Homework 5

In this project, I define Knowledge Base as the dictionary that key is the list of bigrams to describe each page, and value is list of pages in terms of tokens extracted from the page. Based on this design, when the chat bot receives commands from users, it could parse the commands into bigrams and find page that has the most matching bigrams with bigrams from the command.

Due to the sparse of data, the first item in Knowledge Base does not have the bigrams in the top 1000 important terms:



Sample chatbot dialog

Command: “What are the most trending games for iPhones?”

* Parsed into bigrams: [What are, are the, the most, most trending, trending games, games for, for iPhones]

A sample knowledge base:

* Key = [“trending games”, “iOS App”]
* Value = [“Pokemon”, “is”, “the”, “most”, “trending”, “game”, “on”, “app”, “store”]