

BAB 1

PENDAHULUAN

Pada bab ini dijelaskan latar belakang dari penelitian yang memberikan gambaran awal mengenai perolehan informasi blog dan perannya dalam perkembangan teknologi informasi. Tujuan dan ruang lingkup pengerjaan tugas akhir memberikan penjelasan mengenai hasil yang ingin diketahui serta batasan dalam pengerjaan. Selain itu, dijelaskan pula tahapan dalam metodologi penelitian dan sistematika penulisan laporan.

1.1 Latar Belakang

Teknologi informasi yang berkembang dengan pesat membuat internet semakin diminati oleh masyarakat. Melalui internet, masyarakat dapat memperoleh berbagai informasi berupa teks, gambar, ataupun suara dengan mudah dan cepat. Salah satu jenis informasi berbentuk teks yang kini mulai digemari adalah blog. Blog pertama kali muncul ketika pada tahun 1993, Mosaic, salah satu perusahaan *web browser* menerbitkan sebuah halaman pada situs *web* yang dinamakan 'What's New' (Blood, 2000). Halaman tersebut berisi daftar dan *link* ke situs *web* lainnya yang dapat dikunjungi oleh para pengguna *browser*. Seiring dengan kemajuan internet, banyak orang kemudian mulai berinisiatif untuk menampilkan situs *web* yang dianggap menarik di halaman *web* pribadi milik mereka. Gaya penulisan *web* ini kemudian menjadi semakin populer sehingga sebuah nama dinilai perlu diberikan untuk membedakannya dengan gaya penulisan *web* pada umumnya. Jorn Barger, seorang penulis, kemudian menjadi orang pertama yang menggunakan istilah "*web log*" pada tahun 1997. Selanjutnya, pada awal tahun 1999, Peter Merholz, seorang konsultan antarmuka pengguna dan desainer *web*, menggunakan istilah "*wee blog*" pada *weblog* miliknya, yang kemudian disingkat menjadi "blog".

Blog adalah bentuk aplikasi *web* yang menyerupai tulisan-tulisan (yang dimuat sebagai *posting*) yang diperbaharui secara berkala dengan tulisan terbaru berada pada urutan teratas (Brady, 2005). Informasi yang dikandung oleh suatu blog sangat bervariasi. Penulis dari blog (yang dikenal sebagai *blogger*) dapat

menggambarkan kegiatan keseharian mereka, atau menceritakan topik yang lebih spesifik yang sedang diminati, seperti desain *web* atau fotografi. Beberapa blog yang sering dikunjungi merupakan blog dengan topik tertentu (contohnya adalah <http://rovicky.wordpress.com/>, blog mengenai geologi yang cukup populer) atau blog yang berisi opini mengenai peristiwa yang sedang terjadi (contohnya adalah <http://ndorokakung.com/>).

Perolehan informasi adalah pencarian informasi (dokumen) pada keadaan yang tidak terstruktur (teks) dari sebuah koleksi yang besar untuk memenuhi kebutuhan informasi (Manning, 2008). Contoh aplikasi perolehan informasi yang paling umum digunakan adalah mesin pencari *web*. Pengguna *web* dapat melakukan pencarian dokumen *web* dengan memasukkan kata kunci ke dalam mesin pencari *web*. Merujuk pada penelitian yang dilakukan oleh Broder (Broder, 2002), kata kunci (atau umum disebut kueri dalam bidang perolehan informasi) yang umumnya dimasukkan oleh pengguna pada mesin pencari *web* terbagi menjadi tiga macam, yaitu *informational* (penemuan informasi mengenai topik tertentu), *navigational* (penemuan *web* site secara spesifik), dan *transactional* (untuk melakukan aktivitas transaksi dengan menggunakan media *web*).

Perolehan blog (*blog retrieval*) merupakan perolehan informasi yang dikembangkan untuk pencarian dokumen *web* berbentuk blog. Ada beberapa karakteristik yang membedakan perolehan blog dengan perolehan dokumen *web* biasa. Pertama, dari segi kueri Mishne dan de Rijke (Mishne & Rijke, 2006) dalam penelitiannya menemukan bahwa kueri yang dimasukkan pada mesin pencari blog berbeda dengan mesin pencari *web* pada umumnya. Kueri yang bersifat *transactional* dan *navigational* sangat jarang ditemukan pada perolehan informasi blog. Biasanya, kueri yang digunakan pada mesin pencari blog bersifat *informational*, pengguna umumnya menggunakan mesin pencari blog untuk mencari informasi mengenai suatu topik yang sedang diminati (Mishne & Rijke, 2006). Kedua, dari segi struktur dokumen, pencarian blog dapat berupa pencarian beberapa dokumen, karena blog dapat dianggap sebagai satu halaman *web* dengan satu atau banyak tulisan. Ketiga, dari segi struktur bahasa, tidak seperti pada dokumen *web* berupa artikel atau berita, blog biasanya mengalami proses

pemeriksaan yang minimum dengan bahasa yang lebih mengarah pada percakapan dan bersifat tidak teratur (Elsas et.al, 2008). Hal inilah yang kemudian membedakan mesin pencari blog dengan mesin pencari *web* biasa.

