社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2008－GN－68（2）
2008/5/23

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

宮崎 陽司† 河野 泉‡

†NECサービスプラットフォーム研究所 ‡NEC共通基盤ソフトウエア研究所
〒630-0101 奈良県生駒市高山町8916-47
E-mail: †y-miyazaki@bc.jp.nec.com, ‡kohno@ay.jp.nec.com

あらまし 多くの企業で，大量の文書が蓄積，共有されており，文書の再利用が可能となっている．しかし，大量の文書の中から欲しい文書を探すのは容易ではない．検索対象にどのような文書が存在しているのか分からない場合，適切な検索キーワードを設定するのは困難である．この問題を解決するために，我々は，ユーザの検索対象の把握と絞り込みを支援するナビゲーション方式を開発した．本方式では，検索対象に含まれる文書のメタ情報から，検索対象の特徴を表す複数のナビゲーションキーワードを抽出し，提示する．ユーザは，ナビゲーションキーワードを一覧することで検索対象の内容を把握でき，キーワードを選択するだけで検索を進められる．日常的に文書検索システムを利用するユーザに対して評価実験を行ったところ，従来のキーワード検索を多く利用するユーザに比べ，ナビゲーションキーワードの選択回数が多いユーザの方が，文書閲覧回数が少なかった．これは，ナビゲーションキーワードによって検索対象の内容を把握し，検索結果を逐次確認しなくとも検索キーワードを設定できたことを示している．

キーワード 文書検索，ナビゲーション，キーワード，メタ情報

Document Navigation Using Relevant Keywords for Surveying and Refining Search Results

Yoji MIYAZAKI† and Izumi KOHNO‡

† Service Platforms Research Laboratories, NEC ‡ Common Platform Research Laboratories, NEC

8916-47 Takayama-cho, Ikoma-shi, Nara, 630-0101, Japan
E-mail: †y-miyazaki@bc.jp.nec.com, ‡kohno@ay.jp.nec.com

Abstract Recently in many companies, people write and store a lot of documents into databases. They can write documents using examples from databases. However, since the number of stored documents increases by the day, it is difficult to find proper documents in database. To overcome this problem, we developed document navigation system which can support document retrieval with relevant keywords. The system creates relevant keywords based on the search results and displays them as search keyword candidates. A user can survey and refine search results using those keywords. The results of experiments shows that users who use relevant keywords could find proper documents by surveying fewer documents than users who use only conventional keyword search. This result means that users can figure out search results and set adequate search keywords with relevant keywords the system shows.

Keyword document retrieval, navigation, keyword, meta data, retrieval system

1. はじめに

近年企業内のIT化が進んだことにより，社員が作成した文書を，全社あるいは部門毎のサーバで管理するようになってきた．ユーザ間で文書を共有し，文書を再利用することで効率よく文書を作成できる．例えば，ある商品の提案資料を作成する際に，他の社員が作成した提案資料を探し出し，その資料を参考にすることで，効率よく提案資料を作成できる．

共有する文書が大量になってくると，必要な文書を探し出すための検索機能が必要となる．代表的な検索機能としてキーワード検索がある．例えば，ある製品に関する文書を探す場合，製品名を検索キーワードとして入力すると，その製品の名称を含む提案書や設計書，市場調査資料などを検索できる．

しかし，適切な検索キーワードを設定することは難しい．目的の文書を的確に表現するキーワードを思いつかず，抽象的なキーワードを設定した場合は，多数の文書がヒットしてしまう．また，文書中で使われな

いキーワードを設定すると，検索結果が0件になってしまう．

適切に文書を絞り込めず，目的の文書を見つけられなかった場合，ユーザは検索キーワードを修正して，再検索する必要がある．例えば，さらに絞り込むためのキーワードや，別の観点から検索を行うためのキーワードなどを設定し，再検索する．

再検索のための検索キーワードを検討する際，ユーザは，検索結果の内容を確認し，どのような文書が検索結果に含まれているのかを確認しながら，検索キーワードを設定する．しかし，検索結果の文書が多数である場合，各文書の内容を逐次確認して検索結果を把握することは困難である．

また，検索結果が0件である場合，検索対象にどのような文書があるか分からないため，設定可能なキーワードが分からず，文書内に現れないキーワードを設定するなど，適切な検索キーワードを設定できない．そのため，所望の文書を見つけるまで，何度も検索結果の内容確認と検索キーワード設定を繰り返さなければならない．

