A Survey of Automatic Text Summarization Systems, Applications, Challenges, and Future Trends

Uchkempirov Murataly
Tokyo International University
4-42-31 Higashi-Ikebukuro,
Tokyo, 170-0013 Japan

Shoon Lei Phyu
Tokyo International University
4-42-31 Higashi-Ikebukuro,
Tokyo, 170-0013 Japan

Aadhya Lakshmi Raja
Tokyo International University
4-42-31 Higashi-Ikebukuro,
Tokyo, 170-0013 Japan

Kulkarni Parag
Tokyo International University
4-42-31 Higashi-Ikebukuro,
Tokyo, 170-0013 Japan

ABSTRACT
Over the past decades, the amount of information available on the Internet has increased exponentially. This creates the need to obtain valuable information or knowledge from a huge amount of data within a short period of time. The research community has conducted several natural language processing studies to improve automatic text summarization results. Several literature reviews in computational linguistics also aim to provide comprehensive knowledge about text summarization methods, approaches, and techniques. This survey article is distinguished because it provides a holistic review of studies on automatic text summarization advancements, challenges from 1953 to 2023, and future trends. As one of the new trends authors propose a predictive text summarization approach based on predictive analytics techniques. The results and conclusion contribute not only to the academic community, but also to cater to the interests of writers, journalists, and specialists in various fields which use summarization tools regularly.

General Terms
Text summarization, information retrieval, natural language processing, computational linguistics.

Keywords
Automatic text summarization(ATS), automatic text summarization approaches, summarization systems, predictive text summarization.

1. INTRODUCTION
Automatic text summarization (ATS) is a crucial task in today's information-driven world. Text summarization has become increasingly important due to the exponential growth of information available on the Internet[1]. Manually summarizing massive amount of text is quite challenging for humans, which has increased the need for more complex and powerful automatic summarizers. Researchers have been improving automatic text summarization approaches since the 1950s, aiming to create machine-generated summaries that match the quality of human-created summaries[2]. However, despite the progress that has been made in automatic text summarization, there are still challenges that need to be addressed[3]. This paper aims to grab all necessary knowledge regarding automatic text summarization in a limited space. Selection criteria were applied to derive the most reliable open-source, written English, articles published in the last 70 years and contained the phrase "Automatic text summarization" in the title and abstract. These keywords and requirements lead us to select 78 papers including survey and review papers. Review papers were used to derive general information on the ATS systems. Original and single-topic papers also went under review with four research questions: 1) What kinds of approaches and algorithms were used? 2) What are the main findings and what kinds of evaluation metrics are used? 3) Which fields of application and use cases are stated? 4)What kinds of limitations and challenges are mentioned in a paper? All of these answers were structured in the form of a short explanation or table format. It's necessary to point out that electronically available papers mostly lie in the range from 1990'th to 2023. Another specific point of this paper is the outline which is constructed in a way that makes it easy to comprehend for researchers and practitioners using visual representations and plain language. According to an outline of this paper Section 2, discusses the system of text summarization which include the general components of automatic text summarization, its processes, and results. Section 3, presents the review of the 27 open-source papers in terms of domain (12) and language (15) use cases. These papers were derived from Scopus, Science Direct, and Google Scholar within the above-mentioned special period. The section also presents challenges and future directions including the author's proposal for improving and extending ATS systems usability. Section 4, provides the conclusion for conducted survey and future recommendations.

2. AUTOMATIC TEXT SUMMARIZATION SYSTEMS
The concept of automatic text summarization involves the task of condensing a large piece of text into a shorter version while preserving the crucial information and content meaning. The following structure combines the overall structure of text summarization systems[4].

Automatic text summarization systems can be classified based on various criterion [5]. According to current literature, the following flowcharts, tables, and figures are constructed. The paper proposes the classification of ATS based on: text-based, query-based, bidirectional, and knowledge-driven types. These classification criteria help to categorize automatic text summarization systems based on the approach they use.

Text-based text summarization systems analyze the source text itself to generate summaries[1]. They focus on extracting important sentences or phrases from the source text that convey the main information[6].

On the other hand, query-based text summarization systems use user queries to guide the summary generation process[7]. They take into account the specific target information that the user is looking for and generate a summary that is relevant to their query.

Bidirectional text summarization systems consider both the source text and user queries to generate summaries[5]. They aim to provide a summary that not only captures the main information from the source text but also addresses the specific information needs of the user.

Knowledge-driven text summarization[8] systems leverage external knowledge sources, such as ontologies or semantic networks, to enhance the summary generation process. This depends on whether it analyses the source text itself, applies user queries, or bi-directionally use both the source text and user queries or external knowledge sources[8].

Automatic text summarization systems typically consist of components, including input documents, document processing techniques, methods, and additional tools or datasets that improve the overall results. Figure 1 describes the basic components of Text Summarization Systems. These components work together to analyze the input text, identify important sentences, and generate a concise summary.

