

**LAPORAN PROJEK PENYELIDIKAN**  
**JANGKA PENDEK**

---

Penyemakan Ejaan Dokumen Bahasa Melayu  
Secara Interaktif Melalui Web

---

Pusat Pengajian Sains Komputer  
Universiti Sains Malaysia  
11800 USM Pulau Pinang Malaysia

---

Januari 2004

## Daftar Isi

Bab 1	Pengenalan.....	3
1.1	Latar Belakang.....	3
1.2	Objektif.....	3
Bab 2	Skema Projek.....	4
2.1	Skop Projek.....	4
2.1.1	Menyediakan Kamus Bahasa Melayu .....	4
2.1.2	Menyediakan Laman Web Sebagai Antara Muka Pengguna .....	5
2.2	Kajian Ketersauran .....	5
2.2.1	Ketersauran Teknikal.....	5
2.2.2	Ketersauran Ekonomi .....	5
2.2.3	Ketersauran Operasi .....	6
2.3	Pengurusan Projek .....	6
Bab 3	Analisis Sistem .....	7
3.1	Sistem Semasa .....	7
3.1.1	Kajian Sistem Semasa .....	7
3.1.2	Aktiviti Sistem Semasa.....	7
3.1.3	Masalah Dalam Sistem Lama .....	7
3.2	Sistem Baru.....	8
3.2.1	Keperluan Sistem Baru .....	8
3.2.2	Spesifikasi Untuk Sistem Baru .....	9
3.2.3	Senario .....	9
3.3	Input, Output Dan Reka Bentuk Antara Muka .....	11
3.3.1	Antara Muka Dan Papan Cerita.....	11
Bab 4	Perlaksanaan .....	15
4.1	Persekitaran .....	15
4.2	Perancangan Ujian.....	15
4.2.1	Ujian Unit .....	16
4.2.2	Ujian Lain.....	16
4.3	Perancangan Latihan.....	17
Bab 5	Perbincangan Dan Penilaian.....	18
5.1	Kelebihan Sistem.....	18
5.2	Masalah Yang Dihadapi .....	18
5.3	Cadangan Penambahbaikan Sistem .....	19
Bab 6	Kesimpulan.....	20
Bab 7	Rujukan.....	21

## **Bab 1 Pengenalan**

### **1.1 Latar Belakang**

Penyemakan Ejaan Dokumen Bahasa Melayu Secara Interaktif Melalui Web ialah aplikasi yang berasaskan Web yang menyemak ejaan dokumen bahasa Melayu. Aplikasi ini dibina berasaskan pakej yang diperolehi dari Sistem Operasi Linux, iaitu Ispell. Aplikasi ini juga dibina di dalam Sistem Operasi, Linux.

Ispell atau “Interactive Spelling” digunakan untuk menyemak ejaan berasaskan kamus yang digunakan. Sekiranya perkataan tidak wujud di dalam kamus, aplikasi ini akan memberi cadangan dan pilihan bagi membolehkan pengguna membuat pembetulan sekiranya perlu. Disebabkan kamus Bahasa Melayu tidak wujud di dalam Sistem Operasi, Linux, satu kamus Bahasa Melayu telah dihasilkan untuk tujuan ini yang berdasarkan kepada Kamus Dewan Bahasa dan Pustaka.

### **1.2 Objektif**

Objektif utama bagi membangunkan aplikasi ini adalah untuk menyediakan penyemakan ejaan bagi Bahasa Melayu. Di antara objektif-objektif lain adalah:

- Mengurangkan beban pengguna bagi proses penyemakan dokumen.
- Menyediakan antara muka pengguna yang berlandaskan laman Web untuk memudahkan penggunaan aplikasi.
- Mengemas kini sistem yang sedia ada supaya aplikasi ini lebih fleksibel dan ramah pengguna.

## Bab 2 Skema Projek

### 2.1 Skop Projek

Bagi membangunkan aplikasi ini, terdapat dua skop utama iaitu menyediakan kamus Bahasa Melayu (yang belum wujud sebelum ini) dan menyediakan laman Web sebagai antara muka pengguna.

#### 2.1.1 Menyediakan Kamus Bahasa Melayu

Bagi menyediakan kamus Bahasa Melayu ini, ianya dibina berdasarkan perisian yang sedia ada di dalam sistem operasi, Linux iaitu Ispell. Perisian ini akan memeriksa setiap perkataan di dalam teks yang diberikan dan membandingkannya dengan kamus yang digunakan bagi menentukan sama ada perkataan tersebut wujud atau tidak. Perisian ini dibangunkan dengan melibatkan dua perkara; fail affix dan fail kamus.

