Automatic RDF generation using RDF automata algorithm

K. Srihari, Dr. V.P. Arunachalam and Dr. S. Karthik

Lecturer, Department of CSE, Sri Ramakrishna Institute of Technology, Coimbatore, Tamilnadu, India
Supervisor, Principal, SNS College of Technology, Coimbatore, Tamilnadu, India
Joint Supervisor, Dean, Department of CSE, SNS College of Technology, Coimbatore, Tamilnadu, India

ABSTRACT

In this fast moving world searching is the important and vital field where each and every activity involved in it. The semantic search adds special flavor to the normal searching technique. It improves accuracy, amount of time for searching is reduced and the knowledge about searching is also improved. In this paper a technique for the search engine which involves logic based retrieval and relationship are based on RDF. In current RDF the relationship are based on static and manual methods for obtaining relations. We propose a method which automatically generates the logical relation with the help of input URL. The URL is given as input and automatic RDF is generated. The RDF which is generated with the help of URL can be optimized.

Key words: RDF graph, semantic search, algorithms.

Introduction

Problem Identification:

The existing RDF generally created with the help of static or assumed or predicted relations manually made by human. The relations are made based on certain assumptions and certain boundary conditions. This is the main drawback of existing system. In many cases the relationship done is not relevant and meaningless. In such cases a new knowledge based, automatic RDF graph should be generated. The graph generated based on the URL address.

Basic Idea:

Based on the given URL as the input, an automatic graph is generated is the proposed solution. Here a graph (relation based) is drawn or designed. Semantic search engine with URL www.researchsrihari.replicoders.com is created and automatic RDF is generated. So this proposed idea improves relations, meaning and knowledge based semantic search engines

Proposed Solution:

RDF graph Generation:

The current method involves static RDF creation for example, Consider the existing semantic search like swoogle. Here if hotels in India are given as search it will produce a result as 1) Taj 2) Aloft 3) Residency. The searching begins as first control points to a RDF graph where manual relations are designed. It searches the RDF graph and retrieves the word that is typed.

Manual RDF Graph for hotels in India query. Hotels near railway station, hotels near airport. Stars five, three, two

1) So relations are manually designed as said before which is purely based upon the assumptions by the designer not based on quality.
2) Here many quality based information may not be available. So in order to improve quality based relations a new idea is proposed which improves the following parameters.
3) Relationship (R)
4) Quality of web content (Qw)
5) Automatic RDF graph (AR)  
The proposed system gets the URL as input and process the URL in order to get the RDF output.
1) Algorithm for RDF (Automatic)  
3) RDF automata are generated.  
4) Comparison of relationships (R), Qw, between manual and generated. automatic RDF  

Algorithm for RDF (Automatic):
//P parent word  
// C child word  
If P=null, C=max  
If n = max then c=null  
If n=max then Pn = n  
If n =min then Cn=n  
Count n no of words (W1….Wn)  
If Wn occurred n times  
Then Wn = parent  
Wn-1 = Child  
Repeat until all words are traversed  
Parent = n times occurrences of word.  
2nd less max = child of max (p).  
Repeat until parent= Min no of occurrence of words.  

* Relate all the words.  
* The relationship may be for,and ,an,after,before,and n no of relations.  
* So maximum no of relations are obtained, which can be over written in the existing method.  
* R (proposed) always greater than R (existing)  
* matches all the possible relationship between parents and child  
* Tabulate all the relationship as follows.  
* Search for hotels in India is done in www.researchsrihari.replicoders.com and we got many relations that is tabulated as follows.  
* Factors such as hotels near to airport and railway station is done  
* Next the distance between the hotels is taken into account for relationship  
* Stars of hotel is also taken int account  
* These above said relations are automatically made and automatically generated based on the algorithm 1 as described above.  
* Based on the results the following graph is tabulated.  

