

推薦論文

## 出次数制約付き有向グラフを用いた 関連語による文書空間ナビゲーション手法

島田 諭<sup>†1</sup> 福原 知宏<sup>†2</sup> 佐藤 哲司<sup>†1</sup>

探索的検索では、文書集合の特徴を反映させ、多様な関連項目への円滑な遷移を可能とする文書空間ナビゲーションが有効と考えられる。筆者らは、文書集合から抽出される語の共起関係に基づき、語および文書をノードとし、出次数を制約しながら関連項目間にリンクを生成する有向グラフ生成手法を提案している。本論文では、生成される有向グラフを用いて、ユーザが選択した文書または語に対し関連語を提示する文書空間ナビゲーション手法を提案する。新聞記事を用いたユーザ実験の結果、文書集合の特徴を反映させながら、ユーザが入力した語より多様で探索に有用な関連語を提示できることを確認した。

### A Method for Document Navigation in Related Words Using Out-degree Constrained Directed Graphs

SATOSHI SHIMADA,<sup>†1</sup> TOMOHIRO FUKUHARA<sup>†2</sup>  
and TETSUJI SATOH<sup>†1</sup>

In exploratory search, it is thought that document navigation is effective. In this case, document navigation should reflect characteristics of document set and should be able to lead users to various related words or documents by using reduced routes. We have proposed the method that generates out-degree constrained directed graphs based on cooccurrence word graph of document set. In this paper, we propose a document navigation method for presenting related words by using the generated directed graph. We did an experiment that used the newspaper article to compare the words presented by the proposal method with the words input by users. As a result, it was confirmed that the most of words presented by the proposal method is more suitable than the words input by users. Moreover, the words presented by the proposal method contained much more variations than the words input by users.

#### 1. はじめに

現在、広く使用されている、検索語と文書のマッチングを行う情報検索システムでは、ユーザは検索結果を事前に予想し、検索結果に出現する語を検索語として入力する必要がある。しかし、検索の手がかりとなる事実関係や主要なトピックなどをユーザが事前に把握しているとは限らない。さらに、不特定多数のユーザが投稿する電子掲示板やブログなどを対象とした検索では、文書の内容や使われる語が多岐にわたるため、ユーザが事前に検索結果を予想することはより困難になる。

このような、ユーザが内容を予想できない未知の文書集合を対象とした情報探索では、洗練された検索語を用いて多くの適合文書を得ることを志向する従来型検索は有効ではなく、検索結果に応じて検索語を修正しながら検索を繰り返し、その過程で少しずつ適合文書を得ていく探索的検索が必要になる<sup>1),2)</sup>。未知の文書集合における探索的検索を円滑化するには、検索対象の文書集合の特徴、すなわち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ユーザに提示し、ユーザが項目を順に選択することで多様な文書に到達できるナビゲーションが有効である<sup>3)</sup>。

本論文では、ユーザによる検索語の修正や追加を極力必要とせずに文書集合内を遷移可能とする文書空間ナビゲーション手法を提案する。筆者らは、探索に有用である、文書頻度  $df$  が低く反復度  $df_2/df$  が高い語の共起関係に基づき、ノードあたりの出次数を制約しながら、関連語間、関連文書間、および関連する語—文書間にリンクを生成する有向グラフ生成手法を提案している<sup>4),5)</sup>。本論文で提案する文書空間ナビゲーション手法では、上述のグラフ生成手法により生成される有向グラフを用い、ユーザが選択した文書または語に対し、関連語を提示する。提示する関連語の有用性を、提案手法を実装したプロトタイプシステムを用いたユーザ実験により確認する。

以下、2章で本研究で実現する文書空間ナビゲーションの要件をあげ、関連研究と比較し本研究の位置付けを示す。3章で、提案する文書空間ナビゲーション手法の詳細を述べる。

<sup>†1</sup> 筑波大学大学院図書館情報メディア研究科

Graduate School of Library, Information and Media Studies, University of Tsukuba

<sup>†2</sup> 産業技術総合研究所サービス工学研究センター

Center for Service Research, National Institute of Advanced Industrial Science and Technology  
本論文の内容は2009年7月のマルチメディア、分散、協調とモバイル(DICOMO2009)シンポジウムにて報告され、同プログラム委員長により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

4章で、提案手法によりユーザに提示される語の有用性を明らかにするための評価実験について述べ、5章で実験結果について考察し、6章でまとめと今後の展望について述べる。

## 2. 背景

### 2.1 探索的検索を円滑化するための文書空間ナビゲーション

近年、ユーザの試行錯誤的かつ能動的な関与を前提とする情報探索プロセスである探索的検索への関心が高まっている<sup>1)</sup>。Bates は、ユーザの情報探索行動は試行錯誤的で、探索過程で得られた情報がその後の探索行動を変化させることを指摘している<sup>2)</sup>。Marchionini は、ユーザの情報探索行動を分類し、探索的検索には、情報の発見、比較、除外、評価、理解などの側面があることを指摘している<sup>6)</sup>。検索語と文書のマッチングを行う従来の情報検索システムを用いて探索的探索を行うには、ユーザは検索語の修正や追加を繰り返す必要がある。しかし、ユーザが有効な検索語を選定できるためには、検索対象の文書集合の特徴、すなわち特徴的な用語や用語間の関連性を熟知していなければならない。White らは、ユーザが内容を予想できない文書集合（以下、未知の文書集合と記す）を対象とした探索的検索におけるナビゲーションの有用性を指摘している<sup>3)</sup>。

本研究では、探索的検索を円滑化する文書空間ナビゲーションの実現を目指す。このためには、文書集合の特徴を反映させ、多様な関連項目への円滑な遷移を可能とするナビゲーションが必要だと筆者らは考える。これらの要件について、以下に詳述する。

#### (1) 文書集合の特徴の反映

未知の文書集合のナビゲーションでは、シソーラスやオントロジなどの外部知識を用いるだけでなく、語の出現頻度や共起関係など、検索対象の文書集合から直接抽出した特徴量を用いることが必要である。これは、検索対象の文書集合と外部知識の間における内容や記述粒度の差により、ナビゲーション可能な範囲が制限されることを防ぐためである。

#### (2) 多様な関連項目への円滑な遷移

探索的検索の円滑化には、ユーザを多様な関連文書や関連語（以下、関連項目と総称する）へ遷移可能とする必要がある。これは、ユーザが情報の発見、比較、除外を容易に行えるようにするためである。また、ユーザによる検索語の修正や追加を必要とせずに関連項目間を円滑に遷移可能とする必要がある。これは、新たな検索語を想起するユーザの負担を軽減し、情報の評価や理解といった、より高次のプロセスにユーザが専念できるようにするためである。

### 2.2 関連研究

探索的検索に関するユーザの行動については、幅広い観点から分析が行われている。Madrid らは、複雑なリンクが構築されたハイパーテキストの内容をユーザが理解する際に、ページあたりのリンク数よりも、文書を閲覧する順序の影響が大きいことを報告している<sup>7)</sup>。このため、文書間に順序関係が見られない文書集合の閲覧において、ユーザが文書集合全体の内容や、意味的な順序関係を把握するには、適切なナビゲーションによる支援が有効であると考えられる。Amadiou らは、ハイパーテキストの内容の理解において、予備知識の少ないユーザは、ネットワーク型よりも階層型のリンク構造を持つハイパーテキストを用いたほうが知識を得やすいことを指摘している<sup>8)</sup>。しかし、文書間で内容の部分的な重複が多い新聞記事などでは、階層型のナビゲーションを実現することは難しい。このような文書集合に対しては、ユーザが文書を閲覧する順序に一定の統制を加えることのできる、有向グラフに基づくナビゲーションが有用であると筆者らは考える。また、Loizides らは、ユーザの文書閲覧における、画像の有無や文書の長さなどの影響を分析し、テキストのみを表示した場合に最も閲覧に要する時間が短いことを報告している<sup>9)</sup>。このことは、関連語や関連文書など、テキストのみを用いるナビゲーション手法が、シンプルでありながら一定の有用性を持つことを示唆している。

