International Journal of Trend in Scientific Research and Development (IJTSRD)
International Open Access Journal
ISSN No: 2456 - 6470 | www.ijtsrd.com | Volume - 2 | Issue – 3

# Novel Algorithm for Multi-document Summarization using Lexical Concept

**Srashti Garg [1], Dr. Akash Saxena[2]**
[1]M.Tech Scholar, [2]Professor,
[1,2]CompuCom Institute of Information Technology and Management, Jaipur, Rajasthan, India

## ABSTRACT

Text summarization is the necessity of the society as we are surrounded my various documents which if summarized will not only save our time and but also let us to go through more number of documents in the same time. In this paper we presented the a novel approach for multiple document summarization using the lexical chains with taken into concern the adjective, adverbs , nouns etc.. ,for the formation of the lexical chains. Together with that the better approach is used for the tagging which results in better results for recall when compared the results with the base paper.

***Keywords:*** *Text Summarization, WordNet, Multi-document*

## I. INTRODUCTION:

Text summarization has transformed into a vital and advantageous gadget for helping and deciphering text information in the present rapidly creating information age. Enormous extending and basic accessibility of information on the World Wide Web have as of late achieved reviewing the established etymology issue the buildup of information from text reports [1]. This errand is basically an information lessening process. The goal of programmed text summarization is merging the source text into a shorter interpretation securing its information substance and general importance. Text summarization is the procedure of consequently making a compacted variation of a given archive pre-serving its information content. Programmed record summarization is a critical research zone in normal dialect handling (NLP). The innovation of programmed record summarization is creating [1] and may give a response for the information over-trouble issue.

Summary may imply abstract summary, compressed version or authority summary. An abstract is a succinct summary of an exploration article, hypothesis, overview, gathering proceeding or any start to finish examination of a particular subject or instruct, and is frequently used to help the peruser quickly take in the paper's inspiration. [1] When used, an abstract constantly appears toward the beginning of an original copy or typescript, going about as the reason for entry for any given academic paper or patent application.

### WordNet

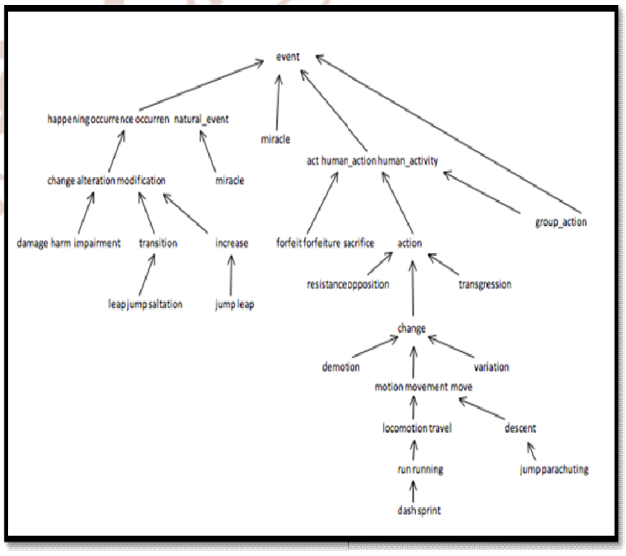



**Figure 1: Wordnet**

WordNet may be a lexical database for the English dialect. It contains or engineers the English words into sets of proportional words suggested as synsets, give short definitions and utilize cases, and records arrangement of relations among these equal word sets or their people. WordNet will in this way be seen as a mix of vocabulary and thesaurus. Despite the fact that it's available to human customers by methods for a web program, its fundamental use is in programmed text examination and fake cognizance applications. The database and programming gadgets are discharged underneath a BSD style allow and are wholeheartedly out there for download from the WordNet website. Both the lexicographic information (etymologist records) and besides the compiler (called crush) for coming to fruition or giving the scattered database are open.

WordNet joins the lexical groups things, verbs, modifiers and intensifiers however neglects social words, determiners and alternative work words.

