

Príhody doktora Vojšiča, epizóda druhá: Jarné upratovanie

Úvod. Doktor Vojšič sa rozhodol, že svoje každoročné jarné upratovanie tentoraz začne tým, že si konečne urobí poriadok vo svojom počítači. Aby ušetril miesto na disku, rozhodol sa, že si svoje cenné údaje s vedeckými experimentmi skomprimuje pomocou algoritmu LZW (Lempel-Ziv-Welch), v knihe kap. 5.5:

<http://index-of.es/Varios-2/Algorithms%204th%20Edition.pdf>

Po čase zistil, že ušetrená kapacita je len veľmi malá a zdá sa mu, že v programe urobil chybu. Potreboval by overiť, či je jeho program správny.

Úloha 5.2. Pomôžte Dr. Vojšičovi overiť správnosť jeho programu. Naprogramujte kompresný algoritmus LZW aj Vy. Váš algoritmus by mal vypočítať počet bajtov v pôvodnom súbore a po skomprimovaní metódou LZW zistiť aj počet bajtov v skomprimovanom súbore.

Predpokladajte, že všetky vstupné súbory obsahujú znaky iba zo štandardnej sady ASCII (teda nie rozšírenej). Pri implementácii algoritmu LZW predpokladajte, že vytváraný slovník bude mať maximálnu kapacitu **256** položiek. Algoritmus by mal vypísať aj kódovaciu tabuľku. Pre kódovanie použite jednu z nasledujúcich alternatív:

- prvých **128** položiek je vyhradených pre jednotlivé bajty vstupného súboru (v ASCII),
- **(nepovinné)** algoritmus upravte tak, aby používal slovník s max. kapacitou 1024 položiek,
- **(nepovinné)** použite kódovanie podľa textovej abecedy so znakmi 'a', ..., 'z', 'A', ..., 'Z', '0', ..., '9' (spolu 62 znakov).

Algoritmus otestujte pre všetky zadané vstupné súbory. (Najprv pre súbor *udaje5_2a.txt*.)