# EXTRACTIVE TEXT SUMMARISATION TECHNIQUES- A SURVEY

**Meena Siwach[1], Suman Mann[2,] Sarthak Jain[3], Jatin Rauthan[4]**

*[1,2] Assistant Professor, Dept of IT, Maharaja Surajmal Institute of Technology, Delhi, India*
*[3,4]Student , Dept of IT , Maharaja Surajmal Institute of Technology, Delhi, India*

---

**Abstract** *There is a huge flow of knowledge on internet nowadays on every topic. Summarizing that information into short form would be beneficial for a lot of users. There is an imminent need to automatically summarise the text to save both time & resources of the users. Automatic Text summarisation is a process through which a synopsis is produced of a very long text into short form which contains only meaningful & useful information of any topic. Text summarisation was first came to use in 1950s. Since then there is a huge interest among the researchers to explore new & modern ways of text summarisation so that summaries produced by these techniques matches with human made synopsis. There are two broad ways of producing summaries -: 1) Abstractive summarisation 2) Extractive summarisation. Methods of abstraction are more complicated as they require Natural Dialect Processing to a large extent so that's why now researchers are aiming more for finding extractive methods trying to get more accurate & useful summaries. Several extractive methods have been applied till now & work is still going on. These methods use Machine Learning, Deep Learning & Optimisation techniques. Through this paper, we've represented an engrossed research of various written summarisation of extractive methods which are currently in working. At last this paper ends with the discussion of future areas where there is more need to improve & which areas are to make better.*

**Key words** *Text Summarisation , Deep Learning , Machine Learning , Natural dialect Processing , Artificial intelligence*

## 1. INTRODUCTION

Automatic Text summarisation creates a condensed form of information which contains only precise information . It just takes important Words or Lines from the whole text & it should be smaller than whole report. It was first started in 1950s & since then there is a vast progress in this field of research. Automatic text summarisation is a tough task & it is necessary to h&le the sentence ordering, redundancy issues etc to make the resultant synopsis short , precise & meaningful. Due to increase in large flow of data online, need of text summarisation has increased a lot in recent years. There are a huge number of reports online & it is obstructive to find predominant information. There is a high chance of redundancy in the text due to high volume of texts on a variety of topics. Text summarisation is must so that we can skip large amount of texts reading & instead study only the predominant part & can save time & resources. The four key goals of this method are information coverage, information importance, information redundancy & text cohesion. In extractive summarisation, some assigned scores to lines in reports & then highly scores of lines are selected to produce synopsis. Synopsis's length depends on the rate of compression. In abstractive summarisation, abstract synopsis is produced in which words or phrases are discrete from the ones in the original report. It uses Natural dialect processing extensively. It is more different than summarisation of extraction.

## 2. RECENT AUTOMATIC TEXT SUMMARISATION EXTRACTIVE METHODS

There are many ways through which we can do Extractive text summarisation. Following are some ways-:

### 2.1. Instructed summariser & latent semantic analysis for summarisation of text

MCBA + GA is worked with the corpus of a special domain & also for online use. When quality of synopsis is the main aim then LSA plus TRM method is preferred. A synopsis of related lines is produced through this method semantically. The methods are dialect-independent. Most of the times, Coherence & Cohesion are missing in the Synopsis. Score function's feature weights are produced by GA does not always result in great performance outcomes for test corpus. In LSA plus TRM method, getting best dimension minimization ratio & LSA effects explanation are comparatively tough. It took more time to calculate SVD. This method employs LSA to get a document's semantic matrix and maintains a connection map for semantic text by using a sentence's semantic representation. It performs better than keyword based approach in single document.

