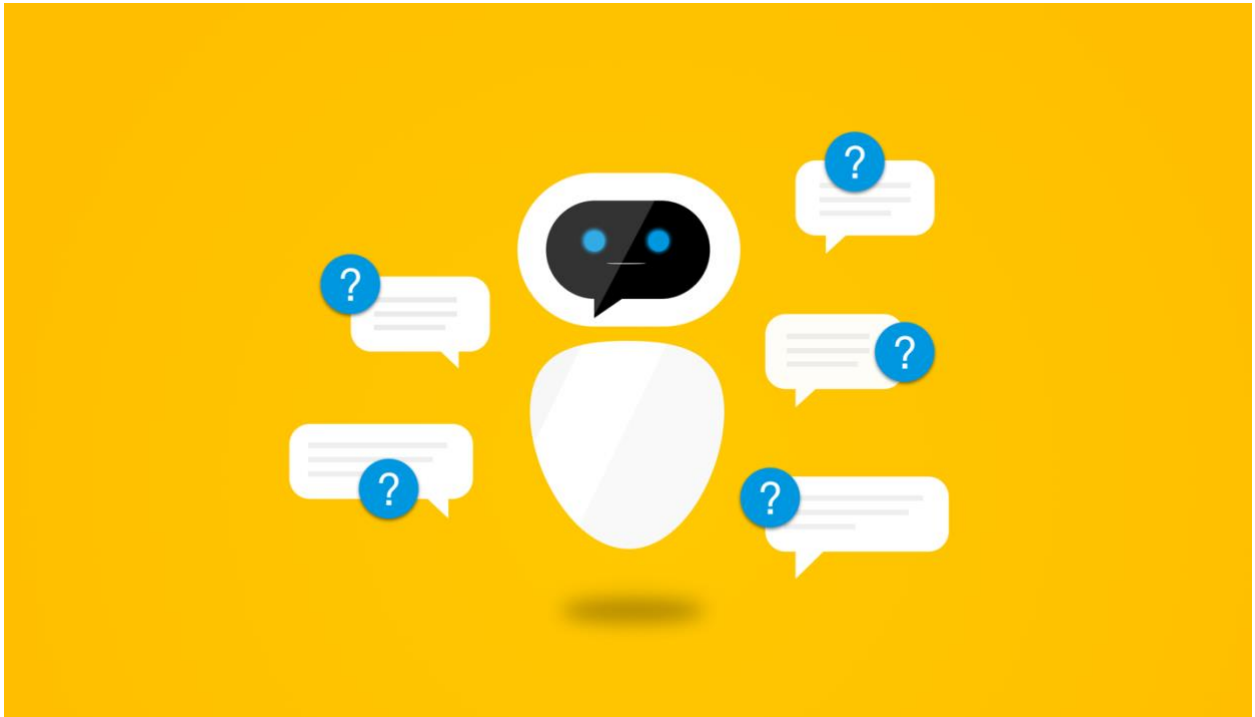


Final Project Report

Implementation of a Chatbot



AIDI 2000-01-Applied Machine Learning

Facilitator: Noopa Jagadeesh

April, 2020

Chutu Li 100765238
Keng Hin Cheong 100711777
Loc Nguyen 100741710

Abstract

We developed a chatbot for the project, which can understand user's questions by using SpaCy's natural language understanding process. This chatbot can not only handle special questions with the questioning words "what", "why", "when", etc., but use keywords in the questions as search criteria to find and extract information from corresponding Wikipedia entries. And lastly, it gives all concise answers to the user. On the other hand, when the chatbot does not understand the out-of-range inputs or simply has no answer for the entry, it will respond accordingly to guide the user to change his/her input. Therefore, this chatbot can certainly be used as a student supportive education system, even for students at their young age.

Implementation of a Chatbot Answering Knowledgeable Questions

1. Introduction

Chat robot (Chatbot) is a computer program for talking via dialogue or text and is ability to simulate human conversation.

Chatbots can be used for practical purposes, such as customer service or information acquisition. Currently, chatbots are part of virtual assistants and can be connected to many organizations' applications, websites, and instant messaging platforms. Non-assistant applications include chat rooms for entertainment purposes, research and specific product promotions, and social robots, etc.

NLP-Natural Language Processing, which is an artificial intelligence technology designed to interpret, recognize and understand user requests in the form of free language. NLP-based chatbots can understand customer queries written in their natural language and immediately respond to them. NLP and other machine learning technologies enable chatbots to conduct most conversations easily and without human assistance. Just ask the chatbot in natural language, the user can easily get all the information, and the chatbot can be perfectly explained with accurate answers.

