

# brief history

(western view)

Chauvet cave  
proto-writing

~20,000 years ago



Sumerian cuneiform  
logographic

~5,000 years ago



Phoenician abjad  
predecessor of alphabet

~3,000 years ago



Latin letters

~2,500 years ago

**ABCDEF**  
**GHIJKLM**  
**NOPQRS**  
**TUVWXYZ**

abstraction

# Text

- Features of Text as representation language
  - abstract
  - general for mental concepts
  - different across population groups (countries, accents, religions,...)
  - linear perception
  - semi-structured (content: grammar, words, sentences, paragraphs,.. ; appearance: typography, calligraphy,..)
  - Legibility **!!!!**

What is the challenge with Text?

Why Text Vis?

## 1.1 Text Visualization

A serious introduction to text visualization has to state that it is not a complete one. Why? When starting to work in the field, researchers are already confronted with the main problem itself, a large collection of documents covering many different aspects related to the subject text. Psychological research e.g. investigates perception and cognition of letters, the psychology of spoken and written language, or the psychology of reading. Linguistics describe inter alia models on language structure, language function, language features, etymology, and linguistic transformations. While both disciplines already fill books and would require introductions by themselves, we did so far not mention visual appearance (typography) or evolution of sign systems. As practical approach, we limit this introduction to key aspects in development of text and text visualizations taking the historic tour (Section 1.1.1), describing psychological backgrounds (Section 1.1.2), and describe landmarks in text visualization (Section 1.1.3). As further simplification we consider written text to stem from an alphabetic system.

### 1.1.1 The historic trail

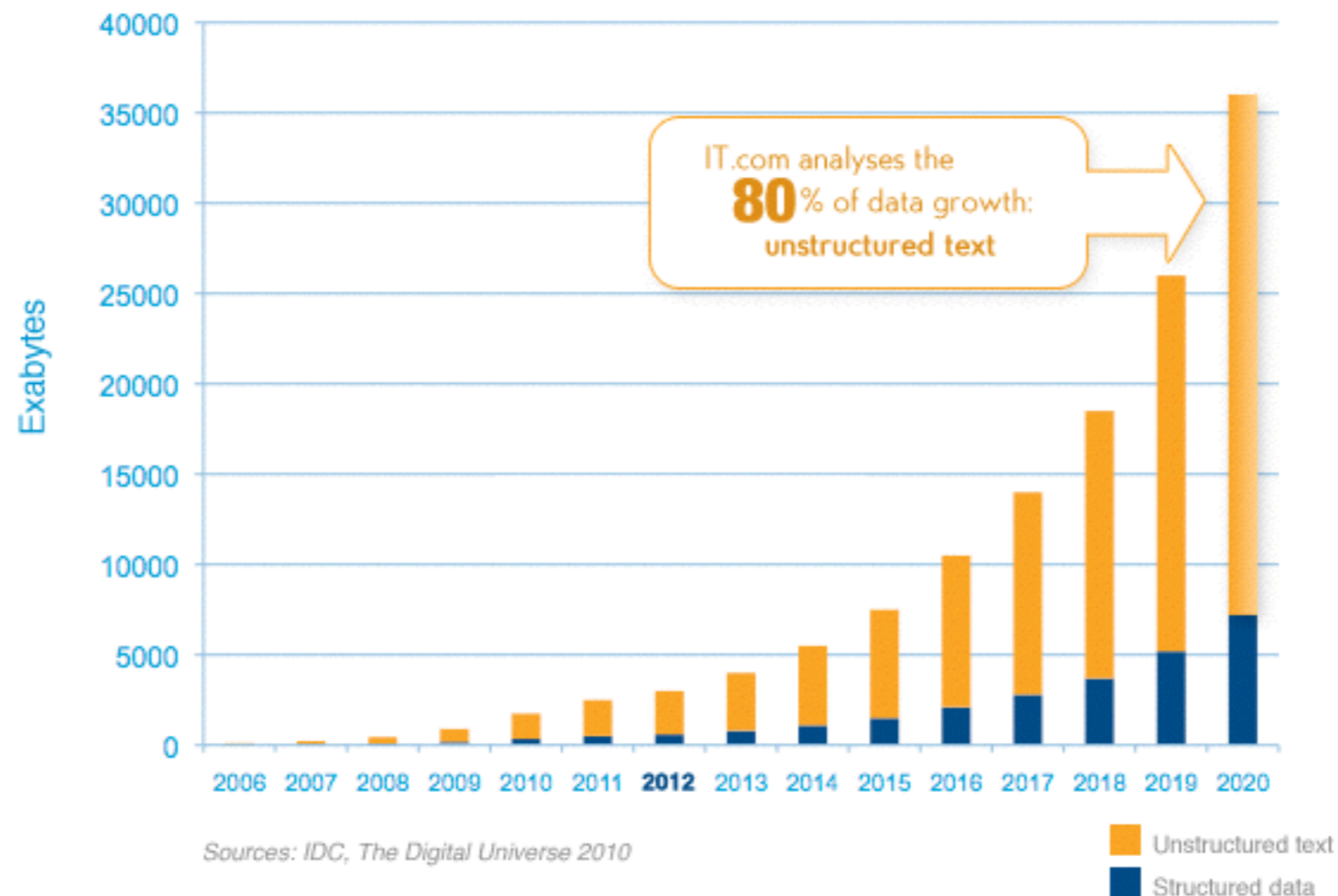
This section relies widely recommendable for further reading. Early humans started representing image and text representation into logographic form. Visual semantics within a language system included 24 signs for writing on papyrus vs. with hieroglyphs to an alphabet ago. Phoenicians have been the first known only-map their ordered set of letters. In Europe, Romans became dominant during the medieval period. The impact on page style and decoration. The industrial revolution. The successors of computers with word-processor and document distribution.

### 1.1.2 The psychological approach

We already discovered that text is nowadays as rapidly produceable and distributable as never before, but we did not throw light on how humans "consume" text. Schönplüg & Schönplüg [SS95] and Rayner & Pollatsek [RP94] provide extensive details on the psychological processes involved in reading which we summarize in this Section.

The consumption of text can be mainly split into reading as the perceptual part and understanding as the cognitive part. For reading, the human visual system performs saccadic eye movement processing lines of text. Each saccade takes on average 20 to 35 ms to bridge a range of 7 to 9

### Worldwide Corporate Data Growth



both references are

The paintings from Chauvet "covered" (Sadier et al. [SDB] story. Divergence between form evolved from pictographic units of meaning elements. Their significance, like the ease of development from most developed 3,000 years ago. Their abjad is excessively, the Greek named

the times of Charlemagne (8th century) printing was already allowed fast reproduction. Or Schnörkel remained as printing machines were content creation. Personal aspects of document production



# Text/Document Visualization

(focused on alphabetical languages)

- Text as Vis
- Vis for Text Documents
- Vis for large Text/Document Corpora
  - for exploring data with visualizations
  - to investigate specific properties
- Text in Vis
- TextVis Specials

# text data type

-no numbers *(implicitly)*

-characters: ASCII

-strings

USASCII code chart

The chart shows a 5-bit binary pattern (b4, b3, b2, b1) for each row and a column index (0-7). The characters are listed in the cells. The bit patterns are: 0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111.

Bits	b4	b3	b2	b1	Column	0	1	2	3	4	5	6	7
0	0	0	0	0	0	NUL	DLE	SP	0	@	P	\	p
0	0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	0	4	EOT	DC4	\$	4	D	T	d	t
0	1	0	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	0	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	0	1	1	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	0	8	BS	CAN	(	8	H	X	h	x
1	0	0	0	1	9	HT	EM	)	9	I	Y	i	y
1	0	0	1	0	10	LF	SUB	*	:	J	Z	j	z
1	0	0	1	1	11	VT	ESC	+	;	K	[	k	{
1	1	0	0	0	12	FF	FS	,	<	L	\	l	
1	1	0	0	1	13	CR	GS	-	=	M	]	m	}
1	1	0	1	0	14	SO	RS	.	>	N	^	n	~
1	1	0	1	1	15	SI	US	/	?	O	_	o	DEL

# text data semantics

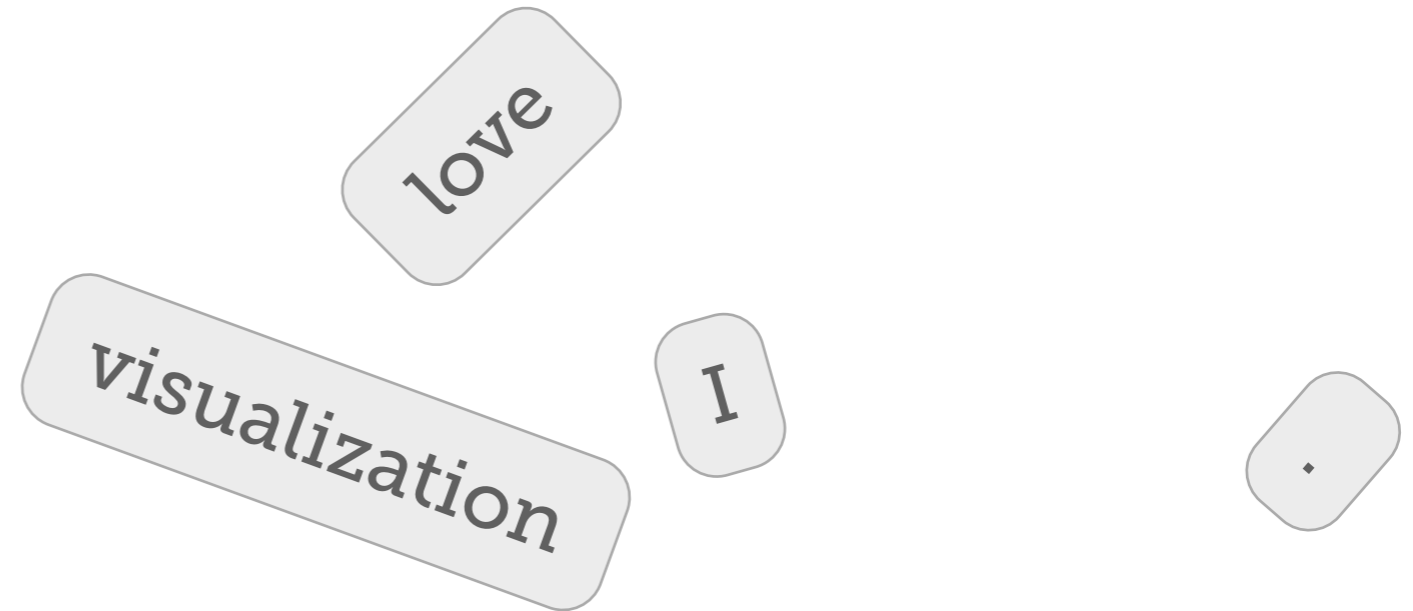
-words

-sentences

-paragraphs

-chapters

-lines



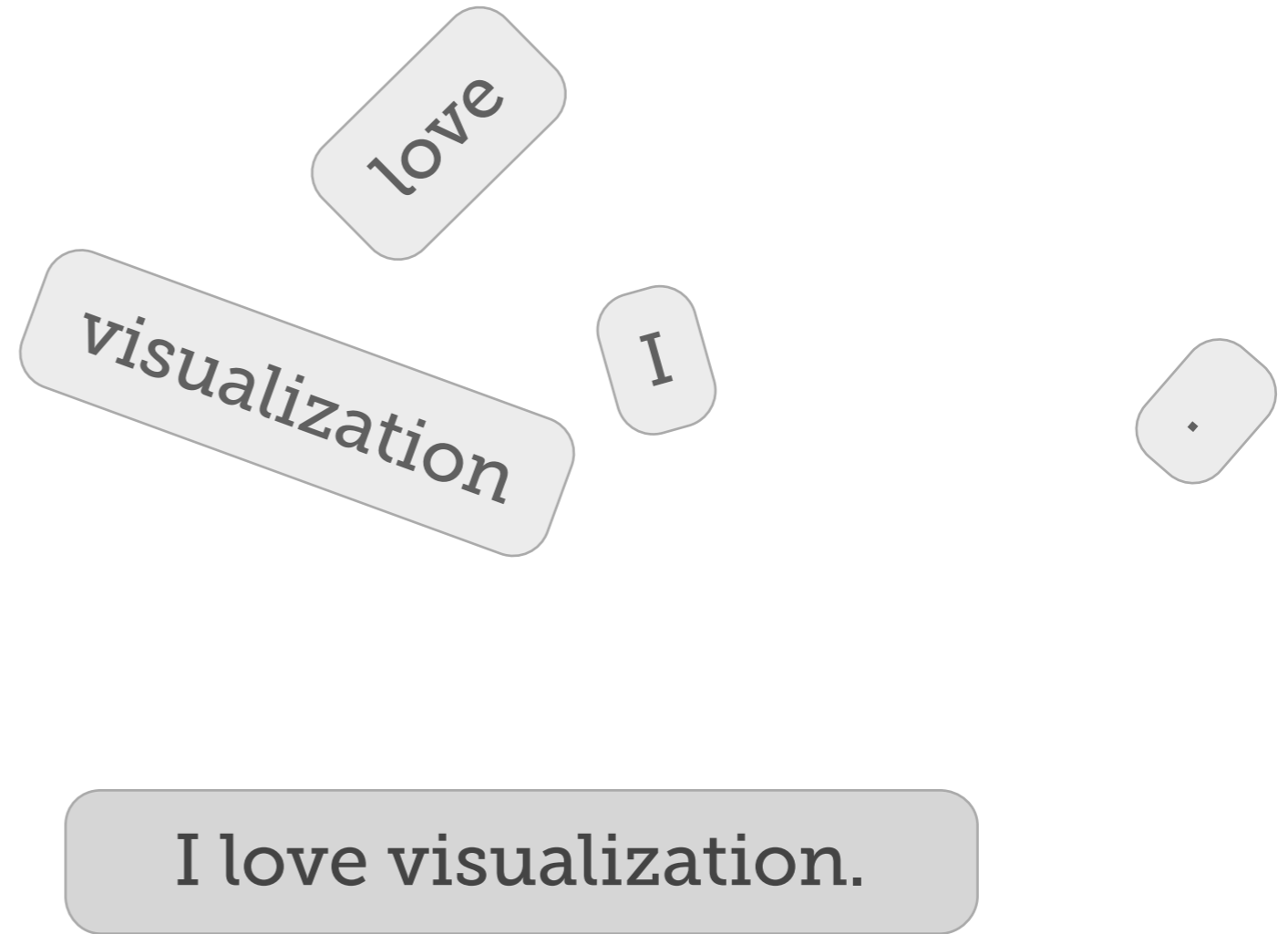
# text data semantics

-words

-sentences

- paragraphs
- chapters

-lines



# text data semantics

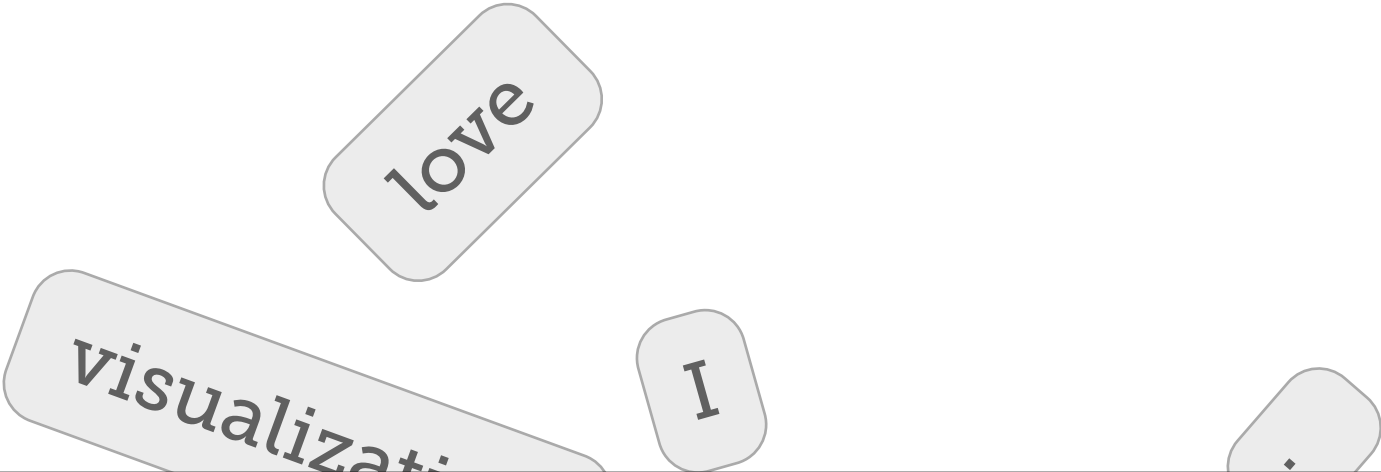
-words

-sentences

-paragraphs

-chapters

-lines



```
16 // displays a data set using parallel coordinates
17
18 // dataset info
19 String dataSet = "cars";
20 String fileName = dataSet + ".csv";
21 boolean cluster = true;
22 FloatTable table;
23 float[][] data;
24
25 // row, column info
26 String[] colNames;
27 int col = 0;
28 int colTot;
29 String[] rowNames;
30 int row = 0;
31 int rowTot;
```

