

Project 1 Report

I. Project Information

My chatbot is built on the topic of traveling in South America, and it contains information regarding 6 countries: Argentina, Brazil, Chile, Colombia, Uruguay, and Paraguay. To build the chatbot, I implemented various NLP techniques which I will further elaborate on in the following sections.

II. Web Crawler

A. Building the Knowledge Base

To create my knowledge base, I created a web scraper that would extract the content off of websites related to my topic of traveling in South America. I used 1 starter url, which was a link to a travel blog, specifically for South America countries. Using the BeautifulSoup library, I scraped off all urls from the starter url and added them to a list. The program then iterates through this list to scrape other urls from each url in the list. Since not all urls were relevant, I filtered the urls to not include any social media, car/hotel/tour booking websites, and countries outside of South America. My starter url often linked to sites that contained links to blogs with content, so I scraped urls on each site until I got the url of the site with content. I also filtered the domain names of the urls to get external links as my starter url and other scraped urls often shared the same domain name. I limited the number of relevant urls collected to 20, and once this limit was met, I extracted all the text enclosed in HTML paragraph tags (<p>) also using BeautifulSoup.

With 20 raw text files, I cleaned each file using NLP techniques like NER, tokenization, and regex expressions. After reading in the text from each file, I applied sentence tokenization on the text and removed any sentences that contained a date. This is where I used NER to identify entities that were labeled as 'DATE' and remove them. Since I scraped content from a blog, the date of when the blog was posted was included in the scraped text, which is not relevant to my knowledge base. Then for text from an internal domain name, I removed sentences that contained irrelevant wording, such as 'copyright', 'privacy', 'commission', 'google', and any references to social media. Similarly, for text scraped from external domain urls, I removed sentences that included social media, company names (like 'cnn', 'warner'), and references to copyright/privacy information. The last text file included public comments on a blog post, so I removed this information by not including any text that followed the last sentence of the blog post. Next, I noticed many words had accented letters (ex: Córdoba), so I applied word tokenization to each sentence and converted these words to have unaccented letters using the Unidecode library. Lastly, as part of the cleaning, I used regex expressions to remove irrelevant numbers, like numbers in a list (ex: 1. , 2.), and extra space before and after

punctuation. This is so each sentence maintains its grammatical structure and is therefore easy to read by humans. The cleaned text is then written to a file, named “clean_file[#].txt”. After cleaning the scraped text, the program outputs 20 cleaned text files.

With the 20 clean files, I then extracted the top 40 important terms using preprocessing techniques and calculated tf-idf frequencies. For preprocessing, the program read each clean file and made all words lowercase. Then, each word was lemmatized and any stopwords, in English and Spanish, and numbers were removed. Lastly, all punctuation was removed using a regex expression. After preprocessing, each text is appended to an array, which represents the corpus of documents. For each document in the corpus, the tf value is calculated, and the idf value is calculated using the corpus array. Using both values, the tf-idf frequency is calculated for each word in every text and stored in a dictionary with the key as the word and the value as the frequency. For each text/document, the dictionary is sorted by value and the top 2 terms are printed. This is done for all 20 texts/documents, so a total of 40 most important words are printed as a result.

Given the top 40 words from the corpus, I selected the 15 most important words based on names of places (cities, countries, attractions, etc.) and activities since these are most relevant to the topic of traveling in South America. Therefore, I chose ['bariloche', 'uruguay', 'salta', 'wine', 'iguazu', 'pedro', 'trek', 'natale', 'colombia', 'beagle', 'fitz', 'car', 'cafayate', 'cuidad', 'cordoba']. For each of the 15 terms, I created a dictionary to store each term's synonyms, hypernyms, and hyponyms, using WordNet synsets, in addition to other words relevant to the term.

To build the knowledge base, I used a dictionary with the key as one of the 15 chosen terms and the value as all sentences within the cleaned text corpus that contains the term or one of its synonyms. The program iterates through each sentence from each of the clean files. After preprocessing the sentence (lemmatization, removing stopwords and numbers), the program checks if one of the important terms or any word from its synonyms is within the sentence. If yes, then the program appends the sentence to the array of sentences in the knowledge base dictionary with the term as the key. After the iteration is complete, the knowledge base dictionary is complete, and the dictionary is pickled and written to a file.

B. Screenshots of Knowledge Base

natale = ['More specifically, you'll see Punta Arenas, Puerto Natales, Torres del Paine National Park, El Calafate, El Chalten/Cerro Torre, and a few sights in Los Glaciares National Park.', 'You should have your W Trek planned out far before you arrive in Puerto Natales.', 'Grab a bus at any time from Punta Arenas to Puerto Natales!', 'Once you arrive in Puerto Natales, you'll be at the bus station about 2km from the center of town.', 'Puerto Natales and Torres del Paine cater very well to the luxury traveler, so there are quite a few amazing places to stay (including one of the top hotels on the continent!', 'Puerto Natales is a beautiful and cozy little town with lots of beautiful sculptures along the water.', 'Most people visit Puerto Natales to get to Torres del Paine, but you can also ride bikes around the town and also take a glacier boat tour.', 'To get out, you must take a catamaran from Paine Grande back to the other side of the park, where busses pick you up to go back to Puerto Natales.', 'Again, busses pick you up right from the catamaran and go straight to Puerto Natales (after stopping at the park entrance first as well).', 'You're back in Puerto Natales - GET SOME SLEEP!', 'You can book another night in the same place you booked for your other night in Puerto Natales.', 'There are several places in Puerto Natales.', 'Your budget and arrival city (ie where you've been before Puerto Natales) may dictate how much of your budget goes towards snacks.']

