

## Model Pengubahan Aksara Hijaiyah Menjadi Aksara Latin dengan Metode Pengeindekan

Hari Murti, Nuk Ghurroh Setyoningrum dan Agung Prihandono

Fakultas Teknologi Informasi, Universitas Stikubank

email : hmurti076@gmail.com, nuke\_s@yahoo.com, agunganak@gmail.com

### Abstrak

Salah satu persoalan dalam bidang komputer yang banyak menjadi perhatian adalah pengubahan bentuk aksara ( *convert font* ). Pengubahan aksara dalam suatu kata atau kalimat digunakan untuk memudahkan dalam pembacaan yang mudah dipahami oleh user yang tidak mengenali *font* tertentu. Pengubahan aksara ini membahas tentang algoritma – algoritma dan teknik – teknik yang diterapkan pada pengalihan aksarannya untuk menghasilkan deskripsi yang dikenali oleh komputer. Pengalihan aksara ini muncul karena kebutuhan efisiensi dalam menganalisis sebuah teks kata atau kalimat yang diinput melalui *textbox* yang disediakan. Dalam input suatu aksara dengan menggunakan *font* tertentu dan akan diubah ke dalam bentuk *font* yang lain membutuhkan suatu cara untuk mempermudah pengubahannya. Maka dari itu guna memudahkan dalam pengalihan aksarannya menggunakan suatu cara, salah satu caranya adalah dengan menggunakan kode ASCII yang telah ada. Penginputan aksara melalui *keyboard* dari *font* tertentu akan dilakukan dengan melakukan pencarian kode ASCII yang melalui tabel ASCII yang tersedia, sehingga pengalihan aksara dapat dilakukan. Kemudian pengalihan aksara tersebut bisa digunakan untuk memudahkan dalam pembacaannya.

**Kata kunci :** Pengubah aksara, Translate

### PENDAHULUAN

Salah satu persoalan dalam bidang komputer yang banyak menjadi perhatian adalah pengubahan bentuk aksara ( *convert font* ). Pengubahan aksara dalam suatu kata atau kalimat digunakan untuk memudahkan dalam pembacaan yang mudah dipahami oleh user yang tidak mengenali *font* tertentu. Pengubahan aksara ini membahas tentang algoritma – algoritma dan teknik – teknik yang diterapkan pada pengalihan aksarannya untuk menghasilkan deskripsi yang dikenali oleh komputer. Pengalihan aksara ini muncul karena kebutuhan efisiensi dalam menganalisis sebuah teks kata atau kalimat yang diinput melalui *textbox* yang disediakan. Dalam input suatu aksara dengan menggunakan *font* tertentu dan akan diubah ke dalam bentuk *font* yang lain membutuhkan suatu cara untuk mempermudah pengubahannya. Maka dari itu guna memudahkan dalam pengalihan aksarannya menggunakan suatu cara, salah satu caranya adalah dengan menggunakan kode ASCII yang telah ada. Penginputan aksara

melalui *keyboard* dari *font* tertentu akan dilakukan dengan melakukan pencarian kode ASCII yang melalui tabel ASCII yang tersedia, sehingga pengalihan aksara dapat dilakukan. Kemudian pengalihan aksara tersebut bisa digunakan untuk memudahkan dalam pembacaannya.

### TINJAUAN PUSTAKA

Penelitian dan tulisan tentang pengalihan aksara telah banyak dilakukan oleh para peneliti. Kevin K. Dan Jonathan G. Memperkenalkan mesin alih aksara yang bisa mengalih aksara dari nama – nama inggris ke aksara Jepang dan suaranya bacanya. Dalam penelitian ini mempergunakan WFS (Weight Finite State Acceptor) dan WFSTs (Weight Finite State Transducers).

Sebuah WFS adalah diagram status / transisi dengan bobot dan symbol dalam transisi – transisi, membuat beberapa barisan output yang serupa sebuah WFSTs adalah WFS dengan sebuah pasangan symbol dalam

transisinya, satu input satu output. (Kevin, 1989).

Aksara arab mempunyai ligetur yaitu dibuat dengan cara menggabungkan paling sedikit dua aksara yang berbeda (Srouji, 1991). Contoh adalah ligatur yang terdiri dari aksara lam dan alif.

Track – 2001 track CLIR memfokuskan dalam pencarian dokumen arabic mempergunakan query – query English, Prancis, dan Arabic (Fedric, 2001). Untuk pencarian dokumen antar bahasa ini Treck – 2001 (Text Retrieval Conference) mempunyai modul alih aksara English, Prancis ke Arabic.

Suprihatin (Suprihatin, 2003), dalam tesisnya telah berhasil mengalihaksarakan tulisan Jawa ke tulisan karakter lain dengan masukan program alihaksara diperoleh dari *keyboard* dan mempergunakan program sofya untuk mendapatkan jenis font karakter.

Sedangkan Anastasia Rita Widiarti (Anastasia Rita Widiarti, 2005), pada tesisnya berhasil mengalihaksarakan tulisan karakter jawa ke tulisan karakter lain, namun masukan untuk program alihaksara diperoleh dari *scan* dokumen dengan mempergunakan program citra dokumen.

Perkembangan internet yang sangat pesat dan kapasistas dokumen online yang besar menjadikan banyak riset yang membahas tentang sistem temu kembali informasi (information retrieval). Yue, dkk (2007) dalam penelitiannya mengusulkan algoritma sistem temu kembali informasi (*information retrieval*) berbasis *query expansion* dan klasifikasi. Algoritma tersebut diinduksi dari *query* yang pendek dan metode pencarian informasi tradisional (*traditional retrieval information method*) yang menghasilkan presisi yang rendah walaupun tingkat recall cukup tinggi. Penelitian ini berusaha untuk mendapatkan lebih banyak dokumen yang relevan dengan mengekspand *query* (*query expansion*) dan klasifikasi dokumen. Hasil dari penelitian dengan algoritma *query expansion* dan klasifikasi dokumen menghasilkan ketepatan dan efisiensi pencarian dokumen dibandingkan dengan metode *query expand* tradisional.

Philipp (2003), mengusulkan penerjemahan frase berbasis model dan algoritma decoding yang memungkinkan untuk mengevaluasi membandingkan beberapa penerjemahan frase berbasis model. Dari eksperimen yang dilakukan menjelaskan mengapa model frase berbasis mengalahkan kata berbasis model. Secara empiris, hasil yang diperoleh menunjukkan bahwa tingkat kinerja tertinggi diperoleh secara relatif sederhana yang berarti belajar heuristik terjemahan frase dari kata berbasis berpihak dan leksikal terjemahan frase pembobotan. Namun, belajar frase lebih dari tiga kata dan belajar frase lebih dari tinggi accuracywordlevel model alignment tidak berdampak kuat pada kinerja.

Mitamura (1999), menyajikan ikhtisar-ikhtisar penyebab dalam mengontrol perancangan bahasa yang implementasi bahasa dikendalikan checker dan penyebaran KANT secara terkendali. Bahasa Inggris untuk mesin penerjemahan multibahasa membahas beberapa kriteria keberhasilan untuk memperkenalkan bahasa yang dikendalikan.

Habash (2007), menjelaskan pendekatan untuk sumber sintaksis bahasa otomatis *preprocessing* dalam konteks Arab-Inggris frase berbasis mesin penerjemahan. Aturan diekstrasi digunakan untuk menyusun ulang sisi bahasa sumber-pelatihan dan data uji.

Pemrosesan, dokumen dipilah menjadi unit-unit yang lebih kecil misalnya berupa kata, frasa atau kalimat. Unit pemrosesan tersebut disebut sebagai token. Parsing merujuk pada proses pengenalan token yang terdapat dalam rangkaian teks (Grossman, 2002). Oleh karena itu bagian dasar dalam parsing adalah algoritma pengambil token dari teks yang disebut tokenizer. Proses ini memerlukan pengetahuan bahasa untuk menangani karakter-karakter khusus, serta menentukan batasan satuan unit dalam dokumen.