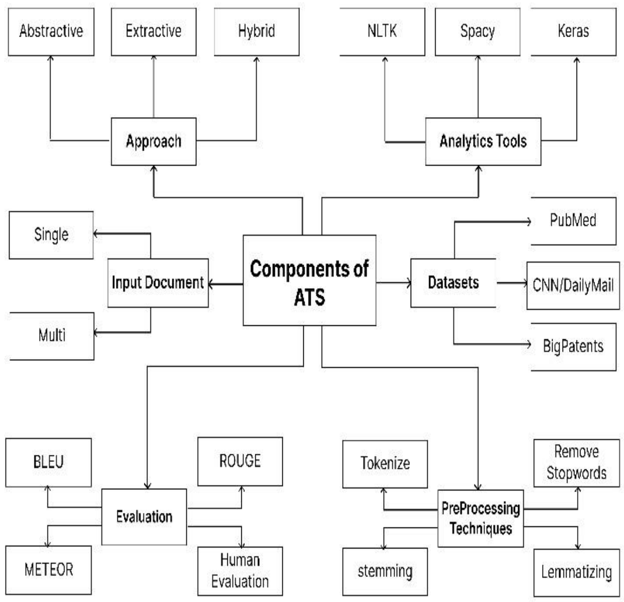

Figure 1. General mind-map of Automatic text summarization components

Different approaches to automatic text summarization can be broadly classified into two categories: extractive and abstractive[8][6][9]. The Extractive summarization approach involves selecting and combining important sentences or phrases from the source text to create a summary[9]. The Abstractive summarization, on the other hand, creates new sentences that capture the essence of the source text[8]. The hybrid approach combines these two methods to create concise, coherent, and informative summaries. However, implementing a hybrid approach requires the integration of different methods and can be computationally intensive[10]. Evaluating the performance of a hybrid system can also be challenging. The main difference in approaches of ATS shown in table 1 along with advantages and challenges.

Table 1. Automatic summarization approaches

	Extractive approach	Abstractive approach	Hybrid approach
Methodology	-Preprocessing -Feature-extraction -Sentence Selection -Post-processing -Evaluation of extractive summary	-Preprocessing -Text Representation-Summary Generation -Post-processing -Evaluation of abstractive summary	-Extraction Stage: -Abstraction Stage -Post-processing -Evaluation of the hybrid summary.
Advantages	-Preserves the original wording and style - Coherence and sequence of the source text. -Requires less computational resources -Less prone to errors - Easier to evaluate.	-Cohesive and human readable summaries -Convey key information in a few sentences - Eliminate ambiguity and redundancy -Generate new content	- Saves important information. - Produce more coherent and readable summaries - The hybrid approach can be adapted to different text types and summarization tasks
Challenges	-Redundancy and repetition -Dependency on sentence scoring -Consistency difficulty	- Content fidelity -Understanding the meaning -Fluency and consistency -Handling ambiguity	-Complexity and integration -Scalability and efficiency -Evaluation and benchmarking -Optimizations and tweaks.

2.1 Datasets for training and evaluation of automatic text summarization

One of the key aspects of developing and evaluating automatic text summarization systems is the availability of suitable datasets[11][8]. Several benchmark datasets such as DUC, TAC, and CNN/Daily Mail have been widely used[8][12] for training and evaluating summarization models. These datasets consist of large corpora of news articles and provide human-generated summaries for comparison and evaluation.

There are several types of datasets exist:

Training data: The dataset serves as a source of labeled training data to train the automatic summarization model. These provide examples of input documents combined with human-generated reference summaries, allowing the model to learn to produce summaries similar to human-generated summaries[13].

Evaluation data: This dataset enables the evaluation of summary models by providing a standardized set of documents and corresponding reference summaries[11][8]. You can test your model against these reference summaries by using evaluation metrics to evaluate your model's performance and compare it to other models.

Benchmarks: Data sets facilitate comparative analysis of aggregate models by providing a common benchmark against which different models can be compared[14][6]. This allows researchers to track progress over time and identify state-of-the-art automatic generalization techniques.

These are widely used dataset in ATS:

Document Understanding Conference (DUC): DUC is one of the first and most widely used benchmark datasets for text summarization[9]. It consists of a collection of news articles and human summaries generated by participants in the DUC summarization task.

Text Analysis Conference (TAC): TAC provides datasets for various text mining tasks such as summarization[9]. Similar to DUC, the TAC dataset contains news articles and related summaries written by people focused on a particular topic or field.

CNN/Daily Mail: The CNN/Daily Mail dataset is derived from online news articles published by CNN and the Daily Mail. It contains a pair of news articles and a summary of several sentences designed to summarize the main points of the article[13].

PubMed/MEDLINE: PubMed/MEDLINE provides commonly used datasets and biomedical literature summaries to summarize problems in the biomedical field. These datasets cover a wide range of medical and healthcare topics[14].

Summary about social networks: Datasets containing social media posts such as tweets and Reddit threads and user-generated summaries are used for social network analysis and sentiment analysis summarization tasks[15].

Big Patent is a new large aggregate dataset containing 1.3 million items of patent documents with human-written summaries[16]. It guides summarization studies to better understand the globality of inputs, and generate summaries with more information and coherent discourse structure.

2.2 Evaluation methods for automatic text summarization

Evaluating the performance of automatic text summarization systems is essential for assessing their quality and effectiveness. Common evaluation metrics include ROUGE, BLEU, and METEOR[17]. These metrics compare the automatically generated summaries with reference human summaries to measure their similarity and coherence. According to[8][9] ROUGE-N Measures the overlap of n-grams (continuous sequences of n words) between the generated summary and the reference summary. Common values for n include 1 (unigram), 2 (bigram), and 3 (trigram). ROUGE-L: Measures the longest common subsequence (LCS) between the generated summary and the reference summary, considering a word sequence. It takes word order into account and rewards summaries that capture the same content in the same order. ROUGE-W: Measures the weighted overlap of N-grams. Longer N-grams are given more weight. This aims to understand the contribution of long phrases to the quality of the summary. BLEU is commonly used in machine translation, but has also been adapted for text summary evaluation. Similar to ROUGE-N, it measures the n-gram overlap between the generated and reference summaries. However, BLEU also penalizes summaries that are too long, preventing the model from producing overly verbose summaries[18]. METEOR evaluates the quality of the generated summaries by considering both exact word matches and semantic similarities. Includes matching stems, synonyms, and paraphrased sentences to more effectively identify semantic similarities between the generated summary and the reference summary.

