FEASIBILITY AND POTENTIALS OF AUTOMATED TRAVEL PLANNING

A THESIS
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
AND COMMUNICATIONS ENGINEERING
THE GRADUATE SCHOOL OF FUNDAMENTAL SCIENCE
AND ENGINEERING
OF WASEDA UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ENGINEERING

July 18th 2022

Chiahua Liu
5120FG49-5

Advisor: Prof. Tetsuya Sakai
Research guidance: Research on Information Access

Abstract

In 2020, global travel and tourism's direct contribution to GDP was around 4.7 trillion U.S. dollars. Travel industry is contributing significantly to countries' overall economy. Certain countries are so dependent on this industry, that over half of their GDP is resulted from this business. Furthermore, based on Satista research, the annual growth rate of the travel industry is approximately 5 percent, showing that it is an expanding business. However, without traveling with a tour group, planning a trip is a tedious and time-consuming process, requiring users to spend hours to do research. Even though the travel industry is a very old and developed field, there is less research on the automated travel planning area. This specific field has not been fully developed and has many future potentials. Additionally, I found out that many existing travel planning services are outdated and do not provide personalized services. This is unfortunate in the era when personalization is heavily valued. If an automated travel recommendation process can be successfully developed and put into practice, it could drastically change the general public's way of planning a trip. Thus this project focuses on exploring the users' acceptability to an automatic travel planning system. Also, this thesis will find out whether people are willing to use such systems, and it will examine people's attitudes and opinions when using an automated planning system. Based on our experiment results, we showed that most people are not satisfied with the current travel planning process, mainly because of it being too time-consuming, troublesome, and not personalized. Additionally, we found that most people are willing to use such a system, and they have high expectations of itineraries generated by algorithms. Many experiment participants suggested that the automatic planning feature is convenient and timesaving, and they prefer the models with automated features over the models without those functions. Through the experiment interviews, we found that settling the expectation-reality gap and the high variation of users' demands is critical for the model's success. Furthermore, I proposed a framework for an automatic itinerary generation system based on the findings in prior works and the experiment.

Contents

Chapter 1

Introduction

This chapter presents the motivation and objective of our research, discusses the research question that complies with the objective, and introduces the overall outline of the thesis.

1.1 Motivation and Objective

The travel industry is a large industry with worldwide revenue of 4.7 trillion U.S. dollars in 2020, as stated by the U.S. Bureau of Economic Analysis [1]. This industry contributes much to many countries' GDP; it accounts for around 3 percent of US GDP in recent years, larger than the agriculture, mining, or utilities industries' GDP contribution. Some countries are highly dependent on revenue generated by the travel industry; its revenue can contribute up to 75 percent of their total GDP. Traveling is also one of the best ways to step out of one's comfort zone, explore new things, and relieve anxiety. Other than consulting with a travel agency, travelers can also choose to plan the trip on their own. In this circumstance, a decent travel recommendation system becomes quite useful for travelers. In recent years, recommendation systems have become increasingly important in many industries. It can enhance user experience, expand services, and boost sales. Many large companies are implementing recommendation systems into their products. However, even though most people like traveling, based on our survey, some do not like the travel planning process. Without consulting with a travel agent, typical travel planning usually consists of searching for attractions, restaurants, hotels, and transportation, along with planning routes and checking the public reviews. The whole planning process, as shown above, proved to be quite time-consuming. We found out that more than 71.7% of the respondents spent over 2 hours on average planning a domestic trip in Taiwan. Furthermore, around 20 percent of the respondents have a negative attitude toward the travel planning procedure. In a world where many services are automated, the way people plan their trips has not changed much; we still need to manually do the research, pick the desired attractions, and schedule the routes. According to Sylejmani and Dika [2], the tourist life cycle consists of pre-trip, on-trip, and post-trip phrases. This thesis will focus on exploring the feasibility of automating the pre-trip phrase.

Rating and review platforms are often used during the travel planning process. The amount of information has grown rapidly as there are more online review platforms and services emerging, which are popular methods for searching and comparing attractions, restaurants, places to stay, etc. Travelers typically need to manually go through the reviews and ratings to thoroughly plan a trip. As mentioned previously, this is quite a time-consuming process. Is there a method that can accelerate the travel planning process? Is there a way that people can quickly be prepared for a vacation without the need to go through the time-consuming "research" process? We proposed that people are gradually able to accept an AI or algorithm to schedule their trips entirely, as people are more exposed to and dependent on AI assistants, such as Siri and Alexa. Since traveling is an old and established industry, there are many platforms that provide reviews and ratings for a number of Points of Interest (POIs)[1]; popular platforms include but are not limited to Google Map, Expedia, TripAdvisor, Yelp, and Hotels.com, etc. Hence, there are much information, such as customer reviews and ratings, for the trip planners to process. Sometimes, individual platforms have drastically different ratings and reviews, making it even harder to decide where to go.

Other than the overwhelming amount of information on the Internet, we also found another issue that is troubling the users when using those popular travel services; most of those popular travel services do not provide enough personalization for recommending POIs. As mentioned by one of our respondents, those travel recommendation services are overly focused on their generality but partially ignore user personalization. This further prolongs the entire travel planning process since users need to filter out the undesired travel spots. The objective of the thesis is to explore and discuss a novel way of travel planning, a new method of travel planning without the need to go through the tiresome "research" process. As there is more information supplied on the Internet, and the techniques of machine learning become more and more sophisticated, computer programs potentially can help people to plan their trips, saving

[1]Sylejmani and Dika's paper introduced the phrase Points of Interest (POIs)

travelers' time. We are going to investigate the feasibility of achieving such a trip recommendation model. Moreover, people's acceptability is also a concern toward this proposal. We need to examine people's attitudes, concerns, and trust toward an automatic travel planning service. Without people's attitudes being examined, developers might waste time developing an actual automatic travel planning application. One of the purposes of this research is to make sure that the public can accept this new way of travel planning. As mentioned previously, people have a high chance of accepting because of the extensive exposures to AI assistants in recent years.

Another objective of this project is to save travelers' time. In the modern world, many people gradually got used to a fast-paced lifestyle. There is no denying that time is one of our most important resources. If merging AI and machine learning into the travel planning process, it is possible to greatly reduce the overall time spent in travel planning.

1.2 Research Questions

This thesis discusses the possibilities and potentials for an automatic travel planning system. To answer this research question, we listed a few sub-questions that need to be answered. The sub-questions are the followings:

1. Is it possible to build a functional automatic itinerary generation system?
2. What approaches are proposed for similar projects?
3. Will people accept such novel ways of scheduling their trips?
4. What are the limitations and concerns when building such models?
5. What can be improved in the current travel planning services and platforms?
6. How to implement an optimal automated travel planning system that is suitable for typical travelers?
7. Is there room to improve from the current travel planning process?

These sub-questions need to be answered before solving the main research question. We first need to launch a basic survey and interview to validate a few presuppositions, making sure that the problem descriptions are legitimate. Second, we will go through literature reviews to discover previous attempts to resolve this problem. The reviews will then be used to compare this project. The findings of those related pieces of literature will also be used to validate and fine-tune the models in this thesis. Lastly, we need to launch an experiment to answer question three to question five. The experiment participants' feedback will give us enough insights to answer those questions.

1.3 Structure of this Thesis

This thesis is divided into six chapters; each provides different angles to tackle the problem. Chapter one gives readers a general description of this project's purposes, as well as the problem descriptions, research questions, and structure of this thesis. Chapter two is the related work chapter, where we will review several pieces of literature related to this topic. We will extract insights from those related works and compare them to this project. Chapter three will be the proposed method chapter. In this chapter, we will discuss the methods that will be used in the main experiment. This chapter will also discuss how and why the experiment is structured in this way based on the results of the pilot experiments. Chapter four will be the experiment results chapter, where the results of the main experiment will be analyzed and discussed. Additionally, utilizing the information gathered in the prior works and the main experiment, a model of an automatic travel planning system will be proposed. Lastly, in Chapter five, based on the results of the experiment, a conclusion will be made. The conclusion will suggest whether the hypothesis is valid or not. We will also provide a quick summary of the whole project.

Chapter 2

Related Works

In this chapter, elements and findings in previous academic papers are analyzed and discussed. One of the purposes of this section is to find differences and similarities between the past works and this project. Furthermore, this chapter will also help us answer research questions one and two, which is "[w]hat approaches are proposed for similar projects?". The findings in the previous works will also aid this project, helping us to formulate and fine-tune the experiments. This chapter will include previous research conducted mainly in Taiwan and the U.S. The papers are mostly written in English.

2.1 Previous Attempts in Building Automatic Planning Systems

After researching through academic platforms, we found a few studies that are related to this topic. In Sylejmani and Dika [2], the authors proposed a standardized method to evaluate travel planning platforms. They suggested that the automatic selection of POIs is an essential criterion for evaluating a travel planning system. In the paper, they spoke highly of the automatic POIs selection function, indicating that it offers significant assistance to tourists. The two main parts of the proposed evaluation framework are "planning functionalities" and "customization level". The planning functionalities are the functions that assist the tourist to prepare more personalized trips, and the customization level is the amount of given ability to choose between predefined options listed in a planning system. Furthermore, the paper selected and introduced a few travel planning services that were providing different levels of automation in their travel planning services.

In Chang, Chang, and Tsai [3], they proposed an algorithm called ATIPS that automatically generates a simple domestic itinerary based on various parameters, including user preference, popularity, cost, distance, and time. They first collected popular tourist sites online using a crawler, then stored the travel spots data in a large MySQL database. When a user initiates a travel planning process, it will trigger a sequence of spot selection algorithms, which go through the spot filtering, score estimation, and spot selection procedures. The authors also recognized that an automatic travel planning system can reduce the planning time and provide a smooth transition for travelers. Based on their observations, they stated that "[travelers] usually travel to a location and then find popular and interesting places nearby to visit while there or during the trip"; this coincides with our assumption of tourists having a common set of behavior, which can then be utilized to build a standardized automatic travel planning system. This research provided a framework for building an automatic travel planning system, which will be helpful in our future work.

In Lu et al. [4], the authors proposed a novel way of automatically planning a trip. They utilized the textual travelogs and geotagged photos from Google's Panoramio service to generate customized short-term travel itineraries. Using this method, they could quickly generate a large number of trip itineraries using the predefined routes provided by the geotagged photos. This could satisfy most tourists' needs since most travelers prefer and tend to visit the popular POIs. Furthermore, the authors provided user customization when scheduling a trip; the users could enter different constraints such as travel duration, travel location, visiting time, and destination preferences. The paper also provided an innovative algorithm for discovering tourist spots and routes from the geotagged photos. Lastly, based on their experiment results, their proposed algorithm and framework were effective in automatically generating personalized itineraries.

In Chen et al. [5], the authors suggested that the most current itinerary generation algorithm and service is only considering the most popular point of interest, lacking personalization and diversification. Thus, the authors proposed a novel algorithm that considered all POIs and heavily utilized users' preferences. They claimed that this problem was a team orienteering problem (TOP), which is an NP-complete problem. The whole problem-solving process went through a MapReduce function to preprocess the dataset, a parallel processing engine to index possible itineraries, a TOP transformation process into a set-packing problem, an approximate algorithm simulation for the set-packing problem, and finally an initialization-adjustment model to generate a result. The authors claimed that their algorithm could generate high-quality and accurate results.

