

# Improving the Quality of the Relevance of the Search for Scientific Publications Based on a Combination of Ranking Methods

Oleksandr Tarasov<sup>a</sup>, Liudmyla Vasylieva<sup>a</sup>, Svitlana Turlakova<sup>b</sup>, Pavlo Sahaida<sup>a</sup>, Aleksandr Karpov<sup>a</sup>

<sup>a</sup> Donbass State Engineering Academy, Akademichna str. 72, Kramatorsk, 84313, Ukraine

<sup>b</sup> Institute of Industrial Economics of the NAS of Ukraine, Marii Kapnist Str. 2, Kyiv, 03057, Ukraine

## Abstract

The study aims to develop a combined method for finding relevant documents using automated Nature Language Processing (NLP), which improves the quality of results based on informative text parameters. To adapt the search method to the text type, it is necessary to combine several algorithms for processing sets of texts. Thus, results of a search in sets of documents should provide the analysis of separate components of the text, especially interesting for the analyst. To achieve this goal, the authors: conducted a brief review of the status of the issue in the field of NLP and full-text search; classified the methods and means of word processing available to Data Science specialists; chose the most promising characteristics of the texts in terms of the purpose of the work. The effectiveness of joint ranking of specialized articles devoted to the solution of specific scientific issues and articles-reviews of the subject area of large volume in terms of their relevance to the analyst's request was studied. The possibility of using the reference text chosen by the analyst as a basic standard for querying and searching and ranking similar scientific and technical documents is considered. Experiments with text sets were performed, which allowed confirming the informativeness of the parameters of text objects selected by the authors and the proposed composition of algorithms for their processing based on a combination of TF-IDF method and relevance ranking method using the distance between term occurrences.

## Keywords

Nature language processing, methods of relevance ranking, information retrieval

## 1. Introduction

Scientific activity is impossible without the analysis of scientific literature in the relevant field of research, because many studies are conducted simultaneously in different countries. Every year the amount of information increases, so in the results of search engines on the Internet the number of links is estimated at hundreds of thousands.

The search for relevant scientific and technical documents by keywords remains an important problem of modern information support of research and educational process. For additional analysis of search results and targeting a large amount of information, browsers provide some features, such as using the logical operator OR when determining keywords in queries. But analysts also need to have additional tools that can be configured and quickly weed out unnecessary information, as well as show what information may be relevant to the topic. To do this, analysts use such an indicator of search quality as the relevance of the text [1, 2]. Relevance is determined in different ways, some of

---

ICST-2023: Information Control Systems & Technologies, September 21-23, 2023, Odesa, Ukraine.

EMAIL: alexandrtar50@gmail.com (O. Tarasov); vasileva.dgma@gmail.com (L. Vasylieva); svitlana.turlakova@gmail.com (S. Turlakova); pavlo.sahaida@gmail.com (P. Sahaida); alexandr.karpov.work@gmail.com (A. Karpov)

ORCID: 0000-0002-0493-1529 (O. Tarasov); 0000-0002-9277-1560 (L. Vasylieva); 0000-0002-3954-8503 (S. Turlakova); 0000-0002-4700-8160 (P. Sahaida); 0000-0003-3901-2992 (A. Karpov)



© 2023 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

the methods are not disclosed by their developers as intellectual property, so research to improve the quality of relevance assessment methods are actual in the field of information retrieval.

For information support of the process of ranking scientific and technical texts, many different parameters are used to describe the properties of text fragments. The ranking of texts depending on the analyst's request is usually based on a model, which provides a comprehensive assessment of the relevance of the entire document and its fragments. To evaluate the model, the assessment of expert groups in the subject area is often used. The ranking accuracy is determined by comparing the results of manual and automatic classification of text sets. At the same time, there are problems with obtaining consistent expert assessments. It must be free from unintentional distortions of the results and misunderstanding of the scientific and technical value of the article. But, with experts' help, classification is a costly undertaking that requires a lot of time to carry out.

The methods for searching for relevant documents with keywords specified by the analyst are cheaper, faster, and free from the above drawbacks. The methods for training classification models are methods for searching for relevant documents with keywords specified by the analyst.

Despite significant advances in full-text search, in the automation of obtaining annotations of documents, the task of extracting the necessary documents for the analyst from large sets of texts remains an important area of research. The disadvantage of existing solutions is the ranking of documents, in terms of request conditions, the totality of document content. This often does not take into account the importance of individual components of the text, which may have a very high relevance and, accordingly, interest the analyst more than whole texts that have questionable relevance due to their volume or quality of presentation. Therefore, the purpose of this study is to develop a method for finding relevant scientific and technical documents using automated nature language processing, which improves the quality of results based on the use of text informative parameters.

