

Popis úlohy:

Program zarovnáva text do bloků a to buď podle daných šířek bloku, nebo podle masky. Pokud slova nejdou naformátovat celá, tak se zalomí po slabikách, pokud to nejde ani po slabikách, tak se vypíše slovo celé i když to znamená porušení formátování.

Vstup a výstup/použití:

Program přijímá právě 3 argumenty a to soubor formátování, vstupní soubor a výstupní soubor, v tomto pořadí. Program po spuštění vypíše text z vstupního souboru zarovnaný podle pravidel z formátování do výstupního souboru. Všechny soubory by měly být v kódování utf-8. Ze vstupního souboru se čtou pouze neblíže znaky, čili nezáleží na zarovnání vstupního souboru.

První řádek souboru udává mód zarovnávání: právě block pro zarovnávání do bloku, nebo právě mask pro zarovnání podle masky, samotné formátování je pak dáno od druhého řádku dál.

Pro zarovnávání do bloku jsou bloky udány dvojicemi celých čísel na samostatných řádcích, první udávající pro kolik řádků daná šířka platí a druhé udávající šířku. Pokud dané bloky nestačí pro vypsání celého vstupního textu, tak se bude opakovat poslední šířka.

Pro zarovnání podle masky má maska následující tvar: O pro mezery, X pro blok znaků. Řádky jsou dány řádky masky. Délka bloků textu a mezer je udána podílem délky vstupního souboru a počtu X v formátování. Více X za sebou se bere jako jeden větší blok.

Popis řešení:

Tvorba bloku probíhá tak, že se přidávají slova dokud je v bloku místo, jakmile místo není, tak se zkusí slovo zalomit, slova se pak vypisují s počtem mezer aby blok měl vždy správnou velikost, pokud nejdou mezery rozmístit zcela rovnoměrně, tak víc mezer budou mít levější slova. Slova co se ani po zalomení nemohou do bloku vejít se vypíší celá.

Block zarovnání pouze postupně čte formátování a vypisuje příslušně široké bloky.

Mask zarovnávání vypisuje bloky textu podobně, ale podle toho jak vyjde délka slov a jejich zalomení nemusí (délka vstupu)/(počet X) pasovat jako délka bloku, pokud je délka moc krátká, tak se nevypíše celý soubor, pokud je moc dlouhá, tak se může celý soubor vypsát před tím, než se dostane na všechny bloky textu, proto program nejdříve zkusí menší velikost bloku a v případě nouze jí zvětší.

Pro zalamování slov podle slabik program používá heuristický algoritmus. Pokud jdou po sobě samohláska, souhláska, samohláska, tak se zalamuje za první samohláskou. Pokud jdou po sobě samohláska, více souhlásek, samohláska, tak se zalamuje uprostřed skupiny souhlásek. Pokud je l nebo r mezi souhláskami, tak se chová jako samohláska. Ch nesmí být rozděleno na c-h.

Popis testovacích dat:

Pro spuštění programu s testovacími daty stačí odkomentovat `RunDemo()` v programu, výsledky budou v `Demo1O.txt`, `Demo2O.txt`, `Demo3O.txt`

První příklad ukazuje použití módu `block` s různými velikostmi bloků. Vstup je dokumentace programu.

Druhý příklad ukazuje mód `block` s hodně malou velikostí bloku a s větším vstupním souborem (60 tisíc slov). Vstup je *Babička* od Boženy Němcové.

Třetí příklad ukazuje mód `mask`, kde maska je střídání textu s mezerami. Vstup je odstavec *Lorem Ipsum*. Program byl zamýšlen pro použití s českým textem, ale funguje relativně dobře i s pseudo-latinou.

Osobní zhodnocení:

Byl jsem příjemně překvapen, jak dobře funguje heuristika pro zalamování slov podle slabik, také jsem spokojen s módem `block`.

Nejsem ale spokojen s zarovnáváním podle masky a problém plyne už ze zadání úlohy, není jasné, nebo nemusí vůbec existovat ideální velikost bloku. Ve většině případů bude zarovnávání fungovat dobře, ale skoro vždy nebudou slova perfektně pasovat do bloků, to se snadno vyřeší tak, že se velikost bloku zvětší, to ale může vést k situaci, kdy se vypíše celý vstup než se dostane na všechny bloky. Já jsem tedy zvolil variantu kdy se vypíše všechn text, ale nemusí se dostat na konec masky. Také je jasné že se nemůže dostat na všechny bloky pokud například bude v masce více X než je písmen ve vstupním souboru. Pokud bych opět dělal zarovnání textu do tvaru podle masky v budoucnu, tak bych určitě zvolil jiný způsob zadání masky a to nejspíše nějaký kde už samotné formátování udává délku bloku.