

Dokumentácia k projektu 3 do predmetu Kódovanie a kompresia dát

Michal Chomo, xchomo01

1 Úvod

V tomto projekte je implementovaná aplikácia, ktorá umožňuje prevod obrázkov vo formáte Graphic Interchange Format (GIF) do formátu bitmap (BMP) s využitím kompresnej metódy Lempel-Ziv-Welch (LZW). LZW je bezstratový kompresný algoritmus založený na LZ78 vyvinutý v roku 1984 Terryom Welchom. Tento algoritmus patrí medzi adaptívne slovníkové metódy, ako slovník sa používa dátová štruktúra trie. Aplikácia používa knižnicu *gif2bmp*, ktorá je taktiež súčasťou projektu a poskytuje potrebné API.

2 Dekódovanie LZW

Pre GIF obrázky sa používa špeciálna varianta LZW, kde sa do slovníku pri inicializácii vložia aj dva špeciálne kódy Clear Code (CC) a End of Information Code (EOI). CC indikuje, že je treba zmazať slovník a začať ho plniť odznovu, EOI indikuje koniec zakódovaných dát.

2.1 Kroky algoritmu

1. Inicializuj slovník - vlož všetky symboly abecedy, CC a EOI
2. Načítaj zo vstupu prvý ukazovateľ *di* do slovníka (vždy sa nájde reťazec)
3. Zapiš reťazec *I* z *di* na výstup
4. Načítaj zo vstupu ďalší ukazovateľ *di* do slovníka
5. Je *di* index CC? Ak áno, reinicializuj slovník a choď na 4, inak pokračuj.
6. Je *di* index EOI? Ak áno, ukonči dekodovanie, inak pokračuj.
7. Je reťazec *J* z *di* v slovníku?
 - Ak áno, zapiš *J* na výstup a pridaj do slovníka $I + J[1]$
 - Ak nie, zapiš $I + I[1]$ na výstup a pridaj do slovníka $I + I[1]$
8. Ulož *J* do *I* a choď na 4.

3 Implementácia

Projekt je implementovaný v jazyku C v štandarde C99. Spracovanie parametrov, otváranie a zatváranie súborov a volanie potrebných funkcií z knižnice je v súbore *main.c*. Knižnica *gif2bmp* pre dekodovanie GIF obrázku, prácu so slovníkom a tvorbu BMP obrázku sa skladá z hlavičkového súboru *gif2bmp.h*, kde sú definované potrebné štruktúry a makrá a deklarované funkcie a zdrojového súboru *gif2bmp.c*, ktorý obsahuje implementáciu funkcií. Vstupom programu je súbor vo formáte GIF zadaný parametrom *-i názov_súboru*. V prípade, že tento parameter nie je zadaný, tak je ako vstup považovaný *stdin*. Výstupom programu je súbor vo formáte BMP zadaný parametrom *-o názov_súboru*. Ak nie je zadaný názov výstupného súboru, výstup sa vypíše na *stdout*. Program taktiež prijíma parameter *-l názov_súboru*, čo je voliteľný súbor výstupnej správy, ktorá obsahuje login autora a veľkosti vstupného avýstupného súboru.