様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２４年　６月１５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 A Dynamic Internet-based Wordbook System to Follow Information Transformation
研究代表者
桝井文人（　Masui　Fumito　）
北見工業大学・工学部・准教授
研究者番号　：80324549

研究成果の概要（和文）：　本研究では，クエリ語に対して説明文や定義文を回答する代わりに WWW から収集した断片知識を使って比喩的に素描し，視覚化提示するインターネット単語帳システムを開発した．いくつかの評価実験を実施した結果，開発システムは，説明文や定義文の代わりとして有効に機能すること，既存の汎用辞書では対応できない新語や固有名詞に対しても有効であることを明らかにするとともに，技術的課題も確認した．

研究成果の概要（英文）：　In this research, we tried to develop an Internet-based word reference system for describing a Japanese word, not with explaining or defining sentences, but with figurative descriptions. Some experiments using a prototype system have been conducted. The responsiveness of the system for hot keywords shows that the outcome of the evaluation had exceeded that of a common dictionary. Furthermore, technical problem of the system had been also defined at the present.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００８年度	1,300,000	390,000	1,690,000
２００９年度	900,000	270,000	1,170,000
２０１０年度	600,000	180,000	780,000
２０１１年度	600,000	180,000	780,000
年度			
総　計	3,400,000	1,020,000	4,420,000

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．研究開始当初の背景

(1) 新しい語やうろ覚えの語について調べたい場合，WWWを知識源として利用する方法が有望だが，的確な情報アクセスを実現するためには，玉石混淆の情報源から要求に適したデータ(玉)を取り出す必要がある．

(2) 特に時事的語彙の認識はその評判や注目度によって動的に遷移するため，これらを的確に把握するためには新たに注目されている語義情報を抽出整理できる速効性と網羅性を備えた知識獲得機構が必要である．

(3) 申請者らは，説明文や語義文を提示する以外にも，これは複数の付箋に書き込まれた情報を総合解釈することで事柄を把握するように，知識断片を網羅的に収集・整理して提示することによってもその目的を達成できるのではないかと考察を進めていた．

(4) パターンを利用してWWWから語義特徴を示す知識断片を高精度かつ効率的に収集できるという知見，複数の格フレームや係り受け関係の性質の利用することによりWWWから常識知識を効率よく獲得する手法，データ集合の統計的性質の分析から得た知見と帰納的学習モデルを組み合わせたデータ分類手法の研究成果を応用すれば，WWWから知識断片を獲得して洗練と整理を施して提示することで，逐次最新の語彙情報を提供して利用者の語彙把握を支援できるという着想に至った．

２．研究の目的

(1) 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組む．開発しようとするシステム(図 1)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した．

① WWW 上の知識遷移に追従し，ことばを素描する知識断片を獲得する機構の構築
② 獲得した知識断片集合を洗練する機構の構築
③ 知識断片を上位概念，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) また，開発しようとするシステムから高い性能を得るために，特に各計算機構の基本性能の確認と技術的課題を明らかにすることも研究の目的とする．

３．研究の方法

(1) まず，WWW から得られる知識断片の特性を把握するための調査を分析を行った．

① 29 個のクエリ語(名詞)を用いて WWW から知識断片を半自動で収集し，得られたデータを分析した．知識断片の尤度を 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．
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知識断片を収集した．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．

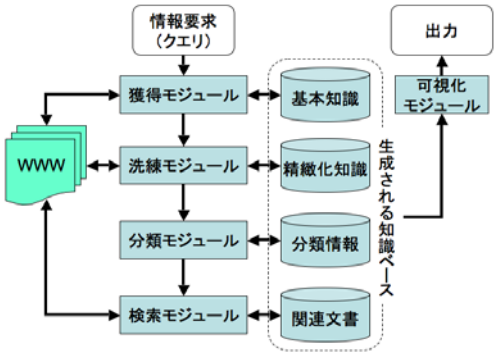

図1 システム構成

(2) 上で得られた知見に基づいて，プロトタイプシステム（Murasaki）を構築し，これを用いて評価実験を実施した．

① 知識獲得モジュールとして，比喩表現を生成する統語パターンを用いて WWW からクエリ語を表現する知識断片を収集し，頻度と局所性を考慮した尤度に基づいてランキングするモジュールを構築した．
② 知識分類モジュールとして，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，知識断片を 3 つのカテゴリに分類するモジュールを構築した．
③ 獲得した知識断片に対して，比喩的関係を利用したフィードバック洗練モジュールを実装した．さらに，常識判断処理と感情判断処理を利用した獲得知識断片のノイズ除去について検討した．
④ 目的や意図に応じて柔軟に対応できる可視化インターフェイス機能を実装し，それぞれ定性的評価によって有効性を確認した．

(3) 各モジュールを設計実装した後，結合評価実験を実施した．

４．研究成果

(1) 調査分析の結果，知識断片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．また，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，刺激語の抽象度と知識断片集合のばらつきや規模に関する知見が得られた．

(3) 知識獲得モジュールの性能については，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

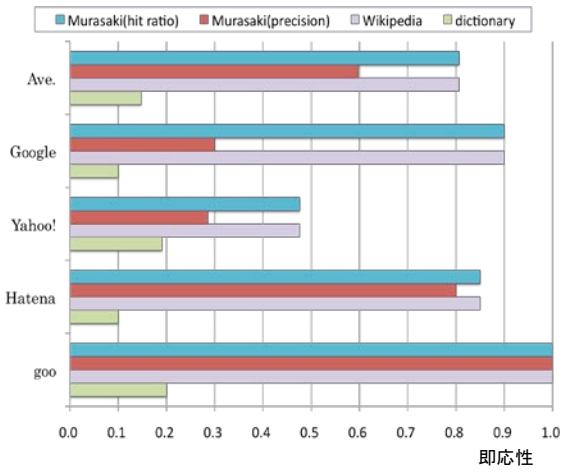

図 2 即応性に関する評価結果

図 3 システムインターフェイス画面

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（図2）．

(4) 知識分類モジュールに関しては，評価実験により約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．分類機構をより精緻化するために，他の二種類の既存手法を拡張利用することで知識断片分類性能を精緻化する手法を開発した．小規模な実験を実施し，その効果と課題を確認した．

