

A Survey on Methods of Text Summarization

Dr. Ashish Kumar Chakraverti¹, Ankit Kumar Pandey², Atul Dhadse³, Shubham Choudhary⁴, Shivam Kumar Pandey⁵

¹Associate Professor, ^{2,3,4,5}U.G. Students

Department of Computer Science and Engineering, School of Engineering and Technology,
Sharda University Greater Noida, Uttar Pradesh, India

Abstract:- Text summarization has become a reduced form that preserves its data content and general meaning. Thanks to the abundance of data we provide and thanks to the advancement of Internet Technologies, text summarization has become an important tool for deciphering text data. Text abstraction techniques can be divided into removable and invisible abstracts. An abridged summarizing method involves selecting sentences that have a high point in the message according to word structures and sentences and pausing them to produce a summary. The value of sentences is chosen according to the main factual and etymological points in the sentences. Unintelligible summaries usually find the basic ideas in a particular record and then convey those ideas in plain language. In this paper, it provides a similar investigation of the different methods of summarizing the texts.

Keywords:- Text summarization, natural language processing, NLP, statistical approach.

I. INTRODUCTION

An outline can be characterized as a text conveyed from no less than one text, which moves significant information to the first text (s), and doesn't surpass part of the first text (s) and when in doubt. not really. The text here is uninhibitedly utilized and can allude to discourse, intuitive media chronicles, hypertext, and the sky is the limit from there. The primary motivation behind the synopsis is to present the principle thoughts in a restricted space. Experiencing the same thing where each sentence in the chronicle of a message has equivalent worth, making summary wouldn't work similarly, as any decline in record size would send a comparing decline in its utilization. Luckily, the substance of the information in the record comes from the blast, and one can hence distinguish the most valuable and immaterial parts. Recognizing instructive parts that hurt others is a fundamental test in rundown [1]. Mechanized text outlines are an element of incorporated conveyance, also, a standard rundown while safeguarding the center information content and vital. All the more as of late, different techniques for programmed synopsis have been created and generally utilized in better places. Programmed deliberation implies that the summed up yield is given when the data is utilized. Albeit the mechanized text synopsis test started during the 1950s at IBM Research Laboratories [2], lately the field of Summary Text Summary has met with amazing advancement because of the web. It is truly difficult to sum up enormous text reports as there is a tremendous measure of information on the web. On the other hand, the web is a rich library that gives a bigger amount of information than is mentioned. Along these lines, it is vital to check out a significant document with countless

available records. The motivation behind text synopsis is to sum up the source text into a decreased inconsistency of information content observing and general importance.

Text summing up methods can be coordinated into a removable and convincing outline. Serious Summary is a clever style that conveys the setting of a message that requires a huge machine for general language the executives, which incorporates sentence designs and word references to separate age. Remove Summary is a system for deciding the units of critical messages (sentences overall) by taking a gander at the worth of the message unit concerning word reference and estimation or by contrasting sentence models [3]. Dynamic techniques give an intricate outline and adjust well to high tension levels while extraction strategies are easy to acclimate to huge sources albeit ensuing edited compositions might be upset. Selective truncations are gotten by removing portions of a significant message like sentences or sections from the message, because of a careful examination of the more disengaged or accumulated parts [4].

Programmed deliberation of text is extremely challenging, because when we as a group sum up a piece of text, generally speaking, we read it completely to advance our understanding, and afterward make a system that mirrors its principle issues. Since PCs require individual information and language abilities, it makes programmed text wrapping an extremely challenging and good for nothing task. There are two particular gatherings of text-arranged contractions and Teaching. The reference rundown provides the client with an essential thought of the message. The length of this outline is roughly 5% of the text given. The informative synopsis outline gives a short outline of the fundamental text. The length of the instructive rundown is roughly 20% of the given text [5]. Also, outline methodologies can be gathered from the source which can be single or different record synopses. In the synopsis of a solitary report, just a single record is incorporated to deliver a rundown. It is a direct and unique method for summing up. Two removable and limitless outline procedures can be utilized to sum up one document. The system of many files is similarly a significant piece of the outline. More than one information source is utilized to produce a synopsis. Numerous electronic clustering designs, for example, stories have been redone since the nullification of many records. Regardless, the charging of many cycles for recording records is substantially more perplexing and complex than single-text methods. The genuine point isn't simply to eliminate the conspicuous redundancy and to bring up the suitable text to sum up yet as well as giving a bizarre and persuading end that the last rundown should be astute and complete in itself.

