

Lembar Pengesahan Pembimbing

Penerapan Algoritma *LALR Parser* Untuk Mendeteksi Struktur Kalimat Bahasa Indonesia



Nama: Mudafiq Riyan Pratama

NIM: 11917222

Yogyakarta, Januari 2017

Pembimbing I

Dr. Sri Kusumadewi, S.Si., M.T

Pembimbing II

Taufiq Hidayat, ST., MCS

**LEMBAR PENGESAHAN PENGUJI**

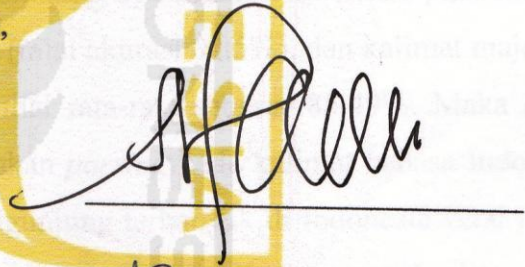
**PENERAPAN ALGORITMA *LALR PARSER* UNTUK MENDETEKSI STRUKTUR  
KALIMAT BAHASA INDONESIA**

Nama: Mudafiq Riyan Pratama  
NIM: 11917222

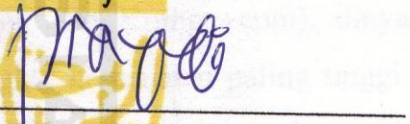
Yogyakarta, Januari 2017

Tim Penguji,

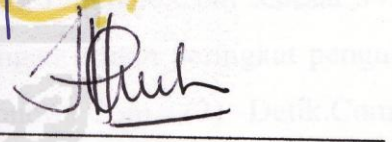
Dr. Sri Kusumadewi, S.Si., M.T  
Ketua



Yudi Prayudi, S.Si., M.Kom  
Anggota I



Taufiq Hidayat, ST., MCS  
Anggota II



Mengetahui,

Ketua Program Pascasarjana Fakultas Teknologi Industri

Universitas Islam Indonesia



  
R. Teduh Dirgahayu, S.T., M.Sc., Ph.D

## **Abstrak**

Penyajian bahasa dalam bentuk tulisan, dapat menyimpan pengetahuan dari generasi ke generasi berikutnya. Salah satu bentuk penyajian bahasa dalam bentuk tulisan adalah berita. Setiap penulisan karya tulis, termasuk berita, hendaknya ditulis dengan tata bahasa yang benar sesuai kaidah bahasa Indonesia, karena akan berpengaruh pada kesempurnaan proses penyampaian ide atau pesan. Pembaca berita melalui media online terus meningkat seiring meningkatnya pengguna internet, sehingga berita online lebih banyak diminati daripada media cetak. Permasalahan yang timbul apakah isi naskah berita online sudah sesuai dengan kaidah tata bahasa Indonesia atau justru sebaliknya. Maka untuk mengujinya dibutuhkan sistem yang dapat mendeteksi kesesuaian tata bahasa. *Parsing* adalah proses untuk memeriksa kebenaran struktur kalimat berdasarkan suatu *grammar* (tata bahasa). Salah satu algoritma untuk parsing kata dalam *Context-Free Grammar* (CFG) adalah *LALR Parser*. *LALR Parser* merupakan salah satu varian dari *LR Parser* yang mampu menangani *grammar*. Pengujian dilakukan pada tiga jenis kalimat yaitu kalimat tunggal, kalimat majemuk setara, dan kalimat majemuk bertingkat untuk mendapatkan tingkat akurasi. Hasilnya adalah *LALR Parser* efektif dalam melakukan parsing terhadap CFG bahasa Indonesia pada kalimat tunggal (nilai akurasi 88.98%), kalimat majemuk setara (nilai akurasi 82.5%), dan kalimat majemuk bertingkat (nilai akurasi 85%), sehingga didapatkan nilai rata-rata sebesar 85.49%. Maka dapat dikatakan bahwa algoritma *LALR Parser* dapat melakukan *parsing* pada kalimat bahasa Indonesia dengan baik dan efektif. Dari 4 situs berita dengan pengunjung terbanyak di Indonesia versi Alexa per Januari 2017 (Tribunnews.com, Detik.com, Liputan6.com, dan Kompas.com), dinyatakan bahwa penulisan naskah berita Tribunnews.Com memiliki tingkat kesesuaian paling tinggi yaitu sebesar 69.31%, kemudian diikuti oleh Detik.Com sebesar 65.08%, Liputan6.Com sebesar 54.31%, dan Kompas.Com sebesar 48.51%. Hal ini berbanding lurus dengan urutan peringkat pengunjung terbanyak versi Alexa per Januari 2017, yaitu (1) Tribunnews.Com, (2) Detik.Com, (3) Liputan6.Com, dan (4) Kompas.Com. Dalam pengujian, sistem mendeteksi salah terhadap kalimat dikarenakan beberapa penyebab, yaitu: (1) kalimat yang diuji memang memiliki struktur kalimat yang tidak sesuai dengan kaidah bahasa Indonesia, (2) unsur pembentuk kalimat tidak tersedia dalam CFG, (3) kata yang terdapat dalam kalimat tidak tersedia dalam *lexicon*.

