

Exploiting temporal changes in query submission
behavior for
improving the search engine result cache
performance

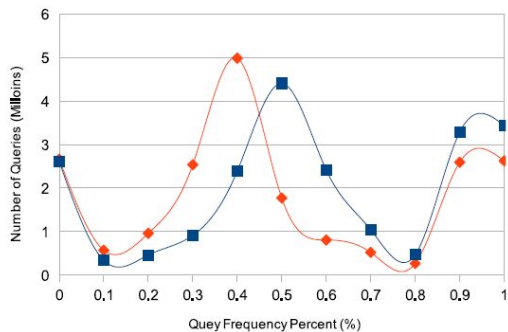
Tayfun Kucukyilmaz

Presented By
Rushikesh Padia

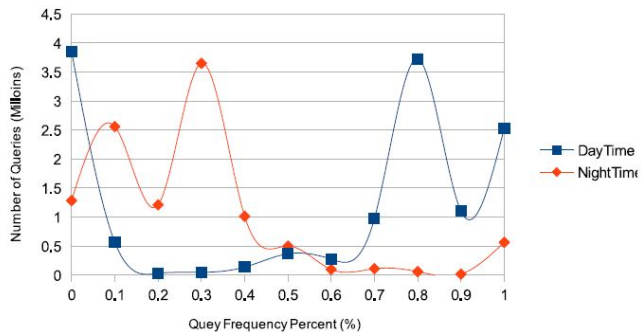
Takeaways

- Improves static part of SD cache (Topical cache - dynamic part).
- 3.3% improvement in SD cache (Topical cache - 3%).
- Key observation - Requests exhibit temporal variations through a day for many queries.
- Large cache sizes, roughly 65% of all distinct query submissions have large frequency fluctuations during day time and night time.

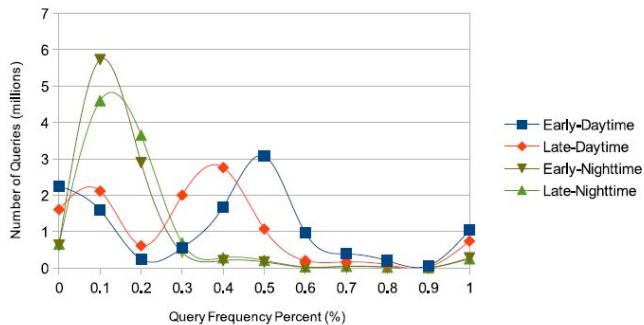
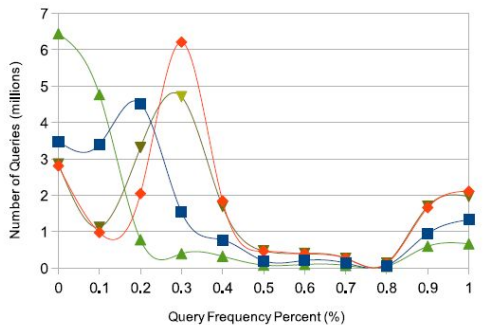
Observations



(a) Daytime and Nighttime (AOL)

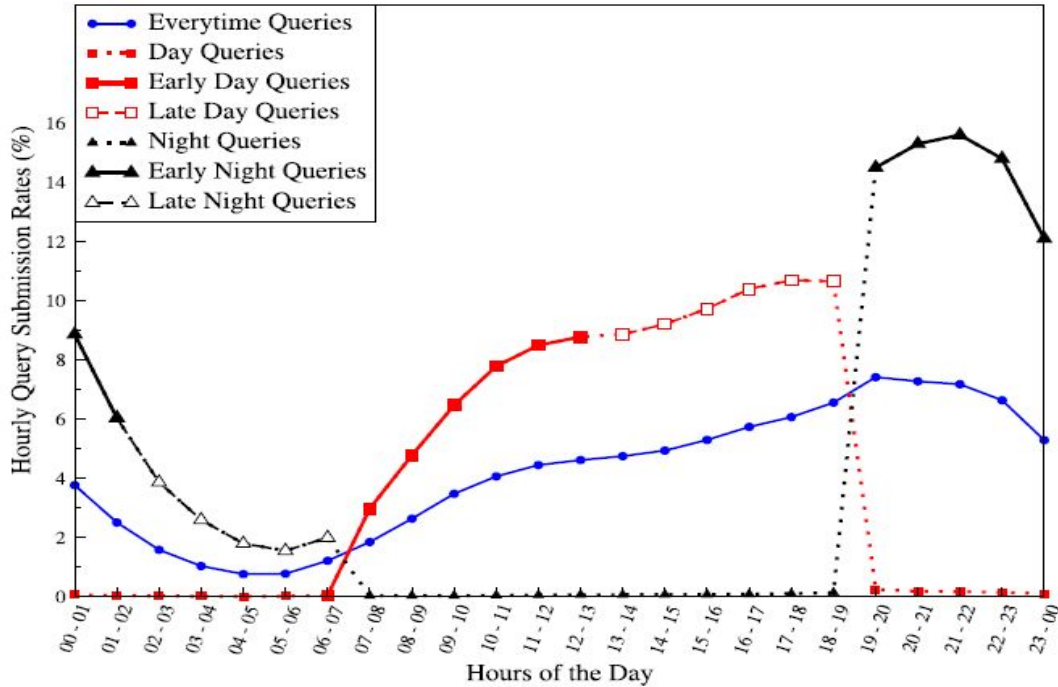


(b) Daytime and Nighttime (MSN)