# text data semantics

- documents
- books
- papers
- webpages
- emails
- twitter post

- corpus: collection of documents





# text data semantics

- documents
- books
- papers
- webpages
- emails
- twitter post

- corpus: *collection of documents*



# Text as Vis

- Typography:
  - typefaces (serif, sans-serif, **bold**, *italic*)
  - point size (10pt, 12pt, 24pt, 36pt.. ) - nowadays: 1/72 inch
  - line length (alignment: left, right, justified)
  - vertical: line spacing (leading)
  - horizontal: spaces between groups of letters (tracking)
  - space between pairs of letters (kerning)
  - combining letters to a glyph ligatures



fi → fi  
fl → fl

ß

# Text as Vis

- Creating a font type is an art which requires profound design knowledge
- .. or it can be a science:

Scientists have developed a way to carve shapes from DNA canvases, including all the letters of the Roman alphabet, emoticons and an eagle's head.

Bryan Wei, a postdoctoral scholar at Harvard Medical School in Boston, Massachusetts, and his colleagues make these shapes out of single strands of DNA just 42 letters long. Each strand is unique, and folds to form a rectangular tile. When mixed, neighbouring tiles stick to each other in a brick-wall pattern, and shorter boundary tiles lock the edges in place. [...]



<http://www.nature.com/news/dna-drawing-with-an-old-twist-1.10742>

# Text as Vis

- Typesetting:
  - letterpress printing
  - Linotype machine
  - digital printing/copying (typewheel, dot-matrix, inkjet, laser)
  - digital text (resolution is key: **s m a l l** -> retina)
- Encoding text for electronic devices:
  - mapping each character to a sequence of bytes
  - Universal Character Set (UTF-[**8**,16,32]) fonts
  - exchange of typeset documents: PostScript and PDF

# Text as Vis

- rules of thumb:
  - limit the use of fonts to only a few typefaces !!
  - use “special” fonts only when appropriate
  - a good resource for fonts in web projects are [google fonts](#)

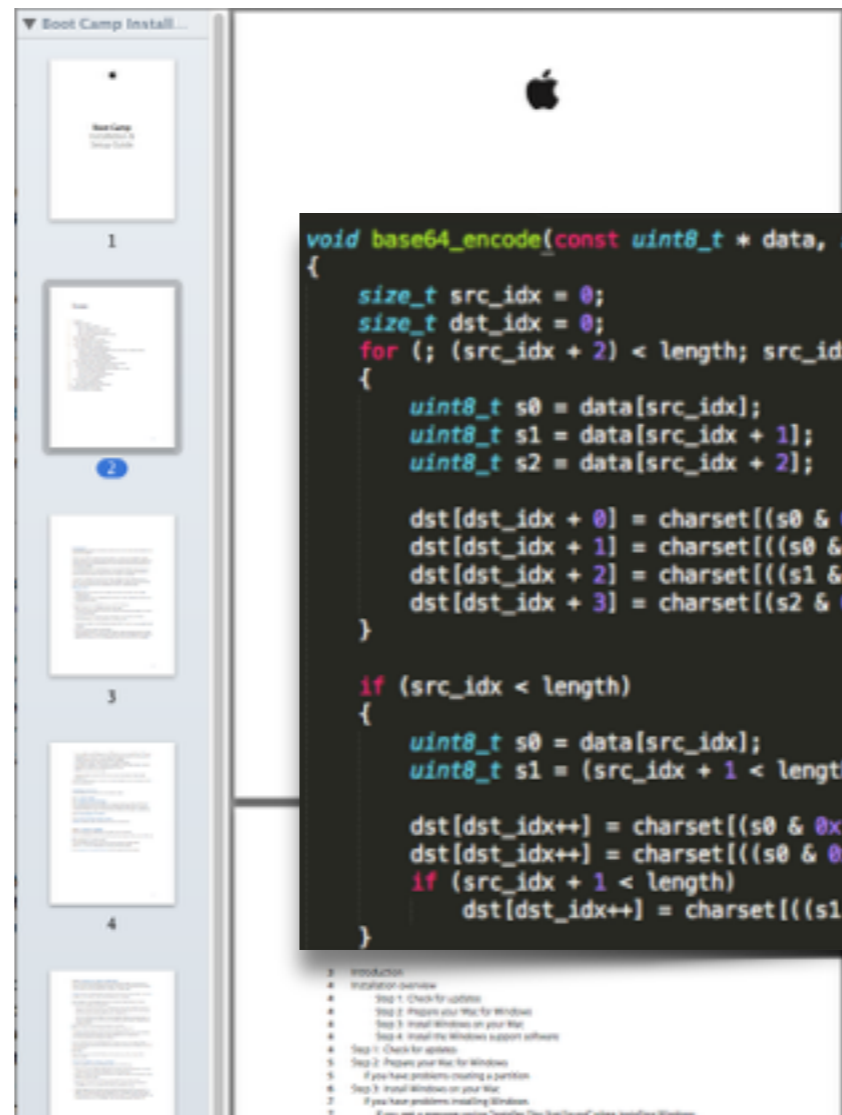




# Visualization for “Raw” Text

- in daily use..  
enriched text - hypertext  
linking (graph navigation)

overview & detail



```
void base64_encode(const uint8_t * data, size_t length, char * dst)
{
    size_t src_idx = 0;
    size_t dst_idx = 0;
    for (; (src_idx + 2) < length; src_idx += 3, dst_idx += 4)
    {
        uint8_t s0 = data[src_idx];
        uint8_t s1 = data[src_idx + 1];
        uint8_t s2 = data[src_idx + 2];

        dst[dst_idx + 0] = charset[(s0 & 0xfc) >> 2];
        dst[dst_idx + 1] = charset[((s0 & 0x03) << 4) | ((s1 & 0xf0) >> 4)];
        dst[dst_idx + 2] = charset[((s1 & 0x0f) << 2) | (s2 & 0xc0) >> 6];
        dst[dst_idx + 3] = charset[(s2 & 0x3f)];
    }

    if (src_idx < length)
    {
        uint8_t s0 = data[src_idx];
        uint8_t s1 = (src_idx + 1 < length) ? data[src_idx + 1] : 0;

        dst[dst_idx++] = charset[(s0 & 0xfc) >> 2];
        dst[dst_idx++] = charset[((s0 & 0x03) << 4) | ((s1 & 0xf0) >> 4)];
        if (src_idx + 1 < length)
            dst[dst_idx++] = charset[((s1 & 0x0f) << 2)];
    }
}
```

- 1. INTRODUCTION
- 2. INSTALLATION OVERVIEW
- 3. STEP 1: CHECK FOR UPDATES
- 4. STEP 2: PREPARE YOUR MAC FOR WINDOWS
- 5. STEP 3: INSTALL WINDOWS ON YOUR MAC
- 6. STEP 4: INSTALL THE WINDOWS SUPPORT SOFTWARE
- 7. STEP 5: CHECK FOR UPDATES
- 8. STEP 6: PREPARE YOUR MAC FOR WINDOWS
- 9. IF YOU HAVE PROBLEMS CREATING A PARTITION
- 10. STEP 7: INSTALL WINDOWS ON YOUR MAC
- 11. IF YOU HAVE PROBLEMS INSTALLING WINDOWS



# Visualization for “Raw” Text

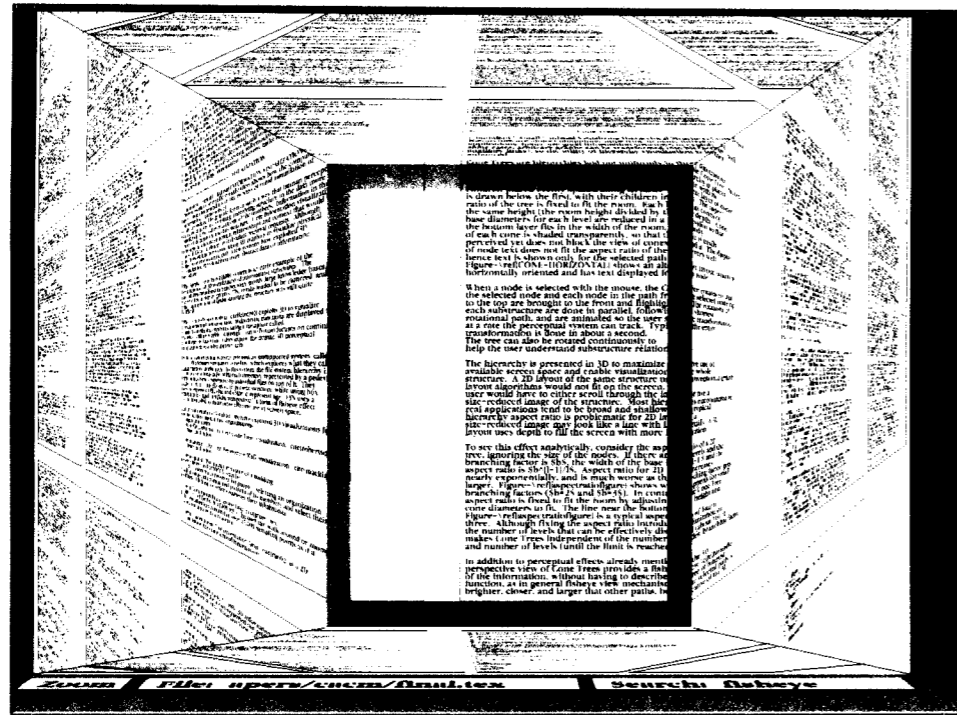


Figure 3: Document Lens with lens pulled toward the user. The resulting truncated pyramid makes text near the lens' edges readable.

Robertson, George G., and Jock D. Mackinlay  
**The document lens**  
*Proceedings of the 6th annual ACM symposium on User interface software and technology*. ACM, 1993.

Eurographics Conference on Visualization (EuroVis) 2012  
S. Bruckner, S. Miksch, and H. Pfister  
(Guest Editors)

Volume 31 (2012), Number 3

## Document Thumbnails with Variable Text Scaling

A. Stoffel and H. Strobel and O. Deussen and D. A. Keim

University of Konstanz, Germany

### Abstract

Document reader applications usually offer an overview of the layout for each page as thumbnail view. Reading the text in these becomes impossible when the font size becomes very small. We improve the readability of these thumbnails using a distortion method, which retains a readable font size of interesting text while shrinking less interesting text further. In contrast to existing approaches, our method preserves the global layout of a page and is able to show context around important terms. We evaluate our technique and show application examples.

### 1. Motivation

#### The user interface of

such as Adobe Reader, consists of a detail view and one or more views for navigation within documents, such as a table of contents and a thumbnail view providing page previews. In addition, most document viewers offer a keyword search functionality where the occurrence of keywords is highlighted in the detail view. However, the navigation views of document viewers (e.g. thumbnails) typically do not show the occurrence of keywords in the documents.

So the user has to step through all occurrences of the keyword within the detail view as scrolling the pages.

To avoid this, we propose to highlight the keywords in the thumbnail view. Using the thumbnail view reduces the

and the user is pointed pages. In addition, thumbnails can be useful for retrieval if the users are trying

know [CVDRH99, DC02]. Due to the small size of text in thumbnails, the highlighting should in addition increase the size of the keywords and their context, at first to make the text better readable and second to allow a simple disambiguation of keywords by their context. For instance, it

about “user” or “user interface” keyword “user” would

The technique we present to create the thumbnails is a general distortion technique for document content that high-

to a user defined interest

The global structure of a page, namely the position of images and columns, is preserved. An example is shown in Figure 1. In the keyword search application, an interest function

submitted to Eurographics Conference on Visualization (EuroVis) (2012)

is used that highlights the keywords and their context. Other applications might use a different interest function, for instance a sentiment score could be used to create thumbnails for sentiment analysis.

### 2. Related Work

Three different techniques are currently used for handling document overview and navigation: abstraction from the document with pixel based representations, thumbnails with different highlighting techniques, and semantic zooming.

A common pixel based technique is TileBars [Hea95], which visualizes the length of documents and the distribution of search terms within these documents with a rectangular pixel-based visualization. Byrd [Byr99] combines the scrollbar of the document view with a pixel visualization of

allowing the user to scroll

rence of the terms. Both techniques do not show the context and a user has to

order to access the context of the search terms.

