

PREDICTIVE TEXT ENTRY MENGGUNAKAN MODEL BIGRAM

Fransiska Dani Hesti Mayasari¹, Toto Suharto², Andrian Rakhmatsyah³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Banyak layanan yang disediakan dalam telepon genggam, tetapi layanan pengiriman pesan singkat/SMS (Short Message Service) merupakan layanan yang paling banyak digunakan. Dan seiring dengan meningkatnya penggunaan SMS, dibutuhkan juga metode/cara penulisan teks yang lebih efektif.

Selama ini hampir seluruh pengguna telepon genggam menggunakan cara manual (multi-tap) dalam menuliskan teks. Hal ini dikarenakan banyak diantara mereka yang tidak mengetahui fungsi prediksi kata atau Predictive Text Entry. Selain itu fungsi prediksi kata yang terdapat pada telepon genggam masih banyak yang menggunakan bahasa Inggris, karena versi bahasa Indonesia dikeluarkan tahun 2006.

Pada tugas akhir ini penulis mencoba mengimplementasikan model bigram dalam memprediksi kata. Sehingga diharapkan dalam menuliskan pesan, kata yang akan ditulis dapat diprediksi berdasarkan kata sebelumnya. Untuk mendapatkan prediksi kata, dilakukan training dahulu terhadap corpus SMS sehingga tiap kata memiliki probabilitas/kemungkinan yang berbeda-beda. Semakin sering suatu kata digunakan, maka semakin tinggi probabilitas atau kemungkinan kata tersebut muncul sebagai prediksi kata.

Analisa keefektifan metode ini dihitung dari nilai KSPC (Keystroker per Character) yang dihitung pada data uji. Jika metode ini menghasilkan nilai KSPC yang lebih kecil dibanding cara manual (multi-tap), maka dapat disimpulkan bahwa metode ini lebih efektif.

Kata Kunci : SMS, Predictive Text Entry, Bigram, KSPC, Word Prediction.

Abstract

There are many services are available in cell phone, but Short Message Service or usually called SMS is the mostly used by many people. Since these service is moving rapidly, it also need an effective new method of writing text.

Along this time, almost cell phone users do the manual method (multitap) in writing text. This is because many of them do not know about Predictive Text Entry. Beside that, these function still using English version since the Indonesian version was launch in 2006.

In this final task, writer try to implement Biagram Model in words prediction. Hopely, in writing text message, we can predict the words according to the previous words. To get the prediction words, we make training to corpus SMS, so each word has different probability. More often a word used, so the probability of prediction word become higher.

The effectiveness analysis of this method is calculated from KSPC (Keystroker per Character) results which is calculated from testing data. If this method produces KSPC result which is smaller than manual method (multi-tap), so can be concluded that this method more effective.

Keywords : SMS, Predictive Text Entry, Bigram, KSPC, Word Prediction

1. Pendahuluan

1.1 Latar belakang

Bagi banyak orang, telepon genggam telah menjadi kebutuhan pokok atau primer dalam berkomunikasi. Dan seiring dengan perkembangan teknologi, telepon genggam juga menyediakan berbagai fitur tambahan. Beberapa fitur memerlukan input atau masukan dari pengguna melalui tombol-tombol ada. Contoh fitur yang paling sering digunakan adalah SMS (*Short Message Service*), fitur yang mampu mengirim pesan kepada penerima. Contoh lainnya adalah fitur *reminder*, *agenda*, *phone book*, dll.

Ukuran telepon genggam yang relatif kecil tidak memungkinkan penggunaan *full-sized keyboards* dimana tiap satu tombol mewakili satu huruf. Dengan demikian, pada umumnya telepon genggam menggunakan 12 tombol yang sering disebut dengan *12-keys keypad*, yang terdiri dari angka 0-9 dan dua tombol tambahan (* dan #). Karakter a-z tersebar pada tombol 2-9, sehingga dalam satu tombol mewakili 3-4 huruf. Karena itu kadang dibutuhkan penekanan tombol beberapa kali untuk setiap karakter yang akan dituliskan (*multiple-tap*). Misal jika ingin menulis huruf "c" maka harus menekan tombol 2 sebanyak tiga kali. Untuk menulis kata "salah" butuh penekanan tombol 77772555244. Selain itu jika ingin menuliskan beberapa karakter yang terdapat dalam satu tombol maka pengguna harus menunggu waktu *timeout* habis, baru kemudian dapat menuliskan karakter selanjutnya. Atau dapat digantikan dengan tombol *next*. Sebagai contoh jika ingin menuliskan kata "cabut" maka harus menekan tombol 222-2-2288-8, dimana karakter "-" menunjukkan waktu *timeout* atau penekanan tombol *next*.

