Enhanced Visualization for Web-Based Summaries

Brent Wenerstrom
Mehmed Kantardzic
from
University of Louisville
Inspiration: Topographical Maps
1. Review document summary.
2. Consider clicking.
3. Click through and view web page.
4. Review next OR rewrite query OR end.
Problem 1: False Skips

• A user skips over a web page containing relevant information.
False Skip Effects

• Not noticeable to the user.
• Matters most in “needle in haystack” situation.
Problem 2: False Clicks

- Clicking on a search result only to find no relevant information.
False Clicks Effects

• Costs
  – Costs the user time.
  – Detracts from user’s experience.
  – Decreases chances of successful search.

• Preventable:
  – Presentation misrepresents content of web page.
Color-Coding Summaries

1. Color-coding query terms
   – Color by frequency.
   – Adding context information at no cost in summary size.

2. Flagging terms
   – Warn the user of topic divergence.
   – Most frequent term colored red when warning.
Color-Coding Example

Query Keyword Usage Extent by Color

   Software Engineering Radio

   Occupational Outlook Handbook (US Gov.)

Topic Departure Warning

A: “episode” – Most frequent page word but unusual for this search.
Query Keyword Coloring

1. Extract Keywords

   Query
   - building
   - a
   - database

2. Extract Web Page Terms

3. Count Term Frequencies

4. Select Color by Frequency

5. Highlight in Summary
Flagging Terms

1. Extract Terms

2. Count Term Frequencies

3. Count Documents Containing Top Term

4. Threshold Percent of Documents

5. Conditionally Flag Top Term in Summary
Text Selection Process of ReClose

- ReClose (Regression + Closeness Centrality)
- Combines 2 sentence rankings to generate summaries.
Experiment

• Compared the following summary types:
  – Google, ReClose and Color-Coded ReClose.

• Participants: 21 graduate and undergraduates in CECS from University of Louisville.

• Survey:
  – Online
  – Each participant saw all 3 summary types in random order.
### Survey Description (Step 1)

<table>
<thead>
<tr>
<th>Title Ignored</th>
<th>Would Click?</th>
<th>Expected Relevant Content</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>image compression</strong></td>
<td>* Data vs information * Data as a means to convey information * Data * Compression ratio CR = 32.7 = 1.11 ... ignored.url/path</td>
<td>None</td>
</tr>
<tr>
<td><strong>Title Ignored</strong></td>
<td>None</td>
<td>Sentences</td>
</tr>
<tr>
<td><strong>Title Ignored</strong></td>
<td>None</td>
<td>Sentences</td>
</tr>
<tr>
<td><strong>Title Ignored</strong></td>
<td>None</td>
<td>Sentences</td>
</tr>
<tr>
<td><strong>Title Ignored</strong></td>
<td>None</td>
<td>Sentences</td>
</tr>
<tr>
<td><strong>Title Ignored</strong></td>
<td>None</td>
<td>Sentences</td>
</tr>
<tr>
<td><strong>Title Ignored</strong></td>
<td>None</td>
<td>Sentences</td>
</tr>
</tbody>
</table>

Submit Expectations
## Survey Description (Step 2)

