Text Visualization

Mohan Raj Rajamanickam

29 Nov 2004

CPSC 533c Presentation

Text Visualization

CPSC 533c Presentation

Mohan Raj Rajamanickam

CPSC 533c Presentation

Text Visualization

Mohan Raj Rajamanickam

23 Nov 2020
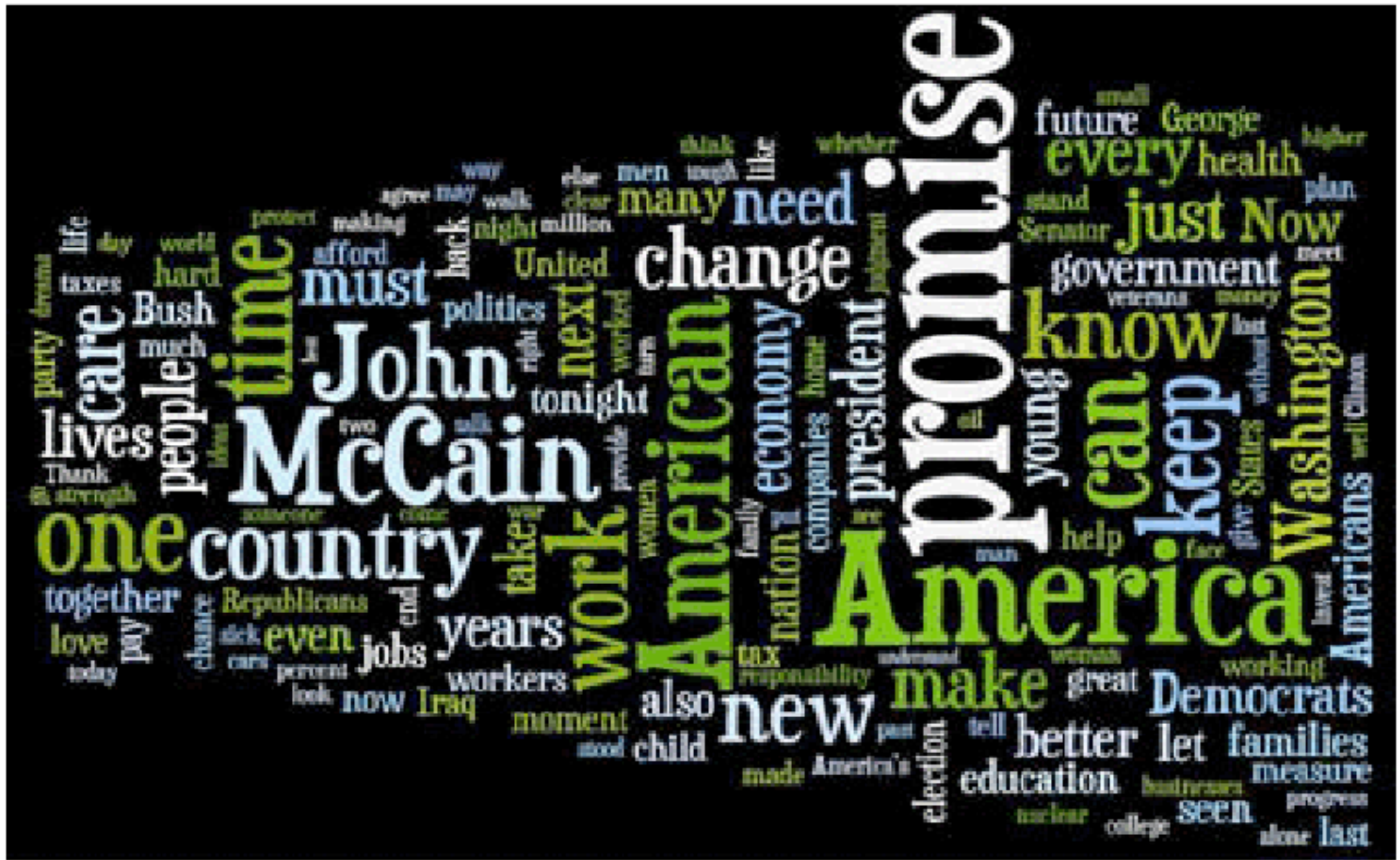
Wordle

- Improvement over Tagclouds

  - Effectiveness of Tagclouds

- wordle = iTagclouds

  - aesthetic combination of layout, type and colors

Tag Cloud of Barack Obama's speech at the Democratic Convention in 2008

Wordle of Barack Obama's speech at the Democratic Convention in 2008

10
hits

for in a child .

?

john edwards calls on us to hope .

