Surveying Wonderland for many more literature visualization techniques

Richard Brath*
Uncharted Software Inc.

Abstract
There are still many potential literature visualizations to be discovered. By focusing on a single text, the author surveys many existing visualizations across research domains, in the wild, and creates new visualizations. 58 techniques are indicated, suggesting a wider variety of visualizations beyond research disciplines.

Keywords: Text visualization, text analysis, natural language processing, digital humanities.

Index Terms: H.5.2 [User Interfaces]: Screen design; I.2.7 [NLP]: Text analysis; I.7.m [Text Processing]: Miscellaneous; I.5.4 [Pattern Recognition]: Text processing; J.5 [Arts and Humanities]: Literature.

1 Introduction
Are current data visualization techniques too narrow for digital humanities? There may be many more literature visualizations in-the-wild than in the visualization research community, and these can be used to inform more visualizations.

The foundations of data visualizations originated from quantitative and categoric data analysis in statistics [1,2], cartography [3] and computer science [4]. There are 400+ visualizations for textual analysis from research [5-8]—but, many extend quantitative visualizations, such as more than 100 variants of graphs. For example, there tend to be few visualizations of letters or full text of paragraphs. Current text visualizations are too constrained. We are in danger of McLuhan’s dictum: “We shape our tools, thereafter our tools shape us.” [9,10]

Broader visualizations enable wider analysis. Many researchers advocate design alternatives. Munzer says “A fundamental principle of design is to consider multiple alternatives then choose the best.” [11] Roberts et al promote the Five Sheet Design Method to foster divergent thinking and facilitate design alternatives [12]. Hinrichs et al argue for novel experimental visualizations to provoke insights, re-interpret knowledge and mediate ideas [13]. Buxton shows the need to ideate through design alternatives [14]. Tohidi indicates multiple designs facilitate design criticisms and help identify design problems [15].

Bertin [3] created 90 different visualizations of a single dataset indicating the breadth of quantitative visualization. Symanzik et al [16] found 24 different visualizations of Titanic data. Expanding on this approach (and [17]), can a single literary dataset aid finding a breadth of text visualizations?

This paper focuses on one fictional book: Lewis Carroll’s Alice's Adventures in Wonderland, as it has been in continuous publication for 150 years, is widely known, is available digitally, and is used in visualizations both inside and outside the visualization research community. The following survey uncovers dozens of visualizations, which, in turn, frames new visualizations, forming an overall contribution of 50+ literature visualization techniques.

2 Survey of Wonderland Visualizations
Using text search, image search and scholar search with the terms visualization, diagram or infographic and variants of the book title returned 41 visualizations with some common themes:

A. Visualization and Humanities Research. The text of Alice is used in now common techniques such as Figure 1) wordtrees [18]; 2) tag clouds [19]; 3) graphs [20,21]; and 4) storylines [22]. It has also been used in the wonderfully experimental TextArc (5) [23]. Many of these visualizations work at the unit of or a word or two at a time. Word size is used to indicate frequency. Associations between words is indicated with lines. Wordtrees extends frequency to sentences. In Storylines, lines encode sequence and character proximity.
Text analysis systems incorporate visualization components with linked interactions, such as Figs. 6) Juxta [24], and 7) Voyant [25], with screenshots of Alice. In addition to word analysis, these systems include full text with keyword markup. Interactions, such as linking and filtering, enable the viewer to narrow in and compare original context. Language departments use text analytics and visuals may accompany publications such as Figs. 8) word concordance aligning and bolding modifiers for skimming; 9) noun topics and classification sources indicated via caps and italics; and 10) location of topic words within the document [all 26].

B. Alice in NLP. In Natural Language Processing (NLP) Alice has been used for low-level syntax and grammar such as 11) parse trees [27]; and 12) other hierarchical displays [28]. Visualizations of terms, topics and relations include 13) graph of terms and related associations; 14) extracted topic words per chapter [both29]; 15) plots of extracted social network metrics per character [30]; 16) lines indicating topic recurrence (persistent homology) [31]; 17) a chart of character frequency [32] and 18) a chart of emotions [33]. These visualizations are utilitarian diagnostics of word statistics and modelled associations with simple representations.

C. Alice in Fine Arts and Computer Graphics. Academics in arts and computer graphics create different layouts and structure, such as, 19) the micro-structure of phonemes across sentences and languages [34]; 20) a tool for drawing non-linear layouts of text [35]; and 21) digital micrography, wherein text flows within expressive shapes [36].

D. Alice Visualized in the Wild. There are many other visualizations beyond formal research and research tools. Some extract quantitative data, such as 22,23) Alice’s changing height [37,38], 24,25) timelines [39,40], or 26) instances of Alice’s dress [41]. These use common visualization techniques such as line charts, timelines and small multiples.

