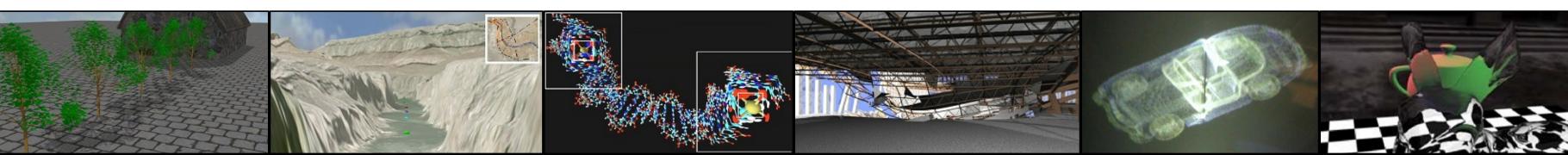
CIS 4930/6930-002 DATA VISUALIZATION

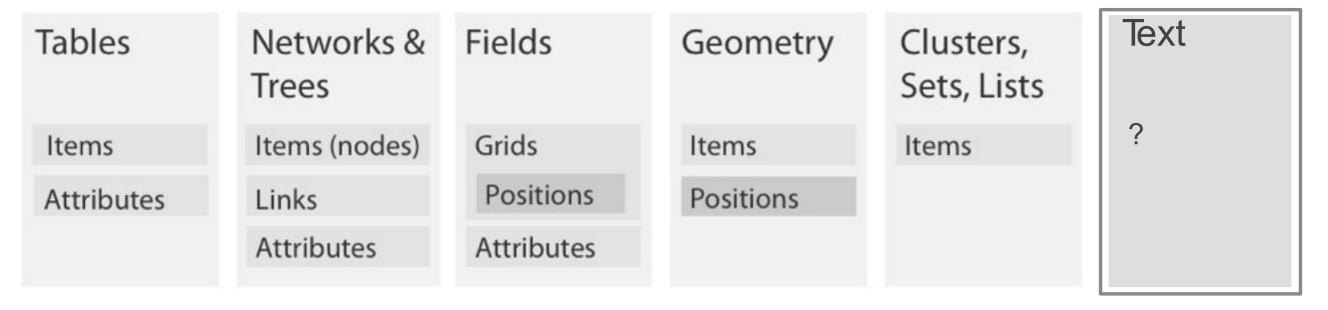


VISUALIZING TEXT

Paul Rosen
Assistant Professor
University of South Florida

slides credits Miriah Meyer (U of Utah)





WHAT DOES IT MEAN TO BE AN "ITEM"?



TEXT DATA TYPE

no numbers (implicitly)
characters (ASCII)
strings

USASCII code chart

5 -	_	_		=_	°°°	°0 ,	٥, ٥	٥,	100	0,	1 10	١,,
5 - 64	b 3	p ⁵	١,	Rowi	0	1	2	3	4	5	6	7
0	0	0	0	0	NUL .	DLE	SP	0	0	Р	`	P
0	0	0	1	1	SOH	DC1	!	1	Α	Q	0	q
0	0	1	0	2	STX	DCS	-	2	В	R	b	,
0	0	1	1	3	ETX	DC3	#	3	С	S	С	5
0	1	0	0	4	EOT	DC4		4	D	Т	d	1
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	8	6	F	٧	f	٧
0	T	1	1	7	BEL	ETB	,	7	G	w	9	w
1	0	0	0	8	BS	CAN	(8	н	×	h	×
T	0	0	1	9	нТ	EM)	9	1	Y	i	у
T	0	1	0	10	LF	SUB	*	:	J	Z	j	2
1	0	1	1	11	VT	ESC	+	:	K	C	k .	(
T	T	0	0	12	FF	FS		<	L	\	1	1
T	1	0	1	13	CR	GS	-	W.	м)	m	}
1	1	1	0	14	so	RS		>	N	^	n	\sim
T	1	1	1	15	SI	us	/	?	0	_	0	DEL



TEXT DATA TYPE

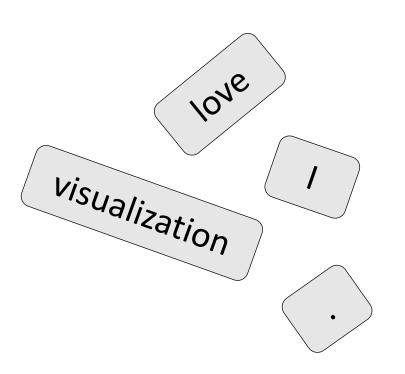
no numbers (implicitly)
characters (ASCII)
strings