## **Kata Kunci**

kalimat, bahasa Indonesia, *parsing*, *lalr parser*, *context-free grammar*, *lexicon*, kalimat tunggal, kalimat majemuk setara, kalimat majemuk bertingkat

## **Abstract**

Presentation of language in written form, may keep knowledge from one generation to the next. One of presentation of language in written form is news. Each writing papers, including news, should be written with proper grammar according to the rules of Indonesian, because it will affect the process of delivering the perfection of an idea or message. News readers through online media continues to increase with increasing internet users, making online news more desirable than the print media. The problem that occurs is whether the content of online news is in conformity with the rules of Indonesian ? then required to test the suitability of the system that can detect grammar. Parsing is the process to check the validity of the sentence structure based on a grammar. One algorithm for parsing words in Context-Free Grammar (CFG) is LALR Parser. LALR Parser is one variant of the LR parser capable of handling grammar. Tests performed on three types of sentences that are single sentence, similar compound sentence, and terraced compound sentences to obtain the level of accuracy. The result is a LALR Parser effective in parsing the CFG Indonesian on a single sentence (value accuracy of 88.98%), similar compound sentences (accuracy value of 82.5%), and terraced compound sentences (accuracy value 85%), so the average value of accuracy is 85.49%. So algorithm LALR Parser to parse sentences Indonesian properly and effectively. From 4 news sites with the most visitors in the Indonesian version of Alexa per January 2017 (Tribunnews.com, Detik.com, Liputan6.com, and Kompas.com), stated that the news script written in Tribunnews.Com have the highest level of conformity that is equal to 69.31% , followed by Detik.Com amounted to 65.08%, Liputan6.Com amounted to 54.31%, and Kompas.Com amounted to 48.51%. It is directly proportional to rank order the most visitors per January 2017 version of Alexa, that are: (1) Tribunnews.Com, (2) Detik.Com, (3) Liputan6.Com, and (4) Kompas.Com. In testing, the system detects a wrong of the sentence is due to several causes, namely: (1) the sentence being tested does have the structure of a sentence that does not comply with the rules of Indonesian, (2) the elements forming the sentence is not available in the CFG, (3) the words contained in sentences are not available in the lexicon.

## **Keywords**

sentence, Indonesian, parsing, parser LALR, context-free grammar, lexicon, single sentence, similar compound sentence, terraced compound sentence

## **Pernyataan Keaslian Tulisan**

Dengan ini saya menyatakan bahwa tesis ini merupakan tulisan asli dari penulis, dan tidak berisi material yang telah diterbitkan sebelumnya atau tulisan dari penulis lain terkecuali referensi atas material tersebut telah disebutkan dalam tesis. Apabila ada kontribusi dari penulis lain dalam tesis ini, maka penulis lain tersebut secara eksplisit telah disebutkan dalam tesis ini.

