A Dictionary of Wisdom and Wit: Learning to Extract Quotable Phrases

Michael Bendersky

University of Massachusetts Amherst

David A. Smith

Quotable Phrase
– A meaningful, memorable, and succinct statement that can be quoted without its original context
– But, not necessarily an already famous quote

Motivation
– Automatically extract quotable phrases for book promotion and advertisement

Our System
– Re-discover forgotten authors and understand what makes a phrase quotable

“There is life in a poet so long as he is quoted.”
Alfred Comyn Lyall, “Studies in Literature and History”

Quotable Phrase Extraction
22,000 English books

Sentence Segmentation
Unsupervised Filtering

Supervised Quotable Phrase Detection

Quotable Phrase Detection – PR Curve

Quotable Phrase Extraction

Sentence Segmentation

Unsupervised Filtering

Supervised Quotable Phrase Detection

Extracted Quotable Phrases

700,000 phrases

Quotability Feature Types
– Words – quote length, capitalization patterns, presence of universal quantifiers and abstract concepts in the quote
– Punctuation – presence of different punctuation types in the quote
– Parts of Speech – parts of speech patterns in the quote

Top Quotability Features

<table>
<thead>
<tr>
<th>Feature Description</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>abstract concepts</td>
<td>+91.64</td>
</tr>
<tr>
<td>universal quantifiers</td>
<td>+61.67</td>
</tr>
<tr>
<td>contains a verb in Past Participle</td>
<td>-60.34</td>
</tr>
<tr>
<td>POS sequences [VB IN PRP]</td>
<td>+39.71</td>
</tr>
<tr>
<td>POS sequences [PRP MD VB]</td>
<td>-38.78</td>
</tr>
<tr>
<td>adjectives</td>
<td>+37.71</td>
</tr>
<tr>
<td>POS sequences [DT NN VBD]</td>
<td>-36.88</td>
</tr>
<tr>
<td>verbs</td>
<td>+35.22</td>
</tr>
</tbody>
</table>

Authors/Books with the highest Quotability Index

1. Henry Drummond
2. Ella Wheeler Wilcox
3. S. D. Gordon
4. Andrew Murray
5. Ralph Waldo Emerson
6. Orison Swett Marden
7. Mary Baker Eddy
8. ‘Abdu’l-Bahá
9. John Hartley
10. Rabindranath Tagore