

Paul Rosen

paul.rosen@utah.edu  
@paulrosenphd  
<https://cspaul.com>

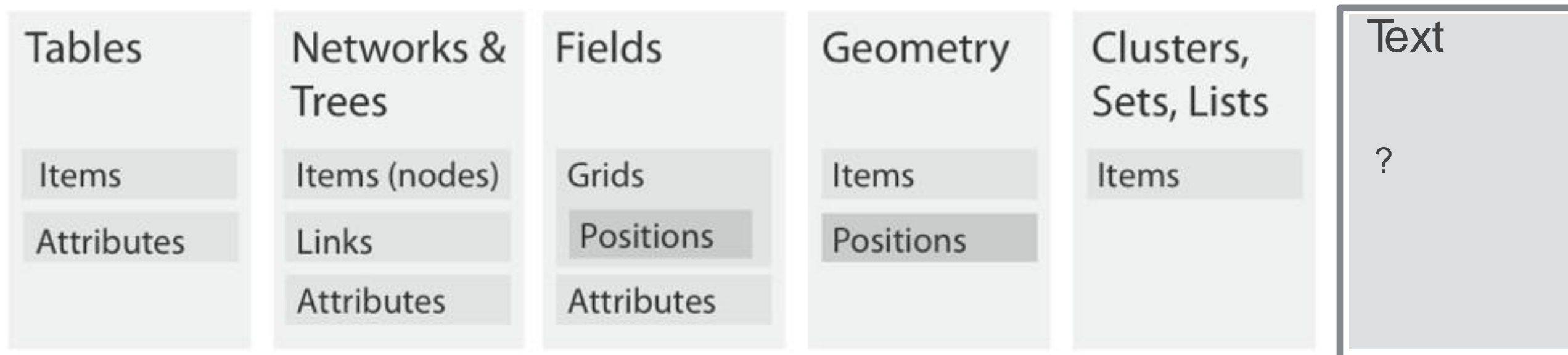


# Visualization for Data Science

## DS-4630 / CS-5630 / CS-6630

VISUALIZING TEXT

# what does it mean to be an “item”?



# text data type

- no numbers (implicitly)
- characters (ASCII)
- strings

USASCII code chart

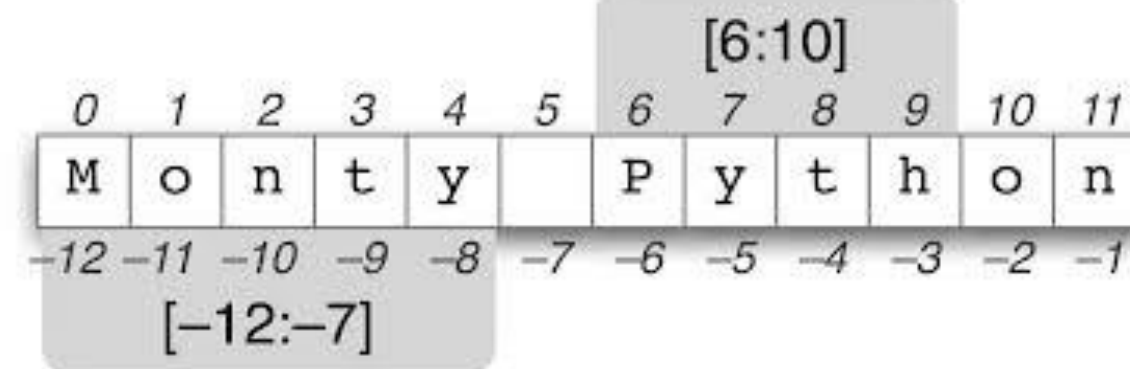
Bits					Column										
b <sub>7</sub>	b <sub>6</sub>	b <sub>5</sub>	b <sub>4</sub>	b <sub>3</sub>	b <sub>2</sub>	b <sub>1</sub>	b <sub>0</sub>	0	1	2	3	4	5	6	7
Row	b <sub>4</sub>	b <sub>3</sub>	b <sub>2</sub>	b <sub>1</sub>	b <sub>0</sub>	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	NUL	DLE	SP	0	@	P	\	p		
0	0	0	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 type