## **2. Work related analysis**

Today, there are many methods that use different approaches to analyze the text. However, the effectiveness of these methods is influenced by the type of specific problem to be solved [3]. But even today there are no methods that could clearly and effectively solve the problem of finding relevant information, despite the fact that many researchers are working on the problem of improving the quality of information classification methods and algorithms for its search and analysis [4]. When analyzing search results, the task is to highlight relevant and irrelevant sources. This raises a number of questions related to the quantitative assessment of the relevance of articles. Relevance assessment is performed by various methods that use different principles of text analysis [5, 6].

The most researched approach to presenting text for its further analysis, clustering and categorization is the use of many pre-selected features and representation of each text object (fragment) as a vector in the multidimensional space of these features [7]. The simplest approach is the bag-of-words model [8], in which each word of the document is considered as a coordinate in the features space, taking two possible values {0, 1}. Problems of identifying informative features and optimizing the process of their selection are discussed in [9]. Several studies have been carried out to ensure effective convolution of expert assessments [10], including convolution based on a fuzzy measure [11].

Linguistic analysis extracts the necessary information from the text to answer the questions asked about the text. Among the methods of such text analysis are the analysis of features using the evaluation of Yule, models of resource semantics (PropBank), and FrameNet, combined methods [12-14].

Statistical analysis deduces certain patterns in determining the location or sequence of words in the text. Statistical analysis is often referred to as a subtype of linguistic analysis. Among the methods of statistical analysis of the text can be distinguished TF-IDF [15], BM25 [16], weight by word pairs [17].

The TF-IDF method [18] and the Confident Weights method [19] during text processing are used to assign the weight of each word from a set of keywords (or for all words in the document) in the range [0, 1]. These weights can be calculated by fairly complex algorithms that determine the

relevance of the word to a particular document. For example, weights can be assigned depending on the lexical distance between words in a document, as proposed in [20].

Consider the algorithms for applying these methods. Their joint application, in our opinion, is promising for solving the tasks set in the framework of this work. The TF-IDF method is a calculation of how important certain words are for a document in relation to other documents [15]. When the term is often used in a particular document, but rarely in other documents, it has great significance for this document. The method includes two parts: determining the TF and IDF values for an individual document and a set of documents.

TF (term frequency - word frequency) is the ratio of the number of occurrences of a word to the total number of words in the document, which is determined by formula (1):

$$tf = \frac{n_k}{n_d}, \quad (1)$$

where  $n_k$  is the number of specified keywords found in the text or document,  $n_d$  – the total number of words in the text.

IDF (inverse document frequency) is the frequency inversion with which a word occurs in collection documents. IDF is determined by formula (2).

$$idf = \log \frac{D}{d}, \quad (2)$$

where  $D$  is the total number of documents in the collection,  $d$  is the number of documents in which the keywords occur.

If you combine these two estimates, you can calculate the weight for each keyword in the document [15]. The weighing model TF-IDF is defined as the product of  $tf$  and  $idf$ , given by formula (3):

$$TFIDF = tf \cdot idf. \quad (3)$$

The method of estimating relevance based on weights for word pairs (WPW) calculates the distance between the same keywords in the text, and based on the ratio of the average distance to the total number of words in the document determines the relevance of the text [17]. The average distance between the same keywords is calculated by formula (4):

$$D = \frac{\sum d}{n_d}, \quad (4)$$

where  $d$  is the distance (number of words) between the same keywords;  
 $n_d$  – the number of obtained distances.

Relevance for the keyword is calculated by formula (5):

$$R_{kw} = 1 - \frac{D}{w}, \quad (5)$$

where  $D$  is the average distance between the same keywords;  
 $w$  – the number of words in the text.

The relevance index  $R_d$  for all keywords for a particular document is calculated by formula (6):

$$R_d = \frac{\sum R_{kw} \cdot k}{n_{kw}}, \quad (6)$$

where  $k$  is the keyword's value factor (if required);  
 $n_{kw}$  – the number of keywords.

The higher the  $R_d$  value, the more relevant the article is considered.

For a collection of documents, relevance is determined by formula (7):

$$R_{kw} = \frac{\sum R_d}{c}, \quad (7)$$

where  $c$  is the total number of documents in the collection.

After pre-processing a set of texts, during which a characteristic vector is formed for each of the documents according to the selected features, the documents are divided into clusters (for the case of unsupervised approach) or category (if classification model or sample given). Cluster analysis collects data and organizes them into groups according to certain characteristics. To do this, in the mode of unsupervised learning using a large set of known machine learning techniques: k-nearest neighbors algorithm, Bayes Classifier, and others [21, 22]. In supervised learning mode, support vector machine (SVM) [23], latent semantic analysis (LSA) [24], Neural Network as classification methods and

others are used for ranking documents. Categorization of scientific and technical articles in the light of the above works is a special task, the solution of which is complicated by specific terminology, the presence of graphic and tabular materials, and a large number of abbreviations [25]. In addition, documents (articles) differ in their forms: a brief report on the study, a full-scale study, review, popular science article. Ranking documents in such conditions often gives an unstable result with questionable relevance.

