Project

Due date: end of semester, TBA on course webpage

Objective: Implement an android mobile app that incorporates data mining and/or information retrieval techniques.

Team size: 1

General Description:

- Please feel free to propose your own topic. Your proposal is subject to approval by the instructor.
- You're advised to design your project/implementation tasks in an incremental manner. Try to get a basic working system done that includes the basic functionalities. Then you can add advanced features based on your interest, motivation, capacity, and time availability.
- Feel free to discuss your project with the instructor and/or TA at the topic/ideas/design/algorithms level. It is your responsibility to find solutions at the implementation level.
- You should start working on the project immediately at least on the mobile app interface part. You should decide your project topic before the spring break.

Example Topic: implement a search engine with a mobile search interface. There are many options of open source search engines that you can choose from to deploy to provide the indexing/searching functionalities at the backend. For the corpus (collection of documents), you may use your personal digital documents (such as course work, notes, assignments etc.) or any interesting datasets you can find from the internet. For example, the following URLs provide many interesting datasets (Wikipedia, Amazon, online reviews, Twitter, Yelp etc.):

https://www.kdnuggets.com/2014/08/interesting-social-media-datasets.html+

https://snap.stanford.edu/data/

Advanced features include but are not limited to:

- search over graphs (consider relationships/edges in query processing and result ranking. Many datasets from the above sources are graph data such as social networks)

- similar results (besides each search result, add a "similar results" button. Clicking the button should return a list of similar results/documents)
- automatic tag generation (for each document, generate a set of tags that best describe the document)
- query suggestion
- document clustering

Other Topic Ideas: If you're seeking a more challenging project topic, there are many sources where you can find inspirations, such as:

https://www.yelp.com/dataset/challenge

https://www.kdd.org/kdd-cup

http://www.sloansportsconference.com/archive/competitive-advantages/

Evaluation: You will be given ~10 minutes to present/demonstrate your work either in class or in person with the instructor.

Submission: Zip your source code, executable and report in a single file and submit to TRACS. In your report (Word format, no lower or upper page limits), describe the various aspects, if applicable, of the project based on your own understanding and situation, such as objective, motivation, design, implementation, pseudo code (not source code), experimental evaluation, datasets, and user manual. Make references (publications, URLs) properly whenever necessary. Be complete yet concise.