ABSTRAK

Anderson Montella Karel

Skripsi

Web Scraping dan Information Retrieval dari Google Scholar, Sinta, ScimagoJr Untuk Pencatatan Otomatis Publikasi Dosen UKP

Setiap publikasi yang dikeluarkan oleh dosen perlu dicatat sebagai kinerja perguruan tinggi, tetapi pencatatan ini masih dilakukan secara manual oleh dosen yang mengeluarkan publikasi itu sendiri. Pencatatan manual ini membuat tugas administrasi dosen bertambah. Selain itu data sitasi yang dicatatat juga sering berubah, yang membuat pencatatan secara manual tidak memungkinkan.

Teknologi web scraping dan headless web browser dapat digunakan untuk mengotomasikan proses pencatatan publikasi, sehingga tugas dosen berkurang. Script web scraping akan dibuat sebagai API yang bisa dipanggil untuk melakukan pencatatan secara otomatis. Data yang didapatkan melalui proses scraping akan disimpan pada database PostgreSQL. Data publikasi yang disimpan dalam database dapat dilihat melalui sebuah aplikasi web sederhana yang dibuat dengan framework Laravel.

Dari pengujian data, sistem ini bisa menemukan dan melakukan pencatatan otomatis pada publikasi dosen yang tercatat pada website sumber, ini ditandai pada kesamaan data abstrak 100%. Tetapi sistem masih mempunyai masalah untuk memberikan kategori yang akurat, ini ditandai dengan kesamaan data yang sangat rendah yaitu 7%. Sistem juga memperlukan id profile dosen pada website sumber untuk bisa memulai proses scraping nya.

Keywords: *web scraping, information retrieval,* google scholar, sinta, scimagojr

ABSTRACT

Anderson Montella Karel

Undergraduate Thesis

Web Scraping and Information Retrieval from Google Scholar, Sinta, ScimagoJr for the Automatic Recording of PCU's Lecturer's Publication

Every publication published by a lecturer needs to be recorded. Right now, the recording process is done manually and done by the lecturer who published the publication. The manual recording adds administrative tasks for the lecturer. The data recorded can also change, like the number of citations a publication has which makes manual recording impossible.

Web scraping and headless web browser technology can be used to automate the publication recording process, reducing lecturer's task. The web scraping script will be made as an API that can be called to do the recording process automatically. The data got from the scraping process will be stored in a PostgreSQL database. The publication data stored in the database can be viewed in a simple web application made with the Laravel framework.

From testing the data, the system can find and record the publication automatically, indicated by the 100% abstract data match. But the system still have problem getting the ISSN, and the publication date, and the system cannot give an accurate category indicated by the low data match, which is 7%. The system also needs the id profile of the lecturer on the resource website to be able to start the scraping process.

Keywords: *Web Scraping, Information Retrieval,* Google Scholar, SINTA, ScimagoJr

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR SEGMEN PROGRAM