Capstone 2 – Milestone Report

Dec. 3, 2017

1. Problem Definition:

The Palm Beach Code School (PBCS - www.palmbeachcodeschool.com) currently offers a sixteen week Web Developer Program. Can they offer a shorter class sufficient to prepare students for some starting web development position?

2. Client:

The Palm Beach Code School (PBCS - www.palmbeachcodeschool.com).

3. Dataset:

The data base for this project is a pandas dataframe scraped from Indeed web developer want ads in the West Palm Beach area.

The following data cleaning steps are required:

a. Read the site html into a BeautifulSoup object.
b. Remove all scripts and style elements.
c. Get the text from the bs4 object.
d. Remove blank lines and ends-of-line.
e. Remove any Unicode characters.
f. Remove all non-alphabetic characters. (This may have to be modified to handle some skills such as Angular2.)
g. Set to lower lase and split them into a list.
h. Remove stop-words.

4. Other potential data sets:

Other web sites could also be scraped.

5. Initial findings:

There are over one hundred web development related jobs within twenty-five miles of West Palm Beach. The next step will be to cluster them to determine if a subset of required skills can be identified.