A Framework For An E-Learning System Based on Semantic Web

Tarek M. Mahmoud
Faculty of science
Computer Sci. Dept.
Minia University
El Minia, Egypt
d.tarek@minia.edu.eg

Tarek Abd-El-Hafeez
Faculty of science
Computer Sci. Dept.
Minia University
El Minia, Egypt
tarek@mu.edu.eg

Ahmed Badawy
Faculty of science
Computer Sci. Dept.
Minia University
El Minia, Egypt
ahmed.badawy2000@mu.edu.eg

Abstract- **E-Learning is efficient, task relevant and just-in-time learning grown from the learning requirements of the new, dynamically changing, distributed business world. The term "Semantic Web" encompasses efforts to build a new WWW architecture that enhances content with formal semantics, which enables better possibilities for navigating through the cyberspace and accessing its contents.**

Semantic Web is a product of Web 2.0 (second generation of web) that is supported with automated semantic agents for processing user data to help the user on ease of use and personalization of services.

The paper describes a framework for the implementation of an e-learning system based on the Semantic Web, using desktop c# application, data visualization tools, HTML web page parser, RDF generating tools, and SPARQL RDF query language. Hopefully we have elucidated the enormous potential of making web content machine-understandable.

I. I. Introduction

During the past few years the World Wide Web has become the biggest and most popular way of communication and information dissemination. It serves as a platform for exchanging various kinds of information, ranging from research papers, and educational content, to multimedia content, software and personal logs. Every day, the web grows by roughly a million electronic pages, adding to the hundreds of millions pages already on-line. Because of its rapid and chaotic growth, the resulting network of information lacks of organization and structure. Users often feel disoriented and get lost in that information overload that continues to expand. [1]

Increasingly, The World Wide Web (WWW) is used to support and facilitate the delivery of teaching and learning materials. This use has progressed from the augmentation of conventional courses through web-based training and distance learning to a newer form of WWW-based education, e-learning. E-learning is not just concerned with providing easy access to learning resources anytime, anywhere, via a repository of learning resources, but is also concerned with supporting such features as personal definition of learning goals, synchronous and asynchronous communication, and collaboration between learners and between learners and instructors . [2]

It is clear that new styles of learning are some of the next challenges for every industry. Learning is a critical support mechanism for organizations to enhance the skills of their employees and thus the overall competitiveness in the new economy. The incredible velocity and volatility of today's markets require just-in-time methods for supporting the need-to-know of employees, partners and distribution paths. It is also clear that this new style of learning will be driven by the requirements of the new economy: efficiency, just-in-time delivery and task relevance.

Time, or the lack of it, is the reason given by most businesses for failing to invest into learning. Therefore, learning processes need to be efficient and just-in-time. Speed requires not only a suitable content of the learning material (highly specified, not too general), but also a powerful mechanism for organizing such material.

ELearning aims at replacing old-fashioned time/place/content predetermined learning with a just-in-time/at work-place/customized/on-demand process of learning. ELearning needs management support in order to define a vision and plan for learning and to integrate learning into daily work. It requires changes in organizational behavior establishing a culture of "learn in the morning, do in the afternoon". Thus, an IT platform, which enables efficient implementation of such a learning infrastructure, is also needed. Current Web based solutions don't meet the above mentioned requirements. Some pitfalls are e.g. information overload, lack of accurate information or content that is not machine-understandable. The new generation of the Web, the so-called Semantic Web, appears as a promising technology for implementing eLearning. The Semantic Web constitutes an environment in which human and machine agents will communicate on a semantic basis [3].

II. E-LEARNING

"E-Learning is just-in-time education integrated with high velocity value chains. It is the delivery of individualized, comprehensive, dynamic learning content in real time, aiding the development of communities of knowledge, linking learners and practitioners with experts" [4]

Standard or traditional learning processes can be characterized by centralized authority (content is selected by the educator), strong push delivery (instructors push knowledge to students), lack of a personalization (content must satisfy the needs of many) and the linear/static learning process (unchanged content).

Differences between training and eLearning are represented in table 1 [4]

Table 1: training VS eLearning

Dimensions	Training	E-Learning
Delivery	Push – Instructor determines agenda	Pull – Student determines agenda
Responsiveness	Anticipatory – Assumes to know the problem	Reactionary – Responds to problem at hand
Access	Linear – Has defined progression of Knowledge	Non-linear – Allows direct access to knowledge in whatever sequence makes sense to the situation at hand
Symmetry	Asymmetric – Training occurs as a separate Activity	Symmetric – Learning occurs as an integrated activity
Modality	Discrete – Training takes place in dedicated chunks with defined starts and stops	Continuous – Learning runs in the parallel to business tasks and never stops
Authority	Centralized – Content is selected from a library of materials developed by the educator	Distributed – Content comes from the interaction of the participants and the educators
Personalization	Mass produced – Content must satisfy the needs of many	Personalized – Content is determined by the individual user's needs and aims to satisfy the needs of every user
Adaptivity	Static – Content organization/taxonomy remains in their originally authored form without regard to environmental changes	Dynamic – Content changes constantly through user input, experiences, new practices, business rules and heuristics

The principle behind eLearning is that the tools and knowledge needed to perform work are moved to the workers – wherever and whoever they are. Simply put, eLearning revolves around people.

III. SEMANTIC WEB

The current WWW is a powerful tool for research and education, but its utility is hampered by the inability of the user to navigate easily the nefarious sources for the information he requires. The Semantic Web is a vision to solve this problem. It is proposed that a new WWW architecture will support not only web content, but also associated formal semantics. [2]

The Semantic Web, in practice, comprises a layered framework: an XML layer for expressing the web content; a Resource Description Framework (RDF) [5] layer for representing the semantics of the content; an ontology layer for describing the vocabulary of the domain; and a logic layer to enable intelligent reasoning with meaningful data. [2]

The meaning of the document content is expressed with RDF that is simply a data model and format that allows the creation of machine-readable data. It comprises a set of triples, i.e., three Universal Resource Identifiers (URIs) that may be used to describe any possible relationship existing between the data – subject object and predicate [6]. Thus, all data stored in the system is easily readable and process able. [2]

RDF is the underlying unified data model for representing semantics. The data model and XML serialization syntax is used for describing resources both on and off the Web. RDF makes use of unique identifiers (URI, Uniform Resource Identifier) for describing metadata. URIs are used to describe things, also called resources, which could represent people, places, documents, images, databases, etc. All RDF applications adopt a common convention for identifying these things. A subset of URI, the Uniform Resource Locator or URL, is concerned with the location and retrieval of resources, while URI is a unique identifier for things or resources that we describe but that may not necessarily be retrievable. However, RDF provides a consistent, standardized way to describe and query internet resources. RDF is based on three core concepts: resources, properties, and statements. A resource is the subject or the object of an RDF description. A property describes a resource; its value is either a literal or a reference to another resource. A statement combines the property of a resource with a value. It can be represented as a subject, a predicate, and an object - as in figure 1- where subject and object may be URIs and predicate is the property linking the object to the subject [7].

