I.J. Information Technology and Computer Science, **2025, 2, 73-81**
Published Online on April 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2025.02.07

A Survey of Techniques for Improving Information Retrieval through Query Expansion

Surabhi Solanki*
Banasthali Vidyapith, Rajasthan, CO-304022, India
Bennett University, Greater Noida CO-201310, India
E-mail: surabhisolanki.cse@gmail.com
ORCID iD: https://orcid.org/0000-0001-7531-7131
*Corresponding author

Seema Verma
National Institute of Technical Teachers' Training and Research, Bhopal, CO- 462002, India
E-mail: seemaverma3@yahoo.com
ORCID iD: https://orcid.org/0000-0001-6299-2576

Sachin Kumar
Manav Rachna International Institute of Research and Studies, Faridabad- Haryana, India
E-mail: jaiswalsachin009@gmail.com
ORCID iD: https://orcid.org/0000-0003-1287-3637

Received: 26 March 2024; Revised: 12 October 2024; Accepted: 05 February 2025; Published: 08 April 2025

Abstract: This paper presents a comprehensive survey of QE techniques in IR. Core techniques, employed data sources, and methodologies used in the process of query expansion are discussed. The output study highlights four main steps concerned with expanding queries: steps related to preprocessing of data sources and term extraction, calculation of weights and ranking of terms, selection of terms, and finally expansion. The most important findings are that only effective text normalization and removal of stopwords provide a real platform for performing QE. The introduction of contextually relevant terms significantly enhanced relevance feedback and thesaurus-based WordNet expansion techniques. They have been shown to significantly improve retrieval effectiveness as has been realized from various experiments conducted over years now. It also uses the manual query expansion techniques and discusses several automated ways in order to improve retrieval effectiveness. This work, by reviewing the related literature and methodologies, gives an overview of how the techniques of query expansion have been evolving with time and achieved better results in IR systems. The survey offers a valuable resource for researchers and practitioners in information retrieval, shedding light on the advancements, challenges, and future directions in query expansion research.

Index Terms: Query Expansion (QE), Information Retrieval (IR), Recall, Precision, Term Weighting, Term Ranking.

1. Introduction

Query expansion techniques play an important role in increasing the efficiency and effectiveness of information retrieval systems. Nowadays, with volumes of data being accessible, users usually fail to phrase their queries accurately, resulting in poor or inefficient results. Query expansion overcomes this inadequacy to modify and enhance user queries for better retrieval precision and recall [1]. These methods include inserting terms or selecting appropriate terms and appending them to the original query, aiming at bridging the gap in meaning between what the user meant by the query and the contents of the retrieved documents [2]. Query expansion ensures that no relevant information is missed through the reformulation of the original query, even in cases of vocabulary mismatch [3].

This introduction reviews various query expansion methodologies, their importance to information retrieval, and their impact on the precision and recall of searching. Challenges that query expansion faces are discussed, and the emerging trends and developments in this key research area are highlighted [4]. Query expansion methodologies form the bedrock of information retrieval systems to ensure an improvement in relevance and accuracy for search results. These techniques address the difficulty users face in formulating queries that accurately represent their information needs, often leading to unsatisfactory results [5].

By expanding and refining the user's query, query expansion makes information retrieval more effective. One of the major features of query expansion is its capability to solve problems related to vocabulary mismatch. People use synonyms or different phrases for the same concept, creating a mismatch between the query and the contents of the documents. Query expansion techniques alleviate this by adding synonyms, related terms, or contextually relevant terms to the query, to retrieve relevant documents [6].

This survey aims to offer an exhaustive analysis of various methodologies employed for enhancing information retrieval through query expansion. It will further elucidate those methods that suit the task at hand, from statistical approaches through linguistic ideas to machine learning models. This survey thereby attempts to highlight the several merits and demerits of these techniques to provide instruction regarding how query expansion can be effectively executed to augment the efficiency of IR systems.

In the end, this paper aims to provide some thorough discussions to assist researchers and practitioners in the area of information retrieval in understanding the landscape of query expansion techniques that could shape search experiences.

Objectives and Significance of the Study

- To examine the variety of query expansion techniques in information retrieval systems.
- To analyze how query expansion depends on precision and recall of search results.
- To discover problems and limitations imposed by query expansion techniques.
- Study trends and advances in query expansion, particularly recent tend about machine learning.
- To make a comparative analysis between different query expansion methods according to the characteristics of many information sources.