探索的検索の円滑化を目的とした文書空間ナビゲーションにおいて、文書集合の特徴の反映と多様な関連項目への円滑な遷移を両立させる手法は、これまで十分検討されてこなかった。Sun らは、ハイパーテキストにおけるトピック検出および文書分類を目的とし、出現語と文書間のリンク構造を用いて代表的なトピックを提示する手法を提案している<sup>10)</sup>。この手法では、文書集合の特徴をユーザに提示することに特化しており、ユーザが関連項目間を探索的に遷移することは想定されていない。宮崎らは、ユーザによる文書集合の俯瞰を目的とし、検索対象の文書集合に付与されたメタデータ中に出現する特徴語をユーザに一覧提示するナビゲーション手法を提案している<sup>11)</sup>。ユーザは検索語を入力することなく文書集合の特徴を把握できるが、メタデータ中に出現する語は一般的な語が多く種類も少ないため、詳細な探索には必ずしも適さない。酒井らは、Web ページを対象とした探索的検索の支援を目的とし、ユーザが気づきにくい関連情報の提示により、情報要求の変化の誘発を狙う検索インタフェースを提案している<sup>12)</sup>。この手法では、検索語の関連情報を Wikipedia<sup>\*1</sup> から抽出し提示することから、適合する項目が Wikipedia 上に存在しない Web ページに対し

\*1 <http://ja.wikipedia.org/>

では有効な提示ができないという問題がある。Buchanan らは、多数の文書の素早い閲覧を支援することを目的に、文書中の任意の選択箇所を基点とし、別の部分にある関連情報への遷移、および基点への復帰を可能とする双方向のナビゲーション手法を提案している<sup>13)</sup>。この手法では、関連の深い項目間における双方向の遷移が円滑化される。ただし、比較的長い文書を順を追って閲覧していくことが前提となっており、探索的に多様な項目間を遷移することは想定されていない。

本論文では、文書空間ナビゲーションにおいて、文書集合の特徴を反映させ、多様な関連項目への遷移を実現するため、検索対象の文書集合から抽出した共起語グラフ (word co-occurrence graph) を用いることを検討する。共起語グラフは、文書集合の特徴すなわち語の出現頻度と共起関係を反映しており、多くのノードに短い距離で到達可能な Small-world グラフ<sup>14)</sup> となることが知られている<sup>15)</sup>。筆者らは、無向グラフである共起語グラフの特性を損なわずに、ナビゲーションに適する有向グラフに変換する手法を提案し、新聞記事および不特定多数のユーザが投稿する質問回答サイトの記事に適用できることを確認している<sup>4)</sup>。

探索的検索における情報の発見、比較、除外には、特定の文書に頻出し、特定のトピックを表現する語 (以下、特定語: distinctive words と記す) および多様な文書に出現し、異なるトピック間を媒介する語 (以下、媒介語: intermediate words と記す) が有用と筆者らは考える。戸田らは、検索結果に含まれる特定語をユーザに提示するナビゲーション手法を提案している<sup>16)</sup>。この手法では、ユーザは提示語を用いて容易に検索結果を絞り込めるが、異なるトピックの発見に有用な語を提示することは考慮されていない。松尾らは、共起語グラフを用いて媒介語を抽出する手法を提案しているが、最短経路探索の計算量を課題としている<sup>17)</sup>。筆者らは、3.1 節で述べる反復度および文書頻度を用いる手法により、経路探索を行わず近似的に特定語と媒介語を抽出できることを確認している<sup>5)</sup>。特定語と媒介語の両方を用いる文書空間ナビゲーション手法は知られておらず、本論文で提案する手法の特徴となっている。

### 3. 出次数制約付き有向グラフを用いた文書空間ナビゲーション手法

本章では、出次数制約付き有向グラフ生成手法<sup>4)</sup> における語の抽出方法および関連項目間へのリンクの生成方法を概説したのち、提案する文書空間ナビゲーションの実現に不可欠なユーザ・インタフェースについて述べる。

本研究で実現する文書空間ナビゲーションの概要を図 1 に示す。図中の 印は特定語を、印は媒介語を表す。(1) 文書集合から 3.1 節で述べる方法により語を抽出し、共起語グ

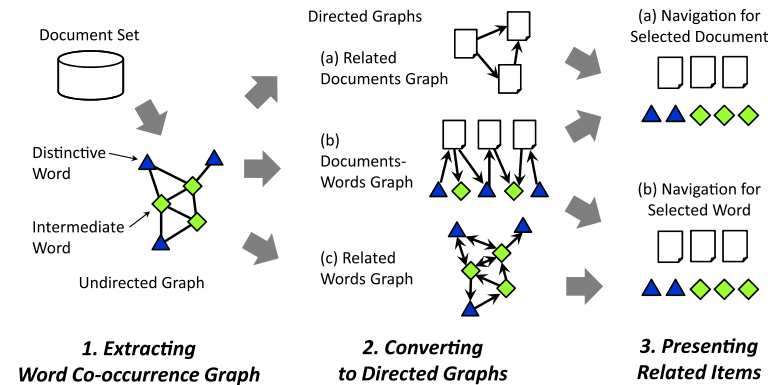


図 1 文書空間ナビゲーションの実現方法

Fig. 1 The flow of processing in proposal document navigation method.

ラフを構築する。(2) 3.2 節で述べる手法により、共起語グラフを (a) 関連文書グラフ、(b) 語—文書グラフ、(c) 関連語グラフに変換する。(3) ユーザが選択した任意の (a) 文書または、(b) 語に対し、生成したグラフを用いて関連文書および関連語を提示する。ユーザは、3.3 節に示すユーザ・インタフェースを通じ、提示された関連文書または関連語の選択を繰り返すことにより、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文書空間内を探索できる。

#### 3.1 文書空間ナビゲーションに用いる語の抽出

文書空間ナビゲーションに用いる語は、反復度 (adaptation)  $df_2/df$  および文書頻度 (document frequency)  $df$  を用いて抽出する。 $df_2$  は、その語が 2 回以上出現する文書数である。 $df_2/df$  は文法的機能を持たず語彙的意味を表す内容語 (content word) において高く<sup>18)</sup>、 $df$  はある分野で共通して用いられる語において高くなる<sup>19)</sup> ことが知られている。

提案手法では、文書集合内で  $df_2/df$  が高い語を、その文書集合における特定語と見なす。これは、特に  $df_2/df$  が高い具体的な語の提示により、ユーザが情報の発見、比較、除外をしやすくなると考えられるためである。また、文書集合内で  $df$  が中程度となる語を、その文書集合における媒介語と見なす。これは、文書集合内である程度共通して使われる語が、異なるトピック間を媒介すると見なせるためである。提案手法における語の区分 (division) を図 2 に示す。区分 I は特定語、区分 II は媒介語を抽出するための区分である。区分 III、IV は、その他の語を抽出するための実装上の区分である。

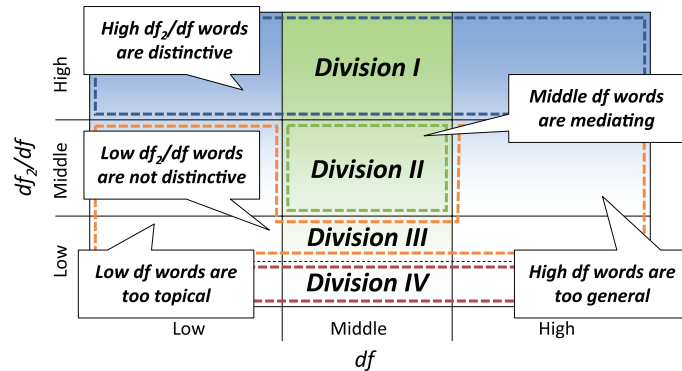


図 2 文書空間ナビゲーションのための語の区分  
Fig.2 The division of words for document navigation.

