

Webcrawler Documentation

1. Creating my Knowledge Base

To create my knowledge base I first created a web crawler that starts at the specified starting link, and scrapes the first viable 15 other pages. Following text formatting and removal of html elements, I break the tokens into unigrams and tokenize the sentences to be saved to a pickle file for later use.

I chose a rap artist I am interested in to start my knowledge base. It saves them to 15 different text files full of the text scraped. Those tokens were broken up as described above. This resulted in me getting the following keywords for my knowledge base:

```
'music', 'jpegmafia', 'album', 'baltimore', 'think', 'song',  
'make', 'feel', 'veteran', 'rap'
```

2. Sample Dialog

Initial dialog 1

Hi, I'm JPEG-BOT, a bot designed to chat about the artist JPEGMafia. I love their music, let's chat about it! What's your name?

```
It's Six, nice to meet you!
```

Hi Six, it's nice to meet you as well. Tell me, what's your favorite song of JPEGMafia's?

```
It's Thug Tears. What about yours?
```

I like that song too! My favorite is [...song...] off of [...album name...]. What's your favorite album by Peggy?

```
I like OP! Demo, what about you?
```

That's a great one!

Initial dialog 2

Hi, I'm JPEG-BOT, a bot designed to chat about the artist JPEGMafia. I love their music, let's chat about it! What's your name?

```
It's Karen, nice to meet you!
```

Hi Karen, it's nice to meet you as well. Tell me, what's your favorite song of JPEGMafia's?

I'm not sure I've heard of that artist? What are some of their songs?

You came to the right bot! I would recommend [...jpegmafia songs...], those all are super good!

Sounds good, I'll check those out!

You should! He's interesting because [...insert fact here...]

That's cool, I've got to check him out!

You bet!