Mining Search Subtopics from Query Logs

Dan Zhu, Jianwei Cui, Jun He, Hongyan Liu, Xiaoyong Du
Key Labs of Data Engineering and Knowledge Engineering, MOE, China.
School of Economics and Management, Tsinghua University
School of Information, Renmin University of China

Motivation

Web queries are usually short and ambiguous. Different users often have different intents for the same query. Understanding different intents of users is a significant problem in information retrieval, which is critical for search engines to provide correct information for users and to enhance user's

Query: Mars

VS.

Key Techniques

Sub-problems of subtopic mining

1. Find related queries
2. Measure relationships between different related queries and find users' intents.

What are related queries?

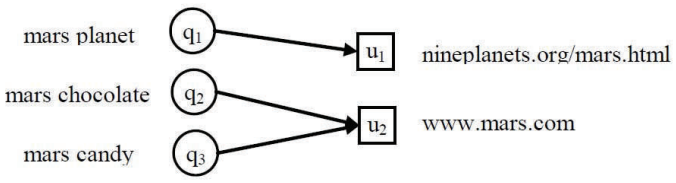
Original Query : mars
Related Queries: mars planet
mars chocolate
mars candy

Major Steps

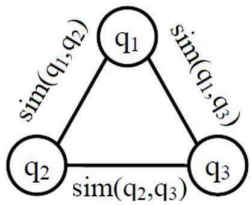
1. Find related queries from

Note: noises
灯红酒绿 (feasting and revelry) vs. 红酒 (red wine)

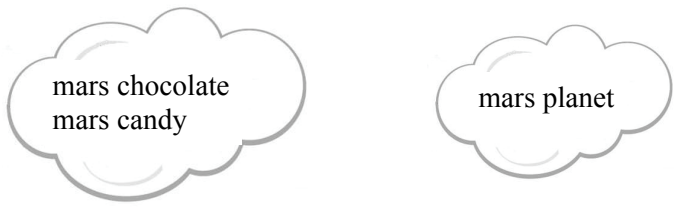
2. Construct query-URL bipartite



3. Measure query similarity based link information by using *SimRank*



4. Cluster related queries to get different topics



Evaluation

Mean I-rec, D-nDCG and D#-nDCG

I-rec@10	D-nDCG@10	D#-nDCG@10
0.4694	0.5620	0.5157
I-rec@20	D-nDCG@20	D#-nDCG@30
0.4926	0.3948	0.4437
I-rec@30	D-nDCG@30	D#-nDCG@30
0.4937	0.3068	0.4010

Top 20 subtopics for query "王子变青蛙" (Prince Turns to Frog)

Intents	English translation
免费看《王子变青蛙》	Watch for free
《王子变青蛙》下载	Download
观看《王子变青蛙》	Watch
免费下载《王子变青蛙》	Download for free
《王子变青蛙》结局	Ending
电视剧：《王子变青蛙》	TV play
《王子变青蛙》16-20集	Episode 16 to 20
在线播放《王子变青蛙》	Watch online
《王子变青蛙》花絮	Tidbits
《王子变青蛙》分集介绍	Introduction to each episode
《王子变青蛙》音乐	Music
《王子变青蛙》的剧本	Script
《王子变青蛙》免费下载	Download for free
《王子变青蛙》电视剧	TV play
《王子变青蛙》最新消息	Latest news
《王子变青蛙》主题曲	Theme song
《王子变青蛙》的歌	Songs
《王子变青蛙》图	Pictures
《王子变青蛙》28集	Episode 28
site:ent.sina.com.cn+<王子变青蛙>	Search in the specified site

Acknowledgement

The work was supported in part by the National Nature Science Foundation of China under Grant No. 70871068 and 70890083 and HGJ project 2010ZX01042-002-002-03.