BİLGİSAYAR BİLİMLERİ VE TEKNOLOJİLERİ DERGİSİ
JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGIES
http://dergipark.org.tr/tr/pub/bibted

Araştırma Makalesi

An Extractive Text Summarization Model for Generating Extended Abstracts of Medical Papers in Turkish

Anıl Kuş*[1], Çiğdem İnan Acı[2]

[1]*Toros University, Computer Technologies Department, Mersin, Türkiye*
[2]*Mersin University, Department of Computer Engineering, Mersin, Türkiye*

Keywords:
Text Summarization
Extended Abstract
Medical paper
COVID-19

ABSTRACT

The rapid growth of technology has led to an increase in the amount of data available in the digital realm. This situation makes it difficult for users to find the information they are looking for within this vast dataset, making it time-consuming. To alleviate this difficulty, automatic text summarization systems have been developed as a more efficient way to access relevant information in texts compared to traditional summarization techniques. This study aims to extract extended summaries of Turkish medical papers written about COVID-19. Although scientific papers already have abstracts, more comprehensive summaries are still needed. To the best of our knowledge, automatic summarization of academic studies related to COVID-19 in the Turkish language has not been done before. A dataset was created by collecting 84 Turkish papers from DergiPark. Extended summaries of 2455 and 1708 characters were obtained using widely used extractive methods such as Term Frequency and LexRank algorithms, respectively. The performance of the text summarization model was evaluated based on Recall, Precision, and F-score criteria, and the algorithms were shown to be effective for Turkish. The results of the study showed similar accuracy rates to previous studies in the literature.

Tıp Makalelerinin Genişletilmiş Özetlerini Oluşturmak İçin Çıkarımsal Bir Türkçe Metin Özetleme Modeli

Anahtar Kelimeler:
Metin Özetleme
Genişletilmiş Özet
Tıp makalesi
COVID-19

ÖZ

Teknolojinin giderek büyümesi, dijital ortamdaki mevcut veri miktarının artmasına neden olmuştur. Bu durum, kullanıcıların bu geniş veri kümesi içinde aradıkları bilgiyi bulmalarını zorlaştırmakta ve zaman alıcı hale getirmektedir. Bu zorluğu hafifletmek için, klasik özetleme tekniklerine kıyasla daha verimli bir şekilde metinlerdeki ilgili bilgiye erişmenin bir yolu olarak otomatik metin özetleme sistemleri geliştirilmiştir. Bu çalışma, COVID-19 hakkında yazılmış Türkçe tıp makalelerinin genişletilmiş özetlerini çıkarmayı amaçlamaktadır. Bilimsel makalelerin hâli hazırda özetleri olmasına rağmen, daha kapsamlı özetlere de ihtiyaç duyulmaktadır. Türkçe dilinde COVID-19 ile ilgili akademik çalışmaların otomatik özetlemesi bildiğimiz kadarıyla daha önce yapılmamıştır. DergiPark'tan 84 adet Türkçe araştırma ve derleme makalesi alınarak bir veri kümesi oluşturulmuştur. Toplanan veri kümesinden, yaygın olarak kullanılan çıkarımsal yöntemlerden olan Terim Frekansı ve LexRank algoritmaları kullanılarak 2455 ve 1708 karakterlik genişletilmiş özetler elde edilmiştir. Metin özetleme modelinin performansı, Duyarlılık, Kesinlik ve F-skoru ölçütlerine göre değerlendirilmiş ve algoritmaların Türkçe için etkili olduğu gösterilmiştir. Çalışmanın sonuçları, literatürdeki önceki çalışmalarla benzer doğruluk oranları göstermiştir.