						US	ASCII	code	chart				
b _{7,6}	ξb ₅ -		_		=-	۰,	°° ,	٥, ٥	٥,	'00	0,	1,0	1,
1	\$ D 5	b 3	p ⁵	١,	Row	0	1	2	3	4	5	6	7
	10	0	0	0	0	NUL .	DLE	SP	0	0	Р	,	P
	0	0	0	1	1	soн	DC1	!	1	Α	Q	0	q
	0	0	1	0	2	STX	DC2	-	2	В	R	b	r
	0	0	1	1	3	ETX	DC3	#	3	C	S	c	5
	0	1	0	0	4	EOT	DC4	1	4	D	Т	d	1
	0	1	0	1	5	ENQ	NAK	%	5	E	U	e	U
	0	1	1	0	6	ACK	SYN	8	6	F	V	f	٧
	0	1	1	1	7	BEL	ETB	'	7	G	w	g	w
	1	0	0	0	8	BS	CAN	(8	н	x	h	x
	1	0	0	1	9	нТ	EM)	9	1	Y	1	у
	1	0	1	0	10	LF	SUB	*	:	J	Z	j	2
	1	0	1	1	11	VT	ESC	+	:	к	1	k.	
	1	1	0	0	12	FF	FS		<	L	'	1	1
	1	1	0	1	13	CR	GS	-		м)	m	}
	1	1 1	11	10	1 14	so	RS	١.	>	l N	_	l n	\sim
								[6:1	0]			_	DEL
	1	2		3	4	5	6	7	8 9	9 10) 11		
2000	0	n		t	У		P	У	t l	1 0	n		
? .	-11	-10	0	-9	-8	-7	-6	-5 ·	4 -	3 -2	2 -1		
	-11 [-	12	:	7]									



TEXT DATA SEMANTICS

words lines sentences paragraphs chapters



I love visualization.



TEXT DATA SEMANTICS

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

Chapter 3 Automatic Mapping of Social Networks of Actors from Text Corpora: Time Series Analysis

James A. Banowski and Neath Capely

Abstract. To use begoeiness whose procedural values security and procedural spin agents of section and to illustrate assumption, social artistics than because the social procedural spin agents of the section of the social security and to be calment of Provident Require, ICFR Mont. (STAM 100.4 CHRMS and CFM Maller Section on the sensitive convergence or some motion. But allowant times in the section of the most interest of the section of

retionic services establish-

Automatic Analysis of Large Text Corpora -A Contribution to Structuring V EB Con numbers

Gotherd Hopes, Unit Quantited, and Christian World

Unjury Lowersty
Computer Science Institute, Natural Language Processing Department
Augmanglate (A): 11. D Add Philapping
Department of Computer Science Institute Computer Science Inst

Almanan. This paper destribus a popul linguistus analysis of large near pear hand an authorities with the size of networking presence obstance, summarized our We request this approach as a state stated the paties and approaching to the size of the depth the secondary generation of collections and will be be put a size of different types of relation that may be found as estimation near law, may state. We conclude that is found that the property of the size of entificand/virigipros/viri editor in supplie setrovito, scator intend virial (27); covario of policial power is policial economiesto, coverpting poplication of

RFR Better S. Groger

201,0

5. Interestual Searching the Wo Learnings and G ded, as single or view of the sound andre with the B One and sound

Introduction and Theoretical Background

Examining the Use of Text Corpora and Online Dictionaries as Learning Tools: Pre-Service Teachers' Perspectives

in Education: 2

10 ha 0.1 m

Semantics-Based Analysis and Navigation of

Teperment of Computer Scatter, 8,1: Leaves, 9-200 Scotter, Salgara Scatter of Schwarzen Stremen, Naminda University Sedia. Dott N. Schwarzen MALLING, a Schwarzen Schriften.

Heterogeneous Text Corpora: The Porpoise News and Blogs Engine

> motion in the disof countries includture studies (Earle III T integration inres are uninterlugguageses and less fithey als not Germlandons, ware as

and Simusia Turkshills

they do not Object to allow, paste as a markers were finance or progressive step applications, suggested by the control of the

Parking Organization of Princip

Southing Tree Corpora with goo.