It was along these lines hard for them to think about every one of the reports and to sum up [6].

II. EXTRACTIVE SUMMARIZATION

As shown earlier, the extractive summarization produce summaries by selecting a small set of sentences in the original text. These summaries contain key information sentences. The information can be a single report or separate archives .To easily understand how summary summaries work, we present three free functions performed by all summaries [7]: 1) We create a temporary display of input text that connects important parts of the text. 2) Extract sentences with a visual representation. 3) Choose a summary with different sentences.

A. INTERMEDIATE REPRESENTATION

Each summary structure creates an intermediate representation of the text which is a summary and is considered significant in view of this representation. There are two types of approaches when it comes to presentation: subject representation and reference representation. Theme presentation methods transform the message into a balanced presentation and clarify the topic in the text. Summary methods based on the topic differ in terms of their complexity and model of expression, and are divided into methods that are repeated for repetition, approximate theme name, invalid semantic testing and Bayesian theme models [7]. We describe the methods of presenting topics in related contexts. Indexes show each sentence as a list of key points of importance such as sentence length, repository, specific expressions, and more.

B. SENTENCE SCORE

At the point where a temporary expression is produced, we give important points in each sentence. The presentation of the point is approaching, sentence points directing how the sentence makes sense to the most important topics of the text. In most index strategies, points are analyzed by combining evidence from various markers. AI techniques are often used to view brand loads.

C. SUMMARY SENTENCES SELECTION

Finally, the summary framework selects the most important k sentences to convey the summary. A few methods use unsatisfactory math to select important sentences and a few methods can change the sentence structure into a problem of simplicity where a variety of sentences are chosen, considering the extent to which they should grow in greater importance and coherence of sentences. Moreover, reduce obvious repetition. There are different elements that ought to be thought about while choosing the key sentences. For example, setting a summary may be helpful in determining its value. The type of record (eg news article, email, sensible paper) is another variation that may influence the choice of sentences.

III. TOPIC REPRESENTATION APPROACHES

In this section we describe some of the most used topic representation approaches.

A. TOPIC WORDS

The topic words method is perhaps the most widely recognized method for addressing the marks of assumption for seeing words that mirror the title of a data record. [8] It was perhaps the earliest task to execute the procedure utilizing tedious impediments to track down metaphorical words in the record and to address the subject of the report. One more formative variety of Luhn's hypothesis was the presentation of [9] the utilization of a halfway rationale test to recognize spellbinding words in a short composing known as the "Title Signature". The utilization of title images as a portrayal of focuses has been a surprising achievement and broadened the exactness of the rundowns of many reports in the media [10]. For more data on the incomplete logging test, see [7].

There are two methods for enlisting the worth of a sentence, for example, the capacity of the quantity of focus it contains, or how much head marks in a sentence. Both accentuation marks are connected to the statement of a similar point, notwithstanding, they can frame various focuses in sentences. The fundamental procedure might be to bring down the focus to longer sentences, as it has more words. The accompanying technique estimates the thickness of the title words.

B. FREQUENCY-DRIVEN APPROACHES

As we reduce the number of words in the title presentations, we can think of the corresponding loads (0 or 1) or the actual (infinite) loads and decide which words are most relevant to the theme. The two most common methods of this classification are: name opportunities and TF IDF (Term Frequency Inverse Document Frequency).

a) WORD PROBABILITY

A simple method for including word reiteration as images of significance is word open doors. Amazing open doors for a word that isn't completely settled forever like the quantity of word events, $f(w)$, isolated by the quantity of the multitude of names of data (either individual report or separate records):

$$P(w) = f(w)/N \quad (1)$$

Vanderwende et al. [75] proposed a SumBasic structure that utilizes just the word that can be utilized to address the meaning of a sentence. In each sentence, S_j , in data, allocate a weight equivalent to the ordinary opportunity of words in a sentence:

$$g(S_j) = \sum w_i \in S_j, P(W_i) / \{w_i | w_i \in S_j\} \quad (2)$$

where $g(S)$ is the trouble of the sentence S_j In the following section, pick the best sentence-observing sentence containing a word that might be the most elevated. This movement guarantees that the main name of the likelihood, which alludes to the subject of the report around then, is recollected by

summation. Then for each word in the chose sentence, the weight is reestablished:

$$P_{new}(w_i) = P_{old}(W_i) P_{old}(W_i) \quad (3)$$

This amendment of the word weight shows that the likelihood that the word that showed up in the contraction is lower than the word that happens once. The previously mentioned choices will change until a rundown of the proper length is reached. The selection of sentences utilized by SumBasic relies upon the technique utilized. Yih et al. [79] utilized coordinating strategy (as a sentence decision) to expand the event of worldwide watchwords in a synopsis is one instance of utilizing a formative methodology.

