

平成 1 7 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 基盤研究(A) 4. 研究期間 平成15年度 ~ 平成18年度
5. 課 題 番 号 1 5 2 0 0 0 1 0
6. 研究課題名 セマンティックWebのための多言語処理

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
0 0 2 0 3 4 8 0	フリガナ ウヱムラ, シュンスケ 植村, 俊亮	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
4 0 2 9 3 3 9 4	フリガナ ミヤザキ, シュン 宮崎, 純	情報科学研究科	助教授
8 0 3 1 4 5 3 2	フリガナ ハタノ, ケンジ 波多野, 賢治	情報科学研究科	助手
2 0 3 5 1 3 2 2	フリガナ マエダ, アキラ 前田, 亮	立命館大学・情報理工学部	助教授
4 0 3 8 8 1 1 1	フリガナ スズキ, ユウ 鈴木, 優	立命館大学・情報理工学部	助手
7 0 3 1 4 5 3 1	フリガナ アマガサ, トシユキ 天笠, 俊之	筑波大学・計算科学研究センター	講師

9. 研究実績の概要(国立情報学研究所でデータベース化するため、600字～800字で記入。図、グラフ等は記載しないこと。)

本年度は、Yahoo に代表される複数の言語版のある Web ディレクトリを利用し、その各カテゴリに登録されている Web ページの中から対訳関係にある単語を自動的に抽出し、検索語を補間する方式を実装し評価した。一般に対訳関係は一つ対一つではなく、一つ対多となる。一つ対多の関係の場合、一つの検索語に対してどの程度まで対訳語を補間すれば検索効率が向上するかという問題に関して、対訳語数による切捨法、対訳語の単語重みによる切捨法を提案し、それぞれの比較検討を行なった。その結果、対訳語数が 2 の場合に最も検索精度が向上することが分かったが、3 以上の場合に急激に精度が落ちる問題点も明らかになった。一方、対訳語の単語重みによる切捨法に関しては、二通りの手法が考えられ、固定の単語重みで切り捨てる方法と、訳語の最大の重み値に対して相対的な重みで切り捨てる方法を比較した。前者は、わずかな閾値の差により急激に精度が落ちる可能性があるが、後者は、閾値の設定において、急激な精度の悪化はなく、最大の重みの訳語の 15 分の 1 程度の重みを持つ語までを検索語として使用すればよいことが分かった。

以上に加えて、単に各文書ごとの検索だけでなく、同一著者の同一内容の文書をクラスタリングすることにより、ユーザが要求する文書を効率よく検索する方法についても検討を行なった。通常、Web ページでは、各ページの著者には詳細・非詳細のコンテンツがあり、単なる PageRank アルゴリズムの適用では、ページのクオリティに関係なくランキングされてしまい、ユーザが望むページがランキング順に上位にくるとは限らない。しかし、各著者、各話題ごとにページをクラスタリングし、クラスタに対して PageRank を適用することにより、ユーザが望むページを含むクラスタが上位にランクされることが期待できる。この手法を、NTCIR-4 NW100G-01 のテストコレクションを用いて評価したところ、多値適合レベルの精度評価指標 WRR ならびに再現率評価指標 DCG の両方において、既存のページ単位の PageRank よりも性能が良いことを示した。本手法は、クラスタリングに特徴ベクトルモデルを利用し統計的に処理しているが、セマンティック Web を利用して、精緻なクラスタリングを行えば、さらに精度のよい Web ページの検索が可能となると考えられる。

成果の公表を見合わせる必要がある場合は、その理由及び差し控え期間等を記入した調書(A4判縦長横書1枚)を添付すること。

10. キーワード

- | | | |
|----------|-----------|-----------|
| (1)多言語処理 | (2)言語横断検索 | (3)データベース |
| (4)情報検索 | (5)ディレクトリ | (6) |
| (7) | (8) | |
- (裏面に続く)

11. 研究発表(平成17年度の研究成果)

〔雑誌論文〕 計(6)件

著者名	論文標 題		
Fuminori Kimura, Akira Maeda, Jun Miyazaki, Shunsuke Uemura	Query Disambiguation for Cross-Language Information Retrieval Using Web Directories		
雑誌名	巻・号	発行年	ページ
Proc. of International Workshop on Challenges in Web Information Retrieval and Integration (WIRI 2005)		2 0 0 5	

著者名	論文標 題		
Kenji Hatano, Hiroko Kinutani, Toshiyuki Amagasa, Yasuhiro Mori, Masatoshi Yoshikawa, Shunsuke Uemura	Analyzing the Properties of XML Fragments Decomposed from the INEX Document Collection		
雑誌名	巻・号	発行年	ページ
Advances in XML Information Retrieval, Lecture Notes in Computer Science	3493	2 0 0 5	168-182

著者名	論文標 題		
Yu Suzuki, Kenji Hatano, Masatoshi Yoshikawa, Shunsuke Uemura	A Normalization Method of Relevance Scores for Combining Ranked Result Lists		
雑誌名	巻・号	発行年	ページ
Proc. of the 8th International Conference on Asian Digital Libraries (ICADL2005)		2 0 0 5	

著者名	論文標 題		
波多野賢治, 絹谷弘子, 吉川正俊, 植村俊亮	XML文書検索システムにおける文書内容の統計量を利用した検索対象部分文書の決定		
雑誌名	巻・号	発行年	ページ
電子情報通信学会論文誌	J89-D-1・3	2 0 0 6	422-431

著者名	論文標 題		
中窪仁, 中島伸介, 波多野賢治, 宮崎純, 植村俊亮	文書集合単位のリンク解析を用いたWebページスコアリング		
雑誌名	巻・号	発行年	ページ
電子情報通信学会第17回データ工学ワークショップ論文集		2 0 0 6	3B-i4

著者名	論文標 題		
中尾 伸章, 天笠 俊之, 的野 晃整, 植村 俊亮	アクセス頻度を考慮したXML文書分割方式の提案		
雑誌名	巻・号	発行年	ページ
日本データベース学会 Letters	4・1	2 0 0 5	113-116

〔図 書〕 計(0)件

著者名	出 版 社		
書 名	発行年	総ページ数	

12. 研究成果による工業所有権の出願・取得状況

計(0)件

工業所有権の名称	発明者	権利者	工業所有権の種類、番号	出願年月日	取得年月日