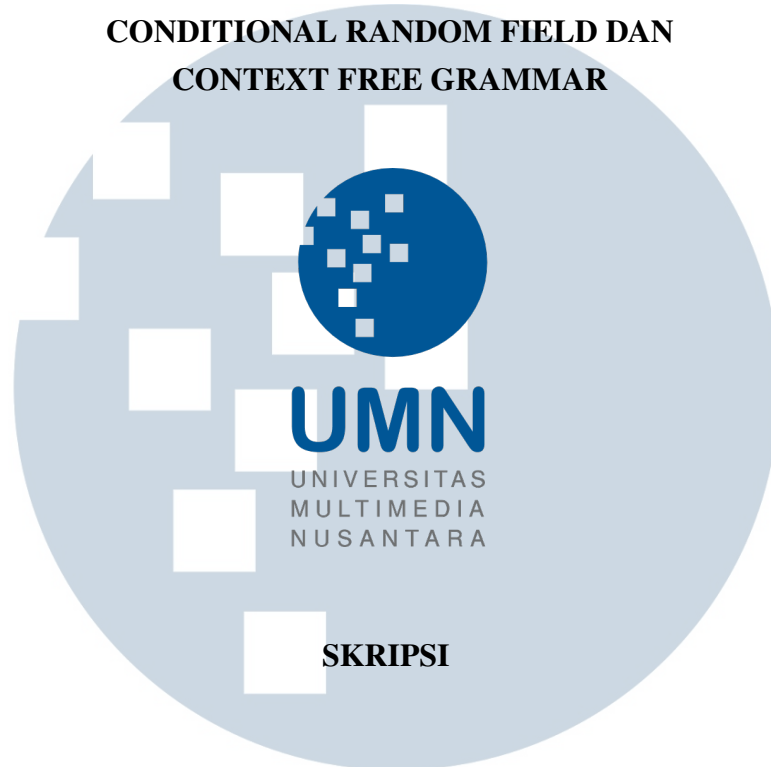


**PENGEMBANGAN U-TAPIS UNTUK PENDETEKSI KESALAHAN
SINTAKSIS KALIMAT BAHASA INDONESIA MENGGUNAKAN
CONDITIONAL RANDOM FIELD DAN
CONTEXT FREE GRAMMAR**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Denn Sebastian Emmanuel

0000036981

UMN

UNIVERSITAS

MULTIMEDIA

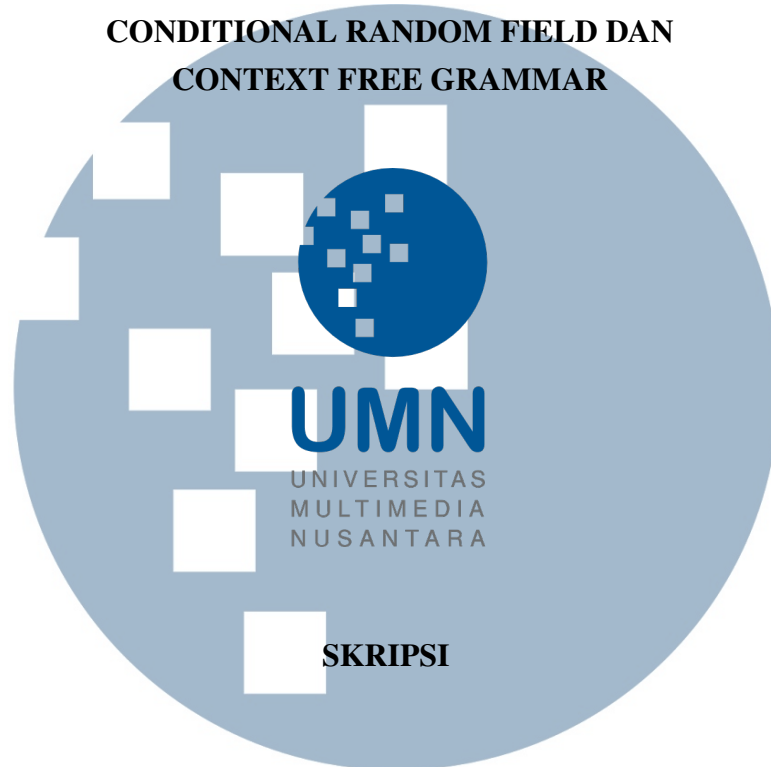
NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

