

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 5月22日現在

機関番号：17102

研究種目：基盤研究（C）

研究期間：2009～2011

課題番号：21500102

研究課題名（和文） 個別化 Web 検索支援システムの開発に関する研究

研究課題名（英文） Study on Developing Personalized Web Retrieval Support Systems

研究代表者

越村 三幸（KOSHIMURA MIYUKI）

九州大学・システム情報科学研究院・助教

研究者番号：30274492

研究成果の概要（和文）：ユーザの意図に適合した検索を行う Web 検索支援システムの構築を目指して、ソーシャルブックマークや ODP(Open Directory Project), Twitter, ブログ, Web 閲覧履歴を用いたユーザの特徴抽出, 及び, Wikipedia を知識源とした関連語抽出を行った。抽出した特徴に基づいたユーザ類似性に着目した情報推薦システムを試作した。また, 時間的な距離に着目した関連単語の抽出を行い, タイムリーな検索支援に有効であることを確かめた。

研究成果の概要（英文）：We conducted the research on extracting user's profiles using social bookmarks, ODP(Open Directory Project), Twitter, blogs and browsing history, and related word extraction from Wikipedia in order to develop Web retrieval support systems considering his/her intent. We developed several Web recommender systems using several similarities between users based on the extracted profiles. We also examined a temporal distance between words, extracted related words based on the distance, and showed the effectiveness of the words for a timely retrieval.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009年度	1,400,000	420,000	1,820,000
2010年度	1,000,000	300,000	1,300,000
2011年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：計算機科学

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：クエリ拡張, ODP, ユーザプロファイル

1. 研究開始当初の背景

(1) Web 空間においては日々ページ数が増加しており, 情報過多あるいは情報氾濫とも言われるほど多くのページが存在するようになった。この中から欲しい情報を探し出す際に多くの人々が利用するシステムが Web 検索エンジンである。しかし, 検索エンジンを利用してユーザが自分自身の検索要求に

合致するページを得られないことも度々ある。多くの場合, ユーザは所望の検索結果が得られるようにクエリに単語を追加(クエリ拡張), あるいはクエリの変更を試みるが, うまくいかないことも多い。

(2) 研究代表者・分担者は関連単語抽出アルゴリズムを考案し, 検索の結果, 得られた Web ページからクエリに関連する語を抽出

し、これを拡張語とする手法を提案し、ユーザの興味のない多くの Web ページを排除することに成功していた。しかし、場合によっては偏った関連語しか抽出できず、検索精度がかえって低下するという問題があり、その原因を分析し、偏りを正す方法の確立が求められていた。

2. 研究の目的

本研究の目的は、Web 検索エンジンに与えるクエリに含まれる語として、ユーザの意図する適切な語を生成し追加することで Web 検索結果の精度を高める Web 検索支援システムを開発することである。

本研究では、代表者及び分担者が提案した関連単語抽出アルゴリズムを核として研究を進める。次のような問題点に対処するために、研究課題を設定する。

- (1) ユーザの負荷軽減：検索精度をあげるには、ユーザからの明示的フィードバックが必要でユーザの負荷は大きい。ユーザの意思や意図、あるいは嗜好を表現するユーザプロフィールを予め作成しておき、これを擬似フィードバックに用いることにより、ユーザの負荷なしで検索精度の向上を目指す（研究課題 1）。ユーザの様々な興味を一つのプロフィールで表現するよりは、複数のプロフィールで表現し、ユーザの状況に応じてプロフィールを自動的に使い分ける仕組みを考案する（研究課題 2）。
- (2) より適切な関連語抽出の実現：オリジナルの関連単語抽出アルゴリズムでは、単語間の文の個数を距離として用いるが、この他にも単語の個数や段落の個数など、様々な距離の与え方が考えられる。距離の与え方が、精度に与える影響を明らかにし、より適切な関連単語の抽出を目指す（研究課題 3）。また、抽出語の自動分類を行い、検索精度の向上を図る（研究課題 4）。さらに、アルゴリズムを適用する文書の拡大を図る（研究課題 5）。

3. 研究の方法

- (1) 研究課題ごとに研究テーマを設定し、各研究テーマに主担当者を割当てた。
- (2) 力（分担者・徳山高専）は、月 1 回程度、九州大学を訪問し、成果発表・研究交流・情報交換を行った。また、最終年度は、内地研究員として、九州大学に常駐したので、密に情報交換ができた。

