CODE DIFF TO PDF

Compared Two Versions of Source Code Listings

Author - me@example.com

Sample
Diff Compiled on September 8, 2019

CONTENTS

Contents

1	Introduction	2
2	src2pdf.tcl	2
3	Removed c file.c	12
4	Changed Path tcl-file.tcl	12
5	Changed Path js-file.js	13
6	Added sql-file.sql	13

1 Introduction

This is an example file that include source code listings from a text file. Red means that the code was deleted. Blue means that code was added.

2 src2pdf.tcl

File path: /home/betsalel/CygwinDocuments/code-diff-to-pdf/src2pdf.tcl

```
#!/usr/bin/tclsh
   # Filename: src2pdf.tcl
   # Copyright (c) 2019, Betsalel (Saul) Williamson, Jordan Henderson (the Authors)
   # All rights reserved.
   # Redistribution and use in source and binary forms, with or without
   # modification, are permitted provided that the following conditions are met:
        * Redistributions of source code must retain the above copyright
          notice, this list of conditions and the following disclaimer.
        * Redistributions in binary form must reproduce the above copyright
          notice, this list of conditions and the following disclaimer in the
          documentation and/or other materials provided with the distribution.
        * Neither the names of the Authors nor the
          names of its contributors may be used to endorse or promote products
14
          derived from this software without specific prior written permission.
15
   # THIS SOFTWARE IS PROVIDED BY THE Authors 'AS IS', AND ANY
   # EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED
   # WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE
   # DISCLAIMED. IN NO EVENT SHALL THE Authors BE LIABLE FOR ANY
   # DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES
   # (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES;
  # LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND
   # ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT
   # (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS
   # SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
27
   #
28
   # Notes:
   # 2019-08-08 Initial creation
   # 2019-08-12 Ran dos2unix on latex template and code example files
     2019-08-18 Added removal or keeping of tmp folder with flag '-t'
32
              Cleaned up line breaks in caption text generation
33
              Added option to output to file with flag '-o fileName'
34
              Cleaned up comments and other misc visual things in script
35
              Added output for progress tracking
              Renamed option variables
   # 2019-08-30 Added input from text file to allow versions of code to be compared.
39
             Fixed issue with calling script from other locations.
   # 2019-08-30.2 Changed logic for generating added, removed, and change paths.
   # 2019-09-08 Changed file name to src2pdf.tcl from diff-src.tcl
              Added clean up proc to handle removing output in case of early exit
43
   # the purpose for this diff tool:
   # take as input files of version X (previous) and of version Y (current)
  # build latex files with sources as listings
  # run latexpand to inline sources
49 # run latexdiff on the latex files version X to Y
```

```
_{50} # build and output the diff as a PDF
51 # exit code 0 success
52 # exit code 1 error, no PDF
# exit code 2 warnings, PDF generated, but there may be issues
54 # exit code 3 script terminated unexpectedly
  # assumptions:
   # 1. files will follow their extensions and the syntax highlighting will not care about RPL
58 # 2. the resultant pdf will have content in the format: title page, TOC, listings with diffs
59 # 3. the input file contains pairs of files with the first file being the previous version
60 # 4. blank lines in the input file means the file was added or removed
# to find files: find . -regex ".*[.][h]?[st]?[cjmq]?[csl]"
64 #
# features that would be nice to have:
# parse headings of file (i.e. license info) to remove common headings and display in introduction
# have a separate TOC of diffs
68 # list of acronymns / dictionary used in code
69 # rolling change list sorted by date
70 # support for comparisons for more than one version
71 # choosing changes by major, minor, or build numbers
# parsing out comments into preabmles
_{73} # combination with Arros system to append figures and descriptions connected with source code
74 # add user information about the changes
75 # compare with diffs and blame for Fossil, Git and other popular systems
76 # right now, this tool is a stepping stone towards generating a PDF version of the complete system
        → modeled in Arros
   # we want to see from version to version what source code is being changed without care for who did
   # in the future, when users are integrated into the system, it will be possible to concern
       → ourselves with tracking who inputs what
   # preamble in latex file
   # don't include titles and latex things that aren't needed (bib, list of tables and figures)
   # run Expect spawn on PDF generation to interact with Latex and output latexdiff command
   # explore the difference between running pdflatex and latexmk, use regular latexdiff and then run
        \hookrightarrow latexmk
83 #
84 # Common errors with latexdiff and latex:
85 # LaTeX Error: File 'ulem.sty' not found.
# missing tex package -> install with 'tlmgr install PGK'
                   -> For SUSE linux use YaST. LaTex packages are prefixed with 'texlive'
87 #
88 # WARNING: Inconsistency in length of input string and parsed string
89 # Issue on Suse linux platform with texlive installed
  # latexdiff at older version. Update to version >= 1.3.
   # Download and install from https://www.ctan.org/tex-archive/support/latexdiff
   # Error with 'src2pdf.tcl' as input due to puts $fp "\\end\{lstlisting\}"
   # Latex assumed end of listing when the '{'s were not escaped
  # see https://blog.tcl.tk/1246 for Tcllib installation
96 package require Tcl 8.6
   package require cmdline
   package require fileutil
