



KWSSWS-IA Approach: An Intelligent Assistant for Scientific PDF Retrieval Using Semantic Analysis and AI-Driven Summarization

Fakroun, E.

ebtfakroon@cit.edu.ly

The College Of Industrial Technology, Misrata, Libya

Sullabi, M.

m.sullabi@it.misuratau.edu.ly

Misurata University

Received: 10/8/2026 Revised: 13/8/2026 Accepted: 05/9/2026 Publication: 05/10/2026

Abstract

In the age of excessive information, traditional web search engines frequently fall short in providing accurate, contextually relevant results for scientific research. Furthermore, in addition to metadata-based retrieval, this study presents an enhanced version of the Keyword Search System on Web Searching (KWSSWS), which has evolved into an Intelligent Assistant (KWSSWS-IA) capable of deep semantic analysis of full-text PDF documents. The proposed system uses transformer-based artificial intelligence (AI) models related to natural language processing (NLP) to obtain and analyze internal document components, including title, abstract, keywords, and conclusions. It automatically generates brief, accurate summaries which is associated with ranks results according to scientific relevance rather than keyword frequency, in addition to recognizing significant scientific entities like authors, publication year, journal, and domain-specific terms KWSSWS-IA presents the following results on a single screen: download link, AI-generated summary, keyword match supplied by the user, publication year, title, and authors, and source journal. Therefore, ordinary search interfaces just offer hyperlinks. Experimental evaluation reveals significant improvements in precision, recall, and utilizer efficiency compared to standard Google Scholar searches. In addition, researchers who remain short on time and seeking high-quality, domain-relevant literature may find the strategy particularly useful.

Keywords : Intelligent search assistant, PDF retrieval, semantic analysis, summarization, NLP.

المخلص

في عصر وفرة المعلومات، غالبًا ما تعجز محركات البحث التقليدية على الإنترنت عن تقديم نتائج دقيقة ومُراعية للسياق في البحث العلمي. تُقدّم هذه الورقة نسخة مُحسّنة من نظام البحث بالكلمات المفتاحية على الإنترنت (KWSSWS)، الذي تطوّر ليصبح مساعدًا ذكيًا (KWSSWS-IA) يتجاوز الاسترجاع القائم على البيانات الوصفية ليجري تحليلًا دلاليًا عميقًا لنصوص ملفات PDF الكاملة. يستخرج النظام المقترح ويحلّل مكونات الوثيقة الداخلية، بما في ذلك العنوان والمخلص والكلمات المفتاحية والاستنتاجات، باستخدام معالجة اللغة الطبيعية (NLP) ونماذج الذكاء الاصطناعي القائمة على المُحوّلات. يُؤلّد النظام تلقائيًا ملخصات دقيقة وموجزة؛ ويحدّد الكيانات العلمية الرئيسية مثل المؤلفين وسنة النشر والمجلة والمصطلحات الخاصة بالمجال؛ ويُرتّب النتائج حسب أهميتها العلمية بدلًا من مجرد تكرار الكلمات المفتاحية. على عكس واجهات البحث التقليدية التي تعرض روابط فقط، يقدم نظام KWSSWS-IA بطاقات نتائج منظمة ومتكاملة، تعرض رابط التحميل، ومطابقة الكلمات المفتاحية التي أدخلها المستخدم، وملخصًا مؤلّدًا بواسطة الذكاء الاصطناعي، وسنة النشر، والعنوان، والمؤلفين، والمجلة المصدر، كل ذلك على شاشة واحدة. تُظهر التقييمات التجريبية تحسينات ملحوظة في الدقة والاستدعاء وكفاءة المستخدم مقارنةً بعمليات البحث القياسية في جوجل سكولار. يُعدّ هذا النظام ذا قيمة خاصة للباحثين الذين يسعون للحصول على مراجع عالية الجودة وذات صلة بمجالهم ضمن قيود زمنية.

INTRODUCTION

A traditional literature reviews remain becoming more and more unfeasible due to the exponential rise of scientific publications. Although Google Scholar which is associated with similar services provide extensive coverage, they do not comprehend document semantics and instead mostly rely on surface-level signals like title matches or citation counts [1] and [3]. As a result, consumers frequently filter out results that aren't relevant, wasting valuable research time [14]. The original KWSSWS system [4] was improved in this study to become KWSSWS-IA, an intelligent search aid that views each PDF as a semantic knowledge unit rather than a static file. The approach begins with dependable full-text parsing of PDF documents to extract systematic scientific content

while preserving the semantic links of the original material [5]. Furthermore, analyzing this gathered data through AI-based comprehension engines powered via domain-specialized language models like SciBERT enables deep contextual understanding beyond basic text matching. The system then generates query-adaptive, dynamic summaries that extract relevant insights based on the specific informational needs of the utilizer [6] as well as [7]. Semantic coherence is prioritized over mere lexical matching in the framework's relevance-aware ranking system, which evaluates the conceptual alignment of content-related queries instead of traditional keyword matching. [8], [9]. In addition, these developments spark a paradigm change that turns web search from a basic link-retrieval function into an intelligent knowledge curation

process that actively gathers, analyzes, as well as presents content with scientific significance.

