MESIN PENERJEMAH BAHASA INGGRIS – INDONESIA BERBASIS JARINGAN SARAF TIRUAN DENGAN MEKANISME ATTENTION MENGGUNAKAN ARSITEKTUR TRANSFORMER

SKRIPSI

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer
pada Departemen Ilmu Komputer/ Informatika

Disusun Oleh :
PRASASTO ADI WISMOYO
24010311130026

DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
2018

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini:

Nama : Prasasto Adi Wismoyo
NIM : 24010311130026
Judul : Mesin Penerjemah Bahasa Inggris – Indonesia Berbasis Jaringan Saraf Tiruan dengan Mekanisme Attention Menggunakan Arsitektur Transformer

Dengan ini saya menyatakan bahwa dalam tugas akhir/ skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 8 Juni 2018

Prasasto Adi Wismoyo
24010311130026

HALAMAN PENGESAHAN

Judul : Mesin Penerjemah Bahasa Inggris – Indonesia Berbasis Jaringan Saraf Tiruan dengan Mekanisme Attention Menggunakan Arsitektur Transformer

Nama : Prasasto Adi Wismoyo

NIM : 24010311130026

Telah diujikan pada sidang tugas akhir pada tanggal 8 Juni 2018 dan dinyatakan lulus pada tanggal **8 Juni 2018**.

Semarang, 5 Juli 2018

Mengetahui,

Ketua Departemen Ilmu Komputer/Informatika
FSM UNDIP

Dr. Retno Kusumaningrum, S.Si., M.Kom.
NIP. 198104202005012001

Panitia Penguji Tugas Akhir
Ketua,

Indra Waspada, S.T., M.T.I.
NIP. 197902122008121002

HALAMAN PENGESAHAN

Judul : Mesin Penerjemah Bahasa Inggris – Indonesia Berbasis Jaringan Saraf Tiruan dengan Mekanisme Attention Menggunakan Arsitektur Transformer

Nama : Prasasto Adi Wismoyo

NIM : 24010311130026

Telah diujikan pada sidang tugas akhir pada tanggal 8 Juni 2018.

Semarang, 5 Juli 2018

Pembimbing

Dr. Retno Kusumaningrum, S.Si., M.Kom.

NIP. 198104202005012001

ABSTRAK

Bahasa Inggris sebagai bahasa pengantar internasional dan bahasa Indonesia sebagai bahasa pengantar nasional memiliki peran yang sangat penting bagi bangsa Indonesia. Sehingga mesin penerjemah (*machine translation*) dirasa sangat membantu dalam proses alih bahasa antara keduanya. Beberapa tahun terakhir mesin penerjemah mengalami perkembangan yang sangat pesat melalui pendekatan mesin penerjemah berbasis jaringan saraf tiruan (*neural machine translation*). Transformer merupakan salah satu arsitektur mesin penerjemah berbasis jaringan saraf tiruan yang mendapatkan hasil *state-of-the-art* dalam beberapa penelitian. Arsitektur ini menggunakan mekanisme *attention* tanpa adanya lapisan *recurrent* maupun *convolution*. Penulis melakukan penelitian untuk mengetahui kualitas terjemahan bahasa Inggris – Indonesia yang dihasilkan oleh Transformer berdasarkan metrik penilaian BLEU. Hasil pengujian menggunakan korpus uji *dev2010, tst2010, tst2017plus,* dan *test.go.id* menunjukkan bahwa model mesin penerjemah mendapatkan skor BLEU 24,83% (*dev2010*) 24,14% (*tst2010*), 27,31% (*tst2017plus*), dan 34,18% (*test.go.id*) untuk arah terjemahan Inggris – Indonesia. Adapun untuk arah terjemahan Indonesia – Inggris, model mendapatkan skor BLEU 90,13% (*dev2010*), 88,68% (*tst2010*), 90,68% (*tst2010*), dan 69,14% (*test.go.id*).

Kata kunci: mesin penerjemah, jaringan saraf tiruan

ABSTRACT

English as an international lingua franca and Bahasa Indonesia as the national lingua franca have significant role for Indonesian people. Therefore machine translation is considered to be very helpful in translating the two languages. In recent years, machine translation is advancing rapidly through neural machine translation approach. Transformer is one of neural machine translation architectures that achieved state-of-the-art results in several researches. That architecture uses attention mechanism without recurrent nor convolution layer. This research was conducted to discover the quality of English-Indonesian translation generated by Transformer based on BLEU metrics. The test results using test corpus *dev2010*, *tst2010*, *tst2017plus*, and *test.go.id* show that the machine translation model achieved BLEU scores 24.83% (*dev2010*) 24.14% (*tst2010*), 27.31% (*tst2017plus*), and 34 , 18% (*test.go.id*) for the English-Indonesian translation direction. As for the Indonesian-English translation direction, the model achieved BLEU scores 90.13% (*dev2010*), 88.68% (*tst2010*), 90.68% (*tst2010*), and 69.14% (*test.go.id*).

