

Domain Specific -Text Summarization using clustering algorithm and Ontology

Prof.R.V .Argiddi¹, Prof.A .R.Kulkarni² and Laxmi Salgare³

Walchand Institute of technology,Solapur.

ABSTRACT

As we Already take the overview of Text Summarization And Ontology.In this paper we will going to see how The Summary comes with the help of Clustering And Ontology Concept . text summarization comes with the help of 3 simple steps:

- 1) Take One text document.
- 2) create new cluster for each paragraph.Cluster is done with the help of ontology.
- 3) Find the most related words from sentence to summarize.

Keywords:- Agglomerative algorithm ,Clustering, ,Ontology.

1.INTRODUCTION

We introduce a clustering and summarization approach that takes advantage of ontology enriched graphical representations of document. Our approach significantly improves the quality of document clusters and understandability of documents through summaries for document cluster. Hierarchical agglomerative clustering algorithms were used for document clustering. The algorithms successively merge the most similar objects based on the pair wise distances between objects until a termination condition holds. This Clustering done with the help of ontology .This ontologies build with the help of Protégé Software [1].

2.PROPOSED METHODOLOGY

The Overall system shows the flow for text summarization:

We have already saw the overall architecture for summarization.

Input Text- Single one document

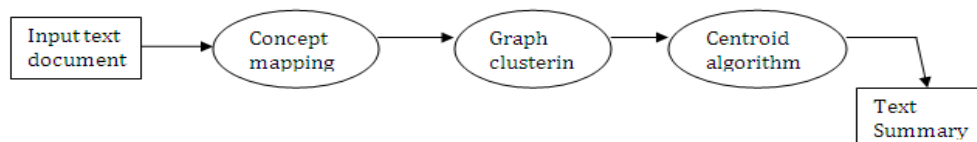


Fig 1.Overall System .Clustering

Agglomerative Clustering algorithm to group similar text information and then text summarization provides condensed text information for the similar text by extracting the most important text content from a similar document set or a document cluster.

Graph Clustering

For each paragraph clustering algorithm generate the one Graph. Graph contains nodes And edges .nodes are sentence and edges how many times each sentence related with another sentence with the help of relatedness of words from that sentence.for ex. Belo figure shows how the clustering is done.

We are using protégé for Graph generation using ontology.

- ex 1) If I play with ball and bat.
- 2) This pitch is very well for batting
- 3) Cricket without player is hard.

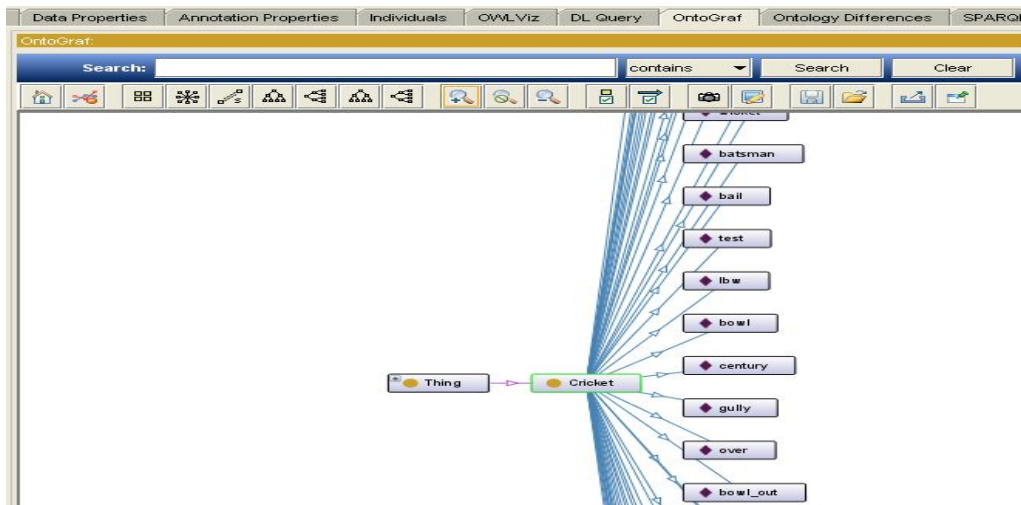


Fig 2. protégé for Graph generation using ontology

From the above three examples we create clusters with the name cluster 1 and cluster 2. Fro above figure Cricket ontology have these words like play, ball, bat, pitch, etc.these words are most related words for creating cluster[3].

