Науковий вісник НЛТУ України
Scientific Bulletin of UNFU
ISSN 1994-7836 (print)
ISSN 2519-2477 (online)
https://doi.org/10.36930/40360304
Article received: 29.03.2026
Article accepted: 27.05.2026
Article published: 29.06.2026
Correspondence author: A. I. Pliusnin artem.pliusnin.pz.2022@lpnu.ua

UDC 004.[02+04+89]

***A. I. Pliusnin, R. M. Malyi***
*Lviv Polytechnic National University, Lviv, Ukraine*

# IMPROVEMENT OF A SEMANTIC MOVIE RETRIEVAL METHOD BASED ON WEIGHTED AGGREGATION OF MULTI-SOURCE EMBEDDINGS

The features of improving semantic movie retrieval were investigated by combining heterogeneous textual sources that describe different aspects of film content. A weighted aggregation approach was analyzed using dense vector representations generated from metadata, user reviews, and English subtitles on a dataset of 100 movies and 50 natural-language queries with predefined relevant movies. Cosine similarity, Recall@5, Hit Rate@5, Mean Reciprocal Rank, bootstrap confidence intervals, and the Wilcoxon signed-rank test were appliedto evaluate standalone and combined configurations. It was established that review-based embeddings achieved the highest standalone effectiveness in semantic movie retrieval, with Recall@5 = 0.63, Hit Rate@5 = 0.90, and MRR = 0.78, outperforming metadata and subtitle embeddings. It was found that reviews are better aligned with natural-language queries because they contain evaluative descriptions, thematic summaries, emotional impressions, and references to narrative elements. It was determined that subtitle embeddings, despite the largest text volume, did not provide the best ranking quality of relevant movies because dialogue contains filler phrases, context-dependent utterances, and semantically neutral fragments. It was confirmed that weighted aggregation of heterogeneous textual representations improved retrieval effectiveness of relevant movies compared with standalone representations: the balanced source-weight configuration configuration achieved the highest mean MRR = 0.89, whereas the reviews-heavy configuration achieved the highest mean Recall@5 = 0.73. Statistical testing of the experimental results was conducted and confirmed significant improvements of weighted configurations over the best standalone textual source for selected metrics; however, the differences between the balanced and reviews-heavy configurations were not statistically significant. It was therefore substantiated that, these configurations should be interpreted as comparable weighted strategies with different retrieval profiles. Query categories were analyzed, and it was established that descriptive queries tend to align with metadata, subjective queries with reviews, and contextual or character-based queries with several sources. The benefits of this study include determining the contribution of metadata, reviews, and subtitles to semantic movie retrieval and confirming the effectiveness of weighted source combination. Future research should extend the dataset, compare several embedding models, and develop adaptive query-dependent weighting strategies.

**Keywords:** information retrieval; vector search; multi-source representation; cosine similarity; Mean Reciprocal Rank.

## Introduction / Вступ

The rapid growth of digital multimedia content, particularly in streaming services, has created a pressing need for effective methods of searching and retrieving relevant items based on natural language queries. Traditional keyword-based search techniques often fail to capture the semantic meaning of user queries, leading to suboptimal retrieval results [11]. Movie search represents a particularly challenging domain, as users often describe films through subjective impressions, plot fragments, character descriptions, or emotional associations rather than precise titles or metadata keywords. This ambiguity demands retrieval systems capable of understanding semantic similarity between free-form queries and diverse textual representations of movies.

Movies are inherently multimodal objects described by structured metadata, user-generated reviews, and dialogue contained in subtitles. Each of these textual sources captures a different semantic facet of a film: metadata encodes factual attributes such as genre, cast, and plot synopsis; reviews reflect subjective audience perception and emotional evaluation; subtitles preserve narrative context, character interactions, and dialogue-level information. The hypothesis motivating this research is that combining dense vector representations from all three sources through weighted aggregation should produce more robust retrieval than relying on any single source alone.

Recent advances in transformer-based text embedding models have made it feasible to encode arbitrary-length text into fixed-size dense vectors that can be compared efficiently via cosine similarity. Modern long-context embedders such as Nomic Embed [4] support input lengths up to 8192 tokens, enabling direct processing of lengthy sources like subtitle transcripts without excessive chunking. Combined with vector similarity search infrastructure [10], these embeddings enable scalable semantic retrieval systems.

This paper proposes a semantic movie retrieval approach that generates dense embeddings for metadata, user re-


**Плюснін Артем Ігорович**, студент, кафедра програмного забезпечення.
**Email:** artem.pliusnin.pz.2022@lpnu.ua; https://orcid.org/0009-0007-7450-9730
**Малий Роман Михайлович**, аспірант, асистент, кафедра програмного забезпечення.
**Email:** roman.m.malyi@lpnu.ua; https://orcid.org/0000-0002-2255-1132
**Цитування за ДСТУ:** Плюснін А. І., Малий Р. М. Удосконалення методу семантичного пошуку фільмів на основі зваженого агрегування багатоджерельних ембедингів. Науковий вісник НЛТУ України. 2026, т. 36, № 3. С. 36–44.
**Citation APA:** Pliusnin, A. I., & Malyi, R. M. (2026). Improvement of a semantic movie retrieval method based on weighted aggregation of multi-source embeddings. *Scientific Bulletin of UNFU*, 36(3), 36–44. https://doi.org/10.36930/40360304

views, and subtitles using a single contrastive embedding model, computes cosine similarity independently for each textual source, and combines them through weighted aggregation into a unified relevance score. The effectiveness of the approach is evaluated on a test set of 50 queries across four semantic categories – descriptive, subjective, contextual, and character-based – using standard information retrieval metrics: Recall@K, Hit Rate@K [5], and Mean Reciprocal Rank [14]. The experimental analysis examines both overall retrieval quality under different weighting configurations and the behavior of individual similarity components at the query level, providing insight into the complementary contribution of each textual source.

*Object of research* – the processes of semantic information retrieval and relevance ranking of content based on natural-language user queries.

*Subject of research* – methods and means of weighted aggregation of similarity scores obtained from different embedding types for improving relevance ranking in semantic movie search.

*The purpose of the research* – to improve the semantic movie retrieval method based on dense vector representations through the weighted aggregation of multi-source embeddings derived from metadata, user reviews, and subtitles in order to increase retrieval effectiveness by combining heterogeneous textual representations.