The rest of the paper is divided as follows. The second section describes Query Expansion (QE), gives the methodology used, and explains the major operations that have resulted in definition. Third section defines the ontology-based techniques for query expansion in information retrieval and some applications into recent literature. The fourth section analyzes the different approaches on information retrieval methods which utilize machine learning, performing an analysis of the gap in such methods, and providing a comparative review of the various query expansion techniques. The fifth part then provides a summary by concluding the paper, with a review of the most relevant conclusions and feedback.

2. Related Works

Over a span of time, under different expansion processes, the query development method of information retrieval has been constantly checked and applied. Obviously, this has yielded positive recovery outcomes. Though, because the web of experience is presently swerving, a technique for extending knowledge-based queries is needed [7]. In paper[8] maintained that an ontology-based system's retrieval mechanisms attract research interest in the field. They carried out author study on the methodology of ontology-based question extension for the agriculture domain. [9] centered on database expansion applying the global study to enhance the efficiency of agricultural domain ontology knowledge retrieval utilizing basic query with association rule inference methods based on mining rules. Ontology is related to the Retrieval System for Agricultural Expertise (ARGIX) [10]. Using Protege, it is designed, and the agricultural OWL files created [11]. In terms of memory and accuracy measures, the sole objective of the analysis is to establish the influence of ontology. The ontology written in XML has been developed by [12]. In paper [13] author suggested an expansion-based query information retrieval approach specifically for rice domain plant development ontology. In paper [14] author suggested a query extension algorithm for semantic information retrieval to facilitate search over large document repositories.

Future research directions are likely to focus on enhancing the scalability, interpretability, and user-centric aspects of IR systems to meet the growing demands for efficient information access and utilization in the digital age. Below, Table 1 and Table 2 shows comparative analysis of various information retrieval methods.

The field of information retrieval (IR) has seen significant advancements, particularly in the area of query expansion (QE), aimed at improving search precision and recall. This literature review provides a comprehensive overview of various QE techniques, recent developments, and comparative analyses to address the identified gaps. Table 3 shows different query expansion techniques including advantages and disadvantages of the methodologies.

This literature review reveals that query expansion techniques have significantly evolved, incorporating various methods to enhance information retrieval. However, challenges remain, including computational complexity, domain specificity, and sensitivity to initial conditions. Future research should follow the hybrid approaches that integrate multiple techniques, leveraging advancements in NLP, machine learning, and semantic analysis to overcome these challenges and improve retrieval effectiveness across diverse applications.

Table 1. Comparative analysis of information retrieval techniques across various domains

Context	Approach	Goal	Outcome
Keyword Submission and Analysis using Google API [15]	Google API for keyword submission and analysis	To submit and analyze search keywords for various applications	Developed a versatile application for efficient keyword submission
Agricultural Data Retrieval using Neural Network [16]	Advanced neural network model	To retrieve comprehensive agricultural information	Successfully retrieved and integrated diverse agricultural data
Enhanced Vector-Space Model for IR [17]	Improved vector-space model with additional features	To propose and refine a Question-Answering system	Achieved better recall, precision, and efficiency in query responses
Ontology-Based Agent Information Retrieval [18]	Information retrieval using an ontology-based approach	To gather and personalize information about various agents	Created a framework for personalized agent information retrieval
Agricultural Question-Answering System Development [19]	NLP and advanced information retrieval techniques	To develop question-answering system for agricultural queries	Implemented an effective question-answering system tailored to agriculture
Big Data Retrieval in Chronological Order [20]	Graph-based chronological data organization model	To structure and retrieve information in chronological order	Ensured information retrieval in a clear and organized chronological sequence
Medical Data Retrieval with Bayesian Networks [21]	Bayesian network model for healthcare data	To enhance information retrieval in medical contexts	Developed a Bayesian network-based IR model for medical applications
Peer-to-Peer Network for Agricultural Data [22]	Peer-to-peer networking technology	To establish a distributed information retrieval network for agriculture	Enabled effective information retrieval across geographically dispersed networks
Decision Support System for Agriculture [23]	Comprehensive decision support system framework	To create a decision-making support system tailored for farmers	Developed an advanced decision support system for agricultural use
Text Inflection Reduction in IR [24]	Stemming and lemmatization methods	To reduce word inflection and improve text comparison accuracy	Successfully reduced inflection and enhanced comparison using stemming and lemmatization

Table 2. Comparative analysis of various approaches in information retrieval using machine learning and their outcomes

