Programming Assignment 5 Report

Name: Kunpeng Xie, Xin Liu

UNH ID: 949610377, 910991617

Webcat Name: kx1005, xl1044

1.

1. Feature data example:

   0 qid:1 1:1 2:0 3:1 4:1 # 1 qid:1 1:.5 2:1 3:.5 4:.5 # 1 qid:1 1:.33 2:0 3:0 4:0 #

2. MAP: 1.0

2.

We didn't find where to download the v2.1 test200 dataset and we tried to use **train.pages.cbor-outlines.cbor** of v2.0 test200 to replace **train.test200.cbor.outlines** but we couldn't get what we expect. So we finally decided to use the v1.5 dataset to get the results.

For the query **Brush%20rabbit**, the contents of the top-ranked paragraph is in **output/paraContent.txt**.

3.

1. The default value should be negative infinity (-INF) because all other scores are negative and our default score should be less than that.
2. Assume Jelinek-Mercer has two feature value: P(Q|d) and P(Q|corpus). When we use learning to rank to train it, it will give us two values. Lamda1 and lamda2. We can use lamda value to substitute value in our equation to see whether lamda value can get good result. And based on it we can optimize our result.