

Project Proposal

Names and NetIDs of our team:

David Newman (davidn4)

Rahul Singhal (rahuls11) (Captain)

Charlie Truong (ctruong4)

What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

Often, popular websites like news.ycombinator.com (Hacker News) are simply presented and lack even rudimentary ways to search and filter the information presented. Hacker News sorts posts with an algorithm that takes into account the age of the post and the number of upvotes it has received.

The problem with this regime is that it fails to take into account the preferences of the user. We propose to solve this problem by soliciting information from the user and incorporating it into a search and/or filtering scheme. We plan to focus on jobs' postings: "Ask HN: Who is hiring?" and "Ask HN: Who wants to be hired?" posts.

We anticipate implementing a web extension that ranks job postings to take into account the users' profession and technology stack preferences.

Briefly describe any datasets, algorithms or techniques you plan to use

Hacker News posts are available through regular SQL queries at:

<https://cloud.google.com/bigquery/public-data>

There is a realtime API available as well: <https://github.com/HackerNews/API>

We plan to explore the use of topic modeling algorithms and associated techniques like BM25, stemming, plsa, etc. to accomplish our goal of surfacing relevant information to the user.

How will you demonstrate that your approach will work as expected?

As compared to the default Hacker News ranking system, our extension will surface more relevant information to the user. This will be demonstrated by comparing the default ranking system to our own.

Which programming language do you plan to use?

Javascript and python.

Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

- Pre-proposal planning meetings: 4 hours x 3 people
- Develop ranking function: 8 hours x 3 people
- Create chrome extension: 8 x 3 people
 - Create deploy pipeline of extension to Chrome store
 - Input form from user
 - Update HackerNews page to reflect ranking result