Prahallad, dkk (2005), transliterasi editor sangat penting untuk memasukkan-dalam skrip bahasa India ke dalam komputer menggunakan keyboard QWERTY. Aplikasi editor transliterasi dalam konteks Universal Digital Library (UDL) termasuk masuknya meta-data dan kamus untuk bahasa india. suatu pendekatan yang sederhana

untuk membangun editor transliterasi untuk bahasa India menggunakan Unicode dan dengan mengambil keuntungan dari mesin rendering nya. menunjukkan kegunaan dari pendekatan berbasis Unicode editor membangun transliterasi untuk bahasa India, dan melaporkan keuntungan membutuhkan sedikit pemeliharaan dan beberapa entri dalam pemetaan meja, dan kemudahan penambahan fitur baru seperti menambahkan surat, untuk skema transliterasi. Kami menunjukkan transliterasi itu editor untuk 9 bahasa India dan juga menjelaskan bagaimana pendekatan ini dapat diadaptasi untuk skrip Arab.

Abas (1998), hambatan utama dalam pengajaran bahasa Arab sebagai bahasa asing adalah adanya teknologi yang tersedia untuk kedua sumber daya guru dan siswa. Langkah pertama dalam mengatasi ketersediaan ini dengan menyediakan rata-rata pengguna komputer kemampuan untuk mengakses Arab di Internet dan untuk memproses kata dalam Bahasa Arab dan sistem transliterasi Arab, terlepas dari komputer mereka saat ini sistem operasi dan versi perangkat lunak. Kemampuan untuk proses kata dalam bahasa Arab dan transliterasi Arab hanya sebagai salah satu yang akan di Inggris kini tersedia dalam teknologi universal seperti Microsoft Word, sehingga kebutuhan untuk obviating membeli prosesor Arab kata yang berdiri sendiri atau memanfaatkan program transliterasi Arab berbeda atau font.

Fedric (2001), mesin transliterasi adalah untuk menuliskan sebuah kata-kata tertulis dalam naskah dengan perkiraan fonetis kesetaraan dalam bahasa lain. Hal ini berguna untuk terjemahan mesin, lintas-bahasa sistem temu kembali informasi, multibahasa teks dan pengolahan pidato. Punjabi Mesin Transliterasi (PMT) adalah kasus khusus dari mesin transliterasi dan merupakan proses mengubah kata dari Shahmukhi (berdasarkan Arab script) untuk Gurmukhi (turunan dari Landa, Sharda dan Takri, skrip lama India benua), dua skrip dari Punjabi, terlepas tentang jenis kata. Mesin Transliterasi Punjabi Sistem menggunakan aturan transliterasi (karakter pemetaan dan aturan dependensi) untuk transliterasi kata-kata Shahmukhi ke Gurmukhi. Sistem PMT dapat

mengalihbahasakan setiap kata ditulis dalam Shahmukhi.

Suprihatin (2003), dalam tesisnya telah berhasil mengalihaksarakan tulisan Jawa ke tulisan karakter lain dengan masukan program alihaksara diperoleh dari *keyboard* dan mempergunakan program *sofy* untuk mendapatkan jenis font karakter.

Widiarti (2005), pada tesisnya berhasil mengalihaksarakan tulisan karakter jawa ke tulisan karakter lain, namun masukan untuk program alihaksara diperoleh dari *scan* dokumen dengan mempergunakan program *citra* dokumen.

## LANDASAN TEORI

### Pengetahuan tentang Penerjemahan (*Translation*)

Pimjai (2007) dalam publikasinya menjelaskan bahwa pengetahuan penerjemahan adalah sebuah konsep yang kompleks dan multidimensi yang menuntut pemahaman yang komprehensif tentang mekanisme, metode, dan pengukuran serta faktor-faktor yang mempengaruhi individu, tingkat kontekstual dan interaksi antara kedua level tersebut.

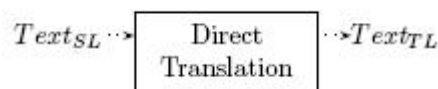
Penerjemahan adalah mengambil sebuah teks di salah satu bahasa (*the source language*) dan menghasilkan sebuah "equivalent" dalam bahasa yang lain (*the target language*). Doug Arnold (1984) memberikan pengertian tentang mesin terjemah sebagai upaya untuk mengotomatisasi semua atau bagian dari kegiatan terjemahan.

Arsitektur terjemahan menurut Doug Arnold (2008) dibagi menjadi dua yaitu klasik dan baru.

#### 1. Arsitektur terjemah secara klasik.

##### a. Direct

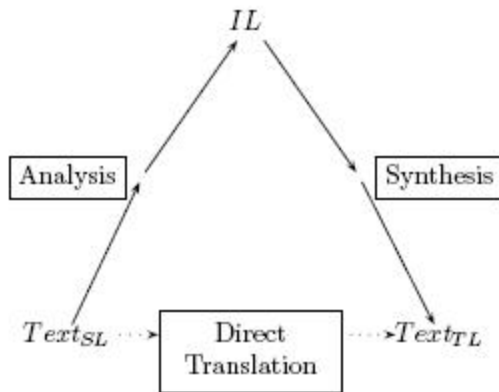
Yaitu penerjemahan yang dilakukan secara langsung dari satu bahasa ke bahasa yang lain. Terlihat pada Gambar 1.



Gambar 1. Direct Tranlation

b. Interlingual

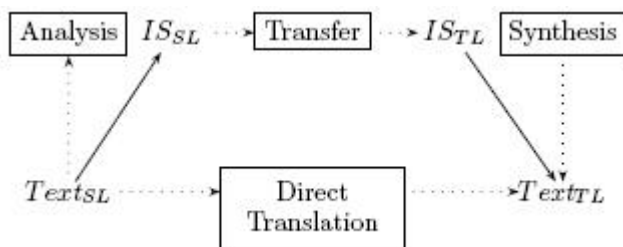
Yaitu penerjemahan yang dilakukan dengan menganalisa teks yang akan diterjemahkan, kemudian melakukan sintesis terhadap teks tersebut. Terlihat pada Gambar 2.



Gambar 2. Interlingual Translation

c. Transfer

Yaitu penerjemahan yang dilakukan dengan menganalisa teks yang akan diterjemahkan kemudian melakukan penyesuaian aturan antar bahasa yang kemudian melakukan sintesis terhadap teks tersebut.



Gambar 3. Transfer Translation

2. Arsitektur terjemah secara baru.

a. Pendekatan Analogi

Pendekatan analogi dapat dilakukan secara statistik dan contoh dasar. Pendekatan secara statistik dikomunikasikan dengan 'noisy chanel' yaitu untuk mengubah suatu kalimat  $f$  ke dalam bahasa tertentu dengan menggunakan rumus,  $(22) \hat{e} = \text{argmax } e\text{Pr}(e) \times \text{Pr}(f|e)$ . Sedangkan pendekatan

secara contoh dasar dilakukan dengan menggunakan database. Sebagai contoh dapat dilihat pada Tabel 1.

Tabel 1. Contoh pendekatan analogi secara contoh dasar

Julie bought a notebook	Julie compró una libreta
Ann read a book on economics	Ann leyó un libro de económica

b. Pendekatan dengan dasar pembatas.

Penerjemahan yang dilakukan dimana terbatas dengan peraturan yang ada dalam bahasa yang diterjemahkan.

3. Transliterasi aksara ke Latin

Rahman (2007), Transliterasi dimaksudkan sebagai pengalihhurufan dari abjad yang satu ke abjad yang lain. Transliterasi Arab-Latin di sini ialah penyalinan huruf-huruf Arab dengan huruf-huruf Latin beserta perangkatnya.