   set exitCode 3;# default unexpected exit
100
101
   set usage ": src2pdf.tcl \[-d\] -i source-list-file.txt \[-o path-to-output.pdf\] \[-t\] \[-v\]\
102
        → noptions:"
```

```
set parameters {
       {i.arg ""
                              "Required argument. Source list file. Contains list of code files. Relative
105
           \hookrightarrow paths must be relative to the file.\n\tFor regular input, each line should contain
           \hookrightarrow relative paths to source files.\n\tFor diff file, the format of file is that odd lines
           \hookrightarrow contain the previous versions of the source code file and even lines contain the next
           \hookrightarrow version. Empty lines mean that the file was removed or added.\n\tDefault:"}
              "./output.pdf" "Optional. Path and filename for result.\n\tDefault: "}
       {o.arg
       {d
                           "Use diff for input source list file."}
       ſν
                        "Optional. Verbose output."}
108
       {t
                        "Optional. Keep temporary Latex output."}
109
    if {[catch {array set options [cmdline::getoptions ::argv $parameters $usage]}]} {
       puts [cmdline::usage $parameters $usage]
114
        exit $exitCode
    } else {
116
       if { $options(v) } {
          puts "Verbose mode on."
118
          set verboseOutput 1
          parray options
      } else {
          set verboseOutput 0
124
      if { $options(t) } {
         puts "Keeping temporary LaTeX output."
         set keepTmpOutput 1
128
      } else {
          set keepTmpOutput 0
129
130
       if { $options(d) } {
          puts "Performing code src diff to pdf."
          set diffMode 1
134
      } else {
         puts "Performing regular src to pdf."
136
          set diffMode 0
      }
138
140
       if {[string length $options(o)] > 0} {
          set outputFile [file normalize $options(o)]
141
         puts "Output file set:\n$outputFile"
142
      } else {
143
          set outputFile [file normalize "./output.pdf"]
144
      }
145
       if {[string length $options(i)] > 0} {
147
          set inputFile [file normalize $options(i)]
148
         puts "Input file set:\n$inputFile"
149
      } else {
150
         puts "Missing source list file.\n\n"
          puts [cmdline::usage $parameters $usage]
153
          exit 1
154
    }
155
156
    proc isLink { aFile } {
157
       return [expr ! [catch {file readlink ${aFile}}]]
```

```
}
160
    if { ! ${tcl_interactive} } {
161
        ;# not being interactive is not enough!
162
        ;# we could be being included which makes us a script
        ;# in which case argv0 would not apply
       set ::myName [info script]
        ; #set myName ${argv0}
    } else {
167
       set ::myName [info script]
168
169
    if { [isLink ${::myName}] } {
170
       set ::myName [file readlink ${::myName}]
    set ::myName [file normalize ${::myName}]
    while {! [file isdirectory $::myName] && ! [catch {file readlink ${::myName}} newName] } {
174
       set ::myName ${newName}
   }
    if { ! [file isdirectory ${::myName}] } {
       set ::myBase [file dirname ${::myName}]
178
    } else {
       set ::myBase ${::myName}
180
181
182
    if { $verboseOutput } {
183
184
       puts "Base:\n${::myBase}"
185
186
    proc printList {list} {
187
      puts "Total files: [llength $list]"
188
      foreach elem $list {
190
         if {[string length $elem] > 0} {
            puts "Opening $elem ..."
            set fp [open $elem r]
193
            set fileData [read $fp]
194
            puts $fileData
195
            close $fp
196
            puts "Closed elem\n'"
         }
199
      }
200
      return
201
   }
202
203
    puts -nonewline "Preparing LaTeX files...
    flush stdout
206
    proc prepTempDir {} {
207
      global ::myBase
208
      global diffMode
209
       ;# del prev, set magic string to avoid collision with existing dir for rm -rf
212
       ;# TODO redo magic number if there was a collision and the tmp dir with
      ;# that number already exists.
      set magicNumber "latex-template-[expr {int(rand()*99999999999) + 1}]"
214
                        "tmp-$magicNumber"
      set tmpDir
      exec rm -rf $tmpDir ;# this is dangerous if $tmpDir is set to another directory
      exec mkdir -p $tmpDir
```

```
set latexPrevDir "$tmpDir/latex-prev/"
219
      set latexCurDir "$tmpDir/latex-cur/"
220
      if { $diffMode } {
         ;# create latex and reference locations
         exec cp -R ${::myBase}/latex-template $latexPrevDir
         exec cp -R ${::myBase}/latex-template $latexCurDir
226
         exec cp -R ${::myBase}/latex-template $tmpDir
228
229
      return [list [file normalize $latexPrevDir] [file normalize $latexCurDir] [file normalize
           → $tmpDir]];
231
232
    set latexRoots [prepTempDir]
233
    set prevRoot [lindex $latexRoots 0]
234
    set curRoot
                  [lindex $latexRoots 1]
    set tmpDir
                  [lindex $latexRoots 2]
    proc cleanUp {keepTmpOutput} {
238
      global tmpDir
239
240
      if {[string length $tmpDir] > 0} {
241
242
         if { $keepTmpOutput } {
            puts "Keeping temporary LaTeX output at:\n$tmpDir"
244
            puts -nonewline "Removing temporary LaTeX output..."
245
            flush stdout
246
            exec rm -r $tmpDir
247
            puts " Done."
248
         }
      }
250
   }
251
252
    proc escapePathForLatex {s} {
253
      global verboseOutput
254
255
      if { $verboseOutput } {
257
         puts "String before escaping: "
         puts $s
258
      }
259
260
       ;# special characters in latex must be escaped
261
       ;#\{}&^$%;#_
      regsub -all {([\\\{\}&\^\$%#_~])} $s {\\\1} s
264
265
       ;# add latex line breaks into paths
266
      regsub -all \{([\]?[\/\&\^\$\#_^\-;'!])\} $s \{\1\ s
267
268
      if { $verboseOutput } {
         puts "String after escaping: "
270
         puts $s
271
      }
272
      return $s
273
   }
274
```