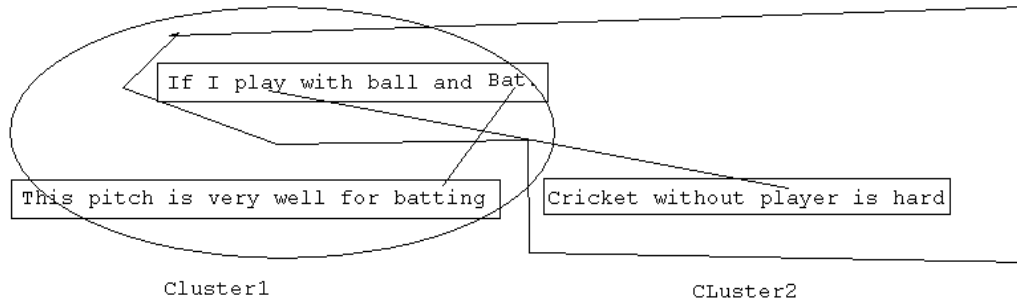


Fig .3 From above fig.2 it should be clear that “If I play with ball and bat”,is taken as summary. Play And bat comes velow in two sentence.so 1st sentence is more threshold value than other two. In Previs Paper Clustering Ontology-enriched Graph Representation method to summarize the text [4].

Executive Summary

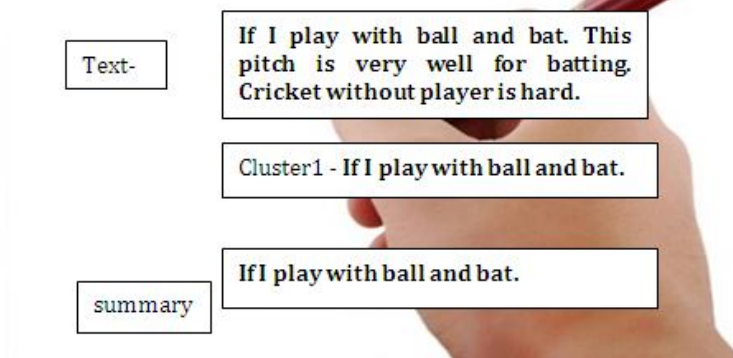


Fig 4. Above Show the execution of summary[5].

3.CONCLUSION

The primary contribution of this paper is we introduce a clustering and summarization approach that takes advantage of ontology enriched graphical representations of document. Our approach significantly improves the quality of document clusters and understandability of documents through summary for document cluster.



REFERANCE

- [1] R. Caves, *Multinational Enterprise and Economic Analysis*, Cambridge University Press, Cambridge, 1982. (book style)
- [2] Erdos, P. and Rényi, A. On the Evolution of Random Graphs. *Publ. Math. Inst. Hungar. Acad. Sci.* 5, 1960, 17-61.
- [3] Guarino, Formal Ontology and Information Systems. In *Proceedings of Formal Ontology and Information Systems*, Trento, Italy. IOS Press, 1998.
- [4] Yoo I., Hu X., and Song I.Y., Clustering Ontology-enriched Graph Representation for Biomedical Documents based on Scale-Free Network Theory, accepted in the *IEEE Conference on Intelligent Systems (IEEE IS'06)*, Sept 4-6, 2006.
- [5] Salton, A. Singhal, M. Mitra, and C. Buckley. 1997. Automatic text structuring and summarization. *Information Processing and Management*,2(32).

AUTHOR



Laxmi Ramchandra Salgare ME Computer Science and Engineering From Walchand Institute Of Technology ,Solapur(2012-2014)