表 1 語の重み付けに用いる反復度および文書頻度の閾値および重み  $w$   
Table 1 Thresholds of  $df_2/df$  and  $df$  for weighting words, and  $w$ .

区分	反復度の範囲	文書頻度の範囲	重み $w$
I (特定語)	$df_2/df \geq 0.6$	$df_2 \geq 3$	10
II (媒介語)	$df_2/df \geq 0.35, df_2/df < 0.6$	$df \geq 10, df_2 < 19$	1
III	$df_2/df \geq 0.1$	$df \geq 4$ (ただし区分 I, II 以外)	0.1
IV	すべて (ただし区分 I, II, III 以外)		0.01

$df < 3$  となる語は除く。

本論文で用いる  $df_2/df$  および  $df$  の閾値、および抽出された語に与える重み  $w$  を、表 1 に示す。ここでは、2 文字以上の漢字またはカタカナからなる文字列、3 文字以上の英数字からなる文字列を語とする。これらの閾値は、308 件のブログ記事からなる文書集合を用いた予備実験により決定している<sup>4)</sup>。文書集合における各語の  $df_2/df$  および  $df$  を区分 I, II, III, IV の順に評価し、最初に一致した区分の重み  $w$  を語の重みとする。

### 3.2 出次数制約付き有向グラフによる関連項目間へのリンク生成

提案手法では、3 種類の有向グラフ、すなわち関連語グラフ、関連文書グラフ、語—文書グラフを生成する。前節で決定した各語の重み  $w$  を用いて、共起関係にある語間、同一の語が出現する文書間、語—文書間の関連度を算出し、これをリンク生成の優先度とする。設定した出次数の範囲内で、リンク生成の優先度の上位の項目間に有向リンクを生成する。

### 関連語グラフ

基点となる語と、それ以外の語との間の関連度を算出し、関連度が上位となる語へのリンクを生成する。関連度は、リンク先の候補となる語の重みに、基点となる語とリンク先の候補となる語の共起文書数を乗算した値とする。基点となる語の重みは考慮しないため、2 語間の関連度は非対称になる。

### 関連文書グラフ

基点となる文書と、それ以外の文書との間の関連度を算出し、関連度が上位となる文書へのリンクを生成する。関連度は、基点となる文書と、リンク先の候補となる文書との間で共起する語集合に含まれる語の重みを総和した値とする。

### 語—文書グラフ

語—文書グラフは、文書から語へのリンクを有するグラフと、語から文書へのリンクを有するグラフで構成する。

まず、基点となる文書と、その文書に出現する語との間の関連度を算出し、関連度が上位となる語へのリンクを生成する。関連度は、基点となる文書において出現する語の重みに、その語の文書頻度を乗算した値を総和した値とする。次に、基点となる語と、その語が出現する文書との間の関連度を算出し、関連度が上位となる文書へのリンクを生成する。関連度は、基点となる語が出現する文書において出現する語の重みを総和した値とする。基点となる語の重みは考慮しない。

### 3.3 ユーザ・インタフェース

提案する文書空間ナビゲーション手法を実装したプロトタイプシステム（以下、提案システムと記す）におけるユーザ・インタフェースの特徴を説明する。特に、本手法の特徴である、関連語および関連文書間の遷移について述べる。

筆者らが提案している有向グラフ生成手法では、ノードあたりの最大出次数が 3 から 10 の範囲において、Small-world 性、リンクの多様性、誘導性を両立させた、ナビゲーションに適するグラフを生成可能であることを確認している<sup>4)</sup>。この知見に基づき、本システムでは最大出次数を 8 として実装した。すなわち、最大で 8 語の関連語および 8 件の関連文書を提示する。なお、本論文では、ユーザ・インタフェースについての直接的な評価は行わないが、人間の短期記憶可能な項目数が  $7 \pm 2$  であることが知られており<sup>20)</sup>、一般にナビゲーションの項目数を  $7 \pm 2$  程度とすることが多い。

提案システムにおける、任意の文書または語に対する関連語および関連文書の提示画面を図 3 に示す。ユーザが選択した文書（図中 A）または語（図中 B）に対し、前節で述べた

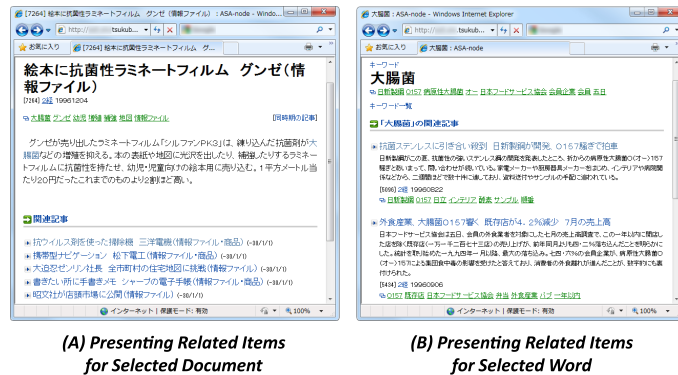


図 3 提案システムにおける任意の文書または語に対する関連語および関連文書の提示画面  
Fig. 3 Screen shot of presenting related words and documents in proposal system.

方法により関連付けられた関連語および関連文書を提示する。なお、任意の語に対する関連項目の提示画面（図中 B）では、関連語 8 語および関連文書 8 件の提示に加え、各々の関連文書に対し各 8 語の関連語を提示する。このため、1 画面内に提示する関連項目は最大 80 (= 8 + 8 + 8 × 8) 件となり、多様な関連項目が一覧できる。

このほか、提案システムには、すべての特定語を  $df$  の降順で一覧表示する画面、簡易な全文検索機能<sup>\*1</sup>、および文書のタイトルを 20 件ずつ日付順に一覧する機能も実装してある。また、初期画面では文書のタイトルを日付の降順で 13 件表示する。ユーザは最初に、これらの機能を用いて、探索の基点となる任意の語または文書を選択する。そこで提示された関連語または関連文書を選択すると、その語または文書に対し、同様に関連語および関連文書が提示される。提示された項目の選択を繰り返すことにより、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、関連項目間を円滑に遷移できる。

#### 4. 評価

本章では、提案する文書空間ナビゲーションの探索的検索における有用性を評価する。特に、提案手法の核となる関連語の提示の有用性をユーザ実験により確認する。以下、システ

\*1 ユーザによる検索語の想起への影響を避けるため、検索語への適合度を考慮せず、部分文字列が一致した文書を新着順に表示する。

ムが提示し、かつユーザが選択した語をシステム提示語、ユーザが想起して入力し検索に用いた語をユーザ入力語と記し、これらをユーザ使用語と総称する。ユーザにより選択された提示語（システム提示語）は一定の有用性を示すと見なせるため、本論文では、システムが提示した語のうち、ユーザが選択した語に限って評価する。

