

# Sistem Koreksi Jawaban Uraian Singkat Otomatis Menggunakan Metode Winnowing

Bagus Nugraha<sup>1</sup>, Ardi Sanjaya<sup>2</sup>, Danar Putra Pamungkas<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail: \*<sup>1</sup>[nbagus61@gmail.com](mailto:nbagus61@gmail.com), <sup>2</sup>[dersky@gmail.com](mailto:dersky@gmail.com) <sup>3</sup>[danar.aflach@gmail.com](mailto:danar.aflach@gmail.com)

**Abstrak** – Dengan diberlakukannya proses belajar mengajar online, dibutuhkan sistem untuk melakukan koreksi untuk jawaban dari soal – soal latihan maupun soal ujian. Rumusan masalah yang dapat disimpulkan adalah bagaimana cara merancang sistem koreksi jawaban uraian singkat otomatis menggunakan metode winnowing. Tujuan penelitian ini adalah untuk membangun sistem koreksi jawaban uraian singkat dan diharapkan dengan adanya sistem koreksi jawaban uraian singkat otomatis menggunakan metode winnowing dapat membantu guru dalam melakukan proses koreksi jawaban uraian singkat. Metode penelitian yang digunakan adalah metode kuantitatif yaitu, metode yang digunakan untuk meneliti pada populasi atau sampel tertentu, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif/statistik, dengan tujuan untuk menguji hipotesis yang telah ditetapkan. Nilai akurasi dari implementasi algoritma winnowing sangat bergantung pada nilai hash, winnow, dan bilangan prima yang digunakan. Dari hal tersebut maka dilakukan 27 skenario pengujian untuk mendapatkan hasil yang mendekati nilai koreksi dari guru tanpa sistem. Dari skenario pengujian yang diimplementasikan pada sistem hasil optimal mengarah pada skenario pengujian 1, 2, dan 3 karena hasil yang paling mendekati nilai koreksi yang diberikan guru. Sistem jawaban uraian singkat menggunakan metode winnowing yang diimplementasikan menghasilkan nilai akurasi rata – rata yang cukup baik yaitu 14,88429 %.

**Kata Kunci** — sistem, jawaban, algoritma winnowing, nilai.

## 1. PENDAHULUAN

Sistem pendidikan di Indonesia pada masa covid - 19 mengalami banyak penyesuaian. Pembelajaran yang semula dilaksanakan secara bertatap muka langsung di sekolah kini harus disesuaikan dengan pelaksanaan pembelajaran secara daring atau online. Pemberian materi, tugas dan soal latihan seluruhnya menggunakan sistem online bahkan pelaksanaan ujian sekolah hingga ujian nasional semua wajib dilaksanakan secara online.

Penelitian yang dilakukan oleh Pradnya pada tahun 2020 sistem penilaian otomatis jawaban esai pada ujian online berbasis web dengan mengimplementasikan algoritma winnowing mampu bekerja dengan baik untuk jawaban yang cukup panjang dengan selisih rata – rata 5,683 % [1].

Dalam penelitian ini peneliti mencoba untuk mengimplementasikan metode winnowing dan query expansion untuk merancang sistem koreksi jawaban uraian singkat otomatis. Metode winnowing digunakan untuk melakukan perekaman dan pengolahan nilai kemiripan dokumen.

Tujuan penelitian ini adalah untuk merancang sistem koreksi jawaban uraian singkat otomatis yang diharapkan dengan adanya sistem koreksi jawaban uraian singkat otomatis menggunakan metode winnowing dan query expansion dapat membantu mempermudah tenaga pendidik dalam melakukan koreksi jawaban uraian singkat.

## 2. METODE PENELITIAN

Sistem penelitian menggunakan deskriptif kuantitatif. Metode kuantitatif adalah metode penelitian yang digunakan untuk meneliti pada populasi atau sampel tertentu, serta proses pengumpulan data menggunakan instrumen penelitian [2].

Metode kuantitatif juga disebut dengan metode survey [3]. Pada penelitian dengan metode survey peneliti memberika pertanyaan kepada beberapa orang yang disebut sebagai responden tentang pendapat mengenai karakteristik dan perilaku suatu objek.

Jadi dapat disimpulkan bahwa penelitian dengan metode kuantitatif adalah penelitian dengan cara melakukan interaksi langsung dengan objek yang akan diteliti.

