21.03.2014.

Zespół testowy grupa 3

Wersja pierwsza

Anna Chomicz

Agata Kacprzak

Damian Roguski

Daniel Brzeziński

Hubert Pietruczuk

Patryk Moreń

Filip Oraczewski

**Specyfikacja**

Celem projektu jest napisanie programy w C# z dziedziny programowania lingwistycznego, którego głównym zadaniem jest oddzielenie liczebników od pozostałych słów języka polskiego. Liczebniki pobierane są z pliku wejściowego (słownika), a następnie zapisywane do dwóch plików tekstowych. Pierwszy zawiera wszystkie posortowane liczebniki poza porządkowymi, które znajdują się w drugim pliku. Pozostałe słowa (nie liczebniki) zapisane są do trzeciego i jednocześnie ostatniego pliku.

Plikiem wejściowym do programu zamówionego przez klienta jest plik o nazwie *odm\_utf8* w formacie tekstowym o rozszerzeniu .txt. Plik zawiera słownik odmian w kodowaniu znaków utf-8. Znajdują się w nim odmiany wszystkich form językowych, w tym: rzeczowników, czasowników, przymiotników i liczebników. Dokument zawiera zarówno wyrazy pisane łącznie, jak i oddzielnie, a także skróty. Słowa w pliku ustawione są według kolejności alfabetycznej.

W wyniku pracy otrzymujemy dwa pliki. Pierwszy z nich zawiera liczebniki główne odmienione przez przypadki, drugi liczebniki porządkowe. Pliki wynikowe muszą być plikami tekstowymi, a ich nazwa powinna być tworzona automatycznie, w następującym formacie:

• *nazwa\_pliku\_wejściowego\_liczebniki\_glowne.txt*

• *nazwa\_pliku\_wejściowego\_liczebniki\_porzadkowe.txt*

Dodatkowo należy utworzyć plik wyjściowy (w formacie SJP.pl), zawierający dane pliku wejściowego, ale bez wierszy rozpoznawanych jako liczebniki. Wyrazy takie jak pół, ćwierć, ćwiartka, jednokrotnie, pięciokrotnie itp. zaliczamy jako liczebniki główne, natomiast pierwszy, drugi itp. jako liczebniki porządkowe.

Pierwszy z plików wyjściowych ma zawierać odpowiednio uporządkowane liczebniki wraz z ich odmianami w rodzaju męskim, żeńskim, nijakim, męskoosobowym i niemęskoosobowym. Każda forma ma być oddzielona od siebie tabulacją. Plik ten ma być posortowany według wybranego wcześniej sposobu, po wartościach lub alfabetycznie.

Opracowana aplikacja powinna rozpoznawać wyjątki odbiegające od standardowej odmiany liczebnika. Podczas przeszukiwania słownika należy uwzględnić : liczebniki ułamkowe (*ćwierć , pół, półtora*), liczebniki nieokreślone (*kilka, kilkanaście, kilkadziesiąt*), wybrane jednostki ilości (*kopa. mendel, gros, tuzin*), liczebniki wielokrotne (*jednokrotnie, dwukrotnie*), zaprzeczenia pisane łącznie (*niejeden*).

Pominąć należy: liczebniki wielorakie (*dwojaki, trojaki*), liczebniki wielowyrazowe (*dwadzieścia dwa, pięćdziesiąt cztery*), liczebniki mnożne (*podwójny, potrójny*), zaprzeczenia pisane rozłącznie (*nie raz*).

Proponuję aby zrobić logowanie do oprogramowania aby zapobiec niechcianemu użytkowaniu oprogramowaniem. Klient nie określił nazwy jednego z plików wyjściowych, który pozbawiony jest wszystkich liczebników, więc proponuję aby nosił on nazwę *Plik\_wejsciowy\_bez\_liczebnikow.txt .*

Klient życzył sobie aby, oprogramowanie wykonywało się nie więcej niż dwa dni, lecz nie określił na jakim sprzęcie ma to oprogramowanie działać, co jest bardzo istotne przy szacowaniu czasu, w jakim ma się wykonać oprogramowanie.

[informacje o licencji]