



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Mengarang merupakan kegiatan ekspresif yang menggunakan bahasa. Kegiatan ini didukung oleh kemampuan seseorang dalam menggunakan bahasa itu sendiri, yang mencakup tata bahasa, pemahaman bahasa dan perbendaharaan kata (Kumara & Andayani., 1997). Hal ini diperkuat dengan penelitian oleh Basuki & Suryani, (2017) dalam jurnalnya yang berjudul Kesulitan Belajar Siswa dalam Pembelajaran Bahasa Indonesia, kemampuan berbahasa Indonesia berarti siswa terampil menggunakan bahasa Indonesia sebagai alat komunikasi. Terampil berbahasa berarti terampil menyimak, berbicara, membaca dan menulis dalam bahasa Indonesia. Namun, kenyataannya siswa mengalami kesulitan pada kompetensi kebahasaan subkompetensi struktur sebanyak 61,07% dan 52,6% pada subkompetensi kosakata. Selain itu, pada kompetensi keterampilan berbahasa siswa mengalami kesulitan belajar pada subkompetensi menulis sebesar 20.02%.

Grammar checker merupakan salah satu teknologi yang digunakan untuk pemeriksaan tata bahasa yang dapat memberikan saran pengoreksian langsung dalam memperbaiki kalimat yang salah menurut tata bahasa. Tujuan utama dari *grammar checker* adalah membuat tulisan dokumen terbebas dari kesalahan tata bahasa (Konchady, 2009).

Penelitian sebelumnya yang dilakukan oleh Syaputra dkk., (2018) pada jurnalnya yang berjudul Perangkat Lunak Identifikasi dan Koreksi Kesalahan

Berbahasa Indonesia pada Karya Ilmiah, membahas tentang pembuatan perangkat lunak yang dapat mendeteksi sekaligus mengoreksi kesalahan ejaan yang disempurnakan (EyD) menggunakan dua algoritma, yaitu algoritma Nazief & Andriani pada proses *stamming* atau mencari *root* dari kata dan algoritma *Lavenshtein distance* yang digunakan dalam mencari pendekatan kata yang sesuai apabila terdapat kesalahan dalam penulisan.

Penelitian serupa juga dilakukan oleh Hariatsyah (2013) memiliki kendala dalam memperbaiki kalimat. Algoritma *Deep Parsing* hanya dapat memisahkan kata berdasarkan spesifikasi tertentu, tidak dapat memisahkan kata secara logika seperti yang dibutuhkan pada imbuhan sisipan ataupun imbuhan akhiran, sehingga program tersebut tidak dapat memperbaiki kalimat majemuk, aktif dan pasif.

Penelitian oleh Alam dkk., (2007) pada jurnalnya yang berjudul *N-gram based Statistical Grammar Checker for Bangla and English*, penggunaan POS (*Part of Speech*) *Tagging* pada penelitannya menghasilkan performa lebih rendah karena banyak kata saling berkaitan satu sama lain, sehingga menghasilkan keluaran kata yang saling tertukar dan salah secara tata bahasa. Hal tersebut dapat diatasi dengan penggunaan *n-gram* dalam menentukan kata mana yang lebih mungkin dihasilkan pada keluaran hasil uji coba. Performa dari pemeriksaan tata bahasa tersebut bergantung pada jumlah *n* dari model *n-gram* yang digunakan untuk setiap panjang rata-rata kalimat dalam bahasa yang digunakan.

Menurut Ariana dkk., (2014), *n-gram* merupakan sebuah metode yang diaplikasikan untuk membangkitkan kata atau karakter. Salah satu keunggulan menggunakan *n-gram* dan bukan suatu kata utuh secara keseluruhan dikarenakan *n-gram* tidak terlalu sensitif terhadap kesalahan penulisan yang terdapat pada suatu

dokumen. Model *n-gram* memiliki keunggulan dalam mengidentifikasi bahasa suatu dokumen yang sifatnya resistan pada berbagai macam kesalahan dalam penulisan, sehingga kesalahan pada sebagian *string* hanya berakibat perbedaan pada *n-gram* (Zaman dkk., 2015). *N-gram* memberikan prediksi kemunculan kata selanjutnya yang lebih akurat berdasarkan probabilitas riwayat kata yang dimasukkan sebelumnya. Probabilitas kemunculan kata *n-gram* dapat digunakan pada aplikasi penulisan seperti *Spelling Correction* dan *Grammatical Error Correction* dengan mencari kesalahan penulisan yang dapat memudahkan pengguna untuk mendeteksi dan mengoreksi kesalahan tersebut (Daniel dkk., 2019).

