Aging effects on Query Flow Graphs for Query Suggestion

Carlos Castillo, Debora Donato

Yahoo! Research
Barcelona, Spain
{chato, debora}@yahoo-inc.com

Ranieri Baraglia, Franco Maria Nardini
Raffaele Perego, Fabrizio Silvestri

Institute of Information Science and Technologies
Italian National Research Council
Pisa, Italy
{name.surname}@isti.cnr.it

ABSTRACT

As users interests change over time, the knowledge extracted from query logs may suffer an aging effect as new interesting topics appear. In order to validate experimentally this hypothesis, we consider the problem of query recommendation. A recent query-log mining approach for query recommendation is based on Query Flow Graphs (QFG). We thus propose an **evaluation of the effects of time on this query recommendation model**. **We build different query flow graphs from the queries belonging to a large query log of a real-world search engine**. Each query flow graph is built on distinct query log segments. Then, **we generate recommendations** on different sets of queries. **Results are assessed both by means of human judgments and by using an automatic evaluator showing that the models inexorably age**.

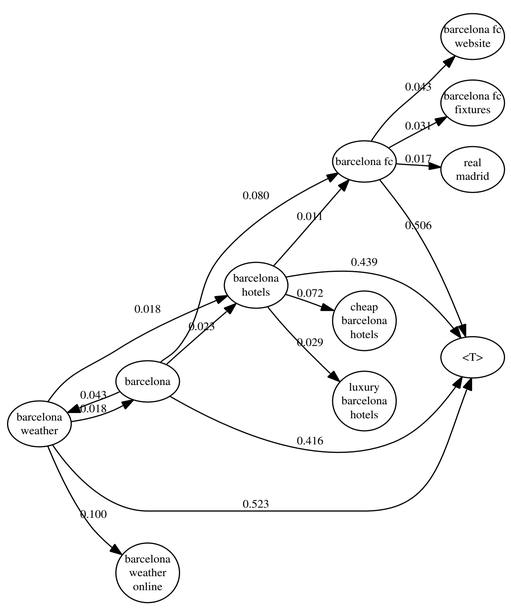
The Query Flow Graph Model

A successfully query-log mining approach for generating useful query recommendation is based on **Query Flow Graphs** (QFGs).

The QFG model **aggregates information** in a query log by providing a **markov-chain representation** of the query reformulation process followed by users trying to satisfy the same information need.

Each query is represented by a **single node** independently of its frequency, or of the number of distinct users who issued it.

The **query recommendation method** is based on a **random walk over a query graph** starting from the query for which we want suggestions.



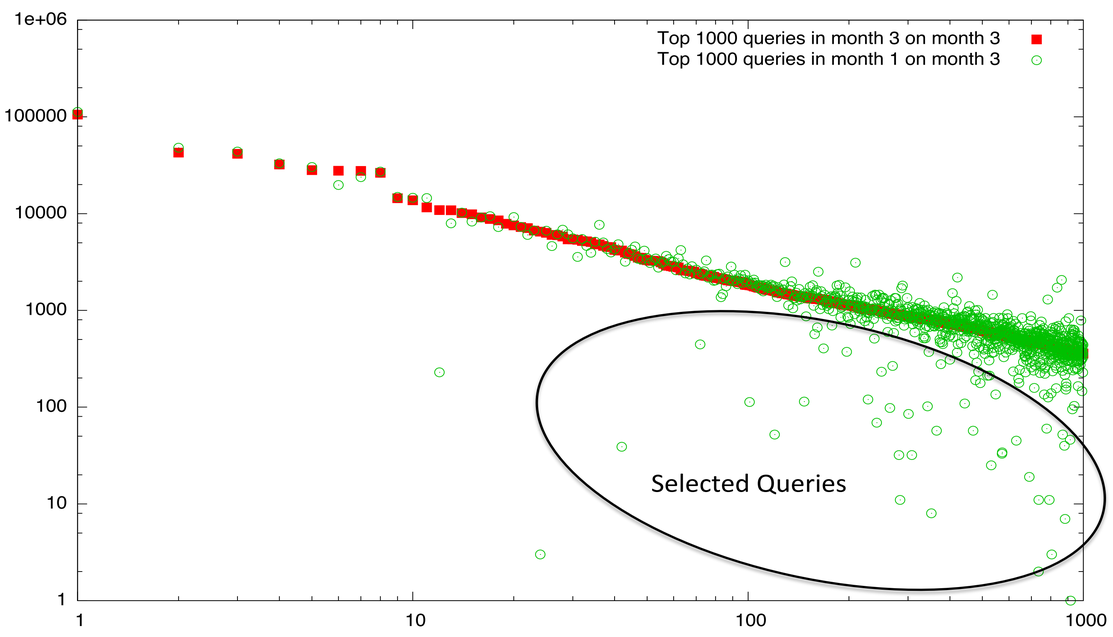
Selecting Test Queries

In order to assess the various reasons why a QFG-based model ages we have considered, for each segment, two classes of queries (F1 and F3).