Words from undefined lexical arrangement that are synonymous are requested into synsets. Synsets incorporates simplex words and besides as collocations incline toward "eat out" and "auto pool". extremely shocking resources of a polysemous word kind are assigned to different synsets. The significance of a synset is more lit up with a succinct shimmer and one or additional utilization cases.

## 2. Related Study

ShwetaSaxena et al[2] Automatic Text Summarization is a fascinating theme for explore. Still it is developing on. Addition of the data is exponentially developing on and it turns out to be excessively hard to discover the right or pertinent data in gigantic measure of data. So it ends up critical for analysts to utilize it for effective recovery of information. Thus Text Summarization assumes a vital part for this issue. Summarization gives the short form for the text record which contains the fundamental context of the archive. Summarization can be characterized into two classifications: Extractive and Abstractive. This paper \presents the extractive summary utilizing lexical fastening approaches. Lexical chains are made by utilizing Knowledge based database i.e. Wordnet. This paper contrasts comes about and the customary strategies and gives better outcomes.

The above talked about strategy is contrasted and the two existing techniques on similar records with same % of summary. These techniques have been assessed on review and exactness highlights. The review factor is more vital which tells that what amount proposed technique is critical in separating the important or precise sentences. Henceforth from the table and graphical portrayal, proposed technique is noteworthy. In a few records proposed calculation needed. The outcomes (summary) of proposed calculation are assessed with the human created summary for each document.

NimishaDheer etal [3] The present innovation of programmed text summarization bestows a critical part in the information recovery (IR) and text order, and it gives the best answer for the information over-burden issue. Text summarization is a procedure of decreasing the span of a text while securing its information content. When thinking about the size and number of records which are accessible on the Internet and from alternate sources, the necessity for an exceptionally proficient device on which produces usable rundowns is clear. They exhibit a superior calculation utilizing lexical chain calculation &WordNet. The calculation one which makes lexical chains that is computationally achievable for the client. Utilizing these lexical chains the client will create a summary, which is considerably more successful contrasted with the arrangements accessible and furthermore nearer to the human produced summary.

H. Gregory Silber etal [4] ,The expanded in the development of the net has brought about enormous measures of information that has turned out to be harder to access with productivity. Web clients require instruments to deal with this gigantic measure of information. The primary objective of this examination is to frame a temperate and compelling device that is ready to condense very huge archives rapidly. This examination shows a direct time algorithmic discount for finding lexical chains that could be a procedure of catching the "suddenness" of an archive. They moreover give distinctive systems for removing and assessment lexical chains. They demonstrate that their method gives comparable outcomes to past examination, however is impressively very more effective. This effectiveness is vital in web seek applications where a few very extensive reports may must be compressed quickly, and where the response time to the end client is extremely essential.

Shape this paper, we have learned and roused by the idea of the lexical chains, and how they are made and connected in the field of the text summarization.

From this paper, we have additionally taken in the idea of how to score the chain and discover the convenience of the chains.

