

BAB I

PENDAHULUAN

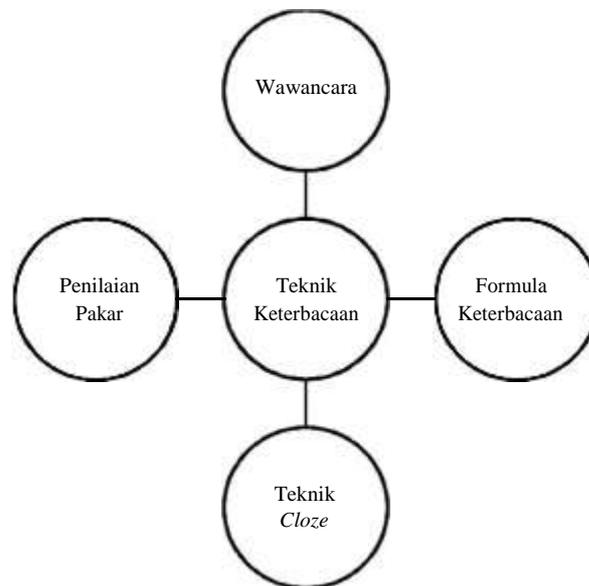
1.1. Latar Belakang

Sumber belajar merupakan salah satu unsur atau sarana prasarana penting dalam kegiatan pembelajaran. Sumber belajar dapat berupa kumpulan teks, suara, video atau gambar yang memiliki pesan untuk disampaikan. Salah satu contoh sumber belajar adalah buku teks. 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, hal tersebut dilakukan untuk menseleksi buku teks pelajaran agar layak digunakan untuk seluruh peserta didik di Indonesia. Dalam peraturan Badan Standar Nasional Pendidikan Nomor: 0041/P/BSNP/VIII/2016 tentang Prosedur Operasi Standar Penyelenggaraan Penilaian Buku Teks Pelajaran dan Buku Panduan Guru Pola “Inisiatif Masyarakat”, salah satu aspek penilaian buku teks pelajaran adalah Bahasa dan keterbacaan.

Aspek Bahasa dan Keterbacaan dalam penilaian terhadap buku teks pelajaran perlu dilakukan karena buku teks pelajaran penting sebagai sumber pengetahuan dan sumber hafalan. Keterbacaan terhadap buku perlu dipertimbangkan kecocokannya dengan tingkat pemahaman peserta didik (Woodbury, 1977). Maka dari itu, aspek Bahasa dan Keterbacaan menjadi salah satu poin yang dinilai dalam penilaian buku teks pelajaran. 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. Keterbacaan merupakan suatu 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* (Gilliland, 1968).



Gambar 1.1. Teknik Keterbacaan

Dari beberapa teknik pengukuran keterbacaan tersebut, teknik *cloze* banyak digunakan dalam pengukuran keterbacaan buku berbahasa Indonesia. Teknik *cloze* digunakan pada penelitian yang menganalisis keterbacaan dalam buku teks pembelajaran tematik terpadu kurikulum 2013 dan hasilnya menunjukkan bahwa buku teks pembelajaran tersebut memiliki tingkat keterbacaan yang bervariasi (Zahro, 2015). Teknik *cloze* digunakan juga pada penelitian yang mengembangkan materi pengajaran dan menggunakan tes *cloze* untuk menganalisis tingkat keterbacaan materi tersebut (Mukhyati & Sriyati, 2016). Teknik *cloze* digunakan untuk mengukur tingkat keterbacaan sebuah teks dan melatih keterampilan serta kemampuan siswa melalui kegiatan pembelajaran (Astuti, Dwi, & Biskoyo, 2000). Teknik *cloze* dianggap teknik yang baik sebagai ukur keterbacaan alat ajar karena pengukurannya melibatkan langsung antara teks dan pembacanya, disamping kelebihan tersebut, terdapat juga kekurangannya yaitu perlu dilakukan uji lapangan yang menyebabkan memakan waktu yang lama pengukuran keterbacaannya.

Selain teknik *cloze* tersebut, 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). Formula keterbacaan banyak digunakan untuk mengukur teks berbahasa Inggris. Terdapat banyak formula keterbacaan, diantaranya ada pada Gambar 1.2.