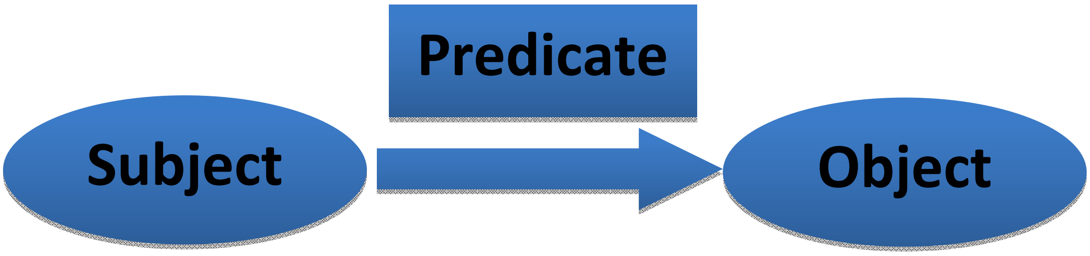

Figure 1: RDF triple

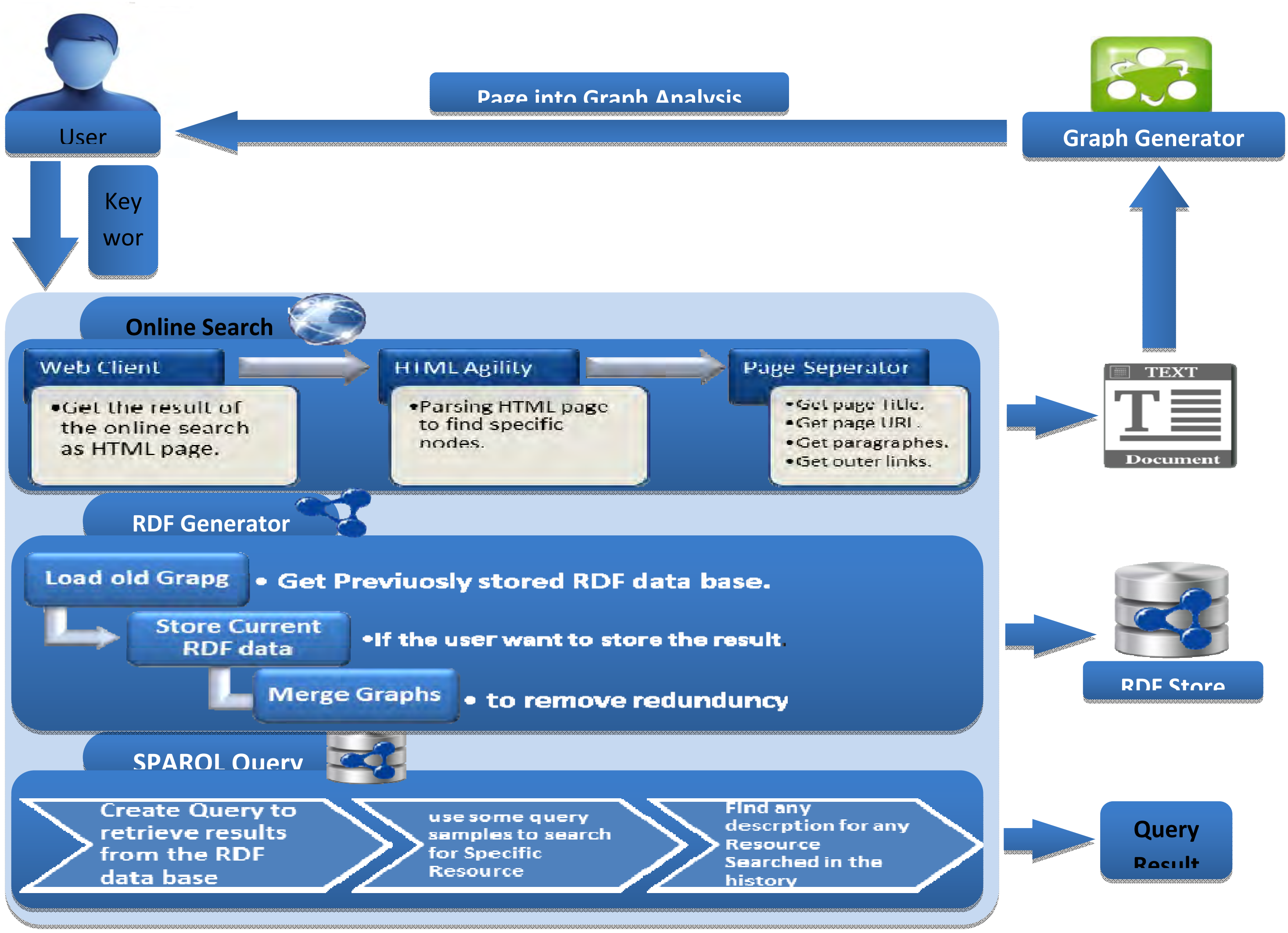

Figure 2: The Proposed System

Thanks to semantic web it makes information more meaningful to people by making it more understandable to machines. It makes current web smart enough to help us organize and evaluate the learning systems. Everybody aware that the more information the better. The architecture of the proposed system is presented in figure 2:

V. SYSTEM DESCRIPTION

The technology and economy of today have intensified the need for more efficient, faster, and larger-scale information management solutions.