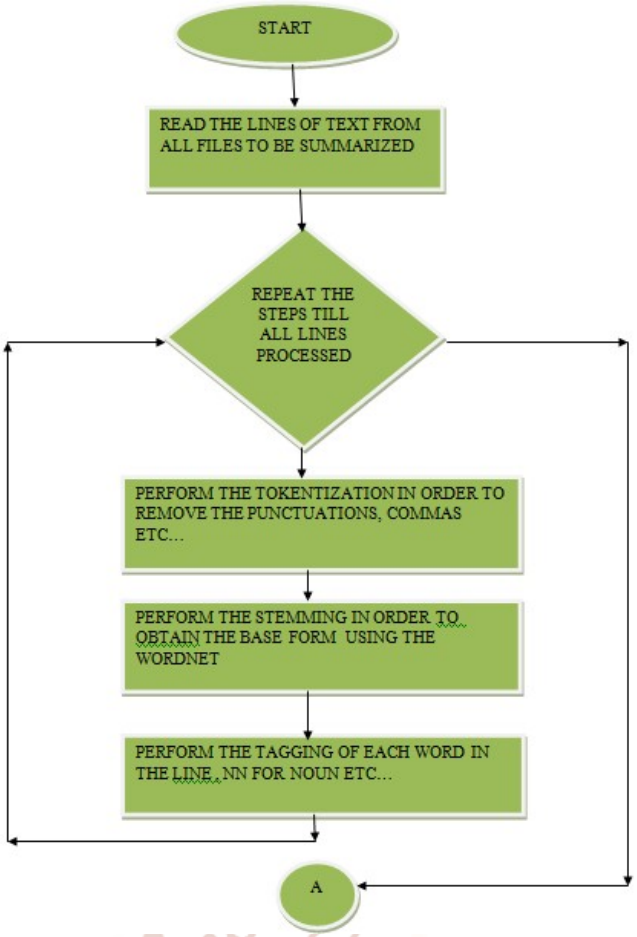
### 3. Problem Description

Text Summarization is constantly required in the each part of our working. It is critical when we are perusing and surfing archives which are very long, investigating through every single sentence of such a report is very unthinkable and these would me be able to number of pages moreover. Thus, summary is dependably the correct answer for that.

Presently, robotized produced summary will likewise assume an extraordinary part. In the base paper "Extractive Automatic Text Summarization through Lexical Chain Method utilizing WordNet Dictionary " by NimishaDheer, ChetanKumar, IEEE 2016 , summary is created utilizing the lexical chains and the idea of word net is additionally utilized. Yet at the same time it experiences some restriction, similar to its review esteem i.e. the quantity of words coordinated with the real summary is additionally low and in the summary age process, it takes excessively time And together , we will jump at the chance to take a shot at the future regions which they have proposed in their paper, Document grouping is the principle advance forward towards the recognizable proof of the different portrayals in a multi-report accumulation. Exact and better similitude measure bestows a vital part during the time spent deciding the general proficiency of the bunches or parts of the record. They have processed the likeness measure and counts in light of the cover of things and formal people, places or things between two words and the portions. We will broaden this comparability measure including verb, intensifier and so on. Multi-archive summarization is as yet an extremely unpredictable and troublesome assignment as proposed by them. We have stretched out our exploration to create the synopses identified with different archives at once..

### 4. Proposed Methodology

Multi-document summarization is a programmed system went for extraction of information from different texts expounded on a similar point. The subsequent summary report permits singular clients, for example, proficient information customers, to rapidly acquaint themselves with information contained in an extensive bunch of documents. In such a way, multi-document summarization frameworks are supplementing the news aggregators playing out the subsequent stage not far off of adapting to information over-burden.



