

Resource Description Framework (RDF) for Organised Searching on Internet

Aditya Tripathi

Abstract

This article focuses on the search performance of the existing search engines on internet. Discusses few solutions for the current searching problems, and suggests the resource description framework (RDF) for organised searching. The paper also includes the usefulness of RDF schema and covers the futuristic vision of RDF.

1. INTRODUCTION

The existing World Wide Web (WWW) has become a platform where one finds information of his/her interest ranging from household to serious scientific literature. But finding information on Web is a matter of luck, often it is like "one puts a hand in the heap of garbage to find a gold coin". One solution to this lies with searching through subject specific information ports or subject gateways. But the question is how to find such ports to take off.

1.1 Understanding the Problem (An Analogy With Real World)

Suppose Mr. Y is to be found out in a group of people. One way of doing is going to everybody and mark the checked ones. Is it a feasible idea? On internet, something like this is the situation with search engines where they produce a huge bunch of uniform resource locators (URLs). It is desired for surfer to go through the long list of URLs. During searching process, checked URLs turn into different colour and user has to spend enormous time in searching for relevant information on the internet.

1.2 Solution in Real World

If we take some parameters to categorise people based on, age, sex, height, etc.,

obviously it becomes easy to find Mr Y. Now the question arises, can we do the same process on web also?

Generating subject specific ports is one of the approaches to tackle this problem. But such ports are soon lost and again it is difficult to find such ports, as these ports will also become the part of internet and get mixed with innumerable other web documents. Definitely inter-disciplinary subjects are a big matter of concern in retrieval.

2. METADATA: IT IS LIBRARIANS' TIME

Librarians are practising use of metadata since the very inception of library. One of its form can be seen in the subject categorisation and description of documents in a catalogue entry. The description of a document is done in terms of its title, author, place, publisher, call number, etc. It seems very easy in traditional form but the question often asked as how to implement it on the internet? Who is the cataloguer? Where to keep such a catalogue?, and is it a feasible solution?

There are thoughts about the problem and different solutions are suggested. A set of 15 elements of Dublin Core (DC) are defined, MARC tags are extended for web documents, usage of metatags are suggested in HTML. But

the problem remains same as how to make search engines understand that a user is asking for a particular document, and even whether the tag set suggested by DC or MARC is sufficient for document description. Even as the standards mentioned above are being widely discussed and pushed, then why did MathML (Mathematical markup language)² and CML (Chemical markup language)³ came into existence?

3. WHAT NEXT?

Generation of new subject specific markup language highlights two aspects of the failure of DC and MARC. Firstly, the non-implementation of existing schemas like DC and MARC or non-standardisation of web-document description. Consequently, most of the user communities are oblivious to these developments. Secondly, subject specific user community may not be very happy with existing schemas like DC or MARC because these schemas cannot define completely the specific information conveyed within the document with few limited tags or in other words it highlights the inefficiency of these schemas to map a specific subject.

However, situation will become worse when each subject domain will have its own markup language schema. Coordinating among these schema is a major problem. Therefore, definitely a platform is required with a mechanism of standard description of document and RDF serves this purpose.

4. DEVELOPMENT OF RDF

Development of RDF started with the initiation of Platform for internet content selection (PICS) project in 1995. PICS was a

rating mechanism about the contents of web pages. The idea was to filter the unwanted set of web pages, which contain foul language, pornographic material, violence, etc. When the project was initiated it was found that it can be used for describing the content of web page. Later on it was made to represent content understandable by machines. The extension of PICS project was PICS next generation (PICSNG), which was later called as resource description framework. The idea was to use semantic nets to describe the web resources.⁴

5. MODELLING DATA IN RDF

Representation of data through RDF is very easy as it follows semantic nets. A simple RDF model has three parts.

- (a) Resource: Any entity which has to be described is known as *resource* which is equivalent to *subject* in normal English grammar. It can be a 'web page' on internet or a 'person' in a society.
- (b) Property: Any characteristic of resource or its attribute which is used for the description of the same is known as *property*, which is equivalent to *predicate* in normal English grammar. For example, a web page can be recognised by 'Title' or a man can be recognised by his 'Name'. So both are attributes for recognition of resource 'web page' and 'person' respectively.
- (c) Value: A *property* must have a value which is equivalent to *Object* in English grammar. Like, the title of DRTC web page is 'Documentation Research and Training Centre', name of a person is 'Ranganathan'.

Fig.1 illustrates simple pictorial data representation in RDF.