The evaluation process consists of 4 major components:

Reference summary: Human-generated reference summaries are collected for a specific set of documents[19]. These summaries serve as a basis for evaluating the quality of the summaries produced.

Generated summary: Automatic summarization systems produce summaries of the same set of documents. These generated summaries are compared to reference summaries using evaluation metrics[18].

Calculating metrics: ATS systems use evaluation metrics such as ROUGE, BLEU, and METEOR to quantify the similarity between the generated and reference summaries[14]. A higher score indicates greater similarity, which can improve the precision of the summary.

Analysis and interpretation: The evaluation results are analyzed to evaluate the performance of the aggregation system[20]. Researchers consider which aspects of the summary are well captured and identify areas for improvement.

2.3 Steps in automatic summarization processes

Generally, the Automatic Text Summarization (ATS) process involves several steps. Here is a combined set of the generalized steps involved in the ATS process[12]:

2.3.1 *Preprocessing of information[4]*

Tokenization: Breaking down the text into individual words or tokens. Sentence segmentation: Splitting the text into individual sentences. Remove stop words: Eliminate common words (e.g. "and", "the") that have no significant meaning. Stemming or lemmatization: Reducing words to their base form to normalize the text (for example: "running" to "run".

2.3.2 *Text rendering transforms words into computable parts*

Vectorization: Convert words or sentences into numerical representations (e.g. TF-IDF, word embeddings). Document representation: Create a numerical representation of the entire document, taking into account the importance of each word or phrase.

2.3.3 *Choosing a summarization approach*

Extractive summarization: Select a subset of sentences directly from the original text based on their relevance and importance. Abstractive summarization: Generate new sentences that capture the essence of the original text using natural language generation techniques. Hybrid summarization: Uses both extractive and abstractive summarization techniques. Feature extraction provides a selection of proper words: Sentence scoring: Assign scores to individual sentences based on various features such as word frequency, position, and similarity to other sentences. Importance estimation: Determine the importance of each sentence in conveying the most important information in the text.

2.3.4 *Summarization generation based on previous steps*

Extractive approach: Select the sentences with the highest scores to form the summary. Abstractive approach: Generate new sentences using techniques such as neural networks, language models or rule-based systems, ensuring coherence and fluency[21].

2.3.5 *Post-processing improves readability of the summary*

Length restriction: Limit the length of the summary to a predefined number of sentences or words. Readability enhancement: Refine the summary to improve readability and coherence. Add formatting: Add bullets, headings, or other formatting to improve the structure and presentation of the summary[22].

2.3.6 Evaluation gives a score for similarity with the original text

ROUGE metrics: Compute ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores to assess the quality of the summary by comparing it to reference summaries. Human evaluation: Obtain feedback from human reviewers to assess the summary's informativeness, coherence, and overall quality[23].

3. REVIEW OF PREVIOUS WORKS IN ATS

Various techniques and algorithms have been developed for automatic text summarization. These include statistical approaches, graph-based methods, and deep learning models. Statistical approaches involve the use of frequency-based methods to identify important sentences, while graph-based methods use graph algorithms to rank the sentences based on their connections within the text. A text document can be represented as a graph in different ways. Nodes represent features, and edges represent relationships between nodes [51]. A multi-level relationship between important features of a text document can also be represented by a semi-graph. The multi vertices property of a semi-graph helps to find linear and non-linear connections between features. Application of semi-graph include the processes of semi-graph construction and sentence extraction in summarization [52]. Deep learning models, on the other hand, utilize neural networks to learn the importance of sentences and generate summaries[24]. We have highlighted approaches, algorithms, fields of applications, and evaluation). metrics. Limitations and challenges mentioned in reviewed papers described in a separated subsection.

Table 2. Review of ATS papers in domain specific applications.

Authors	Approach, Techniques & algorithms	Application domain	Evaluation
Ghanem et al., 2023[20].	Hybrid, Pointer-Generator networks and reinforcement learning to optimize the summary	Social media dialogues.	Accuracy, Precision, Harmonic mean, Error
Afsharizadeh et.al, 2021[25].	Extractive, RNN, LSTM and GRU, along with coreference resolution.	COVID-19 related scientific articles	ROUGE1
Albeer et al. 2022 [26].	Extractive, NLP toolkit for preprocessing with TF-IDF approach	YouTube video transcriptions	ROUGE-Family
Chaves et al., 2022[27].	Extractive, Transformer-based methodologies for summarization	Processing clinical notes and biomedical literature	BERTScore
Chowanda et al., 2017[28].	Abstractive, point-based summarization	Online debate forums	ROUGE
Jindal & Kaur, 2020[29].	Extractive, TFIDF with clustering methods	Software bug reports	Recall, F-score, and pyramid precision
Hartl&Kruschwitz, 2022[30].	Hybrid, CMTR-BERT (Contextual Multi-Text Representations)	News, social media platforms fact-checking	Precision, recall, F1-score, or accuracy.
Kim&Yoon, 2022[31].	Abstractive, Generative adversarial network-based summarization	Patent documents	ROUGE, BLEU
Koniaris et al., 2023[32].	Extractive & Abstractive, Fine-tuning of BERT models	Greek legal documents	Fluency, ROUGE family metrics
Rohil&Magotra, 2022[33].	Extractive & Abstractive, T5-SMALL, Tools4Noobs, and Text Compactor	Biomedical and healthcare	Comparison of 3 summarizer tools
Liu et al.,2021[34]	Generative Pre-Training 2.0. Adabound algorithm & Jaya algorithm	Emergency management research, disaster prevention	ROUGE metrics
Zhong&Wang, 2022[36]	Multitask Learning for Multidomain Adaptation summarization which incorporates with BART	Any domain where there is a lack of annotated data.	ROUGE 1

