

APPENDIX of How we draw texts

A literature review on text visualization tools*

* This work is licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License.

APPENDIX

0.1 Single text visualization

0.1.1 Whole text visualization

1 ▷ Novel Views: Les Misérables - Radial Word Connections by Jeff Clark (2013)

- data: Novel "Les Misérables"
- method: Radial text-line
- description: "A word used in multiple places in a text can be interpreted as a connection between those locations. Depending on the word itself the connection could be in terms of character, setting, activity, mood, or other aspects of the text."

[<http://neoformix.com/2013/NovelViews.html>]

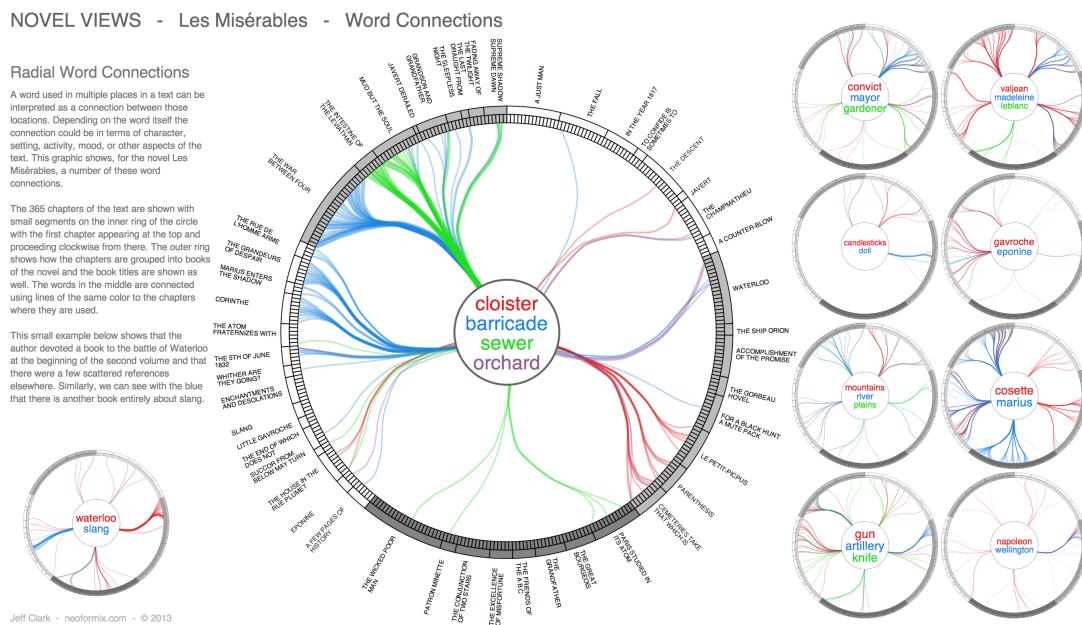


Figure 0.1: Novel Views: Les Misérables - Radial Word Connections by Jeff Clark (2013)

0.1. SINGLE TEXT VISUALIZATION

2▷ Novel Views: Les Misérables - Character Mentions by Jeff Clark (2013)

- data: Novel "Les Misérables".
- method: Horizontal multi text-line.
- description: "it shows where the names of the primary characters are mentioned within the text. Click on any of these images to see larger versions."

[<http://neoformix.com/2013/NovelViews.html>]

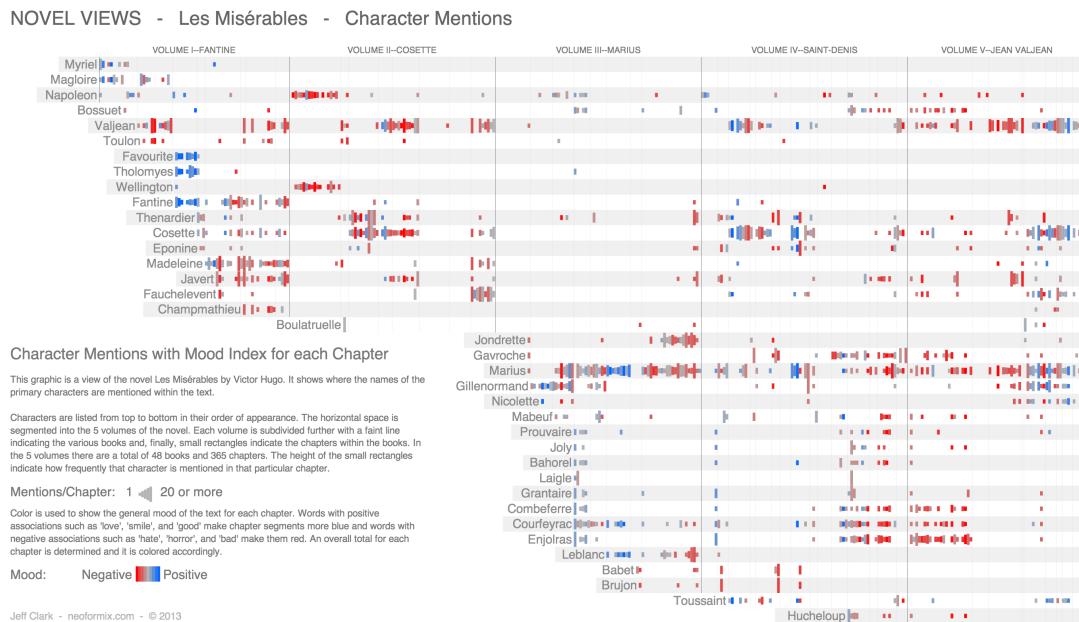


Figure 0.2: Novel Views: Les Misérables - Character Mentions by Jeff Clark (2013)

0.1. SINGLE TEXT VISUALIZATION

3▷ Poem Viewer - (on process) 2013 project Imagery Lens for Visualizing Text Corpora - Oxford e-Research Centre at the University of Oxford by Katharine Coles, Min Chen, Alfie Abdul-Rahman, Chris Johnson, Julie Lein, Eamonn Maguire, Miriah Meyer, and Martin Wynne. (2013)

- data: Poems.
- method: Advanced Text-line.
- description: Poem Viewer is an experimental service for the exploration and analysis of poetry through visualization. Poem Viewer is part of an on-going research project and is currently a work in progress.

[<http://ovii.oerc.ox.ac.uk/PoemVis/>]

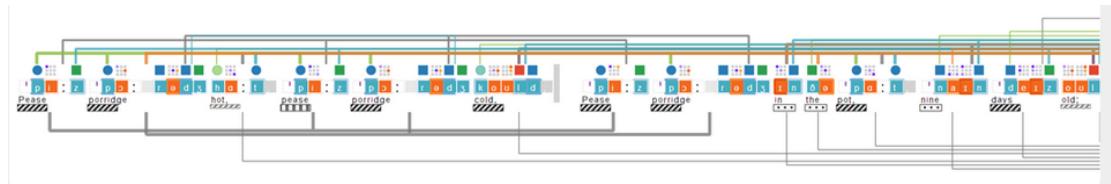


Figure 0.3: Poem Viewer - (on process) 2013 project Imagery Lens for Visualizing Text Corpora - Oxford e-Research Centre at the University of Oxford by Katharine Coles, Min Chen, Alfie Abdul-Rahman, Chris Johnson, Julie Lein, Eamonn Maguire, Miriah Meyer, and Martin Wynne. (2013)

0.1. SINGLE TEXT VISUALIZATION

4▷ State of the Union 2011 - Sentence Bar Diagrams by Jeff Clark (2011)

- data: Speech to text. Obama's State of the Union speech 2011 [N/A]
- method: Sentence Bar Diagrams.
- description: "First we have two Sentence Bar Diagrams for the speeches from 2010 and 2011. Sentence Bar diagrams use color coding to show the topic of the various sentences in the text and bar length to show how long the sentences are."

[<http://neoformix.com/2011/SOTU2011.html>]



Figure 0.4: State of the Union 2011 - Sentence Bar Diagrams by Jeff Clark (2011)

5 ▷ Visualizing Lexical Novelty in Literature by Matthew Hurst (2011)

- data: Literature
- method: Original pixel-block method
- description: Tracking the introduction of new terms in a novel. In the visualization each column represents a chapter and each small block a paragraph of text. The color of the block indicates the % of new words. It is not clear what is the utility of this tool; an specialist in linguistics could say how this visualization can be used.

[http://datamining.typepad.com/data_mining/2011/09/visualizing-lexical-novelty-in-literature.html]

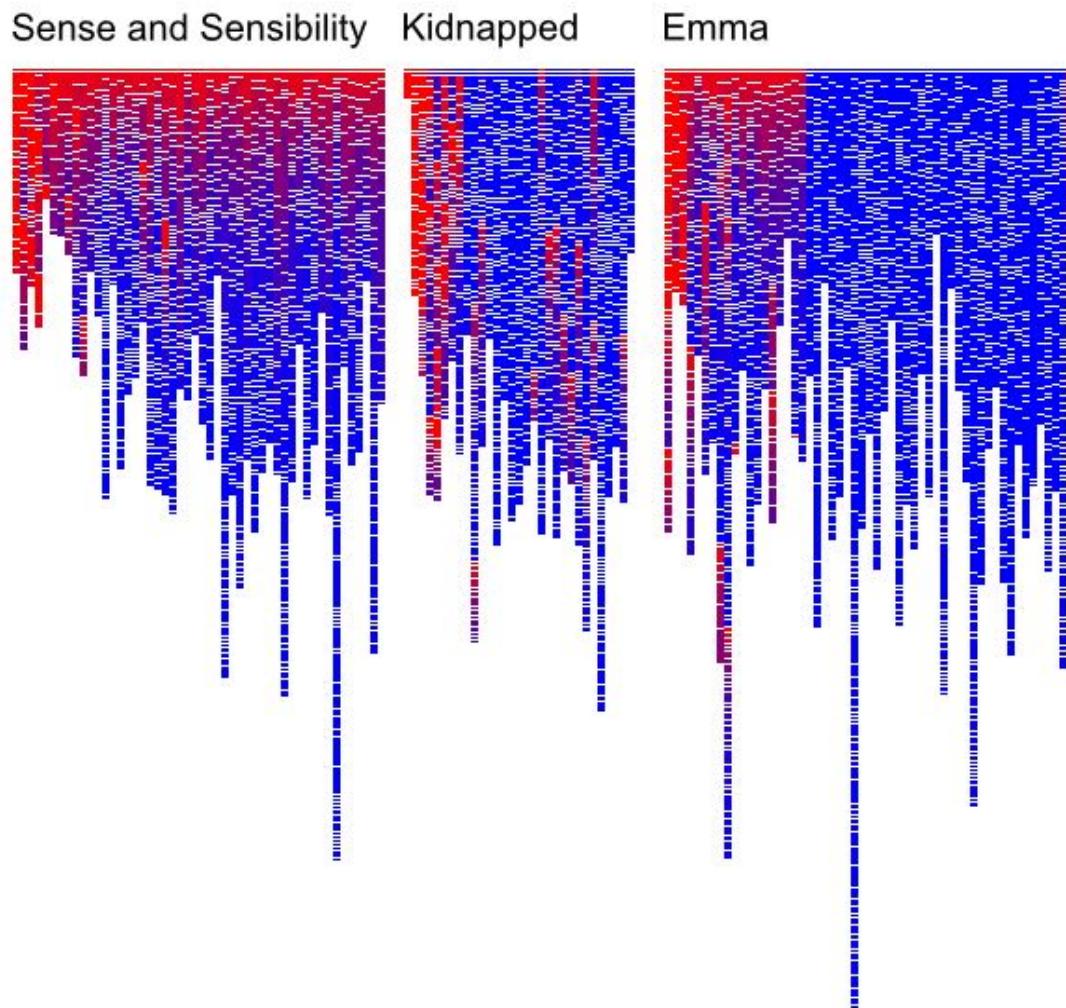


Figure 0.5: Visualizing Lexical Novelty in Literature by Matthew Hurst (2011)

0.1. SINGLE TEXT VISUALIZATION

6▷ On the Origin of Species: The Preservation of Favoured Traces by Ben Fry (2009)

- data: All the versions of Darwin's On the Origin of Species Book.
- method: Matrix of pixel-blocks representing all the text in chunks. Colors are used to show each edition additions.
- description: We often think of scientific ideas, such as Darwin's theory of evolution, as fixed notions that are accepted as finished. In fact, Darwin's On the Origin of Species evolved over the course of several editions he wrote, edited, and updated during his lifetime. The first English edition was approximately 150,000 words and the sixth is a much larger 190,000 words. In the changes are refinements and shifts in ideas whether increasing the weight of a statement, adding details, or even a change in the idea itself.

[<http://benfry.com/traces/>]

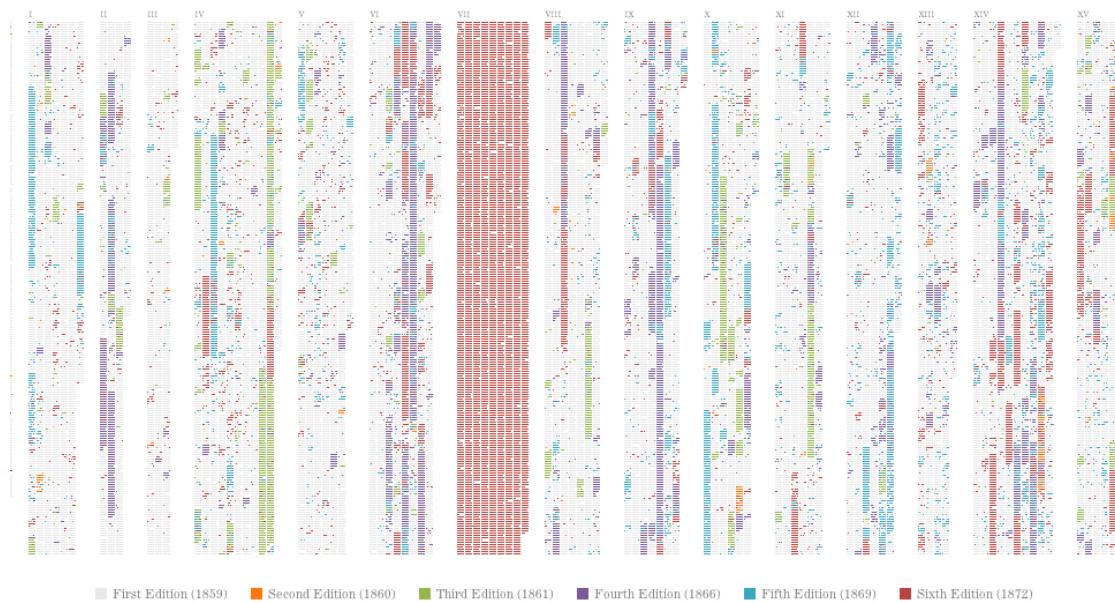


Figure 0.6: On the Origin of Species: The Preservation of Favoured Traces by Ben Fry (2009)

0.1. SINGLE TEXT VISUALIZATION

7 ▷ **Texty, a visualization tool to aid selection of texts from search outputs by Jaume Nualart (2008)**

- data: Ars Electronica Jury Statements (from 1987 to 2007)
- method: Iconic representation of a text
- description: a Texty is an image, an icon that represents the physical distribution of keywords of a text as a flat image. These keywords are grouped in vocabularies, to each of which a color is linked. Texty reveals, the structure, conceptual density and subject matter of a text.
- paper: *Nualart, J. Pérez-Montoro, M (2013). "Texty, a visualization tool to aid selection of texts from search outputs" Information Research.*

[<http://vis.mediaartresearch.at/textass/texty.php>]

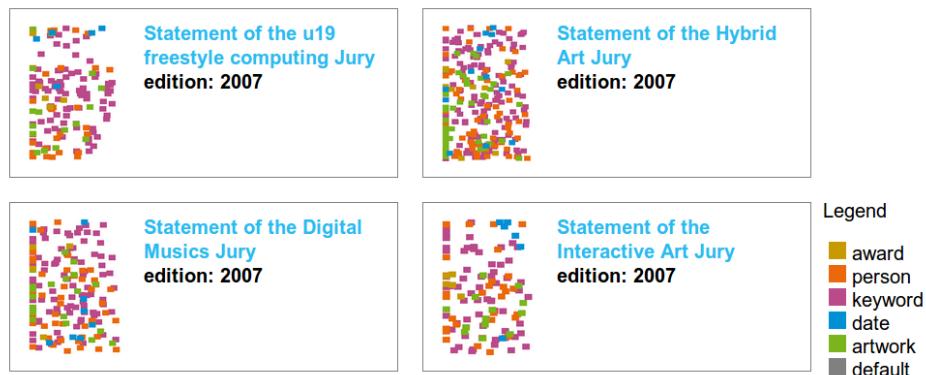


Figure 0.7: Detail of four textys showing texts and the legend from Ars Electronica Jury Statements (2008)

8▷ Bible Cross-References by Chris Harrison (2008)

- data: Bible [Large dataset]
- method: Arc graph + bar graph
- description: "The bar graph that runs along the bottom represents all of the chapters in the Bible. Books alternate in color between white and light gray. The length of each bar denotes the number of verses in the chapter. Each of the 63,779 cross references found in the Bible is depicted by a single arc - the color corresponds to the distance between the two chapters, creating a rainbow-like effect."