##### 4.1 実験の概要

本実験では、2.1 節であげた、探索的検索の円滑化を目的とする文書空間ナビゲーションの 2 要件に対応する以下の 2 項目を評価する。

###### (1) 文書集合の特徴の反映

ユーザが有効な情報探索を行うには、ユーザ使用語の特徴、すなわち語の出現頻度や共起関係の分布が、文書集合全体の特徴と適合していることが重要である。提案する文書空間ナビゲーションの使用により、文書集合の特徴を反映させた探索をユーザが行えることを確認する。システム提示語の文書頻度  $df$  および共起語数  $deg$ <sup>\*2</sup>が、文書集合における分布と同様の分布を示せば、文書集合の特徴を反映させた探索をユーザが行えたとする。

###### (2) 多様な関連語の提示

提案する文書空間ナビゲーションが、多様な関連語を提示でき、かつ探索に有用な語を多く提示できることを確認する。システム提示語の異なり語数が、ユーザ入力語の異なり語数を上回れば、多様な関連語を提示できたといえる。また、提案手法では、3.1 節に示す語の区分 I（特定語）および区分 II（媒介語）の語を、その他の区分の語よりも優先してユーザに提示することから、異なり語数を語の区分ごとと比較し、区分 I および区分 II において、システム提示語の異なり語数が、ユーザ入力語の異なり語数を上回れば、探索に有用な語を多く提示できたとする。

##### 4.2 実験に用いる文書集合

本実験では、新聞の経済面に掲載された 1 年分の記事<sup>\*3</sup>を用いた。記事の見出しと本文を一体として 1 文書と扱い、語および文書の関連付けを行った。文書数は 7,770、抽出された語の数は 20,103 である<sup>\*4</sup>。

抽出された語を  $df$ ,  $df_2/df$ , および  $deg$  の各値で順序付けたときの順位相関を表 2 に示

\*2 ある語と同一文書内で共起する語を、文書集合全体での異なり数で数える。共起語グラフにおけるノードの次数に相当する。

\*3 「朝日新聞記事データ集学術研究用 1996 年版」を用いた。

\*4 実験に用いた環境 (Xeon E5440 2.83 GHz) においては、語の抽出に約 16 秒、グラフ生成に約 4 分を要した。画面表示に要する時間は約 1 秒である。

表 2 実験に用いた文書集合における各指標間の相関行列 (スピアマンの順位相関係数)

Table 2 A correlation coefficient matrix between indexes.

	<i>df</i>	$df_2/df$	<i>deg</i>
<i>df</i>	1.000	0.098	<b>0.837</b>
$df_2/df$		1.000	0.063
<i>deg</i>			1.000

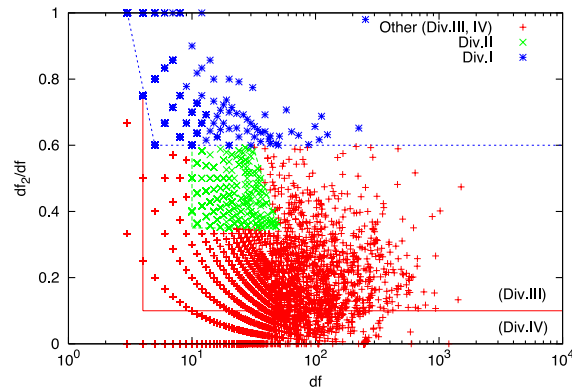


図 4 実験に用いた文書集合における語の文書頻度 *df*, 反復度  $df_2/df$  の分布

Fig. 4 Distribution of document frequency (*df*) and adaptation ( $df_2/df$ ) of all extracted words. The areas of divisioning words in proposal system is shown.

す。また、この文書集合における *df* および  $df_2/df$  の分布を図 4 に示す。*df* と *deg* は相関が高く、*df* が高い語は *deg* も高い傾向がある。*df* と  $df_2/df$  の相関はほとんどないが、*df* が高い語の  $df_2/df$  はやや低くなる傾向がある。

この文書集合に提案手法を適用して得られた語の区分結果および各区分における *df*,  $df_2/df$ , *deg* の平均値を表 3 に示す。本システムが抽出した語のうち、約 2% が区分 I, 約 2% が区分 II として抽出された。媒介語に相当する区分 II は、特定語に相当する区分 I と比較し *deg* の平均値が約 2 倍である。

抽出された語の具体例を以下に示す。区分 I では、経済統計の名称や指標、「以遠権」「免震装置」などの語が抽出された。区分 II では、「勧告」「補助」「受注額」などの語が抽出された。企業名や省庁名などの固有名詞は、主に  $df_2/df$  の高低によって区分 I と II に分かれて抽出された。分割民営化や合併に関する記事で繰り返し言及される「NTT」「新王子(製

表 3 実験に用いた文書集合における語の区分結果および各区分における各指標の平均値

Table 3 Result of word division and means of indexes in each division.

区分	語数	<i>df</i>	$df_2/df$	<i>deg</i>
I (特定語)	405 (2.0%)	10.6	<b>0.751</b>	452.5
II (媒介語)	433 (2.2%)	19.4	0.438	<b>950.2</b>
III	6,915 (34.4%)	17.2	0.139	767.8
IV	12,350 (61.4%)	<b>10.1</b>	0.053	527.9
(全体)	20,103	17.2	0.139	767.6

表 4 実験参加者に課した情報探索のタスク

Table 4 Details of information retrieval tasks using in the experiment.

タスク	実験参加者に提示した指示文
1	この年に猛威を振るった「O157」に関連して、急によく売れるようになった商品や、対策のために売り出された新商品などを網羅的に列挙してください。 また、その商品について書かれた記事もあげてください。
2	この年の経済分野での三大ニュースと思われるトピックを 3 つ見つけて、漢字 4~6 字程度で表現してください。 また、それぞれのトピックを象徴すると思われる記事を 3 件程度ずつあげてください。

1 名以上の分析対象ユーザにより検索語として使用された語句を太字で示した

紙)」は区分 I に、新商品発売や決算発表、市況、主要企業アンケートなどの記事で 1 回のみ出現することが多い「サントリー」「花王」などは区分 II になった。 $df_2/df$  が高いが *df* も高いため区分 III になった語として「大蔵省」「住専」など、*df* が低いが  $df_2/df$  も低いため区分 IV になった語として「財政投融资」「機械受注」などがあつた。

#### 4.3 実験の実施方法

本実験は、情報系の講義を履修する学部 1 年生を対象に、表 4 に示す情報探索タスクを課題として実施した。実験参加者は、表 5 に示す 70 名である。このうち、探索時間が 1 分未満、または操作順序に誤りがあった 24 名を除く 46 名を、本論文における分析対象ユーザ (以下、ユーザと記す) とした。

実験の手順を説明する。本実験では、2 種類のシステム、すなわち 3.3 節で説明した提案システム、および、提案システムから関連項目の提示機能および特定語の一覧画面を除いたシステム (以下、従来システムと記す) を用意した。実験参加者は、提案システムを使用するグループ A および従来システムを使用するグループ B にランダムに振り分けた。グループ A では、システム提示語およびユーザ入力語を使用する。グループ B では、ユーザ入力

表 5 実験におけるユーザ数およびグループへの振り分け  
Table 5 Detail of users and its grouping in the experiment.

グループ	探索方法	ユーザ数		
		分析対象	除外	計
A	システム提示語を選択 および検索語を想起して入力	23	13	36
B	検索語を想起して入力	23	11	34
	計	46	24	70

表 6 実験におけるユーザ使用語の異なり語数  
Table 6 Distinct number of words used in each group and in each task.

