

Hands-On: Visualization

Tamara Munzner
 Department of Computer Science
 University of British Columbia
 IRS Cross-Border Reporting Workshop
 May 2017, Vancouver BC

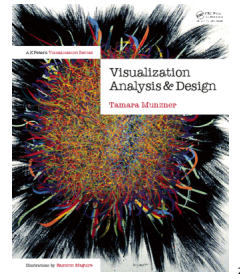
<http://www.cs.ubc.ca/~tmm/talks.html#cb17> @tamaramunzner

Visualization (vis) defined & motivated

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- human in the loop needs the details
 - doesn't know exactly what questions to ask in advance
 - exploration: longterm exploratory data analysis (EDA)
 - presentation: known results
 - stepping stone towards automation: refining, trustbuilding
- external representation: perception vs cognition
- intended task, measurable definitions of effectiveness



more at:
 Visualization Analysis and Design, Chapter 1.
 Munzner. AK Peters Visualization Series, CRC Press, 2014.

Two vignettes

- a tale of two tools created for journalistic use
 - shared frameworks of interdisciplinary methods from my research group
 - thinking about collaboration
 - roles & rewards, for computer scientists & journalists
 - reasoning about visualization design
 - beyond pretty pictures
 - divergent goals & audiences
 - TimeLineCurator: presentation / explanatory
 - Overview: investigation / exploratory

Collaboration incentives

- why do CS/vis people need to understand journalism's problems?
 - we work with you to understand your driving problems
 - we build tools intended to help
 - only works out if we understood the problems deeply enough
 - we observe how you use them
 - if they're good enough
 - CS win: research success stories
 - journalist win: access to better tools
 - we develop guidelines on how to build better tools in general
 - CS win: research progress in visualization

Vignette 1: Vis Tool for Journalistic Presentation

Johanna Fulda @jofu_

Matthew Brehmer @mattbrehmer

Tamara Munzner @tamaramunzner

TimeLineCurator
 Interactive Authoring of Visual Timelines from Unstructured Text

<http://about.timelinecurator.org>
<http://timelinecurator.org>

TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text.
 Fulda, Brehmer, Munzner. IEEE Trans. Visualization and Computer Graphics (Proc IEEE VAST 2015) 22(1):300-309, 2015.

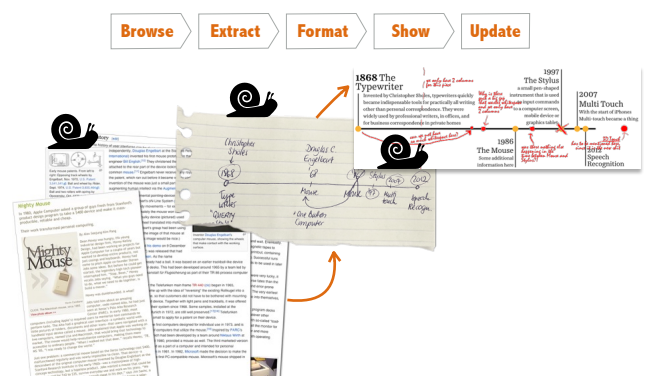
Origin story: Tedium in the newsroom

- Johanna Fulda: interactive infographics developer, Sueddeutsche Zeitung
 - then Munich CS master's student, visiting UBC
- what pain point could we address with interactive visualization?
 - plus some natural language processing (NLP)



<https://vimeo.com/jofu/tlc>

Manual creation process



Structured creation process



Timeline authoring model

time required for each task

	Browse	Extract	Format	Show	Update
Manual Drawing	slow	slow	slow	slow	slow
Structured Creation	slow	slow	slow	automated	fast
TimeLine Curator	fast	automated	automated	fast	fast

The general case for curation

- build for human in the loop as continuing need
 - automatic processing to accelerate not replace
 - assume computational results good but not perfect
 - for the indefinite future!
 - visual feedback to accelerate

Architecture

Extract: Recognition, Normalization

Format: Data set generation

Show: Timeline view

Curate/Update: Search and filter options

Present: Final timeline output

The importance of being brisk

- sexy use case: eureka moment
 - enable what was impossible before
 - vis tools for new insights & discoveries
- workhorse use case: workflow speedup
 - vis tools to accelerate what you're already doing
 - sometimes enables the previously infeasible
- TLC use cases
 - started with speedup use case, for presentation
 - make this doc into a timeline now!
 - two other use cases nudge towards exploration
 - comparison between multiple timelines
 - speculative browsing



