

SmartLibrarian

ISS IPA PROJECT REPORT



ONN WEI CHENG A0092201X

YANG JIESHEN A0003901Y

NIRAV JANAK PARIKH A0213573J

ISA-PM-IPA-2021-01-09-IS02PT

TABLE OF CONTENTS

1. PROBLEM STATEMENT	2
1.1. BUSINESS CASE	2
1.2. OUR PROPOSAL	3
2. SYSTEM OVERVIEW.....	5
2.1. AGENT SYSTEM: CONVERSATIONAL UI.....	5
2.2. RPA: NLB AVAILABILITY CHECKER AND ABSTRACT SUMMARISER.....	5
2.3. RPA: AMAZON RECOMMENDATION SYSTEM AND INTELLIGENT TITLE MATCHER	6
2.4. RPA: EMAIL AGENT	6
3. SYSTEM IMPLEMENTATION.....	7
3.1. AGENT SYSTEM: CONVERSATIONAL UI.....	7
3.2. ROBOTIC PROCESS AUTOMATION.....	7
3.3. LOCAL AI: INTELLIGENT TITLE MATCHER & ABSTRACT SUMMARISER.....	8
4. CONCLUSION.....	9
4.1. SUMMARY OF ACHIEVEMENTS.....	9
4.2. FUTURE IMPROVEMENTS	9

APPENDICES

APPENDIX I:	PROJECT PROPOSAL
APPENDIX II:	SYSTEMS MAPPED TO COURSE OBJECTIVES
APPENDIX III:	SMARTLIBRARIAN INSTALLATION AND USER GUIDE
APPENDIX IV:	SMARTLIBRIRIAN FEATURES AND SCREENSHOTS

1. PROBLEM STATEMENT

What should you read? For both the avid readers and uninitiated, the major retailers, such as Amazon and Kinokuniya, provide a constantly list of best sellers, while blogs and other websites share the must-reads with titles such as “The Top 30 books you must read in your lifetime”. The recommendations are aplenty but how do readers discern between the different recommendations and how do they get hold of the book for free? This is what SmartLibrarian is here to solve.

1.1. BUSINESS CASE

The impact of Covid 19 has turned many businesses online and people toward activities that can be done individually and at home. The online book service market, is not an expectation to this trend, with an annual compounded growth of 6.2% per annum with an expectation to reach 23.8 Billion in 2026.¹

