KABARAK UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE & IT

SMART LIBRARY ASSISTANT

Fredrick Momanyi Morara

A Proposal submitted in partial fulfillment for the requirement of the award of Bachelor of Information Technology Degree of Kabarak University.

August 2025

DECLARATION

I, Fredrick Momanyi Morara, Registration Number INTE/MG/2814/09/22, do hereby declare that this project proposal titled:

"Smart Library Assistant"

is my original work and has not been submitted before for any academic award in this or any other university. All sources of information have been duly acknowledged.

No part of this work ought to be reproduced without my prior consent, or that of Kabarak University.

Signature: _____

Name: Fredrick Momanyi Morara

Date: August __, 2025

COPYRIGHT

ABSTRACT

Traditional library Online Public Access Catalogues (OPACs) often present challenges for users, particularly with topic-based or vaguely formulated queries, leading to inefficient book discovery. This project proposes the design and development of a "Smart Library Assistant," an intelligent chatbot leveraging Natural Language Processing (NLP) to address these limitations. The primary objective is to enhance the book discovery experience by enabling users to interact with their respective institution's library catalogue using natural language. The methodology will involve:(i) investigating current user challenges with OPACs; (ii) designing a robust system architecture integrating NLP for query understanding and information retrieval; (iii) developing a functional prototype capable of interpreting queries related to titles, authors, and thematic content within a specific library's dataset; and (iv) evaluating the prototype's effectiveness and user satisfaction. It is anticipated that the Smart Library Assistant will significantly improve the efficiency of book searches, increase user satisfaction, and demonstrate the practical benefits of AI in modern library services. This work aims to contribute a valuable tool for libraries to deploy for their specific collections and offer insights into the application of conversational AI for information access.

Table of Contents

CHAPTER ONE: INTRODUCTION

1.1 Introduction

Libraries, serving diverse communities including academic institutions and the general public, have long stood as vital repositories of knowledge, providing access to a vast array of information resources. In the digital age, the challenge for libraries is no longer solely about amassing collections, but increasingly about facilitating efficient and intuitive access to these resources. Traditional methods of catalogue searching, while foundational, often present hurdles for users, particularly those unfamiliar with precise bibliographic details or those exploring broad subject areas. This project, titled "Smart Library Assistant," aims to address these challenges by proposing the design and development of an intelligent chatbot. This AI-powered assistant will leverage Natural Language Processing (NLP) to understand user queries, interpret their information needs—even when vaguely expressed—and guide them to relevant books within a given library's active catalogue. This chapter lays the groundwork for the project, outlining the background of the study, defining the problem statement, presenting the project objectives, research questions, highlighting the significance of the study, and detailing its scope and limitations. The overarching goal is to enhance the user experience in library resource discovery, making it more interactive, efficient, and user-friendly across various individual library settings.

1.2 Background of the Study

Libraries play a crucial role in supporting education, research, and lifelong learning by providing access to extensive collections. The primary interface for accessing these collections is typically an Online Public Access Catalogue (OPAC). While OPACs have evolved, many still rely on keyword-based search mechanisms that require users to input specific terms, titles, or author names. Users often struggle when they:

a) Do not know the exact title or author of a book.
b) Are looking for information on a general topic (e.g., "set theory," "machine learning applications in finance") and are unaware of specific book titles covering these areas.
c) Use synonyms or related terms not perfectly matching the catalogue's metadata.

Currently, when users face difficulties with the OPAC in many libraries, their recourse is often to seek assistance from library staff. While librarians provide invaluable expert guidance, their availability can be limited, especially during peak hours or outside standard working times. Furthermore, some users might feel hesitant to approach staff for assistance. This situation indicates a gap in providing continuous, immediate, and intuitive search support in numerous library environments.

The emergence of Artificial Intelligence (AI), particularly in the domain of Natural Language Processing (NLP) and chatbot technology, offers a promising avenue to augment traditional library services. Chatbots can simulate human conversation, understand user intent from natural language queries, and provide instant responses. A "Smart Library Assistant" built on

these technologies could potentially understand the content or themes within books, going beyond simple metadata matching to guide users effectively, even with imprecise queries like "I want to learn about the basics of calculus" or "Which books discuss the history of ancient Rome?"

