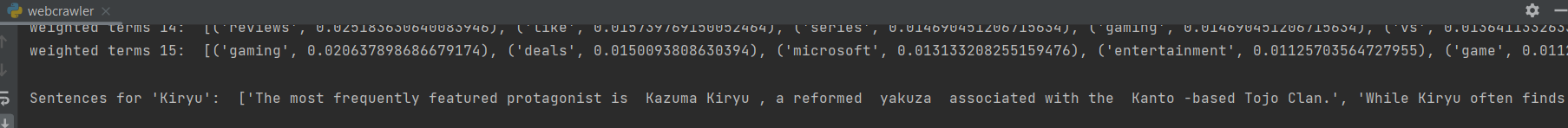
My knowledge base for the web crawler is a simple dictionary with keys that are the 10 most relevant terms for all the given URL pages. The program opens each clean text file and all the text gets sentence tokenized to make it easier to add values to each key. If a given term is found within one of the sentences in the file, then the sentence is added to the list of values for that key. It does this for each sentence in each file.

The top 10 terms that were picked include:

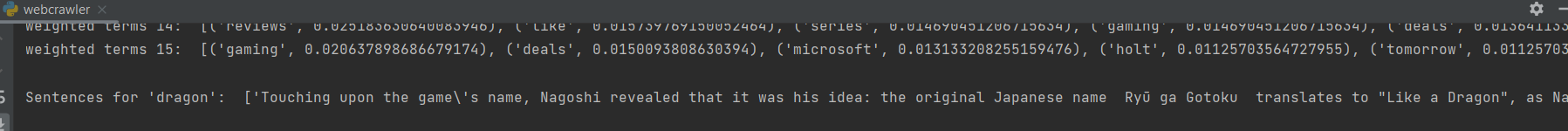
1. Yakuza
2. gaming
3. series
4. Kiryu
5. Kamurocho
6. dragon
7. action
8. Japanese
9. Nagoshi
10. Sega

These are all relevant to the Yakuza / Like a Dragon series that I picked for this assignment.

Preview of the values for key “Kiryu”:



Preview of the values for key “dragon”:



As for sample dialog for a chatbot based on my knowledge base, I’ll put it here:

Chatbot: Hi, I’m Ryu, a fan-bot for the series Yakuza / Like a Dragon. What is your name?

User: My name is Krystian.

Chatbot: Hello Krystian, are you a fan of the Yakuza / Like a Dragon series?

User: Yes, I am a pretty big fan of the series.

Chatbot: Likewise, which is your favorite game in the series?

User: I would have to say Yakuza 0, it has very good gameplay and music.

Chatbot: That’s a very good game, are you familiar with -random fact- ?