

# Instrukcja do laboratorium 2 z ZBP

## 1 Kontenery set i unordered\_set

Zmierzyć wydajność kontenerów `set` i `unordered_set`. Testy należy przeprowadzić na słownikach

- języka angielskiego — <http://sun.aei.polsl.pl/~sdeor/students/zbp/english.dic>,
- języka polskiego — <http://sun.aei.polsl.pl/~sdeor/students/zbp/polish.dic>.

Zmierzyć należy:

- czas wstawiania do kolekcji,
- średni czas wyszukiwania,
- wpływ liczby słów w kolekcji na czas wyszukiwania: dla słownika j. polskiego wykonać pomiary dla 5%, 10%, 20%, 40%, 70%, 100% słów.

## 2 Stworzenie „spellcheckera”

Stworzyć program sprawdzający poprawność tekstów. Słownik powinien być podawany jako parametr. Wyboru struktury danych dla słownika należy dokonać na podstawie wcześniejszych wyników. Program powinien tylko sprawdzać czy słowo z tekstu znajduje się w słowniku czy też nie i wypisywać na wyjście słowa niepoprawne wraz z kilkunastoma znakami, które występują przed i za nim.

Należy uwzględnić fakt, że jeśli:

- słowo w słowniku jest zapisane małymi literami, to w tekście może być zapisane małymi literami, od pierwszej wielkiej litery, bądź wielkimi literami (np. *słowo*),
- słowo w słowniku jest zapisane od wielkiej litery, to w tekście może występować od pierwszej wielkiej litery, bądź być zapisane wielkimi literami (np. *Polska*),
- słowo w słowniku jest inną kombinacją małych i wielkich liter (w tym jest zapisane tylko wielkimi literami), to może występować w tekście tylko w takiej samej formie (np. *EMPiK*, *IBM*).

Uwaga: Można zignorować kwestię tzw. „polskich znaków”.

## 3 Dodanie funkcjonalności sugerowania poprawnych słów\*

Dla każdego błędnego słowa (czyli takiego, które nie występuje w słowniku) zaproponować poprawne wersje opierając się tylko na korekcji pojedynczych błędów maszynowych, czyli błędów, które należą do następujących kategorii:

- pominięcie jednej litery, np. *korekcja* → *korkcja*,
- wstawienie jednej litery, np. *korekcja* → *koreklcja*,
- podmianę jednej litery, np. *korekcja* → *korelcja*,
- zamianę kolejności sąsiednich liter, np. *korekcja* → *korkecja*.

Należy w miarę możliwości stosować poznane algorytmy STL. Kwestie sprawdzania poprawności danych wejściowych i niezależności od języka można całkowicie pominąć — zakładamy język polski.