1.3 Problem Statement

The conventional methods for locating books in libraries, primarily through Online Public Access Catalogues (OPACs), present several challenges for users, leading to inefficiencies and potential underutilization of available resources. Key problems include:

1. Ineffective Topic-Based Search: Users seeking books on a specific topic or concept (e.g., "books on ethical hacking for beginners") often find OPACs restrictive if they don't know relevant keywords or subject headings used in the catalogue.
2. Difficulty with Vague or Incomplete Queries: Users who cannot recall exact titles or authors, or can only describe a book's theme, struggle to retrieve relevant results using traditional search interfaces.
3. Time-Consuming Navigation: Manually browsing through extensive search results or physical shelves when the initial search is not fruitful can be a time-consuming and frustrating experience for users.
4. Limited Accessibility to Human Assistance: Reliance on librarians for complex search assistance is constrained by staff availability, potentially leading to delays or unaddressed queries, especially outside operational hours.

These challenges collectively hinder the user's ability to efficiently discover and access relevant books, impacting their academic, research, or general learning pursuits within the context of their specific library's resources. This project proposes the development of a Smart Library Assistant to mitigate these issues by providing an intelligent, conversational interface that understands and processes natural language queries to facilitate more intuitive and effective book discovery in individual library settings.

1.4 Objectives

1.4.1 Main Objective
The main objective of this project is to design, develop, and evaluate a Smart Library Assistant chatbot that enhances book discovery within library catalogue environments by understanding natural language queries and providing relevant book recommendations from that specific library's collection.

1.4.2 Specific Objectives

i. To investigate the common challenges faced by library users in finding books using existing catalogue search mechanisms in typical library settings.
ii. To design the architecture of a Smart Library Assistant, incorporating Natural Language Processing (NLP) techniques for query understanding and information retrieval from library catalogues.

iii. To develop a functional prototype of the Smart Library Assistant capable of interpreting user queries about book titles, authors, topics, or content themes based on a representative, institution-specific library dataset.

iv. To evaluate the effectiveness and user satisfaction of the Smart Library Assistant in improving the book discovery process compared to traditional methods.

1.5 Research Questions

i. What are the primary difficulties and limitations encountered by users when searching for books in libraries using conventional OPAC systems?

ii. How can Natural Language Processing techniques be effectively integrated into a chatbot architecture to accurately interpret diverse user queries (including title, author, topic, and content-based) for book discovery within a specific library's context?

iii. To what extent can the developed Smart Library Assistant prototype accurately retrieve and recommend relevant books based on natural language queries related to topics or book content, using a representative institution-specific library catalogue?

iv. What is the perceived usability and satisfaction level among users when interacting with the Smart Library Assistant for book discovery tasks?

1.6 Significance of the Study

The development of a Smart Library Assistant holds significant potential benefits for various stakeholders:

For Library Users (e.g., students, researchers, general public): It will provide a more intuitive, efficient, and 24/7 accessible way to find books, especially for topic-based or vaguely formulated queries. This can save time, reduce frustration, and improve access to relevant resources, ultimately supporting their learning, research, and informational needs.

For Libraries: The assistant can enhance their service offerings, improve user engagement with their respective catalogues, and potentially reduce the burden on library staff for routine directional or simple search queries, allowing them to focus on more complex user support. It also positions them as adopters of innovative technologies to better serve their communities.

For the Field of Library and Information Science: This project contributes to the growing body of knowledge on applying AI and NLP in library settings, offering insights into the practical implementation and effectiveness of chatbot technology for information retrieval applicable across different types of libraries.

The realization of this project will demonstrate the tangible advantages of AI-driven solutions in improving information access and user experience in library environments.

1.7 Scope and Limitation of the Study

Scope:

The project will focus on the design, development, and evaluation of a chatbot for book discovery designed for deployment within individual library catalogue systems.

The Smart Library Assistant will interact with a representative digital library catalogue (or a significant, structured subset) specific to an individual library for development and testing purposes.

The core functionality will include understanding user queries related to book titles, authors, specific topics, and general content themes.

The system will be developed as a web-based application, accessible through a standard web browser, intended for integration with a specific library's existing infrastructure.

Natural Language Processing (NLP) techniques will be central to query understanding.

Limitations:

The project will not encompass the full range of library services (e.g., user account management, fines, inter-library loans). Its focus is solely on book discovery and recommendation from the catalogue of the specific library where it is deployed.

The accuracy of the assistant's responses will be dependent on the quality, completeness, and structure of the underlying library catalogue data of the host institution, used for training and operation, and the sophistication of the NLP models implemented.

The initial prototype may have limitations in understanding highly complex, ambiguous, or out-of-scope queries that fall outside the thematic scope of the host library's collection.

The project will be developed within the timeframe and resource constraints of a final year undergraduate project.