[<http://www.chrisharrison.net/index.php/Visualizations/BibleViz>]

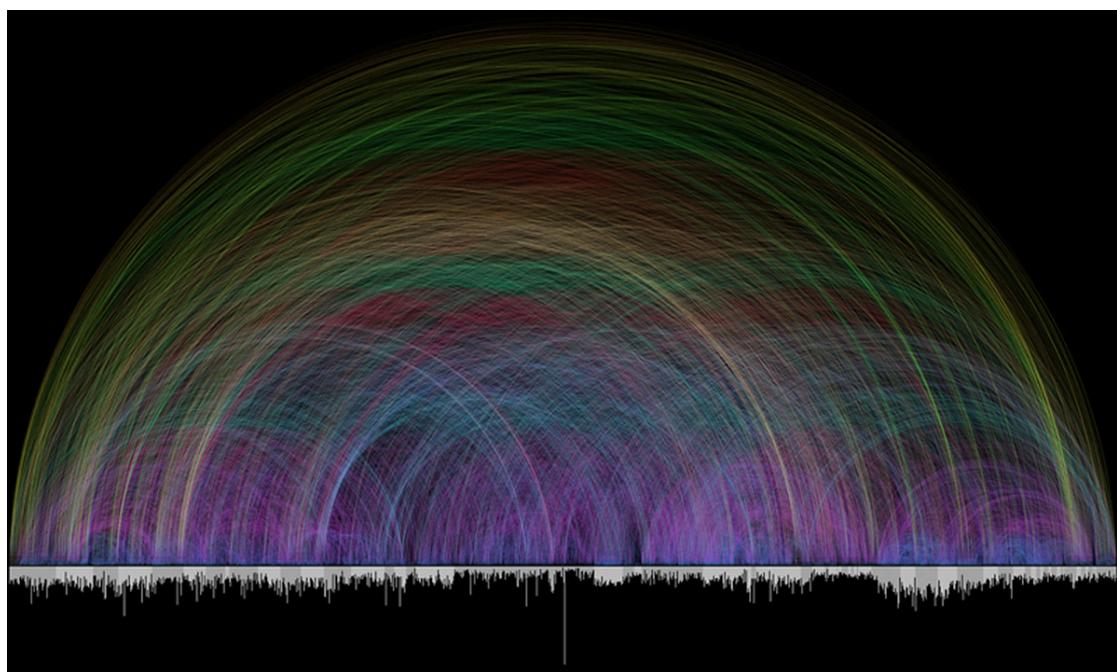


Figure 0.8: Bible Cross-References by Chris Harrison (2008)

9▷ Literature fingerprint by Daniel A. Keim and Daniela Oelke (2007)

- data: Any text, from middle to long
- method: Colored squares representing elements of a text in a sequential way.
- description: Colored squares representing elements of a text in a sequential way.
Description/Comments: the authors develop visualization techniques to show the results of some linguistic analysis that represent a literature fingerprint, demonstrating text authorship.
- paper: *Keim D. A. & Oelke D. (2007). Literature Fingerprinting: A New Method for Visual Literary Analysis. In: Proceedings of the 2007 IEEE Symposium on Visual Analytics Science and Technology (VAST '07). IEEE Computer Society, Washington, DC, USA, 115-122. DOI=10.1109/VAST.2007.4389004 http://dx.doi.org/10.1109/VAST.2007.4389004*

▷ [http://bib.dbvis.de/uploadedFiles/71.pdf]

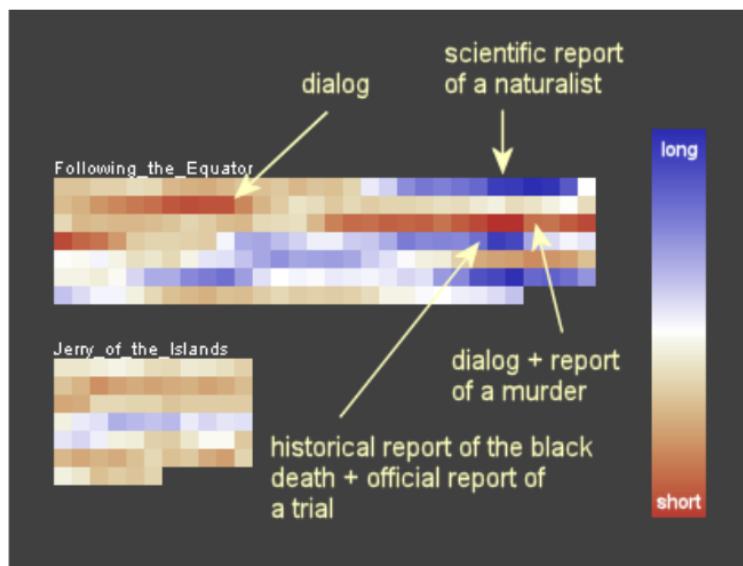


Figure 3: The figure shows the fingerprints of two novels that almost have the same average sentence length. In the detailed view, the different structure of the two novels is revealed. The inhomogeneity of the travelogue *Following the Equator* can be explained with the alternation of dialogs, narrative parts and quoted documents.

Figure 0.9: Sentence length visualization

10 ▷ History Flow by Fernanda Viégas and Martin Wattenberg (2003)

- data: Wikipedia [Small dataset]
- method: Flow chart time-line
- description: “the history flow application charts the evolution of a document as it is edited by many people using a very simple visualization technique. All text segments contributed by the same editor are marked with a unique color. The diagram shows how some editors produce long lasting content, while others don’t. With enough editors contributing to an article, almost every paragraph or even sentence gets modified in the long run.”
- paper: *Viégas, F. B., Wattenberg, M., & Dave, K. (2004, April). Studying cooperation and conflict between authors with history flow visualizations. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 575-582). ACM.* [http://alumni.media.mit.edu/~fviegas/papers/history_flow.pdf]

[<http://hint.fm/projects/historyflow/>]

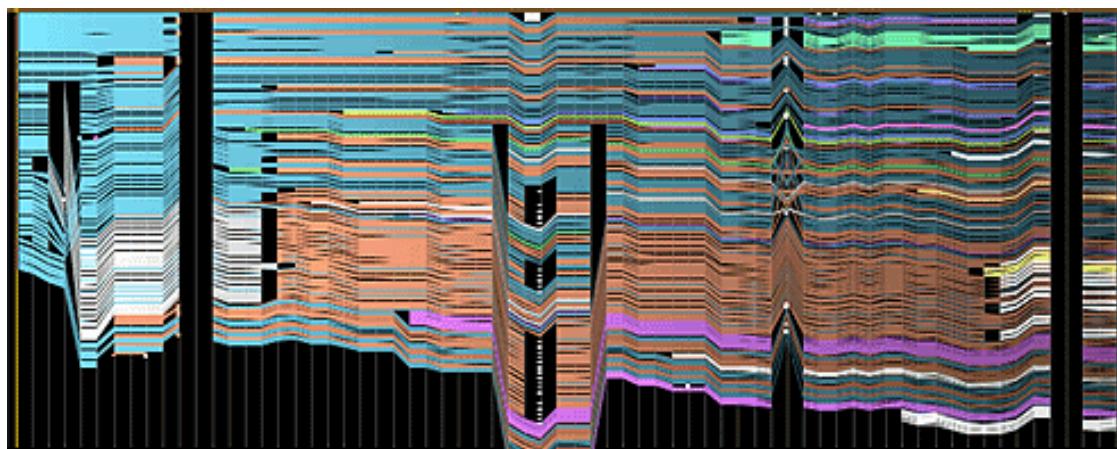


Figure 0.10: History Flow by Fernanda Viégas and Martin Wattenberg researchers at IBM’s Visual Communication Lab (2003)

0.1. SINGLE TEXT VISUALIZATION

11 ▷ Colour-coded chronological sequencing by Joel Deshayé and Peter Stoicheff (2003)

- data: The Sound and the Fury by William Faulkner
- method: Timeline
- description: Case collected by Peter Stoicheff: "This visualization (contributed by Joel Deshayé and Peter Stoicheff) shows a compressed version of the colour-coded "April Seventh, 1928" narrative on the left. The middle bar extracts the narrative of Benjy's present day. The right bar extracts the flashbacks to Caddy's wedding. Although the section seems randomly ordered, within it the present and each flashback reside independently as coherent, chronological sequences."
- paper: *Stoicheff, R.P. "Faulkner's Foreign Levy: Macbeth ,The Sound and the Fury, and Writerhood." The Sound and the Fury: a Hypertext Edition. Ed. Stoicheff, Muri, Deshayé, et al. Updated Mar. 2003. U of Saskatchewan. Accessed 18 Mar. 2003 <http://www.usask.ca/english/faulkner>*

[http://drc.usask.ca/projects/faulkner/main/benjy_spectrum.htm]

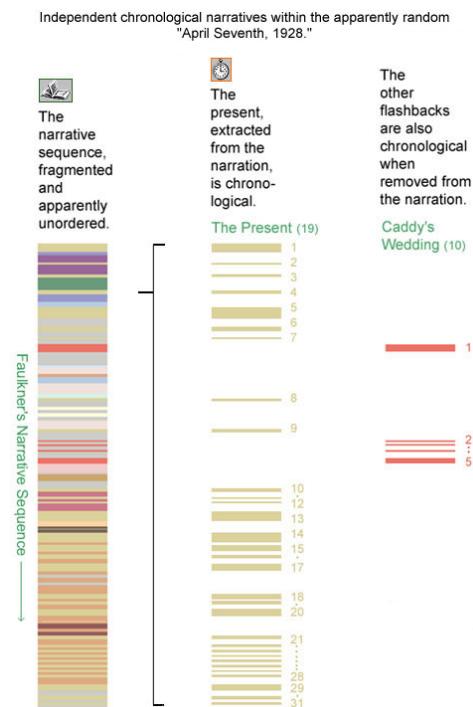


Figure 0.11: Colour-coded chronological sequencing of William Faulkner's novel The Sound and the Fury, by Joel Deshayé and Peter Stoicheff (2003)

12 ▷ 2-D display of time in the novel by Joel Deshaye (2003)

- data: The Sound and the Fury by William Faulkner
- method: Customized 2-D diagram
- description: Case collected by Peter Stoicheff: "This graph (contributed by Joel Deshaye) shows two dimensions of time in The Sound and the Fury: chronological time, and Faulkner's re-ordering of chronological time into the text's narrative sequence. Faulkner scrambles the chronology not only by flashbacks, but also by the non-linear sequence of the novel's sections. Through this graph it becomes clear that the novel follows a conventional *in medias res* structure, whose existence is otherwise obscured by more local narrative complexities."
- paper: *Stoicheff, R.P. "Faulkner's Foreign Levy: Macbeth ,The Sound and the Fury, and Writerhood." The Sound and the Fury: a Hypertext Edition. Ed. Stoicheff, Muri, Deshaye, et al. Updated Mar. 2003. U of Saskatchewan. Accessed 18 Mar. 2003 <http://www.usask.ca/english/faulkner>*

[http://drc.usask.ca/projects/faulkner/main/sf_2d_timegraph.htm]

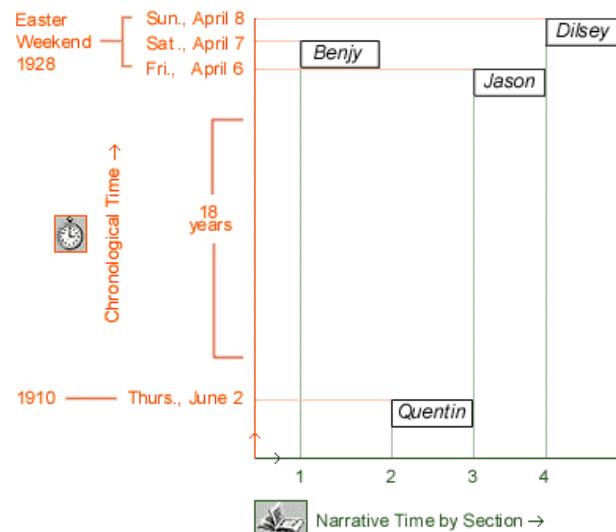


Figure 0.12: 2-D display of time of William Faulkner's novel *The Sound and the Fury*, by Joel Deshaye and Peter Stoicheff (2003)

13 ▷ 3-D display of time in the novel by Joel Deshaye (2003)

- data: The Sound and the Fury by William Faulkner
- method: Highly customized 3D diagram
- description: Case collected by Peter Stoicheff: "This graph (contributed by Joel Deshaye) represents time in three different dimensions. First, there is the sequence of four sections in the novel - the narrative time or *récit* (shown in green). Second, there is the chronology, *l'histoire* (shown in orange), beginning with the earliest recollected date in the novel - Damuddy's death in 1898. These two dimensions show how the conventional view of linear time can be disrupted by a fictional narrative. Third, there is a representation of the proportion of memories of and flashbacks to the past in each section (shown in blue). This last dimension shows how the novel progresses from an emphasis on the past toward an emphasis on the present."
- paper: *Stoicheff, R.P. "Faulkner's Foreign Levy: Macbeth ,The Sound and the Fury, and Writerhood." The Sound and the Fury: a Hypertext Edition. Ed. Stoicheff, Muri, Deshaye, et al. Updated Mar. 2003. U of Saskatchewan. Accessed 18 Mar. 2003 <http://www.usask.ca/english/faulkner>*

[http://drc.usask.ca/projects/faulkner/main/sf_3d_timegraph.htm]