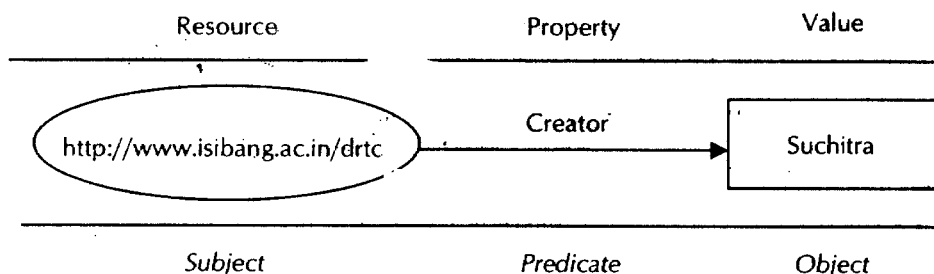


Figure 1. Simple pictorial data representation in RDF

A resource can have an identifier. It is called a Uniform resource identifier (URI). It can be a URL also. A URI means an identifier by which a resource is uniquely identified.

A property itself can be a 'resource', which makes a complex representation model of the data. For example, say Creator of the web page <http://drtc.isibang.ac.in> is 'Suchitra' who is project assistant at DRTC and has an email suchi@isibang.ac.in as represented in Fig. 2.

There may be more than one Creator of the webpage of DRTC. An even more complex system can be generated using other metatag: schemas. Fig. 3 gives the pictorial representation of DRTC site (<http://drtc.isibang.ac.in>).

6. SYNTAX FOR RDF

Basically RDF follows a container package model. A container can have many packages or a package itself can be a container.⁵

RDF follows syntax. The major problem with XML is that anyone can create a set of tags which is a hindrance for the structural standardisation and generation of semantics of a web document. To have standardisation concept of metadata, RDF is a sort of harmonisation of various metadata schemas, in other words it acts as a platform for efficient search and retrieval of web documents.

6.1 Harmonisation: Through Namespace

RDF uses 'namespace' concept for identification of metadata schemas. A 'namespace' is a standard identifier which suffixes a tag to point out to which schema the tag belongs. To use the metadata schema within a document it is must to provide the

location of definition of elements of schema. This can be done using a particular statement within the RDF document.

```
xmlns:dc=http://purl.org/dc/elements/1.0/
```

There, the RDF document uses 'dc' as 'namespace' for the definitions and is available at the URL <http://purl.org/dc/elements/1.0/> where definition of DC elements is given. So, RDF document is supposed to read definitions from the DC website.

In XML, a 'namespace' is represented in the tag as follows:

```
<dc:creator>Suchitra</dc:creator>
```

The above statement means that 'Suchitra' is the creator of mentioned document and the tag <creator> is a Dublin core element. Therefore, in a document one can define as many 'namespaces' as required and use them. But the whole document will be enclosed in a RDF tag, and inside a <rdf:Description> which represents that the document is a RDF document.

For example, a full-fledged RDF entry with DC as metadata schema mentioned in Fig. 3 can be rendered as:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-schema-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:dc="http://purl.org/dc/elements/1.0/">
<rdf:Description about=
http://drtc.isibang.ac.in">
<dc:title>
```

```
Welcome to Home Page of Documentation
Research and Training Centre
```

```
</dc:title>
<dc:creator>Aditya Tripathi</dc:creator>
<dc:language>English</dc:language>
<dc:description>DRTC is an R & D organisation
```

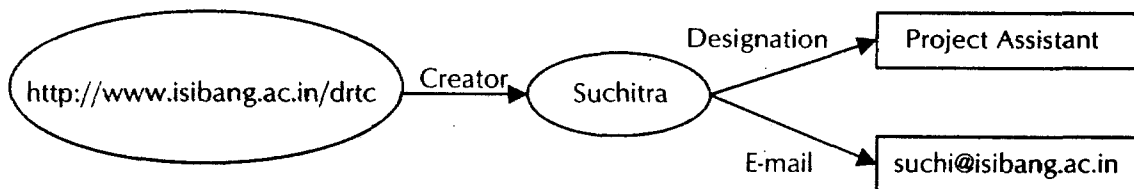


Figure 2. Complex RDF model

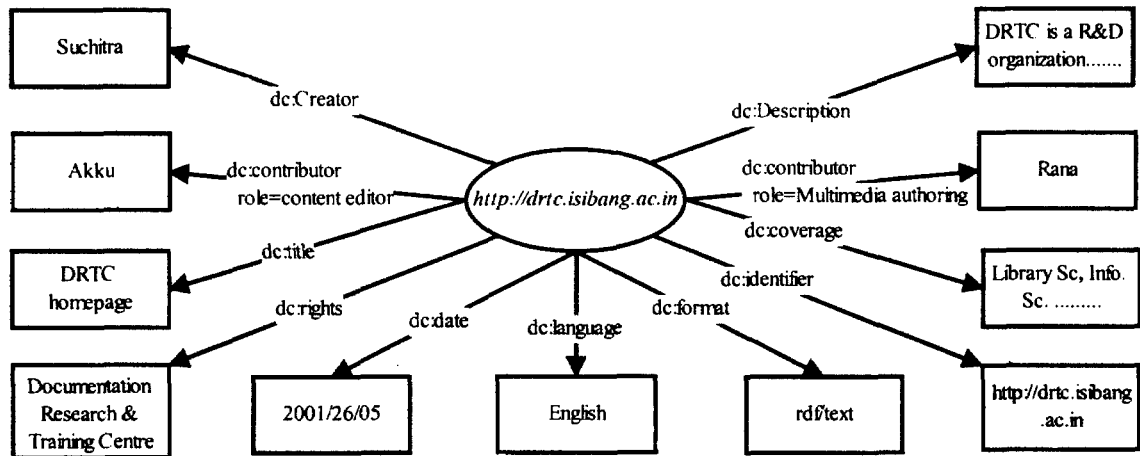


Figure 3. Pictorial representation of RDF format of a site using DC

in the field of Library and Information Science. It was established by Dr. S.R. Ranganathan in 1962 at Bangalore.