Fail affix adalah fail yang mengandungi flag-flag untuk imbuhan perkataan Bahasa Melayu. Sebagai contoh flag M, ianya adalah untuk imbuhan awalan ME. Awalan ini dibahagikan kepada beberapa imbuhan awalan lain mengikut syarat-syarat yang ditentukan seperti MEN MEM, MENG, MENY. Contoh penggunaan, perkataan TEKAN. Apabila ditambahkan dengan awalan ME, maka perkataan tersebut akan menjadi *MENEKAN* iaitu huruf T disingkirkan.

Fail kamus pula dibina berdasarkan kepada fail affix. Di dalam fail kamus terkandung perkataan-perkataan Bahasa Melayu. Perkataan yang disenaraikan di dalam fail kamus ini terbahagi kepada dua bahagian; perkataan asas dan flag-flag yang mewakili imbuhan awalan atau akhiran dalam Bahasa Melayu. Contoh salah satu perkataan adalah, tekan/MKITAYDP. Sebagai contoh, perkataan *MENEKANKAN*. 'MEN' dihasilkan berdasarkan flag M dan 'KAN' berhasilkkan berdasarkan flag K.

### **2.1.2 Menyediakan Laman Web Sebagai Antara Muka Pengguna**

Penyediaan laman Web sebagai antara muka kepada aplikasi ini adalah untuk memudahkan penggunaan aplikasi. Laman Web akan dibangunkan dengan menggunakan skrip PHP dan skrip Perl. Manakala Apache digunakan sebagai pelayan Web. Perisian ini dipilih kerana ianya tidak melibatkan kos yang tinggi dan bersesuaian dengan Sistem Operasi, Linux. Pengguna dihendaki memasukkan teks di dalam laman Web, yang kemudiannya akan dihantar kepada aplikasi Ispell untuk semakan. Setiap kesilapan yang dikesan akan disertakan dengan pilihan bagi tujuan pembetulan kepada pengguna.

## **2.2 Kajian Ketersauran**

### **2.2.1 Ketersauran Teknikal**

Peranti yang diperlukan untuk aplikasi ini ialah sebuah komputer yang mengandungi sambungan Internet. Ini kerana aplikasi ini dijalankan dalam pelayan Linux.

### **2.2.2 Ketersauran Ekonomi**

Dari aspek ekonomi, sistem ini dapat mengurangkan masa untuk penyemakan ejaan dan mengurangkan kesalahan ejaan. Oleh itu ianya dapat mengurangkan masa dan kos. Sebelum aplikasi ini dibangunkan, penyelidikan telah dilakukan bagi menentukan bahawasanya tiada aplikasi seumpamanya dan bagi menentukan tiada persaingan.

### 2.2.3 Ketersauran Operasi

Dari aspek operasi, aplikasi ini tidak sukar untuk digunakan kerana ianya melibatkan penggunaan aplikasi dan bukan menginput data ke dalam sistem. Pengguna hanya perlu menyediakan dokumen untuk disemak dan sistem secara automatik akan menyemak dan memberi pilihan kepada pengguna untuk tujuan pembetulan. Bagi pengguna tadbir, mereka hanya perlu memeriksa sistem secara tetap dan sistem perlu dikemas kini dari semasa ke semasa.

### 2.3 Pengurusan Projek

ID	Task Name	Start	Finish	Duration
1.	Key in all Malay Root words	15/7/02	29/11/02	20 weeks
2.	Improving the Malay Dictionary	2/12/02	24/1/03	8 weeks
3.	Developing the Format Converter	27/1/03	13/6/03	20 weeks
4.	Change the rules in Affix File	16/6/03	25/7/03	6 weeks
5.	Improving the Web Site & Testing	28/7/03	30/9/03	10 weeks

## **Bab 3 Analisis Sistem**

### **3.1 Sistem Semasa**

#### **3.1.1 Kajian Sistem Semasa**

Dalam sistem semasa, aplikasi ini dapat beroperasi dengan baik, di mana ianya dapat menyemak kesilapan ejaan secara interaktif. Walau bagaimanapun aplikasi ini masih tidak memenuhi keperluan pengguna yang mahu aplikasi yang mudah dan cepat untuk digunakan. Sistem lama juga mempunyai beberapa masalah. Masalah utama aplikasi ini ialah tidak dapat menyemak ejaan dengan baik. Ini kerana perkataan di dalam kamus yang dibangunkan masih tidak lengkap dan terdapat beberapa masalah dalam fail affix.