Wu et al. [26] proposed a model of information retrieval based on a relevance score by modeling how a person makes decisions about the relevance of a document. They used related terms in the document in the context of search terms to suggest a new search method. Three principles for combining relevance scores from different parts of a document, which are used to make the final decision on the relevance of a document to a search query, were presented by Kong et al. [27].

The value of the relevance index allows the analyst to break scientific publications into groups with different levels of relevance. This raises the question of the boundaries of relevance. This question is solved by experts in the subject area of scientific publications in accordance with the tasks of the search. In addition, it is necessary to separately consider the relevance of the document to the topic of the request (to the keywords) and the relevance to a specific issue within the studied topic. This requires the use of different groups of keywords that are important for the topic as a whole and allow you to highlight individual issues within this topic. When using several methods for assessing relevance at the same time, it is necessary to ensure consistency of the assessment results obtained by different methods.

We will use in this work the main criteria that characterize information retrieval. The search completeness criterion (Recall ratio) shows how complete the result was given and is calculated by the formula (8):

$$Rc = \frac{a}{a + c}, \quad (8)$$

where  $a$  is the number of relevant documents issued in response to an information request or by keywords;  $c$  is the number of relevant documents in the collection not issued by the system.

The loss factor (Silence ratio) is related to the recall ratio and shows how many documents relevant to a given query or keywords are not retrieved by the search engine. It is calculated by the formula (9):

$$Q = 1 - Rc. \quad (9)$$

The search accuracy coefficient (Precision ratio) shows how accurately the results of relevant documents correspond to the experts' results. This coefficient is calculated by the formula (10):

$$QD = \frac{a}{a + b}, \quad (10)$$

where  $a$  is the number of relevant documents produced by the search engine in response to the information request;  $b$  is the number of irrelevant documents produced by the search engine. The coefficient of search accuracy is related to the coefficient of search noise. Search noise (Noise ratio) is the search engine's output of documents irrelevant to a given query or keywords. It is calculated by the formula (11):

$$S = 1 - D. \quad (11)$$

### 3. Case study

Two basic analysis methods were chosen for the study, which is used to assess the relevance of texts in a set of previously found articles:

- TF-IDF method as a generally accepted method for assessing relevance;
- word pair weighting (WPW) relevance search method as a modification and refinement method.

On their basis, the authors proposed a combined method for assessing the relevance of articles. Consider an algorithm for the combined application of the two methods. First, we get a primary selection of articles, for example, as a result of a search query on the Internet. Then we successively apply the first and the second algorithms for assessing relevance to this list and find the values of the rating scores for all elements in the list. These algorithms sort the found articles by rating according to

their rating models. In the resulting primary list, we select the number of articles for consideration depending on the problem being solved (we restrict the list). At the next stage, we define a group of articles that are included in the lists of relevant articles obtained as a result of the independent work of both methods. We also determine the remaining number of relevant articles of the limited list (which are not included in the general group of relevant articles for the methods). The assessment of the consistency of relevant source definition by two methods can be the relation:

$$K = N_d / C_g, \quad (12)$$

where  $N_d$  is the number of articles identified as relevant by both methods,  $C_g$  is the total number of relevant articles found by different methods in the collection.

The combined method, based on TF-IDF and scales by word pairs, will be defined as the average value of relevance calculated by the methods of TF-IDF and scales by word pairs. These two methods rely on value different ranges when assessing relevance, so we need to bring them to a common range of values.

Bringing the values to the common range for the TF-IDF relevance assessment method is performed as normalization with respect to the maximum (13):

$$R_{tfidf} = \frac{r_{tfidf} - r_{tfidf\_min}}{r_{tfidf\_max} - r_{tfidf\_min}}, \quad (13)$$

where  $0 \leq R_{tfidf} \leq 1$ ;

$r_{tfidf}$  – an indicator of the article's relevance according to the TF-IDF method;

$r_{tfidf\_max}$  – the maximum relevance score according to the TF-IDF method, at which an article or a set of documents is considered relevant;

$r_{tfidf\_min}$  – the minimum relevance score in the set.

For the method of assessing relevance using weights by pairs of words, the normalization is performed similarly to (14):

$$R_{wpw} = \frac{r_{wpw} - r_{wpw\_min}}{r_{wpw\_max} - r_{wpw\_min}}, \quad (14)$$

where  $0 \leq R_{wpw} \leq 1$ ;

$r_{wpw}$  – an indicator of the relevance of the article using the WPW method;

$r_{wpw\_max}$  – the maximum relevance score based on the WPW method, at which an article or set is considered relevant;

$r_{wpw\_min}$  – the minimum relevance score in the set.

