

A Knowledge Representation Language for Arabic Semantic Web Using Resources Description Framework

Dr. Abdul Monem S. Rahma* & Dr. Jamal F. Tawfeq**

Received on: 2/6/2009

Accepted on: 30/6/2010

Abstract

The World Wide Web is the greatest repository of information ever assembled by man. The web was not designed to be processed by machines. The web information is meaningless for computer, and it is very hard to find out what you are looking for. For this reason, it has new challenges a raise to build a Semantic Web infrastructure where documents will be understandable by human and computers. The knowledge creation and development of Semantic Web application has not yet been easy.

In this paper Resources Description Framework (RDF) will be used as a knowledge representation language which allows the representation of an Arabic Sentence in Web environment to be annotated with semantics. And present formalism for the language and discuss the features that make it well suited for the Arabic Web.

Keywords: Semantic Web, Semantic Net, Knowledge Representation, RDF

لغة التمثيل المعرفي الى المواقع الدلالية العربية بأستخدام لغة وصف إطار الموارد

الخلاصة

الشبكة العالمية هي اكبر مستودع للمعلومات اوجد على الاطلاق من قبل الانسان. والتصميم الحالي لشبكة الانترنت ليس مصمم للمعالجة من قبل الحاسبات. لا تستطيع الحاسبة من فهم المعلومات المتوفرة على الشبكة العالمية للمعلومات، ومن الصعب جدا ان نجد ما نبحث عنه. لهذا السبب فقد وجدت تحديات جديدة لبناء البنية التحتية للمواقع الدلالية، حيث ان وثائق الويب الاستدلالي ستكون مفهومة من قبل الانسان والحاسبة. حتى الان خلق المعرفة وانشاء تطبيقات المواقع الدلالية ليست عملية سهلة.

في هذه الورقة، سيتم استخدام RDF كألغة لتمثيل المعرفة ونقدم صيغة مضافة لها، بحيث تسمح بتمثيل الجمل العربية في بيئة الانترنت مقرونة بمعاني الكلمات. ونناقش السمات التي تجعلها مناسبة تماما للمواقع العربية.

1. Introduction

The Web has influenced the way people communicate and collaborate. Publishing information on the Web, make it accessible to anyone with access to the Web, or using the Web as a source of information to derive new

knowledge. The amount of information that is accessible on the Web has increased enormously in a short period of time. This increase of information is a desirable evolution, but it has also made the problems with the Web more evident.

*Computer Science Department, University of Technology/Baghdad

** Science College, University of Al-Nahrain /Baghdad

Everyone that has used the Web to search for information knows that it is not as easy or as fast as one would like it to be. Because the Web was not designed to be processed by machines, intelligent tools must be developed for integration of information that can be extracted from pages including special information that tells a computer how to display a particular piece of text or where to go when a link is clicked. They help the machine to determine what the text means.

Tim Berners-Lee, inventor of the Web has coined the term Semantic Web to describe this approach. Berners-Lee, Hendler and Lassila provide the following definition [1]:

The Semantic Web is not a separated Web but an extension of the current one, in which the information is given a well-defined meaning, better enabling computers and people to work in cooperation.

For this definition, there are two main features to the Semantic Web:-

1. The Semantic Web is not a separated Web, but an extension of the current Web. That means two names to the same thing. The Semantic Web exists in the Web and is a part of the Web at the same time. This makes them inseparable at the Uniform Resource Identifier (URI) level.
2. The name Semantic Web comes from the fact that it represents as a set of semantically and formally interlinked data units there by creating a Semantic Web inside the Web.

2. Knowledge Representation

The Semantic Web depends on the ability to associate formal meaning with content. The field of knowledge representation provides a good starting point for the design of a Semantic Web language because it offers insight into the design and use of languages that attempt to formalize meaning. However, the nature of the Web challenges many of the assumptions of traditional knowledge representation work, and requires us to look at the problem from a new perspective.[2]

A good system for the representation of complex structured knowledge in a particular domain should process the following four properties:

- a. Representation adequacy: the ability to represent all of the kinds of knowledge that are needed in the domain.
- b. Inferential adequacy: the ability to manipulate the representational structures in such a way as to drive new structures corresponding to new knowledge inferred from old knowledge.
- c. Inferential efficiency: the ability to incorporate into the knowledge structure additional information.
- d. Acquisitional efficiency: the ability to acquire new information easily.[3]

One of the reasons that knowledge structures are so important is that they provide a way to represent information about commonly occurring patterns of things.