#### **3.1.2 Masalah Dalam Sistem Lama**

Masalah utama dalam sistem lama adalah ketidakupayaan untuk menyemak sebahagian dari ejaan dengan betul. Puncanya adalah perkataan di dalam fail kamus yang dibangunkan masih kurang menyeluruh dan mempunyai beberapa masalah dalam fail affix. Fail affix ialah fail yang mentakrif semua imbuhan dalam bahasa Melayu.

Sistem semasa hanya dapat menyemak kesalahan ejaan secara interaktif, yang mana pengguna perlu salin(*copy*) bahagian teks daripada fail dokumen dan tampal(*paste*) kepada kotak teks yang telah disediakan untuk menyemak ejaan. Selepas penyemakan ejaan selesai, teks-teks yang telah disemak disalin dan tampal kembali kepada dokumen. Laman Web sistem semasa juga kurang ramah pengguna kerana penyemakan ejaan dilakukan secara interaktif dan bukannya secara automatik.

## 3.2 Sistem Baru

### 3.2.1 Keperluan Sistem Baru

Keperluan untuk membangunkan sistem baru dibahagikan kepada dua bahagian. Bahagian pertama akan menghuraikan keperluan perkakasan dan bahagian kedua menghuraikan keperluan perisian untuk membangunkan sistem baru.

#### I. Keperluan Perkakasan Untuk Sistem Baru:

Bagi Pelayan	Bagi Firewall/Router Linux
<ul style="list-style-type: none"> <li>• Pemproses Pentium III dengan kelajuan 1.7 GHz</li> <li>• 256 MB RDRAM</li> <li>• 30 GB cakera keras</li> <li>• Kad rangkaian 3COM (10/100mbps)</li> <li>• Sambungan Internet</li> <li>• 52X CD-ROM</li> <li>• Monitor, tetikus dan papan kekunci</li> </ul>	<ul style="list-style-type: none"> <li>• Pemproses Pentium III dengan kelajuan 1.4 MHz</li> <li>• 64 MB RDRAM</li> <li>• 10 GB cakera keras</li> <li>• 2 kad rangkaian (10/100 mbps)</li> <li>• 52X CD-ROM</li> <li>• Hub</li> <li>• Monitor, tetikus dan papan kekunci</li> </ul>

#### II. Keperluan Perisian Untuk Sistem Baru:

- a. Keperluan perisian yang minimum untuk pelayan dan "firewall/router" adalah Sistem Operasi, Linux.
- b. Keperluan minimum perisian untuk pembangunan sistem baru adalah
  - Bahasa pengaturcaraan PHP, PERL dan C
  - Pelayan Web Apache
  - Adobe Image Styler

### 3.2.2 Spesifikasi Untuk Sistem Baru

Sistem baru dijangkakan lebih ramah pengguna dan penyemakan ejaan akan menjadi lebih mudah dan fleksibel. Di sini adalah beberapa fungsian untuk sistem baru:

- i. Sistem baru akan membuatkan aplikasi penyemakan ejaan lebih berkesan menyumbang kemudahan penyemakan ejaan bahasa Melayu.
- ii. Masalah dengan fail affix dapat diatasi dengan pembetulan kesalahan peraturan dalam fail tersebut dan memasukkan semua perkataan bahasa Melayu kepada fail kamus.
- iii. Laman Web untuk sistem baru diubah berdasarkan keperluan pengguna yang akan menyokong aplikasi penyemakan ejaan secara interaktif.

### 3.2.3 Senario

#### a) Import Dokumen

Bagi menyemak ejaan, pengguna akan melayari laman Web yang ditentukan. Untuk penyemakan ejaan secara interaktif pengguna perlu salin(*copy*) perkataan dari fail yang hendak disemak dan tampal(*paste*) perkataan kepada ruang penyemakan ejaan aplikasi ini.

#### b) Semak Kesalahan Ejaan

Aplikasi penyemakan ejaan akan menyemak ejaan teks yang telah di salin(*copy*). Setiap perkataan dalam teks akan dibandingkan dengan perkataan yang terdapat di dalam fail kamus. Jika perkataan tersebut wujud dalam kamus, perkataan itu dianggap betul ejaannya, jika sebaliknya perkataan tersebut dianggap yang salah ejaannya. Kemudian pilihan akan diberi supaya pengguna dapat membetulkan kesalahan ejaan. Proses ini akan berterusan sehingga pengguna selesai menyemak kesalahan ejaan.

