

# Mapping Research Trends of Query Expansion in Information Retrieval: A Bibliometric Analysis

Roberto Kaban <sup>1)</sup>

<sup>1)</sup> Teknik Informatika, Institut Teknologi dan Bisnis Indonesia, Deli Serdang, Indonesia  
roberto.kaban@yahoo.com

Submitted : 28 February 2026 | Accepted : 26 March 2026 | Published : 31 March 2026

**Abstract:** This study aims to analyze the development of research on query expansion in the field of information retrieval using a bibliometric approach to understand research trends, distribution, and current research focus. The data were obtained from 676 publications indexed in Scopus during the period from 2020 to February 2026. The research method involves quantitative analysis of annual publication trends, distribution of subject areas, document types, and keyword analysis using VOSviewer to map keyword relationships through co-occurrence analysis, overlay visualization to identify keyword trends, and density visualization to observe the concentration of research topics. The results show fluctuations in the number of publications with a peak occurring in 2025 with 141 publications. The research is dominated by the Computer Science field with 596 publications, and the majority of documents are conference papers with 369 publications. Keyword analysis identifies core topics such as information retrieval with 483 occurrences, query expansion with 354 occurrences, and search engines with 221 occurrences. Recent research trends include large language models, word embedding, and retrieval-augmented generation. The keyword network visualization indicates a shift from traditional methods such as relevance feedback toward modern approaches based on artificial intelligence and machine learning, which are increasingly relevant for improving the effectiveness of information retrieval systems. These findings provide both quantitative and qualitative insights into the evolution of query expansion research. The results also highlight the integration of modern technologies in retrieval practices and provide a foundation for new researchers to identify trends, research gaps, and opportunities for future innovation.

**Keywords:** Query Expansion; Information Retrieval; Bibliometric Analysis; Keyword Co-occurrence; VOSviewer

## INTRODUCTION

The rapid development of information technology has led to an exponential growth of digital data across many areas of life. Information is now available in massive quantities through various platforms such as scientific databases, web search engines, social media, and electronic information systems (Vladlenov, 2023). This situation creates a major challenge in the field of Information Retrieval. The challenge concerns how to find relevant information quickly and accurately from extremely large collections of data (Meliukh et al., 2025; Venkatachalam & Venkatachalam, 2023). Information retrieval systems are designed to help users locate relevant documents based on an information need that is expressed through a query (Bernard & Balog, 2025; Hambarde & Proença, 2023). However, in practice the search process often encounters several limitations that cause retrieval results to be less than optimal.

One of the major problems in search systems is the mismatch between the keywords used by users and the terms that appear in documents (Raj et al., 2025). This phenomenon is known as the vocabulary mismatch problem. Users often employ terms that differ from those used in relevant documents. As a result the system fails to identify documents that actually match the user's information needs (Pan et al., 2025). This issue has become an important concern in research within the field of Information Retrieval because it can reduce both precision and recall in information search systems.

\*Roberto Kaban



To address this problem, various approaches have been developed. One widely used approach is Query Expansion (Baumann & Schoenfeld, 2024). This technique aims to expand the user's initial query by adding other terms that have semantic or contextual relationships with the original keywords (Allahim et al., 2025). By expanding the query, the system can retrieve more relevant documents that were not previously identified by the original query. This approach can be implemented through several methods such as the use of a thesaurus, statistical word analysis, relevance feedback, and machine learning techniques (Kaban et al., 2025b).

In recent years, research on Query Expansion has grown significantly alongside technological advances such as Natural Language Processing, Machine Learning, and Large Language Models. Numerous studies have explored how query expansion techniques can improve the performance of search systems in different domains. These domains include web search, recommendation systems, scientific document retrieval, and social media analysis (Vishwakarma & Kumar, 2025). In addition, the integration of modern methods such as semantic analysis and vector-based word representation has introduced new approaches to improve the effectiveness of the retrieval process.

Although research on Query Expansion has developed rapidly, studies that systematically map the research trends in this field are still relatively limited (Al-Lahham, 2024; Meliukh et al., 2025). Most existing studies focus on developing specific methods or algorithms. Comprehensive analyses that examine publication patterns, dominant research topics, and relationships among concepts in the scientific literature remain limited. Therefore, a literature analysis approach is required to provide a comprehensive overview of research development in this field.

One method that can be used to analyze the development of research is bibliometric analysis. This approach allows researchers to identify patterns of scientific publication, collaboration among researchers, and relationships among research topics based on bibliographic data from scientific databases (Ganti et al., 2025; Kumar, 2025). By using visualization software such as VOSviewer, researchers can map relationships among keywords, author collaboration networks, and the development trends of research topics within a particular field. This analysis provides a visual representation that helps researchers understand the structure of knowledge and the future direction of research.

Based on this background, this study aims to analyze and map research trends related to Query Expansion in the field of Information Retrieval using a bibliometric approach. The analysis is conducted using scientific publication data obtained from the Scopus database. This research seeks to identify the development of publication trends, the distribution of subject areas, document types, and the relationships among keywords that appear in the scientific literature.

Specifically, this study attempts to answer several research questions, namely: (1) how the publication trend of research on Query Expansion in the field of Information Retrieval has evolved over time, (2) which subject areas contribute the most to research related to Query Expansion, (3) what patterns of relationships exist among research keywords in the scientific literature, and (4) what research topics represent the main focus in the current development of Query Expansion studies. The findings of this study are expected to provide a comprehensive overview of the development of research in query expansion and to serve as a reference for researchers in identifying research trends, potential research gaps, and future research opportunities.

