ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, Issue 3, March 2013

Improving ontology and precision recall using ontology model, genetic, greedy algorithm semantic similarity calculation and ontology graph

K.Srihari [1] Dr. V.P.Arunachalam [2] Dr. S Karthik [3]

[1] K.Srihari Lecturer, Department of Computer Science Engineering, Sri Ramakrishna Institute of Technology ,Coimbatore,Tamilnadu,India

[2] Dr. V.P.Arunachalam Supervisor & Principal, SNS College of Technology, Sathy Main Road, Coimbatore,Tamilnadu,India.

[3] Dr. S. Karthik Joint Supervisor, Dean, Department of Computer Science Engineering SNS College of Technology, Sathy Main Road, Coimbatore,Tamilnadu,India.

ABSTRACT

The content is extracted by means of semantic relevancy. The semantic relevancies relate the content of videos based on a certain parameter. The parameter varies between system to system (implementation). The parameter will improve the performance of semantic relevancy and accuracy. This accuracy is obtained after various random experiments. Here a method called concept, sub concept graph method is used to implement the semantic relevancies. A graph algorithm is constructed to improve the relevancies between concepts. The ontology model is created based on the relationship between the vertices. At first relationship between the parent and child are calculated. Then based on all the relationships the diagrammatic representations are done. Based on hit rates the priority of web pages are done and based on the number of relationships the value for the vertices is noted.

1 PROBLEM IDENTIFICATION

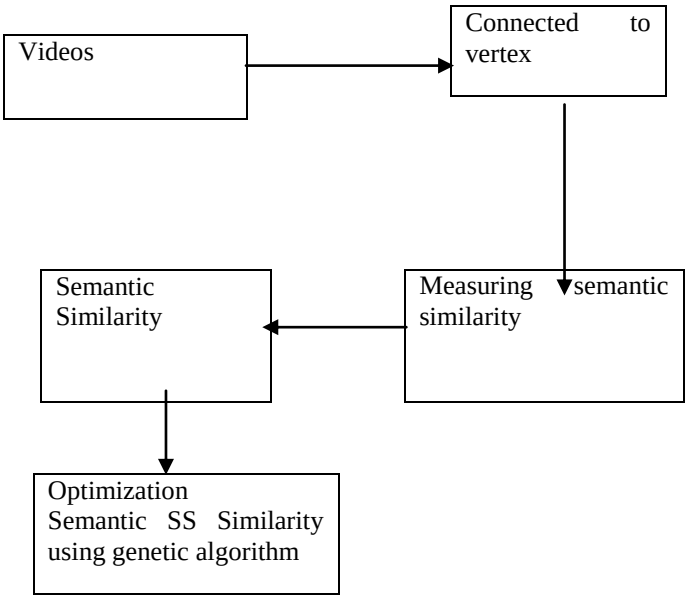
The existing system deals with the fuzzy logic based system where automatic genetic based objects are constructed. The main drawback of this system is the accuracy of the relevant measurement is not measured. The amount of relevancy and amount of objects created using genetic algorithms is main problem in current system. Improving precision and recall are the two main constraints in current web domain.

Here is deal with 2 approaches concept vertex graph (CVG) approach and object genetic measurement (ogm) approach. Figure 1 clearly illustrates the working of proposed semantic search engines.

Working Block Dig:

1. Videos are converted into vertex.
2. Properties of vertex are based on hits.

Figure.1 Final results with high semantic similarity



The proposed solution generally deals with the (i) collecting all related videos in form of

2 ALGORITHM USED

Directed graph i.e. means all videos are represented by a vertex. V1to Vn such that the value of accuracy is measured by the minimum distance between the vertex.

Let us consider:

V1>V2 be 2 vertex such that in a way that the P (V1), P (V2)…….P (Vn)> Propagation of occurrences of vertex depends on user view (rate).

P (V1)>Start vertex, next vertex depends the semantic relevancy is obtained based on the maximum value of the hit rates after compulsion.

Algorithm for ontology constant and semantic similarity

let //V1,V2,V3 be the vertex

//O be the ontology model

//Sm be semantic similarity between videos.

//h(Vn).hit rate of vertex(Video) from Vn-1 videos

//O(S).Optimum similarity of the video.

O created by means of collecting all vertex(video) and based on hit rate.

O=V1->Vn and o € V1 -> Vn

Sm for V1>V2=max hit rate of V2 from V1

Sm for V1>V3=Second max hit rate from V1

Sm for V1.Vn= nth max hit rate from V1.

On the user new.

End The ontology model is created by means of considering videos and their hit rates.

Let us assume that H(V3) = 3 hits

H(V4) from V1= 2 hits

Figure 2 Ontology model

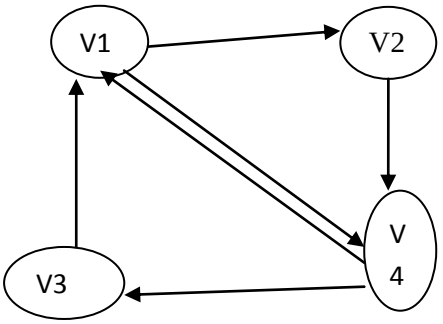


Figure 2 Shows that the vertices with relationship

Similarity between vertexes is calculated by

Sm (V1) =V2, V3, V4 = hits (Max),

Hits (second max), hits (Third max)

Hits (Fourth max) …….. hits (nth max)

Optimum similarity

O(S)= Max{S[V1,[V2,V3]],

S[V2,[V3,V4]],

S[V3,[V4,V1]],

S[V1,[V2,V3,V4]]

O(S)=Max {hits caused by combination}.

Ontology Construction Measurement:

Ontology construction measurement is done by means of hit rates. h1>h2>h3 =V1>V2>V3.The measurements of hit rates generally given by search engines and user views. So based on hit late the ontology graph is done.

Algorithm for ontology graph

// (let olg) Ontology graph model

// V1>Vn Vertex in Ontology graph

// OCM be measuring ontology graph //constructed with V1,V2,…..Vn.

Implies

OCM=Max [h1] - 2nd Max (h2) (v2), n max (hn)(Vn).

end finally ontology is constructed and measured. Genetic algorithm implementation to improve the similarity between moving under objects. The objects presented between the videos are related by mean of genetic algorithm.

6.2.3 Algorithm for genetic optimization

//O1 (V1) is object 1 of video 1 (vertex 1)

// O2 (V2) be object 2 of videos 2 (vertex 2)

// On (Vn) be object n of video (Vn)

// F O1 (V1) Present in On(V2) then

O1 (V1) → parent of On(V2)

O1 (V1) Present in On(Vn) then

O1 (V1) parent of On(Vn)

Else

Check O2 (V2) full On (Vn)

Find parent child relationship between all objects in all video (O1) vertexes.

O1 (V1) present in On(V2) and new O(V2)

On (V2). Then child of O1 (V1) = new

If O1 (V1) present and O new (V2)or O new (V3)

Example for implementing 3 approaches:

1. Ontology Construction

2. Semantic Similarity

3. Genetic algorithm with a restaurant example. Consider hotels present in India (Hotel Taj, Residency, Alloft) Ontology is constructed based on user views.

So V1 is top of vertex since hits are very high.

1. Semantic Similarity is calculated with the formula and optimal solution is obtained.

2. Genetic Algorithm.

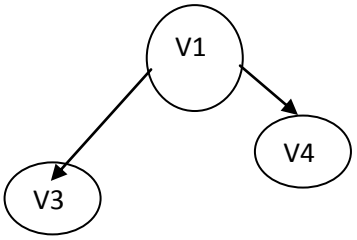
Here V1 is parent because V1 occurs in all searches

O (V2). This child of O1 (V1) =new On(Vn).

End

Figure 3 illustrates genetic optimization proposal technique where the relations are found

Figure 3 Genetic Optimization

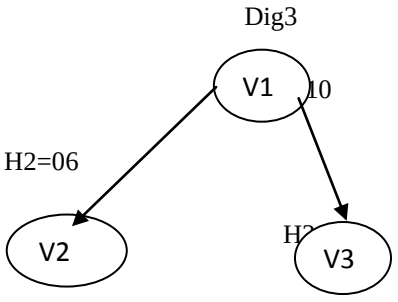


Let V1→ Max hit rate for

Hotel Taj

V2→ 2nd max hit rate for hotel residency ,V3→ 3rd max hit rate for hotel aloft.

