

Michael Zhao

CS 4395.001

15 April 2023

Chatbot Report

Summary

The chatbot that I created helps users with questions regarding the game Hollow Knight. All data was scraped directly from the Hollow Knight Wiki and stored locally to form a knowledge base. The chatbot uses the Rasa framework to interface with the user. This framework allows for an easy way to understand the user inputs and storing user sessions. A custom model is also run underneath the Rasa engine to manually detect user intents and determine the topic of each sentence. After determining the topic and question type, the model will send the relevant information to the user.

System Overview

1. Scraping the data

The data was scraped directly from the Hollow Knight Wiki. I went to the website directory, which listed all pages that were present on the wiki. I took the data from that site and got all the subpages. I then made a request to each URL, fetched the HTML, and used the BeautifulSoup library to parse the HTML files. With BeautifulSoup, I was able to extract the relevant data I needed from the web pages, such as character information, item descriptions, and enemy stats. Once I had the data, I wrote each page to a separate file in raw HTML format that would be easy to work with. This involved creating a new file for each page and giving it a name based on the page's title. I did this to not continuously scrape from the website, and on future runs I could just read the data from the files. I named this folder "output-raw".

2. Parsing the data

To further process the data, I used various techniques to clean and normalize it. For example, I removed unnecessary HTML tags, extracted only the text content, and converted any non-ASCII characters to Unicode. I also used regular expressions to search for and replace any instances of misspelled words or inconsistent formatting. To extract the text, I used BeautifulSoup's `gettext` method to extract only the text. This was also written to its own directory, named "output-data".

3. Forming the Knowledge Base

Here is the first instance of deep learning being used! I used the BART Large model and the zero shot classification pipeline to set up the preprocessing of the data. I took each document, read in the lines as sentences (splitting at the periods), and fed each line into the zero-shot classification pipeline. This allowed me to classify each sentence into a specific intent. The intents consisted of "info", "trivia", "behavior_tactics", and more. Each set of sentences that represent the intent would be grouped and stored together in the knowledge base. The knowledge base was a dictionary of topics (1 for each document), and each topic consisted of a dictionary of intents. To store this knowledge base, I simply created a separate file for each document and stored the dictionary of intents.

4. Training the Mini-Classifier

I created a small model to classify questions into the relevant intent types. This involved writing out sample questions in the "intents.json" file and feeding it into a Keras model. The model consisted of 2 hidden layers with 128 neurons each and 1 output layer with 3 neurons. I simply used a Sequential model with 3 layers and 2 dropout layers in between. I trained this model for 300 epochs using an Adam optimizer. Before I was able to feed my raw intents into the model, I had to tokenize and lemmatize the documents. I used NLTK's `word_tokenize` method to tokenize the words. I then lemmatized each word, converted them to lowercase, and removed all punctuation.

These were all thrown into a set to prevent duplicates. Now that we had a set of words, we needed to convert it to a vector that a model could understand. For this, I used the bag of words model (for its simplicity) to create the training data for the model. This would simply create an array with every word found in the input space and either mark a 0 or 1 if it was present in the current query. This was done for each query, then split up into a training and test set. Now we could feed this into the model, and it generated a decently accurate model that we saved to an h5 file.

5. Rasa Framework

I primarily utilized the Rasa framework to communicate with users as it had pre-built functionality to support user session interactions, infrastructure for adding custom intents, and a nice interactive interface for training the bot with new questions. I also used Rasa's custom actions to manually search my database. Rasa would extract the topic from the given sentence and call my custom action. Rasa also handled conversation points that didn't fully require my knowledge base, such as if the user greets the bot or asks what the bot can do.

6. Retrieving Information from the Knowledge Base

Whenever the custom action got a query, we would process the text and feed it into the secondary model to get the intent. We would have to do the same kind of data parsing that we did when we trained the model. This includes tokenization, lemmatization, and converting the query into the bag of words model. This was fed into the bot and the most likely intent was outputted. In addition to this, we had to consider possible misspellings or slight differences when reading the topic from the user query. To do this, we take a combination of what Rasa tells us and some manual checking to make sure the topic is correct. This is done by extracting the noun phrases from the sentence (I simply used Spacy's `noun_chunks` tool). After doing this, we can match it with the correct topic. We iterate through the set of topics and calculate the Levenshtein distance between the current topic that the user is