Transport Control

Freeze, "Pro-airi James Stage varietà le patricipe de la properti del properti del properti de la properti de la properti de la properti del prope

Lincoleton

The second secon

of springers, but it was arrived and affiliar in immediate a broadfor present that a springers and a death for the present products of a legisla to compare sense. In terms then that one reasons, Throuber, the new compare sense is those the product of the product of the construction o

Constructing a Large Scale Text Corpus Based on the Grid and Trustworthiness

Felling Li 1, Queening Zho', Ports Que', and Goethey C. Roof

⁵Advol of Usepairo Simon & Technology, Southern Lee monty, Smitter, 21508, C. Consentin, Galdrille Southern Strivening, Montagone, St. Alexa, 1971. parties, paterial fraction of the Area, and Smith and area onto.

Windowski. The responsition of a large made corpus to a haid stell. A new approach to designed to informationally found it large upth that corpus with 2 man and other bridge protect based in the informational term. It must give the problems former approach to the supplement of the problems former approach to the supplement of problems, the corporation of the supplement of the supplement of the information of problems and process to the supplement of the supp

AIML Knowledge Base Construction from Text Corpora

Cornell St Copyris, Indials Clief, and New York:

Abstract. This is using (TM) and oraquitational linguisms (Tk) are more put at small; intensive hide where many tools are browning and delice in risk large text on priors and region k to me of an opposit, by reason propagate, in this image, we will oblive an advergable the most because k in the contraction k in the state of the contraction k in t

As a case study, it will be shown how to build the facewings base of an English representated agent for electrical purpose from a dual story that you searned quotelow about characteris, bette sail species of the solicit. A observation of the south Englished and periodological amoust such further suppressionates in othered to the final part of the displace.

Boson (a Oupper - New Plate Equations of Egypter a c forces of Salaranama Massaca Estated dep Stock of Apple 1 Apple had a conditional Supper to the Test Health Co. III

Speciments of Science Assumance in Representationals or proposition controls degli Static di Roma "La Septemer", Ruma, Rufe and Dedictor situat Representati

S.A. Song (St.), July Seed, Seed Tringer, and Musbersonia, NY 67, pp. 577, and an appropriate page.



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 shortciten(chen2000esi) shortcite(larkin1987dsw) shown space

studies successful system table task tasks text textual time translate underlying

usability vibrant visual visualization visually web wide widely





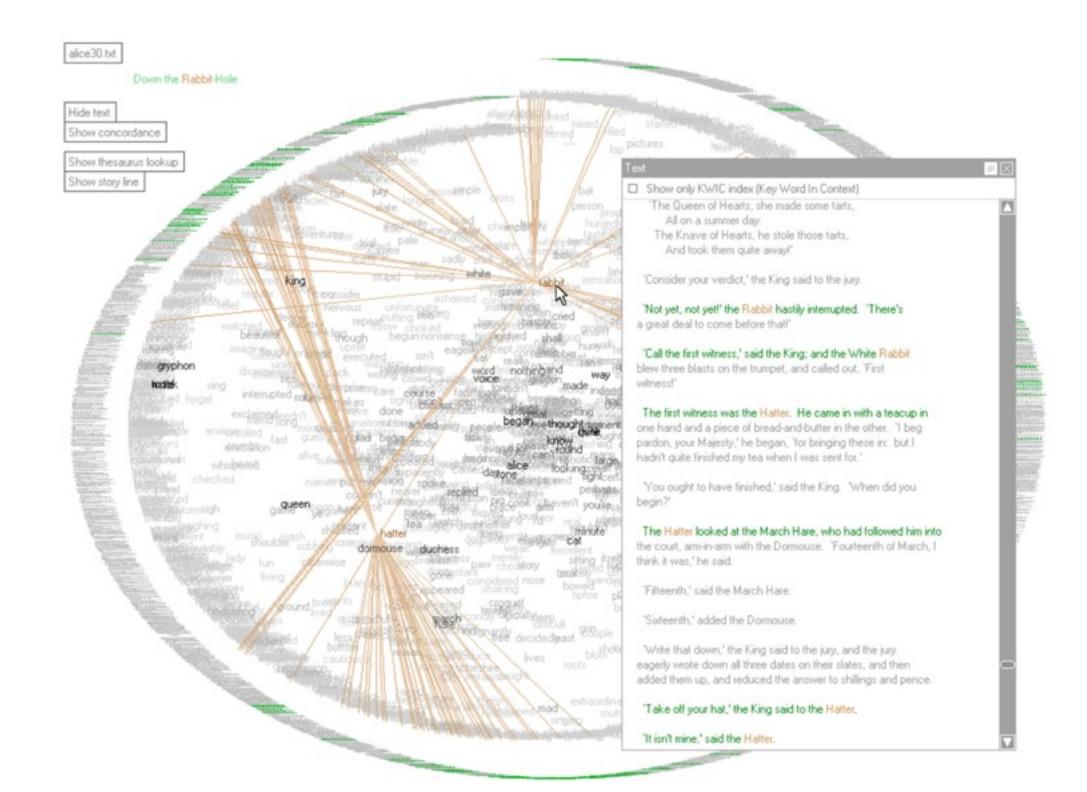
Visualizations: definitions of visualization word tree

