Project Proposal – Intelligent Browsing

- Team Members
  - Trevor Moyer – trevorm4@illinois.edu - Captain
  - Alex Powell – apowell7@illinois.edu
- Topic
  - We are going to do Intelligent Browsing as our overall topic theme. In specific, we are going to apply the bm25 algorithm to hackernews (https://news.ycombinator.com/front) in the form of a browser extension. This site at the moment does not have a great searching feature and requires a lot of manual browsing to go back in time or to figure out specific articles with keywords etc. By taking all the articles as our corpus, we will be able to build an efficient indexer and search over their archive of articles
- Datasets/algorithms
  - We will most likely be using the BM25 algorithm for doing the retrieval. We will be using the hackernews archives as our dataset, which will require manual crawling/scraping,
- Verifying that it works
  - To verify that it works, we will want to test it on a variety of queries, which we can easily do by using keywords from existing articles. We can compare the results to other retrieval methods as well as simply using google to see it's results. The expectation is that our extension will be able to effectively retrieve similar topics/articles to that which google returns.
- Language
  - Since this is a browser extension, the only choice is javascript as it is quite ubiquitous for browser extensions at this point.
- Time Commitment

Topic	Time
Learn javscript frameworks for scraping	5 hours
Build scraper/crawler	15 hours
Implement BM25 retrieval algorithm in Js	10 hours
Link scraper to bm25 retrieval algorithm while taking user input to create extension	10 hours