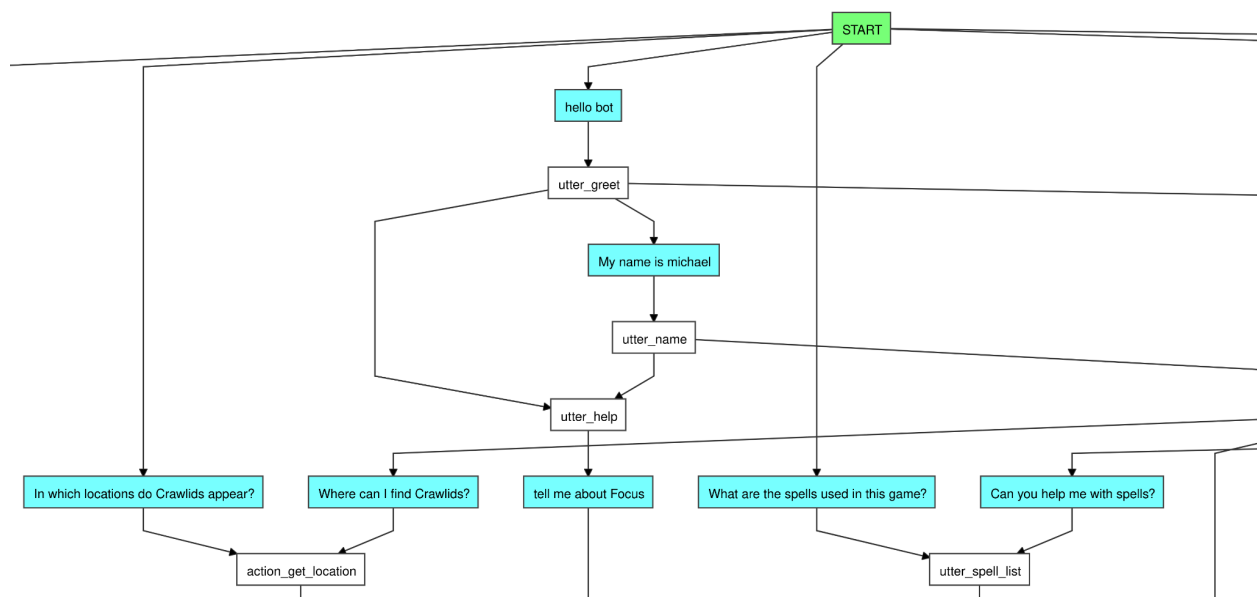
querying for and the actual topic names. We find the topic name that is closest in Levenshtein distance to the query and use that as the topic. I used a simple string similarity algorithm because I found it to be the fastest and easiest to use, without sacrificing much accuracy at all. Now that we had the topic and intent, we could simply do a search through the knowledge base dict to check if the topic had that intent. If it did, we would return the information stored to the user. Otherwise, there would be a default fallback (ie. "Sorry I can't answer that question").

Dialog Flow Tree

Rasa has a nice tool to generate this tree in an HTML format. The interactive version can be found at the following URL:

<https://cdn.michaelzhao.xyz/archive/hollow-knight-chatbot-graph.html>.

Here is an image with a portion of the graph:



Sample Conversation Flow

Your input -> Hello!

Hello! I am a chatbot, designed to help you with anything related to Hollow Knight!

What is your name?

Your input -> My name is Michael

Nice to meet you Michael!

How can I help you today?

Your input -> What is a Primal Aspid?

Primal Aspid Health Drops 35 26 (Spawned by Ascended/Radiant Collector) 6 / 8
None (Colosseum of Fools , Spawned by Ascended/Radiant Collector) Hunter's notes requirement 25 kills. Wilder, ancestral form of the Aspid. How strange that their descendants are so weak. If I have children, I hope they will be stronger than me. Primal Aspids are enemies in Hollow Knight. They are the ancestral form of the Aspids , once thought to be extinct. They can easily be detected by the whistling sound their wings make. Primal Aspid Hunter's Journal entry: "Wilder, ancestral form of the Aspid.

Do you have any more questions about Primal Aspids? I can tell you about general info, location, behavior, tactics, journal entries, or trivia.

Your input -> Where are Primal Aspids located?

Once thought extinct, they have reappeared at the edges of the world. They are often found in groups, assisting other enemies, or in platforming areas where the terrain makes them hard to reach. They are also fought in the Colosseum of Fools , where they should be the highest priority to kill. Location They can be seen throughout all of Kingdom's Edge , except for the southwest portion. They can also be spawned by the Ascended/Radiant versions of the Collector and are present in all the trials at the Colosseum of Fools. Locations in Kingdom's Edge Gallery Idle Spitting In the Colosseum of Fools Spawned by the Collector References.

Do you have any more questions about Primal Aspids? I can tell you about general info, location, behavior, tactics, journal entries, or trivia.

Your input -> Thank you!

Goodbye Michael! Hope I could help you with your questions.

Sample Conversation Flow 2

Your input -> Good morning!

Hello! I am a chatbot, designed to help you with anything related to Hollow Knight!

What is your name?

How can I help you today?

Your input -> My name is Michael

Nice to meet you Michael!

How can I help you today?

Your input -> What can you do?