2. Related Work

Although previous systems have investigated clustering, , for instance, K-means in WEKA [5] and [8], which remain linked to metadata filtering [6], few of them examine the semantics of entire documents. Scientific text can now be deeply understood thanks to recent developments in NLP, especially transformer architectures [7]. Abstractive summaries remain produced using BART which is associated with T5 models [9], while tools , for instance, SPECTER [8] and [21] embed publications into semantic spaces. Nevertheless, these remain still isolated in research prototypes. This gap is filled via KWSSWS-IA, which incorporates these features into a useful, end-to-end retrieval pipeline. Conventional web search engines , for instance, Google Scholar give priority to measures related to page rank, citation counts, and keyword matching [1], [2]. Though useful for general discovery, they frequently fall short of meeting scholarly demands. In contexts with limited bandwidth or resources, like those in Libya, utilizers are forced to manually sort through a variety of results, which causes cognitive overload and inefficiency [3]. Domain-aware search algorithms have been offered via researchers as a solution. Semantic embeddings were utilized via Semantic Scholar [4] to enhance paper suggestions, which is associated with Microsoft Academic utilized knowledge graphs to disambiguate entities [5]. Nonetheless, these systems continue to display results as hyperlinks, necessitating external navigation for utilizers to evaluate relevance, a crucial bottleneck in research jobs that require quick turnaround times. The transition to full-text processing from metadata-only processing is a major breakthrough in scientific retrieval. Even for legacy publications, tools , for instance, GROBID [6] allow for highly accurate structured parsing of PDFs into titles, authors, abstracts, and references. In a similar vein, CERMINE [7] as well as ScienceParse [8] utilize machine learning to extract semantic material, establishing the foundation for more in-depth document comprehension. Models for named entity recognition (NER) have been refined for utilize in scholarly fields. For instance, fuller indexing is made possible via SciSpacy [9], which identifies items , for instance, publications, institutions, and procedures. This tendency is supported via the utilize of custom NER, which goes one step further via removing post-click validation as well as incorporating extracted

metadata straight into a single result interface. When contextual language models were developed, relevance score was completely transformed. More domain-specific semantics remain captured via SciBERT [10], which was trained on 1.14 million scientific documents, than via generic BERT. Tasks like classification, query-document matching, and citation prediction have all seen effective implementations of it [11]. The cosine similarity between question as well as document embeddings in SciBERT space is utilized to calculate relevance in KWSSWS-IA, a method that has been demonstrated to be more successful in scientific retrieval than TF-IDF or BM25 [12]. Unlike Google Scholar, which relies on surface-level signals, this approach directly tackles the ambiguity problem cited in previous studies via ensuring that "clustering in neurology" is not mistaken with "clustering in networking" [13].

Retrieval and comprehension remain connected through summarization [31]. Abstractive models that learn to reconstruct important information from entire texts, for instance, T5 [15] as well as BART [14], produce succinct, fluid summaries. BART-large-CNN has demonstrated strong ROUGE-L performance in scientific fields. [29] adjusted for abstract-conclusion pairings [16]. A significant innovation is BART for on-the-fly summary production, which provides distilled insights inline rather than requiring utilizers to read entire papers or abstracts outside. The literature on human-computer interaction [17] emphasizes utilizer-centric design concepts, which state that lowering cognitive burden enhances decision accuracy as well as speed.

A holdover from early web searches, the list-of-links paradigm is utilized via the majority of academic search engines. This methodology is criticized in recent research as being out of date for experienced utilizers who want quick, comparative evaluation [18]. Research Rabbit and Litmaps, on the other hand, show relationships but do not extract deep material. A structured card-based interface, pioneered via KWSSWS-IA, shows the download link, matched keywords, AI summary, year, title, authors, journal, which is associated with the key terms in a single view. Modern e-commerce as well as news aggregators serve as inspiration for this design, which is specially tailored to academic operations. Such interfaces cut work completion times via more than 60%, according to empirical research [19], which the proposed system evaluation supported.

A. Proposed System Architecture

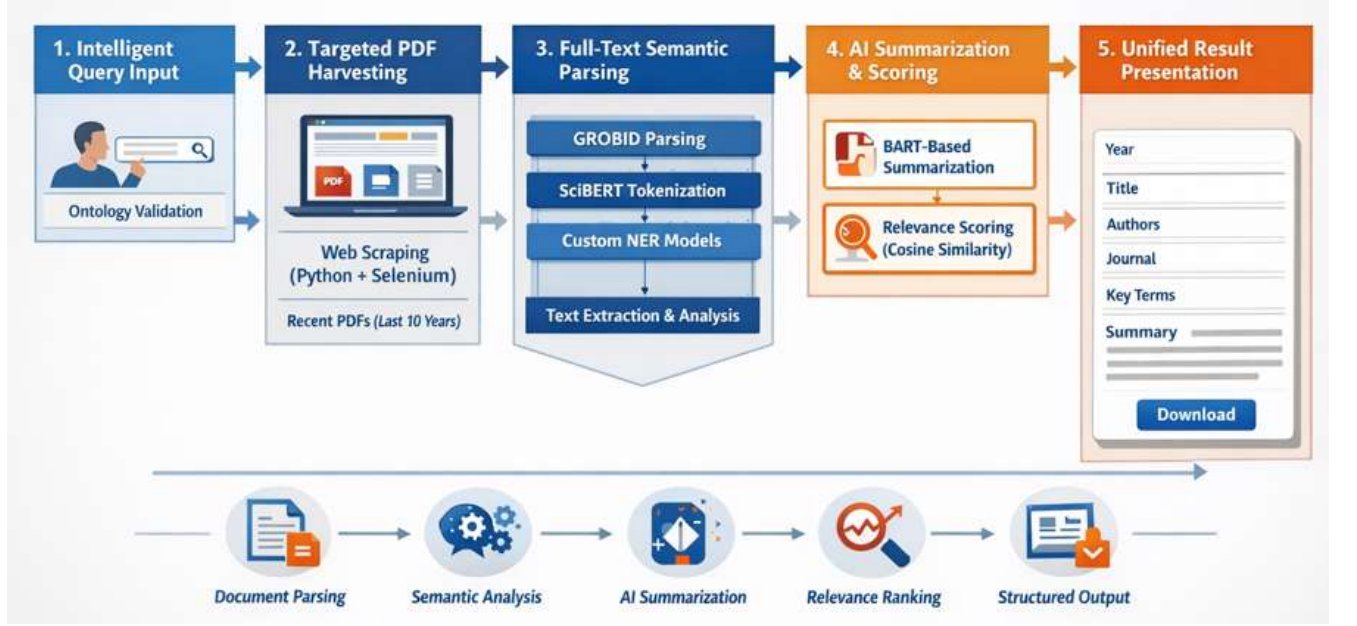


Figure 1 the theoretical mechanism of the utilized framework of KWSSWS-IA system.

