Cost-Aware Strategies for Query Result Caching in Web Search Engines

R. Ozcan, I. S. Altingovde, and Ö. Ulusoy

Presented By: Rushikesh Padia

Overview

1. Query processing costs may significantly vary among different queries
2. Processing cost of a query is not proportional to its popularity (i.e., frequency in the previous logs).

Implies-

1. Cache misses have different, that is, nonuniform, costs in this context.
2. Typical caching policies, solely based on query popularity, can not always minimize the total cost.

Cost function

C_total = C_cpu + C_disk + C_net + C_snip

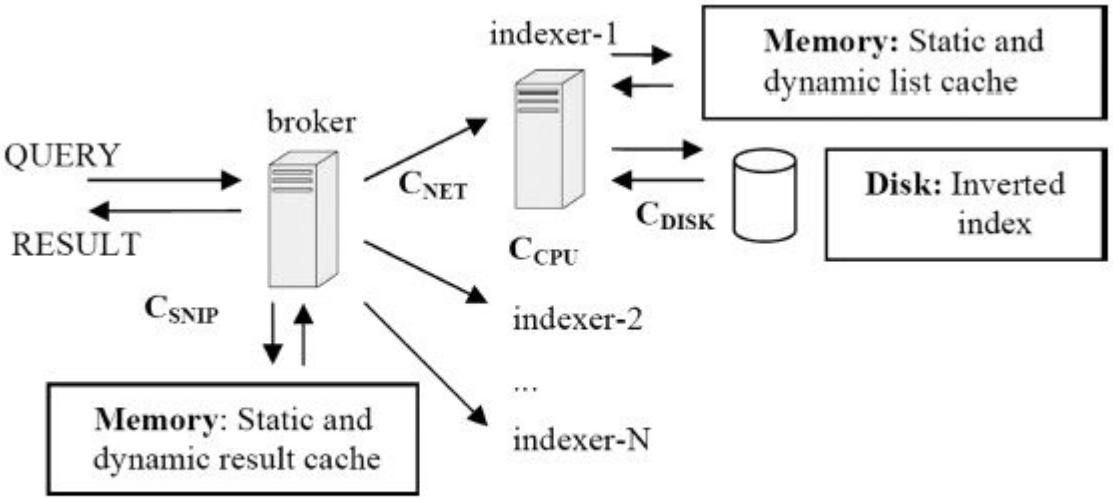
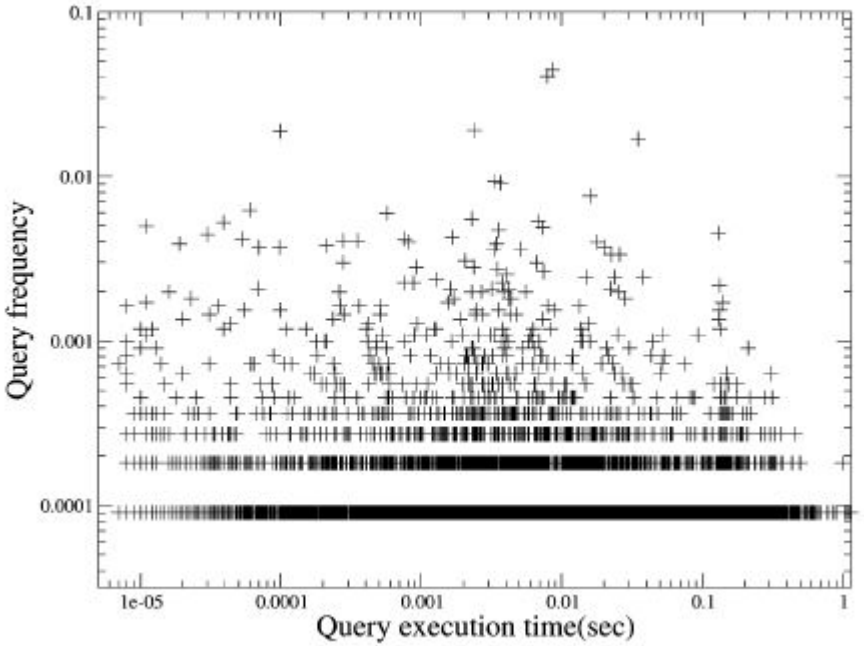


Fig. 1. Query processing in a typical large-scale search engine.

Correlation between frequency and execution time

Low correlation between frequency

and execution time



Cost aware strategies for static cache

1. Most Frequent (MostFreq): Value(q) = Fq.
2. Frequency Then Cost (FreqThenCost): query frequencies in a log follow a power-law distribution. Query as a pair (Fq,Cq). First frequency is used, in case of match, cost is used.
3. Stability Then Cost (StabThenCost): in order to be cached, queries should be frequent and remain frequent over a certain time period.
   a. query with the pair (QFSq,Cq).
   b. QFS is calculated as $QFS_q = \sum_{i=1}^{n} \frac{|f_i - f_\mu|}{f_\mu}$,
   c. Time is divided in ‘n’ intervals, fi….fn is freq in each intervals. fu is mean freq
4. Frequency and Cost (FC_K): Value(q) = Cq*Fq^K , where K> 1. Based on observation higher freq query tends to appear high and lower freq tends to get even lower. Hence emphasis is given to freq term.

Cost aware strategies for dynamic cache

1. LRU, LFU
2. Least Costly Used (LCU) - Least cost cache is becomes the victim
3. Least Frequently and Costly Used (LFCU_K): Same formula of FC_K.
4. Greedy Dual Size (GDS): H_value(q): Cq/Sq + L. Sq is size of page (can be ignored). L is aging factor, 0 at the begin.
    a. Chooses cache with smallest H_value
    b. L is set to evicted item's H_value
    c. When query result is requested again h_val is recalculated (L value might have changed)
5. Greedy Dual Size Frequency (GDSF K): Slightly modified version of GDS. Fq^K * Cq/Sq + L. Frequency is also updated

Cost aware strategies for Hybrid cache

Table III. Hybrid Cache Configurations

Hybrid Cache Configuration	Static Cache Strategy	Dynamic Cache Strategy
Non cost-aware	MostFreq	LRU
Only static cache is cost-aware	FC_K	LRU
Both static and dynamic caches are cost-aware	FC_K	LFCU_K
	FC_K	GDS
	FC_K	GDSF_K

Reference

[1] R. Ozcan, I. S. Altingovde, and Ö. Ulusoy, “Cost-Aware Strategies for Query Result Caching in Web Search Engines,” *ACM Trans. Web*, vol. 5, no. 2, May 2011, doi: 10.1145/1961659.1961663.