Name	Development	Formula
Dale-Chall	Assess the difficulty of comprehending reading materials for students above fourth grade level.	$\text{Grade} = (0.1579 \times \text{percent unfamiliar words}) + (0.0496 \times \text{word/sentence}) + 3.6365$
Flesch-Kincaid	Assess comprehension difficulty of reading materials for upper elementary through secondary grades.	$\text{Grade} = 0.39 (\text{average words/sentence}) + 11.8 (\text{average syllables/word}) - 15.59$
FOG	Used to assess difficulty based on the grade level necessary for comprehending at least 90% of the reading material.	$\text{Grade} = 0.4 [(\text{average words/sentence}) + (\text{percent of hard words})]$
Forcast	Developed to test basic reading competency to enter the U.S. Army and used best for multiple-choice tests and entrance forms.	$\text{Grade} = 20 - (\# \text{ single-syllable words}/10)$
Fry	Assess readability based on a graph and validated with primary and secondary reading materials.	$\text{Grade} = \text{graph average } (\# \text{ sentence, } \# \text{ syllables}) \text{ from three 100-word passages}$
Lexile	Uses both a measure of text difficulty and reader ability to match readers with appropriate text materials.	Based on word frequency and sentence length (rounded to the 10L.)
PSK	Developed an improved formula based on four previously established formulas (Dale-Chall, Flesch, Gunning, and Fairs-Jenkins-Paterson)	$\text{Grade} = 0.0778 (\text{average sentences length}) + 0.455 (\text{number syllables}) - 2.2029$
SMOG	Assess reading difficulty based with the goal of having a simple formula based only on the number of polysyllabic words.	$\text{Grade} = 3 + (\text{square root of the number of polysyllabic word count})$
Spache	Used to assess difficulty of reading materials for students in lower elementary grades (i.e., below third grade).	$\text{Grade} = (0.121 \times \text{word/sentence}) + (0.082 \times \text{percent unfamiliar words}) + 0.659$

Gambar 1.2. Ringkasan Rumus Keterbacaan (Begeny & Greene, 2014)

Dilihat dari Gambar 1.2, seperti yang sudah dijelaskan sebelumnya bahwa formula keterbacaan diukur berdasarkan aspek visual (tulisan) seperti panjang kata, panjang kalimat, panjang kata kompleks dari suatu teks. Ada beberapa penelitian yang menggunakan formula-formula tersebut untuk mengukur keterbacaan. Salah satunya penelitian yang mengukur nilai keterbacaan dengan 10 analisis keterbacaan pada sumber daya internet untuk informasi kesehatan berupa 104 artikel yang diambil dari 10 situs besar web untuk “operasi kanker payudara”, dan hasilnya menunjukkan rata-rata nilai keterbacaan adalah 12,9 dari 10 analisis keterbacaan yaitu Coleman-Liau dengan poin 12,6; Flesch Kincaid dengan poin 12,3; FORCAST dengan poin 11,2; Fry dengan poin 14, Gunning Fog dengan poin 14,4; New Dale-Chall dengan poin 12,4; New Foge dengan poin 11, Raygor dengan poin 14, dan Ukuran Sederhana Gobbledygook 14,3. Hal tersebut menunjukkan bahwa

sumber daya tersebut terlalu sulit untuk dipahami (Vargas, Chuang, Ganor, & Lee, 2014). Penelitian lainnya mengevaluasi tingkat melek kesehatan di 8 situs berbahasa Inggris secara kuantitatif keterbacaan, artikel yang dianalisis adalah 41 artikel epilepsi dari Wikipedia dengan menggunakan Flesch Reading Ease untuk menilai kemudahan pemahaman dan menggunakan Gunning Fox Index, Coleman-Liau Index, Flesch Kincaid Grade Level, Automated Readability Index dan Gobbledygook untuk memperkirakan tingkat kelas akademik yang dibutuhkan, dan hasilnya menunjukkan skor rata-rata FRE dari artikel Wikipedia menunjukkan tingkat keterbacaan yang sulit ($25,6 \pm 9,5$), dengan skala keterbacaan lainnya yang menghasilkan hasil yang sesuai dengan tingkat kelas 14 ($14,3 \pm 1,7$) (Brigo, Otte, Igwe, Tezzon, & Nardon, 2015).