Figure 4 Explains Genetic Algorithm for Restaurant



of V2 and V3. So V1 is parent of V2 and V3.

Ontology Model

Videos→ Vertex → Vertex arranged based on hit rates.Videos (1) →Videos (2) →Video (n) Eg:- Priority sports is searched and the priority as follows:

1. Foot pass
2. Cricket
3. Hockey

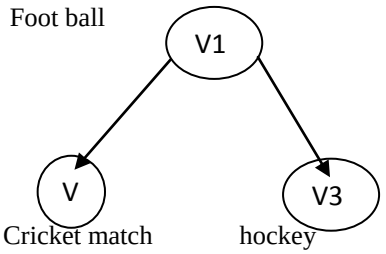
Foot ball no. of hits is 800(app)

Cricket no. of hits is 700(app)

Foot ball and Cricket is 600(app)

Figure 5 Explains Ontology Model for Sports

Dig4



So hence the popular sports are calculated based on the hit rates.

Semantic Similarity between videos.

For the same example the videos are arranged in such a way that the most popular videos are arranged first then the next videos.

V1→V2= Max hits =1500

V2→V3=2nd hits =1400

V3→V1=3rd hits =1300

Based on hits the semantic similarity is calculated.

Genetic Algorithms approach for the objects in the videos

3 COMPARISON OF RESULTS FROM EXISTING AND PROPOSED SOLUTION

The object cricket present in the videos is related with the other videos. Eg. Object football is checked in all the videos in order to generate the genetic object extraction.

Comparison between existing of proposed Experimental approach:

Existing system they have considered precision and recall values. Values for football videos (semantic Similar) existing proposed.

Ontology model comparison:

Existing solution is generally based on rules, here we propose a model based on hit rates. They used beyond connection to create the ontology model.

Here we propose a solution which is retrieved based on hit rates. Rule comparison with ontology.The Comparison of rule schema (existing) with proposed genetic algorithm is shown in table 6.1

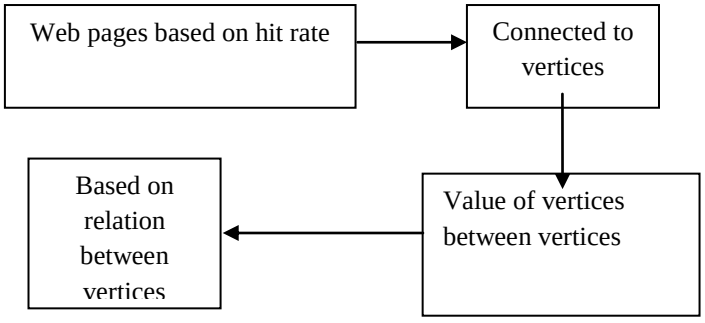
Table1 narrates the difference between rule schema and genetic algorithm

Table 1 Rule Schema Vs Genetic Algorithm

Rule based on logical relations produced by the ontology model designer.	Here hit rate indicates the quality of pages increased and accepted by the user while browsing.
Man made relations.	Quality based relations.
The logical connection is not verified.	Verified by the user comments

Figure 6 clearly shows how relations are made between the vertices.

Figure 6 Functional block diagram



Based on relation between vertices value of vertices based on this rate Functional Block of ontology creation.

Diagrammatic representation of ontology creations.

V1-Based on high number of hits/View by user.

V2-Based on Second high number of hits/view by user

Vn-Based on nth high number of hits/views by user

R1-Maximum number of relations between vertices,1, 4 2

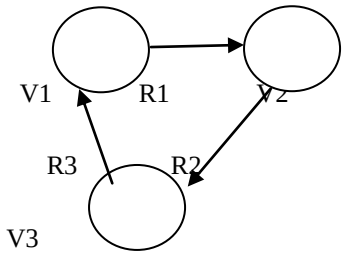
R2-2nd maximum of relation between vertices 2,4,3

R3-3rd maximum number of relations between vertices, 3,4,2.

S.T=Ontology model is given by

Figure 7 clearly shows how the ontology model is created.

Figure 7 a Ontology Creation Model



V1=Maximum hit rate

R1=V1 Relate V2 (Maximum)

R2=V2 Relate V3 (2nd Max)

R3=V3 Relate V1 (3rd Max)

Max1<Max 2<max 3

Relation union (V1, V2)

(Vn Vn) Interaction (V1, V2)

disjoint (V1,V2)

Exnor (V1,V2)

Exnor (V1,V2)

Value.No. of relations

Number of relations increase then value increase else value decrease.

end if.

Algorithm for ontology creation

// V1 be the vertex with maximum number of hits/view by user.

//V2 be the vertex with second maximum number of hits/view by user

//V3 be the vertex with third maximum number of hits/views by user.

//Vn be the nth number of hits/views by user

//R1 be maximum number of relation between V1 & V2.

// R2 be 2nd maximum number of relation between vertices V2 & V3.

R3 be 3rd maximum number of relation between vertices V3 & V1.

Semantic similarity calculations:

The Semantic similarity is calculated by mean of the matrix method of maximum relations. Let us consider for example:(V1,V2,V3)=(14,12,10)

Our primary objective is to increase the number of relations between the vertices. If the relation increases the semantic relevancy gets increased. So our proposed semantic similarity involves the relationship between the vertices has to be increased.

	V1	V2	V3
V1	10	14	10
V2	26	0	12
V3	37	10	0

Figure 7 b

Figure 6.7 b clearly shows how the relations are made between vertices

V1>V3=10 V2.V1=6

V1>V2=14 V2 V2=0

v1 v3=10 v2 v3=12

v3 v1=7, v3 v2=10 v3 v3=0

Procedure:

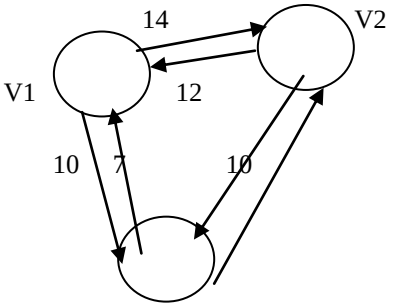
1. First all the vertex and their relations with other vertex is calculated.

2. The relationship between the vertex are calculated.

3. The relationship between the vertex's are calculated

4. If the number of relation is less the proposed alternate path method help us to increase the number of relations between the vertices.V1 V3=10 Alternate path is V1>V2 then V2>V3. figure 8 shows the semantic similarity between the vertices.

Figure 8 Semantic Similarity Relation



Which implies the cost of brand is 14+12=26. The relation are 2 6 11 only. The relation for vertex where the value if 0 is calculated. So maximum numbers of relations are obtained. Relation of V2 V1=22.So probable maximum number of relation are made.

Algorithm for Semantic search similarity calculations

//R be relationship

//V1 Vn be vertex

//R1 Rn be relation for v1 vn

R(v1 Vn) is calculated

if R(Vn1 Vn2)=MAX then stop

else

if R(Vn1 Vn2)=min then

MAX R(Vn1 Vn2)

Relate (Vn1 Vn2)=MAX

then

R(Vn3 Vn4) is selected

Search if

R(Vn3 Vn4)=Min then

Max R(Vn3 Vn4)

Else Stop continue till all vertex visit and maximum relation obtained

End

Semantic similarity optimization using Greedy algorithm Consider the diagram

Figure 9 Greedy Algorithm

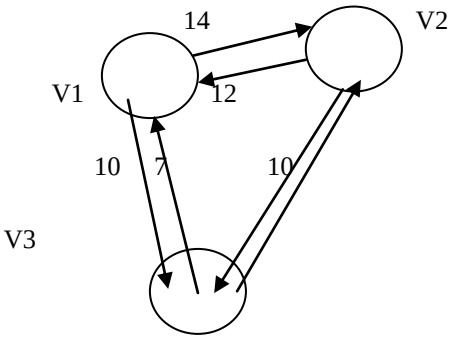


Figure 9 clearly narrates the implementation of greedy algorithm.

