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

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

Posted Date: 14 February 2026

doi: 10.20944/preprints202602.1144.v1

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

Article

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

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), 1113 Sofia, Bulgaria, 8, Acad. G. Bonchev Str.

[2] America University in Bulgaria, 2700 Blagoevgrad, Bulgaria, 1, Georgi Izmirliev Sq.

* 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, as users increasingly struggle to navigate and identify relevant materials in periodical collections. 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 to these challenges; however, digital library environments are often characterized by sparse interaction data, evolving user interests, and the continuous introduction of new resources, which 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

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 [2–4]. They provide centralized and sustainable access to content of different formats and levels of structure, including texts, images, multimedia objects, and data [5–7].

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, limited explicit feedback, and the continuous introduction of new or rarely accessed documents, which complicates the discovery of most relevant resources.

In response to these challenges, artificial intelligence–based personalization and intelligent content discovery approaches have gained increasing attention in digital library research [8–10]. In particular, hybrid approaches that combine collaborative filtering, content-based analysis, and

 doi:10.20944/preprints202602.1144.v1

complementary signals have been identified as a promising direction for addressing data sparsity, cold-start conditions, and popularity bias, while improving robustness and coverage [11–13]. 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 [14,15].

This evolution aims to overcome information overload and to provide personalized services that are key to improving resource discoverability and user satisfaction [16,17]. 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 of user engagement and satisfaction [18,19].

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 [20]. 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 [21]. 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 [22]. 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 [23,24]. 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 [21,23]. 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 [24]. 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 [25,26]. 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 [11,27]. These limitations are exacerbated in large-scale digital libraries, where processing and combining large volumes of data pose significant computational and architectural challenges [23]. Practical implementations, such as the ShareTEC system [28], demonstrate the need for asynchronous

 doi:10.20944/preprints202602.1144.v1

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 [12,17].

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 [8,11,29]. The inability to address this gap leads to reduced user engagement and inefficient use of available resources [17].

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 [8,11,30]. 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 [27,31]. The theoretical basis relying on the principles of machine learning and information retrieval, emphasizes the synergy between user-oriented artificial intelligence and adaptive filtering techniques [8,9].

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 reported using heterogeneous data sources, behavioral analytics, ontologies, and sophisticated algorithmic solutions, a few challenges remain unresolved, among which the problems of sparse data, "cold start," scalability, and the ethical aspects of personalization. This necessitates more robust and interpretable architectures capable of overcoming the limitations of classical collaborative filtering methods.

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 [1,7,10] 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 [2].

In this context previous work [33–35] has noted the advantages of hybrid personalization approaches, which combine multiple complementary techniques, which outperform standalone

 doi:10.20944/preprints202602.1144.v1

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 an asynchronous layer that performs computationally intensive operations and interactive layer that steps on the pre-calculated operational structures and generates personal recommendations with low latency and short response times.

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.

Its construction begins with text normalization (lemmatization for Bulgarian text, removal of stop words, and synonym enrichment) and token-safe segmentation of long documents. Vector representations are derived through a multilingual transformer model, which aims to do two things. On the one hand, it vectorizes the data in a way that preserves the semantics — this is achieved by using contextual embeddings [36,37]. On the other hand, the model also performs data reduction, which eliminates the need for an additional approach to dimensionality reduction [38,39]. Given the substantial volume of data, the texts are processed in batches. Consequently, eliminating the need for dimensionality reduction constitutes a key advantage, as it lowers computational costs. Moreover, as noted in [40], incremental document processing can introduce systemic bias and lead to the degradation of representation quality unless periodic retraining is conducted which considering the amount of data is nearly impossible.

