

氏 名	お 小 山 さとし 聡
学位(専攻分野)	博 士 (情 報 学)
学 位 記 番 号	情 博 第 51 号
学位授与の日付	平 成 14 年 3 月 25 日
学位授与の要件	学 位 規 則 第 4 条 第 1 項 該 当
研究科・専攻	情 報 学 研 究 科 社 会 情 報 学 専 攻
学位論文題目	Query Refinement for Domain-Specific Web Search (ドメイン指向 Web 検索のための質問の精練)

論文調査委員 (主 査) 教 授 石 田 亨 教 授 上 林 彌 彦 教 授 奥 乃 博

論 文 内 容 の 要 旨

本論文は、World Wide Webにおける情報検索の困難を解消する手段としての、特定のドメインに特化した専門の検索システムを、利用者の質問の精練を通して実現するための手法について論じたものであり、6章から構成されている。

第1章は序論であり、本研究の目的と本論文の構成を述べている。

第2章では、Webにおける情報検索の現状とその問題点について述べている。現在一般に用いられている汎用検索エンジンのアーキテクチャと機能、および検索エンジンの評価方法について述べている。汎用検索エンジン利用の実態に関する既存の調査研究を参照しながら、なぜ汎用検索エンジンによる検索が、多くの利用者にとって困難であるのかを、利用者の質問構築能力とWebの多様性の観点から論じている。

第3章では、Webにおける情報検索の困難を解決する手段の一つとして、特定のドメインに関係するページだけが検索できる専門検索エンジンが有効であることを述べ、その新しい構築方法を提案している。ドメイン専門検索エンジンの構築手法として現在採用されている、専用のインデックスを構築する方法および情報フィルタリングを用いる方法では、構築に際し大きなコストが必要となることが指摘されている。そこで、検索隠し味とよばれるドメインに固有のブール式を利用者の質問に付加し、汎用検索エンジンに転送することにより、ドメインに関するWebページだけを取得する新しい専門検索エンジンのアーキテクチャを提案している。提案手法により、大きな設備的・人的コストをかけずに、専門検索エンジンを構築することが可能になっている。また、本手法で検索隠し味として用いられるブール式を抽出するための、決定木を用いたアルゴリズムについて述べている。ここでは小さなキーワード数で、性能の良いブール式を抽出することが必要であり、そのためのブール式の単純化手法を提案している。既存の単純化手法との比較により、提案された単純化手法の優位性を検証している。また、抽出されたブール式の有効性を、料理レシピドメインでの評価実験において、適合率、再現率という情報検索の性能評価指標に基づき、定量的に検証している。

第4章では、特定のドメインに属するWebページ群の中から、利用者が必要とする情報を効率的に検索することを支援する手法を提案している。現在の情報検索エンジンのインタフェースでは、様々なWebの利用形態、特にモバイル環境からの利用に対応できないことを論じている。この問題を解決するために、利用者がシステムとインタラクションを行いながら、質問を精練していく方法が必要であることを述べている。そこで、キーワード間の連想ルールをドメインのWebページから抽出して、質問の精練を行う手法を提案している。特に、連想ルールを、ドメインに固有の検索対象の属性を意識しながら利用することで、効果的な情報ナビゲーションを可能にする手法を提案している。検索の成功のためには、抽出されるルールの質が重要であるため、ページのフォーマットなどに現れるキーワードをクラスタリングを用いて除去する方法、統計的に無意味なルールを検定を用いて除去する方法を提案している。また、提案手法について、レストラン検索のドメインでの評価実験を通してその有効性の検証を行っている。

第5章では、人々の日常生活の様々な面を支援するための、情報エージェントのアーキテクチャについて述べている。ま

ず、デジタルシティに代表されるような、人々の日常生活を支援するための基盤がインターネット上で構築されつつあることが示され、その上での情報エージェントの役割について論じている。さらに、専門検索エンジンの上に構築された情報エージェントの例を参照しながら、日常生活を支援する情報エージェントを実現するためには、複数のドメイン専門の Web 検索システムを組み合わせる動作させる方式が有効であることを述べている。そして、利用者との対話を担当するフロントエンドエージェント、ドメイン専門の情報検索を実施するバックエンドエージェントを協調動作させるマルチエージェントアーキテクチャを提案している。

第6章は結論であり、本論文で得られた成果について要約している。

論文審査の結果の要旨

本論文は、World Wide Web における情報検索の困難を解消する手段としての、特定のドメインに特化した専門の検索システムを、利用者の質問の精錬を通して実現するための研究をまとめたものであり、得られた主な成果は次のとおりである。

- (1) Web からドメインに関係するページだけを検索できる専門検索エンジンを構築するための新しい手法を提案した。“検索隠し味”と呼ばれる、ドメインのページを分類するキーワードのブール式を利用者の質問に付加し、汎用検索エンジンに転送することで、専門検索エンジンを実質的に構築する。この手法により、単純なシステム構成で、応答性能の良いドメイン専門検索エンジンを構築することが可能になった。
- (2) 検索隠し味を Web ページの訓練集合から抽出する機械学習のアルゴリズムを開発した。決定本を用いて抽出された初期の複雑なブール式を、検証集合に対する適合率と再現率の調平均に基づいて単純化することで、適合率と再現率のバランスのとれたブール式を、汎用の検索エンジンに投入可能な小さなキーワード数で抽出できる。この手法を用いれば、分類された Web ページの集合があれば、専門的知識なしに専門検索エンジンを構築できる。料理レシピ検索のドメインで評価実験を行い、本手法で抽出された検索隠し味の有効性を実証した。
- (3) ドメインの Web ページ集合内で利用者の要求を満たすページを検索する問題に、連想ルールを適用する手法を提案した。ドメインに固有の検索対象の属性を考慮して連想ルールを適用することにより、検索の制約条件の緩和や制約条件の追加といった質問の精錬を行うことを可能とした。
- (4) ドメインの Web ページ集合から抽出された連想ルールを、キーワードのクラスタリングと統計的検定を用いて精製する手法を提案した。この手法により、ドメイン内での検索に有効な連想ルールの割合を向上させることができる。レストラン検索のドメインにおいて、提案手法の有効性を実証した。

以上、本論文はドメインを特定した Web 検索を、質問の精錬によって実現する手法をまとめたものであり、学術上、実際上寄与するところが少なくない。

よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成14年2月5日実施した論文内容とそれに関連した試問の結果合格と認めた。