An Automatic Text Summarization: A Systematic Review

Vishwa Patel, Nasseh Tabrizi

East Carolina University,
USA

patelvi17@students.ecu.edu, tabrizim@ecu.edu

Abstract. The 21st century has become a century of information overload, where in fact information related to even one topic (due to its huge volume) takes a lot of time to manually summarize it into few lines. Thus, in order to address this problem, Automatic Text Summarization methods have been developed. Generally, there are two approaches that are currently being used in practice: Extractive and Abstractive methods. In this paper, we have reviewed papers from IEEE and ACM libraries those related to Automatic Text Summarization for the English language.

Keywords. Automatic text summarization, extractive, abstractive, review paper.

1 Introduction

Nowadays, the Internet has become one of the important sources of information. We surf the Internet, to find some information related to a certain topic. But search engines often return an excessive amount of information for the end user to read increasing the need for Automatic Text Summarization (ATS), increased in this modern era. The ATS will not only save time but also provide important insight into a piece of information.

Many years ago scientists started working on ATS but the peak of interest in it started from the year 2000. In the early 21st century, new technologies emerged in the field of Natural Language Processing (NLP) to enhance the capabilities of ATS. ATS falls under NLP and Machine Learning (ML).

The formal definition of ATS is mentioned in this book [11] "Text summarization is the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks)".

A general approach to write summary of a person is that read the whole text first and then try to express the idea with either using same words or sentences in the document or rewrite it.

In either case the most salient idea is captured and expressed. The basic objective of ATS is to create summaries which are as good as human summaries. There are various other methods to extract summaries, but these are the 2 main methods: Extractive and Abstractive Text Summarization [3].

Extractive Text Summarization extracts main keywords or phrases or sentences from the document, combine them and include them in the final summary. Sometimes, the summary generated by this approach can be grammatically erroneous.

While Abstractive Text Summarization totally focuses on generating phrases and/or sentences from scratch (i.e. paraphrasing the sentences in original documents) in order to maintain the key concept alive in summary.

Here the summary generated is free of any grammatical errors, which is an advantage compared to Extractive method.

It is important to note that the summaries generated by this approach are more similar to summaries generated by human (Similar means that the whole document's idea is rewritten using a different set of words). Implementing Abstractive Text Summarization is more difficult compared to Extractive approach.
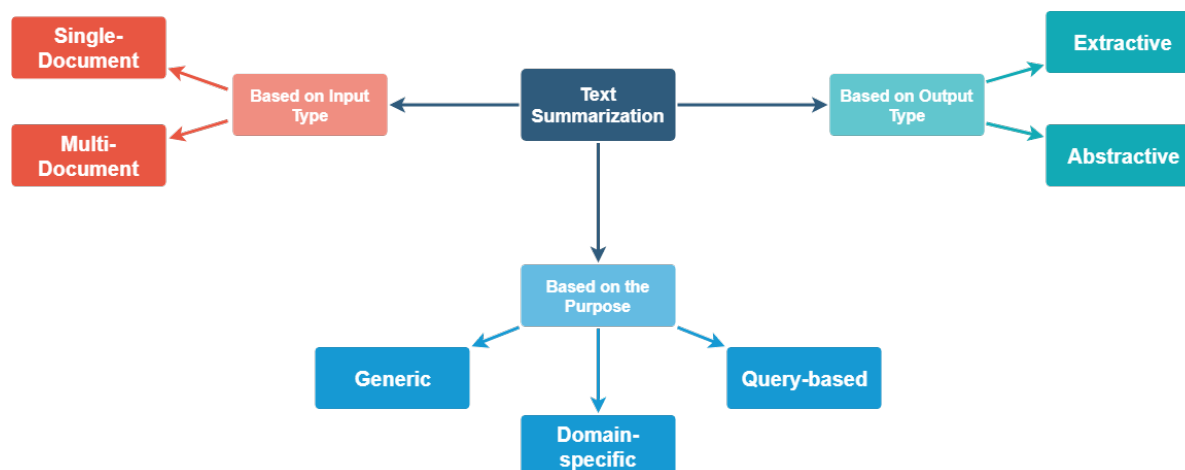
Fig. 1. Types of text summarization

2 Background

Since the idea of ATS emerged years ago, the research got accelerated when tools for NLP and ML with Text Classification, Question-Answers etc became available. Some of the advantages of ATS [3] are listed as follows:

— Reduces time for reading the document.

