

Semantic Web: Improving Web Search Using RDF Instead of XML

GJCST Classification
H.3.3, H.3.7

Fadi Al-Kalani¹, Mamoun .G. Awad², Nabeel Bani Hani³

Abstract- The aim of this research is to present the possible using of XML and RDF in developing the concept of Semantic Web. It is provide a brief introduction to the Semantic Web and defines related terms, in addition to discuss the optimize method to describe web contents using both XML and RDF. However, there have been different work dealing with the same issue, but our approach is more intended to deal with semantic information and produce different standard format. The approach focus on developing a simulate web search engine to describe and emphasize whether to deploy XML or RDF in web search.

keywords- XML, RDF, OWL, W3C.

I. INTRODUCTION

The future and present generation of the Web is often characterized as the "Semantic Web" the concept emphasis that the information will no longer only be intended for human readers, but also for processing by machines, enabling intelligent information services, personalized Web-sites, and semantically empowered search-engines. The Semantic Web requires interoperability on the semantic level as well as semantic interoperability requires standards not only for the syntactic form of documents, but also for the semantic content. Proposals aiming at semantic interoperability are the results of recent W3C standardization efforts, notably Extensible Markup Language (XML)/XML Schema and Resource Description Framework (RDF)/RDF Schema (RDFS) [1]. W3C standard document format for writing and exchanging information on the Web emphasis that XML is mostly concerned with syntax that does not make sense without semantics, and many recent activities aim at adding more semantic capabilities to XML (s.Decker ets 2000). However, RDF is mostly concerned about semantics which is not very useful in a computer system without syntax, and many recent activities aim at providing a syntactic grounding for RDF. In particular, the research introduce positive impact of using RDF instead of XML through the concept "Semantic Web", with the knowledge that many of the researches demonstrated preference RDF for the many reasons which include simplicity, abstract syntax, and providing a data model.

¹ About¹ -Delmon University, Bahrain Email: fadek2000a@yahoo.com

² About² -University of Bahrain, Bahrain Email: mawad@asc.uob.bh

³ About³ - University of Bahrain, Bahrain Email: nbanihani@asc.uob.bh

II. BACKGROUND

Theoretically, many of the researches demonstrated preference RDF rather than XML for following reasons (s.Decker ets 2000):

- The main advantage of RDF over the basic XML is its simplicity. Unlike the order of elements in XML, the order of RDF properties does not matter. RDF offers a very appealing and flexible solution to any web designer.
- RDF has an abstract syntax that reflects a simple graph-based data model, and formal semantics with a rigorously defined notion of entailment providing a basis for well founded deductions in RDF data.
- XML and RDF are the current standards for establishing semantic interoperability on the Web, but XML addresses only document structure. RDF better facilitates interoperation because it provides a data model that can be extended to address sophisticated ontology representation techniques (s.Decker ets 2001).

In Fact, many of these advantages are theoretically described and does not have a clear measure with the concept Semantic web, furthermore those advantages were build on with the assumption of the features of the both languages XML and RDF, in addition, the impact of using RDF rather than XML is not measurable specially through the concept of semantic web. Actually, there have been different previous works dealing with similar problems. However, our proposal is more intended to deal with Semantic information and produce different standard formats. In particular, our main goal is to define the appropriate elements to develop a semantic web using both XML and RDF. Our approach achieve by developing a simulated web search engine to describe and emphasizes the positive role of using RDF rather than XML in web search.

III. RESOURCES AND METHODOLOGY OF THE RESEARCH

Currently most of the Web content is suitable for human use. Furthermore the typical uses of the Web today are information seeking, publishing, and using, searching for people and products, shopping, reviewing catalogues. But, today's Web search face many Limitations summarized as (Payam Barnaghi 2008):

- The Web search results are high recall, low precision.
- Results are highly sensitive to vocabulary.
- Results are single Web page.
- Most of the publishing contents are not structured to allow logical reasoning and query answering.