**PENGEMBANGAN U-TAPIS UNTUK PENDETEKSI KESALAHAN
SINTAKSIS KALIMAT BAHASA INDONESIA MENGGUNAKAN
CONDITIONAL RANDOM FIELD DAN
CONTEXT FREE GRAMMAR**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Denn Sebastian Emmanuel
0000036981

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2023

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Denn Sebastian Emmanuel
Nomor Induk Mahasiswa : 00000036981
Program Studi : Informatika

Skripsi dengan judul:

Pengembangan U-Tapis untuk Pendeteksi Kesalahan Sintaksis Kalimat Bahasa Indonesia Menggunakan Conditional Random Field dan Context Free Grammar

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 5 Juni 2023



(Denn Sebastian Emmanuel)

UNM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

PENGEMBANGAN U-TAPIS UNTUK PENDETEKSI KESALAHAN SINTAKSIS KALIMAT BAHASA INDONESIA MENGGUNAKAN CONDITIONAL RANDOM FIELD DAN CONTEXT FREE GRAMMAR

oleh

Nama : Denn Sebastian Emmanuel
NIM : 00000036981
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Senin, 12 Juni 2023
Pukul 08.00 s/s 10.00 dan dinyatakan
LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang



(Farica Perdana Putri, S.Kom., M.Sc.)

NIDN: 0331019301

Penguji



(Dennis Gunawan, S.Kom., M.Sc)

NIDN: 0320059001

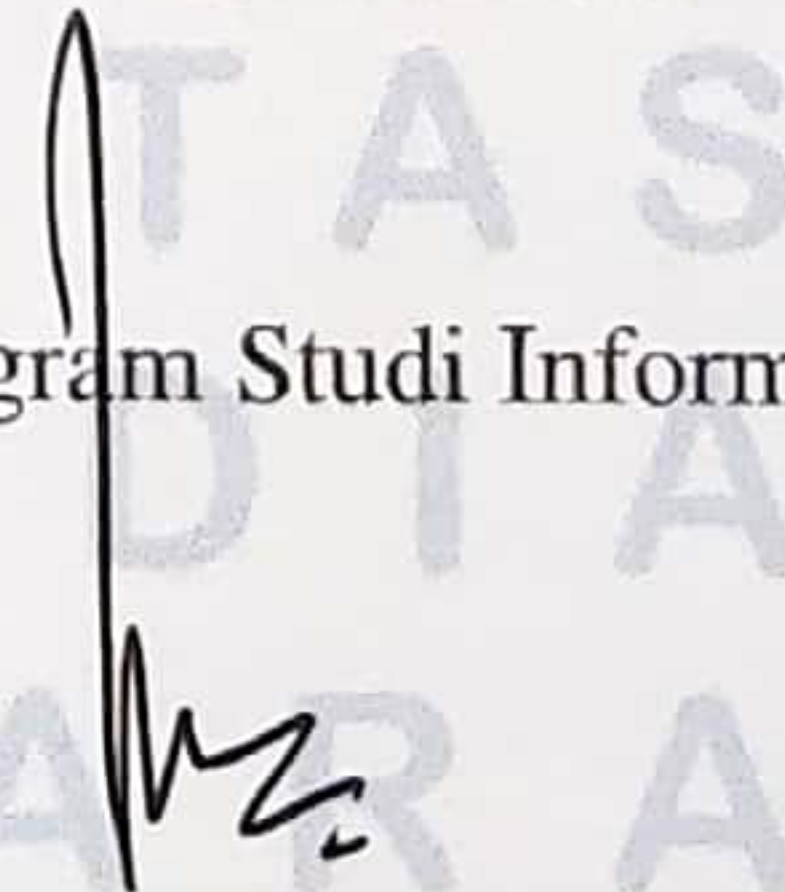
Pembimbing



(Marlinda Vasty Overbeek, S.Kom.,
M.Kom)

NIDN: 0818038501

Ketua Program Studi Informatika,



(Marlinda Vasty Overbeek, S.Kom.,
M.Kom.)

