

Penerapan Algoritma *Enhanced Confix Stripping* dalam Pengukuran Keterbacaan Teks Menggunakan *Gunning Fog Index*

Enhanced Confix Stripping Algorithm Implementation on Measuring Text Readability with Gunning Fog Index

Yessi Nurul Fadzhiah¹, Rasim², Eka Fitrajaya R³

Departemen Pendidikan Ilmu Komputer, Universitas Pendidikan Indonesia
Bandung, Indonesia

¹ yessi.nurul.fadziah@student.upi.edu

^{2,3} {[rasim](mailto:rasim@upi.edu),[ekafitrajaya](mailto:ekafitrajaya@upi.edu)}@upi.edu

Abstrak— Untuk membantu pembaca, penulis dan pengajar dalam memilih bacaan yang mudah untuk dibaca dalam segi sintatik adalah dengan memeriksa tingkat keterbacaan bacaan tersebut. Keterbacaan merupakan ukuran tingkat kesulitan dan kemudahan suatu teks yang dipahami. Pengukuran keterbacaan dapat menggunakan formula keterbacaan atau teknik *cloze*. Penggunaan teknik *cloze* memiliki tingkat akurasi yang baik, namun teknik tersebut dapat memakan waktu pengujian lebih lama, maka dari itu pada penelitian ini digunakan formula keterbacaan untuk mengukur keterbacaan teks berdasarkan jumlah kata, jumlah kalimat dan jumlah kompleksitas kata. Formula Keterbacaan yang digunakan untuk mengukur keterbacaan berbahasa Indonesia adalah *Gunning Fog Index*. Untuk mengetahui validitas Formula *Gunning Fog Index* diperlukan suatu perbandingan hasil antara Formula *Gunning Fog Index* dan Teknik *Cloze*. Hasilnya diketahui bahwa Formula *Gunning Fog Index* memiliki akurasi mencapai 75%. Dalam Bahasa Indonesia, untuk menghitung jumlah kompleksitas kata adalah dengan menghitung jumlah kata serapan. Untuk mendapatkan jumlah kata serapan diterapkan Algoritma *Enhanced Confix Stripping* untuk memenggal kata imbuhan sehingga mendapatkan kata dasar yang selanjutnya kata dasar tersebut digunakan untuk mengecek ke dalam kamus kata serapan agar dapat menghitung jumlah kompleksitas kata. Dalam penelitian ini, Algoritma ini memiliki akurasi mencapai 82%, penerapan Algoritma *Enhanced Confix Stripping* ini masih memiliki kekurangan *overstemming* dan *undertsemming*.

Kata Kunci: Keterbacaan, Gunning Fog Index, Kompleksitas Kata, Algoritma *Enhanced Confix Stripping*

Abstract— To help readers, writers, and teachers on choosing easy-to-read reading materials syntactically is by checking the readability rate of the reading materials. Readability is a measurement of whether a text is hard or easy to be understood. Measuring readability can be done with readability formulas or Cloze technique. Utilization of Cloze technique has a better

accuracy rate, but could take much longer time, and thus readability formula is used on this research to measure a text's readability based on the amount of words, sentences, and word complexities. Readability Formula used to measure the readability on Indonesian is Gunning Fog Index. To find out the Gunning Fog Index Formula's validity, a comparison of Gunning Fog Index's and Cloze technique's result is needed. The result shows that Gunning Fog Index Formula has an accuracy as high as 75%. In Indonesian, to count the amount of word complexities is to count the amount of loanwords. Enhanced Confix Stripping algorithm is used to get the amount of loanwords by detaching affixes to acquire the base form of the word which is then used to check the loanwords dictionary to count the amount of word complexities. In this research, the Enhanced Confix Stripping algorithm has an accuracy as high as 82%, although it suffers from over-stemming and under-stemming.

Keywords: Readability, Gunning Fog Index, Word Complexity, Enhanced Confix Stripping Algorithm.

I. PENDAHULUAN

Sumber belajar merupakan salah satu unsur atau sarana prasarana penting dalam kegiatan pembelajaran. Keberadaan buku teks sebagai sumber maupun bahan ajar menjadi kebutuhan bagi setiap peserta didik agar dapat mencapai kompetensi minimal yang sesuai dengan kemampuan. Badan Standar Nasional Pendidikan (BSNP) telah bekerja sama dengan Pusat Kurikulum dan Perbukuan Balitbang Kementerian Pendidikan dan Kebudayaan dalam penilaian buku teks pelajaran dalam peraturan Badan Standar Nasional Pendidikan Nomor: 0041/P/BSNP/VIII/2016, hal tersebut dilakukan untuk menyeleksi buku teks pelajaran agar layak digunakan untuk seluruh peserta didik di Indonesia. Salah satu aspek penilaian Buku Teks Pelajaran adalah Bahasa dan Keterbacaan.

Keterbacaan terhadap buku perlu dipertimbangkan kecocokannya dengan tingkat pemahaman peserta [1]. Selain itu, ada beberapa buku teks yang informasinya tidak

tersampaikan karena terdapatnya kata-kata yang sulit dipahami. Hal itu mengakibatkan pembaca tidak mengerti dengan isi dari buku teks yang telah dibaca. Mengukur keterbacaan merupakan salah satu solusi dalam penyelesaian masalah-masalah tersebut. Keterbacaan merupakan ukuran tingkat kesulitan dan kemudahan suatu teks yang dipahami oleh peserta didik. Ada beberapa teknik pengukuran keterbacaan yang digunakan para peneliti, diantaranya adalah penilaian subjektif para ahli, metode tanya jawab, formula keterbacaan dan uji rumpang/teknik *cloze* [2].

