

OPINION RETRIEVAL DALAM BAHASA INDONESIA DENGAN UNIFIED GRAPH MODEL

Ade Elisa Damanik¹, Warih Maharani², Ema Rachmawati³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Opini merupakan “expressed statement” yang bisa diucapkan dengan kata-kata, juga bisa dinyatakan dengan isyarat atau dengan cara-cara lain yang mengandung arti dan segera dapat dipahami maksudnya. Opini-opini tersebut dapat dijadikan masukan yang sangat berharga bagi sebuah perusahaan untuk menilai dan meningkatkan pelayanan serta kualitas produknya. Tugas akhir ini bertujuan untuk me-retrieve opini berbahasa Indonesia atau yang dikenal dengan sebutan Opinion Retrieval. Opinion retrieval adalah proses pe-retrieve-an opini pada dokumen, dan dokumen yang diperoleh diurutkan berdasarkan opini mengenai sebuah topik query yang terdapat pada dokumen tersebut. Tugas akhir ini bertujuan untuk menemukan opini yang relevan dengan menggunakan unified graph model dan algoritma HITS untuk menentukan ranking dokumen dari hasil pencarian. Unified graph-model dengan algoritma HITS dipilih karena dapat menggambarkan secara detail hubungan antara word-pair dalam beberapa dokumen, me-retrieve dan menghasilkan opini yang relevan serta mengurutkan dokumen yang relevan tersebut dengan efektif. Algoritma HITS menggunakan dua nilai yaitu authority dan hub. Authority adalah halaman web atau dokumen yang memberikan informasi yang user cari, dan hub adalah halaman web yang memberikan informasi mengenai halaman yang mengandung informasi tersebut. Pengukuran akurasi dari hasil penelitian ini diukur menggunakan metode MOS (Mean Opinion Score).

Kata Kunci : Opinion retrieval, Unified Graph model, HITS

Abstract

Opinion is a “expressed statetement” which can be expressed either by words or by motions and could be understood in short time. In past few years, there are plenty of people expressed their opinion in social media, just like web, twitter, or facebook. Those opinions could become an useful input for a company to measure and increase their service as well as their product quality. This final project aims to retrieve opinion in Indonesian language, or better known as Opinion Retrieval. Opinion Retrieval is a process to retrieve an opinion from one document, and that retrieved document is sorted by opinion about a query topic which contained in that document. The goal of this final project is to discover relevant opinion using Unified Graph Model and HITS Algorithm to determine document ranking form result of related search query. HITS Algorithm is chosen because it can depict relationship between word pair (topic-term and sentiment-word) in detail, retrieve and yield relevant opinion, and effectively sorted those relevant document. HITS Algorithm use two values: authority and hub. Authority is a webpage or document which provide the exact information that user needs. And hub is webpage which offer information about webpages which contain that information. Measurement of accuracy from this research is measured using MOS (Mean Opinion Score) method.

Keywords : Opinion retrieval, Unified Graph model, HITS

1. Pendahuluan

1.1 Latar Belakang Masalah

Menurut William Albig dalam buku *Public Relations*, opini merupakan “*expressed statement*” yang bisa diucapkan dengan kata-kata, juga bisa dinyatakan dengan isyarat atau dengan cara-cara lain yang mengandung arti dan segera dapat dipahami maksudnya. Opini dapat memberikan dampak yang besar bagi masyarakat, seperti dapat mempengaruhi penjualan produk, perubahan kebijakan pemerintah, serta bahkan mempengaruhi suara rakyat dalam pemilu [3]. Dalam beberapa tahun terakhir, banyak orang yang mengekspresikan opini mereka di web, tentang berbagai macam topik seperti produk review, analisis ekonomi, masalah politik dan lain sebagainya. Opini-opini tersebut tidak hanya dapat membantu independen untuk membuat keputusan, tetapi juga memperoleh masukan yang sangat berharga (Pang et al., 2008).

