The object of this study is an improved query expansion method based on context-dependent text sentiment analysis in information retrieval systems for working with databases. Using natural language processing (NLP) methods, in particular contextual embeddings and transformer architectures, this paper focuses on adaptively determining user intent within the submitted query. The study involves analyzing and improving text processing mechanisms using subject-specific filtering to increase the accuracy and relevance of search results. The proposed method demonstrates an increase in the accuracy of context-sensitive models by 6 % compared to baseline approaches. The aggregate F1-measure indicator, which combines precision, completeness, and accuracy, reflects the relevance of the constructed models, showing an increase of 6–8 %. The difference between the least and most effective methods is 16 % in accuracy and 17 % in relevance. The proposed approach overcomes the limitations of static traditional synonym and statistical methods by dynamically interpreting the relationship between tone, context, and domain specificity of content. Improved semantic understanding allows for more accurate matching of extended queries with user goals. This method could be effectively applied in practice in settings where information retrieval systems operate within domain-specific databases. This applies to scenarios in which user queries contain complex, emotionally colored language constructs that require a deeper understanding of context and tone. However, its implementation requires training on high-quality domain-specific datasets with contextual labels that provide accurate adaptation

Keywords: query expansion, natural language processing (NLP), information retrieval (IR), semantic analysis, database

UDC 004.675.2

DOI: 10.15587/1729-4061.2025.322120

QUERY EXPANSION BASED ON CONTEXT-DEPENDENT SENTIMENT ANALYSIS IN DATABASES WITH DOMAIN-SPECIFIC FILTERING

Vasyl Meliukh
Corresponding author
PhD Student*
E-mail: vasylmeliukh430@gmail.com

Kateryna Potapova
PhD, Associate Professor**

Mykola Nalyvaichuk
PhD, Senior Lecturer**

Andrii Dychka
PhD*

*Department of Computer Systems Software***
Department of System Programming and Specialized Computer Systems*
***National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
Beresteyskyi ave., 37, Kyiv, Ukraine, 03056

Received 07.11.2024
Received in revised form 03.01.2025
Accepted date 20.01.2025
Published date 27.02.2025

How to Cite: *Meliukh, V., Potapova, K., 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*

1. Introduction

The unprecedented growth of big data in the field of digital technologies indicates the need to improve the means to ensure accurate and efficient processing of the multifaceted information field. Over the past decades, the need for quick access to target data has stimulated the development of scientific research aimed at optimizing search algorithms and increasing the relevance of results. At the same time, the migration of communication processes to the online space, where data exchange is the basis of the system's functionality, emphasizes the need to transform the methods and mechanisms for processing the complex structure of hidden knowledge. The rapid expansion of the global computer network makes information retrieval an important component of the processes of management and interaction of people with the virtual world. This indicates the need to improve technologies for automatically extracting key information from text to make informed decisions in the changing linguistic landscape of the modern world.

Information Retrieval (IR) is one of the central areas of research in the field of Natural Language Processing (NLP). The NLP field is a discipline that deals with the interpretation of language and its components by means of digital analysis. In the context of query expansion (QE), IR processes, together with artificial intelligence (AI) tools, facilitate the extraction of relevant knowledge from large databases in response to user queries [1]. Search engines are specifically designed to provide an intuitive process of obtaining data by processing multimedia, audio, or text units. However, modern consumers of specialized information face difficulties in accessing accurate search representations inherent in specialized areas, such as economics or finance. Conventional methods of information retrieval (IR) are often unable to take into account the full range of intertextual relationships between words that go beyond simple grammatical constructs. The limitations of classical systems become obvious, especially if only a primitive match by keywords is taken as a basis. This does not make it possible to fully adapt the query to the specific intentions of the user. In particular, without taking into account the subtle tonal nuances of the text, it is difficult to understand irony, sarcasm, and other nuances of language [2, 3].

One of the mechanisms for improving the understanding of the context of queries and, accordingly, narrowing the range of potentially relevant results is to change their formulation and deepen the detail. Query expansion is a method that improves search efficiency by adding related terms or phrases to the initial query. The QE method makes it possible to expand the search area, increase the probability of finding relevant information, and also solve the problem of inaccuracy and incompleteness of search results. Text tone analysis, as a subtask of the NLP field, provides the ability to detect the emotional coloring of the query and classify data into appropriate categories [4]. Ignoring the tone of the query and semantic connections of the surrounding context are the main disadvantages of classical approaches based on synonymous or statistical associations [5].

Wide integration of databases with modern information and search systems increases the resource intensity and requirements for computing power necessary for the development of fast data processing technologies. This emphasizes the importance of the impact of identifying semantic relationships, such as synonymy, antonymy, hyponymy, and hypernymy, between different elements of the text. This approach opens up new prospects in the context of developing recommender systems that would find their application in scientific and professional practice. Inefficient data management could lead to an increase in information noise, which would negatively affect the quality of the results obtained and complicate the decision-making process. A high degree of automation with minimal human supervision provides for effective work in highly specialized industries. A deep understanding of the subject area, where specific concepts and knowledge structures require specialized expertise, is difficult to provide with the help of general intelligent systems. At the same time, the increasing complexity of user queries necessitates the need to take into account not only individual words (metadata) but also the use of a wider field of linguistic analysis.

Therefore, research that focuses on expanding queries based on context and sensitivity to text tone is relevant for improving the accuracy of information retrieval in specialized fields. Of particular note is the potential for practical application in fields with narrow specificity, where general approaches may be less useful. This applies both to reducing the number of search results unrelated to the user's initial intention and to improving the time and quality of decision-making. For example, integrating a query management system with context-oriented analysis tools could allow for differentiating the meaning of the word "help" depending on the situation, indicating a medical procedure, patient care, or emotional support. The increased accuracy of such searches could lead to more informed decisions, optimization of research processes, and operational efficiency in disciplines such as law, finance, science, business, and education. In addition, it would make it possible to individualize search results to the user's needs based on prior knowledge gained during the operation of the system. While maintaining a well-established approach, query expansion by combining related terms with the inclusion of natural language processing methods has the potential to outperform general-purpose systems. Thus, analyzing the relevance of search results to the user's initial intent in a dynamic digital field proves its expediency within the framework of context-sensitive query expansion.

2. Literature review and problem statement

In [6], the BertMasker model was proposed, which combined common and subject-specific features to solve the problem of multi-domain text tone classification. Using token masking to hide specialized or general-purpose structures allowed the model to be trained to distinguish between subject-invariant and subject-specific tone representations. In addition, performance improvements were observed in cross-domain environments when attention mechanisms were integrated, and cross-entropy losses were included. Thus, BertMasker outperformed previous models on reference datasets, including an average accuracy improvement of 1.88 % across all domains, including 6.75 % in the MR domain. However, this study encountered limitations in the form of sensitivity to the masking strategy, which led to little progress in certain domains and datasets. Also, the proposed model lacks an analysis of context-specific nuances, such as the dynamics of emotional coloring, which leaves room for further improvement.

In study [7], the possibility of implementing contextualized embeddings for query expansion is considered. For this purpose, a division of architectural units into unsupervised CEQE and supervised SQET models is introduced. The results obtained highlight the difference between the proposed and existing models in terms of their ability to improve query expansion methods using pseudo-relevant connection and context of candidate terms. Specifically, CEQE and SQET achieve improvements in tasks where recall accuracy plays an important role. Compared to static and probabilistic feedback models, CEQE outperforms them on standard test collections, showing an increase in completeness metrics such as Recall@1000. Furthermore, the supervised SQET model shows an advantage in the internal estimation of term rankings by 6 % but still remains identical to the CEQE model with respect to accuracy-based metrics.

