OKU Fen Bilimleri Enstitüsü Dergisi
8(3): 1024-1041, 2025

OKU Journal of The Institute of Science and Technology, 8(3): 1024-1041, 2025

Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi

Osmaniye Korkut Ata University Journal of The Institute of Science and Technology

Analysis of Turkish Academic Papers in Computer Science Using Natural Language Processing Techniques

Caner KARA[1], Ecir Uğur KÜÇÜKSİLLE[2*], Nazan KEMALOĞLU ALAGÖZ[3*]

[1,3] Department of Computer Technologies, Isparta University of Applied Sciences, Uluborlu Selahattin Karasoy Vocational School, Uluborlu, 32650, Isparta
[2] Department of Computer Engineering, Suleyman Demirel University, Faculty of Engineering and Natural Sciences, 32000, Isparta

[1]https://orcid.org/0000-0002-3100-2899
[2]https://orcid.org/0000-0002-3293-9878
[3]https://orcid.org/0000-0002-6262-4244
*Corresponding author: nazanalagoz@isparta.edu.tr

Research Article

Article History:
Received: 23.09.2024
Accepted: 03.02.2025
Published online: 16.06.2025

Keywords:
Natural language processing
Deep learning
Artificiall intelligence
Research article

ABSTRACT

Scientific articles undergo a rigorous and lengthy process from submission to publication in terms of scope, form, and quality. Despite this review process, publications may contain errors that compromise the integrity of their meaning. However, while it takes time to read and understand even a publication with narrative integrity, it becomes even more difficult to examine the details of a publication whose integrity has not been considered. In this study, 492 full-text Turkish research articles in the field of computer science published in various journals indexed in SCI-EXPANDED, ESCI and TR Index were analyzed. The articles are evaluated by five criteria, which are the integrity of abstract and content, the availability of main concepts in the abstract, the availability of standard sections, the frequency of titles and keywords contained in the content, and the percentage of recent publications in the references. The aim of the study is to reveal the integrity of all sections of the articles examined according to the defined criteria. The findings revealed that in the majority of the articles, abstracts were insufficient to include basic elements such as scope, problem, purpose, method, findings and conclusion, while titles and keywords did not adequately represent the content of the article. In addition, significant deficiencies were identified in terms of the currency of references. This study aims to provide guidance for more effective title, abstract and reference selection in scientific article writing.

Bilgisayar Bilimleri Alanındaki Türkçe Akademik Makalelerin Doğal Dil İşleme Teknikleri Kullanılarak Analizi

Araştırma Makalesi

Makale Tarihçesi:
Geliş tarihi: 23.09.2024
Kabul tarihi:03.02.2025
Online Yayınlanma: 16.06.2025

Anahtar Kelimeler:
Doğal dil işleme
Derin öğrenme
Yapay zeka
Araştırma makalesi

ÖZ

Bilimsel makaleler başvurudan yayına kadar kapsam, biçim ve kalite açısından titiz ve uzun bir süreçten geçer. Bu inceleme sürecine rağmen, yayınlar anlam bütünlüğünü tehlikeye atan hatalar içerebilir. Ancak anlatım bütünlüğü olan bir yayını bile okumak ve anlamak zaman alırken, bütünlüğü dikkate alınmamış bir yayının detaylarını incelemek daha da zorlaşmaktadır. Bu çalışmada, bilgisayar bilimleri alanında SCI-EXPANDED, ESCI ve TR Dizin'de taranan çeşitli dergilerde yayımlanmış 492 tam metin Türkçe araştırma makalesi analiz edilmiştir. Makaleler, özet ve içerik bütünlüğü, özette ana kavramların bulunması, standart bölümlerin bulunması, içerikte yer alan başlık ve anahtar kelimelerin sıklığı ve kaynakçada yeni yayınların bulunma yüzdesi olmak üzere beş kritere göre değerlendirilmiştir. Çalışmanın amacı, belirlenen kriterlere göre incelenen makalelerin tüm bölümlerinin bütünlüğünü ortaya koymaktır. Bulgular, makalelerin büyük çoğunluğunda özetlerin kapsam, problem, amaç, yöntem, bulgular ve sonuç gibi temel unsurları içermekte yetersiz olduğunu, başlık ve anahtar kelimelerin ise makale içeriğini yeterince temsil etmediğini ortaya koymaktadır. Ayrıca, referansların güncelliği açısından önemli eksiklikler tespit edilmiştir. Bu çalışma, bilimsel makale yazımında daha etkili başlık, özet ve referans seçimine yönelik rehberlik sunmayı amaçlamaktadır.

To Cite: Kara C., Küçüksille EU., Kemaloğlu Alagöz N. Analysis of Turkish Academic Papers in Computer Science Using Natural Language Processing Techniques. Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi 2025; 8(3): 1024-1041.

1. Introduction

The analysis and processing of the natural languages that humans use to communicate, by machines using various methods, is called Natural Language Processing (NLP). There are two approaches to NLP in the literature; symbolic NLP and empirical NLP. Empirical NLP is based on experiments and observations using statistics. Statistical analysis was very rational and practical compared to methods using artificial intelligence (AI) until the breakthrough advances in deep learning. Symbolic NLP, on the other hand, uses AI algorithms that convert symbols into numbers in any natural language and process them. The symbolic approach takes into account all the rules and concepts of a natural language. In recent years, with the advances in deep learning, using these two approaches together to benefit from the advantages of both has become a commonly preferred modern approach (Jackson and Moulnier, 2002).