There are many methods of knowledge representation. One of these is

Semantic net, while entities can be described as collection of attributes and associated values. To do this, they are stored as ordered triples, which are usually thought of as having the form

Object , Attribute , Value

Information can be retrieved by specifying values for any two fields from the triples. And we can represent this triple as a set of nodes connected by each other by a set of labeled arc, which represent relationships among the nodes.

As a simple example, consider the Arabic sentence:

قصيدة السيف أصدق أبو تمام نَظَمَ

That mean the (subject) of this sentence is (أبو تمام) and the attribute (property) is (نَظَمَ) and the (object) is (قصيدة السيف أصدق). The sentence above would thus diagrammed as (node and arc diagram) as shown in figure (1).

3. Rdf Language And Model

The Resource Description Framework (RDF) is a graphical language for representing information about resources in the World Wide Web. Information takes the form of a directed graph.[5] Resources are described in terms of properties and property values using RDF statements. Statements are represented as triples, consisting of subject, predicate and object.

RDF offers developers a powerful toolkit for making statements and connecting those statements to derive meaning. RDF offers a different, and in some ways more powerful, framework for data representation than XML or relational databases.

The foundation of RDF is a model for representation named properties

and property value. The RDF model draws on well-established principles from various data representation communities. RDF properties may be thought of as attributes of resources and in this sense correspond to traditional attribute-value pairs. RDF properties also represent relationship between resources and RDF model can there for resemble an entity-relationship diagram.[6]

The RDF data model is a syntax-neutral way of representing RDF expressions. The data model representation is used to evaluate equivalence in meaning. Two RDF expressions are equivalent if and only if their data model representation is the same [7].

RDF is a knowledge representation language which allows the representation of an Arabic sentence in web environment to be annotated with semantics. There are many kinds of statements in Arabic language. In this paper we represent a verb statement in RDF and used it the RDF document.

3.1 Rdf Triple And Model For Arabic Sentence

The basic data model consists of three object types [2][7]:

1. Resources: all things being described by RDF expressions are called resources. A resource may be an entire web page or a part of web page or whole collection of pages, or an object that is not directly accessible via optional anchor ids or anything can have a URI, the extensibility of URIs allows the introduction of identifiers for any entity imaginable.

2. Properties: it is a specific aspect, characteristic, attribute or relation used to describe a resource. Each property has a specific meaning, defines its permitted values, the types of resources it can describe, and its relationship with other properties.
3. Statements: a specific resource together with a named property plus the value of that property for that resource is a RDF statement.

These three individual parts of a statement are called,

1. *Subject*.
2. *Predicate*.
3. *Object*.

The object of a statement (i.e., the property value) can be another resource or it can be a literal; i.e., a resource (specified by a URI) or a simple string or other primitive data type defined by XML. In RDF terms, a *literal* may have content that is XML markup but is not further evaluated by the RDF processor.

There are three steps to represent an Arabic sentence. These steps are:

1. Backing to the previous example figure (1), is a representation of a statement whereby three components of the statement can be replaced by instances of the components to generate a specific statement. Then the sentence

نظم ابو تمام قصيدة السيف اصدق

2. In RDF, this new statement, redefined as RDF triple, can be considered a complete RDF graph because it consists of a complete fact that can be recorded using RDF methodology and that can then be documented using shorthand

techniques. It is using the following to represent a triple:

{*subject, predicate, object*}

Then, the above Arabic statement becomes:

{ قصيدة السيف اصدق, نظم, ابو تمام }

This triple has the following parts:

Subject (Literal)	ابو تمام
Predicate (Property)	نظم
Object (Literal)	قصيدة السيف اصدق

Then the diagram of this sentence is shown in Figure (2):