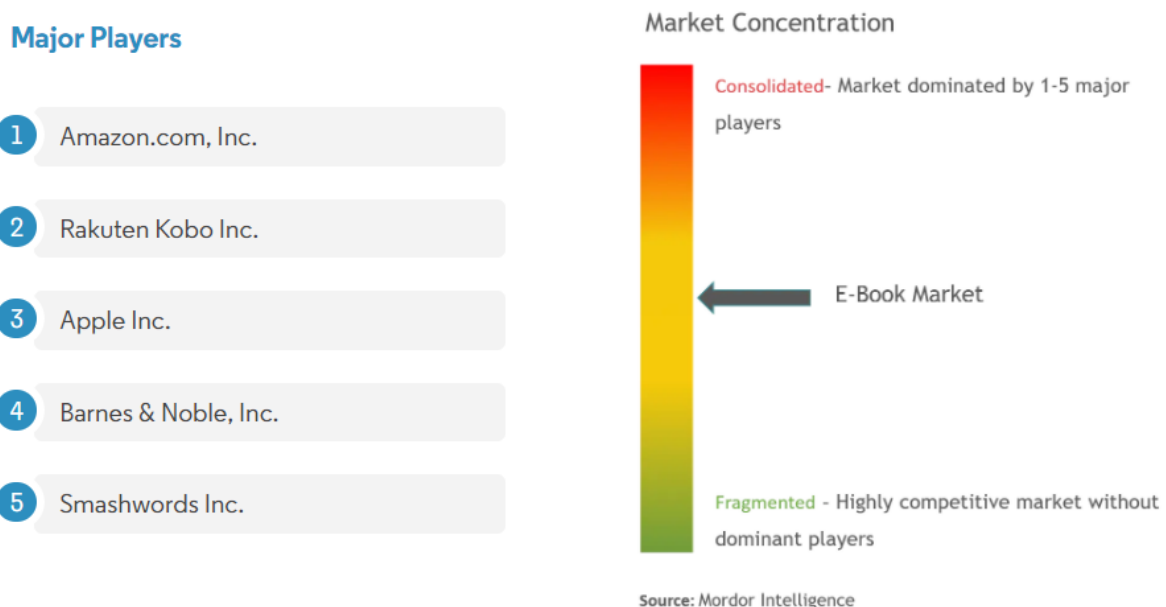


Figure 1: Market Concentration²

Major industry players in the fragmented ebook market that had a foothold in the ebook market also made moves to consolidate their position. For example, Amazon.com, offered the users a free trial of two months on their “Kindle Unlimited” service since June 2020 in an attempt to on board more users.

Most major players in the ebook market seeks to earn revenue by offering ebooks for sale or via a subscription model that offers a spectrum of books for download. The Smart Librarian

¹ https://www.reportlinker.com/p05975423/Global-Online-Book-Services-Market-By-Category-By-Region-Industry-Analysis-and-Forecast.html?utm_source=GNW

² <https://www.mordorintelligence.com/industry-reports/e-book-market>

seeks to disrupt the market by offering the books for free through the NLB ebooks collection. The recommendation system also provides an additional feature to borrow/read the next book.

The application seeks to generate revenue via 2 methods:

Support from NLB/Government

As our application supports the goal of “highlighting the resources of NLB and promote a love of reading for pleasure and development”³, it is highly likely that a collaboration can be achieved between us and NLB. This can in terms of grants or official integration of our application with the NLB website.

Advertisement/Sponsored Recommendation System

Currently our Intelligent Book Borrowing System relies on the Amazon recommendation system for our recommendations. However, with the development of the application, our proprietary system for recommendations can be developed and integrated with sponsored recommendations. Advertisements can also be used to generate revenue for the application.

1.2. OUR PROPOSAL

Smart Librarian is an Intelligent Book Borrowing System that can check for the availability of the books with our National Library Board (NLB) so it can be borrowed for free and also recommend books based on your selections.

The goal of the project will be to deliver a Minimum Viable Product (MVP) of our Intelligent Book Borrowing System, SmartLibrarian with the following objectives:

Activating The Application On The Go

Integrated with Google Assistant, you can check the availability of a book using your smart phone, no matter where you are. Natural language voice and text support is also provided to ensure a simple and natural workflow.

Checking The Availability And Recommendation Of Books

Boosting a huge selection of books and a great association recommendation system, our application taps on the Amazon system using a Robotic Process Automation (RPA) to recommend books for the users. An RPA is also used to check the NLB website for the availability of books. As NLB and Amazon both uses their own proprietary search engine, a matching system is also developed to ensure that the results from both NLB and Amazon are the same.

³ <https://www.nlb.gov.sg/WhatsOn/Programmes/ProgrammesforAdults.aspx>

Notifying The Users On Availability And Recommendations

An email will be sent to the user to inform the user on the availability of the books selected and the associated recommendations. Summarised abstracts of the book and the links to borrow the books in NLB will be included in the email to tease the content of the book and to ensure a seamless borrowing process respectively.