A hybrid model proposed in [8], which combines term selection methods with semantic and genetic filtering, could be an option to overcome these difficulties. Automatic query expansion with a mechanism for reducing data redundancy and filtering irrelevant instances demonstrates a significant increase in performance compared to separate term selection methods. This approach ensures that only the most contextually relevant terms are included in the expanded query. However, despite the instrumental optimization of the selection of relevant extension terms, they rely too much on aggregation and filtering mechanisms. Similarly to the described method, [9] proposes the BERT-QE model, which uses the capabilities of contextualized embeddings. The proposed mechanism increases the efficiency of information retrieval by reranking documents and selecting relevant fragments from high-ranking documents for query expansion. However, the success of BERT-QE strongly depends on the availability of high-quality document fragments, which may be unavailable in domains with sparse or poorly structured data.

A similar problem was addressed in [10], which studied the impact of combining neural network-based word representations with learning methods for ranking. Such a query expansion framework aimed to improve search efficiency by using features derived from word representations and conventional context-based statistics. This fusion of methods showed superiority over conventional pseudo-relevant feedback models, achieving optimal results in search tasks on multiple TREC collections. As an example of a specialized application, one could consider work [11], in which a semantic sequential query expansion (SSDM) model was proposed to improve biomedical document retrieval. The model architecture was based on a combination of Skip-gram-based word embedding with a semantic dependence model (SDM) to encode semantic relationships between terms. However, all the problems related to contextual adaptability provided by modern transformer-based models remain unresolved. The focus on a subject-oriented thesaurus, although it allows for the capture of semantic dependences, does not take into account mechanisms for processing the tonal context within the received documents.

In [12], a two-stage query extension method ReInterpretQE is presented to address the problems of semantic ambiguity and keyword expression limitations. This approach combines query recommendation using a probabilistic algorithm with query interpretation based on the translation of keywords into query subgraphs. This demonstrates that ReInterpretQE improves both the accuracy and recall of query results compared to state-of-the-art methods such as Metadata and K-coupling. As one can see, the emphasis of the study was on optimizing the structural and semantic aspects of queries without extending to contextual nuances. Although static query subgraphs offer efficient implementation in relational databases, there is a limitation in the necessary flexibility for dynamic interpretation to tone-laden documents.

A likely option to overcome these difficulties is to increase or include the role of sentiment as part of feature selection, which is partially considered in study [13]. To improve the performance of text tone analysis, a new Query Expansion Ranking (QER) method is proposed. Nevertheless, tone analysis is considered a static problem, and not one that could adapt to queries in real time, without delving into the nuances of interpretation. Objective difficulties are associated with the excessive coverage of conventional machine learning classifiers, which may not have the depth of transformer architecture in understanding tone-laden contexts. Analysis of the issues in existing studies makes it possible to understand that static feature selection approaches are limited in dynamically adapting to queries or contexts. This affects the reduction of search relevance due to the insufficient subject-oriented focus of the submitted query to process specialized terminology. In particular, incomplete understanding of user intent could be observed in emotionally charged queries. All this gives reason to believe it appropriate to consider the problem of the inability to effectively integrate text tone analysis and dynamic contextual adaptation within the framework of query expansion. This is compounded by the lack of robust filtering mechanisms that would mitigate the variability in recall accuracy of information retrieval systems in specialized contexts.

To address these gaps, it is appropriate to conduct research on a query expansion method that includes context-sensitive text sentiment analysis based on domain filtering. The optimal solution would be to use advanced transformer-based architectures, such as BERT and RoBERTa, to dynamically adapt to the expanded query context. Emotional identification will help better match retrieved documents with user intent. A subject-oriented filtering strategy is associated with improved precision and recall in contexts with dense domain-specific terminology. The importance of considering the need for accurate and relevant information retrieval systems proves its value, especially in scenarios where the user's subjective perception significantly influences the evaluation of results.

3. The aim and objectives of the study

The aim of our study is to improve the method of query expansion by implementing context-sensitive text tone analysis. This will allow us to increase the accuracy and relevance of information retrieval systems for subject-oriented databases. Owing to the adaptive approach to decision-making in information retrieval processes, it is expected to provide personalization to user needs using thematic filtering.

To achieve the goal, the following tasks were set:

– to integrate the component of text tone analysis and detection into the query expansion process, which includes studying the influence of the emotional coloring of the text on the relevance of the results obtained;

– to implement a mechanism of thematic filtering of search results, which ensures effective ranking of documents based on domain-specific terminology.

4. The study materials and methods

The object of our study is the development of query expansion methods that use context-sensitive text tone analysis with thematic filtering. The main hypothesis assumes that the inclusion of context-sensitive text tone analysis could significantly increase relevance, which would be reflected in the accuracy of correct predictions of the machine learning model. Accordingly, the results obtained should translate the values of well-known metrics that demonstrate better search in domain-specific databases. This is explained by the fact that the subject-oriented context of queries affects the interpretation of the terms appearing there. Thus, text tone analysis could provide significant insights into the presence of urgency or seriousness in the user's intentions. NLP-based techniques could facilitate the capture and integration of domain-specific terms and semantic nuances in information search systems.

The main factor of the study is considered to be the measurement of sensitivity to the tonality of the text in user queries, which, in comparison with existing methods, could potentially provide more significant results with minimal training costs. Minimization of computing power is achieved by using deep neural networks based on a family of transformer models that have an extensive set of general knowledge. Refinement and additional training, in turn, makes it possible to adjust the model to specific tasks, while saving computing resources and time for collecting the necessary data is exponentially reduced. This is influenced by the effectiveness of modern large language models (LLMs), such as BERT or GPT, which are in no way inferior to models built from scratch.

We consider the identification of patterns of the influence of emotional coloring on the interpretation of user queries used during machine processing in information systems. To balance information in the domain during query expansion, it is proposed to implement the process of filtering results and compare the performance of context-oriented systems with and without taking into account tone analysis. However, it is worth remembering that the volume of the preliminary data set used for training and evaluating the effectiveness of the developed model is limited. This means that our study approximates the sequence of semantic and syntactic structures in the data sets used for query expansion based on text tone. Also, only three main sentiment states (positive, negative, and neutral) are taken into account to simplify modeling and avoid potential variability when reproducing testing conditions.

This work proposes a methodology for combining natural language processing (NLP) methods, machine learning models, and information retrieval (IR) algorithms to build an effective query expansion structure for practical use in real systems. Correct interpretation of context, recognition of sentiment and consideration of domain-specific nuances require the search for flexible program configurations to learn the best structural solutions. To this end, the design phase includes advanced input processing procedures aimed at analyzing how much the new structure improves the search relevance efficiency according to established metrics and comparing them with existing models. The suitability of the proposed design is determined by the fact that the processing of user queries involves the integration of several linguistic phenomena to solve the complexity of the problem.

As shown in Fig. 1, the organization of the processing pipeline for the model architecture consists of three main stages: "Data Preparation", "Preprocessing", "Model Formation and Query Expansion".

The success of context-aware tonal analysis-based query augmentation depends largely on the availability and quality of data that is robust, representative, and suitable for the task at hand. Our study fills this gap by combining domain-specific datasets with tonality benchmarks. The integration of these sources ensures the model's ability to generalize across different usage cases of language items while maintaining accuracy in domain-specific scenarios.

The datasets were collected from specialized databases that were publicly available in the Kaggle repositories, characterized by the following datasets:

– financial document records: derived from the Reuters-21578 dataset, which contains articles, reports, news, and analytics;

– medical document records: derived from the MIMIC-III dataset, which contains clinical records, patient diagnoses, and hospital discharges;

– cybersecurity document records: obtained from the "Hacker News Corpus" dataset, which contains technical blogs, cybersecurity incident reports, and threat analysis documents.

