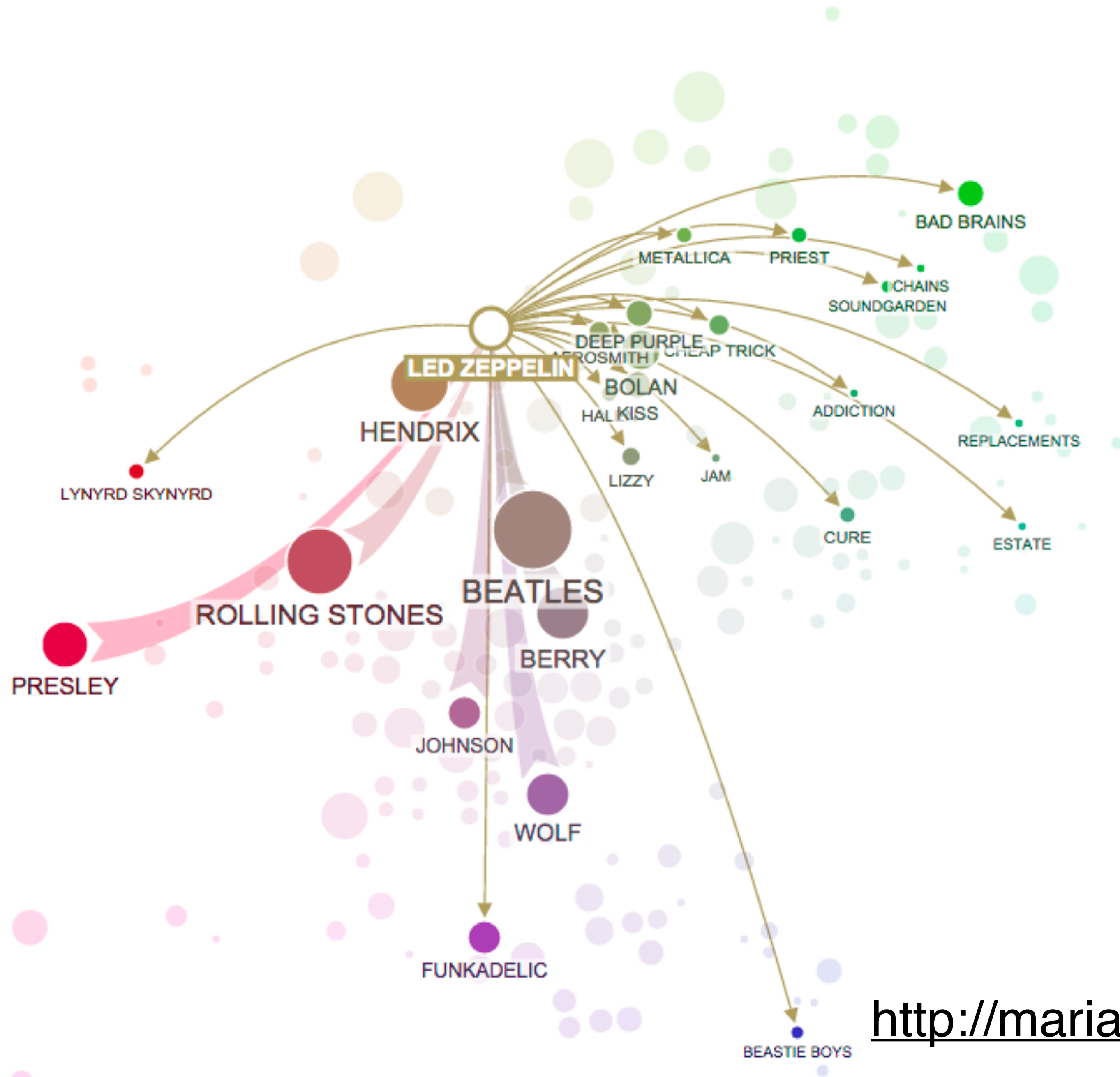


CS-5630 / CS-6630 Visualization for Data Science Text Visualization



Alexander Lex
alex@sci.utah.edu

Design Critique



<https://goo.gl/IDRXDI>

<http://mariandoerk.de/edgemaps/demo/>

Text / Language

Features of Text as representation language

abstract, general

extremely expressive

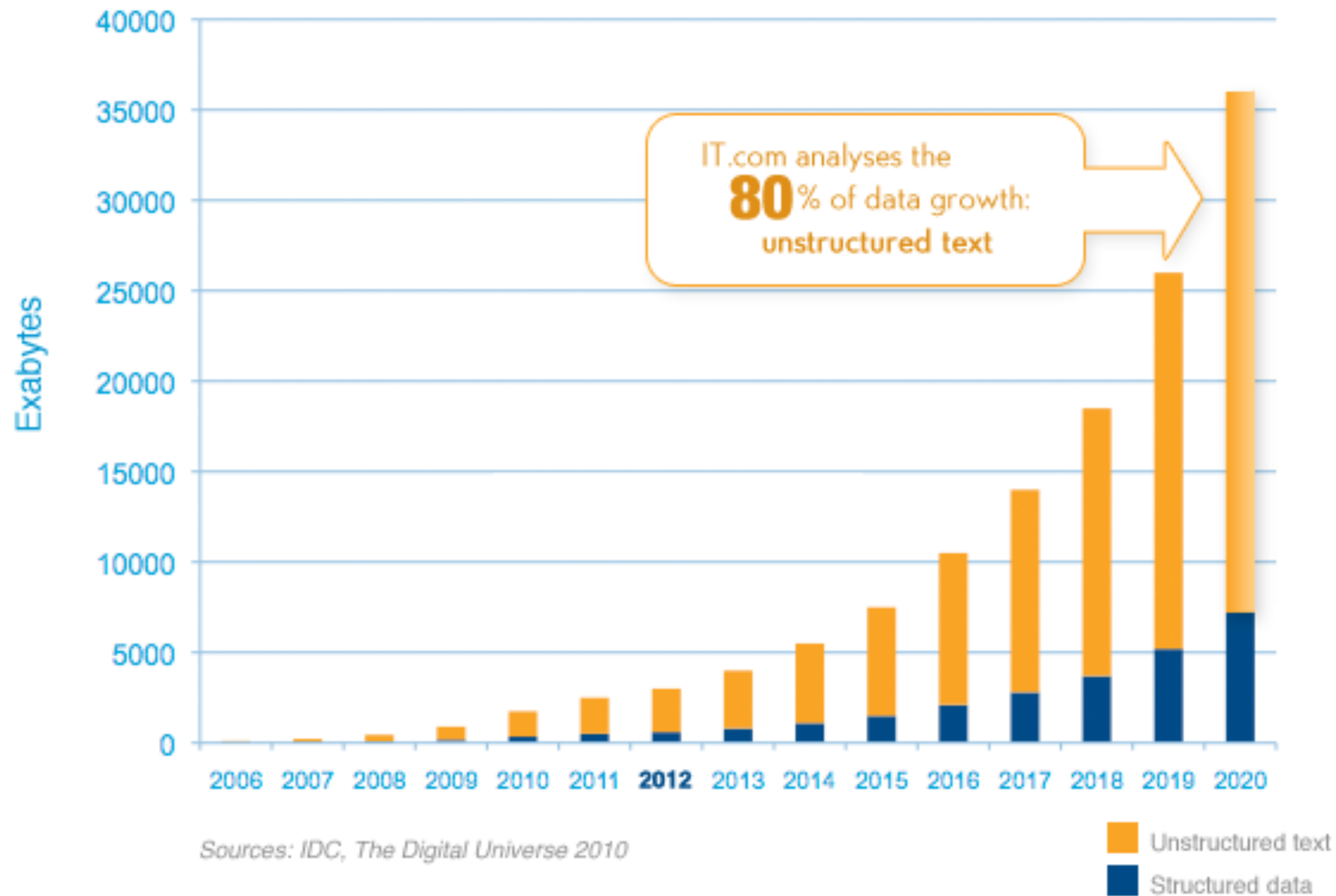
different across population groups
(countries, accents, religions,...)

linear perception

semi-structured (content: grammar, words, sentences,
paragraphs,.. ; appearance: typography, calligraphy,..)

Why Visualize Text?

Worldwide Corporate Data Growth



Design and Text

Typography:

typefaces (serif, sans-serif, **bold**, *italic*)

point size (10pt, 12pt, 24pt, 36pt..) 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

