WJCMS

Wasit Journal of Computer and Mathematics Science

Journal Homepage: https://wjcm.uowasit.edu.iq/index.php/WJCM

e-ISSN: 2788-5879 p-ISSN: 2788-5879

Natural Language Processing For Automatic text summarization [Datasets] - Survey

Alaa Ahmed AL-Banna[1]* and Abeer K AL-Mashhadany[1]

[1]Department of Computer Science, College of Science ,Al-Nahrain University, Baghdad, Iraq

*Corresponding Author: Alaa Ahmed AL-Banna

DOI: https://doi.org/10.31185/wjcm.72

Received: August 2022; Accepted: October 2022; Available online: December 2022

ABSTRACT: Natural language processing has developed significantly recently, which has progressed the text summarization task. It is no longer limited to reducing the text size or obtaining helpful information from a long document only. It has begun to be used in getting answers from summarization, measuring the quality of sentiment analysis systems, research and mining techniques, document categorization, and natural language Inference, which increased the importance of scientific research to get a good summary. This paper reviews the most used datasets in text summarization in different languages and types, with the most effective methods for each dataset. The results are shown using text summarization matrices. The review indicates that the pre-training models achieved the highest results in the summary measures in most of the researchers' works for the datasets. Dataset English made up about 75% of the databases available to researchers due to the extensive use of the English language. Other languages such as Arabic, Hindi, and others suffered from low resources of dataset sources, which limited progress in the academic field.

Keywords: Natural Language Processing, Automatic Text Summarization, Abstractive Text Summarization, Extractive Text Summarization, Text Summarization Datasets

1. INTRODUCTION

Automatic summarization is a challenge in Natural Language Processing (NLP) that involves developing algorithms that can reduce an input text, such as a scientific journal article, to a compacted version containing just the relevant information. The process of summarizing is complex because it involves not only the selection, assessment, collection, and rearrangement of information but also compression, generalization, and paraphrasing [1]. There are several types of summarizations based on the desired type of input and summary. The two major types of summarizations are Extractive and Abstractive summarizations [2]. In extractive summarization, text segments such as sentences or phrases of the most important are extracted from the original text without being modified words and concatenated to generate the summary [3]. Extractive summarization aims to identify the importance of each text sentence and develop a shorter version of the original text that represents it accurately. On the other hand, text summarization based on abstraction uses linguistic methods for understanding and a deeper analysis of the text [4]. An abstract text summary type consists of new sentences generated by paraphrasing or reformulating the extracted content [5]. Abstractive summaries are usually resembling human-written ones because they tend to represent the content and the meaning of the original text more naturally. Although this type of summarization performs more efficiently, its implementation requires great knowledge of deep learning techniques [6]. Also, the type of input documents used for summarizing may vary. The target of summarizing may be to obtain a summary from several documents or one document to obtain a summary text [7]. Single-document summarizers deal with a single source text and generate summaries from it Independently of other documents [8]. On

the other hand, multi-document summarizing is viewed as an extension of a single-document summary. It compiles many documents on the same subject into a single summary. The multi-document summarization task is more complex than summarizing a single document, even if it is lengthy. The difficulty comes from handle within a large set of documents with thematic diversity [9].

1.1 AUTOMATIC TEXT SUMMARIZATION EVALUATION

The quality of the summarized text is evaluated by human assessment and using Natural Language Text Summarization metrics. ROUGE, or" Recall-Oriented Understudy for Gisting Evaluation," is the most popular matrices containing software packages for evaluating automatic summarization and translation in natural language processing. By comparing an automatically produced summary with a reference or a group of contacts (human-produced) [10]. The overlap of unigrams (per word) between the system and reference summaries is referred to as ROUGE-1. In contrast, ROUGE-2 refers to the bigram overlap between the system and reference summaries. And ROUGE-L: Statistics based on the "Longest Common Subsequence (LCS) " The longest common sub-sequence naturally considers sentence-level structural similarities and automatically determines the longest co-occurring in sequence n-grams [11]. In addition to these matrices, human evaluation is adopted based on the correct linguistic rules for the summary, the coherence of the text, etc. [12].

