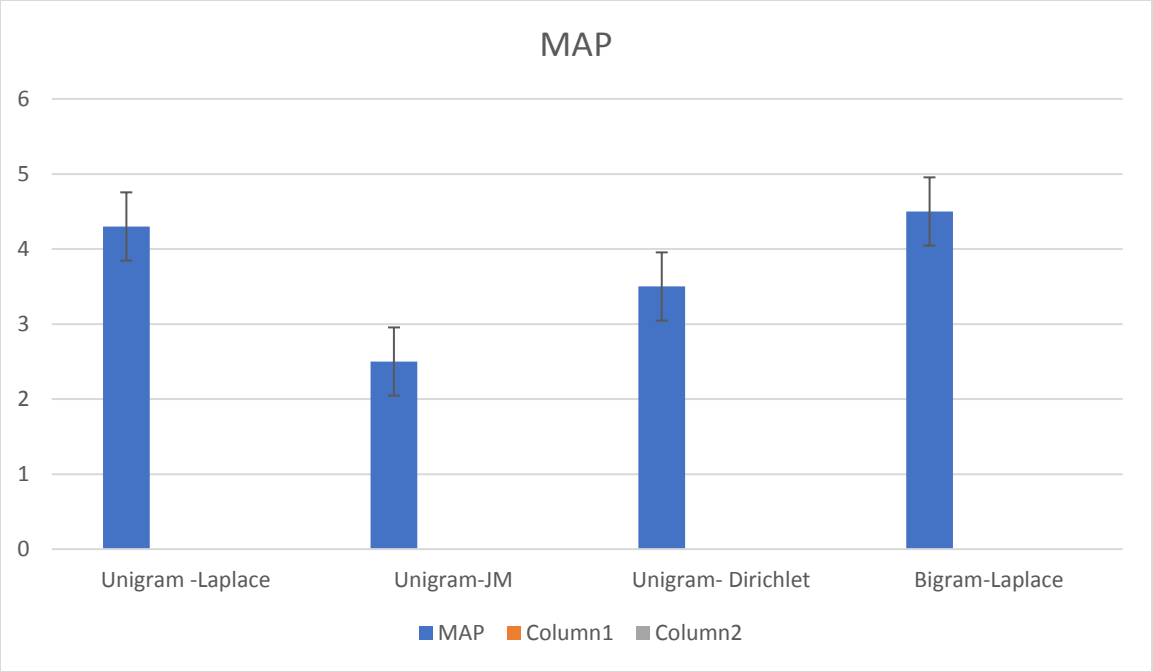
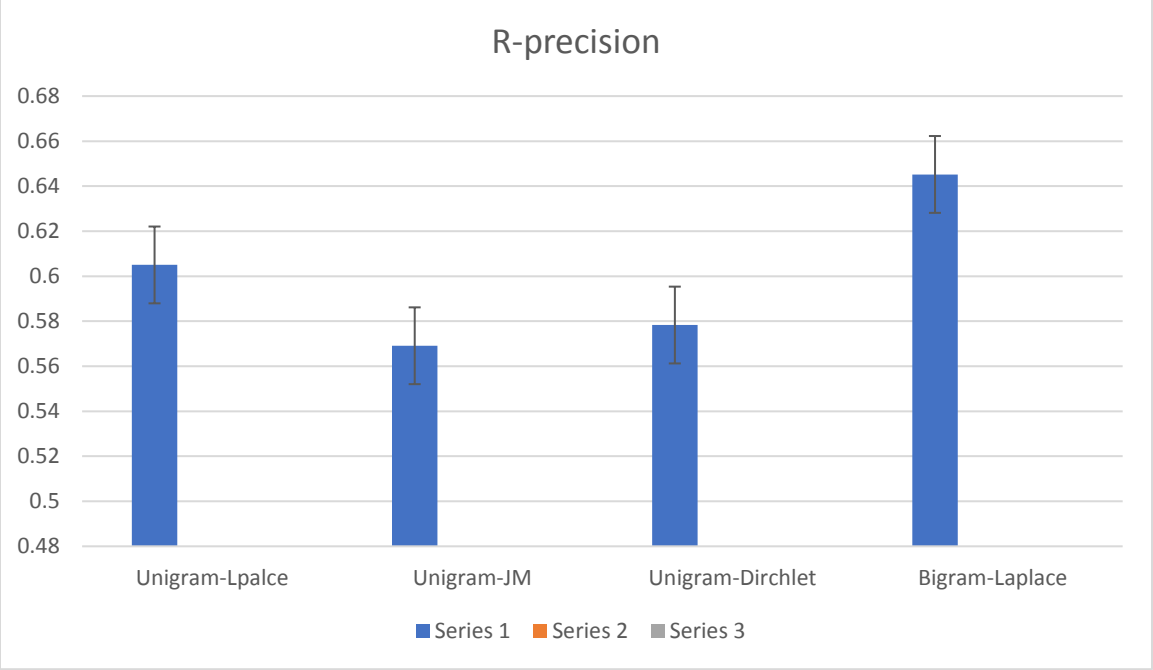
CS 753/853: Topics / Information Retrieval Fall 2019
Programming Assignment 4: Language Models
Results

