Summary Workbench
Unifying Application and Evaluation of Text Summarization Models

Shahbaz Syed  Dominik Schwabe  Martin Potthast

Summary Workbench provides a unified interface to support the development and evaluation of new summarization models.

**Tool Features**
- **Summarization** of long text, web pages, scholarly documents via SOTA models (English only)
- **Evaluation** of model predictions via lexical and semantic overlap based measures with on-demand visualizations
- **Interactive exploration** of any correlation between different evaluation measures
- **Simple integration of new models and measures** as self-contained and reproducible plugins

**Summary Workbench**

- **Summarization**
  - Summarize via multiple models
  - Instant visualization of lexical overlap with the document

- **Evaluation**
  - Evaluate using multiple measures
  - Quantitative evaluation
  - Visualize overlap with the source document or the reference summary
  - Qualitative evaluation

- **Interactive Exploration**
  - Interactive visualization of any correlation between different measures
  - Example: Chris Rock posts selfies after being pulled over three times in last seven weeks. Stopped by the cops again wish me luck, he posted this week.
  - Example: BART Score: -2.97  BLEU: 21.65

**Table**

<table>
<thead>
<tr>
<th>Model</th>
<th>BARTScore</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>BART</td>
<td>-2.97</td>
<td>21.65</td>
</tr>
<tr>
<td>MASS</td>
<td>35.1</td>
<td>10.4</td>
</tr>
<tr>
<td>T5</td>
<td>32.4</td>
<td>10.2</td>
</tr>
</tbody>
</table>

**Document Reference**

- Chris Rock tweets photos of three traffic stops in seven weeks. Stopped by the cops again wish me luck, he posted early Tuesday. Many African-Americans have long bemoaned the phenomenon of being pulled over. Blacks are about 30% more likely to be pulled over by police than whites.