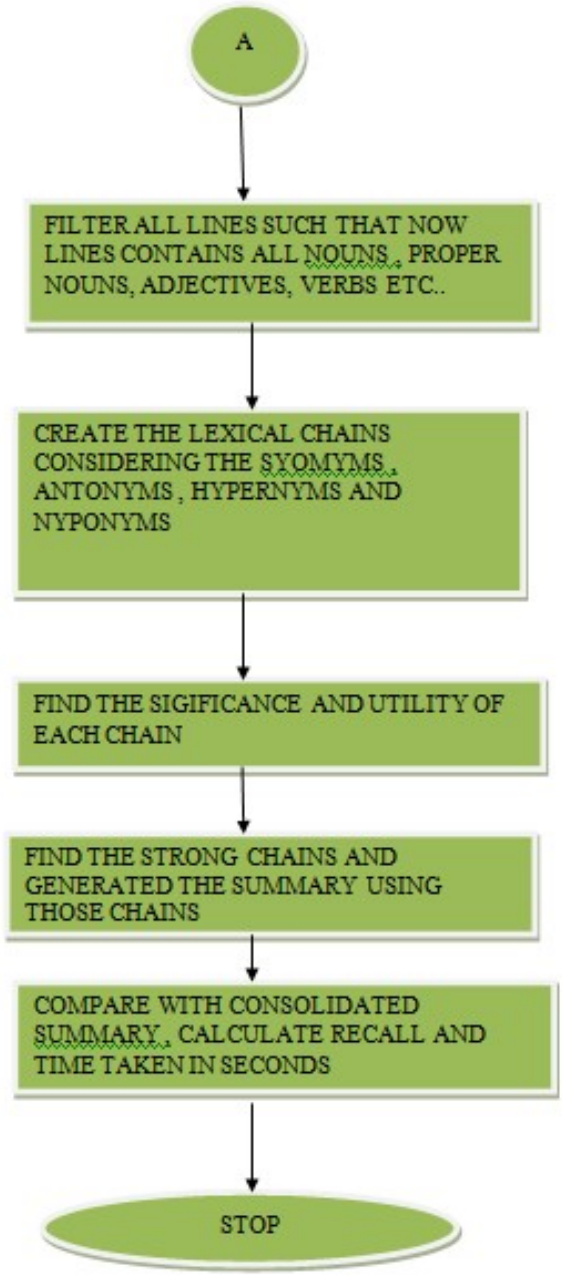



**Figure 2 Proposed Concept Flowchart**

Multi-document summarization is a programmed system went for extraction of information from different texts expounded on a similar point. The subsequent summary report permits singular clients, for example, proficient information customers, to rapidly acquaint themselves with information contained in an extensive bunch of documents. In such a way, multi-document summarization frameworks are supplementing the news aggregators playing out the subsequent stage not far off of adapting to information over-burden.

We have expanded our examination in condensing the different documents at once , with the goal that it will decrease the work stack and will spare the time it getting the total significance or summary of the huge documents.

