

自己評価報告書

平成23年3月31日現在

機関番号：10106

研究種目：基盤研究（C）

研究期間：2008～2011

課題番号：20500833

研究課題名（和文）情報の遷移にダイナミックに追従するインターネット単語帳システムの開発

研究課題名（英文）Development of Internet-based Dynamic Wordbook following Information Transition

研究代表者

榎井 文人（マスイ フミト）

北見工業大学・工学部・准教授

研究者番号：80324549

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学

キーワード：教材情報システム，学習環境，自然言語処理

1. 研究計画の概要

本研究では、「知る・理解する・試す」という学習サイクルのうち、「理解する」課程を支援することを目的として、インターネット単語帳システムの開発に取り組む。開発しようとするシステムは、World Wide Web (WWW)から知識を逐次的に獲得し、ことばの意味や特徴、使われ方などを提示することで利用者の語彙把握課程を支援する。具体的には、以下の項目達成を目指す。

(1) WWW上の知識遷移に追従し、ことばを素描する知識断片(descriptor)を獲得する機構の構築

(2) 獲得したdescriptor集合を洗練する機構の構築

(3) descriptorを上位概念、属性値、例示、比喩などに分類する機構の構築

(4) ユーザの意図や興味に応じて複数の結果提示を可能とする可視化機構の構築

2. 研究の進捗状況

(1) 29個のクエリ(名詞)を用いてWWWからdescriptorを半自動で収集し、得られたデータを分析した。descriptorの尤度を5種類の統計量(頻度、頻度と局所性、独立性、正規性、相互情報量)で表現し、それらの分布と傾向、タイプや多義性との関連を調べた。その結果、descriptor集合の全体的な表現能力、集合規模と表現能力の関係が把握できた。

(2) 被験者90名を用いて心理学実験を実施した。6個の刺激語に対するdescriptorを収集した。6カテゴリ(上位語・主体-特徴・本体-部分・組織-構成要素・比喩・連想)に対する適合度を評価したところ、比喩や上位語

に関して被験者間で高い一致性が認められた他、刺激語の抽象度とdescriptor集合のばらつきや規模に関する知見が得られた。

(3) 実装システム(Murasaki)を構築し、これを用いて評価実験を実施した。実験の結果、新語や注目語に対する即応性(Wikipediaに近い性能が得られた)とランキング性能(上位3位以内には妥当な結果を提示できることがわかった)について十分有効な性能を発揮できることを確認した。

(4) 4種の統語パターンを利用してWWWから適合フィードバックを行い、descriptorを3つのカテゴリに分類する機構を構築した。実験により約60%の分類精度が確認された。さらに、システムログからクラスへの帰属率と支配率を計算し分類結果を精緻化する手法も提案した。

(5) 獲得したdescriptorに対して比喩的關係を利用したフィードバックに基づく洗練モジュールを実装した。さらに、常識判断処理・感情情報処理を組合せた精緻化を検討した。

(6) 獲得したdescriptor集合のランキング結果のグラフ提示や分類結果、該当文書の提示インタフェースを実装した。また、提案手法と感情表現処理を組み合わせ、掲示板上の不適切表現の判別に応用した。

3. 現在までの達成度

②おおむね順調に進展している。

(理由)

計画以上に進展している項目とやや遅れている項目があるが、総合的には順調と判断

できる。以下、項目毎に詳細を述べる。

(1) WWWにおける descriptor の調査、人間判断による descriptor の調査については、当初計画通りのデータ収集およびそれらの分析を終了した。

(2) descriptor 獲得機構は計画通りの性能を達成している。抽出適合率およびランキング性能についてはほぼ計画(再現率 80%)通りの性能を達成しており、新語や注目語に対する即応性については既存辞書の性能を大きく上回り、ヒット率に関しては Wikipedia に近い性能が得られた。この結果は、計画以上の性能を達成しているといえる。

(3) 分類機能は 60%程度の適合率が得られているが、現時点では計画していた性能(適合率 80%)を下回っている。

(4) 洗練機構については、フィードバックモデルの実装と基本性能は確保できているが、精緻化のための常識判断処理・感情情報処理の導入については技術的課題の存在が明らかになり現時点では未達である。

(5) 獲得結果および分類結果のグラフによる可視化インタフェイスに加え、分類結果、文書提示など幾つかの提示インタフェイスの実装も終え、当初計画以上に進展したといえる。

4. 今後の研究の推進方策

今後は以下の項目に注力して研究を推進する予定である。以下に項目毎に方策の詳細を述べる。

(1) 現状に加え、目的や意図に応じて柔軟に対応できるインタフェイス機能(下記)を実装する。

① 提示された descriptor を介して関連する説明文を参照できるインタフェイス

② 得られた descriptor を構造化して視覚化提示するインタフェイス

(2) 分類機構をより精緻化するために、他の手法を複数組み合わせ合わせた新たなアプローチを検討する。また、より詳細な分類を可能とするため、分類機構を拡張する予定である。

(3) 常識判断処理・感情情報処理を応用した洗練機構の実装と精緻化をすすめる予定である。

5. 代表的な研究成果

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 2 件)

- (1) 梶井文人, ジェプカ・ラファウ, 木村泰知, 福本淳一, 荒木健治: “WWW を用いた比喩的素描手法”, 知能と情報(日本知能情報ファジィ学会誌), 査読有, Vol. 21, No. 6, 2010, pp. 707-719.

- (2) Michal PTASZYNSKI, Pawel DYBALA, Tatsuaki MATSUBA, Fumito MASUI, Rafal RZEPKA, Kenji ARAKI and Yoshio MOMOUCHI: “In the Service of Online Order: Tackling Cyber-Bullying with Machine Learning and Affect Analysis”, International Journal of Computational Linguistics Research, 査読有, Vol.1, Issue 3, 2010, pp.135-154.

[学会発表] (計 17 件)

- (1) Fumito MASUI, Rafal RZEPKA, Yasutomu KIMURA, Junichi FUKUMOTO and Kenji ARAKI: “Acquisition of Japanese Word Descriptions from World Wide Web”, International Workshop on Modern Science and Technology 2010, 2010.09.05, Kitami, Japan.
- (2) Tyson Michael ROBERTS, Rafal RZEPKA, Fumito MASUI and Kenji ARAKI: “A Multi-Input Approach for a System for Semantically Relevant Art Creation”, International Workshop on Modern Science and Technology 2010, 2010.09.05, Kitami, Japan.
- (3) Michal PTASZYNSKI, Pawel DYBALA, Tatsuaki MATSUBA, Fumito MASUI, Rafal RZEPKA and Kenji ARAKI: “Machine Learning and Affect Analysis Against Cyber-Bullying”, LaCATODA2010, 2010.03.29, Leicester, England.
- (4) 小島鉄也, 梶井文人, 井須尚紀, 河合敦夫: “精緻な語のディスクリプションモデル構築のための心理学実験と分析”, 電子情報通信学会 NLC 研究会/情報処理学会 NL 研究会, 2009.07.22, 北見市.
- (5) RZEPKA Rafal, 梶井文人, 荒木健治: “The First Challenge to Discover Morality Level In Text Utterances by Using Web Resources”, 第 23 回人工知能学会全国大会, 2009.06.20, 高松市.
- (6) 梶井文人, 川村佳史, 河合敦夫, 井須尚紀: “Web 文書に基づくディスクリプションの提案”, 電子情報通信学会 Web インテリジェンスとインタラクション研究会, 2008.07.19, 淡路市.

[その他]

試作システム「Murasaki」(公開予定)
<http://orion.cs.kitami-it.ac.jp/>