b) TFIDF

Since word process, processes rely upon a rundown of default words so you don't consider it in the title and as concluding which words to place in the stop list is erroneous, there is a requirement for further developed methodologies. Quite possibly the most developed and successful assimilation procedure is to give weight to words TFIDF (Frequency Contrasting Terms Frequency). This rating system surveys the worth of words and recognizes notable words (which ought to be disposed of in the archive) by giving low loads on words from different reports. The heaviness of each word in archive d is estimated as follows:

$$q(w) = fd(w) + \log(|D|/fd(w)) \quad (4)$$

where $fd(w)$ is a term reiteration of the word w in history d , $FD(w)$ number of files containing the word w and $|D|$ number of vaults in collection D . For additional data on TFIDF and other word handling frameworks, see. TFIDF loads rush to process and likewise are great strides for deciding the worth of sentences, so most existing moderators [12,13,14] utilize this technique (or re-utilize some type of it).

c) LATENT SEMANTIC ANALYSIS

Latent Semantic Analysis (LSA), introduced by, a pointless system to reject the declaration of literary semantics by checking perceived words out. Gong and Liu [15] at first proposed a system that utilized the LSA to choose a particular sentence for a solitary, multi-record outline in the media. The LSA system at first fosters a sentence (n by m grid), where every segment is contrasted and something from the data (n words) and each part is connected with a sentence (m sentences). Each piece of a_{ij} lettuce is the heaviness of the word I in the sentence j . Heaps of words are considered a TF IDF procedure and on the off chance that a sentence doesn't have a word the trouble of that word in the sentence is zero. Then, at that point, Special worth rot (SVD) is applied to the network and converts cross section An into three aspects: $A = U \Sigma V$

Matrix U ($n \times m$) speaks with an organization point term network with loads these words. Structure Σ is a skewed framework ($m \times m$) where each line is connected with the intricacy of the topic I . The VT structure is a draft sentence point. The graph $D = \Sigma VT$ shows how far a sentence focuses to a point, and accordingly, the d_{ij} demonstrates the intricacy of the subject I in the sentence j .

Gong and Liu's procedure was to choose one sentence for each point, along these lines, considering the length of the abstract corresponding to the sentences, they had various subjects. This interaction is confounded by the way that the point might require a various sentence to move its information. Subsequently, chose arrangements were proposed to deal with the introduction of LSA-based investigation procedures. One advancement was to utilize the trouble, everything being equal, to finish up the standard of the edge that ought to cover the topic, giving adaptability to having a unique number of sentences. One more technique is displayed in [16]. Steinberger et al. [16] presented a LSA-based approach that accomplishes a completely favored show instead of a remarkable capacity. They comprehend that sentences about the absolute most significant subjects are exceptionally problematic in summing up, after which, while mentioning those sentences they call attention to the intricacy of the sentence as follows:

Let g be the "weight" function, then

$$g(s_i) = \sqrt{\sum_{j=1}^m d_{ij}^2}$$

Fig. 1: Formula of the "weight" function

d) BAYESIAN TOPIC MODELS

Countless current multi-story methodologies have two downsides [19]: 1) They believe sentences to be free of one another, so focuses remembered for records are overlooked. 2) Sentence focuses dissected in most existing techniques frequently don't have an unmistakable conceivable comprehension, and countless sentence focuses are resolved utilizing heuristics.

Models of Bayesia articles are potential models that present and address record titles. They are exceptionally strong and appealing, for the explanation that they discuss information (for instance courses) that are lost in various ways. Their benefit in introducing and tending to points completely enables the improvement of succinct structures that can decide similitudes and examinations between records to be utilized in synopsis [20].

Not with standing title improvement and file show, topic models frequently utilize an

unquestionable rating to get a sentence called Kullbak-Liebler (KL). KL is important for the distinction (contrast) between the two potential financial plans P and Q [21]

KL variety is an intriguing system for tracking down sentences about synopses, as it shows how great normal designs, for example, data records. It shows how the worth of words changes in the system comparable to include, for instance the KL changeability of a legitimate summation and data will be low.