ß

Creating a font type is an art
that requires profound design knowledge

AV Wa
No kerning

AV Wa
Kerning applied

fi → fi

fl → fl

Oscars and Typography

Wrong Movie
announced for Best
Picture

Failure of
Typography

Larger Failures in a
Complicated System



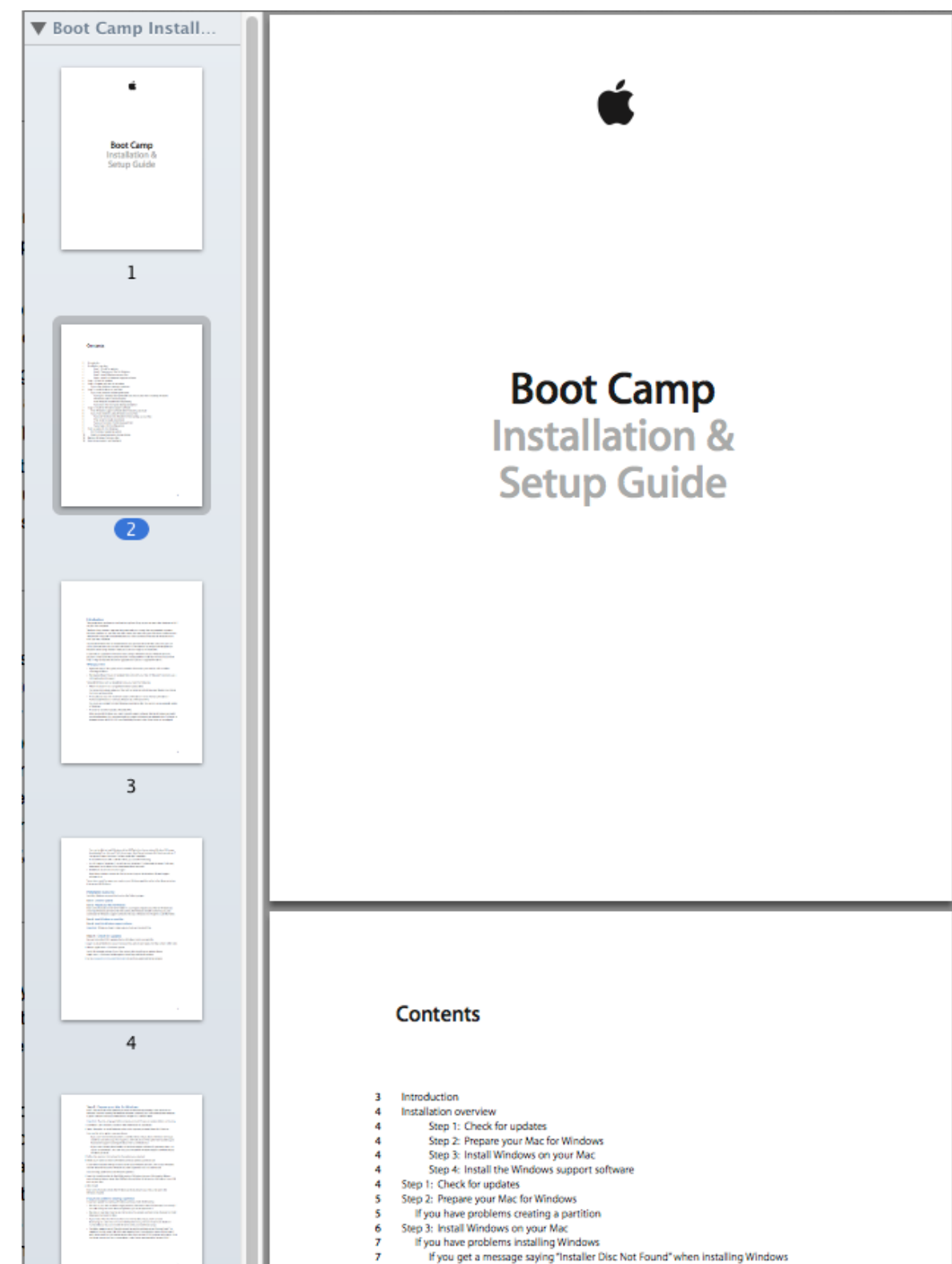
Visualization for “Raw” Text

in daily use..

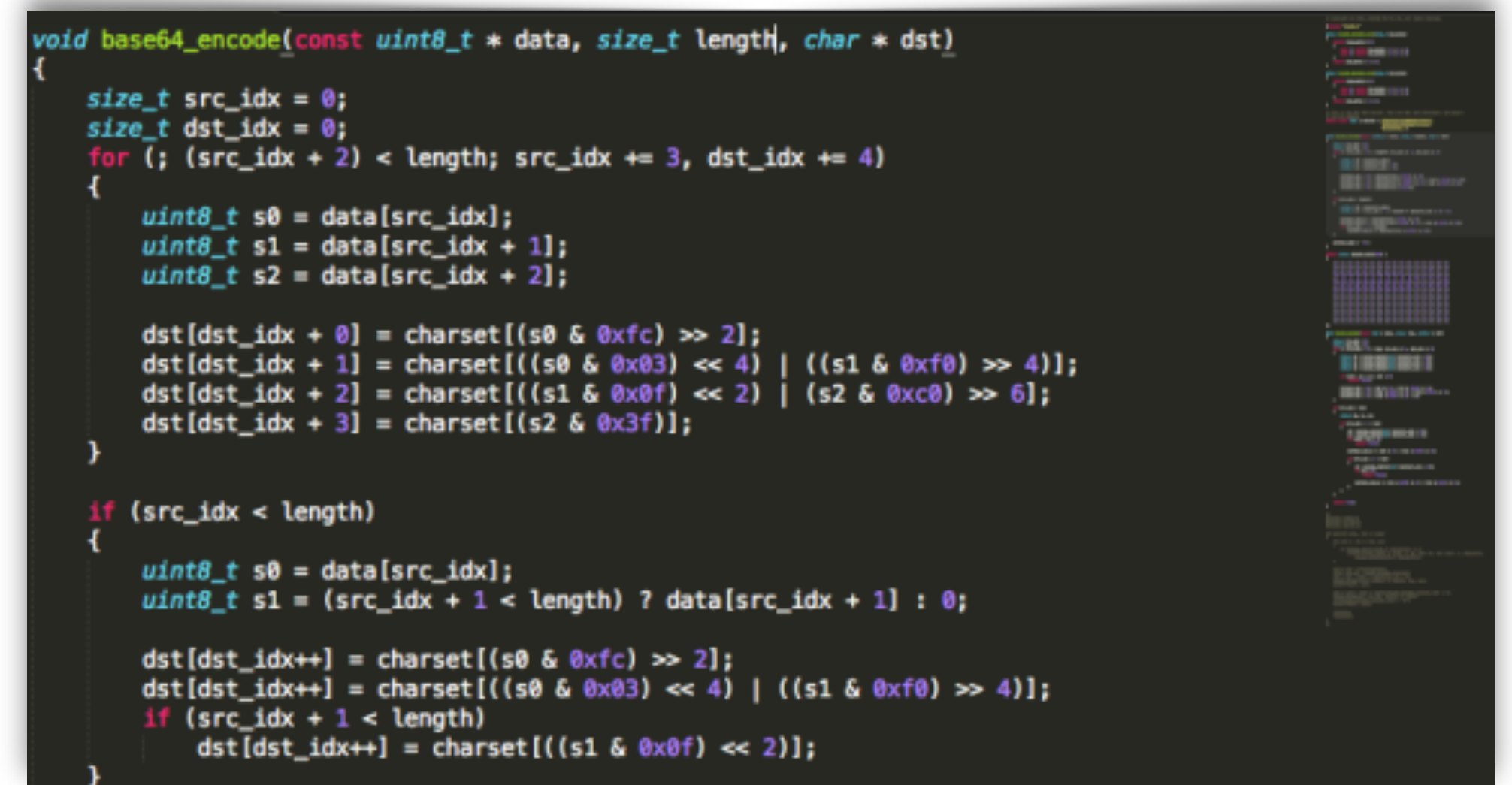
enriched text – hypertext linking (graph navigation)



overview & detail



highlighting semantics



Visualization for “Raw” Text

Document Lens

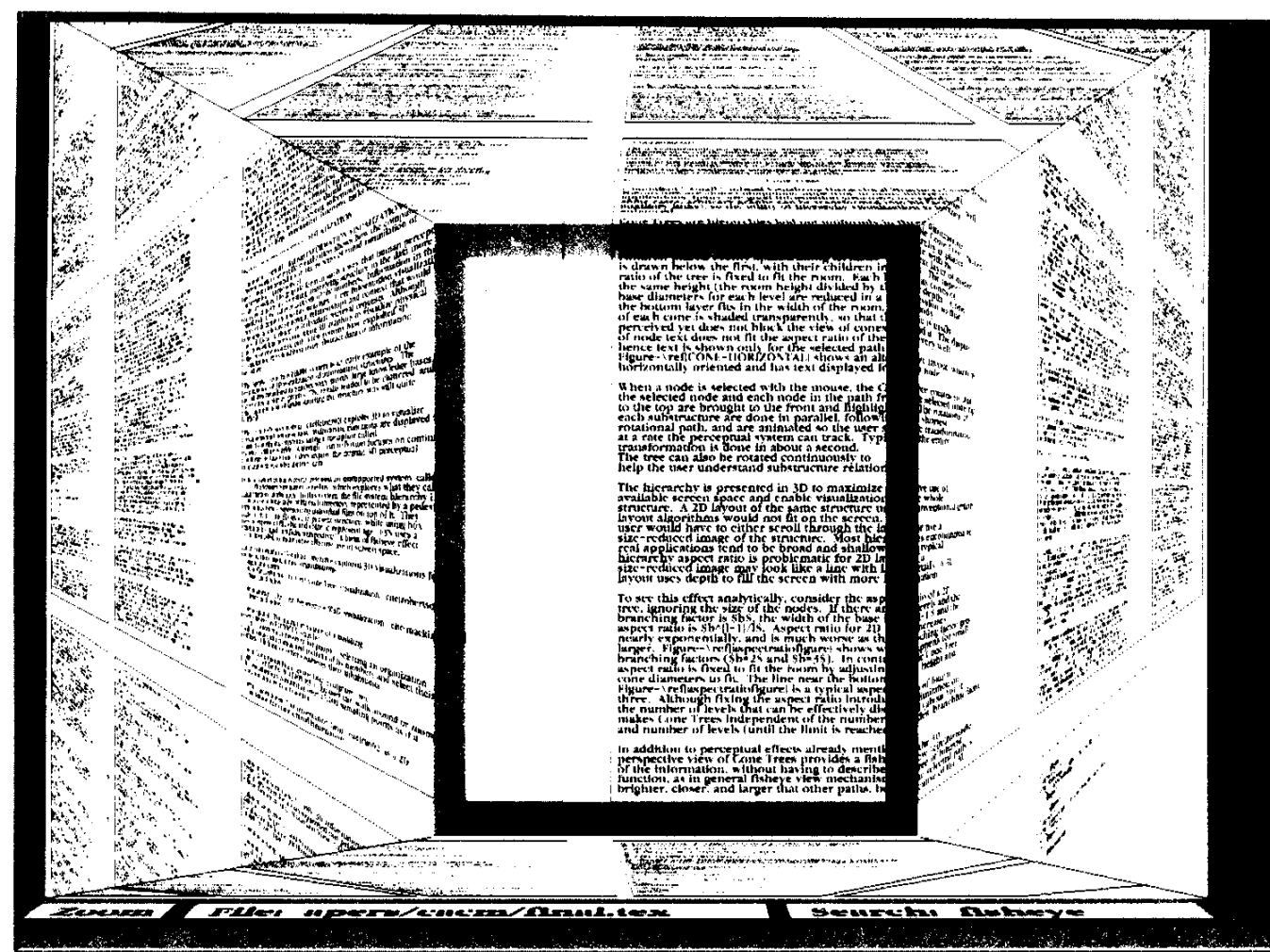


Figure 3: Document Lens with lens pulled toward the user. The resulting truncated pyramid makes text near the lens' edges readable.

