



UNIVERSITI PUTRA MALAYSIA

**SISTEM BANTUAN PEMBINAAN KAMUS
BERASASKAN PANGKALAN DATA TEKS BEBAS**

MUHAMAD TAUFIK BIN ABDULLAH

FSKTM 1992 1



UNIVERSITI PUTRA MALAYSIA

**SISTEM BANTUAN PEMBINAAN KAMUS
BERASASKAN PANGKALAN DATA TEKS BEBAS**

MUHAMAD TAUFIK BIN ABDULLAH

FSKTM 1992 1

**SISTEM BANTUAN PEMBINAAN KAMUS
BERASASKAN PANGKALAN DATA TEKS BEBAS**

MUHAMAD TAUFIK BIN ABDULLAH

Tesis Dikemukakan Kepada
Fakulti Sains Komputer Dan Sistem Maklumat,
Universiti Teknologi Malaysia
Sebagai Memenuhi Syarat Penganugerahan
Ijazah Sarjana Sains Komputer

1992

PENGHARGAAN

Pertama sekali penulis ingin memanjatkan segala pujian ke hadrat Allah s.w.t. yang Maha Pemurah dengan segala nikmat terhadap hamba-hambaNya. Selawat dan salam ke atas junjungan kita, Nabi Muhammad s.a.w.

Penulis juga ingin merakamkan terima kasih yang tidak terhingga kepada En. Safie Mat Yatim dan En. Sarudin Kari selaku penyelia-penyelia yang telah memberikan bimbingan, penyeliaan dan tunjuk ajar sehingga tesis ini dapat disiapkan, dengan izin Allah.

Penghargaan juga diberi kepada sesiapa yang sama ada secara langsung atau tidak langsung membantu dalam projek penyelidikan ini.

ABSTRAK

Sebuah kamus bukan sahaja berguna untuk memberikan maklumat tentang makna dan penggunaan perkataan, tetapi juga dapat dijadikan sebagai sumber rujukan utama tentang perbendaharaan kata sesuatu bahasa. Bidang perkamusan amat luas kerana ianya meliputi pelbagai subdisiplin dalam bidang linguistik termasuklah semantik, sintaksis, morfologi dan sebagainya. Dalam proses pembangunan kamus, penyusunnya perlu menjalankan penyelidikan dan mempunyai kepakaran di dalam pelbagai subdisiplin tersebut.

Tesis ini membincangkan hasil penyelidikan pengkomputeran yang berkaitan dengan bidang perkamusan. Kajian yang dijalankan melibatkan pengkomputeran sebahagian proses pembinaan kamus sebagai satu pendekatan untuk mempercepat dan memudahkan penyusunan kamus. Hasil yang diperolehi ialah algoritma bagi proses analisis morfologi perkataan bahasa Malaysia. Sebuah perisian komputer yang dinamakan *Sistem Bantuan Pembinaan Kamus (SBPK)* juga dibangunkan untuk membantu ahli leksikografi menjalankan tugas penyusunan kamus bahasa Malaysia. Sistem ini menggabungkan kemudahan untuk melaksanakan analisis morfologi, penentuan perkataan, capaian teks daripada pangkalan data teks bebas dan penyuntingan teks. Dengan bantuan sistem ini, dijangka dapat mempercepat dan memudahkan kerja pembinaan sesebuah kamus berbanding dengan kaedah manual yang sedia ada.

ABSTRACT

A dictionary not only provides information of meaning and usage of word but also serves as the main reference of the vocabulary of a language. Lexicography is very wide as it covers various sub disciplines in linguistic including semantics, syntax, morphology, etc. In the dictionary development process, lexicographers should conduct research and knowledgeable in those sub disciplines.

This thesis discussed the result of computerization related to lexicography. The research involved computerizing a part of the dictionary development process as an approach to expedite and facilitate the process. The result was an algorithm for morphology analysis process for the word in Malay language. A software name as *Sistem Bantuan Pembangunan Kamus (SBPK)* was develops to assist the lexicographer to compile bahasa Malaysia dictionary. This system integrated the facilities to execute the morphology analysis, word validation, text retrieval from free text database and text editing. It is expect with the help of this system will speed up and facilitate the compilation of the dictionary compared to the existing manual method.

