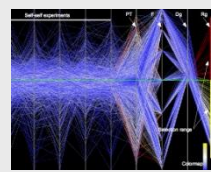


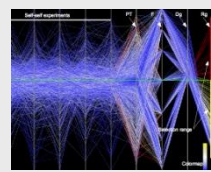
# Informations- visualisierung

Thema:	6. Darstellung von Metadaten und Prozessen
Dozent:	Prof. Dr. Geric Scheuermann scheuermann@informatik.uni-leipzig.de
Sprechstunde:	nach Vereinbarung
Umfang:	2
Prüfungsfach:	Modul Fortgeschrittene Computergraphik Medizininformatik, Angewandte Informatik



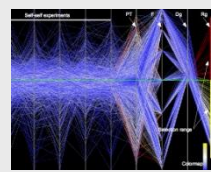
# Wiederholung

- Was ist ein Graph?
- Was heißt adjazent und was ist der Grad eines Knoten?
- Was beschreibt ein Graph/Warum ist er für uns wichtig?
- Was ist ein Zyklus?
- Was ist ein Digraph und was ist nun der Grad eines Knoten?
- Was ist ein DAG?
- Was ist eine Zeichnung, wann ist sie planar?
- Wann ist ein Graph planar?
- Was ist ein Dualgraph?



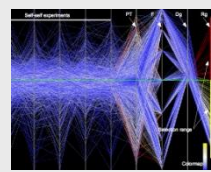
# Wiederholung

- Wann ist ein Graph zusammenhängend?
- Was sind Zusammenhangskomponenten?
- Was sind Schnittknoten und Blöcke?
- Was sind Zeichenkonventionen?
- Was sind Ästhetikkriterien?
- Was ist das Problem dabei?
- Was passiert bei dem Topologie-Form-Metrik-Ansatz?
- Was ist beim hierarchischen Ansatz anders?
- Wie heißen weitere Ansätze?



# Übersicht

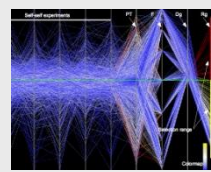
1. Einführung
2. Wahrnehmung von Graphik
3. Statistische Grundlagen
4. Darstellung von Tabellen
5. Darstellung von Graphen
- 6. Darstellung von Metadaten und Prozessen**
7. Interaktion
8. Spezifische Verfahren
9. Visual Analytics
10. Beispielanwendungen



## 6.1 Dokumente

### Visualisierung von Dokumenten

- Ein oder mehrere Textdokumente
- **Einsicht** in die Texte gewinnen
- Textdateiteile oder ganze Textsammlungen werden **in Bilder** umgewandelt
- Beispiele für Dokumentensammlungen sind
  - Digitale Bibliotheken
  - Textdokumente im WWW
  - Zeitungsartikel
  - Wissenschaftliche Artikel
  - Zeugenaussagen in einem Prozess
  - Patente



## 6.1 Dokumente

### Szenarien der Dokumentenvisualisierung

- **Suche(Query):**

**Welche Dokumente** sind wichtig für mein Thema?

Welche anderen Dokumente **sind so nah** an meinem Interessensgebiet, dass ich sie auch ansehen sollte ?

Wie sind meine **Schlüsselworte** in den Dokumenten tatsächlich **verteilt**?

- **Durchsicht (Browsing):**

**Welche Themen** werden in den zugänglichen Dokumenten behandelt?

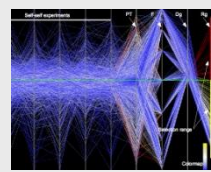
Welche Themen stehen **im Vordergrund**?

Wie ist der **Zusammenhang** der behandelten Themen?

- **Lesen:**

**Welche Teile** des Dokuments soll ich mir ansehen?

*„So much has been written about everything  
that you can not find out anything about it.“ [J. Thurber]*



# 6.1 Dokumente

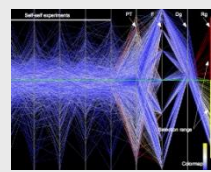
## Dokumentenvisualisierung und Information-Retrieval

### Klassisches Information-Retrieval

- **Unabhängiger Agent** sucht Informationen aufgrund **klarer Vorgaben** und bringt sie dem an der Suche unbeteiligten Anwender

### Dokumentenvisualisierung

- Antwort auf **unklar formulierte** Suchanfrage durch
  - Generierung einer **präzisen Anfrage und Antwort** aus
    - groben **Überblick** über die Dokumente
    - **Detailuntersuchungen** des Anwenders

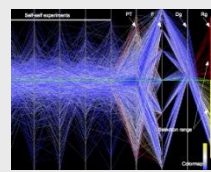


# 6.1 Dokumente

## Vorgehensweise

- Dokumentenvisualisierung idR. in drei Phasen
- **Analyse**
  - Berechnung einer **Statistik 1. Ordnung**: Häufigkeit der Wörter (u. U. mit höherem Gewicht für Überschriften/Zusammenfassungen)
  - Statistiken **höherer Ordnung** bei mehreren Dokumenten
  - **Hochdimensionaler** Vektor für jedes Dokument
- **Algorithmus**
  - Erzeugt effiziente und flexible Repräsentation von Dokumenten durch multivariate Statistik und Projektion
- **Visualisierung**
  - **Grafische** Präsentation
  - Ermöglicht **Exploration** der Dokumente
  - Unterstützt die **interaktive Verfeinerung** der Suchanfrage

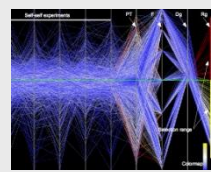




## 6.1 Dokumente

### Probleme der Analyse - Deklination, Konjugation

- Grammatikregeln verändern Wörter, **ohne dass Bedeutung** sich wesentlich ändert: Geld, Geldes, Gelder,...
- **Entfernen der Unterschiede** vor der Suche nach bestimmten Wörtern
  - Wortstammerkennung (stemming)



## 6.1 Dokumente

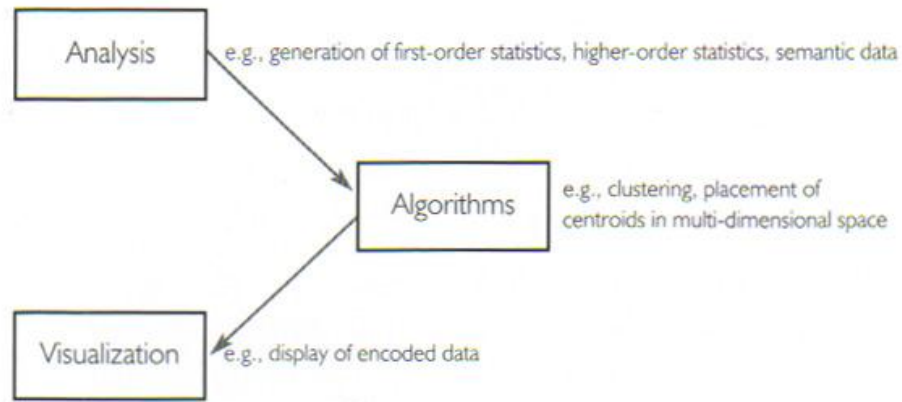
### Probleme der Analyse: Mehrdeutigkeit

- **Mehrfachbedeutungen** von Wörtern: Laser als Segelboot / gebündelter Lichtstrahl
- **Unterschiedliche Wörter** für gleiches Konzept: „gas“ und „patrol“ im AE/BE
- Nur 7 – 18% Wahrscheinlichkeit für **gleiches Wort für das gleiche Konzept** bei zwei Personen

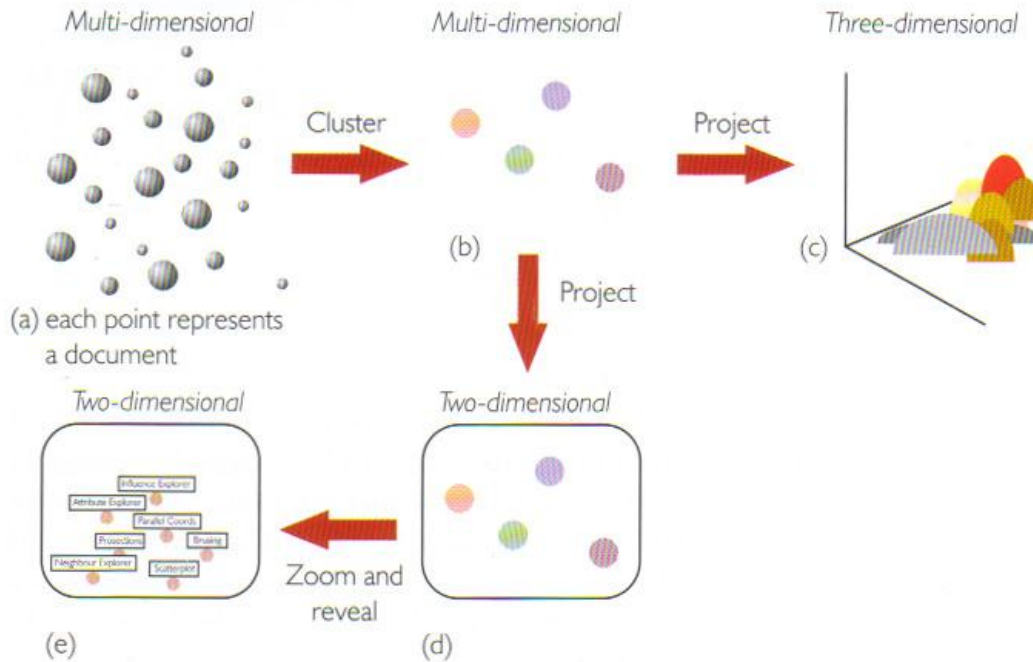
[Barnard, the contributions of applied psychology to the study of human-computer interaction, in Human Factors for Informatics Usability, Shackel, Richardson (Hrsg.), Cambridge University Press, Cambridge, UK, 1991]

- Daher werden oft **typische Dokumente vorgegeben** und dann ähnliche gesucht.
- **Weitere Probleme in Analysephase**: Teil des Information-Retrieval (automatische Sprachverarbeitung)

# 6.1 Dokumente



Stages involved in many schemes for document visualization

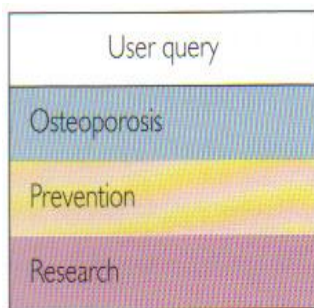


Clustering and projection involved in document visualization

# 6.1 Dokumente: Einzeldokument

## TileBars

- Input ist **Menge von Schlüsselwörtern** und Sammlung von Dokumenten
- **Zuordnung von Farben** zu Schlüsselwörtern
- Erzeugen eines TileBar für jedes Dokument
- Verbindungen (Links) von Farben zu Themen
- Rechts die Häufigkeit der Wörter in den **einzelnen Textbereichen** (Paragraphen, Seiten, Kapiteln)



Specification of a user query by topic words  
Source: Marti Hearst

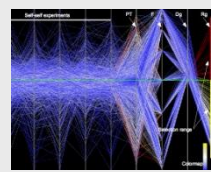


Recent advances in...

A TileBar indicating the relative occurrence of topic words within segments of a single document  
Source: Marti Hearst

[M. Hearst. TileBars: Visualization of Term Distribution Information in Full Text Access. Proceedings ACM SIGCHI 1995, 59-66.]

# 6.1 Dokumente: Einzeldokument



A typical set of TileBars for a collection of documents

Source: Marti Hearst

User Query  
(Enter words for different topics on different lines.)

osteoporosis  
prevention  
research

Run Search    New Query    Quit

Search Limit: 50 100 250 500 1000

Number of Clusters: 3 4 5 8 10

Mode: Tile Bars

Cluster	Titles
[Bar Chart]	FR88513-0157
[Bar Chart]	AP: Groups Seek \$1 Billion a Year for Aging Research
[Bar Chart]	SJMN: WOMEN'S HEALTH LEGISLATION PROPOSED C...
[Bar Chart]	AP: Older Athletes Run For Science
[Bar Chart]	FR: Committee Meetings
[Bar Chart]	FR: October Advisory Committees; Meetings
[Bar Chart]	FR88120-0046
[Bar Chart]	FR: Chronic Disease Burden and Prevention Models; Program
[Bar Chart]	AP: Survey Says Experts Split on Diversion of Funds for AIDS
[Bar Chart]	FR: Consolidated Delegations of Authority for Policy Developm
[Bar Chart]	SJMN: RESEARCH FOR BREAST CANCER IS STUCK IN P

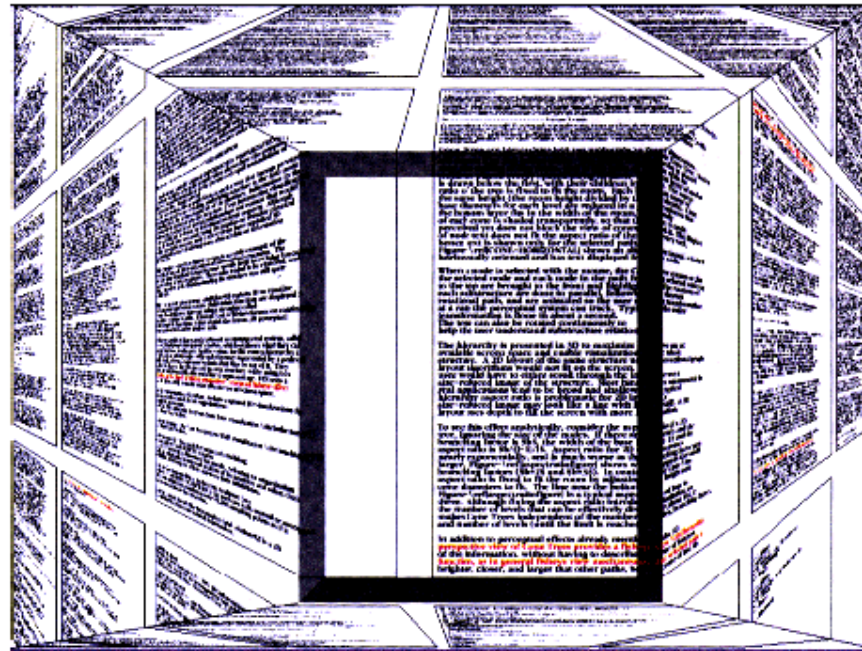
Display of a paragraph, with topic words highlighted in appropriate colours

Source: Marti Hearst

<ITAG tagnum=10>Fortunately, scientific knowledge about this disease has grown, and there is reason for hope. Research is revealing that prevention may be achieved through estrogen replacement therapy for older women and through adequate calcium intake and regular weight-bearing exercise for people of all ages. New approaches to diagnosis and treatment are also under active investigation. For this work to continue and for us to take advantage of the knowledge we have already gained, public awareness of osteoporosis and of the importance of further scientific research is essential.  
</ITAG>

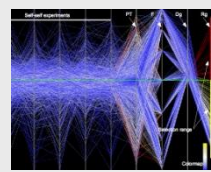
# 6.1 Dokumente: Einzeldokument

## Document-Lens



- Zeigt ein Textdokument aus **mehreren Textseiten**
- Kann **interaktiv durch Sichtfenster** geschoben werden: nicht verzerrt und damit lesbar
- Einzelne Textbausteine **im Kontext** einer Gesamtdokumentation visuell darstellen.