- no numbers (implicitly)
- characters (ASCII)
- strings

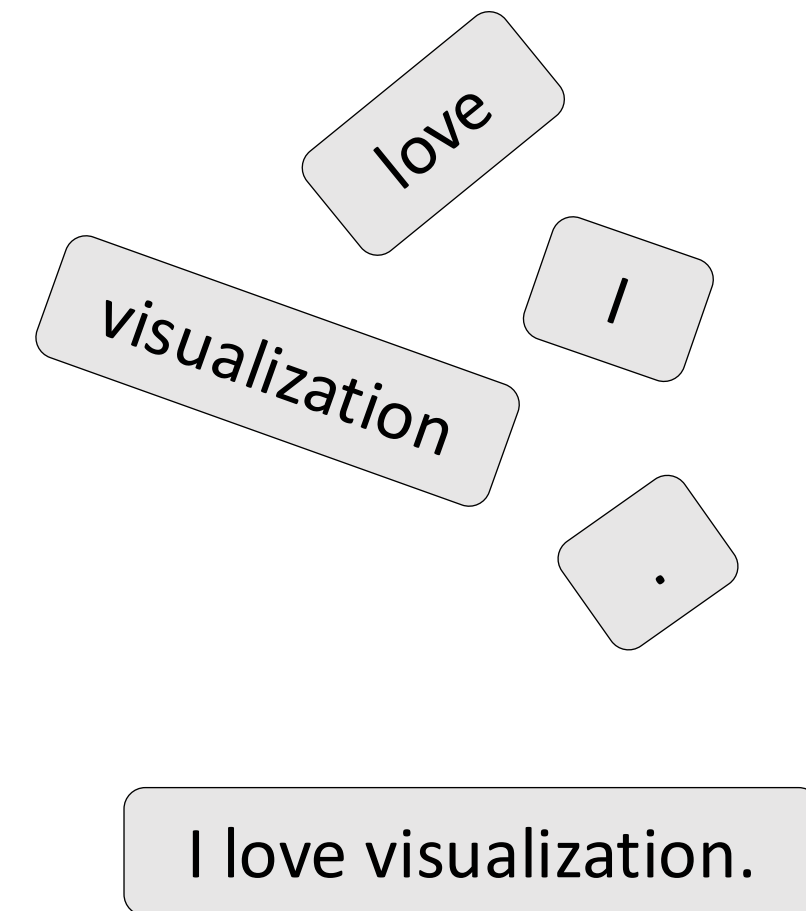
USASCII code chart

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



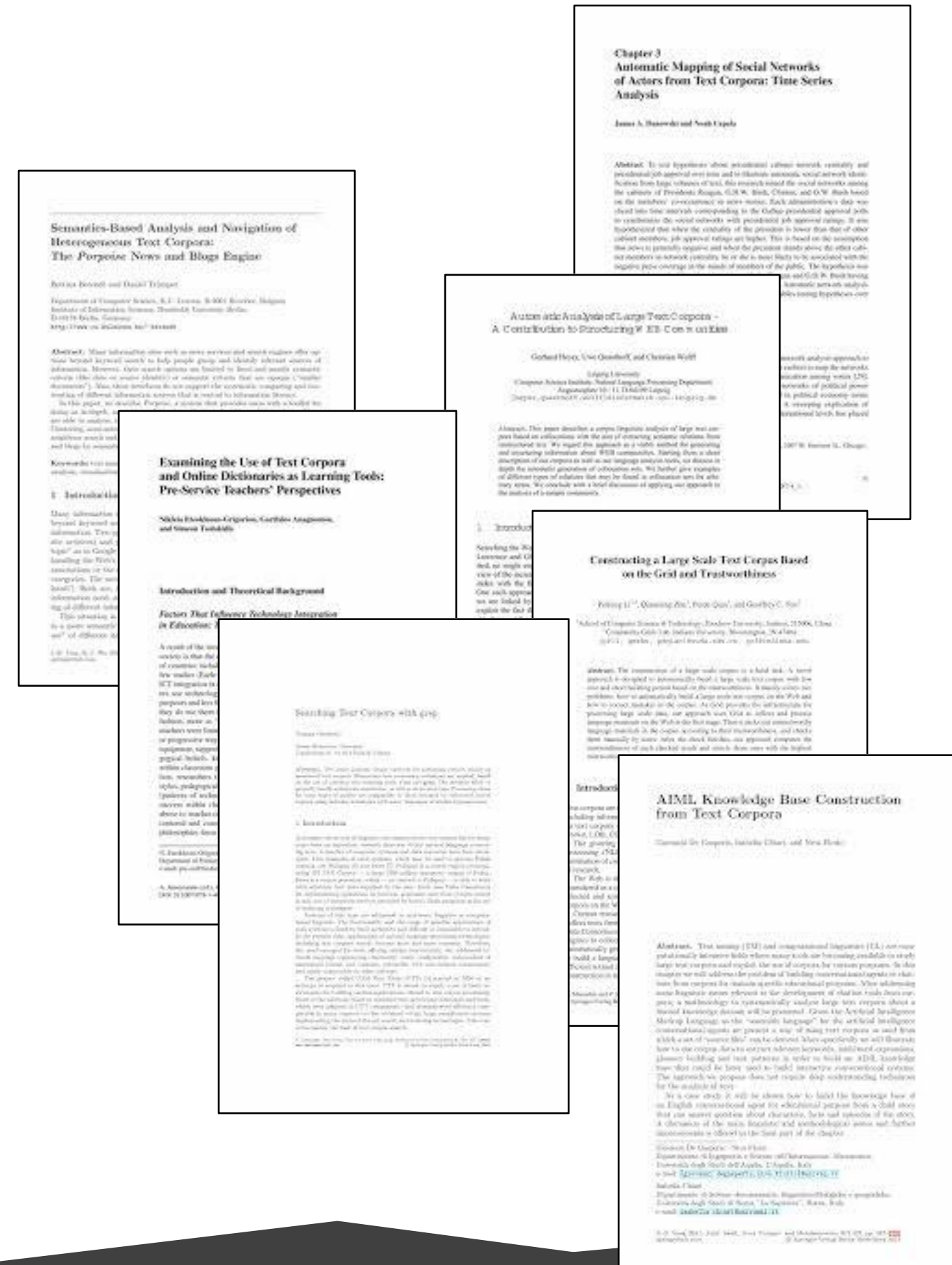
