

Java Project Report 2

Here are the personal reports :

Aida :

I added a search class that uses hashmap for searching in the corpus with its Junit test. I also edited some parts of opennlp part now we can also use URL to get the text, we have a language detector and we can do text analysis in English and German.

Lea :

Lea pushed a simple GUI code and she is working on how to put image as background for elements and making it better

KaiYi:

KaiYi uploaded buttonhandler for GUI part

Daniyal:

Daniyal finished the whole Jsoup - web scraper - task right now the current version :

* Features:

- * - Connect to a Wikipedia page and scrape its contents.
- * - Extract main title and language of the Wikipedia page.
- * - Extract sections of the page including the main content.
- * - Extract text based on specific tags or classes.

*

* Version: V1.3

* Changes:

- * - Added a new constructor that takes a topic and language to generate the Wikipedia URL.

- * - Updated manual and usage instructions.

- * - Improved error handling for the new constructor.

He also created a V1.3 manual pdf for the wikipedia scraper class

Summary :

We decided to divide the tasks like this :

Daniyal's tasks: web scraping (done) and saving to file as XML (in progress)

Lea's task: coding the whole GUI part except the handlers

KaiYi's task: coding the handlers

My tasks: language processing and Creating Jar file