## LITERATURE REVIEW

### Information Retrieval

Information Retrieval is a research field within computer science that focuses on the process of searching for and retrieving relevant information from large collections of documents (Roberts, 2024). Information retrieval systems are designed to assist users in obtaining information that matches their needs, which are typically expressed through a query or search keywords (Pan et al., 2024; Peikos & Pasi, 2024). The rapid development of digital technology has significantly increased the amount of information available across various data repositories. As a result, information retrieval systems have become an essential component in managing digital knowledge (Patel et al., 2024).

In an information retrieval system, users usually enter a query that represents their information needs. The system then performs a matching process between the query and documents stored in a database (Naamha & Abdulmunim, 2024; Pan et al., 2024). This process is conducted using several retrieval models such as the Boolean Model, the Probabilistic Retrieval Model, and the Vector Space Model (Hidri, 2024). The primary objective of these systems is to generate a ranked list of documents that are most relevant to the user's information needs.

One of the main challenges in information retrieval systems is the vocabulary mismatch problem. This problem occurs when the words used by users in their queries differ from the terms used in relevant documents (Zahhar et al., 2025). As a consequence, the system may fail to retrieve documents that actually match the user's

information needs. Therefore, various methods have been developed to improve search quality within information retrieval systems.

### Query Expansion in Information Retrieval

One widely used method to improve the performance of search systems is Query Expansion. This technique aims to enrich the initial query by adding additional words or terms that have semantic relationships with the main keywords (Baumann & Schoenfeld, 2024; Vishwakarma & Kumar, 2025). By adding these related terms, the system can retrieve documents that were not detected by the original query, thereby improving the quality of search results.

Query expansion can be implemented through several approaches. The first approach uses external knowledge sources such as dictionaries, thesauri, or ontologies to identify synonyms of the words used in the query (Kaban et al., 2025a). The second approach relies on statistical analysis of document collections to identify words that frequently co-occur with the main keywords. This method often utilizes term weighting techniques such as TF-IDF to determine relevant terms.

Several studies have also developed machine learning-based query expansion methods. These approaches utilize classification algorithms or language models to generate additional terms that are more contextually relevant (Vishwakarma & Kumar, 2025). Research has shown that query expansion techniques can increase the number of relevant documents retrieved by search systems and improve both precision and recall in the retrieval process.

In addition, advances in artificial intelligence have encouraged the development of query expansion methods based on Large Language Models (Breuer et al., 2025; Pan et al., 2025; Z. Wang & Pei, 2024). This approach enables systems to generate more contextual queries by understanding deeper semantic relationships between words. Such methods have demonstrated improved effectiveness in information retrieval systems across various datasets and search scenarios.

### Bibliometric Analysis in Research Mapping

To understand the development of research within a particular scientific field, many researchers employ a bibliometric analysis approach. Bibliometric analysis is a method used to analyze scientific publications based on metadata such as authors, keywords, publication years, and relationships among studies (Ahmed, 2024; Kumar, 2025). This analysis enables researchers to identify research trends, collaborations among authors, and the most frequently studied research topics. It can also be used to map the structure of knowledge within a research field and identify future research opportunities (Yıldız & Karakuş, 2024).

In bibliometric studies, network visualization is often performed using software such as VOSviewer. This software can visualize relationships among keywords, authors, and documents in the form of research networks. Such visualization helps researchers understand patterns in the development of a particular research field. Previous studies have shown that bibliometric analysis using VOSviewer can produce several types of visualizations, including network visualization, overlay visualization, and density visualization to comprehensively map research trends.

### Previous Studies

Several recent studies have examined the application of Query Expansion to improve the effectiveness of Information Retrieval systems. Research by Liang Wang, Nan Yang, and Furu Wei introduced the Query2doc approach, which utilizes Large Language Models to generate pseudo-documents that expand the original query. Experimental results show that this method can improve retrieval performance by approximately 3% to 15% on benchmark datasets such as MS MARCO and TREC Deep Learning without requiring additional model training (L. Wang et al., 2023).

Another study by Fanghua Ye and colleagues explored the use of large language models to perform informative query rewriting in conversational search environments. Their findings demonstrate that LLM-based query reformulation can significantly improve retrieval performance compared to traditional query reformulation methods, particularly when applied to conversational information retrieval tasks (Ye et al., 2023).

In addition, recent benchmarking research on query expansion techniques for social media information retrieval shows that combining semantic approaches and generative AI models can improve the alignment between query terms and document semantics. The results indicate that semantic expansion techniques provide more relevant search results when compared to traditional keyword-based retrieval approaches (Stathopoulos et al., 2023).

More recent studies have also proposed advanced frameworks for query expansion based on large language models. For example, the MuGI framework generates multiple pseudo-references using language models to enrich query representations and improve retrieval accuracy across both sparse and dense retrieval models. Experimental

evaluations show that integrating contextual information generated by language models can significantly improve retrieval performance (Zhang et al., 2024).

Furthermore, recent survey research highlights that query expansion techniques continue to evolve through the integration of semantic analysis, term weighting strategies, and machine learning methods. These developments demonstrate that incorporating contextual and semantic information can substantially improve the precision and recall of information retrieval systems.

**Research Gap**

Although many recent studies have developed various query expansion techniques to improve the performance of Information Retrieval systems, most of these studies focus primarily on proposing new algorithms or improving retrieval performance using specific datasets. However, research that systematically analyzes the overall development of query expansion research trends remains limited. In particular, comprehensive studies that investigate the global publication trends, subject area distribution, and relationships among research keywords in query expansion research are still relatively scarce.

In addition, most existing studies emphasize algorithm development and system performance evaluation, while relatively few studies attempt to map the structure of knowledge and research evolution in this field through bibliometric analysis. Therefore, this study aims to map research trends related to Query Expansion in the field of Information Retrieval using a bibliometric analysis approach. Publication data are obtained from the Scopus database and analyzed using VOSviewer to identify research patterns, relationships among research topics, and the overall development of research in this field.

