Optimizing ontology and semantic search using genetic and greedy algorithms approach

* 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
harionto@gmail.com

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

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

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 are noted.
[K. Srihari, V. P. Arunachalam, S Karthik. **Optimizing ontology and semantic search using genetic and greedy algorithms approach.** *Life Sci J* 2013;10(1):2914-2921] (ISSN:1097-8135). http://www.lifesciencesite.com. 354

Keywords: ontology, semantics, greedy, algorithm, searching, relations

1. Introduction

The existing system deals with the fuzzy logic based system where automatic genetic based objects are constructed [1].

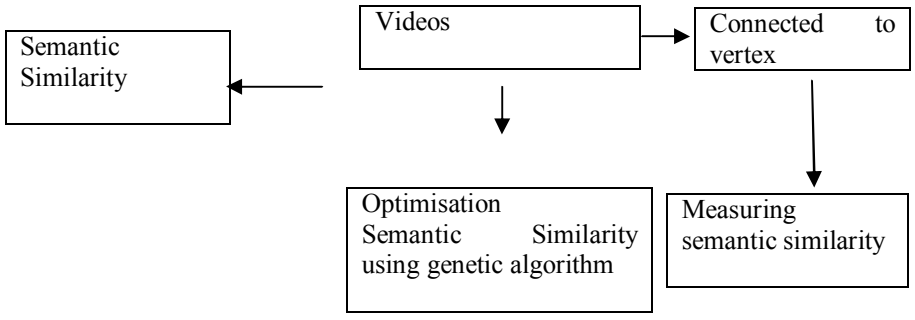
2. Existing system

(1) The main drawback of this system is the accuracy of the relevant measurement is not measured. (2) 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. (3) Here is deal with 2 approaches concept vertex graph (CVG) approach and object genetic measurement (ogm) approach.

3.1Proposed solution:

Working Block Dig:

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



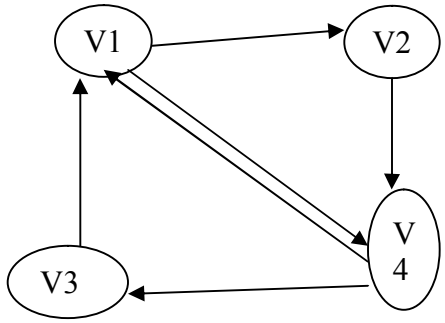
Final result with high semantic similarity.

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

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
Ontology model (dig1)



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
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.

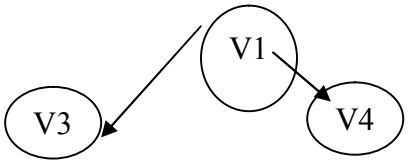
Algorithm for ontology graph

// (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.

Algorithm for genetic optimization

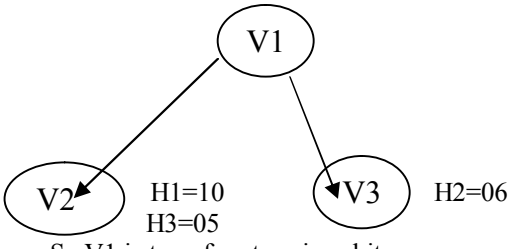
//O1 (V1) is object 1 of video 1 (vertex 1)
// O2 (V2) be object 2 of videos 2 (vertex 2)
// On (Vn) be object n of video (Vn)
// F O1 (V1) Present in On(V2) then
O1 (V1) → parent of On(V2)
O1 (V1) Present in On(Vn) then
O1 (V1) parent of On(Vn)
Else
Check O2 (V2) full On (Vn)
Find parent child relationship 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
O (V2). This child of O1 (V1) =new On(Vn).
End Diagram representation (dig2)



If O1 (V1) Present and O new (V2)
If O1 (V1) present and 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.
Let V1→ Max hit rate for
Hotel Taj
V2→ 2nd max hit rate for hotel residency
V3→ 3rd max hit rate for hotel aloft.