Automated Text Summarization (ATS) is one of the most popular research and application areas of NLP. ATS is categorized according to how the summarized text is generated. Extractive Summarization (ES) is the method that selects essential sentences from the original text. Another method is based on understanding the original text and generating a rephrased summary, called Abstractive Summarization (AS). There are also methods that combine these two methods in various ways, called hybrid summarization methods.

AI algorithms, mostly deep learning methods, are often used in AS. Recurrent Neural Networks (RNNs) with building blocks called Long Short-Term Memories (LSTMs), various techniques such as Encoder and Decoder based Transformers, Attentions, and Bidirectional Encoder Representations of Transformers (BERT) that include Transfer Learning (TL), which uses a pre-training process to reduce training time, are commonly used AI methods in NLP.

This study is similar to the (Teufel,1999; Teufel et. al. 1999; Teufel and Moens, 1999; Feltrim et. al., 2004) as the main subject. Feltrim et al., as in this study, examined articles in the field of computer science, but these articles were published in Portuguese. Also, these articles were evaluated only for the automatic summarization of the abstracts, and for this purpose statistical methods were used. Feltrim et al. is focused on labeling only by summary, instead of a holistic evaluation, which we had done in our study (Feltrim et. al., 2004).

In Teufel's works (Teufel,1999; Teufel et. al. 1999; Teufel and Moens, 1999), English-language articles published in the field of computational linguistics were studied in a similar way. In these publications, the authors focused on automatic and manual tagging according to seven different tags that they identified in the abstract and introduction sections of the articles. Like Feltrim et al., these studies also focused on the abstract and introduction, and the labeling of these sections is emphasized instead of a holistic evaluation of the articles.

In Turkish scientific articles, classification and summarization have generally been studied as in (Kemaloğlu Alagöz, 2022). Apart from our study, no other study has been found that includes all parts of the article, from the abstract to the bibliography, and performs a holistic evaluation with statistical methods and artificial intelligence.

In this article, 492 full-text Turkish research articles in the field of computer science were evaluated. The 492 articles used in our study were randomly selected from TR Index, SCI-EXPANDED and ESCI indexed journals that include Turkish academic articles. These journals were specifically chosen to create a large data set as they have different quality standards and content coverage. The main scoring concept of the study is shown in the Figure 1.

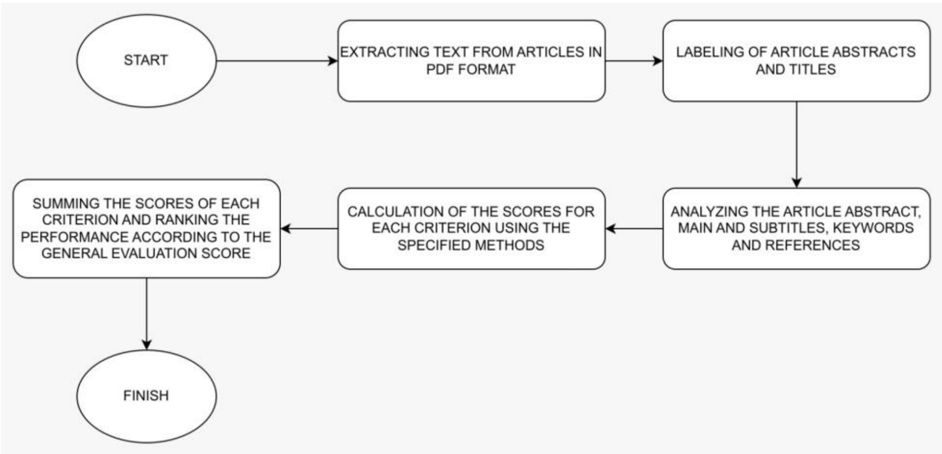

Figure 1. The main scoring concept of the study

While evaluating the articles, five criteria were used to determine the completeness of all sections. These criteria are defined as the integrity of abstract and content, the availability of main concepts in the abstract, the availability of standard headings, the frequency of titles and keywords contained in the content, and the percentage of recent publications in the references. The weights of these criteria were determined intuitively by our labeling experts.The criteria and the scoring value of each criterion are given in Figure 2.