The system includes the following three tiers:

Online search: provides the main search mechanism over the web in order to produce the required information resources and store it into the RDF data base, and generate the Graph# web page structure.

RDF generator: provides the semantic web part of the system where the web page content analyzed and converted into RDF triples and stored into the RDF data base after removing the repeated data.

SPARQL query: provides the semantic query mechanism where the user executes any SPARQL query over the previously stored RDF data base.

In this section we describe some of the main features of our e-learning application that are enabled with Semantic Web technologies.

A main concept of our application is that it enables dynamic Resource creation and extension. The user can enter any key word to search the internet and the result web page will be analyzed and its content will be extracted as in figure 3, so the system can produce the content Visualization graph and generate the RDF triples of the resources result from the search. This means that any user can browse any topic without any constraints, and the result will be always up to date.

Figure 3: Online Search

VI. HOW THE SYSTEM OPERATE:

1-Online Search:

As the previous figure shows, we have an interface that take from the user a keyword to search Wikipedia. The result web page will be analyzed to extract its contents (Paragraphs and related links). Every paragraph has its own number in the page and its internal links. After extracting page contents the system construct a text file to represent the page visual graph with Graph# [8] as shown in figure 4.

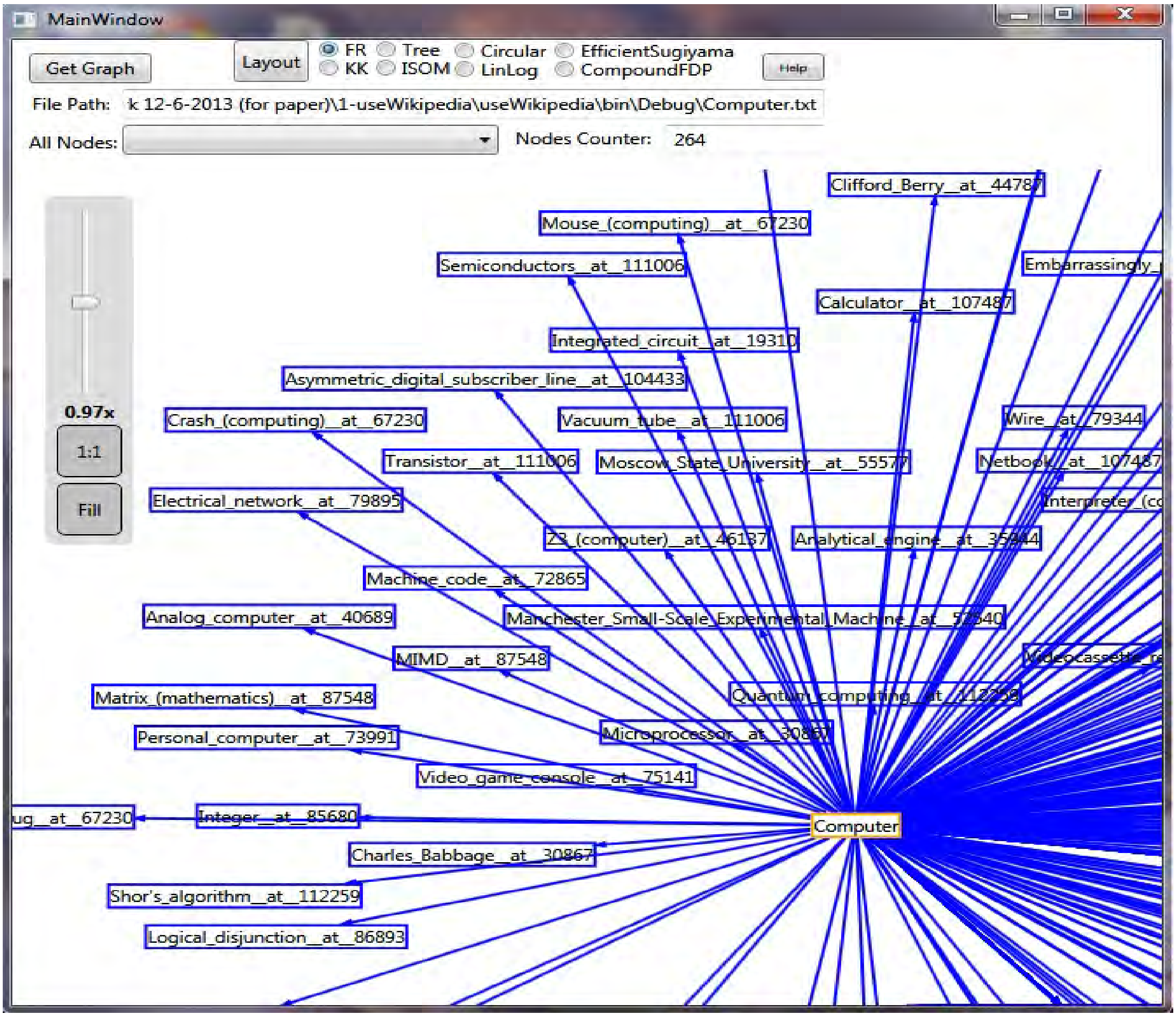

Figure 4: Data Visualization

In the Graph# [8] interface the user can see a set of nodes that are connected together with edges, the root node is the main search topic, and the other nodes are the paragraphs inner links or related topics. If the user R-click on one of the nodes a message box contains the related paragraph will appears as shown in figure 5.