NIDN: 0818038501

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Denn Sebastian Emmanuel
NIM : 00000036981
Program Studi : Informatika
Fakultas : Teknik dan Informatika
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada **Universitas Multimedia Nusantara** Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

**PENGEMBANGAN U-TAPIS UNTUK PENDETEKSI KESALAHAN
SINTAKSIS KALIMAT BAHASA INDONESIA MENGGUNAKAN
CONDITIONAL RANDOM FIELD DAN
CONTEXT FREE GRAMMAR**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 5 Juni 2023

Yang menyatakan



Denn Sebastian Emmanuel

Halaman Persembahan / Motto

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Pengembangan U-Tapis untuk Pendeteksi Kesalahan Sintaksis Kalimat Bahasa Indonesia Menggunakan Conditional Random Field dan Context Free Grammar dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara dan pembimbing skripsi yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan, dan motivasi atas terselesaikannya skripsi ini.
4. Ibu Niknik Mediyawati, S.Pd., M.Hum., selaku Dosen Jurnalistik Universitas Multimedia Nusantara yang telah membimbing saya di bidang tata bahasa Indonesia yang baik dan benar selama pengembangan algoritma U-Tapis ini.
5. Kepada Bapak Abdullah Ahmad, sebagai *Supervisor* dari Perusahaan Tribun News selama pengembangan algoritma U-Tapis ini.
6. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral sehingga saya dapat menyelesaikan skripsi ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 5 Juni 2023



Denn Sebastian Emmanuel

**PENGEMBANGAN U-TAPIS UNTUK PENDETEKSI KESALAHAN
SINTAKSIS KALIMAT BAHASA INDONESIA MENGGUNAKAN
CONDITIONAL RANDOM FIELD DAN
CONTEXT FREE GRAMMAR**

Denn Sebastian Emmanuel

ABSTRAK

Saat pembuatan proyek skripsi ini, proses pembuatan artikel berita baru di Tribun News masih menggunakan teknik semi-digital (*word processing tool* dan *internal website*) dan belum mengimplementasikan teknologi algoritma dan *machine learning*. Hal ini berpotensi menyebabkan artikel berita baru tersebut memiliki kesalahan sintaksis kalimat Bahasa Indonesia. Pengembangan Algoritma U-Tapis Pendeteksi Kesalahan Sintaksis Kalimat Bahasa Indonesia dilakukan untuk membantu memastikan kebenaran sintaksis artikel berita baru di Tribun News. Algoritma ini dapat menerima data artikel berita dan mengembalikan kebenaran sintaksis dari setiap kalimat berita tersebut. Algoritma U-Tapis Pendeteksi Kesalahan Sintaksis Kalimat adalah kombinasi dari algoritma *Conditional Random Field* (CRF) dan *Context Free Grammar* (CFG). Algoritma *Conditional Random Field* (CRF) digunakan untuk melakukan *part-of-speech tagging* dan algoritma *Context Free Grammar* (CFG) digunakan untuk melakukan *parsing* pada hasil *part-of-speech tagging* tersebut serta menguji kebenaran sintaksis setiap kalimat di dalam berita tersebut. Algoritma U-Tapis Pendeteksi Kesalahan Sintaksis Kalimat telah berhasil di-*deploy* dalam bentuk *Python Flask API (Application Programming Interface)*. Hasil evaluasi Algoritma U-Tapis Pendeteksi Kesalahan Sintaksis Kalimat menunjukkan nilai *accuracy* sebesar 0,91, *macro-averaged precision* sebesar 0,83, *macro-averaged recall* sebesar 0,84, *macro-averaged F1-Score* sebesar 0,84, *weighted-averaged precision* sebesar 0,91, *weighted-averaged recall* sebesar 0,91, *weighted-averaged F1-score* sebesar 0,91, dan rata-rata *runtime* sebesar 2,1374 detik per kalimat.