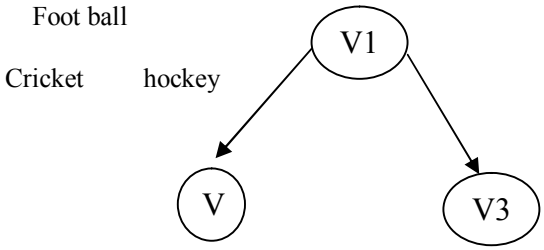
Dig3



So V1 is top of vertex since hits are more.

Semantic Similarity is calculated with the formula and optimal solution is obtained.
Genetic Algorithm.
Here V1 is parent because V1 occurs in all searches 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:

Foot pass
Cricket
Hockey
Foot ball no. of hits is 800(app)
Cricket no. of hits is 700(app)
Foot ball and Cricket is 600(app)
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
4 Comparison of results from existing and proposed solution

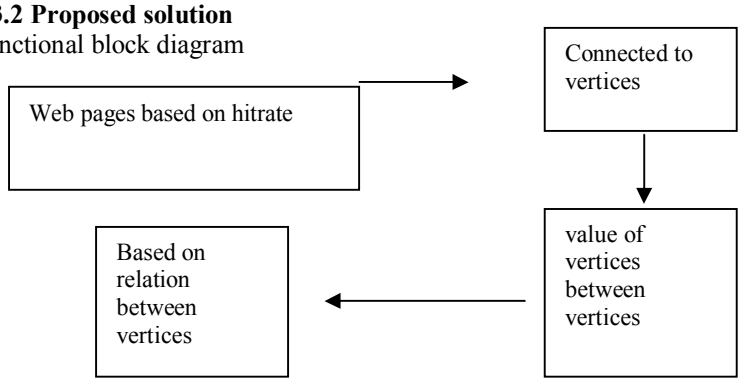
Table 1

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
5 Comparing Rule schema (existing) with proposed genetic algorithm Table 2

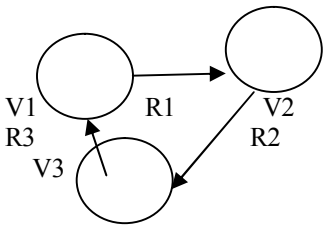
Rule based on logical relations produced by the ontology model designer.	Here hit rate indicates the quality of pages increased and accepted by the user while browsing.
Man made relations.	Quality based relations.
The logical connection is not verified.	Verified by the user comments

3.2 Proposed solution

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 numer 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



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)

	V1	V2	V3
V1 1	0	14	10
V2 2	6	0	12
V3 3	7	10	0

V1>V3=10 V2.V1=6
V1>V2=14 V2 V2=0
v1 v3=10 v2 v3=12

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

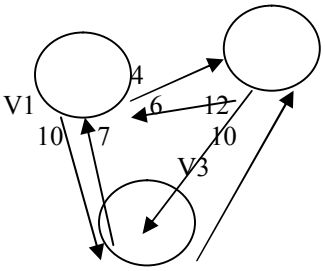
v3 v1=7, v3 v2=10 v3 v3=0

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.

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.

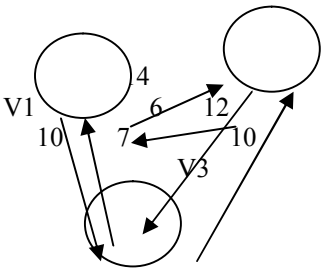


Which implies the cost of brand is 14+12=26. The relation are 2 6 11 only.

The relation for vertex where the value if 0 is calculated. So maximum number 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
Sementic similarity optimization using Greedy algorithm
Consider the diagram



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 travelling 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

Experiment results From experiment we found that more relations arew made (ie)

Test set code	Existing ontology	Proposed solution
01	.69	.90
02	.74	.86
03	.59	.72
04	.73	.79
05	.85	.93
06	.78	.86

semantic retrieval gets increased with more relations.

4 Conclusion and future work:

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.

