

(続紙 1)

京都大学	博士 (情報 学)	氏名	山家 雄介
論文題目	ソーシャルブックマークの特性に基づいたウェブ検索ランキングに関する研究		
(論文内容の要旨)			
本論文における研究では、ウェブ検索エンジンを利用したウェブ情報検索におけるランキングを最適化する目的で、ウェブページの適合性を判断するうえでソーシャルブックマークのデータに着目し、このデータをマイニングすることで既存のページ評価尺度と比較した上で、これまでの検索エンジンでは検索が難しかったページの検索を可能にする手法を提案している。			
具体的には、ソーシャルブックマークにおけるページのブックマーク数や、ブックマークの際に使用されたタグといった情報をページへの一種のアノテーションと見なすことで、これを用いた検索結果の対話的に再ランキングを可能にし、従来は検索結果からの発見が難しかったページの発見を容易にしている。また、ソーシャルブックマークにおいて特定の時期にブックマーク量が増加するページを、その時期に特に有用な情報が載っていると仮定することで、検索者が検索を行う時期に適応した検索結果のランキングを可能にしている。さらに、ページをブックマークしているユーザ間の参照構造を分析し、集団としての特性を明らかにすることで、ページの社会的なインパクトをより正確に反映するモデルと指標を提案している。このために、本論文では、「ソーシャルブックマークの特性分析とそれに基づくWeb 検索結果の対話的再ランキング手法」、「ソーシャルブックマークの周期性発見と時期連動型検索ランキングへの適用」、「参照構造分析に基づいたページへのソーシャルブックマーク数の検証」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、既存の主なページの重要性の評価尺度であるPageRank値とソーシャルブックマークにおけるページのブックマーク数 (SBRank) との関係分析を詳細におこない、SBRankで検索結果を再ランキングすることで、検索ランキングの精度を改善する手法を述べている。その結果、ページのPageRank値とSBRankにはある程度の相関関係があり、これはSBRankによるPageRank値の補完の可能性を示している。次に、分野が異なる複数の検索クエリにおいて、複数の度合いでSBRankを重みづけすることによる再ランキング実験を行っている。その結果、いくつかの検索クエリにおいて適合率および再現率が改善するものの、分野によってはページへのソーシャルブックマークが少ないために再ランキングされるページが限定的であり、提案手法の適用範囲に限界があることを確認している。			

第3章では、ソーシャルブックマークにおけるウェブページのブックマーク数や、ユーザが分類に使用したタグを集約し、それを利用して対話的に再ランキング可能にすることで、いかにユーザがウェブ検索結果から有用なページを発見可能にするかについて述べている。具体的には、ウェブページの有用性に対するリンク構造分析に基づいた既存の主な評価尺度であるPageRankと、ページのソーシャルブックマーク数に対する定量的評価として、両者の値にどのような関係があるかを評価している。また、ソーシャルブックマークユーザがページの分類用キーワードとして利用するタグの使用回数の分布を調査した。その結果は、ソーシャルブックマークユーザはページの内容に関するタグ以外にも、ページの印象や性質といった視点での分類を数多く行っていることを示している。この調査結果を受けて、ページには含まれない閲覧者の印象による検索ランキングを可能とする手法などを提案し、いくつかの興味ある検索結果例を示している。

第4章では、ソーシャルブックマークにおいて周期的にブックマーク数が増加するページを検出して、それをウェブページ検索者の時期による検索意図の変化に応じた時期連動型ランキングに利用するかについて述べている。ここでは、検索意図が検索時期によって変わるような検索クエリにおいて、検索時期においてソーシャルブックマークで周期的にブックマークされるページをクエリログマイニングで用いられている研究を応用することで検出し、その周期の発生回数に基づいてページを重み付けし、再ランキングするというものである。実データによる実験を行った結果、伝統行事に関連したショッピング情報を扱ったページなど、実際に検索時期において特に必要とされ、それ以外の時期にはほとんど有用でないようなページがとくにシーズン中に上位に再ランキングされるという結果となっている。