Dari beberapa teknik pengukuran keterbacaan tersebut, teknik *cloze* banyak digunakan dalam pengukuran keterbacaan buku berbahasa Indonesia. Teknik *cloze* digunakan untuk mengukur tingkat keterbacaan sebuah teks dan melatih keterampilan serta kemampuan siswa melalui kegiatan pembelajaran [3]. Namun teknik *cloze* memiliki kekurangannya yaitu dalam pelaksanaan pengukurannya perlu dilakukan uji lapangan yang menyebabkan dapat memakan waktu yang lama. Selain teknik *cloze*, teknik lain yang digunakan untuk mengukur keterbacaan adalah formula keterbacaan. Teknik *cloze* dan formula keterbacaan memiliki persamaan dan perbedaan. Persamaannya adalah kedua teknik tersebut memiliki fungsi yang sama yaitu menentukan keterbacaan. Perbedaannya adalah formula keterbacaan digunakan untuk menentukan keterbacaan dari aspek visual (tulisan) sedangkan teknik *cloze* digunakan untuk menentukan keterbacaan dari aspek konsep (isi).

Pengukuran keterbacaan dengan formula keterbacaan sudah banyak situs yang menerapkan formula keterbacaan untuk mengukur keterbacaan teks. Seperti situs <http://www.readabilityformulas.com/> dan <https://readable.io/>. Kedua situs tersebut menyediakan media untuk mengukur keterbacaan dari suatu teks. Namun kedua situs tersebut hanya dapat mengukur keterbacaan suatu teks dalam Bahasa Inggris. Sehingga tidak dapat digunakan untuk mengukur keterbacaan suatu teks dari Bahasa lain termasuk Bahasa Indonesia. Hal tersebut dikarenakan aspek kompleksitas kata dalam Bahasa Inggris tentu berbeda dengan kompleksitas kata dalam Bahasa Indonesia. Formula *Gunning Fog Index* telah digunakan untuk mengukur keterbacaan teks berbahasa Indonesia [3].

Kompleksitas kata atau kata sulit dalam Bahasa Indonesia ditentukan dari jumlah kata serapan yang terdapat pada teks yang akan diukur [3]. Untuk membuat sistem pengukuran keterbacaan menggunakan formula *Gunning Fog Index* ini, terdapat perhitungan jumlah kata, jumlah kalimat dan jumlah kata serapan. Untuk menghitung jumlah kata serapan, perlu dilakukan pengecekan ke dalam kamus kata serapan. Kamus kata serapan yang digunakan juga merupakan kata dasar. Dalam Bahasa Indonesia, suatu teks tidak hanya terdiri dari kata dasar. Kata berimbuhan sangat banyak digunakan dalam

Bahasa Indonesia. Sehingga diperlukan suatu metode untuk mendapatkan kata dasar dari kata berimbuhan tersebut yaitu metode *stemming*. Metode *Enhanced Confix Stripping* merupakan metode *stemming* yang terbaru hasil perbaikan dari metode sebelumnya yaitu, *Confix Stripping*.

Oleh karena itu, pada penelitian ini akan dilakukan penerapan Algoritma *Enhanced Confix Stripping* untuk mendapatkan jumlah kata serapan yang akan digunakan untuk mengukur keterbacaan teks berbahasa Indonesia menggunakan *Gunning Fog Index*. Penelitian ini diharapkan menjadi salah satu solusi untuk merekomendasikan buku bacaan teks yang akan digunakan untuk pribadi (diri sendiri), peserta didik (bagi guru) dan target pembaca (bagi penulis).

II. PENELITIAN TERKAIT

Ada beberapa penelitian yang menganalisis tentang keterbacaan suatu teks berbahasa Indonesia. Diantaranya adalah pengimplementasian *Gunning Fog Index* pada uji keterbacaan teks Bahasa Indonesia menggunakan Bahasa pemrograman Python, dengan hasil uji coba terhadap salah satu bacaan untuk D3, S1 dan S2. Lihat pada Gambar 1.



Gambar 1. Hasil Uji Coba

Pada Gambar 1 menunjukkan hasil keterbacaan pada tingkat D3 dengan nilai angka 29,93, tingkat S1 dengan nilai 29,10, dan tingkat S2 dengan nilai 29,93 [4]. Penelitian lainnya menggunakan formula keterbacaan *SMOG*, *Gunning Fog Index*, *Flesch Kincaid* dan *Formula Djoko*. Hasilnya menunjukkan bahwa formula yang digunakan tidak cocok diterapkan untuk mengukur keterbacaan teks berbahasa Indonesia, karena adanya perbedaan indikator jumlah kata sulit [5].

III. TEKNIK CLOZE

Teknik *cloze* atau uji rumpang pertama kali dikenalkan oleh Wilson L. Taylor (1953). Teknik *cloze* merupakan penerapan dari *closure* yang ada pada teori psikologi persepsi Gestalt, merupakan suatu persepsi yang komplit dari gambar atau keadaan yang sebenarnya tidak sempurna [6]. Teknik *cloze* memiliki konsep menjelaskan kecenderungan orang untuk menyempurnakan suatu pola yang tidak lengkap menjadi satu kesatuan yang utuh. Teknik *cloze* dapat digunakan sebagai alat ukur keterbacaan dan alat pengajaran

membaca. Sebagai alat ukur keterbacaan, teknik *cloze* dapat menguji tingkat kesulitan bahan ajar, mengklasifikasikan tingkat baca siswa, dan mengetahui kelayakan wacana sesuai dengan peringkat siswa. Teknik *cloze* merupakan metode yang paling tepat untuk mengukur keterbacaan teks berbahasa Indonesia.

