

AOPH

מגישים: אלירן עמר, חן אוחנה

מנחה: ד"ר עמית בן-בסט

רקע:

תוכנת דחיסה עבור קבצי טקסט, המממשת אלגוריתם ייחודי שיצרנו, אשר יוצר הכנה לטקסט ומאפשר ביצועים טובים יותר של אלגוריתם הופמן.

המערכת מאפשרת דחיסת טקסט באמצעות אלגוריתמים שונים ואוספת מידע על הקבצים עצמם.

תיאור המערכת:

1. בחירת קובץ מסוג txt לדחיסה ובחירת סוג הדחיסה:
 - א. דחיסת הופמן.
 - ב. דחיסת הופמן עם הכנת הטקסט, האלגוריתם שלנו – AOPH.
 - ג. דחיסה באמצעות אלגוריתם למפל-זיו.
2. אלגוריתם דחיסה דינאמי משודרג –

כל תו ייוצג ע"י מספר ראשוני, כך שסדרות תווים עולות ויורדות ייוצגו ע"י מספרים שהם כפל של מספרים ראשוניים אלה. כל מספר שכזה הוא ייצוג של סדרה ייחודית של רצף תווים. לאחר שלב זה מופעל אלגוריתם הופמן על המספרים המייצגים רצפים.
3. בעת ריצת האלגוריתם נאסף מידע על תדירות רצפי התווים בקובץ, מידע אשר נשמר בבסיס הנתונים.
4. המערכת לומדת מתוך המידע המוזן אל בסיס הנתונים לאורך הזמן על סוגי הקבצים ומבצעת סיווג לסוגים שונים (clustering).
5. אם המערכת מצליחה לסווג את הקובץ המורץ לאחד מסוגי הקבצים שנסרקו בעבר, תיחסך הכנסה של מילון קודי ההופמן אל הקובץ הדחוס, מה שיאפשר כיווץ רב עוד יותר.

אמצעים טכנולוגיים:

המערכת נבנתה בסביבת הפיתוח Eclipse בשפת JAVA .

הנתונים נשמרים במסד נתונים מסוג MySQL.

מסקנות ותוצאות:

- כאשר אנו נותנים קוד הופמן לסדרות תווים במקום לתווים בודדים, חל שיפור בדחיסה.
- AOPH משפר את הדחיסה בכ- 20% ביחס לאלגוריתם הופמן.
- האלגוריתם עובד טוב יותר על קבצי טקסט גדולים.