ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

Web Information Retrieval: A Literature Review of Search Engines, Semantic Search, Neural Information Retrieval, and Generative Artificial Intelligence

Celinne Atienza Mendez, Dr. Reagan Ricafort

AMA University Quezon City, Philippines

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150600101

Received: 26 June 2026; Accepted: 01 July 2026; Published: 11 July 2026

ABSTRACT

The growth in digital information available via the World Wide Web has made it more important than ever to develop efficient and intelligent information retrieval solutions. Despite the effectiveness of keyword-based search in matching terms accurately, traditional methods fall short in capturing intent and semantics. As a result, web information retrieval has gone through significant progress due to semantic search, machine learning, deep learning, knowledge graphs, transformers, and generative AI technologies. This literature review focuses on the developments in web information retrieval from 2016 to 2026. It discusses advancements in retrieval models, semantic search techniques, neural information retrieval, recommender systems, large language models, and generative AI-enabled retrieval systems. Major findings include a shift from keyword-based retrieval to context-based, intent-based, and knowledge-based retrieval approaches. Although there have been numerous developments in web information retrieval to ensure enhanced accuracy and user experience, the difficulties associated with scalability, misinformation, bias, interpretability, privacy, and computational efficiency continue to be considerable.

Keywords— web information retrieval, search engines, semantic search, neural information retrieval, deep learning, generative AI, information retrieval systems

INTRODUCTION

Today the Web is the world's largest archive of information. It is billions of web pages, documents, multimedia content, databases, and digital applications which produce an overwhelming amount of data that is growing day after day. This is why the challenge of retrieving relevant information has become one of the key issues for computer science and information systems today.

Information Retrieval (IR) is an operation which implies finding relevant information in large sets of structured or unstructured data on the basis of users' queries. Until recently, web search engines operated mainly through keyword matching, indexing and ranking techniques. The approaches allowed for a breakthrough in information searching but they often failed to cope with vocabulary mismatch, ambiguous queries, and context interpretation problems. Recently researchers noted that semantic search technologies have appeared to address these problems using the concept meaning and the user's intention as a base rather than keyword matching.

Between 2016 and 2026, there have been tremendous developments in artificial intelligence, machine learning, transformers, vector databases, and generative AI which radically changed web information retrieval systems.

This literature review examines contemporary developments in web information retrieval, focusing on semantic search, neural retrieval systems, deep learning models, knowledge-based retrieval, and AI-powered search technologies.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

BACKGROUND LITERATURE

The concept of information retrieval as a research area started with the development of methods of storing, indexing, searching, and retrieving information from large collections of documents. The early information retrieval methods utilized Boolean retrieval models, vector space model, and probabilistic ranking techniques.
Figure 1. Basic Components of a Web Information Retrieval System

The traditional search engines of the World Wide Web used methods based on keyword matching and ranking techniques for determining the relevancy of documents. Though the approach was quite useful in many situations, it had its own limitations when it came to problems of synonymy, polysemy, and vocabulary mismatch. Semantic search was conceived as a new concept that would attempt to interpret the meaning of both the queries and documents. The studies on semantic search are very much concerned with context analysis, ontology reasoning, and knowledge representation.

The history of WebIR is directly associated with changes in search engines. The first search engines used only keyword matching and indexation algorithms based on the information retrieval theory. Nevertheless, as the size of the web grew, search engines began to use ranking algorithms in order to make the search more precise. Modern search engines use various methods such as content, hyperlink analysis, user behavior data, and machine learning techniques in order to give more accurate results. As a result, search engines have become an important tool for searching and knowledge acquisition.

Another innovation that has come with WebIR is the link-based ranking algorithms. The earlier retrieval models did not take into account any other factor apart from the relevance factor of textual documents, while the hyperlinks in the web pages could act as an indication of the authority and relevance of the web pages. Using algorithms such as PageRank and HITS, it has been proven that web pages that receive links from other authoritative web pages are themselves more authoritative and relevant.