Conceivable theme models have as of late arisen in different fields [22 - 28]. The Dormant Dirichlet area (LDA) model is the best pointless class order of topic (s) for an assortment of records. A total LDA overview can be found in [34,35], but the speculative rule is that reports are viewed as an intriguing blend of immaculate subjects, in which the entire setting is a method for conveying words.

The LDA has been generally used to sum up many reports from later. For instance, Daume et al. Proposed BayeSum, a Bayesia rundown model of the inquiry centered structure. Wang et al. [19] presented a Bayesia sentence-based subject model of the most recent variant that utilized both file and term contracts. Their structure accomplished critical execution and beat numerous other outline procedures. Celikyilmaz et al. [30] show a multi-faceted synopsis as an anticipated issue while considering a two-stage model. To start with, they propose a persistent topic model to get themed plans, everything is equivalent. Then, at that point, they surveyed the comparability of their rivals' sentences with the shortenings given to the singular utilizing the genuine volume of the tree-determined sentence. In the second step they utilize these ideas and train the reiteration model that compares to the word reference and fundamental credits of the sentences, and they utilize the model to observe new record sentences (unclear reports) to shape word records.

IV. ABSTRACTIVE TEXT SUMMARIZATION

The theoretical rundown [31] includes grasping the essential ideas and significant information of the framework message and, simultaneously, imparting that information in a brief and clear arrangement. Imperceptible summation procedures can likewise be isolated into two organized based classifications and semantic-based methodologies. Standard based strategies decide key information through files utilizing designs, yield rules and different plans, for instance, tree, reasoning and that's just the beginning.

A. STRUCTURE BASED ABSTRACTIVE TEXT SUMMARIZATION

a) RULE BASED METHODS

The normalized approach [32] comprises of three stages: -

- At first, the reports to be considered are classified by their classifications. Classifications can emerge out of various sources. So the main undertaking is to channel these. Coming up next is an outline of the inquiries in view of these classes. For instance, between various classes, for example, assaults, calamities, prosperity and so forth, taking a gander at the beating class a couple of inquiries can be changed as follows: - What was the arrangement? , when accomplished it work ?, who was impacted ?, What were the outcomes?
- In light of these inquiries, rules are created. Here a couple of action words and items with relative impacts are not completely settled and their construction is precisely recognized.
- The set choice module chooses the best opponent among these.
- Configuration plans for a really long time and utilized for quite a long time of outline sentences.

b) ONTOLOGY METHOD

In this methodology, neighborhood cosmology of information times is reflected in nearby specialists. The subsequent stage is the document the board stage. Corporate names are created in this part [33]. Catchphrases are recognized by arrangement in view of data age. Support degrees connected with different times of nearby way of thinking. Cooperation degree is delivered by fleecy extraction. The impediments to this approach are that they are baffled by the way that neighborhood theory must be seen by nearby specialists. The benefit of this approach is to oversee sketchy data.

c) TREE BASED METHOD

In this approach, preliminary processing is performed of the same sentences using shallow analysis [36]. Then we put those sentences in the predicate-argument structure. Different algorithms can be used to select a common phrase in sentences such as the Theme algorithm. A phrase that conveys the same meaning is selected and we also add some information to it and we will arrange it in a certain order. Finally, the FUF / SURGE language generator can be used to create new abstract sentences by combining and editing a selected common sentence. The use of language generators increases language fluency and also reduces system errors. This feature is a great strength of this approach. The main problem with this method is that the context of the sentences does not include when choosing a common phrase and is an integral part of sentences even if they are not part of a normal phrase.

B. SEMANTIC BASED

Abstractive Summarization Multimodal semantic models capture ideas and arrange interactions between these ideas [39]. These selected ideas are passed on as sentences. This model approves archives and photo records

Multimodal has three stages -

a) SEMANTIC MODAL IDEAS are just words that talk about important data. Ideas are created using the presentation of information in the light of things. Hubs that talk about ideas and connections between these ideas deal with the interactions between them.

b) RATED CONCEPTS Thoughts are tried utilizing a data thickness (ID) system. This organization is utilized to test the similarity of thoughts. The zenith of properties is just important for the credits finished in the semantic model for the absolute number of semantic credits. This gives convincing sentences extra information.