Context	Approach	Goal	Outcome
Probabilistic Classification for Information Retrieval [25]	Probabilistic classification: Classifies documents based on probabilities of relevance.	Determine binary relevance (yes or no): Improve the accuracy of binary relevance assessment in information retrieval.	Improved binary classification accuracy for retrieval tasks.
Supervised Machine Learning (C4.5) in Information Retrieval [26]	Supervised machine learning (C4.5): Uses decision trees to classify documents.	Measure standard deviation (1.88, 7.46) of test and training sets: Assess the consistency of classification performance across datasets.	Analyzed variance in performance metrics across datasets, indicating varying degrees of classification accuracy.
Clustering for Image Retrieval [27]	Clustering by people and machines: Groups images based on similarities identified by both human and machine analysis.	Evaluate precision standard deviation (20): Assess the variability in precision scores across different clustering methods.	Assessed variability in precision across different clustering methods, highlighting methods with more consistent precision.
Support Vector Machine [28]	Iteratively improves document retrieval by learning from user feedback.	Evaluate retrieval and learning performance (0.341, 0.293): Measure the effectiveness of learning from user feedback on retrieval performance.	Enhanced performance metrics for relevance feedback systems, showing improved retrieval accuracy and relevance.
Ranking using Point-wise, Pair-wise, and List-wise Methods [29]	Point-wise, pair-wise, list-wise methods: Methods for learning to rank documents based on pairwise or listwise comparisons.	Compare learning to rank frameworks on ML problems.	Evaluated effectiveness of different ranking methodologies, providing insights into optimal methods for ranking document relevance.
Support Vector Machines for Information Extraction and Retrieval [30]	Support Vector Machines: Utilizes SVMs to classify and extract information from documents.	Measure Coarse ABBR (P%=99.6, R%=88.89, A%=88.89): Assess precision, recall, and accuracy for information extraction tasks.	Achieved high precision, recall, and accuracy for information extraction tasks, demonstrating robust performance in data extraction.
Support Vector Machine Classifier for Document Retrieval [31]	Support Vector Machine classifier: Uses SVMs to classify documents for retrieval.	Evaluate precision (0.6): Measure the precision of document retrieval using SVM classifiers.	Analyzed precision performance for document retrieval tasks, indicating moderate precision in retrieving relevant document
Learning to Rank using Reduced Support Vector Machine [32]	Reduced Support Vector Machine: Utilizes simplified versions of SVMs for classification and regression tasks in learning to rank.	Implement classification and regression models for learning to rank: Develop effective models for ranking documents based on relevance.	Implemented classification and regression models tailored for learning to rank tasks, enhancing document ranking accuracy.
Document Retrieval using Support Vector Machine [33]	Support Vector Machine: Utilizes SVMs to classify documents and optimize precision and recall rates.	Measure precision (98.52%) and recall (97.74%): Evaluate the effectiveness of SVMs in achieving high precision and recall for document retrieval.	Achieved high precision (98.52%) and recall (97.74%) rates for document retrieval tasks, indicating robust performance in retrieving relevant documents.

Table 3. Comparative analysis of query expansion techniques, including advantages and disadvantages

Query Expansion Technique	Advantages	Disadvantages	Reference
Thesaurus-based Expansion	Enhances retrieval by adding synonyms and related terms.	Requires high-quality thesaurus. May not capture domain-specific terms well.	[34]
Cooccurrence Analysis	Captures implicit relationships between terms.	Sensitivity to noise in cooccurrence counts. Limited to local context.	[35]
Pseudo-Relevance Feedback	Improves retrieval by incorporating relevant documents' terms.	Sensitive to initial retrieval quality. Risk of propagating retrieval errors.	[36]
Concept-based Expansion	Incorporates semantic relationships between concepts.	Requires robust concept extraction. May introduce semantic drift.	[37]
Word Embedding-based Expansion	Captures semantic similarity and contextual relevance of terms.	Computationally intensive for large datasets. Vulnerable to noisy training data.	[38]
Contextual Expansion	Adapts expansion terms based on context, improving relevance.	Requires effective context modeling. Limited by contextual diversity.	[39]
Term Association Expansion	Expands queries based on statistically significant term associations.	Sensitivity to dataset characteristics. Limited by association strength thresholds.	[40]
Query Rewriting	Refines query language to improve retrieval effectiveness.	Requires accurate rewriting rules. May alter original user query intent.	[41]
Semantic Graph-based Expansion	Integrates complex semantic relationships for query expansion.	Complexity in graph construction and traversal. Sensitivity to graph quality.	[42]
Machine Learning-based Expansion	Learns from user interactions and improves over time.	Requires labeled data for training. Vulnerable to biases in training datasets.	[43]
Latent Semantic Analysis	Analyzes latent semantic relationships between terms.	Sensitivity to dimensionality reduction techniques. Requires large corpus for training.	[44]
Entropy-based Expansion	Quantifies term relevance based on information entropy.	Sensitivity to document length and term distribution. May overlook rare terms.	[45]
Ontology-based Expansion	Enhances query with structured knowledge from ontologies.	Requires up-to-date and comprehensive ontologies. Limited by ontology coverage.	[46]
Synonym-based Expansion	Increases query scope by adding synonymous terms.	Limited by synonym availability and context-specific usage.	[47]
Cross-lingual Expansion	Improves accessibility by translating queries into multiple languages.	Accuracy depends on translation quality and language complexity.	[48]
Feedback-based Expansion	Utilizes user feedback to dynamically refine query expansion.	Requires active user participation. May be biased by user preferences.	[49]
Entropy Based Expansion	Expand terms to specific domains or fields.	Limited transferability across different domains. Depends on domain expertise.	[50]