***Sorumlu Yazar**

*(anil.kus@toros.edu.tr) ORCID ID 0000-0002-5964-3727
(caci@mersin.edu.tr) ORCID ID 0000-0002-0028-9890

e-ISSN: 2717-8579
Geliş Tarihi: 06/03/2023; Kabul Tarihi: 26/05/2023

1. INTRODUCTION

Natural Language Processing (NLP) is a subfield of Artificial Intelligence (AI) and computer science that deals with the interaction between computers and human languages (Jurafsky and Martin, 2008). The goal of NLP is to enable computers to understand, interpret, and generate human language in a way that is both natural and useful. This is a challenging task, as human language is complex, ambiguous, and context-dependent. To achieve this goal, NLP relies on various techniques from linguistics, computer science, and machine learning.

With the advancement of technology, the quantity of data in the digital area is continually expanding. As the volume of data continues to grow, it becomes increasingly difficult and tedious for users to locate the information they need. Automatic Text Summarization (ATS) systems have been designed to efficiently extract the relevant information from text documents in a more time-effective manner than manual text summarization (Bal and Gunal, 2021).

ATS has a wide range of applications, from news and information extraction to customer feedback analysis, and text-based question answering. As more and more data is generated every day, the need for ATS is becoming increasingly important, and researchers are constantly working to develop new and improved methods for generating accurate and coherent summaries (Bird, Klein and Loper, 2009)

ATS is the process of generating a shorter version of one or more documents while retaining the main idea and key information. It is one of the NLP that deals with the task of condensing a given text to a shorter version while still preserving its essential information. The goal of ATS is to generate a summary that is concise and coherent, yet still conveys the most important information from the original text.

There are two main approaches to ATS: extractive and abstractive. Extractive summarization involves identifying the most important sentences or phrases in the original text and including them in the summary. This is done by using techniques such as keyword extraction, sentence scoring, and clustering. A summary is created by selecting the sentences from the document that best represents that document in the extractive approach. In this type of summarization, the structure of the sentences remains unchanged and the positions of the words are not altered (Kemaloglu Alagöz, 2022). On the other hand, abstractive summarization involves generating new text that is not present in the original document. This is done using techniques such as text generation and natural language generation. Abstractive methods are more complex and require a deeper understanding of the content and context of the original text. But they can generate a more human-like summary, including a paraphrase of the original text. Abstractive summarization is a complex process that involves shortening and rephrasing the original text through the use of linguistic features, making it more difficult to implement than extractive summarization. However, despite ongoing research efforts, it has not yet reached a fully mature level. (Bal and Gunal, 2021).

Both summarization approaches have their advantages and limitations. Extractive methods are generally considered to be more reliable and accurate, but they can be limited by the quality and specificity of the original text. On the other hand, abstractive methods can generate more fluent and natural-sounding summaries, but they are more challenging to implement and can be prone to errors and inaccuracies (Bird, Klein & Loper, 2009).

The research on ATS in the Turkish language commenced during the 1990s, with the study conducted by Oflazer and Kuruoz (Akulker, 2019). Turkish is among the most commonly used 20 languages in the world. The Turkish language has distinctive characteristics compared to well-studied languages in the literature, such as English, Spanish, and German (Safaya et al., 2022). It has agglutinative morphology which means that various new words can be derived by adding suffixes to a root of a word. Working with an agglutinative language such as Turkish is a real and important research issue in the context of ATS. In contrast to the other languages, there are not enough studies done on ATS in the Turkish language (Akulker, 2019). Recently, the number of studies on the Turkish language has started to increase due to its importance. One of the frequently carried out studies is the summarization of Turkish news texts. However, the studies on the summarization of academic papers are very limited. Reading scientific articles is considered to be very important in the healthcare industry and research. Successful summaries can save a significant amount of time for the reader. Although scientific papers already have abstracts, it is also necessary to have more comprehensive summaries. For researchers, it is a requirement to keep track of articles that fall within their area of interest. The challenge today is to be able to select documents that are truly relevant to their research topic from the vast amounts of information that are easily accessible within a short amount of time and to allocate sufficient time to actually read the selected documents (Celik, 2021).