In Friggstad et al. [6], the authors suggested that many previous algorithms suffer from two problems for multi-day trips, which are unbalanced itineraries and recommending an excessive number of low-tier POIs. Their solution is to maximize the trip quality of the itinerary's worst day. Using this method, the authors claimed that "[the algorithms] do not let the travel experience be downgraded by the existence of a full day visiting low-quality attractions" and "keeping daily itineraries local and not revisiting the same neighborhoods across days". Furthermore, they evaluated the model using both numerical and human rater experiments, getting positive results in both types of experiments.

De Choudhury et al. [7] is another research paper related to the automatic travel planning system. First, the authors acknowledged that the travel planning process is laborious, and they were also finding a method to reduce the travel planning time. They proposed an algorithm that utilized the "latent source reflecting geo-temporal breadcrumbs left by millions of tourists". The "social breadcrumbs" are photos or maps that contain certain geographical and semantic metadata. From those social breadcrumbs, they could extract essential information such as the locations, length of stay, transit time, and POIs' popularity. Then, using that information, they could generate itineraries based on users' input constraints. Lastly, through Amazon Mechanical Turk, they evaluated the model by comparing the generated itineraries with some popular professional bus tours in the same target cities. Based on 450 responses from the AMK, they concluded that their approach could generate high-quality itineraries that could match professionally made tours in certain cities.

2.2 Related Research Trends And Overviews

Based on these prior works, we have demonstrated that it is completely possible to build such systems, which answered our research sub-question two. Many researchers have tried various innovative methods to automatically generate itineraries. Automatically recommending POIs, which is one of the essential components of building an automatic itinerary generation system, has been developed for a long time. Throughout the past, researchers have been trying different algorithms to recommend travel POIs to travelers. Developers have implemented the matrix decomposition algorithms, clustering algorithms, nearest neighbor algorithms, neural networks, Naive Bayes, Gradient boosting algorithms, and collaborative filtering algorithms to tackle the travel POIs recommendation problem. In recent decades, deep learning, even though not specifically designed for travel recommendation systems, has gained popularity for solving this issue. Compared to the traditional machine learning algorithms presented in Foote [8] and Korbut [9], it proved to be accurate, flexible, and bear high robustness to natural variations. Because of its immense potential, more researchers are investigating its effectiveness on travel recommendation systems. Although traditional ML techniques are still performing quite well, deep neural networks are gradually replacing traditional ML algorithms in research directions and large service platforms. In Chetana et al. [10], the authors introduced several algorithms that are considered state-of-the-art. The authors first listed the RS techniques in figure 2.1, and they explained each technique's advantages and disadvantages. In that paper, a novel technique, Hybrid Collaborative Filtering, stands out from the rest. As mentioned by the authors, this technique combines the advantages of both the model-based approach and the neighborhood-based (memory-based) approach. It mitigates the issues when only using either of the two approaches.

The research directions for personalized POIs recommendation systems have not changed significantly. Based on our research, we found that researchers have been studying personalized POIs recommendations since the millennium year. Many innovative approaches have been developed and tested. In the earlier researches, scholars used collaborative filtering to make personalized POIs predictions. This approach assumes that groups of users have similar sets of preferences. One critical issue of CF is that it can not handle fresh items, e.g. the cold start problem. Other than that, CF provides outstanding performance and features in early personalized travel recommendation systems. In recent years, as mentioned previously, researchers have been slowly moving their attention toward building a deep learning model for personalized travel RS. Furthermore, scholars have tried various kinds of data to select the recommending POIs. The most common data source is the social media data, such as geo-tagged photos or travelogs. Scholars can derive some essential information from the data. For instance, if one can collect sufficient geo-tagged photos of an area, he or she can get much information from the POIs, such as their levels of popularity, locations, lengths of stays, sequences of visits, etc. With the information, a basic travel

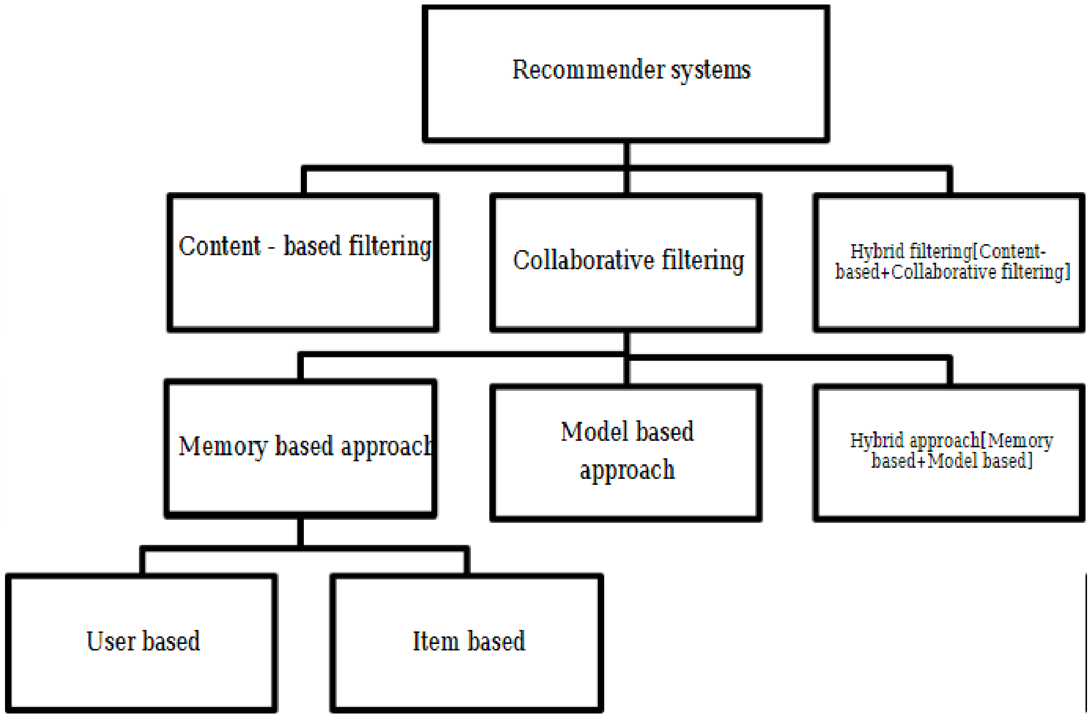

FIGURE 2.1: Common recommendation system techniques, provided by Chetana and Ibrahim

recommendation system can easily be built based on POIs' popularity. Other than using social media data, some research asked professional tourists to comment and tag the POIs. Then, the model used this metadata to recommend POIs to travelers.

2.3 Research Comparison

As we can see, researchers have been studying and developing travel recommendation systems for many years. Most of them focused on developing innovative algorithms and utilizing different data sources. However, there were less research on the actual users. Since the users are the ones who are using the end products, it is important to study their behaviors and opinions as well. Studying users' opinions can help developers develop products and services that more closely fit the users' demands. Furthermore, most past research focused on recommending POIs based on users' preferences in a certain location, somewhat ignoring the itinerary construction action, which proved to be one of the most time-consuming procedures in the whole travel planning process. During the itinerary construction process, the travelers need to select the POIs, decide the travel sequence, arrange the travel route, determine the length of stay, etc. It is possible to shorten this process by using an automatic itinerary generation model. Recommending an itinerary instead of suggesting POIs is a relatively novel idea for travel recommendation systems. If the model can successfully be applied to some travel services or websites, it can bring significant convenience to the tourists, and it will greatly change the way people plan their leisure trips.

Chapter 3

Proposed Method

This chapter of the thesis will introduce the proposed method for testing the hypotheses. It will also explain the design train of thoughts in detail; the experiment is designed based on several pilot experiments. The experiment models, survey questions, and interview contexts will be displayed in this chapter of the thesis.

3.1 Pre-Experiment Thought Process

Before the main experiment, we have done a set of smaller experiments; their purpose is to adjust and guide the main experiment's questions and structure, acting as a research pivot; also, we need to validate our basic hypothesis, to construct a more in-depth experiment. One of the smaller experiments is a general survey regarding travel behaviors. This survey used both English and Traditional Chinese. The questions are listed in the following,

- Have you ever planned a trip (with family or friends)? (MC)
- In the past 6 month, have you used any travel planning app? (MC)
- Which aspect do you mainly consider when deciding an attraction/restaurant to visit? (MC,SA)
- Average time spent in trip planning? (MC)
- Used travel applications? (MC,SA)
- The degree of personalization in the current travel planning app? (MC)
- How satisfied with the above travel planning app? (MC)

- Please elaborate on your choice of the last question. (SA)
- Travel planning is tedious? (MC)
- Route planning part of the travel planning is the most troublesome part (MC)
- When planning a trip, which is the most time consuming part? (finding attractions, checking ratings/comments, finding hotels, route planning, etc) (MC,SA)
- Have you ever felt puzzled about where to visit in an unfamiliar city because you did not plan a trip beforehand? (MC)
- During a trip, in terms of saving time and money, degree of difficulty to plan the optimal route of the whole trip? (MC)
- The level of trust to AIs? (Such as Siri, Alexa, Google Assistant..etc) (MC)
- Would you accept an AI automatically planning your trip? (MC)
- Do you think that automatic trip planning is a novel procedure? (MC)
- Do you trust AI's choice of attractions/restaurant/route? (MC)
- In terms of contents, level of expectation from AI's travel plans? (MC)
- Would you like your AI generated plans very personalized or just follow the overall public's preference? (MC)
- What is the most important feature of a travel planning app? (MC,SA)
- Please describe your ideal travel planning app. (SA)

As can be seen, this survey contains multiple-choice questions (MC), short answer questions (SA), or a combination of both. The purpose of this survey is to learn people's habits and opinions while planning a trip. Our conclusion from this survey is that most travelers spent a lot of time manually researching on POIs and constructing itineraries, and they found the idea of automatic travel planning attractive and novel. For this survey, there are in total of 51 respondents; the respondents' ages range from fourteen to thirty. Notably, over 72.5% of the respondents spent over 2 to 4 hours planning an ordinary trip, and over 37.2% of them spent over 4 to 6 hours in planning. Based on the results, the top three most spending travel sub-tasks were selecting for

POIs, route planning, and picking hotels, respectively. This shows that our assumption of "travel planning is a time-consumable process" is valid. They might spend even longer time planning a more lengthy trip. The majority of the respondents had a situation where they felt puzzled during a trip because of not planning ahead. This shows that they recognize the importance of planning travel ahead. Most of them (84.3%) used Google Maps to plan their trips, which means that they did not utilize any application that was equipped with more sophisticated recommendation systems; this might be one of the reasons why travelers spent so much time in travel planning. Furthermore, we found an interesting opinion about travelers' attitudes toward the planning process. Even though they did not think that the whole travel planning process is tedious (64.7%), many of them felt that parts of the planning process were troublesome (43.2%), such as route planning and researching for POIs. Over 39.2% of the respondents agreed that optimizing routes between POIs, or deciding the optimal travel sequence, was difficult. Optimizing travel routes might be a tough task for travelers. Furthermore, we also found that many people select POIs based on their feelings (49%) instead of utilizing statistical data such as online ratings and reviews. We found that travelers use Google Maps as their only travel planning tool; based on respondents' comments, we believed that this was caused by Google Maps' wide and general features. Even though Google Maps does not offer very personalized recommendation features, it provides a wide range of features, such as showing routes, comments, ratings, POIs, hotels, prices, and a lot more information. Users believed that it was convenient even though it did not provide personalized recommendation services. Thus, when developing applications, it might be a good idea to link some functions to Google Maps or to utilize the information provided by it.

