Instability in Search Engine Results
Lessons learnt in the context of horizon scanning

Marco Palomino
Tim Taylor
Geoff McBride
Richard Owen
Outline

• What is horizon scanning?
• How do we scan the horizon?
  – Web-based horizon scanning
• Search engine usage
  – A day-by-day example
• Google’s results instability
• Conclusions
What is horizon scanning?

• **Horizon scanning** is the systematic search for information to identify early signs of new and emerging risks and opportunities that might become important in the future.

• It is the search for **anything new** that will be important.
Approaches

• Traditional approach: Emerging issues are identified through “experts”.

• Web-based approach: Emerging issues are identified through the World Wide Web.
Web-based approach

Iteration

Information retrieval: keyword selection and search

Analysis of information: categorisation and archival

Communication of results: report

Interface with further tools: e.g., risk and opportunity analysis

Web-based horizon scanning

Decision support: e.g., resource allocation and policy

Analysis of information: categorisation and archival
Search results: weak signals

- Known issues
- New, emerging issues

Number of documents
Google results

Query: eHealth; breakthrough

Number of results

02/08/2012 03/08/2012 04/08/2012 05/08/2012 06/08/2012 07/08/2012 08/08/2012 09/08/2012 10/08/2012
Google results

Query: eHealth; breakthrough

Number of results

Google WUI

Google API

03/08/2012
17/08/2012
01/09/2012
16/09/2012
01/10/2012
16/10/2012
31/10/2012
## Same day instability

<table>
<thead>
<tr>
<th>Query</th>
<th>Google’s WUI 05/12/2012</th>
<th>Google’s WUI 05/12/2012</th>
<th>Google’s API 05/12/2012</th>
<th>Google’s API 05/12/2012</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>09:45</td>
<td>17:45</td>
<td>09:40</td>
<td>17:40</td>
</tr>
<tr>
<td>ehealth; breakthrough</td>
<td>167,000</td>
<td>167,000</td>
<td>21,500</td>
<td>21,400</td>
</tr>
<tr>
<td>telemed; breakthrough</td>
<td>377</td>
<td>197</td>
<td>669</td>
<td>665</td>
</tr>
<tr>
<td>ehealth; &quot;closer to reality&quot;</td>
<td>3,340</td>
<td>3,470</td>
<td>470</td>
<td>469</td>
</tr>
<tr>
<td>telemed; &quot;closer to reality&quot;</td>
<td>12,300</td>
<td>12,300</td>
<td>457</td>
<td>457</td>
</tr>
<tr>
<td>ehealth; &quot;paves the way&quot;</td>
<td>9,350</td>
<td>8,830</td>
<td>1,460</td>
<td>1,460</td>
</tr>
<tr>
<td>telemed; &quot;paves the way&quot;</td>
<td>22,800</td>
<td>22,800</td>
<td>1,140</td>
<td>1,170</td>
</tr>
<tr>
<td>ehealth; &quot;previously impossible&quot;</td>
<td>82</td>
<td>82</td>
<td>103</td>
<td>103</td>
</tr>
<tr>
<td>telemed; &quot;previously impossible&quot;</td>
<td>61</td>
<td>603</td>
<td>501</td>
<td>502</td>
</tr>
<tr>
<td>ehealth; revolutionary</td>
<td>21,200</td>
<td>21,200</td>
<td>34,200</td>
<td>34,400</td>
</tr>
<tr>
<td>telemed; revolutionary</td>
<td>364</td>
<td>594</td>
<td>1,740</td>
<td>1,750</td>
</tr>
<tr>
<td>ehealth; unprecedented</td>
<td>12,100</td>
<td>54,200</td>
<td>9,650</td>
<td>9,650</td>
</tr>
<tr>
<td>telemed; unprecedented</td>
<td>589</td>
<td>585</td>
<td>621</td>
<td>623</td>
</tr>
</tbody>
</table>
Conclusions

• Google’s API offers less documents than Google’s WUI, but the documents are more recent.
• The same behaviour shown by Google’s interfaces appears to be shown by Bing’s interfaces.