[G.G. Robertson, J.D. Mackinlay. The Document Lens. Proceedings ACM Symp' User Interface Software and Technology, 101-108]



## 6.1 Dokumente: Einzeldokument

### Sequential-Document-Visualization

- Interpretation eines Dokuments lokal (extrem: an jedem Wort) als (Einheits-) **Vektor im Raum  $\mathbb{R}^{\#\text{Wörter}}$**
- **Glätten mit Tiefpassfilter** (z. B. Gaussfilter) und erhält „locally weighted bag of words“ (“lowbow”)
  - Darstellung: Kurve in diesem hochdimensionalem Raum
- Erkennung von **Themenwechsel im Dokument** über Betrag des Gradienten der Kurve
- **Variieren des Glättungsfilters**: Position oder Standardabweichung
- Betrag der Krümmung (2. Ableitung) nutzbar
- Kurve selbst über Hauptkomponentenanalyse / Multidimensionale Skalierung darstellen
- **Vergleich von Kurven** erlaubt die Darstellung mehrerer Dokumente

[Y. Mao, J.V. Dillon, G. Lebanon. Sequential Document Visualization. Proceedings IEEE Infovis 2007, IEEE TVCG 13(6):1208-1215]

# 6.1 Dokumente: Einzeldokument

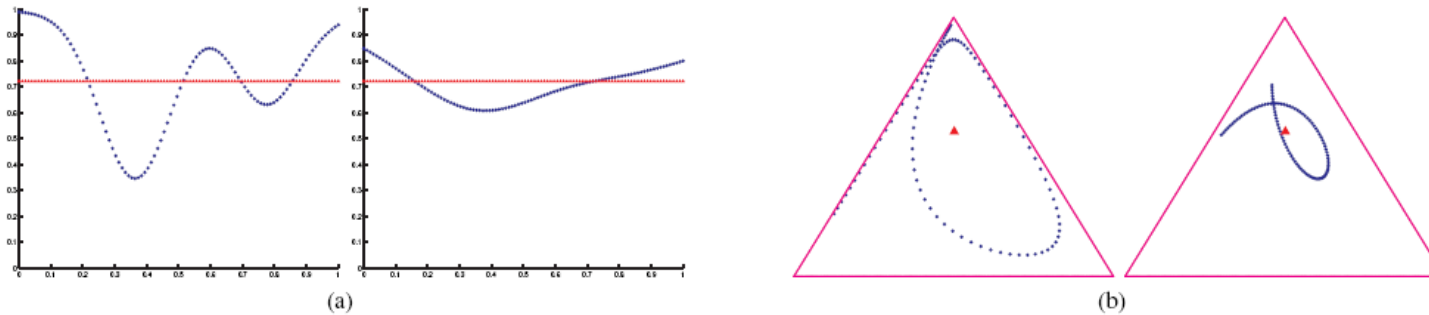
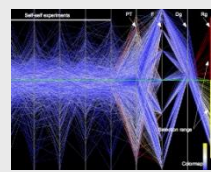


Fig. 2. (a) The curve  $\gamma^{(\sigma)}(z)$  in  $\mathbb{P}_{\{1,2\}}$  is visualized by graphing  $[\gamma_{\mu}^{(\sigma)}(z)]_1$  as a function of  $\mu/N$ . (b) The curve  $\gamma^{(\sigma)}(w)$  in  $\mathbb{P}_{\{1,2,3\}}$  is visualized by graphing it on the 2-D triangular space representing  $\mathbb{P}_{\{1,2,3\}}$  (see Figure 1). In both cases  $\sigma/N = 1/10$  (left) and  $\sigma/N = 2/10$  (right) were used. Increasing  $\sigma$  causes the curves to be more smooth and to shrink towards the red line (a) or red triangle (b) which denote the histograms or the degenerate curves  $\gamma^{(\infty)}(z), \gamma^{(\infty)}(w)$ .

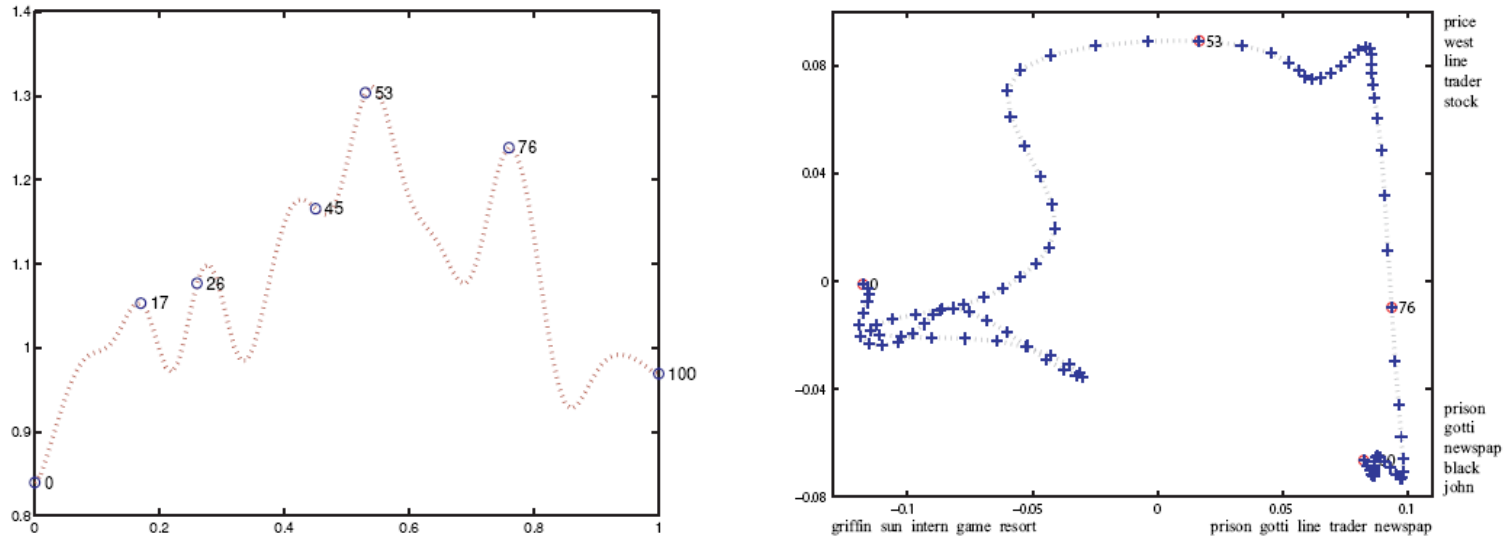


Fig. 3. Speed  $\|\dot{\gamma}_{\mu}^{(\sigma)}\|$  (left,  $\sigma/N = 0.064$ ) and 2D embedding using PCA (right,  $\sigma/N = 0.026$ ) of the lowbow curve representing the three successive RCV1 stories of varying lengths as a function of  $\mu/N$ . The words appearing to the right and bottom of the right panel correspond to the words whose projection on the principle components give the highest (or lowest) values providing a convenient visual summary for the user.



# 6.1 Dokumente: Einzeldokument

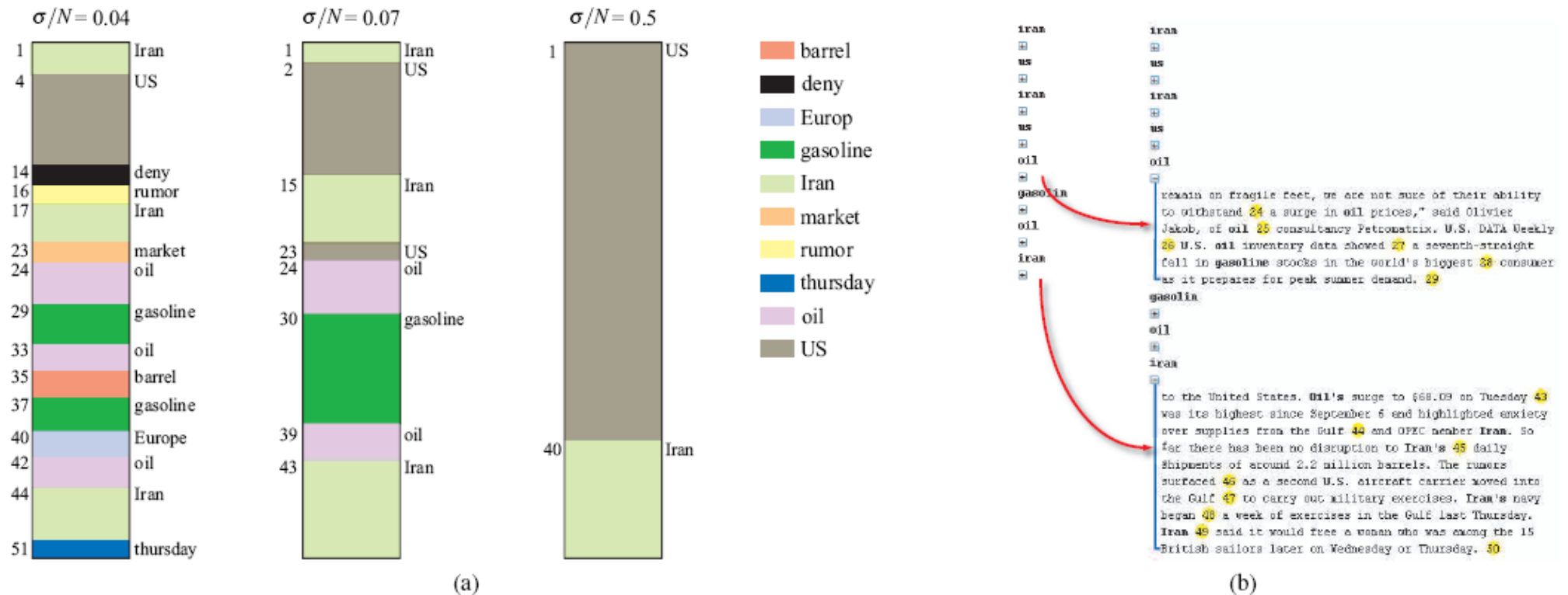


Fig. 6. Documents mapped from lowbow curves generated with three different values of  $\sigma/N$ . Some words are capitalized in the graph for improved readability. (a) Color bar approach. (b) Text folding approach.

# 6.1 Dokumente: Einzeldokument

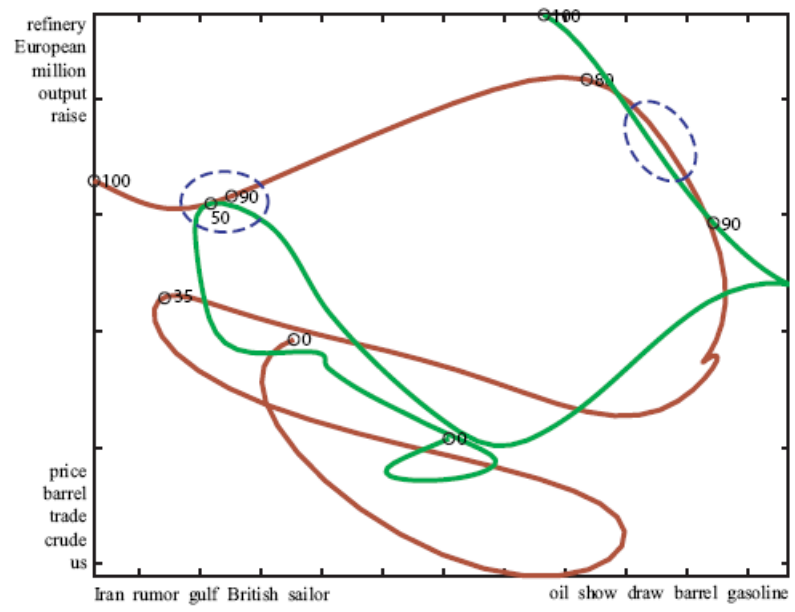


Fig. 7. Two dimensional PCA projection of the lowbow curves representing two similar news stories (see Section 4.4 for more details) using a VARIMAX rotation of PCA ( $\sigma/N = 0.04$ ).

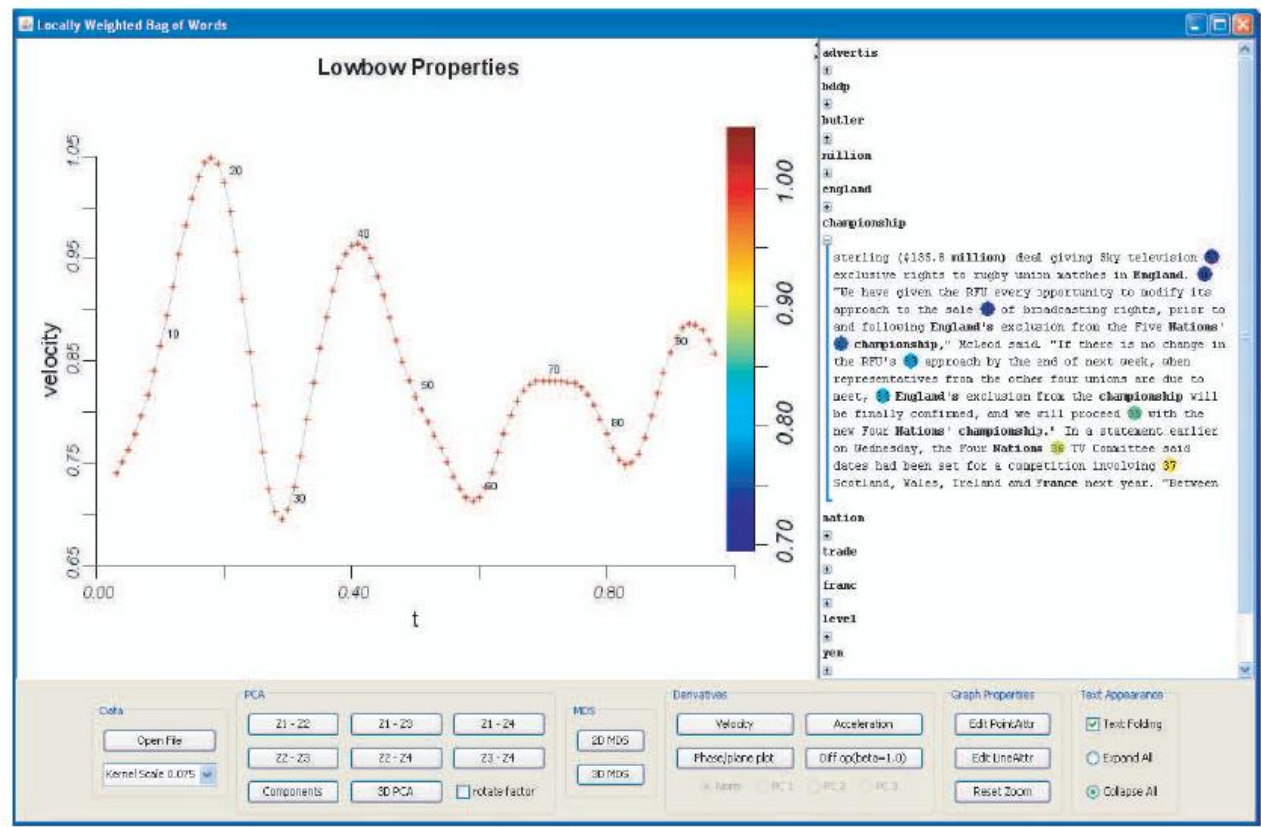


Fig. 8. Screen shot of the toolkit.