3. Then the Figure (2) can be represented as the RDF/XML is:

```
<rdf:RDF>
  <rdf:Description about=" ابو تمام">
    <نظم/> <قصيدة السيف- اصدق/> <نظم/>
  </rdf:Description>
</rdf:RDF>
```

3.2 Dublin Core

The Dublin Core is a set of 15 elements believed to be broadly applicable to describing web resources to enable their discovery. An important consideration in the development of the Dublin Core was to allow simple descriptions, but also to provide the ability to qualify descriptions in order to provide both domain specific elaboration and descriptive precision.[8]

4. Metadata And Rdf

Metadata is Structured Document about document, objects, article and poem when properly implemented metadata can unambiguously describe information resource, enhancing information retrieval and enabling accurate matches to be done.

Metadata are used to describe data, information and any resource.

Metadata can describe just about anything you find on a computer. Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage information resource[10][11]. In this paper the information resource is poem. T. B. Lee give a well-formed definition to metadata as [6][7]:

Definition:

Metadata is machine understandable information about web resource or other things.

Poem metadata is structured information about poem, such as the poet, poem name, rhyme and prosody that were using in the poem.

Metadata can also include information about a poet that says the poem. Metadata can also include information about verses of the poem, and the date of the poem.

Throughout the search in this paper, it was founding that, there are three principles for poem metadata:

1. Metadata is essential to identify poem.
2. Ownership metadata must never be removed.
3. Metadata must be written in formats that are understood by all.

If all application writes metadata using open standards that are fully consistent and interoperable, then information will be able to move seamlessly between different system and environments.

5. Rdf Vocabulary Schema

The RDF schema (RDFS) is a general-purpose language for representing information in the web.

This specification describes how to use RDF to describe RDF vocabularies.[9] A Resource Description Foundation (RDF) vocabulary is a defined set of predicates that can be used in an application.

A vocabulary is similar to a schema in that it defines a set of elements that can be used in an application. However, a vocabulary also helps you to define the domain and range of a predicate.

RDF vocabularies can describe relationships between vocabulary items from multiple vocabularies that have been developed independently.

6. The Proposed Rdf Poem Schema

In this paper, the RDF Vocabulary schemas are used to describe these characteristics of the poem and much information about poet. Those schemata are passes the validation of W3C at site <http://www.w3.org/RDF/Validator/>.

These schemas are separated into:

a. Modify Dublin Core (DC) Schema.

It is a general schema for identify original works, like article, painting, poem, book...etc. it contains properties such as, creator, title, date of publication...etc. in this paper, not all DC properties are used. These properties are:

- i. Title: the name given to the poem.
- ii. Creator: poet name or URL that can be further described with other schemas.
- iii. Subject: the topic of the content of the poem. It could be one keyword or more to describe the poem.

