

A Chatbot-Based System for Graduate Admission Information Retrieval in Taiwan

李宜恩
國立政治大學
111208001

鄭光希
國立政治大學
112306003

唐湘怡
國立政治大學
111703005

晏天德
國立政治大學
112303081

Abstract

This study proposes a chatbot-based system that uses web crawlers and OCR to collect and standardize admission data, enabling natural language queries through a LINE Official Account. A FastAPI backend handles intent classification and retrieval with a lightweight BM25-style ranking. Users can quickly access deadlines, quotas, and requirements, and use features such as calendar integration, bookmarking, and comparison between different programs. The system aims to improve information retrieval efficiency and user experience, bridging the gap between fragmented data and practical access through conversational AI.

1. Introduction

Graduate school admission information in Taiwan is displayed in various ways and formats, making manual searching inefficient and error-prone. While chatbots have been widely applied in education, most of them focus on tutoring rather than administrative queries. This research proposes a chatbot-based system that automatically extracts and standardizes admission data, enabling students to access and compare key information through natural language queries, thereby improving retrieval efficiency and user experience.

1.1. Motivation

In Taiwan, graduate school admissions through recommendation or screening play a crucial role in higher education. However, information regarding these admissions is scattered across various university websites and documents, often in heterogeneous formats such as PDFs, HTML pages, or images. Students spend significant time manually searching and comparing admission details such as deadlines, available seats, required documents, and interview schedules, which increases the risk of missing critical updates.

With recent advancements in artificial intelligence and natural language processing, chatbots have been widely adopted in education, showing positive impacts on

information retrieval efficiency and user experience. This motivates the development of a chatbot-based system that can streamline the process of accessing graduate admission information across institutions.

1.2. Problem Definition

This research aims to address the following challenges:

- Information fragmentation across multiple school websites.
- Heterogeneous formats, making automated extraction difficult.
- Low retrieval efficiency, as students cannot query data using natural language.
- Lack of comparison tools for cross-institutional analysis.

The problem can thus be defined as: How to design a natural language interface chatbot capable of extracting, standardizing, and responding to queries on graduate school admission information across universities.

1.3. Existing Solutions and Literature

Currently, students rely on manual browsing or social platform to gather admission information, which is time-consuming and incomplete. While every official platforms provide aggregated admission documents, they are often static and lack interactivity.

In academia, prior work has explored chatbots in education. For instance, a systematic review highlighted the effectiveness of chatbots in supporting learning and information provision, though most focus on tutoring rather than administrative queries [2]. Retrieval-augmented generation has been proposed to enhance answer accuracy and explainability, and its potential in educational applications has been validated [3]. Other studies also emphasized the multi-role potential of chatbots in education [4], but applications specifically targeting graduate admission information retrieval remain limited. These findings reveal both the promise of chatbots and the existing research gap that this study intends to fill.

1.4. Proposed Approach

The proposed solution consists of some main components:

- a. Information collection and standardization using web crawlers and OCR to automatically extract admission data, stored in a structured database with fields such as program, quota, deadlines, required documents, and interviews.
- b. Conversational retrieval system combining keyword search and semantic retrieval within a RAG framework, enabling natural language queries with links to official sources.

Through this approach, the system aims to create a reliable and interactive platform for graduate admission queries, bridging the gap between existing literature and practical application.

2. Method

We will use a LINE Official Account as the entry point. The backend adopts FastAPI (Python 3.11) to provide a callback webhook. After verifying the X-Line-Signature, it performs intent classification and data retrieval. The data layer uses Google Sheets as the single source of truth. On startup, the service loads the entire sheet, constructs strongly typed records and an inverted index, and uses Redis to cache both query results and user state. A background job refreshes the snapshot every five minutes and increments a version number. Cache keys include this version to prevent stale reads. If a refresh fails, the system falls back to the previous snapshot.

Query understanding is rule-first, the text is normalized with NFKC, full-/half-width variants are unified, and synonyms are mapped (e.g., 台/臺; 資工/CS, different Chinese writing of “Taiwan” or different way of calling the major). Based on keywords, the system identifies four intents: deadlines, requirements (English threshold/documents), quotas, and general browsing. Retrieval combines the inverted index with a lightweight BM25-style weighting (with higher weights for department and program name). For deadline queries, the earliest deadline is used as the tie-breaker, and for other intents, record update time is used for freshness. Each response returns at most five items and includes a page token (containing the normalized query, offset, and version). Page results use version + intent + query + page as the Redis key, and hot queries are periodically prewarmed.

Responses are dynamically assembled with LINE Flex templates. A bubble presents the university, department/program, and deadline (color-coded by days remaining: “urgent” within 7 days, “soon” within 30 days, and “closed” if past), along with quota, English requirement, and required documents. The footer provides actions for “Open Official Page,” “Add to Calendar,” “Bookmark,” “Compare,” and “Next Page.” Missing fields are handled with placeholders, and string lengths are

controlled to comply with LINE limits. Extended features include two calendar integration options (Google Calendar deep link or ICS download), bookmarks (Redis Sets keyed by an anonymized user identifier), and side-by-side comparison of up to three items.

3. Expected Results

We expect the proposed system to deliver three major outcomes: (i) core information query functions, (ii) extended features, and (iii) interaction scenarios.

The chatbot will provide users with essential admission information through keyword-based queries. Specifically, it will return: (a) important dates such as deadlines and document submission deadlines, (b) admission quotas for each program, and (c) links to the complete application guide.



Figure 1: When users enter the name of the graduate school, the chatbot will provide the information.

Beyond the core query functions, the system will support additional features to improve usability and personalization: (a) calendar integration, allowing users to directly add critical deadlines to Google Calendar for automated reminders; (b) a bookmarking function, enabling quick access to saved departments of interest; and (c) program comparison, providing side-by-side comparisons of two to three departments for informed decision-making.

To illustrate the expected results, we present two representative interaction scenarios of the proposed LINE chatbot: (i) a user queries key admission information by sending a message, and (ii) a user saves a department of interest to the bookmark list.

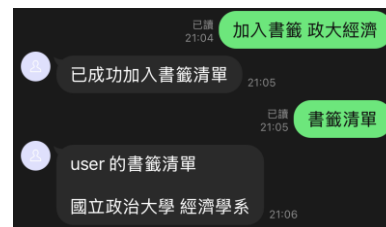


Figure 2: Users can save the department they are interested into their own bookmark list.

References

- [1] Shawn. Over 10,000 users in one month: LINE bot creation and sharing for university department search , 2022. [Online] (<https://jcshawn.com/dpsearch-bot-development-journal/>)
- [2] Labadze, L., Nelimarkka, M. Role of AI chatbots in education: A systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), 1–28, 2023.
- [3] Li, X., Zhang, H.,and Wang, Y. Retrieval-augmented generation for educational application. *Patterns*, vol. 6, no. 5, article 100578, 2025
- [4] Chen, Y., Liu, Z., Huang, C. Chatbots in education: A systematic review of objectives, evaluation methods, and challenges. *Computers & Education*, vol. 222, article 105091, 2025