- **Vergleich von Kurven** erlaubt die Darstellung mehrerer Dokumente.

# 6.1 Dokumente: Ausgewählte Dokumente

## Data-Mountain

- Viele Dokumente auf einem Datenberg angeordnet
- Animiert mit Audio „fliegen“ Dokumente an anderen vorbei **in Vordergrund**
- **Räumliche Gedächtnis** des Menschen hilft beim Wiederfinden.

[G. Robertson, M. Czerwinski, K. Larson, D.C. Robbins, D. Thiel, M. van Dantzich. Data Mountain: Using Spatial Memory for Document Management. Proceedings ACM UIST 1998, 153-162.]



Figure 1: Data Mountain with 100 web pages.



## 6.1 Dokumente: Ganzer Korpus

### Galaxien

- Analyse erzeugt hochdimensionaler Vektor der Worthäufigkeiten
- Ähnlichkeit als Skalarprodukt definiert: Clustering mit Schwerpunktbildung
- Vorgabe eines Referenzdokumentes möglich
- Ergebnis in Ebene visualisieren

[Wise, Thoma, Pennock, Lantrip, Pottier, Schur, Crow, Visualizing the non-visual: Spatial Analysis and Interaction with Information from Text Documents, Proceedings IEEE InfoVis 1995, 76-82]

Zur Anzeige wird der QuickTime™  
Dekompressor „  
benötigt.

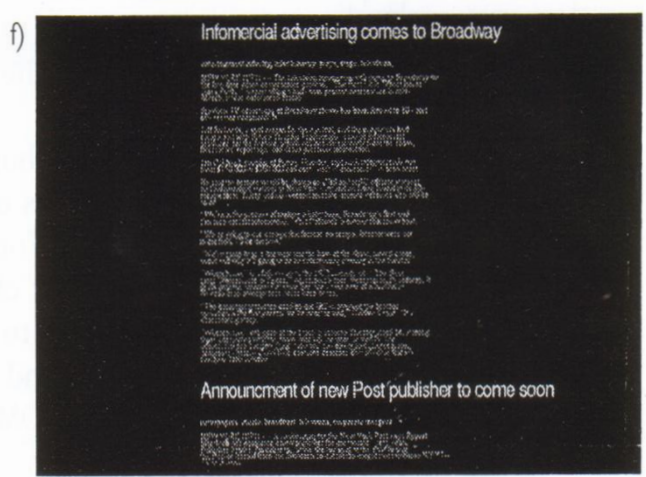
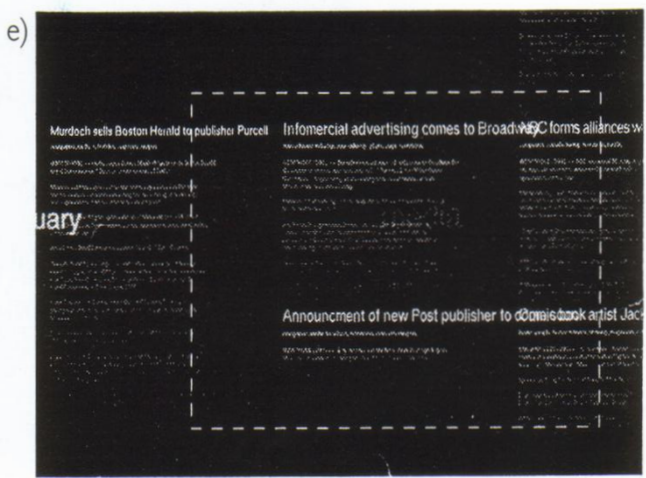
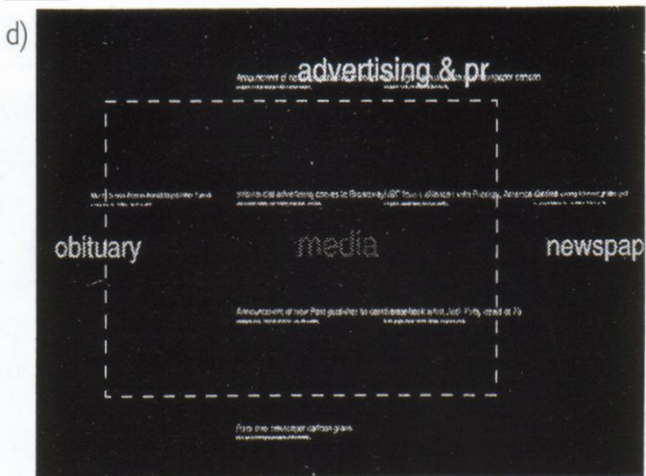
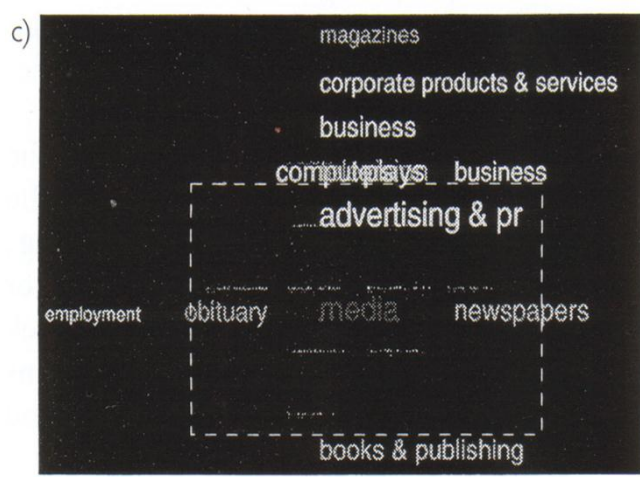
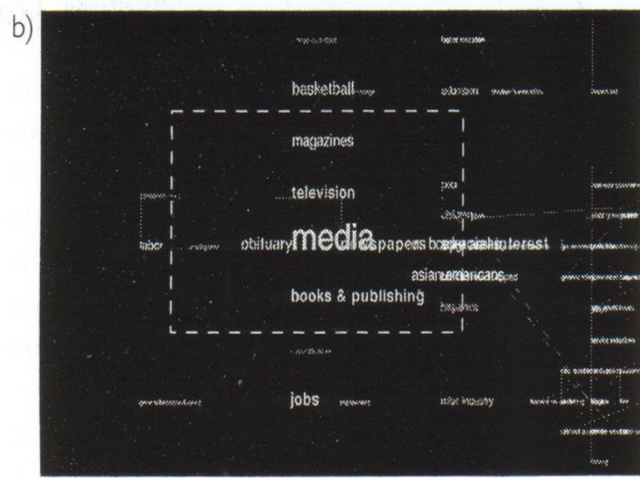
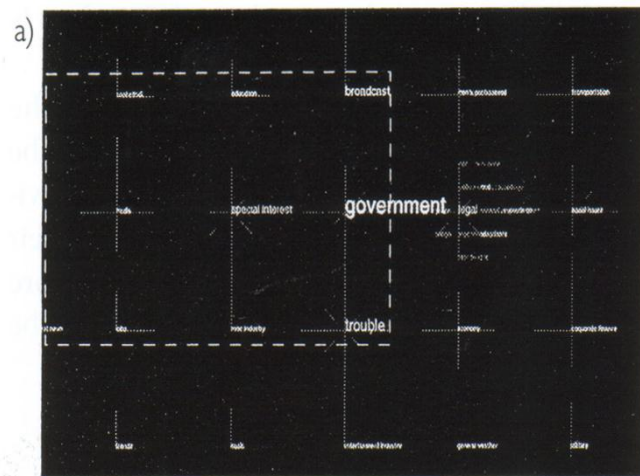
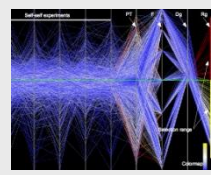


## 6.1 Dokumente: Ganzer Korpus

### Galaxy of News

- Erzeugt **Netzwerk von Symbolen** (etwa Schlüsselwörter, Zeitpunkte oder Orte) mit einer Abstandsmetrik
- **Legt diese Symbole in eine Ebene**, die der Anwender mit einem Fenster durchfahren kann
- **Dritte Raumdimension zoomed** bis zu den Artikeln
- Legt viel Wert auf **schnelle, kontinuierliche Bewegung**, um dem Anwender eine kontinuierliche Sicht zu gewähren

[Rennison, Galaxy of News: An Approach to Visualizing und Understanding Expansive News Landscapes, Proceedings ACM Symposium User Interface Software and Technology 1994, 3-12]



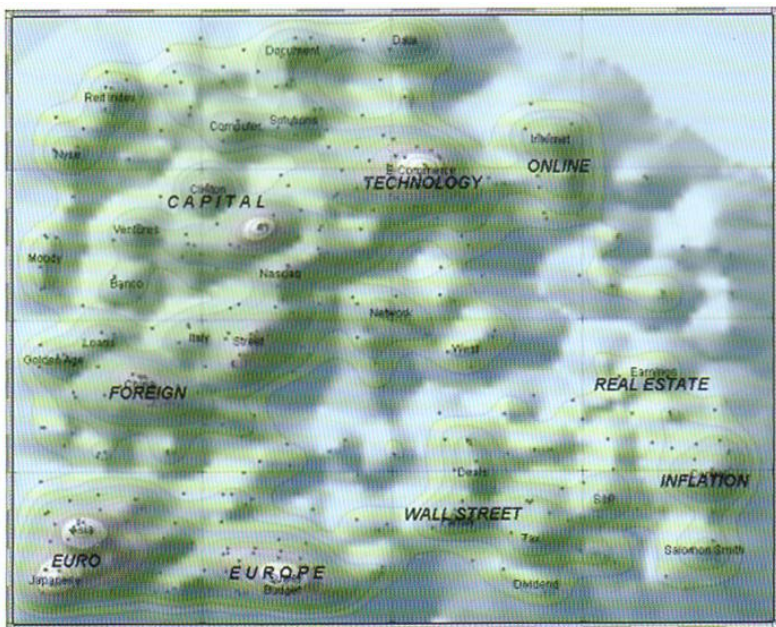
Successive views presented to a user of the Galaxy of News system

Source: Remnison (1994). Remnison © 1994 Association for Computing Machinery, Inc. Reprinted by permission

# 6.1 Dokumente: Ganzer Korpus

## ThemeScapes

- Anzahl der nahen Dokumente als Höhe definieren
- Repräsentation der Gebiete durch Gipfel: 3D-Landschaft
- Beispiel: 700 Artikeln aus Finanzbereich:

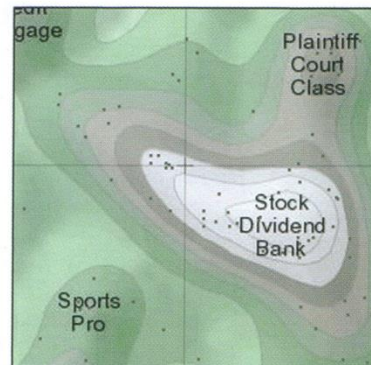


A themescape of articles related to the financial industry

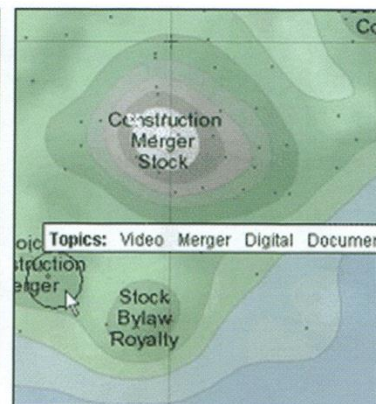
Source: [www.cartia.com](http://www.cartia.com)

Facilities available to a themescape user

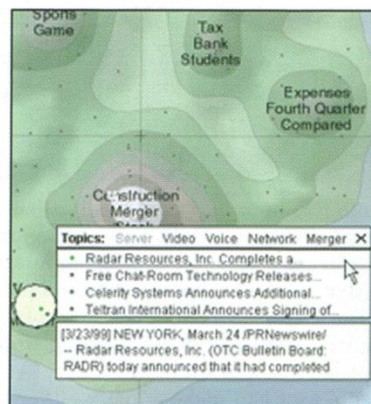
Source: [www.cartia.com](http://www.cartia.com)



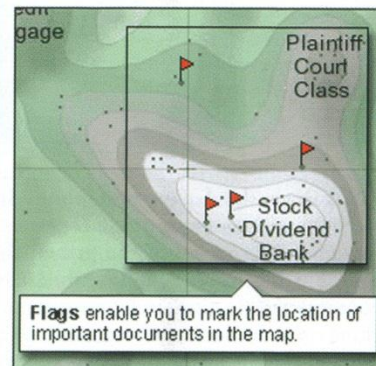
(a)



(b)



(c)



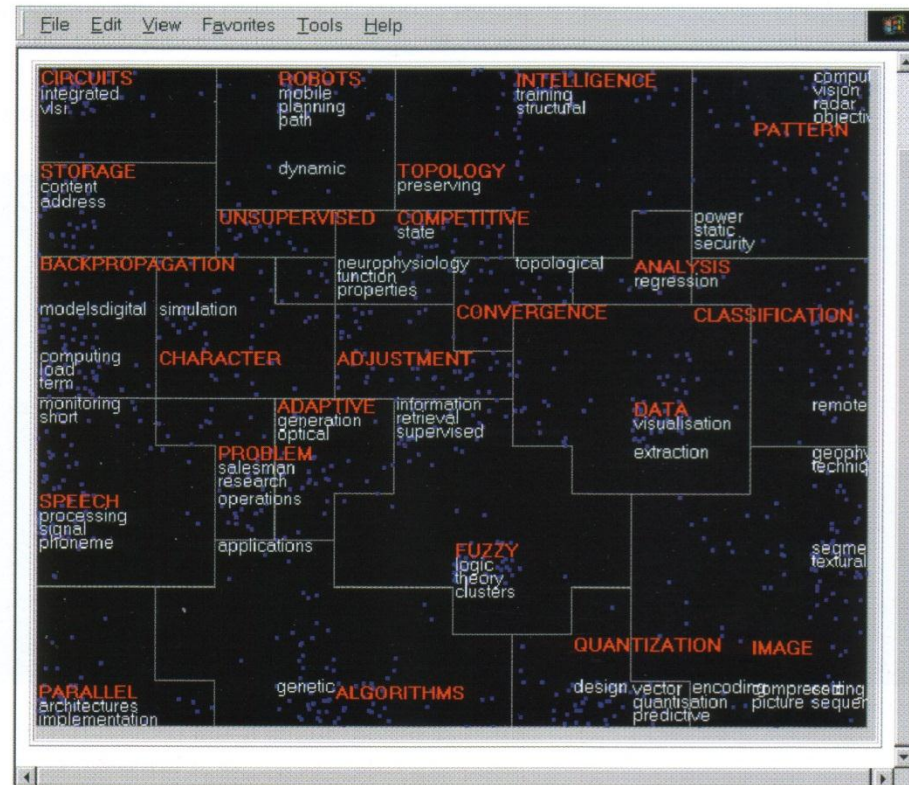
(d)

## Informationsvisualis

# 6.1 Dokumente: Ganzer Korpus

## Kohonen-Maps

- Nutzung von **Selbstorganisationsprinzipien neuronaler Netze** zur Dokumentvisualisierung
- Für  $m$  Objekte (Dokumente) mit  $n$  Merkmalen (Schlüsselwörter in Überschriften und Zusammenfassungen) erhält man **2D-Netz**
- Regionen mit ähnlichen Objekten definieren und darstellen



A Kohonen map representation of a collection of documents

Source: Xia Lin





## 6.1 Dokumente: Ganzer Korpus

### DEPICT

- **Kohonen-Maps** zur Anordnung von Dokumenten (Beispiel basiert auf 17700 AP Newswire Dokumenten)
- Knoten der Karte werden **anhand ihrer Vektorwerte gruppiert**
  - Clustering
  - Skalarprodukt als Metrik mit Schwellwert für Single-, Average- or Complete-Linkage
- Berechnung von **Cluster-Zentrum** („Centroid“): Label des am nächsten gelegenen Einzelwortes
- Anzeige dieser Worte
- **Geordnete Liste der Dokumente** durch Auswahl von einem oder mehreren dieser Themen (nach Ähnlichkeit - Skalarprodukt)

[D.A. Rushall, M.R. Ilgen. DEPICT: Documents Evaluated as Pictures. Proceedings IEEE Information Visualization 1996, 100-107.]

