

File 20100608.0930: Notes from Probabilistic Redaction status meeting this morning:

The 5.0 regression test after Beta 1 CT&E is not going well this week. The entire IV&V team is in Denver, and they worked all weekend, some of them (Kori Phillips mostly, as Test Director) worked all night. The reason is said to be that they took out all the web GUI stuff that worked and replaced it two months ago with untested new web GUI code.

On the subject of Probabilistic Redaction, Tom reports that he is on schedule to have the 'do something with it' part of the prototype working by the end of this month. It will replace redacted phrases with X's or black boxes or something.

I reported on some experiments with PDF and PostScript. I generated a simple one-page document in Microsoft Word and printed it to CutePDF Writer. Next, I installed Ghostscript in Cygwin (had to update Cygwin setup first), which should provide `pdf2ps` as a shell script. I want to run that on the simple PDF file and see if I can extract the text, the margin dimensions, and the paragraph dimensions from the generated PostScript.

Next, I want to parse out the page numbers, page header and footer, and other page details.

Finally, I want to try re-generating the PostScript code with redacted text in it, and finally run `ps2pdf` to regenerate a PDF file. I will continue the experiments tomorrow (it's late and I am typing up today's notes before I got to sleep). Reading Group tomorrow early.

References