iv. Description: Describe the content of poem.	</rdf:Property>
v. Publisher: The person or institution responsible for making available.	<rdf:Property rdf:ID="subject"> <label xml:lang="en">Subject</label> <label xml:lang="ar">الموضوع</label> <SubPropertyOf
vi. Contributor: not used.	
vii. Date: The date and time the poem was written.	
viii. Format: not used.	
ix. Identifier: not used.	rdf:resource="http://purl.org/dc/elements/1.1/subject" /> </rdf:Property>
x. Relation: A reference to a related for series of poem.	
xi. Coverage: not used.	
xii. Rights: Information about rights held in and over the poem.	<rdf:Property rdf:ID="description"> <label xml:lang="en">Description</label> <label xml:lang="ar">الوصف</label> <SubPropertyOf
xiii. Source: not used.	
xiv. Language: not used.	
Below is a modify Dublin Core Schema, and we used an Arabic Vocabulary.	rdf:resource="http://purl.org/dc/elements/1.1/description" /> </rdf:Property>
<rdf:RDF	
xmlns="http://www.w3.org/TR/1999/PR-rdf-schema-19990303#"	
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">	<rdf:Property rdf:ID="publisher"> <label xml:lang="en">Publisher</label> <label xml:lang="ar">الناشر</label> <subPropertyOf
<rdf:Property rdf:ID="title">	
<label	
xml:lang="en">Title</label>	
<label	
xml:lang="ar">عنوان</label>	rdf:resource="http://purl.org/dc/elements/1.1/publisher" /> </rdf:Property>
<SubPropertyOf	
rdf:resource="http://purl.org/dc/elements/1.1/title" />	</rdf:Property>
</rdf:Property>	
<rdf:Property rdf:ID="creator">	<rdf:Property rdf:ID="date"> <label xml:lang="en">Date</label> <label xml:lang="ar">التاريخ</label> <subPropertyOf
<label	
xml:lang="en">poet</label>	
<label	
xml:lang="ar">شاعر</label>	
<SubPropertyOf	
rdf:resource="http://purl.org/dc/elements/1.1/Creator" />	rdf:resource="http://purl.org/dc/elements/1.1/date" />

```

</rdf:Property>
<rdf:Property rdf:ID="relation">
  <label
xml:lang="en">Series</label>
  <label
xml:lang="ar">سلسلة</label>
  <subPropertyOf

rdf:resource="http://purl.org/dc/eleme
nts/1.1/relation" />
</rdf:Property>

<rdf:Property rdf:ID="rights">
  <label
xml:lang="en">Rights</label>
  <label
xml:lang="ar">الحقوق</label>
  <subPropertyOf

rdf:resource="http://purl.org/dc/eleme
nts/1.1/rights" />
</rdf:Property>
</rdf:RDF>

```

b. Artistry Schema

This schema is the content of the artistry characteristics to the poem. There are many properties to describe the poem. These properties are:

- i. Poem: the poet name.
- ii. Rhyme: the last letter for each verse in the poem.
- iii. Prosody: the type of the prosody.
- iv. Verse: one line in the poem.
- v. Description: describe the meaning of the verse.
- vi. Type: the type of the poetry.

Below is an artistry schema, using an Arabic vocabulary:

```
<rdf:RDF
```

```

xmlns:rdf="http://www.w3.or
g/1999/02/22-rdf-syntax-ns"
xmlns="http://www.w3.org/20
00/01/rdf-schema#"
xmlns:artistry="">
<Class rdf:ID="Artistry-Data">
  <comment xml:lang="en">A
class that represents artistry data
about a poem</comment>
  <comment xml:lang="ar">الفئة
التي تمثل البيانات الفنية حول القصيدة
</comment>
</Class>
<rdf:Property rdf:ID="poem">
  <label
xml:lang="en">Poem</label>
  <label
xml:lang="ar">القصيدة</label>
  <comment xml:lang="ar">اسم
القصيدة</comment>
  <domain rdf:resource=#Artistry-
data"/>
</rdf:Property>
<rdf:Property rdf:ID="rhyme">
  <label
xml:lang="en">Rhyme</label>
  <label
xml:lang="ar">القافية</label>
  <comment xml:lang="ar">نوع
القافية</comment>
  <domain rdf:resource=#Artistry-
data"/>
</rdf:Property>
<rdf:Property rdf:ID="prosody">
  <label
xml:lang="en">Prosody</label>
  <label
xml:lang="ar">العروض</label>
  <comment xml:lang="ar">البحر
الذي بنيت على وزنه
القصيدة</comment>
  <domain rdf:resource=#Artistry-
data"/>
</rdf:Property>

```

```

<rdf:Property rdf:ID="verses">
  <label
xml:lang="en">Verses</label>
  <label      xml:lang="ar">بيت
شعر</label>
  <comment xml:lang="ar">بيت شعر واحد</comment>
  <comment xml:lang="ar">من القصيدة</comment>
  <domain      rdf:resource=#Artistry-
data"/>
</rdf:Property>
<rdf:Property rdf:ID="meaning">
  <label
xml:lang="en">Meaning</label>
  <label
xml:lang="ar">المعنى</label>
  <comment xml:lang="ar">بيت شعر واحد</comment>
  <comment xml:lang="ar">معنى من القصيدة</comment>
  <domain      rdf:resource=#Artistry-
data"/>
</rdf:Property>
</rdf:RDF>

