BAB 1
PENDAHULUAN

1.1 Latar Belakang

Setiap tahunnya jumlah kosakata Bahasa Indonesia berkembang dengan pesat, hal ini dapat dilihat pada bertambahnya muatan lema dalam kamus bahasa Indonesia dari satu edisi ke edisi berikutnya. Dalam kurun waktu 20 tahun Kamus Besar Bahasa Indonesia sejak saat diterbitkannya pada edisi pertama tahun 1988 memuat 62.000 lema, mengalami penambahan hingga 91.000 lema pada edisi ke empat tahun 2008 yang sampai saat ini dipakai. Hal tersebut di atas menunjukan bahwa kamus tersebut selalu memutahirkan kandungan lemanya. Perkembangan kamus saat ini tidak hanya berupa buku, namun dapat dijumpai berupa kamus digital yang dapat diakses secara online. Jenis-jenis kamus online yang berkembang hingga saat ini antara lain; KBBI Online Kemdikbud, KBBI-Universitas Indonesia, Kateglo dan lain-lain. Pengembangan kamus online Bahasa Indonesia pembangunannya tidak disertai dengan resource atau sumber daya, dengan kata lain pengguna hanya memakai dan tidak mempunyai hak akses untuk mengubah atau menambahkan. Keberadaan sumber daya kamus bagi pengembang dibutuhkan, karena selain membantu dalam proses pengembangan kamus online itu sendiri sumber daya kamus digital juga dapat dimanfaatkan pada penelitian yang berhubungan dengan NLP.

Natural Language Processing (NLP) merupakan salah satu cabang ilmu AI yang berfokus pada pengolahan bahasa natural [**1**]. Tujuan dalam bidang Natural Language ini adalah melakukan proses pembuatan model komputasi dari bahasa, sehingga dapat terjadi suatu interaksi antara manusia dengan komputer dengan perantaraan bahasa alami. Sumber daya kamus berguna selain sebagai data masukan seperti kata dasar, jenis kata, *stopword*, korpus dan lain sebagainya juga dalam pengklasifikasian kelas kata dalam proses *Question-*

Answering, language generator, information extraction, summarization, machine translation dan lain-lain.

POS Tag merupakan suatu cara pengkategorian kelas kata, seperti kata benda, kata kerja, kata sifat, dll. Jenis kata adalah golongan kata dalam satuan bahasa berdasarkan kategori bentuk fungsi dan makna dalam susunan sistem gramatikal. Kelas kata atau jenis kata terbagi menjadi beberapa golongan tergantung narasumber yang digunakan dan untuk kebutuhan apa penggunaannya. Untuk menyusun kalimat yang baik dan benar dengan berdasarkan pola-pola kalimat baku, pemakai bahasa harus mengenal jenis dan fungsi kelas kata dan untuk pengkategorian kata dibutuhkan kelas kata.

Berdasarkan pada kebutuhan sumber daya kamus yang bersifat open source khususnya bahasa Indonesia, maka pembangunan kamus jenis kata sebagai sumber daya NLP diharapkan dapat menjadi salah satu sumber daya yang dapat dipakai dalam pengembangan teknologi di bidang bahasa. Pembangunan kamus jenis kata dilakukan dengan pendekatan bidang NLP.

1.2 Identifikasi Masalah

Identifikasi masalah menurut latar belakang di atas adalah pengembangan kamus *online* pembangunannya tidak disertai dengan *resource* atau sumber daya dibutuhkannya sumber daya kamus sebagai *resource* NLP Bahasa Indonesia yang bersifat *open source*. Hal tersebut di atas mengakibatkan keterbatasan dalam penggunaan bagi pengembang di bidang NLP.

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah implementasi pembangunan kamus jenis kata sebagai sumber daya NLP Bahasa Indonesia. Adapun tujuannya adalah menyediakan sumber daya jenis kata bagi penelitian NLP Bahasa Indonesia bersifat *open source* dengan beberapa format seperti *.txt*, *.csv*, dan *.json*.

