https://jurnaleeccis.ub.ac.id/
e-ISSN(Online): 2460-8122

Jurnal EECCIS 19, No. 3, December 2025
pp 190-196

Implementation of Semantic Similarity for Book Search in Digital Library

M. Royhan Daffa [1], Muhammad Ainul Yaqin [1]
[1] Department of Informatics Engineering, Maulana Malik Ibrahim State Islamic University of Malang, Malang, Indonesia
Email: royhandf@gmail.com, yaqinov@ti.uin-malang.ac.id

***Abstract*—This study discusses improving the relevance of book searches in digital libraries by applying semantic similarity. This study aims to improve search relevance and contextual understanding by integrating the TF-IDF method for keyword selection and Wu Palmer for WordNet-based similarity measurement. Tests were conducted with 25 different queries on 25,000 digital books in five scenarios. The test results show that the most optimal scenario is a combination of 3 TF-IDF terms, which produces the highest average similarity score among semantic methods, ranging from 87.52% (1-word queries) to 52.4% (5+ word queries). It is evident that increasing the number of TF-IDF terms (5 or 10) or removing them consistently decreases relevance. In comparison, non-semantic literal search achieves 100% match for short queries but fails completely in understanding the context for longer queries (down to 53.5%). This study proves that the combination of the top 3 TF-IDF terms with semantic similarity is the most balanced and effective approach in improving the relevance of search results, especially for short to medium queries. These results contribute to the development of smarter and more contextual digital library search systems.**

***Index Terms*—Semantic similarity, digital library, TF-IDF, Wu Palmer, book search, WordNet.**

I. INTRODUCTION

Libraries have transformed from physical storage institutions to digital information service providers in the digital era. Libraries serve as a force in preserving and disseminating knowledge and culture that develops along with human information needs [1]. However, libraries face significant challenges in managing and presenting the ever increasing digital information.

One of the main challenges digital libraries facing is the increasing amount of information available. As a result, users often have difficulty finding books that suit their needs amidst the abundance of information. Conventional keyword-based library systems usually fail to understand the context and meaning behind user search queries, resulting in less accurate and relevant results [2]. This becomes a problem when the keywords entered differ from the terms used in the book's contents, even though both have the same meaning.

To overcome this problem, applying semantic similarity in book search systems is crucial. Semantic similarity allows the system to understand the meaning and context of a search query based on keyword matching and through metadata or concepts from the book. Book metadata such as title, author, and keywords can be used to find books relevant to user queries [3]. This approach can increase the relevance of search results and make it easier for users to find the information they need.

This research aims to improve the relevance between queries and book search results in digital libraries through the use of semantic similarity. This study integrates the TF IDF method for keyword weighting with the Wu Palmer method for measuring semantic similarity. This integration of these two methods is an innovative approach that has not been widely explored in the context of digital libraries and is expected to produce more accurate and contextual searches.

Book search systems in digital libraries must consider both keyword matching and semantic relationships between concepts. This approach can broaden search coverage and improve the user experience in finding relevant information. Furthermore, this study uses book metadata and tables of contents as processed data for search and leverages WordNet to improve semantic understanding.

In previous literature, the concept of semantic similarity has been widely studied and applied in various research contexts. For example, the TF-IDF method was applied to weight words in ranking Arabic documents [4]. On the other hand, the Wu-Palmer method was used to measure the degree of conceptual similarity between entities. Its application has been tested in comparing sentence meanings to understand semantic similarity [5] and in analyzing business process models [6].

The novelty of this research lies in the integration of TF IDF and Wu Palmer to address weaknesses in the system's semantic understanding and the noise that arises from the lack of keyword selection. This study applies TF-IDF to determine the most representative keywords before Wu Palmer expands the search conceptually. This combination is a key contribution, enabling the system to understand the meaning of queries more deeply. This approach is believed to significantly improve the relevance of search results compared to literal search systems.

