



Article

A Hybrid Approach for Personalized and Intelligent Content Recommendation in Digital Libraries

Emanuela Mitreva, Desislava Paneva-Marinova, Vladimir Georgiev, Alexandra Nikolova and Radoslav Pavlov

Special Issue

Intelligent Interaction in Cultural Heritage





Edited by

Dr. Tibor Szkaliczki, Prof. Dr. Desislava Paneva-Marinova, Dr. Detelin Luchev,
Dr. Nektarios Moutzidis, Prof. Dr. János Demetrovics and Prof. Dr. Radoslav Pavlov



Article

A Hybrid Approach for Personalized and Intelligent Content Recommendation in Digital Libraries

Emanuela Mitreva ¹, Desislava Paneva-Marinova ^{1,*}, Vladimir Georgiev ², Alexandra Nikolova ¹
and Radoslav Pavlov ¹

¹ Institute of Mathematics and Informatics, Bulgarian Academy of Sciences (IMI-BAS), 8, Acad. G. Bonchev Str., 1113 Sofia, Bulgaria; emitreva@gmail.com (E.M.); a.nikolova@math.bas.bg (A.N.); radko@cc.bas.bg (R.P.)

² Computer Science Department, America University in Bulgaria, 1, Georgi Izmirliov Sq., 2700 Blagoevgrad, Bulgaria; vgeorgiev@aubg.edu

* Correspondence: dessi@cc.bas.bg

Abstract

The rapid digitization of cultural heritage materials has led to the substantial growth of digital library collections, particularly large and heterogeneous archives of periodicals. This expansion has intensified challenges related to content discovery, accessibility, and user engagement. Users increasingly struggle to navigate large periodical collections and identify relevant materials. In this context, intelligent interaction with cultural content has become an essential aspect of effectively accessing and utilizing resources in modern digital libraries, highlighting the need for adaptive and user-oriented mechanisms that support navigation and discovery. Artificial intelligence-driven personalization offers promising solutions. However, digital library environments often contain sparse interaction data, evolving user interests, and continuously growing collections. These characteristics limit the effectiveness of standalone content-based or collaborative approaches. This work proposes an integrated personalization approach that combines behavioral interaction data with semantic relationships between documents to support adaptive content delivery in digital libraries. The approach facilitates the discovery of both established and newly digitized or rarely accessed materials, supporting more effective access, exploration, and reuse of large and diverse digital library collections.

Keywords: digital libraries; item-based collaborative filtering; content-based filtering; personalized content



Academic Editor: Jose María Alvarez Rodríguez

Received: 11 February 2026

Revised: 11 March 2026

Accepted: 12 March 2026

Published: 13 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

In recent decades, digital transformation has profoundly reshaped the ways in which scientific and cultural heritage resources are preserved, disseminated, and accessed. The rapid growth of digital collections, together with the availability of broad remote access, has made digital libraries a cornerstone of scientific research, education, and public engagement [1]. Digital libraries are increasingly understood not as just digitized repositories, but as intelligent and adaptive knowledge management environments that integrate resources, infrastructure, and services to support diverse user needs [2,3]. They provide centralized and sustainable access to content of different formats and levels of structure, including texts, images, multimedia objects, and data [4–6].

However, the expanding scale and thematic diversity of digital cultural heritage collections introduce significant challenges related to information overload and hinder the effective content discovery of specific resources. Users are often confronted with large volumes of heterogeneous materials and limited explicit feedback. The continuous introduction of new or rarely accessed documents further complicates the discovery of relevant resources.

In response to these challenges, artificial intelligence-based personalization and intelligent content discovery approaches have gained increasing attention in digital library research [7,8]. In particular, hybrid approaches that combine collaborative filtering, content-based analysis, and complementary signals have been identified as a promising direction for addressing data sparsity, cold-start conditions, and popularity bias, while improving robustness and coverage [9–11]. By leveraging behavioral data and access logs, such approaches support more adaptive navigation and exploration of digital libraries, improve the visibility of underutilized content, and contribute to more effective use and exploitation of rich and diverse digital cultural heritage collections [12,13].

This evolution aims to overcome information overload and to provide personalized services that are key to improving resource discoverability and user satisfaction [14,15]. Creating personalized intelligent content is establishing itself as a key mechanism for adapting content to the individual needs, interests, and behavior of the user, that can lead to a significant increase in user engagement and satisfaction [16,17].

The goal of this paper is to propose and evaluate a hybrid approach for intelligent and personalized content discovery in digital libraries, designed to support effective user-oriented access in large and heterogeneous cultural heritage collections. The proposed approach integrates behavioral interaction data with semantic content relationships to improve discovery quality under sparse and evolving interaction conditions.

The following sections present an overview of current research directions in the field of personalized and intelligent content delivery in digital libraries. The related work section reviews recent advances and identifies existing research gaps. Subsequently, the proposed hybrid approach is described in detail, followed by the description and discussion of verification and validation experiments proving the adequacy and effectiveness of the approach.

2. Related Work

Research in the field of intelligent access and discovery mechanisms for digital libraries is based on classic approaches, initially established in e-commerce, where products are recommended based on previous user behavior, such as search and purchase history [18]. These principles are gradually being transferred and adapted to content management and learning systems, where two main methodological directions are emerging: content-based filtering and collaborative filtering.

Content-based filtering generates recommendations by analyzing the similarity between objects and the individual user profile, built based on previous interactions with the system [19]. The effectiveness of this approach is highly dependent on the availability of sufficient data for the specific user and on the quality of the details of the resources, often implemented through metadata and semantic links [20]. A significant limitation is the dependence on already known resources, which can lead to a narrow scope of recommendations, especially in the absence of explicit ratings [21,22]. In such cases, the system suggests formally similar but not necessarily relevant items, a shortcoming that can be partially offset by including ratings or implicit indicators of interest, such as time spent on a given resource.

Unlike this individually oriented approach, collaborative filtering is based on the collective behavior of multiple users and generates recommendations by identifying similarities between user profiles or between items [19,21]. Its main assumption is that users with similar preferences in the past are likely to have similar interests in the future. The literature distinguishes between two main variants: user-based collaborative filtering and item-based collaborative filtering [22]. The advantage of this approach is its ability to provide more diverse and personalized recommendations without requiring in-depth knowledge of the content of the items themselves.

Despite its widespread use, collaborative filtering faces a number of well-known limitations, among which sparse data and cold start problems dominate, significantly degrading the quality of recommendations and the user experience [23,24]. Existing techniques often fail to make effective use of implicit user actions and rely on a single data source, which limits their accuracy and adaptability [9,25]. These limitations are exacerbated in large-scale digital libraries, where processing and combining large volumes of data pose significant computational and architectural challenges [21]. Practical implementations, such as the ShareTEC system [26], demonstrate the need for asynchronous processing and separate storage of behavioral data to ensure timely and scalable recommendation delivery.

In response to these problems, there is growing interest in hybrid approaches that combine content-based and collaborative filtering, as well as multiple sources of information. However, there are divergent views in the literature on the optimal way to integrate the two approaches. While some studies emphasize the advantages of hybrid models in terms of the balance between accuracy and diversity, others highlight issues related to the scalability and interpretability of these solutions [10,15].

