

BAB I

PENDAHULUAN

1.1. Latar Belakang

Blog merupakan salah satu penyumbang informasi yang notabene dibuat oleh seorang *blogger*. Data *ictwatch* tahun 2013, jumlah pengguna *blog* di Indonesia kurang lebih 3,3 juta orang . Hal ini berbanding lurus dengan jumlah informasi di *internet* yang semakin berkembang. Sehingga banyaknya informasi yang terus bermunculan dari media *blog* tersebut muncul kendala yang akan dihadapi yaitu kesulitan mengetahui *niche blog* serta belum adanya kategori-kategori dalam pencarian *blog* tertentu yang memudahkan para pencari berita khususnya melalui media *blog*. Untuk mendapatkan informasi yang tepat dan cepat diperlukan alat bantu *search engine* yang terdapat pencarian informasi khusus dari *blog* dengan cara mengetahui *niche blog* tersebut kemudian dimasukkan dalam kategori-kategori yang telah ditentukan sebelumnya .Kategorisasi *blog* yang merupakan suatu proses pengklasifikasian dokumen ke dalam satu atau lebih kategori yang telah didefinisikan sebelumnya atau ke dalam kelas-kelas dari dokumen-dokumen yang sama, telah menjadi suatu teknik kunci dalam penanganan dan pengorganisasian data yang berupa teks. Teknik-teknik kategorisasi *blog* digunakan secara umum untuk mengklasifikasikan *niche blog*, menemukan informasi yang menarik di WWW.

Beberapa *search engine* telah memanfaatkan *web crawler* untuk memperoleh informasi tersebut. *Web crawler* merupakan sebuah perangkat lunak yang digunakan untuk menjelajahi halaman-halaman di *internet* dan akan mengambil informasi-informasi yang tersedia di halaman *web* tersebut. *Web crawler* juga dapat digunakan untuk mencari informasi dengan topik tertentu sehingga disebut dengan *topical web crawler*. [1]

Aplikasi tugas akhir yang akan dibangun ini menggunakan *web crawler* untuk melakukan pencarian ke halaman-halaman *blog* yang terdapat di *internet* untuk mencari data-data mengenai *niche blog*. Tidak semua *website/situs* akan dikunjungi oleh *web crawler* ini, hanya *blog-blog* berdomain tertentu yang telah ditentukan sebelumnya yang akan menjadi tujuan *crawling*. Setelah data yang diinginkan telah ditemukan,

maka data tersebut akan disimpan di *database* dan akan digunakan untuk proses selanjutnya.

1.2. Perumusan Masalah

Dari latar belakang yang telah dijelaskan dapat dimunculkan suatu permasalahan yaitu bagaimana menerapkan regular expression untuk mengelompokkan informasi dari media blog dengan banyaknya jumlah informasi pada media *blog* yang terus bertambah.

1.3. Tujuan dan Manfaat

Tujuan dari penelitian ini adalah menerapkan sebuah konsep *web crawler* menggunakan *Regular Expression* yang mampu menentukan *niche blog* dan mengklasifikasikan blog sesuai kategori-kategori yang telah ditentukan. Sehingga dengan adanya pengklasifikasian terhadap blog sesuai kategori-kategori yang telah terbentuk ini mempercepat proses pencarian berita dari media blog pada khususnya dengan cara mengelompokkan dan menentukan blog sesuai dengan *niche blog*.

1.4. Batasan Masalah

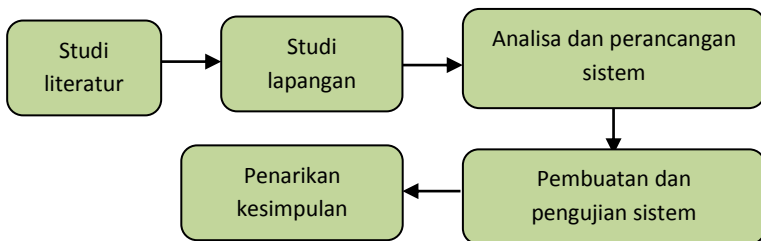
Penelitian tugas akhir ini memberikan batasan masalah yang mampu membuat arah penelitian lebih jelas yang diuraikan sebagai berikut.