This research is expected to make practical contributions to the development of more effective search systems in digital libraries. By understanding the context of queries, the system directly assists users who struggle to find books due to a lack of specific keywords, enabling them to locate relevant information more quickly and intuitively. This not only speeds up the information retrieval process but also significantly improves the quality and scope of contextually relevant results. The results of this research can serve as a basis for libraries to implement semantic similarity technology, thereby improving the quality of their information retrieval services.

II. RESEARCH METHODOLOGY

The data source used is a collection of 25,000 English

language digital books sourced from the website https://www.oapen.org. The selection of English-language data was a technical necessity, based on the Wu Palmer method's reliance on the WordNet lexical database, which ensures accurate semantic similarity calculations. The data processed for each book consists of metadata and the table of contents.

The query matching process specifically prioritizes text from the book's metadata as the primary basis for calculations. However, if a description in the metadata is unavailable, the system uses the table of contents as an alternative data source. This approach ensures that each book in the collection can be consistently analyzed for relevance to the query.

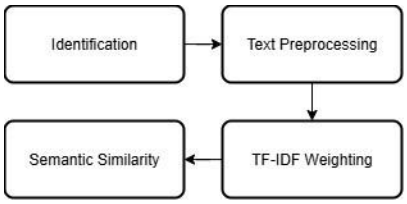

Fig. 1. Research process

The research process, as illustrated in Figure 1, includes identification, text preprocessing, TF-IDF weighting, and semantic similarity calculations. Standard deviation is then used in the analysis stage to measure variation in the similarity results.

The system's workflow begins when a user enters a search query. The query undergoes standard preprocessing steps, including case folding, tokenization, stopword removal, and lemmatization, to produce a clean set of keyword tokens. The system then systematically iterates through each book in the dataset.

For each book, text from the metadata (or table of contents if necessary) is extracted, and TF-IDF scores are calculated for all the words within it. Based on the applicable test scenario, the system then selects N terms (3, 5, or 10) with the highest TF-IDF scores as the core semantic representation of the book. Keyword tokens from the query are then compared to these N selected terms using the Wu Palmer method, where all resulting similarity scores are aggregated and averaged to produce a single final similarity score for the book. Once this iteration process is complete for all books, the system sorts all books by their final similarity scores in descending order and presents the sorted list as search results to the user.

A. Identification

Identification is the first step in the sentence similarity calculation process. This stage involves identifying the query and book metadata. If the book has a description, it will be used. However, if the description is unavailable or empty, the book's table of contents will be used as an alternative. This process ensures that the most relevant and available data is processed. Figure 2 shows the flowchart of the book identification process.

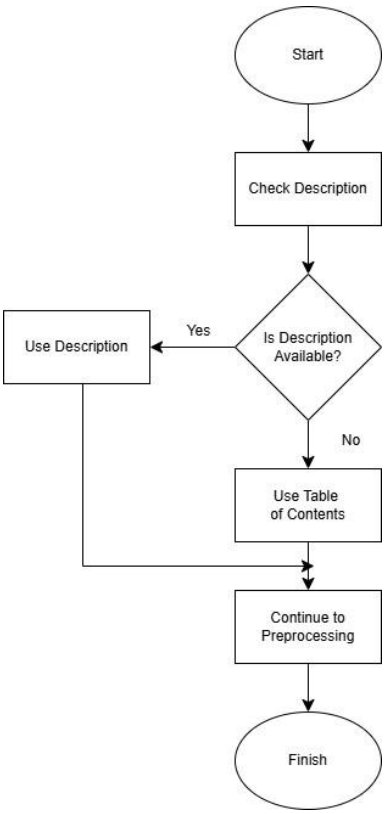

Fig. 2. Flowchart checking description

B. Text Preprocessing

Text preprocessing is an important stage in text data processing. At this stage, the text data will be transformed into a more suitable form, producing high-quality text information that is ready for use in the following process [7]. Text Preprocessing has four stages: case folding, tokenizing, stopword removal, and lemmatization.

1. Case Folding

Case folding is the process of converting capital letters to lowercase and removing punctuation marks, such as periods, commas, and single quotation marks. This process is important in text processing because it helps simplify the text and ensures consistency during further analysis.

