deal with data on the web and can be enriched using annotation. Semantic annotation is not only just used for enriching the data model information, but also it can be one of the useful solutions to help semi-automatic or even automatic systems interoperability.

However, many Semantic Web annotation tools developed and have proved their success in multiple languages, while Arabic is the one of the worldwide spoken language in the world that still facing difficulties especially in annotation.

The goal of this thesis is to study the idea beyond annotation and how annotation could be developed and used in the semantic web, trying to answer the question: is the Arabic Language could be used as an annotating language? Our thesis presents a new Semantic Annotation Architecture and tool that read web RSS News and annotate them semantically which can facilitate the reading of full metadata to be a source of information for different users through adding extra data for a specific word, concept and topic. Our tool provides users with the ability to update metadata retrieved from the triple store based on the ontology created. Our tool supports the Arabic language and annotating the Arabic text semantically to enrich web content in order to be more meaningful and understandable by machines.