In this context, there is a clear research gap related to the systematic understanding of how access logs and user interaction data can be used to improve collaborative filtering algorithms based on items in order to overcome the limitations mentioned [6,9,27]. The inability to address this gap leads to reduced user engagement and inefficient use of available resources [15].

The conceptual framework underlying contemporary research views access logs as structured repositories of data on user access and interaction. When integrated with collaborative filtering based on items, they can significantly increase the relevance and personalization of recommendations [6,9,28]. This framework combines user behavior analysis, techniques for dealing with sparse data, and hybrid recommendation strategies with the overall goal of improving services in digital libraries [25,29]. The theoretical basis relies on principles from machine learning and information retrieval. These approaches emphasize the synergy between user-oriented artificial intelligence and adaptive filtering techniques [6,7].

In summary, the analysis of existing research outlines a clear trend in the development of systems that provide intelligent personalized content fitted for the needs of the users for digital libraries towards hybrid models that integrate access logs, item-based collaborative filtering, and content-based approaches to increase the degree of personalization, improve search accuracy, and enhance user engagement. Despite the progress when using heterogeneous data sources, behavioral analytics, ontologies, and advanced algorithms, several challenges remain unresolved. These include sparse data, cold-start situations, scalability, and ethical aspects of personalization. This necessitates more robust and interpretable architectures capable of overcoming the limitations of classical collaborative filtering methods.

Notable mentions of recent research in recommender systems are deep learning architectures such as transformer-based sequential recommendation models and graph neural network approaches [30,31]. Transformer-based models employ attention mechanisms to capture contextual relationships in sequential interaction data, while graph-based methods represent user-item interactions as structured networks and apply graph neural networks to learn complex relational patterns [30,31]. These approaches have demonstrated strong performance in large-scale recommendation environments with dense interaction data [30,31]. However, such models typically require substantial training data and computational resources. In digital library environments, where interaction data is often relatively sparse and collections evolve continuously, hybrid approaches that combine behavioral signals, semantic similarity, and popularity-based mechanisms remain a practical and effective solution.

In this context, the present study adopts a hybrid methodological approach that combines item-based collaborative filtering with a content-based component. The content part is constructed on the integration of global thematic proximity, local matches between documents, thematic distribution of content, and analysis of the factual context, including named entities such as personalities, places, and other significant objects. This combination aims to systematically address the limitations identified in the literature and achieve a more accurate, explainable, and practically applicable recommendation that conceptually and methodologically builds on contemporary hybrid paradigms in the field of digital libraries.

3. Proposed Approach

The digital library platform used in this study (National Library “Ivan Vazov”, Plovdiv [32]) is a web-based system designed for the management, organization, and access of heterogeneous digital cultural heritage resources. It supports the storage and description of diverse content types and provides advanced search, grouping, and analytical functionalities that facilitate access to large and complex collections [1]. Prior studies [1,8] have focused on enhancing content management and exploratory access through flexible organization and usage analysis, establishing a solid foundation for user-oriented interaction with digital cultural heritage content. Building on this foundation, the present work extends the platform’s capabilities by incorporating artificial intelligence-based personalized content discovery mechanisms, aiming to further improve access, usability, and adaptive interaction with rich and diverse cultural heritage collections [1].

In this context, previous works [33,34] have noted the advantages of hybrid personalization approaches, which combine multiple complementary techniques and outperform standalone methods that rely on a single source of evidence. Individual approaches often exhibit inherent limitations, particularly in environments characterized by sparse interaction data, heterogeneous content, and continuous collection growth. By contrast, hybrid methods can mitigate these limitations and provide more robust, scalable, and context-aware personalization.

In this regard, the proposed approach applies a hybrid AI-driven personalization framework that integrates three complementary sources of information: (1) item-based collaborative filtering on a sparse “user-document” matrix extracted from filtered access logs; (2) content transfer via a pre-computed document similarity matrix constructed from semantic representations, thematic profiles and a factual layer of named entities; and (3) adaptive contribution of global popularity in cases of sparse or missing history.

The architecture is divided into two layers. The asynchronous layer performs computationally intensive operations, while the interactive layer relies on precomputed structures to generate personalized recommendations with low latency. In the asynchronous layer, resource-intensive tasks such as document preprocessing, embedding generation, similarity matrix construction, clustering, and aggregation of interaction logs are executed offline and processed iteratively as new data becomes available. This design allows the system to incrementally update internal representations without interrupting the recommendation service. The interactive layer operates exclusively on these precomputed structures, including the document similarity matrix, the sparse user-document interaction matrix, and the global popularity vector. As a result, recommendation generation requires only lightweight sparse operations and nearest-neighbor retrieval, ensuring fast response times while maintaining scalability for growing digital library collections. The trade-off of this architecture is that the interactive component may temporarily operate on slightly outdated data between asynchronous update cycles; however, this compromise enables stable performance and low-latency recommendation delivery while still allowing the model to incorporate newly available data through periodic updates.

3.1. Content Layer

Within the proposed hybrid architecture, the content layer constitutes a core component responsible for capturing semantic relationships among documents. Central to this layer is a document similarity matrix, which provides a stable and content-driven signal that remains independent of user interaction or behavioral data. This component is particularly important in digital library environments where interaction data may be sparse, incomplete, or unavailable. Given the specific characteristics of the underlying corpus—namely, periodical publications with heterogeneous topics and temporal structure—the construction of the similarity matrix presents distinct challenges and requires tailored modeling choices.

The similarity matrix is constructed by combining two complementary sources of information that capture different aspects of document relatedness. One of the information sources used in the similarity model is the factual context of documents represented through named entities. Named entities such as persons, locations, institutions, and historically significant objects are automatically extracted from the document texts using a named entity recognition (NER) pipeline adapted for Bulgarian language content. The extraction process employs two pretrained NER models compatible with the Bulgarian language, available through the Hugging Face model repository. Because individual pretrained models did not provide sufficiently reliable entity extraction, their outputs are combined to improve coverage and robustness.

The extracted entities are stored as structured metadata associated with each resource in the digital library and are later used to compute named entity-level similarity between documents using the Jaccard index, which captures overlaps in factual references between texts. This named entity-based similarity forms one component of the overall document similarity model. In the current configuration, the named entity similarity contributes to the final similarity score through a weighted aggregation of the different similarity signals.

After named entities are extracted, the process continues with text normalization, including Bulgarian lemmatization, stop-word removal, synonym enrichment, and token-safe segmentation of long documents. The textual representation of documents is then generated using a multilingual transformer-based sentence embedding model (MiniLM) available through the Hugging Face repository. The model produces contextual embeddings that

capture semantic relationships between documents while simultaneously transforming the original text into compact fixed-size vector representations [35,36].

Because the embedding model already produces compact representations, additional dimensionality reduction techniques such as PCA or SVD are not required [37,38]. Avoiding these transformations is particularly beneficial in the proposed architecture, where documents are processed incrementally in batches during the offline preprocessing stage. Dimensionality reduction methods typically require recomputation over the full document collection whenever new data is incorporated. Moreover, as noted in [39], repeated dimensionality reduction on incrementally processed data may introduce systemic bias and degrade representation quality unless periodic retraining is performed, which is impractical for large-scale collections. By generating compact embeddings directly, the system reduces preprocessing complexity while maintaining stable semantic representations suitable for similarity computation.