2. Tokenization

Tokenization is the process of breaking down sentences into smaller units, known as tokens. Typically, this process uses spaces as separators to mark word boundaries. This allows longer phrases to be broken down into simpler, more easily analyzed parts.

3. Stopword Removal

Stopword removal removes frequently occurring words that lack significant meaning. Each tokenized word is compared against a list of stopwords, which are common words deemed to provide no helpful information. Examples of stopwords include "a", "and", "that", "for", "to," and others. These words are removed to help focus the analysis on more informative words.

4. Lemmatization

The last stage in text preprocessing is lemmatization. Lemmatization is the process of changing words into their basic forms. Unlike stemming, which simply cuts off endings, lemmatization considers the form of parts of speech such as nouns, verbs, adjectives, and adverbs to get an accurate basic

form. For example, the verb "running" can be changed to "run". This process ensures that word variations can be recognized during analysis.

C. TF-IDF Weighting

TF-IDF (Term Frequency Inverse Document Frequency) is a method used to measure the importance of a word in a sentence in the context of a larger collection of sentences [8]. TF-IDF has two components: TF (Term Frequency) and IDF (Inverse Document Frequency) values.

The TF value measures the frequency of a word's occurrence in a document. Since the length of each document varies, the TF value generally divides the number of occurrences of a word by the total number of words in the document, as shown in equation (1):

$$\text{tf} = \frac{\text{number of word occurrences in the document}}{\text{total number of words in the document}} \quad (1)$$

Meanwhile, the IDF value measures the importance of a word in the overall document corpus. The smaller the IDF value, the lower the word's importance, as it appears in many documents. The IDF is calculated using equation (2):

$$\text{idf} = \log\left(\frac{\text{number of documents in the corpus}}{\text{number of document containing the word}}\right) \quad (2)$$

The TF-IDF value is obtained by multiplying the TF value and the IDF value, as shown in equation (3):

$$\text{tfidf} = \text{tf x idf} \quad (3)$$

By combining these two values, TF-IDF can produce a high score that reflects a word's relevance to a document within the context of the entire corpus. This method assigns a low weight to words that are commonly found across many documents. Conversely, words that are specific to a document receive a high weight [9].

Words weighted using TF-IDF are then further analyzed to calculate semantic similarity. This weighting aids the book search process by highlighting the most relevant words. In book searches, TF-IDF functions to identify the most representative keywords. This enables the system to find books relevant to the user's query. In its implementation, the TF-IDF calculation utilizes the Scikit learn library in Python, using default parameters, where L2 normalization is automatically applied to the resulting vector.

D. Semantic Similarity

Semantic similarity is a measure of how close or similar the meanings are between two words, phrases, or documents. This concept can measure the similarity of meanings in the context of short texts [10]. In Natural Language Processing (NLP), semantic similarity is crucial because it is utilized in various applications, including information retrieval systems, plagiarism detection, and sentiment analysis [11].

To handle words with multiple meanings, this system does not just take a single meaning, but instead takes all possible meanings (synsets) of each word being compared. The system then calculates a Wu Palmer similarity score for all possible meaning pairs between the two words. It selects the one with the highest similarity score as the final result. This approach ensures that the strongest semantic relationship between two words is identified, rather than being limited to the most common meaning.

In this study, semantic similarity was calculated using the Wu Palmer method. This method measures the similarity between two concepts by considering the depth of nodes in the same hierarchical structure and the relationships between their edges [12].

$$sim_{wup}(w_1, w_2) = \frac{2\ x\ depth\ (LCS)}{depth(w_1) + depth\ (w_2)} \quad (4)$$

In equation (4), the LCS (Lowest Common Subsumer) is the smallest common subsumer between the first and second words. The terms depth(w1) and depth(w2) represent the depths of the first and second words in WordNet. The Wu Palmer calculation results range from 0 to 1, indicating how closely the two words are semantically related. This approach considers the semantic meaning of hypernyms, hyponyms, and meronyms [13].

