**Latar belakang**

* Menulis sebagai alat komunikasi dan informasi telah dikenal sejak lama. Pada masa itu, manusia hanya dapat berkomunikasi melalui tulisan gambar dan simbol pada benda-benda alam seperti batu, kayu, dan dinding gua.
* Awal memasuki tahun masehi, manusia baru mulai menemukan kertas. Ketika teknologi semakin maju kegiatan tulis menulis telah beralih ke media digital seperti menggunakan komputer, smartphone, dan internet, sehingga menggantikan penggunaan kertas dalam kegiatan menulis.
* Peralihan media tulis menulis ke media digital memberikan keuntungan dalam efisiensi dan aksesibilitas informasi. Meskipun kemajuan teknologi telah memudahkan kegiatan menulis, tidak terhindarkan bahwa kesalahan penulisan seperti typo masih sering terjadi. Salah satu faktor penyebab kesalahan ejaan pada tulisan ini, yaitu ketidaktepatan pengetikan atau kurangnya pengetahuan tentang bahasa baku.
* Microsoft Office Word sebagai aplikasi pengolah kata yang sering digunakan, umumnya menyediakan fitur pengecekan ejaan yang dapat mengidentifikasi kesalahan ejaan dalam teks. Tetapi Microsoft Office Word masih terbatas dalam kemampuannya untuk mendeteksi kesalahan pada istilah-istilah khusus atau terminologi di bidang tertentu seperti di bidang komputer dan informatika. Dalam penulisan jurnal ilmiah di mana ketelitian diperlukan, harus diartikan bahwa terminologi atau istilah yang digunakan telah dikejar dengan benar. Proses pengecekan ejaan dengan cara manual ini dapat menjadi tugas yang memakan waktu dan berisiko terlewatnya beberapa kesalahan.
* Algoritma damerau levenshtein distance telah banyak digunakan misalnya dalam penelitian yang dilakukan oleh Wibawa (2020) tentang DLD dengan pendekatan empiris untuk koreksi ejaan bahasa indonesia. Dalam penelitiannya algoritma Damerau Levenshtein Distance dapat diterapkan untuk sistem koreksi ejaan. Namun, algoritma ini masih memiliki keterbatasan dalam mengatasi kesalahan yang disebabkan oleh hilangnya spasi atau kata yang saling berdempetan.
* Oleh karena itu maka diperlukan suatu algoritma lain yang dapat mengidentifikasi kata yang berdempetan tersebut. Salah satu algoritma yang dapat digunakan adalah algoritma boyer moore. Algoritma ini merupakan metode pencarian string yang efisien di mana melakukan pencarian kata terhadap suatu teks yang dimulai dari arah kanan ke kiri. Dan memiliki loncatan pergeseran karakter yang besar.

**Damerau levenshtein distance**

* Algoritma Damerau-Levenshtein Distance adalah metode untuk mengukur seberapa berbeda dua string berdasarkan jumlah operasi minimum yang diperlukan untuk mengubah satu string menjadi string lainnya. Algoritma Damerau Levenshtein Distance merupakan pengembangan dari algoritma Levenshtein Distance.
* Algoritma Damerau Levenshtein Distance melakukan operasi perbandingan kata-kata dengan memperhitungkan empat jenis kesalahan pengetikan yaitu penyisipan (insertion), penghapusan (deletion), penggantian (substitution), dan penukaran untuk dua karakter yang bersebelahan (transposition).
* 1. Operasi insertion merupakan operasi penyisipan dengan menyisipkan karakter pada indeks tertentu untuk menyamakan string sumber (i) string target (j). Pada indeks ke-4, dilakukan penyisipan karakter ‘n’ pada string sumber untuk menyamakan string tersebut terhadap string target.
* 2. Operasi deletion merupakan operasi penghapusan dengan menghapus karakter pada indeks tertentu untuk menyamakan string sumber (i) dan string target (j). Pada indeks ke-5, dilakukan penghapusan karakter ‘n’ pada string sumber untuk menyamakan string sumber tersebut string target.
* 3. Operasi substitution merupakan operasi penggantian dengan mengganti karakter pada indeks tertentu untuk menyamakan string sumber (i) dan string target (j). Pada indeks ke-4, dilakukan penggantian karakter ‘m’ menjadi ‘n’ pada string sumber untuk menyamakan string tersebut terhadap string target.
* 4. Operasi transposition merupakan operasi penukaran dengan menukar karakter berurutan pada indeks tertentu untuk menyamakan string sumber (i) dan string target (j). Pada indeks ke-4 dan ke-5, dilakukan penukaran karakter berurutan ‘g’ dan ‘n’ pada string sumber untuk menyamakan string tersebut terhadap string target.