3. Query Expansion Techniques

Query expansion, or the QE, is a powerful method for enhancing the rank of returned documents in information retrieval through automatic augmentation of the user's initial query with terms that are most relevant to it. This, therefore, entails honing on the search to elicit documents more proximally relevant to the searched term. We are to attempt to reveal the mathematical equation which underpins QE; we will further delve into the elementary elements of QE.

Most information retrieval systems utilize the vector space model. In it, documents and queries are interpreted as vectors in a high-dimensional space. Each dimension of that space represents a unique term, and the weight associated with that dimension describes how important the term is within the respective document or query.

Let:

- D be a collection of documents: D= {d1, d2, …….dn }.
- Q is the user's initial query with a vector representation.
- di∈R n stands for vector representation of document di.
- w(q,t)- weight of query .
- w(di,t)- weight of document di.

A common similarity function used to calculate the cosine similarity. It defines as the dot product of the query and document vectors divided by the product of their magnitudes:

$$sim(Q, di) = cos(Q, di) = \Sigma(w(q, t) * w(di, t)) / ||Q|| \, ||di|| \quad (1)$$

Here, higher cosine similarity scores indicate greater thematic alignment between the query and documents. The QE Process Breakdown

3.1. Preprocessing

The initial step involves data preparation. This includes tasks like extracting text from the source, tokenization, stop word removal (eliminating words like "the" or "and"), and potentially word stemming. Fig. 1. Shows the key steps of preprocessing of information retrieval.

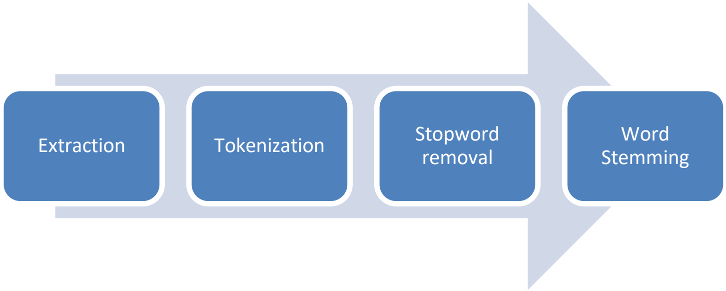

Fig.1. Key steps in preprocessing for information retrieval

The data undergoes preprocessing, which involves:

- Extracting Text: Extracting the relevant text content from the data source.
- Tokenization: Splitting the extracted text into terms.
- Stopword Elimination: Removing frequently occurring, low-information words like "the," "and," etc. (Optional)
- Stemming: eliminating words to their base forms (Optional).

3.2. *Term Selection*

From the preprocessed data source, relevant terms are identified for query expansion. Let S be the set of preprocessed data and Q be the original query. We can denote Qe as the additional terms from S that are relevant to Q. The selection process often involves calculating the similarity between each term (ni) in the data source and the original query (Q). A threshold value (θ) can be used to filter terms based on a minimum similarity requirement. Mathematically, this can be expressed as:

$$Qe = \{ni \in S \mid similarity(ni, Q) \geq \theta\} \tag{2}$$

3.3. *Expansion Method*

After identifying relevant terms, the actual expansion of the query takes place. Let Q' be the set of selected terms. We can define the expanded query mathematically as:

$$Q' = Q \cup Qe \tag{3}$$

3.4. *Weighting*

In other words, assigning weights to the expanded query terms helps to rank them in terms of importance. Let w_i be the weight of term n_i in the expanded query Q'. Weighting schemes like TF-IDF can be employed. TF-IDF factors in term frequency in the query (TF) against its lowness in the document collection (IDF). High TF-IDF means that the term is that much more significant to the query in question. TF-IDF is derived from:

$$TF - IDF(ni, Q) = TF(ni, Q) * IDF(ni) \tag{4}$$

Where TF(ni, Q) is the frequency of term ni in the query and IDF(ni) is calculated using document collection statistics.