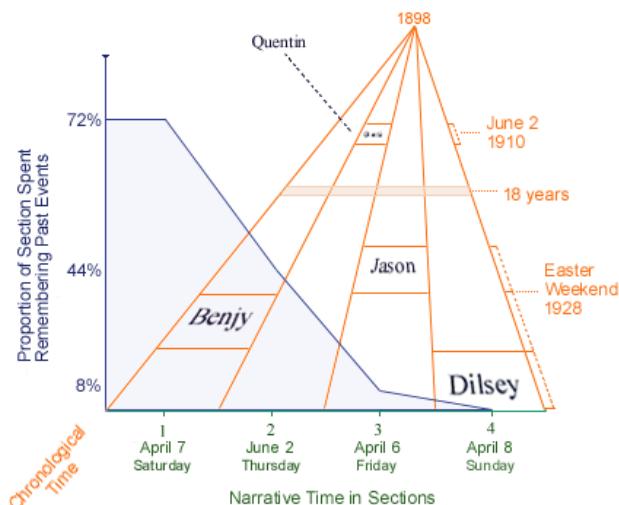


Figure 0.13: 3-D display of time of William Faulkner's novel The Sound and the Fury, by Joel Deshaye and Peter Stoicheff (2003)

14 ▷ Arc diagram Wattenberg by Martin Wattenberg (2002)

- data: Any text or string [Small dataset]
- method: Arc diagram
- description: Arc diagram is capable of representing complex patterns of repetition in string data. Arc diagrams improve over previous methods such as dotplots because they scale efficiently for strings that contain many instances of the same subsequence.
- paper: *Wattenberg, M. (2002). Arc diagrams: Visualizing structure in strings. In Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on (pp. 110-116). IEEE.*

▷ [[http://domino.watson.ibm.com/cambridge/research.nsf/58bac2a2a6b05a1285256b30005b3953/e2a83c4986332d4785256ca7006cb621/\\$FILE/TR2002-11.pdf](http://domino.watson.ibm.com/cambridge/research.nsf/58bac2a2a6b05a1285256b30005b3953/e2a83c4986332d4785256ca7006cb621/$FILE/TR2002-11.pdf)]



Figure 0.14: Arc diagram by Martin Wattenberg (2002)

0.1. SINGLE TEXT VISUALIZATION

15 ▷ TileBars: Visualization of Term Distribution Information in Full Text Information Access by Marti A. Hearst (1995)

- data: search results from information retrieval systems. Applied to PubMed results.
- method: iconic representation of a text
- description: “TileBars, which provides a compact and informative iconic representation of the documents’ contents with respect to the query terms. The goal is to simultaneously indicate: the relative length of the document, the frequency of the term sets in the document, and the distribution of the term sets with respect to the document and to each other. Each large rectangle indicates a document, and each square within the document represents a TextTile. The darker the tile, the more frequent the term (white indicates 0, black indicates 8 or more instances)”
- paper: Hearst, M. 1995. “TileBars: visualization of term distribution information in full text information access.” Proceedings of the SIGCHI conference on Human http://dl.acm.org/citation.cfm?id=223912 (March 26, 2013).

[http://dl.acm.org/citation.cfm?id=223912]

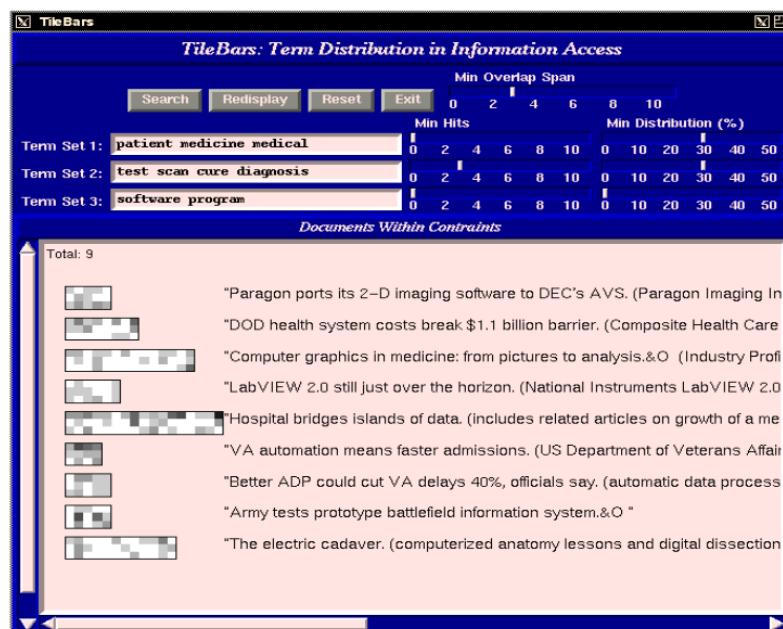


Figure 0.15: TileBar search on (patient medicine medical AND test scan cure diagnosis AND software program) with stricter distribution constraints.

0.1. SINGLE TEXT VISUALIZATION

0.1.2 Part text visualization

16▷ Novel Views: Les Misérables - Characteristic Verbs by Jeff Clark (2013)

- data: Novel "Les Misérables" [N/A]
- method: Tables + condensed bar graphs
- description: "the verbs used together with character names in a novel can provide a glimpse into the personalities and actions of that character. For the primary people in the novel Les Misérables this graphic illustrates their characteristic verbs."

[<http://neoformix.com/2013/NovelViews.html>]

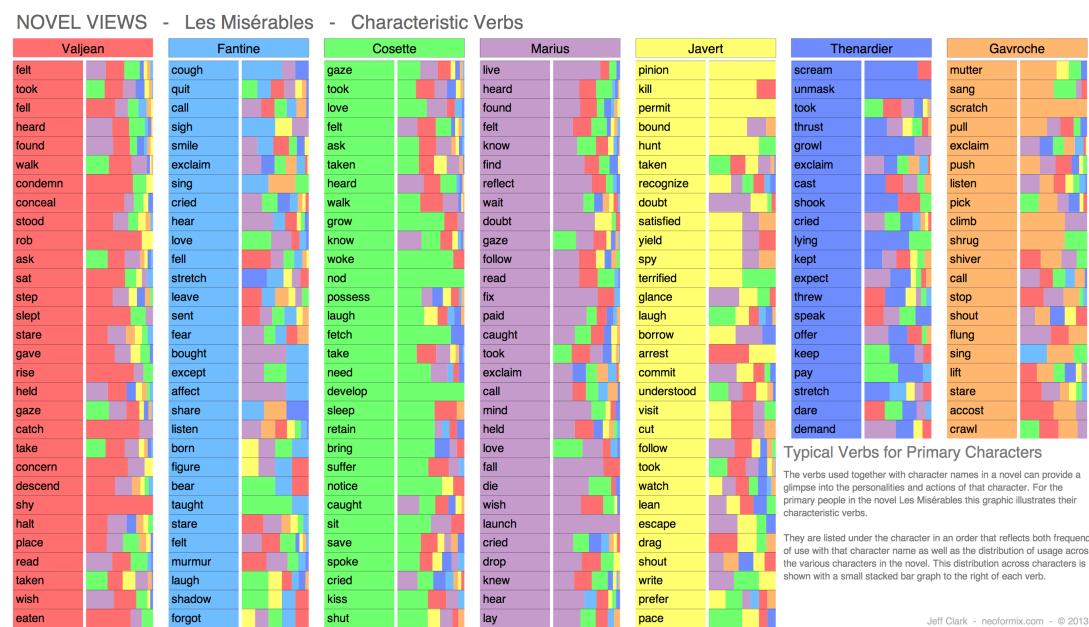


Figure 0.16: Novel Views: Les Misérables - Characteristic Verbs by Jeff Clark (2013)

17 ▷ Wordle by Jonathan Feinberg (2009)

- data: any text
- method: Word cloud
- description: Wordle is a toy for generating “word clouds” from text that you provide. The clouds give greater prominence to words that appear more frequently in the source text.
- paper: *Viegas, F. B., Wattenberg, M., & Feinberg, J. (2009). Participatory visualization with wordle. Visualization and Computer Graphics, IEEE Transactions on, 15(6), 1137-1144.*

▷ [<http://cyber-kap.blogspot.com.au/2011/04/top-10-sites-for-creating-word-clouds.html>]

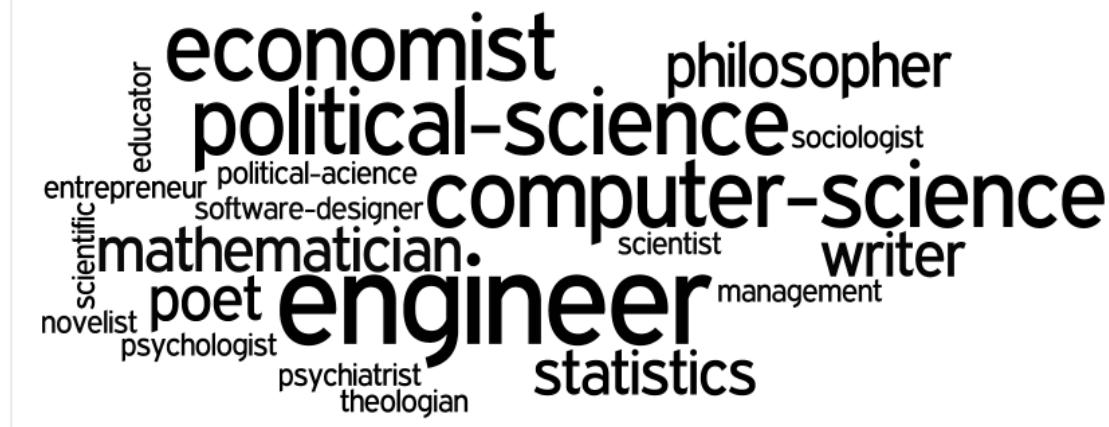


Figure 0.17: Wordle showing the professions of inventors of data visualization tools

18 ▷ Docuburst by C. Collins, S. Carpendale , and G. Penn (2009)

- data: any text, specially used with books. It is a visualization of Wordnet lexical database.
- method: Radial, space-filling layout of hyponymy (IS-A relation)
- description: DocuBurst is the first visualization of document content which takes advantage of the human-created structure in lexical databases. The authors used an accepted design paradigm to generate visualizations which improve the usability and utility of WordNet as the backbone for document content visualization. A radial, space-filling layout of hyponymy (IS-A relation) is presented with interactive techniques of zoom, filter, and details-on-demand for the task of document visualization. The techniques can be generalized to multiple documents.
- paper: *C. Collins, S. Carpendale, and G. Penn, “DocuBurst: Visualizing Document Content Using Language Structure,” Computer Graphics Forum (Proc. of the Eurographics/IEEE-VGTC Symposium on Visualization (EuroVis)), vol. 28, iss. 3, pp. 1039-1046, 2009.*

▷ [<http://vialab.science.uoit.ca/portfolio/docuburst-visualizing-document-content-using-language-structure>]



Figure 0.18: Screenshot of docuburst interactive interface.

19 ▷ Phrase Nets by Frank van Ham, Martin Wattenberg and Fernanda B. Viégas (2009)

- data: Any text. Jane Austen's novel "Pride and Prejudice." [N/A]
- method: Network of sized-words, some connected
- description: "Phrase Nets use a simple form of pattern matching to provide multiple views of the concepts contained a book, speech, or poem. The image below is a word graph made from Jane Austen's novel "Pride and Prejudice." The program has drawn a network of words, where two words are connected if they appear together in a phrase of the form "X and Y":"
- paper: *Van Ham, F., Wattenberg, M., & Viégas, F. B. (2009). Mapping text with phrase nets. Visualization and Computer Graphics, IEEE Transactions on, 15(6), 1169-1176.*

[<http://hint.fm/projects/phrasenet/>]

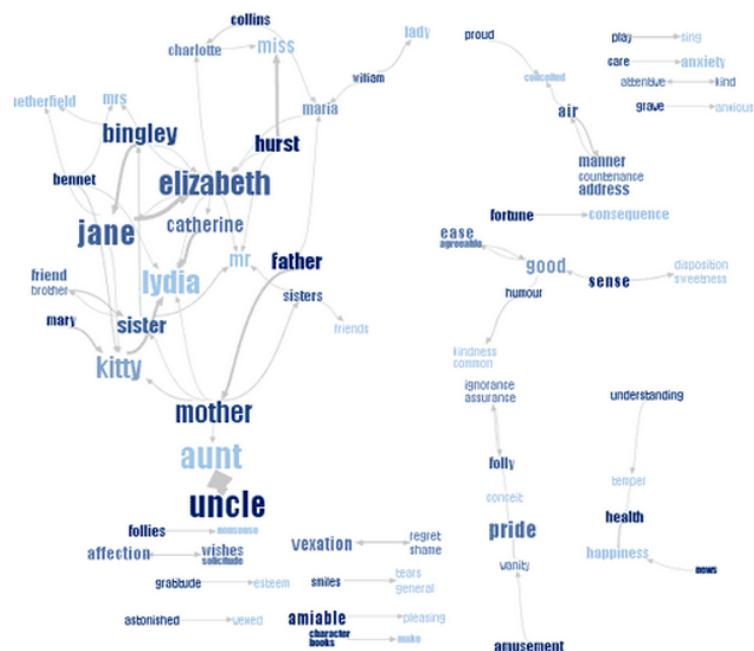


Figure 0.19: Phrase Nets of Jane Austen's novel "Pride and Prejudice." by Frank van Ham, Martin Wattenberg and Fernanda B. Viégas (2009)

20 ▷ Word Spectrum: Visualizing Google's Bi-Gram Data by Chris Harrison (2008)

- data: The huge Google bi-gram dataset [Large dataset]
- method: Word spectrum
- description: “Using Google’s enormous bigram dataset, I produced a series of visualizations that explore word associations. Each visualization pits two primary terms against each other. Then, the use frequency of words that follow these two terms are analyzed. For example, “war memorial” occurs 531,205 times, while “peace memorial” occurs only 25,699. A position for each word is generated by looking at the ratio of the two frequencies. If they are equal, the word is placed in the middle of the scale. However, if there is an imbalance in the uses, the word is drawn towards the more frequently related term. This process is repeated for thousands of other word combinations, creating a spectrum of word associations. Font size is based on an inverse power function (uniquely set for each visualization, so you can’t compare across pieces). Vertical positioning is random.”