1.4 Batasan Masalah

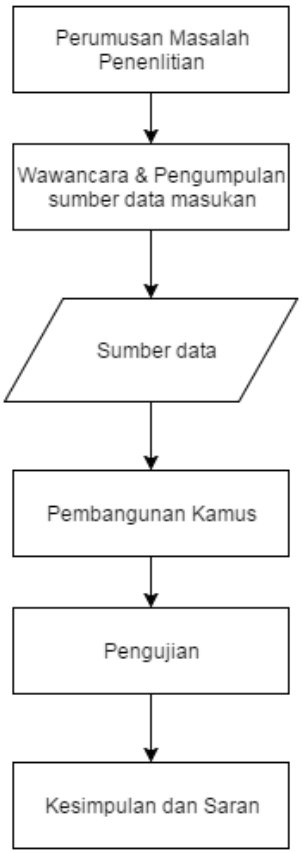
Agar permasalahan yang dirumuskan dapat lebih fokus, maka dalam penelitian ini dibatasi dalam hal-hal sebagai berikut:

1. Data yang diolah dalam pembangunan kamus jenis kata adalah sekumpulan kata yang diambil dari Kamus Besar Bahasa Indonesia dengan format .txt
2. Jenis kata yang diolah terdiri dari 7 jenis kata,[1] diantaranya; Kata kerja (verba), Kata benda (nomina), Kata ganti (pronomina), Kata bilangan (numeralia), Kata sifat (adjektiva), Kata keterangan (adverbia), Kata sandang (partikel).
3. Kamus jenis kata yang dibangun hanya mengolah kata dasar dan turunannya, tidak mengolah kata majemuk.
4. Keluaran dari pembangunan sistem adalah kamus jenis kata dengan format *.txt,.csv*, dan *.json*.

1.5 Metodologi Penelitian

Jenis penelitian yang dipakai bersifat deskriptif merupakan yang mempelajari masalah-massalah yang ada serta tata cara kerja yang berlaku [**2**]. Pada metode ini peneliti menghimpun data, menyusunnya secara sistematis, faktual dan cermat. Metode penelitian ini memiliki dua tahapan yaitu pengumpulan data dan pembangunan perangkat lunak.

Peneliti menggunakan metode ini karena sesuai dengan penelitan pada pembangunan kamus online yang akan dibangun berdasarkan pada fakta-fakta yang didapatkan pada pengumpulan data.



Gambar 1. 1 Skema Metode Penelitian Deskriptif

1.5.1 Metode Pengumpulan Data

Teknik yang digunakan untuk mengumpulkan data adalah sebagai berikut:

1. Wawancara

   Teknik pengumpulan data dengan melakukan wawancara berupa tanya jawab langsung dengan pihak terkait dengan objek yang diteliti, dalam hal ini adalah melakukan tanya jawab dengan pihak yang ahli di bidang bahasa khususnya Bahasa Indonesia diantaranya Balai Bahasa Bandung, Laboratorium Leksikologi dan Leksikografi di Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia.

2. Studi Pustaka

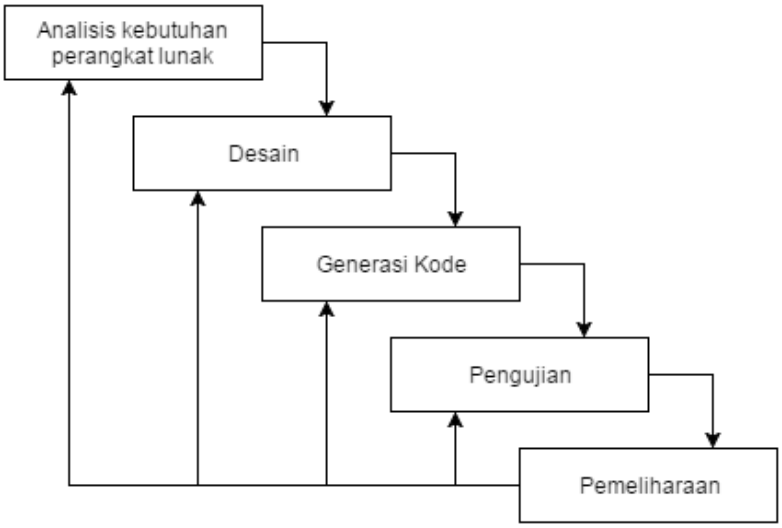
   Teknik pengumpulan data dengan cara mengambil kata dan jenis kata dari KBBI dan buku-buku.

3. Observasi

   Teknik pengumpulan data dengan mengadakan pengamatan atau penelitian langsung terhadap objek penelitian. Dalam hal ini peneliti mengumpulkan data dengan mengambil sumber dari Kamus Besar Bahasa Indonesia-Badan Pengembangan dan Pembinaan Bahasa Kementrian Pendidikan dan Kebudayaan.

