Introduction

- Users don’t read privacy policies due to their lengthy nature and the presence of legal jargon.
- A brief overview (e.g. table of contents) can enhance the readability of privacy policies.
- Text in privacy policies is organized into different paragraphs based on the topicality.
- An overview can be constructed by generating a title for each of the paragraphs of the policy.

Dynamic Title Generation

- Dynamic title generation creates a content-tailored description of the paragraph.
- Search related tasks typically go through the entire document to find relevant material. Titles can facilitate a narrower search on relevant sections, thereby increasing the accuracy and speed of information retrieval tasks.

Challenges of Title Generation

- Modeling title generation as a supervised sequence to sequence learning task requires a paragraph-title dataset.
- Privacy policy text is highly cohesive, and it retains high semantic similarity between sentences making text generation difficult.

Contribution

- We generate a sequence of titles for privacy policy text to enhance the readability of privacy policies.
- We leverage the document structure of web privacy policies to create a paragraph-title dataset.
- We train neural deep learning based encoder-decoder models that generate titles for the given privacy policy text dynamically.

Dataset

Web privacy policies contain a header followed by a paragraph of text.

- Transformer model performs the best as it models the long-range dependencies well.
- Most of the errors are due to the presence of company-specific and lengthy titles.

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>ROUGE-1</th>
<th>Semantic Similarity</th>
<th>Fidelity</th>
<th>Fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline 1: Seq2Seq</td>
<td>34.61</td>
<td>30.32</td>
<td>39%</td>
<td>61%</td>
</tr>
<tr>
<td>Baseline 2: S2S+Attention</td>
<td>35.60</td>
<td>35.60</td>
<td>58%</td>
<td>87%</td>
</tr>
<tr>
<td>Transformer – Sentence</td>
<td>35.96</td>
<td>47.31</td>
<td>63%</td>
<td>90%</td>
</tr>
<tr>
<td>Baseline 2 – Paragraph</td>
<td>26.32</td>
<td>24.72</td>
<td>19%</td>
<td>67%</td>
</tr>
<tr>
<td>Transformer – Paragraph</td>
<td>28.01</td>
<td>36.31</td>
<td>48%</td>
<td>91%</td>
</tr>
</tbody>
</table>

Future Work

- Apply language model corrections to the decoder to improve the fluency of the outputs.
- Conduct a usability study of privacy policies with sections prefaced by generated titles.