2. OVERVIEW OF THE DATASETS

The dataset is the first step to getting a well-trained model to perform a specific task in artificial intelligence [13]. When summarizing texts, it is necessary to look at the available datasets, languages, types of functions, and the latest methods that led to obtaining a well-trained model. Below we review the most famous datasets that have been worked with a summary table of information:-

1) Multi-Document Summarization on "Wikipedia Current Events Portal (WCEP)"

The WCEP dataset for multi-document summarization comprises human-written summaries of news events from the ("WCEP"), each linked with a cluster of news articles. These stories are based on sources mentioned by WCEP editors and automatically collected from Common Crawl News and supplemented with items acquired automatically from t (WCEP) [14]. The best result in the dataset shown in [15] is by using the PRIMERA model by result Rouge-1=46.1 Rouge-2=25.2 Rouge-L=37.9. Multi-Document Summarization captures useful information and filters out superfluous information to summarize a series of documents. Extractive and abstractive multi-document summarization are popular. Extractive summarizing systems extract significant samples, phrases, or sections from texts, whereas abstractive summarization systems paraphrase the information [9].

2) DUC 2004 (Document Understanding Conferences)

The DUC2004 dataset is a multi-document summarization dataset used for testing. It comprises 500 news stories, each with four human-written summaries. It comprises 50 clusters of "Text REtrieval Conference (TREC) " documents drawn from the following collections: "AP newswire, 1998-2000; New York Times newswire, 1998-2000; and Xinhua News Agency (English version) "through 1996-2000. Each cluster had an average of ten papers [16]. in [17] Summarization values got their highest levels: ROUGE-1 = 28.18 ROUGE-2 = 8.49 ROUGE-L = 23.81.

3) CNN/Daily Mail

The "CNN/Daily Mail" dataset is used for summarizing text. Questions (with one of the entities obscured) and associated portions from CNN and Daily Mail news articles were manually constructed to train the algorithm to answer a fill-in-the-blank inquiry. Authors have made available scripts to crawl, extract, and produce passage and question pairs from various resources. The hands specify 11,487 test pairs, 13368 validation pairs, and 286,817 training pairs. Documents in training average 766 words and 29.74 sentences in length, whereas their summaries only use 53 comments and 3.72 sentences [18].in the latest studies in 2022 [19] using “Pegasus 2B + SLiC”, Y. Zhao and others get the best result to text summarization task in this data set, Rouge-1= 47.97, Rouge-2= 24.18, Rouge-L= 44.88.

4) PubMed ("Public/Publisher MEDLINE")

There are a total of 19717 studies on diabetes in the PubMed dataset, which may be broken down into three categories. There are a total of 44338 references in the citation web. A word vector from a vocabulary with 500 different terms is used to define each article in the dataset according to its TF/IDF importance [20]. By using transformers architectures [21] to get the best value to abstractive text summarization in ROUGE-1 =50.95, ROUGE-2= 21.93, and ROUGE-L= 45.61.

5) arXiv ("Arxiv HEP-TH (high energy physics theory citation graph")

There are 27,770 publications included in the dataset that makes up the "Arxiv HEP-TH" citation graph, and the total number of edges is 352,807. A directed edge connects two nodes in the network if and only if paper I refers to paper J. There is no indication in the graph of any papers outside the dataset being referenced or cited by any documents inside it. Specifically, the data includes articles published between January 1993 and April 2003. (124 months) [22]. the transformer

was applied to summarize the text in [21] and reached a ratio of ROUGE-1 =50.95, ROUGE-2 =21.93 ROUGE-L =45.61.