**c) Di Akhir Semakan**

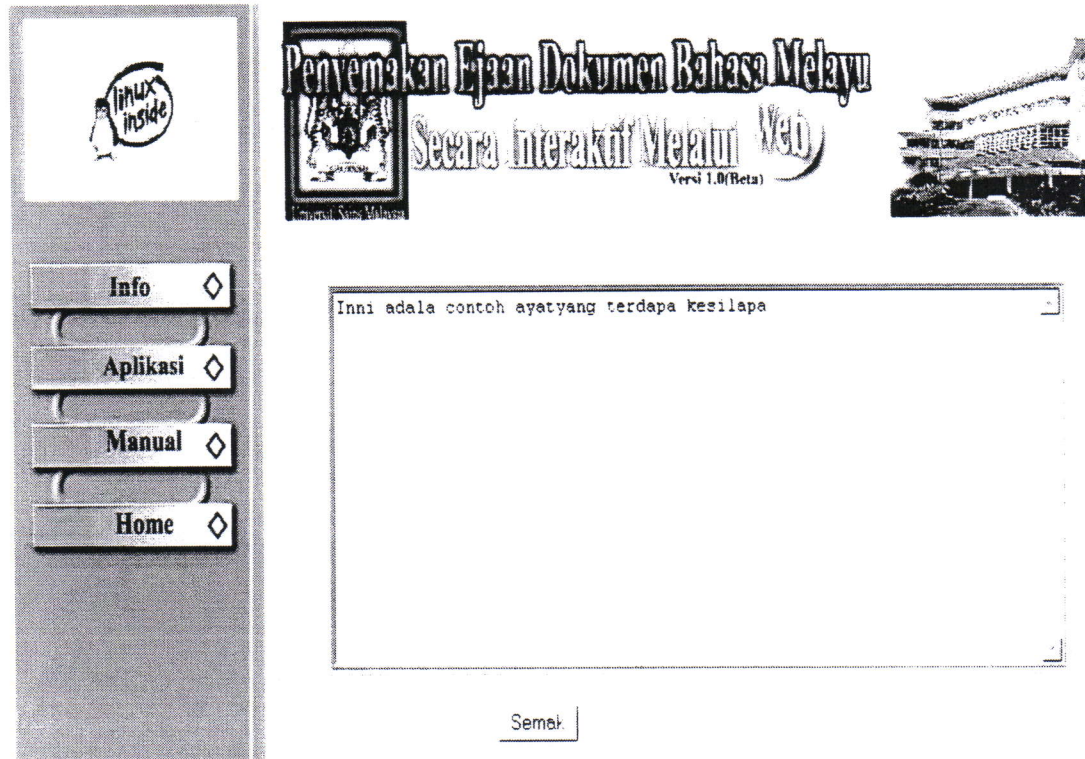
Selepas pengguna telah selesai menyemak ejaan secara interaktif, perkataan yang telah disemak akan di salin(*copy*) dan tampal(*paste*) ke fail asal.

**3.2.3.1 Strategi Reka Bentuk**

Pemilihan perisian untuk projek ini berdasarkan beberapa kelebihan yang dapat memenuhi keperluan untuk membangunkan projek. Di antara kelebihan-kelebihan tersebut:

- a) Pelayan Web, Apache adalah perisian percuma dan bersumber terbuka. Apache juga sokong saluran berbilang dan sokong kepada berbilang sistem operasi lain. Ianya mempunyai ciri-ciri keselamatan dan mampu berfungsi sebagai pelayan proksi HTTP
- b) Bahasa pengaturcaraan yang dipilih seperti PHP, Perl dan C. merupakan bahasa pengaturcaraan peringkat tinggi yang mempunyai kesamaan yang mempermudah proses pembangunan aplikasi ini.

### 3.3 Input, Output Dan Reka Bentuk Antara Muka



Gambar rajah I. Pengguna memasukkan teks dan tekan butang "Semak".