colombia = ['This sleepy little Caribbean gem on Colombia's coast is only accessible via air or sea, which makes it that much more charming to visit.', 'I came into contact with Capurgana Colombia because of my San Blas islands tour from Panama to Colombia with San Blas adventures, and although at first skeptical of ending in such a random small village, I quickly understood that the charm of this place is even more reason to book a speedboat tour between the two countries.', 'Capurgana was a warm welcome to Colombia and a nice introduction to the simple way of life that they live in this coastal Caribbean village.', 'I was there in low season so it was quite empty, but from what I have heard the sleepy little town of Capurgana Colombia comes quite alive during the holidays - over New Year's in particular.', 'There are a few flights a week from the teeny tiny little airstrip in town, to bigger Colombian cities Medellin and Cartagena.', 'Other than that, you can take a bus from Medellin or Cartagena to Necocli, where you can catch a speedboat across the big bay to reach Capurgana.', 'If you hike north from Capurgana's beaches, you will find a lovely coastal track that follows the breaking waves along the hilly jungle.', 'Sapzurro is an even more quaint and colorful little beach town just over the hill from Capurgana; it could be a destination in itself!', 'There's a short hike - maybe an hour and a bit - to Sapzurro from Capurgana.', 'Sapzurro has lots of cozy hammock hostels, a lovely white sand beach, some little restaurants, and even more trails to continue through the jungle.', 'La Miel is a must-do when in the area.', 'It's just over the hill from Sapzurro, so you can make one big trek out of the two if you would like.', 'This is where you can see signs that point to Panama in one direction, and Colombia in the other.', 'Another awesome hike in Capurgana is El Cielo.', 'You can reach El Cielo if you take the path right next to the airstrip and continue into the jungle - you can't miss the signs!', 'Inside El Cielo you can chill inside a natural jacuzzi/pool, hike up to the top of a tall hill to get a fantastic viewpoint of the jungle (and all the way to the ocean!', 'Capurgana is also a very popular place to go scuba diving out in the pristine Caribbean waters.', 'If you walk the opposite direction from the village - south - you can reach another popular rocky beach called Aguacate.', 'Capurgana has LOTS of food options all over town... but the food is really hit and miss.', 'On one of the main roads right past the Colombia immigration office (which you will need to go to if you have just arrived to Capurgana Colombia from Panama) there will be a little table outside a small stall near the road (with pictures of fruit and a colorful menu on the wall).', 'It's always good when it's a local's holiday destination:) And yeahhh, worth it just for the coco locos: P San Blas Sailing Panama to Colombia: Speedboat or Sailboat Trip?', 'It was my first ever solo backpacking trip and I decided to kick it off in the country of cocaine and cartels: Colombia.', 'That goes for pretty much every country that I've been to, but especially so for places like Colombia, Brazil, and Argentina.', 'You can get cheap flights in some countries, like Colombia, but more often than not, a night bus will beat it at comfort and price.', 'I was in traffic for almost an hour and ended up only paying about \$7 in C

pedro = ['This beautiful road goes through a gorge called the Quebrada del Toro and up into the Atacama highlands - a beautiful watercolor desert that's very high in elevation.', 'It takes you through the Toro Gorge (Quebrada del Toro) and up into the Puna Highlands (Puna de Atacama) which are all colorful and gorgeous.', 'The driest desert in the world, San Pedro de Atacama is home to incredible geysers, salt lakes, volcanos, altiplanic lagoons, other-worldly landscapes, and a crazy amount of adventures.', 'During my time there, I did quite a few San Pedro de Atacama tours to some of the most popular sites around the area.', 'Most of the awesome things to do in San Pedro de Atacama are within a few hours of the main town, making it a perfect hub for excursions and points of interest in the Atacama desert.', 'Once you arrive in Atacama you will understand it runs on tours.', 'Whether you book in advance or not, it would be wise to make a schedule of morning/evening/night tours tailored to your own interests and timings when visiting San Pedro de Atacama Chile, and book tours from there (choose from my lists below).', 'I'll let you know all the tours that are available, the ones that I did, possible other ways to see certain places, and finally, recommend an itinerary for your trip to San Pedro de Atacama Chile.', 'Now, I will break the best tours in San Pedro de Atacama Chile into times of day.', 'No matter which company you go with, they will run at the same time of day because certain things to do in San Pedro de Atacama are better in the sunrise or afternoon/evening time.', 'Extra Costs: CLP \$5,000 to enter Rainbow Valley Timings: About 7:30/8am to 1/1:30pm EVERYONE must do this tour when visiting San Pedro de Atacama Chile!', 'Extra Costs: CLP \$30,000 General or discounted after 2pm M-F Timings: About 1:30pm to 5:30/6pm or 7:30/8am to noon San Pedro de Atacama Chile is known to have the best stargazing in the WORLD.', 'Luckily for you, you can book multi-day San Pedro de Atacama tours where everything is already sorted for you.', 'There are amazing shops lining the streets of town, and also a bigger market near the main square (Plaza de San Pedro de Atacama) and also a bigger artisanal market nearer to the bus station.', 'Atacama has tons of accommodation options on all ends of the budget scale.', 'Day 1 Afternoon: Moon Valley Tour Night: Stargazing Tour Day 2 Morning: Altiplanic Lagoons + Atacama Salt Flats Tour Afternoon: Cejar Lagoon + Los Ojos + Laguna Tebinquiche Sunset Tour Day 3 Morning: El Tatio Geysers Tour Well, I hope this has helped you plan (or inspired you to plan) your trip to San Pedro de Atacama, Chile.', 'Have you been to Atacama?']