**METHOD**

This study employs a bibliometric analysis approach to map the development of research on Query Expansion in the field of Information Retrieval. Bibliometric analysis is a quantitative method used to examine scientific publications through statistical measurements of various publication metadata such as authors, keywords, research fields, and relationships among documents. This approach enables researchers to understand patterns of research development, the structure of knowledge, and emerging research trends within a particular scientific domain.

The scientific publication data were obtained from the Scopus database because it is one of the largest sources of scholarly literature that includes numerous reputable international journals. The literature search process was conducted using keywords related to the research topic, namely “query expansion” and “information retrieval.” These keywords were entered into the search field to identify articles relevant to the research topic.

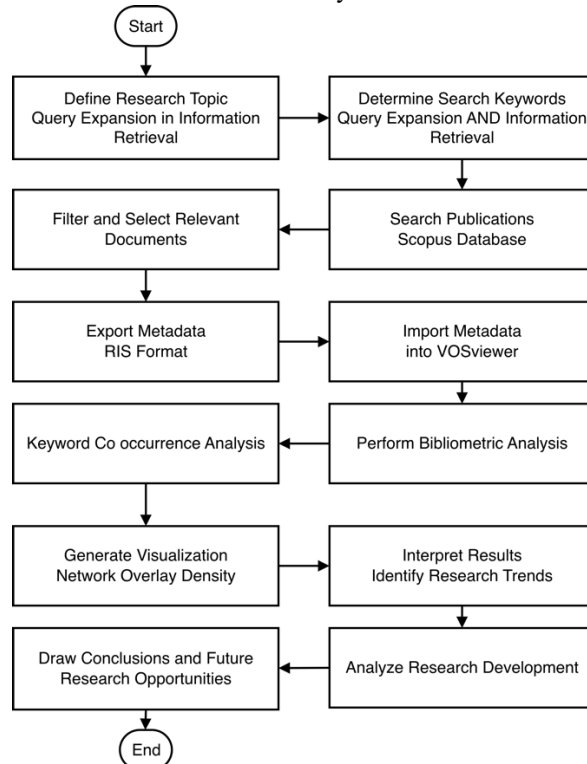


Fig. 1 Research Methodology Flowchart

\*Roberto Kaban



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

The search results were then filtered to ensure that the analyzed documents were truly relevant to the research topic. The metadata of the retrieved documents were exported in RIS format, which contains bibliographic information such as author names, document titles, publication years, keywords, and publication sources. The metadata file was subsequently used as the input for the bibliometric analysis process.

Data analysis in this study was carried out using the software VOSviewer. This software is used to visualize relationships among bibliographic elements in the form of networks through network visualization. VOSviewer is capable of mapping relationships among research keywords, collaborations among authors, and connections among research topics within a particular field. The resulting visualizations enable researchers to identify clusters of related research topics and understand the direction of future research development.

The research procedure consists of several main stages. The first stage involves searching for literature in the Scopus database using the predetermined keywords. The second stage includes selecting relevant articles and downloading their metadata. The third stage involves performing bibliometric analysis using VOSviewer to map relationships among keywords and identify research trends in the area of Query Expansion. The final stage is the interpretation of the visualization results to understand research patterns and the development of topics in Information Retrieval.

To clarify the research workflow, the research process can be illustrated using a flowchart that shows the stages from literature search to data analysis. The diagram describes the process of identifying articles in the Scopus database, exporting metadata, analyzing the data using VOSviewer, and interpreting the results of the bibliometric analysis. Through these stages, this study provides a systematic overview of the development of research on Query Expansion within the field of Information Retrieval.

The method used in this research is expected to contribute to understanding the structure of research development and identifying dominant research topic trends. In addition, the results of the analysis can serve as a foundation for determining future research directions in the areas of Query Expansion and Information Retrieval.

## RESULT

The results of this study are presented based on a bibliometric analysis of 676 scientific publications related to Query Expansion in the field of Information Retrieval. The publication data were obtained from the Scopus database. Table 1 presents several examples of recent publications that are relevant to the topic of this study.

Table 1  
Recent Publications on Query Expansion in Information Retrieval

No	Authors	Year	Title
1	Li R.; Cui M.; Wang B.; Wang Y.; Miao Q.	2026	Query Expansion with Topic-Aware In-Context Learning and Vocabulary Projection for Open-Domain Dense Retrieval
2	Malone D.	2026	Topic-Specific Corpus Compilation: A Componential Approach to Query Formulation
3	Ortiz-Viso B.; Gutiérrez-Batista K.; Ruiz M.D.; Martin-Bautista M.J.	2026	Schema-Based Inference for Query Expansion and Completion over Knowledge Graphs
4	Azad H.K.	2026	Query-Based Automatic Text Summarization Using Query Expansion Approach
5	Alier M.; García-Peñalvo F.J.; Casañ M.J.; Pereira J.; Granados A.	2026	Closing the Acronym Gap in Educational Retrieval with Ontology-Guided Augmentation
6	Miah M.H.; Gupta S.K.; Yali L.; Alsekait D.M.; Alzu'bi S.; Nabil A.; Chowdhury H.R.; Kanon M.	2026	AI-Based Decision-Making System for Tooling Design of Aircraft Product Assembly with Developed Knowledge Retrieval Algorithm
7	Komenda A.; Makowski M.; Can E.; Prucker P.; Busch F.; Wachter A.; Weller D.; Kim S.H.; Ziegelmayr S.; Bressemer K.; Adams L.C.	2026	Development and Evaluation of a Retrieval-Augmented Generation System for Radiology Guidelines
8	Keller J.; Fröbe M.; Hendriksen G.; Alexander D.; Potthast M.; Schaer P.	2026	Simplified Longitudinal Retrieval Experiments: A Case Study on Query Expansion and Document Boosting