Effective query expansion requires careful processing to standardize, eliminate redundant information, and ensure proper input quality for machine learning models. This is a fundamental step in transforming raw text into structured formats that enable advanced natural language processing (NLP) techniques. The first step in processing is to clean the text of uninformative elements such as tags, URLs, or special characters. This is followed by text normalization to bring all words into a single representation. The output is divided into individual units, represented by words or phrases, while preserving their stem or root form (e.g., "walking"→"walk"). The lemmatization process is prioritized for domain-specific datasets to preserve contextual accuracy. Potential data sparsity issues are addressed by vocabulary reduction, where uninformative and high-frequency words (e.g., "and," "the") are ignored.

Tone annotation is divided into three main categories, including positive, negative, and neutral text moods. Initial labeling of text tokens is performed using pre-trained mood classifiers. The process is characterized by fine-tuning of ready-made models using vector representations of subject datasets, facilitating the capture of semantic nuances of the text. For greater control over the accuracy and relevance of datasets, the possibility of manual annotation and expansion of tones according to thematic-specific contexts, such as urgency or confidence, is envisaged.

Words or phrases with explicit emotional content were identified using existing sentiment lexicons and augmented with a custom vocabulary. The methodologies used for feature extraction were based on NLP techniques and augmented to meet the unique tone requirements of the text in a given domain. Tools such as SentiWordNet, VADER were selected based on their proven reliability in tone detection. Domain-specific lexicons and ontologies provided the necessary adaptability for specialized contexts. Numerical representations of the intensity of the sentiment (positive, negative, or neutral) associated with terms or sentences were calculated using a combination of rule-based and machine learning methods. Polarity estimation included lexicon-based tools (VADER) and transformer-based models (BERT with tone fine-tuning). Contextual adjustments were applied to refine ratings based on modifiers such as negation and reinforcement (e. g., "bad," "very promising").

To align feature extraction across specialized domains, a two-step process was implemented, involving data corpus analysis and matching against existing ontologies. Domain-specific corpora were analyzed to identify unique tonal markers not found in general-purpose sentiment lexicons. For example, the inclusion of features such as "critically low" or "positive prognosis" correlates with phrases specific to the healthcare domain. Domain ontologies were used to establish semantic hierarchies and relationships between terms, improving the understanding of contextual tone. For example, in legal texts, the term "liability" could have a negative connotation depending on the context. The extracted features were structured into vectorized representations suitable for machine learning models.

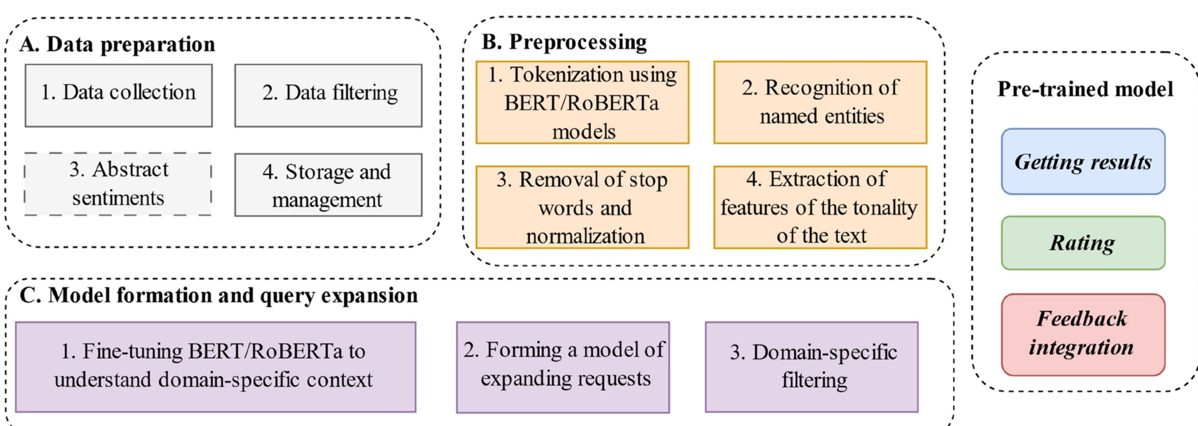

Fig. 1. Query expansion pipeline architecture with domain-specific data processing and sentiment analysis

A key component of our study is the selection of appropriate pre-trained sentiment analysis models for contextual tone detection and analysis. The models were evaluated based on their architecture, performance, and suitability for specialized tasks. The BERT model, which is based on a transformer architecture, is known for its bidirectional context understanding, which significantly affects the capture of complex language relationships. An improved variation of the model, called RoBERTa, was optimized by training on larger datasets and longer sequences, making it more suitable for complex, subject-specific datasets. The selection process was based on assessing their ability to fine-tune with adaptation to subject-specific data. RoBERTa demonstrated exceptional flexibility due to robust training optimization, making it ideal for such specialized industries.

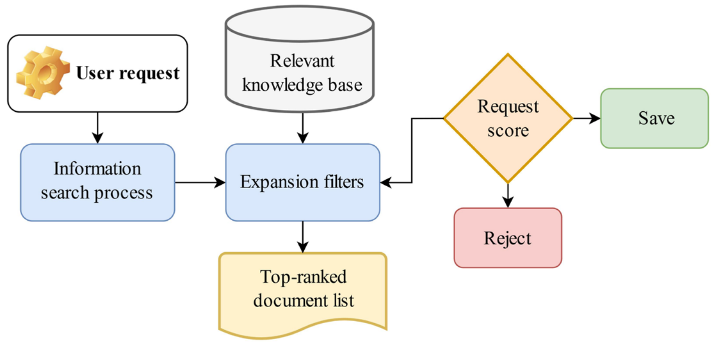

Fig. 2. Query expansion workflow using fine-tuned BERT/RoBERTa with context-sensitive sentiment filtering

The model training stage begins with fine-tuning BERT/RoBERTa on already trained data to adapt the model to specific tasks within the chosen domain. After generating a list of extension terms that match the context and sentiment of the original query, the extracted sentiment scores retain only those terms that match the underlying intent of the query. After receiving the results from the tuned model, the extended query is ranked based on the relevance score of the match between the retrieved document and the extended query terms (Fig. 2). The learning rate was dynamically adjusted during fine-tuning to prevent overfitting while maximizing the model's ability to generalize from unseen data. An additional dropout layer and early stopping mechanisms were used to ensure that the model maintained robustness without memorization rather than adapting to the training data. The model output is evaluated using accuracy, completeness, and other relevance metrics. If the situation requires it, user feedback is collected, and model parameters are iteratively adjusted to fine-tune the model to improve future extensions.

A summary of the main algorithm is shown in Fig. 3 below.