Greedy method

Normal path

v1 v2=14

V2 V3=12

V3 V1=7

V3 V2=10

V2 V1=6

V1 V3=10

Special path

To travel from V1 V3 normal path given 10.when we travel vertex V1 to V2 to V3

V1 V2 V3=14+12=26

Special Task 1

To travel from V3 V2 normal path gives 10 but when we travel V3 V1 V2=7+14=21

Special Task 2

To travel from V2 V1 normal path gives 10 but when we travel V2 V3 V1=12+7=19

Special Task 3

To travel from V1 V2normal path gives 14.In special path V1 V3 V2=10+10=20

Special Task 4

To travel from V2 V3normal path gives 12.In special path

V2 V1 V3=6+10=16

Special Task 5

To travel from V3 V1 normal path gives 7.In special path V3 V2 V1=10+6=16

A traveling from normal path the value is high when compared to special path technique. This technique helps us to make more relationship. The optimization of semantic search can also done by greedy algorithm.

Greedy Algorithm

//let V1 Vn be vertex

Task1 Task n be greedy algorithm task

R1 Rn be relation

Traverse from Vn1 Vn2 Then

The value is calculated

If Value <MAX value

Special Task (Greedy) implemented

Sub vertex traversal implemented

Vn1 Vn2=Vn1 Vn3 Vn2

Greedy special task=Max path for vertex

Max.path=Max.Relations R1 Rn

From experiment we found that more relations arew made (ie) semantic retrieval gets increased with more relations.

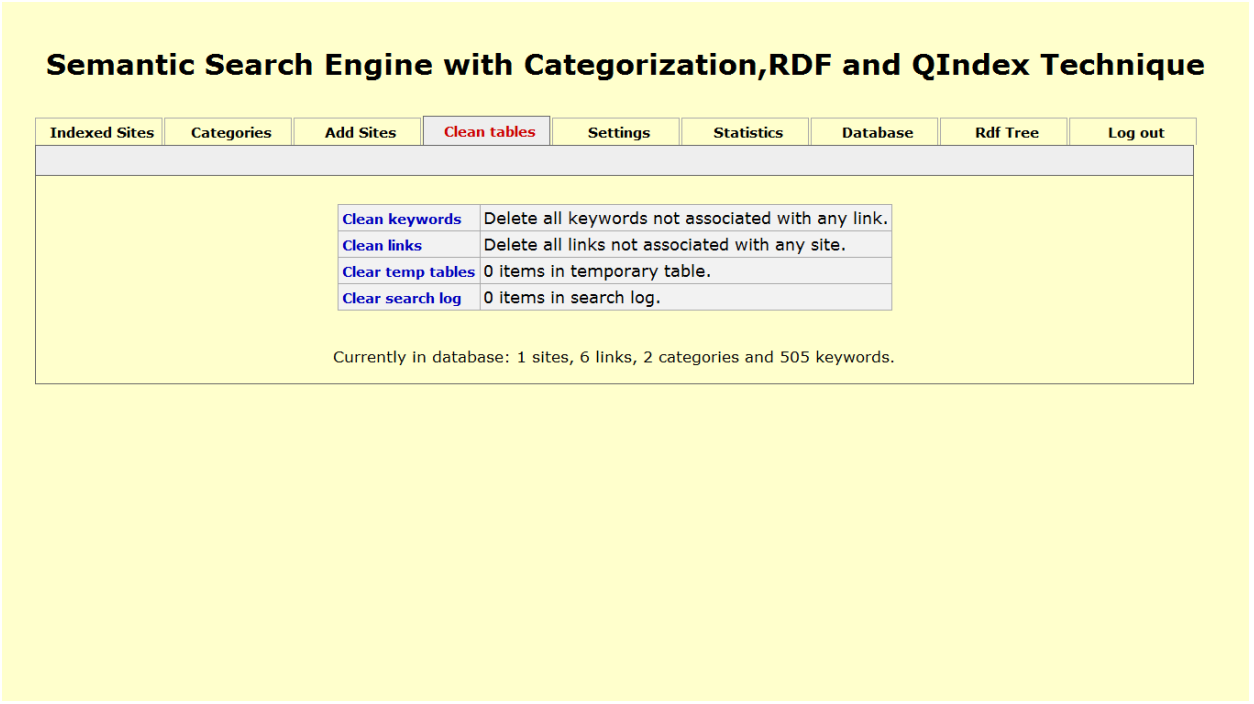
Thus the ontology model and semantic similarity for that model is designed so that the semantic relation of the vertex is increased and the increased similarity results in the optimal ontology semantic search. This optimal solution is verified and improved by the means of greedy algorithm. In future we have decided to implement the solution for more vertex and more relation and to complex system will be designed for improving optimal semantic search.Hence we proposed a method of ontology construction, semantic similarity measurement and genetic algorithm for object. This method is a competitive method in real world market. In future we have decided to construct a similar type of ontology, semantic and genetic approach to certain current real world problems and to find solution. Thus the ontology model and semantic similarity for that model is designed so that the semantic relation of the vertex is increased and the increased similarity results in the optimal ontology semantic search. This optimal solution is verified and improved by the means of greedy algorithm. In future we have derived to implement the solution for more vertex and more relation and to complex system will be designed for improving optimal semantic search.

4 EXPERIMENTAL RESULTS

Clean Tables Option

The clean table's option consists of 4 categories such as clean keywords, clean links, clear temp tables and clear search log.Figure 10 produces clean tabled result.

Figure 10 Clean Tables



Clean keywords

It deletes all the keywords that are not associated with any link.

Clean links

It deletes all links that are not associated with any site.

Clear temp tables

It clears the contents of temp tables.