To achieve this *purpose*, the following main *research objectives are identified*:

- to analyze existing approaches to semantic information retrieval based on dense vector embeddings and identify the limitations of single-source textual representations, which will enable the formation of a theoretical basis for combining heterogeneous semantic sources in movie retrieval systems;
- to develop a procedure for generating movie embeddings from heterogeneous textual sources, including metadata, user reviews, and subtitles, taking into account their structural and semantic differences, which will give an opportunity to compare the individual contribution of each source to retrieval effectiveness;
- to formalize a weighted aggregation method for combining similarity scores obtained from different embedding types into a unified relevance score, which will ensure controlled evaluation of the influence of weighting coefficients on movie ranking quality;
- to experimentally evaluate the proposed method using standard information retrieval metrics, including Recall@5, Hit Rate@5, and Mean Reciprocal Rank, and compare it with retrieval based on individual embedding types, which will help determine whether the integration of heterogeneous textual sources improves semantic search effectiveness;
- to investigate the dependence of weighting effectiveness on the semantic type of user query and formulate recommendations for the practical application of the proposed method, which will enable the selection of appropriate weighting strategies for different retrieval scenarios.

The implementation of these objectives provides a methodological basis for improving semantic retrieval systems by integrating structured, subjective, and contextual textual information into a unified relevance ranking mechanism. This will make it possible to evaluate the contribution of each textual source to retrieval effectiveness, determine the influence of weighting coefficients on ranking quality, and identify query-dependent patterns in the use of metadata, reviews, and subtitles. As a result, the proposed approach can support the development of more flexible retrieval systems capable of adapting semantic source contributions to different user query types and search objectives.

**Analysis of recent research and publications.** In the study [11], the authors proposed Sentence-BERT, a modification of the BERT architecture that uses siamese and triplet network structures to derive semantically meaningful sentence embeddings. The key contribution of this work lies in demonstrating that fixed-size dense vectors produced by fine-tuned transformers can be compared efficiently via cosine similarity, reducing pairwise sentence comparison from hours of cross-encoder inference to seconds of vector computation. This approach established the encode-once, search-with-cosine paradigm that forms the architectural foundation for modern semantic retrieval systems, including the embedding pipeline employed in the present study.

In the work [4], the authors introduced SimCSE, a contrastive learning framework for sentence embeddings that operates in both unsupervised and supervised settings. The unsupervised variant uses dropout as a minimal data augmentation strategy, treating two forward passes of the same input as positive pairs, while the supervised variant leverages natural language inference pairs. The InfoNCE-based contrastive training recipe proposed in this work has become the standard methodology adopted by virtually all subsequent text embedding models, including E5 and Nomic Embed, which is directly used in the present research.

In the study [10], the authors developed Nomic Embed, a fully open and reproducible long-context text embedder with an 8192-token context window. The model was trained using a contrastive learning pipeline on curated web pairs and demonstrated performance surpassing comparable commercial models such as OpenAI Ada-002 on both the MTEB benchmark and the LoCo long-context evaluation. The extended input length capacity of this model is particularly critical for the present study, as it enables direct processing of lengthy movie subtitle documents with reduced chunking, thereby preserving more contextual coherence in the resulting embeddings.

In the research [5], the authors investigated a hybrid movie recommendation system that fuses collaborative filtering with content-based filtering on movie synopses and user review text. The study demonstrated that combining user-generated review data with structured metadata through TF-IDF representations improves recommendation accuracy compared to using either source independently. This finding directly motivates the present work, which extends the same multi-source principle from sparse representations to dense contrastive embeddings and adds subtitle text as a third complementary source.

In the work [14], the authors proposed DeepCoNN, a deep learning model that jointly represents users and items through parallel convolutional neural networks applied to review text. The study confirmed that dense representations derived from review content outperform rating-only collaborative filtering models, establishing that user reviews carry rich semantic information beyond numerical scores. This conclusion supports the inclusion of review-based embeddings as one of the three textual sources in the present retrieval approach, particularly for capturing subjective and evaluative aspects of movie content.

In the study [2], the authors introduced NARRE, a neural attentional rating regression model that assigns learned attention weights to individual reviews when constructing item and user representations. The key finding of this work is that reviews contribute unequally to the overall representation – some reviews are significantly more informative

than others. This empirical observation provides direct justification for weighted aggregation strategies when combining per-review or per-source embeddings, as adopted in the present study where different textual sources receive different weighting coefficients in the final relevance score.

In the research [3], the authors directly investigated the value of subtitle text for improving movie recommendations by combining BPR collaborative filtering with LDA topic models and neural embeddings derived from subtitle data. The experiments were conducted on a large-scale dataset linking MovieLens-20M ratings with OpenSubtitles and IMDb metadata, and the results demonstrated substantial recommendation improvements, particularly for cold-start items where traditional collaborative signals are unavailable. This work provides the strongest empirical precedent for treating subtitles as a first-class textual source in movie retrieval systems and directly motivates the subtitle-based embedding component of the present approach.

In the work [9], the authors proposed a late-fusion approach for content-based recommendation of BBC television programmes, in which subtitle embeddings generated via LSI and Doc2Vec, audio features, and metadata vectors are combined through weighted averaging of cosine similarity matrices. The study demonstrated that weighted multi-source fusion outperforms individual representations and that the optimal weighting depends on the type of content being retrieved. This architecture represents the closest methodological precedent to the weighted cosine aggregation strategy employed in the present study, where metadata, review, and subtitle similarity scores are combined with configurable weighting coefficients to produce a unified relevance score.

*Research gap.* Despite the progress reviewed above, existing approaches typically utilize only one or two textual sources and often rely on sparse representations or older embedding models. No prior work has systematically combined metadata, user reviews, and subtitles as three complementary dense embedding sources generated by a single modern long-context contrastive embedder, fused through weighted cosine aggregation, and evaluated under standard retrieval metrics across different semantic query types. The present study addresses this gap.

**Materials and methods of research.** The study applies methods of semantic text representation, comparative experimental analysis, and information retrieval evaluation. The research was conducted on a dataset of 100 movies with metadata, user reviews, and English subtitles. Semantic similarity was computed using cosine similarity. Retrieval effectiveness was assessed using Recall@5, Hit Rate@5, Mean Reciprocal Rank, and statistical processing of query-level results.

## Research results and their discussion / Результати дослідження та їх обговорення

The experimental study aims to evaluate the effectiveness of different semantic representations of movies based on textual data and to analyze their contribution to search performance. In particular, the study investigates the impact of three types of embeddings derived from metadata, user reviews, and subtitles. The main objective of the research is to evaluate the influence of combining heterogeneous semantic representations on retrieval effectiveness. In particular, the study investigates how weighted aggregation affects ranking quality under different query conditions and aims to identify weighting strategies that provide stable and effective semantic matching across different types of search queries.

