Mid Term Examination 2nd Semester 2017/2018

CSH403 – NATURAL LANGUAGE PROCESSING

Monday, March 5th 2018 10.15 – 12.15 WIB (120')

Lecturer: ADE

= ***Individual, Close Book and Note*** =

Guidelines

- Read the questions carefully.
- Give clear and sufficiently detail answer.
- You may use ballpoint or pencil on writing the answers.
- Pray before the exam. ☺

Student's Name:	Student's ID Number:	Class:	Room:	Score:
……	……	IF-……	……	

Please write the following sentence:

I am answering all the questions independently and honestly. If I disobey the rules, I am willing to accept sanctions

Student's Sign:

……

……

……

……

Natural Language Processing Mid Term Examination

Competency List:

Competency	Subcompetency
[CLO 1] Able to build and evaluate a system based on language modeling and POSTagging	[CLO 1.1] Building language modeling probability component
	[CLO 1.2] Evaluating language modeling
	[CLO 1.3] HMM, calculating emission and transition probability, and likelihood
[CLO 2] Able to build and evaluate a system based on syantactic parsing	[CLO 2.1] Two views of syntactic parsing
	[CLO 2.2] Parsing algorithm: top-down vs bottom up
	[CLO 2.3] Parsing using CYK algorithm [CLO 2.4] Parser evaluation

1. **[CLO 1.1] Building language modeling probability component [15]**
   Suppose we will build a language model over the following training corpus:
   Ayah berangkat ke kantor tadi pagi.
   Ibu berangkat ke Jakarta tadi malam.
   Ayah dan ibu akan pulang ke rumah malam ini.

a. Train a **unigram** language model using maximum likelihood estimation. What are the probabilities?
b. Train a **bigram** language model using maximum likelihood estimation. What are the probabilities?
   Answer:

2. [CLO 1.2] Evaluating Language Modeling [10]

One method to evaluate language modeling is by calculating the perplexity. Given a test sentence “Ibu pulang dari kantor malam ini”, what is the perplexity of unigram and bigram language model that you have built in answer #1?

3. **[CLO 1.3] HMM - POSTagging [25]**

   Explain how the HMM based method could be used to solve a POSTagging task! Give an example case to support your explanation (design a simple corpus, build the transition and emission table, give a test sentence, etc)!

4. **[CLO 2.1] Two views of syntactic parsing [10]**

   What are the differences between constituent/phrase structure and dependency structure?

5. **[CLO 2.2] Parsing algorithm: top down vs bottom up [15 Poin]**

   Given following CFG:

   S -> NP VP

   VP -> V NP | V NP PP

   PP -> P NP

   V -> "melihat" | "makan" | "berjalan"