After vectorizing the data, several complementary aspects of proximity are calculated and aggregated:

- Global semantic similarity—average cosine similarity between document embedding vectors.
- Local fragment similarity—a best-match measure capturing correspondences between document segments, particularly useful for long or heterogeneous periodical texts.
- Thematic similarity—derived from Fuzzy C-Means clustering applied to the document embedding space, where documents are represented by soft membership vectors across multiple thematic clusters.
- Factual similarity—based on overlaps between the extracted named entities.

The final similarity score between documents is obtained through a weighted linear aggregation of these components. The coefficients controlling the contribution of each similarity signal were determined empirically during the development of the model through an ablation study designed to evaluate the impact of the individual components. A detailed analysis of the parameterization of the similarity model is presented in a complementary study focusing on multi-component similarity modelling.

The resulting similarity values are organized into a symmetric similarity matrix with a stable shared document index, ensuring compatibility with other operational structures used in the recommendation process. The similarity matrix plays a dual role in the personalization mechanism. First, it enables content transfer, allowing a limited or sparse user history to be expanded with semantically related resources based on proximity to previously accessed documents. Second, it supports cold-start situations for new items: even without recorded interactions, a document can be recommended if it is semantically close to resources associated with user interests.

Additional benefits include improved robustness against popularity bias by introducing thematic diversity in mixed profiles and increased explainability, as recommendations can be traced to contributions from global semantics, local matches, thematic similarity, and named entities. The overall process of constructing the similarity matrix and combining the similarity components is illustrated in Figure 1.

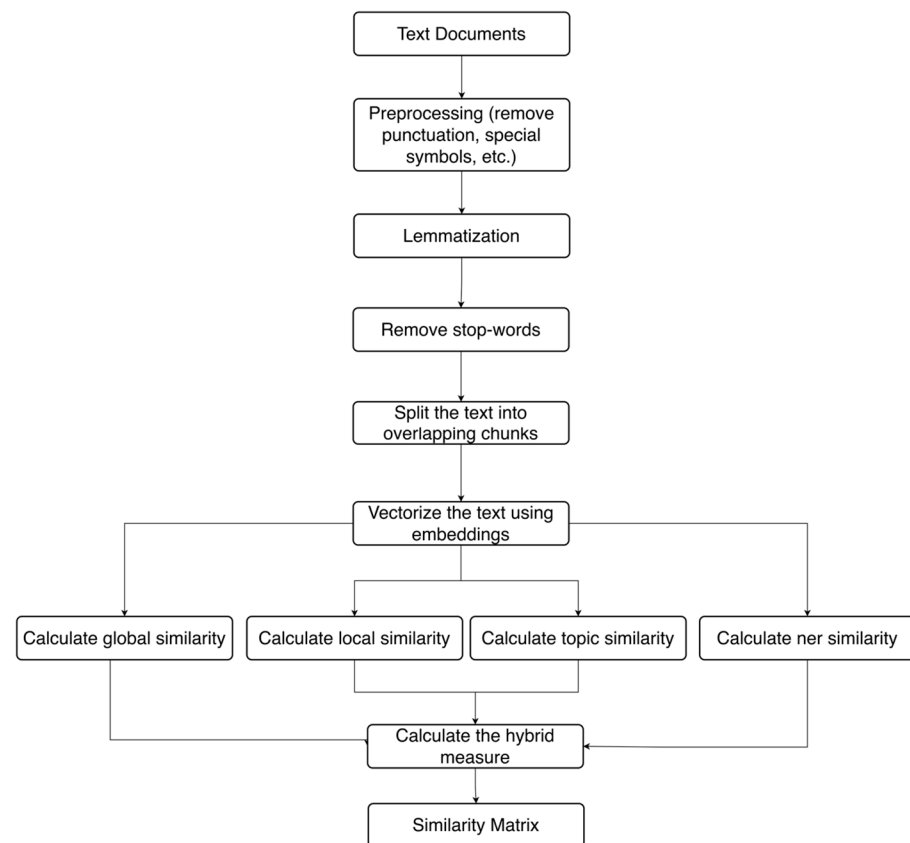


Figure 1. The process of creating similarity matrix for content layer.

3.2. Behavioral Layer

The behavioral layer represents the second core component of the proposed hybrid architecture and is responsible for modeling user interaction patterns within the digital library environment. This layer transforms resource access logs into a formal and inherently sparse representation of user preferences, capturing implicit behavioral signals that complement the content-based similarity and popularity indicators. By incorporating observed usage behavior, the behavioral layer contributes an adaptive signal that reflects evolving user interests and interaction dynamics.

From an architectural perspective, the behavioral layer is also organized according to a two-tier design. Computationally intensive operations, such as log processing, filtering, aggregation, and the construction of matrices and indexing structures, are executed asynchronously and outside the critical request-processing path. The interactive component operates exclusively on precomputed representations, enabling low-latency responses at runtime. This design reduces reliance on extensive individual interaction histories by leveraging the content signal, supports scalability through asynchronous computation, and enhances interpretability by maintaining a clear separation between content-based, behavioral, and popularity-driven contributions.

The analysis of interaction logs begins with filtering and cleaning. Administrative, system, and automated requests are removed, and the logs are normalized into a minimal and unambiguous format. The purpose of transforming them is to obtain a reliable and compact representation of actual behavior, which can be used as an implicit assessment of interest and serve as input to the hybrid model for generating personalized recommendations.

Due to the lack of explicit ratings, the behavioral signal is derived from the number of validated accesses of a “user-document” pair. To avoid oversensitivity to navigational repetitions and anomalies, the raw counts are converted into weights using a monotonically increasing but saturating function (single access—base weight; subsequent access—decreasing increment up to an upper bound). The resulting implicit weights reflect “confidence” in interest and are suitable for sparse linear algebraic treatment.

$$w_{u,i} = \begin{cases} 0, & c_{u,i} \leq 0 \\ \min(1 + 0.3 * (c_{u,i} - 1), 2), & c_{u,i} > 0 \end{cases} \quad (1)$$

Based on these weights, a sparse “user-document” matrix is created, where the non-zero elements are only the observed interactions. Due to its extreme sparsity, the matrix is stored in Compressed Sparse Row (CSR) format, which preserves only the non-zero values and their coordinates and allows for efficient multiplication, sorting, and incremental updates on standard hardware. Maps are maintained simultaneously for unambiguous mapping of user and document identifiers to matrix indices. The column index of the “user-document” matrix is synchronized with the index of the documents used by the similarity matrix, which ensures compatibility in hybrid calculations.

In parallel, a popularity vector is calculated that summarizes the audience’s interest in each document (including anonymous sessions) by aggregating the implicit weights and normalizing them in the interval [0, 1]. This signal is used adaptively as a “fallback mechanism” in cases of sparse history or anonymous users and stabilizes the ranking without dominating when the personal signal is sufficient.