Tahapan untuk melakukan teknik *cloze* [3] adalah sebagai berikut :

1. Memilih wacana yang relatif sempurna, yang tidak bergantung pada informasi sebelumnya
2. Melakukan penghilangan kata ke-n tanpa memperhatikan arti dan fungsi kata-kata itu
3. Mengganti bagian-bagian yang dihilangkan tersebut dengan tanda garis lurus datar yang sama panjangnya
4. Memberi Salinan dari semua bagian yang direproduksi kepada siswa
5. Menggiatkan siswa untuk berusaha mengisi semua delisi dengan pertanyaan-pertanyaan dari konteks atau kata-kata sisanya
6. Menyediakan waktu yang relatif cukup untuk memberi kesempatan kepada siswa dalam menyelesaikan tugasnya

Perhitungan dari hasil prosedur tersebut dihitung dengan rumus :

$$Cloze\ test = \frac{Jawaban\ Benar}{Jumlah\ Soal} \times 100\%.....(1)$$

Jawaban dapat dinilai benar jika diisi dengan kata yang sama dengan teks asli. Penilaian *cloze test* yang digunakan adalah menurut Earl F. Rankin dan Josph W. Culhane, dengan kriteria persentase pada Tabel 2 [7].

TABEL 1
KRITERIA PERSENTASE PENILAIAN KETERBACAAN DENGAN
TEKNIK CLOZE

Persentase	Kriteria
> 60%	Mudah
41% - 60%	Sedang
< 40%	Sulit

IV. GUNNING FOG INDEX

Gunning Fog Index adalah salah satu formula keterbacaan yang mengukur tingkat keterbacaan teks menjadi 3 kategori, mudah, sedang atau sulit. Formula ini dikembangkan oleh Asosiasi Robert Gunning pada tahun 1944. Pada awalnya, uji coba keterbacaan wacana hanya bagi teks berbahasa Inggris, namun kemudian dikembangkan secara universal sehingga dapat diterapkan ke Bahasa non Inggris [8]. Penilaian *Gunning Fog Index* menggunakan formula sebagai berikut :

$$GFI = 0.4 \left(\frac{A}{k} + \frac{100}{A} S \right).....(2)$$

Dimana keterangan persamaan (2),

GFI : *Gunning Fog Index*

A : Jumlah Kata

k : Jumlah Kalimat

S : Jumlah Kompleksitas Kata

Algoritma *Gunning Fog Index* adalah sebagai berikut :

1. Pilihlah teks yang terdiri dari kurang lebih 100 kata
2. Tentukan dan hitunglah rata-rata panjang sebuah kalimat terdiri dari berapa kata
3. Hitung kata-kata kompleks, termasuk kata
4. Tambahkan panjang rata-rata kalimat dengan presentase kata-kata sukar
5. Kalikan dengan 0.4

Gunning mengelompokkan kategori kalimat menjadi empat [3]. Pengelompokan itu adalah sebagai berikut :

1. Jika indeks keterbacaannya di bawah 8, kalimat dianggap mudah dipahami (tingkat keterbacaan tinggi).
2. Jika indeks keterbacaannya antara 8 dan 9, kalimat masih dikategorikan mudah (tingkat keterbacaan tinggi).
3. Jika indeks keterbacaannya antara 10 dan 11, kalimat dikategorikan dapat dipahami (tingkat keterbacaan sedang).
4. Jika indeks keterbacaannya di atas 11, kalimat dianggap sulit dipahami (tingkat keterbacaannya rendah).

Dalam Bahasa Indonesia banyak sekali kata turunan yang terdiri atas tiga suku kata atau lebih. Kata-kata turunan seperti itu masih dapat dimengerti dengan mudah oleh masyarakat awam sekalipun. Oleh karena itu, yang termasuk dalam kata sulit dalam penelitian ini adalah kata-kata serapan dari Bahasa asing dan istilah-istilah teknis dalam bidang tertentu [3].

Pada Gambar 2, merupakan salah satu contoh teks yang masuk ke kategori mudah. Pengukurannya mengikuti langkah-langkah yang telah disebutkan sebelumnya. Seluruh indikator yang telah dihitung jumlahnya dimasukan ke dalam persamaan (1). Lihat pada Tabel 2. Hasil pengukuran yang lain dapat dilihat contoh teks pada Gambar 3 merupakan salah satu teks yang masuk ke kategori sulit. Pengukurannya dapat dilihat pada Tabel 3.

TABEL 2
HASIL PENGUKURAN TEKS 1

Jumlah Kata	Jumlah Kalimat	Jumlah Kata Serapan	Skor GFI	Kategori
144	24	14	6,288889	Mudah

TABEL 3
HASIL PENGUKURAN TEKS 2

Jumlah Kata	Jumlah Kalimat	Jumlah Kata Serapan	Skor GFI	Kategori
262	27	47	11,05705	Sulit

Pengalaman yang Tak Akan Kulupakan

Sudah tiga hari, aku berbaring di tempat tidur. Wajahku pucat dan suhu badanku turun naik. Menjelang malam, suhu badanku tinggi sekali.