uruguay = ['It's a pretty popular tour in Buenos Aires to take a boat ride up to the Tigre Delta - the mouth of the Rio de la Plata, the 'largest river in the world' that separates Argentina from Uruguay.', 'Mar del Plata, or 'la Plata,' is where Portenos, as people from Buenos Aires are called, tend to vacation (that is, if they aren't ferrying across to Uruguay).', 'The names of quite a few neighboring countries, such as Paraguay and Uruguay, also come from Guaraní roots.', 'More specifically, you'll see Punta Arenas, Puerto Natales, Torres del Paine National Park, El Calafate, El Chalten/Cerro Torre, and a few sights in Los Glaciares National Park.', 'Once you arrive in Punta Arenas, settle in and rest!', 'There are very cheap flights to Punta Arenas from Santiago, the capital of Chile.', 'From there, you can get another cheap flight down to Punta Arenas.', 'Hotel Almasur Punta Arenas (Prev.)', 'There are a few hostels in Punta Arenas for backpackers.', 'This tour will take you on a ferry to Porvenir from Punta Arenas, to the penguins for a few hours, and the up to some historical sites in San Gregorio and a few other points of interest.', 'Grab a bus at any time from Punta Arenas to Puerto Natales!', 'The bus station is right in town in Punta Arenas, and is within walking distance from any accommodation in the center.', 'Many of Uruguay's main cities and points of interest lie along its coast, which begins along the largest river in the world, the Rio de la Plata.', 'Uruguay's coast stretches all the way from its border with Argentina to the Atlantic Ocean, finishing at its border with Brazil.', 'I'm sure there are lots of beautiful little farming towns to see in the middle of the country, but this Uruguay road trip will focus on the long stretch of coast between Montevideo and Punta del Diablo.', 'There are 17 amazing road trip stops, including Punta del Este, and recommendations on where would be good to spend a night or two.', 'You can also start in Colonia del Sacramento if you want, for a complete schedule of beautiful places to visit in Uruguay.', 'Uruguay is accessible via a few international airports, and also by ferry from Buenos Aires.', 'Fly: You can fly into Montevideo or Punta del Este from many international airports.', 'This route will start in Montevideo and go through all the places to visit in Uruguay along the coast to Punta del Diablo.', 'Before you begin driving, you'll want to know a few tips for driving in Uruguay.', 'To make the most of your Uruguay road trip, I recommend taking the coastal routes as much as possible.', 'These Uruguay road trip stops are listed in order, but it's up to you to decide how much time you would like to see it all.', 'Montevideo is full of beautiful coastlines, little pedestrian streets, and busy markets serving up the best Uruguayan fare (and LOTS of amazing meats).', 'The first turnout before Atlantida is the way to El Aguila, a beautiful eagle sculpture with a dolphin's body, that actually used to be the artist's home (see photo below).', 'It was made by an Italian artist who was infatuated with the Uruguayan coast, and retired to his eccentric self-built home on the hill.', 'Atlantida has a beautiful Rambla along the water, with more sea-view parks where locals hang out and share meals with their friends.', 'After driving along the coast past Atlantida and enjoying the view of a creek, you'll have to get back on the highway for a little while (make sure you have cash for tolls - have you read my tips for driving in Uruguay?).', 'Its definitely one of the most unique places to visit in Uruguay!', 'This Castle was actually made to be the living quarters of a Uruguayan businessman named Francisco Piria.', 'He is famous for taking a trip to Europe and coming home with the inspiration to create the next town on this Uruguay road trip itinerary, Piriapolis, and to encourage tourism of the beautiful Uruguayan coast.', 'Piriapolis is a gorgeous little beach town with a lovely beach overlooked by a few hills clad with expensive homes.', 'You get a wonderful view of Piriapolis and down the Uruguayan coast, out to sea (or out across the Rio de la Plata as they say), over the rest of the point (Punta Colorado), and even all the way up to Punta del Este.', 'This is definitely one of my very favorite places to visit in Uruguay and a must-see on your Uruguay road trip.', 'Casaapueblo is a massive hotel (Club Hotel Casaapueblo - you can stay there too!), 'You can also hike down the hill next to Casaapueblo for even more sweeping views of the building and the sea.', 'Punta del Este is much more than just a road trip stop, as it's one of the most popular places to visit in Uruguay.', 'The rest of these Uruguay road trip stops can be completed in a jam-packed day, or you can take your time checking out different places to visit in Uruguay along the coast.', 'The best places to stay the night if you want are Cabo Polonia, La Pedrera, and Punta del Diablo.', 'This bridge is a fun quick first stop out of Punta del Este.', 'La Barra is a super trendy area just outside Punta del Este (just across the Puente de la Barra) with hip outdoor bars and restaurants.', 'This is actually where a lot of the most famous nightlife in Punta del Este actually takes place.', 'Punta del Diablo is yet another beachside community with sand-swept roads, but this one has a higher concentration of waterside bars, restaurants, and clubs that come alive in the summertime.', 'Renting a car in Uruguay is relatively easy, with a few big companies like Hertz and Budget in each big city.', 'I did a coastal road trip in Uruguay, from Mon

C. Important terms

15 terms used for knowledge base: ['bariloche', 'uruguay', 'salta', 'wine', 'iguazu', 'pedro', 'trek', 'natale', 'colombia', 'beagle', 'fitz', 'car', 'cafayate', 'cuidad', 'cordoba']

40 important terms after tfidf calculation:

- 1: [('glacier', 0.008013501304547173), ('bariloche', 0.007716277694758179)]
- 2: [('jujuy', 0.012600742173272103), ('catamarca', 0.011470123205675501)]
- 3: [('wine', 0.02548264241766026), ('bodega', 0.023572684282340627)]
- 4: [('salta', 0.03756807271785588), ('jujuy', 0.019894887134325176)]
- 5: [('iguazu', 0.03530763514600715), ('fall', 0.029105614325124055)]
- 6: [('pedro', 0.029577047259917957), ('atacama', 0.027810953066725912)]
- 7: [('bariloche', 0.021693907788762484), ('angostura', 0.013558692367976553)]
- 8: [('trek', 0.019161842244211216), ('natale', 0.015555967761343348)]
- 9: [('temperature', 0.021145460945714727), ('snow', 0.02064869820560177)]
- 10: [('capurgana', 0.03627647661823037), ('colombia', 0.024562680703363)]
- 11: [('ciudad', 0.019926908959012522), ('kebab', 0.019926908959012522)]
- 12: [('uruguay', 0.022170777341541867), ('punta', 0.020315075164789753)]
- 13: [('uruguay', 0.030096077839197103), ('car', 0.026423010199185148)]
- 14: [('punta', 0.0326314250004456), ('peninsula', 0.015606333695865287)]
- 15: [('montevideo', 0.018140113381466157), ('uruguay', 0.015587683011939961)]
- 16: [('beagle', 0.027536464373424243), ('channel', 0.027536464373424243)]
- 17: [('needle', 0.041012359136572286), ('fitz', 0.029134022523148093)]
- 18: [('cordoba', 0.06066688085134874), ('slide', 0.014337654007094375)]
- 19: [('solo', 0.0288588369563288), ('south', 0.023974524123280546)]
- 20: [('campsite', 0.062483353410033), ('mainly', 0.02157225006571998)]

III. Chatbot

A. System Description

The chatbot, supported by the knowledge base, answers user's queries about traveling in South America. It covers the following 6 countries: Argentina, Brazil, Chile, Colombia, Uruguay, and Paraguay. To develop the chatbot, I utilized the NLP techniques, such as lemmatization, tokenization, synsets, tfidf vectorization, cosine similarity, POS tagging, and NER. The chatbot is built as a continuous program until the user ends the conversation, and with every dialogue, the user input is checked to determine if they want to end the conversation. A user model is created as a dictionary, and it stores personal information of the user from the conversation, such as their name, countries they want to visit, likes, and dislikes. The keys of the dictionary are the category names (ex: 'name', 'country', 'like', 'dislike'), and the values are the information provided by the user.

The first dialogue is the greeting, where the chatbot introduces itself and asks the user for their name. The program then parses the user's input using the spaCy library to POS tag each token and identify a proper noun provided by the user. If a proper noun is identified then the conversation continues, otherwise the chatbot continues to ask the user for their name. The user's name is then stored in the user model under the key 'name'.

The second dialogue is asking the user which countries they'd like to visit. The program parses the user's input and uses spaCy's NER feature to identify all entities/tokens labeled as 'GPE' (geopolitical entities). After getting all GPE entities, the program checks to see if all the user's locations are supported by the chatbot (6 countries mentioned above). If none of the user's locations are supported, the program requests the user for their country preferences again and also displays the supported countries. If one or multiple supported countries are identified within the user's input, the conversation continues. The user's preferences for the countries are stored in the user model under the key 'country'.

The third dialogue asks the user for their travel experience level. There are 3 options: beginner, intermediate, and advanced. The program checks the user's input to identify whether one of these levels is in the user's input. If no level or multiple levels are provided by the user, the program continues to ask the user until one valid experience level is provided. Once their travel experience level is identified, it is stored in the user model with the key 'experience', and the conversation continues.

The fourth dialogue asks the user for things they like when traveling. This includes any activities, places, or things they like in general. The program creates a list of synonyms for the word dislike to identify any dislikes the user mentions for this question. If they do, the program informs them to only list their likes. The program then parses the user's input with spaCy's POS tagging to identify any nouns and verbs. As long as the user provides at least one verb or noun, the program continues, otherwise the question is repeated for the user to provide an input again. Once a valid input is determined, any verbs found are stored in the user model under the key 'activities' and any nouns are

stored under the key 'likes'. The program avoids adding the verb "like" to the user model.

Similar to the fourth dialogue, the fifth dialogue asks the user for their dislikes while traveling. The program parses the input using spaCy's POS tagging to identify nouns and verbs. As long as one verb or noun is given, the program continues and stores all nouns and verbs in the user model under the key 'dislikes'. If no noun or verb is identified, the question is repeated until a valid response is received from the user.

Given all this information, the chatbot provides the user a suggestion based on their country preferences, things they like, and activities they like. Based on these preferences, the program goes to the knowledge base dictionary and looks through all sentences with keywords that match the user's preference. For example, if the user's choice of countries is Brazil, Argentina, and Chile, the program will look through the sentences (values) of key terms: [bariloche, salta, cafayate, cordoba, fitz, beagle, iguazu, natale, pedro, fitz]. [bariloche, salta, cafayate, cordoba, fitz, beagle], [iguazu], and [natale, pedro, fitz] are all names of cities/attractions in Argentina, Brazil, and Chile respectively. Then, the tfidf vectorization is applied to the user preference (a string concatenated with the user's countries, activities, and likes). With all the corresponding sentences from the related terms, tfidf vectorization is applied to each sentence. After the 2 vectors are computed, cosine similarity is applied to determine how similar each sentence is to the user's preferences. The sentence with the maximum cosine similarity score and that doesn't contain any of the user's dislikes is chosen as the response displayed to the user. If the cosine similarity value is less than 0.5, the program prints a clarifying message stating the chatbot does not have information to provide a tailored response and provides the next closest response.

After the suggestion is provided, the user can continue discussing the suggestion provided or give other queries to the chatbot. The program follows a similar process as stated before to provide the output. The only difference is it performs tfidf vectorization on the user's query rather than their preferences to provide the most specific response. Cosine similarity is used again to find the best possible match to their query from the knowledge base. If the cosine similarity value is less than the threshold of 0.5, then the clarifying message and the next closest response is provided.

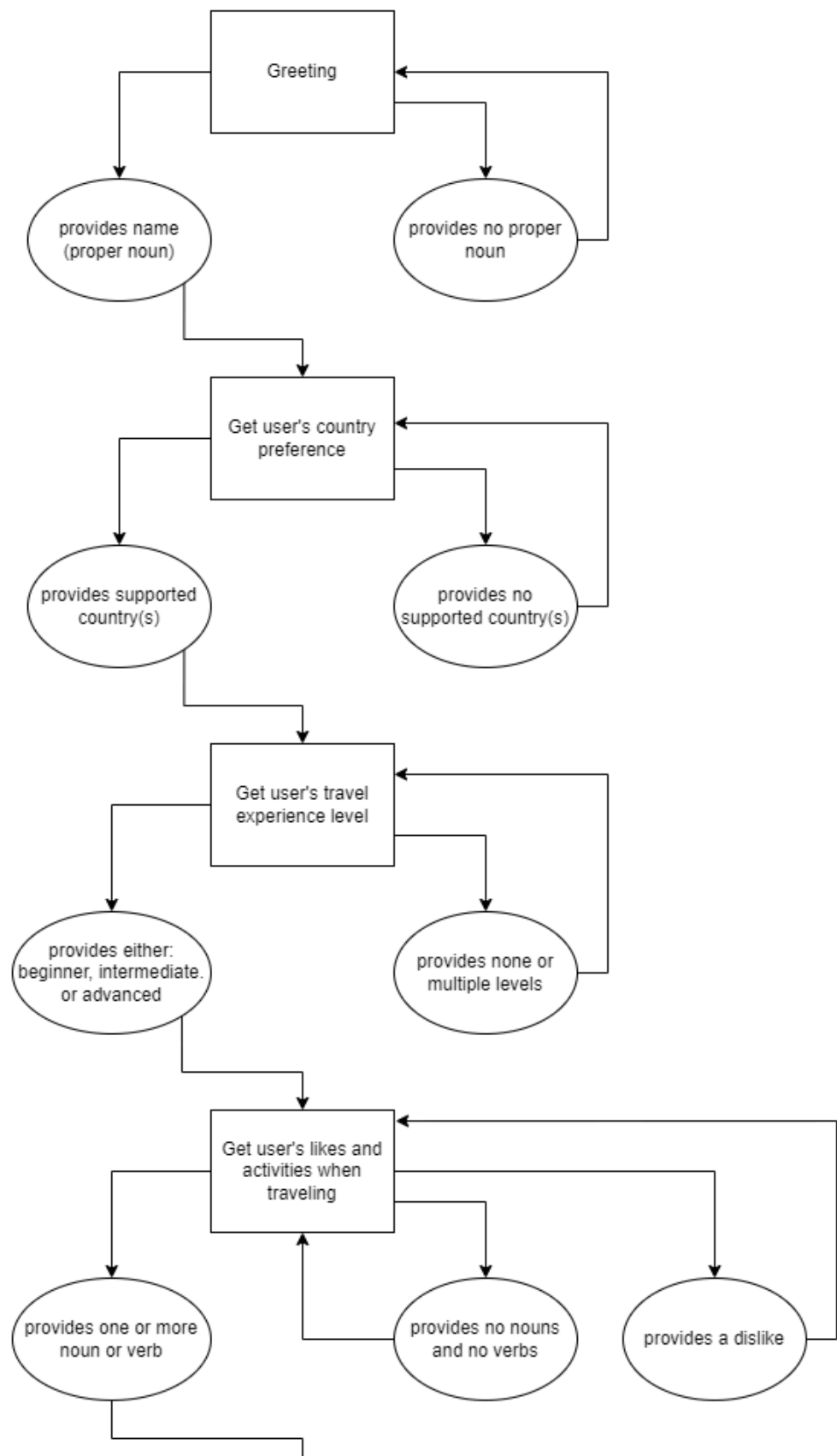
The user is free to continue the conversation as long as they like. Entering 'end' or 'bye' will stop the conversation and program. After the conversation ends, the user preferences stored in the dictionary are written to a user model text file.

