

# Ontology construction and query indexing technique to improve query indexing time, precision and number of good pages visited

K.Srihari<sup>1</sup> Dr. V.P.Arunachalam<sup>2</sup> Dr. S Karthik<sup>3</sup>

<sup>1</sup> K.Srihari Lecturer, Department of Computer Science Engineering, Sri Ramakrishna Institute of Technology ,Coimbatore,Tamilnadu,India

<sup>2</sup> Dr. V.P.Arunachalam Supervisor & Principal, SNS College of Technology, Sathy Main Road, Coimbatore,Tamilnadu,India.

<sup>3</sup> Dr. S. Karthik Joint Supervisor, Dean, Department of Computer Science Engineering SNS College of Technology, Sathy Main Road, Coimbatore,Tamilnadu,India.

## Abstract

For The existing search engines have many issues. In this research we have considered the issue of query processing in hakia search engines. Current search is done on ranking based query processing means displaying results based on hit rates . We have decided to propose a solution based on semantic search. The query processing is done and implemented using semantic search Technique.

### 1Problem Identification

. In hakia such engine query indexing is done. Here too much query is taken under consideration and those queries are processed without any standard methodology. Here no proper indexing of query is done [4]. Some propose a standard query indexing Technique in

issues in the existing system. We propose an optimal solution in order to minimize those issues.

order to implement the query traffic control and semantic query retrieval . The current search (semantic) is done and implemented to solve the queries in private data [4]. We propose a solution for query indexing in private data. The existing problems are precision, time, delay factors and number of good pages visited, increase in query retrieval time are some of the

### 2 PROPOSED SOLUTION

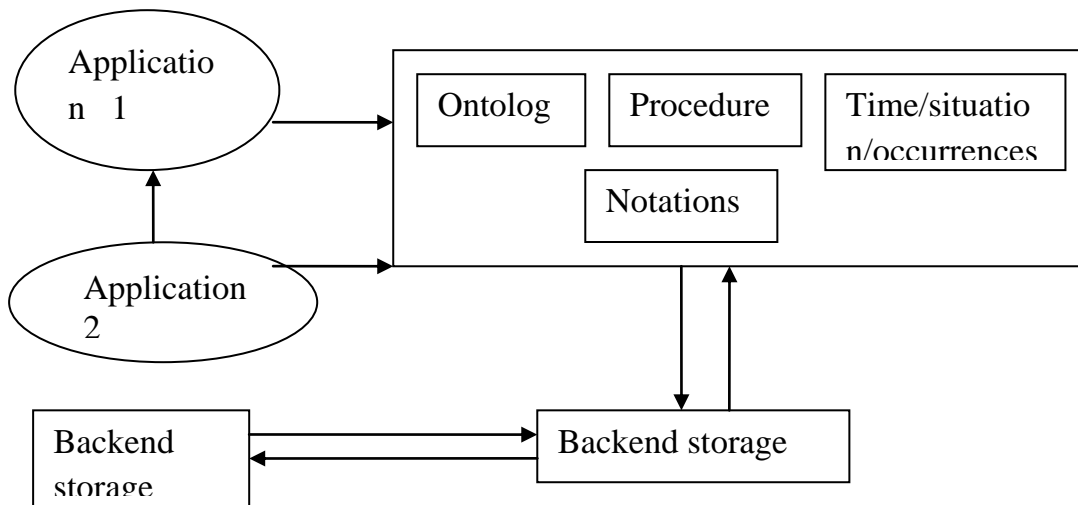
The proposed solution involves 2 main phases.

Ontology construction  
 Query indexing Technique.

order to improve the query  
 indexing, precision and reduced  
 query retrieval time.

Figure 5.1 illustrates the  
 ontology construction model in

**Figure 1 Ontology construction**



The general ontology is  
 constructed based on 4 levels (or)  
 components.

Real time applications.

Objects present in ontology

Backend-storage

Backend –storage API.

**Ontology server block diagrams:**

The application generally deals  
 with the application where ontology has to  
 be implemented. Time / Occurrences /  
 Situation explain the control of the  
 occurrences of words to be searched. The  
 notations describe the parameters (or) the

representation of the word in difference  
 word’s but of same meaning. Procedures  
 deal with the search mechanism. Backend  
 storage deals with the data storage. And  
 API deals with the software compatibility.  
 The existing ontology’s are based on the  
 certain model where relationship is made

with each other. Here we propose an architecture based on decision tree.

And API deals with the software compatibility. The existing ontology's are based on the certain model where relationship is made with each other. Here we propose an architecture based on decision tree. We propose a ontology construction technique using the decision tree based concept.

### 2.1 Algorithm Used

The QDEX method has two embedded operations namely, indexing and extraction method. The indexing is done using the query index method and then the database and the data related to that link is saved as a small part.

#### Query Indexing Technique Algorithm

$Wt(x, y) = Wt1$

If (percentage =80) or

If (percentage=not exact) then

$Wt(x, y) = Wt2$

If (percentage >100) or

If (percentage = greater than exact) then

$Wt(x, y) = Wt3$  end if end for end for

Graph K=Get match (G)

Query match (Set 1, Set 2)

Graph  $G = \text{null graph}$   
( $V1+V2, \Sigma$ )

$Ve1 \leftarrow \text{Set 1}$

$Ve2 \leftarrow \text{Set 2}$

( $Wt1, Wt2, Wt3$ )  $\leftarrow$  Weights  $Ve1, 2$   
For every concept X in  $Ve1$  and  $Ve2$  do  
percentage = Percentage of match (x, y) if  
percentage =0 or fail then add edge (f, g)  
to Graph G.

If (percentage =100) or

If (percentage =cent) then

If (K=nu11) then

Zero match and returns negative result.