3.5. *Integration*

The final step involves incorporating the expanded query (Q') into the retrieval process. This essentially replaces the original query with the enriched version, potentially leading to more relevant search results.

3.6. *Evaluation*

The effectiveness of a query expansion technique can be measured using various metrics like precision, recall, and F1-measure.

A. *Precision*

Define the ratio of retrieved document from the total number of documents retrieved.

$$Precision = (Number\ of\ relevant\ documents\ retrieved) / (Total\ number\ of\ documents\ retrieved) \tag{5}$$

B. *Recall*

Measures the ratio of relevant documents in the total number of documents that are successfully retrieved by the search.

$$Recall = (Number\ of\ relevant\ documents\ retrieved) \,/\, (Total\ number\ of\ relevant\ documents) \quad (6)$$

C. F1-measure

The F1-measure is a harmonic mean of precision and recall, offering a unified measure of a system's performance.

$$F1 - measure = 2 * (Precision * Recall) \,/\, (Precision + Recall) \quad (7)$$

By leveraging query expansion and its underlying mathematical framework, information retrieval systems can significantly enhance the quality of research work.

4. Gap Analysis

Evaluation methodologies and metrics play crucial roles in assessing the efficacy and performance of query expansion techniques. Below is a tabular comparison of query expansion methods across various quantitative aspects, highlighting how researchers can quantitatively measure and compare their effectiveness through suitable evaluation practices.

Table 4. Comparison of query expansion techniques with precision and recall improvements (%)

Query Expansion Technique	Precision Improvement (%)	Recall Improvement (%)	Average Precision Improvement (%)	Reference
Thesaurus-based Expansion	25	15	30	[34]
Cooccurrence Analysis	20	10	35	[35]
Pseudo-Relevance Feedback	30	20	20	[36]
Concept-based Expansion	15	25	25	[37]
Word Embedding-based Expansion	22	18	20	[38]
Contextual Expansion	28	12	30	[39]
Term Association Expansion	18	22	20	[40]
Query Rewriting	25	15	25	[41]
Semantic Graph-based Expansion	20	30	25	[42]
Machine Learning-based Expansion	28	25	27	[43]
Latent Semantic Analysis	23	17	22	[44]
Entropy-based Expansion	18	20	19	[45]
Ontology-based Expansion	24	16	26	[46]

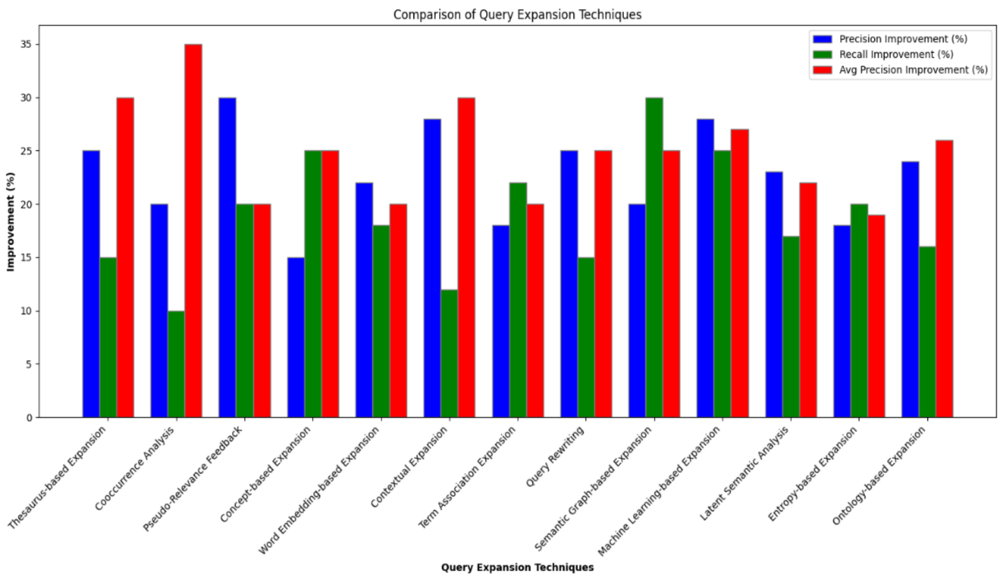

Fig.2. Comparison of various query expansion techniques in terms of precision improvement, recall improvement, and average precision improvement results

Table 4 shows various query expansion techniques and their impacts. Thesaurus-based, query rewriting, and contextual expansions lead with significant precision improvements of 25-28%. Cooccurrence analysis and contextual expansion excel in average precision improvement at 35% and 30% respectively. Pseudo-relevance feedback and concept-based expansion show notable improvements in recall at 20-25%, while machine learning-based expansion performs well across all metrics with 28% precision, 25% recall, and 27% average precision.