The result of asynchronous processing (shown in Figure 2) includes three operational structures: (i) a sparse “user-document” matrix, (ii) a global popularity vector, and (iii) index maps for users and documents. These are published in compact file formats for fast loading from the interactive layer and are maintained incrementally: new interactions update the relevant rows and popularity components; adding/changing a document is reflected through the common index, without a complete recalculation.

This procedure ensures reproducibility, traceability, and operational stability with growing volume. In the interactive layer of the “personalized recommendations” functionality, the system aggregates three sources of information: (1) content transfer of history through the similarity matrix, (2) behavioral proximity, implicitly encoded in the sparse “user-document” matrix, and (3) adaptive contribution from the global popularity vector, active in sparse data. The evaluation is decomposable into components, which facilitates explainability (“why” a given resource is recommended) and ensures predictable behavior in edge cases: cold start for users (global popularity prevails, and the similarity matrix assists in the current context), exhausted history (smooth fallback to global popularity), thematic shift (a single new interaction acts as an “anchor” through the similarity matrix), poorly represented behavioral data (the content component dominates).

Built this way, the behavioral layer provides a compact, scalable, and explainable representation of actual interactions. Its consistency with the similarity matrix and document indexes allows for hybrid ranking with low latency and stability as data volume increases, while maintaining clarity in the rationale for each recommendation.

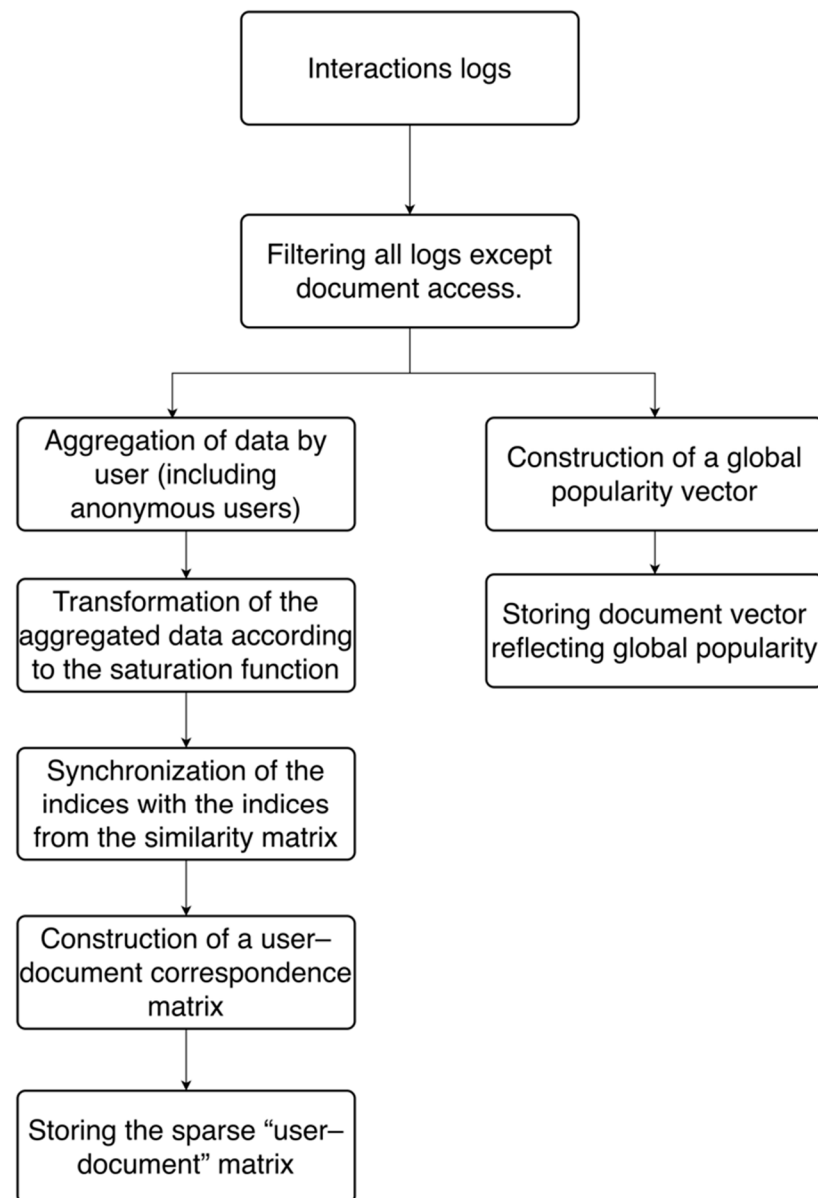


Figure 2. The process of creating document-user matrix and global popularity vector.

3.3. Integration of the Behavioral and Content Layer

The integration of content-based and behavioral signals is realized through a unified scoring mechanism that combines semantic proximity between documents, observed user interaction patterns, and global popularity indicators. This formulation corresponds to an item-based collaborative filtering approach augmented with content-based similarity, allowing complementary sources of evidence to be jointly exploited within a single personalization scheme. The resulting combination is formalized by the following formula:

$$\text{score}(u, d) = \sum_{i \in \text{history}(u)} w_{u,i} \times S(i, d) + \beta(u) \times p(d) \quad (2)$$

The popularity component $p(d)$ is not applied with a fixed global weight. Instead, its contribution is controlled through an adaptive coefficient $\beta(u)$, which determines the relative influence of popularity in the final score. The value of $\beta(u)$ depends on the amount of interaction data available for the user. When the total interaction weight $\sum_{i \in \text{history}(u)} w_{u,i}$ is below a predefined threshold H_{\min} , the popularity signal contributes to the recommendation score to stabilize the results in situations with limited interaction

history. As the interaction history grows and sufficient behavioral information becomes available, $\beta(u)$ is set to zero so that the recommendation score is determined entirely by the behavioral similarity component. In this way, the popularity signal functions primarily as a cold-start support mechanism rather than as a permanent influence on personalized recommendations. The parameters H_{\min} and the base coefficient $\beta(u)$ act as system configuration parameters that define when the popularity signal becomes active.

The mechanism for generating personalized recommendations according to the formula works in two main modes. The first mode is activated for users who have a history of interactions with the system, where the algorithm attempts to generate personalized suggestions that are close to the user's current history. The operational procedure includes the following steps:

1. A set of candidate documents is formed by aggregating the k -nearest neighbors of each document in the user's interaction history according to the similarity matrix. In the current configuration, $k = 5$, meaning that the five most similar documents are considered during candidate generation.
2. For each candidate document, a weighted score is calculated relative to the history of the viewed and similar documents.
3. Documents already viewed by the user are excluded.
4. The results are sorted by score, and the top k are returned.

In this mode, it is possible that the available history has exhausted the possible recommendations, in which case the fallback mechanism is automatically activated, and globally popular documents are displayed, which is explicitly noted.

The second mode of operation is for unregistered users or those without an accumulated interaction history, in which case the fallback mechanism is activated, and globally popular documents are displayed. Those cases are displayed in Figure 3.