## 5. Test Results

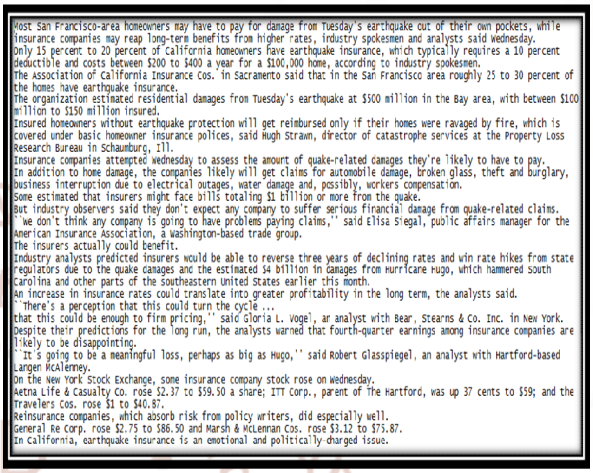

Most San Francisco-area homeowners may have to pay for damage from Tuesday's earthquake out of their own pockets, while insurance companies may reap long-term benefits from higher rates, industry spokesmen and analysts said Wednesday.
Only 15 percent to 20 percent of California homeowners have earthquake insurance, which typically requires a 10 percent deductible and costs between $200 to $400 a year for a $100,000 home, according to industry spokesmen.
The Association of California Insurance Cos. in Sacramento said that in the San Francisco area roughly 25 to 30 percent of the homes have earthquake insurance.
The organization estimated residential damages from Tuesday's earthquake at $500 million in the Bay area, with between $100 million to $150 million insured.
Insured homeowners without earthquake protection will get reimbursed only if their homes were ravaged by fire, which is covered under basic homeowner insurance polices, said Hugh Strawn, director of catastrophe services at the Property Loss Research Bureau in Schaumburg, Ill.
Insurance companies attempted Wednesday to assess the amount of quake-related damages they're likely to have to pay.
In addition to home damage, the companies likely will get claims for automobile damage, broken glass, theft and burglary, business interruption due to electrical outages, water damage and, possibly, workers compensation.
Some estimated that insurers might face bills totaling $1 billion or more from the quake.
But industry observers said they don't expect any company to suffer serious financial damage from quake-related claims.
"We don't think any company is going to have problems paying claims," said Elisa Siegal, public affairs manager for the American Insurance Association, a Washington-based trade group.
The insurers actually could benefit.
Industry analysts predicted insurers would be able to reverse three years of declining rates and win rate hikes from state regulators due to the quake damages and the estimated $4 billion in damages from Hurricane Hugo, which hammered South Carolina and other parts of the southeastern United States earlier this month.
An increase in insurance rates could translate into greater profitability in the long term, the analysts said.
"There's a perception that this could turn the cycle ...
that this could be enough to firm pricing," said Gloria L. Vogel, an analyst with Bear, Stearns & Co. Inc. in New York.
Despite their predictions for the long run, the analysts warned that fourth-quarter earnings among insurance companies are likely to be disappointing.
"It's going to be a meaningful loss, perhaps as big as Hugo," said Robert Glasspiegel, an analyst with Hartford-based Langen McAlenney.
On the New York Stock Exchange, some insurance company stock rose on Wednesday.
Aetna Life & Casualty Co. rose $2.37 to $59.50 a share; ITT Corp., parent of The Hartford, was up 37 cents to $59; and the Travelers Cos. rose $1 to $40.87.
Reinsurance companies, which absorb risk from policy writers, did especially well.
General Re Corp. rose $2.75 to $86.50 and Marsh & McLennan Cos. rose $3.12 to $75.87.
In California, earthquake insurance is an emotional and politically charged issue.

**Figure 3** Sample Document **1**

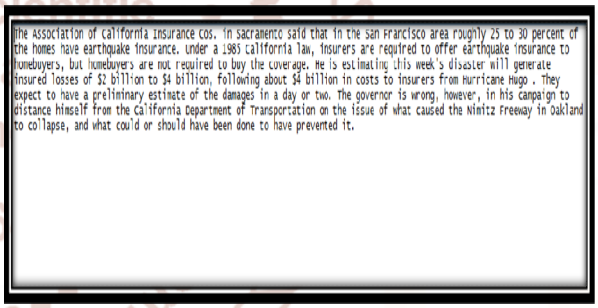

The Association of California Insurance Cos. in Sacramento said that in the San Francisco area roughly 25 to 30 percent of the homes have earthquake insurance. Under a 1985 California law, insurers are required to offer earthquake insurance to homebuyers, but homebuyers are not required to buy the coverage. He is estimating this week's disaster will generate insured losses of $2 billion to $4 billion, following about $4 billion in costs to insurers from Hurricane Hugo . They expect to have a preliminary estimate of the damages in a day or two. The governor is wrong, however, in his campaign to distance himself from the California Department of Transportation on the issue of what caused the Nimitz Freeway in Oakland to collapse, and what could or should have been done to have prevented it.

