

Bab 1 Pendahuluan

1.1 Latar Belakang

Bahasa merupakan salah satu aspek penting dalam kehidupan manusia karena bahasa adalah alat untuk berkomunikasi dengan manusia lain. Bahasa dapat disajikan dalam bentuk lisan maupun tulisan. Penyajian bahasa dalam bentuk tulisan, dapat menyimpan pengetahuan dari generasi ke generasi berikutnya. Penyajian bahasa dalam bentuk tulisan dapat berupa tulisan buku, novel, majalah, artikel ilmiah, berita, dan lain sebagainya.

Berita dapat disajikan dalam 3 media, yaitu media elektronik seperti radio dan televisi, media cetak seperti koran dan majalah, serta media online seperti *kompas.com*, *detik.com*, dan lain-lain. Setiap penulisan karya tulis, termasuk berita, baik cetak maupun online, hendaknya ditulis dengan tata bahasa yang benar sesuai kaidah bahasa Indonesia, karena akan berpengaruh pada kesempurnaan proses penyampaian ide atau pesan. Seperti dikutip dari H. Rosihan Anwar dalam bukunya yang berjudul “Bahasa Jurnalistik Indonesia dan Komposisi”, mengatakan bahwa “bahasa jurnalistik didasarkan pada bahasa baku, tetap pada kaidah-kaidah tata bahasa, harus tetap memperhatikan ejaan yang benar”. Seorang ahli bahasa, David Wilkins, mengatakan bahwa “tanpa tata bahasa, sangat sedikit yang bisa disampaikan, dan tanpa kosakata, tidak ada yang bisa disampaikan”.

Akan tetapi, banyak ditemukan bahwa penulisan isi berita yang tidak sesuai dengan tata bahasa. Berdasarkan penelitian yang dilakukan oleh S.N. Wahyuni (2016), didapatkan sebanyak 46% kesalahan dari segi gramatikal pada penulisan berita pada rubrik *Citizen Journalism Tribun Bali*, kesalahan yang terjadi yaitu karena kesalahan ejaan, penggunaan kalimat tidak efektif, penggunaan kata tidak baku, dan kaidah tata bahasa normatif. Sedangkan pada surat kabar harian pagi *Postmetro Padang*, juga didapatkan beberapa kesalahan penulisan, salah satunya adalah kesalahan penyusunan kalimat (Fatmi Amsir, 2012).

Penyampaian berita melalui media online terus meningkat seiring dengan meningkatnya pengguna internet, sehingga berita online seperti *Tribunnews.com*, *detik.com*, dan lain-lain lebih banyak diminati oleh pembaca daripada media cetak. Berdasarkan Alexa per Januari 2017, peringkat 4 besar situs berita online di Indonesia yaitu secara terurut: *Tribunnews.com*, *Detik.com*, *Liputan6.com*, *Kompas.com*. Berdasarkan data tersebut, 4 besar situs berita online yang banyak

diminati oleh masyarakat Indonesia ini, apakah isi naskah beritanya sudah sesuai dengan kaidah atau tata bahasa Indonesia?

Mengingat pentingnya pemahaman tentang kalimat efektif pada penulisan naskah berita, maka perlu adanya kajian khusus tentang komputasi linguistik untuk pemeriksaan struktur kalimat dalam bahasa Indonesia. Dan metode yang digunakan adalah algoritma *LALR Parser* karena metode tersebut sesuai dengan karakteristik dari penulisan kalimat bahasa Indonesia yang memproses kata dari kiri-ke-kanan.

Salah satu bidang dalam dunia komputer yang mengkaji tentang interaksi bahasa manusia dengan komputer adalah *Natural Language Processing* (NLP). NLP atau pengolahan bahasa alami merupakan salah satu bidang dari ilmu komputer, *artificial intelligence*, dan komputasi linguistik yang berkaitan dengan interaksi antara komputer dan manusia melalui bahasa alami (R. Raman Nair & L. Sulochana Devi, 2013). Dalam memproses bahasa alami manusia, perlu adanya proses yang dapat melakukan analisis sintaksis terhadap struktur kalimat. Salah satu metode dalam menganalisis sintaksis kalimat adalah *parsing*.

Parsing adalah proses untuk memeriksa kebenaran struktur kalimat berdasarkan suatu *grammar* (tata bahasa) dan untuk memperoleh struktur sintaksisnya. Secara umum, *parsing* meliputi alur pemetaan kalimat terkait struktur sintaksisnya. (Magdum P. G. & Kodavade D. V., 2013). Salah satu algoritma untuk mem-*parsing* kata dalam *Context-Free Grammar* (CFG) adalah *LALR Parser*.