At present, conversational chatbots have created many applications, such as intelligent customer service applications in various industries. The conversational chatbot, as a typical application scenario of artificial intelligence, is also a very challenging task, not only involving the representation, extraction, reasoning and application of multiple heterogeneous knowledge

(such as language knowledge, domain knowledge, common sense knowledge, etc.) It also involves the comprehensive utilization of other artificial intelligence core technologies (such as user portraits, dialogue management, etc.) including natural language understanding.

Some chatbots for education purposes make learning simpler and more efficient in response to the problems faced by students.

2. Industry selection – a knowledge and information chatbot

With the continuous advancement of chatbot technology, the application of chatbots appears in almost any imaginable field. Recently, robots have begun to enter the field of education, and the benefits it brings is foreseeably to be profound.

The general advantage of robots is that they can respond quickly and available at any time.

In education, time, funds and resources are usually limited. Teachers have discovered that providing chatbots for students has many benefits. The more tasks a robot can handle, the more time educators can focus on their main role: education.

We intend to provide a basic chatbot to help children quickly acquire basic knowledge and tips when they encounter new knowledge or new terms during reading and other learning processes.

3. Analysis on popular chatbot building platforms

We will look at this year's (2020) top 5 platforms of NLP Chatbot:

IntelliTicks

IntelliTicks is one of the fresh and exciting AI dialogue platforms that have appeared in recent years. Companies around the world are deploying the IntelliTicks platform to engage and attract potential customers. IntelliTicks Chatbot uses natural language processing (NLP) to answer repeated customer inquiries instantly, without manual operations.

Through end-to-end automation, IntelliTicks Chatbot can interact with visitors instantly. Its Ai-Powered Chatbot comes with manual fallback support, which can transfer control of the conversation to a human agent when the chat bot cannot understand complex customer queries. Enterprises can design custom chatbots according to their own needs and set up the dialogue process.

- Deeper API integration, such as Zapier
- "Pay in chat" option
- Has a chat menu option
- With manual takeover option-yes
- Suitable for SMEs
- Drag and drop function, no coding knowledge required
- Analysis dashboard
- Free plan - yes
- Functional restrictions under the free plan-no
- Deployment environment: mobile, website, login page, Facebook Messenger
- Live Chat - Yes
- Conforms to GDPR

Google dialog flow

Dialogflow is an artificial intelligence software used to create chatbots to attract online visitors. It handles automatic dialogue based on NLP. It can be used for AI-based text and voice chat

robots. Dialogflow combines Google's machine learning expertise and products, such as Google Cloud Speech-to-Text. Dialogflow is a Google service running on Google Cloud Platform that allows you to scale to hundreds of millions of users. Dialogflow is the most widely used tool for building operations for more than 400 million Google Assistant devices. Below are its features:

- Multi-lingual
- Support Google Assistant, Alexa, Cortana, Facebook Messenger
- Inline code editor
- Cross-platform support
- Analysis
- Agent creation and management
- Entity
- Free plan

Wit.AI

Wit.ai makes it easy for developers to build applications and devices with which they can talk or text. Our vision is to provide developers with an open and extensible natural language platform. Wit.ai learns the human language through every interaction and makes the most of the community: what is learned is shared among developers. Its features:

- Supports more than 50 languages
- Free
- Support text and voice
- Conforms to GDPR
- When your application opens, the community can access your intents, entities, and verified expressions, but the logs cannot.
- Can also be configured by non-technical personnel
- Support mobile wearable devices, home automation, wearable devices

Haptik

Haptik is an enterprise conversational AI platform. The most advanced algorithm can add intelligence to the robot. A simple and powerful tool for designing, building, and maintaining chatbots-the dashboard can view reports on chat metrics and receive conversation overview. Integrate with any existing or custom CRM system. Below are its features:

- Voice Robot
- Multi-lingual
- Based on NLP
- LiveChat support
- Analysis dashboard
- Prices are not disclosed
- Not suitable for SMEs

Kore.AI

The industry-leading enterprise-level end-to-end conversational AI robot program platform for designing, creating, training, testing and analyzing AI and NLP chat robots for all popular consumer and enterprise communication channels. Its features:

- Pre-built integration and authentication with recording systems (SAP, Salesforce, ServiceNow, etc.)
- Give enterprises complete control over voice engine training, auditability and compliance
- Support includes IP phones, wearable devices, and other gadgets that support voice functions, such as Amazon 's Echo, etc.
- Develop websites and mobile applications through Kore.ai 's platform SDK
- Workforce collaboration tools, including Sk

After the review, we decided to build the chatbot from scratch rather than building it on a platform, as this way we will learn more about the working mechanism.