[no response]

Your input -> What spells are there in the game

There are many spells that you can use to attack in Hollow Knight! Some spells include Vengeful Spirit, Desolate Dive, and Howling Wraiths. Do you have a question about a specific spell?

Your input -> Can you tell me more information about Howling Wraiths?

Inventory Prompt Howling Wraiths Blast foes with screaming SOUL. The Wraiths requires sic SOUL to be conjured. Strike enemies to gather SOUL. Spells will deplete SOUL. Howling Wraiths Abyss Shriek Howling Wraiths is a Spell in Hollow Knight. The Spell has a wide range allowing for some hits even when the target is at an upper corner in relation to the Knight. Like Vengeful Spirit, this Spell can stall the Knight in mid-air, though it does not send them in a direction opposite of the force. Modifiers Shaman Stone increases the damage dealt by Howling Wraiths by 50%. Spell Twister lowers the SOUL cost to cast Howling Wraiths to 24 SOUL. Damage Values Not all three bursts might be able to hit an enemy at once, depending on enemy movement and hitbox. Using the Dream Nail on this corpse yields the line: ". " After obtaining Howling Wraiths, the Knight's SOUL is completely refilled.

Do you have any more questions about Howling Wraiths? I can tell you about general info, location, behavior, tactics, journal entries, or trivia.

Your input -> Are you really a bot?

I am a chatbot, designed to help you with anything related to Hollow Knight!

Your input -> Goodbye

Goodbye Michael! Hope I could help you with your questions.

Appendix for Knowledge Base

The appendix can be found at:

<https://cdn.michaelzhao.xyz/archive/hollow-knight-chatbot-kb-summary.txt>

Here is a sample of the appendix:

Dreamshield:

info: Dreamshield Defensive charm once wielded by a tribe that could...
behavior_tactics: Conjures a shield that follows the bearer and attempts to...
location: Effect Conjures a shield that slowly rotates around the Knight's...
weapon_damage: Attacking with the Nail causes the shield to bash outward...
trivia: Trivia Dreamshield's pedestal The pedestal where the Charm is found...

Oro_%26_Mato:

info: Nailmasters Oro & Mato Hunter's Journal Idle Theme (Phase One)...
location: "Gods by toil and nail bound, Brothers sworn to guard...
behavior_tactics: Behaviour and Tactics Phase 1 Nailmaster Oro uses the following...
weapon_damage: The first slash is a backhand swipe, and the second...
spell_upgrade: This allows safe healing during any attack except a Dash...
trivia: 10G Trivia Despite the bad blood of Oro and Mato's...

Isma%27s_Grove:

location: Marked Unmarked Connection City of Tears Kingdom's Edge Fungal Wastes...
info: White Palace Aren't these waterways thrilling? A labyrinth of pipes...
behavior_tactics: All trash carried by the water eventually ends up in...
trivia: Inhabitants Points of Interest Inhabitants Characters Fluke Hermit The

Godseeker...

journal_entry: Durandoo Hunter's Journal entry: "Where does this water come from?..."

Mushroom_Clan:

info: Strength in the shared self. Strength in the mind united...
location: Lore Tablet in the Fungal Wastes The lands of Hallownest...
behavior_tactics: These territorial creatures possess a hive mind and developed their...
journal_entry: Shrumeling Hunter's Journal entry: "Young fungal creature. ..."

Pilflip:

info: Pilflip Hunter's Journal Flipped Health Drops 30 5 / 6...
behavior_tactics: Odd creature that adapts to changing situations by flipping over...
location: Adaptability is an important survival skill. Location They are only...
trivia: Locations in the Royal Waterways Gallery Patrolling Attacking Trivia Its...

The_Five_Great_Knights:

info: What a towering figure he was, though surprisingly soft spoken...

location: Information on the individual Great Knights is sparse, with Relic...
behavior_tactics: She dissipates after the Knight delivers the Delicate Flower to...
journal_entry: Ogrim mentions that he misses Hegemol's humour. When he is...
weapon_damage: While Hegemol was sleeping, his armoured shell was stolen by...
trivia: Once a knight, now lost to puerile obsession: The Dung...

Appendix for Sample User Model

```
{  
  user_id: "auto-generated ID for user"  
  name: "name of user that the hello intent will prompt for"  
  topic: "last topic visited"  
}
```

Example of user model:

```
{  
  user_id: "f28f299de923jf32fsavb3q"  
  name: "Michael"  
  topic: "Abyss Shriek"  
}
```

Evaluation of the Chatbot

The chatbot works decently well for simple questions. When users ask broad questions, it will answer with general information. When the user asks about a specific topic, the chatbot will give the user information about that topic. There are a couple of issues with the bot unfortunately. I wasn't able to figure out how to have it retain and use contextual information for future queries, so you have to specify the topic for each question you ask. It does remember the user's name and uses it in templates, but all those cases have to be written out in the training set.