

平成 19 年度科学研究費補助金実績報告書（研究実績報告書）

1. 機関番号 1 4 6 0 3 2. 研究機関名 奈良先端科学技術大学院大学
3. 研究種目名 若手研究(B) 4. 研究期間 平成 18 年度 ～ 平成 20 年度
5. 課題番号 1 8 7 0 0 1 4 8

6. 研究課題名 逆系列アラインメント問題の条件付確率場による解法と情報抽出への適用

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
9 0 3 1 1 5 8 9	<small>カガナ シンボ, マサシ</small> 新保, 仁	情報科学研究科	助教

8. 研究分担者(所属研究機関名については、研究代表者の所属研究機関と異なる場合のみ記入すること。)

研究者番号	研究分担者名	所属研究機関名・部局名	職名
.....	<small>カガナ</small>		
.....	<small>カガナ</small>		
.....	<small>カガナ</small>		
.....	<small>カガナ</small>		
.....	<small>カガナ</small>		

9. 研究実績の概要(国立情報学研究所でデータベース化するため、600字～800字で記入。図、グラフ等は記載しないこと。)

下欄には、当該年度に実施した研究の成果について、その具体的内容、意義、重要性等を、交付申請書に記載した「研究の目的」、「研究実施計画」に照らし、600字～800字で、できるだけ分かりやすく記述すること。また、国立情報学研究所でデータベース化するため、図、グラフ等は記載しないこと。

本年度の成果は以下の通り。

- 昨年度考案した機械学習に基づく並列句解析手法をさらに発展させるとともに、C++ によるコードと用いているデータ構造を改良し、当初のプロトタイプ実装比3倍の速度向上を得た。
- 学習用訓練データに対するタグ付け作業負担を軽減するため、並列句両端のみのタグ付けデータを用いても十分な解析精度が得られる手法を目指しているが、このような不確定なデータに対処する手法を新たに考案し、英語医学論文アブストラクト集に適用し新手法の有効性を確認した。
- 各種の素性（解析に用いる手がかり）の有効性について調査し、コーパスから収集した単語共起頻度や、類語辞書を取り扱う枠組をコードに組み込んだ。これらについては次年度実験を行い有効性を証明する予定である。
- 医療アブストラクト要約システムの実現に向け、重要文抽出とトピック分割を組み合わせてアブストラクト全体の要約を行う手法を考案した。Document Understanding Conference が公開している標準的な文書自動要約ベンチマーク問題に適用したところ、考案手法が既存法を凌駕する性能を持つことがわかった。
- 日本語への適用にむけ、京都テキストコーパスに対して名詞並列句範囲のタグ付けを行った。

※ 成果の公表を見合わせる必要がある場合は、その理由及び差し控え期間等を記入した調書(A4 判縦長横書 1 枚)を添付すること。

10. キーワード

- | | | |
|-----------|------------|-------------|
| (1) 並列句解析 | (2) 構文解析 | (3) アラインメント |
| (4) 医学文献 | (5) 自然言語処理 | (6) |
| (7) | (8) | (裏面に続く) |

11. 研究発表（平成19年度の研究成果）

〔雑誌論文〕 計（1）件

著者名	論文標題			
原一夫, 新保仁, 松本裕治	アラインメントと機械学習を応用した並列句解析: 医学生物学論文からの情報抽出に向けて			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
人工知能学会論文誌	有	22	2008	248-255

〔学会発表〕 計（3）件

発表者名	発表標題		
H. Bhandari, M. Shimbo, T. Ito, Y. Matsumoto	Generic Text Summarization Using Probabilistic Latent Semantic Indexing		
学会等名	発表年月日	発表場所	
International Joint Conference on Natural Language Processing	2008.1	Hyderabad, India	

発表者名	発表標題		
小町守, 工藤拓, 新保仁, 松本裕治	カーネル法を用いた意味的類似度の定義とブートストラップの一般化		
学会等名	発表年月日	発表場所	
言語処理学会第14回年次大会	2008.3	東京	

発表者名	発表標題		
新保仁	カーネル法によるリンク・引用解析		
学会等名	発表年月日	発表場所	
人工知能学会・人工知能基本問題研究会	2007.7	湯布院	

〔図書〕 計（0）件

著者名	出版社		
書名	発行年	総ページ数	

12. 研究成果による産業財産権の出願・取得状況

〔出願〕 計（0）件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	出願年月日	国内・外国の別

〔取得〕 計（0）件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	取得年月日	国内・外国の別

13. 備考

※ 研究者又は所属研究機関が作成した研究内容又は研究成果に関するwebページがある場合は、URLを記載すること。

--