Other than exploring users' habits in planning a trip, we also studied people's opinions toward an automatic travel planning system. We found that most travelers held a neutral attitude toward this automated system while they suggested that they generally trusted commercial AI such as Siri and Google Assistant. 41.2% of the respondents showed that they trusted AI such as Siri while only 21.5% of them would accept an AI automatically plan a trip for them. A potential reason for this phenomenon is that people do not believe that these AI can handle such complicated tasks. This is comprehensible since planning a trip consists of many sub-tasks, such as researching for POIs and route planning. We found that respondents generally trusted that AI can handle the sub-tasks of the whole travel planning process, but they held a slightly negative attitude toward an AI planning the entire trip. This suggests that people recognized the complexity of trip planning, and they doubted AI's ability in completing the complicated information integration tasks. However, based on our research, these sets of tasks are often fixed, which means that the procedure and structure of travel planning do not change much. With the help of current machine learning techniques, it is completely possible to formulate algorithms to handle these fixed tasks. Although they generally held a neutral attitude toward automated systems, 54.9% of them found this idea novel and interesting. Also, they had a broadly high level of expectation from AI-generated travel plans; over 45.1% of the respondents looked forward to such features. The respondents preferred very personalized travel plans over plans with only the most popular POIs. This means that the ultimate application needs to take care of providing personalization of the service. In the last question where we asked "please describe your ideal travel planning app" to the respondents, 20 out of 51 respondents commented that it would be convenient if an application can automatically schedule the routes or generate an itinerary based on certain personal criteria. This shows that an automated travel planning system has the potential to become people's ideal choice when planning trips if it is implemented correctly. But what features will be essential for such a system? The main experiment will try to answer this question.

3.2 Main Experiment Method

The experiment is separated into two main parts; we called them the online group and the interview group. Even though they used the same experiment models and questions, they have different purposes. The online group's main purpose is to collect travelers' opinions and thoughts on the automatic itinerary generation system. Their responses will be collected and analyzed using statistical techniques. We can also find out respondents' general attitudes toward this system. The interview group's purpose is to observe users' behaviors while operating the experiment models, collecting data like the model's usability and user interactions. Furthermore, since a face-to-face experiment is used for the interview group, we can ask more detailed questions, such as their thoughts, concerns, and recommendations. In general, even though some of the questions between the two groups are similar, we can obtain different types of data from the two.

Due to the widespread COVID-19 virus, we chose to use an online "experiment package" to be our main method to perform the online group experiment; with this method, we can potentially get more respondents. The experiment package includes a few operable Pidoco[1] prototype models, an all-in-one google forms survey, and an experiment guiding document. We provided both English and Traditional Chinese versions for all of these materials. The participants need to follow the instructions specified on the guiding document to finish the experiment. Pidoco is an online tool that allows developers to create interactive low-fidelity prototypes and clickable wireframes. It allows us to create several prototype models without the need to implement all the functions. Even though the prototype models do not contain all the functions of a fully functional prospective application, it is sufficient to validate our hypothesis. To attract people to participate in this experiment, we prepared twenty-five $200 7-11 vouchers as gifts, and we will randomly select the lottery winners from the participants. The online group experiment took 15 days to collect the responses.

In the guiding document, there are in total of five sections. A snippet of the guiding document will be shown in figure 3.1. The first section prompts players to enter their basic information, such as names, email, age, professions, marriage status, etc. Experiment sections two to four are the main part, where the participants would operate the Pidoco models for five minutes and answer the questions. This means that one Pidoco model will be paired with a set of survey questions, in a total of three Pidoco model-survey pairs. We will also further explain the Pidoco models in the following paragraphs. In section five, the survey would prompt participants to answer a few more general questions related to a travel recommendation system. Participants would spend around fifteen minutes finishing the entire experiment process. We sent out the experiment package on several public media platforms, such as Facebook, Instagram, Line, etc. The prospective participants mostly come from Taiwan. To avoid biases in the responses, we made some changes to the experiment package. First, in the survey questions' options, if the options are non-ordinal, the options will be randomly presented to the participants when they are answering the survey. Furthermore, we transformed the experiment package into three versions to avoid sequence bias; each version used a different sequence for presenting the three Pidoco models. Even though it is unlikely to get a perfectly even amount of responses in each package version, this method will at least mitigate the response sequence bias. Furthermore, since we are not able to ensure participants' seriousness while filling out the survey, we insert a few small quizzes in the survey questions. For instance, in the same section, there might be multiple questions that are asking about the same subject. If the

[1] https://pidoco.com/en

participants answered differently in those questions, we would consider the response invalid. Also, if a response was suspected to be illegitimate, that response would be dropped, ensuring accurate results. The Pidoco models used in this experiment are similar to mobile application prototypes; the main purpose of using Pidoco models is to evaluate certain design ideas and application functions. The models simulate applications with or without certain functions, such as the ability to automatically generate an itinerary or the ability to optimize routes. An example of our Pidoco model can be seen in figure 3.2. The participants can perform some operations on the models, such as dragging, clicking, browsing, etc. The models are not fully functional since the models do not contain actual recommendation algorithms; only certain functions and buttons are implemented. This means that the models have predefined pages to show to the participants, and the users' input would not influence the model recommendations or outputs. The model uses graphs and text to simulate the proposed functions, such as the ability to automatically generate routes and itineraries. Even though the models are not actual working applications, they are sufficient for validating our hypothesis.

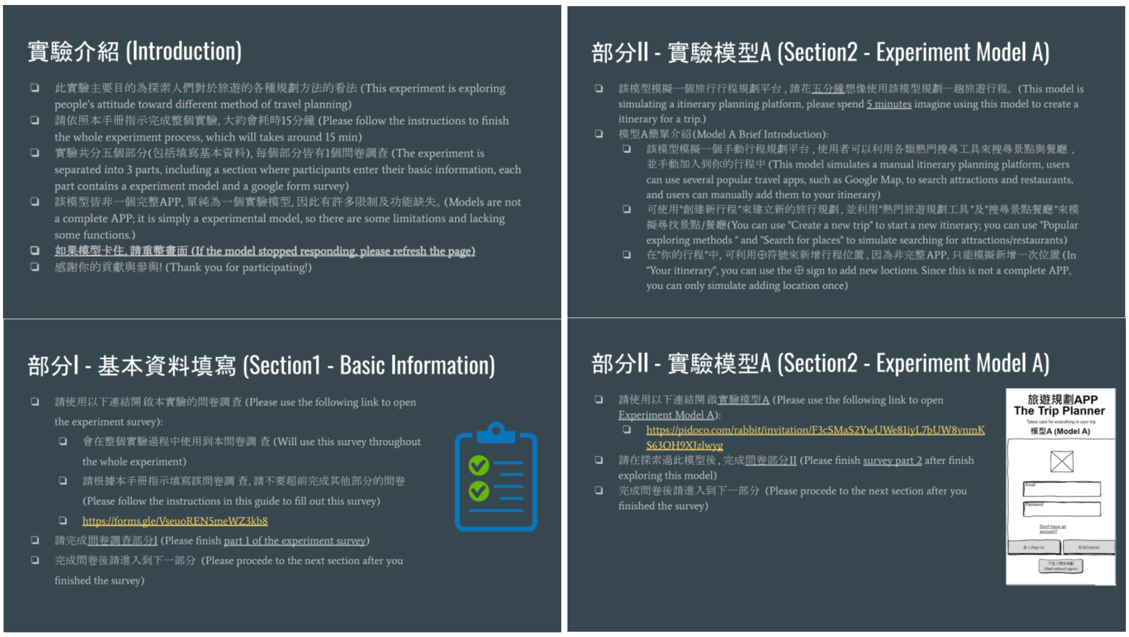

FIGURE 3.1: A snippet of the experiment guiding document

There are in total three groups in this experiment, which are the control group,

FIGURE 3.2: An example of the Pidoco model

the experiment group, and a special semi-experiment group. The control group represents an application that contains no automatic planning functions; this is also how most people plan their trips, i.e, only using the most popular tools like Google Map to find the POIs. In this model, users need to manually search for POIs, add them to the itinerary, select desired start and end times, and decide the travel sequence. The second group is the semi-experiment group; the model is simulating the interface of a common travel recommendation application. We specially added this group to get more information about trip planners' opinions when using typical travel recommendation APPs. The model will automatically recommend POIs based on users' preferences; the users only need to click and select the POIs, and the selected travel spots will be added to the itinerary. However, this model does not provide fully automatic itinerary generation. The users still need to manually enter the length of stay, decide on travel sequence, and optimize routes. The last model is the experiment model, which is the main proposed model in this research. Compared to the other two models, the model presents the ability to provide fully automatic itinerary generation based on users' preferences. When initiating an itinerary generation process, the users need to first enter their "trip settings", such as starting location, ending location, time, date, trip purpose, number of people, desired theme, trip pace, etc. After selecting the options, the model will prompt the user to select a few restaurants and POIs that he or she desires. The model then should use this information to create an itinerary; the travel sequence, length of stay, and routes will be adequately arranged, ensuring a smooth and optimized trip. Even after the itinerary is generated, the users still can change its contents, such as POIs, sequence, etc. This feature provides some level of malleability and control to the users.

In Table 3.1, we listed the questions that we asked the respondents; those questions mainly are related to their basic demographic information, such as their age, profession, and whether they have planned a trip. In Table 3.2, this set of questions is paired with the Pidoco models; each model will have a set of these questions. The purpose of these questions is to find out people's opinions, ratings, and attitudes toward a specific model. These questions mainly answer research questions 3 and 5. Table 3.3 are the additional questions asking some general questions about the three models. Some questions are short answer questions, and most of them are 5-points Likert Scale questions, in which options 1 to 5 are "strongly disagree", "disagree", "neutral", "agree", "strongly agree", respectively.

Similar procedures are also used in the interview experiment group. Due to the pandemic, we also chose to conduct this part of the experiment online. The interview experiment group's main purpose is to answer research questions 3, 4, and 6. We host the interviews mainly on Zoom, Skype, and Discord. During the experiment, if the participants were willing to do so, we asked them to open their webcams and to share their screens. In this way, we could observe their behaviors when they were operating the models. After they were ready, we would ask them to open the same guide document that was used in the online experiment group. They needed to follow the instructions on the document and finish the survey. While they were filling out the survey, we would observe their behaviors and record the arising questions. After they finished the survey, we would ask them about their thoughts, opinions, and recommendations of the models. Their responses would then be recorded.

