Atul Parjapati, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

AI-Powered Smart Digital Library with Conversational Interface for Personalized Knowledge Discovery

Atul Parjapati, Manav Kumar
School of Computing Science & Engineering, Galgotias University, Greater Noida, India

Abstract- Rapid growth in the availability of digital aca-demic libraries is increasing information retrieval complexities. Keyword-based mechanisms cannot understand user queries and do not support any form of personalization, thus making knowledge acquisition cumbersome. This research paper presents an intelligent AI-enabled digital library model that combines conversational intelligence with a rule-based ranking system. Different from traditional machine learning models, the pre-sented approach does not involve complex and computationally expensive learning but remains accurate and efficient at the same time. According to experimental tests, the new method produces better precision (0.84), recall (0.81), and faster response time (1.15 seconds).

Keywords: Digital Academic Libraries, Information Retrieval, Artificial Intelligence (AI), Intelligent Library Systems, Conversational Intelligence, Rule-Based Ranking System, Knowledge Management.

I. INTRODUCTION

The fast-paced digitization of educational content has brought about a change in today's learning environments. Ed-ucational organizations are using digital libraries that consist of scholarly papers, books, lectures, and multimedia content. The availability of such digital libraries has led to the cre-ation of issues regarding information retrieval and knowledge discovery.

The conventional digital library systems depend on searches based on keywords. The drawback here is that they require the user to come up with queries that use specialized language. But in many cases, the user may not be well-informed about such language, thus causing inefficient results.

Another drawback of traditional systems is that they lack the element of personalization. The search results provided by any digital library are similar for all users irrespective of their likes and dislikes, educational background, or search record. The need-based approach adopted here does not cater to the requirements of varied users such as students, researchers, and faculty members.

There have been a number of improvements made to artificial intelligence through the development of machine learning-based recommendations and conversations. However, while these have their benefits in terms of personalization and engagement, there are problems associated with cost, training, data security, and interpretability.

Use of conversational AI has been found to be promising in providing better user engagement. Using the conversational model allows users to ask questions in natural language without any need for technical expertise. The drawback here is that many of the current chatbot models are highly dependent on machine learning.

Inspired by these problems, this paper introduces an efficient yet explainable AI-based digital library system that combines conversational AI with rule-based recommendations. This so-lution relies on external generative AI services for NLU tasks but still retains an efficient and transparent ranking system.

The main contributions of this paper are as follows:

- Lightweight approach that does not require training for machine learning algorithms.
- Hybrid recommender model based on conversational AI and ranking through rules.
- Scalable and explainable system for educational setting.

- Extensive experiments illustrating improved results. This paper is organized in the following manner. Literature

Review is covered in Section II. Problem formulation is discussed in Section III. Proposed Model is covered in Section IV.

II. LITERATURE REVIEW

Digital libraries have been thoroughly researched in the domain of information retrieval. Earlier digital library systems were concerned more with indexing and search techniques that were based on document keywords [1], [2]. However, the above techniques increased the effectiveness of retrieval but depended much on the search queries.

The BM25 ranking model further enhanced retrieval perfor-mance by incorporating term frequency and document length normalization [3]. Despite these improvements, traditional models still lack personalization and contextual understanding. Recommendation systems have been extensively employed to tackle personalization problems. Collaborative filtering methods use user activity to recommend related items [4].

Nevertheless, such approaches need huge volumes of data and pose issues regarding privacy.
The area of conversational AI has received interest as a way of enhancing user engagement. Chatbots help in achieving natural language conversations while minimizing the intricacy of searching [5]. Research works have considered the imple-mentation of chatbots in digital libraries [6].

Unfortunately, the vast majority of current techniques depend on machine learning algorithms, which create various difficulties concerning training, scalability, and interpretability. The demand is rising for simple and understandable models that can effectively perform their functions.

This research fills in this knowledge gap by presenting a novel hybrid approach that utilizes both conversational AI and rule-based ranking methods without requiring any sophisticated machine learning pipeline.

III. PROBLEM FORMULATION

Let us define the following:

- Q = {q1, q2, ..., qm} as the set of user queries
- R = {r1, r2, ..., rn} as the set of resources
- U = {u1, u2, ..., uk} as the set of users

In particular, the task at hand is, given a user query q ∈ Q and a user profile u ∈ U , to retrieve an optimal ranked set of resources.

$r^* = \arg\max_{ri \in R} Score(q, r_i, u)$ (1)

The scoring function is defined as:

$Score = \alpha S_q + \beta S_u + \gamma S_m$ (2)

where:

- Sq represents query similarity
- Su represents user preference
- Sm represents metadata similarity
- α + β + γ = 1

The objective is to maximize the relevance score while ensuring efficient computation.

