'C' What's Trending – Checkpoint 1

Nick Melnyk, Brady Pacha, Sam Stewart

Data Collected:

We have collected the first page of job postings on m.monster.com for 40 topics (~800 postings). We chose to use the mobile site because it has more consistent id tags, making it easier to get and parse pages.

Data to Collect:

Increase number of job postings per topic (some topics are so small i.e. quantum algorithms that there are only a handful of posts).
Increase number of topics (currently 40).

Problems/Issues:

Python has a very interesting platform for crawling. Some of the pages that are returned to us (sans html) are formatted poorly (unclear word boundaries, end of sentences, some words at the end of a sentence run into each other, etc.). We may be missing several languages due to this issue (still working on resolving).

Results:

The graph below shows the zipfian distribution of languages/skills. It is interesting to see the top languages/skills. We have also captured the occurences for each topic, and will be interested to determine if there are natural clusters of languages/skills within topics (i.e. game development similar to artificial intelligence in programming languages).

Occurence of Languages/Technical Skills in Job Postings