

ANALISIS PERBANDINGAN MODEL BIGRAM DAN TRIGRAM PADA PREDICTIVE TEXT ENTRY

Fransiska Vatma Paramita E.u¹, Andrian Rakhmatsyah², Retno Novi Dayawati³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Layanan pesan singkat/sms (Short Message Service) merupakan layanan pada telepon genggam yang paling sering digunakan. Dari layanan sms ini kemudian dikembangkan aplikasi predictive text entry untuk mengurangi jumlah penekanan tombol sehingga penulisan pesan dirasakan lebih efektif. Implementasi predictive text entry pada layanan sms umumnya menggunakan model unigram. Namun dewasa ini sudah mulai dikembangkan aplikasi predictive text entry menggunakan model bigram dan trigram sebagai solusi baru untuk memprediksi kata pada suatu inputan teks.

Pada tugas akhir ini, penulis mencoba membandingkan model bigram dan trigram dari segi memori yang digunakan dan waktu yang dibutuhkan untuk memprediksi kata serta KSPC (Keystroke per Character). Dari hasil analisis didapatkan kesimpulan bahwa dari sisi KSPC dan waktu yang diperlukan untuk prediksi, model trigram lebih baik daripada model bigram. Sedangkan untuk penggunaan memori untuk prediksi, sangat tergantung pada struktur kamus pada saat proses prediksi berlangsung

Kata Kunci : SMS, Predictive Text Entry, Unigram, Bigram, Trigram, KSPC.

Abstract

Short Message Service is a service that frequently used in handphone. From this service, predictive text entry application is developed to reduce the number of a button pressed so writing a message will be more effective. Implementation of predictive text entry in SMS service usually use unigram model. Nowadays predictive text entry has been developed by using bigram and trigram model as new solution to predict a word in a text entry.

In this final project, writer tries to compare between bigram and trigram model from memory that used, time that required to predicting the words and KSPC (Keystroke per Character) point of view. From this analyze can be conclude that from KSPC factor and time that required to predict, trigram model is better than bigram model. From memory that used to predict, it depends on data structure while predicting.

Keywords : SMS, Predictive Text Entry, Unigram, Bigram, Trigram, KSPC.

Telkom
University

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 yang 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*, dan *phone book*.

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* dengan model n-gram. *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. Penggunaan model bigram diterapkan untuk memprediksi kata yang akan dituliskan berdasarkan satu kata sebelumnya sedangkan model trigram berdasarkan dua kata sebelumnya. Karena jumlah kata yang digunakan untuk prediksi berbeda, dapat memungkinkan nilai KSPC (*Keystroke per Character*), waktu yang dibutuhkan dan memori yang digunakan untuk prediksi kata pada masing-masing model berbeda.

1.2 Perumusan masalah

Permasalahan yang dijadikan objek penelitian dan pengembangan tugas akhir ini adalah

1. Merancang pembentukan kamus untuk data yang digunakan, serta perubahan isi kamus saat aplikasi digunakan.

2. Merancang proses pencarian kandidat kata berdasarkan *previous word*, urutan tombol yang ditekan, serta frekuensi penggunaan tiap kata, yang mengacu pada kamus.
3. Menganalisis penggunaan model *bigram* dan *trigram* pada *predictive text entry* dari nilai KSPC yang dihasilkan.
4. Menganalisis penggunaan model bigram dan trigram dari waktu yang diperlukan untuk prediksi kata
5. Menganalisis penggunaan model bigram dan trigram dari memori yang digunakan untuk proses prediksi kata

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 dengan data yang digunakan.
3. Untuk kata yang tidak terdapat pada kamus (*Out Of Vocabulary*), maka pengguna dapat menuliskannya dengan cara *multi-tap*.
4. Model yang dibandingkan adalah model *bigram* dan *trigram*, kedua model ini juga dibandingkan dengan model *unigram* yang merupakan model yang digunakan pada aplikasi *predictive text entry* saat ini.

1.3 Tujuan

Tujuan dari Tugas Akhir ini adalah sebagai berikut

1. Penerapan model *bigram* dan *trigram* dalam memprediksi kata yang akan ditulis sebagai metode *Predictive Text Entry*. Penggunaan model *bigram* dan *trigram* pada *Predictive Text Entry* diharapkan mampu mengurangi jumlah penekanan tombol dalam penulisan teks terutama yang menggunakan *12-keys keypad*.
2. Menganalisis penggunaan model *bigram* dan *trigram* pada *Predictive Text Entry* ini berdasarkan nilai KSPC yang dihasilkan. KSPC merupakan rata-rata penekanan tombol yang dilakukan untuk setiap karakter dari suatu kata.
3. Menganalisis penggunaan model bigram dan trigram dari waktu yang diperlukan untuk prediksi kata
4. Menganalisis penggunaan model bigram dan trigram dari jumlah memori yang digunakan selama proses prediksi kata

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*, *trigram*, dan KSPC 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* dan *trigram*, perancangan kamus, serta perhitungan nilai KSPC untuk metode ini. Tahap ini dilakukan selama proses penyusunan Tugas Akhir.