KANDUNGAN

Muka Surat

Judul	
Perakuan	
Penghargaan	i
Abstrak	ii
Abstract	iii
Kandungan	iv
Senarai Jadual	viii
Senarai Rajah	ix
Bab I Pengenalan	1
1.1 Latar Belakang Kajian	1
1.2 Pernyataan Masalah	2
1.3 Matlamat Penyelidikan	4
1.4 Struktur Tesis	4
Bab II Morfologi, Perkataan Dan Kamus	6
2.1 Pendahuluan	6
2.2 Morfologi Bahasa Malaysia	7
2.2.1 Imbuhan	8
2.2.2 Perkataan	9
2.3 Pembentukan Kata Terbitan	10
2.3.1 Kata Terbitan Berawalan	10
2.3.2 Kata Terbitan Berakhiran	12

2.3.3	Kata Terbitan Berapitan	13
2.3.4	Kata Terbitan Bersisipan	14
2.4	Pembinaan Kamus	15
2.4.1	Pentakrifan Makna Perkataan	16
2.4.2	Penyusunan Perkataan	17
Bab III	Pangkalan Data Teks Bebas	19
3.1	Pendahuluan	19
3.2	Sistem Pangkalan Data Teks Bebas	20
3.2.1	Pengindeksan Dokumen	20
3.2.2	Capaian Dokumen	22
Bab IV	Pendekatan Pembangunan	
	Sistem Bantuan Pembinaan Kamus	24
4.1	Pendahuluan	24
4.2	Penyediaan Data Rujukan	25
4.2.1	Senarai Imbuhan-Huruf	25
4.2.2	Matriks Imbuhan-Huruf	26
4.2.3	Automaton Keadaan Terhingga	27
4.2.3.1	Imbuhan Awalan	31
4.2.3.2	Imbuhan Akhiran	32
4.2.3.3	Imbuhan Apitan	33
4.2.3.4	Partikel	36
4.2.3.5	Kata Ganti Singkat	37
4.2.4	Jadual Peralihan Keadaan	38
4.3	Proses Penghuraian Perkataan	41
4.3.1	Analisis Perkataan	42
4.3.2	Analisis Imbuhan	43

4.3.2.1	Ujian Imbuhan Khas	43
4.3.2.2	Ujian Partikel Dan Kata Ganti Singkat	45
4.3.2.3	Ujian Matriks Imbuhan-Huruf	46
4.3.2.4	Ujian Satu Suku Kata	47
4.4	Proses Pembentukan Kata Terbitan	48
4.4.1	Terbitan Matriks Imbuhan-Huruf	49
4.4.2	Terbitan Satu Suku Kata	49
4.4.3	Terbitan Imbuhan Khas	50
4.4.4	Terbitan Perkataan Khas	50
4.5	Proses Penentusahan Perkataan	51
Bab V	Sistem Bantuan Pembinaan Kamus	53
5.1	Pendahuluan	53
5.2	Reka Bentuk Sistem	53
5.3	Modul Analisis Morfologi	55
5.3.1	Penyediaan Matriks Imbuhan-Huruf	56
5.3.2	Penghuraian Dan Pembentukan Kata	57
5.3.2.1	Penyediaan Perkataan	57
5.3.2.2	Penghuraian Perkataan	59
5.3.2.3	Pembentukan Kata Terbitan	60
5.4	Modul Penentusahan Perkataan	60
5.5	Modul Paparan Dokumen	62
5.6	Modul Penyunting Teks	63
5.7	Antara Muka Sistem Dan Pengguna	64
Bab VI	Penutup	65
6.1	Pencapaian Kajian	65
6.2	Kesimpulan	67

Rujukan	70
Lampiran I : Senarai Huruf Vokal Dan Konsonan	72
Lampiran II : Senarai Imbuhan-Huruf	74
Lampiran III : Matriks Imbuhan-Huruf	76
Lampiraan IV : Jadual Peralihan Keadaan	77
Lampiraan V : Algoritma Analisis Perkataan	83
Lampiraan VI : Algoritma Analisis Imbuhan	84
Lampiran VII : Algoritma Pembentukan Kata Terbitan	87
Lampiran VIII : Antara Muka Sistem-Pengguna	90

