IDEAL Spreadsheet

Tony Ardura, Austin Burnett, Rex Lacy, Shawn Neumann
IDEAL is utilized to provide visualization aids for researchers making use of large web archive files. This information will be processed through Hadoop with a spreadsheet-like interface. Why would we do things this way?
How Will We Do This?

- Doing a Web Application.
  1. RESTful Web Service
  2. Front End Interface
  3. Integration with Hadoop / Pig / Solr
RESTful Web Service

Request Types:

1. GET
2. POST
3. PUT
4. DELETE
What tools will we use?
Front End Interface

- Using HTML / CSS, jQuery and javafx

Interface will consist of selectable archives
Filter options to restrict output
Selectable list items take user directly to result
An early idea for a filter tool

Addresses filter needs
But not client specs

Interface design is truly an iterative process.
More current version:
Web page of results
Ability to continue search
Ability to further refine

<table>
<thead>
<tr>
<th>Subject</th>
<th>People Talking</th>
<th>Images Related</th>
<th>Tweets Related</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floods, China</td>
<td>Obama</td>
<td></td>
<td><a href="https://twitter.com/GlobalFloodNews/status/440318944302231552">https://twitter.com/GlobalFloodNews/status/440318944302231552</a></td>
</tr>
</tbody>
</table>
Backend

- WARC files
- Apache Pig
- Concept extraction
Questions?