

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 4 日現在

機関番号：34310

研究種目：挑戦的萌芽研究

研究期間：2013～2014

課題番号：25540150

研究課題名(和文)レファレンスデータを利用した検索システム利用者の検索行動補助機構に関する研究

研究課題名(英文)Information Seeking Behavior Suggestion System using Collaborative Reference Database

研究代表者

波多野 賢治 (Hatano, Kenji)

同志社大学・文化情報学部・准教授

研究者番号：80314532

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：本研究では、既に文書形式で記録として存在している利用者の検索システムを用いた手順を活用した利用者の検索行動の補助機構の提案を行った。具体的には、需要の多い質問に対してあらかじめ必要な資料を準備・作成し、文書として保存されたものから利用者の情報行為を抽出、知識ベース化することで、利用者の情報要求に対し、どのような過程を経て、情報の発見に至ったかを示すことができるようになる。

最も意義があった研究成果は、文書に記録として残されている検索行動から情報行為を抽出する際は、階層的な係り受け解析を行うことでトピック推定を正確に行うことができるようになり、抽出される行為の正確性が向上した点である。

研究成果の概要(英文)：In this research, we propose a method for suggesting information seeking behavior to search engine users using Collaborative Reference Database operated by National Diet Library (NDL). The reason why we do not utilize access log of search engines is that Collaborative Reference Database includes IT specialist's searching procedures to get useful information, so that it is useful for the search engine users to show them, we think. As a result, the search engine users can learn how to access such the useful information to emulate the capabilities of the IT specialist without their knowledge.

One of the most impressive research outcome is our novel dependency parsing method based on hierarchical structure of Japanese Language. It can help to precisely extract information seeking behaviors from Collaborative Reference Database compared with conventional approaches.

研究分野：情報検索

キーワード：検索行動補助 ルールマイニング 知識ベース

1. 研究開始当初の背景

World Wide Web (WWW) の発展は社会に爆発的な量のデータを提供しており、情報化社会において必要不可欠な存在となっている。WWW 上にデータ量が増えれば増えるほど、社会にとって必要な情報を効果的に検索することが困難となるため、通常はそうした情報を効率よく検索するために Google や Yahoo! といった検索エンジンを利用することになる。

こうした利用者の要求に応えるべく、検索エンジンを支える要素技術も年々変化を見せている。例えば、検索エンジンの検索結果には、利用者の情報要求に合致した Web 文書であることが認識し易いように要約文の一種であるスニペットが含まれているが、利用者の検索傾向を把握し良質なスニペットを生成するための研究は既に始まっている。また他にも、情報技術を使い慣れていない利用者のために、検索システムへアクセスした時間、アクセス元 IP アドレス、アクセス先 URL、検索キーワードといった検索システムに蓄えられるアクセスログを用いて、利用者が入力した検索キーワードに対して自動的に検索質問拡張を行ったり、関連語を利用した新しい質問を提示したりする研究も行われている。

しかしながら、これらの研究はすべて利用者の念頭にある検索要求に対して、容易にキーワードを思い出すような利用者を対象としており、検索要求をキーワードに具現化できない利用者を対象としてこなかったという根本的な問題が残されている。

2. 研究の目的

本研究の目的は、前述した問題に対し、検索システムの検索行動補助機構を実装する際にこれまで多くの研究者が取り組んできた検索システムのアクセスログを用いるのではなく、レファレンス記録と呼ばれる需要の多い質問に対してあらかじめ人が必要な資料を準備・作成し、記録として保存されたデータを解析することで、利用者が抱えている情報要求に対し、どのような過程を経て情報の発見に至ったかを示す検索システムの検索行動補助機構を提案することにある。

この補助機構は、検索初心者にありがちな「何から検索してよいかわからない」といった状況を回避することを目的としており、この補助機構を使い続けることで、検索初心者はどのような過程を経て未知の情報に到達すればよいのかその指針を理解することができるようになる。その結果、最終的には提案する補助機構により、情報弱者を生む一つの原因となっている情報技術を活用できない状態の利用者を減少させることにつなげることを目的としている。

3. 研究の方法

本研究の斬新性は、検索システムの検索行

動補助機構を実装する際に、これまで多くの研究者が取り組んできた検索システムのアクセスログを用いるのではなく、レファレンス記録を用いる点にある。

レファレンス記録とは、需要の多い質問に対してあらかじめ必要な資料を人が準備・作成し、その使用履歴等を記録として保存されているものを指し、主に図書館において利用者が学習・研究・調査を目的として必要な情報・資料などを求めた際に利活用されるものである。つまり、レファレンス記録そのものを人の知識と考えることができ、この知識を利活用することこそが本研究の斬新性に他ならない。

そもそも情報と知識は互いに関連語ではあるが、前者が利用者に対して何らかの影響を与えているものを指しているのに対し、後者はデータや情報を体系的にまとめ将来の判断や思考に活用できるものを指している。そういう意味で、これまでの研究で盛んに行われてきた研究における検索システムのアクセスログ利用は、その利用者にとっての情報を活用しようとした研究と言え、知識を活用しようとしている本研究とは内容が大きく異なる。

またいわゆる「情報ピラミッド」に示されているように、知恵が物事の理を悟り適切に処理する能力であると考ええると、人間は知恵を利用してデータや情報などから知識を獲得している。これはつまり、知恵や知識がなければデータや情報を活用できないことを示唆しており、特に研究目的で述べた念頭にある検索要求をキーワードに具現化できないような利用者に対しては、検索行動の初期段階のうちには利用者に影響を与えている可能性がある情報を検索させる行為を補助するよりは、知識自体を提示して利用者にそれを模倣させるほうが「何をしてよいかわからない」という検索初心者にとっては有益であることを意味している。なぜなら、人間が知識や技能を獲得するプロセスには様々な形態が考えられるが、「学ぶ(まなぶ)」という言葉と「まねぶ」が同源である事が言われているように、模倣を通じて学習することは、知識獲得、技能獲得において本質的役割を担っているためである。

その一方で、知識を有しない計算機を使用して、データや情報から知識を発見することは非常に難しいという意味で、本研究は非常にチャレンジングなテーマであるともいえる。しかしビジネスの世界においては、現場の問題などの早期発見・効率化・改善にビジネスフローの可視化が行われているように、レファレンス記録自体を一種の問題解決のための過程と考え、その過程を利用者に提示することは、「何をしてよいかわからない」という利用者自身の現状や自分自身が抱えている問題の認識に利用することができる。そのため、提示された問題解決の過程を利用者が模倣することで、問題からの脱却の第一

歩につながる可能性があることを考えると、本研究を実施することは情報技術を活用できない利用者を減少させることに大きく貢献することは間違いない。

以上のように、これまで行われてきた研究のトレンドを追わず、敢えて計算機で扱うことが難しいレファレンス記録を用いた知識抽出とその提示を行おうとする新たな観点こそが、本研究の斬新でチャレンジ性に富んだといえる点である。

4. 研究成果

本研究の最も大きな成果は、検索行動補助機構の構築そのものではなく、その前処理段階におけるレファレンス記録からの行動抽出の際に行われる処理である。

具体的には、レファレンス記録から、いつ、誰が、何を、どうしたかという人間の検索行動を抽出する処理があるが、この処理を行う際に必要となる日本語構文解析器では、特に一文が長くなる場合に誤解析が生じる場合があり、うまく人間の検索行動を抽出することができないという問題が生じていた。

そこで、我々の研究では、これまでの日本語構文解析器が、基本的にある二つの形態素間の共起関係を用いて係り受け関係を統計的学習の結果を用いて抽出してきたが、複数の形態素間の共起関係から導かれる階層的関係を用いて係り受け関係を抽出する手法を提案することで、特に長文の構文解析時の誤解析の発生を抑制できるように改善できた点は大きな成果であると言える。その結果、発表した国内外の会議では平成 25 年度情報処理学会関西支部大会学生奨励賞と IARIA ICDS 2014 Best Paper Award を授与され、行ってきた研究成果が高く認められたと言える。

しかしながら、本研究の目的であった検索行動補助機構の構築までは出来なかった点は大きな問題であるように考えている。もちろん、構築出来なかったのではなく補助機構の構築までは実施できたが、補助機構により推薦された検索行動をただ真似るだけでは、検索初心者が実際に欲しい情報を得ることができなかった。

この問題点の解決は、人間の行動問題の分析と修正を目的として発達してきた応用行動分析 (Applied Behavior Analysis: ABA) 分野では盛んに研究されているという報告がなされている。ただ、ABA を適用する際には人間の行動の汎化が必要であり、その汎化を目指すとなると、人間の検索行動の知識ベース化が必須となると思われる。そのため、人間の検索行動の知識ベース化をする際に、その質的保証をどのように行うのか、また多くの検索行動から知識ベースを構築する際の手間をどのように軽減していくのかについて、今後、考えて行かねばならないと思われる。

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1件)

- Kazuki Ono and Kenji Hatano: "A Dependency Parsing and Its Application using Hierarchical Structure in Japanese Language", International Journal on Advances in Internet Technology, 査読有, Vol.7, No.3 and 4, pp.193-204, December 2014.

[学会発表](計10件)

- 上月 悠也, 波多野 賢治: "心理学定義に基づいた感情辞書構築とその評価", 情報処理学会第 77 回全国大会, 京都, 3N-3, March 18th, 2015 .
- 眞田 和枝, 塩井 隆円, 波多野 賢治: "概念辞書グラフの経路に基づく語の直上概念決定法", 情報処理学会第 77 回全国大会, 京都, 3N-6, March 18th, 2015 .
- Qian Meng and Kenji Hatano, "Visualizing Basic Words Chosen by Latent Dirichlet Allocation for Serendipitous Recommendation", Proceedings of 2014 3rd IIAI International Conference on Advanced Applied Informatics (IIAI-AAI 2014), 福岡, pp.819-824, September 3rd, 2014.
- Kazuki Hagiwara, Kazuki Ono, and Kenji Hatano, "Extracting Experiences Using Dependency Parsing on Japanese E-Commerce Websites", Proceedings of 2014 3rd IIAI International Conference on Advanced Applied Informatics (IIAI-AAI 2014), 福岡, pp.813-818, September 3rd, 2014.
- Kazuki Ono and Kenji Hatano: "A Dependency Parsing Method Based on the Hierarchical Structure in Japanese Language", Proceedings of The Eighth International Conference on Digital Society (ICDS 2014), バルセロナ (スペイン), pp.203-208, March 27th, 2014.
- 萩原 一貴, 大野 一樹, 波多野 賢治: "品詞間の係り受けに着目した体験情報抽出手法の提案", DEIM 2014 Proceedings, 兵庫, E6-6, March 4th, 2014 .
- 大野 一樹, 波多野 賢治: "係り受け関係の階層化に基づいた構文木モデルを利用したトピック推定手法の提案", DEIM 2014 Proceedings, 兵庫, C3-5, March 3rd, 2014 .
- 大野 一樹, 波多野 賢治: "係り受け関係の階層化とその共起に基づいた構文木モデルを利用した構文解析手法の提案", 情報処理学会研究報告, 鹿児島, Vol.2013-NL-214, No.6, pp.1-6, November 14th, 2013.
- 大野 一樹, 波多野 賢治: "係り受け関係の階層化に基づいた構文木モデルによる

る構文解析手法の提案", 情報処理学会
関西支部 2013 年度支部大会, 大阪, F-03,
September 25th, 2013.

- Hironao Yomegane, Saori Kitahara, and
Kenji Hatano, ``Extracting Information
Seeking Behavior from The Collaborative
Reference Database'', Proceedings of 2013
Second IIAI International Conference on
Advanced Applied Informatics (IIAI-AAI
2013), 島根, pp.301-305, September 4th,
2013.

〔図書〕(計 0 件)

〔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 :
出願年月日 :
取得年月日 :
国内外の別 :

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

波多野 賢治 (HATANO, KENJI)

同志社大学・文化情報学部・准教授

研究者番号 : 80314532

(2) 研究分担者

(3) 連携研究者