— Makes selection process easier while searching for documents or research papers.

— Summaries generated by ATS are less biased compared to humans.

— Personalized summaries are more useful in question-answering systems as they provide personalized information.

ATS has many applications in almost all fields, for example summarizing news. It is also useful in medical field, where long medical history of a patient can be summarized in few words which saves lot of time and also helps doctors to understand patient's condition easily.

Authors of the book [11] provide the daily useful applications of ATS which are described as follows:

— Headlines (from around the world).

— Outlines (notes for students).

— Minutes (of the meeting).

— Previews (of movies).

— Synopses (soap opera listings).

— Reviews (of a book, CD, movie, etc.).

— Digests (TV guide).

— Biography (resumes, obituaries).

— Abridgments (Shakespeare for children).

— bulletins (weather forecasts/stock market reports).

— Sound bites (politicians on a current issue).

— Histories (chronologies of salient events).

When the Internet resurfaced in 2000, data started to expand. The following two evaluation programs (conferences related to Text) were established by National Institute of Standard and Technology (NIST), Document Understanding Conferences (DUC) [18] and Text Analysis Conference (TAC) [13] in the USA. They both provide data related to summaries.

3 Types of Text Summarization

Text Summarization can be classified into many different categories. Figure 1 illustrates the different types of Text Summarization [3].

3.1 Based on Output Type

There are 2 types of Text Summarization based on Output type[3]:

— Extractive Text Summarization.

— Abstractive Text Summarization.

3.1.1 Extractive Text Summarization

Extractive Text Summarization where important sentences are selected from the given document and then are included in the summary. Most of the text summarization tools available are Extractive in nature. Below mentioned are some of the online Tools available:

— TextSummarization.

— Resoomer.

— Text Compactor.

— SummarizeBot.

The full list of ATS tools available at [16]. The process of analyzing the document is fairly straight forward.

First the information (text) is pre-processing step where all words are converted into either lower-or-upper-case letters, stop words are eliminated and remaining words are converted into their root forms. Next step is to extract different features based on which next step will be decided. Some of these features are:

— Length of the Sentence.

— The frequency of the word.

— The most appearing word in sentence.

— Number of characters in sentence.

Now, based on these features, using sentence scoring all the sentences will be placed either in descending or ascending order. In the last step, the sentences which have the highest value will be selected for the summary.

Figure 2 represents the steps of the process of Extractive Summarization. Authors in [9] paper have implemented ATS for multi-document, which contains headings for each document.
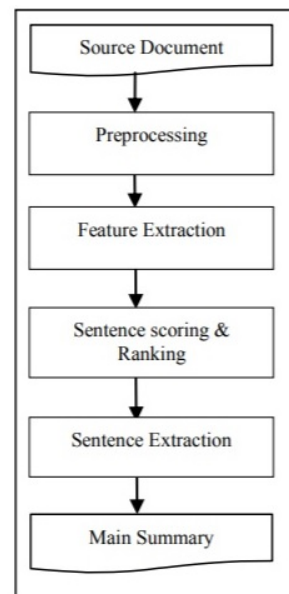

Fig. 2. Steps of extractive text summarization

3.1.2 Abstractive Text Summarization

Abstractive Text Summarization, tries to mimic human summary by generating new phrases or sentences in order to offer a more coherent summary to the user. This approach sounds more appealing because it is the same approach that any human use in order to summarize the given text.

The drawback of this approach is that its practical implementation is more challenging compared to Extractive approach. Thus, most of the tools and research have focused on Extractive approach. Recently, researchers are using deep learning models for Abstractive approaches and are achieving good results.

These approaches are inspired by Machine Translation problem. The authors [2] presented the attentional Recurrent Neural Network (RNN) encoder-decoder model which was used in Machine Translation, which produced excellent performance. Researchers then formed Text Summarization problem as a Sequence – to – Sequence Learning.