QF of a query is percentage of submissions performed within the corresponding day-partition

Observation

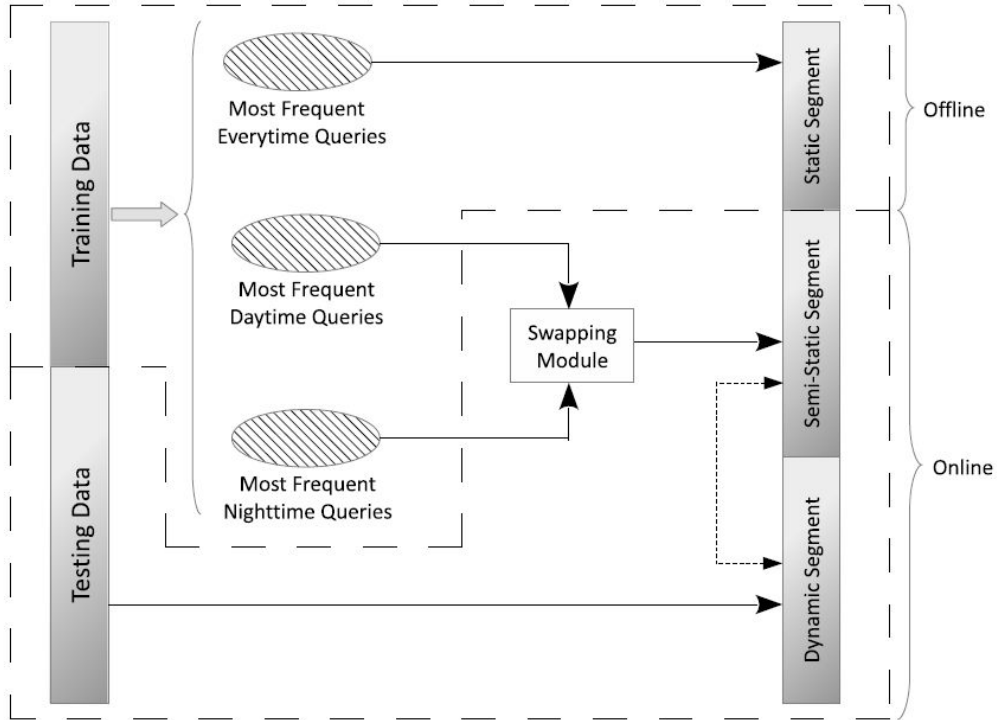


- Night-queries have larger submission rate during the night hours. Day-queries have a larger submission rate during day hours
- Sharp shift in query submissions during 18:00 PM and 19:00 PM. Possible reason being end of business hours.

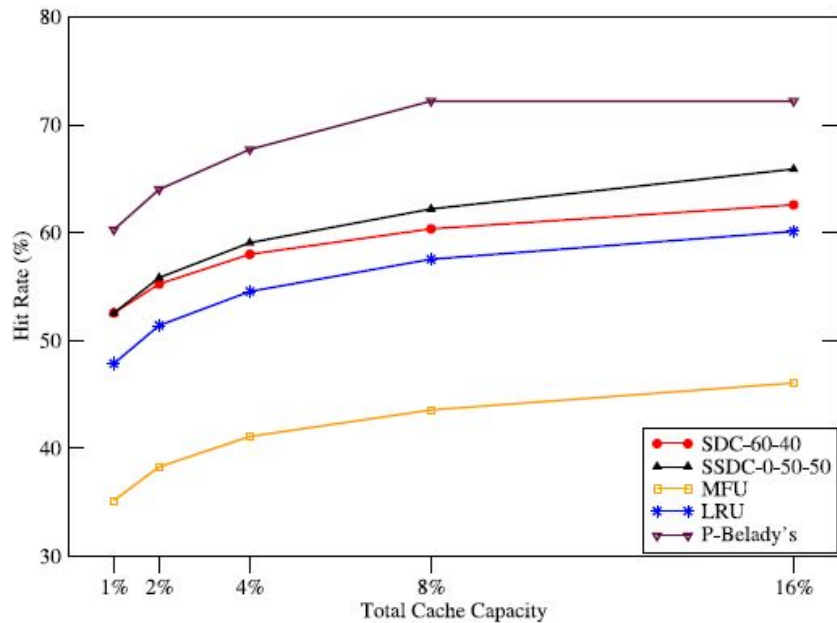
Mechanism

- Uses normalized temporal query frequency (NT-Freq) - hourly submission rate of a query.
- 3 NT-Freq values are calculated
 - Hourly query frequency collected for 24 h (i.e., query frequency divided by 24)
 - During day hours (i.e., hourly query frequency collected during 07:00 AM until 19:00 PM)
 - During night hours (i.e., hourly query frequency collected only during 19:00 PM until 07:00 AM)
- NT-Freq values collected for 24 h are placed in the static segment in an offline fashion.
- NT-Freq day time are and night time are swapped between day and night

