

Totalna wyszukiwarka Wikipedii

Raport 1

1. Skład grupy:

- a. Szymon ApolinarSKI
- b. Mateusz Buśko
- c. Błażej Kubiński (lider)

2. Założenia zadania

- a. Wpisane zapytanie zwraca wyniki dla całej historii polskiej Wikipedii (łącznie z dyskusją).
- b. Możliwość zawężenia do określonego czasu.

3. Co zostało zrobione

- a. Pobrano i pobieżnie przejrzano stosunkowo nowy dump polskiej Wikipedii zawierający wszystkie zmiany (niestety nie udało się dotrzeć do sekcji zawierających dyskusje).
- b. Znaleziono szybki i prosty sposób na zaimportowanie dumpa do Solra (informacje [tutaj](#)).
- c. Repozytorium zostało założone na serwisie <http://github.com>
- d. Do zarządzania projektem wykorzystamy własnego [Redmine'a](#):
(login: *filipg* hasło: *isi-2014*). Projekt został już utworzony.

4. Napotkane problemy:

- a. Po wstępnym rozpoznaniu zadania jedynym problemem wydaje się brak dostatecznie dużego dysku - jedna z czterech części dumpu całkowitej wersji wikipedii po rozpakowaniu zajmuje ok. 120 GB danych, które i tak trzeba jeszcze wrzucić do Solra.
- b. Po rozmowie z administratorami WMI, dowiedzieliśmy się, że możemy otrzymać jedynie maszynę wirtualną z niewielką powierzchnią dyskową (zaledwie kilka GB).
- c. Nie udało się znaleźć starszych dumpów, które byłyby mniejsze.

5. Co zostanie zrobione:

- a. Dostosowanie indeksowania Solra do potrzeb zadania.
- b. Zindeksowanie wszystkich wpisów z określonego zakresu.
- c. Rozpoczęcie prac nad frontendem umożliwiającym wyszukiwanie i przeglądanie wyników.
- d. Konfiguracja Solra, aby zwracał odpowiednio przygotowane wyniki.

6. Zobowiązanie do wykonania zadania:

- a. Warunki takie, jak w punkcie 2 pod warunkiem uzyskania dostępu do maszyny z dużą pojemnością dyskową
- b. Minimalistyczny, aczkolwiek działający serwis umożliwiający przeglądanie wyników
- c. Możliwość podglądu zmian dla danego artykułu