The experiments were conducted on a dataset consisting of 100 movies, each associated with structured metadata, user-generated reviews, and English-language subtitles. To assess retrieval quality, a test set of 50 queries was constructed, where each query corresponds to a predefined relevant movie. This setup allows evaluating the ability of the system to correctly identify the most relevant item under different query conditions.

The experimental evaluation focuses on three main aspects. First, the overall quality of semantic search is analyzed using ranking-based evaluation metrics. Second, the influence of weighting coefficients assigned to different embedding types is examined in order to determine their contribution to the final relevance score. Third, the behavior of the model is studied for different types of queries, including descriptive, subjective, and context-based formulations.

Such an experimental design makes it possible to systematically assess both the individual and combined effectiveness of embedding-based representations, as well as to identify patterns in their contribution to semantic similarity estimation.

***Embedding Generation and Time Analysis.*** Semantic representations of movies were generated using a transformer-based embedding model *nomic-embed-text* (*768-dimensional embedding space, 8192-token context window, 137M parameters, F16 quantization*) deployed locally via the Ollama inference framework. The model produces fixed-length dense vector representations for arbitrary textual input and is designed for semantic similarity and retrieval tasks. All embeddings were generated under identical model settings to ensure consistency and comparability of results.

For each movie, three types of textual documents were constructed prior to embedding generation. The metadata-based document was formed by concatenating structured attributes, including the movie title, release year, genres, cast members, character names, directors, production countries, duration, rating, and a textual description. These elements were combined into a single coherent textual representation that preserves the structure and semantic relationships between attributes. Due to the relatively small size of metadata, the entire document was processed as a single input without additional segmentation, which ensured stable and efficient embedding generation.

The review-based representation was constructed from user-generated reviews. To ensure consistency across movies, up to ten non-empty user reviews were selected for each movie in their original retrieval order provided by the data source. Unlike metadata, reviews introduce variability in language, style, and content, which increases the complexity of processing.Instead of concatenating all reviews into a single document, embeddings were generated independently for each review. This approach avoids excessive input length and preserves the semantic contribution of each review. The final movie-level representation was obtained by averaging individual review embeddings:

$$E = \frac{1}{n}\sum_{i=1}^{n} e_i , \qquad (1)$$

where $E$ denotes the resulting embedding, $e_i$ denotes the embedding of the $i$-th textual element, and $n$ denotes the number of aggregated elements. This aggregation method reduces the influence of noise and extreme opinions while

preserving the overall semantic direction of audience perception.

The subtitle-based embedding was generated from English subtitle text associated with each movie. The subtitle data were preprocessed to remove timestamps, formatting tags, and other non-textual elements, followed by whitespace normalization. Due to the input length limitations of the embedding model, the subtitle text was segmented into sequential fixed-length fragments of up to 4000 characters. Each fragment was processed independently to generate an embedding vector, and the final subtitle representation was obtained by averaging all fragment embeddings according to equation (1).

The computational cost of embedding generation was evaluated by measuring the total processing time and the average time per movie for each embedding type. Measurements were performed under identical hardware conditions (Apple M2 CPU, integrated GPU, 8 GB RAM) and within the same software environment to ensure comparability (Table 1).

**Table 1.** Computational time required to generate movie embeddings for different textual data sources / Обчислений час, необхідний для формування векторних подань фільмів за різними текстовими джерелами

| Embedding type | Total generation time, s | Average time per movie, s |
|---|---|---|
| Metadata | 21.407 | 0.214 |
| Reviews | 284.141 | 2.841 |
| Subtitles | 1522.845 | 15.228 |

The results indicate a substantial difference in embedding generation time depending on the type of input data. Metadata-based embeddings required 21.407 s in total, with an average of 0.214 s per movie, which confirms their high computational efficiency due to the compact and structured nature of the input. Review-based embeddings required significantly more time, reaching 284.141 s in total and 2.841 s per movie, as multiple independent inputs must be processed and aggregated for each instance. Subtitle-based embeddings demonstrated the highest computational cost, with a total generation time of 1522.845 s and an average of 15.228 s per movie. This increase is explained by the large volume of textual data and the need to process multiple fragments for each movie.

The obtained results clearly demonstrate the existence of a trade-off between computational efficiency and semantic expressiveness. Metadata embeddings provide fast processing but capture only general characteristics of movies. Review-based embeddings introduce additional semantic information related to user perception, at the cost of increased processing time. Subtitle-based embeddings, despite being the most computationally expensive, provide the most detailed representation of contextual and narrative information.

***Experiment 1. Search Quality Evaluation.*** The first experiment evaluates the effectiveness of semantic search using different embedding types and their combinations. The objective of the experiment is to verify the hypothesis that combining multiple semantic representations improves retrieval quality compared to using individual embeddings.

The evaluation was performed on a test set consisting of 50 query scenarios. Each query was manually designed to represent a specific semantic description of movies, including plot-based, character-based, subjective, and contextual formulations. For each query, a predefined set of relevant movies was assigned. The number of relevant movies differs across test cases, which makes it possible to evaluate not only whether the system retrieves a single expected item, but also whether it ranks semantically related movies within the top results. This setup allows assessing retrieval quality by analyzing the presence and positions of relevant movies in the ranked list.

For each query, a semantic embedding was generated using the same embedding model as for movie representations. The similarity between the query embedding and each movie embedding was computed independently for metadata, reviews, and subtitles using cosine similarity:

$$Sim(q,m) = \frac{q \cdot m}{\|q\|\|m\|}, \quad (2)$$

where $q$ denotes the query embedding and $m$ denotes the movie embedding.

The retrieved movies were ranked in descending order according to similarity scores. To evaluate the quality of the ranking, three standard information retrieval metrics were applied: Hit Rate@K, and Mean Reciprocal Rank (MRR). These metrics provide complementary perspectives on retrieval effectiveness by measuring the completeness, success rate, and ranking position of relevant results.

Recall@K measures the proportion of relevant movies that appear within the first $K$ retrieved results. This metric reflects the ability of the retrieval system to cover the complete set of relevant items for a given query and is defined by equation (3):

$$\mathrm{Re}call@K = \frac{R_K}{R}, \quad (3)$$

where $R_K$ – denotes the number of relevant movies retrieved within the top $K$ positions, and $R$ denotes the total number of relevant movies associated with the query.

A higher Recall@K value indicates better retrieval completeness, meaning that the system successfully retrieves a larger portion of semantically relevant movies. This metric is particularly important in scenarios where multiple relevant results exist and their coverage is essential.

