

ANAPHORA RESOLUTION MENGGUNAKAN METODE MITKOV'S ROBUST KNOWLEDGE-POOR ALGORITHM (STUDI KASUS CUSTOMER REVIEW)

Andi Nur Fitri Afifah¹, Z.k. Abdurahman Baizal², Sri Suryani³

¹Ilmu Komputasi, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Anaphora Resolution adalah proses menentukan sebuah objek atau noun pada kalimat sebelumnya yang menggantikan sebuah kata ganti. Anaphora resolution memiliki fungsi untuk menentukan noun di kalimat sebelumnya yang menggantikan kata ganti "it". Kata "it" seperti ini sering ditemukan dalam kalimat-kalimat sederhana atau sebuah teks dokumen, contohnya pada review konsumen di Amazon.com.

Dalam tugas akhir ini, penulis mencoba menerapkan Mitkov's Robust Knowledge-Poor Algorithm pada data review yang ada di Amazon.com, baik review panjang maupun pendek. Data diambil dari studi kasus review dengan harapan dengan menggunakan text real, sistem yang dibangun bisa membantu penelitian dalam bidang opinion mining untuk lebih lanjut. Hasil dari percobaan dalam tugas akhir ini menunjukkan bahwa sistem dapat mendeteksi noun phrase pengganti kata "it" untuk review pendek dengan success rate terbesar 80% dan review panjang dengan success rate terbesar 76.5%.

Kata Kunci : Anaphora Resolution, Mitkov's Robust Knowledge-Poor Algorithm, Review

Abstract

Anaphora resolution is a process to determine an object or a noun in the previous sentence that replaces a pronoun. Anaphora resolution has a function to determine the noun in the previous sentence that replaces the word "it". The word "it" as is often found in simple sentence or a text document, for example on consumer review at Amazon.com.

In this thesis, the authors tries to apply the Mitkov's Robust Knowledge-Poor Algorithm on a review data on Amazon.com, both long and short review. The data is taken from a customer review in hopes of using real text, the system could help further research in the field of opinion mining. The result of the experiments shows that the system can detect the noun phrase that substitute the word "it" with the highest success rate of 80% for the short review and the highest success rate of 76.5% for the long review.

Keywords : Anaphora Resolution, Mitkov's Robust Knowledge-Poor Algorithm, Review

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Anaphora Resolution adalah sebuah cara untuk menunjukkan kata pengganti ke sebuah objek yang dimaksud di kalimat sebelumnya. *Anaphora* atau biasa disebut *anaphor* merupakan kata ganti seperti “*it*” yang mengacu pada suatu objek (*antecedent*) di kalimat sebelumnya [8]. Kasus *anaphora resolution* banyak ditemukan dalam kehidupan sehari-hari, salah satunya adalah *review* yang ada di Amazon.com seperti “*I bought this phone for my wife. She loves it. From my perspective, it is a good Android-based-phone that has lived up to and exceeded expectations. One motivation to get this phone was that it would work with an existing SIM card and handles data using WiFi only.*” Pada *review* tersebut, terdapat kata “*it*” yang mengacu pada suatu benda tertentu. Untuk mengetahui kata ganti benda dari kata “*it*” tersebut, diperlukan *anaphora resolution*.

Dalam tugas akhir ini, penulis mencoba menerapkan metode *Mitkov's Approach Knowledge-Poor Algorithm* [10] untuk mencari *noun phrase* yang menggantikan kata “*it*” pada *review* yang ada di Amazon, baik *review* panjang maupun *review* pendek. *Mitkov's Approach Knowledge-Poor Algorithm* merupakan sebuah metode untuk mencari *noun phrase* pengganti “*it*” dengan cara memberikan *point-point* untuk setiap kandidat *noun phrase* yang terpilih yang sesuai dengan metode yang ada dalam *Mitkov's Approach Knowledge-Poor Algorithm*.