One thought is related with various thoughts and these connections are counted. Checking this computation assists us with acknowledging how significant this thought is.

c) SENTENCE GENERATION When the ideas are appraised utilizing ID lattice the following stage is to produce sentences utilizing parsing procedures.

C. INFORMATION ITEM BASED METHOD

In this system, as opposed to making hypothetical from sentences of the data archive, it is made from hypothetical depiction of the data record. The hypothetical depiction is just an information thing which is the humblest part of information in a text. The framework [40] used in his technique was proposed concerning the Text Analysis Conference (TAC) 2010 for multi-file summary of data. The modules of this construction are: Information thing recuperation, sentence age, sentence decision and abstract age. In the Information Item (INIT) recuperation stage, subject-activity word object fundamentally increments are formed by semantic examination of text got done with the assistance of parser. During semantic assessment, activity word's subject and article are removed. In sentence age stage, the sentences are made using a language generator. In the accompanying stage for instance sentence decision stage, positioning of each sentence is done in view of the typical record repeat (DF) score. At last in the summary age stage, significantly situated sentences are coordinated and reasonable is made with authentic Arranging.

From this strategy, a short, sane, information rich and less monotonous summary can be molded. Ignoring such countless advantages, this method has in like manner various obstructions. While making syntactic and critical sentences, various huge information things get excused. In view of which, the phonetic idea of the resultant diagram gets diminished.

D. SEMANTIC GRAPH BASED METHODS

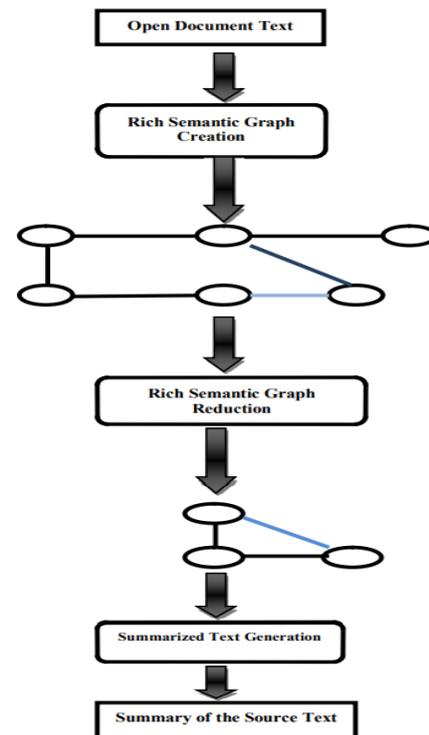


Fig. 2: Semantic graph based method work flow

The essential objective of this technique is creating a synopsis by making a semantic graph called rich semantic outline (RSG) [38]. As shown in Figure The semantic chart approach contains three phases:

- The essential stage tends to enter reports using rich semantic outlines (RSG). In RSG, the activity words and things of the information report are addressed as graph centers and the edges contrast with semantic and topological relations between them.
- The ensuing stage lessens the main graph to a more diminished chart using heuristic standards
- The third stage makes an abstractive outline. The potential gain of this technique is that it conveys less redundant likewise, semantically right sentences. The injury of this methodology is that it is confined to a single document and not different records.

V. CONCLUSION

In text rundown, the best test is to recuperate critical information from given basic sources including site pages, any record, and informational index. An effective synopsis ought to be conveyed by text outline strategies using less time and less obvious dullness. The two procedures for synopsis produce promising results. They basically achieve the results surprisingly. From the above pondered systems, the advantages and cutoff points of each methodology is created underneath. The outline created by the Rule based methodology is of high information thickness anyway it is incredibly long work since all of the rules and models are made genuinely.

In the procedure for Ontology, treatment of uncertain data is possible which is ridiculous in clear space mysticism. Issue with this method is that principle region experts would be capable describe the transcendentalism of the space which is monotonous. In the Tree based procedure, the idea of framework gets enhanced record of the use of language generators. Figuratively speaking, the issue with this method is that the essential setting of the sentences gets excused while getting the assembly of expressions. The Multimodal semantic model procedure creates a hypothetical synopsis wherein it consolidates abstract data as well as graphical data and accordingly, gives a splendid result. Issue with, This procedure is that evaluation is to be done genuinely.

In the Data thing based method, the decision of accommodating information is done. In light of picked information, the sentences and diagrams are made. This system gives somewhat, conscious and information rich synopsis. Issue with this method is that sporadically important information gets excused while the improvement of huge and grammatically right sentences which diminishes the etymological idea of summary. The Semantic outline method, Sentences formed are less tedious as well as semantically right. Anyway, this procedure is limited to only a solitary chronicle.