TABLE 3.1: Main experiment Section 1 Questions

Q. No.	Survey Question	Research Q.
1	What is your age? (MC)	N/A
2	What gender do you identify as? (MC)	N/A
3	Are you married? (MC)	N/A
4	What is the highest degree or level of education you have completed? (MC)	N/A
5	Employment status/profession (MC)	N/A
6	Do you like traveling? (MC)	N/A
7	When going out with a group, do you like to be the one planning the trip?	(MC) N/A
8	In the past 2 years, have you ever planned a trip (with family or friends)? (MC)	N/A N/A
9	In the past 2 years, how many trips/travel have you planned? (MC)	N/A
10	Do you agree that travel planning is a time-consuming process? (MC)	7
11	Average time spent in trip planning (MC)	7
12	Do you have the habit of using AI assistants (Siri, Google Assistant)? (MC)	3
13	What tools/services do you use to plan a trip or explore attractions and restaurants? (MC,SA)	N/A
14	With the help of modern technologies, do you believe that there are more efficient methods to plan trips? (MC)	3

TABLE 3.2: Main experiment survey section 2-4 questions

Q. No.	Survey Question	Research Q.
1	What is your level of understanding of the model's functions and purpose? (MC)	N/A
2	Based on the model's usability, convenience and its functions, what is your overall rating for this model/APP? (MC)	3
3	With the help of modern technologies, do you think that it is possible to create this model/app? (MC)	3
4	Do you think that this app/model is useful for your travel planning process? (MC)	3
5	If the model is fully functional, will you use this model/app to plan your trip? (MC)	3
6	The level of personalization in this model/APP	3
7	Can this model/APP save your time in the travel planning process? (MC)	3
8	This model will learn users' preferences and recommend attractions/restaurants locations. Do you like this mechanism? (MC)	3 3
9	This model will automatically decide travel sequences and optimize routes based on your selections' locations. Do you like this mechanism? (MC)	3 3
10	Do you believe that algorithms can learn your preferences from your travel spot selections? (MC)	3
11	Do you believe that algorithms can generate a satisfactory travel sequence and route for you? (MC)	3
12	Do you trust AI's choice of attractions/restaurant/route? (MC)	3
13	Do you think that this model has enough usability? (MC)	5
14	Do you think that this model has enough personalization? (MC)	5
15	Do you think that this model gives enough freedom to users? (MC)	5
16	In this travel planning simulation, which part is the hardest for you? (MC)	5
17	In this travel planning simulation, which part is the easiest for you? (MC)	5
18	In your perspective, what functions are missing from this model? How to improve this model? (SA)	5

TABLE 3.3: Main experiment Section 5 Questions

Q. No.	Survey Question	Research Q.
1	If you are going to plan a trip, which simulation model/app will you use? (MC)	3
2	The reason for your choice in the last question? (SA)	3
3	Do you trust the travel itinerary generated by AI/algorithms? (MC)	3
4	What do you think is the most important function for a travel planning service/platform? (MC)	6
5	Do you think that automatically generating itineraries is a novel idea? (MC)	3
6	In terms of contents, level of expectation from AI's travel plans (MC)	3
7	If Model C successfully developed into a functional service/app, will you consider using its automatic itinerary generation function? (MC)	3
8	The reason for your choice in the last question? (SA)	5
9	Please describe your ideal travel planning app/service (SA)	6
10	Overall suggestions and thoughts? (SA)	N/A

Chapter 4

Experiment Results

In this chapter, we will list and analyze the results gathered in the main experiment. We filtered out the invalid responses and analyzed the data using statistical tools. We will also list the findings of the interview experiment group. Lastly, an automatic itinerary recommendation system's general model structure will be proposed. The model structure will contain a blueprint of a travel RS designed by utilizing the experiment results.

4.1 Online Group Results and Analysis

As mentioned previously, the experiment package is separated into three versions with different question sequences to avoid bias. For the convenience of statistical analysis, we merged answers from the three surveys into a single dataset.

Table 4.1 is the average scores from responses in survey section 1, using the evaluation method provided in Decker [11]. These questions are the same across all three versions of the experiment. Section 1 is mainly obtaining respondents' basic demographics, such as age, profession, and gender. It also helps us select the respondents who have planned a trip recently. The response "scores" are results from the 5-point Likert scale questions; score 1 usually means "strongly disagree" and score 5 means "strongly agree". For the online experiment group, there are in total of 28 respondents; after filtering out the invalid responses, there are 21 responses for this experiment. The remaining respondents all have at least planned a trip in the past two years and held a positive attitude toward traveling. The respondents' ages range from 12 to 65, with an average of 28.2 years old. 57.1% of them are male and 42.9% are female. In the past two years, on average, the participants planned their trips 2.78 times. Also, based on

TABLE 4.1: Survey section 1 results

No.	Survey Question	AVG. Score
1.	When going out with a group, do you like to be the one planning the trip?	3.17
2.	In the past 2 years, how many trips/travels have you planned?	2.78
3.	Do you agree that travel planning is a time-consuming process?	3.83
4.	Average time spent in trip planning	4.45
5.	Do you have the habit of using AI assistants (Siri, Google Assistant)?	2.22
6.	With the help of modern technologies, do you believe that there are more efficient methods to plan trips?	4.28

the responses, the participants agreed that travel planning is a time-consuming process, with an average score of 3.83; they spent 4.45 hours planning a single trip on average. Furthermore, most of them agreed that there are more efficient ways to plan trips, which means that there is room to improve the current travel applications or services; the score for this subject is 4.28.

Table 4.2 is the average score comparison between the three models. The statistical significance is marked beside the average scores. We conducted ANOVA tests to evaluate the statistical significance of the scores. There is no significance code marked besides Model A's scores because it is used as the baseline. For instance, the significance code marked in Model B is computed using the values in both Model A and Model B; similarly, the code marked in Model C is computed using the values in both Model A and Model C. After the ANOVA test, we conducted the Tukey's Post Hoc Test, but the pairs showed no significance. This may be caused by the relatively small sample size of this experiment or the limited options and structure of the 5-point Likert scale. Other than the Tukey's test, based on the ANOVA test results, we can see that most of the Model B and C's average scores are statistically different from Model A's.

As mentioned before, Model A is the control group, mainly simulating an ordinary application without any automated features or personalized recommendations; users needed to manually search for POIs and insert them into the itinerary. Model B is the semi-experiment group; it is simulating applications that provide some level of automation and personalization. In this case, Model B demonstrates that it can recommend personalized POIs based on users' preferences, but parts of the whole planning process are not automated. For instance, users still needed to decide the length of stay, travel sequence, and routes manually. Model C is the main experiment group. It simulates a fully automatic travel planning system. It takes users' preferences and constraints as inputs and the model will generate a full itinerary. Users could still modify the contents of the itinerary; they could also choose to use the itinerary directly during their trips.

In this table, we can see that most of the respondents understand the models' functions and purposes, getting scores of around 4.1 in all three models. For the overall model rating question, the participants rated Model A a score of 3.5. Model B and Model C slightly outperform Model A by 0.11 and 0.22. This shows that the participants gave slightly higher ratings in Model B and Model C. Also, slightly more respondents thought that Model B and Model C are more useful than Model A. Other than that, based on the average scores in the table, we can see that Model B and Model C slightly outperform Model A in most of the fields, including personalization, convenience, usability, and control. Furthermore, Model B and Model C's scores are very similar, meaning that they have matching performances. Model A, even though slightly underperforming the other two models, still gained positive ratings in most questions. In a nutshell, Model B and Model C, the models with automated features, gained slightly higher popularity in most fields compared to Model A did.

Other than the Likert questions in Table 4.2, there are also some short answers and multiple-choice questions regarding the three models. For the question "in this travel planning simulation, which part is the hardest for you", respondents answered "optimize travel routes" the most in Model A, "researching travel spots" the most in Model B; notably, in Model C, none of the options' responses stand out. For the question "in this travel planning simulation, which part is the easiest for you", respondents answered "finding attractions /restaurants" and "deciding visiting sequence" the most in Model A, "finding attractions/ restaurants" the most in both Model B and Model C. This shows that finding POIs is usually the easiest part with or without the automated features. Other than that, we found an interesting shift in the selected answers. In Model A, only 14.3% of the participants suggested that "making the itinerary" is the easiest. However, in Model B, 23.8% of the people thought that making the itinerary was the easiest part. Additionally, in Model A, only 9.5% of the people chose "making the itinerary" as the easiest, but in Model C, 23.8% of the people chose that as the easiest part of travel planning. These two changes in the "easiest part" coincide with the models' added features. Model B, compared to Model A, added the feature to assist users in quickly constructing the itineraries. Likewise, Model C added the feature to optimize travel routes. These phenomenons suggest that the participants noticed and liked the added features in Model B and C.

There are some questions asked specifically for Model C. The questions can be found in Table 4.3. These questions are mainly evaluating respondents' trust and preferences toward AI and algorithms handling their trips. We can see that the average scores of these questions range from 3.94 to 4.33, which approximately means "agree". From these results, we can first see that respondents liked the model's ability to learn users' preferences and to automatically construct the itineraries. Second, the results show that the participants saw algorithms and AI as trustworthy assistants. The majority of them trusted their abilities to perform various travel planning tasks.

Questions in section 5 of the survey were mostly assessing travelers' opinions and thoughts on an automatic travel planning system. Some questions would also ask participants to leave a reason or a comment. In the question "If you are going to plan a trip, which simulation model/app will you use", 42.9% chose Model C, 19.0% chose Model A, 19.0% chose Model B, and 19.0% chose "neither". This data demonstrates that the majority of the participants liked the automatic itinerary generation process in Model C. There was a follow-up question right after this question, asking "The reason for your choice in the last question?". For the respondents who chose Model C in the last question, many of them suggested that the methods in Model C are more convenient and can construct an itinerary more quickly. Also, we found that 19% of the respondents stated that they like the detailed travel options (users' constraints options). This suggests that users prefer high controllability in using travel planning applications. Similarly, a few participants stated that they like the personalized trip; the automatic preference learning feature and the detailed user constraints allow users to construct personalized trips while using the automatic itinerary generation feature. For the question "do you trust the travel itinerary generated by AI/algorithms?", 7 chose "strongly believe", 9 chose "agree", 5 chose "neutral", and none of them chose "disagree" and "strongly disagree". This result concurs with the responses to the

TABLE 4.2: Survey responses averages regarding models' usage

No.	Survey Question	Model A	Model B	Model C
1.	What is your level of understanding of the model's functions and purpose?	4.00	4.11***	4.17**
2.	Based on the model's usability, convenience and its functions, what is your overall rating for this model/APP?	3.50	3.61*	3.72*
3.	With the help of modern technologies, do you think that it is possible to create this model/app?	4.61	4.56***	4.50***
4.	Do you think that this app/model is useful for your travel planning process?	3.89	4.06*	4.00
5.	If the model is fully functional, will you use this model/app to plan your trip?	3.67	4.06.	4.00..
6.	The level of personalization in this model/APP	3.72	3.78.	4.11
7.	Can this model/APP save your time in the travel planning process?	3.50	3.67*	3.61.
8.	Do you think that this model has enough usability?	3.28	3.67	3.72
9.	Do you think that this model has enough personalization?	3.72	3.83*	4.11**
10.	Do you think that this model gives enough freedom to users?	3.94	3.89	3.94

Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '..' 0.1 '.' 0.2

TABLE 4.3: Survey question specifically for Model C

No.	Survey Question	AVG. Score
1.	This model will learn users' preferences and recommend attractions/restaurants locations. Do you like this mechanism?	3.94
2.	This model will automatically decide travel sequences and optimize routes based on your selections' locations. Do you like this mechanism?	4.11
3.	Do you believe that algorithms can learn your preferences from your travel spot selections?	4.06
4.	Do you believe that algorithms can generate a satisfactory travel itinerary for you?	4.00
5.	Do you trust AI's choice of attractions/ restaurant/route?	4.06

Model C questions; the results show that the participants generally trusted the algorithms' ability to properly handle their trips. Similarly, the same concept applies to other questions in Table 4.4; the results from these questions further confirm that users trusted AI's performances. In the question "what do you think is the most important function for a travel planning service/platform?", we found various kinds of answers related to travel planning. The most common answers were "convenient", "provide POIs quickly and accurately", "recommended itineraries", "assisting travel research", and "save money". These aspects should be carefully considered when building an actual travel recommendation application. We also found that users had high expectations from the automatically generated results. Moreover, in the question "if Model C successfully developed into a functional service/app, will you consider using its automatic itinerary generation function?", none choose "strongly disagree", 2 chose "disagree", 6 chose "neutral", 9 chose "agree", 4 chose "strongly agree". This shows that 61.9% of the participants were willing to try out an automatic travel planning service. In the question "please describe your ideal travel planning app/service", we found some notable responses like "finish planning a trip in 30 minutes and don't need to use too much brain power", "automatically recommend travel spots and decide the travel sequence for me", "recommend the best route for the trip, recommend good travel spots, and be able to change the itinerary when the situation changes", "the application should be able to take care of the travel budget", "the application should display

TABLE 4.4: Survey section 5 results

No.	Survey Question	AVG. Score
1.	Do you trust the travel itinerary generated by AI/algorithms?	4.10
2.	Do you think that automatically generating itineraries is a novel idea?	4.06
3.	In terms of contents, level of expectation from AI's travel plans	4.06
4.	If Model C successfully developed into a functional service/app, will you consider using its automatic itinerary generation function?	3.89

all the information at once", "an application that can generate an itinerary quickly", "contain other people's shared itineraries, and I can copy those itineraries directly". The responses demonstrate that the users demand an application that can quickly construct itineraries, and the users want the application to have high customizability and elasticity. The responses also confirm that people have various demands, such as budget control, money-saving, collaboration, quickly planning trips, POIs comparison, application integration, accurate alarms, information gathering, etc.

