

BAB 4

IMPLEMENTASI DAN PENGUJIAN

4.1. Perancangan Sistem

Pada proses ini akan dijelaskan mengenai beberapa tahapan pengujian yang akan dilakukan terhadap sistem ekstraksi informasi menggunakan algoritma GHMM.

4.1.1. Implementasi Perangkat Keras

Berdasarkan percobaan sebelumnya terhadap beberapa perangkat keras untuk menangani proses ekstraksi informasi menggunakan algoritma GHMM. Berikut ini merupakan implementasi perangkat keras yang digunakan dapat dilihat pada table 4.1

Tabel 4.1 Implementasi Perangkat Keras

No	Perangkat Keras	Spesifikasi
1.	<i>Processor</i>	Intel(R) Core(TM) i3-2370M CPU @ 2.40GHz
2.	RAM	4 GB

Perangkat keras pada table 4.1 digunakan untuk melakukan pengujian terhadap sistem ekstraksi informasi. Perangkat keras tersebut merupakan perangkat keras minimum yang telah ditentukan.

4.1.2. Implementasi Perangkat Lunak

Perangkat lunak yang digunakan pada sistem ekstraksi informasi menggunakan algoritma GHMM berbasis platform web terdapat beberapa perangkat lunak pilihan. Berikut implementasi perangkat lunak dapat dilihat pada table 4.2.

Tabel 4.2 Implementasi Perangkat Lunak

No	Perangkat Keras	Spesifikasi
1.	Sistem operasi	Windows
2.	Web Browser	<i>Google Chrome/ Uc Browser</i>

Pada tabel 4.2 perangkat lunak yang digunakan untuk melakukan pengujian berdasarkan analisi yang telah dilakukan sebelumnya.