# text data semantics

- words
- lines
- sentences
- paragraphs
- chapters

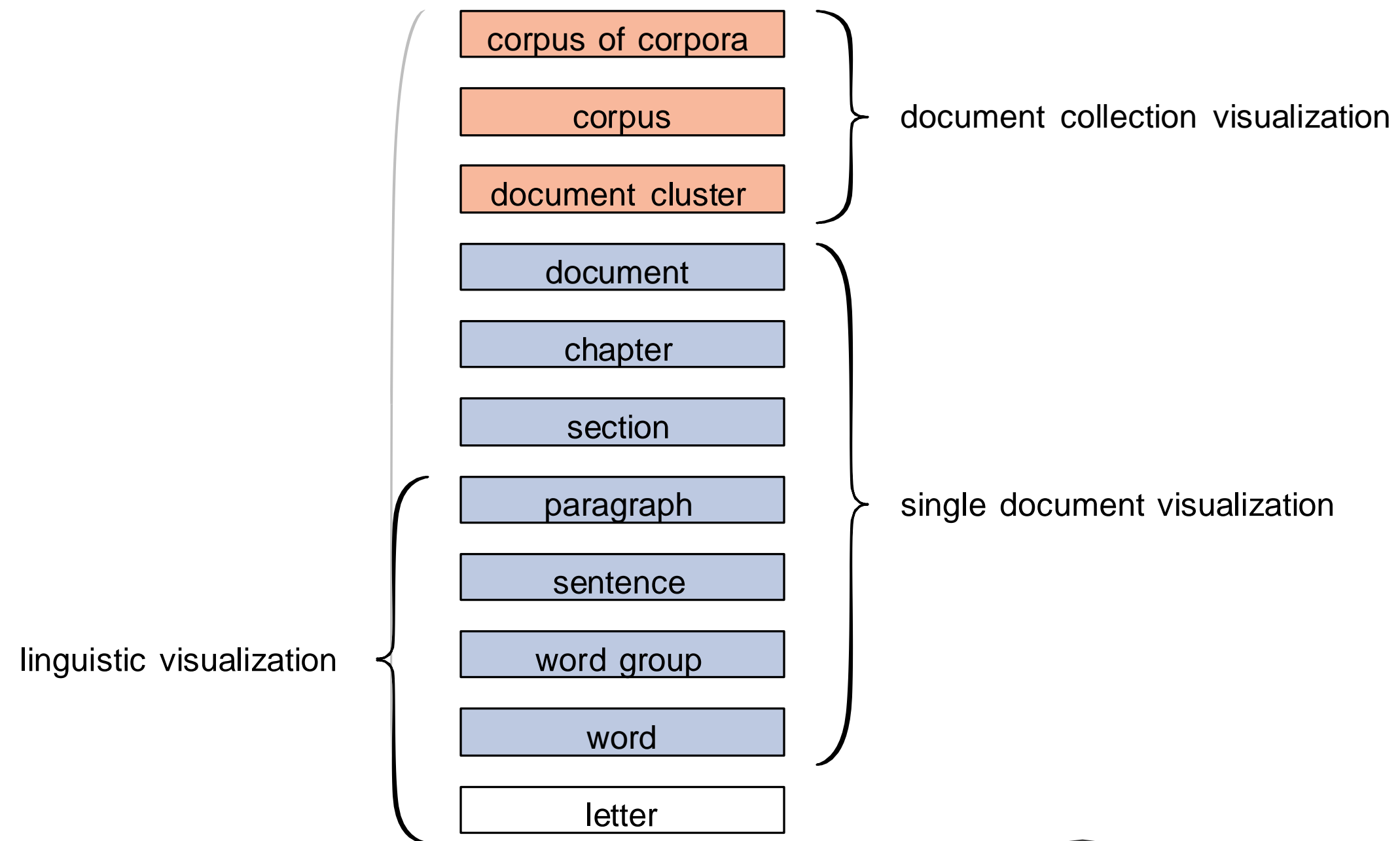


# text data semantics

- documents: books, papers, webpages, e-mails, twitter post
- corpus: collection of documents