1.5.2 Metode Pembangunan Perangkat Lunak

Teknik analisis data yang digunakan untuk implementasi pembangunan kamus jenis kata bahasa Indonesia menggunakan model *waterfall sommerville* [**3**]karena model ini melakukan pendekatan secara sistematis dan berurutan dalam pembangunan perangkat lunak meliputi proses sebagai berikut:



Gambar 1. 2 Diagram *Waterfall Sommervile* [3**]**

1. Analisis Kebutuhan Perangkat Lunak

   Pada tahap ini dilakukan pengumpulan kebutuhan penelitian secara lengkap tentang informasi sumber jenis kata bahasa Indonesia yang diambil dari Kamus Besar Bahasa Indonesia, Kamus Besar Bahasa Indonesia-Badan

Bahasa. Data yang diambil adalah kumpulan jenis kata Bahasa Indonesia dari Badan Bahasa dengan format *.txt*,

2. Desain

Setelah kebutuhan data dan pemroses telah dikumpulkan, maka tahap selanjutnya adalah perancangan arsitektur sistem secara keseluruhan. Untuk kebutuhan perangkat lunak dalam pengolahan data diperlukan software diantaranya penyunting kode sumber yang berjalan pada operasi sistem windows dalam hal ini peneliti menggunakan notepad++ dengan bahasa pemrograman PHP. Sumber data yang diolah berformat *.txt*, MySQL sebagai pengolah *database,* Sistem operasi yang digunakan penulis yaitu windows 7. Kebutuhan perangkat keras yang dibutuhkan mengolah data penulis menggunakan PC dual core, dengan hardisk 10GB sebagai ruang penyimpanan.

3. Generasi Kode

Pada tahap ini desain program yang telah dibuat kemudian diimplementasikan ke dalam bentuk kode bahasa pemrograman diantaranya data dengan format txt, selanjutnya adalah penghilangan simbol, digit, blank line, spasi kosong, dan sebagainya sehingga membentuk pola kata jenis kata dengan bahasa pemrograman PHP. Data yang telah membentuk pola selanjutnya dimasukan ke dalam *database* menggunakan *MySQL* sehingga siap digunakan untuk menjadi sebagai sumber daya jenis kata bahasa Indonesia

4. Pengujian

Tahap ini digunakan untuk program akan diintegrasikan satu sama lain dengan membuat kamus jenis kata dengan mengambil kata dan jenis kata dari *database* yang telah dibuat. Apabila program yang dibuat berjalan dengan sesuai harapan maka, sumber daya jenis kata bahasa Indonesia siap diedarkan dengan berbagai format.

5. Pemeliharaan

Dalam tahap ini, sumber daya kata dasar Bahasa Indonesia sudah siap digunakan. Selain itu juga memperbaiki *error* yang tidak ditemukan pada tahap pembuatan. Dalam tahap ini juga dilakukan pengembangan sistem seperti penambahan fitur dan fungsi baru yang mungkin muncul kemudian sesuai dengan kebutuhan pengguna.

1.6 Sistematika Penulisan

Sistematika dari penulisan tugas akhir dibagi ke dalam lima bab, dimana dalam tiap babnya terbagi menjadi beberapa bagian yang merupakan satu kesatuan dan saling terkait.

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, dan sistematika penulisan dari penelitian kamu jenis kata Bahasa Indonesia.

BAB II LANDASAN TEORI

Bab ini berisi tentang teori-teori pendukung dan teori-teori yang menjadi landasan dalam topik penelitian yang dilakukan, seperti Kamus Besar Bahasa Indonesia, metode yang digunakan, perangkat lunak, dan sebagainya.

BAB III ANALISIS DAN PERANCANGAN

Berisi tentang tahapan analisis sistem, analisis kebutuhan fungsional dan nonfungsional, serta perancangan sistem yang mencakup perancangan antar muka, alur data secara keseluruhan.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Berisi tentang tahapan yang dilakukan dalam penelitian secara garis besar dimulai dari tahap masukan data sampai pada keluaran yang dihasilkan, pada tahap ini sumber jenis kata bahasa Indonesia akan diuji apakah telah sesuai dengan kriteria tujuan pembangunan kamus jenis kata.

BAB V KESIMPULAN DAN SARAN

Pada bab ini membahas tentang kesimpulan dari penelitian dan saran yang dapat dijadikan masukan untuk pengembangan penelitian pada kasus yang sama selanjutnya.