User-oriented approaches have also been considered in the area of WebIR. Apart from being influenced by the algorithms, the efficiency of retrieval is also dependent on the way users construct their queries and interact with the search engines. Some of the functions such as query expansion, advanced searching, personalization and relevance feedback make it possible for the users to customize their search and thus retrieve more relevant results. Thus, it is clear that user behavior has become an important aspect in the design of web search engines.
The development of WebIR has also been impacted by the progress in machine learning and artificial intelligence. The current research emphasizes the development of neural ranking models that use deep learning for comprehending semantic similarities between queries and documents. In contrast to conventional keyword-based systems, the new approach is able to comprehend context and resolve the problem of vocabulary mismatch between queries and websites.

Another crucial trend in WebIR is the use of temporal information within the ranking processes. The dynamic nature of web information makes the timeliness of the information very important. The use of temporal information retrieval models takes into consideration the date of publication of information, its freshness, and changing user interest to provide more relevant information. This trend is applicable in news retrieval, social media searching, and event information search.

Moreover, the issue of information quality and credibility on the web is another area that has been addressed by the research community. As a result of the huge quantity of user-generated and unverified information available on the web, there is a growing trend of including trust and quality evaluation components in retrieval systems.
Recently, due to the advancements in the domain of deep learning and natural language processing, neural retrieval models have been introduced in information retrieval.

METHODOLOGY

This study uses the method of structured narrative literature review, where PRISMA approach was included to enhance the quality of the research process. Even though this paper does not represent the systematic literature

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

review, the presence of the clear selection criteria will reduce possible biases, and the research process will become reproducible.

Literature Review was conducted from 2016 to 2026 and focused on peer-reviewed journals, conference papers, survey papers, and reliable academic sources in the field of web information retrieval. The literature sources were taken from reliable digital libraries such as IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, Google Scholar, and arXiv.

The review process consists of four steps:

1.Selection of suitable literature sources through the use of keywords such as web information retrieval, semantic search, neural information retrieval, dense retrieval, transformer model, vector database, retrieval augmented generation, and generative AI.

2. Screening process where the duplicates from the gathered literature were deleted, and those not directly relating to web information retrieval technology were excluded.

3. Assessment of the eligibility of the selected studies based on their publication quality, relevance to the research objectives, publication date, and adequacy of technical analysis.

4. Inclusion and synthesis where the selected sources were categorized into several categories, such as traditional retrieval models, semantic search, neural information retrieval, transformer retrieval, cloud retrieval, and generative AI.

Unlike summarization of existing literature sources, this study also analyzes the developments and evolution of the techniques, along with identifying common patterns, research gaps, and technical challenges associated with different generations of information retrieval systems.

LITERATURE REVIEW

Traditional Information Retrieval Models

The conventional information retrieval systems depended heavily on keyword-based systems. Some popular methods for information retrieval were Boolean retrieval, vector space method, and probability ranking principles.

The retrieval systems used inverted indices to retrieve the documents that contained certain terms. Although they were highly scalable and efficient, such keyword-based systems did not have the capability of understanding the semantic relationship among the concepts. The researchers found vocabulary mismatch as one of the biggest problems that affected the retrieval process. The documents might contain different vocabularies than those used in the query.

Semantic Search and Knowledge-Based Retrieval

Semantic search is one of the greatest innovations in the domain of web information retrieval. As opposed to conventional search techniques, semantic search focuses on the meaning behind a user query.

There is evidence to suggest that semantic search engines make use of ontology, knowledge graphs, linked data, entity recognition, and contextual reasoning to increase their effectiveness in retrieving information. Such engines identify relationships between concepts and return relevant responses based on semantics and not just lexical proximity.

According to recent findings, semantic search has been found to greatly increase the efficiency of retrieval in complicated search scenarios in which users can formulate concepts through different vocabularies. Concepts of knowledge representation also contribute to efficient semantic search by allowing machines to understand conceptual relationships in large volumes of data.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

Neural Information Retrieval

Deep Learning gave rise to a new approach called Neural Information Retrieval (Neural IR). Neural IR makes use of artificial neural networks for acquiring representations of queries and documents from the raw data itself.