4. 研究成果

- (1) 研究発表：15 編の雑誌論文（査読有の国際会議論文を含む）、19 件の学会発表を行った。

- (2) 推薦理由を提示する情報推薦システムの試作を帰納論理プログラミングによって行い、明確かつ定性的な推薦理由を提示できることを確かめた。

- (3) Wikipedia の「リンク構造」「カテゴリ」「編集者情報」などを素性とし、ランキング学習により関連語を抽出する手法を実現し、人間の感覚により近い関連語の抽出が可能であることを確かめた。

- (4) ソーシャルブックマークに付与されているタグに着目して、ユーザ間の関連度を抽出し、これを Web ページ推薦に用いる手法を提案した。

- (5) Twitter のリスト構造からユーザの特徴を抽出し、それを基にフォローするユーザを推薦するシステムを作成した。

- (6) リアルタイム性のある Twitter のツイートを利用し、関連単語抽出アルゴリズムにおける単語間距離をツイート間の時間的距離に置き換えることにより、リアル性の高い関連語の抽出を行った。

- (7) マルチエージェントモデルに基づき、ユーザの意図を汲み取る検索モデルの提案を行い、その効果を検証した。

- (8) Web ページの検索を協力者の検索に置き換え、スマートグリッドの核となる電力制御手法であるデマンドレスポンスを実現する手法の提案を行った。

- (9) 学生奨励賞（情報処理学会第 150 回データベースシステム研究会）対象論文「Twitter 上の流行語先取り発言者検出システムの開発」

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 15 件）

- 1) Y. Takiguchi, K. Kurakado, T. Oishi, M. Koshimura, H. Fujita, R. Hasegawa, Evaluating Reranking Methods based on Link Co-occurrence and Category in Wikipedia, ICAART 2012, 査読有, Vol.1, pp. 277-282, 2012.
- 2) K. Takahashi, T. Matsuzaki, T. Mine, T. Kawamura, K. Sugahara, Protection of Personal Information based on User Preference, IJNCAA, 査読有, 1(3), pp. 820-832, 2011.
- 4) H. Yu, T. Mine, M. Amamiya, Towards User Intent Based Searching, IEEE TrustCom-11, 査読有, pp. 1400-1407, 2011.
- 5) 力 規晃, 越村 三幸, 藤田 博, 長谷川 隆三, 推薦理由を考慮した情報推薦システム, システム情報科学紀要, 査読有, 16(2), pp. 75-81, 2011.
- 6) 小林 寛武, 峯 恒憲, ユーザフィードバ

ックとクエリ学習手法の複数コミュニティ上での評価, 人工知能学会論文誌, 査読有, 26 巻, 1 号, pp.97-106, 2011.

7) K. Kurakado, T. Oishi, R. Hasegawa, H. Fujita, M. Koshimura, Evaluating Reranking Methods Using Wikipedia Features, ICAART 2011, 査読有, Vol.1, pp.376-381, 2011.

8) W. Shirakihara, T. Oishi, R. Hasegawa, H. Fujita, M. Koshimura, Trendspotter Detection System for Twitter, ICAART 2011, 査読有, Vol.1, pp.625-628, 2011.

9) N. Chikara, M. Koshimura, H. Fujita, R. Hasegawa, Rule Extraction from Blog Using Inductive Logic Programming, WebPRES 2010, 査読有, pp.269-272, 2010.

10) K. Hori, T. Oishi, T. Mine, R. Hasegawa, H. Fujita, M. Koshimura, Related Word Extraction from Wikipedia for WEB Retrieval Assistance, ICAART 2010, 査読有, Vol.2, pp.192-199, 2010.

11) T. Oishi, T. Mine, R. Hasegawa, H. Fujita, M. Koshimura, ACCDS 2009, 査読有, pp.41-56, 2009.

12) T. Mine, H. Kobayashi, Applying user feedback and query learning methods to multiple communities, PRIMA 2009, 査読有, pp.276-291, 2009.

13) H. Yu, T. Mine, M. Amamiya, Semantic ACP2P Information Retrieval Method, SSIRI 2009, 査読有, pp.297-302, 2009.

14) T. Mine, A. Kogo, S. Amamiya, M. Amamiya, Refinement of the ACP2P by Sharing User-Feedbacks and Learning Query-Responder-Agent Relationships, AAMAS 2009, 査読有, pp.1341-1342.

〔学会発表〕(計19件)

1) 白木原 渉, 大石 哲也, 越村 三幸, 藤田 博, 長谷川 隆三, 時間的距離に着目した Twitter からの関連単語抽出, 情報処理学会第105回 IFAT 研究発表会, 2012年1月21日, 福岡大学.

2) 田代 祐一, 大石 哲也, 越村 三幸, 藤田 博, 長谷川 隆三, ソーシャルブックマークにおける有用なユーザの発見, 情報処理学会第151回 DBS 研究会, 2010年11月13日, 早稲田大学.

3) 力 規晃, 越村 三幸, 藤田 博, 長谷川 隆三, 帰納学習を用いたブログからのルール抽出, 平成 22 年 電気学会 電子・情報・システム部門大会, 2010年9月3日, 熊本大学.

4) 倉門 浩二, 大石 哲也, 長谷川 隆三, 藤田 博, 越村 三幸, Wikipedia のリンク共起とカテゴリに基づくリランキング手法, 情報処理学会第99回 IFAT 研究会, 2010年8月4

日, 青山学院大学.

5) 白木原 渉, 大石 哲也, 長谷川 隆三, 藤田 博, 越村 三幸, Twitter 上の流行語先取り発言者検出システムの開発, 情報処理学会第150回 DBS 研究会, 2010年8月4日, 青山学院大学.

6) 峯 恒憲, 小林 寛武, P2P 型情報検索のためのユーザフィードバックの活用とクエリ送信先学習法の評価, 電子情報通信学会「人工知能と知識処理研究会」, 2010年6月25日, 関西学院大学東京キャンパス.

7) 白木原 渉, 倉門 浩二, 大石 哲也, 越村 三幸, 藤田 博, 長谷川 隆三, Twitter の流行語発言者の抽出に基づくフォロワー推薦システムの開発, 情報処理学会(第72回)全国大会, 2010年3月11日, 東京大学.

8) 力 規晃, 越村 三幸, 藤田 博, 長谷川 隆三, 帰納論理プログラミングを用いたブログからのルール抽出, 人工知能学会(第87回)知識ベース研究会, 2010年1月29日, NTT 武蔵野研究開発センター.

9) 中村 徹, 大石 哲也, 越村 三幸, 藤田 博, 長谷川 隆三, ODP を利用したユーザプロフィール作成と個別化検索システムの評価, 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」, 2009年7月5日, 広島市立大学.

10) 倉門 浩二, 大石 哲也, 長谷川 隆三, 藤田 博, 越村 三幸, 関連単語抽出アルゴリズムの改良と評価, 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」, 2009年7月5日, 広島市立大学.

11) 田代 祐一, 大石 哲也, 力 規晃, 越村 三幸, 藤田 博, 長谷川 隆三, 帰納論理プログラミングを用いた Web ページ評価ルールの抽出とその評価, 電子情報通信学会「人工知能と知識処理研究会」, 2009年5月15日, 関西学院大学東京丸の内キャンパス.

〔図書〕(計1件)

峯 恒憲, 「知識ベース」7群7編2章5節「連邦アーキテクチャ」, 同群同編3章12節「仲介(メディエーション)」を担当, 2011.

〔産業財産権〕

○出願状況(計0件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

○取得状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

[その他]

6. 研究組織

(1) 研究代表者

越村 三幸 (KOSHIMURA MIYUKI)
九州大学・システム情報科学研究院・助教
研究者番号：30274492

(2) 研究分担者

長谷川 隆三 (HASEGAWA RYUZO)
九州大学・システム情報科学研究院・教授
研究者番号：20274483

藤田 博 (FUJITA HIROSHI)
九州大学・システム情報科学研究院・准教授
研究者番号：70284552

峯 恒憲 (MINE TSUNENORI)
九州大学・システム情報科学研究院・准教授
研究者番号：30243851

力 規晃 (CHIKARA NORIAKI)
徳山工業高等専門学校・情報電子工学科・助手
研究者番号：50290804