

GENERATION MODEL PADA OPINION RETRIEVAL UNTUK DOKUMEN BERBAHASA INDONESIA

Ahmad Rodya Amali¹, Ema Rachmawati², Arie Ardiyanti Suryani³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Opinion retrieval merupakan suatu proses untuk memperoleh dokumen berdasarkan kata kunci pencarian yang diinputkan oleh pengguna. Opinion retrieval berbeda dengan information retrieval yang hanya memperhatikan relevansi dokumen. Pada opinion retrieval, selain dokumen yang diperoleh harus relevan, opini yang terdapat pada dokumen tersebut juga harus diperhatikan. Untuk itu diperlukan suatu metode yang benar-benar efektif dalam menentukan perolehan dokumen, agar dokumen tersebut tidak sebatas mengandung opini, tetapi opini yang berkaitan dengan kata kunci pencarian.

Pada Tugas Akhir ini, sistem opinion retrieval diimplementasikan pada koleksi dokumen berbahasa Indonesia dengan menggunakan generation model, yaitu suatu pemodelan pada opinion retrieval yang menyatukan skor relevansi dokumen dengan skor opini pada dokumen menggunakan kombinasi kuadrat. Pemecahan masalah dilakukan secara bertahap, yaitu dengan pengumpulan data berupa koleksi dokumen, penentuan daftar kata sentimen, document pre-processing, implementasi, kemudian mengukur akurasi dari hasil yang didapat.

Berdasarkan hasil pengujian menggunakan metode Mean Opinion Score (MOS), sistem opinion retrieval yang dibangun menggunakan generation model dapat menjamin hasil pencarian opini pada dokumen berbahasa Indonesia dengan akurasi sebesar 91,2%.

Kata Kunci : opinion retrieval, information retrieval, generation model

Abstract

Opinion retrieval is a process for obtaining documents based on query entered by user. Opinion retrieval is different from information retrieval which only pays attention to document relevance. In opinion retrieval, beside of retrieved documents should be relevant, opinions which contained in document should also be considered. So we need a method which is truly effective in determining retrieved documents, so the documents aren't only contained opinion, but opinion which is relevant to query.

In this final project, opinion retrieval system is implemented on Indonesian document collection using generation model, which is a opinion retrieval model uniting document relevance score with document opinion score using quadratic combination. Problem solving is done in few stages, there are collecting data of document collection, determining list of sentiment words, document pre-processing, implementation, and then measure accuracy of the result.

Based on test result using Mean Opinion Score (MOS) method, opinion retrieval system which built using generation model can guarantee the results of opinion retrieval on Indonesian document with an accuracy of 91.2%.

Keywords : opinion retrieval, information retrieval, generation model

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.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan beberapa poin, yaitu sebagai berikut.

1. Sistem *opinion retrieval* menggunakan *generation model* dapat diimplementasikan pada dokumen berbahasa Indonesia.
2. Parameter λ berfungsi sebagai penentu pengaruh skor opini terhadap perolehan dokumen. Semakin besar nilai λ , semakin kecil pengaruh skor opini terhadap perolehan dokumen, begitu juga sebaliknya.
3. *Window W* berfungsi sebagai parameter yang menentukan keterkaitan *query* terhadap kata sentimen. Semakin besar *window W*, semakin banyak jumlah kemunculan kata sentimen, namun penambahan skor opini untuk setiap kemunculannya menjadi kecil.
4. Berdasarkan keseluruhan aspek yang dinilai pada pengujian MOS, perolehan dan perangkingan dokumen yang dihasilkan adalah sebesar 91,2%.

5.2 Saran

Berdasarkan kesimpulan penelitian, penulis merekomendasikan agar pengujian sebaiknya dilakukan menggunakan dataset yang berukuran lebih besar, baik untuk koleksi dokumen maupun kamus sentimen Bahasa Indonesia, agar hasil perolehan dan perangkingan dokumen menjadi lebih bervariasi.

6. Daftar Pustaka

- [1] Eguchi, K., Lavrenko, V. 2006. *Sentiment Retrieval Using Generative Models*. National Institute of Informatics at Empirical Methods on Natural Language Processing (EMNLP).
- [2] Esuli, A., Sebastiani, F. 2005. *Determining the Semantic Orientation of Terms through Gloss Classification*. The Institute of Information Science and Technologies at CIKM.
- [3] He, B., Macdonald, C. 2008. *An Effective Statistical Approach to Blog Post Opinion Retrieval*. Glasgow: University of Glasgow.
- [4] Huang, X., Croft, W. 2009. *A Unified Relevance Model for Opinion Retrieval*. Shanghai: Fudan University.
- [5] Lafferty, J., Zhai, C. 2002. *Probabilistic Relevance Models Based on Document and Query Generation*. Pittsburgh: Carnegie Mellon University.
- [6] Liao, X., Cao, D., Tan, S., Liu, Y., Ding, G., Cheng, X. 2006. *Combining Language Model with Sentiment Analysis for Opinion Retrieval of Blog-Post*. Beijing: Chinese Academy of Sciences.
- [7] Manning, Christopher D. 2008. *Introduction to Information Retrieval*. New York: Cambridge University Press.
- [8] Mishne, Gilad. 2006. *Multiple Ranking Strategies for Opinion Retrieval in Blogs*. University of Amsterdam at the 2006 TREC Blog Track.
- [9] Padmapriya, A., Maheswaran, S. 2012. *Opinion Search and Retrieval from WWW*. Karaikudi: Alagappa University.
- [10] Pang, B., Lee, L. 2008. *Opinion Mining and Sentiment Analysis*. New York: Cornell University.
- [11] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M. 2011. *Lexicon-Based Methods for Sentiment Analysis*. Burnaby: Simon Fraser University.
- [12] Vania, C., Adriani, M. 2009. *Opinion Retrieval on Indonesian Blogs*. Depok: Universitas Indonesia.
- [13] Zhai, Chengxiang. 2007. *A Brief Review of Information Retrieval Models*. Urbana-Champaign: UIUC.

- [14] Zhang, M., Ye, X. 2008. *A Generation Model to Unify Topic Relevance and Lexicon-based Sentiment for Opinion Retrieval*. Beijing: Tsinghua University.
- [15] Zhang, W., Yu, C., Meng, W. 2007. *Opinion Retrieval from Blogs*. Chicago: University of Illinois.

