Hyun Do Jung

Hjung35

CS 410

ChenXiang Zhai

**CS410 Final Project: Real-time popular search word ranking function (Page Rank)**

**Introduction**

From the experience in my life, majority of major worldwide webpages are currently holding a real-time popular search word ranking function on their home page. This is mainly for user convenience along with a good method to bring more users into such webpages. Such ranking function helps convenient search by just clicking on the word that they were trying to find while they can see which words are popular at the current time. This helpful function can be applied to our department / academic page also, especially on registration period when majority of students use webpages very often and heavily.

**The function of the tool**

Such popular word ranking as known as popular keywords indicator will have popular keywords that were searched through Illinois homepage so that top 10 frequent searched keywords will be popped up at the side banner side. This indicator will be updated theoretically each time any keyword is being searched up, but for now educational purpose, so it can be updated after 10 keywords for example.

**Who will benefit from such a tool?**

This will be beneficial to the not only students but also faculties that may show the trend of keywords. Faculties can follow up and see the trend through appearing keywords that is shown in this indicator and see if what topics are students interested in. Students do not have to bookmark or step by step search through their hot topic through each department website, for example, DARS Web login site etc. They can simply click on the keyword and just like other search engines do, they get proper information from it without putting effort.

**Does this kind of tools already exist? If similar tools exist, how is your tool different from them? Would people care about the difference?**

I am sure there are lots of recommendation system existed in different languages and with different tool systems they are using. This would be not much different from such existing systems, but the point is to give more helpful information or help on users who would use the tool wisely without putting minor / tedious efforts for such jobs.

**What existing resources can you use?**

There are numerous researches already done on major searching websites. Using google and human resources will be far enough for me to learn for this project. I proceed to search up on google with the keyword ‘page ranking’, ‘page ranking algorithm’ and ‘keyword indicator’.

**Work Allocation**

Since I had to do this project by myself, from the beginning, everything was very tough especially on implementation and learning a new language for such implementation. I have seen that different countries use different methods on their portals even though they work like same search engine mainly.

**Algorithm behind the scene(vague concept)**

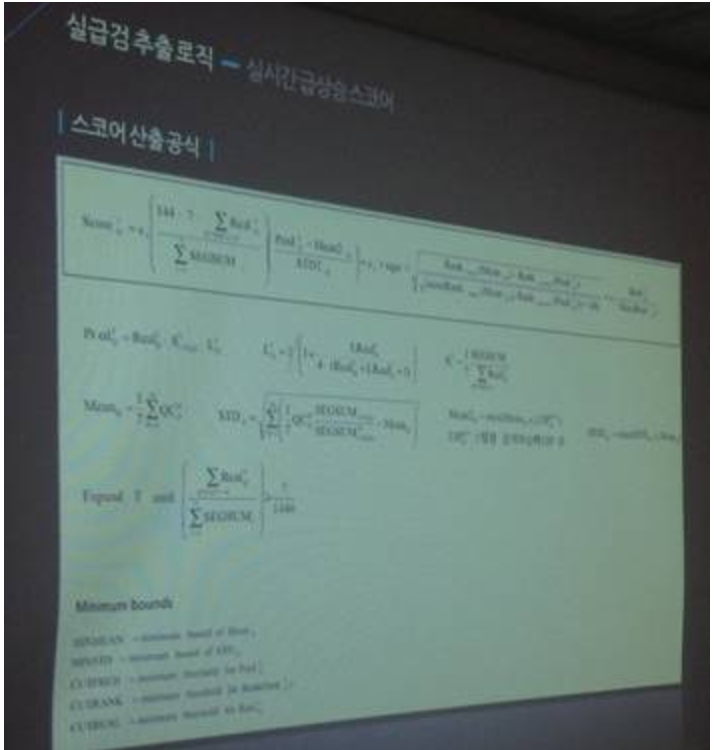
Since this is a something beyond of search or research due to its knowledge property issue and its competitive environment in the industry, there were only limited number of sources I can found. Thus, this is not an exact algorithm to implement such feature but the basic idea is close enough to explain. Every 15 seconds, it collects all searched queries from numerous servers and to make these queries useful data, use the search log we have been using. By using the such search log, we extract ‘expected scores’ which is how much this keyword has been searched at specific time. Such expected score is not fixed, but changes each time section and per date, thus changing 15 seconds. Thus, the top 10 results from calculating the difference between current search frequency on such keyword and expected score from the existing search log. The following is the figure that briefly shows the algorithm in mathematical way.

**Figure 1-1. Brief algorithm how the popular keywords indicator works**

**Figure 1-2. How to get Number of Expectation**

Summing up, the score in figure 1 is calculated per 15 seconds and the number of observation is for 10 minutes at base, but it gets longer when the search amount decreased at dawn or night. All the standard deviation, number of observation and number of expectation are set its minimum to prevent failure indication.

As far as I know, the Specific algorithm Naver’s real time popular keywords indicator was once opened to the public quarterly last few years. The following figure is blurry but a photo taken from the public conference held by Naver that shows the specific mathematical algorithm they are using for such function implementation.



**Figure 2. A photo for specific mathematical algorithm**

**What I proposed to get done**

After the research of page ranking and its algorithm along with real-time keywords ranking indicator, I wanted to have a sample coding of what I have done similar to the real-life example. This way I was going to explain better and in depth by showing the example I make. Research had done for two weeks by myself, but felt like it was not enough though.

**What I could not have done**

Learning a new language to implement such function with at least features was not easy as I expected. It required too much knowledge for such implementation so I have decided to turn this into research project so at least I can prove that I have researched through such algorithm along with the basic algorithm working from page rank and the function called popular keywords indicator like page rank.

**Conclusion**

CS 410 has been a great opportunity for me to learn Text Information System theoretically and coding wise. Implementations in M.Ps were very helpful yet, a little bit bored to do it. Instead I thought any real life example that can be seen around our lives will be much more interesting for students to get interested and learn faster and better, at least one out of total four M.Ps. In this case, I would extend my project to be designed as future CS410 M.Ps This way, students like myself would learn much better and hang tight on the interest that we have been seen on such topics.