

Shaun McThomas 13828643
Sean Letzer 24073320
Sean King 82425468
2/22/16

Search Engine - Milestone 1

1. There were 148311 documents used to create the index.
2. There were 526473 words.
3. Sample Index:

termToTermIDMap (map from term to term id) :

{accessible=6, hallvisit=13, welcomefacts=10, standards=5, site=1,
figuresdonald=11, diversity=15, schooldeans=9, web=4, schoolequity=14,
browser=2, bren=12, contact=16, supports=3, departments=17, device=8,
internet=7}

termIDToTermMap (map from term id to term):

{1=site, 2=browser, 3=supports, 4=web, 5=standards, 6=accessible, 7=internet,
8=device, 9=schooldeans, 10=welcomefacts, 11=figuresdonald, 12=bren,
13=hallvisit, 14=schoolequity, 15=diversity, 16=contact, 17=departments}

docIDToTermIDToTermCountMap (map for document id to a map from term id to term count in that document):

{1={1=1, 2=2, 3=1, 4=1, 5=1, 6=1, 7=1, 8=1, 9=1}, 2={16=1, 17=1, 9=1, 10=1,
11=1, 12=2, 13=1, 14=1, 15=1}, 3={1=1, 2=2, 3=1, 4=1, 5=1, 6=1, 7=1, 8=1, 9=1},
4={1=1, 2=2, 3=1, 4=1, 5=1, 6=1, 7=1, 8=1, 9=1}, 5={16=1, 17=1, 9=1, 10=1,
11=1, 12=2, 13=1, 14=1, 15=1}}

termIDToDocIDToTermCountMap (map from document id to term count in that document):

{1={1=1, 3=1, 4=1}, 2={1=2, 3=2, 4=2}, 3={1=1, 3=1, 4=1}, 4={1=1, 3=1, 4=1},
5={1=1, 3=1, 4=1}, 6={1=1, 3=1, 4=1}, 7={1=1, 3=1, 4=1}, 8={1=1, 3=1, 4=1},
9={1=1, 2=1, 3=1, 4=1, 5=1}, 10={2=1, 5=1}, 11={2=1, 5=1}, 12={2=2, 5=2},
13={2=1, 5=1}, 14={2=1, 5=1}, 15={2=1, 5=1}, 16={2=1, 5=1}, 17={2=1, 5=1}}

termIDToDocIDToTFNormilizedMap (map for term id to a map from document id to term Frequency Normalized to add to ~1 in that document) :

{1={1=0.1, 3=0.1, 4=0.1}, 2={1=0.2, 3=0.2, 4=0.2}, 3={1=0.1, 3=0.1, 4=0.1},
4={1=0.1, 3=0.1, 4=0.1}, 5={1=0.1, 3=0.1, 4=0.1}, 6={1=0.1, 3=0.1, 4=0.1},

7={1=0.1, 3=0.1, 4=0.1}, 8={1=0.1, 3=0.1, 4=0.1}, 9={1=0.1, 2=0.1, 3=0.1, 4=0.1, 5=0.1}, 10={2=0.1, 5=0.1}, 11={2=0.1, 5=0.1}, 12={2=0.2, 5=0.2}, 13={2=0.1, 5=0.1}, 14={2=0.1, 5=0.1}, 15={2=0.1, 5=0.1}, 16={2=0.1, 5=0.1}, 17={2=0.1, 5=0.1}}

termIDToDocIDToWTFMap (map from document Id to a logarithmic Weighted term Frequency in that document):

{1={1=1.0, 3=1.0, 4=1.0}, 2={1=1.30103, 3=1.30103, 4=1.30103}, 3={1=1.0, 3=1.0, 4=1.0}, 4={1=1.0, 3=1.0, 4=1.0}, 5={1=1.0, 3=1.0, 4=1.0}, 6={1=1.0, 3=1.0, 4=1.0}, 7={1=1.0, 3=1.0, 4=1.0}, 8={1=1.0, 3=1.0, 4=1.0}, 9={1=1.0, 2=1.0, 3=1.0, 4=1.0, 5=1.0}, 10={2=1.0, 5=1.0}, 11={2=1.0, 5=1.0}, 12={2=1.30103, 5=1.30103}, 13={2=1.0, 5=1.0}, 14={2=1.0, 5=1.0}, 15={2=1.0, 5=1.0}, 16={2=1.0, 5=1.0}, 17={2=1.0, 5=1.0}}

termIDToIdfMap(map from term id to it's inverse document frequency):

{1=0.22184874, 2=0.22184874, 3=0.22184874, 4=0.22184874, 5=0.22184874, 6=0.22184874, 7=0.22184874, 8=0.22184874, 9=0.0, 10=0.39794, 11=0.39794, 12=0.39794, 13=0.39794, 14=0.39794, 15=0.39794, 16=0.39794, 17=0.39794}

4. The index file is 1213107.0 kilobytes.

5.The index file is created in 0:8:20 hours:minutes:seconds and written to disk in in 0:17:28 hours:minutes:seconds total time is 0:25:49 hours:minutes:seconds