Research Article/Araştırma Makalesi

GAZİ
JOURNAL OF ENGINEERING SCIENCES

Performance Evaluation of the Extractive Methods in Automatic Text Summarization Using Medical Papers

Anil Kus[*a], Cigdem Inan Aci[b]

Submitted: 16.11.2023 Revised: 19.12.2023 Accepted: 19.12.2023 doi:10.30855/gmbd.0705S02

Keywords: Automatic text summarization, extractive method, scientific papers, health sciences

[a,*] Toros University, Dept. of Computer Technologies, 33140 - Mersin, Türkiye
Orcid: 0000-0002-5964-3727
e mail: anil.kus@toros.edu.tr

[b] Mersin University, Engineering Faculty, Dept. of Computer Engineering 33343 - Mersin, Türkiye
Orcid: 0000-0002-0028-9890

[*]Corresponding author: anil.kus@toros.edu.tr

ABSTRACT

The rapid development of technology has resulted in a surge in the volume of digital data available. This situation creates a problem for users who need assistance in locating specific information within this massive collection of data, resulting in a time-consuming process. Automatic Text Summarization systems have been developed as a more effective solution than traditional summarization techniques to address this problem and improve user access to relevant information. It is well known that researchers in the health sciences find it difficult to keep up with the latest literature due to their busy schedules. This study aims to produce comprehensive abstracts of Turkish-language scientific papers in the field of health sciences. Although abstracts of scientific papers are already available, more thorough summaries are still needed. To the best of our knowledge, no previous attempt has been made to automatically summarize Turkish language health science papers. For this purpose, a dataset of 105 Turkish papers was collected from DergiPark. Term Frequency, Term Frequency-Inverse Document Frequency, Latent Semantic Analysis, TextRank, and Latent Dirichlet Allocation algorithms were chosen as extractive text summarization methods due to their frequent use in this field. The performance of the text summarization models was evaluated using recall, precision, and F-score metrics, and the algorithms gave satisfactory results for Turkish.

Çıkarımsal Otomatik Metin Özetleme Yöntemlerinin Tıp Makaleleri Kullanılarak Performans Değerlendirmesi

ÖZ

Teknolojinin hızlı gelişimi, mevcut dijital veri hacminin artmasına neden olmuştur. Bu durum, bu devasa veri koleksiyonu içinde belirli bilgilerin yerini bulma konusunda yardıma ihtiyaç duyan kullanıcılar için sorun yaratarak zaman alıcı bir süreçle sonuçlanır. Otomatik Metin Özetleme sistemleri, bu sorunu gidermek ve kullanıcının ilgili bilgiye erişimini iyileştirmek için geleneksel özetleme tekniklerinden daha etkili bir çözüm olarak geliştirilmiştir. Sağlık bilimleri alanındaki araştırmacıların yoğun programları nedeniyle güncel literatürü takip etmekte zorlandıkları bilinmektedir. Bu çalışma, sağlık bilimleri alanında Türkçe dilindeki bilimsel makalelerin kapsamlı özetlerini üretmeyi amaçlamaktadır. Bilimsel makalelerin özetleri hâlihazırda mevcut olmasına rağmen, daha kapsamlı özetlere hâlâ ihtiyaç vardır. Bildiğimiz kadarıyla, Türkçe sağlık bilimi makalelerini otomatik olarak özetlemek için daha önce herhangi bir girişimde bulunulmamıştır. Bu amaçla, DergiPark'tan 105 Türkçe makaleden oluşan bir veri kümesi toplanmıştır. Çıkarımsal metin özetleme yöntemleri olarak Terim Frekansı, Terim Frekansı-Tersine Doküman Frekansı, Gizli Anlamsal Analiz, TextRank ve Gizli Dirichlet Ayırımı algoritmaları bu alanda sıklıkla kullanılmaları nedeniyle seçilmiştir. Metin özetleme modellerinin başarımı duyarlılık, kesinlik ve F-skor metrikleri kullanılarak değerlendirilmiş ve algoritmalar Türkçe için tatmin edici sonuçlar vermiştir.

