

## PJC – Semestrální práce

### 1. Zadání

Typický problém s jednoduchou paralelizací. Na vstupu je sada souborů, na výstupu je seřazený výpis obsažených slov a jejich četnost.

### 2. Popis implementace

Samotná metoda počítání výskytu slov je shodná pro obě řešení – multi-thread i single-thread. Soubor je transformován na stream a prochází se slovo po slově. Ta jsou ukládána do mapy, která využívá slovo jako klíč a jako hodnota je využita četnost slova.

Tento základní princip je zabalen ve dvou různých implementacích. Tou první je single-thread, která postupně prochází soubory předané jako parametry, pro každý počítá četnost slov, ty následně sloučí a vypíše.

Druhou implementací je multi-thread. Ten otevírá každý z předaných souborů ve vlastním vlákne. Výsledky z nich nakonec sloučí a vypíše.

Určitou slabinou použité implementace více vláken je, že program neposkytuje žádné zrychlení při zpracování jednoho souboru.

Z důvodu využití mapy je výpis četnosti slov řazen abecedně, nikoliv od slova s nejvyšším výskytem, po to nejméně časté. Výstup je ale formátován tak, aby jej bylo možné (v případě přesměrování do textového souboru) jednoduše importovat do programu Excel.

Fungování je detailněji popsáno pomocí komentářů v samotném kódu.

### 3. Ukázkové řešení

Jako vstup pro toto řešení je využit text českého ekumenického překladu Bible s odstraněnou diakritikou. Tento text je dán funkci třikrát, pro demonstraci paralelní implementace.

V případě single-thread implementace trval výpočet 41,8782 sekund, v případě multi-thread řešení 23,2215 sekund. Referenčnímu řešení v MATLAB trvala stejná úloha 2068,4042 sekund. Testovací sestava používá z CPU Intel Core i7 4700HQ (4C/8T).

Výše uvedený soubor, příslušný výstup programu, stejně jako několik dalších, je součástí tohoto projektu.