End if,

Let (x, y) denote high weight edge in G

### 3 EXPERIMENTAL RESULTS

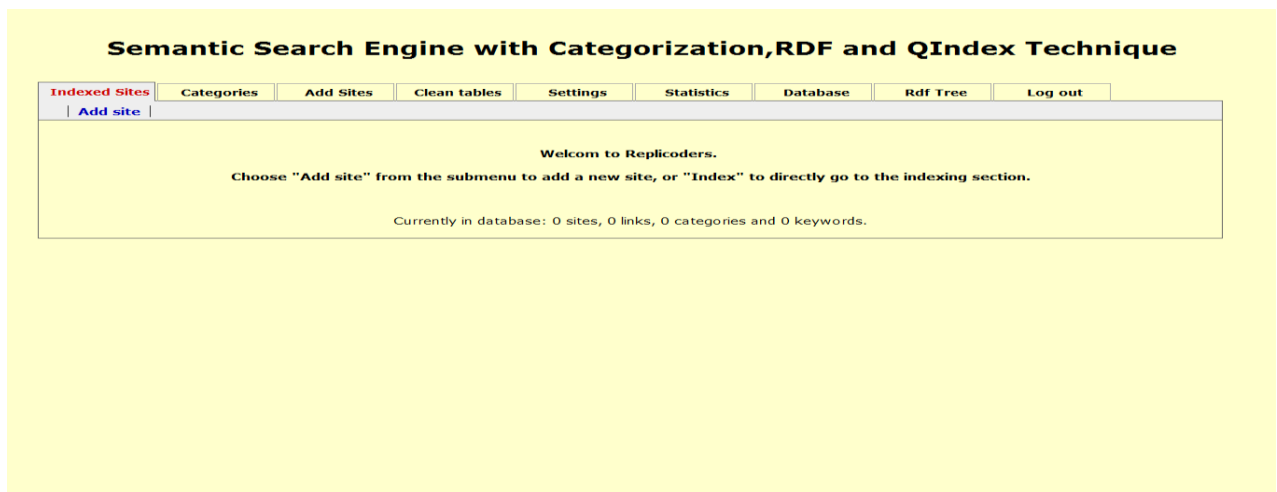
#### ADMIN PANEL:

##### Admin Panel Overview

- The admin panel consists of the 8 fields such as indexed sites, categories; add sites, clean tables, settings, statistics, database and RDF tree.

- The panel will highlight the first option by default.
- The text present in red color is the category and the text present in blue color is the sub category option. Figure 5.2 clearly provides the detail about the admin panel.

**Figure 2 Admin Panel**



#### Adding a Site

To add a site in the search index, the title, URL and the description has to be

mentioned. Figure 5.3 provides details about the addition of sites

Figure 3 Add Site

**Semantic Search Engine with Categorization,RDF and QIndex Technique**

<b>Indexed Sites</b>	Categories	Add Sites	Clean tables	Settings	Statistics	Database	Rdf Tree	Log out
----------------------	------------	-----------	--------------	----------	------------	----------	----------	---------

Add a site

URL:

Title:

Short description:

Category:

click to parse rdf

Currently in database: 0 sites, 0 links, 0 categories and 0 keywords.

### After Adding Sites Option:

After adding a site, the site can be edited, deleted or indexed. The site can be used to browse for pages or statistics can

be viewed. Figure 5.4 shows about the addition of the web site to the main page.

Figure 4 Result of Add Site

**Semantic Search Engine with Categorization,RDF and QIndex Technique**

<b>Indexed Sites</b>	Categories	Add Sites	Clean tables	Settings	Statistics	Database	Rdf Tree	Log out
----------------------	------------	-----------	--------------	----------	------------	----------	----------	---------

URL:	<a href="http://www.replicoders.com/">http://www.replicoders.com/</a>	<a href="#">Edit</a>
Title:	<b>Replicoders-Official Site</b>	<a href="#">Index</a>
Description:	An leading software company	<a href="#">Browse pages</a>
Last indexed:	Not indexed	<a href="#">Delete</a>
		<a href="#">Stats</a>

Currently in database: 1 sites, 0 links, 0 categories and 0 keywords.

### Index Option

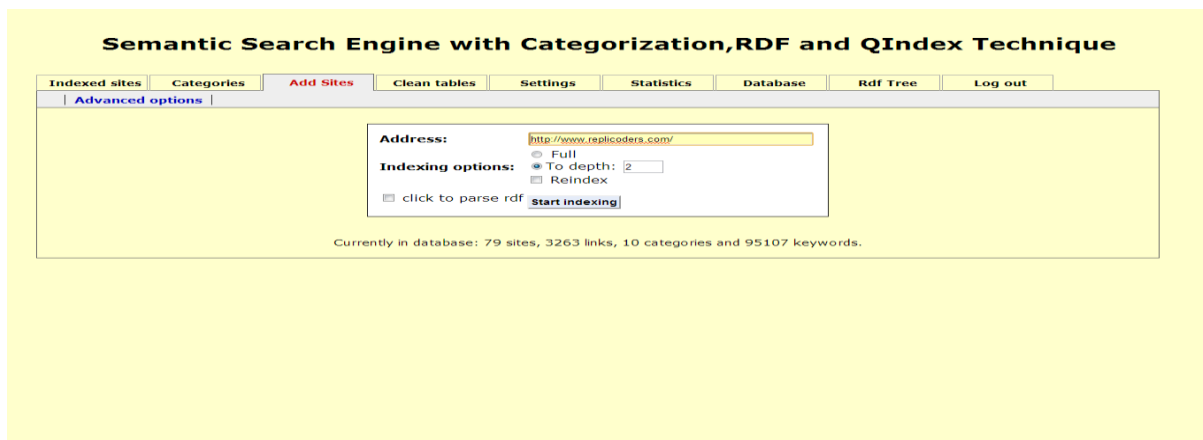
Here the indexing options such as full, depth can be mentioned, the options to parse RDF is also mentioned. Some

advanced options is also present, it has 2 options such as, the url to be included and

the URL that should not be included.

Figure 5.5 shows the indexed sites

**Figure 5 Indexing Sites**



After indexing the sites, the output will be like this.