6) XSum (Extreme Summarization)

Extreme Summarizing (XSum) is a dataset for testing abstract single-document summarization methods. The objective is to write a brand-fresh, catchy one-sentence summary that explains the article's subject. A one-sentence summary is provided for each of the 226,711 news stories in the collection. Many different topics are covered in this compilation of BBC stories that span from 2010 to 2017. (e.g., "News, Politics, Sports, Weather, Business, Technology, Science, Health, Family, Education, Entertainment, and Arts"). 204,045 (90%), 11,332 (5%) and 11,334 (5%) documents make up the official training, validation, and test sets, respectively [23]. In [19], researchers work on this dataset and introduce sequence likelihood calibration (SLiC). This makes Decoding unnecessary, and the quality of decoding candidates rises dramatically independent of the decoding method. Exceed SOTA results on various generation tasks, including abstractive summarization, question creation, abstractive question answering, and data-to-text generation, even with small models. Pegasus 2B +SLiC achieved result up to ROUGE-1 =49.77, ROUGE-2 =27.09, and ROUGE-L =42.08.

7) MentSum ("Mental Health Summarization Dataset")

Mental health is still major one in the field of public health. Many people today are turning to internet forums and social media to discuss their struggles with mental health, vent their emotions, and connect with like-minded and trained professionals. The length of the articles may vary, but it is helpful to offer a brief yet relevant description so that the counselors can go through it quickly. MentSum includes over 24k hand-picked user posts from Reddit and their short user-written summary (called TL; DR) in English from 43 mental health subreddits to facilitate research into the summarization of online posts related to mental health. This domain-specific dataset could be of interest not only for generating short summaries on Reddit but also for generating summaries of posts on dedicated mental health forums such as Reachout. in 2022 [24]. This paper [25]works on this dataset and evaluates extractive and abstractive state-of-the-art summarization baselines [24]. Using the BART model, better values in this research are "Rouge-1 =29.13, Rouge-2 =7.98, and Rouge-L =20.27".

8) OrangeSum

OrangeSum is an extreme summarizing dataset that focuses on a single document to summarize it. It includes two tasks, title and abstract. Title and abstract assignments have an average ground truth summary length of 11.42 and 32.12 words, whereas 315- and 350-word documents have similar average sizes. Creating a French language version of the XSum dataset was the impetus behind OrangeSum. For models to do well in OrangeSum, they need to be more abstract than they need to be on the historical CNN, Daily Mail, and NY Times datasets. Extracting article titles and abstracts from the orange Actu website led to the development of OrangeSum. Pages scraped span over a decade, from February 2011 to September 2020, and may be roughly categorized under five broad headings: France, the globe, politics, automobiles, and society. Health, environment, people, culture, media, high technology, unusual ("insolite" in French), and miscellaneous are the eight divisions of the society event [26] .in the 2021 paper [27] provides "BARThez", the first large-scale pre-trained seq2seq model for French, in this paper. "BARThez" is particularly well-suited for generative jobs because of its BART foundation.BARThez is very competitive with cutting-edge BERT-based French language versions such as CamemBERT and FlauBERT.and also proceed to train a multilingual BART on BARThez' corpus and show that the resultant model, mBARThez, greatly improves BARThez' generative performance.

9) BookSum (Books Summarization)

BookSum is a data set library for summarizing books and other lengthy texts. Novels, dramas, and tales are some of the literary works included in this dataset, which also has highly abstractive, human-written summaries at the paragraph, chapter, and book levels. Longer texts, non-trivial causal, temporal relationships, and complex discourse structures are only some of the obstacles that summarizing algorithms must overcome in this dataset. So far, BookSum has summarized 142,753 paragraphs, 12,293 chapters, and 436 volumes [28]. Using BART-LS in [29], the measurements ROUGE-1=38.5, ROUGE-2=10.3, and ROUGE-L=36.4.