3.1 Domain-specific applications

The application of automatic text summarization extends to various domains, including journalism, research, and information retrieval[13]. In journalism, it is used for creating concise news articles and improving the efficiency of news aggregation platforms. In research, automatic summarization aids in digesting and comprehending large volumes of academic papers and articles. Furthermore, in information retrieval, text summarization facilitates efficient content organization and retrieval in digital libraries and web search engines. Table 2 shows use cases in several areas, as well as commonly used methods, approaches, and algorithms based on selected articles. Evaluation processes are mainly carried out using quantitative methods where human evaluation is overlooked.

3.2 Use cases in different Languages

Automatic text summarization has been applied to different languages, catering to a diverse range of linguistic needs[18]. It has been utilized in multilingual settings to provide summaries in various languages, allowing for wider accessibility and reach. Moreover, with the increasing demand for localization and multilingual content, automatic summarization plays a vital role in overcoming language barriers and enabling effective communication across different language speakers and readers[37]. We derived the important aspects of ATS utilized in different widely used languages.

The paper “Automatic Text Summarization for Urdu Roman Language by Using Fuzzy Logic" was proposed by Ali[38]. Paper’s model for *the Urdu Roman Language* using Fuzzy Logic performs better than previous models. The study compared machine-generated summaries using ROUGE and BLEU Score methods as evaluation metrics.

Mendoza et al.,[39] proposed the paper "Ground Truth *Spanish* Automatic Extractive Text Summarization Bounds". The five state-of-the-art methods and five systems, as well as four heuristics, were used. The concordance heuristic has shown a 66% level of agreement between experts using ROUGE.
Golovizhina and Kotelnikov provide comparisons of Extractive and Abstractive Methods for Russian texts[40]. The algorithms used in the experiments were ruT5-large, mBART, ruT5-base, LexRank, ruGPT3Large, TextRank, and ruGPT3Small. Evaluation scores utilized were ROUGE-N, ROUGE-L, BLEU, METEOR, and BERTScore. Authors stated the fact that most existing work uses only extractive methods, whereas abstractive methods may offer more concise and human-like summaries.
Heidary at al.[41] presented an Extractive summarization approach with feature selection based on text structure analysis in the Persian language. They warned that the statistical methods may produce incoherent and inconsistent summaries due to reliance on statistical features only.
For the Hindi language, Jain et al.[42] used Real Coded Genetic Algorithm. Features such as sentence similarity and Named Entity Recognition are utilized alongside other linguistic parameters. The ATS method showed improvements with a summary reduction of 65% for the Hindi health data corpus. Jovi D'Silva and Uzzal Sharma [43]utilized unsupervised machine learning for the summarization of Konkani Texts using K-means with the Elbow Method. The paper found that summaries created using three clusters were better than those with two clusters.
Most of the research was done in the Indonesian language. Authors Nurul Khotimah et al.,[37] provided a Review of the papers. Their research has shown several papers in automatic text summarization that use *a Statistical approach* using Genetic Algorithm; Algebraic approaches like Latent Semantic Analysis, Non-Negative Matrix Factorization, and Singular Value Decomposition; Term Frequency-Inverse Document Frequency and Vector Space Model are also mentioned as part of the methods used in Indonesian language. Lin et al.,[44] proposed a simple but effective method for Indonesian Automatic text summarization. The proposed method based on the Light Gradient Boosting Machine regression model (LightGBM) was found to be more applicable to Indonesian documents. According to Maylawati et al.,[45] a feature-based approach along with sequential pattern mining methods SumBasic, SentenceScoring enhanced the quality of Indonesian automatic text summarization. Gunawan et al.,[46] have observed the Performance of the TextRank Algorithm on Automatic Text Summarization for Bahasa Indonesia. TextRank algorithm for text summarization in Bahasa Indonesia TextRank is a graph-based model, which does not depend on the language. The study found that there is much room for improvement in text summarization for Bahasa Indonesia as both TextRank and the modified TextRank algorithms did not demonstrate exemplary performance.
For the Chinese language were found several papers including Liu et al.,[34] used ATS for emergency Domain using Generative Pre-Training 2.0., which utilizes the Adabound algorithm for optimization to avoid issues with extreme learning rates and ensure convergence to the global minimum. Huai et al.,[35] propose a new topic-based automatic summarization method specifically for Chinese short text. It combines topic words and TF-IDF to score sentences in the original text data, selecting the highest-scored sentence as the topic sentence. Kuo and Huang[51] showed examples of the Extraction of key-senses/sense-patterns discovery and Key Sentences via Word Sense Identification in the Chinese language. They employed a fuzzy transaction method for sentence representation and measured the summary quality using information-retrieval criteria.
Fejer and Omar[48] utilized combined clustering methods to group documents and applied key-phrase extraction to each cluster. Then several similarity algorithms identified the most important sentences and extracted one sentence from groups of similar sentences. The proposed model aims to improve Arabic text summarization and achieves an accuracy of 43.4%. Tanfouri et al., 2021[49] proposed an automatic Arabic text summarization system based on an extractive approach and genetic algorithms, using the Essex Arabic Summaries Corpus.