# 6.1 Dokumente: Ganzer Korpus

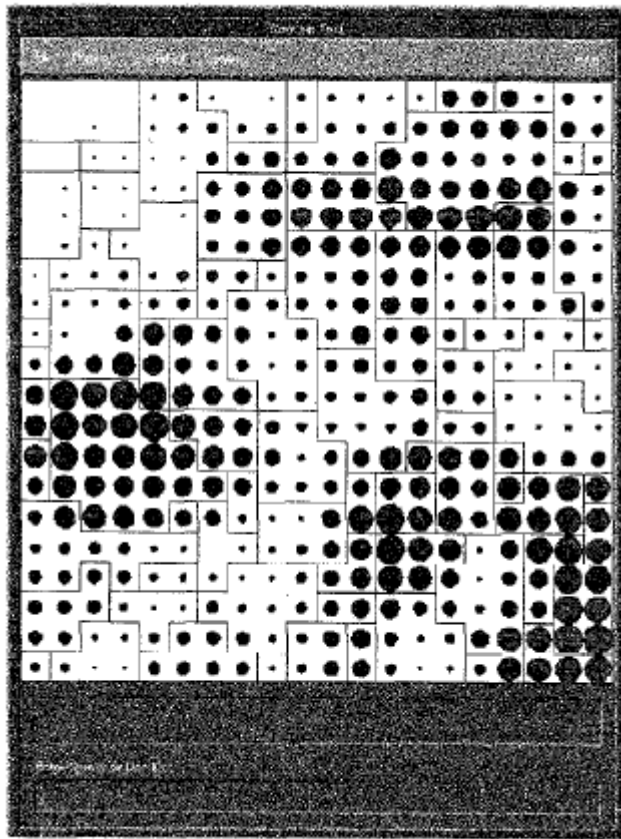


Figure 4a. The automatically generated regions

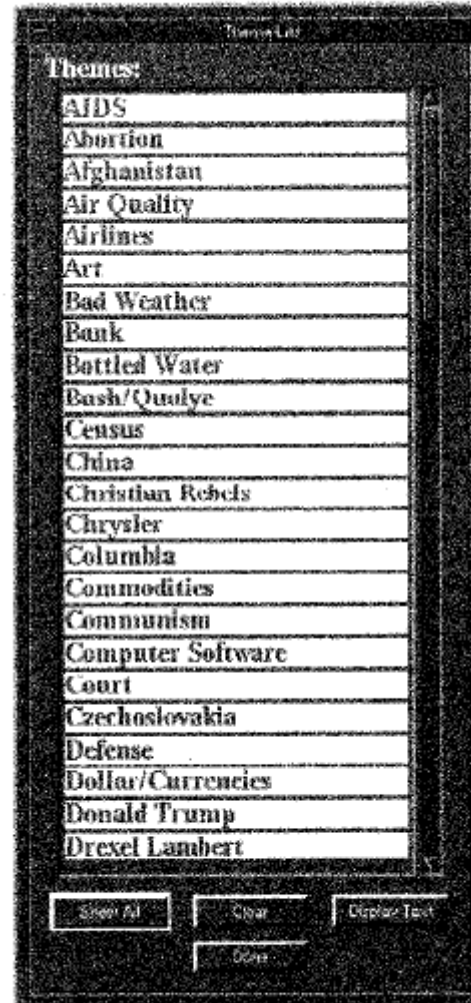


Figure 4b. The automatically generated labels.

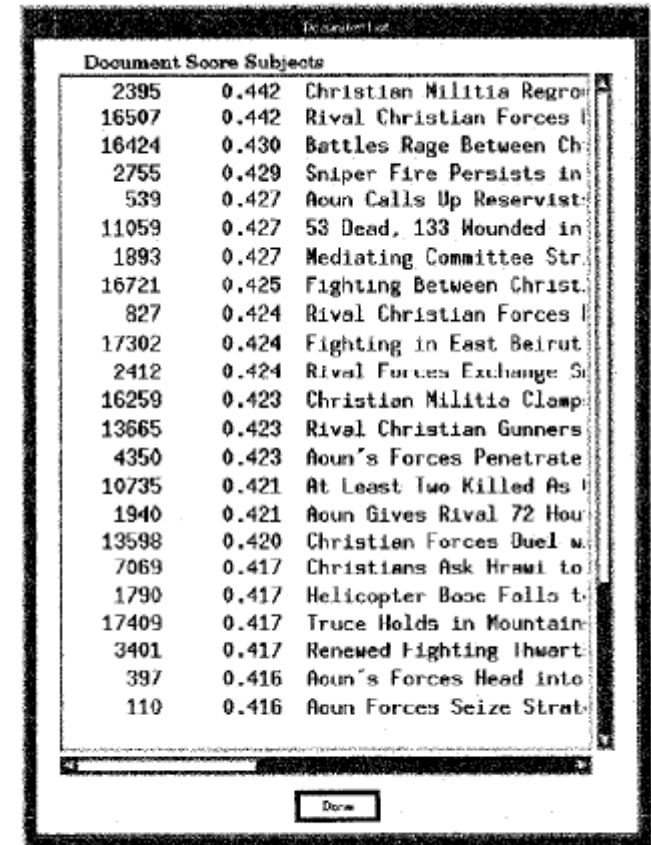
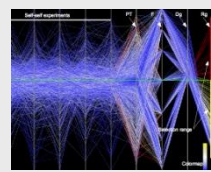


Figure 5. A ranked list of documents.



## 6.1 Dokumente: Ganzer Korpus

### DEPICT: Zeitliche Analyse

- Zeitliche Betrachtung wenn Dokumente **Datum tragen**
- **Vektor aus Zeitscheibe bilden**, der dann animiert über die SOM-Karte läuft. (Bei Newswire-Daten zeigt sich dann beispielsweise der übliche Rhythmus von Werktagen und Wochenende sehr deutlich.)

# 6.1 Dokumente: Zeitabhängiger Korpus

## ThemeRiver

- Sammlung wird **in einzelne Zeitscheiben aufgeteilt** (Dokumente mit Zeitstempel)
- **Beherrschende Themen** (Begriffe) für jede Scheibe ermitteln
- **Reduktion** der Themenzahl
- **Zeitliche Verlauf** der Anzahl Artikel pro Thema wird als **Fluss mit Teilströmen** verschiedener Farben dargestellt
- **Ähnliche Themen erhalten ähnliche Farben** und werden nebeneinander angeordnet

[S. Havre, E. Hetzler, P. Whitney, L. Nowell. ThemeRiver: Visualizing Thematic Changes in Large Document Collections. IEEE TVCG 8(1):9-20, 2002.]

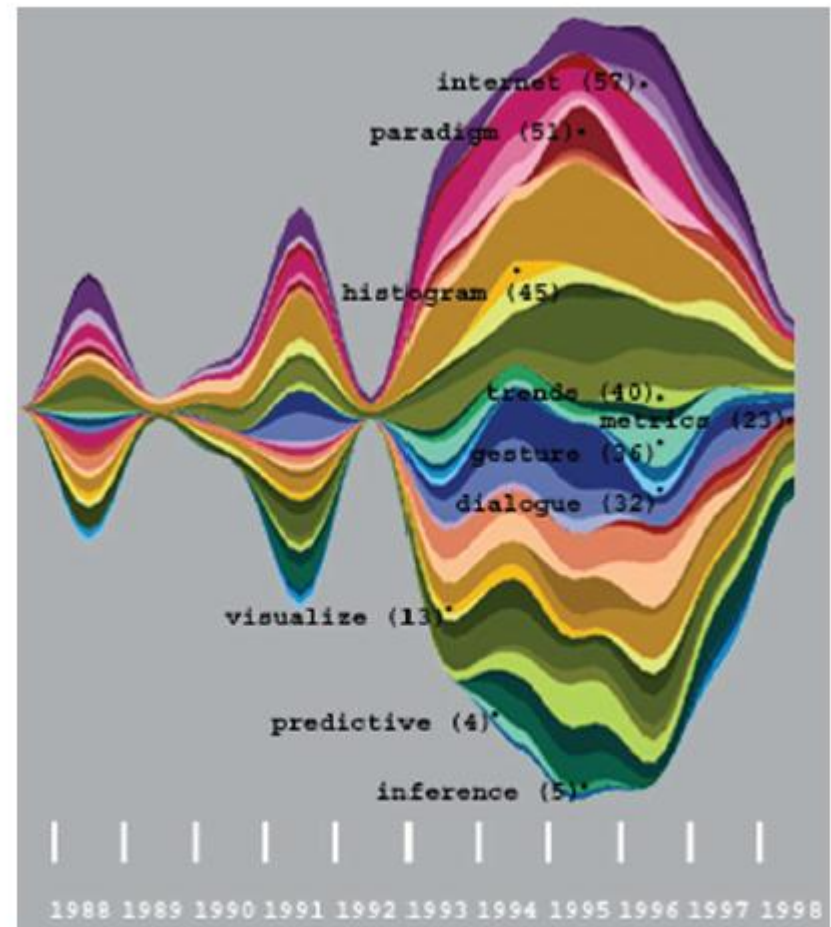


Fig. 1. ThemeRiver depicts thematic changes over time in a collection of patents from one company.

# 6.1 Dokumente: Zeitabhängiger Korpus

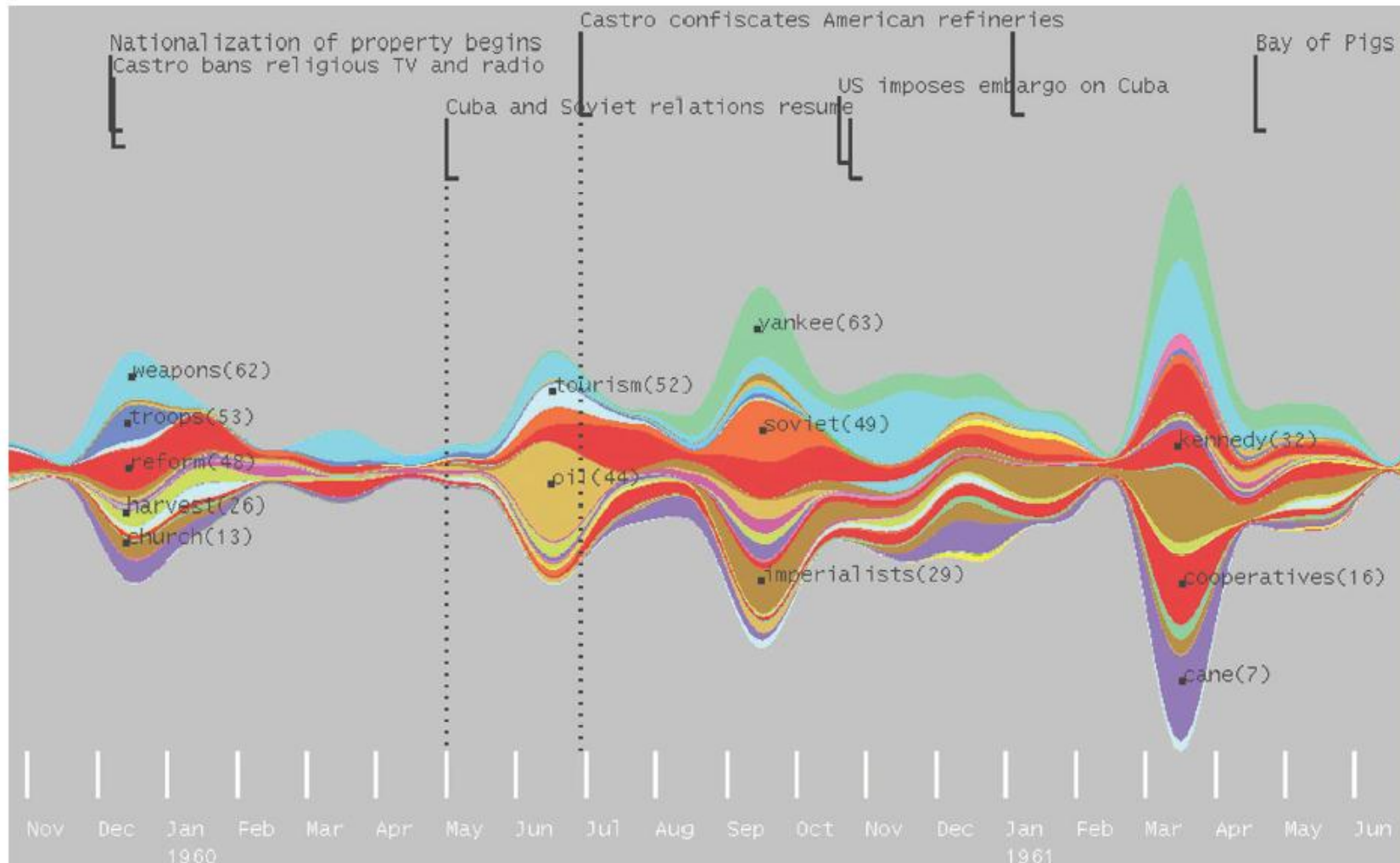


Fig. 2. ThemeRiver uses a river metaphor to represent themes in a collection of Fidel Castro's speeches, interviews, and articles from the end of 1959 to mid-1961.