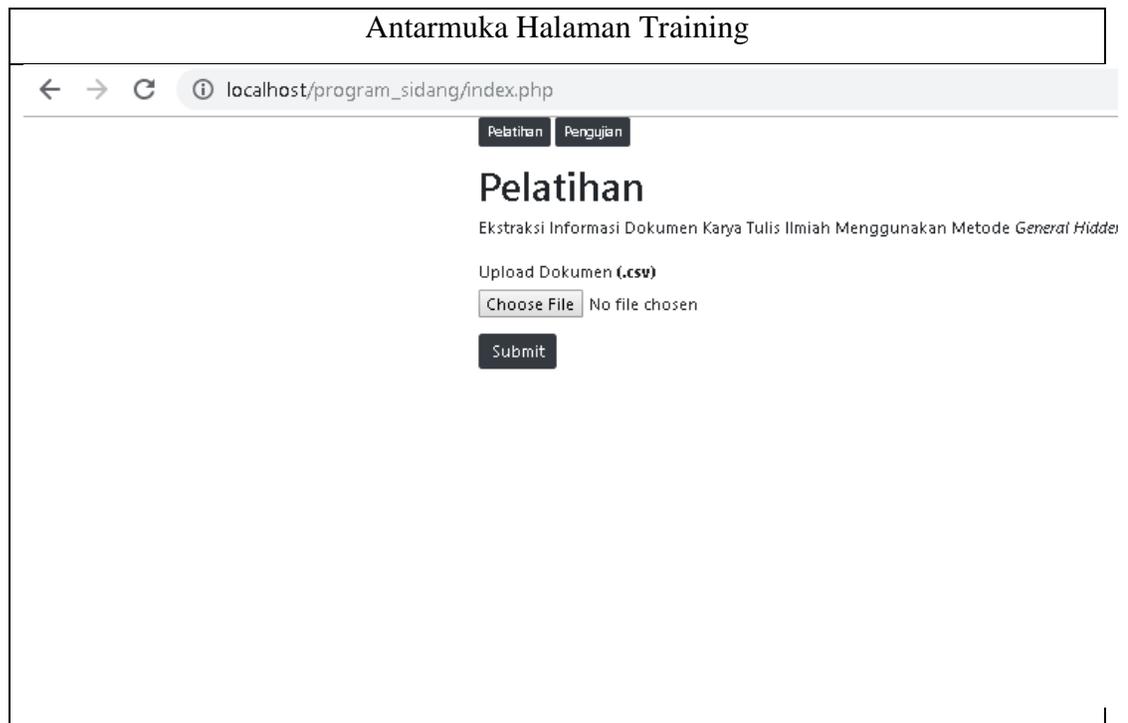
4.1.3. Implementasi Antarmuka

Perancangan antarmuka yang telah dirancang sebelumnya pada 3.5 diimplementasikan terhadap sistem ekstraksi informasi menggunakan algoritma GHMM yang akan dibangun. Berikut ini implementasi antarmuka sistem ekstraksi informasi dapat dilihat pada tabel 4.3.

Tabel 4.3 Implementasi Antarmuka

No	Nama Antarmuka	Keterangan	Nama Berkas
1	<i>Training</i>	Menampilkan halaman proses <i>training</i>	index.php
2	<i>Testing</i>	Menampilkan halaman proses <i>testing</i>	Index2.php

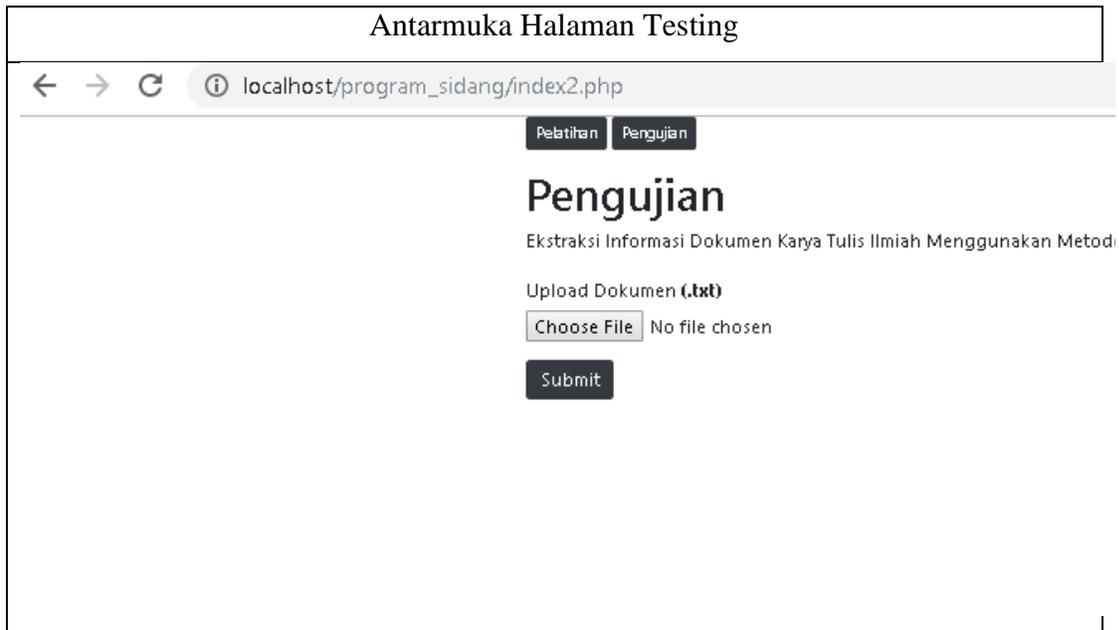
Pada tabel 4.3 terdapat dua halaman antarmuka sistem ekstraksi informasi menggunakan algoritma GHMM. Kedua halaman tersebut telah menjadi kecukupan untuk pengguna dalam memahami proses yang berjalan pada sistem ekstraksi informasi. Berikut antarmuka pada halaman training.



Gambar 4.1 Antarmuka Halaman Training

Pada gambar 4.1 terdapat tombol choose file yaitu untuk memilih file yang akan di unggah. File yang diunggah harus berbentuk CSV, halaman training akan menampilkan data masukan training, preprocessing data training dan training GHMM.

