

## BAB 4

### IMPLEMENTASI DAN PENGUJIAN

#### 4.1. Implementasi Sistem

Pada proses ini akan dijelaskan mengenai beberapa tahapan pengujian yang akan dilakukan terhadap sistem ekstraksi informasi menggunakan algoritma *CRF* berdasarkan yang telah dilakukan sebelumnya.

##### 4.1.1. Implementasi Perangkat Keras

Berdasarkan percobaan sebelumnya terhadap beberapa perangkat keras untuk menangani proses ekstraksi informasi menggunakan algoritma *CRF*. 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 *CRF* 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 analisis yang telah dilakukan sebelumnya pada 3.3.1.2.

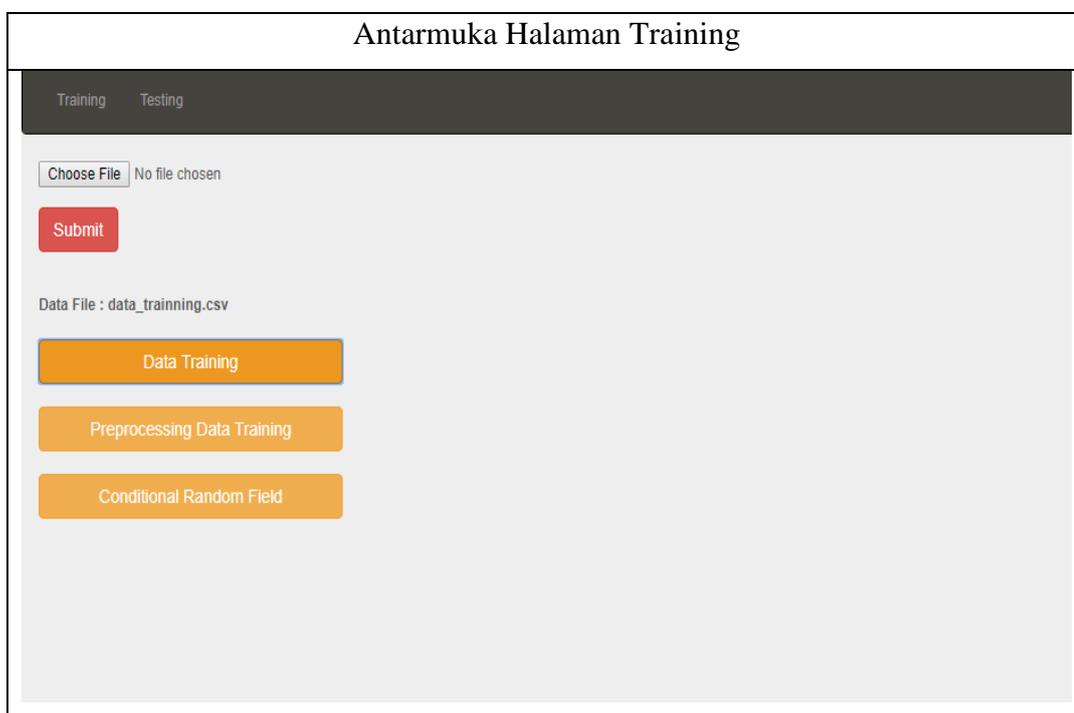
#### 4.1.3. Implementasi Antarmuka

Perancangan antarmuka yang telah dirancang sebelumnya pada 3.5 diimplementasikan terhadap sistem ekstraksi informasi menggunakan algoritma *CRF* 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>	testing.php