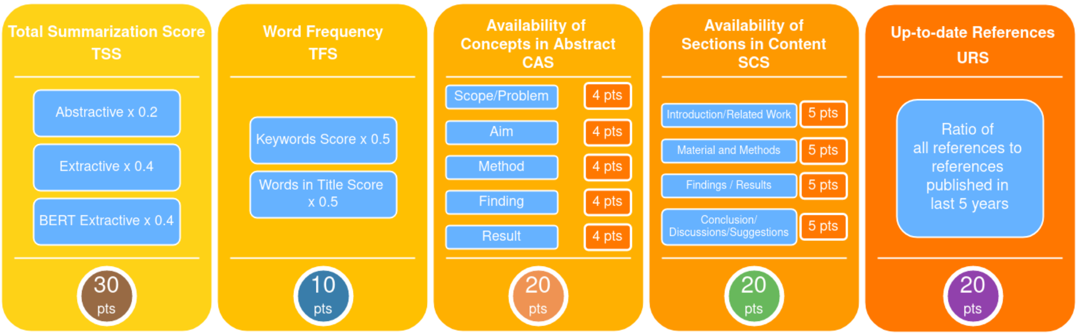

Figure 2. Criteria and overall score distribution

2. Material and Method

2.1. Natural Language Processing (NLP)

NLP is the process of making the desired inferences from a pre-processed text written in any natural language and producing an answer using various AI methods. In everyday life, search engines, voice assistants on phones, language translation tools, etc. are NLP applications. The field began in the 1950s with rule-based systems, and in the 1990s with statistical methods. Between 2000 and 2014, ML methods and now DL methods and their applications have been used to process text data (Campesato, 2021). To analyze the text and get successful results, some preprocessing should be done, such as cleaning the text from punctuation and stop words, lowercasing all letters, tokenizing, stemming, and lemmatizing the text. Stop words are defined as words in the text that have less meaning compared to other words. For example, conjunctions are stop words. Tokenization is the process of breaking the text into smaller pieces. The division into words or sentences is determined by the algorithm or purpose of the processing. Stemming is the process of leaving only the stem of a word and deleting the rest.

Term Frequency (TF) is one of the statistical approaches in NLP. The importance of a word is determined by the number of times it occurs in the text. Term Frequency - Inverse Document Frequency (TF-IDF) is a technique that gives an opinion about how important this word is for this text (Zhou et al., 2024).

Word2Vec is a word embedding method used in NLP applications. In this method, words are represented by vectors according to their mutual meanings. Similar words are represented with similar vectors in distance, and unrelated words are represented with a greater distance in vector space (Mallik and Kumar, 2024). CBOW (Continuous Bag of Words) and Skip-gram are two opposite methods of Word2Vec. CBOW tries to find the word in the middle, while Skip-gram tries to find the words between the input words (Chen et al., 2024).

The field of Natural Language Processing (NLP) has made significant progress in automatic text summarization and text classification models, especially after the introduction of deep learning techniques. Automatic text summarization is classified into two different areas: extractive and abstractive summarization. Extractive summarization involves selecting key sentences or phrases from the source text, while abstractive summarization generates new sentences with the same meaning as the original text (Alsuhaibani, 2023; Shafiq et al., 2023). Recent work has focused on improving these methodologies using deep learning frameworks.

Shopnil (2024) presents a novel text summarization approach that uses K-means clustering and sentence embedding without relying on traditional word embedding techniques. This method demonstrates the flexibility of deep learning in effectively summarizing large volumes of text.

Gao and Karuna (2023) investigated the use of Recurrent Neural Networks (RNNs) for automatic abstractive summarization, emphasizing the importance of preserving the underlying meaning of the text while vectorizing words and sentences. In the same vein, Shafiq et al. (2023) argue that creating coherent summaries requires a deep understanding of the semantics of the text. Sun et al. (2021) also proposed the integration of attention mechanisms and deep reinforcement learning to optimize the summarization process. The use of transducer-based models such as BERT has revolutionized summarization tasks by providing robust contextual embeddings that improve the quality of the generated summaries (Vaswani et al., 2017; Teng, 2023).

2.2. Automatic Text Summarization

Automatic text summarization is the process of identifying the important points of text content and separating them from the rest of the content. There are three main approaches used for automatic summarisation. These are extractive summarizing, which is the approach of selecting the sentences representing the main idea in the text and summarizing them without changing them, and abstractive summarizing, which is the approach of perceiving the main idea in the text and expressing it with new sentences. The approach formed by using these two approaches together is known as hybrid summarization. (Erhandı et al., 2020).

In the field of text classification, deep learning models have outperformed traditional machine learning approaches on several tasks, including sentiment analysis and topic categorization. Minaee et al. (2021) presented a comprehensive review detailing the architectures and performance metrics of more than 150 deep learning models. In their comprehensive review, the authors found that implementing more sophisticated models that utilize large datasets and complex neural networks will yield better results. In recent years, advances in deep learning models have brought significant innovations in the field of text summarization. In particular, models such as BART, PEGASUS and T5 are capable of generating more natural and human-like summaries by capturing the meaning of texts. BART performs strongly in language modeling and sequence prediction tasks using both encoder and decoder structures (Lewis et al., 2020). PEGASUS is a model specifically designed for summarization and is notable for its ability to extract main sentences from input text (Zhang et al., 2020). The T5 model reduces text processing tasks to a "transformation" problem and is flexible, so it can be used effectively for summarization and other natural language processing tasks (Raffel et al., 2020). However, these models have important limitations. For example, in morphologically rich languages such as Turkish, the effectiveness of these pre-trained models may not be at the desired level due to the limited data sets. Moreover, the high computational costs required for training these models pose challenges in terms of both time and resources.

Moreover, the integration of advanced techniques such as transfer learning has been crucial in improving classification results. This approach allows pre-trained models on large datasets to be fine-tuned for specific tasks, significantly reducing the amount of labeled data and training time required for training (Singh & Mahmood, 2021). The effectiveness of these models is further supported by the systematic review conducted by Tahseen, which highlights the growing need for automated systems that can efficiently process and classify large amounts of textual data (Tahseen, 2021).

2.3. BERT

BERT, translated into Turkish as 'Bidirectional Encoding Representations from Transformers', is a pre-trained language detection model developed by Google Artificial Intelligence. After pre-training the model, a second training, called fine-tuning, is applied to the desired target. This modelling technique with two-step training is called Transfer Learning. Thanks to the pre-trained models, it is possible to obtain high successful results in a shorter time with less data.