2. Analisis kebutuhan dan Perancangan Perangkat Lunak

Tahapan ini meliputi analisis kebutuhan dan perancangan perangkat lunak *Predictive Text Entry* menggunakan model *bigram* dan *trigram*, yaitu:

- a. Perancangan kamus yang digunakan
 - b. Perancangan pencarian dalam kamus dalam memprediksi kata.
3. Implementasi
Mengimplementasikan tree dan hashtable untuk merepresentasikan model *bigram* dan *trigram* 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 Analisis Perbandingan Model Bigram dan Trigram pada Predictive Text Entry adalah :

1. Penggunaan model bigram dalam penulisan pesan lebih baik (dari sisi jumlah penekanan tombol) jika dibandingkan penulisan pesan dengan model trigram. Kedua model tersebut akan menjadi lebih baik lagi jika digabungkan dengan model unigram.
2. Dari keenam kombinasi tersebut, dilihat dari sisi jumlah penekanan tombol, gabungan model unigram, bigram dan trigram merupakan model yang paling baik dari model lainnya, sedangkan model trigram merupakan model yang paling buruk. Urutan model dari yang paling baik sampai paling buruk sebagai berikut :
 - unigram-bigram-trigram
 - unigram-bigram
 - unigram-trigram
 - bigram-trigram
 - bigram
 - trigram
3. Dari sisi waktu yang diperlukan untuk prediksi, penggunaan model trigram lebih baik daripada model bigram. Namun kedua model ini akan menjadi buruk jika digabungkan. Urutan model dari yang paling baik sampai paling buruk sebagai berikut (berkebalikan dengan urutan jika dilihat dari sisi jumlah penekanan tombol):
 - trigram
 - bigram
 - bigram-trigram
 - unigram-trigram
 - unigram-bigram
 - unigram-bigram-trigram
4. Dari sisi memori yang digunakan, baik model bigram maupun trigram sangat tergantung pada struktur kamus pada saat proses prediksi berlangsung.

5.2 Saran

Saran-saran yang dapat penulis uraikan untuk keperluan analisis selanjutnya adalah:

1. Untuk penelitian mendatang perlu dilakukan analisis predictive text entry yang diimplementasikan langsung pada telepon genggam.
2. Analisis predictive text entry pada inputan keyboard QWERTY.
3. Karena tujuan predictive text entry adalah mengurangi jumlah penekanan tombol, maka gabungan model unigram, bigram dan trigram yang lebih baik untuk implementasi pada telepon genggam. Namun karena semakin

banyak model yang diimplementasikan, hal ini dapat menyebabkan kamus menjadi semakin besar. Besar kamus tersebut dapat dibatasi dengan menentukan jumlah kata maksimal yang bisa ditampung. Sehingga, untuk kata yang jarang digunakan (frekuensi atau usesnya kecil) dapat dibuang dari kamus.



Daftar pustaka

- [1] Carmignani, Nicola, 2006, *Predicting Words and Sentences using Statistical Model*, Departement or Computer Science University of Pisa.
- [2] Gregory W. Lesh, Bryan J. Moulton, *Effects of N-Gram Order and Training Text Size On Word Prediction*, Enkidu Research, Inc, Lockport, New York, USA.
- [3] "Hashtable". <http://www.oopweb.com> didownload pada tanggal 23 April 2008.
- [4] 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.
- [5] Mayasari, Fransiska Dani Hesti. 2007. *Predictive Text Entry Menggunakan Model Bigram*. Departemen Teknik Informatika Sekolah Tinggi Teknologi Telkom, Bandung.
- [6] M.D. Dunlop, A. Crossan, 2000, *Predictive Text Entry Methods for Mobile Phones*, *Personal Technologies*, pp. 134-143.
- [7] Scott MacKenzie, Heddy Kober, 2007, *LetterWise: Prefix-based Disambiguation for Mobile Text Input*.
- [8] M. Fachrurrozim, Yusliani, N., 2006, *Modul Praktikum "Struktur Data" Laboratorium Dasar Komputer Program Ilmu Komputer Universitas Sriwijaya*.
- [9] Munir, Rinaldi. *Bahan Kuliah IF2151 Strategi Algoritmik "BFS dan DFS"*.
- [10] Sima'an, Khalil, *Language and Speech Processing*, Institute For Logic, Language and Computation, Universiteit van Amsterdam.