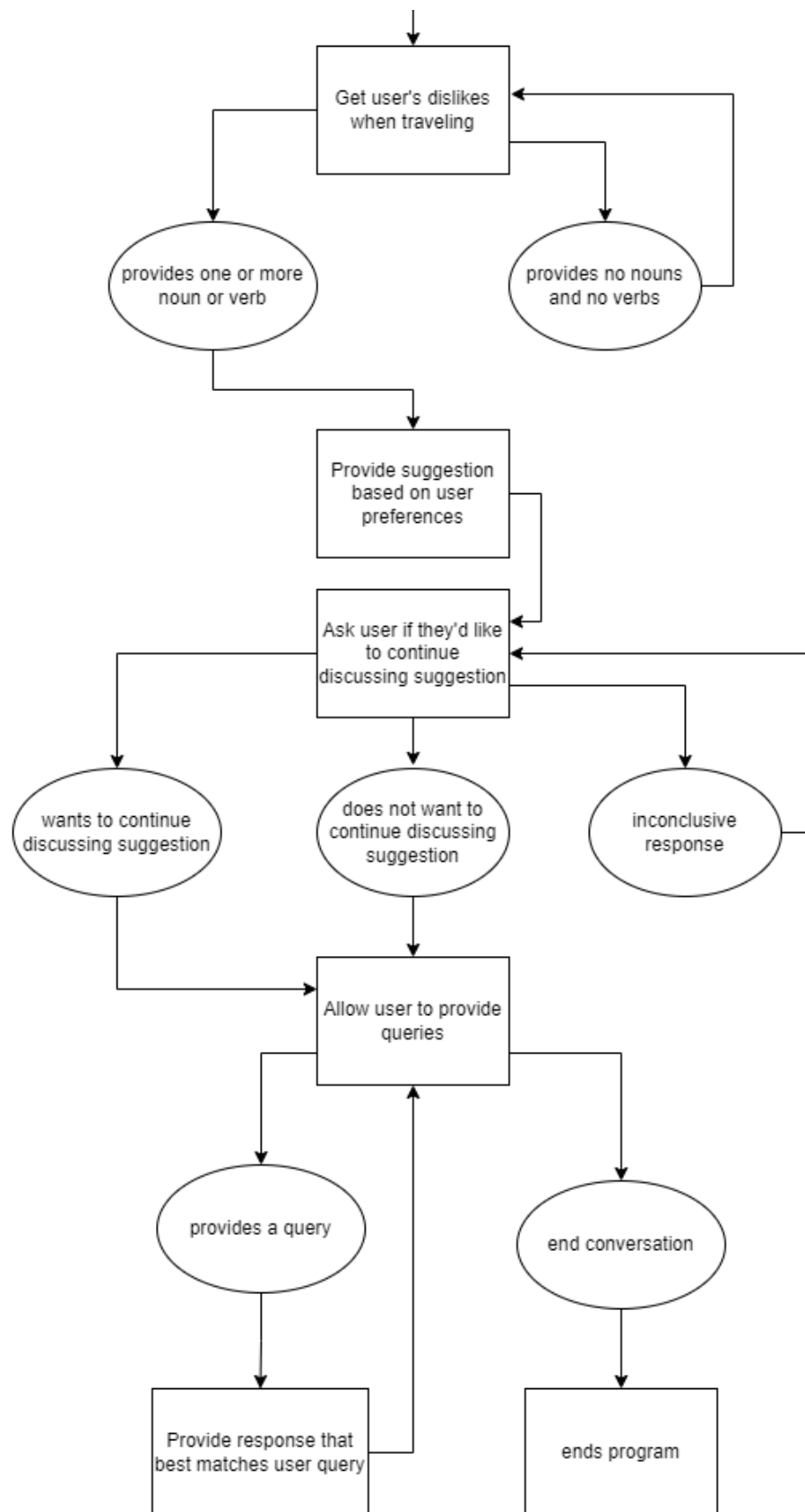
B. Dialog Tree

Key:

Chat Bot

User





C. Sample dialog interactions

Dialog 1:

```
CatalinaCB: Hola! I'm Catalina, your South America travel guide chatbot. (To end the conversation, please type 'bye' or 'end')
CatalinaCB: What's your name?
My name is Sarah
CatalinaCB: Hi Sarah! If you're looking to travel to South America, you've come to the right place! What country(s) are you looking to visit in South America?
I want to visit Argentina and Brazil
CatalinaCB: Perfect! Argentina, Brazil are beautiful places to visit! What would you rate your travel experience level: beginner, intermediate, advanced?
I'm a beginner
CatalinaCB: Awesome! Glad you're ready to explore Sarah! What things do you like when traveling (activities, places, etc.)?
I like skiing and nature
CatalinaCB: Skiing sounds like fun! Now, what things do you dislike when traveling?
museums
CatalinaCB: Thank you for sharing! Based on your responses, I'll give you a suggestion:
CatalinaCB: Cordoba is one of the most central places to visit in Argentina and a good stopover hub between Mendoza and Salta/Buenos Aires.
CatalinaCB: Would you like to continue discussing this?
yes
CatalinaCB: Ok! What can I help you with Sarah?
best skiing spots in Argentina
CatalinaCB: Here are the best places to visit in Argentina's south, a mostly mountainous, chilly, and scenic region including Patagonia.
CatalinaCB: Feel free to ask me anything else!
best skiing spots
CatalinaCB: But the skiing is rather remarkable and Cerro Castor, about 45 minutes from Ushuaia, offers world-class skiing with some of the best powder around, relatively few people on the slopes, seven restaurants and luxe cabins on-site that are the perfect accommodation choice for families.
CatalinaCB: What else would you like to know?
nature sites in Brazil
CatalinaCB: Tierra Turismo is the best option for getting you out and about into nature.
CatalinaCB: What other information can I provide for you?
what is Tierra Turismo
CatalinaCB: Tierra Turismo is the best option for getting you out and about into nature.
CatalinaCB: What else can I help you with?
end
CatalinaCB: Thanks for chatting with me! Hope to see you again:)
```

Dialog 2:

```
CatalinaCB: Hey there! I'm Catalina, a chatbot for visiting South America. (To end the conversation, please type 'bye' or 'end')
CatalinaCB: What's your name?
Bob
CatalinaCB: Hi Bob! If you're looking to travel to South America, you've come to the right place! What country(s) are you looking to visit in South America?
Uruguay
CatalinaCB: Perfect! Uruguay is a beautiful place to visit! What would you rate your travel experience level: beginner, intermediate, advanced?
I think I'm an intermediate
CatalinaCB: Awesome! Glad you're ready to explore Bob! What things do you like when traveling (activities, places, etc.)?
eating, roadtrips
CatalinaCB: Eating sounds like fun! Now, what things do you dislike when traveling?
walking
CatalinaCB: Thank you for sharing! Based on your responses, I'll give you a suggestion:
CatalinaCB: During my time in Colonia del Sacramento, Montevideo, Punta del Este, and road tripping up the coast, I put together the perfect Uruguay itinerary with the best places to see, stay, eat, and even watch the sunset all over Uruguay.
CatalinaCB: Would you like to continue discussing this?
yes
CatalinaCB: Ok! What can I help you with Bob?
museums in Uruguay
CatalinaCB: It's a pretty popular tour in Buenos Aires to take a boat ride up to the Tigre Delta – the mouth of the Rio de La Plata, the 'largest river in the world' that separates Argentina from Uruguay.
CatalinaCB: What else would you like to know?
best beaches
CatalinaCB: This itinerary is for those who want to make the most of their time in this country, accomplish the best things to do and see in Uruguay from big cities to beach towns, and make sure to explore as much as possible.
CatalinaCB: What else can I help you with?
best beach towns
CatalinaCB: This itinerary is for those who want to make the most of their time in this country, accomplish the best things to do and see in Uruguay from big cities to beach towns, and make sure to explore as much as possible.
CatalinaCB: Feel free to ask me anything else!
beach town
CatalinaCB: Vacation Rentals: There are also some nice Airbnb options in Punta del Este, La Barra, and even Jose Ignacio, a tiny little beach town full of beautiful massive houses.
CatalinaCB: What else would you like to know?
surfing in beach town
CatalinaCB: Vacation Rentals: There are also some nice Airbnb options in Punta del Este, La Barra, and even Jose Ignacio, a tiny little beach town full of beautiful massive houses.
CatalinaCB: Is there anything else you're curious about?
end
CatalinaCB: Thanks for chatting with me! Hope to see you again:)
```

