Learning to Search More Deeply

**Problem:**
Google search provides unrepresentative search results in terms of race or gender, and it fails to provide satisfactory results to minority consumers.

**Goal:**
- Imitating Google Search Engine
- Quantifying Human Inputs
- Incorporating Human Inputs to Construct New Search Engine
- Experimenting New Engine and Comparing with Google Search

**Imitating Google Image Search**
We web scraped search results and used machine learning algorithms to determine the importance of each feature.

**Human Input**
We performed sentiment analysis to quantify public opinions from Twitter and used community-based crawling and seeding to collect information relevant to minority groups.

**Better Search Engine?**
We will conduct Surveys to compare Google’s search result and our search engine and gather feedback.
Model Output & Analysis

• **Important Features:**
  - Domain Links
  - URL MozRank
  - Domain MozRank
  - Keyword In Page Text
  - Keyword In Domain

• **Domain information** is the most important feature in Google’s ranking algorithm.
Product & Future Endeavors

We transformed search into discovery by reverse engineering Google’s search to provide a platform that incorporates human input from Twitter and expert opinions to construct an innovative search engine.

Future Endeavors:

- We want to incorporate more features Google uses, including image recognition.
- We want to obtain the sentiment score of the domain name in combination with keywords on Twitter to obtain public opinion as a feature to provide human input into the image ranking.
- We want to create a seeded search method geared toward specific communities that combines out web scrapped data, Twitter sentiment analysis and researched minority related sites.

Shampoo Case Study:

1. Head & Shoulders
2. Johnson’s
3. Argan Oil of Morocco
4. Pantene
5. Aussie

With Twitter Input

1. Herbal Essences
2. Aveeno
3. Aveda
4. Sexy Hair
5. Pantene