LALR Parser merupakan salah satu varian dari *LR Parser* yang mampu menangani *grammar* (Gupta, Shivani, & Yadav, Sunita, 2013). Alfred V. Aho dan Jeffery D. Ullman (1977) mengatakan bahwa *LALR Parser* atau *Look-ahead LR Parser* adalah versi sederhana dari *Canonical LR Parser* dan *LALR Parser* tidak melakukan *backtracking* (pengulangan pencarian), oleh karena itu metode tersebut cukup efisien untuk melakukan proses *parsing* dari kiri-ke-kanan dari sebuah inputan.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah:

1. Bagaimana merancang CFG (*Context-Free Grammar*) pada bahasa Indonesia berdasarkan jenis kalimatnya: kalimat tunggal, kalimat majemuk setara, dan kalimat majemuk bertingkat?
2. Bagaimana menerapkan, menguji, dan menghitung tingkat akurasi algoritma *LALR Parser* dalam mendeteksi struktur kalimat bahasa Indonesia?
3. Bagaimana menguji penulisan naskah berita online menggunakan algoritma *LALR Parser*?
4. Situs berita online manakah yang sesuai dengan tata bahasa Indonesia berdasarkan hasil pengujian menggunakan algoritma *LALR Parser*?

1.3 Batasan Masalah

Agar penelitian lebih terarah, maka perlu adanya batasan masalah, diantaranya:

1. Penelitian tidak membahas hal yang berkaitan dengan pendeteksian kesalahan penulisan dan pembenahan kata yang salah.
2. Dapat mendeteksi jenis kalimat: kalimat tunggal, kalimat majemuk setara, dan kalimat majemuk bertingkat.
3. Pola kalimat yang diperiksa berdasarkan unsur: Subyek, Predikat, Obyek, Pelengkap, dan Keterangan.
4. CFG (*Context-Free Grammar*) dibuat berdasarkan jenisnya, yaitu kalimat tunggal, kalimat majemuk setara, dan kalimat majemuk bertingkat. Terdapat 6 pola kalimat tunggal, 20 pola kalimat majemuk setara, dan 12 pola kalimat majemuk bertingkat.
5. Diujikan pada 3 jenis kalimat yang merujuk pada buku-buku bahasa Indonesia.
6. Menguji kesesuaian tata bahasa pada penulisan naskah berita online dari situs Tribunnews.Com, Detik.Com, Liputan6.Com, dan Kompas.Com.

1.4 Tujuan

Tujuan dari penelitian ini adalah:

1. Membuat CFG pada bahasa Indonesia berdasarkan jenis kalimatnya, yaitu kalimat tunggal, kalimat majemuk setara, dan kalimat majemuk bertingkat.
2. Menerapkan dan melakukan uji coba algoritma *LALR Parser* dalam mendeteksi struktur kalimat bahasa Indonesia.
3. Menghitung tingkat akurasi dan kinerja penggunaan algoritma *LALR Parser* dalam melakukan pendeteksian struktur kalimat bahasa Indonesia.
4. Menguji penulisan naskah berita online menggunakan algoritma *LALR Parser*.
5. Menganalisis kesesuaian tata bahasa Indonesia pada penulisan naskah berita online.

1.5 Manfaat

Manfaat penelitian ini adalah:

1. Memberikan pengetahuan mengenai bagaimana merancang *Context-Free Grammar* (CFG) pada kalimat bahasa Indonesia.
2. Memberikan pengetahuan tentang teknik *parsing* kata dalam kalimat bahasa Indonesia.
3. Memberikan pengetahuan tentang kinerja algoritma *LALR Parser* dalam mendeteksi pola kalimat bahasa Indonesia.
4. Memberikan informasi mengenai berita online mana yang tata bahasanya sesuai dengan kaidah bahasa Indonesia.

1.6 Metode Penelitian

Secara garis besar, tahapan dalam penyelesaian penelitian ini antara lain:

1. Studi Literatur

Melakukan studi pustaka dengan mencari literatur dan referensi tentang pola kalimat bahasa Indonesia, teknik mengurai kata dalam kalimat, dan algoritma *LALR Parser*.

2. Desain Sistem

Pada tahap ini, membuat rancangan sistem, mulai dari perancangan alur sistem secara keseluruhan, perancangan *lexicon* sebagai kamus kata bahasa Indonesia, yang kemudian dilanjutkan tahap perancangan alur sistem dengan menerapkan algoritma *LALR Parser* pada perangkat lunak. Keluaran dari desain sistem ini adalah berupa diagram dan *flowchart*.

3. Implementasi

Implementasi yang dilakukan adalah pembuatan perangkat lunak yang nantinya digunakan untuk pengujian kinerja algoritma *LALR Parser* dalam melakukan pendeteksian pola kalimat bahasa Indonesia.

4. Pengujian

Tahap pengujian ini dilakukan dengan memasukkan berbagai macam kalimat, baik itu kalimat tunggal, kalimat majemuk setara, maupun majemuk bertingkat untuk dilakukan pendeteksian pola kalimat yang terkandung dalam kalimat tersebut.

5. Analisa Hasil

Setelah dilakukan ujia coba untuk pendeteksian pola kalimat, maka pada tahap ini dilakukan analisis terhadap kinerja dan kesesuaian dari algoritma *LALR Parser* dalam melakukan pendeteksian pola kalimat bahasa Indonesia.

6. Pelaporan

Tahap pelaporan merupakan penulisan atau pendokumentasian penelitian dari tahap awal sampai dengan hasil dari penelitian.

1.7 Review Penelitian

Puji Priyono Triwibowo (2010) melakukan penelitian untuk menentukan jeda pada pengucapan kalimat bahasa Indonesia. Metode yang digunakan adalah *Lexer Parser*. Didapatkan hasil bahwa metode *Lexer Parser* yang diterapkan pada CFG (*Context-Free Grammar*) dapat digunakan sebagai penentu jeda pada pengucapan kalimat bahasa Indonesia berdasarkan struktur kalimatnya. Akan tetapi struktur kalimat yang diterapkan dengan metode tersebut hanya terbatas pada kalimat dasar atau kalimat tunggal.

Dalam pemeriksaan bahasa Indonesia terdapat sebuah tata bahasa yang digunakan sebagai standar untuk mengetahui bahwa bahasa yang digunakan tersebut sudah benar atau belum. Didalam

tata bahasa terdapat beberapa cara atau metode untuk menyusun sebuah kalimat yang benar. Cara sederhana yang digunakan yaitu dengan menentukan kalimat berdasarkan pola kalimat SPOK (Subyek-Predikat-Obyek-Keterangan). Salah satu algoritma yang dapat digunakan untuk memeriksa sebuah kalimat yaitu *Left-Corner Parsing*, yang merupakan gabungan dari algoritma *Top-Down Parsing* dan *Bottom-Up Parsing* (Vita Meriati Pandiangan, 2015). Akan tetapi pada penelitian yang dilakukan oleh Vita Meriati Pandiangan (2015) tersebut tidak mampu mendeteksi secara otomatis terhadap struktur kalimat bebas, hanya mampu membenarkan penulisan kata yang mengalami kesalahan morfologi (bentuk penulisan) dalam kalimat. Selain itu, struktur kalimatnya hanya terbatas pada SPOK atau bisa disebut sebagai kalimat dasar (Vita Meriati Pandiangan, 2015).

Salah satu algoritma untuk mem-*parsing* kata dalam *Context-Free Grammar* (CFG) adalah *LALR Parser*. *LALR Parser* merupakan salah satu varian dari *LR Parser* yang mampu menangani *grammar*. Metode *LALR Parser* digunakan untuk mengekstrak konteks dari sebuah link. Konteks dari link tersebut digunakan untuk menentukan halaman web yang sangat berguna untuk mengindeks dokumen. Shivani Gupta menggunakan *LALR Parser* karena kemampuannya yang lebih unggul dibandingkan varian LR yang lain yaitu SLR dan LR (Gupta, Shivani, & Yadav, Sunita, 2013).

Alfred V. Aho dan Jeffery D. Ullman (1977) mengatakan bahwa *LALR Parser* atau *Look-ahead LR Parser* adalah versi sederhana dari *Canonical LR Parser* dan *LALR Parser* tidak melakukan *backtracking* (pengulangan pencarian), sehingga metode tersebut cukup efisien untuk melakukan proses *parsing* dari kiri-ke-kanan dari sebuah masukan. Sehingga metode tersebut sesuai dengan karakteristik penulisan kalimat bahasa Indonesia yang melakukan proses dari kiri-ke-kanan.

Dalam penerapan algoritma *LALR Parser* membutuhkan *generator*, dan memang pada penelitian-penelitian sebelumnya, semua penerapan *LALR* menggunakan *generator* tertentu. Pada penelitian Shivani Gupta (2013) menggunakan *Bison Parser* yang merupakan *generator parsing* yang umum digunakan untuk mengubah deskripsi *grammar* pada *LALR* ke dalam program C. *Bison Parser* membantu untuk mengubah aturan produksi ke dalam *state* (Gupta, Shivani, & Yadav, Sunita, 2013). Terdapat *generator parser* seperti *Yacc*, *Bison*, dan *Cup* yang merupakan jenis *generator LR Parser* yang digunakan untuk *grammar* pada *LALR* (AbdelLatif, A. M., Kamel, A., & Bahgat, R., 2015). Untuk program *Java*, terdapat *Cup Parser Generator*, yaitu sistem untuk menghasilkan *LALR Parser* dari spesifikasi yang sederhana. *Cup* ditulis di *Java* yang menggunakan spesifikasi yang tertanam pada *Java*, serta menghasilkan *parser* yang diimplementasikan di *Java* (Hudson, Scott E., 1999).

Dari review penelitian yang sudah diuraikan diatas, dapat ditarik kesimpulan yang ditunjukkan pada tabel 1.1 berikut ini.

Tabel 1.1 Review Penelitian

Tahun	Peneliti	Objek dan Metode	Keterangan
2010	Puji Priyono Triwibowo	Penentu jeda pada pengucapan kalimat bahasa Indonesia menggunakan metode <i>Lexer Parser</i>	Penelitian hanya terbatas pada kalimat tunggal atau kalimat dasar
2015	Vita Meriati Pandiangan	Pemeriksaan tata bahasa pada kalimat bahasa Indonesia menggunakan metode <i>Left-Corner Parsing</i>	Hanya mampu membenarkan kesalahan morfologi penulisan kata dalam kalimat. Dan struktur kalimatnya hanya terbatas pada kalimat dasar.
2013	Shivani Gupta & Yadav Sunita	Ekstraksi <i>link</i> dari halaman web menggunakan metode <i>LALR Parser</i>	Penerapannya pada konteks dari sebuah link web, bukan pada bahasa.
1977	Alfred V. Aho & Jeffery D. Ullman	<i>Principals of Compiler Design</i>	Mengatakan bahwa <i>LALR Parser</i> merupakan metode yang efisien untuk melakukan parsing dari kiri-ke-kanan sebuah masukan. Sehingga sangat cocok untuk kalimat bahasa Indonesia.
1999	Scott E. Hudson	<i>CUP Parser Generator for Java</i>	Untuk program <i>Java</i> , terdapat <i>Cup Parser Generator</i> , yaitu sistem untuk menghasilkan <i>LALR Parser</i> dari spesifikasi yang sederhana.
<p>Dengan merujuk dari penelitian sebelumnya dan teori yang sudah ada, maka pada penelitian ini objek yang diambil adalah kalimat bahasa Indonesia yang mampu menangani semua jenis kalimat, baik kalimat tunggal, kalimat majemuk setara, maupun kalimat majemuk bertingkat dengan menggunakan algoritma <i>LALR Parser</i>. Dalam penerapan metode <i>LALR Parser</i>, <i>generator</i> yang digunakan adalah <i>Cup</i> karena generator tersebut cocok untuk bahasa pemrograman <i>Java</i>.</p>			

1.8 Sistematika Penulisan

Sistematika penulisan dalam penyusunan penelitian ini terdiri dari beberapa bab, yaitu:

- 1) BAB I Pendahuluan, pada bab ini berisi tentang latar belakang, permasalahan, batasan masalah, tujuan, manfaat, metode penelitian, dan review penelitian.
- 2) BAB II Landasan Teori, pada bab ini berisi tinjauan kepustakaan terhadap teori-teori yang relevan dan terkait dari subyek permasalahan dan penyusunan penelitian, yaitu teori tentang: Kalimat, *Natural Language Processing* (NLP), *Grammar*, *Context-Free Grammar* (CFG), *LALR Parser*.
- 3) BAB III Metodologi Penelitian, pada bab ini berisi metode penelitian yang dilakukan, data yang diperlukan, teknik pengumpulan data, dan rancangan terhadap sistem yang akan dibangun.

- 4) BAB IV Hasil dan Pembahasan, pada bab ini berisi tentang hasil realisasi dari rancangan dan memuat uraian ketercapaian tujuan serta hasil analisis terhadap permasalahan yang telah diuji.
- 5) BAB V Kesimpulan dan Saran, pada bab ini memuat rangkuman dari seluruh hasil penelitian dan memuat saran untuk penelitian lebih lanjut.
- 6) Daftar Pustaka, berisi daftar referensi yang digunakan sebagai penunjang dalam menyelesaikan penelitian.