275

```
proc getLanguageFromFile {fileName} {
       global verboseOutput
277
278
       set fe [file extension $fileName]
      if { $verboseOutput } {
         puts "File extension: $fe"
281
      set language ""
284
      switch $fe {
285
          .c {
286
            set language "C"
287
         }
          .sql {
290
             set language "SQL"
291
          .js {
292
             set language "ECMAScript"
293
         }
294
          .tcl {
             set language "tcl"
297
         default {}
298
299
300
301
      return $language
302
303
    proc addCodeToFile {texFilePath sectionText bodyText codeFilePath} {
304
      global verboseOutput
305
307
       set fp [open $texFilePath a+]
      puts $fp "\\section\{$sectionText\}"
      puts $fp "$bodyText"
       ;# puts $fp "\\label{sec:$captionText}"
       ;# label=code:$fileName
311
       ;# TODO ensure that the equivalent files have the same label to allow references to work
312
           \hookrightarrow correctly
      set fp1 [open $codeFilePath r]
313
314
      set fileData [read $fp1]
315
      close $fp1
      puts $fp "\begin\{lstlisting\}\[language=[getLanguageFromFile $codeFilePath]\]"
      puts $fp $fileData
317
      puts $fp "\\end\{lstlisting\}"
318
      close $fp
319
      if { $verboseOutput } {
         puts "Wrote data into tex file..."
322
323
324
      return:
    }
325
326
327
    proc processDiffInputFile {inputFile prevRoot curRoot} {
328
      global verboseOutput
329
       if { $verboseOutput } {
330
         puts "\nProcessing input file..."
331
332
      if {[catch {exec dos2unix -q $inputFile} result]} {
```

```
puts "Error with converting line endings to Unix."
            puts "Information about error: $::errorInfo"
335
            puts $result
336
            cleanUp 1
            exit 1
      }
339
      set fpInputFile [open $inputFile]
      set inputFileLines [split [read $fpInputFile] "\n"]
342
      if { $verboseOutput } {
343
         puts "File lines: [llength $inputFileLines]"
344
         puts "Input files $inputFileLines"
345
      }
347
      close $fpInputFile;
348
349
       ;# Change path to the input file source to ensure relative paths work
350
      set curDir [exec pwd];
351
      cd [file dirname $inputFile]
      for { set i 0} {$i < [llength $inputFileLines]} {incr i 2} {</pre>
          ;# check to see if file was added/removed
356
         set norm1 [lindex $inputFileLines $i]
357
         set norm1 [file normalize $norm1]
358
359
         set norm2 [lindex $inputFileLines [expr $i+1]]
361
         set norm2 [file normalize $norm2]
362
          ;## both norm1 and norm2 cant be empty
363
          set norm1Empty [expr ![string length $norm1]]
364
         set norm2Empty [expr ![string length $norm2]]
365
          if {$norm1Empty && $norm2Empty} {
            puts "Error with input file. Cannot have two blank lines $i and [expr $i+1]."
               cleanUp 1
369
            exit 1
         }
372
          if {!($norm1Empty)} {
374
            set prevPathForLatex [escapePathForLatex $norm1]
            set prevFileName [file tail $norm1]
         if {!($norm2Empty)} {
378
            set curPathForLatex [escapePathForLatex $norm2]
            set curFileName [file tail $norm2]
         }
381
382
          ;# if added grab second line and insert to first that this was added
383
          if {$norm1Empty} {
384
            if { $verboseOutput } {
385
               puts "added:\n$norm2"
            }
388
            addCodeToFile "$prevRoot/body/01-code.tex" "Added $curFileName" "New file: $curPathForLatex

→ " "$norm2"

            ;# cur
            addCodeToFile "$curRoot/body/01-code.tex" "Added $curFileName" "New file: $curPathForLatex"
```

```
\hookrightarrow \texttt{"$norm2"}
392
          ;# if removed grab first line and insert to next that this was removed
393
         } elseif {$norm2Empty} {
394
             if { $verboseOutput } {
               puts "removed:\n$norm1"