Penggunaan *multi-tap* dan adanya waktu *timeout* mengurangi keefektifan penulisan pesan karena dibutuhkan banyak penekanan. Salah satu solusi yang dapat diterapkan untuk menghilangkan *multi-tap* dan waktu *timeout* adalah dengan penerapan *predictive text entry* menggunakan model *bigram*. *Predictive text entry* menggunakan kamus yang dapat mencari kata sesuai dengan urutan tombol yang ditekan. Sebagai contoh jika ingin menuliskan kata "salah" maka penekanan yang dibutuhkan 72524. Sedangkan penggunaan model *bigram* diterapkan untuk memprediksi kata yang akan dituliskan berdasarkan satu kata sebelumnya. Dengan diterapkannya *predictive text entry* dengan menggunakan model *bigram* diharapkan dapat mengurangi jumlah penekanan tombol saat menuliskan pesan sehingga penulisan pesan pada telepon genggam menjadi lebih efisien.

1.2 Perumusan masalah

Permasalahan yang dijadikan objek penelitian dan pengembangan tugas akhir ini adalah :

1. Pembentukan kamus dan hashtable untuk data latih yang digunakan, serta perubahan isi kamus dan hashtable saat aplikasi digunakan.
2. Pencarian kandidat kata berdasarkan *previous word*, urutan tombol yang ditekan, serta frekuensi penggunaan tiap kata, yang mengacu pada kamus.

3. Menganalisis penggunaan model bigram pada *predictive text entry* untuk Bahasa Indonesia dari nilai KSPC (Keystroke per Character) yang dihasilkan, yang nantinya akan dibandingkan dengan nilai KSPC dari penggunaan *multi-tap*.

Dalam penelitian tugas akhir ini, objek penelitian dibatasi dengan ruang lingkup sebagai berikut :

1. Prediksi hanya dilakukan untuk huruf saja, tidak untuk angka maupun tanda baca.
2. Daftar kata pada kamus disesuaikan data latih yang digunakan.
3. Untuk kata yang tidak terdapat pada kamus (Out Of Vocabulary), maka pengguna dapat menuliskannya dengan cara multi-tap.

1.3 Tujuan

Tujuan dari Tugas Akhir ini adalah sebagai berikut :

1. Penerapan model *bigram* dalam memprediksi kata yang akan ditulis sebagai salah satu metode Predictive Text Entry untuk bahasa Indonesia sehingga penulisan teks menjadi lebih efisien. Penggunaan model *bigram* pada Predictive Text Entry diharapkan mampu mengurangi jumlah penekanan tombol dalam penulisan teks terutama yang menggunakan *12-keys keypad*.
2. Menganalisa penggunaan model *bigram* pada *Predictive Text Entry* ini berdasarkan nilai KSPC yang dihasilkan. KSPC merupakan rata-rata penekanan tombol yang dilakukan untuk setiap karakter dari suatu kata. Dengan demikian nilai KSPC yang diharapkan dari penggunaan model *bigram* adalah lebih kecil dari nilai KSPC *multi-tap*.

1.4 Metodologi penyelesaian masalah

Metodologi pembahasan yang digunakan dalam penelitian Tugas Akhir ini adalah:

1. Studi Literatur
 - a. Pencarian referensi
Tahapan ini meliputi pencarian referensi yang berhubungan dengan model *bigram*, *hash table*, *tree*, serta KSPC (Keystroke per Character) serta hal-hal lain yang berhubungan dengan Tugas Akhir ini yang bertujuan memberikan gambaran, detail, dan dasar teori yang jelas.
 - b. Pendalaman materi
Mempelajari dan memahami model *bigram*, pencarian kata menggunakan *hash table*, perancangan kamus, serta perhitungan nilai KSPC untuk metode ini. Tahap ini dilakukan selama proses penyusunan Tugas Akhir.
2. Analisa kebutuhan dan Perancangan Perangkat Lunak
Tahapan ini meliputi analisa kebutuhan dan perancangan perangkat lunak *Predictive Text Entry* untuk Bahasa Indonesia menggunakan model *Bigram*, yaitu:
 - a. Perancangan kamus yang digunakan

- b. Perancangan *hash table*.
 - c. Perancangan pencarian dalam kamus dan hash table dalam memprediksi kata.
3. Implementasi
Mengimplementasikan tree dan hashtable untuk merepresentasikan model bigram dalam memprediksi kata.
4. Evaluasi
Akan dilakukan pengujian terhadap hasil implementasi dengan memasukkan beberapa input teks dan akan dilihat solusi yang dihasilkan. Evaluasi terhadap objek penelitian meliputi diskusi dan konsultasi dengan dosen pembimbing dan pihak lain untuk penyempurnaan pengerjaan tugas akhir ini
5. Pembuatan laporan tugas akhir dan kesimpulan akhir.