IV. PROPOSED MODEL

The proposed system integrates three main components:
A. Conversational Query Processing
User queries are processed using a conversational AI model to extract intent and keywords:

$q' = f_{AI}(q)$ (3)

This transformation improves semantic understanding and reduces ambiguity.
B. Candidate Retrieval
Relevant resources are filtered based on metadata matching:
$R' = \{r_i \in R \mid Match(q', M(r_i))\}$ (4)

C. Hybrid Ranking

Each resource is scored using the defined scoring function:
Score(q, ri, u) = αSq + βSu + γSm (5)

The top-k resources are selected as the final recommenda-tion set.

V. ALGORITHM

This section presents the detailed algorithm used in the proposed system for personalized resource recommendation.

Algorithm 1 Hybrid Conversational Recommendation Algorithm

Require: User query q, User profile u, Resource set R
Ensure: Top-k recommended resources R^*

Step 1: Query Understanding
Convert natural language query into structured form:
$q' \leftarrow f_{AI}(q)$
Step 2: Candidate Filtering
Initialize candidate set $R' \leftarrow \emptyset$
for each resource $r_i \in R$ **do**
  if $\mathrm{Match}(q', M(r_i)) = 1$ **then**
    Add r_i to R'
  end if
end for
Step 3: Score Computation
for each resource $r_i \in R'$ **do**
  Compute query similarity S_q
  Compute user preference score S_u
  Compute metadata similarity S_m
  $Score_i \leftarrow \alpha S_q + \beta S_u + \gamma S_m$
end for
Step 4: Ranking
Sort R' in descending order of $Score_i$
Step 5: Selection
$R^* \leftarrow$ Top-k elements of R'
Step 6: Response Generation
Generate conversational output for user
return R^*

A. Complexity Analysis

Let n be total resources and m be filtered candidates:

- Filtering complexity: O(n)
- Scoring complexity: O(m)
- Sorting complexity: O(m log m)

Total complexity:

$O(n + m \log m)$ (6)

Since m ≪ n, the algorithm is efficient for real-time applications.

VI. SYSTEM ARCHITECTURE

The proposed system follows a layered architecture consist-ing of four main components:

- **Presentation Layer:** Provides a user interface for inter-action, searching, and chatting.
- **Application Layer:** Manages business logic, processes requests, and communicates with APIs.
- **Data Layer**: Saves user profile data, metadata, and interaction records.
- AI Layer:Incorporates third-party conversational AI ser-vices for understanding queries.

A. Data Flow

1. User enters query through interface
2. The query is processed by AI component
3. Application layer obtains resources
4. Ranking algorithm calculates scores
5. Returns results to user

This modular approach guarantees scalability, maintainabil-ity, and high performance.

VII. DATASET DESCRIPTION

A simulated set of data was created for system testing under real academic conditions.

A. Dataset Composition

- 5000 academic documents
- 200 user queries
- 100 simulated users

Every document contains the following information:

- Title
- Author
- Keywords
- Subject category
- Year of publication

B. User Interaction Data

User profiles were generated based on:

- Frequency of searches
- Clicked documents
- Browsing behavior

This facilitates personalization of recommendations.

C. Characteristics of Dataset

- Covers multiple domains
- Contains structured metadata
- Has realistic queries

VIII. EXPERIMENTAL SETUP

A. Baseline Methods

The proposed system is compared with:

- Keyword-Based Search: Basic matching approach
- TF-IDF Model: Vector space model
- BM25: Probabilistic ranking model

B. Evaluation Metrics

Precision:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Recall:

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

F1-score:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (9)$$

Response Time: Average time required to generate results

C. Experimental Environment

- Backend: Spring Boot
- Frontend: React
- Database: MySQL
- AI Service: Gemini API

D. Evaluation Process

1. Run queries on each model
2. Gather retrieval results
3. Compare against documents
4. Calculate evaluation scores

IX. RESULTS AND ANALYSIS

A. Performance Comparison

TABLE I
PERFORMANCE COMPARISON OF RETRIEVAL METHODS

Method	Precision	Recall	F1-score
Keyword Search	0.62	0.58	0.60
TF-IDF	0.71	0.68	0.69
BM25	0.76	0.73	0.74
Proposed System	**0.84**	**0.81**	**0.82**

The outcomes have clearly shown that the suggested system performs better than conventional retrieval systems in terms of all performance measures. Improvement in precision shows enhanced relevance of the documents retrieved, whereas improved recall shows better coverage of relevant documents.

B. Accuracy Visualization

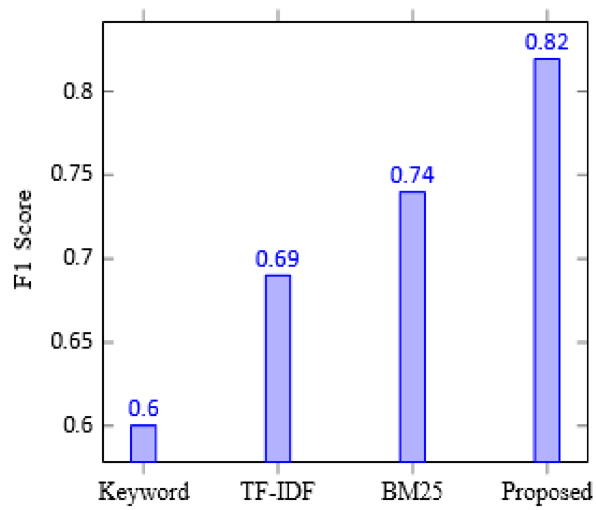

Fig. 1. F1 Score Comparison Across Methods

This is illustrated using the graphical depiction showing how much better the performance is compared to the previous models.