Pebakuan pedoman transliterasi Arab-Latin ini disusun dengan prinsip sebagai berikut:

1. Sejalan dengan Ejaan yang Disempurnakan
2. Huruf Arab yang belum ada padanannya dalam huruf Latin dicarikan padanan dengan cara memberi tambahan tanda diakritik, dengan dasar "satu fonem satu lambang".
3. Pedoman transliterasi ini diperuntukkan bagi masyarakat umum.

Hal-hal yang dirumuskan secara konkrit dalam pedoman transliterasi Arab-Latin ini meliputi:

1. Konsonan
2. Vokal (tunggal dan rangkap)
3. Maddah
4. Ta'marbutah
5. Syaddah
6. Kata sandang (di depan huruf syamsiah dan qamariah)
7. Hamzah

- 8. Penulisan kata
- 9. Huruf Kapital
- 10. Tajwid

**4. Font Arab**

Kalimat dalam huruf arab pada dasarnya merupakan string / deretan simbol huruf arab, maka untuk itu dalam penulisan kata dalam bahasa arab dapat terbentuk apabila beberapa huruf arab digabungkan menjadi satu agar menjadi sebuah rangkaian kata yang dapat dibaca menurut Bawani (1987).

Huruf arab ini dapat digunakan dengan mengambil pada Multilanguage yang ada pada sistem operasi windows. Untuk mengambil language Arabic ini dengan membuka “regional and language options” pada control panel, pilih details, pilih add, kemudian pilih language yang dikehendaki. Jika sudah, maka language dapat digunakan untuk menuliskan huruf arab.

**a. Cara Menulis Huruf Arab**

Menurut Susiyanto (2002), penulisan huruf arab dapat mengikuti kaidah – kaidah berikut :

- 1. Penulisan huruf arab dimulai dari kanan ke kiri.
- 2. Jumlah huruf arab atau huruf hijaiyah ada 28 huruf.
- 3. Huruf-huruf itu ada yang dapat menyambung dan disambung, ada yang bisa disambung tetapi tidak bisa menyambung. Masing-masing mempunyai bentuk huruf sesuai posisinya (di depan, tengah, belakang atau terpisah). Di antara huruf-huruf itu terdapat beberapa huruf yang dapat disambung dan menyambung dan beberapa huruf yang hanya dapat disambung.
- 4. Semua huruf Arab adalah konsonan, termasuk alif, wawu dan ya (sering disebut huruf *illat*), maka mereka memerlukan tanda vokal (sakal).

**b. Huruf Arab ( Huruf Hijaiyah ) dan Cara Penulisan**

Susiyanto (2002), dalam penulisan huruf Arab perlu menguasai huruf-huruf Arab berikut

bunyinya. Berikut ini adalah Tabel 2. nama huruf Arab beserta letaknya dan bunyinya.

Tabel 2. Tabel Nama Huruf Arab Beserta Letak dan Bunyi

contoh	Di akhir	Di tengah	Di awal	Berdiri sendiri	Bunyi	Nama
ا ا ا	ا	ا	ا	ا	-	Alif
ب ب ب	ب	ب	ب	ب	b	Ba
ت ت ت	ت	ت	ت	ت	t	Ta
ث ث ث	ث	ث	ث	ث	ts	Tsa
ج ج ج	ج	ج	ج	ج	j	Jim
ح ح ح	ح	ح	ح	ح	ch	Cha
خ خ خ	خ	خ	خ	خ	kh	Kho
د د د	د	د	د	د	d	Dal
ذ ذ ذ	ذ	ذ	ذ	ذ	dz	Dzal
ر ر ر	ر	ر	ر	ر	r	Ra
ز ز ز	ز	ز	ز	ز	z	Za
س س س	س	س	س	س	s	Sin
ش ش ش	ش	ش	ش	ش	sv	Svin
ص ص ص	ص	ص	ص	ص	sh	Shod
ض ض ض	ض	ض	ض	ض	dh	Dhad
ط ط ط	ط	ط	ط	ط	th	Tha
ظ ظ ظ	ظ	ظ	ظ	ظ	zh	Zho
ع ع ع	ع	ع	ع	ع	'	'ain
غ غ غ	غ	غ	غ	غ	gh	Ghain
ف ف ف	ف	ف	ف	ف	f	Fa
ق ق ق	ق	ق	ق	ق	q	Qaf
ك ك ك	ك	ك	ك	ك	k	Kaf
ل ل ل	ل	ل	ل	ل	l	Lam
م م م	م	م	م	م	m	Min
ن ن ن	ن	ن	ن	ن	n	Nun
و و و	و	و	و	و	w	Wawu
ه ه ه	ه	ه	ه	ه	h	Ha
	لا	لا	لا	لا	l	Lam alif
	ء	ء	ء	ء	'	Hamzah
ي ي ي	ي	ي	ي	ي	y	Ya

**c. Cara Menyambung Huruf Arab**

Susiyanto (2002), Di antara 28 huruf hijaiyah terdapat huruf yang dapat disambung dan menyambung, ada yang dapat disambung tetapi tak dapat menyambung. Di bawah ini adalah huruf-huruf yang dapat disambung tetapi tak dapat menyambung.

ا د ذ ر ز و

Selain enam huruf di atas, semua huruf dapat menyambung dan disambung. Huruf hijaiyah yang dapat disambung terlihat pada Tabel 3.

Tabel 3. Contoh Penyambungan Huruf Arab

sambung	pisah	Sambung	pisah
حَمَلٌ	ح م ل	اِخْتٌ	ا خ ذ
بَيْن	ب ي ن	سَأَل	س أ ل
قَام	ق ا م	رَزَق	ر ز ق
نَصَح	ن ص ح	جَوِب	ج و ب
سَهْل	س ه ل	وَاحِد	و ا ح د
كَبَس	ك ب س	سُرُو	س ر و

**d. Konversi font**

Pada dasarnya, komputer hanya berurusan dengan angka. Mereka menyimpan huruf dan karakter lain dengan menetapkan nomor untuk masing-masing karakter. Sebelum Unicode ditemukan, ada ratusan sistem pengkodean berbeda untuk menetapkan angka-angka ini. Tidak ada encoding bisa berisi karakter tunggal cukup: misalnya, Uni Eropa sendiri membutuhkan pengkodean yang berbeda untuk menutup semua bahasa nya. Bahkan untuk satu bahasa seperti bahasa Inggris tidak ada pengkodean tunggal adalah cukup untuk semua huruf, tanda baca, dan simbol-simbol teknis umum digunakan (<http://www.unicode.org/standard/WhatIsUnicode.html>).

Sistem-sistem encoding juga bertentangan satu sama lain. Artinya, dua pengkodean dapat menggunakan nomor yang sama untuk dua karakter yang berbeda, atau menggunakan nomor yang berbeda untuk karakter yang sama. Setiap komputer yang diberikan (terutama server) harus mendukung pengkodean yang berbeda, namun setiap kali data yang melewati antara pengkodean berbeda atau platform data selalu menjalankan resiko penyalahgunaan data.

Unicode menyediakan nomor unik untuk setiap karakter, tidak peduli platform, program, atau bahasanya. *The Unicode Standard* telah disesuaikan oleh para pemimpin industri seperti Apple, HP, IBM, mail dengan, Microsoft, Oracle, SAP, Sun, Sybase, Unisys dan banyak lainnya. Unicode diperlukan oleh standar modern seperti XML, Java, ECMAScript (JavaScript), LDAP, CORBA 3.0, WML, dll, dan merupakan cara resmi untuk menerapkan ISO / IEC 10646. Hal ini didukung dalam banyak sistem operasi, semua browser modern, dan banyak produk lainnya. Munculnya *Standard Unicode*, dan ketersediaan alat pendukung itu, adalah salah satu tren paling

signifikan pada teknologi perangkat lunak secara global.

