

氏名	高見眞也
学位(専攻分野)	博士(情報学)
学位記番号	情博第298号
学位授与の日付	平成20年3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	情報学研究科社会情報学専攻
学位論文題目	検索目的に応じたスニペットの生成・拡張によるウェブ検索結果の最適化に関する研究

論文調査委員 (主査) 教授 田中克己 教授 石田 亨 教授 吉川正俊

論文内容の要旨

本論文では、ウェブ検索エンジンを利用したウェブ情報検索において、検索されたウェブページの適合性を判断するうえで重要な概要文(スニペット)に着目し、利用者の検索意図に応じたスニペットの生成・提示を行うことで検索結果の最適化を実現する手法について述べている。

ウェブ検索エンジンを利用したウェブ情報検索では、インターネットを利用する人口の増加による検索目的の複雑化とウェブ上に散在する情報の多様化により、利用者が入力した検索質問だけで最適な検索結果を得ることはますます難しくなっている。また、刻一刻と変化する流動的なウェブ情報空間を検索するには、従来のデータベースを対象とした検索技術だけではなく、ウェブ情報の半構造的性、社会性に着目したアプローチが有効な場合もある。そのため、ウェブ情報検索における検索精度向上のために、ランキングアルゴリズムの改善だけではなく、検索質問の修正や拡張による再検索支援や、検索結果の分類やクラスタリング、検索履歴を利用した個人適応化などを行うことで検索結果の最適化を目指す研究が行われている。

本論文における研究では、このようなウェブ検索エンジンを利用したウェブ情報検索において、ウェブページの適合性を判断するうえで重要なスニペットに着目し、スニペットへの情報付加やスニペットの動的な生成方法を提案している。

具体的には、ウェブページの主題と検索語を含む部分との内容の一致度を視覚化して提示することにより正解ページの選択に関する利用者の判断支援を行ったり、スニペットの生成方法を、内容の包括要約性と対象の集合依存性を軸として分類し利用者の検索意図に応じたスニペットを生成・提示することにより、検索結果中から利用者が適合性の高いページを選択できるよう支援する手法を提案している。このために、本論文では、「ウェブページに対する定量的評価の視覚化」、「生成方法の分類に基づくスニペットの動的再生成」、「検索結果を利用した関連・因果性をもつ話題語の獲得」という3つのテーマについて研究を行い、提案した手法の評価を行っている。

本論文は6章から構成されている。その概要は以下の通りである。

第1章は序論であり、本論文の研究の背景、本論文の研究を行うに至った動機、および、本論文の研究の全体の概要について述べている。

第2章では、第3～5章において述べるそれぞれの研究テーマについて、特に深く関連する従来の研究について説明し、本研究の位置付けを行っている。

第3章では、ウェブページの適合性を推測するための判断材料としてのスニペットの機能をいかに強化するかについて述べている。具体的には、ウェブページに対する定量的評価として、ウェブページの主題と、ウェブページ内の利用者が指定した検索語を含む部分との内容の一致度を、トピック-クエリ適合度として評価し、この情報を視覚化して提示する手法を提案している。また、正解ページのスニペットと類似するスニペットをもつウェブページは、別の正解ページである可能性が高いという仮説のもとで、トピック-クエリ適合度やスニペット同士の類似度が高い場合に正解ページとなる可能性の高

い検索質問クラスを同定することを試みている。そして、特定の検索質問クラスの場合、トピックークエリ適合度やスニペット同士の類似度が検索結果で視覚化されると、ウェブページの特徴を推測するための判断材料になり得ることを示している。

第4章では、ウェブページの断片としてのスニペットを検索目的に応じた形でいかに動的に生成するかについて述べている。ここでは、スニペットの生成において考慮すべき、内容の包括要約性および対象の集合依存性を軸として、I型からIV型までの4種のスニペット生成手法を提案している。さらに、それぞれのスニペットを4種類の重要語により具体的に生成する手法を提案している。次に、様々な視点で検索語に関する情報を獲得することを目的とした4種類の検索質問クラスを用意し、スニペット・タイプの違いが検索精度に与える影響について被験者実験を行っている。そして、それぞれの検索質問クラスに適したスニペットの特性を確認し、検索質問クラスによっては、ウェブページの主題を示すスニペットや、他のウェブページとの関係を考慮したスニペットの場合に、高い検索精度が得られることを示している。