SENARAI JADUAL

	<u>Muka Surat</u>
Jadual 2.1 : Imbuhan Awalan Bagi Kata Nama	11
Jadual 2.2 : Imbuhan Awalan Bagi Kata Kerja	11
Jadual 2.3 : Imbuhan Awalan Bagi Kata Adjektif	12
Jadual 2.4 : Imbuhan Akhiran Bagi Kata Nama	12
Jadual 2.5 : Imbuhan Akhiran Bagi Kata Kerja	12
Jadual 2.6 : Imbuhan Apitan Bagi Kata Nama	13
Jadual 2.7 : Imbuhan Apitan Bagi Kata Kerja	14
Jadual 2.8 : Imbuhan Apitan Bagi Kata Adjektif	14
Jadual 2.9 : Imbuhan Sisipan Bagi Kata Nama	14
Jadual 2.10 : Imbuhan Sisipan Bagi Kata Adjektif	15
Jadual 4.1 : Pemecahan Imbuhan Apitan	33
Jadual 4.2 : Pemilihan Gambar Rajah Bahagian Akhiran Bagi Apitan	36
Jadual 4.3 : Contoh Jadual Peralihan Keadaan	39
Jadual 4.4 : Imbuhan Khas	44
Jadual 4.5 : Perubahan Huruf Pertama Perkataan	44
Jadual 4.6 : Kemungkinan Huruf Pertama Bagi Kata Dasar	45
Jadual 4.7 : Imbuhan Perkataan Satu Suku Kata	48
Jadual 4.8 : Perkataan Khas	51

SENARAI RAJAH

	<u>Muka Surat</u>
Rajah 3.1 : Struktur Pangkalan Data	21
Rajah 3.2 : Reka Bentuk Umum Subsistem Capaian Dokumen	23
Rajah 4.1 : Gambar Rajah Keadaan (Kaedah Kiri Ke Kanan)	29
Rajah 4.2 : Gambar Rajah Keadaan (Kaedah Kanan Ke Kiri)	30
Rajah 4.3 : Gambar Rajah Keadaan Imbuhan Awalan	32
Rajah 4.4 : Gambar Rajah Keadaan Imbuhan Akhiran	33
Rajah 4.5 : Gambar Rajah Keadaan Imbuhan Apitan (Awalan)	34
Rajah 4.6 : Gambar Rajah Keadaan Imbuhan Apitan (Akhiran I)	35
Rajah 4.7 : Gambar Rajah Keadaan Imbuhan Apitan (Akhiran II)	35
Rajah 4.8 : Gambar Rajah Keadaan Imbuhan Apitan (Akhiran III)	35
Rajah 4.9 : Gambar Rajah Keadaan Imbuhan Partikel	37
Rajah 4.10 : Gambar Rajah Keadaan Imbuhan Kata Ganti Singkat (Awalan)	37
Rajah 4.11 : Gambar Rajah Keadaan Imbuhan Kata Ganti Singkat (Akhiran)	38
Rajah 5.1 : Hubungan Antara SPDTB Dan SBPK	54
Rajah 5.2 : Modul Analisis Morfologi	56
Rajah 5.3 : Hubungan Subindeks, Fail-Fail Indeks Dan Dokumen	58
Rajah 5.4 : Modul Penentusahan Perkataan	61
Rajah 5.5 : Modul Paparan Dokumen	63
Rajah 5.6 : Modul Penyunting Teks	64

BAB I

PENGENALAN

1.1 Latar Belakang Kajian

Kegunaan asas bagi sesebuah kamus ialah membolehkan pengguna mendapatkan maklumat tentang makna atau takrif bagi perkataan. Sehubungan dengan itu, adalah menjadi fungsi utama bagi sesebuah kamus untuk merakamkan perbendaharaan kata dalam sesuatu bahasa dan memberikan panduan tentang penggunaan bagi kata-kata tersebut [Asmah 1987].