BERTScore basically calculates the matches between the reference sentence (x) and the candidate sentence (ˆ x) by utilising the cosine similarity between two sentences. With the BERT algorithm, words separated from sentences are obtained. The BERT Score metric is measured by the F1 score (Equation 3), which is calculated from the recall (Equation 1) and precision (Equation 2) values (Kemaloğlu Alagöz, 2022).

$$R_{BERT} = \frac{1}{|x|} \sum_{x_i \in x} \max x_i^T \hat{x}_j \ , \ (\hat{x}_j \in \hat{x}) \tag{1}$$

$$P_{BERT} = \frac{1}{|\hat{x}|} \sum_{\hat{x}_j \in \hat{x}} \max x_i^T \hat{x}_j \ , \ (x_i \in x) \tag{2}$$

$$F_{BERT} = 2 \frac{P_{BERT} \cdot R_{BERT}}{P_{BERT} + R_{BERT}} \tag{3}$$

3. Findings

3.1. Collecting Dataset

In this study, academic papers in computer science written in Turkish are used, which are served by Dergipark, a process management system and data repository for electronic academic papers in Turkey. Papers published in SCI-EXPANDED, ESCI and TR-indexed electronic journals are especially preferred. TR Index (Formerly National Databases-UVT) is a national citation index in Turkey, which can be searched on the website, containing articles from national, peer-reviewed, scientific journals at the end of science and social science topics, and bibliographic/full-text information from TUBITAK

(Scientific and Technological Research Council of Turkey). TR Index service, TUBITAK ULAKBIM (National Academic Network and Information Center) by link (STRCT, 2000). Table 1 shows the representative names of the journals and the indexes in which they are searched. The journals and articles selected for review have been renamed without using their full names in accordance with ethical values. The papers evaluated from each journal vary because of the limited availability of papers that provide the expected conditions.

Table 1. Evaluated papers published by electronic journals

Journal	SCI –E.	ESCI	TR Index	Count
Journal A	-	-	+	128
Journal B	+	-	+	73
Journal C	-	-	+	53
Journal D	-	-	+	48
Journal E	-	+	-	40
Journal F	-	-	+	38
Journal G	-	+	+	30
Journal H	-	-	+	21
Journal I	-	-	+	13
Journal J	-	-	+	10
Journal K	-	-	+	9
Journal L	-	-	+	6
Journal M	-	-	+	5
Journal N	-	-	+	4
Journal O	-	-	+	3
Journal P	-	-	+	3
Journal R	-	-	+	3
Journal S	-	-	+	3
Journal T	-	-	+	1
Journal U	-	-	+	1

3.2. The Integrity of Abstract and Content

As a first step, the articles in the journals listed in Table 1 are downloaded as PDFs, and the plain text in the PDFs is extracted using the X-PDF library in PHP. In order to tag the corresponding concepts in the abstracts and standard sections of the abstracted text, a platform is built and published on the web. Three Ph.D. specialists in computer science separately labeled the sentences of the abstracts and divided the articles into standard sections. The highest selected label in the specialists' count is considered as the final evaluation of that abstract sentence or section. If there is no conclusion, the concept of the sentence or the standardized heading for the section is discussed by the experts until there is a conclusive one. Figure 3 shows the main concepts in the abstracts. The text of the articles in the figure is blurred to protect the anonymity of the articles and journals. These concepts are explained in items as follows:

Scope/Problem - Sentences that refer to the scope of the work or define the problem and related work.

Purpose - Sentences that refer to the main purpose of the work.

Method - Sentences that refer to the scope of the work or define the problem and related work.

Findings - Sentences related to the main purpose of the work.

Result - Sentences that refer to the scope of the work or define the problem, as well as related work.

Figure 3. Interface for marking abstract concepts in the platform

The order of writing scientific articles has been developed for easier understanding by the readers and the IMRAD (Introduction, Methods, Results, and Discussion) model has been accepted by almost all segments. In this context, each journal imposes formal constraints that must be followed when writing articles nowadays. While in some journals these restrictions are very strict, in others the author is free to determine the subject headings of the article.

Figure 4 shows the standard main sections of an article. The text of the articles in the figure is blurred to protect the anonymity of the articles and journals. These sections are as follows

- Introduction/Related Work
- Methods
- Results
- Conclusion/Discussion/Suggestion

