

BAB I

PENDAHULUAN

1.1 Latar belakang

1.1.1 Latar belakang umum

Bahasa merupakan alat komunikasi yang penting bagi manusia, karena dengan menggunakan bahasa kita dapat mengetahui dan mengetahui informasi yang dibutuhkan. Dengan bahasa juga manusia dapat mengekspresikan diri, menyampaikan kritik dan pendapat, pikiran serta keinginannya. Bahasa juga merupakan alat yang digunakan untuk sosialisasi dan adaptasi sosial antara manusia satu dengan manusia lainnya agar kehidupannya semakin berkembang.

Dari segi bahasa, Indonesia adalah salah satu negara terkaya diseluruh dunia. Dari jumlah total lima sampai enam ribu bahasa di dunia di Nusantara ini terdapat enam ratus bahasa, dan bahkan lebih, yang digunakan oleh sekitar 215 juta penutur. Bahasa merupakan alat utama, modal utama, dan satu satunya cara untuk berkomunikasi.

Indonesia terletak di persilangan Asia Selatan dan Asia Timur, yang dari dahulu hingga sekarang dikunjungi oleh para pedagang-pedagang, sebagai akibat migrasi tersebut, pergeseran bahasa tidak dapat dihindarkan. Pergeseran bahasa adalah perubahan secara tetap dalam pilihan bahasa seseorang untuk keperluan sehari hari. Dengan seringnya pergeseran bahasa akan menimbulkan kepunahan bahasa yaitu hilangnya bahasa persatuan atau daerah akibat dari penuturnya menggunakan bahasa lain.

Bahasa Jawa adalah salah satu bahasa komunikasi yang digunakan secara khusus di lingkungan etnis Jawa. Bahasa ini merupakan bahasa pergaulan, yang digunakan untuk berinteraksi antarindividu dan memungkinkan terjadinya

komunikasi dan perpindahan informasi. Bahasa Jawa yang dulu merupakan bahasa yang besar, dengan bertambahnya waktu, penggunaannya semakin berkurang. Saat ini para kaum muda di Pulau Jawa, khususnya yang masih di usia sekolah, sebagian besar tidak menguasai bahasa Jawa. Hal ini disebabkan oleh gencarnya serbuan beragam budaya asing dan arus informasi yang masuk melalui bermacam sarana seperti televisi dan lain-lain.

Untuk mencegah terjadinya kepunahan bahasa atau biasa disebut dengan istilah kematian bahasa, maka sudah seharusnya masyarakat, khususnya etnis Jawa menyadari pentingnya pemeliharaan bahasa daerah.

Pemeliharaan bahasa daerah berarti melindungi bahasa daerah agar tetap memainkan peran dalam kehidupan masyarakat, pada tatanan kehidupan masa kini dan masa mendatang. Pemeliharaan ini berupa pengembangan bahasa daerah agar mampu memenuhi tuntutan masyarakat pendukungnya, di antaranya meliputi pemekaran kosakata dan kodifikasi berupa penyusunan pedoman ejaan, kamus, dan tata bahasa. [2]

Dari beberapa usaha pemeliharaan bahasa, kamus merupakan salah satu sarana yang dapat membantu siapa saja yang ingin mempelajari suatu bahasa. Di dalam sebuah kamus seseorang dapat mengakses bahasa yang diinginkan dengan mudah. Didukung dengan kemajuan teknologi, kamus dapat diciptakan dengan memanfaatkan perkembangan teknologi, berupa mesin penerjemah untuk menyesuaikan dengan kebutuhan dalam era ini.

Mesin penerjemah merupakan alat yang dapat digunakan dalam pemeliharaan bahasa sebagai wujud kesadaran akan terjadinya kepunahan bahasa daerah. Dalam menerjemahkan kalimat dari satu bahasa ke bahasa yang lain, mesin penerjemah memiliki beberapa metode diantaranya adalah *example-based*, *case-based*, *rule-based*, dan *statistic-based*. Metode *example-based* yaitu penerjemah yang menghasilkan terjemahan terbaru mengacu kepada hasil terjemahan sebelumnya yang tersimpan di dalam basis data sistem. Metode kedua adalah *case-based*, konsep dari sistem ini adalah menerjemahkan kalimat dengan memakai pola/*case* yang serupa dengan pola yang disimpan dalam *set of case*. Metode *rule-based*, aturan yang ada di dalam sistem dibuat oleh pembuat

sistem berdasarkan pengetahuan linguistik. Sedangkan metode *statistic-based* merupakan metode sistem translasi berdasarkan ilmu statistika dengan menggunakan model pembelajaran dari sekumpulan kalimat (korpus) sebagai data latihnya.