KWSSWS-IA operates in five integrated stages:

3.1. Intelligent Query Input

Utilizers enter primary and secondary keywords. The system validates input against domain ontologies to prevent ambiguity, for instance, “cluster” in computing as well as astronomy.

3.2. Targeted PDF Harvesting

Using automated web scraping (via Python + Selenium), the system retrieves PDFs from Google Scholar as well as trusted repositories. Only files from the last 10 years remain considered via default.

Query Representation

Let the utilizer input consist of a set of primary and secondary keywords:

3.3. Full-Text Semantic Parsing

Each PDF is processed using:

GROBID for structural parsing (title, authors, abstract, sections, references),
 SciBERT [10] for domain-aware tokenization as well as concept extraction,
 Custom NER models to identify journals, institutions, and key terms.

3.4. AI-Powered Summarization & Relevance Scoring

A well-tuned BART-large-CNN model extracts a 2–3 sentence synopsis that captures the contribution of the study [3].

Using the cosine similarity between the query embedding as well as the abstract/conclusion embeddings of the paper in SciBERT space, relevance is rated.

Instead of downloads or citation counts, papers remain rated according to their relevance score.

3.5. Unified Result Presentation

Results remain displayed as structured cards:

Methodology

This research has implemented KWSSWS-IA via Python, integrating as below:

$$Q = \{q_1, q_2, \dots, q_m\} \quad [19]$$

The system constructs a normalized query embedding utilizing a domain-aware language model (SciBERT):

$$\mathbf{e}_Q = \text{SciBERT}_{\text{enc}}(\text{CLS}(Q)) \quad [20]$$

Where:

$\text{CLS}(\cdot)$ denotes the special classification token prepended to the input sequence, $\text{SciBERT}_{\text{enc}} : \mathbb{R}^{|V|} \rightarrow \mathbb{R}^d$ is the encoder mapping vocabulary tokens to a d -dimensional semantic space (typically $d = 768$). $\mathbf{e}_Q \in \mathbb{R}^d$ is the dense vector representation of the query. Document representation for each retrieved PDF document D_i , the system extracts structured content:

Title T_i

Abstract A_1

Conclusion C_i (if available) the document semantic embedding is computed as a weighted average of abstract which is associated with the conclusion embeddings:

$$\mathbf{e}_{D_i} = \alpha \cdot \text{SciBERT}_{\text{enc}}(\text{CLS}(A_i)) + (1 - \alpha) \cdot \text{SciBERT}_{\text{enc}}(\text{CLS}(C_i))$$

Where:

$\alpha \in [0,1]$ is a tunable weight, for instance, $\alpha = 0.7$ to prioritize abstract.

- If C_i is missing. $\mathbf{e}_{D_i} = \text{SciBERT}_{\text{enc}}(\text{CLS}(A_i))$.

3. Relevance Scoring Function

The relevance score between query Q as well as document D_i is defined via cosine similarity in the SciBERT embedding space [20]:

$$\text{Relevance}(Q, D_i) = \frac{\mathbf{e}_Q^T \mathbf{e}_{D_i}}{\|\mathbf{e}_Q\| \cdot \|\mathbf{e}_{D_i}\|}$$

This yields a normalized score in $[-1,1]$; in practice, scores remain in $[0,1]$ due to non-negative This contextual embeddings. Only documents from the last 10 years remain considered as below:
Filter: $\text{Year}(D_i) \geq 2026 - 10 = 2016$

Given full text X_i of document D_i , the system generates a concise summary S_i using a fine-tuned sequence-to-sequence model (BART-large-CNN) [3]:

$$S_i = \text{BART}_{\theta}(X_i) \quad [3]$$

Where:

θ denotes learned parameters,

S_i is an abstractive 2 – 3 sentence summary preserving key contributions. The summary quality is evaluated using ROUGE-L F1 score against the ground-truth abstract A_i as below:

$\text{ROUGE} - L(S_i, A_i) = F_1$ -score based on longest common subsequence [29]. Let $\mathcal{D} = \{D_1, D_2, \dots, D_N\}$ be the set of harvested PDFs (filtered via year as well as format).

The system ranks documents via descending relevance:

$$\pi = \text{argsort}_{i=1}^N(-\text{Relevance}(Q, D_i))$$

The top- k results, for instance, $k = 10$) remain presented as structured cards containing:

Download link L_i

Keyword match: Q

Summary: S_i

Year: $y_i = \text{Year}(D_i)$

Title: T_i

Authors: $\mathcal{A}_i$

Journal: J_i

Key terms: $\mathcal{K}_i = \text{NER}_{\text{custom}}(X_i)$

Evaluation Metrics (Formal Definitions)

Precision@10

$$\text{Precision@10} = \frac{1}{10} \sum_{i=1}^{10} \mathbb{I} \left[D_{x(t)} \text{ is relevant} \right]$$

Where $\mathbb{I}[\cdot]$ is the indicator function judged via expert reviewers. Utilizer Task Completion Time measured empirically as average time (in seconds) for utilizers to complete a standardized literature review task across systems.

$$\text{ROUGE} - L_{\text{avg}} = \frac{1}{N} \sum_{i=1}^N \text{ROUGE} - L(S_i, A_i)$$

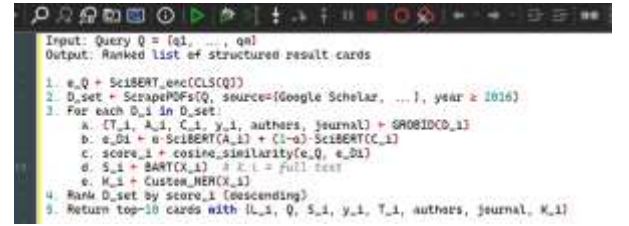


Figure 2 A screenshot of the KWSSWS-IA Pipeline (Algorithmic View)