# Text Units Hierarchy



single document



# Tag Cloud / Word cloud

abstract accepted analogue applications applying attuned bar burgeoning challenging  
 chapters chart collections combine communicate conducted convert data date difficult  
 discussed earlier effectively end evaluation evocative familiar field focus focused form  
 general goal graph highly human hundreds ideas images improve  
**information** innovative insight kinds line makes means  
 meta-analysis nature new numbers order ost perceive perceptual points positive  
 problems providing purpose range rapidly read reading reasons representations **results**  
 retrieval robust **search** shorten[chen2000esi] shortcite[larkin1987dsw] shown space  
 studies successful system table task tasks **text** textual time translate underlying  
 usability vibrant **visual** **visualization** visually web wide widely

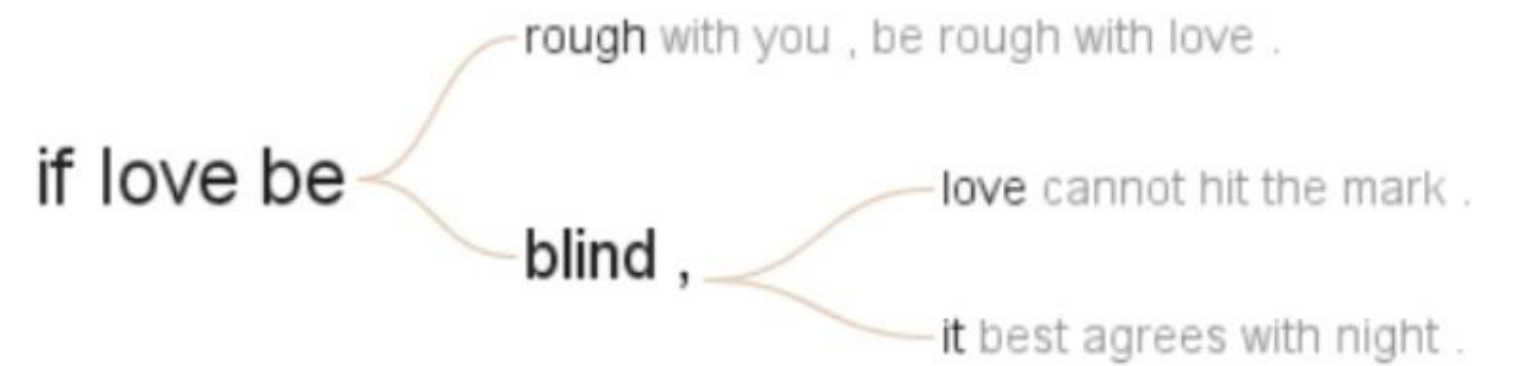


# Word Tree

## 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



[Wattenberg 2008]

# Visualizations : definitions of visualization word tree

Uploaded by: mhalle

Created at: Wednesday May 21 2008, 11:37 PM

Tags: text

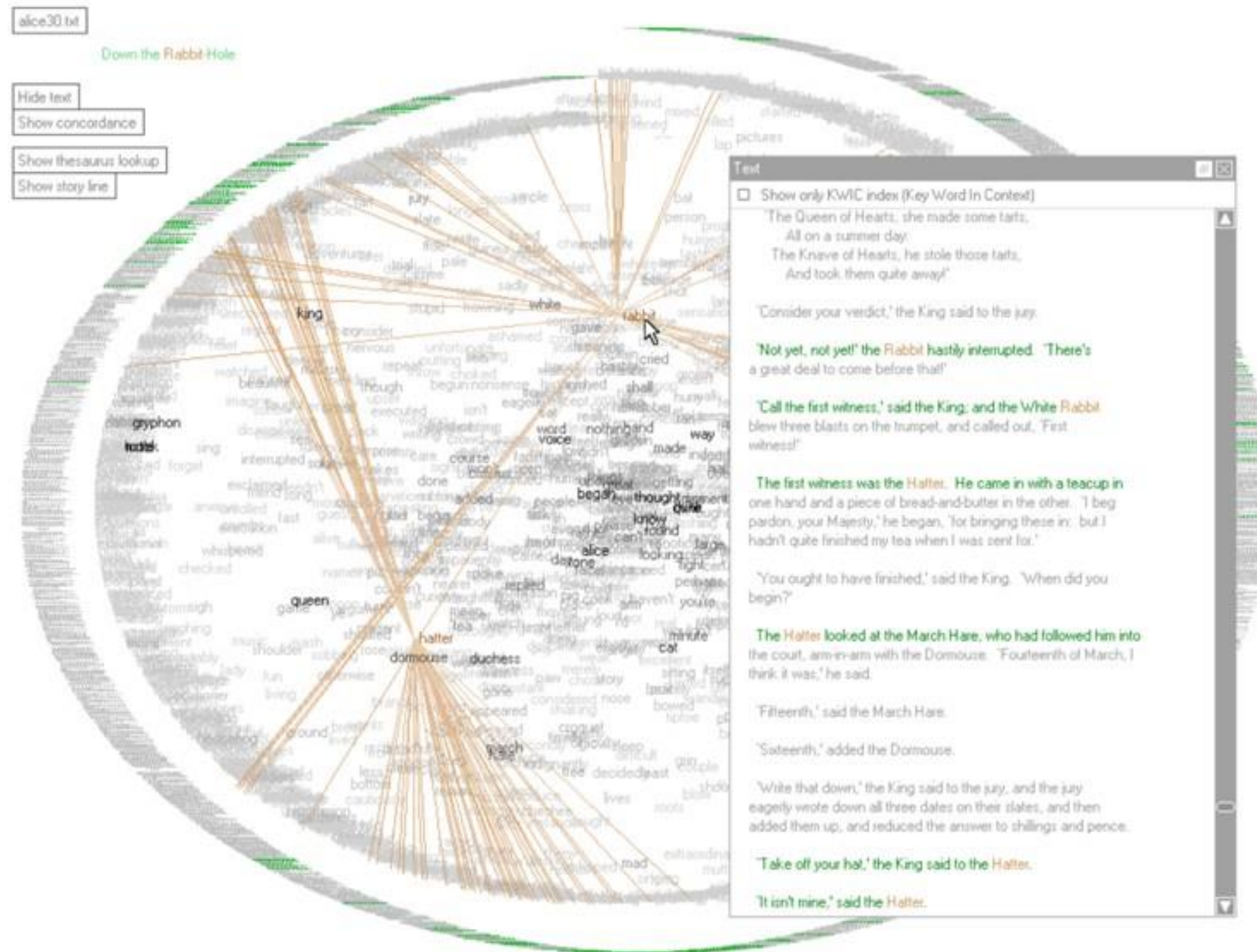
Search    Back   Forward    Start    End   Occurrence Order   Clicks Will Zoom

