**Information Retrieval/Web Search**

**Program: Master’s Data Science**

Describe two of the most frustrating experiences you have had recently with a major search engine. Think of possible ways to solve these problems. You only need to describe conceptually the solutions.

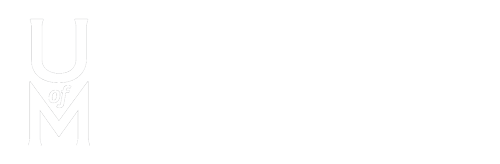
Krishnaveni Mandadi

UID:U00829899

Email:kmandadi@memphis.edu

Assignment #1

Due Date: August 30, 2021

[](https://www.memphis.edu/)

**Abstract:**

This assignment discuss two [exasperate](https://www.bing.com/search?q=define+exasperate&FORM=DCTRQY) problems that we are facing in informational retrieval with major search engines and finding out the methods to decipher them.

**Introduction**:

Web engines are faced with more difficult problems to maintain the quality of their performance due to rapid development in internet users and more information continuously comes into the network and the number of websites and web pages immediately increase.

We start with the high frequent problems that I observed in search engines are.

**Viable sources of information:**

Many people don’t understand the importance of continual link building and few weblinks are in different languages though they represent same topic as their content. It is very important that you get links from quality websites, It takes time in order to other websites to recognize your website’s content and link to it. There are link building techniques that you can use, such as doing link outreach–asking website owners to link to your website. That can be just as easy as sending an email or filling out a contact form on a website, asking for a link. However, You have to know which links to request.

**Ranking:** Search engines provide unwanted information shown on top and relevant information is in bottom or sometimes not even top 10 results and instead we get the most popular or the one with the best marketing agent. Current state of search engines, they have a very large number of indexed pages to search from. So, every search depends on the keywords defined by the user. But ranking of the page largely involves the use of context of the query which is generated by the user search history and pages search itself.

**Solutions:**

1. **Cross Language Information Retrieval**

Monolingual Search engines search in the language of user provided query. What it means that the search results can only be retrieved if the information is present in the query language. However, the research to handle those issues is also being conducted and is usually known as Cross Language Information Retrieval (CLIR).

CLIR requires the ability to represent and match information in the same representation space even if the query and the document collection are in different languages. The fundamental problem in CLIR is to match terms in different languages that describe the same or a similar meaning. The strategy of mapping between different language representations is usually machine translation. In CLIR, this translation process can be in several ways.

* Dictionary-based CLIR techniques
* Parallel corpora based CLIR techniques
* Comparable corpora based CLIR techniques
* Machine translator based CLIR techniques

1. **Learning to link images with their descriptions**

Many real-world applications don’t involve only one data modality. Web pages, for example, contain text, images, links to other pages, videos, ads, style, etc. Restricting oneself to using only one modality would involve losing all the information contained in the others. Searching across different data sources other than text: images, videos.  needs better understanding of the images, videos. May need to generate embeddings for each of them and try matching it with the user query.

**Information retrieval,** As explained before, we tackle the problem of multimodal information retrieval. Given one view (image or text), we want to retrieve the most relevant other view from a database. Mathematically, we want to build a system that computes a similarity measure between an image and a text. To do so, we use the machine learning general framework: using a dataset of aligned images and textual descriptions, we try to learn from this dataset a way to compute a similarity between those views. Many algorithms have tried to learn a similarity measure between an image and a text, and the goal here is not to provide an exhaustive bibliography on the topic. Therefore, we will present two different methods.

**A linear mapping between pre-learned embeddings**

Suppose you have learned a way to represent sentences in a “ low “ dimensional space, such as the position of a sentence in this space encodes some high-level information. This space provides a fixed-size embedding for sentences.

**Conclusion:**

I have briefly discussed couple of challenges faced in the development of the search engine and solutions to handle the problem.