Unicode adalah sistem pengkodean standar untuk komputer untuk menampilkan teks dan simbol dari semua sistem tulisan di seluruh dunia. Unicode dikoordinasi oleh Konsorsium Unicode. Ada beberapa Unicode pengkodean: yang paling populer adalah UTF-8, contoh lain adalah UTF-7 dan UTF-16. UTF-8 menggunakan pengkodean karakter variabel-panjang, dan semua kode karakter dasar Latin identik dengan ASCII. Pada sistem windows teks editor yang mendukung Unicode adalah ultraedit (File → Conversions → 'ASCII to UTF-8' or 'ASCII to Unicode (16-Bit)'). Namun dalam penulisan karakter arab ini perubahan ASCII ke Unicode menggunakan 8-bit.

Melalui penginstalan karakter arab pada sistem windows menghasilkan inputan string dalam huruf arab, bisa mempergunakan papan ketik, adapun Gambar papan ketik ditunjukkan pada Gambar 4.

Q ض	W ص	E ث	R ق	T ف	Y غ	U ع	I ه	O خ	P ح	 <	 >
A ش	S س	D ي	F ب	G ل	H أ	J ت	K ن	L م	;	‘	Ente r
Z ذ	X ظ	C ف	V ر	B لا	N أ	M ة	,	.	/	shif t	
Alt	~	Space						In s	Alt		

Gambar 4. Penulisan Huruf Arab dan Harokat Pada Papan Ketik

Penyimpanan karakter arab ke dalam suatu data, sebenarnya bukan karakter yang dimaksudkan melainkan hasil dari encoding dari karakter satu ke karakter yang lain. Dalam hal ini perubahan karakter arab menjadi karakter latin melalui encoding dari *unicode*. Sehingga nilai *unicode* yang dihasilkan terlihat pada tabel 4.

Tabel 4. Tabel Huruf Arab dan Nilai Ascii

Huruf Arab	Unicode
ا	199
ب	200
ك	228
ت	203
ج	204
ح	205
خ	206
د	207
ذ	208
ر	209
ز	210
س	211
ش	212
ص	213
ض	214
ط	216
ظ	217
ع	218
ف	219
ق	221
ك	222
ل	223
م	225
ن	227
و	228
ه	230
و	201
لا	225, 199
ء	193
ي	237
ـ	243
ـ	240
ـ	246
ـ	242
ـ	245
ـ	241
ـ	250
ـ	248

### 1. Analisa Sistem

Analisis sistem adalah proses untuk menterjemahkan suatu sistem yang utuh ke dalam bagian-bagian / komponen-komponen untuk mengidentifikasi dan mengevaluasi kebutuhan, permasalahan dan hambatan yang mungkin terjadi sehingga dapat diadakan perbaikan. Analisis sistem merupakan hal yang perlu diperhatikan dalam pembuatan sistem. Tahap analisis dapat membantu untuk mengetahui kebutuhan dan kemampuan apa yang diperlukan pada sistem tersebut. Hasil analisis sistem berupa rancangan sistem yang nantinya akan dibuat.

Dari analisis yang dilakukan untuk pengindeksan aksara yang akan dibuat, diharapkan mempunyai kemampuan sebagai berikut :

- a. Sistem dapat membaca text dalam bentuk tulisan arab yang diinput melalui papan ketik atau *keyboard*.
- b. Sistem yang dibuat diharapkan dapat menyimpan data hasil indexing dalam bentuk struktur inverted file.
- c. Sistem yang dibuat harus memiliki kemampuan untuk melakukan proses query yang diinputkan oleh user (pemakai).
- d. Sistem harus bisa melakukan pengubahan bentuk huruf dari huruf arab ke dalam huruf latin.

### 2. Kebutuhan Sistem

Sebuah pembelajaran bahasa yang baik disamping bisa memberikan kemudahan bagi pengguna juga harus bisa memenuhi kebutuhan sistem serta menyediakan informasi yang relevan bagi user.

#### a. Bahan Penelitian

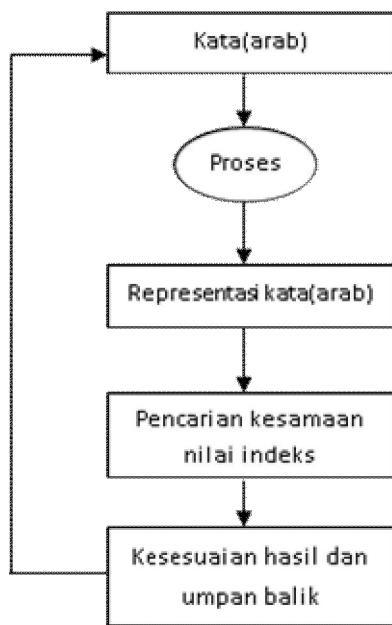
Penelitian ini membutuhkan bahan berupa text dalam bahasa arab. Text yang digunakan adalah text dalam bentuk font Arabic yang dimaksudkan untuk mempermudah pengolahan. Sedangkan pengambilan text arab bersumber dari ayat-ayat yang terkandung dalam Al-Qur'an terutama pada juz pertama sebagai contoh penyimpanan text. Untuk penginputan text dilakukan dengan menggunakan papan ketik atau *keyboard*, sehingga dibutuhkan sistem

## ANALISA DAN PERANCANGAN SISTEM

operasi yang mendukung penulisan text menggunakan bahasa arab.

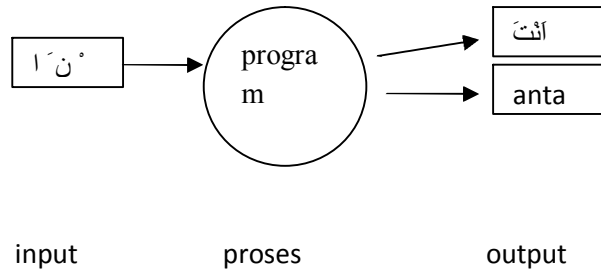
**b. Cara Kerja Perangkat Lunak**

Secara garis besar proses pengindeksan dilakukan dengan memanfaatkan nilai ascii yang digunakan untuk mengenali kombinasi dari huruf hijaiyah dan harokat yang melekat pada huruf hijaiyah tersebut. Untuk urutan dari proses tersebut dapat dilihat pada Gambar 5 yang menjelaskan tentang cara kerja dari sistem transliteration.



Gambar 5. Proses Pengindeksan Karakter

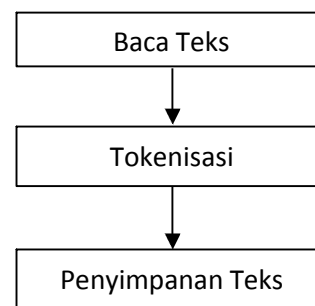
Penginputan data dapat dilakukan dengan memasukkan huruf arab melalui *keyboard*. Proses penginputan diawali dari huruf hijaiyah yang kemudian diikuti oleh harokat yang melekat pada huruf tersebut. Huruf akan menyambung sebelum penggunaan spasi untuk memisahkan penulisan kata, kecuali huruf – huruf tertentu yang memang tidak dapat disambung pada awal penulisan. Penginputan huruf akan diproses oleh program yaitu proses pengubahan bentuk huruf yang disebut juga dengan proses *convert* dan proses penerjemahan dari inputan kata atau proses *translate*. Dari pemrosesan yang dilakukan oleh program akan menghasilkan pengeluaran berupa penulisan kata dalam bentuk huruf arab, pembacaan kata dalam bentuk huruf latin dan arti dari inputan kata. Hal ini dapat ditunjukkan pada Gambar 6.