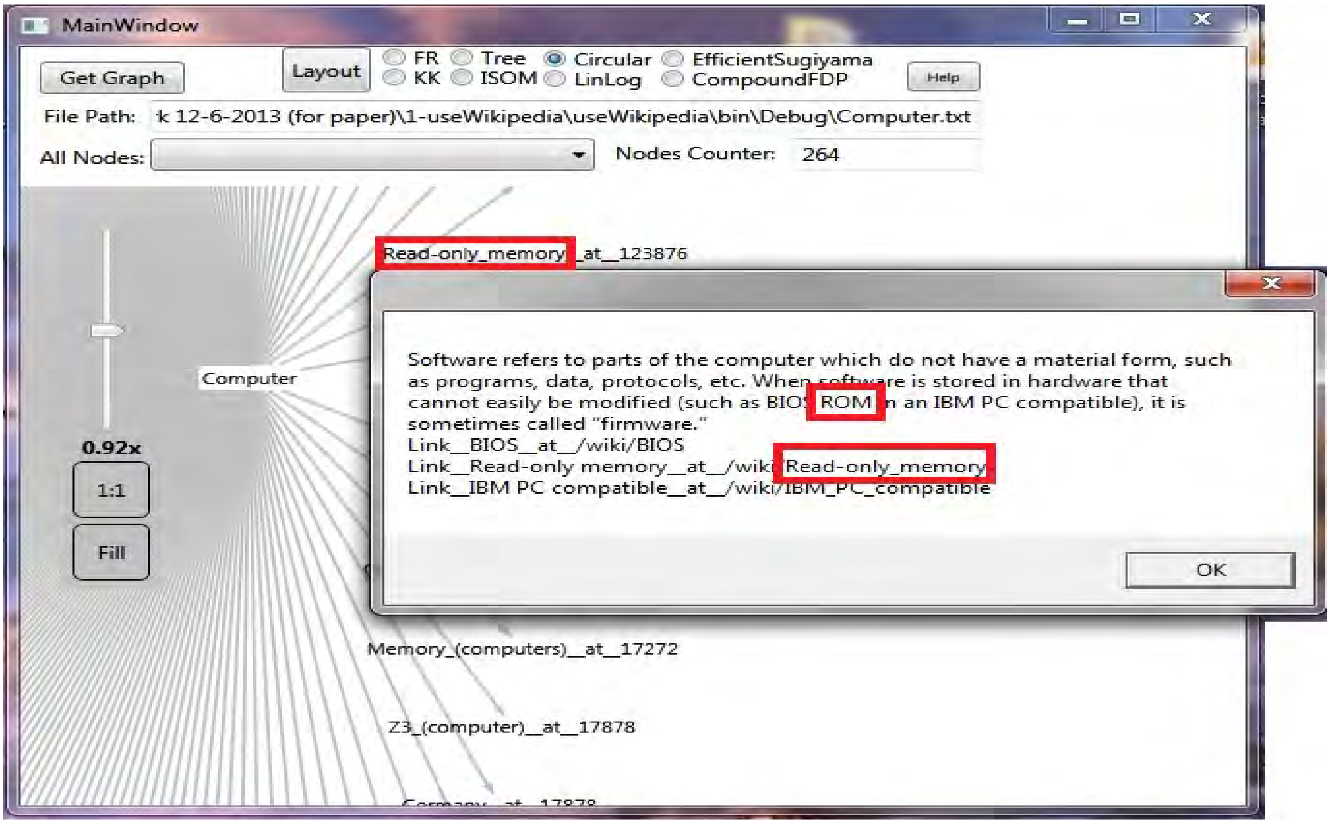

Figure 5: Select node from graph

The user also can select any node from the nodes dropdown list as shown in figure 6.

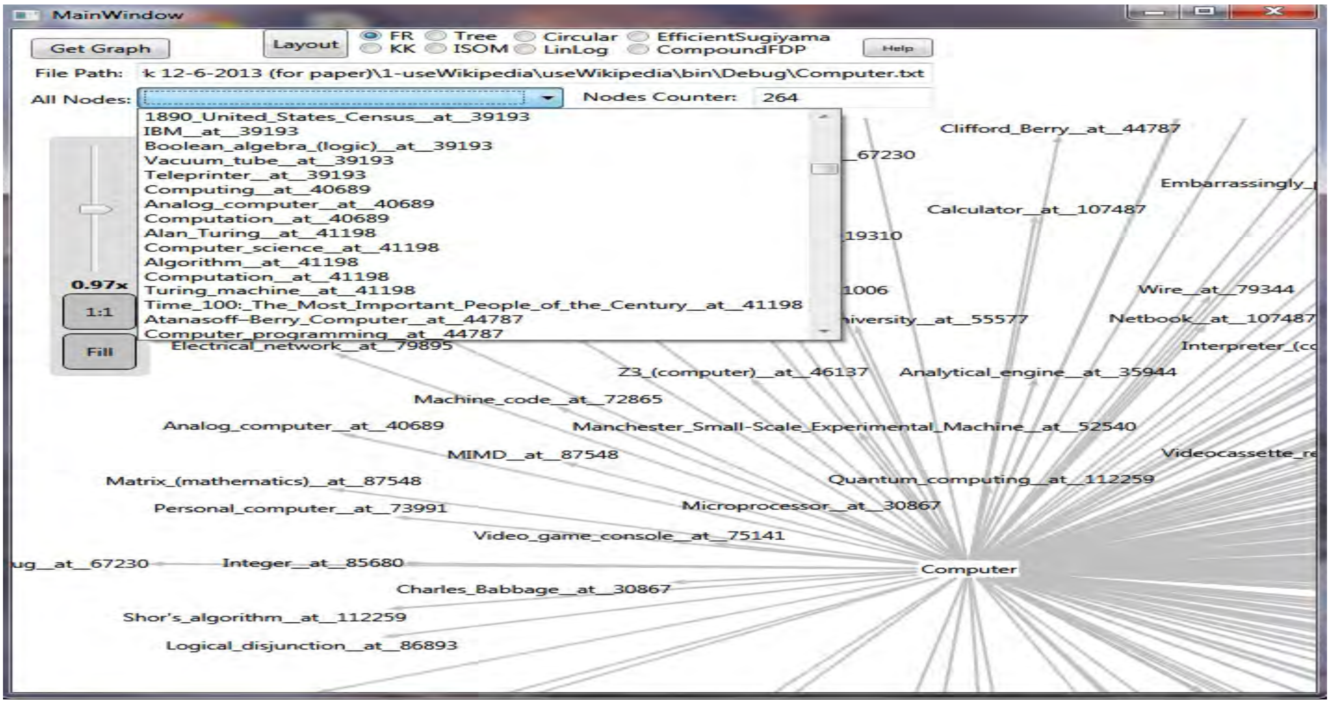

Figure 6: select node from nodes list

2-RDF generator

This part of the system represents the semantic web part, where the extracted content converted into RDF triples using dotnetRDF [9] and stored into the RDF data base. The page can be represented as the graph shows in figure

7.