9	Pednekar S.; Vegda D.C.	2026	Fine-Tuning SBERT for Semantic Search in Marathi News with Synonym-Based Positive Pairs
10	Marchi E.; Boldi P.; Casiraghi E.; Zapperi S.; La Porta C.A.M.	2026	Revisiting Minamata Disease through Computational Phenotypic Similarity Analysis
11	Solanki S.; Pandey R.; Patel P.; Kankariya D.P.; Kund G.W.; Verma A.B.	2026	Transforming Query Expansion: A Deep Learning Innovation for Next-Generation Information Retrieval
12	Yoshihara N.; Yamamoto T.; Shoji Y.	2026	Expanding Aspect Queries into Review Sentence Fragments for Product Comparison via LLM-Generated Synthetic Reviews
13	Vysotska V.; Klich L.; Nagachevska O.; Shvets M.	2026	Transformational Question-Answering System Based on Ukrainian Legal Documents Utilizing Large Language Models
14	Panchumarthi L.Y.; Nunna S.K.; Parchuri L.; Padmanabuni S.; Kallam Y.R.	2026	Improving Information Retrieval Task with Asymmetric Divergence from Randomness in Query Expansion
15	Jung S.; Yu S.J.; Yi B.K.	2026	Clinical Terminology Mapping Service Based on Information Retrieval
676	...	...	...

All publication data were subsequently analyzed using the VOSviewer software to map research trends, identify relationships among keywords, and examine the distribution of publications in the field of Query Expansion.

### Annual Publication Trends

Figure 2 illustrates the annual trend in the number of publications related to Query Expansion research in the field of Information Retrieval during the period from 2020 to February 2026. The publication data were obtained from the Scopus database. Based on the figure, the number of publications shows fluctuations across the observed years. In 2020, there were 103 publications, and the number increased to 140 publications in 2021. However, the number of publications declined in 2022 to 82 publications. In 2023 and 2024, the number of publications remained relatively stable, with 84 and 92 publications respectively.

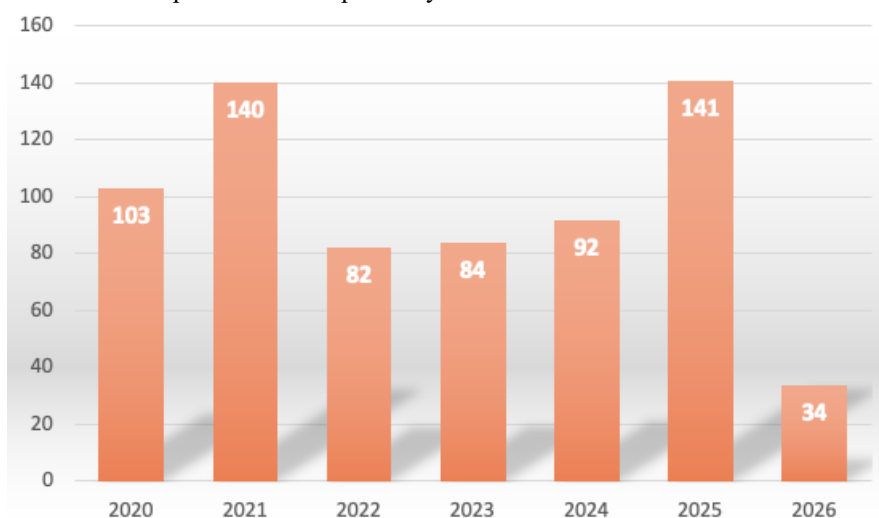


Fig. 2. Annual Publication Trends on Query Expansion in Information Retrieval (2020–February 2026)

A significant increase occurred again in 2025, when the number of publications reached 141, which represents the highest number during the observation period. Meanwhile, in 2026, a total of 34 publications were recorded. It should be noted that the data for 2026 only cover publications indexed up to February 2026, and therefore the number of publications may continue to increase by the end of the year.

### Distribution by Subject Area

Based on the subject area analysis, the majority of publications related to Query Expansion originate from the field of Computer Science with a total of 596 publications. This field represents the primary contributor because

\*Roberto Kaban



research on query expansion is closely related to the development of algorithms, information retrieval systems, and computational techniques. The second largest contributor is Engineering with 161 publications, followed by Mathematics with 137 publications, Social Sciences with 114 publications, and Decision Sciences with 88 publications. Table 2 presents the distribution of publications based on subject areas.

Table 2  
Distribution of Publications by Subject Area

Subject Area	Number of Publications	Subject Area	Number of Publications
Computer Science	596	Materials Science	15
Engineering	161	Physics and Astronomy	14
Mathematics	137	Energy	13
Social Sciences	114	Health Professions	11
Decision Sciences	88	Multidisciplinary	8
Medicine	37	Biochemistry, Genetics and Molecular Biology	7
Arts and Humanities	34	Environmental Science	7
Business, Management and Accounting	31	Chemical Engineering	6
Materials Science	15	Neuroscience	5
Physics and Astronomy	14	Agricultural and Biological Sciences	4
Energy	13	Earth and Planetary Sciences	4
Health Professions	11	Economics, Econometrics and Finance	3
Multidisciplinary	8	Chemistry	2
Biochemistry, Genetics and Molecular Biology	7	Immunology and Microbiology	1
Environmental Science	7	Nursing	1
Chemical Engineering	6	Pharmacology, Toxicology and Pharmaceutics	1

Research contributions also originate from several other fields such as Medicine with 37 publications, Arts and Humanities with 34 publications, and Business, Management and Accounting with 31 publications. Additional contributions are observed in Materials Science with 15 publications, Physics and Astronomy with 14 publications, Energy with 13 publications, and Health Professions with 11 publications.