Orang tuaku membawaku ke klinik yang tidak jauh dari rumah. Dokter memeriksa detak jantungku dengan stetoskop. Aku diberi obat. Meskipun begitu, kesehatanku belum juga pulih. Suhu badanku masih belum stabil. Badanku kadang-kadang menggigil kedinginan.

Kesehatanku semakin memburuk. Aku tidak enak makan. Makanan terasa pahit di lidahku. Aku hanya makan bubur satu atau dua sendok saja. Itupun sering aku muntahkan lagi.

Orang tua tentu saja semakin cemas. Aku kembali dibawa ke dokter. Dokter memeriksa mulut, mata, perut, dan dadaku.

Selesai pemeriksaan, dokter memanggil orang tuaku. Dokter menduga aku terkena tifus. Ia menganjurkan agar aku dirawat di rumah sakit. Akhirnya, aku dirawat selama satu minggu di sana. Guru dan teman-temanku menjengukku. Pengalaman itu tidak akan pernah kulupakan. Kini, sadarlah aku bahwa kesehatan itu sangat penting.

Gambar. 2 Teks 1

Lebih Baik Setelah Ikut Sekolah Sepak Bola

Hobiku bermain sepak bola. Hobiku itu bermula dari kesenanganku melihat pertandingan sepak bola di televisi. Hampir dapat dipastikan, tidak ada satu pertandingan pun yang terlewatkan. Bahkan, tengah malam pun aku pasti menontonnya.

Setiap sore, bersama-sama teman sekampung, aku bermain sepak bola di lapangan dekat sawah. Rasanya, senang sekali saat harus berebut bola. Apalagi jika aku dapat memasukkan bola ke gawang lawan.

Suatu hari, Ayah memasukkan aku ke sekolah sepak bola. Ayah mengatakan bahwa di sekolah sepak bola aku akan mengenal dengan baik seluk-beluk sepak bola. Jadinya, aku tidak sekadar bermain mengejar-ngejar bola. Setelah itu, dengan berlatih di sekolah sepak bola, hobiku bermain sepak bola dapat lebih terarah. Aku sangat senang.

Di sekolah sepak bola ini, aku merasakan latihan sepak bola yang berbeda. Di sekolah ini, aku mendapat pengetahuan tentang aturan bermain yang benar. Aku juga mendapat teman baru. Selain itu, aku juga mendapat bimbingan dan latihan yang benar. Kata pelatihku, pak Hamin, bermain sepak bola itu tidak cukup hanya menendang, menyundul, dan berlari. Semua ada caranya yang benar.

Pelatih tak lupa mengajarkan tentang kerja sama tim. Tim harus kompak, tidak boleh jalan sendiri-sendiri. Bermain sepak bola kan tidak sendirian, tetapi dengan orang banyak. Jadi, kita harus mengandalkan kerjasama yang baik dalam bermain sepak bola.

Di sekolah sepak bola ini, aku juga mendapat latihan fisik secara rutin. Ada satu hal yang menyenangkan bagiku, aku bisa bersaing secara sportif. Setelah ikut sekolah sepak bola, aku merasakan perbedaan dengan sebelum sekolah. Pokoknya, dengan ikut sekolah sepak bola, cara bermainku menjadi lebih bagus. Jika berada di lapangan, aku tidak sekadar berlari atau asal merebut bola.

Gambar. 3 Teks 2

V. ENHANCED CONFIX STRIPPING

Algoritma *Enhanced Confix Stripping* adalah pengembangan dari algoritma sebelumnya yaitu *Confix Stripping*. Algoritma ini menghasilkan performa yang baik dengan memiliki jenis kesalahan *stemming* lebih sedikit dari algoritma sebelumnya. Kesalahan dari Algoritma *Confix Stripping* adalah kurangnya aturan pemenggalan, adanya elemen beberapa kata dasar yang sama seperti partikel imbuhan sehingga terjadinya *overstemming* dan

understemming. Untuk memperbaiki kesalahan-kesalahan tersebut terdapat solusi seperti perlu dilakukan modifikasi dan penambahan aturan pemenggalan, menambahkan suatu algoritma tambahan [9]. Pada algoritma *Enhanced Confix Stripping* terdapat algoritma tambahan berupa proses pengembalian akhiran apabila proses recoding sebelumnya gagal dan penambahan aturan pemenggalan seperti pada Tabel 4 dengan keterangan C:huruf konsonan; V:huruf vokal; A:huruf vokal atau konsonan;P:partikel dari suatu kata.