Then we substitute the normalized values into formula (15) and obtain an assessment of article relevance based on the combined use of two assessment methods.

The relevance index  $Rs$  for the combined method, based on TF-IDF and scales by word pairs can be determined by formula (16), if there are preferences in the use of methods for a particular search:

$$Rs = \frac{R_{tfidf} + R_{wpw}}{2}, \quad (15)$$

$$Rs = \frac{k_1 \cdot R_{tfidf} + k_2 \cdot R_{wpw}}{2}, \quad (16)$$

where  $k_i$  – situational coefficients of confidence in the quality of the methods.

In this study, expert assessments were used to be able to evaluate the results of the selected methods of relevance assessment. Expert data is obtained by processing the ratings of articles by experts whose publications are related to the selected topic.

The comparison will be based on the effectiveness of information retrieval. The efficiency of information retrieval - evaluation of the quality of search in information retrieval systems. The main criteria characterizing the information search are the completeness of the search, search accuracy, information loss, and search noise (formulas (8) – (11)).

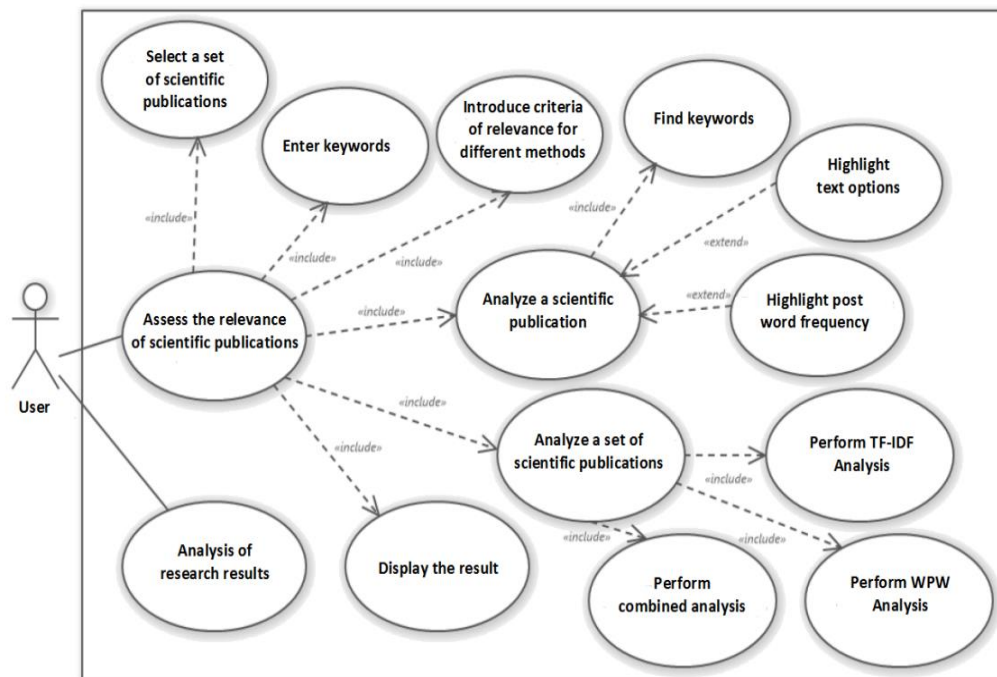
According to the developed research plan, a thematic area was selected and keywords were allocated for it: "search", "relevance", "key". According to these keywords, twenty scientific publications were found in the Google Scholar search engine. Expert analysis was then conducted to determine the relevant publications for this collection. Relevance indicators for the analyzed relevance assessment methods were also determined (presented in Table 1). To get the results of the

selected methods of assessing the degree of relevance, a software product was created, the main functions of which are presented using the precedence diagram in Figure 1.

**Table 1**

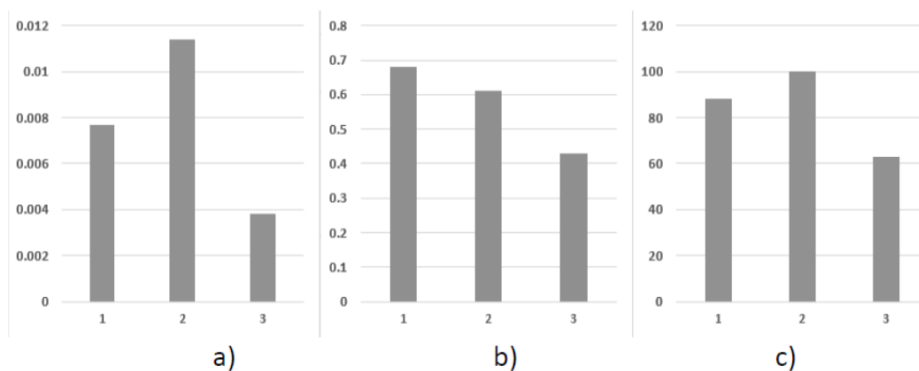
Accepted values of the relevant evaluation criteria for the selected methods