Overall, most participants liked the concept of automatically generating itineraries. Most of them thought that the automatic travel recommendation system is a novel idea, and almost none of them have tried such service or application before. They had high expectations of its features and were willing to try out if the application exists. Also, the results show that most of them trusted the algorithm's ability to handle such tasks; only a few respondents did not believe that algorithms can truly assist their trip planning process. Furthermore, the majority of the respondents thought that Model C, our most automated model, was the best model out of the three models. And many of them mentioned that the mechanism in Model C could help them reduce travel planning time. In other words, users saw that mechanism as convenient and time-saving. Additionally, in some short answer questions, a few participants specifically mentioned that "quickly generating itineraries", "saving my time during the planning process", and "helping me gather and process information" as essential features of their ideal travel applications.

4.2 Interview Group Results and Analysis

In addition to the online group experiment, we also conducted the interview group experiment to obtain more detailed data. As mentioned previously, the interview group experiment asked the participants to go through the same experiment process in the online group, but we would observe the participants operate the models and fill out the survey questions. For this experiment group, there are in total of 7 participants, and their ages range from 23 to 55. All of the participants have earned Bachelor's Degrees or higher. And more importantly, all of them said that they like traveling.

In this paragraph, we will focus on participants' recommendations and comments instead of the numerical data presented in the previous paragraph. Thus, we asked the participants to elaborate on questions like "if you are going to plan a trip, which simulation model/app will you use?", "what do you think is the most important function for a travel planning service/platform?", "if Model C successfully developed into a functional service/app, will you consider using its automatic itinerary generation function?", and "overall suggestions and thoughts?". These questions mainly come from section five of the survey. In the first question, the majority of the participants stated that the automatic features in Model C are convenient, and they suggest that this feature can help them reduce travel planning time. A participant chose Model A in this question; he said that "planning the travel myself will always be faster than using travel recommendation applications". This shows that some people might want to have full control when planning a trip, or a few people simply do not trust that the application can handle such a task. Another participant said "I would like this feature, but there are simply no such applications out there currently". This suggests that either there is a limited amount of such service, or the participant "does not believe that there is such service". This response also shows that people are generally willing to try out an automatic travel service, but they could not find such services easily. Also, a respondent stated that "selecting a restaurant for a trip may be more complicated than you think" This shows that travelers' demands may be complicated, and it might be difficult to include all the options in an application. This is a downside of AI trip planning since it is not possible to include every options and preferences. When building an actual application, this should be carefully managed. Another participant, who had planned numerous trips for a large organization, mentioned a similar matter. She stated "people's preferences are complicated, and I believe that it is impossible to include all the people's preferences in an application". This statement concurs with another participant's response. This means that application developers need to find another method to make up for this downside. Regarding Model C's users' usability, we found some of the participants stating that "it is more fluent using the model", "it is more convenient to use", "the flow of the application is more clear", "it is a novel design". But we also receive comments like "there is room for improvement", "it is a bit confusing on what to do next", and a participant even specifically said that "the UI and UX need improvement". Overall, the participants suggested that the usability is "good" and "usable", but there is room for improvement.

I had a detailed conversation with one of the respondents; he is quite experienced in self-planned custom trips, and he showed high interest in this system. For the question, "will you use an automatic itinerary generation system for your trip", he said that he would use it if the system has high customizability and changeability. He suggested that it is important to be able to modify the itinerary after its generation since there will always be unexpected events during a trip. This shows that the application should include a route recalculation feature and numerous kinds of user constraints. Also, for the question "do you believe that algorithms can generate a satisfactory travel itinerary for you", he stated that "sophisticated algorithms definitely can achieve such a task, but it is different in real-world scenarios, it is unrealistic to have so many options in an application to cover all of my demands". Thus, he said that it is possible to generate a satisfactory itinerary, but generating a perfect itinerary is "difficult to achieve". He also said that "when using this application in unfamiliar places, even though it cannot perfectly match my ideal choices, it can still speed up my itinerary creation process". He suggested that the application is not perfect, but if it is built properly, it can still speed up the travel planning process. Moreover, for the question "do you trust the travel itinerary generated by AI/algorithms", he stated "I have high expectations from AI-generated itineraries", "sometimes algorithms know you better than you do", and "there are so many travel platforms and database; algorithms and crawlers can always search faster and better". Similar to other participants, he had high expectations of AI's ability; he thought that algorithms can do better than humans since it has access to a large amount of information and can "think" faster. To conclude this discussion, he had high expectations of the algorithm-generated itineraries, but he had several concerns about this itinerary generation method. He suggested that it is unrealistic to meet everyone's demands, so the itineraries are deemed to be "imperfect"; however, even though they are not perfect, they can still assist his travel planning by giving him usable "itineraries recommendations".

In a nutshell, we found a few interesting opinions and concerns about an automatic travel planning system in this chapter. First, in both online and interview experiment groups, people generally trusted the ability of algorithms and AI; they expected that algorithms should always perform well, and evidence shows that they were willing to try out these kinds of services. We also found that people had high expectations of the AI-generated itineraries. This is not necessarily a good thing for developing an automatic travel planning application. High expectation means that the models need to be exceptionally good and can meet most of their demands, but in reality, this is hard to achieve since it is unrealistic to handle all the preferences. It is not that the model is incapable of learning all the preferences parameters but there is simply not enough space in an application to prompt for all the preferences options; it is also unreasonable and unrealistic to do so. This raises a serious issue for this kinds of services, especially when this is a novel idea for travelers. The discrepancy between the users' expectations and reality may cause the users to stop using such services on their first try. When a user tries out this automatic planning service, based on our results, he or she might generally have high expectations of the generated results. Due to limited options and arithmetic errors, the generated itinerary might not be able to match his or her expectation. This can result in disappointment in such services, which is often called the "expectation and reality gap" demonstrated in Scott [12], causing users to stop using the application. This might potentially be the reason why this service is not popular among travelers. We came up with a solution that might mitigate this issue. First, all the POIs should have detailed "labels" or notes presenting their travel information. For instance, a museum should have a label or note stating "suitable for people who love literature and art", and a national park should have a label showing "suitable for people who love nature and exploration". This method can avoid excessive expectations of the itinerary, lowering the level of expectation to a reasonable level. This can help to solve the gap between expectation and reality. Furthermore, the application should demonstrate that the automatic generation feature is more of a travel planning assistant than a replacement for travel agents. The users need to understand that the itineraries are not perfect and still need some manual modification to become a "thorough" itinerary. This further lowers the users' expectations, avoiding bad first impressions when first using this kinds of services. This also means that the model needs to offer high malleability and changeability to the users. This allows users to modify the itinerary when situations change during a trip or the users are not satisfied with the recommended POIs. Also, in the model's usability examination, even though most of the participants had good ratings for the models, some respondents stated that the interface and the general structure needed improvements. Since this model used a relatively new concept, the interfaces and the applications' flow should be designed as intuitively as possible.

4.3 Proposed Model

In section 4.3, we proposed a model that utilized the information gathered in the prior works and the experiments. Some of the model's designs referred to Chang, Chang, and Tsai [3], which is introduced in the related works chapter. The purpose of the model is to generate personalized itineraries for users, utilizing information from various sources. As we can see in figure 4.1, the overall model design is separated into three main parts, database structure, itinerary generation, and user interface.

In the user interface part, the user application allows users to enter constraints, request itineraries generation, and give feedback. The feedback will be used to update the data in the "POIs data" and "User profiles" databases. Also, similar to ATIPS, when a user first uses this application, he or she must register an account and go through a "preference testing" procedure. In this procedure, the user needs to pick a few POIs and restaurants that he or she is interested in. This can help the application build the initial user profile. In this method, the model can produce more personalized itineraries and mitigate the cold-start problem to some level. Also, when a user browses or selects a travel location, it will send a signal to the Google API and the web crawlers, requesting it to collect the most up-to-date POI information, such as ratings, location, and recommended length of stay, etc. Also, after the model has recommended an itinerary for the user, he or she can enter feedback for the recommendations.

The itinerary generation part consists of four main algorithms, which are POI selection, information processing, route optimization, and itinerary generation. The POI selection algorithm has been developed for a long time. As shown in the related work chapter, many different kinds of techniques and models focus on solving this task. A state-of-the-art recommendation technique suitable for this task is the Model-Based Collaborative filtering technique mentioned in Chetana et al. [10]; this technique combined the memory-based technique and the model-based technique. This novel technique has a few advantages over the traditional recommendation system techniques. Based on the article, it can perform better when there is sparse data in the user-item matrix, and it has better scalability. Generally, this approach is more accurate compared to the individual model-based approach or the memory-based approach. This technique may be suitable for our task of predicting the POIs. First, when first building the application, the user data will not be sufficient, which means the prediction will not perform well in this case. However, if we can gather enough information through either the online crawlers or the professionally selected POIs, we can build the "model part" of the Model-Based Collaborative filtering Model. Second, since we aimed to design a personalized and detailed itinerary generation system, we will use many parameters when building the deep learning model. As mentioned by Chetana and Ibrahim, such a model will be difficult to build and fine-tune. Overall, it will be easier to build the POIs selection algorithm using this novel technique. Using the algorithm, we can obtain the top N POIs that are suitable for a specific user, and we will further arrange and filter the POIs in the following procedures. The POIs selection in this part will be more general and does not consider most of the users' constraints, avoiding the model from becoming too complex. Figure 4.2 shows an architecture of the model-based collaborative filtering technique, provided by Chetana and Ibrahim. The next algorithm in the itinerary generation procedure is the information processing part. It will utilize the crawlers and the database to quickly gather miscellaneous information needed for constructing an itinerary. This information includes price level, recommended length of stay, theme category, location, business hours, POI type, characteristics, etc. The algorithm will then match the gathered information with the user's constraints, selecting the valid POIs. The unused POIs will be saved as "candidate POIs", which will be presented to the user if he or she is not satisfied with the current recommendations. After picking out the candidate POIs, the next step is to decide the travel sequence and optimize the routes. This algorithm will take the POIs' location, business hours, and recommended length of stay as input, and output a candidate travel sequence. As mentioned in the related works, the travel sequence can greatly influence a traveler's impression of a trip. According to Friggstad et al. [6], for a multi-day trip, the authors suggested that we should maximize the quality of the trip's worst day. By doing this, we can obtain a better and more balanced trip across multiple days. On the other hand, in Bergeron, Fallu, and Roy [13], the authors stated "our results indicate that perceived quality and trust were found to be influenced by the initial impression, whereas satisfaction was mostly impacted by the final impression". This indicates that both first and the last impression are equally important if we want to create a perfect trip for users. These two articles suggested that, for a multi-day trip, even though the first and last days' impressions are crucial to travelers' experience, we can not disregard the quality of the trips in the middle. The generated trip should not exist a great quality disparity between the days, and we should focus on the "quality balance" throughout the multi-day trip, or else there will be a significant drop in travelers' overall experience. Even if the user is only creating a single-day trip, the algorithm should still apply the same concept to the generated trip. To achieve this, we should give a score to each of the selected POIs based on their popularity and the concordance between the POIs' characteristics and the user's constraints. Using the scores, we can create a balanced multi-day trip while allocating slightly better POIs for the first and the last day. Lastly, the itinerary generation procedure will combine all the generated information and recommend an itinerary to the user.