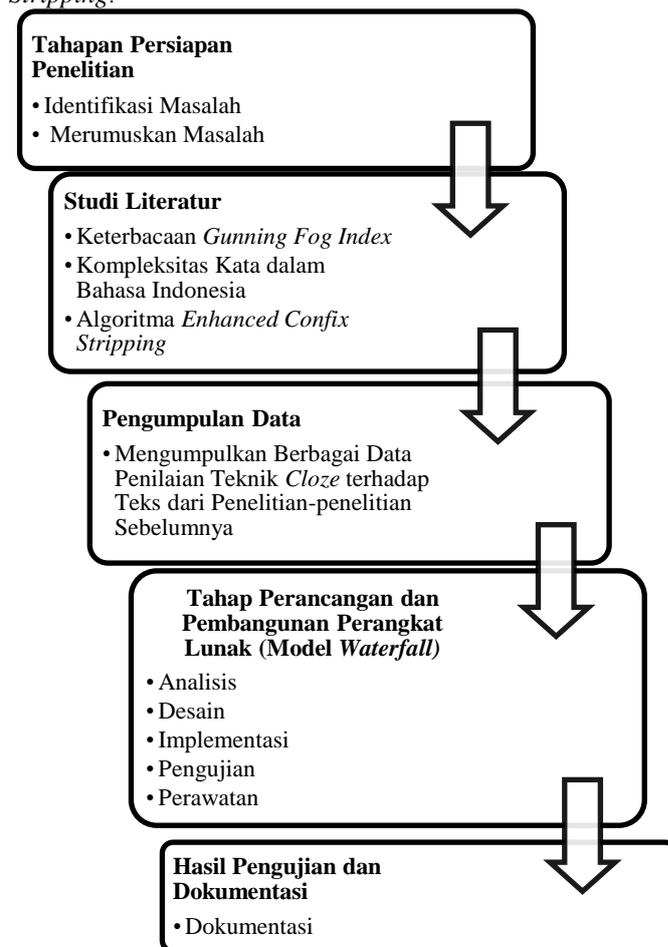
TABEL 4. ATURAN PEMENGGALAN ECS

Aturan	Format Kata	Pemenggalan
1	berV...	ber-V
2	berCAP...	ber-CAP, dimana C!='r' & P!='er'
3	berCAerV	ber-CaerV... dimana C!='r'
4	belajar	bel-ajar
5	beC1erC2...	be-C1erC2... dimana C1!={'r','l'}
6	terV...	ter-V... te-rV...
7	terCerV...	ter-CerV... dimana C!='r'
8	terCP...	ter-CP... dimana C!='r' dan P!='er'
9	teC1erC2...	te-C1erC2... dimana C1!='r'
10	me{l r w y}V...	me-{l r w y}V...
11	mem{b f v}...	mem-{b f v}...
12	mempe...	mem-pe...
13	mem{rV V}...	me-m{rV V}... me-p{rV V}...
14	men{c d j z s}...	men-{c d j z s}...
15	menV...	me-nV... me-tV
16	meng{g h q k}...	meng-{g h q k}...
17	mengV...	meng-V... meng-kV... (mengV-... jika V='e')
18	menyV...	meny-sV...
19	mempA...	mem-pA... dengan A!='e'
20	pe{w y}V...	pe-{w y}V...
21	perV...	per-V... pe-rV...
22	perCAP	per-CAP... dimana C!='r' dan P!='er'
23	perCAerV...	per-CAerV... dimana C!='r'
24	pem{b f V}...	pem-{b f V}...
25	pem{rV V}...	pe-m{rV V}... pe-p{rV V}...
26	pen{c d j z}...	pen-{c d j z}...

27	penV...	pe-nV... pe-tV...
28	pengC...	peng-C..
29	pengV...	peng-V... peng-kV.. (pengV-... jika V='e')
30	penyV...	peny-sV...
31	pelV...	pe-lV... kecuali "pelajar" yang menghasilkan "ajar"
32	peCerV...	per-erV... dimana C!={r w y l m n}
33	peCP...	pe-CP... dimana C!={r w y l m n} dan P!='er'
34	terC1erC2...	ter-C1erC2... dimana C1!='r'
35	peC1erC2...	pe-C1erC2... dimana C1!={r w y l m n}

VI. METODOLOGI

Gambar 4 menjelaskan tahapan pada penelitian ini. Tahapan pertama adalah persiapan penelitian yang meliputi identifikasi masalah dan perumusan masalah. Tahapan kedua studi literatur meliputi dasar teori yang digunakan pada penelitian ini seperti keterbacaan teks *Gunning Fog Index*, kompleksitas kata dalam Bahasa Indonesia, Algoritma *Enhanced Confix Stripping*.



Gambar 2. Tahapan Penelitian

Tahapan ketiga adalah pengumpulan data. Data yang dikumpulkan berupa data penilaian keterbacaan teks menggunakan teknik *cloze*. Tahapan keempat adalah tahapan pengembangan perangkat lunak, model yang digunakan pada penelitian ini adalah model *waterfall*, meliputi analisis kebutuhan sistem mengenai aturan stemming kata Bahasa Indonesia dan formula perhitungan keterbacaan, setelah analisis dilakukan maka merancang antarmuka sistem yang sesuai dengan kebutuhan sistem yang telah dianalisis, kemudian di implementasikan berupa kode program, alur yang dilakukan pada tahap ini dapat dilihat dari Gambar 5, setelah sistem telah dibuat maka dilakukan pengujian untuk mengetahui hasil dari penerapan Algoritma *Enhanced Confix Striping* dalam pengukuran keterbacaan teks menggunakan *Gunning Fog Index*.

Pada Gambar 5, merupakan *flowchart* sistem yang ada pada penelitian ini. Berikut adalah penjelasan dari *flowchart* sistem :

1. Inputan Teks, Pengguna menginputkan teks yang akan diukur keterbacaannya.
2. *Tokenization*, Teks tersebut akan dibersihkan dari angka dan simbol yang tidak dibutuhkan dalam pengukuran keterbacaan dan teks dibagi menjadi perkata untuk dilakukan pengecekan di tahap selanjutnya.
3. *Stemming* ECS, kata yang telah dibagi di cek ke dalam kamus kata dasar. Jika kata tidak ditemukan didalam kamus kata dasar, maka kata tersebut dilakukan pemotongan partikel imbuhan hingga ditemukan kata dasar.
4. Cek kamus kata serapan, kata hasil dari *stemming* di cek satu-persatu ke dalam kamus kata serapan. Jika kata tersebut adalah kata serapan maka dijumlahkan. Dan didapatkan jumlah kata serapan.
5. Hitung hasil keterbacaan, Jumlah kata serapan dimasukkan ke dalam formula *Gunning Fog Index*.
6. Menampilkan hasil keterbacaan



Gambar. 3 Flowchart Sistem

VII. HASIL DAN PEMBAHASAN PENELITIAN

Pengumpulan data yang digunakan dalam penelitian ini berupa penilaian keterbacaan dengan teknik *cloze*. Data tersebut digunakan untuk sebagai pembandingan dan validasi formula keterbacaan *Gunning Fog Index*. Data diambil dari penelitian-penelitian sebelumnya [10]. Data dapat dilihat di Tabel 5.