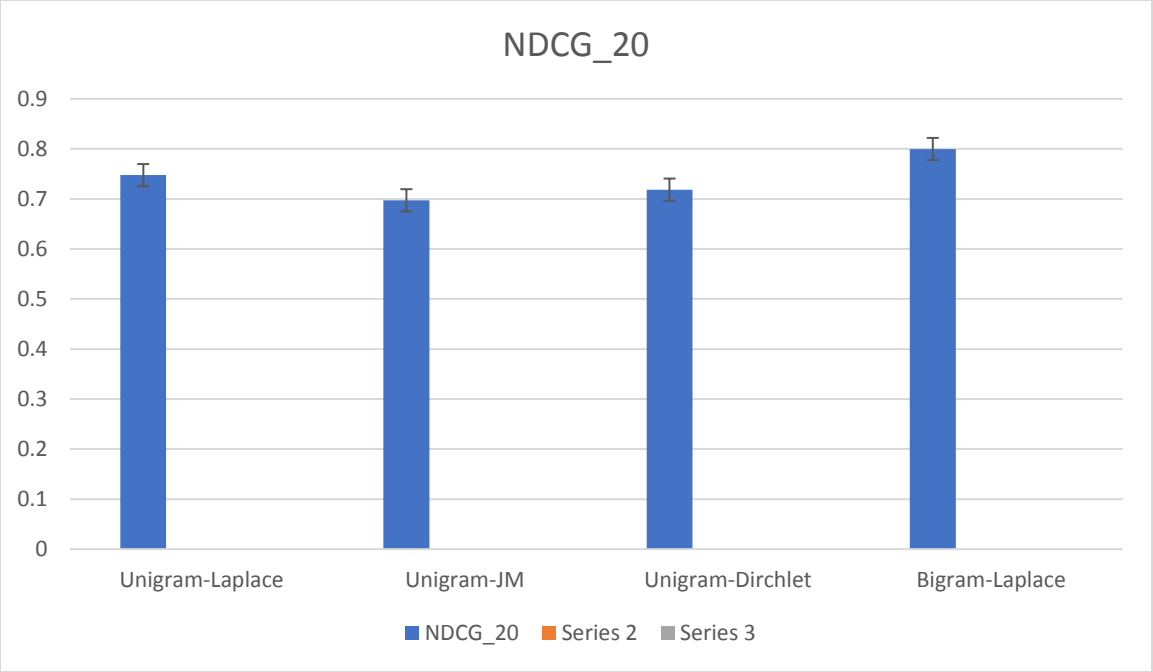
Team Members:
Medhini Shankar Narayan
Akhila Bezawada
Sai ArvindReddy DesiReddy

1.Implement Unigram Language Models with Smoothing
Implement your own unigram Query Likelihood ranking model (using unigrams) for Lucene by
extending the appropriate similarity class and setting it as similarity for the searcher.
In particular implement the following variants:

	MAP	**R-precision**	**NDCG_20**
Unigram-Laplace	0.5965 +- 0.031	0.6050 +- 0.022	0.7476 +- 0.004
Unigram – JM	0.5528 +- 0.031	0.5691 +- 0.022	0.6973 +- 0.004
Unigram-Dirichlet	0.5636 +- 0.031	0.5783 +- 0.022	0.7186 +- 0.004
Bigram-Laplace	0.6470 +- 0.031	0.6452 +- 0.022	0.7998 +- 0.004


MAP
6
5
4
3
2
1
0
Unigram -Laplace
Unigram-JM
Unigram- Dirichlet
Bigram-Laplace
MAP
Column1
Column2


R-precision
0.68
0.66
0.64
0.62
0.6
0.58
0.56
0.54
0.52
0.5
0.48
Unigram-Lpalce
Unigram-JM
Unigram-Dirchlet
Bigram-Laplace
Series 1
Series 2
Series 3



1. We have used English- analyser
2. Unigram laplace smoothing performs better when compared to others.
3. Language models berform better when compared to tf-idf . (Unigram – JM)
4. Search quality has increased compared to Standard analyser when we use English analyser and we have got better results for Evaluation measures compared to previous. The compleity has been reduced since we we removed stop words.
5. On using Standard method, we see standard errors are low and even evaluation measures are slightly less.
6. Among the precision and recall, precision serves to be a better evaluation measure