Text and Document Visualization

Hendrik Strobelt - hstrobelt@seas.harvard.edu

housing day 2015

CS 171 - Visualization

HARVARD
School of Engineering
and Applied Sciences

/Users/hen> whoami

Dipl. (TU Dresden)

Dr. (Uni Konstanz)

PostDoc (Harvard SEAS)

PostDoc (NYU Poly)

Text Visualization




Visualization for Sciences


Layout


This Week

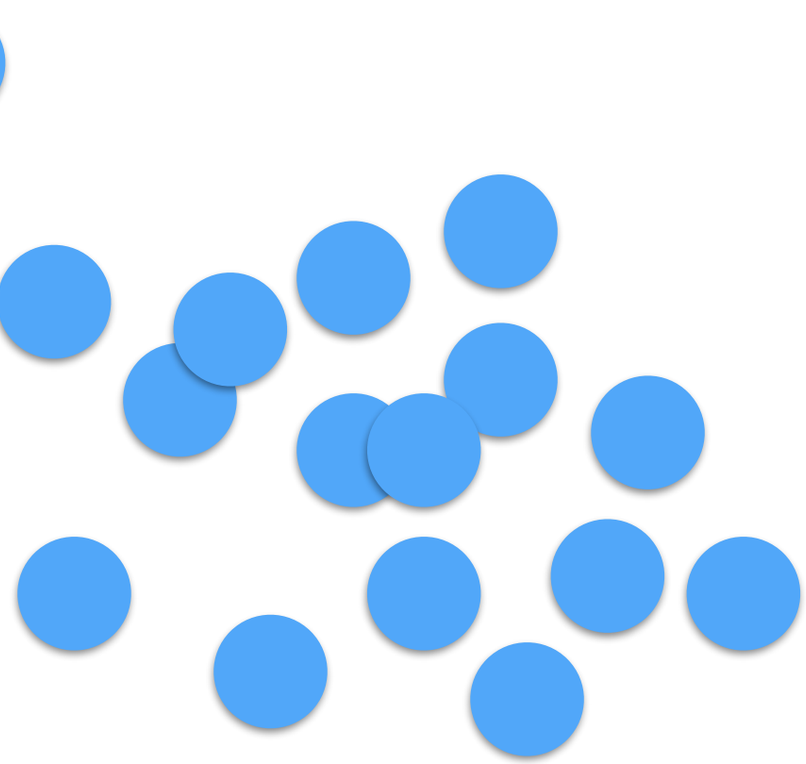
- HW2 (due to **FRIDAY** — 11:59 pm):
    - include design studio solutions
- Section 6 special **TODAY** at 4pm MD G125

A little experiment

Task: How many dots?

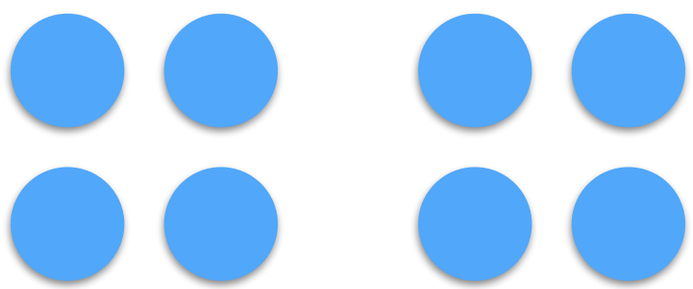
thanks to Martin Krzywinski

A little experiment

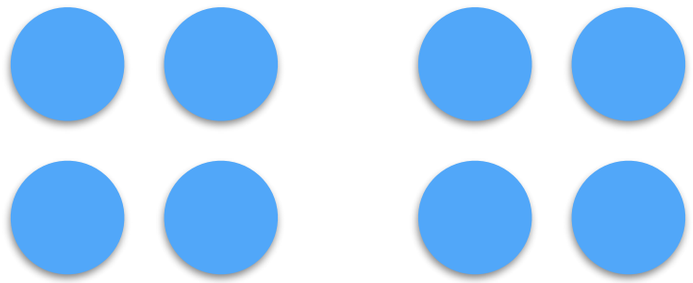


thanks to Martin Krzywinski

A little experiment



Task: How many dots?



thanks to Martin Krzywinski

brief history

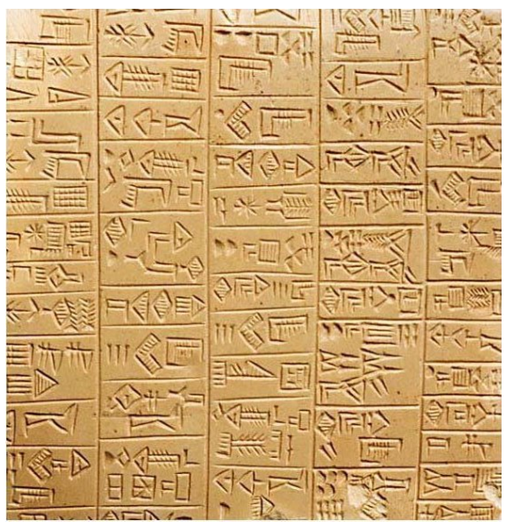
(western view)

Chauvet cave
proto-writing

~20,000 years ago

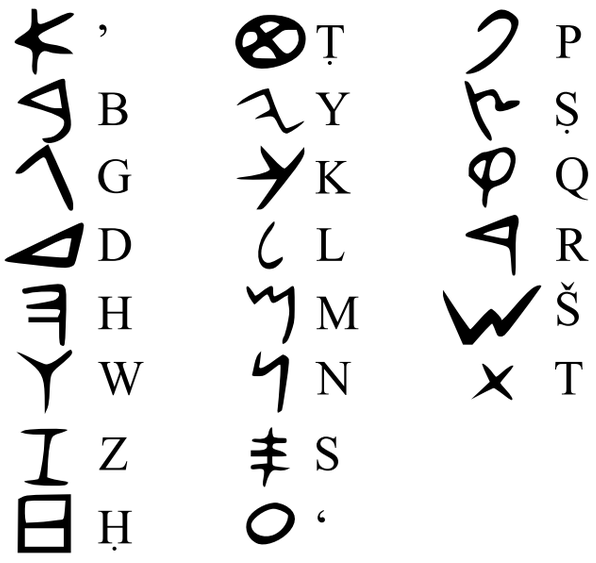
Sumerian cuneiform
logographic

~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?

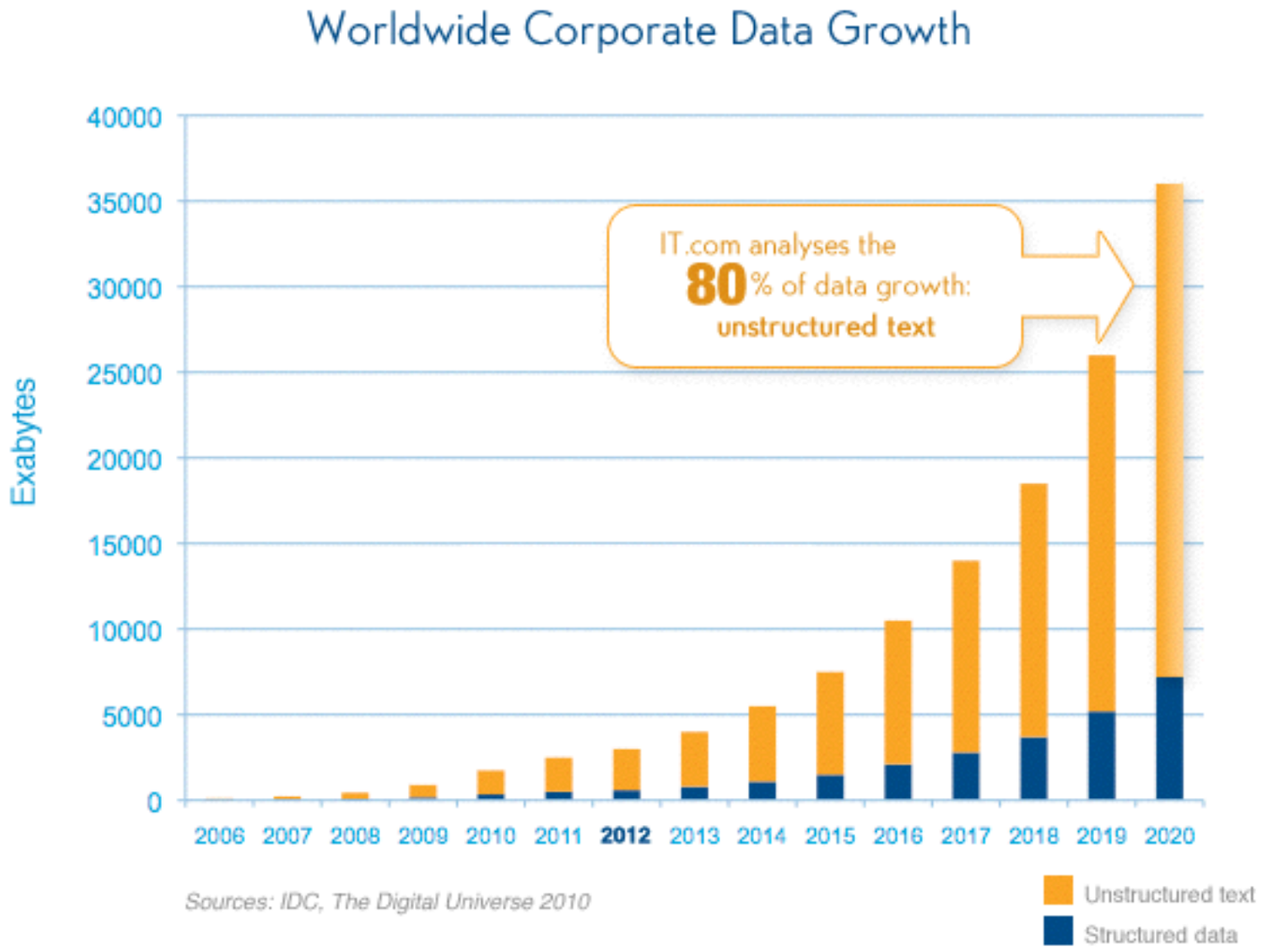
Why Text Vis?

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 cover- ing 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 in- ter 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 men- tion 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 psycho- logical backgrounds (Section 1.1.2), and describe landmarks in text visualiza- tion (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 furth
Early humans started rep
cave [CHQ+06] date at
+12]). These paintings r
image and text rep- resei
into logographic form. V
(semantics) within a lang
system included 24 signs
writing on papyrus vs. w
hieroglyphs to an alphab
ago. Phoenicians have b
the first known only-map
their ordered set of lette
In Europe, Romans beca
century) and the medieva
developed during the 8th
The impact on page styl
decoration. The indus- t
invented. The successors
computers with word-pr
and document distributi



oth references are

. The paintings from Chauvet
covered" (Sadier et al. [SDB
story. Divergence between
rm evolved from pictographic
est units of meaning
e- mic elements. Their sign
stances, like the ease of
of development from
iest developed 3,000 years
anean cultures. Their abjad is
cessively, the Greek named

he times of Charlemagne (8th
hile printing was already
allowed fast reproduction.
ls or Schnörkel remained as
lating machines were
ontent creation. Personal
s of document production

1.1.2 The psychological approach

We already discovered that text is nowadays as rapidly produceable and dis- tributable as never before, but we did not throw light on how humans "consume" text. Schönpflug & Schönpflug [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 saccade1 takes on average 20 to 35 ms to bridge a range of 7 to 9

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 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

ß

fi → fi

fl → fl

Text as Vis

- Creating a font type is an art which requires profound design knowledge

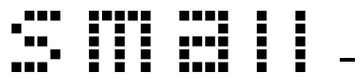
- .. or it can be a science:

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:  -> 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

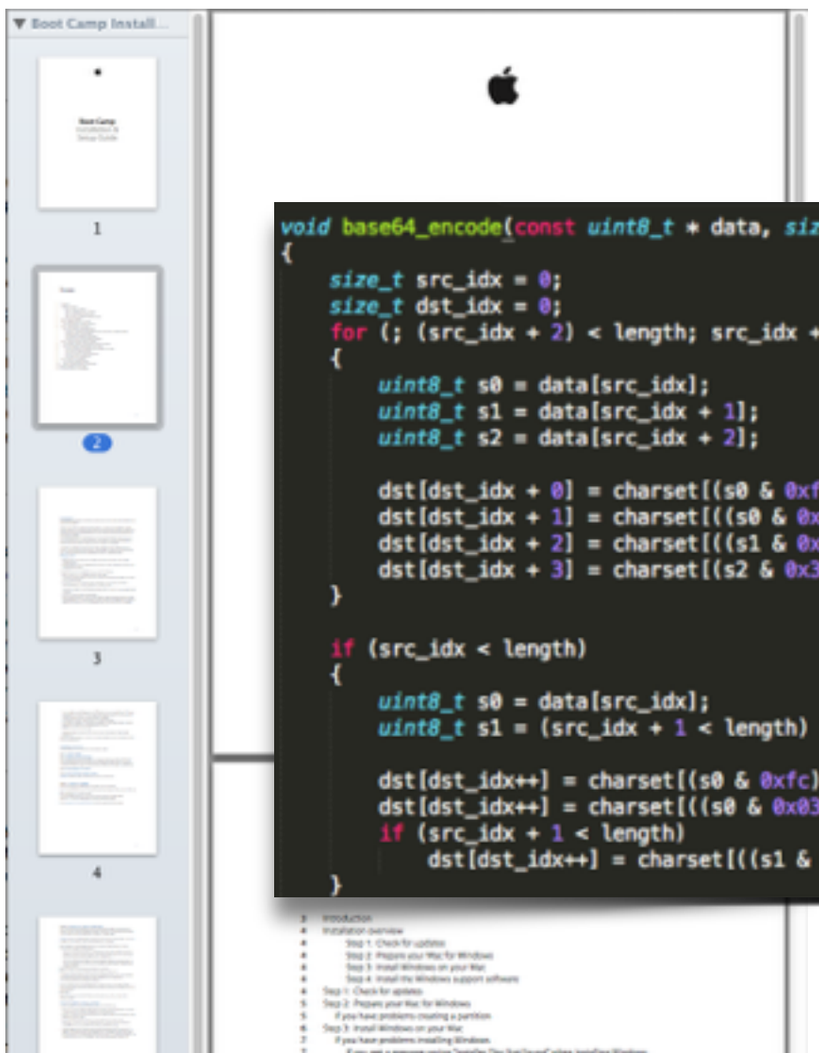
- rules of thumb:
  - limit the use of fonts to only a few typefaces !!
  - use “special” fonts only when appropriate
  - a good resource for fonts in web projects are google fonts