63 hits

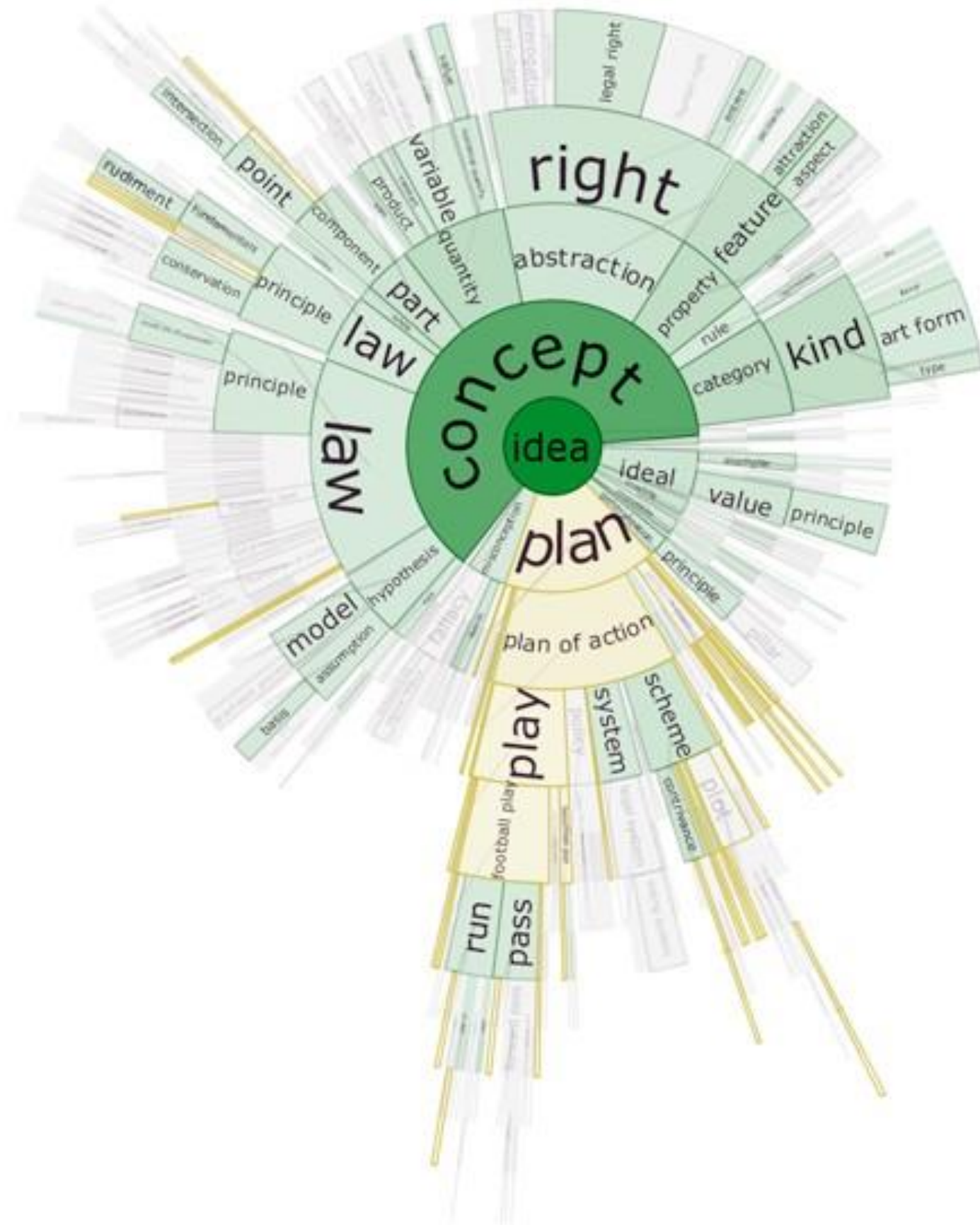
**visualization**

- is**
  - the
    - process of
      - representing abstract business or scientific data or images that can only be understood by the meaning of the data ( www.igppa.com) - transforming information into a visual form, enabling users to observe the information.
    - graphical
      - presentation of information with the goal of providing the viewer with a qualitative understanding of the information content ( selfline word).
    - use of
      - computer - supported, interactive, visual representations of data that include graphical, cognitive (see detail, 1999).
      - interactive
        - dynamic representations, typically visual, of abstract data to enhance cognition.
        - visual representations of data to simplify cognition.
    - study
      - visual representations to transfer knowledge between abstracted persons ( burland and reiser, 2004).
      - compact graphical presentation and user interface for manipulating large numbers of items possibly selected from a larger database.
  - to
    - render visible the abstract relationships that are more understandable, understood, analyzed, developed, fundamental ideas leading to generalizations for real applications.
  - a
    - method of
      - computing - presenting data or information in non-traditional, interactive graphical forms.
    - complex research areas.
    - specialized visualization.
    - part of computer graphics, which is from a subfield of computer science ( vogel, 2002).
    - Here from a method of computing.
    - defined as follows: visualization is the use of interactive visual representations ( such as simple cognition).
    - visualizing of abstract data.
    - arg to focus on creating images, diagrams, or animations to communicate a message ( wikipedia, 2008).
    - Full turning transaction data and currency information into charts and graphs.
    - visual - computer - aided design ( cad) to create screen images like 3d models that can be viewed from any angle and which can part of the greater field of visualization.
- is formed in a vision, image, or picture of something not visible or presented in light or other abstraction;
- right to the building; or mapping; of data to a representation that can be perceived;
- often a method for easing the viewer's
- produces ( interactive) visual representations of abstract data to enhance human cognition and perception, thus enabling the viewer to gain knowledge about the material ( vobbe ), is the continuous view of abstract data through the use of interactive visual interfaces ( vobbe and, 2000);
- enable users to make decisions, decisions, or explorations about patterns ( brett, chander, gag, vobbe ) groups of items,
- enable to efficiently map data variables onto visual elements in order to create graphic representations ( gen et al, 2002);
- enable graphics of structures, diagrams to show the structure of information and improve the ease of access to large data repositories,
- enable for display structures of data, for example, bar charts, pie charts, pie charts, combinations of them, for example,
- techniques
  - ( for example, 3d animation) ( vobbe and, 2004).
  - include video-like fading of data, layering data, using web page of 3-dimension space, using coding techniques to provide new spaces for users.
- visual - data representations of the content, or meaning, of information.
- information visualization typically deals with numerical, navigable, and high - dimensional data ( chen, 2005);
- which deals with physically - based data.
- typically deal with numerical, navigable, and high - dimensional data ( chen, 2005).
- in
  - computer programs for data collecting, transforming and representing abstract data in a form that facilitates human interaction for exploration and understanding.
- real - time, two - dimensional perception of patterns and structural relations in the abstract data ( vobbe and, 2004).
- for interactivity and dynamics of the visual representation.
- general, and the goal is to make them applicable to abstract data of any size whatever while still retaining high interactivity.
- regard to visual forms, auditory and other sensory representations are also of interest ( wikipedia, 2008).
- visualization
  - is a part of computer graphics, which is from a subfield of computer science ( vogel, 2002).
  - enable researchers, computer graphics, image processing, high performance computing, and other areas.
- make use of what is called external cognition.
- of
  - abstract data
    - is not straightforward.
    - data from any source.
- enable users to interact with a visualization, which deals with physically - based data.
- abstract, navigable data ( kay and miller, 2004);
- enable data with an inherent flexibility in components ( brett and miller, 2004);
- research in computer graphics, image processing, high performance computing, and other areas.
- uses the computer to
  - display real - world objects that cannot normally be seen, such as the shapes of molecules, oil and fuel dynamics and weather patterns.
  - convert classic picture form.
  - require intensive computing resources, and the supercomputer centers and national laboratories throughout the world are doing, of the benefits of such work by ( gen, 2004).
  - but is widely used ( gen, 2004).
  - and it is only related problems such as computational design or fluid dynamics analysis.

