Bab 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Question Generation(selanjutnya disingkat QG) adalah subsclass Natural Language Processing(selanjutnya disingkat NLP) yang memainkan peran penting dalam lingkup pembelajaran, pencarian informasi, serta dalam penerapan lainnya [1]. Penelitian terkait QG sudah banyak dilakukan, baik penerapannya dalam Berbahasa Inggris [1] [2] [3] [4] maupun Berbahasa Indonesia [5] [6] [7] [8] [9] [10] [11]. Pada dasarnya kata tanya atau interogatif sendiri dibagi menjadi dua; interogatif dengan jawaban Ya/Tidak dan juga interogatif dengan jawaban informatif [12]. Namun, adapula yang membagi interogatif menjadi tiga; Interogatif dengan jawaban Ya/Tidak, Interogatif dengan jawaban informatif, serta interogatif dengan jawaban retorik [12]. Interogatif sendiri dapat diakhiri dengan kata ingkaran seperti belum, bukan, dan tidak, disebut juga kalimat tanya embelan [13].

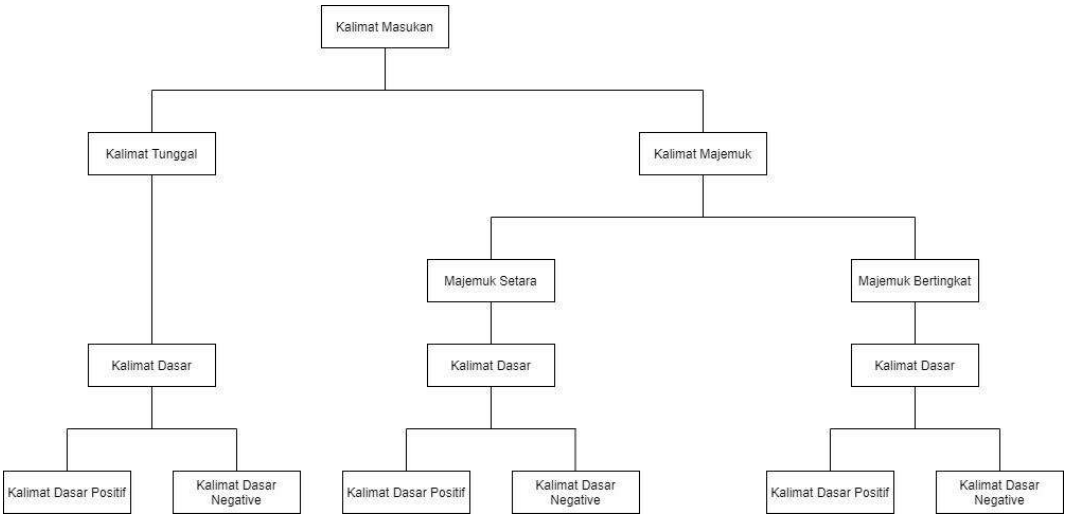
Berdasarkan penelitian sebelumnya, telah dilakukan penelitian pembangkit pertanyaan pada kalimat biasa [9] [10], kalimat majemuk [9], keterkaitan kalimat [11], serta pengembangan pada kalimat tanya non-factoid [8]. Pada penelitian pembangkit pertanyaan kalimat biasa dan majemuk diketahui menghasilkan akurasi 64,52%, pada penelitian pengembangan pembangkit pertanyaan pada kalimat tanya non-factoid sendiri menghasilkan nilai akurasi sebesar 68,84% [8], sedangkan pada penelitian keterkaitan kalimat diperoleh hasil 90,80%. Hal ini terjadi akibat kurang mendalamnya proses penelitian mengenai proses pengenalan entitas bernama(NER) dan jenis kata (POS), semantik antar kalimat yang saling berkaitan, aturan pada pengklasifikasian kata, aplikasi dan atau data penunjang seperti kamus [8] [9] [11]. Adapula identifikasi masalah yang dilakukan dengan metode observasi [5] [6] [7] [8] [9] [10] [11] serta keterkaitan dengan Studi Dokumen [12] [13] antara lain: kurang mendalamnya proses penelitian mengenai proses pengenalan entitas bernama(NER) dan jenis kata(POS), kurangnya pengklasifikasian jenis kalimat tanya positif dan negatif berdasarkan jenis kalimat dasar yang telah ditentukan, dibutuhkannya kalimat tanya yang merujuk pada kebeneran kalimat masukan, pada

predikat, serta pada kondisi subjek pada kalimat masukan, kurangnya pola frasa yang berakibat pada proses identifikasi klausa.