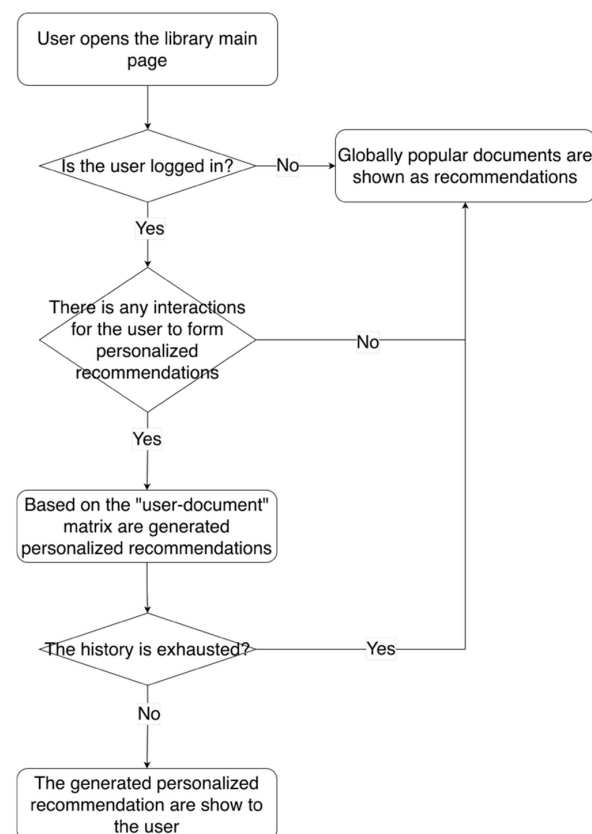


Figure 3. Basic flow for displaying personalized recommendations.

3.4. System Implementation

The implementation of the proposed hybrid recommendation framework follows the architecture described in Sections 3.1–3.3. Document texts undergo normalization, including Bulgarian lemmatization, stop-word removal, synonym enrichment, and token-safe segmentation of long documents prior to embedding generation. Semantic representations are produced using a multilingual transformer-based sentence embedding model available through the Hugging Face ecosystem.

User interaction signals are derived from filtered access logs and transformed into implicit preference weights according to Equation (1). The resulting sparse user–document interaction matrix is stored in Compressed Sparse Row (CSR) format. During recommendation generation, candidate documents are obtained from the k -nearest neighbors of the user’s interaction history, and final ranking scores are computed using the hybrid scoring function defined in Equation (2). Configuration parameters used in the similarity model and recommendation process were determined empirically during system development and remain fixed across all experiments reported in this study.

From a computational perspective, the architecture separates resource-intensive preprocessing from the interactive recommendation stage. Operations such as embedding generation, similarity matrix construction, clustering, and log aggregation are executed asynchronously in an offline layer and processed in batches. At runtime, the interactive component operates exclusively on precomputed structures, including the document similarity matrix, the sparse user-document matrix, and the global popularity vector. Recommendation generation therefore consists primarily of lightweight sparse vector operations and nearest-neighbor retrieval.

In terms of algorithmic complexity, the offline preprocessing stage includes document embedding generation and similarity matrix construction. For a collection of N documents with embedding dimension d , embedding generation scales approximately as $O(N \cdot d)$, while pairwise similarity computation for the similarity matrix has theoretical complexity $O(N^2)$. Processing of interaction logs and construction of the user-document matrix scale linearly with the number of log events E , resulting in complexity $O(E)$. Since these operations are executed asynchronously and updated periodically, they do not affect the runtime performance of the recommendation service.

Empirical measurements were conducted using the evaluation dataset consisting of 500 documents, 870 users, and 26,000 interaction events. The most computationally intensive preprocessing step is the extraction of named entities from document texts. Using the NER pipeline described in Section 3.1, processing 500 documents required approximately 20 min (≈ 2.46 s per document). However, this operation is performed only once for each document during ingestion and is required only for newly added resources. Subsequent similarity computations reuse the extracted entities and therefore do not repeat the full NER processing.

The remaining stages of the offline preprocessing pipeline, including semantic embedding generation, clustering, and construction of the document similarity matrix, required approximately 4 min for the evaluation dataset (500 documents). The resulting similarity matrix and associated semantic vectors occupy approximately 1–2 MB of storage in this configuration, thus have a small memory footprint. Aggregation of interaction logs and construction of the sparse user-document interaction matrix required approximately 0.15 s for the evaluation dataset (500 documents). A larger-scale test was performed, processing 5,000,000 log events, and the test completed in 15.8 s, demonstrating stable performance under significantly larger workloads. Due to the use of sparse matrix representations, the resulting interaction structures require only a small fraction of the storage required by dense matrices.

Because candidate generation uses a fixed number of nearest neighbors, the run-time complexity per recommendation request grows primarily with the size of the user interaction history rather than with the size of the full document collection. In practice, recommendation generation involves only sparse vector operations and nearest-neighbor retrieval, with observed response times below 50 ms per request. This architectural separation between offline computation and lightweight online inference enables the system to scale to larger digital library collections while maintaining low-latency recommendation responses.

4. Experimental Setup, Results and Discussion

This section presents the experimental setup and discusses the results obtained from a series of evaluation scenarios designed to assess the behavior and robustness of the proposed approach. Both typical usage conditions and critical edge cases are examined to verify that the system responds in a stable and predictable manner across a range of realistic interaction patterns.

The formulation embodies a hybrid approach by jointly exploiting complementary signals: content-based similarity ensures semantic relevance and covers new or rarely visited documents; behavioral signals personalize results based on user history; and popularity provides a stable fallback in cold-start or sparse-data scenarios without overriding strong personal signals.

The evaluation is organized into two complementary parts. First, a quantitative evaluation based on real interaction logs is conducted to measure recommendation performance under normal system usage. This is followed by a set of controlled synthetic scenarios designed to examine the behavior of the hybrid model under specific edge cases that are difficult to systematically reproduce in real datasets.

4.1. Quantitative Evaluation Using Real Interaction Logs

The quantitative evaluation of the hybrid approach and the comparison with pure collaborative filtering and content-based filtering were conducted using interaction logs collected from the digital library platform of the National Library “Ivan Vazov” in Plovdiv. The logs record document access events generated during normal system operation and include the document identifier, the user identifier (when available; anonymous interactions do not contain a user identifier), and the timestamp of the access. Each record corresponds to an event (login, access of resource, add resource, delete resource, etc.) and represents an implicit interaction between a user and a specific document or the platform.

For the purposes of controlled experimentation, the evaluation was performed on a representative subset of the digital library collection. In preliminary experiments, several document subsets ranging between 200 and 1000 documents were used to analyze the behavior of the recommendation methods under different collection sizes and to assess the impact of collection size on document similarity calculation and similarity matrix construction.

In the evaluation presented in this study, a subset of 500 documents was selected from the digital library collection, and the interaction logs were filtered to include only user interactions (access of resources) associated with these documents. The subset was selected in order to enable controlled experimentation and efficient similarity computation, while preserving realistic interaction patterns derived from the operational system logs.

