Categorization & Visualization: The Basics

• Ramana Rao, CTO & SVP
• Inxight Software, Inc

www.ramanarao.com
www.inxight.com
Beyond Search & Browse

- **Search**
  - Precise, but brittle ... leaves users searching, not finding ...

- **Browse**
  - Robust, but vague ... leaves users wandering & lost, not found ...

- **Opportunity is to blend Browsing & Search**
  - Categorization
  - Information Extraction
  - Information Visualization
Automatic Categorization

What:
Subject Categorization classifies textual documents into categories based on what they are about.

Why:
Increases Efficiency & Effectiveness of Systems and People that utilize the content
Information Extraction

Information extraction is about pulling elements out of documents and collections that guides the more intelligent use of content.

Often characterized as metadata that provides context.

Types of metadata include:
- Noun phrases
- Named entities (e.g., people, companies, places, products)
- Key sentences
- Concepts and topic relationships
- Similarity between documents, paragraphs and phrases
Optimism that Wall Street is indeed emerging from its slump sent technology stocks higher Thursday, adding to the previous session's triple-digit surge. Blue chips struggled to keep up, fluctuating in light profit-taking. ``The market's beginning to buy the scenario that the interest rate cuts by the Federal Reserve are going to help,'' said Gregory Nie, technical analyst at First Union Securities.
Information Visualization

♦ The Role of Content Visualization
  • Provides maps of large content spaces … and also the means for getting to specific documents or items
  • Thus support early or organizing processes like orientation, assessing, survey, etc. … as well as tune very focused processes like direct walk navigation

♦ Nature of the Solution
  • Leverage our visual/spatial skills
  • Like browsing, but shows much more, maps not just pages
  • Can eliminate mechanical overheads of browsing
  • Can integrate with searching more tightly

♦ Two key types of Content Visualizations
  • Content Terrain Maps
  • Wide Widgets
Content Terrain Maps

- Analogy to geographic maps
- Organize on 2-d surface with Terrain created by contours, colors, regions, based on:
  - Human Design
  - Categorization
  - Automatic Clustering based on Document Similarity
  - Metadata
Wide Widgets

- high bandwidth widgets for interacting with large collections
- arranged on a spine
  - Hierarchical Cone Tree, Spiral Calendar, Hyperbolic Tree Browser
  - Temporal Perspective Wall, Time Lens
  - Pages - Document Lens, Web Books
  - Calendars - Spiral Calendar
  - Tabular - Table Lens, Time Lattice
To be continued ...

- rao@inxight.com
  - Don’t hesitate to write ...
- www.ramanarao.com
  - Papers from talks
  - Information Flow newsletter
- www.inxight.com
  - White papers
  - Demos & free downloads