Kata kunci: *Conditional Random Field*, *Context Free Grammar*, sintaksis kalimat bahasa Indonesia, Tribun News, U-Tapis

DEVELOPMENT OF U-TAPIS FOR INDONESIAN SENTENCE SYNTAX ERROR DETECTION USING CONDITIONAL RANDOM FIELD AND CONTEXT FREE GRAMMAR

Denn Sebastian Emmanuel

ABSTRACT

During the development of this thesis project, Tribun News had still been using a semi-digital method (word processing tool and internal website) and hadn't implemented algorithm and machine learning technology in their news making process. This could potentially lead to the occurrences of Indonesian sentence syntax errors in newly created news article. The development of U-Tapis Indonesian Sentence Syntax Error Detection Algorithm is done to help ensure the syntatic correctness of Tribun News' newly created news article. This algorithm can accept a news article text data and return the syntatic correctness of each news sentence. U-Tapis Indonesian Sentence Syntax Error Detection Algorithm is a combination of Conditional Random Field (CRF) algorithm and Context Free Grammar (CFG) algorithm. Conditional Random Field (CRF) algorithm is used to perform part-of-speech tagging and Context Free Grammar (CFG) algorithm is used to parse the results of the part-of-speech tagging and test the syntatic correctness of each news sentence. U-Tapis Indonesian Sentence Syntax Error Detection Algorithm has been successfully deployed as a Python Flask API (Application Programming Interface). The evaluation result of U-Tapis Indonesian Sentence Syntax Error Detection Algorithm shows an accuracy value of 0.91, macro-averaged precision value of 0.83, macro-averaged recall value of 0.84, macro-averaged F1-score value of 0.84, weighted-averaged precision value of 0.91, weighted-averaged recall value of 0.91, weighted-averaged F1-score value of 0.91, and average runtime value of 2.1374 seconds per sentence.

Keywords: *Conditional Random Field, Context Free Grammar, Indonesian sentence syntax, Tribun News, U-Tapis*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR KODE	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Permasalahan	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	6
2.1 Tribun News	6
2.2 Sintaksis Kalimat Bahasa Indonesia	6
2.3 Kata	7
2.3.1 Nomina	7
2.3.2 Verba	8
2.3.3 Adjektiva	9
2.3.4 Adverbia	10
2.3.5 Preposisi	11
2.3.6 Pronomina	12
2.3.7 Numeralia	13
2.3.8 Artikula/Artikel	14
2.3.9 Konjungsi	15
2.3.10 Interjeksi	26
2.3.11 Partikel	28
2.4 Frasa	29
2.4.1 Frasa Nominal	30
2.4.2 Frasa Verbal	35
2.4.3 Frasa Adjektival	35
2.4.4 Frasa Preposisional	36
2.4.5 Frasa Numeral	37
2.4.6 Frasa Pronominal	37
2.4.7 Frasa Adverbial	38
2.5 Klausa	38
2.6 Kalimat	39
2.6.1 Unsur Kalimat	39
2.6.2 Jenis Kalimat	42
2.7 Aposisi dan Suplementasi	44

2.7.1	Aposisi	44
2.7.2	Suplementasi	45
2.8	Text Normalization	45
2.8.1	Tokenization	45
2.8.2	Word Normalization	46
2.8.3	Sentence Segmentation	47
2.9	Conditional Random Field	48
2.10	Context Free Grammar	51
2.11	Teknik Parsing	52
2.11.1	Top-Down Parsing	52
2.11.2	Bottom-Up Parsing	52
2.11.3	Left-Corner Parsing	53
2.11.4	Contoh Top-Down Parsing, Bottom-Up Parsing, dan Left-Corner Parsing	53
2.12	Accuracy, Precision, Recall, F1-Score	57
2.12.1	Confusion Matrix	57
2.12.2	Accuracy	59
2.12.3	Precision dan Recall	59
2.12.4	F1-Score atau F-Measure	61
2.12.5	Macro-Averaged dan Weighted Averaged	61
2.13	Run Time	63
2.14	Python Flask Web Framework	64
BAB 3	METODOLOGI PENELITIAN	65
3.1	Pengumpulan Data	65
3.2	Data Preprocessing	66
3.3	Pendefinisian <i>Class</i> atau Tag	71
3.4	Pengembangan Algoritma Context Free Grammar	71
3.5	Pelabelan Training Data	72
3.6	Training Conditional Random Field	72
3.7	Evaluasi Model Conditional Random Field (CRF) dan <i>Rule-Rule</i> Context Free Grammar (CFG)	74
3.8	Deployment sebagai Application Programming Interface/API	74
3.9	Testing Application Programming Interface/API	76
BAB 4	HASIL DAN DISKUSI	78
4.1	Spesifikasi Sistem	78
4.2	Pengumpulan Data	79
4.3	Pengembangan Preprocessing Function	80
4.4	Pendefinisian Class/Tag untuk Conditional Random Field dan Context Free Grammar	92
4.5	Pengembangan Algoritma Context Free Grammar	95
4.6	Pelabelan Training Data	98
4.7	Training Model Conditional Random Field	100
4.8	Deployment sebagai Application Programming Interface/API	104
4.9	Hasil Evaluasi Model Conditional Random Field dan Algoritma Context Free Grammar	110
4.10	Hasil Evaluasi Run Time Application Programming Interface/API	116
BAB 5	SIMPULAN DAN SARAN	121
5.1	Simpulan	121
5.2	Saran	121
DAFTAR PUSTAKA		123