The database part consists of two data storage and one algorithm. The "Information compiling" algorithm will process the information gathered from various online sources. Static information such as business hours and POI name will be written directly to the database. The non-static information, such as ratings, reviews, and theme categories, will require further processing. For processing the ratings, a potential suitable technique is a "weighted score" method. Since different online sources have different credibility, each of them should have different weights. For instance, Google Maps' ratings and reviews should weigh more than other online sources' since most people see Google Maps as an authoritative and accurate guide. With the predefined weights, we can calculate an overall score for a specific POI, which will be used in the POI selection algorithm. Additionally, the algorithm needs to be capable of categorizing POIs' themes and labeling their characteristics. These two pieces of information are essential to recommend personalized spots and improve system usability. On the other hand, the two data storages are the POIs data and the User profiles. The POIs data storage stores the information gathered and processed in the above information compiling algorithm. It allows for faster access to the POI processed data and adjustment according the users' feedback. The user profiles data contains the users' basic information, travel preferences, and feedback. The information will be used to generate recommended POIs for users. In figure 4.3, a proposed data structure is shown.

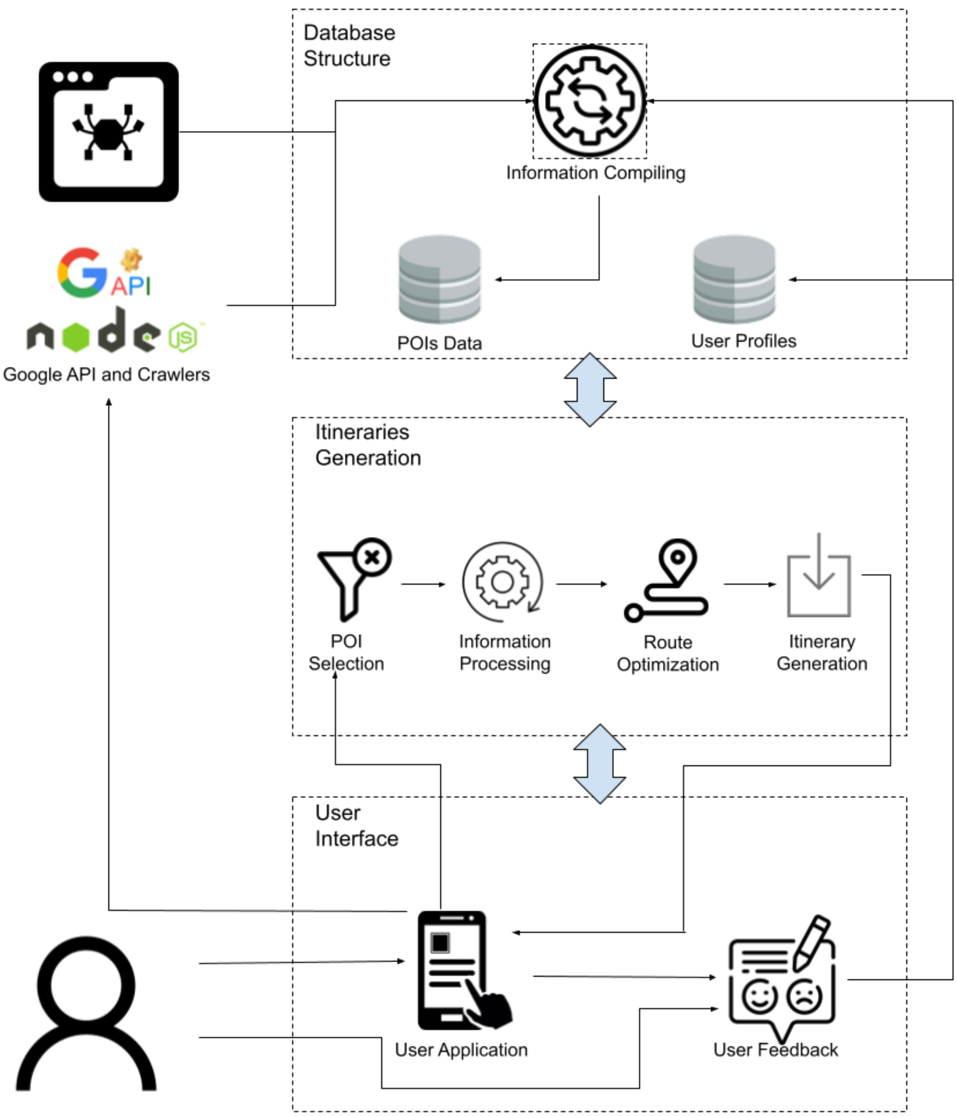

FIGURE 4.1: Proposed model design structure

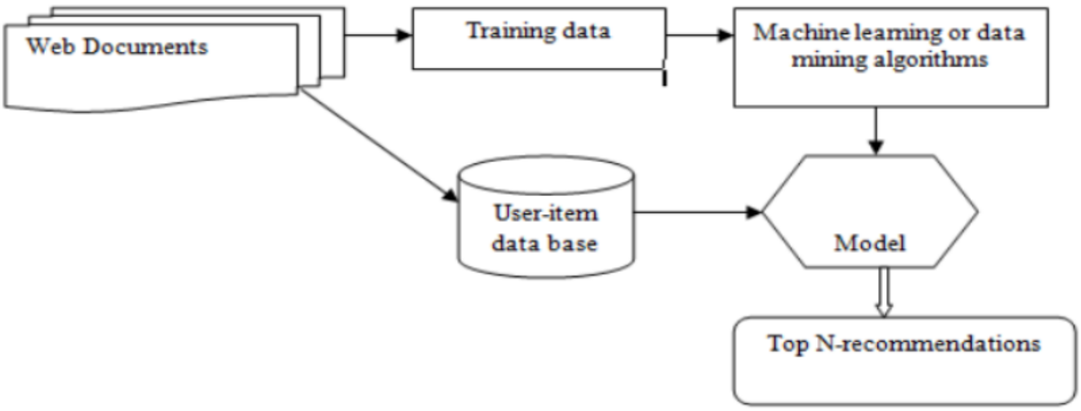

FIGURE 4.2: An architecture for model-based collaborative filtering, provided by Chetana and Ibrahim

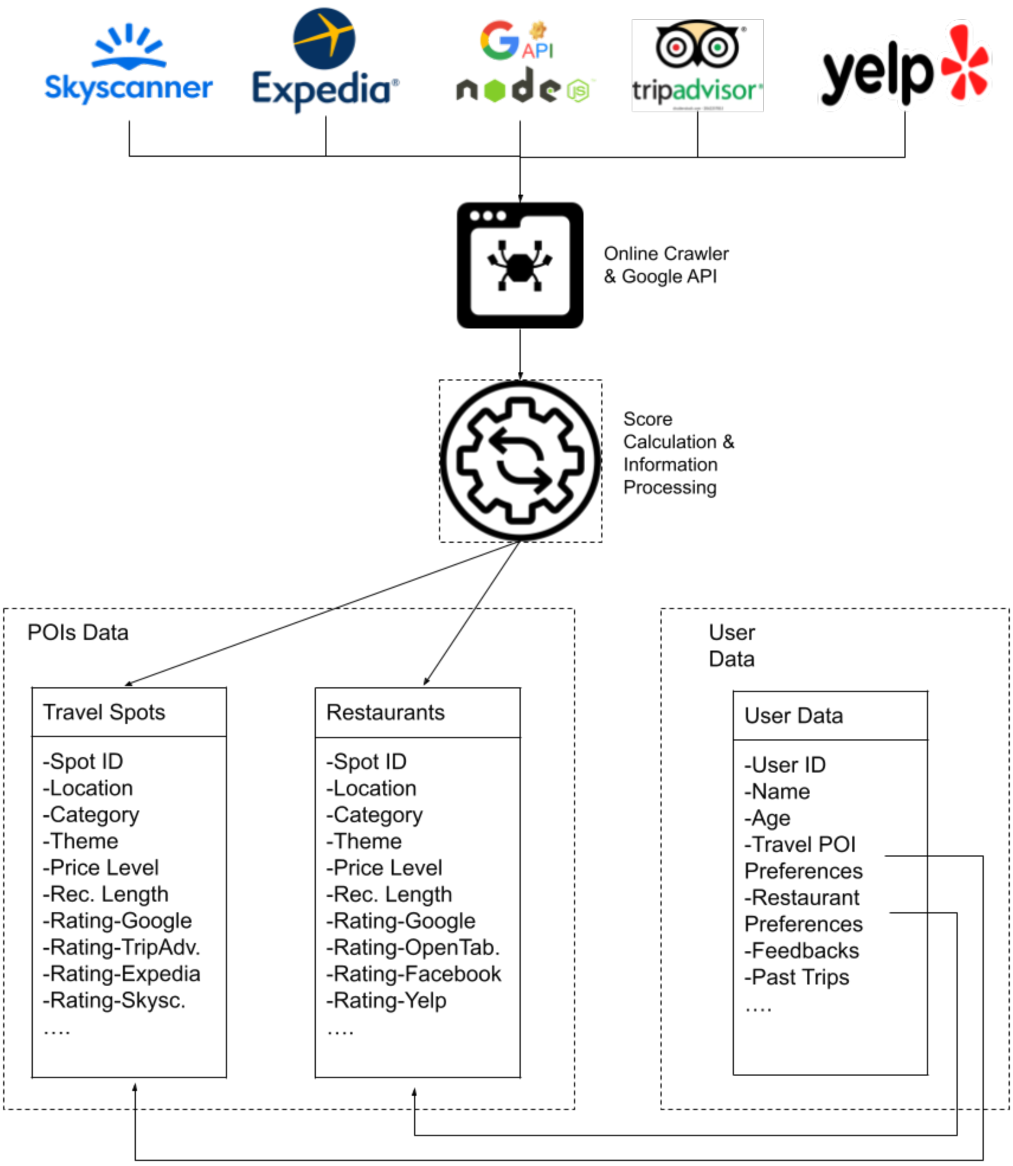

FIGURE 4.3: The proposed data structure

Chapter 5

Conclusion

This chapter of the thesis will conclude all the findings and solutions that we have studied and examined. These include the findings in the related works, the pivot research results, the online group experiment results, the interview group experiment results, and a proposed model that utilized all of the findings in this thesis. In this chapter, we will also discuss the issues that we encountered during this research. Furthermore, we will discuss the potential applications of this proposed model and some potential future work.

5.1 Conclusion of This Project

In this thesis, we have explored the possibilities of an automatic travel planning system (or an automatic itinerary generation system), mainly focusing on users' acceptability and feasibility of such systems. The key points of this thesis are the followings,

- We showed that most people are not satisfied with the current travel planning process, mainly because it is too time-consuming, troublesome, and not personalized.
- The experiment results show that most people generally accept this novel method of planning their trips, and they have high expectations of the generated itineraries.
- We showed that the automated planning feature is convenient and timesaving for travelers, and they prefer the models with automated features over the models without those functions.
- Travelers' demands differ greatly; thus, to build an acceptable system, it needs to offer high user customization, provide more personalization, and deliberately lower users' expectations to avoid the expectation-reality gap.