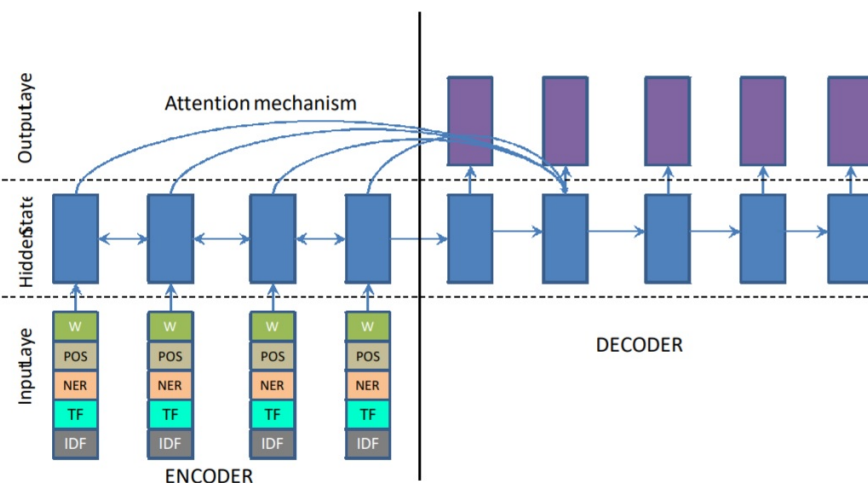Other authors [12] have used the same model "Encoder-Decoder Sequence-to-Sequence

Fig. 3. Architecture of encoder-decoder sequence-to-sequence RNN

RNN" and used it in order to obtain a summary which is an Abstractive method. Figure 3 is an "Encoder-Decoder Sequence-to-Sequence RNN" architecture.

3.2 Based on Input Type

There are 2 types of Text Summarization based on Input types [3]:

— Single-document.

— Multi-document.

3.2.1 Single-Document

This type of text summarization is very useful for summarizing the single document. It is useful for summarizing short articles or single pdf, or word document. Many of the early summarization systems dealt with single document summarization.

Text Summarizer [17] is an online tool which summarizes a single document, it takes URL or text as input and summarize the input document into several sentences.

It will accept the input text and will find the most important sentences and will include them into the final summary.

3.2.2 Multi-Document

It gives summary which is generated from multiple text documents. If multiple documents on related to specific news, or articles are given to multi-document text document then it will be able to create a concise overview of the important events.

This type of ATS is useful when user needs to reduce overall unnecessary information, because these multiple documents of articles about the same events can contain several sentences that are repeated.

3.3 Based on Purpose

There are 3 text summarization techniques based on its purpose [3]:

— Generic.

— Domain specific.

— Query based.

3.3.1 Generic

This type of text summarization is general in application, where it does not make any assumption regarding the domain of article or the content of the text. It treats all the inputs as equal.

For example, generating headlines of news articles, generating a summary of news, summarizing a person's biography, or summarizing sound-bites of politicians, celebrities, entrepreneurs etc. Most of the work that has been done in ATS field, is related to generic text summarization.

The authors [15] has developed ATS which summarizes news articles. These articles can consist of news from different categories. Any ATS which is not explicitly designed for a specific domain or topic, falls under this category.

3.3.2 Domain-Specific

This type of text summarization is different than generic type, domain means the topic of the text. Domain-specific means that the model uses domain-specific knowledge along with the input text, it helps in producing a more accurate summary.

An example of this could be a text summarization model which uses a heart(cardiology) related knowledge, or computer science related knowledge. The main benefit here is that domain-specific knowledge helps model to understand the context of the text and can extract more important sentences which are related to the field.

3.3.3 Query Based

Query based means that user gives the query to text summarization tool which then retrieves information related to that query. This type of tool is mainly used for natural language question-answers. The goal here is to extract personalized summary based on user needs.

For an instance, if a article is related to "John", "Car" and user wants to extract summary which is related towards "john" then ATS will retrieve summary which is related to it.

4 Current Research in Automatic Text Summarization

The authors [8] presented an ATS system which summarizes Wikipedia articles using an Extractive Approach. They first perform preprocessing step, where the text is tokenized, porter-stemming is applied, and 10 different features are extracted (f1 - f10) and given as the input to neural network with one hidden layer and one output layer. Output scores ranges from 0-1.

This score is proportional to the importance of the sentence. These scores are then used to generate summary. Windows Word 2007 is used to generate summary for the same article. Summary generated from Microsoft Word 2007 is referred to as "Reference Summary".

Both summaries (Reference Summary and System Generated Summary) is then used to evaluate model performance, and precision, recall and f1-score are calculated. Model performs best if it uses the only f9 feature with f-1 score of 0.223. Similarly, f7 has lowest f1-score of 0.055.

Other authors [6], presented a 4 dimensional graph model for ATS. Graph models show the relationship between the sentences in the text, which is valuable for ATS tasks. They used the TextRank algorithm to evaluate in the context of Extractive Text Summarization.