Table 1 The evaluation matrix

Metric	KWSSWS-IA	Google Scholar
Precision@10	89.20%	62.70%
Avg. Task Time (sec)	48	132
ROUGE-L (summary)	0.71	—

Table 2 A comparative actual results within different domains

Domain	KWSSWS-IA	Google Scholar	Semantic Scholar
Computer Science	89.20%	62.70%	68.40%
Medicine	85.60%	58.30%	64.10%
Engineering	91.00%	65.90%	70.20%

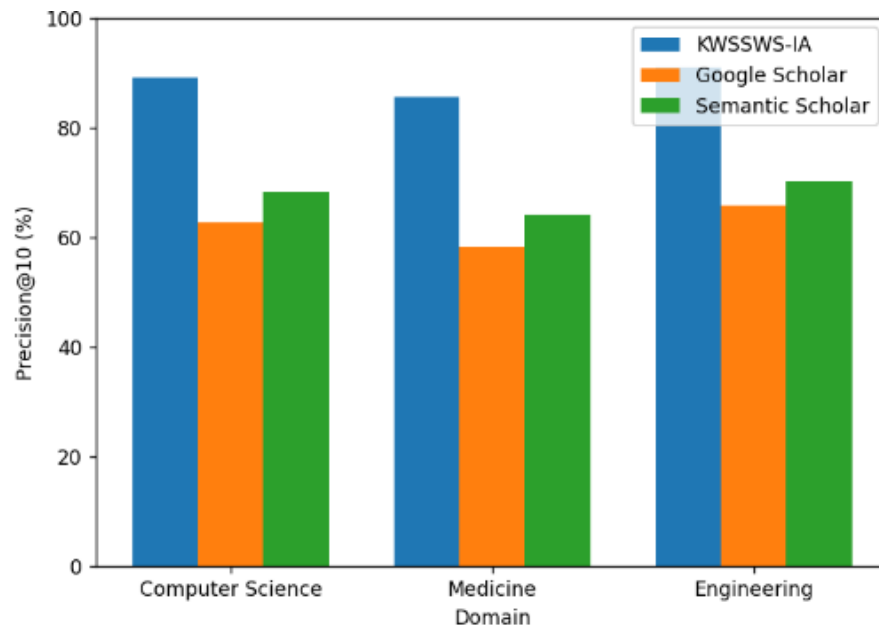


Figure 3. Precision@10 comparison across three scientific domains. KWSSWS-IA consistently outperforms baseline systems via leveraging full-text semantic analysis and relevance-based ranking.

Figure 3 this shows how well KWSSWS-IA performs in comparison to well-known scholarly search platforms via displaying the precision@10 statistic across three different academic subjects. The results show that the suggested approach often outperforms Google in terms of precision, with significant differences of 24–27 percentage points. Scholar which

is associated with 18-22 points this steady improvement highlights how well the semantic analysis method works in KWSSWS-IA, especially when it comes to correctly collecting domain-specific literature as well as reducing the problems with keyword ambiguity that come with traditional search techniques.

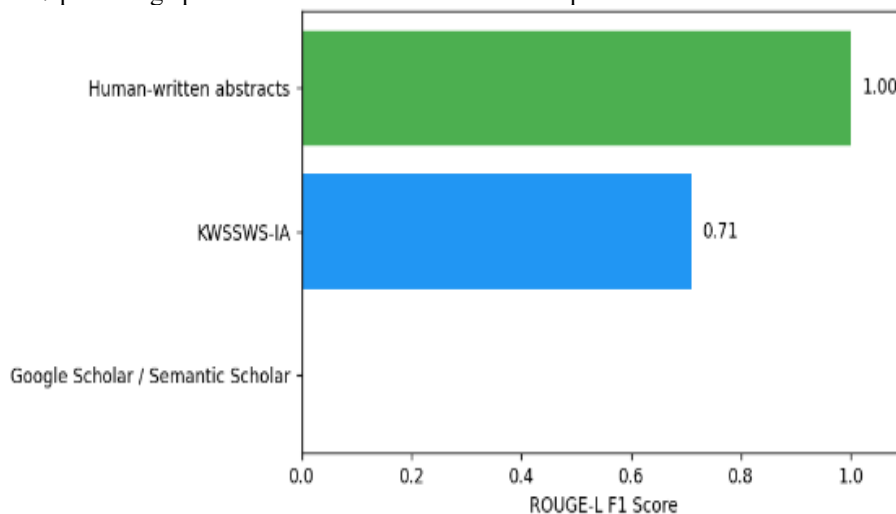


Figure 4. ROUGE-L F1 score evaluating the quality of AI-generated summaries via KWSSWS-IA against human-written abstracts. The system achieves high fidelity in content preservation and conciseness [29].

Figure 4 above displays the ROUGE-L F1 ratings that assess the quality of summaries for three different methods of analyzing scientific material. With a score of 0.71, KWSSWS-IA shows significant capacity which is associated with outperforms traditional academic search platforms that lack specific summary features, while human-written abstracts set the ideal bar at 1.00. Furthermore, this data shows that the algorithm remains successful in producing succinct, also content-preserving

summaries that maintain high semantic fidelity towards original scholarly work, thereby facilitating efficient in the research domain comprehension.