# 6.1 Dokumente: Zeitabhängiger Korpus

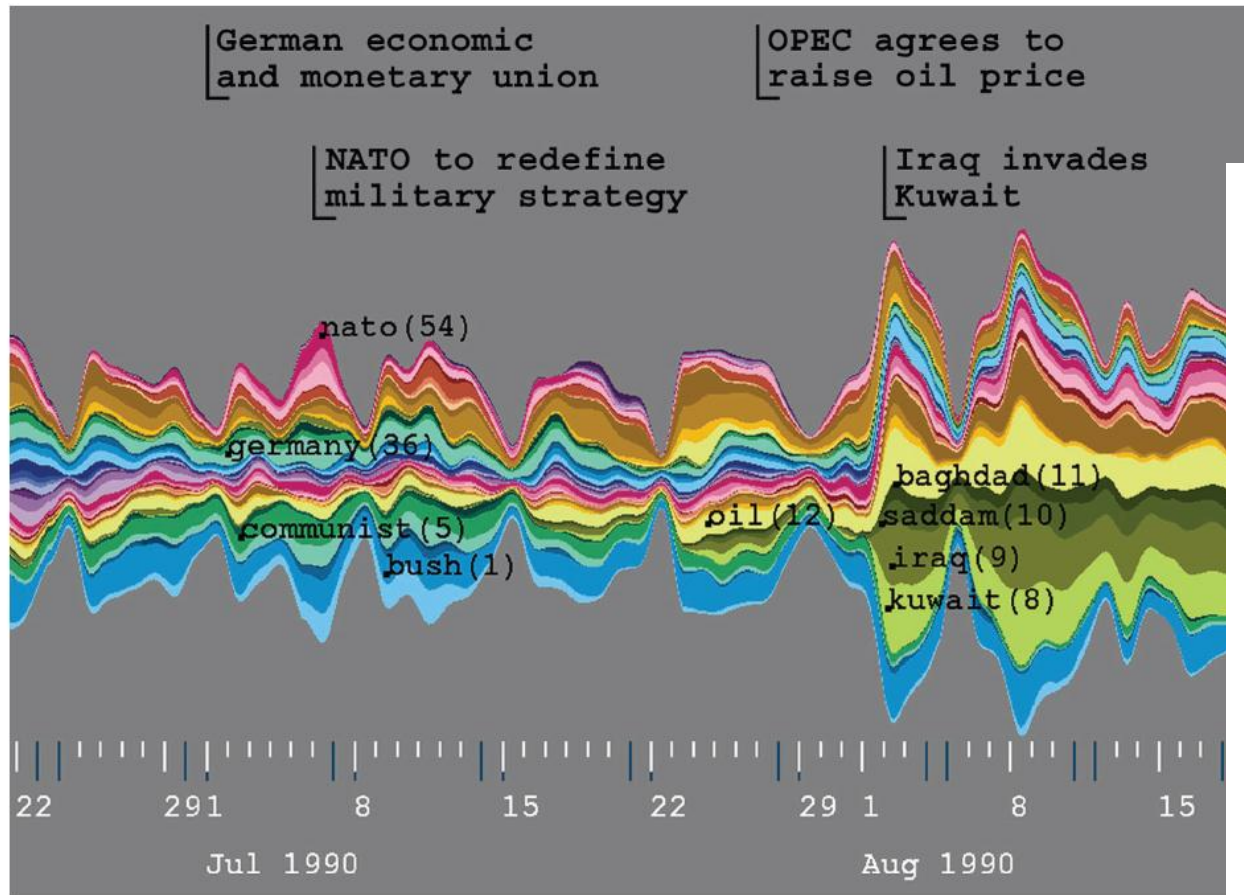


Fig. 3. ThemeRiver visualization of Associated Press news wire stores from July and early August 1990.

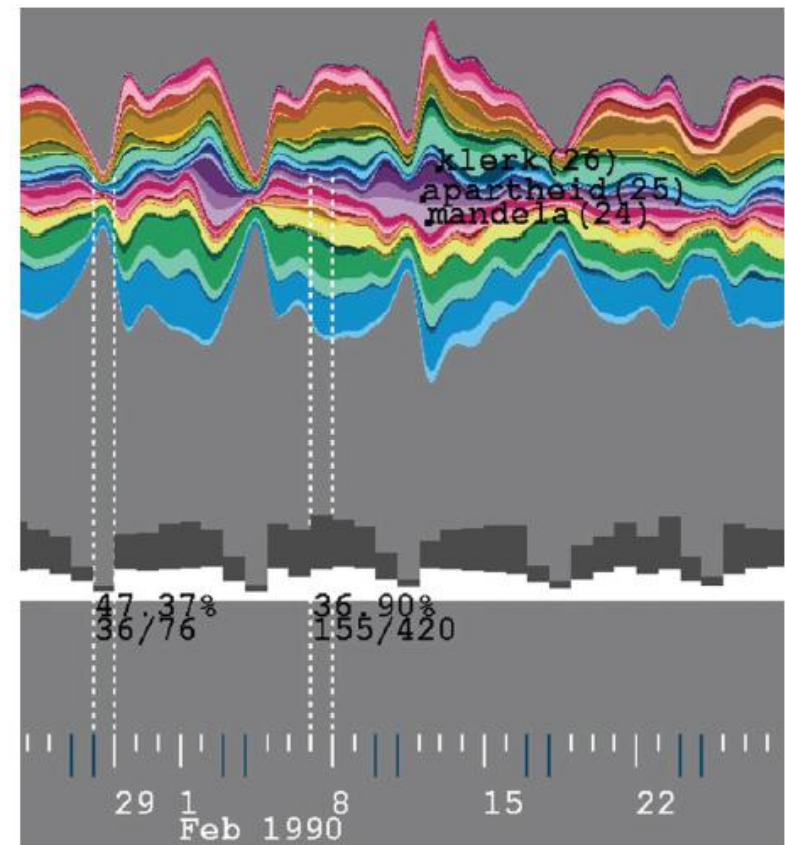
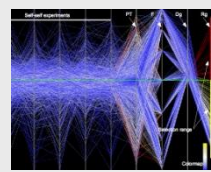


Fig. 5. The addition of a histogram to ThemeRiver reveals that news is light on Sundays, not that themes shift.



## 6.1 Dokumente: Zeitabhängiger Korpus

### Sequential-Patterns

- Suchen nach **wiederkehrenden Folgen** von Themen („Sequential-Patterns“) in größere Dokumentensammlungen mit Zeitstempel
- Dabei geht es um eine **zeitliche Folge von Themen A->B->C->D**, die mit hoher Wahrscheinlichkeit in dieser Reihenfolge auftritt (mit Sprüngen/Füller)
- Diese “Sequential-Patterns” lassen sich in einem **Zeit-Thema-Diagramm** darstellen und visuell erfassen
- Visualisierung dient
  - Zur Information über den **zeitlichen Ablauf** dieser Folgen
  - Um zu erkennen, ob dieses Muster nur in einer **bestimmten Zeitspanne** auftritt oder **regelmäßig** wiederkehrt

[P.C. Wong, W. Cowley, H. Foote, E. Jurrus, J. Thomas. Visualizing Sequential Patterns for Text Mining. Proceedings IEEE Information Visualization Symposium 2000, 1-5.]

# 6.1 Dokumente: Zeitabhängiger Korpus

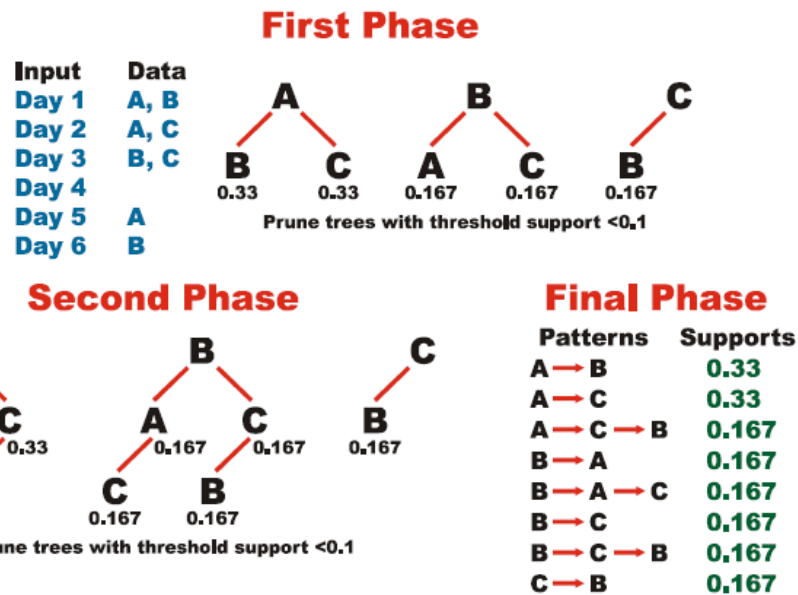


Figure 2: An example to illustrate the mining of sequential patterns from a small time-series dataset. The input data are blue. The output patterns and their supports are green. The mining task completes after two phases of matching and pruning.

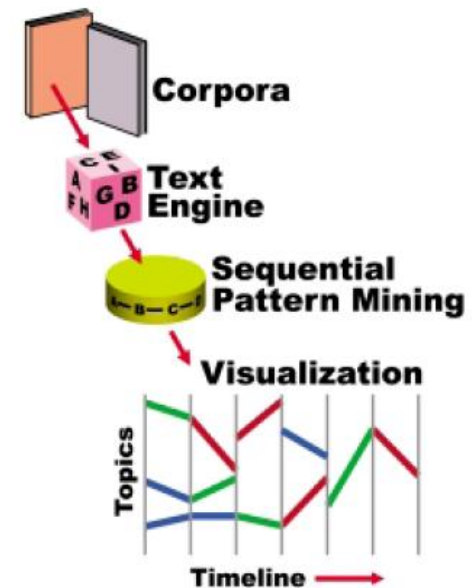


Figure 3: An overview of our visual data mining system in the discovery of sequential patterns of a corpus.



# 6.1 Dokumente: Zeitabhängiger Korpus

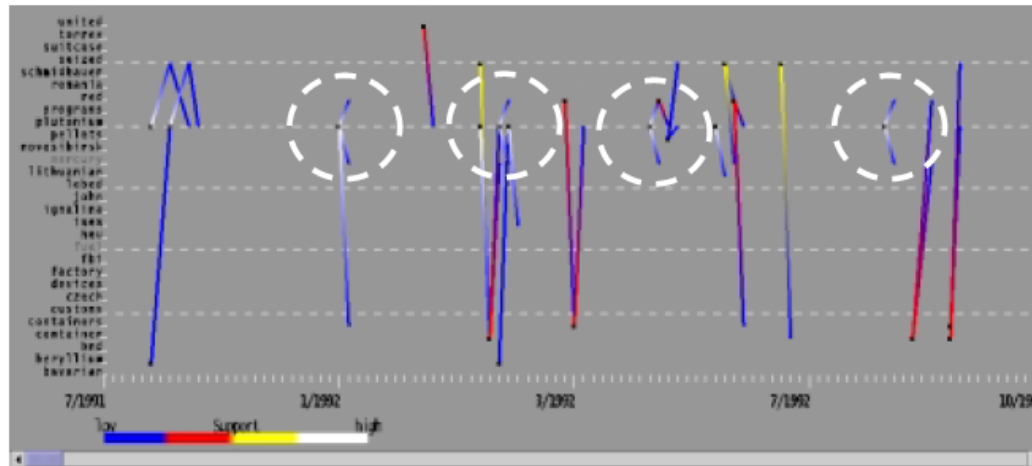


Figure 4: A snapshot of the visualization system. The topics are listed on the left-hand side. Colored lines are used to connect the topics to form a sequential pattern. The beginning of a pattern is marked by a black dot. Four basic colors – shown at the bottom – are used to indicate the support of a pattern and its sub-patterns. Shading is turned on to blend the colors between two sub-pattern support values. The four white dashed circles mark the four appearances of the same two patterns within an 8-month period from Jan 92 to Aug 92.

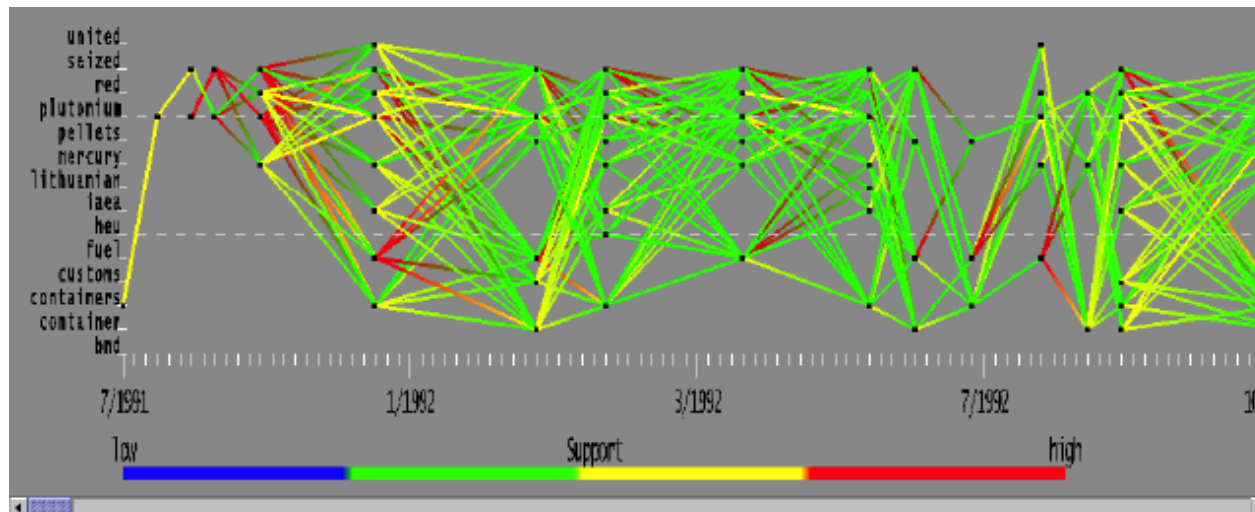
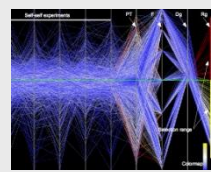


Figure 5: A visualization snapshot of sequential patterns using the monthly binning topic sets.