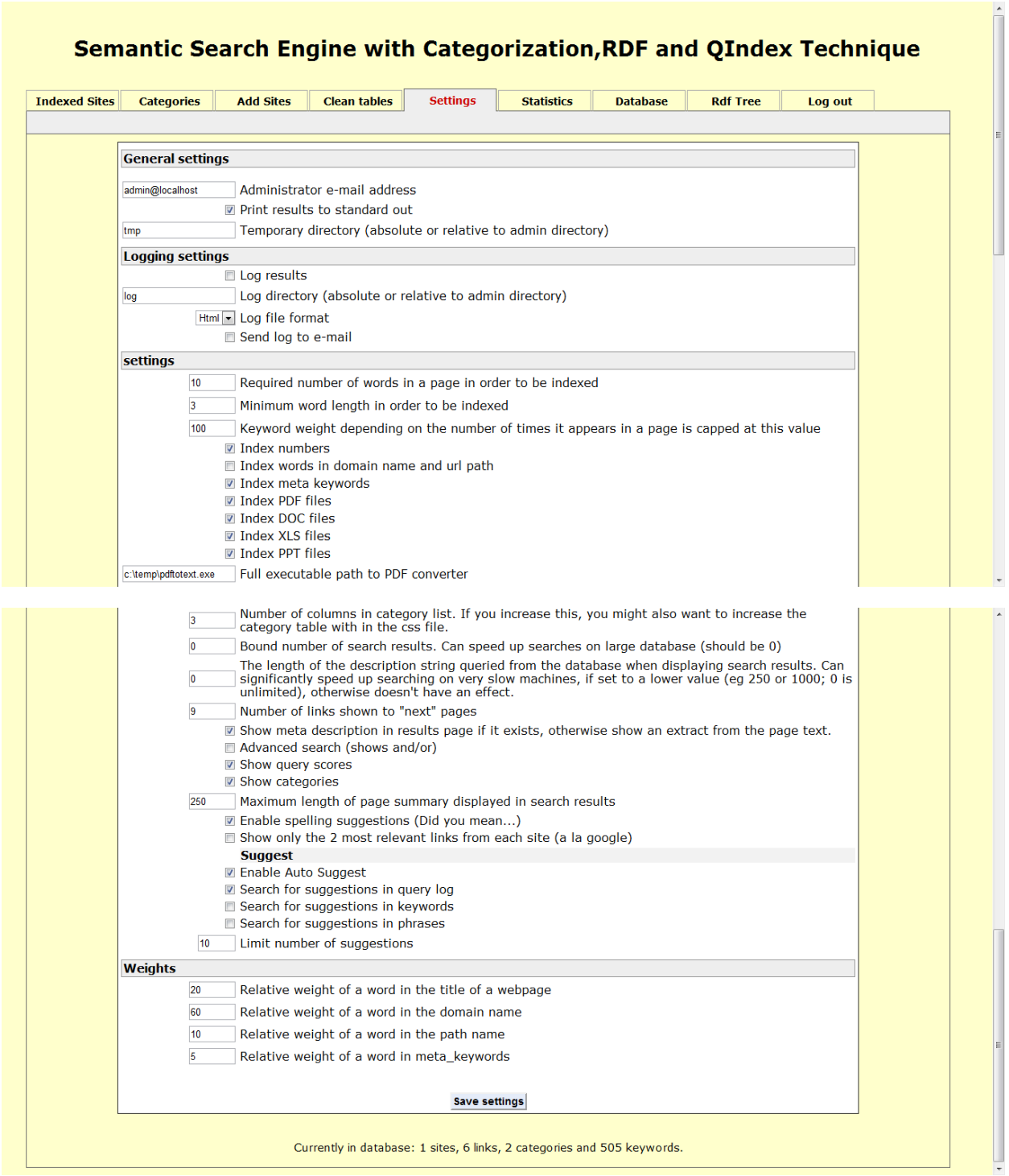
Clear search log

It clears the items present in search log.

Settings Option

The settings option has four categories namely general settings, logging settings, general settings, and weights. These settings can be changed or kept as by default settings. Figure 11 Shows about the Panel Settings

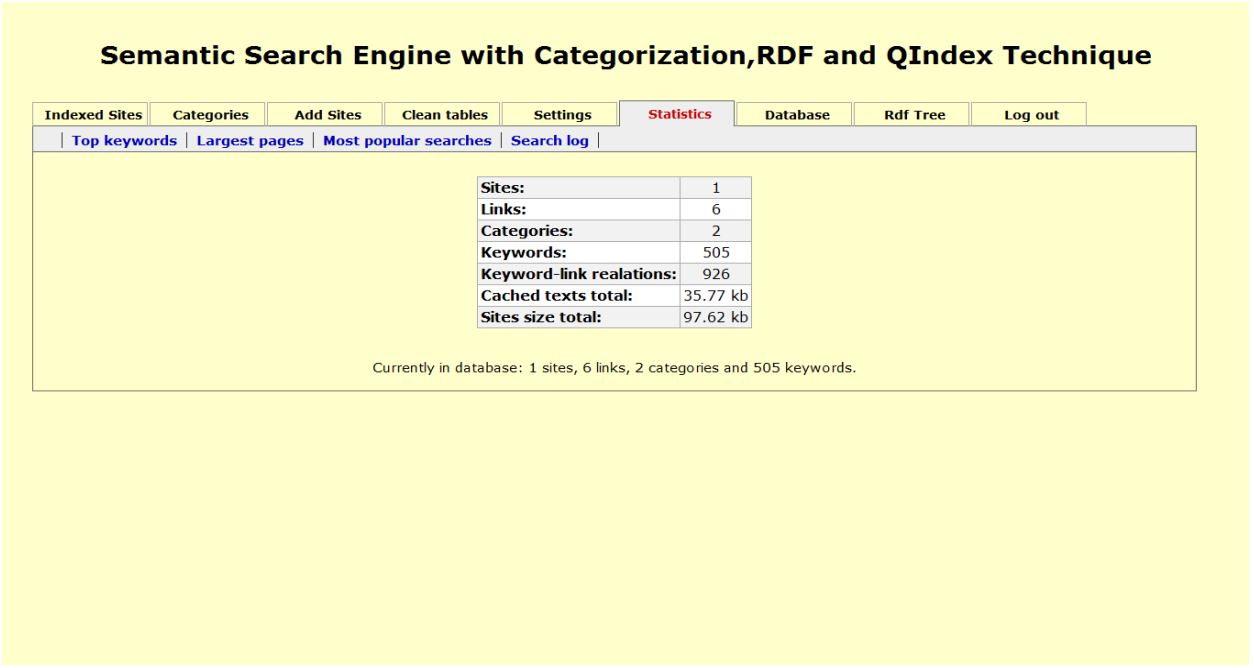
Figure 11 Panel Settings



Statistics

The statistics option consists of four options namely, top keywords, largest pages, most popular searches and search log. Figure 6.12 Shows the Statistics about the search engine

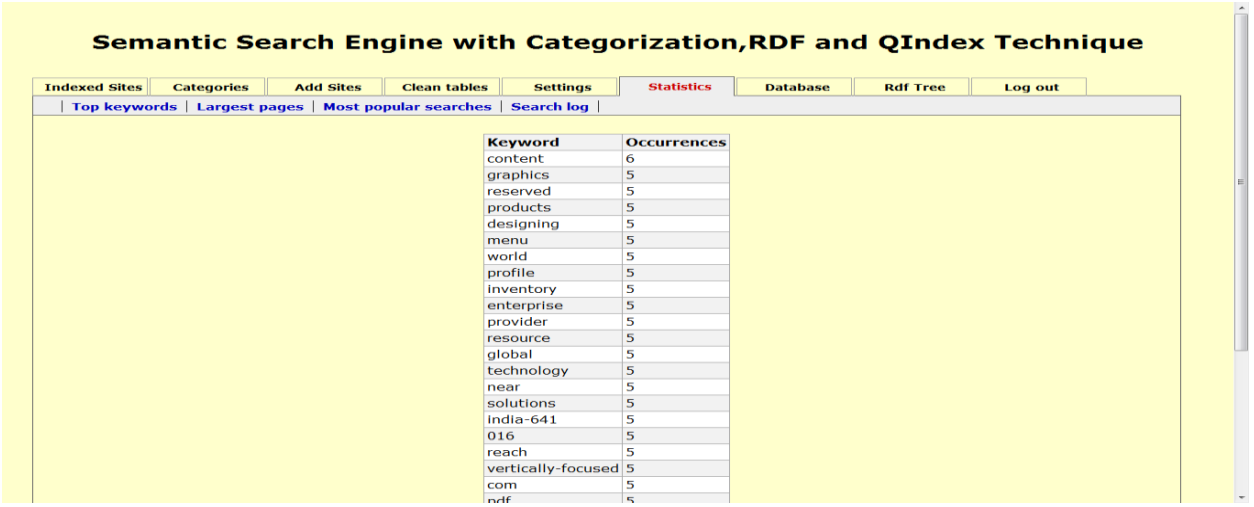
Figure 12 Statistics about the search engine



Top Keywords:

It displays the table containing keywords and its occurrences. The keyword with the maximum number of occurrence stays at the top while the keyword with minimum number of occurrence stays at the bottom. figure 13 produces about top keywords after searching.

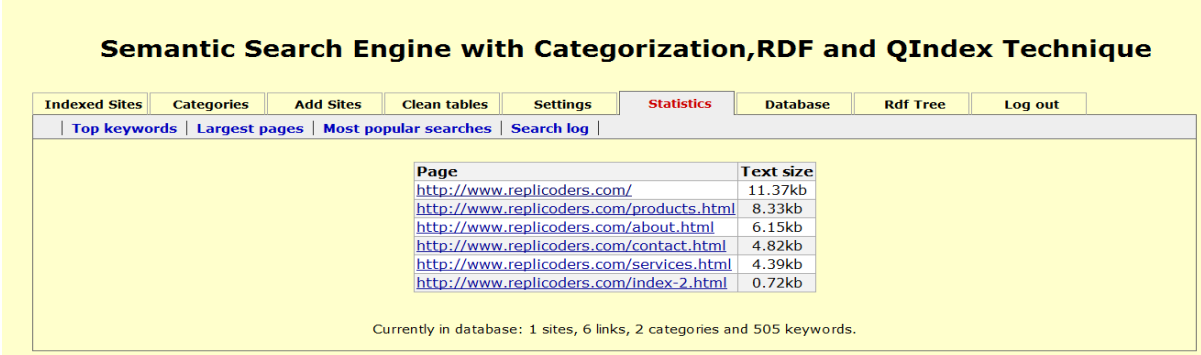
Figure 13 Statistics-Top Keywords



Largest Pages Option:

The pages with the size are at the top option and others follow correspondingly. Figure14 produces results about large pages.