10) arXiv Summarization Dataset

For more than 30 years, ArXiv has served the public and scientific communities by offering open access to scholarly publications in fields ranging from physics to computer science and everything in between, including "math, statistics, electrical engineering, quantitative biology, and economics". This vast repository of data provides substantial but, at times, bewildering depth. This dataset is a free, open pipeline on Kaggle to the machine-readable arXiv dataset: a library of 1.7 million articles containing important information such as article titles, authors, categories, abstracts, full-text PDFs, and more, to help make the arXiv more accessible. To enable new use cases that can lead to the exploration of more profitable machine learning techniques that combine multi-modal features for applications such as trend analysis, paper recommender engines, category prediction, co-citation networks, knowledge graph construction, and semantic search interfaces. ArXiv is a community-supported, jointly financed resource created by Paul Ginsparg in 1991 and maintained and controlled by Cornell University [30]. Savelieva and others In [31], using Facet-Aware Modeling Improving Unsupervised Extractive Summarization, the best result is ROGUE-1=40.92 ROUGE-2=13.75 ROUGE-L =35.56.

11) WikiHow Dataset

"WikiHow" is a dataset of over 230,000 article-summary pairs mined from a wiki-based knowledge repository and created by various human writers. The topics covered and the writing styles represented in these pieces have high diversity [32]. Using state-of-the-art NLP in [33] makes abstractive summaries of "recorded instructional films ranging from gardening and cooking to software configuration and sports". The model is pre-trained on a few large cross-domain datasets in written and spoken English using transfer learning. ROUGE measurement is used to evaluate, and this paper has achieved the best results and used BertSum by relying on this dataset, as it was ROUGE-1 =35.91, ROUGE-2 =13.9, ROUGE-L =34.82.

12) Urdu News Dataset

Over a million news articles from the fields of business, economics, science and technology, entertainment, and sports are included in this collection. This dataset is helpful for numerous Urdu NLP applications since its four unique categories were carefully selected to eliminate ambiguity. For many NLP, Machine/Deep Learning tasks, including "text processing, classification, summarization, named entity recognition, topic modeling, and text generation", the dataset A Large-Scale News Dataset for Urdu Text Processing is the only dataset in the Urdu language that is currently available—created this dataset.in 2021 and so far, no studies have been done on this dataset in text summarization [34].

13) Bengali News Articles ("IndicNLP")

Since the previous several decades, "natural language processing (NLP)" has been used extensively in studying Western languages, especially the English language. Language processing research on the eastern counterpart, particularly the languages of the Indian subcontinent, needs to be increased. Western languages have access to a wealth of dictionaries, WordNet, and related resources. This data collection, which has been cleaned and comes with a train and test set to compare your classification and summarization models against, contains 14k news items. A collection of Bengali news items make up this dataset. It may be applied to problems involving classification and language modeling [35]. Although there has been a substantial amount of critical study on abstractive summary in English, just a few works have been done on Bengali abstractive news summarization. Paper [36] described a seq2seq-based Long Short-Term Memory (LSTM) network architecture focused on the encoder-decoder. The suggested system uses the attention-based model to construct a long sequence of words. The summary was evaluated subjectively and statistically, and its results were compared to other published results. Mechanism of attention demonstrated a considerable improvement in state-of-the-art human assessment ratings.

14) Hindi Text Short and Large Summarization Corpus

The "Hindi Text Short and Large Summarization Corpus" is a collection of 180k articles with headlines and summaries taken from Hindi news websites. This is a first-of-its-kind Hindi Dataset that may be used to benchmark algorithms for Hindi text summarization. This does not include articles from this dataset, which is being released concurrently with this Dataset. The dataset preserves the articles' original punctuation, numerals, and other formatting [37]. In 2022 [38] executed the Hindi text summarization, which had received relatively little attention. A machine learning model has been developed and evaluated using around 100,000 data samples, resulting in highly accurate summaries that benefit society. The model has an F-Score of 58% and a Rouge Score of 67.5%. Pandas, NumPy, sklearn, and other libraries are utilized.LSTM, word embedding and seq2seq are used to train the data model.