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)];
    }
```

Visualization for “Raw” Text

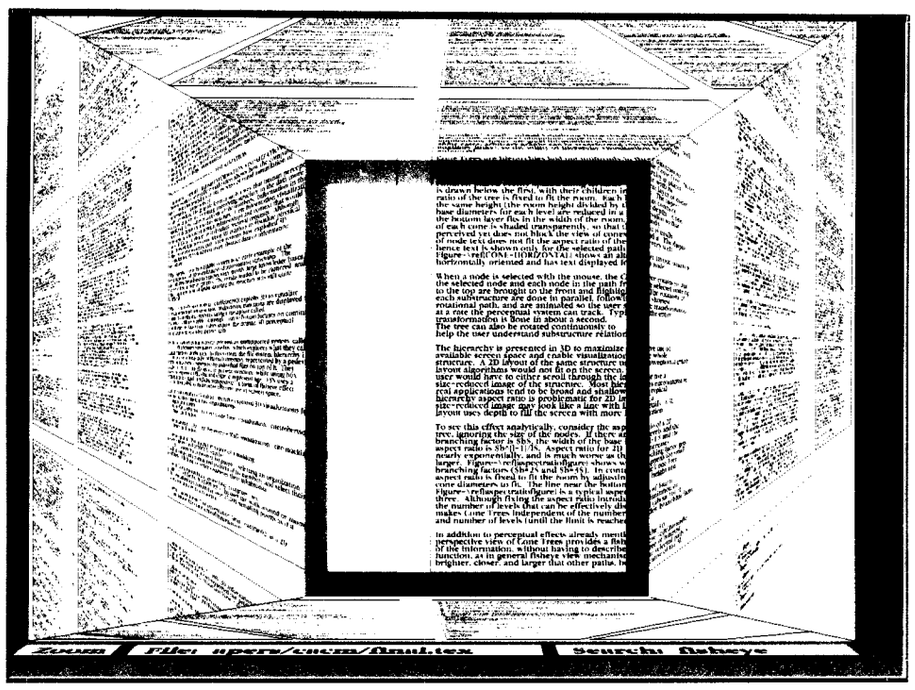


Figure 3: Document Lens with lens pulled toward the user. The resulting truncated pyramid makes text near the lens' edges readable.

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. Strobelt 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 viewer 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 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.

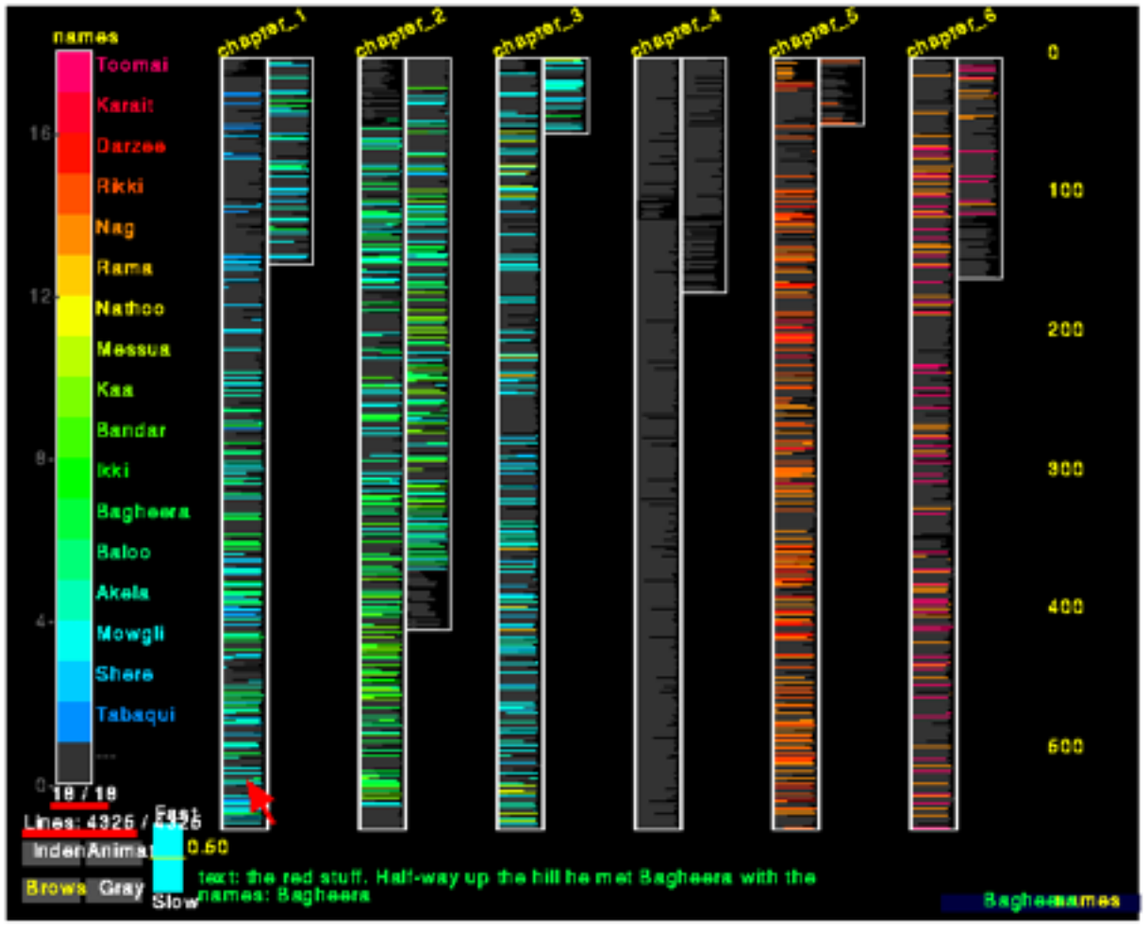
A common pixel based technique is TileBars [Hea95], which visualizes the length of documents and the distribution of search terms within these documents with a rectangular pixel-based visualization. Byrd [Byr99] combines the scrollbar of the document view with a pixel visualization of allowing the user to scroll rence of the terms. Both techniques do not show the context of the search terms and a user has to open the detail view in 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 highlgiht search terms by rendering them in a readable size with a semi-transparently colored background above of the original thumbnail. Woodruff et al. [WRM*02] pre-

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

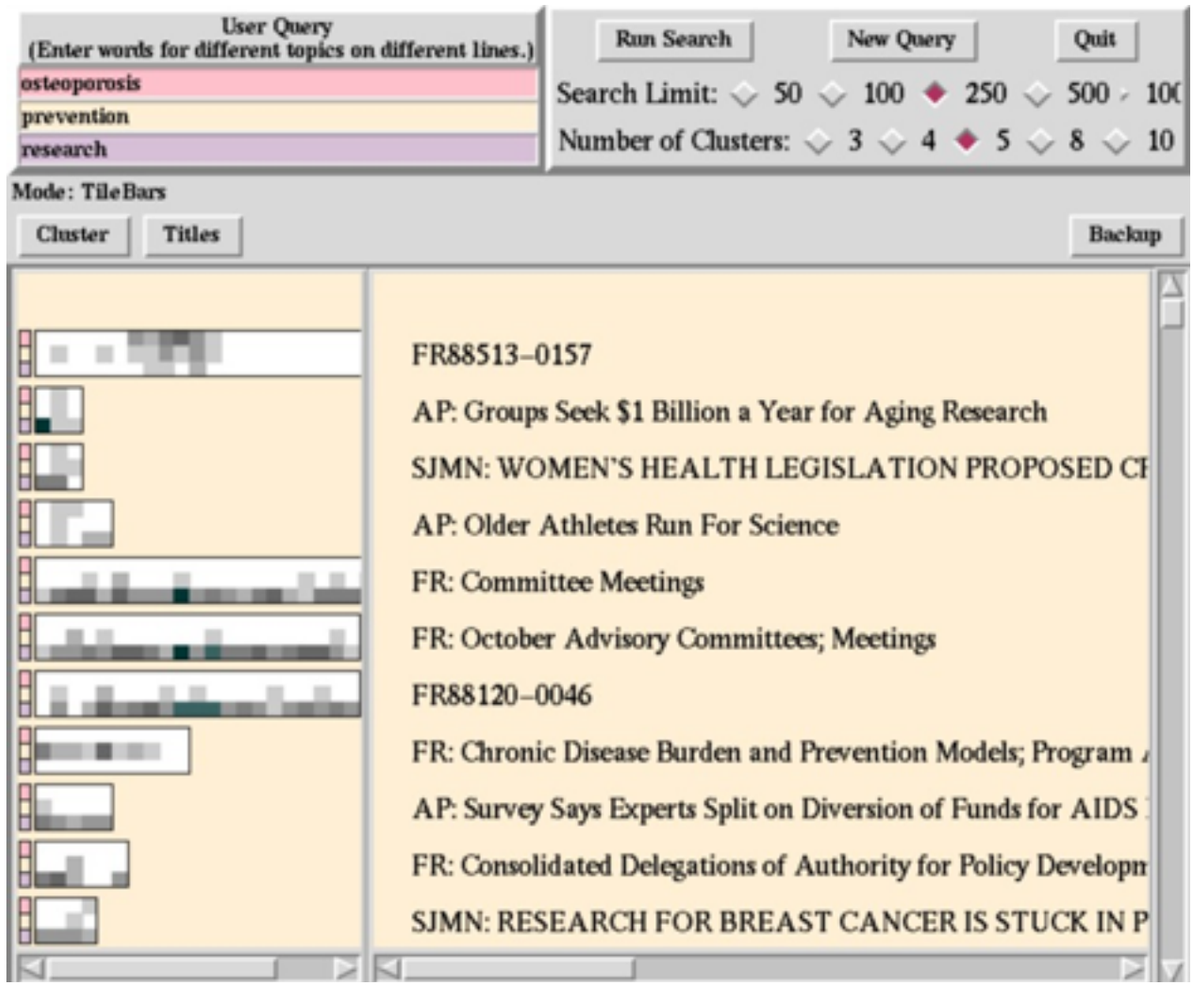
Document Thumbnails with Variable Text Scaling
A. Stoffel, H. Strobelt, O. Deussen, D. A. Keim
Computer Graphics Forum, volume 31 issue 3 pp.

Visualization for “Raw” Text



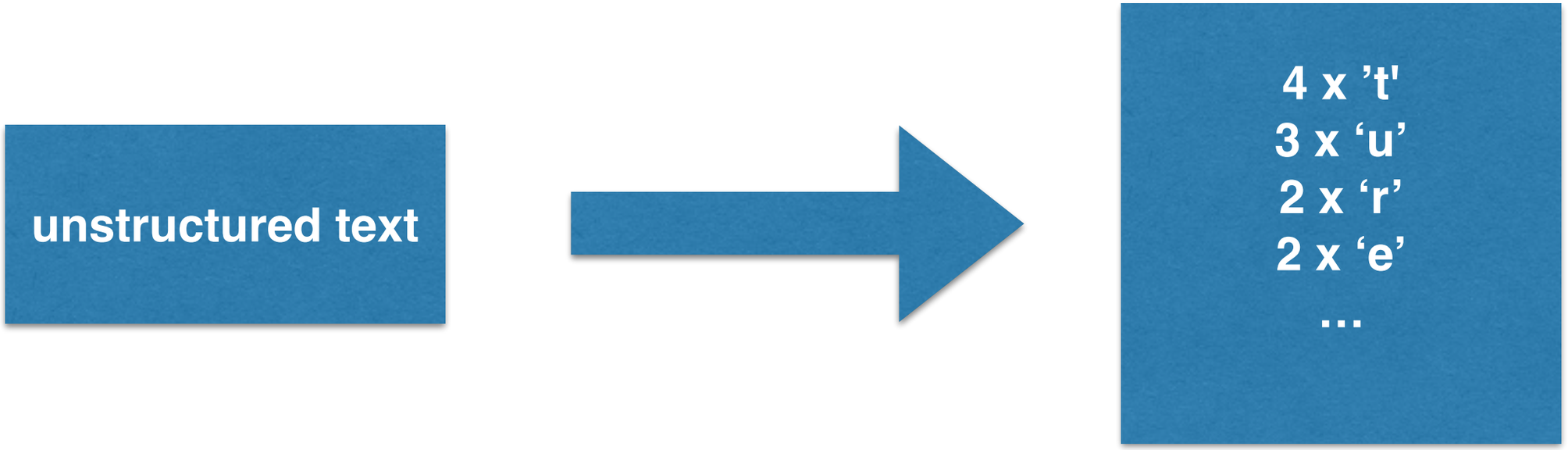
SeeSoft

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,..



Structured Text Features

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

	princess	dragon	castle
doc1	1	1	1
doc2	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. Yanukovych 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?*

http://en.wikipedia.org/wiki/List_of_linguistic_example_sentences

Was that irony? - Nooo

Profanity sucks. (14)

Be more or less specific. (15)

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

excerpt from Rules of Writing

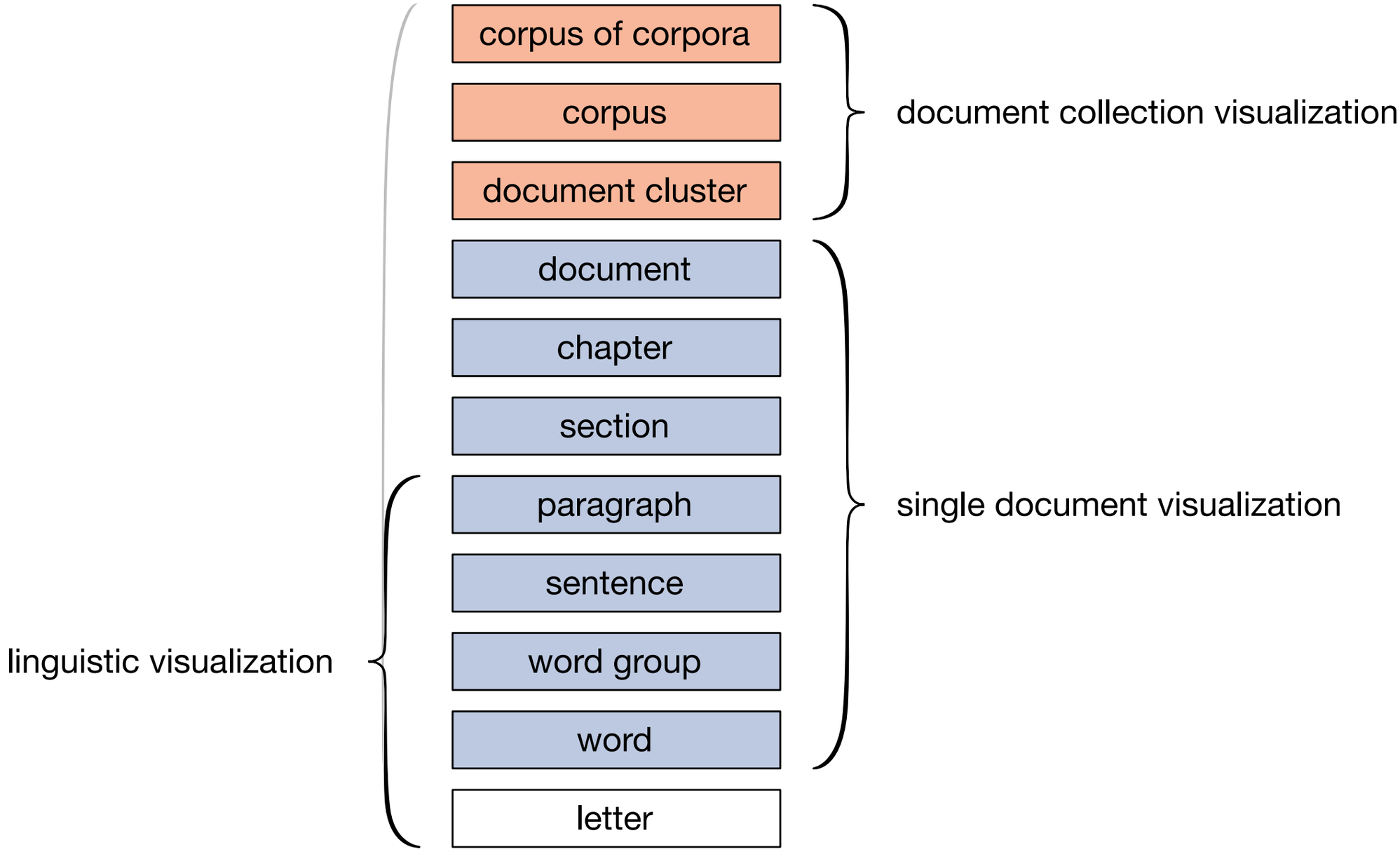
by Frank L. Visco (June 1986 in Writers' digest)

Thinking about..

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

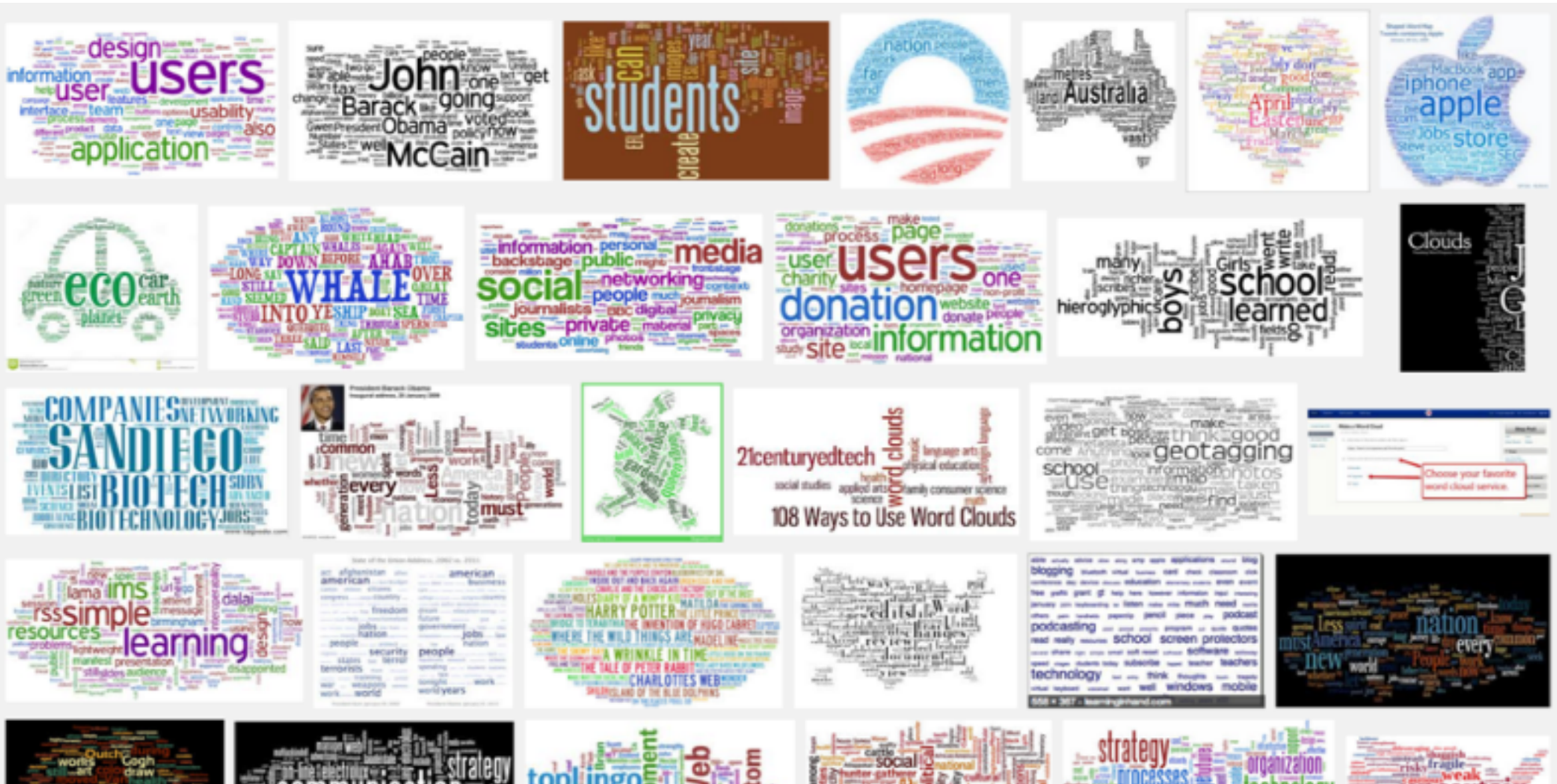
	princess	dragon	castle
doc1	1	1	1