While the underlying model aims for general applicability to different libraries, the primary data collection and testing will be conducted with a specific sample of users and the catalogue of a particular library. This may influence the initial tuning and each new deployment in a different library would require configuration and potentially re-training or fine-tuning with that library's specific catalogue data

The system, in any given deployment, will be intentionally constrained to suggest only books available within that specific library's physical or licensed digital collection. It will not suggest resources from other libraries, even if aware of them through a potentially shared underlying knowledge base or cache.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

This chapter embarks on a review of scholarly literature to establish the academic context for the "Smart Library Assistant." The project envisions an intelligent conversational system designed to simplify how users within specific institutions discover books in their library's collection. By leveraging the power of Natural Language Processing (NLP) and modern Artificial Intelligence, particularly pre-trained Large Language Models (LLMs), this initiative seeks to transcend the limitations of traditional library search interfaces. This review will explore the well-documented challenges users face with conventional Online Public Access Catalogues, setting the stage for more intuitive solutions. It will then delve into the growing application of conversational AI and chatbots in enhancing information access and user support. Crucially, the discussion will cover how NLP, especially the capabilities of LLMs, can interpret user needs expressed in natural language and how these models can be effectively grounded with information from specific library catalogues to ensure relevant, institution-specific results. Architectural considerations, such as efficient data management for these AI systems and strategies for optimizing performance and cost through caching and shared knowledge, will also be examined. Finally, established methods for evaluating such systems will be outlined. The aim is to build a compelling case for the proposed Smart Library Assistant by connecting its core ideas to established research and highlighting its potential to offer a more effective and user-centric book discovery experience.

2.2 Review of Objective One: To investigate the common challenges faced by library users in finding books using existing catalogue search mechanisms in typical library settings.

The challenges library users encounter with existing catalogue search mechanisms, primarily Online Public Access Catalogues, are well-documented in library and information science literature. A foundational issue, consistently reported since early OPAC evaluations, is the "vocabulary problem" or "keyword barrier." As Borgman (1996) highlighted in a seminal paper, users often struggle to match their natural search terminology with the controlled vocabularies and specific metadata terms utilized by Online Public Access Catalogues. This discrepancy frequently results in failed searches or an overwhelming number of irrelevant results, particularly for users unfamiliar with precise bibliographic details or formal subject headings. The work of Gross & Taylor (2005) further underscored how reliance on controlled vocabulary, while intended to improve precision, can negatively impact recall when users employ different keywords.

Beyond vocabulary mismatches, studies like Novotny's (2004) protocol analysis revealed significant user difficulties in formulating effective search queries and navigating OPAC interfaces, often leading to frustration and abandonment of the search process. These systems often lack robust support for natural language queries, a feature users increasingly expect due to their experiences with web search engines (Head & Eisenberg, 2010 – *conceptual reference*

to user expectations shaped by web search). This is particularly evident in topic-based or exploratory searches, where users may not have a specific title or author in mind. Traditional OPACs typically struggle to interpret such vague queries or support the iterative, conversational search process that users might prefer (Antell & Engel, 2006). This body of literature collectively points to a persistent gap in providing intuitive, effective search support, especially for users seeking information thematically within a library's collection.

2.3 Review of Objective Two: To design the architecture of a Smart Library Assistant, incorporating Natural Language Processing (NLP) techniques for query understanding and information retrieval from library catalogues.

Designing an effective Smart Library Assistant necessitates an architecture that leverages modern Natural Language Processing (NLP) techniques to overcome the limitations identified in traditional systems. The core of such an architecture involves using conversational AI, or chatbots, as an interactive front-end. Literature supports the use of chatbots in service environments, including libraries, to provide more accessible and engaging user support (Shawar & Atwell, 2007; Mckie & Narayan, 2019). The proposed system intends to utilize a pre-trained Large Language Model (LLM), like Google Gemini, for its core language understanding capabilities. The power of LLMs, as demonstrated in landmark research such as that by Brown et al. (2020) on GPT-3, lies in their ability to comprehend and generate human-like text, enabling sophisticated interpretation of user queries with minimal task-specific fine-tuning.

For such an LLM to provide accurate and relevant book recommendations from a *specific library's catalogue*, it must be effectively grounded with that institution's data. The concept of Retrieval Augmented Generation (RAG), as detailed by Lewis et al. (2020), is highly pertinent here. This approach involves the LLM first retrieving information from an external, curated knowledge source – in this case, a processed version of the specific library's catalogue – before generating a response. This ensures that the assistant's suggestions are contextually relevant and limited to the institution's actual holdings. Key NLP tasks facilitated by the LLM within this architecture include intent recognition (discerning the user's goal), named entity recognition (identifying titles, authors, topics in the query), and semantic matching (understanding the meaning behind the user's words to connect them to relevant books, beyond mere keyword overlap) (Jurafsky & Martin, 2023 – *latest edition/draft for comprehensive NLP concepts*). The architecture must therefore include robust mechanisms for creating and maintaining this institution-specific knowledge base, potentially using automated data ingestion pipelines to process raw catalogue data (Abadi et al., 2016 – *conceptual link for managed data processing*), and efficient caching strategies to optimize performance and manage costs associated with LLM API usage (Baeza-Yates & Ribeiro-Neto, 2011, Chapter on Caching).