The motivation of this study is to summarize scientific medical papers in Turkish about COVID-19 Pandemic using ATS methods, which have rarely worked in previous studies so far. 84 papers have been collected as the dataset from Dergipark (URL-1). The Term Frequency (TF) and LexRank algorithms, which include extractive methods and are widely used, were selected for summarization. The obtained results show an F-score of 0.52 for the TF method and 0.51 for the LexRank approach

which means that we have achieved an average level of success compared to Turkish ATS studies in the literature.

The rest of the paper is organized as follows: In Section 2, previous studies on ATS are explained. The ATS models are given in detail in Section 3. In Section 4, the results are given. Section 5 concludes the paper.

2. RELATED STUDIES

In this section, a literature review was conducted on papers published in the last seven years on ATS in Turkish is presented.

Table 1. Summary of the previous studies on extractive text summarization

Author	Year	Metric	Accuracy
Kemaloglu Alagöz	2022	BertScore+ ROUGE	0.88
Bal and Gunal	2021	Accuracy	0.78
Akulker	2019	ROUGE	0.86
Torun and Inner	2018	Human Test	0.68
Kaynar, Isik and Gormez	2017	ROUGE	0.45
Demirci, Karabudak, and Ilgen	2017	ROUGE	0.43
Hatipoglu and Omurca	2016	ROUGE	0.60

Kemaloglu Alagöz developed an ATS study that was conducted on Turkish studies in the field of computer science. In addition to pre-processing techniques prevalent in literature, a novel, format-specific pre-processing function was developed. Deep Belief Networks (DBN) were used for summarization. To assess the performance of the developed system, a customized variant of the pre-trained NLP model Bidirectional Encoder Representations from Transformers (BERT) was employed. After summarization with BERT Extractive Summarizer and DBN, the generated summaries were compared using a specialized comparison metric of BERT called BERTScore. Results showed that the system achieved an F-score of 88% in generating the summary of an article (Kemaloglu Alagöz, 2022).

Bal and Gunal developed a new extractive ATS model. In order to evaluate the efficiency of commonly utilized attributes for ATS in Turkish, sequential attribute selection methods are employed. The study was carried out using three different sets of data. The first dataset consists of 100 texts in the categories of economy, art, and sports. The second dataset consists of 20 texts from various categories. Eight different features were utilized, represented numerically using 0 and 1. These features can be described as similarity to the first sentence, similarity to the last sentence, location, length, frequency of usage, usage of question marks and exclamation marks, number of numerical characters, and number of proper nouns. A decision tree was employed for classification, with sentences in the dataset labeled as "in summary" or "not in summary." If the majority of evaluators had decided the sentence would be a summary, the sentence was labeled as "in summary" and otherwise labeled as "not in summary. It is concluded that the accuracy of the model reached %80,84, and the presented model is considered to be efficient in ATS for the future (Bal and Gunal, 2021).

Akulker proposed an extractive ATS system that was specifically developed for the Turkish language. The system employs a statistical-based TF-IDF algorithm, as well as a hybrid approach that combines TF-IDF with the graph-based PageRank algorithm. The study primarily aims to assess the feasibility and efficacy of these algorithms for processing Turkish documents. In addition, the TF-IDF and the TF-IDF with PageRank (Hybrid) systems have been evaluated and compared against each other using ROUGE metrics during the co-selection evaluation process. According to the evaluation results, the performance of the system is dependent on both the threshold and the specific summarization algorithms employed. It was hypothesized that the precision, recall, and F-score values would improve with higher thresholds for both the TF-IDF and Hybrid systems, as human evaluators were not constrained in their selection of sentences during the summarization process, and thus, It was anticipated that summaries generated by humans would encompass a greater number of sentences than those produced using a lower threshold. The findings show that the average F-score values of the Hybrid system are better than those of the TF-IDF system, even at lower thresholds. Additionally, both systems tend to exhibit improved average F-score values with higher threshold summarization (Akulker, 2019).