They used CNN dataset for evaluation. Their model improves the TextRank algorithm overall(better precision, recall and f-measure) by improving 34.87% in relation to the similarity model. Here is the list of 4 dimensions which were used to create the graph:

— **Similarity**, It measures the overlapping content between pairs of sentences. If it exceeds a threshold score which is selected by the user, then edge between the sentence pair is created.
— **Semantic Similarity**, It employs ontology conceptual relations such as synonyms, hyponym and hypernym. Then sentences must first be represented as vectors with words and the semantic similarity scores for each pair of words using WordNet must be calculated.

— **Coreference resolution** It is the process by which they identified the noun that was referring to the same entity. There are 3 forms of coreference: named, nominal or pronominal.

— **Discourse Relations** It is used to highlight the relevant relationships in the text.

The authors [1] proposed a Query-oriented ATS using Sentence Extraction technique. First the input text is pre-processed (Tokenization, Stop Words Removal, Stemming and POS tagging), then 11 features were extracted from the input text.

The first set of features are used to identify informative sentences and the second set of appropriate features will help to extract the query relevant sentences. Based on those features, each sentence was scored, and used DUC-2007 dataset for training and evaluation purpose.

Min-Yuh Day and Chao-Yu Chen [4] proposed an AI approach for ATS. They have developed ATS with 3 different models: Statistical, Machine Learning and Deep Learning Models. They used Essay titles and abstracts as their dataset.

Using the Essay abstracts as input, it is inputted into all 3 models, and a headline for essay is generated by all 3 models. Then all 3 generated title summaries are evaluated by using ROUGE evaluation metric, and then best fitting title is selected.

The authors in [5] published an article presents an unsupervised extractive approach based on graphs. This method constructs an undirected weighted graph from the original text by adding a vertex for each sentence and calculates a weighted edge between each pair of sentences that are based on a similarity/dissimilarity criterion.

A ranking algorithm is applied and most important sentences based on their corresponding rank are identified. They used DUC-2002 dataset for their analysis. Results are then evaluated using ROUGE-1 using different distance measures like LSA, TextRank, Correlation, Cosine, Euclidean etc.

Other authors [7] introduced an ATS which is based on an unsupervised graph based ranking model. This model builds a graph by collecting words and their lexical relationships from the document. They collect a subset of high rank and low rank words. Sentences are extracted based on how many high rank words are present in it.

Collecting such sentences leads to generating a summary of the document. Here authors have focused on using ATS for people who are visually challenged or visual loss. They have tested the proposed system on NIPS(Neural Information Processing System) Dataset. They have focused on Single Document Summarization.

5 Research Methodologies

The purpose of this study is to investigate the trends in which the Automatic Text Summarization (ATS) for English language have progressed by doing research on published articles and to gain intuition on the current direction of ATS.

The first step was researching for "Automatic Text Summarization" from different databases. There were 160+ papers published in our selected databases. As our goal is to study the current trends in ATS field, so the research was limited to past 7 years (2012-2019), and omitted any type of book and early available articles.

Therefore, we started our search in 2 databases: *IEEE*, *ACM*, using this query: "Automatic Text Summarization" (with quotes). Among all the related papers, some papers were not related to our topic, for instance, some paper had built ATS for different languages like Arabic, Hindi etc.

Since our focus is on ATS for the English Language, these papers were excluded from our research. For writing information about ATS, we have used several articles and some research papers.
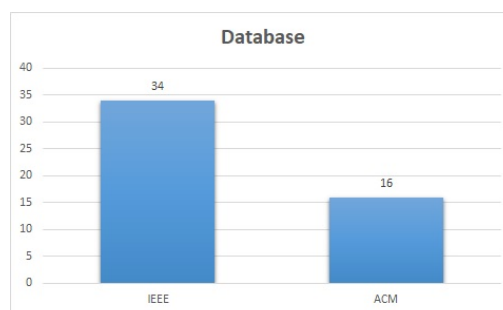

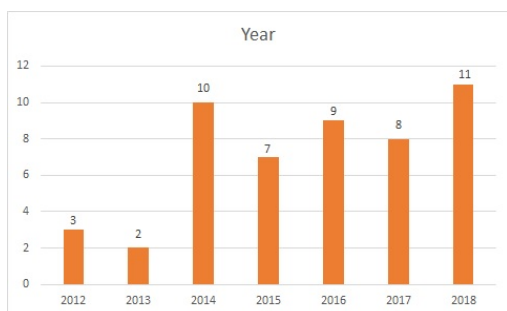**Fig. 4.** Distribution of papers by database

Fig. 5. Distribution of papers by year of publication

6 Classification of Papers

The selected papers from 2 different databases were classified them in different categories. The details are described below:

6.1 Distribution by Database

We have collected 50 research papers from 2 different databases: *IEEE* and *ACM*. 34 papers were found from IEEE (68%) and, 16 papers from ACM (32%) see *Figure-4*. All the research papers are conference paper which is related to our topic "Automatic Text Summarization".