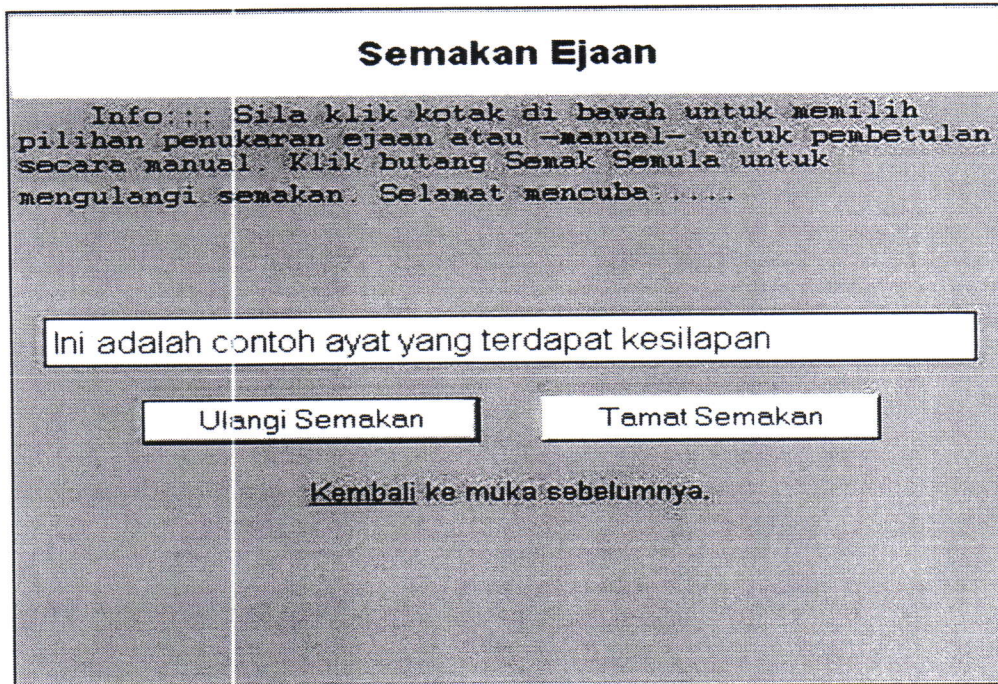
**Semakan Ejaan**

Info::: Sila klik kotak di bawah untuk memilih pilihan penukaran ejaan atau -manual- untuk pembetulan secara manual. Klik butang Semak Semula untuk mengulangi semakan. Selamat mencuba.....

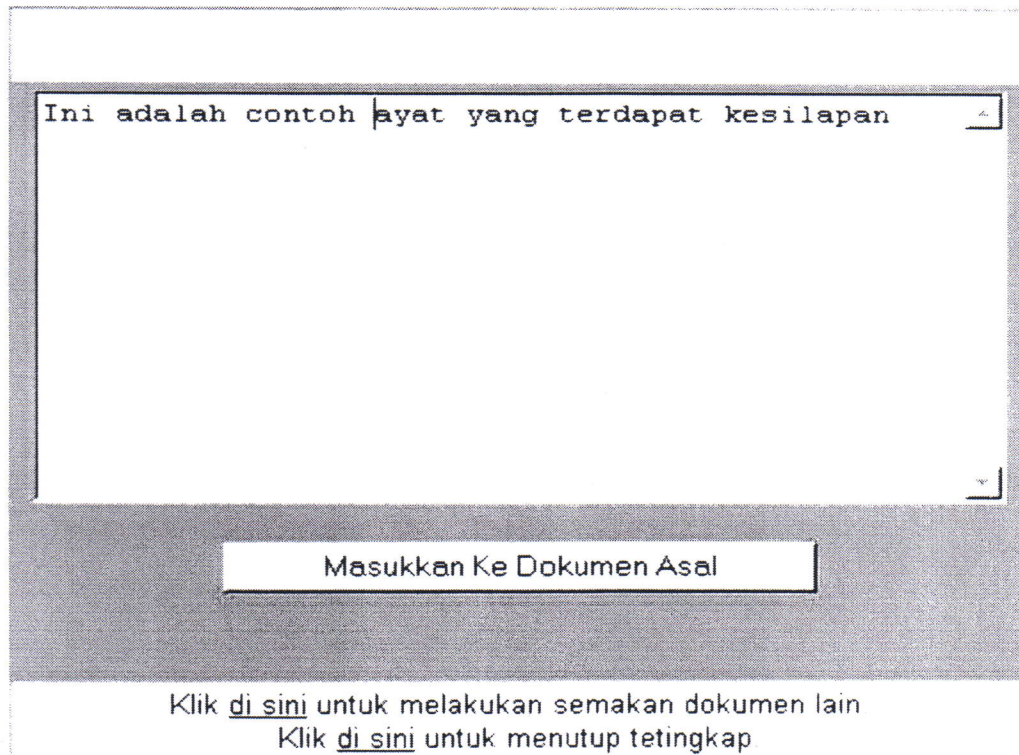
In  adalah  contoh ayat yang

[Kembali ke muka sebelumnya.](#)

