

# BAB 1

## PENDAHULUAN

Bab ini membahas mengenai latar belakang penelitian, perumusan masalah, tujuan dan manfaat dilakukannya penelitian, ruang lingkup penelitian, dan terakhir metodologi yang digunakan.

### 1.1 Latar Belakang

Berbagai penelitian yang berkaitan dengan *Information Retrieval* dan *Natural Language Processing* mulai muncul menggunakan Bahasa Indonesia. Informasi mengenai kata dan artinya sangat dibutuhkan. Sayangnya, *resource* yang dimiliki berbasis Bahasa Indonesia masih sangat terbatas. Informasi tersebut umumnya disimpan dalam kamus digital seperti WordNet. Penelitian ini dilaksanakan untuk mengembangkan *resource* dalam Bahasa Indonesia.

WordNet adalah kamus digital yang dapat digunakan untuk menunjang penelitian di bidang *Information Retrieval* dan *Natural Language Processing*. WordNet yang paling sering digunakan dalam berbagai penelitian adalah WordNet Princeton berbahasa Inggris yang dibentuk secara manual oleh ahli linguistik. Setiap *entry* pada WordNet disimpan dalam bentuk set sinonim atau biasa disebut *synset* dan arti dari *synset* tersebut atau biasa disebut *sense*. Informasi lain yang disimpan dalam suatu *synset* adalah relasi antar *synset*. Relasi semantik yang tersimpan dalam WordNet adalah *synonym*, *antonym*, *hypernym*, *hyponym*, *holonym*, *meronym*, *troponym*, dan *entailment*.

Beberapa penelitian telah dilakukan untuk membangun WordNet Bahasa Indonesia. WordNet Bahasa Indonesia yang telah ada adalah Indonesian WordNet (IWN) yang dibuat oleh Fakultas Ilmu Komputer Universitas Indonesia (Fasilkom UI) dan WordNet Bahasa yang dibuat oleh Nanyang Technology University (NTU). Salah satu kelemahan kedua WordNet tersebut adalah ukurannya yang masih sangat terbatas. Selain itu, informasi mengenai relasi kata juga belum dapat tersimpan secara baik. Kedua WordNet memetakan *synset* Bahasa Indonesia ke WordNet Princeton dan memanfaatkan relasi yang di dalamnya. Mengetahui hal tersebut, dilakukan penelitian yang dapat mengekstrak relasi antar kata dengan hanya menggunakan korpus Bahasa Indonesia. Proses ekstraksi berjalan secara cepat dan data yang dihasilkan berjumlah besar. Korpus yang dihasilkan diharapkan dapat berguna

untuk penelitian selanjutnya.

Relasi kata adalah salah satu hal penting yang perlu diketahui jika ingin mengetahui hubungan antar kata secara semantik. Informasi yang berkaitan dengan semantik atau arti kata sulit diperoleh tanpa adanya pengetahuan sebelumnya. Kata-kata yang mirip secara leksikal belum tentu berelasi secara semantik. Sementara kata-kata yang tidak memiliki kesamaan secara leksikal bisa memiliki arti yang mirip atau berhubungan secara semantik. Korpus relasi yang dibuat diharapkan dapat membantu memperoleh informasi tersebut. Selain itu, pengetahuan mengenai relasi kata dapat dimanfaatkan dalam berbagai penelitian lain seperti *question answering* (Ravichandran dan Hovy, 2002), *information extraction*, dan *anaphora resolution*.

Melihat adanya kebutuhan akan korpus relasi kata, dilakukanlah penelitian *word relation extraction*. Penelitian ini berusaha mengekstrak kata berdasarkan relasi tertentu dari suatu dokumen sehingga dihasilkan korpus relasi kata. Penelitian kali ini akan fokus pada relasi kata *hypernym-hyponym*. Keduanya menyatakan relasi antara kata yang lebih umum (*hypernym*) dengan kata yang lebih khusus (*hyponym*). Metode yang digunakan adalah *pattern matching* dengan memanfaatkan korpus Wikipedia Bahasa Indonesia. Wikipedia memuat banyak kata dari berbagai domain sehingga dapat dimanfaatkan untuk membuat *pattern* yang general serta menghasilkan korpus relasi berukuran besar.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, pertanyaan yang menjadi rumusan penelitian adalah sebagai berikut.

1. Bagaimana cara membangun korpus relasi secara cepat dan berkualitas baik secara otomatis?
2. Apakah metode *pattern matching* baik dilakukan untuk ekstraksi relasi Bahasa Indonesia?
3. Bagaimana cara mengevaluasi *pattern* dan korpus relasi dari hasil eksperimen?

### **1.3 Tujuan dan Manfaat Penelitian**

### **1.4 Tahapan Penelitian**

### **1.5 Ruang Lingkup Penelitian**

### **1.6 Sistematika Penulisan**

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
- Bab 2 LANDASAN TEORI
- Bab 3 APLIKASI YANG DIGUNAKAN
- Bab 4 PERANCANGAN IMPLEMENTASI DAN ANALISIS
- Bab 5 IMPLEMENTASI DAN PENGUJIAN
- Bab 6 HASIL IMPLEMENTASI DAN EVALUASI
- Bab 7 KESIMPULAN DAN SARAN

## **DAFTAR REFERENSI**

Ravichandran, D. dan Hovy, E. (2002). Learning surface text patterns for a question answering system.