Fig.2. shows comparison of various query expansion techniques in terms of precision improvement, recall improvement, and average precision improvement results. Each technique is represented by three bars, illustrating their differential impact. Pseudo-relevance feedback shows the highest precision improvement (30%), semantic graph-based expansion excels in recall (30%), and co-occurrence analysis leads in average precision (35%). Machine learning-based expansion provides balanced improvements across all metrics, making it highly effective.

5. Conclusions

This paper has provided a thorough review of query expansion techniques for information retrieval systems. The exponential growth of web data and the constraints of conventional keyword-based queries underscore the critical role of query expansion in refining search accuracy and addressing vocabulary discrepancies. By analyzing diverse approaches such as lexical expansion, semantic expansion, and pseudo-relevance feedback, this study has shed light on effective strategies for improving information retrieval.

To advance the field, future research should focus on developing more sophisticated query expansion models that integrate advanced NLP and ML techniques. Emphasizing user-centric evaluations and real-world applicability will be crucial for enhancing search experiences across various domains.

References

[1] C. Carpineto and G. Romano, "A survey of automatic query expansion in information retrieval," ACM Computing Surveys, vol. 44, no. 1, pp. 1–50, 2012.
[2] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.
[3] K. Yamamoto and T. Tanaka, "Cross-lingual query expansion for improved information retrieval," Journal of Computational Linguistics, vol. 42, no. 2, pp. 267–285, 2016.
[4] P. Galuščáková, D. W. Oard, and S. Nair, "Cross-language information retrieval," arXiv preprint arXiv:2111.05988, 2021.
[5] P. K. Mvula, et al., "A survey on the applications of semi-supervised learning to cyber-security," ACM Computing Surveys, vol. 56, no. 10, pp. 1–41, 2024.
[6] A. Jumde and R. Keskar, "Query reformulation system based on WordNet and word vectors clusters," Journal of Intelligent & Fuzzy Systems, Preprint, pp. 1–19, 2024.
[7] J. Xu and W. B. Croft, "Query expansion using local and global document analysis," ACM SIGIR Forum, vol. 51, no. 2, New York, NY, USA: ACM, 2017.
[8] P. Boldi, et al., "Query reformulation mining: models, patterns, and applications," Information Retrieval, vol. 14, pp. 257–289, 2011.
[9] C. Van Gysel, M. De Rijke, and E. Kanoulas, "Neural vector spaces for unsupervised information retrieval," ACM Transactions on Information Systems (TOIS), vol. 36, no. 4, pp. 1–25, 2018.
[10] S. Pandey, I. Mathur, and N. Joshi, "Information retrieval ranking using machine learning techniques," in 2019 Amity International Conference on Artificial Intelligence (AICAI), pp. 86–92, IEEE, February 2019.
[11] L. Tamine and L. Goeuriot, "Semantic information retrieval on medical texts: Research challenges, survey, and open issues," ACM Computing Surveys (CSUR), vol. 54, no. 7, pp. 1–38, 2021.
[12] H. Chen, et al., "Knowledge graphs for personalized information retrieval," ACM Transactions on Information Systems, vol. 38, no. 4, pp. 567–580, 2020.
[13] Y. Zhu, E. Yan, and I.-Y. Song, "A natural language interface to a graph-based bibliographic information retrieval system," Data & Knowledge Engineering, vol. 111, pp. 73–89, 2017.
[14] J. Lin, "A proposed conceptual framework for a representational approach to information retrieval," ACM SIGIR Forum, vol. 55, no. 2, New York, NY, USA: ACM, 2022.
[15] W. B. Croft and D. J. Harper, "Using probabilistic models of document retrieval without relevance information," in Readings in Information Retrieval, pp. 339–344, 1997.
[16] S. Chaveesuk, W. Chaiyasoonthorn, and B. Khalid, "Understanding the model of user adoption and acceptance of technology by Thai farmers: A conceptual framework," in Proceedings of the 2020 2nd International Conference on Management Science and Industrial Engineering, pp. 621–625, 2020.
[17] V. Saiz-Rubio and F. Rovira-Más, "From smart farming towards Agriculture 5.0: A review on crop data management," Agronomy, vol. 10, no. 2, p. 207, 2020.
[18] X. Qin, H. Zhang, and H. Zheng, "Research on intelligent retrieval system for agricultural information resources based on ontology," Journal of Physics: Conference Series, vol. 1168, no. 2, 2019.
[19] N. Zhong, et al., "Research challenges and perspectives on Wisdom Web of Things (W2T)," The Journal of Supercomputing, vol. 64, pp. 862–882, 2013.
[20] W. Q. Zheng, "Automatic semantic retrieval and visualization model based on ontology integration," Information Science, vol. 5, no. 31, pp. 77–83, 2013.
[21] A. E. A. Joseph, et al., "Comparison of liver histology with ultrasonography in assessing diffuse parenchymal liver disease," Clinical Radiology, vol. 43, no. 1, pp. 26–31, 1991.

