WEBOT

The wedding planner

Saniya Sharma(20162020) Mansi Singh(20162034) M Chandravati(20162046) Yojana Birje(20162130)

Table Of Contents

- **>**Introduction
- ➤ Project Objective
- **≻**Data
- ➤ Techniques used
- **≻**Architecture
- ➤ Sample Output
- **≻**Challenges
- ➤ Future work

Introduction

One of a trend that has shaped a previous year and definitely is going to influence the year 2017 is the increasing popularity of chat bots. Chatbots is a program that simulates a real conversation with users by using artificial intelligence. A chatbot (also known as a talkbot, chatterbot, Bot, IM bot, interactive agent, or Artificial Conversational Entity) is a computer program which conducts a conversation via auditory or textual methods. Such programs are often designed to convincingly simulate how a human would behave as a conversational partner, thereby passing the Turing test. Chatbots are typically used in dialog systems for various practical purposes including customer service or information acquisition. Some chatterbots use sophisticated natural language processing systems, but many simpler systems scan for keywords within the input, then pull a reply with the most matching keywords, or the most similar wording pattern, from a database.

The rise of artificial intelligence technologies brings chatbots to the new level. Together with natural language processing Al brings up to 90% to an accuracy of machine parsing and understanding of requests. Another significant factor here is sophistication notifications that take into account context of the situation and are always on across devices.

Chatbots change the form of communication between the brand and consumer. Of course, some of them will stay with us longer, and another are going to evolve further. Anyway, there are no doubts that both users and brands can get benefits from this technology. Use cases of chatbots are shopping, payments, restaurants and delivery, healthcare, customer support etc.

We have worked on a chatbot whose aim is to assist user in wedding planning, thus coined WEBOT.

Webot interacts with the user in english language and tries to understand the type of wedding the user wants. For now we have worked on hindu, christian, muslim weddings. Webot is designed in such a way that other types can be easily added. Once the type of wedding finalised Webot needs things like location, date, budget etc to organise the wedding successfully. That can be provided by user directly or Webot can prompt user these details.

Objective of Project

We have a built a chatbot that performs mixed interaction, that is, it replies to user as well as initiates a conversation with user if required. The chatbot that we have built is goal oriented, that is we are working on specific task and we process the sentences to obtain information to perform that task. Our task is to plan a wedding.

Webot acts as wedding planner, assistant and Intermediary.

Following tasks are performed by Webot:

Webot as assistant: Webot collects information from the client and uses the information to provide valid suggestions. In case client

misses any valuable information Webot suggests by Itself.

<u>Webot as planner</u>: based on the previously provided data and the available suggestions webot organises the event .Webot

also

maintains context to see which events have been

planned

and to what extent they have been planned.

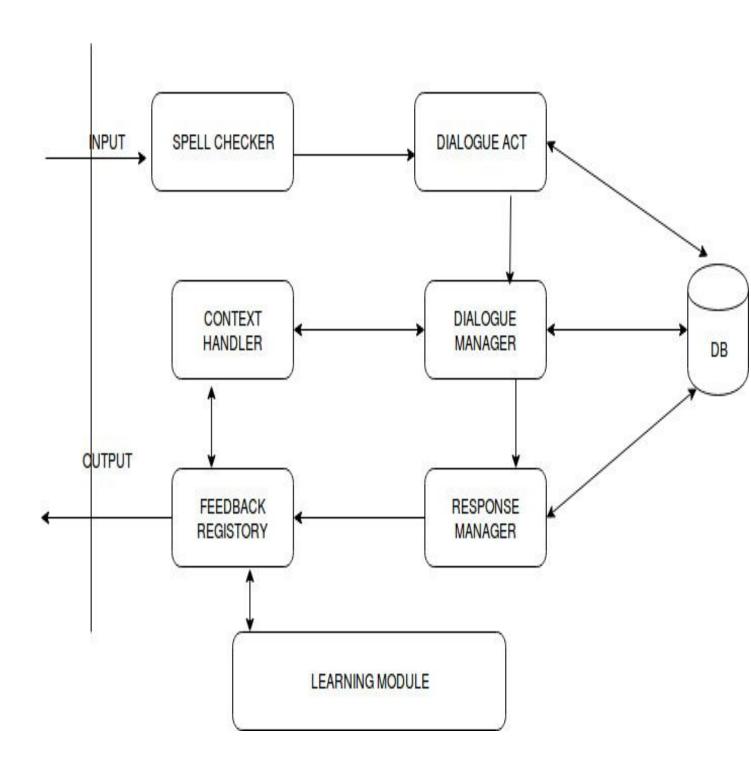
<u>Intermediary</u>: Webot keeps a data set of possible vendors .Vendors

