

ABSTRAK

Sendy Andrian Sugianto:

Skripsi

Pembuatan Aplikasi *Predictive Text* Menggunakan Metode *N-Gram-Based*

Penggunaan teknologi dalam pengetikan sudah sangat berkembang. Bermulai dari mesin ketik, lalu mengalami perubahan kemajuan teknologi menjadi komputer, dan yang paling marak adalah telepon genggam. Semua itu adalah media pengetikkan yang masih berkembang sampai saat ini. Teknologi tidak hanya berkembang pada media pengetikkannya saja, tetapi juga berkembang pada sistem proses kata. Sistem-sistem prediksi pun banyak bermunculan untuk membantu mempercepat proses pengetikkan.

Oleh karena itu, pada skripsi ini dilakukan pembuatan aplikasi untuk melakukan prediksi kata dengan menggunakan *n-gram* sebagai metode dasar melakukan proses prediksi. Proses dimulai dengan memecah kata per kata dan mengelompokkannya sesuai dengan *language model*. Kemudian, dilakukan proses *scoring* untuk menentukan kata mana yang sesuai untuk menjadi pilihan prediksi kata.

Hasil pengujian menunjukkan bahwa metode *n-gram* sebagai metode dasar dalam proses prediksi sangatlah membantu pemilahan kata, sehingga proses prediksi menjadi lebih efektif, mampu menghasilkan prediksi efektif di atas 20% dari total prediksi yang terjadi. *Keystroke saving* yang dapat dihasilkan dapat mencapai 50% tergantung dari data *training* yang digunakan. Selain dari pada metode *n-gram* sendiri, pengaturan bobot untuk masing-masing *score* kata juga sangat mempengaruhi proses prediksi kata.

Kata Kunci:

Predictive Text, N-Gram, Keystroke Saving.

ABSTRAK

Sendy Andrian Sugianto:

Thesis

Implementation of Predictive Text Using N-Gram-Based

The Typing technology has been highly developed. From the used of typewriters, then it changes into the computer technology, and the most prevalent is mobile device. They are the typing technology that still involving until today. Technology is not only about the typing method, but also thrives on word processing systems. There are many words prediction that appear to help human speed up their typing.

Therefore, this thesis making an application to perform word prediction using n-gram based method to do the prediction process. The process begins by parsing the words and grouping them according to the language model. Then, scoring process determine which words are the appropriate choice of word prediction.

The test result says that n-gram based method is very helpful in the process of sorting word prediction, so that prediction process became more effective, and able to produce effective prediction above 20% of the total prediction. Keystroke savings that can be generated can reach 50% depending on the training data are used. Other than n-gram based method, setting of the weights for each scoring method also affect the process of word prediction.

Keywords:

Predictive Text, N-Gram, Keystroke Saving.

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	iii
KATA PENGANTAR	iv
ABSTRAK.....	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xiv
DAFTAR SEGMENT PROGRAM	xvi
DAFTAR LAMPIRAN.....	xvii
1. PENDAHULUAN	Error! Bookmark not defined.
1.1. Latar Belakang Masalah	Error! Bookmark not defined.
1.2. Perumusan Masalah.....	1
1.3. Tujuan Skripsi	Error! Bookmark not defined.
1.4. Ruang Lingkup	Error! Bookmark not defined.
1.5. Metodologi Penelitian	2
1.6. Sistematika Penulisan.....	3
2. LANDASAN TEORI.....	4
2.1. <i>Predictive Text</i>	4
2.2. <i>N-Gram</i>	Error! Bookmark not defined.
2.3. <i>Language Model</i>	9
2.4. <i>Semantic Affinity</i>	Error! Bookmark not defined.
3. DESAIN SISTEM.....	13
3.1. Garis Besar Sistem Kerja Perangkat Lunak	13
3.1.1. <i>Load File</i>	15
3.1.2. Proses <i>Parsing Data</i>	15
3.1.3. Proses Prediksi.....	18
3.1.4. Proses <i>Scoring</i>	19
3.1.5. Pengolahan <i>Output</i>	22
3.2. Desain <i>Interface</i> pada Aplikasi	23