.

i'm not talking about blind optimism here — the almost willful ignorance that thinks un

hope

slaves sitting around a fire singing freedom songs ; the hope of immigrants setting ou

immigrants setting out for distant shores ; the hope of a young naval lieutenant brave

of

young naval lieutenant bravely patrolling the mekong delta ; the hope of

a

millworker's son who dares to defy the odds ; the hope of a skinny kid wi

skinny kid with a funny name who believes that america has a place for

!

Wordtree of Obama's speech

What is Wordle useful for ?

information data design color New visual INFORMATION ENVISIONING map maps one also time colors York flatland work value right reading timetable two space diagram small line paper color Tokyo pp grid lines displays viewers graphical see even areas train trains Paul Mrs Visual University computer image display effects Galileo names numbers Albers complexity used light sun day strong satellites movement large order notation may John like surface spots William statistical designs elements Library edition first white rather Note micro/macro fine Paris three area stars chart analysis book graphic motion better Color often detail show various methods shown sequence text translated must every point means along now Edward times number music around much little view method top seen another without Redawn Information schedule images many left Art shows individual Japanese eye several different make complex Robert Jupiter dance des pattern years table sunspots

What is Wordle useful for ?

- Gives a good Overview

  - as reminders, reinforces existing knowledge, as 'memento' (~50% users of wordle)

  - as learning aids

    – used by educators

- Emotional visualization !

  - marriage vows

What is Wordle not useful for ?

- Standalone learning tool
  - prior knowledge required
- Discovering new information
- Interactive visualization

Design implications of Wordle

- Creating mass visualizations
  - Low cost of entry
  - Portability or Sharing
  - Aesthetics ~= Usability

Types of Text Visualizations

- Visualizing relationships among huge document collections
  - Targetted more at experts
  - Discovering previously unknown information is key

- Visualizing structure of a document
  - Unlikely to discover new knowledge
  - act as aids to reinforce existing knowledge

- Visualizing literature and citation relationships

German economic
and monetary union

OPEC agrees to
raise oil price

NATO to redefine
military strategy

Iraq invades
Kuwait

nato(54)

germany(36)

baghdad(11)

communist(5)

oil(12)

saddam(10)

bush(1)

iraq(9)

kuwait(8)

22   291   8   15   22   29 1   8   15
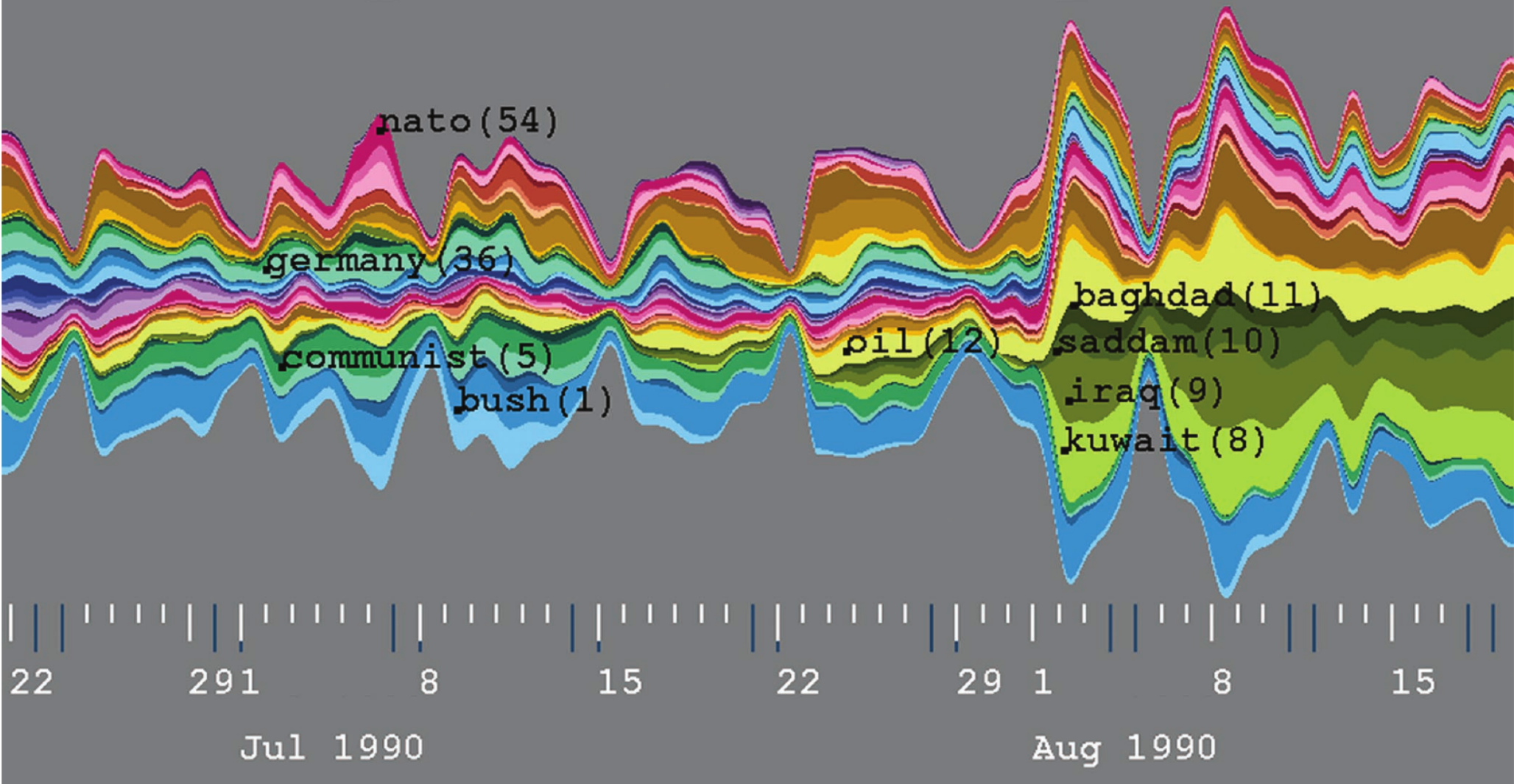
Jul 1990