            }
             ;# prev
            addCodeToFile "$prevRoot/body/01-code.tex" "Removed $prevFileName" "Removed file:
399
                 \hookrightarrow $prevPathForLatex" "$norm1"
             :# cur
400
            addCodeToFile "$curRoot/body/01-code.tex" "Removed $prevFileName" "Removed file:
401
                 \hookrightarrow $prevPathForLatex" "$norm1"
403
          ;# if directory path doesn't equal then add change path
          } elseif {[string compare [file dirname $norm1] [file dirname $norm2]] != 0} {
404
            if { $verboseOutput } {
405
               puts "path changed:\n$norm1\n$norm2"
            }
             ;# prev
             addCodeToFile "$prevRoot/body/01-code.tex" "Changed Path $prevFileName" "Changed file path
                 ;# cur
            addCodeToFile "$curRoot/body/01-code.tex" "Changed Path $curFileName" "Changed file path
411
                 \hookrightarrow from: prevPathForLatex to <math>curPathForLatex" "$norm2"
412
         } else {
413
             ;# prev
414
            addCodeToFile "$prevRoot/body/01-code.tex" "$prevFileName" "File path: $prevPathForLatex" "
                 \hookrightarrow $norm1"
             ;# cur
            addCodeToFile "$curRoot/body/01-code.tex" "$curFileName" "File path: $curPathForLatex" "
                 \hookrightarrow $norm2"
         }
      }
      cd $curDir
421
       if { $verboseOutput } {
422
         puts "Finished looping through files..."
423
424
      }
425
426
      return;
   }
427
428
    proc processRegularInputFile {inputFile tmpDir} {
429
       global verboseOutput
       if { $verboseOutput } {
432
         puts "\nProcessing input file..."
433
434
      if {[catch {exec dos2unix -q $inputFile} result]} {
          puts "Error with converting line endings to Unix."
436
            puts "Information about error: $::errorInfo"
            puts $result
438
            cleanUp 1
439
             exit 1
440
      }
441
442
       set fpInputFile [open $inputFile]
```

```
set inputFileLines [split [read $fpInputFile] "\n"]
      if { $verboseOutput } {
         puts "File lines: [llength $inputFileLines]"
         puts "Input files $inputFileLines"
447
      }
      close $fpInputFile;
       ;# Change path to the input file source to ensure relative paths work
452
      set curDir [exec pwd];
      cd [file dirname $inputFile]
454
      for { set i 0} {$i < [llength $inputFileLines]} {incr i} {</pre>
         ;# check to see if file was added/removed
458
         set norm1 [lindex $inputFileLines $i]
459
         set norm1 [file normalize $norm1]
460
         ;## both norm1 and norm2 cant be empty
462
         set norm1Empty [expr ![string length $norm1]]
         if {$norm1Empty} {
465
            puts "Error with input file. Cannot have blank line $i."
466
            exit 1
467
         }
468
469
         set pathForLatex [escapePathForLatex $norm1]
471
         set fileName [file tail $norm1]
         addCodeToFile "$tmpDir/latex-template/body/01-code.tex" "$fileName" "File path: $pathForLatex"
472
              \hookrightarrow "$norm1"
      }
474
      cd $curDir
      if { $verboseOutput } {
477
         puts "Finished looping through files..."
478
479
480
      return;
481
482
   }
483
   if { $diffMode } {
484
      processDiffInputFile $inputFile $prevRoot $curRoot
485
   } else {
      processRegularInputFile $inputFile $tmpDir
487
   }
488
   proc inlineLatex {latexRoot} {
490
      global verboseOutput
491
      set prevDir [pwd]
492
      set inlineMain "main2.tex"
493
      set defaultMain "main.tex"
494
      set defaultBib "references.bib"; ;#TODO still not sure if this needs to be a bbl or bib file
      cd $latexRoot
      exec latexpand --keep-comments --expand-bbl $defaultBib $defaultMain > $inlineMain
497
      if { $verboseOutput } {
498
         puts "latexpand --keep-comments --expand-bbl $defaultBib $defaultMain > $inlineMain"
499
      }
500
```