タスク	グループ	使用語の種類	異なり語数		計
			$df \geq 3$	$df < 3$	
1	A	システム提示語	42	0	42
	B	ユーザ入力語	19	23	42
2	A	システム提示語	122	0	122
	B	ユーザ入力語	77	28	105

タスク指示文中に出現する語は除いた。

語のみを使用する。

実験参加者に課した情報探索タスクについて説明する。本実験では、情報探索行動の評価で一般的である主題検索 (subject search)<sup>21),22)</sup> を対象に、探索の範囲や詳細度が異なる 2 種類のタスクを設定した。タスク 1 は、特定の主題に関して主題検索を行うタスクである。タスク 2 は、文書集合に含まれる主題を発見した後、その主題に関して主題検索を行い、主題の重要度を判断する複合的なタスクである。実験参加者には、タスク 1、タスク 2 の順に、各 15 分程度の時間をかけて探索するよう指示した。なお、探索の終了は参加者各自の判断による。

#### 4.4 実験結果

本節では、ユーザ使用語について述べたのち、2.1 節であげた、探索的検索の円滑化を目的とする文書空間ナビゲーションの 2 要件に対応し、実験結果を分析する。

本実験におけるユーザ使用語の統計を表 6 に示す。ユーザが能動的に選択または想起して入力した語に着目して分析するため、タスクの指示文中に出現する語句は除外した。グループ A では、システム提示語のみを分析対象とする。グループ B のユーザ入力語における表記揺れは、システム提示語と一致するよう修正した。異なり語数は、同一タスクではグループ間の差は小さい。グループ B では、 $df < 3$  の語も使用され、特にタスク 1 では

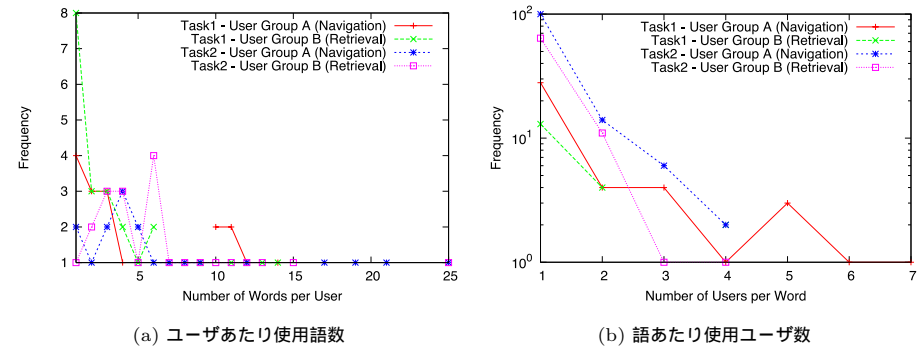


図 5 実験におけるユーザあたり使用語数および語あたり使用ユーザ数の分布  
Fig. 5 Distributions of number of words used in the experiment.

$df \geq 3$  の語より多かった。以下、提案システムで提示する  $df \geq 3$  の語を分析対象とする。

ユーザあたり使用語数の分布を図 5 (a) に示す。また、語あたり使用ユーザ数の分布を図 5 (b) に示す。ユーザあたり使用語数の最大は、タスク 1 では両グループとも 14、タスク 2 では同様に 25 となり、同一タスクではグループ間に差は見られない。語あたり使用ユーザ数の最大は 7 であり、ユーザ間での使用語の重なりは小さい。

#### (1) 文書集合の特徴の反映

提案する文書空間ナビゲーションの使用により、文書集合の特徴を反映させた探索をユーザが行えることを確認するため、システム提示語における文書頻度  $df$  および共起語数  $deg$  の分布を、文書集合における分布と比較した。

ユーザ使用語における  $df$  の分布を文書集合における分布と比較した結果を図 6 に示す。横軸は  $df$ 、縦軸は  $df$  の値が  $k$  である語の数である。文書集合全体では、語の  $df$  はベキ分布を示し、 $df$  の最大は  $10^3$  付近である。いずれのタスクでも、システム提示語の  $df$  の分布は文書集合と同様、すなわち  $df$  の低い語の使用頻度が高く、 $df$  の高い語の使用頻度は少ないという分布を示した。なお、タスク 2 におけるユーザ入力語のピークは  $10^2$  付近にあり、 $df$  の値が高い語が、文書集合全体では少ないにもかかわらず多用される傾向があることが明らかになった。

ユーザ使用語における  $deg$  の分布を文書集合における分布と比較した結果を図 7 に示す。横軸は  $deg$ 、縦軸は  $deg$  の値が  $k$  である語の数である。文書集合全体では、語の  $deg$  は山型の分布を示し、 $deg$  が  $10^3$  程度の語が多い。ユーザ使用語における  $deg$  の分布は、タスク

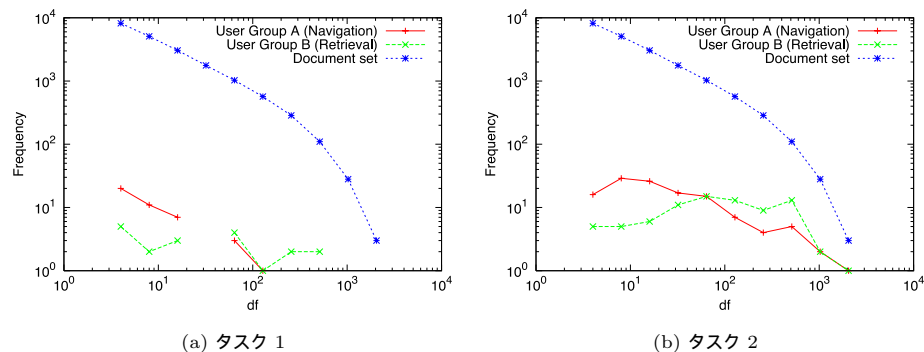


図 6 実験におけるユーザ入力語とシステム提示語の文書頻度  $df$  の分布  
Fig. 6 Distributions of  $df$  of words used in the experiment.

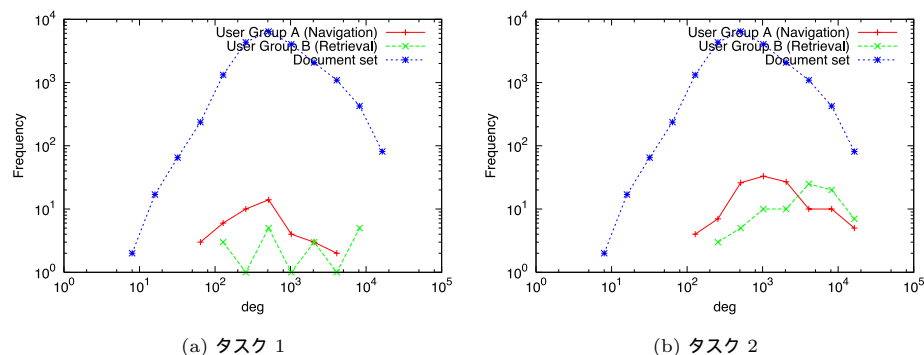


図 7 実験におけるユーザ入力語とシステム提示語の共起語数  $deg$  の分布  
Fig. 7 Distributions of  $deg$  of words used in the experiment.

