

Drugi raport z prac nad projektem

“Wyszukiwarka dziwnych wydarzeń”

1. Opis prac nad projektem

W chwili obecnej jesteśmy na etapie konfiguracji i instalacji Solr'a oraz pozyskiwania materiałów źródłowych dla naszej wyszukiwarki. Artykuły scrawlowane są już wyczyszczone i jak najbardziej nadają się jako dane wejściowe do silnika Solr. W najbliższych planach mamy dalsze crawlowanie, oraz ukończenie prac nad instalacją Solra, aby można było załadować do niego zgromadzone materiały.

2. Zadania wykonane w ostatnim tygodniu

- Scrawlowanie wikipedii (około 500 artykułów)
- Scrawlowanie strony niewiarygodne.pl (około 4500 artykułów)
- Rozpoczęcie pracy nad konfiguracją i instalacją Solr'a

3. Plan na następny tydzień

- Konfiguracja Solra
- Utworzenie kolekcji danych dla Solra
- Zbudowanie bazy danych na podstawie scrawlowanych artykułów
- Dalsze crawlowanie
- Rozpoczęcie prac nad front-endem do Solra