15) Arabic News articles from Aljazeera net

Natural Language Processing (often known as NLP) is a highly researched area of machine learning. Recent years have seen significant development, paving the way for this field's expansion into widespread use across various contexts. Today, natural language processing (NLP) is used in several contexts, including but not limited to: social media platforms, search engines, translation apps, Chabot helpers, and many more. Progress and outcomes, however, vary from language to language. Most ML systems only care about English and ignore other languages, particularly Arabic. A primary cause of this is the scarcity of relevant datasets. This data collection comprises around 5870 Arabic-language news stories extracted from the aljazeera.net website [39]. An abstractive Arabic text summarization model based on RNN is proposed and used in this dataset in this study [40]. A multilayer encoder and a single-layer decoder are included. Encoder layers use bidirectional long short-term memory, whereas decoder levels use unidirectional long short-term memory. The evaluation shows that the model produced is the best, achieving ROUGE1=38.4.

16) COVID-19 Open Research Dataset Challenge (CORD-19)

In response to COVID-19, the White House and prominent academic organizations created the COVID-19 Open Research Dataset (CORD-19). CORD-19 has almost 1,000,000 academic publications on COVID-19, SARS-CoV-2, and similar coronaviruses, including over 400,000 in full text. This free dataset is supplied to the worldwide research community to utilize natural language processing, and other AI approaches to develop new insights in the battle against this deadly illness. The tremendous increase in coronavirus literature makes it challenging for medical researchers to stay up [41]. In 2022 paper [42] proposes a hybrid, unsupervised, abstractive approach that walks through a document, generating salient textual fragments representing its key points. We then select the most important sentences of the paper

by choosing the most similar sentences to the generated texts, calculated using BERTScore.and this method achieved ROUGE-1 =41.02, ROUGE-2=13.79, ROUGE-L =37.25, in the text summarization measures', which are considered the best among the works submitted on this dataset.

17) Scientific Document Summarization (SciTLDR)

A new multi-target dataset of 5.4K TLDRs over 3.2K papers. SciTLDR contains both author-written and expert-derived TLDRs in computer science. The latter are collected using a novel annotation protocol that produces high-quality summaries while minimizing the annotation burden [43]. In this dataset train =1992, valid =618, test=619.The first paper published in this data set is [44] introducing TLDR generation for scientific papers and releasing SCITLDR, a multi-target dataset of TLDR-paper pairs. The result is RPOGE-1=43.8, ROUGE-2=20.9 ROUGE-L=35.5 Using pre-trained models. This data set is considered one of the difficult types due to the small number of samples and the difficulty of training them, which is why researchers often resort to pre-training and fine-tuning methods.

18) ScisummNet Corpus

This massive corpus may be used to train citation-based summarization algorithms for scientific papers, opening new avenues of inquiry into supervised approaches. This dataset includes 1,000 samples. Has been organized since 2014 in computational linguistics and NLP papers [45]. Group of researchers in [46] using BertSum, Continual BERT, Adapter-based BERT, and SummaRuNNer in PubMed and ScisummNet datasets. Bertrum outperforms other pre-trained models used on the ScisummNet dataset, archives result to ROUGE-1= 33.0, ROUGE-2 = 13.4, ROUGE-L= 31.6, but PubMed is better In terms of assessment scores because of phrases use of complex and specialized medical terminologies rather than ScisummNet's general scientific phrases.

19) SumArabic

SumArabic is a dataset for abstractive text summarization in Arabic. The information comes from two Arabic news websites: emaratalyoum.com and emaratalyoum.com. - www.almamlakatv.com.The data is divided into four sets: training, testing, validation, and out-of-domain. Each split has the following examples: 75,817 training, 4,121 validations, and 4,174 tests. 652 out-of-domain, Total: 84,764 [47].This dataset is one of the latest additions to 2022 in the Arabic language, which does not contain academic studies on it yet.

