

Project: **X**Word

Sabastian Mugazambi & Frederik Roenn Stensaeth

Repository: https://bitbucket.org/mugazambis/cs257_webapp

With this project we hope to serve a subset of the crossword solving community, as well as the demographic that use online dictionaries. The crossword solving community that we are serving is believed to be older and established in their lives; however, young to the extent that they are on the Internet. The users of online dictionaries range from school children to well established adults. We believe there is a slight overlap between the dictionary user base and the crossword solving user base.

We want to create an application that allows crossword solvers to look up words associated with a particular word in a crossword puzzle, as well as definitions of words. The application will allow them to narrow down the possible solutions and perhaps enlighten them of possible solutions that they had not thought about yet. Associated words will be categorized by noun, verb, adverb, and adjective. We envision the a useful feature of the application to crossword solvers will be the ability to look for words of certain sizes and with particular letters in particular positions. Further, if there is time we would like to explore the possibility of generating a crossword puzzle or word map available to the user to print or download. Synonyms are also quite useful to crossword puzzle solvers, thus we hope to add this feature as well, if time allows for it.

The dataset we will be using comes from the Word Associations Network (<http://wordassociations.net>). The dataset appears to be under Copy Right protection; however, previous students have used this data in the past so we see no reason why we would not also be allowed to use this data. Also, despite being under Copy Right we are not using the content of the website for commercial use, which is generally accepted. However, just to be certain we sent an email to the host of the website for confirmation – currently waiting for a reply. By means of web scraping the actual dataset will be obtained – both students in the group have experience with scraping and see no major concerns for doing the scrape due to simple and consistent HTML structure of the site. The Word Associations Network's dataset contains an extensive list of words and words that are associated with other words. The synonym dataset will be obtained from Thesaurus.com (<http://www.thesaurus.com/browse/>) using a web scraping script that we will have to write. The site uses a predictable url pattern, so the script will not need to be complex. The Thesaurus site states that the content is not to be used for commercial purposes, but this does not affect us as this project is not commercial. The Thesaurus dataset is a list of synonyms for different words.

Features (most important to least important):

- Ability to look up words associated with a given word.
- Get definition of a word.
- Look up words that fall within a specific vocabulary classification.
- Narrow the search results according to word length, letter positions, and meaning.
- Find synonyms.
- Generate a crossword puzzle or word map for download.
- Generate the solution to a generated crossword puzzle or word map for download.