

**Universität Stuttgart**

# Die Bedeutung von Lexika für die visuelle Text- und Dokumentanalyse

Steffen Koch, Thomas Ertl

6. Arbeitstreffen des Netzwerks Internetlexikografie

Mannheim, 20.11.2013



Institut für Visualisierung  
und Interaktive Systeme

# Ein Bild sagt mehr als 1000 Worte!?



[Die niederländischen Sprichwörter von Pieter Bruegel dem Älteren, 1559, Quelle: [www.wikipedia.de](http://www.wikipedia.de)]

# Ein Bild sagt mehr als 1000 Worte!?

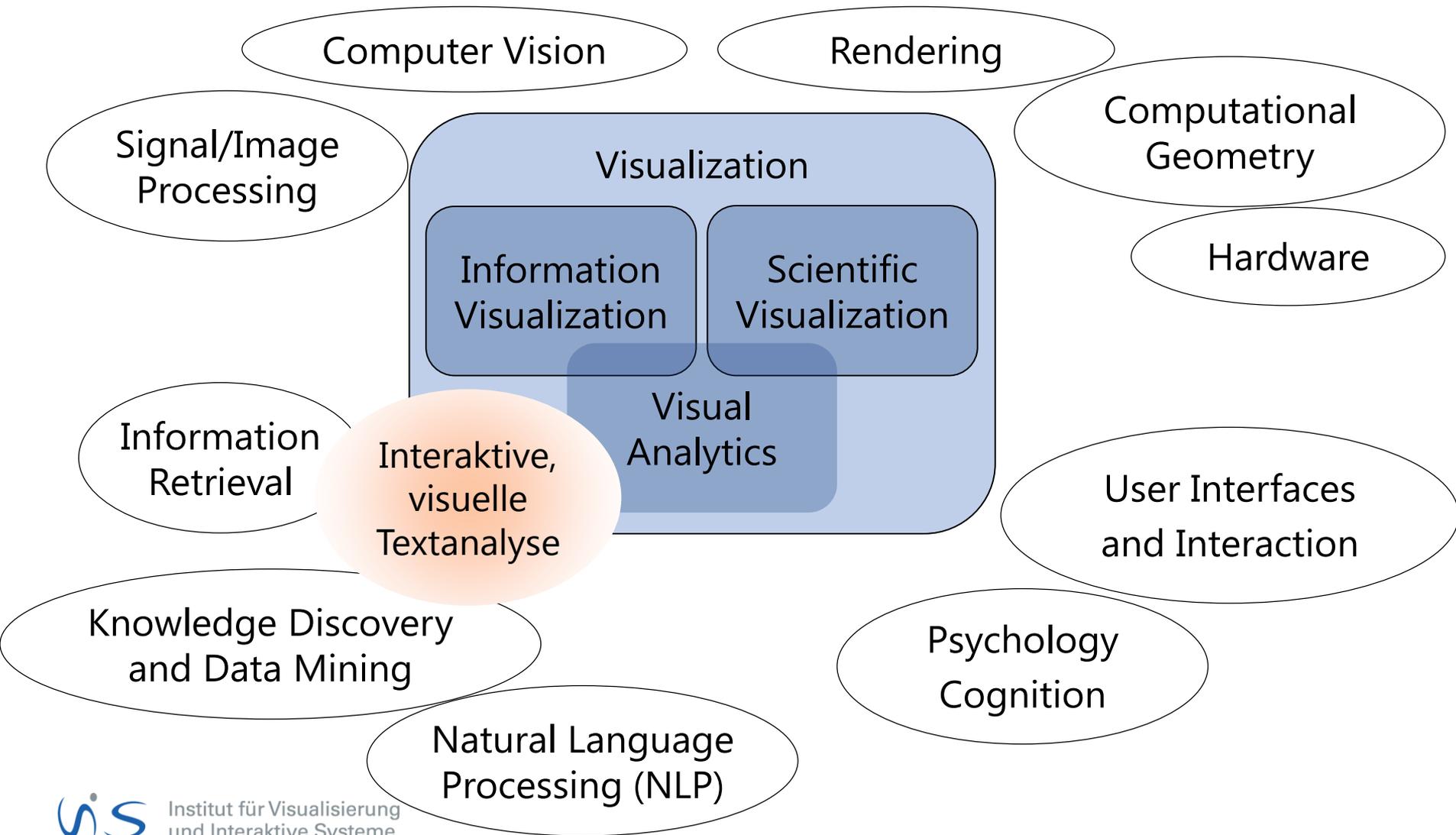
- Bruegels Gemälde zeigt mehr als 100 niederländische Sprichwörter
- 80 sind nachfolgend aufgeführt (in 965 Worten)

