

平成22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                特定領域研究                4. 研究期間      平成18年度～平成22年度
5. 課題番号 1 8 0 6 1 0 0 5
6. 研究課題名                書き言葉コーパスの自動アノテーションの研究

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
1 0 2 1 1 5 7 5	マツモト ユウジ 松本 裕治	情報科学研究科	教授

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
2 0 1 9 7 8 7 5	トクナガ タケノブ 徳永 健伸	東京工業大学・大学院情報理工学研究科	教授
6 0 2 7 2 6 8 9	イヌイ ケンタロウ 乾 健太郎	東北大学大学院・情報科学研究科	教授
0 0 3 5 7 7 6 6	ハシダ ヨウイチ 橋田 浩一	独立行政法人産業技術総合研究所・サービス工学研究センター	次長
8 0 3 7 9 5 2 8	アサハラ マサユキ 浅原 正幸	情報科学研究科	助教
1 0 3 4 5 3 8 2	ハシモト タイイチ 橋本 泰一	東京工業大学・総合プロジェクト支援センター	特任准教授

9. 研究実績の概要

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。

本特定領域研究で構築されるコーパスに対して、様々なアノテーションを施すための自動言語解析ツールとアノテーション支援およびコーパス利用ツールの構築を研究目的とした。日本語コーパスへの形態素情報の付与については、電子化辞書班とデータ班が担当することになっており、我々の研究グループ（ツール班）では、形態素情報より上のアノテーションを担当し、そのための様々な言語解析ツール、アノテーション支援ツールの構築、および、コーパスへの具体的なアノテーション作業を実施した。

構築したツールの主なものは、自動言語解析ツールとしては、日本語係り受け解析、固有表現解析、述語項構造解析、照応・共参照解析、モダリティ解析ツールがあり、これらの解析ツールを機械学習を用いて構築するため、および、性能評価のため、それぞれに対応するタグ（アノテーション）付きコーパスを構築するとともに、自然言語解析ツールとして実装した。

コーパスアノテーションの支援ツールとしては、形態素、文節、係り受け解析に特化したコーパス管理ツール「茶器」に並列構造のアノテーション機能や文末情報の挿入・削除機能などを実装し、想定したアノテーションのすべてを単独で実現できるシステムとして完成させた。汎用のコーパスアノテーションツール「Slate」を Web ブラウザ上で動作可能なシステムとして構築した。また、様々なタグ付きコーパスやコーパス構築ツールを相互運用するためのツールを構築した。

10. キーワード

- |             |             |            |
|-------------|-------------|------------|
| (1) コーパス    | (2) 形態素解析   | (3) 統語解析   |
| (4) 述語項構造解析 | (5) アノテーション | (6) 自然言語解析 |
| (7) 照応解析    | (8) 談話解析    | (裏面に続く)    |

11. 研究発表（平成22年度の研究成果）

〔雑誌論文〕 計（ 3 ）件    うち査読付論文 計（ 3 ）件

著者名	論文標題			
飯田龍, 小町守, 井之上直也, 乾健太郎, 松本裕治	述語項構造と照応関係のアノテーション: NAISTテキストコーパス構築の経験から			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
自然言語処理	有	17-2	2010	25-50

著者名	論文標題			
Ai Azuma and Yuji Matsumoto	A generalization of forward-backward algorithm			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
Transactions of the Japanese Society for Artificial Intelligence	有	25-3	2010	494-503

著者名	論文標題			
吉川克正, 平尾努, リーデル セバスチャン, 浅原正幸, 松本裕治	共参照関係を用いた医学生物学文書中の事象抽出			
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	26-2	2011	318-323

〔学会発表〕 計（ 1 ）件    うち招待講演 計（ 1 ）件

発表者名	発表標題	
松本裕治	単語依存構造と単語共起による言語解析と理解	
学会等名	発表年月日	発表場所
言語科学会第12回年次国際大会	2010年6月26日	電気通信大学, 東京

〔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。

--