TimeLineCurator: Speculative Browsing

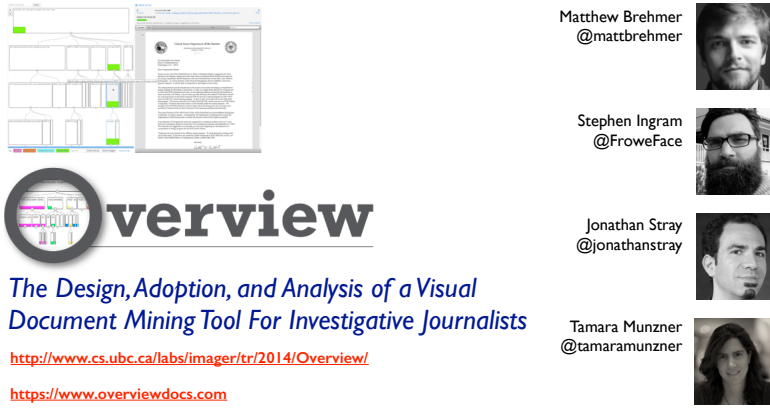
speculative browsing

<https://vimeo.com/jofu/tlc>

Try it out!

<http://timelinecurator.org>

Vignette 2: Vis Tool for Investigative Reporting

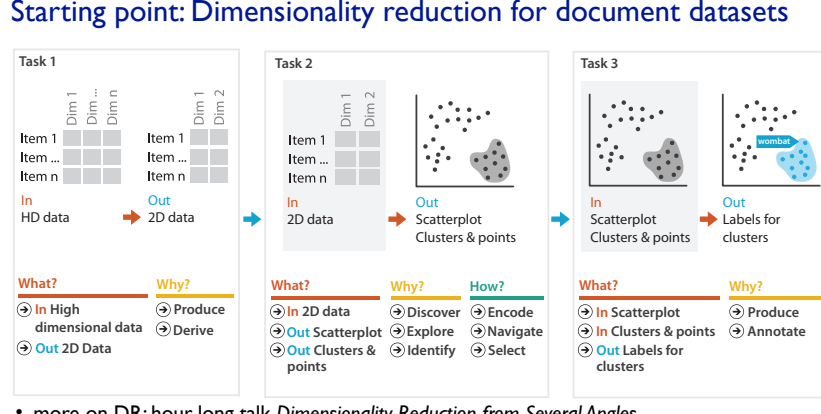


Overview
The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists

<http://www.cs.ubc.ca/labs/imager/tr/2014/Overview/>
<https://www.overviewdocs.com>

Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool For Investigative Journalists. Brehmer, Ingram, Stray, and, Munzner. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2014), 20(12):2271-2280, 2014.

Matthew Brehmer @mattbrehmer
Stephen Ingram @FroweFace
Jonathan Stray @jonathanstray
Tamara Munzner @tamaramunzner

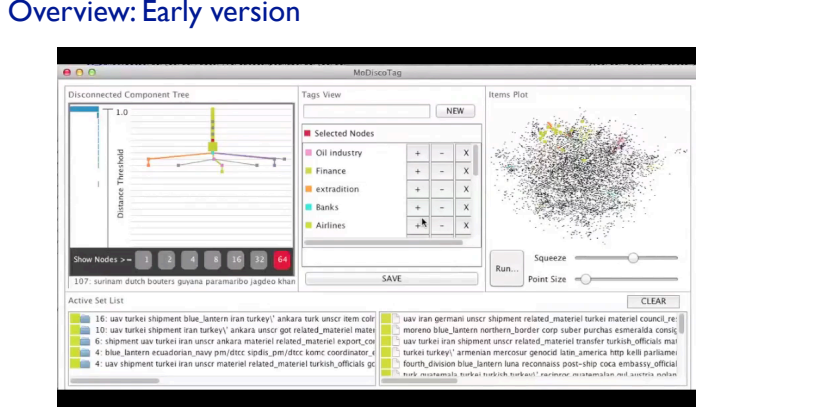


Origin story: WikiLeaks meets Glimmer

- WikiLeaks: hacker-journalist Jonathan Stray analyzing Iraq warlogs
 - one instance of general problem: Too Many Documents
 - conjectured that existing label classification falls short of showing all meaningful structure in data
 - friendly action, criminal incident, ...
 - he had some NLP, needed better vis tools
- Glimmer: multilevel dimensionality reduction algorithm
 - scalability to 30K documents and terms