3.3 Limitations and challenges

While significant advancements have been made in the field of automatic text summarization, there are still challenges that need to be addressed[50]. Some of the ongoing challenges include improving the coherence and readability of generated summaries, handling noisy and informal text, and addressing the issue of abstractive summarization quality. The 27 papers mentioned above show several specific challenges. We have grouped all of these challenges of ATS regarding summarization techniques, and approaches in the following lists:

3.3.1 *Challenges in domain-specific applications*

Automatic short text summarization (ASTS) methods face inefficiencies in summarizing short text on social media, states Ghanem et al.[20], such as noisy and informal language, abbreviations and acronyms, context understanding, dynamic language and trends, nested conversations and answers. This also includes sentiment and emotion analysis, data volume and real-time processing, data sparsity and imbalanced datasets, ethical and biased content.
Some problems are domain-specific, Chaves et al.,[31] for example, summarizing biomedical texts involves growing and complex requirements in the domain;
Chowanda et al., [28]remark on these challenges of the ATS field which include missing contemporary techniques and insufficient detail for recent studies while remaining a time-consuming process. Quality assurance of meeting summarization and dealing with the various complex layers of spoken dialogue impose growing difficulties;
Jindal & Kaur[29] mentioned about internal threats that might include potential inaccuracies related to splitting paragraphs into sentences and variances in annotation by different annotators while external threats pertain to the reliance on specific corpora for employing the proposed technique. It can be mitigated as the approach does not necessitate a trained model;
The variety and volume of data sources, the need for context-aware analysis, potential biases in the datasets, and the complexity of natural language present another group of difficulties in model performances[30].
The study of Kim&Yoon [31] mentions that Multi Patent Document Summaries (MPDS) may contain incorrect grammar and verb forms, affecting the fluency and suitability of the generated summaries. There are also challenges with evaluation metrics and domain adaptation due to the specialized language of patents. Longer input lengths present computational challenges for neural models.
According to Koniaris et al.,[32]. several challenges defined to text summarization in legal documents due to their complexity, specialized vocabulary, structured format, extensive length, and reference to authoritative texts. The
legally binding nature of their content, emphasizes the need for accurate summaries to avoid significant legal consequences.

There are challenges[33] related to the usability, reliability, and usefulness of automatic summarization systems in clinical settings.
Zhong&Wang[36] stated the need to make the model insensitive to domain-specific features that may interfere with the summarization process, and achieve a balance of performance across different source and target domains;

3.3.2 Challenges of ATS in different languages:

The paper [34]mentions that the Chinese automatic text summarization faces difficulties due to a lack of large-scale, high-quality datasets, and the complexity of accurately capturing semantic information.
Another paper mentions about underdevelopment of methodologies specific to the Russian language[40], as most research focuses on English.
The papers reviewed in[41] mention the difficulty for machines to gain a deep understanding of text content based on syntactic and semantic structure as humans do.
Regarding the challenges the authors of the paper[37] mentioned of dealing with homonyms and polysemes, and the trade-off between summary conciseness and the retention of critical information.
The paper acknowledges the complexity of feature generation for Hindi ATS[42].
The paper[43] concludes that unsupervised learning for text summarization can face challenges related to the accuracy and coherence of generated summaries without language-dependent domain knowledge or training corpora.
The paper[44] mentions limitations such as the scarcity of corpus for abstractive summarization in low-resource languages like Indonesian, which led the researchers to focus on extractive summarization.
The complexities of the Indonesian language in NLP, the robustness of the algorithms, data sparsity, and low computational efficiency posed some challenges and limitations[45].
At the same time, several challenges were mentioned by authors[35], including the complexity and noise inherent in the microblog texts of Weibo, the brevity of texts (limited to 140 words or less), and the difficulty in applying traditional summarization methods, which are more suitable for longer texts, to the concise and often noisy Chinese short text.
The paper notes that Arabic natural language processing[48] suffers from a lack of advanced tools and resources, which hinders the advancement of Arabic text summarization research. Because the studies on Arabic text summarization started much later compared to English, and there has been little research done. Some of the challenges[49] mentioned were processing linguistic nuances, maintaining context, and ensuring the coherence of the produced summaries in arabic texts.

3.4 Predictive summarization as a new trend for overcoming challenges

Future directions in automatic text summarization may involve leveraging multimodal information, integrating domain-specific knowledge, and enhancing the adaptability of summarization systems to dynamic and evolving content sources.
Taking into account all of these problems and limitations, the authors of this article propose a new direction called "Predictive text summarization" shown in figure 2 that may be useful to ATS users. This involves providing possible decision scenarios based on input data. This can be clearly shown as follows.