Sudah banyak situs yang sudah menerapkan formula keterbacaan untuk mengukur keterbacaan teks. Seperti situs <https://readable.io/> dan <http://www.readabilityformulas.com/>. Kedua situs tersebut menyediakan media untuk mengukur keterbacaan dari suatu teks. Namun kedua situs tersebut hanya dapat mengukur keterbacaan suatu teks dalam Bahasa Inggris. Oleh karena itu, situs tersebut tidak dapat digunakan untuk mengukur keterbacaan suatu teks dari Bahasa lain termasuk Bahasa Indonesia. Ada beberapa penelitian yang menerapkan formula keterbacaan untuk mengukur keterbacaan teks berbahasa Indonesia. Salah satunya penelitian tentang mengukur keterbacaan situs biomassa Indonesia dan hasilnya menunjukkan bahwa formula keterbacaan tidak dapat digunakan untuk Bahasa Indonesia karena variabel keterbacaannya berbeda dengan Bahasa Inggris (Biddinika, et al., 2016). Khusus formula Spache dan Dale-Chall, formula tersebut tidak dapat digunakan untuk Bahasa Indonesia, karena presentase kata sulit pada kedua formula tersebut diambil dari daftar kata sulit yang sudah ditetapkan dan daftar tersebut terdiri dari kata berbahasa Inggris. Disamping itu, seperti pada penelitian sebelumnya formula keterbacaan kurang kompatibel dengan teks Bahasa Indonesia, hal tersebut dikarenakan adanya indikator kompleksitas kata/persentase kata sulit. Kompleksitas kata dalam Bahasa Inggris tentu berbeda dengan kompleksitas kata dalam Bahasa Indonesia.

Dalam Bahasa Inggris, yang disebut kompleksitas kata adalah kata yang memiliki 3 atau lebih suku kata. Sedangkan dalam Bahasa Indonesia, kata yang memiliki 3 atau lebih suku kata masih dapat dipahami dengan mudah oleh masyarakat awam sekalipun karena kata tersebut biasanya merupakan kata turunan dari kata dasar. Maka dari itu, kata sulit dalam Bahasa Indonesia ditentukan dari jumlah kata serapan yang terdapat pada teks yang akan diukur (Astuti, Dwi, & Biskoyo, 2000).

Untuk memenuhi indikator kompleksitas kata pada formula keterbacaan, yaitu kata serapan, diperlukan perhitungan jumlah kata serapan dengan mengecek satu persatu kata dari suatu teks yang akan diukur. Jika teks tersebut hanya terdiri dari kata dasar, dapat secara langsung di cek ke dalam kamus data. Namun dalam Bahasa Indonesia, terdapat kata turunan. Kata turunan adalah kata dasar yang terdapat imbuhan baik awalan, sisipan maupun akhiran (Triwarsih, 2014). Contoh kata serapan yang berimbuhan lihat pada Tabel 1.1.

Tabel 1.1. Contoh Kata Serapan Berimbuhan

No	Kata Serapan (Dasar)	Kata Serapan (Berimbuhan)	Bahasa
1	Aplikasi	Mengaplikasikan	Bahasa Inggris
2	Potret	Memotret	Bahasa Belanda
3	Palsu	Memalsukan	Bahasa Portugis
4	Cuci	Mencuci	Sansekerta

Untuk mendapatkan kata dasar dari kata turunan seperti pada Tabel 1.1, diperlukan suatu metode yaitu Metode *Stemming*. *Stemming* adalah suatu metode yang mengubah suatu kata yang memiliki imbuhan menjadi kata dasar. Ada beberapa penelitian yang menggunakan metode *stemming* untuk berbagai kasus. Diantaranya penelitian tentang penilaian mendetail mengenai status proses *stemming* dalam aplikasi pencarian informasi dengan menelusuri *historynya* dan menunjukkan bahwa metode *stemming* dapat memperbaiki sebuah informasi (Moral, Antonio, Imbert, & Ramirez, 2014). Penelitian lainnya tentang menyelidiki kinerja algoritma *stemming* melalui studi pengguna dan hasilnya menunjukkan bahwa kemampuan dengan

menggunakan algoritma Nazief dan Andriani mencapai 93% sedangkan dengan perbaikan yang diusulkan oleh penelitian ini kemampuannya mencapai 95% (Asian, Williams, & Tahaghoghi, 2005). Penelitian sebelumnya mengembangkan algoritma *Enhanced Confix Stripping* untuk melakukan *stemming* pada kasus klasterisasi dokumen baru berbahasa Indonesia dan hasilnya menunjukkan bahwa algoritma *Enhanced Confix Stripping* dapat menyelesaikan permasalahan pada algoritma *Confix Stripping* dan mengurangi ukuran hingga 32,66%, sedangkan *confix stripping stemmer* hanya mengurangi 30,95% pada kasus klasterisasi dokumen (Arifin, Mahendra, & Ciptaningtyas, 2009).

