Vocabulary Elimination Analysis

After examining the w(q, t) numbers that I typically get for this corpus, I experimented with a range of threshold from 0.1 to 1.4. Based on the experimental results (vocab_elimination_experiments.csv where the first set of threshold is for default rank strategy and the rest of them for OKAPI rank strategy), I have selected the threshold value of 1.1 and 0.4 for default and okapi respectively. Following summarized the results and provide justification for selecting the threshold:

	Exact	Vocabulary Elimination
Default Strategy (Threshold=1.1)		
MAP	0.378	0.392
Mean Response Time (ns)	0.262	0.158
Throughput	3.802	6.322
Okapi Strategy (Threshold=0.4)		
MAP	0.389	0.390
Mean Response Time (ns)	1.232	0.307
Throughput	0.811	3.255