2. SYSTEM OVERVIEW

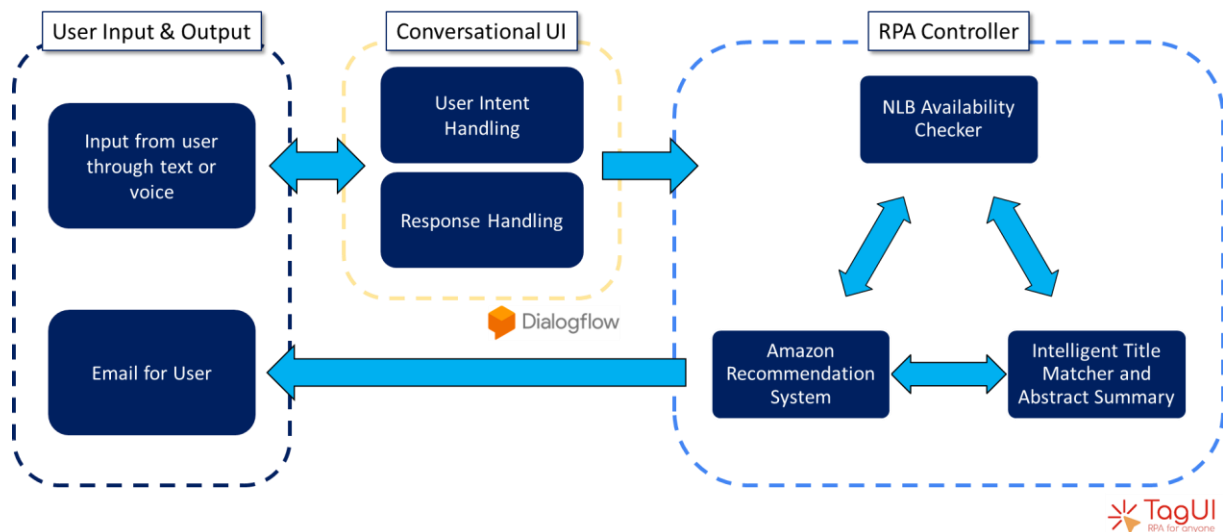


Figure 2: System Overview

An overview of the system is given in **Figure 2**. The following sub-sections summarise the major features of the application.

2.1. AGENT SYSTEM: CONVERSATIONAL UI

Once the agent is activated using Google Assistant or Google Dialog Flow using their phone or desktop respectively, the conversational UI will request for book titles to be provided. The UI allows for more than one book to be searched at one time by asking for more titles to be provided. The search and recommendation process will start once the user finishes his list of books, or when a maximum of three books are given.

If the user put in a request while the system is still processing their previous request, the conversational UI will also be able to inform the user that the system is busy. This is important as it will take a while for the RPA to scrape the information from the internet.

2.2. RPA: NLB AVAILABILITY CHECKER AND ABSTRACT SUMMARISER

Based on the books that are being selected by the user through the Conversational UI, an RPA will be used to scrape the NLB website for the availability and other information of the titles across all book types, including ebook and audiobooks. An intelligent abstract summariser is also used to summarise the abstract into a punchline of one or two sentences (150 characters) to give a flavour of the book when the email is sent. If the user is interested to find out more, the user will be able to follow the link to read the entire abstract.

2.3. RPA: AMAZON RECOMMENDATION SYSTEM AND INTELLIGENT TITLE MATCHER

To provide for an intelligent recommendation based on the book that the user is interested to read. We made use of the amazon associative recommendation system to give suggestions to the user on the next books to borrow. It is determined that the associative purchases made by people who also bought the book selected by the user served as good recommendations as there lies a commitment to purchase, as compared to associative views made by the same people. It is also observed that these recommendations are likely to be of similar genre and types instead of just being from the same author, which might not be very informative.

As NLB and Amazon both have their own proprietary search engine, we realised that the same search term might not give rise to the same results. Therefore, an Intelligent Title Matcher is used to compare between the results given from NLB and Amazon to ensure the same books are being compared for recommendations.

2.4. RPA: EMAIL AGENT

To summarise all the information and recommendations that is found by the application. An email is sent to the user either through the default email or via an email set by the user during setup. With this email, the user is able to verify their interest on the book using the summary provided while having an easy call to action with the borrowing link provided. Recommendations related to the books selected are also provided in the same email.

3. SYSTEM IMPLEMENTATION

3.1. AGENT SYSTEM: CONVERSATIONAL UI

The Conversational UI made use of intent and slot detection of Google Dialogflow to enable natural language conversations. As part of our MVP solution, there are currently two custom intents, namely the WelcomeIntent and GetBookTitleIntent, and one default fallback intent. When the user first activates the application, the WelcomeIntent will be activated to get start the conversation.

After which, the GetBookTitleIntent will be activated to accept book titles as slots. Based on the training phases given, the GetBookTitleIntent can discern the book titles from conversations. Follow-up intents from GetBookTitleIntent will continue to handle responses until the user finishes his list of books, or when a maximum of three book titles are given.

The Conversational UI will then inform the user that an email containing the information about the request will be sent. The context and slots will subsequently be refreshed to accept a new request. This will trigger the RPA segment in a parallel thread to acquire information about the books and send the user an email. While this thread is running, the agent will reject any new queries triggered by the user until the RPA thread is completed and terminated.