Opinion retrieval adalah proses *pe-retrieve-an* opini pada dokumen, dan dokumen yang diperoleh diurutkan berdasarkan opini mengenai sebuah topik *query* yang terdapat pada dokumen tersebut. Berbagai penelitian telah dilakukan untuk menghasilkan suatu mesin pencari opini yang efektif, diantaranya adalah “*Opinion Retrieval from Blogs*” oleh University of Illinois at Chicago dan Binghamton University.

Tugas akhir ini berfokus pada *opinion retrieval* berbahasa Indonesia. Penelitian *opinion retrieval* berbahasa Indonesia diperlukan mengingat adanya perbedaan penulisan antara bahasa Indonesia dan Inggris. Penelitian ini bertujuan untuk menemukan sekumpulan opini yang tidak hanya mengandung *query keyword* atau target tetapi juga opini yang relevan. Pada sebuah dokumen bisa terdapat lebih dari satu opini, namun tidak semua opini yang ada di dokumen relevan dan sesuai dengan target [1]. Oleh karena itu dibutuhkan suatu metode untuk menentukan opini mana saja yang relevan sesuai dengan target, sehingga dapat diperoleh informasi dari hubungan antara opini dan target tersebut serta hubungan antar opini dalam suatu dokumen. Kedua jenis informasi di atas

dikombinasikan dan dideskripsikan dengan *unified graph-model* yang menerapkan algoritma *HITS*. Unified graph-model dengan algoritma *HITS* dipilih karena dapat menggambarkan secara detail hubungan antara *word pair* (*topic-term* dan *sentiment-word*) dalam beberapa dokumen, meretrieve dan menghasilkan opini yang relevan serta mengurutkan dokumen yang relevan tersebut dengan efektif [1].

1.2 Perumusan Masalah

Berdasarkan pada latar belakang di atas, permasalahan yang akan diuraikan dan diteliti adalah:

1. bagaimana cara menerapkan *Unified Graph Model* dengan algoritma *HITS* pada *Opinion retrieval*;
2. bagaimana mengetahui akurasi *Unified Graph Model* dengan algoritma *HITS* menggunakan Mean Opinion Score (MOS).

1.3 Batasan Masalah

Adapun batasan masalah untuk tugas akhir ini adalah:

1. dokumen yang akan dianalisis adalah blog berbahasa Indonesia yang memiliki ejaan yang benar (EYD);
2. data akan dianalisis secara offline learning, yaitu sistem tidak terhubung ke internet;
3. penentuan topik di setiap kalimat dilakukan secara manual (belum terotomatisasi).

1.4 Tujuan

Tujuan yang ingin dicapai dalam pengerjaan Tugas Akhir ini adalah sebagai berikut:

1. menerapkan *Unified Graph Model* dengan algoritma *HITS* pada *Opinion retrieval*;
2. mengevaluasi hasil akurasi *Unified Graph Model* dan Algoritma *HITS*.

1.5 Metodologi penyelesaian masalah

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

1. Studi Literatur

Mempelajari literature-literatur yang relevan dengan permasalahan *information retrieval*, konsep dan teknik-teknik *opinion retrieval* khususnya *unified graph model*.

2. Pengumpulan Data

Pengumpulan data dilakukan secara manual yaitu berupa komentar yang berasal dari <http://students.ittelkom.ac.id/web/forumphp>.

3. Pre-processing

Tahapan yang dilakukan dari dokumen pre-processing adalah *case folding*, *tokenization*, *stopword removal*, dan *stemming*.

4. Analisis

Pada tahap ini dilakukan analisa *unified graph model* dengan algoritma *HITS* terhadap akurasi perolehan opini dan perengkingan dokumen dari hasil pencarian.

5. 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.

Telkom
University

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan percobaan yang dilakukan di atas, dapat disimpulkan beberapa poin sebagai berikut:

1. *Unified Graph Model* dapat diimplementasikan untuk meretrieve kalimat opini berbahasa Indonesia.
2. Algoritma *HITS* dapat diterapkan dalam perengkingan dokumen berbahasa Indonesia.
3. Panjang atau banyaknya kata di dalam kalimat tidak mempengaruhi tingkat kerelevanan dari kalimat opini yang diretrieve.
4. *Unified Graph Model* dapat menangani kalimat opini dengan topik yang implisit.
5. Rata-rata skor MOS sudah cukup baik yaitu 4.37 yang dinilai dari tiga aspek yaitu keakuratan topik, kerelevanan opini, dan keakuratan perengkingan.

