Results

| Full Name | UNH id | Webcat Username |
| --- | --- | --- |
| Austin Fishbaugh | 986886495 | amf2015 |
| Bindu Kumari | 978952599 | bk1044 |
| Daniel Lamkin | 952018235 | Dtl2000 |
| Kevin Zhang |  | zkx0804 |

# 1 Implement TF-IDF

## Which variant performs best?

Of our three variants, only one of them was able to be analyzed by trec\_eval; the other two have run files, but they give errors when trying to analyze them with trec\_eval. This would suggest an issue with our implementations of those two variants, but we were unable to get the evaluations done by submit time.

The ranking system that we were able to get results for was **TF-IDF (anc.apc).** Since we do not have other measures to compare it to, I will compare it to lucenes scorer to see how it compares.

## How does it compare to Lucene?

### TF-IDF (anc.apc)

|  |  |  |  |
| --- | --- | --- | --- |
| Statistic | MAP | Precision @ R | NDCG@20 |
| Mean | 0.319403046 | .359947 | 0.445314 |
| Standard Error | 0.0203 | .0201 | 0.0231 |

### Lucene Scoring

|  |  |  |  |
| --- | --- | --- | --- |
| Statistic | MAP | Precision @ R | NDCG@20 |
| Mean | 0.577585 | 0.359947 | 0.741686 |
| Standard Error | 0.0184 | .0201 | 0.0159 |

When looking at the two results from above, it is obvious that Lucenes scoring function gives a higher mean for the MAP and [NDCG@20](mailto:NDCG@20) scores, as well as boasting a better standard error. This would imply that Lucenes scoring function is doing a better job at returning relevant documents, at least with the given test documents for this sample.

## The Standard Error Method

The standard error statistics above suggest that Lucenes scoring function has less of a percent error than our implementation of TF-IDF (anc.apc). This would suggest that, on average, Lucenes ranking function would produce results closer to what the user was actually looking for.