Understanding Reader Backtracking in Online News Articles

Uzi Smadja, Max Grusky, Yoav Artzi, Mor Naaman

What is backtracking?
The action of scrolling back in a browser while reading an online news article.

Types of backtracking (BT) events
- Simple
- Continued
- Full

Why might backtracking occur?
Readers re-read a piece of text that they did not fully understand, indicating readability issues.

Data
Obtained from Chartbeat
- 1.4M sessions
- 26k users
- 8k online news articles
- two major online publications: long-form news magazine (LONG), short-form news website (SHRT)

Are some pages more prone to backtracking?
Full backtrack events are more common, especially on long-form pages.

Number of backtracking events for long and short articles:

<table>
<thead>
<tr>
<th></th>
<th>Users</th>
<th>Pages</th>
<th>Sessions</th>
<th>Simple BT</th>
<th>Cont. BT</th>
<th>Full BT</th>
</tr>
</thead>
<tbody>
<tr>
<td>LONG</td>
<td>15,949</td>
<td>3,968</td>
<td>694,728</td>
<td>57,897</td>
<td>28,840</td>
<td>246,787</td>
</tr>
<tr>
<td>SHRT</td>
<td>11,117</td>
<td>3,957</td>
<td>763,335</td>
<td>136,153</td>
<td>49,211</td>
<td>127,421</td>
</tr>
</tbody>
</table>

Mean number of backtracking events per reading session:

What textual features predict backtracking?

Lexical: readability measures & simple text properties
Entity-density: proportions of named entities
Part-of-speech density: proportions of nouns
Coreference chains: characteristics of coreference chains

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>F1</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>All features</td>
<td>0.838</td>
<td>0.839</td>
<td>0.901</td>
</tr>
<tr>
<td>Length only (control)</td>
<td>0.747</td>
<td>0.739</td>
<td>0.809</td>
</tr>
</tbody>
</table>

Find in the full paper:
- Feature ablation showing relative contribution of each family of features
- How we control for different types of readers and topics, finding no effect on the prediction task

What does this mean for Web content and readability?

- Full backtrack events may be an indication of readability issues in online news.
- Large-scale Web signals can help define improved readability measures.

Learn more:
bite.ly/backtracking-www
Contact: uzi.smadja@gmail.com
Support: Yahoo! Research, NSF IIS-1840751