

**PENERAPAN RECURRENT NEURAL NETWORK (RNN) PADA
TRANSLITERASI TEKS ARAB KE TEKS LATIN INDONESIA
BERBASIS OPTICAL CHARACTER RECOGNITION (OCR)**

TUGAS AKHIR

Oleh :

NOVIANI

NIM. 1810817120014



PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS LAMBUNG MANGKURAT

BANJARMASIN

2024

**PENERAPAN RECURRENT NEURAL NETWORK (RNN) PADA
TRANSLITERASI TEKS ARAB KE TEKS LATIN INDONESIA
BERBASIS OPTICAL CHARACTER RECOGNITION (OCR)**

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat Sarjana Strata-1 Teknologi Informasi

Oleh :

NOVIANI

NIM. 1810817120014



PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS LAMBUNG MANGKURAT

BANJARMASIN, JANUARI 2024

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Noviani
NIM : 1810817120014
Fakultas : Teknik
Program Studi : Teknologi Informasi
Judul Tugas Akhir : Penerapan Recurrent Neural Network (RNN) pada
Transliterasi Teks Arab ke Teks Latin Indonesia
Berbasis Optical Character Recognition (OCR)
Pembimbing Utama : Andreyan Rizky Baskara, S.Kom., M.Kom
Pembimbing Pendamping : Muti'a Maulida, S.Kom, M.T.I

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya, juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar rujukan.

Banjarmasin, 15 Desember 2023

Handwritten signature of Noviani in black ink, featuring stylized initials and the name 'Novi' with a star above the 'i'.

NOVIANI

NIM. 1810817120014

LEMBAR PERSETUJUAN

PERSETUJUAN TUGAS AKHIR

PENERAPAN RECURRENT NEURAL NETWORK (RNN) PADA
TRANSLITERASI TEKS ARAB KE TEKS LATIN INDONESIA BERBASIS
OPTICAL CHARACTER RECOGNITION (OCR)

OLEH
NOVIANI
NIM. 1810817120014

Telah diperiksa dan terpenuhi semua persyaratan akademik, administrasi, dan
disetujui untuk dipertahankan di hadapan dewan penguji

Banjarmasin, 6 Desember 2023

Pembimbing Utama,



Andreyan Rizky Baskara, S.Kom., M.Kom
NIP. 19930703 201903 1 011

Pembimbing Pendamping,



Muti'a Maulida, S.Kom, M.T.I
NIP. 19881027 201903 2 013

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

SKRIPSI PROGRAM STUDI S-1 TEKNOLOGI INFORMASI

Penerapan Recurrent Neural Network (RNN) pada Transliterasi Teks Arab ke
Teks Latin Indonesia Berbasis Optical Character Recognition (OCR)

oleh

Noviani (1810817120014)

Telah dipertahankan di depan Tim Penguji pada 2 Januari 2024 dan dinyatakan

LULUS

Komite Penguji :

Ketua : Eka Setya Wijaya, S.T., M.Kom.
NIP 198205082008011010

Anggota 1 : Nurul Fathanah Mustamin, S.Pd., M.T.
NIP 199110252019032018

Anggota 2 : Dr. Ir. Yulena Sari, S.Kom., M.Kom.
NIP 198411202015042002

Pembimbing Utama : Andreyan Rizky Baskara, S.Kom., M.Kom.
NIP 199307032019031011

Pembimbing Pendamping : Muti'a Maulida, S.Kom., M.T.I.
NIP 198810272019032013

Banjarbaru, 19 JAN 2023

diketahui dan disahkan oleh:

Wakil Dekan Bidang Akademik
Fakultas Teknik ULM,



Dzakiyahud, S.T., M.T.
NIP 401071998021001

Koordinator Program Studi
S-1 Teknologi Informasi,

Andreyan Rizky Baskara, S.Kom., M.Kom.
NIP 199307032019031011

ABSTRAK

Berdasarkan hasil riset Perguruan Tinggi Ilmu Quran (PTIQ) ada 65 persen dari 229 juta jiwa penduduk Indonesia beragama Islam belum mampu membaca tulisan arab. Transliterasi atau alih aksara dalam Kamus Besar Bahasa Indonesia adalah pengalihan suatu jenis aksara atau huruf ke aksara atau huruf yang lain. Di Indonesia, transliterasi banyak digunakan pada tulisan berbahasa arab yang biasanya ditemukan di buku-buku keagamaan, namun tidak semua tulisan berbahasa arab disertai transliterasi. Penelitian ini berguna untuk permasalahan transliterasi teks arab ke teks latin dengan menggunakan pedoman transliterasi latin Indonesia. Teknologi yang dapat melakukan *input* secara otomatis tanpa harus diketikkan dengan menggunakan OCR juga diperlukan. Mesin OCR yang dapat digunakan adalah Tesseract OCR karena mendukung berbagai Bahasa terutama Bahasa Arab dan dapat digunakan dengan mudah. Penelitian mengenai mesin penerjemah dengan metode *Recurrent Neural Network* (RNN) telah banyak dilakukan, yang mana model yang optimal adalah berbasis LSTM untuk proses transliterasi. Penelitian ini dimulai dengan pengolahan *dataset* teks arab dan teks latin sesuai pedoman di Indonesia dengan *dataset* 5000 kata yang dibagi menjadi tiga bentuk pembagian *training* dan *testing* yaitu 90:10 (4500:500), 80:20 (4000:1000), 70:30 (3500:1500). Dan dihasilkan *BLEU Score* model yang paling optimal adalah pembagian 90:10 dengan nilai 49.95%.

Kata Kunci: Arab, *BLEU Score*, Latin, *LSTM*, *RNN*, *Tesseract OCR*, Transliterasi

ABSTRACT

Based on research results from Perguruan Tinggi Ilmu Quran (PTIQ), 65 percent of Indonesia's 229 million Muslim population cannot read Arabic text. Transliteration or transfer of characters in the Big Indonesian Dictionary is the transfer of one type of script or letters to another script or letters. In Indonesia, transliteration is widely used in Arabic writing which is usually found in religious books, but not all Arabic writing is accompanied by transliteration. This research is useful for the problem of transliterating Arabic text into Latin text using Indonesian Latin transliteration guidelines. Technology that can input automatically without having to type using OCR is also needed. The OCR engine that can be used is Tesseract OCR because it supports various languages, especially Arabic and can be used easily. Research has been carried out on machine translation using the Recurrent Neural Network (RNN) method, where the optimal model is LSTM-based for the transliteration process. This research began with processing a dataset of Arabic text and Latin text according to Indonesian guidelines with a dataset of 5000 words which was divided into three forms of training and testing, namely 90:10 (4500:500), 80:20 (4000:1000), 70:30 (3500:1500). And the most optimal BLEU Score model produced was a 90:10 division with a value of 49.95%.

Keywords: Arabic, BLEU Score, Latin, LSTM, RNN, Tesseract OCR, Transliteration

HALAMAN PERSEMBAHAN

Penulis mempersembahkan Tugas Akhir ini kepada:

1. Orang tua dan saudara yang telah memberikan dukungan baik itu bantuan moral, materi dan juga doa selama proses pengerjaan Tugas Akhir ini.
2. Bapak Andreyan Rizky Baskara, S.Kom., M.Kom., selaku Dosen Pembimbing Utama, Pembimbing Akademik dan juga Koordinator Program Studi yang memberikan kesempatan waktunya untuk memberikan bimbingan, arahan dan dukungan kepada penulis dari awal hingga selesai pengerjaan Tugas Akhir ini.
3. Ibu Muti'a Maulida, S.Kom, M.T.I., selaku Pembimbing Pendamping yang memberikan kesempatan waktunya untuk memberikan bimbingan, arahan dan dukungan kepada penulis dari awal hingga selesai pengerjaan Tugas Akhir ini.
4. Seluruh Dosen beserta Staf Program Studi Teknologi Informasi yang turut membantu dan mengarahkan dari awal perkuliahan hingga selesai pengerjaan Tugas Akhir ini.
5. Teman-teman mahasiswa Program Studi Teknologi Informasi baik itu dari Angkatan 2018, kemudian kakak-kakak angkatan 2017 ke atas hingga adik-adik Angkatan 2019 hingga 2023 yang turut memberikan dukungan dan motivasi, serta pembelajarannya selama penulis menempuh Pendidikan di Program Studi Teknologi Informasi ini.
6. Mentor dan Guru serta Kakak-kakak Ideologis yang telah memberikan pendampingan, bimbingan, doa dan pelajaran serta penguatan spiritual selama pengerjaan Tugas Akhir ini.
7. Teman-teman, kakak-kakak dan Adik-adik Dakwah Kampus dan se-organisasi mahasiswa yang telah memberikan dukungan, doa dan Pelajaran selama penulis menempuh Pendidikan Sarjana ini.

