

平成 27 年度自然言語処理 定期試験練習問題

実施日：平成 27 年 12 月 24 日(木)
時間：60 分(持ち込み不可)
担当：CS 学部 亀田弘之

問題 0 ^{がくせきばんごう} 学籍番号・^{しめい} 氏名を^{しよてい} 所定の^{かしょ} 箇所に^{ただ} 正しくかつ^{ていねい} 丁寧に^か 書くこと [10 点]。

Write down your student number and name in the proper box.

問題 1 ^{つぎ} 次の文の^{ぶん} 構造をそれぞれ^{こうぞう} 分析し^{ぶんせき} 説明しなさい。

Analyze and explain the following sentences, respectively.

- (1) 文 1 (Sentence 1) : Jonathan flies.
- (2) 文 2 (Sentence 2) : Thomas runs fast.
- (3) 文 3 (Sentence 3) : Time flies like an arrow.

問題 2 ^{つぎ} 次の形式文法 G について答えよ。

^{ぶんぽう} 文法 G = $\langle \{ \text{文}, \text{主語}, \text{動詞} \}, \{ \text{Charlie, speaks} \}, \text{文}, \{ \text{文} \rightarrow \text{主語}, \text{動詞}, \text{主語} \rightarrow \text{Charlie}, \text{動詞} \rightarrow \text{speaks} \} \rangle$

Answer the questions about the grammar below:

Grammar G = $\langle \{ \text{Sentence, Subject, Verb} \}, \{ \text{Charlie, speaks} \}, \text{Sentence}, \{ \text{Sentence} \rightarrow \text{Subject Verb}, \text{Subject} \rightarrow \text{Charlie}, \text{Verb} \rightarrow \text{speaks} \} \rangle$

- (1) 開始記号は何か。Which is the start symbol of this grammar?
- (2) 非終端記号をすべて列挙せよ。List up all non-terminal symbols.
- (3) 終端記号をすべて列挙せよ。List up all terminal symbols.
- (4) 書き換え規則は全部でいくつあるか。How many rules are there in this grammar?
- (5) この文法により定まる言語 L は何か。What language named L is defined by the grammar G?

(注) 言語とは文の集合である。

Note: Language is defined as a collection of sentences.

問題 3 下記の文章は、統語解析プログラム作成手順に関するものである。空欄部分を適切な用語で埋めよ。Sentences described below show a procedure of building syntactic analyzer. Fill in the blanks with appropriate words.

『工科大のイブラヒム君は試験勉強のために統語解析のプログラムを作った。対象言語 L は、 $L = \{ \text{子犬が眠っている} \}$ とした。

Ibrahim-kun wrote a syntactic analyzer to study for term-exam. He set a language $L = \{ \text{子犬が眠っている (A puppy is sleeping.)} \}$

手順 1) 文の解析

Step 1) sentence analyze

子犬 が 眠っている
名詞 助詞 動詞
主語 述語

文

Note) 名詞: noun, 助詞: postposition

動詞: verb, 主語: subject,

述語: predicate, 文: sentence

手順 2) ^{ぶんぽう} 文法の設定

Step 2) grammar setting

まず、語彙すなわち終端記号の集合 V_t を作成：

$V_t = \{ \text{子犬}, \square, \text{眠っている} \}$

次に、開始記号を決定： \square

次に、非終端記号を決定： $V_n = \{ \text{文}, \text{主語}, \text{述語}, \square, \text{助詞}, \text{動詞} \}$

最後に、上記の情報（用語）を利用して、書き換え規則群を作成。

First, write down vocabulary V_t , i.e., a collection of words:

$V_t = \{ \text{子犬}, \square, \text{眠っている} \}$

Next, decide non-terminal symbols: $V_n = \{ \text{文}, \text{主語}, \text{述語}, \square, \text{助詞}, \text{動詞} \}$

Finally, write down a set of rewriting rules in terms described above.

手順 3) 文法の知識を Prolog 言語で表現する。

Step 3) Describe knowledge of the grammar in Prolog language. 』

問題 4 上記の問題 2 で定義されている文法が定める言語 L の文を受理するプログラムを作成しなさい。

Write a program which accepts a language defined by the grammar of the Problem 2 described above.