If any sites related to it, is already present in the database. It shows the message for the site as already in database.

If any sites already indexed come again for indexing, then the page is considered as a duplicate one. Figure 5.6 clearly illustrates the indexed sites.

**Figure 6 Results of Indexed Sites**



After clicking back to admin button it refers and goes to the first page.

Figure 7 explains about the after indexing of sites.

**Figure 7 After indexing the site**

Semantic Search Engine with Categorization,RDF and QIndex Technique								
Indexed sites	Categories	Add Sites	Clean tables	Settings	Statistics	Database	Rdf Tree	Log out
Add site   Reindex all								
Site name	Site url	Last indexed						
social networking site	<a href="http://www.facebook.com/">http://www.facebook.com/</a>	Not indexed	Options					
article	<a href="http://www.articlestoreprint.net/triplify/post/204948">http://www.articlestoreprint.net/triplify/post/204948</a>	2012-06-21	Options					
articles	<a href="http://www.articlesbase.com/education-articles/">http://www.articlesbase.com/education-articles/</a>	2012-06-21	Options					
benefits of education	<a href="http://gibill.va.gov/">http://gibill.va.gov/</a>	2012-06-21	Options					
blog about railway tickets	<a href="http://www.blogdefined.com/3115/how-to-book-railways-tickets-online/">http://www.blogdefined.com/3115/how-to-book-railways-tickets-online/</a>	2012-06-21	Options					
book rail ticket	<a href="http://www.bookrailticket.com/tatkal-train-ticket-booking-tips-and-tricks">http://www.bookrailticket.com/tatkal-train-ticket-booking-tips-and-tricks</a>	2012-06-21	Options					
bus tickets	<a href="http://www.ticketvala.com/">http://www.ticketvala.com/</a>	2012-06-21	Options					
data mining	<a href="http://dataminingwarehousing.blogspot.com/">http://dataminingwarehousing.blogspot.com/</a>	2012-06-21	Options					
data mining	<a href="http://www.ehow.com/how_2106922_data-mine.html">http://www.ehow.com/how_2106922_data-mine.html</a>	2012-06-21	Options					
data mining techniques	<a href="http://www.statsoft.com/textbook/data-mining-techniques/">http://www.statsoft.com/textbook/data-mining-techniques/</a>	2012-06-21	Options					
document delivery	<a href="http://www.healthnet.com/">http://www.healthnet.com/</a>	2012-06-21	Options					
education	<a href="http://www.education.org/">http://www.education.org/</a>	2012-06-21	Options					
education center	<a href="http://www.theeducationcenter.com/">http://www.theeducationcenter.com/</a>	2012-06-21	Options					
education dictionary	<a href="http://www.thefreedictionary.com/education">http://www.thefreedictionary.com/education</a>	2012-06-21	Options					
education news	<a href="http://www.nytimes.com/pages/education/index.html">http://www.nytimes.com/pages/education/index.html</a>	2012-06-21	Options					
Flight ticket	<a href="http://www.delta.com/planning_reservations/plan_flight/book_flights/index.jsp">http://www.delta.com/planning_reservations/plan_flight/book_flights/index.jsp</a>	2012-06-21	Options					
florida education	<a href="http://www.fldoe.org/">http://www.fldoe.org/</a>	2012-06-21	Options					
food habits influencing health	<a href="http://eating.health.com/">http://eating.health.com/</a>	2012-06-21	Options					
Hakia	<a href="http://hakia.com/">http://hakia.com/</a>	2012-06-21	Options					
Hakia information	<a href="http://company.hakia.com/whatis.html">http://company.hakia.com/whatis.html</a>	2012-06-21	Options					
health fitness	<a href="http://healthfitnesssites.com/">http://healthfitnesssites.com/</a>	2012-06-21	Options					
Health information	<a href="http://health.gov/">http://health.gov/</a>	2012-06-21	Options					

#### 4 Graph-Comparison Results

The table 1,2,.3 and the figure 8,9 and 10 clearly justifies that the proposed semantic search engine decreases the query indexing time when it is compared to lexxe,hakia,swoogle.. The table 4,5, 6, 7 and the figures 11, 12, 13 produces optimized retrieval time. The table 8, .9, 10, 11 and the figure 14,.15, 16 clearly shows that precision is optimized when compared to existing search engines like lexxe,swoogle,hakia.

**Query Indexing (number of queries vs total query indexing time)**

**Table 1 Query Indexing -Semantic Query Indexing time)  
Search Engine vs Lexxe (Decrease in**

category	Semantic search engine		Lexxe	
	Total number of queries	Total query indexing time(seconds)	Total number of queries	query indexing time (seconds)
Replicoders	2	0.00124	2	0.150
Common	31	0.01354	20	0.345
Health	105	0.81202	52	0.405
Travel	163	0.79851	36	0.234
Education	578	1.10954	74	0.313

**Figure 8 Total number of queries vs total query indexing time (Semantic Search Engine vs Lexxe) ( Decrease in Query Indexing time)**