### 2.1 Sistem Koreksi

Sistem adalah kumpulan beberapa elemen yang memiliki fungsi dan aturan tertentu yang digunakan untuk menyelesaikan suatu masalah. Prosedur – prosedur yang saling berhubungan dan berkumpul dalam suatu jaringan kerja untuk menyelesaikan sebuah permasalahan disebut sebagai sistem [4].

Koreksi adalah suatu cara untuk meninjau, memeriksa, dan melakukan perbaikan terhadap sesuatu. Koreksi dilakukan untuk mengetahui tingkat

kesalahan yang terjadi dan untuk mendapatkan kebenaran serta kesesuaian informasi yang masuk dengan informasi yang sudah dimiliki.

Otomatis adalah segala hal dan kegiatan manusia yang dapat dilakukan serta diselesaikan menggunakan suatu sistem dan mesin. Otomatisasi adalah kreasi dan penerapan teknologi untuk melakukan pemantauan dan kontrol produksi serta pengiriman produk dan layanan [5]

Dari pembahasan diatas dapat ditarik kesimpulan bahwa sistem koreksi otomatis adalah suatu sistem yang digunakan untuk meninjau hasil dan memeriksa kebenaran terhadap suatu hal dengan menerapkan metode serta prinsip kreasi teknologi..

## 2.2 Algoritma Winnowing

Metode *winnowing* adalah sebuah proses untuk melakukan deteksi terhadap nilai sebuah dokumen agar dapat diolah secara sistematis. Algoritma *winnowing* digunakan untuk melakukan proses *document fingerprinting* dan menggunakan penghitungan nilai *k-gram* dan *hashing* dalam implementasinya [6].

Implementasi algoritma *winnowing* meliputi beberapa proses, antara lain *hashing* dan penghitungan nilai *n-gram*.

### A. Proses hashing

Hashing adalah proses mengubah data kata yang memiliki variabel string menjadi nilai sesuai dengan ASCII (american standard code for information interchange) dalam bentuk array.

Proses hashing merupakan proses untuk melakukan konversi karakter dengan variabel *string* menjadi sebuah nilai dengan panjang tetap serta dapat mewakili karakter dari *string* asli [7]

Adapun rumus yang digunakan untuk melakukan proses *hashing* adalah

$$H(C_1 \dots C_n) = C_1 * b^{(n-1)} + C_2 * b^{(n-2)} + \dots + C_{(n-1)} * b + C_n$$

- c : nilai ASCII karakter
- b : basis bilangan prima (ditentukan oleh user)
- n : banyaknya karakter (ke - n)

contoh perhitungan nilai *hash* pada kata "saya" menggunakan basis bilangan prima (2) :

$$H(saya) = \text{ascii}(s) * 2^{(4-1)} + \text{ascii}(a) * 2^{(4-2)} + \text{ascii}(y) * 2^{(4-3)} + \text{ascii}(a) * 2^0$$

$$= 115 * 2^3 + 97 * 2^2 + 121 * 2^1 + 97 * 2^0 = 1647$$

## B. Menghitung nilai N-Gram

Nilai *n-gram* adalah nilai yang akan dijadikan pedoman proses perulangan untuk menentukan bentuk dari kata yang akan diolah. Pada proses ini data awal akan diubah menjadi kalimat baru sesuai dengan nilai yang telah ditentukan.

Contoh : saya main bola.  
N – gram : 4  
Hasil :  
[saya],[ayam],[yama],[amai],[main],  
[ainb],[inbo],[nbol],[bola]

## 2.3 Python

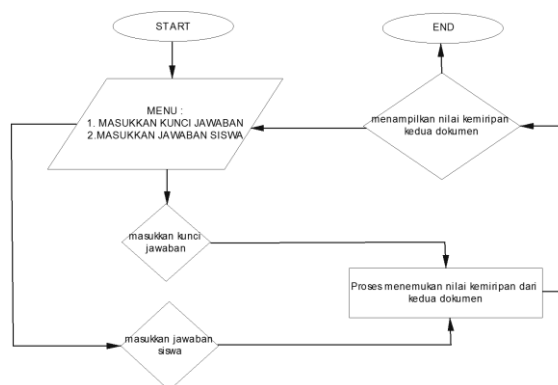
Python adalah bahasa pemrograman yang pada implementasinya berbasis objek. Bahasa pemrograman python adalah *object oriented programming* yang bersifat *high level* yaitu bahasa yang mudah untuk diterjemahkan digunakan mendekati dengan bahasa manusia [9]. Python dapat digunakan untuk merancang berbagai macam sistem seperti *mobile programming*, CLI (*command line interface*) , desktop dan game.

