

The SEO Game

Search engines use the *content* of the document, among other things, to calculate a relevance score for a query. This score is secret, known only to the search engine. The search engine then ranks the documents in descending order of score.

The object of this game is to create a document (a web page) that ranks highly for some search query. This query will be some made-up word we will invent in class. During class on Oct 14, Nov 4, and Nov 18, we will check the document rankings for that query in Google and Bing. You will get 1 point for every web page (your classmates' pages) that yours outranks.

Rules: 1) No more than 2 incoming links from other pages.
2) If your page is not ranked #1, change it.

Each time we check the rankings, you will calculate 2 metrics: 1) The tf-idf score and 2) the kl-divergence between the previous version and the current version. tf-idf is the "traditional" document scoring formula used in information retrieval; kl-divergence measures how much the words in your document have changed.

Tips:

1. Start early! It takes a while for search engines to index your page. If your page is not indexed when we check the rankings, you won't get any points. That would be sad.
2. Use the "Submit a URL" feature on Google and Bing to help get your page indexed.
3. You can use your UW students account to host your page. The reputation of the UW domain might boost your score.
4. <http://guides.seomoz.org/beginners-guide-to-search-engine-optimization>

Grading:

30 points – Based on effort and creativity. Write down *everything* you tried in order to rank higher, even if it doesn't work. Mention why you think it can increase the document's score if it's not obvious. Bullet points are fine.

30 points – List your tf-idf and kl-divergence for the 3 days we check rankings. You will not have a kl-divergence score for the first day, so there should be 5 scores in total. Show your work; your calculations should be self-contained.

30-ish points – The total number of web pages (other students' pages) that you outrank on Bing and Google on the 3 days we check rankings. You can get up to $3 \times 2 \times N$ points for N teams.

Due Dates:

- Oct 14 tf-idf score (don't turn in yet), page indexed by Google and Bing
 Nov 4 tf-idf, either kl-divergence or entropy (don't turn in yet)
 Nov 18 List of everything you tried; all tf-idf and kl-divergence (or entropy) including today's