Guidelines for Effective Usage of Text Highlighting Techniques

Hendrik Strobelt, Daniela Oelke, Bum Chul Kwon, Tobias Schreck, Hanspeter Pfister

presented by Jordon Johnson
Many text vis tools...

http://textvis.lnu.se/
... but sometimes need to read text with annotations (WHY)

**bold font** and **yellow background**

same, shedding gallons of tears, until there was a large pool all round her, about four inches deep and reaching half down the hall. After a time she heard a little pattering of feet in the distance, and she hastily dried her eyes to see what was coming. It was the White Rabbit returning, splendidly dressed, with a pair of white kid gloves in one hand and a large fan in the other: he came trotting along in a great hurry, muttering to himself as he came, 'Oh! the Duchess, the Duchess! Oh! I won't she be savage if I've kept her waiting!' Alice felt so desperate that she was ready to ask help of any one; so, when the Rabbit came near her, she began, in a low, timid voice, 'If you please, sir—' 'The Rabbit started violently, dropped the white kid gloves and the fan, and skidded away into the darkness as hard as he could go. Alice took up the fan and gloves, and, as the hall was very hot, she kept fanning herself all the time she went on talking. 'Dear, dear! How queer everything is to-day! And yesterday things went on as usual. I wonder if I've been changed in the night? Let me think: was I the same when I got up this morning? I almost think I can remember feeling a little different. But if I'm not the same, the next question is, Who in the world am I? Ah, that's the great puzzle! And she began thinking over all the children she knew that were of the same age as herself, to see if she could have been changed for any of them. 'I'm sure I'm not Ada,' she said, 'for her hair goes in such long ringlets, and mine doesn't go in ringlets at all; and I'm sure I can't be Mabel, for I know all sorts of things, and she, oh! she knows such a very little! Besides, she's She, and I'm I, and—oh dear! how puzzling! It all is! I'll try if I know all the things I used to know. Let me see: four times five is twelve, and four times six is thirteen, and four times seven is—oh dear! I shall never get to twenty at that rate! However, the Multiplication Table does not signify: let's try Geography. London is the capital of Paris, and'

**extra spacing and italics**

, and seemed to her to wink with one of its little eyes, but it said nothing. 'Perhaps it doesn't understand English,' thought Alice; 'I dare say it's a French mouse, come over with William the Conqueror.' (For, with all her knowledge of History, Alice had no very clear notion how long ago anything had happened.) So she began again: 'Ouest ma chatte?' which was the first sentence in her French lesson-book. The Mouse gave a sudden leap out of the water, and seemed to quiver all over with fright. 'Oh, I beg your pardon!' cried Alice hastily, afraid that she had hurt the poor animal's feelings. 'I quite forgot you didn't like cats.' 'Not like cats!' cried the Mouse, in a shrill, passionate voice. 'Would YOU like cats if you were me?' 'Well, perhaps not,' said Alice in a soothing tone; 'don't be angry about it. And yet I wish I could show you our cat Dinah: I think you'd take a fancy to it if you could only see her. She's such a dear quiet thing; Alice went on, half to herself, as she swam lazily about in the pool, and she sits purring so nicely by the fire, licking her paws and washing her face—and she's such a nice soft thing to nurse—and she's such a capital one for catching mice—oh, I beg your pardon!' cried Alice again; for this time the Mouse was bristling all over, and she felt certain it must be really offended. 'We won't talk about any more if you'd rather not.' 'We indeed!' cried the Mouse, who was trembling down to the end of his tail. 'As if I would talk on such a subject! Our family always HATED cats: nasty, low, vulgar
Design study...-ish

• Elicits requirements from domain experts
  – separate interviews with 5 NLP experts
• Carries out user studies to evaluate techniques
• All evaluated techniques have been in use for decades
  – similar to a study of the relative effectiveness of different marks and channels
Requirements (WHAT)

Annotations can be:

- **statistical**
  - word length
- **syntactic**
  - parts-of-speech
- **semantic**
  - sentiment tags
- **structural**
  - page margins
- **domain-specific**
  - proper names

- **categorical**
- **ordered**
- **quantitative**
- **boolean**
- of any textual scope
- overlapping
Pop-out is key

Characters/words are marks that are fairly densely packed and regularly spaced, and that already make use of some visual channels.

