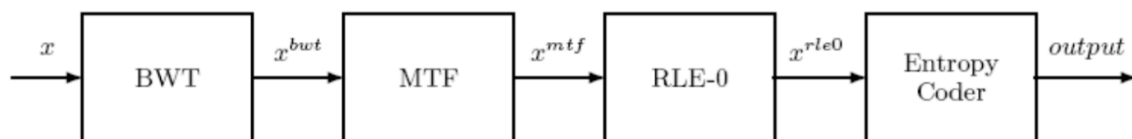


Dokumentace k projektu z KKO

Lukáš Dibdák, xdibda00

Celý projekt bych rozdělil do 5 částí, které budu specifikovat níže.



1. Načtení souborů a kontrola argumentů

V této části byly zkontrolovány vstupní argumenty. Vstupní argumenty byli zkontrolovány pomocí funkce **getopt**. Vstupní argumenty mohou být tyto:

-h	vyvolání nápovědy
-c	komprimace vstupního souboru
-x	dekomprimace vstupního souboru
-i	specifikace umístění vstupního souboru
-o	specifikace umístění výstupního souboru
-l	specifikace umístění logovacího souboru

Následuje otevření souborů a předání těchto souborů ke zpracování. V programu je tato část zpracovávána v souboru **main.cpp**.

2. BWT

Burrows-Wheelerova transformace spočívá ve vytvoření n permutací vstupního řetězce o délce n postupným přerovnáváním prvního znaku na konec řetězce. Tyto permutace se následně seřadí a z posledního znaku každé permutace se vytvoří konečný řetězec. Vstupní řetězec v tomto projektu musí být označen znakem ukončující řetězec 'l'.

Kódování i dekódování je umístěno v souborech **BWT.cpp** a **BWT.h**

3. MTF

V této části je inicializována tabulka symbolů. Následně se prochází vstup (v tomto případě část MTF navazuje na část BWT, vstupem je tedy výstup z B-W transformace)

znak po znaku, získává se index znaku v tabulce a aktualizuje se tabulka, kde znak na daném indexu je přesunut na začátek.

Kódování i dekódování je umístěno v souborech **MTF.cpp** a **MTF.h**

4. RLE

Run-length encoding je bezeztrátová komprese, která kóduje větší výskyt stejných vstupních znaků za sebou do zkráceného formátu. V projektu je tato komprese použita u výskytu více než 3 stejných znaků za sebou. Pokud bylo nalezeno na vstupu (tím je výstup z MTF) posloupnost více než 3 znaků za sebou, pak byli zakódováni v tomto formátu: @ <počet znaků> <znak>.

Kódování i dekódování je umístěno v souborech **RLE.cpp** a **RLE.h**

5. Entropy Coder

Bohužel z časových důvodů nebyla tato část implementována.

6. Testování

Bohužel z časových důvodů nemohlo být použito pokročilé testování. Program byl testován od části 1 do části 4 (Entropy Code nebyl implementován) na jednoduchých a krátkých řetězcích.

Toto testování je možné ověřit pomocí přeložení příkazem **make** a následně pomocí příkazu **make compress**, které provede kompresi řetězce s ukončovacím znakem: **testovaciretezec!** přes části BWT -> MTF -> RLE a následně dekódováním pomocí **make decompress**, kdy se ze výstupního řetězce RLE přes MTF a BWT kód dekomprimuje zpět na původní řetězec.

7. Závěr

Projekt hodnotím jako velice zajímavý. Problematika kompresních algoritmů mě velice zaujala a práce mě bavila. Bohužel kvůli pracovnímu vytížení a následné nemoci jsem nemohl implementaci a následné pokročilé testování dokončit v daném čase.