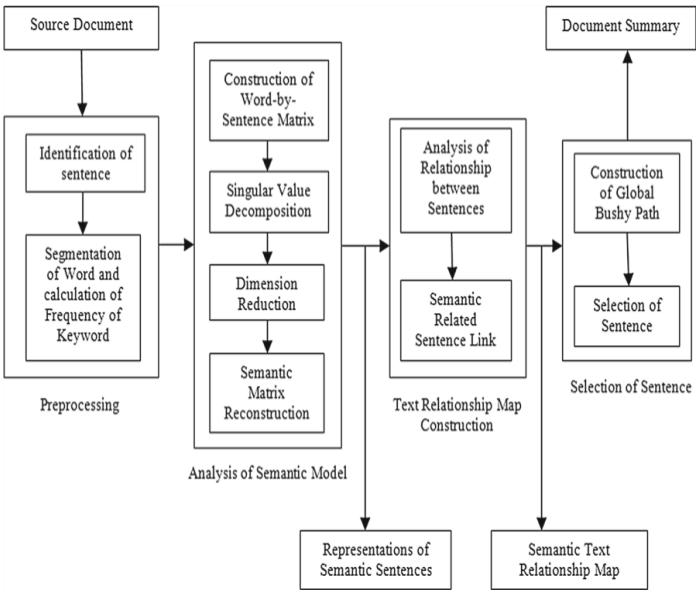


Fig- 1. Complete procedure of LSA plus TRM approach

### 2.2. Data extraction via line based abstraction approach

This method focuses on textual continuity like coherence in text & the lexical cohesion for generating synopsis . Larger level reports can be very easily acknowledged by this given method & it shows better human perception & its also important in enhancing retrieval performances. In this given method, only the coherence in causality is considered while causality, spatiality & temporality are also used for showing episodes of behaviour in a discourse.

### 2.3. Extraction of sentence using contextual knowledge & summarisation of text based on Statics

This method is used when reports of discrete dialects needs to be Condensed as the method is dialectindependent. So the greatest strength of the approach is that its dialect -independent. Memory capacity & much processor is not required for extracting significant lines. Also reports without the title can be condensed with the method.

### 2.4. Email summarisation using cohesion in conversion & subjective opinions

This method is used to summarise email taking into fact conversation of structure in the emails & the subjective phases and word they carry . Emails however can be Condensed which aids the users have a fast view through the past dialogues through emails in a short time. By connecting subjective views into approach, approach efficacy is improved way better.

### 2.5. Text summarisation via complex network method

This method can be needed when it is used to summarise reports of discrete dialects & a large no. of the linguistic assets aren't accessible.So this method is dialect-independent. The extract is produced by employing deep linguistic knowledge. This complex network ideas give discrete complementary views of the network.

### 2.6. Creation of automation general report synopsis using non -ve matrix factorisation

The method is used to do general text summarisation when training synopsis aren't accessible to train the approach & when semantic properties required to be found systematically. The method take out more useful lines & the subtopics available in the report can be located systematically. Any training data isn't required as it's a method of unsupervision.

### 2.7. Text of automatic summarisation using GA, MR, FFNN, PNM & GMN based Prototypes

This method is used if required to employ the statistical ways for summarisation of text. When we require to use the trainable summariser & we've training synopsis in the special dialect but we desire to summarise reports in different dialect. By the method, Prototypes can be instructed on the certain data dialect & can be tested on another dialect's data. All properties used are dialect independent except +ve & negative key word.

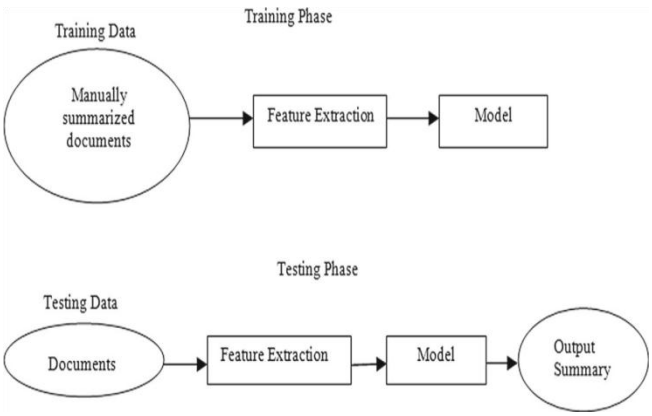


Fig – 2. The proposed Automatic Summarisation Model

### 2.8. Query based summarisation of many reports by using regression prototypes

This method is used when its used to do ML based summarisation of various reports by using query based properties. Its also used when its required to make data sets pseudo training from the human synopsis to guess

scoring of lines. Regression Prototypes give good outcomes than identification & finding to rank approaches. A good mapping function is produced between characteristic vector & the sentence significance score. MMR method is used to remove redundancy from synopsis.