Fenomena blog yang semakin berkembang memunculkan kebutuhan akan mesin pencari blog yang efektif. Berbagai penelitian telah dilakukan untuk menghasilkan suatu mesin pencari dokumen blog yang efektif, diantaranya adalah penelitian mesin pencari blog berbahasa Inggris yang cukup banyak dilakukan oleh berbagai institusi pendidikan pada konferensi TREC (*Text Retrieval Conference*)¹ yaitu Blog Track sejak tahun 2006 hingga sekarang. Ada berbagai aspek yang diteliti dalam pencarian blog, diantaranya adalah pencarian opini pada tulisan blog, klasifikasi opini pada tulisan blog (negatif, positif, atau keduanya), pencarian blog yang berisikan suatu topik tertentu (*feed search*), dan pendeteksian blog berupa *spam* (*splogs*).

Salah satu aspek yang paling menarik dan diminati adalah menemukan opini mengenai suatu topik pada blog. Penelitian pencarian opini pada blog dikembangkan untuk merepresentasikan pencarian blog oleh pengguna yang cenderung membutuhkan informasi berupa opini daripada fakta. Pencarian opini melalui blog dapat membantu dalam menemukan pendapat publik mengenai suatu target, yang kemudian dapat digunakan untuk memantau merek suatu produk, analisis media, dan menelusuri kebutuhan pasar (Macdonald, Ounis, & Soboroff, 2007).

Blog yang bisa diterbitkan dan diperbaharui dengan cepat dan mudah turut menarik perhatian sebagian masyarakat Indonesia. Adanya acara tahunan Pesta Blogger yang merupakan acara berkumpulnya para blogger se-Indonesia menunjukkan perkembangan blog di Indonesia yang cukup pesat. Hal ini kemudian mendorong adanya penelitian mengenai mesin pencari blog berbahasa Indonesia. Penelitian mesin pencari blog berbahasa Indonesia diperlukan mengingat adanya perbedaan penulisan antara blog berbahasa Indonesia dan Inggris. Penelitian ini diharapkan mampu membantu memenuhi kebutuhan akan

¹ <http://trec.nist.gov/>

informasi berupa opini publik akan fenomena-fenomena yang sedang terjadi di masyarakat, khususnya di Indonesia.

1.2 Permasalahan

Hingga kini, penelitian perolehan dokumen blog berbahasa Inggris telah berkembang sangat pesat. Beberapa diantaranya adalah penelitian pencarian opini pada tulisan blog (Zhang, Yu, & Meng, 2007), klasifikasi opini (Lin & Chen, 2007), pendeteksian blog *spam* (Java et.al, 2007), dan pencarian blog (Elsas et.al, 2008). Penelitian pencarian blog untuk bahasa Indonesia belum pernah dilakukan, namun untuk bahasa Inggris ada dua pendekatan yang umumnya digunakan, yaitu pendekatan statistik (Hui, Si, & Callan, 2007) dan *machine learning* (Zhang, Yu, & Meng, 2007).

Pada penelitian ini, penulis ingin meneliti bagaimana memperoleh dokumen blog yang mengandung opini mengenai suatu topik sesuai dengan kebutuhan pengguna dengan menggunakan pendekatan statistik. Penelitian dilakukan pada dokumen blog berbahasa Indonesia.

1.3 Tujuan

Tujuan utama dari tugas akhir ini adalah mempelajari teknik pencarian opini pada dokumen blog berbahasa Indonesia dengan menggunakan pendekatan statistik. Secara lebih detail, tujuan dari penelitian ini antara lain adalah:

1. Meneliti teknik pemrosesan kueri yang sesuai untuk perolehan dokumen blog berbahasa Indonesia.
2. Meneliti teknik perbaikan kueri yang tepat untuk meningkatkan akurasi perolehan dokumen blog berbahasa Indonesia.
3. Meneliti pencarian opini mengenai suatu topik yang sesuai dengan kebutuhan pengguna pada blog berbahasa Indonesia.

