

**Deanship of Graduate Studies  
Al-Quds University**



**The Semantic Web Framework (SWF)**

**Ahmad Yahya Mohammed Nada**

**M.Sc. Thesis**

**Jerusalem – Palestine**

**1431/2010**

# **The Semantic Web Framework (SWF)**

Prepared By:

**Ahmad Yahya Mohammed Nada**

Supervisor:

**Dr. Badie Sartawi**

A thesis Submitted in Partial fulfillment of requirements for the degree  
of Master of Computer Science Department/ Master of Computer  
Science Program/ Faculty of Graduated Studies -  
Al-Quds University

1431/2010

Al-Quds University  
Deanship of Graduate Studies  
M.Sc. In Computer Science



## **Thesis Approval**

### **The Semantic Web Framework (SWF)**

Prepared By: Ahmad Yahya Mohammed Nada  
Registration No: 20714011

Supervisor: Dr. Badie Sartawi

Master thesis submitted and accepted, Date: June 5, 2010.

The names and signatures of the examining committee members are as follow:

|                      |                   |                  |
|----------------------|-------------------|------------------|
| 1- Head of committee | Dr. Badie Sartawi | Signature: ..... |
| 2- Internal Examiner | Dr. Raid Zaghal   | Signature: ..... |
| 3- External Examiner | Dr. Muna Matar    | Signature: ..... |

**Jerusalem – Palestine**

**1431/2010**

## **Declaration**

I certify that this thesis submitted for the degree of Master, is the result of my own research, except where otherwise acknowledged, and that this thesis (or any part of the same) has not been submitted for higher degree to any other university or institution.

Signed: .....

Ahmad Yahya Mohammed Nada

... /... /2010

## **Acknowledgments**

After thanking Allah .....

I would like to thank my advisor, Dr. Badie Sartawi, for his guidance throughout my years at university. He has prepared me well as a researcher in computer science, prepared me in computer science academics skills, and has worked hard to promote my work .....

Al-Quds University, I would like to thank my university specially the staff of Computer Science Department for their facilities over study period .....

Special thank to my father Dr. Yahya Nada, for his encouragement, support and advice to complete the master degree .....

Thanks to my honor students who have helped me in some stages of the work, Imad Kawasmi, and Mohamad Hamayel .....

I must thank my family, who has provided love and support over all of these years. My mother and father who have helped me get to where I am now. My brothers and sisters Somaya and her husband Ayob, Mohammed my reference in technology, Abd elnaser, Samya, Zakarya and Heba .....

Finally, most importantly, I want to thank my beautiful love Mryam for her love and supportive ideas and notes. She never leaves me when I worked in the preparation of this thesis, and always offered words of encouragement .....

## ملخص

من المهم ايجاد آلية لتنظيم موارد الويب واعادة هيكلتها بالطريقة الاكثر فعالية للحصول على افضل استفادة منها، هذه الاهمية مبينة على مدى شعبية هذا العالم الرقمي ومدى الاعتماد عليه من قبل مستخدميه الذين يمارسون انشطتهم من خلاله، سواء كانوا أناسا أو وكلاء الكترونيين أذكيا.

عدة حلول تم وضعها للوصول الى هذا التنظيم المهم، ومن بينها ما قام مخترع الويب بتقديمه وهي نظرية تعنى بتحول الويب الى قاعدة بيانات مترابطة مشتركة بين جميع شبكات العالم، تحت مسمى الويب الدلالي (Semantic Web)، وقام العديد من الباحثين في هذا المجال بتقديم جهود كبيرة لتحويل هذه النظرية الى واقع ولكن حتى الآن لم يتمكنوا من وضع اطار عام لها.

هدف الرسالة ايجاد اطار جديد وموحد للويب الدلالي يدعى (SWF)، SWF عبارة عن نظام يقدم اطار لادارة محتويات الويب والمساعدة في تنظيم موارده لتحقيق الحد الاقصى من اعادة استخدامها، اطار الويب الدلالي المقترح في هذه الرسالة يحل المشاكل المتعلقة بمجالات الويب المغلقة و لغات النمذجة المعقدة و مشاكل تحقيق محتويات الويب على انها كيانات ويب ومشاكل مشاركة البيانات الوصفية في الويب.

