

# 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 extraction* dan *matching* baik dilakukan untuk ekstraksi relasi kata Bahasa Indonesia?
3. Bagaimana cara mengevaluasi *pattern* dan korpus relasi kata dari hasil eksperimen?

## 1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian *word relation extraction* ini adalah membangun korpus relasi kata Bahasa Indonesia berukuran besar dan berkualitas baik secara otomatis. Selain

itu, ingin diketahui pula apakah metode *pattern extraction* dan *matching* baik digunakan untuk mengekstrak relasi kata Bahasa Indonesia. Diharapkan korpus relasi kata yang dihasilkan dapat menunjang berbagai penelitian selanjutnya.

Penelitian ini juga diharapkan dapat memotivasi adanya penelitian selanjutnya di bidang *Language Resource Development*, terutama pembangunan WordNet Bahasa Indonesia. Penelitian mengenai ekstraksi relasi kata berikutnya dengan berbagai metode lain diharapkan terus dilaksanakan sehingga Bahasa Indonesia memiliki *resource* yang semakin baik.

## 1.4 Ruang Lingkup Penelitian

Penelitian ini hanya fokus pada pembuatan korpus pasangan kata dengan relasi *hypernym* dan *hyponym* Bahasa Indonesia. Kelas kata yang menjadi fokus penelitian adalah kata benda (*noun*). Pengembangan korpus dilakukan secara *semi-supervised* dengan metode *pattern matching*. *Pattern* yang dibuat masih terbatas hanya merupakan *pattern* leksikal. Evaluasi akan dilakukan pada *pattern* dan korpus pasangan kata yang dihasilkan. Proses evaluasi korpus pasangan kata dilakukan menggunakan teknik *random sampling*. Data yang digunakan untuk pembuatan *pattern* maupun untuk ekstraksi korpus pasangan kata baru adalah Wikipedia Bahasa Indonesia. Pemilihan relasi *hypernym-hyponym* dalam penelitian karena memiliki berbagai manfaat. Relasi tersebut dapat mengidentifikasi *Proper Noun* baru yang belum diketahui maknanya.

## 1.5 Tahapan Penelitian

Proses penelitian dilakukan dalam beberapa tahapan sebagai berikut.

### 1. Studi Literatur

Pada tahap ini, dilakukan pembelajaran mengenai penelitian-penelitian sebelumnya yang telah dilakukan di bidang *word relation extraction* sehingga diketahui langkah yang perlu diambil selanjutnya.

### 2. Perumusan Masalah

Perumusan masalah dilakukan untuk mendefinisikan masalah yang ingin diselesaikan, tujuan penelitian, dan hasil yang diharapkan sehingga proses penelitian dapat berjalan dengan baik.

### 3. Perancangan Penelitian

Setelah diketahui hasil yang ingin dicapai, dirancang tahap-tahap eksperimen

secara terstruktur. Hal-hal yang diperhatikan mulai dari pengumpulan korpus awal (*seed*), *pre-processing* dokumen, perancangan implementasi *pattern extraction matching*, hingga proses evaluasi.

#### 4. Implementasi

Implementasi dilaksanakan sesuai dengan rancangan penelitian untuk menjawab rumusan masalah. Segala hasil yang ditemukan digunakan untuk terus memperbaiki metode dan teknik penelitian sehingga didapatkan hasil yang semakin baik.

#### 5. Analisis dan Kesimpulan

Tahap terakhir dari penelitian ini adalah menganalisis korpus pasangan kata relasi yang dihasilkan. Pertanyaan dari perumusan masalah dijawab, kemudian ditarik kesimpulan.

## BAB 2

### TINJAUAN PUSTAKA

Pada bab ini, dijelaskan mengenai studi literatur yang dilakukan. Studi literatur yang dilakukan digunakan sebagai dasar konsep dan teknik penelitian. Dipaparkan pula berbagai istilah dan metode yang digunakan dalam penelitian.

#### 2.1 WordNet

WordNet adalah kamus leksikal yang tersimpan secara digital dan digunakan untuk berbagai keperluan komputasi (Miller, 1995). Pembuatan WordNet dilatarbelakangi keperluan mendapatkan *sense* atau arti semantik suatu kata. Informasi tersebut perlu disimpan dan dapat dibaca oleh mesin. WordNet pertama dibuat oleh George A. Miller pada tahun 1995 berbasis Bahasa Inggris dan sekarang dikenal dengan nama Princeton WordNet (PWN). WordNet menyimpan informasi dalam bentuk database dimana setiap entry-nya adalah pasangan *synset* dan arti semantiknya (*sense*). Set sinonim (*synset*) adalah himpunan kata yang memiliki arti yang sama atau saling berelasi *synonym*.

WordNet mengandung beberapa kelas kata seperti kata benda (*noun*), kata kerja (*textitverb*), kata sifat (*textitadjective*), dan kata keterangan (*adverb*). WordNet juga menyimpan informasi mengenai relasi semantik antar *synset*. Relasi yang disimpan adalah *synonymy*, *antonymy*, *hyponymy*, *hypernym*, *meronymy*, *holonymy*, *troponymy*, dan *entailment*.

