

מכללה אקדמית הדסה

החוג למדעי המחשב

תרגיל #10 בקורס: תכנות מודולארי

קלט פלט ושימוש בקבצים חיצוניים

בתרגיל זה יש שלוש שאלות. עליכם לתכנת את תשובותיכם בקבצים ששםם הוא `ex10a.cc, ex10b.cc, ex10c.cc`.

תכנית a: הצפנת קובץ

בתכנית זאת יהיה עליכם לבצע הצפנה (פרימיטיבית) של אותיות לטיניות קטנות בלבד, כלומר בתחום `a..z`, המופיעות בקובץ (תווים שאינם אותיות לטיניות קטנות לא יוצפנו, הם יועתקו כמות שהם מהקלט לפלט, בפרט נכון הדבר לרווחים ומעברי שורה).

התכנית שלכם תקרא מהשתמש שמות של שלושה קבצים:

א. קובץ הכולל מפתח הצפנה

ב. קובץ הכולל קלט אותו יש להצפין

ג. קובץ עליו ייכתב הקלט המוצפן

הקובץ הכולל את מפתח ההצפנה יכלול זוגות של תווים: איזה תו יש להצפין, ובאמצעות איזה תו. דוגמה אפשרית לקובץ: `ae bc cd db ea`
קלט זה מורה ש:

א. את `a` יש להצפין ע"י החלפתה ב: `e`

ב. את `b` יש להצפין ע"י החלפתה ב: `c`

ג. את `c` יש להצפין ע"י החלפתה ב: `d`

ד. את `d` יש להצפין ע"י החלפתה ב: `b`

ה. את `e` יש להצפין ע"י החלפתה ב: `a`

נניח שהקלט תקין בכל מובן שהוא.

כמו כן, אות שלא הוזכרה בקלט, יש 'להצפין' ע"י החלפתה באותה אות, או במלים אחרות, אות זו אינה מוצפנת. (אות שלא מוצפנת גם לא יכולה להצפין אות אחרת, שכן עת נקבל אותה בטקסט המוצפן לא נדע האם היא מייצגת 'את עצמה' או מצפינה אות אחרת).

קובץ הקלט השני, יכלול טקסט אותו יש להצפין באמצעות מפתח ההצפנה, תוך שאת תוצאת ההצפנה יש לכתוב על קובץ פלט (וכן על הפלט הסטנדרטי).
לדוגמה: אם קובץ הקלט הוא:

`a bb ccc!`
`12 ddddd eeee x`

אזי קובץ הפלט יהיה:

`e cc ddd!`
`12 bbbbbb aaaa x`

פלט זה יישלח במקביל גם לפלט הסטנדרטי (ולא רק לקובץ הפלט).

הערות

- א. בכל השאלות בתרגיל, הניחו כי אורך שם קובץ הקלט יהיה קטן ממאה תווים.
ב. התיישרו עם פתרון בייס בשאלה: מה קורה עת המשתמש מזין שמות קבצים שגויים.

תכנית b: פענוח קובץ שהוצפן ע"י תכנית a

תכנית זאת קוראת מהמשתמש שמות של שני קבצים:

א. קובץ הכולל טקסט שהוצפן ע"י תכנית a.

ב. קובץ לא מוצפן, הכולל טקסט 'סתם'.

התכנית תקרא את הקובץ המוצפן, ועבור כל אות בתחום a..z בקובץ תספור כמה פעמים היא הופיע בקובץ. בדוגמה שלנו נקבל: {a= 4, b= 5, c= 2, d= 3, e= 1, x= 1}

עתה התכנית תקרא את הקובץ הכולל טקסט 'סתם' ותספור כמה פעמים מופיעה בו כל אות. למשל, אם הקובץ 'הסתמי' הינו:

aaa bbbbbb ccc ccc ddd ddd dd eee eee e x

אזי היא תגלה ש: {a= 3, b= 5, c= 6, d= 8, e= 7, x= 1}

עתה התכנית תניח שאם בטקסט הסתמי d הוא התו השכיח ביותר, ואם בטקסט המוצפן b הוא התו הנפוץ ביותר, אזי d הוצפן על-ידי b. באופן דומה: אם בטקסט הסתמי התו השני בשכיחותו הוא e, ואם בטקסט המוצפן התו השני בשכיחותו הוא a אזי אפשר להסיק ש: e הוצפן ע"י a. כלומר על סמך השכיחות היחסית של אותיות בטקסט הסתמי, ובטקסט המוצפן, התכנית תסיק איזה אות מצפינה איזה אות. כך התכנית תבנה את מפתח הפיענוח.

עתה התכנית תקרא פעם שניה את הטקסט המוצפן, תפענח אותו, ואת הפלט תשלח לפלט הסטנדרטי.

אם שני תווים מופיעים אותו מספר פעמים בקובץ הסתמי ובקובץ המוצפן (בפרט אפס פעמים) אזי נניח שהראשון ביניהם בסדר ה-ASCII מצפין את הראשון. לדוגמה: אם הקובץ המוצפן הוא: a a b c והקובץ הסתמי הוא: bbb אזי ניתן להסיק ש: a הצפין את b (ולכן את aa יש לפענח כ: bb). לגבי הפיענוח של b, c התכנית תניח ש: b הצפין את a, וש: c הצפין את עצמו, שכן השכיחות של שניהם בקובץ הסתמי היא זהה: אפס (לכן b יפוענח ע"י התו הראשון שלא פיענח שום תו אחר, כלומר: a, ו: c ע"י התו הבא שטרם פיענח אף תו אחר, כלומר: c).