```

7. Example

Based on those RDF vocabulary schemata the metadata to explain the poem for Abi-Tamam about the death of his son Muhammad. The verses is

ولا تحسبن الموت عاراً فأنا رأينا المنايا قد
أصبنَ مُحَمَّدًا

The RDF description for these verses is:

```

<? xml version='1.0' encoding='ISO-
8859-1'?>
  <rdf:RDF
xmlns:RDF="http://www.w3.org/1999
/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/TR/19
99/PR-rdf-schema-19990303#"
xmlns:pd0=
"http://www.w3.org/2000/PoemRDF/d
c-1-0#"

```

```

xmlns:pd1=
"http://www.w3.org/2000/PoemRDF/a
rtistry#">
  <rdf:Description rdf:about="">
    <pd0:الشاعر > أبو تمام</pd0:الشاعر >
    <pd0:العنوان > يرثي ابنه</pd0:العنوان >
    <pd0:الموضوع > العنوان</pd0:الموضوع >
    <pd0:الوصف > قال هذه القصيدة عند</pd0:الوصف >
    <pd1:القافية > الوصف</pd1:القافية >
    <pd1:الطويل > المقول</pd1:الطويل >
    <pd1:المعنى > العروس</pd1:المعنى >
    <pd1:بيت شعر > بيت</pd1:بيت شعر >
    <pd1:عارة > بيت</pd1:عارة >
    <pd1:المعنى > شعر</pd1:المعنى >
    <pd1:المعنى > لا يعد الوت هو العار</pd1:المعنى >
    <pd1:المعنى > لأن الموت قد أصاب محمد</pd1:المعنى >
  </rdf:Description>
</rdf:RDF>

```

8. Conclusions

The Arabic web sites are expanding every day. There is now a huge information publishing on the web in an Arabic language. There are now vast recourses of Arabic web site available on it.

Using metadata about the Arabic web site help us to find exact information wanted and become a processable by machine.

RDF is a formal specification of how to represent the metadata as subject and object and relation between them.

There are two schema suggested in this paper to describe an Arabic poem. These schemas are:

1. Modify Dublin Core (DC) schema.
2. Artistry schema.

3. References

- [1] T.Berners-Lee, J.Hendler, and O.Lasilla, "The Semantic Web". *Scientific American*, May 2001.
- [2] Shelley Powers, "Practical RDF", O'Reilly & Associates, United States of America, 2003.
- [3] Elaine Rich, "Artificial Intelligence", McGraw-Hill book co., Singapore, 1984. [2] Passin T. B., "Explorer's guide to the Semantic Web", Manning Publications Co., ISBN 1932394206, June 2004.
- [4] George Luger, "Artificial Intelligence, structures and strategies for complex problem solving", Addison Wesley, England, 2002.
- [5] VJ Carey, "Semantic web concepts and tools in Bioconductor", 2006 July 28
- [6] Gustavsson B., "On the Semantic Web language", Växjö University, School of Mathematics and Systems Engineering, 2001.
- [7] Ora Lassila, Ralph R. Swick, "Resource Description Framework (RDF) Model and Syntax Specification", URL: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>, 22 February 1999 [4] Dr.Michael L.Brodie, "Information Technology. OntoWeb", Innsbruck, Austria, December 16-18,2002,Version January 24,2003.
- [8] Dan Brickley, R.V. Guha, and Andrew Layman, *Resource description Framework (RDF) Schema Specification, W3C, 1998,*
- [9] Dan Brickley, *RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation 10 February 2004*
- [10] National Information Standards Organization, "Understanding Metadata", Niso Press, 2004, URL www.niso.org
- [11] Lee T. B., "Metadata Architecture", 6 Jan 1997, URL: <http://www.W3C.org/DesignIssues/semantic.html>.



Figure (1) Semantic Net for Arabic Sentence

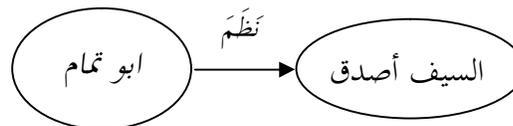


Figure (2) Simple Node and Arc Diagram