[<http://www.chrisharrison.net/index.php/Visualizations/WordSpectrum>]

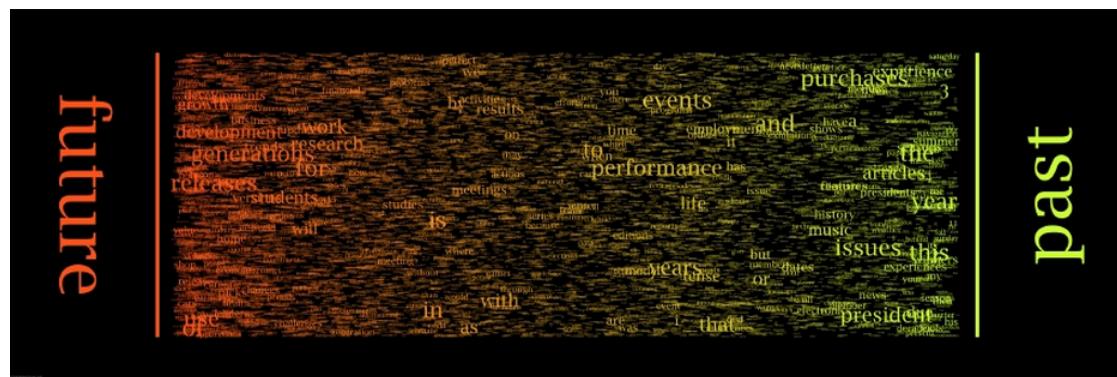


Figure 0.20: ▷ Word Spectrum: Visualizing Google's Bi-Gram Data by Chris Harrison (2008)

0.1. SINGLE TEXT VISUALIZATION

21 ▷ Word Associations Visualizing Google's Bi-Gram Data by Chris Harrison (2008)

- data: The huge Google bi-gram dataset [Large dataset]
- method: Words rays
- description: “words are bucketed into one of 25 different rays. Each of these represent a different tendency of use (ranging from 0 to 100% in 4% intervals). Words are sorted by decreasing frequency within each ray. I render as many words as can fit onto the canvas. There is a nice visual analogy at play - the “lean” of each ray represents the strength of the tendency towards one of the two terms. As in the word spectrum visualization, font size is based on a inverse power function (uniquely set for each visualization, so you can’t compare across pieces). Common words (a, the, for, as, etc.) are not shown.”

[<http://www.chrisharrison.net/index.php/Visualizations/WordAssociations>]

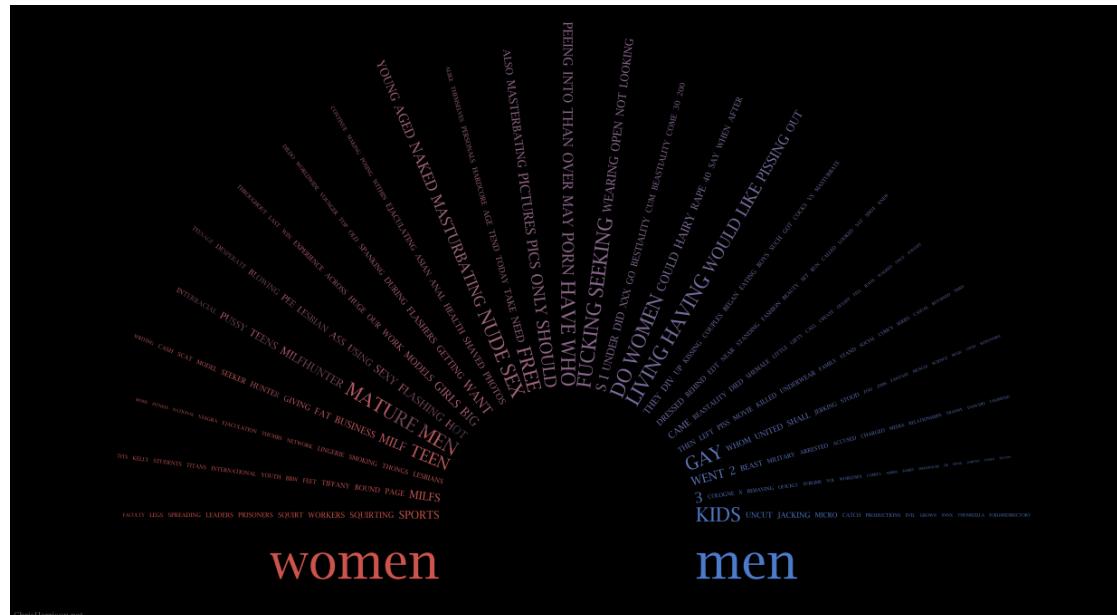


Figure 0.21: Word Associations Visualizing Google's Bi-Gram Data by Chris Harrison (2008)

0.1. SINGLE TEXT VISUALIZATION

22 ▷ Document Arc Diagrams by Jeff Clark (2007)

- data: any text
- method: Arc diagram
- description: Document Arc Diagrams illustrate the similarity structure within a text document by drawing arcs connecting segments of a document that share similar vocabulary.

▷ [http://www.neoformix.com/2007/DocumentArcDiagrams.html]

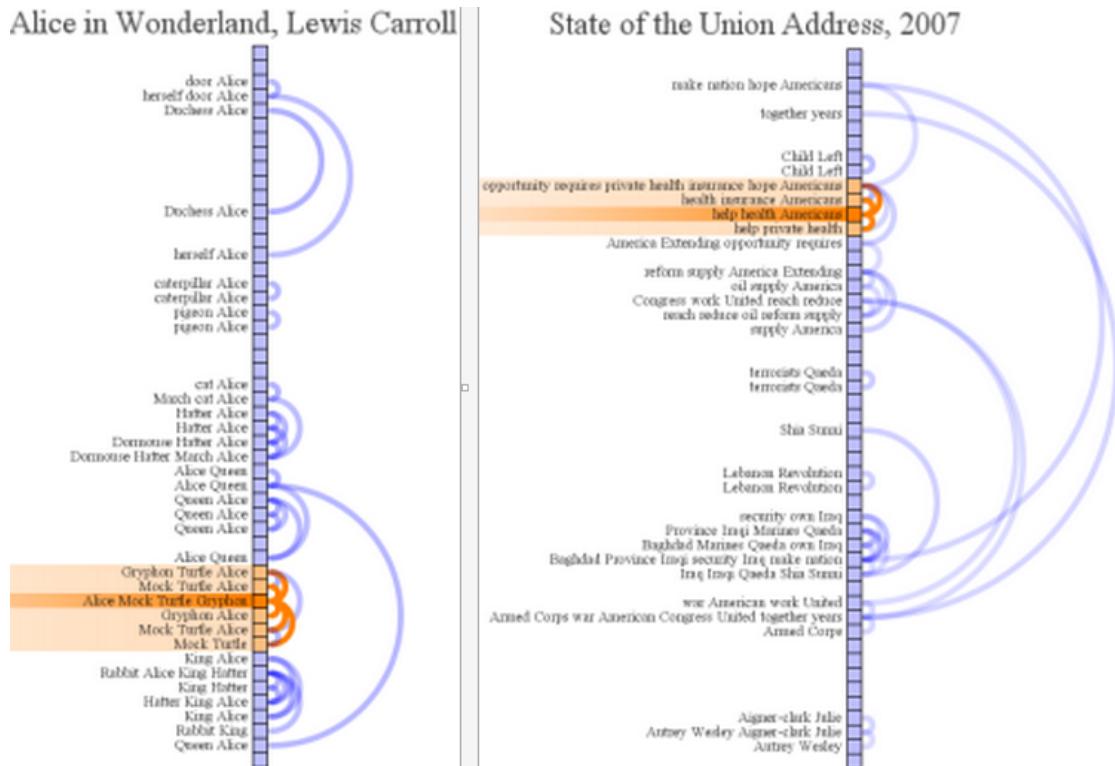


Figure 0.22: Jeff Clark version of arc diagrams

23 ▷ Gist icons by P. DeCamp, A. Frid-Jimenez, J. Guiness, D. Roy (2005)

- data: any body of literature: 1) The complete set of approximately 7 million USPTO patents; 2) Enron email data set comprised of 500,000 emails; 3) A collection of computer generated speech transcripts from MIT Media Lab Symposia.
- method: interactive radial histogram
- description: The shape contains the semantic profile of a single document where the peaks and valleys are defined by the relatedness of words or concepts to that document.
- paper: *Decamp P., Frid-Jimenez A., Guiness J., Roy D.: Gist icons: Seeing meaning in large bodies of literature. In Proc. of IEEE Symp. on Information Visualization, Poster Session (Oct. 2005).*

▷ [<http://media.mit.edu/cogmac/publications/IEEEIcons.pdf>]

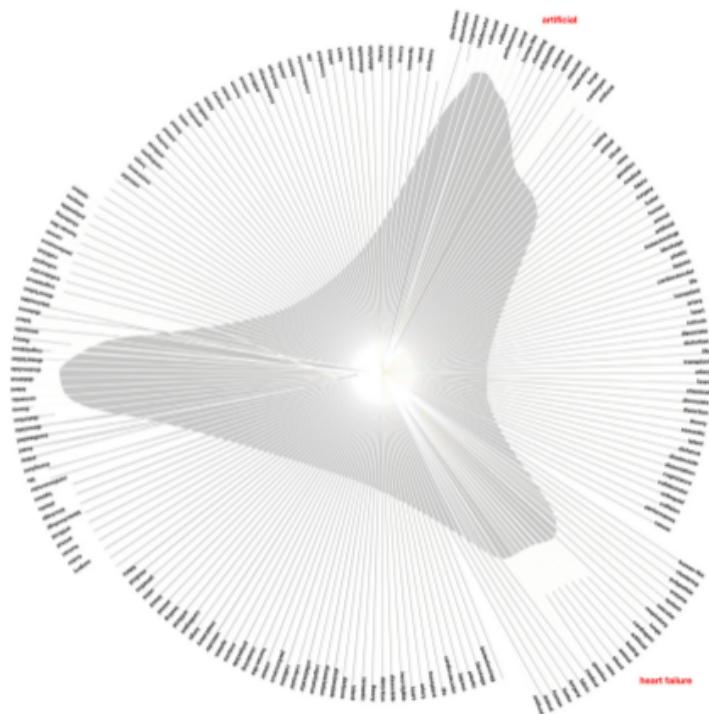


Figure 0.23: Gist icons by P. DeCamp, A. Frid-Jimenez, J. Guiness, D. Roy (2005)

0.2. TEXTS COLLECTION VISUALIZATIONS

0.2 Texts collection visualizations

24 ▷ Novel Views: Les Misérables - Segment Word Clouds by Jeff Clark (2013)

- data: Novel "Les Misérables" [N/A]
- method: Word cloud
- description: "a series of small word clouds for each book within the novel are shown."

[<http://neoformix.com/2013/NovelViews.html>]

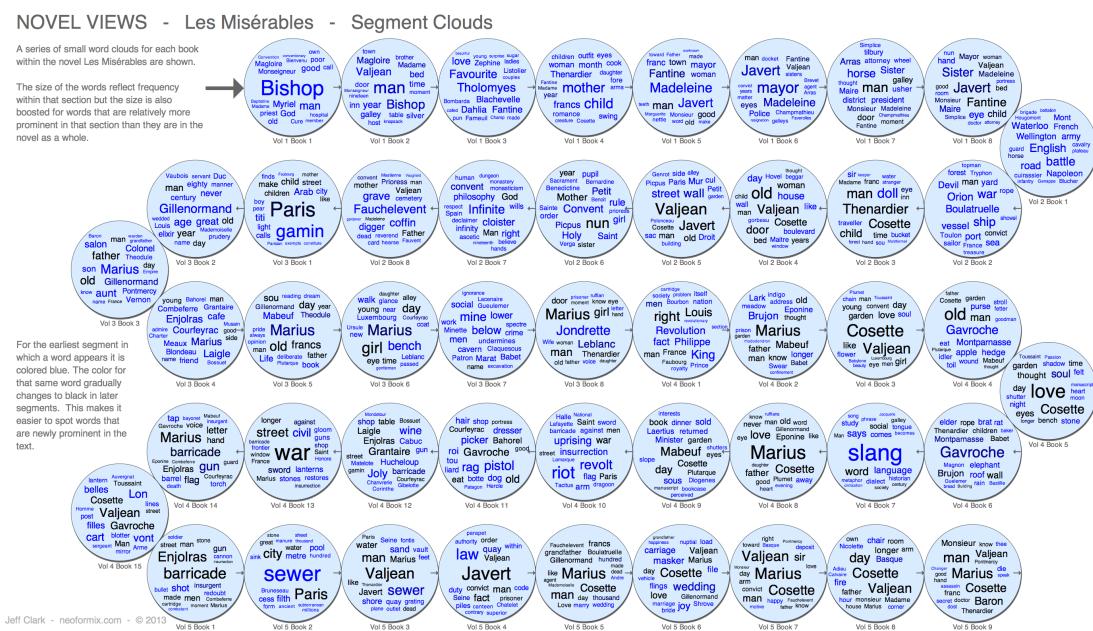


Figure 0.24: Novel Views: Les Misérables - Segment Word Clouds by Jeff Clark (2013)

25 ▷ Grimm's Fairy Tale Network by Jeff Clark (2013)

- data: 62 stories of the Grimms's Fairy Tales [Small dataset]
- method: 2-dimension networks
- description: "the graphic below is a simple network showing which stories are connected through the use of a common vocabulary. There are three different strengths of connection shown and I've tried to minimize the usual 'hairball' nature of these types of diagrams by only showing the top three connections for a story."

[<http://neoformix.com/2013/GrimmNetwork.html>]

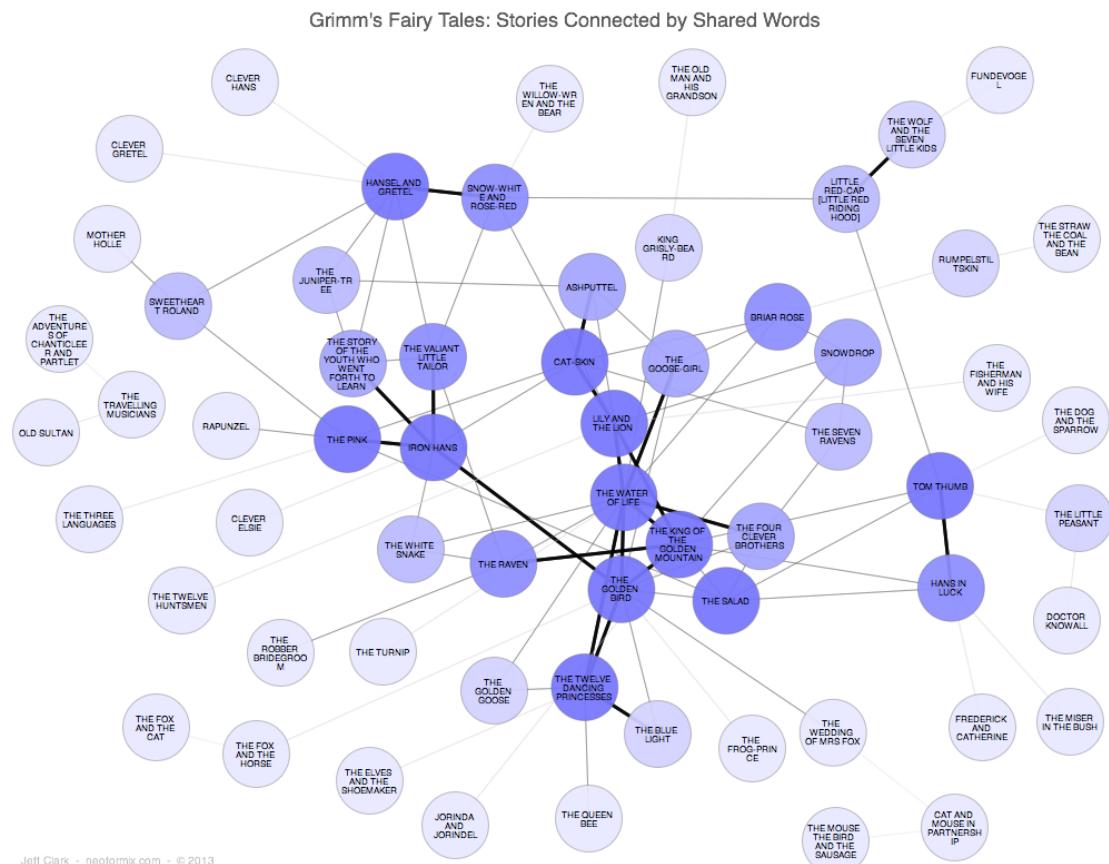


Figure 0.25: Grimm's Fairy Tale Network by Jeff Clark (2013)

26 ▷ Spot by Jeff Clark (2012)

- data: Twitter [Small dataset]
- method: Multi visualization of tweets based on a search query.
- description: "Spot is an interactive real-time Twitter visualization that uses a particle metaphor to represent tweets. The tweet particles are called spots and get organized in various configurations to illustrate information about the topic of interest."