It should be noted that restricting the candidate space to a document subset may slightly influence the absolute values of ranking metrics, since the recommendation task is performed over a smaller item set. However, in highly sparse digital library interaction datasets, controlled subsets are frequently used in order to maintain sufficient interaction

density for meaningful evaluation. In this case, the selected subset preserves the original user behavior patterns while enabling reproducible experimentation and stable comparison between the evaluated recommendation approaches.

The dataset used for quantitative evaluation contains 870 users, approximately 12,000 interaction events, and 500 documents. As is typical for digital library environments, the resulting user-document interaction matrix is highly sparse because each user interacts with only a small fraction of the available resources. Such sparsity is characteristic of real digital library interaction datasets and represents a realistic condition for evaluating recommendation systems in cultural heritage collections. Anonymous access events were excluded from the quantitative evaluation because the evaluation protocol requires identifiable user histories to construct user profiles and generate recommendations.

To assess the recommendation quality of the proposed approach and the baseline methods, a leave-one-out evaluation protocol was applied, which is commonly used in recommender system research. For each user with multiple interactions, the most recent interaction was reserved as the test instance. The remaining interactions were used to construct the user profile. The recommender system then generated a ranked list of candidate documents for each user while excluding the documents already present in the user's interaction history.

Three recommendation approaches were evaluated:

- Collaborative filtering (CF) based on item-based similarity derived from the user-document interaction matrix.
- Content-based (CB) based on the semantic similarity matrix constructed from document representations.
- Hybrid approach, corresponding to the proposed approach that integrates collaborative filtering signals, semantic document similarity, and the global popularity component.

The evaluation focuses on classical collaborative filtering and content-based baselines because the objective of this work is to analyze the behavior of the proposed hybrid integration of behavioral and semantic signals in digital library environments. More complex recommendation models typically require substantially larger interaction datasets and different training assumptions, which fall outside the scope of the present system-oriented study.

For each user, the system generated a ranked list of the top 10 recommended documents, and the recommendation quality was evaluated by comparing the list with the held-out interaction from the test set.

Recommendation performance was assessed using three widely used ranking metrics in recommender system evaluation. Hit Rate at 10 (HR@10) [21] measures whether the relevant document appears among the top 10 recommended items; if the held-out document is present in the recommendation list, the recommendation is considered successful. Precision at 10 (Precision@10) [21] measures the proportion of relevant items among the top 10 recommendations and reflects how frequently relevant documents appear within the recommendation list. Normalized Discounted Cumulative Gain at 10 (NDCG@10) evaluates the ranking quality of the recommendation list by assigning higher scores when relevant documents appear closer to the top of the ranked list [40]. These metrics are commonly used for evaluating recommender systems operating on implicit feedback data [12,21,40].

The quantitative results obtained from the evaluation on real interaction logs are summarized in Table 1.

Table 1. Quantitative evaluation results for collaborative filtering, content-based, and hybrid approaches on real digital library interaction logs.

Model	Precision@10	HR@10	NDCG@10
Collaborative Filtering	0.0306	0.3057	0.1775
Content-Based	0.0051	0.0506	0.0237
Hybrid (Proposed Approach)	0.0321	0.3207	0.1888

The results show that the hybrid approach achieves the slightly higher values across all evaluated ranking metrics than the other approaches. Although the improvement over the collaborative filtering baseline is modest, it is consistent across Precision@10, HR@10, and NDCG@10. This behavior reflects the complementary nature of the hybrid formulation, where behavioral signals capture collective usage patterns while semantic similarity contributes additional contextual information. The content-based method alone shows considerably lower performance, indicating that semantic similarity by itself does not sufficiently capture the collective reading patterns reflected in user interactions. Collaborative filtering captures these behavioral patterns but can be affected by sparse interaction histories. By combining both sources of information, the hybrid model maintains the strengths of collaborative filtering while improving robustness in sparse interaction scenarios.

While evaluation on real interaction logs provides quantitative evidence of recommendation performance under normal usage conditions, some behaviors of recommender systems are difficult to observe using real datasets alone. Situations such as cold-start users, isolated thematic clusters, abrupt shifts in user interests, or the introduction of new documents without interaction history occur relatively infrequently and cannot always be systematically reproduced in real usage logs.

For this reason, additional controlled synthetic evaluation scenarios were constructed. These scenarios make it possible to examine how the proposed hybrid approach behaves under well-defined conditions that represent common edge cases in digital library environments. The following subsection therefore focuses on analyzing these scenarios to illustrate the situations where the hybrid formulation provides its main advantages.

4.2. Synthetic Evaluation Scenarios for Edge Cases

To validate the correct behavior of the hybrid approach and to isolate random influences from real data, a synthetic set with three thematic groups of documents (clusters A, B, and isolated C) has been constructed. A strong co-occurrence was deliberately introduced between documents in clusters A and B, meaning that they are read together by the same users. Cluster C remains without co-occurrences in order to observe system behavior when no neighbors exist. A document with a high number of anonymous visits was also added, which functions as a “globally popular” reference point in the absence of interaction history. In all experiments, the standard restriction applies that text documents already viewed by a given user are excluded from the possible recommendations. The resulting lists are compared with the expected behavior according to the scenario.

The following scenarios were tested and demonstrated:

1. “Cold start” for a user. The profile represents a new or anonymous visitor with no registered interaction history; its vector in the “user-document” matrix is zero. The resilience of the system in the complete absence of personal signals, as well as its ability to deliver useful and consistent results, is evaluated. This scenario is critical in real-world settings, where a substantial proportion of users are new or unregistered, and empty result lists negatively affect trust and retention. The expected and confirmed result is a switch to a fallback strategy based on global popularity

without simulating personalization; recommendations are clearly marked as “popular” rather than “personal,” thus maintaining transparency and basic usefulness until there are enough personal signals to form personalized recommendations.

2. Exhausted isolated cluster. The user has consumed all available documents in a closed, loosely connected topic (e.g., cluster B) that has no joint readings with other topics. The test checks whether the model correctly recognizes a “semantic dead end” and avoids trivial or cyclical suggestions of already seen content. The scenario is critical for systems with thematic “pockets” where the joint signal does not lead to new neighbors. The confirmed expectation is that the system recognizes the lack of valid personalized candidates and triggers a controlled fallback mechanism (global popularity) by offering relevant but non-personalized alternatives outside the niche, without “inventing” personalization.
3. Shift in interest. The user has a dominant history in one topic (e.g., cluster A) but performs a single, new interaction in a different topic (cluster B). Here, the model’s sensitivity to new signals and its ability to reorient recommendations without ignoring accumulated history are tested. The scenario is important because real interests are dynamic and require rapid adaptation during thematic transitions. The expected result is for the single signal to act as an “anchor”: the hybrid approach brings out candidates from the new topic through content proximity and available joint readings, while limiting the dominance of the old topic; the list is filled mainly with resources from cluster B.
4. Mixed profile. The user has a balanced history distributed between two or more independent thematic clusters (e.g., A and B). The aggregating behavior of the model is tested: that the contributions from different interests are combined proportionally, without one cluster suppressing another and without “leakage” to unrelated topics. The scenario is important for maintaining diversity and avoiding the “tunnel effect.” The result is a balanced list of recommendations in which the relative share of elements from A and B reflects the strength of historical signals; the system demonstrates stable mixing of content and collaborative contributions and be resistant to asymmetric but non-priority variations.
5. “Cold start” for items. The synthetic data that was generated also includes a test resource that is a new document with no observed readers and, therefore, no collaborative links. The goal is to verify whether the content component of the hybrid approach overcomes the lack of collaborative readings and makes the document visible to an audience with similar thematic preferences. The scenario is key to introducing new resources and reducing their time to being offered as recommended items. The result is that the new document will appear in the personalized list of recommendations for users which previously read texts that are semantically close, with a moderately lower initial weight. A control check with the purely collaborative method shows the absence of this document in the recommendations.
6. Scarce behavioral history. The profile includes interactions with rare, lightly read resources that have few shared readers; the collaborative signal is weak or unstable. The test measures the robustness of the model under conditions of sparse data and the ability of the content layer to provide semantic relevance. This scenario is particularly relevant for institutional collections with a large number of niche resources and heterogeneous consumption patterns. The expected result is a leading role for content proximity (the similarity matrix) in forming recommendations that remain thematically consistent; in the purely collaborative control variant, the fallback mechanism is activated, and popular but less thematically close elements dominate.

