

Pixel-oriented Visualization for Analyzing Classical Latin Texts

Bharathi Asokarajan*

June Abbas†

Sam Huskey‡

Chris Weaver§

The Digital Latin Library Project, The University of Oklahoma

ABSTRACT

Classics scholars working on ancient manuscripts view each text as a treasure trove of puzzles waiting to be solved. With the help of interactive features and query capabilities, new visualization designs can help scholars improve the efficiency of analysis and quickly identify patterns and distinct characteristics of details in the text. We are developing a new pixel-based text visualization techniques to summarize and analyze the complex structure of primary texts with their rich apparatus metadata in an accessible and comparable fashion. In this poster we will present how pixel-based representations can link the influences of past manuscripts on a scholar's reproduction of lost ancient texts. We will also describe our ongoing work to incorporate new focus+context and other advanced features to help scholars drill deeply into texts.

Keywords: pixel-based, text analysis, data visualization

1 INTRODUCTION

Classics scholars who work with large ancient texts often start with open-ended questions and then narrow down their search to focus on interesting patterns and draw conclusions from them. Text analysis tools including visualizations are of increasing interest in the classics community and in the digital humanities more broadly. In particular, classics scholars work to reconstruct ancient texts in Latin, Greek, and other languages through a combination of close reading and fine-grained linguistic comparison. A *critical edition* presents an editor's reconstruction of a text that includes an assessment of errors and alterations made by scribes throughout the manuscript history of the text, as well as the editor's conjectures based on a variety of trusted sources.

We are working in direct collaboration with scholars in the Department of Classics & Letters at The University of Oklahoma to understand the processes involved in textual criticism and the information presented in critical editions. Typically, footnotes at the bottom of each page record the editor's reasoning behind the reconstructed primary text. This *critical apparatus* is encoded compactly and is often difficult to interpret. Entries in the apparatus contain several types of information including: the location of reconstructed elements in the primary text, such as the chapter and line number of a chosen word; the important textual *variants* that appear in past documents, such as the alternate words chosen in prior editions; and the *witnesses* who chose the variants, such as the scribes copying, hands correcting, and editors reconstructing.

Classics scholars use the information in a critical apparatus to find key points of interest in the primary text, examine linguistic and stylistic choices made over a text's history, analyze the variant choices made by groups of witnesses, and create new editions based on background knowledge and linguistic judgement. Consequently, visualization tools to help classics scholars must be able to

summarize lengthy primary text while at the same time highlight sequential and other structural patterns that exist in the witness and variant information contained in an accompanying apparatus.

Toward meeting these needs, we are developing a visual text analysis tool for scholars of classical Latin using pixel-based visualization techniques. Pixel-based visualization has been popularized by Keim [3] and others for the display of large amounts of multidimensional data within the limited constraints of screen size and human visual acuity. Pixel-based approaches map individual data values into single pixels (or small regions of pixels) as a way to show large amounts of data in limited space. Pixels are typically colored as a function of some quantitative data attribute. Variation in the placement and coloring of pixels allows users to view data in different ways. Notable pixel-oriented visualization techniques span a variety of applications domains and include those used in literature fingerprinting [4] in linguistics, the Information Mural [2] in computer software analysis, OnSet [5] in metabolomics, and the Sequence Surveyor [1] in bioinformatics. We are exploring different ways to represent data, support expressive querying, and maintain interactive immediacy in pixel-based visualizations of text. In our current focus on classical Latin, we are particularly interested in information about text reconstructions within and across scales from individual words to lines to pages to entire works.

2 PIXEL-BASED TEXT ANALYSIS TOOL

The display scalability of pixel-based techniques is well-suited to visualization of critical editions of Latin texts, which can span hundreds of pages and involve thousands of variant readings from dozens of witnesses. For example, the apparatus in Giarratano's critical edition of *Calpurnius Siculus* dissects approximately 1200 variant readings from over 50 witnesses for the 94 lines (spanning 7 pages at an average of 6 words per line) of the first poem alone [6].

We are using the *Improvise* visualization environment [7] to design and implement an interactive pixel-oriented visual analytics tool for exploring and analyzing occurrences of variants across witnesses in critical edition texts. Figure 1 shows the primary text and a pixel-based visualization of the critical apparatus arranged side by side. The horizontal dimension displays the annotated words (*lemmata*) in the primary text in reading order from left to right. Along the vertical are the names of the witnesses (scribes and editors) who used variants of those words in their respective versions of the text. Each "pixel" indicates the presence of a variant for each word and witness combination. Pixel coloring is customized to the analytic purpose; currently, the Levenshtein edit distance between each variant and its corresponding word in the primary text is mapped into a brightness gradient with cyan as the base hue. Brighter pixels indicate higher lexicographic similarities (smaller edit distances).

When a column of pixels is represented by the same color, it is an indication that many witnesses agree on a variant (and that the editor thinks this is worth noting). Conversely, a column showing many pixels of different colors indicates the existence of multiple variants of a word in different copies of the work, suggesting that that point in the text warrants significant scholarly attention.

