תרגיל מעבדה 1

1. כתבו סקריפט אשר מפצל את הספר "אליסה בארץ הפלאות" (נמצא בתיקייה lab1 bash) לפרקים- כלומר צרו תיקייה חדשה אשר תכיל קובץ נפרד עבור כל אחד מהפרקים בספר. את כל הניתוחים יש לבצע לפי מילים בתצורתם באותיות קטנות (lower), ללא סימני פיסוק וללא מילות קישור נפוצות. רשימת מילות הקישור נמצאת בהמשך דף ההוראות. awk + bash
2. עבור כל פרק בספר ועבור הספר כולו:
   a) מצאו את צמד המילים שמופיע הכי הרבה פעמים בצמידות (ללא חשיבות לסדר) בשורה. הדפיסו את צמד המילים לפי סדר אלפביתי (פונקציית עזר לשימושכם נמצאת בהמשך דף ההוראות). awk
   b) מצאו את המילה שמופיעה הכי הרבה פעמים בתור המילה הראשונה בשורה. awk
3. חשבו מהו מספר המילים הממוצע בכל שורה והדפיסו את מספר השורות הקצרות מהממוצע. awk
4. חשבו את המיקום הממוצע של אליס בשורה. (לדוגמא במשפט "אליס בארץ הפלאות" המיקום הוא 1 ובמשפט "ראיתי את אליס" המיקום הוא 3). שורה שלא מכילה את המילה אליס לא תילקח בחשבון בחישוב. שימו לב שמילה יכולה להופיע מספר פעמים בשורה ויש לקחת בחשבון את כל המופעים. awk

דגשים והנחיות

- בהורדת סימני פיסוק ו stop words אין להוסיף תו אחר במקומם.
- שימו לב שרשימת ה stop words היא באותיות קטנות וללא סימני פיסוק (כמו גרש למשל).
- אין להתחשב בשורות ריקות (שורות שמכילות רק רווחים או tab) בחישובים.
- יש לוודא ששורה מתחילה בתו כלשהו ולא ברווח או tab.
- בחלוקת הספר לפרקים, פרק ייחשב כל מה שמופיע לאחר המילה "chapter" .

הוראות הגשה

- ההגשה היא בזוגות. רק אחד מבני הזוג נדרש להגיש את התרגיל במודל.
- לפני ההגשה, חובה לבדוק את התרגיל במעבדת ההוראה ולא בסביבה אחרת.
- הנכם מתבקשים להגיש אך ורק את קובץ הסקריפט שלכם (סיומת sh).
- שם הקובץ חייב להיות hw1_lab_xxxxxxxx_yyyyyyyy.sh כאשר xxxxxxxx ו-yyyyyyyy הם מספרי תעודות הזהות של המגישים, כולל ספרת ביקורת.
- מצורף פורמט הפלט הרצוי. הקפידו להדפיס את תשובותיכם בפורמט המדויק משום שתתבצע בדיקה אוטומטית להשוואת הפלט שאתם מדפיסים מול הפלט הרצוי.
- בקובץ הפלט לדוגמה שקיבלתם, שורות 1-6 מכילות את התשובות הנכונות, על מנת שתוכלו לבדוק את עצמכם.
- תרגיל בית שלא יוגש לפי הוראות ההגשה – לא ייבדק.

רשימת מילות קישור נפוצות (stop words):

stop_words = (a about above across after afterwards again against all almost alone along already also although always am among amongst amoungst amount an and another any anyhow anyone anything anyway anywhere are around as at back be became because become becomes becoming been before beforehand behind being below beside besides between beyond bill both bottom but by call can cannot cant co computer con could couldnt cry de describe detail do done down due during each eg eight either eleven else elsewhere empty enough etc even ever every everyone everything everywhere except few fifteen fify fill find fire first five for former formerly forty found four from front full further get give go had has hasnt have he hence her here hereafter hereby herein hereupon hers herself him himself his how however hundred i ie if in inc indeed interest into is it its itse keep last latter latterly least less ltd made many may me meanwhile might mill mine more moreover most mostly move much must my myse name namely neither never nevertheless next nine no nobody none noone nor not nothing now nowhere of off often on once one only onto or other others otherwise our ours ourselves out over own part per perhaps please put rather re same see seem seemed seeming seems serious several she should show side since sincere six sixty so some somehow someone something sometime sometimes somewhere still such system take ten than that the their them themselves then thence there thereafter thereby therefore therein thereupon these they thick thin third this those though three through throughout thru thus to together too top toward towards twelve twenty two un under until up upon us very via was we well were what whatever when whence whenever where whereafter whereas whereby wherein whereupon wherever whether which while whither who whoever whole whom whose why will with within without would yet you your yours yourself yourselves)

פונקציה להדפסת צמד מילים לפי הסדר האלפביתי שלהן:

Add to the awk script (e.g. before the BEGIN{})

```
function sorted(a,b){
        if (a < b) return a " "  b;
        else return b " " a;
}
```

סדר:

1. Lowercase. Awk lowercase
2. Remove punctuation(sed\awk)
3. Stop words. Sed
4. Remove empty lines/ remove trailing spaces. Sed