More varied are visualizations related to characters and relationships, such as Figure 27) a interactive lap book for teaching [42]; 28) a pop-up book, summarizing Alice into six scenes [43]; 29) itemized entities, e.g. color-coded characters, places and events [44]; 30) infographics of themes, characters and relations and motifs [45]; 31,32) graph diagrams showing key characters and relations [46,47]; and 33,34) mind-maps of characters, setting, problems and the solution [48,49].

Plot and themes include examples following chapter sequence, such as 35) a bar chart, one bar per chapter [50]; 36) a wheel where each wedge is a chapter, extending by number of sections in the chapter, color coded by the theme(s) per section [51]; 37,38) plot diagrams, summarizing steps leading to climax and resolution[52,53]; and 39) a multivariate design with rings per chapter; bars indicating content measures of puns, logic, rhyme, etc., and color-coded character dots [54]. Some of these visualizations attempt to uncover structure –
Designers have created unique layouts in limited edition books, such as variation in type size and layout [57]; and superimposed extracts of logical inversions [58]. The typographic manipulation creates a secondary reading across paragraphs and chapters while maintaining the original text.

**Summary.** These art and in-the-wild examples may be questionable (accurate? legitimate?), but they do use visualization cues such as color-coding of words or backgrounds; organizing groups of entities; and using connectors. Unlike prior visualizations, these examples tend to be text heavy ranging down to phonemes and up to abstractive descriptions of characters, extracted quotes, etc., as summarized in Table 1. Typographic enrichment, such as bold, italics, are also more common in the wild.

### Table 1: Analysis of prior Alice visualizations

| Domain          | Number of Visualizations | Number of visualizations by text use in plot area | Visual encoding |
|-----------------|--------------------------|--------------------------------------------------|-----------------|
| Vis & Humanities| 10                       | 3 / 4 / 6 / 3                                  | 2 / 7           |
| NLP             | 8                        | 0 / 3 / 0                                     | 0 / 4           |
| Art & CG        | 3                        | 1 / 0 / 0                                     | 2 / 2           |
| In the Wild     | 20                       | 3 / 4 / 9 / 2                                  | 8 / 12          |
| **TOTAL**       | **41**                   | **8 / 1 / 13 / 9 / 8 / 2**                    | **12 / 25 / 10** |

### 3 ADDITIONAL VISUALIZATIONS

The prior 41 visualizations indicate many ways to visualize a text—are there more? Other novel visualization techniques exist which have not covered Alice. E.g., parallel tag clouds [59] and self-organizing maps [60], but largely focus on words and associations by proximity or explicit linking. Tendril [61] creates a 3D structure based on hyperlinks for an interactive branching reading. But few research visualizations are similar to Figures 19-41.

To consider a wider range of visualizations for literature is to reconsider what can be visualized and how it can be visualized, i.e. a design space exploration. A design space is the set of all the parameters for design variation—in visualization this includes data such as categories, quantities and literal text; visual attributes such as color, size, bold and italic; scope ranging individual letters to chapters; and various layouts—as suggested by Table 1. Many authors have defined the design space [3-4,62-68], the maximal set of design space definitions is considered here. Using Table 1’s outline of text in the plot area, consider additional visualizations of Alice:

**A. Sub-words.** Figure 19 (and prior work e.g. [69]) indicates potential for sub-word visualization. Figure 42 shows neologisms, words deliberately misspelled by Carroll, either lengthened with added letters (shown with high x-height) or shortened (removed letters with low x-height). 43 shows a song from Alice, with a shifting baseline to indicate note pitch, letter-width to indicate note duration, red
text to indicate rhyme phoneme, and yellow box to indicate Carroll’s edits of the original. Sub-word visualization is under-explored: it could be used to indicate stylistic devices such as alliteration, assonance, consonance, cacophony, etc.

**B. Words and Word Pairs.** Many text visualization techniques extract words then visualize (e.g. Figs. 2-5,9-10). These words may be used as literal markers in a visualization such as 44, depicting emotions per character, based on an NLP emotion lexicon [70]. Visualizations that show all labels (Figs 5,9,43,44) do not require slow interaction to reveal labels. Instead, the labels can be read quickly with a shift in attention, yielding faster scanning [68] and potential for serendipitous discovery [71].

Labels can encode information via font, color etc., such as Fig. 9, or: 45 underline length to indicate word frequency—a more compact, more perceptually accurate representation than a word cloud; 46 oblique angle to indicate character sentiment (extracted via NLP), with the most negative character sloping steep left (Gryphon) to most positive sloping steep right (Puppy)[17], and 47 characters over time with character type indicated by typeface (e.g. birds and cards tend to occur together). 48 shows the most common bigrams in Alice: left is the first word in the pair, the following list of words are the most frequent following words, with font weight indicating frequency of occurrence of the pair, e.g., “oh dear” is more frequent than “oh my”, which is more frequent than “oh how”. Pair analysis can do more, e.g. nouns and nearby modifiers [72].