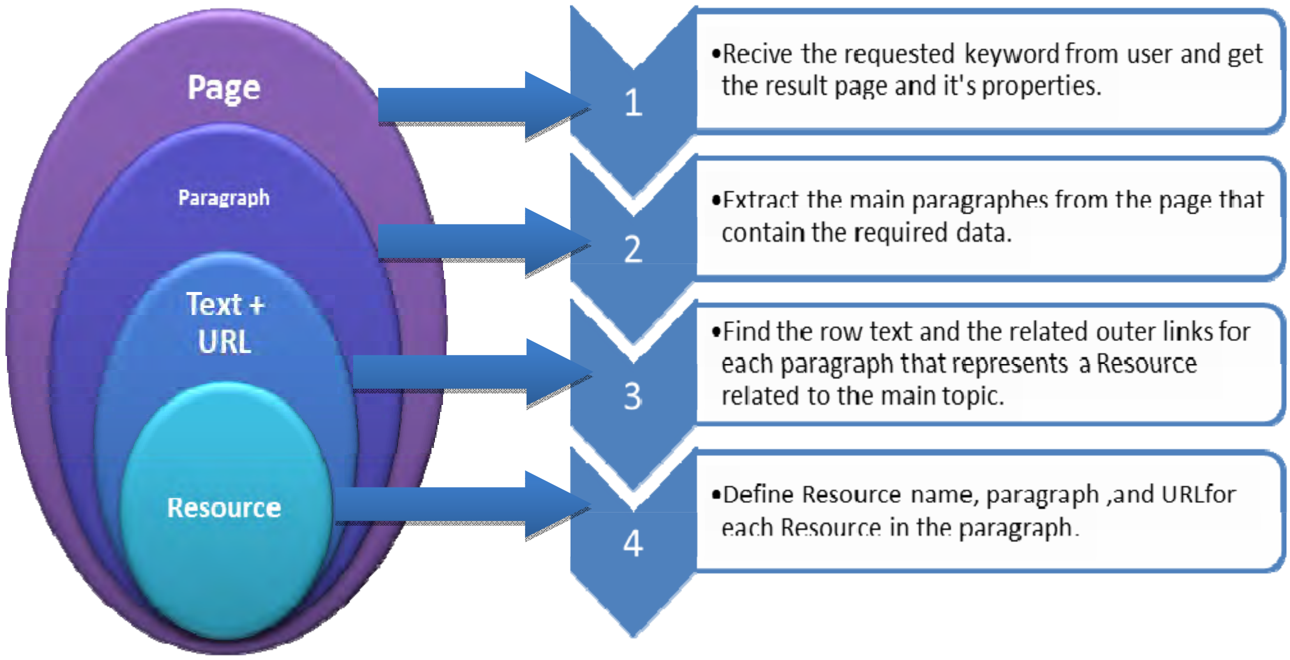

Figure 7: page representation

The RDF triples that the system produces to represent a page are shown in table 2:

Table 2: RDF triples

	Subject	Predicate	Object
1	<Resource class>	<Resource has URL property>	<Resource URL>
Ex:	<http://----.org/Resource> <http://----.org/ResourceURL> <http://en.wikipedia.org/wiki/Special:Search/Computing>		
2	< Resource URL >	< Resource has Description property >	<Resource Description>
Ex:	<http://en.wikipedia.org/wiki/Special:Search/Computing> <http://----.org/ResourceDescription> "Computing is any goal-oriented activity…..”		
3	<Resource URL >	< Resource has Title property >	<” Resource Title”>
Ex:	<http://en.wikipedia.org/wiki/Special:Search/Computing> <http://----.org/ResourceTitle> "Computing – Wikipedia…”		
4	<Resource URL>	<Resource has paragraph property>	<paragraph URL>
Ex:	<http://en.wikipedia.org/wiki/Special:Search/Computing> <http://----.org/paragraphURL> <http://en.wikipedia.org/wiki/Special:Search/Computing/12290>		
5	<paragraph URL>	<paragraph has internal link property>	<internal link URL>
Ex:	<http://en.wikipedia.org/wiki/Special:Search/Computing/12290> <http://badawi.org/contentURL> <http://en.wikipedia.org/wiki/Computer>		
6	<paragraph URL>	<paragraph has text property>	<”paragraph text”>
Ex:	<http://en.wikipedia.org/wiki/Special:Search/Computing/12290> <http://badawi.org/paragraphText> "Computing is any goal-oriented activity”		

The structure in figure 8 can sum up the previous triples.

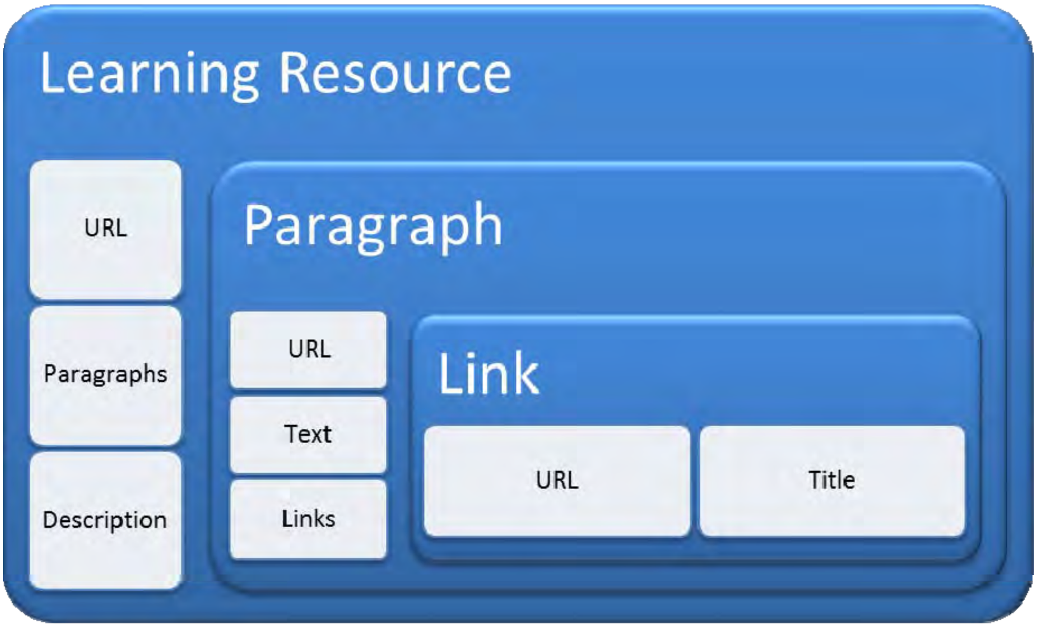

Figure 8: triples structure

2-SPARQL query:

RDF is a directed, labeled graph data format for representing information in the Web. The SPARQL query language for RDF can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware. SPARQL contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports extensible value testing and constraining queries by source RDF graph. The results of SPARQL queries can be results sets or RDF graphs. [10]

In this part the system use the “SPARQL” query language for RDF to execute any query over the previously stored data, using this technique the user can revise his search history, or other students can find data without online access, so that the program can learn from user’s search history.

Some SPARQL queries could be used in the system are listed in table 3.

Table 3: some SPARQL queries

	SPARQL query	Description
1	SELECT ?s WHERE { ?url bad:ResourceTitle ?s}	Get all Resource titles
2	SELECT ?s WHERE { ?url bad:ResourceTitle ?s FILTER (regex(?s, 'ss'))}	Search for Resource titles that contain the string “ss”
3	SELECT ?page ?text WHERE { ?page bad:paragraphText ?text FILTER (regex(?text, 'ss'))}	Search for Paragraph that contain the string “ss”
4	SELECT ?Resource ?Description WHERE { ?url bad:ResourceTitle ?Resource FILTER (regex(?Resource, 'ss')) . ?url bad:ResourceDescription ?Description }	Search for Resource description that contain the string “ss”

There are many SPARQL queries stored in the system ready for user use; as well the user can create his query as shown in figure 9.

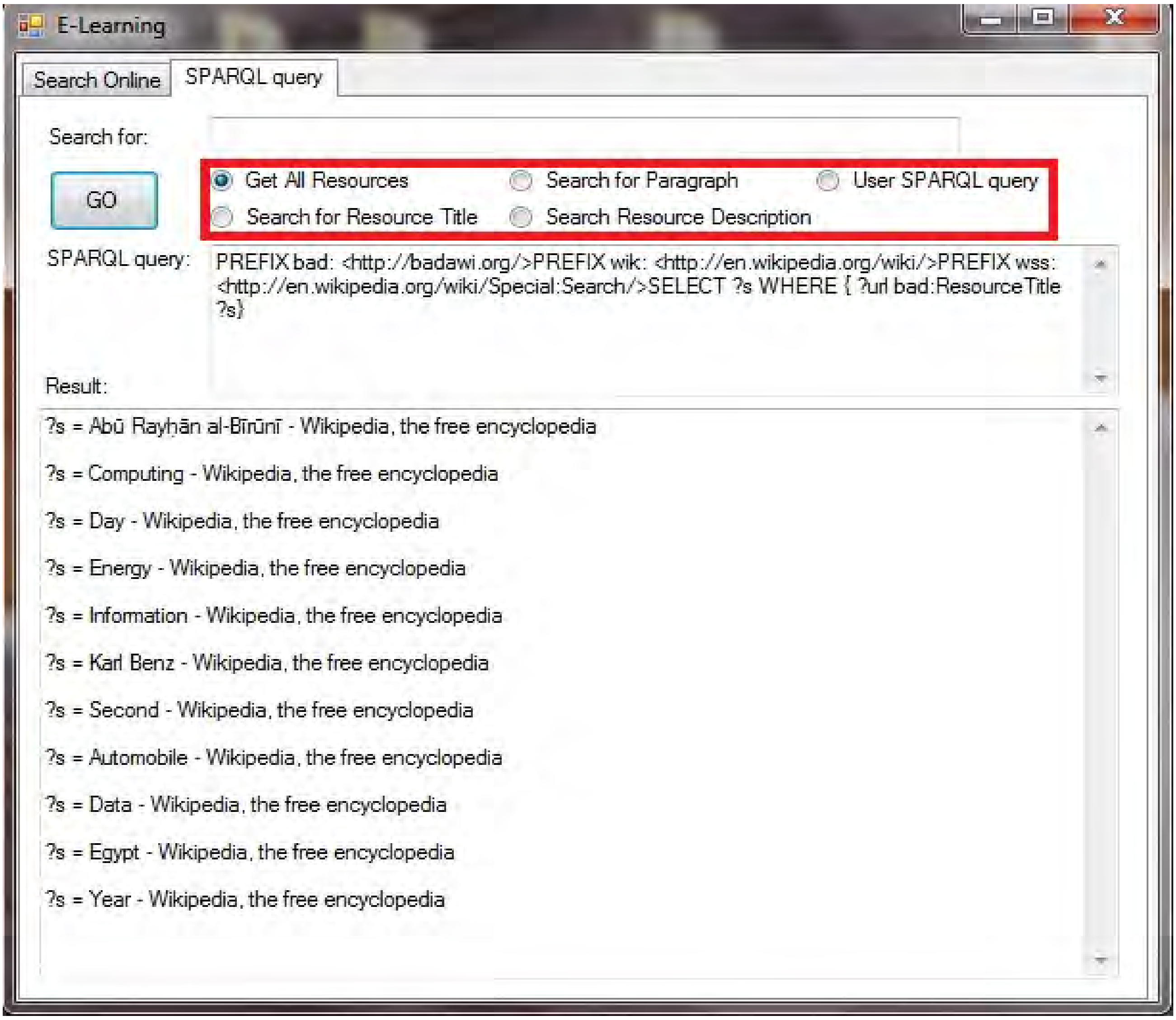

Figure 9: SPARQL query interface

VII. EXPERIMENTAL RESULT

In this section we present a table (table 4) contain number of online search operations, the keyword the user searched for, inner links, paragraphs, generated triples, and old triples in the RDF data base.