6.2 Distribution by Publication Year

We found 160+papers while firing query related to our topic. We restricted our search to the past 7 years. As we can see from *Figure-5* that initially in 2012-2013 there were not much research being conducted, while from 2014 there was an increase in published papers. From the data we can clearly say there were 2 years having the highest number of publications, 11 publications in 2018, while 10 publications in 2014.

6.3 Distribution by Type of ATS

6.3.1 Based on Output Type

As mentioned earlier in the paper, there are 2 types of ATS based on Output type: Extractive Text Summarization and Abstractive Text Summarization.

We have increased one more category as "Hybrid" where researchers have used both Text Summarization technique and combined them to generate summaries. We classified our papers based on these types, Figure 6 shows that among the papers, 46 Extractive type, 2 Abstractive and 2 Hybrid.

6.3.2 Based on Input Type

As, there are 2 types of Text Summarization based on Input type: Single and Multi Text Summarizations. Figure 7 represents that 46 of papers were based on Single Document ATS, while 4 were based on Multi Document ATS. Most of the researchers focus on summarizing single input document.

7 Dataset

Among all papers, DUC[18] dataset was most popular among the research studies. DUC offers single document articles with handwritten summaries. These summaries are also referred to as "Gold Summary", which is used to compare the resultant summary obtained by Text Summarization. DUC has many different datasets according to year wise, starting from 2001-2007.

The second most popular dataset is CNN dataset, which consists of news and/or articles from CNN website. There are other datasets which were used in papers which consists of (but not limited to) Gigaword, Elsevier Articles, Opinonis and Daily Mail.
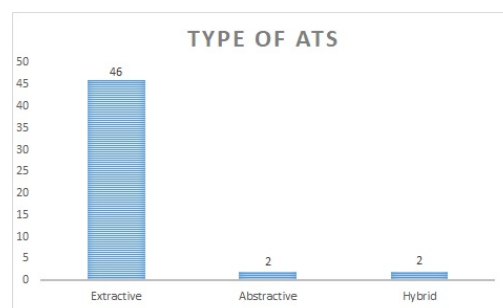

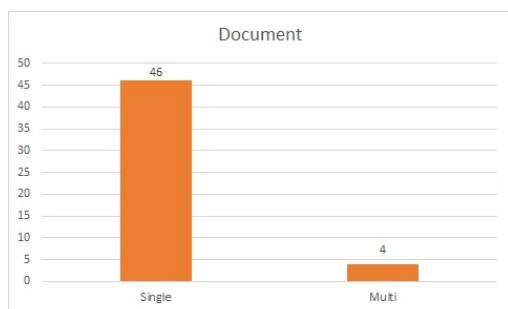**Fig. 6.** Distribution of papers by output type of text summarization

Fig. 7. Distribution of papers by input type of text summarization

8 Evaluation Technique

Main evaluation methodology used is ROUGE evaluation [10]. ROUGE stands for Recall-Oriented Understudy. It is a set of metrics for evaluating automatic test summarization of text and also for machine translation.

Basically, it compares 2 different types of summaries, Automatically Produced Summary by ATS and Set of Reference Summary (which is typically produced by humans). Another evaluation methodology used for evaluation is F-1 measure, where Precision and Recall is calculated.

F1 score or *F-measure* is used to calculate the accuracy of a certain system. F1 score calculation uses both, Precision (p) and Recall (r). Precision is the fraction of the summary that is correct.

Recall is the fraction of the correct (model) summary that is outputted. Some papers, however, did not use any evaluation metrics to check the accuracy.

Precision in the Context of ROUGE, we are actually measuring how much of the ATS summary was actually relevant or needed? While Recall means that how much of the reference summary is the ATS summary recovering or capturing?

$$\text{Precision} = \frac{\text{number\_of\_overlapping\_words}}{\text{total\_words\_in\_system\_summary}}, \quad (1)$$

$$\text{Recall} = \frac{\text{number\_of\_overlapping\_words}}{\text{total\_words\_in\_reference\_summary}}. \quad (2)$$

Besides Precision and Recall, there are 3 other evaluation metrices:

— ROUGE-N.

— ROUGE-L.

— ROUGE-S.

8.1 ROUGE-N

This ROUGE package [10] is used to measure unigrams, bi-grams, trigrams and higher order n-grams overlap. For example,ROUGE-1 refers to unigrams, whereas ROUGE-2 refers to bigrams, ROUGE-3 as trigrams etc.

They use both summaries, system summary and reference summary to calculate overlap of unigrams, or bigrams or trigrams or any high order n-grams [14].

8.2 ROUGE-L

This measures [10] longest matching sequence of words using LCS. The advantage of using LCS is that it does not require consecutive matches but in sequence matches that reflects sentence level word order.

It automatically includes the longest in-sequence common n-grams, that's why there is no need of defining predefined n-gram length [14].

8.3 ROUGE-S

Skip-gram measures the overlap of word pairs that can have maximum n gaps between words. For example, skip-bigram measures the overlap of word pairs that can have a maximum of two gaps between words [14, 10].

9 Conclusion

In this paper, we have researched 50 papers from IEEE and ACM databases in Automatic Text Summarization. We described different type of ATS based on input, output and purpose. Current research studies are also discussed in the paper.

We distributed the collected papers in various categories like by year, input type, output type and database and discussed various databases used by researchers. Finally, the most used for evaluation metric, ROUGE is explained along with its different metrics.

Acknowledgment

References

1. **Afsharizadeh, M., Ebrahimpour-Komleh, H., Bagheri, A. (2018).** Query-oriented text summarization using sentence extraction technique. 2018 4th International Conference on Web Research (ICWR), IEEE, pp. 128–132.

2. **Bahdanau, D., Cho, K., Bengio, Y. (2014).** Neural machine translation by jointly learning to align and translate.

3. **Chauhan, K. (2018).** Unsupervised text summarization using sentence embeddings.

4. **Day, M. Y., Chen, C. Y. (2018).** Artificial intelligence for automatic text summarization. 2018 IEEE International Conference on Information Reuse and Integration (IRI), pp. 478–484.

5. **de la Peña Sarracén, G. L., Rosso, P. (2018).** Automatic text summarization based on betweenness centrality. Proceedings of the 5th Spanish Conference on Information Retrieval, pp. 11.

6. **Ferreira, R., Freitas, F., Cabral, L. d. S., Lins, R. D., Lima, R., França, G., Simskez, S. J., Favaro, L. (2013).** A four dimension graph model for automatic text summarization. 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), volume 1, pp. 389–396.

7. **Hamid, F., Tarau, P. (2014).** Text summarization as an assistive technology. Proceedings of the 7th International Conference on Pervasive Technologies Related to Assistive Environments, Association for Computing Machinery, pp. 60.

8. **Hingu, D., Shah, D., Udmale, S. (2015).** Automatic text summarization of Wikipedia articles view document. 2015 International Conference on Communication, Information & Computing Technology, IEEE, pp. 1–4.

9. **Krishnaveni, P., Balasundaram, S. R. (2017).** Automatic text summarization by local scoring and ranking for improving coherence. 2017 International Conference on Computing Methodologies and Communication, pp. 59–64.

10. **Lin, C. Y. (2004).** ROUGE: A package for automatic evaluation of summaries. Text Summarization Branches Out, Association for Computational Linguistics, pp. 74–81.

11. **Mani, I., Maybury, M. T. (1999).** Advances in automatic text summarization. MIT press.

12. **Nallapati, R., Zhou, B., dos Santos, C. N., Gulcehre, C., Xiang, B. (2016).** Abstractive text summarization using sequence-to-sequence RNNs and beyond.

13. **NIST (2018).** Text Analysis Conference.

14. **RxNLP (2018).** ROUGE Evaluation Metrics.

15. **Sethi, P., Sonawane, S., Khanwalker, S., Keskar, R. B. (2017).** Automatic text summarization of news articles. 2017 International Conference on Big Data, IoT and Data Science (BID), pp. 23–29.

16. **Softsonic (2018).** List of Online Automatic Test Summarization Tools.

17. **Text Summarizer (2018).** Manual of Text Summarization.

18. **Voorhees, E. (2002).** Document understanding conferences website.