

2 つの CSV ファイルの自然結合

問題

カレントディレクトリにある 2 つの CSV ファイルを自然結合した CSV ファイルを作成する Python の関数 `natural_join(infile1, infile2, outfile)` を実装せよ。[10 点]

仕様

引数:

- `infile1` 1 つめの CSV ファイル名 (中身は UTF-8)。ヘッダー行はカンマ区切りされた属性名。
- `infile2` 2 つめの CSV ファイル名 (中身は UTF-8)。ヘッダー行はカンマ区切りされた属性名。
- `outfile` 出力先の CSV ファイル名 (中身は UTF-8)。ヘッダー行はカンマ区切りされた属性名。

戻り値:

なし

2 つの CSV ファイルの自然結合とは、両方のファイルに共通の属性について、その値が等しい行どうしを横に連結した行からなる CSV ファイルである。ただし、共通の属性については出力ファイルでは 1 つめの CSV ファイルに由来する値のみ出力するものとする。

[11 月 9 日追記]

- (1) 出力 CSV ファイルにおける属性の順番は、2 番目の CSV ファイル中の共通属性を除外する以外は、与えられた CSV ファイルにある順番を守ってください。
- (2) 自然結合の際に、片方の CSV ファイルの行が、もう片方の CSV ファイルの複数行に対応することは無いものとします。

例 1: 共通属性が 1 つの場合

student.csv の中身

student_id	name
123	工大太郎
124	野々市花子
125	休学三郎

grade.csv の中身

student_id	hw1	hw2
123	8	10
124	10	9

natural_join('student.csv', 'grade.csv', 'output.csv')を実行後の output.csv の中身

student_id	name	hw1	hw2
123	工大太郎	8	10
124	野々市花子	10	9

例 2: 共通属性が 2 つの場合

student.csv の中身

class	no	name
2EP1	10	工大太郎
2EP1	35	野々市花子
2EP3	12	休学三郎

grade.csv の中身

record_id	class	no	hw1	hw2
1001	2EP1	10	8	10
1002	2EP1	35	10	9

natural_join('student.csv', 'grade.csv', 'output.csv')を実行後の output.csv の中身

class	no	name	record_id	hw1	hw2
2EP1	10	工大太郎	1001	8	10
2EP1	35	野々市花子	1002	10	9

natural_join('grade.csv', 'student.csv', 'output.csv')を実行後の output.csv の中身

record_id	class	no	hw1	hw2	name
1001	2EP1	10	8	10	工大太郎
1002	2EP1	35	10	9	野々市花子

注意

- 入力ファイルの最後には複数の空行があるかも知れないが、それは無視して処理すること。
- 逆に出力ファイルには空行を入れないこと。
- 出力ファイルにおける行の順番は特に指定しないが、属性名からなるヘッダ行は 1 行目に書くこと。
- 必要に応じて追加の関数を定義して用いてもよい。
- csv ライブラリをはじめとする標準ライブラリ (<https://docs.python.jp/3/library/>) は自由に使ってよい。
- **まずは共通属性が 1 つの場合にうまく動作させることを目標にすること。[この部分の配点を 8 点とします]**
→ [11 月 9 日追記] newton 課題 2 の `build_dict` に沿った方法で実装する場合は、共通属性の数について
の場合分けをせずに、まとめて面倒を見た方が見通し良くプログラムできると思います。

提出方法

e シラバスにて .py の拡張子のファイル 1 つをアップロード（課題のエントリーは登録済み）。

提出期限は 11 月 23 日（金）深夜を予定。

できる限り早く取りかかること！私がやっても 30-40 分かかりました。