1.1.2 Latar belakang khusus

Dalam pembangunan mesin penerjemah ini akan digunakan metode *statistic-based* yang merupakan suatu paradigma dari mesin penerjemah dimana penerjemahan dilakukan berbasis model statistik dengan parameter-parameter yang diturunkan dari analisis paralel korpus. Mesin penerjemah statistik ini didasarkan pada pandangan bahwa setiap kalimat dalam bahasa memiliki kemungkinan terjemahan dalam bahasa lain. Sebuah kalimat dapat diterjemahkan dari satu bahasa ke dalam bahasa yang lain dalam banyak kemungkinan cara. Metode terjemahan statistik mengambil pandangan bahwa setiap kalimat dalam bahasa target adalah terjemahan yang dimungkinkan dari kalimat masukan.

Kelebihan utama dari mesin penerjemah berbasis statistik adalah bersifat lebih umum sehingga untuk mengembangkannya tidak diperlukan pengetahuan bahasa, cukup dengan bermodalkan korpus paralel antara dua bahasa yang akan diterjemahkan. Korpus paralel menjadi penting karena semakin besar jumlah kalimat yang tersimpan dalam korpus paralel maka semakin baik pula kualitas terjemahan yang dihasilkan.

Dalam pembangunannya mesin penerjemahan statistik ini disusun oleh beberapa komponen utama yaitu *Language Model*, *Translation Model* dan decoder. Telah disediakan banyak toolkit untuk mendukung seluruh proses tersebut. Namun demikian, di Indonesia belum banyak penelitian mengenai pengembangan mesin penerjemah statistik berbasis frasa ini, untuk itu penulis tertarik mengkaji lebih lanjut dalam sebuah penelitian berjudul " Mesin Penerjemah Bahasa Indonesia-Jawa Menggunakan Metode Statistik Berbasis Frasa ".

1.2 Rumusan masalah

Sehubungan dengan latar belakang seperti yang diuraikan sebelumnya, maka penulis merumuskan permasalahan dalam penelitian yaitu :

1. Indonesia memiliki keberagaman bahasa yang tinggi, namun karena letaknya di persimpangan benua mengakibatkan mudahnya bahasa asing masuk dan menggeser bahasa daerah di Indonesia.
2. Pergeseran bahasa dan tingginya tingkat migrasi menjadi penyebab utama punahnya suatu bahasa daerah karena ditinggalkan penuturnya.
3. Gencarnya serbuan beragam budaya asing dan arus informasi yang masuk terhadap anak usia sekolah melalui bermacam sarana seperti televisi dan lain-lain menjadi penyebab berkurangnya penggunaan bahasa Jawa.

1.3 Batasan masalah

Agar pembahasan penelitian ini tidak menyimpang dari apa yang telah dirumuskan, maka diperlukan batasan-batasan. Batasan-batasan dalam penelitian ini adalah:

1. Mesin penerjemah statistik yang akan dibangun hanya menerjemahkan bahasa Indonesia dan bahasa Jawa secara satu arah.
2. Jumlah kata atau kalimat yang tersedia dibatasi hanya beberapa kata atau kumpulan kata.
3. Tingkat ketepatan dalam menerjemahkan data uji dievaluasi oleh software BLEU.
4. Proses Pengujian dilakukan menggunakan corpus yang telah disusun, dengan Alkitab terjemahan Jawa dan Indonesia yang digunakan merupakan terjemahan yang valid.

1.4 Tujuan penelitian

Hasil dari penelitian ini diharapkan dapat :

1. Menghindarkan bahasa Jawa dari fenomena pergeseran bahasa yang memicu kepunahan bahasa.

2. Memberikan sarana bagi penutur yang ingin menuturkan bahasa Jawa.
3. Membantu anak usia sekolah untuk memahami dan menggunakan bahasa Jawa.

1.5 Manfaat penelitian

Dari hasil penelitian ini, penulis menguraikan beberapa manfaat yang dapat digunakan antara lain :

1.5.1 Bagi pengguna

Hasil dari penelitian ini dapat digunakan sebagai alat bantu untuk mempelajari bahasa Jawa, membantu masyarakat dalam proses interaksi lintas suku, dan sebagai bentuk kesadaran untuk menghindarkan bahasa Jawa dari pergeseran bahasa.

1.5.2 Bagi akademik

Dari hasil penelitian ini dapat digunakan sebagai salah satu tolak ukur keberhasilan bagi akademik maupun mahasiswa dalam pengembangan proses belajar mengajar.

1.5.3 Bagi penulis

1. Dapat menerapkan mata kuliah kecerdasan buatan yang selama ini di pelajari dalam perkuliahan.
2. Dapat menerapkan pengembangan perangkat lunak yang selama ini dipelajari selama perkuliahan.
3. Dapat memperluas wawasan penulis dalam merancang suatu sistem guna memenuhi syarat meraih gelar dalam ilmu komputer pada jurusan Teknik Informatika di Universitas Dian Nuswantoro Semarang.