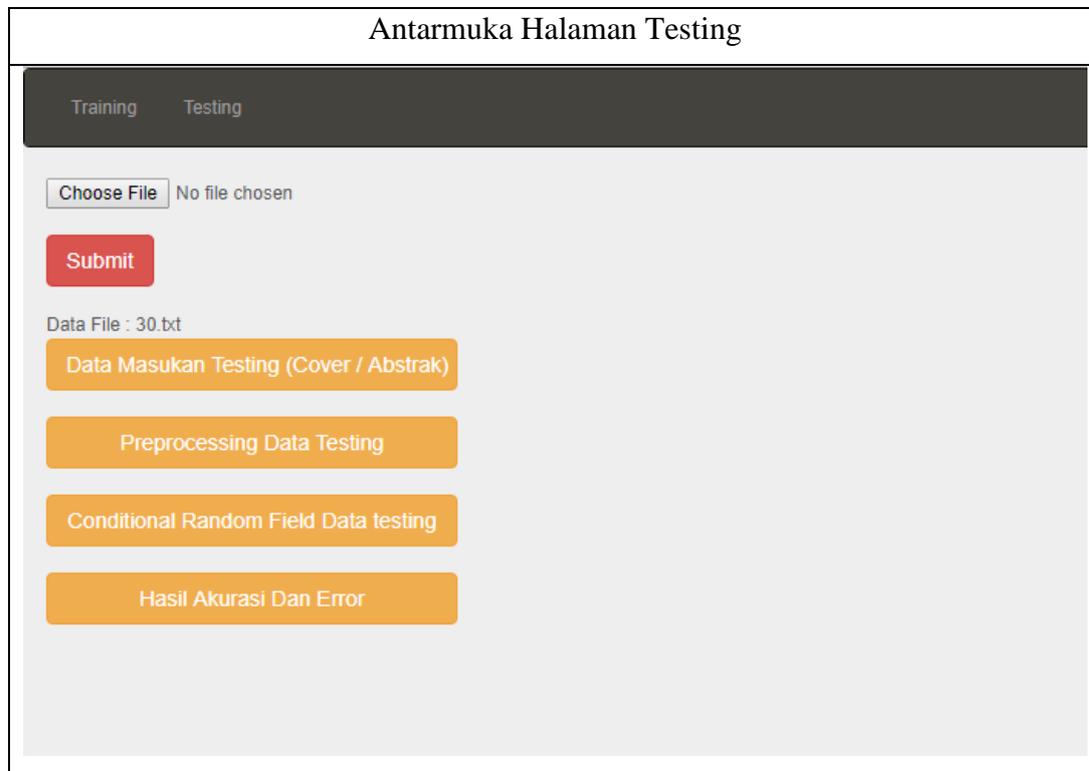
Pada tabel 4.3 terdapat dua halaman antarmuka sistem ekstraksi informasi menggunakan algoritma *CRF*. Kedua halaman tersebut telah menjadi kecukupan untuk pengguna dalam memahami proses yang berjalan pada sistem ekstraksi informasi. Berikut antarmuka pada halaman training.



**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 *CRF*.

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



**Gambar 4.2 Antarmuka Halaman Testing**

Pada gambar 4.2 antarmuka halaman 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 CRF 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*, maupun terhadap proses yang terjadi. Berikut skenario pengujian fungsionalitas dapat 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	<i>Black Box</i>
		Ekstraksi Fitur	
3	<i>Training CRF</i>	Menghitung Node & Edge Potensial	<i>Black Box</i>
		Menghitung <i>Forward</i> & <i>Backward Pass</i>	
		Menghitung Probabilitas Node & Edge	
		Menghitung Gradien	
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	<i>Black Box</i>
		Tokenisasi	
		Ekstraksi Fitur	
7	<i>Testing LVQ</i>	Menghitung Node & Edge Potensial	<i>Black Box</i>
		Menghitung <i>Maximal Forward Pass</i>	
		<i>Backtracking</i>	
8	Evaluasi	Hasil Pengujian Token-Kelas	<i>Black Box</i>
		Hasil Pengujian Kelas-Token	

#### 4.2.1.2. Skenario Pengujian Akurasi dan Error

Skenario pengujian nilai akurasi dan error memvalidasi kelas token antara data asli dan kelas prediksi. Validasi dilakukan antara *data\_testing.csv* hasil

plotting diluar sistem sebagai data asli dan hasil pengujian kelas-token sebagai data prediksi. Jumlah dokumen yang diuji sebanyak 40 dokumen bervariasi.

#### **4.2.2. Pengujian**

Pengujian dilakukan terhadap fungsionalitas serta nilai akurasi dan error pada sistem ekstraksi informasi menggunakan algoritma *Conditional Random Field* (CRF).

##### **4.2.2.1. Pengujian Fungsionalitas**

Pengujian fungsionalitas yang telah dilakukan menggunakan metode *blackbox*. Metode tersebut memiliki konsep untuk melakukan validasi terhadap sistem dengan status dapat diterima dan tidak diterima. Berikut pengujian fungsionalitas dapat dilihat pada tabel.

