

BAB I Pendahuluan

Pada bab ini permasalahan dan latar belakang yang menjadi awal mula penelitian dijabarkan dalam subbab latar belakang dan rumusan masalah serta tujuan dari penelitian dijelaskan dalam subbab tujuan penelitian serta bab ini memuat tentang metodologi penelitian yang digunakan dalam penelitian.

1.1 Latar Belakang

Salah satu penelitian mengenai *Natural Language Processing* (NLP) yang relatif penting ialah penelitian mengenai kesamaan semantik, kesamaan semantik adalah studi mengenai kesamaan arti dari ekspresi linguistik [3]. Kesamaan semantik penting untuk diteliti karena dengan kesamaan semantik komputer dapat mengidentifikasi kalimat atau perintah mana saja yang memiliki maksud atau makna yang sama selain itu kesamaan semantik adalah penelitian yang penting karena dapat diaplikasikan di berbagai macam bidang *natural language processing*, *information retrieval* atau bidang edukasi seperti *question answering* [4]. Penelitian mendasar mengenai kesamaan semantik yaitu dengan mencari tahu kesamaan semantik antarkalimat pendek. Contohnya kalimat 1: *I wanna call my mom*, kalimat 2: *I want to contact my mother*. Manusia akan dengan mudah mengidentifikasi bahwa kedua kalimat tersebut memiliki makna yang sama tetapi tidak dengan komputer. Kalimat pendek dalam konteks penelitian ini adalah kalimat yang relatif pendek, hanya satu baris, hanya memiliki subjek, predikat, objek atau minimal hanya memiliki subjek dan predikat saja.

Menghitung atau mencari tahu kesamaan semantik antarkalimat pendek dibutuhkan metode yang sesuai dalam mengidentifikasinya. Banyak metode yang dapat digunakan dalam mengidentifikasi kesamaan semantik antarkalimat pendek. Masalah utama dalam mengidentifikasi kesamaan semantik ialah jika penggunaan kata-kata di dalam pasangan kalimat sangat beragam dibutuhkan metode yang tepat untuk mengidentifikasinya. Salah satu metode yang relatif mendasar dan mempunyai keunggulan dalam keringanan komputasi ialah dengan menggunakan *alignment* atau penyejajaran kata-kata dalam kalimat tersebut. *Alignment* merupakan salah satu metode yang sederhana, akurasi tinggi dan cepat sesuai dengan tujuan pembuatan sistemnya [5]. Selain sederhana, cepat dan akurasi tinggi, *alignment* memiliki banyak fitur yang bisa digunakan sesuai dengan kebutuhan sistem, pemilihan fitur yang tepat diperlukan untuk menghasilkan performansi yang tinggi pada setiap kasus termasuk dalam mengidentifikasi kalimat yang menggunakan kata-kata yang beragam. Untuk mengukur kesamaan semantik dalam bahasa yang sama *monolingual alignment* merupakan representasi metode yang relatif sesuai karena melakukan penyejajaran kata-kata dalam bahasa yang sama. Metode *monolingual alignment* adalah metode yang melakukan penyejajaran kata-kata dalam kalimat dengan bahasa yang sama. Kata-kata dalam kalimat akan diidentifikasi urutan dan posisinya sehingga dapat diukur kesamaannya dengan menggunakan fitur-fitur yang tepat.

1.2 Perumusan Masalah

Berdasarkan latar belakang bahwa interaksi manusia dengan komputer semakin sulit untuk dipisahkan dan didukung oleh adanya NLP sehingga menarik untuk diteliti maka rumusan masalah yang muncul adalah:

1. Bagaimana pengaruh fitur-fitur *monolingual alignment* dalam melakukan *alignment* pada pasangan kalimat bahasa Inggris?
2. Bagaimana performansi *monolingual alignment* dalam melakukan *alignment*?
3. Bagaimana hasil dari penggunaan metode *monolingual alignment* untuk mengidentifikasi kesamaan semantik dari pasangan kalimat bahasa Inggris?

1.3 Tujuan

Berdasarkan latar belakang dan rumusan masalah yang telah diuraikan diatas maka tujuan dari adanya penelitian ini adalah:

1. Menganalisis pengaruh penggunaan fitur-fitur *monolingual alignment* dalam melakukan *alignment* pada pasangan kalimat dalam bahasa Inggris.
2. Mengetahui performansi metode *monolingual alignment* dalam melakukan *alignment*.
3. Mengimplementasikan metode *monolingual alignment* untuk mengetahui kesamaan semantik dari pasangan kalimat bahasa Inggris.

1.4 Batasan Masalah

Untuk menyelesaikan penelitian ini ada beberapa hal yang menjadi batasan agar penelitian ini lebih terarah dan tidak terlalu menyimpang dari inti permasalahan maka batasan penelitian ini adalah:

1. Mencari kesamaan sepasang kalimat dalam bahasa Inggris dengan metode *monolingual alignment* dengan fitur *align identical word*, *align PPDB*, *align word sequences*, *align named entities*, *align content words (dependency)*, *align content words using surrounding words (text neighbor)*, *align stop words*, *align PPDB Extended*.
2. Bahasa yang digunakan sebagai kalimat yang akan dibandingkan adalah bahasa Inggris.
3. Dataset yang digunakan adalah data yang tersedia di SemEval ¹ dan dari Microsoft Research.
4. Keringanan dan kecepatan *alignment* diluar lingkup penelitian dan tidak dipertimbangkan dalam penelitian ini.

¹<http://alt.qcri.org/semeval2015/>

1.5 Metodologi Penelitian

Untuk menyelesaikan masalah dibutuhkan tahapan dalam pemecahan masalah tersebut. Metodologi yang dilakukan dalam menyelesaikan masalah adalah sebagai berikut:

1. Identifikasi Masalah

Pada tahap ini yaitu tahap mengidentifikasi masalah yang sedang menjadi tren untuk diselesaikan dengan cara membaca beberapa paper (studi literatur) yang dipublikasikan dalam tiga tahun terakhir.

2. Pengumpulan Data

Tahap pengumpulan data akan dilakukan setelah selesai mengidentifikasi masalah dan menentukan metode yang tepat berdasarkan hasil dari studi literature, dengan begitu maka penulis dapat menentukan jenis data yang tepat untuk digunakan dalam penelitian dan mulai melakukan pencarian data di SemEval tahun 2013, 2014, 2015 dan Microsoft Research² (MSR).

3. Perancangan Sistem

Berdasarkan data yang telah dikumpulkan sistem akan dirancang menyesuaikan dengan bentuk permasalahan dan jenis data sehingga dapat menghasilkan solusi terbaik.

4. Implementasi

Tahap yang dilakukan untuk membangun system dengan metode *monolingual alignment* agar dapat menguji data dan menghasilkan nilai korelasi yang diharapkan untuk menentukan kesamaan semantic dua buah kalimat.

5. Pengujian Sistem

Setelah selesai tahap implementasi maka dilakukan pengujian terhadap data dan hasilnya dianalisis.

6. Analisis dan Kesimpulan

Tahapan terakhir dalam penelitian ini yaitu menganalisis hasil dari sistem yang telah dibangun dan menarik kesimpulan dari hasil penelitian.

²http://ixa2.si.ehu.es/stswiki/index.php/Main_Page