The main goal of text tone detection in the query expansion method is to supplement the terms embedded in the user's query that would correlate with the overall thematic direction. Dynamic term scoring is characterized by a scoring mechanism to rank potential query expansion terms based on their semantic relevance and agreement with the detected tone. Tone alignment was quantified using the weighted cosine similarity between the query tone vector and the candidate terms in the vector space. To measure the similarity between the original and expanded queries, the direction alignment between the two vectors is evaluated by checking how far the expanded query is contextually distant from the original, as shown below:

$$Cosine\ similarity = \frac{A * B}{|A||B|}, \tag{1}$$

where A is the vector representation of the original query;

B is the vector representation of the expanded query.

```
# Function to process and expand query based on sentiment
def expand_query(query, domain_data):
    # Preprocess and tokenize user query
    query_tokens = tokenizer.tokenize(query)
    # Extract initial candidate terms for expansion from domain data
    candidate_terms = Counter()
    for doc in domain_data:
        doc_tokens = tokenizer.tokenize(doc['text'])
        doc_sentiment = doc['sentiment']  # Precomputed sentiment score for domain data document
        # Include terms matching query context and sentiment
        if doc_sentiment == get_sentiment(query):
            candidate_terms.update(doc_tokens)

    # Rank candidate terms based on relevance to the query context
    # Calculate cosine similarity between query and candidate terms
    query_vector = np.mean([bert_model(**tokenizer(query, return_tensors='pt'))[0].detach().numpy()], axis=0)
    candidate_scores = {}
    for term, freq in candidate_terms.items():
        term_vector = np.mean([bert_model(**tokenizer(term, return_tensors='pt'))[0].detach().numpy()], axis=0)
        similarity = cosine_similarity(query_vector, term_vector)
        candidate_scores[term] = similarity * freq

    # Select top-n terms for query expansion
    top_terms = sorted(candidate_scores, key=candidate_scores.get, reverse=True)[:5]  # Top 5 terms
    # Expand query
    expanded_query = query + " " + " ".join(top_terms)
    # Retrieve relevant documents using the expanded query
    expanded_results = retrieve_relevant_documents(expanded_query, domain_data)

    return expanded_results
```

Fig. 3. Implementing the query expansion algorithm

For example, a query with a cautionary tone preferred terms with negative sentiment over neutral or positive terms. Balancing semantic relevance and tone consistency requires a dual-criteria framework. First, it is necessary to assign higher weight to terms that exactly match the detected tone. Second, it is necessary to include a penalty for semantically relevant terms that are inconsistent in sentiment representation with the original query tone. An attentional mechanism inspired by neural network architecture was used to highlight terms with the highest mood match and semantic relevance during the expansion process.

The increasing complexity of information retrieval systems has led to a growing demand for domain-specific filtering mechanisms that align query results with user intent. While conventional filtering approaches rely heavily on keyword and metadata matching, they often fail to take into account the tonal or contextual nuances embedded in user queries. Sentiment analysis models were used to classify user queries into tone categories. These classifications directly informed the filtering logic, ensuring a match between the emotional intent of the query and the sentiment of the results.

A pre-trained sentiment analysis model (e.g., BERT or RoBERTa) was used to detect the tone of each query. Filtering thresholds were dynamically adjusted based on the tone classification, taking into account both emotional content and semantic relevance to optimize the selection process. The filtering algorithm combined sentiment scoring with content relevance metrics such as term frequency, inverse document frequency (TF-IDF), and cosine similarity. The results were ranked using a weighted scoring system that prioritized tone alignment without compromising semantic coherence. A hybrid weighting mechanism was implemented to balance tone relevance and domain specificity. This ensured that the top results were both emotionally and contextually aligned with the users' needs. The equation representing the implementation of the dynamic weighting strategy is as follows:

$$R = \alpha * W_t + \beta * W_d, \tag{2}$$

where W_t is a weighted tone matching score;

W_d is a weighted domain relevance score;

α is a parameter reflecting the relative importance of tone;

β is a parameter reflecting the relative relevance of the domain.

The values of α and β were optimized through cross-validation, maximizing performance metrics such as mean reciprocal rank (MRR) and normalized discounted cumulative gain (nDCG). To continuously improve the ranking mechanism, it is possible to extend the functionality to include an iterative feedback loop. Adjusting the weight parameters and improving the consistency over time could be implemented during the user interaction with the system (e.g., clicks, time spent on results).

To improve the contextual matching of query terms with database content, domain-specific ontologies were developed. These ontologies served as structured knowledge representations that encompassed entities, relationships, and terminology specific to the target domain. Terms retrieved from user queries were mapped to ontology nodes to identify semantically related concepts. Ontologies were used to expand queries with synonyms, broader terms, or related concepts that matched the user's intent. Tone-sensitive mappings ensured that emotionally aligned terms were prioritized during query expansion and filtering.

The metrics chosen to evaluate the effectiveness of information retrieval in terms of the quality of expanded queries from the database are based on the following evaluation equations:

1. *Precision* ensures that only documents relevant to the original query are returned from the database, limiting the level of misinformation. Equation (3) measures the proportion of retrieved documents that are relevant, as follows:

$$Precision = \frac{TP}{TP + FP}, \tag{3}$$

where *TP* (True Positives) – correctly extracted relevant documents;

FP (False Positives) – incorrectly extracted irrelevant documents.

2. *Recall* measures the efficiency of identifying all relevant documents in the database, ensuring that the expansion process captures a wide set of documents. Equation (4) measures the proportion of documents that were extracted from the total number of relevant documents available, as follows:

$$Recall = \frac{TP}{TP + FN}, \tag{4}$$

where *FN* (FalseNegatives) are relevant documents that were not extracted.

3. *Accuracy.* Although less valuable, it is still useful for overall assessment of the model's performance in terms of the balance between extracting relevant documents and avoiding irrelevant ones. Equation (5) measures the proportion of correctly extracted documents, including relevant and irrelevant ones, out of the total number of extracted documents:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}, \tag{5}$$

where *FN* (False Negatives) are relevant documents that were not retrieved.

4. *F*1-score measures the impact of both irrelevant documents (false positives) and missing relevant documents (false negatives) on the trade-off between accuracy and completeness. Equation (6) measures the harmonic mean of accuracy and completeness as follows:

$$F1 = 2 * \frac{precision * recall}{precision + recall}. \tag{6}$$

5. Mean Average Precision (*MAP*) evaluates the quality of the document ranking produced by the extended query. Equation (7) measures the average accuracy across multiple queries by taking the average of mean precision for each individual query:

$$MAP = \frac{1}{|D_r|} \sum_{k=1}^{N} precision_k * RDI, \tag{7}$$

where D_r is the total number of relevant documents in the database for a given query;

RDI (Relevant Document Indicator) is an indicator equal to 1 if the document at position k is relevant and 0 if it is not relevant.

Google Colaboratory served as the main computing platform, providing an interactive environment based on Jupyter Notebook with access to cloud resources. The CPU processor was used for lightweight models such as TF-IDF and Word-2Vec, which required less computational effort. The T4 GPU was used to train and infer transformer-based models (such as

BERT and RoBERTa), which allowed for efficient processing of large data sets and significantly reduced model training time. The Python programming language was chosen due to its extensive ecosystem of machine learning, data analysis, and natural language processing libraries.

Matplotlib and Seaborn libraries provided clear and informative visual representations of performance indicators and data distributions, facilitating interpretation and representation of results. Conceptual diagrams were modeled using the free online software draw.io.