20) TalkSumm

TalkSumm is a dataset that contains 1705 automatically-generated summaries papers in science from ACL, NAACL, EMNLP, SIGDIAL (2015-2018), and ICML (2017-2018). The dataset contains titles, URLs, and corresponding summaries. [48]This study presents a unique way for automatically producing summaries for scientific publications based on recordings of lectures at scientific conferences, suggesting that such presentations provide a cohesive and short summary of the paper's content and can serve as the foundation for effective summaries. It compiled a collection of paper summaries from 1716 publications and their accompanying videos. A model trained on this dataset outperforms models trained on a manually constructed dataset of summaries. Furthermore, human specialists confirmed the quality of our this summary [49].

Table 1. Summarizes the basic information of the dataset used to text summarization from Language of dataset ,content ,Input document Single or Multiple and Referring to the best research and methods that achieved the highest percentages in summary matrices

Dataset Name	Dataset Language	Dataset Field	Input Type Single/Multiple	Output types/Extractive	Author Ref. / Methods	Result
1. Multi-Document Summarization on WCEP [14]	English	news articles	Multi documents	Extractive	W. Xiao et al. [14]\ Primer	ROUGE-1=46.1 ROUGE-2=25.2 ROUGE-L=37.9
2. DUC 2004 [16].	English	news articles	Multi documents	Extreme/	S. Shen et al. [17]\ wise sentence	ROUGE1=28.18 ROUGE-2=8.49 ROUGEL=23.81

Continued on next page

Table 1 continued

3. CNN/Daily Mail [18]	English	news articles	Single document		Y. Zhao et al. [19]\Pegasus 2B +SLiC	ROUGE-1=47.97 ROUGE-2=24.18 ROUGE-L 44.88
4. PubMed [20]	English	scientific	Single document		B. Pang et al. [21]\Transformer	ROUGE-1=50.95 ROUGE-2=21.93 ROUGE-L=45.61
5.arXiv ("Arxiv HEP-TH (high energy physics theory) citation graph) " [22]	English	scientific	Single document		B. Pang et al. [21]\Transformer	ROUGE-1=50.95 ROUGE-2=21.93 ROUGE-L=45.61
6. XSum [23]	English	News ("Politics, Sports, Weather, Business, Technology, Science, Health, Family, Education, and Arts")	Single document	Extreme/	Y. Z o et al. [19]\Pegasus 2B +SLiC	ROUGE-1=49.77 ROUGE-2=27.09 ROUGE-L=42.08
7.MentSum ("Mental Health Summarization Dataset") [24]	English	Mental health	Single document	Extreme/	S. Sotudeh et al. [25]\ BART	ROUGE-1=29.13 ROUGE-2=7.98 ROUGE-L=20.27
8. OrangeSum [26]	French	society	Single document		M. K. Eddine et al. [27]\BARThez'	ROUGE-1=15.49 ROUGE-2=5.82 ROUGE-L=13.05
9.BookSum [28]	English	literary works	Single document		W. Xiong et al. [29]\ BART-LS	ROUGE-1=38.5 ROUGE-2=10.3 ROUGE-L=36.4

Continued on next page

Table 1 continued

10.arXiv Summarization Dataset [30]	English	scientific	Single documents	Extractive	Liang et al. [31]\Facet-Aware Modeling	ROGUE-1=40.92 ROUGE-2=13.75 ROUGE-L=35.56
11.WikiHow [32]	English	article	Single documents		Savelieva et al. [33]\ BertSum	ROUGE-1 =35.91 ROUGE-2 =13.9 ROUGE-L =34.82
12. Urdu News Dataset [34]	Urdu	news articles	single documents	-	-	-
13. Bengali News Dataset [35]	Bengali	news articles	Single documents		Bhattacharjee et al. [36]\ LSTM	Humans Evaluation
14. Hindi Text Short and Large Summarization Corpus [37]	Hindi	Hindi news	Single documents		Shah et al. [38]\ LSTM+ word embedding	F-Score =58% Rouge =67.5 %
15. Arabic News articles from Al-jazeera.net [39]	Arabic	News articles	Single documents		Suleiman et al . [40]\ RNNs	ROUGE-1=38.4
16. COVID-19 Open Research Dataset Challenge (CORD-19) [41]	English	Scientific	Single documents		Bishop et al. [42]\ GenCompareSum	ROUGE-1= 41.02 ROUGE-2 = 13.7 ROUGE-L= 37.25
17.Scientific Document Summarization ("SciTLDR") [43]	English	Scientific	Single documents	Extreme/	Cachola et al. [44]\ CATTS	-
18.ScisummNet Corpus [45]	English	Scientific	Single documents		Park et al. [46]\ BertSum	ROUGE-1= 33.0 ROUGE-2 = 13.4 ROUGE-L= 31.6
19. SumArabic [47]	Arabic	News	Single documents		-	-
20. TalkSumm [48]	English	Scientific	Single documents		Lev et al. [49]\	-