Maka dari itu, pada penelitian ini akan menerapkan algoritma *Enhanced Confix Stripping* (ECS) dalam pengukuran keterbacaan teks menggunakan Formula *Gunning Fog Index* (GFI) pada teks berbahasa Indonesia. 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) agar pengguna dapat mendapatkan informasi atau pesan yang telah disampaikan dalam buku yang telah dibacanya.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang sudah dipaparkan diatas, maka rumusan masalah yang dapat dirumuskan adalah sebagai berikut :

1. Bagaimana akurasi formula keterbacaan GFI untuk teks berbahasa Indonesia?
2. Bagaimana hasil *stemming* algoritma *Enhanced Confix Stripping* setelah diterapkan untuk mengukur keterbacaan teks?
3. Bagaimana membuat sistem pengukuran keterbacaan teks?

1.3. Tujuan Penelitian

Dari rumusan masalah yang telah diuraikan di atas, ada beberapa tujuan dari penelitian ini, yaitu :

1. Untuk mengetahui akurasi formula keterbacaan GFI untuk teks berbahasa Indonesia.

2. Untuk mengetahui hasil stemming dari Algoritma *Enhanced Confix Stripping*.
3. Untuk membuat sistem untuk pengukuran keterbacaan teks.

1.4. Batasan Masalah

Adapun batasan masalahnya, sebagai berikut :

1. Aturan kata berimbuhan dibatasi oleh Pedoman Umum Ejaan Bahasa Indonesia.
2. Sumber belajar yang dihitung keterbacaannya berupa teks.
3. Teks yang diproses hanya alphabet, tidak termasuk simbol dan angka.
4. Teks yang diinputkan ke dalam sistem berupa isian form.

Adapun kata kunci yang terdapat pada penelitian ini, sebagai berikut:

1. *Enhanced Confix Stripping Stemmer* : Suatu algoritma *stemming* yang dikembangkan dari algoritma sebelumnya yaitu *Confix Stripping*.
2. Kompleksitas Kata : Jumlah kata serapan dalam suatu teks.
3. *Gunning Fog Index* : Suatu formula keterbacaan untuk menentukan tingkat kesulitan.

1.6. Struktur Organisasi Skripsi

Untuk mempermudah dan memahami pembahasan penulisan penelitian ini, maka struktur organisasi skripsi ini disusun sebagai berikut:

BAB I. PENDAHULUAN

Berisi pembahasan masalah umum yang diangkat pada penelitian, di dalamnya terdapat latar belakang, rumusan masalah, tujuan penelitian, batasan masalah dan sistematika penulisan tentang penerapan Algoritma *Enhanced Confix Stripping* (ECS) dalam pengukuran keterbacaan suatu teks menggunakan *Gunning Fog Index*

BAB II. KAJIAN PUSTAKA

Berisi mengenai kajian teori yang digunakan di dalam penelitian. Pada bab ini akan dibahas dasar teori mengenai Keterbacaan, Kompleksitas Kata,

Kata Berimbuhan Bahasa Indonesia, Algoritma *Enhanced Confix Stripping* (ECS).

BAB III. METODOLOGI PENELITIAN

Berisi dasar teori mengenai metodologi yang digunakan untuk melakukan penelitian, metodologi meliputi desain penelitian yang membahas tahapan penelitian yang akan dilakukan dan model pengembangan perangkat lunak yang digunakan adalah *waterfall*, alat dan bahan penelitian.

BAB IV. HASIL PENELITIAN DAN PEMBAHASAN

Berisi penjabaran hasil penelitian yang telah dilakukan dan analisisnya. Semua pertanyaan mengenai rumusan masalah yang diangkat dalam tema, dibahas di bab ini.

BAB V. PENUTUP

Berisi kesimpulan dan saran bagi peneliti selanjutnya dari hasil penelitian yang telah dilakukan.