

# 平成 24 年度自然言語処理定期試験練習問題

配布日：平成 25 年 1 月 18 日(月)4 限  
試験時間(予定)：60 分(持ち込み不可)  
担当：CS 学部 亀田弘之

**問題 0** 学籍番号・氏名を所定の箇所に正しくかつ丁寧に書くこと[10 点]。

**問題 1** 言語  $L = \{ \text{山田さんは美しい。山田さんは楽しい。} \}$  に対して、次の (1) ~ (2) に答えよ[30 点]。ただし、“**田中さん**” は名詞、“**は**” は助詞、“**美しい**” と “**楽しい**” は形容詞として答えなさい。

(1) 言語  $L$  を定める形式文法  $G$  を作成せよ[10 点]。

(ヒント 終端記号, 非終端記号, 書き換え規則等の用語を使って表現せよ。)

(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。[20 点]

(コメント Prolog で表現することを試みよ。)

(【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 2** 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[40 点]。

(1) TF(Term Frequency) という単語の意味を説明しなさい[5 点]。

(2) DF(Document Frequency) という単語の意味を説明しなさい[5 点]。

(3) TF-IDF 法の考え方とともに、TF-IDF 法の問題点を指摘せよ。[10 点]。

(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」で検索するとき、もっとも関連性の高い文書はどれか？ TF-IDF 法の観点からその理由も答えよ。必要に応じて計算式も示すこと。[20 点]

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

(【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 3** 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[20 点]。

ただし、(A) なぜ必要なのか？[5 点] (B) 誰に対してメリットがあるのか？[5 点]

(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？[10 点]

についても言及すること[20 点]。

(【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 4** 次の用語のうち 1 つを選び、知るところを 200 文字程度で述べよ[20 点]。

(1) 自然言語(定義・種類等) (2) 形態素(定義・例等) (3) 形態素解析(処理方法と長所短所等) (4) 構文解析(処理方法と長所短所等) (5) 言語の意味(概念、意味属性、シニフィエ、シニフィアン等) (6) 感情処理(意義、基本的考え等) (7) 未知語(定義、種類、意義、処理方法等) (8) コーパス(定義、種類、意義) (9) 言語処理ツールソフトウェア(実例の紹介等) (10) 人間における言語処理過程 (10) 言語表現の曖昧性(ambiguity)(定義、種類、例等)

(【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