The Wu Palmer method was chosen as the semantic similarity measure in this study because it does not require an additional corpus for statistical calculations, unlike Resnik's, making it more practical to implement. Meanwhile, more sophisticated models such as BERT were not chosen due to their high computational requirements, which are beyond the scope of this study. Therefore, Wu Palmer was considered a suitable choice for this study.

E. Standard Deviation

Standard deviation is a statistical measure that indicates how spread out or varied the data in a dataset are. The larger the standard deviation, the greater the variation of the data from the mean. A smaller standard deviation indicates that the data are more clustered or consistent around the mean [14].

This study uses standard deviation to measure the consistency of the results obtained from semantic similarity measurements. By calculating the standard deviation, researchers can assess the extent to which these results vary, which provides insight into the stability and reliability of the applied method. The standard deviation is calculated using equation (5):

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \quad (5)$$

Where:

- σ = Population standard deviation
- N = Number of data points in the population
- x_i = *i*-th data value
- μ = Population mean

III. Results and Discussion

The search system was tested using 25 different queries, ranging from simple to complex. Each query was processed against 25,000 digital books. Table 1 presents the complete list of queries used in the test, organized by the number of words in each query.

TABLE I
LIST OF EXPERIMENTAL QUERIES

Experiment	Query
1-word	law community economy conflict education
2-word	public policy media systems artificial intelligence supply chain world history
3-word	world history international trade regulation war and genocide philosophy of language political resistance movements gender and society
4-word	gender and society international law and policy digital platforms and innovation regeneration and moral identity financial risk management techniques education for social justice
5+-word	human rights law and global regulation globalization impact on community and society governance and democracy in global crisis strategic investment in emerging markets sustainable ecosystem management for clean energy

This study analyzed each query in five different scenarios to evaluate the effectiveness of various methods for calculating semantic similarity. The goal was to identify the approach that produced the most relevant and optimal results. The following sections detail the analysis of the results for each scenario:

A. Semantic Similarity with 3 TF-IDF Terms

The first scenario calculated semantic similarity involving the top 3 terms from the TF-IDF weighting. This approach is designed to test the hypothesis that concise keyword selection can provide a strong semantic signal without introducing noise. The average similarity results and standard deviation are presented in Table 2.

TABLE II
RESULTS OF SCENARIO 1 (3 TF-IDF TERMS)

Query	Average Similarity	Standard deviation
1-word	87,52%	13,9%
2-word	70,85%	27,05%
3-word	68,84%	27,91%
4-word	60,1%	31,55%
5+-word	52,4%	35,05%

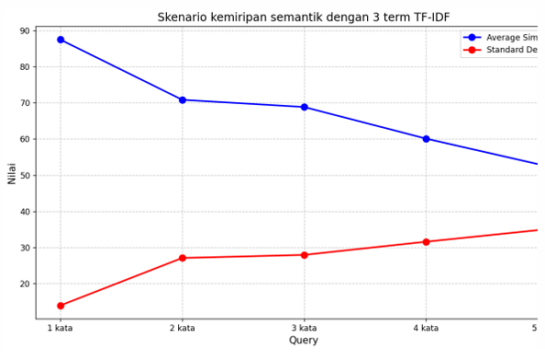

Fig. 3. Graph of the similarity and 3 terms

Figure 3 shows a graph of the average similarity values and standard deviations presented in Table 2. The highest average similarity was found for 1-word queries (87.52%), while the lowest was found for queries with 5+ words (52.4%). While effective for short, queries with broad meanings, system performance declined for more specific queries. One reason for this is the potential for term mismatches.

Conversely, the standard deviation increased from 13.9% to 35.05% as query length increased. This increase indicates less consistent results for complex queries. This may occur because different books may only match a subset of the words in a long query, leading to significant variation in scores across search results. Thus, this scenario proved optimal for focused queries, but faced consistency challenges for longer and more complex queries.