F1 is the set of the 30 queries that are among the 1,000 most frequent queries in the first month (M1) but **whose frequency has had the greater drop in the last month covered by the query log (test log)**.

F3 is the set of the 30 queries among the 1,000 most frequent queries in the test log **whose frequency has the greater drop in the first part of the log M1.**

We **generate recommendations** on this two sets of queries **assessing their quality using an user study**.

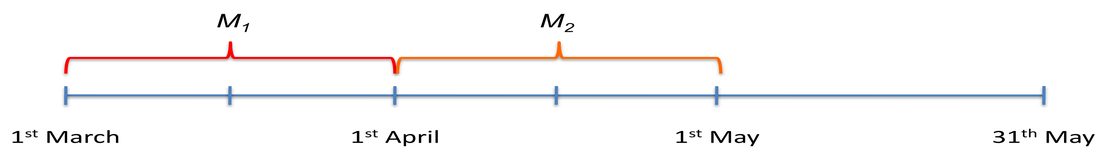


Assessing the aging effect

Our experiments have been conducted on the **AOL query log** (**20M queries for 650K users**, **from 1st March, 2006 to 31st May, 2006**).

To assess the aging effects we conducted several experiments to evaluate the impact of different factors.

The log has been split into three different segments. Two of them have been used for training (M1, M2) and the third one for testing (test log). **Two QFG has been built**. **One using M1 and one using M2**.



Results

Here the suggestions and their relative score, computed for queries from F1 and F3 on M1 and M2.

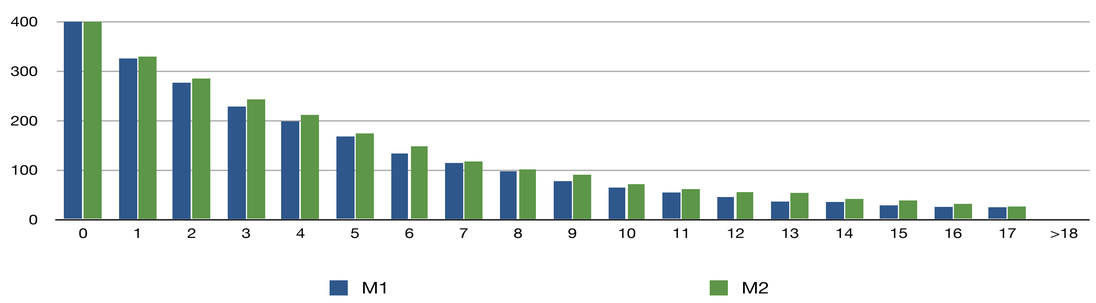
Going down in the list, **we observe that less useful suggestions are associated with low score values.**

Query Set	Query	$\mathcal{M}_1$		$\mathcal{M}_2$	
$\mathcal{F}_3$	da vinci	49743	da vinci's self portched black and white	73219	da vinci and math
		47294	the vitruvian man	33769	da vinci biography
		35362	last supper da vinci	31383	da vinci code on portrait
		31307	leonardo da vinci	29565	'ying machines
		30234	post it	28432	inventions by leonardo da vinci
		30234	handshape 20stories	26003	leonardo da vinci paintings
				23343	friends church
				23343	jerry c website
$\mathcal{F}_1$	harley davidson	5097	harley davidson ny	5749	harley davidson premium sound system owners manual
		2652	american harley davidson	3859	automatic motorcycles
		2615	2002 harley davidson ultra classic	3635	harley davidson credit
		2602	adamec harley davidson	3618	cherokee harley davidson
		2341	air ↓ght	2103	harley davidson sporster
		2341	928 zip code	1965	2002 harley davidson classic
		2341	antispy ware	1394	regions banking
				1394	aol email only
				1394	adultactioncamcom

We make the following hypothesis: **when a QFG-based query recommender system gives the same score to consecutive suggestions, these recommendations and the following ones having a lower score are very likely to be less useful**.

We will use the hypothesis to derive an **automatic evaluation methodology to assess the usefulness of suggestions**. We perform the automatic evaluation on the **400 most frequent queries in the third month (test log)**.

To highlight more the **aging effect** we show the total number of queries **having at least a certain number of useful recommendation.** For example, the **third bucket** shows how many **queries have at least three useful suggestions.**



We can observe that a model trained on M2 has a larger percentage of queries for which the number of useful suggestions is at least 4.

This confirms our hypothesis that QFG-based recommendation models age.

filtering threshold	average number of useful suggestions on $\mathcal{M}_1$	average number of useful suggestions on $\mathcal{M}_2$
0	2.84	2.91
0.5	5.85	6.23