Gambar rajah II. Pengguna memilih perkataan yang betul dan menekan butang Ulangi Semakan selepas kesalahan ejaan telah diperbetulkan.



Gambar rajah III. Pengguna menekan butang Tamat Semakan apabila penyemakan ejaan selesai.



Gambar rajah IV. Pengguna menekan butang Masukkan Ke Dokumen Asal untuk kembali kepada Laman utama aplikasi

## **Bab 4      Perlaksanaan**

### **4.1 Persekitaran**

Selaras dengan perkembangan semasa, aplikasi ini dibangunkan dengan menggunakan teknologi Web dan kemudahan Internet.

### **4.2 Perancangan Ujian**

Dalam perancangan ujian ke atas sistem yang telah dibangunkan, terdapat beberapa isu atau kriteria tertentu yang perlu diambil kira sebelum ujian dijalankan. Kriteria-kriteria tersebut adalah:

- Mengenal pasti keperluan sistem sebelum ujian dilakukan dan mengambil kira ciri-ciri lain seperti kesesuaian, kegunaan dan ketahanan sistem.
- Menyatakan objektif ujian sistem dengan jelas seperti lingkungan ujian ke atas sistem, pengesanan kerosakan yang berterusan dan mengenal pasti ujian tersebut efektif ke atas sistem dan sebagainya.
- Melaksanakan strategi penambahbaikan yang berterusan dalam proses ujian sistem.

#### **4.2.1 Ujian Unit**

Sebelum menjalankan ujian unit ke atas sistem, beberapa perkara perlu dipertimbangkan seperti modul antara muka. Modul tersebut diuji untuk memastikan perjalanan maklumat secara dalaman atau luaran seperti yang dirancang. Struktur data juga perlu diperiksa supaya proses pengaturcaraan yang melibatkan data disimpan dalam perhubungan dan lingkungan yang betul sepanjang langkah pelaksanaan dilaksanakan. Selepas itu keadaan kekangan juga diuji untuk memastikan proses setiap modul berjalan dengan betul dalam lingkungan kekangan sama ada mengehadkan dan mengetatkan sebarang proses. Setiap laluan yang tidak berkait dengan struktur kawalan dipraktikkan untuk memastikan semua pernyataan atur cara dalam setiap modul dilaksanakan walaupun hanya sekali. Akhir sekali, kesemua penyemakan kesalahan diuji ke atas sistem.

#### **4.2.2 Ujian Lain**

Ujian-ujian lain yang dilakukan ke atas sistem adalah yang berkaitan dengan ujian pemulihan, ujian keselamatan dan ujian prestasi. Ujian pemulihan dilaksanakan dengan menyediakan "*back-up*" dan memastikan "*back-up*" ini dapat dijalankan sama seperti sistem sebenar, jika terdapat sebarang masalah yang terjadi kepada sistem. Ujian prestasi dilakukan dengan menguji kepantasan sistem memproses data dan keberkesanan proses yang dilakukan supaya lebih efektif.

### **4.3 Perancangan Latihan**

Bagi memastikan sistem dapat digunakan oleh semua pengguna, setiap komputer mestilah berada di dalam persekitaran Intranet. Pengguna hanya perlu mempunyai pelayar seperti Internet Explorer dan Netscape Navigator untuk melayari laman Web.

Bagi tujuan latihan, sebuah manual yang mudah difahami dihasilkan bagi membantu pengguna menggunakan sistem ini. Memandangkan sistem yang dibina adalah mesra pengguna maka diharapkan pengguna mampu menggunakannya berpandukan manual tanpa sebarang latihan praktikal.