Gambar 6. Aliran Proses Bekerjanya Program

**3. Pengindeksan**

Proses pengindeksan bertujuan untuk mengenali kombinasi huruf arab dan harokatnya yang kemudian disimpan beserta artinya. Dalam proses pengenalan kombinasi huruf arab dan harokatnya dilakukan penyimpanan dalam bentuk array dengan cara memisah huruf per huruf dan harokat per harokat sebagai string yang dikonversikan kedalam bentuk ascii. Bentuk array yang digunakan adalah array satu dimensi dengan panjang sebanyak nilai ascii pada kata yang akan disimpan. Sebagai contoh bila satu kata terdiri dari tiga huruf sehingga panjang array yang diperlukan adalah dua kali lipat dari jumlah huruf, hal ini berkaitan dengan jumlah harokat yang melekat pada huruf. Urutan proses pengindeksan dapat dilihat pada Gambar 7.



Gambar 7. Alur Pengindeksan Teks Arab

**a. Baca teks**

Dalam tahap pembacaan teks ini yang diinputkan berupa huruf hijaiyah dan harokat. Proses pembacaan teks, ketika ditekan sebuah tombol pada papan ketik, sistem akan membaca karakter sesuai dengan tombol yang telah ditekan kedalam jenis font Arabic.

**b. Tokenisasi**



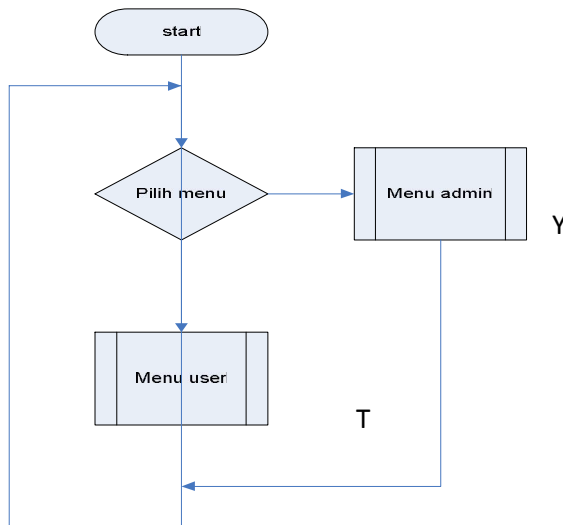
Tokenisasi digunakan untuk memisahkan text menjadi unit – unit kecil berupa huruf ataupun kalimat. Pengertian huruf ataupun kalimat dalam bahasa arab ini memiliki arti sama dengan huruf ataupun kata dalam bahasa indonesia. Token dilakukan berdasarkan penggunaan spasi.

**c. Penyimpan teks**

Proses penyimpanan teks dilakukan dengan merekam semua karakter ascii yang diinputkan melalui papan ketik kedalam table indexing. Setiap teks yang masuk di simpan dengan indeks secara otomatis bertambah nilai indeksnya atau *auto increment*.

**4. Flowchart Pengindeksan Karakter**

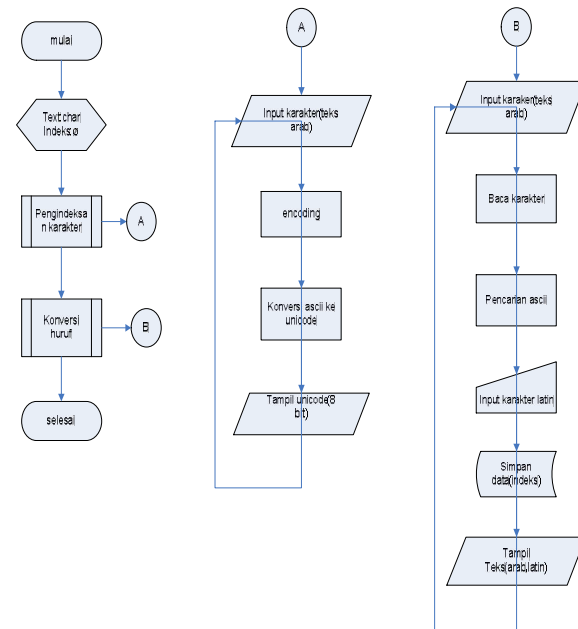
Proses pengindeksan teks arab akan melalui beberapa tahapan yaitu proses konversi tekt arab dan pengindeksan teks arab seperti yang telah ditunjukkan pada gambar 8, dimana Seorang user akan memilih proses yang telah disediakan yaitu *convert*. Pada pilihan *convert* maka user akan masuk ke dalam proses pengubahan huruf arab menjadi huruf latin. Dari proses tersebut, maka pemilihan dibagi menjadi dua menu yaitu menu untuk user admin dan user pengguna biasa. Dapat terlihat dari Gambar 8.



Gambar 8. Skema Sistem Penerjemahan Teks Arab

Pemilihan menu pada sistem dibagi menjadi dua yaitu menu yang akan digunakan oleh admin untuk pengolahan data dan menu yang digunakan oleh pengguna atau user yang

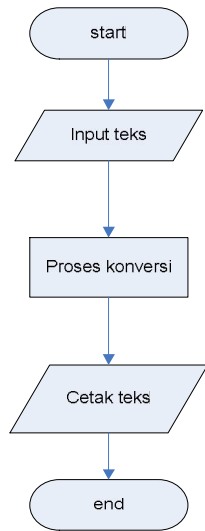
membutuhkan sistem pengindeksan ini. Pada menu admin, seorang user admin akan memilih proses konversi karakter, pengaturan pengaturan penindeksan karakter. Dapat terlihat pada Gambar 9. Skema menu yang dilakukan oleh user admin.



Gambar 9. Skema Menu User Admin

**5. Konversi Huruf**

Konversi huruf merupakan proses pengubahan bentuk huruf dari suatu inputan sehingga menghasilkan keluaran dengan bentuk huruf yang berbeda. Dalam proses konversi huruf ini mengubah bentuk huruf arab menjadi huruf latin dengan memanfaatkan kode ascii. Kode ascii pada huruf arab yang dihasilkan disamakan dengan kode ascii pada huruf latin sehingga pada proses penyimpanan data yang tersimpan adalah karakter ascii pada nilai latin. Setelah proses penyimpanan selesai maka data yang ditampilkan berupa teks dalam huruf latin. Konversi huruf ini dilakukan dengan pengkodean nomor pada karakter, dikarenakan pengkodean setiap karakter pada setiap platform berbeda sehingga dibutuhkan perubahan pada nomor kode tiap karakter.

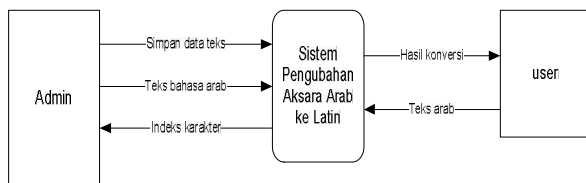


Gambar 10. Skema Pengubahan Pada User Biasa

### 6. Diagram Alir Data

Diagram alir data (Data Flow Diagram) adalah representasi grafis dari aliran data yang melewati sistem. Perancangan dengan Diagram Alir Data dimulai dengan adanya diagram konteks (context diagram). Dengan diagram konteks akan terlihat siapa yang berinteraksi dengan sistem dan apa respon sistem terhadap interaksi tersebut. Dalam penelitian ini diagram alir data dirancang konteks diagram dan diagram level 1.

Seperti pada Gambar 11, terlihat bahwa *user* dibagi menjadi dua yaitu user admin dan user biasa, admin dapat memasukkan teks arab ke dalam sistem. Teks arab itulah yang akan diproses oleh sistem hingga menghasilkan teks dalam bentuk latin beserta arti. Sedangkan user biasa dapat menuliskan suatu teks arab yang kemudian diproses oleh sistem hingga user mendapatkan pembacaan karakter.