3 INTERACTION AND QUERY CAPABILITIES

A scholar can examine variations in pixel placement and coloring to identify interesting patterns and outliers. Many words in an edition

*e-mail: bharathi.asokarajan@ou.edu

†e-mail: jmabbas@ou.edu

‡e-mail: huskey@ou.edu

§e-mail: weaver@cs.ou.edu

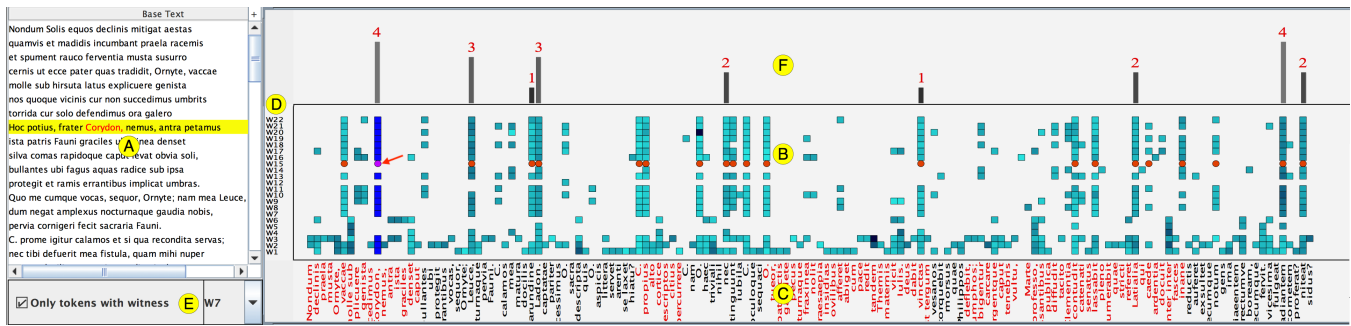


Figure 1: Pixel-based visualization tool for analyzing variants of words that occur across diverse versions of a classic Latin text: (a) edited text; (b) variants by word; (c) words in reading order; (d) witnesses; (e) filtering options; (f) edit distances of variants used by a selected witness (W7).

have no variants noted in the apparatus. A checkbox allows filtering to display only the words in the text that have variants. This filter helps to reduce sparseness in the pixel view, especially over longer primary texts. It also makes it easier to get an overview of the full set of variant readings and their occurrences over a text. To aid in navigation and lookup, words along the horizontal are colored in red (as opposed to black) when they occur in the first line of a page. In the filtered pixel view, this feature helps the user to check the approximate proportion of variants present on each page.

A few of the tasks that a scholar can perform using the tool are:

- Examine the pattern of variants of different words for a given witness. Hovering over a pixel highlights all the pixels for the witness in the same row; see red arrow in Figure 1. Red pixels correspond to words for which that witness provides a variant.
- Examine the pattern of variants over different witness for a given word. Hovering over a pixel also highlights all pixels in the same column. Dark blue pixels mark the word's witnesses.
- Brush the word to see it in context. Clicking a pixel highlights the corresponding word (in red) and the line in which it occurs (in yellow) in the primary text view. This feature exploits the familiarity of the primary text view to help the user navigate the less natural arrangement of text in the pixel-based view.
- Examine the amount of variation over the text. A bar graph above the pixel view acts as a supplement to display the edit distance of words for a chosen witness (selected in a dropdown box). This allow a scholar to fix a particular witness as a benchmark for comparing variation with other witnesses.

4 ONGOING DESIGN

Our next goal is to incorporate more advanced query features to highlight all words in the primary text based on the characteristic (e.g., word length, part of speech) of the pixel selection in the

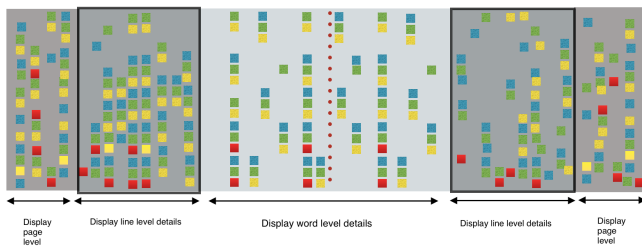


Figure 2: Design of a future pixel-based visualization that offers multiple levels of focus+context for exploring variants across text scales.

graphical display. Sorting and reordering the pixels based on the similarity of variants recorded by scribes will be a key feature of this text analysis tool. We are also extending the current pixel view to represent text at multiple levels of granularity such as chapters, pages, lines, and words, using the multi-level focus+context representation shown in Figure 2. The line and page level sections on the left and right act as increasingly aggregated overviews of the text. A pannable central level view provides a detailed view of individual words and their variants.

The pixel-based visual text analysis tool will help classics scholars perform textual criticism more effectively and efficiently. Correlating witnesses and grouping them in to categories will help scholars analyze prior versions of the text and propose improved reconstructions. The pixel-based view also serves as a way to share observations about textual variation including as a tool for teaching texts to students of Latin. As a part of the ongoing Digital Latin Library project, the pixel-oriented text views will be integrated with storyline and other visualization techniques to provide a full-featured desktop tool for analyzing critical editions and helping Latin scholars perform textual criticism on ancient texts.

ACKNOWLEDGEMENTS

This work was supported as a part of the Digital Latin Library project by a grant from the Andrew W. Mellon Foundation.

REFERENCES

- [1] D. Albers, C. Dewey, and M. Gleicher. Sequence surveyor: Leveraging overview for scalable genomic alignment visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2392–2401, December 2011.
- [2] D. F. Jerding and J. T. Stasko. The information mural: A technique for displaying and navigating large information spaces. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 43–50. IEEE, October 1995.
- [3] D. A. Keim. Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):59–78, January-March 2000.
- [4] D. A. Keim and D. Oelke. Literature fingerprinting: A new method for visual literary analysis. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 115–122, Sacramento, CA, October 2007. IEEE.
- [5] R. Sadana, T. Major, A. Dove, and J. Stasko. OnSet: A visualization technique for large-scale binary set data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1993–2002, December 2014.
- [6] Titus Calpurnius Siculus and Caesar Giarratano. *Calpurnii Et Nemesiani Bucolica*. 1910.
- [7] C. E. Weaver. *Improvise: A User Interface for Interactive Construction of Highly-Coordinated Visualizations*. PhD thesis, University of Wisconsin–Madison, Madison, WI, June 2006.