## 2.4 Perancangan Sistem

Perancangan sistem adalah tahap untuk mengetahui apa saja yang dibutuhkan untuk membangun sebuah sistem. Perancangan sistem juga merupakan tahap untuk melakukan perancangan alur kerja sistem dan tampilan sistem.

### A. Flowchart Sistem

Untuk mendapatkan nilai kemiripan dari data kunci jawaban dan data jawaban siswa digunakan diagram alir atau *flowchart*. Diagram alir dari sistem yang dirancang ditunjukkan oleh gambar 2.1



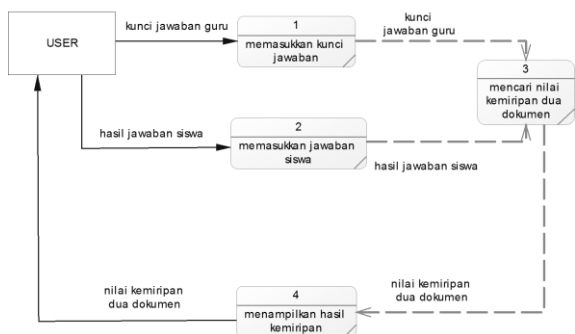
Gambar 2.1 Flowchart Sistem

Proses diawali dari user melakukan input data kunci jawaban, selanjutnya user melakukan input data jawaban siswa. Jika kedua data sudah dimasukkan maka proses berikutnya adalah sistem

akan melakukan proses mencari nilai kemiripan dari kedua dokumen. Proses terakhir adalah sistem akan menampilkan nilai kemiripan dari kedua dokumen dan proses selesai.

### B. Data Flow Diagram

Data flow diagram merupakan sebuah diagram yang menggambarkan aliran data dari suatu sistem yang dirancang. Aliran data pada sistem yang dirancang ditunjukkan dengan gambar 2.2.



Gambar 2.2 Data Flow Diagram Sistem

Langkah pertama user melakukan input data pada sistem, data yang dimasukkan adalah data kunci jawaban yang dimiliki oleh guru. Selanjutnya user memasukkan data jawaban yang diperoleh dari siswa. Selanjutnya kedua data akan diolah oleh sistem menjadi data nilai kemiripan. Data nilai kemiripan akan ditampilkan kembali pada user dan proses sistem selesai.

## 3. HASIL DAN PEMBAHASAN

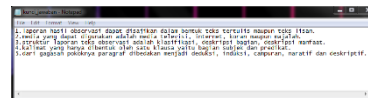
Hasil yang diharapkan dari sistem koreksi jawaban uraian singkat adalah nilai akhir tingkat kemiripan yang diperoleh mendekati nilai koreksi yang dilakukan oleh guru tanpa menggunakan sistem. Agar bisa memperoleh hasil akurasi yang mendekati nilai yang dilakukan oleh guru tanpa sistem maka dilakukan skenario pengujian.

### 3.1 Preprocessing Data

Tahap preprocessing data adalah proses mengolah data awal menjadi data yang siap untuk diolah pada proses inti, preprocessing data terdiri dari proses menyiapkan data kunci jawaban, dan data jawaban yang dimasukkan oleh siswa.

#### A. Data Kunci Jawaban

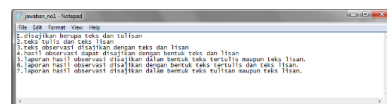
Data kunci jawaban adalah kunci jawaban yang dimiliki oleh guru yang akan menjadi acuan dalam melakukan koreksi pada jawaban siswa. Data kunci jawaban disimpan pada penyimpanan internal komputer menggunakan format '.txt'. adapun data kunci jawaban ditunjukkan pada gambar 3.1



Gambar 3.1 Data Kunci Jawaban

#### B. Data Jawaban Siswa

Data jawaban siswa adalah data jawaban yang dimasukkan oleh siswa pada saat mengerjakan soal latihan. Data jawaban siswa disimpan pada penyimpanan internal komputer menggunakan format '.txt'. adapun data jawaban siswa ditunjukkan pada gambar 3.2



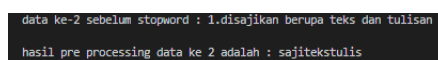
Gambar 3.2 Data Jawaban Siswa

### 3.2 Proses Inti

Pada proses inti data awal yang telah disiapkan diolah menggunakan metode *winnowing*. Proses inti terdiri dari beberapa tahapan. Tahapan yang dilakukan pada proses inti adalah :