Table 4: Experimental Results

Search number	Searched keyword	Inner links	Page paragraphs	Previous RDF triples	New RDF triples
1	Science	320	64	0	462
2	knowledge	114	30	462	178
3	universe	507	71	640	664
4	scientist	153	19	1304	197
5	philosophy	622	90	1501	820
6	physics	397	62	2321	519
7	geology	277	71	2840	428
8	biology	332	48	3268	433
9	theory	115	33	3701	186
10	astronomy	401	87	3887	582
11	technology	265	49	4469	369
12	Socrates	183	57	4838	311
13	Aristotle	269	97	5149	468
14	teleological	66	25	5617	121
15	actualization	29	25	5738	88
16	Ethics	164	61	5826	297
17	Ptolemy	131	25	6123	185
18	Galileo	280	67	6308	441
19	energy	293	123	6749	510
20	species	138	58	7259	289
21	atoms	466	82	7548	637
22	Einstein	298	96	8185	499
23	idealism	201	80	8684	370
24	accuracy	64	30	9054	127
25	methodology	2	5	9181	15
26	supernatural	31	12	9196	59
27	algebra	166	47	9255	261
28	geometry	263	27	9516	327
29	calculus	190	60	9843	317
30	topology	73	34	10160	213
31	computer science	171	38	10373	189
32	history	220	65	10562	360
33	human	690	89	10922	886
34	culture	371	118	11808	643
35	law	433	72	12451	599
36	politics	223	42	13050	321
37	art	252	59	13371	390

We draw a graph contain for each search operation number of links, number of paragraphs, and number of generated RDF triples - in figure 10 – which show that each operation save a lot of time and effort for the user, without the system this operation may take hours and with our system it takes only minutes.

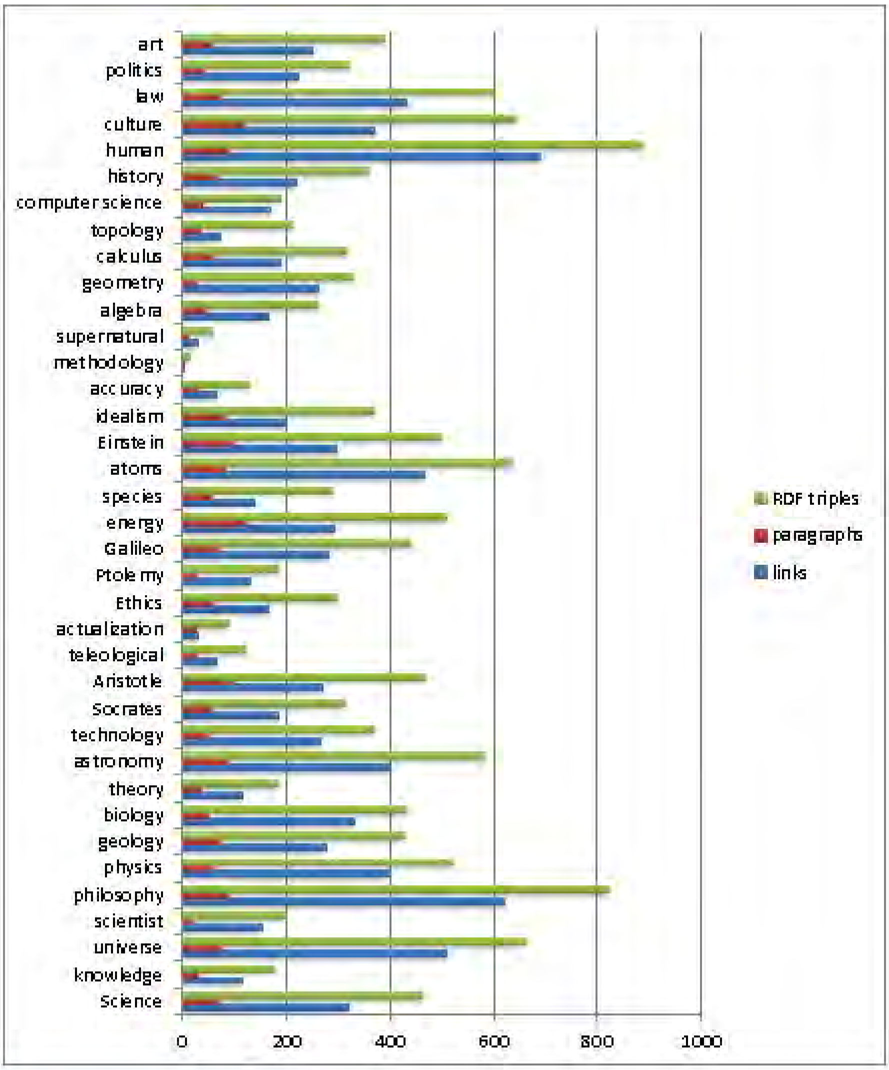

Figure 10: search keyword and generated data

Figure 10: search keyword and generated data

The next advantage of the system that it learns from the search history of users, we show in the next graph – figure 11- that with each new search operation that the number of triples in the RDF data base increases with a high rate, which create a huge data base while the users find what they need. The system RDF data base can be used if the internet connection not found.