Hit Rate@K evaluates whether at least one relevant movie appears within the first $K$ retrieved results. Unlike Recall@K, which measures the proportion of retrieved relevant items, Hit Rate@K focuses on the success of retrieving any relevant result. The metric is defined by equation (4):

$$HR@K = \frac{1}{N}\sum_{i=1}^{N} Hit_t@K, \quad (4)$$

where $N$ denotes the total number of test queries, and $Hit_i@K$ takes the value 1 if at least one relevant movie is found within the top $K$ results for the $i$-th query, and 0 otherwise.

This metric reflects the practical usability of the search system, since in many retrieval scenarios the presence of at least one relevant result in the top positions is sufficient for successful search.

Mean Reciprocal Rank (MRR) evaluates the ranking quality by considering the position of the first relevant result. The metric is based on the reciprocal rank of the first correctly retrieved movie and is defined by equation (5):

$$MRR = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{rank_i}, \quad (5)$$

where $rank_i$ denotes the rank position of the first relevant movie for the $i$-th query.

Higher MRR values indicate that relevant results appear earlier in the ranked list, which reflects better ranking precision from the user's perspective. This metric is particularly

important in semantic search systems, where the position of the first relevant result directly affects the efficiency of information retrieval.

The first stage of the experiment evaluates the retrieval effectiveness of individual embedding types. Three independent configurations were tested: metadata-only, reviews-only, and subtitles-only retrieval. This comparison makes it possible to identify the semantic contribution of each representation independently and determine which textual source provides the highest retrieval quality under different query conditions. The evaluation was performed using Recall@5, Hit Rate@5, and Mean Reciprocal Rank, where the top five retrieved results were considered for each query. The obtained results for individual embedding types are presented in Table 2.

**Table 2.** Retrieval effectiveness of individual embedding types according to Recall@5, Hit Rate@5, and Mean Reciprocal Rank / Ефективність пошуку для окремих типів векторних подань за метриками Recall@5, Hit Rate@5 та Mean Reciprocal Rank

| Embedding type | Recall@5 | Hit Rate@5 | MRR |
|---|---|---|---|
| Metadata | 0.53 | 0.88 | 0.67 |
| Reviews | 0.63 | 0.9 | 0.78 |
| Subtitles | 0.55 | 0.9 | 0.65 |

The results presented in Table 2 show differences in retrieval effectiveness across individual embedding types. Metadata-based embeddings achieved Recall@K = 0.53, Hit Rate@K = 0.88, and MRR = 0.67. These results indicate stable retrieval performance for queries related to structured semantic attributes such as genre, plot, and cast information.

Review-based embeddings demonstrated the highest effectiveness among standalone representations, with Recall@K = 0.63, Hit Rate@K = 0.90, and MRR = 0.78. The higher values indicate that review texts provide richer semantic information for retrieval, as they contain descriptive and evaluative content closer to natural user queries.

Subtitle-based embeddings achieved Recall@K = 0.55, Hit Rate@K = 0.90, and MRR = 0.65. Despite the higher contextual density of subtitle data, their ranking effectiveness remained lower than that of review-based embeddings, which indicates the presence of redundant or semantically neutral information in dialogue sequences.

The comparative analysis of individual embedding types shows that review-based embeddings provide the most effective standalone semantic representation for movie retrieval tasks. Metadata embeddings preserve stable performance due to their structured nature, while subtitle embeddings provide broader contextual coverage but lower ranking precision. These observations indicate that different textual sources capture complementary semantic properties, which supports the hypothesis that combining them may improve retrieval effectiveness.

The second stage of the experiment evaluates combined embeddings using weighted similarity aggregation. For each movie, the final relevance score was computed as a weighted sum of similarity values obtained from metadata, reviews, and subtitles according to equation:

$$S = w_m S_m + w_r S_r + w_s S_s \,, \tag{6}$$

where $S_m$, $S_r$ and $S_s$ denote similarity values for metadata, reviews and subtitles, and $w_m$, $w_r$ and $w_s$ denote the corresponding weighting coefficients satisfying the normalization condition $w_m + w_r + w_s = 1$.

To evaluate the influence of weighted similarity aggregation on retrieval effectiveness, a coarse sensitivity analysis of weight distributions was performed. The balanced configuration was used as a neutral baseline, assigning approximately equal contribution to all embedding types. In addition, three source-dominant configurations were examined, where one embedding type received the dominant weight of 0.6, while the remaining sources received 0.2 each. This symmetric weighting scheme preserves the normalization condition defined in equation (6) and allows isolating the effect of increasing the contribution of each semantic source under comparable conditions. The selected weight values were not considered optimal model parameters, but controlled experimental configurations for evaluating the sensitivity of retrieval quality to semantic emphasis. The retrieval results for weighted embedding configurations are summarized in Table 3.

**Table 3.** Retrieval effectiveness under different weighting configurations of combined embeddings / Ефективність пошуку за різних конфігурацій зважування комбінованих векторних подань

| Configuration | Weight distribution ($w_m$: $w_r$: $w_s$) | Recall@5 | Hit Rate@5 | MRR |
|---|---|---|---|---|
| Balanced | 0.33: 0.33: 0.33 | 0.7 | 0.98 | 0.89 |
| Metadata-heavy | 0.6: 0.2: 0.2 | 0.65 | 0.94 | 0.84 |
| Reviews-heavy | 0.2: 0.6: 0.2 | 0.73 | 0.98 | 0.88 |
| Subtitles-heavy | 0.2: 0.2: 0.6 | 0.68 | 0.98 | 0.85 |

The results presented in Table 3 demonstrate that combining multiple semantic representations improves retrieval effectiveness compared to standalone embeddings. The balanced configuration achieved the highest mean MRR, with Recall@5 = 0.70, Hit Rate@5 = 0.98, and MRR = 0.89. Compared to the best standalone configuration, namely review-based embeddings, the balanced weighting strategy improved Recall@5 by 11.1 % and MRR by 14.1 %, which indicates the effectiveness of semantic integration.

The metadata-heavy configuration achieved lower performance among the combined variants, with Recall@5 = 0.65, Hit Rate@5 = 0.94, and MRR = 0.84, indicating that emphasizing structured semantic attributes alone limits retrieval flexibility. The reviews-heavy configuration achieved the highest mean Recall@5 = 0.73, demonstrating the strongest retrieval completeness, while the subtitles-heavy configuration showed intermediate performance, with Recall@5 = 0.68, Hit Rate@5 = 0.98, and MRR = 0.85.

To assess the statistical stability of the obtained retrieval metrics, bootstrap 95 % confidence intervals were calculated for all standalone and weighted configurations. The confidence intervals were estimated at the query level using non-parametric bootstrap resampling. The results are presented in Table 4.

