

研究種目：若手研究(B)

研究期間：2008～2009

課題番号：20700128

研究課題名(和文) 情報探索支援を目的としたハイブリッド情報資源オントロジー

研究課題名(英文) An Hybrid Ontology of Information Resources for Information Retrieval

研究代表者

清田 陽司 (KIYOTA YOJI)

東京大学・情報基盤センター・助教

研究者番号：10401316

研究成果の概要(和文)：情報要求の具体化を支援する情報探索システムの実現にあたって必要となる技術を開発し、図書館の現場で実運用評価した。被験者のうち78%が有用なヒントを得られるという結果を得た。

研究成果の概要(英文)：This study developed basic methods for an information navigation system which supports clarification of information needs, and evaluated the methods at libraries. The result showed that 78% of subjects could get useful hints for information retrieval.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,400,000	420,000	1,820,000
2009年度	1,300,000	390,000	1,690,000
年度			
年度			
年度			
総計	2,700,000	810,000	3,510,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：情報検索システム、Web情報検索、情報図書館学、質問応答システム、自然言語処理、ユーザインタフェース、オントロジー工学

1. 研究開始当初の背景

人が情報探索を行うプロセスは、人の持っている情報要求が徐々に具体化するプロセスとしてとらえることができる。このプロセスをTaylorは以下の4階層に分類している：(Q1)直感的要求(visceral need)：漠然と知識が欠けていることを認識しているが、どんな知識が欠けているかはわからない
(Q2)意識された要求(conscious need)：どんな知識が欠けているかはほぼわかるが、どん

な情報が必要かは明確に説明できない

(Q3)形式化された要求(formalized need)：欠けている知識が明確にわかり論理的な質問ができる

(Q4)調整済みの要求(compromised need)：検索を行う対象(質問を投げかける相手や検索システム)の知識構造や理解力を想定し、相手の制約を認識して検索を行うことができる

一方、情報探索を工学的アプローチによつ

て研究する立場からは、さまざまな情報検索システムが提案されてきているが、ほとんどの研究は Taylor の階層における Q4 を前提としており、Q1～Q3 は対象とされていない。テキスト検索システムに関する評価型ワークショップである TREC などでは共通の検索課題を設定するが、評価の客観性を確保する必要性から、課題として設定される検索質問は必然的に Q4 のレベルのものとなる。また、ユーザインタフェースの研究分野では Q3 から Q4 へのプロセスが扱われている。例えば、利用者の入力した質問キーワードに関連するキーワードの提示についての研究などがある。

しかし、Q1～Q3 のプロセスをコンピュータシステムで扱う研究はほとんど存在しない。その大きな理由としては、情報要求が具体化するプロセスを形式化するのが困難だということ、プロセスを客観的に評価する方法論が確立していないことがあげられる。しかし、人が必要な情報を得られないのはそもそも Q1～Q3 の階層から Q4 に具体化できないのが原因となっていることが多い。

では、必要な情報を効率的に見つけられる人はどのようにして「情報要求の具体化」を行っているのだろうか？ この命題を考える際に、文献探索の専門家である図書館員が行っている情報探索支援行動が参考になる。一般に、図書館のレファレンスサービスにおける質問回答の際、図書館員は利用者の情報要求を「主題－探索事項」の形式に当てはめて分析しているといわれている。例えば、「杜若の読み方を教えてください」という質問では、主題が「杜若」、探索事項が「読み方」となる。図書館員は、利用者の調べたい主題や探索事項が不明確な場合にさまざまな角度からインタビューを行い、徐々に利用者の情報要求を明確にしていく。さらに主題や探索事項に関連するさまざまな情報資源に関する豊富な知識を活用することによって、利用者に適切な資料を提示することができる。

図書館員が行っている情報探索支援をシステムで模倣する先駆的試みとしては、京都大学において開発されたレファレンスサービスシステムが存在する。このシステムは、図書館目録の件名標目表を用いて利用者質問文から書架分類(経済、政治、工学など)を導出することで、図書館利用者の情報要求のうち「主題」の推定をある程度実現している。しかし、件名標目表に含まれている件名は数万件にすぎないため、利用者がもつ多様な情報要求に対応することは難しかった。

2. 研究の目的

(1) 図書館員が利用者の情報要求を具体化するために行っている「主題－探索事項」の枠組みに沿った質問分析をシステムで模倣す

るために必要となる知識構造を明らかにする。具体的には、Web や図書館において利用可能なさまざまな情報資源を「主題－探索事項」の枠組みに沿ってシステム上で整理するためのオントロジーを設計する。ここで、情報資源の多様性への柔軟性をもたせるためにフォークソノミーの考え方を導入する。

(2) 自宅のパソコンや図書館のキオスク端末で自ら検索を行う一般利用者を対象として効率的に情報探索支援を行うことのできるユーザインタフェースのあるべき姿を明らかにする。具体的には、「主題－探索事項」を軸として、上述の知識構造を活用して対話的に情報要求の具体化を行うことができるインタフェースを開発する。

(3) 上述の「知識構造」「ユーザインタフェース」がどれだけ利用者の情報要求具体化に寄与したかを客観的に評価するための方法論を明らかにする。具体的には、プロトタイプシステムを実際に運用して得られるログデータを評価するための多角的な指標を開発する。

3. 研究の方法

(1) 「情報要求の具体化」を目的としたハイブリッドオントロジーを適切に構築するためには、日常の業務において利用者の情報要求に向き合っている図書館員の知見が不可欠である。そのため、本研究では参考調査の現場で豊富な経験をもっている図書館員の方からの支援を受ける。

(2) 本研究の目的上、一般の利用者に広く活用してもらうことが不可欠となるため、本研究において構築するプロトタイプシステムには Web ベースのユーザインタフェースをもたせる。すでに平成 18～19 年の若手研究(B)で開発した情報探索支援システム

「Reference Navigator」の Web ユーザインタフェース構築において協力関係にある東京電機大学増田研究室と引き続き研究協力を行う。また、図書館員の方にプロトタイプの試用をお願いし、いただいた意見を参考にシステムの改善を図る。

4. 研究成果

(1) Wikipedia カテゴリと図書館分類体系を「主題」「探索事項」の 2 つの観点から結びつける手法について研究を行い、「神奈川県経済」「アメリカの政治家」などの精密な分類を活用する手法を実装した。

(2) 国内の複数の大学図書館において本システムを実際に運用し、図書館員や一般利用者からの意見を収集した。

(3) 「Wikipedia カテゴリと図書館分類体系の関係を示すテーマグラフの描画機能」を本システムに実装した。

(4) 本システムの各機能が利用者の情報探

索においてどのように利用されているかを分析するため、ログデータの収集・分析を行った。

(5) 本システムのうち、Wikipedia カテゴリネットワークを解析する部分を独立したコンポーネントとして抽出し、Wikipedia 軽量解析ツールとして一般に公開した。

(6) 実運用システムの実装を行った。具体的には、国立国会図書館 (NDL) が保有する情報探索向けのさまざまなデータベースを統合して提供するポータルサイト「リサーチ・ナビ」の検索システムの設計・開発に取り組んだ。NDL は一般国民向けの情報探索支援サービスとして、蔵書検索システム NDL-OPAC、近代デジタルライブラリ、参考書誌紹介、調べ方案内、レファレンス協同データベースなど、多数のデータベースサービスを運営しているが、それぞれのデータベースがほとんど連携しておらず、利用者にとっての使いにくさが指摘されていた。リサーチ・ナビは NDL の Web サイトにおいて 2009 年 5 月に公開され (<http://rnavi.ndl.go.jp/>)、ライブラリアンを中心としたユーザ層に、文献調査を行うための基本的なツールとして受け入れられている。

(7) リサーチ・ナビの設計・開発において得られた知見を、国際学会や招待講演などで発表した。具体的には、デジタルライブラリ分野の国際会議 JCDL2009 をはじめ、図書館総合展、京都大学図書館機構講演会、日本図書館研究会、生物医学図書館員研究会など、ライブラリアンが多数集まる機会における招待講演を行った。また、NDL が発行する雑誌「参考書誌研究」に、リサーチ・ナビ検索システムの技術解説記事を寄稿した。

(8) リサーチ・ナビに実装されたテーマグラフをはじめとする機能の評価実験を行った。具体的には、100 名の被験者を対象としてリサーチ・ナビを実際に利用してもらい、アンケートを収集した。その結果、テーマグラフについては 78% のユーザが有用なヒントを得られたと回答した。また、72% のユーザが「Google や Yahoo! では得られない有用な情報を得た」と回答した。

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 7 件)

- ① Yoji Kiyota, Noriyuki Tamura, Satoshi Sakai, Hiroshi Nakagawa, Hidetaka Masuda, Automated Subject Induction from Query Keywords through Wikipedia Categories and Subject Headings, Proceedings of The Sixth International Conference on Language

Resource and Evaluation (LREC 2008), 査読有, L08, 2008, p. 1255

- ② 増田英孝, 清田陽司, 中川裕志, 自動レファレンスサービスにむけて, 情報の科学と技術, 査読有, Vol. 58, 2008, pp. 347-352
- ③ 清田陽司, レファレンスサービスをパワーアップするレファレンスナビの開発 ~図書館情報資源と Web 情報資源の統合的利用~, 短期大学図書館研究, 査読無, Vol. 28, 2008, pp. 51-56
- ④ 清田陽司, 対話的情報探索支援のためのオントロジー ~Wikipedia と図書館件名標目表の統合~, 短期大学図書館研究, 査読無, Vol. 28, 2008, pp. 135-138
- ⑤ Yoji Kiyota, Hiroshi Nakagawa, Satoshi Sakai, Tatsuya Mori, Hidetaka Masuda, Exploitation of the Wikipedia Category System for Enhancing the Value of LCSH, Proceedings of the 9th ACM/IEEE-CS joint conference on Digital Libraries, 査読有, 2009, pp. 411-412
- ⑥ 清田陽司, リサーチ・ナビ検索システムの技術, 参考書誌研究, 査読無, Vol. 71, 2009, pp. 33-53
- ⑦ 清田陽司, Wikipedia を活用した新たな情報ナビゲーションシステムの提案, 薬学図書館, 査読無, Vol. 55, 2010, pp. 51-59

[学会発表] (計 2 1)

- ① Yoji Kiyota, Fusion of Libraries and the Web: Subject-based Information Retrieval in the Web 2.0 Era, The 2008 annual meeting of the Committee on Japanese Materials (CJM), Council on East Asian Libraries (CEAL), Association for Asian Studies (AAS), 2008 年 4 月, Atlanta, Georgia, USA
- ② 清田陽司, 対話的情報探索支援のための統合分類体系の構築, 第 7 回情報科学技術フォーラム (FIT2008) イベント企画「情報爆発時代の自然言語処理の新展開 ー大規模ウェブリソースは対話を賢くするか?ー」, 2008 年 9 月, 神奈川県藤沢市
- ③ 森竜也, 増田英孝, 清田陽司, 中川裕志, Wikipedia データ抽出ツール WikIE, NLP 若手の会 第 3 回シンポジウム, 2008 年 9 月, 静岡県熱海市
- ④ 清田陽司, レファレンスサービスをパワーアップする Web2.0 型レファレンスナビの開発, 第 10 回図書館総合展 フォーラム企画, 2008 年 11 月, 神奈川県横浜市

- ⑤ 森竜也, 増田英孝, 清田陽司, 中川裕志, Wikipedia エントリ構造抽出ツール: Wik-IE, セマンティックウェブとオントロジー研究会 第 20 回 Wikipedia ワークショップ, 2009 年 1 月, 東京都文京区
- ⑥ 中野幹生, 緒方淳, 清田陽司, 東中竜一郎, 翠輝久, 【パネル討論】音声インタフェースにおける Web テキスト処理技術の利用, 情報処理学会 第 191 回自然言語処理研究会 (NL-191-12), 2009 年 5 月, 東京都
- ⑦ 清田陽司, 図書館分類体系と Wikipedia を統合した情報探索支援システムの開発, 日本図書館研究会情報組織化研究グループ/情報知識学会関西部会 合同研究会, 2009 年 7 月, 大阪市
- ⑧ 清田陽司, Wikipedia をいかに使いこなすか? —知識抽出、情報ナビゲーション、そしてトピック発見—, 第 114 回 Ku-librarians 勉強会, 2009 年 8 月, 京都市
- ⑨ 清田陽司, Wikipedia と図書館情報資源のマッシュアップ, 第 36 回生物医学図書館員研究会, 2009 年 9 月, 東京都
- ⑩ 清田陽司, リサーチ・ナビ検索システムの技術, 第 11 回図書館総合展/学術情報オープンサミット 2009 フォーラム企画, 2009 年 11 月, 神奈川県横浜市
- ⑪ 清田陽司, 学生向けレファレンス支援ツールの可能性, 11 回図書館総合展/学術情報オープンサミット 2009 ミニ・フォーラム&プレゼンテーション企画, 2009 年 11 月, 神奈川県横浜市
- ⑫ 中崎寛之, 阿部佑亮, 宇津呂武仁, 河田容英, 福原知宏, 神門典子, 吉岡真治, 中川裕志, 清田陽司, 特定トピックの日英ブログ収集・分析・類型化: 事例研究, 情報処理学会 第 194 回自然言語処理研究会 (NL-194-9), 2009 年 11 月, 愛媛県松山市
- ⑬ 佐藤由紀, 横本大輔, 中崎寛之, 宇津呂武仁, 吉岡真治, 福原知宏, 神門典子, 中川裕志, 清田陽司, Wikipedia を介した関連ニュース・ブログの対応付け: Wikipedia エントリの分析, 情報処理学会 第 194 回自然言語処理研究会 (NL-194-10), 2009 年 11 月, 愛媛県松山市
- ⑭ 坂井哲, 増田英孝, 清田陽司, 中川裕志, 国立国会図書館リサーチ・ナビにおけるテーマグラフの生成, 情報処理学会 第 96 回情報学基礎研究会 (FI-96-5), 2009 年 11 月, 東京都
- ⑮ 清田陽司, 知識体系の新たな融合: 情報探索と件名標目表の活用をめぐって, 京都大学図書館機構 平成 21 年度第 2 回

講演会「次世代 OPAC を考える 一目録情報の視点から」, 2009 年 11 月, 京都市

- ⑯ 清田陽司, Apache Hadoop による大規模データの研究活用事例, 産学連携共同研究シンポジウム「クラウド・コンピューティングと大学発ベンチャー」, 2010 年 1 月, 東京都
- ⑰ 森竜也, 増田英孝, 清田陽司, 中川裕志, Wikipedia を活用した言語間差異比較システムの提案, 第 2 回データ工学と情報マネジメントに関するフォーラム/第 8 回日本データベース学会年次大会 (DEIM2010), 2010 年 3 月, 兵庫県
- ⑱ 野田陽平, 清田陽司, 中川裕志, ネットワーク構造を利用した Wikipedia からの意外性のある情報の抽出, 情報処理学会 第 72 回全国大会 講演論文集 (3K-4), 2010 年 3 月, 東京都
- ⑲ 坂井哲, 増田英孝, 清田陽司, 中川裕志, Wikipedia と図書館情報資源による調べ方自動提示システム, 情報処理学会 第 72 回全国大会 講演論文集 (3K-5), 2010 年 3 月, 東京都
- ⑳ 森竜也, 増田英孝, 中川裕志, 清田陽司, Wikipedia における言語間の差異マイニング, 情報処理学会 第 72 回全国大会 講演論文集 (5ZN-6), 2010 年 3 月, 東京都
- 21 清田陽司, Hadoop による大規模分散データ処理, 産業技術大学院大学 InfoTalk 第 16 回, 2010 年 3 月, 東京都

〔図書〕(計 0 件)

〔産業財産権〕

○出願状況 (計 0 件)

○取得状況 (計 0 件)

〔その他〕

ホームページ等

<http://www.r.dl.itc.u-tokyo.ac.jp/~kiyota/>

6. 研究組織

(1) 研究代表者

清田 陽司 (KIYOTA YOJI)

東京大学・情報基盤センター・助教

研究者番号: 10401316

(2) 研究分担者

(3) 連携研究者