Aug 1990

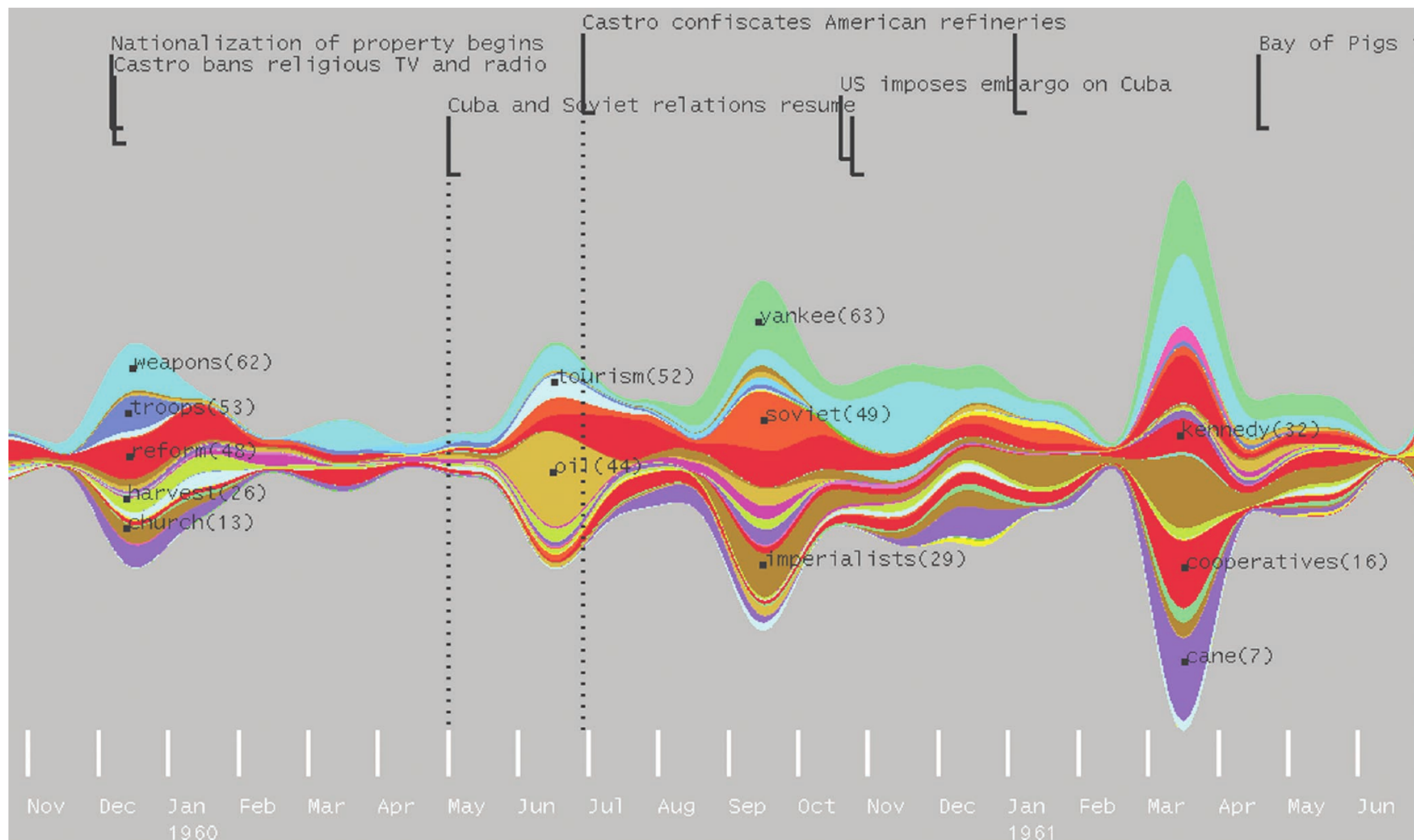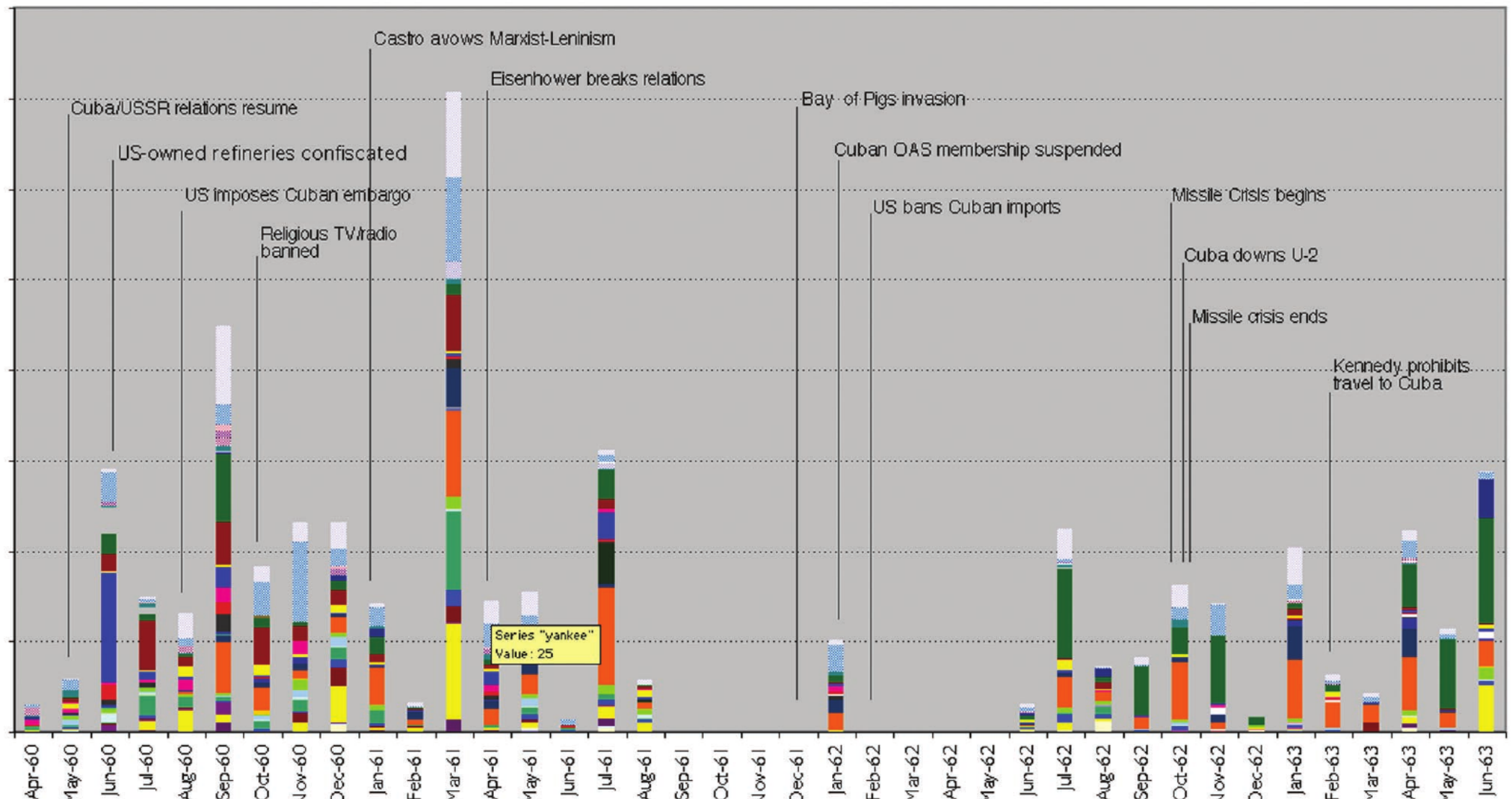Fig. 3. ThemeRiver visualization of Associated Press news wire stores from July and early August 1990.

Fig. 2. ThemeRiver uses a river metaphor to represent themes in a collection of Fidel Castro's speeches, interviews, and articles from the end of 1959 to mid-1961.