Smaller contributions are found in fields such as Multidisciplinary with 8 publications, Biochemistry, Genetics and Molecular Biology with 7 publications, Environmental Science with 7 publications, and Chemical Engineering with 6 publications. Other fields such as Neuroscience with 5 publications, Agricultural and Biological Sciences with 4 publications, and Earth and Planetary Sciences with 4 publications also demonstrate the relevance of query expansion research across various application domains.

More limited contributions are identified in Economics, Econometrics and Finance with 3 publications, Chemistry with 2 publications, and several other fields including Immunology and Microbiology, Nursing, and Pharmacology, Toxicology and Pharmaceutics, each with one publication. This distribution indicates that research on Query Expansion is not only developing within computational disciplines but is also increasingly being applied across various scientific domains.

### Distribution by Document Type

The analysis of document types shows that the majority of publications related to Query Expansion in the field of Information Retrieval are published as conference papers, with a total of 369 publications. This is followed by journal articles with 270 publications. Other document types include book chapters and conference reviews, each with 12 publications, and review papers with 10 publications. In addition, there are a small number of publications in the form of books with 2 publications and editorials with 1 publication. Table 3 presents the distribution of publications based on document type.

Table 3

\*Roberto Kaban



Distribution of Publications by Document Type

Document Type	Number of Publications
Conference Paper	369
Article	270
Book Chapter	12
Conference Review	12
Review	10
Book	2
Editorial	1

This distribution indicates that conferences serve as one of the primary channels for disseminating recent research findings in the field of Query Expansion. Conferences often provide a platform for researchers to present emerging ideas, experimental results, and preliminary studies. Meanwhile, journal articles are generally used to publish more comprehensive and in-depth research contributions within the broader domain of Information Retrieval.

**Keyword Analysis**

Keyword analysis was conducted to identify the main research topics and the development trends related to Query Expansion in the field of Information Retrieval. Based on the extraction of keywords from the analyzed publications indexed in the Scopus database, several keywords were identified as having a high frequency of occurrence. Table 4 presents the most frequently occurring keywords in the analyzed publications.

Table 4  
Most Frequent Keywords in Query Expansion Research

Keyword	Occurrences	Keyword	Occurrences	Keyword	Occurrences	Keyword	Occurrences
Information Retrieval	483	Relevance Feedback	22	Information Storage and Retrieval	11	Retrieval Efficiency	8
Query Expansion	354	Retrieval Models	22	Knowledge Based Systems	11	Search System	8
Search Engines	221	Retrieval Performance	22	Metadata	11	Semantic Retrieval	8
Semantics	176	Automatic Query Expansion	21	Social Networking (Online)	11	Signal Encoding	8
Query Processing	171	Benchmarking	21	Text Retrieval	11	Text Generations	8
Expansion	117	Knowledge Management	21	Transformer Modeling	11	Clustering Algorithms	7
Language Model	99	State of the Art	21	Economic and Social Effects	10	Computer Vision	7
Natural Language Processing Systems	93	Query Expansion Techniques	20	Fuzzy Logic	10	Digital Libraries	7

\*Roberto Kaban



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Information Retrieval Systems	77	Retrieval Methods	20	Machine Learning	10	Domain Knowledge	7
Embeddings	76	Websites	20	Mismatch Problems	10	Extraction	7
Structured Query Language	73	Word Embeddings	20	Modeling Languages	10	Intelligent Systems	7
Query Languages	70	Article	19	Query Expansion Models	10	Latent Semantic Analysis	7
Computational Linguistics	60	Knowledge Graphs	19	Semantic Information Retrieval	10	Nearest Neighbor Search	7
Deep Learning	53	Query Terms	19	Thesauri	10	Neural Networks	7
Ontology	50	Classification (of Information)	18	Behavioral Research	9	Ranking Model	7
Information-retrieval Systems	49	Information Management	18	Codes (Symbols)	9	Recommender Systems	7
Natural Language Processing	46	Image Retrieval	17	Cross-language Information Retrieval	9	Reinforcement Learning	7
Re-ranking	46	Information Systems	17	Feedback	9	Large Scales	7
Large Language Model	45	Machine Learning	17	Image Enhancement	9	Matchings	7
Pseudo-relevance Feedbacks	44	Model-based OPC	17	LLM	9	Clusterings	7
Performance	42	Online Searching	17	Optimisations	9	Graph-based	7
Document Retrieval	37	Query Documents	17	Precision	9	Fine Grained	7
Semantic Web	36	Retrieval-augmented Generation	17	Precision Medicine	9		
Word Embedding	35	Semantic Similarity	17	Translation (Languages)	9		
Learning Systems	33	Simple++	17	Zero-shot Learning	9		
Question Answering	32	Learning to Rank	16	Adversarial Machine Learning	8		
Artificial Intelligence	31	Decision Making	15	Biomedical Information Retrieval	8		
Text Processing	30	Efficiency	15	BM25	8		

\*Roberto Kaban



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



Based on the network visualization, the keywords “query expansion” and Information Retrieval appear as the largest nodes and are positioned at the center of the network. This indicates that these two concepts are the most dominant topics and have strong connections with many other keywords in the research field.

