Linus Fackler

Portfolio: Finding or Building a Corpus

How I created my knowledge base

I scraped texts from multiple websites and found the 40 most important words that have something to do with the topic that is talked about in the websites. I have narrowed those 40 words down to 10 (I actually combined some words, like 'Led' and 'Zeppelin', just because they go together. So technically it's more than 10 words.) I have created sample sentences using those words, sentence-tokenized these and pickled the resulting list.
Here are the top 10 terms:

- wembley
- play
- culture
- edition
- led zeppelin

- abbey road
- remastered
- brain damage
- waters
- anniversary

Here is a screenshot of this hardcoded knowledge base:

```
def buildKnowledgeBase():
    # top 10 manually picked terms:
    # wembley, play, culture, edition, led zeppelin, abbey road, remastered, brain damage, waters, anniversary

    # sample sentences:
    text = "Hi, I'm Syd, a user-created fan-bot of psychedelic/progressive rock band Pink Floyd. What's your name? Hi <name>, do you like Pink Floyd? Awesome, me too. What's your favorite song? Mine is Brain
    Damage. That's sad to hear. Have you listened to their Album 'The Dark Side of the Moon' yet? It had its 50th anniversary this year. That's a good song. If you like Pink Floyd, you should play other albums
    like Led Zeppelin II or Abbey Road. You should give it a try! You can find the remastered edition on spotify. This album has influenced culture a lot."

    sents = sent_tokenize(text)
    return sents
```

Sample dialog with a chatbot

Here is a sample dialog I would like to create with a chatbot based on my knowledge base:

Sample dialog 1:

Hi, I'm Syd, a user-created fan-bot of psychedelic/progressive rock band Pink Floyd. What's your name?

Hi Syd! My name is Linus.

Hi Linus, do you like Pink Floyd?

Yes, it is my favorite band.

Awesome, me too. What's your favorite song? Mine is Brain Damage.

My favorite Pink Floyd song is Time.

That's a good song. If you like Pink Floyd, you should play other albums like Led Zeppelin II or Abbey Road.

> Thanks for the suggestions. I will check them out.

Sample dialog 2:

Hi, I'm Syd, a user-created fan-bot of psychedelic/progressive rock band Pink Floyd. What's your name?

> Hi Syd! My name is Linus.

Hi Linus, do you like Pink Floyd?

> No.

That's sad to hear. Have you listened to their Album 'The Dark Side of the Moon' yet? It had its 50th anniversary this year.

> No, I have never listened to this before.

You should give it a try! You can find the remastered edition on spotify. This album has influenced culture a lot.