B. Semantic Similarity with 5 TF-IDF Terms

The second scenario was designed to examine the impact of expanding the keyword base on search relevance. In this test, the number of keywords was increased to 5 terms TF IDF. The primary goal was to determine whether a broader keyword set could more comprehensively capture the context of the book or instead introduce less relevant terms. The results of this experiment are presented in Table 3, which displays the mean similarity and standard deviation for various query lengths.

TABLE III
RESULTS OF SCENARIO 2 (5 TF-IDF TERMS)

Query	Average Similarity	Standard deviation
1-word	76,93%	19,39%
2-word	62,25%	29,52%
3-word	59,96%	26,75%
4-word	51,57%	31,52%
5+-word	43,35%	33,44%

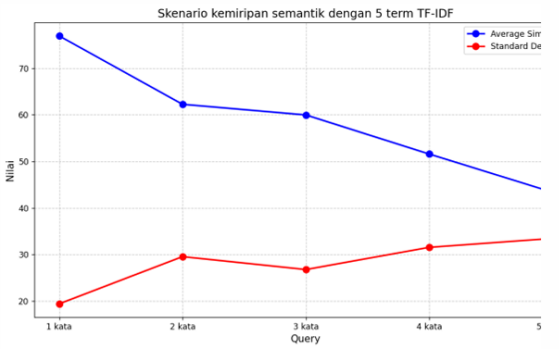

Fig. 4. Graph of the similarity and 5 terms

Figure 4 visualizes the data from Table 3, showing the mean similarity values and standard deviations for each query length. The mean similarity value decreases with increasing query length, from 76.93% (1 word) to 43.35% (5+ words). This decrease is due to the complexity of longer queries, making it more difficult to find a full semantic match for the five TF-IDF terms from the book.

Conversely, the standard deviation increases from 19.39% to 33.44%, indicating greater variation in results for long queries. However, there is a slight decrease in the standard deviation for the three-word query (26.75%), indicating that three words may be sufficient to form a more stable context for matching the five book terms. Overall, these results suggest

that increasing the number of terms to five does not effectively improve relevance and may actually introduce less representative terms, thus degrading search performance compared to Scenario 1.

C. Semantic Similarity with 10 TF-IDF Terms

The third scenario tests the upper bound of keyword expansion, aiming to maximize semantic coverage. This is achieved by calculating similarity using 10 TF-IDF terms. This test evaluates whether a maximalist approach to keyword selection improves performance or degrades search quality due to the introduction of significant noise from less important terms. The calculation results for this scenario are shown in Table 4.

TABLE IV
RESULTS OF SCENARIO 3 (10 TF-IDF TERMS)

Query	Average Similarity	Standard deviation
1-word	63,41%	28,67%
2-word	52,19%	30,4%
3-word	47,13%	30,9%
4-word	43,46%	30,5%
5+-word	37,26%	31,01%

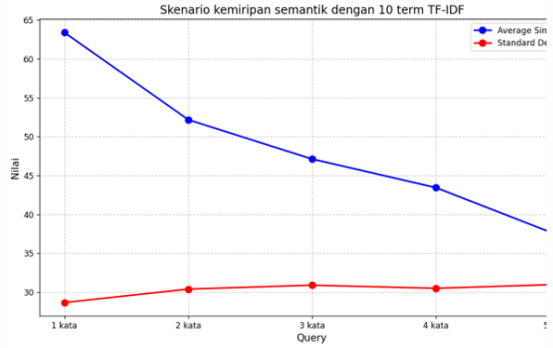

Fig. 5. Graph of the similarity and 10 terms

Figure 5 shows a visualization of the average similarity and standard deviation results from Table 4. It can be seen that the average similarity again experienced a significant decrease, from 63.41% (1 word) to 37.26% (5+ words). At the same time, the standard deviation remained relatively high and stable, within the range of 28% to 31%. This decrease confirms that increasing the number of terms to 10 does not guarantee increased relevance, especially for highly complex queries. This occurs because many of the 10 terms taken from the book are likely not representative of the specific context of the query. A high and stable standard deviation indicates that the score variation between search results is consistently considerable, not necessarily a uniformly good level of relevance. It can be concluded that using 10 terms actually introduces too much noise from less important words, causing the system to lose focus and significantly degrade search performance.