**Table 4.** Bootstrap 95 % confidence intervals for retrieval effectiveness under all configurations / Bootstrap 95 % довірчі інтервали для ефективності пошуку за всіма конфігураціями

| Configuration | Recall@5 | 95 % CI | Hit Rate@5 | 95 % CI | MRR | 95 % CI |
|---|---|---|---|---|---|---|
| Metadata | 0.53 | [0.45–0.62] | 0.88 | [0.78–0.96] | 0.67 | [0.57–0.77] |
| Reviews | 0.63 | [0.54–0.72] | 0.9 | [0.82–0.98] | 0.78 | [0.68–0.87] |
| Subtitles | 0.55 | [0.47–0.64] | 0.9 | [0.82–0.98] | 0.65 | [0.55–0.75] |
| Balanced | 0.7 | [0.62–0.77] | 0.98 | [0.94–1] | 0.89 | [0.82–0.95] |
| Metadata-heavy | 0.65 | [0.57–0.73] | 0.94 | [0.86–1] | 0.84 | [0.76–0.92] |
| Reviews-heavy | 0.73 | [0.65–0.8] | 0.98 | [0.94–1] | 0.88 | [0.82–0.95] |
| Subtitles-heavy | 0.68 | [0.6–0.75] | 0.98 | [0.94–1] | 0.85 | [0.78–0.92] |

The bootstrap confidence intervals confirm that weighted configurations generally demonstrate higher retrieval effectiveness than standalone embeddings. The balanced and reviews-heavy configurations show the strongest results among the tested variants. However, their MRR confidence intervals overlap completely ([0.82; 0.95] in both cases), which indicates that the difference between MRR = 0.89 for the balanced configuration and MRR = 0.88 for the reviews-heavy configuration should be interpreted as a small difference in mean values rather than as clear evidence of the superiority of one weighting strategy over the other.

To further evaluate whether the observed differences between selected configurations are statistically significant, pairwise comparisons were performed using the Wilcoxon signed-rank test. The compared configurations were selected according to their role in the experimental results presented in Tables 2 and 3. The reviews-only configuration was included as the strongest standalone baseline, while the balanced and reviews-heavy configurations were selected as the two best-performing weighted variants, since the balanced configuration achieved the highest mean MRR and the reviews-heavy configuration achieved the highest mean Recall@5. Therefore, the statistical comparison focused on whether weighted aggregation improves retrieval quality over the strongest individual source and whether the two leading weighted strategies differ significantly from each other. The test was applied to query-level Recall@5 and reciprocal rank values, since all configurations were evaluated on the same test queries. The results are presented in Table 5.

**Table 5.** Pairwise Wilcoxon signed-rank test results for selected retrieval configurations / Результати парного критерію Вілкоксона для вибраних конфігурацій пошуку

| Comparison | Metric | Mean difference | p-value | Interpretation |
|---|---|---|---|---|
| Balanced & Reviews | Recall@5 | 0.069 | 0.073 | not significant |
| Balanced & Reviews | MRR | 0.114 | 0.04 | significant |
| Reviews-heavy & Reviews | Recall@5 | 0.095 | 0.013 | significant |
| Reviews-heavy & Reviews | MRR | 0.108 | 0.029 | significant |
| Balanced & Reviews-heavy | Recall@5 | -0.027 | 0.345 | not significant |
| Balanced & Reviews-heavy | MRR | 0.006 | 0.859 | not significant |

The Wilcoxon signed-rank test showed that the balanced configuration significantly outperformed the reviews-only configuration in terms of MRR ($p = 0.040$), while the difference in Recall@5 was not statistically significant ($p = 0.073$). The reviews-heavy configuration significantly outperformed reviews-only for both Recall@5 ($p = 0.013$) and MRR ($p = 0.029$), confirming that weighted aggregation provides statistically supported improvements over the strongest standalone embedding source.

However, the differences between the balanced and reviews-heavy configurations were not statistically significant for either Recall@5 ($p = 0.345$) or MRR ($p = 0.859$). Therefore, although the balanced configuration achieved the highest mean MRR and the reviews-heavy configuration achieved the highest mean Recall@5, these two weighted strategies should be interpreted as comparable under the current experimental conditions. The comparative ranking quality across all standalone and combined configurations is illustrated in Figure 1.

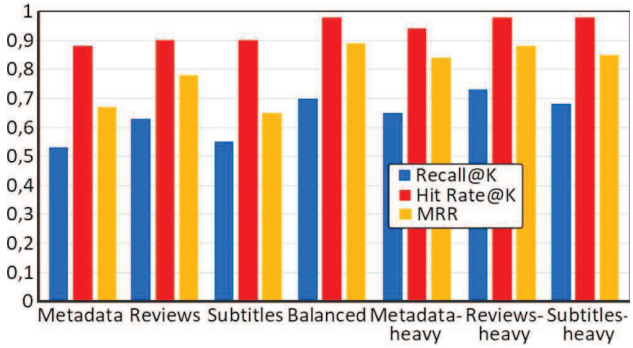


**Fig. 1.** Comparative Mean Reciprocal Rank values for standalone and combined embedding configurations / Порівняльні значення Mean Reciprocal Rank для окремих і комбінованих конфігурацій векторних подань

The results of the first experiment confirm that combining semantic representations from multiple textual sources improves search quality compared with standalone embeddings. At the same time, statistical processing indicates that the strongest weighted configurations have comparable effectiveness, which establishes the basis for the second experiment, where the influence of weighting coefficients is analyzed at the level of individual query – movie pairs.

***Experiment 2. Analysis of Weight Influence and Query Types.*** The second experiment investigates the influence of weighting strategies on semantic similarity under fixed query – movie pairs. Unlike the first experiment, which evaluates overall retrieval effectiveness across the complete test set, this experiment focuses on the behavior of individual similarity components and their contribution to the final relevance score under different semantic query types.

For each experimental case, a target movie and a semantically relevant query were selected. The similarity between the query embedding and the corresponding movie embeddings was computed independently for metadata, reviews, and subtitles using cosine similarity according to equation (2). These similarity values represent the semantic contribution of each embedding type and are denoted as $S_m$, $S_r$ and $S_s$ respectively.

To analyze the effect of weighting strategies, three weighted configurations were applied: metadata-heavy, reviews-heavy, and subtitles-heavy. Each configuration increases the contribution of one embedding type while preserving the contribution of the remaining sources. The final relevance score for each configuration was computed according to equation (6). The experimental results are summarized in Table 6. The queries were divided into four semantic categories: descriptive, subjective, contextual, and character-based. Descriptive queries represent general plot and thematic descriptions and are expected to align with metadata embeddings. Subjective queries contain evaluative or emotional formulations and are expected to correlate with review embeddings. Contextual queries describe specific scenes or situations and are associated with subtitle embeddings. Character-based queries focus on character interactions and may involve both metadata and subtitle representations.

