As a concrete deliverable of this project, besides the code itself, you must submit a report containing answers to the following questions:

1. How many unique pages did you find? Uniqueness is established by the URL, but discarding the fragment part. So, for example, *http://www.ics.uci.edu#aaa* and *http://www.ics.uci.edu#bbb* are the same URL.
2. What is the longest page in terms of number of words? (HTML markup doesn’t count as words)
3. What are the 50 most common words in the entire set of pages? (**Ignore English stop words**, which can be found, for example, [here](https://www.ranks.nl/stopwords)
4. [(Links to an external site.)](https://www.ranks.nl/stopwords)
5. ) Submit the list of common words ordered by frequency.
6. How many subdomains did you find in in the ics.uci.edu domain? Submit the list of subdomains ordered alphabetically and the number of unique pages detected in each subdomain. The content of this list should be lines containing *URL***,** *number,* for example:  
   http://vision.ics.uci.edu, 10 (not the actual number here)

**What to submit**: a zip file containing your modified crawler code and the report.