AUTOMATED QUESTION ANSWERING
WITH CHATBOTS PROJECT PLAN

SUNG TRACY HIU YANG (3035369522)
KONG YAN TUNG NATALIE (3035367031)
NYEONGMIN LEE (3035401370)

SUPERVISOR

DR. ANTHONY T.C. TAM
1. **Project background**  
   1.1. Current Situation  
   1.2. Problem Identification  
      1.2.1. Delay in solving users’ questions  
      1.2.2. Huge Demand of Manpower  
   1.3 Our Solution

2. **Objective and Scope**  
   2.1. Objective  
   2.2. Scope

3. **Methodology**

4. **Schedule and Milestone**

5. **References**
1. PROJECT BACKGROUND

The HKU library (hereafter referred to as HKUL) was first established in October 1912. It was located in two rooms in the Main Building of HKU back in that time [1]. With over a hundred years of development, the Libraries now contain seven components, including the Main Library and six branch libraries. The Libraries provide a large range of information resources with collections exceeds 3 million volumes [1]. It includes books, journals and audio-visual materials. With the rapid development of technology, the Libraries have also rapidly expanded its collection of electronic resources with ebooks, e-journals and e-news. Apart from providing resources for students, alumni and staff to study and borrow, the Libraries also provide a conducive environment for study, research and knowledge exchange. There are a variety of learning spaces and study rooms available for users to self study or do discussions. In recent years, the libraries have also created new areas inside the libraries to meet the increasing needs of the digital world and rising demands of users. Ingenium is a new spaced that opened for the University Community to have interaction with the latest technologies. Think Tank and Social Space are also some of the learning spaces that were built to handle the demands.

1.1. CURRENT SITUATION

With the increasing services provided by the Libraries, the Libraries receive more and more questions regarding the services from users. For requesting a material, users may first reserve it in the online requesting system and get the material later. As for booking a study room or a study table in the library, users may login to the online booking system and reserve the space. These methods enable users to access and request these resources on and off campus. For checking the location of the book, users may use the devices inside the library or check it in the online requesting system. The Libraries have also offered a few channels to users to seek help whenever they face any questions about the library. Users may either visit the librarians who are on duty at the moment, email their questions to the librarian, call or WhatsApp the librarian. For users who would like to have a more advanced information research, they may also use the Research Consultation Service. However, this service is only offered to HKU staff or postgraduate students.

1.2. PROBLEM IDENTIFICATION

1.2.1. DELAY IN SOLVING USERS’ QUESTIONS

Both the librarians who are on duty or the librarians who received calls and reply on WhatsApp are only available on their working hours. When users would like to seek help outside the opening hours of the Libraries, they may fail doing that and have to wait till the next day. This system could not help users who have urgent requests and need immediate help.
1.2.2. HUGE DEMAND OF MANPOWER

Librarians may not be able to handle multiple users at the same time. Some questions are repeated and very general. On the other hand, some users may have more complicated questions and the librarian may have to spend more time helping them. The Libraries will have to allocate more manpower and they may have to recruit more librarians. The cost on manpower will increase. However, it is not possible in the reality as the numbers of human resources of the library are limited. Thus, the process of handling repeated questions is time-consuming and inefficient.

1.3 OUR SOLUTION

The HKUL Chatbot will be incorporated into the HKUL website to achieve an automated answering system that can optimise the readiness of the resources on the website. The Chatbot will understand the context of the questions that users made through a trained NLP model, then prompt an interactive human-like answer to the user.

The answers will be provided to them in forms of replies with the relevant information or directions to the corresponding HKUL page.

2. OBJECTIVE AND SCOPE

2.1. OBJECTIVE

The main objective of this project is to build a chatbot that can understand, interpret and provide answers to questions on HKUL from the library users or the public at any time. Ultimately, the Chatbot aims to provide accurate and immediate responses to users through an interactive and friendly way that aligns with the HKUL image.

2.2. SCOPE

The initial scope is set to be all static information that can be found on the HKUL website. (https://lib.hku.hk/)

In particular, the chatbot will be able to articulate static information available on the HKUL website. The chatbot will also record user questions that need follow-up actions from the librarians.

Through studying the HKUL website structure, we identified three areas that the questions can be categorised into. First, the general questions that subjects to specific branch libraries under HKUL, they
include the libraries’ basic information (e.g. location, hours and contacts). Second, the policy information which differs by users, such as borrowing duration, membership fee. The chatbot will ask the user type on questions identified in this category first before providing them the answer. Third and the last category consists of generic questions which can be the HKUL publications, donations (Giving to HKUL) and document searching (Find@HKUL).

The chatbot is also designed to cover document searching and availability checking through the prebuilt API to the Integrated Library System which does not require a login action. Primarily, the project does not include solutions that involves logging in to user accounts. The scope is subject to change upon further discussion with the HKUL staff to include user accounts access for book or room reservations and filling in e-Forms in later stages.

3. METHODOLOGY

This section will introduce the methodology needed to achieve the project objectives presented in the previous section. The essential technologies this project will use are natural language processing (NLP) including natural language understanding (NLU) and natural language generation (NLG) and machine learning (ML). NLP is to transform natural language into structured data for machines to interact with human language [2]. NLU is to let machines understand human language while NLG is to produce similar to human language by machines. ML is to train machines to find patterns from data without specific instructions given.