doc2	0	0	1

Text Units Hierarchy

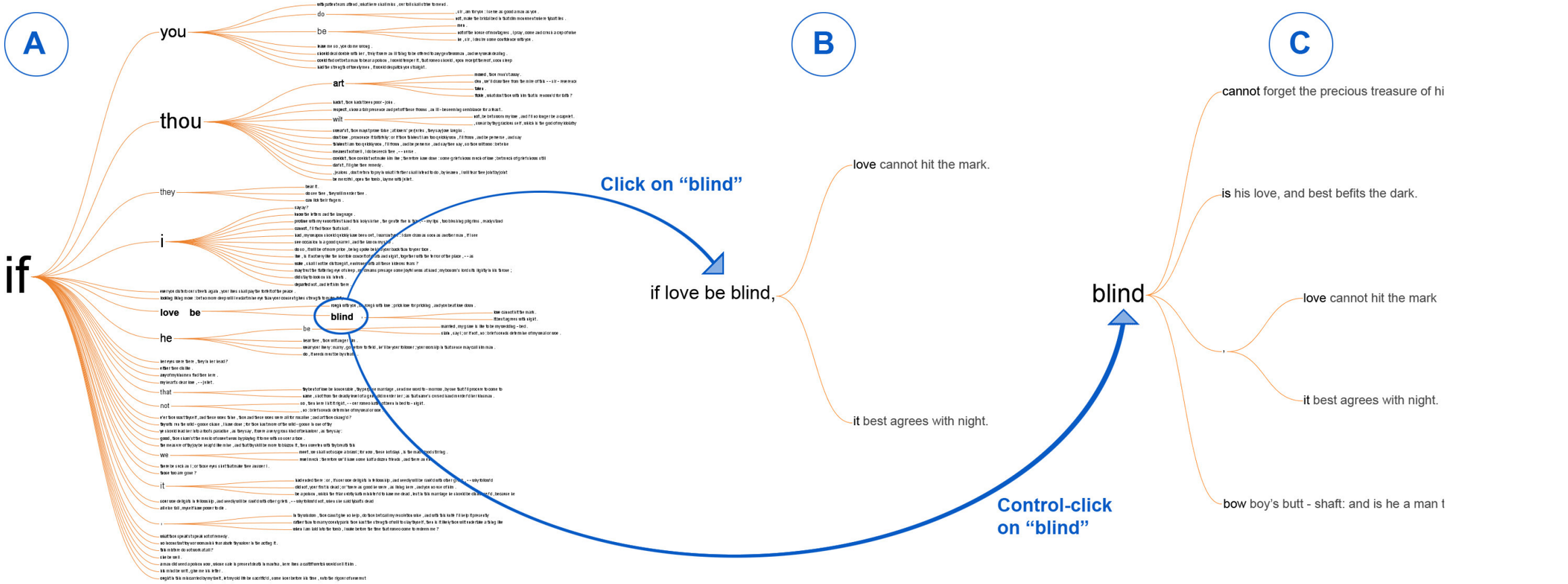


Vis for Text Documents

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



Vis for Text Documents



The word tree, an interactive visual concordance

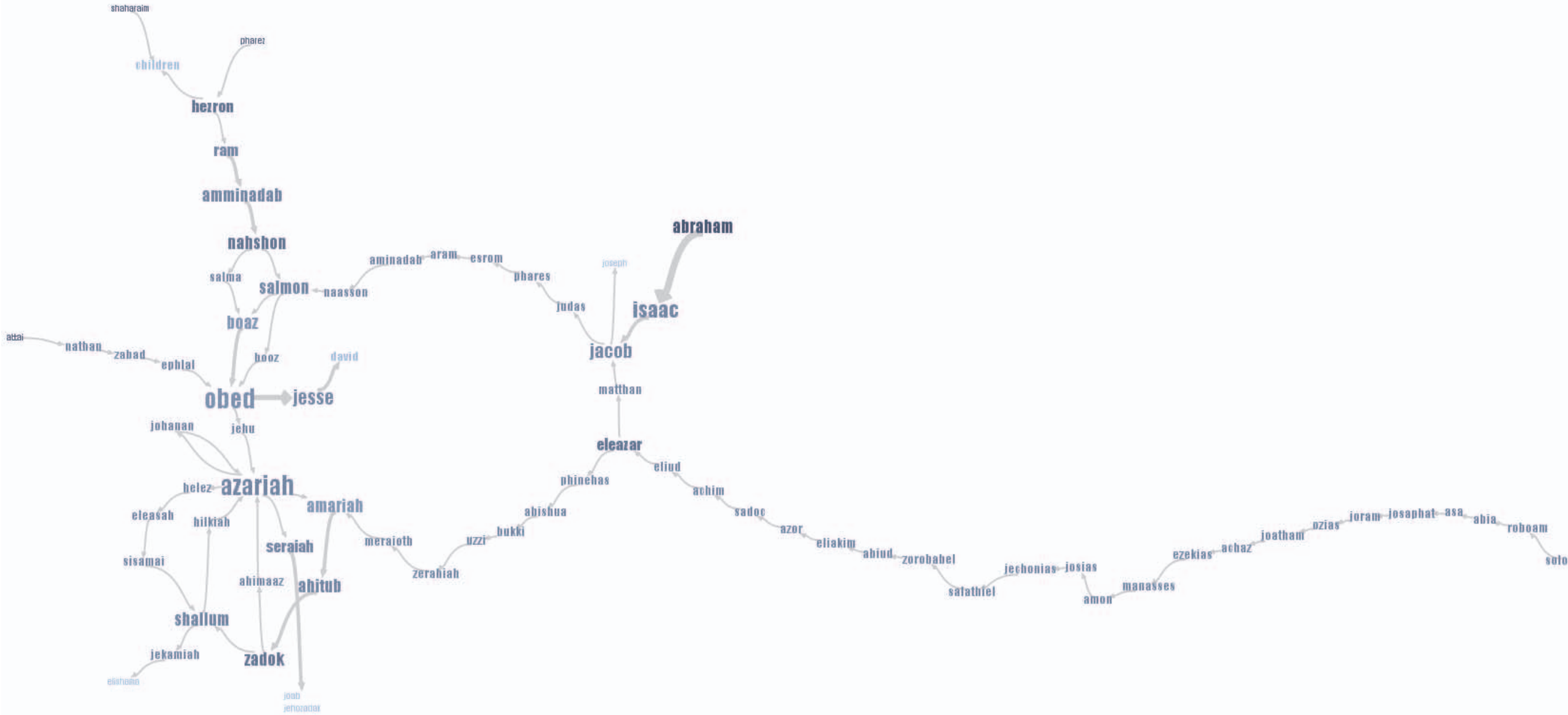
M Wattenberg, FB Viégas

Visualization and Computer Graphics, IEEE Transactions on 14 (6), 1221-1228

http://www.bobdylan.com/us/songs/blowin-wind

Vis for Text Documents

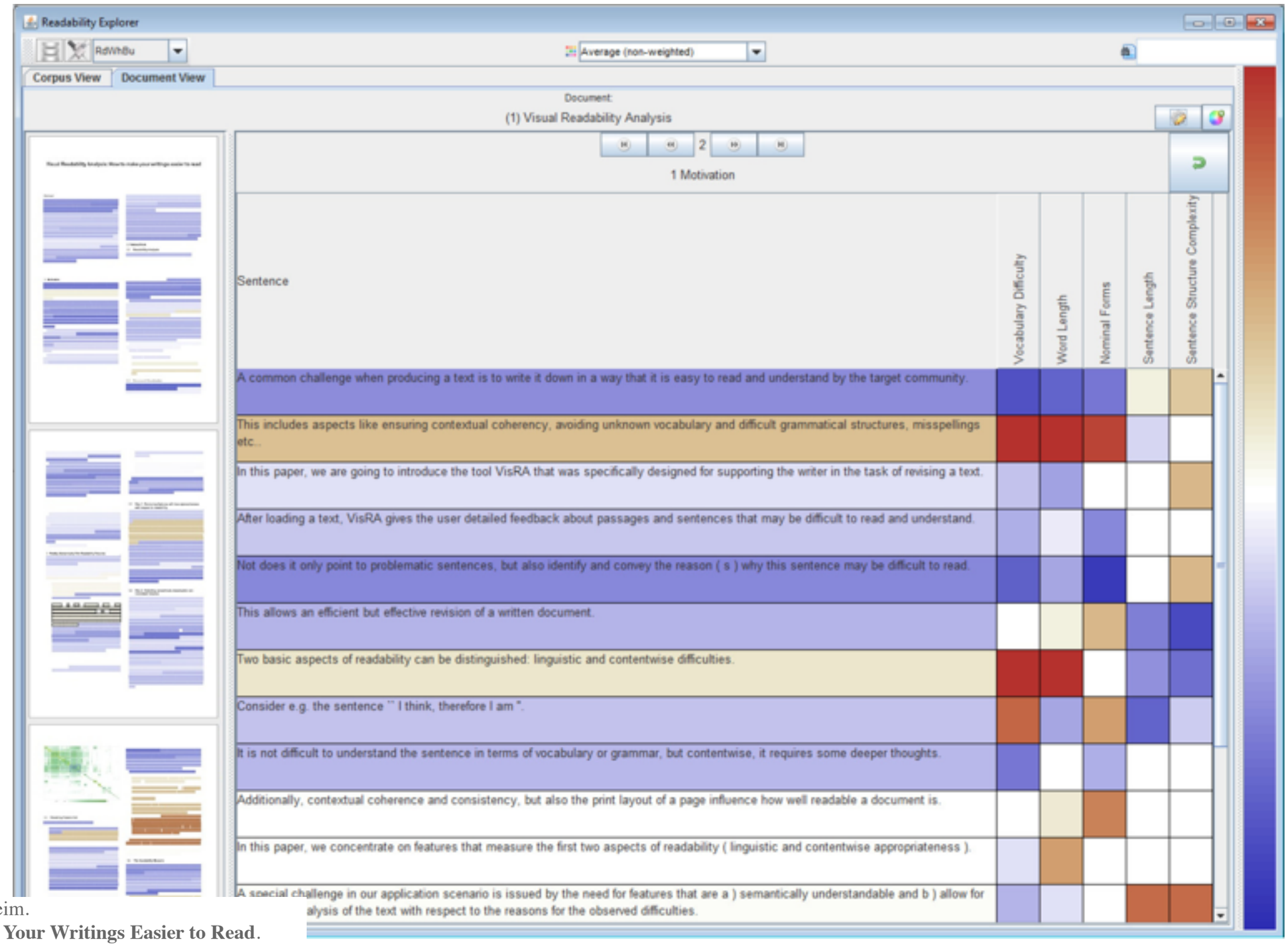
Frank van Ham, Martin Wattenberg, and Fernanda B. Viegas.
Mapping Text with Phrase Nets.
IEEE Transactions on Visualization and Computer Graphics 15, 6 (November 2009)


shaharaim
pharez
children
hezron
ram
amminadab
nahshon
salma
salmon
naasson
aminadab
aram
esrom
phares
judas
joseph
abraham
isaac
jacob
matthan
eleazar
eliud
achim
sadoc
azor
eliakim
abiud
zorobabel
salathiel
jechonias
josias
amon
manasses
ezekias
achaz
joatham
ozias
joram
josaphat
asa
abia
roboam
phinehas
abishua
bukki
uzzi
meraioth
zerahiah
amariah
azariah
helez
eleasah
hilkiah
seraiah
sisamai
ahimaaz
ahitub
shallum
jekamiah
zadok
elishama
joab
jehozadak
attai
nathan
zabad
ephlal
obed
jesse
booz
david
johanan
jehu

Vis for Text Documents

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

Vis for Language Analysis



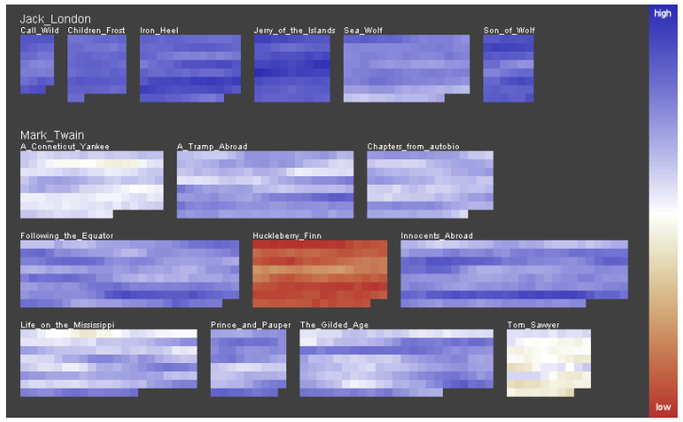
D. Oelke, D. Spretke, A. Stoffel and D. A. Keim.
Visual Readability Analysis: How to Make Your Writings Easier to Read.
IEEE Transactions on Visualization and Computer Graphics, 18(5):662-674, 2012.

Vis for Language Analysis

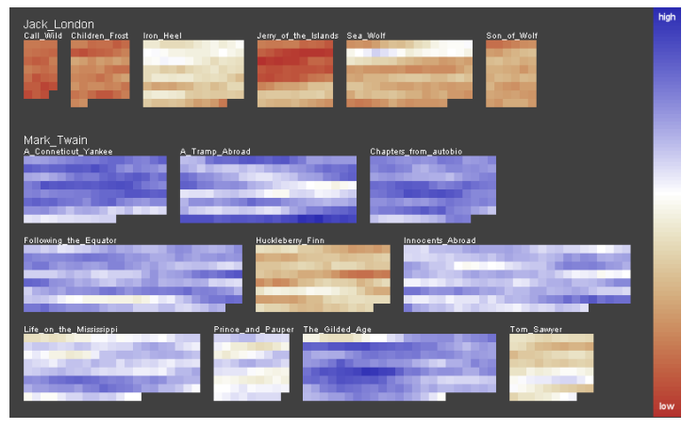
- Literature fingerprints:

“ 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.”

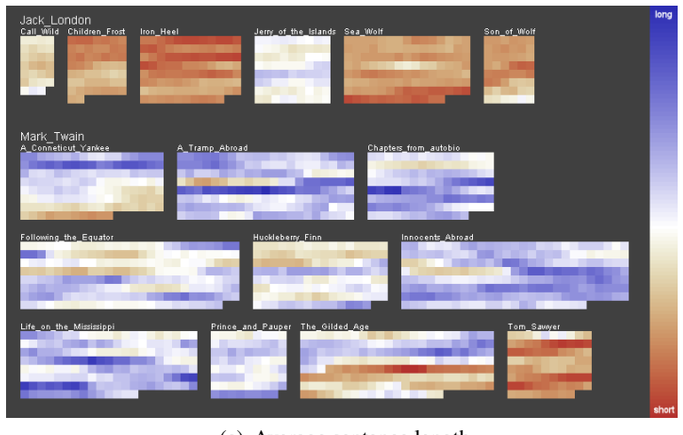
*method to measure the vocabulary richness



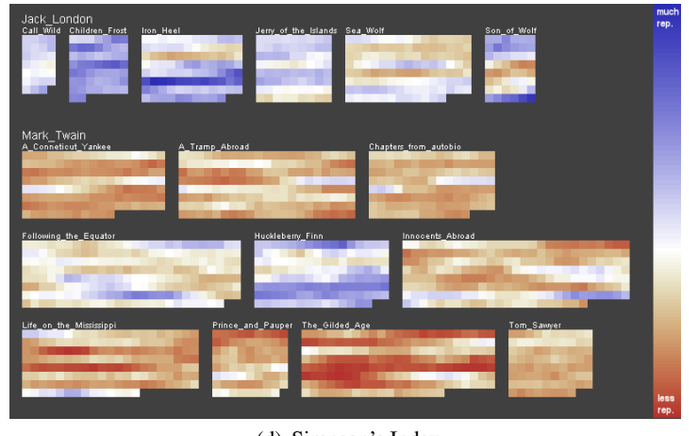
(a) Function words (First Dimension after PCA)



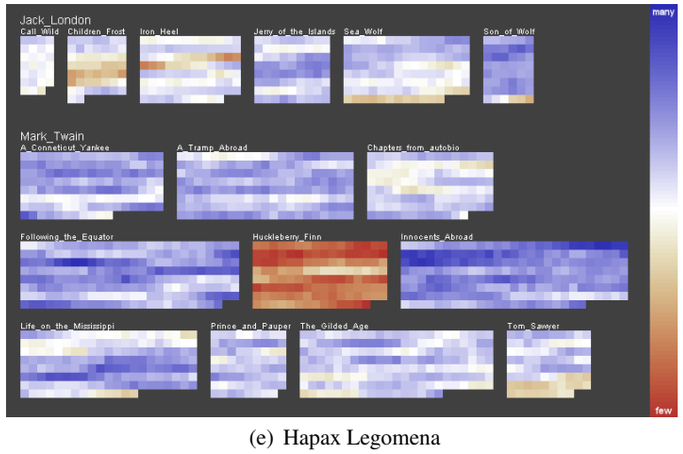
(b) Function words (Second Dimension after PCA)



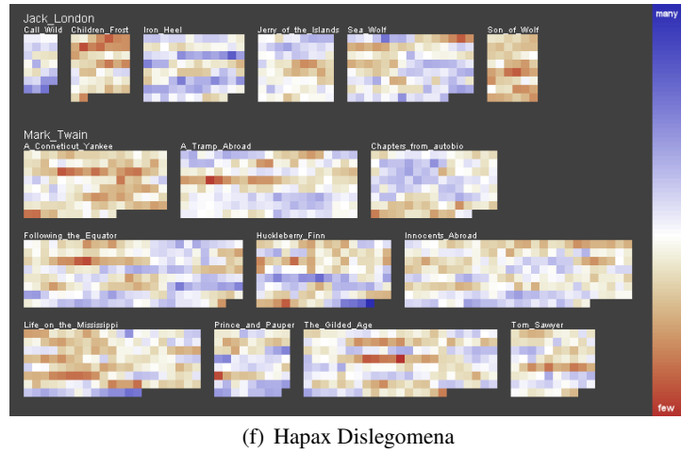
(c) Average sentence length



(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-word to project documents w.r.t. text similarity into a landscape

- (only) one example

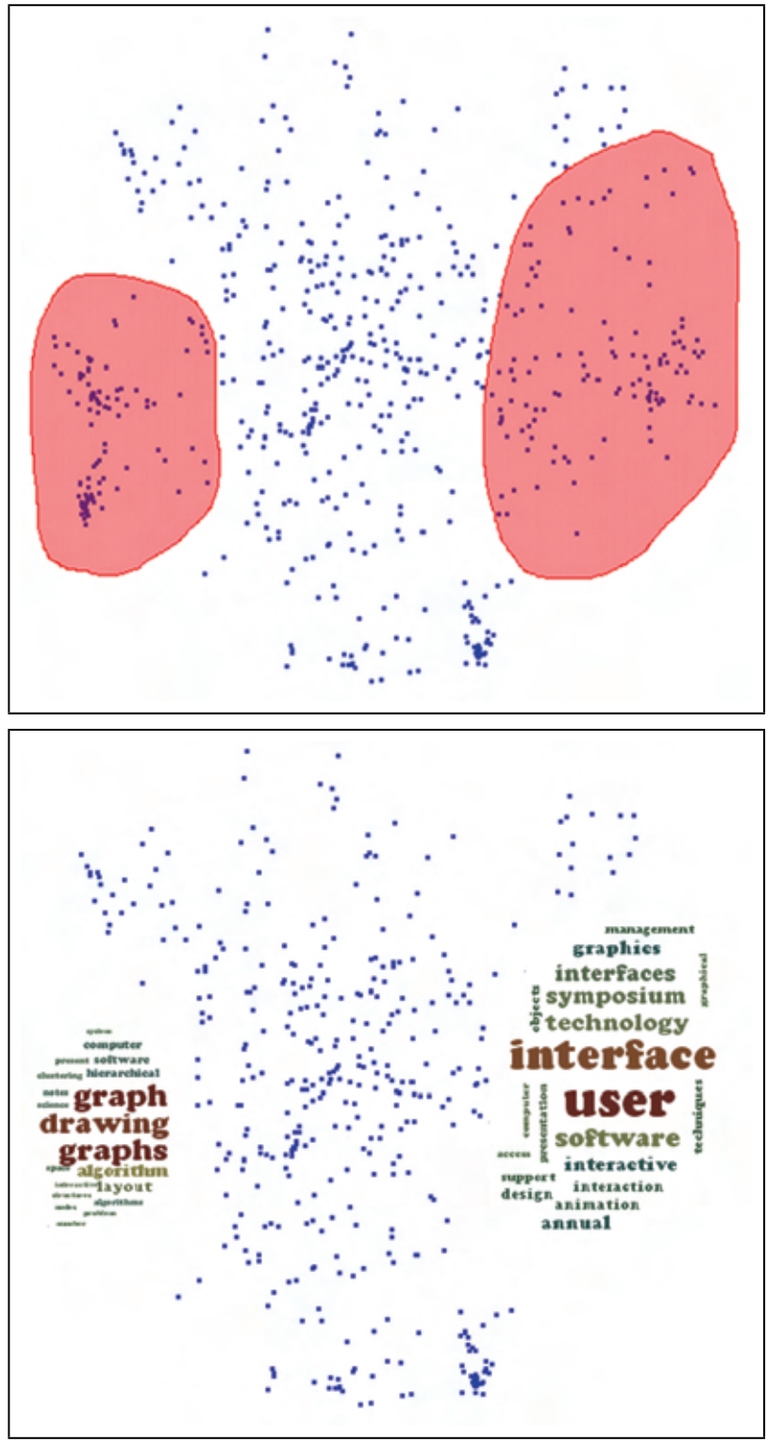
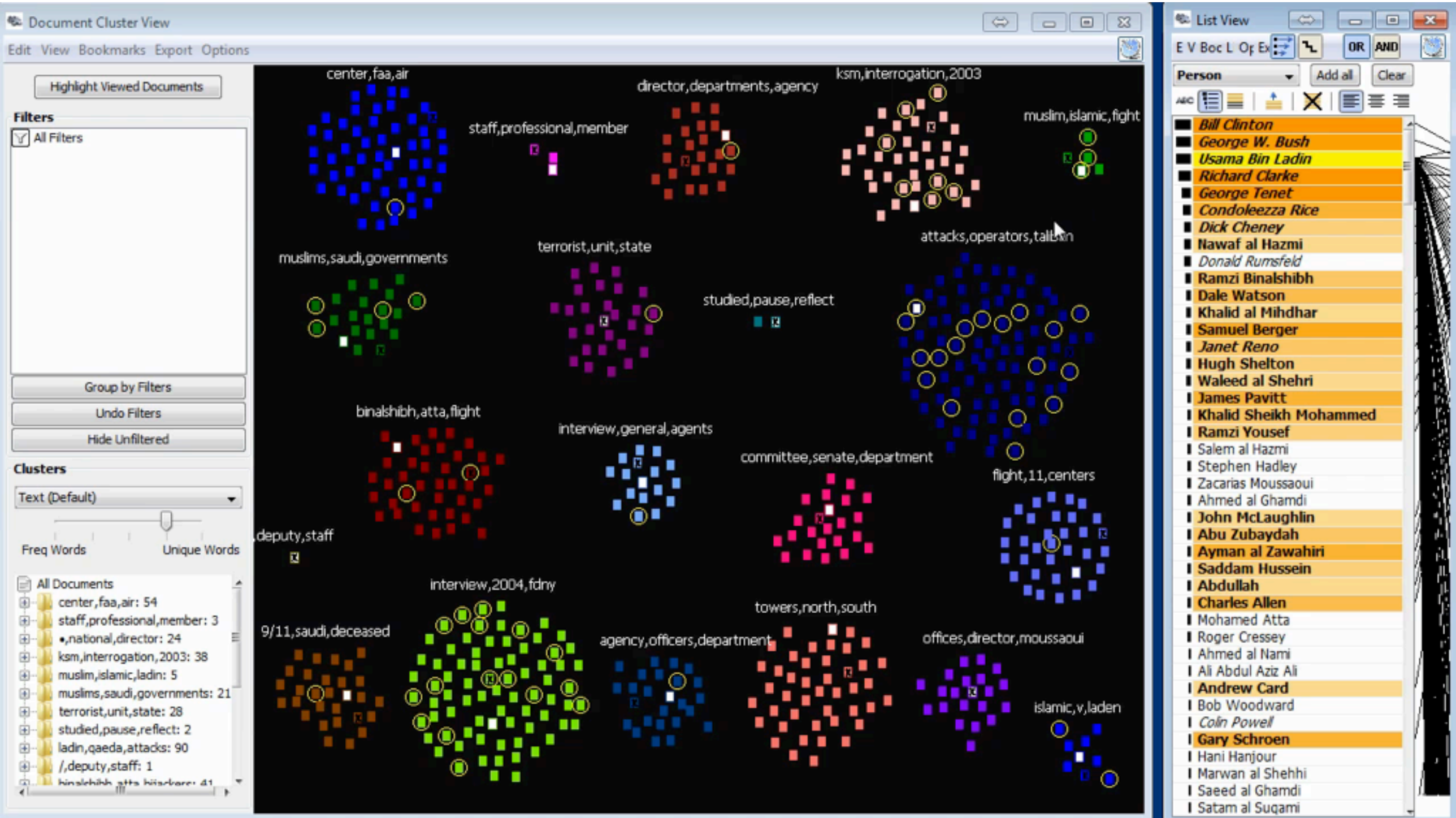