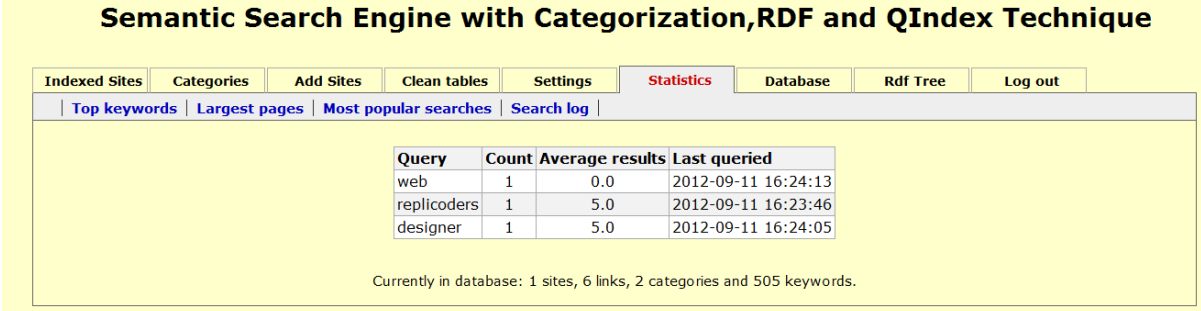
Figure 14 Statistics-Largest Pages



Most Popular Searches:

It consists of the table having the contents such as query, count, last queried, average results. The content with the most popular searches can be viewed.figure 15 produces results about most popular search pages.

Figure 15 Statistics- Most Popular Searches



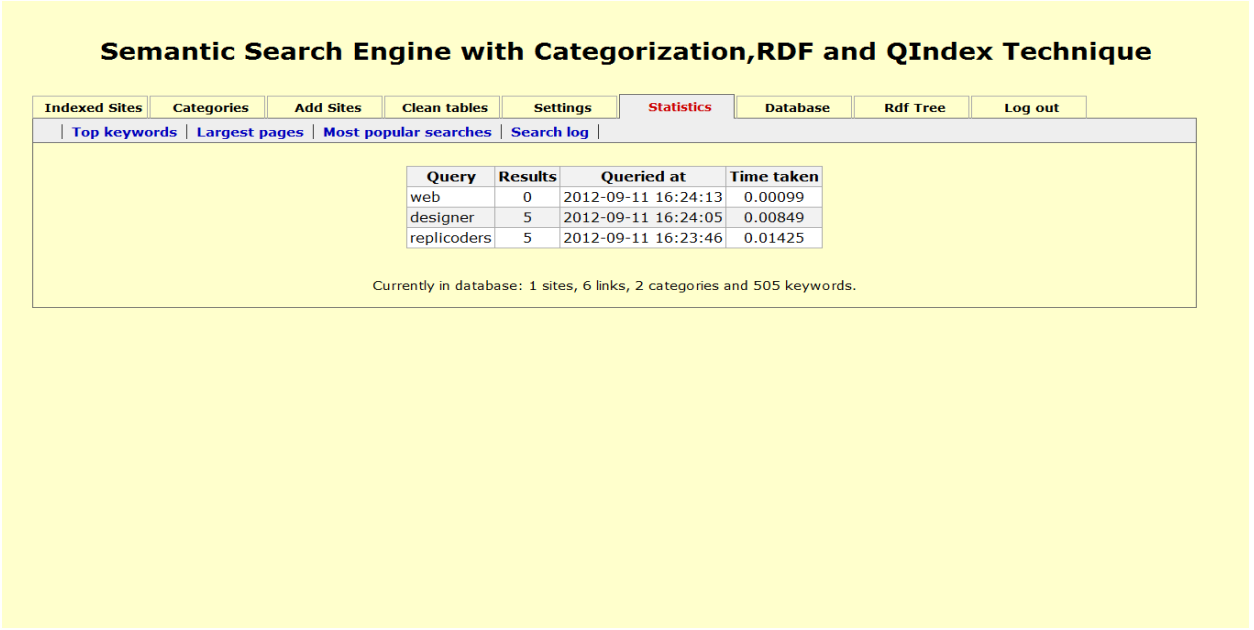
Search Log:

It displays the information about the queries asked by the user. It displays the information of the following contents such as

- Query,
- Results returned,
- Time and
- Total time taken to display the content.

Figure 16 Statistics- Search Log

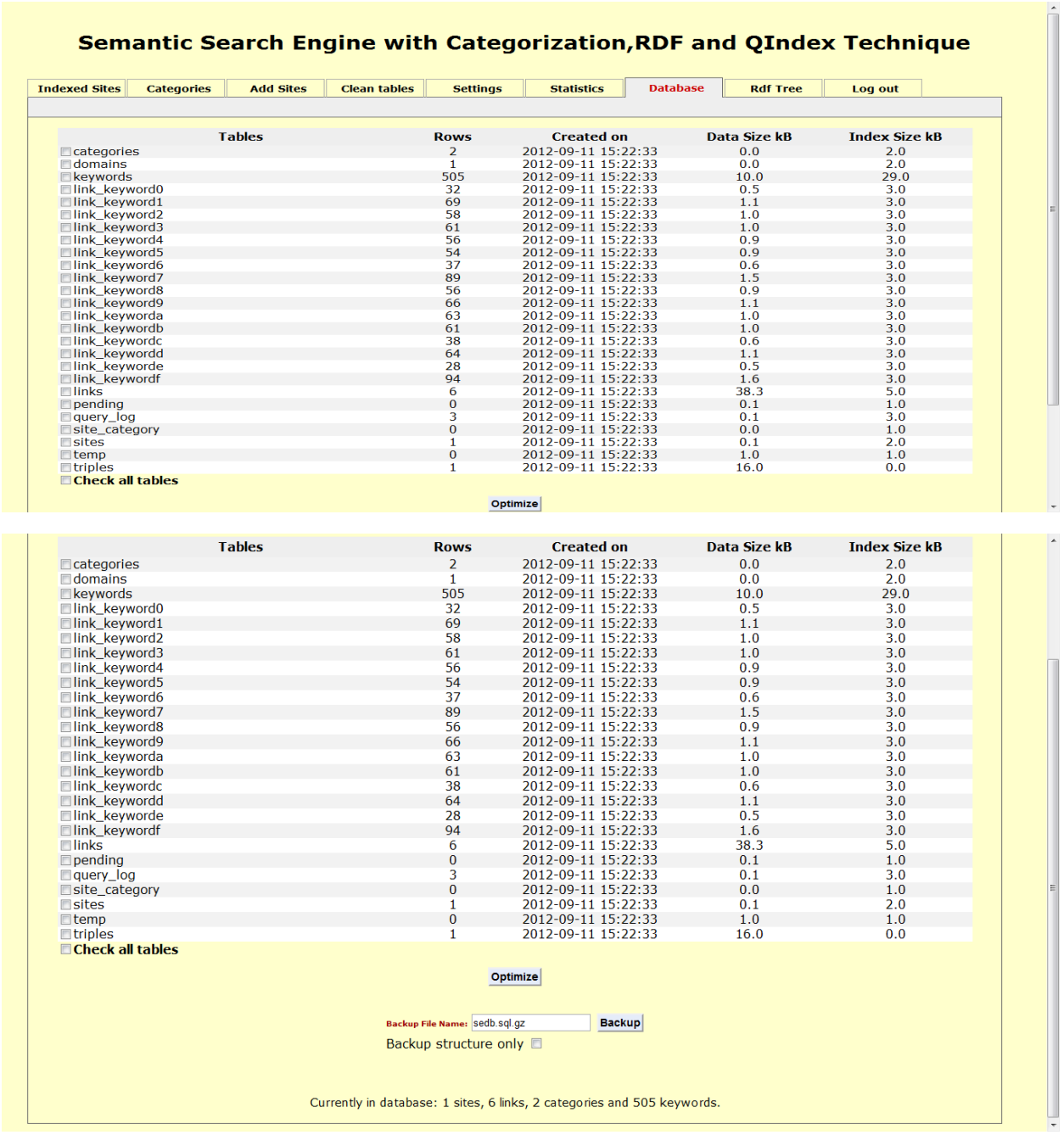
Figure 16 produces results about the search logs.