In summary, these six scenarios cover the full spectrum of main cases especially the edge cases—lack or exhaustion of personal data, dynamics and competition of interests, lack of co-readings for new or rare resources—and validate that the hybrid model responds predictably, transparently, and stably in conditions typical of digital libraries.

The evaluation summarizes two types of checks: (1) qualitative—whether the lists match the expected logic of the scenario, and (2) quantitative—share of results in the first k suggestions and relative contribution of the two clusters in a mixed profile. Thus, the test results demonstrate that the module generates appropriate personalized recommendations, remains useful in the absence of data, and reacts predictably when interests change.

5. Conclusions and Future Work

The growing volume and diversity of digitized resources in modern digital libraries make it increasingly difficult for users to locate relevant documents within large cultural heritage collections. Addressing this challenge, a hybrid approach for intelligent and personalized interaction was presented, aimed at improving content discovery and user-oriented access in digital library environments. The approach integrates multiple sources of information to support effective navigation and discovery, particularly in the presence of sparse interaction data and continuously expanding collections, while remaining robust and scalable for practical use.

Empirical evaluation using real interaction logs and scenario-driven synthetic tests shows that the hybrid approach produces stable recommendation behavior and achieves modest but consistent improvements in ranking performance compared with the individual collaborative filtering and content-based baselines. The explicit separation of content, behavior, and popularity also enhances explainability: each recommendation can be traced to its constituent contributions, supporting user trust and explainability in production environments.

Future work will enrich interaction logs with signals such as time spent and depth of access, enabling the model to distinguish strong interest from weak or negative interactions rather than treating all accesses as implicitly positive. While incremental model updates are already supported, future work will also explore temporal weighting of interactions to better reflect evolving user interests. At the system level, approximate nearest-neighbor indices will be introduced to further reduce latency and infrastructure costs.

Author Contributions: Conceptualization, investigation and methodology, E.M. and D.P.-M.; Create software, E.M. and V.G.; Validation, E.M. and A.N.; writing—original draft, E.M. and D.P.-M.; Manuscript review and contribution to final version, E.M., D.P.-M., V.G., A.N. and R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded partly by CLaDA-BG, the Bulgarian National Interdisciplinary Research e-Infrastructure for Resources and Technologies in favour of the Bulgarian Language and Cultural Heritage, part of the EU infrastructures CLARIN and DARIAH, <https://clada-bg.eu/bg/> (accessed on 21 January 2026), Grant number D01-97/26.06.2025, financed by the Bulgarian Ministry of Education and Science, Funding Procedure: The National Science Infrastructure Roadmap 2020–2027 (NSIF).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The digital periodicals and interactions logs were provided by National Plovdiv Library “Ivan Vazov” digital library.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Goynov, M.; Luchev, D.; Paneva-Marinova, D.; Pavlov, R.; Rangochev, K. Towards Providing Analytical Services in the Web-Based Platform for Intelligent Cultural Content Management CultIS. *TEM J.* **2025**, *14*, 2946–2952. [\[CrossRef\]](#)
- Liu, Z.; Shao, B. A systematic review of library services platforms research and research agenda. *Libr. Inf. Sci. Res.* **2024**, *46*, 101325. [\[CrossRef\]](#)
- Borgman, C.L. Libraries, Digital Libraries, and Data: Forty years, Four Challenges. *Portal Libr. Acad.* **2025**, *25*, 39–58. [\[CrossRef\]](#)
- Fekadu, M.; Alemneh, D. Digital Library Models: A Systematic Review. In *Sustainability and Empowerment in the Context of Digital Libraries: 26th International Conference on Asia-Pacific Digital Libraries (December 2024)*; Springer: Berlin/Heidelberg, Germany, 2024. [\[CrossRef\]](#)
- Owusu-Ansah, C.R.A. Digital Information and Library Services in ODDE. In *Handbook of Open, Distance and Digital Education*; Springer: Singapore, 2022. [\[CrossRef\]](#)
- Troussas, C.; Krouska, A.; Koliarakis, A.; Sgouropoulou, C. Harnessing the power of user-centric artificial intelligence: Customized recommendations and personalization in hybrid recommender systems. *Computers* **2023**, *12*, 109. [\[CrossRef\]](#)
- Adewojo, A.A.; Dunmade, A.O. From big data to intelligent libraries: Leveraging analytics for enhanced user experiences. *Bus. Inf. Rev.* **2024**, *41*, 104–109. [\[CrossRef\]](#)
- Goynov, M.; Luchev, D.; Paneva-Marinova, D.; Senka, G.; Rangochev, K.; Pavlova, L.; Pavlov, R.; Zlatkov, L. CultIS: Web-based Platform for Intelligent Cultural Content Management. *Digit. Present. Preserv. Cult. Sci. Herit.* **2024**, *14*, 19–36. [\[CrossRef\]](#)
- Koliarakis, A.; Krouska, A.; Troussas, C.; Sgouropoulou, C. Modified collaborative filtering for hybrid recommender systems and personalized search: The case of digital library. In Proceedings of the 2022 17th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP), Corfu, Greece, 3–4 November 2022; pp. 1–6. [\[CrossRef\]](#)
- Tursunov, T.; Kaibassova, D.; Kashkimbayeva, N. Comparative Analysis of Recommendation Algorithms: Collaborative, Content-Based and Hybrid Approaches. In Proceedings of the 2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, 14–16 May 2025; pp. 1–5. [\[CrossRef\]](#)
- Yuan, H.; Hernandez, A.A. User cold start problem in recommendation systems: A systematic review. *IEEE Access* **2021**, *11*, 136958–136977. [\[CrossRef\]](#)
- Fayyaz, Z.; Ebrahimian, M.; Nawara, D.; Ibrahim, A. Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *Appl. Sci.* **2020**, *10*, 7748. [\[CrossRef\]](#)
- Liqiang, H.; Quan, L. Design of Resource Recommendation Model for Personalized Learning in the Era of Big Data. In Proceedings of the 2019 Annual Meeting on Management Engineering in Annual Meeting on Management Engineering, Kuala Lumpur, Malaysia, 8–10 December 2019; pp. 181–187. [\[CrossRef\]](#)
- Rana, P.; Singh, D. Navigating the information labyrinth: A smart search system for libraries. *Libr. Hi Tech News* **2025**, *42*, 6–10. [\[CrossRef\]](#)
- Kumar, V.A.; Chidambaram, M. Personalization and User Behavior Analysis in Digital Libraries: A Systematic Review. *Acad. Res. J. Sci. Technol. (ARJST)* **2025**, *2*, 37–43. [\[CrossRef\]](#)
- Speciale, A.; Vallero, G.; Vassio, L.; Mellia, M. Recommendation systems in libraries: An application with heterogeneous data sources. *arXiv* **2023**, arXiv:2303.11746. [\[CrossRef\]](#)
- Liang, Y.; Wang, J. Intelligent library recommendation based on GNN and attention networks. *J. Comput. Methods Sci. Eng.* **2025**, 14727978251364472. [\[CrossRef\]](#)
- Liao, M.; Sundar, S.S.; Walther, J. User trust in recommendation systems: A comparison of content-based, collaborative and demographic filtering. In Proceedings of the CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April–5 May 2022. [\[CrossRef\]](#)
- Liao, M.; Sundar, S.S. When e-commerce personalization systems show and tell: Investigating the relative persuasive appeal of content-based versus collaborative filtering. *J. Advert.* **2022**, *51*, 1–16. [\[CrossRef\]](#)
- Stoikov, J. Using Conditional Probability for Discovering Semantic Relationships between Named Entities in Cultural Heritage Data. *Digit. Present. Preserv. Cult. Sci. Herit.* **2021**, *11*, 77–88. [\[CrossRef\]](#)
- Schafer, J.B.; Frankowski, D.; Herlocker, J.; Shen, S. Collaborative filtering recommender systems. The adaptive web: Methods and strategies of web personalization. In *The Adaptive Web. Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4321, pp. 291–324. [\[CrossRef\]](#)
- Kapembe, S.; Quenum, J.G. A Personalised Hybrid Learning Object Recommender System. In Proceedings of the 11th International Conference on Management of Digital EcoSystems, Limassol, Cyprus, 12–14 November 2019; pp. 242–249. [\[CrossRef\]](#)
- Ibrahim, I.I.; Heryadi, Y.; Qomariyah, N.N.; Budiharto, W. Improving Library Recommender Systems with a Hybrid Interaction-ALS Approach under Data Sparsity. In Proceedings of the 2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA), Antalya, Turkiye, 7–9 August 2025; pp. 1–6. [\[CrossRef\]](#)

