

הקדמה:**LZ77**

האלגוריתם משיג דחיסת נתונים על ידי החלפת מחרוזות מסוימות במופעים חוזרים במקור הנתונים, במצביע למיקום קודם של אותה מחרוזת בתוך הנתונים המקוריים, ומספק גם את אורך המחרוזת המשותפת. הרעיון בבסיס הקידוד הוא שכל חלק בקובץ הקידוד הוא המחרוזת הארוכה ביותר שנמצאה קודם לאותו המקום בקובץ המקורי, בנוסף לתו אחד. במילים אחרות, האלגוריתם מנסה לאתר מחרוזות מסוימות שנראות כמו מחרוזות אחרות שכבר מופיעות בטקסט, ולהשתמש באותן מחרוזות כדי להציע דרך קצרה יותר להציג את הנתונים. זה נעשה על ידי הצבת המיקום (offset) של המחרוזת הקיימת ואורך המחרוזת במיקום הנוכחי. בסך הכל, האלגוריתם מנסה ליצור קוד דחיסה שיכיל מידע פחות מהקובץ המקורי, על ידי הזיהוי והשימוש החכם במחרוזות משותפות שחוזרות במקור הנתונים.

LZSS

LZSS הוא אלגוריתם דחיסת נתונים ללא אובדן של מידע, האלגוריתם מבוסס על האלגוריתם המוכר LZ77. האלגוריתם מחליף מחרוזות של סמלים בהתייחסות למיקום המילון של אותה מחרוזת. ישנם הבדלים בין ה LZ77 ל LZSS, ההבדל העיקרי בין LZ77 ל-LZSS הוא שב-LZ77 הפניה למילון יכולה להיות למעשה ארוכה מהמחרוזת שהיא מחליפה. ב-LZSS, הפניות כאלה מושמטות אם האורך קטן מנקודה מסוימת.

Huffman

אלגוריתם לקידוד סימנים, כגון תווי טקסט, ללא אובדן נתונים. הקוד שייך למשפחה שימושית של קודים המכונה קודי תחליות ובמשפחה זו הוא הקוד המספק דחיסת נתונים מרבית, כלומר מאחסן את הסימנים במספר מזערי של סיביות, על פי הקריטריון מסוים. השיטה מתבססת על הקצאת אורך משתנה לסימנים על פי שכיחותם, כך שסימן נפוץ יוצג באמצעות מספר קטן של סיביות.

תיאור הפרויקט:

בפרויקט זה מימשנו את אלגוריתם LZSS, LZ77 ו-Huffman. הדחיסה באלגוריתמים אלו היא "עד הסוף", כולל דחיסה שניונית באלגוריתמים. האלגוריתם מקבל קובץ ויוצר קובץ דחוס, או לחילופין מקבל קופץ שנדחס ויפענח אותו לקובץ המקורי. בחירת הקובץ, הפרמטרים והאלגוריתם נעשים דרך ממשק GUI שמימשנו.

הסבר על המחלקות:

המחלקה GUI - זו המחלקה הראשית המייצגת את הממשק הגרפי של הפרויקט. היא מורשת מ-JFrame, שהוא המרכיב הבסיסי ליצירת חלון גרפי בממשק Swing. המחלקה מגדירה את המראה של החלון ופרמטרים נוספים כמו גודל החלון וכו'. המחלקה מכילה מספר פאנלים שנוצרים בתוך החלון הגרפי. כל פאנל מכיל אלמנטים גרפיים, כמו תיבות טקסט (נתיב הקובץ), כפתורי Compress ו Decompress, חלונות לבחירת Bytes וכפתור לבחירת האלגוריתם.

המחלקה LZ77Compress – המחלקה LZ77Compress מייצגת את אלגוריתם הדחיסה LZ77. היא מקודדת טקסט טקסטואלי על ידי החלפת מחרוזות חוזרות במצביע למיקום קודם ואורך החלפת המחרוזת. המחלקה מנהלת חלון חיפוש buffer, מוצאת התאמות בחלון ודוחסת את הנתונים.

המחלקה LZ77Decompressor - המחלקה LZ77Decompressor מבצעת פענוח של נתונים שעברו דחיסה באמצעות אלגוריתם הדחיסה LZ77. היא מפענחת את הנתונים הדחוסים ושומרת את הנתונים המפוענחים. המחלקה מבצעת קריאה מהקובץ הדחוס ומשמשת להבנת הנתונים הדחוסים באמצעות המידע שנוצר בתהליך הדחיסה.

המחלקה LZSSCompressor – זו מחלקה המבצעת דחיסת נתונים באלגוריתם LZSS. המחלקה מקבלת כקלט את נתיב הקובץ שיש לדחוס, נתיב לקובץ הדחוס שממנו ייצר הקומפרסור, גודל החלון, אורך ההתאמה המינימלית ואורך ההתאמה המרבי. בשלב הדחיסה, המחלקה קוראת את הנתונים מהקובץ המקורי, בונה מחרוזות וביטים נתונים לפי האלגוריתם. היא בונה את נתוני המקור באמצעות חלונות חיפוש ומציאת התאמות ביניהם.

המחלקה LZSSDecompressor - זו מחלקה המבצעת שחזור של הקובץ הדחוס לגודלו המקורי באלגוריתם LZSS. היא טוענת את נתוני הביטים הדחוסים מהקובץ הדחוס ובונה מחרוזות ונתוני מקור בהתאם לפי הפרטים שנדחסו על ידי הקומפרסור. הפרטים כוללים את גודל החלון, אורך ההתאמה המקסימלית והאורך של חלון הדחיסה. המחלקה מבצעת פרוש של הנתונים הדחוסים לפי הנתונים המקוריים באמצעות חלון חיפוש, ומשתמשת במחלקה PatternMatch לניתוח ושמירת ההתאמות. התוצאה היא שחזור המידע לנתונים המקוריים.

המחלקה PatternMatch - המחלקה "PatternMatch" מכילה משתנים ופונקציות שמשמשות לזיהוי וניתוח של תופעות דומות במחרוזות. המחלקה נוצרה למענה על הצורך לאחזור תבניות מחרוזות כחלק מתהליך דחיסת מידע, וניתוח זיהוי התבניות במחרוזות.

המחלקה Node - מייצגת רכיב בעץ למטרת קידוד ופענוח מידע באמצעות קוד הופמן. הצמתים בעץ מכילים את התדירות של התווים או הסימנים שהם מייצגים, והמבנה זה בעץ מאפשר יצירת קודים שמתחילים בתווים/סימנים פשוטים ונמשכים לשכבות עמוקות יותר מסמלות קודים ארוכים יותר.

המחלקה MyComparator - המחלקה MyComparator מגדירה איך לבצע השוואה בין שני אובייקטים מסוג Node כאשר יש צורך למיין אוסף של אובייקטים מסוג זה. ההשוואה נעשית על פי ערך התדירות (freq) של הנודים.

File Name	Original size	LZSS	LZ77	LZSS and huffman
Input.txt	3156KB	9KB	10225KB	25KB
Words_alpha.txt	4136KB	2089KB	13441KB	2163KB
Story.txt	15826KB	23KB	51435KB	9295KB
photo1.TIFF	257KB	288KB	833KB	288KB
img_original .bmp	118KB	2KB	3KB	1.6KB
Compression.mp4	15868KB	17618KB	17618KB	28028KB
example_JPG_1MB.jpg	1019KB	1136KB	1750KB	1136KB
OnTheOrigin.txt	1027KB	708KB	797KB	708KB

צילומי מסך:

