

ABSTRAK

Penggunaan bahasa daerah saat ini semakin tergerus karena sistem pendidikan di banyak sekolah karena kurang mengadopsi pemeliharaan nilai-nilai budaya lokal dan tradisional dan justru terlena oleh hal-hal yang terlalu berbau internasional tanpa melalui filterisasi sehingga generasi muda saat ini lebih sering menggunakan bahasa Indonesia. Untuk melestarikan bahasa Sunda, maka diperlukan pengumpulan kata secara digital. Pengumpulan kata bahasa Sunda secara digital pernah dilakukan pada proyek akhir yang dikerjakan oleh Novihana Nurani Setiawan dengan judul Pembangunan Basis Data Leksikal Wordnet Bahasa Sunda. Pada proyek akhir tersebut pengumpulan kata bahasa Sunda masih dilakukan dengan cara memasukkan kata satu per satu akibatnya memerlukan waktu yang lama untuk pengumpulan kata dan butuh pengguna yang banyak, sedangkan pengguna yang memasukkan kata tidak dibatasi pada pengguna dari suku Sunda saja. Oleh karenanya dibangunlah Aplikasi Ekstraksi Kata Bahasa Sunda. Aplikasi Ekstraksi Kata Bahasa Sunda merupakan sebuah aplikasi yang berguna untuk melakukan pengumpulan kata pada halaman web wikipedia bahasa Sunda dan bahasa Indonesia. Kata yang telah dikumpulkan nantinya akan dibandingkan dengan kamus *online* untuk mengetahui arti setiap kata. Dalam pembuatan Aplikasi Ekstraksi Kata Bahasa Sunda akan menggunakan PHP dengan *library* HTML *Simple* DOM. Kata yang terkumpul dari hasil *crawling* diklasifikasikan ke dalam kata Sunda dan kata bahasa Indonesia.

Kata Kunci: *Crawler*, *Sinonim set*, Sunda, PHP, HTML *Simple* DOM