<table>
<thead>
<tr>
<th>Title Ignored</th>
<th>Page: ... Data vs information • Data as a means to convey information • Data ...</th>
<th>No</th>
<th>Sentences</th>
<th>link</th>
<th>None</th>
<th>Sentences</th>
<th>Paragraphs</th>
<th>Pages</th>
<th>Book</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword: ... 0.02 = 2.7bits • Compression ratio CR = 32.7 = 1.11 ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title Ignored</th>
<th>Page: Added to queue Lecture -18 DCT Quantization and Limitations by nptelhrd ...</th>
<th>No</th>
<th>Sentences</th>
<th>link</th>
<th>None</th>
<th>Sentences</th>
<th>Paragraphs</th>
<th>Pages</th>
<th>Book</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword: Prof Sengupta made image compression so easy ... Image ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title Ignored</th>
<th>Page: ... order σ1 ≥ σ2 ≥ ... ≥ σr &gt; σr+1 = · · · = σp = 0, Introduction to SVD ...</th>
<th>Yes</th>
<th>Pages</th>
<th>link</th>
<th>None</th>
<th>Sentences</th>
<th>Paragraphs</th>
<th>Pages</th>
<th>Book</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword: Close Quit Image Compression using Singular Value ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title Ignored</th>
<th>Page: In computer science and information theory, data compression or source ...</th>
<th>Yes</th>
<th>Pages</th>
<th>link</th>
<th>None</th>
<th>Sentences</th>
<th>Paragraphs</th>
<th>Pages</th>
<th>Book</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword: JPEG image compression works in part by “rounding off” some of this ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title Ignored</th>
<th>Page: all news &gt;&gt; LizardTech © Unveils GeoExpress 8 at GEOINT 2010 Symposium ...</th>
<th>No</th>
<th>None</th>
<th>link</th>
<th>None</th>
<th>Sentences</th>
<th>Paragraphs</th>
<th>Pages</th>
<th>Book</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword: ... lossless and lossy compression, is released at the 2010 ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Survey Results

- Color-coded ReClose summaries led to better click precision (fewer false clicks).

<table>
<thead>
<tr>
<th>Approach</th>
<th>Click Precision</th>
<th>Click Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>66% (38/58)</td>
<td>60% (38/63)</td>
</tr>
<tr>
<td>ReClose</td>
<td>75% (39/52)</td>
<td>64% (39/61)</td>
</tr>
<tr>
<td>Color-Coded</td>
<td>80% (49/61)</td>
<td>70% (49/70)</td>
</tr>
</tbody>
</table>

- No statistically significant difference.

- Click Precision = $|\{\text{Relevant}\} \cap \{\text{Clicked}\}| / |\{\text{Clicked}\}|$

- Click Recall = $|\{\text{Relevant}\} \cap \{\text{Clicked}\}| / |\{\text{Relevant}\}|$
Assessment of Color-Coding Keywords

• No noticeable difference found for counts from 0-59.
• 60+ keyword counts led to accurately higher expectations.
Assessment of Color-Coding Keywords

No noticeable difference found for counts from 0-59.

60+ keyword counts led to accurately higher expectations.

Google Summary for [logic gates]
Note that logic gates are not always required because simple logic functions can be ... A truth table is a good way to show the function of a logic gate. ...
Assessment of Color-Coding Keywords

Google Summary for [logic gates]
Note that logic gates are not always required because simple logic functions can be ... A truth table is a good way to show the function of a logic gate. ...

Color-Coded ReClose Summary for [logic gates]
Page: The symbols 0 (false) and 1 (true) are usually used in truth tables. ...
Keyword: Logic gates process signals which represent true or false. ...

- 66+ keyword counts led to accurately higher expectations.
- Logic occurs 20 times.
- Gates occurs 107 times.
Assessment of Flagging Terms

- ReClose vs. Color-Coded: nearly identical text.
- Flaggable – top page term was uncommon.
- Unflagged results led to higher click precision.

<table>
<thead>
<tr>
<th>ReClose</th>
<th>Click Precision</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flaggable</td>
<td>70%</td>
<td>24/53 (45%)</td>
</tr>
<tr>
<td>Not Flaggable</td>
<td>78%</td>
<td>37/52 (71%)</td>
</tr>
<tr>
<td>Color-Coded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flaggable</td>
<td>57%</td>
<td>25/52 (48%)</td>
</tr>
<tr>
<td>Not Flaggable</td>
<td>87%</td>
<td>45/53 (85%)</td>
</tr>
</tbody>
</table>
Color-Coded ReClose Summarized

• Additional useful information without expanding listing size.
• Color-coded ReClose summaries led to higher click precision.
• Each color-coding technique helped in specific situations.
• Contributions:
  – Two new techniques for color coding terms in search.
Color Blindness (vischeck.com)

Original

Deuteranope

Tritanope

Protanope