

Hurtownie Danych

- projekt aplikacji ETL

Grupa dziekańska:

KrDUIs1011

Prowadząca:

mgr Katarzyna Wójcik

Autorzy:

Data Paweł

Konopka Dawid

Dubiel Kamil

Adres: aplikacji: <http://pablod.usermd.net/etlprojekt>

1. Technologie.

a) PHP w wersji 5+

+**SIMPLE HTML DOM** - Parser HTML

<http://simplehtmldom.sourceforge.net>

b) HTML5 wraz CSS3

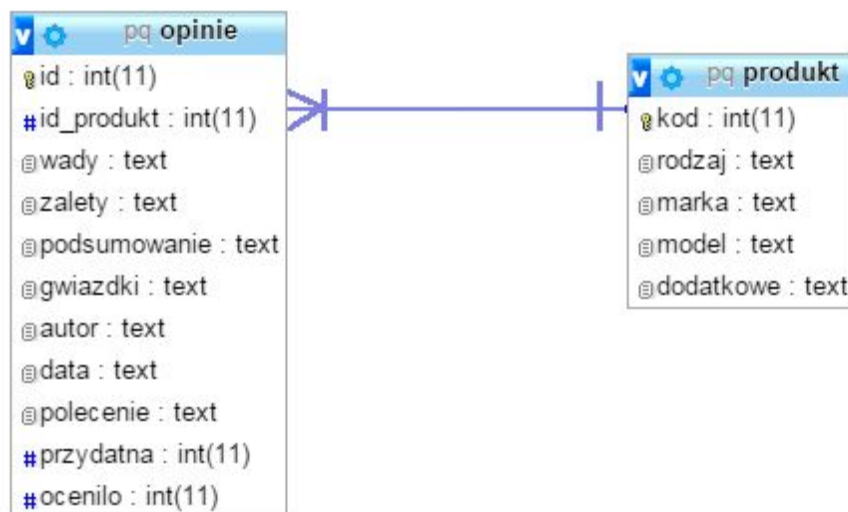
+ Bootstrap3 - Framework do HTML'a i CSS'a

<http://getbootstrap.com/>

c) MySQL 5.4 - Baza danych

d) Pliki aplikacji wraz z bazą danych przechowywane są na serwerze mydevil.net

2. Baza danych



3. Środowisko uruchomieniowe

Do uruchomienia aplikacji wymagany jest komputer lub smartphone z dostępem do internetu i systemem operacyjnym wyposażonym w jedną z wiodących przeglądarek internetowych.

4. Kody źródłowe wraz z objaśnieniem.

index.php

Wyświetlanie formularza, usuwanie zmiennych z poprzedniej sesji.

kod.php

*Odbiór formularza z kodem produktu.
Weryfikacja poprawności formularza.
Wywoływanie funkcji sprawdzającej istnienie produktu o konkretnym kodzie.
Rozpoczynanie procesu ETL.*

extraction.php

Pobranie danych o konkretnym produkcie z ceneo.

transformation.php

*Transformacja opinii odnoszących się do wczytanego produktu.
Ze względu na długi proces parsowania większej ilości opinii proces podzielony jest na "strony" zawierające po 10 opinii.*

load.php

Ładowanie bazy danymi na temat produktu i powiązanymi z nim opiniami.

products.php

Wyświetla listę produktów znajdujących się w bazie danych z odnośnikami do powiązanych opinii.

opinions.php

Lista wszystkich opinii powiązanych z wybranym produktem

clean.php

Czyszczenie bazy danych.

connect.php

Plik odpowiedzialny za logowanie do serwera MySQL

funkcje.php

Biblioteka funkcji wykorzystywanych w programie

countReviews(\$kod)

Opis:

Funkcja oblicza ilość opinii powiązanych z produktem na stronie internetowej.

Argumenty:

\$kod - Kod produktu

Wartość zwracana:

Ilość opinii powiązanych z produktem

extraction(\$id)

Opis:

Funkcja pobiera dane na temat produktu powiązanego z kodem

Argumenty:

\$id - Kod produktu

Wartość zwracana:

Pobrane z ceneo.pl dane dotyczące produktu

formularz()

Opis:

Generator formularza służącego do wprowadzania kodów produktów

Wartość zwracana:

Kod HTML formularza do kodu produktu

loadOpinions(\$kod, \$page, \$offset)

Opis:

Funkcja ładująca do bazy opinie na temat produktu wraz z powiązaniem ich z produktem

Argumenty:

\$kod - Kod produktu

\$page - Numer strony z opiniami

\$offset - Ilość opinii na danej stronie, zmienna ta pozwala na uniknięcie pobierania "nieistniejących" opinii w przypadku gdy ilość opinii na stronie jest mniejsza niż 10 (np. ostatnia strona)

loadProduct(\$kod)

Opis:

Funkcja ładuje dane na temat produktu do bazy danych.

Argumenty:

\$kod - Kod produktu

transformation(\$kod, \$page)

Opis:

Funkcja transformująca opinie o załadowanym produkcie

Argumenty:

\$kod - Kod produktu

\$page - Numer strony z opiniami

Wartość zwracana:

Tablica zawierająca opinie z danej strony wraz z etykietami.

transformationShow(\$kod, \$page, \$offset)

Opis:

Funkcja odpowiada z wyświetlanie pobranych z danej strony opinii dotyczących produktu

Argumenty:

\$kod - Kod Produktu

\$page - Numer strony z opiniami

\$offset - Ilość opinii na danej stronie, zmienna ta pozwala na uniknięcie pobierania "nieistniejących" opinii w przypadku gdy ilość opinii na stronie jest mniejsza niż 10 (np. ostatnia strona)