Evaluation/Method	TF-IDF	WPW	Combined method
Bad	$R_m < 0,004$	$R_m < 0,3$	$R_m < 0,3$
Average	$0,004 < R_m < 0,01$	$0,3 < R_m < 0,5$	$0,3 < R_m < 0,5$
Good	$R_m > 0,01$	$R_m > 0,5$	$R_m > 0,5$



**Figure 1:** Diagram of precedents for the research process of relevance assessment methods

This set of publications was analyzed for relevance to keywords with the help of the created computer application "Article Analyzer". The choice of keywords has a significant impact on the relevance of the articles in the search. The results of the relevance assessment for each keyword, which were obtained using the selected methods, are presented in Figure 2. The methods used show different word relevance and the degree of influence on the relevance score of the articles, which is due to the different models of article relevance assessment.



**Figure 2:** Relevance scores for each keyword (1 – relevance, 2 – search, 3 – keyword), obtained by: a) TF-IDF method, b) WPW method; c) combined method

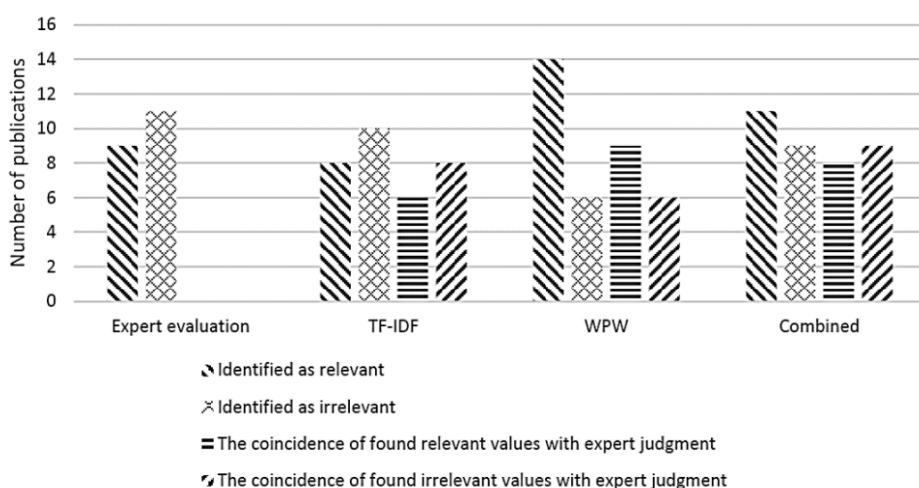
After obtaining the results of the "Article Analyzer", we can compare the accuracy of the results of the selected methods with the expert evaluation using the criteria for the effectiveness of information retrieval. Comparative results of finding relevant publications by the methods are presented in Table 2.

**Table 2**  
Results of determination of relevant and irrelevant publications by the selected methods

Methods	Identified as relevant	Identified as irrelevant	The coincidence of found relevant values with expert judgment	The coincidence of found irrelevant values with expert judgment
Expert evaluation	9	11	–	–
TF-IDF	8	10	6	8
WPW	14	6	9	6
Combined	11	9	8	9

The results of the publication relevance assessment for each method are presented in **Ошибка! Источник ссылки не найден..**

Values of information retrieval efficiency estimates for the selected methods based on the criteria, which are determined by formulas (8, 9, 15, 16), are shown in **Ошибка! Источник ссылки не найден..**



**Figure 3:** Number of found relevant publications for each method

**Table 3**  
Information search efficiency criteria for the selected methods

Methods / Performance criteria	TF-IDF	WPW	Combined method
Recall ratio $R_c$	0,67	1,00	0,89
Silence ratio $Q$	0,33	0	0,11
Precision ratio $D$	0,75	0,64	0,73
Noise ratio $S$	0,25	0,36	0,27

According to the analysis of the results, we can see that the combined method was the most accurate, because it has the greatest accuracy of search query results and the least search noise. A more complete search result is given by the WPW-based relevance method, because it has the least

loss of information, but also the greatest search noise. All three algorithms produced more than half of all relevant publications that matched the expert opinion.

To increase the recall ratio  $R_c$  by the combined method, one of the solution options is to change the ranges for the values of the relevance evaluation criteria  $R_m$  for the applied methods (Table 1). Harmonization of the ranges for determining the relevant articles by TF-IDF and WPW methods is also necessary to exclude contradictory results as much as possible.