5. Kesimpulan dan Saran

5.1 Kesimpulan

Kesimpulan yang dapat diambil dari Tugas Akhir Predictive text entry menggunakan model bigram adalah :

1. Model bigram dapat digunakan dalam memprediksi kata berdasarkan satu kata sebelumnya.
2. Penggunaan model bigram dalam penulisan pesan lebih baik (dari sisi jumlah penekanan tombol) jika dibandingkan penulisan pesan secara multi-tap.
3. Ada tidaknya prediksi kata ditentukan oleh ada tidaknya kata tersebut dalam kamus tree ataupun hashtable. Tidak adanya kata yang ingin dituliskan didalam kamus atau hashtable akan memperbesar nilai KSPC.
4. Jika semua kata yang dituliskan tidak terdapat dalam kamus, maka nilai KSPC model bigram akan sama dengan multi-tap.

5.2 Saran

Saran-saran yang dapat penulis uraikan untuk keperluan perangkat lunak ini selanjutnya adalah:

1. Implementasi model n-gram yang lain, misalnya model *trigram* (3-gram) ataupun *quadrigram* (4-gram)
2. Pada penelitian mendatang perlu dilakukan modifikasi dengan metode lain misalnya *syntactic word prediction* yang memperhatikan gramatikal dari suatu kata serta *semantic word prediction* yang memperhatikan striktur kata dalam kalimat.

Daftar pustaka

- [1] Amrullah, Afif, *Unified Modeling Language*.
- [2] Chadenat, O., Raballand, T., 2002, *Predictive Text for Mobile Device with Reduced Keyboards*.
- [3] Dunlop, M.D., Crossan, A., 1999, *Dictionary based text entry method for mobile phones. In Proceedings of the Second Workshop on Human Computer Interaction with Mobile Devices*.
- [4] Dunlop, M.D., Crossan, A., 2000, *Predictive Text Entry Methods for Mobile Phones. Personal Technologies*, pp. 134-143
- [5] Fachrurrozim, M., Yusliani, N., 2006, *Modul Praktikum " Struktur Data" Laboratorium Dasar Komputer Program Ilmu Komputer Universitas Sriwijaya*.
- [6] Fazly, A., 2002, *The Use of Syntax in Word Completion Utilities*. Master dissertation. Canada: University of Toronto.
- [7] Fowler, Martin, 2005, *UML Distilled Edisi 3 "Panduan Singkat Bahasa Pemodelan Objek Standar"*.
- [8] Hasselgren, J., Montnemery, E., Svensson, M., 2003, *HMS: A predictive text entry method using bigrams. In Extended Abstracts of 10th Conference of the European Chapter of the Association of Computational Linguistics* (pp. 43-49)
- [9] Janssen, Maarten, 2005, *Lexical vs. Dictionary Databases.: design choices of the MorDebe system. East-West Encounter: second international conference on Meaning-Text Theory*. Moscow, Russia.
- [10] Jurafsky, D., Martin, James H., 2006, *An Introduction to Speech Recognition, Computational Linguistics and Natural Language Processing*.
- [11] Liyanthy, M., Lidya, L., Shabaruddin, I., 2004, *Prototipe Model Kamus bahasa Indonesia-bahasa Sunda dan bahasa Sunda-bahasa Indonesia Berbasis Komputer*.
- [12] MacKenzie, I.S., Kober, H., Smith, D., Jones, T., & Skepner, E. (2001). *LetterWise: Prefix-based disambiguation for mobile text input. In Proceedings of the 14th annual ACM Symposium on User Interface software and technology*, (pp. 111-120)
- [13] MacKenzie, I.S., 2002, *KSPC (Keystrokes per Character) as a Characteristic of Text Entry Techniques*, In *Proc. Of 4th Int'l Symp. On HCI with Mobile Devices*, Springer, Germany, pp. 195-210.
- [14] Munir, Rinaldi, *Bahan Kuliah IF2151 Strategi Algoritmik "BFS dan DFS"*.
- [15] Sorensen, K., Springael, J., 2003, *Optimisation of SMS Keyboards with a Dictionary Using Local Search*.
- [16] "Hashtable", <http://www.oopweb.com> didownload pada tanggal 29 Agustus 2007