Uploaded by: mhalle Created at: Wednesday May 21 2008, 11:37 PM

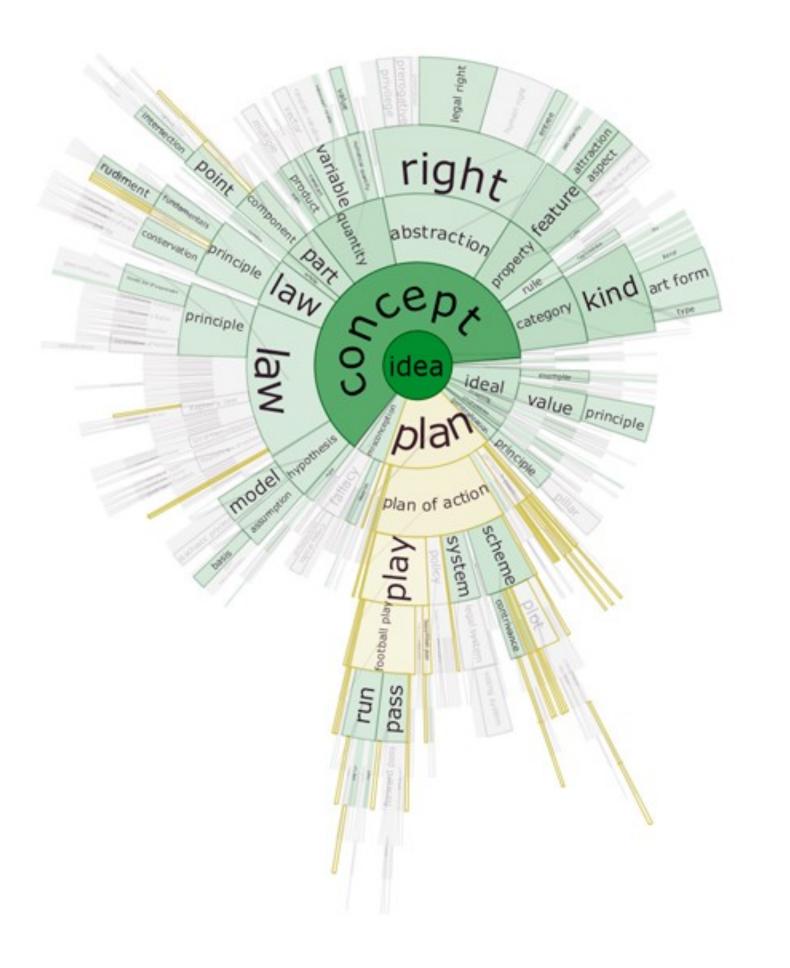
Tags: text



Text Arc



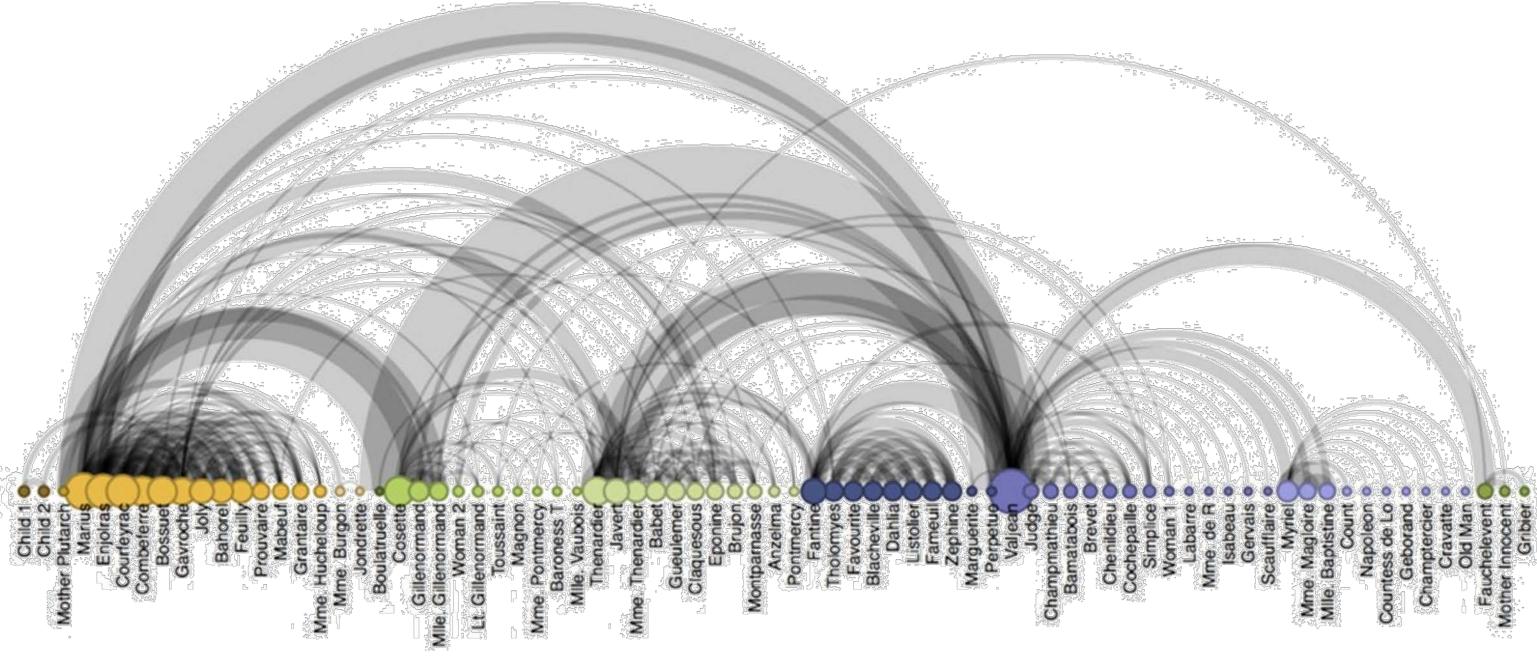
DOCUBURST





ARC DIAGRAM







COLLECTION OF DOCUMENTS