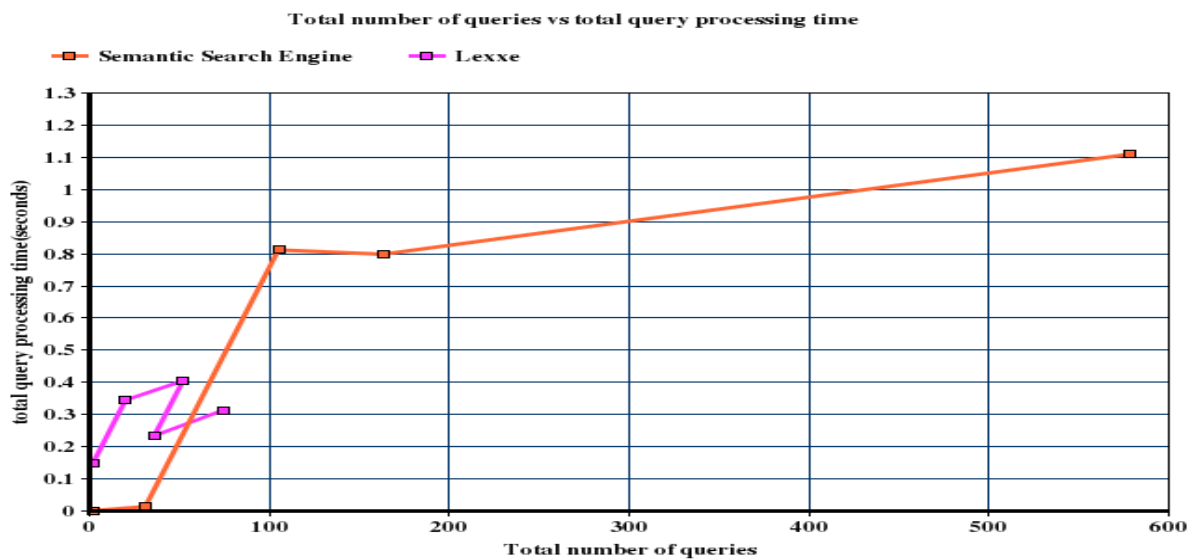


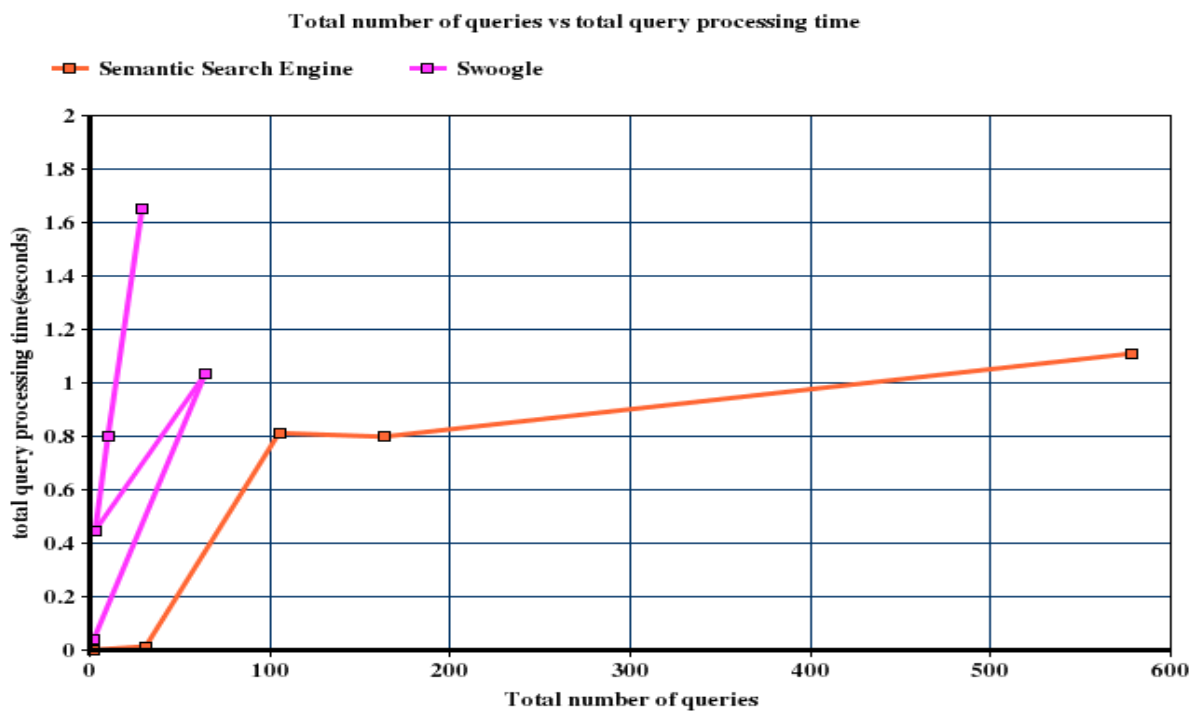


Table 2 Query indexing -Semantic Search Engine vs Swoogle

(Decrease in Query Indexing time)

category	Semantic search engine		Swoogle	
	Total number of queries	query indexing time (seconds)	Total number of queries	query indexing time(seconds)
Replicoders	2	0.00124	2	0.038
Common	31	0.01354	64	1.033
Health	105	0.81202	3	0.448
Travel	163	0.79851	10	0.798
Education	578	1.10954	29	1.652

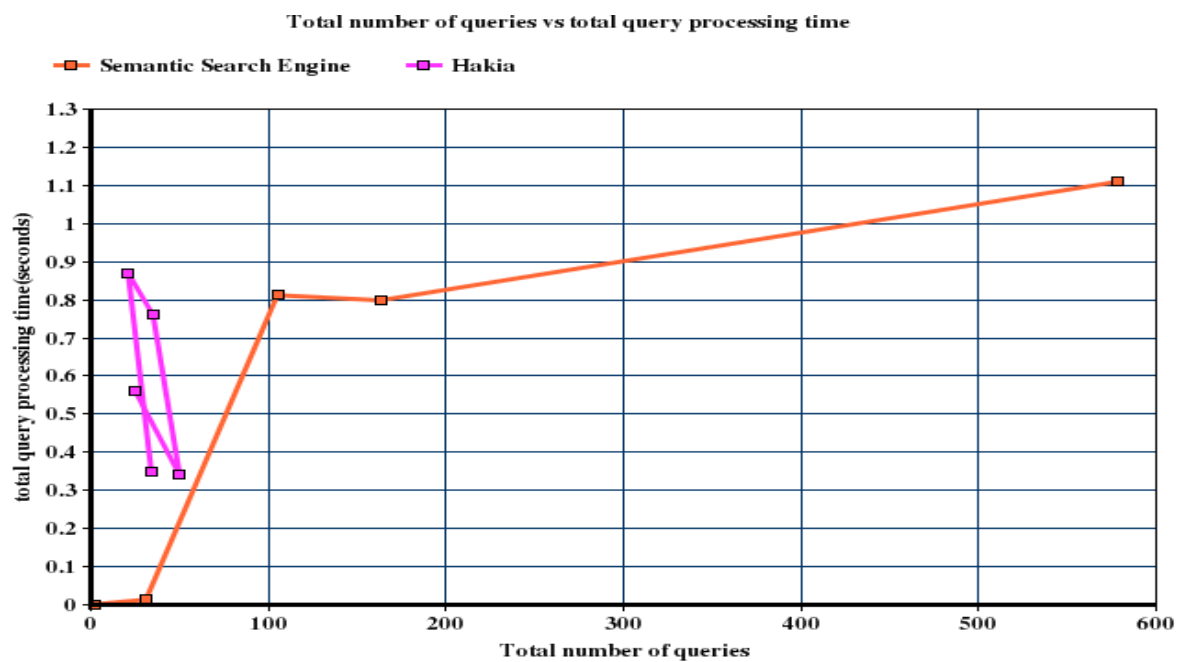
Figure 9 Total number of queries vs total query indexing time (Semantic Search Engine vs Swoogle) ( Decrease in Query Indexing time)