3.2. ROBOTIC PROCESS AUTOMATION

After the request have been processed by the Conversation UI, a separate thread will be used to run the RPA Robot. The titles of the books requested will be searched first in the NLB website for the closest match. As NLB website shows different book types as different book units, the RPA will scrape the website to look for all the book types with the same title. This will ensure that all available book types, such as ebook and audiobooks, will also be available for the user. Other information such as ratings, abstracts, copies available will also be scrapped.

Subsequently, the exact book title will be searched in Amazon to extract its associative recommendations. However, as the amazon search operates slightly differently as compared to the NLB search the exact match might not appear as the first search result. Therefore, the first few books from the Amazon site will be extracted and compared with the search results of NLB using our Intelligent Title Matcher to find the correct book for the recommendations to be extracted.

All the information extracted using the RPA process will be consolidated in an email and sent to the user using an RPA robot. If no response is found during the search, the email will also be sent to inform the user. The user is able to define an email address during set up by placing their email in a txt setting file.

3.3. LOCAL AI: INTELLIGENT TITLE MATCHER & ABSTRACT SUMMARISER

The **Intelligent Title Matcher** is used to compare the book titles between NLB & Amazon. This is required as for marketing reasons Book Titles change somewhat between countries, editions & formats. In some cases, the words in the title are slightly rearranged or the words in subtitle text may be somewhat altered. Hence, a simple string compare of titles does not work. And for books with long titles, there could be other books that share a couple of common words in their titles (but not all) and are essentially completely different book. As such, a simple bag of words comparison may also be misleading. What humans do is to understand the 'meaning' of the book title and compare book titles based on this meaning. We need something similar to compare and find the best matching titles.

Traditional approaches to NLP, such as one-hot encoding and bag-of-words models, while useful for some tasks, do not capture information about a word's meaning or context. Instead, we use word-vectors which represents the word's distributed weight across dimensions. Each dimension represents a meaning and the word's numerical weight on that dimension captures the closeness of its association with and to that meaning. Thus, the semantics of the word are embedded across the dimensions of the vector. We then append these word vectors to create a *title vector* which represents the semantic meaning of the book title.

First, we tokenize the book title into individual words and then strip all the stop words to extract the key words in the book title. For creating this word vectors, we have used GenSim model pre-trained on Google News dataset. Using this model, we generate a word vector for each of these title key words. These generated word vectors are then appended to create the *title vectors* ie a vectorized representation of the book title. Finally, we measure the cosine similarity between these title-vectors to identify the closest matching title & hence recommend the correct book.

We also use an **Intelligent Abstract Summariser** to summarise the book's abstract from Amazon & NLB into a couple of sentences that captures the essence of the book in our email to the user.

Text summarization condenses a document into brief statements using mathematical methods. The transformers based NLP model called BERT provides a significant improvement on conventional approaches such as TextRank in the quality of summaries, combining context with the most important sentences.

We use the Extractive Summarization model which utilizes the raw words, sentences or phrases of the text and outputs a summarization, using only the content from the source. For this, we leverage the Summarizer which is a pre-trained BERT model and provides extractive text summaries. The summarizer can be configured for the length of the summary required (150 characters in this case).

Note, since these pre-trained Bert's Summarizer & Gensim's Word2Vec model are quite large they take several minutes to load into memory. Hence, we pre-load these models at the start to ensure acceptable response times when interacting with the user.

4. CONCLUSION

4.1. SUMMARY OF ACHIEVEMENTS

Integrated with Google Assistant and Dialogflow, the Smart Librarian allows the user to check the availability of books on the go with natural language. The inbuilt RPA robot taps on the resources of NLB and Amazon to provide information on the availability of the books and related recommendations, respectively. All essential information, including abstracts and links are summarised in the email to ensure a seamless borrowing process. With this, the Smart Librarian have accomplished all the objectives that have set out as an MVP.

4.2. FUTURE IMPROVEMENTS

Despite a successful implementation for a minimum viable product, there are areas for future improvements.

Firstly, the current solution involves multiple steps of setting up before the solution is ready for use. This can be improved on by further development to integrate the various functions such as ngrok and DialogFlow into a single solution. Furthermore, the solution can be hosted on a server to reduce setup time and steps for a better user experience.

