Subject: Re: String near match algorithm
Posted by koldo on Mon, 28 Dec 2009 15:25:58 GMT
View Forum Message <> Reply to Message

Hello all

Definitely Didier String comparison and BSDiff binary file comparison are not comparable, but indeed they are interesting.

Implemented BSDiff in Functions4U, there have been compared original and slightly changed files: theide.exe, a .xls file, a .doc file and a .txt file.

For all of them following appears the size of new file, compressed and bsdiff patch.

This is not a serious benchmark but it seems that, if size really matters, the patch files are much smaller than the .7z compressed files, so BSDiff is doing it very well for binary and for text files too (200 times better in this case) .

See how:

- TheIde.exe size comparison
New version: 4,854,784 bytes
New version .7z compressed: 2,032,605 bytes (42% size reduction)
BSDiff patch file: 284,739 bytes (6% size reduction)
-> So BSDiff file is 7 times smaller than .7z file

- .xls file size comparison
New version: 10,399,232 bytes
New version .7z compressed: 692,397 bytes (7% size reduction)
BSDiff patch file: 3,744 bytes (0.04% size reduction)
-> So BSDiff file is 184 times smaller than .7z file

- .doc file size comparison
New version: 259,072 bytes
New version .7z compressed: 150,212 bytes (58% size reduction)
BSDiff patch file: 3,123 bytes (1% size reduction)
-> So BSDiff file is 48 times smaller than .7z file

- .txt file size comparison
New version: 1,248,891 bytes
New version .7z compressed: 34,461 bytes (3% size reduction)
BSDiff patch file: 174 bytes (0.01% size reduction)
-> So BSDiff file is 198 times smaller than .7z file

Best regards
Koldo