Berdasarkan masalah yang telah diketahui serta identifikasi masalah yang ada, ditentukan solusi dengan cara diskusi kelompok dengan pihak-pihak terkait. Hasil diskusi tersebut sebagai berikut:

1. Melakukan pengklasifikasian kalimat tanya ya/tidak, serta kalimat tanya klarifikasi/konfirmasi.
2. Penambahan kalimat tanya "Apa" yang merujuk pada perdikat.
3. Penambahan kalimat tanya "Bagaimana" yang merujuk pada kondisi subjek.
4. Melakukan perbaikan pada postag dengan menambahkan proses membandingkan kamus kata dasar dan Aplikasi Indonesia NER [14] [15].
5. Penambahan pola Frasa determinan.

Proses pengklasifikasian sendiri dapat dilihat pada Gambar 1.



Gambar 1.1 Klasifikasi Kalimat

1.2 Identifikasi Masalah

Berdasarkan masalah yang telah diketahui serta identifikasi masalah yang ada, ditentukan masalah sebagai berikut:

1. Kurang mendalamnya proses penelitian mengenai proses pengenalan entitas bernama(NER) dan jenis kata (POS).
2. Kurangnya pola frasa yang merujuk pada determinan.
3. Dibutuhkannya pertanyaan yang merujuk pada predikat.
4. Dibutuhkannya pertanyaan yang merujuk pada kondisi subjek.
5. Dibutuhkannya kalimat tanya untuk menanyakan kebenaran pada kalimat masukan, atau pertanyaan Ya/Tidak.

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah sebagai pengembangan pada penelitian pembangkit pertanyaan sebelumnya [8] [9] [10] [11] dengan tujuan meningkatkan nilai akurasi pada relevansi kalimat masukan dengan kalimat tanya dan mengembangkan kalimat-kalimat tanya yang akan dibangkitkan, sehingga dapat meningkatkan kuantitas pertanyaan yang dibangkitkan.

1.4 Batasan Masalah

Adapula batasan masalah pada penelitian ini yang dibagi menjadi tiga bagian, anta lain masukan, proses, dan keluaran.

1) Sistem yang dibuat dapat memproses dengan ketentuan masukan sebagai berikut.
    a. Data masukan berupa teks dengan format *.txt*.
    b. Teks masukan menggunakan Bahasa Indonesia baku dan mengikuti Ejaan Yang Disempurnakan(EYD).
    c. Teks masukan tidak mengandung simbol *emoticon*.
    d. Teks masukan diambil dari buku elektronik sekolah dari Kemendikbud RI.

2) Proses sistem yang dapat dilakukan dan tidak dapat dilakukan antara lain sebagai berikut.
   a. Kalimat majemuk yang dapat diproses adalah kalimat majemuk setara dan bertingkat.
   b. Kalimat majemuk diproses dengan mencari kata penghubung.
   c. Kata penghubung untuk kalimat majemuk dikelompokkan sesuai jenis hubungan.
   d. Kamus kata diambil dari penelitian sebelumnya.
   e. Kalimat dasar akan menentukan jenis kalimat negatif atau positif.
   f. Sistem tidak dapat mendeteksi kata-kata ambigu, salah penulisan, atau bahasa asing.
   g. Sistem tidak mendeteksi keterkaitan antar kalimat.
   h. Kalimat tanya yang akan dibangkitkan sesuai dengan proses identifikasi kalimat tanya nonfactoid.
   i. Kalimat tanya factoid dibagi menjadi dua bagian yaitu kalimat tanya factoid bukan ingkaran dan kalimat tanya factoid ingkaran

