



Exploiting Lightweight Structure in Text: Multiple-Selection Editing and Outlier Finding

Robert Miller

Carnegie Mellon University

Thursday February 28

3:00PM

Olsen 311

Multiple selections, though heavily used in file managers and drawing editors, are virtually nonexistent in text editors. This talk will describe how multiple selections can be used to automate repetitive text editing. A multiple selection is inferred from positive and negative examples given by the user. The selection is then used for typing, deleting, copying, pasting, or other editing. Multiple selection editing has been evaluated by user studies and shown to be fast and usable by novices. Selection inference required only 1.26 examples per selection in the user study, closely approaching the one-example ideal.

When users handle large amounts of data, however, errors can be hard to notice. "Outlier finding" is a new way to reduce errors by drawing the user's attention to inconsistent data that may indicate errors. I have developed an outlier finder for text that can suggest both false positives and false negatives in a multiple selection. When integrated into the multiple-selection editor and tested in a user study, outlier finding reduced errors.

To be effective, selection inference and outlier finding depend on "lightweight structure," a new way to represent and manipulate text structure with an extensible, reusable library of patterns and parsers. Lightweight structure has applications not only to text editing, but also to web browsing, program transformation, and semistructured databases.

These techniques are implemented in LAPIS, a freely-available, open-source text editor/web browser written in Java (<http://www.cs.cmu.edu/~rcm/lapis>).