

מבוא לתכנות מערכות
תרגיל בית מספר 8
(בנוס)

נושא: עבודה עם קבצי טקסט, קבצים בינאריים וביטים

סמסטר אביב 2020-21

תאריך הגשה: 22.07.21
הגשה בזוגות

בהצלחה!

מטרת התרגיל

עבודה עם קבצי טקסט, קבצים בינאריים וביטים.

תאור התרגיל

כתוב תוכנית המבצעת החלפת סדר המילים במשפטים המופיעים בקובץ טקסט נתון, מדפיסה אותן בסדר שונה ודוחסת את הקובץ.

קלט התוכנית

- קובץ טקסט המכיל משפטים - אוסף המילים המופרדות ברווחים.
- כל משפט מסתיים בסימן נקודה .
- משפט יכול להכיל עד 1000 תווים.
- כל שורה בקובץ יכולה להכיל יותר ממשפט אחד.
- משפט יכול להתפרס על יותר משורה אחת.

א). בשלב ראשון יש ליצור קובץ חדש ששמו כשמו של הקובץ המקורי בתוספת הסימון "rpl". המכיל את החלפת סדר המילים במשפטים המופיעים בקובץ טקסט נתון. כל משפט בקובץ החדש יתחיל משורה חדשה.
למשל, המשפט I love Matam Course. מקובץ טקסט נתון יוחלף במשפט
Course Matam love I.

ב). בשלב שני נרצה להצפין את הקובץ שהתקבל בסעיף א'. לשם כך יש ליצור קובץ חדש ששמו כשמו של הקובץ המקורי בתוספת הסימון "bend". המכיל את התווים של הקובץ מסעיף א' באופן הבא: בהתחלה יופיע התו הראשון של הקובץ, אחריו התו האחרון, אחר זה התו השני, ואחריו התו שלפני התו האחרון וכך סידורגין הדבר יימשך עד אשר נגיע לאמצע הקובץ.

הערות:

1). אין להניח כי מספר התווים בקובץ הוא זוגי או אי-זוגי.

- (2). יש להתייחס אל התו "\n" כאל כל תו אחר
(3). לפתרון יעיל, מומלץ להשתמש בפונקציית ftell

ג). נרצה לצמצם את המקום הנדרש לייצוג מחרוזת בת n תווים, שבה מספר התווים השונים הוא לכל היותר 16. כלומר המקום שנדרש ל n בתים כדי לשמור אותה ע"פי קוד ascii, יצטמצם ל- $n/2$.
לכל תו נצמיד מספר בין 0 ל- 15 (4 ביטים) כך שבכל byte יהיו שני תווים.
למשל, המחרוזת abba תצטמצם להיות: 00000000100010000
אם נסתכל בסדרת התווים השונים מהמחרוזת המקורית, מסודרים לפי סדר הופעתם הראשונה, הערך שיוצמד לכל תו יהיה בסדרה זו (החל מ- 0).
למשל, אם המחרוזת המקורית היא "acba",
ל- a, יוצמד הערך 0.
ל- c, יוצמד הערך 1.
ל- b, יוצמד הערך 2.
ולכן המחרוזת הדחוסה תראה כך:
00000000100100000

שים לב:

- 1). המחרוזת לא מסתיימת ב "\0".
- 2). אם בשורה מסוימת של קובץ text יש יותר מ-16 תווים שונים, יש להודיע כי לא ניתן לבצע דחיסה לפי האלגוריתם הנתון ולסיים את התוכנית.
כתוב תוכנית המקבלת כפרמטר שם של קובץ טקסט שבנית בסעיף א'. ידוע שהקובץ הנ"ל מורכב משורות באורך זוגי של תווים, שבכל שורה יש לכל היותר 16 תווים שונים.
על התוכנית להמיר קובץ זה לקובץ בינארי:
עבור כל שורה בקובץ הטקסט, ישמרו בקובץ הבינארי המיוצר הנתונים הבאים:
1. מספר מטיפוס unsigned char המייצג את מספר התווים השונים בשורה.
2. רצף (באורך זה) של ה-charים המכיל את סדרת התווים השונים בשורה, מסודרים לפי סדר הופעתם הראשונה באותה שורה.
3. רצף של בתים המכילים את הצמצום של השורה (כפי שתואר לעיל).
שמו של הקובץ המיוצר הוא כשמו של קובץ הקלט, בתוספת " rds." בסוף.

שים לב:

יש להקפיד על שחרורים של שטחים שאינם בשימוש, כמו גם על הקצאת שטחים חדשים. יש להשתמש בפונקציות free, realloc, calloc, malloc. ניתן להשתמש בכל הפונקציות מ- string.h.

יש לקמפל ולהריץ את התוכנית ב LINUX. שים לב: תוכנית שלא תתקמפל במערכת הפעלה LINUX תקבל ציון 0!

ערעורים:

יש להפנות לבודק **אביתר בהלקר** בלבד ע"י שליחת מייל evyatarbhal@gmail.com **תוך שבוע** ממועד פרסום הציונים. בכותרת המייל יש לציין: "ערעור במת"ם עבודת בית מס' X ע"י +ת.ז. של הסטודנטים". יש לקחת בחשבון שבעת הערעור העבודה נבדקת מחדש וכתוצאה מהערעור ציון העבודה עלול להשתנות (יכול לעלות או לרדת). התשובה לערעור תהיה סופית ולא ניתנת לערעור נוסף.

בהצלחה!