1 のユーザ入力語を除き、いずれも山型の分布を示した。いずれのタスクでも、システム提示語の  $deg$  の分布は文書集合と同様、すなわちピークが  $10^3$  程度となる分布を示した。なお、タスク 2 におけるユーザ入力語のピークは  $10^4$  付近にあり、 $deg$  の値が高い語が、文書集合全体では少ないにもかかわらず多用される傾向があることが明らかになった。

(2) 多様な関連語の提示

提案する文書空間ナビゲーションが、多様な関連語を提示できることを確認するため、システム提示語とユーザ入力語の異なり語数を比較した。また、探索に有用な語を多く提示で

表 7 ユーザ使用語における語の区分ごとの異なり語数および  $df$ ,  $df_2/df$ ,  $deg$  の平均値  
Table 7 Averaged  $df$ ,  $df_2/df$  and  $deg$  of words in each division.

グループ 語の区分	語数	タスク 1			タスク 2			
		$df$	$df_2/df$	$deg$	語数	$df$	$df_2/df$	$deg$
A (全体)	42	10.1	0.251	478.9	122	64.6	0.457	1,780.6
I (特定語)	6	7.8	0.711	402.3	25	26.2	0.701	1,031.9
II (媒介語)	3	10.3	0.388	545.7	42	21.0	0.419	1,236.5
III, IV	33	10.5	0.155	486.7	55	115.4	0.376	2,536.3
B (全体)	19	80.7	0.231	2,019.0	77	144.9	0.264	3,731.8
I (特定語)	1	10.0	0.600	366.0	1	4.0	0.750	321.0
II (媒介語)	1	10.0	0.400	261.0	3	24.0	0.384	1,444.3
III, IV	17	89.1	0.199	2,219.6	73	151.8	0.253	3,872.5

きることを確認するため、3.1 節に示す語の区分別に異なり語数を比較した。

システム提示語とユーザ入力語における異なり語数を語の区分ごとに比較した結果を表 7 に示す。全体では、いずれのタスクでも、システム提示語の異なり語数は、ユーザ入力語の異なり語数の約 2 倍だった。語の区分ごとでは、いずれのタスクでも、システム提示語における区分 I および II の異なり語数は、同一タスクかつ同一区分のユーザ入力語の異なり語数を上回った。また、区分 III および IV の異なり語数の割合に着目すると、タスク 2 では、ユーザ入力語の 94.8% と大きいのが、システム提示語では 45.1% と小さい。

参考として、 $df$ ,  $df_2/df$ ,  $deg$  の平均値を表 7 にあわせて示す。いずれのタスクでも、システム提示語はユーザ入力語と比較し、 $df$  の平均が低く、 $df_2/df$  の平均が高く、 $deg$  の平均が低い。タスク 2 では、タスク 1 よりも  $df$  および  $deg$  が高くなる傾向があるが、タスク 2 におけるシステム提示語の  $df$  および  $deg$  の平均は、いずれもタスク 1 におけるユーザ入力語の平均を上回ることはなかった。これは、提案システムが、探索的検索における関連文書の発見に有用と考えられる  $df$  および  $deg$  が中程度である語を提示でき、かつユーザにより選択されたことを意味する。

ユーザ使用語の文書頻度  $df$  および反復度  $df_2/df$  の分布を図 8 に示す。いずれのタスクでも、システム提示語は、 $df$  が  $10^1$  付近で、 $df_2/df$  が高い領域を中心に分布しているのに対し、ユーザ入力語は、 $df$  が  $10^2$  付近で、 $df_2/df$  が低い領域を中心に分布している。また、これらの中間の領域にある語は、両グループのユーザによって使用されていた。

以上の結果から、提案する文書空間ナビゲーションの使用により、文書集合の特徴を反映させた探索をユーザが行えること、および、提案する文書空間ナビゲーションが多様な関連語を提示できることを確認した。



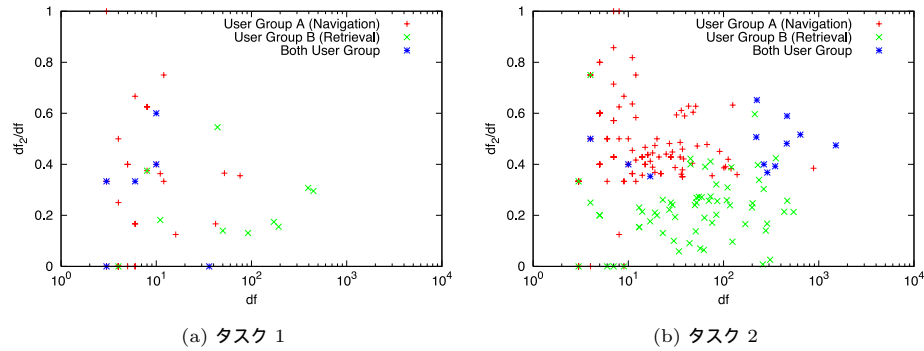


図 8 実験におけるユーザ使用語の文書頻度  $df$  および反復度  $df_2/df$  の分布  
 Fig. 8 Distributions of  $df$  and  $df_2/df$  of words used in the experiment.

## 5. 考 察

本章では、探索的検索の円滑化における提案手法の有用性について考察する。まず、4章で行った評価実験の結果に基づき、システム提示語の有用性について考察する。次に、実験参加者による主観評価について考察する。

### 5.1 システム提示語の有用性

4.4 節で詳述したように、(1) システム提示語の文書頻度  $df$  および共起語数  $deg$  が文書集合における分布と同様の分布を示し、(2) システム提示語の異なり語数がユーザ入力語の異なり語数を上回り、かつ 3.1 節に示す語の区分 I および区分 II において、システム提示語の異なり語数がユーザ入力語の異なり語数を上回った。これらの結果から、提案する文書空間ナビゲーションにおける関連語の提示が、2.1 節であげた文書空間ナビゲーションの要件を満たすことを確認した。

語の各区分における使用ユーザ数が上位の語を表 8 に示す。各タスクで両グループに共通する使用語を除くと、システム提示語を用いるグループ A では、ユーザ入力語を用いるグループ B よりも、具体的な内容を表す語が多い。区分 III, IV の語に着目すると、タスク 1 でグループ B の 2 名が使用した「発売」の  $df$  は 394、 $df_2/df$  は 0.307、 $deg$  は 4,782 であり、実験に用いた文書集合中では一般的な語であり探索に有用な語とはいえない。これに対し、グループ A の 5 名が使用した「弁当」の  $df$  は 6、 $df_2/df$  は 0.333、 $deg$  は 231 である。 $df_2/df$  は同程度でありながら、 $df$  および  $deg$  の差が大きい。このような傾向は、本

表 8 ユーザ実験におけるユーザ使用語および使用ユーザ数  
 Table 8 A list of words and numbers of users per words.

