

# BAB 1

## PENDAHULUAN

Bab ini membahas latar belakang penelitian, perumusan masalah, tujuan penelitian, dan ruang lingkup penelitian. Bagian latar belakang berisi penjelasan perkembangan perolehan informasi secara umum. Bagian perumusan masalah berisi penjelasan mengenai hal yang ingin diselidiki dan alasannya. Bagian tujuan berisi penjabaran tujuan dari penelitian ini. Bagian ruang lingkup berisi penjelasan batasan kerja penulis pada penelitian ini.

### 1.1 Latar Belakang

Untuk memenuhi kebutuhan manusia dalam mendapatkan informasi yang banyak dalam waktu yang singkat, teknologi informasi dikembangkan dengan semakin cepat. Saat ini, informasi menjadi lebih mudah didapat melalui internet. Pencarian informasi melalui internet biasanya dilakukan dengan menggunakan mesin pencari (*search engine*) yang ada di internet. Beberapa mesin pencari yang tersedia di internet antara lain adalah Google<sup>1</sup>, Yahoo!<sup>2</sup>, dan Vivisimo<sup>3</sup>.

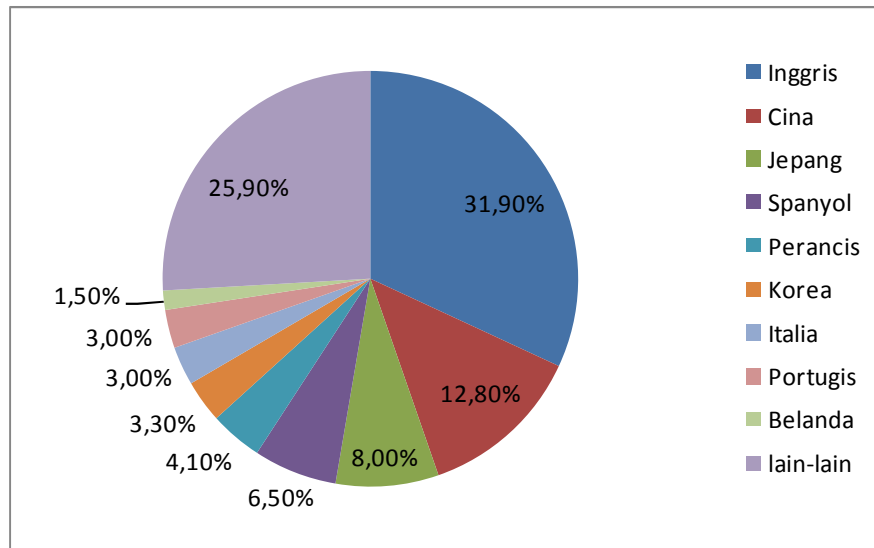
Berbagai informasi atau dokumen yang dapat diakses melalui internet tidak hanya terdiri dari satu bahasa saja. Hal tersebut karena jutaan pemakai internet berasal dari berbagai macam bangsa dengan berbagai macam bahasa. Menurut data statistik tahun 2005 [Hart, 2006], pemakai internet yang berbahasa Inggris ada sekitar 31.9%. Hal tersebut berarti bahwa 68.1% pemakai internet lainnya adalah orang-orang yang berkomunikasi dengan bahasa selain Inggris. Informasi lebih rinci mengenai bahasa pengakses internet terdapat dalam Gambar 1.1.

---

<sup>1</sup> <http://www.google.com/>

<sup>2</sup> <http://www.yahoo.com/>

<sup>3</sup> <http://www.vivisomo.com/>



**Gambar 1.1 Data statistik pengguna internet berdasarkan bahasa**

Dari Gambar 1.1 terlihat banyak dari para pemakai internet yang menyediakan dan mengakses informasi di internet dengan berbagai bahasa yang mereka mengerti, selain bahasa Inggris.

Manusia ingin mendapatkan informasi sebanyak mungkin tanpa terhalang oleh faktor bahasa. Oleh karena itu, di berbagai negara dikembangkan penelitian-penelitian mengenai pengambilan informasi atau dokumen-dokumen dengan bahasa yang berbeda-beda tetapi berisi informasi atau topik yang sejenis.

Cara untuk mendapatkan informasi biasanya dengan menggunakan kueri. Misalnya untuk mencari informasi di internet kita mengetik kueri pada mesin pencari. Kueri mengandung kata kunci atau kata yang dianggap mewakili topik atau informasi penting yang ingin dicari. Mesin pencari kemudian memberikan daftar dokumen yang dianggap sesuai dengan kueri yang diberikan. Jika kueri yang diberikan adalah kata berbahasa Indonesia, biasanya hasil pencarian yang keluar adalah dokumen-dokumen dalam bahasa Indonesia juga. Untuk mengambil dokumen dalam bahasa yang berbeda dari bahasa yang digunakan pada kueri, maka dilakukan perolehan informasi lintas bahasa.

Perolehan atau pengambilan informasi lintas bahasa dapat berarti mengambil informasi atau dokumen yang ditulis dalam suatu bahasa yang berbeda dari bahasa

pada kata kunci. Perolehan informasi lintas bahasa dapat pula berarti mengambil dokumen dalam beberapa bahasa yang berbeda.

Perolehan informasi lintas bahasa dapat dilakukan dengan menerjemahkan kata kunci ke bahasa yang sama dengan dokumen yang dicari, atau dengan menerjemahkan dokumen ke bahasa yang sama dengan kueri [Talvenssaari et al., 2006]. Menerjemahkan kueri ke bahasa dokumen yang dicari akan lebih mudah, selain karena kata yang harus diterjemahkan lebih sedikit, kueri hanya mementingkan arti dan konteks dari kata. Sedangkan jika menerjemahkan dokumen, selain kata yang harus diterjemahkan lebih banyak, juga harus diperhatikan susunan kata yang benar dan aturan bahasa lainnya. Namun, dokumen yang diterjemahkan tentunya akan lebih memudahkan pencari informasi, karena pencari informasi akan lebih mengerti isi dokumen yang bahasanya sama dengan bahasa yang digunakan pada kueri.

Kebutuhan untuk mengakses informasi dalam bahasa yang berbeda-beda mendorong kami untuk mengembangkan sistem perolehan informasi lintas bahasa. Penelitian yang pernah dilakukan saat ini adalah perolehan informasi lintas bahasa Indonesia-Inggris. Hal tersebut karena bahasa Inggris adalah bahasa internasional yang populer dan telah menjadi mata pelajaran wajib di Indonesia.

Pemerintah juga mulai menyadari pentingnya menyediakan informasi dalam bahasa selain bahasa Indonesia. Beberapa situs web pemerintah dan perusahaan Indonesia mulai dilengkapi dengan halaman web berbahasa Inggris.

## **1.2 Perumusan Masalah**

Teknik perolehan informasi lintas bahasa yang pernah dikembangkan antara lain dengan menggunakan mesin penerjemah dan menggunakan kamus dwi-bahasa. Namun, mesin penerjemah sulit untuk dikembangkan karena membutuhkan banyak sumber daya dan menggunakan pengetahuan bahasa. Kamus dwi-bahasa juga mempunyai beberapa kelemahan, antara lain satu kata dapat mempunyai

banyak arti. Arti yang satu dapat berbeda dari arti yang lain. Selain itu kamus dwi-bahasa Inggris-Indonesia yang bagus masih sangat terbatas jumlahnya.

Oleh karena itu, pada penelitian ini penulis ingin mengembangkan perolehan informasi lintas bahasa berdasarkan korpus. Korpus yang digunakan adalah korpus sebanding. Berikut ini adalah masalah yang ingin diselidiki oleh penulis pada penelitian ini:

1. Bagaimana cara mendapatkan sumber korpus dalam bahasa Indonesia dan bahasa Inggris melalui internet?
2. Bagaimana cara memasang dokumen Inggris yang sebanding dengan dokumen Indonesia?
3. Bagaimana cara mendapatkan daftar kata Inggris-Indonesia dari korpus yang sebanding dengan menggunakan teknik statistik  $\phi$  square dan teknik statistik koefisien korelasi Pearson?
4. Apakah daftar kata Inggris-Indonesia yang dihasilkan dari korpus yang sebanding dapat menerjemahkan kueri Inggris ke dalam bahasa Indonesia sedemikian sehingga bisa mendapatkan dokumen Indonesia dengan tingkat relevansi yang tinggi?

### **1.3 Tujuan**

Tujuan dari penelitian ini antara lain adalah sebagai berikut:

5. Meneliti kinerja perolehan informasi lintas bahasa Indonesia-Inggris dengan menggunakan korpus yang sebanding melalui teknik statistik  $\phi$  square dan koefisien korelasi Pearson untuk membuat kamus persamaan kata Inggris-Indonesia.
6. Membandingkan teknik statistik  $\phi$  square dan koefisien korelasi Pearson dengan kinerja perolehan informasi lintas bahasa menggunakan kamus dwi-bahasa dan mesin penerjemah Transtool.
7. Menambah koleksi korpus yang sebanding yang dapat digunakan untuk penelitian selanjutnya.

## **1.4 Ruang Lingkup**

Ruang lingkup dari penelitian ini antara lain adalah:

1. Menggunakan koleksi dokumen berbahasa Indonesia dan Inggris yang diperoleh dari internet.
2. Mencari pasangan dokumen Indonesia-Inggris yang berisi topik yang sama atau mirip (korpus yang sebanding).
3. Membuat daftar kata Inggris-Indonesia dari korpus yang sebanding yang ada.
4. Mengevaluasi kinerja daftar kata Inggris-Indonesia yang telah dibuat.

## **1.5 Metodologi Penelitian**

Penelitian ini dilakukan dengan menggunakan metodologi penelitian eksperimental.

Dalam penelitian ini, penulis melakukan hal-hal sebagai berikut:

1. Studi literatur

Pada tahap ini penulis mempelajari jurnal-jurnal yang meneliti perolehan informasi lintas bahasa, terutama yang menggunakan korpus paralel atau korpus yang sebanding.

2. Melakukan eksperimen

Hal-hal yang dilakukan penulis dalam eksperimen adalah:

- mengumpulkan koleksi dokumen yang sebanding dari internet,
- membuat daftar kata Inggris-Indonesia dari korpus (koleksi dokumen) yang sebanding dengan menggunakan teknik penghitungan statistik.

### 3. Melakukan analisis

Pada tahap ini penulis melakukan analisis atas hasil eksperimen yang telah dilakukan untuk menilai tingkat keberhasilan dari teknik yang dicoba pada eksperimen.

## 1.6 Sistematika Penulisan

Laporan ini ditulis dengan sistematika penulisan sebagai berikut:

- **BAB 1 PENDAHULUAN**  
Bab ini berisi latar belakang penelitian, perumusan masalah, tujuan penelitian, metodologi penelitian, dan ruang lingkup penelitian.
- **BAB 2 LANDASAN TEORI**  
Bab ini berisi penjelasan teori yang menjadi dasar penelitian ini dan penjelasan mengenai alat bantu yang digunakan dalam penelitian ini.
- **BAB 3 EKSPERIMEN**  
Bab ini berisi penjabaran hal-hal yang dilakukan dalam penelitian ini.
- **BAB 4 HASIL DAN ANALISIS**  
Bab ini berisi hasil-hasil yang didapat dari eksperimen yang dilakukan dan juga analisis dari hasil eksperimen tersebut.
- **BAB 5 PENUTUP**  
Bab ini berisi kesimpulan dari penelitian yang telah dilakukan. Bab ini juga berisi saran dari penulis untuk penelitian selanjutnya.