

---

UNIVERSITI SAINS MALAYSIA

Peperiksaan Semester Pertama  
Sidang Akademik 2002/2003

September 2002

**CCS503 – Pemprosesan Dokumen Cerdas**

**CSC514 – Pemprosesan Bahasa Tabii**

Masa : 3 jam

---

**ARAHAN KEPADA CALON:**

- Sila pastikan bahawa kertas peperiksaan ini mengandungi **EMPAT** soalan di dalam **ENAM** muka surat yang bercetak sebelum anda memulakan peperiksaan ini.
  - Jawab **SEMUA** soalan.
  - Peperiksaan ini akan dijalankan secara 'Open Book'.
  - Anda boleh memilih untuk menjawab semua soalan dalam Bahasa Malaysia atau Bahasa Inggeris.
-

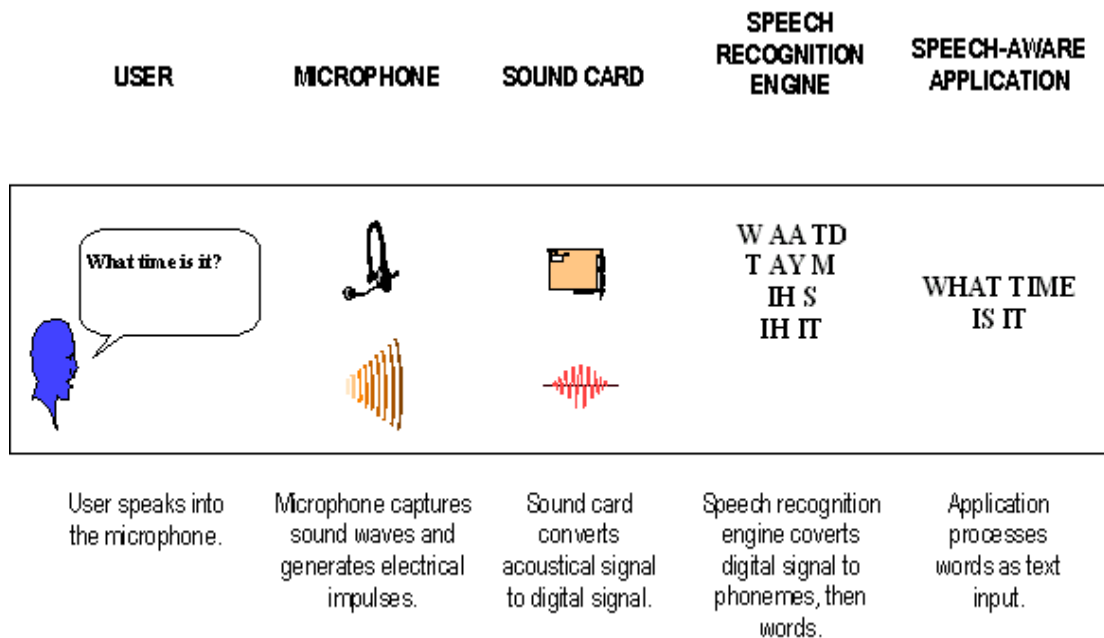
1. Diberikan ayat berikut:

"What time is it?"

- (a) Berapakah fonem yang wujud dalam ayat di atas ? Tuliskan sebutan ayat di atas dalam simbol IPA.

[25/100]

- (b) Pengecaman pertuturan (speech recognition) atau pertuturan ke teks (speech-to-text) melibatkan proses menangkap dan mendigitkan gelombang bunyi, menukarkan mereka ke unit bahasa asas atau fonem, menjanakan perkataan daripada fonem, dan menganalisa perkataan berasaskan konteks untuk memastikan ejaan yang tepat untuk perkataan yang serupa cara sebutannya (misalnya 'write' dan 'right'). Gambarajah berikut memberikan gambaran dan deskripsi secara umum tentang proses-proses tersebut.



Bincangkan secara menyeluruh bagaimana pelbagai alat NLP yang diperkenalkan dalam kursus ini dapat digunakan untuk meningkatkan keberkesanan 'speech recognition engine' seperti yang disebutkan di atas.

[50/100]

- (c) Dalam Bahasa Inggeris, cara umum untuk menjanakan bentuk 'plural' perkataan jenis kata nama ialah menambahkan 's' di belakang perkataan tersebut. 's' ini akan disebut secara berlainan (i.e. /z/ or /s/) pada konteks yang berbeza. Misalnya, bentuk sebutan 'plural' untuk '*bid*' ialah [bɪdz] dan bentuk sebutan 'plural' untuk '*bit*' ialah [bɪts]. Janakan petua fonologi yang sesuai untuk fenomena ini.