Secondly, more search categories can be added to the application. Categories such as author, genre, book length etc, can be included to give a more wholistic search experience. The agent can even learn the user's profile based on search history and recommend new books for the user.

Thirdly, a response back from the chatbot on the summary of search results would complement the email with more details. This function is not viable with Google DialogFlow due to its limitations. To improve on this, another chatbot with the required capabilities will need to be developed and integrated.

Lastly, a full RPA to help the user reserve the book in NLB is an area that can be expanded into. The current agent stops at sending the link to the NLB through the email. But this can be further enhanced by seeking approval from the user, followed by reserving the book in NLB under the user's account.

APPENDIX I: PROJECT PROPOSAL

Name & Date of proposal: 30 January 2021 - NUS ISS, Intelligent Software Agents - Practice Module
Project Title: ISS ISA Project – Smart Librarian
Sponsor/Client: <i>(Name, Address, Telephone No. and Contact Name)</i> Institute of Systems Science (ISS) at 25 Heng Mui Keng Terrace, Singapore NATIONAL UNIVERSITY OF SINGAPORE (NUS) Contact: Mr. GU ZHAN / Lecturer & Consultant Telephone No.: 65-6516 8021 Email: zhan.gu@nus.edu.sg
Background/Aims/Objectives: Smart Librarian is an Intelligent Book Borrowing System that can check for the availability of the books with our National Library Board (NLB) so it can be borrowed for free and also recommend books based on your selections.
Requirements Overview: The goal of the project will be to deliver a Minimum Viable Product (MVP) of our Intelligent Book Borrowing System, SmartLibrarian with the following objectives: <ul style="list-style-type: none">• Activating the Application On The Go• Checking the Availability And Recommendation Of Books• Notifying the Users On Availability And Recommendations
Resource Requirements (please list Hardware, Software and any other resources) Software proposed for consideration: <ul style="list-style-type: none">• Chat-bots: Google Assistant• Cognitive systems: Google Dialog Flow• Robotic Process Automation: TagUI• Pertained machine learning models: Gensim Word2Vec & Bert Summarizer• Machine learning use cases: Extractive Summarizer, Semantic Cosine Similarity• Application container: NGROK to establish http tunnel link between Chatbot & Intelligent Agents
Number of Learner Interns required: (Please specify their tasks if possible) <ul style="list-style-type: none">• ONN WEI CHENG A0092201X• YANG JIESHEN A0003901Y• NIRAV JANAK PARIKH A0213573J

APPENDIX II: SYSTEMS MAPPED TO COURSE OBJECTIVES

System	Relevant Course Objectives
Conversational UI	Single Agent System implemented using Google Dialogflow
NLB Availability Checker	RPA Robot implemented using TagUI
Amazon Recommendation System	RPA Robot implemented using TagUI
Intelligent Title Matching	Local AI implemented using GenSim Word@Vec model pre-trained using Google News dataset
Intelligent Abstract Summary	Local AI implemented using Extractive Summarizer from the pre-trained BERT model.

APPENDIX III: SMARTLIBRARIAN INSTALLATION AND USER GUIDE

Installation Guide

1. Clone Github repository to location of choice
Note: Make sure file path does not have any spaces.
2. Download GoogleNews-vectors-negative300.bin from
<https://drive.google.com/file/d/0B7XkCwpI5KDYNINUTTISS21pQmM/edit>
Note: This file is 3GB big
3. Extract(unzip) and place GoogleNews-vectors-negative300.bin into “DialogFlow” folder
4. Install system with Python 3.8.2.
5. Run “install.bat” to install all required python packages
6. If using your own account in Dialogflow, and not the credentials provided, import SmartLibrarian agent through the zip file provided in Github into Dialogflow to setup the agent.
7. If test is to be done on Google Assistant, integration setup with Google Assistant needs to be done under “Integrations” tab.
DialogFlow link: <https://dialogflow.cloud.google.com/>

For the provided credentials, all setup in Dialogflow is done, including Google Assistant integration