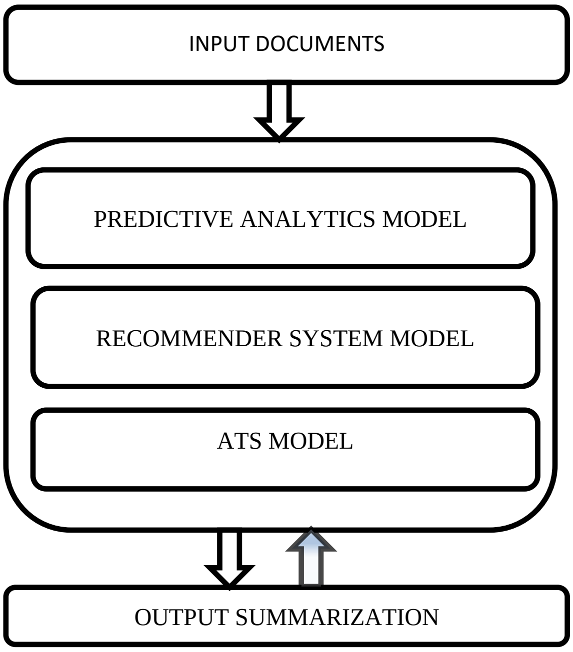

Figure 2. Model of Predictive text summarization system

It consists of both an input document and an output document in the form of a generalized summary version, which is accompanied by an analytics component providing possible scenarios based on the previous experience of the trained models.
These models can be improved for use in various areas of human activity, including business management, public administration, and solving social problems. For example, a young man could create a career plan based on the advantages and disadvantages of the career path, extracted from his resume or queries. Predictive automatic text summarization can provide recommendations from selecting priority professions for youth to preventing epidemic and endemic diseases in specific areas based on weather forecasts. Companies could use this concept to predict human resources attrition, productivity and creativity of the personnel based on resume and cover letters. That would be another challenge for professionals in the Artificial intelligence field.

4. CONCLUSION

The survey on automatic text summarization provides a comprehensive analysis of the concepts, applications of ATS, and challenges in this field. It discusses the components and steps involved in automatic text summarization, as well as different approaches, techniques, and datasets used in this domain. The survey also highlights the importance of evaluation methods for assessing the quality of automatic text summarization systems. The authors analyzed selected 27 papers in terms of their usage of ATS in specific domains or another language than English. Results of the survey show necessities for human evaluation methods for assessing readability and coherence of the text. Derived limitations and challenges were presented. Future directions and trends proposed by the authors conclude the paper. The new term **"Predictive text summarization"** and its draft model were provided. Further studies needed for implementing predictive summarization concept in different areas for increasing decision making processes.

5. REFERENCES

[1] G. Sharma and D. Sharma, "Automatic Text Summarization Methods: A Comprehensive Review," *SN Computer Science*, vol. 4, no. 1, Oct. 2022, doi: 10.1007/s42979-022-01446-w.

[2] Y. J. Kumar, O. S. Goh, H. Basiron, N. H. Choon and P. C. Suppiah. "A Review on Automatic Text Summarization Approaches". *Journal of Computer Science.* vol. 12. no. 4. pp. 178-190. Apr. 2016. 10.3844/jcssp.2016.178.190.

[3] I. Mani, *Automatic summarization*. 2001. doi: 10.1075/nlp.3.

[4] M. Allahyari *et al.* "Text Summarization Techniques: A Brief Survey". *International Journal of Advanced Computer Science* and Applications. vol. 8. no. 10. Jan. 2017. 10.14569/ijacsa.2017.081052.

[5] L. Basyal and M. M. Sanghvi, "Text summarization using large language models: a comparative study of MPT-7B-Instruct, Falcon-7b-Instruct, and OpenAI Chat-GPT models," *arXiv (Cornell University)*, Oct. 2023, doi: 10.48550/arxiv.2310.10449.

[6] A. Rajasekaran and R. Varalakshmi, "Review on automatic text summarization," *International Journal of Engineering & Technology*, vol. 7, no. 3.3, p. 456, Jun. 2018, doi: 10.14419/ijet.v7i2.33.14210.

[7] S. Gupta and S. Gupta. "Abstractive summarization: An overview of the state of the art". *Expert Systems with Applications*. vol. 121. pp. 49-65. May. 2019. 0.1016/j.eswa.2018.12.011.

[8] O. Klymenko, D. A. Braun, and F. Matthes, Automatic Text Summarization: A State-of-the-Art Review. *Proceedings of the 22nd International Conference on Enterprise Information Systems - (Volume 1)* May 5-7, 2020 648-655,2020. doi: 10.5220/0009723306480655.

[9] D. Yadav, J. Desai, and A. K. Yadav, "Automatic Text Summarization Methods: A Comprehensive Review," *arXiv (Cornell University)*, Mar. 2022, doi:10.48550/arxiv.2204.01849.

[10] B. Khan, Z. Shah, M. Usman, I. B. Khan, and B. Niazi, "Exploring the Landscape of Automatic Text Summarization: A Comprehensive survey," *IEEE Access*, vol. 11, pp. 109819–109840, Jan. 2023, doi: 10.1109/access.2023.3322188.

[11] W. H. Alquliti and N. Binti, "Convolutional Neural Network based for Automatic Text Summarization," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 4, Jan. 2019, doi: 10.14569/ijacsa.2019.0100424.

[12] F. M. Alliheibi, A. Omar, and N. Al-Horais, "An evaluation of automatic text summarization of news articles: the case of three online Arabic text summary generators," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, Jan. 2021, doi: 10.14569/ijacsa.2021.0120513.

[13] A. P. Widyassari *et al.*, "Review of automatic text summarization techniques & methods," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, pp. 1029–1046, Apr. 2022, doi: 10.1016/j.jksuci.2020.05.006

[14] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan, and Md. M. Kabir, "A survey of automatic text Summarization: progress, process and challenges," *IEEE Access*, vol. 9, pp. 156043–156070, Jan. 2021, doi: 10.1109/access.2021.3129786.