הערה

ראוי לפתוח את הקובץ המוצפן פעם אחת בלבד בתכנית (ולא לפתחו, לסגרו ואחר כך לפתחו בשנית)

הסבר נוסף

כאמור: עלינו לפענח קובץ מוצפן.

איך נעשה זאת: ראשית, נקרא מקובץ רגיל (לא מוצפן) את השכיחות של האותיות בו. נניח שנגלה ש: A היא הכי שכיחה, D היא השנייה בשכיחותה, ו- C היא זו שמופיעה הכי פחות פעמים.

עתה נקרא את הקובץ המוצפן.

נניח שבו נגלה ש: X היא השכיחה ביותר W היא השנייה בשכיחותה ו- Z היא הכי פחות שכיחה.

על כן נניח ש: X מצפינה את A, W מצפינה את D, ו-Z מצפינה את C.

לכן כדי לפענח את הקובץ המוצפן, בכל מקום שנקרא את X נדפיס את A, בכל מקום שנקרא את W נדפיס את D, בכל מקום שנקרא את Z נדפיס את C, וזה יהיה הפיענוח.

הרעיון הוא כזה:
על כל תו בקובץ המוצפן, אנו שואלים: את מי הוא הצפין, או במלים אחרות: איך יש לפענח אותו?

אם המוצפן הוא: aa b c
והסתמי הוא: bbb

כלומר יש לנו שלושה תווים בקובץ המוצפן שיש לפענח:

מי שמופיע בשכיחות הגבוהה ביותר בקובץ המוצפן, יפענח ע"י מי שמופיע בשכיחות הגבוהה ביותר בקובץ הסתמי:
ולכן: a **הצפין** את b או את a יש לפענח ע"י b.
והמשפט:
(ולכן את aa יש לפענח כ: bb)
הוא נכון

כלומר את a בקובץ המוצפן אנו כבר יודעים איך יש לפענח: ע"י b.

עתה נשאל לגבי יתר התווים בקובץ המוצפן (c, b), את מי הם הצפינו, ולכן איך יש לפענח אותם:
את מי מצפין b שמופיע בקובץ המוצפן? או איך יש לפענח אותו?
b הצפין את a
. a הוא התו הראשון שלא קבענו מי הצפין אותו, או במלים אחרות את מי הוא מפענח.

ואת מי הצפין c שמופיע בקובץ המוצפן? במלים אחרות: איך יש לפענח אותו?
אנו כבר יודעים שאת a נפענח ע"י b, את b נפענח ע"י a, ולכן את c נפענח ע"י התו הראשון שלא מפענח עוד אף תו, כלומר את c נפענח ע"י c.

בכל השאלות בתרגיל, הניחו כי אורך שם קובץ הקלט יהיה קטן ממאה תווים

תכנית c: מיזוג של שני קבצים

מיזוג היא פעולה הקולטת שתי סדרות ממוינות, ומצרפת אותן לכדי סדרה ממוינת אחת (בזמן לינארי בגודל הקלט).

הסדרות	שתי	File1	File2	לדוגמה: מיזוג של (אחת בכל שורה)
		Dana 80 70 90	Rama 70	
5 7 7 12 14 15		Rama 60 80	Yosi 1 2 3 4	
6 8 12		Sam 6		

ייתן את הסדרה:

5 6 7 7 8 12 12 14 15

כתבו תכנית הקוראת מהמשתמש שמות של שני קובצי קלט, ושם של קובץ פלט. קובצי הקלט כוללים שמות של סטודנטים וציוניהם, והם **ממוינים לפי שם הסטודנט**.

שורה בודדת בכל קובץ תראה:

Dana 80 70 90

כלומר לתלמידה Dana יש שלושה ציונים: 80 70 90.

כל שורה בכל קובץ תכיל מחרוזת ראשונה שהינה מזהה הסטודנט (כדוגמת: Dana), ואחר סדרה באורך כלשהו של מספרים שלמים שמציינים את ציוני הסטודנט.

על התכנית למזג את שני הקבצים, באופן שאם סטודנט מופיע בשניהם אזי תהיה לו שורה יחידה בפלט. השורה תכלול ראשית את ציוניו כפי שהופיעו בקובץ הקלט הראשון, ואחר את הציונים מקובץ הקלט השני. את התוצאה יש לכתוב על קובץ הפלט, וכן על הפלט הסטנדרטי.

לדוגמה: אם קובצי הקלט הם:

אזי קובץ הפלט יהיה:

Dana 80 70 90

Rama 60 80 70

Sam 6

Yosi 1 2 3 4

הערות

א. ניתן להניח כי מזהה של תלמיד יהיה מחרוזת באורך > 50 תווים.

ב. ניתן להניח כי בסוף השורה האחרונה מצוי התו '\n'.

ג. אינם צריכים להיות מודאגים מהבדלים בין הפתרון שלכם לפתרון ביי"ס שנגרמים מרווחים (למרות שאין סיבה שיהיו כאלה).

תרגילים נוספים שלא להגשה

תרגיל בסיסי פשוט

כתבו תכנית הקוראת מהמשתמש שמות של שני קבצים: קובץ קלט וקובץ פלט. התכנית תפתח את הקצים, תקרא את קובץ הקלט עד תומו, ותדפיס על קובץ הפלט:

א. כמה תווים יש בקובץ הקלט

ב. כמה מחרוזות יש בו
ג. כמה שורות הוא כולל
לדוגמה עבור קובץ הקלט הבא :

Yosi 17 abc!
!?ab

יוצג הפלט :

18 4 2

הערה : בספירת התווים נספר גם ה : \n.

בכל השאלות בתרגיל, הניחו כי אורך שם קובץ הקלט יהיה קטן ממאה תווים