To test the conclusions of the first set of articles for which the relevance scores were determined, the second set of articles was selected. It was devoted to one of the technical issues in the field of mechanical engineering: the processes of severe plastic deformation of metals. The articles were selected from Internet searches based on a number of keywords highlighted by experts (ultrafine, grained, plastic, deformation, materials). In addition, an additional group of keywords related to the search topic was selected for analysis. This makes it possible to assess the relative importance of keywords for assessing relevance using different methods.

The articles were selected into groups based on several specific issues related to the same topic. These articles were treated as one group. The next group included several articles that focused on reviews of research in the field, results achieved, and forecasts of technology development. These articles had a different structure in that they addressed some issues related to the research topic in each article.

The adopted TF-IDF, WPW, and combined methods were applied to these two groups. The articles from the two groups were then combined into one mixed group and the analysis was repeated. The results obtained for each keyword are shown in Table 4.

**Table 4**

Keyword relevance values determined for the selected evaluation methods and groups of articles with different structures

Methods for assessing the relevance	Keywords				
	ultrafine	grained	plastic	deformation	materials
<b>The first set of articles exploring selected issues</b>					
TF-IDF	0,0015	0,0012	0,0043	0,0190	0,0141
WPW	0,47	0,43	0,84	0,95	0,95
Combined method	54,4	49,2	71,7	100,0	100,0
<b>The second set of review articles analyzing the results achieved (review articles)</b>					
TF-IDF	0,0009	0,0008	0,0036	0,0100	0,0145
$Del_{TF-IDF\ 1-2}$	0,67	0,50	0,19	0,90	-0,03
WPW	0,67	0,66	0,95	0,97	0,98
$Del_{WPW\ 1-2}$	-0,30	-0,35	-0,12	-0,02	-0,03
Combined method	54,6	53,8	68,0	100,0	100,0
$Del_{Comb\ 1-2}$	0,00	-0,08	0,06	0,00	0,00
<b>A third set of articles, including articles from the previous two groups</b>					
TF-IDF	0,0017	0,0013	0,0046	0,0246	0,0171
$Del_{TF-IDF\ 1-3}$	-0,13	-0,08	-0,07	-0,29	-0,21
WPW	0,51	0,48	0,86	0,96	0,96
$Del_{WPW\ 1-3}$	-0,08	-0,10	-0,03	0,00	-0,01
Combined method	58,7	56,5	73,7	100,0	100,0
$Del_{Comb\ 1-3}$	-0,08	-0,15	-0,03	0,00	0,00

Relative change in word relevance scores for the  $i$ -th method and three groups of articles, determined by the formula:



$$Del_{i,1-j} = (R_{i,1} - R_{i,j})/R_{i,1},$$

where  $R_{i,1}$  is the word relevance value for the  $i$ -th method and the first group of articles,  $R_{i,j}$  is the word relevance value for the  $i$ -th method and the  $j$ -th group of articles,  $j = 1,2,3$ .

The analysis of changes in the relevance parameter for the first and second groups of articles showed that the scores of articles devoted to the solution of specific issues when using the TF-IDF method are generally higher than the scores of review articles. The WPW method for assessing relevance showed opposite results.

In this case, the scores for the articles devoted to the analysis of the accumulated information on the selected topic, in general, are higher than for the articles devoted to the solution of specific issues. The results make it possible to conclude that the combined application of these methods can increase the reliability of the assessment of the relevance of scientific articles if they have a different structure.

The selection of keywords is of great importance. For the correct selection of relevant sources, it is advisable to define two groups of keywords. One of them will define the general topic of the search, and the second one will define the specific study. At the same time, the second group of keywords can be used to rank the publications in the primary search results.

As noted, an indicator of method consistency is the ratio of the number of relevant articles that are selected by the two methods to the total number of relevant articles in the set (formula 12). By changing the number of articles after ranking in one of the lists, it is possible to change this quality indicator to increase the consistency of the scores.

## 4. Discussion

The simultaneous application of two methods to one set of articles showed that each of the considered methods has certain advantages. For example, the initial frequency analysis of articles in the TF-IDF method allows us to determine the keywords for which we should continue the search. Therefore, the task is to find the conditions for the most effective joint application of methods to improve search efficiency and assess the relevance of articles.

At the first stage, a comparison was made between TF-IDF, WPW and expert assessments. This was done in order to evaluate the efficiency of these algorithms in terms of recall ratio and precision ratio of the search. Based on the analysis of the selected set of articles and their expert assessments, the ranges of criteria values for the selected methods were adopted, which divide the high, medium and low level of relevance of the articles.

After analyzing the results of the application of two separate methods, as well as the use of the combined method, it was found that the use of individual methods and their combination does not allow to fully select those articles that, in the opinion of experts, are relevant. However, the application of the considered methods significantly reduces the time for assessing the relevance of articles, which makes it possible to recommend them for practical application.