2.4 Review of Objective Three: To develop a functional prototype of the Smart Library Assistant capable of interpreting user queries about book titles, authors, topics, or content themes based on a representative, institution-specific library dataset.

The development of a functional prototype directly translates the architectural design into a working system. Central to this objective is the LLM's ability to interpret a diverse range of user queries—from specific title/author requests to broader thematic explorations—all within the context of the institution-specific library dataset. This involves not only matching query terms to metadata but also understanding the *content themes* of the books. Modern LLMs have shown strong capabilities in tasks like text summarization and key topic extraction (Brown et al., 2020), which can be leveraged during the data ingestion phase to enrich the institutional knowledge base with semantic information about each book. This enriched data then allows the LLM, through the RAG approach, to respond more intelligently to content-based queries.

When users pose ambiguous or very broad topic queries, the system might need to engage in clarification dialogues. Literature on dialogue management in conversational systems provides frameworks for how systems can ask clarifying questions to narrow down the user's intent effectively (Jurafsky & Martin, 2023). While the LLM can use its general world knowledge to better understand the *topic* a user is interested in (e.g., if it's a niche subject), any subsequent book recommendations must be strictly derived from the institution's curated knowledge base. This careful delineation between general topic understanding and institution-specific resource recommendation is critical, aligning with principles from closed-domain question answering systems where responses are confined to a specific corpus (Chen & Yih, 2020). The prototype's design will also incorporate local caching mechanisms to store results of frequent queries within that institution, thereby improving response times and system efficiency (Tanenbaum & Van Steen, 2007), and potentially benefit from a shared, aggregated knowledge layer to inform initial topic-book associations, drawing parallels to concepts like transfer learning (Pan & Yang, 2009).

2.5 Review of Objective Four: To evaluate the effectiveness and user satisfaction of the Smart Library Assistant in improving the book discovery process compared to traditional methods.

Evaluating the developed Smart Library Assistant requires a comprehensive approach, assessing both its technical efficacy in information retrieval and the overall user experience. Standard Information Retrieval (IR) metrics such as Precision, Recall, and F1-score will be crucial for measuring the accuracy and completeness of the book recommendations provided by the system from the specific institution's collection (Manning, Raghavan, & Schütze, 2008, Chapter 8). These metrics will quantify how well the assistant retrieves relevant items and avoids irrelevant ones.

Beyond these objective IR measures, user-centric evaluation methodologies are essential for a system designed for direct user interaction. The System Usability Scale (SUS), a widely

validated questionnaire, offers a standardized way to measure users' subjective perception of the system's usability (Brooke, 1996). Comparative task-based evaluations, where users perform defined book discovery tasks using both the Smart Library Assistant and the institution's existing OPAC, will provide valuable data on task completion rates, time taken to complete tasks, and perceived ease of use (Nielsen, 1994). Qualitative data gathered through methods like think-aloud protocols during user testing and post-task interviews will offer deeper insights into user thought processes, points of satisfaction or frustration, and suggestions for improvement. Furthermore, specific frameworks for evaluating the quality of conversational AI systems, such as those proposed by Radziwill & Benton (2017), can be adapted to assess aspects like the assistant's ability to understand queries, the coherence of its responses, and its overall perceived helpfulness and intelligence.

2.6 Conceptual Framework / Concept Map

The literature review reveals a clear trajectory: traditional library OPACs present significant usability challenges (Borgman, 1996; Novotny, 2004), creating a need for more intuitive information discovery tools. Conversational AI, powered by sophisticated LLMs (Brown et al., 2020; Shawar & Atwell, 2007), offers a promising avenue. The effectiveness of such systems in specialized domains hinges on grounding them with accurate, context-specific knowledge, for which retrieval-augmented approaches are key (Lewis et al., 2020). The proposed Smart Library Assistant conceptualizes a system that leverages these principles, integrating an LLM with an institution-specific knowledge base derived from the library's catalogue. This core knowledge is supported by efficient data ingestion pipelines and multi-layered caching strategies designed to enhance performance and manage resources (Baeza-Yates & Ribeiro-Neto, 2011). The expected outcome is a marked improvement in the user's ability to discover relevant books within their own library's collection, leading to increased user satisfaction and more effective resource utilization. The success of this approach will be measured using a combination of established IR and HCI evaluation metrics (Manning et al., 2008; Brooke, 1996).

Conceptual Framework for Smart Library Assistant (Chapter 2)