Anahtar Kelimeler: Otomatik metin özetleme, çıkarımsal metot, bilimsel makaleler, sağlık bilimleri

To cite this article: A. Kus and C.I. Aci, "Performance Evaluation of the Extractive Methods in Automatic Text Summarization Using Medical Papers," *Gazi Journal of Engineering Sciences,* vol.9, no.4, pp. 14-22, 2023. doi:10.30855/gmbd.0705S02

1. Introduction

The health sciences are facing a rapid growth of information, and there is an urgent need for effective methods to extract useful information from massive amounts of textual data. Many important papers have been published in Turkish scientific journals, which have contributed to medical research and healthcare procedures. However, manually reviewing and analyzing these papers takes a lot of time and effort [1]. Therefore, it is essential to develop Automatic Text Summarization (ATS) approaches that are specifically adapted to Turkish medical publications to facilitate the efficient extraction of critical information and accelerate the dissemination of knowledge. In this paper, we present an approach to automatically summarize health science papers to provide researchers, healthcare professionals, and other stakeholders with concise and informative summaries that capture the essential aspects of the original texts. Using Natural Language Processing (NLP) and machine learning techniques, our proposed method aims to improve information retrieval, increase research efficiency, and contribute to advances in the medical field.

There are two main methods used in summarization tasks: extractive and abstractive: Extractive summarization systems produce summaries by directly selecting and combining the most important sentences from the source documents [2]. Abstractive summarization systems, on the other hand, aim to create new words and phrases that are not present in the original documents, using text rewriting operations such as substitution and reordering [3]. The extractive summarization method evaluates the sentences in the text and takes into account the factors that determine the importance of these sentences. For example, factors such as the frequency of a sentence, its inclusion of keywords, its relationship to other sentences, or its relationship to other important elements in the text are evaluated. Using this information, the most important sentences or paragraphs are selected and a summary is created from these selections. The extractive method uses the words in the text you present to summarize. The main difference with abstractive summarizing is that the summary is limited to the words in the original text. When the previous studies were examined, it was found that the success rate of summaries made using extractive methods was more successful than summaries made using abstractive methods [4]. Turkish is an agglutinative language belonging to the Ural-Altaic family. Therefore, some difficulties arise in the field of NLP due to these characteristics. Each suffix added to a word causes the creation of new words with different meanings [5]. Although these limitations have been partially overcome with the increasing number of recent studies, the progress made in English-based research has not been fully replicated. In this regard, the advantages offered by the structural simplicity of the English language and its wide range of applications have played an important role [4].

The motivation of this study is the need to quickly retrieve accurate information in the health field, given the time constraints of academics. While abstracts of scientific papers are currently accessible, there remains a demand for more comprehensive summaries. In addition, there is a lack of research on ATS in the Turkish language. This study aims to develop an ATS system for Turkish scientific publications in the field of health, using algorithms that include extractive summarization methods in Turkish. The proposed system includes a total of five different extractive methods, including four different statistics-based algorithms and one graph-based algorithm. The reason for choosing these methods is to observe the possible effects of summarization algorithms on Turkish, to compare the summarization success of the methods used, and also to determine the success of ATS systems in academic publications written in the field of health.

The contributions of this study can be listed as follows:

- An original dataset of 105 medical papers was collected from well-known Turkish health journals in Dergipark [6].
- Extractive methods such as Term Frequency (TF), Term Frequency-Inverse Document Frequency (TF-IDF), Latent Semantic Analysis (LSA), TextRank, and Latent Dirichlet Allocation (LDA) algorithms were used as ATS methods.
- The expanded summaries showed promising results close to human-generated summaries. It is believed that this study will contribute to Turkish NLP studies in terms of dataset and application.