4. Design and Building the Chatbot

For the designing, the pipeline of the model will be described along with constraints and other solution to improve it.

User input is used as a question. Next, the tag function will be called to employ the SpaCy library and features to assign a part-of-speech tag to each word in the question. The part-of-speech and dependency tags, the question type, subject and detail are recorded from the question.

After that, the response function is then called, which uses the subject to search for a corresponding content in Wikipedia page. After searching through the page, it matches the most relevant sentences based on the question type of the question.

For instance, for "when" type questions, it searches for a "time" entity based on SpaCy's Named Entity Recognition. If a sentence matches the question, it is added to a list of 4 answers which are displayed to the user. The user can choose the most accurate answer, which is then recorded and stored with the question on the file Recorded Answers

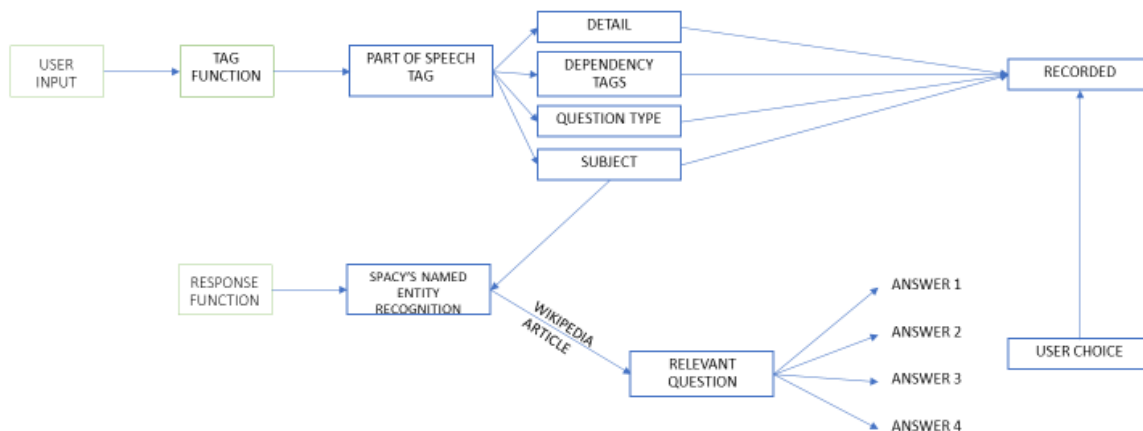


Figure 1 Chatbot pipeline

5. Discussions on the functioning of the chatbot

The limitations of the chatbot

The chatbot fails on questions such as “How long is the Nile River?” or “How large is Asia?”, where “how” is used to ask about the extent of something, due to the fact that the bot is unable to detect and record adjectives as the detail of the question.

My Response:

- 1)
iteru, meaning "river", Coptic: $\phi\iota\alpha\rho\omega$) is a major north-flowing river in northeastern Africa, and is the longest river in Africa and the disputed longest river in the world , as the Brazilian government says that the Amazon River is longer than the Nile.
- 2)
The Nile, which is about 6,650 km (4,130 mi)
- 3)
long, is an "international" river as its drainage basin covers eleven countries: Tanzania, Uganda, Rwanda, Burundi, the Democratic Republic of the Congo, Kenya, Ethiopia, Eritrea, South Sudan, Republic of the Sudan, and Egypt.
- 4)
In particular, the Nile is the primary water source of Egypt and Sudan.

Which answer do you think is the most accurate one?
Type in the answer number on the list above:

Figure 2 Question: How long is the Nile River?

The bot should give a numerical value associated with the adjective in the question. The bot also currently doesn't answer 'why' or 'how' questions very well because it can't use SpaCy's named entity recognition to search for an answer in Wikipedia. It can only search for keywords such as “because” or “using” and question detail in Wikipedia. This means that if the Wikipedia article does not contain a sentence with the keywords or the detail, then it can't return an answer. Another issue is that searching for detail in Wikipedia articles is not always effective, because articles can often use synonyms such as “constructed” instead of “built”. This makes the bot unable to find exact words in Wikipedia. One way to solve this would be to use a form of thesaurus to search for keywords and words with similar meanings to the keywords in articles.

