Projekt zaliczeniowy

Proces ETL

Grupa projektowa: Projekt02

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Imię | Nazwisko | Numer albumu | Grupa dziekańska | Wkład w prace nad projektem[[1]](#footnote-1) | Udział procentowy |
| Jakub | Rak | 213601 | WZISS2-1112 | Testowanie, instrukcja użytkownika, dokumentacja techniczna. | 30% |
| Marcin | Stanuch | 213597 | WZISS2-1112 | Testowanie, stworzenie klas do pobierania oraz transformacji produktów, aplikacja GUI. | 35% |
| Maciej | Małochleb | 213722 | WZISS2-1112 | Testowanie, stworzenie algorytmów i klasy do zasilania bazy danych. | 35% |

|  |
| --- |
| \_\_/70 pkt |

Spis treści

[Cel dokumentacji 3](#_Toc531894724)

[Wymagania systemowe programu 3](#_Toc531894725)

[Wymagania sprzętowe programu 3](#_Toc531894726)

[Rozpoczęcie działania programu 4](#_Toc531894727)

[Interfejs użytkownika 4](#_Toc531894728)

[Proces ETL 5](#_Toc531894729)

[Ekstrakcja danych 5](#_Toc531894730)

[Transformacja 5](#_Toc531894731)

[Załadowanie danych 6](#_Toc531894732)

[Automatyczny proces ETL 6](#_Toc531894733)

[Dodatkowe akcje w programie VubuETL 7](#_Toc531894734)

[Wyświetlanie dostępnej bazy danych 7](#_Toc531894735)

[Eksport danych do formatu CSV 7](#_Toc531894736)

[Czyszczenie bazy danych 7](#_Toc531894737)

# Cel dokumentacji

Niniejsza dokumentacja powstała jako instrukcja obsługi dla użytkowników programu VubuETL. Program służy do ekstrakcji, transformacji i załadowania danych do bazy, które pochodzą ze strony sklepu Vubu (<https://vubu.pl/>). Dane dotyczą ofert sprzedaży ubrań i zawierają między innymi nazwy i numery produktów, ceny, rozmiary.

## Wymagania systemowe programu

Program działa pod systemem Windows (Vista lub nowszy).

Dodatkowo, należy zainstalować pakiet plików Visual Studio 2017 dostępny na stronie:

<https://support.microsoft.com/pl-pl/help/2977003/the-latest-supported-visual-c-downloads>.

## Wymagania sprzętowe programu

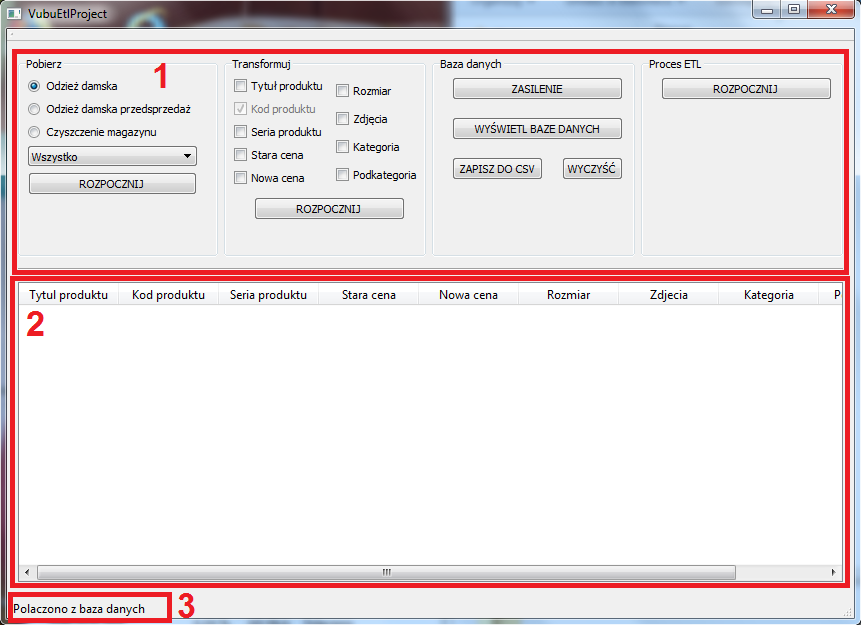
Minimalne wymagania sprzętowe:

* 150 MB przestrzeni dysku twardego
* 20 MB wolnej pamięci RAM
* procesora 500 MHz
* Szerokopasmowego łącza z siecią Internet o przepustowości 1 Mb/s

# Rozpoczęcie działania programu

Aby rozpocząć działanie z programem, należy uruchomić plik VubuEtlProject.exe dostępny w folderze Release. Prawidłowo działający program wyświetli interfejs użytkownika.

## Interfejs użytkownika



**Rys. 1.:** Interfejs użytkownika

Interfejs użytkownika składa się z trzech części:

1 – akcje użytkownika

2 – szczegóły działania programu

3 – komunikaty poprawności

# Proces ETL

Proces ETL polega na ekstrakcji (ang. Extract), transformacji (ang. Transform) i załadowania (ang. Load) danych z jednego systemu do drugiego. Ze względu na różne struktury systemów danych proces ten jest nieoczywisty i wymaga dodatkowego narzędzia w postaci programu lub zbioru programów służących do tego celu.

## Ekstrakcja danych

W tym procesie dane są pobierane z systemu pierwotnego a następne załadowane do pamięci podręcznej programu wykonującego proces ETL.

