Work Progress
Demo: KNN Search with Incremental Query Answering

Jaouhara Chanchaf

September 29th, 2022

Summary

Action Items for Jaouhara
Done:

AI 3	Add precision to recall/query-time plot.
AI 3	Find out why Kashif crashes when searching over large datasets.

In progress:

AI 1	Write a pseudo code for Parallel Incremental Query Answering.
AI 4	Run kashif on larger datasets. Cannot run process as root. Error: "Sorry, user jaouhara.chanchaf is not allowed to execute 'bin/dstree ...' on server server150012".

Action Items for Karima

AI 1	Change Hercules code to support incremental query answering.

Querytime, recall and precision evaluation

Experiment on 100k tables, 494k columns, 5M vectors.
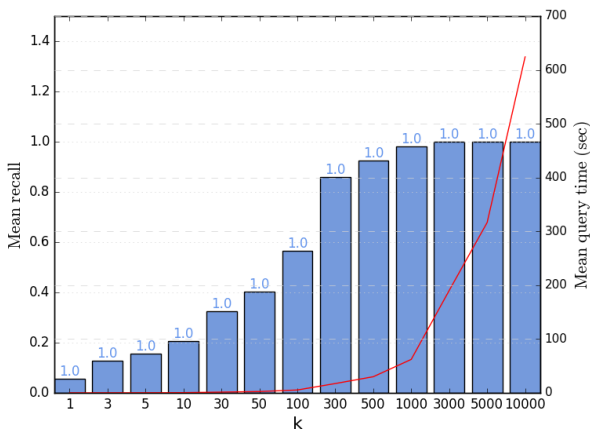10 Queries of size [50 - 100].

Figure: Kashif Average Recall

WDC Web Table Corpus 2015 - Statistics about English-language Relational Subset

Statistics:

#tables	50,820,165
#columns	235,087,091
datasize	69 GB

Column type distribution:

String	46.9%
Numeric	51.8%
Date	1.2%
Others	0.1%