Dengan ini saya juga menyatakan bahwa segala kontribusi dari pihak lain terhadap tesis ini, termasuk bantuan analisis statistik, desain survei, analisis data, prosedur teknis yang bersifat signifikan, dan segala bentuk aktivitas penelitian yang dipergunakan atau dilaporkan dalam tesis ini telah secara eksplisit disebutkan dalam tesis ini.

Segala bentuk hak cipta yang terdapat dalam material dokumen tesis ini berada dalam kepemilikan pemilik hak cipta masing-masing. Untuk material yang membutuhkan izin, saya juga telah mendapatkan izin dari pemilik hak cipta untuk menggunakan material tersebut dalam tesis ini.

Yogyakarta, Januari 2017



Mudafiq Riyan Pratama, S.Kom

**Publikasi Selama Masa Studi**

**Publikasi yang menjadi bagian dari tesis**

“Tidak ada publikasi yang menjadi bagian dari tesis”



**Kontribusi Yang Diberikan Oleh Pihak Lain Dalam Tesis Ini**

**“Tidak ada kontribusi dari pihak lain”**

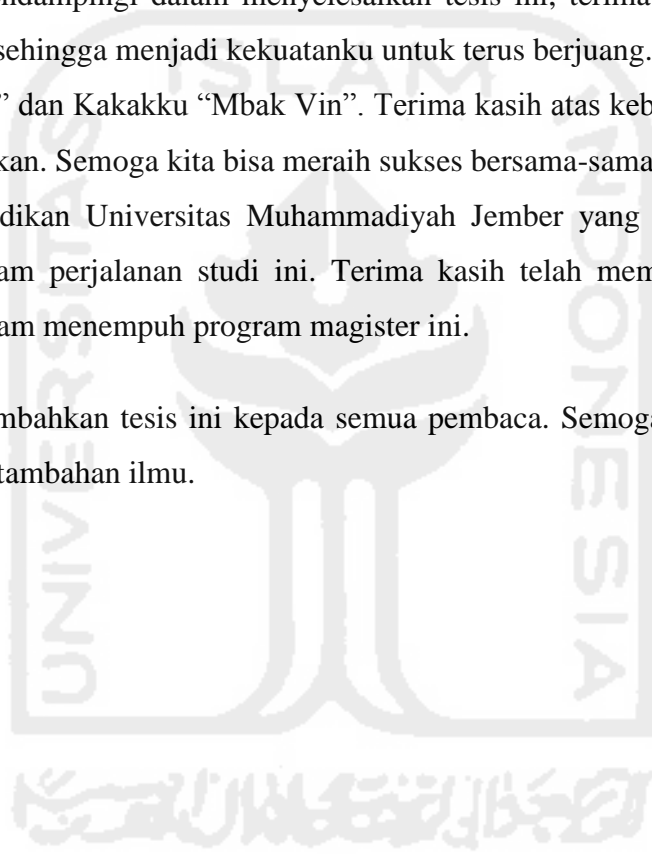


## **Halaman Persembahan**

Kupersembahkan karyaku ini kepada orang-orang yang selalu memberi dukungan dan semangat untuk menyelesaikan tesis ini, yaitu:

1. Kedua orang tuaku tercinta, “Nurkholis” dan “Susriyanti” dan kedua mertuaku. Doa tulus kepada anakmu ini seperti air yang terus mengalir. Terima kasih atas pengorbanan dan motivasi yang selalu diberikan selama ini.
2. Istriku tercinta Arsy Rachma Yuanita dan anakku tercinta Azarine Dzikra Faranisa. Merekalah yang menjadi semangatku hingga dapat menyelesaikan tesis ini. Terima kasih sudah sabar mendampingi dalam menyelesaikan tesis ini, terima kasih doa dan semangat yang diberikan sehingga menjadi kekuatanku untuk terus berjuang.
3. Adikku “Gygas” dan Kakakku “Mbak Vin”. Terima kasih atas kebersamaan, dukungan, dan doa yang diberikan. Semoga kita bisa meraih sukses bersama-sama.
4. Lembaga pendidikan Universitas Muhammadiyah Jember merupakan hal yang tak terpisahkan dalam perjalanan studi ini. Terima kasih telah memberikan kesempatan dan pembiayaan dalam menempuh program magister ini.