After vectorizing the data, several complementary aspects of proximity are calculated and aggregated: (1) average cosine proximity between document vectors (global semantics), (2) "best-match" between fragments (local matches, relevant for long/short texts and especially useful for periodicals) and (3) thematic profiles from Fuzzy C-Means, which capture the distributed membership to topics. In parallel, a sparse representation of the named entities, extracted from the texts and stored as metadata, is constructed, on which the Jaccard index is evaluated. The resulting components are integrated into a symmetric matrix with a stable shared index, which ensures compatibility with other operational structures. The role of the similarity matrix in personalization is twofold. On the one hand, it allows for content transfer: a limited or "sparse" user history can be expanded to semantically similar resources through proximity assessments to already viewed documents. On the other hand, the matrix addresses the "cold start" for new items: even in the absence of interactions, a new document can be suggested if it is content-wise close to resources with established interest from a user. Additional benefits include stabilization against popularity bias (introducing thematic diversity in mixed profiles) and increased explainability by decomposing contributions into global semantics, local matches, topics, and named entities. The full process of creating the similarity matrix and combining the different components of the similarity measure is shown on Figure 1.

 doi:10.20944/preprints202602.1144.v1

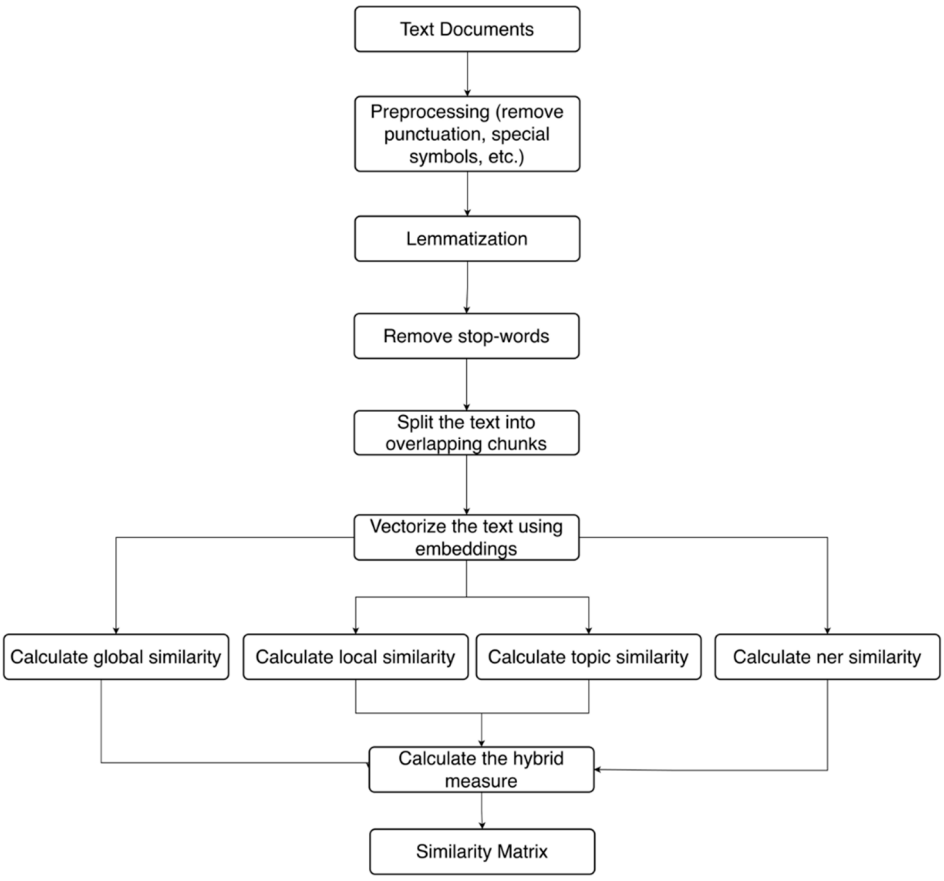


Figure 1. The process of creating similarity matrix for content layer.

