NAMED ENTITY RECOGNITION BERITA BAHASA INDONESIA DENGAN POLYGLOT

SKRIPSI

disusun oleh

Barep Setiyadi

19.21.1330

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2021

NAMED ENTITY RECOGNITION BERITA BAHASA INDONESIA DENGAN POLYGLOT

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika

disusun oleh

Barep Setiyadi

19.21.1330

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021

PERSETUJUAN

SKRIPSI

***NAMED ENTITY RECOGNITION* BERITA BAHASA INDONESIA DENGAN POLYGLOT**

yang dipersiapkan dan disusun oleh

Barep Setiyadi

19.21.1330

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 5 Agustus 2021

Dosen Pembimbing,

Mardhiya Hayaty, S.T., M.Kom.

NIK. 190302108

PENGESAHAN

SKRIPSI

***NAMED ENTITY RECOGNITION* BERITA BAHASA INDONESIA DENGAN POLYGLOT**

yang dipersiapkan dan disusun oleh

Barep Setiyadi

19.21.1330

telah dipertahankan di depan Dewan Penguji

pada tanggal 19 Juli 2021

Susunan Dewan Penguji

Nama Penguji	**Tanda Tangan**
Anggit Dwi Hartanto, M.Kom. **NIK. 190302163**	_____
Bayu Setiaji, M.Kom. **NIK. 190302216**	_____
Mardhiya Hayaty, S.T., M.Kom. **NIK. 190302108**	_____

Skripsi ini telah diterima sebagai salah satu persyaratan

untuk memperoleh gelar Sarjana Komputer

Tanggal 19 Juli 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom.

NIK. 190302096

PERNYATAAN

PERNYATAAN

Saya yang bertanda tangan di bawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan orang lain untuk memperoleh gelar akademis di suatu Institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Tegal, 20 Agustus 2021

Barep Setiyadi

NIM. 19.21.1330

MOTTO

”Melakukan sesuatu selalu dengan alasan”

PERSEMBAHAN

Segala puji bagi Allah SWT yang telah mencurahkan rahmat dan karuniaNya kepada makhluk-makhluknya. Sholawat serta salam tidak lupa kita curahkan kepada junjungan nabi besar kita Nabi Muhammad SAW yang kita nantikan syafaatnya di hari kiamat kelak.

Alhamdulillah, penulis ucapkan syukur kehadirat Allah SWT karena atas kehendakNya-lah penulis dapat menyelesaikan laporan skripsi yang berjudul "**NAMED ENTITY RECOGNITION BERITA BAHASA INDONESIA DENGAN POLYGLOT**". Tidak lupa penulis persembahkan karya tulis ini untuk:

1. Kedua orang tua tercinta, yang senantiasa memberikan kasih sayang dan juga doa yang tak ada batasnya dan yang selalu mendidik tanpa bosannya, semoga selalu dalam keadaan sehat dan selalu berada dalam lindungan-Nya.
2. Ibu Mardhiya Hayati. yang telah membimbing hingga skripsi ini dapat diselesaikan.
3. Tuan R, Tuan E, Tuan F, Tuan D yang telah membimbing selama ini.
4. Rekan-rekan yang telah mendukung.

KATA PENGANTAR

Puji syukur kepada Allah SWT atas rahmat dan hidayah-Nya, penulis dapat menyelesaikan skripsi yang berjudul "**NAMED ENTITY RECOGNITION BERITA BAHASA INDONESIA DENGAN POLYGLOT**" dengan lancar.