第5章では、より高度な集合依存性を考慮したスニペット生成を行うために、検索語と関連する話題語の獲得について述べている。特定の時区間や空間条件のもとで、相関関係または因果関係をもつ話題語の組合せ（社会的知識）を獲得する対象としてブログ記事に着目し、時系列的な話題共起性とコミュニティ依存性に注目したコミュニティ評価型モデルを提案している。また、既知の因果関係をもつ話題語の組合せが、提案した評価モデルの特徴を有していることを確認し、ウェブ検索エンジンの検索結果を利用して、何らかの社会的知識が獲得できる可能性を示している。

最後に、第6章では、本研究で得られた研究成果をまとめ、さらに今後の展開について述べている。

論文審査の結果の要旨

本論文は、ウェブ検索エンジンを利用したウェブ情報検索において、ウェブページの適合性を判断するうえで重要な概要文（スニペット）に着目し、利用者の検索意図に応じたスニペットの生成・提示を行うことで検索結果の最適化を実現する手法について検討したものである。

ウェブ検索エンジンを利用したウェブ情報検索では、インターネットを利用する人口の増加による検索目的の複雑化とウェブ上に散在する情報の多様化により、利用者が入力した検索質問だけで最適な検索結果を得ることはますます難しくなっている。学位申請者は、検索されたウェブページの主題と、そのページ内の検索語を含む部分との、内容の一致度をトピックークエリ適合度として視覚化したり、スニペットの生成において、内容の包括要約性と対象の集合依存性を軸とした分類を行うことで検索目的に応じたスニペットを生成・提示する手法を提案している。さらに、それら提案手法について、有効性検証のための評価実験を行い、興味ある結果を提示している。

具体的には、本論文の研究によって得られた成果は以下のように要約される。

1. ウェブページの主題と、ページ内の検索語を含む部分との内容の一致度を、トピックークエリ適合度として評価する方式を提案し、得られたトピックークエリ適合度を視覚化してスニペットとあわせて提示する手法を提案した。また、正解ページのスニペットと類似するスニペットをもつウェブページは、別の正解ページである可能性が高いという仮説のもと、トピックークエリ適合度やスニペット同士の類似度が高い場合に正解ページとなる可能性の高い検索質問クラスを同定することを試みた。そして、特定の検索質問クラスの場合、トピックークエリ適合度やスニペット同士の類似度を検索結果上で視覚化して提示することが、ウェブページの適合性の推測に有用となることを示した。
2. 内容の包括要約性および対象の集合依存性を軸として、生成方法によりI型からIV型までのタイプにスニペットを分類し、それぞれのスニペットを4種類の重要語により生成する手法を提案した。次に、検索意図の異なる4種類の検索質問クラスを用意し、スニペット・タイプの違いが検索精度に与える影響について被験者実験を行った。そして、それぞれの検索質問クラスに適したスニペットの特性を確認し、検索質問クラスによっては、ウェブページの主題を示す要約型のスニペットや、他のウェブページとの関係を考慮したスニペットの場合に高い検索精度が得られることを示した。
3. 特定の時区間や空間条件のもとで、相関関係または因果関係をもつ話題語の組合せ（社会的知識）を獲得する対象としてブログ記事に着目し、時系列的な話題共起性とコミュニティ依存性に注目したコミュニティ評価型モデルを提案

した。また、既知の因果関係をもつ話題語の組合せが、提案した評価モデルの特徴を有していることを確認し、ウェブ検索エンジンの検索結果を利用して、何らかの社会的知識が獲得できる可能性を示した。

以上、本論文は、ウェブ情報検索におけるスニペットの特性と検索質問クラスとの関係をいくつかの視点から評価し、スニペットの生成方法に着目することで検索意図に応じた検索結果を提示できる手法を提案しており、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20年2月6日に論文内容とそれに関連した事項について試問を行った結果、合格と認めた。