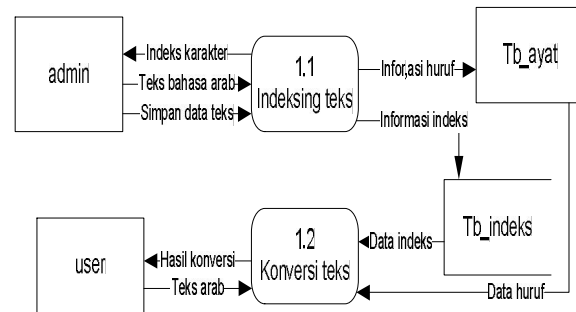


Gambar 11. Diagram Context Sistem Pengindeksan Arab

Gambar 12, memperlihatkan diagram alir dokumen level 1 sistem pengindeksan Arab terdiri dari proses indexing karakter, konversi teks. Admin yang telah memasukkan teks bahasa

ke dalam sistem akan diolah dengan mengambil tiap karakter yang telah dimasukkan. Karakter tersebut akan dibaca oleh sistem sebagai nilai kode karakter dan disimpan menjadi indeks karakter. Sehingga ketika seorang user akan menggunakan sistem, yang terbaca adalah nilai indeks karakter yang tersimpan. Indeks karakter akan digunakan dalam pengubahan teks sehingga menghasilkan konversi teks. Konversi teks yang dihasilkan akan digunakan dalam proses pembacaan karakter yang disimpan menjadi tabel *tb\_indeks*.

Proses indexing karakter dapat lebih dijelaskan menjadi proses konversi karakter dan simpan arti. Konversi karakter dilakukan dengan melakukan *encoding* terhadap tiap karakter yang kemudian melakukan pengubahan dari kode ASCII menjadi *Unicode* dengan menggunakan 1 byte yaitu 8 bit. Diagram alir data level 2 tentang proses indexing karakter, ditunjukkan pada Gambar 12.

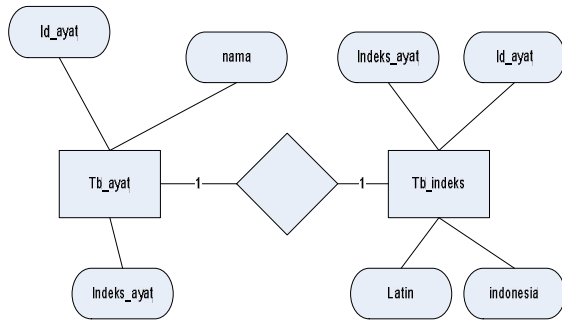


Gambar 12. Skema DFD Level 2 Tentang Penjelasan Proses Indexing Karakter

### 7. Perancangan Database

Perancangan database ini kami menggunakan alat bantu berupa entity relationship diagram (ERD) dan rancangan antar tabel. Entity Relationship Diagram (ERD) digunakan untuk memudahkan struktur data dan hubungan antar data, karena hal ini relatif kompleks. Dengan Entity Relationship Diagram (ERD) dapat dilakukan pengujian model dengan mengabaikan proses yang harus dilakukan.

Dalam rancangan basis data pada Proses penerjemahan yang dikembangkan dalam penelitian ini Entity Relationship Diagram (ERD) atau Diagram Hubungan Entitas ditunjukkan pada Gambar 13.



Gambar 13. Entity Relationship Diagram(ERD)

Gambar 12 Diagram alir data level 2 menghasilkan dua tabel untuk dapat digunakan pada saat pembuatan aplikasi sistem penerjemahan bahasa Arab, yaitu tabel ayat dan tabel Indeks. Tabel ayat digunakan untuk menyimpan ayat-ayat yang dihasilkan setelah melakukan pengindeksan karakter dan indeks terhadap per kata dari ayat.

Tabel 5 : Struktur Tabel Ayat

Field name	Type	Size	Key
Id_ayat	Alpha	2	*
Nama	Alpha	20	
Indeks_ayat	Short	3	

Pada tabel 6 tabel indeks digunakan untuk menyimpan transaksi pengindeksan teks, yang tersimpan dalam indeks yaitu penyimpanan dari tiap karakter yang dimasukkan.

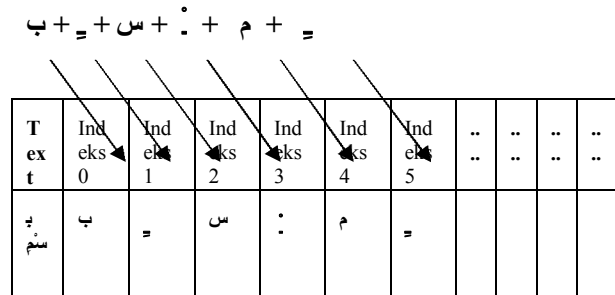
Tabel 6 : Struktur Tabel Indeks

Field name	Type	Size	Key
Id_index	Alpha	5	*
Id_ayat	Alpha	2	*
Indeks_kar	Alpha	20	
Latin	Alpha	15	
Indonesia	Alpha	50	

Penyimpanan text dalam database akan dipisah – pisah sebanyak inputan yang dimasukkan. Sebagai contoh text بِسْمِ akan diproses untuk penyimpanan huruf dan harokat dalam array.

بِسْمِ = ب + س + م +

kata tersebut terdiri dari 6 item yg diinput maka akan disimpan dalam array sebanyak 6 field dalam database.



Gambar 14. Proses Penyimpanan Text Arabic

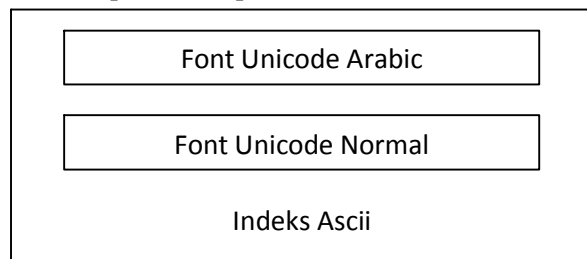
Gambar 14. Bentuk array yang digunakan adalah array satu dimensi dengan panjang sebanyak nilai ascii pada kata yang akan disimpan. Penyimpanan huruf dalam array hanya digunakan untuk prosesnya saja, kemudian yang disimpan kedalam database adalah nilai ascii yang telah dikonversi dari array tersebut. Jadi satu inputan text merupakan satu array, kalau text ada dua huruf berarti nantinya ada dua indeks (0 dan 1, array 1 kali 2), kalau hurufnya ada tiga maka array 1 kali 3 yaitu array indeks pertama(0), array indeks 2(1), array indeks 3(2), dan seterusnya.

### 8. Perancangan Antarmuka

Rancangan antar muka digunakan sebagai media interaksi antara user dengan sistem yang akan dikembangkan.

#### a. Rancangan Antarmuka untuk Indeksing

Dalam melakukan pengindeksan, admin akan memilih di mana letak text tersimpan, sehingga dapat digunakan pada saat pencarian data, dapat dilihat pada Gambar 15.