الخدمة الاولى التي يقوم الاطار بتقديمها هي Resource Description Framework (RDF) Generation وتهدف الى تحول عملية انشاء ملفات ال RDF الى عملية اوتوماتيكية يقوم بها النظام عن طريق الوصول الى الموقع الالكتروني و استخدام آلية تحديد المجال للموقع الالكتروني لاجاد البيانات المهمة و نبذ البيانات غير المهمة. آلية تحديد المجال للموقع الالكتروني تمكن النظام من تحديد نوع الموقع و استخدام المفردات الدلالية لاستخلاص النصوص من جميع صفحاته، ومن ثم تعينتها بنموذج ال RDF المخصص للمجال والمحضر مسبقا في النظام والذي يمكن النظر اليه على انه قالب للبيانات. لقد تم المقارنة بين العملية الآلية والعملية التقليدية في خلق ال RDF وتم اكتشاف ان العملية الآلية تستهلك ثوان من الوقت بينما العملية التقليدية تستهلك دقائق من الوقت.

خدمة لغة الاستعلام عن الويب الذي يقدمها الاطار تسهل البحث في الويب باستخدام لغة SQL معدلة. لقد قمنا بتقليل الوقت الذي يستهلكه تعلم لغة استعلام جديدة للاطار عن طريق استخدام لغة ال SQL، واختيرت هذه اللغة بالتحديد لما تتصف به من عالمية واستخدامها من قبل العديد من العمليات التجارية واستخدامها على نطاق واسع أكثر من أي لغة أخرى. الجانب المهم في هذه الخدمة هو ان المستعلم غير مضطر لتزويد النظام بالكلمات الصحيحة في البحث، وذلك لان النظام المطبق يستخدم محلل لغوي ذكي Linguistics Analyzer لفهم اجزاء جملة ال SQL، كما ويستخدم نظام التعلم الذكي Learning system ليقوم بسؤال المستعلم عن الكلمات غير المفهومة او الواضحة في الجملة.

تصنيف ملفات ال RDF هي خدمة مهمة اخرى في نظام الويب الدلالي، نظام التصنيف مهم جدا لمحركات البحث في تحسين النتائج، وذلك لانه يساعد في استرجاع النتائج المتوقع انها الاقرب لمعايير البحث المطلوبة. الاطار يطبق صيغة

تصنيف جديدة تعتمد على محتويات صفحات الموقع وعلى تاريخ ملفات ال RDF الخاصة بالموقع، وتستخدم الصيغة الجديدة قيمة التصنيف القديم مضافا اليها قيمة معقولة وذلك لاحتواء التصنيف الجديد ضمن مدى محدد.

ينفذ نظام ال SWF ميزات ذكية مثل المحلل اللغوي الذكي الذي يقوم بفهم جمل البحث، ونظام التعلم الذكي الذي يطلب مساعدة المستعلم عندما لا يستطيع النظام تحديد بعض البيانات المهمة اثناء تنفيذ الخدمات التي تم ذكرها، والحالات التي يكلب فيه النظام المساعدة تضمن: المساعدة في تحديد مجال الموقع الالكتروني اثناء عملية اشاء ملفات ال RDF، والمساعدة في تحديد مرادفات كلمات جمل البحث عند عدم القدرة على ايجادها في قاعدة المعرفة الخاصة بالنظام، والمساعدة في توضيح الكلمات الغامضة في جمل البحث.

يقدم نظام ال SWF ميزة جديدة هي تلخيص مواقع الويب، الفكرة من هذه الخدمة هي تلخيص ملف ال RDF المنشأ للموقع وتقديم فقرة مختصرة وقصيرة عن اجزاء الموقع المهمة.

يمكن لنظام ال SWF ان يغطي كامل الويب، لكن في هذه الرسالة قمنا بالتركيز على مجال التعليم وجعله نطاق التنفيذ لدينا، وقمنا بتطوير تطبيق على شبكة الانترنت باستخدام Microsoft C#.Net 2008 و Microsoft SQL Server 2008.

## **Abstract**

It is important to find a mechanism that organizes and restructures the web resources in the most effective way to get the best benefit from it; this importance comes from the dependability and popularity of such digitalized world where human and intelligent agents practice their activities.

Many solutions were put to resolve this organization, one of the most valuable solutions was suggested by the WWW creator Tim Berners-Lee, and his vision is to convert the entire web into integrated database across the world's networks, under the name of "Semantic Web". Researchers who adopted this vision actively participated to make this dream true, but until now no standardized framework is globalized.

The purpose of this thesis was to find a new and standard Semantic Web Framework (SWF), the SWF is a system that provides a framework to manage the web contents and help in organizing its resources to achieve the maximum reusability of them; many researchers had worked on this field but it is agreed that no global framework is set yet. The suggested SWF covers the problems areas of closed domains, complexity of modeling languages, realizing web entities and Metadata sharing.

The Resource Description Framework (RDF) Generation in the main service in the SWF, aims to automate the tradition RDF creation process by accessing the website pages and uses domain-oriented approach to find relevant contents and eliminate unwanted data, the domain-oriented approach enables the system to detect the website type and use semantics to extract the text from all the website's pages, then fill the suitable prepared RDF model which can be considered as RDF document template. The automated and the manual generation of the RDF documents have been compared and discovered that the automated process consumes time in seconds but the traditional process consumes time in minutes.

