Information Visualization in IR

The next step beyond text-based interfaces

• **Visualizing Internal Document Substructure**
  • **Show where the content of interest resides**
    • If only a paragraph on page 37 is relevant, then we want to pinpoint this.

• **Distribution of relevant information in a document**
  • Aid to selecting documents
  • TextTiling
  • DotPlot

• **Visualizing Document/WebPage Clusters and Relationships**
  • Graphical Elements of Information Browsing Systems
TextTiling/TileBars Visualization of Term Distributions


Term Set 1: law legal attorney lawsuit
Term Set 2: network lan

TileBars

Regression testing handling hardware and software
Toll fraud includes related article on MCI Communications
In conversation Teleglobe Canada Inc Pres and CQ
Deregulation indicates a healthy satellite services
The last word letters to the editor letter to the editor
What's wrong with network licensing includes related
Letters letter to the editor
Protecting information now vital Law Viewpoint column
Letters O
Loose LIPS sink ships logical inferences per second
Document management eases file control market
Connectivity O
Document managers bring law and order Software
When users write their own applications O
Time lapse may scuttle Xerox claim includes related
Insider revisited AI Insider column
Vendors offering more remedies for file buildup do
Laser Lite Apple's new personal LaserWriters Hard
Hacker's handicap Michael Colvin MP has won su
TextTiling/TileBars Visualization of Term Distributions


Tool for visualizing terms or term-set matches within a document

Block density proportional to weighted term frequency (TF-IDF)
Fixed-length blocks (e.g. 1000 words), so total length indicates document length:
TextTiling/TileBars Visualization of Term Distributions

Doc. 1

Doc. 2
TextTiling/TileBars Visualization of Term Distributions

Overlap of Topics in this region

Brief segue to topic #2

Document #2 is shorter
### TextTiling/TileBars Visualization of Term Distributions

#### TileBars: Term Distribution in Information Access

<table>
<thead>
<tr>
<th>Min Overlap Span</th>
<th>Min Hits</th>
<th>Min Distribution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 2 4 6 8 10</td>
<td>0 10 20 30 40 50</td>
</tr>
<tr>
<td>2</td>
<td>0 2 4 6 8 10</td>
<td>0 10 20 30 40 50</td>
</tr>
<tr>
<td>4</td>
<td>0 2 4 6 8 10</td>
<td>0 10 20 30 40 50</td>
</tr>
<tr>
<td>6</td>
<td>0 2 4 6 8 10</td>
<td>0 10 20 30 40 50</td>
</tr>
<tr>
<td>8</td>
<td>0 2 4 6 8 10</td>
<td>0 10 20 30 40 50</td>
</tr>
<tr>
<td>10</td>
<td>0 2 4 6 8 10</td>
<td>0 10 20 30 40 50</td>
</tr>
</tbody>
</table>

#### Documents Within Constraints

- Patient medicine medical
- Test scan cure diagnosis
- Software program

```
"Paragon ports its 2-D imaging software to DEC’s AVS. (Paragon Imaging Inc.)"
"DOD health system costs break $1.1 billion barrier. (Composite Health Care Systems Inc.)"
"Computer graphics in medicine: from pictures to analysis. (Industry Prof..."
"LabVIEW 2.0 still just over the horizon. (National Instruments LabVIEW 2.0)"
"Hospital bridges islands of data. (includes related articles on growth of a..."
"VA automation means faster admissions. (US Department of Veterans Affairs)"
"Better ADP could cut VA delays 40%, officials say. (automatic data process..."
"Army tests prototype battlefield information system. (Army)
"The electric cadaver. (computerized anatomy lessons and digital dissection..."
```
TextTiling/TileBars Visualization of Term Distributions