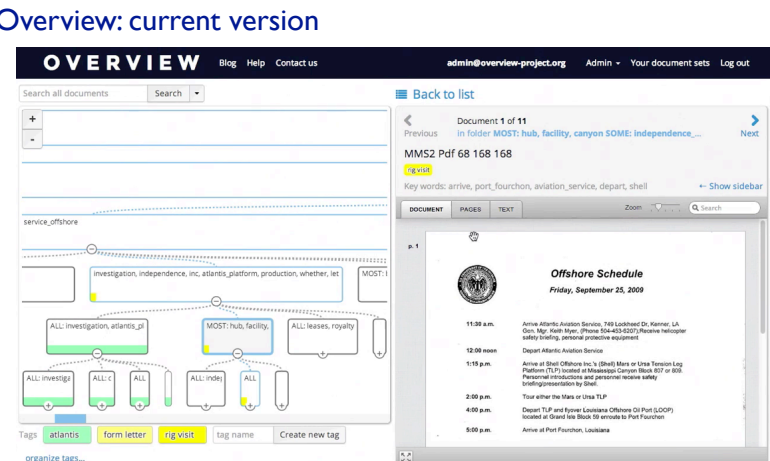
[Glimmer: Multilevel MDS on the GPU. Ingram, Munzner, Olano. IEEE TVCG 15(2):249-261, 2009.]

Overview: Early version



<http://www.cs.ubc.ca/labs/imager/tr/2012/modiscotag>

Overview: current version



Deploy in the real world

Case Study	#1	#2	#3	#4	#5	#6
Document Collection	4,500 pages from FOIA	5,996 emails from FOIA	8,680 pages from FOIA	1,278 survey comments	4,653 emails from FOIA	1,680 bills
Question	What did security contractors do during Iraq war?	Were municipal police funds mismanaged?	Were Paul Ryan's campaign statements hypocritical?	What is the gun ownership debate about?	Was gov't response to emergency incident effective?	Did gov't fail to pass bills addressing police misconduct?

Deploy in the real world, understand user goals

Case Study	#1	#2	#3	#4	#5	#6
Document Collection	4,500 pages from FOIA	5,996 emails from FOIA	8,680 pages from FOIA	1,278 survey comments	4,653 emails from FOIA	1,680 bills
Question	What did security contractors do during Iraq war?	Were municipal police funds mismanaged?	Were Paul Ryan's campaign statements hypocritical?	What is the gun ownership debate about?	Was gov't response to emergency incident effective?	Did gov't fail to pass bills addressing police misconduct?
Insight	the trend story: getting the big picture	find the needle in the haystack / smoking gun	wheat vs chaff: filtering out irrelevant material	categorize and count: turning docs into data	wheat vs chaff: filtering out irrelevant material	prove haystack contains no needles!

<https://blog.overviewdocs.com/2014/01/09/what-do-journalists-do-with-documents-the-different-kinds-of-document-driven-stories/>

Evolution across levels

- evolution of task abstraction
 - task 1: generate hypotheses → explore → summarize
 - obviously you can't read everything; speed up with tool for categorizing and counting
 - task 2: verify hypotheses → locate → identify
 - you really do read each doc; speed up with tool to keep track of findings
- evolution of data abstraction & idioms
 - arrange cluster tree to emphasize nodes vs links
 - new vis insight: DR scatterplot less effective than cluster tree vis + tagging
 - better affordance for systematic traversal of document cluster hierarchy

early → current

Try it out!

<https://www.overviewdocs.com/>

<https://blog.overviewdocs.com/completed-stories/>

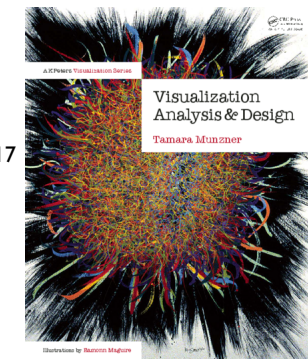
Discussion

- how many of you have jumped into data journalism?
 - what kind of tools have you used?
- have any of you grappled with timelines?
 - what kind of tools did you use?
- have any of you grappled with large document collections?
 - what kind of story did you have in mind?

More Information

- this talk
www.cs.ubc.ca/~tmm/talks.html#cb17
- book
<http://www.cs.ubc.ca/~tmm/vadbook>
– 20% off promo code, book+ebook combo: HVN17
– <http://www.crcpress.com/product/isbn/9781466508910>
- papers, videos, software, talks, courses
<http://www.cs.ubc.ca/group/infovis>
<http://www.cs.ubc.ca/~tmm>

@tamaramunzner



Visualization Analysis and Design. Munzner. A K Peters Visualization Series, CRC Press, Visualization Series, 2014.