

1. Pendahuluan

1.1 Latar Belakang Masalah

Dalam beberapa tahun terakhir banyak pengguna Internet yang menggunakan media *website* untuk berbagi informasi, berbagi pengalaman, berdiskusi, dan mengemukakan pendapat mereka sehingga banyak opini-opini yang terdapat pada media *website*, baik itu berupa forum, blog, situs penjualan, maupun halaman web lainnya. Opini-opini yang ada pada media *website* tersebut dapat digunakan sebagai referensi baik untuk memperoleh rekomendasi produk ataupun mengetahui pandangan masyarakat dalam suatu topik tertentu. Tentu saja untuk memperoleh opini-opini tersebut akan membutuhkan banyak waktu jika pengguna harus bersusah payah mencarinya di setiap halaman web.

Opinion retrieval adalah proses pengambilan dokumen yang memerlukan dokumen-dokumen untuk diambil dan diurutkan berdasarkan opininya mengenai topik tertentu [15]. Dengan menggunakan sistem *opinion retrieval*, pengguna dapat mencari dan menemukan dokumen, bisa berupa halaman web, yang berisi opini-opini berdasarkan kata kunci yang diinginkan. Berbagai penelitian telah dilakukan untuk menghasilkan sistem *opinion retrieval* yang efektif, diantaranya adalah “*Opinion Search and Retrieval from WWW*” yang dipublikasikan oleh IOSR Journal of Computer Engineering pada tahun 2012.

Terdapat suatu kendala dalam proses pencarian opini, khususnya pada dokumen berbahasa Indonesia dikarenakan adanya perbedaan penulisan antara Bahasa Indonesia dan Inggris. Tidak semua opini yang ada pada suatu dokumen berkaitan dengan kata kunci tertentu. Hal ini mengakibatkan kemungkinan terjadinya perolehan dokumen yang relevan, tetapi opini-opini di dalamnya tidak berkaitan dengan kata kunci.

Pada penelitian ini, relevansi topik dan opini pada dokumen berbahasa Indonesia ditentukan dan dikombinasikan menggunakan *generation model*. Pemodelan ini dipilih karena dapat menentukan opini-opini, termasuk opini berbahasa Indonesia, yang berkaitan dengan kata kunci, sehingga hanya opini-

opini yang relevan saja yang diambil untuk diakumulasikan menjadi skor opini dari suatu dokumen. Skor relevansi topik dan opini pada pemodelan ini disatukan menggunakan kombinasi kuadrat. Hal ini mengakibatkan skor opini pada dokumen memiliki pengaruh yang signifikan terhadap skor akhir dari dokumen tersebut.

1.2 Perumusan Masalah

Adapun permasalahan yang dibahas dalam Tugas Akhir ini adalah sebagai berikut.

1. Bagaimana membangun sistem *opinion retrieval* yang dapat menentukan keterkaitan opini-opini yang ada pada dokumen dengan kata kunci pencarian?
2. Bagaimana mengimplementasikan sistem yang dibangun untuk digunakan pada koleksi dokumen berbahasa Indonesia?

1.3 Batasan Masalah

Adapun batasan masalah dalam pengerjaan Tugas Akhir ini adalah sebagai berikut.

1. Koleksi dokumen yang diuji adalah dokumen berbahasa Indonesia yang sesuai dengan EYD.
2. Koleksi dokumen berupa teks yang berasal dari halaman web yang mengandung opini.
3. Sistem menggunakan kamus sentimen untuk menentukan opini pada dokumen. Dikarenakan kurangnya dataset kamus sentimen berbahasa Indonesia, kamus yang menjadi acuan adalah kumpulan kata sifat yang berasal dari Wordnet Bahasa.
4. Pengujian dilakukan secara *offline*, tanpa terkoneksi dengan Internet.

1.4 Tujuan

Tujuan dari pengerjaan Tugas Akhir ini adalah sebagai berikut.

1. Mengimplementasikan *generation model* pada *opinion retrieval* untuk digunakan pada koleksi dokumen berbahasa Indonesia.
2. Mengevaluasi tingkat akurasi dari sistem yang dibangun.

1.5 Metodologi Penyelesaian Masalah

Penyelesaian masalah dilakukan dalam beberapa tahap, secara garis besar adalah sebagai berikut.

1. Studi Literatur

Mempelajari literature-literatur yang berkaitan dengan metode yang digunakan, seperti permasalahan, konsep, dan teknik-teknik yang ada pada *opinion retrieval*, khususnya *generation model*.

2. Pengumpulan Data

Pengumpulan data berupa koleksi dokumen dan kamus sentimen dilakukan secara manual. Koleksi dokumen berasal dari tulisan-tulisan yang ada di <http://students.itelkom.ac.id/web/forum.php>. Sedangkan kamus sentimen, berupa kumpulan kata sifat Bahasa Indonesia, berasal dari Wordnet Bahasa yang dapat diunduh di <http://wn-msa.sourceforge.net>.

3. Pre-processing Dokumen

Tahapan yang dilakukan pada *pre-processing* dokumen adalah *cleaning*, *case folding*, tokenisasi, *stemming*, dan *indexing*.

4. Implementasi

Pada tahap ini dilakukan implementasi *generation model* hingga didapatkan hasil perolehan dokumen yang terurut berdasarkan skor akhirnya.

5. Analisis

Dilakukan analisa terhadap perolehan dan pengurutan dokumen yang dihasilkan serta mengukur tingkat akurasinya.

6. Pengambilan Kesimpulan dan Pembuatan Laporan

Pada tahap ini dilakukan pengambilan kesimpulan terhadap analisa yang telah dilakukan dan pembuatan laporan tugas akhir untuk mendokumentasikan tahap-tahap kegiatan.