```

</dc:description>
<dc:format>rdf/text</dc:format>
<dc:date>2001/26/05</dc:date>
<dc:coverage>Library Science, Information
Science, Activities of DRTC
</dc:coverage>
<dc:contributors role="Multimedia authori-
ng">Rana</dc:contributors>
<dc:contributors role="Content
Editor">Akku</dc:contributors>
<dc:identifier>http://drtc.isibang.ac.in</dc:identi-
fier>
<dc:rights>Documentation Research and
Training Centre</dc:rights>
</rdf:Description>
</rdf:RDF>

```

RDF has already defined few tags for the document description. One of the tags is 'Description', which is represented as <:Description>, and shows that the packages inside the <rdf:Description> is the description of the document. XML provides the facility of defining attributes. So <rdf:Description> has an attribute about which says that this RDF document is a description of 'http://drtc.isibang.ac.in'.

7. IMPORTANCE OF RDF FOR FUTURE OF WEB

Currently the web is a mixture of noise and useful information. It is often difficult to extract the relevant information out of it. The model of RDF has been created keeping a broad idea about the future of the web.

- (a) There is no semantic value in current search engines. One of the objectives of RDF is to make machines understand the semantics of the data within a web document. That means tomorrow if one searches for the term 'computer drivers' the search engine will not retrieve information as, 'taxi drivers', 'truck drivers' or 'screw-drivers'.
- (b) Tomorrow's web will be having a wide range of applications. One can 'switch on' the home microwave machine from his/her office. The appliance will have intelligent agents which will function according to the program for which they are programmed. For example, if the washing machine could not identify a particular material of cloth, the agent will collect the data from the server of the manufacturer and find the type of washing the clothe requires. All these functions will be done by the agents which will work behind the scene. It is also true that all house appliances will also have the IP address and not only that, they will have all

the data which will distinguish them from rest of the world and this will be done using metadata and in RDF format. Perhaps, this is a far-reaching scenario.

8. HOW LIBRARIES ARE BENEFITED

Spread of internet has made almost a sort of compulsion for the libraries to put their collection on the web or on local intranet. Definitely RDF provides librarians a unique way to describe their collection and library both.

Since RDF uses XML encoding, it can be very beneficial for the data interchange among the libraries and extraction of data can be done from databases and promises standardisation for data interchange.

9. CONCLUSION

It is true that internet is the biggest find of the century, but its unmanageable growth has played havoc in retrieval. Particularly to those who are in the profession of information storage, retrieval and dissemination it poses a big problem. It is equally true that this growth has made many libraries and information centres to go online but the danger is always

there that a useful collection may get mixed with unwanted collection and thus discovery of information becomes the primary issue. RDF is a format which can be used for efficient search and retrieval. It has been initiated out of the PICS project and just 2-3 years old, it means still it has a long way to go. Already tools are available on the web which can be used to write RDF documents without much hassle.

REFERENCES

1. Dublin core metadata initiative.
<http://dublincore.org/documents/1998/09/dces/#>
2. Mathematical Markup Language (MathML) 1.01 Specification
<http://www.w3.org/TR/REC-MathML/>
3. CML: Chemical Markup Language.
<http://www.ch.ic.ac.uk/cml/>
4. Miller, Eric. An introduction to the resource description framework.
<http://www.dlib.org/dlib/may98/miller/05miller.html>
5. Metadata: An overview.
<http://www.nla.gov.au/nla/staffpaper/cathro3.html>

Contributor: Shri Aditya Tripathi is Junior Research fellow at Documentation Research & Training Centre, Bangalore.
--