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah memberikan limpahan rahmat dan hidayah-nya kepada penulis sehingga dapat menyelesaikan Tugas Akhir dengan judul “Penerapan Recurrent Neural Network (RNN) Pada Transliterasi Teks Arab ke Teks Latin Indonesia Berbasis Optical Character Recognition (OCR)”. Tugas Akhir ini disusun untuk memenuhi salah satu syarat untuk memperoleh gelar Sarjana Strata-1 Teknologi Informasi di Fakultas Teknik Universitas Lambung Mangkurat, Banjarmasin. Penulis ingin memberikan ucapan terima kasih sebesar-besarnya kepada semua pihak yang telah memberikan bantuan, dukungan, bimbingan, dan motivasi bagi penulis dalam menyelesaikan Tugas akhir ini, terutama kepada :

1. Orang tua dan saudara kandung yang sangat penulis syukuri kehadirannya di dunia ini menyertai penulis dalam memberikan kebermanfaatan dan memaksimalkan potensi sehingga penulis dapat mulai berkuliah hingga menyusun Tugas Akhir ini dengan proses yang penuh hikmah.
2. Rektor Universitas Lambung Mangkurat, Bapak Prof. Dr. Ahmad, S.E., M.Si. yang memimpin dan memajemen jalannya seluruh perkuliahan yang ada di Universitas Lambung Mangkurat.
3. Dekan Fakultas Teknik, Bapak Prof. Dr. Ir. Irphan Fitriani Radam, S.T., M.T., IPU yang telah memberikan layanan terbaik dalam perkuliahan, terkhusus pada pelaksanaan Tugas Akhir di lingkungan Fakultas Teknik.
4. Ketua Program Studi Teknologi Informasi, Pembimbing Akademik, sekaligus Dosen Pembimbing Utama, Bapak Andreyan Rizky Baskara, S.Kom., M.Kom., yang telah memberikan arahan dan motivasi kepada penulis dari awal perkuliahan sampai dengan penyelesaian Tugas Akhir ini.
5. Pembimbing Pendamping, Ibu Muti’a Maulida, S.Kom, M.T.I., yang memberikan kesempatan waktunya untuk memberikan bimbingan, arahan dan dukungan kepada penulis dari awal hingga selesai pengerjaan Tugas Akhir ini.
6. Dosen-dosen beserta staf Program Studi Teknologi Informasi yang telah mengarahkan dan teman-teman yang membantu dalam proses penyelesaian Tugas Akhir.

7. Seluruh mahasiswa Program Studi Teknologi Informasi yang telah membantu penulis dalam membentuk lingkungan perkuliahan yang memberi semangat bagi penulis dalam menyelesaikan Tugas Akhir ini.
8. Mentor, Guru, Kakak-Kakak serta Adik-Adik Ideologis baik itu dalam Dakwah Kampus, Organisasi Mahasiswa, maupun rekan dalam kegiatan kebermanfaatan yang telah membentuk lingkungan perkuliahan yang memberi semangat bagi penulis dalam menyelesaikan Tugas Akhir ini.

Penulis menyampaikan terimakasih juga kepada semua pihak yang turut membantu dalam penyelesaian laporan Tugas Akhir ini. Tugas Akhir ini telah disusun dengan optimal dan masihlah jauh dari kesempurnaan. Penulis menerima saran dan kritik yang membangun agar Tugas Akhir ini dapat memberikan banyak manfaat bagi semua orang.

Banjarmasin, Desember 2023

Penulis,

A handwritten signature in black ink, appearing to be 'Noviani', with a small star above the 'i'.

Noviani

DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR PERNYATAAN	iii
LEMBAR PERSETUJUAN	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
<i>ABSTRACT</i>	vii
HALAMAN PERSEMBAHAN	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvi
DAFTAR RIWAYAT HIDUP	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA	7
2.1 Landasan Teori	7
2.1.1 Huruf Arab	7
2.1.2 Translasi dan Transliterasi	8
2.1.3 Transliterasi Arab-Latin Indonesia	8