Figure 5: *A user can interactively draw a region (polygon) containing a subset of documents of interest (top figure). Keywords are extracted from the selected document and their corresponding word could is built inside the user-defined region (bottom figure).*

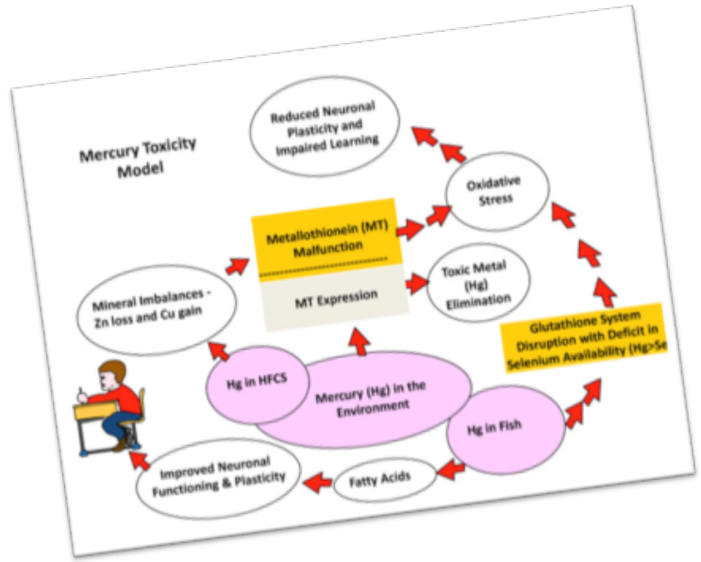
Fernando V. Paulovich, Franklina M. B. Toledo, Guilherme P. Telles, Rosane Minghim, and Luis Gustavo Nonato.
Semantic Wordification of Document Collections.
Comp. Graph. Forum 31, 3pt3 (June 2012)

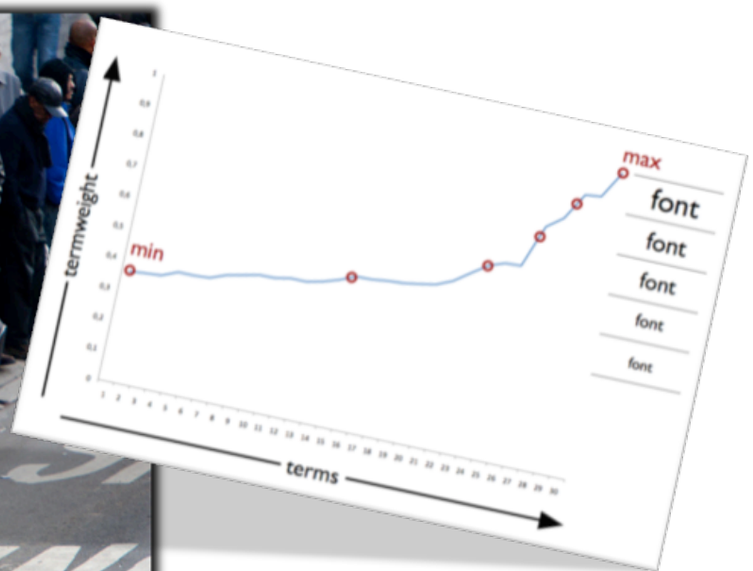
Visual Analytics for Large Text Corpora (example JigSaw)



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
  - Semanticons:

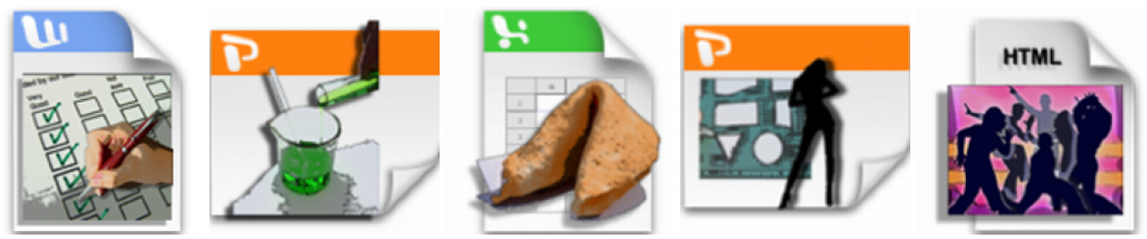
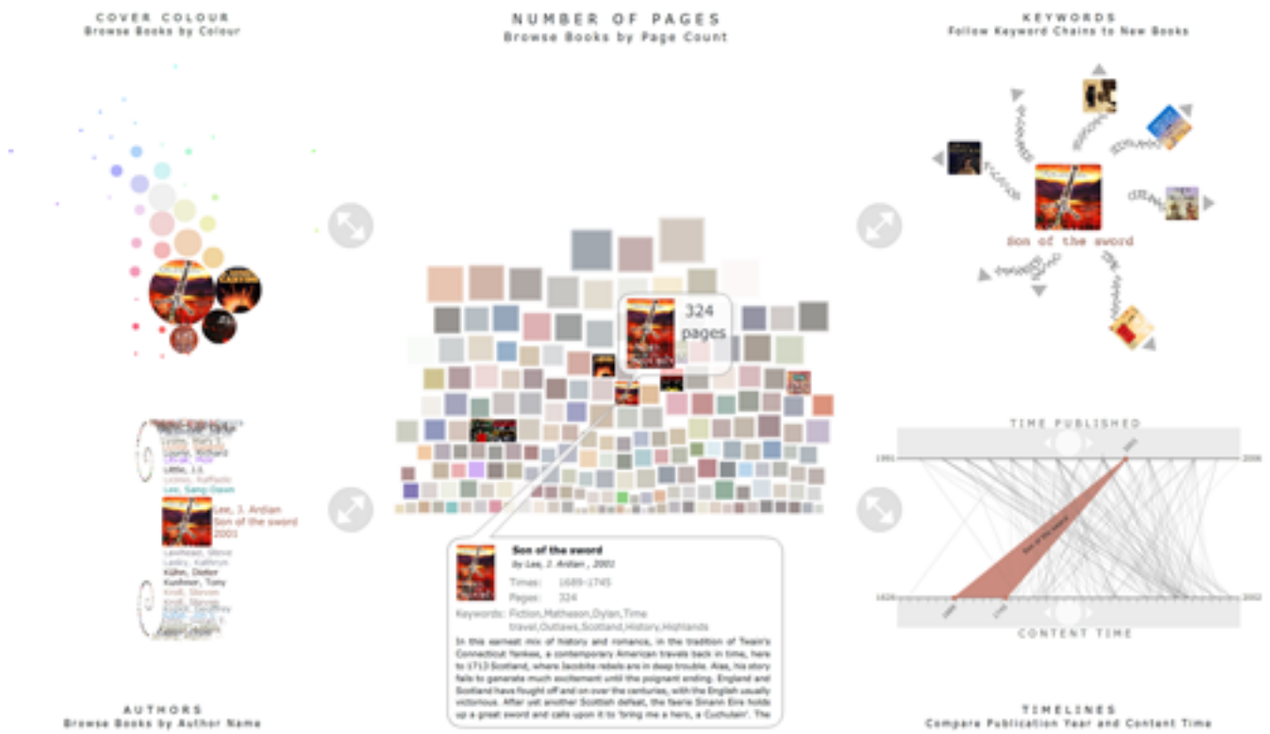


Figure 1: Semanticons *generated by our system for various filenames.*

Semanticons: 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

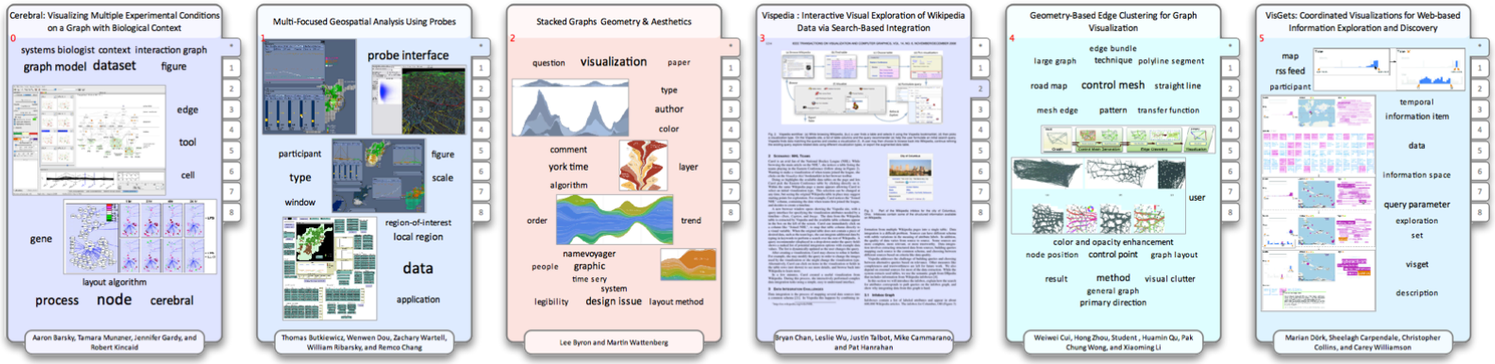
- summarize scientific documents using important terms and important figures

- 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

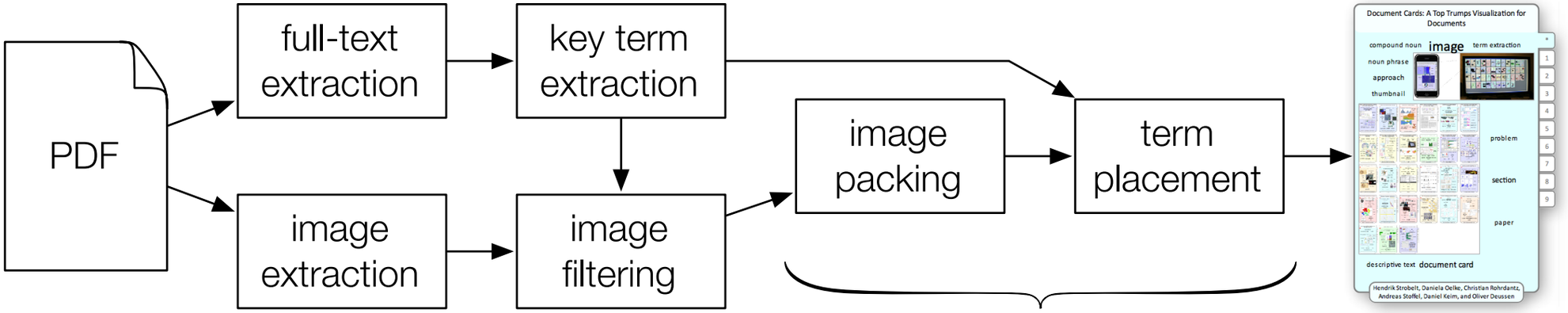
Document Cards: A Top Trumps Visualization for Documents
H. Strobelt, D. Oelke, C. Rohrdantz, A. Stoffel, O. Deussen, D. Keim
IEEE Transactions on Visualization and Computer Graphics (TVCG - InfoVis), 2009

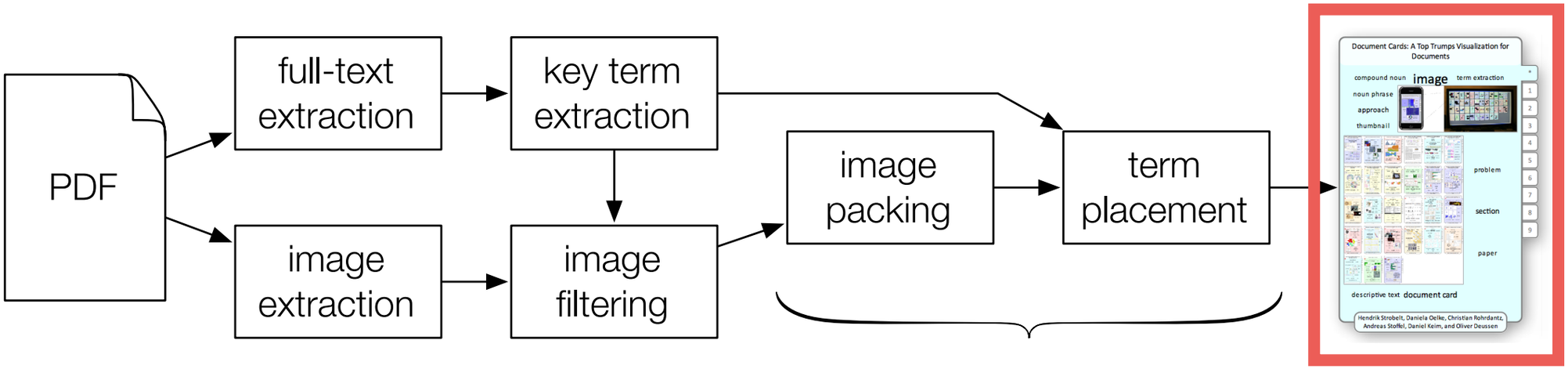
http://documentcards.hs8.de

DocumentCards


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 Gardy, and Robert Kincaid
Multi-Focused Geospatial Analysis Using Probes
probe interface
participant
type
window
figure
scale
region-of-interest
local region
data
application
Thomas Butkiewicz, Wenwen Dou, Zachary Wartell, William Ribarsky, and Remco Chang
Stacked Graphs Geometry & Aesthetics
question visualization paper
type
author
color
comment
york time
algorithm
layer
order
trend
namevoyager
people graphic
time sery
system
legibility design issue layout method
Lee Byron and Martin Wattenberg
Vispedia : Interactive Visual Exploration of Wikipedia Data via Search-Based Integration
Bryan Chan, Leslie Wu, Justin Talbot, Mike Cammarano, and Pat Hanrahan
Geometry-Based Edge Clustering for Graph Visualization
edge bundle
large graph technique polyline segment
road map control mesh straight line
mesh edge pattern transfer function
user
color and opacity enhancement
