

מבני נתונים ואלגוריתמים 88-280

תרגיל 4

תאריך הגשה: 2.12.17, שעה 23:55

הוראות הגשה:

יש להגיש את התרגיל דרך האתר <https://submit.cs.biu.ac.il/cgi-bin/welcome.cgi>

יש לציין בתחילת כל קובץ בהערה שם, מספר תרגיל ות"ז!

יש להגיש קובץ בשם ex2.py.

מיון קובץ csv

בתרגיל זה עליכם לממש אלגוריתם מיון אשר ממיין קובץ csv (Comma Separated Values) לפי שדה.

מה זה קובץ csv?

Comma-separated values (CSV) הוא צורה לשמירת טבלאות בטקסט פשוט. בפורמט זה כל רשומה מופרדת באמצעות שבירת שורה וכל רשומה מורכבת משדות המופרדים באמצעות תו מוגדר, כאשר בדרך כלל נעשה שימוש בפסיק או בטאב פורמט זה נקרא (tsv) בדרך כלל לכל התאים יש מספר שדות זהה. (לפי וויקיפדיה)

ניתן ליצא את הקבצים הללו לתוך תאים על ידי תוכנת גליון אלקטרוני, כמו אקסל, או להכניסם כמו שהם לתוך מסד נתונים. לרוב הם משמשים להעברת אינפורמציה מסד נתונים.

דוגמה:

כל שורה בקובץ ה CSV-ייתבת לתאר יישות נפרדת, כגון שם, מאכל אוהב וחיה אוהבה. בדוגמה הבאה השורה הראשונה היא הכותרת.

שם, מאכל אוהב, חיה אוהבה

יוסי, פיצה, כלב

משה, סושי, ארנב

אילנית, פלאפל, חתול

במילים אחרות: נרצה למיין לפי עמודה.

הוספתי קובץ לדוגמה. כמו כן, הוספתי קישור למידע על ספריית csv בפייתון.

מהלך התוכנית:

הכנתי קובץ csv שעליכם לקרוא ממנו שנמצא בשרת של האוניברסיטה. ניתן להניח שהקובץ אכן קיים בתיקייה. עליכם לקבל שדה ולמיין לפיו. אם השדה מורכב ממחרוזות, יש למיין בסדר אלפביתי (סדר לקסיקוגרפי). על המיון להיות מיון יציב. לא ניתן להניח שהשדה אכן קיים בקובץ. את התוצאות יש להדפיס על המסך (עם פוקדת print).

קלט: שם של קובץ אותו תצטרכו למיין, שם של שדה שלפיו נרצה למיין.

פלט: השורות של הקובץ ממוינות. יש להדפיס שדה1, שדה2, שדה3.....

אין להדפיס את הכותרת.

לא ניתן להניח שהשדה בקלט אכן קיים בקובץ! במידה והשדה לא קיים יש להדפיס: "Error" ולקבל שם של שדה נוסף.

ניתן להניח שלבסוף אתן שם תקין והתוכנית תסיים את ריצתה.

דוגמה :

נניח לקובץ הבא קוראים "csv.example":

והקובץ נראה כך:

	A	B	C
1	name	favorite food	age
2	moshe	pizza	21
3	adi	sushi	34
4	avi	sushi	15

עליכם לכתוב תוכנית שקוראת מהקובץ וממיינת לפי שדה כלשהו.

דוגמאות ריצה(שחור הוא הקלט, אדום הוא הפלט):

1.

example.csv,grade

Error

age

avi,sushi,15

moshe,pizza,21

adi,sushi,34

2.

Example.csv,favorite food

moshe,pizza,21

adi,sushi,34

avi,sushi,15

הגבלות:

1. אין להשתמש בפקודה sorted או sort.

2. על המיין להיות ב- $O(n \log n)$

3. על המיין להיות יציב.

4. אין להניח דבר על מספר השדות בקובץ!