[<http://neoformix.com/2012/IntroducingSpot.html>]

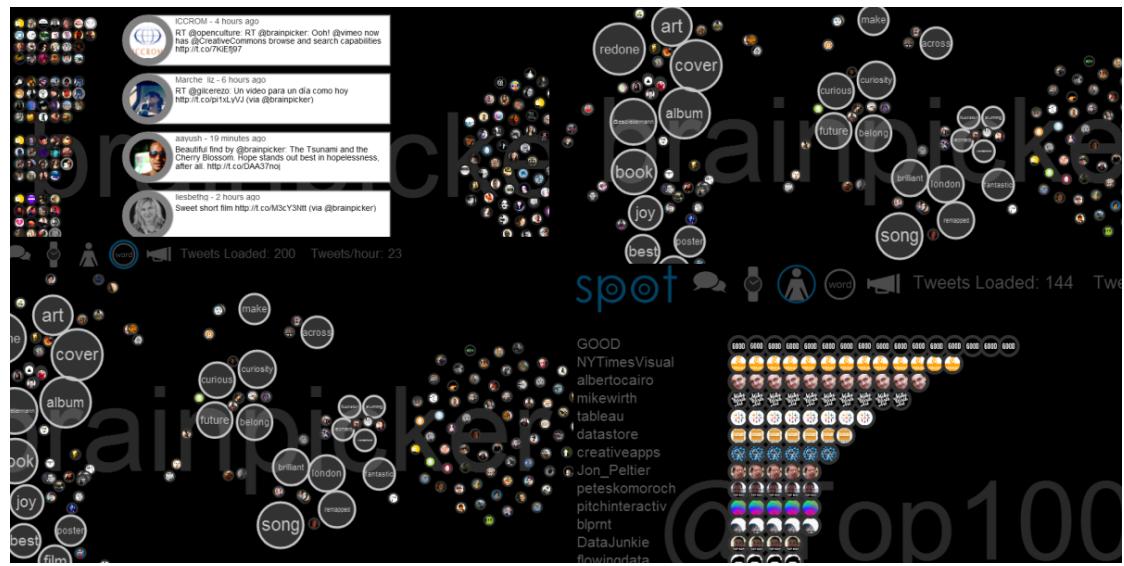


Figure 0.26: Spot by Jeff Clark (2012)

0.2. TEXTS COLLECTION VISUALIZATIONS

27 ▷ Word storm by Quim Castella and Charles Sutton (2012)

- data: Papers from ICML - International Conference on Machine Learning June 26July 1, 2012 Edinburgh, Scotland [Small dataset]
- method: Word cloud with fixed word position
- description: visualization of the conference session in the form of a word storm, which is a group of word clouds. The clouds are arranged so that if the same word appears in two clouds, it is in the same position. This is intended to make it easier to see the difference between clouds.
- paper: *Castella, Q., & Sutton, C. (2013). Word Storms: Multiples of Word Clouds for Visual Comparison of Documents. arXiv preprint arXiv:1301.0503.*

[<http://icml.cc/2012/whatson-all/>]

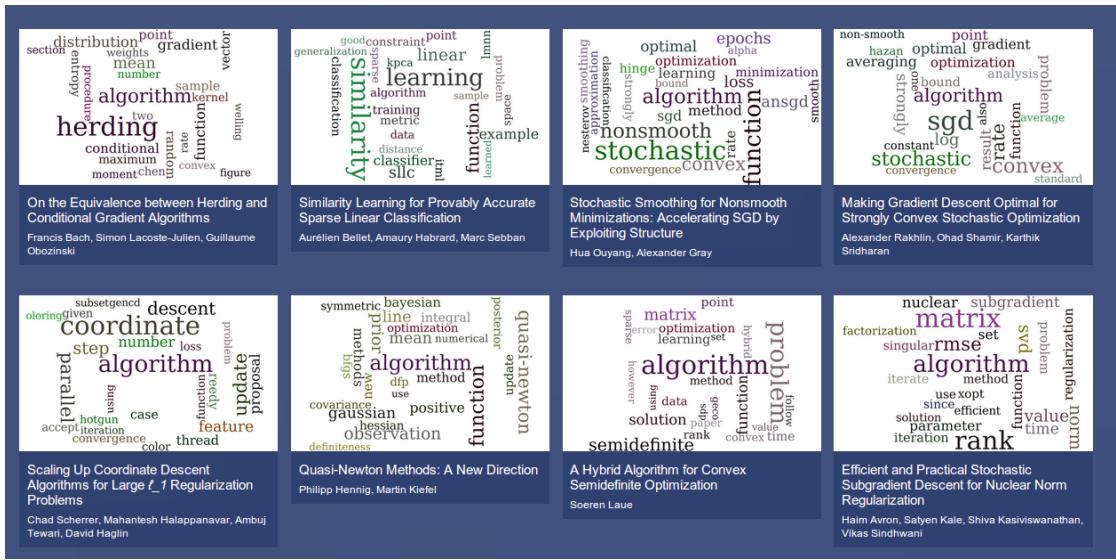


Figure 0.27: Word storm by Quim Castella and Charles Sutton (2012)

28 ▷ Topic Networks in Proust - Topology by Elijah Meeks, Jeff Drouin (2011)

- data: Marcel Proust texts [Large dataset]
- method: Topic model network
- description: “this is document-topic network representation. This visualization shows the relationships abd topics of a collection of documents. Visually this is represented by the distances between documents, topics and documents-topic.”

[<https://dhs.stanford.edu/algorithmic-literacy/topic-networks-in-proust/>]

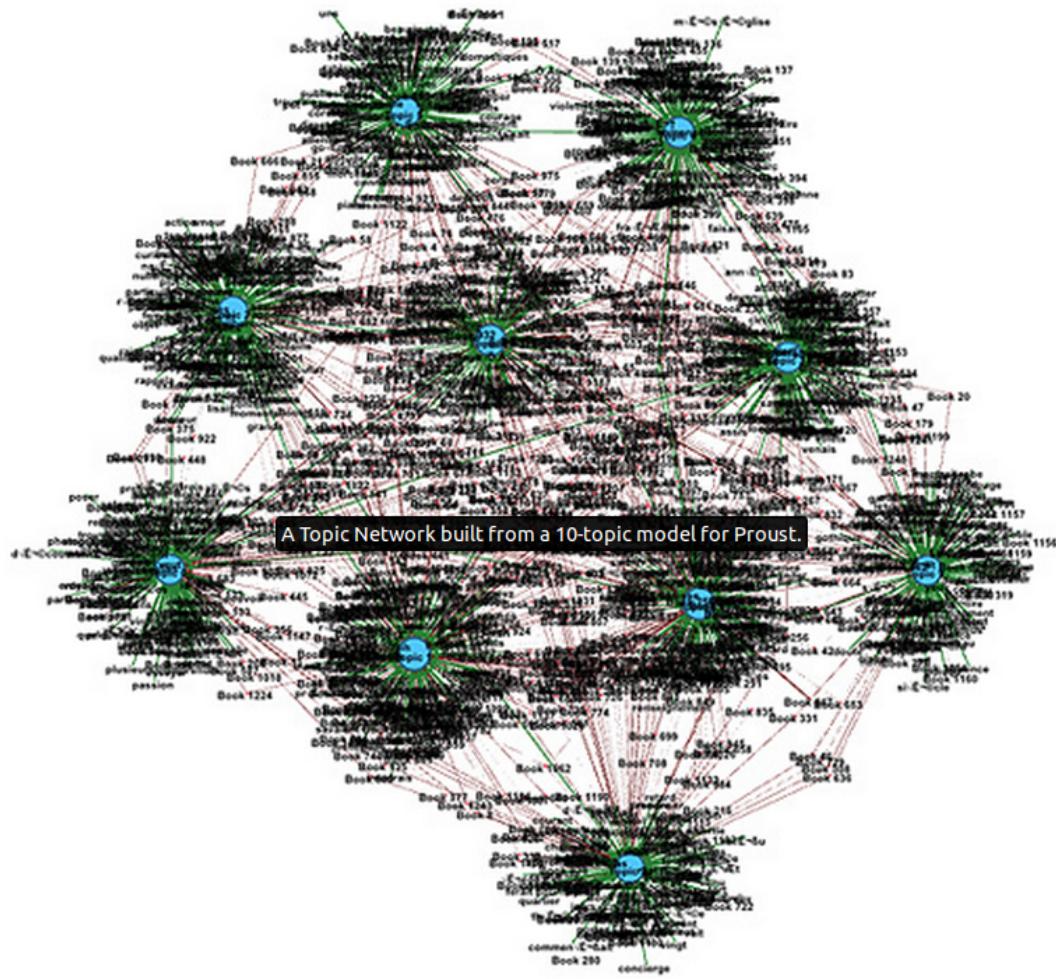


Figure 0.28: Topic Networks in Proust - Topology by Elijah Meeks, Jeff Drouin (2011)

29 ▷ Notabilia. 100 Longest Article for Deletion [AfD] discussions on Wikipedia by Dario Taraborelli, Giovanni Luca Ciampaglia (data and analysis) and Moritz Stefaner (visualization). (2010)

- data: 100 Longest Article for Deletion [AfD] of Wikipedia [Small dataset]
- method: Original. Bouquet of edition lines
- description: “visualization of debate deletion of entries in Wikipedia. Each time a user joins an AfD discussion and recommends to keep, merge, or redirect the article a green segment leaning towards the left is added. Each time a user recommends to delete the article a red segment leaning towards the right is added. As the discussion progresses, the length of the segments as well as the angle slowly decay.”

[<http://notabilia.net/>]

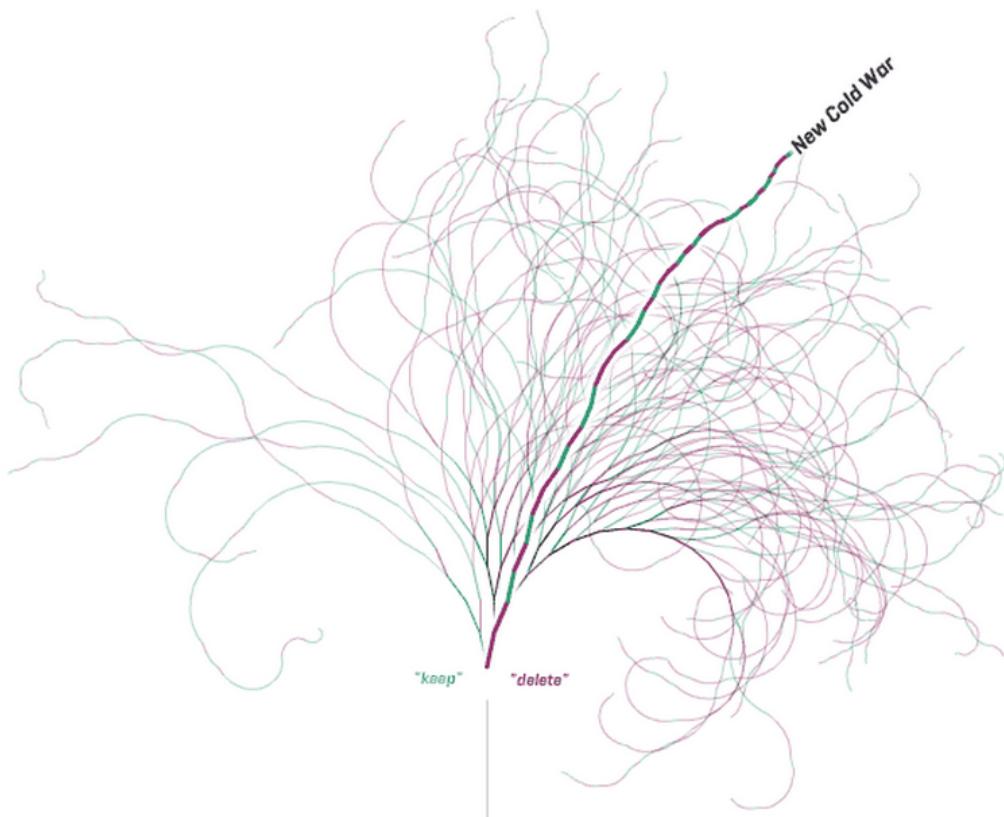


Figure 0.29: Notabilia. 100 Longest Article for Deletion [AfD] discussions on Wikipedia by Dario Taraborelli, Giovanni Luca Ciampaglia (data and analysis) and Moritz Stefaner (visualization). (2010)

30 ▷ X by Y by Moritz Stefaner (2009)

- data: Almost 40.000 submissions to the Prix Ars Electronica, from the early beginnings in 1987 up to 2009 [Large dataset]
- method: Visual spots in groups and colors.
- description: “X by Y visualizes all submissions to the Prix Ars Electronica, from the early beginnings in 1987 up to 2009. The goal is to characterize the ”ars world“ in quantitative terms. A series of diagrams groups and juxtaposes the submissions by years, categories, prizes and countries. The graphics are composed of little dots (each representing a single submission) to provide a visual scale for the statistical statements and thematize the relation of the totality and the individual.”
- Book reference: *Stefaner, M., Ferré, S., Perugini, S., Koren, J., & Zhang, Y. (2009). User interface design. In Dynamic Taxonomies and Faceted Search (pp. 75-112). Springer Berlin Heidelberg.*

[<http://moritz.stefaner.eu/projects/x-by-y/>]

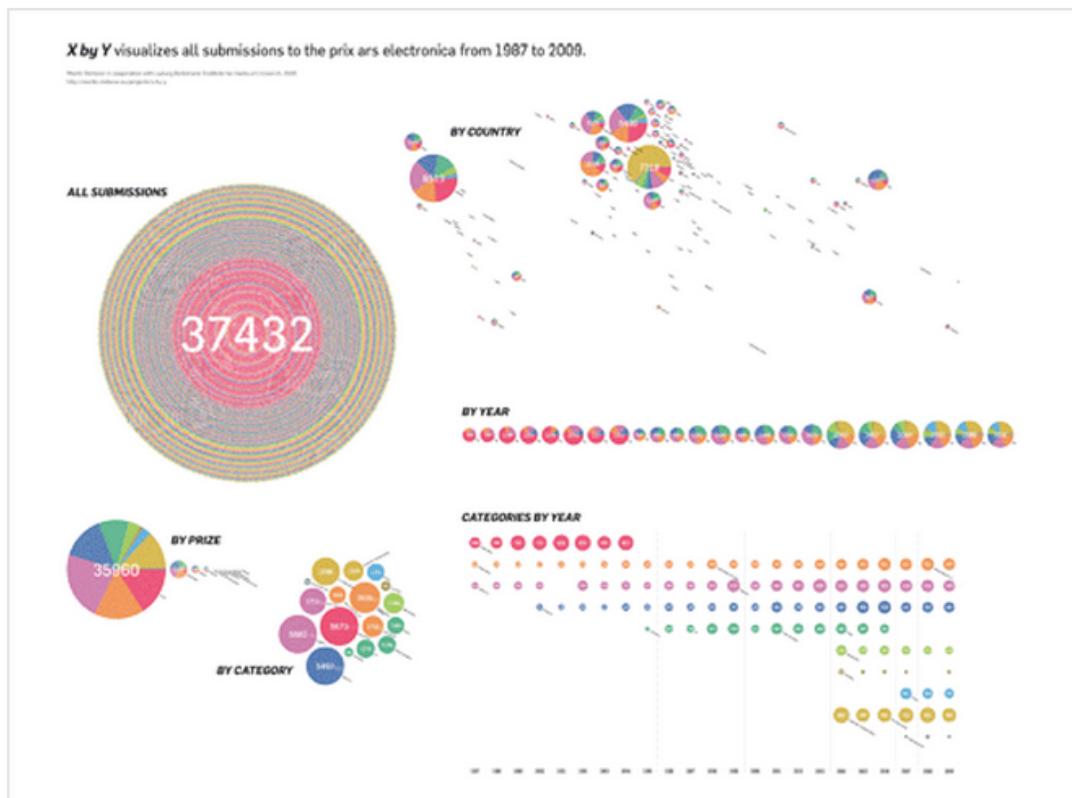


Figure 0.30: X by Y by Moritz Stefaner (2009)

0.2. TEXTS COLLECTION VISUALIZATIONS

31 ▷ Search Clock by Chris Harrison (2008)

- data: Search engine queries [Large dataset]
- method: Clock with queries
- description: "I was curious to see if data from search engines would support my anecdotal observations. I built a simple clock-like visualization that displays the top search terms over a 24-hour period. Displaying search terms in a cyclical layout (like a clock) allows continuous examination of trends that would otherwise be broken up. The data I had access to was both large and noisy. In response, I combined hourly data into week or year averages."