DAFTAR GAMBAR

Gambar 2.1	Logo Tribun News	6
Gambar 2.2	Diagram alir <i>Tokenization</i>	46
Gambar 2.3	Diagram alir <i>Word Normalization</i>	47
Gambar 2.4	Diagram alir <i>Sentence Segmentation</i>	48
Gambar 2.5	Diagram alir <i>Conditional Random Field</i>	49
Gambar 2.6	Diagram alir <i>Context Free Grammar</i>	51
Gambar 2.7	Contoh <i>Top-Down Parsing</i>	55
Gambar 2.8	Contoh <i>Bottom-Up Parsing</i>	56
Gambar 2.9	Contoh <i>Left-Corner Parsing</i>	57
Gambar 2.10	Confusion matrix dua <i>class</i>	58
Gambar 2.11	Confusion matrix tiga <i>class</i>	58
Gambar 2.12	Rumus-rumus pada confusion matrix dua <i>class</i>	60
Gambar 2.13	Rumus-rumus pada confusion matrix tiga <i>class</i>	61
Gambar 3.1	Diagram alir <i>metodologi penelitian</i>	65
Gambar 3.2	Diagram alir <i>preprocessing function</i> untuk CFG dan CRF	67
Gambar 3.3	Diagram alir <i>preprocessing function</i> untuk API	67
Gambar 3.4	Flowchart <i>Application Programming Interface (API)</i> dari U-Tapis Pendeteksi Kesalahan Sintaksis Kalimat	76
Gambar 3.5	Tampilan aplikasi Postman	77
Gambar 4.1	Tampilan <i>rule-rule Context Free Grammar</i> (tampilan 1)	97
Gambar 4.2	Tampilan <i>rule-rule Context Free Grammar</i> (tampilan 2)	98
Gambar 4.3	Tampilan hasil pelabelan <i>training data</i> (tampilan 1)	99
Gambar 4.4	Tampilan hasil pelabelan <i>training data</i> (tampilan 2)	100
Gambar 4.5	Tampilan contoh hasil CRF (tampilan 1)	104
Gambar 4.6	Tampilan contoh hasil CRF (tampilan 2)	104
Gambar 4.7	Tampilan API U-Tapis saat diaktifkan	105
Gambar 4.8	<i>API request</i> menggunakan <i>Postman</i>	106
Gambar 4.9	<i>API request</i> menggunakan <i>Postman (bool-only)</i>	106
Gambar 4.10	<i>API request</i> menggunakan <i>Postman (false-only)</i>	106
Gambar 4.11	Contoh data berita yang diterima API	107
Gambar 4.12	Hasil <i>preprocessing</i> untuk algoritma CRF dan CFG	107
Gambar 4.13	Hasil <i>preprocessing</i> untuk <i>API response</i>	107
Gambar 4.14	Hasil <i>part-of-speech tagging</i> dan <i>parsing</i>	108
Gambar 4.15	Contoh <i>API response</i> tanpa <i>query parameter</i>	109
Gambar 4.16	Contoh <i>API response</i> dengan <i>query parameter bool-only</i>	109
Gambar 4.17	Contoh <i>API response</i> dengan <i>query parameter false-only</i>	109
Gambar 4.18	Tampilan <i>test data</i> algoritma CRF dan CFG	110
Gambar 4.19	<i>Confusion Matrix</i> hasil evaluasi algoritma CRF	111
Gambar 4.20	<i>Confusion Matrix</i> hasil evaluasi algoritma CFG	114
Gambar 4.21	<i>Confusion Matrix</i> hasil evaluasi integrasi algoritma CRF dan CFG	115