node position control point graph layout
result method visual clutter
general graph
primary direction
Weiwei Cui, Hong Zhou, Student , Huamin Qu, Pak Chung Wong, and Xiaoming Li
VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery
map
rss feed
participant
temporal
information item
data
information space
query parameter
exploration
set
visget
description
Marian Dörk, Sheelagh Carpendale, Christopher Collins, and Carey Williamson
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
visual query language
visualization
design
task
participant
poll data ring system
data point
Geoffrey M. Draper, and Richard F. Riesenfeld
Exploration of Networks Using Overview+Detail with Constraint-based Cooperative Layout
layout method
route
focal node
tool
user lod
edge rout
placement constraint
level
high quality layout
primary graph
large network
detailed view
part
system
uml class diagram
display
model
layout technique
cluster position structure
focus node
Tim Dwyer, Kim Marriott, Falk Schreiber, Peter J. Stuckey, Michael Woodward and Michael Wybrow
Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation
visual exploration
data dimension
query
prototype implementation
visual representation
cameras
digital camera dataset
tool
figure
navigation
overview
grand tour
range
system
scatterplot matrix
user
plane
operation method order
Niklas Elmqvist, Pierre Dragicevic, and Jean-Daniel Fekete
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, Jock D. Mackinlay, Chris Stolte, and Manecesh Agrawala
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, Jock D. Mackinlay, Chris Stolte, and Maneesh Agrawala
Improving the Readability of Clustered Social Networks using Node Duplication
representation success rate
social network time clonenode
central actor
analysis
clonelink
visualization
duplicate
test
community
noduplication
significant effect
cluster
duplication link
participant splitlink readability
Nathalie Henry, Anastasia Bezerianos, and Jean-Daniel Fekete
EMDialog: Bringing Information Visualization into the Museum
touch emily carr
interaction design
node
installation
visual appeal
statement
set
tree diagram
perspective
information visualization
cut section
museum visitor
tree ring
public space
data representation people museum context
Uta Hinrichs, Holly Schmidt, and Sheelagh Carpendale