**C. Sentences and Paragraphs.** Many rhetorical devices are associated with sentence construction and repetition of phrases across sentences. Fig. 51 is a syntax diagram [73], similar to a word tree plus loops for repetition. Font weight indicates frequency of successive phrases. A railway diagram can represent actual text or generative text (e.g. auto-completion or transformer models). This diagram indicates text that Carroll wrote, e.g., “will you walk a little faster”; or used to construct phrases that Carroll did not write, e.g., “will you, won’t you, will you walk a little faster.” 52 is a kelp fusion underlay [74], highlighting and linking repetitions of two or more words, similar to Poemage [75]. Beyond logical inversions in this dialogue are many
more repetitions, such as "the same thing" and "you might just as well say," verbally aligning the mad partiers against Alice.

With prose, layout is arbitrary. Interactive layouts can be adjusted, for example, to align repetitions. Unlike a word tree or keywords-in-context (KWIC), the full linear sequence of text remains, and can align words defined by the user or automated, e.g. key nouns.

Fig. 55 shows NLP statistics and annotation counts per chapter. Each chapter and a portion of opening text are shown on each line. Stats are shown as colored bars, bold or underline: e.g. Pig and Pepper is the most disgusting chapter.

Fig. 56 is an adjacency matrix indicating dialogue from one character (vertically) to another (horizontally). Each cell contains the dialogue. Cells are expanded if the next cell to the right would otherwise be empty, and some cells are truncated. Colored content indicates repeated word sets: the Duchess repeats "moral of that is"; the Queen says variants of "off with his/her/their head".

D. Chapters and Document. Document navigation can be aided with enhanced chapters, paragraphs, or lines. With posters, 4k displays, or blankets, the entire text of Alice can be displayed, e.g. Fig. 57. For sizable paragraphs, a noun and verb are extracted, enlarged and set in red (e.g. "Alice peeped" or "Mouse patted"). Chapter titles are presented even larger behind the text. 58 shows the text of Alice, with font size successively increased for 173 popular Alice quotations cited on the Internet, making sentences stand-out such as "Who in the world am I? Ah, that's the great puzzle," or, "We're all mad here."

4 Discussion
These 58 examples, from research, in the wild and constructed, visualize Alice many ways. So what?

Common visualizations such as line charts, bar charts, graphs and scatterplots can be extended further, e.g. prose text, formats, alternate layouts. Marks can range beyond words down to subwords, sentences, paragraphs and documents. The breadth allows the text to be deconstructed and analyzed, e.g. for neologisms, prosody, emotion, sentiment, timelines, character types, character popularity, character descriptions, catch phrases, key nouns, logical fallacies, summarization, landmark extraction, and the potential for much more.
Reviewing these with seven visualization designers (advanced degrees and industry experience in visualization, interaction design, media design, and graphic design), included responses: “I had no idea the range of possible solutions could be so broad,” “It’s interesting what you can do when you cast a wider net,” and, “I really like how there are opportunities to visually manipulate text more than words.” There was also discussion regarding usability with consensus that interaction can aid use.

Some examples split text, manipulate layout and superimpose text, creating different readings, and interrupting linear reading. This is countered that post-modernists create multiple readings though type and layout manipulation, as seen in periodicals such as Ray Gun, Octavo, and Emigre.

5 Conclusion
Most important, this review and the constructions shows there are many different ways to visualize a text, beyond existing visualizations. Each technique extracts, highlights or modifies aspects of the text, beyond existing visualizations. Each technique shows there are many different ways to visualize a text. Some examples split text, manipulate layout and superimpose text, creating different readings, and interrupting linear reading. This is countered that post-modernists create multiple readings through type and layout manipulation, as seen in periodicals such as Ray Gun, Octavo, and Emigre.

Acknowledgements
The author's images are open source, licensed under CC BY SA 4.0 and available on the Internet.