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.

Document Thumbnails with Variable Text Scaling
A. Stoffel, H. Strobel, O. Deussen, D. A. Keim
Computer Graphics Forum, volume 31 issue 3 pp.

Visualizing Search Results

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 pre-views. 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 im-

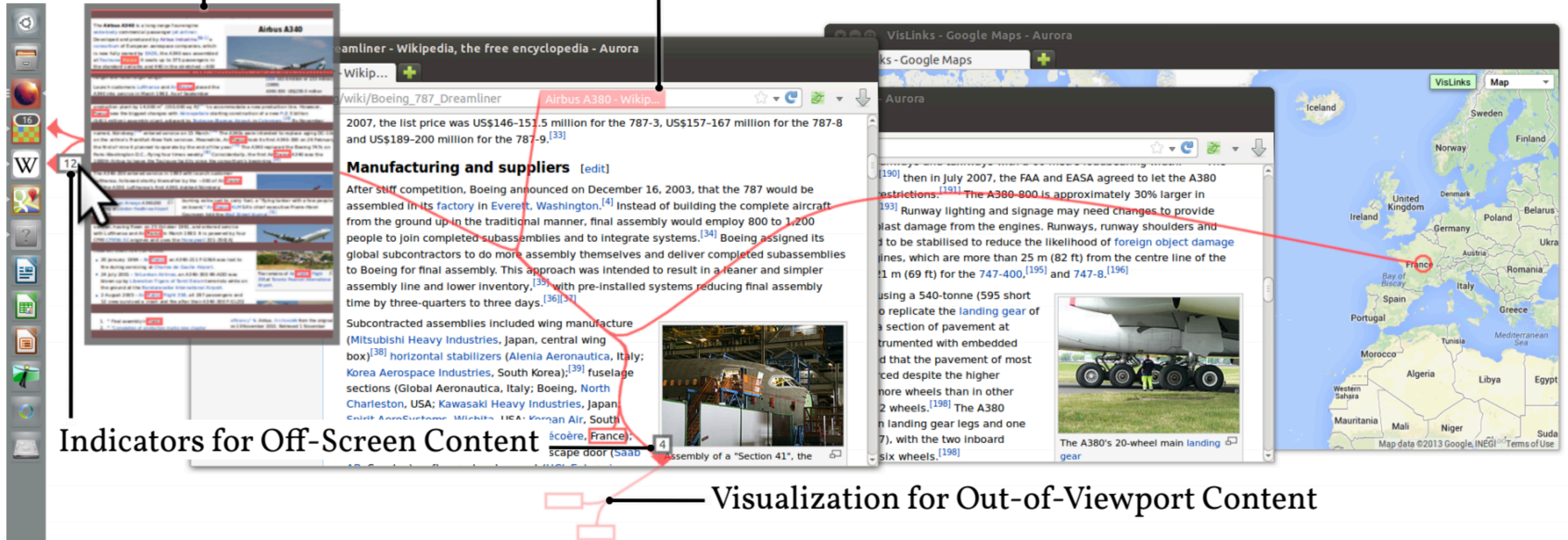
is used that highlights the keywords and their context. Other applications might use a different interest function, for instance a sentiment score could be used to create thumbnails for sentiment analysis.

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 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

Visualizing Hidden Text

Smart Preview for Off-Screen Content

Indicators for Occluded Windows



Working with Text

unstructured text



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

...

structured data

Structured Text Features

simple counts (bag of words)
used for similarity measures

	princess	dragon	castle
doc1	1	1	1
doc2	0	0	1

Processing to Derive Features

Typical steps are:

- cleaning (regular expressions)

- sentence splitting

- change to lower case

- stopword removal (most frequent words in a language)

- stemming

- POS tagging (part of speech)

- noun chunking

- NER (name entity recognition)

- deep parsing - try to “understand” text.

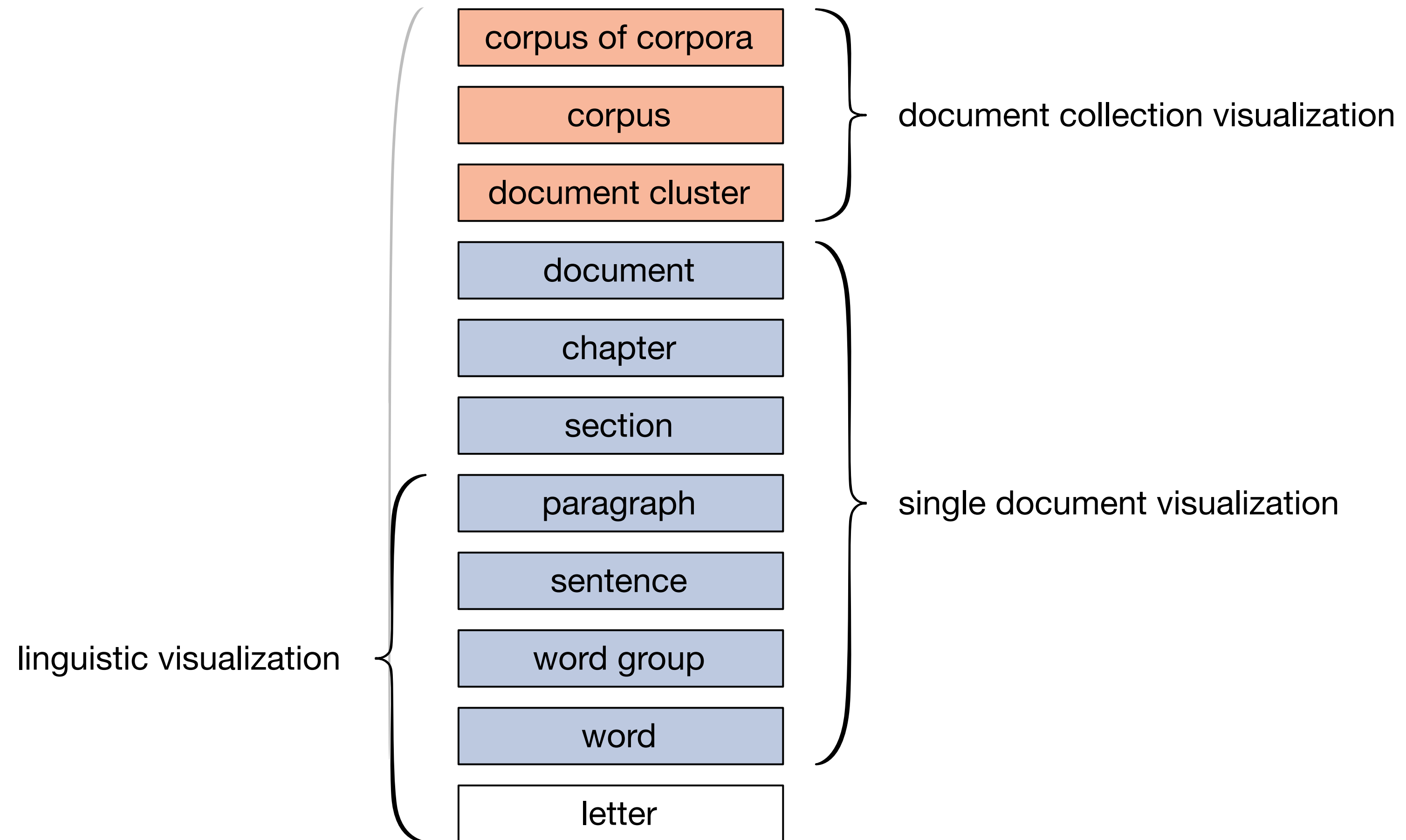
Text features are complicated

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?

