

平成 19 年度自然言語処理定期試験問題

実施日:平成 20 年 1 月 21 日(月)1限

試験会場:講実 A402 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 $L = \{ \text{Students have a book. Teachers drive a car.} \}$ に対して、次の (1) ~ (3) に答えよ。

(1) 言語 L を定める文法 G を作成せよ。[20 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。

(1) TF-IDF 法の計算式(定義式)を書け。[10 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? 計算式・途中結果も記すこと。[20 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) このような表は転地インデックス(索引)とも呼ばれる。

問題3 NLP (自然言語処理) の新しい新プロケーションを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]

以上