

ABSTRAK

Internet merupakan kebutuhan pokok bagi banyak orang di era informasi. Informasi yang terdapat di internet sangat banyak. Search engine merupakan sistem yang mampu melakukan pencarian dan pengambilan informasi berdasarkan kata kunci yang dimasukkan oleh pengguna. Blog merupakan salah satu penyumbang informasi yang notabene dibuat oleh para *blogger*.

Jumlah *blogger* di Indonesia semakin bertambah, berbanding lurus dengan jumlah informasi di internet. Dari kondisi tersebut diperlukan mekanisme mempercepat proses pengambilan informasi dari pemilahan blog berdasarkan kategori yang telah ditentukan sebelumnya. Web crawler aplikasi yang berfungsi mengambil informasi yang diperlukan dengan cara menjelajahi halaman situs blog kemudian disimpan dalam database.

Proses pengambilan halaman web dilakukan pengecekan bahasa menggunakan *regular expression* dan daftar *stopword* untuk tujuh kategori yang ditentukan berdasarkan *niche blog* : tips dan trik komputer, kuliner nusantara, sepak bola, petualangan indonesia, kesehatan Ibu dan anak, handphone dan gadget, fashion wanita.

Niche blog adalah aktivitas *blogging* yang dilakukan oleh para *blogger* untuk membuat blog bertopik spesifik. Dengan mengaplikasikan konsep *web crawler*, penelitian ini mampu menentukan *niche blog* sesuai dengan tujuh kategori yang telah ditentukan dengan mengujikan 10 kata kunci dan daftar *stopword*. Dari hasil ujicoba, dapat diambil kesimpulan penggunaan 3 kunci mendapatkan hasil paling maksimal.

Keyword: *Niche Blog, Regular Expression, Web Crawler, Search Engine*