Text Units Hierarchy



Types of Text Visualizations

Document Visualization

Corpus Visualization

Visualization for NLP

Creativity Support

Document Visualization

Wordle vs Tag Cloud

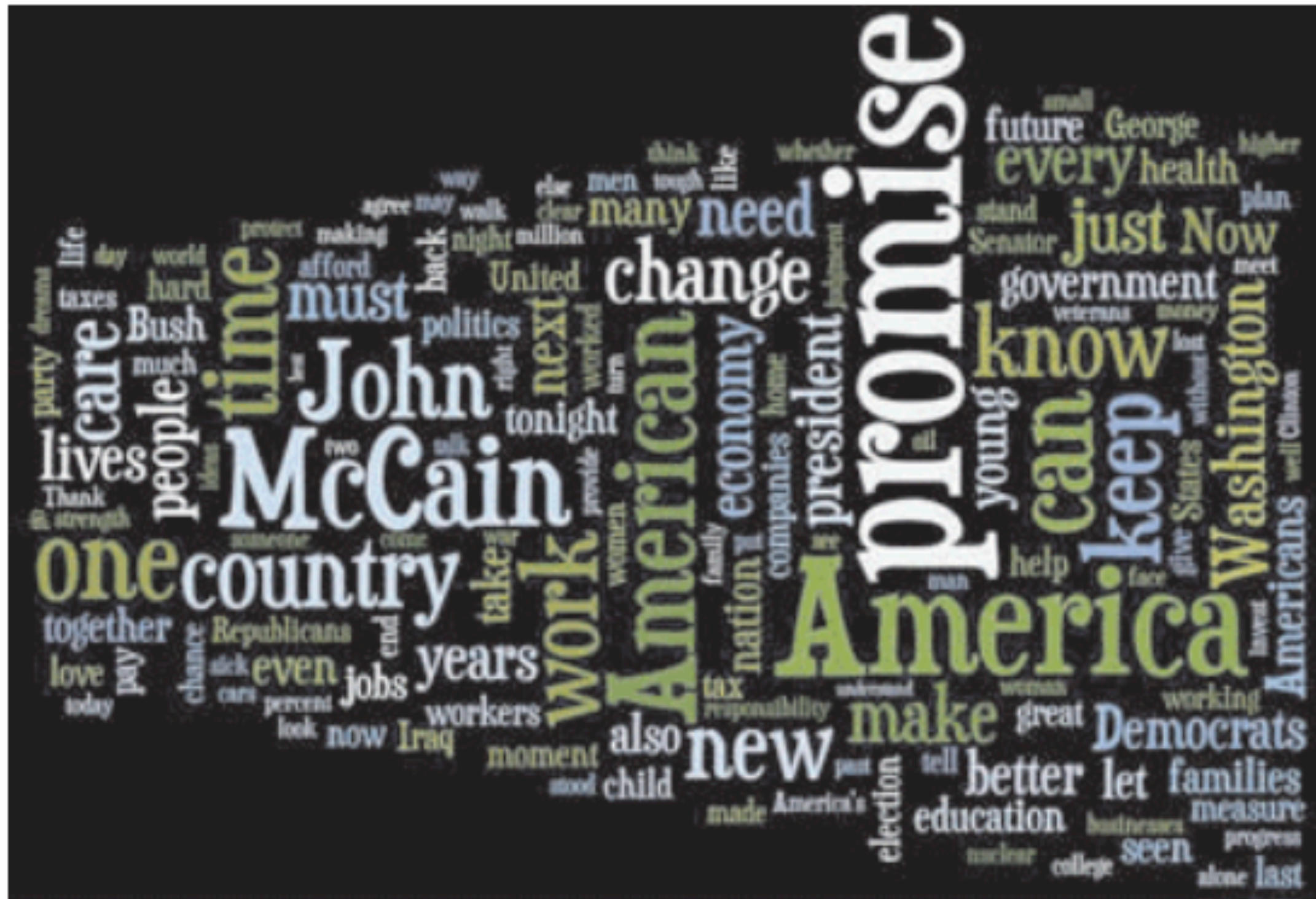


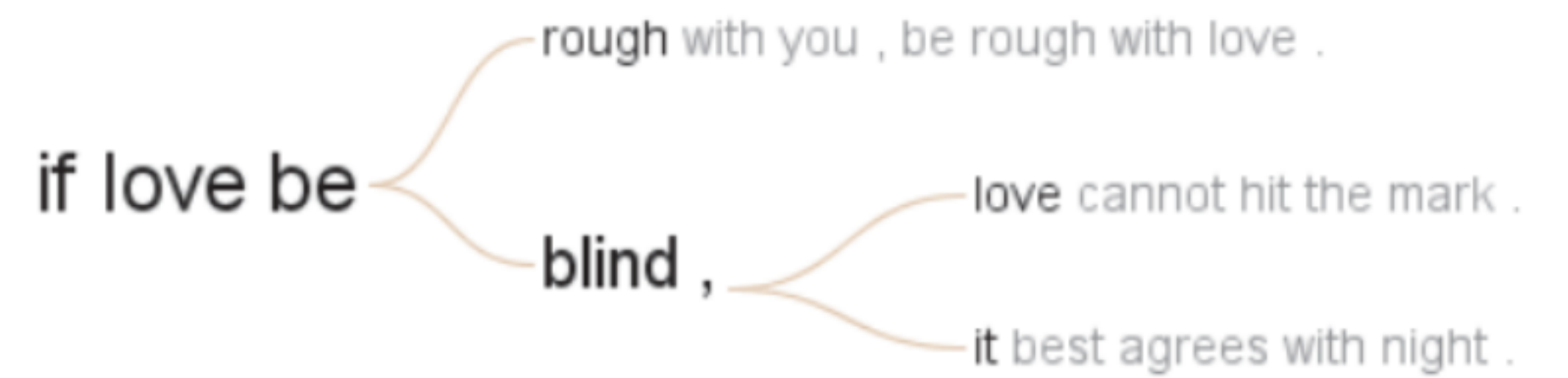
Fig 2: Wordle vs. Tag Cloud of Barack Obama's speech at the Democratic Convention in 2008.

Word Tree

Text

if love be rough with you , be rough with love .
if love be blind , love cannot hit the mark .
if love be blind , it best agrees with night .

WordTree



PhraseNets

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:

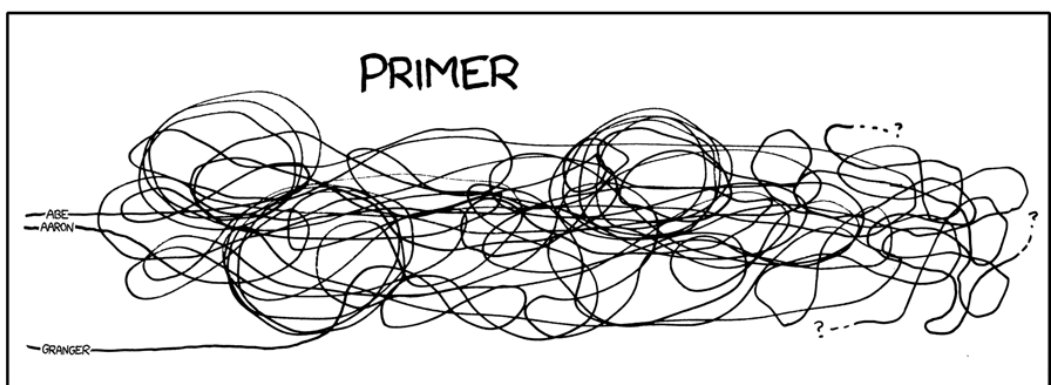
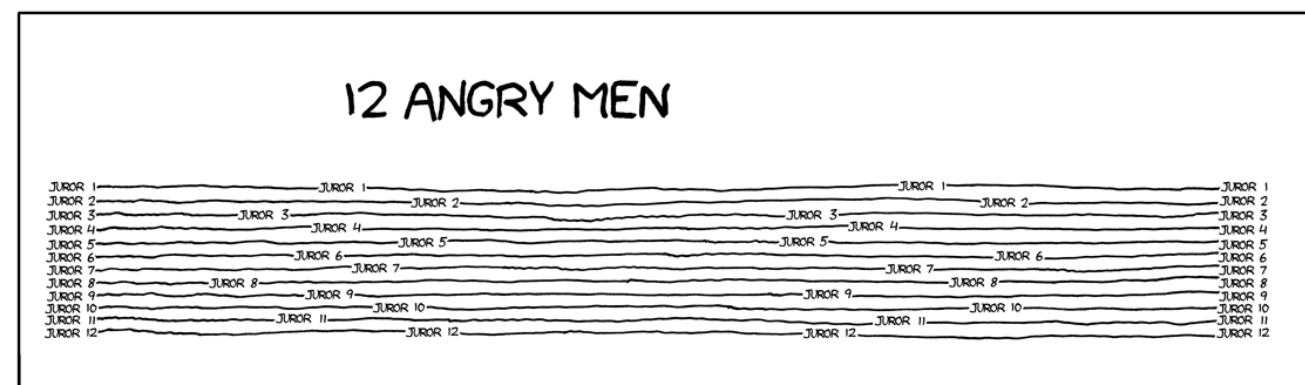
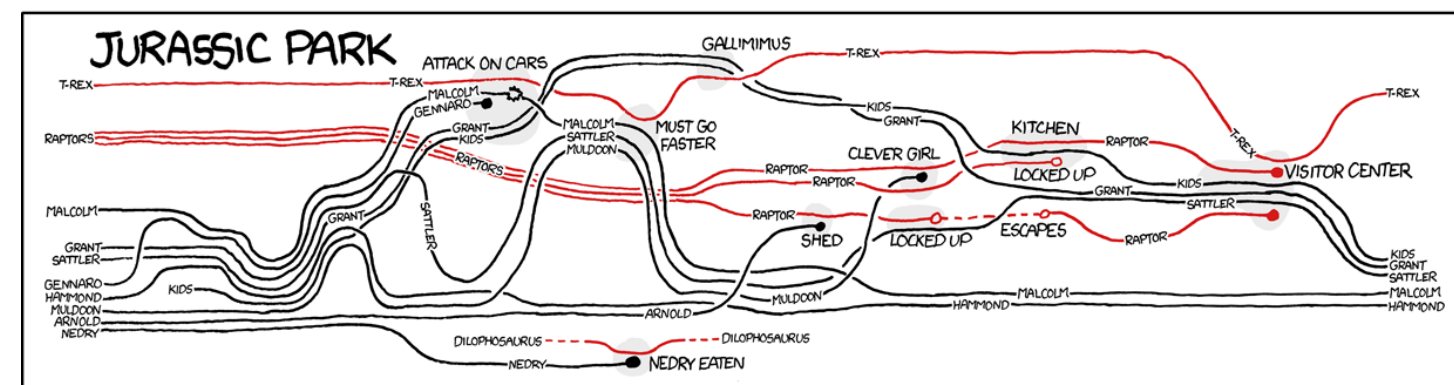
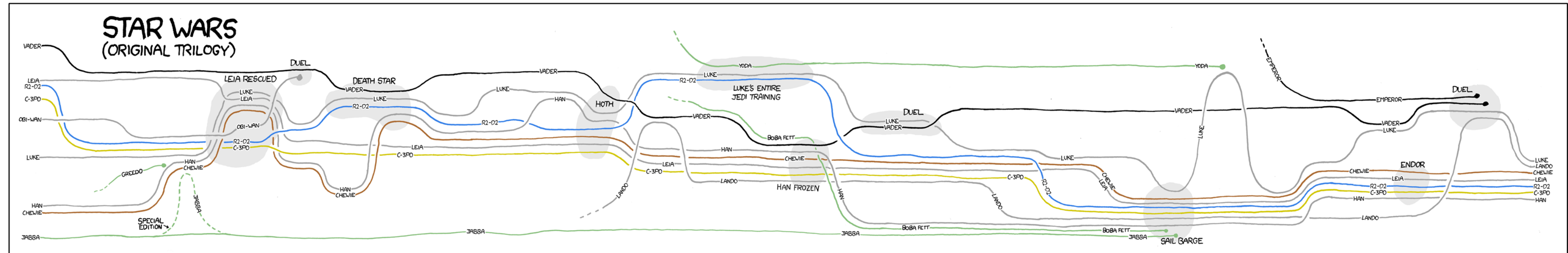
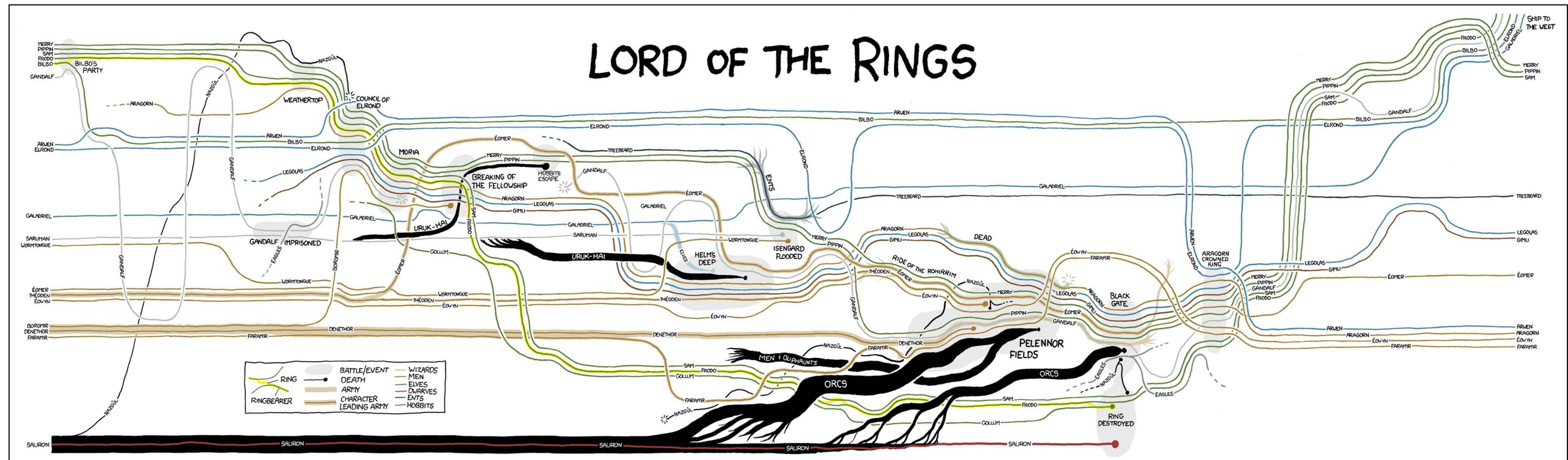
pride → **impertinence**

Frank van Ham, Martin Wattenberg, and Fernanda B. Viegas.

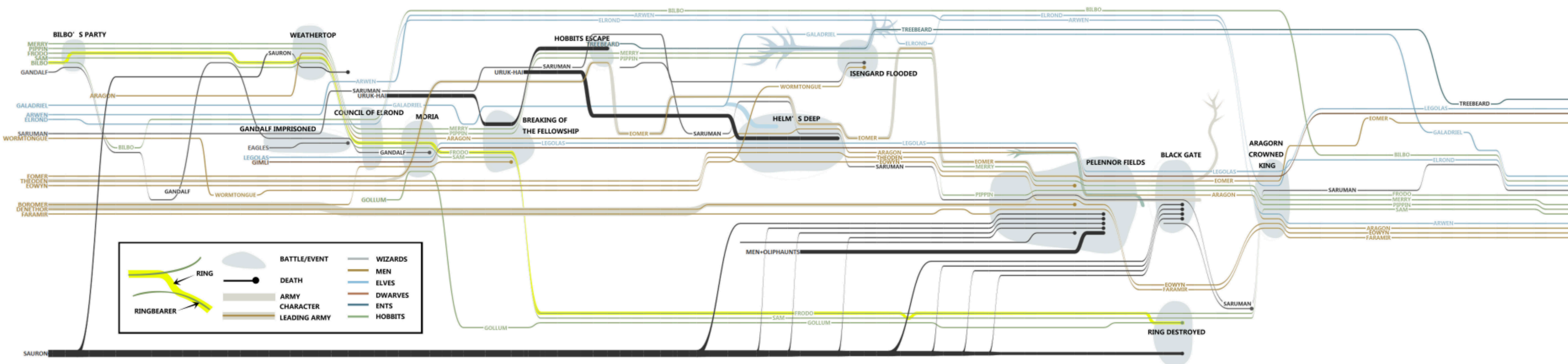
Mapping Text with Phrase Nets.

IEEE Transactions on Visualization and Computer Graphics 15, 6 (November 2009)

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.



StoryFlow: Tracking the Evolution of Stories



Visualizing Corpora

Text Corpora

Varied Goals:

Discover interesting documents

Summarize Documents

Classify Documents

Extract Facts (Intelligence Analysis)

Rich Information:

Document Metadata

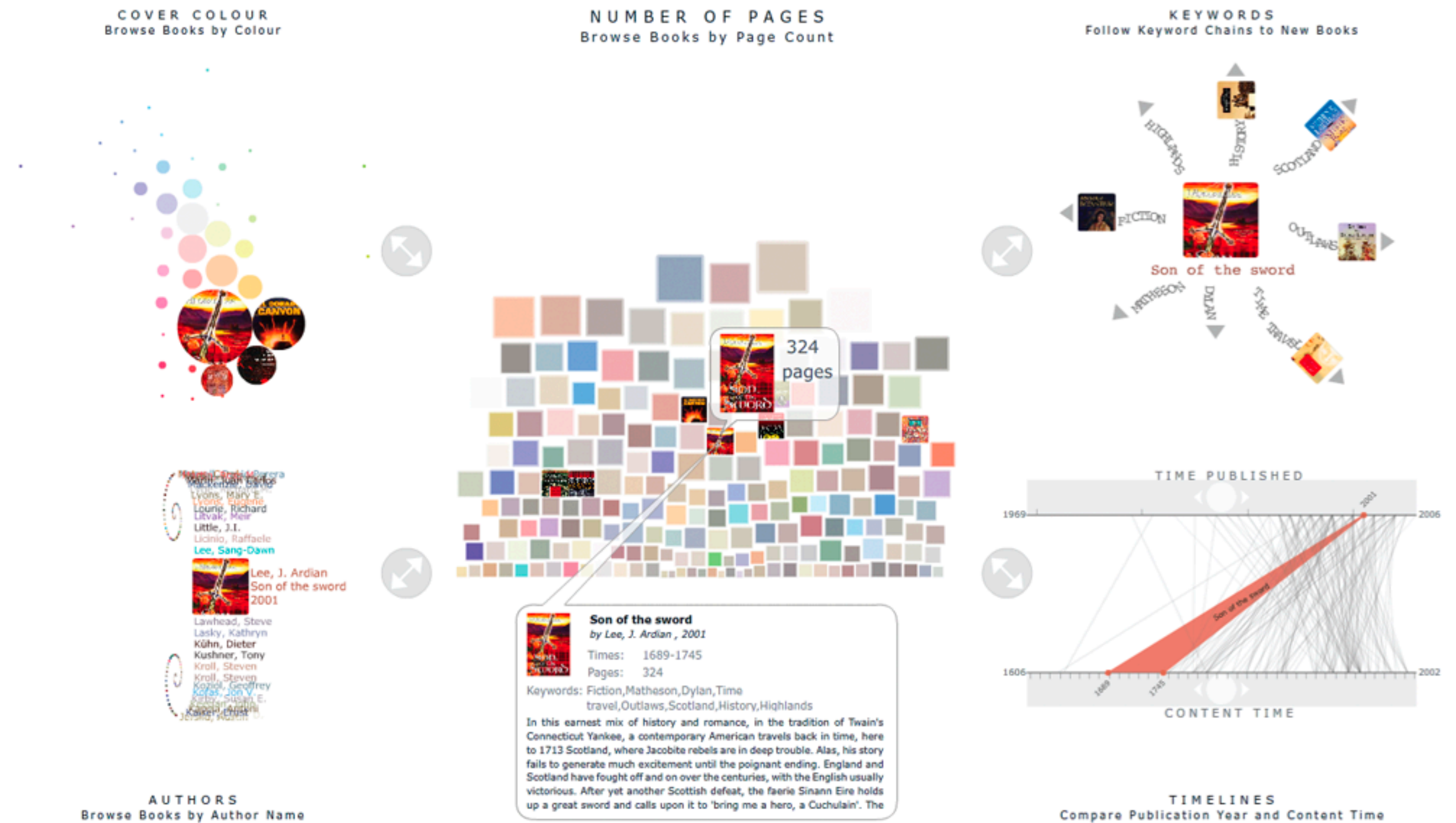
Authors, date, type,

Paragraphs, figures...