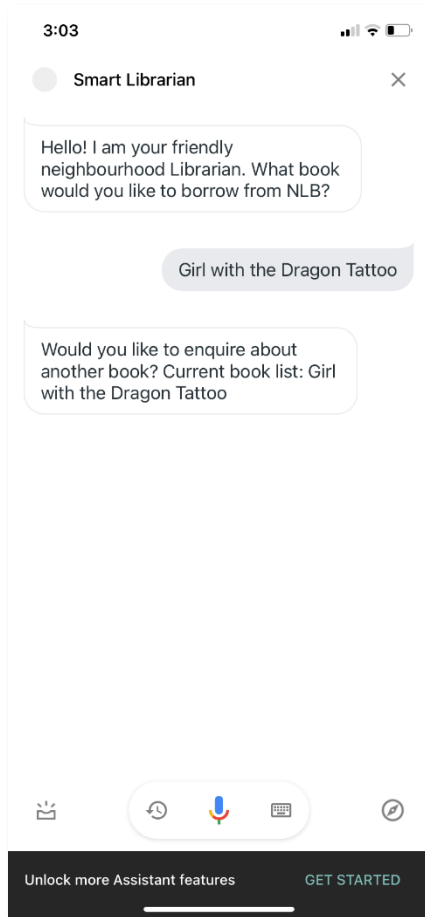
8. Log in to Gmail account using tagui browser which is used to send the NLB email

User Guide

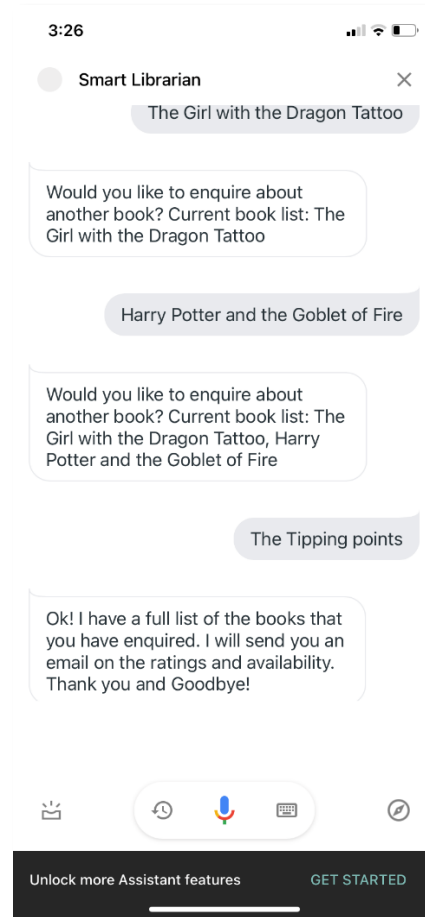
1. Run “ngrok.bat”. Extract https tunnel link
2. Paste https link into Dialogflow fulfillment. Press save.
<https://dialogflow.cloud.google.com/#/agent/librarian-qusu/fulfillment>
3. Run “run.bat”
Note: Webhook takes ~6mins to initialize the models
4. SmartLibrarian is ready for use!
5. **[Test Platform]** Try it out using DialogFlow test console or, download Google Assistant on your device, log in using respective credentials, and initialize the agent by typing/saying “Talk to Smart Librarian”
6. **[Email Recipient]** By default, email recipient is set to the Gmail account that is logged in. To change the email recipient of the results of the NLB search, go to “DialogFlow” folder and add in the email address in “email_recipient.txt”

