Finding or Building a Corpus

1.

Knowledge base was created by manually choosing 10 relevant words after looking at tf-idf values of every word in a given document. These 10 words were placed in their own list, and a list of lists was created to account for all 15 documents.

```
doc1most = ["nintendo", "price", "regular", "mario", "dlc", "hollywood", "metroid", "studios", "universal", "switch"]
doc2most = ["wii", "released", "launched", "players", "u", "super", "advance", "titles", "ds", "pokémon"]
doc3most = ["fortune", "luck", "heaven", "yamazaki", "connection", "nin", "gorges", "kanji", "ten", "plausible"]
doc4most = ["beauty", "deck", "hanafuda", "cards", "photo", "seasons", "continues", "natural", "blossoms", "traditional"]
doc5most = ["cameo", "earn", "commission", "york", "nintendo", "links", "creator", "encourages", "playable", "extensive"]
doc6most = ["hiroshi", "sekiryo", "feature", "deal", "atari", "conservative", "haskell", "yamauchi", "yokoi", "cartridges"]
doc7most = ["insider", "user", "warioware", "love", "score", "produces", "club", "device", "shape", "person"]
doc8most = ["oldest", "collection", "box", "harrah", "earliest", "items", "mistake", "bill", "closing", "marufuku"]
doc9most = ["calender", "item", "promotional", "shame", "bad", "torn", "national", "great", "kanji", "days"]
doc10most = ["kimishima", "wii", "iwata", "president", "global", "device", "health", "benefits", "politics", "yamauchi"]
doc11most = ["pokémon", "movie", "eshop", "download", "iwata", "kimishima", "furukawa", "switch", "yamauchi", "president"]
doc12most = ["princess", "sold", "bowser", "worldwide", "super", "insider", "system", "video", "million", "zelda"]
doc13most = ["yokoi", "yamauchi", "source", "portable", "firm", "salaryman", "popeye", "sharp", "gunpei", "adventure"]
doc14most = ["laughs", "asks", "development", "research", "iwata", "watch", "color", "design", "system", "block"]
doc15most = ["laughs", "watch", "display", "ball", "title", "asks", "calculator", "iwata", "idea", "research"]
```

The list of lists was pickled into a file known as 'list.p'. This can be unpickled to access the list of lists, where an entry is a list of 10 relevant words corresponding to the document of the same position. Ex: list.p[0] corresponds to output1.txt

2.

Sample dialogue with ChatBot:

Person – What are popular properties of Nintendo?
ChatBot – Pokémon, Mario, Zelda, Metroid

Person – Who are some presidents of Nintendo?
ChatBot – Yamauchi, Iwata, Kimishima, Furukawa

Person – What are some systems of Nintendo?
ChatBot – Wii, DS, Super, Switch