Friday, 2 September 2011

Multiple-file LaTeX diff

One major pain point I've faced while finishing my thesis is finding a nice way to show my supervisor exactly what revisions I'd made since our last chat. One thing that made a big difference to this process was to use latexdiff. One of the limitations of that tool is that it doesn't support documents that span multiple files, a feature used heavily in my thesis template.

The work-around was to write a small python script to glue the files back together, and so here is a script to flatten LaTeX files

```python
#!/usr/bin/python
import sys
import os
import re

inputPattern = re.compile('\\input{(.*)}')

def flattenLatex( rootFilename ):
    dirpath, filename = os.path.split(rootFilename)
    with open(rootFilename,'r') as fh:
        for line in fh:
            match = inputPattern.search( line )
            if match:
                newFile = match.group(1)
                if not newFile.endswith('tex'):
                    newFile += '.tex'
                flattenLatex( os.path.join(dirpath,newFile) )
            else:
                sys.stdout.write(line)

if __name__ == "__main__":
    flattenLatex( sys.argv[1] )
```

Which ends up being called like this:

```
# merge multiple files into the old and current versions of the document
flatten-latex ${DIFFTREE}/thesis.tex > old.tex
flatten-latex ${WORKINGTREE}/thesis.tex > cur.tex

# produce the marked up document
latexdiff old.tex cur.tex > tmp.tex

# fix line ending problem introduced by latexdiff
sed 's/^M//' tmp.tex > diff.tex
```

Posted by andrew at 17:36

Labels: python latex

3 comments:

Jörg 15 February 2012 at 23:56

Hi, this script is exactly what I need. But I have no experience in Python. I have installed Python, but I am at a less what to do with the script and how to call it. Could you give me a little tutorial please? Thank you!

Reply

סולי וישקאוצן28 October 2012 at 19:35

latexdiff also has a --flatten parameter.
But sadly it doesn't work recursively, only for the first level of includes
.
Another solution I would suggest is to write a script that diffs all files between two directories representing two revisions, and outputting a difference directory.

Reply

KGs~ 28 August 2013 at 00:25

thanks a lot from Paraguay!

It just work :)