The results show a clear dependency between query type and dominant semantic source. Descriptive queries demonstrated higher metadata similarity in most cases, confirming the effectiveness of structured semantic information for general movie descriptions. Subjective queries consistently produced the highest review-based similarity values, indicating that user-generated reviews provide the strongest semantic alignment for evaluative formulations.

**Table 6.** Similarity analysis of query-movie pairs under different weighting configurations / Аналіз подібності пар "запит-фільм" за різних конфігурацій зважування

| Movie | Query type | $S_m$ | $S_r$ | $S_s$ | Metadata-heavy | Reviews-heavy | Subtitles-heavy |
|---|---|---|---|---|---|---|---|
| Interstellar | Descriptive | 0.76 | 0.64 | 0.71 | 0.72 | 0.68 | 0.7 |
| Oppenheimer | Descriptive | 0.71 | 0.73 | 0.73 | 0.72 | 0.73 | 0.73 |
| Arrival | Descriptive | 0.66 | 0.67 | 0.61 | 0.65 | 0.65 | 0.63 |
| The Batman | Subjective | 0.53 | 0.73 | 0.6 | 0.58 | 0.66 | 0.61 |
| Whiplash | Subjective | 0.5 | 0.7 | 0.52 | 0.54 | 0.62 | 0.56 |
| La La Land | Subjective | 0.65 | 0.71 | 0.66 | 0.66 | 0.69 | 0.67 |
| Inception | Contextual | 0.63 | 0.76 | 0.69 | 0.67 | 0.72 | 0.69 |
| The Matrix | Contextual | 0.57 | 0.61 | 0.48 | 0.56 | 0.59 | 0.52 |
| 12 Angry Men | Contextual | 0.55 | 0.58 | 0.56 | 0.56 | 0.57 | 0.57 |
| The Dark Knight | Character-based | 0.65 | 0.6 | 0.58 | 0.63 | 0.61 | 0.6 |
| Spider-Man: No Way Home | Character-based | 0.67 | 0.63 | 0.68 | 0.66 | 0.65 | 0.67 |
| Dune: Part Two | Character-based | 0.65 | 0.57 | 0.59 | 0.64 | 0.6 | 0.66 |

Contextual queries showed mixed behavior. Although subtitle embeddings were expected to dominate, review embeddings produced higher similarity values in several cases. This indicates that user reviews also preserve contextual semantic information through references to important scenes and narrative elements. Character-based queries demonstrated distributed semantic dominance between metadata and subtitles. Metadata embeddings achieved higher similarity for structured character information, while subtitle embeddings contributed more strongly when character interactions were explicitly represented in dialogue.

The weighted similarity analysis confirms that the highest final score generally corresponds to the weighting configuration that emphasizes the dominant semantic source. Metadata-heavy weighting improved descriptive queries, reviews-heavy weighting improved subjective and contextual queries, and subtitles-heavy weighting showed higher effectiveness for some character-based queries. The distribution of dominant embedding sources across query categories is illustrated in Figure 2.

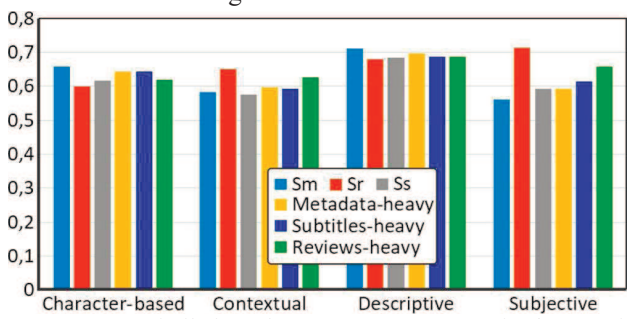


**Fig. 2.** Average similarity values across query categories / Середні значення подібності для різних категорій запитів

The results of the second experiment confirm that retrieval effectiveness depends on the semantic structure of the query. Different embedding types demonstrate higher effectiveness for different query categories, which supports the use of adaptive weighting strategies for semantic retrieval.

*Interpretation and Discussion of Results.* Previous studies have shown the effectiveness of dense embeddings for semantic retrieval and the potential of textual sources for content-based recommendation. However, existing approaches usually focus on one dominant representation or do not systematically evaluate the contribution of several textual sources across query types. The present study addresses this limitation by combining metadata, user reviews, and subtitles as complementary dense embedding sources fused through weighted cosine aggregation.

***Differential effectiveness of individual textual sources.*** Review-based embeddings demonstrated the highest standalone retrieval effectiveness, with Recall@5 = 0.63 and MRR = 0.78, outperforming metadata-based embeddings (0.53/0.67) and subtitle-based embeddings (0.55/0.65). This result can be explained by linguistic register alignment: user reviews are closer to natural-language search queries than structured metadata or raw subtitles. Reviews contain thematic summaries, evaluative judgments, emotional descriptions, and references to memorable scenes, which makes them semantically close to the way users formulate search requests.

Metadata embeddings remain useful for descriptive and factual queries because they encode structured attributes such as genres, cast, production details, and plot summaries. However, their compact and attribute-oriented nature limits their effectiveness for subjective or impressionistic queries. Subtitle embeddings, despite being based on the largest textual source, achieved lower standalone ranking effectiveness. This indicates that text volume does not directly determine retrieval utility. Subtitles contain dialogue-level information, including filler utterances, incomplete references, and context-dependent exchanges, which may dilute the movie-level semantic signal when averaged across fragments.

***The Recall-MRR trade-off.*** An important result is the divergence between the reviews-heavy and balanced weighting configurations in terms of mean metric values. The reviews-heavy configuration achieved the highest mean Recall@5, whereas the balanced configuration achieved the highest mean MRR (0.89). This indicates a potential trade-off between retrieval completeness and ranking precision. However, statistical testing showed that the differences between these two configurations were not significant for either Recall@5 or MRR. Therefore, this trade-off should be interpreted as a tendency in mean values rather than as a statistically confirmed superiority of one weighted strategy over the other.

Reviews-heavy weighting increases the influence of broad thematic and evaluative information, which may help retrieve more relevant movies within the top five results. However, the same semantic breadth may reduce top-rank discrimination, because thematically similar but less precise matches can receive high similarity scores. Balanced weighting combines review semantics with metadata and subtitle signals, which may stabilize the ranking of the first relevant result. Since the balanced and reviews-heavy configurations demonstrated comparable statistical effectiveness, the choice between them should depend on the target retrieval objective: broader top-K coverage or stronger emphasis on the first relevant result.