- a. Mesin crawler yang digunakan adalah RISE Project yang disesuaikan dengan kebutuhan penelitian.
- b. Menjadikan sebuah sistem yang dapat mengelompokkan blog berdasarkan tujuh kategori yang telah ditentukan yaitu :
 1. Tips & Trik Komputer
 2. Kuliner Nusantara
 3. Sepak Bola
 4. Petualangan Indonesia
 5. Kesehatan Ibu & Anak
 6. Handphone & Gadget
 7. Fashion Wanita
- c. Blog yang di *crawl* adalah blog dengan engine :

1. Wordpress & Wordpress Self Hosting
 2. Blogspot
 3. Blogdetik.
 4. Tumblr
 5. Opera
 6. Joomla
 7. Drupal
 8. Blogfam
 9. Kompasiana
 10. Dagdigdug
 11. Posterous
- d. Blog yang di *crawl* adalah blog yang ter-*index* di mesin pencari seperti : *google, yahoo, atau bing*.

1.5. Metodologi Penelitian

Terdapat beberapa tahap dari metodologi yang di terapkan pada penelitian ini. Diantaranya dari studi literatur, studi lapangan, analisa dan perancangan sistem, implementasi dan pengujian, dan yang terakhir penarikan kesimpulan. Berikut adalah alur serta penjelasan dari masing-masing alur tersebut.



Gambar 1.1 Alur metodologi penelitian

1.5.1. Studi Literatur

Sebelum melakukan penelitian ini, dilakukan pembelajaran materi mengenai *information retrieval, search engine, webcrawler, regular expression*, pengelompokan *niche blog*.

1.5.2. Studi Lapangan

Diketahui dari uji coba program yang telah di implementasikan sebelumnya oleh Arifin(2011). Menjalankan *web crawler* pada waktu-waktu tertentu dan uji coba dengan 5-60 *thread*. Proses yang terjadi diantaranya adalah pengambilan web dengan pengecekan bahasa terlebih dahulu dengan *regular expression*. Setelah halaman *web* sudah tersimpan, *web crawler* membersihkan dari *tag-tag html* dan *stoplist* yang ada. Sehingga data bersih yang tersimpan adalah isi web yang bersih. Dan uji coba ekstraksi informasi web untuk menghasilkan bagian-bagian web yang sudah dikelompokkan menurut perbandingan akar dari pohon yang terbentuk dari laman web.

Selain itu, uji coba program sebelumnya yang telah dilakukan oleh Qomaruddin(2013). Ekstraksi informasi *simple tree matching* berhasil diimplementasikan ke dalam *web crawler* bahasa Indonesia. Hasil ekstraksi konten yang didapatkan dari *web crawler* bahasa Indonesia ini diperoleh rata-rata nilai *precision* 100% dan *recall* 89.31365%.

1.5.3. Analisa dan Perancangan Sistem

Merancang sistem dan menganalisa kebutuhan apa saja yang akan dibutuhkan untuk membuat aplikasi di penelitian ini. Baik berupa modul-modul internal program dan perangkat keras (*hardware*) yang dipergunakan. Setelah mendapatkan semua bahan yang dibutuhkan mulailah perancangan sistem.

1.5.4. Pembuatan dan Pengujian Sistem

Mengimplementasikan rancangan sistem yang sudah di buat dalam bentuk fisik sekaligus melakukan pengujian. Perbandingan analisis berbeda di perangkat keras dan setelah itu baru ditarik sebuah kesimpulan.

1.5.5. Penarikan Kesimpulan

Pada tahap terakhir dalam penelitian ini adalah penarikan kesimpulan. Beberapa hal penting dapat dicatat dan disusun di dalam sebuah laporan akhir.

1.6. Sistematika

Sistematika penulisan karya ilmiah ini dibagi menjadi beberapa bab yang disusun sesuai tahap pengerjaan. Berikut kami jelaskan per poin:

BAB I : PENDAHULUAN

Pada bagian ini terdiri dari : latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penulisan.

BAB II : TINJAUAN PUSAKA

Pada bagian ini dijelaskan mengenai hal-hal yang melandasi permasalahan yang diangkat sebagai pedoman untuk menyelesaikannya.

BAB III : PERANCANGAN SISTEM

Pada bagian ini akan dijelaskan pemodelan sistem yang akan dibangun. Modul-modul yang dibutuhkan sistem didesain secara lengkap, sehingga tujuan tercapai secara maksimal.

BAB IV : SIMULASI DAN ANALISIS

Pada bab ini menjelaskan tentang implementasi program yang telah dikerjakan. Meliputi penjelasan tentang arsitektur sistem, desain sistem, desain interface, dan menguraikan tentang hasil uji coba serta analisa terhadap sistem.

BAB V : KESIMPULAN

Bab yang berisi kesimpulan dan saran yang dihasilkan dari sistem yang telah diujicobakan.

(Halaman ini sengaja dikosongkan)