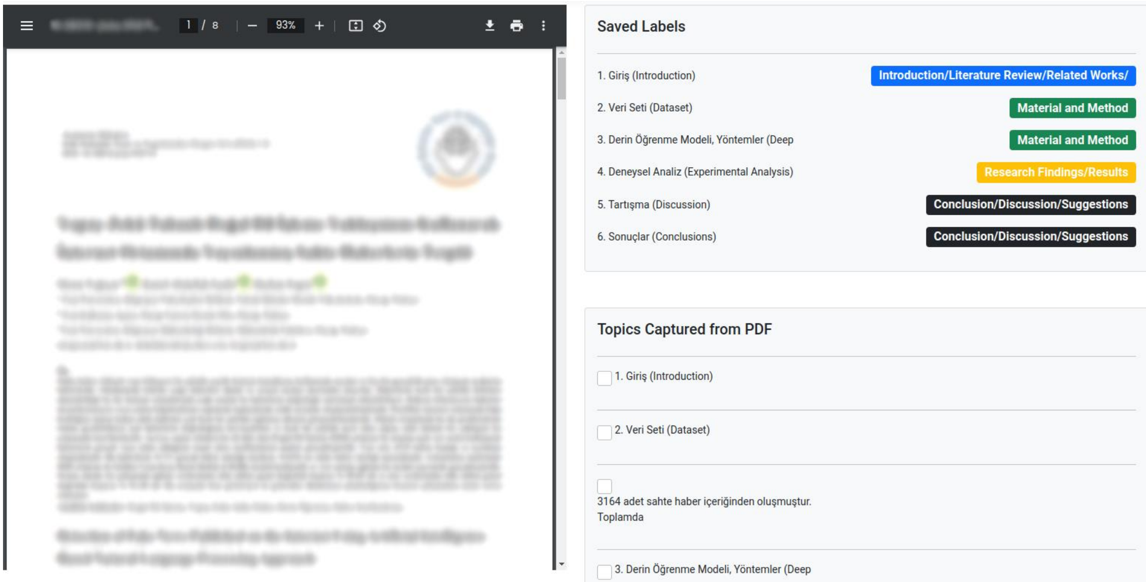

Figure 4. Interface for marking main sections in the platform

According to Gastel & Day, an abstract of a scientific study should be written in the past tense, state the scope and purpose of the research, define the method and methodology, summarize the research findings, and state the main results (Gastel and Day, 2022). Except in special cases, the abstract should not be cited. An abstract is a projection of the entire article content, which is expected to include elements from the main headings of the article (Gastel and Day, 2022).

In order to analyse the extent to which the articles obtained overlap with the existing abstracts, the extent to which they overlap with the abstracts of the articles was analysed using the accepted summarisation methods in the literature. For the analysis, two different inferential (Miller, 2022; Kemik, 2020) and one abstractive (Veyssier, 2020) summarisation models available in the literature were used. For inferential summarisation, the BERT Inferential Summarisation model was preferred. For this model, all articles were given to the system in order and the BERT Model was asked to create a summary containing an equal number of sentences with the main summary of the article. Afterwards, the extent to which these two summaries overlap was determined by looking at the BERT Score similarity.

Scores were calculated for each article using the outcome of the model (Kemik, 2020) and F-scores were obtained using the BERT score metric. Equation 4 gives the formulation of the Extractive Summarization Score (ES).

$$\mathrm{ES} = \mathrm{F} - \mathrm{Score}(\mathrm{ES}) * 0.4 * 30 \tag{4}$$

Scores were calculated for each article using the outcome of the model (Miller, 2022) and F-scores were obtained using the BERT score metric. Equation 5 gives the formulation of the Extractive Summarization Score (EBS) using BERT.

$$EBS = F - Score(EBS) * 0.4 * 30 \tag{5}$$

Scores were calculated for each article using the outcome of the model (Veyssier, 2020) and F-scores were obtained using the BERT score metric. Equation 6 gives the formulation of the Abstractive Summarization Score (AS).

$$AS = F - Score(AS) * 0.2 * 30 \tag{6}$$

The Total Summarization Score (TSS) is the sum of the three summarization scores (Equations 4, 5 and 6) of the article (Equation 7).

$$TSS = ES + EBS + AS \tag{7}$$

The integrity of the abstract and content is considered the most important criterion because abstracts are read first by researchers and are an important part that gives clues about the whole content. That's why this criterion represents 30% of the OS (Overall Score) and is called TSS (Total Summarization Score). Two extractive summarization methods equally share 80% of TSS and the abstractive summarization method forms 20% of TSS. Due to the challenges of abstractive summarization in Turkish, abstractive summarization affects OS less than extractive summarization methods.

3.3. The Availability of Main Concepts in Abstract

An abstract of a scientific article should contain sentences that address all the main concepts (Gastel and Day, 2022) In this study, the absence of one or more scope/problem, aim, method, finding, or result concepts in the abstract, leading to inconsistency, is considered a negative point of the article. This criterion affects 20% of OS and is called CAS. The availability of each concept adds 4% to the OS. No points are awarded for the absence of a concept. Table 4 shows the results of the evaluation of the availability of the main concepts in the articles.

Table 4. The availability of main concepts in abstract

Availability	SP	A	M	F	R
Available	421	374	447	199	322
N/A	71	118	45	293	170

SP: Scope/Problem label, M: Method label, F: Finding label, R: Result label

According to the evaluation, there was no scope/problem label in 71 articles, 118 articles did not have a main purpose, 45 articles were unable to mention the method of the work, 293 articles lacked to mention the findings of the research and 170 articles did not have judgments inferred from the findings. Most of the abstracts focused on scope/problem and method rather than findings, results, and final judgments. There were two articles with 4 points, 54 articles with 8 points, 158 articles with 12 points, 210 articles with 16 points, and 68 articles with 20 points.

3.4. The Availability of Standard Sections

In the absence of journal restrictions, some of the standard sections have not been used or have been combined with another standard section. An effectively presented article should not have missing sections or unnecessarily exaggerated headings. This criterion evaluates whether or not the article has these standard sections. The availability of standard sections criterion affects 20% of the OS and is called SCS. The availability of sections labeled according to the IMRAD standard is added to the OS by 5%. Nothing is added to OS if the article does not contain the standard section. In Table 5, the number of articles with standard sections was reported as present and absent. It is observed that every article had an Introduction/Related Works section, one article did not have a Materials and Methods section, the

Results section was missing in 81 articles, and three articles did not have a Conclusion/Discussion/Suggestions section.