[<http://www.chrisharrison.net/index.php/Visualizations/SearchClock>]

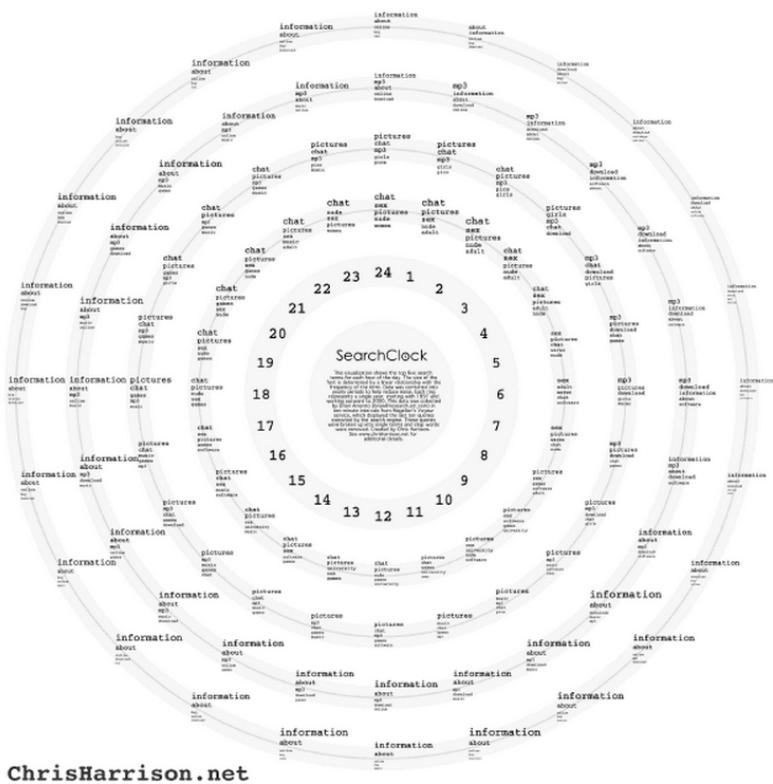


Figure 0.31: Search Clock by Chris Harrison (2008)

32 ▷ Digg Rings by Chris Harrison (2008)

- data: Digg data - top 10 most digg stories of the day for May 24, 2007 to May 23, 2008 [Large dataset]
- method: Tree-ring-like visualizations
- description: “using the Digg API, I grabbed the top 10 most-dugg stories of the day (by midnight) for the past year - May 24, 2007 to May 23, 2008. I then rendered a series of tree-ring-like visualizations (moving outwards in time). Rings are colored according to Digg’s eight top-level categorizations (see key at bottom of page). Ring thickness is linearly proportional to the number of diggs the story received. I also made a pair of visualizations using Digg’s entire archive, which goes back to December 1, 2004.”

[<http://www.chrisharrison.net/index.php/Visualizations/DiggRings>]

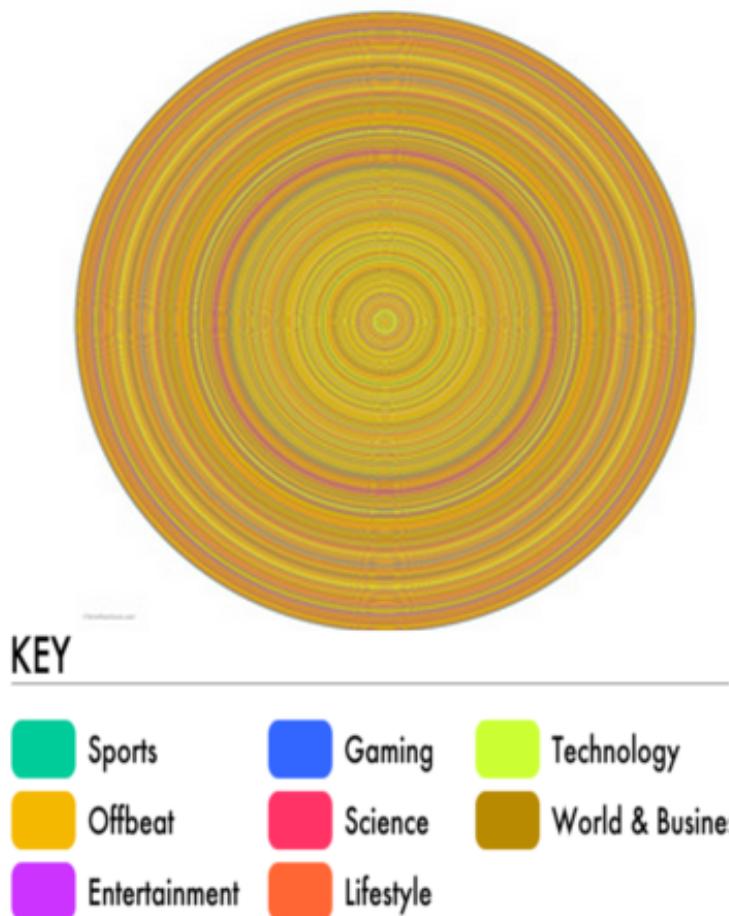


Figure 0.32: Digg Rings by Chris Harrison (2008)

0.2. TEXTS COLLECTION VISUALIZATIONS

33 ▷ Royal Society Archive by Chris Harrison (2008)

- data: titles of papers of the Royal Society Archive (1665-2005) [Large dataset]
- method: time-line
- description: "The Royal Society recently provided access to an archive of papers published in the scientific academy's prestigious journals. Some 25 thousand scholarly works are represented, which date from 1665 to 2005. Many notable scientific minds are represented, including Isaac Newton, Michael Faraday and Charles Darwin. This interesting data set was ripe for some visual tinkering. The database I used was put together by Brian Amento and Mike Yang of AT&T Labs."

[<http://www.chrisharrison.net/index.php/Visualizations/RoyalSociety>]



Figure 0.33: Detail of the visualization Royal Society Archive by Chris Harrison (2008)

34 ▷ WikiViz: Visualizing Wikipedia by Chris Harrison (2007)

- data: Wikipedia [Large dataset]
- method: network
- description: "Wikipedia is an interesting dataset for visualization. As an encyclopedia, its articles span millions of topics. Being a human edited entity, connections between topics are diverse, interesting, and sometimes perplexing - five hops takes you from subatomic particles to Snoop Dog. Wikipedia is revealing in how humans organize data and how interconnected seemingly unrelated topics can be."

[<http://www.chrisharrison.net/index.php/Visualizations/WikiViz>]

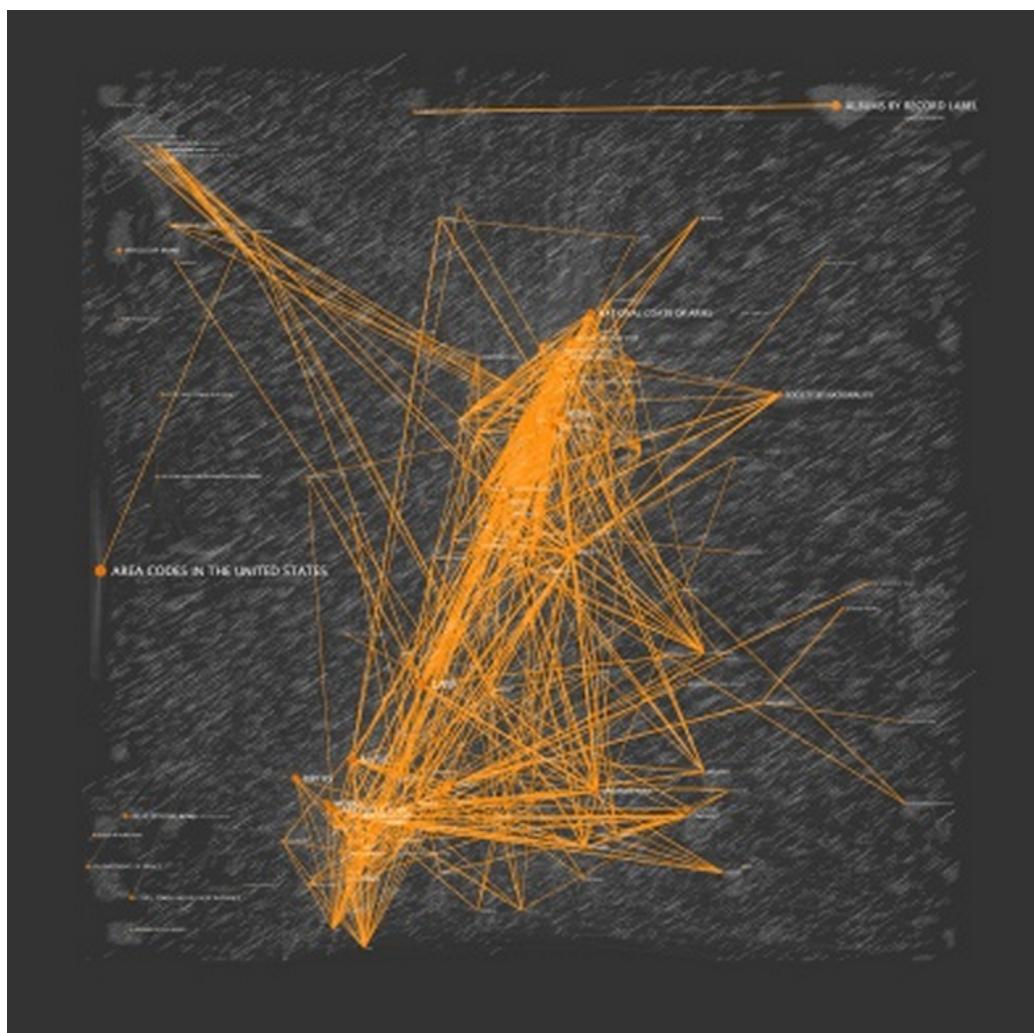


Figure 0.34: WikiViz: Visualizing Wikipedia by Chris Harrison (2007)

35 ▷ Area by Jaume Nualart (2007)

- data: Any collection of texts with parameterizable meta-data to discrete values [Small dataset]
- method: Database representation and browsing tool.
- description: AREA represents the whole dataset, it gives a size-overview of the data, is a data browser, is a data visualization, is a whole data understanding, is interactive, data can be filtered, some time using Area teaches about the main characteristics of the represented data.

[<http://nualart.com/area2>]



Figure 0.35: Area by Jaume Nualart (2007)

36 ▷ Visualizing Activity on Wikipedia with Chromograms by M. Wattenberg, F.B. Viégas, and K. Hollenbach (2004)

- data: Wikipedia [Large dataset]
- method: Chromograms
- description: a chromogram describes the unique edit pattern of a Wikipedian over time by categorizing and color coding edits. Chromograms for distinct persons (and bots!) can be markedly dissimilar.
- Wattenberg, M., Viégas, F. B., & Hollenbach, K. (2007). *Visualizing activity on Wikipedia with chromograms*. In *Human-Computer Interaction-INTERACT 2007* (pp. 272-287). Springer Berlin Heidelberg.

[http://pensivepuffin.com/dwmcphd/syllabi/info447_wi12/readings/wk09-Organizing/wattenberg.Chromogram.INTERACT08.pdf]

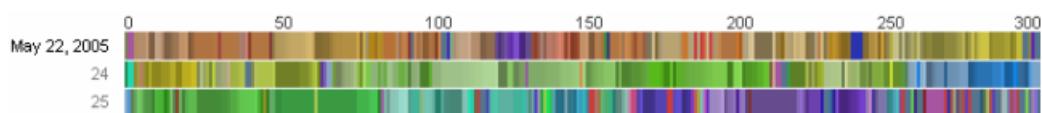


Figure 0.36: Visualizing Activity on Wikipedia with Chromograms by M. Wattenberg, F.B. Viégas, and K. Hollenbach (2004)

37 ▷ Kartoo/Ujiko by Laurent Baleydier and Nicholas Baleydier (2001)

- data: Internet web pages [Small dataset]
- method: Search results as a map of web pages with tagged edges
- description: “Kartoo and Ujiko have been the more advanced search engine interfaces for a wide use. The results were presented as a map of web pages with tagged edges. Kartoo (later Ujiko) was norm in 2001 and shut down in 2010.”