[15] I. S. Blekanov, N. Tarasov and S. S. Bodrunova. "Transformer-Based Abstractive Summarization for Reddit and Twitter: Single Posts vs. Comment Pools in Three Languages". *Multidisciplinary Digital Publishing Institute.* vol. 14. no. 3. pp. 69-69. Feb. 2022. https://doi.org/10.3390/fi14030069.

[16] E. Sharma, C. Li and L. Wang. "BIGPATENT: A Large-Scale Dataset for Abstractive and Coherent Summarization". *Cornell University*. Jun. 2019. https://doi.org/10.48550/arxiv.1906.03741.

[17] K. Jacobs, "Automatic text Summarization: Overview and challenges," *Data Blogger*, Feb. 16, 2024. https://www.data-blogger.com/automatic-text-summarization-overview-and-challenges/

[18] M. M. Haque, S. Pervin and Z. Begum. "Literature Review of Automatic Multiple Documents Text Summarization". *International Journal of Innovation and Applied Studies.* vol. 3. no. 1. pp. 121-129. May. 2013.

[19] N. Stiennon et al. "Learning to summarize from human feedback". *Cornell University*. Sep. 2020. https://doi.org/10.48550/arxiv.2009.01325.

[20] F. A. Ghanem, M. C. Padma and R. Alkhatib. "Automatic Short Text Summarization Techniques in Social Media Platforms". *Multidisciplinary Digital Publishing Institute*. vol. 15. no. 9. pp. 311-311. Sep. 2023. https://doi.org/10.3390/fi15090311.

[21] A. Aries, D. E. Zegour and W. K. Hidouci. "Automatic text summarization: What has been done and what has to be done". *arXiv.org.* Apr. 2019. 10.48550/arXiv.1904.00688.

[22] D. O. Cajueiro *et al.*, "A comprehensive review of automatic text summarization techniques: method, data, evaluation and coding," *arXiv.org*, Jan. 04, 2023. https://arxiv.org/abs/2301.03403

[23] P. Lemberger. "Deep Learning Models for Automatic Summarization". *arXiv (Cornell University).* May. 2020. 10.48550/arXiv.2005.11988.

[24] D. Tsirmpas, I. Gkionis and I. Mademlis. "Neural Natural Language Processing for Long Texts: A Survey of the State-of-the-Art". *arXiv (Cornell University).* May. 2023. 10.48550/arxiv.2305.16259.

[25] M. Afsharizadeh, H. Ebrahimpour-Komleh and A. Bagheri. "Automatic Text Summarization of COVID-19 Research Articles Using Recurrent Neural Networks and Coreference Resolution". *Knowledge E.* Feb. 2021. https://doi.org/10.18502/fbt.v7i4.5321.

[26] R.A.Albeer, H. F. AL-Shahad,H J. Aleqabie and N. D. Al-shakarchy. "Automatic summarization of YouTube video transcription text using term frequency-inverse document frequency". *Institute of Advanced Engineering and Science(IAES).*vol.26.no.3.pp.1512-1512.Jun. 2022 https://doi.org/10.11591/ijeecs.v26.i3.pp1512-1519.

[27] A. Chaves-Villota, C. Y. Kesiku and B. Garcia-Zapirain. "Automatic Text Summarization of Biomedical Text Data: A Systematic Review". Multidisciplinary Digital

Publishing Institute. vol. 13. no. 8. pp. 393-393. Aug. 2022. https://doi.org/10.3390/info13080393.

[28] A. D. Chowanda, A. R. Sanyoto, D. Suhartono and C. J. Setiadi. "Automatic Debate Text Summarization in Online Debate Forum". Elsevier BV. vol. 116. pp. 11-19. Jan. 2017. https://doi.org/10.1016/j.procs.2017.10.003.

[29] S. G. Jindal and A. Kaur. "Automatic Keyword and Sentence-Based Text Summarization for Software Bug Reports". Institute of Electrical and Electronics Engineers. vol. 8. pp. 65352-65370. Jan. 2020. https://doi.org/10.1109/access.2020.2985222.

[30] P. Hartl and U. Kruschwitz. "Applying Automatic Text Summarization for Fake News Detection". Cornell University. Apr. 2022. https://doi.org/10.48550/arxiv.2204.01841.

[31] S. Kim and B. Yoon. "Multi-document summarization for patent documents based on generative adversarial network". Elsevier BV. vol. 207. pp. 117983-117983. Nov. 2022. https://doi.org/10.1016/j.eswa.2022.117983.

[32] M. Koniaris, D. Galanis, E. Giannini and P. Tsanakas. "Evaluation of Automatic Legal Text Summarization Techniques for Greek Case Law". Multidisciplinary Digital Publishing Institute. vol. 14. no. 4. pp. 250-250. Apr. 2023. https://doi.org/10.3390/info14040250.

[33] M. K. Rohil and V. Magotra. "An exploratory study of automatic text summarization in biomedical and healthcare domain". Elsevier BV. vol. 2. pp. 100058-100058. Nov. 2022. https://doi.org/10.1016/j.health.2022.100058.

[34] M. Liu, Z. Wang and L. Wang. "Automatic Chinese Text Summarization for Emergency Domain". IOP Publishing. vol. 1754. no. 1. pp. 012213-012213. Feb. 2021. https://doi.org/10.1088/1742-6596/1754/1/012213.

