

IR Term Project Proposal

B03705011 李羽浩 B03705031 洪念綺 B03705035 謝惠羽

Title: Demand for Labor in Industries Mining via Online Job Postings

Inspiration

It is easy to get a good view of overall demand for labor in a particular industry from a job market. Although online job postings can be a richer, more easily-accessible source of information than traditional job postings, such as those found in printed newspapers, the vast amount of information could be really messy and may not be well-organized. Thus we believe these kinds of data would be a good indicator of the demand for certain professions, job titles, or skills if they are well classified and studied.

Problem and Solution

We are trying to find the demand for labor through a specific online job postings dataset, and thereby help universities with curriculum development as well as identify skills that are most frequently required by employers, and how the distribution of necessary skills changes over time. Lastly, we hope to make recommendations for job seekers and employers on the future trend of a particular industry.

We take a bottom-up approach. First, we identify the skill keywords which are strong indicators for all job postings by analyzing job requirements. Then we separate job postings into groups using clustering. In detail, we look into job titles and job descriptions and match them with high similarity keywords. Next we list the required skills for each given job (group). Therefore the relationship between skills and jobs would be clear. At last, we observe **the need in industries of all time**, and also show **the variation of essential skills with the course of time**.

Dataset

We are using an open dataset consists of **19,000 job postings in the early 2000s** (more precisely, 2004-2015) that were posted through the **Armenian human resource portal Career Center**. The reason for choosing this dataset is that it is well-structured and contains 24 variables including detailed time information (although some fields of the posting are not necessarily filled out by the client). Moreover, it is already cleaned by removing unrelated jobs which is very helpful!