APPENDIX IV: SMARTLIBRARIAN FEATURES AND SCREENSHOTS

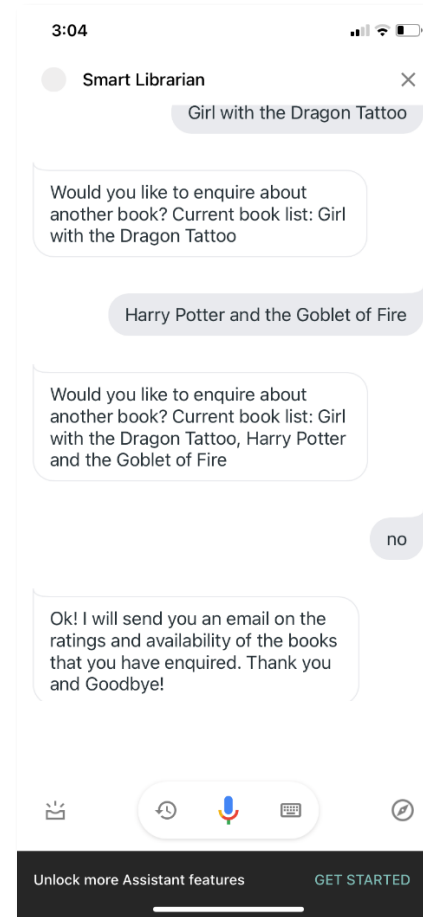
Google assistant on mobile



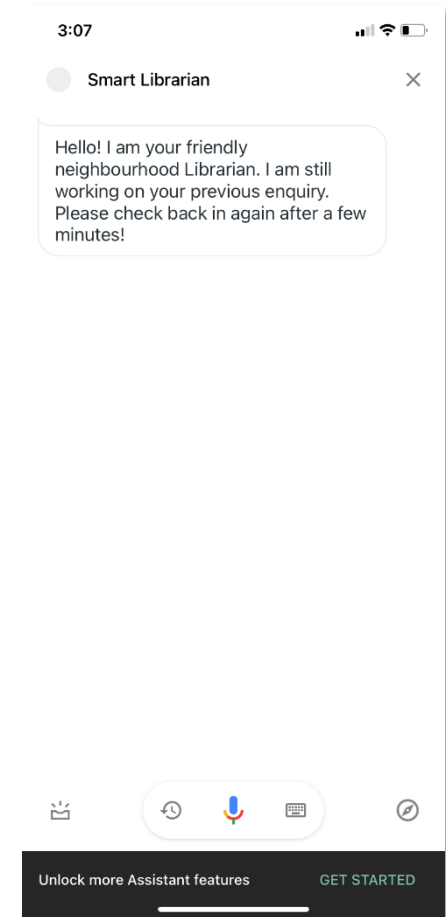
SmartLibrarian starts off with a simple introduction and asks user for a book title to search for



One way to start RPA for data acquisition is when 3 books have been requested by the user




Second way to is when user decides not to add on anymore book titles



If another request is triggered while the first enquiry is ongoing data acquisition, SmartLibrarian will not function

Email Layout

[SmartLibrarian] NLB Information for The Girl with the Dragon Tattoo

 **ISS IPA** <iss.ipa.2021@gmail.com>
to me ▾

Hello there,

BOOK 1: "The Girl with the Dragon Tattoo"

Subtitle: Millennium Trilogy, Book 1
Author: Stieg Larsson
Book Type: ebook
Ratings: 4.3
Punchline: Harriet Vanger, a scion of one of Sweden's wealthiest families disappeared over forty years ago.
Availability: 0 of 7 copies available
Link: <https://nlb.overdrive.com/media/158318>

Let's see what information Amazon has about a similar book!
Title: The Girl with the Dragon Tattoo: Book 1 of the Millennium Trilogy
Author: Stieg Larsson
Ratings: 4.5
Punchline: Harriet Vanger, a scion of one of Sweden's wealthiest families disappeared over forty years ago.

People who bought this book also bought these other books:
1: Girl Who Kicked The Hornet's Nest, The

BOOK 2: "Harry Potter and the Goblet of Fire"

Subtitle: Harry Potter Series, Book 4

NLB Information

Amazon Information

BOOK 2: "Harry Potter and the Goblet of Fire"

Subtitle: Harry Potter Series, Book 4
Author: J. K. RowlingOlly Moss
Book Type: Audiobook
Ratings: 4.7
Punchline: The Triwizard Tournament is to be held at Hogwarts. Then at Hallowe'en, when the Goblet of Fire is lit, the champions are chosen.
Availability: 50 of 185 copies available
Link: <https://nlb.overdrive.com/media/791486>

Subtitle: Harry Potter Series, Book 4
Author: J. K. RowlingOlly Moss
Book Type: ebook
Ratings: 4.6
Punchline: The Triwizard Tournament is to be held at Hogwarts. Then at Hallowe'en, when the Goblet of Fire is lit, the champions are chosen.
Availability: 39 of 165 copies available
Link: <https://nlb.overdrive.com/media/790959>

Let's see what information Amazon has about a similar book!
Title: Harry Potter and the Goblet of Fire: 04
Author: J K Rowling, Mary Grandpré
Ratings: 4.9
Punchline: If I knew what the book was about I would surely tell you.

People who bought this book also bought these other books:
1: Harry Potter and the Half-Blood Prince

Multiple Book Types in NLB