The rest of the paper is organized as follows: Section 2 presents the previous studies on ATS. Section 3 presents the details of materials and methods. Section 4 presents the experimental results and discussions. The conclusion is given in Section 5.

2. Literature Review

The first text summarization studies were carried out in the 1950s, and many of the terms used today were first published in these early studies. The concept of "term frequency (TF)" is still the most important part of today's text summarization studies and is considered the basis of the algorithms used [7].

Turkish ATS studies were conducted after the 2000s. In his dissertation, Güran [8] created 2 new Turkish datasets for this purpose. The first dataset includes 130 news documents, which were collected from different news websites, and 130 summary documents, which were extracted by 3 different people who read these documents. The other dataset contains 20 news documents, which are shorter than the first dataset, and 20 summary documents of these documents produced by 30 different people. In the study by Kaynar et al. [9], text summarization methods based on sentence extraction were discussed and an attempt was made to determine the effectiveness of attributes in producing summaries using a genetic algorithm. The dataset used in the study includes Turkish news texts, consisting of 120 documents, and summaries of these news. In the study of Torun and İnner [10], 12,000 Turkish news items were used for the ATS. Using inference-based methods, the texts were divided into sentences, and abbreviations were not taken into account. The news items in the dataset are summarized in five sentences. Karcıoğlu and Yaşa [11] presented a method that uses genetic algorithms to generate ATS models. Sentence selection based on genetic algorithms was used for summarization, after which the summary was created and evaluated using a fitness function.

Table 1 summarizes the Turkish ATS studies developed using extractive methods and their performance results.

Table 1. Previous studies on Turkish ATS using extractive methods

Author	**Year**	**Metrics**	**The best results**
Güran [8]	2013	F-score, ROUGE	0.59 / 0.65
Kaynar et al. [9]	2017	ROUGE	0.72
Torun and İnner [10]	2018	ROUGE	0.68
Karcıoğlu and Yaşa [11]	2020	F-score, ROUGE	0.56 / 0.57

3. Material and Methods

This section explains the ATS architecture used in the study, the dataset prepared, and the summarization algorithms used.

3.1. The extractive ATS model

The extractive method aims to summarize the main points and the main idea by understanding the meaning of a text. When summarizing using the extractive method, information is extracted directly from the text and used to create a shorter summary. The basis of the extractive method is based on analyzing the structure of the text, identifying key words and recognizing important sentences or paragraphs. The algorithm evaluates the sentences in the text and takes into account the factors that determine the importance of these sentences. Figure 1 shows the components of the extractive ATS architecture used in this study.

In our study, we have used five different algorithms that are considered to be among the most widely used extractive methods. All algorithms used the pre-processed version of the texts to be summarized. Since extractive-based methods use the words of the original text, it was observed that the pre-processing stage is very important and has a direct impact on the success of the summary. As can be seen in Figure 1, the created data set goes through normalization and pre-processing stages, respectively. Texts translated into a certain format and suitable for summarization are fed into the appropriate algorithm to obtain word scores and sentence scores based on these scores. Finally, a summary output is generated based on the analysis of these scores, depending on the characteristics of the algorithm.