Masing-masing proses yang ditampilkan pada halaman training memiliki sub proses. Setelah antarmuka training sistem ekstraksi informasi menggunakan algoritma GHMM dapat diimplementasikan, berikut ini adalah halaman testing yang diimplementasikan pada sistem ekstraksi informasi.



Gambar 4.2 Antarmuka Halaman Testing

Pada gambar 4.2 antarmuka halamn testing terdapat tombol untuk mengunggah file sampul dan abstrak berbentuk TXT. Setelah file diunggah sistem akan menampilkan beberapa proses yang terdapat pada antarmuka halaman testing. Proses yang terdapat pada halaman testing meliputi data masukan testing, preprocessing data testing, testing GHMM dan evaluasi.

4.2. Pengujian Sistem

Pengujian sistem dilakukan dengan 2 tahapan, yaitu membuat skenario pengujian, kemudian melakukan pengujian berdasarkan skenario yang telah dibuat. Pengujian dilakukan terhadap fungsionalitas sistem serta nilai akurasi error.

4.2.1. Skenario Pengujian

Pengujian yang dilakukan sistem melalui 2 bagian, diantaranya pengujian fungsionalitas dan nilai akurasi serta error. Suatu metode digunakan pada pengujian fungsionalitas, sedangkan untuk pengujian nilai akurasi dan error dihitung dengan konsep kelas-token.

4.2.1.1. Skenario Pengujian Fungsionalitas

Metode yang digunakan untuk membuat skenario pengujian fungsionalitas adalah metode black box. Metode ini berfokus pada nilai output yang diberikan oleh input, mauput terhadap proses yang terjadi. Berikut skenario pengujian fungsionalitas dapa dilihat pada tabel 4.4.

Tabel 4.4 Skenario Pengujian Fungsionalitas

No	Proses	Rincian Pengujian	Jenis
1	Menampilkan data masukan <i>preprocessing data training</i>	-	<i>Black Box</i>
2	<i>Preprocessing data training</i>	Tokenisasi Ekstraksi Fitur	<i>Black Box</i>
3	<i>Training GHMM</i>	Menghitung Probabilitas awal Menghitung <i>probabilitas transisi</i> Menghitung Probabilitas emisi Menghitung observasi sequence Menghitung Observasi Sequeunce lower dan upper Menghitung nilai Generalisasi Probabilitas awal Menghitung nilai Generalisasi Probabilitas Transisi Menghitung nilai Generalisasi Probabilitas Emisi	<i>Black Box</i>
4	Menampilkan bobot akhir	-	<i>Black Box</i>
5	Menampilkan data masukan <i>preprocessing data testing</i>	-	<i>Black Box</i>
6	<i>Preprocessing data testing</i>	Segmentasi Data Tokenisasi Ekstraksi Fitur	<i>Black Box</i>
7	<i>Testing GHMM</i>	Menghitung inisialisasi Menghitung <i>induksi</i> <i>Terminasi</i> <i>Backtracking</i>	<i>Black Box</i>

8	Evaluasi	Hasil Pengujian Token-Kelas	<i>Black Box</i>
		Hasil Pengujian Kelas-Token	

4.2.2. Pengujian Akurasi token-kelas

Pengujian nilai akurasi dan error untuk memvalidasi kelas-token antara data asli dan data prediksi. Validasi dilakukan antara data_testing.csv dari hasil plotting diluar sistem sebagai data asli dan data hasil pengujian kelas-token sebagai data prediksi. Pengujian dilakukan sebanyak 40 dokumen, dan keseluruhan dokumen yang telah dilakukan pengujian akan dihitung nilai rata-rata akurasi dan error. Berikut hasil pengujian nilai akurasi dan error dapat dilihat pada Tabel 4.5.