### 2.9. Maximum coverage & minimum redundancy in text summarisation

This method can easily be applied when there's a requirement to employ an optimisation method for resolving problem of summarisation. When there is not enough training synopsis & when goal is to take significant matter of report with negligible redundancy. The method is a generic text summarisation method of un supervision so it doesn't need training synopsis. It can produce the synopsis consisting of predominant matter with negligible redundancy.

### 2.11. Sentence assessment scoring methods for extractive text summarisation

This method requires to be mentioned to have an understanding into literature in the last 10 years for obtaining association with several methods of summarisation of texts & to understand how the qualitative & quantitative evaluation is done on 15 algorithms of line marking. Due to this method we get friendly with qualitative & quantitative evaluation of line algorithms scores & some helpful directions are issued for increasing line marking results.

### 2.12. Correlations exploration among many terms using a graph based summariser, GRAPHSUM

This method is used to find connections between some terms available in the report by using association rules. This approach doesn't rely on semantic-based prototypes advancement & do a negligible numbers of dialect dependent tasks so its versatile & movable & can be employed with reports belonging to the context of discrete application.

### 2.13 Integrating numerous level of dialect analysis for handling redundancy in summarisation of text

This method is used to observe redundancy through 3 different levels of dialect analysis like lexical, syntactic & semantic. The method wastes information redundancy & produces a synopsis from information non redundancy & further the redundant information helps in finding the important lines.

### 2.14. Emerging optimisation algorithm for summarising various reports

This method is used to do optimisation-based general report summarisation & to attain highest coverage of matter with less redundant data. The method decreases redundancy in synopsis , chooses significant lines from the report & includes predominant matter of the original report.

### 2.15. Summarisation of multiple reports through a hybrid machine learning Prototype

This method is used to produce abstracts of discrete dialects & it can further be employed when there's accessibility of the training abstracts to try the trainable summariser evolved from the union of ML algos. All the given text properties employed in this method are dialect independent. This criteria of property extraction is employed in the method gives a chance to use the numbers of forms on the base of dialect & the text type.

### 2.16. Sentence level improving grouping using technique ranking based for theme based summarisation

This method is used to carry out theme based on the multi-report summarisation s.t. lines are grouped on the themes basis. The method produces large standard line groups on theme basis & an altered MMR alike method is needed to check redundancy in the multi-report summarisation.

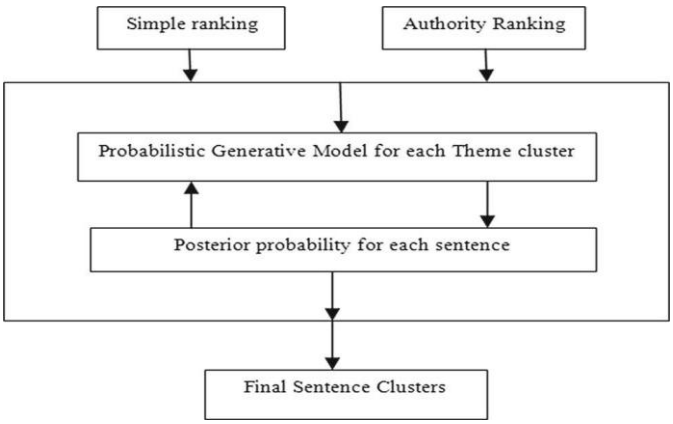


Fig – 3. Ranking based Line grouping Approach

### 2.17. Sparse optimisation based compressive report summarisation

This method is used to attain compressive summarisation that gives better outcome as compared to the extraction original approaches based on the information construction. This given approach is unsupervised completely so it needs no data training.

### 2.18. Sub modular Mixtures based summarisation of many-report topic hierarchy

This method is used to produce Wikipedia disambiguation pages for the article set based on discrete topics but with same headings. The method can summarise big group of label into compact , achievable & more relevant label sets.

### 2.19. Summarising many reports using approach combination

The given method is used as it can very much increase the matter quality by merging synopsis produced from discrete approaches. The given method of merging abstracts from discrete approaches aids to increase the matter quality. Also, this method can merge abstracts produced by any approaches.

### 2.20. Phrase-based compressive cross-dialect summarisation

This method is used to aid users get the key ideas of these reports written in the specific dialect that they ain't friendly with. Further despite of not using any syntactic data, given approach keeps good fluency and grammaticality.