Diharapkan sistem yang dibuat dalam tugas akhir ini dapat membantu penelitian dibidang *opinion mining* salah satunya pada *sentence level*, karena sampai saat ini banyak penelitian dibidang *opinion mining* yang tidak melibatkan *anaphora resolution* [4, 6, 7, 14]. Dalam *sentence level* permasalahan yang umum ditemukan adalah dokumen yang diproses per-kalimat, bukan sebagai kesatuan dokumen sehingga ketika menemukan *sentence level* seperti “*I like reading a novel. It's better than reading a comic.*”, kata “*it*” dalam kalimat kedua tidak secara eksplisit menyebutkan satu objek yaitu novel, objek tersebut diganti

menjadi kata ganti “it” sehingga kata “it” pada kalimat kedua diabaikan oleh sistem. *Anaphora resolution* dapat menjadi *tools* yang membantu mengganti kata “it” tersebut dengan objek yang dimaksud, sehingga tidak ada lagi objek yang diabaikan oleh sistem.

1.2 PERUMUSAN MASALAH

Rumusan masalah dalam tugas akhir ini adalah sebagai berikut:

1. Bagaimana penerapan *Anaphora Resolution* pada *review* konsumen dengan menggunakan *Mitkov's Approach knowledge-poor algorithm*.
2. Bagaimana performansi sistem yang dibangun dari sisi akurasi melalui *success rate*-nya.

1.3 TUJUAN PENELITIAN

Tujuan penulisan tugas akhir ini adalah sebagai berikut:

1. Menganalisis dan mengimplementasikan *Anaphora Resolution* dengan menggunakan *Mitkov's Approach Knowledge-poor Algorithm* pada studi kasus *review* konsumen.
2. Menganalisis performansi sistem yang telah dibangun dari sisi akurasi melalui *success rate*-nya.

1.4 BATASAN MASALAH

Batasan masalah yang digunakan dalam penulisan tugas akhir ini adalah sebagai berikut:

1. Data yang digunakan merupakan opini konsumen dari *social media*.
2. Data yang digunakan berbahasa Inggris.
3. Kalimat yang bisa menjadi data adalah kalimat yang benar, jelas, dan lengkap berdasarkan *grammar* manual yang dilakukan oleh manusia.
4. Kalimat yang digunakan memiliki satu kata “it” dan satu *noun phrase*.
5. Satu data atau *review* memiliki paling tidak satu sampai delapan kalimat.
6. Kata “it” yang dideteksi oleh sistem hanya kata “it” pertama dalam sebuah dokumen.

1.5 METODOLOGI PENELITIAN

Dalam penyelesaian tugas akhir ini, ada beberapa metodologi penelitian yang digunakan, diantaranya:

1. Studi Literature

Mencari, mengumpulkan, dan mempelajari informasi mengenai *Anaphora Resolution* dari berbagai sumber seperti buku, jurnal, *paper*, dan internet sebagai landasan teori dalam pengerjaan dan penyusunan tugas akhir ini.

2. Pengumpulan Data

Mencari dan mengumpulkan data dari opini konsumen di <http://www.amazon.com/>. Data berikutnya akan digunakan untuk pengujian dalam tahap implementasi.

3. Preprocessing Data

Mengolah data mentah menggunakan *casefolding*, *cleaning*, *POS Tagging*, dan *tokenization*.

4. Perancangan dan Implementasi Sistem

Menentukan model dan perancangan sistem yang akan digunakan untuk perhitungan data dan analisis serta diimplementasikan kedalam sebuah program.

5. Analisis Implementasi Sistem

Menganalisis data yang telah diimplementasi pada sistem yang telah dibuat.

6. Pembuatan Laporan Tugas Akhir

Mendokumentasikan tugas akhir ini kedalam bentuk laporan tertulis yang memuat analisis dan penarikan kesimpulan berdasarkan dengan model dan perancangan yang telah dilakukan.

1.6 SISTEMATIKA PENULISAN

Penulisan laporan tugas akhir disusun berdasarkan sistematika sebagai berikut:

BAB I Pendahuluan

Bab ini menjelaskan tentang latar belakang masalah, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, dan sistematika penulisan dalam penyelesaian laporan tugas akhir ini.

BAB II Dasar Teori

Bab ini menjelaskan tentang dasar teori yang dapat mendukung penyelesaian tugas akhir ini. Dasar teori tersebut berhubungan dengan *Anaphora Resolution*.

BAB III Pemodelan dan Perancangan Sistem

Bab ini menjelaskan semua proses pembangunan sistem sesuai dengan masalah yang dibahas mulai dari pengumpulan data sampai implementasi sistem yang telah dibuat.

