Jiaying Wei - Softdes hw3 - philsophy game reflection

Part 3: Project Writeup / Reflection

1. Project Overview: What were you trying to accomplish? What was your general approach?

    For project three I was trying to do something quirky with wikipedia, and immediately my thought went to the Wikipedia Hitler game and the Wikipedia Philosophy game. I decided to pick the Wikipedia Philsophy game because it seemed like a relatively simple implementation of data mining since pattern had a function for Wikipedia that specifically stripped out links. However as I finished my first iteration, it became very apparent that what pattern thinks are Wikipedia's first links and what I think Wikipedia's first links were, are very different. Then with the help of Chris I decided to pull from the source code instead and write code that finds the first relevant paragraph and filters out the links that I want.

2. Implementation: How does your code work? What libraries did you use? How would someone (for instance a NINJA) run your code? What data structures (e.g. lists, dictionaries) did you use in your program and why?

3. My code works by asking for user input for what article they want to start on, and then pulls the source code for that specific Wikipedia article. I then look through the source code and iterated through the <p> tags, until I found one with valid links inside. It then parses the stuff inside the <p> tags using the Beautiful Soup library, which pulled out all the <a href> tags and put it into a lsit. I then filtered the <a href> tags for things that followed the Wikipedia article link format. Finally I iterate through the list to find a link that has not been previously clicked before and then return the name of that article. The program then inputs that article and repeats until the article title returned is 'Philosophy'.

4. Include some examples of your program's output.

    Wikipedia plays philosophy for you:
    Pick a random article: New York Times
    Newspaper...
    Periodical literature...
    Published...
    Dissemination...
    Communication...
    Information...
    Data...
    Set (mathematics)...
    Mathematics...
    Quantity...

Property (philosophy)...
Modern philosophy...
Philosophy...
OMG Philosophy!
Your journey here:
['New York Times', 'Newspaper', 'Periodical literature', 'Published', 'Dissemination', 'Communication', 'Information', 'Data', 'Set (mathematics)', 'Mathematics', 'Quantity', 'Property (philosophy)', 'Modern philosophy', 'Philosophy']

5. Reflection: from a process point of view, what went well? what could you improve? Was your project appropriately scoped? Did you have a good plan for unit testing?

Well, it taught me about how to filter text and get rid of parts I don't want to check. It actually works pretty well and plays the philosophy game like a human. I suppose instead of rerouting the looped articles I should have just printed that it wasn't possible to reach philosophy from that page. I think it was appropriately scoped because I learned how to approach a problem from another angle once I reached a dead end and it made me consider additions to the problem, like how to deal with links that have previously clicked, what to do if a <p> tag didn't have a valid link, and how to get rid of parenthesis.