501

```
cd $prevDir
      return $inlineMain;
503
504
505
    if { $diffMode } {
       ;# inline latex to make it easier to do the diff
      set inline [inlineLatex $prevRoot]
      set inline2 [inlineLatex $curRoot]
510
      set inline [inlineLatex "$tmpDir/latex-template"]
511
512
513
    if { $verboseOutput } {
514
515
      puts "Finished inlining files..."
516
517
    puts " Done."; ;# preparing latex files
518
519
    # generate pdf
520
    # uses latexdiff-vc run in the temp folder
    cd "$tmpDir"
523
524
    # this was the original diff latex, but produces latex output
525
    # the following program latexdiff-vc is preferred because it creates direct to PDF
526
    # manual diff to .tex
   # exec latexdiff "$prevRoot/$inline" "$curRoot/$inline2" > "$tmpDir/main2-diff.tex"
    # spawn latexdiff-vc --verbose --pdf "$prevRoot/$inline" "$curRoot/$inline2"
530
   puts "If the following step hangs (more than 2 minutes) enter 'x' and hit enter."
531
   puts -nonewline "Generating PDF...
   flush stdout
    if { $diffMode } {
       ;# inline latex to make it easier to do the diff
536
       ;# TODO investigate if there will be security issues with input sources being malicious from TCL
537
           → or Latex
      set cmd "latexdiff-vc --pdf \"$prevRoot/$inline\" \"$curRoot/$inline2\""
538
      set tmpOutputPdf "main2-diff.pdf"
539
   } else {
541
      set cmd "latexmk -pdf \"$tmpDir/latex-template/$inline\""
      set tmpOutputPdf "main2.pdf"
542
   }
543
544
545
    if { $verboseOutput } {
      puts "Running command to generate PDF:"
      puts $cmd
548
549
550
    if {[catch {exec {*}}$cmd} result]} {
551
552
       ;# report errors and warnings,
553
       ;# ideally there shouldn't be any because we are generating everything here
554
       ;# but LaTex generation has many warnings that are a nuisance
555
       ;# input code files shouldn't break anything
556
      ;# todo change to expect script to allow interaction with above program in case something
557
           \hookrightarrow happens
       ;# as of now the program halts if there is a latex warning that requries user engagement
```

```
puts "\nInformation about error: $::errorInfo\n\n"
       if {[file exists "$tmpDir/$tmpOutputPdf"]} {
561
         puts "Warning on pdf generation!"
562
         set exitCode 2
       } else {
         puts stderr "Error!"
         set exitCode 1
567
568
    } else {
569
       set exitCode 0
570
   }
571
573
    puts " Done."; ;# generating PDF
574
    if {[file exists "$tmpDir/$tmpOutputPdf"]} {
      exec mv "$tmpDir/$tmpOutputPdf" $outputFile
      if {[file exists "$outputFile"]} {
         puts "PDF created at:\n$outputFile"
580
         set exitCode 0
581
      } else {
582
         puts "Error generating output..."
583
         set exitCode 1
584
   } else {
587
      puts "Error generating output..."
588
      set exitCode 1
    cleanUp $keepTmpOutput
593
    exit $exitCode
594
```

