

תרגיל בית מס. 2

מטרת התרגיל: כתבו תכנית המדמה את התכנית הבנויה ¹grep. הפקודה מקבלת מחרוזת ושם של קובץ, ומדפיסה את כל השורות בקובץ המכילות את המחרוזת. לדוגמא, ניקח את הקובץ הבא:

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;
Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,
And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.
I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.

אם שמו של הקובץ לעיל הוא TheRoadNotTaken.txt, ונריץ את פקודת ה-grep הבאה, התוצאה תהיה כדלקמן:

```
rsivan@ubuntu:~$ grep the TheRoadNotTaken.txt
To where it bent in the undergrowth;
Then took the other, as just as fair,
And having perhaps the better claim,
Though as for that the passing there
Had worn them really about the same,
Oh, I kept the first for another day!
I took the one less traveled by,
And that has made all the difference.
rsivan@ubuntu:~$
```

(הצבעים בטקסט לעיל אינם חלק מהעבודה של grep: שמתי אותם כדי להקל על ההבנה של הדוגמא. באדום מציינים המופעים של מחרוזת החיפוש "the". הצבע הירוק מציין את ה-prompt של ה-shell, ואינו חלק מההקלדה של המשתמש או הפלט של grep.)

¹ שמה של הפקודה grep הוא מוזר. מקורו בתמלילן ישן בשם ed שהיה מקובל בימיה הראשונים של Unix. בימים ההם משאבי המחשבים לא אפשרו עריכה בגוף הטקסט, כפי שמקובל היום: כל הוראה פעלה רק על השורה הנוכחית. ההוראה re str (ה re הם ראשי תיבות של Regular Expression) ציינה שיש לחפש את המחרוזת str בשורה זו. אם הוסיפו g (בשביל Global) בהתחלה, הכוונה היתה לחפש בכל שורות הקובץ. ואם הוסיפו p (בשביל Print) בסוף, הכוונה היתה להדפיס את כל השורות שנמצאו. מכאן g-re-p...

שימו לב ש-"Then" (בתחילת שורה 6 של השיר), שזהה באיות למחרוזת שאותה חיפשנו, איננה בין התוצאות, לעומת "other" (באותה שורה) שכן: "The" ו-"the" הן מחרוזות שונות ב-Unix.

אני ממליץ לכם להריץ את הפקודה `grep` על כמה קבצים (כרגיל ב-Unix, אפשר גם כמה ביחד) וראו מה היא יודעת לעשות.

התכנית מקבלת לפחות פרמטר אחד: המחרוזת אותה יש לחפש. כל הפרמטרים הבאים הם שמות של קבצים בהם יש לחפש את המחרוזת הנתונה. קובץ מצויין על ידי `path`, יחסי או מוחלט. אם לא מצויינים קבצים כלל, החיפוש נעשה על **הקלט הסטנדרטי**.

יש לקרוא כל קובץ המצויין בין הפרמטרים לתכנית (או את הקלט הסטנדרטי, אם לא צויינו קבצים), שורה אחר שורה. כל שורה המכילה מופע אחד או יותר של המחרוזת שצויינה בפרמטר הראשון לתכנית, מודפסת ללא שנוי לפלט הסטנדרטי. אחרי שהטיפול בקובץ מסתיים, יש להפיק שורה נוספת המכילה את שם הקובץ ומספר השורות שנמצא שהכילו את המחרוזת, כך:

TheRoadNotTaken.txt: 8 lines found

אתם רשאים להשתמש ב-`printf` כדי לכתוב לפלט הסטנדרטי, אבל על כל פעולות הקריאה והכתיבה האחרות להשתמש רק בקריאות המערכת עליהן למדנו (`close`, `write`, `read`, `open`). כמו כן, אתם רשאים להשתמש בכל הפונקציות המוגדרות בקובץ `string.h` (כמו `strlen`, `strcmp` וגם הגרסאות מוגבלות האורך, כמו `strncmp`, שמקבלות פרמטר נוסף המגביל את אורך ההשוואה). נצלו את התיעוד המובנה בתוך Unix: הפקודה `man` (קיצור של Manual) מספקת הסברים על כל פקודה, קריאת מערכת, פונקצית ספרייה או סתם נושא שקשור בבצוע מטלות ב-Unix. פשוט קראו ל-"`man` <שם>" על מנת לקבל הסבר על המטרה ודרך השימוש ב-<שם>.

`man` כולל דפי הסבר גם לפקודות (כאלו שניתן להפעיל מה-shell, כמו `ls`), קריאות מערכת (כמו `open` או `write`) ופונקציות של C (כמו `printf` או `isspace`). קורה שלכמה מהן שם זהה, למשל `chmod` (שהיא גם פקודה וגם קריאת מערכת) או `dialog`, שהיא גם פקודה וגם פונקציה של C. במקרה זה אפשר לכוון את `man` לפרק המתאים ע"י הדגל `-S`:

```
$ man -S1 chmod -   chmod הפקודה
$ man -S2 chmod -   chmod קריאת המערכת

$ man -S1 dialog -   dialog הפקודה
$ man -S3 dialog -   dialog פונקציה של C
```

את התרגיל יש לבצע בזוגות בבית על מחשב אישי או על מחשב במעבדה במכללה. יש לקבץ את קוד התכנית ואת הפלט שלה באמצעות `RAR`, `ZIP` או תוכנה דומה. שם הקובץ שתיצרו צריך להיות בפורמט הבא:

Israel_Israeli_123456789_Ploni_Almoni_987654321_Ex2.zip

כלומר: השם באנגלית ומספר הזהות של כל אחד מהמגישים, והמחרוזת "Ex2", מופרדים באמצעות קו תחתון ביניהם. את הקובץ (היחיד) יש להעלות באמצעות moodle לאתר הקורס. נא ציינו, בנוסף, את שמותיכם (בעברית) ואת מספרי הזהות שלכם בגוף ההגשה בקובץ `ReadMe.txt`. את התרגיל יש להגיש עד יום

בהצלחה!