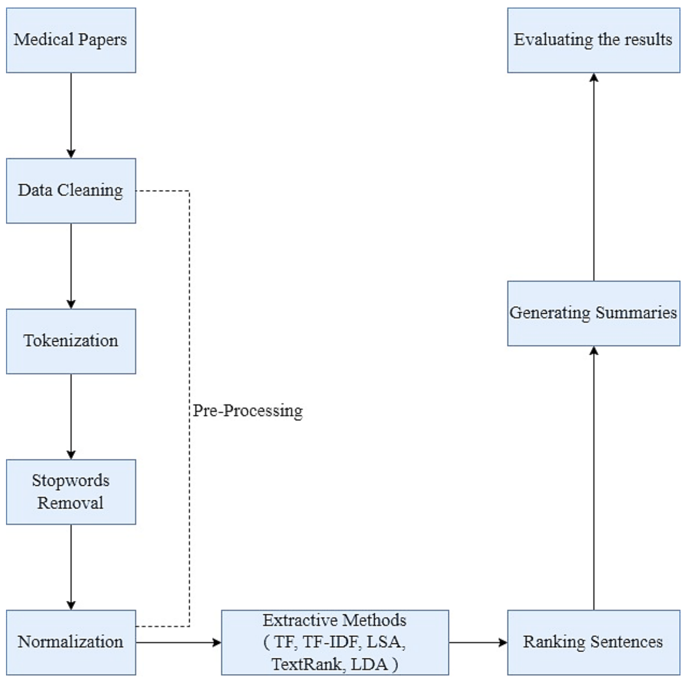

Figure 1. The architecture of the extractive ATS model

3.2. The dataset

ATS is a field where both extractive and abstractive methods are used, and it is well known that the dataset is critical to these methods. The dataset plays an important role in training and determining the performance of summarization models. However, it is a fact that Turkish datasets for ATS studies are limited. Qualified Turkish datasets obtained from local sources will contribute to significant progress in the field of ATS by providing a more accurate and effective summary of Turkish texts.

In this study, a dataset consisting of 105 medical research articles, reviews, and case studies was created for automatic summarizing. The dataset was created using the Dergipark [6] website, which is one of Turkey's academic research and publishing platforms and also provides an online publishing infrastructure supported by universities and other academic institutions. The papers used in the dataset were selected from the medical journals of Ankara University [12], Çukurova University [13], Ege University [14], Gazi University [15], Hacettepe University [16], Istanbul University [17], Mersin University [18] and Samsun University [19]. The selected papers were conducted in the last 5 years in the fields of gynecological diseases, brain diseases, eye diseases, heart diseases, and pediatric diseases. The reason these journals were selected for the dataset is that they are published by reputable medical faculties in Turkey.

Among the 105 articles in the dataset, the shortest document contains 6531 characters, the longest document contains 30785 characters, and the average document length is 17465.10 characters. The shortest of the original summaries is 636 characters, the longest is 2321 characters, and the average is 1501.58 characters. The shortest of the summaries produced is 1247 characters, the longest is 8137 characters, and the average is 2862.53 characters.

3.2. Pre-processing

Pre-processing steps are used in a variety of text mining applications, including text categorization and ATS; this step is essential for ATS [20]. Unwanted characters and incorrect data ordering are often encountered during the data collection process, which can lead to unwanted problems [21]. During the pre-processing phase of this study, the parts of the dataset were reviewed to determine their relevance to the abstract. As a result, the papers were initially split into two parts. The dataset was stripped of those elements that wouldn't be covered by the abstract, including the bibliography, English language abstract, author, and journal information. The original abstracts in Turkish, which were used to evaluate the results by comparing them with those of the ATS, were included in the second section.

The first step is data cleaning, which removes certain characters, such as numerals and punctuation, which are often considered non-essential. The second step is tokenization. In this step, a text corpus or sentence is broken down into smaller elements, such as words, phrases, or n-grams. Tokenization makes the dataset more manageable, adaptable, and amenable to analysis. It is essential for improving the accuracy of NLP modeling and analysis by making the linguistic and semantic structures of the data more visible. In order to implement ATS algorithms, it is crucial to remove stop words, which are high-frequency functional words, from a text corpus. In Turkish, words such as "ile", "şu", "fakat", "yani", etc. are considered irrelevant for certain NLP tasks. In our study, the Natural Language Toolkit (NLTK) was used to remove all stop words.

The last step is normalization for the pre-processing stage. All texts were converted to lower case, words were converted to their root form and unnecessary characters such as white spaces and short lines were removed.

3.3. Term Frequency (TF)

TF is a key concept in the extractive summarization approach. In this approach, the first step is to separate all words based on their root forms and organize them in tables based on their occurrence in the text. The second step is to calculate the frequency of each word, excluding common stop words [22]. These word frequency scores are then divided by the maximum frequency, which represents the overall score of the document. This result indicates the relative importance of the words. Next, sentence scores are calculated based on the length of the sentences. Finally, words with high scores are included in the summarization section by constructing sentences around them.