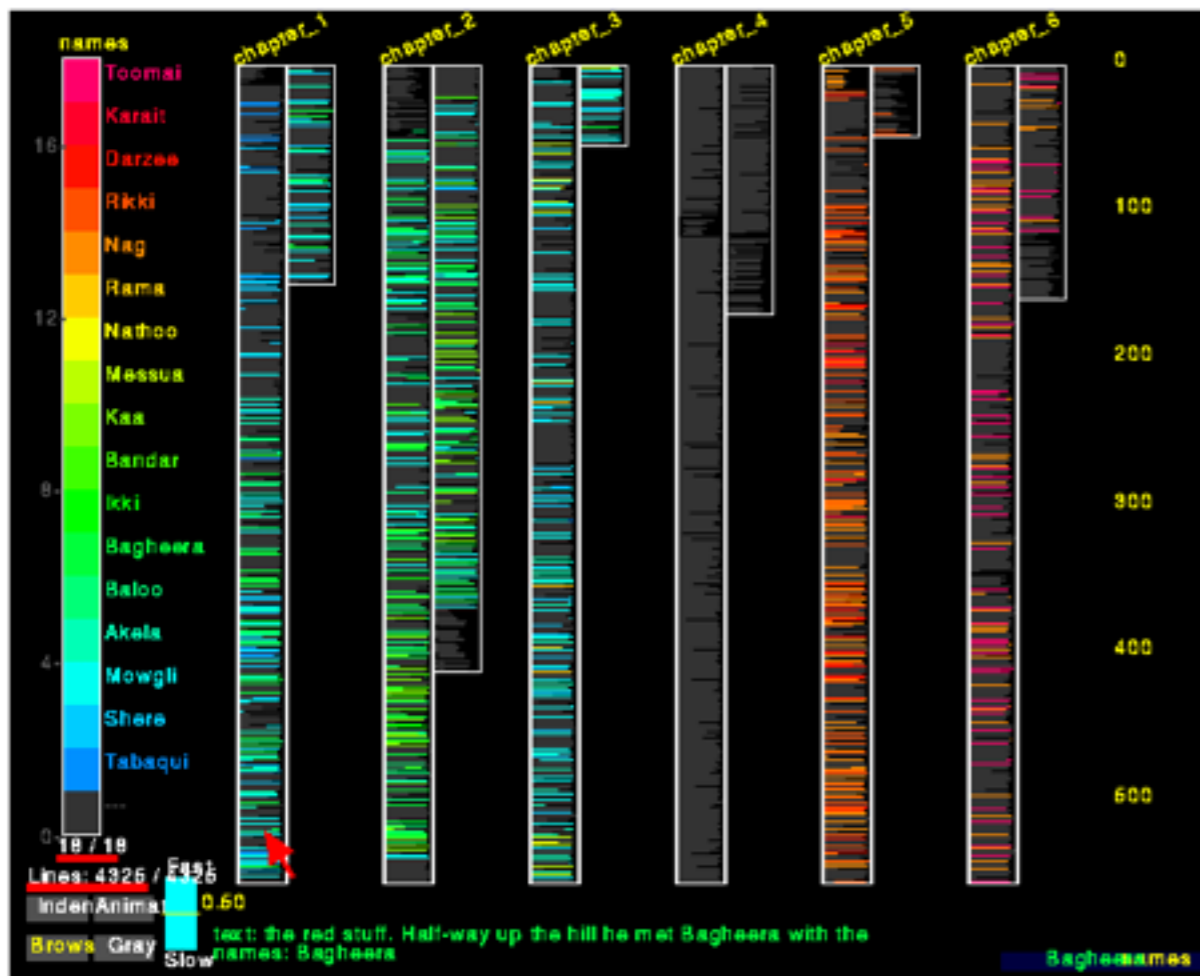
Thumbnails, small version of the document or page, are commonly used for overview and navigation. The space-filling thumbnail approach of Cockburn et al. [CGA06] avoids scrolling in the overview of a document, by positioning the thumbnails of all pages on a grid on the screen and resizing the thumbnails to fit the window size. Suh et al. [SWRG02] combined the thumbnails with popouts, which highlight search terms by rendering them in a readable size with a semi-transparently colored background above of the original thumbnail. Woodruff et al. [WRM02] pre-

## Document Thumbnails with Variable Text Scaling

A. Stoffel, H. Strobel, O. Deussen, D. A. Keim

*Computer Graphics Forum*, volume 31 issue 3 pp.

# Visualization for “Raw” Text

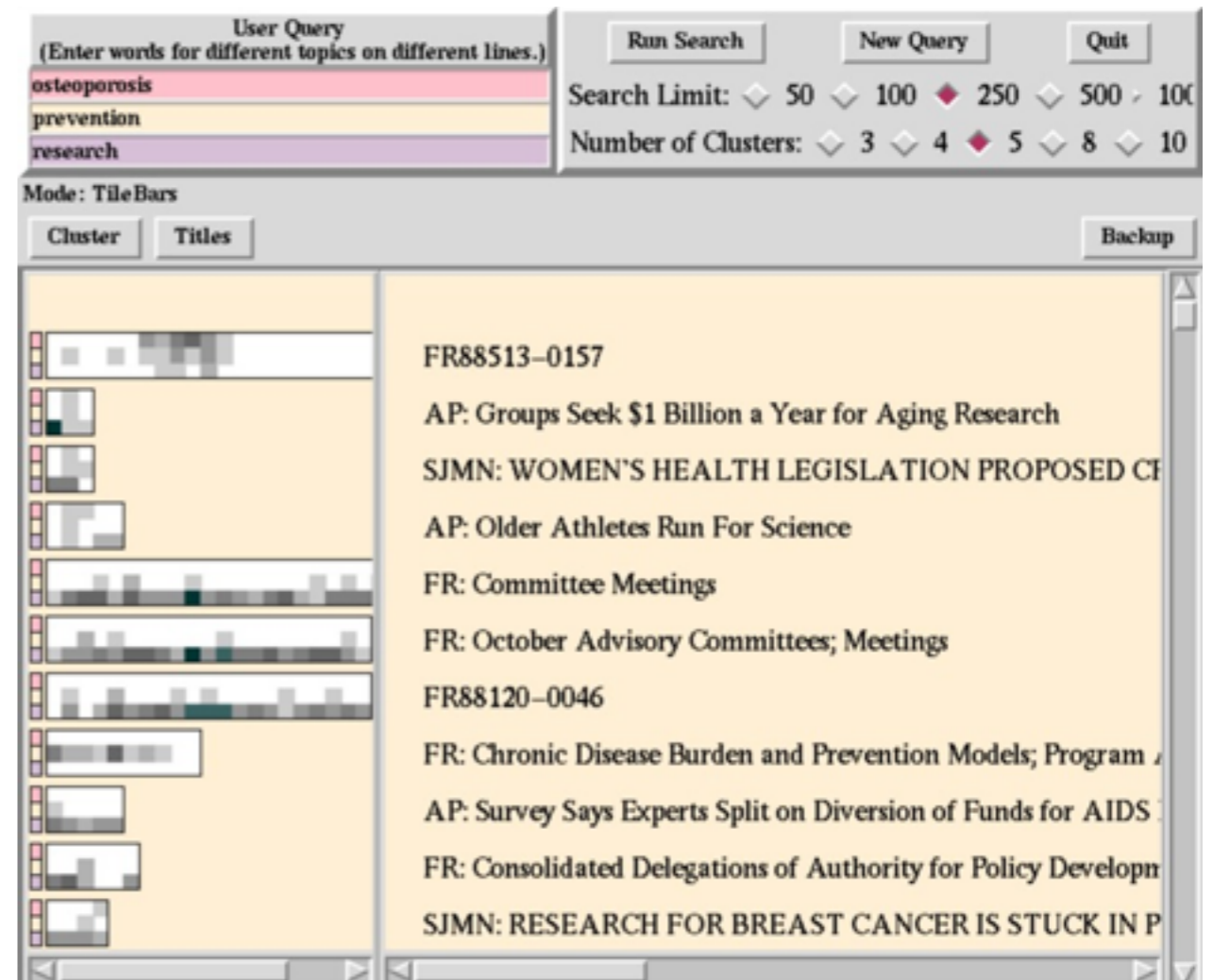


SeeSoft

Stephen G. Eick.

Graphically displaying text.

*Journal of Computational and Graphical Statistics*, 3(2):127-142, June 1994.



TileBars: Visualization of Term Distribution Information in Full Text

Marti Hearst

*Information Access, Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), Denver, CO, 1995*

Visualizing text (features)  
requires a transformation step:  
discretization, aggregation, normalization,...

unstructured text



4 x 't'  
3 x 'u'  
2 x 'r'  
2 x 'e'  
...

structured data

# Structured Text Features

- simple counts
- or a bag of words (used for similarity measures):

	<b>princess</b>	<b>dragon</b>	<b>castle</b>
<b>doc1</b>	1	1	1
<b>doc2</b>	0	0	1

# Typical Steps of Processing to derive Text Features

- Large collections require pre-processing of text to extract information and align text. Typical steps are:
  - cleaning (regular expressions)
  - sentence splitting
  - change to lower case
  - stopword removal (most frequent words in a language)
  - stemming - demo porter stemmer
  - POS tagging (part of speech) - demo
  - noun chunking
  - NER (name entity recognition) - demo opencalais
  - deep parsing - try to “understand” text.

# Sample Text

KIEV, Ukraine — Struggling to reach a deal to form a new majority coalition in Parliament, and under excruciating pressure because of a looming economic disaster, the Ukrainian lawmakers temporarily running the country on Tuesday delayed until Thursday the naming of an acting prime minister and a provisional government.

The delay underscored the extreme difficulty that lawmakers now face in rebuilding the collapsed government left behind when President Viktor F. Yanukovich fled Kiev on Saturday and was removed from power in a vote supported by some members of his own party.

The three main opposition parties, which share little in common politically, have been in fierce negotiations, not just among themselves, but also with civic activists and other groups representing the many constituencies involved in Ukraine's three months of civic uprising.

Arseniy P. Yatsenyuk, the leader in Parliament of the Fatherland Party and a leading contender to serve as acting prime minister, pleaded with colleagues to swiftly reach an agreement on the designation of an interim government, which is needed to formally request emergency economic assistance from the International Monetary Fund.



# Text features are complicated

- Be aware!! text understanding can be hard:
  - *Toilet out of order. Please use floor below.*
  - *“One morning I shot an elephant in my pajamas. How he got in my pajamas, I don't know.”*
  - *Did you ever hear the story about the blind carpenter who picked up his hammer and saw?*

# Was that irony? - Noooo

Profanity sucks. (14)

Be more or less specific. (15)

Analogies in writing are like feathers on a snake. (19)

*excerpt from Rules of Writing*

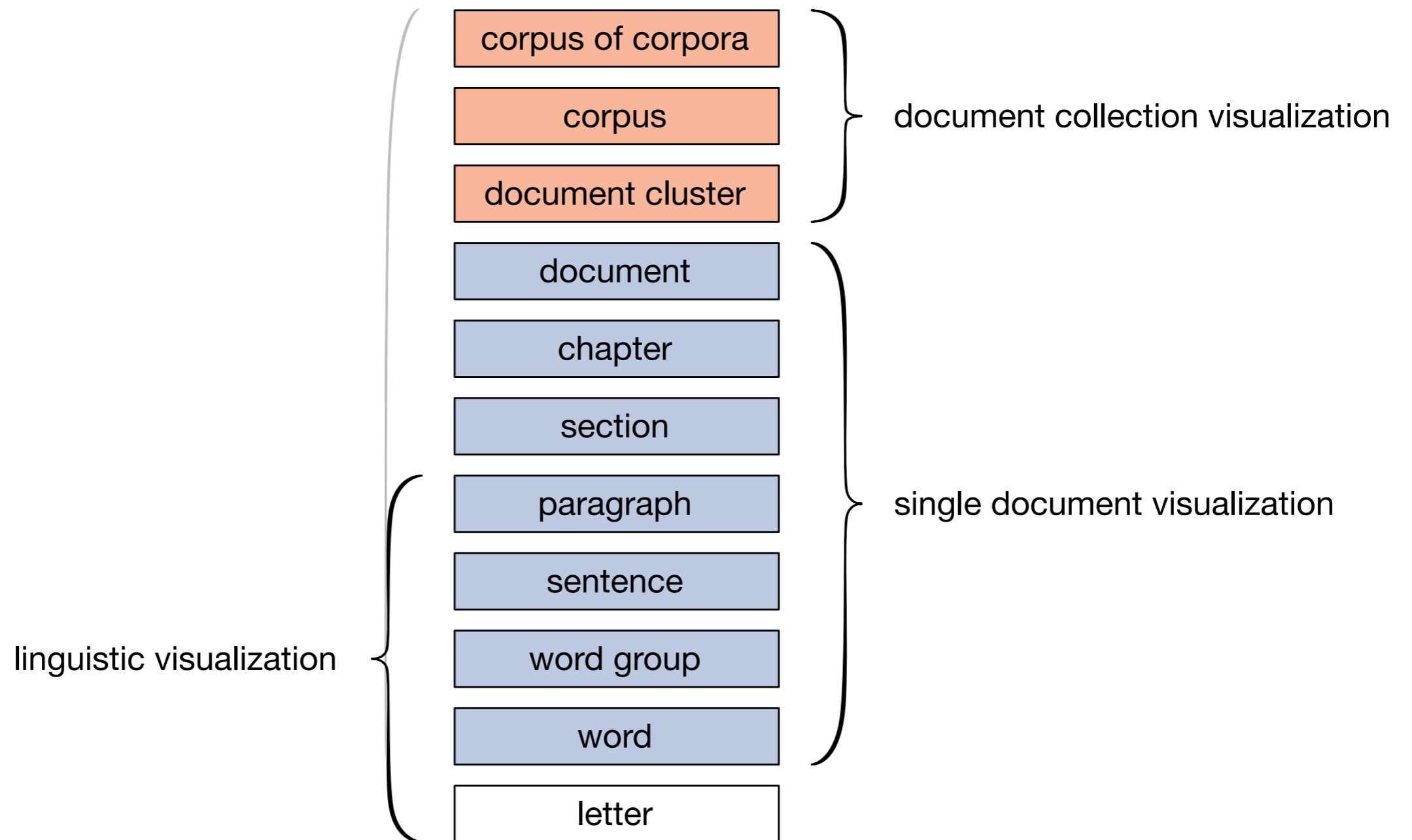
*by Frank L. Visco (June 1986 in Writers' digest)*

# Thinking about..

- or a bag of words (used for similarity measures):

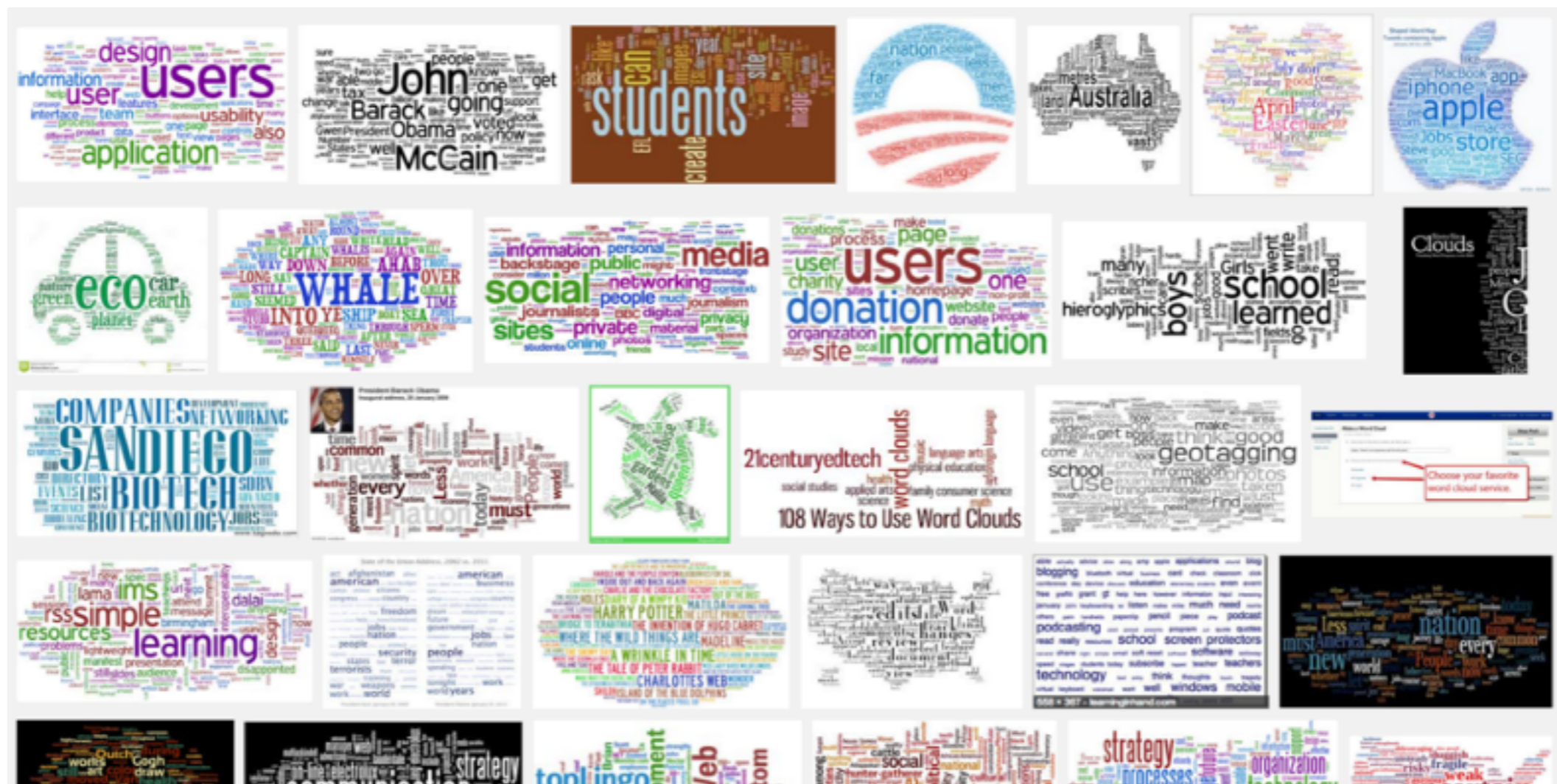
	<b>princess</b>	<b>dragon</b>	<b>castle</b>
<b>doc1</b>	1	1	1
<b>doc2</b>	0	0	1

# Text Units Hierarchy



# Vis for Text Documents

- TagClouds : <http://www.flickr.com/photos/tags/>
- WordCloud (popular) — <http://www.wordle.net>







# Vis for Text Documents

1 You create the word sequence filter:  
**WORD1** and **WORD2**

2 Many Eyes finds this word relationship in Jane Austen's text:

Her manners were pronounced to be very bad indeed,  
a mixture of **pride and impertinence**; she had no  
conversation, no stile, no taste, no beauty.