HTML is the language used to display graphics and text, but the contents describes cannot be processed by the machine. The semantic web is a tool to address this issue by introducing XML (extensible markup language), RDF (resource description framework), RDFS (RDF Schema) and OWL (web ontology language) to describe web contents that enable automated information access.

Initial work in this field of research is carried out by introducing the objectives of the research that will be achieved by developing application file system that offer specifically suited search and processing methods for given address file. In detail each file would contain information that describes common biography of apple. Deeply, results would have contextual analysis of the represented information and knowledge. However these aims should be addressed by the following steps:

- Develop two separate XML files contains common biography of apple. Typically, the property of the root apple consists of hotel, restaurant, hardware, and fruit. In addition, both files would contains information which represents summarize descriptions of the web contents. the structure of the first file shown in (fig. 1) where used to restore the data of apple with the categories elements defined as classification, name, and address. While the second file shown in (fig. 2) organized elements with name given as hotel, restaurant, hardware, and fruit. In fact, both files would present same information with different document structure.

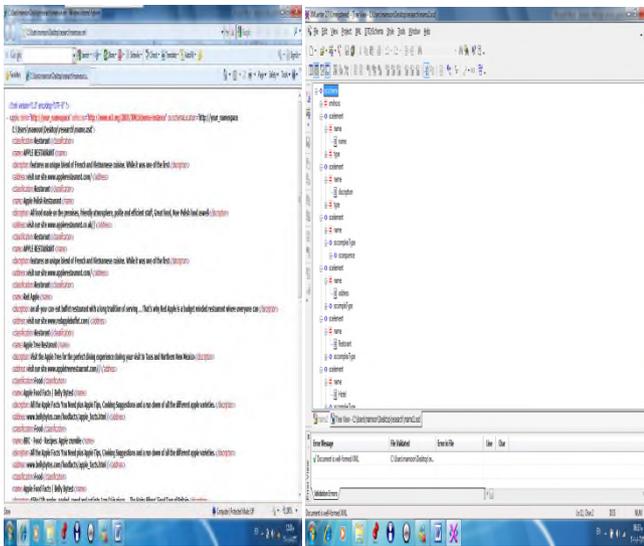


Fig. 1. XML source code and schema for the first file

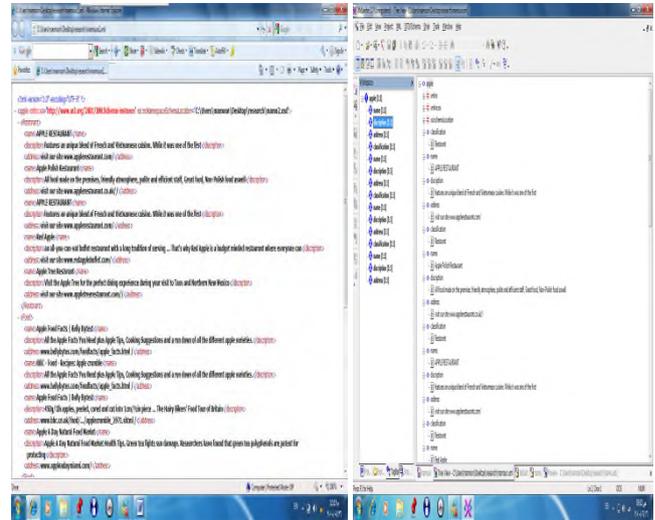


Fig. 2. XML source code and schema for the second file

- Develop java script files used to retrieve data from both XML and RDF file and display the results in web pages format.
- In order to complete the idea about our application's capabilities we present RDF file depicted in (fig. 3) to restore the same information described in XML files.

