

תרגיל 3, מבוא לתכנות מערכות, אביב 2022

הגשה בזוגות או ביחידים דרך המודל
התרגיל הוא להגשה עד ליום רביעי, 12/4/2022 בשעה 23:59

בתרגיל זה נחלק את הטקסט של הספר מובי דיק לפרקים!

זכויות היוצרים של הספר שנכתב ב-1951 כבר פגו, ואפשר להוריד את המילים באופן חופשי מ-[github](https://github.com), ואפילו לראות צילום של המקור. בתרגיל זה תפרקו את טקסט הספר לפרקים, ותבקשו להדפיס סטטיסטיקות פשוטות על כל פרק ועל כל הספר. במודל תוכלו למצוא קובץ ארכיב בשם `mobydick.zip` המכיל את טקסט הספר ופלט לדוגמא.

מטרת התרגיל היא לכתוב תוכנית ש:

1. מפרקת את הקובץ הארוך (15k שורות) לקובץ אחד לכל פרק.
2. מדפיסה עבור כל פרק את מספר השורות, המילים והתווים.
3. מדפיסה את סך כל השורות, המילים והתווים בכל הספר.

בניית התוכנית

ממשו את הפונקציות הבאות בקובץ אחד. השתמשו בפונקציות שממשתם כבר לצורך ביצוע משימות והמנעו ככל האפשר משכפול קוד. הפונקציות מסודרות בסדר קושי עולה אז מומלץ לממש אותן לפי הסדר ולבדוק כל פונקציה בנפרד.

אין להשתמש בפונקציות מתוך `string.h`! מומלץ בחם לבדוק כל פונקציה שממשתם לפני שאתם עוברים לפונקציה הבאה.

1. `my_strlen`: מימוש שלכם לפונקציה `strlen`, המקבלת מחרוזת ומחזירה את אורכה.
2. `my_strcat`: מימוש שלכם לפונקציה `strcat`, המקבלת שני מחרוזות `dest`, `src` ומשרשרת המחרוזת את `src` בסוף של `dest`. (הפונקציה אינה מקצה מקום, ומניח שיש בסוף `dest` מקום פנוי מוקצה).
3. `starts_with`: המקבלת מחרוזת `s` ומחרוזת `prefix` לחיפוש, ומחזירה `true` אם `s` מתחילה ב-`prefix`.
4. `num_words`: המקבלת מחרוזת ומחזירה את מספר המילים במחרוזת. מילה מוגדרת כרצף הכי ארוך של תווים שאינם `white-space`, כלומר אינם 'נ' או 'ז' או רווח.
5. `get_chapter_file_name`: המקבלת מחרוזת `s`, מחרוזת `prefix`, ומחרוזת `suffix`, ומחזירה שם של קובץ. לשם כך היא מנרמלת את `s`, ומשרשרת אותו עם `prefix` לפני ועם `suffix` אחרי. פעולת הנרמול של `s` כוללת את הורדת כל תווי ה-`white-space` מסופו של `s`, והחלפת תווי ה-`white-space` הנותרים בתו מינוס. לדוגמא, עבור `s="CHAPTER I \n"`, `prefix="Moby-"`, ו-`suffix=".txt"`, הפונקציה תחזיר את המחרוזת `"Moby-CHAPTER-I.txt"`.

אסור לפונקציה לשנות את המחרוזות שהיא מקבלת (`s`, `prefix`, `suffix`), ועליה להקצות זיכרון בשביל התשובה.

6. `main`: מקבלת, דרך שורת הפקודה, 2 או 3 ארגומנטים (מלבד הפקודה עצמה):
 1. שם קובץ הקלט של הספר
 2. ה-`output_prefix`
 3. ה-`output_suffix` (זה ארגומנט אופציונלי, עם ברירת מחדל `".txt"`).
- הפונקציה פותחת את קובץ הקלט לקריאה, וקוראת אותו שורה שורה, כאשר השורות מועתקות לקובץ הפלט המתאים.
- בהתחלה לקובץ (`get_chapter_file_name("PREFIX", output_prefix, output_suffix)`) וכל פעם שנתקלת בשורה מתחילה ב-"CHAPTER" מחליפה את קובץ הפלט ל-`get_chapter_file_name(current_line, output_prefix, output_suffix)` כאשר `current_line` הינו תוכן השורה הנוכחית.

- לאחר כל החלפת קובץ פלט, וגם בסוף התוכנית יש להדפיס הסטטיסטיקות על הקובץ הכוללת את שם הקובץ, ואת מספר השורות, מילים ותווים בקובץ בפורמט הבא:
 Moby-PREFACE.txt : 969 lines, 4362 words, 24706 characters
 Moby-CHAPTER-I.txt : 286 lines, 2302 words, 12644 characters
 Moby-CHAPTER-II.txt : 191 lines, 1507 words, 8256 characters
- בסוף התוכנית, יש להדפיס בנוסף את הסטטיסטיקה על כל הקובץ, בפורמט הבא:
 TOTAL : 15604 lines, 115314 words, 643210 characters
- הערה: את שם הקובץ יש להדפיס בפורמט 30s-% כדי שההדפסות יצאו מעומדות יפה.

דגשים

- הקפידו לשחרר כל זיכרון שהקצתם דינמית (או שהוקצה עבורכם!)
- אין להניח דבר על אורך השורה המקסימלי בקובץ. יש להשתמש ב-getline לקריאת שורות.
- הקפידו לבדוק שהקצאות זכרון הצליחו. במידה ולא, הדפיסו הודעת שגיאה מפורטת עם מספר הבתים שניסיתם להקצות וצאו מיידית עם ערך שאינו 0.
- הקפידו לבדוק שפתיחת קבצים הצליחה. אם לא, הדפיסו הודעת שגיאה מתאימה וצאו מהתוכנית.
- הקפידו לסגור קבצים שפתחתם
- אין להשתמש ב-string.h
- ממשו פונקציה, בידקו אותה, ורק אז עברו לפונקציה הבאה! למי שינסה לממש ולבדוק את הכל בבת אחת מובטחות שעות רבות של סבל ואומללות.

הגשה

ממשו את כל הפונקציות באותו הקובץ main.c. הגישו את הקובץ main.c **לא דחוי** דרך המודל. אם אתם מגישים בזוג הגישו את התרגיל רק פעם אחת, דרך החשבון של אחד מכם (לא משנה מי), אבל הקפידו לרשום שמות בכותרת בראש הקובץ.

דוגמת הרצה

ניתן למצוא דוגמת לפלט הרצה במודל.

בהצלחה!