BAB IV Analisis Hasil Pengujian

Bab ini menjelaskan hasil analisis dari penerapan metode, pemodelan, dan perancangan sistem yang telah dibuat di bab sebelumnya.

BAB V Kesimpulan dan Saran

Bab ini berisi kesimpulan dari analisis yang telah dilakukan dan saran untuk pembaca agar dapat mengembangkan tugas akhir ini menjadi lebih baik.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pengujian pada bab sebelumnya, dapat disimpulkan:

1. *Mitkov's Robust Knowledge-Poor Algorithm* dapat diterapkan pada data *text* pendek atau data *text* panjang karena hasil *success rate* yang didapat tidak signifikan.
2. Proses *POS Tagging* dalam sistem cukup berpengaruh terhadap hasil *success rate* karena ada beberapa *noun* yang seharusnya terdeteksi sebagai *noun*, namun tidak terdeteksi saat proses *POS Tagging* sehingga kandidat *noun* menjadi kurang dan ada kemungkinan *noun* yang tidak terdeteksi tersebut merupakan *noun* pengganti “*it*” yang dicari.
3. Proses *number and gender agreement* juga berpengaruh terhadap hasil *success rate* sistem, karena *number and gender agreement* yang ada pada kalimat dibaca sebagai *noun*, sehingga kandidat *noun* sebenarnya akan semakin banyak dan membuat peluang *noun* sebenarnya untuk dipilih oleh sistem menjadi semakin kecil.

5.2 Saran

Berdasarkan percobaan yang dilakukan pada bab sebelumnya, sebaiknya perlu diperhatikan beberapa poin sebagai berikut:

1. Memodifikasi algoritma sehingga tidak hanya dapat mencari *noun* pengganti “*it*” tapi dapat mencari *noun* pengganti *this, his, one*, dan kata ganti lainnya.
2. Menggunakan *tools POS Tagger* yang lebih bagus, sehingga tidak ada lagi *noun* yang tidak terdeteksi untuk meningkatkan performansi sistem.

DAFTAR PUSTAKA

- [1] Baron, Dennis E. 1987. *Grammar and Gender*. Yale University Press.
- [2] Devi, S. L., Hendrickx, I. 2011. *Anaphora Processing and Applications - 8th Discourse Anaphora and Anaphor Resolution Colloquium*, DAARC 2011, Faro, Portugal.
- [3] Devore, J. L. 1991. *Probability and Statistic for Engineering in The Sciences*. Third Edition. Pacific Grove California.
- [4] Hu, Minqing., Liu, Bing. 2004. *Mining and Summarizing Customer Reviews*. Department of Coumputer Science. Chicago. To appear in KDD '04
- [5] IsmailMidi. 2012. *Part of Speech*. <http://ismailmidi.com/english-546404-part-of-speech.html>. Diakses pada 24 Desember 2012.
- [6] Jindal, N &Liu, Bing. 2006. *Identifying Comparative Sentences in Text Documents*. University of Illions at Chicago: Department of Computer Science. In proceeding of SIGIR '06, Seattle, Washington, USA
- [7] Jindal, N & Liu, Bing. 2006. *Mining Comparative Sentences and Relations*. University of Illions at Chicago: Department of Computer Science. In proceeding of American Association for Artificial Intelligence.
- [8] Lust, B. 1986. *Studies in the Acquisition of Anaphora: Defining the constraints*. University of California
- [9] M.C. Marnee, C.D. Manning. 2008. *The Stanford typed Dependencies Representation*, in: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation, CrossParser '08 Coling, pp. 1-8.
- [10] Mitkov, R. *Anaphora Resolution*, First published in Great Britain in 2002
- [11] Pyle, Dorian. 1999. *Data Preparation for Data Mining*. San Fransisco. Morgan Kaufmann Publisher, Inc.
- [12] Stanford Tagger. <http://nlp.stanford.edu/downloads/tagger.shtml>. Diakses pada 26 September 2013.
- [13] University of Memphis. 2010. *Anaphora Resolution*. <http://www.slideshare.net/kancho/anaphora-resolution>. Diakses pada 24 Desember 2012.

- [14] Xu, Kaiquan., Liao, Stephen Shaoyi. 2010. *Mining comparative opinions from customer reviews for Competitive Intelligence*. *Decision Support Systems*, 50 (2011): 743-754