4.	IMPLEMENTASI SISTEM.....	Error! Bookmark not defined.
4.1.	Implementasi Aplikasi yang Digunakan	Error! Bookmark not defined.
4.2.	Struktur <i>Class</i> dari Perangkat Lunak	Error! Bookmark not defined.
4.2.1.	Fungsi <i>ClearData</i>	27
4.2.2.	Fungsi <i>CopyData</i>	27
4.2.3.	Fungsi <i>SaveData</i>	28
4.2.4.	Fungsi <i>LoadData</i>	28
4.2.5.	Fungsi <i>AddData</i>	29
4.2.6.	Fungsi <i>CleaningString</i>	30
4.2.7.	Fungsi <i>ParseString</i>	31
4.2.8.	Fungsi <i>GetLastNWord</i>	31
4.2.9.	Fungsi <i>CountWordInCorpus</i>	32
4.2.10.	Fungsi <i>GetTotalCount</i>	33
4.2.11.	Fungsi <i>CountFrequencyScore</i>	33
4.2.12.	Fungsi <i>CountSemanticScore</i>	Error! Bookmark not defined.
4.2.13.	Fungsi <i>CountNgramScore</i>	34
4.2.14.	Fungsi <i>Predictive</i>	35
4.2.15.	Fungsi <i>FilterPredictiveData</i>	36
4.2.16.	Fungsi <i>SortData</i>	36
4.2.17.	Fungsi <i>PrintData</i>	37
5.	PENGUJIAN SISTEM	38
5.1.	Sistem Aplikasi.....	38
5.2.	Pengujian Validasi Sistem.....	40
5.2.1.	Pengujian pada <i>Bigram</i>	40
5.2.2.	Pengujian pada <i>Trigram</i>	48
5.3.	Pengujian Bobot pada Tiap Metode <i>Scoring</i>	57
5.3.1.	Data Pengujian.....	58
5.3.1.1.	Data Satu.....	58
5.3.1.2.	Data Dua	59
5.3.1.3.	Data Tiga	60
5.3.1.4.	Data Empat	61
5.3.1.5.	Data Lima	61
5.3.1.6.	Data Enam	62
5.3.1.7.	Data Tujuh	63
5.3.1.8.	Data Delapan	63
5.3.1.9.	Data Sembilan.....	64
5.3.1.10.	Data Sepuluh.....	64
5.3.2.	Hasil Pengujian.....	65
5.3.2.1.	Pengujian dengan Data <i>Training</i> dan Data <i>Input</i> yang sama	65
5.3.2.2.	Pengujian dengan Data <i>Training</i> dan Data <i>Input</i> yang berbeda.....	71
5.3.2.2.1.	Pengujian Pertama	71
5.3.2.2.2.	Pengujian Kedua.....	77
5.4.	Pengujian <i>Hardware</i>	82

6. KESIMPULAN DAN SARAN.....	86
6.1. Kesimpulan.....	86
6.2. Saran.....	86
DAFTAR REFERENSI	88

LAMPIRAN

DAFTAR TABEL

2.1.	Hasil Pembentukan Model <i>N-Gram</i> Menggunakan <i>Bigram</i> pada Dokumen Pertama.....	7
2.2.	Hasil Pembentukan Model <i>N-Gram</i> Menggunakan <i>Bigram</i> pada Dokumen Kedua	8
2.3.	Hasil <i>Bigram</i> dari Kedua Dokumen yang Sudah Digabung dan Diurutkan Sesuai Dengan Jumlah Frekuensi Kemunculannya.....	9
2.4.	Tabel <i>Bigram</i> dan <i>Trigram</i>	Error! Bookmark not defined.
4.1.	Tabel Fungsi pada <i>Class</i> <i>CNGram</i>	28
5.1.	Isi Data <i>Language Model Bigram</i>	41
5.2.	Hasil Perhitungan <i>Frequency Score</i> Manual Kasus Pertama <i>Bigram</i>	42
5.3.	Hasil Perhitungan <i>Semantic Score</i> Manual Kasus Pertama <i>Bigram</i>	42
5.4.	Hasil Perhitungan <i>Ngram Score</i> Manual Kasus Pertama <i>Bigram</i>	43
5.5.	Hasil Total <i>Score</i> Manual Kasus Pertama <i>Bigram</i>	43
5.6.	Hasil Perhitungan <i>Frequency Score</i> Manual Kasus Kedua <i>Bigram</i>	45
5.7.	Hasil Perhitungan <i>Semantic Score</i> Manual Kasus Kedua <i>Bigram</i>	45
5.8.	Hasil Perhitungan <i>Ngram Score</i> Manual Kasus Kedua <i>Bigram</i>	46
5.9.	Hasil Total <i>Score</i> Manual Kasus Kedua <i>Bigram</i>	47
5.10.	Tabel Hasil Proses <i>Filtering</i> Kata Prediksi Kasus Kedua <i>Bigram</i>	48
5.11.	Isi Data <i>Language Model Trigram</i>	49
5.12.	Hasil Perhitungan <i>Frequency Score</i> Manual Kasus Pertama <i>Trigram</i>	50
5.13.	Hasil Perhitungan <i>Semantic Score</i> Manual Kasus Pertama <i>Trigram</i>	50

