Overview

- The project deliverables are due by 11:59:59 PM on Monday, November 6, 2017.
- Project deliverables will count as 15% of your final course grade.
- (v1.2) Please email me your deliverables, one email from each team; be sure to CC all of your team members. You can email me URLs to shared documents, if you would like.

Project Specifications

For your team project, you will start by organizing your team and coming up with both an architecture and a test plan. Note that no code should be written at this stage of the project.

Overall, projects will be in support of building a search engine. There are six major components of the search engine we will build, and since each team is focusing on only one of these six components, the interfaces between each component is crucial. Using this approach, we should be able to swap in different components from different teams with minimal to no problems.

The six major search engine components are:

- **Crawling**: This component is primarily responsible for identifying and acquiring documents for the search engine.

- **Text Transformation**: This component is primarily responsible for transforming raw text from crawled documents into parsed sets of words, $n$-grams, etc.

- **Link Analysis**: This component is primarily responsible for extracting links from crawled documents and using such links to support crawling, indexing, ranking, etc.

- **Indexing**: This component is primarily responsible for constructing and maintaining one or more inverted indexes of all crawled documents.

- **Ranking**: This component is primarily responsible for calculating document scores based on a specific ranking model.

- **Querying**: This component is primarily responsible for interacting with users by transforming user queries into specific search queries, then presenting search results back to users.
Team Assignments

CRAWLING:

TEAM XX: Deslandes, Samuel Augusto
        Dunn, Chandler
        Ekhalikar, Rasika
        Tranate, Jedrick

TEAM L: Braunschweig, Erica
        Dzialo, Stephen
        Mahoney, Justin
        Wenzel, Aidan

TEAM 7: Garber, Logan
        Grogan, Joseph
        Steigerwald, John
        Wolf, Benjamin

LINK ANALYSIS:

TEAM K: Dong, Yixin
        Ferritto, Anthony
        Mogielnicki, Sarah
        Phanse, Adeet

TEAM QQ: Khedr, Sara
         Neupane, Enosh
         Yang, Xini

TEXT TRANSFORMATION:

TEAM G: Lawson, Eric
        Shvedsky, Joseph
        Tsang, Marcus
        Waclawik, Sean

TEAM Z: Cattarin, Lee
        Schmitter, Charles
        Wimer, Zachary

INDEXING:

TEAM Q: Gou, Jiashun
        Martinez, Andy
        Segreto, Anthony
        Stakoff, Ethan

TEAM RHINO: Gay, Daniel
            Lima, Anjin
            Valmas, Christopher

TEAM Y: Chen, Sensen
        George, Christopher
        Mawby, Matthew
        Pelaez, Tyler

RANKING:

TEAM THORN: Chuang, Matthew
            De Leon, Kevin
            Umlor, Eugene
            Yan, Jianfeng

TEAM X: Abney, Elijah
        Robinson, Mark
        Yale, Andrew
        Zhou, Jingfei

TEAM B: Keating-Wood, Reilly
        Leung, Alvin
        Postler, Sean

QUERYING:

TEAM BB: Conway, Casey
         Dahlin, Taylor
         Fox, Ethan
         Park, Tae

TEAM Zzz: Gilman, Matthew
          Gleason, Carter
          Greaves, Eryka
          Mottola, Brian
Architecture and Design

Document the overall architecture and design of your specific search engine component. In doing so, be sure to describe the major sub-components, language(s) you plan to use, expected inputs and outputs, etc. You will need to identify the architectural divisions through which your component will communicate with other components of the entire search engine.

Once you have a well-documented design, identify who will be responsible for implementing what (but do not write any code yet!).

Test Plans

In addition to unit testing, we will use a test-driven approach for each of our search engine components. Therefore, systematically describe all tests that your component should be able to pass. At this stage, we are only considering “black-box testing,” since we do not yet have any code written.

For each test in your test plan, describe the initial state of the system, the required input(s) and/or user action(s), and the expected behavior and expected output(s). Be as comprehensive as possible to ensure that a system that passes all of these tests is of high quality.

Quality Metrics

As part of your project, please define at least six quality metrics that you will measure over the duration of your project.

- At least two metrics must focus on search engine quality (i.e., the resulting product).
- At least two metrics must focus on your software development process.
- At least two metrics must focus on your (non-existent) code base.

Coding Standards

Even though you will not be writing code yet, you need to decide on what language(s) and corresponding coding standards you will agree on.

For each language you plan to use, document which compiler(s) or interpreter(s) you will use (including version numbers), what naming conventions you will use for variables, modules, functions, classes, methods, etc., and what comment headers or other similar documentation you will use for each major component of your implementation.

Team Presentations (Round 1)

Team presentations will occur in class on October 26, October 30, and November 2. Each team will present material from this first set of deliverables. The purpose of your presentation will be to effectively communicate your ideas and also garner feedback from your peers.