**Table 3 Query Indexing -Semantic Search Engine vs Hakia ( Decrease in Query Indexing time)**

category	Semantic search engine		Hakia	
	Total number of queries	query indexing time (seconds)	Total number of queries	query indexing time (seconds)
Replicoders	2	0.00124	25	0.56
Common	31	0.01354	50	0.34
Health	105	0.81202	35	0.76
Travel	163	0.79851	21	0.87
Education	578	1.10954	34	0.35

**Figure 10 Total number of queries vs total query indexing time (Semantic Search Engine vs Hakia) (Decrease in Query Indexing time)**



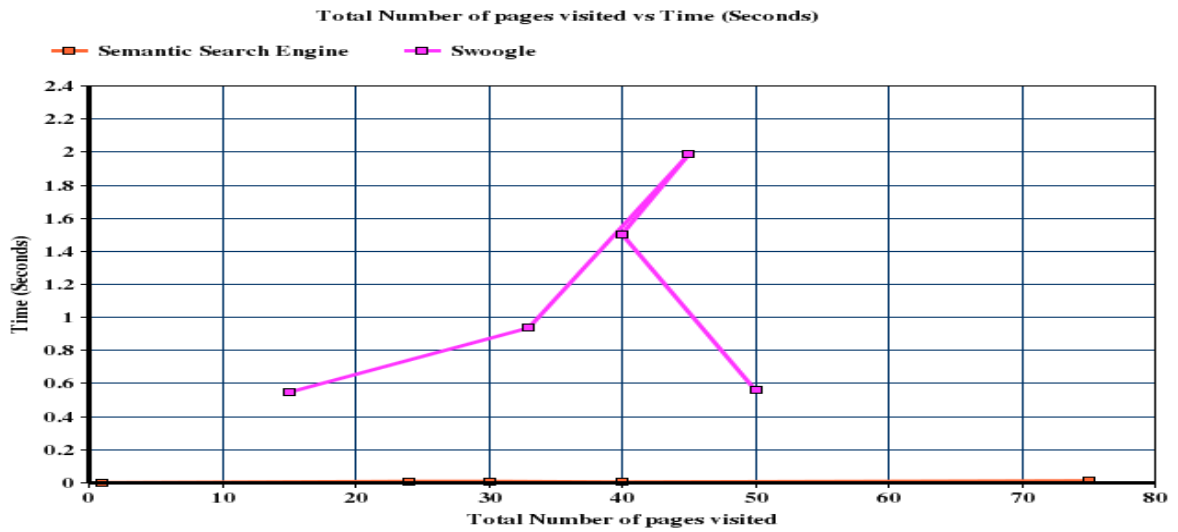
**Table 4 Total Number of pages visited vs Time (Seconds) (Decrease in retrieval time)**

category	Semantic search engine		Swoogle		Hakia		Lexxe	
	no of pages visited	Time (seconds)	no of pages visited	Time (seconds)	no of pages visited	Time (seconds)	no of pages visited	Time (seconds)
Replicoders	1	0.00133	10	0.547	10	0.2	15	0.150
Common	24	0.00887	27	0.937	25	0.03	33	0.176
Health	30	0.00907	21	1.986	20	0.1	45	0.138
Travel	40	0.00518	41	1.503	35	0.22	40	0.245
Education	75	0.01216	35	0.561	30	0.15	50	0.320

**Table 5 Total Number of pages visited vs Time (Seconds) (Semantic search engine vs Swoogle) (Decrease in retrieval time)****Performance of Semantic search engine vs Swoogle**

category	Semantic search engine		Swoogle	
	no of pages visited	Time (seconds)	no of pages visited	Time (seconds)
Replicoders	1	0.00133	10	0.547
Common	24	0.00887	27	0.937
Health	30	0.00907	21	1.986
Travel	40	0.00518	41	1.503
Education	75	0.01216	35	0.561

**Figure 11 Total Number of pages visited vs Time (Seconds) (Semantic search engine vs Swoogle) (Decrease in retrieval time)**

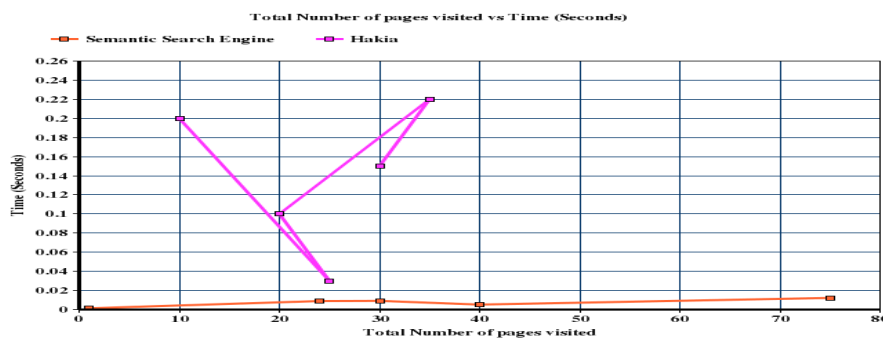


**Table 6 Total Number of pages visited vs Time (Seconds) (Semantic search engine vs Hakia) (Decrease in retrieval time)**