#### A. Menghapus Kata dan Karakter Tidak Penting.

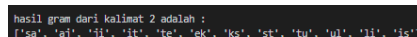
Kata dan karakter yang tidak penting akan membuat sistem berjalan lambat, maka kata dan karakter yang tidak penting harus dihapus. Kata dan karakter tidak penting yang akan dihapus adalah tanda baca, angka, spasi kosong, kata imbuhan dan kata yang sudah ditetapkan sebagai kata *stopword* atau kata tidak relevan pada library python. Hasil yang diperoleh pada proses ini ditunjukkan oleh gambar 3.3



Gambar 3.3 Menghapus Kata dan Karakter Tidak Penting

#### B. Menghitung Nilai N-Gram

Nilai n-gram adalah nilai yang akan dijadikan pedoman pada proses perulangan untuk menentukan bentuk dari kalimat yang akan diolah. Hasil dari proses menghitung nilai N-Gram ditunjukkan pada gambar 3.4



Gambar 3.4 Nilai N-Gram

#### C. Menghitung Nilai Hash

Menghitung nilai hash adalah proses mengubah data kata yang memiliki variabel string menjadi nilai sesuai dengan ASCII (american standard code for information interchange) dalam bentuk array agar dapat dilakukan penghitungan nilai secara matematis. Hasil proses menghitung nilai hash ditunjukkan oleh gambar 3.5

```
Hasil Hash kalimat 2 :
[115, 97, 106, 105, 116, 101, 107, 115, 116, 117, 108, 105]
```

Gambar 3.5 Nilai Hash

#### D. Menghitung Nilai Fingerprint

Proses menghitung nilai fingerprint adalah proses menghitung nilai terkecil yang diambil dari data hasil proses menghitung hash. Hasil menghitung nilai fingerprint ditunjukkan pada gambar 3.6

```
Data Fingerprinting 2 :
[97, 97, 105, 105, 101, 101, 107, 115, 116, 108, 105]
```

Gambar 3.6 Nilai Fingerprint

#### E. Mencari Nilai Fingerprint yang Sama

Nilai fingerprint yang diperoleh dari kedua dokumen akan dicari nilai yang sama. Nilai fingerprint yang sama akan menjadi acuan untuk menentukan nilai kemiripan dari kedua dokumen yang diproses. Hasil fingerprint yang sama dari kedua dokumen ditunjukkan pada gambar 3.7

```
Fingerprint sama :
[97, 97, 101, 101, 105, 105, 105, 107, 108, 115, 116]
```

Gambar 3.7 Nilai Fingerprint Sama

#### F. Menghitung Nilai Kemiripan

Proses yang terakhir adalah proses menghitung nilai kemiripan dari kedua dokumen. Nilai kemiripan kedua dokumen akan ditampilkan dalam bentuk persentase kemiripan. Hasil nilai kemiripan ditunjukkan gambar 3.8

```
Tingkat Kemiripan Kedua Dokumen Adalah : 55.00000000000001%
```

Gambar 3.8 Nilai Kemiripan

### 3.3 Skenario Pengujian

Peneliti menggunakan 27 skenario pengujian dengan membatasi nilai gram 2, 3 serta 4, nilai window 2,3 serta 4 dan nilai bilangan prima 2,3 serta 5. Berikut disajikan data kombinasi untuk skenario pengujian pada tabel 1 :

Tabel 1 Skenario Pengujian

Skenario	Nilai Gram	Nilai Window	Nilai Bilangan Prima
1	2	2	2
2	2	2	3
3	2	2	5
4	2	3	2
5	2	3	3
6	2	3	5
7	2	4	2
8	2	4	3
9	2	4	5

10	3	2	2
11	3	2	3
12	3	2	5
13	3	3	2
14	3	3	3
15	3	3	5
16	3	4	2
17	3	4	3
18	3	4	5
19	4	2	2
20	4	2	3
21	4	2	5
22	4	3	2
23	4	3	3
24	4	3	5
25	4	4	2
26	4	4	3
27	4	4	5

Untuk memudahkan dalam analisa, peneliti menggunakan skala nilai 100 untuk masing-masing butir jawaban hasil nilai koreksi jawaban dari aplikasi yang dibuat dan dari guru selaku pakar yang akan digunakan sebagai pembanding. Berikut disajikan cuplikan data koreksi jawaban dari pakar pada tabel 2 :

Tabel 2 Data Koreksi Jawaban Dari Pakar

ID	J1	J2	J3	J4	J5
1	60	75	60	60	75
2	60	75	100	60	50
3	90	90	90	60	60
4	100	50	75	100	75
5	100	55	55	60	100
6	100	55	60	75	50
7	95	60	90	90	50

### 3.4 Hasil Akhir

Berikut disajikan tabel 3 yang merupakan cuplikan hasil pengujian pada skenario 1 dan tabel 4 yang merupakan selisih terhadap nilai koreksi dari pakar :

Tabel 3. Cuplikan Hasil Skenario 1

ID Siswa	J1	J2	J3	J4	J5
1	55.0	79.31	97.87	45.83	87.80
2	64.0	96.0	100.0	90.47	68.75

3	80.0	94.28	100.0	88.0	75.86
4	86.11	86.66	76.47	100.0	87.5
5	100.0	85.71	97.5	58.33	100.0
6	100.0	75.0	97.87	88.0	70.0
7	100.0	76.92	71.08	100.0	73.33

Tabel 4 Selisih Skenario 1 Terhadap Data Pakar

ID Siswa	J1	J2	J3	J4	J5
1	5	4.31	37.87	14.17	12.8
2	4	21	0	30.47	18.75
3	10	4.28	10	28	15.86
4	13.89	36.66	1.47	0	12.5
5	0	30.71	42.5	1.67	0
6	0	20	37.87	13	20
7	5	16.92	18.92	10	23.33

Dari hasil pengujian berdasarkan skenario percobaan yang ditentukan, didapati hasil bahwa terdapat selisih nilai yang dihasilkan antara hasil koreksi dari aplikasi yang dibuat terhadap hasil koreksi yang dilakukan oleh pakar. Tabel 4 menyajikan data selisih pada skenario 1. Hasil yang ideal adalah nilai selisih yang mendekati nol. Nilai selisih tertinggi terdapat pada hasil koreksi pada ID Siswa 5 jawaban nomor 3 dengan nilai selisih 42,5. Berikut adalah data dari jawaban siswa tersebut dan jawaban versi pakar :

Jawaban siswa = struktur teks observasi adalah deskripsi, dan klasifikasi.

Kunci jawaban = struktur laporan teks observasi adalah klasifikasi, deskripsi bagian, deskripsi manfaat

Berdasarkan analisa aplikasi memberikan nilai 97,5 dan dari pakar memberi nilai 55. Nilai 97,5 yang diberikan oleh aplikasi memang tinggi yang mana memang terdapat banyak kata yang sama sehingga diperoleh nilai tinggi namun dari sisi penilaian pakar diberikan nilai 55 dikarenakan pakar tidak hanya menilai dari sisi kesamaan kata namun juga ada faktor kemiripan makna. Pada kunci jawaban terdapat kata deskripsi bagian dan deskripsi manfaat sementara pada jawaban siswa hanya terdapat kata deskripsi. Berikut disajikan data selisih rata-rata pada masing-masing skenario pada tabel 5 :

Tabel 5 Selisih Rata-Rata Masing-Masing skenario Terhadap Nilai Jawaban Pakar

Skenario	Rata-Rata Tiap Jawaban					Rata-Rata
	J1	J2	J3	J4	J5	
1	5	4.31	37.87	14.17	12.8	
2	4	21	0	30.47	18.75	
3	10	4.28	10	28	15.86	
4	13.89	36.66	1.47	0	12.5	
5	0	30.71	42.5	1.67	0	
6	0	20	37.87	13	20	
7	5	16.92	18.92	10	23.33	

