Interactive visualization for opportunistic exploration of large document collections

Simon Lehmann, Ulrich Schwanecke, Ralf Doerner

Information Systems 35 (2010)

PAWS meeting
4/6/2010

Motivations

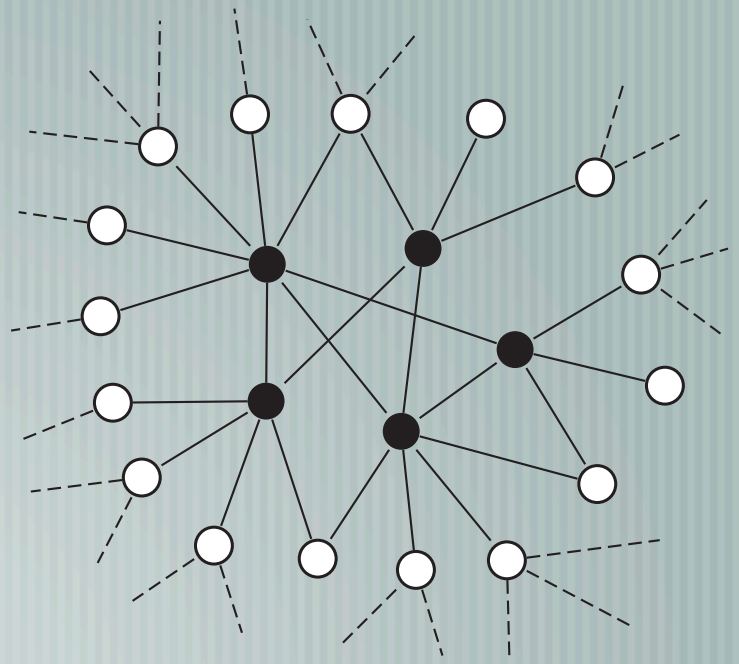
- Visualize...
  - Complex structure of highly cross-referenced articles
- Help users to opportunistically explore the information
  - Choosing potentially interesting articles
  - Highlight more important articles

Opportunistic exploration

- Exploration
  - Start with already known information
  - Read initial search/browsing result
  - Learn about the subject encounter additional cross-reference
  - New navigational opportunities

Wivi – Idea

- Visualize Wikipedia navigation
- Start from the first article
- Show navigational tips
  - To the articles linked from the first section
    - Of visited articles



Wivi – Degree of Interest of unvisited articles

- Relative Degree of Interest (DOI)
  - Based on the history of the article graph
- A-priori-importance (API)
- Distance between article and the current focus (D)
- DOI = API - D

Wivi – Degree of Interest of unvisited articles

- API – More inbound links from already visited articles → more important

$$API(v) = \frac{d_G(v)}{\Delta(G)}$$

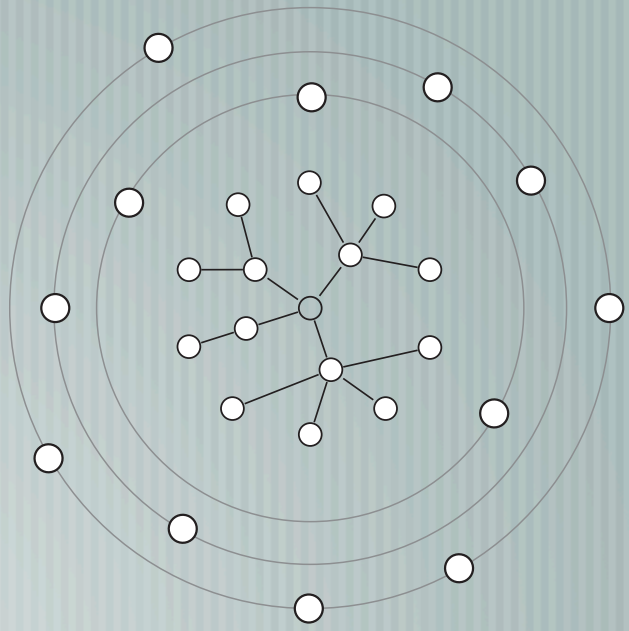
- D – temporal distance; by the number of articles a user has visited since the last visit to that article

$$D(v) = \frac{1}{d_G(v)A(G)} \sum_{v_i \in N_G(v)} a(v_i)$$

- For every unvisited vertex *v* of the graph *G*, the *DOI* function assigns a degree of interest to that vertex depending on the already visited vertices [-1,1]