Torun and Inner developed a dataset of 12,000 Turkish news articles in order to utilize the ATS system. The extraction method was employed for summarization and the texts were broken down into sentences, with abbreviations excluded from evaluation. The resulting summaries were condensed to a maximum of 5 sentences. News items that were shorter than the 5 sentences captured by the news collection tool, as well as those containing only visual content, were not considered in the evaluation. For the detection of similarity, a similar approach was proposed to the traditional methods used in ATS processes. The frequency information of words was consulted for the identification of keywords from the summarized texts (Torun and Inner, 2018).

Kaynar, Isik, and Gormez proposed a genetic algorithm-based sentence extraction method for ATS. The dataset used in the study consists of 120 Turkish news texts and their summaries. 80 documents were trained with the help of a genetic algorithm, the best weight values for the features

were determined, and then these weights were used to summarize 40 test documents, and the results were compared with the original summaries. In the study, the following steps were applied in sequence: cleaning from unnecessary words, tokenization, determining title and content, and selecting summary sentences. In this direction, 8 different features were extracted. After calculating the features of the sentences, these features were combined using a specific function to obtain the sentence score. When these scores were compared, the highest score was used in the summary sentence. Of the 120 Turkish news included in the dataset, 80 were used for training and 40 for testing. After testing, accuracy rates of 84% were reached. The system, which determined the weights through the Genetic Algorithm, has shown success with a serious accuracy rate (Kaynar, Isik, and Gormez, 2017).

Demirci, Karabudak, and Ilgen developed a summarization system for long documents in Turkish. To cluster articles based on their topics, they utilized the cosine similarity method after collecting newspaper articles dynamically from web pages via Real Simple Syndication. The Latent semantic analysis algorithm was employed for sentence scoring. To assess the performance of the system,3four news domains were utilized, each consisting of 20, 30, 20, and 36 documents. The researchers used the ROUGE evaluation metric to compare their system's performance with manually generated summaries. The summaries were created with the assistance of 15 human evaluators. The system achieved an average recall and precision rate of 43%. The authors reported that the system's performance decreased when summarizing long papers and increased when the summarization rate was increased (Demirci, Karabudak, and Ilgen, 2017).

Hatipoglu and Omurca designed a mobile text summarization system that utilized Turkish articles from Wikipedia sources for summarization purposes. The system incorporated an Analytical Hierarchical Process (AHP) algorithm to combine structural and semantic features scores and calculate an overall score for sentences. To assess its effectiveness, the automated summaries were compared with the ones created by humans using precision and recall metrics. The study concluded that the proposed summarization method held a lot of potential in generating an understandable summary of Turkish Wikipedia articles (Hatipoglu and Omurca, 2016).

In this study, when the obtained ROUGE results are compared with the previous studies, it was observed that average success was achieved. BERT, TF-IDF, and PageRank (hybrid) algorithms were applied as an extra to the ATS models in studies that achieved more successful results in the literature.

3. MATERIAL AND METHOD

In this section, we present the details of the dataset and the algorithms utilized for ATS.

3.1 The Dataset

Since there are numerous studies, especially in the field of health, it takes a lot of time for the readers, and it is seen as a need to summarize in this field. During the COVID-19 pandemic, it was crucial to research to find a solution to the disease. Based on this, 84 papers about the pandemic which obtained from the most popular academic journals in Turkish and published on DergiPark. The journals used in the dataset are considered to be pioneers in this field such as the Journal of Istanbul Faculty of Medicine, Hacettepe University Faculty of Health Sciences Journal, Aegean Journal of Medical Sciences, and Cukurova Medical Journal.