Then we selected two more sets of articles that are relevant to the topic under study, but have a different structure of information presentation. The first group of articles was devoted to the study of individual issues in the subject area. The second group was devoted to the synthesis of information on the topic and contained reviews of the research. The difference between these articles is that review articles include material on a number of aspects of the consideration of the issue under study. Information on individual issues is concentrated in separate sections that make up the article as a whole. They are naturally relevant in the opinion of the experts. However, in this case, the use of frequency methods may not show that the articles are relevant, due to the small relative number of keywords in the total volume of the text. The WPW method allows you to highlight the parts of review articles that are devoted to the questions of interest and have an appropriate density of keywords in a separate part of the general text.

A way to improve the consistency of the methods in this case can be to expand the range of relevance values (for TF-IDF), so that frequent methods also do not exclude diverse articles from the list of relevant sources. In this case, the percentage of sources that both methods have determined to be relevant can be normalized. This allows you to control the process of highlighting relevant sources by setting the percentage of losses. The rule for expanding the range can be a such condition that the

group of relevant sources, determined by both the first and the second methods (formula 12), should include, for example, 95% of articles.

In general, the use of two or more methods, taking into account the peculiarities of the structure of scientific articles, will make it possible to more reasonably highlight relevant publications in the study of individual issues related to the general research topic. In particular, differences in the evaluations of articles by different methods can make it possible to single out and group articles with different structures of information presentation. For example, it is advisable to separate some articles that are devoted to specific issues and analysis of research carried out over a certain period, experimental and theoretical research, etc.

## 5. Conclusion

Based on the analysis and classification of text analysis methods, two methods for assessing relevance were selected: the TF-IDF method and the assessment method based on WPW. The possibility of their joint use as a combined method for assessing the relevance of articles was also considered. Based on the use of dependencies for the criteria for evaluating the effectiveness of information retrieval, a software-methodical complex was built to study the assessment of relevance by the selected methods.

A selection of a set of publications was made, and their expert assessment was obtained per the search topic, based on which a comparative analysis of the results of the implemented methods was made. It is advisable to rank articles by relevance sequentially using more general and specific terms within the search topic. A study of the relevance of the first set articles was carried out, according to which it was found that the proposed combined method for assessing relevance has a precision ratio value of  $D = 0,73$  at a search noise ratio of up to  $S = 0,27$ , recall ratio  $Rc = 0,89$  and silence ratio only  $Q = 0.11$ . This result improves the precision ratio for the WPW method by 12% and recall ratio of the search for the TF-IDF method by 25%.

From this, we can conclude that for tasks requiring an increase in the reliability of a search result, the proposed combined method for assessing relevance can be used.

An algorithm is proposed for the joint application of two methods and for assessing the consistency of determining the relevant sources by two methods. The algorithm allows us to control the degree of consistency of the methods for assessing relevance based on changing the range of values of the relevance of the methods. This makes it possible to take into account the peculiarities of the models for determining the relevance by different methods when considering articles with different structures of information presentation.

The analysis of determining the relevance of keywords for different methods and types of articles showed that the estimates for articles with different structure for the same method can vary widely, which should be taken into account when using them.

It is advisable to rank articles by relevance sequentially using at first more general and then more specific terms within the search topic.

The combined approach has some potential to expand the functionality of search management based on the application of various methods. The combined application of the methods will make it possible to use their advantages more fully for the search for scientific articles. The development of this algorithm requires a clearer statistical justification of the threshold values for the formation of groups of articles with different relevance.

## 6. References

- [1] D. Gunawan, C. A. Sembiring, M. A. Budiman, The Implementation of Cosine Similarity to Calculate Text Relevance between Two Documents, *Journal of Physics: Conference Series*, 978 1 (2018) 012120. doi:10.1088/1742-6596/978/1/012120.
- [2] D. E. Losada, J. Parapar, A. Barreiro. When to stop making relevance judgments? A study of stopping methods for building information retrieval test collections, *Journal of the Association for Information Science and Technology*, 70 1 (2019) 49-60.