Wivi – Goal of visualization

- Provide a representation of
  - The previously visited articles
  - Their connections to the current article
  - Show navigational hint/ possible future



Wivi – Goal of visualization

- Avoid clutter – only the edges connected to the currently read article are shown
- Show texts too – Users do not want to use graphical representations alone for navigation
- Transition animation – efficiently perceive the changes and maintain their mental map

Demo

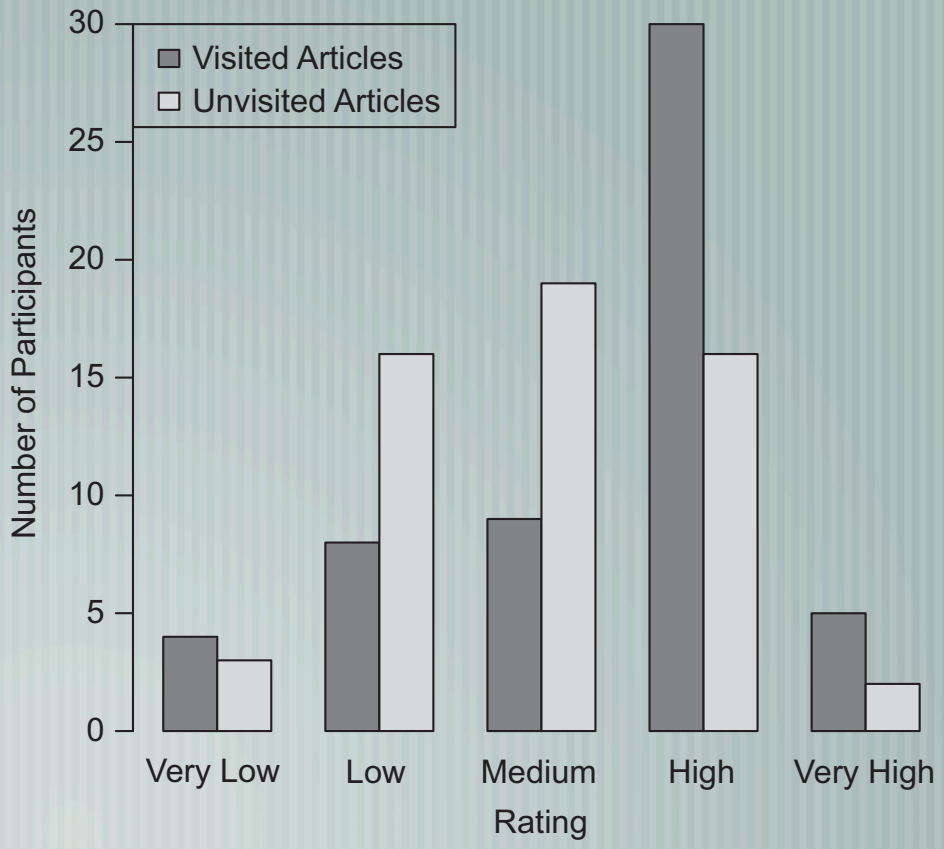
- http://wivi.slashslash.de

Evaluation

- Anonymous remote usability test
- Task – (1) Search for subject subjects are interested (2) random selection
- Allowed to perform tasks up to 5 times
- 14 days, 157 people
- 56 people remained after filtering
  - Spent more than 100s
  - Finished the test by filling in the exit questionnaire
- Subjective evaluation only