3.4. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF is a technique commonly used in NLP and information retrieval to evaluate the importance of words in a document. Unlike the TF method, which only considers the frequency of a term within a document, TF-IDF considers both the TF and its IDF [7]. In the TF-IDF approach, the TF still represents how often a term occurs within a given document. However, it is further normalized by dividing it by the total number of terms in the document [23]. It is calculated by taking the logarithm of the ratio between the total number of documents in the collection and the number of documents containing the specific term. The IDF value increases for terms that appear in fewer documents, indicating their potential importance.

3.5. Latent Semantic Analysis (LSA)

LSA is another common and powerful approach to summarizing text using the extractive method. LSA works by creating a mathematical model that represents the relationships between words and documents in a corpus. It uses a dimensionality reduction technique, such as singular value decomposition, to identify the underlying latent semantic structure. By capturing the co-occurrence patterns of words across multiple documents, LSA can reveal the semantic similarity and relatedness between different terms [24].

3.6. TextRank

TextRank is an extractive method used in NLP to identify important words in a text. It represents the text as a graph, with each word or sentence represented as a node and their relationships captured by connections. TextRank calculates importance scores based on these connections and identifies significant keywords within the text [25]. We used some parameters such as iterations and a damping factor to increase the accuracy of the summarizations. A range of 50 to 100 iterations is considered for high performance by determining how many times the algorithm updates the importance scores. A common value, 0.85, is used for the damping factor, which allows us to control the transition probabilities between word or sentence nodes.

3.7. Latent Dirichlet Allocation (LDA)

LDA is a generative statistical model used for topic modeling in NLP. It automatically discovers and extracts the underlying thematic structure in a collection of documents, providing insights into large text corpora [26]. The length of the summary is a crucial parameter as it determines the number of most representative words or phrases to be selected from each topic to form the summary. The study determines the optimal number of words based on the desired level of conciseness and information retention.

4. Results and Discussion

This section presents the study's performance metrics, development environment, and the performance results of each extractive method.

4.1. Performance metrics

Evaluating the effectiveness of ATS systems is crucial, but challenging due to the subjective nature of summarization quality. Analytical evaluation is difficult because of this subjective nature. This study used ROUGE (Recall-Oriented Understudy for Gisting Evaluation) and F-score evaluation metrics, which are commonly used in ATS studies. We chose these metrics to ensure comparability with previous studies, as they are commonly used in the literature [27].

The ROUGE family of metrics is commonly used to evaluate the quality of summarization. ROUGE-N measures similarity based on a given gram size N. For instance, ROUGE-1 assesses single-word summarization, while ROUGE-2 assesses double-word summarization. ROUGE-L is a similarity measure based on the longest common subsequence [28]. The ROUGE metrics evaluate the similarity between the abstract and the reference abstract. The study requires extended summaries, so the maximum value for n was set to 3. The formula for ROUGE-N is given by Eq. (1).

$$ROUGE - N = \frac{Count\ of\ matching\ n{-}grams\ in\ generated\ summary}{Count\ of\ n{-}grams\ in\ reference\ summary} \tag{1}$$

The study presented ATS models, and their accuracy was evaluated using recall and precision values. The F-score value was obtained accordingly to evaluate the similarity between system-generated summaries and actual summaries. The precision determines the accuracy level between requested documents and system-generated answers. Precision calculation is done using the formula:

$$Precision = True\ Positives\ /\ (True\ Positives + False\ Positives) \tag{2}$$

Recall is used for documents that are relevant to the user's query and are recalled by the system. Recall score is calculated using the formula:

$$Recall = True\ Positives\ /\ (True\ Positives + True\ Negatives) \tag{3}$$

F-score is obtained by combining precision and recall. To calculate the F-score, first calculate the average precision and recall using the following formula:

$$F - score = (Recall * Precision)\ /\ (Precision + Recall) \tag{4}$$

All of the Extractive ATS models have been developed in Python/Anaconda using the Zemberek [29] and SpaCy [30] libraries.