3. CONCLUSION

This review concludes that the dataset's field in summarizing texts is focused on news and scientific aspects, especially the abstract type. The available datasets lack diversity in topics, primarily literary, artistic, and some fields of science. The dataset available in a large size can be used to train the model, fine-tune a small emerging data set, and get good results. After the significant development in text summarization techniques, getting a good summary has become the beginning of other tasks related to NLP.

FUNDING

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] D. Suleiman and A. Awajan, "Deep learning based abstractive text summarization: approaches, datasets, evaluation measures, and challenges," 2020. 2020-2020.
[2] N. Munot and S. S. Govilkar, "Comparative study of text summarization methods," *International Journal of Computer Applications*, vol. 102, no. 12, 2014.
[3] V. Gupta and G. S, "A survey of text summarization extractive techniques," *Journal of emerging technologies in web intelligence*, vol. 2, no. 3, pp. 258–268, 2010.
[4] A. Sinha, A. Yadav, and A. Gahlot, "Extractive text summarization using neural networks," 2018.
[5] W. T. Hsu, C. K. Lin, M. Y. Lee, K. Min, J. Tang, and M. Sun, "A unified model for extractive and abstractive summarization using inconsistency loss," 2018.
[6] R. Nallapati, B. Zhou, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence rnns and beyond," 2016.
[7] [c. Y Lin and E. Hovy, "From single to multi-document summarization," *Proceedings of the 40th annual meeting of the association for computational linguistics*, pp. 457–464, 2002.
[8] H. Christian, M. P. Agus, and D. Suhartono, "Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF)," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 7, no. 4, pp. 285–294, 2016.
[9] Mutlu, E. A. Begum, M. A. Sezer, and Akcayol, "Multi-document extractive text summarization: A comparative assessment on features," *Knowledge-Based Systems*, vol. 183, pp. 104848–104848, 2019.
[10] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," *Text summarization branches out*, pp. 74–81, 2004.
[11] M. Ozsoy, Gulcin, I. F. N. Alpaslan, and Cicekli, "Text summarization using latent semantic analysis," *Journal of Information Science*, vol. 37, no. 4, pp. 405–417, 2011.
[12] Y. Gong and X. Liu, "Generic text summarization using relevance measure and latent semantic analysis," *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 19–25, 2001.
[13] A. Wissner-Gross, Edge.com. Retrieved 8 January 2016.
[14] D. G. G. . J. G. G. Ifrim and Dataset, "Chris Hokamp; Nghia The Pham." https://paperswithcode.com/dataset/wcep.
[15] W. Xiao, I. Beltagy, G. Carenini, and A. Cohan, "Primer: Pyramid-based masked sentence pre-training for multi-document summarization." 2021.
[16] "DUC 2004 Dataset." https://paperswithcode.com/dataset/duc-2004.
[17] S. Shen, Y. Zhao, Z. Liu, and M. Sun, "Neural headline generation with sentence-wise optimization," 2016.
[18] R. Nallapati and . B.-W. Zhou, "Cicero Nogueira dos santos; Caglar Gulcehre; Bing Xiang, CNN/Daily Mail Dataset." https://paperswithcode.com/dataset/cnn-daily-mail-1.