D. Semantic Similarity without TF-IDF

The fourth scenario serves as a control test to demonstrate the crucial role of keyword selection. In this scenario, the TF-IDF selection mechanism is wholly removed, and the system compares the query against all words in the book's metadata. This test aims explicitly to measure the extent of performance degradation without relevant keyword filtering. The results of the similarity calculations and standard deviations for this test are presented in Table 5.

TABLE V
RESULTS OF SCENARIO 4 (WITHOUT TF-IDF)

Query	Average Similarity	Standard deviation
1-word	42,14%	30,31%
2-word	34,63%	28,15%
3-word	31,79%	27,71%
4-word	28,77%	27,16%
5+-word	25,6%	27,13%

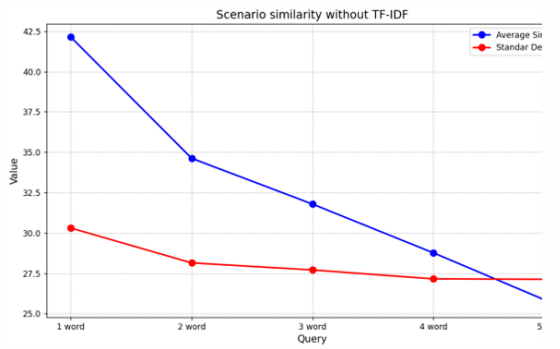

Fig. 6. Graph of the similarity without TF-IDF

Figure 6 visualizes the average similarity and standard deviation results from Table 5. The average similarity score was the lowest of all semantic scenarios, dropping dramatically from 57.33% (1 word) to 29.1% (5+ words), while the standard deviation remained relatively stable at around 27-28%. This low similarity score clearly demonstrates the crucial role of TF-IDF as a filter.

Without selection, the system becomes less selective and includes many irrelevant, common words in the calculation. This results in a low proportion of relevant words to the total words in the metadata, drastically reducing the focus on matching meaning. The stability of the standard deviation does not indicate good results, but rather indicates that the noise from common words tends to be evenly distributed throughout the book. These results demonstrate that the keyword selection process using TF IDF is a fundamental step, and without it, the relevance of search results becomes meaningless.

E. Literal Search

The fifth scenario uses a literal search, which operates without semantic interpretation or term weighting. This approach was intentionally included to serve as a crucial baseline or control group in the study. With this baseline, the performance of the semantic-based scenario can be more effectively measured to demonstrate its added value. The system mechanism in this scenario calculates a percentage score based on the word match ratio between the query and metadata, ensuring no variation in scores that would preclude the calculation of standard deviation. The results of the average match calculation for this scenario are presented in Table 6.

TABLE VI
RESULTS OF SCENARIO 5 (LITERAL SEARCH)

Query	Average Similarity
1-word	100%
2-word	100%
3-word	75,5%
4-word	61,67%
5+-word	53,5%

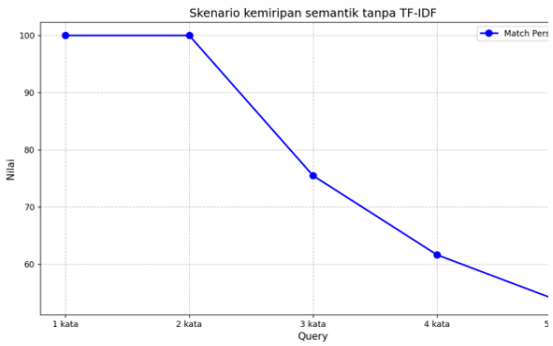

Fig. 7. Graph of the literal search

Figure 7 visualizes the data from Table 6. Perfect matches (100%) are achieved for 1- and 2-word queries, but drop sharply to 75.5% (3 words) and reach 53.5% for queries with 5+ words. The high scores for short queries occur because common words are easily found explicitly in the metadata. In contrast, for long queries, the probability of finding all words exactly is very low, resulting in only partial match scores. These results highlight a fundamental weakness of non-semantic search: while effective for simple, exact-match queries, it fails to comprehend context, synonyms, or semantic relationships, rendering it unreliable for complex search needs.

