

## PANDUAN PEMBUATAN & ANOTASI PATTERN

Made Nindyatama N | 1306381622 | Ilmu Komputer 2013

### LATAR BELAKANG

*Natural Language Processing* (NLP) adalah salah satu cabang ilmu komputer yang berusaha melakukan pemrosesan data berbahasa manusia menggunakan mesin (komputer). Salah satu cabang dalam NLP adalah *Language Resource Development* atau pengembangan *resource* dimana *resource* (korpus) linguistik dikembangkan agar dapat dimanfaatkan dalam berbagai penelitian lain. *Resource* yang dimiliki Bahasa Indonesia masih sangat terbatas. Salah satu informasi yang dibutuhkan adalah pengetahuan mengenai relasi semantik antar kata. Topik skripsi ini adalah “Ekstraksi Relasi Semantik Kata secara Otomatis dengan Pattern Matching”.

**Relasi semantik** kata adalah hubungan antar kata berdasarkan arti kata (*sense*). Beberapa relasi semantik sebuah kata seperti *synonym*, *antonym*, *hypernym-hyponym*, dan *meronym-holonym*. Pada penelitian ini difokuskan pada relasi *hypernym-hyponym*. *Hyponym* adalah relasi yang menyatakan hubungan kata yang lebih khusus. Sementara untuk kata yang lebih umum disebut *hypernym*. Sebagai contoh ‘kucing’, ‘kelinci’, dan ‘ikan’ adalah *hyponym* sementara ‘binatang’ adalah *hypernym*.

Metode yang digunakan pada penelitian ini adalah **Pattern Extraction** dan **Matching**. Dari suatu text berukuran besar, akan dibentuk secara otomatis *pattern* yang merepresentasikan relasi tersebut. Text yang digunakan untuk proses ekstraksi adalah dokumen-dokumen Wikipedia. *Pattern* yang dibentuk adalah *pattern lexical* yang berarti hanya melihat struktur leksikal sebuah kalimat. Sebagai contoh terdapat sebuah *pattern* berikut.

<hyponym> adalah <hypernym> yang

- <hyponym> : merepresentasikan kata yang akan ditandai sebagai relasi hyponym
- <hypernym> : merepresentasikan kata yang akan ditandai sebagai relasi hypernym

Dari *pattern* diatas, jika diterapkan pada artikel Wikipedia (<https://id.wikipedia.org/wiki/Anjing>) akan menghasilkan pasangan kata: (hyponym: anjing, hypernym: mamalia). *Pattern* yang dihasilkan oleh sistem berjumlah banyak dan tidak semuanya valid. Untuk mengetahui kualitas *pattern*, perlu dilakukan evaluasi secara manual oleh manusia.

### PEMBUATAN PATTERN MANUAL

Pada penelitian ini, Anda diminta untuk membuat *pattern* secara manual. Pembuatan *pattern* manual dilakukan untuk membandingkan apakah *pattern* yang dihasilkan sistem sesuai dengan *pattern* yang dibuat oleh anotator yang ahli di bidang linguistik. Berikut adalah deskripsi pekerjaan untuk pembuatan *pattern* manual.

Anotator membuat secara manual *pattern* yang dianggap dapat mengekstraksi kata berelasi *hypernym-hyponym*. Proses ini memerlukan anotator yang ahli di bidang linguistik. *Pattern* yang dibuat secara manual berjumlah sebanyak mungkin. Untuk mempermudah, pembuatan *pattern* bisa menggunakan referensi artikel-artikel Wikipedia yang juga digunakan pada sistem.

Berikut ditampilkan contoh beberapa *pattern* yang dihasilkan dari hasil ekstraksi oleh sistem.

<start> <hyponym> adalah <hypernym>
<hypernym> termasuk <hyponym>
<hyponym> adalah <hypernym> yang

<hyponym> adalah sebuah <hypernym>

<hypernym> khususnya <hyponym>

Keterangan: simbol <start> sebagai penanda awal kalimat

### ANOTASI PATTERN

Setelah membuat *pattern*, anotator diminta untuk melakukan penilaian (anotasi) terhadap *pattern* yang dihasilkan secara otomatis oleh sistem. Proses evaluasi *pattern* yang dilakukan secara otomatis oleh sistem hanya berdasarkan frekuensi (kemunculan) *pattern* tersebut. Sistem belum dapat melakukan evaluasi mengenai semantik *pattern* yang terbentuk, apakah benar *pattern* tersebut sesuai untuk mengekstrak relasi *hypernym-hyponym*. Berikut adalah deskripsi pekerjaan untuk menganotasi *pattern* yang dihasilkan sistem.

Anotator diberi data yang berisi kumpulan *pattern* yang dihasilkan oleh sistem. Kemudian, anotator **memberi nilai kebaikan** untuk setiap *pattern*. Nilai kebaikan dilihat berdasarkan dua parameter berikut.

- berapa **jumlah** kata *hypernym-hyponym* **benar** yang dapat dibentuk oleh *pattern* tersebut.
- Berapa **jumlah** kata *hypernym-hyponym* **salah** yang dapat dibentuk dari *pattern* tersebut.

Tiga nilai anotasi untuk mengukur jumlah kata *hypernym-hyponym* benar/salah adalah 1 (berjumlah sedikit), 2 (berjumlah sedang), 3 (berjumlah banyak). Berikut adalah contoh *pattern* yang perlu dianotasi.

No	Pattern	Jumlah kata benar (B)	Jumlah kata salah (S)
1	Salah satu <hypernym> adalah <hyponym>	2	3
2	<hyponym> adalah suatu <hypernym>	3	1
3	<hyponym> adalah <hypernym>	3	2

Keterangan: Nilai di atas hanya contoh dan ditulis secara *random*

Dari contoh diatas dapat dikatakan bahwa *pattern* 'salah satu <hypernym> adalah <hyponym>' dapat menghasilkan pasangan kata *hypernym-hyponym* benar berjumlah sedang dan pasangan kata salah berjumlah banyak. Untuk memperkuat penilaian Anda, silahkan tuliskan beberapa **contoh pasangan kata *hypernym-hyponym* salah** yang dapat dibentuk dari *pattern* tersebut.

Anda dapat memberi **komentar mengenai *pattern*** yang terbentuk. Komentar bisa mengenai apakah *pattern* tersebut baik/tidak, apakah *pattern* terlalu pendek/panjang, apakah *pattern* seharusnya mengandung kata-kata tertentu, dll.

- Terima kasih -