Tabel 4.5 Pengujian Nilai Akurasi Token-Kelas

No	Nama Dokumen	Akurasi
1	1.txt	65%
2	2.txt	77%
3	3.txt	43%
4	4.txt	72%
5	5.txt	58%
6	6.txt	66%
7	7.txt	66%
8	8.txt	73%
9	9.txt	72%
10	10.txt	53%
11	11.txt	72%
12	12.txt	72%
13	13.txt	65%
14	14.txt	52%
15	15.txt	72%

16	16.txt	80%
17	17.txt	52%
18	18.txt	55%
19	19.txt	63%
20	20.txt	71%
21	21.txt	53%
22	22.txt	56%
23	23.txt	74%
24	24.txt	60%
25	25.txt	69%
26	26.txt	65%
27	27.txt	61%
28	28.txt	66%
29	29.txt	55%
30	30.txt	74%
31	31.txt	57%
32	32.txt	54%
33	33.txt	73%
34	34.txt	60%
35	35.txt	64%
36	36.txt	68%
37	37.txt	69%
38	38.txt	57%
39	39.txt	69%
40	40.txt	70%
	Rata-rata	64%

4.2.3. Pengujian Akurasi Kelas-Token

Pengujian Kelas-Token dilakukan berdasarkan hasil *plotting* per kelas yang sudah dilakukan sebelumnya. Berikut hasil pengujian kelas-token dapat dilihat pada Tabel 4.6

Tabel 4.6 Pengujian Nilai Akurasi dari Kelas-Token

No	Nama Dokumen	Akurasi
1	1.txt	6%
2	2.txt	0%
3	3.txt	0%
4	4.txt	0%
5	5.txt	0%
6	6.txt	0%
7	7.txt	0%
8	8.txt	0%
9	9.txt	0%
10	10.txt	0%
11	11.txt	6%
12	12.txt	0%
13	13.txt	0%
14	14.txt	6%
15	15.txt	0%
16	16.txt	0%
17	17.txt	0%
18	18.txt	6%
19	19.txt	0%
20	20.txt	0%

21	21.txt	0%
22	22.txt	0%
23	23.txt	0%
24	24.txt	0%
25	25.txt	0%
26	26.txt	0%
27	27.txt	0%
28	28.txt	0%
29	29.txt	0%
30	30.txt	0%
31	31.txt	0%
32	32.txt	0%
33	33.txt	0%
34	34.txt	0%
35	35.txt	0%
36	36.txt	6%
37	37.txt	0%
38	38.txt	0%
39	39.txt	0%
40	40.txt	0%
	Rata-rata	1%

4.2.4. Analisis Hasil Pengujian

Pengujian nilai akurasi dan error untuk memvalidasi antara token-kelas pada data pembandingan dan data prediksi. Validasi dilakukan antara data_testing.csv dari hasil plotting diluar sistem sebagai data asli dan data hasil pengujian kelas-token sebagai data prediksinya.

Nilai tinggi rendahnya akurasi dan error yang diperoleh memiliki beberapa penyebab. Berikut analisis yang dilakukan terhadap hasil pengujian akurasi dan error dengan konsep token-kelas.

Analisis hasil pengujian dengan konsep token-kelas :

1. Karena data masukan yang digunakan oleh sistem berupa file txt sehingga terdapat kekurangan. seharusnya dapat menggunakan format .docx supaya dapat menggunakan fitur tambahan seperti mendeteksi bold, italic.
2. Kenaikan ataupun penurunan nilai akurasi dan error disebabkan oleh kualitas dari file PDF yang dihasilkan pada proses scanning sebelumnya.
3. Nilai akurasi yang rendah diperoleh karena pada dokumen testing terdapat beberapa simbol tidak beraturan contohnya seperti adanya token yang tidak jelas "'V-'" ataupun "!'PVW'!" yang dihasilkan oleh proses konversi sebelumnya.
4. Berdasarkan pengamatan yang dilakukan, hasil dari prediksi token kelasmenjadi rendah disebabkan karena nilai tertinggi dari perhitungan induksi selalu pada kelas 14 (abstrak)