**Performance of Semantic search engine vs Hakia**

category	Semantic search engine		Hakia	
	no of pages visited	Time (seconds)	no of pages visited	Time (seconds)
Replicoders	1	0.00133	10	0.2
Common	24	0.00887	25	0.03
Health	30	0.00907	20	0.1
Travel	40	0.00518	35	0.22
Education	75	0.01216	30	0.15

**Figure 12 Total Number of pages visited vs Time (Seconds) (Semantic search engine vs Hakia) (Decrease in retrieval time)**

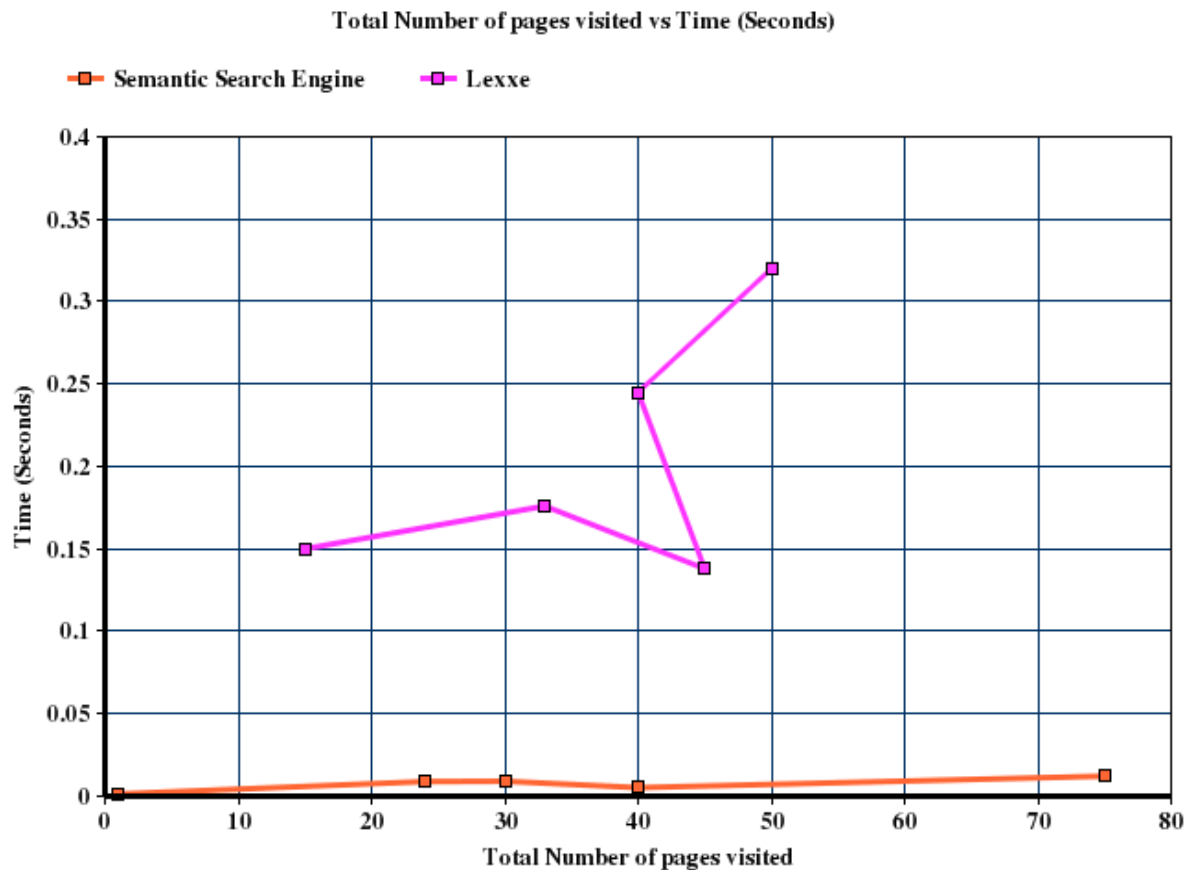


**Table 7 Total Number of pages visited vs Time (Seconds) (Semantic search engine vs Lexxe) (Decrease in retrieval time)**

**Performance of Semantic search engine vs Lexxe**

category	Semantic search engine		Lexxe	
	no of pages visited	Time (seconds)	no of pages visited	Time (seconds)
Replicoders	1	0.00133	15	0.150
Common	24	0.00887	33	0.176
Health	30	0.00907	45	0.138
Travel	40	0.00518	40	0.245
Education	75	0.01216	50	0.320

**Figure 13 Total Number of pages visited vs Time (Seconds) (Semantic search engine vs Lexxe) (Decrease in retrieval time)**