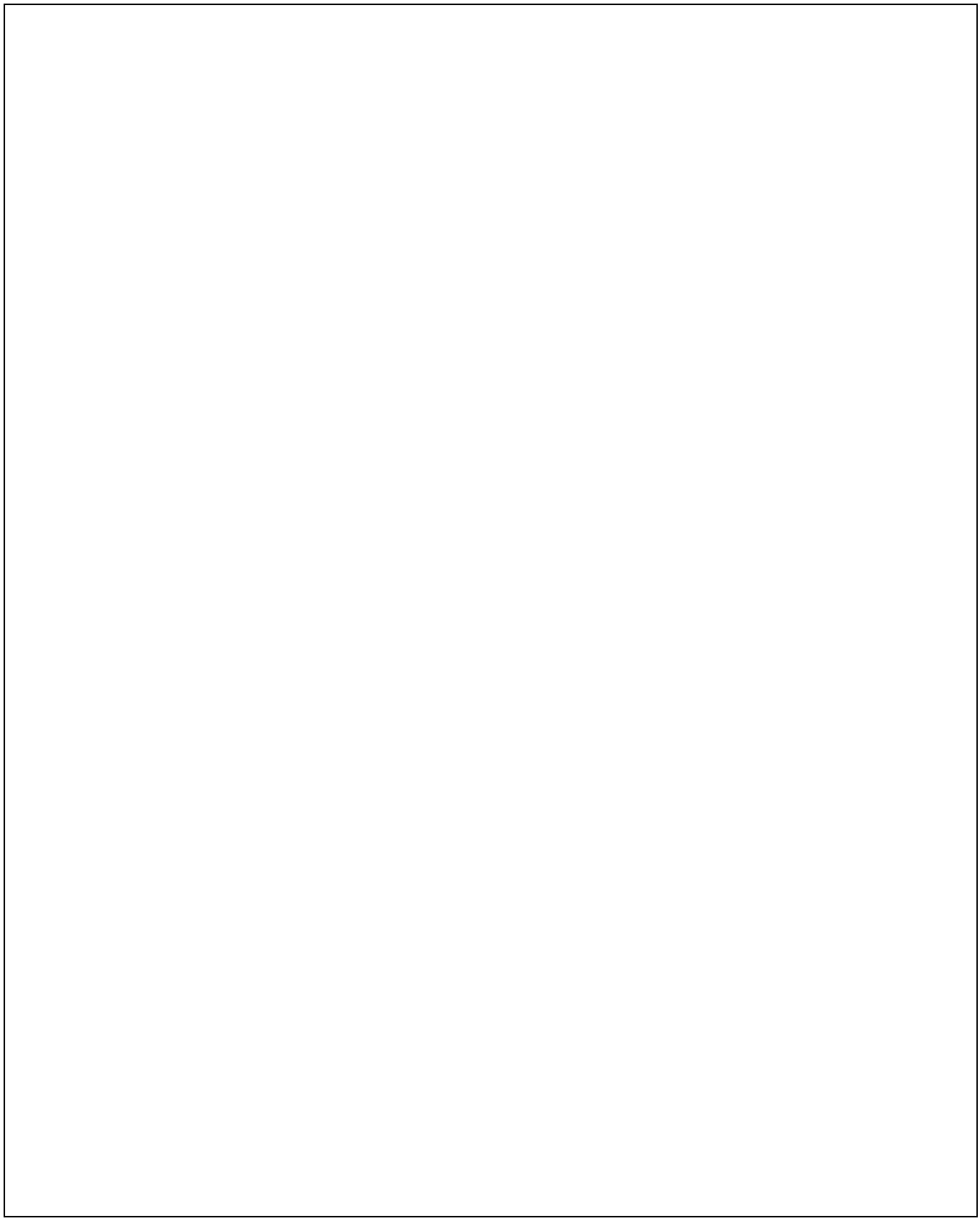
   NP -> "Andi" | "Ani" | "Budi" | Det N | Det N PP

   Det -> "sebuah" | "seorang" | "si" | "sang" | "seekor"

   N -> "orang" | "anjing" | "kucing" | "teleskop" | "taman"

   P -> "di" | "pada" | "oleh" | "dengan"

   Parse "*Andi melihat seekor kucing dengan teleskop*" using top down and bottom up algorithm!



6. [CLO 2.3] CYK Parsing [15]

Given the following grammar:
S -> Aux NP VP
S -> VP
VP -> Verb NP | VP NP | VP PP | Verb
NP -> P NP | Det N
Verb -> book | flight
N -> flight | Jakarta
Aux -> does
P -> in | an | to
Det -> a | the

Convert the grammar into CNF! Parse "*Book the flight to Jakarta*" using CYK algorithm!

Answer:

7. [CLO 2.4] Parser Evaluation [10]

Given the following gold constituent:
S(0, 4), NP(0, 2), VP(2, 4), PP(3, 4)
The constituent that was obtained from proposed system:
S(0, 4), NP(0, 1), VP(1, 4), PP(3, 4)
Evaluate the parser performance by calculating the precision, recall, and F-Measure!