A survey literature on Neural IR pinpointed the evolution of deep neural architectures as a breakthrough in information retrieval research. Neural models are capable of automatically learning semantic representations, thereby eliminating the need for feature engineering.

Further research showed that neural information retrieval models are more effective than conventional models in many retrieval tasks, especially in handling semantic similarity and context. Neural information retrieval models are broadly divided into sparse retrieval models, dense retrieval models, and hybrid retrieval models.

Dense retrieval models make use of vector representations for documents and queries.

Transformer Models and Semantic Search

Transformers have now emerged as the prevailing architecture in information retrieval systems. The BERT, RoBERTa, T5, and GPT models make use of self-attention techniques to learn the contextual relationship of text.

According to recent surveys, the precision and recall of semantic search based on transformer models is higher than previous neural methods. The contextual embeddings generated by these models are able to detect the semantic relationship between queries and documents.

DPR and vector search systems have become increasingly important in retrieval systems. Vector databases allow for the efficient storing and retrieving of semantic embeddings, thus enabling semantic search.

The transformer retrieval systems have been used in web search, enterprise search, question answering, recommendation systems, and scientific information retrieval.

Big Data and Cloud-Based Information Retrieval

With the advent of big data, many problems have arisen for the information retrieval system. The contemporary information retrieval environment needs to deal with large scale heterogeneous dynamic, and continuously evolving databases.

The cloud-based retrieval architecture has been found to be quite efficient in handling large-scale information retrieval environments. They make use of distributed computation power, elastic storage infrastructure, and scalable indexing facilities.

There is evidence to suggest that cloud-based retrieval environments offer higher levels of scalability and fault tolerance. Nevertheless, issues related to privacy, security, and latency cannot be overlooked.

Generative Artificial Intelligence and Retrieval-Augmented Generation

One of the key trends occurring between 2023 and 2026 is the implementation of generative AI in information retrieval systems.

The large language models, like the GPT models, completely changed the nature of communication between users and information retrieval systems by switching the search result generation from the ranking of the document list to the synthesis of information into a response.

According to the latest studies in the area, there is an increasing tendency to combine retrieval and generative models using the approach of Retrieval-Augmented Generation (RAG). In this case, the retrieval component finds sources for the search, and the generator generates the response in natural language using retrieved data.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

Modern discourse in this area also emphasizes the transition from keyword-based search to concept-based and semantic understanding approaches.

Comparative Analysis of Information Retrieval Approaches

Web search has evolved over time from a process that focused on lexical matching to one that incorporates context and knowledge-based retrieval. There are various strengths and weaknesses associated with each of the methods.

Keyword-based retrieval is efficient from a computational point of view. This method works well when dealing with large numbers of documents that have been indexed. The method's major limitation lies in vocabulary mismatch.

On the other hand, semantic retrieval uses ontologies, knowledge graphs, and contextual reasoning to help comprehend the meaning of user queries. Semantic retrieval yields more relevant search results as compared to keyword-based retrieval, especially when dealing with complicated searches. However, building and managing semantic resources requires hard work and extensive knowledge in that domain.

Neural information retrieval enhances retrieval performance by generating dense vector representations of queries and documents using deep learning techniques. Methods that employ dense vectors perform exceptionally well at identifying semantic similarities between queries and documents beyond lexical matching. The main problem with dense retrieval is that it is highly computationally expensive.

More recently, however, generative AI algorithms, especially those known as Retrieval-Augmented Generation (RAG), have revolutionized the concept of information retrieval through the integration of document retrieval and natural language generation processes. Instead of showing the users a list of ranked documents, the systems integrate the retrieved information into meaningful answers. This technique offers better performance for users but is associated with several issues such as hallucinations, explainability, factual accuracy, and high computational power.

In general, today's research shows that the most optimal information retrieval approaches are hybrid architectures integrating keyword-based search, semantic representation, neural information retrieval, and generative models.

Challenges in Modern Web Information Retrieval

Despite substantial advancements, several challenges remain.

Scalability

The expanding amount of content on the web necessitates a need for retrieval systems that can index and search large amounts of data efficiently.

