

# Place Value: Word Position Shifts Vital To Search Dynamics

Rishiraj Saha Roy, Anusha Suresh and Niloy Ganguly  
IIT Kharagpur, India

Monojit Choudhury  
Microsoft Research India

## ABSTRACT

With fast changing information needs in today's world, it is imperative that search engines precisely understand and exploit temporal changes in Web queries. In this work, we look at shifts in preferred positions of segments in queries over an interval of four years. We find that such shifts can predict key changes in usage patterns, and explain the observed increase in query lengths. Our findings indicate that recording positional statistics can be vital for understanding user intent in Web search queries.

**Contact:**  
**Rishiraj Saha Roy**  
**rishiraj.saharoy@gmail.com**  
**Phone: 0-2194076766**

## DATA

**2006:** AOL USA Query Log (Pass et al., InfoScale 2006), 12.8M Sample (August 2006)

**2010:** Bing Australia Query Log, 11.9M Sample (May 2010)

## METHOD

Queries in both the logs are segmented using a state-of-the-art segmentation algorithm (Saha Roy et al., SIGIR 2012). Segments with the highest co-occurrence counts (top-2000) are labeled as intent (Yin and Shah, WWW 2010; Yu and Ren, CIKM 2012), and the rest as content. For each segment  $s$  that appears in both logs, we compute the query beginning probability ( $P_b(s)$ ), ending probability  $P_e(s)$  and that of occurring in the middle  $P_m(s)$ , and the occurrence probability in the log,  $P_{occ}(s)$ . Positional trends are now identified in each of the following four classes.

### CONTENT IN 2006, INTENT IN 2010

$P_b(s)$  drops  
Dominant Trend

393/445

#Support Segments

| Segment | Example query in 2006 | Example query in 2010   |
|---------|-----------------------|-------------------------|
| youtube | youtube videos        | new visions 31 youtube  |
| xbox    | xbox logo             | sonic the hedgehog xbox |

- ❖ 2006: Segments mostly issued as navigational queries for internal searches or as informational queries
- ❖ 2010: Appended with content words as search engines can handle direct queries now
- ❖ Manifold increase in frequency → user-guided SE rules
- ❖ Positional dynamics vital to query intent detection!

### INTENT IN 2006, CONTENT IN 2010

$P_e(s)$  drops  
Dominant Trend

481/576

#Support Segments

| Segment      | Example query in 2006        | Example query in 2010      |
|--------------|------------------------------|----------------------------|
| yellow pages | granger indiana yellow pages | yellow pages usa wikipedia |
| motels       | maryland motels              | motels for sale brisbane   |

- ❖ 2006: High popularity leads to intent labels
- ❖ 2010: Usage becoming obsolete, esoteric interests
- ❖ Error analysis: *what to do*, *cheat codes* and *official site* labeled as content in 2010 due to drops in co-occurrence counts
- ❖ But relative positions still indicate “intent”-ness!

### CONTENT IN 2006, CONTENT IN 2010

$P_b(s)$  drops or  $P_e(s)$  drops  
Dominant Trend

9210/9293

#Support Segments

| Segment      | Example query in 2006   | Example query in 2010                         |
|--------------|-------------------------|---|
| epilepsy     | treatments for epilepsy | light sensitive epilepsy light emitting diode |
| harry potter | harry potter            | harry potter game photos quidditch            |

- ❖ Segments are mostly entities or classes
- ❖ 2006: Several standalone queries or with basic intent
- ❖ 2010: Increased specificity of user needs leads to (multiple) intent words added to left or right
- ❖ This leads to increased mean (distinct) query lengths: 3.5 words in 2006 to 3.98 words in 2010!

### INTENT IN 2006, INTENT IN 2010

$P_b(s)$  rises or  $P_e(s)$  rises  
Dominant Trend

1234/1253

#Support Segments

| Segment    | Example query in 2006 | Example query in 2010               |
|------------|-----------------------|-------------------------------------|
| download   | realplayer download   | realplayer beta free download       |
| meaning of | meaning of bipolar    | meaning of bipolar for young people |

- ❖ Generally increased segment frequency from 2006 to 2010 – users specifying intent more often
- ❖ Mostly appear to right of content – content segments are conceived first in user model of query formulation
- ❖ Stacking of intent a major factor in increased mean query lengths!