## 6.1 Dokumente: Suchergebnisse

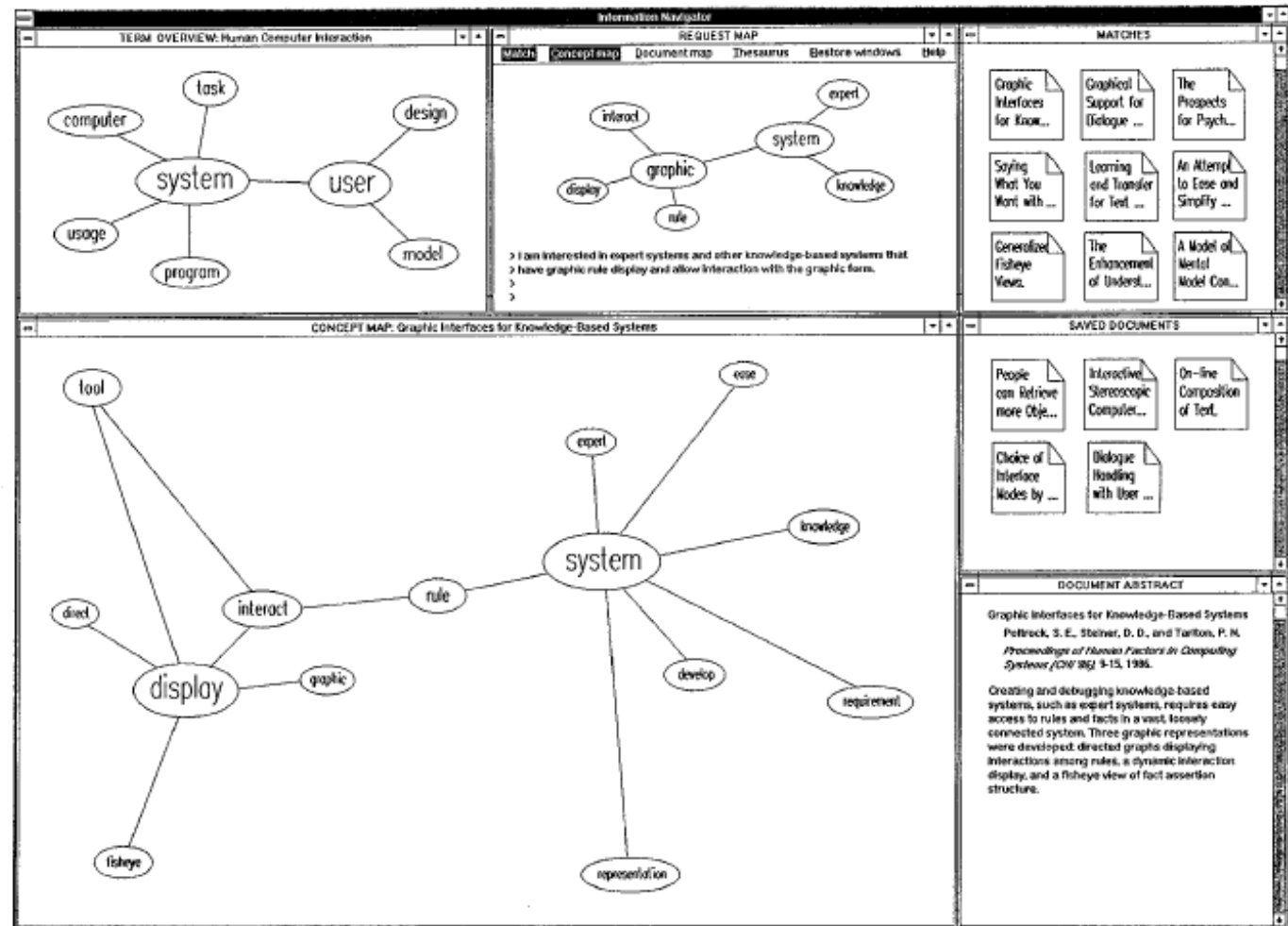
### Information-Navigator

- Anfrage in natürlicher Sprache: **Dokument analysieren** und **wesentliche Begriffe** herausuchen
- **Ähnlichkeitsnetzwerk (Graph) der Suchbegriffe** durch Analyse gemeinsam auftretender Begriffe
- Reduktion auf die **signifikante Kanten** empfehlenswert
- **Graph** kann zur Präzisierung der Suchanfragen **manipuliert** werden
- Thesaurusnutzung zum besseren Verständnis des Themengebiets
- Anzeige des aktuellen Dokuments über
  - Zusammenfassung (bei wissenschaftlichen Artikeln einfach)
  - Begriffsnetzwerk des Dokuments
- Darstellung von Teilen des gesamten **Ähnlichkeitsnetzwerkes aller Dokumente**
- IdR. Wird **Netzwerk zu groß**: zu viele Dokumente

[R. H. Fowler, W. A. L. Fowler, B. A. Wilson. Integrating Query, Thesaurus, and Documents through a Common Visual Representation. Proceedings of 14<sup>th</sup> Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1991, 142-151.]

# 6.1 Dokumente: Suchergebnisse

## Navigator



**Figure 1:** A typical display. The upper windows display 1) an overview diagram of the associative term thesaurus, 2) the text of a natural language query with its visual representation, and 3) an ordered list of documents retrieved in a search. Below the retrieved document icons are 4) documents saved by the user. The lower windows show 5) a document abstract network and 6) the document abstract's text.



## 6.1 Dokumente: Suchergebnisse

### InfoCrystals

- Typischer Fall: Suche nach **vielen Begriffen**, aber **nicht alle Begriffe in allen Dokumenten**
- Suche nach **geeigneter Untermenge** ist idR. schwierig
- InfoCrystals sucht nach **allen möglichen UND-Verknüpfungen** der gegebenen Begriffe gleichzeitig
- Passende Dokumentenanzahl **numerisch** oder Kuchendiagramm anzeigen
- **Anzahl Begriffe** definiert Seiten des „Kristalls“
- UND-Verknüpfungen mit gleicher Begriffsanzahl liegen **auf einem Kreis**
- Kreise sind konzentrisch angeordnet
- Gleichzeitige Auswahl (ODER) mehrerer UND-Varianten ergibt **boolesche Anfragekombination** (disjunktive Normalform)
  - Beachte: Dokument erscheint nur in UND-Verknüpfung mit wenigen Begriffen, wenn es die anderen Begriffe nicht enthält

[A. Spoerri. InfoCrystal: A visual tool for information retrieval. Proceedings IEEE Visualization, 1993, 156-163.]

# 6.1 Dokumente: Suchergebnisse

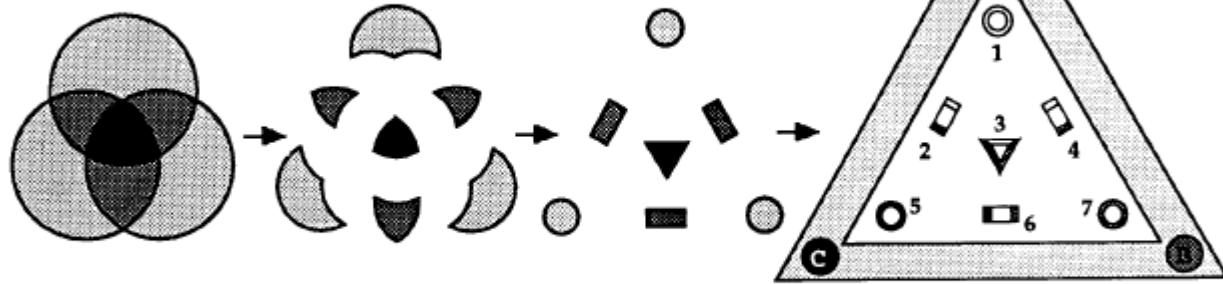


Figure 1: shows how to transform a Venn diagram into an iconic display, called the *InfoCrystal*, which represents all the possible Boolean queries involving its inputs in a normal form (see section 2.2). The interior icons have the following Boolean meanings: 1 = (A and (not (B or C)), 2 = (A and C and (not B)), 3 = (A and B and C), 4 = (A and B and (not C)), 5 = (C and (not (A or B))), 6 = (B and C and (not A)), 7 = (B and (not (A or C))).

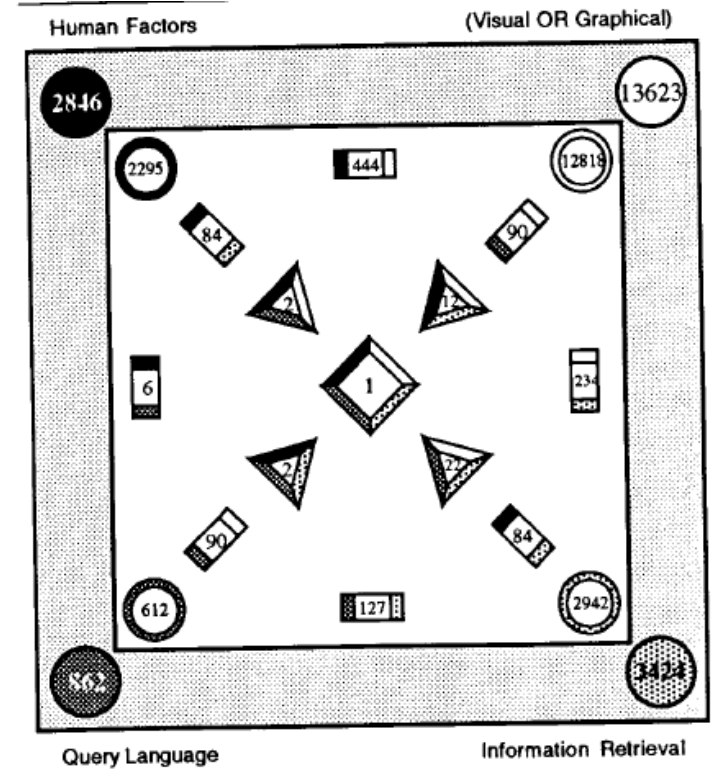
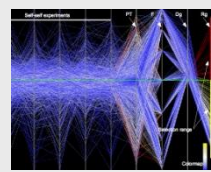


Figure 2: The number associated with an icon indicates how many of the retrieved documents satisfy the conditions represented by it.

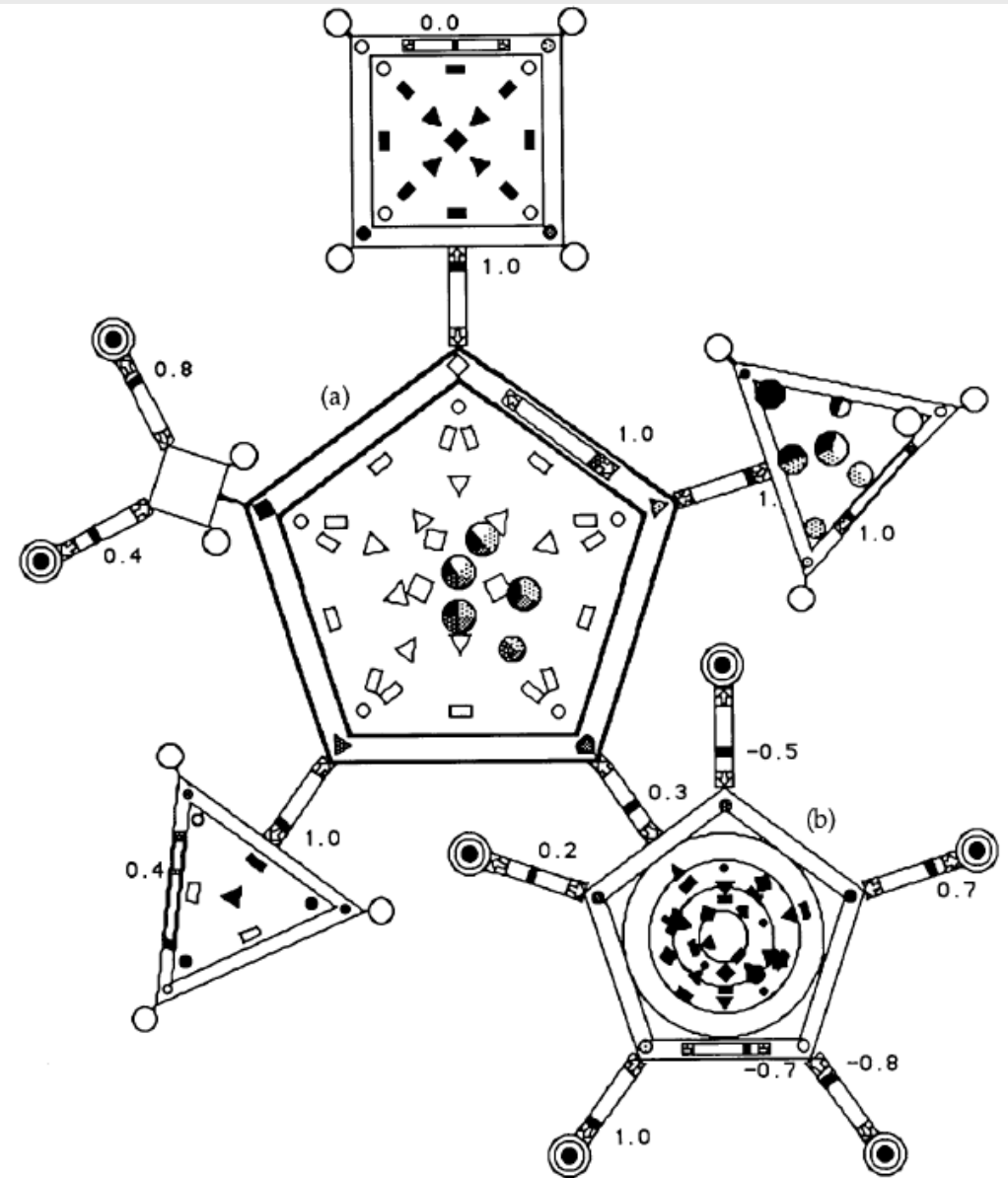


## 6.1 Dokumente: Suchergebnisse

### Komplexe Suchanfragen mit InfoCrystals

- Menge der Dokumente durch **Auswahl gewünschter Icons** (bei jedem InfoCrystal)
- **Komplexe Anfragen** durch mehrere InfoCrystals als Input für ein InfoCrystal
  - Nutzung von Gewichten für **vektorielle Ähnlichkeit** statt boolescher Logik (Eingabe per Slider zwischen -1 und 1)
  - **Schwellenwert** zur Dokumentauswahl erforderlich

# 6.1 Dokumente: Suchergebnisse



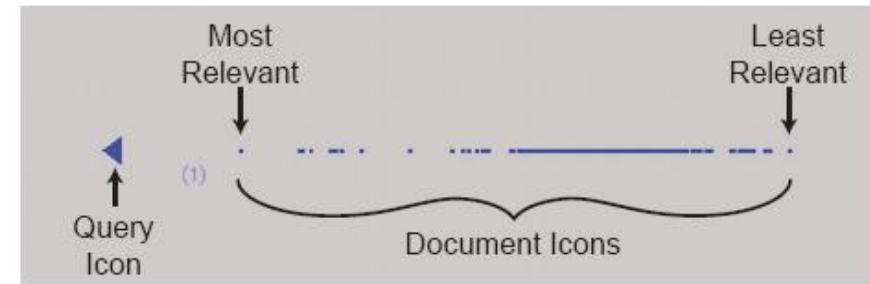
**Figure 3:** shows how the InfoCrystals can be organized in a hierarchical structure. Users can interactively change the way the InfoCrystals filter their inputs and they can dynamically observe how the information coming in through the circular InfoCrystals is propagated through the query structure. The interior icons that are shown in solid black indicate that they have been selected by the user to define the output of an InfoCrystal. Some of the InfoCrystals are displayed only as an outline, but the user can just click on them to view them in full detail. (a) Shows the top-level InfoCrystal, using the *rank layout* principle, where the selected icons are rendered, using a pie chart representation, to emphasize the quantitative information associated with them. (b) Shows the *bull's-eye layout* of the interior icons when the input weights are set to -0.5, 0.7, 0.8, 1.0 and 0.2; and it shows which icons will be selected if the threshold is set to -0.7 (in this case all of them).