[25/100]

2. Soalan ini mempunyai dua bahagian. Kedua-dua bahagian MESTI dijawab.

- (a) Kata nama boleh ditulis dalam bentuk tunggal (= *singular form*) ataupun bentuk jamak (= *plural form*), misalnya. *cat ~ cats*, *house ~ houses*. Dalam bahasa Inggeris, makna kejamakan boleh dinyatakan dengan menggunakan pelbagai imbuhan (= 'inflection').

|                           |                          |
|---------------------------|--------------------------|
| <i>baby ~ babies</i>      | <i>maze ~ mazes</i>      |
| <i>box ~ boxes</i>        | <i>rash ~ rashes</i>     |
| <i>bush ~ bushes</i>      | <i>shelf ~ shelves</i>   |
| <i>buzz ~ buzzes</i>      | <i>spy ~ spies</i>       |
| <i>cat ~ cats</i>         | <i>switch ~ switches</i> |
| <i>chicken ~ chickens</i> | <i>tax ~ taxes</i>       |
| <i>echo ~ echoes</i>      | <i>toad ~ toads</i>      |
| <i>frog ~ frogs</i>       | <i>tomato ~ tomatoes</i> |
| <i>glass ~ glasses</i>    | <i>watch ~ watches</i>   |
| <i>house ~ houses</i>     | <i>wife ~ wives</i>      |
| <i>knife ~ knives</i>     | <i>wolf ~ wolves</i>     |

- (i) Dengan menggunakan HANYA data yang diberikan di sini, tentukan imbuhan-imbuhan yang berkenaan dalam bahasa Inggeris. [5/100]
- (ii) Imbuhan-imbuhan ini hanya digunakan dalam keadaan tertentu. Sekarang tuliskan petua-petua supaya bentuk jamak yang sesuai dalam bahasa Inggeris dapat dijanakan. Petua-petua anda harus ditulis dalam format seperti berikut. Nyatakan syarat-syarat petua jika ada.

imbuhan\_tunggal → imbuhan\_jamak | syarat

[40/100]

- (iii) Nyatakan tertib petua-petua ini dalam perisian.

[5/100]

- (b) Dalam bahasa Spanyol, kata nama boleh juga ditulis dalam bentuk tunggal (= *singular form*) ataupun bentuk jamak (= *plural form*), misalnya *libro ~ libros*, *iglu ~ iglúes*. Sebagaimana dalam bahasa Inggeris, makna kejamakan boleh dinyatakan dengan menggunakan pelbagai imbuhan (= 'inflection'). [Dalam bahasa Sepanyol, tekanan (= 'accent') pada sukukata dinyatakan dengan tanda ', seperti dalam *el bambú* 'buluh', *el ladrón* 'perompak' dan *el lápiz* 'pensel'.]

|                       |               |                          |                         |
|-----------------------|---------------|--------------------------|-------------------------|
| <i>el bambú</i>       | 'buluh'       | <i>los bambúes</i>       | 'buluh-buluh'           |
| <i>el corredor</i>    | 'pelari'      | <i>los corredores</i>    | 'pelari-pelari'         |
| <i>el hombre</i>      | 'lelaki'      | <i>los hombres</i>       | 'lelaki-lelaki'         |
| <i>el iglú</i>        | 'iglu'        | <i>los iglúes</i>        | 'iglu-iglu'             |
| <i>el joven</i>       | 'pemuda'      | <i>los jóvenes</i>       | 'pemuda-pemuda'         |
| <i>el ladrón</i>      | 'perompak'    | <i>los ladrones</i>      | 'perompak-perompak'     |
| <i>el lápiz</i>       | 'pensel'      | <i>los lápices</i>       | 'pensel-pensel'         |
| <i>el libro</i>       | 'buku'        | <i>los libros</i>        | 'buku-buku'             |
| <i>el lunes</i>       | 'hari Isnin'  | <i>los lunes</i>         | 'hari-hari Isnin'       |
| <i>el mujer</i>       | 'wanita'      | <i>las mujeres</i>       | 'wanita-wanita'         |
| <i>el nivel</i>       | 'aras'        | <i>los niveles</i>       | 'aras-aras'             |
| <i>el padre</i>       | 'ibu bapa'    | <i>los padres</i>        | 'ibu bapa-ibu bapa'     |
| <i>el papel</i>       | 'kertas'      | <i>los papeles</i>       | 'kertas-kertas'         |
| <i>el paraguas</i>    | 'payung'      | <i>los paraguas</i>      | 'payung-payung'         |
| <i>el pollo</i>       | 'ayam'        | <i>los pollos</i>        | 'ayam-ayam'             |
| <i>el régimen</i>     | 'rejim'       | <i>los regímenes</i>     | 'rejim-rejim'           |
| <i>el rubí</i>        | 'batu delima' | <i>los rubíes</i>        | 'batu-batu delima'      |
| <i>la bici</i>        | 'basikal'     | <i>las bicis</i>         | 'basikal-basikal'       |
| <i>la casa</i>        | 'rumah'       | <i>las casas</i>         | 'rumah-rumah'           |
| <i>la ciudad</i>      | 'bandaraya'   | <i>las ciudades</i>      | 'bandaraya-bandaraya'   |
| <i>la flor</i>        | 'bunga'       | <i>las flores</i>        | 'bunga-bunga'           |
| <i>la nación</i>      | 'negara'      | <i>las naciones</i>      | 'negara-negara'         |
| <i>la pluma</i>       | 'pen'         | <i>las plumas</i>        | 'pen-pen'               |
| <i>la universidad</i> | 'universiti'  | <i>las universidades</i> | 'universiti-universiti' |
| <i>una vez</i>        | 'satu kali'   | <i>algunas veces</i>     | 'beberapa kali'         |