1. Den Teufel aufs Kissen binden (überwindet starrköpfig selbst den Teufel)
2. Ein Pfeilerbeißer (scheinheiliger Heuchler)
3. Sie trägt Feuer in der einen, Wasser in der anderen Hand (doppelzüngige Person)
4. Mit dem Kopf durch die Wand (stur und unbedacht Unmögliches versuchen)
5. Scher sie ab, doch schinde sie nicht (nicht um jeden Preis seinen Vorteil suchen)
6. Der eine schert Schafe, der andere Ferkel (der eine lebt im Überfluss der andere in Not)
7. Geduldig wie ein Lamm
8. Dem Mann einen blauen Mantel umhängen (ihn betrügen)
9. Er schüttet den Brunnen zu, nachdem das Kalb ertrunken ist (es wird erst etwas unternommen, wenn es zu spät ist)
10. Rosen (Perlen) vor die Säue werfen (Verschwendung an Unwürdige)
11. Wer durch die Welt will, muss sich krümmen (wer etwas werden will, muss sich anpassen)
12. Er lässt die Welt am Daumen tanzen (er lässt alles nach seiner Pfeife tanzen)
13. Jeder reißt sich um seinen Vorteil
14. Wer seinen Brei verschüttet hat, kann nicht wieder alles einsammeln (der Schaden ist nicht wieder gutzumachen)
15. Sich an der Seite des Geldbeutels festhalten (die Liebe hängt auf der Seite des Geldbeutels)
16. Eine Harke ohne Stiel (etwas Unbrauchbares)
17. Er geht von einem Brot zum anderen (er kommt mit seinem Geld nicht aus)
18. Er sucht das Beilchen (einen Vorwand suchen); er kann sein Licht leuchten lassen (er kann zeigen was er kann)
19. Er brät den Hering wegen des Rogens 1: Eine Kleinigkeit opfern, um einen Vorteil zu erreichen (Mit der Wurst nach der Speckseite werfen). 2: Dort brät der Hering nicht (Es läuft nicht nach Wunsch). 3: geringer Vorteil
20. Hier zieht die Sau den Zapfen raus (Nachlässigkeit rächt sich)
21. Er hängt der Katze die Schelle an (ein Unternehmen ausposaunen und dadurch gefährden)
22. Bis an die Zähne bewaffnet sein
23. Die eine trocknet, was die andere spinnt (etwas weitertratschen)
24. Das Schwein ist durch den Bauch gestochen (etwas ist unwiderrufflich entschieden)
25. Zwei Hunde an einem Bein (Knochen) kommen selten überein (sich erbittert um eine Sache streiten)
26. Er bindet Gott einen flächsernen Bart um (Betrug unter der Maske der Scheinheiligkeit)
27. Sie greift nach dem Hühnerrei und lässt das Gänseei fahren (unüberlegt eine schlechte Wahl treffen)
28. Er versucht das Maul weiter aufzureißen als eine Ofenklappe (er überschätzt seine Fähigkeiten)
29. Er fällt durch den Korb (einen Korb/eine Abfuhr bekommen); er hängt zwischen Himmel und Erde (er befindet sich in einer misslichen Lage)
30. Er sitzt zwischen zwei Stühlen (Er kann keine Entscheidung treffen/ er steht auf keiner Seite)
31. Da hängt die Schere heraus (Beutelschneiderei)
32. Sich um ungelegte Eier kümmern
33. Er trägt das Licht im Korb an den Tag (seine Zeit nutzlos vertun)
34. Kerzen für den Teufel anzünden (sich jedermann zu Freunden machen)
35. Beim Teufel in die Beichte gehen (seinen Feinden Geheimnisse anvertrauen)
36. Was nützt ein schöner Teller, wenn nichts drauf ist?;
37. Er fängt die Fische mit den Händen (er nimmt die Fische aus dem Netz, das die anderen auslegen/er ist ein Schlaumeier)
38. Auf glühenden Kohlen sitzen
39. Verkehrte Welt
40. Er schießt auf die Welt (er verachtet die Welt)
41. Die größten Narren bekommen die besten Karten
42. Wie die Karten fallen (wie man Glück hat)
43. Er pisst gegen den Mond (er strengt sich an, Unmögliches zu erreichen)
44. Jemanden über den Löffel barbieren (mit jemandem Spott treiben)
45. Hinter dem Netz fischen (die Gelegenheit verpassen)
46. Der große Fisch frisst den kleinen
47. Des Nachbarn Wohlergehen mir Kummer macht, mich kränkt, dass die Sonne ins Wasser lacht (Missgunst)
48. Gegen den Strom schwimmen (gegen die Mehrheitsmeinung sein)
49. Den Aal am Schwanz packen (eine schwierige Sache unternehmen)
50. Er schaut durch die Finger (bei einem Vorteil schaut man nicht so genau hin, ob es mit rechten Dingen zugeht)
51. Da hängt das Messer (jemand wird herausgefordert)
52. Da stehen Holzschuhe (man wartet vergeblich)
53. Ein Loch im Dach haben
54. Einen Pfeil dem anderen nachsenden (falsche Beharrlichkeit), seine Pfeile verschossen haben (nichts mehr in Reserve haben)
55. Es wächst zum Fenster hinaus (es kann nicht geheim gehalten werden)
56. Er spielt auf dem Pranger (im Unrecht nicht noch die Aufmerksamkeit auf sich lenken)
57. Vom Ochs auf den Esel fallen (schlechte Geschäfte machen)
58. Er reibt den Hintern an der Tür (sich über etwas hinwegsetzen), sein Päckchen tragen (sein Unglück tragen)
59. Sie schießen durch dasselbe Loch (unzertrennlige Kumpanen); es hängt wie ein Scheißhaus über dem Graben (unzweideutige Angelegenheit)
60. Er wirft sein Geld ins Wasser (sein Geld zum Fenster hinauswerfen, unnütz vergeuden)
61. Seine Kutte über den Zaun hängen (etwas Neues anfangen, ohne zu wissen, was einen erwartet)
62. Er sieht die Bären tanzen (er sieht vor Hunger etwas, was nicht da ist); andere Deutung: Wilde Bären sind gern beieinander (wilde Tiere vertragen sich besser als Menschen)
63. Unterm Besen getraut (ohne Kirchensegen zusammenleben); da steckt der Besen raus (die Herrschaft ist nicht zuhause)
64. Da ist das Dach mit Fladen bedeckt (es herrscht Überfluss)
65. Wenn das Gatter offen ist, laufen die Schweine ins Korn (ohne Aufsicht geht alles drunter und drüber)
66. Wer Feuer frisst, schießt Funken (wer Gefährliches unternimmt, braucht sich über die Folgen nicht zu wundern)
67. Sein Mäntelchen nach dem Wind hängen (Opportunismus)
68. Er küsst den Ring (übertriebene Ehrfurcht)
69. Sie schaut dem Storch nach (seine Zeit vertun)
70. Die Federn in den Wind geschüttet (planlos und erfolglos Arbeiten)
71. Zwei Fliegen mit einer Klappe schlagen (übertriebener Ehrgeiz)
72. Ihm ist gleich, wessen Haus brennt, wenn er sich nur am Feuer erwärmen kann (jede Gelegenheit für einen Vorteil nutzen)
73. Er schleift den Klotz hinter sich her (sich für eine verlorene/unnutzige Sache abrackern)
74. Rossäpfel sind keine Feigen (man soll sich nichts weismachen lassen)
75. Wenn ein Blinder den andern führt, fallen beide in den Graben (wenn ein Unwissender den anderen führt, gibt es ein Unglück)
76. Die Reise ist noch nicht zu End, wenn man Kirch und Turm erkennt (Das Ziel einer Unternehmung wird erst mit Abschluss des Erfolgs erreicht)
77. Vor dem Wind segeln (unter günstigen Umständen leicht einen Erfolg haben)
78. Ich bin zum Gänsehüten nicht berufen, so lass ich Gänse Gänse sein
79. Er beschießt den Galgen (er fürchtet sich vor keiner Strafe)
80. Wo Aas ist, fliegen Krähen

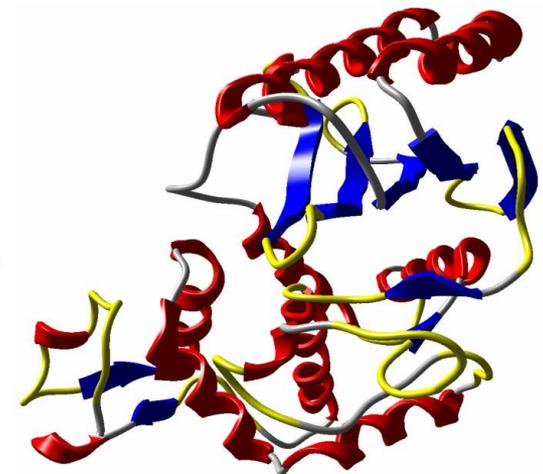
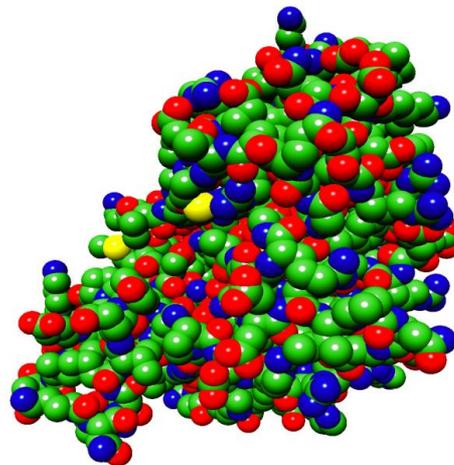
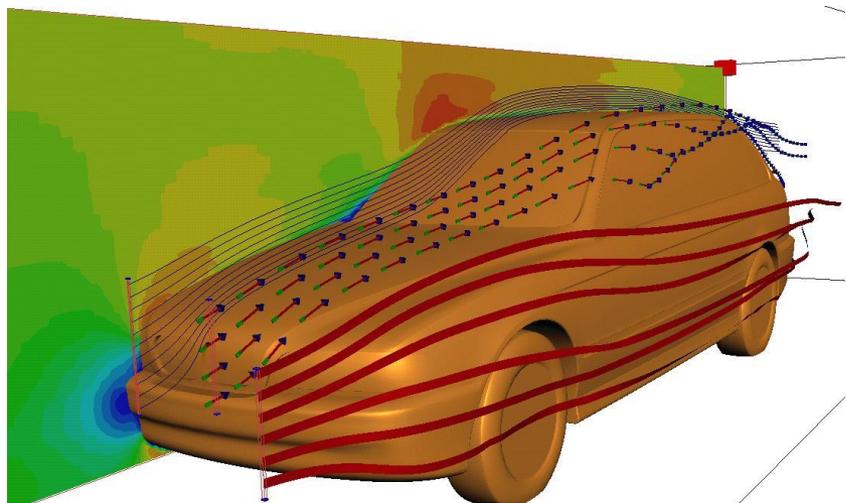
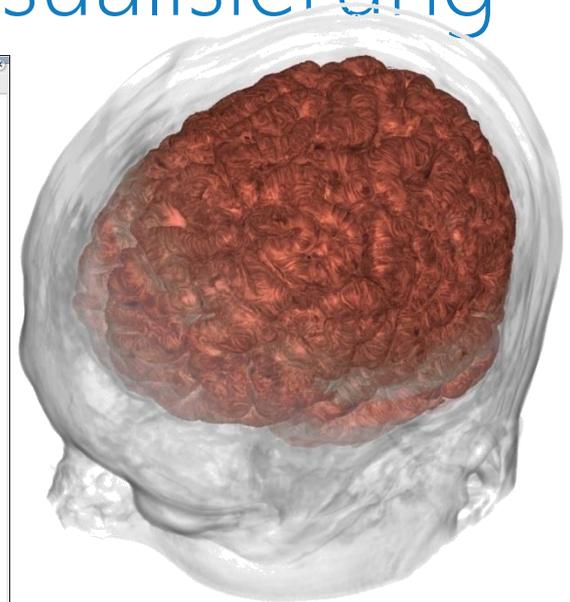
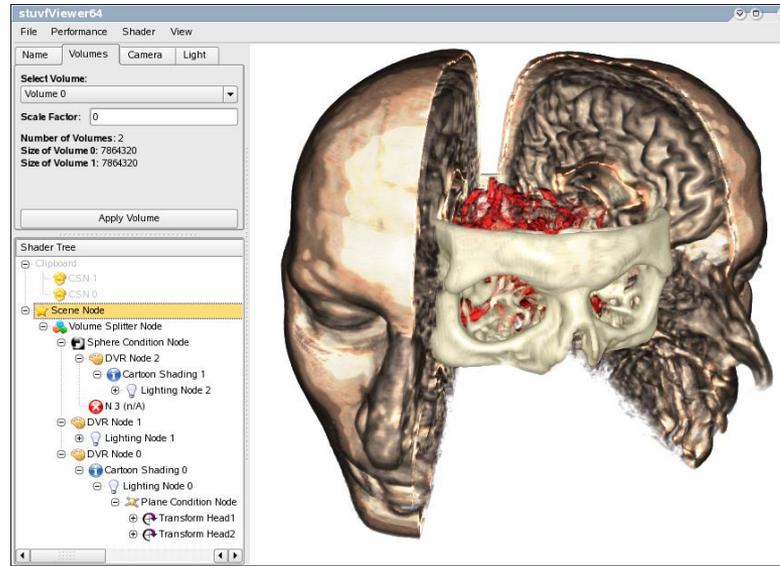
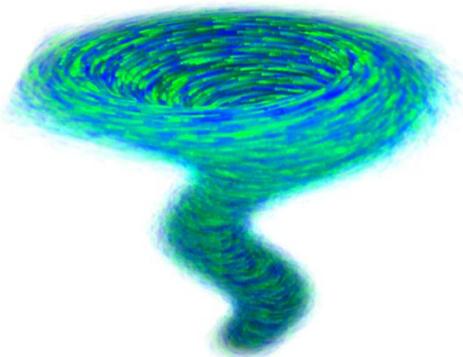
 **Manchmal ja!**