Informations

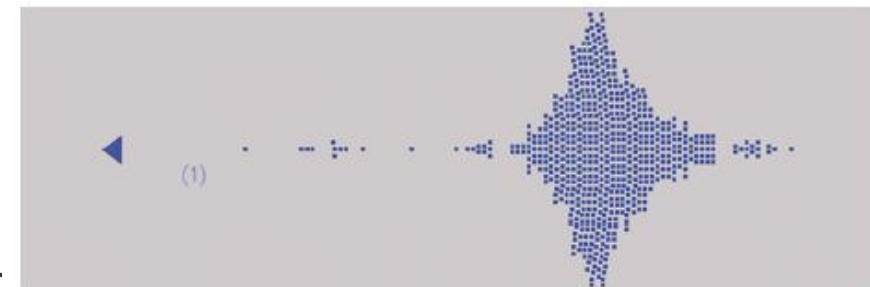
## 6.1 Dokumente: Suchergebnisse

### Darstellung von Suchergebnissen mit Sparkler

- Sparkler **positioniert die Suchanfrage(n)** im Zentrum eines Kreises
- Dokumente werden **längs eines radiales Strahls** entsprechend ihrer **Distanz** zu den Anfragetermen ausgelegt
- Zur **Vermeidung von Overplotting** (Überdeckung):
  - Dokumente mit gleichem Abstand werden um den Strahl **gestreut** (gleicher Radius)
  - Mehrere Anfrage ergeben so **Kreissegmente mit Achsen** und **gestreuten Dokumenten**



a.



b.

Figure 1. Document icons are plotted in relation to the query icon (a) overplotted along the invisible query line and (b) spread radially

[S. Havre, E. Hetzler, K. Perrine, E. Jurrus, N. Miller. Interactive Visualization of Multiple Query Results. Proceedings IEEE Information Visualization 2001, 105-112.]



# 6.1 Dokumente: Suchergebnisse

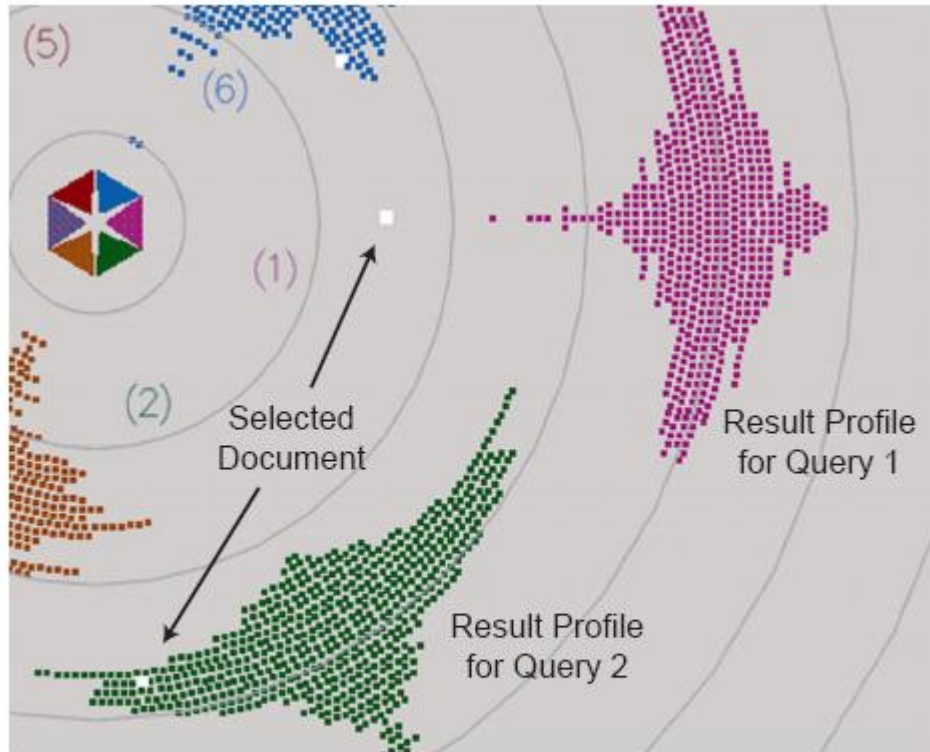


Figure 2. Zoomed-in view of two expanded result profiles and one selected document

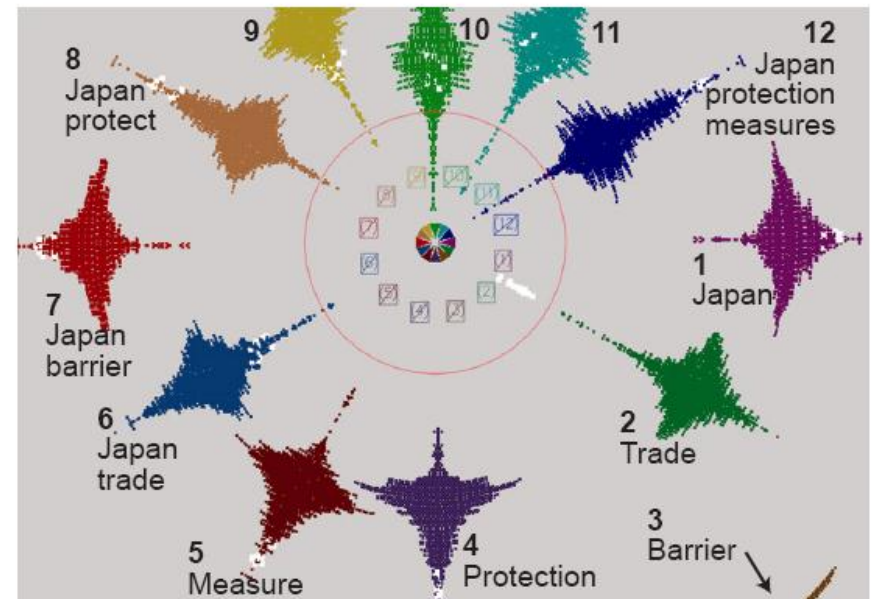


Figure 5. We select document icons inside the ring along the Query 2 result profile. These same documents are also highlighted in the other sets.

# 6.1 Dokumente: Assoziationen

## Darstellung von Assoziationsregeln

- **Aufstellen von Regeln** wie Begriffe zusammenhängen:  
Inwieweit tritt ein Begriff auf, wenn bereits andere Begriffe auftreten
  - **Bessere Beschreibung** von Themen
  - **Reduktion Anzahl** der relevanten Wörter (Dimensionen)
- Erzeugt **Hypergraph über Wörtern**  
(Wörter sind Knoten, Regeln sind Hyperkanten von notwendigen Worten zu dann auftretenden Worten.)
- **Inzidenzmatrix** als Möglichkeit der Darstellung

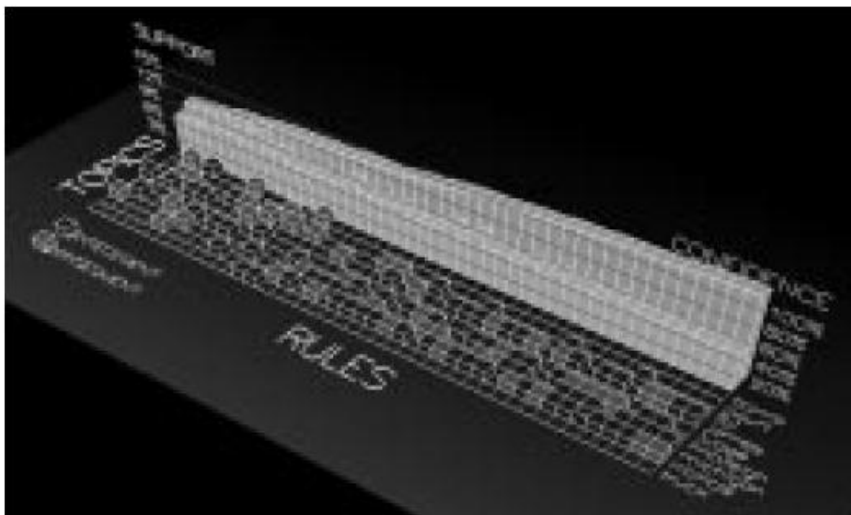


Figure 8: Rules are sorted by confidence values. See CP2.

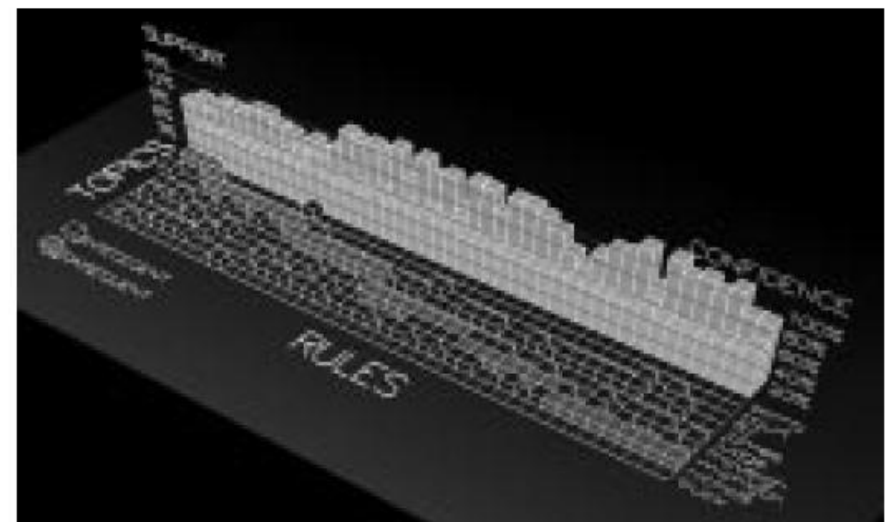


Figure 9: Rules are sorted by consequent topics. See CP3.

# 6.1 Dokumente: User-Interface

## History

- Anwender betrachtet **einzelne Dokumente** aus Dokumentensammlungen
- Darstellung des Wegs über einzelne Dokumente als **Bewegung über die Zeit** darstellen  $\Rightarrow$  Geschichte (History)
- Funktionen zur **raschen Rückkehr** zu vorher besuchten Dokumenten

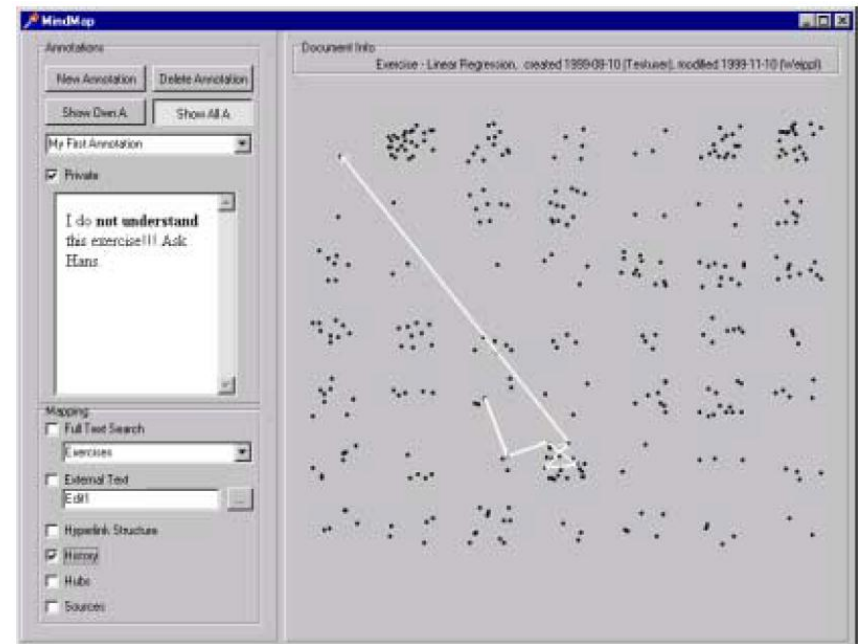
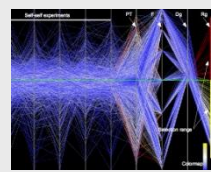


Figure 2: The 'History' function shows the last six pages that have been read and the one currently open. The white line connects them so that the user can see his trail through the information landscape.



## 6.2 Videodaten

- **Nicht nur** Textdokumente als Analysegegenstand
- **Intuitiver und schneller Zugang** zu großen Mengen an Videodaten wichtig
- Interaktive Suche über **Visualisierung**
  - Bei ungenauen Vorgaben nützlich
  - Beispiele im Folgenden