F. Analysis

Of the five scenarios tested, Scenario 1 (using 3 TF-IDF terms) was the most optimal approach among the semantic methods evaluated. With an average similarity of 87.52% on single-word queries, this scenario consistently outperformed Scenarios 2 (76.93%), 3 (63.41%), and 4 (57.33%). Although Scenario 5 (literal search) achieved 100% matching on short queries, this advantage was only literal and failed in understanding context, which was the primary objective of this study.

A comparison between Scenarios 1, 2, and 3 revealed a significant trend: increasing the number of TF-IDF terms consistently decreased relevance. This demonstrates that TF-IDF's primary function in this framework is to act as a filter, reducing noise. Using 3 terms proved sufficient to capture the semantic core, while 5 or 10 terms began to introduce less relevant words and interfere with the calculation. The failure of Scenario 4, which eliminated TF-IDF and produced the lowest similarity score, provides the most substantial evidence for the importance of the keyword selection stage.

In terms of consistency, scenario 1 also showed the lowest standard deviation for short queries (13.9%), indicating more stable results compared to other semantic scenarios. Therefore, it can be concluded that the combination of 3 terms TF-IDF with Wu Palmer provides the best balance between relevance, understanding meaning, and search focus, effectively addressing the research problem formulation.

G. Practical Implications

The practical implications of this research are its potential to improve the quality of information retrieval across the digital library ecosystem. This method is not limited to one type of system. However, it can be embedded in various platforms to enhance retrieval, such as institutional repositories, digital archives, online public access catalogs, and other similar platforms.

Its implementation can be realized as a server-side (backend) semantic search module, which serves as an advanced search option. However, its application to Indonesian-language data collections requires technical adaptations, particularly the replacement of English WordNet with Indonesian, as well as adjustments to other linguistic tools, such as stop words and lemmatization. Nevertheless, this study successfully demonstrated the effectiveness of the proposed framework and can serve as a foundation for building more contextually aware searches across various digital library services.

IV. CONCLUSION

This study successfully demonstrated that the application of semantic similarity, especially when combined with the top 3 TF-IDF terms (Scenario 1), effectively improves the relevance of book search results in digital libraries. The system is able to understand the meaning relationship between words, not just perform literal matching. This is reflected in the high relevance results, where scenario 1 consistently provides the highest similarity score among other semantic methods, with a peak in single-word queries reaching 87.52%.

However, it was found that the system's effectiveness is affected by query length. Shorter queries tend to yield higher and more consistent relevance, while longer and more complex queries can potentially decrease relevance. This suggests that the combination of TF-IDF and Wu Palmer provides an optimal solution for improving conceptual relevance, but challenges remain in handling multi-layered queries.

The research problem formulation aims to test semantic similarity methods to improve search relevance. Because the tested method produces a similarity score as a representation of relevance, the most logical evaluation is to measure this output directly. Therefore, this study uses the average similarity score to assess the method's effectiveness in capturing conceptual relationships, and the standard deviation to measure the consistency of the results. This evaluation approach differs from standard metrics such as precision, recall, and F1-score, which require the availability of a ground truth dataset. Since this study focuses on the implementation and analysis of the semantic similarity method itself, rather than on validation against external references, the process of creating a ground truth dataset falls outside the established scope.

A priority step for future research is to conduct a more comprehensive evaluation. This can be achieved by first building a ground truth dataset, which allows for measuring system performance using standard information retrieval metrics such as precision, recall, and F1-score. Once baseline

metrics have been validated, further exploration can be directed at improving the semantic model by implementing word embeddings (e.g., Word2Vec or BERT), testing more sophisticated weighting schemes such as BM25, and applying Topic Modeling (LDA) for more abstract conceptual analysis. Finally, larger-scale validation involving real users will provide important insights into the system's performance in the real world.