**Figure 4** Standard Summary DataSet **1**

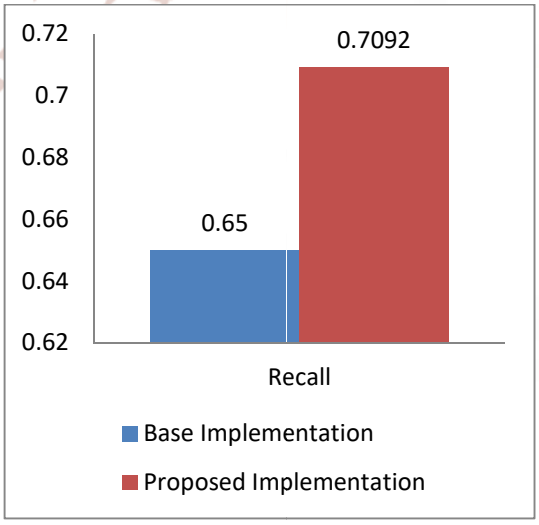



**Figure 5 Graphical** Comparison For Dataset 1 on Recall basis

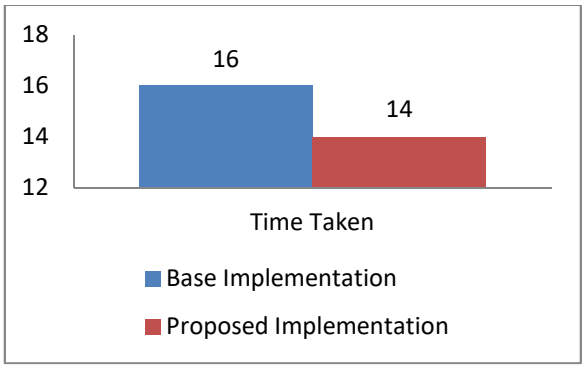



**Figure 6** Graphical Comparison For Dataset 1 on Time basis

This graphs in Figure 5 and Figure 6 also show the histogram based comparison on the basis of the recall and time taken for Dataset 1respectively.

In the table 6.1 we have compare the efficiency of the both the base and the proposed algorithm on the basis of the recall and time taken. In the Dataset 1 , the percentage match with the standard summary is .65 i.e. 65% similarity and that for the proposed is .70.92 i.e. 70.92%. And the time taken by base is 16 seconds to complete the process and proposed work complete that in the 14 seconds.

| | Base Implementation | Proposed Implementation |
|---|---|---|
| Recall | .65 | .7092 |
| Time Taken | 16 | 14 |

**Table 1:** LTB division results

## Conclusion

The document summarization issue is a critical issue because of its effect on the information recovery techniques and additionally on the effectiveness of the basic leadership forms, and especially in the period of Big Data Analysis. In spite of the fact that great sort of text summarization systems and calculations are created there's a prerequisite for growing new ways to deal with supply exact and solid document summaries that may endure varieties in document attributes.

In this postulation, we displayed a technique to discover the lexical chains as an effective transitional portrayal of our document. Alongside WordNet API, our technique additionally incorporated the things, pronoun. descriptive word, verb and so forth in the calculation of lexical chains. Furthermore, the measurable counts in our proposed philosophy brought about the better yield when contrasted with the base paper.

## References

1) SurajitKarmakar, Tanvi Lad, HitenChothani,"A Review Paper on Extractive Techniques of Text Summarization",International Research Journal of Computer Science (IRJCS),2015
2) ShwetaSaxena , AkashSaxena, PhD ,"An Efficient Method based on Lexical Chains for Automatic Text Summarization",International Journal of Computer Applications (0975 – 8887) Volume 144 – No.1, June 2016
3) NimishaDheer Mr. Chetan Kumar ,"Extractive Automatic Text Summarization through Lexical Chain Method using WordNet Dictionary", IEEE 2016
4) H. Gregory Silber Kathleen F. McCoy,"Efficient Text Summarization Using Lexical Chains",International Journal of Research in Engineering and Technology ,2013
5) Kupiec, J., Pedersen, J., and Chen, F,"A trainable document summarizer. In Proceedings SIGIR", USA,1995.
6) Lin, C.-Y. andHovy, E.,"Identifying topics by position", In Proceedings of the Fifth conference on Applied natural language processing, USA, 1997.
7) Conroy, J. M. and O'leary, D. P. ,"Text summarization via hidden markov models",In Proceedings of SIGIR ,USA,2001.
8) Osborne, M.,"Using maximum entropy for sentence extraction",In Proceedings of the ACL Workshop on Automatic Summarization, May 2015
9) Nenkova, A. , "Automatic text summarization of newswire: Lessons learned from the document understanding conference". In Proceedings of AAAI 2005,USA,2005.
10) ReginaBarzilay and Michael Elhadad,Using Lexical Chains for Text Summarization ,,University of Israil , 2013
11) Nikita Munot,Sharvari S. Govilkar ,Comparative Study of Text Summarization Methods,International Journal of Computer Applications (0975 – 8887) Volume 102– No.12, September 2014
12) A.R.Kulkarni,S.S.Apte , "An Automatic Text Summarization Using Lexical Cohesion And Correlation Of Sentences ",Ijret: International Journal of Research in Engineering and Technology ,2014