DAFTAR TABEL

Tabel 4.1	<i>Class-class</i> CRF dan CFG	92
Tabel 4.2	<i>Accuracy, Precision, Recall, F1-Score</i> dari CRF	112
Tabel 4.3	<i>Accuracy, Precision, Recall, F1-Score</i> dari CFG	114
Tabel 4.4	<i>Accuracy, Precision, Recall, F1-Score</i> dari CRF dan CFG .	115
Tabel 4.5	<i>Runtime</i> dari API U-Tapis	116



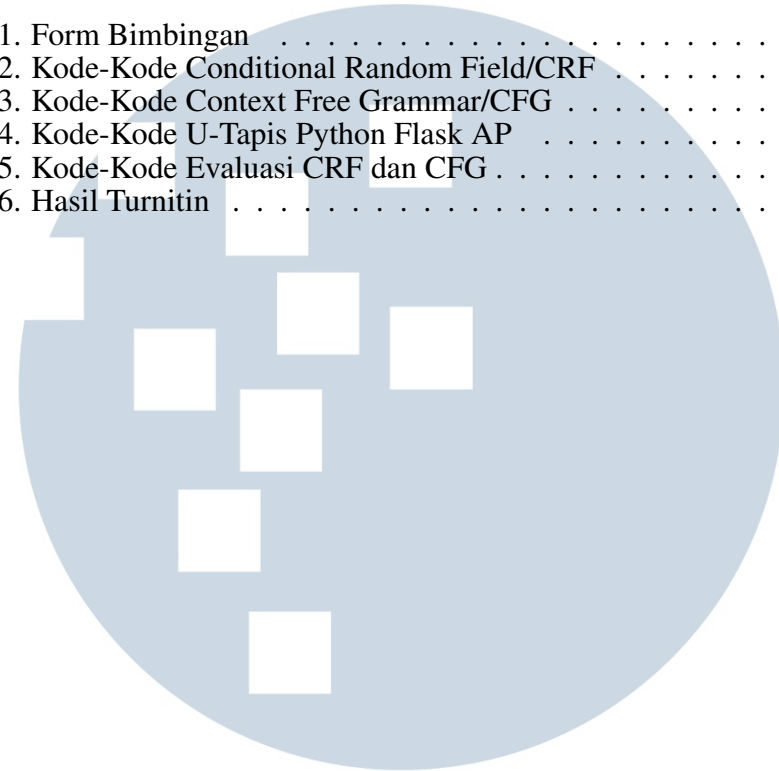
DAFTAR KODE

2.1	<i>Rule-rule</i> contoh teknik parsing	53
3.1	<i>JSON API response</i>	74
3.2	<i>Bool-Only JSON API response</i>	75
3.3	<i>False-Only JSON API response</i>	75
4.1	Kode <i>preprocessing function</i> untuk CRF dan CFG	80
4.2	Kode <i>preprocessing function</i> untuk API	86
4.3	Kode <i>Left-Corner Parsing</i> CFG	95
4.4	<i>Feature function</i> CRF	100



DAFTAR LAMPIRAN

Lampiran 1. Form Bimbingan	124
Lampiran 2. Kode-Kode Conditional Random Field/CRF	127
Lampiran 3. Kode-Kode Context Free Grammar/CFG	127
Lampiran 4. Kode-Kode U-Tapis Python Flask AP	127
Lampiran 5. Kode-Kode Evaluasi CRF dan CFG	127
Lampiran 6. Hasil Turnitin	128



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA