Understanding Temporal Query Dynamics

Anagha Kulkarni
Carnegie Mellon University

Jaime Teevan, Krysta M. Svore, Susan Dumais
Microsoft Research

WSDM 2011
Hong Kong, Feb. 9-12, 2011
Motivation

Web search is strongly influenced by time

- Queries’ popularity changes over time
- Document content changes over time
- Query intent (what’s relevant to the user) changes over time
Take away:
Changes in query popularity and document content can signal changes in query intent.
Example: *march madness*

http://ncaa.com becomes relevant *during*
Example: *march madness*

http://mahalo.com/march-madness is always about March Madness and is relevant *before*, *during*, and *after*.
**Our Study**

- 100 temporal queries, 20 documents per query

<table>
<thead>
<tr>
<th>Year</th>
<th>Type of change targeted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Δ query popularity</td>
</tr>
<tr>
<td>2009</td>
<td>Queries that spiked during this time period last year <em>(april fools, tax extension)</em></td>
</tr>
<tr>
<td>2010</td>
<td>Queries related to scheduled upcoming events <em>(ipad, crystal bowersox)</em></td>
</tr>
</tbody>
</table>
Data Collected

March 25 – May 28, 2010

Query
- Frequency (daily)

Document
- Number of clicks (per query, daily)
- Result rank (per query, daily)
- Crawl of document content (daily)

Query Intent
- Human relevance judgments (weekly)
Changes in Query Popularity

Methods

- Measure query frequency over time
- Characterize query popularity curves
  - Number of spikes / Periodicity / Shape / Trend
**Example: Query Popularity**

Query Popularity - march madness

*march madness*

Multiple spikes / not periodic / castle / downward trend

**During tournament**

**After tournament**
**Example:** Query Popularity

**Query Popularity - dancing with the stars**

*dancing with the stars*

Multiple spikes / periodic / castle / flat

![Query Popularity Chart](image)
Example: Query Popularity

**Query Popularity - mlb**

*mlb*

Multiple spikes / not periodic / castle / upward trend

Pre-season  During season
Example: Query Popularity

Query Popularity - miley cyrus

miley cyrus
1 spike / not periodic / wedge / flat

New music video
Example: Query Popularity

Query Popularity - earthquake

earthquake
1 spike / not periodic / right sail/ flat

Baja CA and Mexico quake
Example: Query Popularity

Query Popularity - tax extension

tax extension
1 spike / not periodic / left sail / flat

Tax day
Findings:
Query Popularity

- Most queries in study
  - One or more spikes
  - Not periodic
  - Wedge
  - Show temporal trend
Changes in Document Content

Methods

- Crawl documents daily
- Measures of content change
  - *Query-dependent*: Term frequency (TF) of query words
  - *Query-independent*: Dice coefficient over words

\[
Dice(W_i, W_j) = \frac{2|W_i \cap W_j|}{|W_i| + |W_j|}
\]
Example: Document Content

Content Change - march madness

During tournament  After tournament

http://www.mahalo.com/march-madness
http://www.ncaa.com/
Findings: Document Content

- 61% pages show query-related (TF) change
- 95% pages changed by less than 15% (Dice)
- Periodic and multiple-spike queries are more likely to have content-changed documents
Changes in Query Intent

Methods

- Measures of intent change
  - Explicit human relevance judgments
    - Count number of top-labeled documents per query
  - Entropy of clicked results
    - \[ CE(q) = - \sum_{u \in P_c(q)} p_c(u|q) \cdot \log p_c(u|q) \]
Example: Query Intent

Human Judgments - march madness

During: scores/schedules
After: general info/planning

<table>
<thead>
<tr>
<th>Date</th>
<th>Human Relevance Judgments</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-25</td>
<td>0</td>
</tr>
<tr>
<td>4-01</td>
<td>0</td>
</tr>
<tr>
<td>4-09</td>
<td>0</td>
</tr>
<tr>
<td>4-15</td>
<td>0</td>
</tr>
<tr>
<td>4-22</td>
<td>0</td>
</tr>
<tr>
<td>4-29</td>
<td>0</td>
</tr>
<tr>
<td>5-07</td>
<td>0</td>
</tr>
<tr>
<td>5-13</td>
<td>0</td>
</tr>
<tr>
<td>5-21</td>
<td>0</td>
</tr>
<tr>
<td>5-28</td>
<td>0</td>
</tr>
</tbody>
</table>

- http://www.ncaa.com/
- http://www.cbssports.com/collegebasketball
- http://www.mahalo.com/march-madness
- http://mmod.ncaa.com/
Findings: Query Intent

- Click entropy is correlated with
  - Spikes
    - 1 spike has lower click entropy than Multiple spikes
  - Shape
    - Sail (taxes online, pgatour) has significantly lower click entropy than wedge (robert pattinson, kristen stewart)
  - Trend
    - Downwards has lower click entropy than flat (microsoft update)
Summary

Changes in query popularity and document content are related to changes in query intent

- Periodicity
- $\Delta$Dice
- $\Delta$TF
- Human Relevance
Future Directions

Understanding query intent can be used to improve
- Captions, Ranking, Results display

Patterns of query intent change
- **Zoom**: *march madness* – intent changes from general to specific and back to general
- **Shift**: *opening day* – different intents throughout year
- **Static**: *easter ideas* – same intent throughout year
Questions?