Pandas and NumPy were used to organize and manipulate structured data and support numerical computations. Pre-trained transformer models such as BERT and RoBERTa were accessed through the Huggins Face Transformers API. This allowed us to implement and fine-tune the models for specialized tasks in the study using the PyTorch framework. The NLTK (Natural Language Toolkit) library was used for basic pre-processing tasks, including tokenization, stopword removal, stemming, and lemmatization. spaCy was used as an additional library for efficient text processing and syntactic analysis, which helps with complex NLP tasks. Static modeling and calculation of evaluation metrics including precision, completeness, accuracy, and F1-score relied on the Scikit-learn library.

```
Document ID: test/14826
Query: ASIAN EXPORTERS FEAR DAMAGE FROM U.S.-JAPAN RIFT
Text:   Mounting trade friction between the
  U.S. And Japan has raised fears among many of Asia's exporting
  nations that the row could inflict far-reachin...
--------------------------------------------------------------------------------
Document ID: test/14828
Query: CHINA DAILY SAYS VERMIN EAT 7-12 PCT GRAIN STOCKS
Text:   A survey of 19 provinces and seven cities
  showed vermin consume between seven and 12 pct of China's grain
  stocks, the China Daily said.
      It...
--------------------------------------------------------------------------------
Document ID: test/14829
Query: JAPAN TO REVISE LONG-TERM ENERGY DEMAND DOWNWARDS
Text:   The Ministry of International Trade and
  Industry (MITI) will revise its long-term energy supply/demand
  outlook by August to meet a forecast down...
--------------------------------------------------------------------------------
Document ID: test/14832
Query: THAI TRADE DEFICIT WIDENS IN FIRST QUARTER
Text:   Thailand's trade deficit widened to 4.5
  billion baht in the first quarter of 1987 from 2.1 billion a
  year ago, the Business Economics Department said.
...
--------------------------------------------------------------------------------
Document ID: test/14833
Query: INDONESIA SEES CPO PRICE RISING SHARPLY
Text:   Indonesia expects crude palm oil (CPO)
  prices to rise sharply to between 450 and 550 dlrs a tonne FOB
  sometime this year because of better European demand...
--------------------------------------------------------------------------------
```

Fig. 4. Document record structure for tone analysis

5. Results of investigating the process of subject-oriented contextual query expansion integrated with text sentiment analysis

5. 1. Integration of text sentiment analysis into the contextual query expansion model

Determining the text sentiment required to generate alternative phrases and additions to the initial query requires processing data sets that contain relevant information for each document. Important for identifying contextual relationships is the availability of headings to the text content for further analysis. As shown in Fig. 4, each scenario is divided into training and test data sets.

Based on the extracted structured document data, the document's emotional coloration is assessed. For this purpose, a pre-trained CardiffNLP Twitter RoBERTa sentiment analysis model is used, which classifies the tone into positive, neutral, and negative classes. This step provides text data that is the basis for both query expansion and context-sensitive tone analysis. Each document is assigned all three sentiment labels using a softmax function to calculate probabilities. The dominance score for each entry identifies the emotional bias inherent in the text sequences, as shown in Fig. 5.

For a more weighted representation of tonal shifts, contextual sentiment representations are combined with rule-based instruments dictated by stable ontologies and dictionaries (SentiWordNet, VADER). This allows us to analyze each word separately and calculate a score by hybrid tone normalization of both each entry as a whole and taking into account the gradient distribution of each individual unit (Fig. 6).

To integrate text sentiment analysis into the contextual query expansion model, specialized datasets (from finance, medicine, cybersecurity) were used, which allowed us to adapt the models to specific topics. Based on the resulting data, a comparative analysis of models integrated into the tonal context of subject analysis for query expansion was conducted. Conventional TF-IDF expansion uses a basic statistical approach to query expansion based on term frequency. Queries are enriched with words with high TF-IDF that correspond to the document domain. Word2Vec-based expansion uses a pre-trained Word2Vec model to select terms similar in meaning to those in the query. On the other hand, the use of standard language models (BERT or RoBERTa) generates the closest results to the submitted query without taking into account the text sentiment context. The improvement consists in selecting words for query expansion depending on the document sentiment context (for example, positive, negative, neutral) and adjusting them to the query context.

Table 1 demonstrates the context-aware models outperform the baseline models by 6–8 % on each evaluation metric. The context-aware model integrating RoBERTa's transformer architecture and sentiment analysis shows the best results to date. This demonstrates superior contextual understanding in terms of query expansion with an accuracy of 86 % and an F1-score of 84 %, likely due to RoBERTa's extensive pre-training on large datasets.

Conventional query expansion methods, including orthodox TF-IDF algorithms or even more advanced contextual embeddings like Word2Vec, are significantly worse than their context-aware counterparts. Applying advanced knowledge incorporation to already established concepts provides a significant increase in the return of relevant answers, such that the difference in F1-score between the least effective and most powerful models reaches 17 %.

Fig. 7 demonstrates that the conventional expansion based on TF-IDF or Word2Vec is inferior to the more advanced contextual word representation models BERT and RoBERTa.

