

平成 20 年度自然言語処理定期試験問題

実施日:平成 21 年 1 月 26 日(月)1限

試験会場:講実 A403 試験時間:60 分 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 $L = \{ \text{Students had a book. Teachers drove a car.} \}$ に対して、次の (1) ~ (2) に答えよ[30 点]。

(1) 言語 L を定める文法 G を作成せよ。[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 表情報検索における TF-IDF 法について以下の (1) と (2) に答えよ[20 点]。

(1) TF-IDF 法の計算式 (定義式) を書け。[5 点]

(2) 以下のような 4 つの文書がある。いま、キーワード「言語」と「翻訳」との AND 検索を行ったとき、もっとも関連性の高い文書はどれか? [15 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(2) |
| 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。

(注 2) このような表は転置インデックス(索引)とも呼ばれる。

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、なぜ必要なのか? 誰に対してメリットがあるのか? 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、知るところを述べよ[10 点]。

(1)自然言語 (2)形態素 (3)形態素解析 (4)構文解析

問題5 日本語文「飛行機が飛ぶ」に対して、以下のような構文解析プログラムを作成した。これを形態素解析もあわせて実行することのできるように改良せよ[10 点]。

文(A, C, 文(_主語, _述語)) :- 主語(A, B, _主語), 述語(B, C, _述語).
主語(A, C, 主語(_名詞, _助詞)) :- 名詞(A, B, _名詞), 助詞(B, C, _助詞).
述語(A, B, 述語(_動詞)) :- 動詞(A, B, _動詞).
名詞([飛行機], [], 名詞(飛行機)).
助詞([が], [], 助詞(が)).
動詞([飛ぶ],[], 動詞(飛ぶ)).

以 上