# Text Arc

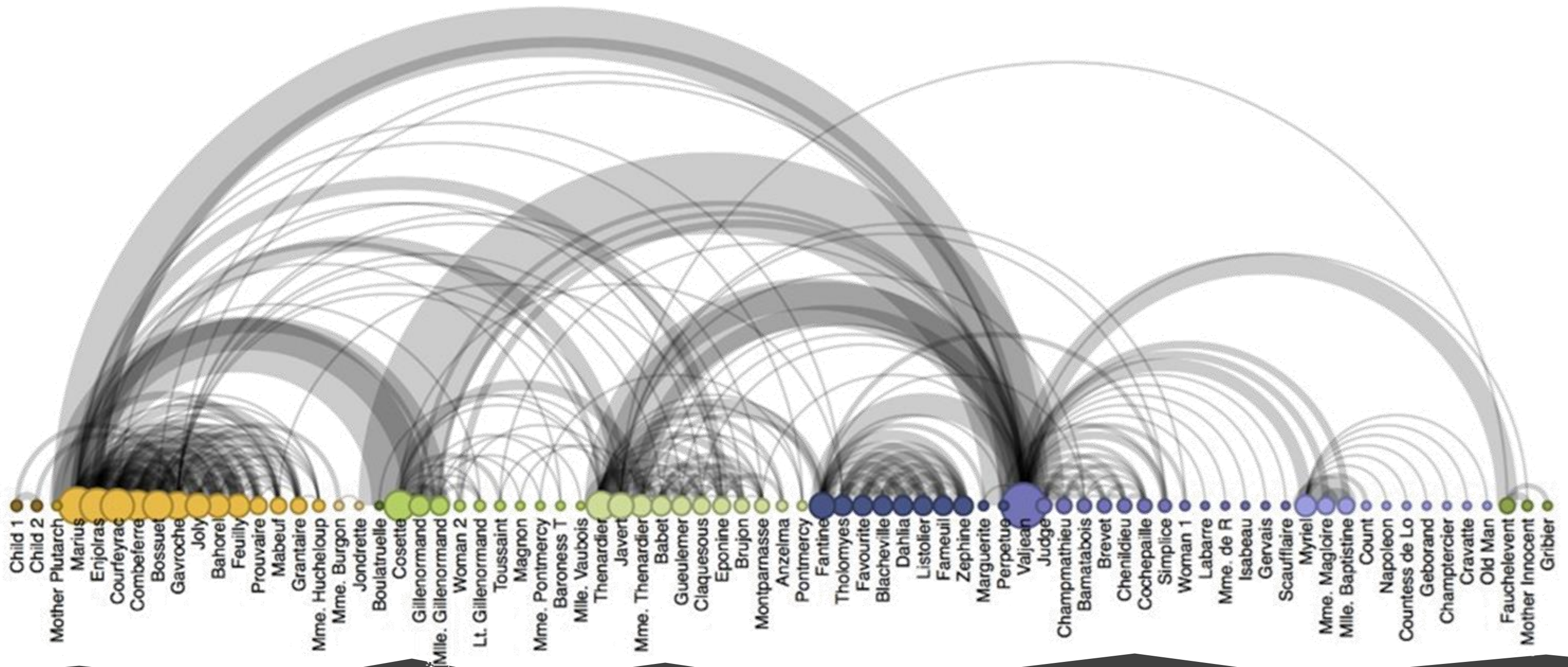


# DocuBurst

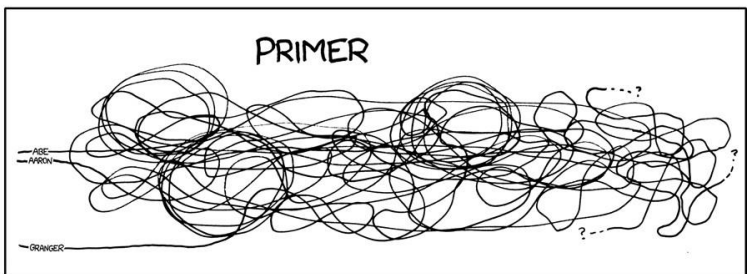
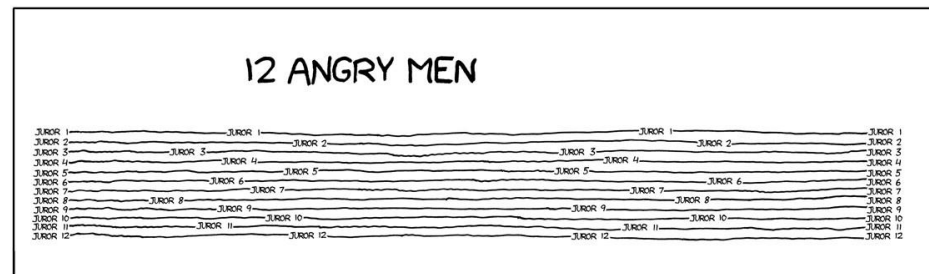
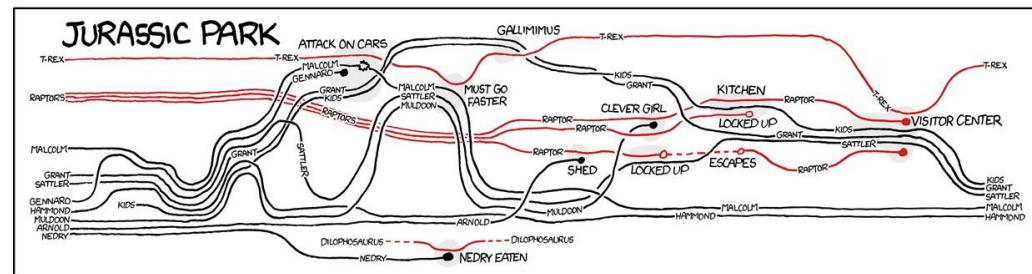
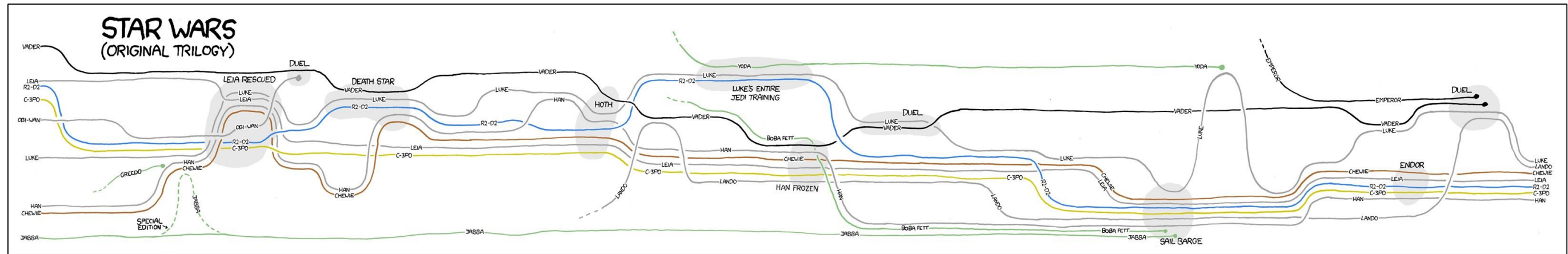
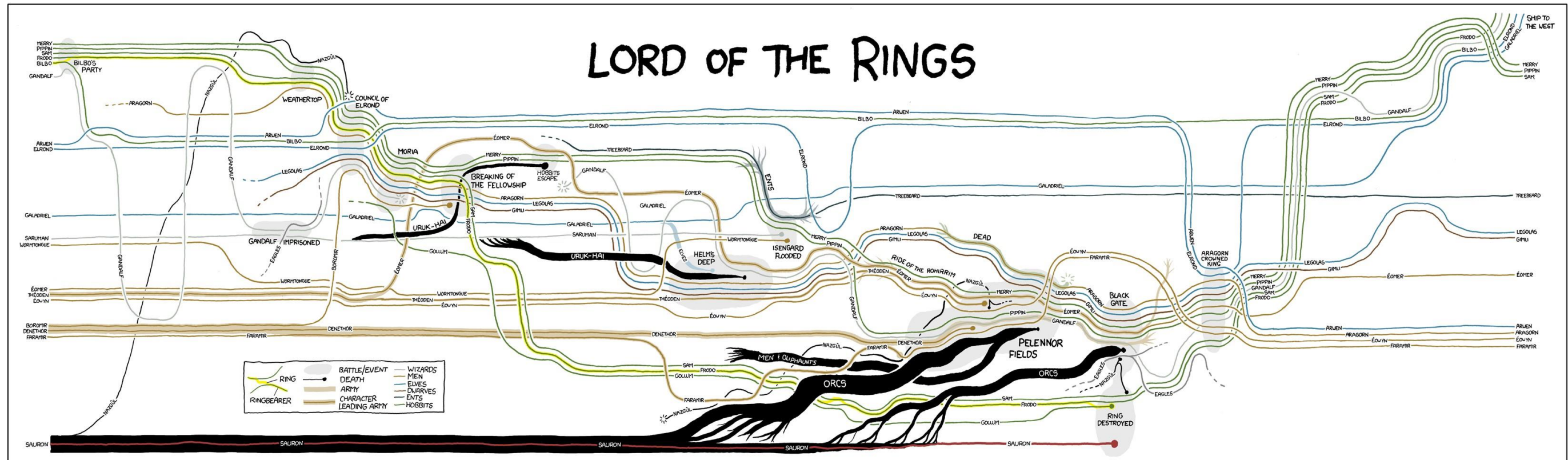


