Information Retrieval - Assignment II - Retrieval System
Fall Semester 2015

fgmehlin@student.ethz.ch
malvaro@student.ethz.ch

December 2, 2015

1 Models

For this project we implemented the two following models :

- Term-based : TF-IDF

- Language Model : Maximum Likelihood + Smoothing

2 Metrics

We obtained the following metrics for the training queries :

	TF-IDF	MLE + Smoothing
Average Precision	0.09725	0.15300
Average Recall	0.04901	0.06403
Average F1	0.05818	0.07776
MAP	0.03466	0.06769

3 Remarks

- We bounded the AP denominator as MIN((TP+FN), 100).

- We did not use stemming as it did not improve the score

- Computation takes between 21 and 25 minutes

- We unzipped all the subzips, namely zips-1.zip to zips-7.zip in a folder called zips

- The main method is located in the file "RetrievalSystem.scala" (the qrels/topics/zips paths might need to be adapted)