5.14. Hasil Perhitungan <i>Ngram Score</i> Manual Kasus Pertama <i>Trigram</i>	51
5.15. Hasil Total <i>Score</i> Manual Kasus Pertama <i>Trigram</i>	51
5.16. Hasil Perhitungan <i>Frequency Score</i> Manual Kasus Kedua <i>Trigram</i>	52
5.17. Hasil Perhitungan <i>Semantic Score</i> Manual Kasus Kedua <i>Trigram</i>	53
5.18. Hasil Perhitungan <i>Ngram Score</i> Manual Kasus Kedua <i>Trigram</i>	53
5.19. Hasil Total <i>Score</i> Manual Kasus Kedua <i>Trigram</i>	55
5.20. Tabel Hasil Proses <i>Filtering</i> Kata Prediksi Kasus Kedua <i>Trigram</i>	56
5.21. Formula Pengujian <i>Scoring</i>	58
5.22. Hasil Penghitungan Formula <i>Scoring</i> Pertama Pada <i>Bigram</i>	65
5.23. Hasil Penghitungan Formula <i>Scoring</i> Kedua Pada <i>Bigram</i>	66
5.24. Hasil Penghitungan Formula <i>Scoring</i> Ketiga Pada <i>Bigram</i>	67
5.25. Hasil Penghitungan Formula <i>Scoring</i> Keempat Pada <i>Bigram</i>	67
5.26. Hasil Penghitungan Formula <i>Scoring</i> Pertama Pada <i>Trigram</i>	68
5.27. Hasil Penghitungan Formula <i>Scoring</i> Kedua Pada <i>Trigram</i>	68
5.28. Hasil Penghitungan Formula <i>Scoring</i> Ketiga Pada <i>Trigram</i>	69
5.29. Hasil Penghitungan Formula <i>Scoring</i> Keempat Pada <i>Trigram</i>	69
5.30. Hasil Pengelompokan Data <i>Score</i> Terbesar pada <i>Bigram</i>	70
5.31. Hasil Pengelompokan Data <i>Score</i> Terbesar pada <i>Trigram</i>	71
5.32. Hasil Penghitungan Formula <i>Scoring</i> Pertama Pada <i>Bigram</i>	72

5.33. Hasil Penghitungan Formula <i>Scoring</i> Kedua Pada <i>Bigram</i>	72
5.34. Hasil Penghitungan Formula <i>Scoring</i> Ketiga Pada <i>Bigram</i>	73
5.35. Hasil Penghitungan Formula <i>Scoring</i> Keempat Pada <i>Bigram</i>	73
5.36. Hasil Penghitungan Rata-Rata Nilai Tiap Formula Pada <i>Bigram</i>	74
5.37. Hasil Penghitungan Formula <i>Scoring</i> Pertama Pada <i>Trigram</i>	74
5.38. Hasil Penghitungan Formula <i>Scoring</i> Kedua Pada <i>Trigram</i>	75
5.39. Hasil Penghitungan Formula <i>Scoring</i> Ketiga Pada <i>Trigram</i>	75
5.40. Hasil Penghitungan Formula <i>Scoring</i> Keempat Pada <i>Trigram</i>	76
5.41. Hasil Penghitungan Rata-Rata Nilai Tiap Formula Pada <i>Trigram</i>	76
5.42. Hasil Penghitungan Formula <i>Scoring</i> Pertama Pada <i>Bigram</i>	77
5.43. Hasil Penghitungan Formula <i>Scoring</i> Kedua Pada <i>Bigram</i>	78
5.44. Hasil Penghitungan Formula <i>Scoring</i> Ketiga Pada <i>Bigram</i>	78
5.45. Hasil Penghitungan Formula <i>Scoring</i> Keempat Pada <i>Bigram</i>	79
5.46. Hasil Penghitungan Rata-Rata Nilai Tiap Formula Pada <i>Bigram</i>	79
5.47. Hasil Penghitungan Formula <i>Scoring</i> Pertama Pada <i>Trigram</i>	80
5.48. Hasil Penghitungan Formula <i>Scoring</i> Kedua Pada <i>Trigram</i>	80
5.49. Hasil Penghitungan Formula <i>Scoring</i> Ketiga Pada <i>Trigram</i>	81
5.50. Hasil Penghitungan Formula <i>Scoring</i> Keempat Pada <i>Trigram</i>	81
5.51. Hasil Penghitungan Rata-Rata Nilai Tiap Formula Pada <i>Trigram</i>	82
5.52. Spesifikasi <i>Hardware</i> Pengujian	83

