

IT עבודה מספר 5

אדם טרסוב 321396822 גיא הולצמן 04049362

1. טבלת סיכום:

זמן ריצה של כל אחד פונקציות (בשניות)									
ENT	DIC	COM	DEC	אנטרופיה אמפירית (מס' ביטים לאות)	מס' ביטים ממוצע לאות בקובץ הדחוס	גודל של קובץ דחוס (ביטים)	מס' ביטים ממוצע לאות בקובץ מקורי	גודל קובץ מקורי (ביטים)	שיטת קידוד
2.22	2.22	3.55	1.17	סימבול = אות אחת	4.18	2745848	8	5250816	Huffman-1
1.3	1.38	2.11	0.88	סימבול = 2 אותיות	3.90	2556656	8	5250816	Huffman-2
2.36	4.41	1.63	0.85	סימבול = 3 אותיות	3.64	2387680	8	5250816	Huffman-3
2.22	2.24	3.71	1.46	סימבול = אות אחת	4.58	3003608	8	5250816	Shannon
150 ms		150 ms		-----	2.88	1892112	8	5250816	7-Zip

2. הסבר התוצאות:

- לגבי הדחוס והפורס: ככל שהסימבול מכיל יותר תוים, כך פעולת הדחיסה והפרוסה לוקחות פחות זמן
- בחישוב האנטרופיה קיבלנו מינימום ב 2 תוים. אפשר להסביר את זה בזמני גישה למערך ובגודלו של המערך. כניראה שלקח זמן רב יותר לגשת למערך המילון ב 3 תוים, אבל לקח זמן רב יותר לגשת לקובץ בתו אחד, בשני תוים, זמן הגישה לשניהם הייתה מהירה יותר
- ביצירת המילון, יש צורך בחישוב של אנטרופיה, על כן זמן יצירת המילון קטן מזמן חישוב האנטרופיה. מלבד זאת, יצירת המילון ב 3 תוים לקחה זמן רב הרבה יותר מיצירת האנטרופיה, זאת משום שיצירת העץ הבינארי שם לקחה חלק משמעותי יותר.
- ככל שהשתמשנו בבולק גדול יותר כך קיבלנו כיווץ טוב יותר, זה לא מפתיע מכיוון שטקסט אינו חסר זיכרון. כמו כן, במקרה האידיאלי היינו יכולים לשלוח את כל המידע בביט בודד, בהינתן שהמילון שלנו היה מכיר רק ערך אחד שכולל את כל הטקסט. כלומר, בחישוב הכיווץ במקרה זה, לא התחשבנו בכלל בגודלו של המילון ועל כן חישוב זה לוקה בנסיבות מסוימות.
- תוכנת 7ZIP לקחה זמן קטן משמעותית מכל מה שתיכננו ממספר סיבות: מחישוב הכיווץ ניתן לראות שהיא משתמשת בבולקים גדולים יותר מ 3 תוים. סיבה שניה היא שזו תוכנית שרצה בשפה תחתית ומשתמשת במשאבי המחשב בצורה יעילה יותר. סיה שלישית היא שכניראה יש שיטות הרבה יותר יעילות לכיווץ (אבל עדיין לא למדנו אותם)