

UNIVERSITI SAINS MALAYSIA

Peperiksaan Semester Pertama
Sidang 1989/90

Oktober/November 1989

CSI504 - Pemprosesan Bahasa Manusia

Masa : [3 jam]

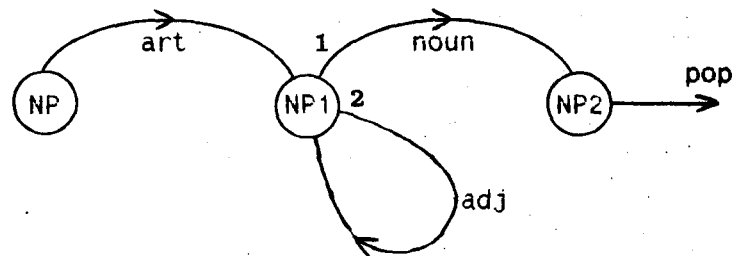
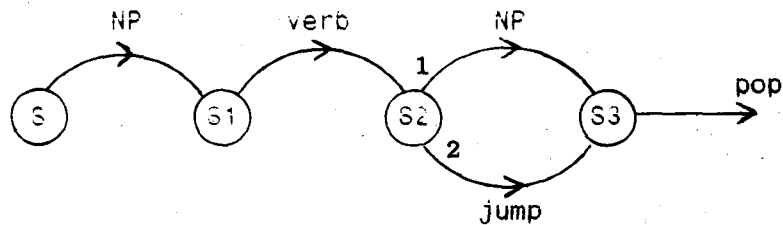
Sila pastikan bahawa kertas peperiksaan ini mengandungi 5 muka surat yang bercetak sebelum anda memulakan peperiksaan ini.

Jawab SEMUA soalan di dalam Bahagian A, dan DUA (2) soalan daripada Bahagian B.

Semua soalan mestilah dijawab dalam Bahasa Malaysia.

Bahagian A. Jawab semua TIGA (3) soalan di dalam bahagian ini.

1. Satu rangkaian peralihan imbuan (augmented transition network ATN) untuk analisis ayat di dalam Bahasa Inggeris ditunjukkan di bawah.



...2/-

Arc	Test	Actions
NP/1	none	DET ← * NUM ← NUM*
NP1/1	NUM ∩ NUM*	HEAD ←-* NUM ← NUM ∩ NUM*
NP1/2	none	ADJS ← Append {ADJS,*}
S/1	none	SUBJ ←-*
S1/1	NUM _{SUBJ} ∩ NUM*	MAIN-V ←-* NUM ← NUM _{SUBJ} ∩ NUM*
S2/1		OBJ ←-*

(a) Terangkan proses penghuraian (parsing) yang akan dilakukan ke atas ayat Bahasa Inggeris.

"A COMPUTER PROCESSES THESE DATA"

dengan menggunakan ATN di atas dan kamus leksikon berikut:

- A: (ART NUM {3s})
- COMPUTER: (NOUN NUM {3s})
- PROCESSES: (VERB NUM {3s})
- THESE: (ART NUM {3p})
- DATA: (NOUN NUM {3s 3p})

Jawapan anda mesti menunjukkan langkah-langkah terperinci yang diambil di dalam proses penghuraian itu.

(40/100)

(b) Tentukan perubahan yang mesti dilaksanakan ke atas ATN di atas supaya meliputi sifat-sifat linguistik berikut:

- (i) kategori sintaks nama NAME (msl. AHMAD, CHONG) dan gatinama PRONOUN (msl. HE, IT);
- (ii) kategori sintaks MODAL (msl. CAN, SHOULD);
- (iii) frasa sendi nama (prepositional phrase) yang wujud di kedua-dua rangkaian S dan NP.

(60/100)

...3/-

2. Nahu Bebas Konteks (context free grammar) dan Klausa Horn

Suatu nahu bebas konteks digambarkan di bawah.

S <-- NP VP
VP <-- verb NP
VP <-- verb NP PP
NP <-- art NP1
NP <-- NP1
NP1 <-- noun
NP1 <-- noun PP
NP1 <-- adj NP1
PP <-- prep NP

- (a) Terangkan satu penghuraian bawah-atas (bottom-up) untuk ayat dan leksikon di bawah dengan menggunakan nahu bebas konteks tersebut di atas.

"THE COMPUTER READS A PORTION OF THE DATA IN THE FAST SMALL DEVICE"

THE: art
COMPUTER: noun
READS: verb
A: art
PORTION: noun
OF: prep
DATA: noun
IN: prep
FAST: adj
SMALL: adj
DEVICE: noun

[Perhatian: Mungkin lebih daripada satu pohon huraian (parse tree) dihasilkan].

(30/100)

- (b) Tentukan petua-petua (rules) tambahan yang diperlukan di dalam nahu bebas konteks di atas untuk memproses ayat-ayat berikut:

"WE WROTE THREE LARGE COBOL PROGRAMS"
"THE COMPUTER CRASHED"

...4/-

WE: pronoun
WROTE: verb
THREE: cardinal
LARGE: adj
COBOL: name
PROGRAMS: noun
CRASHED: verb

(40/100)

- (c) Tulis satu nahu di dalam perwakilan klausa Horn yang setara dengan nahu bebas konteks di atas.

(30/100)

3. Sistem Penulisan Nahu Jenis Penghasilan
(Production-type Grammar Writing System)

- (a) Struktur pohon bercatatan (annotated tree structure) adalah salah satu cara perwakilan maklumat teks/ayat. Terangkan jenis-jenis maklumat linguistik yang boleh disimpan di dalam struktur ini dengan menggunakan beberapa contoh tertentu.

(30/100)

- (b) Bincangkan isu-isu yang berkaitan dengan **strategi kawalan** (control strategy) dan **petua nahu** (grammar rules).

(50/100)

- (c) Apakah kebaikan dan keburukan sebuah sistem penulisan nahu yang bersifat penghasilan 'rule-based'?

(20/100)

...5/-

Bahagian B. Jawab DUA (2) soalan di dalam bahagian ini.

4. Sistem pemprosesan bahasa asli (NLP) biasanya menggunakan berbagai jenis pengetahuan (knowledge), termasuk pengetahuan linguistik, pengetahuan domain dan pengetahuan dunia. Pengetahuan linguistik juga mengandungi beberapa sub-kategori termasuk pengetahuan morfologi, sintaks dan semantiks.

Terangkan bagaimana jenis-jenis pengetahuan ini digunakan di dalam sistem terjemahan mesin (MT) dan sistem antaramuka pangkalan data bahasa asli (NLDFE). Gunakan contoh-contoh tertentu untuk menyokong jawapan anda.

(100/100)

5. Terjemahan Mesin

Tulis nota ringkas untuk setiap topik berikut:

- (a) Pendekatan pemindahan (transfer) dan 'interlingual'.

(30/100)

- (b) Bahasa khas untuk pemrograman linguistik (specialised languages for linguistic programming).

(40/100)

- (c) Apakah sifat-sifat yang dianggap penting untuk menentukan kebolehan sebuah sistem terjemahan mesin (MT) yang praktik?

(30/100)

6. Tulis satu karangan pendek (lebih kurang 500 perkataan) untuk salah satu daripada topik berikut:

- (a) Konsep semantik 'case' dan 'valency'.
(b) Senibina antaramuka pangkalan data bahasa asli (NLDFE).
(c) Penghuraian 'top-down/bottom-up', 'deterministic/non deterministic'.
(d) Penjanaan Teks (text generation).

(100/100)