Design & Engineering of Intelligent Information Systems: PI6

Instructed by *Dr. Eric Nyberg*

Due on October 19, 2015

Vivian Robison vrobison

The design pattern I used most closely matches the composite pattern. This was primarily motivated by the template structure and diagram provided. The fact that there was a "composite ranker" heavily suggested the composite pattern would be a suitable choice. The way in which the two individual rankers are created and then composed is an example of the builder pattern, as the composite ranker is designed in a way that is independent of the type of rankers it is composed of, limiting the number of constructors. There are two individual rankers, ngram and other, but also a composite ranker that combines the scores of both. The composite ranker delegates the main scoring work to the child components, and aggregates the returned scores. At the client level in PassageRankingWriter, all three rankers can be treated in much the same way because all have a score() method.

My non-ngram ranker attempts to select passages that contain the subject of the question. It uses the Stanford dependency parser to parse the question, and identifies a subject as tagged with some variant as "nsubj". The score of a passage is 1 if it contains the subject string, 0 if not. This is not designed to be used in isolation, but to theoretically boost the ngram ranker performance by increasing the relative importance of the passage containing the subject unigram. The passages are not all complete sentences, so I didn't attempt parsing them, only the questions. Below I report F1 scores of the various rankers on a set of ten questions and their associated passages:

Table 1: F1 scores

ranker	Macro-average F1	Micro-average F1
ngram	0.19047619047619052	0.171482683982684
other	0.14285714285714282	0.13874098124098125
composite	0.20238095238095238	0.18398268398268397

Since these measures are calculated using only a small set of ten questions, they aren't particularly representative of actual system performance. On some random sets of ten, the composite ranker performed identically to the ngram ranker alone, but casual investigation of performance by changing the seed didn't turn up any sets where the composite ranker performed worse. This suggests that the subject identification theory and technique are reasonably sound, and with the classification threshold set as low as five, aren't causing any harm.