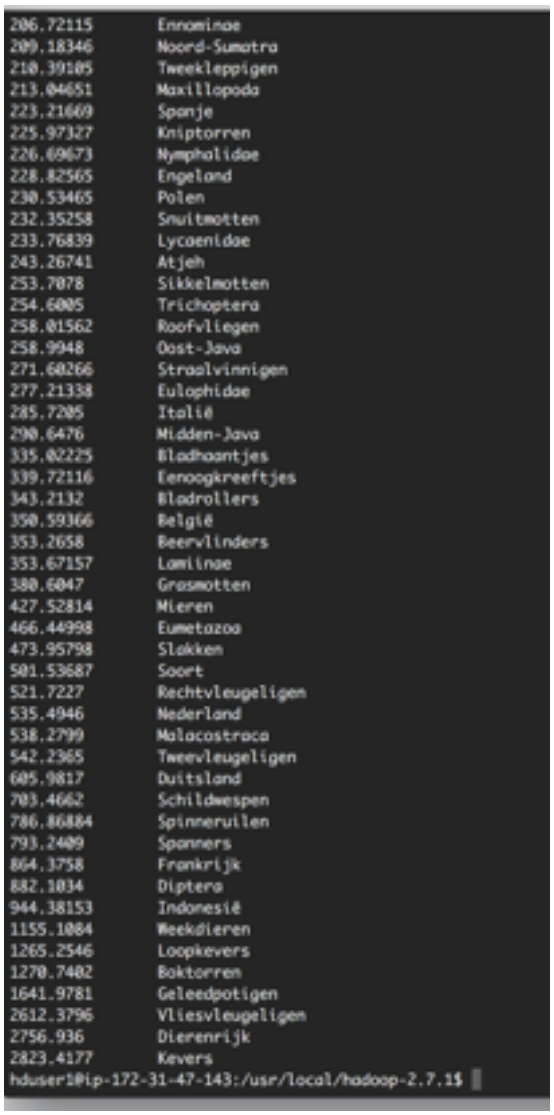
Eli Sorey
CS11 Hadoop
4/27/2016

Homework Three

Most of the top pagerank pages are on insects. At first, this is bewildering — I certainly never associated the Dutch with insects — but it actually makes sense after some thought. Pagerank measures, in a sense, the degree to which pages are linked to from other pages. Due to Linnaeus' binomial nomenclature, pages on life forms tend to link to other pages about life forms (kingdoms link to phyla, phyla to orders, and so on). This causes the broad area of wiki pages on life forms to have a very high density of links. It turns out that beetles form the largest order in the animal kingdom. This means that lots and lots of pages link to the beetles page, either from the lower levels of classification (species, genus, etc) or the upper levels (kingdom, phyla, etc). Thus, it is no surprise that the page for beetles is highly ranked. Similarly, many of the other top pages on insects refer to very large groups within the Linnean system.

Here are the top ranking pages for 2 iterations and 10 iterations, respectively:



This information can be used to determine “central hubs” on the Wikipedia network. From this, one can draw conclusions about the topic areas that have the best Wikipedia coverage, for example. Another example (contrived, as long as Wikipedia keeps its word) would be determining ad pricing based on the wiki page the ad appears on; the higher the page rank of a page, the more value its ad space is.