

Zadania rekrutacyjne

Zadanie 1. Pobranie i parsowanie danych HTML

Sparsowanie i wyciągnięcie danych atomowych produktów ze strony http://www.smartbay.pl/15-monitory_lcd/.

Należy wykonać zapytanie do wskazanej strony internetowej, następnie zwróconą treść należy sparsować a poszczególne wyniki zapisać do pliku CSV.

Uruchomienie skryptu powinno:

1. wykonać request do strony (np. poprzez cURL)
2. zapisać zwróconą odpowiedź do pliku HTML
3. sparsować dane do postaci atomowej (np. poprzez Simple HTML DOM)
4. wygenerować pliki CSV, gdzie pojedynczy wiersz będzie reprezentował pojedynczy produkt. Szkic tabeli:

Nazwa produktu	Cena od	Liczba sklepów	URL
Philips 223V5LSB	328,57	12	http://www.smartbay.pl/ceny/14184687-philips_223v5lsb.html
...
...

Uwaga! Dane pobrać można bez symulacji użycia przeglądarki, a co za tym idzie obsługi JavaScript (użyć można np. CURL'a).

Zadanie 2.1. Analiza i pokazanie danych z mySQL

Dane wejściowe dostępne są w plikach CSV lub tutaj:

https://docs.google.com/spreadsheets/d/1_d-6YR346AZd3ky8aDLVPvrqEWoG3Nn0BLcmQziBC_s/edit#gid=0

Załadować dane wejściowe do tabel bazy danych:

```
CREATE TABLE `products` (  
  `product_id` INT(11) NOT NULL AUTO_INCREMENT,  
  `product_date` DATE NOT NULL,  
  `producer_id` INT(11) NULL,  
  `product_url` VARCHAR(300) NOT NULL,  
  `product_name` VARCHAR(300) NOT NULL,  
  `product_price` DECIMAL(11,2) NULL DEFAULT NULL,  
  `product_price_old` DECIMAL(11,2) NULL DEFAULT NULL,  
  `product_reviews` INT(11) NULL DEFAULT NULL,  
  `product_rating` FLOAT NULL DEFAULT NULL,  
  PRIMARY KEY (`product_id`),  
  INDEX `product_date_producer_id` (`producer_id`, `product_date`)  
)  
COLLATE='utf8_general_ci'  
ENGINE=InnoDB;
```

```
CREATE TABLE `producers` (
  `producer_id` INT(11) NOT NULL AUTO_INCREMENT,
  `producer_name` VARCHAR(64) NOT NULL,
  PRIMARY KEY (`producer_id`)
)
COLLATE='utf8_general_ci'
ENGINE=InnoDB;
```

Następnie przygotować skrypt, który wygeneruje następującą tabelkę w kodzie HTML.

Daną wejściową mają być data od oraz data do. Przyjąć można, że dane są dla każdego dnia (tj. dane mają ciągłość i nie zdarzy się, że są dane dla 2017-03-28 oraz 2017-03-30, a nie ma danych dla dnia 2017-03-29).

Producer	% products share	Avg products
Producer A	34%	7
Producer B	22%	4
...		
Other	14%	3

Gdzie w pierwszej kolumnie pojawi się nazwa producenta, w drugiej kolumnie % produktów danego producenta w stosunku do wszystkich produktów, w kolumnie trzeciej wyświetlić się ma średnia liczba produktów danego producenta dziennie.

Produkty nie posiadające nadanego numeru ID producenta (`producer_id = NULL`) należy zakwalifikować jako "Other" i przedstawić w zestawieniu.

Zadanie 2.2 Rozwinięcie wizualizacji danych

Rozwinięcie skryptu z zadania 2.1, polegające na analizowaniu danych w rozbiciu na poszczególne dni, w formie tabelki:

Producer	2017-03-28	2017-03-29	2017-03-30
Producer A	47%	37%	23%
Producer B	18%	21%	27%
...			
Other	12%	11%	18%