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://APPLE_POLISH_RESTORANT"
  xmlns:rd="http://APPLE_RED_RESTORANT">
  <rdf:Description rdf:about="APPLE_POLISH RESTORANT">
  <dc:description>features an unique blend of French and Vietnamese cuisine. While it
  was one of the first</dc:description>
  <dc:address>visit our site www.applerestaurant.com</dc:address>
  </rdf:Description>

  <rdf:Description rdf:about="APPLE_RED RESTORANT">
  <rd:description>an all-you-can-eat buffet restaurant with a long tradition of serving ...
  That's why Red Apple is a budget minded restaurant where everyone
  can</rd:description>
  <rd:address>visit our site www.redapplebuffet.com</rd:address>
  </rdf:Description>
</rdf:RDF>
    
```

Fig. 3. RDF source code

As mentioned, our application process is based on the concept of Information Retrieval, where a structured XML file and RDF are used as input, and where the result is a presentation document in HTML format with using of the transformation language java script.

IV. RESULTS OF RESEARCH

The following part of our research examines and describes the output that has shown after the process of search using the input which representing XML and RDF files.

- *Why not XML*

The wise of develop two different structures XML files is attempt to answer the question: Is the web search limitation is a result of using XML in describing web content? Or if we think more in XML we can gain better results? Although, we used the two different structure XML files, mostly we found UN satisfactions results. The UN satisfactions results appear through a single result only or the Web search results are high recall

- a) *Single result*

When uncompleted or the information is missed as a process of any search the impression of UN certainly and UN satisfaction of the process of the search would express. On the one hand the information is available; on the other hand complete information is not displayed. Using of our developed simulated web search engine, we attempt to reach to specific information about APPLE RESTAURANT meaning that we need information about restaurant related to apple. According to that the equation assumed to be (word_search= "APPLE RESTAURANT"). Surprisingly, the result shown in (fig. 4) is totally uncompleted since there is more information that contains the word search.

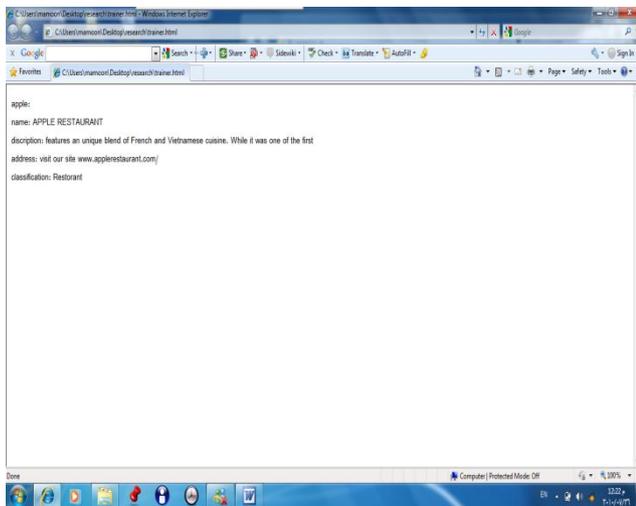


Fig.4. Single result displayed

Despite the fact that there are more results described APPLE RESTAURANT, but the result displayed show that this is the only information that is available. Essentially, we refer such of these results to the structure of XML which face Some of the problems with using attributes, XML Attributes are difficult to read and maintain, in addition it cannot contain multiple values as well as it cannot contain tree structures .those facts allow us to use elements for data and use attributes for information that is not relevant to the data. In other words XML is not support the idea that the attribute used to describe the element which allows us to give more information about each element and then specifically reach to a target element. Even using the two XML files described; the result is closely, Thus because XML structure

enforce the element to be isolate from it description. In fact mostly we found this drawback result in many of library and bookshop web sites. For example unless typing the same book name, we wouldn't get a target result. Basically, we may refer the limitation search to the extremely using of XML in developing of library and bookshop web sites.

- b) *High recall results*

The drawback characteristic of using attributes in XML described may cause another web search limitation represented in displaying many of the results that are not related to the target search. In our developed simulated web search engine we need to reach to the specific information about "Apple Hotel In the world ". Originally, the word search follow classification "Hotel", user expect to only get a fully information about Hotels in the world. In fact the results appear in (fig. 5) show information about Foods, Hardware, Restaurants and Hotels that may dispersal the user.



