西安电子科技大学
XIDIAN UNIVERSITY

Cash实验汇报

——汇报人：张中俊

实验策略

- to evaluate the performance of our scheme, we implement Cash's OXT scheme

- The data set which we use in our experiment is derived from Wikipedia, each web page is treated as a separate document

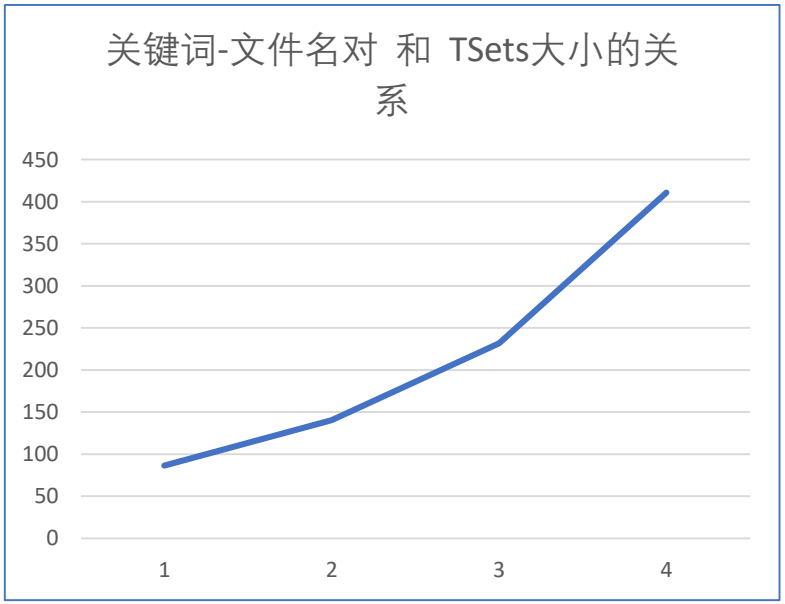
- use redis to storage keyword-filename pairs
- use MySql to storage TSets
- use Bloom filter to storage XSets（参考了Cash的做法）

- Intel Xeon CPU E5-1603 2.8GHz x4
- 16GB RAM
- Ubuntu 14.04 LTS

索引大小

- we use Bloom filter to store XSet. Since the size of a Bloom Filter is determined by the max number of elements which will be insert to it, we only evaluate the size of TSet.
- the size of Bloom Filter with 10,000,000 elements is 11.758 MB

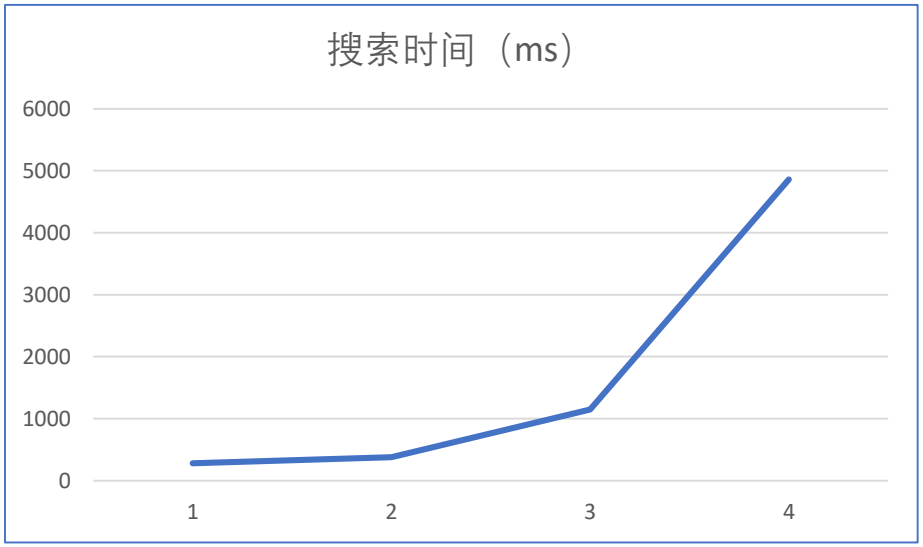
number of keywords	size of TSets(MB)
2^{16}	86.58
2^{17}	140.63
2^{18}	231.70
2^{19}	410.84
2^{20}	
2^{21}	



TSet实例化

- randomly choose 3 keywords send to server to retrieve in encrypted database.
- Note that the time below do not contain token generation time

number of keywords	searching time(ms)
2^{16}	279.45
2^{17}	378.41
2^{18}	1148.46
2^{19}	4859.77
2^{20}	
2^{21}	



- 备注：因为实验的工作站正在建立索引，所以这里的数据是我用我的笔记本仿真的，但是趋势应该是一致的

思考

- 搜索时间其实和w1包含的文件个数有关，选取查询的关键词时候是不是应该注意
- Cash的实验中是通过控制结果集中文件名的个数来区分的，见下图：

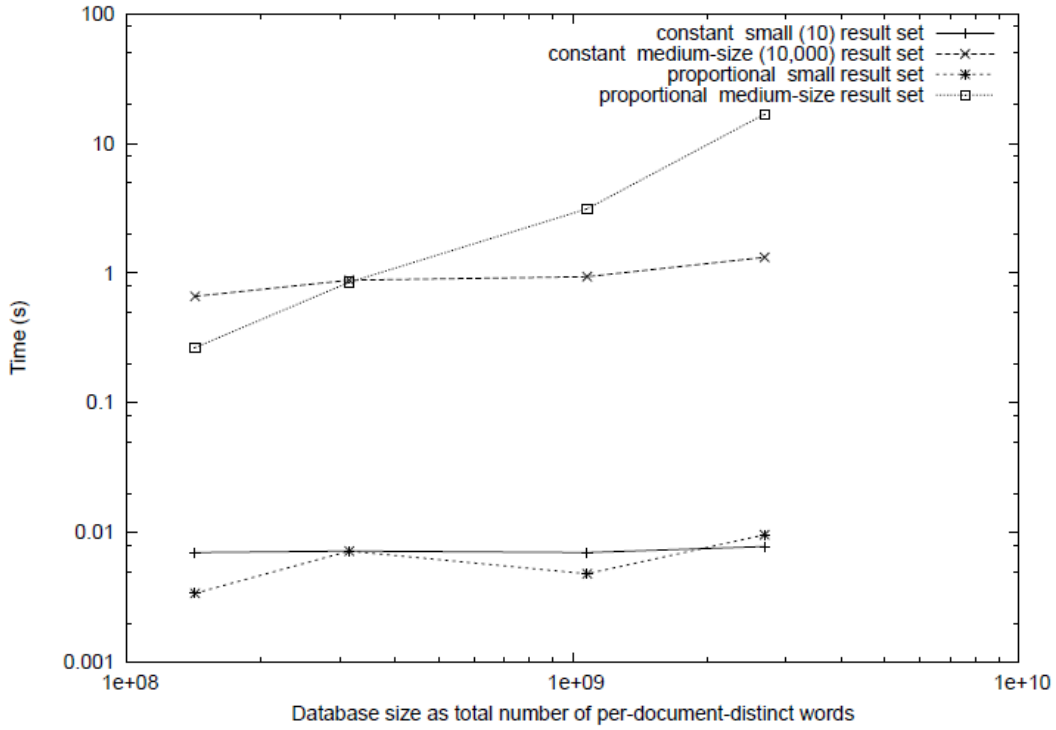


Figure 6: Clueweb09 Performance Measurement: Scaling Database Size