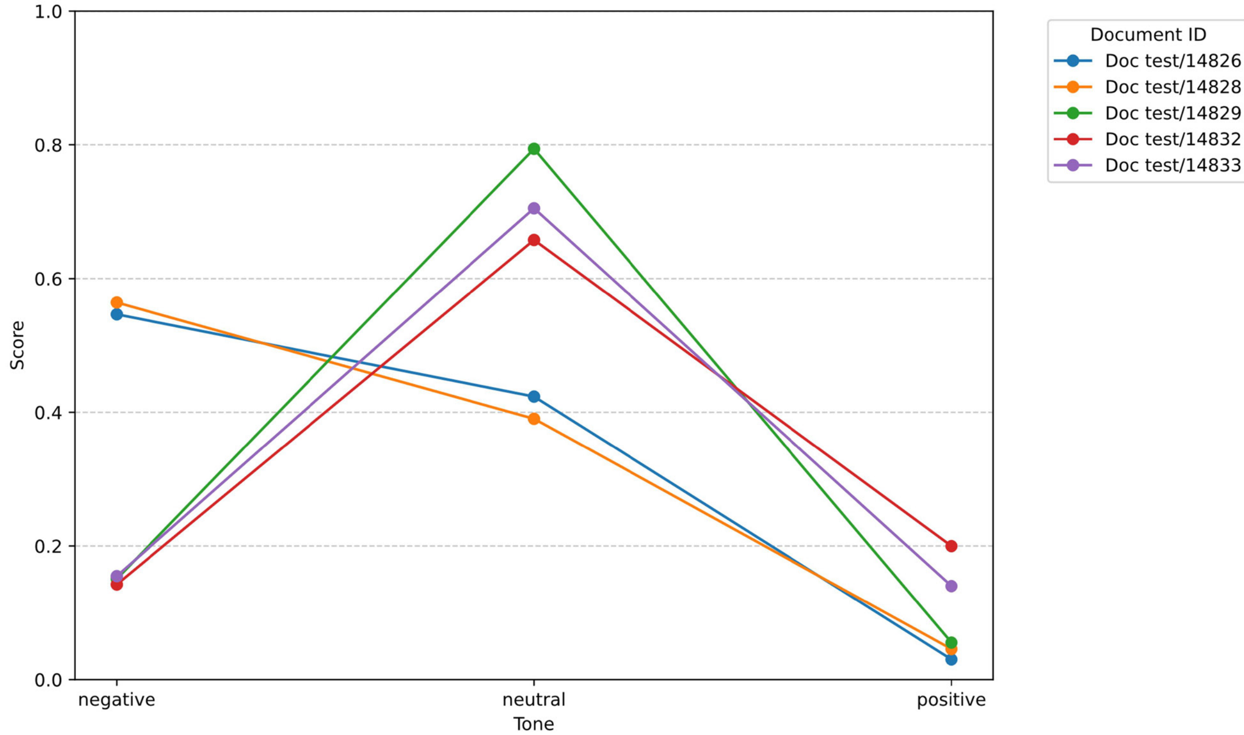

Fig. 5. Probability distribution of the tonal representation of dataset records

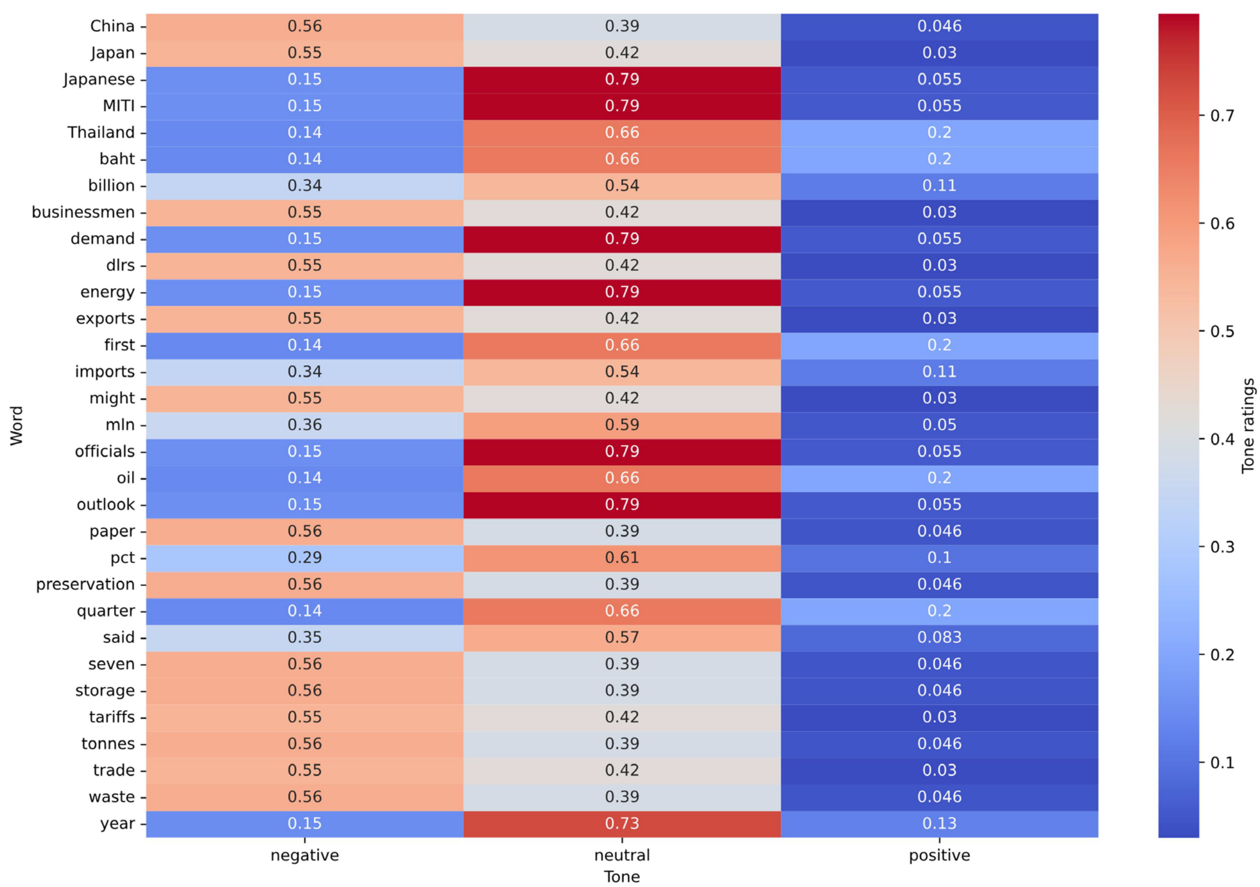

Fig. 6. Map of the distribution of tone ratings within a single text representation of a document

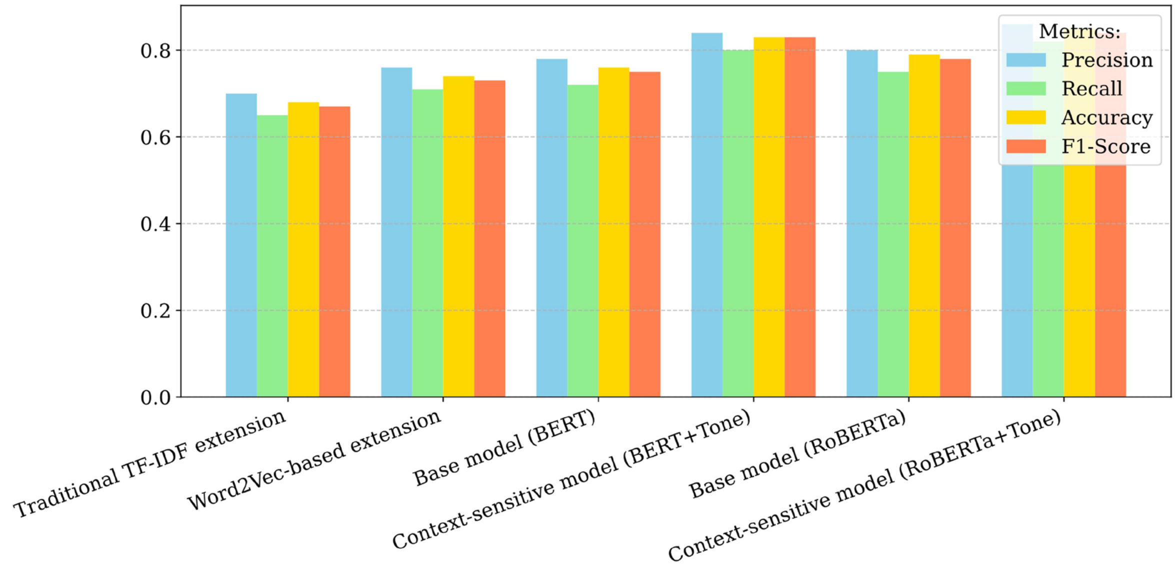

Fig. 7. Comparison of performance metrics of query expansion performance evaluation models

Table 1

Performance assessment across all domains

Model Type	Precision	Recall	Accuracy	F1-score
Traditional TF-IDF Extension	0.70	0.65	0.68	0.67
Word2Vec-Based Extension	0.76	0.71	0.74	0.73
Baseline Model (BERT)	0.78	0.72	0.76	0.75
Context-Aware Model (BERT+Tone)	0.84	0.80	0.83	0.83
Baseline Model (RoBERTa)	0.80	0.75	0.79	0.78
Context-Aware Model (RoBERTa+Tone)	0.86	0.82	0.85	0.84

5. 2. Implementation of the subject-oriented filterf ing mechanism

The results of the implementation of subject-oriented filtering mechanisms are focused on improving the relevance of search results in context-oriented systems by using subject terminology. The performance indicators – accuracy, recall, precision and F1 score – were evaluated in three areas: finance, healthcare, and cybersecurity (Fig. 8).

The performance assessment was performed separately for each domain-specific dataset, covering the disciplines of finance, healthcare, and cybersecurity (Fig. 8). Our results show that the model tuned to expand the cybersecurity queries achieved the highest performance on each metric. Most likely, the superior performance is associated with more clearly defined sentiment polarities, which helps classify positive or negative threats to the user. The healthcare domain, on the other hand, showed slightly lower results with a 3.4 % difference in F1 score due to the inherent complexity of labeling medical records when extracting applicable records for the user recommendation system. Overall, the recorded improvement in accuracy and F1 score reached 5–7 % across all domains compared to the baseline. This indicates that the model captured more relevant results, characterized by a completeness metric in the process of maintaining high accuracy in sentiment-based expansion.

Considering not only the query domain, but also its length, it becomes obvious that the performance of longer queries with richer semantic content will be more advantageous than dry factual statements (Table 2).

More complex queries are particularly practical for context-aware extension in terms of exploring domains filled with subjective documents, such as finance or cybersecurity. However, while precision and completeness decreased in queries with 1–3 tokens, accuracy remained unchanged due to the model's ability to handle specific contexts quite effectively.