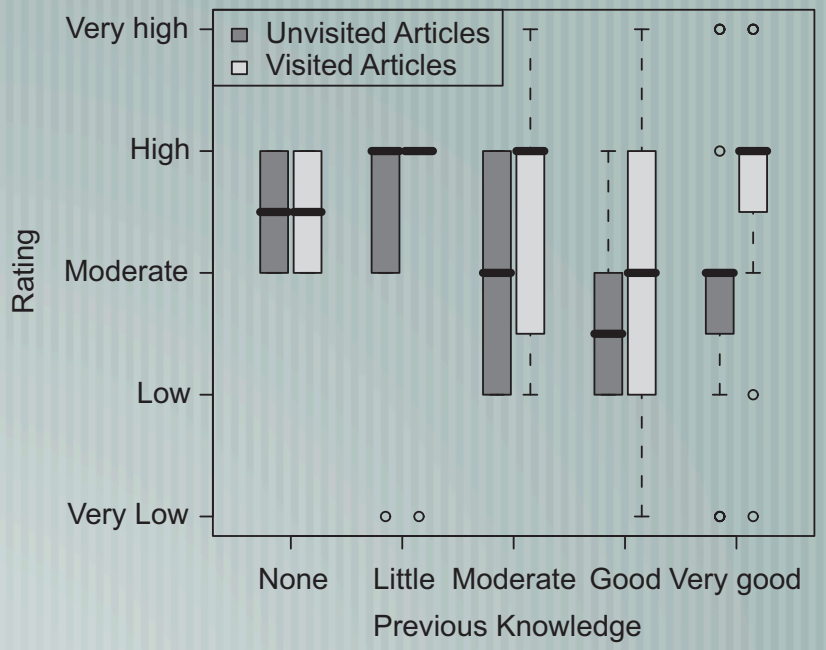
Results

- 69.6% of the participants – easy to understand and use
- Unvisited articles were not easily found
- Invisibility of edges (?)



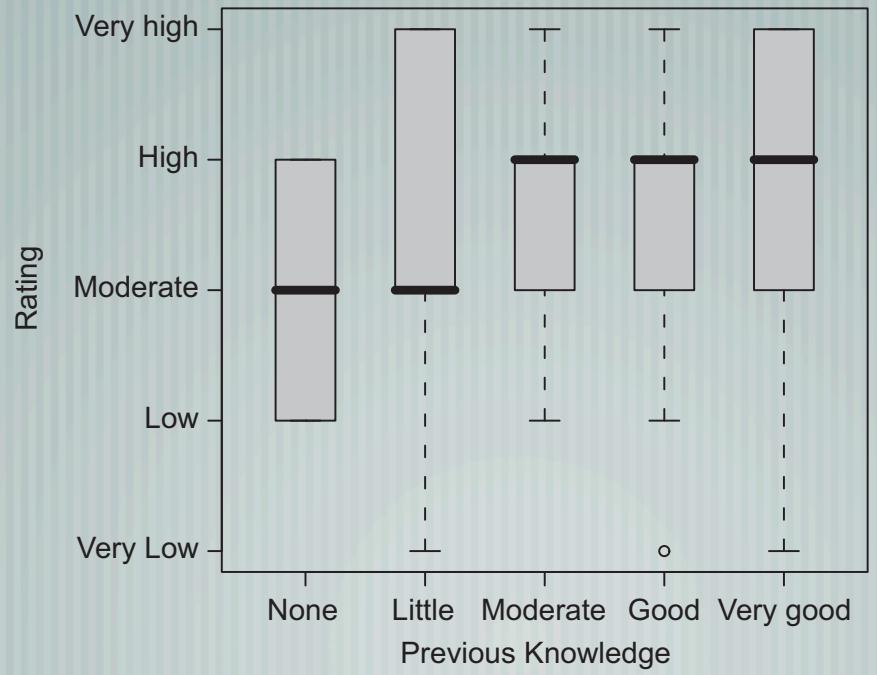
Results

- Separate subjects by previous knowledge – K(low), K(high)
- K(low) – rating between visited/ unvisited – similar
- K(high) – visited article – higher rating
- Wivi suitable for researching a new subject



Results

K(high) liked Wivi more



Conclusions

- Opportunistic exploratory visualization
- Personalized navigation (?) based on Wikipedia link structure and visited articles
- Interest-based

(1) InfoSky: A system for visual exploration of very large, hierarchically structured knowledge spaces
(2) Evaluating Information Visualization

Wolfgang Kienreich, et al.

Proceedings der GI Workshopwoche LLWA-Workshop der Fachgruppe FGWM (2003)