The visualization also reveals the formation of several research clusters that represent different thematic areas within query expansion studies. Some clusters highlight classical approaches such as relevance feedback and pseudo-relevance feedback. Other clusters demonstrate the emergence of modern approaches involving artificial intelligence technologies, including word embedding, language models, and Large Language Models. These findings indicate a shift in research trends from traditional methods toward more advanced approaches based on machine learning and semantic representation techniques for improving the performance of information retrieval systems.

### Keyword Trend Analysis

In addition to network analysis, this study also utilizes the overlay visualization feature in VOSviewer to identify the evolution of research trends over time. This visualization illustrates the changes in research focus based on the publication year of keywords appearing in the analyzed documents obtained from the Scopus database. In this map, the color of each node represents the average publication year of the keyword. Darker colors indicate relatively older research topics, while brighter colors represent more recent research themes.

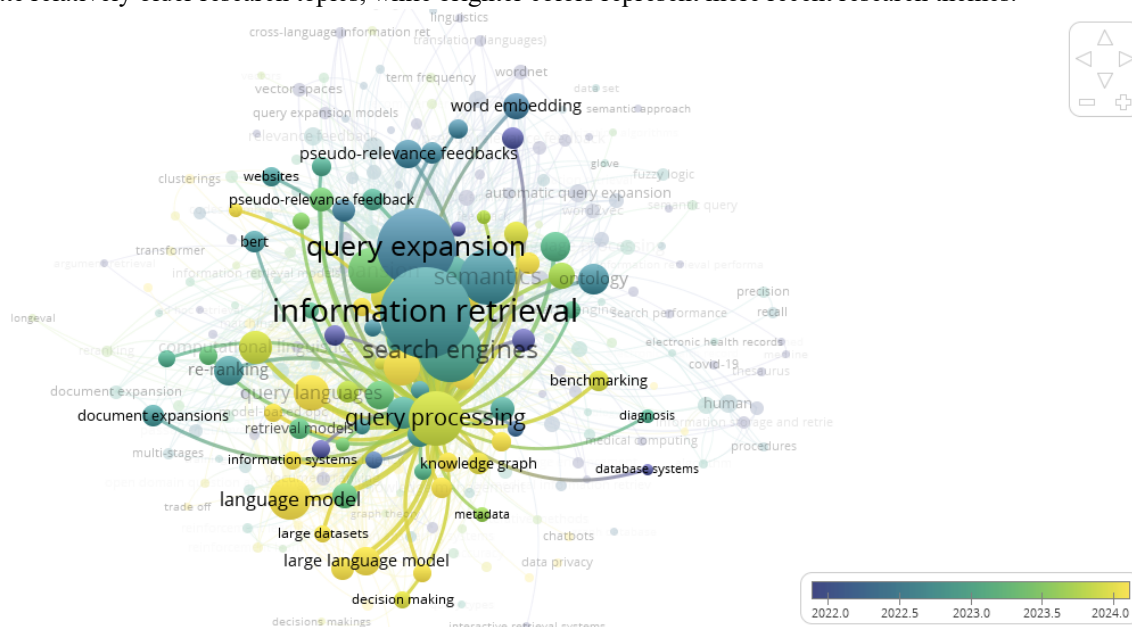


Fig. 4 Overlay Visualization of Keyword Trends

Based on Figure 4, several keywords such as Information Retrieval, Query Expansion, and search engines appear as core topics that have long been the primary focus of research and continue to maintain strong relationships with many other topics. These keywords represent the fundamental concepts that form the basis of research in this field.

Meanwhile, keywords represented with brighter colors indicate more recent research trends. Examples include Large Language Models, language models, and word embedding. The emergence of these keywords suggests that research on query expansion is increasingly incorporating artificial intelligence approaches and modern machine learning techniques. This visualization indicates a shift in research focus from traditional query expansion methods toward approaches that emphasize language models and semantic representation in the development of advanced information retrieval systems.

### Research Topic Density Analysis

In addition to network and temporal trend visualizations, bibliometric analysis was also conducted using the density visualization feature in VOSviewer to identify the level of density and concentration of research topics within the field of Query Expansion. This visualization highlights the areas of research that are most frequently discussed in scientific publications obtained from the Scopus database. In the density map, brighter colors such as yellow indicate keywords with a high frequency of occurrence, whereas darker colors represent keywords with lower frequencies.

\*Roberto Kaban



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



indicate that query expansion research is also being applied in various contexts, including user behavior analysis and information processing in digital social systems.

In addition to subject area distribution, the analysis of document types provides insight into the characteristics of publications in this research domain. The results show that most documents are published as conference papers, totaling 369 publications, followed by journal articles with 270 publications. Other document types such as book chapters, conference reviews, and review papers have relatively smaller numbers. The dominance of conference publications suggests that innovations in the field of query expansion develop rapidly and are often first presented in scientific conferences before being published in more comprehensive forms in academic journals.

Keyword analysis also provides an overview of the main research focus in this field. Based on the frequency of keyword occurrences, Information Retrieval with 483 occurrences and Query Expansion with 354 occurrences emerge as the most dominant topics. Other keywords such as search engines with 221 occurrences, semantics with 176 occurrences, and query processing with 171 occurrences also appear frequently. These findings indicate that query expansion research is strongly related to the development of information search systems aimed at improving the relevance of search results through more effective query processing.

In addition to these core keywords, several keywords related to recent technological developments also appear prominently, including Deep Learning, word embeddings, knowledge graphs, and Large Language Models. The emergence of these keywords indicates that research on query expansion increasingly integrates machine learning and artificial intelligence techniques to enhance the performance of information retrieval systems. This trend reflects an evolution in research approaches from traditional statistical and rule-based methods toward data-driven and semantic representation approaches.