On the Visualization of Social and other Scale-Free Networks
scale-free network
simplification method
shortest path
node degree
visualization
layout
time
distance matrix
weighted graph
geodesic cluster
gprime
original graph
edge filter
edge metric power-law graph hub node
Yuntao Jia, Jared Hoberock, Michael Garland, and John C. Hart
A Framework of Interaction Costs in Information Visualization
visualization
user intent study focus-lock mode
paper interaction cost participant
item
time interface operation
menu
focus lock framework evaluation
section interaction target
overview window subject gulf
Heidi Lam
Distributed Cognition as a Theoretical Framework for Information Visualization
cognitive property process endangered specy
infovis system cognition insight
internal representation
Rule1 Rule2 Rule3
Int Int Int
Int Int Ext
Int Ext Ext
developer cognitive system
report theory artifact
cognitive science analyst document view
Zhicheng Liu, Nancy J. Nersessian, and John T. Stasko
Particle-Based Labeling: Fast Point-Feature Labeling without Obscuring Other Visual Features
section virtual particle method
problem
distant labele
conflict particle
figure
labeling step
solution
point-feature
visual element
number
table
time result collision map
Martin Luboschik, Heidrun Schumann and Hilko Cords
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 Muelder and Kwan-Liu Ma, Senior
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
Effectiveness of Animation in Trend Visualization
group trace line
visualization
data point
user study task
presentation
participant
animation and trace
George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, and John Stasko
Viz-A-Vis: Toward Visualizing Video through Computer Vision
condition
pattern recognition
computer vision
aggregate
activity table
activity cube
raw data
video
system
heat map
frame
viz-a-vis
level
time
model
view
motion
aggregation
abstraction step figure
difference
Mario Romero, Jay Summet, John Stasko, and Gregory Abowd
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
method
layout
focus line
balloon focus original treemap task
Ying Tu and Han-Wei Shen
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
The Word Tree, an Interactive Visual Concordance
tree structure
common word user time
option data design
tag cloud