5.53. Hasil Pengujian Pertama *Timing Process* pada *Hardware* yang berbeda84

5.54. Hasil Pengujian Kedua *Timing Process* pada *Hardware* yang berbeda.....84

DAFTAR GAMBAR

3.1.	Desain Sistem Kerja <i>Predictive Text</i>	14
3.2.	Diagram Alur Proses <i>Load File</i>	15
3.3.	Diagram Alur Proses <i>Parsing Data</i>	Error! Bookmark not defined.
3.4.	Diagram Alur Proses <i>Cleaning String</i>	17
3.5.	Diagram Alur Pengelompokan Kata Berdasarkan Metode <i>N-Gram</i>	17
3.6.	Diagram Alur Proses Prediksi	18
3.7.	Diagram Alur Proses <i>Get Input Data</i>	18
3.8.	Diagram Alur Proses <i>Searching dan Scoring</i>	19
3.9.	Diagram Alur Proses <i>Scoring</i>	20
3.10.	Diagram Alur <i>Frequency Scoring</i>	20
3.11.	Diagram Alur <i>Semantic Scoring</i>	21
3.12.	Diagram Alur <i>Ngram Scoring</i>	22
3.13.	Diagram Alur Pengolahan <i>Output</i>	23
3.14.	Desain <i>Interface Form</i> Awal	23
3.15.	Desain <i>Interface Form</i> Aplikasi	24
3.16.	Desain <i>Interface Form Score</i>	24
5.1.	Tampilan <i>Form</i> Awal	38
5.2.	Tampilan Form Aplikasi Saat <i>Window Predictive Text</i> Muncul.....	39
5.3.	Tampilan <i>Form Score</i>	39
5.4.	Proses <i>Input</i> Kalimat pada Pengujian Validasi Sistem.....	40
5.5.	Contoh Pengujian Kasus Pertama pada <i>Bigram</i>	42
5.6.	Contoh Pengujian Kasus Kedua pada <i>Bigram</i>	44

5.7. Contoh Pengujian Kasus Pertama pada <i>Trigram</i>	50
5.8. Contoh Pengujian Kasus Kedua pada <i>Trigram</i>	52
5.9. Data Cerita Pendek “Kerasukan Arwah Penasaran”	58
5.10. Gambar Proses <i>Input</i> Cerita Pendek untuk Data Satu	59
5.11. Gambar Proses Pengetikan Kembali Cerita Pendek untuk Data Satu	59
5.12. Data Cerita Pendek “Teknologi Canggih”	60
5.13. Data Opini Ekonomi Politik	60
5.14. Data Berita tentang Honda	61
5.15. Data Artikel “Proyek Baru Trio Harry Potter”	62
5.16. Data Artikel “Cara Alami Usir Ketombe”	62
5.17. Data Cerita Pendek Misteri Bunyi Gamelan	63
5.18. Data Berita Gadget Mozilla.....	63
5.19. Data Berita Jokowi Kebut Bangun Rusun	64
5.20. Data Artikel Opini	65

DAFTAR SEGMENT PROGRAM

4.1. Variabel yang Ada pada <i>Class</i> CNGram	26
4.2. <i>Source code</i> fungsi <i>ClearData</i>	27
4.3. <i>Source code</i> fungsi <i>CopyData</i>	28
4.4. <i>Source code</i> fungsi <i>SaveData</i>	28
4.5. <i>Source code</i> fungsi <i>LoadData</i>	28
4.6. <i>Source code</i> fungsi <i>AddData</i>	29
4.7. <i>Source code</i> fungsi <i>CleaningString</i>	30
4.8. <i>Source code</i> fungsi <i>RemoveSymbol</i>	30
4.9. <i>Source code</i> fungsi <i>IsAlphabet</i>	30
4.10. <i>Source code</i> fungsi <i>ParseString</i>	31
4.11. <i>Source code</i> fungsi <i>GetLastNWord</i>	31
4.12. <i>Source code</i> fungsi <i>CountWordInCorpus</i>	32
4.13. <i>Source code</i> fungsi <i>GetTotalCount</i>	33
4.14. <i>Source code</i> fungsi <i>CountFrequencyScore</i>	33
4.15. <i>Source code</i> fungsi <i>CountSemanticScore</i>	33
4.16. <i>Source code</i> fungsi <i>CountNgramScore</i>	34
4.17. <i>Source code</i> fungsi <i>createNgram</i>	35
4.18. <i>Source code</i> fungsi <i>Predictive</i>	35
4.19. <i>Source code</i> fungsi <i>FilterPredictiveData</i>	36
4.20. <i>Source code</i> fungsi <i>SortData</i>	36
4.21. <i>Source code</i> fungsi <i>PrintData</i>	37

DAFTAR LAMPIRAN

1. Hasil Analisis Penggunaan Model *N-gram* Terhadap *Language Model*90