3 Many Eyes creates the word graph:

**pride** → **impertinence**



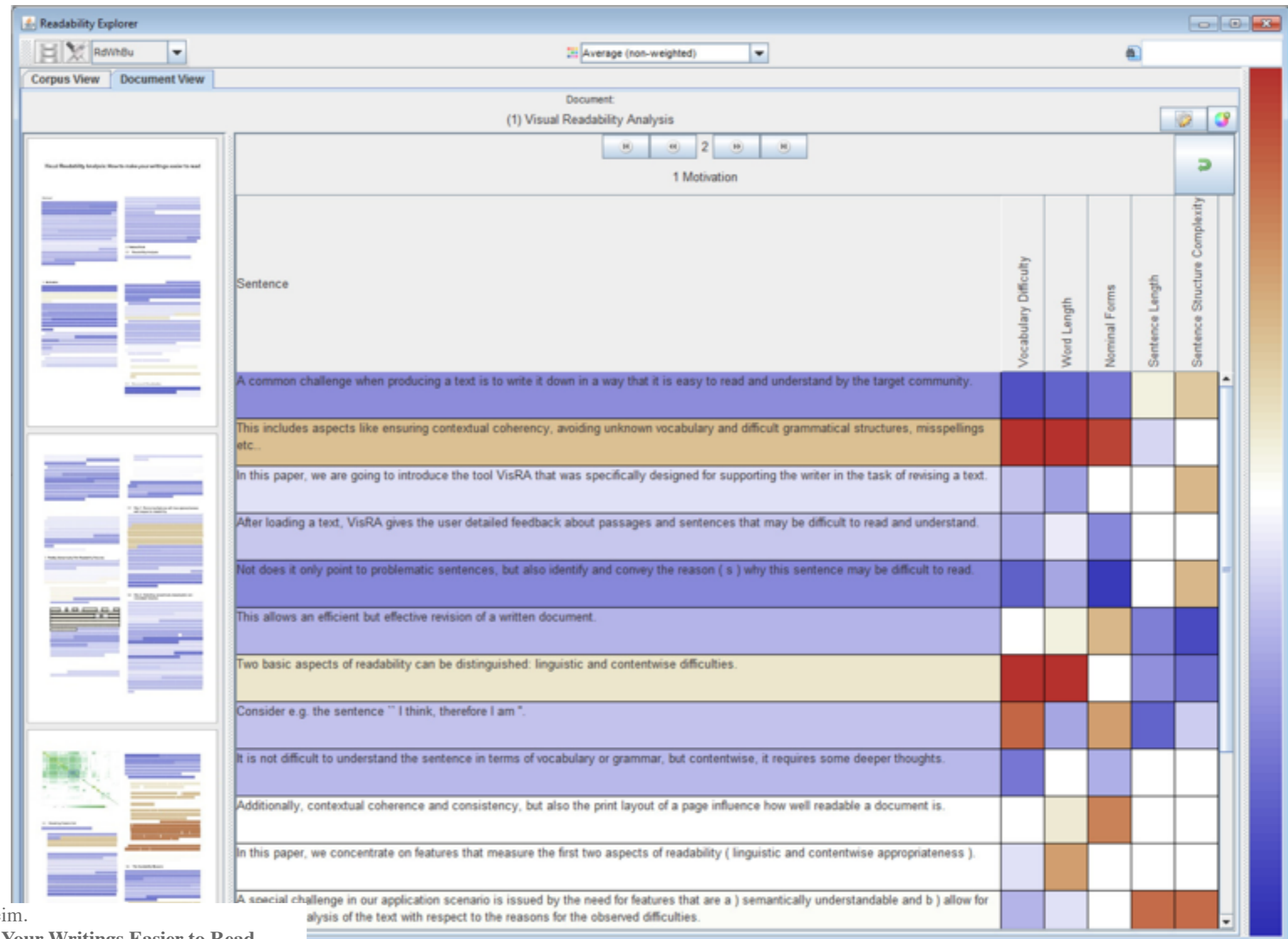
# Vis for Text Documents

- DocuBurst : <http://vialab.science.uoit.ca/docuburst/>
- based on: WordNet, see the network

The screenshot displays the DocuBurst web application interface. At the top, there is a navigation bar with links for 'DocuBurst', 'Help', 'Administrate', 'Consent Form', and 'Embed'. The main content area is divided into three sections:

- Search Panel (Left):** Contains a search bar, a filter input, and a 'Sense Details' section. The 'Sense Details' section shows 'No Synset Selected'. Below this is a 'Root' search bar, a 'Sense' dropdown menu set to 'All', and a 'Word Not Found' message. A 'Suggested' section lists words: 'thing', 'change', 'content', 'message', 'part'. At the bottom, there are options for 'Colour by: Single Node' and 'Depth: 2'.
- Radial Word Network (Center):** A circular network diagram with 'animal' at the center. It branches out into categories like 'dog', 'fish', 'bird', 'cat', and 'gray'. Each category further branches into related terms, such as 'puppy', 'cur', 'by dog', 'lark', 'shark', 'sole', 'jack', 'grunt', 'writing', 'sole', 'lark', 'stark', 'lark', 'raven', 'maggie', 'owl', 'turkey', 'pigeon', 'duck', 'goose', 'lark', 'pigeon', 'duck', 'goose', 'lark', 'pigeon', 'duck', 'goose'.
- Word Cloud (Right):** A collection of words of varying sizes and orientations. Prominent words include 'William the Conqueror', 'Alice', 'Caterpillar', 'Mary Ann', 'King and Queen', 'Canterbury', 'head', 'it\_would\_', 'executioner', 'And the Gryphon', 'judge', 'Queen', 'King, the Queen', 'on the bank', 'patriotic Archbishop', 'Edgar Atheling', 'bank', 'choking', 'King, and the Queen', 'Canterbury', 'head, it\_would\_'.

# Vis for Language Analysis



D. Oelke, D. Spretke, A. Stoffel and D. A. Keim.

**Visual Readability Analysis: How to Make Your Writings Easier to Read.**

*IEEE Transactions on Visualization and Computer Graphics*, 18(5):662-674, 2012.

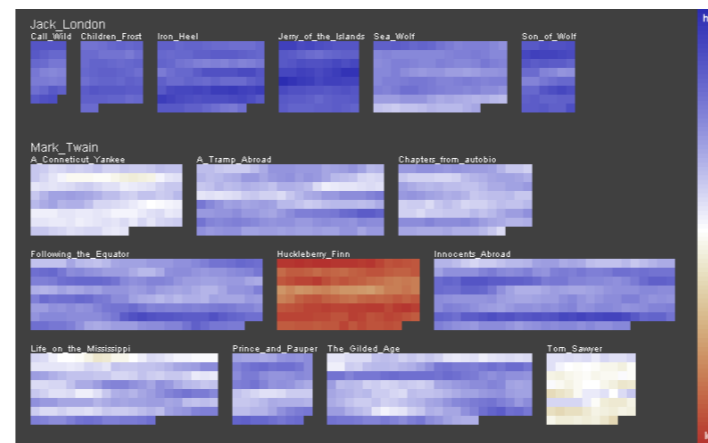
# Vis for Language Analysis

- Literature fingerprints:

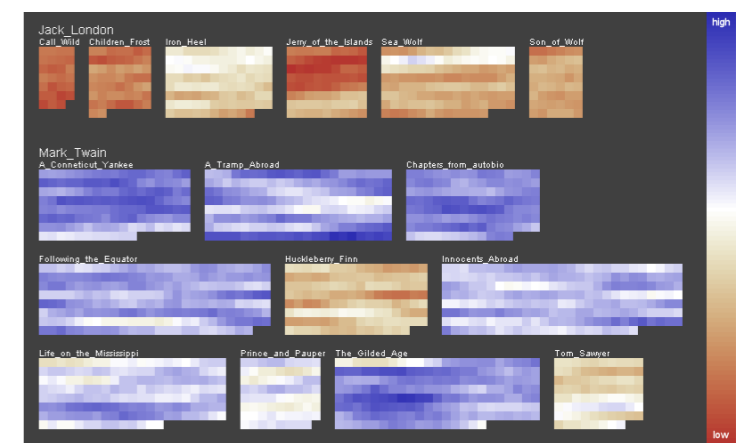
“Fingerprints of books of Mark Twain and Jack London. Different measures for authorship attribution are tested. If a measure is able to discriminate between the two authors, the visualizations of the books that are written by the same author will equal each other more than the visualizations of books written by different authors. It can easily be seen that this is not true for every measure (e.g. Hapax Dislegomena\*).

Furthermore, it is interesting to observe that the book *Huckleberry Finn* sticks out in a number of measures as if it is not written by Mark Twain.”

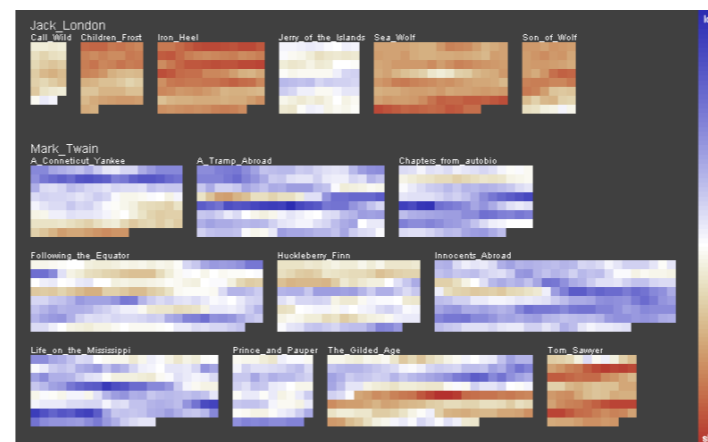
\*method to measure the vocabulary richness



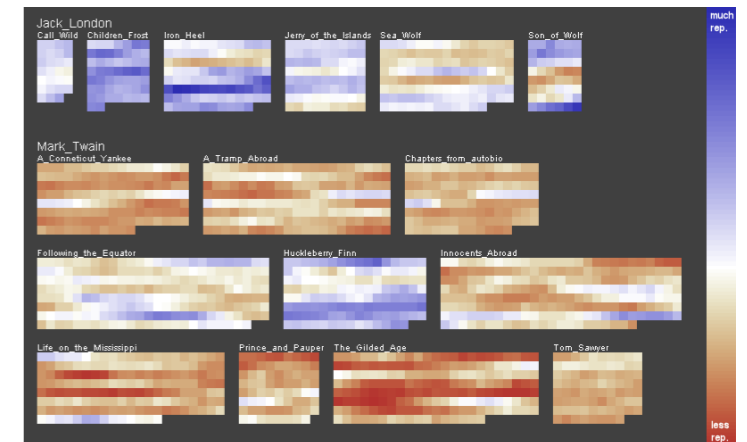
(a) Function words (First Dimension after PCA)



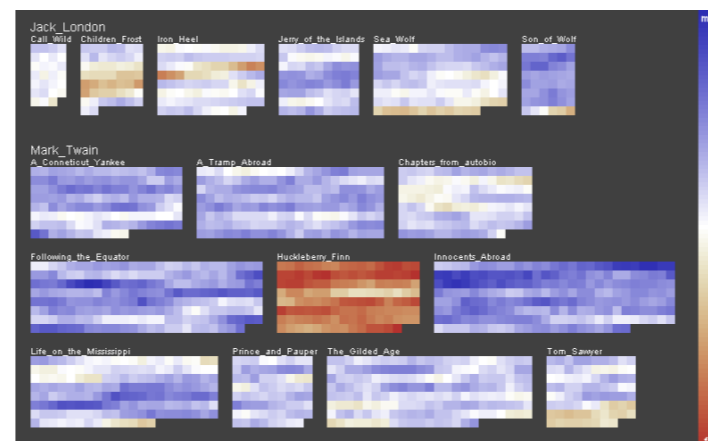
(b) Function words (Second Dimension after PCA)