Keith Andrews

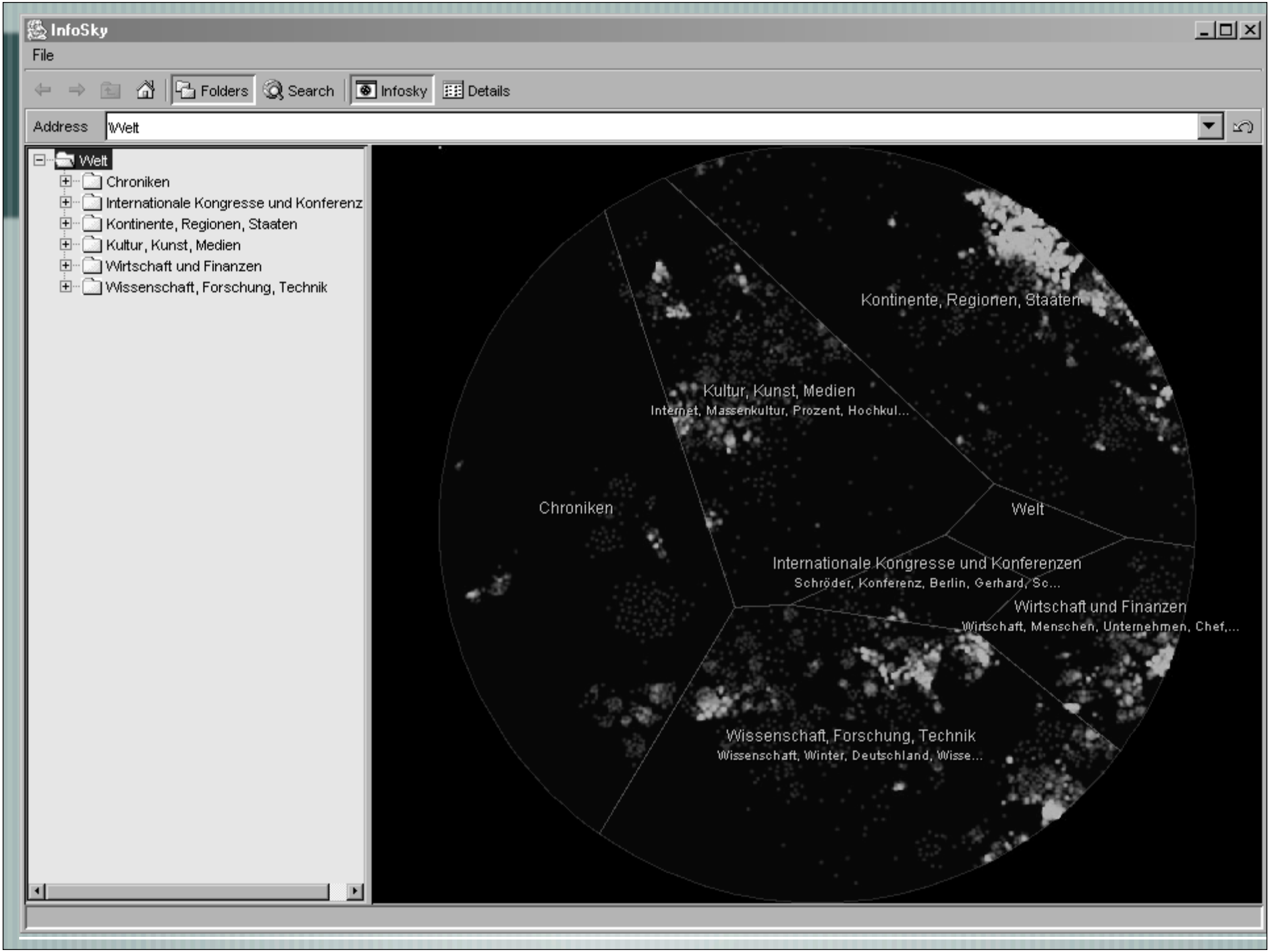
AVI 2006 BELIV Workshop

Motivations

- Very large, hierarchically structured document visualization
- Requirements
  - Scalability
  - Hierarchy plus similarity
  - Focus plus context
  - Query plus exploration