[Quelle: [www.wikipedia.de](http://www.wikipedia.de)]

# Gebiete der Visualisierung



# Beispiele Wissenschaftliche Visualisierung

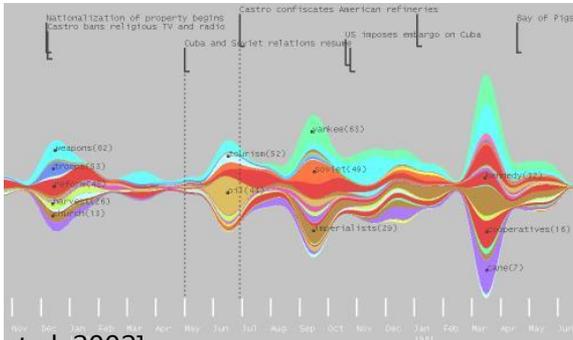


# Beispiele Wissenschaftliche Visualisierung

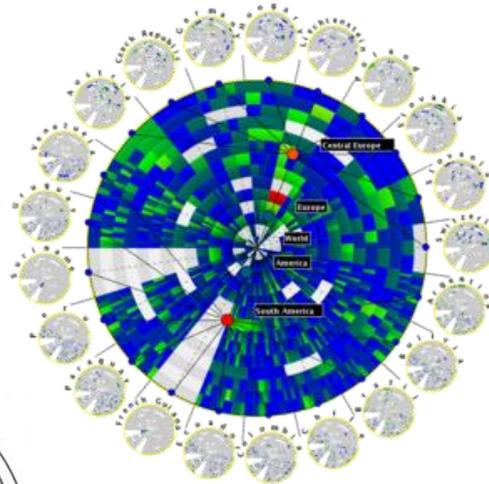
- Hier werden gemessene oder simulierte Daten mit räumlich (zeitlicher) Komponente „direkt“ visualisiert
- Hier trifft der Spruch offensichtlich zu!
- Keiner will, dass der Gehirnchirurg die OP mit Hilfe großer Zahlentabellen oder einer Textbeschreibung plant/ausführt, sondern mit einem CT-Bild

Visualisierung ist hier von Vorteil

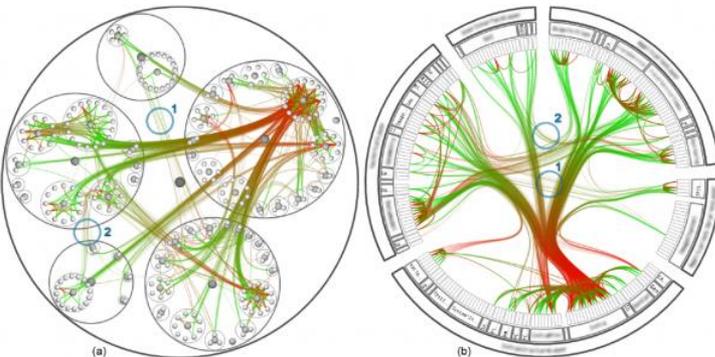
# Beispiele Informationsvisualisierung



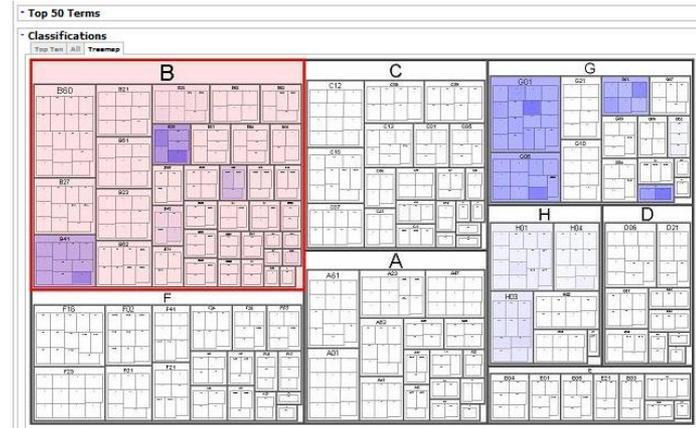
[Havre et al. 2002]



[Burch, Diehl 2008]



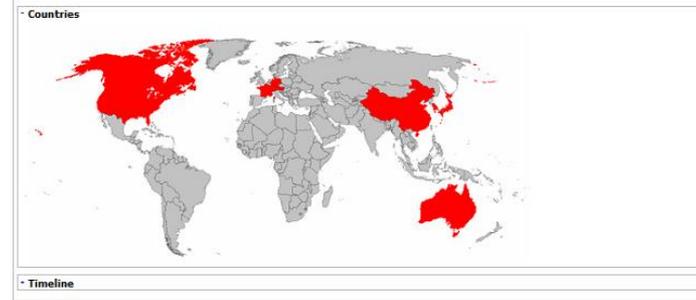
[Holten 2009]