searches

available must register with themselves .Webot

for the available vendors from the websites and stores them in its database for future use.

Model



Spell Checker

The model takes input and first it performs spell checking on the input. This helps to understand the intent of the user correctly. This module detects if there is a spelling mistake in the input and maps it to word that is nearest to it.

Dialogue Act

Dialogue act tries to identify the type of input the user has entered, that is whether it is a greeting, request, information, suggestion, reply etc.

Dialogue Manager

Based on the type of dialogue the dialogue manager decides which module should be called, context handler or response manager.

Response Manager

Response manager generates the appropriate response based on the type of input.

Context Handler

Context handler maintains the context of the conversation that we have had with the user, that is, the events that have been planned and that are yet to be planned.

Registry

Registry_ all the feedbacks,inputs,responses saved for learning model to train and test.

Data

Set of possible inputs that user can give

We have created the data for possible that user can give. For that we have generated various inputs for every category:

- → Suggestion what is your choice, which one will you suggest
- → Request what are the best venues for the event in the city ,venues around city
- → Response any dj's in town,good venues in town
- → Greeting
 Greetings such as hi,hello,good morning etc
- → Info
 iam planning for a hindu occasion,iam planning for a muslim occasion

Data Related to wedding organization

Database preparation

Libraries used

- Selenium
- webdriver
- contextlib

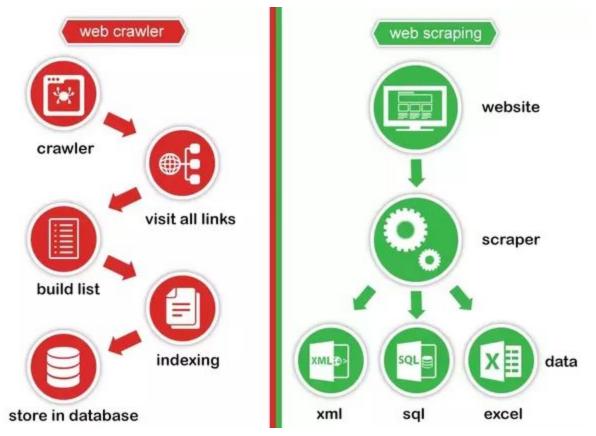
- Lxml
- Urllib2
- Html

Data is obtained using 3 phases.

> CRAWLING

In order to get data from a website programmatically, you need a program that can take a URL as an input, read through the underlying code and extract the data into either a spreadsheet, JSON feed or other structured data format you can use. These programs which can be written in almost any language are generally referred to as web scrapers. So, crawling is a way of generating a list of URLs from your scraped pages. So, How do crawlers work?

Crawlers are URL discovery tools. You give them a webpage to start from and they will follow all the links they can find on that page. If the links they follow lead them to a page they haven't been to before, they will follow all the links on that page as well. And so on in a loop.



Sample Crawler and scraper

For our project we chose a popular online travel information website "wedmegood.com". It's a dynamic web page.

A dynamic web page is a web page that displays different content each time it's viewed. For example, the page may change with the time of day, the user that accesses the webpage, or the type of user interaction. There are two types of dynamic web pages.

The first phase of crawling is done upto 3 levels:

Extracted all the urls related to services from the seed page .
 Services like : makeup, venues, djs etc

```
https://www.wedmegood.com/vendors/all/wedding-venues/
wedding-venues
                                https://www.wedmegood.com/vendors/all/wedding-photographers/
wedding-photographers
bridal-makeup
                        https://www.wedmegood.com/vendors/all/bridal-makeup/
bridal-wear
                        https://www.wedmegood.com/vendors/all/bridal-wear/
groom-wear
                        https://www.wedmegood.com/vendors/all/groom-wear/
                                https://www.wedmegood.com/vendors/all/wedding-decorators/
wedding-decorators
                        https://www.wedmegood.com/vendors/all/wedding-cards/
wedding-cards
wedding-favors
                        https://www.wedmegood.com/vendors/all/wedding-favors/
                                https://www.wedmegood.com/vendors/all/wedding-jewellery/
wedding-jewellery
                                https://www.wedmegood.com/vendors/all/wedding-accessories/
wedding-accessories
trousseau-packers
                                https://www.wedmegood.com/vendors/all/trousseau-packers/
wedding-cakes
                        https://www.wedmegood.com/vendors/all/wedding-cakes/
mehendi-artists
                        https://www.wedmegood.com/vendors/all/mehendi-artists/
                                https://www.wedmegood.com/vendors/all/sangeet-choreographers/
sangeet-choreographers
                https://www.wedmegood.com/vendors/all/djs/
dis
                                https://www.wedmegood.com/vendors/all/wedding-catering/
wedding-catering
                                https://www.wedmegood.com/vendors/all/wedding-videography/
wedding-videography
```

Fed these urls crawler to crawl all the vendor urls for each service . vendors are the services providers

```
https://www.wedmegood.com/profile/Divya-Reddy-21575/
https://www.wedmegood.com/profile/Dolly-J-Bridal-Collection-313/
https://www.wedmegood.com/profile/Anushree-Reddy-2378/
https://www.wedmegood.com/profile/Dalmia-Fashions-5912/
https://www.wedmegood.com/profile/MOHA-122576/
https://www.wedmegood.com/profile/DARSHI-SHAH-BHAVIN-TRIVEDI-28261/
https://www.wedmegood.com/profile/Nalli-Silks-3279/
https://www.wedmegood.com/profile/Neerus-Emporio-11959/
https://www.wedmegood.com/profile/Saachi-Thahryamal-33660/
https://www.wedmegood.com/profile/Sudhir-Bhai-Saree-Wale-Chandni-Chowk-5398/
https://www.wedmegood.com/profile/Varun-Bahl-88739/
https://www.wedmegood.com/profile/Varuna-Jithesh-1576/
https://www.wedmegood.com/profile/PLUMTIN-180263/
https://www.wedmegood.com/profile/Purva-Couture-24961/
https://www.wedmegood.com/profile/Ashwin-Thiyagarajan-Bridal-Wear-44627/
https://www.wedmegood.com/profile/abu-sandeep-khosla-235/
https://www.wedmegood.com/profile/Pallavi-Design-Studio-25538/
https://www.wedmegood.com/profile/Prathyusha-Garimella-19952/
https://www.wedmegood.com/profile/Riya-Kodali-Design-House-73373/
https://www.wedmegood.com/profile/Jiya-by-Veer-Design-Studio-11634/
https://www.wedmegood.com/profile/Ram-Kishan-Sarees-5904/
https://www.wedmegood.com/profile/Vasansi-21374/
https://www.wedmegood.com/profile/Chamee-and-Palak-147/
https://www.wedmegood.com/profile/Honii-Sandhu-Bridal-Couture-25072/
https://www.wedmegood.com/profile/Studio-Bandana-20962/
https://www.wedmegood.com/profile/Tek-Chand-Arjit-Goel-97419/
```

Feed these links to crawler to get the content related to the vendor and store it in DB Riantas-176757! DelhiNCR! 5.0!! ***what can I say about this super talented girl! her collection is so exquisite, exclusive n outrightly gorgeous. Every piece she cr***...**Read full review***eates so unique n breathtaking I got my engagement gown n Mehendi outfit designed from her. Both the outfits turned out like a dream! Luckily I didn't have to burn a big hole in my pocket for such exclusive unique stuff! she has a long long way to go! ***Less>>> Divya-Reddy-21575! Hyderabad! 5.0!! ***Being the bride and not worrying about outfits is not so common. But I had that privilege! All my clothing for my wedding events w***...***Read full review***ere by Divya Reddy and I couldn't have asked for anything better. Every outfit made me feel like a princess. The splendid detailing that goes into the making of the apparel amazed me. There was not a single event where I did not get compliments for my attire. All credits to DR!

> SCRAPING

Data scraping used for extracting data from websites. Web scraping software may access the World Wide Web directly using the Hypertext Transfer Protocol, or through a web browser. Scraping a web page involves fetching it and extracting data from it. The content of a page may be parsed, searched, reformatted, its data copied into a spreadsheet, and so on.

Extracted information about the vendor from the page

- Vendor name
- Location
- Vendor ratings
- Budget
- Summary
- Reviews

> DATABASE

Choosing the right database:

We know that the information internet generates is much large than yesterday. Also, that the enterprise capable of extract the value of that data will adopt better strategic decisions and will offer their customers the best products. Therefore, store and analyze the information as quickly as possible will be key. In this area arises MongoDB.

MongoDB is the leading NoSQL database, Advantages:

- It can store any type of data: structured, semi-structured and polymorphic.
- Huge performance in data process and scalability.
- It can process the large amount of information that we generate
- It supports the current requirements of the applications
- It empowers enterprises to be more agile growing quickly
- It is a document oriented database. This means that in only one document is capable of store all the information required for your product. In the fields of the document we can allocate any type of information including arrays and embedded documents. This allows documents to have a very rich and flexible structure.
- We can shape your schema design on-the-fly, leading to much faster development. This is possible because we can do it from the code side, without the need to spending our time in administrative database tasks, as it would be in relational databases.
- Developers have all the same functionalities they have in their RDBMS. So, considering our above requirements, we chose to go for MongoDB. The above scrapped information is then stored in a structured format for the project use.

Sample collection output in MongoDB

This data is used for suggesting the user about the vendor, and helps in planning based on the user's interest.

Techniques Used

DA: Dialogue Act

We annotated about 500 utterances for DAs, and the annotated utterances were used to train the DA classifier. We used deep neural network to classify dialogue acts into 7 categories.

Categories: request (req), positive feedback (pfb), negative feedback (nfb), information (info), suggestion (sug), yes-no type (yn), greeting (greetings).

Pre-processing

We padded queries with 0 to maximum length for uniform representation

Word embedding layer:

We keras word embedding layer with Glove 100 dimensional pre trained word embedding to represent queries.

Training parameters are as follows:

Activation function - softmax

Loss function - categorical cross entropy

Dropout - 20%

No. of epochs = 50

Sentence similarity

The annotated information is used when a given user utterance matched to the corresponding examples . In this process , DA is used to reduce the search space. For example, when a user inputs an utterance that is req type question, the system would not search examples that have info type sentences . After the DA matching, the system finds more similar examples to the user utterance by using the POS-tagged tokens.

The following steps are executed:

- 1. Define a set S that has POS-tagged tokens (T 1, T 2,..., T n) of the user utterance as elements.
- 2. Search examples that have all of elements of S.
- 3. If no example is searched, then remove the token Ti from S that has the lowest POS priority. If any examples are searched, then return the example set E and end the steps.
- 4. Adjust the Set S as S {Ti}, and go back to the step 2).