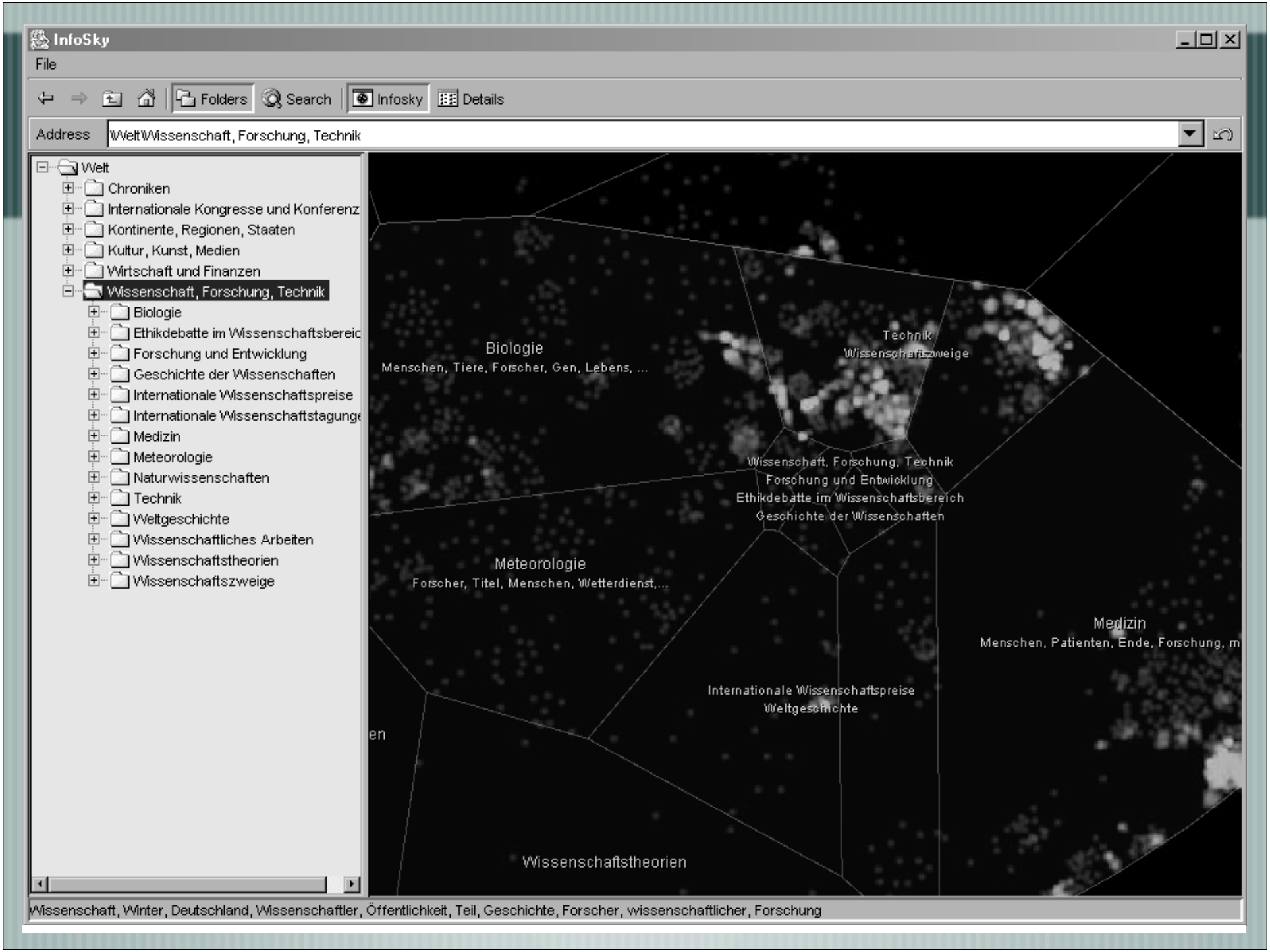
InfoSky

- KnowledgeScope
- Galaxy and telescope metaphor
- Similarity-based visualization – document layout and clusters
  - Documents – stars; Force-directed layout
  - Collection (cluster) – constellation; partitioning


InfoSky
File
Folders
Search
Infosky
Details
Address
Welt
Welt
Chroniken
Internationale Kongresse und Konferenz
Kontinente, Regionen, Staaten
Kultur, Kunst, Medien
Wirtschaft und Finanzen
Wissenschaft, Forschung, Technik
Kontinente, Regionen, Staaten
Kultur, Kunst, Medien
Internet, Massenkultur, Prozent, Hochkul...
Chroniken
Welt
Internationale Kongresse und Konferenzen
Schröder, Konferenz, Berlin, Gerhard, Sc...
Wirtschaft und Finanzen
Wirtschaft, Menschen, Unternehmen, Chef,...
Wissenschaft, Forschung, Technik
Wissenschaft, Winter, Deutschland, Wisse...