Fig.5. Many results displayed

Because we don't specify to each element it description so we couldn't specify a target word search. In other words the limitation appeared is a result of XML couldn't assign attribute to the element?

- *Why RDF*

As recognizable, XML itself is not concerned with meaning; however the standard of XML doesn't indicate how to derive a fact from a document. In fact XML documents are not useful for understandable but display documents.

RDF provides some basic level of meaning particularly, RDF is designed to represent knowledge in a distributed world, means RDF is particularly concerned with meaning. The second key aspect of RDF is that it works well for distributed information by establishing some relationships between documents that allow RDF applications files to put together .It does this in two ways, first by linking documents together by the common vocabularies they use, and second by allowing any document to use any vocabulary (<http://www.rdfabout.com/intro/>). Insider structure, RDF has XML syntax, consist of triples or sentences it consists of a resources, a property, and a value. In other sequence Object -> Attribute-> Value triples. Every resource has a URI

which can be a URL (a web address) or a some other kind of identifier; the Properties which are special kind of resources; used to describe relations between resources; and lastly Statements that is an object-attribute-value triple [9]. Back to our developed simulated web search engine, according to (fig. 6) we convert part of XML file into RDF with the same given information.

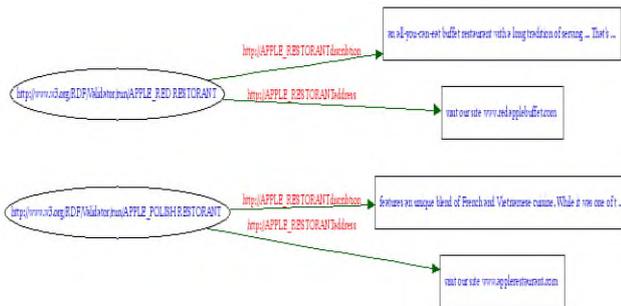


Fig.6. Graphical presentation of RDF file

Notable, we can easily reach to a particular and fully information depending on target word. For example, if we are looking for the word search "APPLE RESTAURANT" then the results that appear will contain only all the information describing "APPLE RESTAURANT".

V. CONCLUSIONS

XML is mostly concerned with syntax that does not make sense without semantics and can specify the structure of documents, not the meaning of the document contents. While Resource Description Framework (RDF) provides a standard for describing resources on the Web which gets us into metadata (data about data) and that is where things start getting particularly Semantic and exponentially more exciting.

VI. REFERENCES

- 1) Stefan Decker, Frank van Harmelen, Jeen Broekstra, Michael Erdmann, Dieter Fensel, Ian Horrocks, Michel Klein & Sergey Melnik. (2001). The Semantic Web - on the respective roles of XML and RDF, 13, 44, 53231.
- 2) Stefan Decker, Sergey Melnik, Frank Van Harmelen, Dieter Fensel, Michel Klein, Jeen Broekstra, Michael Erdmann & Ian Horrocks. (2000, Sep./Oct.). The Semantic Web: The Roles of XML and RDF, IEEE Internet Computing, vol. 4(5), pp. 63-74.
- 3) Martha M. Yee, (2009, June). Can Bibliographic Data be Put Directly onto the Semantic Web? Information Technology and Libraries, University of California.
- 4) Enrique Chavarriaga & José A. Maciac. (2009). A model-driven approach to building modern Semantic Web-Based User Interfaces, Elsevier Ltd.

- 5) Nazaraf Shah, Chen-Fang Tsai¹, Milko Marinov², Joshua Cooper³, Pavel Vitliemov⁴ & Kuo-Ming Chao. (2009) Ontological On-line Analytical Processing for Integrating Energy Sensor Data", IETE Technical Review, Vol 26(5).
- 6) Jorge Martinez-Gil, Enrique Alba & Jos_e F. Aldana-Montes. Optimizing Ontology Alignments by Using Genetic Algorithms, Boulevard Louis Pasteur s/n 29071 M_alaga (Spain).
- 7) Payam .M. Barnaghi. (2007, April) Automatic Hypermedia Presentation Generation System, University of Malaya.
- 8) Payam Barnaghi. (2008). Introduction to the Semantic Web the University of Nottingham, Malaysia Campus.
- 9) W3C Semantic Web. Retrieved from: <http://www.w3.org/2001/sw/>
<http://www.w3.org/TR/rdf-concepts/>