Keywords: machine translation, artificial neural network

KATA PENGANTAR

Segala puji syukur bagi Tuhan Yang Maha Esa atas karunia-Nya yang diberikan kepada penulis sehingga dapat menyelesaikan penulisan laporan tugas akhir yang berjudul "Mesin Penerjemah Bahasa Inggris – Indonesia Berbasis Jaringan Saraf Tiruan dengan Mekanisme Attention Menggunakan Arsitektur Transformer". Laporan tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Jurusan Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

Dalam penyusunan laporan ini tentulah banyak mendapat bimbingan dan bantuan dari berbagai pihak. Untuk itu, pada kesempatan ini penulis mengucapkan rasa hormat dan terima kasih kepada:

1. Ibu Dr. Retno Kusumaningrum, S.Si., M.Kom., selaku Ketua Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang, sekaligus menjadi Dosen Pembimbing yang telah membantu dalam proses bimbingan hingga selesainya laporan tugas akhir ini.
2. Bapak Helmie Arif Wibawa, S.Si., M.Cs., selaku Koordinator Tugas Akhir sekaligus menjadi Penguji Tugas Akhir.
3. Bapak Indra Waspada, S.T., M.T.I., selaku Ketua Penguji Tugas Akhir.
4. Bapak Sudarwanto, S.H. dan Ibu Adi Sulistyomurti, selaku orang tua yang terus memberikan dukungan dan do'a dalam menyelesaikan tugas akhir ini.
5. Semua pihak yang telah membantu kelancaran dalam pelaksanaan tugas akhir yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam laporan ini masih banyak kekurangan baik dari segi materi ataupun dalam penyajiannya karena keterbatasan kemampuan dan pengetahuan penulis. Oleh karena itu, kritik dan saran sangat penulis harapkan. Semoga laporan ini dapat bermanfaat bagi pembaca pada umumnya dan penulis pada khususnya.

Semarang, 8 Juni 2018

Penulis,

Prasasto Adi Wismoyo

24010311130026

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

1.1. Latar Belakang

Keberagaman suku dan budaya di Indonesia memberikan pengaruh terhadap bahasa yang digunakan sebagai sarana komunikasi dalam kehidupan bermasyarakat. Untuk menjembatani perbedaan bahasa antarsuku, antardaerah, dan antarbudaya, ditetapkanlah bahasa Indonesia sebagai bahasa pengantar (*lingua franca*) demi kelancaran interaksi sosial dan integrasi bangsa. Asal mula bahasa Indonesia berpangkal dari bahasa Melayu, bahasa yang ketika itu menjadi bahasa perdagangan bagi para saudagar di wilayah Nusantara (Mohamed and Harahap 2013).

Selain bahasa Indonesia dan bahasa daerah, bahasa asing juga digunakan dan diajarkan di Indonesia untuk mendukung komunikasi antarbangsa dalam ruang lingkup internasional. Bahasa asing yang banyak digunakan dan diajarkan di Indonesia adalah bahasa Inggris dan bahasa Arab. Bahasa Inggris merupakan bahasa pengantar dalam komunikasi beragam tema pembicaraan di kancah internasional. Munculnya bahasa Inggris sebagai bahasa pengantar internasional tak lepas dari peran kekuasaan politik dan perdagangan Kerajaan Inggris pada abad 19 dan 20 (Kloss 1967). Sementara itu, Bahasa Arab umumnya digunakan sebagai bahasa pengantar dalam komunikasi antar umat Islam, khususnya komunikasi yang berkaitan dengan tema keagamaan (Arifin and Yundiafi 2014; Kloss 1967).

Merupakan suatu kelaziman bahwa bahasa pengantar digunakan dalam pertukaran informasi di segala bidang seperti perdagangan, pendidikan, teknologi, diplomatik, dan lain sebagainya. Sehingga penggunaan bahasa Indonesia sebagai bahasa pengantar nasional serta bahasa Inggris sebagai bahasa pengantar internasional berimplikasi terhadap tingginya kebutuhan alih bahasa di antara keduanya. Kamus Inggris – Indonesia menjadi acuan primer untuk menemukan kata atau terminologi yang belum dimengerti ketika memahami atau menerjemahkan teks informasi. Penggunaan kamus cetak pada era digital sudah semakin tergantikan oleh kamus digital yang memiliki beberapa keunggulan seperti portabilitas tinggi, kemudahan pencarian lema dan keterbaruan entri pada basis data. Alat bantu penerjemahan selain kamus dwibahasa dapat berupa thesaurus, memori penerjemahan (*translation memory*), pemeriksa ejaan (*spell checker*), pemeriksa tata bahasa (*grammar checker*), maupun mesin penerjemah (*machine translation*).