Revisions, annotations, comments,

Visualize Library Catalogs

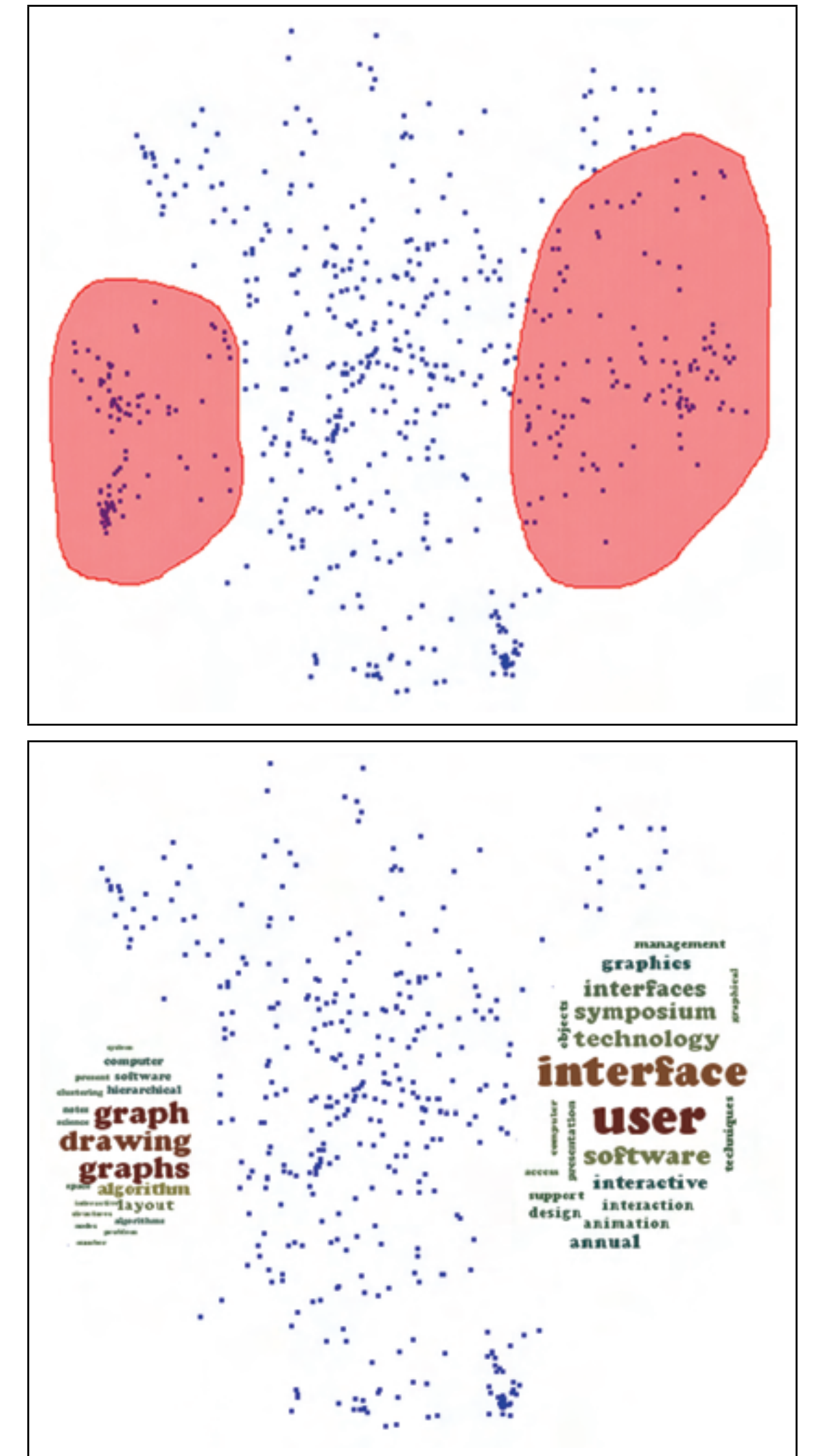
Bohemian Bookshelf



Corpora: MDS Approaches

use bag-of-words to project documents w.r.t. text similarity into a landscape

(only) one example



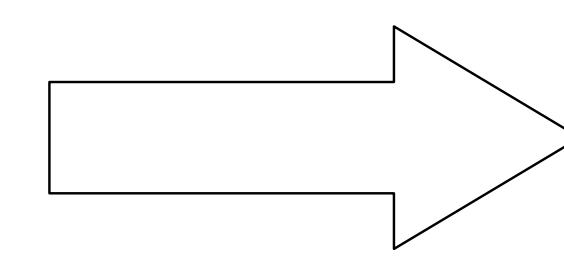
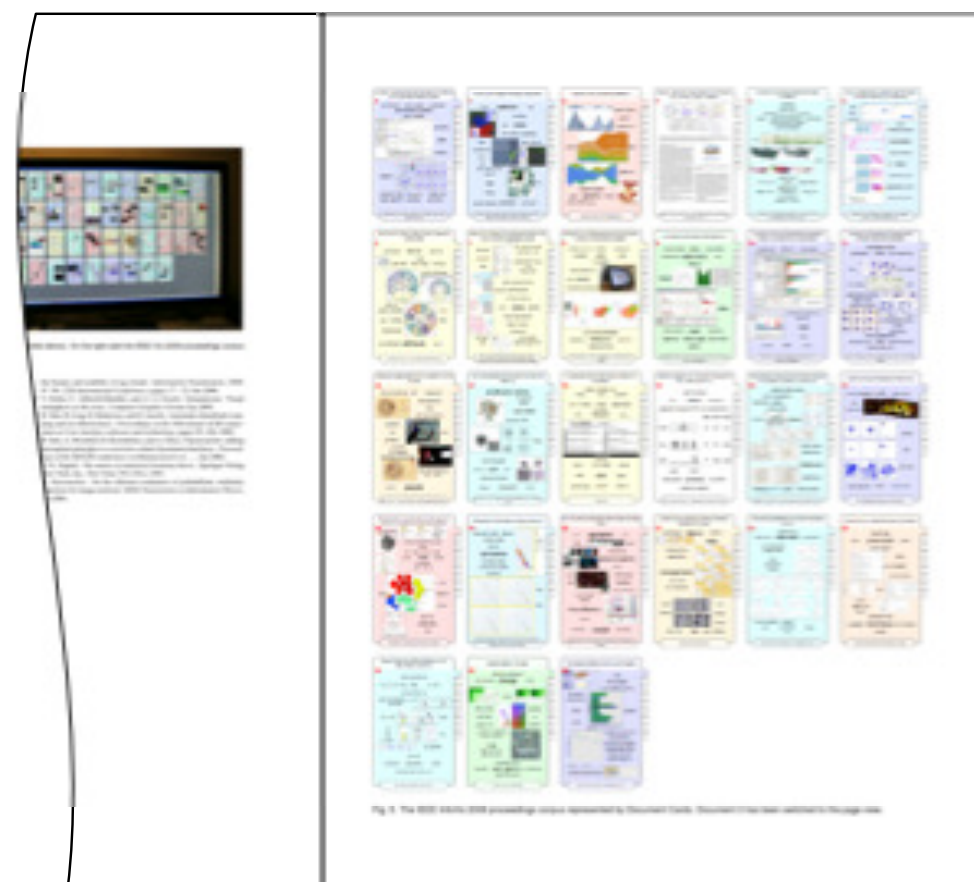
Fernando V. Paulovich, Franklina M. B. Toledo, Guilherme P. Telles, Rosane Minghim, and Luis Gustavo Nonato.
Semantic Wordification of Document Collections.
Comp. Graph. Forum 31, 3pt3 (June 2012)

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).