search term
romeo and juliet
title
love
token
word tree
eye
context number
suffix tree branch blind
king james bible
Martin Wattenberg and Fernanda B. Viégas
Evaluating the Use of Data Transformation for Information Visualization
comment performance distribution task
system step error rate
non-transformed data
choice technique
study
experiment
house
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
Spatially Ordered Treemaps
readability score
node order data geographic location
angular change color space
position
sequence distortion
scene type unit size
problem
histomap layout
treemap node
spatial layout
hill
beach
image
layout algorithm
relationship consistency arrangement
displacement vector
Jo Wood, and Jason Dykes
The Shaping of Information by Visual Metaphors
information
visual metaphor
verbal metaphor
type
response time user
understand
incompatible question
study
correct response
participant
task question
session
difference
visual representation
Caroline Ziemkiewicz and Robert Kosara

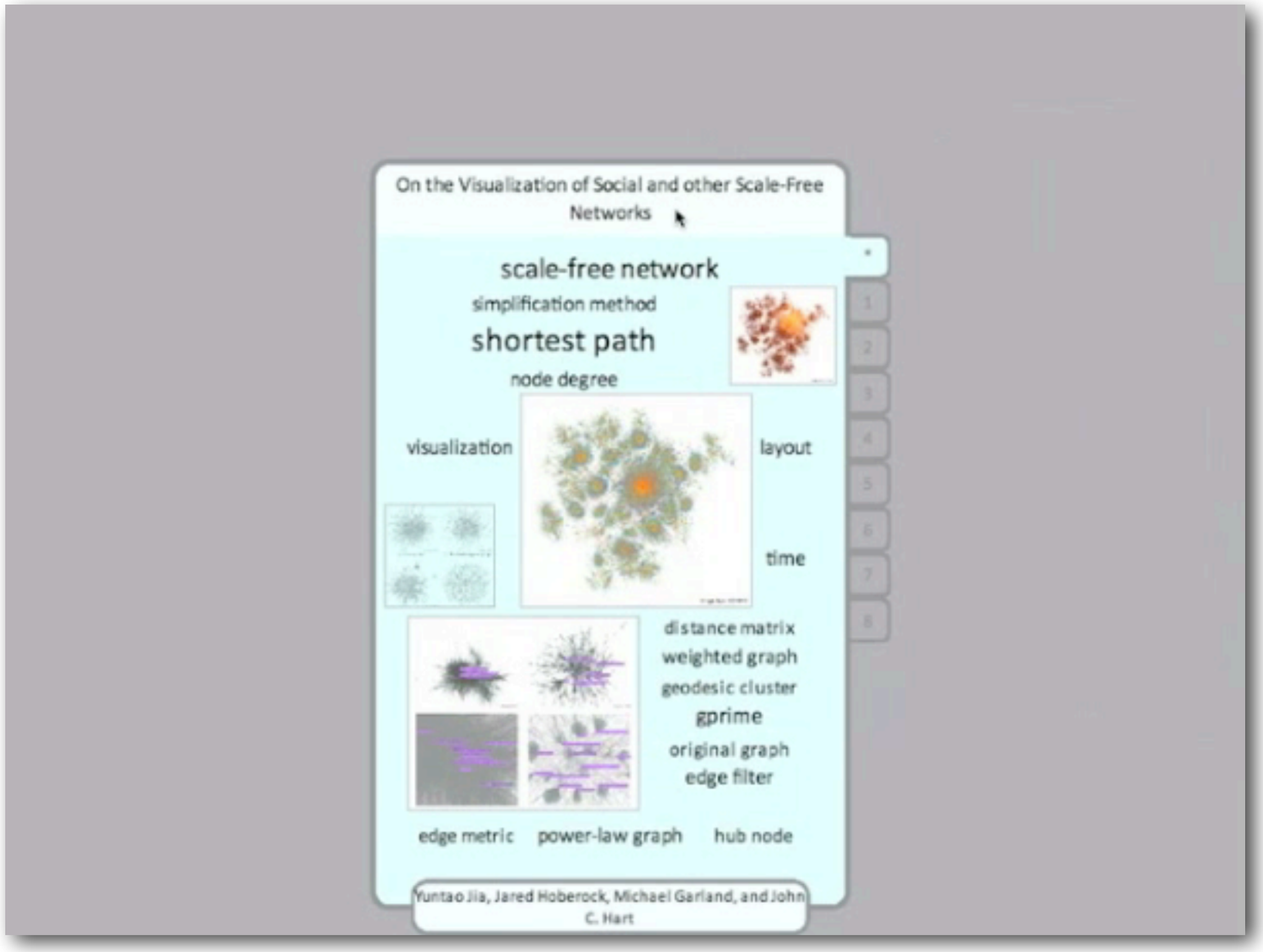
DC - pipeline


PDF
full-text extraction
key term extraction
image extraction
image filtering
image packing
term placement
Document Cards: A Top Trumps Visualization for Documents
compound noun
image
term extraction
noun phrase
approach
thumbnail
problem
section
paper
descriptive text
document card
Hendrik Strobelt, Daniela Oelke, Christian Rohrdantz, Andreas Stoffel, Daniel Keim, and Oliver Deussen



Interaction:

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



DiTop

corpus of corpora
corpus
document cluster
document
chapter
section
paragraph
sentence
word group
word
letter

corpus of corpora
corpus
document cluster
document
chapter
section
paragraph
sentence
word group
word
letter

corpus of corpora
corpus
document cluster
document
chapter
section
paragraph
sentence
word group
word
letter

time

Compare Corpora

- Compare topics between text collections

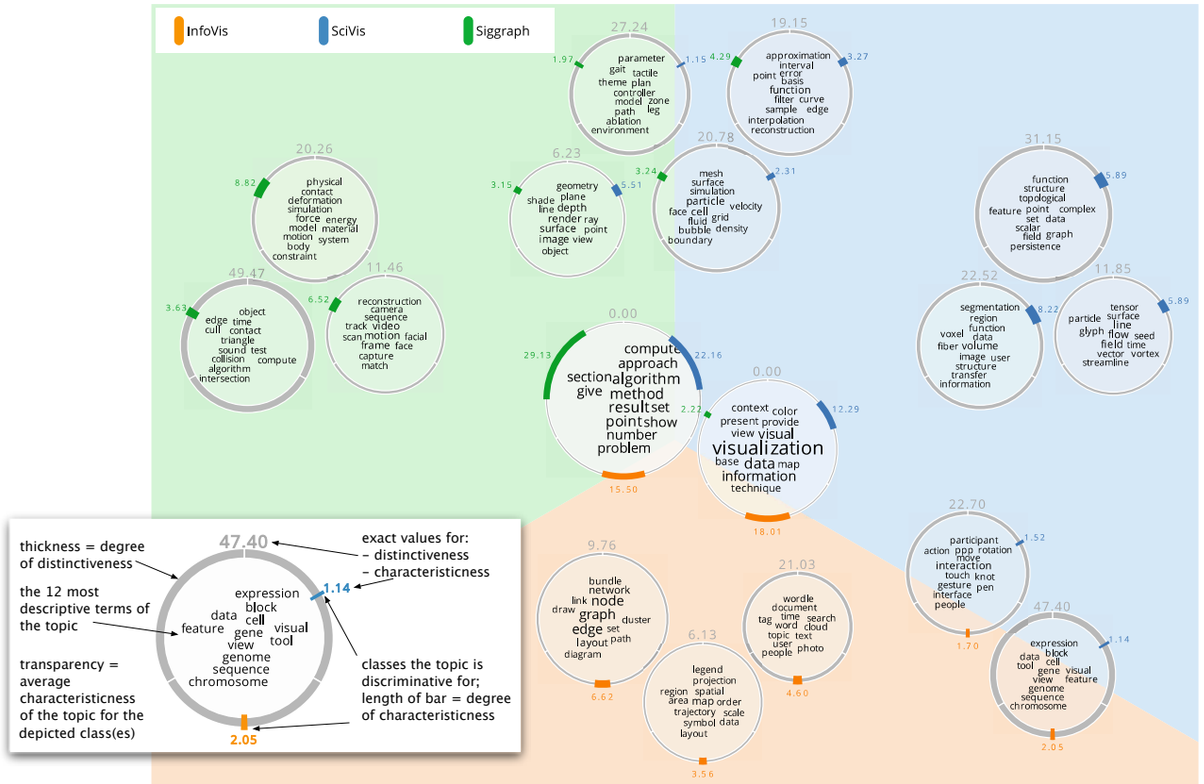
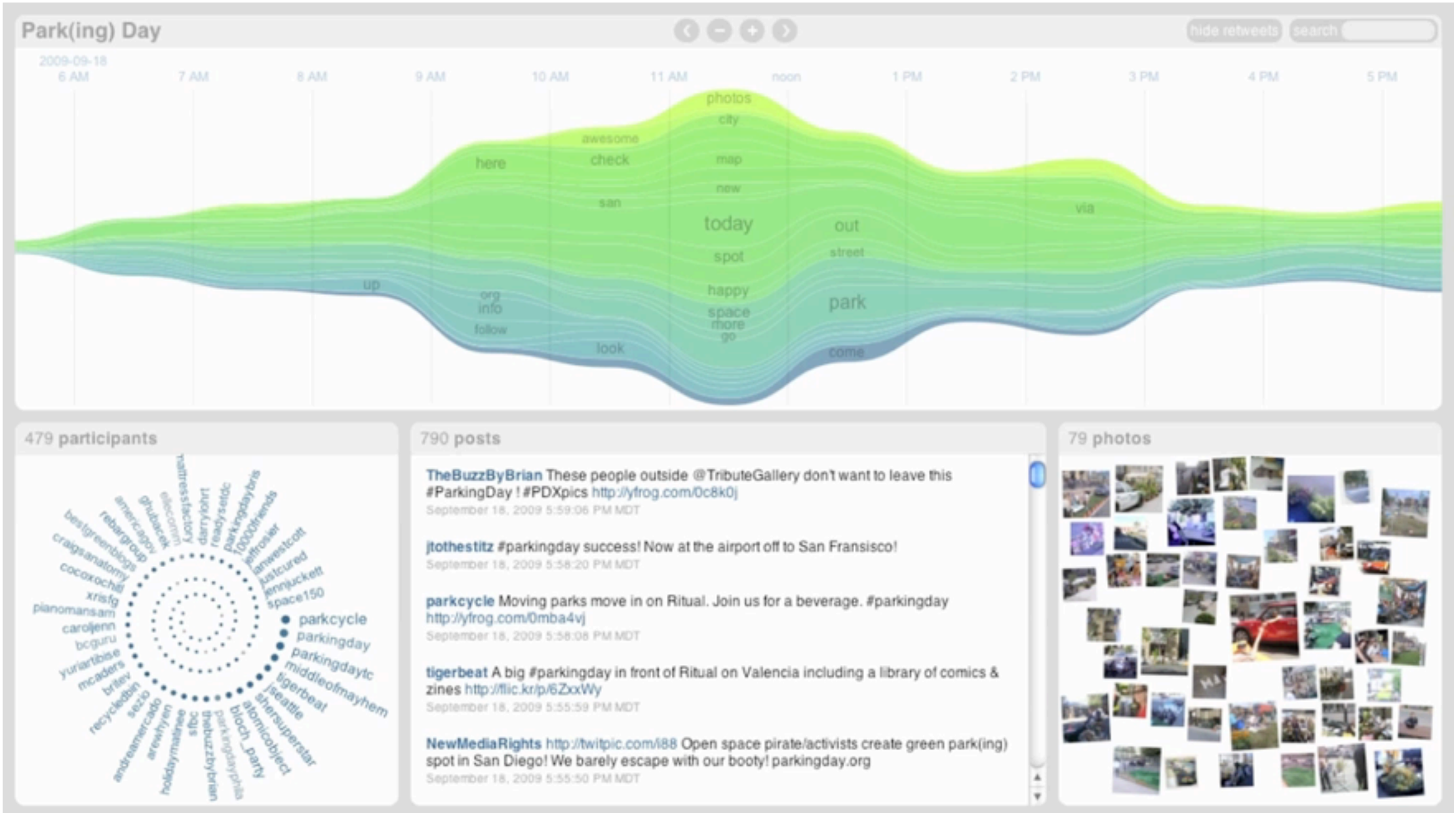


Figure 1: Comparison of 495 papers of InfoVis, SciVis, and Siggraph (discrimination threshold = 6, number of topics = 30)

Comparative Exploration of Document Collections: a Visual Analytics Approach (http://ditop.hs8.de)

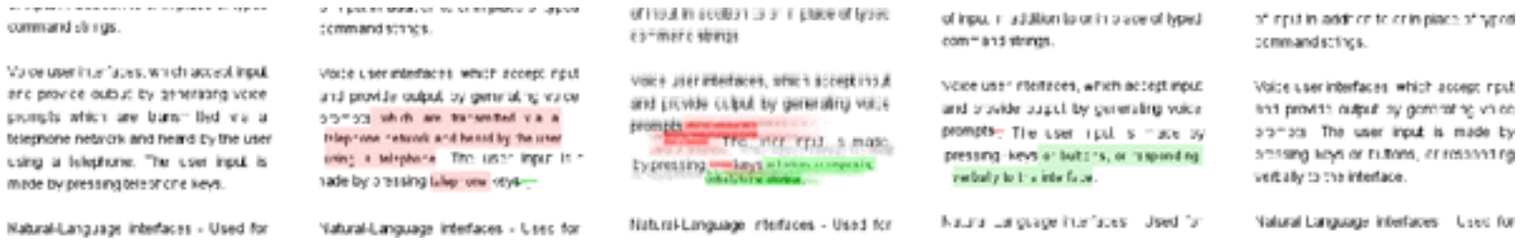