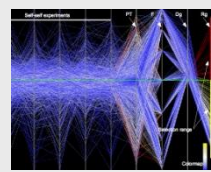


## 6.2 Videodaten

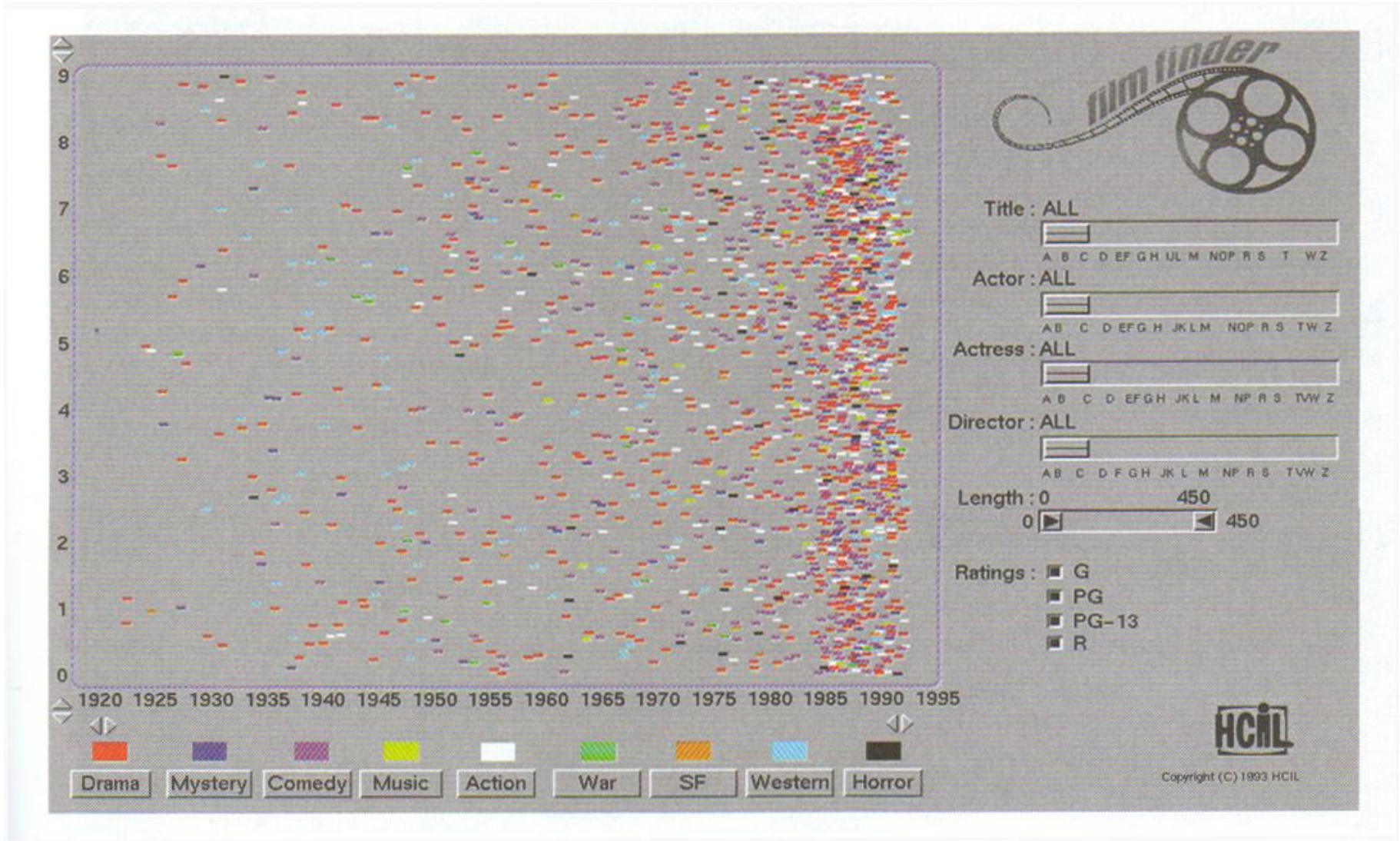
### Filmdaten

- **Entscheidungshilfe** bei der Auswahl von Filmen: FilmFinder
- **Filmdatenbank** enthält folgende Attribute:  
Erscheinungsjahr, Popularität, Kategorie, Titel, Hauptdarsteller, Hauptdarstellerin, Regisseur, Länge in Minuten, Jahrgangsempfehlung.
  - **Erscheinungsjahr** auf x-Achse abbilden
  - **Popularität** werden auf y-Achse ab abbilden
  - Film symbolisiert durch kleines Rechteck
  - Farbe beschreibt **Kategorie**
- Suche nach anderen Merkmalen einschränken (Slider)
- Weniger als 25 Ergebnisse: Titel erscheint
- Nach Filmauswahl: Alle Angaben werden in Extrafenster gelistet

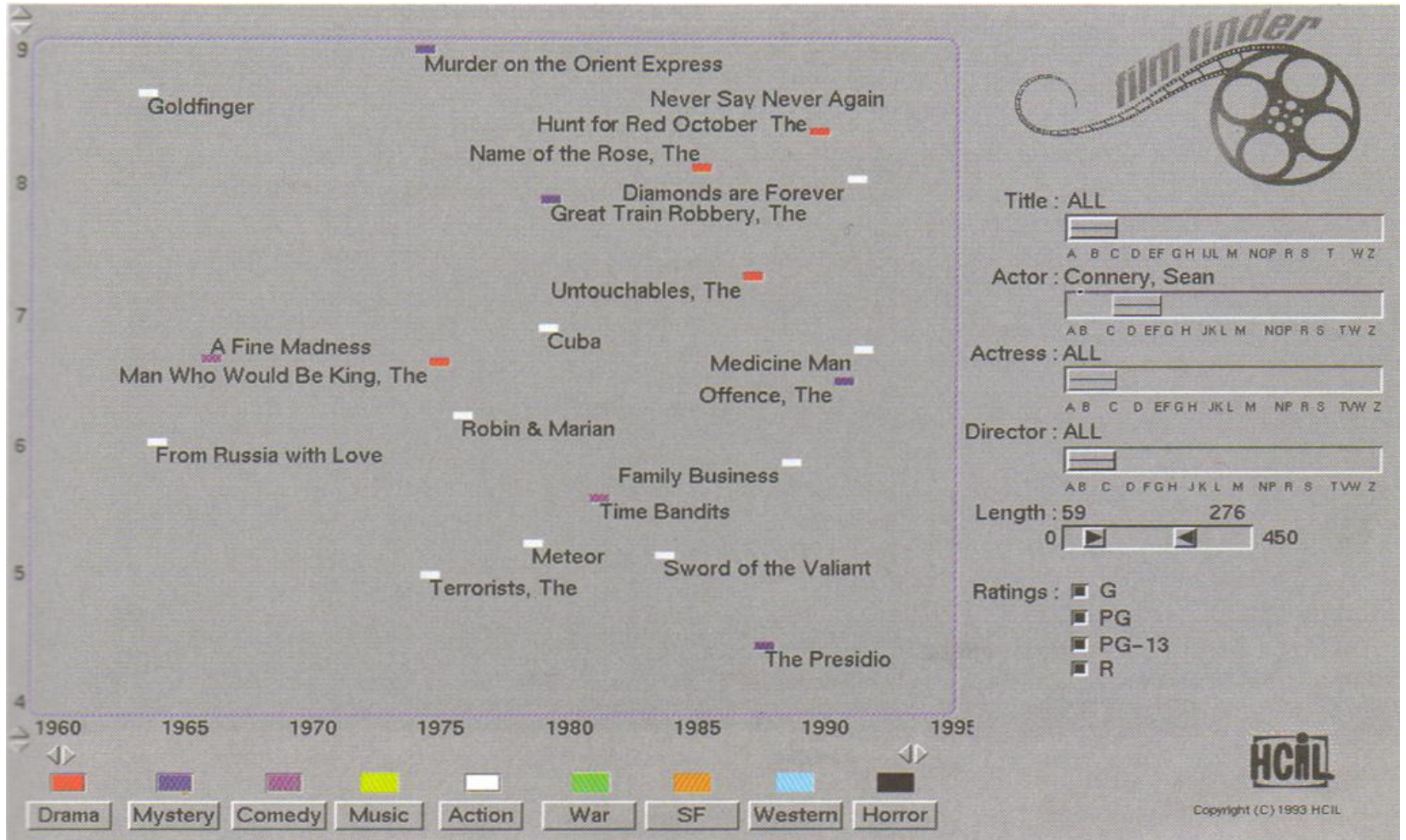
[Ahlberg, Schneiderman, Visual Information Seeking: Tight Compling of Dynamic Query Filters with Starfields Displays, in Readings in Information Visualization, Card, Mac Kinlay, Schneiderman (Hrsg.), Morgan Kaufmann Publishers, San Francisco, CA, USA, 1999]



# 6.2 Videodaten



# 6.2 Videodaten



# 6.2 Videodaten

**filmfinder**

**Search Filters:**  
 Title : ALL  
 Actor : Perkins, Anthony  
 Actress : Bergman, Ingrid  
 Director : ALL  
 Length : 59 / 276 / 450  
 Ratings :  G  PG  PG-13  R

**Genre Legend:**  
 Drama (Red) Mystery (Blue) Comedy (Purple) Music (Yellow) Action (White) War (Green) SF (Orange) Western (Light Blue) Horror (Black)

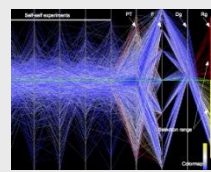
**Movie Data from Scatter Plot:**

Year (approx)	Title	Genre
1964	Goldfinger	Action
1974	Murder on the Orient Express	Mystery
1976	Never Say Never Again	Action
1977	Hunt for Red October, The	Action
1978	Name of the Rose, The	Mystery
1980	Diamonds are Forever	Action
1980	Great Train Robbery, The	Action
1987	Untouchables, The	Action
1990	Cuba	Action
1991	Medicine Man	Action
1992	A Fine Madness	Comedy
1993	Man Who Would Be King, The	Action
1994	Offence, The	Mystery

**Pop-up: Murder on the Orient Express**  
 Director: Lumet, Sidney Year: 1974  
 Country: USA Language: English  
 Actors: Balsam, Martin; Cassel, Jean-Pierre; Perkins, Anthony; Connery, Sean; Gielaud, John  
 Actresses: Bacall, Lauren; Bergman, Ingrid; Bisset, Jacquelin; Hiller, Wendy

**HCIL**  
 Copyright (C) 1993 HCIL



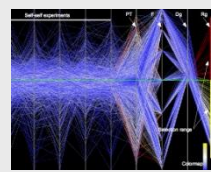


## 6.2 Videodaten

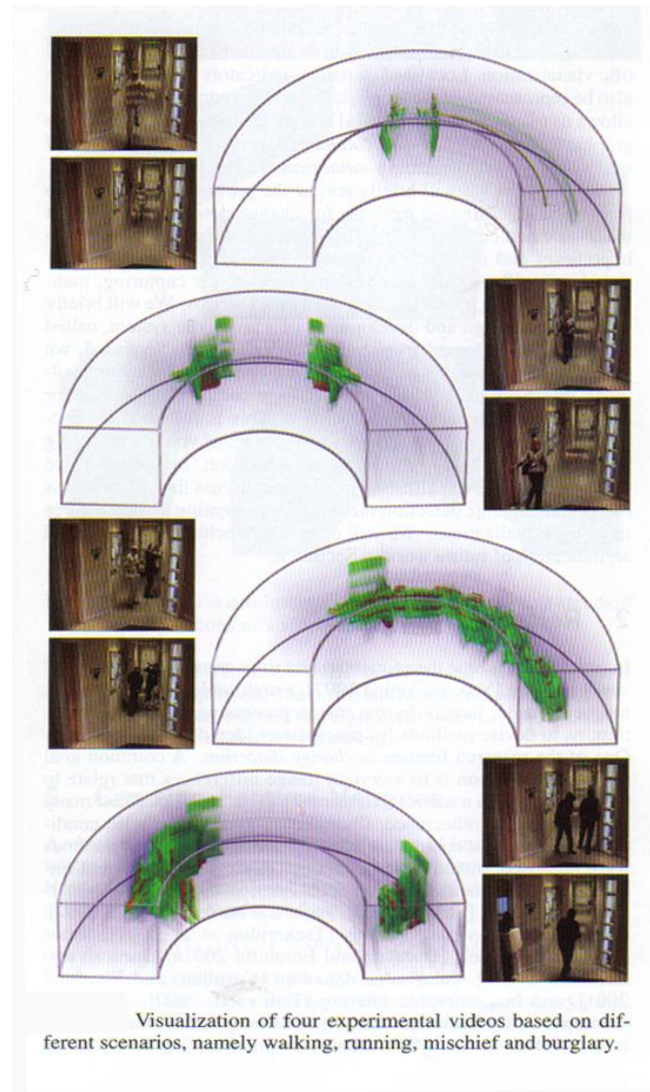
### Video-Visualization

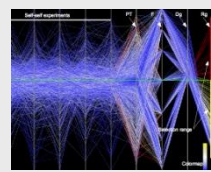
- Videodaten stellen **sehr große Datenmenge** dar
- **Ohne Metadaten (DB)** ist Suche nach Auffälligkeiten oder kurzen Abschnitten sehr mühsam oder praktisch unmöglich
- Darstellung des **Videos als Volumen**
- Handhabbare Darstellung durch **Differenzbildung**
- Verschiedener Farbräume
- Unterschiedliche Metriken
- Alternative Transferfunktionen

[Daniel, Chen, Video Visualization, Proceedings IEEE Visualization 2003, 409-416]

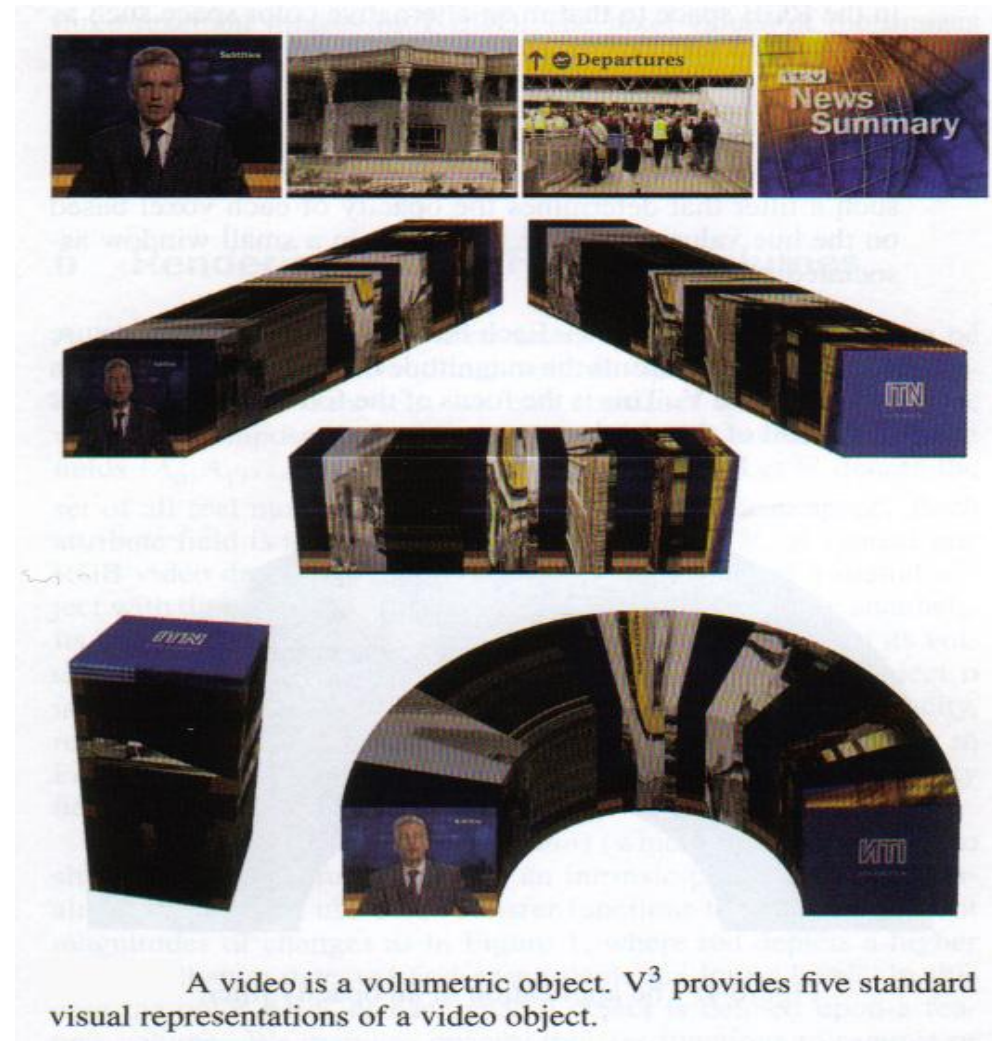


## 6.2 Videodaten





## 6.2 Videodaten



## 6.2 Videodaten