- Collins, Carpendale, Penn 2008

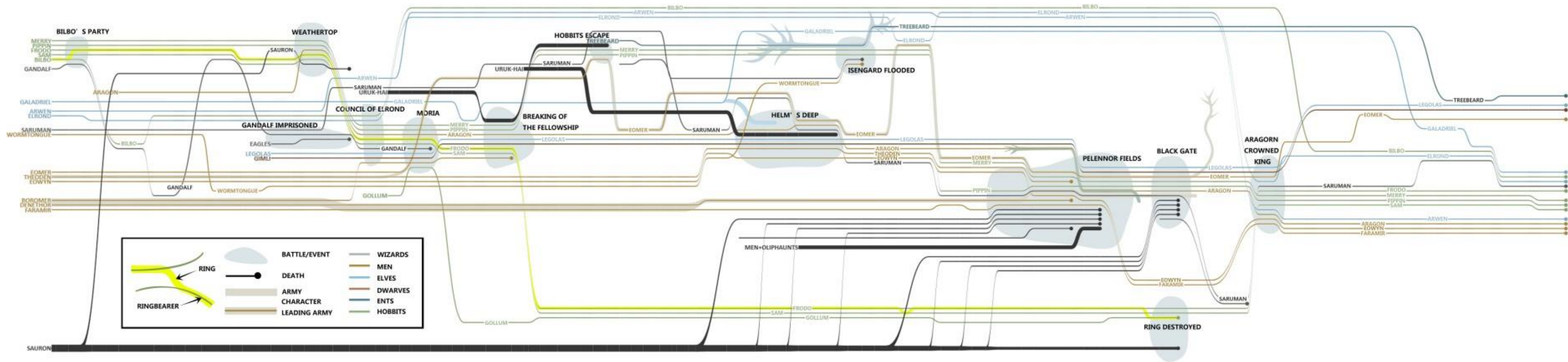
# Arc Diagram



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





collection of documents

# Text Corpora

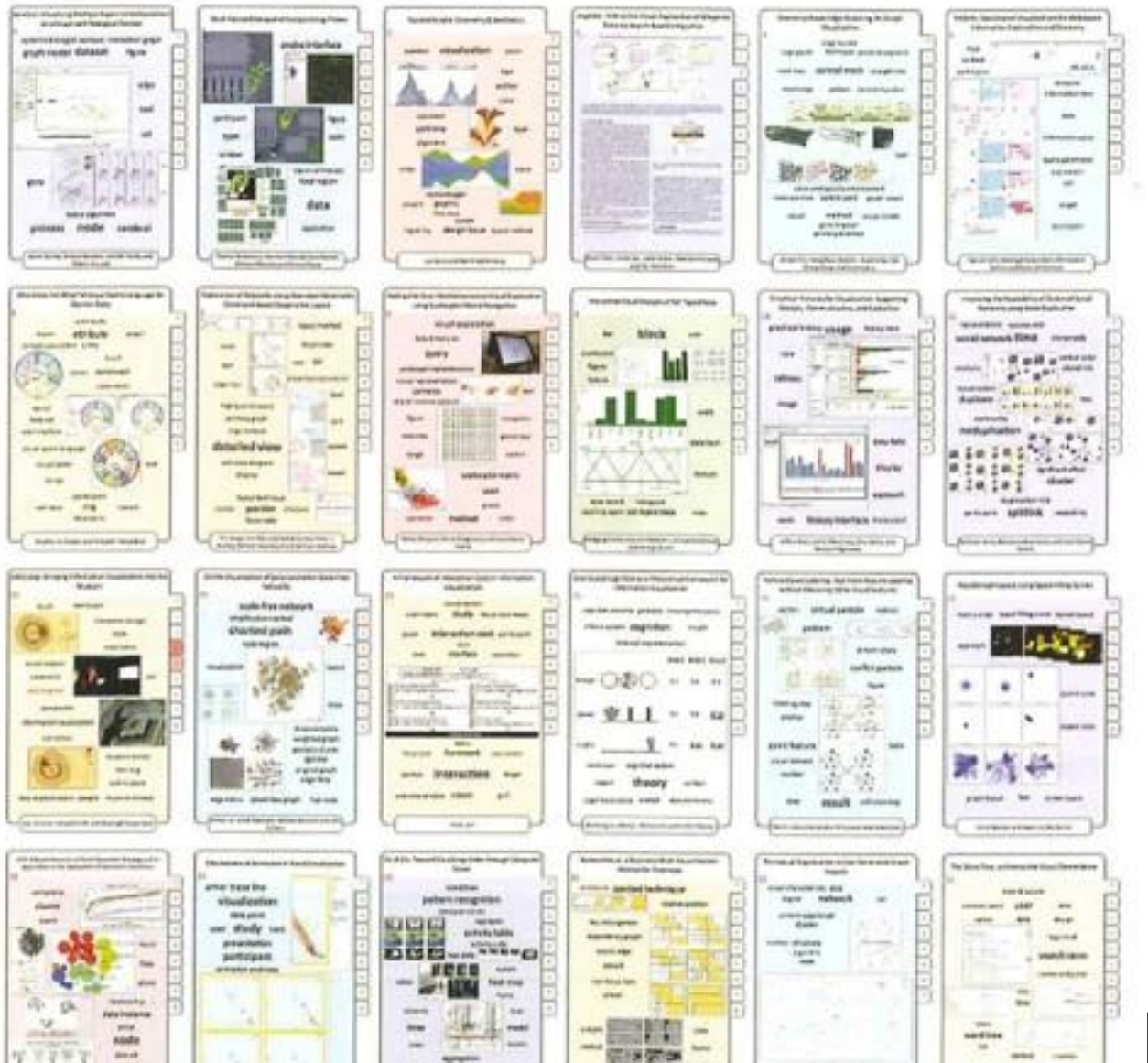
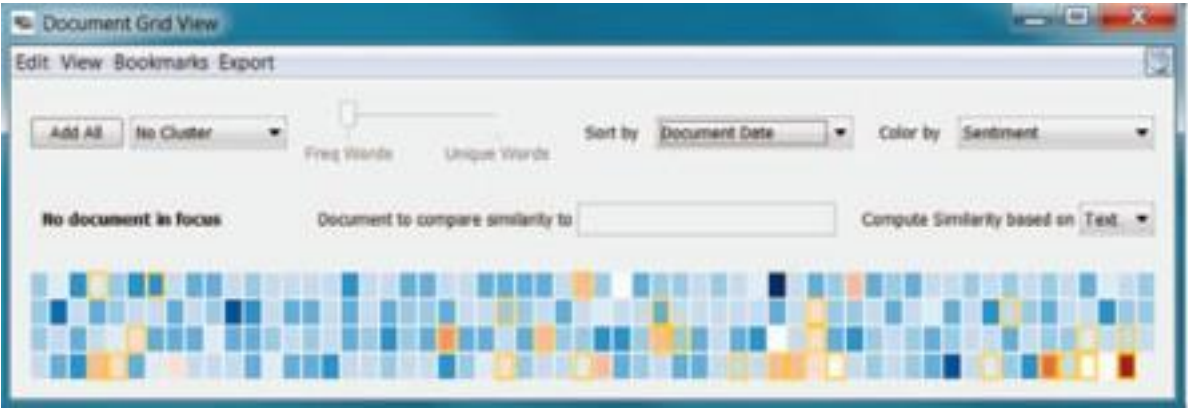
- Varied Goals:

- Discover interesting documents Summarize Documents Classify Documents
- Extract Facts (Intelligence Analysis)

- Rich Information:

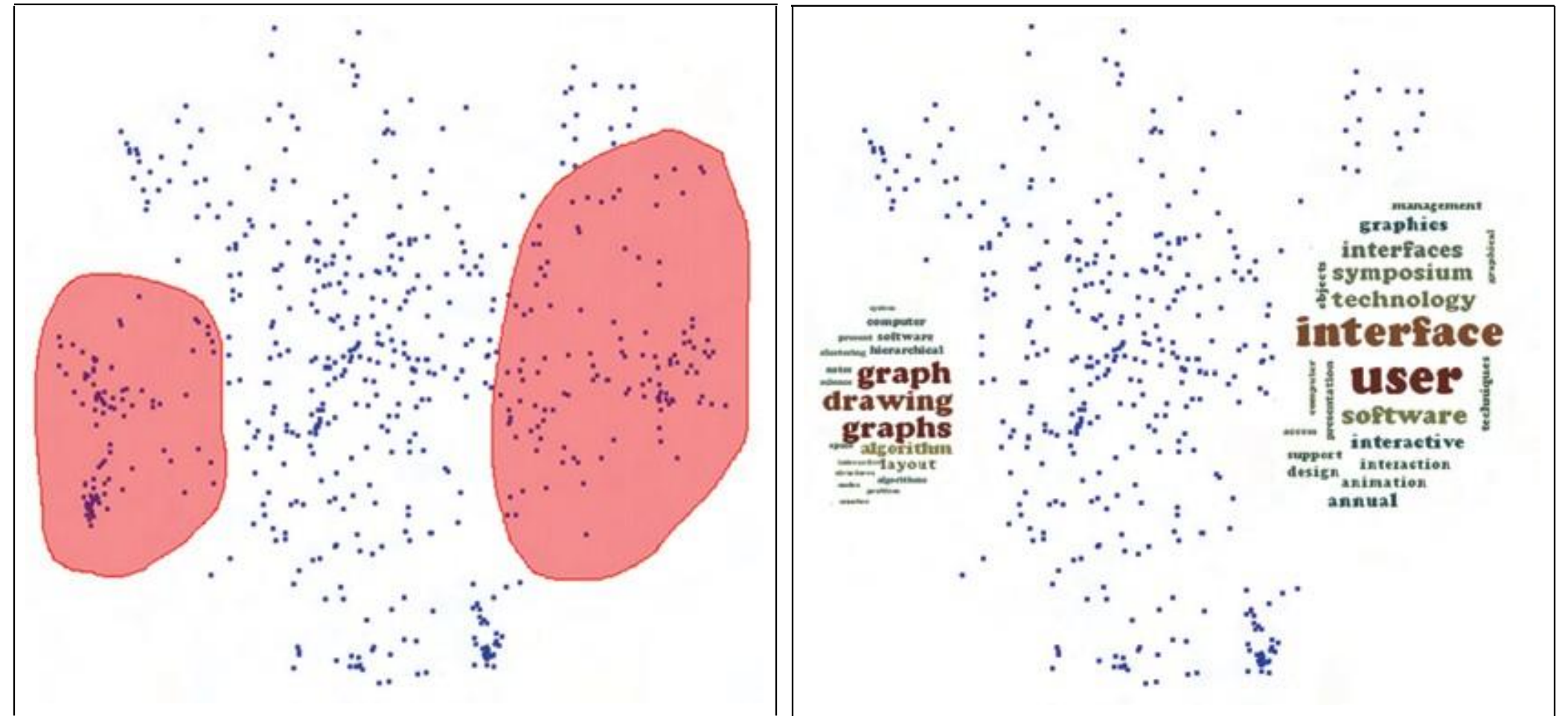
- Document Metadata Authors, date, type,
- Paragraphs, figures...
- Revisions, annotations, comments,

# Document Cards (small multiples)

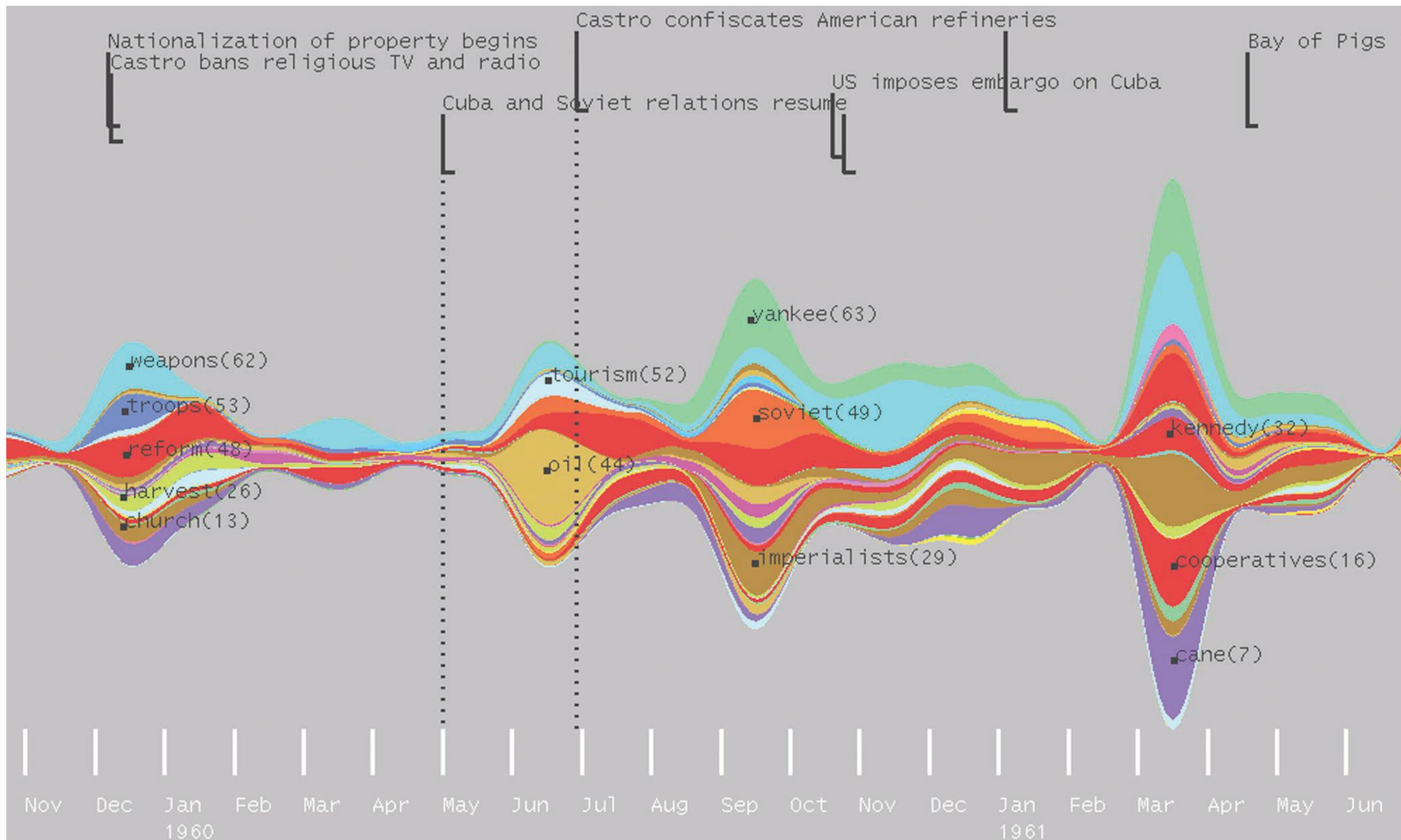


# MDS Approaches

- 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).



# Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora

Christopher Collins  
Fernanda B. Viégas  
Martin Wattenberg

0:01 / 7:40



## Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora



Christopher Collins

Subscribe 26

2,768

+ Add to   Share   ... More

12   0

Uploaded on Aug 19, 2009

This video accompanies a full publication which appeared in the Proceedings of the 2009 IEEE Symposium on Visual Analytics Science and

# Jigsaw: Many Linked Views

## **Visual Analytics Support for Intelligence Analysis Case Study: The 9/11 Report**

Carsten Görg  
Youn-ah Kang  
Zhicheng Liu  
John Stasko



Information Interfaces Group  
Georgia Institute of Technology

# <http://textvis.lnu.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

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

Visualization Tasks

- Star
- Save
- List
- Eye
- Zoom
- Pan
- ...

Data

Source

- File
- Folder
- Upload

Properties

- Info
- Refresh
- ...