**Tabel 4.5 Pengujian Fungsionalitas**

No	Nama Aktifitas	Data Masukan	Keluaran yang diharapkan	Hasil Pengamatan	Kesimpulan
1	Menampilkan data masukan <i>preprocessing data training</i>	<i>data_training.csv</i>	Setiap data pada <i>data_training.csv</i> muncul dalam bentuk tabel <i>data masukan preprocessing training.</i>	Tidak data yang hilang ketika disajikan dalam bentuk tabel	[√] Diterima [ ] Tidak Diterima
2	Tokenisasi ( <i>Preprocessing data training</i> )	Barisan data dari tabel data masukan <i>preprocessing training</i>	Data telah berbentuk barisan token, disertai dengan kelasnya.	Terdapat beberapa simbol tidak beraturan pada sebagian Dokumen	[√] Diterima [ ] Tidak Diterima
3	Ekstraksi Fitur ( <i>Preprocessing data training</i> )	Token kata, angka, dan simbol	Setiap token kata, angka, dan simbol telah memiliki masing – masing bobot fitur	Token kata, angka, dan simbol berhasil diberikan masing-masing bobot fitur	[√] Diterima [ ] Tidak Diterima
4	Menghitung Node & Edge Potensial	Hasil Ekstraksi Fitur Data Training	Nilai node & edge potensial yang benar dan sesuai.	Menampilkan Nilai node dan edge potensial yang benar dan sesuai.	[√] Diterima [ ] Tidak Diterima
5	Menghitung <i>Forward</i> & <i>Backward Pass</i>	Nilai Node & Edge Potensial	Nilai <i>forward</i> & <i>backward pass</i> yang benar.	Menampilkan nilai <i>forward</i> & <i>backward pass</i> yang benar.	[√] Diterima [ ] Tidak Diterima
6	Menghitung Probabilitas Node & Edge	Nilai Node & Edge Potensial, nilai <i>forward</i> & <i>backward</i>	Nilai Probabilitas node & edge yang benar dan sesuai.	Menampilkan nilai Probabilitas node & edge yang benar dan sesuai.	[√] Diterima [ ] Tidak Diterima
7	Hitung Gradien	Hasil ekstraksi fitur, probabilitas <i>node</i> dan <i>edge</i>	Nilai gradien dan parameter yang benar, berdasarkan hasil ekstraksi fitur serta nilai probabilitas <i>node</i> dan <i>edge</i> yang tepat.	Menampilkan Nilai gradien dan parameter yang benar, berdasarkan hasil ekstraksi fitur serta nilai probabilitas <i>node</i> dan <i>edge</i> yang tepat	[√] Diterima [ ] Tidak Diterima
8	Pembaharuan Bobot	Nilai parameter	Bobot akhir diperoleh dari proses <i>training.</i>	Parameter Berhasil diperbaharui	[√] Diterima [ ] Tidak Diterima
9	Menampilkan bobot akhir	<i>Database</i> bobot_akhir	Bobot akhir ditampilkan dalam bentuk tabel.	<i>Database</i> bobot akhir berhasil diakses oleh sistem untuk	[√] Diterima [ ] Tidak Diterima

				menampilkan data bobot akhir	
10	Menampilkan data masukan <i>preprocessing data testing</i>	<i>File</i> sampul.txt dan <i>file</i> abstrak.txt	Setiap data pada <i>file</i> sampul.txt dan abstrak.txt muncul dalam bentuk <i>textfield data masukan preprocessing testing</i> .	Tidak data yang hilang ketika disajikan dalam bentuk <i>textfield</i>	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Tidak Diterima
11	Segmentasi Data	<i>File</i> sampul.txt dan <i>file</i> abstrak.txt	Menghasilkan barisan data per segmen.	Segmen data per baris berhasil ditampilkan pada sistem	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Tidak Diterima
12	Tokenisasi ( <i>Preprocessing data testing</i> )	Hasil barisan per segmen	Data telah berbentuk barisan token	Terdapat beberapa simbol tidak beraturan pada sebagian dokumen	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Tidak Diterima
13	Ekstraksi Fitur ( <i>Preprocessing data testing</i> )	Token kata, angka, dan simbol.	Setiap token kata, angka, dan simbol telah memiliki masing – masing bobot fitur	Token kata, angka, dan simbol berhasil diberikan masing-masing bobot fitur	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Tidak Diterima
14	Menampilkan bobot akhir ( <i>Training CRF</i> )	<i>Database</i> bobot_akhir	Bobot akhir ditampilkan dalam bentuk tabel.	<i>Database</i> bobot akhir berhasil diakses oleh sistem untuk menampilkan data bobot akhir	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Tidak Diterima
15	Hitung <i>node</i> dan <i>edge</i> potensial	Hasil ekstraksi fitur data <i>testing</i> , nilai parameter	Nilai <i>node</i> dan <i>edge</i> potensial yang benar, dan sesuai dengan hasil ekstraksi fitur	Menampilkan Nilai <i>node</i> dan <i>edge</i> potensial yang benar, dan sesuai dengan hasil ekstraksi fitur	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Tidak Diterima
16	Hitung <i>maximal forward pass</i>	<i>Node</i> dan <i>edge</i> potensial	Nilai <i>maximal forward pass</i> yang benar dan mengacu pada nilai <i>node</i> dan <i>edge</i> potensial yang tepat	Menampilkan Nilai <i>maximal forward pass</i> yang benar dan mengacu pada nilai <i>node</i> dan <i>edge</i> potensial yang tepat	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Tidak Diterima
17	Hitung <i>backtracking pass</i>	<i>Node</i> potensial, nilai <i>maximal forward pass</i>	Nilai <i>backtracking pass</i> yang benar, dan mengacu pada nilai <i>node</i> potensial dan <i>maximal forward pass</i> yang tepat	Menampilkan Nilai <i>backtracking pass</i> yang benar, dan mengacu pada nilai <i>node</i> potensial dan <i>maximal forward pass</i> yang tepat	<input checked="" type="checkbox"/> Diterima <input type="checkbox"/> Tidak Diterima
18	Hasil Pengujian	<i>data_testing..csv</i>	Menampilkan hasil pengujian	Hasil dari setiap token kata, angka,	<input checked="" type="checkbox"/> Diterima