The Web Query language is a service that is provided by this system facilitates the web searching using modified SQL. We had reduced the time to learn new framework query by using the SQL language; it was chosen because it is standardized, used by many business processes and widely used more than any other language, the important aspect of this service



that the SQL requestor do not have to provide the correct words to the system; because the applied SWF uses Intelligent Linguistics Analyzer system to understand the statement parts, also it uses intelligent Learning system to ask the requestor for help about the parts of the statement which it couldn't understand.

RDF Ranking System is another significant service in the SWF, the ranking is important for the search engines to optimize the results; because it helps in returning the results that are predicted to be the closest ones to the search criteria, the SWF applies new updating rank formula that depends on the website pages contents and on the RDF documents history, the formula includes the old rank added to it a reasonable value to limits the ranks in a managed range.

The SWF carries out intelligent features such as the Linguistics Analyzer that understand the users' SQL statements, and the Learning System that learns from the requester when the SWF system cannot determine some of core data during the services execution, the cases when the Learning system asks for help includes: asking about the domain in the RDF generating, asking for keywords synonyms when it does not find them in the knowledgebase through the SQL execution, asking for the unclear keywords meanings when executing the SQL statement.

The SWF brings out a new feature of Auto-Web Summarizing; the idea of this feature is to summarize the generated RDF document and provides a brief and short paragraph of the website main parts.

The SWF can cover the entire web, but in this thesis we took the educational domain as the implementation scope using Microsoft C#.Net 2008 and Microsoft SQL Server 2008 to develop a web-based application.

## Table of Contents

|  |     |
|--|-----|
| Declaration. . . . .                                       | i   |
| Acknowledgment. . . . .                                    | ii  |
| Abstract in Arabic. . . . .                                | iii |
| Abstract. . . . .  | v   |
| Table of Contents. . . . .                                 | vii |
| List of Figures. . . . .                                   | x   |
| List of Tables. . . . .                                    | xi  |
| List of Appendices. . . . .                                | xii |
| <br>   |     |
| Chapter 1 Introduction. . . . .                            | 1   |
| 1.1 Introduction. . . . .                                  | 1   |
| 1.2 Problems and Motivation. . . . .                       | 2   |
| 1.3 Contributions to the Field. . . . .                    | 3   |
| 1.4 Outline. . . . .                                       | 4   |
| <br>   |     |
| Chapter 2 Background. . . . .                              | 5   |
| 2.1 The Web. . . . .                                       | 5   |
| 2.2 The Semantic web. . . . .                              | 6   |
| 2.3 The RDF. . . . .                                       | 8   |
| 2.4 SPARQL Query Language for RDF. . . . .                 | 9   |
| 2.5 Web Ranking. . . . .                                   | 10  |
| 2.6 Related Work. . . . .                                  | 13  |
| 2.6.1 Related Projects. . . . .                            | 13  |
| 2.6.2 Comparison between SWF and Other Frameworks. . . . . | 15  |
| 2.6.3 Our Work. . . . .                                    | 16  |

|           |  |    |
|-----------|--|----|
| Chapter 3 | Semantic Web Framework Model. . . . .          | 17 |
| 3.1       | Introduction. . . . .                          | 17 |
| 3.2       | SWF Overall Approach and Architecture. . . . . | 17 |
| 3.3       | The RDF Generator Model. . . . .               | 20 |
| 3.3.1     | Introduction. . . . .                          | 20 |
| 3.3.2     | Websites type detection. . . . .               | 21 |
| 3.3.3     | Websites classification. . . . .               | 21 |
| 3.3.4     | RDF Models. . . . .                            | 22 |
| 3.3.5     | Semantic Extraction. . . . .                   | 23 |
| 3.3.6     | SWF Learning. . . . .                          | 25 |
| 3.4       | The RDF Query Language for SWF. . . . .        | 26 |
| 3.5       | Ranking System for SWF. . . . .                | 28 |
| 3.6       | Discussion. . . . .                            | 29 |
| Chapter 4 | SWF System Design. . . . .                     | 31 |
| 4.1       | Introduction. . . . .                          | 31 |
| 4.2       | Overall Description. . . . .                   | 31 |
| 4.3       | Functional Requirements. . . . .               | 33 |
| 4.4       | External Interface Requirements. . . . .       | 34 |
| 4.5       | Non-Functional Requirements. . . . .           | 35 |
| 4.6       | The System Design. . . . .                     | 36 |
| 4.6.1     | SWF Flowcharts. . . . .                        | 36 |
| 4.6.2     | SWF Use Case Diagram. . . . .                  | 40 |
| 4.6.3     | Narrative Use Cases. . . . .                   | 41 |
| 4.6.4     | Database Entity Relationship Diagram. . . . .  | 60 |
| 4.6.5     | Database Entities Specifications. . . . .      | 60 |

|           |  |    |
|-----------|--|----|
| Chapter 5 | Results and Evaluation. . . . .                                | 64 |
| 5.1       | Results. . . . .   | 64 |
| 5.1       | Evaluation. . . . .  | 65 |
| Chapter 6 | Conclusion and Future Works. . . . .                           | 74 |
| 6.1       | Conclusion. . . . .  | 74 |
| 6.2       | Recommendations and Future Works. . . . .                      | 75 |
|           | References. . . . .  | 76 |
|           | Appendices. . . . .  | 81 |
|           | Appendix 1: An Example of RDF Document for Alquds.edu. . . . . | 81 |
|           | Appendix 2: W3C Document License. . . . .                      | 86 |
|           | Appendix 3: SWF User Manual. . . . .                           | 88 |
|           | Appendix 4: Publications. . . . .                              | 98 |

## List of Figures

|            |   |    |
|------------|---|----|
| Figure 2.1 | Semantic Web Layer Cake. . . . .                        | 7  |
| Figure 2.2 | Basic form of a semantic triple. . . . .                | 8  |
| Figure 2.3 | RDF triples. . . . .                                    | 9  |
| Figure 2.4 | Sample RDF statement. . . . .                           | 9  |
| Figure 2.5 | Hits Algorithm Authority and Hub. . . . .               | 10 |
| Figure 2.6 | Hits Update Roles. . . . .                              | 11 |
| Figure 3.1 | SWF architecture. . . . .                               | 18 |
| Figure 3.2 | The current status of the Web. . . . .                  | 19 |
| Figure 3.3 | The Web after using the Semantic Web Framework. . . . . | 19 |
| Figure 3.4 | RDF Generator model. . . . .                            | 20 |
| Figure 3.5 | Update rank algorithm. . . . .                          | 29 |
| Figure 3.6 | The SWF System. . . . .                                 | 29 |
| Figure 4.1 | RDF Flowchart. . . . .                                  | 36 |
| Figure 4.2 | Query Flowchart. . . . .                                | 37 |
| Figure 4.3 | Web summary Flowchart. . . . .                          | 38 |
| Figure 4.4 | Update rank. . . . .                                    | 39 |
| Figure 4.5 | SWF Use Case Diagram. . . . .                           | 40 |
| Figure 4.6 | SWF ERD. . . . .  | 60 |

## List of Tables

|            |  |    |
|------------|--|----|
| Table 2.1  | Comparison Between SWF and the Other Frameworks. . . . .                     | 15 |
| Table 4.1  | User Classes and characteristics. . . . .                                    | 32 |
| Table 4.2  | RDFs Table. . . . .  | 60 |
| Table 4.3  | Domains Table. . . . .   | 61 |
| Table 4.4  | Properties Table. . . . .  | 61 |
| Table 4.5  | RDF_Properties Table. . . . .  | 62 |
| Table 4.6  | Domain_ Properties Table. . . . .  | 62 |
| Table 4.7  | KeyTypes Table. . . . .  | 63 |
| Table 4.8  | Knowledge base Table. . . . .  | 63 |
| Table 5.1  | Evaluation of Creating RDF Manually and Automatically. . . . .               | 65 |
| Table 5.2  | The distribution of the study sample according to the Academic Rank. . . .   | 67 |
| Table 5.3  | The distribution of the study sample according to the Number of Publications | 67 |
| Table 5.4  | The study tool depending on the dimensions of the study. . . . .             | 67 |
| Table 5.5  | Balance of the percentages of responses. . . . .                             | 69 |
| Table 5.6  | Averages and percentages of the first pivot. . . . .                         | 69 |
| Table 5.7  | Averages and percentages of the second pivot. . . . .                        | 70 |
| Table 5.8  | Averages and percentages of the third pivot. . . . .                         | 71 |
| Table 5.9  | The results of (ANOVA) test to academic rank variable. . . . .               | 71 |
| Table 5.10 | The results of (ANOVA) test to Number of Publications variable. . . . .      | 71 |

## List of Appendices

|            |  |    |
|------------|--|----|
| Appendix 1 | An Example of RDF Document for Alquds.edu. . . . . | 81 |
| Appendix 2 | W3C Document License. . . . .                      | 86 |
| Appendix 3 | User Manual. . . . .                               | 88 |
| Appendix 4 | Publications. . . . .                              | 98 |