

BAB 1

PENDAHULUAN

Pada bab ini akan dibahas deskripsi umum penelitian, yang terdiri dari lima bagian, yakni latar belakang permasalahan, tujuan penelitian, ruang lingkup penelitian, metodologi penelitian, dan sistematika penulisan laporan penelitian.

1.1 Latar Belakang

Di era teknologi seperti saat ini, internet berkembang pesat. Data yang tersimpan di internet pun sangat besar, hingga mencapai satuan *terabyte*. Data ini terdiri atas berbagai bentuk, yakni teks, gambar, dan suara. Di samping itu, perkembangan media televisi dan radio yang pesat pun turut menambah ketersediaan data, umumnya berupa gambar dan suara. Setiap data mengandung informasi penting yang perlu diketahui, tidak terkecuali data suara, misalnya berita yang sering disiarkan stasiun televisi maupun radio. Masalahnya adalah bagaimana cara mengolah data suara yang ada, sehingga diperoleh informasi yang diinginkan secara tepat dan cepat. Hal inilah yang menjadi alasan pentingnya sistem pengenalan suara otomatis.

Sama seperti data teks, data suara pun mengandung informasi tertentu yang perlu diketahui. Secara alami, seseorang dapat mengetahui informasi dari suatu data suara dengan cara mendengarnya terlebih dahulu. Mendengarkan satu data suara mungkin bukan pekerjaan yang sulit, namun jika data suara yang tersedia berjumlah sangat besar, ratusan atau bahkan ribuan, maka mendengarkannya satu per satu membutuhkan banyak tenaga dan waktu yang lama.

Selain itu, pengolahan data suara melibatkan sinyal-sinyal bunyi atau akustik yang tidak mudah untuk diproses. Hal tersebut berbeda dari data teks yang memiliki representasi kode ASCII untuk setiap simbolnya, sehingga pengolahannya jauh lebih mudah, cukup dengan memanfaatkan aplikasi *word processing*.

Dari segi ukuran, data suara umumnya berukuran lebih besar daripada teks untuk kandungan informasi yang sama. Hal ini mengakibatkan pemakaian memori yang besar pula untuk menyimpan data suara dalam jumlah besar. Dengan demikian, tingkat portabilitas data menjadi lebih rendah. Bandingkan dengan dokumen teks yang jauh lebih hemat pemakaian memorinya dan lebih fleksibel dalam hal portabilitas.

Menghadapi berbagai masalah terkait pengolahan data suara, diperlukan suatu cara pengorganisasian data suara untuk mengurangi masalah-masalah tersebut. Salah satu cara adalah mengubah data suara menjadi teks. Selanjutnya, pengolahan cukup dilakukan terhadap teks. Sebagai contoh, jika seseorang ingin mencari rekaman berita mengenai kerusuhan Poso, maka ia cukup mencarinya ke dalam teks dengan memanfaatkan program sederhana, tanpa perlu mencari langsung ke rekaman-rekaman berita yang tersedia. Dengan menggunakan teks, pengorganisasian data suara menjadi lebih mudah dan cepat.

Konversi suara menjadi teks juga dibutuhkan pada aplikasi-aplikasi yang diperuntukkan bagi penderita cacat fisik, misalnya lumpuh. Penderita cacat fisik merasa sangat terbantu dalam menjalankan suatu aplikasi cukup dengan perintah suara dibandingkan harus mengetikkan perintahnya ke dalam sistem. Sebagai contoh, penderita cacat fisik cukup mengatakan “buka pintu” untuk membuka pintu rumahnya. Pada kasus ini, teks hasil konversi suara diproses lebih lanjut, sehingga sistem mampu bereaksi sesuai perintah yang diucapkan pengguna.

Kebutuhan lain akan konversi suara menjadi teks adalah pada bidang hukum. Para profesional di bidang hukum membutuhkan informasi diproses dalam bentuk digital. Hal ini bertujuan untuk mengurangi waktu perputaran (*turnaround time*) dokumen di antara pihak-pihak yang berkepentingan. Dengan adanya sistem pengenalan suara otomatis, seorang pengacara dapat mendiktekan informasi baru yang diperoleh dari kliennya melalui sistem ini, sehingga proses pencatatan atau dokumentasi kasus menjadi lebih cepat. Pendiktean (*dictation*) pun dapat dilakukan dari luar kantor, sehingga lebih fleksibel dan efisien.

Dokumen yang dihasilkan pun sudah berbentuk digital, sehingga saat diperlukan dapat segera disampaikan ke pihak lain secara elektronik.

Saat ini, penelitian di bidang sistem pengenalan suara (SPS) untuk beberapa bahasa telah berkembang pesat, misalnya Bahasa Inggris, Perancis, Thailand, Jepang, dan lain-lain. Suatu aplikasi SPS Bahasa Inggris, yakni Dragon NaturallySpeaking 9, bahkan menyatakan akurasi mencapai 99%¹. Penelitian SPS untuk Bahasa Perancis oleh J. L. Gauvain dkk [GAUV07] telah mencapai akurasi sebesar 95%. Sinaporn Suebvisai dkk [SUEB05] dari Thailand pun telah mengembangkan SPS untuk Bahasa Thai dengan akurasi sekitar 80%. SPS untuk Bahasa Jepang juga telah dikembangkan oleh Satoru Tsuge dkk [TSUG02] dengan akurasi maksimum yang dapat diperoleh adalah 90,4%. Bahkan Slovenia pun sudah memulai penelitian SPS untuk Bahasa Slovenia oleh France Mihelič dkk [MIHE00] dengan akurasi rata-rata 69,32%. Untuk Bahasa Indonesia, penelitian SPS sudah dimulai, namun masih berupa langkah awal [LEST06] [SAKT08].