1.4 Ruang Lingkup

Ruang lingkup dari pengerjaan tugas akhir ini adalah sebagai berikut.

1. Penelitian yang dilakukan adalah perolehan opini pada dokumen blog.

2. Penelitian menggunakan dokumen blog berbahasa Indonesia berupa sebuah tulisan blog beserta komentar terkait (*permalinks*) dan blog yang dapat berisi satu atau lebih tulisan (*feeds*).
3. Penelitian menggunakan pendekatan statistik yang difokuskan pada teknik pemrosesan kueri yang tepat, penggunaan teknik perbaikan kueri untuk meningkatkan akurasi pencarian dokumen, dan teknik pengurutan kembali dokumen untuk memperoleh dokumen yang mengandung opini.

1.5 Metodologi Penelitian

Berikut ini adalah metodologi penelitian yang dilakukan selama pelaksanaan tugas akhir.

1. Studi literatur
Tahap ini dilakukan untuk mempelajari perolehan informasi blog, mengetahui penelitian yang sudah dilakukan serta metode-metode yang sudah pernah digunakan.
2. Persiapan dokumen blog dan aplikasi
Tahap ini dilakukan untuk mengumpulkan dokumen blog berbahasa Indonesia yang akan digunakan dalam penelitian. Selain itu dalam tahap ini juga dilakukan pembelajaran terhadap aplikasi-aplikasi yang akan digunakan di dalam penelitian.
3. Eksperimen
Pada tahap ini dilakukan eksperimen dengan menggunakan korpus dokumen yang telah disiapkan dan aplikasi yang telah dipelajari. Eksperimen dilakukan dengan melakukan pengindeksan dokumen, menerapkan berbagai teknik pembobotan kueri dan teknik perbaikan kueri. Selain itu, juga dilakukan pengurutan kembali dokumen yang relevan untuk memperoleh dokumen yang berisi opini.
4. Evaluasi dan Analisis Hasil
Pada tahap ini dilakukan evaluasi terhadap hasil eksperimen. Evaluasi dilakukan pada setiap eksperimen yang dilakukan untuk menentukan teknik perolehan opini pada dokumen blog yang paling tepat.

1.6 Sistematika Penulisan

Sistematika penulisan laporan mengikuti tahapan-tahapan yang dilakukan untuk menyelesaikan tugas akhir, sebagai berikut:

1. BAB I PENDAHULUAN

Pada bab ini dijelaskan latar belakang dari penelitian yang memberikan gambaran awal mengenai perolehan informasi blog dan perannya dalam perkembangan teknologi informasi, permasalahan, tujuan dan ruang lingkup pengerjaan tugas akhir yang memberikan penjelasan mengenai hasil yang ingin diketahui serta batasan dalam pengerjaan dan tahapan dalam metodologi penelitian dan sistematika penulisan laporan.

2. BAB II LANDASAN TEORI

Pada bab ini dijelaskan landasan teori dari topik perolehan dokumen blog dan metode yang digunakan dalam tugas akhir untuk melakukan pencarian dokumen blog berbahasa Indonesia. Pembahasan ini dimulai dengan penjelasan mengenai perolehan informasi dan evaluasi yang dilakukan pada perolehan informasi. Selain itu, akan dibahas juga mengenai penelitian serupa yang pernah dilakukan sebelumnya serta fokus dari penelitian pada tugas akhir ini.

3. BAB III EKSPERIMEN

Pada bab ini dijelaskan berbagai macam data dan aplikasi yang digunakan serta skenario eksperimen yang dilakukan. Eksperimen dilakukan dengan melakukan pengindeksan dan perolehan dokumen pada penelitian. Hasil dari tahap perolehan kemudian akan dievaluasi untuk menemukan metode pencarian yang tepat. Metode yang memberikan hasil paling akurat kemudian akan diproses hasilnya. Dokumen-dokumen yang relevan akan diproses kembali untuk menghasilkan urutan dokumen berdasarkan relevansi dan opini yang terkandung di dalam dokumen.

4. BAB IV HASIL EKSPERIMEN DAN ANALISIS

Pada bab ini dijelaskan hasil eksperimen dan analisis terhadap hasil eksperimen. Selanjutnya hasil evaluasi akan dianalisis untuk mengetahui perbandingan berbagai teknik perolehan dokumen blog.

5. BAB V PENUTUP

Bab ini merupakan penutup dari laporan tugas akhir yang berisi kesimpulan dan kendala dari percobaan yang dilakukan. Selain itu, pada subbab terakhir juga diberikan saran-saran untuk pengembangan lebih lanjut dalam penelitian pencarian dokumen blog berbahasa Indonesia.