Explainability

Deep learning and transformer-based retrieval systems

operate as black boxes, meaning there is an absence of clarity regarding the reasons certain results are generated.

Bias and Fairness

There are chances that machine learning models might be biased due to the training data used, which affects the output of the retrieval process.

Misinformation

Generative AI models could generate false information in cases where retrieval systems are not efficient enough.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

Privacy and Security

Personalized retrieval systems necessitate large amounts of data about users, posing challenges related to data privacy and security.

DISCUSSION

Leadership and Governance

Web information retrieval technology literature clearly shows the evolution of retrieval technologies over the last decade. Keyword-based search has established the basics of search but the drawbacks in keyword-based search have led to the creation of semantic and neural retrieval technologies.

The development of semantic technologies helped in improving the quality of search because the search systems were able to comprehend the concept behind user queries. In addition, neural information retrieval helped to improve the technology even more through learned semantics that could solve the vocabulary mismatch problem.

The latest technologies in search systems are Transformer-based models that make use of the context behind the search queries to provide an effective search solution. Vector databases and dense retrieval technologies have made retrieval technologies more scalable and meaningful.

The latest development in web information retrieval technology is generative AI technology, which has made retrieval technology help users to synthesize and explain the knowledge retrieved.

CONCLUSION

There have been notable changes in web information retrieval from 2016 to 2026. Innovations in semantic search, neural retrieval, transformers, vector database technology, and generative artificial intelligence have brought fundamental change to information discovery, ranking, and presentation.

The literature suggests that in the future, retrieval systems will be increasingly focused on semantic analysis, contextual inference, personalization, and AI-driven knowledge creation. Although innovations help make information retrieval more effective and user-friendly, the issues of scalability, interpretability, fairness, misinformation, and privacy should not be ignored.

In the future, it is important to work on trustworthy AI retrieval systems, interpretability of the search, multimodal retrieval, and robust retrieval-enhanced generation systems.

While this literature review covers the major advances in web information retrieval made in the last decade, there are still some directions for future research. It would be beneficial to conduct a more thorough systematic review following the entire PRISMA approach to have a higher level of methodological rigor and reproducibility. Future studies would also gain from the quantitative comparison of retrieval performance conducted with the help of established benchmarks like MS MARCO, BEIR, or TREC corpora. The comparison of keyword-based, semantic, neural, and generative retrieval would give more insights into their relative advantages and disadvantages and show more practical aspects of them.

Moreover, future reviews would be able to make use of the real-life examples in domains such as healthcare, digital libraries, enterprise search, and e-commerce to have a stronger empirical basis for analysis.

REFERENCES

1. Chowdhury, T. (2026). *Semantic Search With Vector Database: A Comprehensive Review of Models, Indexing and Applications*. The Eastasouth Journal of Information System and Computer Science, 3(3), 1–15.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VI, June 2026

