Presenting: Visual Readability Analysis: How to Make Your Writings Easier to Read

By: Daniela Oelke et al.
Originally at VAST 2010

Presented by Michael Shah
Duration: 10-15 minutes
November 7, 2012
Motivation: Helping people write to serve target readers

- Contextual coherency
- Avoiding using unknown vocabulary
- Find Misspellings
How to measure readability?

- **Statistical Properties**
  - Word Length/syllable count
  - Words per sentence
  - easy versus difficult words (based on syllable count, frequency of use, special rules, etc.)
  - Sentence Structure - Depth/Breadth of parse tree
  - Coherence (See [3])
Example of Flesch-Kincaid Readability Test

\[
206.835 - 1.015 \left( \frac{\text{total words}}{\text{total sentences}} \right) - 84.6 \left( \frac{\text{total syllables}}{\text{total words}} \right)
\]

<table>
<thead>
<tr>
<th>Score</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>90.0–100.0</td>
<td>easily understood by an average 11-year-old student</td>
</tr>
<tr>
<td>60.0–70.0</td>
<td>easily understood by 13- to 15-year-old students</td>
</tr>
<tr>
<td>0.0–30.0</td>
<td>best understood by university graduates</td>
</tr>
</tbody>
</table>

http://en.wikipedia.org/wiki/Flesch%E2%80%93Kincaid_readability_test
Author Presents: Semi-automatic Readability Analysis

● Automated algorithms predict what indicators to use for readability
  ○ 141 candidate readability features
● Redundancy is measured with correlation measures between features
  ○ Features that are most expressive are saved
● User then inspects
Figure 1: Correlation matrix of the features remaining after removing the ones with low expressiveness. As can be seen, some features are highly correlated to each other measuring the same aspect of readability.
Figure 2 - Normalizing Selected Features weights in readability score

- Overall, on a per sentence level we can get a readability score
VisRA

- Used to show not only what is hard to read, but why
- **Literature Fingerprinting** technique
  - Color is mapped to text features
- **Three Views Presented**
  - Corpus View - Gives Overview Fig. 3a
  - Block View - Sections with Scores Fig. 3b
  - Detail View - Per sentence view Fig. 3c
Figure 3 - 3 Levels of Granularity

Figure 3: Screenshot of the VisRA tool on 3 different aggregation levels. (a) Corpus View (b) Block View (c) Detail View. To display single features, the colormap is generated as described in section 3.4 and figure 2.
Figure 7 - Revising in Detail View

- Example Revisions from draft of this paper
- Each sentence is measured based on five readability metrics
Case Studies

- Revising authors paper
- Revising Chapters of a book
  - Shows off scalability features using fingerprinting/pixel map technique
- Comparative Analysis of documents
  - Election Agenda's of German Parliament
    ■ See Figure 8a
Revisions of Paper Workflow (Fig. 6)

Figure 6: Revision of our own paper. (a) The first four pages of the paper as structure thumbnails before the revision. (b) Detail view for one of the sections. (c) Structure thumbnails of the same pages after the revision.
Case Study: Document Comparison

Figure 8: Visual Analysis of 8 election agendas from the elections of the German parliament in 2009.

(c) Feature: Word Length
Thoughts

- Can we build a readability measure based on multiple sentences? How does this change our system?
- In the end, only five metrics are used, is this sufficient?
- Would have liked to see a stronger evaluation
Why I choose the paper

- Code Analysis
- Software Visualization

http://en.wikipedia.org/wiki/Parse_tree
http://www.idi.ntnu.no/~progspr/handouts/03-handouts.pdf
Software Visualization in the Large

Relation to Provenance

● How do users make revisions over time?
  ○ What is there thought process?
  ○ Can we have insight into how they explain their ideas?
  ○ We definitely have a visual tool that can be used to map the complexity of ideas expressed over time.

● It may be interesting to combine this tool and do a readability analysis on user annotations within other systems
Presenting: Visual Readability Analysis: How to Make Your Writings Easier to Read

By: Daniela Oelke et al.
Originally at VAST 2010

Presented by Michael Shah
Duration: 10-15 minutes
November 7, 2012