6.4.4 Database

The database displays the table name, rows present, size, creation date and index size. The tables can be optimized by using optimization option. The whole database can be taken as a backup for further reference. Figure 17 generates results about the data bases.

Figure 17 Database



5 GRAPH-COMPARISON RESULTS

The comparison result of the proposed search engine vs other search engine is compared and the graph is generated based on the range of values.

Table 2,.3 and figure 18 and 19 shows that precision and recall has been improved when it is compared to the football events and an outstanding results are obtained. Here proposed system is compared with [2]Yildirim, Y., *et al.*

Table 2 Optimization of Ontology (Increase in precision and recall)

Name	Precision	Recall	Precision (Proposed System)	Recall (Proposed System)
Pass(event)	87.5	70	88.2	72
Side kick(event)	100	50	100	56
Shot(goal even)	100	100	100	100
Average	90	69	96.3	76

Figure 18 Graph comparison of Optimization of Ontology (Increase in precision and recall)

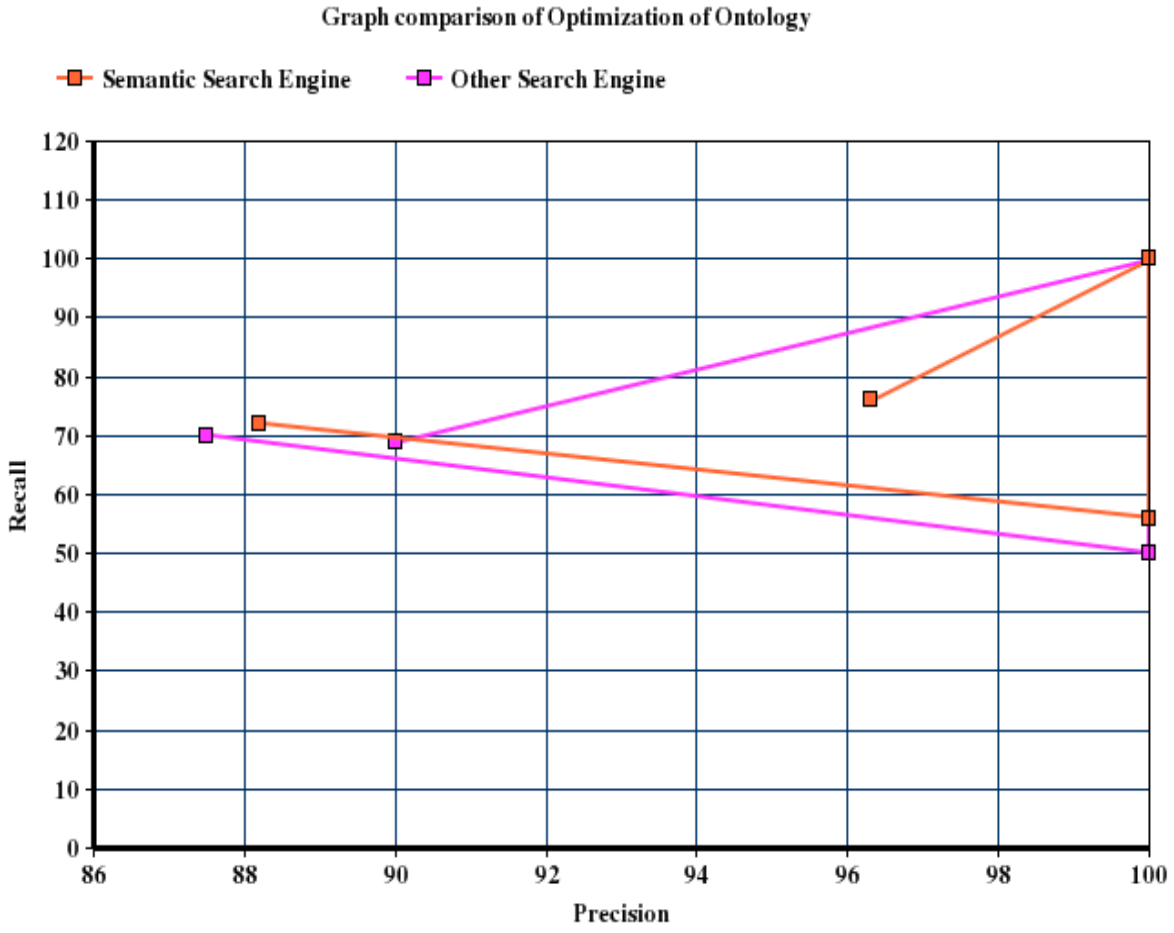


Table 3 Improving Precision and Recall (Increase in precision and recall)

Name	Precision	Recall	Precision (Proposed System)	Recall (Proposed System)
Pass(event)	88	73	91.2	75
Side kick(event)	100	56	100	67
Shot(goal even)	100	100	100	100
Average	93	73	97	78

Figure 19 Graph comparison of Improving Precision and Recall (Increase in precision and recall)

