

Report 2

Database

- to store data for processing a PostgreSQL database was set up (Version 9.5 on Linux Ubuntu 16.04)

Dictionaries

- page title information are only provided for the last 3 months: <https://dumps.wikimedia.org/other/pagetitles/>
- wikipedia dumps could only be found for the last months: https://dumps.wikimedia.org/other/static_html_dumps/
- there are however page title dumps some very early and very specific points in time 2007 -2008: https://dumps.wikimedia.org/other/static_html_dumps/
- it was not possible to obtain information on page history of that kind that would allow to build a graph of old page titles or reconstruct page merges, splits and so on
- dictionary tables are of the following form:

```
CREATE TABLE IF NOT EXISTS dict_de (  
    page_id serial,  
    page_name text unique not null  
);
```