References
[1] Tukey, J. Exploratory Data Analysis, Addison-Wesley, 1970.
[2] Cleveland WS, McGill R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American statistical association. 1984 Sep 1;79(387):531-54.
[3] Bertin, J. Sémiologie graphique, Gauthier-Villars, Paris, 1967.
[4] Mackinlay J. Automating the design of graphical presentations of relational information. ACM TOG. 1986 5(2):110-41.
[5] Kucher, K. and Kerren, A. 2015, April. Text visualization techniques: Taxonomy, visual survey, and community insights. In 2015 IEEE Pacific visualization symposium (pp. 117-121). IEEE.
[6] Alharbi, M. and Laramee, R.S., 2019. Sos textvis: An extended survey of surveys on text visualization. Computers, 8(1), p.17.
[7] Puretskiy, A.A., Shutt, G.L. and Berry, M.W., 2010. Survey of text visualization techniques. Text mining: applications and theory, pp.105-127.
[8] Nualart-Vilaplana, J., Pérez-Montoro, M. and Whitelaw, M., 2014. How we draw texts: a review of approaches to text visualization and exploration. El profesional de la información, 23(3).
[9] Culkín, J. A schoolman’s guide to Marshall McLuhan. Saturday Review, March 18, 1967, 51–53, 70–72.
[10] Heritage Minutes: Marshall McLuhan, last modified Feb. 17, 2016. historicacanada.ca/content/heritage-minutes/marshall-mcluhan
[11] Munzner T. Visualization analysis and design. CRC press; 2014.
[12] Roberts JC, Headland C, Ritoso P. Sketching designs using the five design-sheet methodology. IEEE TVCG. 2015 Aug 12,22(1):419-28.
[13] Hinrichs U, Forlinski S, Moynihan B. In defense of sandcastles: Research thinking through visualization in digital humanities. Digital Scholarship in the Humanities. 2019 Dec 1;34(Supplement_1):i80-99.
[14] Buxton B. Sketching user experiences: getting the design right and the right design. Morgan Kaufmann; 2010 Jul 28.
[15] Tohidi M, Buxton W, Baecker R, Sellen A. Getting the right design and the design right. In Proceedings of the SIGCHI conference on Human Factors in computing systems 2006 Apr 22.
[16] Symanzik, J. Friendly M, and Onder O. The Unsinkable Titanic Data. 2018. Extended version of Visualizing the Titanic Disaster. Significance. 2019 Feb;16(1):14-9.
[17] Brath, R. Visualizing with Text. CRC Press. 2021.
[18] Davies, J. Word Tree [with Alice in Wonderland]. jason-davies.com/wordtree?source=alice-in-wonderland.txt&prefix=dear [Wattementberg M. Viégas FB. The word tree, an interactive visual concordance. IEEE TVCG. 2008 Oct 24;14(6):1221-8].
[19] Wolfram. Word Cloud examples [using Alice in Wonderland]. wolfram.com/language/11/new-visualization-domains/oriented-word-clouds.html.
[20] Semantic Knowledge. Gephi GEXF Exports in Tropes. semantic-knowledge.com/doc/V81/text-analysis/gephi-gexf-exports.html.
[21] Paranyushkin D. Addresses to the Federal Assembly of the Russian Federation by Russian presidents, 2008–2012: Comparative analysis. Russian Journal of Communication. 2013 Dec 1,5(3):265-74.
[22] Tanahashi, Y., and Kwan-Liu Ma. “Design considerations for optimizing storyline visualizations.” IEEE TVCG 18.12 (2012): 2679-2688.
[23] Paley WB. TextArc: Showing word frequency and distribution in text. Poster IEEE Symposium on Information Visualization 2002.
[24] Juxta. Alice: Wonderland vs. Underground. juxta-commons.org/shares/QjM409, juxta-softeware.org
[25] Senghor, L. Alice’s Adventures After Wonderland: Visualizing Alice in the Digital Era. Visual Learning: Transforming the Liberal Arts Conference, 2018. See also: slideplayer.com/slide/3575003 and lateogorman.org/text-analysis/voyant-tools
[26] Hrdličková, J. A Corpus Stylistic Perspective on Lewis Carroll’s Alice’s Adventures in Wonderland, Thesis, Department of English Language and Didactics, Univerzita Karlova v Praze, 2015.
[27] Brennan, J.R, Dyer C., Kuncoro A., Hale JT. Localizing syntactic predictions using recurrent neural network grammars, Neuro-psychologia, Volume 146, 2020, 107479, ISSN 0028-3932.
[28] Jettka, D. and Stührenberg M. “Visualization of concurrent markup: From trees to graphs, from 2D to 3D.” In Proceedings of Balisage: The Markup Conference 2011. Balisage, vol. 7 (2011).
[29] Thys, F. Al in wonderland. SAS blogs. 2017 Jun 23.
[30] Agarwal A, Corvalan A, Jensen J, Rambow O. Social network analysis of Alice in wonderland. In NAACL-HLT 2012 Workshop on computational linguistics for literature 2012 Jun (pp. 88-96).
[31] Zhu X. Persistent homology: An introduction and a new text representation for natural language processing. In Twenty-Third International Joint Conference on Artificial Intelligence 2013.
[32] Langit, L. Visualizing Alice in Wonderland – Wolfram Alpha Pro. 2012 Feb 12. lynmlangt.com/2012/02/12/visualizing-alice-in-wonderlandwolframalphapro/.
[33] Maharjan, S., Kar, S., Montes, M., González, F., Solorio, T. (2018). Letting Emotions Flow: Success Prediction by Modeling the Flow of Emotions in Books. 259-265. 10.18653/v1/N18-2042.