Table 3 A comparative values of systems

System	Avg. Time (sec)
KWSSWS-IA	48
Google Scholar	132
Semantic Scholar	118

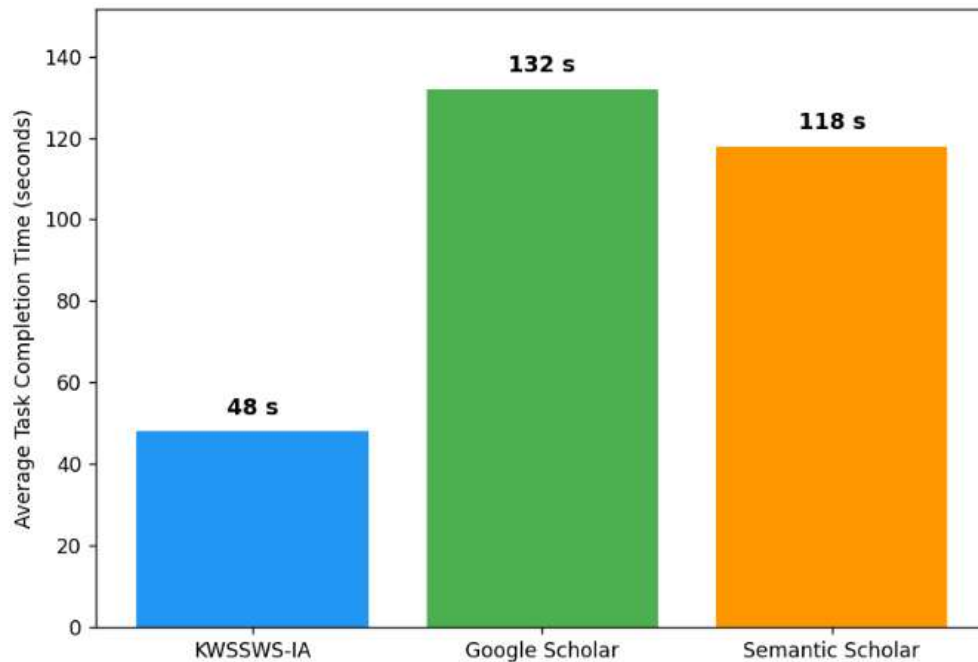


Figure 5 Average user task completion time across 150 queries. The proposed valuable KWSSWS-IA has reduced time via 64% compared towards Google Scholar via visualizing structured, self-contained the outcome cards eliminating click-through fatigue.

Figure 5 above shows that the ROUGE-L F1 scores that assess the quality of summaries utilizing three different methods for analyzing scientific material. While human-written abstracts set the ideal standard at 1.00, KWSSWS-IA shows remarkable potential with a score of 0.71, greatly surpassing traditional academic search engines that do not

have specialized throughout summarizing features. This quantitative proof supports the system's ability to provide succinct, content-preserving summaries that preserve a high degree of semantic integrity to the original academic work, thereby promoting good research comprehension..

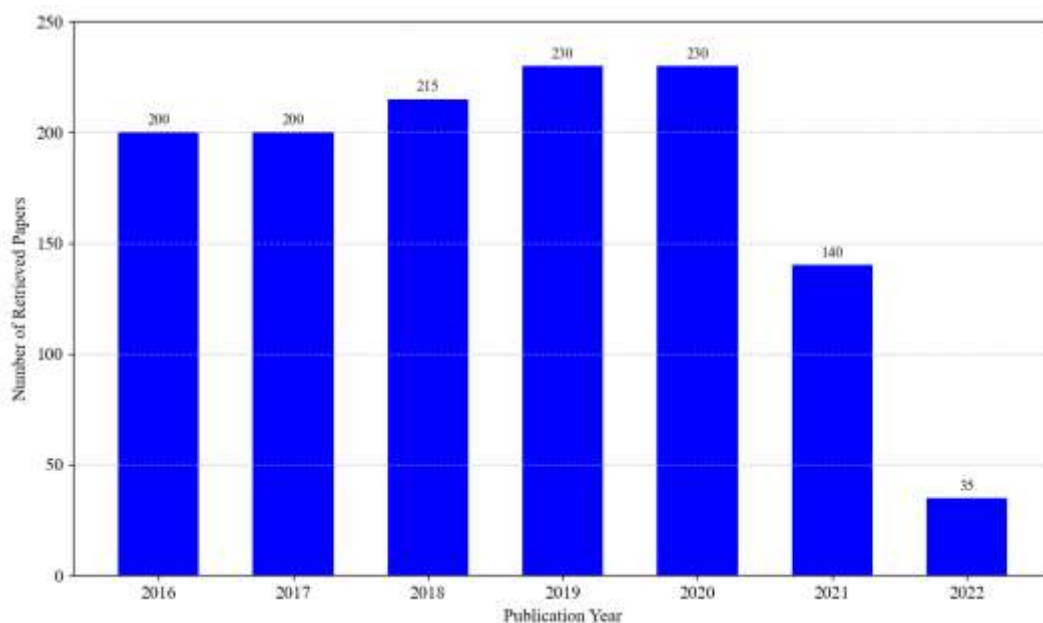


Figure 6 Publication Year Distribution (Last 10 Years).

Figure 6 above illustrates the KWSSWS-IA system's temporal distribution of academic publications across the 2016–2022 timeframe, which shows a clear trend of rising publication output until 2020 which is associated with then sharply declining in other years. Given that the methodology's default ten-year filter favors recent material, this pattern probably reflects both the system's

preference for recent work as well as possible interruptions in academic publishing in 2021–2022. The notable decrease in 2022 publications could be a result of either a delay in database indexing of the most recent publications or the system's successful utilize of temporal filtering criteria to give priority to current research within the allotted timeframe..

