

Specyfikacja

Wersja trzecia

Adam Kostusiak, Agata Kacprzak, Anna Chomicz, Artur Chojnacki, Damian Roguski, Daniel Brzeziński, Ewelina Detkiewicz, Filip Oraczewski, Grzegorz Podbielski, Hubert Pietruczuk, Jarosław Szczepański, Karolina Wakulak, Martyna Grotek, Patryk Moreń, Paweł Kmieciak

1. Cel projektu

Celem projektu jest napisanie programu, którego zadaniem jest rozdzielenie linii z wyrazami z podanego programowi pliku wejściowego do osobnych plików wyjściowych. Kryterium rozdzielenia mają być liczebniki. Program ma być aplikacją konsolową, uruchamianą w sposób wsadowy. Oznacza to, że wszystkie parametry zostaną podane w trakcie uruchomienia programu. Nie ma żadnej interakcji pomiędzy użytkownikiem, a programem.

2. Dane wejściowe

Plikiem wejściowym z liniami, które będą rozdzielone w programie ma mieć nazwę „plik_wejscowy.txt”. Plik ten 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. Formy fleksyjne słowa są oddzielone od siebie przecinkiem i spacją.

Dane wejściowe powinny zawierać listę liczebników, które mamy uwzględnić jako liczebniki główne i porządkowe. Lista ta powinna zawierać nie tylko słowną wartość, lecz także ich wartość liczbową.

Opracowana aplikacja powinna wyszukać liczebniki główne jednowyrazowe (*pięć, siedemnaście*). Podczas przeszukiwania słownika należy uwzględnić również: liczebniki ułamkowe (*ćwierć, pół, półtora*), liczebniki nieokreślone (*kilka, kilkanaście, kilkadziesiąt*), staropolskie jednostki (*kopa, mendel, 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*), liczebniki zaprzeczenia pisane rozłącznie (*nie raz*).

Dane wejściowe zostaną dostarczone przez klienta.

3. Pierwszy plik wyjściowy

Pierwszy z plików wyjściowych ma zawierać odpowiednio uporządkowane liczebniki główne. Ma mieć nazwę „plik_wejsciovwy_liczebniki_glowne.txt”.

W związku z brakiem tabel końcówek określających formy liczebników (*rodzaj żeński, męski, nijaki, niemęskoosobowy, niemęskoosobowy*) program jedynie przepisuje całe linie, tak jak były one podane w pliku wejściowym.

Plik ten ma być posortowany według wybranego wcześniej sposobu: po wartościach lub alfabetycznie.

4. Drugi plik wyjściowy

Drugi plik ma zawierać liczebniki porządkowe. Mają być one przepisane całymi liniami, tak jak są w pliku wejściowym.

5. Trzeci plik wyjściowy

Dodatkowo należy utworzyć plik wyjściowy (w formacie SJP.pl), zawierający dane pliku wejściowego, ale bez wierszy rozpoznawanych jako liczebniki. Ma zostać zapisany jako „plik_wejsciovwy_bez_liczebnikow.txt .”

6. Czas działania

Oprogramowanie ma przetworzyć maksymalny plik wejściowy (ze strony SJP.pl) w mniej niż dwa dni na sprzęcie znajdującym się w pracowni komputerowej (s.3/18 b.34 na terenie uczelni SGGW).

7. Funkcje dodatkowe

Program ma wyświetlić informacje o autorach oraz informację o licencji.
Program powinien sygnalizować postęp pracy.

8. Licencja

Program będzie tworzony na warunkach Powszechnej Licencji Publicznej GNU, wydanej przez Fundację Wolnego Oprogramowania - według wersji 2-giej tej Licencji lub którejś z późniejszych wersji.

Oryginalna wersja angielska licencji: <http://www.gnu.org/licenses/old-licenses/gpl-2.0.txt>

Polska wersja licencji: <http://gnu.org.pl/text/licencja-gnu.html>