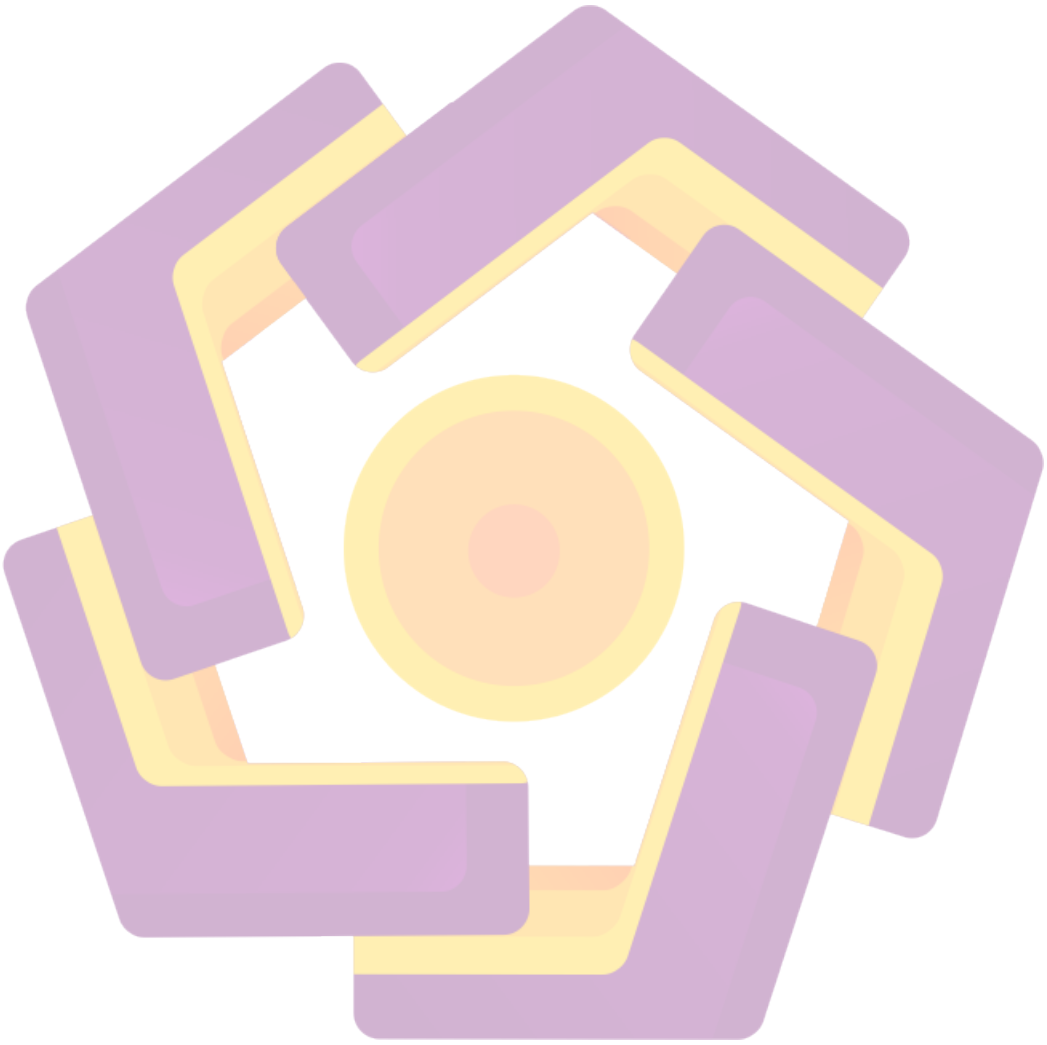
Laporan ini disusun sebagai salah satu syarat kelulusan program S1 Informatika Universitas Amikom Yogyakarta. Dalam penyusunan laporan ini penulis mendapat bantuan dari berbagai pihak. Penulis ingin mengucapkan terima kasih kepada para pihak yang telah membantuk dalam penulisan laporan skripsi ini. Maka dari itu penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Mardhiya Hayaty, S.T., M.Kom. selaku dosen pembimbing yang telah memberikan bimbingan dan arahan sehingga skripsi ini selesai.
4. Dewan Penguji dan segenap Dosen Universitas Amikom Yogyakarta yang telah berbagi ilmu dan pengalamannya.
5. Kedua orang tua yang selalu mendoakan, memberikan semangat dan dukungan moril.
6. Penulis sumber bacaan, jurnal dan makalah yang penulis jadikan referensi dalam penulisan laporan skripsi ini.

Penulis menyadari bahwa masih ada banyak kekurangan di dalam laporan ini. Namun penulis berharap laporan skripsi ini dapat memberikan manfaat pada para pembaca sekalian.