Table 2. Paper distribution according to journals in the dataset

Journal Name	# of papers
Suleyman Demirel University Journal of Health Sciences Institute	6
Journal of Samsun Health Sciences	10
Journal Of Health Sciences	5
Health Care Academician Journal	11
Online Turkish Journal of Health Sciences	9
Mersin University Journal of Health Sciences	9
Journal of Anatolia Nursing and Health Sciences	5
Journal of Istanbul Faculty of Medicine	2
Hacettepe University Faculty of Health Sciences Journal	8
Gazi Journal of Health Sciences	4
Aegean Journal of Medical Sciences	3
Cukurova Medical Journal	12

3.2 Pre-processing

Preprocessing is a critical step that should be done after the normalization stage when summarizing the text. During the data acquisition process, it is not uncommon to encounter unwanted characters or incorrect data ordering, which can lead to unacceptable issues (Horasan and Bilen, 2020).

The sections of the dataset in this study were examined in the preprocessing according to whether they were active in the abstract or not. As a result of this review, first of all, the papers were divided into two sections. The sections (i.e.

bibliography, abstract in English, author and journal information) that will not be included in the summary have been removed from the dataset. The second one is the original abstracts (i.e. abstract in Turkish) which are used to evaluate results and compare them to the ATS's results.

The preprocessing steps used in this study are given as follows:

Data cleaning: For NLP applications, certain characters such as numeric characters, punctuation marks, etc. are generally considered non-essential and may be removed or ignored during the pre-processing stage. However, directly removing them might not be enough for the data-cleaning step (Karayigit, Aci and Akdagli, 2021).

Tokenization: This step involves dividing a text corpus or sentence into smaller elements, such as words, phrases, or n-grams. The dataset became more manageable, modifiable, and analyzable. It is significant for enhancing the precision of NLP modeling and analysis by making the linguistic and semantic structures of the data more discernible.

Stop-words removal: Stop-words removal is a common preprocessing step in NLP that involves removing high-frequency, function words from a text corpus. In the Turkish language, these words, such as "ve", "veya", "fakat", "yani" etc., are considered irrelevant for certain NLP tasks. In our study, Natural Language Toolkit (NLTK) was used to remove all stop-words.

Normalization: All texts were converted to lowercase, converted words to their root form and removed characters such as whitespaces, short lines, etc. which are not useful.

3.3 Term Frequency (TF)

TF is one of the main concepts of the extractive summarization method. In this method, firstly we need separate all the words according to their roots and put them in tables where they occur throughout the text. The second step is to calculate all word frequencies by eliminating the stopwords. Then, word frequency scores will be divided by maximum frequency which represents the total score of the paper. The result shows the rate of the words. Some keywords such as "COVID-19", "pandemics" or "Corona" are multiplied by 5 due to an increase in word frequency and selection of these words. After that, sentence scores are calculated by determining sentence lengths. And the words which have high scores will be included in the sentences of the summarization part. In the experiments, it was seen that a summary of a maximum of 30 sentences had the highest ROUGE value.

TF(t) = (Number of times term t appears in a document) / (Total number of terms in the document) (1)

Upon examination of the most successful summary generated using the TF method, it can be seen that the original paper consists of 33,369 characters, while the summary generated by the system is 2,460 characters long and the original abstract is 2,493 characters long. Therefore, it can be said that the summary generated using the TF method is almost as long as the original abstract. The F-score of the summarized text was found to as 0.52.

In the study, the studies published on coronavirus are summarized and the ten publications with the highest ROUGE value are as in table 3.

Table 3. Precision, Recall, and F-score of summaries that have top 10 scores using the TF method.

#	Recall	Precision	F-score
1	0.46	0.61	0.52
2	0.46	0.44	0.45
3	0.40	0.49	0.44
4	0.51	0.38	0.44
5	0.55	0.36	0.44
6	0.40	0.47	0.43
7	0.53	0.35	0.42
8	0.63	0.31	0.42
9	0.36	0.48	0.41
10	0.49	0.34	0.40

3.4 LexRank

LexRank is a graph-based algorithm for ATS. The algorithm creates a graph structure to represent the semantic similarity among sentences in the text and then performs a PageRank analysis on the graph to identify the most important sentences or paragraphs in the text. The PageRank algorithm is commonly used to measure the importance of a website in search engine results. By using the LexRank algorithm, the summary generated will be more meaningful and accurate, as it takes into account the semantic similarity between sentences. The algorithm is widely used for ATS and generally produces good results.