To make highlighting detectable, need to maximize pop-out.
Common highlighting techniques (HOW)

<table>
<thead>
<tr>
<th>Technique</th>
<th>Use</th>
<th>Typical variations</th>
<th>Used in our studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Font color</td>
<td>c q</td>
<td>Saturation, luminance, hue</td>
<td>Red color (rgb (227, 26, 28))</td>
</tr>
<tr>
<td>Background color</td>
<td>c q</td>
<td>Saturation, luminance, hue</td>
<td>Bright yellow (rgb (255, 255, 50))</td>
</tr>
<tr>
<td>Underlined</td>
<td>c q</td>
<td>Styles, thicknesses</td>
<td>Single underline</td>
</tr>
<tr>
<td>Font size</td>
<td>- q</td>
<td>% increase</td>
<td>150% increase</td>
</tr>
<tr>
<td>Font style</td>
<td>--</td>
<td>Italics, subscript,...</td>
<td>Italics</td>
</tr>
<tr>
<td>Font weight</td>
<td>--</td>
<td>Font weight</td>
<td>bold font</td>
</tr>
<tr>
<td>Rectangular border</td>
<td>c q</td>
<td>Styles of border, lines, thickness</td>
<td>Single border</td>
</tr>
<tr>
<td>Spaced out font</td>
<td>- q</td>
<td>Letter spacing</td>
<td>5px spacing</td>
</tr>
<tr>
<td>Text shadow</td>
<td>--</td>
<td>Offset, intensity,...</td>
<td>CSS: text-shadow: 4px 4px 3px rgb(50, 50, 50);</td>
</tr>
<tr>
<td>Font family</td>
<td>(c)</td>
<td>Sans-serif, Times, Helvetica,...</td>
<td>—</td>
</tr>
<tr>
<td>CAPITALIZATION</td>
<td>--</td>
<td>Small caps, large caps</td>
<td>—</td>
</tr>
<tr>
<td>Strike through</td>
<td>--</td>
<td>True, false</td>
<td>—</td>
</tr>
<tr>
<td><em>Blinking</em></td>
<td>--</td>
<td>True, false</td>
<td>—</td>
</tr>
</tbody>
</table>

- Each technique can also encode boolean features (scope of paper limited to this consideration)
- 9 techniques used in user studies
3 User Studies

• Performed using Amazon Mechanical Turk
• Analysis techniques: ANOVA and Tukey HSD
• Unwanted variation
  – Individual difference: normalized each participant’s responses with respect to their performance range
  – Learning curve: discarded first trials in first study, added training trials in others
  – Fatigue effects: not observed
Study 1: Ranking Techniques

- **Goal**: rank techniques with respect to pop-out
- 673 words, 20 randomly highlighted
  - Find as many highlighted words as possible within a time limit
- 45 participants
- 3 trials per technique (27 trials total) per participant
  - trials ordered randomly
Study 1 - results

<table>
<thead>
<tr>
<th>Technique</th>
<th>Perf. Rank</th>
<th>Mean/StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>font size</td>
<td>A</td>
<td>0.86 (0.12)</td>
</tr>
<tr>
<td>border</td>
<td>A B</td>
<td>0.84 (0.14)</td>
</tr>
<tr>
<td>background</td>
<td>B C</td>
<td>0.78 (0.14)</td>
</tr>
<tr>
<td>red</td>
<td>C</td>
<td>0.76 (0.16)</td>
</tr>
<tr>
<td>bold</td>
<td>C</td>
<td>0.74 (0.15)</td>
</tr>
<tr>
<td>shadow</td>
<td>C</td>
<td>0.71 (0.15)</td>
</tr>
<tr>
<td>underlined</td>
<td>D</td>
<td>0.58 (0.18)</td>
</tr>
<tr>
<td>spacing</td>
<td>D</td>
<td>0.55 (0.23)</td>
</tr>
<tr>
<td>italic</td>
<td>E</td>
<td>0.15 (0.14)</td>
</tr>
</tbody>
</table>
Study 1 - discussion

Possible explanations of strong results:

• Increased font size: sticks out from cap line, fill white space
• Border: makes the target appear larger
• Colour: strong pop-out effect
  – background may outperform text colour because coloured area is larger
Study 1 - discussion

Possible explanations of weak results:

• Letter spacing: already a normal feature of text
• Italics: slanted character features already found in text
Study 2: Search with Distractor

- **Goal:** determine how different techniques (A,B) *interfere* when used in the same text
  - Is relative strength of techniques a factor?
- 20 highlighted words for each of A, B, A+B
  - must choose words highlighted only with A
- 30 participants
- All pairs of techniques tried (72 trials total) per participant
Study 2 - results

