

Project proposal

Question 1

For my final project, what I want to accomplish is a few things. For context, currently I'm enrolled in CS 260, which is a web development course. And as such, I have to make a website as a "creative project" and for this project, I'm going to try to create a Book of Mormon parallel reader. An example of what I'm kind of meaning to do can be found [here](#), this example is with the Bible, however. I also think that I'll probably just limit this to two languages at a time, for the sake of simplicity. I think that this will help me be able to process large amounts of text files and to be able to display them in a proper format.

Question 2

The data that I'm planning on using is going to come from the Church of Jesus Christ of Latter-day Saint's website [found here](#), specifically under the section of scripture study, and for the text of the Book of Mormon. It seems that the way this content is organized is really nice in that inside the URL itself it specifies from where each chapter comes, and also which language it has, which is going to be the most helpful tool, I think.

Question 3

In order to accomplish this task, I need to be able to scrape large amounts of text from the Church's website. Looking ahead to some of the lesson plans on web-scraping, I think that I'll probably use the BeautifulSoup package. I think that I also need to be careful in scraping the data to avoid the footnotes, because those little letters will just show up as plain text, and so I'll need a way to figure out how to scrape just the plain text from the site. Maybe at some point I'll integrate the hyperlinks into the site but that's something that I don't really need to do for this class.

I think that I'll start out with trying to do six languages (Arabic, Chinese (Simplified), English, French, Russian, and Spanish). I chose these languages because these are the six official languages of the United Nations. I also think that these will be a good set of languages to start with for a few reasons. In the case of Arabic (although I know that some are not fond of the translation), this will be good practice for me to deal with right-to-left (RTL) languages, and additionally the Book of Mormon is only available in pdf form, and so I would be able to get some good practice with that. Additionally, Russian and Chinese give me the chance to deal with languages that don't use the Roman alphabet, and would force me to use understand Unicode better.

In order to test for precision and accuracy, once I've parsed the files, I'll select a number of chapters from each language (maybe 4 or 5) and see if they: (a) eliminated all of the footnotes, (b) see if the verse numbers match up with the corresponding text, and (c) just check for any abnormalities (html tags being left-over, etc).

Question 4

Programming skills that this will require are at the bare minimum I need to learn how to use BeautifulSoup. Additionally, I'll need to learn how to use some sort of python package in order to read in data from a pdf, and find my own way to parse that into verses. Once those have all been parsed into verses, I'd need to format this data into JSON, and from there I would be able to export that into a database to be kept of my web server, and then how I format it in the webpages and on the website goes beyond the scope of this class, I believe.

PS

If you would rather that this project be something that's like, totally original and not "double-dipping" with another class, that's totally fine, just let me know.

665 Words