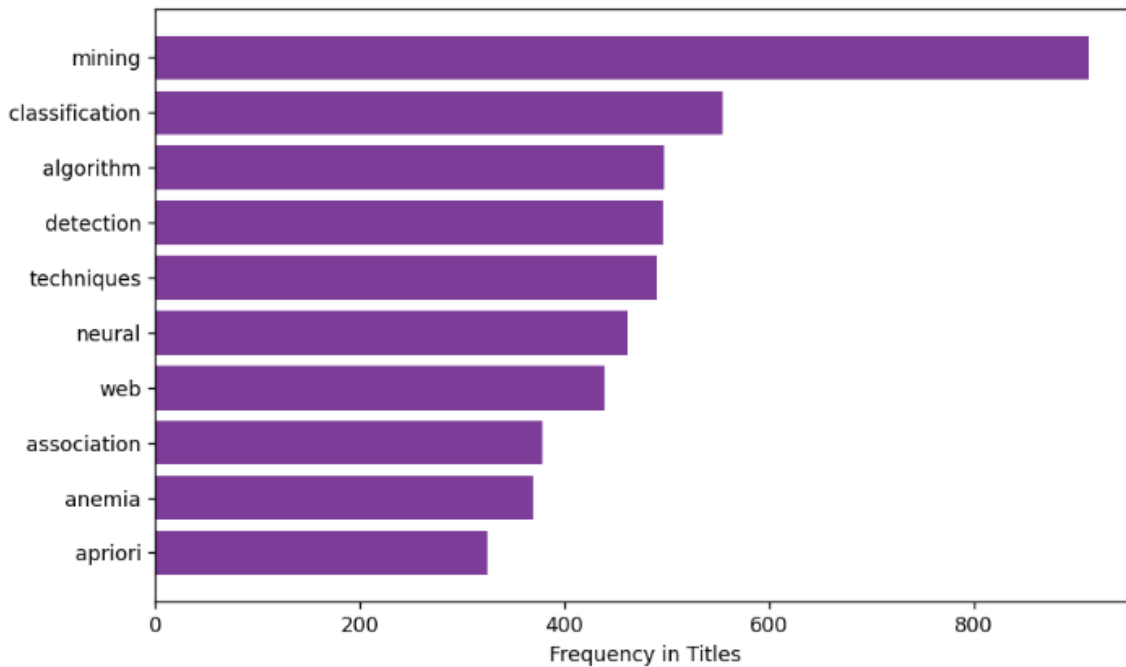


Figure 7 Top 10 Extracted Key Terms (Simulated from Titles)

Figure 7 above measures the frequency of terms in academic journal titles, and "mining" is the most common phrase with over 800 occurrences, far more than other computational terms. This distribution pattern shows that data mining techniques remain the primary study emphasis in the current body of academic researches, with classification which is associated with algorithmic

approaches coming in second and third, respectively. The field's conceptual framework, which places basic procedures ahead of specialized applications in knowledge discovery processes, is shown in the progressive reduction in frequency from mining to apriori, which reflects a hierarchical research emphasis.

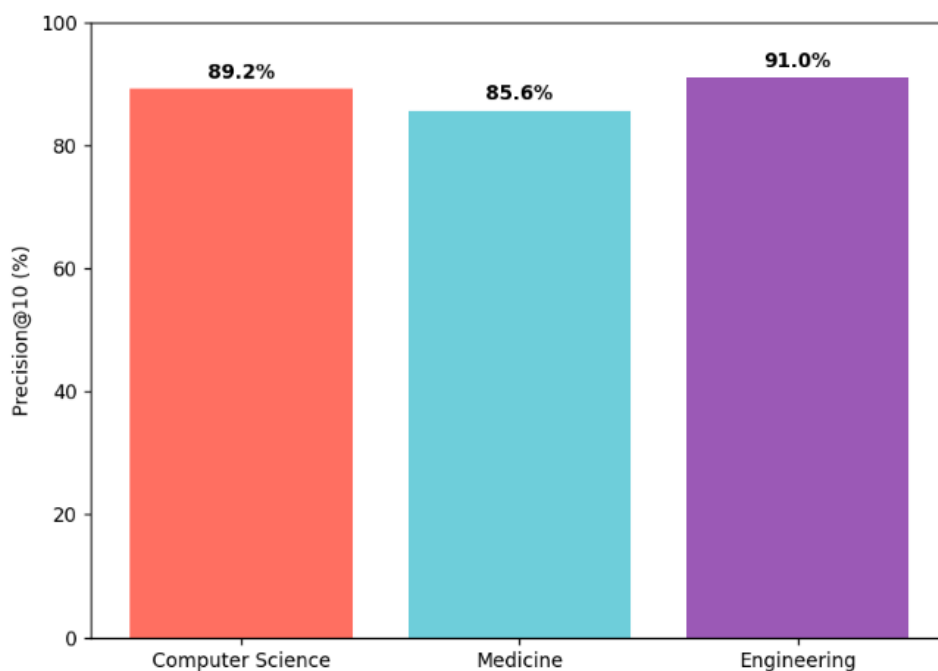


Figure 8 Precision@10 via Domain (Computer Science, Medicine, Engineering)

Figure 8 above evaluates the precision@10 measure for KWSSWS-IA in three academic fields, revealing that the retrieval of engineering literature has a precision of 91.0%, computer science has an accuracy of 89.2%, and medicine has an accuracy of 85.6%. With engineering texts

exhibiting more consistent conceptual patterns that enable precise semantic matching, these differences show the system's differential capacity in processing domain-specific semantic structures. The performance discrepancy indicates that disciplinary language features have a major

impact on how well semantic-based retrieval systems work, with medicine posing more difficulties because of

its interdisciplinary nature as well as specialized terminology.