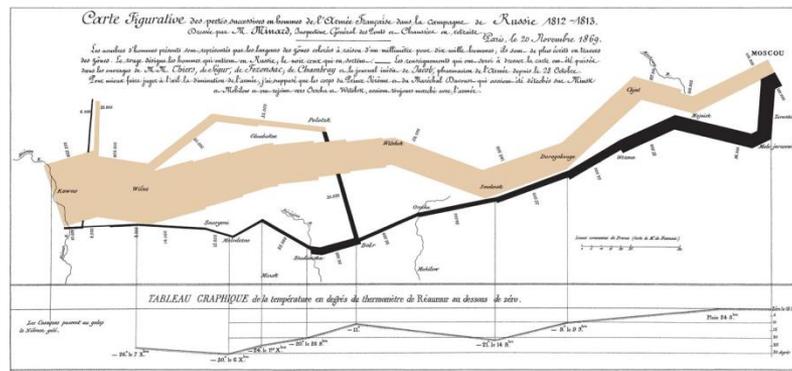
**Applicants**

**Patents**

Publ. Number	Country	Kindcode	Publ. Date	Appl. Number	Title	Exportent
EP0208861A1	EP	B1	19 Feb 1992	EP8720311A	Optisches Datenaufzeichnungsgerät	<a href="#">View</a>
EP0289289A1	EP	B1	20 Jul 1994	EP8830394A	3D-Malzeuerrichtung für einen Laser in einem e...	<a href="#">View</a>
EP0404886A1	EP	B1	10 Apr 1996	EP9120079A	Laser-Verrichtung	<a href="#">View</a>
EP0404886A1	EP	B1	31 Jun 1998	EP9120079A	Vorrichtung zur Regelung eines Aufzeichnungsabstrahl...	<a href="#">View</a>
EP0404886A1	EP	B1	22 Jul 1998	EP9030009A	ACTIVATION SIMULTANÉE DE CAPTES OPTIQUES ENTRAÎNÉES	<a href="#">View</a>
EP0404886A1	EP	B1	29 Nov 1998	EP9120079A	Laserfokussierendes Verrichtung	<a href="#">View</a>
EP0404886A1	EP	B1	19 Nov 2003	EP9120079A	Laser-Verrichtung	<a href="#">View</a>
EP0404886A1	EP	B1	31 Okt 2003	EP9120079A	Laserfokussierendes Verrichtung und mit dieser...	<a href="#">View</a>
EP0404886A1	EP	B1	20 Dec 2006	EP0228510A	Mehrstrahlige Malzeuerrichtung und optische (...)	<a href="#">View</a>

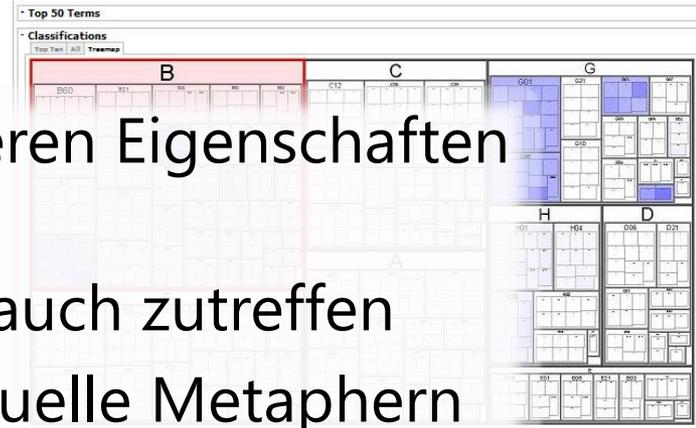
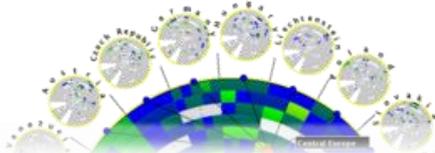
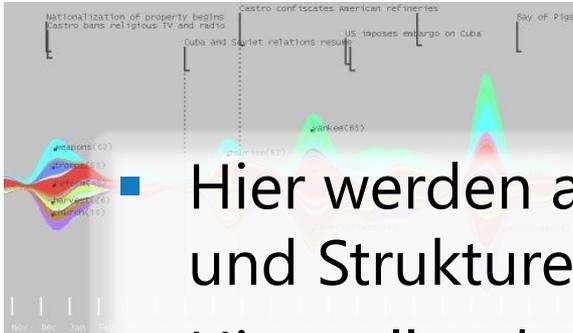


**Timeline**



[Minard 1869]

# Beispiele Informationsvisualisierung



■ Hier werden abstrakte Daten, deren Eigenschaften und Strukturen visualisiert

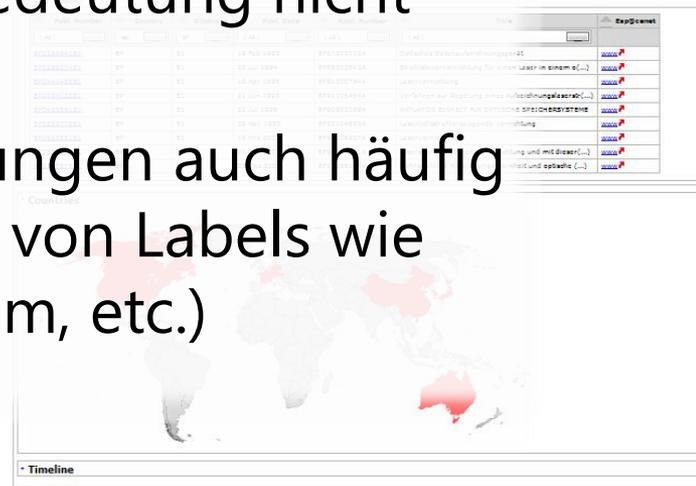
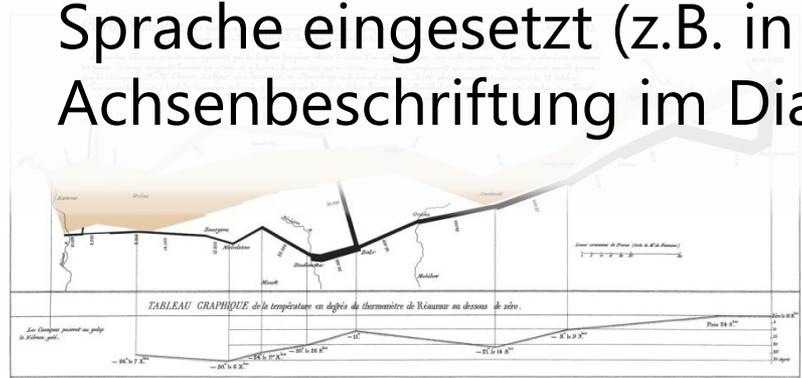
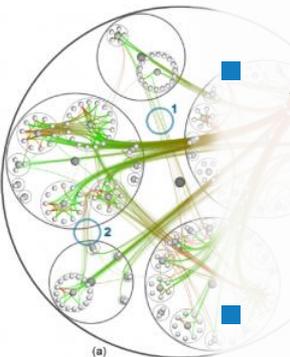
■ Hier sollte der Spruch natürlich auch zutreffen

■ Aber: Visualisierungen deren visuelle Metaphern müssen gelernt werden, da die Bedeutung nicht immer intuitiv verständlich ist

■ Daher wird in diesen Visualisierungen auch häufig Sprache eingesetzt (z.B. in Form von Labels wie Achsenbeschriftung im Diagramm, etc.)

[Havre et al. 2009]

[Holten, van Wijk 2009]



# Visualisierungsziele

## **National Science Foundation ViSC Report 1987:**

**Visualization** is a method of computing. It transforms **the symbolic into the geometric**, enabling researchers to observe their simulations and computations. Visualization offers a method for **seeing the unseen**.

- Präsentation/Kommunikation
  - Überblick geben
  - Komplexe Information zugänglich machen
  - Entscheidungshilfen geben
- Interaktive Exploration
  - Information filtern
  - Visuelle Suche
- Interaktive, visuelle Analyse
  - Visuelle Informationsextraktion
  - Nicht offensichtliche Zusammenhänge visuell begreifbar machen
  - Hypothesen überprüfen
  - Steuerung von automatischen Verfahren

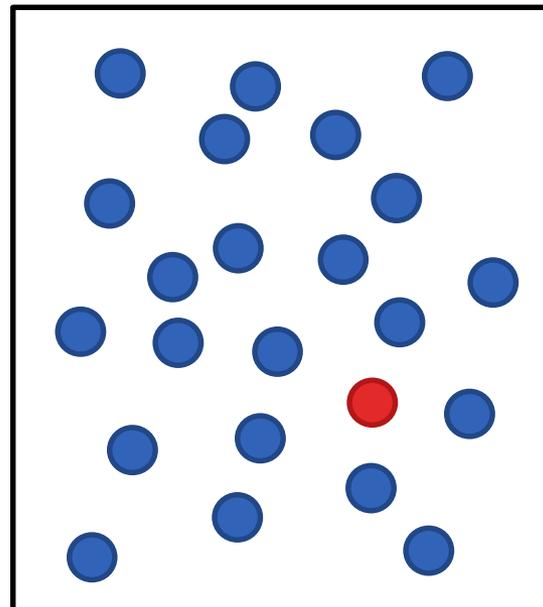
“The purpose of computing is insight not numbers” (Hamming 1962)