2. Chowdhury, T. (2026). *Cloud-Based Information Retrieval for Big Data: A Survey of Architectures and Scalability Challenge*. The Eastasouth Journal of Information System and Computer Science, 3(3), 1–14.
3. Ezhilarasi, K., & Kalavathy, G. M. (2018). *Literature Survey: Analysis on Semantic Web Information Retrieval Methodologies*. Proceedings of PECTEAM 2018.
4. Guo, J., Cai, Y., Fan, Y., Sun, F., Zhang, R., & Cheng, X. (2022). *Semantic Models for the First-Stage Retrieval: A Comprehensive Review*. ACM Transactions on Information Systems, 40(4), 1–42.
5. Kamil, M., & Çakir, D. (2025). *Advances in Transformer-Based Semantic Search: Techniques, Benchmarks, and Future Directions*. Turkish Journal of Mathematics and Computer Science, 17(1), 1–25.
6. Martinez-Rodriguez, J. L., Hogan, A., & Lopez-Arevalo, I. (2020). *Information Extraction Meets the Semantic Web: A Survey*. Semantic Web Journal, 11(2), 255–335.
7. Patil, S. D., & Aalam, Z. (2026). *A Critical Review of Information Retrieval Techniques: Current Trends and Challenges*. International Journal of Informatics and Communication Technology, 15(2), 456–464.
8. Rashid, J., & Nisar, M. W. (2016). *A Study on Semantic Searching, Semantic Search Engines and Technologies Used for Semantic Search Engines*. International Journal of Information Technology and Computer Science, 8(10), 82–89.
9. Zhang, Y., Rahman, M. M., Braylan, A., Dang, B., Chang, H. L., Kim, H., McNamara, Q., Angert, A., Banner, E., Khetan, V., McDonnell, T., Nguyen, A. T., Xu, D., Wallace, B. C., & Lease, M. (2016). *Neural Information Retrieval: A Literature Review*. arXiv Preprint arXiv:1611.06792.
10. Zhang, Y., Altingovde, I. S., Karagoz, P., Rahman, M. M., Braylan, A., Dang, B., Chang, H. L., Kim, H., McNamara, Q., Angert, A., Banner, E., Khetan, V., McDonnell, T., Nguyen, A. T., Xu, D., Wallace, B. C., & Lease, M. (2018). *Neural Information Retrieval: at the End of the Early Years*. Information Retrieval Journal, 21(2–3), 111–182.
11. Asadi, S., & Jamali, H. R. (2004). *Shifts in Search Engine Development: A Review of Past, Present and Future Trends in Research on Search Engines*. Webology, 1(2).
12. Baeza-Yates, R. (2003). *Information Retrieval in the Web: Beyond Current Search Engines*. International Journal of Approximate Reasoning, 34(2–3), 97–104. https://doi.org/10.1016/j.ijar.2003.07.002
13. Choudhary, L., & Burdak, B. S. (2012). *Role of Ranking Algorithms for Information Retrieval*. arXiv. https://arxiv.org/abs/1208.1926
14. Guo, J., Fan, Y., Pang, L., Yang, L., Ai, Q., Zamani, H., Wu, C., Croft, W. B., & Cheng, X. (2019). *A Deep Look Into Neural Ranking Models for Information Retrieval*. arXiv. https://arxiv.org/abs/1903.06902
15. Hiwale, K., More, P., & Nayake, Y. (2024). *A Evolution and Impact of Web Search Engines: A Comprehensive Review*. Engineering and Technology Journal, 9(6).
16. Langville, A. N., & Meyer, C. D. (2005). *A Survey of Eigenvector Methods for Web Information Retrieval*. SIAM Review, 47(1), 135–161. https://doi.org/10.1137/S0036144503424786
17. Lewandowski, D. (2005). *Web Searching, Search Engines and Information Retrieval*. Information Services & Use, 25(3–4), 137–147. https://doi.org/10.3233/ISU-2005-253-402
18. Mitra, B., & Craswell, N. (2017). *Neural Models for Information Retrieval*. arXiv. https://arxiv.org/abs/1705.01509
19. Moulahi, B., Tamine, L., & Ben Yahia, S. (2016). *When Time Meets Information Retrieval: Past Proposals, Current Plans and Future Trends*. Journal of Information Science, 42(6), 794–818. https://doi.org/10.1177/0165551515607277
20. Pokorný, J. (2004). *Web Searching and Information Retrieval*. Computing in Science & Engineering, 6(4), 43–48. https://doi.org/10.1109/MCSE.2004.24
21. Sanderson, M., & Croft, W. B. (2012). *The History of Information Retrieval Research*. Proceedings of the IEEE, 100(Special Centennial Issue), 1444–1451. https://doi.org/10.1109/JPROC.2012.2189916
22. Srinivasa, S., & Bhatt, P. C. P. (2002). *Introduction to Web Information Retrieval: A User Perspective*. Resonance, 7(6), 29–41. https://doi.org/10.1007/BF02834389
23. Ramachandran, S., Paulraj, S., Joseph,S., & Ramaraj, V. (2009). *Enhanced Trustworthy and High-Quality Information Retrieval System for Web Search Engines*. arXiv. https://arxiv.org/abs/0911.0914