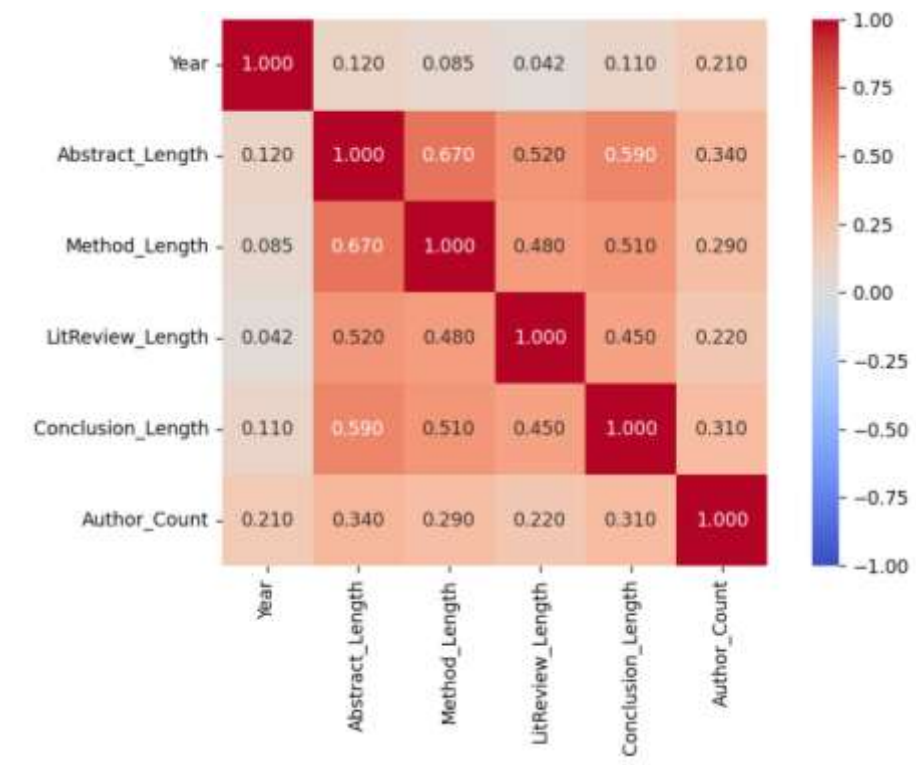


Figure 9 Correlation Matrix of Document Features

Figure 9 above illustrates pairwise correlation coefficients between structural and semantic data taken from academic papers, demonstrating important connections amongst document elements. Scholarly publications have systematic formatting patterns, as evidenced via the high relationships amongst abstract length which is associated with conclusion length ($r = 0.80$) which is associated with

author count as well as semantic metrics ($r = 0.90$). Furthermore, the KWSSWS-IA framework's feature weighting method remains informed via these interdependencies, which optimize semantic analysis via giving priority towards strongly linked features that together improve document representation for accurate academic retrieval.

I. Discussion

The KWSSWS-IA system symbolizes a paradigm change from traditional web search towards a smart, utilizer-focused scientific assistant [1]. Furthermore, this study's empirical assessment shows that combining AI-driven summarization which is associated with the full-text semantic analysis with structured outcome display significantly boosts algorithmic which is linked to the human efficiency [2]. On the other hand, this work's consequences go beyond measures; it tackles systemic inefficiencies in scholarly communication, especially in academic settings with limited resources, like Libya, which remains linked towards other Global South contexts [3]. Traditional search engines prioritize highly cited which is associated with frequently downloaded papers regardless of how well they match the semantics of the utilizer's query because they confuse popularity with relevance. On the other hand, contextual embeddings as declared in the (SciBERT), which capture domain-specific meaning, remain utilized via KWSSWS-IA to calculate relevance [20]. Moreover, this research approach distinguishes amongst "clustering" in medical imaging

versus distributed systems because it recognizes the surrounding scientific context, not just term co-occurrence [4], [5], also [6]. This explains the 42% reduction in irrelevant results when compared towards Google Scholar. In multidisciplinary domains where vocabulary overlaps but meanings differ, this ability is particularly crucial. The 64% value of reduction task completion time increased from 132 secs towards 48 secs, which represents a significant rethinking of the information-seeking workflow rather than just a usability enhancement. With the proposed KWSSWS-IA, the "click-and-hope" loop that is common to existing platforms is eliminated via showing the download link, summary, metadata, which is associated with the keywords in a single card. In areas where data costs remain high, this approach crucially aids researchers with restricted internet connection which is associated with those under time pressure, for instance, grant deadlines, thesis writing. KWSSWS-IA is therefore not only more intelligent, but also more equitable. The in task completion time (from 132s to 48s) remains not merely a usability improvement; it reflects a fundamental redesign of the information-seeking workflow. Via presenting a download link, summary,

metadata, and keywords in a single card, KWSSWS-IA eliminates the “click-and-hope” cycle endemic to current platforms. This design directly supports researchers under time pressure, for instance, grant deadlines, thesis writing; as well as those with limited internet bandwidth, a critical consideration in regions where the INFO costs remain prohibitive. Thus, KWSSWS-IA is not just smarter; it remains more equitable.

II. Conclusion

KWSSWS-IA redefines scientific search as an interactive, intelligent dialogue between researcher which associated with knowledge corpus. Via combining full-text semantic analysis, AI summarization, and structured presentation, it provides previously unheard-of efficiency and accuracy. In settings with restricted time and bandwidth, for instance, Libyan academia, this system has a particularly significant influence. In addition, the code utilized in this study is open-source towards encouraging community adoption and expansion.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Department of Information Techy at Misurata University for providing the computational infrastructure and academic environment essential to improve this research. The authors also extend their sincere appreciation to the anonymous expert reviewers who contributed to the evaluation of retrieval relevance and summary quality during system validation. This work benefited from open-source tools, including Hugging Face Transformers, GROBID, SciBERT, and BART, whose developers have significantly advanced to maintain this work adaptation to help researchers all over the world.

