Team Name: LZL
Team Member: Guanhua Li(guanhua2@illinois.edu),Ruoyu Zhu(ruoyuz3@illinois.edu),Ziqi Li(ziqili3@illinois.edu)
Team Project: Improving ExpertSearch System
Primary Language: Python
Team Leader: Ruoyu Zhu

- What is the function of the tool?

  Given a university website as the root URL, our tool is able to classify the correct faculty directory pages. Then we are able to classify the correct faculty home webpage according to the directory pages.

  Thus, our tool can further extract useful information from faculty webpages(Bios, Awards, Teaching Courses, Email...)

- Who will benefit from such a tool?

  College students that want to get detailed information about certain professors.

- Does this kind of tools already exist? If similar tools exist, how is your tool different from them? Would people care about the difference?

  Our tool will try to achieve higher accuracy and efficiency.

- What existing resources can you use?

  In order to build classifier, Python packages such as scikit-learn and Pytorch

  In order to extract useful features and information from given URL website, Python packages such as Selenium and Requests/Bs4

- What techniques/algorithms will you use to develop the tool? (It's fine if you just mention some vague idea.)

  Using Web Crawler technique to get some features (or URL features) for building and training classifiers (Decision Trees → Random Forest Classifier). Classify the correct Faculty Directory pages and Faculty Home Webpages.

Using Text Retrieval Technique learned from CS410 courses to get relevant useful text from Faculty Home Webpages.

- How will you demonstrate the usefulness of your tool?

    It can efficiently get detailed and correct faculty information.

- A very rough timeline to show when you expect to finish what. (The timeline doesn't have to be accurate.)

    By November, Finish building the basic classifier for Faculty Directory pages and the training process for it, and starting to evaluate the classifier for further improvement.

    By 11/10, Finish building the Basic classifier for Faculty Home Webpages, starting the evaluation process, at the same time, starting building proper text retrieval algorithm for extracting useful info from Faculty Homepages.

    By 11/15 Finish all the classifiers, and make sure the high accuracy of each classifier.

    By December, Finish the whole project