Extractive text rundown moreover makes extraordinary summaries. It produces remarkable results when the data text isn't incredibly extended. The issue with this method for synopsis age is that it produces texts that are very machine-like in nature. It at times makes grammatically wrong sentences. Despite the fact that estimations considering Extractive text outline are not very perplexing in nature and are easy to complete. Anyway the strategy for programmed outline is an old test, the experts these days are getting more skewed towards abstractive synopsis techniques rather than extractive rundown techniques. This is in light of the fact that, abstractive rundown procedures produce more clear, less overabundance and information rich outline. Making calculated using abstractive rundown procedures is an inconvenient endeavor since it requires more semantic and etymological examination.

REFERENCES

- [1.] Dragomir R Radev, Eduard Hovy, and Kathleen McKeown. 2002. Introduction to the special issue on summarization. *Computational linguistics* 28, 4 (2002), 399–408.
- [2.] H. P. Luhn, "The automatic creation of literature abstracts", *IBM Journal of Research and Development*, vol. 2, pp. 159-165, 1958
- [3.] U. Hahn, & I. Mani, "The challenges of automatic summarization" *IEEE-Computer*, 33(11), 29–36, 2000.
- [4.] V. Gupta, G.S. Lehal, "A survey of Text Summarization Extractive Techniques", *Journal of Emerging Technologies in Web Intelligence*, Vol. 2, No.3, August 2010.
- [5.] S.A. Babar, P. D. Patil, "Improving Performance of Text Summarization", *International Conference on Information and Communication Technologies, ICICT*, 2014.
- [6.] Sherry, P. Bhatia, "A Survey to Automatic Summarization Techniques", *International Journal of Engineering Research and General Science* Volume 3, Issue 5, September-October, 2015.
- [7.] Ani Nenkova and Kathleen McKeown. 2012. A survey of text summarization techniques. In *Mining Text Data*. Springer, 43–76.
- [8.] Hans Peter Luhn. 1958. The automatic creation of literature abstracts. *IBM Journal of research and development* 2, 2 (1958), 159–165.
- [9.] Ted Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational linguistics* 19, 1 (1993), 61–74.
- [10.] Sanda Harabagiu and Finley Lacatusu. 2005. Topic themes for multi-document summarization. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 202–209.
- [11.] Wen-tau Yih, Joshua Goodman, Lucy Vanderwende, and Hisami Suzuki. 2007. Multi-Document Summarization by Maximizing Informative Content-Words. In *IJCAI*, Vol. 2007. 20th.
- [12.] Rasim M Alguliev, Ramiz M Aliguliyev, Makrufa S Hajirahimova, and Chingiz A Mehdiyev. 2011. MCMR: Maximum coverage and minimum redundant text summarization model. *Expert Systems with Applications* 38, 12 (2011), 14514–14522.
- [13.] Rasim M Alguliev, Ramiz M Aliguliyev, and Nijat R Isazade. 2013. Multiple documents summarization based on evolutionary optimization algorithm. *Expert Systems with Applications* 40, 5 (2013), 1675–1689.
- [14.] Günes Erkan and Dragomir R Radev. 2004. LexRank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Intell. Res.(JAIR)* 22, 1 (2004), 457–479.
- [15.] Yihong Gong and Xin Liu. 2001. Generic text summarization using relevance measure and latent semantic analysis. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 19–25.
- [16.] Josef Steinberger, Massimo Poesio, Mijail A Kabadjov, and Karel Ježek. 2007. Two uses of anaphora resolution in summarization. *Information Processing & Management* 43, 6 (2007), 1663–1680.
- [17.] Ben Hachey, Gabriel Murray, and David Reitter. 2006. Dimensionality reduction aids term co-occurrence based multi-document summarization. In *Proceedings of the workshop on task-focused summarization and question answering*. Association for Computational Linguistics, 1–7.
- [18.] Dingding Wang, Shenghuo Zhu, Tao Li, and Yihong Gong. 2009. Multi-document summarization using sentence-based topic models. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*. Association for Computational Linguistics, 297–300.
- [19.] Makbule Gulcin Ozsoy, Ilyas Cicekli, and Ferda Nur Alpaslan. 2010. Text summarization of Turkish texts using latent semantic analysis. In *Proceedings of the 23rd international conference on computational linguistics*. Association for Computational Linguistics, 869–876.