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 process of analyzing the logs, generated from the interactions between the user and the system, starts with filtering and cleaning the logs from administrative, system, and automated requests and normalizing to a minimal, unambiguously interpretable 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

 doi:10.20944/preprints202602.1144.v1

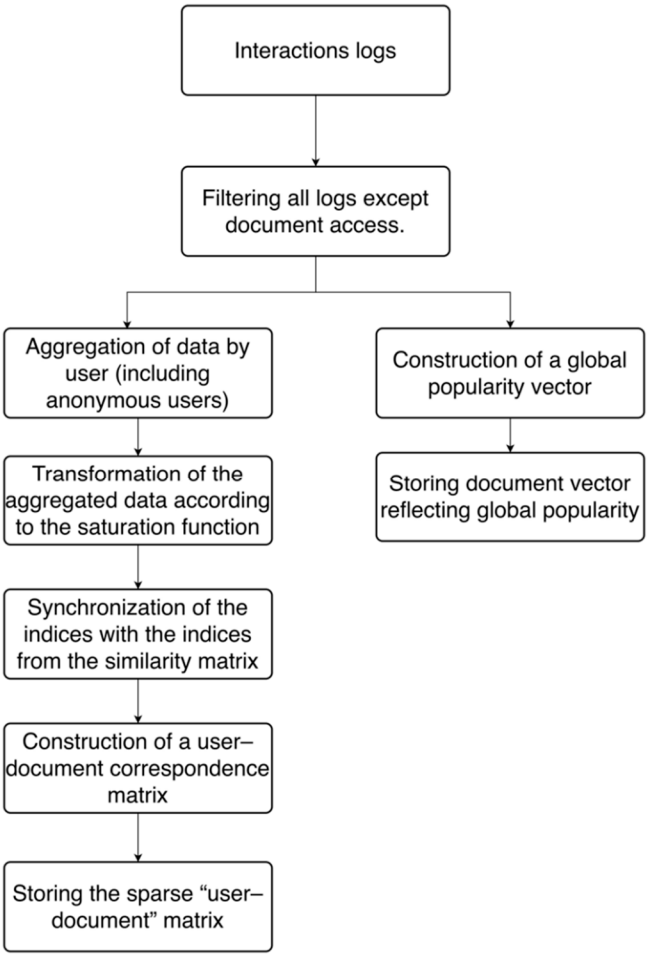
function (single access - base weight; subsequent access - decreasing increment up to an upper bound). The resulting implicit weights reflect "confidence" in interest and are suitable for sparse linear algebraic treatment.

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

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 row 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.



 doi:10.20944/preprints202602.1144.v1

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

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 consumption. 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.

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:

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

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. 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. If this is not the case, the operational procedure includes the following steps:

1. A set of candidates is formed by combining the k-nearest neighbors according to the similarity matrix of each document in the user's history.
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.

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.