Tegal, 20 Agustus 2021

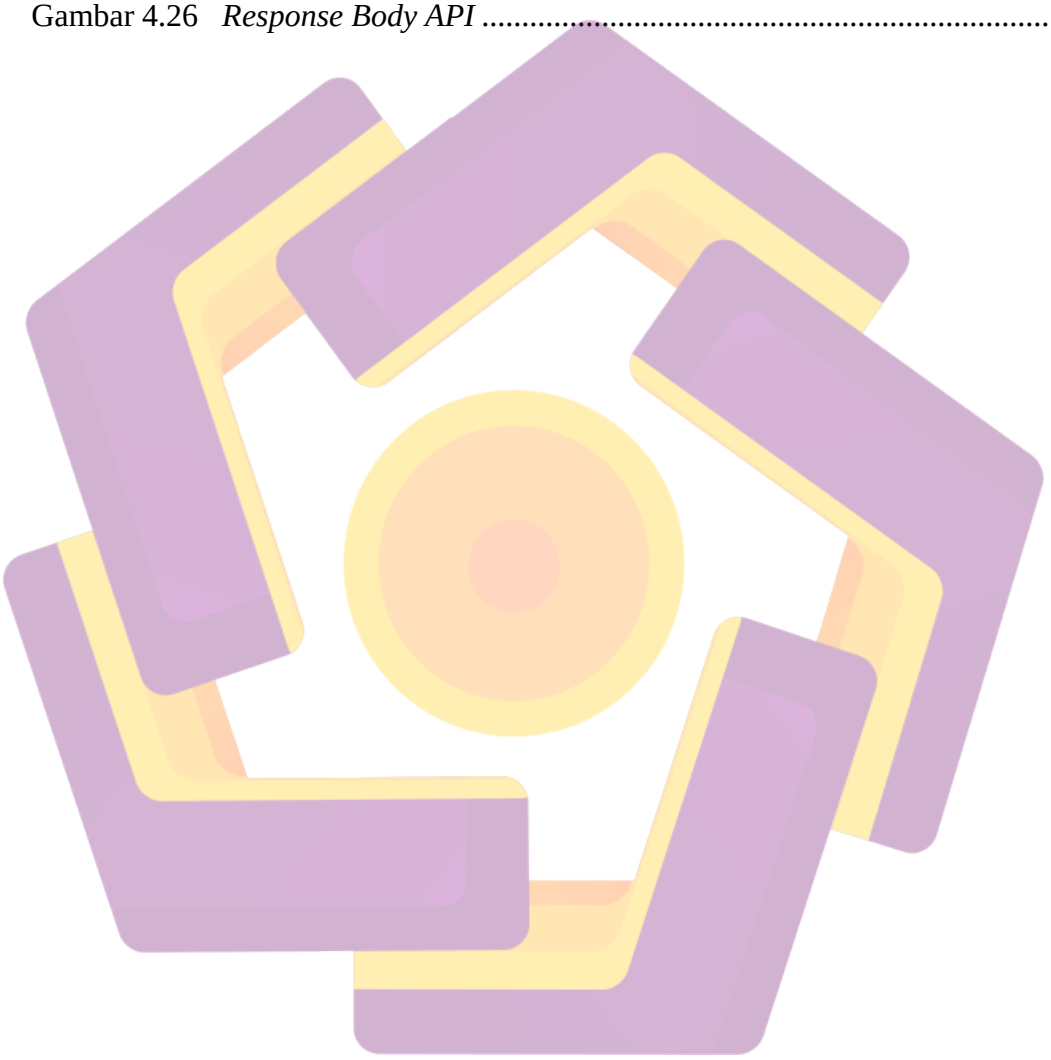
Barep Setiyadi
19.21.1330



DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR



INTISARI

Dalam *Natural Language Processing(NLP*), *Named Entity Recognition* merupakan sub-bahasan yang cukup banyak digunakan untuk penelitian. Tugas utama dari *Named Entity Recognition* yaitu membantu mengidentifikasi dan mendeteksi nama entitas dari suatu kata yang terdapat dalam kalimat.

Sumber data yang digunakan yaitu berita Bahasa Indonesia yang berasal dari 3 media *online* faktual, yaitu detik.com, kompas.com, dan tribunews.com. Kata yang terdapat pada berita Bahasa Indonesia dapat merujuk nama entitas orang, lokasi, atau organisasi. Dengan mengintegrasikan antara *crawler* berita, *Elasticsearch* sebagai alat untuk menyimpan data berita yang sudah diambil dan *Polyglot-NER* menggunakan *REST API*, maka dimungkinkan untuk membuka ruang kolaborasi antara pengembang sistem analisa media lainya.

Saran yang disampaikan penulis adalah menerapkan *REST API* sebagai jembatan komunikasi antar sistem, diharapkan mampu mengintegrasikan antar sistem crawler berita dengan sistem *NER* Bahasa Indonesia dengan tingkat akurasi diatas 80%.

Kata Kunci: *Natural Language Processing, Named Entity Recognition, REST-API, Polyglot, Elasticsearch.*

ABSTRAK

In Natural Language Processing (NLP), Named Entity Recognition is a sub-discussion that is quite widely used for research. The main task of Named Entity Recognition is to help identify and detect the name of the entity from a word contained in a sentence.

The data sources used are Indonesian news originating from 3 factual online media, namely detik.com, kompas.com, and tribunews.com. Words in Indonesian news can refer to the name of the entity, person, location, or organization. By integrating news crawlers, Elasticsearch as a tool to store retrieved news data and Polyglot-NER using the REST API, it is possible to open up space for collaboration between developers of other media analysis systems.

The suggestion submitted by the author is to apply the REST API as a communication bridge between systems, which is expected to be able to integrate news crawler systems with the Indonesian NER system with an accuracy rate above 80%.

Keywords**:** *Natural Language Processing, Named Entity Recognition, REST-API, Polyglot, Elasticsearch.*