<table>
<thead>
<tr>
<th>Table 1:</th>
<th>Parent (P)</th>
<th>Child (C)</th>
<th>Relation (and)</th>
<th>Relation (the)</th>
<th>Relation (for)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>C1</td>
<td>P1C4</td>
<td>P1C4</td>
<td>P1C1</td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>C2</td>
<td>P1C3</td>
<td>P1C3</td>
<td>P2C3</td>
<td>P4P3</td>
</tr>
<tr>
<td>P3</td>
<td>C3</td>
<td>P1P2</td>
<td>P4P2</td>
<td>P1P2</td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>C4</td>
<td>P1P4</td>
<td>P1P4</td>
<td>P3C4</td>
<td></td>
</tr>
</tbody>
</table>

*Here P1….Pn denotes the parent as designed based on algorithms and C1…Cn are child designed based on the algorithm with respect to hotel domain.  
* After finding relation, the relations are ordered to find the maximum count.  
* If pncn= max order from PnCn till last relation.  

RDF automata are generated  
* The maximum relation obtained from table1 is elaborated and a graph is generated  
* Maximum relation  
* Second maximum relation is calculated  
* Third maximum relation is calculated  
* Nth maximum is calculated.  
* So automatic RDF is generated based on maximum relationship.  

Case Study:  
* RDF (automatic) for certain URL obtained from the result of searching doneon www.researchsrihari.replicoders.com

Experimental Results:  
RDF tree from above said web site.
The automatic RDF generation for railways ontology and yahoo ontology are displayed above respectively are formed automatically with the help of automatic RDF algorithm. The results are referred with graph (relations are made) www.rhizomik.net [4]. So combined graph from RDF automata algorithm and www.rhizomik.net [4] are displayed above. The problem with the the existing RDF the relations are assumed and predefined. So automatic RDF is created. The comparison results are explained as follows.

Comparison of relationships (R), Qw between manual and generated automatic RDF.

So here swoogle is considered for comparing the proposed algorithm and concepts. The parameters such as relationship R, Qw are compared and the difference between the two is found with the help of graph and tabular column. The tabular column is as follows.

![Diagram]

Fig. 2: [4]
Table 2:

<table>
<thead>
<tr>
<th>category</th>
<th>Proposed RDF automata</th>
<th>Swoogle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of good pages</td>
<td>Relationship R</td>
</tr>
<tr>
<td>Replicoders (proposed)</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Common</td>
<td>5</td>
<td>25.2</td>
</tr>
<tr>
<td>Health</td>
<td>10</td>
<td>33.3</td>
</tr>
<tr>
<td>Travel</td>
<td>20</td>
<td>50.0</td>
</tr>
<tr>
<td>Education</td>
<td>50</td>
<td>66.6</td>
</tr>
</tbody>
</table>

Fig. 3:

Table 3:

<table>
<thead>
<tr>
<th>category</th>
<th>Proposed RDF automata</th>
<th>Swoogle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of good pages</td>
<td>Qw</td>
</tr>
<tr>
<td>Replicoders (proposed)</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Common</td>
<td>5</td>
<td>25.2</td>
</tr>
<tr>
<td>Health</td>
<td>10</td>
<td>33.3</td>
</tr>
<tr>
<td>Travel</td>
<td>20</td>
<td>50.0</td>
</tr>
<tr>
<td>Education</td>
<td>50</td>
<td>66.6</td>
</tr>
</tbody>
</table>

Fig. 4:
Conclusion:

Thus an automatic RDF is generated for limited websites. In future a more improved procedure will be developed for complex web pages to generate RDF graph. Nearly 100 web sites will be taken into account and RDF graph is generated. Here in this proposal three parameters are improved such as Relationship, quality of web pages, and RDF automatic generation.

References

Dipt di ing, applicata university of salerno, Fisciano, Italy. IEEE transaction on systems, man and cybernetics Part A: System and humans.
www.rhizomik.net - RDF reference for relations.