

Deskripsi Dokumen: <http://lib.ui.ac.id/opac/themes/libri2/detail.jsp?id=41970&lokasi=12>

---

Abstrak

#### ABSTRAK

Nama : Mercia

Program Studi : Ilmu Komputer

Judul : Metode Pembentukan Korpus Teks dengan Memanfaatkan

Twitter: Studi Kasus Bahasa Jawa

Indonesia memiliki banyak bahasa daerah yang perlu dilestarikan keberadaannya. Salah satu cara untuk dapat melestarikan bahasa daerah adalah dengan membuat korpus teks bahasa daerah. Twitter sebagai sumber aliran data dari berbagai bahasa di dunia dapat dimanfaatkan sebagai sumber korpus teks. Terdapat dua API Twitter yang dapat digunakan untuk melakukan crawling tweet suatu bahasa daerah, yaitu Streaming API dan REST API. Streaming API dapat digunakan untuk mengoleksi tweet secara realtime berdasarkan kata kunci pada parameter track. REST API dapat digunakan untuk mencari tweet sampai tujuh hari terakhir dengan suatu parameter query dan mencari tweet-tweet dari timeline seorang pengguna Twitter. Tugas akhir ini melaporkan pemanfaatan kedua API Twitter agar dapat menemukan konfigurasi yang dapat menghasilkan tweet sebanyakbanyaknya, presisi yang paling bagus, dan yang memiliki keseimbangan terbaik antara nilai presisi dan jumlah tweet yang dihasilkan. Selain itu, juga akan terdapat informasi mengenai resource korpus bahasa daerah yang berhasil penulis buat dan dapat diunduh secara gratis.

Kata Kunci:

Bahasa daerah, crawling, tweet, timeline, REST API, Streaming API, query, track, korpus teks.