With these steps, the system can find the examples that have similar user utterances. Once the examples are selected, the system calculates detailed similarities between the current user utterance and user utterances in the selected examples. For the POS-tagged tokens $S = (T\ 1, T\ 2, ..., T\ n)$ of the utterance given by the real user and the tokens $S = (T\ E\ 1, T\ E\ 2, ..., T\ E\ x)$ of the user utterance in an example, the similarity between them is calculated like this:

$$Similarity(S, S_E)$$

$$= \frac{\sum M(T_i, S_E)w(T_i) + \sum M(T_{Ei}, S)w(T_{Ei})}{\sum w(T_i) + \sum w(T_{Ei})}$$

$$= \frac{2\sum M(T_i, S_E)w(T_i)}{\sum w(T_i) + \sum w(T_{Ei})}$$

Pos Taggings

In corpus linguistics, part-of-speech tagging, also called grammatical tagging or word-category disambiguation, is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition and its context—i.e., its relationship with adjacent and related words in a phrase, sentence, or paragraph. A simplified form of this is commonly taught to school-age children, in the identification of words as nouns, verbs, adjectives, adverbs, etc.

USE OF POS TAGS IN WEBOT:

- The incoming queries from user were tokenized into sentences, using NLTK string tokenizer.
- Each tokenized sentences was sent to NLTK pos tagger, which resulted in

POS tagged sentence.

- Study of POS tags in sentence helped recognize named entities of the query.

For Eg. I want to plan a wedding in Hyderabad.

Here wedding will be NOUN , Hyderabad is also a NOUN .

The lags relevant to our model is

- NN Noun, singular or mass
- ❖ CD Cardinal number
- NNS Noun, plural
- NNP Proper Noun, singular
- VBN Verb, past participle

CD: For budget related information and date Rest for extracting services, events, location etc.

Example

Input: I want to plan a wedding in Hyderabad

Output: we get POS tagger:

[('i', 'NN'), ('want', 'VBP'), ('to', 'TO'), ('plan', 'VB'), ('a', 'DT'), ('wedding', 'NN'), ('in', 'IN'), ('hyderabad', 'NN')]

Thesaurus

In general usage, a **thesaurus** is a reference work that lists words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms), in contrast to a dictionary, which provides definitions for words .

To generate all the synonyms we have refered thesaurus.com and some of the local word like synonyms of mehendi we have used local dictionaries for knowledge building.

We have also considered similar meaning words

Ex: Both bus and taxi refers to service transportation

All 'marriage', 'wedding', 'nuptial', 'matrimony' refer to wedding only.

```
['venue']=['place of wedding','venue','place of marriage','place of nuptials','place']
['dj']=['jockey','dj','music','djs','jockies']
['card']=['card','invitation']
['cake']=['cake','pastry']
['photo']=['photo','photography','shutterbug','photographer']
['catering']=['caterer','food','buffet','dinner','lunch','catering']
['video']=['video','coverage','videography']
['makeup']=['makeup','parlour','beautician']
['decoration']=['decoration','decorator']
['jewellery']=['jewellery','bangel','necklace','ring']
['transportaion']=['bus','car','taxi','transportation','flights','aeroplanes','transport']
['choreographer']=['choreographer']
['accessories']=['accessories','purse','shoe']
['accomodation']=['accomodation','stay','hotel','lodge']
['trousseau']=['gift','trousseau','wrapper']
['clothing']=['clothing','wear','shopping malls','shopping']
```

Spell Checker

Spell Checker was used in Webot to provide good response to user even if user happens to write wrong spelling in query. Wrongly spelled words may occur due to i) incomplete knowledge of spelling ii) Mistyping words when typing fast/or using keyboard.