# Sprache und Visualisierung

- Ist es sinnvoll Sprache/Text mit Hilfe Visualisierung zu repräsentieren? Wenn ja, in welchen Fällen?
- Gemeinsamkeiten
  - 80% der Informationsaufnahme erfolgt beim (gesunden) Menschen über den Sehsinn
  - Sprache und Visualisierung können zur Kommunikation genutzt werden

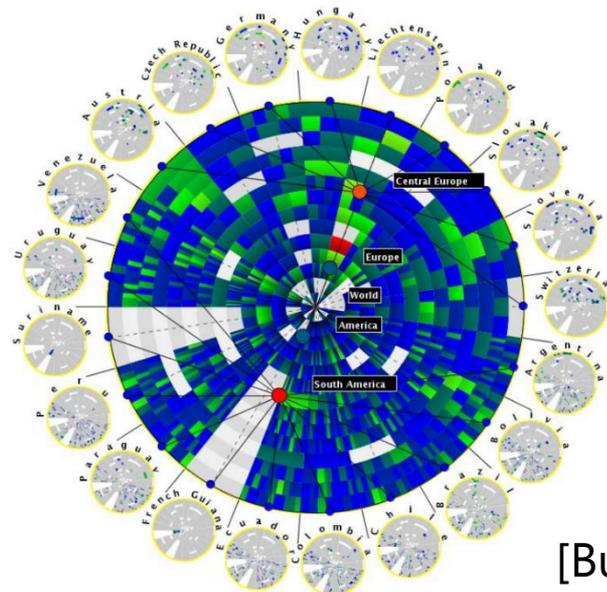
# Sprache und Visualisierung

- Unterschiede
  - Während natürliche Sprache sequentiell aufgenommen wird, können statische Visualisierungen parallel (evtl. präattentiv) wahrgenommen werden



# Sprache und Visualisierung

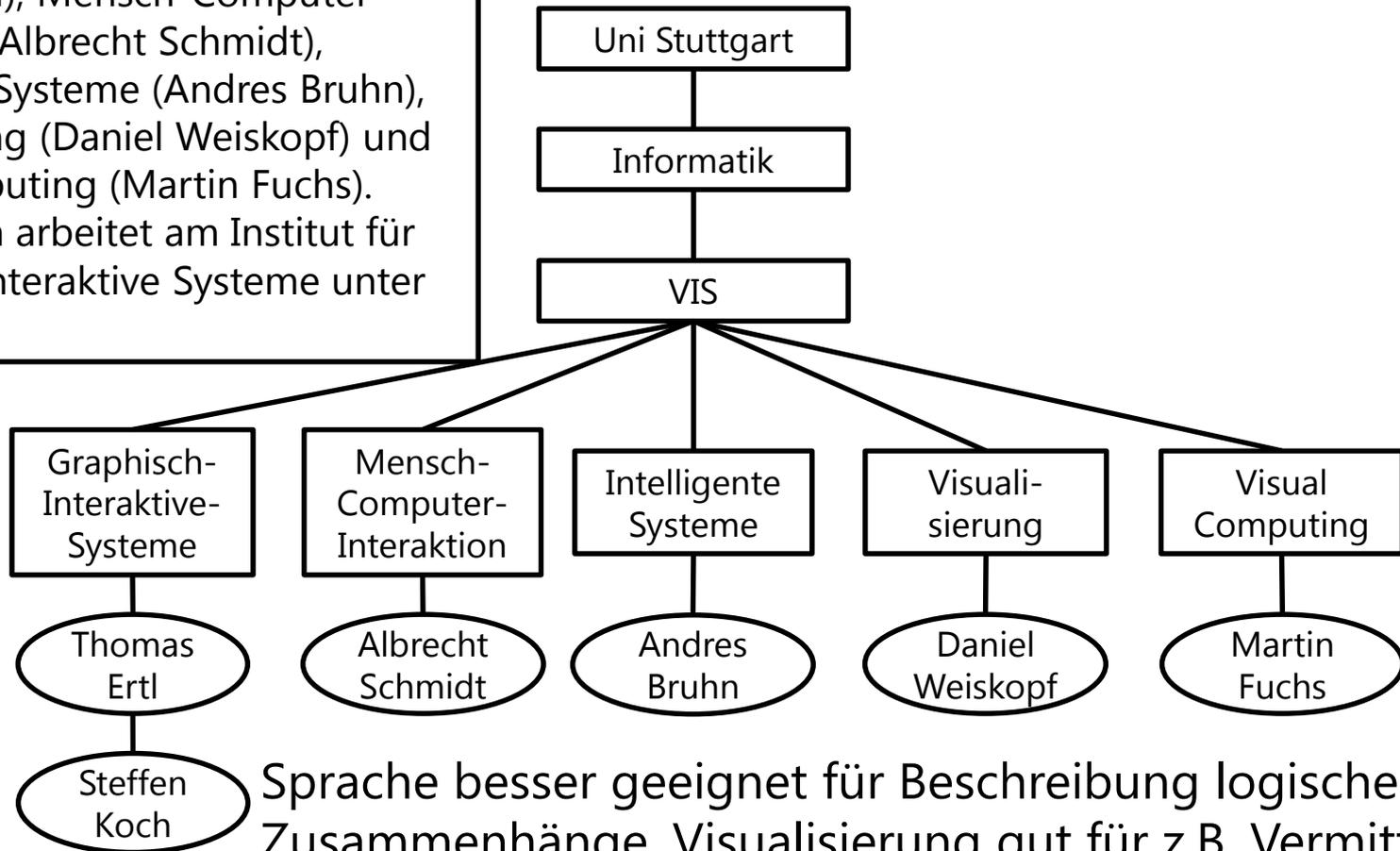
- Unterschiede
  - Während natürliche Sprache sequentiell aufgenommen wird, können statische Visualisierungen parallel (evtl. präattentiv) wahrgenommen werden
  - Sprache ist das am weitesten entwickelte und verbreitete Symbolsystem, für „abstrakte Visualisierung“ gilt das nicht im selben Maße



[Burch, Diehl 2008]

# Sprache und Visualisierung

Das VIS ist ein Informatikinstitut der Uni Stuttgart mit fünf Abteilungen: Graphisch-Interaktive Systeme (Thomas Ertl), Mensch-Computer-Interaktion (Albrecht Schmidt), Intelligente Systeme (Andres Bruhn), Visualisierung (Daniel Weiskopf) und Visual Computing (Martin Fuchs). Steffen Koch arbeitet am Institut für Graphisch-Interaktive Systeme unter Thomas Ertl.



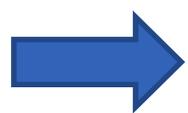
Sprache besser geeignet für Beschreibung logischer Zusammenhänge, Visualisierung gut für z.B. Vermittlung struktureller Eigenschaften

# Sprache und Visualisierung

- Unterschiede
  - Während natürliche Sprache sequentiell aufgenommen wird, können statische Visualisierungen parallel (evtl. präattentiv) wahrgenommen werden
  - Sprache ist das am weitesten entwickelte und verbreitete Symbolsystem, für „abstrakte Visualisierung“ gilt das nicht im selben Maße
  - Sprache besser geeignet für Beschreibung logischer Zusammenhänge, Visualisierung gut für z.B. Vermittlung struktureller Eigenschaften
  - Visualisierung kommt nicht ohne Textbestandteile aus, Sprache aber ohne Visualisierung

# Visuelle Textanalyse

- Ist dann erfolgversprechend, wenn abstrakte, aggregierte, Zusammenhänge dargestellt werden sollen
- Insbesondere interaktive Visualisierung stellt mächtiges Werkzeug dar um Analysen flexibel und „benutzergetrieben“ durchzuführen
- Visual Analytics verbindet die interaktive Visualisierung mit (semi-)automatischer Sprachanalyse