- Using the findings in both the prior works and the experiment, a framework for an automatic travel planning system is proposed in section 4.3.

In a nutshell, we not only demonstrated that an automatic travel planning system is feasible but also showed that most people accept this novel method of travel planning. First, we initiated some pivot research, ensuring the validity of some essential hypotheses and assumptions. The results of the pivot research guided the formation of the main experiment. Second, for the main experiment, we separated them into an online group experiment and an interview group experiment. In the online group experiment, the participants will follow a guide to examine the models and fill out the surveys. In the interview group experiment, when they are under the experiment process, we will observe their behaviors and answer any questions. Also, we would ask some more detailed questions in addition to the preset questions, getting more detailed comments and recommendations. Based on the online group results, we showed that respondents are generally willing to try out this novel planning approach, and they hold high expectations of such an automatic system. And the majority of the respondents showed that Model C, the model with the most automated features, is the best model among the three. This means that people saw the automatic planning features as timesaving and convenient, and they believed that the algorithms are completely capable of completing such tasks. Many of the participants suggested that Model B and Model C made the travel planning procedure easier. On the other hand, in the interview group experiment, we got many valuable comments and recommendations from the respondents. Similar to the results found in the online group, the respondents also showed high interest and expectation in the generated itineraries. However, the high expectation might cause the expectation-reality gap, which can lead to users stopping using such services on their first try. To avoid this issue, we came up with a solution to lower the users' expectations to a reasonable level, including using "POIs labels" and emphasizing the assistive property of the automatic recommendation features. This prevents a bad first impression, which is disastrous, especially for novel applications. To further mitigate this issue, the developed application should provide high malleability and changeability, allowing the itineraries to adapt to the various preferences that are not covered by the application options. Lastly, after the main experiment, we proposed a framework of an automatic travel recommendation system. The construction of the proposed model utilized the information and findings in the main experiment. The model is our ideal image of an automatic travel recommendation system, backed by the prior researches and the results of the main experiments. we also mentioned a few key points about users' behaviors that should be carefully considered and handled.

5.2 Future Works and Potential Applications

This project provides developers with general insights into people's opinions toward automatically generated trips. Since this project only built simulation models instead of actual functional travel applications for the experiments, there might be some minor discrepancy between our results and the users' actual thoughts on the system. To obtain more accurate results, we need to construct a working automatic travel recommendation application using the proposed model in section 4.3. However, building such an application will not be easy. As we can see, there are many sub-tasks in the whole travel planning procedure. We estimated that the application will take several months with several developers' efforts. With the actual program, we can measure people's opinions and acceptability more accurately in another experiment. Also, we will continue to expand the capabilities of the model, including multi-platform integration, utilizing data from various sources, more personalized design, etc.

If we can successfully build the proposed model, we can not only learn more about people's thoughts in this field but also apply the model in various situations. For instance, other than using the model in a typical personal travel planning application, the model can also be used by travel agency companies to provide their customers with unique trips. Traditionally, when building a customized trip for their customers, the travel agency still needs to manually go through the ordinary travel planning process. But with the help of our proposed model, a candidate itinerary can be generated quickly, reducing their planning time.

Acknowledgements

This research was undertaken with the support from The Real Sakai Laboratory, Waseda University. I would like to express my thanks of gratitude to members in The Real Sakai Laboratory, who helped me and gave me many useful pieces of advice for continuing the research. Also, I would like to thank Professor Tetsuya Sakai, who gave me this great opportunity and lots of supports to do research on this topic.

Bibliography

[1] *Travel and tourism*. URL: https://www.bea.gov.

[2] Kadri Sylejmani and Agni Dika. "A SURVEY ON TOURIST TRIP PLANNING SYSTEMS". In: *International Journal of Arts and Sciences, Gottenheim, Germany* ISSN: 1944-6934 (Apr. 2011), pp 13 –26.

[3] Hsien-Tsung Chang, Yi-Ming Chang, and Meng-Tze Tsai. *ATIPS: Automatic Travel Itinerary Planning System for domestic areas*. 2015. URL: https://www.hindawi.com/journals/cin/2016/1281379/.

[4] Xin Lu et al. "Photo2Trip: Generating Travel Routes from Geo-Tagged Photos for Trip Planning". In: *Proceedings of the 18th ACM International Conference on Multimedia*. MM '10. Firenze, Italy: Association for Computing Machinery, 2010, 143–152. ISBN: 9781605589336. DOI: 10.1145/1873951.1873972. URL: https://doi.org/10.1145/1873951.1873972.

[5] Gang Chen et al. "Automatic Itinerary Planning for Traveling Services". In: *IEEE Transactions on Knowledge and Data Engineering* 26.3 (2014), pp. 514–527. DOI: 10.1109/TKDE.2013.46.

[6] Zachary Friggstad et al. "Orienteering Algorithms for Generating Travel Itineraries". In: *WSDM*. 2018.

[7] Munmun De Choudhury et al. "Automatic Construction of Travel Itineraries Using Social Breadcrumbs". In: *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*. HT '10. Toronto, Ontario, Canada: Association for Computing Machinery, 2010, 35–44. ISBN: 9781450300414. DOI: 10.1145/1810617.1810626. URL: https://doi.org/10.1145/1810617.1810626.

[8] Keith D. Foote. *A brief history of machine learning*. 2021. URL: https://www.dataversity.net/a-brief-history-of-machine-learning/.

[9] Daniil Korbut. *Recommendation system algorithms: An overview*. URL: https://www.kdnuggets.com/2017/08/recommendation-system-algorithms-overview.html.

[10] Lakshmi Chetana et al. "Personalization Using Recommendation Systems: State-of-the Art Review". In: 11 (Aug. 2020), pp. 56–74.

[11] Fred Decker. *How to average Likert Scales*. 2019. URL: https://sciencing.com/average-likert-scales-6181662.html.

[12] Elizabeth Scott. *How to Manage Your Expectations vs Reality*. 2022. URL: https://www.verywellmind.com/expectation-vs-reality-trap-4570968.

[13] Jasmin Bergeron, Jean-Mathieu Fallu, and Jasmin Roy. "A Comparison of the Effects of the First Impression and the Last Impression in a Selling Context". In: *Recherche et Applications en Marketing (English Edition)* 23.2 (2008), pp. 19–36. DOI: 10.1177/205157070802300202. eprint: https://doi.org/10.1177/205157070802300202. URL: https://doi.org/10.1177/205157070802300202.