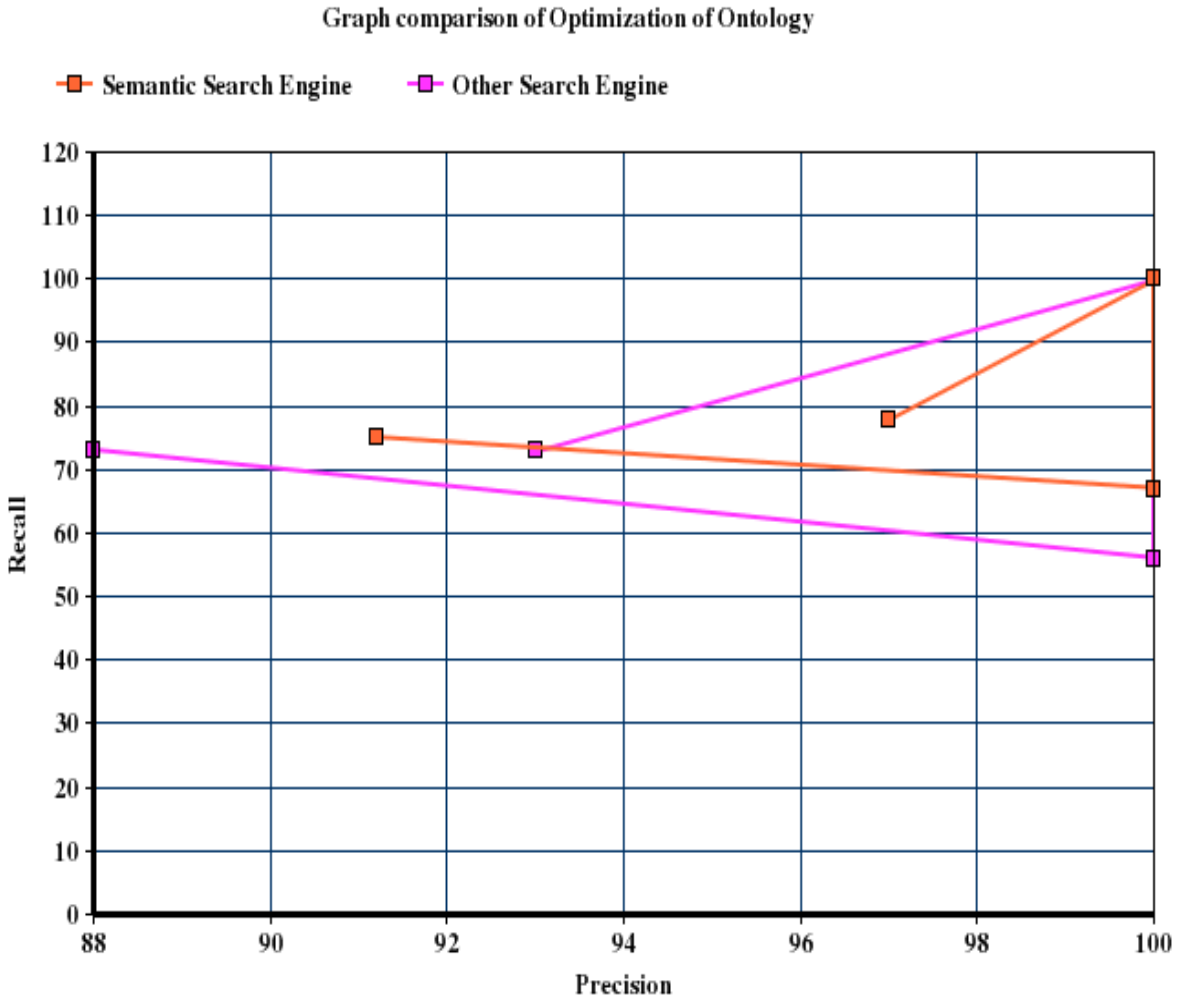


Table 4,.5,6,.7 and figure 19,20,21,22 clearly shows that precision is inproved when the proposed search engine is compared to the swoogle,hakia and lexxe.

Table 4 Total Number of pages visited vs Precision (percent) (Increase in precision)

category	Semantic search engine		Swoogle		Hakia		Lexxe	
	no of pages visited	precision	no of pages visited	precision	no of pages visited	precision	no of pages visited	precision
Replicoders	1	100	10	10	10	33.33	15	6.66
Common	24	25.2	27	34.1	25	25.0	33	30.33
Health	30	33.3	21	49.5	20	14.2	45	55.55
Travel	40	50.0	41	44.2	35	24.1	40	37.5
Education	75	66.6	35	55.6	30	15.7	50	36

Table 5 Total Number of pages visited vs Precision (percent)(Semantic search engine vs Swoogle) (Increase in precision)

Performance of Semantic search engine vs Swoogle

category	Semantic search engine		Swoogle	
	no of pages visited	precision	no of pages visited	precision
Replicoders	1	100	10	10
Common	24	25.2	27	34.1
Health	30	33.3	21	49.5
Travel	40	50.0	41	44.2
Education	75	66.6	35	55.6

Figure 20 Total Number of pages visited vs Precision (percent)(Semantic search engine vs Swoogle) (Increase in precision)

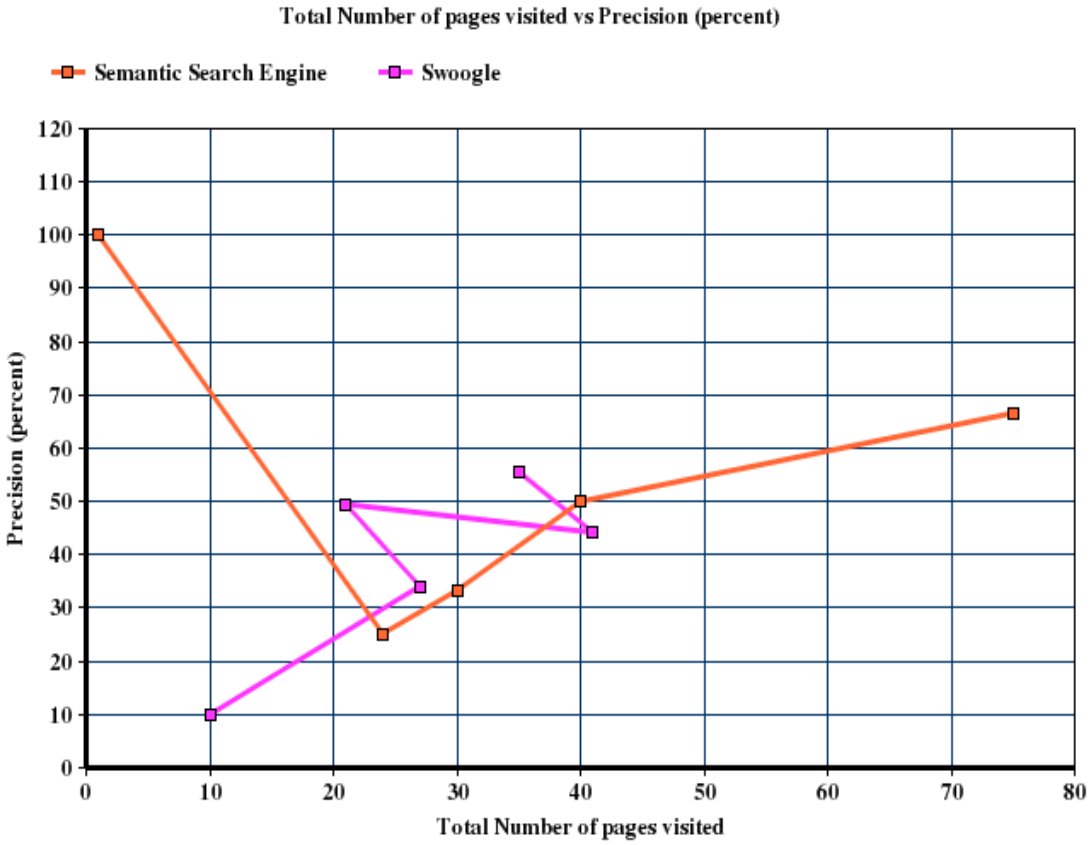


Table 6 Total Number of pages visited vs Precision (percent) (Semantic search engine vs Hakia) (Increase in precision)

Performance of Semantic search engine vs Hakia

category	**Semantic search engine**		**Hakia**	
	no of pages visited	**precision**	**no of pages visited**	**precision**
Replicoders	1	100	10	33.33
Common	24	25.2	25	25.0
Health	30	33.3	20	14.2
Travel	40	50.0	35	24.1
Education	75	66.6	30	15.7

Figure 21Total Number of pages visited vs Precision (percent) (Semantic search engine vs Hakia) (Increase in precision)