24. Pan, T. Personalized Recommendation Service in University Libraries using Hybrid Collaborative Filtering Recommendation System. In Proceedings of the 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), Hassan, India, 23–24 August 2024; pp. 1–6. [\[CrossRef\]](#)
25. Guo, P.; Nasir, M.K.M.; Xu, Y. Collaborative Filtering Recommender System for Online Learning Resources with Integrated Dynamic Time Weighting and Trust Value Calculation. *TEM J.* **2024**, *13*, 1352–1361. [\[CrossRef\]](#)
26. Stefanov, K.; Boychev, P.; Stefanova, E.; Georgiev, A. Digital Libraries in Teacher Education. In Proceedings of the Fortieth Jubilee Spring Conference of the Union of Bulgarian Mathematicians, Borovetz, Bulgaria, 5–9 April 2011.
27. Surekha, A.; Gouni, R.; Gorripati, S.K.; Rachapudi, V.; Devi, S.A.; Angadi, A. A Comprehensive Hybrid Implicit and Explicit Item-Based Collaborative Filtering Approach with Bayesian Personalized Ranking for Enhancing Book Recommendations. In *Smart Factories for Industry 5.0 Transformation*; Scrivener Publishing LLC: Beverly, MA, USA, 2025; pp. 89–104. [\[CrossRef\]](#)
28. Verma, M.; Rawal, A. An enhanced item-based collaborative filtering approach for book recommender system design. *ECS Trans.* **2022**, *107*, 15439–15449. [\[CrossRef\]](#)
29. Ifada, N.; Rahmatullah, A.; Rachman, F.H. Leveraging Hybrid Semantic Ontology-based Model for Library Book Recommendation System. In Proceedings of the 2024 9th International Conference on Information Technology and Digital Applications (ICITDA), Nilai, Negeri Sembilan, Malaysia, 7–8 November 2024; pp. 1–6. [\[CrossRef\]](#)
30. Sun, F.; Liu, J.; Wu, J.; Pei, C.; Lin, X.; Ou, W.; Jiang, P. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM 2019), New York, NY, USA, 3–7 November 2019; pp. 1441–1450. [\[CrossRef\]](#)
31. Ying, R.; He, R.; Chen, K.; Eksombatchai, P.; Hamilton, W.L.; Leskovec, J. Graph Convolutional Neural Networks for Web-Scale Recommender Systems. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2018), London, UK, 19–23 August 2018; pp. 974–983. [\[CrossRef\]](#)
32. National Library “Ivan Vazov” Plovdiv. Available online: <https://digital.libplovdiv.com/bg> (accessed on 21 January 2026).
33. Christozov, D.; Mitreva, E. Trust in Learning from Big Data: The Two Sides of the Same Coin. *Issues Inf. Syst.* **2020**, *21*, 147–152. [\[CrossRef\]](#)
34. Mitreva, E.; Nikolova, A.; Georgiev, V.; Gigova, A. Personalization approaches for cultural heritage study. *Digit. Present. Preserv. Cult. Sci. Herit. Conf. Proc.* **2023**, *13*, 181–188. [\[CrossRef\]](#)
35. Alkaabi, H.; Jasim, A.K.; Darroudi, A. From Static to Contextual: A Survey of Embedding Advances in NLP. *PERFECT J. Smart Algorithms* **2025**, *2*, 57–66. [\[CrossRef\]](#)
36. Das, K.; Kamlash; Abid, F. Advancements in Word Embeddings: A Comprehensive Survey and Analysis. *Proc. Pak. Acad. Sci. A Phys. Comput. Sci.* **2024**, *61*, 227–245. [\[CrossRef\]](#)
37. Ashraf, M.; Anowar, F.; Setu, J.H.; Chowdhury, A.I.; Ahmed, E.; Islam, A. A Survey on Dimensionality Reduction Techniques for Time-Series Data. *IEEE Access* **2023**, *11*, 42909–42923. [\[CrossRef\]](#)
38. Chang, Y.-C.I. A Survey: Potential Dimensionality Reduction Methods. *arXiv* **2025**, arXiv:2502.11036. [\[CrossRef\]](#)
39. Shahzad, M.; Barzamini, H.; Wilson, J.; Alhoori, H.; Rahimi, M. Dynamic domain analysis for predicting concept drift in engineering AI-enabled software. *J. Data Inf. Sci.* **2025**, *10*, 124–151. [\[CrossRef\]](#)
40. Järvelin, K.; Kekäläinen, J. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst. (TOIS)* **2002**, *20*, 422–446. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.