[19] Y. Zhao, M. Khalman, R. Joshi, S. Narayan, M. Saleh, and P. J. Liu, "Calibrating Sequence likelihood Improves Conditional Language Generation." 2022-2022.
[20] datadiscovery. nlm. nih. gov, ed. 2021.
[21] B. Pang, E. Nijkamp, W. Kryściński, S. Savarese, Y. Zhou, and C. Xiong, "Long Document Summarization with Top-down and Bottom-up Inference." 2022-2022.
[22] "Arxiv dataset." https://paperswithcode.com/dataset/arxiv.
[23] S. Narayan, . Shay, B. C. M. Lapata, and X. Dataset
[24] S. S. N. Goharian and . Z. Young *Mental Health Summarization (MentSum*, 2022.
[25] S. Sotudeh, N. Goharian, and Z. Young *MentSum: A Resource for Exploring Summarization of Mental Health Online Posts*, pp. 2022–2022.
[26] M. K. E. Antoine and J. P. Tixier
[27] M. K. Eddine, A. J, . P. Tixier, and M. Vazirgiannis *Barthez: a skilled pretrained french sequence-to-sequence model*, 2020.
[28] W. K. N. Rajani
[29] W. Xiong, A. Gupta, S. Toshniwal, Y. Mehdad, and W. T. Yih *Adapting Pretrained Text-to-Text Models for Long Text Sequences*, pp. 2022–2022.
[30] A. Cohan, . F. D. D. S. K. T. Bui, and . S. K. W. C. N. Goharian 2021.
[31] X. Liang, S. Wu, M. Li, and Z. Li, "Improving unsupervised extractive summarization with facet-aware modeling," *The Association for Computational Linguistics Findings: ACL-IJCNLP 2021*, pp. 1685–1697, 2021.
[32] M. Koupaee and . W. Y. Wang *WikiHow Dataset*, 2021.
[33] A. Savelieva, B. Au-Yeung, and V. Ramani *Abstractive summarization of spoken and written instructions with BERT*, 2020.
[34] S. Shahane *Urdu News Dataset*, 2021.
[35] Prithwirajsust *Bengali News Summarization Dataset*, 2020.
[36] P. Bhattacharjee, A. Mallick, and S. Islam, "Bengali abstractive news summarization (BANS): a neural attention approach," in *Proceedings of International Conference on Trends in Computational and Cognitive Engineering*, pp. 2021–2021, Springer.
[37] Gaurav *Hindi Text Short Summarization Corpus*, 2020.
[38] A. Shah, D. Zanzmera, and K. Mehta, "Deep Learning based Automatic Hindi Text Summarization," *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 2022–2022.
[39] A. Rhouati *Arabic News articles from Aljazeera*, 2020.
[40] D. Suleiman and A. Awajan, "Multilayer encoder and single-layer decoder for abstractive Arabic text summarization," *Knowledge-Based Systems*, vol. 237, pp. 107791–107791, 2022.
[41] 2020.
[42] J. Bishop, Q. Xie, and S. Ananiadou, "GenCompareSum: a hybrid unsupervised summarization method using salience," *Proceedings of the 21st Workshop on Biomedical Language Processing*, pp. 220–240, 2022.

[43] A. Raj *Scientific Document Summarization (SciTLDR-A)*, 2022.
[44] I. Cachola, K. Lo, A. Cohan, and D. S. Weld *TLDR: Extreme summarization of scientific documents*, 2020.
[45] M. Yasunaga, . J. K. R. Z. Alexander, R. F. I. L. D. F. Dragomir, and R. Radev 2021.
[46] J. Park and Won *Continual bert: Continual learning for adaptive extractive summarization of covid-19 literature*, 2020.
[47] M. Bani and Almarjeh 2022.
[48] G. Lev, . M. S.-S. J. H. A. J. D. Konopnicki, and T. Dataset
[49] G. Lev, M. Shmueli-Scheuer, J. Herzig, A. Jerbi, and D. Konopnicki *Talksumm: A dataset and scalable annotation method for scientific paper summarization based on conference talks*, 2019.