TTDS_cw2 report

This report consists of two projects. The first project is to calculating the following measures of six different IR systems: **P@10**, **R@50**, **r-precision**, **MAP**, **nDCG@10**, **nDCG10**, **nDCG20**. The second project is to build a text classification model for classifying text.

IR EVALUATION

The following table is the score of each system on measures.

	p@10	R@50	r-Precision	AP	nDCG@10	nDCG@20
s1	0.39	0.834	0.401	0.4	0.363	0.485
s2	0.22	0.867	0.253	0.3	0.2	0.246
s3	0.41	0.767	0.449	0.451	0.42	0.511
s4	0.08	0.189	0.049	0.075	0.069	0.076
s5	0.41	0.767	0.358	0.364	0.333	0.424
s6	0.41	0.767	0.449	0.445	0.4	0.49

The following table is about the best system according to each measure.

	P@10	R@50	r-Precision	AP	nDCG@10	nDCG@20
Best	s3, s5, s6	s2	s3, s6	s3	s3	s3
Value	0.41	0.867	0.449	0.451	0.420	0.511
p-value	0.888	0.7026	1	0.967	0.883	0.868

First of all, we rank the system according to the score under the measures. We think the first place is better than the second place. We use p-value for validation. Assuming the No.1 system is better than the No.2 system in table 1. To test this assumption, we calculate the p-value of the No.1 system and the No. 2 system, displaying the results in table 2. According to table 2, we can see that the p-values are all greater than 0.05, so that means our hypothesis is correct.