Castro avows Marxist-Leninism

Eisenhower breaks relations

Cuba/USSR relations resume

US-owned refineries confiscated

US imposes Cuban embargo

Religious TV/radio banned

Bay of Pigs invasion

Cuban OAS membership suspended

US bans Cuban imports

Missile Crisis begins

Cuba downs U-2

Missile crisis ends

Kennedy prohibits travel to Cuba

Series "yankee"
Value : 25

Apr-60 | May-60 | Jun-60 | Jul-60 | Aug-60 | Sep-60 | Oct-60 | Nov-60 | Dec-60 | Jan-61 | Feb-61 | Mar-61 | Apr-61 | May-61 | Jun-61 | Jul-61 | Aug-61 | Sep-61 | Oct-61 | Nov-61 | Dec-61 | Jan-62 | Feb-62 | Mar-62 | Apr-62 | May-62 | Jun-62 | Jul-62 | Aug-62 | Sep-62 | Oct-62 | Nov-62 | Dec-62 | Jan-63 | Feb-63 | Mar-63 | Apr-63 | May-63 | Jun-63

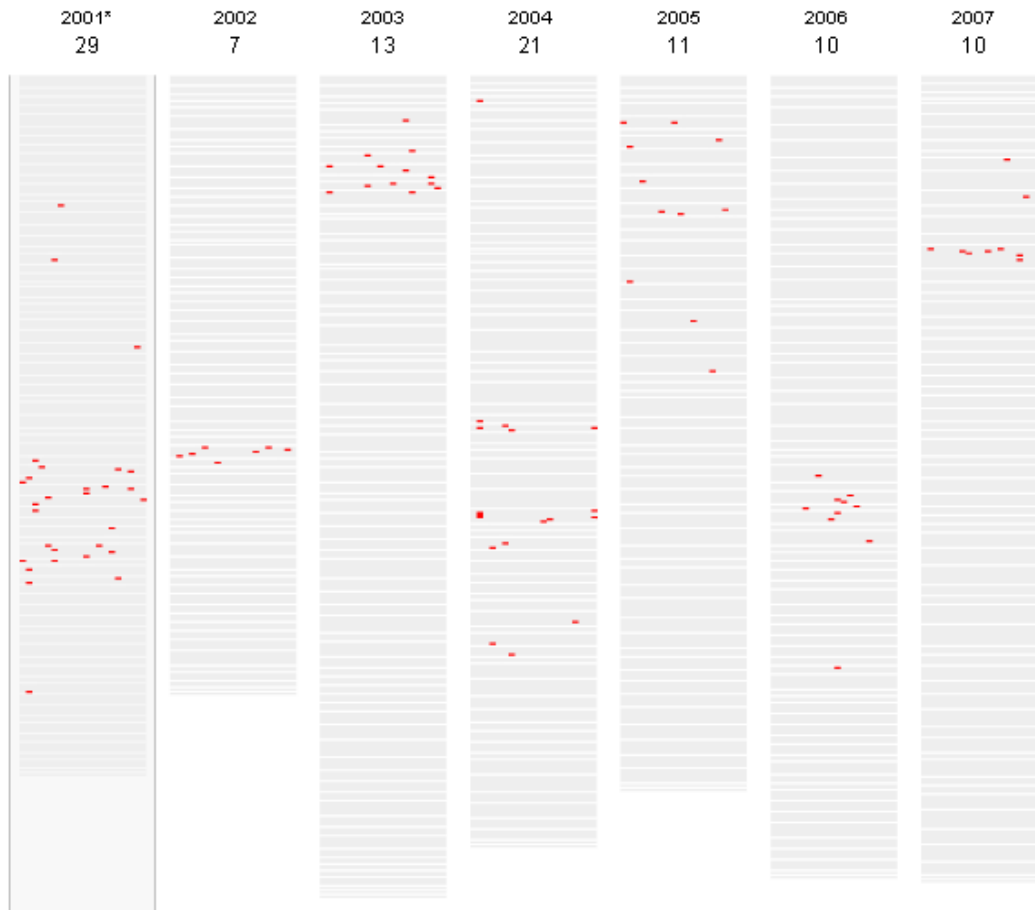| angola | anpp | athletes | beer | brazil | brezhnev | camaguey | cane | carter | cdr |
ceausescu	chernobyl	chile	church	clinton	coffee	cooperatives	czechoslovakia	debt	dengue
elections	fao	gdr	gorbachev	grenada	guinean	harvest	holguin	holstein	imperialists
jamaican	johnson	kennedy	krushchev	lard	mafia	mandela	manley	matanzas	mexico
minint	missiles	nicaragua	oas	oil	oriente	potatoes	reagan	reform	soviet
sugarcane	toure	tourism	troops	ubpc	ujc	vaccine	venezuelan	vietnam	vietnamese
vote	votes	weapons	yankee						

READ 2007 SPEECH | FEEDBACK

The 2007 State of the Union Address

Over the years, President Bush's State of the Union address has averaged almost 5,000 words each, meaning the the President has delivered over 34,000 words. Some words appear frequently while others appear only sporadically. Use the tools below to analyze what Mr. Bush has said.

