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/TileBars
  - DotPlot

- **Visualizing Document/WebPage Clusters and Relationships**
  - Graphical Elements of Information Browsing Systems
DOT PLOT (addl points)

To Show:
- Repetition
- Regions of self similarity (same **TOPIC** blocks)

![Diagram of a dot plot showing the following:
- Diagonal lines indicate repetition.
- Dense blocks indicate self-similarity.
- The main diagonal is highlighted.](image)
Figure 1. Dotplot Browser.
Motivation for Block Formation

-- CS466 Lecture XXI --
Figure 6. Dense Versus Sparse Features.
Figure 2. Features in Synthesized Dotplots.
Figure 3. Dotplot of Two DNA Sequences (7,000 Nucleotides).
Figure 4. Four AP News Stories (3,000 words).
Region Mapping (Image Compression)

F: max [cell[i], cell[j]] +=

Color or Blue
Compressing to single pixel value

Dot represents 1000x row character block

Cut off threshold for Display if Black or White

-- CS466 Lecture XXI --
Figure 7. Combination of Diagonals and Squares.
Figure 5. Three Years of Hansards (37 million words).
Figure 14. Computing a Hansard Plot With Different Values of T.
Figure 8. Six Chapters of Microsoft Manuals in Seven Languages (3.3 Million Words) With Color Map.
Figure 9. Three Thousand Lines of Code.
Figure 10. Six Hundred Lines of Code (detail of Figure 1).
Figure 11. Shrinking Diagonals.
Figure 12. Three Thousand Four Hundred Lines of Code.
Figure 13. Three Thousand Four Hundred Author Attributes.
Figure 15. Histogram of Values in Figure 9's f-Image.
Figure 1: The dotplot of four concatenated Wall Street Journal articles.
Figure 2: The outside density plot of the same articles.
Document Segmentation via DotPlot Approaches

Figure 1: The dotplot of four concatenated *Wall Street Journal* articles.

Figure 2: The outside density plot of the same articles.
Figure 1: The dotplot of four concatenated *Wall Street Journal* articles.

Figure 2: The outside density plot of the same articles.
Document Segmentation via DotPlot Approaches

Figure 1: The dotplot of four concatenated *Wall Street Journal* articles.

Figure 2: The outside density plot of the same articles.
Document Segmentation via DotPlot Approaches

Figure 1: The dotplot of four concatenated *Wall Street Journal* articles.

Figure 2: The outside density plot of the same articles.
Document Segmentation via DotPlot Approaches

Predicted and actual document region boundaries.
Predicted = local minima
Actual = vertical lines.