Information Retrieval Evaluation Exercise

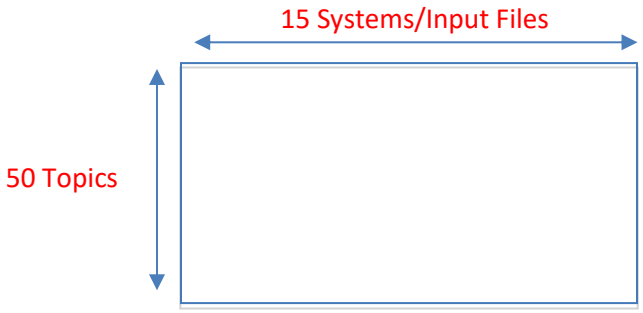
Instructions:

1. Students will be given a set of files (containing about 15 input files of system runs). A system run simulates the ranked documents of a search engine.

2. Using the TREC 8 Adhoc track ***qrels* files** (relevance judgments files) that can be downloaded from the TREC website https://pages.nist.gov/trec-browser/trec8/adhoc/data/ :
    a. Students are required to compute the system/input files evaluation scores (also known as performance/effectiveness scores) using the metrics **precision@10** for a set of systems (per system-per topic scores).
    b. Then, these per system-per topic scores should be averaged to compute the **overall precision@10 scores** for each of the given system.
    c. Finally, **both (i) correlation coefficient and (ii) significance testing** should be computed on these system scores.
    d. Correlation coefficient is to compare how similar are the system performance rankings when two different metrics are used in evaluating systems eg. precision@10 vs. precision@1
    e. Significance testing is done between pairs of systems/input files to assess if the difference is by chance or if the difference is significant.

3. Students are required to generate:

    a. Scores of Precision@10 for the given systems (for each topic and also average over all topics)
        i. E.g. matrix of scores



    b. Correlation coefficient values for system rankings
    c. Significance testing results (p-value<0.05) between pairs of systems. How many pairs of systems are significantly different.

NOTE: For computation of these scores above, a statistical package can be used called the R programming. It is open source. See http://www.r-project.org/. Or you may use any programming language/software.