Currently, the bot searches for keywords in each sentence in the article, then returns the first sentence with the keywords as an answer. This can lead to inaccuracies in answers if the keyword is a common word. One way to improve this would be to use a scoring system. The bot would search through every sentence in the article's summary, then assign a score to each sentence based on how suitable the sentence is for an answer. For example, a sentence with more keywords in it would have a higher score and would therefore be selected as the answer. The bot can only answer questions starting with what, why, where, when, who and how, and fails on questions such as "Are whale mammals?". One way to solve this would be to introduce a new question type for questions with yes/no answers. The bot could for example, check if "whales" and "mammals" are in the first sentence of the article, then return "yes" as an answer.

```
Ask me something, my boss! I will try my best to answer you! (type "quit" to exit): Ar
e whale mammals?
```

```
-----
Analysis of Your Entry:
```

```
Are AUX VBP ROOT Are ROOT
whale ADJ JJ amod mammals attr
mammals NOUN NNS attr Are ROOT
? PUNCT . punct Are ROOT
```

```
-----
My Response:
```

```
Sorry! That's not something I can answer for! Please try to ask me something else...
```

```
Ask me something, my boss! I will try my best to answer you! (type "quit" to exit):
```

Figure 3 The bot can only answer questions starting with what, why, where, when, who and how, and fails on questions such as "Are whale mammals?"

The correct functioning and performance of the chatbot

The bot can answer questions starting with what, why, where, when, who and how, and provides terse and correct answers. Here below are examples of its performance.

Trial question 1: "What is the weather?" and the chatbot's outcome:

Ask me something, my boss! I will try my best to answer you! (type "quit" to exit): What is the weather?

Analysis of Your Entry:

What PRON WP attr is ROOT
is AUX VBZ ROOT is ROOT
the DET DT det weather nsubj
weather NOUN NN nsubj is ROOT
? PUNCT . punct is ROOT

Question type: what
Subject: weather
Detail: is

My Response:

1)
Weather is the state of the atmosphere, describing for example the degree to which it is hot or cold, wet or dry, calm or stormy, clear or cloudy.

Figure 4 The chatbot's answer to the question: "What is the weather?"

Trial question 2: "When is the Canada day?" and the chatbot's outcome:

Question type: when
Subject: canada day
Detail: is

My Response:

1)
A federal statutory holiday, it celebrates the anniversary of July 1, 1867, the effective date of the Constitution Act, 1867 (then called the British North America Act, 1867), which united the three separate colonies of the Province of Canada, Nova Scotia, and New Brunswick into a single Dominion within the British Empire called Canada.

2)
Originally called Dominion Day (French: Le Jour de la Confédération), the holiday was renamed in 1982, the year the Canada Act was passed.

Which answer do you think is the most accurate one?
Type in the answer number on the list above:

Figure 5 The chatbot's answer to the question "When is the Canada day?"

From Figure 5, it can be seen that the chatbot performs well with correct answers as long as the questioning words are provided. Meanwhile, Figure 6 below shows that the user can choose the best answer on the list, and the best answer along with the question will be saved into a text file called "Recorded Answers".

```
1)
A federal statutory holiday, it celebrates the anniversary of July 1, 1867, the effective date of the Constitution Act, 1867 (then called the British North America Act, 1867), which united the three separate colonies of the Province of Canada, Nova Scotia, and New Brunswick into a single Dominion within the British Empire called Canada.

2)
Originally called Dominion Day (French: Le Jour de la Confédération), the holiday was renamed in 1982, the year the Canada Act was passed.

Which answer do you think is the most accurate one?
Type in the answer number on the list above: 1

OK, boss! Then my final answer for your question is:
A federal statutory holiday, it celebrates the anniversary of July 1, 1867, the effective date of the Constitution Act, 1867 (then called the British North America Act, 1867), which united the three separate colonies of the Province of Canada, Nova Scotia, and New Brunswick into a single Dominion within the British Empire called Canada.

Ask me something, my boss! I will try my best to answer you! (type "quit" to exit):

```

Figure 6 The chatbot's answer to the question "When is the Canada day?"

6. Conclusion

After some testings on the chatbot, we can conclude that we have successfully developed a chatbot that can help children quickly acquire basic knowledge for their reading and studying. The chatbot uses SpaCy natural language process to understand questions input by the user. This chatbot can not only handle special questions with the questioning words "what", "why", "when", etc., but use keywords in the questions as search criteria to find and extract information from corresponding Wikipedia entries. And lastly, it gives all concise answers to the user. On the other hand, when the chatbot does not understand the out-of-range inputs or simply has no answer for the entry, it will respond accordingly to guide the user to change his/her input. Therefore, this chatbot can certainly be used as a student supportive education system, even for students at their young age. For future improvements of the chatbot, we could try to expand the range of the knowledge referencing of the chatbot, so that it can cover broader range of inputs.