The results of the keyword co-occurrence analysis reveal the conceptual structure of the research field, which forms several thematic clusters. The network visualization shows that the keywords “query expansion” and Information Retrieval are located at the center of the network with the largest node sizes, indicating the central role of these concepts in the analyzed research domain. The clusters that emerge also demonstrate differences in research approaches. Some clusters represent classical methods such as relevance feedback and pseudo-relevance feedback, while other clusters represent modern approaches involving technologies such as word embedding, language models, and Large Language Models.

Furthermore, the overlay visualization highlights the development of research trends over time. The visualization shows that classical topics such as information retrieval, query expansion, and search engines have long been the primary focus of research. However, several keywords that appear in brighter colors in the visualization, such as large language models, language models, and word embedding, represent more recent research trends. This indicates that query expansion research is increasingly evolving by incorporating modern artificial intelligence technologies to enhance contextual and semantic understanding in information search processes.

Meanwhile, the density visualization results indicate that topics such as Information Retrieval, Query Expansion, and search engines have the highest density levels in the research network. This confirms that these three topics form the core of research in this field. On the other hand, topics such as Large Language Models, data privacy, and chatbots have lower density levels, suggesting that these topics are relatively new and have not yet been extensively explored. This situation opens opportunities for future research to explore the integration of generative artificial intelligence technologies with query expansion methods.

Overall, the analysis results indicate that research on Query Expansion continues to evolve along with advances in Information Retrieval technologies and artificial intelligence. The integration of classical methods such as relevance feedback with modern approaches based on machine learning and large language models represents one of the most promising research directions for improving the quality and effectiveness of information retrieval systems in the future.

## CONCLUSION

This study analyzes the development of research on Query Expansion in the field of Information Retrieval through a bibliometric approach based on 676 scientific publications obtained from the Scopus database. The analysis was conducted using the VOSviewer software to identify publication trends, subject area distribution, document types, and relationships among research keywords. The results show that the number of publications fluctuated during the period 2020–2026, with the highest increase occurring in 2025. In terms of subject areas, the research is dominated by Computer Science, followed by Engineering, Mathematics, and Social Sciences. Furthermore, most of the publications are conference papers, indicating that innovations in the field of query expansion are rapidly disseminated through scientific conference forums.

Keyword analysis reveals that the main research topics include Information Retrieval, Query Expansion, and search engines, which represent the core concepts in the development of information retrieval systems. Network

visualization and research trend analysis also indicate a shift from classical approaches such as relevance feedback toward more modern artificial intelligence-based methods, including word embedding, Deep Learning, and Large Language Models. This study provides an overview of the structure and development trends of research on query expansion. It also highlights potential future research opportunities related to the integration of query expansion techniques with artificial intelligence technologies to improve the effectiveness and performance of information retrieval systems.

## REFERENCES

- Ahmed, M. (2024). Bibliometrix: An Easy Yet Powerful Approach for Quantitative and Qualitative Analyses of Scholarly Literature. *Information Research Communications*, 1(1), 43–45. <https://doi.org/10.5530/irc.1.1.7>
- Al-Lahham, Y. (2024). *Improved Arabic Query Expansion using Word Embedding*. <https://doi.org/10.21203/rs.3.rs-4065010/v1>
- Allahim, A., Cherif, A., & Imine, A. (2025). Semantic approaches for query expansion: Taxonomy, challenges, and future research directions. *PeerJ Computer Science*, 11, e2664. <https://doi.org/10.7717/peerj-cs.2664>
- Baumann, O., & Schoenfeld, M. (2024). *PSQE: Personalized Semantic Query Expansion for user-centric query disambiguation*. <https://doi.org/10.21203/rs.3.rs-4178030/v1>
- Bernard, N., & Balog, K. (2025). A Systematic Review of Fairness, Accountability, Transparency, and Ethics in Information Retrieval. *ACM Computing Surveys*, 57(6), 1–29. <https://doi.org/10.1145/3637211>
- Breuer, T., Frihat, S., Fuhr, N., Lewandowski, D., Schaer, P., & Schenkel, R. (2025). Large Language Models for Information Retrieval: Challenges and Chances. *Datenbank-Spektrum*, 25(2), 71–81. <https://doi.org/10.1007/s13222-025-00503-x>
- Ganti, L., Persaud, N. A., & Stead, T. S. (2025). Bibliometric analysis methods for the medical literature. *Academic Medicine & Surgery*. <https://doi.org/10.62186/001c.129134>
- Hambarde, K. A., & Proença, H. (2023). Information Retrieval: Recent Advances and Beyond. *IEEE Access*, 11, 76581–76604. <https://doi.org/10.1109/ACCESS.2023.3295776>
- Hidri, M. (2024). Learning-Based Models for Building User Profiles for Personalized Information Access. *Interdisciplinary Journal of Information, Knowledge, and Management*, 19, 010. <https://doi.org/10.28945/5275>
- Kaban, R., Sihombing, P., Efendi, S., & Lydia, M. S. (2025a). Enhancing Retrieval Performance in Social Media Using Corpus-Based Query Expansion. *2025 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, 1–6. <https://doi.org/10.1109/AIMS66189.2025.11229497>
- Kaban, R., Sihombing, P., Efendi, S., & Lydia, M. S. (2025b). Enhancing retrieval performance in social media with corpus-based query expansion using bidirectional encoder representations from transformers. *Eastern-European Journal of Enterprise Technologies*, 5(2 (137)), 70–83. <https://doi.org/10.15587/1729-4061.2025.340258>
- Kumar, R. (2025). Bibliometric Analysis: Comprehensive Insights into Tools, Techniques, Applications, and Solutions for Research Excellence. *Spectrum of Engineering and Management Sciences*, 3(1), 45–62. <https://doi.org/10.31181/sems31202535k>
- Meliukh, V., Potapova, E., Nalyvaichuk, M., & Dychka, A. (2025). Query expansion based on context-dependent sentiment analysis in databases with domain-specific filtering. *Eastern-European Journal of Enterprise Technologies*, 1(2 (133)), 6–17. <https://doi.org/10.15587/1729-4061.2025.322120>
- Naamha, E. Q., & Abdulmunim, M. E. (2024). Web Page Ranking Based on Text Content and Link Information Using Data Mining Techniques. *ARO-THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY*, 12(1), 29–40. <https://doi.org/10.14500/aro.11397>
- Pan, M., Liu, Y., Chen, J., Huang, E. A., & Huang, J. X. (2024). A multi-dimensional semantic pseudo-relevance feedback framework for information retrieval. *Scientific Reports*, 14(1), 31806. <https://doi.org/10.1038/s41598-024-82871-0>
- Pan, M., Xiong, W., Zhou, S., Gao, M., & Chen, J. (2025). LLM-Based Query Expansion with Gaussian Kernel Semantic Enhancement for Dense Retrieval. *Electronics*, 14(9), 1744. <https://doi.org/10.3390/electronics14091744>
- Patel, V., Hiran, D., & Dangarwala, K. (2024). Recent Trends of Information Retrieval System: Review Based on IR Models and Applications. In V. K. Gunjan & J. M. Zurada (Eds.), *Proceedings of 4th*