Mesin penerjemah merupakan perangkat lunak penerjemah bahasa yang proses penerjemahannya dilakukan oleh mesin (komputer) secara otomatis dan hasilnya dapat langsung disajikan kepada pengguna. Opsi penggunaan mesin penerjemah lebih umum dipilih karena cepat dan praktis dalam melakukan penerjemahan teks, berbanding terbalik dengan penerjemahan oleh manusia yang memerlukan waktu lama. Meskipun penerjemah manusia yang berkompeten akan menghasilkan penerjemahan yang lebih baik dan dalam beberapa keadaan tak tergantikan oleh penerjemah otomatis, namun ada banyak kasus di mana cukup untuk diterjemahkan menggunakan mesin penerjemah. Selain itu, hasil penerjemahan oleh mesin pun tidak serta merta digunakan begitu saja, namun disunting kembali oleh manusia. Sehingga penerjemah manusia berkualifikasi tinggi pun juga menggunakan mesin penerjemah untuk mempercepat proses penerjemahan.

Penelitian terkait mesin penerjemah telah banyak dilakukan dan termasuk bidang penelitian yang sangat aktif. Namun penelitian mesin penerjemah bahasa Inggris – Indonesia tergolong masih sedikit ditemui terutama metode-metode berbasis jaringan saraf tiruan yang merupakan *state-of-the-art* untuk mesin penerjemah. *Neural machine translation* merupakan istilah yang digunakan untuk menyebut metode mesin penerjemah dengan arsitektur berbasis jaringan saraf tiruan. Metode-metode yang paling umum digunakan adalah RNN (*Recurrent Neural* Network) , CNN (*Convolutional Neural Network*), dan *attention*. Pada tugas akhir ini, *attention* dipilih karena model dasarnya sudah cukup untuk menghasilkan terjemahan yang baik serta kompleksitas waktu komputasi dan ruang memori yang relatif lebih rendah dibandingkan metode lainnya (Chen et al. 2018; Vaswani et al. 2017). Arsitektur *Transformer* adalah arsitektur mesin penerjemah yang menerapkan mekanisme *attention* tanpa adanya tambahan RNN ataupun CNN (Vaswani et al. 2017).

1.2. Rumusan Masalah

Berdasarkan penjelasan latar belakang yang telah disampaikan, dapat disusun rumusan masalah yaitu bagaimana menerapkan metode jaringan saraf tiruan dengan mekanisme attention menggunakan arsitektur Transformer pada mesin penerjemah Bahasa Inggris – Indonesia.

1.3. Tujuan dan Manfaat

Tujuan yang ingin dicapai dari penelitian tugas akhir ini adalah mengetahui kualitas model penerjemahan Bahasa Inggris – Indonesia berbasis jaringan saraf tiruan dengan mekanisme *attention* menggunakan arsitektur *Transformer* berdasarkan matrik skor BLEU.

Adapun manfaat yang diharapkan dari penelitian tugas akhir ini adalah dapat meningkatkan peran mesin penerjemah dalam kegiatan masyarakat, penelitian, maupun industri.

1.4. Ruang Lingkup

Ruang lingkup penelitian tugas akhir ini adalah sebagai berikut :

1. Korpus latih berupa korpus paralel bahasa Inggris – Indonesia dari PANL-BPPT, IWSLT17, dan OpenSubtitles2018.
2. Korpus uji berupa korpus paralel dari IWSLT17 dan korpus uji yang disusun oleh penulis.
3. Metode penelitian berbasis jaringan saraf tiruan dengan mekanisme *attention* menggunakan arsitektur Transformer.
4. Evaluasi kualitas model penerjemahan menggunakan matrik skor BLEU.
5. Output hasil terjemahan berupa *plain text* dengan pemisah kalimat hasil terjemahan berupa baris baru.

1.5. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam laporan tugas akhir ini terbagi menjadi beberapa pokok bahasan, yaitu:

BAB I PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan laporan.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan mengenai teori-teori serta istilah-istilah yang digunakan dalam penelitian tugas akhir meliputi perkembangan penelitian mesin penerjemah bahasa Inggris - Indonesia ,jaringan saraf tiruan, *machine translation*, *neural machine translation*, korpus, tokenisasi,

subword, *embedding*, *positional encoding*, mekanisme *attention*, dan BLEU.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan yang dilakukan dalam penelitian tugas akhir. Tahapan tersebut meliputi pengumpulan data, prapengolahan, arsitektur, pelatihan, dan evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil eksperimen yang telah dilakukan berupa implementasi model, pengujian *hyperparameter*, serta pengujian model akhir.

BAB V PENUTUP

Bab ini merupakan kesimpulan dari bab-bab yang dibahas sebelumnya dan saran bagi penelitian selanjutnya.