DocumentCards

summarize scientific documents using important terms and important figures

represent the document's content as a mix of figure and text



Document Cards: A Top Trumps Visualization for Documents

compound noun image term extraction

noun phrase

approach

thumbnail

1

2

3

4

5

6

7

8

9

problem

section

paper

descriptive text document card

Hendrik Strobelt, Daniela Oelke, Christian Rohrdantz, Andreas Stoffel, Daniel Keim, and Oliver Deussen

0 Cerebral: Visualizing Multiple Experimental Conditions on a Graph with Biological Context
systems biologist context interaction graph graph model dataset figure edge tool cell gene layout algorithm process node cerebral
Aaron Barsky, Tamara Munzner, Jennifer Gerdy, and Robert Kincaid

1 Multi-Focused Geospatial Analysis Using Probes
probe interface participant type window region-of-interest local region data application
Thomas Bullock, Wenwen Dou, Zachary Warren, William Ribarsky, and Ramco Chang

2 Stacked Graphs: Geometry & Aesthetics
question visualization paper type author color comment york time layer order namevayer graphic trend people time sey system legibility design issue layout method
Lee Byron and Martin Wattenberg

3 Vispedia: Interactive Visual Exploration of Wikipedia Data via Search-Based Integration
Byan Chan, Leslie Wu, Justin Talbot, Mike Comarum, and Pat Hanrahan

4 Geometry-Based Edge Clustering for Graph Visualization
edge bundle technique polyline segment large graph control mesh straight line mesh edge pattern transfer function user node position control point graph layout result method visual clutter general graph primary direction
Weiwei Cai, Hong Zhou, Student, Huanmin Qi, Pak Chung Wong, and Xiaoming Li

5 VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery
map rrs feed participant temporal information item data information space query parameter exploration set visget description
Marlan Dirk, Sheelagh Corpendale, Christopher Collins, and Carey Williamson

6 Who Votes for What? A Visual Query Language for Opinion Data
user study report attribute paper sample population entity result sector opinion poll state street typical data set user interface visualization visual query language design participant ring system data point
Geoffrey M. Draper and Richard F. Riesenfeld

7 Exploration of Networks Using Overview+Detail with Constraint-based Cooperative Layout
layout method route focal node user lod placement constraint edge rout level high quality layout primary graph large network part system detailed view uml class diagram display model layout technique cluster position structure focus node
Tim Davies, Kim Marriott, Falk Schreiber, Peter J. Stuckey, Michael Woodward and Michael Wyllow

8 Rolling the Dice: Multidimensional Visual Exploration Using Scatterplot Matrix Navigation
visual exploration data dimension query prototype implementation visual representation cameras digital camera dataset figure overview grand tour range scatterplot matrix user operation method plane order
Miklas Dimpolt, Pierre Dragicevic, and Jean-Daniel Fekete

9 Interactive Visual Analysis of Set-Typed Data
bar block user scatterplot figure feature width data item dataset data record histogram washing agent set-typed data view
Wolfgang Freiler, Kresimir Markotić, Computer Society, and Helwig Hauser

10 Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation
graphical history usage history item rule tableau image data field display approach event history interface history tool
Jeffrey Heer, Jack D. Mackinlay, Chris Stolte, and Maneesh Agrawala

11 Improving the Readability of Clustered Social Networks Using Node Duplication
representation success rate social network time clonode analysis visualization duplicate community noduplication significant effect cluster duplication link participant splitlink readability
Nathalie Henry, Anastasia Bejarano, and Jean-Daniel Fekete

12 EMDialog: Bringing Information Visualization into the Museum
touch emily carr interaction design node installation visual appeal statement tree diagram perspective information visualization cut section museum visitor tree ring public space data representation people museum context
Uta Hinrichs, Holly Schmidt, and Sheelagh Corpendale

13 On the Visualization of Social and other Scale-Free Networks
scale-free network shortest path node degree visualization layout time distance matrix weighted graph geodesic cluster grime original graph edge filter edge metric power-law graph hub node
Nuntao Jia, Jared Hoberock, Michael Garland, and John C. Hart

14 A Framework of Interaction Costs in Information Visualization
visualization user intent study focus-lock mode paper interaction cost participant time interface operation menu framework evaluation focus lock section interaction target overview window subject gulf
Heidi Lam

15 Distributed Cognition as a Theoretical Framework for Information Visualization
cognitive property process endangered spacy infovis system cognition insight internal representation Raki! Raki! Raki! Developer cognitive system report theory artifact cognitive science analyst document view
Dichang Li, Nancy J. Nersessian, and John T. Stasko

16 Particle-Based Labeling: Fast Point-Feature Labeling without Obscuring Other Visual Features
section virtual particle method distant particle conflict particle figure labeling step solution point-feature visual element number time result collision map
Martin Luboschik, Heidrun Schumann and Hilke Conds

17 Rapid Graph Layout Using Space Filling Curves
matrix order space filling curve based layout approach peano curve aspect ratio graph layout fast screen space
Chris Maulder and Kwan-Liu Ma, Senior

18 HIPP: A Novel Hierarchical Point Placement Strategy and its Application to the Exploration of Document Collections
complexity cluster space result hipp plane relationship data instance group node data set approach
Fernando V. Paulovich and Rosane Minghim

19 Effectiveness of Animation in Trend Visualization
group trace line visualization data point user study task presentation participant animation and trace
George Robertson, Roland Fernandez, Daniel Fisher, Benjamin Lee, and John Stasko

20 Viz-A-Vis: Toward Visualizing Video through Computer Vision
condition pattern recognition computer vision aggregate activity table activity cube law data system heat map frame level model motion time view aggregation step figure difference
Mario Romero, Jay Sumner, John Stasko, and Gregory Abowd

21 Balloon Focus: a Seamless Multi-Focus+Context Method for Treemaps
enclosure context technique relative position foci enlargement dependency graph elastic edge result non-focus item player subject view layout focus line balloon focus original treemap task
Ying Tu and Han-Wai Shen

22 Perceptual Organization in User-Generated Graph Layouts
visual characteristic: data degree network set uniform edge length cluster number structure algorithm node human observer edge crosse condition study
Frank van Ham and Bernice E. Rogowitz, Senior

23 The Word Tree, an Interactive Visual Concordance
tree structure common word user time option data design tag cloud search term title love romeo and juliet token word tree eye context number suffix tree branch blind king james bible
Martin Wattenberg and Fernanda B. Viégas

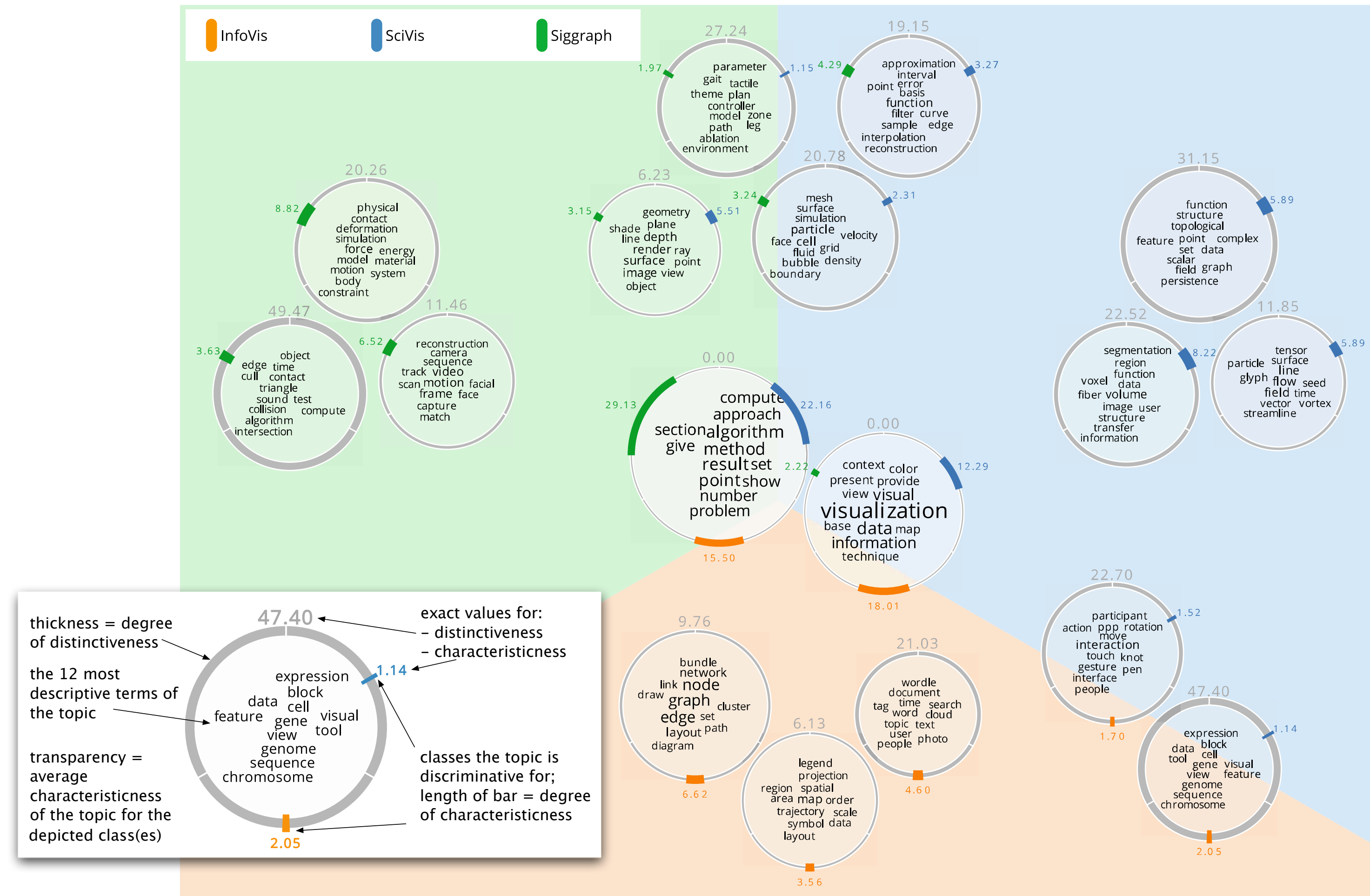
24 Evaluating the Use of Data Transformation for Information Visualization
comment performance distribution task system step error rate non-transformed data choice technique study experiment noise context figure data set data transformation anova test analytic task data property criteria visualization participant benefit impact task type
Zhen Wen and Michelle X. Zhou

25 Spatially Ordered Treemaps
readability score node order data geographic location angular change color space position sequence scene type distortion unit size problem histogram layout treemap node spatial layout image layout algorithm relationship consistency arrangement displacement vector
Jo Wood, and Jason Dykes