Christopher Collins Fernanda B. Viégas Martin Wattenberg

0:01 / 7:40

Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora



Christopher Collins

2,768





Add to Share ••• More

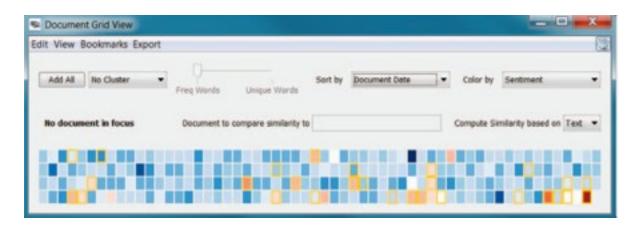








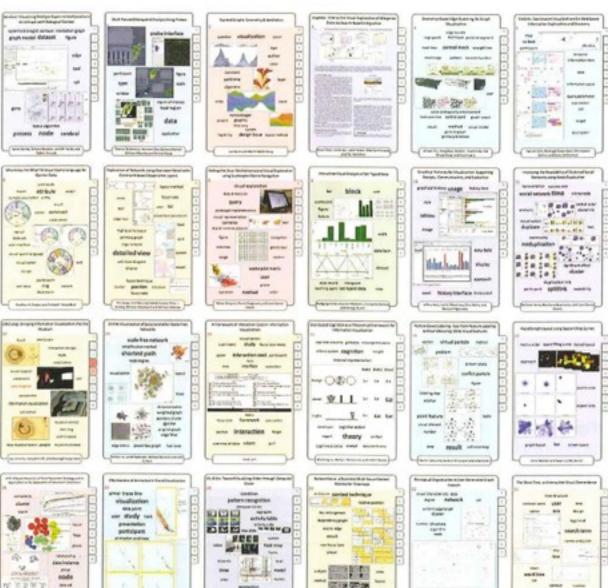




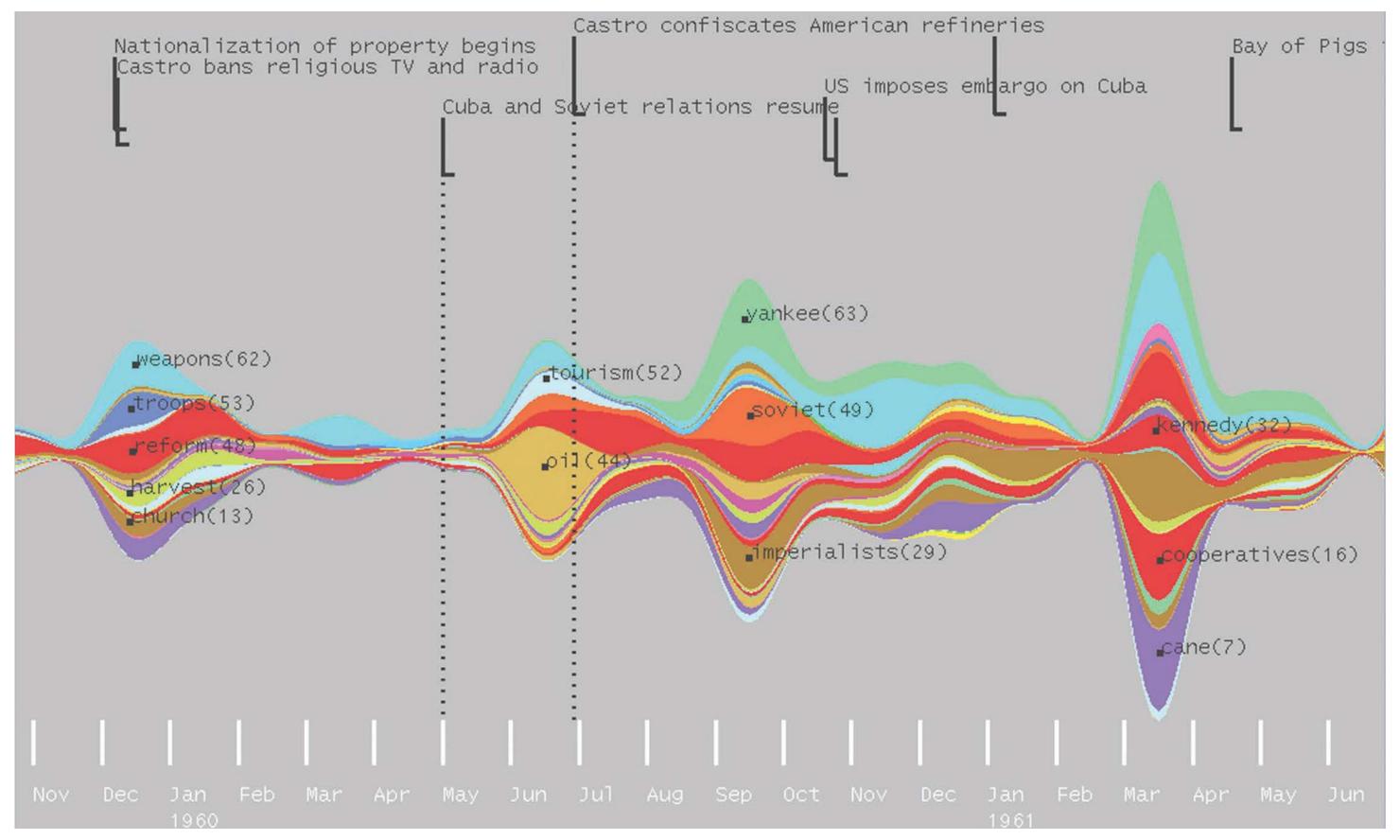
DOCUMENT CARDS

(small multiples)









2

THEMERIVER - HAVRE, HETZLER, WHITNEY, NOWELL 2002

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



