

Information Retrieval. University of Virginia, Fall 2014

Homework 1. Essay: What is Missing in Current Search Engine Services. Written by Christian Kümmerle.

In my opinion, first, one has to distinguish between two fields of interest of search engines, which overlap in a certain extent, but which become more and more conflicting in those days.

Even though the representatives of contemporary companies as Google or Microsoft, which offer products with technology coming from the field of Information Retrieval, often claim that they want to change the world for the better by their products, their main *raison d'être* is still economic success, earning money and – even if this might be in the eyes of the CEOs only in the benefit of the public, since this would enable them to create the next Google Maps or Youtube – gaining power.

Thus, if an average citizen uses Google Search for finding a clip on a video sharing website, the search results will most likely favor YouTube links over those of competitors, even if the gullible user is not aware of that at all. Another danger is, in my point of view, the excessive personalization that allows powerful companies as Google or Facebook to use the combined information (meta data, search queries, GPS information and social network communication) collected about a user to show them a product which he or she is most likely to click at, but which a well-considered person would not necessarily regard as a piece of neutral "information": If a social network proved Republican is looking for online news or just scrolling through Facebook, it is probable that he will be shown news from Fox News and not those from news chains with a diverging political orientation.

A good information retrieval system should, in my opinion, provide easily accessible and reliable options of de-personalization for the user. Like that, different personal preferences are met.

A feature classical search engines are lacking but which I find really useful is the thematic clustering of retrieved websites similar to what is already implemented in the open source search engine Carrot2. One crucial point here for usefulness will probably be the underlying clustering algorithms that are ordering the documents in meaningful groups. As a mathematician, I would find it very interesting to have a closer look at how to find meaningful clusters in – abstractly spoken – a set of points in a certain metric space, for example the \mathbb{R}^d , where d could be the number of (tractable) principal components that was previously determined.

Another crucial factor in the success of such a feature is the visualization. I personally like the depiction of the different thematic groups or clusters in a ring chart, as present in Carrot2, that gives an appropriate representation for the different groups. On the other hand, this idea of a search engine can be enhanced by reconciling it with current result representation features of the market leaders Google and Bing as knowledge graph, picture or map results.