3) Sistem dapat menghasilkan keluaran sebagai berikut.
   a. Pertanyaan akan diklasifikasikan berdasarkan jenis-jenis kalimat tanya nonfactoid.
   b. Pertanyaan factoid akan membangkitkan pertanyaan bukan ingkaran dan juga pertanyaan ingkaran.
   c. Pertanyaan negatif dan positif akan tergantung pada jenis kalimat factoid.
   d. Pertanyaan akan dibangkitkan pada induk kalimat majemuk serta anak kalimat tunggal atau hanya pada induk kalimat tunggal.

1.5 Metodologi Penelitian

Metode dalam penelitian ini terdapat 6 tahap antara lain sebagai berikut.

1. Analisis masalah dan solusi pada sistem sebelumnya

   Dimana pada tahap ini akan dicari masalah serta solusi guna memperbaiki dan atau meningkatkan sistem yang akan dibuat pada penelitian ini.
2. Analisis tata bahasa Indonesia

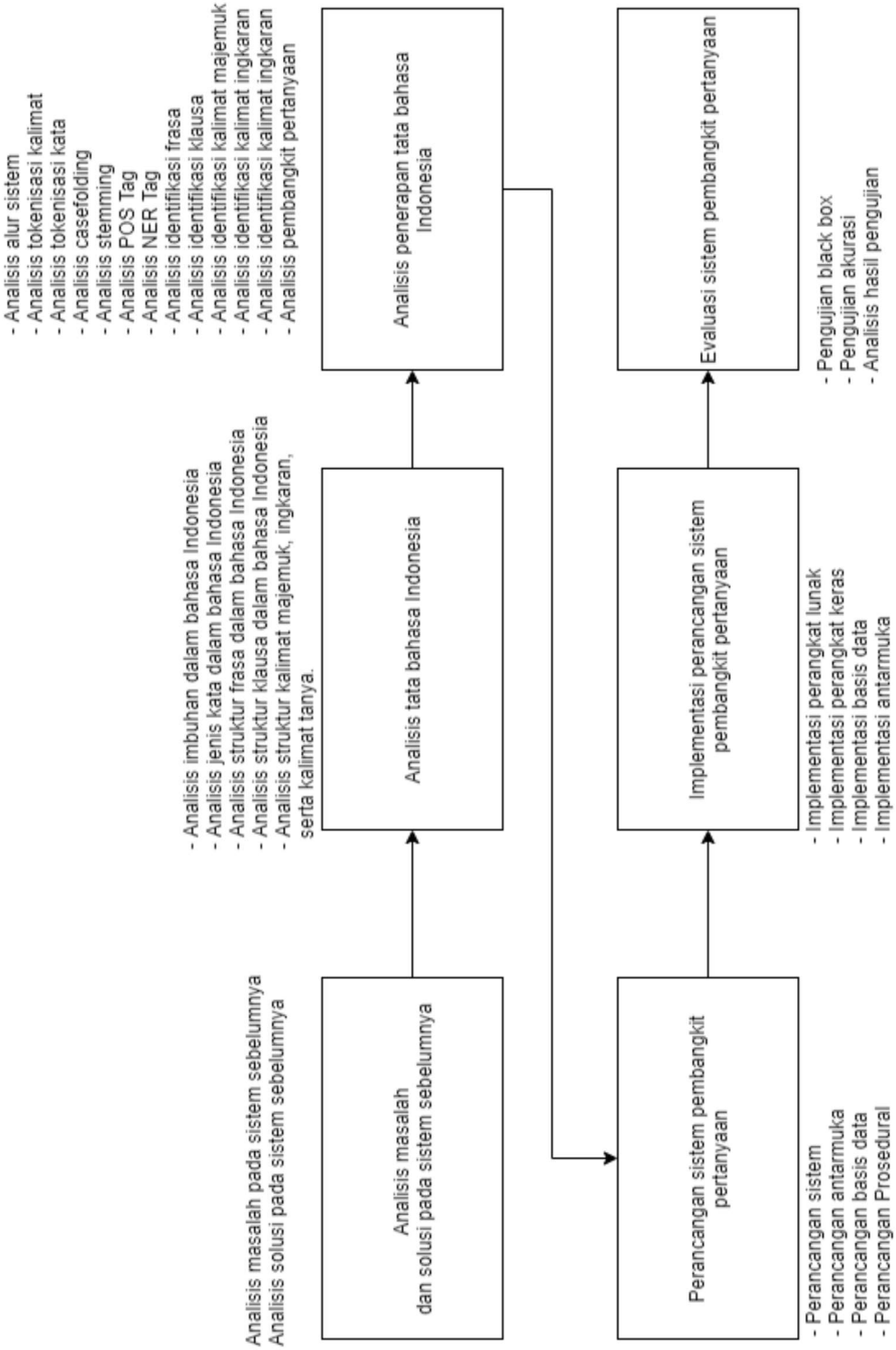
   Pada tahapan ini akan dilakukan analisis tata bahasa indonesia yang nantinya hasil analisis pada tahapan ini akan dipergunakan untuk diterapkan pada sistem.
3. Analisis penerapan tata bahasa Indonesia

   Pada tahap ini hasil analisis tata bahasa Indonesia akan dipetakan sesuai dengan kebutuhan sistem yang akan dibangun, serta akan dilakukan analisis alur proses dari setiap tahapan yang akan diterapkan pada sistem.
4. Perancangan sistem pembangkit pertanyaan

   Pada tahap ini dilakukan perancangan sistem dengan menggunakan metode UML, selain itu dilakukan juga proses perancangan prosedural yang digambarkan menggunakan flowchart.
5. Implementasi perancangan sistem

   Pada tahap ini dipaparkan mengenai hasil implementasi sistem yang telah dibangun pada penelitian ini.
6. Evaluasi sistem

   Terakhir adalah evaluasi sistem yang dilakukan adalah melakukan uji coba pada sistem secara black box, lalu perhitungan akurasi pada sistem yang telah dibangun guna penarikan kesimpulan dan saran.



Gambar 1.2 Metodologi Penelitian.

1.5.1 Metode Pengumpulan Data

Pada penelitian ini metode pengumpulan data dilakukan dengan menggunakan beberapa metode sebagai berikut.

1. Studi Literatur, metode ini dilakukan pada saat melakukan analisis pada penelitian yang telah dilakukan serta pengumpulan informasi dan analisis mengenai aturan tata bahasa pada buku-buku serta jurnal yang ada, serta.
2. Diskusi Kelompok, metode ini dilakukan guna membahasa kekurangan serta langkah-langkah yang perlu dilakukan dalam pengembangan penelitian pembangkit pertanyaan ini.
3. Wawancara Terstruktur, dilakukan guna melengkapi informasi yang kurang serta mengetahui kekurangan pada penelitian ini.

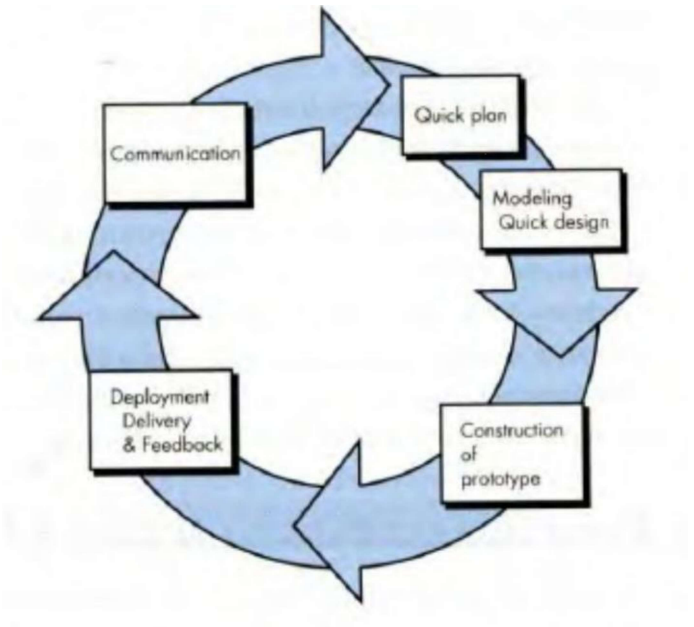
1.5.2 Metode Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak yang digunakan pada penelitian ini adalah metode *prototyping*. Berdasarkan Rani dan Anna, metode *prototyping* adalah suatu teknik dalam mengumpulkan informasi tertentu mengenai kebutuhan informasi pengguna secara cepat dan berfokus pada penyajian aspek-aspek perangkat lunak sehingga nampak bagi pelanggan atau pengguna [16]. Penelitian ini menggunakan metode *prototyping* guna mempermudah proses pengembangan, perbaikan, serta analisis kekurangan dan kelebihan yang dihasilkan dari penelitian.

Pada metode *prototyping* terdapat langkah-langkah yang perlu dilakukan, antara lain:

1. *Communication.*
2. *Quick Planning.*
3. *Quick Modeling.*
4. *Construction.*
5. *Deployment*.

Untuk lebih jelasnya langkah-langkah dapat dilihat pada Gambar 2.



Gambar 1.3 Model Pengembangan Perangkat Lunak *Prototyping*.

1.6 Sistematika Penulisan

Sistematika penulisan pada penelitian ini disusun dengan maksud memberikan gambaran umum mengenai penelitian pengembangan pembangkit pertanyaan yang dilakukan. Sistematika penulisan laporan tugas akhir ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Pada Bab 1 Pendahuluan akan diuraikan mengenai latar belakang masalah mengenai penelitian sebelumnya mengenai pembangkit pertanyaan yang secara khusus dalam Bahasa Indonesia. Melakukan identifikasi masalah pada penelitian sebelumnya, lalu menentukan maksud dan tujuan dari penelitian yang akan dilakukan terkait masalah yang telah diidentifikasi. Menjelaskan batasan-batasan masalah yang ada pada sistem yang akan dibangun. Menentukan metode-metode yang digunakan selama penelitian serta menjabarkan sistematika penulisan.

BAB 2 LANDASAN TEORI

Pada Bab 2 Landasan Teori akan memaparkan mengenai teori-teori serta aturan-aturan yang digunakan dan berkaitan dengan penelitian ini serta digunakan sebagai acuan untuk analisis dalam perancangan. Teori yang dimaksud adalah teori mengenai tata Bahasa Indonesia yang mencakup kata, kalimat, jenis kata, jenis kalimat, frasa, klausa. Selain itu akan dipaparkan juga mengenai *preprocessing* yang digunakan, pembangkit pertanyaan, UML, serta aplikasi-aplikasi yang digunakan.

BAB 3 ANALISIS DAN PERANCANGAN

Pada Bab 3 Analisis dan Perancangan akan menguraikan mengenai analisis masalah, analisis solusi, analisis basis data, analisis kebutuhan fungsional dan non-fungsional, analisis pengujian, serta akan memaparkan mengenai perancangan sistem yang akan diterapkan mulai dari UML, perancangan basis data, perancangan prosedural, perancangan tampilan antar muka, serta perancangan jaringan semantik.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Pada Bab 4 Implementasi dan Pengujian akan dipaparkan mengenai implementasi yang dilakukan berdasarkan perancangan yang telah ditentukan serta memaparkan pengujian sistem berdasarkan hasil implementasi. Pengujian dilakukan terhadap relevansi pertanyaan yang dibangkitkan dengan teks masukan yang diberikan.

BAB 5 KESIMPULAN DAN SARAN

Pada Bab 5 Kesimpulan dan Saran akan dipaparkan hasil kesimpulan dari penelitian pengembangan kalimat tanya dan ingkaran pada *template* pembangkit pertanyaan otomatis untuk teks berbahasa Indonesia berdasarkan hasil analisis sintaksis dan juga saran untuk penelitian selanjutnya mengenai bidang yang sama untuk melengkapi kekurangan dari penelitian yang telah dilakukan.