## **Bab 5 Perbincangan Dan Penilaian**

### **5.1 Kelebihan Sistem**

- Sistem yang dibangunkan ini memberi banyak kelebihan dan faedah kepada pengguna berbanding sistem manual yang masih dilaksanakan sehingga kini kerana sistem baru yang dilaksanakan atas pembaharuan sistem prototaip juga lebih kemas dan teratur dan pengintegrasian yang dilakukan lebih memudahkan pengguna.
- Kamus yang diperbaiki dapat meliputi lebih banyak perkataan dalam Bahasa Melayu dan terkini. Hasilnya penyemakan dapat dilakukan dengan lebih menyeluruh, efektif dan tepat.
- Pembangunan laman Web sebagai antara muka pengguna, banyak membantu mempermudah proses menyemakkan dan juga memperluaskan penggunaan aplikasi di kalangan pengguna kerana ianya mudah dan mesra pengguna.

### **5.2 Masalah Yang Dihadapi**

Walaupun telah banyak pembaruan dibuat di dalam aplikasi ini masih terdapat beberapa perkara yang tidak dapat diselesaikan. Masalah utama berkaitan dengan struktur kamus yang digunakan atau dibangunkan dan juga bahasa itu sendiri. Lebih banyak penyelidikan perlukan dijalankan bagi memperbaiki masalah tersebut. Di antara masalah tersebut adalah:

- Data kamus yang dibangunkan perlu dikemas kini sekiranya fail affix diubah. Ini merumitkan kerana data kamus melibatkan data yang besar.

- Ada beberapa perkataan tidak dapat dimasukkan ke dalam kamus data kerana ada ruang kosong di antara perkataan tersebut.
- Wujudnya masalah dengan fail *melayu.aff*. Terdapat masalah yang tidak dapat diselesaikan kerana penyelesaiannya akan menimbulkan masalah pada yang lain. Contohnya:

*#awalan ME*

*flag M:* > *MEM*

*F* > *memfokus*

*-F. MEM (memikirkan)*

Flag M bagi huruf pertama perkataan F dengan F tidak boleh letak dua peraturan pada masa yang sama kerana konflik akan berlaku.

### 5.3 Cadangan Penambahbaikan Sistem

Berikut disenaraikan beberapa cadangan yang mampu memperbaiki sistem:

- Kamus data haruslah dikemas kini dari semasa ke semasa kerana kata dasar Melayu terlalu luas.
- Mengubahsuai fail *melayu.aff* supaya membolehkan kamus menerima perkataan yang berulang.
- Membiarkan satu peraturan di dalam fail flag dan menaip seluruh perkataan bagi peraturan lain di dalam kamus tersebut.

## **Bab 6      Kesimpulan**

Penyemakan Ejaan Dokumen Bahasa Melayu Secara Interaktif Melalui Web diwujudkan bagi memudahkan pengguna untuk menyemak dokumen Bahasa Melayu. Memandangkan pada ketika ini amat sukar untuk mendapatkan satu aplikasi yang sedemikian rupa, maka satu alternatif baru haruslah disediakan. Aplikasi yang dibangunkan akan dikemas kini dari semasa ke semasa. Sehingga kini aplikasi ini telah banyak dikemas kini dan diubahsuai agar aplikasi ini menjadi lebih fleksibel dan mudah digunakan.

Walaupun aplikasi ini berjalan dengan baik, namun ada kemungkinan terdapat beberapa masalah yang tidak disedari. Oleh yang demikian aplikasi ini perlu dikemas kini dari semasa ke semasa. Komen dan cadangan daripada pengguna amat penting bagi menilai sejauh mana keberkesanan dan kemampuan aplikasi ke atas mereka. Diharap aplikasi ini mampu memberi manfaat kepada pengguna.

## Bab 7 Rujukan

- [1] Satzanger, Jackson and Burd. *System Analysis and Design In A Changing World*. Thomson Learning, 2000.
- [2] Satzanger & Orvik. *The Object Approach-Concept, System Development, and Modeling with UML*. 2<sup>nd</sup> Edition. Thomson Learning, 2001.
- [3] Dr Abdullah Embong, *Sistem Pangkalan Data: Konsep Asas, Reka Bentuk dan Pelaksanaan, Tradisi Ilmu*, 2000.
- [4] Stig Sæther Bakken, Alexander Aulbach, Egon Schmid, Jim Winstead, Lars Torben Wilson, Rasmus Lerdorf, Andrei Zmievski, Jouni Ahto. *PHP – PHP Manual*. <http://www.php.net/manual/en/>
- [5] *Perl Monks – The Monastery Gates*. <http://perlmonks.org>
- [6] *Perldoc.com – Perl Documentation*. <http://www.perldoc.com>