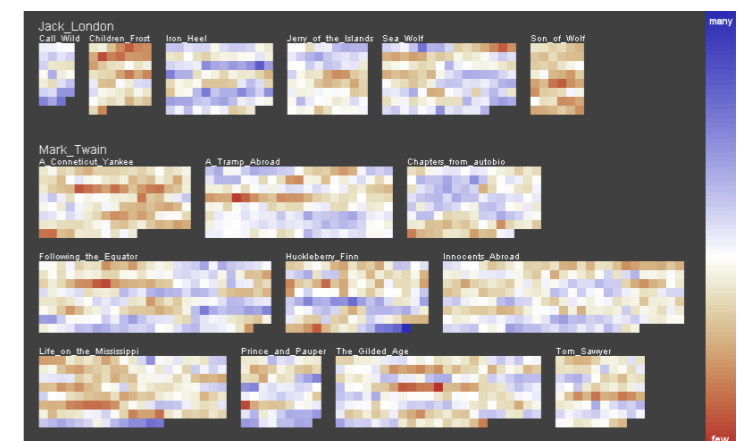
(c) Average sentence length



(d) Simpson's Index



(e) Hapax Legomena



(f) Hapax Dislegomena

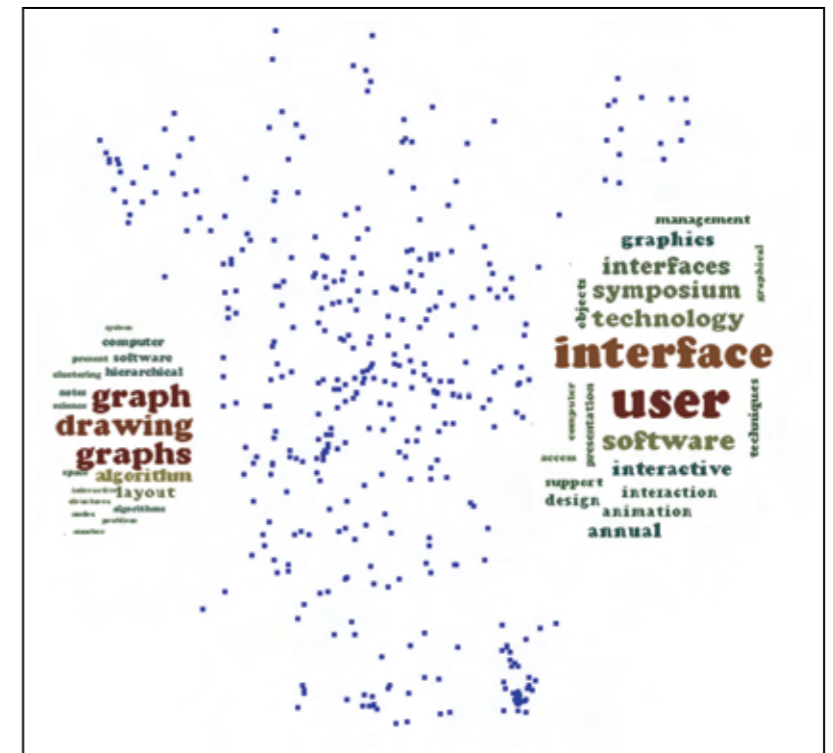
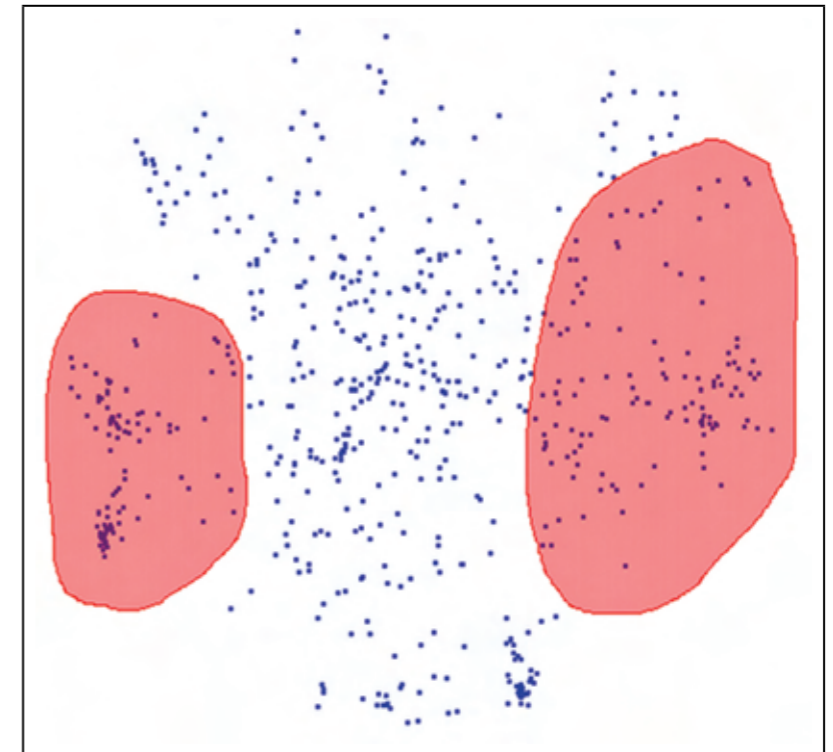
Daniel A. Keim and Daniela Oelke.

**Literature Fingerprinting: A New Method for Visual Literary Analysis.**  
*Proceedings of the 2007 IEEE Symposium on Visual Analytics Science and Technology (VAST '07)*



# Visualization for Large Text Corpora

- use bag-of-words to project documents w.r.t. text similarity into a landscape
- (only) one example



**Figure 5:** A user can interactively draw a region (polygon) containing a subset of documents of interest (top figure). Keywords are extracted from the selected document and their corresponding word cloud is built inside the user-defined region (bottom figure).

Fernando V. Paulovich, Franklina M. B. Toledo, Guilherme P. Telles, Rosane Minghim, and Luis Gustavo Nonato.

**Semantic Wordification of Document Collections.**

*Comp. Graph. Forum* 31, 3pt3 (June 2012)



# Visual Analytics for Large Text Corpora (example JigSaw)



# Vis for Large Document Collections

- documents contain more information than just text:

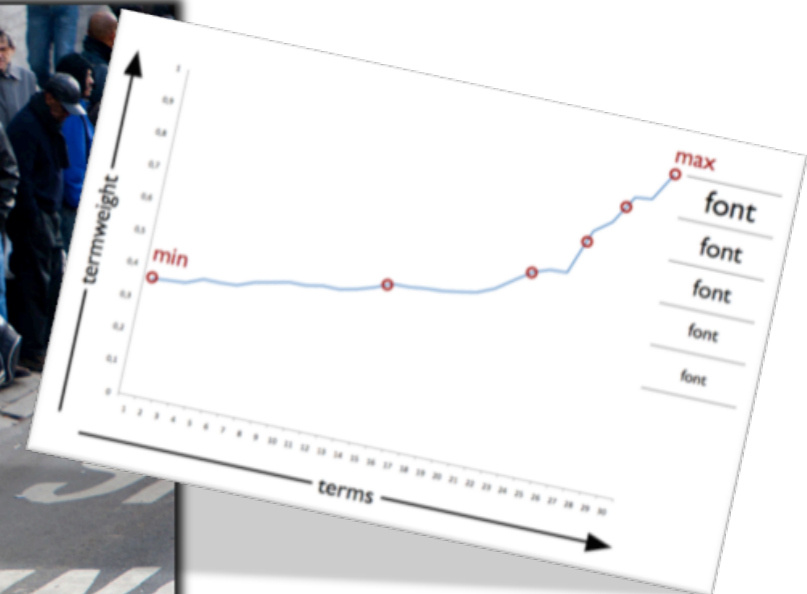
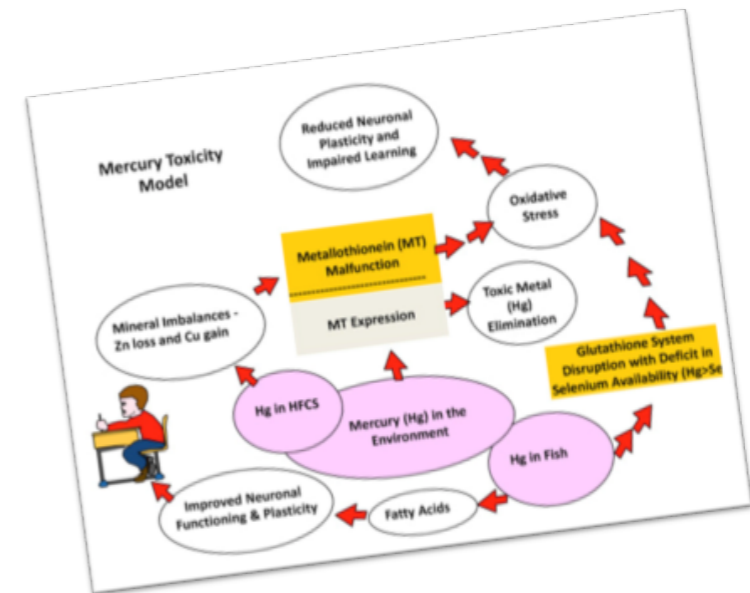
- meta information
- structure (paragraphs, text boxes,..)
- figurative content:

- parallel perception

- compact

- multi-lingual

- empathy



# Vis for Large Document Collections

- (only) three examples:
  - Bohemian bookshelf
  - DocumentCards
  - Semantics:



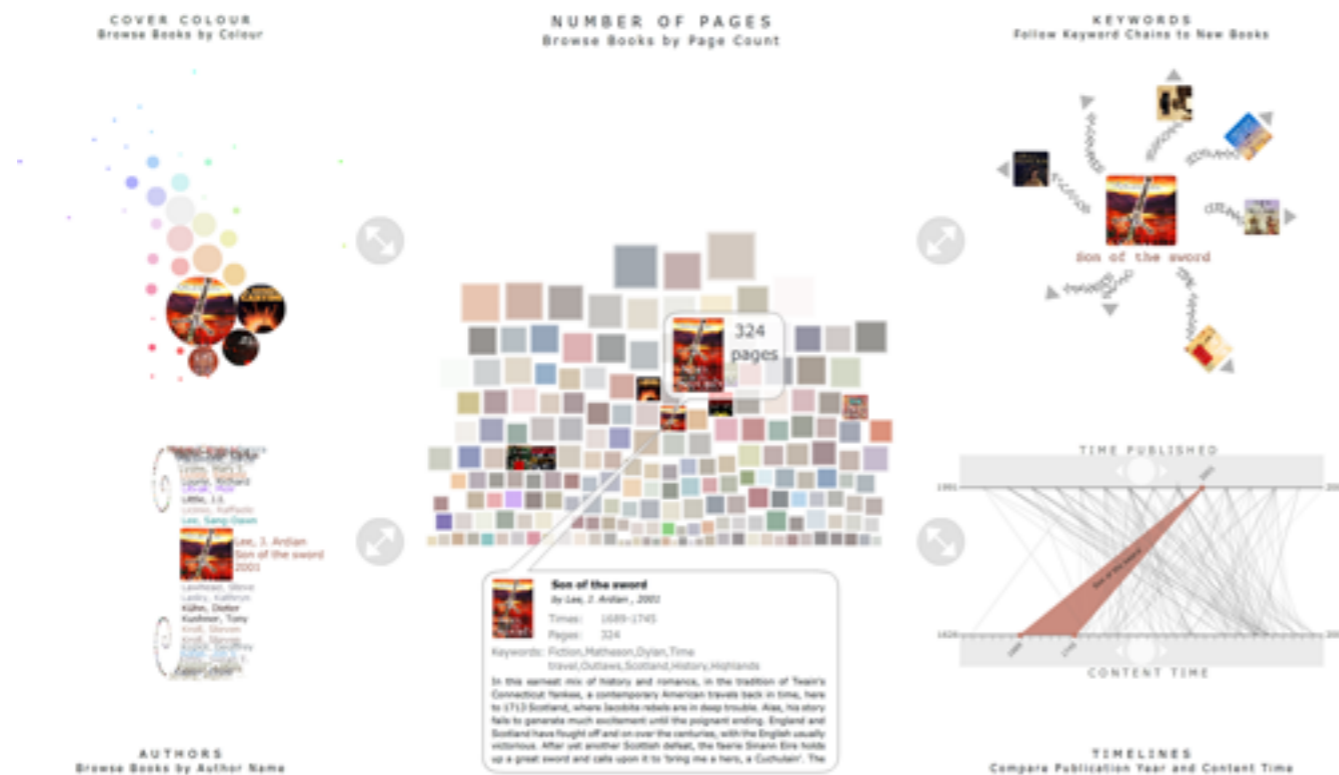
Figure 1: Semantics generated by our system for various filenames.

## **Semantics: Visual Metaphors as File Icons**

Vidya Setlur, Conrad Albrecht-Buehler, Amy A. Gooch,  
Sam Rossoff, Bruce Gooch



# Vis for Large Document Collections



[webpage with video](#)

Alice Thudt, Uta Hinrichs and Sheelagh Carpendale.

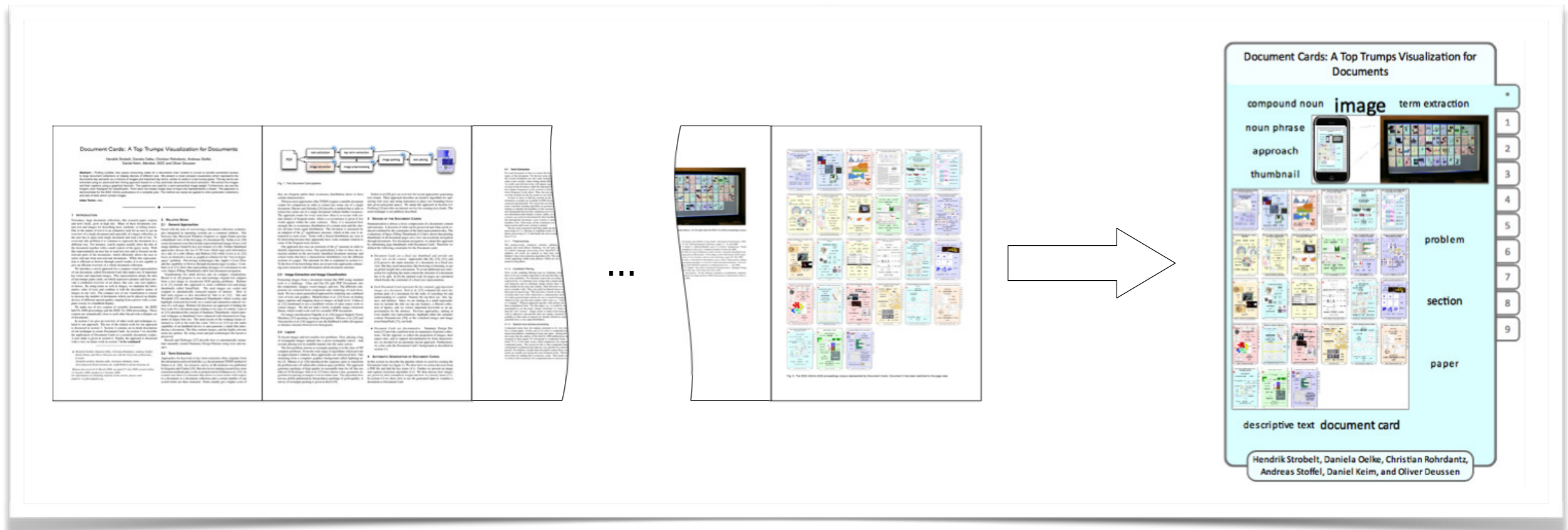
**The Bohemian Bookshelf: Supporting Serendipitous Book Discoveries through Information Visualization.**

*CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012*

# DocumentCards

- summarize scientific documents using important terms and important figures
- design considerations:
  - Document Cards are fixed size thumbnails that are self-explanatory
  - Document Cards represent the document's content as a mixture of figure and textual representatives
  - Document Cards should be discriminative and should have a high recognizability