4.2. Performance results

A total of 105 different papers published in the field of health were used in this study to obtain extended automatic summaries. Table 2 shows the recall and precision values, Figure 2 shows the F-score values and Figure 3 shows the ROUGE-3 results obtained from the TF, TF-IDF, TextRank, LSA, and LDA algorithms used in our study.

Looking at the results, the most successful summary produced by the TF method achieved an F-score of 0.52, indicating an average level of success compared to previous studies. While the results of the TF-IDF method had an F-score of 0.38, the LSA approach produced an F-score of 0.46. It was found that the maximum achievable F-score was 0.51 for summaries generated by the TextRank algorithm. The summaries produced by LDA had an F-score of 0.55, giving them the highest performance score in the study.

Table 2. Precision and Recall values of the ATS algorithms

Algorithm	Precision	Recall
TF	0.55	0.50
TF-IDF	0.52	0.30
LSA	0.53	0.41
TextRank	0.60	0.44
LDA	0.87	0.40

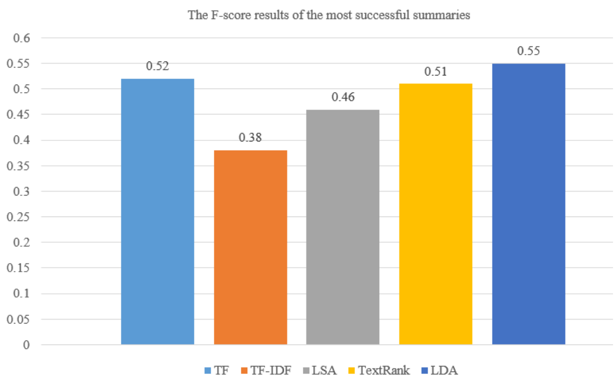

Figure 2. F-score values of the extractive ATS methods

The comparison showed that LDA emerged as the most successful approach. The reason for the superior performance of LDA can be attributed to its ability to capture the underlying topics in a collection of documents. By using the probabilistic topic modelling approach, LDA effectively identifies the main themes and concepts present in the text, allowing for more accurate and coherent summarization [24]. LDA considers the distribution of words across topics, which helps to extract important and representative phrases or sentences for summarization. This ensures that the generated summaries are more informative and concise. Furthermore, LDA is capable of handling large and diverse datasets, making it suitable for different document types and domains [31].

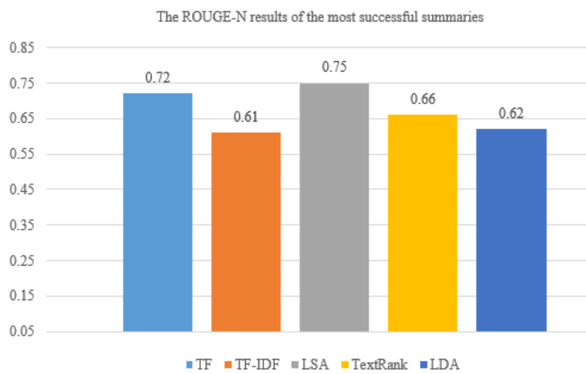

Figure 3. ROUGE-3 values of the extractive ATS methods

4. Conclusions

As part of this research, a system was developed to automatically collect and summarize health-related texts. The system was used to summarize a total of 105 different papers. Performance analysis of the ATS system using five different extraction methods. To the best of our knowledge, there is no study in the literature that systematically collates and summarizes existing work in the field of health. It is widely recognized in the healthcare field that a comprehensive summary of existing studies would be extremely valuable to professionals, providing them with a comprehensive understanding of the subject and aiding their work.