Aby uruchomić proces ekstrakcji w programie VubuETL, należy wybrać jedną z 3 podstawowych kategorii sklepu vubu.pl (Odzież damska, Odzież damska przedsprzedaż oraz Czyszczenie magazynu) oraz jednej z dostępnych opcji szczegółowych, które zawężają zakres ofert sklepu. Możliwa jest także ekstrakcja całego działu głównego.

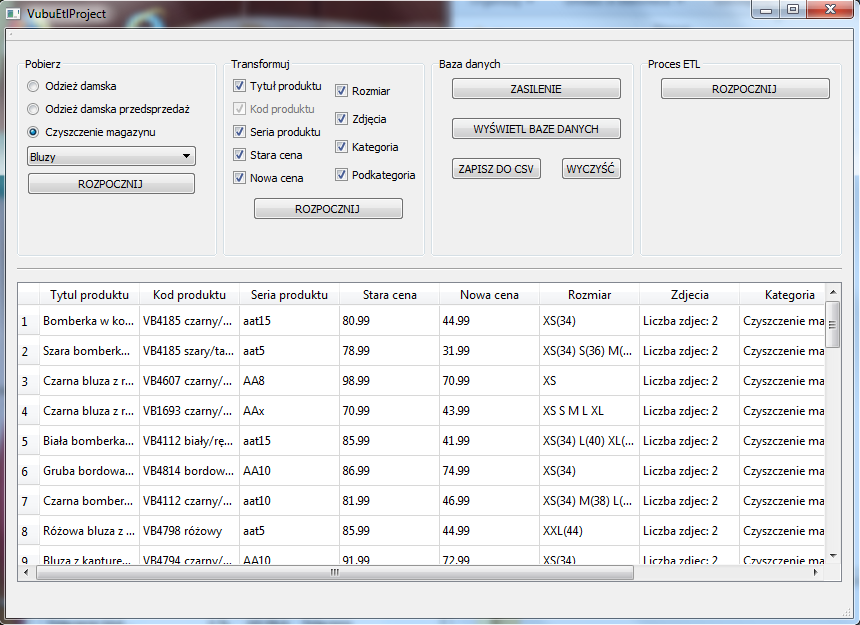
Czas pobierania ofert jest najdłuższym z procesów w programie i trwa on około 1 sekundę na każdą z ofert. Przykładowo, próbując dokonać ekstrakcji Odzieży damskiej w przedsprzedaży, wszystkie oferty, proces ten potrwa około 9 minut. Zakończenie procesu będzie zaznaczone odpowiednim komunikatem w polu komunikatów poprawności.

## Transformacja

Transformacja polega na zmianie jednego typu danych (wejściowych) na drugi (wyjściowy), przy czym dane wyjściowe muszą być zgodne z formatem danych systemu docelowego.

ProgramVubuETL proces ten realizuje poprzez opcje dostępne w polu „Transformuj”. Użytkownik może wybrać jakie elementy dotyczące ofert go interesują. Wszystkie wybrane należy zaznaczyć i później kliknąć w przycisk „Rozpocznij”. Proces zakończy się wyświetleniem wszystkich wybranych wartości w formie tabelarycznej w polu szczegółów działania programu.

Transformacja nie jest możliwa bez wcześniejszej ekstrakcji danych.



**Rys. 2.:** Przykład danych po transformacji

## Załadowanie danych

Proces załadowania danych polega na zapisaniu wybranych, wcześniej przygotowanych, rekordów do bazy danych systemu docelowego.

Program VubuETL umożliwia ten proces poprzez przycisk „Zasilenie” w kategorii „Baza danych” interfejsu użytkownika.

## Automatyczny proces ETL

Częstym rozwiązaniem jest wykonanie całego procesu w sposób w pełni zautomatyzowany. W takim przypadku poszczególne części procesu ETL wykonywane są synchronicznie po sobie, bez ingerencji człowieka.

W programie VubuETL istnieje możliwość wykonania automatycznego procesu ETL. Jeśli użytkownik wybierze kategorię danych do ekstrakcji oraz interesujące pola transformacji, może za pomocą przycisku „Rozpocznij” w polu „Proces ETL” dokonać automatycznego przejścia przez wszystkie elementy procesu ETL. Po zakończeniu procesu wyświetlony będzie odpowiedni komunikat.

# Dodatkowe akcje w programie VubuETL

Program VubuETL zawiera kilka dodatkowych funkcji pozwalających na usprawnienie procesów obróbki pozyskanych danych.

## Wyświetlanie dostępnej bazy danych

Po kliknięciu przycisku „Wyświetl bazę danych” w kategorii „Baza danych”, wszystkie zapisane wcześniej rekordy zostaną wpisane w formie tabelarycznej w polu szczegółów działania programu.

## Eksport danych do formatu CSV

Po kliknięciu przycisku „Zapisz do CSV” w kategorii „Baza danych”, wszystkie zapisane wcześniej rekordy zostaną zapisane w oddzielnym pliku w formacie CSV. Nazwa i ścieżka pliku na dysku są zdefiniowane przez użytkownika w odpowiednim oknie dialogowym.

## Czyszczenie bazy danych

Po kliknięciu przycisku „Wyczyść” w kategorii „Baza danych”, wszystkie zapisane wcześniej rekordy zostaną usunięte z bazy danych.

UWAGA! Ten proces jest nieodwracalny!

1. proszę wymienić konkretne zadania [↑](#footnote-ref-1)