Weaker techniques did not expect improvements

<table>
<thead>
<tr>
<th>distractor technique --&gt;</th>
<th>fs</th>
<th>bo</th>
<th>bg</th>
<th>red</th>
<th>bold</th>
<th>sha</th>
<th>und</th>
<th>spa</th>
<th>it</th>
</tr>
</thead>
<tbody>
<tr>
<td>font size</td>
<td>-15.4</td>
<td>-10.1</td>
<td>-4.6</td>
<td>-74.8</td>
<td>-12.5</td>
<td>-33.5</td>
<td>-92.9</td>
<td>-62.1</td>
<td></td>
</tr>
<tr>
<td>border</td>
<td>-27.1</td>
<td>-6.3</td>
<td>-5.8</td>
<td>-8.8</td>
<td>-10.6</td>
<td>-66.4</td>
<td>-42.8</td>
<td>-59.9</td>
<td></td>
</tr>
<tr>
<td>background</td>
<td>-13.5</td>
<td>-16.0</td>
<td>-17.5</td>
<td>-6.8</td>
<td>-14.5</td>
<td>-26.1</td>
<td>-40.0</td>
<td>-50.0</td>
<td></td>
</tr>
<tr>
<td>red</td>
<td>-17.2</td>
<td>-9.7</td>
<td>2.7</td>
<td>-16.5</td>
<td>-19.9</td>
<td>-30.5</td>
<td>-39.4</td>
<td>-48.8</td>
<td></td>
</tr>
<tr>
<td>bold</td>
<td>-68.6</td>
<td>-15.5</td>
<td>0.3</td>
<td>3.3</td>
<td>-15.1</td>
<td>-21.1</td>
<td>-29.9</td>
<td>-43.2</td>
<td></td>
</tr>
<tr>
<td>shadow</td>
<td>-20.1</td>
<td>-10.4</td>
<td>-1.7</td>
<td>-1.3</td>
<td>-13.4</td>
<td>-65.4</td>
<td>-23.8</td>
<td>-73.3</td>
<td></td>
</tr>
<tr>
<td>underlined</td>
<td>-22.8</td>
<td>-25.5</td>
<td>3.0</td>
<td>7.3</td>
<td>-6.9</td>
<td>-10.6</td>
<td>-37.3</td>
<td>-40.4</td>
<td></td>
</tr>
<tr>
<td>spacing</td>
<td>-56.0</td>
<td>-45.3</td>
<td>-6.4</td>
<td>-4.5</td>
<td>-30.3</td>
<td>-21.8</td>
<td>-44.6</td>
<td>-97.3</td>
<td></td>
</tr>
<tr>
<td>italic</td>
<td>23.2</td>
<td>35.6</td>
<td>48.3</td>
<td>37.6</td>
<td>31.9</td>
<td>28.8</td>
<td>15.6</td>
<td>2.5</td>
<td></td>
</tr>
</tbody>
</table>
Study 2 - results

Fig. 8: Absolute performance values of Study 2 (referenced as Matrix M2).
## Study 2 - results

<table>
<thead>
<tr>
<th>Technique</th>
<th>Perf. Rank</th>
<th>Mean/StDev</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>border</td>
<td>A</td>
<td>0.67 (0.22)</td>
<td>-0.17 (-20%)</td>
</tr>
<tr>
<td>font size</td>
<td>A B</td>
<td>0.65 (0.25)</td>
<td>-0.21 (-24%)</td>
</tr>
<tr>
<td>background</td>
<td>A B</td>
<td>0.64 (0.19)</td>
<td>-0.14 (-18%)</td>
</tr>
<tr>
<td>red</td>
<td>A B</td>
<td>0.63 (0.20)</td>
<td>-0.13 (-17%)</td>
</tr>
<tr>
<td>bold</td>
<td>B C</td>
<td>0.62 (0.19)</td>
<td>-0.12 (-16%)</td>
</tr>
<tr>
<td>shadow</td>
<td>C</td>
<td>0.58 (0.22)</td>
<td>-0.13 (-18%)</td>
</tr>
<tr>
<td>underlined</td>
<td>D</td>
<td>0.51 (0.20)</td>
<td>-0.07 (-12%)</td>
</tr>
<tr>
<td>spacing</td>
<td>E</td>
<td>0.41 (0.20)</td>
<td>-0.14 (-25%)</td>
</tr>
<tr>
<td>italic</td>
<td>F</td>
<td>0.22 (0.14)</td>
<td>+0.07 (+47%)</td>
</tr>
</tbody>
</table>

Fig. 5: Performance rank of target highlighting with a distractor (Study 2). The column *Deviation* reports the Deviation of the Mean Score from Study 1 (Percentage Change of Mean Score from Study 1). See caption of Figure 3 for how to read the *Perf. Rank* column.
Study 3: Visual Conjunctive Search

• **Goal:** How strong is a combination of techniques (A,B) compared to each alone?
• 20 highlighted words for each of A, B, A+B
  – must choose only A+B
• 24 participants
• All pairs of techniques tried (36 trials total) per participant
Study 3 - results

results similar to study 2

<table>
<thead>
<tr>
<th>Font Size</th>
<th>bo</th>
<th>bg</th>
<th>red</th>
<th>bold</th>
<th>sha</th>
<th>und</th>
<th>spa</th>
<th>it</th>
</tr>
</thead>
<tbody>
<tr>
<td>fs</td>
<td>-16.4</td>
<td>-9.7</td>
<td>-9.2</td>
<td>-32.4</td>
<td>-15.4</td>
<td>-17.7</td>
<td>-56.3</td>
<td>-71.5</td>
</tr>
<tr>
<td>border</td>
<td>-13.7</td>
<td>-21.9</td>
<td>-5.6</td>
<td>-14.3</td>
<td>-32.5</td>
<td>-77.1</td>
<td>-13.9</td>
<td>-74.9</td>
</tr>
<tr>
<td>background</td>
<td>0.5</td>
<td>-13.2</td>
<td>-34.7</td>
<td>-8.1</td>
<td>-37.5</td>
<td>-42.7</td>
<td>-23.8</td>
<td>-64.7</td>
</tr>
<tr>
<td>red</td>
<td>3.5</td>
<td>4.5</td>
<td>-31.2</td>
<td>-26.5</td>
<td>-39.2</td>
<td>-34.3</td>
<td>-30.6</td>
<td>-69.4</td>
</tr>
<tr>
<td>bold</td>
<td>-13.9</td>
<td>-0.7</td>
<td>-2.6</td>
<td>-23.2</td>
<td>-28.2</td>
<td>-34.7</td>
<td>-35.2</td>
<td>-64.5</td>
</tr>
<tr>
<td>shadow</td>
<td>4.8</td>
<td>-12.0</td>
<td>-25.2</td>
<td>-30.0</td>
<td>-23.0</td>
<td>-78.3</td>
<td>-43.3</td>
<td>-105.0</td>
</tr>
<tr>
<td>underlined</td>
<td>20.6</td>
<td>-22.3</td>
<td>-6.1</td>
<td>-2.5</td>
<td>-5.6</td>
<td>-45.7</td>
<td>8.9</td>
<td>-97.5</td>
</tr>
<tr>
<td>spacing</td>
<td>0.0</td>
<td>25.4</td>
<td>12.7</td>
<td>5.5</td>
<td>-0.4</td>
<td>-11.0</td>
<td>13.6</td>
<td>-73.9</td>
</tr>
<tr>
<td>italic</td>
<td>70.1</td>
<td>68.8</td>
<td>68.3</td>
<td>66.6</td>
<td>66.7</td>
<td>56.7</td>
<td>59.9</td>
<td>52.6</td>
</tr>
</tbody>
</table>

Only underlined + spacing showed improvement over both individually
Study 3 - results

![Heatmap showing study 3 results](image)

**Fig. 9:** Absolute performance values of Study 3 (referenced as Matrix M3).
Guidelines

Scenarios:
• Only one feature should be highlighted
• Both features should have the same visibility; conjunctive visual search is not important
• Conjunction of features is more important than each individually
• One feature is significantly more important than the other
• Both features should have the same visibility; their conjunction should be easy to see
Only one feature

Choose a technique with strong pop-out

**Examples:**

- Font size
- Borders
- Yellow background
Same visibility; conjunction unimportant

Choose techniques with strong pop-out that do not significantly interfere with each other

Examples:
• Bold + yellow background
• Border + red
• Font size + yellow background
• Font size + border
Conjunction of features more important than each individually

Choose techniques that scored high in visual conjunction test

Examples:
• Border + red
• Font size + red
• Font size + yellow background
One feature significantly more important than the other

Choose techniques such that one has significantly higher pop-out

Examples:
- Yellow background + spacing
- Font size + underlined
- Border + italics
Same visibility, easy-to-see conjunction

Choose techniques with strong pop-out that do not significantly interfere with each other, whose conjunction is easy to see

**Examples:**

- Border + red
- Font size + yellow background
- Yellow background + bold
Discussion/Future Work

Increase **scope**

- Combinations of more than two techniques
- Include more techniques (e.g. different colour combinations)
- Include categorical/ordered/quantitative data
- Include tasks that require context/analysis
- Consider overlay visualizations
Comments/Critiques

• The guidelines for some scenarios are very similar, and multiple examples cover multiple scenarios
  – 3 studies for 5 scenarios
  – Some scenario refactoring would not be amiss
• I would have liked to see a larger scope
  – The authors don’t misrepresent the scope
  – A larger scope would be a lot more work
  – **BUT** a larger set of matrices might reveal more clusters to fit the scenarios better
Comments/Critiques

• I would have liked to see a statement of expected results, based on existing understanding of marks and channels
Are there any questions?