Kamus boleh dianggap sebagai 'authority' atau pemberi kata putus dalam aspek tertentu mengenai bahasa, termasuklah tentang ejaan, makna kata dan penggunaan kata. Sebagai contoh, untuk menyelesaikan perselisihan faham tentang penggunaan sesuatu kata, biasanya kamus akan menjadi pegangan di samping merujuk kepada ahli bahasa. Ahli bahasa juga harus berpegang kepada 'authority' yang sudah berakar umbi, iaitu kamus [Asmah 1987]. Sekiranya kamus yang digunakan tidak lengkap kandungannya atau tidak dapat memberikan makna yang tepat, ini boleh menambahkan lagi kekeliruan tentang penggunaan bahasa yang betul dan seragam.

Kerja pembangunan sesebuah kamus yang lengkap bagi sesuatu bahasa merupakan kerja yang rumit dan memakan masa. Penyusunan kamus perlu dilakukan dengan penuh tanggungjawab terutama di dalam menentukan makna dan penggunaan kata dengan tepat. Proses pembangunannya melibatkan penyelidikan dan pengetahuan tertentu, terutamanya yang mempunyai perkaitan secara langsung dengan bidang

linguistik.

Buat masa ini proses penyusunan kamus di negara ini dijalankan secara manual. Kerja penyusunan kamus biasanya dimulai dengan usaha mengumpul perkataan daripada pelbagai jenis bahan terbitan seperti kamus (yang sedia ada), buku, majalah, akhbar dan sebagainya. Bahan-bahan terbitan ini akan menjadi sumber rujukan utama tentang perbendaharaan kata bagi bahasa yang berkenaan dan sebagai rujukan di dalam menentukan makna perkataan, iaitu untuk mentakrifkan perkataan-perkataan yang telah dikumpulkan.

1.2 Pernyataan Masalah

Di dalam proses penyusunan kamus terdapat beberapa masalah utama yang dapat dikenalpasti iaitu:

(i) *Mengumpul Perkataan*

Masalah pengumpulan perkataan untuk mendapatkan perbendaharaan kata bagi bahasa yang berkenaan daripada bahan-bahan terbitan adalah di antara masalah utama di dalam proses penyusunan kamus. Masalah ini timbul disebabkan oleh sesebuah kamus yang lengkap memerlukan bahan-bahan terbitan dengan kuantiti yang besar untuk dijadikan sebagai rujukan.

(ii) *Pentakrifan Makna*

Di dalam pentakrifan makna bagi setiap perkataan, sekumpulan teks diperlukan untuk mengkaji konteks penggunaan bagi perkataan tersebut.

Masalahnya ialah untuk mencari dan mendapatkan teks yang berkenaan di dalam bahan-bahan terbitan yang telah dikumpulkan.

(iii) *Mengumpul Kata Entri*

Penyusun kamus biasanya akan mengumpulkan perkataan-perkataan yang akan dijadikan kata entri mengikut kumpulan kata. Kumpulan kata ini terdiri daripada kata dasar¹ dan diikuti dengan kata-kata terbitan daripada kata dasar tersebut. Di dalam menjalankan pengumpulan kata ini, penyusun kamus perlu melakukan analisis morfologi bagi perkataan-perkataan tersebut untuk mendapatkan kata dasarnya. Daripada kata dasar tersebut juga, penyusun kamus akan menjalankan proses pembentukan kata-kata terbitan baru untuk dikumpulkan bersama dengan kata dasar tersebut.

Masalah-masalah yang telah disenaraikan di atas akan dikaji untuk diselesaikan dengan bantuan sistem komputer. Misalnya di dalam pengumpulan perkataan, dengan bantuan sistem komputer, lebih banyak bahan rujukan (teks) yang dapat digunakan. Ini dapat membantu melengkapkan lagi kamus yang akan dihasilkan. Begitu juga di dalam masalah pencarian teks, proses carian yang lebih mudah dan pantas dapat dilakukan dengan penggunaan sistem komputer. Di dalam proses analisis morfologi, sistem berkomputer dapat membantu dengan mengautomatiskan proses penghuraian perkataan (bagi mendapatkan kata dasar) dan proses pembentukan kata-kata terbitan. Dengan bantuan sistem berkomputer, dijangkakan masa yang diperlukan untuk membangunkan kamus akan dapat disingkatkan lagi berbanding dengan kaedah manual yang digunakan.

¹ Kata dasar adalah kata yang boleh menerima penambahan imbuhan [Asmah, 1986].