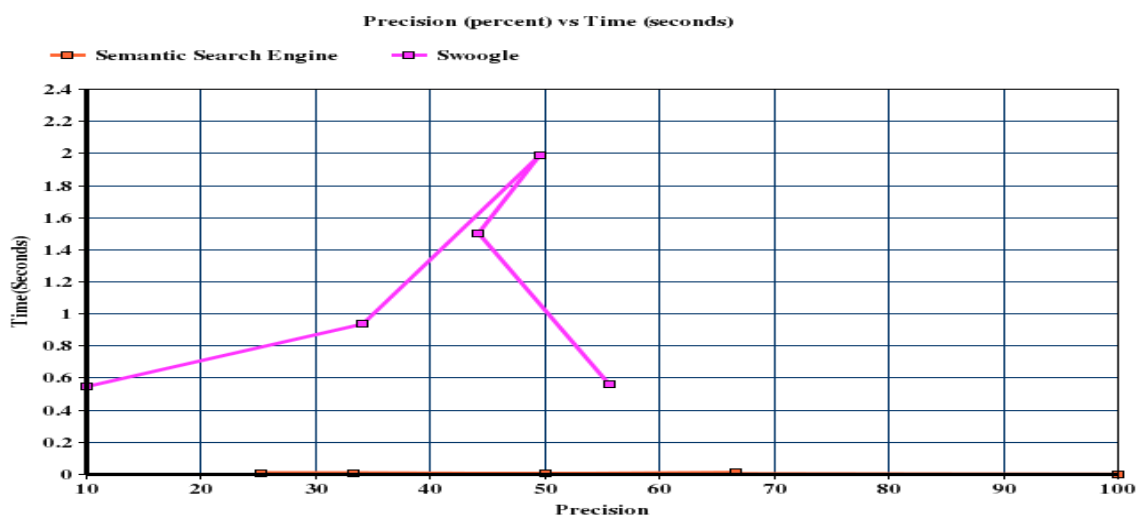


**Table 8 Precision (percent) vs Time (seconds) (Increase in precision)**

category	Semantic search engine		Swoogle		Hakia		Lexxe	
	Precision	Time (seconds)	Precision	Time (seconds)	Precision	Time (seconds)	Precision	Time (seconds)
Replicoders	100	0.00133	10	0.547	33.33	0.2	6.66	0.150
Common	25.2	0.00887	34.1	0.937	25.0	0.03	30.33	0.176
Health	33.3	0.00907	49.5	1.986	14.2	0.1	55.55	0.138
Travel	50.0	0.00518	44.2	1.503	24.1	0.22	37.5	0.245
Education	66.6	0.01216	55.6	0.561	15.7	0.15	36	0.320

**Table 9 Precision (percent) vs Time (seconds)(Semantic search engine vs Swoogle) (Increase in precision)****Performance of Semantic search engine vs Swoogle**

category	Semantic search engine		Swoogle	
	Precision	Time (seconds)	Precision	Time (seconds)
Replicoders	100	0.00133	10	0.547
Common	25.2	0.00887	34.1	0.937
Health	33.3	0.00907	49.5	1.986
Travel	50.0	0.00518	44.2	1.503
Education	66.6	0.01216	55.6	0.561

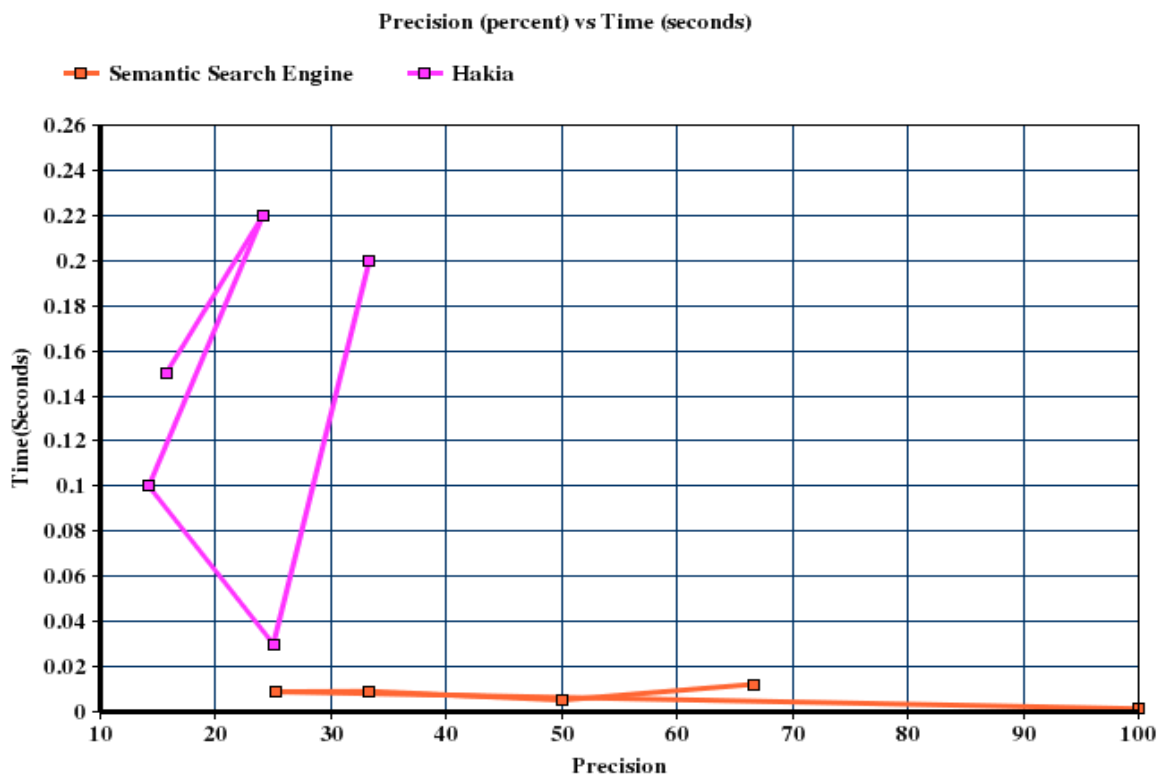
**Figure 14 Precision (percent) vs Time (seconds)(Semantic search engine vs Swoogle) (Increase in precision)**

**Table 10 Precision (percent) vs Time (seconds)(Semantic search engine vs Hakia)  
(Increase in precision)**

**Performance of Semantic search engine vs Hakia**

category	Semantic search engine		Hakia	
	Precision	Time (seconds)	Precision	Time (seconds)
Replicoders	100	0.00133	33.33	0.2
Common	25.2	0.00887	25.0	0.03
Health	33.3	0.00907	14.2	0.1
Travel	50.0	0.00518	24.1	0.22
Education	66.6	0.01216	15.7	0.15