D. Oelke, H. Strobelt, C. Rohrdantz, I. Gurevych, and O. Deussen

Vis for Time-Evolving Document Collections



Marian Dörk, Daniel Gruen, Carey Williamson, and Sheelagh Carpendale. **A Visual Backchannel for Large-Scale Events**.
TVCG: Transactions on Visualization and Computer Graphics (Proceedings Information Visualization 2010

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 explo- ration of revision histories by combining text animated tran- sitions with simple navigation and visualization tools. We finally describe a user study showing that smooth text anima- tion allows users to track changes in the evolution of textual documents more effectively than flipping pages.”

Video on the webpage

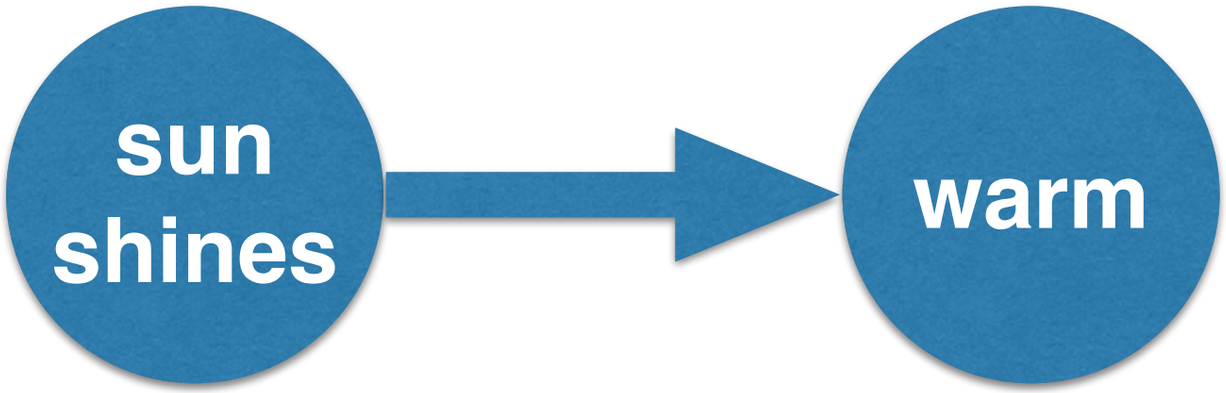
Chevalier, F., Dragicevic, P., Bezerianos, A., and Fekete, J.
Using text animated transitions to support navigation in document histories.
Proceedings of the 28th international Conference on Human Factors in Computing Systems CHI '10

The Role of Text in Vis

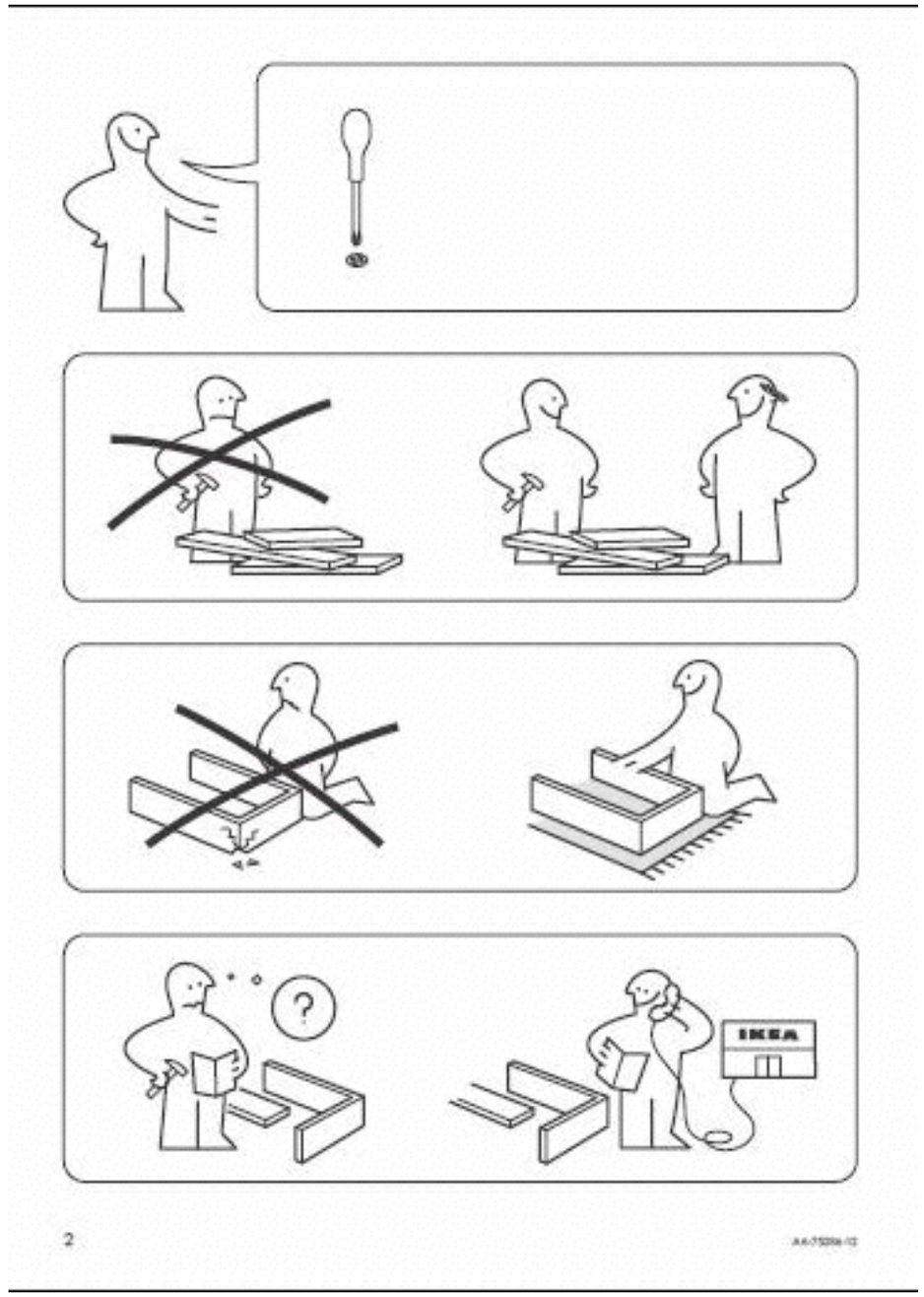
Text in Vis

- Non-Example: Ikea

- Labels:



- Map Legends



Text in Vis Storytelling

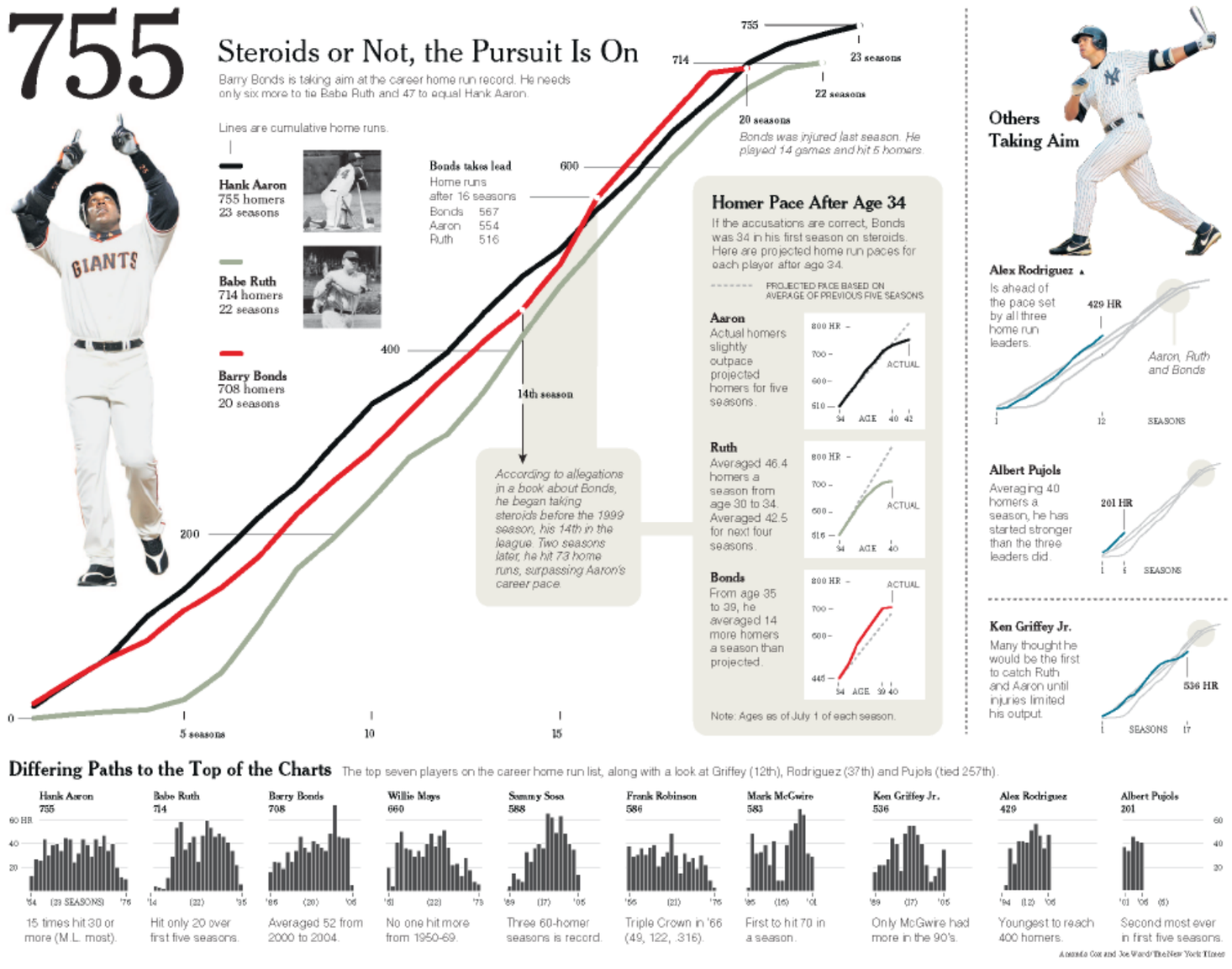


Fig. 1. Steroids Or Not, the Pursuit is On. New York Times.

Narrative Visualization: Telling Stories with Data
Edward Segel, Jeffrey Heer
IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 2010

TextVis Specials

Vis for Text Translation

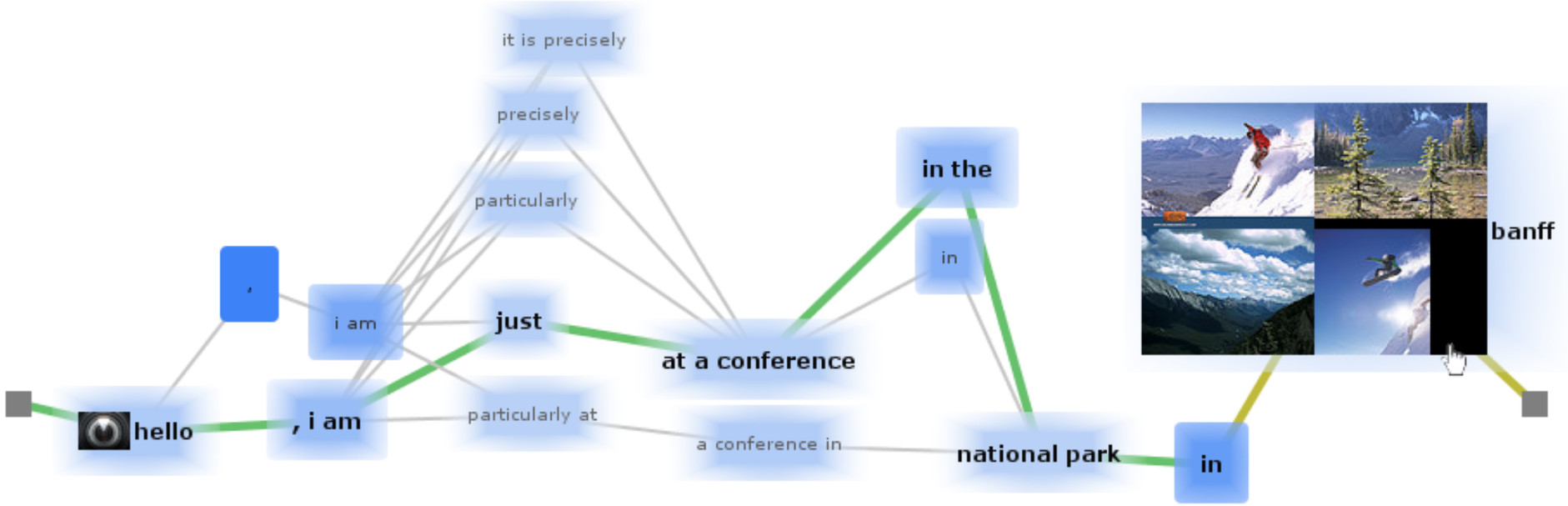
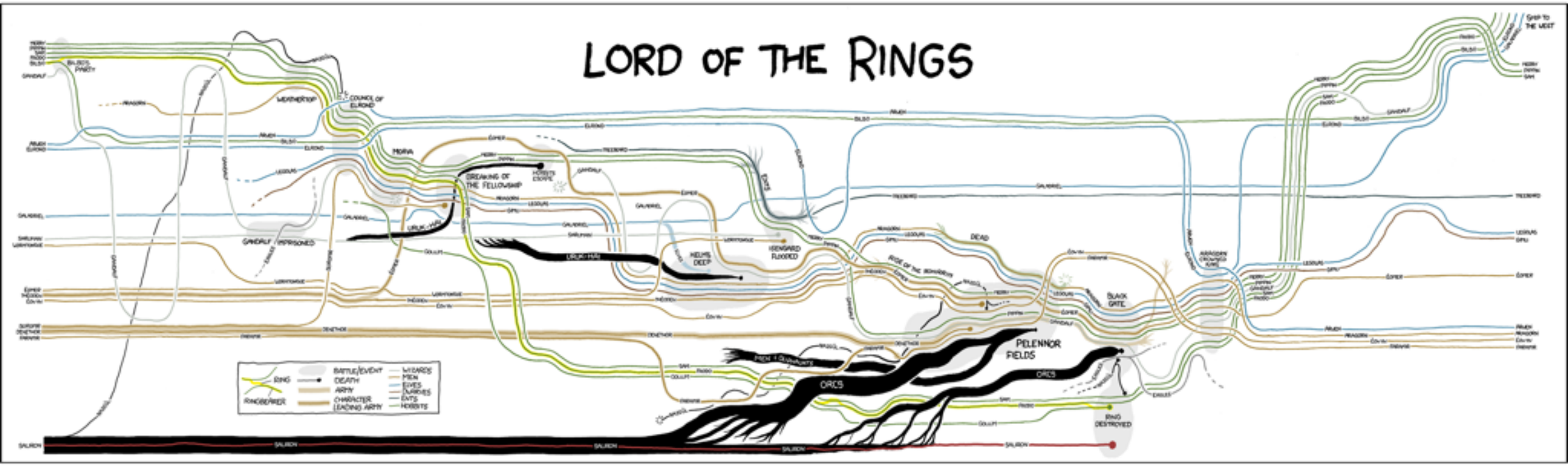
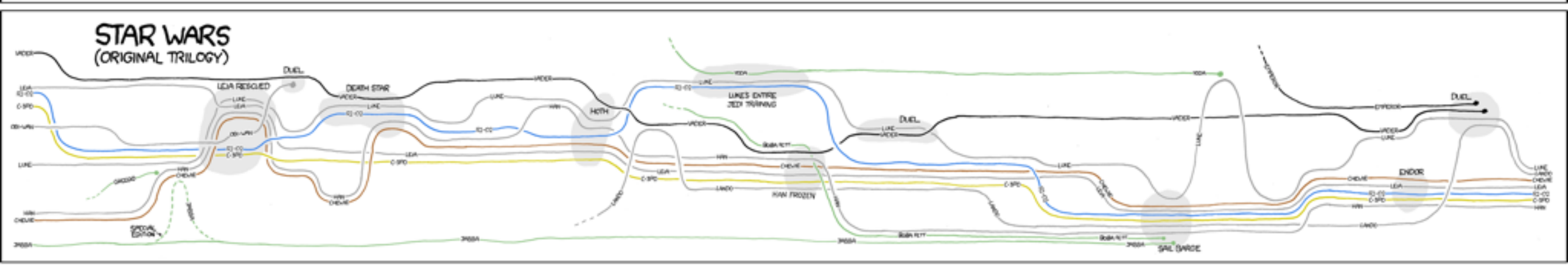


Figure 6: Translation lattice for the German sentence, “Hallo, ich bin gerade auf einer Konferenz im Nationalpark in Banff.” The statistically-identified best path (along the bottom) was incorrect and has been repaired. Photo nodes provide an alternative representation for words not in the translation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

C. Collins, S. Carpendale, and G. Penn
Visualization of Uncertainty in Lattices to Support Decision-Making
Proc. of Eurographics/IEEE VGTC Symposium on Visualization (EuroVis), Norrköping, Sweden, 2007

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.





https://xkcd.com/657/

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 convert- ing 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.lnu.se/