A Survey on Temporal Web Search Experience

Hideo Joho\textsuperscript{1}, Adam Jatowt\textsuperscript{2}, and Roi Blanco\textsuperscript{3}
\textsuperscript{1}University of Tsukuba, Japan
\textsuperscript{2}Kyoto University, Japan
\textsuperscript{3}Yahoo! Research, Barcelona, Spain
Outline

• Motivation
• Data Collection Method
• Gross Statistics
• Correlation → Breakdown
• (Some) Qualitative Results
• Discussion
• Summary
Motivation

• Much of research on temporal search focuses on log analyses or system development
• A broader and deeper understanding of people’s temporal information seeking behaviour can facilitate the community
• Let’s ask people!
Data Collection Method 1

• Critical Incident Analysis
  – Asking people to describe a prominent experience
  – Usually carried out via interviews
  – Evans & Chi [5] applied to questionnaires

• 18 Questions
  – Most recent web search experience
  – 15 closed and 3 open questions (Broad scope)
  – Contexts, Information Needs, Process, Outcomes
Data Collection Method 2

• 110 Participants in Japan
  – 11 Females + 11 Males / 20s – 60s Blocks
  – Filtered by a search experience within 3 days
  – 48% single, 52% married
  – Self-reported occupations
    • Office Workers (36), Homemakers (21) Part-time workers (15), University Students (8), Others (16), Managers (2)

• Mostly quantitative analyses so far
Context: When

Mostly fresh memories

- < 1 hour: 43%
- < 1 day: 38%
- Yesterday: 15%
- < 2 days: 3%
- > 2 days: 1%
Context: Where

- Home: 79%
- Office: 17%
- Transport: 2%
- Outdoor: 2%

Very indoor
Context: How

- Desktop PC: 35%
- Laptop/Tablet: 50%
- Mobile: 15%

Mixed devices
Info Needs: Target Time

Recency is important
Info Needs: Seasonal Interests

76% is some kind of temporal needs
Info Needs: Fresh, Re-find, Tech

Table 5: Freshness, re-findability, and technicality of information needs.

<table>
<thead>
<tr>
<th>Freshness</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshness was important</td>
<td>67</td>
<td>60.9%</td>
</tr>
<tr>
<td>Freshness was not so important</td>
<td>43</td>
<td>39.1%</td>
</tr>
<tr>
<td>Total</td>
<td>110</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Re-finding</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have searched before</td>
<td>72</td>
<td>65.5%</td>
</tr>
<tr>
<td>Have never searched before</td>
<td>38</td>
<td>34.5%</td>
</tr>
<tr>
<td>Total</td>
<td>110</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Technicality</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Searching technical information</td>
<td>16</td>
<td>14.5%</td>
</tr>
<tr>
<td>Searching general information</td>
<td>94</td>
<td>85.5%</td>
</tr>
<tr>
<td>Total</td>
<td>110</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
Process: Search Time

- < 5 min: 45%
- 5 - 10 min: 31%
- 10 - 30 min: 3%
- 30 min - 1h: 3%
- > 1h: 3%
- N/A: 4%

Nothing unusual?
Process: Difficulty

2/3 still have a difficulty (in general)
Outcome: Found?

Yet managed to find relevant information
Outcome: Satisfaction

Still 17% are not happy
Outcome: Information Use

35% was not used within a day

<table>
<thead>
<tr>
<th>Time Frame</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (Now)</td>
<td>28</td>
</tr>
<tr>
<td>&lt; 1D</td>
<td>18</td>
</tr>
<tr>
<td>&lt; nD</td>
<td>26</td>
</tr>
<tr>
<td>&lt; 1W</td>
<td>2</td>
</tr>
<tr>
<td>&lt; nW</td>
<td>9</td>
</tr>
<tr>
<td>&lt; 1M</td>
<td>2</td>
</tr>
<tr>
<td>&lt; nM</td>
<td>2</td>
</tr>
<tr>
<td>&lt; 1Y</td>
<td>1</td>
</tr>
<tr>
<td>&lt; nY</td>
<td>1</td>
</tr>
<tr>
<td>No Plan</td>
<td>21</td>
</tr>
</tbody>
</table>
Correlation

Table 10: Gross Correlation analyses by Spearman. Question number are based on Appendix A. Significant correlation ($p \leq .05$) is highlighted.