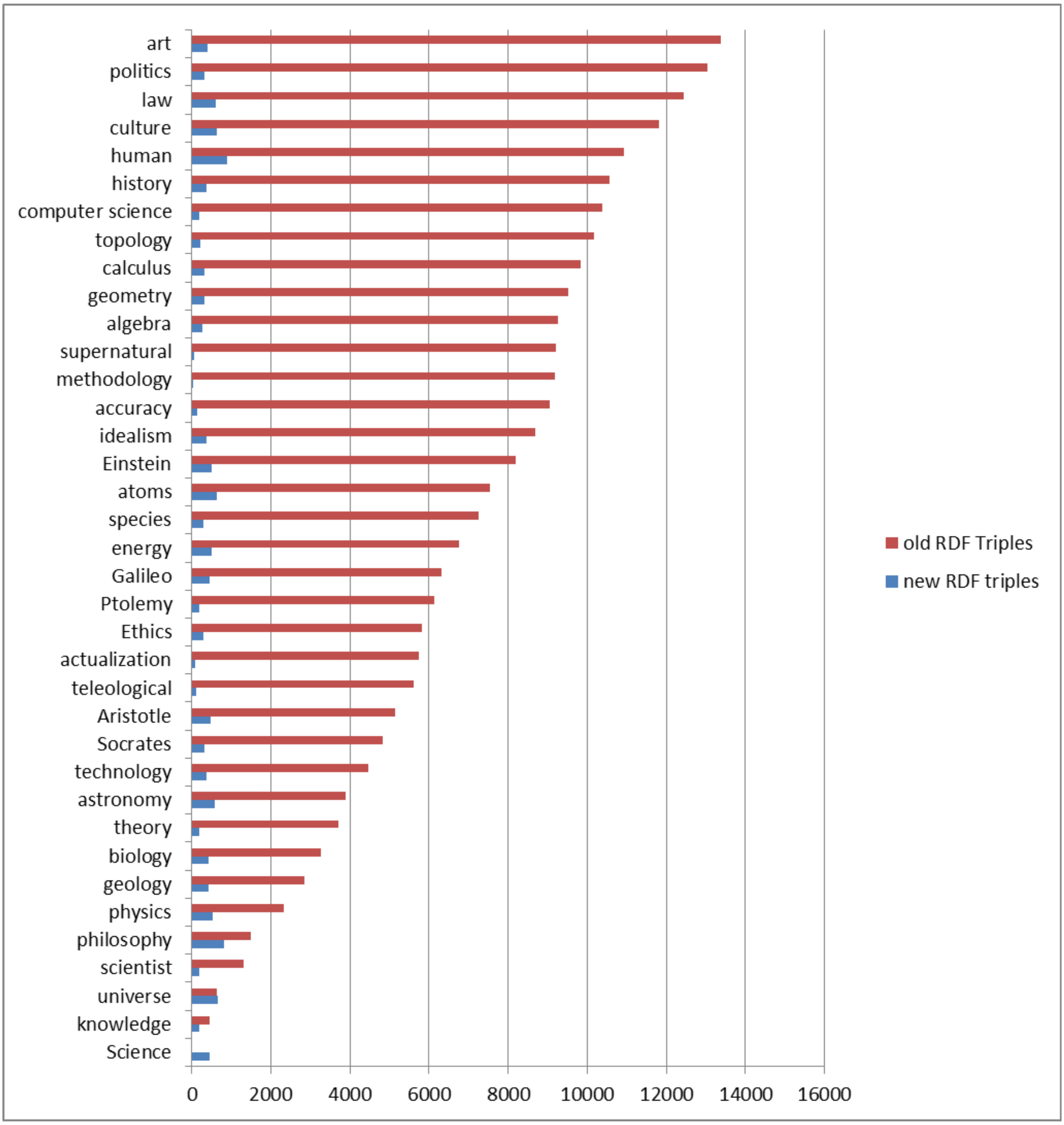

Figure 11: search operations and the increase in the RDF data base

VIII. Discussion & Conclusion

This paper presents a system that can help everyone needs help to find specific information without time or effort consuming, We believe that centralized education, in other words personalization depending students' learning style, is the key on educational field. In situations where distance learning or e-learning is applied, personalization appears as a problem since it's not a desirable way to teach and also not a desirable way of learning in a restricted way of education in any topic. Also a dynamic and semantic program has to be setup for minimizing the gap between traditional learning and e-learning. [11]

The system shows that in very short time the user can gain a great amount of knowledge, with data visualization tools he/she can decide whether this is the data he/she needs or not.

The system also shows that it can built an enormous data base of RDF triples from zero using only users search result, that can be used later without any internet connection.

Considering the fact that learners might have different learning styles and capabilities from the others, displaying and forcing students or learners to get the useful information out of the same content, will not be a good solution for all in e-learning. At this point, there comes the idea of building a semantic structure [12]. The semantic structure of the system provide speed access to the required information, and separate the web pages into smaller contents -Resources- that can be used with other resources or in a different learning topics.

In addition, the system provide an up-to-date information without any data redundancy in the RDF data base, which makes the best use of the web pages retrieved from the search result.

IX. References

[1] MAHENDRA THAKUR, G. S. P. Performance Based Novel Techniques for Semantic Web. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, p. 1694-0814, January 2012.

[2] GRADINAROVA, B. . Z. O. . G. M. An e-Learning Application Based on the, v. 210, 2006.

[3] BERNERS-LEE, T. What the Semantic Web can represent, September 1998. Disponivel em: <http://www.w3.org/DesignIssues/RDFnot.html>. Acesso em: 25 jun. 2013.

[4] DRUCKER, P. Need to Know: Integrating e-Learning with High Velocity Value Chains, 2000.

[5] BRAY, T. What is RDF? O'Reilly xml.com, 2001. Disponivel em: <http://www.xml.com/pub/a/2001/01/24/rdf.html>. Acesso em: 25 jun. 2013.

[6] BRAY, T. RDF and Metadata. O'Reilly xml.com, 1998. Disponivel em: <http://www.xml.com/pub/a/98/06/rdf.html>. Acesso em: 25 jun. 2013.

[7] B.HEMANTH KUMAR, M. S. P. B. An Implementation of Semantic Web System for Information retrieval using J2EE Technologies, india, v. 3, n. 6, 2011. ISSN 0975-3397.

[8] GRAPH#. graph sharp. CodePlexProject Hosting for Open Source Software, 2010. Disponivel em: <http://graphsharp.codeplex.com/>. Acesso em: 25 jun. 2013.

[9] ROB VESSE, R. M. Z. K. A. G. M. T. P. dot net RDF. dot net RDF, 6 May 2013. Disponivel em: <http://www.dotnetrdf.org/>. Acesso em: 25 jun. 2013.

[10] ERIC PRUD'HOMMEAUX, A. S. H.-P. L. B. w3c. SPARQL Query Language for RDF, 26 March 2013. Disponivel em: <http://www.w3.org/TR/rdf-sparql-query/>. Acesso em: 25 jun. 2013.

[11] OGUZ MUSTAPASA, D. K. . A. K. . A. Y. . U. Implementation of Semantic Web Mining on E-Learning, Istanbul-Turkey, 2010.

[12] HARPER, K. C. . C. K. . &. Y. D. C. Distance learning, virtual classrooms, and teaching pedagogy in the Internet Environment, 2004.