Mathematically, the LexRank algorithm works by representing sentences in a text as nodes in a graph and then using similarity measures such as cosine similarity to determine the edges between the nodes. Cosine similarity is the fundamental measure utilized for evaluating content similarity. It is widely preferred as one of the most common methods for determining the similarity between two texts by comparing them (Celik, 2021). The similarity measure between two sentences is calculated as the dot product of their vector representations. Once the graph is constructed, the LexRank algorithm uses the PageRank algorithm, which is based on a random walk model, to assign a score to each sentence. The PageRank algorithm

works by iteratively updating the score of each sentence in the graph based on the scores of the sentences that are linked to it. The final scores assigned to each sentence by the algorithm represent the importance of that sentence in the text. The highest-scored sentences are then selected to create the summary of the text.

In this study, the LexRank algorithm was applied with sentence weighting, and the parameter for sorting the highest-scoring sentences was set to a minimum of 10. A commonly used threshold value of 0.1-0.2 was chosen to determine the similarity measure, which has led to variations in obtaining broader summaries. When weighing the words based on their frequency, specifically those considered as keywords, they were multiplied by a coefficient of 2 or 3 depending on the context.

The original text consists of 24,492 characters, while the system generated a summary of 962 characters in summaries generated by the LexRank algorithm. The original summary length in this study is 1,567 characters. The performance of the generated summary was calculated as 0.51 F-score.

In Table 4, the ROUGE results of the summaries of the papers are given with the LexRank algorithm.

Table 4. Precision, recall, and F-score of summaries that have top 10 scores using the LexRank algorithm.

#	Recall	Precision	F-score
1	0.43	0.62	0.51
2	0.47	0.52	0.49
3	0.38	0.54	0.45
4	0.41	0.46	0.43
5	0.42	0.42	0.42
6	0.44	0.41	0.42
7	0.53	0.35	0.42
8	0.37	0.48	0.42
9	0.38	0.44	0.41
10	0.38	0.43	0.40

3.5 Design and Implementation

This section provides a detailed account of the technical specifications involved in the development process of the ATS model, along with an explanation of the working principles of the model which utilizes the extracting method with TF and LexRank systems. The subsequent sections will elaborate on each module developed for the summarization process, detailing each step comprehensively.

The ATS model was created by leveraging the Python programming language in the Anaconda Integrated Development Environment, utilizing the Spyder Framework. The hardware setup that was employed throughout the development phase comprised an Intel Core i5-7200 CPU with a processor speed of 2.50 GHZ, 8192 MB of RAM, and 500 GB of Hard Disk space.

Figure 1 illustrates a comprehensive overview of the ATS model architecture that we developed along with a step-by-step demonstration.

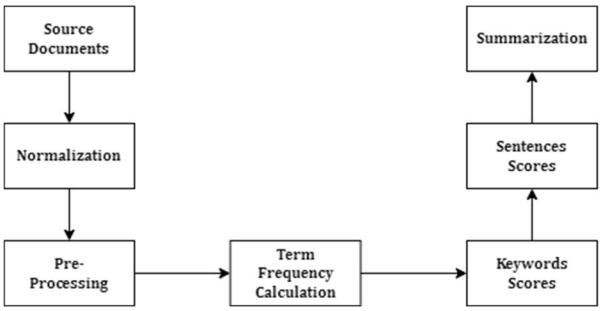

Fig.1. The architecture of the ATS model with TF and keyword scores

4 RESULTS AND DISCUSSIONS

Evaluating the performance of ATS systems is a crucial step in the development and improvement of these systems. The evaluation of ATS systems is a challenging task as it is difficult to quantify the quality of a summary, as it is a subjective task.