### 2.21. Re-assessment of automatic summarisation using 192 and BLEU variants of ROUGE

These assessment of summarisation approach is performed to get which kind of summarisation metric surpasses others. These assessment result corrected incorrect assumption of readers as these result display that the higher forms of summarisation metric is discrete from the ones highly recommended earlier.

## 3. CONCLUSION

Text summarisation is an amazing research field & it has many uses. This paper's aim is to aware budding scientists about past development in the text summarisation field & the future's possibilities. Through this paper we provide a fine report for new scientists for getting a deep insight about text summarisation issues & development. Recent automatic text summarisation extractive methods are talked in this paper. All the pros & cons along with the basic knowledge & basic technique for all the methods will help the readers a lot to learn about significance of each technique. Through this paper, researchers will get future directions that will aid them in enhancing synopsis production approaches through which progress is done in this field continuously.

## 4. REFERENCES

[1] E. Reddy B (2011) Information extraction by an abstractive text summarisation for an Indian regional dialect. In: Natural dialect processing & knowledge engineering (NLP-KE), 2011 7th internationalAliguliyev RM (2009) A new sentence similarity measure & sentence based extractive technique for automatic text summarisation. Expert Syst Appl 36(4):7764–7772

[2] A RM, Aliguliyev RM, Isazade NR (2013) Multiple reports summarisation based on evolutionary optimisation algorithm. Expert Syst Appl 40:1675–1689 .2012.09.014

[3] A Hajirahimova MS, Mehdiyev CA (2011) MCMR: maximum coverage & minimum redundant text summarisation Prototype. Expert Syst Appl 38:14514–14522. doi:10.1016/j.eswa. 2011.05.033

[4] M, Martins AF (2013) Fast & robust compressive summarisation with dual decomposition & multi-task learning. In: ACL (1), pp 196–206

[5] E, Gonzalo J, Penas A, Verdejo F (2005) QARLA: a framework for the assessment of text summarisation approachs. In: ACL '05: proceedings of the 43rd annual meeting on association for computational linguistics, pp 280–289

[6] Amati G (2003) Probability Prototypes for information retrieval based on divergence from r&omness. University of Glasgow.

[7] M, Atkinson M, Steinberger J et al. (2010) NewsGist: a multilingual statistical news summariser.Lecture notes in computer science (including including subseries lecture notes in artificial intelligence & lecture notes in bioinformatics) 6323 LNAI, pp 591–594. doi:10.1007/978-3-642-15939-8_40

[8] R, Foster J, Roturier J (2014) Semantic role labelling with minimal resources: experiments with french. In: Lexical & computational semantics (*SEM 2014), p 87

[9] Kallimani JS, Srinivasa KG, conference on IEEE, pp 319–322

[10] Kedzie C, McKeown K, Diaz F (2015) Predicting salient updates for disaster summarisation. In: Proceedings of the 53rd annual meeting of the ACL & the 7th international conference on natural dialect processing. pp 1608– 1617

[11] Khan A, Salim N, Jaya Kumar Y (2015) A framework for multi-report abstractive summarisation based on semantic role labelling. Appl Soft Comput 30:737–747. doi:10.1016/j.asoc.2015.01.070

[12] Kikuchi Y, Hirao T, Takamura H, Okumura M, Nagata M (2014) Single report summarisation based on nested tree structure. In: Proceedings of the 52nd annual meeting of the association for computational linguistics, vol 2, pp 315–320

[13] Suman Mann, Deepa Gupta, Yukti Arora, Shivanka Priyanka Chugh, Akash Gupta, Smart Hospitals Using Artificial Intelligence and Internet of Things for COVID-19 Pandemic,chapter in Smart Healthcare Monitoring Using IoT with 5G, 2021

[14] Meena Siwach, Suman Mann, Anomaly detection for web log data : A Survey,IEEE Conference,2022.

[15] Meena Siwach, Suman Mann, Anomaly detection for web log data analysis using improved PCA Technique, Journal of information and optimization Science. 131-141, DOI: 10.1080/02522667.2022.2037283

[16] D. Gupta, S. K. Jha and S. Mann Maharaja Surajmal, " Internet Crimes-It's Analysis and Prevention Approaches," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2021, pp. 1-4, doi: 10.1109/ICRITO51393.2021.9596396