Memoplex (Browser)++: A Semantic Document Browser

Clarence Chan
CPSC 533
Nov. 12, 2006

The problem

- Directed search: find docs easily if
  - we have the document title
  - we know exactly what keywords are involved
  - ***We know what we're looking for***
- ***What if we don't exactly know?***
  - ***look through a list of titles***
  - ***"feel around", browse***

Navigation, browsing

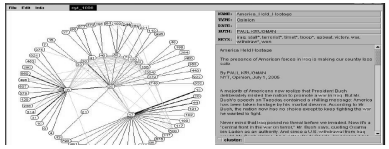
- When we don't know what we want ...
  - we start off with an inexact query
  - we look around the vicinity
  - ***We browse***
- ***Useful for learning about an area***
  - ***orient yourself w.r.t. to a local landmark***
  - ***browsing: relative navigation***
  - ***contrast with searching (absolute)***

Navigation, browsing

- Analogy:
  - Directed search: Look up exact item in library call number system
  - Browsing: Go to a shelf and look around at various books
- What if we combine these?
  - Go to a bookshelf in a general call number range
  - Browse around various titles and books that are related

Memoplex Browser, v.1

- Originally a 533 project from a previous year
  - Built on top of Mike Huggett's Memoplex server
  - Backend: Semantic network of documents
  - Frontend: node-link graph of network
  - Nodes are individual documents



Memoplex Browser, v.1: Existing issues

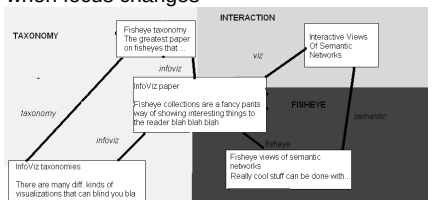
- Doesn't visualize the actual documents
- Too many nodes at once
- Meaning of edges is unclear (what concept links two documents?)
- *Isn't a Google search easier and more effective?*

Memoplex (Browser)++: Proposed solution

- Visualize just a few nodes at a time
  - Important to see document text in browsing!
- Cluster documents along diff. dimensions
- Colour + spatial re-alignment of connected nodes
- Arguably, Google search is better ...
  - Run a study!

Memoplex++: Proposed solution

- Show only elements one hop away
- fade adjacent nodes in and out as necessary when focus changes



Memoplex++: Current progress

- Document corpus acquired, basics of old Memoplex browser understood
  - Throwing out much of GUI, starting from scratch
- Basic node filtering, labeling with document text accomplished
  - Not very elegant: recalculates every time, have to figure out Prefuse expression syntax
- Working on smart way of rotating edges into appropriate position based on existing clusters

Memoplex++: Current progress

- Figuring out what to do with documents I want to hide
  - Show edges, but not node text?
  - Smaller nodes
    - Again, must learn filtering language
    - Or just hard-code?
- "Clustering" documents according to keywords
  - Writing naïve algo to parse keywords
  - What to do when no keywords are present?