[Search] [or choose a word here. ▾]

Use of the phrase "Tax" in past State of the Union Addresses

2001*	2002	2003	2004	2005	2006	2007
29	7	13	21	11	10	10

Compared with other words

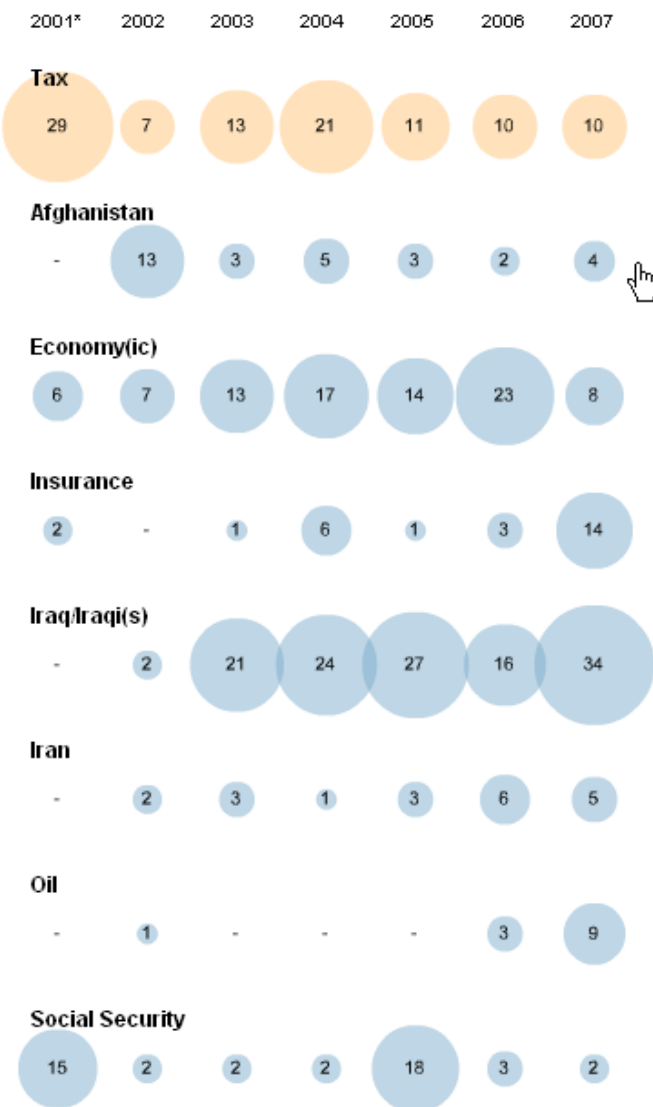|	2001*	2002	2003	2004	2005	2006	2007
Tax	29	7	13	21	11	10	10
Afghanistan	-	13	3	5	3	2	4
Economy(ic)	6	7	13	17	14	23	8
Insurance	2	-	1	6	1	3	14
Iraq/Iraqi(s)	-	2	21	24	27	16	34
Iran	-	2	3	1	3	6	5
Oil	-	1	-	-	-	3	9
Social Security	15	2	2	2	18	3	2

The word in context

Next Instance of 'Tax'

I believe in local control of schools. We should not, and we will not, run public schools from Washington, D.C. Yet when the federal government spends **TAX** dollars, we must insist on results. Children should be tested on basic reading and math skills every year between grades three and eight. Measuring is the only way to know whether all our children are learning. And I want to know, because I refuse to leave any child behind in America.

-- 2001 (Paragraph 14 of 73)

SeeSoft-style visualization

* As a newly elected president, Mr. Bush did not deliver a formal State of the Union address in 2001. His Feb. 27 speech to a joint session of Congress was analogous to the State of the Union, but without the title.

Ben Werschkul/The New York Times

Themeriver

- can get good overview with help of markers

Drawbacks:

- not much useful without prior knowledge on the topics/markers

- scalability of number of topics - reuse of colors

- less avenues for user interaction

  - high cost of processing

- evaluation with 2 users

Docuburst

- What if
  - tagclouds + lexical semantic hierarchy = ?