<table>
<thead>
<tr>
<th></th>
<th>Q2 Device</th>
<th>Q3 Location</th>
<th>Q5 Target</th>
<th>Q6 Season</th>
<th>Q7 Specific</th>
<th>Q8 Fresh</th>
<th>Q9 Re-find</th>
<th>Q12 Clarity</th>
<th>Q13 Use</th>
<th>Q14 Time</th>
<th>Q15 Found</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2 Device to search</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3 Location of search</td>
<td>-0.04</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5 Target time of information</td>
<td>-0.14</td>
<td>0.22</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6 Seasonal needs</td>
<td>-0.22</td>
<td>0.08</td>
<td>0.20</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7 Specificity of needs</td>
<td>0.08</td>
<td>-0.01</td>
<td>-0.20</td>
<td>-0.20</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8 Freshness of needs</td>
<td>-0.07</td>
<td>-0.06</td>
<td>-0.04</td>
<td>0.11</td>
<td>0.30</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9 Re-findability</td>
<td>-0.08</td>
<td>0.04</td>
<td>0.06</td>
<td>0.18</td>
<td>-0.03</td>
<td>0.24</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q12 Clarity of needs</td>
<td>0.05</td>
<td>0.03</td>
<td>-0.04</td>
<td>-0.10</td>
<td>-0.09</td>
<td>-0.05</td>
<td>-0.07</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q13 Information Use</td>
<td>-0.08</td>
<td>-0.14</td>
<td>-0.07</td>
<td>-0.04</td>
<td>0.01</td>
<td>-0.09</td>
<td>-0.08</td>
<td>0.24</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q14 Time taken to search</td>
<td>-0.02</td>
<td>-0.14</td>
<td>0.03</td>
<td>-0.08</td>
<td>0.06</td>
<td>-0.11</td>
<td>-0.22</td>
<td>0.27</td>
<td>0.16</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Q15 Rel doc found</td>
<td>-0.12</td>
<td>0.00</td>
<td>-0.03</td>
<td>0.05</td>
<td>0.13</td>
<td>0.11</td>
<td>-0.14</td>
<td>0.39</td>
<td>0.19</td>
<td>0.25</td>
<td>1.00</td>
</tr>
<tr>
<td>Q18 Satisfaction</td>
<td>-0.15</td>
<td>0.10</td>
<td>-0.02</td>
<td>-0.02</td>
<td>0.11</td>
<td>-0.07</td>
<td>-0.06</td>
<td>0.39</td>
<td>0.09</td>
<td>0.26</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Focusing on some of significant pairs

2013/05/13
Location vs. Target Time

Recent information when outside or stops

No future information sought from offices

2013/05/13
Technical information tends to be long-term or recent interests
Freshness vs. Technicality

Freshness is important for general not technical
Re-finding vs. Freshness

Re-finding is likely to prefer fresh information

Cannot tell for first-time search
Re-finding vs. Seasonal

First time search is likely to prefer recent info

Cannot tell for re-finding search
Technicallity vs. Target Time

No future information sought by technical needs
Qualitative analyses

• A lot of TV is going on
• A lot of Internet browsing is going on
• Weak information needs
• Future temporal expression can be difficult
  – Project X is supposed to complete by the end of January 2015.
  – Temporal incidents in fictions
Summary

• Questionnaire-based CIA was carried out to gain broader & deeper understanding of temporal information seeking behaviour
• Results reinforcing the importance of recency needs
• Seasonal interests, technicality, target time, re-finding, and freshness can all interplay to affect people’s temporal search behaviour.
• How can we deal with the gap between search and use?
• Smaller proportion of future information search is natural cause or lack of advancement?


