Displaying unbiased information in the Living Voters Guide

CSE 573 Project by Allison Obourn
The Living Voters Guide

- Provides information about ballot measures
- Promotes deliberation
Pro and Con Lists

- Site suggests pros and cons to users
- The pros and cons displayed affect what the user puts in their list
- Selection must be unbiased
Goals

- High level:
  - Suggest unbiased groups of points
  - Suggest points covering the topic space

- Specific:
  - Create a summary of the points in the system
  - Cluster the points and pick a point from each cluster

The intent of this initiative is good; but the specific requirements are rigid and unrealistic on the given time scale and current budget.

To comply with 1163, the state would need to hire three additional auditors and five additional fraud investigators that it can't afford.
My project

• Create similarity matrices based on different parts of the text

• Evaluate which works best

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.220446049250</td>
<td>0.177667721355</td>
<td>0.17278314860</td>
<td>0.163033491353</td>
<td></td>
</tr>
<tr>
<td>0.123060320609</td>
<td>0.2366009557440</td>
<td>0.188274539339</td>
<td>0.262898047091</td>
<td></td>
</tr>
<tr>
<td>0.131345771030</td>
<td>0.2693995340550</td>
<td>0.254049644135</td>
<td>0.140338427424</td>
<td></td>
</tr>
<tr>
<td>0.332347629498</td>
<td>0.2827661665480</td>
<td>0.1969316628130</td>
<td>0.251764726721</td>
<td></td>
</tr>
<tr>
<td>0.177667721355</td>
<td>-2.220446049250</td>
<td>0.164801292224</td>
<td>0.14693501793</td>
<td></td>
</tr>
<tr>
<td>0.137118495245</td>
<td>0.2374770003140</td>
<td>0.194451821588</td>
<td>0.296954501463</td>
<td></td>
</tr>
<tr>
<td>0.154140039915</td>
<td>0.247182626573</td>
<td>0.315748057909</td>
<td>0.12208753828</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Method

- Bag of words approach
  - Tf-idf: term frequency – inverse document frequency

- Inputs
  - A short list of stopwords removed
  - Synonyms added and a short list of stopwords removed
  - Only nouns and verbs considered in similarity metric
  - Synonyms added and only nouns and verbs considered in similarity metric
• Which text input causes Tf-idf to produce the similarity matrix closest to what a person would produce?

• Method: comparing resulting matrices with matrix generated by people voting for point similarities

• Visualize results by using the matrices to cluster the points
  o The generated clusters are very different
  o All of the results are bad!