**Figure 15 Precision (percent) vs Time (seconds)(Semantic search engine vs Hakia)  
(Increase in precision)**

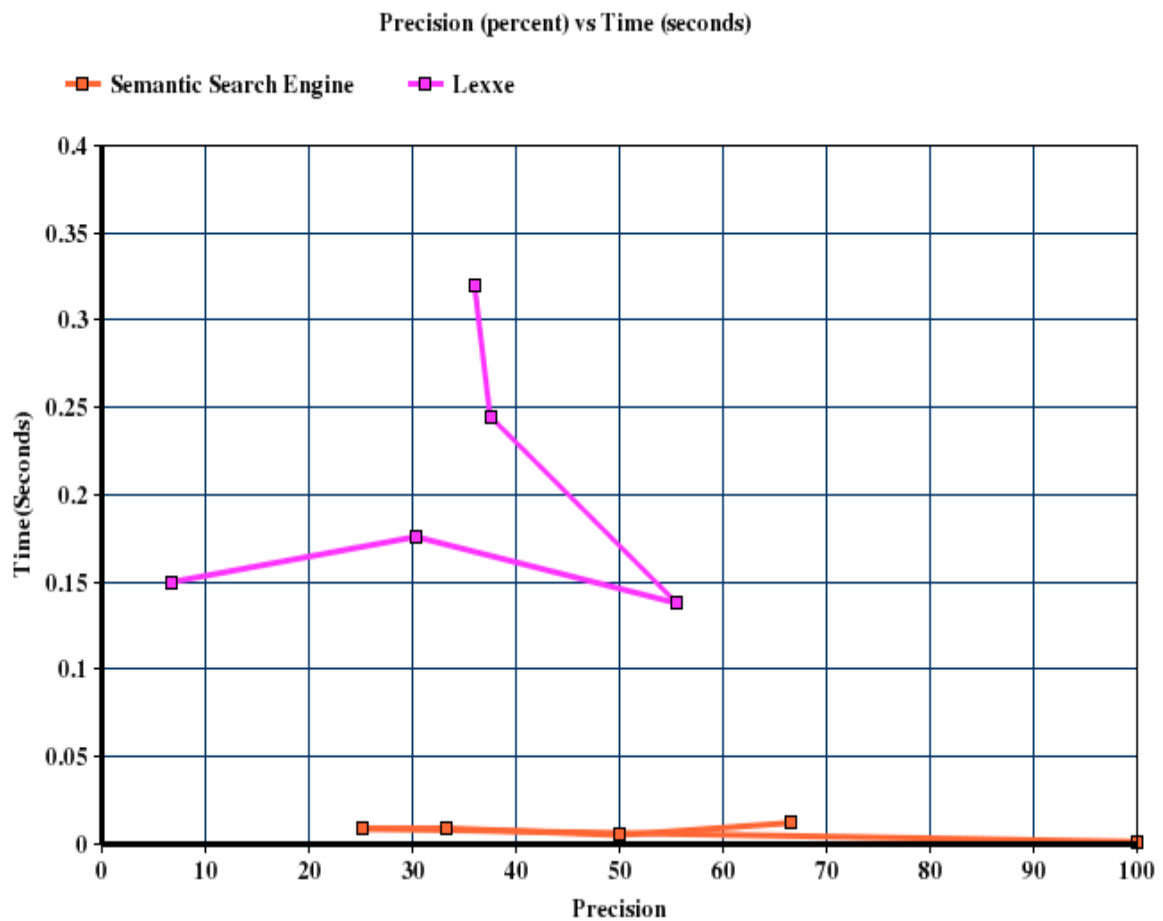


**Table 11 Precision (percent) vs Time (seconds)(Semantic search engine vs Lexxe)  
(Increase in precision)**

**Performance of Semantic search engine vs Lexxe**

category	Semantic search engine		Lexxe	
	Precision	Time (seconds)	Precision	Time (seconds)
Replicoders	100	0.00133	6.66	0.150
Common	25.2	0.00887	30.33	0.176
Health	33.3	0.00907	55.55	0.138
Travel	50.0	0.00518	37.5	0.245
Education	66.6	0.01216	36	0.320

**Figure 16 Precision (percent) vs Time (seconds)(Semantic search engine vs Lexxe)  
(Increase in precision)**





The above tables and graphs clearly shows that proposed search engines produces improved values in query indexing time, total no of pages visited, precision, time, and delay factors and no of good pages visited

## 5 Conclusion

This chapter gives a brief description of queries and the files related to it. Here indexing of a query, ontology, precision and number of good pages visited are optimized and compared with various search engines.

## 6 References

- [1] Bollegala, D., Matsuo, Y., Ishizuka, M., "A Web Search Engine-Based Approach to Measure Semantic Similarity between Words", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 23, Pp. 977 – 990, 2011.
- [2] Yildirim, Y., Yazici, A., Yilmaz, T., "Automatic Semantic Content Extraction in Videos using a Fuzzy Ontology and Rule-Based Model", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, Pp. 47 – 61, 2011.
- [3] Ibanez, M.J., Fabra, J., Alvarez, P., Ezpeleta, J., "Model Checking Analysis of

Semantically Annotated Business Processes", *IEEE Transactions on Man and Cybernetics, Part A: Systems and Humans*, Vol. 42, Pp. 854 – 867, 2012.

[4] Rebollo-Monedero, D., Domingo-Ferrer, J., Forne, J., "Query Profile Obfuscation by Means of Optimal Query Exchange between Users", *IEEE Transactions on Dependable and Secure Computing*, Vol. 9, Pp. 641 - 654, 2012.

[5] Papagelis, A., Zaroliagis, C., "A Collaborative Decentralized Approach to Web Search", *IEEE Transactions on and Cybernetics, Part A: Systems and Humans*, Vol. 42, Pp. 1271-1290, 2012.

[6] Bozzon, A., Brambilla, M., Ceri, S., Quarteroni, S., "A Framework for Integrating, Exploring, and Searching Location-Based Web Data", *IEEE Internet Computing*, Vol. 15, Pp. 24 - 31, 2011.