5 References:

[1] Automatic Semantic Content Extraction in Videos using a Fuzzy Ontology and Rule-Based ModelYakup Yildirim, Adnan Yazici, *Senior Member, IEEE,* and Turgay Yilmaz, *Student Member, IEEE* IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING September 2011

[2] ontology extractioin for knowledge reuse: The e-learning perspective.Gaeta M Orciuoli,F paolozzi,S.Dipt di ing,applicata university of salerno, Fisciano,Italy. IEEE transaction on systems , man and cybernetics Part A : System and humans July 2011.

[3]Distributed Management of Concurrent Web ServiceTransactions Mohammad Alrifai, Peter Dolog, Wolf-Tilo Balke, and Wolfgang Nejdl. IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 2, NO. 4, OCTOBERDECEMBER 2009

[4] (2007) RDF Semantics, W3C Recommendation 10February [Online].Available:http://www.w3.org/TR/2004/ REC-rdf-mt-20040210/

[5] (2007) RDF Vocabulary Description Language 1.0: RDF Schema,W3C Recommendation 10 February2004.[Online]. Available:http://www.w3.org/TR/2004/REC-rdf-schema- 20040210/

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

[7] K. Anyanwu, A. Maduko, and A. Sheth,
"SemRank:Ranking Complex relationship search results on the Semantic Web," in Proc WWW, 2005.

[8] C. Faloutsos, K. S. McCurley, and A. Tomkins, "Fast Discovery of Connection Subgraphs," in Proc Tenth ACM SIGKDD Conference,2004.

[9] C. Ramakrishnan, W. H. Milnor, M. Perry, and A. Sheth,"Discovering informative connection subgraphs in multirelational graphs," SIGKDD Explorations,vol. 7,no 2,pp. 56-63,2005.

[10] S. Mukherjee and B. Bamba, "BioPatentMiner: An information retrieval system for biomedical patents," in Proc thirtieth VLDB Conference,2004.

[11] B. Aleman Meza, M. Nagarjan, C. Ramakrishnan, L Ding, P. Kolari, A. Sheth, I. Arpinar, A. Joshi, and T. Finin, "Semantic Analytics on Social Networks: Experiences in
addressing the problem of Conflict of Interest detection," in Proc WWW,2006.

[12] M. Janik and K. Kochut, "BRAHMS: A WorkBench RDF Store And High Performance Memory System for Semantic Association Discovery," in Proc Fourth International Semantic Web Conference ISWC,2005.

13] (2008) Knoesis Center, Wright State University.[Online].Available: http://knoesis.wright.edu/library/ontologies/sweto /220

[14] Q. T. Tho, S. C. Hui, A. C. M. Fong, and T. H. Cao,"Automatic fuzzy ontology generation for semantic web," IEEE Transactions on Knowledge and Data Engineering,vol.18, no.6, 2006, pp. 842-856.

[15] Muhammad Abulaish, and Lipika Dey, "Interoperability among distributed overlapping ontologies – a fuzzy ontology framework," in: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Hong Kong, 2006, pp.397-403.

[16] Calegari Silvia, and Ciucci Davide, "Fuzzy ontology and fuzzy-OWL in the KAON project," in: Proceedings of 2007 IEEE International Conference on Fuzzy Systems Conference, London, UK, 2007, pp.1-6.

[17] Y. C. Jiang, Z. Z. Shi, Y. Tang, and J. Wang, "Fuzzydescription logic for semantics representation of the Semantic Web," Journal of Software, vol. 18, no. 6, 2007, pp. 1257−1269. (in Chinese)

[18] D. Beckett, and B. McBride, "RDF/XML syntax specification," W3C, 2004, URL http://www.w3.org/TR/rdfsyntax- grammar/.

[19] John W.T. Lee, and Alex K.S. Wong, "Information retrieval based on semantic query on RDF annotated resources," in: Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics, 2004, pp.3220-3225.

[20] D. Brickley, R. Guha, and B. McBride, "RDF vocabulary description language 1.0: RDF schema," W3C, 2004, URL http://www.w3.org/TR/rdf-schema/.

1/11/2013