REFERENCES

[1] B. Ayu *et al.*, "Perkembangan dan Peran OPAC Pada Aplikasi CIP (Cerah Informasi Pustaka) Untuk Temu Kembali Informasi di Perpustakaan Universitas Tridinanti Palembang," *Jurnal Multidisipliner Bharasumba*, vol. 2, no. 04, pp. 296–315, Oct. 2023.

[2] Rauzatunnisa and Suherman, "Sistem Pengolahan Bahan Pustaka dan Dampaknya Terhadap Temu Balik Koleksi di Dinas Perpustakaan dan Kearsipan Kabupaten Aceh Utara," *LIBRIA*, vol. 14, no. 1, pp. 86–97, Jul. 2022.

[3] E. Lacic *et al.*, "Evaluating Tag Recommendations for E-Book Annotation Using a Semantic Similarity Metric," Aug. 2019.

[4] S. Anggraini, D. Purwitasari, and A. Z. Arifin, "Pengukuran Kemiripan berbasis Leksikal dan Semantik untuk Perangkingan Dokumen Berbahasa Arab," *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 4, no. 2, pp. 134–145, Aug. 2022.

[5] G. Ulta Abriania and M. A. Yaqin, "Implementasi Metode Semantic Similarity untuk Pengukuran Kemiripan Makna antar Kalimat," *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 1, no. 2, pp. 47–57, Dec. 2019.

[6] S. M. Pamungkas, I. Aqimuddin, C. Gunawan, M. A. Yaqin, and Abd. C. Fauzan, "Analisis Kemiripan Model Proses Bisnis PMBoK dan Scrum menggunakan Metode Jaccard Coefficient Similarity dan Semantic Similarity," *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 5, no. 2, pp. 53–64, Aug. 2023.

[7] S. Khairunnisa, Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, pp. 406–414, 2021.

[8] S. Ferdiana Kusuma, T. Badriyah, P. Wibowo, R. Faradisa, and S. Huda, "Optimalisasi Fitur Pencarian Pada E-Catalog Menggunakan Query Expansion Dan Algoritma TF-IDF," *Techno. Com*, vol. 22, no. 3, pp. 703–711, 2023.

[9] N. Putri Husain, A. Febriana Syam, and R. Mustikosari, "Analisis Sentimen Ulasan Pengguna Tiktok pada Google Play Store Berbasis TF-IDF dan Support Vector Machine," *Journal of System and Computer Engineering*, vol. 5, no. 1, pp. 91–102, Jan. 2024.

[10] D. Jatnika, M. A. Bijaksana, and A. A. Suryani, "Word2Vec Model Analysis for Semantic Similarities in English Words," *Procedia Comput. Sci.*, vol. 157, pp. 160–167, Jan. 2019.

[11] H. Ezzikouri, Y. Madani, M. Erritali, and M. Oukessou, "A New Approach for Calculating Semantic Similarity between Words Using WordNet and Set Theory," *Procedia Comput. Sci.*, vol. 151, pp. 1261–1265, Jan. 2019.

[12] T. T. M. Phan, A. T. Duong, and P. T. M. Tran, "WordNet-based Computation of Semantic Similarity Between Two Vietnamese Nouns," *Technium Soc. Sci. J.*, vol. 43, p. 532, 2023.

[13] F. Husein Wattiheluw and R. Sarno, "Developing Word Sense Disambiguation Corpuses Using Word2vec and Wu Palmer for Disambiguation," *In 2018 International Seminar on Application for Technology of Information and Communication*, pp. 244–248, Nov. 2018.

[14] W. W. Pribadi, A. Yunus, and A. S. Wiguna, "Perbandingan Metode K-Means Euclidean Distance dan Manhattan Distance pada Penentuan Zonasi Covid-19 di Kabupaten Malang," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 493–500, Aug. 2022.