Hierarchy & structure

- Navigation
  - Zooming – navigate vertically within the hierarchy
  - Panning – explore across a single level
- Collections – similar ones are placed close to each other
  - Polygons are partition – Voronoi diagram
- Tree view is synced

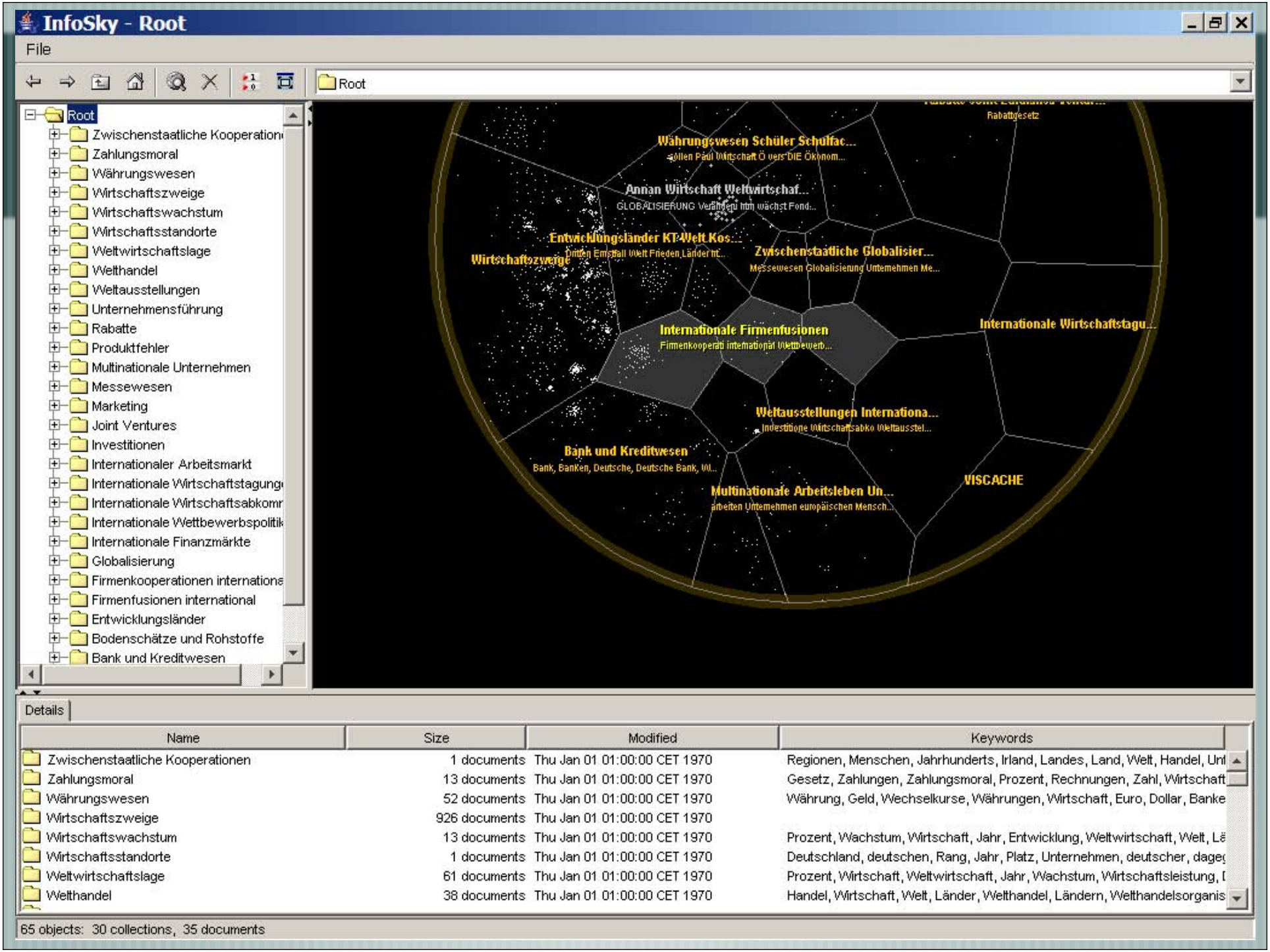

InfoSky
File
Folders
Search
Infosky
Details
Address
Welt\Wissenschaft, Forschung, Technik
Welt
Chroniken
Internationale Kongresse und Konferenz
Kontinente, Regionen, Staaten
Kultur, Kunst, Medien
Wirtschaft und Finanzen
Wissenschaft, Forschung, Technik
Biologie
Ethikdebatte im Wissenschaftsbereic
Forschung und Entwicklung
Geschichte der Wissenschaften
Internationale Wissenschaftspreise
Internationale Wissenschaftstagunge
Medizin
Meteorologie
Naturwissenschaften
Technik
Weltgeschichte
Wissenschaftliches Arbeiten
Wissenschaftstheorien
Wissenschaftszweige
Biologie
Menschen, Tiere, Forscher, Gen, Lebens, ...
Technik
Wissenschaftszweige
Wissenschaft, Forschung, Technik
Forschung und Entwicklung
Ethikdebatte im Wissenschaftsbereich
Geschichte der Wissenschaften
Meteorologie
Forscher, Titel, Menschen, Wetterdienst,...
Medizin
Menschen, Patienten, Ende, Forschung, m
Internationale Wissenschaftspreise
Weltgeschichte
en
Wissenschaftstheorien
Wissenschaft, Winter, Deutschland, Wissenschaftler, Öffentlichkeit, Teil, Geschichte, Forscher, wissenschaftlicher, Forschung

Data and implementation

- 109,000 German language news articles
- 6,900 collections and sub-collections (manual)
- 15 levels deep

Evaluation

- Formative Testing (Think aloud)
- Formal Experiment (2002)
  - Users preferred tree view (familiarity)
    - InfoSky was significantly slower
- Experiment (2004) – Version 2
  - No significant difference; improved (?)


InfoSky - Root
File
Root
Root
Zwischenstaatliche Kooperation
Zahlungsmoral
Währungswesen
Wirtschaftszweige
Wirtschaftswachstum
Wirtschaftsstandorte
Weltwirtschaftslage
Welthandel
Weltausstellungen
Unternehmensführung
Rabatte
Produktfehler
Multinationale Unternehmen
