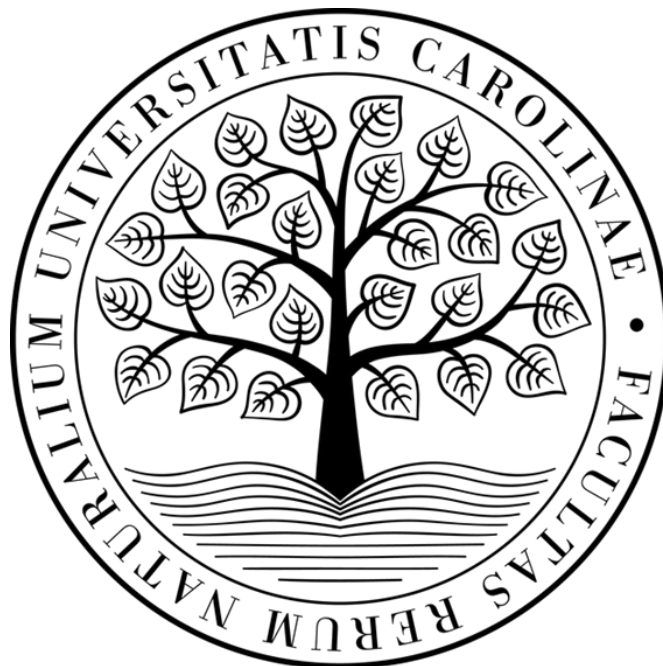


Univerzita Karlova

Přírodovědecká fakulta



ÚVOD DO PROGRAMOVÁNÍ

Zkouška – příklad 84

RLE komprese

Ludmila Vítková

3 BGEKA

Praha, 2023

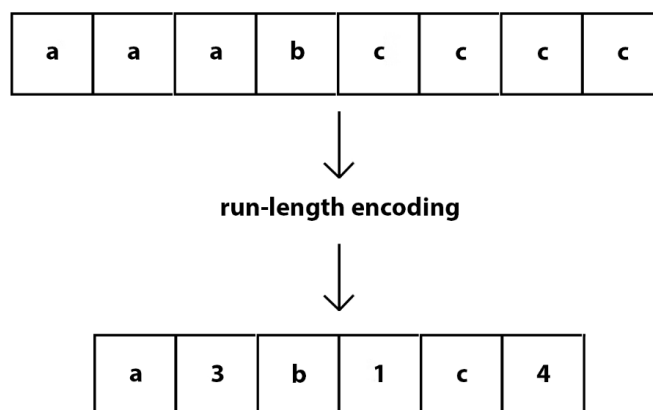
Zadání

Pro zadaná data zobrazující hodnoty odstínu šedi (hodnoty na intervalu 0 až 255) proveďte RLE kompresi.

Popis a rozbor problému

RLE komprese (což je zkratka pro Run Length Encoding) je jedním z nejjednodušších kompresních algoritmů. Tato komprese nahrazuje opakující se řetězce hodnotou udávající počet opakování a samotnou hodnotou znaku.

RLE komprese může pracovat s různými znaky například s písmeny (*Obrázek 1*), nebo s číselnými hodnotami jako v případě tohoto programu.



Obrázek 1: příklad RLE komprese

Použitý algoritmus

RLE komprese komprimuje data tak, že prochází data a v případě že za sebou najde dvě stejné hodnoty připočítá si jeho opakování. To se opakuje, dokud nenarazí na další hodnotu, která se už liší. V tomto případě zůstanou za sebou uložené dvě hodnoty, první uvádějící samotnou číselnou hodnotu z intervalu 0–255 z původních dat a druhou, která uvádí její počet opakování za sebou. Takto projede všechny prvky v souboru a výsledkem je komprimovaný soubor.

Struktura programu

Program se skládá z jedné třídy *Komprese*, součástí jsou metody:

- `__init__(self)` definuje datové složky
 - o `self.__text` – vstupní data
 - o `self.compressed` – komprimovaná data
- `encode(self)` – provádí kompresi dat
- `compressed(self)`

Průběh programu

Program otvírá soubor obsahující data hodnoty od 0 do 255 obsahující odstíny šedi. Během otvírání souboru zároveň program kontroluje, jestli daný soubor existuje a má

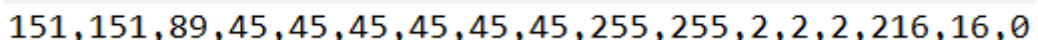
k němu uživatel přístup. Kontroluje také správný zápis dat ve vstupním souboru – například zda, se v něm neobjevují písmena, či jiné znaky než číslice, se kterými program nedokáže správně pracovat. Kontroluje také, zda soubor není prázdný. V případě nesplnění některé s uvedených podmínek program na problém upozorní a okamžitě se ukončí.

Následně program postupně hledá jednotlivá prvky zadaných dat (číselné hodnoty). Program ověřuje, zda jsou data opravdu zadána v intervalu 0 až 255. Pokud data nejsou zadána tímto způsobem, program na to upozorní a ukončí se. Pokud jsou nesou data správné hodnoty, prochází dále program hodnoty a porovnává je s následující hodnotou, v případě, že jsou hodnoty stejné uloží se do proměnné *count* nová hodnota, která vzniká tak že se ke stávající hodnotě připočte 1. Pokud už následující hodnota není stejná, uloží se konečná hodnota proměnné *count* společně s hodnotou prvku, a následně program vyskočí z cyklu a přechází na další prvek.

Komprimovaná data se nakonec zapisují do nového výstupního souboru.

Vstupní data

Program přijímá vstupní data z textového souboru nesoucí název *vstup_RLE.txt*. Soubor bude obsahovat hodnoty představující odstín šedi (celá čísla v intervalu 0 až 225). Hodnoty musí být v souboru zapsány za sebou v jednom řádku a odděleny čárkou.



151,151,89,45,45,45,45,45,45,255,255,2,2,2,216,16,0

Obrázek 2: ukázka vstupních dat

Výstupní data

Do výstupního souboru s názvem *vystup_RLE.txt* ukládají komprimovaná data. V komprimovaných datech první hodnota udává hodnotu intervalu, druhé číslo udává počet opakování.

Možná zlepšení

Tento program je vytvořen čistě pro komprimování dat podle RLE komprese. Pro lepší využití a plnou funkčnost programu by bylo vhodné vytvořit algoritmus i pro dekomprimaci zkomprimovaných dat, také za pomoci metody RLE komprese.

Zdroje:

api-video (2023): What is run-length encoding <https://api.video/what-is/run-length-encoding> (cit. 12. 2. 2023)