Project topic: **ExpertSearch: Extracting relevant information from faculty bios**
Group name: Remarkable Scientists
Members:

- Mike Pigott (mpigott2@illinois.edu)
- Shuopeng Zhou (sz46@illinois.edu)
- Amitha Supragna Sandur (asandur2@illinois.edu)

1. What is the function of the tool?
   The function of the tool is to make faculty information more accessible to users. Our aim is to improve and expand the existing ExpertSearch system.

2. Who will benefit from such a tool?

   Users of the ExpertSearch system.

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

   The ExpertSearch system already leverages Named Entity Extraction, but can be inaccurate. For example, not all professors are listed when searching for "Data Mining." In addition, sometimes the existing Named Entity extraction tool retrieves the wrong name from the data set.

   Users will care about getting more accurate and complete information.

4. What existing resources can you use?

   - SpaCy allows for building a custom Named Entity Extraction model.
   - DBPedia Spotlight scrapes Wikipedia pages and collects structured content.
   - Wikipedia provides an introduction to Named Entity Extraction that we can leverage.

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

   We expect to heavily leverage Named Entity Extraction for faculty names. The existing system uses regular expressions to find e-mail addresses; we may inspect the work it already does, or come up with our own. Finally, we will investigate using topic mining to extract topics from the bios.

6. How will you demonstrate the usefulness of your tool?

By comparing it with the existing ExpertSearch system and showing all the improvements we make.

7. A very rough timeline to show when you expect to finish what. (The timeline doesn't have to be accurate.)
   By mid-Nov:
   - Train Models for Named Entity Extraction.
   - Run the existing code of ExpertSearch website.
   - Work on implementing new features for the system.
   - Submit the progress report.

   By first week of Dec:
   - Complete training the models for Topic Mining.
   - Complete tweaking and verifying the models against the new data sets.
   - Wire in the resulting model to the existing ExpertSearch system.
   - Work on software code submission along with documentation.