1	2.84	19.13	15.83	9.38	14.75	12.38
2	2.84	19.13	15.83	9.38	14.75	12.38
3	2.84	19.13	15.83	9.38	14.75	12.38
4	2.79	21.34	13.96	7.10	15.05	12.05
5	2.79	21.34	13.96	7.10	15.05	12.05
6	2.79	21.34	13.96	7.10	15.05	12.05
7	1.77	20.91	15.51	4.86	13.28	11.27
8	1.77	20.91	15.51	4.86	13.28	11.27
9	1.77	20.91	15.51	4.86	13.28	11.27
10	16.97	7.83	2.37	11.30	17.22	11.14
11	16.97	7.83	2.37	11.30	17.22	11.14
12	16.97	7.83	2.37	11.30	17.22	11.14
13	16.21	11.46	0.06	11.14	19.03	11.58
14	16.21	11.46	0.06	11.14	19.03	11.58
15	16.21	11.46	0.06	11.14	19.03	11.58
16	14.90	14.34	0.30	13.36	19.80	12.54
17	14.90	14.34	0.30	13.36	19.80	12.54
18	14.90	14.34	0.30	13.36	19.80	12.54
19	18.95	15.00	2.72	22.07	21.69	16.09
20	18.95	15.00	2.72	22.07	21.69	16.09
21	18.95	15.00	2.72	22.07	21.69	16.09
22	17.41	18.75	0.31	23.84	24.02	16.87
23	17.41	18.75	0.31	23.84	24.02	16.87
24	17.41	18.75	0.31	23.84	24.02	16.87
25	17.44	20.96	1.29	29.01	27.55	19.25
26	17.44	20.96	1.29	29.01	27.55	19.25
27	17.44	20.96	1.29	29.01	27.55	19.25

#### 4. KESIMPULAN

Dari pembahasan yang dilakukan dan hasil yang diperoleh dapat diambil kesimpulan sebagai berikut :

1. Sistem koreksi jawaban uraian singkat menggunakan metode Winnowing pada skenario 10,11,12 memiliki selisih rata-rata paling kecil dibandingkan dengan skenario yang lain yaitu sebesar 11,14.
2. Nilai akhir algoritma winnowing sangat bergantung pada nilai hash, winnow dan bilangan prima yang digunakan.
3. Dari sisi akurasi kemiripan kalimat algoritma Winnowing memiliki hasil kemiripan yang tinggi namun apabila dibandingkan dengan nilai dari pakar yang dipengaruhi faktor penilaian terhadap kemiripan makna.

#### 5. SARAN

Perancangan sistem koreksi jawaban uraian singkat otomatis ini tentu masih jauh dari kata sempurna maka dari itu saran dan masukan yang membangun sangat dibutuhkan. Adapun saran yang diberikan oleh penulis adalah :

1. Sistem yang diimplementasikan membutuhkan lebih banyak data yang beragam untuk meningkatkan hasil akurasi yang didapatkan.
2. Diperlukan tampilan antarmuka yang baik untuk mendukung sistem agar lebih mudah dioperasikan.
3. Menggunakan metode lain sebagai metode pendamping algoritma Winnowing agar dapat mengolah data dalam bentuk yang lain.

#### DAFTAR PUSTAKA

- [1] Pradnya, W. M. (2020). *Implementasi Algoritma Winnowing Pada Sistem Penilaian Otomatis Jawaban Esai Pada Ujian Online Berbasis Web*. Jurnal Teknik Komputer, 6(2), 169-175
- [2] Sugiyono.2018. *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta. Hal.15
- [3] Neuman, W. L. 2003. *Social Research Methods: Qualitative and Quantitative Approaches 5th Edition USA: A and B*.
- [4] Kristanto, A. 2003. *Analisa dan Perancangan Sistem Informasi dan Aplikasinya*. Yrama Widya, Malang. Harliyan, *Analisa Dan Perancangan Sistemn Informasi Kependudukan Desa*.Hal.55
- [5] *International Society of Automation. 2018. Functional Safety–Safety Instrumented Systems for the Process Industry Sector–Part 1: Framework, Definitions, System, Hardware and Application Programming Requirements (IEC 61511–1: 2016+ AMD1: 2017 CSV, IDT)*. International Society of Automation, Research Triangle Park.
- [6] Schleimer, S., Wilkerson, D. S., & Aiken, A. (2003, June). Winnowing: local algorithms for document fingerprinting. In Proceedings of the 2003 ACM SIGMOD international conference on Management of data. Hal.76-85.
- [7] Simarmata J.,dkk. 2021. *Teknologi Informasi dan Multimedia*. Yogyakarta: Yayasan Kita Menulis.Hal.25
- [8] Laxmi, M. D., & Indriati, M. A. F. 2018. *Query Expansion Pada Sistem Temu Kembali Informasi Berbahasa Indonesia Dengan Metode Pembobotan TF-IDF Dan Algoritme Cosine Similarity Berbasis Wordnet*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X.
- [9] Sembiring F., Erfina A.2020.*Bahasa Ular untuk Pemrograman Python*.Solok: INSAN CENDEKIA MANDIRI. Hal.1