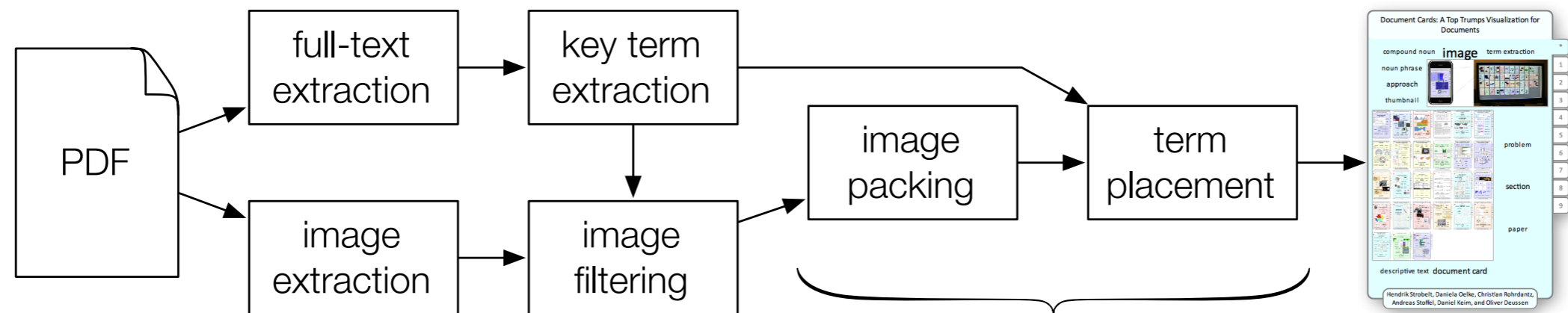
# DocumentCards

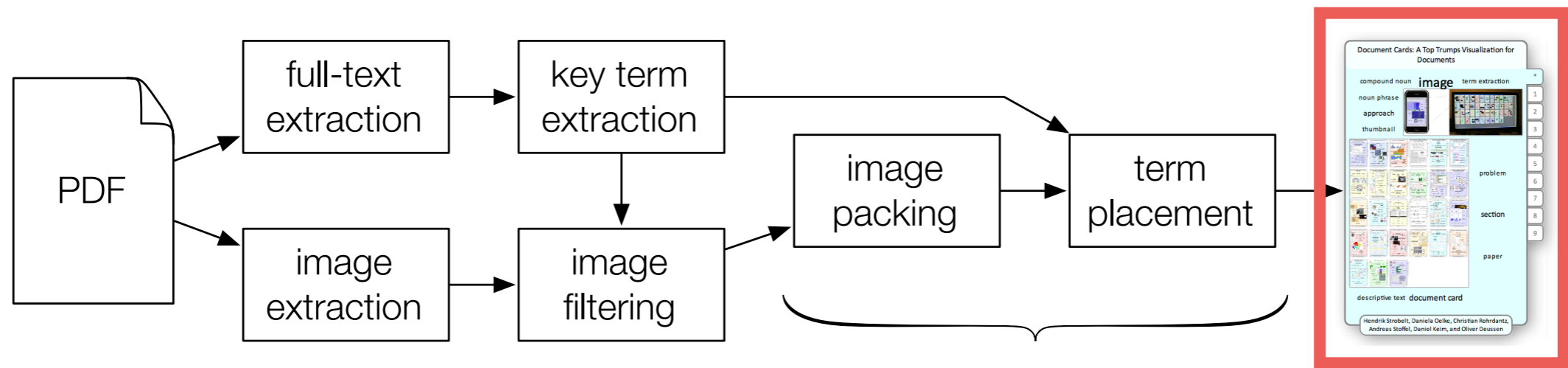




<p><b>Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context</b></p> <p>systems biologist context interaction graph graph model dataset figure</p> <p>edge tool cell</p> <p>gene</p> <p>layout algorithm</p> <p>process node cerebral</p> <p>Aaron Barsky, Tamara Munzner, Jennifer Gandy, and Robert Kincaid</p>	<p><b>Multi-Focused Geospatial Analysis Using Probes</b></p> <p>probe interface</p> <p>participant figure scale</p> <p>window region-of-interest local region</p> <p>data application</p> <p>Thomas Burkiewicz, Wenwen Dou, Zachary Wartell, William Ribarsky, and Remco Chang</p>	<p><b>Stacked Graphs: Geometry &amp; Aesthetics</b></p> <p>question visualization paper</p> <p>type author color</p> <p>comment york time algorithm layer</p> <p>order trend</p> <p>name/voyager graphic time story system</p> <p>legibility design issue layout method</p> <p>Lee Byron and Martin Wattenberg</p>	<p><b>Vispedia: Interactive Visual Exploration of Wikipedia Data via Search-Based Integration</b></p> <p>edge bundle technique polyline segment</p> <p>road map control mesh straight line</p> <p>mesh edge pattern transfer function</p> <p>user</p> <p>color and opacity enhancement node position control point graph layout</p> <p>result method visual cluster general graph primary direction</p> <p>Weiwei Cui, Hong Zhou, Student, Huamin Qu, Pak Chung Wong, and Xiaoming Li</p>	<p><b>VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery</b></p> <p>map rss feed participant</p> <p>temporal information item</p> <p>data</p> <p>information space</p> <p>query parameter exploration set</p> <p>visget</p> <p>description</p> <p>Marian Dirk, Sheelagh Carpendale, Christopher Collins, and Carey Williamson</p>	
<p><b>Who Votes For What? A Visual Query Language for Opinion Data</b></p> <p>user study paper</p> <p>report attribute</p> <p>sample population entity result</p> <p>sector opinion poll state street</p> <p>typical data set user interface</p> <p>visual query language visualization design</p> <p>participant poll data ring system data point</p> <p>Geoffrey M. Draper and Richard F. Riesenfeld</p>	<p><b>Exploration of Networks Using Overview+Detail with Constraint-based Cooperative Layout</b></p> <p>layout method</p> <p>route focal node</p> <p>tool user lod</p> <p>edge rout placement constraint</p> <p>high quality layout primary graph large network</p> <p>part system</p> <p>detailed view</p> <p>uml class diagram display</p> <p>layout technique cluster position structure focus node</p> <p>Tim Dwyer, Kim Marriott, Falk Schreiber, Peter J. Stuckey, Michael Woodward and Michael Wybrow</p>	<p><b>Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation</b></p> <p>visual exploration</p> <p>data dimension</p> <p>query</p> <p>prototype implementation visual representation cameras</p> <p>figure navigation grand tour system</p> <p>range scatterplot matrix user</p> <p>operation method order</p> <p>Niklas Elmquist, Pierre Dragicevic, and Jean-Daniel Fekete</p>	<p><b>Interactive Visual Analysis of Set-Typed Data</b></p> <p>bar block user</p> <p>scatterplot figure feature</p> <p>width data item dataset</p> <p>data record histogram washing agent set-typed data view</p> <p>Wolfgang Freiler, Kresimir Matkovic, Computer Society, and Helwig Hauser</p>	<p><b>Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation</b></p> <p>graphical history usage history item</p> <p>rule tableau image</p> <p>data field display approach</p> <p>event history interface history tool</p> <p>Jeffrey Heer, Jack D. Mackinlay, Chris Stolte, and Maneesh Agrawala</p>	<p><b>Improving the Readability of Clustered Social Networks using Node Duplication</b></p> <p>representation success rate clone node</p> <p>social network time central actor clone link</p> <p>analysis visualization duplicate test</p> <p>community noduplication</p> <p>significant effect cluster</p> <p>duplication link participant splittlink readability</p> <p>Nathalie Henry, Anastasia Bearman, and Jean-Daniel Fekete</p>
<p><b>EMDialog: Bringing Information Visualization into the Museum</b></p> <p>touch emily carr</p> <p>interaction design node installation</p> <p>visual appeal statement tree diagram perspective information visualization cut section</p> <p>museum visitor tree ring public space</p> <p>data representation people museum context</p> <p>Uta Hinrichs, Holly Schmidt, and Sheelagh Carpendale</p>	<p><b>On the Visualization of Social and other Scale-Free Networks</b></p> <p>scale-free network</p> <p>simplification method shortest path node degree</p> <p>visualization layout</p> <p>time</p> <p>distance matrix weighted graph geodesic cluster grime original graph edge filter</p> <p>edge metric power-law graph hub node</p> <p>Nantao Jia, Jared Hoberock, Michael Garland, and John C. Hart</p>	<p><b>A Framework of Interaction Costs in Information Visualization</b></p> <p>visualization</p> <p>user intent study focus-lock mode</p> <p>paper interaction cost participant</p> <p>time item interface operation</p> <p>focus lock framework evaluation</p> <p>section interaction target</p> <p>overview window subject gulf</p> <p>Heidi Lam</p>	<p><b>Distributed Cognition as a Theoretical Framework for Information Visualization</b></p> <p>cognitive property process endangered species</p> <p>infosys system cognition insight</p> <p>internal representation</p> <p>changer Inst Inst Inst</p> <p>Developer Inst Inst EXI</p> <p>Coffee Inst EXI EXI</p> <p>developer cognitive system</p> <p>report theory artifact</p> <p>cognitive science analyst document view</p> <p>Zhicheng Liu, Nancy J. Nersessian, and John T. Stasko</p>	<p><b>Particle-Based Labeling: Fast Point-Feature Labeling without Obscuring Other Visual Features</b></p> <p>section virtual particle method</p> <p>problem distant labels conflict particle figure</p> <p>labeling step solution</p> <p>point-feature visual element number</p> <p>time result collision map</p> <p>Martin Lubschik, Heidrun Schumann and Hilko Cordt</p>	<p><b>Rapid Graph Layout Using Space Filling Curves</b></p> <p>matrix order space filling curve based layout</p> <p>approach</p> <p>peano curve</p> <p>aspect ratio</p> <p>graph layout fast screen space</p> <p>Chris Mueller and Kwan-Liu Ma, Senior</p>
<p><b>HPP: A Novel Hierarchical Point Placement Strategy and its Application to the Exploration of Document Collections</b></p> <p>complexity cluster space</p> <p>result hipp plane</p> <p>relationship data instance group node data set approach</p> <p>Fernando V. Paulovich and Rosane Minghim</p>	<p><b>Effectiveness of Animation in Trend Visualization</b></p> <p>group trace line visualization data point</p> <p>user study task presentation participant animation and trace</p> <p>George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, and John Stasko</p>	<p><b>Viz-A-Vis: Toward Visualizing Video through Computer Vision</b></p> <p>condition pattern recognition computer vision</p> <p>aggregate activity table activity cube raw data</p> <p>video heat map frame</p> <p>viz-a-vis level model motion</p> <p>view aggregation step figure</p> <p>abstraction difference</p> <p>Mario Romero, Jay Sumner, John Stasko, and Gregory Abowd</p>	<p><b>Balloon Focus: a Seamless Multi-Focus+Context Method for Treemaps</b></p> <p>enclosure context technique</p> <p>relative position</p> <p>foci enlargement dependency graph elastic edge result non-focus item player</p> <p>subject view layout</p> <p>method balloon focus original treemap task</p> <p>Ying Tu and Han-Wei Shen</p>	<p><b>Perceptual Organization in User-Generated Graph Layouts</b></p> <p>visual characteristic data degree network set</p> <p>uniform edge length cluster</p> <p>number structure algorithm node</p> <p>human observer edge crosse condition study</p> <p>Frank van Ham and Bernice E. Rogowitz, Senior</p>	<p><b>The Word Tree, an Interactive Visual Concordance</b></p> <p>tree structure common word user time</p> <p>option data design</p> <p>tag cloud search term</p> <p>romeo and juliet</p> <p>title love</p> <p>token word tree eye context number</p> <p>suffix tree branch blind</p> <p>king james bible</p> <p>Martin Wattenberg and Fernanda B. Viégas</p>
<p><b>Evaluating the Use of Data Transformation for Information Visualization</b></p> <p>comment performance distribution task</p> <p>system step error rate</p> <p>non-transformed data</p> <p>choice technique</p> <p>study experiment house context figure data set</p> <p>data transformation</p> <p>anova test analytic task data property criteria visualization participant benefit impact task type</p> <p>Zhen Wen and Michelle X. Zhou</p>	<p><b>Spatially Ordered Treemaps</b></p> <p>readability score</p> <p>node order data geographic location angular change color space</p> <p>position</p> <p>sequence scene type distortion unit size</p> <p>histomap layout treemap node</p> <p>spatial layout</p> <p>image layout algorithm consistency arrangement displacement vector</p> <p>Jo Wood, and Jason Dykes</p>	<p><b>The Shaping of Information by Visual Metaphors</b></p> <p>information visual metaphor verbal metaphor</p> <p>type response time user</p> <p>understand</p> <p>incompatible question study correct response participant task question session difference</p> <p>visual representation</p> <p>Caroline Zemke and Robert Kosara</p>			