- International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications* (Vol. 873, pp. 619–629). Springer Nature Singapore. [https://doi.org/10.1007/978-981-99-9442-7\\_51](https://doi.org/10.1007/978-981-99-9442-7_51)
- Peikos, G., & Pasi, G. (2024). A systematic review of multidimensional relevance estimation in information retrieval. *WIREs Data Mining and Knowledge Discovery*, 14(5), e1541. <https://doi.org/10.1002/widm.1541>
- Raj, G. D., Mukherjee, S., Robin, C. R. R., & Jasmine, R. L. (2025). An Intelligent Feature Concatenation Process-Based Effective Query Expansion for Patent Retrieval Approach Using Optimal Bi-clustering and Enhanced Social Engineering Optimizer. *International Journal of Computational Intelligence Systems*, 18(1), 259. <https://doi.org/10.1007/s44196-025-00963-9>
- Roberts, K. (2024). Information Retrieval. In H. Xu & D. Demner Fushman (Eds.), *Natural Language Processing in Biomedicine* (pp. 195–230). Springer International Publishing. [https://doi.org/10.1007/978-3-031-55865-8\\_8](https://doi.org/10.1007/978-3-031-55865-8_8)
- Stathopoulos, E. A., Karageorgiadis, A. I., Kokkalas, A., Diplaris, S., Vrochidis, S., & Kompatsiaris, I. (2023). A Query Expansion Benchmark on Social Media Information Retrieval: Which Methodology Performs Best and Aligns with Semantics? *Computers*, 12(6), 119. <https://doi.org/10.3390/computers12060119>
- Venkatachalam, C., & Venkatachalam, S. (2023). *Optimal Intelligent Information Retrieval and Reliable Storage Scheme for Cloud Environment And E-Learning Big Data Analytics*. In Review. <https://doi.org/10.21203/rs.3.rs-2545685/v1>
- Vishwakarma, D., & Kumar, S. (2025). Fine-Tuned BERT Algorithm-Based Automatic Query Expansion for Enhancing Document Retrieval System. *Cognitive Computation*, 17(1), 23. <https://doi.org/10.1007/s12559-024-10354-5>
- Vladlenov, D. (2023). MODERN METHODS OF APPLYING SCIENTIFIC THEORIES. *Proceedings of the X International Scientific and Practical Conference*, 1–481. <https://doi.org/10.46299/ISG.P.2023.1.10>
- Wang, L., Yang, N., & Wei, F. (2023). Query2doc: Query Expansion with Large Language Models. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 9414–9423. <https://doi.org/10.18653/v1/2023.emnlp-main.585>
- Wang, Z., & Pei, Q. (2024). Dense Retrieval Systems with LLM-Based Query Expansion. *2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 682–686. <https://doi.org/10.1109/WI-IAT62293.2024.00110>
- Ye, F., Fang, M., Li, S., & Yilmaz, E. (2023). Enhancing Conversational Search: Large Language Model-Aided Informative Query Rewriting. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 5985–6006. <https://doi.org/10.18653/v1/2023.findings-emnlp.398>
- Yıldız, M., & Karakuş, T. (2024). Bibliometric Analysis in Scientific Research Using R: A Review of Scopus and Web of Science Databases. *Journal of Data Applications*, 0(2), 31–46. <https://doi.org/10.26650/JODA.1462396>
- Zahhar, S., Mellouli, N., & Rodrigues, C. (2025). Leveraging Sentence-Transformers to Overcome Query-Document Vocabulary Mismatch in Information Retrieval. In M. Barhamgi, H. Wang, X. Wang, E. Aïmeur, M. Mrissa, B. Chikhaoui, K. Boukadi, R. Grati, & Z. Maamar (Eds.), *Web Information Systems Engineering – WISE 2024 PhD Symposium, Demos and Workshops* (Vol. 15463, pp. 101–110). Springer Nature Singapore. [https://doi.org/10.1007/978-981-96-1483-7\\_8](https://doi.org/10.1007/978-981-96-1483-7_8)
- Zhang, L., Wu, Y., Yang, Q., & Nie, J.-Y. (2024). *Exploring the Best Practices of Query Expansion with Large Language Models* (arXiv:2401.06311). arXiv. <https://doi.org/10.48550/arXiv.2401.06311>