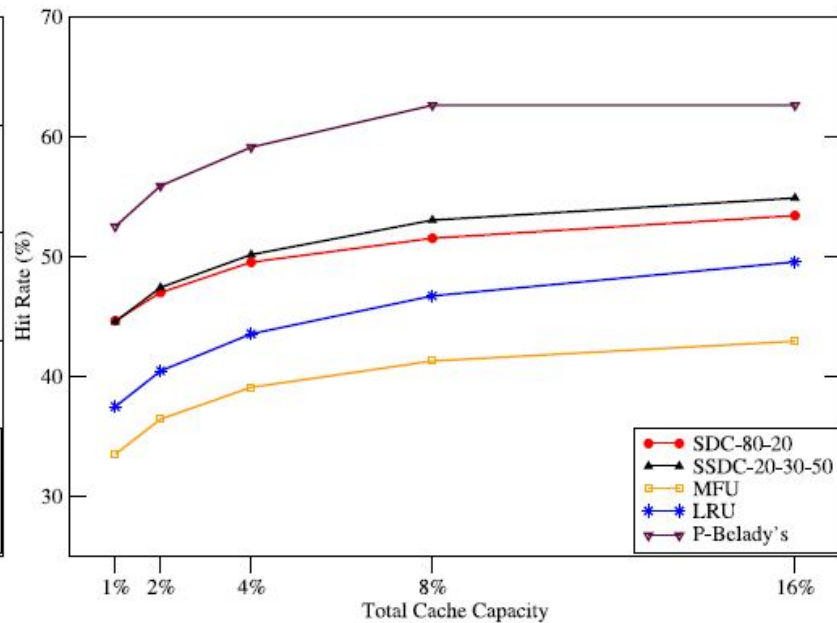
Architecture



Results



(a) AOL dataset



(b) MSN dataset

Results

Comparison of the hit rates of SDC and SSDC for the AOL dataset (two partitions).

Caching technique	Total cache capacity (%)				
	%1	%2	%4	%8	%16
SDC-80-20	53.036	55.578	58.011	60.378	62.498
SSDC(2)-70-10-20	52.953	55.500	58.067	60.381	63.556
SSDC(2)-60-20-20	52.991	55.582	58.215	60.865	63.558
SSDC(2)-50-30-20	53.029	55.700	58.401	61.072	63.657
SSDC(2)-40-40-20	53.100	55.779	58.592	61.250	63.834
SSDC(2)-30-50-20	53.135	55.841	58.788	61.408	64.347
SSDC(2)-20-60-20	53.150	55.898	58.921	61.632	64.932
SSDC(2)-10-70-20	53.156	55.932	59.010	61.854	65.390
SSDC(2)-0-80-20	53.160	55.944	59.036	61.947	65.617
SDC-60-40	52.559	55.235	57.990	60.350	62.575
SSDC(2)-50-10-40	52.542	55.289	58.051	60.625	62.584
SSDC(2)-40-20-40	52.660	55.562	58.394	60.950	63.601
SSDC(2)-30-30-40	52.749	55.741	58.667	61.494	64.356
SSDC(2)-20-40-40	52.802	55.872	58.933	61.861	64.679
SSDC(2)-10-50-40	52.824	55.946	59.084	62.116	65.025
SSDC(2)-0-60-40	52.835	55.969	59.138	62.278	65.266
SDC-50-50	52.175	54.959	57.765	60.260	62.535
SSDC(2)-40-10-50	52.244	55.132	57.972	60.501	63.045
SSDC(2)-30-20-50	52.407	55.473	58.405	61.147	63.753
SSDC(2)-20-30-50	52.487	55.671	58.778	61.652	64.748
SSDC(2)-10-40-50	52.525	55.789	58.990	61.998	65.479
SSDC(2)-0-50-50	52.540	55.819	59.067	62.190	65.885
SDC-40-60	51.732	54.621	57.457	60.067	62.465
SSDC(2)-30-10-60	51.947	54.945	57.821	60.520	63.181
SSDC(2)-20-20-60	52.115	55.332	58.401	61.354	64.049
SSDC(2)-10-30-60	52.197	55.505	58.741	62.004	64.856
SSDC(2)-0-40-60	52.221	55.562	58.845	62.272	65.473
SDC-20-80	50.325	53.482	56.446	59.237	61.826
SSDC(2)-10-10-80	51.067	54.479	57.867	61.074	64.007
SSDC(2)-0-20-80	51.170	54.672	58.284	61.876	65.480