The experiments showed that the most frequent words and keywords did not always accurately reflect the subject matter of the document. It was observed that titles or keywords improved the results and better represented the documents. In addition, due to the limited studies on Turkish ATS, the dataset used in this research will contribute to researchers studying the Turkish language. The results showed that the summaries generated by the system were similar to the original texts. The LDA method showed an improved performance of the ATS. However, the LSA and TF-IDF methods produced less successful results compared to the LDA method.

When analyzing the results of our summarization, we found that the summaries can be considered successful. However, due to the calculation of success rates based on the original summary, they may appear less successful than they actually are. The subjective nature of summarization is considered to be one of the most challenging aspects of this field. In addition, medical papers are particularly difficult to summarize due to their rich content of technical terms, analysis results, and mathematical and formulaic expressions.

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

References

[1] J. P. Andersen, M. W. Nielsen, N. L. Simone, R. E. Lewiss, and R. Jagsi, "COVID-19 medical papers have fewer women first authors than expected," eLife, vol. 9, June 2020. doi:10.7554/elife.58807

[2] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," *arxiv.org, April 2017* [Online]. Available: arXiv, https://arxiv.org/abs/1704.04368. [Accessed: 15 Dec. 2023].

[3} S. Narayan, S. B. Cohen, and M. Lapata, "Ranking sentences for extractive summarization with reinforcement learning," *arxiv.org, April 2018* [Online]. Available: arXiv, https://arxiv.org/abs/1802.08636. [Accessed: 15 Dec. 2023].

[4] E. Erdağı, "Extractive based automatic text summarization in Turkish texts," Ph.D. dissertation, Maltepe University, İstanbul, Turkey, 2023.

[5] Ö. E. Gündoğdu and N. Duru, "Turkish text summarization and methods," in *Proc. of 18th Academic Computing Conference -AB 2016, Aydın, Turkey, January 30-February 5, 2016,* pp. 69–76.

[6] DergiPark, *DergiPark Akademik* [Online]. Available: DergiPark Akademik, https://dergipark.org.tr/tr/. [Accessed: 16 Dec. 2023].

[7] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: A literature survey," International Journal on Digital Libraries, vol. 17, no. 4, pp. 305–338, July 2015. doi:10.1007/s00799-015-0156-0

[8] A. Güran, "Automatic text summarization system," Ph.D. dissertation, Yıldız Technical University, Istanbul, Turkey, 2013.

[9] O. Kaynar, Y. Işık, Y. Görmez, and F. Demirkoparan "Genetic algorithm based sentence extraction for automatic text summarization," Journal of Management Information Systems, vol. 3, no. 2, pp. 62-75, December 2017.

[10] H. Torun and A. B. Inner, "Detecting similar news by summarizing Turkish news," in *Proc. of 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2018,* pp. 1-4. doi: 10.1109/SIU.2018.8404826

[11] A. A. Karcioğlu and A. C. Yaşa, "Automatic summary extraction in texts using genetic algorithms,"in *Proc. of 2020 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Turkey, 2020,* pp. 1-4. doi: 10.1109/SIU49456.2020.9302205

[12] Ankara Journal of Health Services Homepage, *DergiPark Akademik* [Online]. Available: DergiPark Akademik, https://dergipark.org.tr/tr/pub/ashd. [Accessed: 16 Dec. 2023].

[13] Çukurova Medical Journal Homepage, *DergiPark Akademik* [Online]. Available: DergiPark Akademik, https://dergipark.org.tr/tr/pub/cumj. [Accessed: 16 Dec. 2023].

[14] Ege Journal of Medicine Homepage, *DergiPark Akademik* [Online]. Available: DergiPark Akademik, https://dergipark.org.tr/tr/pub/egetbd. [Accessed: 16 Dec. 2023].

[15] Gazi Journal of Health Sciences Homepage, *DergiPark Akademik* [Online]. Available: DergiPark Akademik, https://dergipark.org.tr/tr/pub/gsbdergi. [Accessed: 16 Dec. 2023].