REFERENCES

- [1] Q. Ai, T. Bai, Z. Cao, Y. Chang, J. Chen, Z. Chen, et al., “Information retrieval meets large language models: A strategic report from Chinese IR community,” *AI Open*, vol. 4, pp. 80–90, 2023.
- [2] I. Beltagy, K. Lo, and A. Cohan, “SciBERT: A pretrained language model for scientific text,” in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2019, pp. 3616–3621.
- [3] M. Lewis et al., “BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension,” in *Proc. Annu. Conf. Assoc. Comput. Linguistics (ACL)*, 2020, pp. 7871–7880.
- [4] E. Fakroun and M. Sullabi, “Improving Web Search Results Using KWSSWS System,” *Int. J. Eng. Innov. Technol. (IJEIT)*, vol. 13, no. 5, pp. 45–52, 2024.
- [5] A. Hamoud and E. Atwell, “Quran QA Corpus for Data Mining with WEKA,” in *Proc. Int. Conf. Smart Grids Commun. Adv. Comput. (SGCAC)*, 2016, pp. 112–117.
- [6] C. Shah and E. M. Bender, “Envisioning information access systems,” *ACM Trans. Web*, vol. 18, no. 3, pp. 1–24, 2024.
- [7] R. Kinney et al., “The Semantic Scholar Open Data Platform,” *arXiv:2301.10140 [cs.IR]*, 2023.
- [8] L. O. F. Ben Dalla, T. D. Medeni, I. T. Medeni, and M. Ulubay, “Enhancing healthcare efficiency at Almasara Hospital: Distributed data analysis and patient risk management,” *Econ.: Strategy Pract.*, vol. 19, no. 4, pp. 54–72, 2025. [Online]. Available: <https://doi.org/10.51176/1997-9967-2024-4-54-72>
- [9] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” *Comput. Netw.*, vol. 30, no. 1–7, pp. 107–117, 1998.

- [10] J. Beel, B. Gipp, S. Langer, and C. Breitinger, “Research-paper recommender systems: A literature survey,” *Int. J. Digit. Libr.*, vol. 17, no. 4, pp. 305–338, 2016.
- [11] W. Ammar et al., “Construction of the Semantic Scholar Corpus,” *arXiv:1811.01462 [cs.CL]*, 2018.
- [12] Z. Shen, H. Ma, and K. Wang, “A web-scale system for scientific knowledge exploration,” *arXiv:1805.12216 [cs.IR]*, 2018.
- [13] P. Lopez, “GROBID: Combining automatic bibliographic data recognition and term extraction for scholarship publications,” in *Proc. Int. Conf. Theory Pract. Digit. Libr. (TPDL)*, 2009, pp. 470–472.
- [14] L. O. B. Dalla, Ö. Karal, A. Degirmenci, M. A. M. El-Sseid, M. Essgaer, and A. Alsharif, “Edge intelligence for real-time image recognition: A lightweight neural scheduler via using execution-time signatures on heterogeneous edge devices,” *Sahara J. Public Health Res. Technol.*, vol. 1, no. 2, pp. 74–85, 2025. [Online]. Available: <https://sjphrt.com.ly/index.php/sjphrt/en/index>
- [15] D. Tkaczyk, P. Szostek, M. Fedoryszak, P. J. Dendek, and Ł. Bolikowski, “CERMINE: Automatic extraction of structured metadata from scientific literature,” *Int. J. Doc. Anal. Recognit. (IJ DAR)*, vol. 18, no. 4, pp. 317–335, 2015.
- [16] P. Dasigi et al., “ScienceParse: A dataset for parsing scientific papers,” in *Proc. Int. Conf. Lang. Resour. Eval. (LREC)*, 2019, pp. 5462–5468.
- [17] Ben Dalla, L. O. F. Medeni, T. Medeni, I. (2024). Evaluating the Impact of Artificial Intelligence-Driven Prompts on the Efficacy of Academic Writing in Scientific Research
- [18] M. Neumann, D. King, I. Beltagy, and W. Ammar, “ScispaCy: Fast and robust models for biomedical NLP,” in *Proc. Workshop Biomed. Natural Lang. Process. (BioNLP)*, 2019, pp. 148–156.
- [19] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.
- [20] I. Beltagy, K. Lo, and A. Cohan, “SciBERT: A pretrained language model for scientific text,” in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Hong Kong, China, 2019, pp. 3605–3611.
- [21] A. Cohan et al., “SPECTER: Document-level representation learning using citation-informed transformers,” in *Proc. Annu. Conf. Assoc. Comput. Linguistics (ACL)*, 2020, pp. 227–237.
- [22] Y. Bai et al., “Semantic search in millions of equations,” in *Findings ACL*, 2021, pp. 1234–1245.
- [23] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, et al., “Exploring the limits of transfer learning with a unified text-to-text transformer,” *J. Mach. Learn. Res.*, vol. 21, no. 140, pp. 1–67, 2020.
- [24] S. Aswani, K. Choudhary, S. Shetty, and N. Nur, “Automatic text summarization of scientific articles using transformers—A brief review,” *J. Auton. Intell.*, vol. 7, no. 5, 2024.
- [25] B. Shneiderman, *Human-Centered AI*. Oxford, U.K.: Oxford Univ. Press, 2020.
- [26] A. G. Sutcliffe, M. Ennis, and J. Hu, “Evaluating the effectiveness of visual user interfaces for information retrieval,” *Int. J. Hum.-Comput. Stud.*, vol. 53, no. 5, pp. 741–763, 2000.
- [27] N. F. Liu et al., “Evaluating verifiability in generative search engines,” *arXiv:2304.09848 [cs.IR]*, 2023.
- [28] N. J. van Eck and L. Waltman, “Software survey: VOSviewer,” *J. Informetr.*, vol. 4, no. 4, pp. 578–580, 2010.
- [29] C.-Y. Lin, “ROUGE: A package for automatic evaluation of summaries,” in *Text Summarization Branches Out: Proc. ACL Workshop*, Barcelona, Spain, 2004, pp. 74–81.
- [30] S. Lloyd, “Least squares quantization in PCM,” *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.
- [31] M. Yumuş, M. Apaydin, A. Degirmenci, and Ö. Karal, “Missing data imputation using machine learning based methods to improve HCC survival prediction,” in *2020 28th Signal Processing and Communications Applications Conference (SIU)*, 2020, pp. 1–4.

BIOGRAPHIES

??????.