タスク	1		2	
	A	B	A	B
グループ	システム提示語	ユーザ入力語	システム提示語	ユーザ入力語
I (特定語)	ミドリ十字 日新製鋼 健保組合	5 3 3	日新製鋼 NTT 第三分野 国民総支出	1 4 3 2
II (媒介語)	ヨーグルト ハム 危機管理	4 2 1	ヨーグルト 株会社解禁 整備新幹線 ベンチャー	1 3 2 2
III, IV	大腸菌 病原性大腸菌 弁当 前年比五割増 消費回復 錠剤	7 6 5 5 3 3	大腸菌 食中毒 鮮度 発売 感染症 病原性大腸菌	4 4 2 2 2 2
			改革 景気回復宣言 金融行政 企業 個人消費 経済審	4 3 3 3 3 2
			金融検査監督庁 経済協力 ビッグバン 補正予算 改革 事件 大蔵省 景気 消費税 規制緩和	1 2 1 1 4 3 2 2 2 2

$df < 3$  の語、およびタスク指示文中に出現する語は除いた。

実験で全般的に観察された。一方、探索に有用と考えられる、 $df$  が低く  $df_2/df$  が高い語は、システム提示語に多く含まれていた。これは、ユーザの検索語想起において、 $df$  や  $deg$  が低い語の想起が困難であることを示唆している。同時に、システム提示語とユーザ入力語の違いによらず、ユーザが検索に使用する語の  $df_2/df$  は、ある程度高かった。これは、 $df_2/df$  が高い、すなわち具体的な内容を表す語の識別がユーザにとって容易であることを示唆している。このため、 $df_2/df$  が高い語を提示することで、ユーザが違和感なく語を選択でき、再検索を促すことができたと考えられる。

本実験では、探索の終了はユーザの判断に委ねたため、ユーザ間での探索時間の差は大きい。本論文では、タスク 1 およびタスク 2 の両方で、各 1 分以上の探索を行ったユーザに限定して分析している。多くのユーザは、15 分から 20 分程度で探索を終了していたが、タスク 2 に 1 時間以上かけて取り組んだユーザも数名いた。ユーザごとの異なり到達文書数の最大は 40、平均は 11.3 から 15.0 の範囲であり、タスクやグループによる差はほとんどなかった。ユーザごとの探索時間の差が大きかった背景として、ユーザによるシステム提示語の使用に、手がかりを求めて次々に別の語を使用していく場合と、文書の閲覧に時間をかけながら補助的にシステム提示語を使用する場合とがあったことが考えられる。

### 5.2 実験参加者による主観評価

表 5 に示した実験参加者 70 名による、提案システムの機能および関連語の提示方法に対

表 9 提案システムの機能に対する実験参加者の主観評価  
Table 9 Subjective evaluations for functions of proposal system.

機能	関連語の提示		関連記事の提示		特定語の一覧		全文検索	
観点	平均 (標準偏差)		平均 (標準偏差)		平均 (標準偏差)		平均 (標準偏差)	
使いやすさ	0.32	(0.62)	<b>0.33</b>	(0.65)	0.08	(0.71)	-0.07	(0.64)
探索の始めやすさ	0.38	(0.64)	0.23	(0.65)	0.26	(0.67)	<b>0.04</b>	(0.65)
探索の続けやすさ	0.26	(0.72)	0.30	(0.72)	0.06	(0.69)	-0.13	(0.68)
提示内容の意外性	<b>0.49</b>	(0.56)	0.29	(0.70)	<b>0.30</b>	(0.73)	-0.12	(0.70)

表 10 関連語の提示方法に対する実験参加者の主観評価  
Table 10 Subjective evaluations for presentation methods of related words.

観点	内容の適切性		分量の適切性		提示方法の適切性		提示の必要性	
語の提示方法	平均 (標準偏差)		平均 (標準偏差)		平均 (標準偏差)		平均 (標準偏差)	
記事に対する関連語の提示	0.23	(0.51)	<b>0.27</b>	(0.57)	0.28	(0.58)	<b>0.42</b>	(0.61)
本文中での関連語リンク	<b>0.26</b>	(0.62)	0.08	(0.63)	<b>0.29</b>	(0.65)	0.19	(0.65)
語に対する関連語の提示	0.22	(0.56)	0.09	(0.61)	0.21	(0.54)	0.23	(0.64)
特定語の一覧	0.11	(0.61)	0.12	(0.63)	0.14	(0.62)	0.26	(0.62)

する主観評価を行った。提案システムに実装した 4 つの機能について、4 つの観点で 4 段階評価を行った結果を表 9 に示す。回答の選択肢「そう思わない」「どちらかといえばそう思わない」「どちらかといえばそう思う」「そう思う」に対応する得点を  $-1, -0.5, 0.5, 1$  とし、平均および標準偏差を算出した。関連語の提示機能は、すべての観点で高い得点を得ている。特に、提示内容の意外性では他の機能より得点が高い。特定語の一覧機能は、探索の始めやすさ、提示内容の意外性については得点が高い一方、使いやすさ、探索の続けやすさでは得点が低い。全文検索機能は、否定的な評価が大勢を占めているが、探索の始めやすさについては一定数の肯定的な評価があった。次に、提案システムで実装した 4 種類の語の提示方法について、上述と同様に 4 段階評価を行った結果を表 10 に示す。記事に対する関連語の提示は、すべての観点で高い得点を得ている。特に、提示の必要性については他の提示方法より得点が高い。本文中での関連語リンクおよび語に対する関連語の提示は、分量の適切性で得点が低い。特定語の一覧は、提示の必要性を除く各観点で得点が低い。

提案システムの操作感に関する自由記述では、システム提示語の妥当性には不満があるものの、関連語の提示機能はあったほうがよい、とする意見が複数あった。具体的には、「関連する記事を調べるといよりは、新しいことの発見の方が強い気がする。タスク 2 のようなものだと関連語はほとんど意味がなく結局は自分で記事を見つけていくようになっていた」「必要もないのに(いい意味で)興味をひかれてしまう」「目的がはっきりしていない

ネットサーフに便利だ。関連語の提示は、目的ある調べごとには、向かない可能性もある。クリックしてみることで、話がそれていきやすいから(原文どおり)という意見があった。提案手法で提示する関連語は、必ずしも類似度の高い文書に共通して出現する語に限定されないため、ユーザに戸惑いが生じたケースもあったと考えられる。また、「三大ニュースといわれると、どんな単語を調べればいいのか分からないし、重要度が分からなくて結局いろいろ記事を見て、しかもよく分からないという結果になってしまった(同)とする感想もあったが、これはユーザが文書集合の内容を把握できていない状態でも、多様な文書にアクセスできたことを意味している。一方、すべての特定語を一覧表示する画面については、特定語は 405 あり、数が多く見にくいという意見が複数あった。また、重要度が分からない、50 音順で表示してほしい、といった意見もあったが、これは  $df$  の降順で表示していることをユーザには示さなかったためと考えられる。このことは、文書から独立して特定語だけを一覧提示しても、ユーザには分かりにくいということを示唆しており、文書を特定できない検索初期の段階における語の提示方法については、さらに検討が必要である。

## 6. おわりに

本論文では、ユーザによる検索語の入力なしで文書集合内を遷移可能とする、出次数制約付き有向グラフ生成手法を用いた文書空間ナビゲーション手法を提案した。提案手法は、文書集合から抽出される語の共起関係に基づき、ユーザが選択した語または文書に対し、関連語および関連文書を提示する手法である。

探索的検索を円滑化する文書空間ナビゲーションにおいては、ナビゲーション可能な範囲が制限されることを防ぐため、語の共起関係や出現頻度など、検索対象の文書集合から直接抽出した特徴量のみを用いることが必要である。ユーザが情報の発見、比較、除外を容易に行えるようにするため、多様な関連語や関連文書へ遷移可能とすることが必要である。また、ユーザが情報の評価や理解といった高次のプロセスに専念できるようにするため、検索語の入力なしで文書空間内を探索できることが必要である。

本論文では、提案する文書空間ナビゲーションにおける関連語の提示が、(1) 文書集合の特徴を反映させ、(2) 多様な関連語を提示できることを確認するため、ユーザ実験を行った。その結果、(1) システム提示語の文書頻度  $df$  および共起語数  $deg$  が文書集合における分布と同様の分布を示し、(2) システム提示語の異なり語数がユーザ入力語の異なり語数を上回り、かつ 3.1 節に示す語の区分 I および区分 II において、システム提示語の異なり語数がユーザ入力語の異なり語数を上回った。これらの結果から、提案する文書空間ナビゲーション