Appendix A

More Figures and Tables

A.1 Screenshots of the Experiment Pidoco Models and the Guiding Documents

FIGURE A.1: Model A - Application Login and Main Page

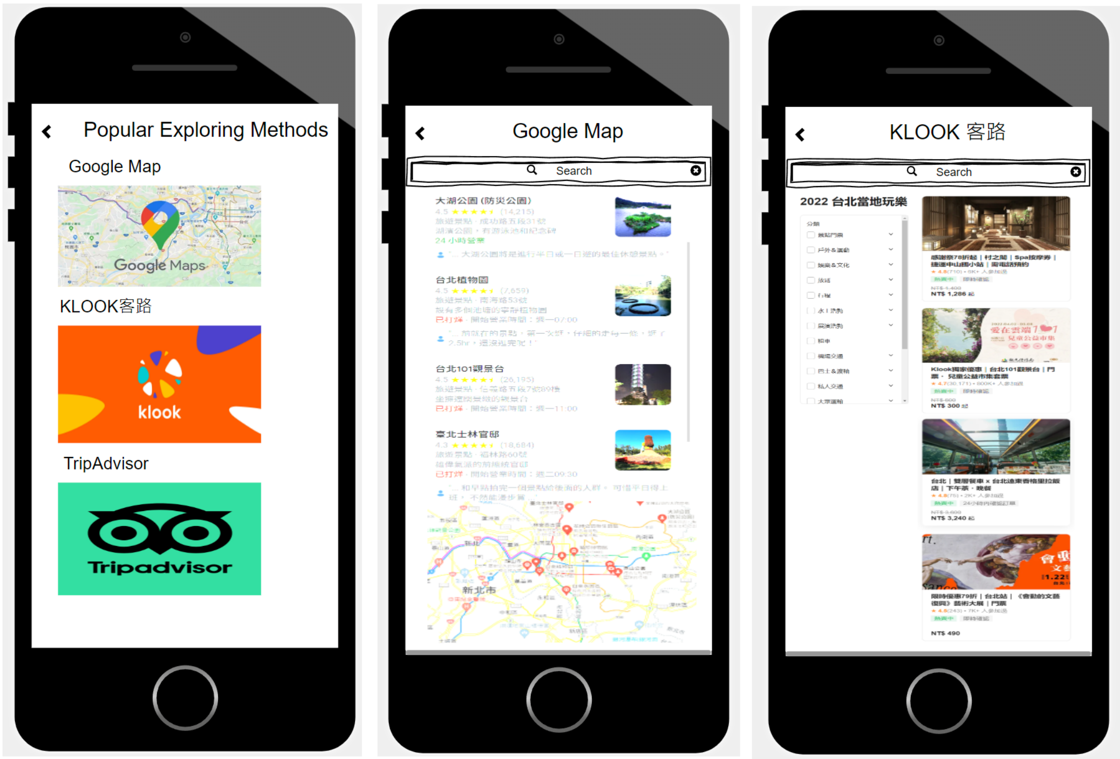

FIGURE A.2: Model A - Integrated POI Searching Tools

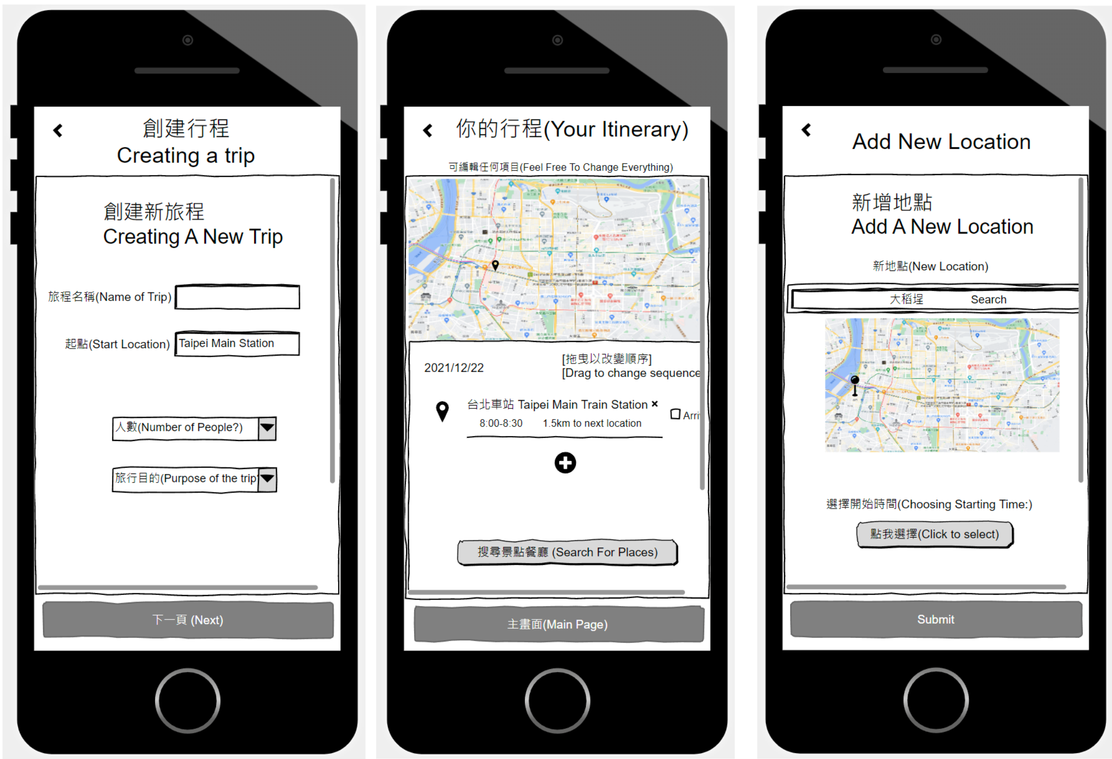

FIGURE A.3: Model A - Procedures of Creating a trip I

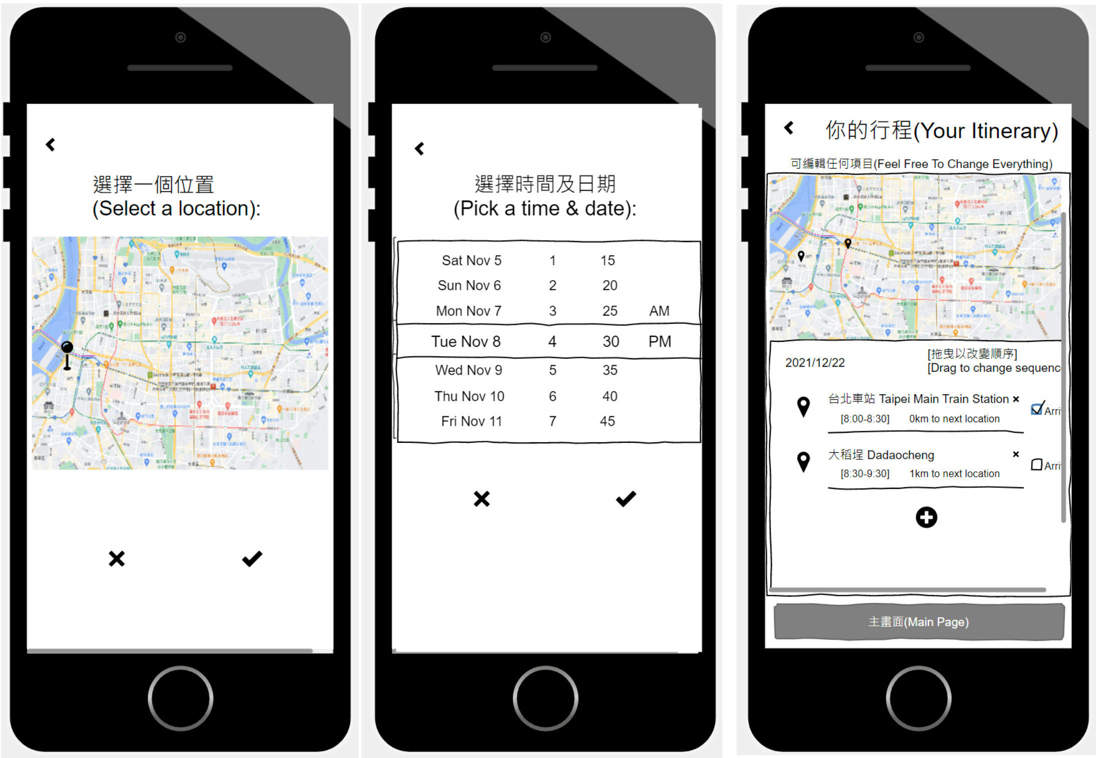

FIGURE A.4: Model A - Procedures of Creating a trip II

FIGURE A.5: Model B - Login Page and The Preference Learning Procedures

FIGURE A.6: Model B - Procedures of Creating a trip I

FIGURE A.7: Model B - Procedures of Creating a trip II

FIGURE A.8: Model C - Login Page and The Preference Learning Procedures

FIGURE A.9: Model C - Procedures of Creating a trip I

FIGURE A.10: Model C - Procedures of Creating a trip II

FIGURE A.11: Guide Document - Page 1

FIGURE A.12: Guide Document - Page 2

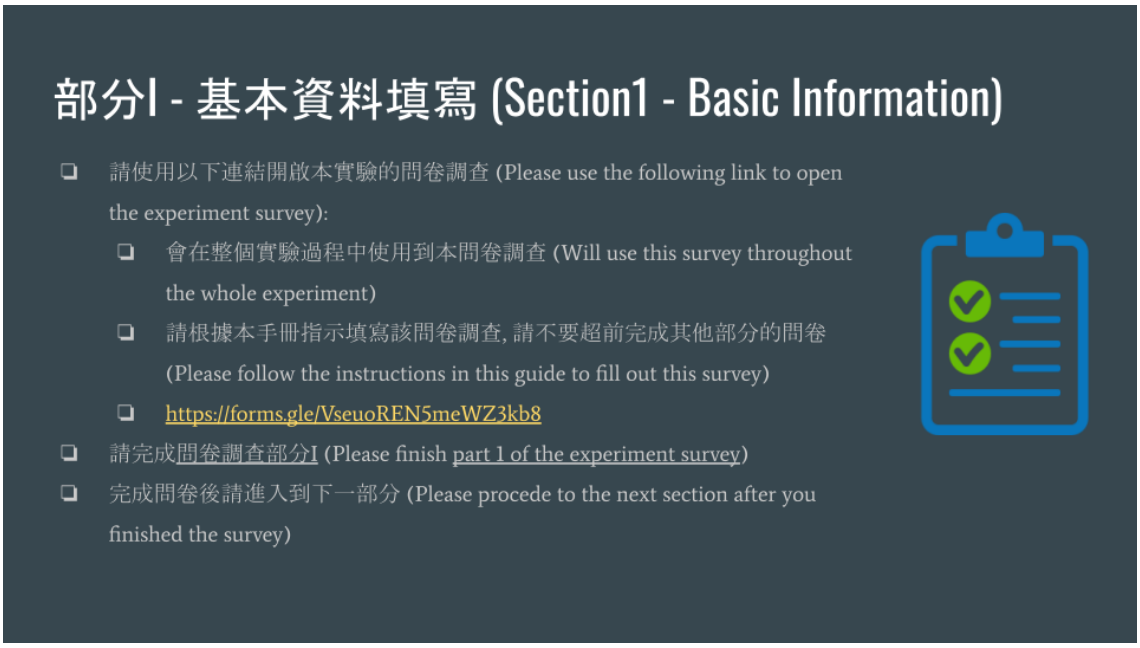

FIGURE A.13: Guide Document - Page 3

FIGURE A.14: Guide Document - Page 4

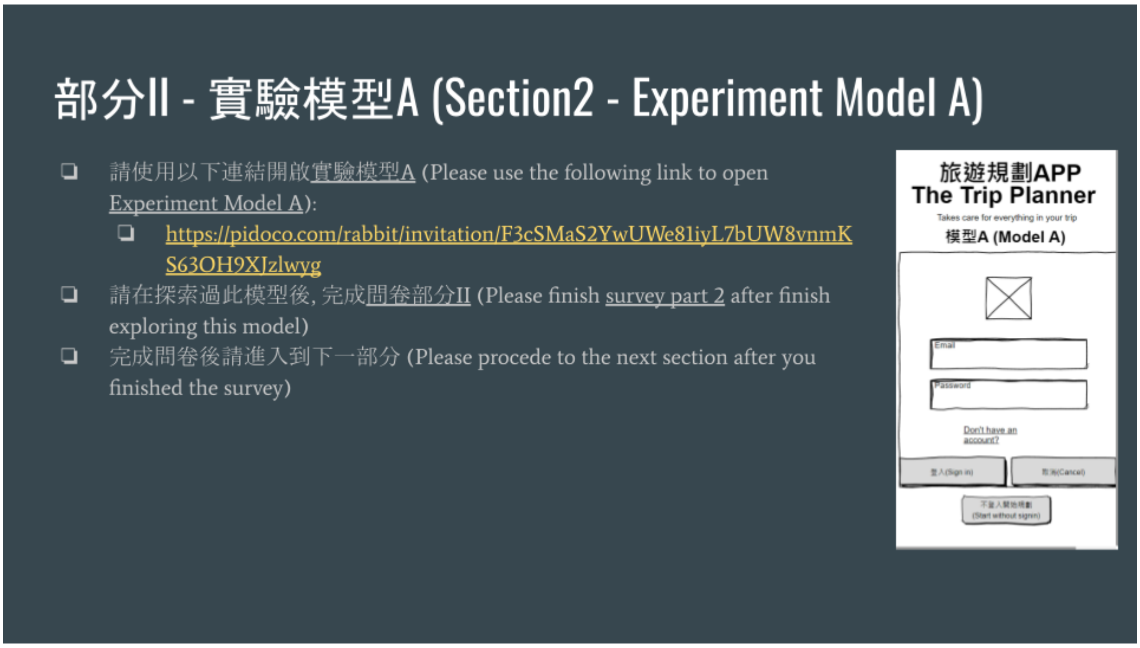

FIGURE A.15: Guide Document - Page 5

FIGURE A.16: Guide Document - Page 6

FIGURE A.17: Guide Document - Page 7

FIGURE A.18: Guide Document - Page 8

FIGURE A.19: Guide Document - Page 9

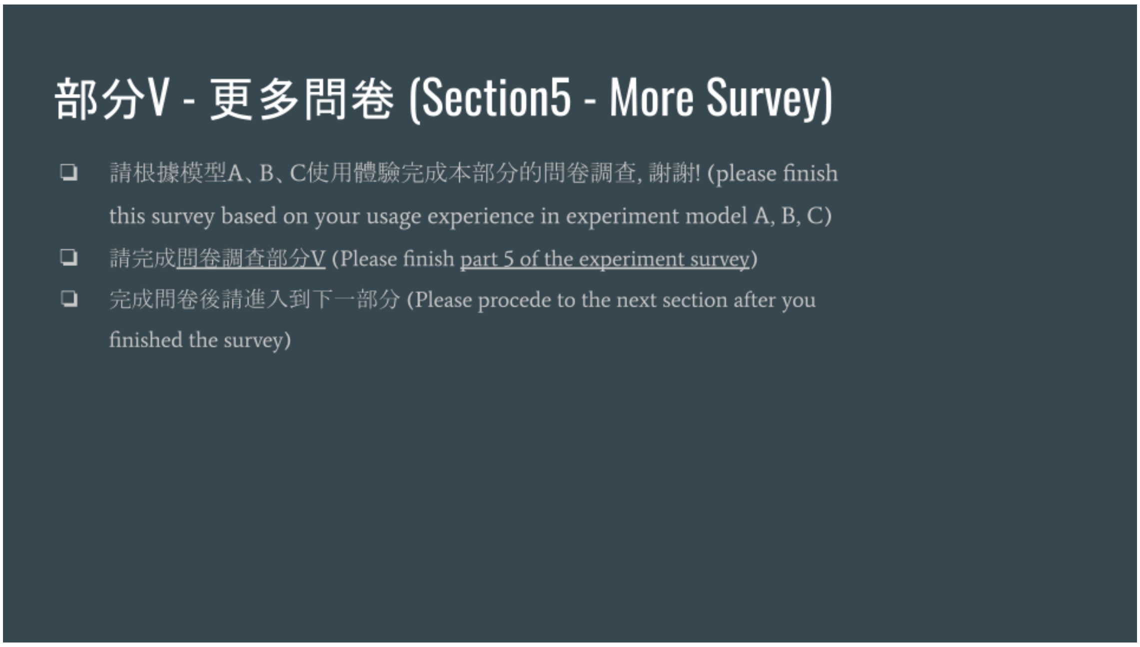

FIGURE A.20: Guide Document - Page 10

FIGURE A.21: Guide Document - Page 11