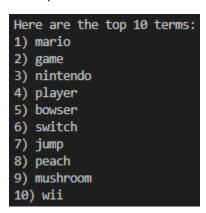# Knowledge Base from Web Scrape

Creating the Knowledge Base

The web crawler program used the starter URL https://en.wikipedia.org/wiki/Super_Mario to crawl the web for relevant links. Once it scraped all text off each web page that was relevant to *Super Mario*, it stored each page's raw text into a new file, for a total of 20 files. It then cleaned up the scraped text from each page, such as by removing redundant words and extra spaces, and stored the processed version of the text into a new file, for a total of 20 output files for the 20 input files. The output files containing the clean text for each page were used to extract 40 important terms by calculating the term frequency count of each word in the file and sorting each word based on the count of their occurrences in all the pages. From the 40 important terms extracted from all the relevant files, 10 most important terms were determined manually. The top 10 terms and all sentences, from the clean text files, containing each of the 10 terms were then used to build a searchable knowledge base of facts that a chatbot may share related to the terms.

The top 10 terms added to the knowledge base for *Super Mario*:

```
Here are the top 10 terms:
1) mario
2) game
3) nintendo
4) player
5) bowser
6) switch
7) jump
8) peach
9) mushroom
10) wii
```

The knowledge base is a Python dictionary, where each key is a term from the top 10 terms and the corresponding value for each key is a list of all sentences containing the term. For example:

```
'mario': ['the game features sandbox-like elements similar to super mario 64 and super mario sunshine, and revolves
mario encounters bowser on the latter's airship.", "the story begins in the skies above peach's castle, where mario
kidnapped tiara, cappy's sister, to use her as peach's wedding headdress.", "together, the two decide to team up to

'game': ['the game features sandbox-like elements similar to super mario 64 and super mario sunshine, and revolves around mario exploring
'unique to this game, mario is able to throw cappy onto certain enemies and objects, allowing mario to take control of them and use their
'mario can also dress in different costumes, several being based on the previous games.', 'the game supports two-player co-op, allowing a

'peach': ["the story begins in the skies above peach's castle, where mario and bowser are in the middle of fighting
band that was previously stolen from tostarena on peach.', 'mario, cappy, peach, and tiara are reunited but soon dis
cter called uncle amiibo, which unlocks more costumes, including ones based on luigi, wario, waluigi, and diddy kong

'mushroom': ["after the credits, the odyssey lands in the mushroom kingdom near peach's castle, where cappy says that
with mario to collect more power moons.", 'they do so, going through the mushroom kingdom, as well as rabbit ridge of
om no moe-eyes stone beings that resemble moai statues.', "sand kingdom lake kingdom snow kingdom bowser's kingdom yes
```

The dictionary was pickled into a file. The knowledge base may be utilized by a chatbot after unpickling the file containing the dictionary of facts.

Below is a sample dialog with a chatbot based on the knowledge base of facts for the 10 terms.

Chatbot: Welcome! Have you played a Super Mario game lately?

User: Yes, I have.

Chatbot: What is one of your favorite Super Mario games?

User: I've recently completed Super Mario Sunshine, and I enjoyed it a lot.

Chatbot: Super Mario Odyssey and Super Mario 64 have similar features to Super Mario Sunshine. Have you played those before? You might like them.

User: No, I haven't. Can you tell me more about Super Mario Odyssey?

Chatbot: Sure. In Super Mario Odyssey, Mario explores the world with his new partner, Cappy. Unique to this game, Mario is able to throw Cappy onto certain enemies and objects, allowing Mario to take control of them and use their abilities. The game also supports two-player co-op, allowing a second player to control Cappy independently from Mario. Does that seem interesting to you?

User: Great! Now I know what game my friend and I can play next.