3 Removed c file.c

Removed file: /home/betsalel/CygwinDocuments/code-diff-to-pdf/example/v1/c file.c

```
// Version 1 of file

#include <stdio.h>
int main( int argc, const char* argv[] )

{
    // this is an added comment
    printf( "I changed too!\n");
}
```

4 Changed Path tcl-file.tcl

Changed file path from: /home/betsalel/CygwinDocuments/code-diff-to-pdf/example/v1/folder with space at end/tcl-file.tcl to /home/betsalel/CygwinDocuments/code-diff-to-pdf/example/v2/

folder with space at end/tcl-file.tcl

```
#!/usr/bin/tclsh

puts "Hello World"

puts "One last change..."

# example comment

DIF >

puts "A very long line of text. A very long line of text. A very long line of text.

A very long line of text. A very long line of text. A very long line of text.

A very long line of text. A very long line of text. A very long line of text.

A very long line of text. A very long line of text. A very long line of text.

A very long line of text. A very long line of text. A very long line of text.
```

5 Changed Path js-file.js

Changed file path from: /home/betsalel/CygwinDocuments/code-diff-to-pdf/example/v1/folder\$with#Chars%;'!&/js-file.js to /home/betsalel/CygwinDocuments/code-diff-to-pdf/example/v2/folder with space at end/js-file.js

```
console.log( 'Hello, world!' );
// JUST FOR THE HECK OF IT! I'LL CHANGE TOO :-)
```

6 Added sql-file.sql

 $New \quad file: \quad /home/betsalel/CygwinDocuments/code-diff-to-pdf/example/v2/folder\$with\#Chars\%;'1\&/sql-file.sql$

```
DROP TABLE HELLO_WORLD;

CREATE TABLE HELLO_WORLD(

ID INT NOT NULL,

PRIMARY KEY (ID)

);
```