- [20.] Inderjeet Mani and Eric Bloedorn. 1999. Summarizing similarities and differences among related documents. *Information Retrieval* 1, 1-2 (1999), 35–67.
- [21.] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* (1951), 79–86.
- [22.] Mehdi Allahyari and Krys Kochut. 2015. Automatic topic labeling using ontology-based topic models. In *Machine Learning and Applications (ICMLA)*, 2015 IEEE 14th International Conference on. IEEE, 259–264.
- [23.] Mehdi Allahyari and Krys Kochut. 2016. Discovering Coherent Topics with Entity Topic Models. In *Web Intelligence (WI)*, 2016 IEEE/WIC/ACM International Conference on. IEEE, 26–33.
- [24.] Mehdi Allahyari and Krys Kochut. 2016. Semantic Context-Aware Recommendation via Topic Models Leveraging Linked Open Data. In *the International Conference on Web Information Systems Engineering*. Springer, 263–277.
- [25.] Mehdi Allahyari and Krys Kochut. 2016. Semantic Tagging Using Topic Models Exploiting Wikipedia Category Network. In *Semantic Computing (ICSC)*, 2016 IEEE Tenth International Conference on. IEEE, 63–70.
- [26.] Freddy Chong Tat Chua and Sitaram Asur. 2013. Automatic Summarization of Events from Social Media.. In *ICWSM*.
- [27.] John Hannon, Kevin McCarthy, James Lynch, and Barry Smyth. 2011. Personalized and automatic social summarization of events in video. In *Proceedings of the 16th international conference on Intelligent user interfaces*. ACM, 335–338.
- [28.] Liu Na, Li Ming-xia, Lu Ying, Tang Xiao-jun, Wang Hai-wen, and Xiao Peng. 2014. Mixture of topic model for multi-document summarization. In *Control and Decision Conference (2014 CCDC)*, The 26th Chinese. IEEE, 5168–5172.
- [29.] Zhaochun Ren, Shangsong Liang, Edgar Meij, and Maarten de Rijke. 2013. Personalized time-aware tweets summarization. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 513–522.
- [30.] Asli Celikyilmaz and Dilek Hakkani-Tur. 2010. A hybrid hierarchical model for multi-document summarization. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 815–824.
- [31.] Atif Khan, Naomie Salim, “A Review on Abstractive Summarization Methods” Faculty of Computing, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia, E-mail: katif2@utm.my, naomie@utm.my.
- [32.] Pierre-Etienne Genest, Guy Lapalme Rali-Diro, “Fully Abstractive Approach to Guided Summarization” Université de Montréal P.O. Box 6128, Succ. CentreVille Montréal, Québec Canada, H3C 3J7
- [33.] C.-S. Lee, et al., "A fuzzy ontology and its application to news summarization," *Systems, Man, and Cybernetics, Part B: Cybernetics*, IEEE Transactions on, vol. 35, pp. 859-880, 2005.P.E.
- [34.] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3 (2003), 993–1022.
- [35.] Mark Steyvers and Tom Griffiths. 2007. Probabilistic topic models. *Handbook of latent semantic analysis* 427, 7 (2007), 424–440.
- [36.] Pierre-Etienne Genest, Guy Lapalme Rali-Diro, “Framework for Abstractive Summarization Using Text-to-Text Generation” Université de Montréal P.O. Box 6128, Succ. Centre-Ville Montréal, Québec Canada, H3C 3J7
- [37.] Vishal Gupta, Gurpreet Singh Lehal, “A Survey of text summarization of extractive techniques” University institute of engineering and Technology, Computer Science & Engineering, Punjab University, Chandigarh, India, E-mail: vishal@pu.ac.in, gslehal@yahoo.com.
- [38.] Ibrahim F. Moawad, Mostafa Aref, “Semantic Graph Reduction Approach for Abstractive Text Summarization” Information Systems Dept. Faculty of Computer and Information Sciences, Ain shams University Cairo, Egypt.
- [39.] Charles F. Greenbacker, “Towards a Framework for Abstractive Summarization of Multimodal Documents” Dept of Computer and Information Sciences University of Delaware Newark, U.S.A.
- [40.] Genest and G. Lapalme, "Framework for abstractive summarization using text-to- text generation," in *Proceedings of the Workshop on Monolingual Text-To Text Generation*, 2011, pp. 64-73.