Messewesen
Marketing
Joint Ventures
Investitionen
Internationaler Arbeitsmarkt
Internationale Wirtschaftstagung
Internationale Wirtschaftsabkomr
Internationale Wettbewerbspolitik
Internationale Finanzmärkte
Globalisierung
Firmenkooperationen internationa
Firmenfusionen international
Entwicklungsländer
Bodenschätze und Rohstoffe
Bank und Kreditwesen
Rabattgesetz
Währungswesen Schüler Schulfac...
Annan Wirtschaft Weltwirtschaf...
Entwicklungsländer KT-Welt Kos...
Wirtschaftszweige
Zwischenstaatliche Globalisier...
Internationale Firmenfusionen
Internationale Wirtschaftstagu...
Weltausstellungen Internationa...
Bank und Kreditwesen
Multinationale Arbeitsleben Un...
VISCACHE
Details
Name
Size
Modified
Keywords
Zwischenstaatliche Kooperationen 1 documents Thu Jan 01 01:00:00 CET 1970 Regionen, Menschen, Jahrhunderts, Irland, Landes, Land, Welt, Handel,
Zahlungsmoral 13 documents Thu Jan 01 01:00:00 CET 1970 Gesetz, Zahlungen, Zahlungsmoral, Prozent, Rechnungen, Zahl, Wirtschaft
Währungswesen 52 documents Thu Jan 01 01:00:00 CET 1970 Währung, Geld, Wechselkurse, Währungen, Wirtschaft, Euro, Dollar, Banke
Wirtschaftszweige 926 documents Thu Jan 01 01:00:00 CET 1970
Wirtschaftswachstum 13 documents Thu Jan 01 01:00:00 CET 1970 Prozent, Wachstum, Wirtschaft, Jahr, Entwicklung, Weltwirtschaft, Welt,
Wirtschaftsstandorte 1 documents Thu Jan 01 01:00:00 CET 1970 Deutschland, deutschen, Rang, Jahr, Platz, Unternehmen, deutscher,
Weltwirtschaftslage 61 documents Thu Jan 01 01:00:00 CET 1970 Prozent, Wirtschaft, Weltwirtschaft, Jahr, Wachstum, Wirtschaftsleistung,
Welthandel 38 documents Thu Jan 01 01:00:00 CET 1970 Handel, Wirtschaft, Welt, Länder, Welthandel, Ländern, Welthandelsorganis
65 objects: 30 collections, 35 documents

Evaluation

- In 2006, 32 test users
- Compares 4 hierarchical methods
  - Tree view, information pyramid browser, tree map browser, hyperbolic browser
- Using automated HVTE testing environment
- 8 tasks divided into overview (2), search (2), count (2), compare (2) tests