5.2 Saran

1. Proses pengambilan dataset dapat dikembangkan secara online, sehingga proses pengambilan data dapat dilakukan secara otomatis.
2. Dataset yang digunakan sebaiknya menggunakan dataset berbahasa Indonesia dimana ada *query* yang sudah ditentukan beserta *relevance judgment*-nya, sehingga dapat dihitung akurasi secara otomatis dan tidak perlu melakukan *survey* ke responden yang membutuhkan banyak waktu.
3. Proses penentuan topik pada setiap kalimat dapat dilakukan secara otomatis.

6. Daftar Pustaka

- [1] Binyang Li, Lanjun Zhou, Shi Feng, Kam-Fai Wong. *A Unified Graph Model for Sentence-based Opinion Retrieval*, The Chinese University of Hong Kong.
- [2] Budianto, Arifin Zainal Agus, Lili Suhadi. Perancangan dan Pembuatan Perangkat Lunak Penelusur Web (Web Crawler) Menggunakan Algoritma Pagerank. Teknik Informatika, Institut Teknologi Sepuluh November Surabaya, 2003.
- [3] Clement Yu, Weiyi Meng. 2011. In *Collaborative Research Handling Negations and Temporal Aspects for Opinion Retrieval*.
- [4] Erwin Ashari Hariyono Muhammad, Wahyudi. *Customer Information Gathering menggunakan Metode Temu Kembali Informasi dengan Model Ruang Vektor*. Teknik Informatika Universitas Islam Indonesia, 2005.
- [5] Herwansyah, Adhit. (2009). *Aplikasi Pengkategorian Dokumen dan Pengukuran Tingkat Similaritas Dokumen menggunakan Kata Kunci pada Dokumen Penulisan Ilmiah Universitas Gunadarma*. Jakarta: Universitas Gunadarma.
- [6] Jahna Otterbacher, Gunes Erkan, and Dragomir R. Radev. 2005. Using random walks for question-focused sentence retrieval. In *EMNLP '05, Proceedings of 2005 Conference on Empirical Methods in Natural Language Processing*.
- [7] Jon M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5): 604-632.
- [8] Jon M. Kleinberg. "Lecture #4: HITS Algorithm - Hubs and Authorities on the Internet." <http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture4/lecture4.html> (diakses tanggal 30 Desember 2012)
- [9] Khairina Indah. *Studi World Web sebagai Sebuah Graf Berarah*. Bandung: Institut Teknologi Bandung.
- [10] Manning, Christopher D. 2008. *Introduction to Information Retrieval*. Newyork ; Cambridge Univ. Press.

- [11] Pratiwi, A. Y. (2010). *Implementasi Model Ruang Vektor Pada Sistem Temu Balik Informasi Dokumen Berbahasa Indonesia*. Bandung: Universitas Pendidikan Indonesia.
- [12] Signorini Alessio. A survey Ranking Algorithms. <http://www.cs.uiowa.edu/~cremer/courses/cs2/SignoriniRankingAlgSurvey.pdf>.
- [13] Susanto, S. (2006). *Pengklasifikasian Artikel Berita Berbahasa Indonesia Secara Otomatis Menggunakan Naive Bayes Classifier*. Depok: Universitas Indonesia.
- [14] Wei Zhang, Clement Yu, Weiyi Meng. *Opinion Retrieval from Blogs*, Department of Computer Science Binghamton University.
- [15] Wibowo Adi, Handojo Andreas, Taliwang Minardi. *Recommender System di Perpustakaan Universitas Kristen Petra menggunakan Rocchio Relevance Feedback dan Cosine Similarity*. Surabaya: Universitas Kristen Petra.
- [16] Wicaksono Alfian & Purwarianti Ayu. 2010, *HMM Based Part-of-Speech Tagger for Bahasa Indonesia*, Institut teknologi Bandung.