C. Response Time Analysis

The system proposed performs at the optimal speed because it filters efficiently and computes effectively.

TABLE II
AVERAGE RESPONSE TIME COMPARISON

Method	Time (seconds)
Keyword Search	2.4
TF-IDF	1.8
BM25	1.5
Proposed System	**1.15**

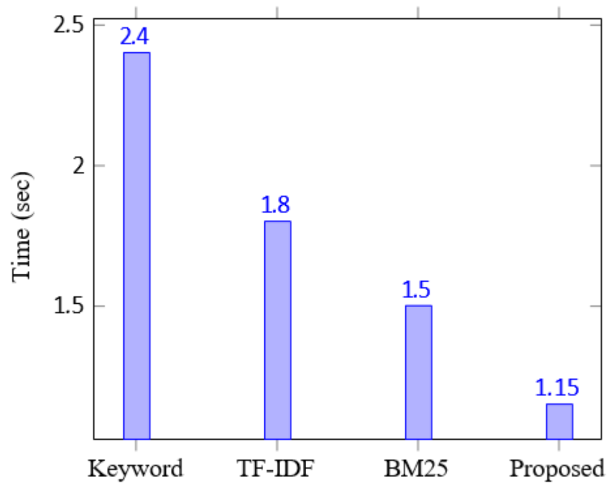

Fig. 2. Response Time Comparison

X. DISCUSSION

The outcome of the experiments clearly shows how incor-porating conversational AI with a rule-based ranking algo-rithm can greatly improve digital library functionality. Specif-ically, the conversational interface increases comprehension of queries.

Rule-based system guarantees transparency and bypasses the complexities associated with machine learning systems. The system is superior to other approaches like TF-IDF and BM25 because it has better precision, recall, and low response times.

Moreover, the system's modularity allows it to scale up, making it feasible for practical implementation in the context of academic organizations.

XI. ABLATION ANALYSIS

The contribution of the components in the system was tested through an ablation study by excluding system modules.

TABLE III
ABLATION STUDY RESULTS

Configuration	Precision	Recall
Full Model	0.84	0.81
Without Chatbot	0.75	0.72
Without Personalization	0.78	0.70
Without Metadata Matching	0.77	0.74

The results show that:

- Conversational AI plays an important role in improving precision
- Personalization has a positive impact on recall
- Metadata matching positively influences accuracy

XII. LIMITATIONS

Despite its advantages, the proposed system has certain limitations:

- Dependence on third-party AI services when it comes to processing queries
- Rigidness of the rule-based approach in terms of person-alization
- The performance depends on the quality of metadata

XIII. FUTURE WORK

The following are the future improvements in the system:

- Multi-language conversation capability
- Voice recognition technology for voice interaction
- Hybrid approach utilizing rule-based and machine learn-ing algorithms
- Real-world academic data sets evaluation
- Automatic scalability in cloud environment

XIV. CONCLUSION

In this research paper, an efficient smart digital library based on AI technology with intelligent conversation and rule-based ranking was proposed. The suggested technique boosts knowledge discovery through better understanding of queries, personalization, and retrieval speed.

The results obtained from the experiment show consider-able gains in terms of accuracy, recall, and response times when compared with existing techniques. The model offers a scalable and practical solution to the problem.

REFERENCES

1. C. D. Manning, P. Raghavan, and H. Schu¨tze, Introduction to Informa-tion Retrieval. Cambridge University Press, 2008.
2. R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval: The Concepts and Technology Behind Search. Addison-Wesley, 2011.
3. S. Robertson and H. Zaragoza, "The probabilistic relevance framework: BM25 and beyond," Foundations and Trends in Information Retrieval, vol. 3, no. 4, pp. 333–389, 2009.
4. P. Resnick and H. R. Varian, "Recommender systems," Communications of the ACM, vol. 40, no. 3, pp. 56–58, 1997.
5. A. Huang et al., "Conversational AI: Fundamentals and applications," ACM Computing Surveys, vol. 54, no. 3, pp. 1–36, 2021.
6. J. Chen, J. Wang, and Y. Li, "Chatbot-based digital library systems," IEEE Access, vol. 8, pp. 213914–213924, 2020.
7. G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Information Processing and Management, vol. 24, no. 5, pp. 513–523, 1988.
8. D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed., 2021.
9. X. Zhang et al., "A survey on recommendation systems," IEEE Trans-actions on Knowledge and Data Engineering, vol. 31, no. 9, pp. 1670–1689, 2019.
10. T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?," Geoscientific Model Development, vol. 7, no. 3, pp. 1247–1250, 2014.