

平成 22 年度自然言語処理定期試験問題

実施日:平成 23 年 1 月 24 日(月)1限(9:00-10:00)

試験時間:60 分(持ち込み不可) 担当: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 $L = \{\text{The boys loved animals. The girls loved flowers.}\}$ に対して、次の (1) ~ (2) に答えよ[30 点]。ただし、the は限定詞、boys, girls, animals および flowers は名詞、loved は動詞である。

(1) 言語 L を定める形式文法 G を作成せよ。つまり、非終端記号の集合 V_n 、終端記号の集合 V_t 、開始記号 σ および書き換え規則群 P をそれぞれ決定しなさい[10 点]
なお、必要に応じて以下の用語を適宜使用しなさい。

s : sentence (文) subj : subject (主語) obj : object (目的語)
np : noun phrase (名詞句) vp : verb phrase (動詞句) v : verb (動詞)
det : determinant (限定詞) art : 冠詞(article) n : noun (名詞)
prpn : proper noun (固有名詞)

(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]

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[20 点]。

- (1) TF-IDF 法の基本的な考え方を 200 文字程度で述べよ。[10 点]
(2) 以下のような 4 つの文書がある。いま、キーワード「翻訳」で検索を行うとき、もっとも関連性の高い文書はどれか? その理由とともに答えよ。[10 点]

事実:

表. 文書とそこに含まれるキーワード

文書名	含まれるキーワードとその個数
文書1	言語(2)、コンピュータ(1)、翻訳(3)
文書2	コンピュータ(1)、翻訳(2)
文書3	言語(2)、翻訳(3)、知識(2)
文書4	翻訳(2)、知識(1)

(注1) キーワードの後のカッコ付き数字は、そのキーワードの出現回数を表している。

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。
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]

問題4 次の用語のうち 1 つを選び、知るところを 500~800 文字程度で述べよ[20 点]。

- (1)自然言語 (2)形態素 (3)形態素解析 (4)構文解析
(5)言語の意味 (6)意味処理 (7)未知語 (8)コーパス
(9)言語処理ツールソフトウェア (10)人間における言語処理過程

以 上