Firstly for this project, past QnA data from whatsapp-a-librarian system will be analyzed. The data is not yet received and will be provided by HKUL. Python, NLTK library and jupyter notebook will be used to preprocess data and order words by frequency to prioritize questions to be handled in this project. NLTK allows to tokenize, remove stop words, and classify text [3].

Secondly, NLU model will be built with Dialogflow. Dialogflow is a google ML chatbot framework which allows to train the machine to find intents, information the user wants, and entities, supporting information needed to give response. Intents and entities can be set manually and multiple entities can be matched to one intent. Also, some entities can be ‘required’ which is the mandatory entity for this intent. If the input question does not provide it, added ‘prompt’ of this entity will ask again to get such information [4]. Then, input question can be matched to intent and entities manually or Dialogflow can automatically match them if there are enough past input questions. Any manual edition is also available for any wrong prediction.
Database structure will be organized along with building NLU model for efficient database search. Since information posted on the HKUL website is not databased, members will construct database mainly categorized by library branches and user types. The Python BeautifulSoup web crawling method has been considered, yet due to the complicated structure and lengthy data on HKUL website, the manual input is predicted to be more suitable in this case as keywords can be identified more efficiently. Regardless, the web crawling method will be testified in the early phase of data collection. Besides database, library item search engine Find@HKUL will be used for document searching.

After building NLU model and database successfully, responses for each intent will be matched. Responses on Dialogflow can be either built-in or dynamic by connecting to fulfillment. Built-in response is used for static response that does not require data from database, such as simple greetings. The responses that need data from database will use fulfillment in Dialogflow [5]. Fulfillment allows to write JavaScript code on its inline editor to call API, imported database as an API for this project, get data from it and print response including such data [6]. The logs of code execution and model training can be viewed in Firebase, an application development platform connected to Dialogflow fulfillment inline editor. Calling external API is not supported in free plan on Dialogflow. Blaze plan which will be charged by service usage but has minimum limit which is predicted to be more than enough for this project for free of charge [7]. Billing information required for this service will be further discussed with HKUL.

When the model can understand the questions and get corresponding responses, it now needs to be connected to users. An agent will be needed to connect the client to the chatbot to send the user’s input and receive the Chatbot’s responses. DialogFlow offers the option to call the built chatbot through an API. The implementation and deployment of the chatbot will be handled with a Python script [8] on Git.

The final product of the Chatbot will be integrated on the HKUL website with an easy to use and clear user interface. The chatbot user interface will be built utilizing a convention open-source UI template BotUI [9] where it can be further customised with SCSS to display the HKUL chatbot interface. To create a friendly and down-to-earth impression, the user interface will be designed to be interactive using cartoons and icons. This can be achieved through JavaScript function on the UI script. The agent and the interface file will be stored at the HKUL server machine in order to process the web request and also facilitate future maintenance on the chatbot.
## 4. SCHEDULE AND MILESTONE

<table>
<thead>
<tr>
<th>Date</th>
<th>Milestone</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sep. 29, 2019</td>
<td></td>
<td><strong>Deliverables of Phase 1 (inception)</strong></td>
</tr>
<tr>
<td>Oct. 15, 2019</td>
<td></td>
<td>- Analyze past QnA data from HKUL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Testify BeautifulSoup on Data Scraping</td>
</tr>
<tr>
<td>Nov. 15, 2019</td>
<td>M1</td>
<td>Train question NLU model with Dialogflow</td>
</tr>
<tr>
<td>Dec. 31, 2019</td>
<td>M2</td>
<td>- Design and complete constructing the database</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Connect NLU model to database (preliminary implementation)</td>
</tr>
<tr>
<td>Jan. 13, 2020</td>
<td></td>
<td><strong>First Presentation</strong></td>
</tr>
<tr>
<td>Feb. 2, 2020</td>
<td></td>
<td><strong>Deliverables of Phase 2 (elaboration)</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Complete interim report</td>
</tr>
<tr>
<td>Feb. 29, 2020</td>
<td>M3</td>
<td>Build dynamic response</td>
</tr>
<tr>
<td>Mar. 15, 2020</td>
<td>M4</td>
<td>Integration testing (question-response accuracy)</td>
</tr>
<tr>
<td>Mar. 31, 2020</td>
<td>M5</td>
<td>Implement on HKUL website (UI/UX)</td>
</tr>
<tr>
<td>Apr. 10, 2020</td>
<td>M6</td>
<td>Final testing</td>
</tr>
<tr>
<td>Apr. 19, 2020</td>
<td></td>
<td><strong>Deliverables of Phase 3 (construction)</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Finish final report</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Prepare for presentation</td>
</tr>
<tr>
<td>Apr. 20, 2020</td>
<td></td>
<td><strong>Final presentation</strong></td>
</tr>
<tr>
<td>May 5, 2020</td>
<td></td>
<td><strong>Project exhibition</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Prepare poster for exhibition</td>
</tr>
</tbody>
</table>
5. REFERENCES