TABEL 5
DATA TEKNIK CLOZE

No	Judul Teks	Skor Teknik Cloze	Kategori
1	Asyiknya Ikut Lomba Pramuka	48,79%	Sedang
2	Senangnya Ikut Persami	42,50%	Sedang
3	Capung Besi yang Gesit dan Pemberani	29,60 %	Sulit
4	Lebih Baik Setelah Ikut Sekolah Sepak Bolra	48,25 %	Sedang
5	Pengalaman yang Tidak Akan Kulupakan	64,90 %	Mudah
6	Badmini Dulu Baru Badminton	44,49 %	Sedang
7	Mari Menabung di Koperasi	51,59%	Sedang
8	Kegemaran Perlu di Kembangkan	68,87%	Mudah

Pengukuran keterbacaan menggunakan formula keterbacaan *Gunning Fog Index* memerlukan suatu perbandingan untuk mengetahui hasil uji coba formula tersebut apakah dapat digunakan untuk mengukur keterbacaan teks berbahasa Indonesia atau tidak. Dengan teks yang sama dengan data uji, masing-masing teks dihitung jumlah kata, jumlah kalimat dan jumlah kata serapan dan dimasukkan ke dalam persamaan (2). Lihat pada Tabel 6.

TABEL 6
HASIL PERHITUNGAN GFI

Teks	Jumlah Kata	Jumlah Kalimat	Jumlah Kata Sulit	Skor GFI	Kategori
1	262	30	39	9,447532	Sedang
2	284	32	41	9,324648	Sedang
3	216	21	22	8,18836	Mudah
4	262	27	47	11,05705	Sulit
5	144	24	14	6,288889	Mudah
6	169	13	19	9,697041	Sedang
7	251	21	33	10,03992	sedang
8	226	26	21	7,193737	Mudah

Dari hasil perhitungan tersebut, perlu dibandingkan dengan teknik *cloze* untuk mengetahui tingkat akurasi dari formula *Gunning Fog Index*. Perbandingannya dapat dilihat di Tabel 7.

Untuk menghitung akurasi formula *Gunning Fog Index* maka digunakan rumus pada persamaan (3). Hasilnya menunjukkan bahwa akurasi formula *Gunning Fog Index* berdasarkan perhitungan 8 teks tersebut adalah 75%. Dari hasil tersebut, dapat disimpulkan bahwa formula *Gunning Fog Index* dapat digunakan untuk mengukur keterbacaan teks berbahasa Indonesia.

$$Akurasi = \frac{Jumlah\ Data\ Yang\ Sama}{Jumlah\ Semua\ Data} \times 100\%.....(3)$$

TABEL 7
PERBANDINGAN HASIL KETERBACAAN TEKNIK CLOZE DAN GFI

Teks ke	Cloze Tezt	GFI
1	Sedang	Sedang
2	Sedang	Sedang
3	Sulit	Mudah
4	Sedang	Sulit
5	Mudah	Mudah
6	Sedang	Sedang
7	Sedang	sedang
8	Mudah	Mudah

Sistem yang dibangun untuk mendukung penelitian ini menggunakan algoritma atau metode *Enhanced Confix Stripping*. Algoritma tersebut digunakan untuk mendapatkan kata dasar yang berguna untuk pengecekan ke dalam kamus kata serapan. Hasil dari pengecekan tersebut digunakan untuk menghitung jumlah kata serapan sebagai salah satu indikator dari pengukuran keterbacaan formula *Gunning Fog Index*. Sistem yang dibangun merupakan sebuah sistem berbasis web.

Antarmuka sistem yang dibuat dapat dilihat sebagai berikut:

1. Antarmuka Halaman Ukur

Pada Gambar 6 merupakan antarmuka form input teks untuk input teks yang akan diukur keterbacaannya oleh pengguna. Inputannya terdiri dari Judul, Topik dan Teks.

2. Antarmuka Hasil Pengukuran Keterbacaan

Pada Gambar 7 merupakan antarmuka hasil keterbacaan teks untuk menampilkan hasil dari perhitungan keterbacaan dari teks yang telah diinputkan sebelumnya. Gambar 8 merupakan antarmuka hasil keterbacaan teks yang menunjukan indikator keterbacaan. Pada Gambar 9 merupakan antarmuka hasil keterbacaan teks yang menunjukkan kata serapan.

3. Antarmuka Daftar Penilaian Keterbacaan

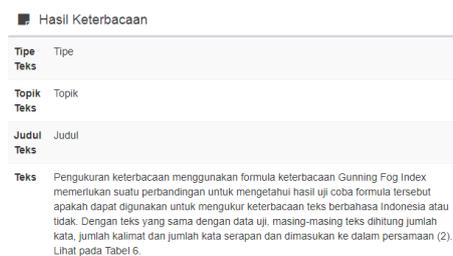
Pada Gambar 10 merupakan antarmuka daftar penilaian keterbacaan teks setelah hasil keterbacaan disimpan oleh pengguna.