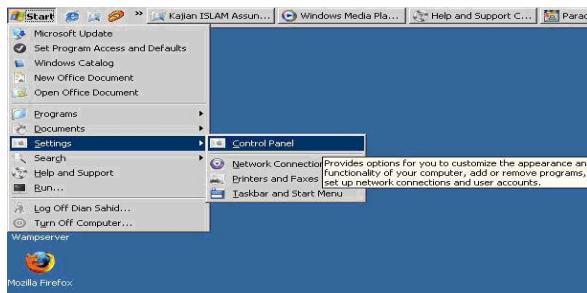
Gambar 15. Rancangan Antarmuka Pengindeksan

## IMPLEMENTASI SISTEM

### 1. Penginstalan Bahasa Arab Pada Windows XP

Sebelum memulai implementasi sistem multimedia text dengan menggunakan bahasa arab, terlebih dahulu melakukan instalasi bahasa arab pada perangkat computer. Langkah – langkah penginstalan dapat dilakukan sebagai berikut:

- a. siapkan CD instalasi windows XP, kemudian Klik “START”, “Settings”, “Control Panel”



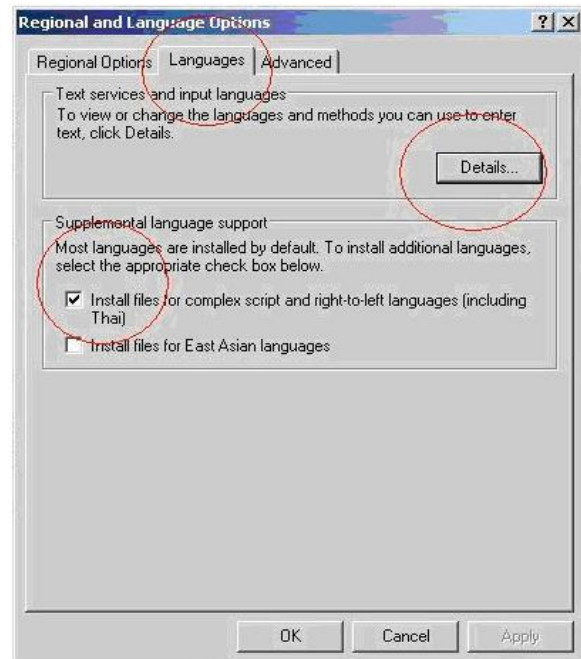
Gambar 16. Setting Instalasi Bahasa Arab

- b. Kemudian Double Klik “Regional and Language Options”



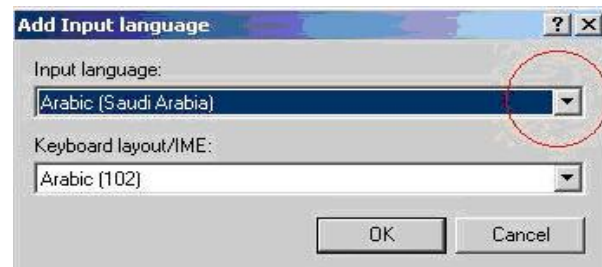
Gambar 17. Regional and Language Options pada Windows XP

- c. Pilih tab “Languages”, berikan tanda centang pada “Install Files for Complex Script and ....” dan klik Apply, masukkan CD instalasi windows XP jika pada langkah ini si Windows minta CD installer, hindari jika windows minta Reboot karena akan memperlama proses setting.



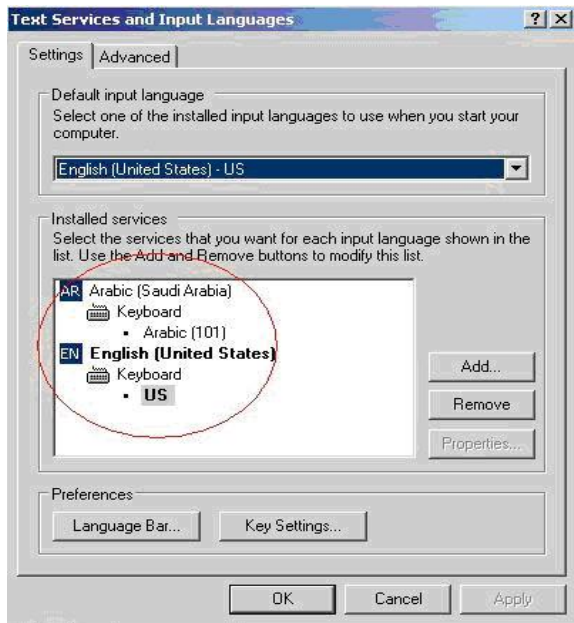
Gambar 18. Setting language Arabic

- d. Klik “Details...” dan pada jendela berikutnya klik “Add...” . Pilih “Input Language” dengan “Arabic (Saudi Arabia)” dan klik OK.



Gambar 19. Input Language Arabic

- e. Munculnya jenis keyboard pada jendela “Installed Service” menandakan jenis bahasa yang disupport oleh keyboard untuk pengetikan, klik OK hingga akhir dan bila perlu restart windows.



Gambar 20. Penginstalan Bahasa Arab Support dengan Keyboard

## 2. Implementasi Penulisan Huruf arab dan Harokat

Perpindahan pengetikan dari bahasa Inggris (latin) ke bahasa arab dan sebaliknya dengan penekanan tombol Alt + Shift, ditandai dengan icon “Language Bar” pada Taskbar yang berubah-ubah. Pemberian harokat pada bahasa arab dengan menekan Shift + ... , misalnya ingin memberi harokat dhomah, yaitu dengan menekan Shift + E.

Tampilan keyboard visual pada monitor, fungsi dari tool ini adalah sebagai visual dari keyboard.



Gambar 21. Tampilan Keyboard Visual dengan font latin

Perubahan ke font arab, tekan tombol alt + shift, maka tampilan akan berubah seperti pada gambar 22.



Gambar 22. Tampilan Keyboard Visual dengan font arab

Untuk mengetik harakat fathah, kasrah, fathatain dan seterusnya silakan tekan tombol shift, maka akan terlihat tanda tersebut.



Gambar 23. Tampilan Keyboard Visual dengan penekanan tombol shift

## 3. Implementasi Sistem

Pada tahap implementasi ini akan dilakukan pengujian terhadap sistem yang dibuat apakah benar-benar sudah memenuhi *requirement* dalam analisis & perancangan sistem atau tidak.

Spesifikasi hardware dan perangkat lunak yang digunakan untuk merancang dan membuat sistem ini menggunakan processor intel core 2 duo T5600 @1,83 Hz, memori 1 GHz. Sistem operasi menggunakan Windows XP, penyimpanan data menggunakan paradox, user interface dan bahasa pemrograman yang digunakan Borland Delphi.

## 4. Batasan Implementasi

Perancangan untuk sistem yang dibuat ini sebenarnya ada 2 tingkatan pengguna yaitu end user dan administrator. End user hanya memiliki hak untuk melakukan membaca konversi teks terhadap koleksi data dokumen yang dimiliki oleh sistem informasi. Sedangkan administrator





belum dapat diterjemahkan secara sempurna (terjemahan terbalik tata bahasanya).

## HASIL PENELITIAN DAN PEMBAHASAN

### 1. Pengujian

Setelah tahap implementasi dilakukan pada pembahasan sebelumnya, maka diperlukan pengujian atau testing dan evaluasi terhadap sistem yang dibuat.

Dalam proses penindeksan dilakukan dengan mengetik melalui media *keyboard* sesuai teks box yang telah disediakan.

### 2. Pembahasan Penindeksan Arab

Pada Gambar 26 dalam uji coba ini menggunakan penyimpanan kata yang dibuat seperti kamus ayat Al-qur'an sebanyak record yang diambilkan dari ayat-ayat Al-Qur'an yang disimpan dalam tabel indexing.



