**CS5560 Knowledge Discovery Management**

**Lab Assignment #1**

**Lab Submission before September 10th 2018 (11 AM).**

Submit a short lab report including screenshots and source code & data to GitHub. Post your GitHub URL through Lab 1 form at https://goo.gl/forms/MP2n7LQGmm2atsZo2

1. Check [Project Details](https://docs.google.com/spreadsheets/d/1erN0HKj-S-nrS-YXY01W-dunvN6Nf557sVVBVUYban8/edit#gid=1683369177) in the google sheets for Potential Data Sources.
2. Mine 10 unique publications abstracts relevant to your *“Project Topic”* from the Data Sources (Use [Abstract Download Code](https://umkc.app.box.com/file/315302184730))
3. Report Data Statistics (e.g. Year of Publication, Number of terms presented in the paper, Number of images & graphs)
4. Perform Basic NLP (Tokenization, Lemmatization) and provide the statistics
5. Perform Valid Word Filtering & Valid Medical Word Filtering and provide the statistics

Note: Statistics can be done each paper or all together or both.

**Submission Guidelines (for all students):**

1. Submit your source code and documentation to GitHub and represent the work through wiki page properly (submit your screenshots as well. The screenshot should have both the code and the output)
2. Comment your code appropriately.
3. Use the following google link to submit your assignment

(Lab Submission Link#) : https://goo.gl/forms/MP2n7LQGmm2atsZo2

***Cheating, plagiarism, disruptive behavior and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL:*** [*https://catalog.umkc.edu/special-notices/academic-honesty/*](https://catalog.umkc.edu/special-notices/academic-honesty/)