ンにおける関連語の提示が、文書空間ナビゲーションの要件を満たすことを確認した。

本論文における評価実験では、実験参加者が内容を予想しにくいと考えられる文書集合として、新聞記事を用いた。提案手法は、文書集合から抽出される共起語グラフの Small-world 性を仮定している。このため、共起語グラフの特性が大きく異なる文書集合への適用は、今後の課題である。提案手法を電子掲示板やブログなどに適用する際の課題としては、表記の揺れへの対応、文書を適切な長さにするなどの前処理が必要になると考えられる。今後、本論文では詳細な評価を行わなかった関連文書の提示、および評価実験において記録されたユーザの探索経路や閲覧文書などの詳細な評価を行う予定である。

謝辞 本研究は科研費(21500091)の助成を受けたものである。ここに記して謝意を示す。

### 参 考 文 献

- 1) White, R., Kales, B., Ducker, S. and Schraefel, M.: Supporting exploratory search, *Comm. ACM*, Vol.49, No.4, pp.36–39 (2006).
- 2) Bates, M.J.: The design of browsing and berry-picking techniques for the online search interface, *Online Review*, Vol.13, No.5, pp.407–424 (1989).
- 3) White, R.W., Marchionini, G. and Muresan, G.: Editorial: Evaluating exploratory search systems, *Inf. Process. Manage.*, Vol.44, No.2, pp.433–436 (2008).
- 4) 島田 諭, 福原知宏, 佐藤哲司: 文書空間ナビゲーションのための出次数制約付き有向グラフ生成手法, *情報処理学会論文誌: データベース*, Vol.3, No.2, pp.111–122 (2010).
- 5) 島田 諭, 福原知宏, 佐藤哲司: 文書集合の特性を考慮した包括的 Web ナビゲーション, *情報処理学会マルチメディア, 分散, 協調とモバイルシンポジウム DICOMO2009 論文集*, 1B-3, pp.47–54 (2009).
- 6) Marchionini, G.: Exploratory search: From finding to understanding, *Comm. ACM*, Vol.49, No.4, pp.41–46 (2006).
- 7) Madrid, R.I., Oostendorp, H.V. and Melguizo, M.C.P.: The effects of the number of links and navigation support on cognitive load and learning with hypertext: The mediating role of reading order, *Computers in Human Behavior*, Vol.25, pp.66–75 (2009).
- 8) Amadiou, F., Tricot, A. and Mariné, C.: Interaction between prior knowledge and concept-map structure on hypertext comprehension, coherence of reading orders and disorientation, *Interacting with Computers*, Vol.22, No.2, pp.88–97 (2010).
- 9) Loizides, F. and Buchanan, G.: An empirical study of user navigation during document triage, *Proc. 13th European Conference on Research and Advanced Technology for Digital Libraries, ECDL'09*, Berlin, Heidelberg, Springer-Verlag, pp.138–149 (2009).
- 10) Sun, C., Gao, B., Cao, Z. and Li, H.: HTM: A topic model for hypertexts,

*Proc. Conference on Empirical Methods in Natural Language Processing, EMNLP '08*, Morristown, NJ, USA, Association for Computational Linguistics, pp.514–522 (2008).

- 11) 宮崎陽司, 河野 泉: 検索対象の特徴を表すキーワードを用いた文書検索ナビゲーション(オフィスインフォメーションシステム, グループウェア及び一般), *電子情報通信学会技術研究報告, OIS, オフィスインフォメーションシステム*, Vol.108, No.53, pp.7–12 (2008).
- 12) 酒井哲也, 小山田浩史, 野上謙一, 北村仁美, 梶浦正浩, 東美奈子, 野中由美子, 小野雅也, 菊池 豊: クリックスルーに基づく探検型検索サイトの設計と開発, 第7回情報科学技術フォーラム(FIT 2008)講演論文集, 第2分冊, pp.1–4 (2008).
- 13) Buchanan, G. and Owen, T.: Improving navigation interaction in digital documents, *Proc. 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '08)*, New York, NY, USA, ACM, pp.389–392 (2008).
- 14) Watts, D. and Strogatz, S.: Collective dynamics of 'small-world' networks, *Nature*, Vol.393, No.6684, pp.440–442 (1998).
- 15) Ferrer, R. and Sole, R.V.: The small world of human language, *Proc. Royal Society of London Series B, Biological Sciences*, Vol.268, pp.2261–2265 (2001).
- 16) 戸田浩之, 中渡瀬秀一, 片岡良治: 特徴的な固有表現を用いたラベル指向ナビゲーション手法の提案, *情報処理学会論文誌: データベース*, Vol.46, No.13, pp.40–52 (2005).
- 17) 松尾 豊, 大澤幸生, 石塚 満: Small World 構造に基づく文書からのキーワード抽出, *情報処理学会論文誌*, Vol.43, No.6, pp.1825–1833 (2002).
- 18) Takeda, Y. and Umemura, K.: Selecting indexing strings using adaptation, *SIGIR '02: Proc. 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, ACM, pp.427–428 (2002).
- 19) 小峰 恒, 山田剛一, 絹川博之, 中川裕志: 文書頻度と節長を利用した図書概要縮約方式, *NII Journal*, Vol.8, pp.23–33 (2004).
- 20) Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information, *Psychological Review*, Vol.63, No.2, pp.81–97 (1956).
- 21) Matthews, J.R., Lawrence, G.S. and Ferguson, D.: *Using Online Catalogs: A Nationwide Survey*, Neal-Schuman Publishers, Inc., New York, NY, USA (1983).
- 22) Kim, K.S. and Allen, B.: Cognitive and task influences on Web searching behavior, *Journal of the American Society for Information Science and Technology*, Vol.53, No.2, pp.109–119 (2002).

(平成 22 年 8 月 20 日受付)

(平成 23 年 1 月 14 日採録)

## 推薦文

現在の情報検索では、ユーザが想起しにくい情報へのアクセスが困難であり、ユーザが情報を俯瞰しにくいという問題がある。本論文では、文書集合から得られた特徴語をたどることで文書集合内を網羅的に探索できるナビゲーションシステムを提案している。網羅的な探索に有効と考えられる媒介中心性の高い語を、利用者は想起できないが、システムが提示した語の中から選択することはできることを利用者実験により示している。新規性、有用性のいずれにも優れている論文と考えられ、推薦論文に値すると判断した。

(マルチメディア, 分散, 協調とモバイル (DICOMO2009) シンポジウム  
プログラム委員長 土井美和子)



島田 諭 (学生会員)

2004年東洋大学社会学部メディアコミュニケーション学科卒業。2009年筑波大学大学院図書館情報メディア研究科博士前期課程修了。同年より同研究科博士後期課程に在籍。グループウェア, テキストマイニング, Webナビゲーションに関する研究に従事。日本データベース学会学生会員。



福原 知宏 (正会員)

2003年奈良先端科学技術大学院大学情報科学研究科博士後期課程単位取得認定退学。工学博士。科学技術振興機構社会技術研究開発センター研究員, 東京大学人工物工学研究センター特任助教を経て, 2010年産業技術総合研究所サービス工学研究センター特別研究員。Webテキストマイニング, スпамフィルタリングの研究に従事。



佐藤 哲司 (正会員)

1980年山梨大学工学部電子工学科卒業。同年日本電信電話公社武蔵野電気通信研究所に入所。1994年工学博士(大阪大学)取得。NTTコミュニケーション科学基礎研究所を経て, 2007年4月より現職。情報アクセス技術, 社会インタラクションに関する研究に従事。電子情報通信学会会員。