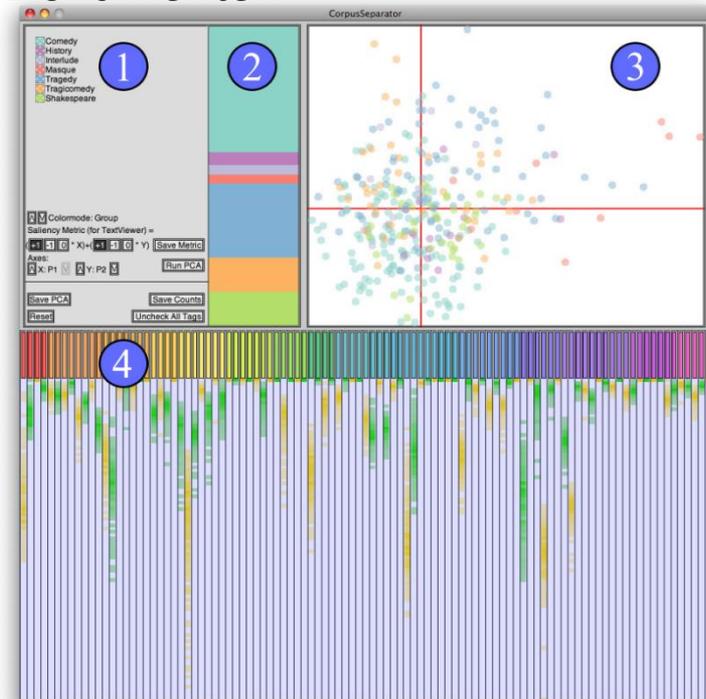
Mächtigeren Analysemöglichkeiten  
(auch auf großen und sehr großen Datenbeständen)

“Overview first, zoom and filter, then details-on-demand.”  
[Shneiderman, 1996]

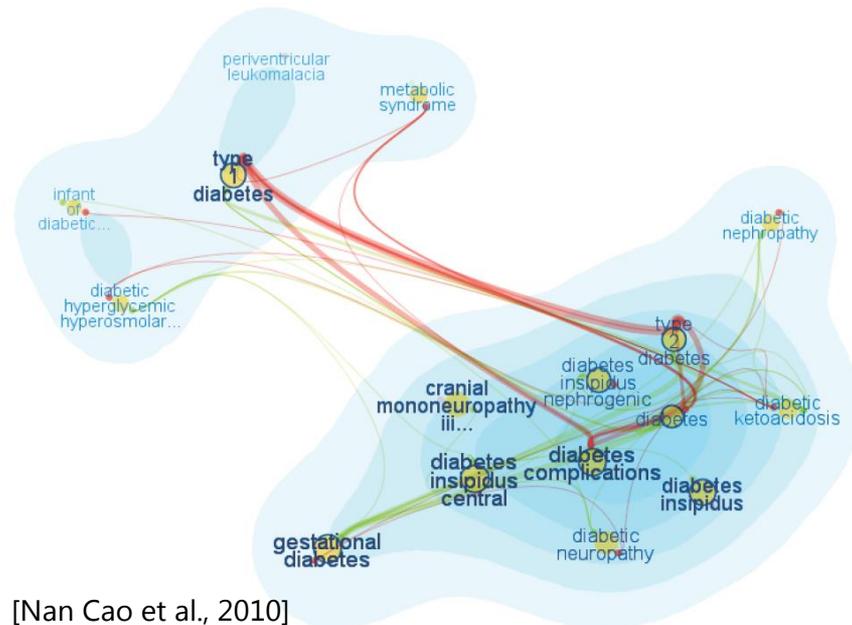
# Textvisualisierung (Overview first,...)

Ansätze für Dokumentkolektionen:

- Dokument Clustering (über Metadaten oder Inhalt)
- Klassifikation von Dokumenten
- Suche und Auswahl von Gruppen von Dokumenten



[Correll et al., 2011]



[Nan Cao et al., 2010]

Ziele:

- Überblick über (Teilaspekte) von Dokumentkolektionen ermöglichen
- Grobe Strukturen zwischen Dokumenten aufzeigen



# Textvisualisierung (... , then details on demand)

- Visualisierung von Strukturen im „Quelltext“
- Ziele: Verifikation von Hypothesen

The screenshot shows the 'Annotated Patent Viewer' interface for patent EP0045183A2. The main text area contains the patent description, with several terms highlighted in yellow and red arrows indicating semantic relationships. For example, 'recording\_material' is linked to 'stabilizing\_layer' via a 'hasPart' relationship. Other terms like 'information\_recording\_layer', 'memory\_medium', and 'laser\_beam' are also highlighted. A 'Contents' sidebar on the right lists sections like 'Abstract', 'Description', and 'Claims', with 'Description' selected. The bottom of the page shows a figure caption: 'Fig. 1 is a cross-sectional view of one form of a recording\_material according to the present in'.

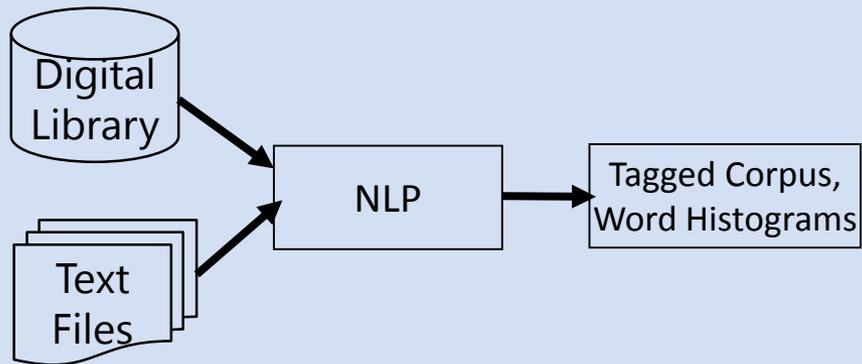
The screenshot shows a 'TextViewer' window displaying a text document. On the right side, there is a vertical visualization consisting of a blue bar and a red line graph, likely representing a structural or semantic analysis of the text. The text itself is a mix of English and Greek, with some words highlighted in red. The interface includes a scroll bar and a search bar at the top.

[Correll et al., 2011]

# Visual Analytics

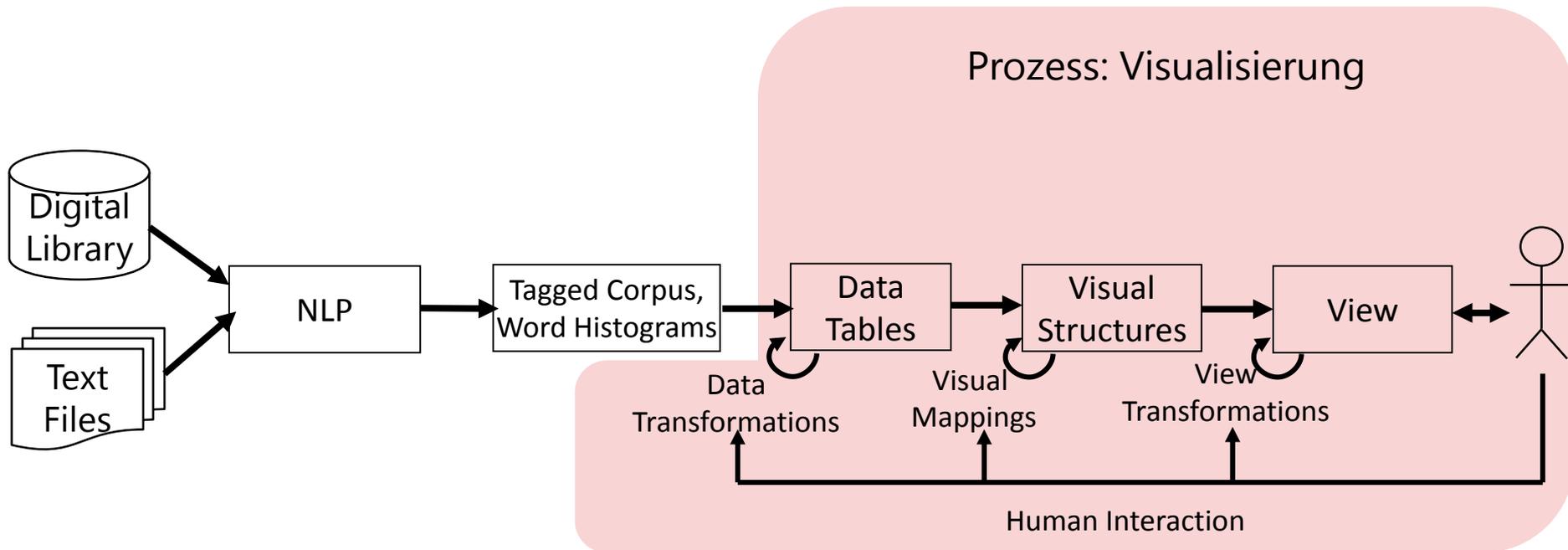
- Sprachanalyse als Vorverarbeitungsschritt

## Prozess: Textanalyse



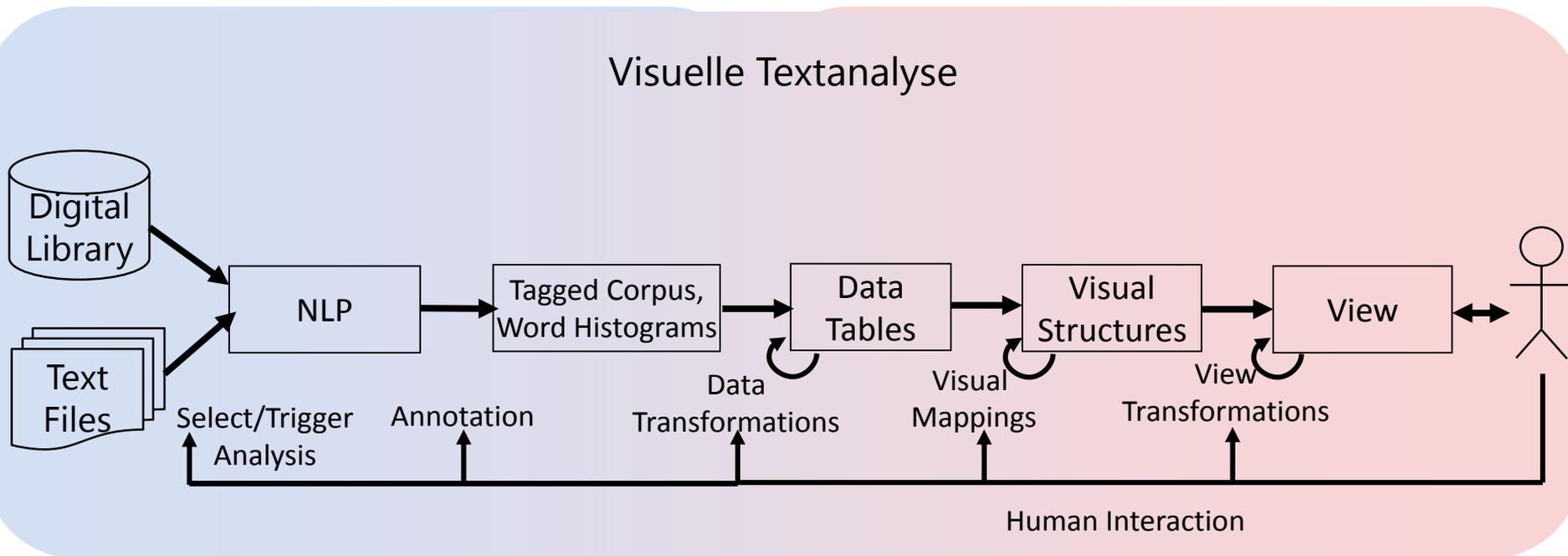
# Visual Analytics

- Sprachanalyse als Vorverarbeitungsschritt
- Entkoppelt von Visualisierungsschritt
- Direkte Einflussnahme auf die Sprachanalyse durch Interaktion nicht möglich



# Visual Analytics

➔ Kopplung zum Visual Analytics System



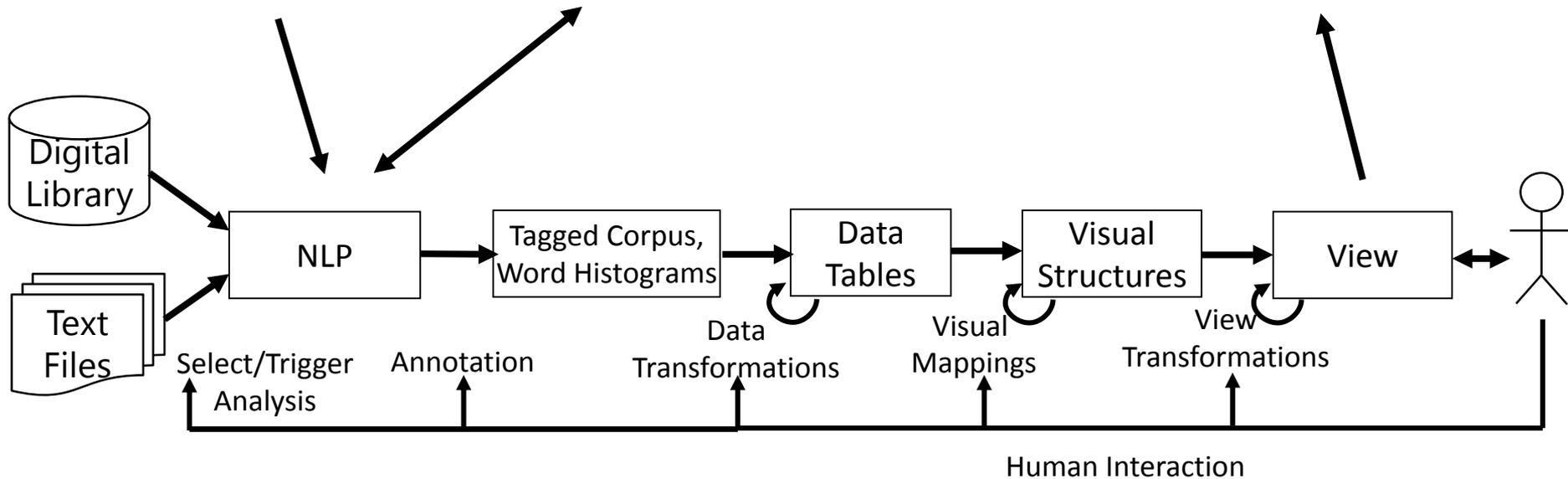
# Einsatzgebiete elektronischer Lexika

- Als interaktives Nachschlagewerk

Als kontrolliert erstellte  
elektronische Ressource

Als automatisch erzeugte  
elektronische Ressource

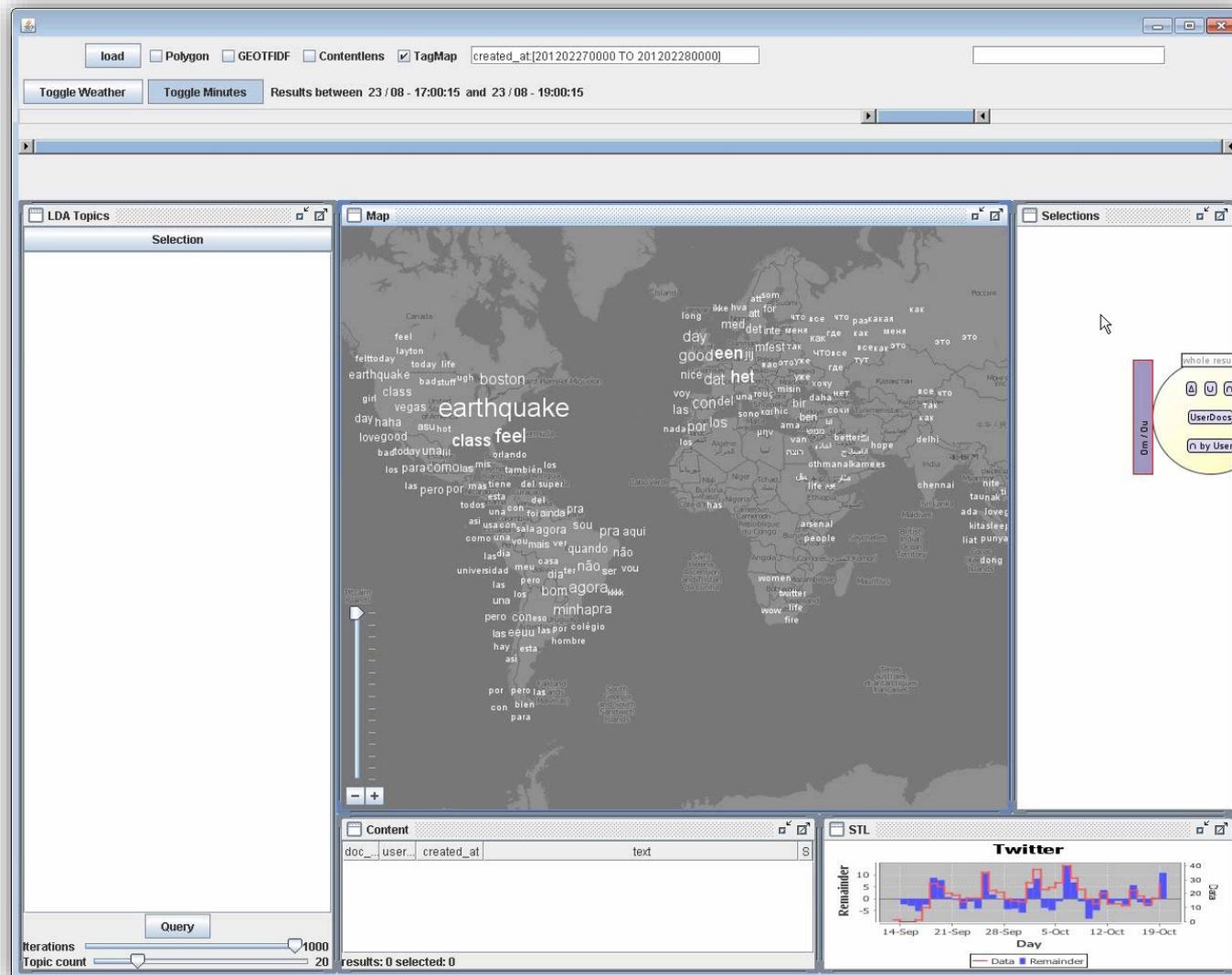
Als visuelle,  
Erklärung



# Lexikon als visuelle Erklärung

- Visualisierung allein genügt oft nicht, um Kontexte von Textanalyseergebnissen zu transportieren
- Einfache Text-Labels in Visualisierungen genügen zur Erläuterung von Textanalyse auch nicht immer
- Zusätzliche Erläuterung/Annotation der Labels kann helfen
- Visualisiert werden auf die Problemstellung passende Wörterbuchsinstanzen (immer automatisch erzeugt)
- „Sortierung“ ist abgestimmt auf die Aufgabe
- Für diese Sichtweise ist eine sehr vage Definition eines Lexikon notwendig: Sammlung von Begriffen/Wortformen/Stichwörtern, denen eine Erklärung, eine Kategorie oder ein Wert zugeordnet ist

# Beispiel „ScatterBlogs“



# Automatisch erzeugtes Wörterbuch

„love“

# Automatisch erzeugtes Wörterbuch

„amore“

# Automatisch erzeugtes Wörterbuch



# Automatisch erzeugtes Wörterbuch



**„liefe“**

# Beispiel Interaktive Klassifikation

**Interactive SVM Training for Document Classification**

File Annotations Selection Tools

corpus

Cluster View of the Most Uncertain Examples

Classifier Visualization, iteration: 3

Document Content / Titles

**title:** Jigsaw: Supporting Investigative Analysis through Interactive Visualization  
**authors:** John Stasko, Carsten Görg, Zhicheng Liu, Kanupriya Singhal  
**corpus id:** 1042

Investigative analysts who work with collections of text documents connect embedded threads of evidence in order to formulate hypotheses about plans and activities of potential interest. As the number of documents and the corresponding number of concepts and entities within the documents grow larger, sense-making processes become more and more difficult for the analysts. We have developed a visual analytic system called Jigsaw that represents documents and their entities visually in order to help analysts examine reports more efficiently and develop theories about potential actions more quickly. Jigsaw provides multiple coordinated views of document entities with a

Back to List

Term Weights

**ten highest changes in weight:**

sequential patterns mine timeline vast manage obtain pattern powerful quickly

**ten highest positive weights:**

text corpus documents document information corpora topic knowledge analysts collections

**ten highest negative weights:**

ambient algorithms multimedia ability spectrum proxy graph astrophysical social conservation

**Training Impact**

Estimated Progress of Training due to Labeled Examples:

Label current selection as

irrelevant clear relevant

Train Classifier with current Labels

**Labeled Documents**

relevant lucene	non relevant lucene	
relevant	non relevant	
doc id	authors	title
252	Weiwei Cui, Shi...	TextFlow: Towa...
841	Daniela Oelke, ...	Visual Evaluati...
249	Ana M. Cuadros...	Point Placeme...
420	Elizabeth C. Hot...	Turning the Dye...

**Classifier History**

model	iterat...	pos / neg la...	pos / neg cl...
my model	0	7 / 14	8 / 1321
model nr 1	1	9 / 14	14 / 1315
model ...	2	11 / 14	25 / 1304
mod...	3	12 / 14	40 / 1289

load selected model

**Hints**

**Legend**

- Black: Currently under the point
- Light Blue: Unlabeled
- Green: Will be on the relevant s
- Red: Will be on the irrelevant
- Purple: Model point of the class
- Triangle: Labeled in this round

# Zusammenfassung

- Visualisierung kann visuelle Text- und Dokumentanalyse unterstützen
- Insbesondere hinsichtlich Aggregation und Explikation struktureller Merkmale
- „Lexika“ spielen auf mehreren Ebenen eine wichtige Rolle für die interaktive Analyse von Texten - auch zur Erläuterung von Analyseergebnissen, Zuständen von automatischen Verfahren und zur Zusammenfassung von Dokumenten

# Kontakt

- Steffen Koch  
[steffen.koch@vis.uni-stuttgart.de](mailto:steffen.koch@vis.uni-stuttgart.de)  
<http://www.vis.uni-stuttgart.de/>
- Thomas Ertl  
[thomas.ertl@vis.uni-stuttgart.de](mailto:thomas.ertl@vis.uni-stuttgart.de)  
<http://vis.uni-stuttgart.de>

# Quellen

Nichtreferenzierte Bilder wurden am Institut für Visualisierung und Interaktive Systeme erzeugt

- [Bruegel 1559] Pieter Bruegel der Ältere, „Die Niederländischen Sprichwörter“, [http://de.wikipedia.org/wiki/Die\\_niederl%C3%A4ndischen\\_Sprichw%C3%B6rter](http://de.wikipedia.org/wiki/Die_niederl%C3%A4ndischen_Sprichw%C3%B6rter) , 1559
- [Burch, Diehl 2008] Burch, Diehl, „TimeRadarTrees: Visualizing Dynamic Compound Digraphs.“ Computer Graphics Forum, vol.27, pp.823-830, 2008
- [Cao, et al., 2010] Nan Cao; Jimeng Sun; Yu-Ru Lin; Gotz, D.; Shixia Liu; Huamin Qu, „FacetAtlas: Multifaceted Visualization for Rich Text Corpora,“ *Visualization and Computer Graphics, IEEE Transactions on* , vol.16, no.6, pp.1172-1181, Nov.-Dec. 2010
- [Correll et al., 2011] Correll, Michael, Michael Witmore, and Michael Gleicher. „Exploring collections of tagged text for literary scholarship.“ *Computer Graphics Forum*. vol. 30, no. 3, 2011
- [van Ham et al., 2009] Van Ham, Frank, Martin Wattenberg, and Fernanda B. Viégas. „Mapping text with phrase nets.“ *Visualization and Computer Graphics, IEEE Transactions on*, vol.15, no.6, pp1169-1176, 2009
- [Havre, et al. 2000] Havre, S.; Hetzler, B.; Nowell, L., "ThemeRiver: visualizing theme changes over time," *InfoVis, IEEE Symposium on* , pp.115-123, 2000
- [Hotlen 2006] Holten, D., "Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data," *Visualization and Computer Graphics, IEEE Transactions on* , vol.12, no.5, pp.741-748, Sept.-Oct. 2006
- [Minard 1869] von Charles Minard, Visualisierung des Feldzugs von Napoleon, <http://commons.wikimedia.org/wiki/File:Minard.png> , 1869
- [Shneiderman, 1996] Shneiderman, Ben. "The eyes have it: A task by data type taxonomy for information visualizations." *Visual Languages, 1996. Proceedings., IEEE Symposium on*. IEEE, 1996.
- [Ware, 2000] Ware, „*Information visualization*“, Morgan Kaufmann, 2000.