***Limitations.*** Several limitations should be considered when interpreting the obtained results. First, the experimental dataset contains 100 movies, which is sufficient for

controlled comparison but limited for generalizing the findings to large-scale movie retrieval systems. Second, the use of English-only subtitles may introduce linguistic and cultural bias, since subtitle availability and quality are higher for English-language or internationally distributed films. Third, the experiments were conducted using a single embedding model, so the stability of the results across different embedding architectures was not evaluated. Finally, the test queries were manually constructed, which ensures controlled evaluation conditions but may introduce bias related to the query designer's assumptions about relevance and query formulation. Future work should address these limitations by using larger multilingual datasets, comparing several embedding models, and validating the approach on user-generated queries.

**Discussion of research results.** In the work [8], the authors showed that pretrained sentence embeddings improve movie identification from short user descriptions compared with sparse keyword-based methods. The present results support the value of dense semantic representations but show that retrieval effectiveness depends not only on the embedding model, but also on the integration strategy used for multiple textual sources.

In the research [7], the authors proposed a multi-aspect reviewed-item retrieval framework that decomposes queries into aspect-level sub-queries and aggregates bi-encoder query-review similarities to item-level scores through late fusion. The present study arrives at a conceptually similar conclusion: aggregating similarity signals from multiple textual facets improves retrieval. However, the proposed approach achieves this through a simpler weighted-sum mechanism without query decomposition, demonstrating that even straightforward fusion strategies can yield measurable gains when applied to semantically complementary sources.

In the study [1], the authors developed a Mixture-of-Embedding-Experts model for story-based movie retrieval that fuses subtitle, visual, and audio embeddings through learned expert-specific weights. The present work differs by operating exclusively in the textual domain and using a single embedding model for all three sources rather than separate modality-specific experts. Despite this simpler design, the observed improvements from weighted aggregation support the usefulness of multi-source combination in text-only retrieval settings.

In the work [12], the authors introduced a hybrid document retrieval framework combining FAISS and HNSWlib indexing for multi-vector search in high-dimensional spaces. While the present study does not employ distributed indexing infrastructure, it addresses the same underlying challenge of comparing a query against multiple embedding vectors per item. The weighted aggregation of per-source cosine similarities used here can be viewed as a lightweight alternative to multi-vector retrieval for moderate-scale datasets.

In the research [6], the authors demonstrated that dual-encoder dense embeddings trained with in-batch contrastive negatives significantly outperform BM25 on passage retrieval tasks, establishing the dense retrieval paradigm. The present study extends this paradigm from single-document retrieval to multi-source item retrieval, where each movie is represented by three independent dense vectors rather than one vector representation.

In the study [13], the authors proposed E5, a general-purpose text embedder trained through weakly supervised contrastive pre-training and evaluated on diverse retrieval benchmarks. The present results show that the advantages of modern dense embeddings can be transferred to the domain-specific task of movie retrieval when heterogeneous textual sources are combined through weighted aggregation.

The discussion confirms that the proposed approach contributes to semantic movie retrieval by combining three textual sources within a single weighted aggregation framework and evaluating their behavior across different query categories. The statistically supported improvements of combined configurations over standalone sources indicate that multi-source semantic aggregation can improve retrieval quality, while the query-level analysis shows that the contribution of each source depends on the semantic structure of the query.

So, based on the results of the work performed, it is possible to formulate the following scientific novelty and practical significance of the research results.

*Scientific novelty of the obtained research results* – the method of semantic movie retrieval based on dense vector representations has been improved by introducing weighted aggregation of multi-source embeddings generated from metadata, user reviews, and subtitles, which, unlike approaches based on a single textual source increased MRR from 0,78 to 0,89 by considering the contribution of multiple heterogeneous textual sources that describe different aspects of movie content.

*Practical significance of the research results* – the results of the study can be applied in the design of content-based semantic retrieval systems for selecting weighting strategies according to the target retrieval objective, in particular, balanced aggregation can be used for ranking-precision-oriented search, reviews-heavy weighting for discovery-oriented retrieval with higher top-K coverage, and metadata-heavy weighting for structured-query scenarios where factual attribute matching dominates user intent.

## Conclusions / Висновки

The effectiveness of semantic movie retrieval was improved by developing and evaluating a weighted aggregation method for combining heterogeneous textual embeddings generated from metadata, user reviews, and subtitles. The conclusions should be correlated with the objectives and the data analysed.

1. Analyzed current scientific approaches to semantic information retrieval based on dense vector representations and identified the limitation of single-source movie representations. It was established that metadata, user reviews, and subtitles describe different semantic aspects of movie content and therefore should be considered as complementary sources for relevance ranking.
2. Developed a procedure for generating dense movie embeddings from three heterogeneous textual sources: metadata, user reviews, and English subtitles. Metadata embeddings represent structured factual information, review embeddings reflect subjective and evaluative semantic content, and subtitle embeddings preserve contextual and narrative information.
3. Formalized a weighted aggregation method for combining similarity scores obtained from different embedding types into a unified relevance score. The method enables controlled regulation of metadata-, review-, and subtitle-based contributions during semantic ranking and allows comparison of balanced and source-dominant weighting configurations.
4. Conducted an experimental evaluation on a dataset of 100 movies and 50 natural-language queries using Recall@5, Hit Rate@5, Mean Reciprocal Rank, bootstrap confidence intervals, and the Wilcoxon signed-rank test. It was established that review-based embeddings achieved the highest

standalone effectiveness with Recall@5 = 0.63, Hit Rate@5 = 0.90, and MRR = 0.78.

5. Determined that weighted aggregation improves retrieval effectiveness compared with standalone embedding configurations. The balanced configuration achieved the highest ranking precision with MRR = 0.89, whereas the reviews-heavy configuration achieved the highest retrieval completeness with Recall@5 = 0.73, which indicates a difference between ranking-oriented and coverage-oriented retrieval objectives.
6. Established that the contribution of textual sources depends on the semantic type of query. Descriptive queries tend to align with metadata embeddings, subjective queries with review-based embeddings, while contextual and character-based queries demonstrate mixed source contribution, which substantiates the need for adaptive weighting strategies in semantic movie retrieval systems.
7. Outlined further research directions related to extending the dataset, using multilingual subtitles and queries, comparing several embedding models, and developing adaptive mechanisms for selecting weighting coefficients according to query type and retrieval objective.

