

Information Retrieval Index Term Weighting Models Comparison

TF-IDF, OkapiBM25 and DFI

Fatih KIYIKÇI, 170709032

Global warming

tfidf -> 10.37, 30.36, 20.21 10.4 30.49 5.34 5.44 5.11 7.27 5.30 5.17 2.4
bm25 -> 30.36 30.49 20.21 10.37 5.34 10.4 5.44 5.11 7.27 5.30 5.17 2.4
dfi -> '30.36', '10.37', '30.49', '20.21', '5.34', '5.44', '5.11', '10.4', '7.27', '5.30'
apache solr -> '10.37', '30.36', '30.49', '20.21', '7.27', '10.4', '5.34', '5.30', '5.44', '5.11'

champions league

tfidf -> '12.11', '12.8', '12.7', '8.29', '28.14', '8.17', '12.37', '28.50', '14.33', '26.5'
bm25 -> '12.7', '12.11', '28.14', '14.33', '8.17', '12.8', '28.50', '28.43', '28.25', '28.5'
dfi -> '12.7', '12.11', '28.14', '12.8', '14.33', '8.17', '28.43', '28.50', '28.25', '12.37'
apache solr -> '12.11', '12.8', '8.29', '28.14', '8.17', '26.26', '28.9', '12.7', '12.42', '12.37'

racism

tfidf -> '3.3', '4.35', '30.5', '5.46', '13.47', '13.48', '3.30', '4.9', '26.42', '4.38'
bm25 -> '3.3', '30.5', '4.35', '5.46', '13.47', '13.48', '3.30', '4.9', '26.42', '4.38'
dfi -> '3.3', '30.5', '4.35', '5.46', '13.47', '13.48', '3.30', '4.9', '26.42', '4.38'
#apache solr-> '3.3', '4.35', '30.5', '13.48', '13.47', '5.46', '26.42', '3.30', '4.38', '4.9'

alcohol and drug abuse

tfidf -> '23.20', '16.22', '4.9', '3.1', '11.50', '8.36', '8.20', '15.10', '11.23', '16.44'
bm25 -> '23.20', '16.22', '4.9', '3.1', '11.50', '8.36', '8.20', '11.23', '16.44', '15.10'
dfi -> '23.20', '16.22', '4.9', '11.50', '3.1', '8.36', '8.20', '11.23', '3.28', '15.10'
apache solr -> '23.20', '16.22', '3.1', '4.9', '11.50', '8.20', '8.36', '15.10', '11.23', '3.28'

exercising for kids

tfidf -> '23.26', '23.40', '23.29', '23.35', '23.44', '23.14', '23.28', '23.38', '23.43', '23.10'
bm25 -> '23.26', '23.40', '23.29', '23.35', '23.44', '23.28', '23.14', '23.38', '23.43', '23.10'
dfi -> '23.40', '23.26', '23.29', '23.35', '23.28', '23.44', '23.14', '23.38', '23.10', '23.43'

apache solr -> '23.35', '23.40', '23.14', '23.38', '23.26', '23.44', '23.29', '23.28', '23.10', '23.43'

obama vs clinton election

tfidf -> '4.7', '5.35', '5.24', '10.29', '1.26', '10.17', '5.42', '10.15', '10.18', '10.11'

bm25 -> '4.7', '5.35', '5.24', '1.26', '5.42', '5.48', '10.17', '10.29', '10.15', '5.27'

dfi -> '4.7', '5.24', '5.35', '1.26', '5.42', '5.48', '10.17', '10.43', '5.27', '10.18'

apache solr -> '4.7', '5.35', '5.24', '5.42', '1.26', '10.43', '5.48', '5.43', '5.50', '10.11'

health insurance

tfidf -> '24.31', '24.32', '25.41', '24.29', '25.9', '25.16', '25.27', '25.48', '25.7', '25.17'

bm25 -> '24.32', '24.31', '25.41', '25.9', '24.29', '25.48', '25.8', '25.39', '25.16', '25.27'

dfi -> '24.32', '24.31', '24.29', '25.41', '25.9', '25.8', '25.48', '25.27', '25.7', '24.26'

apache solr -> '24.31', '24.32', '25.17', '25.27', '25.30', '25.7', '25.24', '25.32', '25.33', '25.48'

breast cancer risk for women

tfidf -> '24.48', '24.49', '24.50', '24.47', '24.23', '24.6', '24.7', '9.37', '24.2', '11.35'

bm25 -> '24.48', '24.49', '24.50', '24.23', '24.47', '11.35', '24.7', '24.6', '9.37', '24.3'

dfi -> '24.48', '24.49', '24.50', '24.23', '24.7', '9.37', '24.47', '11.35', '24.2', '24.3'

apache solr -> '24.48', '24.49', '24.50', '24.23', '9.37', '24.47', '16.48', '9.1', '24.7', '24.2'

cancer risk for smokers

tfidf -> '24.7', '27.35', '24.34', '24.3', '24.48', '24.49', '24.50', '24.2', '27.19', '9.50'

bm25 -> '24.3', '24.7', '27.35', '24.34', '29.40', '24.39', '11.35', '24.6', '24.48', '24.49'

dfi -> '24.3', '24.7', '27.35', '29.40', '24.34', '24.39', '11.35', '24.48', '24.49', '24.50'

apache solr -> '24.3', '24.7', '27.35', '29.40', '24.34', '24.39', '16.14', '24.48', '24.49', '24.50'

alzheimer risk

tfidf -> '27.28', '11.47', '9.30', '11.29', '11.50', '27.47', '9.27', '11.17', '11.25', '9.33'

bm25 -> '27.28', '9.30', '11.47', '11.29', '11.17', '9.27', '11.50', '27.47', '9.33', '11.25'

dfi -> '27.28', '11.47', '9.30', '11.29', '11.17', '11.50', '9.27', '27.47', '9.33', '11.25'

apache solr -> '27.28', '9.30', '11.47', '11.29', '11.50', '9.16', '9.27', '11.17', '27.47', '11.25'

Mean Average Precision scores for models:

tfidf: 0.7776666666666666

okapiBM25: 0.7712380952380953

DFI: 0.7813333333333334

DFI seems to have a little edge over the other 2 models.