(5) 常識判断および感情情報処理を利用した知識洗練については，予備実験から効果が限定的であるという結論を得た．このことから，知識断片獲得モジュールのランキング性能が十分有効に機能していることが示唆された．

(7) 可視化インターフェイスモジュールとして，以下の三種類の機構を実装した．

① 獲得した知識断片をスコアによってソートし，その重要度をグラフで表示するインターフェイスを実装した(図 3)．
②提示された知識断片を介して関連する説明文を参照できるインターフェイスを実装した（図 3)．
③ クエリと知識断片を色付きの円で表示し，重要度を円の直径で表現するインターフェイスを実装した．

(8) 本研究で実装したシステムの一部を利用した応用研究にも着手した．

① 情報検索に対する比喩的素描手法を応用した検索クエリ拡張手法を提案し有効性を確認した．掲示板における不適切表現の判別手法を用いて知識断片のノイズ除去について検討したが，(1) と同様の理由で効果はかなり限定されることがわかった．
② 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．

(9)モジュールを結合し，その動作検証と性能評価を行った結果，バックエンドで用いる検索エンジンを複数用いることで知識断片候補推定の性能が向上することがわかった．また，開発したシステムが，語彙知識の把握に対する有効性について定性的に評価を実施し，ユーザに対する未知語については有効であるという結論も得た．

(10) 今後の展望として，以下に示すように，開発システムの精緻化に関するものと，開発システムあるいはその一部を応用した新たな研究の展開が考えられる．

① 獲得した知識断片集合の分類モジュールおよび洗練モジュールについて，分類精度の精緻化および詳細化，洗練モジュールについては，ノイズ除去性能の高度化などへの取組みが必要である．
② 本システムを解析エンジンとした比喩的連想に基づく柔軟なクエリ語拡張手法を発展させることにより，うろ覚えなどクエリ語を明示できない場合のクエリ語顕在化支援技術の提案が可能であると思われる．
③ 本システムのログ情報を利用して，時系列による知識断片の変化を提示する手法，共有される知識断片を利用することにより従来とは異なる観点から語の類似性を判定する手法の提案が可能であると思われる．

５．主な発表論文等
（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. 桝井文人，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）, Vol. 21, No. 6, pp.707-719, 2010.

〔学会発表〕（計 25 件）

1. 岩城秀則, 桝井文人: “WWW から獲得した語の比喩的素描表現の上位下位関係精緻化に関する一考察”, 言語処理学会第 18 回年次大会, 2012.03.15, 広島市
2. 久保真哉，桝井文人，福本淳一: “比喩的関係を利用した検索クエリ拡張手法”, 言語処理学会第 18 回年次大会, 2012.03.15, 広島市
3. 長谷川恭佑，桝井文人，後藤文太朗: “比喩的素描を用いた類似語推論およびその視覚化インターフェイスの構築”, 情報処理学会第 67 回全国大会, 2012.03.07, 名古屋市
4. 久保真哉，桝井文人，福本淳一: “喩える関係を利用した検索クエリ拡張に関する一考察”, 人工知能学会情報編纂研究会 第 5 回研究会, 2011.07.01, 東京都
5. 松葉達明，桝井文人，河合敦夫，井須尚紀: “学校非公式サイトにおける有害情報検出を目的とした極性判定モデルに関する研究”, 言語処理学会第 17 回年次大会, 2011.03.08, 豊橋市
6. 桝井文人，久保真哉，福本淳一: “比喩表現による検索手法の構想”, 人工知能学会情報編纂研究会, 2010.09.05, 草津市
7. Fumito MASUI, Rafal RZEPKA, Yasutomo 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
8. 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
9. Michal PTASZYNSKI, Pawel DYBALA, Tasuaki MATSUBA, Fumito MASUI, Rafal RZEPKA and Kenji ARAKI: “Machine Learning and Affect Analysis Against Cyber-Bullying”, LaCATODA2010, 2010.03.29, Leicester, England
10. 小島鉄也, 桝井文人, 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
11. RZEPKA Rafal, 桝井文人，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
12. 桝井文人，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
13. Fumito MASUI, Yoshifumi KAWAMURA, Jun’uchi FUKUMOTO and Naoki ISU: “MURASAKI: WWW-based Word Sense Description System”, International Conference ITC-CSCC2008, 2008.07.07, Shimonoseki, Japan.

〔その他〕
ホームページ等
プロトタイプシステム”Murasaki”
http://orion.cs.kitami-it.ac.jp/
（ユーザ登録制による公開）

６．研究組織
(1)研究代表者
桝井 文人（ MASUI Fumito ）
北見工業大学・工学部・准教授
研究者番号：80324549

(2)研究分担者
ジェプカ・ラファウ（ Rzepka Rafal ）
北海道大学大学院・情報科学研究科・助教
研究者番号：80396316

(3)研究分担者
木村 泰知（ KIMURA Yasutomo ）
小樽商科大学・商学部・准教授
研究者番号：50400073

(4)研究協力者
荒木 建治（ ARAKI Kenji ）
北海道大学大学院・情報科学研究科・教授
研究者番号：50202742

(5)研究協力者
プタシンスキ・ミハウ（PTASZYNSKI Michal）
北海学園大学・工学部・研究員
研究者番号：なし