Tilder

- What is the function of the tool?
  - It is supposed to aid in summarizing long text data into short, quick points *that are relevant and important*. More specifically, we are interested in trying to create a list of important topics from the text and generate a dependency list of the topics to each other.
- Who will benefit from such a tool?
  - We think this can benefit anyone who has to analyze long text data. To be more specific, students and professors often have to read long essays and try to grasp the most important information. This tool can greatly benefit them by trying to list the overall topics.
- Does this kind of tool already exist? If similar tools exist, how is your tool different from them? Would people care about the difference?
  - Yes, this tool does exist. It is different in the sense that we are trying to take leverage of topic dependencies. We are trying to structure the content in an organized fashion.
- What existing resources can you use?
  - We plan to use MeTA to experiment in our initial stages (prototype)
  - Eventually, we will move to another library or try to construct it by hand
- What techniques/algorithms will you use to develop the tool?
  - We are going to use web scraping to collect the data (selenium)
  - Our algorithm will try to extract a list of topics from the text based on how often the word is repeated (some variation of term frequency)
  - From this list, we will construct a graph (not sure on exact details)
- How will you demonstrate the usefulness of your tool?
  - We will apply the tool to Coursera and show that it can extract the necessary data and return the proper information.
- A **very rough** timeline:
  - Initial experimentation (2 weeks)
  - Web scraping (1 weeks)
  - Developing a good prototype (1 week)
  - Improving/perfecting topic extraction (1 week)
  - Building a topic dependency graph (2 weeks)
  - Testing on CS410 Coursera website (1 week)
  - Wrapping things up / putting it all together (1 week)

- Who is the coordinator of the project?
  - Christian Cygnus
- Names and emails:
  - Shashank Saxena <ssaxen4@illinois.edu>
  - Pranay Prakash <pgpraka2@illinois.edu>
  - Christian Cygnus <cygnus2@illinois.edu>