- [3] S. Buttcher, C. L. Clarke, G. V. Cormack, *Information retrieval: Implementing and evaluating search engines*. Mit Press, 2016.
- [4] S. Ranaei, A. Suominen, A. Porter, T. Kässi, *Application of text-analytics in quantitative study of science and technology*. Springer Handbook of Science and Technology Indicators, 2019.
- [5] X. Li, Y. Liu and J. Mao, Understanding the role of human-inspired heuristics for retrieval models. *Frontiers of Computer Science*, 16 (2022) 1-11.
- [6] A. Kanwal, A. W. Septyanto, M. H. G. Muhammad, R. A. Said, M. Farrukh and M. Ibrahim, Adaptively Intelligent Meta-search Engine with Minimum Edit Distance, in: 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2022, pp. 1-7. doi: 10.1109/ICBATS54253.2022.9759088.
- [7] M. D. Shermis, J. C. Burstein, (Eds.) *Automated essay scoring: A cross-disciplinary perspective*. Routledge, 2003. doi: 10.4324/9781410606860.
- [8] T. Joachims, *Learning to classify text using support vector machines*. Springer Science & Business Media, 2002. doi:10.1007/978-1-4615-0907-3.
- [9] Yi Bong-Jun, L. Do-Gil, R. Hae-Chang, The Effects of Feature Optimization on High-Dimensional Essay Data, *Mathematical Problems in Engineering*, 2015 (2015). doi:10.1155/2015/421642.
- [10] O. Tarasov, L. Vasylieva, O. Altukhov, V. Anosov, Automation of the synthesis of new design solutions based on the requirements for the functionality of the created object, in: *Nine International Conference: Information Control Systems Technologies (ICST-2020)*, volume 2711, CEUR-WS.org, 2020, pp. 161–175. URL: <http://ceur-ws.org/Vol-2711/paper13.pdf>.
- [11] P. Sahaida, "Model and Method of Processing Partial Estimates During Intelligent Data Processing Based on Fuzzy Measure," in: *2020 IEEE KhPI Week on Advanced Technology (KhPIWeek)*, 2020, pp. 114-118, doi: 10.1109/KhPIWeek51551.2020.9250134.
- [12] G. Yule, *The statistical study of literary vocabulary*. Cambridge: Cambridge University Press, 1944.
- [13] M. Palmer, P. Kingsbury, D. Gildea, The Proposition Bank: An Annotated Corpus of Semantic Roles, *Computational Linguistics*, 31 (2005) 71–106. doi:10.1162/0891201053630264.
- [14] C. J. Fillmore, C. R. Johnson, M. R. Petruck, Background to FrameNet, *International journal of lexicography*, 16 3 (2003) 235–250.
- [15] Q. Shahzad, R. Ali, Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents, *International Journal of Computer Applications*, 181 (2018) 25–29. doi:10.5120/ijca2018917395.
- [16] S. Jimenez, S. P. Cucerzan, F. A. Gonzalez, A. Gelbukh, G. Dueñas, BM25-CTF: Improving TF and IDF factors in BM25 by using collection term frequencies, *Journal of Intelligent and Fuzzy Systems*, (2018) 2887-2899. Doi:10.3233/JIFS-169475.
- [17] D. Hawking, P. Thistlewaite, Relevance weighting using distance between term occurrences, *ANU Research Publications*, (1996) 20. URL: <http://hdl.handle.net/1885/40762>.
- [18] G. Salton, C. Buckley, Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing and Management*, (1988) 513–523.
- [19] P. Soucy, Mineau G. W. Beyond, TFIDF Weighting for Text Categorization in the Vector Space Model, in: *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, 2005, pp. 1130–1135.
- [20] C. Buckley, G. Salton, J. Allan, Automatic retrieval with locality information using SMART. in: D. K. Harman, editor, *Proceedings of TREC-1*, Gaithersburg MD, November 1992, 1992, pp. 59-72.
- [21] L. M. Rudner, T. Liang, Automated essay scoring using Bayes' theorem, *The Journal of Technology, Learning, and Assessment*, 1 2 (2002) 1–22.
- [22] C. Liang, B. Zhang, X. Gong, M. Li, H. Guo, R. Li, A Survey of Automated Essay Scoring System Based on Naive Bayes Classifier, in: *Artificial Intelligence in Education and Teaching Assessment*, 2021, pp. 247-250.
- [23] H. Chen, B. He, T. Luo, B. Li. A ranked-based learning approach to automated essay scoring, in: *Proceedings of the 2nd International Conference on Cloud and Green Computing (CGC '12)*, IEEE, 2012, pp. 448–455. doi:10.1007/978-3-319-52836-6\_40

- [24] P. W. Foltz, D. Laham, T. K. Landauer, The intelligent essay assessor: applications to educational technology, *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, 1 2 (1999) 939-944.
- [25] S. Valenti, F. Neri, A. Cucchiarelli, An overview of current research on automated essay grading, *Journal of Information Technology Education* 2 (2003) 319-330. doi: 10.28945/331.
- [26] H.C. Wu, R.W. Luk, K.F. Wong, K. Kwok, A retrospective study of a hybrid document-context based retrieval model. *Inf. Process. Manag.* 43 5 (2007) 1308–1331.
- [27] Y.K. Kong, R. Luk, W. Lam, K.S. Ho, F.L. Chung, Passage-based retrieval based on parameterized fuzzy operators, in: *The SIGIR 2004 Workshop on Mathematical/Formal Methods for Information Retrieval* (2004).