 doi:10.20944/preprints202602.1144.v1

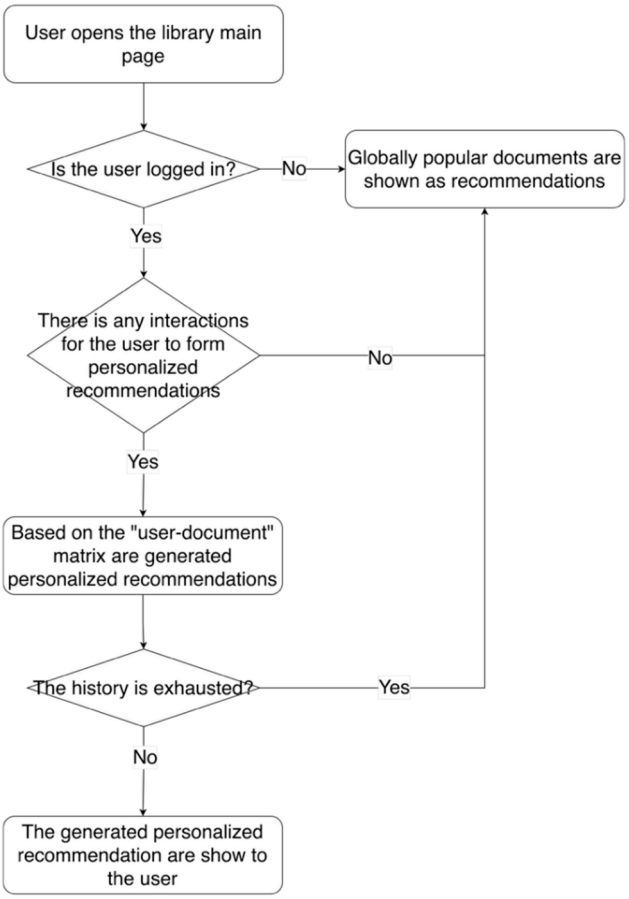


Figure 3. Basic flow for displaying personalized recommendations.

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 in order to verify that the system responds in a stable and predictable manner across a range of realistic interaction patterns. The evaluated scenarios focus on conditions that are particularly relevant for large and heterogeneous digital library collections.

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.

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 has been deliberately introduced between the documents in A and B (they are read together by the same users), while B remains without co-occurrences to observe the reaction in the absence of neighbors. 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

 doi:10.20944/preprints202602.1144.v1

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 of 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.

 doi:10.20944/preprints202602.1144.v1

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 give a clear picture that the module create proper personalized recommendation, remains useful in the absence of data, and reacts predictably when interests change.

5. Conclusion 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 validation using scenario-driven synthetic tests shows that the hybrid approach exceeds item-based collaborative filtering and content-only baselines in terms of stability and relevance, mitigating popularity-driven spurious recommendations while maintaining coverage when user interaction history is sparse or unavailable. 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., D.P.-M.; Create software, E.M., V.G.; Validation, E.M., A.N.; writing—original draft, E. M., D.P.-M.; Manuscript review and contribution to final version, E.M., D.P.-M., V.G., A.N., 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/, 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.

Acknowledgments: Not applicable.

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

References

1. 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 Journal* **2025**, *vol. 14, no. 4*, 2946-2952. 10.18421/TEM144-05