問題 5 言語 $L = \{ \text{太郎が歌う。花子が踊る。} \}$ に対して、次の (1) と (2) に答えよ [30 点]。ただし、“太郎” と “花子” は名詞、“が” は助詞、“歌う” と “踊る” は動詞として答えなさい。

Answer the question (1) and (2) below about a language $L = \{ \text{太郎が歌う。花子が踊る。} \}$, where “太郎” and “花子” are nouns, “が” postposition, and “歌う” and “踊る” verbs.

(1) 言語 L を定める形式文法 G を作成したい。

We want to define a formal grammar to define the language L .

(ア) 終端記号の集合（語彙） V_t を書き下しなさい。

Write down V_t , i.e., a set of terminal symbols (vocabulary).

(イ) 非終端記号の集合 V_n を書き下しなさい。

Write down V_n , i.e., a set of non-terminal symbols

(ウ) 開始記号 σ を書き下しなさい。

Write down the start symbol σ .

(エ) 書き換え規則をすべて書き下しなさい。

List up all rewriting rules.

(2) 上記 (1) で作成した文法 G に基づき、言語 L の文を対象とする構文解析プログラムを作成せよ。

Write a program to parse all sentences in L , based on the grammar G of (1).

(【出題意図】 1. 言語分析ができ、それに基づき形式文法を自力で定義できる。2. 形式文法に基づきプログラムが書ける。)

問題 6 情報検索における TF-IDF 法について以下の (1) と (2) に答えよ。

Answer the questions (1) and (2) about the TF-IDF method of Information retrieval.

(1) TF-IDF 法の基本的な考え方を 200 文字程度で述べよ。

Explain a basic idea of the TF-IDF method in around 200 words.

(2) 以下のような 5 つの文書がある。いま、キーワード「未知語」で検索を行うとき、あなたはどの文書を最優先の検索結果とするか？ その理由とともに答えよ。

Given 5 documents as described in a table below. Which document should be the first candidate to show as a retrieval result, when a keyword “未知語 (unknown word)” is given to retriever.

表. 文書とキーワード出現回数

	文書 ₁	文書 ₂	文書 ₃	文書 ₄	文書 ₅
未知語	10	10	10	10	10
単語	4	7	12	9	20
形態素	7	9	20	10	15
文字	6	2	1	8	1

Table. Documents and the # of occurrences of the words in them

	Doc ₁	Doc ₂	Doc ₃	Doc ₄	Doc ₅
unknown word	10	10	10	10	10
word	4	7	12	9	20
Morpheme	7	9	20	10	15
Character	6	2	1	8	1

(【出題意図】 1. TF-IDF 法の意味を説明しかつ使用することができる。)

問題 7 NLP (自然言語処理) の新しいアプリケーションを 1 つ提案しなさい。

ただし、(A) なぜ必要なのか？ (B) 誰に対してメリットがあるのか？

(C) そのアプリケーション実現のために最も重要な要素技術は何か？

についても言及すること。

Propose a novel NLP (Natural Language Processing) application with explanation of (A) why it is needed, (B) to whom it is useful, and (C) what element technology is needed to implement the application you propose.

(【出題意図】 1. NLP 技術のアプリケーションを説得力ある形で提案することができる。)

問題 8 次の用語から 1つを選び、知るところを 100～150 文字程度で簡潔に述べよ。

Select a single term out of the followings, and explain what you know about it in 100—150 words.

- (1)自然言語(natural languages)
- (2)形態素(morphemes)
- (3)形態素解析(morpheme analysis)
- (4)構文解析(parsing)
- (5)言語の意味(semantic of language)
- (6)意味処理(meaning processing)
- (7)未知語(unknown words)
- (8)コーパス(corpora)
- (9)言語処理ツールソフトウェア(NLP tool software)
- (10)人間における言語処理過程(natural language processing processes in human)

(【出題意図】 1. NLP に関する用語・概念を自分の言葉で説明することができる。)

以上