26 The Shaping of Information by Visual Metaphors
information visual metaphor verbal metaphor type response time understand incompatible question study correct response task question session difference visual representation participant
Caroline Zemkiewicz and Robert Kosara

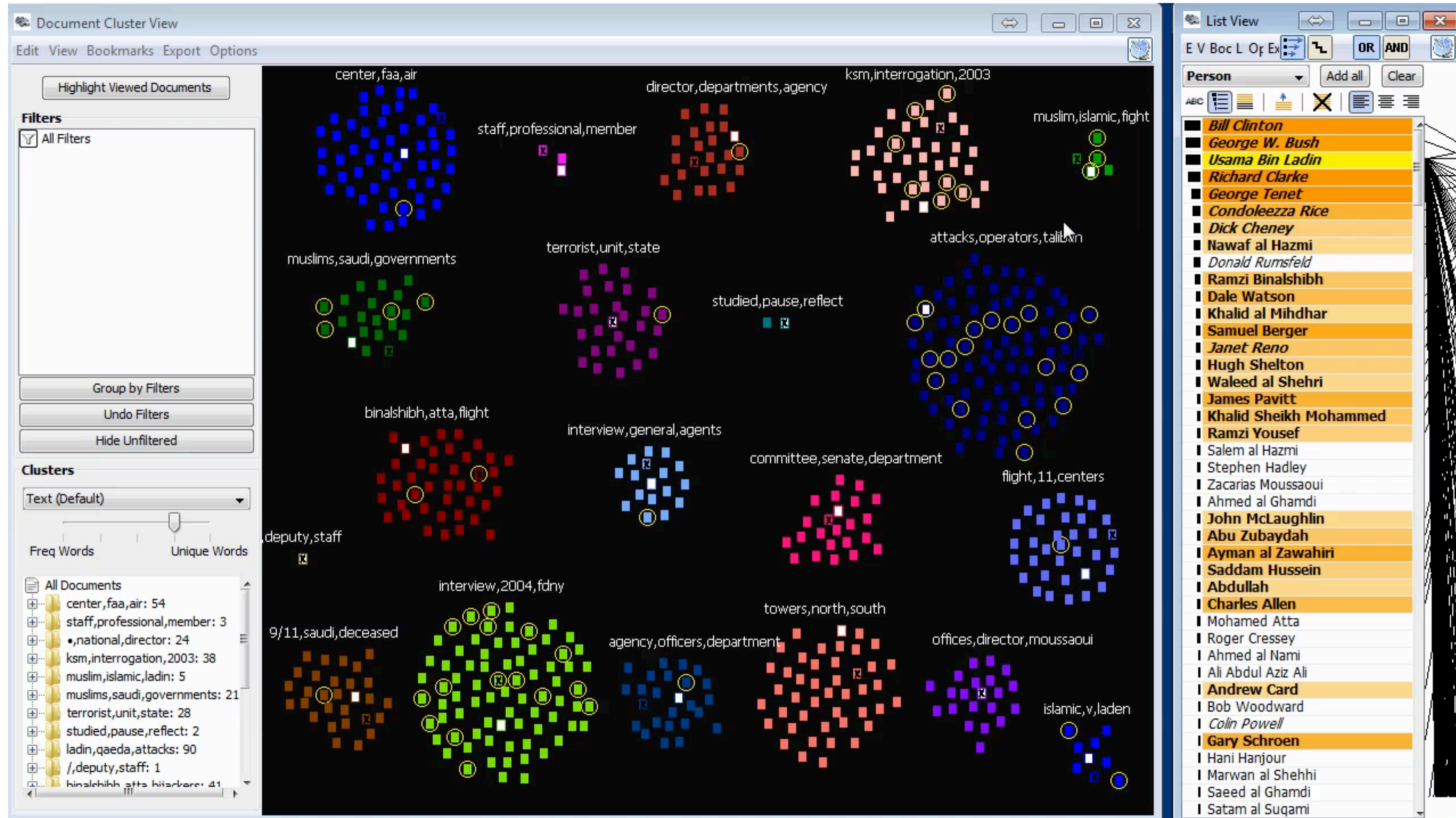
Compare Corpora

Compare topics between text collections



Comparative Exploration of Document Collections: a Visual Analytics Approach (<http://ditop.hs8.de>) Figure 1: Comparison of 495 papers of InfoVis, SciVis, and Siggraph (discrimination threshold = 6, number of topics = 30)
 D. Oelke, H. Strobel, C. Rohrdantz, I. Gurevych, and O. Deussen

JigSaw – Intelligence Analysis



Extracting and Linking Info From Documents

The image displays a 'Concept Graph' interface in a Mozilla Firefox browser. The central part of the interface is a network diagram with nodes and relationships. The nodes include 'POK', 'Silvia Marek', 'Elian Karel', 'Hank Fluss', 'Sten Sanjorge', and 'GASTech'. Relationships are labeled as 'leader', 'prior leader', 'father - son', 'Chief Operating Officer', and 'CEO'. Some nodes have small circular icons with text like '10 year history', '5 year report clean', and 'found'. The interface also has a sidebar on the left with options like '+ Add Concept' and 'Select All Concepts and Relations'. Below the graph, there are two document excerpts. The first excerpt is from a document titled 'ear historical document clean' and contains text about the formation of an SMO and the POK's agenda. The second excerpt is from a document titled '5 year report clean' and contains text about the POK's history and its role as a grassroots movement. A red line connects a highlighted sentence in the first excerpt to the 'Hank Fluss' node in the graph. A blue line connects a highlighted sentence in the second excerpt to the 'POK' node in the graph.

Concept Graph - Mozilla Firefox
file:///home/tom/Documents/hidden-content/links/addons/concept-graph/index.html

Concept Graph

+ Add Concept Shift+A
Select All Concepts and Relations Control+A

ear historical document clean - Mozilla Firefox
/home/tom/Dropbox/master/va

One of the critical steps for the formation of an SMO is to establish an identity that will help bring their message to the citizenry and the government. Osvaldo proposed to the activists they form a social movement organization with an identity brand and a specific agenda: To bring clean water to Elodis and clean up the contamination in the River. The group formalized their identity with a name, the Protectors of Kronos (POK), and a logo, consisting of an open right hand within a white circle on a black background.

Osvaldo reached out to an international agency specializing in clear water for communities, Wellness for All (WFA). The WFA Project Manager Joclyn Reynolds began formal scientific testing of the Tiskele River water, and advised the POK to engage the GASTech company regarding the issue of water contamination.

Members of the POK repeatedly requested meetings with GASTech representatives, but received nothing but denials for several months. This continued until Hank Fluss, the Chief Operating Officer at GASTech, agreed to a meeting with Bodrogi. The meeting took place outside range of media, and involved only Fluss, Bodrogi and Osvaldo. Bodrogi reported he felt encouraged by the seriousness with which Fluss took the POK agenda, and told the POK he would take their issues back to the CEO of GASTech, Sten Sanjorge, Jr.

Events Take a Turn for the Worse

Up to this point, the POK had primarily used statistics about health issues and names of toxins in their agenda, and then on August 18 1998 ten-year old Juliana Vann, daughter of Lemual and Neske Vann, died of leukemia associated with benzene toxicity.

10 year history

5 year report clean - Mozilla Firefox
file:///home/tom/Dropbox/master/vast14-n

History of the Protectors of Kronos
A Psycorps Analysis Brief
By Fredrick N. Wagner and Westley B. Andrews
January 2009

The Protectors of Kronos (POK) is a political activist movement that seven citizens concerned about contamination from drilling at the Tisk the POK has grown under the charismatic leadership of Elian Karel to with an estimated membership of 200-300 people.

This report summarizes the history of the POK and assesses the likely

The POK as a Grassroots Movement (1997-2001)

The protectors of Kronos emerged from the Elodis township, a rural 6500 persons that lies 25 km from Abila, capital city of Kronos. The primarily engaged in floodplain farming which is dependent upon the In early 1997, citizens of Elodis began to be concerned about an abn occurrence of illnesses such as cancer, birth defects, respiratory illness: diseases, in addition to a marked decrease in crop yield. When the El take action on the citizen's call for investigation into possible contami grassroots group of seven citizens formed with the goal of bringing th Kronos government.

The grassroots organization coalesced under the leadership of Henk E who had joined the group after his wife had become ill with cardiopu consistent with ethylene glycol contamination. Bodrogi was a newly

Collection of Tweets



Visualization for NLP

GLTR: Statistical Detection and Visualization of Generated Text, Gehrmann, Strobel Rush: <http://gltr.io/dist/index.html>

LSTMVis: Visual Analysis for Recurrent Neural Networks, Strobel et al.: <http://lstm.seas.harvard.edu/>

Visual Exploration of Semantic Relationships in Neural Word Embeddings. Liu et al.

Visualization for Creativity Support

Poemage: Visualizing the Sonic Topology of a
Poem. McCurdy et al. [http://www.sci.utah.edu/~nmccurdy/
Poemage/](http://www.sci.utah.edu/~nmccurdy/Poemage/)

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

Text Visualization Browser

A Visual Survey of Text Visualization Techniques

Provided by ISOVIS group

[About](#) [Add entry](#) [Contact](#)

Techniques displayed: **141**

Search:

Time filter: 1976 2014

Analytic Tasks

- Sum
- Alert
- Like
- Share
- Refresh
- Print
- ...

Visualization Tasks

- Star
- Save
- Sort
- Hide
- Zoom
- ...

Data

Source

- File
- Folder
- Share

Properties

- Info
- Refresh
- Settings
- ...