[<http://en.Wikipedia.org/wiki/Kartoo>]

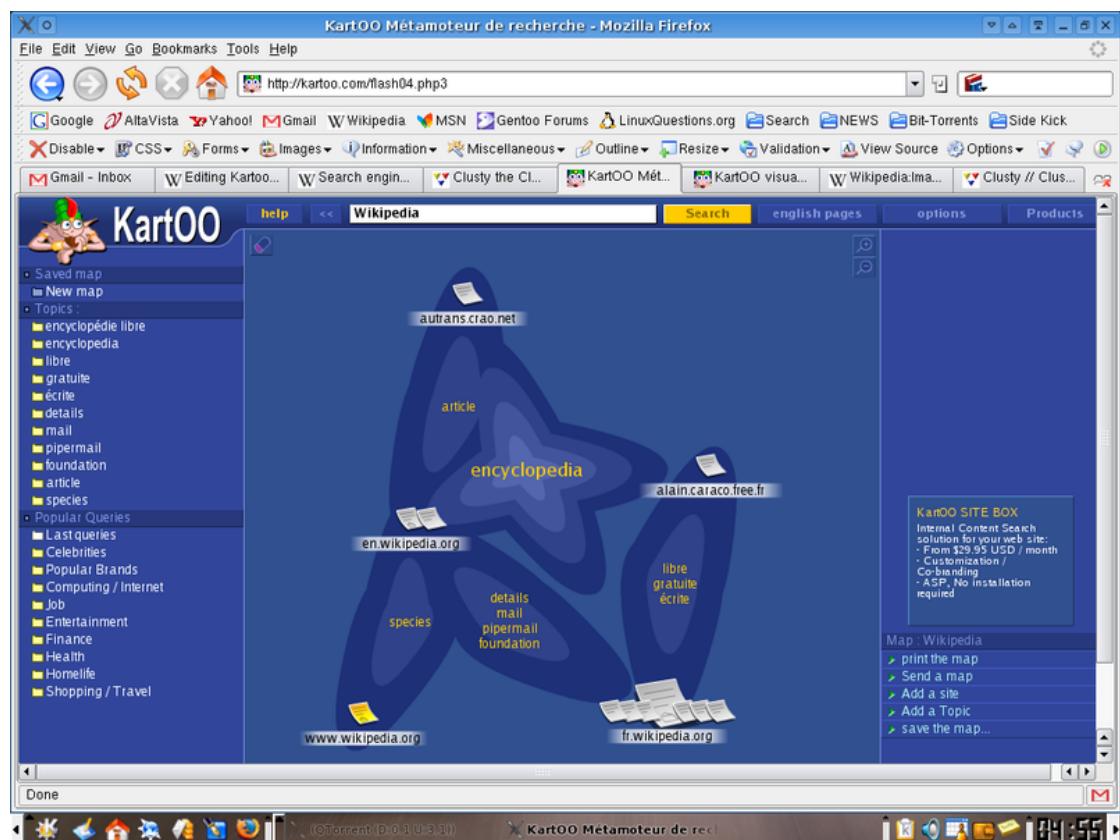


Figure 0.37: Kartoo/Ujiko by Laurent Baleydier and Nicholas Baleydier (2001)

38 ▷ Touchgraph by TouchGraph, LLC. (2001)

- data: Search engine results [Large dataset]
- method: Network visualization of search results
- description: TouchGraph was founded in 2001 with the creation of the original visual browser for Google. Since then, millions of people have used TouchGraph's tools to discover the relationships contained in Google, Amazon, Wikis, and other popular information sources.

[<http://www.touchgraph.com/seo>]

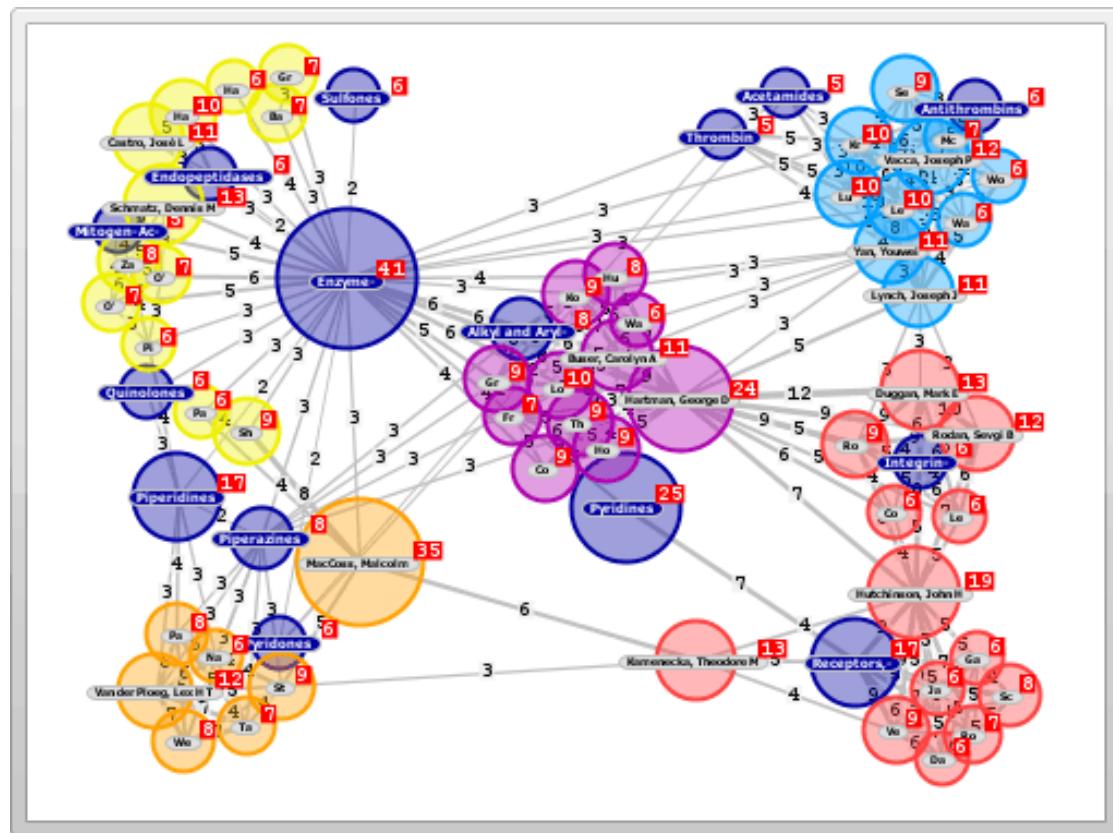


Figure 0.38: Touchgraph by TouchGraph, LLC. (2001)

39 ▷ HotSauce by Ramanathan V. Guha (1996)

- data: Websites relationships written in Meta Content Framework (MCF), a precursor of RDF. [Small dataset]
- method: 3D interactive network
- description: Ramanathan V. Guha: “HotSauce worked as a plug-in to an existing browser so that when a hyperlink to a MCF-enabled website was selected the user was dropped into a first-person perspective view of the Web. It was a videogame view with Web pages floating as brightly colored blocks in an infinite black space, something like the view from a starship cockpit navigating through some strange asteroid field. It was easy to fly into and around the space, using the mouse to guide the direction of flight and holding down buttons to go forwards and backwards. A page could be accessed by simply double-clicking on the relevant block. A 3D immersive environment where I could move around an Antarcti.ca like space, moving things around, interacting with others in that space. And I should be able to do this wherever I am!”

[http://mappa.mundi.net/maps/maps_018/]

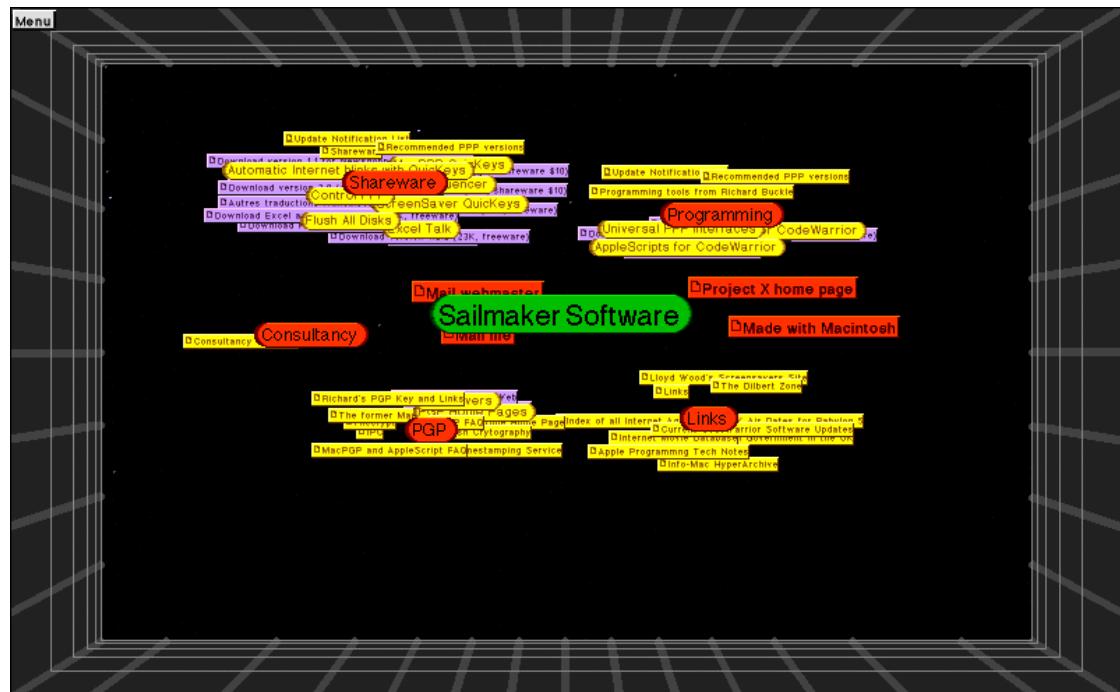


Figure 0.39: HotSauce by Ramanathan V. Guha (1996)

0.2. TEXTS COLLECTION VISUALIZATIONS

0.2.2 Collections of aggregations visualization

40 ▷ Grimm's Fairy Tale Metrics by Jeff Clark (2013)

- data: 62 stories of the Grimms's Fairy Tales [Small dataset]
- method: Sortable matrix with links to the dataset
- description: very complete metric analysis of the 62 stories of the Grimms brothers. This is a high quality example of a tool for linguistics analysis.

[<http://neoformix.com/2013/GrimmStoryMetrics.html>]

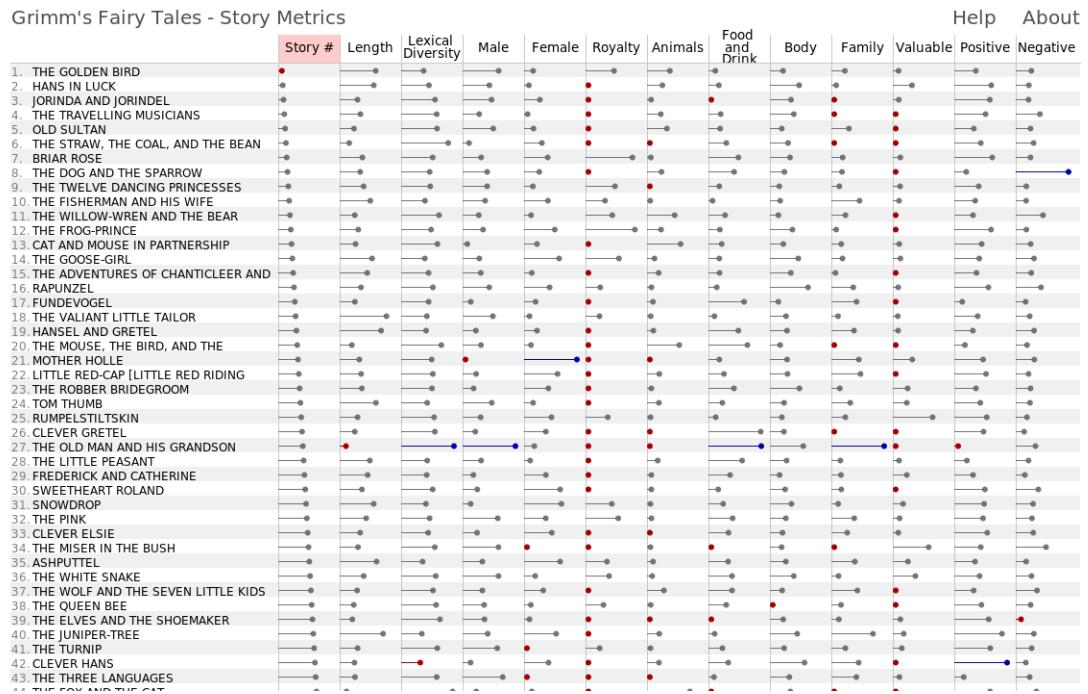


Figure 0.40: Grimm's Fairy Tale Metrics by Jeff Clark (2013)

41 ▷ Termite: Visualization Techniques for Assessing Textual Topic Models by Jason Chuang, Christopher D. Manning, Jeffrey Heer (2012)

- data: any text corpora [Large dataset]
- method: Matrix/tabular view
- description: it is matrix view to support the assessment of topical term distributions and enable the comparison of latent topics. And, in general, it is a visualization of topic models and their terms distribution.
- paper: *Chuang, J., Manning, C. D., & Heer, J. (2012, May). Termite: Visualization techniques for assessing textual topic models. In Proceedings of the International Working Conference on Advanced Visual Interfaces (pp. 74-77). ACM.*

[<http://vis.stanford.edu/papers/termite>]

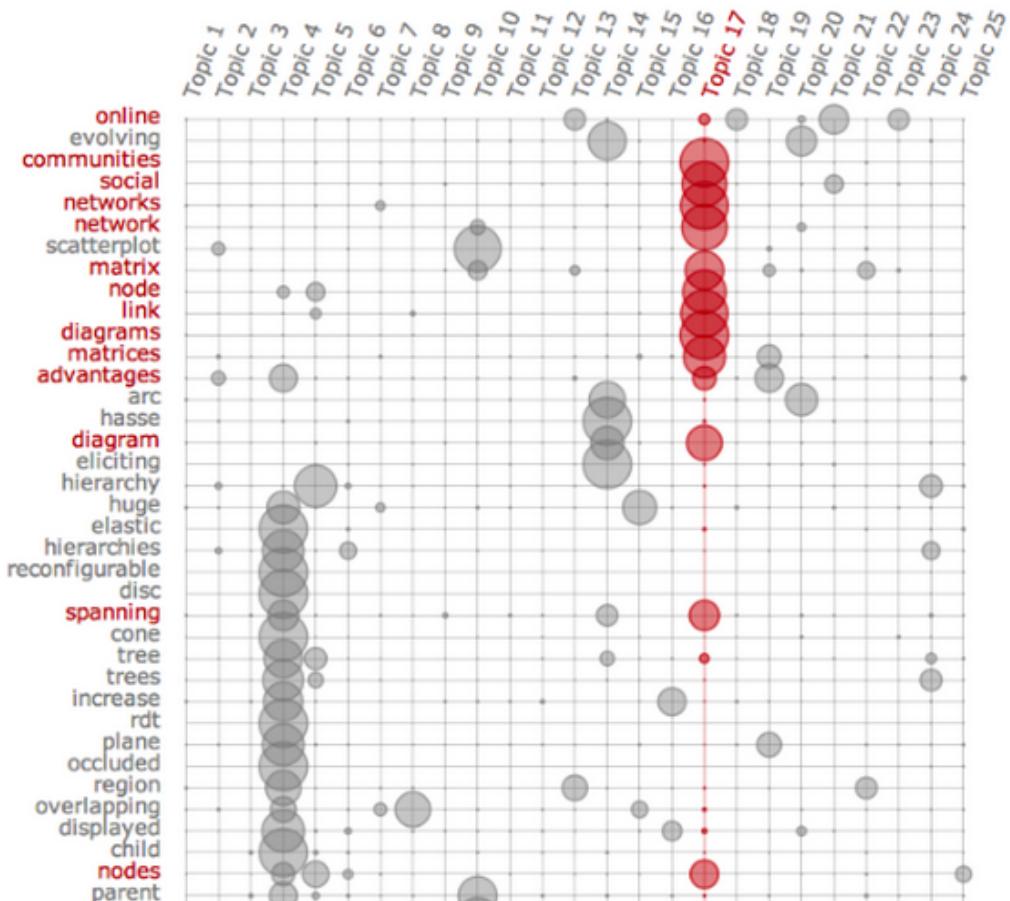


Figure 0.41: Termite: Visualization Techniques for Assessing Textual Topic Models by Jason Chuang, Christopher D. Manning, Jeffrey Heer (2012)