Table 5. The availability of standard sections in content

Availability	I/ RW	MM	R	C/D/S
Available	492	491	411	489
N/A	0	1	81	3

I/RW: Introduction/Related Work, MM: Material and Method, R: Results, C/D/S: Conclusion/Discussion/Suggestions

It is also seen that the Results and Conclusion/Discussion/Suggestions sections were mostly mixed. It is found that there were 85 articles with 15 points, 407 articles with 20 top points.

3.5. The Availability of Keyword Frequencies

Keywords are chosen from the terms related to the work that are not included in the title or are slightly different. The purpose of defining keywords in an article is to increase the accessibility of the work in search results. A consistent scientific paper is expected to contain an optimal level of keyword repetition. When scoring by keyword frequency, the sum of keyword frequencies in each article was divided by the total number of words in the article during the article evaluation process. Table 6 shows the distribution of keywords according to their repetition percentages in the article.

Table 6. Percentage of keyword frequency in articles

Percentage rate (%)	Article count
8-100	0
7-8	1
6-7	3
5-6	7
4-5	13
3-4	39
2-3	88
1-2	140
0-1	201

Thus, SKWF (Keyword Frequency Score), the total percentage frequency of keywords in the article compared to the article was obtained. In the content of all articles the word "olarak" was found to be the word with the highest frequency (15194 times). The frequency value of the word with the highest frequency was proportional to the total number of words in the whole article, and the percentage of the highest frequency value (unit frequency) was found. The value of optimal SKWF is obtained by multiplying this value by the number of keywords in the article. Given the values found, the Keyword Frequency Score KWFS was obtained using Equation 8.

$$KWFS = \frac{SKWF_{(Optimum)} - SKWF}{SKWF_{(Optimum)}} * 5 \tag{8}$$

In this equation, the higher SKWF does not mean the higher KWFS. To show the difference, the article with the largest SKWF and the largest KWFS is given in Table 7.

Table 7. The largest SKWF and the largest KWFS comparison

ID	FS	KC	UF	Opt SKWF	KWFS
432	4.967	5	0.99	4.95	4.982
106	7.012	5	0.99	4.95	2.916

ID: Record Number, FS: Frequency Sum, KC: Keyword Count, UF: Unit Frequency, Opt. SKWF: Optimum Sum of Keywords Frequency, KWFS: Keyword Frequency Score

The distribution of keywords in the article criterion affects 5% of the OS. On the other hand, article titles are the most searched parameter. That's why words in a title should be the most defining parts of a scientific paper. The content of a consistent article should contain significant repetition of these words. When scoring according to the frequency of the words in the title, the sum of the frequencies of the words that make up the title of each article is divided by the total number of words in the article during the evaluation process of the article. This gives the STWF, the total percentage frequency of the words in the title of the article compared to the whole article.

Unlike the evaluation performed for keywords, stop words in the title and article content are neglected. The frequency value of the word with the highest frequency in the content of all articles was proportional to the total number of words in the whole article, and the percentage of the highest frequency value (unit frequency) was found. The value of optimal STWF was obtained by multiplying the unit frequency value by the number of words in the title of the article. Given the values found, the Title Words Frequency Score TWFS was obtained using Equation 9.

$$TWFS = \frac{SKWF_{(Optimum)} - SKWF}{SKWF_{(Optimum)}} * 5 \tag{9}$$

Table 8. Percentage of article title word frequency

Percentage rate (%)	Article count
22-100	0
20-22	1
15-20	8
10-15	63
5-10	279
0-5	141

Table 8 shows the distribution of title words according to their repetition percentages in the article. In Equation 6, the larger STWF does not mean the larger TWFS. To show the difference, the article with the largest STWF and the article with the largest TWFS are shown in Table 9.

Table 9. The largest SKWF and the largest KWFS comparison

ID	FS	TWC	UF	Opt. STWF	TWFS
233	9.993	10	1	10	4.997
122	21.943	14	1	14	2.163

ID: Record Number, FS: Frequency Sum, TWC: Keyword Count, UF: Unit Frequency, Opt. STWF: Optimum Sum of Title Words Frequency, TWFS: Title Word Frequency Score

The distribution of title words in the article criterion affects 5% of the OS. In Equation 10, the Sum of Keywords Frequency Score (KWFS) and Title Words Frequency Score (TWFS) are equal to Total Frequency Score (TFS).

$$TFS = KWFS + TWFS \tag{10}$$

3.5. The Percentage of Recent Publications in the References

The publication date of other publications from which scientific research benefits gives serious clues as to the current research. For this reason, one of the study's criteria was to extract the distribution of the publication dates of the sources in the bibliography section and to gradually score them according to the percentage of recent articles in this distribution. According to the specified criteria, PHP regex was used to determine what percentage of the references in the bibliography were published in the last five years according to the publication date of the evaluated article, and a share of 20% of the overall evaluation was distributed to each evaluated article according to this percentage.

$$URS = \frac{URC}{TRC} * 20 \tag{11}$$

In Equation 11, it is shown that the Up-to-date Reference Score (URS) is equal to the multiplication of 20 and the Up-to-date Reference Count (URC) by the Total Reference Count (TRC). At least half of the references published in the last five years are observed in 178 articles (Table 10).

