

# 1. Pendahuluan

## Latar Belakang

Bahasa Inggris mempunyai banyak kata yang berulang dalam berbagai tulisan ataupun kata-kata yang masih mempunyai keterhubungan atau topik yang saling berkaitan. Oleh karena dibuat implementasi mengenai konsep Interpretable (STS) pada potongan kata berbahasa Inggris. Tujuannya untuk melakukan pengukuran kesamaan makna dengan menentukan label keterhubungan dan penghitungan skor kesamaan berdasarkan metode chunk alignment.

*Interpretable Semantic Textual Similarity (ISTS)* merupakan metode untuk melakukan penilaian kesamaan teks semantik dari pasangan kalimat. Dalam *Interpretable STS (ISTS)*[3] kesamaan dihitung pada tingkat potongan, dan cara kerja sistemnya adalah meminta untuk memberikan jenis hubungan antara dua potongan sebagai interpretasi dari kesamaan. Dalam melakukan pengukuran kesamaan antar kata yang dilakukan yaitu berpanduan pada SemEval2016 Task2, karena SemEval2016 Task2 merupakan salah satu sarana penelitian mengenai pengukuran kesamaan semantik yang telah menjadi acuan para peneliti dibidang linguistik.

Untuk mencari kesamaan atau keterkaitan semantik dapat digunakan metode *Chunking Otomatis*, *Chunk Alignment* dan *Semantic Relation Prediction*. Karena metode *Chunking Otomatis*, *Chunk Alignment* dan *Semantic Relation Prediction* digunakan untuk menghitung kesamaan semantik. Kesamaan semantik didasarkan pada pohon linguistik yang masih dalam satu pohon linguistik. Berdasarkan masalah tersebut pada tugas akhir ini metode yang digunakan adalah konten informasi dalam mengukur kesamaan semantik untuk mengetahui tingkat keakuratan kesamaan antara 2 buah data teks. Metode *Chunking Otomatis*, *Chunk Alignment* dan *Semantic Relation Prediction* digunakan karena dapat mengukur kesamaan semantik dengan akurasi yang baik. Pengukuran dilakukan bukan berdasar pada satu struktur analogi tetapi dari hasil kombinasi dari struktur hirarki pada ontologi dengan menggunakan aktual kata pada suatu dataset yang besar.

## Topik dan Batasannya

Berdasarkan latar belakang masalah tersebut, maka permasalahan yang dapat dirumuskan sebagai berikut.

1. Bagaimana tahapan proses yang dilakukan untuk menentukan kesamaan makna antar potongan kata bahasa Inggris dengan metode *Chunking Otomatis*, *Chunk Alignment* dan *Semantic Relation Prediction*?
2. Bagaimana hasil analisis terhadap pengukuran kesamaan makna terhadap potongan kata dalam bahasa Inggris dengan metode *Chunking Otomatis*, *Chunk Alignment* dan *Semantic Relation Prediction*?
3. Bagaimana evaluasi dari pengukuran kesamaan makna antar kedua potongan kata bahasa Inggris dengan metode *Chunking Otomatis*, *Chunk Alignment* dan *Semantic Relation Prediction*?

## Tujuan

Berdasarkan rumusan masalah tersebut, maka tujuan yang dicapai adalah sebagai berikut.

1. Memahami serta menerapkan tahapan secara lengkap proses untuk menentukan kesamaan makna antar potongan kata bahasa Inggris dengan metode *Chunking Otomatis*, *Chunk Alignment* dan *Semantic Relation Prediction*.
2. Memahami bagaimana hasil analisis terhadap pengukuran kesamaan makna terhadap potongan kata bahasa Inggris metode *Chunking Otomatis*, *Chunk Alignment* dan *Semantic Relation Prediction*.
3. Menganalisis evaluasi pengukuran kesamaan makna antar kedua potongan kata bahasa Inggris dengan metode *Chunking Otomatis*, *Chunk Alignment* dan *Semantic Relation Prediction*.

## Batasan Masalah

Adapun batasan masalah dari Tugas Akhir ini adalah:

1. Pengukuran kesamaan makna hanya digunakan untuk teks yang relatif pendek, hal ini berdasarkan hasil penelitian yang ada di SemEval2016 Task2 untuk melakukan perhitungan nilai kesamaan hanya baru bisa digunakan untuk teks yang relatif pendek.
2. Penelitian ini hanya untuk mengetahui tingkat kesamaan makna dengan menentukan jenis label dan skor kesamaan dari setiap alignment potongan kata bahasa Inggris.
3. Data Input pasangan kata telah dilakukan chunk secara manual(di luar sistem)