Gambar 26. Contoh Penjelasan Dari Pengeindeksan Teks Arab

Indeks yang dilakukan pada saat kata teks arab dimasukkan:

11	12	13	14	15	16	17	18
199	243	225	250	205	243	227	2
225	246	225	248	229	246	0	
209	243	200	246	248	0	0	
199	225	250	218	243	225	243	2
199	243	225	209	248	243	205	2
199	225	209	248	243	205	246	
227	243	225	246	223	246	0	
237	243	230	250	227	246	0	

Gambar 27. Contoh Penjelasan Dari Pengeindeksan Teks Arab

Kata “مَلِك” terdapat 6 karakter yang telah dimasukkan yaitu “م + ل + ك + . + +” masing-masing memiliki nilai ascii secara berurutan 227, 243, 225, 246, 223, 246” sehingga pada saat penerjemahan yang akan diproses pencarian adalah nilai ascii yang disimpan dalam indeks. Teks yang diindeks dalam tabel indexing akan disimpan dengan indeks yang bertambah secara sendirinya atau *auto increment* yang dapat dilihat dari Gambar 6.6. sehingga nilai indeks

yang tersedia akan semakin bertambah sesuai jumlah input yang dimasukkan dalam teks arab.

## PENUTUP

### 1. Kesimpulan

Dari hasil penelitian yang telah dilakukan dapat disimpulkan hal-hal sebagai berikut:

- Aplikasi mampu membantu pengguna dengan tidak melakukan penerjemahan keyword yang diinputkan pada saat melakukan pencarian ayat yang berbeda bahasa.
- Aplikasi mampu menampilkan hasil konversi ascii dari bentuk arab ke bentuk latin, yang nantinya akan digunakan pada saat konversi text.
- Sistem yang dikembangkan mampu untuk menyimpan representasi text dalam bentuk inverted file yang menyimpan term-term.

### 2. Saran

Keterbatasan pengetahuan dan waktu yang tersedia penulis menyadari bahwa masih banyak terdapat kekurangan dalam sistem ini terutama metode penerjemahan dan pencarian ayat Al-qur'an yang digunakan dan harus dibenahi diantaranya adalah :

- Aplikasi nantinya dalam penerjemahan text masih berdasarkan pemisahan kata dengan menggunakan spasi, sehingga perlu dikembangkan lagi dengan menggunakan pencocokkan dari setiap huruf yang dimasukkan.
- Aplikasi penerjemahan dan pencarian text dilakukan dengan menggunakan ascii sebagai konversi text dari bentuk arab ke bentuk latin, sehingga perlu dikembangkan dengan menggunakan Unicode yang lebih umum oleh pengguna dengan berbagai macam bahasa.
- Bagi peneliti lain yang berniat mengembangkan sistem penerjemahan dan pencarian ayat Al-qur'an disarankan menggunakan metode yang diperluas sehingga menghasilkan terjemahan dan pencarian yang lebih baik.
- Pengembangan pencarian dari ayat Al-qur'an ini dapat dikembangkan menjadi

pencarian dokumen surat-surat yang terkandung dalam Al-qur'an.

- e. Database yang digunakan hendaknya, database yang lebih fleksibel sehingga dapat memudahkan dalam pengambilan data.

#### DAFTAR PUSTAKA

- Abbas, M, (1998), "*Punjabi Machine Transliteration*", [www.aclweb.org/anthology/P/P06/P06-1143.pdf](http://www.aclweb.org/anthology/P/P06/P06-1143.pdf). (Akses terakhir tanggal 20 April 2010).
- Arnorld, D, 1984, "*Machine Translation*", [http://privatewww.essex.ac.uk/~doug/index\\_5.html](http://privatewww.essex.ac.uk/~doug/index_5.html). (Akses terakhir tanggal 20 April 2010).
- Fedric, C. G., 2001, "*The Trec – 2001 Cross language Information Retrieval Track: Searching Arabic Using English, French or Querries*", <http://terpconnect.umd.edu/~oard/pdf/trec01ov.pdf>. (Akses terakhir tanggal 23 Mei 2009).
- Grossman, D. *IR Book*. [http://www.ir.iit.edu/~dagr/cs529/files/ir\\_book](http://www.ir.iit.edu/~dagr/cs529/files/ir_book). (Akses tanggal 2 September 2009)
- Habash, N, 2007, "Syntactic Preprocessing for Statistical Machine Translation", [www.mt-archive.info/MTS-2007-Habash-1.pdf](http://www.mt-archive.info/MTS-2007-Habash-1.pdf). (Akses terakhir tanggal 20 April 2010)
- Imaduddin, 2005, "*Tata Bahasa Arab Sistematis*", Yogyakarta: Nuansa Aksara Group.
- Kevin, K and Jonathan G., 1998, "*Machine Transliteration*", <http://acl.ldc.upenn.edu/J/J98/J98-4003.pdf>. (Akses terakhir tanggal 23 mei 2009).
- Mandhany, N, A., 2006, "*MULTILINGUAL COMPUTING WITH ARABIC AND ARABIC TRANSLITERATION Arabicizing Windows Applications to Read and Write Arabic & Solutions for the Transliteration Quagmire Faced by Arabic- Script Languages*" [http://www.lib.uchicago.edu/e/su/mideast/Multilingual\\_Computing\\_with\\_Arabic\\_and\\_Arabic\\_Transliteration.pdf](http://www.lib.uchicago.edu/e/su/mideast/Multilingual_Computing_with_Arabic_and_Arabic_Transliteration.pdf). (Akses terakhir tanggal 8 Juni 2009).
- Mitamura, T, 1999, "*Controller Language for Multilingual Machine Translation*", [www.lti.cs.cmu.edu/Research/Kant/PDF/MTSummit99.pdf](http://www.lti.cs.cmu.edu/Research/Kant/PDF/MTSummit99.pdf). (akses terakhir tanggal 20 April 2010)
- Philip, K , 2003, "*Statistikal Phrase-Based Translation*", <http://www.isi.edu/~marcu/papers/phrases-hlt2003.pdf>. (akses terakhir tanggal 20 April 2010).
- Pimjai, S, (2007), "*Knowledge Translation: Introduction to models, Strategies, and Measures*". [www.naric.com/public/reSearch/ReSearchVol3no1.pdf](http://www.naric.com/public/reSearch/ReSearchVol3no1.pdf). (akses terakhir tanggal 20 april 2010).
- Prahallad, L, (2005), "*A simple approach for building transliteration editors for Indian languages*", [www.ulib.org/conference/2005/17.pdf](http://www.ulib.org/conference/2005/17.pdf). (akses terakhir tanggal 20 april 2010).
- Rahman, S.A. 2007. *Teach Yourself Arabic: a Modern and Step by Step Approach*. New Delhi: Goodword Book.
- Salton, G. (1968), *A Comparison Between Manual and Automatic Indexing Methods*. Technical Report No. 68-11. Department of Computer Science. Cornell University, Ithaca, N.Y.
- Salton, G. (1989), *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley.
- Srouji, J. and Berry. D, 1993, "*Arabic with Ditroff / ffortid*", [http://se.uwaterloo.ca/~dberry/FTP\\_SITE/reprints.journals.conferences/keshide.journal.paper.pdf](http://se.uwaterloo.ca/~dberry/FTP_SITE/reprints.journals.conferences/keshide.journal.paper.pdf). (akses terakhir tanggal 23 Mei 2009).
- Suprihatin, 2003, "*Penerapan Finite Automata Dalam Mengalihaksarakan Tulisan Aksara Jawa ke Tulisan Huruf latin. Tesis S2 Ilmu Komputer*", Universitas Gadjah Mada, Yogyakarta.



- Susiyanto, Azib. 2002. *Kitabiah: Pedoman menulis Huruf Arab dan Al Quran Sistem 5 Jam*. Yogyakarta: Pustaka Kitabiah.
- Widiarti, R, A., 2005, ” *Analisis Citra Dokumen: Konsep Dan Implementasinya ( Studi Kasus pada Buku Sastra Jawa” Menak Sorangan”)*. Tesis S2 Ilmu Komputer”, Universitas Gadjah Mada, Yogyakarta.
- Yue, W., 2005, *Using Query Expansion and Classification for Information Retrieval*, College of Computer and Communication, Hunan University ChangSha, Hunan Province, 410082,China.