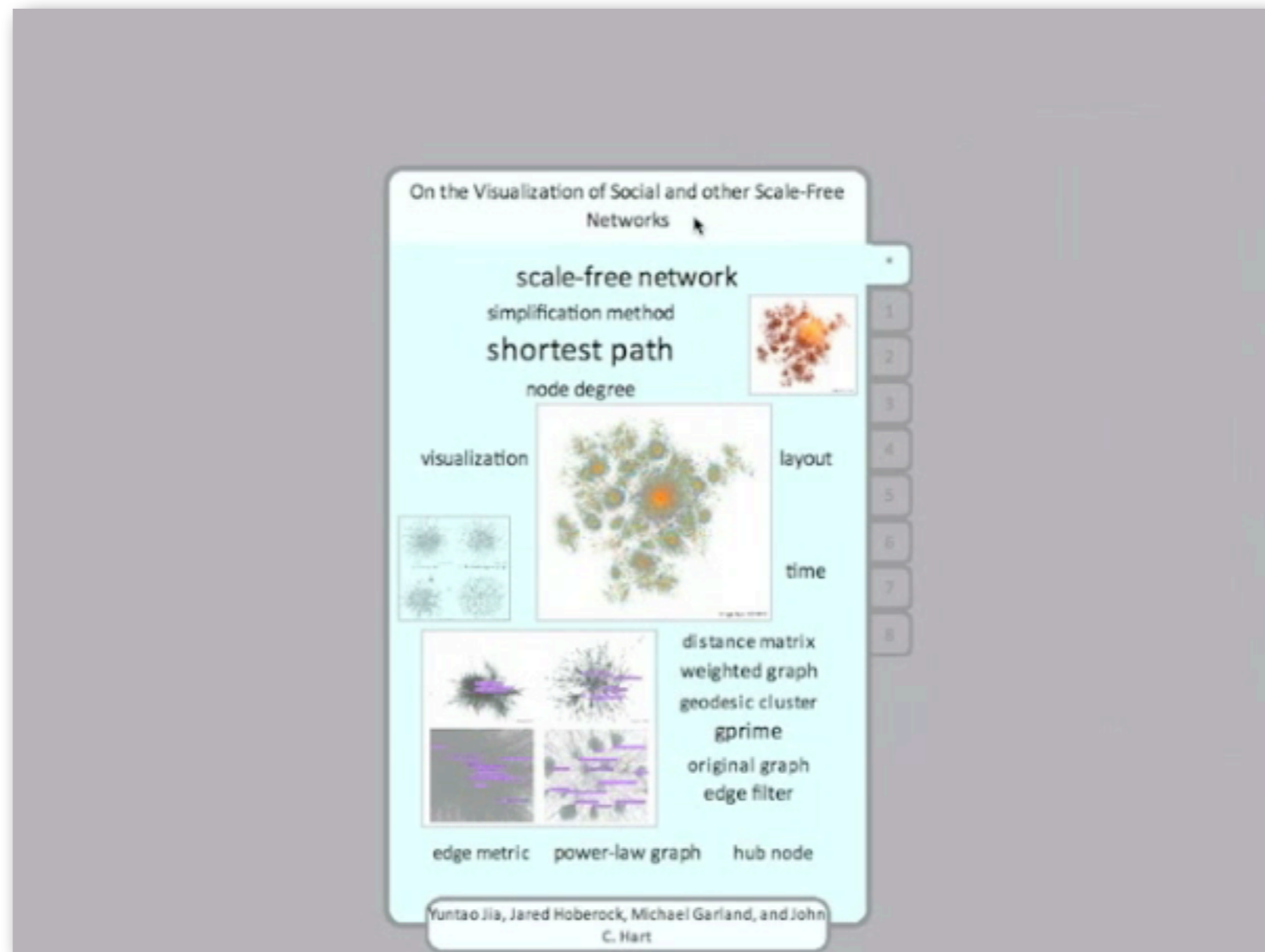
# DC - pipeline



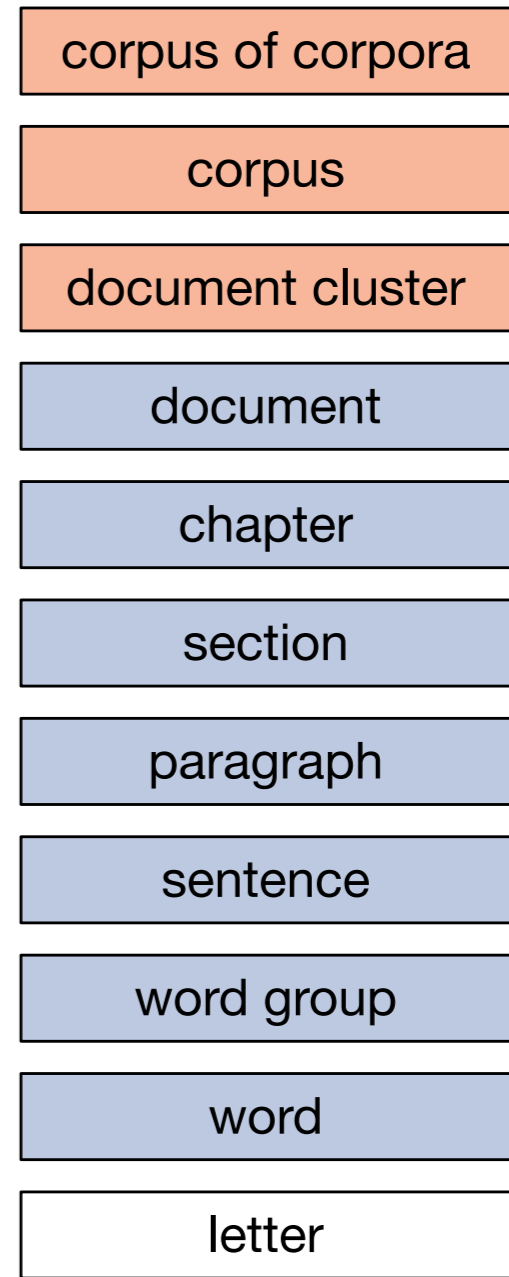
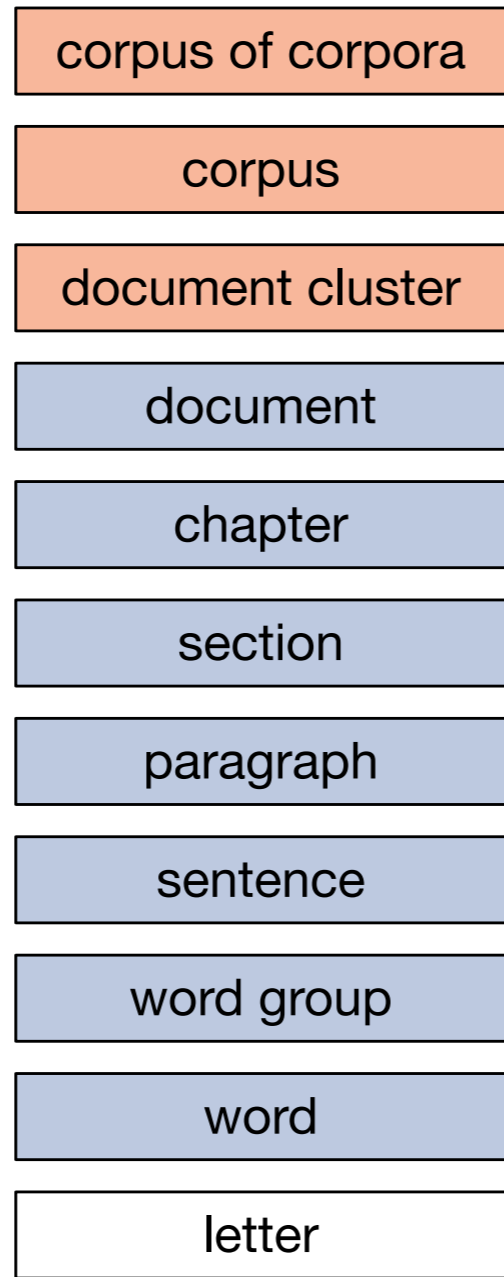
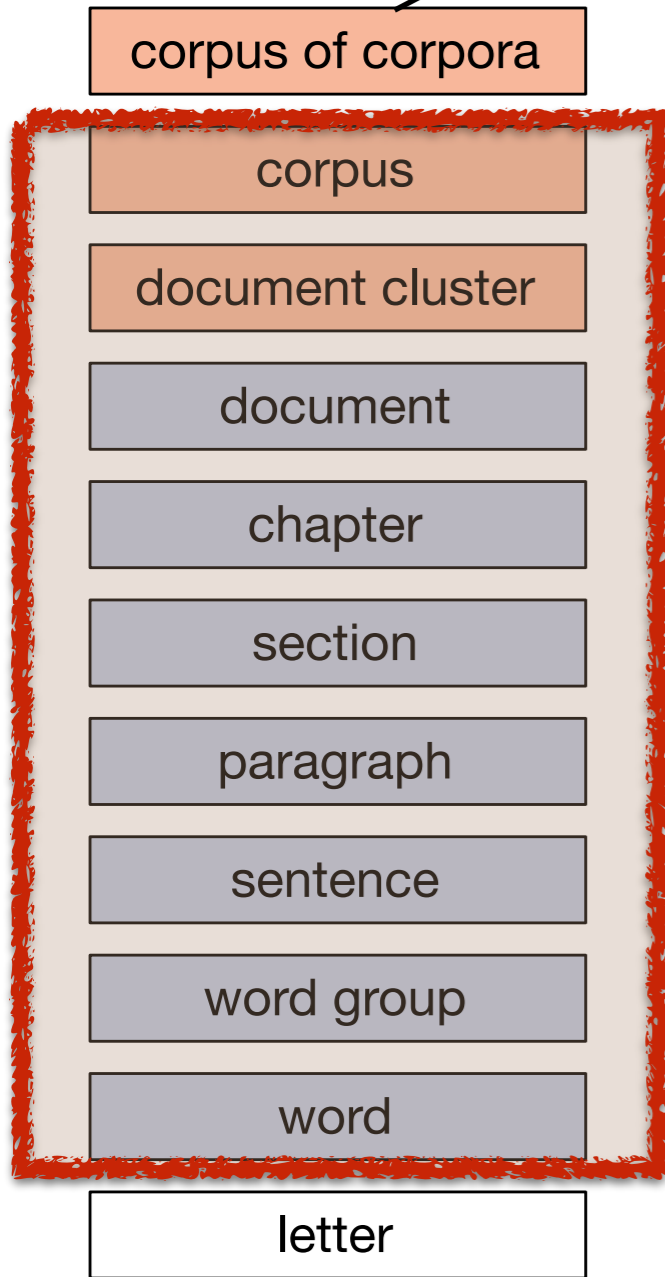


## Interaction:

- caption tooltip
- abstract tooltip
- move to orig. Pos.
- page switch
- term highlighting



# DiTop



# Compare Corpora

- Compare topics between text collections

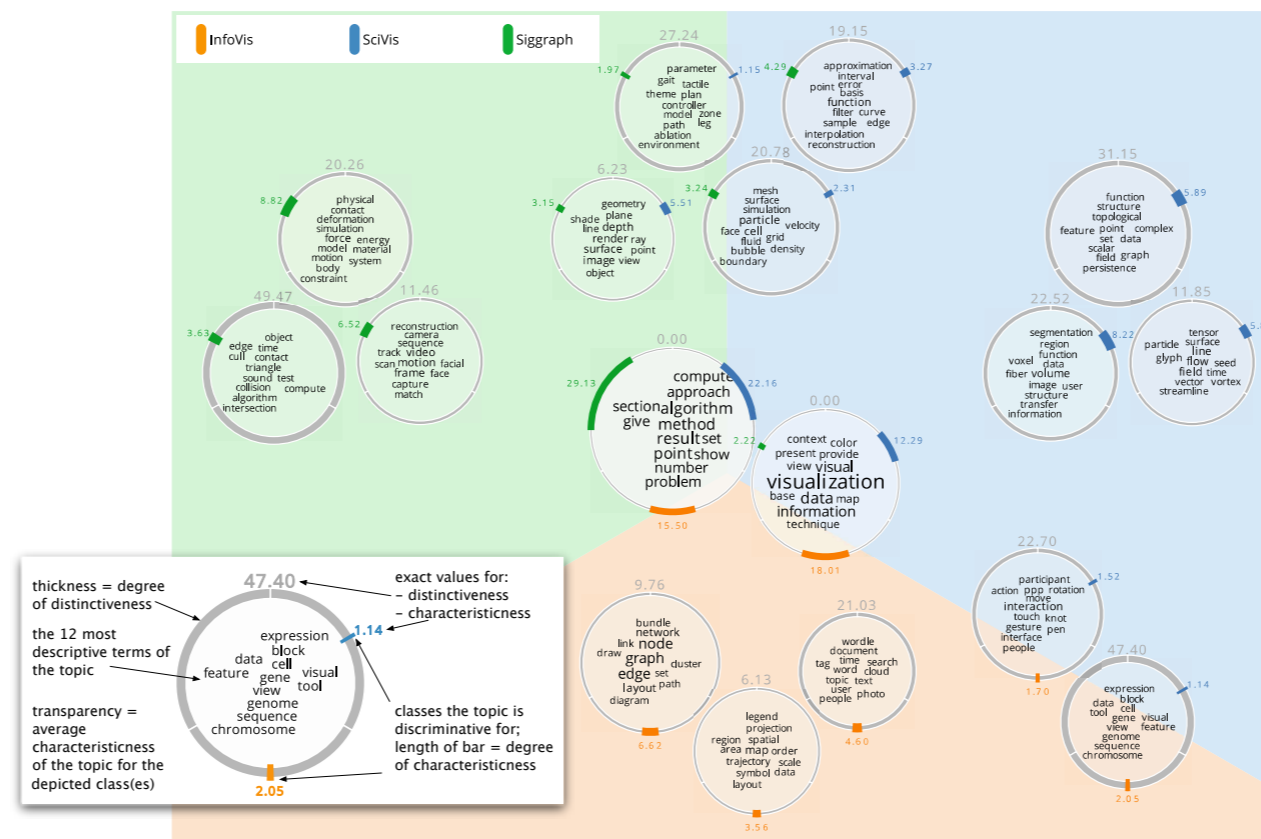


Figure 1: Comparison of 495 papers of InfoVis, SciVis, and Siggraph (discrimination threshold = 6, number of topics = 30)

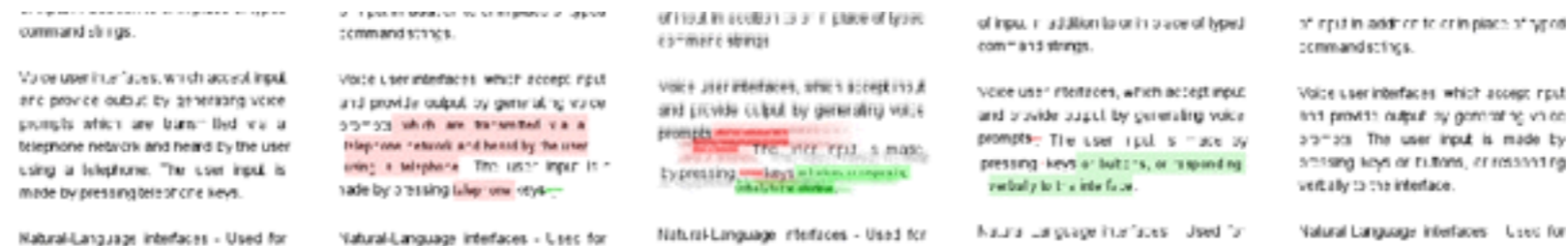


# Vis for Time-Evolving Document Collections





# Vis for Time Evolving Texts



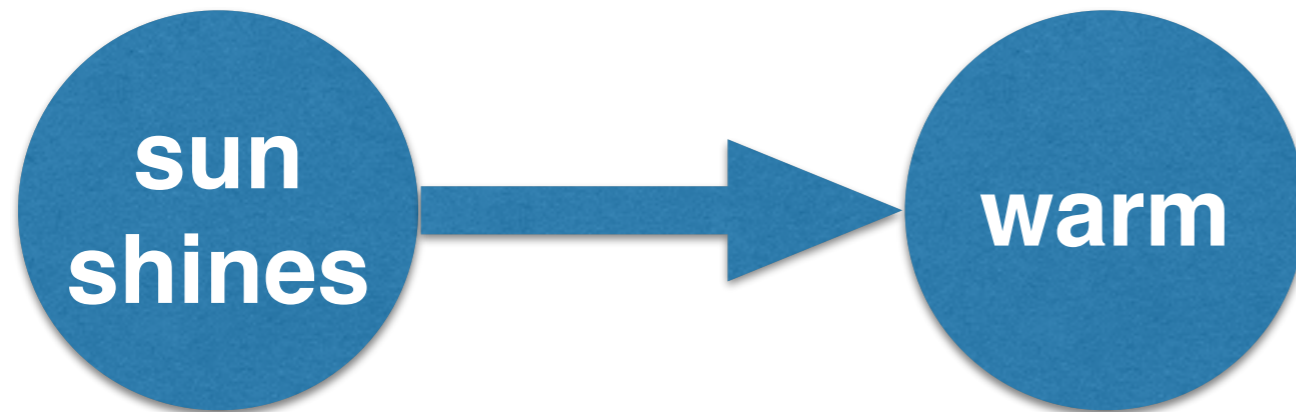
“This article examines the benefits of using text animated transitions for navigating in the revision history of textual documents. We propose an animation technique for smoothly transitioning between different text revisions, then present the Diffamation system. Diffamation supports rapid exploration of revision histories by combining text animated transitions with simple navigation and visualization tools. We finally describe a user study showing that smooth text animation allows users to track changes in the evolution of textual documents more effectively than flipping pages.”

**Video on the [webpage](#)**

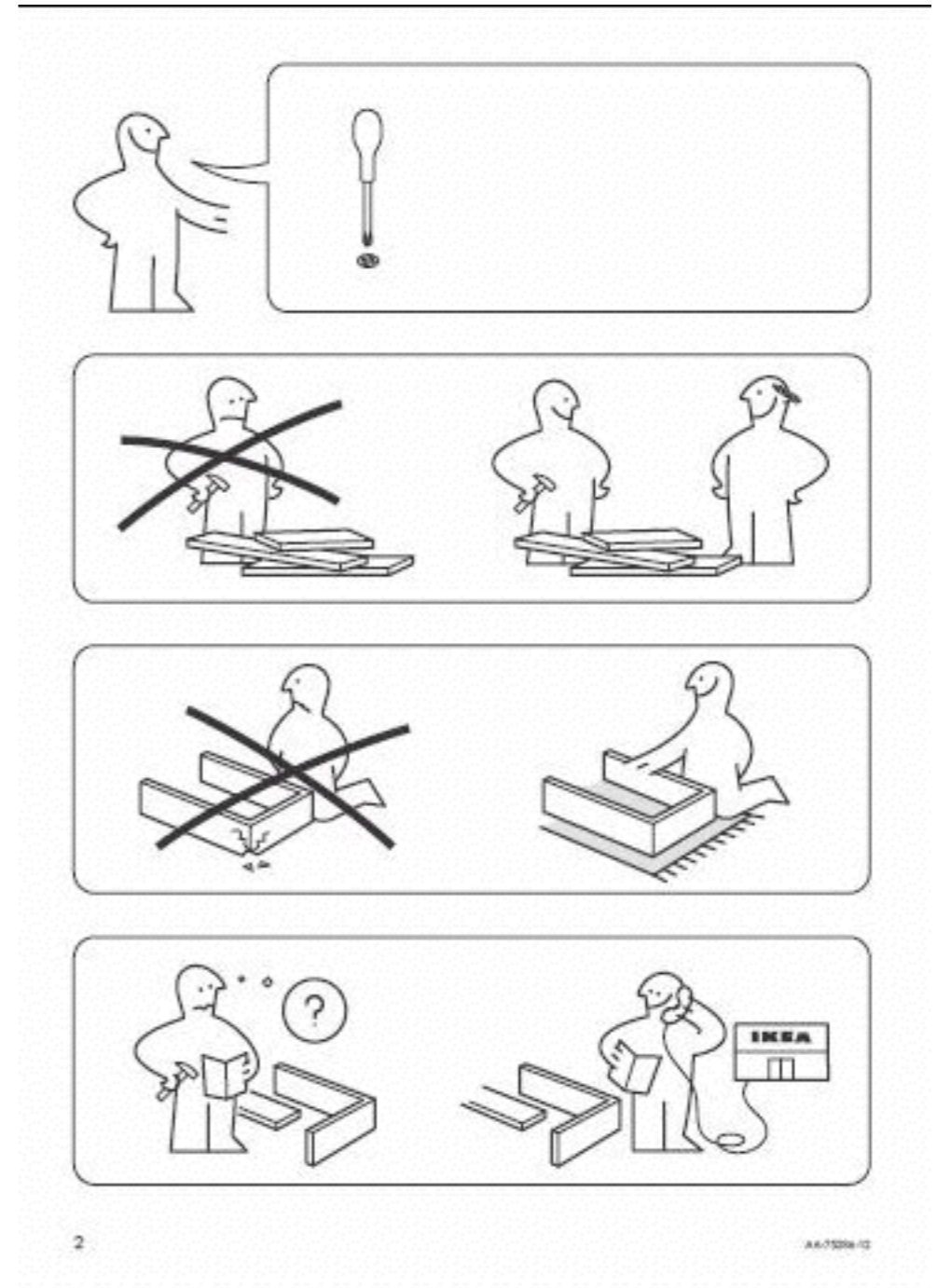
# The Role of Text in Vis

# Text in Vis

- Non-Example: Ikea
- Labels:



- Map Legends



# Text in Vis Storytelling

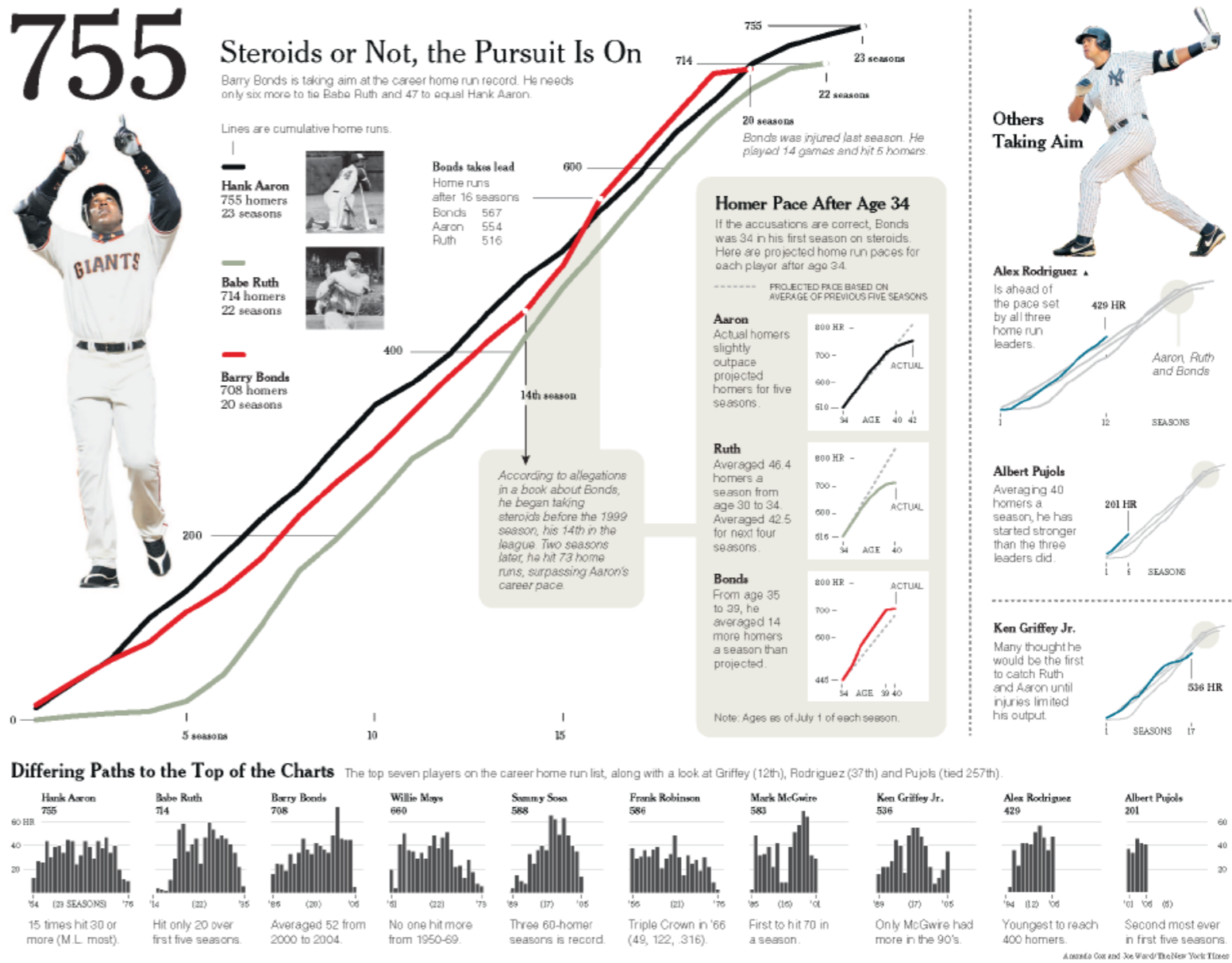


Fig. 1. Steroids Or Not, the Pursuit is On. New York Times.

**Narrative Visualization: Telling Stories with Data**

Edward Segel, Jeffrey Heer

IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 2010

# TextVis Specials

# Vis for Text Translation

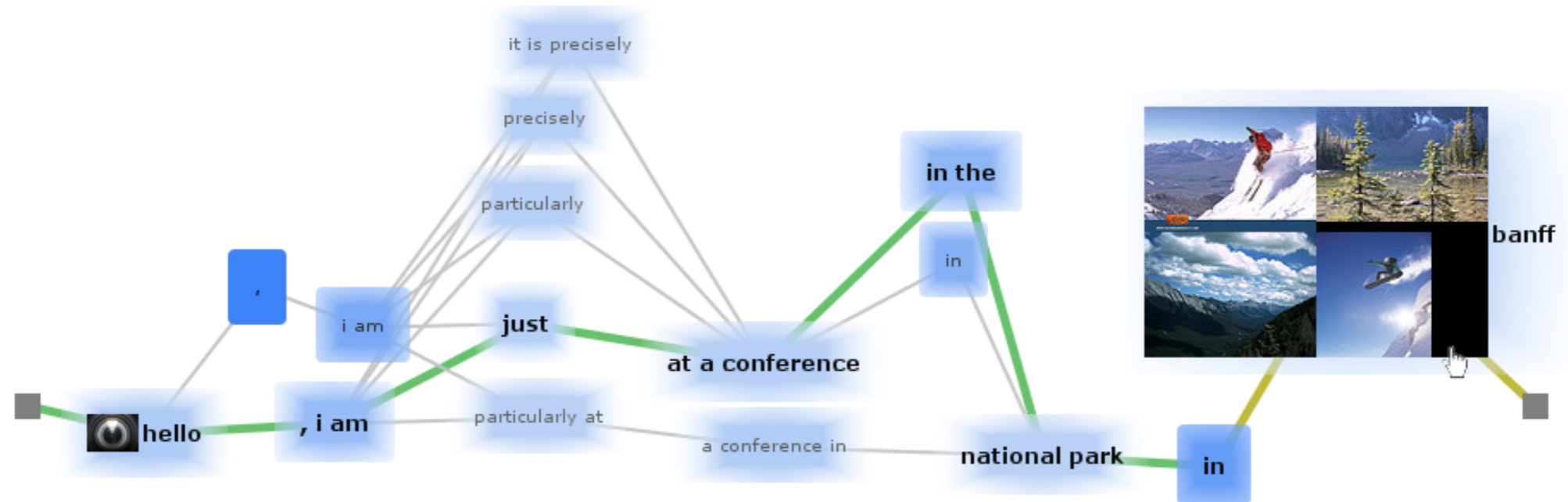
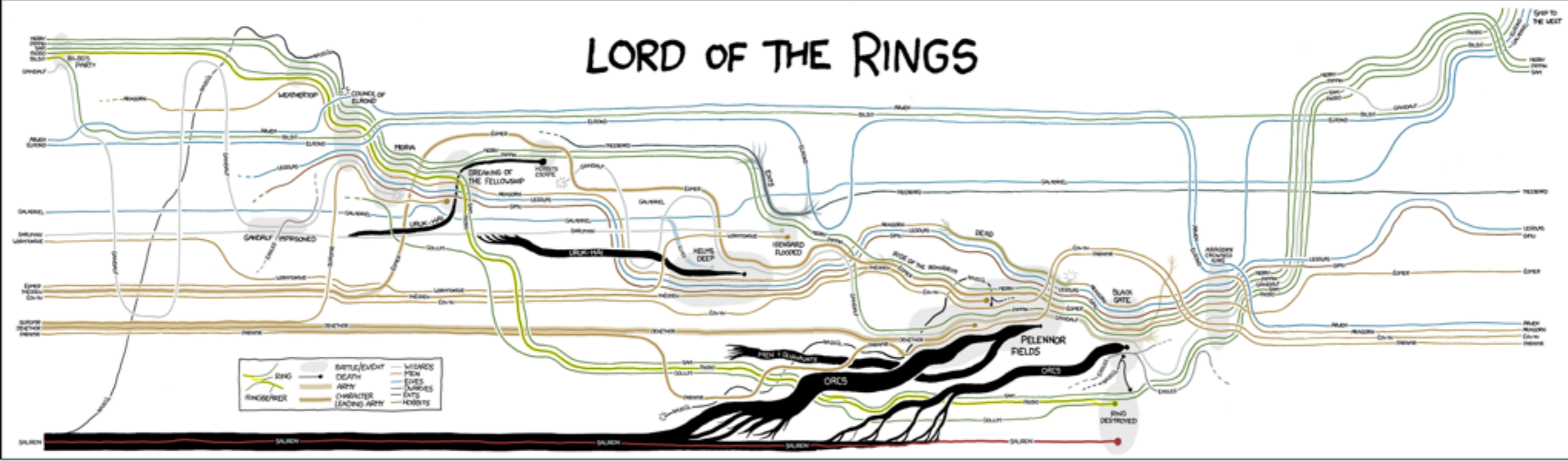


Figure 6: Translation lattice for the German sentence, “Hallo, ich bin gerade auf einer Konferenz im Nationalpark in Banff.” The statistically-identified best path (along the bottom) was incorrect and has been repaired. Photo nodes provide an alternative representation for words not in the translation vocabulary. Mouse over expands the node and reveals four photos, while other nodes move away to avoid occlusion.

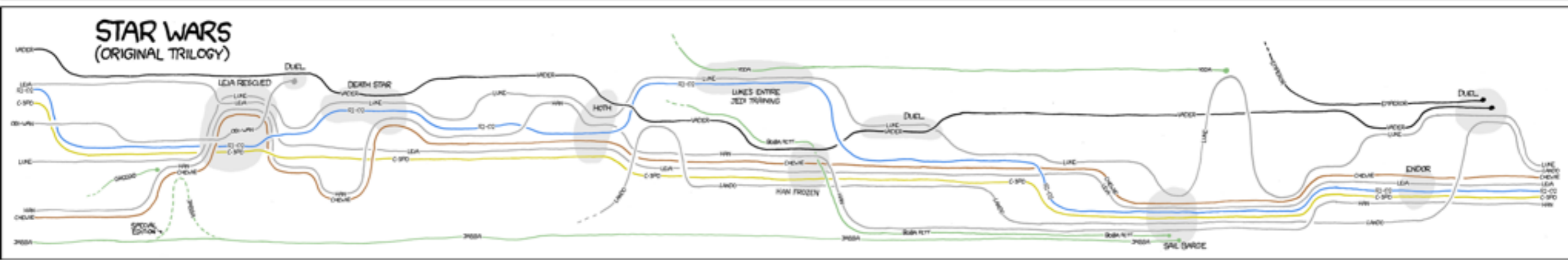


THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS.  
THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE  
LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.

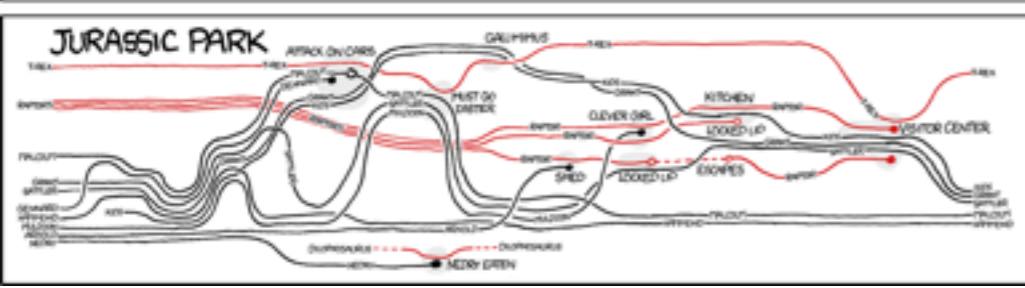
# LORD OF THE RINGS



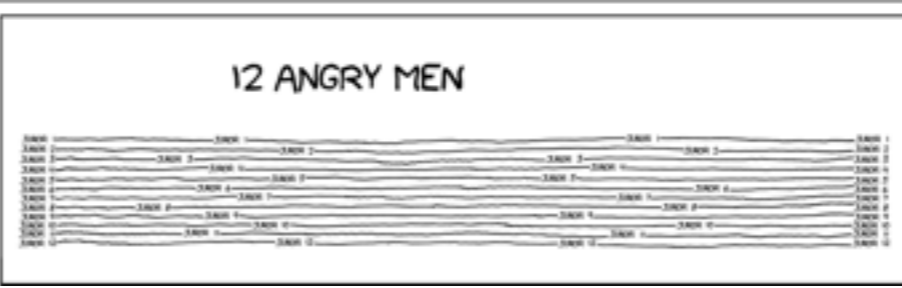
# STAR WARS (ORIGINAL TRILOGY)



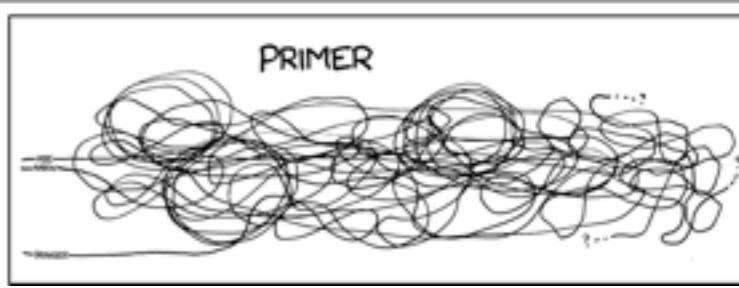
# JURASSIC PARK



# 12 ANGRY MEN



# PRIMER



# Text to Vis conversion

“Natural language is an easy and effective medium for describing visual ideas and mental images. Thus, we foresee the emergence of language-based 3D scene generation systems to let ordinary users quickly create 3D scenes without having to learn special software, acquire artistic skills, or even touch a desktop window-oriented interface. WordsEye is such a system for automatically converting text into representative 3D scenes. WordsEye relies on a large database of 3D models and poses to depict entities and actions. Every 3D model can have associated shape displacements, spatial tags, and functional properties to be used in the depiction process.”



Figure 1: *John uses the crossbow. He rides the horse by the store. The store is under the large willow. The small allosaurus is in front of the horse. The dinosaur faces John. A gigantic teacup is in front of the store. The dinosaur is in front of the horse. The gigantic mushroom is in the teacup. The castle is to the right of the store.*

Bob Coyne and Richard Sproat. 2001.

**WordsEye: an automatic text-to-scene conversion system**

*Proceedings of the 28th annual conference on Computer graphics and interactive techniques (SIGGRAPH '01)*

# Further TextVis..

- ... on topic modeling
- ... for text exploration (human computer interaction)
- ... for search results
- ... linguistic features (e.g. vowel harmony)
- ... source code
- ... for sentiment analysis
- ... **SO MUCH MORE !!**



# <http://textvis.inu.se/>

## Text Visualization Browser

A Visual Survey of Text Visualization Techniques

Provided by ISOVIS group

About Add entry

Techniques displayed:

141

Search:

Time filter:

1976

2014

Analytic Tasks



Visualization Tasks



Data

Source



Properties