[16]Journal of Hacettepe University Faculty of Health Sciences Homepage, *DergiPark Akademik* [Online]. Available: DergiPark Akademik, https://dergipark.org.tr/tr/pub/husbfd. [Accessed: 16 Dec. 2023].

[17] Journal of Istanbul Faculty of Medicine Homepage, *DergiPark Akademik* [Online]. Available: DergiPark Akademik https://dergipark.org.tr/tr/pub/iuitfd. [Accessed: 16 Dec. 2023].

[18] Mersin University Journal of Health Sciences Homepage, *DergiPark Akademik* [Online]. Available: DergiPark Akademik https://dergipark.org.tr/tr/pub/mersinsbd. [Accessed: 16 Dec. 2023].

[19] Journal of Samsun Health Sciences Homepage, *DergiPark Akademik* [Online]. Available: DergiPark Akademik https://dergipark.org.tr/tr/pub/jshs. [Accessed: 16 Dec. 2023].

[20] S. Bal, "New methods for improving the performance of extractive Turkish text summarization," Ph.D. dissertation, Eskisehir Osmangazi University, Eskisehir, Turkey, 2022.

[21] F. Horasan F and B. Bilen "Extractive text summarization system for news texts," International Journal of Applied Mathematics Electronics and Computers, vol. 8, no. 4, pp. 179-184, December 2020. doi:10.18100/ijamec.800905

[22] N. Abdi Omar, "A user based comparative study of automatic text summarization," M.Sc. Dissertation, Institute of Science and Technology, Kocaeli University, Kocaeli, Turkey, 2018.

[23] E. Akulker, "Extractive text summarization for Turkish using TF-IDF and pagerank algorithms," Ph.D. dissertation, The Graduate School of Natural and Applied Sciences of Atılım University, İstanbul, Turkey, 2019.

[24] M. G. Ozsoy, F. N. Alpaslan, and I. Cicekli, "Text summarization using latent semantic analysis," Journal of Information Science, vol. 37, no. 4, pp. 405–417, June 2011. doi:10.1177/0165551511408848

[25] V. Gulati, D. Kumar, D. E. Popescu, and J. D. Hemanth, "Extractive article summarization using integrated TextRank and BM25+ algorithm," Electronics, vol. 12, no. 2, p. 372, January 2023. doi:10.3390/electronics12020372

[26] M. Kar, S. Nunes, and C. Ribeiro, "Summarization of changes in dynamic text collections using Latent Dirichlet Allocation model," Information Processing & Management, vol. 51, no. 6, pp. 809–833, November 2015. doi:10.1016/j.ipm.2015.06.002

[27] S.W. Kim and J.M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes," Human-Centric Computing and Information Sciences, vol. 9, no. 1, August 2019. doi:10.1186/s13673-019-0192-7

[28] M. Zhang, C. Li, M. Wan, X. Zhang, and Q. Zhao, "ROUGE-SEM: Better evaluation of summarization using ROUGE combined with semantics," Expert Systems with Applications, vol. 237, p. 121364, March 2024. doi:10.1016/j.eswa.2023.121364

[29] M. D. Akın and A. A. Akın, "Türk dilleri için açık kaynaklı doğal dil işleme kütüphanesi: ZEMBEREK," EMO Elektrik Mühendisliği Dergisi, vol. 431, no. 1, pp. 38–44, August 2007.

[30] B. Srinivasa-Desikan, *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Packt Publishing Ltd, 2018, pp. 250-287.

[31] R. Rani and D. K. Lobiyal, "An extractive text summarization approach using tagged-LDA based topic modeling," Multimedia Tools and Applications, vol. 80, pp. 3275-3305, January 2021. doi:10.1007/s11042-020-09549-3

** This paper was presented at the 5th International Conference on Artificial Intelligence and Applied Mathematics in Engineering (ICAIAME 2023) and the abstract was published as an e-book.*