- (i) Dengan menggunakan HANYA data yang diberikan di atas, tentukan imbuhan-imbuhan yang berkenaan dalam bahasa Sepanyol.

[5/100]

- (ii) Imbuhan-imbuhan ini hanya digunakan dalam keadaan tertentu. Sekarang tuliskan petua-petua supaya bentuk jamak yang sesuai dalam bahasa Sepanyol dapat dijanakan. Petua-petua anda harus ditulis dalam format seperti berikut. Nyatakan syarat-syarat petua jika ada.

imbuhan\_tunggal → imbuhan\_jamak | syarat

[40/100]

- (iii) Nyatakan tertib petua-petua ini dalam perisian.

[5/100]

## 3. Diberikan nahu berikut:

$S \rightarrow \text{noun}(\text{SEM0}) \text{ VP}(\text{SEM0})$   
 $\text{VP}(\text{SEM0}) \rightarrow \text{verb}(\text{Comps}, \text{SEM0}, \text{SEM1}) \text{ VerbComps}(\text{Comps}, \text{SEM1})$   
 $\text{VerbComps}(\text{none}, \text{SEM1}) \rightarrow \emptyset$   
 $\text{VerbComps}(\text{noun}, \text{SEM1}) \rightarrow \text{noun}(\text{SEM1})$   
 $\text{noun}(\text{SEM0}) \rightarrow \{\text{noun}, \text{SEM0}\}$   
 $\text{noun}(\text{SEM1}) \rightarrow \{\text{noun}, \text{SEM1}\}$   
 $\text{verb}(\text{Comps}, \text{SEM0}, \text{SEM1}) \rightarrow \{\text{verb}, \text{Comps}, \text{SEM0}, \text{SEM1}\}$

leksikon:

cows : noun, +ANIMATE  
grass : noun, -ANIMATE  
eat : verb, noun, +ANIMATE, -ANIMATE  
sleeps : verb, none, +ANIMATE, -

- (a) Senaraikan semua ayat yang tepat secara semantic tetapi tidak gramatis (dalam Bahasa Inggeris) yang dapat dijanakan oleh nahu di atas. [20/100]
- (b) Kembangkan nahu dan leksikon seperti yang diberikan di atas supaya hanya ayat yang gramatis (dalam Bahasa Inggeris) sahaja dapat dijanakan. [30/100]
- (c) Senaraikan semua ayat yang tepat secara semantic dan gramatis (dalam Bahasa Inggeris) yang dapat dijanakan oleh nahu seperti yang diberikan dalam 3(b). [15/100]
- (d) Janakan suatu 'chart' yang dihasilkan oleh teknik 'bottom-up chart parsing' bagi sesuatu ayat terpanjang yang disenaraikan dalam 3(c). Anda perlu tunjukkan nilai semua pembolehubah yang terlibat di dalam chart yang dijanakan. [35/100]

4. (a) Untuk setiap alat NLP berikut, jelaskan fungsinya dan berikan suatu contoh input/output.

(i) 'Spelling checker'

(ii) 'Inflectional morphological analyzer'

(iii) 'Part of speech tagger'

(iv) 'Parser'

(v) 'Word sense tagger'

[40/100]

(b) Bincangkan secara menyeluruh bagaimana alat NLP dalam 4(a) dapat digunakan untuk membangun aplikasi NLP berikut.

(i) 'Search engine'

(ii) 'Grammar checker'

(iii) 'Intelligent helpdesk system (or call center)'

[60/100]