Dialog 3:

```
CatalinaCB: Hello fellow explorer! I'm Catalina, a travel guide chatbot for South America. (To end the conversation, please type 'bye' or 'end')
CatalinaCB: What's your name?
Hi! I'm Mary
CatalinaCB: Hi Mary! If you're looking to travel to South America, you've come to the right place! What country(s) are you looking to visit in South America?
Chile, Colombia, and Paraguay
CatalinaCB: Perfect! Chile, Colombia, Paraguay are beautiful places to visit! What would you rate your travel experience level: beginner, intermediate, advanced?
advanced
CatalinaCB: Awesome! Glad you're ready to explore Mary! What things do you like when traveling (activities, places, etc.)?
hiking, ice cream, relaxing
CatalinaCB: Hiking sounds like fun! Now, what things do you dislike when traveling?
traffic, crowds
CatalinaCB: Thank you for sharing! Based on your responses, I'll give you a suggestion:
CatalinaCB: Sorry, I don't have enough information for your request, but this is the closest response I have:
CatalinaCB: Whether you book in advance or not, it would be wise to make a schedule of morning/evening/night tours tailored to your own interests and timings when visiting San Pedro de Atacama Chile, and book tours from there (choose from my lists below).
CatalinaCB: Would you like to continue discussing this?
yes
CatalinaCB: Ok! What can I help you with Mary?
What is San Pedro de Atacama
CatalinaCB: The driest desert in the world, San Pedro de Atacama is home to incredible geysers, salt lakes, volcanos, altiplanic lagoons, otherworldly landscapes, and a crazy amount of adventures.
CatalinaCB: What else would you like to know?
best ice cream
CatalinaCB: Now, I will break the best tours in San Pedro de Atacama Chile into times of day.
CatalinaCB: What else can I help you with?
hiking in mountains
CatalinaCB: Laguna de Los Tres is located at the base of one of the most amazing mountains in the world, Fitz Roy, and about a 10km hike out of El Chalten.
CatalinaCB: Is there anything else you're curious about?
end
CatalinaCB: Thanks for chatting with me! Hope to see you again:)
```

D. Appendix for User Models

User Model Dialog 1:

```
{'name': 'Sarah',
 'country': ['argentina', 'brazil'],
 'experience': 'beginner',
 'likes': ['skiing', 'nature'],
 'dislikes': ['museums']}
```

User Model Dialog 2:

```
{'name': 'Bob', 'country':
 ['uruguay'], 'experience':
 'intermediate', 'likes':
 ['roadtrips', 'eating'],
 'dislikes': ['walking']}
```

User Model Dialog 3:

```
{'name': 'Mary', 'country':  
['chile', 'colombia',  
'paraguay'], 'experience':  
'advance', 'likes': ['ice',  
'cream', 'hiking', 'relaxing'],  
'dislikes': ['traffic',  
'crowds']}
```

E. Appendix for Knowledge Base

Same knowledge base used for crawler and chatbot

F. Evaluation

In terms of strengths, the chatbot is able to provide responses related to the user's query to some extent. Given user preferences that closely match content from the websites, the chatbot is able to select a sentence and provide a suggestion that relates to the user's query. I noticed most responses contained at least one word from the user's preferences or queries.

In terms of weakness, when parsing the user input to get the user's name, it does not pick up on all proper nouns. For example, when I tried my own name, it didn't recognize it as a proper noun. Additionally, the spaCy NER model is not able to identify all 'GPE' entities. The model could not identify "Paraguay" and "Chile" as countries, so I had to hard code the parsing for these countries. Also, I noticed when Uruguay was listed before Argentina, the model would only pick up Uruguay as a country and not Argentina. Because of this weakness, I hard coded the parsing for Argentina as well in order to capture the user's response.

Another weakness I noticed was when the user query contained some information along with a country name, the response the chatbot provided was less relevant than when the user only mentions the information in the query. This is because the program is trying to match sentences that contain the user's information and the country name, even though it's looking through the knowledge for that specific country already, which makes it harder to identify relevant sentences. Lastly, since the knowledge of the chatbot is limited, it can repeat responses it provides as there is no other information it can access to provide more tailored responses.

Comparing my evaluation to others, I surveyed 4 other people and I asked the following questions on a 1-5 Likert scale (1 - lowest rating, 5-highest rating:

The chatbot was easy to interact with.

How relevant were the chatbot's responses to your queries?

The chatbot personalized its responses based on my input.

I would use this chatbot again for travel-related inquiries.

I would recommend this chatbot to others for their South America travel inquiries.

All participants mentioned the chatbot was very easy to use (all 5s). However, participants thought the chatbot response relevance was average (mean of scores: 3). All participants thought the chatbot did well to personalize its responses based on input (all 4s). Lastly, the average score for using the chatbot again and recommending it was 3.5.

Overall, my evaluation aligned with the participants' evaluation as the chatbot was able to answer the user queries to some extent but had some limitations.