Different evaluation measures can be applied to assess the performance of ATS systems. Some of the most commonly used evaluation measures include Recall-Oriented Understudy for Gisting Evaluation (ROUGE), Metric for Evaluation of Translation with Explicit Ordering (METEOR), and Consensus-Based Image Description Evaluation (CIDEr). It includes measures to automatically determine the quality of a summary by comparing it to other (ideal) summaries created by humans. It is observed that the evaluation using ROUGE is the most popular in studies conducted in this field. In our study, the ROUGE value was also used for summary evaluations.

To assess the performance of our ATS study, we utilized the most frequently employed evaluation metrics in the field, which are recall, precision, and the F-score derived from these values.

The four components of the confusion matrix serve as the foundation for evaluating the performance of a classifier. True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN) (Karayigit, Aci and Akdagli, 2021). TP refers to the number of sentences in the generated summary that are also present in the reference summary. FN refers to the number of sentences in the reference summary that are not present in the generated summary. FP refers to the number of sentences in the generated summary that are not present in the reference summary. TN refers to the number of sentences in the original text that are not

present in either the reference summary or the generated summary.

Based on these definitions, precision, recall, and F-score can be calculated using the following formulas:

Precision (P) = TP / (TP + FP) (2)
Recall (RC) = TP / (TP + FN) (3)
F-score = 2 * (P * RC) / (P + RC) (4)

Regarding the TF method and LexRank, we can see that the success of the summarization scores is similar, but LexRank provides shorter summaries. Figure 2 shows the most successful summaries acquired with these methods.

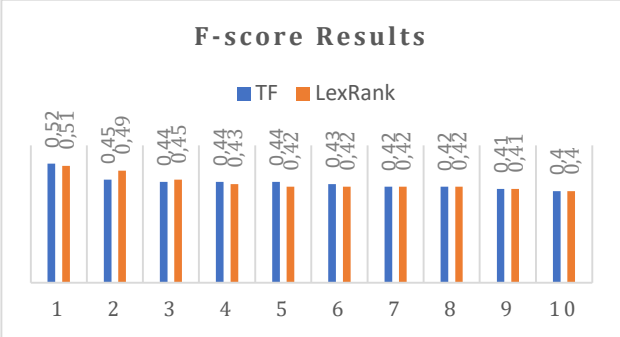

Fig.2. F-score results of two methods

When comparing the two methods, the lengths of the top 10 summaries with the highest F-score are as follows:

- The TF method produces summaries with an average length of 2,455 characters.
- The LexRank algorithm produces summaries with an average length of 1,708 characters.

Based on this result, it can be said that LexRank achieves the same level of performance with shorter summaries. However, the shortness of the summaries does not necessarily mean that they are semantically successful. This comparison is based on the similarity to the original text.

The ranking or frequency values in the BOW (Bag-of-Words) representations obtained from the two methods have shown similarity. The reason for the differences is attributed to the fact that the TF method operates on word-level weights, while the LexRank method operates on sentence-level weights. However, both approaches share similarities in terms of applying normalization and pre-processing steps, as well as extracting summaries based on frequency.

The reason for the similar performance of the summaries obtained using LexRank and TF methods is believed to be due to several factors. Firstly, both methods are based on word frequency, making them frequency-based approaches. Secondly, both LexRank and TF methods fall into the category of extractive summarization techniques. Instead of generating new sentences, they extract important information directly from the original text. Furthermore, both LexRank and TF methods prioritize the original text content. They aim to capture and highlight important information present in the original text.

5 CONCLUSION

This paper presents the performance analysis of the Turkish ATS system that applies two different methods. Increasing keywords scores with the TF method is more successful and its results are promising. To the best of current academic knowledge, there is no study available in the literature that systematically collates and summarizes the existing papers on the subject of COVID-19. In the field of health, it is widely recognized that a comprehensive summary of existing studies would be extremely valuable for professionals working in this field, as it would provide them with a comprehensive understanding of the current state of knowledge on the subject and aid them in their work.