Table 10. Percentage of article title word frequency

Percentage rate (%)	Article count
90-100	10
80-89	14
70-79	27
60-69	45
50-59	82
0-49	314

4. Results

The articles were sorted in descending order by their OS. The articles with the highest and lowest scores are listed in Table 11 and Table 12.

Table 11. Top 5 articles with the highest OS

Rep. Name	ID	TSS	CAS	SCS	TFS	URS	OS
Journal D	358	15.422	20	20	9.478	20	84.900
Journal A	137	18.631	20	20	3.054	16.800	78.485
Journal D	316	12.357	20	20	7.945	16.667	76.969
Journal A	180	15.632	20	16	8.265	16.667	76.564
Journal D	281	10.756	20	20	6.334	17.241	74.331

Table 12. Bottom 5 articles with the lowest OS

Rep. Name	ID	TSS	CAS	SCS	TFS	URS	OS
Journal L	473	6.405	20	12	3.753	0	42.158
Journal E	320	7.013	20	8	3.286	3.125	41.424
Journal F	315	8.833	15	12	5.230	0	41.063
Journal A	195	6.987	15	8	5.211	5.714	40.912
Journal H	375	7.976	20	8	4.842	0	40.818

Table 13 shows the average scores of the articles grouped by the journals in which they were published.

Table 13. Average scores of the journals

Rep. Name	Article Count	Average Score
Journal P	3	64.928
Journal K	9	61.383
Journal T	1	61.255
Journal C	53	60.828
Journal D	48	60.511

Scientific article titles are an important element that summarizes the content of the study and aims to attract the reader's attention. In the field of natural language processing (NLP), acrostics (e.g. names of models such as BART, PEGASUS, T5) are frequently used. Such acrostics offer the advantage of increasing memorability while summarizing technical details. However, overuse of such words, which are not always directly linked to the content, can prevent the title from fully representing the study. Alternatively, choosing descriptive and highly representative words that better reflect the content can make the work easier to find and understand.

In this context, it is clear that the choice of title requires a balance. On the one hand, short and expressive phrases such as acrostics are important to engage the reader and create a certain brand perception. On the other hand, accurate and effective representation of the content is critical for titles to provide direct information to the reader. Therefore, while utilizing the advantages of acrostics in the selection of titles in natural language processing studies, it is necessary to ensure that the title clearly reflects the scientific content.

5. Conclusions and Discussions

In the literature, no publication is directly related to the semantic integrity of academic publications. This may be due to the fact that each journal accepts articles to be published with different scope, format, and expertise values. In this study, it was observed that each journal, even in different volumes of the same journal, had different formal characteristics, and these differences were perhaps the most challenging part of the study.

The results of the evaluation reveal striking results in terms of the format of the articles examined and the elements that should be included in the article expression. As a result of the overall evaluation, the highest score was given to the article with registration number 358 with 84,900 points, and the lowest score was given to the article with registration number 197 with 34,733 points. Journals P, K, and T stand out as the most successful journals with average article scores of 64.928, 61.383, and 61.255, respectively.

In the summarization part, the low similarity between the abstract generated after applying the abstractor, summarizer and BERT-driven summarizer and the original abstract reveals that the abstract sentences that are expected to represent the relevant article section have poor representativeness. This situation shows that the abstract section is not written carefully in Turkish computer science research articles.

In addition, it was found that only 68 of the 492 articles included in the evaluation included all the concepts in the abstracts labeled by the experts, while 85 articles did not include any of the main headings. In addition, it was found that only 68 of the 492 articles included in the evaluation contained all concepts in the abstracts labeled by the experts, and 85 articles did not contain any of the main headings. The experts involved in the evaluation during the research process expressed the opinion that the IMRAD standard used in scientific publications shortens the process of reading and understanding the article. As a result, it was concluded that the Abstract section of the majority of the articles reviewed was partially or largely devoid of terms indicating the scope, problem, purpose, method, results, and conclusion. It was found that in 11 articles none of the keywords were included in the article content, and in the bibliography of 17 articles there was no reference within the last five years from the date of publication.

As a result of the developments in the IT sector and digitization, scientists perform literature searches on the various platforms where scientific publication databases are located, searching for the studies they are interested in. From time to time, scientists have difficulties in finding the appropriate keywords, sometimes do not find what they are looking for, have to sift through many irrelevant publications, or can only reach all the publications on the relevant subject by trying several different search methods. This situation leads to the fact that important works encountered during the working period are discovered much later and often too late, wasting effort, energy and time. It is estimated that this study will raise awareness for the use of more accurate and effective titles and keywords in future academic studies, more effective writing of abstracts, and more consistent content creation.

Future studies will consider automating the labeling of the concepts that make up the abstract using appropriate machine learning algorithms, developing the application of the work done in this study as an evaluation tool, and applying it in different languages.

Conflict of Interest

The article authors declares that there is no conflict of interest.

Author's Contributions

The authors contributed equally to the study.

References

Alsuhaibani M. Fine-Tuned Pegasus: Exploring the performance of the transformer-based model on a diverse text summarization dataset. 9th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'23), August 03-05 2023, London.

BERT Extractive Summarizer. 2020; https://github.com/dmmiller612/bert-extractive-summarizer. (Access Date: 11.02.2025)

Campesato O. Natural language processing fundamentals for developers. Herndon: Mercury Learning and Information; 2021.