We had a dump of all the data we collected from our travel website. This dump was used to form the list of words in my dataset, say myVocabulary. Spell Checker worked to find Edit Distance of a wrongly typed word from the words in myVocabulary. Our implementation used a spell checker to find nearest word with edit distance 2 from mispelled word.

iam planning for A wedding in hderabad : Query before spell checker iam planning for a wedding in hyderabad : Query after spell checker

Architecture

Once user provides input in the form of a query it is first sent for spellchecking . Any corrections required are performed here . After that the corrected input is passed to Dialogue act to get the type of the input

Dialogue Act

Dialogue Manager

Once we have the dialogue and its act dialogue manager takes decision according to the dialogue act. It is the central part connecting all the important modules of our architecture.

Type of actions:

- 1. NO information type: Dialogue of the form Greetings, yes or no type, Negative Feedback. Response handler handles these dialogues.
- 2. Contains information: Dialogue of the form Information, Request, Suggestion, Positive feedback. Extract information from these sentences, update the client frame and take proper action like passing the information to context handler so that current state of client can be updated or call response handler to generate response.

For Information related queries extract information get binding information

What is binding Information of a dialogue?

Suggest few good chocolate cake suppliers in Hyderabad .

Here the binding information for the sentence is :

Location: hyderabad

Service: cake

Flavour: Chocolate

- Based on the dialogue act output get the type of dialogue.
- Get most relevant queries from the set based on POS taggings .
 using sentence similarity .
- Now check whether binding information matches output the best matched dialogue.
 - Get POS taggings of the dialogue
 - NN Noun, singular or mass
 - CD Cardinal number
 - NNS Noun , plural
 - NNP Proper Noun, singular
 - ❖ VBN Verb , past participle
 - Use Synonyms list to identify the information

Extracted informations include:

- Event: Wedding/Reception/Sangeet/Mehendi/Engagement/Istikhara etc
- Wedding type: Hindu/Muslim/Christian etc
- Location : List of places
- Budget
- Service : All the services provided by the Webot
- Date

Handling the Dialogue

- 1. Greeting, Yes / No, Negative feedback: Pass it to response handler.
- 2. Information: Updations in the frame, pass it to context handler.
- 3. Positive feedback: Get the feedback if any.
- 4. Suggestion, Request: Pass all the information to the response handler.

Context Handler:

It maintains the information about events(e.g. Reception , haldi, sangeet etc)

and its services(e.g catering, clothing,djs etc) and also maintains budget and

location of services like djs, catering, jewellery etc.

It maintains current context like event and its service.

We use class to store all the information . We create a class called Events and every time for a new event we create object with the name of the event . We followed the below structure :

```
Events (Class objects)----->
(Reception,sangeet etc)
Services (Frames)----->
(Catering - S1, clothing - S2 etc)
S1 ==> [ location , Budget ]
S2 ==> [ location , Budget ]
```

Flow of the information:

Dialogue Manager -----> Context Handler ----> Dialogue Manager

Input to context Handler:

Dialogue manager gives input to the context handler, dialogue manager parse the current input and give information like what is the name of the event, location etc to the context handler.

Output from the context handler to the dialogue manager:

We maintain information about the local event, local service and details of the service like location, budget. There is a dictionary to maintain the information of the events.

We are maintaining completed, current, and not required services information and to maintain these information, we used flags.

Flag -1 ⇒ initial value

Flag $0 \Rightarrow$ It indicates that this service is not required.

Flag 1⇒ It indicates that the service is started.

Flag $2 \Rightarrow$ it indicates that the service is completed.