**Boyer moore**

* Algoritma Boyer Moore adalah salah satu algoritma untuk mencari suatu string di dalam teks, dibuat oleh R.M Boyer dan J.S Moore. Ide utama Algoritma ini adalah mencari string dengan melakukan pembandingan karakter mulai dari karakter paling kanan dari string yang dicari.
* Boyer Moore menggunakan dua heuristik untuk memutuskan seberapa jauh karakter melompat yaitu bad-character heuristic (preBmBc), juga sering disebut tabel Occurrence Heuristic, dan heuristik yang good-suffix heuristic (premis), juga disebut tabel Match Heuristic.
* Nilai tabel yang paling besar yang akan digunakan untuk menentukan lompatan karakter.
* Pada contoh ini pencocokan dilakukan dari posisi paling akhir/kanan pattern dan dapat dilihat bahwa karakter “X” pada pattern “LINUX” tidak cocok dengan karakter “R” pada teks “BOYER” dan karakter “R” tidak pernah ada dalam pattern “LINUX” yang dicari sehingga pattern “LINUX” dapat diesel melewati teks “BOYER” sehingga posisi seperti.
* Dari hal tsb menunjukkan bahwa algoritma boyer moore memiliki pergeseran karakter yang besar sehingga mempercepat pencarian pattern. Karena dengan hanya memeriksa sedikit karakter, dapat langsung diketahui bahwa pattern yang dicari tidak ditemukan dan dapat diesel ke posisi berikutnya
* **Prembmc**
* Prosedur preBmBc (Pre-Compute Bad Character) digunakan untuk menentukan pergeseran pola ketika terjadi ketidakcocokan antara karakter dalam pola dan teks. Hasil dari prosedur preBmBc kemudian disimpan di dalam tabel Occurence Heuristic.
* **Premis**
* Prosedur ini bertujuan untuk membangun tabel yang disebut “good suffix table”. Tabel ini digunakan untuk menentukan pergeseran pola ketika sebagian pola cocok dengan teks, tetapi kemudian terjadi ketidakcocokan karakter.
* Sebelum menjalankan prosedur BmGs, prosedur suffix dijalankan terlebih dulu pada pattern. Fungsi dari prosedur suffix adalah memeriksa kecocokan sejumlah karakter yang dimulai dari karakter terakhir/terkanan dengan sejumlah karakter yang dimulai dari setiap karakter yang lebih kiri dari karakter terkanan.
* Pada contoh berikut kata “DETEKSI” sebagai pattern akan dilakukan pencarian atau pencocokan string pada teks “TEKDETKDETEKSITEK”. Langkah pertama adalah menjalankan procedure preBmBc dan prosedur premis. Kemudian hasilnya disimpan pada tabel Occurence Heuristic dan Match Heuristic.

**Pengumpulan data**

* 1. Studi Literatur, dilakukan dengan cara mengumpulkan referensi yang berkaitan dengan topik penelitian. Referensi dalam hal ini merupakan artikel dan dokumen terkait dengan implementasi Algoritma Boyer Moore, dan Damerau Levenshtein Distance yang bersumber dari buku, paper, jurnal, skripsi, dan lain sebagainya.
* 2. Untuk memenuhi kebutuhan data dalam penelitian ini, digunakan data kosa kata KBBI (Kamus Besar Bahasa Indonesia) yang diperoleh dari https://github.com/damzaky/kumpulan-kata-bahasa-indonesia-KBBI.git. Data juga berasal dari buku “Kamus Istilah Komputer dan Informatika“ oleh Banding Maseleno untuk mendapatkan istilah-istilah di bidang komputer dan informatika. Selain itu, draft jurnal sebagai data uji bersumber dari jurnal Animator UHO.

**Preprocessing text**

* tahap awal yang dilakukan untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut.
* 1. Cleaning : tahap pembersihan noise atau suatu bentuk yang nantinya dapat mengganggu proses pengolahan kata. Noise yang dimaksud yaitu karakter selain huruf alfabet, misalnya tag, HTML, link, script, dan sebagainya.
* 2. Case folding : preprocessing proses case folding bertujuan untuk mengubah semua huruf pada dokumen teks menjadi huruf kecil.
* 3. Tokenizing : proses tokenisasi yang berfungsi untuk memecah dokumen menjadi beberapa kata-kata yang disebut token.
* 4. Stemming : Stemming adalah proses pemetaan dan penguraian bentuk kata menjadi bentuk kata dasarnya. Proses stemming merupakan proses mengubah kata berimbuhan menjadi kata dasar

**Dictionary lookup**

* Metode dictionary lookup merupakan metode yang digunakan dalam penentuan non-word error. Proses yang dilakukan adalah pengecekan kata yang terdaftar dalam kamus atau tidak, jika tidak terdapat dalam kamus maka dianggap sebagai non-word. Cara ini adalah cara yang efektif untuk penentuan kata yang termasuk adanya kesalahan penulisan atau tidak.

**Black Box Testing**

* Black box testing merupakan metode perangkat lunak untuk menunjukkan kesalahan pada sistem aplikasi. Black box testing hanya menguji fungsionalitas sistem tanpa mengetahui secara rinci struktur program dari suatu sistem. Dalam black box testing, fokus utama berada pada input dan output yang dieksilkan sistem, bagaimana sistem menanggapi input yang diberikan

**Confusion matrix**

* Confusion matrix merupakan suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi.
* Pengolahan nilai-nilai yang ada pada kolom matriks True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP) maka dapat diketahui nilai accuracy, precision, dan recall