WHAT IS CS35L HOMEWORK 3

An informal specification

Here's the contents of files a and b:

File a	File b
A	B
C	C
D	D
E	E
F	F
G	Z

If you run the command "comm a b", here is the output:

A		
	B	
		C
		D
		E
		F
G		
	Z	

Column 1 contains strings unique to file 1, column 2 contains strings unique to file 2, and column 3 contains strings that are in both files.

Running the comm command with any of the following flags "-1" "-2" "-3" will remove the corresponding column from the output. So "comm -2 a b" would remove the second column from the output, producing this:

A	
	C
	D
	E
	F
G	

There's one thing about these files and the output of comm: Their content is all alphabetically sorted. According to the spec, if input isn't sorted, then behavior is undefined.

Eggert wants us to add a flag “-u” that specifically handles cases in which the input files are not sorted. The logic for the “-u” flag is completely different, so you can just write the not “-u” part without worrying about things being unsorted.

Here’s my suggested pseudocode:

Main:

    If “-u” flag

        Do “-u” flag stuff

    Else

        Do not “-u” flag stuff

In other words, just write two different code paths. It’s better that way, I promise.

For the “-u” flag, the second column comes after the first and third columns, so even for the sorted files from before the output would be different:

A

C
D
E
F

G

B
Z

Since the files are assumed to be not sorted due to “-u”, the logic is going to be different. Eggert wants “-u” to have all of the output be in the order of the first file, so for the files below

File a	File b
A	G
C	B
D	D
Z	E
G	A
B	B

the output would be:

A

C

D

Z

G
B

E
B

If you call comm or comm.py with “-“ as one of the filenames, then you read from stdin instead of reading from a file. You don’t really have to change much code, instead of doing

contents = file.readlines() or whatever method you were using, just use

contents = sys.stdin.readlines()