	Token-Kelas	dan hasil klasifikasi token kata, angka, dan simbol.	token-kelas, nilai akurasi dan error.	dan simbol yang terklasifikasikan berhasil ditampilkan dan menghasilkan nilai akurasi dan error	<input type="checkbox"/> Tidak Diterima
--	-------------	--	---------------------------------------	---	---

#### 4.2.2.2. Pengujian Nilai Akurasi dan Error

Pengujian dilakukan dengan konsep kelas-token, seperti pada 3.2 sebanyak 40 dokumen *testing*. Keseluruhan dokumen yang telah memiliki nilai akurasi akan dihitung rata-rata nilai akurasi dan errornya. Berikut ini pengujian nilai akurasi dan error dapat dilihat pada tabel 4.6.

**Tabel 4.6 Pengujian Nilai Akurasi dan Error dengan Konsep Kelas-Token**

No	Nama Dokumen	Akurasi
1	1.txt	5%
2	2.txt	3%
3	3.txt	6%
4	4.txt	5%
5	5.txt	4%
6	6.txt	5%
7	7.txt	2%
8	8.txt	1%
9	9.txt	4%
10	10.txt	6%
11	11.txt	3%
12	12.txt	3%
13	13.txt	4%
14	14.txt	4%
15	15.txt	3%
16	16.txt	2%
17	17.txt	6%
18	18.txt	5%
19	19.txt	6%
20	20.txt	5%
21	21.txt	5%
22	22.txt	5%
23	23.txt	2%
24	24.txt	4%
25	25.txt	7%
26	26.txt	6%
27	27.txt	5%

28	28.txt	3%
29	29.txt	5%
30	30.txt	4%
31	31.txt	5%
32	32.txt	5%
33	33.txt	2%
34	34.txt	6%
35	35.txt	6%
36	36.txt	8%
37	37.txt	5%
38	38.txt	2%
39	39.txt	5%
40	40.txt	6%
<b>Rata-Rata</b>		4%

#### 4.2.3. Analisis Hasil Pengujian

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. Dapat dilihat pada lembar lampiran Tabel Persentase Akurasi Error Token-Kelas.

Analisis hasil pengujian dengan konsep token-kelas :

1. Karena data masukan yang digunakan oleh sistem adalah file teks, maka terdapat keterbatasan pada fitur yang digunakan.
2. Kenaikan ataupun penurunan nilai akurasi dan error disebabkan oleh kualitas *file PDF* yang dihasilkan pada proses *scanning* sebelumnya.
3. Nilai akurasi yang rendah diperoleh karena pada dokumen *testing* terdapat beberapa simbol tidak beraturan yang dihasilkan oleh proses konversi sebelumnya.
4. Berdasarkan pengamatan yang dilakukan, dampak yang mempengaruhi pada penurunan nilai akurasi selalu disebabkan oleh fitur *ORGANIZATION* yang dikhususkan untuk memberikan pembobotan pada kategori Program Studi, Fakultas, dan Universitas. Karena, pada algoritma *CRF* memiliki

jaringan kompetitif, maka jika Program Studi diasumsikan menang dengan pembobotan yang sama antara Fakultas dan Universitas, Fakultas dan Universitas akan terklasifikasikan pada kelas yang sama seperti Program Studi.

5. Minimnya penggunaan ekstraksi fitur yang dapat mengenali beberapa token secara spesifik terhadap banyaknya jumlah kelas, yaitu 16 kelas, sehingga untuk memberikan bobot pada setiap token tidak signifikan.