4. Antarmuka Kamus Kata Serapan

Pada Gambar 11 merupakan antarmuka daftar kamus kata serapan untuk melihat kamus kata serapan yang terdiri kata tersebut dan Bahasa yang diserap.



Gambar 4. Antarmuka Halaman Ukur.



Gambar 5. Antarmuka Hasil Keterbacaan

Indikator Keterbacaan	
Jumlah Kata	109
Jumlah Kalimat	12
Jumlah Kompleksitas Kata	13
Skor Keterbacaan	8.4039755351682
Tingkat Kesulitan (Kategori)	Mudah

Gambar 6. Antarmuka Hasil Indikator Keterbacaan

#	Kata Serapan	Bahasa
1	wajah	Arab
2	alam	Arab
3	wajah	Arab
4	wajah	Arab
5	wajah	Arab
6	wajah	Arab
7	wajah	Arab
8	alam	Arab
9	pos	Belanda
10	pos	Belanda
11	pos	Belanda
12	pos	Belanda
13	hasil	Arab

Gambar 7. Antarmuka Hasil Kata Serapan

No	Tipe Bacaan	Topik Bacaan	Judul Bacaan	Skor Keterbacaan	Aksi
1	Narasi	Lingkungan Hidup	Asyiknya ikut Pramuka	8.39	Detail Skor
2	Deskripsi	Deskripsi	Deskripsi	0.00	Detail Skor
3	Deskripsi	Lingkungan Hidup	Badminton dulu	11.26	Detail Skor
4	Deskripsi	Lingkungan Hidup	Badminton dulu	11.26	Detail Skor
5	Tipe	Topik	Capung Besi yang Gesit dan Pemberani	6.72	Detail Skor
6	Narasi	Pengalaman Hidup	Pengalaman yang Tak Akan Kulupakan	5.76	Detail Skor

Gambar 8. Antarmuka Daftar Penilaian Keterbacaan

Hasil pengujian penerapan algoritma Enhanced Confix Stripping pada sistem dapat dilihat dari Tabel 8.

Dari 28 pengujian tersebut, terdapat 5 aturan yang gagal dalam pemotongannya. Hal tersebut dikarenakan hasil pemotongan loop pertama merupakan kata dasar sehingga tidak sesuai dengan yang seharusnya dipotong. Dari pengujian tersebut, dihitung akurasi menggunakan persamaan (3). Dan hasilnya menunjukkan bahwa penerapan Algoritma Enhanced Confix Stripping memiliki tingkat akurasi mencapai 82%. Hal itu dikarenakan masih terjadinya *overstemming* dan *understemming*. *Overstemming* dan *understemming* terjadi karena hasil pemotongan partikel imbuhan bukan hasil yang seharusnya. Kata yang dihasilkan dari proses stemming merupakan bukan kata yang dimaksud, namun kata tersebut terdapat pada kamus kata dasar.

No	Kata Serapan	Bahasa
1	abad	Arab
2	abadi	Arab
3	abdul	Arab
4	abjad	Arab
5	nasrani	Arab
6	absah	Arab
7	adab	Arab
8	adat	Arab
9	adil	Arab
10	ahad	Arab

Gambar 9. Antarmuka Kamus Kata Serapan

TABEL 8
HASIL PENGUJIAN STEMMING

No	Kata Berimbuhan	Hasil Stemming ECS	Kata Dasar	Keterangan
1	Beraksi	Aksi	Aksi	Berhasil
2	Bersantai	Santai	Santai	Berhasil
3	Belajar	Ajar	Ajar	Berhasil
4	Bekerja	Kerja	Kerja	Berhasil
5	Terambil	Ambil	Ambil	Berhasil
6	Terperah	Perah	Perah	Berhasil
7	terkejut	Kejut	Kejut	Berhasil
8	Melawan	lawan	lawan	berhasil
9	Membantu	bantu	Bantu	Berhasil
10	Mempekerjakan	mempekerjakan	kerja	Gagal
11	Memroses	proses	proses	Berhasil
12	Mencangkul	Cangkul	Cangkul	Berhasil
13	Menikam	Tikam	Tikam	Berhasil
14	Mengganggu	Ganggu	Ganggu	Berhasil
15	Mengaduk	Mengaduk	Aduk	Gagal
16	Menyaku	Saku	Saku	Berhasil
17	memprakasai	Memprakasai	prakasa	Gagal
18	Pewaris	Waris	Waris	Berhasil
19	Perusak	Perusak	Rusak	Gagal
20	Pertanda	Tanda	Tanda	Berhasil
21	Pembantu	Bantu	Bantu	Berhasil
22	Pemroses	Proses	Proses	Berhasil
23	Pencangkul	Cangkul	Cangkul	Berhasil
24	Penawar	Tawar	Tawar	Berhasil
25	Penggila	Gila	Gila	Berhasil
26	Pengampun	Ampun	Ampun	Berhasil
27	Penyangga	Angga	Sangga	Gagal
28	pelamar	Lamar	Lamar	Berhasil

VIII. KESIMPULAN

Sistem pengukuran keterbacaan ini merupakan sebuah sistem agar dapat memberikan suatu rekomendasi yang sesuai dengan kebutuhan. Berdasarkan hasil analisa dan pengujian yang

telah dilakukan, maka dapat diambil kesimpulan antara lain:

1. Formula Keterbacaan *Gunning Fog Index* hanya dapat mencapai akurasi 75%. Formula Keterbacaan *Gunning Fog Index* cukup baik diterapkan untuk mengukur keterbacaan teks berbahasa Indonesia.
2. Algoritma *Enhanced Confix Stripping* masih melakukan kesalahan-kesalahan seperti *understemming* dan *overstemming*. Hal tersebut dikarenakan terdapat 1 kata yang memiliki makna yang berbeda. Penerapan Algoritma *Enhanced Confix Stripping* pada penelitian ini memiliki tingkat akurasi mencapai 82%. Penerapan dalam pengukuran keterbacaan teks berbahasa Indonesia baik.
3. Sistem berhasil dibuat untuk mengukur keterbacaan teks berbahasa Indonesia.

REFERENSI

- [1] R. B. Woodbury, "On Book Reviewing," *American Anthropologist*, vol. 79, no. 3, pp. 551-554, September 1977.
- [2] J. Gilliland, "The Concept of Readability," *Journal of Reading*, pp. 24-29, 1968.
- [3] Astuti, W. Dwi and K. Biskoyo, Keterbacaan Kalimat Bahasa Indonesia dalam Buku Pelajaran SLTP, Jakarta: Pusat Bahasa, 2000.
- [4] D. Saptono, T. M. Sampurna, T. Wahyu R.N. and Fitrianiingsih, "Implementasi Algoritma Gunning Fog Index Pada Uji Keterbacaan (Readability Test) Bahasa Indonesia Menggunakan Bahasa Pemrograman Python," in *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2013*, Semarang, 2013.
- [5] M. K. Biddinika, R. P. Lestari, B. Indrawan, K. Yoshikawa, K. Tokimatsu and F. Takahashi, "Measuring the readability of Indonesian biomass websites: The ease of understanding biomass energy information on websites in the Indonesian language," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 1349-1357, 2016.
- [6] A. Rutledge, "Andy Rutledge," Agustus 2009. [Online]. Available: <http://www.andyrutledge.com/closure.php>. [Accessed 29 Oktober 2017].
- [7] V. Christanti, D. S. Naga and C. Benedicta, "PENGUKURAN TARAF SUKAR BACAAN DENGAN LEXILE FRAMEWORK DAN GUNNING FOG INDEX," *Jurnal Teknik dan Ilmu Komputer*, vol. 6, no. 22, pp. 199-216, 2017.
- [8] "Readability Formulas," The Gunning's Fog Index (or FOG) Readability Formula, [Online]. Available: <http://www.readabilityformulas.com/gunning-fog-readability-formula.php>. [Accessed 27 Oktober 2017].
- [9] A. Arifin, Mahendra and Ciptaningtyas, "Enhanced Confix Stripping Stemmer and Ants Algorithm for Classifying News Document in Indonesian Language," in *Proceeding of International Conference on Information & Communication Technology and Systems (ICTS)*, 2009.
- [10] A. Syukron, "Keterbacaan Wacana dalam Buku Teks Bahasa dan Sastra Indonesia untuk Sekolah Dasar Kelas 4 Terbitan Erlangga Berdasarkan Teknik Cloze," Universitas Jember, Jember, 2013.
- [11] N. H. Zahro, "ANALISIS TINGKAT KETERBACAAN DALAM BUKU TEKS PEMBELAJARAN TEMATIK TERPADU KURIKULUM 2013," *NOSI*, pp. 176 - 185, 2015.
- [12] M. Mukhyati and S. Sriyati, "Pengembangan Bahan Ajar Perubahan Lingkungan Berbasis Realitas Lokal dan Literasi Lingkungan," in *Prosiding Seminar Biologi*, 2016.
- [13] J. C. Begeny and D. J. Greene, "Can Readability Formulas Be Used To Successfully Gauge Difficulty Of Reading Materials?," *Psychology in the Schools*, pp. 198-215, 2014.
- [14] C. R. Vargas, D. J. Chuang, O. Ganor and B. T. Lee, "Readability of Online Patient Resources for the Surgical Treatment of Breast Cancer," *Surgery*, 2014.
- [15] F. Brigo, W. M. Otte, S. C. Igwe, F. Tezzon and R. Nardon, "Clearly written, easily comprehended? The readability of websites," *Epilepsy & Behavior*, pp. 35-39, 2015.
- [16] C. Moral, A. d. Antonio, R. Imbert and J. Ramirez, "A survey of stemming algorithms in information retrieval," *Information Research*, vol. 19, no. 1, 2014.
- [17] J. Asian, H. E. Williams and S. M. M. Tahaghoghi, "Stemming Indonesian," in *In Proceedings of the Twenty-eighth Australasian conference on Computer Science*, 2005.
- [18] M. Adriani, J. Asian, B. Nazief, S. Tahaghoghi and H. E. Williams, "Stemming Indonesian: A Confix-Stripping," *ACM Transactions on Asian Language Information Processing (TALIP)*, 2007.
- [19] Suladi, W. D. Astuti and K. Biskoyo, Keterbacaan Kalimat Bahasa Indonesia dalam Buku Pelajaran SLTP, Jakarta: Pusat Bahasa, 2000.
- [20] A. Rutledge, "Closure," 30 Agustus 2009. [Online]. Available: <http://www.andyrutledge.com/closure.php>.