Table 2

Impact of query length and complexity

Query length	Precision	Recall	Accuracy	F1 score
Short queries (1–3 tokens)	0.81	0.76	0.79	0.78
Medium queries (4–7 tokens)	0.85	0.81	0.84	0.83
Long queries (8+ tokens)	0.88	0.84	0.86	0.86

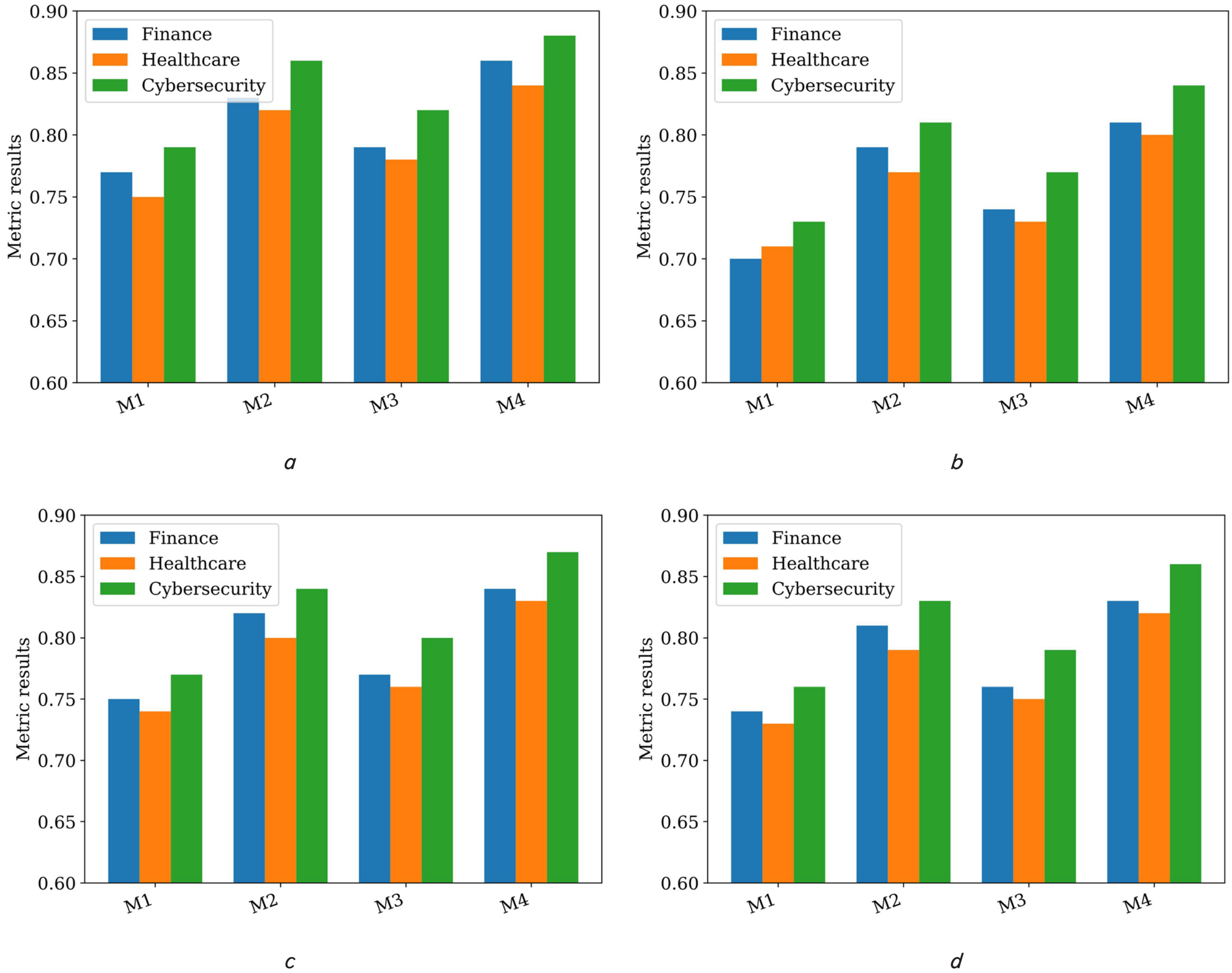

Fig. 8. Comparative analysis of performance evaluation in the financial, medical, and cybersecurity industries using models where M1 is the base model (BERT), M2 is the context-sensitive model (BERT+tonality), M3 is the base model (RoBERTa), M4 is the context-sensitive model (RoBERTa+tonality) for the metrics: *a* — precision; *b* — recall; *c* — accuracy; *d* — F1-Score

6. Discussion of results related to the process of subject-oriented contextual query expansion integrated with text tone analysis

Our study demonstrates the significant advantages of context-sensitive text tone analysis for increasing the relevance of document extraction. Focusing on specialized domains in which there may be variability in tone and emotional coloring proves the influence of user intent on database search mechanisms. Compared with conventional methods, the proposed model demonstrates the ability to match query semantics with tonal variability, increasing the accuracy factor in subjective queries. This addresses the limitations noted in previous works, such as the inability to distinguish between general text and local features, as highlighted in [6].

Table 1 illustrates the scoring format for domain-specific queries, where the emotional content of the text is integrated into the expansion of the original queries across domains. The table provides examples from the finance, healthcare, and cybersecurity domains, where the original queries are analyzed for their tone (e.g., negative, neutral) and augmented with additional relevant keywords. These expanded keywords reflect the nuances of user intent in each domain, which could be influenced by tone and context. For example, in finance, the query "Stock market crash" is associated with a negative connotation and is expanded with terms such as "economic recession" and "financial crisis." Similarly, in healthcare, the neutral query "COVID-19 vaccine approval" is expanded with terms such as "vaccine approval," "vaccine trial," and "FDA approval." In cybersecurity, the negative query "Data Breach Notification" has been expanded with keywords such as "Cyberattack Alert," "Security Breach Disclosure," and "Privacy Breach Alert." These expansions highlight how tone analysis helps enrich query terms, improving the search engine's ability to retrieve more relevant documents based on the user's specific context and intent.

As shown in Table 2, the improved method outperforms the baseline models in all metrics. For example, the F1-score in the sentiment BERT model reached 0.83, which outperformed the baseline BERT with a value of 0.75 by 8 %. RoBERTa demonstrates a similar trend of improving performance by 17 % compared to conventional extension methods. These improvements reflect the enhanced ability of the proposed approach to take into account context dependence, which leads to more accurate query expansion and document retrieval. Comparing these results with those of other studies, the improvements proposed in our study become even more obvious. For example, in a comparative study [6], the BertMasker model is presented, which improves domain invariance by masking irrelevant tokens, which leads to a moderate increase in accuracy by 0.94 % in multi-domain classification. Furthermore, the masking strategy in BertMasker reduces domain discrimination by more than 12 %, which is a less non ticeable trade-off due to its direct inclusion of sentiment analysis. Domain-specific comparisons further highlight these differences. In the financial domain, BertMasker's domain-aware representation strategy improves overall sentiment classification, but does not clearly improve search accuracy for sentiment-driven queries, as observed in the improved method.

It is worth comparing the approach to query expansion using context-sensitive sentiment analysis with the hybrid model presented in [7]. This model uses rank aggregation and semantic filtering combined with genetic algorithms to optimize query expansion terms. While it improves accuracy and completeness by filtering out irrelevant terms, it mainly focuses on static relevance feedback, without taking into account tonal variations in the text. In contrast, the improved method incorporates sentiment markers to dynamically adjust query terms based on contextual sentiment, achieving better performance, particularly in domains such as finance and cybersecurity.

Fig. 5 demonstrates that the financial domain highlights the effectiveness of the context-sensitive model in processing text with contrasting polarity, with an accuracy increase from 0.77 in the basic BERT to 0.84 in the context-sensitive RoBERTa. This outperforms the hybrid approach that relies on static semantic filtering. Similarly, cybersecurity presents another significant benchmark, with an accuracy increase from 0.79 in the basic BERT to 0.88 in the context-sensitive RoBERTa, exceeding a 7 % gain. This shows the ability to handle tonal nuances and domain-specific language better than the semantic filtering used in the aforementioned work.