 doi:10.20944/preprints202602.1144.v1

2. Paneva-Marinova, D.; Minev, D.; Krachanov, I.; Luchev, D.; Goynov, M.; Zlatkov, L.; Pavlov, R.; Pavlova, L. The Power of Intelligent Content Curation and Context-Based Digital Library Content Usage for Research and E-Learning (The Case of National Library "Ivan Vazov" – Plovdiv, Bulgaria). In Proceedings of the 16th annual International Technology, Education and Development Conference, IATED (March 2022). 10.21125/inted.2022.1097
3. Liu, Z.; Shao, B. A systematic review of library services platforms research and research agenda. *Library & Information Science Research* **2024**, *vol. 46, no. 4.* 10.1016/j.lisr.2024.101325
4. Borgman, C. L. Libraries, Digital Libraries, and Data: Forty years, Four Challenges. *portal: Libraries and the Academy* **2025**, *vol. 25, no. 3*, 39-58. 10.1353/pla.2025.a964199
5. 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). https://doi.org/10.1007/978-981-96-0865-2_7
6. Owusu-Ansah, C. R. A. Digital Information and Library Services in ODDE. In Handbook of Open, Distance and Digital Education (2022). https://doi.org/10.1007/978-981-19-0351-9_45-2.
7. Paneva-Marinova, D. ; Goynov, M.; Zlatkov, L.; Luchev, D.; Pavlov, R.; Pavlova, L. Implementation of Humanities and Social Sciences Data Storage, Retrieval and Curation Environment for National Library "Ivan Vazov" – Plovdiv Needs. In Proceedings of International ClaDA-BG Conference (2021)
8. Troussas, C.; Krouska, A.; Koliarakis, A.; Sgouropoulou, C. Harnessing the power of user-centric artificial intelligence: Customized recommendations and personalization in hybrid recommender systems. *Artificial Intelligence Models, Tools and Applications with A Social and Semantic Impact* **2023**, *vol. 12, no. 5*. https://doi.org/10.3390/computers12050109
9. Adewojo, A. A.; Dunmade, A. O. From big data to intelligent libraries: Leveraging analytics for enhanced user experiences. *Business Information Review* **2024**, *vol. 41, no. 3*, 104-109. 10.1177/02663821241264707
10. 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. *Digital Presentation and Preservation of Cultural and Scientific Heritage* **2024,** *vol. 14*, 19-36. https://doi.org/10.55630/dipp.2024.14.1
11. Koliarakis, A.; Krouska, A.; Troussas, C.; Sgouropoulou, C. Modified collaborative filtering for hybrid recommender systems and personalized search: The case of digital library. *17th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP)* **2022**, pp. 1-6. 10.1109/SMAP56125.2022.9942020
12. Tursunov, T.; Kaibassova, D.; Kashkimbayeva, N. Comparative Analysis of Recommendation Algorithms: Collaborative, Content-Based and Hybrid Approaches. In 2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST) (2025). pp. 1-5. 10.1109/SIST61657.2025.11139153
13. Yuan, H.; Hernandez, A. A. User cold start problem in recommendation systems: A systematic review. *IEEE access* **2021**, vol. 11, pp. 136958-136977. 10.1109/ACCESS.2023.3338705
14. Fayyaz, Z.; Ebrahimian, M.; Nawara, D.; Ibrahim, A. Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *Applied Sciences* **2020**, vol. 10, no. 21. https://doi.org/10.3390/app10217748
15. 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 (2019), pp. 181-187. https://doi.org/10.1145/3377672.3378054
16. Rana, P.; Singh, D. Navigating the information labyrinth: a smart search system for libraries. *Library Hi Tech News* **2025**, *vol. 42, no. 7*, pp. 6-10. https://doi.org/10.1108/LHTN-10-2024-0188
17. Kumar, V. A.; Chidambaram, M. Personalization and User Behavior Analysis in Digital Libraries: A Systematic Review. *Academic Research Journal of Science and Technology (ARJST)* **2025**, *vol. 2, no. 2*, pp. 37-43. 10.63300/arjst0202202505
18. Speciale, A.; Vallero, G.; Vassio, L.; Mellia, M. Recommendation systems in libraries: an application with heterogeneous data sources. *arXiv preprint arXiv:2303.11746* **2023**. 10.48550/arXiv.2303.11746
19. Liang, Y.; Wang, J. Intelligent library recommendation based on GNN and attention networks. *Journal of Computational Methods in Sciences and Engineering* **2025**. 10.1177/14727978251364472

