

---

## Callimachus: A Virtual Archivist for Electronic Markup Projects

---

**Jeff Smith** ([jeffs@smithicus.com](mailto:jeffs@smithicus.com))

University of Saskatchewan

**Joel Deshaye** ([joel.deshaye@usask.ca](mailto:joel.deshaye@usask.ca))

University of Saskatchewan

**Peter Stoicheff** ([peter.stoicheff@usask.ca](mailto:peter.stoicheff@usask.ca))

University of Saskatchewan

---

The field of electronic text editing in the Humanities has been somewhat polarized along boundaries of preferred technology. Some feel that relational databases provide a more robust and powerful representation scheme while others cleave to the expressive power and transferability of XML.

The *Callimachus* project at the University of Saskatchewan was conceived as a way to have our cake and eat it too - merging the robust scalability of formal database technologies with the expressive power and Humanist-friendly accessibility of HTML and XML schema.

First applied to the hypertext edition of Faulkner's *The Sound and the Fury*, *Callimachus* delivers an easy-to-use schema development and implementation environment that eliminates the need for client-side editing, document management and revision control systems.

We propose to present our experiences with *The Sound and the Fury*, sharing the lessons learned and some of the surprising and unanticipated scholarly results that were achievable only with a system that allowed us to change our minds repeatedly and reconceive our schema as we moved along.

*Callimachus* was designed to allow the possibility of using a true database application to markup Faulkner's 1929 novel on a token (or word) level. Using free software tools such as *MySQL*, we built a Web interface with the database, thereby circumventing the need for client-side software, and enabling more than one editor to alter the database simultaneously. By storing each word of the novel in a separate record, we avoided the problems caused by imposing a strict hierarchy (like TEI) on a literary text. The new approach enables us to layer and overlap tags without fear of corrupting a strictly structured markup; we can also use conventional data-mining algorithms to reveal unforeseen relationships between tagged elements in the text.

These relationships are crucial in understanding any literary text that builds meaning through association and structure. Faulkner's *The Sound and the Fury* is a prime example. With this custom database and interface, we can discover when and how a concept appears in the novel. We can discover which characters dwell on what concepts and to what extent. We can discover how many words (how much narrative space) characters use when talking or thinking about specific topics. We can display relationships with charts and graphs computed with any combination of variables. And, to share our data, we can easily script the software to transform our database in conformance with the TEI or any other schema.

Before the invention of *Callimachus*, the first version of the hypertext edition of Faulkner's novel used HTML and JavaScript to visualize the complicated and apparently disordered narrative in the book's first chapter. Faulkner, telling this part of the story through the mind of an 'idiot', normally provides only obscure clues to mark the mnemonic flashing from one event to another. The narrative does not follow the chronological sequence of events in the novel. However, using HTML and JavaScript to tag each event, we built an interface that links events in the narrative sequence with events in a chronologically correct version of the text. For the first time, readers could reorient themselves in the chronology by clicking a button, leaving behind the much more confusing original narrative.

The hypertext edition helped us clarify our understanding of the novel and yielded some surprising results. We knew that the 'idiot', narrator, named Benjy, would relive an event (such as his grandmother's death), would trigger a sequence of flashbacks, and would often repeatedly return to that initial event. Benjy's memory of his grandmother's death is interrupted 17 times by other flashbacks. When we isolate this event from the interruptions, we notice that it is transmitted chronologically. Hidden in the chaos of so many relived events are small, coherent, chronological narratives.

This archive (<http://www.usask.ca/english/faulkner>) was recently called "one of the best applications of the true potential of hypertext to date" (Neyt 140). However, there were only a few surprises yielded through the approach of this early edition. In general, we knew what to expect; the manual markup in HTML and JavaScript made our innovative display possible, but without search functions or sophisticated computing, we had to draw graphs manually and show proportions based on manual word counts. At the ACH/ALLC conference in 2001, we saw the promise of the TEI and, very slowly, began rethinking our approach. The *Callimachus* archive structure is the result.

The point of *Callimachus* (named after the Greek poet and grammarian who was the chief librarian at Alexandria) is to free Humanities researchers from the burden of having to specify their destination before starting their journey. And we do so in

a way that allows each participating project to take advantage of analysis tools that might have been originally created for a different text.

In addition to providing a universally accessible web-based editing infrastructure, *Callimachus* offers powerful analysis tools: on-the-fly visualization to produce graphs of the relationships inherent in the text; data mining to help identify textual relationships that were not immediately apparent; and translation to allow the user to transform the text into arbitrary formats (such as HTML, TEI or other XML schema) for exchange with other parties.

*Callimachus* is designed to grow and adapt with the user, but without invalidating his or her previous work. The user does not have to learn another markup language or data formalization in order to begin exploring the text with state-of-the-art analysis tools. We prefer to leave the construction of hierarchical representations schemes until after we've learned what those relationships are, rather than presupposing what we are going to find in order to begin.

## Bibliography

Neyt, Vincent. "Review of Stoicheff, Muri, Deshayé, et al. (eds.): *The Sound and the Fury: A Hypertext Edition*." *Literary and Linguistic Computing* 19.1 (2004): 137-143.