Chen YM., Lin P., Yeh EH., Yang SR., Lu R. CPBW: A change-point-detection and bag-of-words based mechanism utilizing smartphone triaxial accelerometer data for driver identification. IEEE Internet of Things Journal 2024; 11(18): 29766-29780.

Erhandı B., Text summarization with deep learning. Sakarya University Institute of Science and Technology, Master Thesis, Sakarya, Turkiye, 2020.

Feltrim VD., Pelizzoni JM., Teufel S., Das Graças Volpe Nunes M., Aluísio SM. Applying argumentative zoning in an automatic critiquer of academic writing. 17th Brazilian Symposium on Artificial Intelligence, September 29-Ocotber 1 2004, Sao Luis.

Gastel B., Day RA. How to write and publish a scientific paper. New York: Bloomsbury Publishing; 2022.

Jackson P., Moulnier I. Natural language processing for online applications: Text retrieval, extraction and categorization Chapter 1: Natural language processing. John Benjamins Publishing Company 2002; 1-17.

Karuna G., Akshith M., Dinesh PS., Vardhan BV., Bisht YS., Narsaiah MN. Automated abstractive text summarization using deep learning. 15th International Conference on Materials Processing and Characterization (ICMPC 2023), September 5- 8 2023, England.

Kemaloğlu Alagöz N. Automatic text summarization with deep learning. Süleyman Demirel University Institute of Science and Technology, PhD. Thesis, Isparta, Turkiye, 2022.

Kemik Doğal Dil İşleme Grubu 2020; http://www.kemik.yildiz.edu.tr/. (Access Date: 11.02.2025)

Lewis M., Liu Y., Goyal N., Ghazvininejad M., Mohamed A., Levy O., Stoyanov V., Zettlemoyer L. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461. 2020.

Mallik A., Kumar S. Word2Vec and LSTM based deep learning technique for context-free fake news detection. Multimedia Tools and Applications 2024; 83(1): 919-940.

Mcculloch WS., Pitts W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 1943; 5: 115–133.

Minaee S., Kalchbrenner N., Cambria E., Nikzad N., Chenaghlu M., Gao J. Deep learning--based text classification: a comprehensive review. ACM computing surveys (CSUR) 2021; 54(3): 1-40.

Mwandau B. Investigating keystroke dynamics as a two-factor biometric security. Strathmore University, Faculty of Information Technology, PhD. Thesis, Nairobi, Kenya, 2018.

Raffel C., Shazeer N., Roberts A., Lee K., Narang S., Matena M., Zhou Y., Liu PJ. Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research 2020; 21(140): 1-67.

Russel SJ., Norvig P. Artificial intelligence: A modern approach. New Jersey: Pearson Education Publishing; 2013.

Türkiye Bilimsel ve Teknolojik Araştırma Kurumu. 2000; https://trdizin.gov.tr. . (Access Date: 11.02.2025)

Shafiq N., Hamid I., Asif M., Nawaz Q., Aljuaid H., Ali H. Abstractive text summarization of low-resourced languages using deep learning. PeerJ Computer. Science 2023; 9: e1176.

Shaik T., Tao X., Li Y., Dann C., McDonald J., Redmond P., Galligan L. A review of the trends and challenges in adopting natural language processing methods for education feedback analysis. IEEE Access 2022; 10: 56720-56739.

Shopnil TA., Das A., Saha D. Automatic text summarization using deep learning methods. Authorea Preprints 2024.

Singh S., Mahmood A. The NLP cookbook: Modern recipes for transformer based deep learning architectures. IEEE Access 2021; 9: 68675–68702.

Sun G., Wang Z., Zhao J. Automatic text summarization using deep reinforcement learning and beyond. Information Technology and Control 2021; 50(3): 458-469.

Tahseen R., Omer U., Farooq MS., Adnan F. Text summarization techniques using natural language processing: A systematic literature Review. VFAST Transactions on Software Engineering 2021; 9(4): 102-108.

Teng K., Qiang B., Wang Y., Yang X., Wang Y., Wang C. Abstractive text summarization model based on BERT vectorization and bidirectional decoding. International Conference on Cyber Security, Artificial Intelligence, and Digital Economy (CSAIDE 2023), March 3-5 2023, China.

Teufel S. Argumentative zoning: information extraction from scientific text. University of Edinburgh, PhD. Thesis, Scotland, 1999.

Teufel S., Carletta J., Moens M. An annotation scheme for discourse-level argumentation in research articles. Ninth Conference of the European Chapter of the Association for Computational Linguistics, June 8-12 1999, 110-117, Norway.

Teufel S., Moens M. Discourse-level argumentation in scientific articles: Human and automatic annotation. In Proceedings of the ACL99 Workshop on Standards and Tools for Discourse Tagging 1999; 10: 977035-977051.

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez AN., Kaiser Ł., Polosukhin I. Attention is all you need. 31st International Conference on Neural Information Processing Systems, December 4 – 9 2017, Cambridge.

Veyssier L. Abstractive text summarization model in Keras. 2020; https://github.com/LaurentVeyssier/Abstractive-Text-Summarization-model-in-Keras. (Access Date: 11.02.2025)

Zhang J., Zhao Y., Saleh M., Liu P. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. 37th International Conference on Machine Learning, July 13-18 2020, Austria.

Zhou J., Ye Z., Zhang S., Geng Z., Han N., Yang T. Investigating response behavior through TF-IDF and Word2vec text analysis: A case study of PISA 2012 problem-solving process data. Heliyon 2024; 10(16): e35945.