[35] T. Huai, H. M. Wang, Y. Zhao, Y. Tian and N. Al-Nabhan. "Topic-based automatic summarization algorithm for Chinese short text". Arizona State University. vol. 17. no. 4. pp. 3582-3600. Jan. 2020. https://doi.org/10.3934/mbe.2020202.

[36] J. Zhong and Z. Wang. "MTL-DAS: Automatic Text Summarization for Domain Adaptation". Hindawi Publishing Corporation. vol. 2022. pp. 1-10. Jun. 2022. https://doi.org/10.1155/2022/4851828.

[37] N. Khotimah, A. W. P, B. Andreas and A. S. Girsang. "A Review Paper on Automatic Text Summarization in Indonesia Language". vol. 11. no. 8. pp. 89-96. Aug. 2021. https://doi.org/10.46338/ijetae0821_11

[38] Z. A. Ali. "Automatic Text Summarization for Urdu Roman Language by Using Fuzzy Logic". vol. 3. no. 2. pp. 23-23. Aug. 2021. https://doi.org/10.32629/jai.v3i2.273.

[39] G. A. M. Mendoza, Y. Ledeneva, R. Hernández, M. Alexandrov and Á. Hernández-Castañeda. "Ground Truth Spanish Automatic Extractive Text Summarization Bounds". National Polytechnic Institute. vol. 24. no. 3. Sep. 2020. https://doi.org/10.13053/cys-24-3-3484.

[40] V. Goloviznina and E. Kotelnikov. "Automatic Summarization of Russian Texts: Comparison of Extractive and Abstractive Methods". Jun. 2022. https://doi.org/10.28995/2075-7182-2022-21-223-235.

[41] E. Heidary, H. Parv 飆, S. Nejatian, K. Bagherifard and V. Rezaie. "Automatic Persian Text Summarization Using Linguistic Features from Text Structure Analysis". *Tech Science Press.* vol. 69. no. 3. pp. 2845-2861. Jan. 2021. https://doi.org/10.32604/cmc.2021.014361.

[42] A. Jain, A. Arora, J. Morato, D. Yadav and K. V. Kumar. "Automatic Text Summarization for Hindi Using Real Coded Genetic Algorithm". *Multidisciplinary Digital Publishing Institute.* vol. 12. no. 13. pp. 6584-6584. Jun. 2022. https://doi.org/10.3390/app12136584.

[43] J. D'Silva and U. Sharma. "Unsupervised Automatic Text Summarization of Konkani Texts using K-means with Elbow Method". *International Journal of Engineering Research and Technology.* vol. 13. no. 9. pp. 2380-2380. Sep. 2020. https://doi.org/10.37624/ijert/13.9.2020.2380-2384.

[44] N. Lin, J. Li and S. Jiang. "A simple but effective method for Indonesian automatic text summarization". *Taylor & Francis.* vol. 34. no. 1. pp. 29-43. Jun. 2021. https://doi.org/10.1080/09540091.2021.1937942.

[45] D. S. Maylawati, Y. J. Kumar and F. Kasmin. "Feature-based approach and sequential pattern mining to enhance quality of Indonesian automatic text summarization". *Institute of Advanced Engineering and Science (IAES*). vol. 30. no. 3. pp. 1795-1795. Jun. 2023. https://doi.org/10.11591/ijeecs.v30.i3.pp1795-1804.

[46] D. Gunawan, D. Witarsyah, D. Syamsuar, A. Amalia, -. Abdurrohman and R. F. Rahmat. "Observing the Performance of the TextRank Algorithm on Automatic Text Summarization for Bahasa Indonesia". *Insight Society.* vol. 13. no. 3. pp. 1147-1147. Jun. 2023. https://doi.org/10.18517/ijaseit.13.3.14988.

[47] Y. Kuo and H. Huang. "Automatic Extraction of Key Sentences via Word Sense Identification for Chinese Text Summarization". *Fuji Technology Press Ltd*. vol. 11. no. 4. pp. 416-422. Apr. 2007. https://doi.org/10.20965/jaciii.2007.p0416.

[48] H. N. Fejer and N. Omar. "Automatic Multi-Document Arabic Text Summarization Using Clustering and Keyphrase Extraction". *Journal of Artificial Intelligence* vol. 8. no. 1. pp. 1-9. Dec. 2014. https://doi.org/10.3923/jai.2015.1.9.

[49] I. Tanfouri, G. Tlik and F. Jarray. "An automatic arabic text summarization system based on genetic algorithms". *Elsevier BV.* vol. 189. pp. 195-202. Jan. 2021. https://doi.org/10.1016/j.procs.2021.05.083.

[50] S. H. B. Sri and S. R. Dutta. "A Survey on Automatic Text Summarization Techniques". *Journal of Physics: Conference Series*. vol. 2040. no. 1. pp. 012044-012044. Oct. 2021. 10.1088/1742-6596/2040/1/012044.

[51] S. Sonawane and P. A. Kulkarni, "Graph based representation and analysis of text document: A survey of techniques," *International Journal of Computer Applications*, vol. 96, no. 19, pp. 1–8, Jun. 2014. doi:10.5120/16899-6972

[52] S. Sonawane, P. Kulkarni, C. Deshpande, and B. Athawale, "Extractive summarization using semi-graph (ESSG)," *Evolving Systems*, vol. 10, no. 3, pp. 409–424, Jul. 2018. doi:10.1007/s12530-018-9246-8