0.2. TEXTS COLLECTION VISUALIZATIONS

42 ▷ Pediameter by Müller-Birn, Benedix and Hantke (2011)

- data: Wikipedia [N/A]
- method: time-line + arduino indicator
- description: a live visualization of Wikipedia Edits using pixel block in a time-line and an arduino indicator

[<http://l3q.de/pediameter/>]

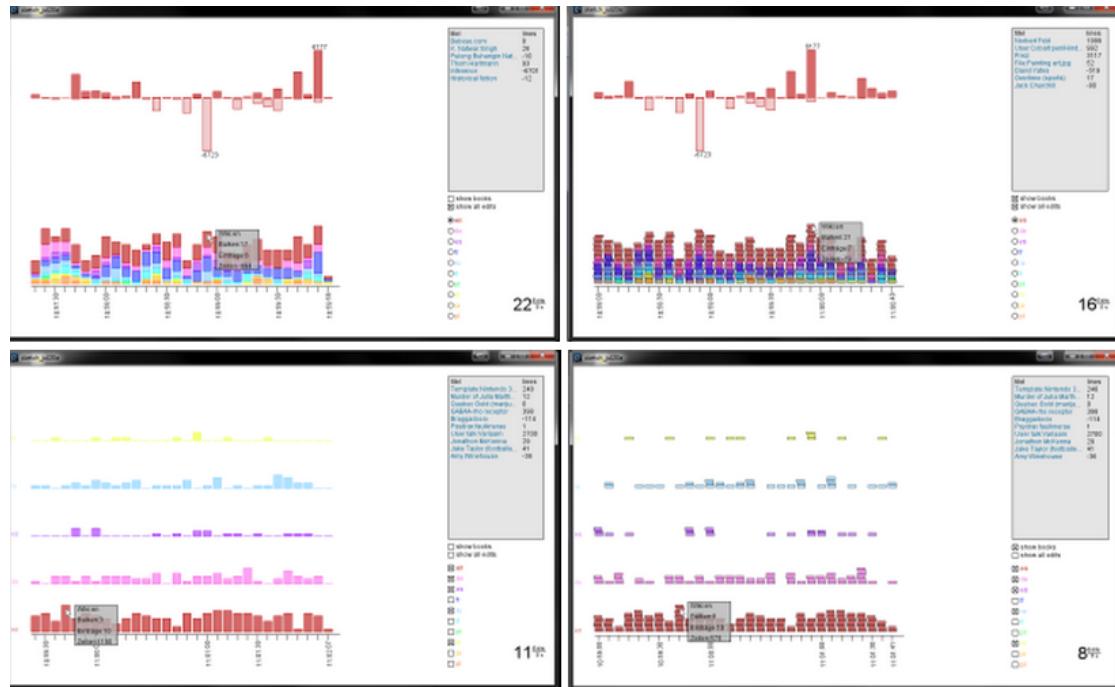


Figure 0.42: Pediameter by Müller-Birn, Benedix and Hantke (2011)

44 ▷ Web Trigrams: Visualizing Google's Tri-Gram Data by Chris Harrison (2008)

- data: Google n-gram dataset (2006) [large dataset]
- method: specific. Connected trees
- description: "As soon as I got my hands on the data, I quickly got to work on some straight forward visualizations. The first type compares two sets of trigrams, each starting with a different word. One visualization compares 'He' with 'She', while the other uses 'I' and 'You'. In the case of the 'He' vs. 'She', the top 120 trigrams for each were identified. The frequencies of the second word in the trigrams were combined and sorted, and rendered in decreasing frequency-of-use order. A similar process was used to create a ranking for the third (and final) word in the trigrams. Words are sized according to the square root of their use frequencies. The color-coded lines act like paths (a tree structure), enumerating all of the trigrams. The process was identical for the 'I' and 'You' version, except that only the top 75 trigrams were used."

[<http://www.chrisharrison.net/index.php/Visualizations/WebTrigrams>]

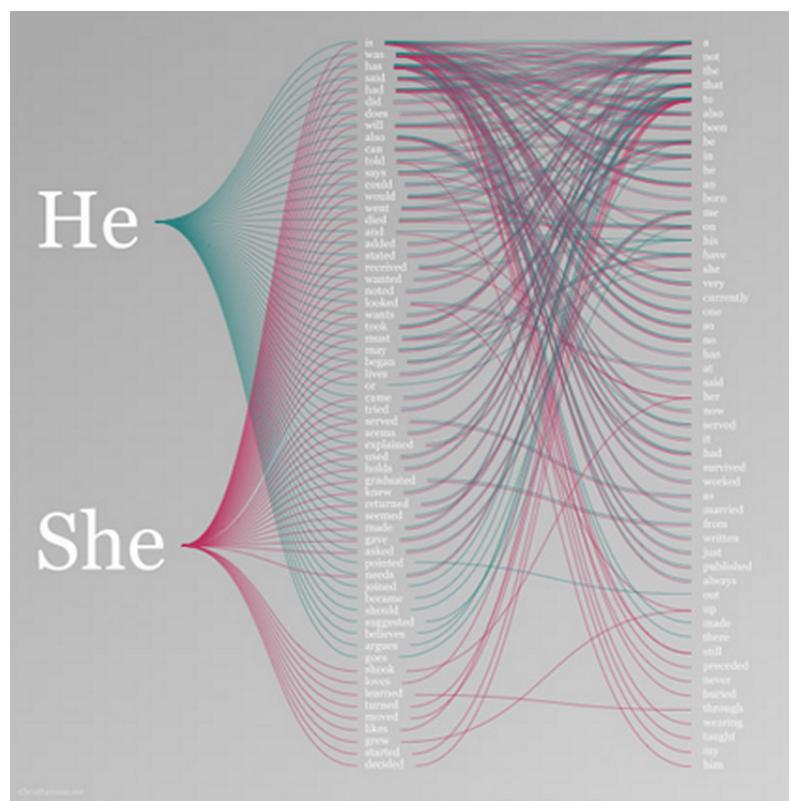


Figure 0.44: Differences in how the he and she subjects are used.

0.2. TEXTS COLLECTION VISUALIZATIONS

45 ▷ **FeaturedLens by A. Don, E. Zheleva, M. Gregory, S. Tarkan, L. Auvil, T. (2007)**

- data: President Bush's eight annual speeches (2001-2007)
- method: rich interface for search, visualize and explore texts
- description: FeatureLens visualizes a text collection at several levels of granularity and enables users to explore interesting text patterns. The current implementation focuses on frequent itemsets of n-grams, as they capture the repetition of exact or similar expressions in the collection. Users can find meaningful co-occurrences of text patterns by visualizing them within and across documents in the collection. This also permits users to identify the temporal evolution of usage such as increasing, decreasing or sudden appearance of text patterns. The interface could be used to explore other text features as well.
- paper: *Don A., Zheleva E., Gregory M., Tarkan S., Auvil L., Clement T., Shneiderman B., Plaisant C.: Discovering interesting usage patterns in text collections: Integrating text mining with visualization. In Proc. of the Conf. on Information and Knowledge Management (2007).*

[<http://hcil2.cs.umd.edu/trs/2007-08/2007-08.pdf>]

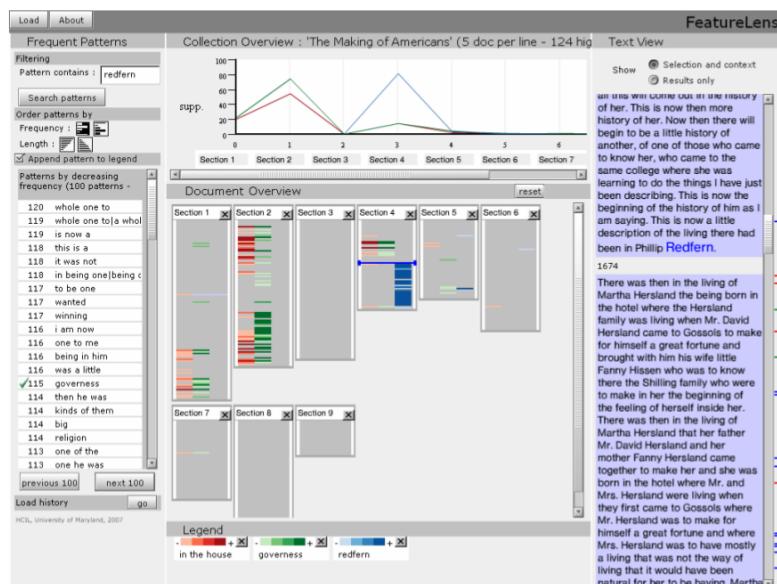


Figure 0.45: Screenshot of the FeatureLens interface.

0.2. TEXTS COLLECTION VISUALIZATIONS

46 ▷ Newsmap by Marcos Weskamp (2004)

- data: Google news [Small dataset]
- method: Treemap
- description: "Newsmap is an application that visually reflects the constantly changing landscape of the Google News news aggregator. Google News automatically groups news stories with similar content and places them based on algorithmic results into clusters. In Newsmap, the size of each cell is determined by the amount of related articles that exist inside each news cluster that the Google News Aggregator presents. In that way users can quickly identify which news stories have been given the most coverage, viewing the map by region, topic or time"

[<http://marumushi.com/projects/newsmap>]



Figure 0.46: Newsmap by Marcos Weskamp (2004)

0.2. TEXTS COLLECTION VISUALIZATIONS

47 ▷ TheMail by Fernanda B. Viégas, Scott Golder, Judith Donath (2006)

- data: Collection of emails conversation. [Small dataset]
- method: Specific interface. Time line with columns of words
- description: "a visualization that portrays relationships using the interaction histories preserved in email archives. Using the content of exchanged messages, it shows the words that characterize ones correspondence with an individual and how they change over the period of the relationship."
- paper: *Viégas, F. B., Golder, S., & Donath, J. (2006, April). Visualizing email content: portraying relationships from conversational histories. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 979-988). ACM.*

[http://smg.media.mit.edu/papers/Viegas/themail/viegas_themail.pdf]

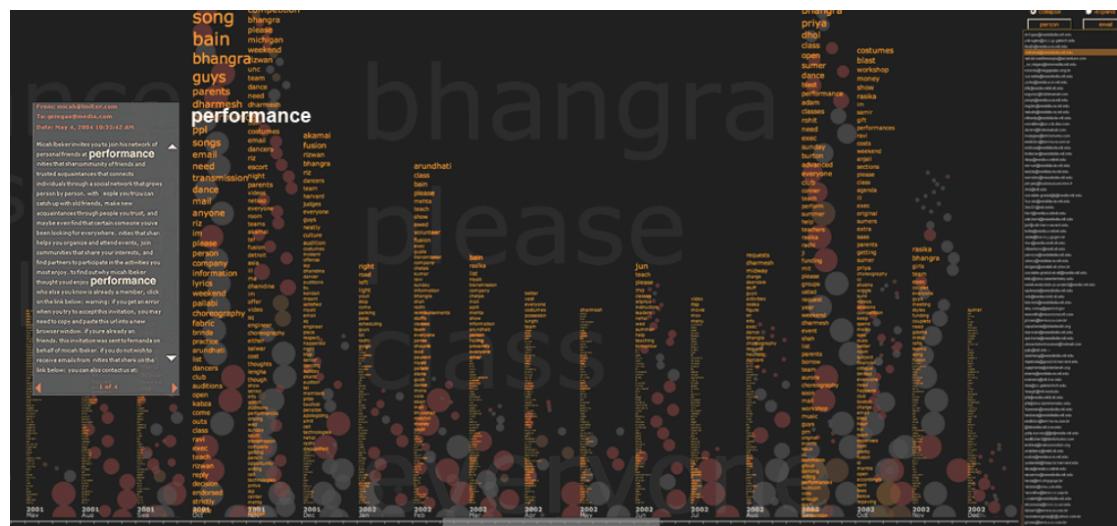


Figure 0.47: TheMail by Fernanda B. Viégas, Scott Golder, Judith Donath (2006)

48 ▷ WebBook by Stuart K. Card, George G. Robertson, and William York (1996)

- data: Search engine results [Small dataset]
- method: 3D interactive book of HTML pages
- description: WebBook was a dynamic multimedia contents constructor from lists of Internet pages. It is very inspiring the reading of the 1996 paper to understand the history of the world wide web.
- paper: *Card, S. K., Robertson, G. G., & York, W. (1996, April). The WebBook and the Web Forager: an information workspace for the World-Wide Web. In Proceedings of the SIGCHI conference on Human factors in computing systems: common ground (pp. 111-ff). ACM.*

[<http://www.sigchi.org/chi96/proceedings/papers/Card/skc1txt.html>]



Figure 0.48: WebBook by Stuart K. Card, George G. Robertson, and William York (1996)

49 ▷ Dotplot Applications by Jonathan Helfman (1994)

- data: Any text [Large dataset]
- method: Dotplot
- description: a bit minimalistic and very effective method to check similarity in large amount of texts, including multilanguage texts or computer code.
- paper: *Helfman, J. (1996). Dotplot patterns: a literal look at pattern languages. Theory and Practice of Object Systems, 2(1), 31-41.*

[<http://imagebeat.com/index.php?id=17>]

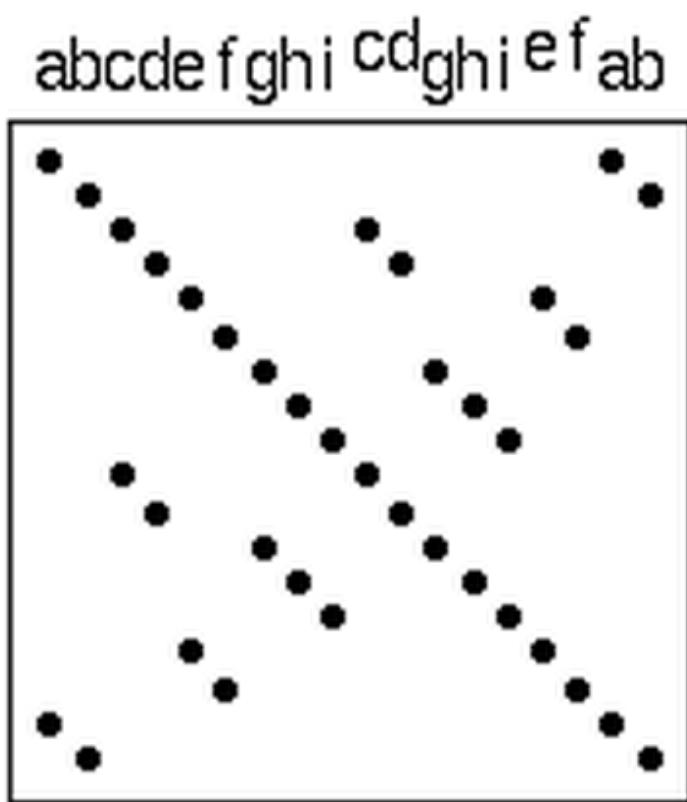


Figure 0.49: Dotplot Applications by Jonathan Helfman (1994)