1.3 Matlamat Penyelidikan

Secara amnya matlamat kajian ini adalah untuk menghasilkan satu perisian bagi memudahkan ahli bahasa di dalam menjalankan penyelidikan dan pembinaan kamus bahasa Malaysia. Matlamat kajian ini secara khususnya adalah seperti berikut:

- (i) mendapatkan algoritma bagi proses analisis morfologi perkataan bahasa Malaysia untuk dilaksanakan secara berkomputer.
- (ii) membangunkan satu perisian yang dapat menyediakan kepada pengguna kemudahan untuk melaksanakan proses analisis morfologi, penentuan perkataan, capaian teks daripada pangkalan data teks bebas dan penyuntingan teks.

1.4 Struktur Tesis

Tesis ini dibahagikan kepada enam bab. Bab I membincangkan latar belakang masalah pembinaan kamus serta matlamat penyelidikan dan pembangunan *Sistem Bantuan Pembinaan Kamus*. Bab II pula membincangkan kajian tentang morfologi, perkataan dan pembinaan kamus bahasa Malaysia. Bab III membincangkan tentang teks bebas dan *Sistem Pangkalan Data Teks Bebas* yang merupakan inti (kernel) kepada sistem yang dibangunkan.

Seterusnya Bab IV mengemukakan pendekatan dan kaedah-kaedah yang digunakan untuk mereka bentuk *Sistem Bantuan Pembinaan Kamus*. Bab V pula menerangkan secara terperinci reka bentuk *Sistem Bantuan Pembinaan Kamus*, termasuk modul-modul yang membentuk sistem tersebut. Bab VI yang merupakan bahagian terakhir

tesis, merumuskan hasil penyelidikan yang telah dijalankan dan mengemukakan cadangan untuk penyelidikan selanjutnya.

BAB II

MORFOLOGI, PERKATAAN DAN KAMUS

2.1 Pendahuluan

Kamus Dewan [DBP 1989] mentakrifkan kamus sebagai buku yang mengandungi kata-kata dan lain-lain (yang disusun menurut abjad) dengan keterangan mengenai makna perkataan dan sebagainya.

Sejarah perkamusan Melayu bermula dengan kamus dwibahasa iaitu hasil daripada keinginan untuk mengetahui bahasa Melayu oleh orang-orang asing yang datang ke sini. Sejarah pembangunan kamus Melayu bermula kira-kira 500 tahun yang lalu. Kamus pertama adalah dalam bentuk dwibahasa yang terdiri daripada senarai 482 kata entri Melayu-Cina yang telah dikumpulkan di antara tahun 1403 dan 1511 [Asmah 1987].

Selain dari kamus dwibahasa, terdapat beberapa jenis lagi kamus iaitu termasuklah kamus ekabahasa (satu bahasa), kamus pelbagai bahasa (lebih dari dua bahasa), kamus peribahasa dan simpulan bahasa serta kamus istilah teknikal.

Para penyusun kamus mempunyai tanggungjawab yang berat di dalam memperbaiki mutu kamus Melayu memandangkan bahasa ini sedang mengalami perkembangan yang pesat, terutama dengan pemasukan dan penciptaan perkataan-perkataan baru. Ini penting kerana kamus berwibawa mengubah penglihatan dunia (world-view) para penutur bahasa berkenaan [Asmah 1987]. Untuk melaksanakan tugas ini,

penyelidik atau penyusun kamus perlu mempunyai pengetahuan yang mendalam terutama tentang morfologi bagi bahasa yang berkaitan.

2.2 Morfologi Bahasa Malaysia

Morfologi merupakan bidang ilmu bahasa yang mengkaji *struktur*, *bentuk* dan *penggolongan kata*. Dengan *struktur kata* dimaksudkan susunan bentuk bunyi ujaran atau lambang (tulisan) yang menjadi unit bahasa yang bermakna. *Bentuk kata* pula adalah rupa unit tatabahasa sama ada berbentuk tunggal atau hasil daripada proses pengimbuhan, pemajmukan dan penggandaan. *Penggolongan kata* ialah proses menjeniskan perkataan berdasarkan keserupaan bentuk dan/atau fungsi dengan anggota lain dalam golongan yang sama [Safiah 1989].

