

氏名(本籍)	島田諭(静岡県)
学位の種類	博士(情報学)
学位記番号	博甲第6273号
学位授与年月日	平成24年3月23日
学位授与の要件	学位規則第4条第1項該当
審査研究科	図書館情報メディア研究科
学位論文題目	関連項目間の探索的な遷移を支援する文書空間ナビゲーションの研究
主査	筑波大学教授 佐藤哲司
副査	筑波大学教授 田中和世
副査	筑波大学准教授 森嶋厚行
副査	筑波大学教授 石塚英弘
副査	静岡大学教授 石川博

## 論文の内容の要旨

インターネット上の情報検索エンジンは、知りたいこと分からないことを調べるのに極めて有効であり、多くのユーザが日々の生活の中で活用している。検索エンジンの基本原理は、ユーザが入力した検索語と適合度が高い文書、すなわち、ウェブページを適合度順に出力することであり、入力した検索語が適切であるかどうか、所望の結果を得られることの鍵となっている。これまで情報検索が対象としてきた新聞記事や特許などの文書集合では、使用される語彙が予め選別(語彙統制)されており、ユーザは効率的な検索が可能な検索語を比較的容易に想起することができた。企業などが提供するウェブページについても、ある程度は統制された語彙で文書が書かれていると期待できたが、一般の生活者が発信する情報(UGC: User-Generated Contents)では、その時々状況に応じて新しい語が作り出されるなど語彙の制限がなく、使用されている用語をユーザが事前に把握することは難しい。新聞記事のように語彙が統制された文書集合であっても、特段の予備知識を持たないユーザにとっては、UGCテキストへのアクセスと同様に、的確な検索語を想起することは困難をともなう。

本研究は、上述のように、ユーザが適切な検索語を想起することが困難な未知の文書集合を対象とした情報探索を支援する手法の実現を目的としている。内容を予想できない未知の文書集合からユーザが所望の情報を得るには、検索語を入力する従来型検索よりも、検索結果に応じて検索語を修正しながら検索を繰り返す探索的検索(Exploratory Search)が有効であるとの先行研究が知られている。そこでは、ユーザの入力に基づく試行錯誤では無駄が多く、探索の成否もユーザのスキルに大きく依存するとの報告がなされている。このような課題に対処することを目的として、本論文では、検索対象の文書集合において特徴的に使われる用語および用語間の関連性を文書集合から抽出・ユーザに提示し、ユーザは提示された項目を次々に選択していくことで多様な文書に到達できる文書空間ナビゲーション手法を提案している。

本論文全体を通した論旨の展開は以下である。まず、文書集合内を遷移可能とする文書空間ナビゲーションの基盤となる有向グラフを、探索の対象となる文書集合から生成する。出現する文書数が比較的少なく、特定の文書で頻繁に使用されている語を文書集合から抽出し、それらの語の共起関係に基づいて、関連する

語間、関連する文書間、および関連する語-文書間にリンクを生成する。共起関係に基づいて生成されたリンクは無向グラフであるから、ユーザを探索結果へとナビゲーションするために、共起語グラフの特性を可能な限り保持しながら有向グラフに変換する。この時、同時に提示する関連項目数、すなわちノードの出次数は可能な限り抑制する。生成した有向グラフを用いて利用者実験を行い、提案する文書空間ナビゲーション手法の有効性を検証している。

本論文は全5章で構成されている。以下、2章から5章までの概要を記述する。

第2章では、本研究に関連する先行研究を、探索的検索に関する研究、文書空間内のナビゲーションに関する研究、および、ユーザインタフェースに関する研究に大別して整理し、本研究の位置づけを明らかにしている。一般に、各文書中における語の共起関係をエッジとする共起語グラフはスモールワールド (Small-World) 性を示すことが知られており、この性質を失うことなく文書空間ナビゲーションに適した有向グラフへと共起関係グラフを変換することが、本研究の主たる課題であることを述べている。

第3章では、文章集合中に出現する語彙をノードとし、文章間での共起関係をリンクとする共起関係グラフの構築法、ならびに、各語彙がそれぞれの文書中に出現する頻度情報に基づいて有向グラフに変換する手法を明らかにしている。ナビゲーションは、ノードから出力されるリンクをユーザが選択して行うことから、このリンクの数を3から10程度の範囲となるように、共起語グラフの特性を維持したままグラフを書き換える手法を提案し、実装評価を行っている。

第4章では、前章で構築した有向グラフに基づいて、提示される関連語あるいは関連文書をユーザが選択する操作を繰り返す文書空間ナビゲーション手法を提案し、その有効性を利用者実験によって検証している。提案法により、ユーザは自ら新たな検索語を想起して入力することなく、提示される語あるいは文書をクリックして選択する操作だけで、多様な関連項目へと網羅的に文書空間内を遷移できることを確認している。

第5章では、本研究の成果のまとめとしての結論を述べ、今後の研究展開の方向、さらにユーザの探索行動ログの多面的な分析と合わせて検討することで、より広範な情報探索行動を支援できる可能性についても言及している。

## 審査の結果の要旨

インターネットの普及・発展に伴い、あらゆる情報へのアクセスが容易となったことは明らかであり、キーワードを入力して関連する文書を検索する全文検索エンジンの果たしている役割は計り知れない。しかしながら、著者が指摘しているように、検索者が適切な検索キーワードを想起できない限り、所望の情報を得ることは困難であることも事実である。この困難さは、検索対象である文書集合に対する予備知識を持たないユーザにとっては致命的であり、デジタルデバイドの解消という観点からも、多くの研究開発がなされている。また、複雑な情報要求には、単発の情報検索では満足する結果が得られないことから、検索結果に応じて検索語を修正しながら検索を繰り返す探索的検索を高度化する研究も盛んに行われている。

本論文で著者は、検索対象とする文書集合中に出現する語彙の出現頻度に着目している。特定の文書に頻出し他の文書での出現頻度が低い語は検索に有効な語であることは広く知られているところであるが、著者の着眼点は、中庸の頻度で、ある程度の文書に出現する語が、文書間を関連づける語彙として有効であるとする点である。この着眼点を核として、このような語彙をノードとするハイパーリンク構造を生成する手法を明らかにしており、本論文で提案されている手法に新規性を認めることができる。また、ユーザは、システムが提示する語を選択することで情報探索を行うことができ、これまでの検索エンジンでは不可欠であった検索語の入力を必要としないことから、提案法の有効性・有用性についても、十分な価値を認めることができる。以上述べたように、本論文で提案されているナビゲーション手法は、提示された語の選択という簡

便な操作で、関連する文書間の遷移、あるいは、文書空間内の網羅的な散策を行うことができる手法であり、探索的検索の領域において新たな地平を拓く橋頭堡になると期待でき、本論文の研究意義は十分に認められる。

本研究の特徴と認められる主要な成果を以下に示す。まず、文書集合中に出現する語彙の頻度と共起に関して、語彙統制がなされている新聞記事の集合と、生活者が自分の言葉で投稿している質問回答サイトの記事集合とを対象に、実践的かつ詳細に分析を行い、いずれの文書集合に於いてもスモールワールド性があることを確認している。スモールワールド性は、本論文の立脚点である「多くのノードに短い距離で到達できる」ことを示す重要な論拠となっていることから、本研究が対象としている、日本語文書集合を対象に、実際にスモールワールド性を確認できたことは大きな成果であると認められる。

スモールワールド性を示すグラフは、個々のノード（語彙や文書）から出るエッジ（アウトリンク）の本数がべき乗則となることから、このままではアウトリンク数が多くてユーザが選択できないノードが存在する。そこで、本論文では、スモールワールド性をできる限り損なわずにアウトリンク数を数本から十数本に制限するグラフの変換手法を提案している。情報探索者がリンクをたどりながら、文書集合内を網羅的に探索するナビゲーションにおいては、探索者が選べる程度のリンク本数であることは重要な要件であり、このようなグラフ変換手法が存在することを示したことは重要な成果であるといえる。

最後に、本論文では、新聞記事1年分（経済面）の記事を対象に、提案法を適用してグラフ構造を生成・変換し、数十名の実験協力者による情報探索実験を行っている。実験タスクは、あらかじめ探索のゴールが与えられない「3大ニュースを見つける」ことと、膨大な情報に埋もれてしまう可能性が高いマイナーな情報を見つける「O157（病原性大腸菌）の流行によって、売れるようになった商品は何か」の2題である。いずれのタスクも、従来のキーワード入力型情報検索では検索が難しいタスクであり、いずれのタスクにおいても提案手法の有効性が確認されている。

以上述べたように、本論文では、探索対象とする文書集合内に出現する特徴的な語彙をノードとしたグラフ構造の構成法を提案し、実際のデータを用いた利用者実験により有効性を検証していることから、提案法に対して一定の信頼性を認めることができる。一方で、提案手法の適用限界、特に、大規模な文書集合を対象とした場合に、現実的な時間でナビゲーションに適したグラフを構築可能であることを定量的に示すことも必要であろう。また、探索的検索という技術的課題において、本論文が目的としている文書空間の網羅的な把握の位置づけについても、より深い考察がなされることが望ましい。しかしながら、探索的検索に関する研究は緒についたばかりであり、本研究が探索的検索に対する一つの研究指針を示していることは明らかであり、本論文は博士論文として十分な内容を持つものと判断される。

平成24年1月16日、図書館情報メディア研究科学学位論文審査委員会において、審査委員全員出席のもと、本論文について著者に説明を求めた後、関連事項について質疑応答を行った。引き続き、「図書館情報メディア研究科博士後期課程の学位論文の審査に関する内規」第12項第12号に基づく最終試験を行い、審議の結果、審査委員全員一致で合格と判定された。

よって、著者は博士（情報学）の学位を受けるに十分な資格を有するものと認める。