TextViz

Vaidyanath Divya
Why TextViz?

- Text visualizations of large corpora are hard to comprehend
- Underlying text processing techniques play an important role in making the visual efficient
- It's nice to have a platform to visualize the text in various ways
Project Goal

Create a one stop efficient text visualization platform to:

- Gather basic text statistics
- Understand the text corpus
- Help data analysts get an idea of the text to further apply models - Text mining
Components

Data input from user (Html/CSS/JS)
- Single text document
- Multiple documents

Text processing (Python/Flask)
- Basic text statistics
- Summarize text
- Similarity between text documents
- Topic detection
Components continued . . .

Visual element (Html/CSS/JS)

- Word Clouds - With contextual component
- Occurrence of words in a text
- Cluster of topics (Bubble charts)
Existing techniques/ Prior work

- [http://www.chrisharrison.net/index.php/Visualizations/BibleViz](http://www.chrisharrison.net/index.php/Visualizations/BibleViz)
- [https://gramener.com/mahabharatha/](https://gramener.com/mahabharatha/)
Demo
<table>
<thead>
<tr>
<th>Date</th>
<th>Task</th>
<th>Deliverable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov 18 2013</td>
<td>Research / Initial design</td>
<td>storyboard / basic framework</td>
</tr>
<tr>
<td>Nov 21 2013</td>
<td>Complete research and build backend text processing algorithm</td>
<td>Initial backend algorithm in python</td>
</tr>
<tr>
<td>Nov 25 2013</td>
<td>Experiment with various visual displays</td>
<td>Initial frontend layout</td>
</tr>
<tr>
<td>Nov 30 2013</td>
<td>Sync frontend and backend</td>
<td>Initial complete tool</td>
</tr>
<tr>
<td>Dec 2 2013</td>
<td>Quality testing and modifications</td>
<td>Final model/tool</td>
</tr>
<tr>
<td>Dec 6 2013</td>
<td>Complete writeup</td>
<td>Project complete</td>
</tr>
</tbody>
</table>
Thank you

Any Questions ?