Adanya kebutuhan akan sistem pengenalan suara otomatis serta SPS untuk bahasa-bahasa lain yang telah berkembang pesat menjadi faktor-faktor yang memotivasi penulis untuk melakukan penelitian SPS untuk Bahasa Indonesia.

1.2 Tujuan Penelitian

Ada 5 tujuan dari penelitian ini, antara lain:

1. melakukan proses pembelajaran dan eksperimen untuk menentukan himpunan bunyi, atau yang dikenal dengan alofon, dalam Bahasa Indonesia yang paling tepat, dalam arti mengurangi ambiguitas sistem dalam proses pengenalan suara,
2. menyusun kamus pelafalan kata-kata Bahasa Indonesia berdasarkan himpunan alofon yang telah ditentukan,
3. membuat model Bahasa Indonesia untuk mendukung sistem pengenalan suara otomatis,

¹ http://www.nuancestore.com/v2.0-img/operations/scansoft/site/html/dragon/stndrd_9.htm

4. melakukan pengujian sejumlah berkas suara Bahasa Indonesia dengan menggunakan sistem pengenalan suara otomatis, serta
5. mengukur dan mengevaluasi kinerja sistem pengenalan suara otomatis berdasarkan hasil pengujian tersebut, sehingga dapat diperoleh himpunan alofon yang paling tepat untuk Bahasa Indonesia, dalam arti yang memiliki ambiguitas paling rendah dalam proses pengenalan suara.

1.3 Ruang Lingkup Penelitian

Suatu sistem pengenalan suara otomatis memerlukan dua model dalam proses mengenali suara, yakni model akustik dan model bahasa. Fokus penelitian ini adalah pembuatan model akustik yang baik. Kualitas model akustik sangat bergantung pada ketepatan pendefinisian alofon atau bunyi dalam Bahasa Indonesia yang dipahami oleh komputer. Untuk mendapatkan himpunan alofon, atau dikenal dengan kamus fonetik, yang paling tepat untuk Bahasa Indonesia, penulis melakukan berbagai eksperimen, mulai dari pelatihan hingga pengujian, menggunakan beberapa jenis kamus fonetik. Caranya adalah membandingkan akurasi hasil eksperimen dari berbagai kamus fonetik yang ada. Kamus fonetik yang menyebabkan akurasi paling tinggi dari sejumlah eksperimen dianggap sebagai kamus fonetik yang paling sesuai untuk Bahasa Indonesia.

Data yang digunakan dalam penelitian ini terbatas pada data suara rekaman melalui telepon dan beberapa rekaman siaran berita.

1.4 Metodologi Penelitian

Berikut ini adalah metodologi penelitian yang diterapkan selama pelaksanaan tugas akhir:

1. Studi Literatur

Tahap ini dilakukan untuk mengetahui seluk-beluk mengenai topik sistem pengenalan suara, baik alur proses pengenalan suara, model-model yang dibutuhkan, metode atau pendekatan untuk memperoleh model-

model tersebut, maupun perkembangan bidang pengenalan suara untuk berbagai bahasa saat ini, dari berbagai sumber, seperti karya tulis ilmiah, jurnal, buku, dan lain-lain.

2. Pembersihan Data Suara

Pada tahap ini dilakukan pembersihan data suara, baik memperbaiki berkas suara maupun memperbaiki transkripsinya karena penulis menemukan cukup banyak ketidaksesuaian antara berkas suara dengan transkripsinya. Pembersihan hanya dilakukan pada data suara untuk pelatihan.

3. Eksperimen dan Pengukuran Kinerja

Pada tahap ini, penulis melakukan eksperimen pengenalan suara dengan berbagai kamus fonetik untuk Bahasa Indonesia, kemudian mengukur kinerja sistem pengenalan suara tersebut dengan menghitung persentase akurasi.

4. Analisis Hasil Eksperimen

Pada tahap ini, penulis melakukan analisis terhadap sejumlah hasil eksperimen, tentunya dengan kamus fonetik yang berbeda-beda, untuk mengetahui kamus fonetik apa yang paling tepat untuk pendefinisian bunyi dalam Bahasa Indonesia. Secara numerik, hal ini bisa dilihat dari persentase akurasi yang tertinggi di antara persentase akurasi sejumlah eksperimen tersebut.

1.5 Sistematika Penulisan Laporan

Laporan tugas akhir ini disusun dengan sistematika sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini akan membahas latar belakang permasalahan, tujuan penelitian, ruang lingkup penelitian, metodologi penelitian, dan sistematika penulisan laporan.

BAB 2 LANDASAN TEORI

Bab ini akan membahas teori-teori terkait pelaksanaan penelitian ini, yang terdiri dari deskripsi mengenai Sistem Pengenalan Suara (SPS), model akustik, kamus fonetik, model bahasa, SPS untuk Bahasa Indonesia, evaluasi SPS, dan penelitian terkait SPS Bahasa Indonesia.

BAB 3 EKSPERIMEN

Bab ini akan mengulas data-data yang akan digunakan dalam eksperimen, tahap-tahap atau skenario eksperimen, dan parameter keberhasilannya.

BAB 4 HASIL EKSPERIMEN DAN ANALISIS

Bab ini akan menjabarkan hasil eksperimen, analisis terhadap hasil eksperimen tersebut, analisis kesalahan dari hasil eksperimen, dan usaha-usaha perbaikan yang dilakukan untuk meningkatkan kinerja SPS Bahasa Indonesia.

BAB 5 PENUTUP

Bab ini akan menjabarkan kesimpulan dari hasil penelitian dan saran untuk penelitian selanjutnya, sehingga dapat diperoleh hasil yang lebih baik.