Despite these achievements, overextension remains a common problem, as excessive query terms could degrade accuracy. Future integration of dynamic precomputing methods, as proposed in BERT-QE [8], could optimize the performance of our model while maintaining its adaptability to tonal variability. Combining sentiment-aware mechanisms with precomputing could further improve scalability and performance in real-time search systems.

The integration of sentiment analysis into query expansion remains underexplored, with most previous work treating sentiment as an auxiliary feature rather than a core component. As could be seen from Table 2, where a 5–8 % increase in F1 score is observed, embedding sentiment and linguistic features together increases the semantic coherence of expanded queries, a progressive approach not observed in conventional IR systems. This distinction positions the presented query expansion framework as a more adaptive and domain-aware search tool compared to existing models. However, our study has several limitations that should be taken into account.

It is worth noting that the improved method has minor advantages in highly technical domains, where the use of neutral language prevails. In contrast, domains that are filled with linguistic units with subjective emotional tone perform better under these conditions. That is why, when tuning information systems for specialized domains, it is worth considering the limited universality of the proposed method. Thus, the CEQE study [9] emphasizes stability in different datasets, such as TREC, while the contextual-tone method addresses the problems of mood variability in real data, emphasizing adaptability compared to static approaches.

Unlike CEQE, which mainly uses context-dependent embeddings for pseudo-relevant feedback, our improved method directly integrates tone analysis. This adds unique improvements in relevance for subjective or mood-driven queries, especially in domains such as finance and healthcare. CEQE achieves 18–31 % improvement in average accuracy on some datasets, emphasizing completeness. The study demonstrates 8–17 % improvement in F1-score, emphasizing accuracy improvement in tone-sensitive domains. SQET uses supervised learning to classify extension terms, achieving 6 % improvement in term ranking accuracy. On the other hand, the considered context-aware method with tone integration adapts this by embedding mood markers in the context-sensitive architecture, improving adaptability in subjective domains.

A potential drawback of the method is the overexpansion of queries, which could underestimate accuracy by extracting irrelevant documents. One way to mitigate the demonstrated problem was to design a mechanism for dynamic calibration of the obtained results. Thus, documents that did not fully correspond to the subject matter were not taken into account in the final assessment of the model's performance. However, the limited granularity of tone, which leads to simplified interpretations, could be represented through more detailed emotional scales or by increasing the dimensionality of the studied models.

Future research should build on these findings, addressing such limitations as the dynamic evolution of language and the inclusion of multimodal data. In addition, hybrid models that are able to balance technical accuracy and emotional nuance could expand applicability to broader domains, supporting a personalized search experience. Long-term stability testing using constantly updated data is expected to maintain the high performance of the models.

Although the dynamic evolution of language patterns to user requirements has the potential to improve the robustness of the model in real-world applications. Complex models require a detailed approach to optimizing training methods and hyperparameters, which requires high computational costs. Despite this, the development of hybrid architectures that could handle multi-subject data is important to enhance universal applicability. However, effective evaluation metrics for hybrid systems remain understudied, which complicates the validation of the developed product. Adapting practical systems in real time will require scalable and resource-efficient architectures. In addition, the ability of the method to generalize information to perform the task depends on the completeness and compatibility of the provided data.

7. Conclusions

1. Adding a sentiment analysis component to the basic query expansion framework introduced an additional variable during the process of calculating the efficiency of the presented models. This allowed us to increase the relevance by 6–8 % in those areas where tasks based on the extraction of opinions or analysis of customer feedback were the majority in the specificity of query interpretation. As a result, the nesting of advanced transformer model architectures improves the accuracy of queries and the relevance of extracted results in tonal-contrast scenarios. The improved method demonstrates the best relevance results when using the Ro-

BERTa-based architecture, although the increase in accuracy and F1-score increases only by 1–2 %.

2. The filtering mechanism has demonstrated a stable increase in performance in all studied areas, adjusting to domain-specific terminology. The improvement is characterized by an improvement in accuracy and completeness, which in turn increases the overall performance of the system. Even in the context of excessive term saturation in the medical field, transformative models with deep contextual knowledge could mitigate these limitations due to the strong contrast between general and professional content. At the same time, the length of the query affects the increase in recognition efficiency, where larger average and long queries make it possible to increase the F1 metric value by 3 % and 8 %, respectively.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Funding

The study was conducted without financial support.

Data availability

The data will be provided upon reasonable request.

Use of artificial intelligence

The authors used artificial intelligence technologies within acceptable limits to provide their own verified data, which is described in the research methodology section.

References

1. Cai, F., de Rijke, M. (2016). A Survey of Query Auto Completion in Information Retrieval. Foundations and Trends® in Information Retrieval, 10 (4), 273–363. https://doi.org/10.1561/1500000055
2. Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation, 28 (1), 11–21. https://doi.org/10.1108/eb026526
3. Manning, C. D., Raghavan, P., Schütze, H. (2008). Introduction to Information Retrieval. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511809071
4. Hu, Z., Dychka, I., Potapova, K., Meliukh, V. (2024). Augmenting Sentiment Analysis Prediction in Binary Text Classification through Advanced Natural Language Processing Models and Classifiers. International Journal of Information Technology and Computer Science, 16 (2), 16–31. https://doi.org/10.5815/ijitcs.2024.02.02
5. Mikolov, T., Chen, K., Corrado, G. Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv. https://doi.org/10.48550/arXiv.1301.3781
6. Yuan, J., Zhao, Y., Qin, B. (2022). Learning to share by masking the non-shared for multi-domain sentiment classification. International Journal of Machine Learning and Cybernetics, 13 (9), 2711–2724. https://doi.org/10.1007/s13042-022-01556-0
7. Naseri, S., Dalton, J., Yates, A., Allan, J. (2022). CEQE to SQET: A study of contextualized embeddings for query expansion. Information Retrieval Journal, 25 (2), 184–208. https://doi.org/10.1007/s10791-022-09405-y
8. Singh, J. (2017). Ranks Aggregation and Semantic Genetic Approach based Hybrid Model for Query Expansion. International Journal of Computational Intelligence Systems, 10 (1), 34. https://doi.org/10.2991/ijcis.2017.10.1.4
9. Zheng, Z., Hui, K., He, B., Han, X., Sun, L., Yates, A. (2020). BERT-QE: Contextualized Query Expansion for Document Re-ranking. Findings of the Association for Computational Linguistics: EMNLP 2020. https://doi.org/10.18653/v1/2020.findings-emnlp.424
10. Xu, B., Lin, H., Lin, Y., Yang, L., Xu, K. (2018). Improving Pseudo-Relevance Feedback With Neural Network-Based Word Representations. IEEE Access, 6, 62152–62165. https://doi.org/10.1109/access.2018.2876425
11. Fang, F., Zhang, B.-W., Yin, X.-C. (2018). Semantic Sequential Query Expansion for Biomedical Article Search. IEEE Access, 6, 45448–45457. https://doi.org/10.1109/access.2018.2861869
12. Wang, Y., Wang, N., Zhou, L. (2017). Keyword Query Expansion Paradigm Based on Recommendation and Interpretation in Relational Databases. Scientific Programming, 2017, 1–12. https://doi.org/10.1155/2017/7613026
13. Parlar, T., Özel, S. A., Song, F. (2018). QER: a new feature selection method for sentiment analysis. Human-Centric Computing and Information Sciences, 8 (1). https://doi.org/10.1186/s13673-018-0135-8