第5章では、あるページがどの程度社会的にインパクトがあったかを測る指標としてページのソーシャルブックマーク数が利用されているものの、実際には影響力のあるブックマーカーによる影響が大きい可能性を指摘し、その影響がどの程度を測るモデルを提案している。具体的には、ソーシャルブックマークにおけるユーザ同士のブックマークの購読関係と同一ウェブページのブックマークの時間間隔から、ウェブページをブックマークしたユーザ間の参照構造を推定し、この構造を基にジェームズ・スロヴィッキーが提唱するよい集合知を形成な集団の特徴の3要件である「多様性」「独立性」そして「分散性」を評価しブックマーク数を補正することで社会的なインパクトの指標として近似する、という手法を提案している。実データによる評価の結果、ブックマーク数が同程度であり、話題も似通っているページであっても、資料性が高いものと速報性が高いものとでユーザ間の参照構造が異なっていることを示し、単純なウェブページへのソーシャルブックマーク数を補完し実際の社会的なインパクトを推定する指標としての可能性を示すことに成功している。

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の検索結果のランキングの際にページの重要度および検索クエリとの関連度を推定する情報源として、ソーシャルブックマークに蓄積されているページへのブックマークデータに着目し、これをマイニングすることでランキング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蔓延するスパム行為などにより、有用なページが埋もれてしまい、ページ間のリンク関係分析など既存のアプローチによるページの重要度や関連度の推定がますます難しくなっている。そこで学位申請者は、ソーシャルブックマークというウェブサービスに着目し、ブックマークとして蓄積されたページへの評価やタグによる分類を分析・利用することで、既存のウェブ検索結果を対話的に再ランキングし有用なページを発見する手法や、特定の時期に需要が増すページを、そのソーシャルブックマーク数が周期的に増加する現象を検出することで発見して検索時期に連動した検索ランキングに適用する手法や、ソーシャルブックマークユーザ間の参照構造を分析することでブックマーク数の社会的なインパクトを検証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. ブックマークの際にページの閲覧者の印象に相当するタグが利用されていることやブックマーク増加の時間的な特徴に着目し、これをウェブ結果の再ランキングで利用することで、ページの印象や注目度の時間的特性といった属性にもとづいてウェブ検索結果を対話的な再ランキング可能にする手法を提案した。評価実験により、閲覧者に特定の印象を与えるようなページが、通常のウェブ検索エンジンのみでは発見が困難でも、提案手法では容易になることを示した。
2. 人々がいろいろな生活情報を調べる際にウェブ検索エンジンを利用するが、その際に必要とする情報は検索が行われた時期によって異なることに着目し、ページごとの社会的需要の周期性をソーシャルブックマーク数の増加の周期性で近似することで、検索者が検索を行った時期に応じて検索ランキングを最適化する手法を提案した。評価実験の結果、実際に周期的に需要が増加すると考えられるページが、関連する検索クエリにおいて上位に再ランキングされることを示した。
3. ページのソーシャルブックマーク数はそのページの社会的なインパクトの度合いを示す指標の1つと考えられるが、実際には一部の影響力のあるブックマーカーが他のブックマーカーに影響を与えることでページのブックマーク数が増加する点に着目し、この影響の度合いを定量化することで実質的な社会的インパクトとしてブックマーク数を補正する手法を提案した。具体的にはユーザのページのブックマーク時間と互いの購読関係から参照構造を推定し、集団の多様性、独立性、分散性をそれぞれ評価するモデルを提案した。実データによる実験の結果、ブックマーク数が同程度であり同分野の内容のページであっても、そのページのコンテンツの性質によって参照構造が異なっているケースを発見し、より社会的なインパクトを反映する指標としての可能性を示した。

以上、本論文は、ソーシャルブックマークの特性を分析して明らかにした上で、

ページの重要度および関連度の評価を補完する新たな指標を複数提案し、それらを検索者が対話的に再ランキング可能にする手法を提案している。本論文に関連する国際会議論文は、Google Scholarによると学位申請時点で100件以上の被引用があり、学術上、實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23年2月21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。