Wordnet

- ## Sets of synonyms
  - ### {bank, financial institution}
  - ### {bank, slope, riverside}

- ## Sets are connected through IS-A relationship
  {robin, redbreast} IS-A

  {bird} IS-A

  {animal, animate_being} IS-A

  {organism, life_form, living_thing} IS-A
  {entity}

games→game
taken→take

absolute,noun,10
chair,noun,2
moment,noun,11
game,noun,30
reality,noun,3
take,verb,13
represent,verb,17
...
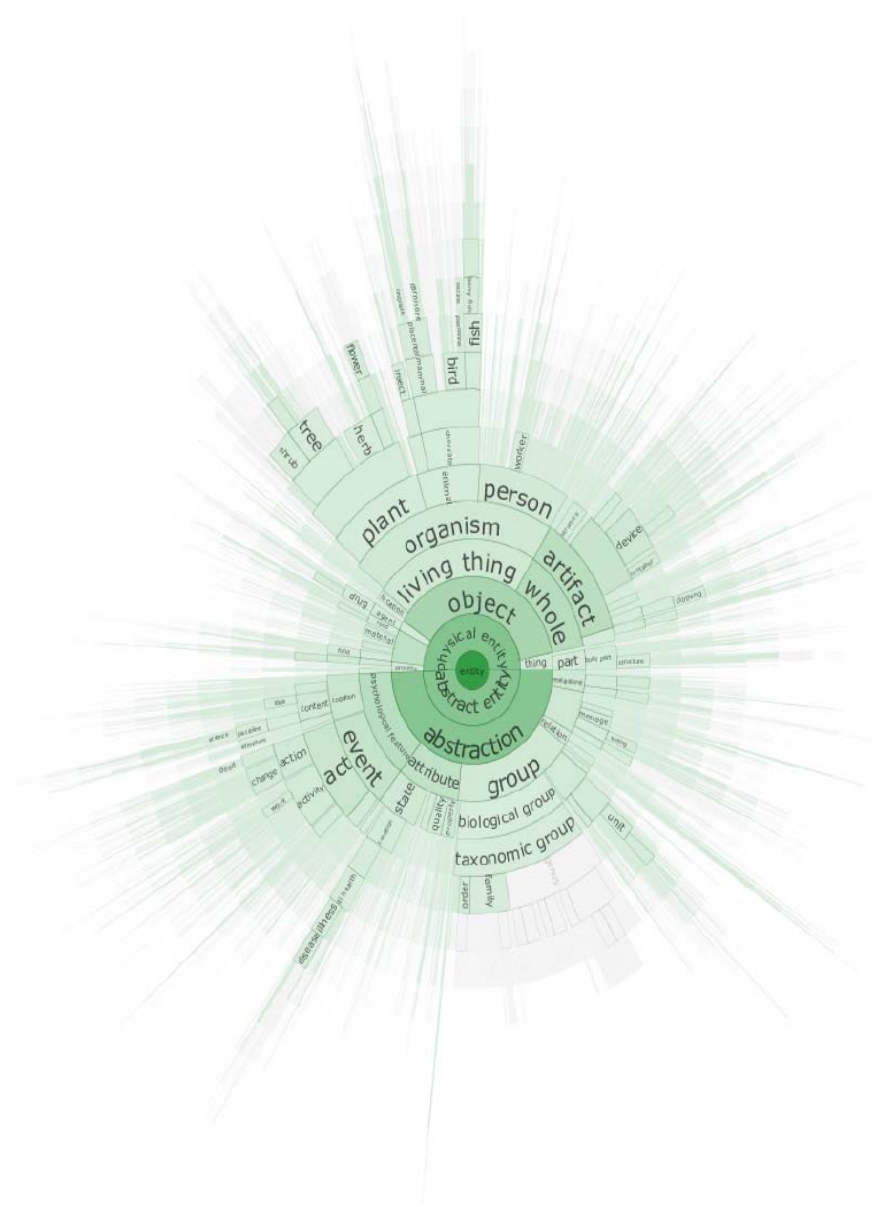
game IS activity
chair IS furniture

Source: Eurovis Docuburst presentation

Physical phenomenon sunburst diagram. Central node: physical phenomenon.

First ring categories: atmospheric phenomenon, electrical phenomenon, energy, force, optical phenomenon, pressure, conduction.

Major segments include:

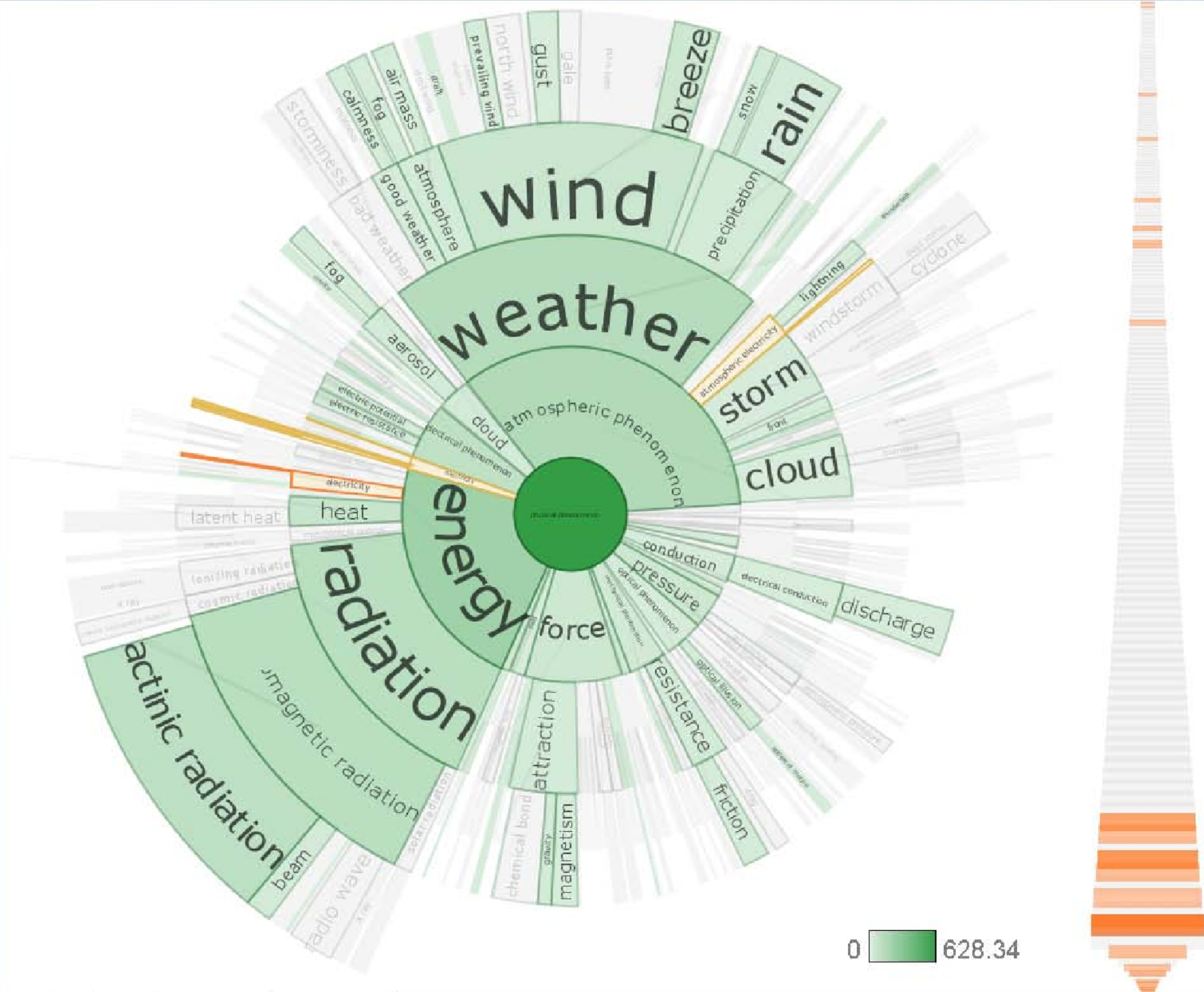- **weather**: wind, breeze, rain, snow, precipitation, atmosphere, good weather, bad weather, calmness, fog, air mass, north wind, prevailing wind, gust, gale, storminess
- **storm**: lightning, windstorm, cyclone
- **cloud**, **fog**, **aerosol**, **cloud**
- **energy**: heat, latent heat
- **radiation**: actinic radiation, beam, electromagnetic radiation, ionizing radiation, cosmic radiation, x ray, radio wave
- **electricity**: electric potential, electric resistance
- **force**: attraction, magnetism, chemical bond, gravity, resistance, friction
- **pressure**, **conduction**: electrical conduction, discharge
- **optical phenomenon**

Color scale: 0 ——— 628.34

Docuburst

- Answers the question: 'What is this document about ?'

- Goal: to bring the exploratory experience of physical libraries to digital libraries

  - Idea of small multiples: library interfaces could be enhanced with arrays of Docuburst icons

- Drawbacks:

  - Higher cognitive load, as well as system load

Conclusion

- Nominal, categorical variables
  - difficult to visualize
    - no inherent ordering
- Text
  - categorical nature +
  - high dimensionality
  - = challeging to visualize

References

- DocuBurst: Visualizing Document Content using Language Structure. Christopher Collins, Sheelagh Carpendale, Gerald Penn. EuroVis 2009.

- Participatory Visualization with Wordle. Fernanda B. Viegas, Martin Wattenberg, Jonathan Feinberg. InfoVis 2009.

- ThemeRiver: Visualizing Thematic Changes in Large Document Collections Susan Havre, Elizabeth Hetzler, Paul Whitney, Lucy Nowell. IEEE Transactions on Visualization and Computer Graphics, 8(1), pp 9-20, 2002.

- SEARCH USER INTERFACES, PUBLISHED BY CAMBRIDGE UNIVERSITY PRESS. COPYRIGHT © 2009 BY MARTI A. HEARST.
  http://searchuserinterfaces.com/book