Bahasa Indonesia merupakan bahasa persatuan yang sudah ada sejak 1928 dengan kerapatan dan ikrar oleh para pemuda seluruh Indonesia yang dikenal dengan Sumpah Pemuda. Penelitian ini akan menguji dengan jumlah *n* sebanyak 4 (*Quadgram*) pada pemeriksaan tata bahasa. Bentuk *quadgram* dipilih karena memiliki angka akurasi 96% pada penelitian yang dilakukan oleh Hanami., dkk (2015) dalam memberi harakat pada kalimat Bahasa Arab. Selain itu, *quadgram* juga dinilai memiliki tingkat efektifitas yang lebih tinggi dibandingkan nilai *n* lainnya seperti *bigram* dan *trigram* karena memiliki nilai *false positif* yang lebih sedikit serta akurasi yang paling besar (Vernanda, 2018). Hasil dari keluaran akan memberitahu letak kesalahan dalam penulisan dan memberikan saran perbaikan yang berguna dalam pembelajaran siswa. Keluaran ini diharapkan dapat membantu siswa yang memiliki kesulitan dalam menulis.

Berdasarkan latar belakang tersebut, maka dibuatlah judul tugas akhir “Implementasi N-gram Model pada Aplikasi Grammar Checker Bahasa Indonesia”.

Metode akan diimplementasikan menggunakan *framework CodeIgniter* serta berbasis *website*.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, dirumuskan masalah pada penelitian ini, yaitu bagaimana cara mengimplementasikan *n-gram* model pada aplikasi *grammar checker* bahasa Indonesia?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Aplikasi hanya melakukan pemeriksaan terhadap Bahasa Indonesia.
2. Kata pada kalimat yang diperiksa minimal berjumlah 4 kata.
3. Tanda baca titik (.), koma (,), tanda seru (!), dan tanda tanya (?) akan dihilangkan selama proses pemeriksaan berlangsung.
4. Aplikasi hanya memberitahu letak kesalahan pada kata dalam kalimat dan memberikan saran perbaikan pada kata yang salah tersebut.
5. Data kalimat yang digunakan untuk *training corpus* didapatkan dari beberapa buku sekolah dasar yang diambil dari buku sekolah elektronik sekolah dasar dalam situs buku.kemendikbud.go.id dan annibuku.com.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah mengimplementasikan *n-gram* model pada aplikasi *grammar checker* bahasa Indonesia.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah membantu proses belajar bahasa Indonesia lebih mudah bagi pelajar dengan mengetahui kesalahan dan memberikan saran perbaikan tata bahasa dengan menggunakan aplikasi ini.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penulisan ini adalah sebagai berikut.

1. BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah yang memuat hal-hal yang mengantarkan pada pokok permasalahan, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

2. BAB II LANDASAN TEORI

Bab ini berisi uraian teori terkait dengan penelitian yang dilakukan, yakni mengenai metode *n-gram*, *grammar checker*, dan bahasa Indonesia.

3. BAB III METODOLOGI DAN PERANCANGAN SISTEM

Bab ini berisi tentang metode penelitian dan perancangan sistem yang terdiri dari *unified modelling language* (*use case diagram*, *activity diagram*, *sequence diagram*, dan *class diagram*) dan rancangan antarmuka dari aplikasi.

4. BAB IV IMPEMENTASI DAN PENGUJIAN

Bab ini berisi impelementasi sistem, hasil tangkapan layar aplikasi yang dibuat, uji coba, dan hasil uji coba.

5. BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan dari penelitian yang telah dilakukan dan

saran yang diberikan untuk penelitian selanjutnya berdasarkan penelitian yang telah dilakukan.