In this study, a system was designed to automatically collect and summarize texts written about the coronavirus. The system was used to summarize 84 different papers. Similarity detection was performed on the summarized texts. The similarity between the original abstracts and the summaries generated by the system was calculated using the ROUGE value.

According to the results, it was observed that the summaries generated by the system were similar to the original texts. Furthermore, the length of the summaries in terms of character count was also analyzed. It was determined that in the TF method, as the length of the summary increased, the success of the summary also increased. In contrast, in the LexRank method, shorter summaries were found to be more successful compared to the TF method.

In the respect to the experiments, we can explain that the most frequent words and keywords do not always reflect the subject of the document correctly. It has been observed that the titles or keywords increase the results and reflect the documents more properly. In addition, since there is not enough study on Turkish ATS, the dataset will contribute to researchers who study the Turkish language.

In future work, we plan to apply other summarization methods, especially abstractive methods using machine learning and deep learning algorithms. We believe that in order to reach more reliable results preprocessing steps must be done carefully. To ensure the effectiveness of summarized texts, they should be compared to a more dependable standard. Although human evaluation may require more resources, it is widely considered to provide a more accurate assessment than comparing the summarized text to the original abstract.

REFERENCES

Akulker, E. (2019). Extractive Text Summarization For Turkish Using Tf-Idf And Pagerank Algorithms (Doctoral dissertation). *The Graduate School Of Natural And Applıed Scıences Of Atılım Unıversity. Turkey.*

Bal, S. and Sora Gunal, E. (2021). A New Model On Automatic Text Summarization For Turkish. *Eskisehir Technical University Journal Of Science And Technology A- Applied Sciences And Engineering, 22(2),* 189–198.

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with python. O'Reilly Media.

Celik Ozkan, A. E. (2021). Structured Abstract Extraction System for Turkish Academic Publications (Doctoral dissertation). *Hacettepe University, Turkey.*

Demirci F., Karabudak, E. and Ilgen, B. (2017). Multi-Document Summarization for Turkish News. *International Artificial Intelligence and Data Processing Symposium (IDAP),* pp. 1-5.

Hatipoglu, A. and Omurca, S.I. (2016). A Turkish Wikipedia Text Summarization System for Mobile Devices. *I.J. Information Technology and Computer Science, vol.1*, pp. 1–10.

Horasan, F. And Bilen, B. (2020). Extractive Text Summarization Systems For News Texts. *International Journal Of Applied Mathematics Electronics and Computers,* 8(4), 179-184.

Jurafsky, D., and Martin, J. H. (2008). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition (2nd ed.). Prentice Hall.

Karayigit, H., İnan Aci, C. and Akdagli, A. (2021). Detecting abusive Instagram comments in Turkish using convolutional Neural network and machine learning methods. *Expert Systems with Applications, 174(March).*

Kaynar, O., Emre Isik, Y. and Gormez, Y. (2017). Genetic Algorithm Based Sentence Extraction for Automatic Text Summarization. *Journal of Management Information Systems*, 3 (2) , 62-75.

Kemaloglu Alagöz, N. (2022). Automatic Text Summarızatıon Wıth Deep Learning (Doctoral dissertation). *Suleyman Demirel University, Turkey.*

Safaya, A., Kurtulus, E., Goktogan, A. and Yuret, D. (2022). Mukayese: Turkish NLP Strikes Back, 846–863.

Torun, H. and Inner, A. B. (2018). Detecting similar news by summarizing Turkish news. *26th IEEE Signal Processing and Communications Applications Conference*, SIU 2018, 1-4.

URL-1: https://dergipark.org.tr/tr/pub/page/about [last accessed: 2023/02/06]