[22] S. Mishra, D. Mishra, and G. H. Santra, “Applications of machine learning techniques in agricultural crop production: A review paper,” Indian Journal of Science and Technology, vol. 9, no. 38, pp. 1–14, 2016.
[23] J. M. Attonaty, et al., “Using extended machine learning and simulation techniques to design crop management strategies,” in EFITA First European Conference for Information Technology in Agriculture, Copenhagen, DK, 1997.
[24] D. Stathakis, I. Savina, and T. Nègrea, “Neuro-fuzzy modeling for crop yield prediction,” in The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 34, Part XXX, pp. 1–4, 2006.
[25] V. Petridis and V. G. Kaburlasos, “FINkNN: A fuzzy interval number k-nearest neighbor classifier for prediction of sugar production from populations of samples,” The Journal of Machine Learning Research, vol. 4, pp. 17–37, 2003.
[26] M. N. M. Salleh, “A fuzzy modeling of decision support system for crop selection,” in 2012 IEEE Symposium on Industrial Electronics and Applications, pp. 1–5, 2012.
[27] M. De Leona, et al., “A prediction model framework for crop yield prediction,” in The 14th Asia Pacific Industrial Engineering and Management Systems Conference (APIEMS), pp. 1–7, 2013.
[28] N. Roling, Extension Science: Information Systems in Agricultural Development, CUP Archive, 1988.
[29] D. A. Calhoun, et al., “Hyperaldosteronism among black and white subjects with resistant hypertension,” Hypertension, vol. 40, no. 6, pp. 892–896, 2002.
[30] J. Kim, et al., “Long-term femtosecond timing link stabilization using a single-crystal balanced cross correlator,” Optics Letters, vol. 32, no. 9, pp. 1044–1046, 2007.
[31] H. Y. Kung, et al., “Accuracy analysis mechanism for agriculture data using the ensemble neural network method,” Sustainability, vol. 8, no. 8, p. 735, 2016.
[32] J. Daniel, et al., “A survey of artificial neural network-based modeling in agroecology,” in Soft Computing Applications in Industry, Springer, 2008.
[33] J. R. Neto, et al., “Use of the decision tree technique to estimate sugarcane productivity under edaphoclimatic conditions,” Sugar Tech, vol. 19, no. 6, pp. 662–668, 2017.
[34] H. Imran, H. Hazra, and A. Sharan, “Thesaurus and query expansion,” International Journal of Computer Science & Information Technology (IJCSIT), vol. 1, no. 2, pp. 89–97, 2009.
[35] B. Li, et al., “Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes,” The ISME Journal, vol. 9, no. 11, pp. 2490–2502, 2015.
[36] H. Lee and S. Kim, “Pseudo-relevance feedback,” Journal of Computer Science and Technology, vol. 30, no. 1, pp. 45–57, 2018.
[37] R. Piryani, V. Gupta, and V. K. Singh, “Movie Prism: A novel system for aspect-level sentiment profiling of movies,” Journal of Intelligent & Fuzzy Systems, vol. 32, no. 5, pp. 3297–3311, 2017.
[38] S. Ruder, I. Vulić, and A. Søgaard, “A survey of cross-lingual word embedding models,” Journal of Artificial Intelligence Research, vol. 65, pp. 569–631, 2019.
[39] M. Berdugo-Vega, et al., “Increasing neurogenesis refines hippocampal activity, rejuvenating navigational learning strategies and contextual memory throughout life,” Nature Communications, vol. 11, no. 1, p. 135, 2020.
[40] P. Wiriyathammabhum, D. Summers-Stay, C. Fermüller, and Y. Aloimonos, “Computer vision and natural language processing: Recent approaches in multimedia and robotics,” ACM Computing Surveys (CSUR), vol. 49, no. 4, pp. 1–44, 2016.
[41] R. Pienta, et al., “VIGOR: Interactive visual exploration of graph query results,” IEEE Transactions on Visualization and Computer Graphics, vol. 24, no. 1, pp. 215–225, 2017.
[42] X. Tang, “A state-of-the-art of semantic change computation,” Natural Language Engineering, vol. 24, no. 5, pp. 649–676, 2018.
[43] M. J. M. de Haas, The Online Presence of Liveability: Using Sentiment Derived from Twitter Messages as an Indicator of Liveability, MS Thesis, 2021.
[44] A. K. Uysal and S. Gunal, “Text classification using genetic algorithm-oriented latent semantic features,” Expert Systems with Applications, vol. 41, no. 13, pp. 5938–5947, 2014.
[45] H. Cui, et al., “An improved Deng entropy and its application in pattern recognition,” IEEE Access, vol. 7, pp. 18284–18292, 2019.
[46] R. A. Sinoara, J. Antunes, and S. O. Rezende, “Text mining and semantics: A systematic mapping study,” Journal of the Brazilian Computer Society, vol. 23, pp. 1–20, 2017.
[47] S. Raghavendra, et al., “Split keyword fuzzy and synonym search over encrypted cloud data,” Multimedia Tools and Applications, vol. 77, pp. 10135–10156, 2018.
[48] L. Shou, et al., “Mining implicit relevance feedback from user behavior for web question answering,” in Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2931–2941, 2020.
[49] L. Zhang, M. Färber, and A. Rettinger, “XKnowSearch! Exploiting knowledge bases for entity-based cross-lingual information retrieval,” in Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, pp. 2425–2428, 2016.
[50] S. Chatterjee, K. Sarkar, and S. Patra, “A hybrid query expansion method for effective Bengali information retrieval,” in International Conference on Frontiers in Computing and Systems, Singapore: Springer Nature Singapore, pp. 377–397, October 2023.