Sample Conversation

```
HI Webot
                                Welcome to Webot !!!
i want to plan a hindu wedding
                                Thanks for the information.
iam planning for reception in hyderabad
                                Thanks for the information.
suggest good venues
                                Below are the most suitable vendors for venue
suggest good venues in hderabad
                                Below are the most suitable vendors for venue
 Vendor: jrc-conventions-24971 Ratings :5.0
 Vendor: princeton-convention-center-21119 Ratings :5.0
 Vendor: novotel-hyderabad-23206 Ratings :4.5
 Vendor: fort-grand-31093 Ratings :4.8
 Vendor: marigold-hotel-15653 Ratings :4.0
 Vendor: katriya-hotel-17825 Ratings :5.0
 Vendor: siddh-convention-17978 Ratings :4.5
 Vendor: the-park-37864 Ratings :5.0
i like princeton
 Vendor: princeton-convention-center-21119 Ratings :5.0
are the ratings good
                                        Yes, this is the best available option.
okay
```

```
Vendor: eat-drink-design-37980 Ratings:5.0
Vendor: le-kitchen-23622 Ratings:5.0
Vendor: crumbs-to-gourmet-16645 Ratings:5.0
 Vendor: m-talwar-catering-services-16682 Ratings :5.0
is the service reliable
                                               Yes, this is the best available option.
i like crumbs
Vendor: crumbs-to-gourmet-16645 Ratings :5.0
i want to plan engagement
iam planning for engagement
                                     Thanks for the information.
suggest cake shops in mumbai
                                     Below are the most suitable vendors for cake
 Vendor: deliciae-patisserie-115 Ratings :5.0
Vendor: lil-white-oven-20454 Ratings :5.0
Vendor: mishmash-30226 Ratings :5.0
 Vendor: haagen-dazs-14581 Ratings :5.0
Vendor: happy-16829 Ratings :5.0
 Vendor: wedding-cake-by-celestial-37952 Ratings :5.0
 Vendor: the-cake-lady-by-neeta-gurbaxani-37938 Ratings :5.0
is the service reliable
                                               Yes, this is the best available option.
i like deliciae
iam planning for hindu wedding
                                    Thanks for the information.
iam planning for reception n delhi
                                    Thanks for the information.
                                    Hi which service you want to avail !!!
suggest venues in hyderabad
                                    Below are the most suitable vendors for venue
 Vendor: jrc-conventions-24971 Ratings :5.0
 Vendor: princeton-convention-center-21119 Ratings :5.0
 Vendor: novotel-hyderabad-23206 Ratings :4.5
 Vendor: fort-grand-31093 Ratings :4.8
 Vendor: marigold-hotel-15653 Ratings :4.0
 Vendor: katriya-hotel-17825 Ratings :5.0
 Vendor: siddh-convention-17978 Ratings :4.5
 Vendor: the-park-37864 Ratings :5.0
                                    Hi are you busy
```

Hi are you busy

Below are the most suitable vendors for catering

princeton-convention-center-21119 : Selected vendor

suggest food centers in mumbai

i like princeton

Vendor: mini-punjab-catering-22051 Ratings :5.0

Vendor: little-food-co-788 Ratings :5.0

```
HI Webot
                                                 Welcome to Webot !!!
i want to plan a hindu wedding
                                                 Thanks for the information.
iam planning for reception in hyderabad
                                                 Thanks for the information.
suggest good venues
                                                 Below are the most suitable vendors for venue
suggest good venues in hderabad
                                                 Below are the most suitable vendors for venue
 Vendor: jrc-conventions-24971 Ratings :5.0
 Vendor: Jrc-conventions-24971 Ratings:5.0
Vendor: princeton-convention-center-21119 Ratings:5.0
Vendor: novotel-hyderabad-23206 Ratings:4.5
Vendor: fort-grand-31093 Ratings:4.8
Vendor: marigold-hotel-15653 Ratings:4.0
Vendor: katriya-hotel-17825 Ratings:5.0
Vendor: siddh-convention-17978 Ratings:4.5
Vendor: the-park-37864 Ratings:5.0
i like princeton
 Vendor: princeton-convention-center-21119 Ratings :5.0
are the ratings good
                                                             Yes, this is the best available option.
okay
```

Challenges

Data

Named-Entity recognition

Handling transition of states

Handling wrong spellings for more than 2 distance

• Future work

Handling chat abbreviations

Mixed languages