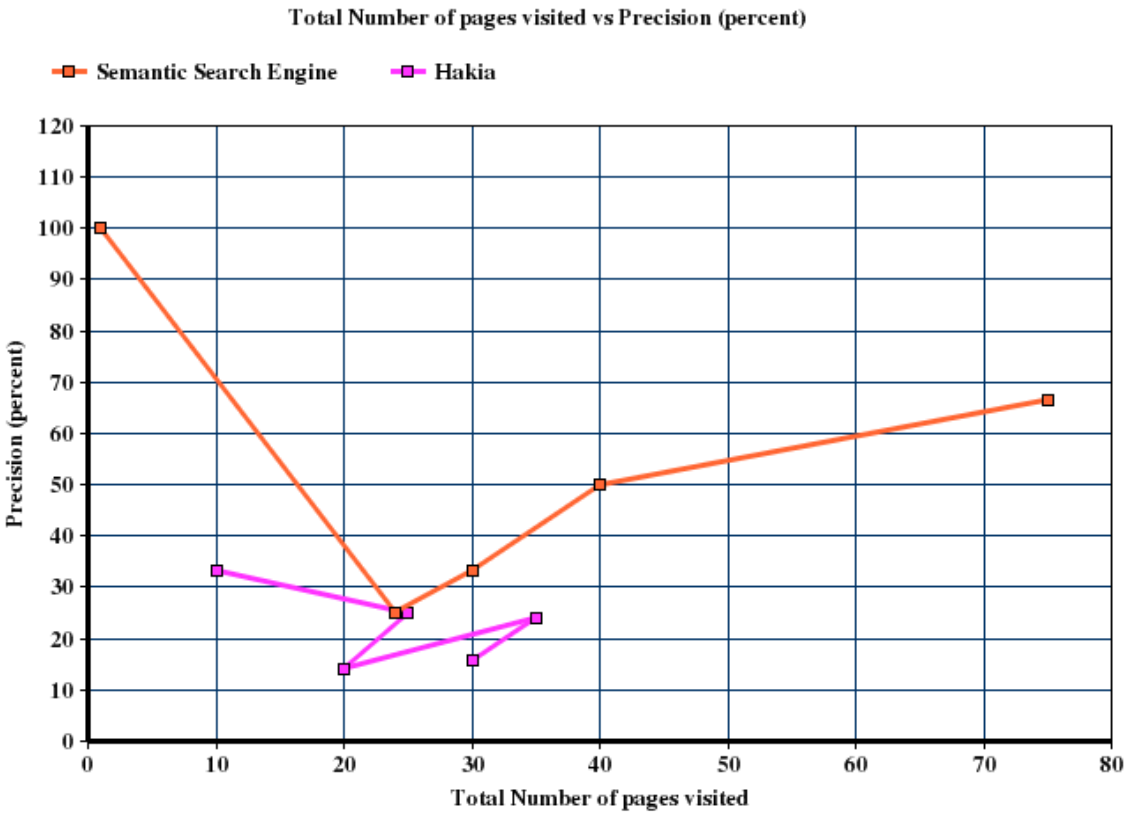
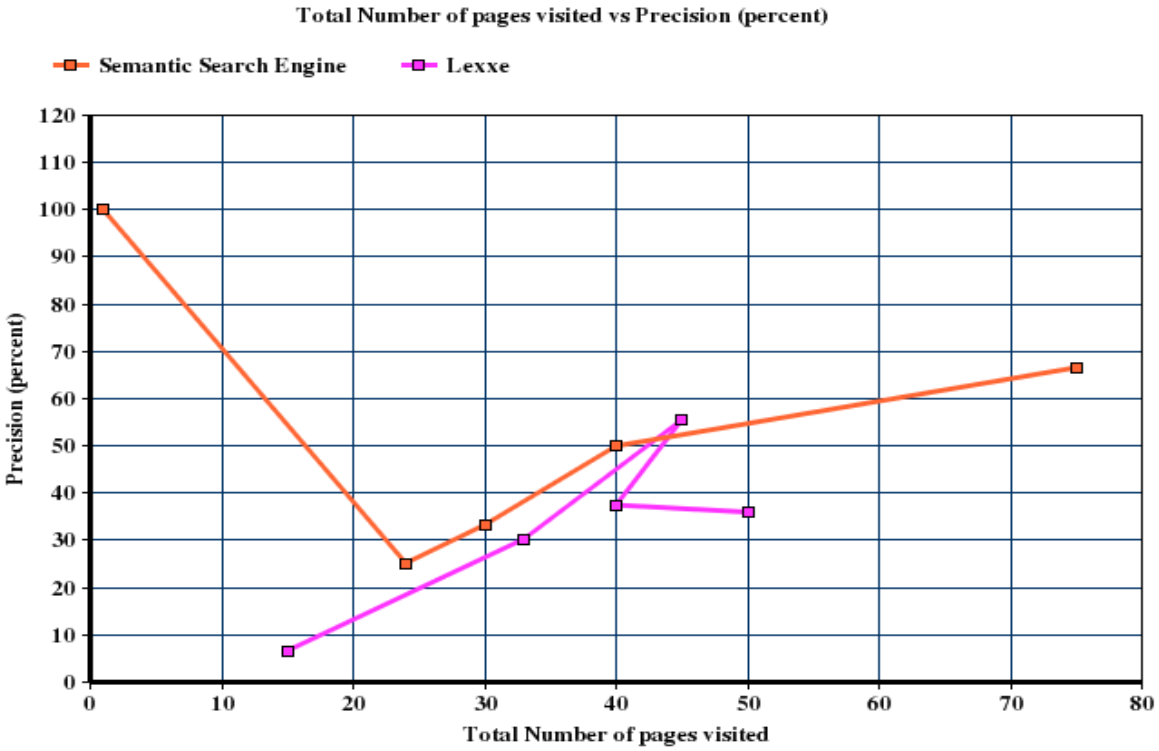


Table 7 Total Number of pages visited vs Precision (percent) (Semantic search engine vs Lexxe) (Increase in precision)

Performance of Semantic search engine vs Lexxe

category	Semantic search engine		Lexxe	
	no of pages visited	precision	no of pages visited	precision
Replicoders	1	100	15	6.66
Common	24	25.2	33	30.33
Health	30	33.3	45	55.55
Travel	40	50.0	40	37.5
Education	75	66.6	50	36

Figure 22 Total Number of pages visited vs Precision (percent) (Semantic search engine vs Lexxe) (Increase in precision)



6 Conclusions

Thus the proposed search engine improves the values of precision and recall, ontology, No of good pages visited and results are compared with the existing semantic search engines.

It generally gives an overview about constructing ontology and improving precision and recall using ontology model, genetic and greedy algorithm, semantic similarity calculation and ontology graph or model

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.

[7] Singh, R., Dhingra, D., Arora, A., "SCHISM—A Web search engine using semantic taxonomy", IEEE Potentials, Vol. 29, Pp. 36 - 40, 2010.

[8] Fergus, R., Li Fei-Fei, Perona, P., Zisserman, A., "Learning Object Categories From Internet Image Searches", IEEE Proceedings, Vol. 98, Pp. 1453 -1466,2010.

[9] García, R., Gil, R., Gimeno, J.M., Granollers, T., López, J.M., Oliva, M., Pascual, A., "Semantic wiki for quality management in software development projects", IET Software, Vol. 4, Pp. 386 - 395, 2010.

[10] Lamberti, F., Sanna, A., Demartini, C., "A Relation-Based Page Rank Algorithm for Semantic Web Search Engines", IEEE Transactions on Knowledge and Data Engineering, Vol. 21, Pp. 123 – 136, 2009.

[11] Halpin, H., Thompson, H.S., "Social Meaning on the Web: From Wittgenstein to Search Engines", IEEE Intelligent Systems, Vol. 24, Pp. 27 – 31, 2009.

[12] S. Latha Shanmugavadivu, Dr. M. Rajaram, "Semantic Based Multiple Web Search Engine", Vol. 02, No. 05, Pp. 1722-1728, 2010.

[13] Tim Finin and Li Ding, "Search Engines for Semantic Web Knowledge", Proceedings of Xtech 2006: Building Web 2.0, 2006.

[14] Alatrash, M., Hao Ying , Dews, P., Ming Dong , Massanari, R.M., "Ranking biomedical literature search result based on relevance feedback using fuzzy logic and Unified Medical Language System" , 2012 Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS), Pp. 1-6,2012.

[15] Rabiyathul basariya, A., Jannath Nisha, O.S., "Semantic based query optimization using ontology for knowledge management", International Conference on Advances in Engineering, Science and Management (ICAESM), Pp. 332-336,2012.

[16] Gousia Tabassum, A.Poongodai, "An Ontology Based Search for Relevant Pages Using Semantic Web", International Journal of Advanced Engineering Sciences and Technologies, Vol.11, Pp. 106 – 110, 2011.

[17] Sheetal A. Takale, Sushma S. Nandgaonkar, "Measuring Semantic Similarity Between Words Using Web Documents", International Journal Of Advanced Computer Science And Applications, Vol. 1, No.4, 2010.

[18] Gaeta, M., Orciuoli, F., Paolozzi, S., Salerno, S., "ontology extraction for knowledge reuse: The e-learning perspective",

IEEE transaction on systems, man and cybernetics Part A: System and humans, Vol.41, Pp. 798-809, 2011.

[19] Alrifai, M., Peter Dolog, Wolf-Tilo Balke, and Wolfgang Nejdl, "Distributed Management of Concurrent Web Service Transactions", IEEE Transactions On Services Computing, Vol. 2, No. 4, 2009.

[20] www.rhizomik.net - RDF reference for relations.

[21] RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation 10 February2004.[Online]. Available: http://www.w3.org/TR/2004/REC-rdf-schema-20040210/

[22] K. Anyanwu and A. Sheth, "ρ-Queries: enabling querying for semantic associations on the Semantic Web," in Proc WWW, Pp. 690-699, 2003.