

1. Pendahuluan

Latar Belakang

Kalimat dalam bahasa mempunyai beragam bentuk dan struktur kata didalamnya. Keberagaman tersebut membuat beberapa kalimat sebenarnya memiliki beberapa kata yang sebenarnya memiliki makna yang sama. Karena itu penelitian di bidang linguistik ini bisa dilakukan untuk mencari skor tingkat kesamaan makna antar teks. Konsep yang akan dipakai untuk perhitungan dan penentuan kesamaan antar teks adalah Interpretable Semantics Textual Similarity. Interpretable Semantics Textual Similarity merupakan landasan dari SemEval 2016 Task 2. Konsep tersebut dipakai untuk dasar melakukan perhitungan kesamaan makna antar teks yang dibandingkan dalam Tugas Akhir ini.

Sejauh pengamatan penulis, terdapat riset yang menggunakan konsep Interpretable Semantics Textual Similarity dengan chunking manual. Untuk itu penulis mencoba mengembangkan konsep tersebut dengan metode berbeda yaitu menggunakan IXA Pipelines dan chunking otomatis. IXA Pipelines dipilih karena sistemnya yang *ready to use* sehingga tidak diperlukan instalasi yang terlalu rumit, serta nilai performansi yang tinggi berdasarkan dari beberapa penelitian yang telah dilakukan menggunakan IXA Pipelines. Proses yang akan dilakukan adalah membandingkan chunk dengan proses chunking otomatis dan alignment. Setelah melakukan Proses tersebut, dilakukan perbandingan kembali dengan Gold Standart. Pembangunan sistem di dalamnya akan terdapat teks yang dibandingkan dengan teks lainnya sehingga akan memunculkan nilai keterkaitannya. Dengan penelitian ini akan didapatkan informasi tentang keterkaitan teks yang menjadi perbandingan serta tingkat akurasinya

Topik & Batasan

Berdasarkan latar belakang diatas, maka permasalahan yang dapat dirumuskan adalah sebagai berikut.

1. Bagaimana Tahapan Proses yang dilakukan untuk mendapatkan data chunking secara otomatis melalui IXA Pipeline sebagai Natural Language Processing Tools ?
2. Bagaimana Hasil dan analisa terhadap output kesamaan teks dan tingkat akurasinya ?

Tujuan

Tujuan yang ingin didapatkan penulis dalam tugas akhir yang dikerjakan ini adalah Mengetahui proses chunking otomatis untuk mendapatkan data chunking teks dan Mendapatkan output kesamaan teks dan skor Evaluation Metrics dengan membandingkan data sistem dan gold standart.