2.1.4	<i>Deep Learning dalam Arabic Natural Language Processing</i>	15
2.1.5	<i>Optical Character Recognition</i>	16
2.1.6	Tesseract OCR	16
2.1.7	<i>Neural Machine Translation and Transliteration Method</i>	17
2.1.8	<i>BLEU Score</i>	19
2.2	Penelitian Terkait	21
2.2.1	<i>Arabic Optical Character Recognition Software: A Review</i>	21
2.2.2	<i>Experimental Evaluation of Arabic OCR Systems</i>	22
2.2.3	<i>Machine Transliteration and Transliterated Text Retrieval: A Survey</i>	22
2.2.4	<i>Neural Machine Transliteration: Preliminary Results</i>	23
2.2.5	<i>A Deep Learning Based Approach to Transliteration</i>	24
2.2.6	<i>Phrase Based and Neural Network Translation for Text Transliteration from Arabic to Indonesian</i>	25
2.2.7	<i>A Unified Model for Arabizi Detection and Transliteration using Sequence-to Sequence Models</i>	25
2.2.8	<i>Detection and Recognition of Hindi Text From Natural Scenes and Its Transliteration to English</i>	26
2.3	Kerangka Pemikiran	30
BAB III METODOLOGI PENELITIAN		31
3.1	Alat dan Bahan Penelitian	31
3.1.1	Alat.....	31
3.1.2	Bahan Penelitian.....	32
3.2	Alur Penelitian.....	32
3.2.1	Identifikasi Masalah.....	32
3.2.2	Studi Literatur	32
3.2.3	Rancangan Penelitian.....	33

3.2.4	Pengumpulan Data	34
3.2.5	Implementasi Penelitian	36
3.2.6	Evaluasi dan Perbaikan	37
BAB IV HASIL DAN PEMBAHASAN		38
4.1	Hasil Pengumpulan Data	38
4.1.1	<i>Dataset</i>	38
4.2	<i>Preprocessing Data</i>	39
4.2.1	<i>Case Folding</i>	39
4.2.2	<i>Tokenization</i>	40
4.3	Pemodelan dan Pelatihan.....	42
4.3.1	Pembuatan Model.....	42
4.3.2	Optimasi	43
4.4	Evaluasi	44
4.5	Implementasi Sistem	46
BAB V KESIMPULAN DAN SARAN		51
5.1	Kesimpulan.....	51
5.2	Saran.....	51
DAFTAR PUSTAKA.....		52
LAMPIRAN.....		55

DAFTAR TABEL

Tabel 2. 1 Contoh Perbedaan Translasi dan Transliterasi.....	8
Tabel 2. 2 Transliterasi Konsonan	9
Tabel 2. 3 Transliterasi Vokal Tunggal	10
Tabel 2. 4 Transliterasi Vokal Rangkap.....	11
Tabel 2. 5 Transliterasi Maddah	11
Tabel 2. 6 Contoh Perbandingan Hasil dan Referensi Transliterasi	20
Tabel 2. 7 Ringkasan Penelitian Terkait	27
Tabel 3. 1 Alat Penelitian.....	31
Tabel 3. 2 Sampel Dataset Transliterasi dari teks Al-Qur'an	35
Tabel 3. 3 Interpretasi Skor BLEU [27].....	37
Tabel 4. 1 Dataset Teks Arab - Teks Transliterasi Latin	38
Tabel 4. 2 Token Karakter Data	40
Tabel 4. 3 Hasil Perbandingan Evaluasi BLEU Scores	45
Tabel 4. 5 Hasil Transliterasi	49

DAFTAR GAMBAR

Gambar 2. 1 Tantangan Bahasa Arab dalam Computer Science [9].....	15
Gambar 2. 2 Tahapan dalam Optical Character Recognition [21].....	16
Gambar 2. 3 Proses Transliterasi [12].....	17
Gambar 2. 4 Struktur Seq2seq Encoder dan Decoder [12].....	18
Gambar 2. 5 Kerangka Pemikiran.....	30
Gambar 3. 1 Alur Penelitian.....	32
Gambar 3. 2 Rancangan Alur Penelitian.....	33
Gambar 3. 3 Gambar berisi teks Arab yang diambil dari Internet.....	35
Gambar 3. 4 Tahapan Implementasi.....	36
Gambar 3. 5 Rancangan Antarmuka Prototipe Sistem.....	37
Gambar 4. 1 Arsitektur Model RNN.....	42
Gambar 4. 2 Ringkasan Model.....	43
Gambar 4. 3 User Interface.....	47
Gambar 4. 4 Tampilan Aplikasi Input Gambar.....	47
Gambar 4. 5 Tampilan Aplikasi Hasil Deteksi Teks.....	48
Gambar 4. 6 Tampilan Aplikasi Hasil Transliterasi.....	48

DAFTAR LAMPIRAN

Lampiran 1. Source Code Model Transliterasi	55
Lampiran 2. Source Code Implementasi Sistem	62
Lampiran 3. Lembar Konsultasi.....	65