Akhir kata, saya persembahkan tesis ini kepada semua pembaca. Semoga tesis ini dapat memberi manfaat, inspirasi, dan tambahan ilmu.





## Kata Pengantar

*Alhamdulillah*, segala puji syukur kehadiran Allah *subhanahu wa ta'ala* atas segala nikmat, rahmat, dan hidayah yang begitu besar sehingga penulis dapat menyelesaikan tesis dengan judul “Penerapan Algoritma *LALR Parser* Untuk Mendeteksi Struktur Kalimat Bahasa Indonesia”, yang merupakan salah satu syarat kelulusan program Magister Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia.

Laporan tesis ini membahas tentang kinerja algoritma *LALR Parser* dalam mendeteksi struktur kalimat bahasa Indonesia, baik itu kalimat tunggal, kalimat majemuk setara, maupun kalimat majemuk bertingkat. Banyak tahapan-tahapan yang ditempuh dalam penelitian ini, salah satunya adalah tahapan perancangan. Perancangan yang dilakukan antara lain perancangan pada *lexikon* sebagai kamus kata, perancangan CFG (*Context-Free Grammar*) sebagai inti dari pendeteksian struktur kalimat yang didalamnya terdapat unsur-unsur pembentuk kalimat, serta perancangan dari algoritma *LALR Parser* yang sebagai metode yang melakukan *parsing* kalimat terhadap CFG. Penelitian ini selain menguji tingkat akurasi dari tiga jenis kalimat (kalimat tunggal, kalimat majemuk setara, dan kalimat majemuk bertingkat), juga menguji kesesuaian naskah berita terhadap tata bahasa Indonesia menggunakan algoritma *LALR Parser*, dan diujikan pada 4 besar situs berita online di Indonesia menurut versi Alexa per Januari 2017, yaitu Tribunnews.com, Detik.com, Liputan.com, dan Kompas.com.

Dalam penyelesaian tesis ini, penulis mendapatkan banyak bantuan dari berbagai pihak, baik langsung ataupun tak langsung. Atas berbagai bantuan itu, penulis ucapkan terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Sri Kusumadewi, S.Si., M.T dan Bapak Taufiq Hidayat, ST., MCS selaku dosen pembimbing yang telah memberikan semangat, ilmu, dan pikirannya dalam penyelesaian tesis ini.
2. Bapak Yudi Prayudi, S.Si., M.Kom selaku dosen penguji yang banyak memberikan masukan-masukan terkait dengan penelitian yang dikerjakan ini.
3. Istri, Ayah, Bunda, Mertua, Kakak, Adik, serta keluarga besar penulis yang selalu memberikan doa dan motivasi yang tak henti-hentinya.
4. Universitas Muhammadiyah Jember yang telah memberikan kesempatan kepada penulis untuk melanjutkan studi strata 2 di Magister Teknik Informatika Universitas Islam Indonesia.
5. Semua teman-teman seperjuangan dari Universitas Muhammadiyah Jember yang selalu membantu, memberikan semangat, motivasi, dan doa hingga tesis ini terselesaikan.
6. Semua pihak yang telah memberikan dukungan terhadap penyelesaian tesis ini yang tidak bisa disebutkan satu-persatu.

Penulis menyadari bahwa dalam penulisan tesis ini masih banyak kekurangan dan jauh dari sempurna. Oleh karena itu, penulis siap menerima setiap masukan dan kritik yang diberikan demi penyempurnaan penelitian ini. Penulis juga menyadari bahwa tesis ini mungkin tidak mencerminkan kualitas seorang penyandang gelar magister. Namun, setidaknya semua proses telah dilalui dengan penuh kesungguhan. Akhir kata, semoga ada inspirasi dan manfaat yang bisa ditarik hikmahnya dalam karya sederhana ini.

Yogyakarta, Januari 2017

Penulis