HVTE - Hierarchical Visualisation Testing Environment

[1] HVS Test
User 1
User 2
User 3
User 4
User 5
User 6
User 7
User 8
User 9
User 10
User 11
User 12
User 13
User 14
User 15
User 16

Frage

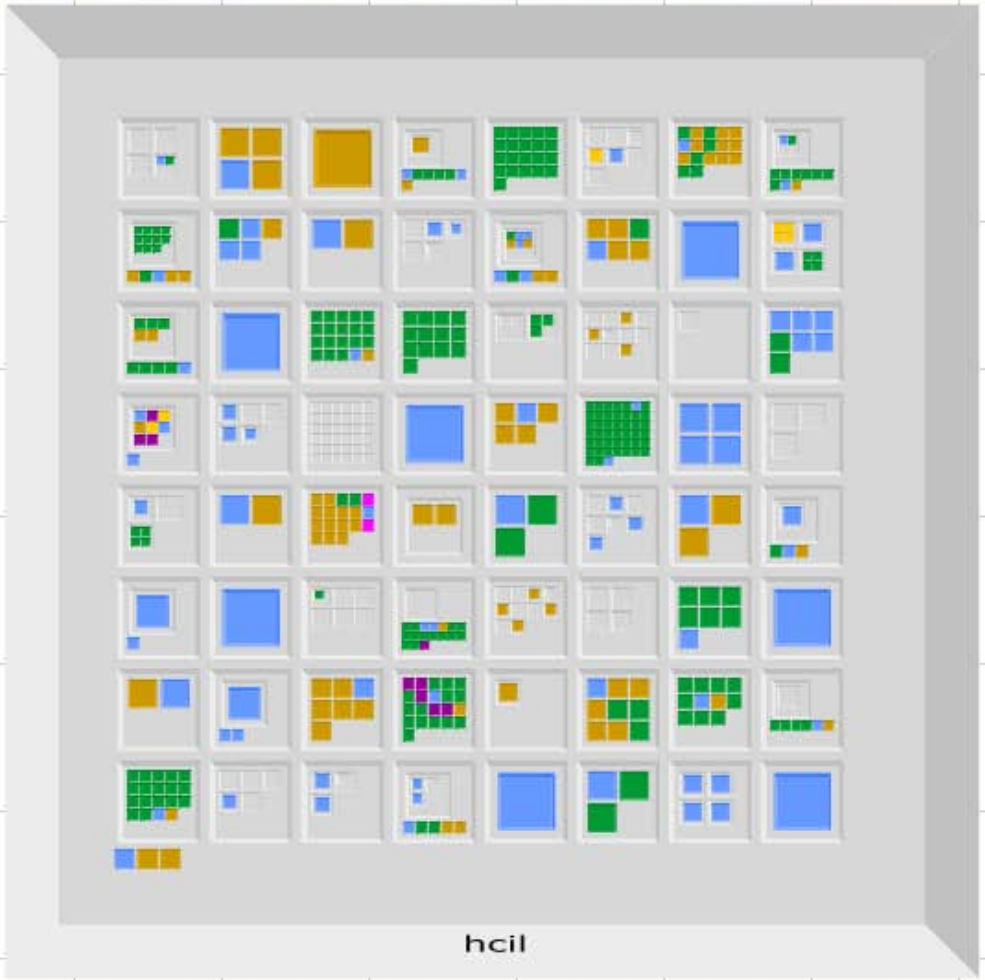
Aufgabe #9: Überblick

Aufgabe:
Finden Sie im Ordner "treemaps" (/hcil/treemaps) den tiefsten Unterordner. Schreiben Sie den Namen des Ordners in das Antwortfeld rechts und drücken Sie anschliessend "Weiter...".

Antwort

Weiter.....

Information Pyramids



Evaluation

- No difference between 4 systems
- Except, treemap was significantly faster than hyperbolic browser (counting test)
- Subjects preferred tree view method significantly to others

Conclusion

- Even if performance data show no significant differences, users significantly prefer the tree view
- Users will apparently need a great deal of persuading to move from a familiar trusted interface to a new, unfamiliar one
- **Suggested Strategies**
  - Looking for experienced analysts rather than CS students (subjects)
  - Providing extensive training
  - Formulating more involved tasks