Unit-unit tatabahasa yang menjadi unsur perkataan disebut morfem. Morfem merupakan unit yang terkecil dalam bahasa yang berfungsi gramatis atau yang menjalankan tugas nahu [Safiah 1989]. Setiap perkataan terbentuk daripada satu atau lebih morfem. Proses pembentukan kata dalam sesuatu bahasa merupakan proses penyusunan morfem menurut peraturan sistem morfologi bahasa tersebut.

Morfem boleh dibahagikan kepada dua iaitu morfem bebas dan morfem terikat. Morfem bebas adalah morfem yang boleh berdiri sendiri sebagai satu kata, misalnya "saya", "itu", "ayam", "rumah" dan lain-lain. Morfem terikat pula merupakan morfem yang memerlukan bantuan morfem lain untuk membentuk kata. Sebagai contoh, "ber" dalam perkataan "berfikir" merupakan morfem terikat yang memerlukan morfem bebas "fikir" untuk membentuk kata terbitan tersebut.

2.2.1 Imbuan

Imbuan ialah unit-unit bahasa tertentu yang ditambahkan pada bentuk-bentuk lain yang menyebabkan perubahan makna nahunya [Safiah 1989]. Jenis-jenis imbuan di dalam bahasa Malaysia adalah seperti berikut:

(i) *Awalan*

Awalan ialah imbuan yang ditambahkan pada bahagian hadapan kata dasar. Misalnya “ber” dalam “berjalan”.

(ii) *Akhiran*

Akhiran merupakan imbuan yang ditambahkan pada bahagian belakang kata dasar. Misalnya “an” dalam “lautan”.

(iii) *Apitan*

Apitan adalah imbuan yang ditambahkan serentak pada bahagian hadapan dan belakang kata dasar. Misalnya “ke-an” dalam “kedudukan”.

(iv) *Sisipan*

Sisipan ialah imbuan yang diselitkan di antara unsur-unsur kata dasar (selepas konsonan pertama). Misalnya “em” dalam “gemuruh”.

2.2.2 Perkataan

Perkataan ialah satu ujaran yang bebas dan mengandungi makna, ianya terdiri daripada satu atau gabungan beberapa bentuk bunyi bahasa [Safiah 1989]. Perkataan mengandungi berbagai-bagai bentuk yang dapat dijeniskan seperti berikut:

(i) *Kata Tunggal*

Kata tunggal merupakan bentuk kata yang terdiri daripada hanya satu bentuk kata dasar, ianya tidak menerima sebarang bentuk imbuhan atau kata dasar yang lain. Misalnya *saya, itu, ini* dll.

(ii) *Kata Terbitan*

Kata terbitan adalah bentuk kata yang mengandungi kata dasar yang mendapat atau telah menerima imbuhan (sama ada awalan, akhiran, apitan atau sisipan). Misalnya *dipukul* (di + pukul), *makanan* (makan + an), *mengakui* (meng + aku + i), *gemuruh* (g + em + uruh) dll.

(iii) *Kata Majmuk*

Kata majmuk ialah bentuk kata yang wujud apabila dua atau lebih kata dasar dirangkaikan menjadi satu kesatuan sintaksis yang membawa makna tertentu. Misalnya *air mata, kerani kanan, kata kunci* dll.

(iv) *Kata Ganda*

Kata ganda merupakan bentuk kata yang dihasilkan dengan menggandakan atau mengulangi kata dasar sama ada secara keseluruhan atau pada bahagian-bahagian tertentu dan dengan imbuhan atau tanpa imbuhan. Misalnya *kanak-kanak, lelaki, bertubi-tubi* dll.

2.3 Pembentukan Kata Terbitan

Proses pembentukan kata terbitan adalah proses pembentukan kata yang terpenting dalam bahasa Malaysia, kerana bahasa Malaysia merupakan bahasa jenis aglutinatif atau bahasa yang bersifat derivatif. Bahasa jenis ini membentuk kebanyakan perkataannya melalui proses pengimbuhan iaitu penggabungan morfem kata dasar dengan morfem imbuhan secara rentetan mendatar [Safiah 1989].

2.3.1 Kata Terbitan Berawalan

Imbuhan awalan dapat wujud dalam kata nama terbitan, kata kerja terbitan dan kata adjektif terbitan. Berikut adalah senarai imbuhan awalan dan contoh penggunaannya: