

氏 名	かざ 風 ま かず ひろ 間 一 洋
学位(専攻分野)	博 士 (情 報 学)
学 位 記 番 号	情 博 第 183 号
学位授与の日付	平 成 17 年 11 月 24 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 シ ス テ ム 科 学 専 攻
学位論文題目	Web 空間における情報と人間の関連性分析とその応用

論文調査委員 (主 査) 教授 片 井 修 教授 熊 本 博 光 教授 石 田 亨

論 文 内 容 の 要 旨

Web 情報検索とは、World Wide Web で公開されている膨大な情報から、目的の情報を探し出す行為である。この Web 情報検索に関する研究は、インターネット上の分散して格納されたデータの分析・利用を支えるフレームワークとして非常に期待を集めている。Web 情報検索を困難にする要因は、扱う情報が非常に膨大であるにもかかわらず統制されておらず種々雑多であることと、ユーザが情報を探す手段は基本的に検索式に限定されることである。つまり、少数の単語から構成される検索式を用いて、膨大な情報から目的の情報を発見するためには、情報検索システム側でユーザを支援しなければならない。

本論文は、膨大な情報の中からの目的とする情報の発見の支援を目的として、Web マイニングに関して行ってきた研究をまとめたものである。本研究の重要な点は、自然言語処理技術の範囲を超えて、インターネット上の人間の行動に着目し、情報と人間の間の意味のある相関関係や規則を探し出し、情報検索支援に利用したことであると言える本論文は、7章で構成されている。

第1章は序論であり、本研究の背景及び目的、および従来の Web 情報検索に対する本研究の位置づけを述べている。

第2章では、Web マイニングに関する基礎として、Web 利用マイニング、Web 構造マイニング、Web 内容マイニングに関する基本概念を述べ、システム設計における評価基準を与えている。

第3章では、Web 空間における情報閲覧者の情報閲覧履歴を分析し、情報閲覧者と情報の相関関係を求めて、Web 情報検索の支援に適用する手法を提案している。実際には、プロキシサーバのログファイル中の URL の2-gram に着目して導出した HTML 文書の相関関係の度合いを表す相関度から、あるコミュニティ中の HTML 文書の関連グラフを作成しており、評価においては実際のデータを使用して、相関度が文書間の関連性を示す良い指標であることを証明している。

第4章では、Web 空間における情報作成者に関する経験則から文書構造を推定し、検索結果を複数のレベルでグループ化可能にすると共に、その文書群の索引となるインデックスページを提示することで、情報検索を効率化する手法を提案している。さらに、実際に Web サーチエンジンとして実装・公開し、その運用履歴を分析することで、有効性を示している。

第5章では、Web 空間における情報作成者が作成したハイパーリンクの構造を分析し、Web ディレクトリの各カテゴリに適合する権威ある Web サイトを発見し、さらにその Web サイトについて述べている文章群の中で最も適した文章を説明文として抽出することで、Web ディレクトリを自動的に拡張する手法を提案している。評価では、関連する権威ある Web サイトの発見、および説明文発見について被験者を用いた実験をおこなうだけでなく、実際に Web ディレクトリとして実装・公開して、その運用履歴を分析することで、システムの有効性を示している。

第6章は、実世界の実体と Web 空間の中の固有表現を結びつけることで、ユーザの検索過程を支援する実世界指向の情報探索手法を提案している。実際には、特に固有表現として人名に着目しており、収集した大量の Web ページ群から姓と名の組を取り出し、その共起から人間関係を求めているが、精度を向上するために、共起の判定をページ内の狭い範囲に限

定したり、影響度と呼ぶ独自の中心性の指標を用いてキーパーソンだけに限定している。実際に実装されたシステムでは、内部にサーチエンジンを持っており、検索用の GUI と人間関係可視化の 2 種類のインタフェースを用いて情報を探索できることが示されている。さらに、人名分布、人間関係ネットワークの評価と、さまざまなトピックに対して情報探索の有効性が検討されている。

第 7 章は結論であり、本研究で得られた成果をまとめ、今後の課題について述べている。

論文審査の結果の要旨

本研究は、膨大な Web 情報の中からの目的とする情報の発見の支援を目指したものであり、Web 情報の探索閲覧履歴、Web のハイパーリンクの構造、Web ページ内の固有表現などの複数の側面から、Web 情報の情報探索を支援する手法を提案・実装し、実際の利用履歴や Web 情報を用いて評価することで有効性を示している。得られた主な成果は以下のよう

1. Web 上の情報を閲覧した履歴の中の URL の隣接関係に着目し、それを統計的に処理することで情報閲覧者側から見た Web 文書の関係グラフを導出し、コミュニティ内の重要な関連情報や情報探索経路の提示に使用する可能性を示した。
2. Web 空間の情報作成者に関する経験則を用いた文書構造の推定に基づいてサーチエンジンの検索結果をグループ化して検索結果リスト長を短縮するとともに、その文書群の索引となるインデックスページを利用者に提示することで、情報探索を効率化できることを示した。
3. Web 空間の情報作成者が作成したハイパーリンクの構造を分析することで、与えられた特定のトピックに関する文書群に内容が類似し、かつ有用だとみなされている Web サイトを発見する手法を提案し、その有効性を示した。
4. ある Web サイトについての説明文を編集者が人手で作成するかわりに、Web サイトを集めて紹介しているリンク集からある特定のパターンに着目することで紹介文を収集し、複雑な自然言語処理をおこなわずに簡単な文字や単語出現と文字列長に基づいて説明文として適切な紹介文を抽出する手法を提案し、その有効性を示した。
5. 検索結果の Web ページ群から人名を抽出し、Web ページ内の人名の近接関係や人名の Web 空間上の分布を考慮することで、検索語が示すトピックに関する人間関係を抽出する方法を提案し、さらに抽出された人間関係を Web 情報の情報探索に使用できることを示した。

以上のように、本論文は情報と人間の関連性を Web 空間の中に見いだす先進的な方法とその効果を明らかにするものであり、情報学の展開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17年10月26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。