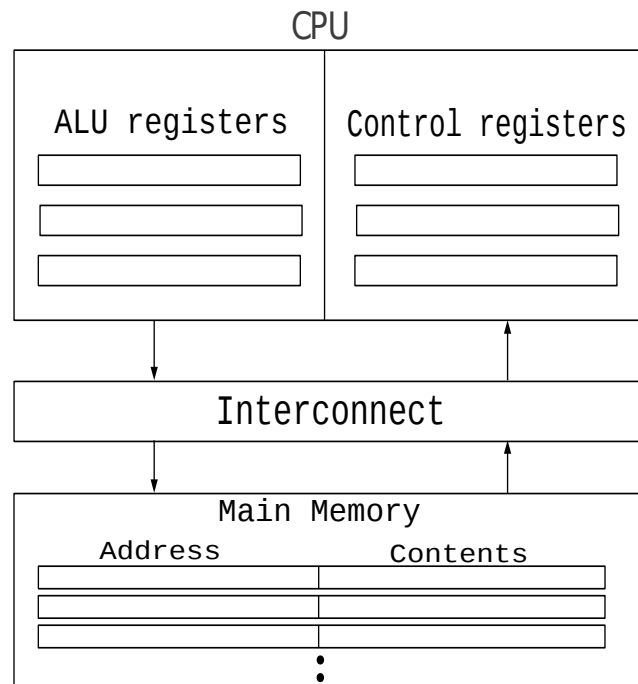
##### 2.1.1 Indonesian WordNet

##### 2.1.2 Klasifikasi X

Figure dalam enum dan dua sitasi sekaligus (Buyya, 1999; Jones et al., 2002) :

1. ***Bold Italic***

Penjelasan..... Untuk gambarannya dapat dilihat di Gambar 2.1.



**Gambar 2.1:** Arsitektur klasik von Neumann

Sumber gambar terinspirasi dari: (Pressman, 2010)

## 2. *Sesuatu banget*

Penjelasan.....

## 2.2 *Section in Eng*

Hal pertama yang mungkin ditanyakan adalah bagaimana membuat huruf tercetak tebal, miring, atau memiliki garis bawah. Pada Texmaker, Anda bisa melakukan hal ini seperti halnya saat mengubah dokumen dengan LO Writer. Namun jika tetap masih tertarik dengan cara lain, ini dia:

- **Bold**  
Gunakan perintah `\textbf{}` atau `\bo{}`.
- *Italic*  
Gunakan perintah `\textit{}` atau `\f{}`.
- Underline  
Gunakan perintah `\underline{}`.

- *Overline*  
Gunakan perintah `\overline`.
- *superscript*  
Gunakan perintah `\{ }`.
- *subscript*  
Gunakan perintah `\_ { }`.

Perintah `\f` dan `\bo` hanya dapat digunakan jika package uithesis digunakan.

### 2.2.1 Pengertian *Section in Eng*

### 2.2.2 Next Subsection *Section in Eng*

## 2.3 *Keatas lagi*

Contoh cite yang ga ada ?. Cite author Neal, cite tahun 2004, cite mention Guarddin (2010), dan cite di akhir kalimat (Mell dan Grance, 2009).

### 2.3.1 *Masuk lagi*

Footnote example nih : MPICH <sup>1</sup>, LAM/MPI <sup>2</sup>, dan OpenMPI <sup>3</sup> (McGuire, 2010). MPI-3 sedang dalam tahap perencanaan <sup>4</sup>. Fungsi-fungsi tersebut berada di tabel 2.1. (Contoh tabel).

**Tabel 2.1:** Fungsi fundamental MPI

No.	Nama Fungsi	Penjelasan
1	MPI_Init	Memulai kode MPI
2	MPI_Finalize	Mengakhiri kode MPI
3	MPI_Comm_size	Menentukan jumlah proses
4	MPI_Comm_rank	Menentukan label proses
5	MPI_Send	Mengirim pesan
6	MPI_Recv	Menerima pesan

Sumber tabel: taro sitasi disini, if i were u

<sup>1</sup><http://www.mpich.org/>

<sup>2</sup><http://www.lam-mpi.org/>

<sup>3</sup>[www.open-mpi.org](http://www.open-mpi.org)

<sup>4</sup>[http://meetings.mpi-forum.org/MPI\\_3.0\\_main\\_page.php](http://meetings.mpi-forum.org/MPI_3.0_main_page.php)

## DAFTAR REFERENSI

- Buyya, R. (1999). *High Performance Cluster Computing: Architectures and systems*. High Performance Cluster Computing. Prentice Hall PTR.
- Guarddin, G. (2010). Percobaan Kompresi Menggunakan MPIBZIP2 pada Cluster Hastinapura. Peronal Communication.
- Jones, J. P., Lifka, D., Nitzberg, B., dan Tannenbaum, T. (2002). Cluster workload management. In Sterling, T., editor, *Beowulf cluster computing with Linux*, pages 301–306. MIT Press, Cambridge, MA, USA.
- McGuire, T. J. (2010). A gentle way of introducing multi-core programming into the curriculum: tutorial presentation. *J. Comput. Sci. Coll.*, 26(1):124–125.
- Mell, P. dan Grance, T. (2009). The NIST definition of cloud computing. Technical report, National Institute of Standards and Technology, Information Technology Laboratory. [Diakses 17 April 2013].
- Neal, R. (2010). Cloud computing brings savings in energy, maintenance costs. <http://www.federaltimes.com/article/20100323/IT03/3230304/Cloud-computing-brings-savings-energy-maintenance-costs>. [Diakses 1 April 2013].
- Oxford Dictionaries (2010). <http://oxforddictionaries.com/definition/english/parallel>. [Diakses 1 April 2013].
- Pressman, R. (2010). *Software engineering: a practitioner's approach*. McGraw-Hill higher education. McGraw-Hill Higher Education.
- Ravichandran, D. dan Hovy, E. (2002). Learning surface text patterns for a question answering system.
- Treese, W. (2004). How to build a supercomputer. *netWorker*, 8(4):15–18.