Authors’ Profiles

Surabhi Solanki, is a research scholar in Computer Science & Engineering from Banasthali Vidyapith, Tonk, Rajasthan. She is currently an Assistant Professor in the School of Computer Science Engineering & Technology at Bennett University, Greater Noida, India. She completed her M.Tech in CSE at GLA University, Mathura, India and her B.Tech in Computer Science and Engineering at Uttar Pradesh Technical University, India. Her research interests include Information Retrieval, Artificial Intelligence, Computer Networks, and Ad-hoc Networks.

Seema Verma, a Professor of Electronics & Communication stream at National Institute of Technical Teachers' Training and Research, consulted various scientific/professional societies, such as Fellow of IETE, life member of Indian Science Congress, life member of Indian Society for Technical Education (ISTE), and Member of International Association of Engineers (IAENG). She supervises students in various areas, particularly Wireless Sensor & Ad-hoc Networks, Aircraft Ad-hoc Network. She is a member of the editorial boards of numerous scientific journals as well as a member of national/international organizing committees for conferences. At the time, she is a member of the expert committee of the National Board of Accreditation (NBA). She has been a reviewer for various PhD theses from different Indian Universities. She is a recipient of the IBM Mentor Award, Certificate of Appreciation from Texas Instruments, Certificate of Excellence in Education, and a Digital Seal Award from the Institute of Women of Aviation Worldwide [IWOAW].

Sachin Kumar, an Assistant Professor in Computer Science & Engineering at Parul University, Vadodara. He has completed his B.Tech in Computer Science & Engineering from AKTU, Lucknow, India in 2011. He obtained his M.Tech degree in Computer Science from Baba Saheb Bhimrao Ambedkar University, Lucknow, India in 2016. He has completed his diploma in Statistics and computing in 2022 from Banaras Hindu University. He has done JRF and SRF from Banaras Hindu University. He has completed his doctoral studies from Veer Bahadur Singh Purvanchal University in 2024 in the area of Machine Learning, Deep Learning and Cloud Computing.

How to cite this paper: Surabhi Solanki, Seema Verma, Sachin Kumar, "A Survey of Techniques for Improving Information Retrieval through Query Expansion", International Journal of Information Technology and Computer Science(IJITCS), Vol.17, No.2, pp.73-81, 2025. DOI:10.5815/ijitcs.2025.02.07