 doi:10.20944/preprints202602.1144.v1

20. Liao, M.; Sundar, S. S.; Walther, J. User trust in recommendation systems: A comparison of content-based, collaborative and demographic filtering. In CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, USA (April 2021). https://doi.org/10.1145/3491102.3501936
21. Liao, M.; Sundar, S. S. When e-commerce personalization systems show and tell: Investigating the relative persuasive appeal of content-based versus collaborative filtering. *Journal of Advertising* **2022**, *vol. 51, no. 1*, pp. 1-16. 10.1080/00913367.2021.1887013
22. Stoikov, J. Using Conditional Probability for Discovering Semantic Relationships between Named Entities in Cultural Heritage Data. *Digital Presentation and Preservation of Cultural and Scientific Heritage* **2021**, *vol. 11*, pp. 77-88. https://doi.org/10.55630/dipp.2021.11.7
23. 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,* 2007, vol. 4321, pp. 291-324. https://doi.org/10.1007/978-3-540-72079-9_9
24. Kapembe, S.; Quenum; J. G. A Personalised Hybrid Learning Object Recommender System. In 11th International Conference on Management of Digital EcoSystems (2019), pp. 242 – 249. https://doi.org/10.1145/3297662.3365810
25. Ibrahim, I. I.; Heryadi, Y.; Qomariyah, N. N.; Budiharto, W. Improving Library Recommender Systems with a Hybrid Interaction-ALS Approach under Data Sparsity. In 2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA), Antalya, Turkiye (2025). pp. 1-6. 10.1109/ACDSA65407.2025.11166363.
26. Pan, T. Personalized Recommendation Service in University Libraries using Hybrid Collaborative Filtering Recommendation System. In 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), Hassan, India (2024). pp. 1-6. 10.1109/IACIS61494.2024.10721676
27. 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 Journal* **2024**, *vol. 13, no. 2*, pp. 1352-1361. 10.18421/TEM132-49
28. Stefanov, K.; Boychev, P.; Stefanova, E.; Georgiev, A. Digital Libraries in Teacher Education. In Fortieth Jubilee Spring Conference of the Union of Bulgarian Mathematicians (2011).
29. 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,* 2025, pp. 89-104. 10.1002/9781394200467.ch5
30. Verma, M.; Rawal, A. An enhanced item-based collaborative filtering approach for book recommender system design. *ECS transactions* **2022**, *vol. 107, no. 1*, 15439-15449. 10.1149/10701.15439ecst
31. Ifada, N.; Rahmatullah, A.; Rachman, F. H. Leveraging Hybrid Semantic Ontology-based Model for Library Book Recommendation System. In 2024 9th International Conference on Information Technology and Digital Applications (ICITDA), Nilai, Negeri Sembilan, Malaysia (2024). pp. 1-6. 10.1109/ICITDA64560.2024.10809808
32. National Library "Ivan Vazov" Plovdiv. Available online: https://digital.libplovdiv.com/bg.
33. Christozov, D.; Mitreva, E. Trust in Learning from Big Data: The Two Sides of the Same Coin. *Issues in Information Systems* **2020**, vol. 21, no. 1. https://doi.org/10.48009/1_iis_2020_147-152
34. Mitreva, E.; Nikolova, A.; Georgiev, V. Web mining techniques applicable for cultural heritage observations. *Digital Presentation and Preservation of Cultural and Scientific Heritage: Conference Proceedings* **2021**, *vol. 11*, p. 254–259. https://doi.org/10.55630/dipp.2021.11.22
35. Mitreva, E.; Nikolova, A.; Georgiev, V.; Gigova, A. Personalization approaches for cultural heritage study. *Digital Presentation and Preservation of Cultural and Scientific Heritage: Conference Proceedings* **2023**, *vol. 13*, p. 181–188. https://doi.org/10.55630/dipp.2023.13.17
36. Alkaabi, H.; Jasim, A. K.; Darroudi, A. From Static to Contextual: A Survey of Embedding Advances in NLP. *PERFECT: Journal of Smart Algorithms* **2025**, *vol. 2, no. 2*, pp. 57-66. 10.62671/perfect.v2i2.77
37. Das, K.; Kamlish; Abid, F. Advancements in Word Embeddings: A Comprehensive Survey and Analysis. In Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences (2024), vol. 61, no. 3. 10.53560/PPASA(61-3)842

 doi:10.20944/preprints202602.1144.v1

38. 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**, *vol. 11*, pp. 42909-42923. 10.1109/ACCESS.2023.3269693
39. Chang, Y.-c. I. A Survey: Potential Dimensionality Reduction Methods. *arXiv:2502.11036* **2025**. 10.48550/arXiv.2502.11036
40. Shahzad, M.; Barzamini, H.; Wilson, J.; Alhoori, H.; Rahimi, M. Dynamic domain analysis for predicting concept drift in engineering AI-enabled software. *Journal of Data and Information Science* **2025**, *vol. 10, no. 2*, pp. 124-151. https://doi.org/10.2478/jdis-2025-0020