SWIT – Projekt

Piotr Wachowicz 12204

1. Opis celu:

Program na wejściu dostaje rozpoznany skan hasła ze słownika. Zakładamy, że słownik jest wyłącznie w języku polskim, np. słownik frazeologiczny. Hasło może być „zanieczyszczone”, tzn. proces rozpoznawania skanu nie wyprodukował „idealnego” rozwiązania i w wyniku mogą się pojawić nadmiarowe symbole, np. kropki. Po kliknięciu na słowo będzie możliwość wyszukania tego słowa w słowniku haseł (słownik haseł jest częścią programu). Znalezione hasło zostanie wyświetlone w przyjazny dla użytkownika sposób.

Podstawowym utrudnieniem jest to, że słowa nie będą w postaci podstawowej, konieczne będzie sprowadzenie słowa do postaci podstawowej i wyszukanie zmodyfikowanego w ten sposób słowa w słowniku.

1. Opis pliku wejściowego:

Program na wejściu otrzymuje skan pojedynczego hasła z słownika w wersji papierowej. Skan powinien być „dobrej” jakości (minimum 600dpi), oraz w miarę możliwości nie zanieczyszczony błędami skanowania.

1. Opis procesu wyszukiwania haseł w słowniku.

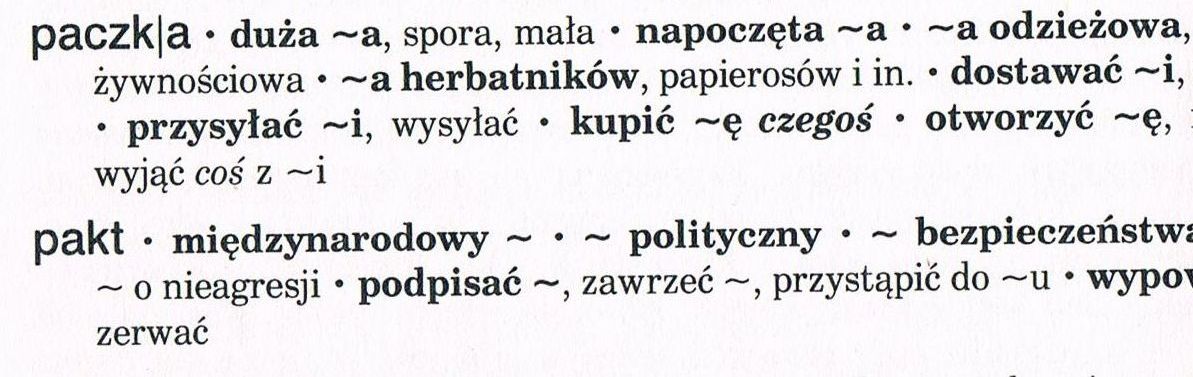
Program po wczytaniu wcześniej przygotowanego pliku z obrazem umożliwia jego rozpoznanie i przetworzenie do wersji cyfrowej. Odbywa się to przy użyciu darmowej biblioteki OCR TESSERACT. Tak przetworzony tekst można dodatkowo poddać „oczyszczeniu” po błędnym rozpoznawaniu.

Program posiada również wyszukiwarkę haseł w wbudowanych słownikach.

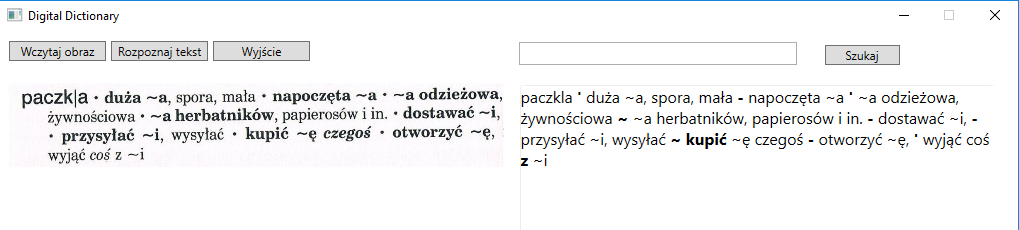
Podczas startu wyszukiwania najpierw słowo sprowadzane jest do postaci podstawowej – sprowadzane do rdzenia. Wykorzystywana jest do tego biblioteka morfologik. Następnie ładowane są słowniki w formacie .dict do pamięci komputera i następnie wyszukiwane jest słowo. Wyszukane słowo wyświetlane jest na ekranie w przyjazny dla użytkownika sposób.

1. Przykłady:

Obraz załadowany do programu:



Rozpoznany tekst:



Wyszukanie przykładowego hasła w słownikach:

