

平成21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 14603 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 18061005
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
10211575	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
20197875	徳永 健伸	東京工業大学・大学院情報理工学研究科	教授
60272689	乾 健太郎	情報科学研究科	准教授
00357766	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
10345382	橋本 泰一	東京工業大学・統合研究院	特任准教授
80379528	浅原 正幸	情報科学研究科	助教

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。

コーパス管理ツール「茶器」のデータベース仕様を見直し、.NETフレームワーク上での再構築を行った。統語解析における並列構造のアノテーション仕様について検討し、並列構造の範囲に関するアノテーションと係り受け構造を表示し、修正を行うことができるTreeEditインタフェースを完成させた。述語項構造解析と事象間の時間関係解析に全域的な情報を利用することで精度向上を実現した。また、照応解析について、先行詞候補に制限を与えることで精度を犠牲にせず効率改善が可能であることを示した。

汎用アノテーションツールSLATの機能を拡張し、Ver. 2.0の基本設計と一部の実装を完了した。旧版に比べ、アノテーション工程の管理まで視野に入れてデータベース・スキーマを再設計し、また、クライアントの処理の大幅な高速化をはかった。

さまざまな書式のコーパスデータに対して統合的なアクセスを可能にするためのアーキテクチャを設計し、これを単一の種類のコーパスデータに関して実装することにより一般に多様なデータを統合するための準備を行なった。

20年度作成した拡張固有表現コーパス（白書、書籍、Yahoo!知恵袋）に対して、タグ付けの見直しを行い、さらに新聞(380文書)、雑誌(79文書)に対してタグ付けを行った。また、20年度に作成したコーパスを用いて、機械学習アルゴリズムの一つであるCRFをベースに固有表現認識ツールを開発した。評価実験を行ったところ、精度約80%、再現率約46%、F値約60%という結果を得た。

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) 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21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			
吉川克正, リーデル セバスチャン, 浅原正幸, 松本裕治	Markov Logicを利用した時間的順序関係の同時推論			
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	24	2009	521-530

著者名	論文標題			
大熊秀治, 原一夫, 新保仁, 松本裕治	バイパス付きアラインメントグラフを用いた日本語並列句検出と範囲同定			
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	25	2010	206-214

著者名	論文標題			
渡邊陽太郎, 浅原正幸, 松本裕治	述語語義と意味役割の結合学習のための構造予測モデル			
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	25	2010	252-261

〔学会発表〕 計（11）件 うち招待講演 計（0）件

発表者名	発表標題	
Tokunaga Takenobu	Aspects of Language Resource Management: Creation and Utilisation	
学会等名	発表年月日	発表場所
2nd FLReNet Forum	2010年2月11日	Barcelona, Spain

発表者名	発表標題	
Yotaro Watanabe, Masayuki Asahara, and Yuji Matsumoto	Multilingual Syntactic-Semantic Dependency Parsing with Three-Stage Approximate Max-Margin Linear Models	
学会等名	発表年月日	発表場所
In Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL)	2009年6月4日	Boulder, CO, USA

発表者名	発表標題	
Katsumasa Yoshikawa, Sebastian Riedel, Masayuki Asahara, and Yuji Matsumoto	Jointly Identifying Temporal Relations with Markov Logic	
学会等名	発表年月日	発表場所
In Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2009)	2009年8月3日	Singapore

発表者名	発表標 題	
Ryu Iida, Kentaro Inui, and Yuji Matsumoto	Capturing Saliency with a Trainable Cache Model for Zero-anaphora Resolution	
学 会 等 名	発表年月日	発 表 場 所
In Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4 th International Joint Conference on Natural Language Processing(ACL-IJCNLP 2009)	2009年8月4日	Singapore

発表者名	発表標 題	
Hideharu Okuma, Kazuo Hara, Masashi Shimbo, and Yuji Matsumoto	Bypassed Alignment Graph for Learning Coordination in Japanese Sentences	
学 会 等 名	発表年月日	発 表 場 所
In Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4 th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2009)	2009年8月4日	Singapore

発表者名	発表標 題	
Ai Azuma and Yuji Matsumoto	A Generalization of Forward-backward Algorithm	
学 会 等 名	発表年月日	発 表 場 所
In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases(ECML/PKDD)	2009年9月8日	Bled, Slovenia

発表者名	発表標 題	
渡邊陽太郎, 浅原正幸, 松本裕治	オンライン最大マージン学習アルゴリズムに基づく多言語依存構造-述語項構造解析	
学 会 等 名	発表年月日	発 表 場 所
情報処理学会研究報告, 自然言語処理研究会, 2009-NL-192, No.2	2009年7月22日	北海道北見市

発表者名	発表標 題	
Koiti Hasida	WG2 PWI24617-6: SemAF-Discourse Relations.	
学 会 等 名	発表年月日	発 表 場 所
ISO/TC37/WG Meeting	2010年1月16日	Cyty University Hong Kong

発表者名	発表標 題	
橋本泰一, 中村俊一	拡張固有表現タグ付きコーパスの構築- 白書, 書籍, Yahoo!知恵袋コアデータ -	
学 会 等 名	発表年月日	発 表 場 所
言語処理学会第16回年次大会	2010年3月11日	東京都文京区

発表者名	発表標題	
渡邊陽太郎, 浅原正幸, 松本裕治	述語語義と意味役割の結合学習のための構造予測モデル	
学会等名	発表年月日	発表場所
言語処理学会第16回年次大会	2010年3月11日	東京都文京区

発表者名	発表標題	
東藍, 新保仁, 松本裕治	半環に基づく前向き後ろ向きアルゴリズムの一般化	
学会等名	発表年月日	発表場所
人工知能学会データマイニングと統計数理研究会資料, SIG-DMSM-A903	2010年3月30日	東京都立川市

〔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。

<http://sourceforge.jp/projects/chaki/>