

File comparison algorithms and utilities 'diff', 'patch' and 'diff3' are widely used in programming for the purpose of code comparison, and in many version control systems. Despite the usefulness, the differences and patches produced by the tools are strictly line-oriented, which complicates processing of differently formatted data, such as free flowing text, markup, and various other formats where line breaks are not crucial. This thesis describes and implements a customizable version of these tools, which allows the user to specify an arbitrary tokenization of the input, thus allowing easy diffing, patching and change-merging of content not supported by the traditional diff. Additionally, the thesis describes a newly appearing challenge of managing the whitespace in the patches, and demonstrates the functionality on a practical use-case that can not be performed with the current diff utilities.