## References


1. Bain, M., Nagrani, A., Brown, A., & Zisserman, A. (2020). Condensed movies: Story based retrieval with contextual embeddings. *Proceedings of the Asian Conference on Computer Vision* (*ACCV 2020*). 460–479. https://doi.org/10.48550/arXiv.2005.04208
2. Chen, C., Zhang, M., Liu, Y., & Ma, S. (2018). Neural Attentional Rating Regression with Review-level Explanations. *Proceedings of the 2018 World Wide Web Conference* (*WWW 18*). *2021*, pp. 1583–1592. https://doi.org/10.1145/3178876.3186070
3. Eden, S., Livne, A., Shalom, O. S., Shapira, B., & Jannach, D. (2022). Investigating the Value of Subtitles for Improved Movie Recommendations. *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (*UMAP 22*). *2022*, pp. 99–109. https://doi.org/10.1145/3503252.3531291
4. Gao, T., Yao, X., & Chen, D. (2021). SIMCSE: Simple Contrastive Learning of Sentence Embeddings. *arXiv* (*Cornell University*). https://doi.org/10.48550/arxiv.2104.08821
5. Iliopoulou, K., Kanavos, A., Ilias, A., Makris, C., & Vonitsanos, G. (2020). Improving movie recommendation systems filtering by exploiting User-Based reviews and movie synopses. In *IFIP advances in information and communication technology* (pp. 187–199). https://doi.org/10.1007/978-3-030-49190-1_17
6. Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). Dense passage retrieval for open-domain question answering. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (*EMNLP*). *2020,* 6769–6781. https://doi.org/10.48550/arXiv.2004.04906
7. Korikov, A., Saad, G., Baron, E., Khan, M., Shah, M., & Sanner, S. (2024). Multi-Aspect Reviewed-Item retrieval via LLM query decomposition and aspect fusion. *arXiv* (*Cornell University*). https://doi.org/10.48550/arxiv.2408.00878
8. Mustafa, A., Mheidat, H., & Shatnawi, A. (2024). A semantic similarity search engine for movies. *International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering, 14*(*6*), article ID 7137. https://doi.org/10.11591/ijece.v14i6.pp7137-7144
9. Nazir, S., Cagali, T., Newell, C., & Sadrzadeh, M. (2020). Cosine similarity of multimodal content vectors for TV programmes. *arXiv* (*Cornell University*). https://doi.org/10.48550/arxiv.2009.11129
10. Nussbaum, Z., Morris, J. X., Duderstadt, B., & Mulyar, A. (2024). Nomic Embed: Training a reproducible long context text embedder. *arXiv* (*Cornell University*). https://doi.org/10.48550/arxiv.2402.01613
11. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *TUbilio* (*Technical University of Darmstadt*). https://doi.org/10.48550/arxiv.1908.10084
12. VectorSearch: Enhancing Document Retrieval with Semantic Embeddings and Optimized Search. (n.d.). https://arxiv.org/html/2409.17383v1#S5
13. Wang, L., Yang, N., Huang, X., Jiao, B., Yang, L., Jiang, D., Majumder, R., & Wei, F. (2022). Text embeddings by Weakly-Supervised contrastive pre-training. *arXiv* (*Cornell University*). https://doi.org/10.48550/arxiv.2212.03533
14. Zheng, L., Noroozi, V., & Yu, P. S. (2017). Joint deep modeling of users and items using reviews for recommendation. *arXiv* (*Cornell University*). https://doi.org/10.48550/arxiv.1701.04783



***А. І. Плюснін, Р. М. Малий***

*Національний університет "Львівська Політехніка", м. Львів, Україна*


## УДОСКОНАЛЕННЯ МЕТОДУ СЕМАНТИЧНОГО ПОШУКУ ФІЛЬМІВ НА ОСНОВІ ЗВАЖЕНОГО АГРЕГУВАННЯ БАГАТОДЖЕРЕЛЬНИХ ЕМБЕДИНГІВ


Досліджено особливості підвищення ефективності семантичного пошуку фільмів поєднанням різнорідних текстових джерел, які описують різні аспекти кіноконтенту. Перевірено підхід до зваженого агрегування щільних векторних подань, сформованих з метаданих, користувацьких відгуків та англомовних субтитрів, на наборі даних зі 100 фільмів і 50 природномовних запитів із наперед визначеними релевантними фільмами. В експерименті застосовано косинусну подібність, Recall@5, Hit Rate@5, Mean Reciprocal Rank, bootstrap-довірчі інтервали та критерій Вілкоксона для оцінювання окремих і комбінованих конфігурацій. Встановлено, що векторні подання на основі відгуків забезпечили найвищу ефективність семантичного пошуку фільмів серед окремих джерел: Recall@5 = 0,63, Hit Rate@5 = 0,90 і MRR = 0,78, перевищивши результати метаданих і субтитрів. Це свідчить про те, що відгуки краще узгоджуються з природномовними запитами, оскільки містять оцінкові описи, тематичні узагальнення, емоційні враження та згадки про сюжетні елементи. Визначено, що субтитри, попри найбільший обсяг тексту, не забезпечили найкращої якості ранжування релевантних фільмів через наявність діалогових заповнювачів, контекстно залежних висловлювань і семантично нейтральних фрагментів. Підтверджено, що зважене агрегування різнорідних текстових подань покращило ефективність пошуку релевантних фільмів порівняно з окремими поданнями: збалансована конфігурація вагомостей джерел досягла найвищого середнього MRR = 0,89, тоді як конфігурація з домінуванням відгуків досягла найвищого середнього Recall@5 = 0,73. Проведено статистичне тестування результатів експерименту, яке підтвердило значні покращення зважених конфігурацій порівняно з найкращим окремим текстовим джерелом для окремих метрик; однак відмінності між збалансованою конфігурацією та конфігурацією з домінуванням відгуків не були статистично значущими. Отже, ці конфігурації доцільно розглядати як порівнянні зважені стратегії з різними профілями пошуку. Проаналізовано категорії запитів і встановлено, що описові запити тяжіють до метаданих, суб'єктивні до відгуків, а контекстні та персонажно орієнтовані запити залежать від кількох джерел. Практична цінність дослідження полягає у визначенні внеску метаданих, відгуків і субтитрів у семантичний пошук фільмів та підтвердженні ефективності зваженого поєднання джерел. Окреслені подальші етапи дослідження доцільно спрямувати на розширення набору даних, порівняння кількох embedding-моделей і розроблення адаптивних стратегій зважування залежно від типу запиту.



**Ключові слова:** інформаційний пошук; векторний пошук; багатоджерельне подання; косинусна подібність; середній взаємний ранг.