そこで我々は，検索キーワードの設定を支援する検索ナビゲーション方式を開発した．本方式では，検索対象の特徴を表すナビゲーションキーワードを提示することによって，ユーザは容易に検索対象を把握できる．また，ユーザは，ナビゲーションキーワードを選択するだけで，簡単に検索を進められる．

本稿では，2節でユーザの検索行動を分析し，文書検索における問題点や，問題点を解決するための要件を述べる．3節では要件を満たすナビゲーション方式を提案し，4節で方式を実装したシステムについて説明する．5節で試作システムを用いた評価実験について述べ，提案方式によって効率よく検索を行えることを示す．

2. 検索行動分析

本節では，ユーザの検索行動を分析し，従来のキーワード検索における問題点を挙げ，文書検索システムの要件について述べる．

2.1. 検索行動モデル

我々は，図1(a)に示すように，文書検索時のユーザの行動が，検索キーワード設定，検索結果調査，検索結果理解の3つのフェーズから構成され，文書を検索する際は，各フェーズを順次繰り返し行っていると考えた．

まず所望の文書の特徴を考え，検索キーワードを設定する（検索キーワード設定）．検索結果を得ると，各文書のタイトルや作成者，作成日，内容などを調査し（検索結果調査），検索結果がどのような文書で構成されているかを理解したり，自分の要求にあった文書に

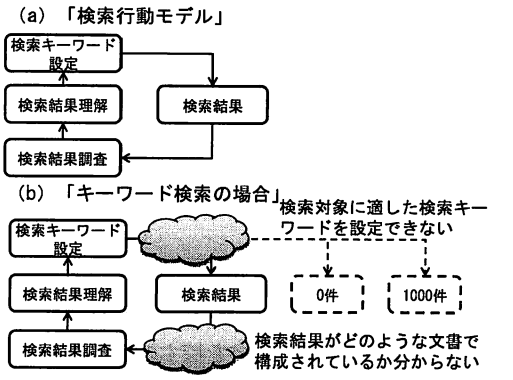


図1 検索行動モデル

絞り込めているかどうかを確認したりする（検索結果理解）．そして，検索結果が多すぎたり，少なすぎたり，あるいは目的とする文書が含まれなかったりする場合は，検索キーワードを再設定して検索を行う（検索キーワード設定）．

2.2. 従来の検索システムの問題

従来のキーワード検索では，適切な検索キーワードを設定できず，所望の文書をうまく見つけられないという問題があった．これは，図1(b)に示すように，検索の各フェーズにおいて「検索対象がどのような文書で構成されているか分からない」「検索対象に適した検索キーワードを設定できない」ためであると考える．以下に，各問題の詳細を説明する．

問題1 **検索対象がどのような文書で構成されているか分からない（検索結果調査・理解フェーズ）**

所望の文書を表すキーワードを思いつかない時や，ある技術領域に関する文書を探す時などは，曖昧な検索キーワードを設定することが多い．その結果，広い範囲に関する文書が検索されるため，多くの文書を検索結果として得ることが多い．例えば，ユビキタスに関連した製品を探したいが，製品名を覚えていない場合は，「ユビキタス」を検索キーワードとして検索を行う．しかし，「ユビキタス」は一般的なキーワードであるために，多くの文書が検索される．さらに絞り込むための詳細なキーワードを考えるために，検索結果を調査しようとするが，検索結果の文書数が多いために，各文書を逐次確認していくことは非常に面倒である．

問題2 **検索対象に適した検索キーワードを設定できない（検索キーワード設定フェーズ）**

所望の文書を的確に表現し，十分に絞り込める検索キーワードを設定できれば，1度の検索で所望の文書にたどり着ける．しかし，文書中に含まれないキーワードを設定したり，文書毎に「NEC」と「日本電気」のような表現の揺れがあったりする場合には，設定したキーワードで目的の文書を検索できない場合がある．

その結果，検索キーワードを何度も切り替えながら，検索していかなければならない．

2.3. 文書検索システムの要件

前節で述べた問題を解決するシステムの要件として，次の2要件を設定した．

要件1　検索対象の特徴を示す情報を提示する

問題1を解決するためには，検索対象の調査，理解を支援する必要がある．ユーザが逐次文書を確認する手間を省けるように，検索対象を容易に概観できる情報を提示する必要がある．

要件2　確実に絞り込めるキーワードを提示する

問題2を解決するためには，検索対象を絞り込み可能なキーワードをユーザに提示することが必要である．検索に利用可能なキーワードを予め提示できれば，ユーザは，それらのキーワードの中から，自分の検索目的に合致するキーワードを利用して，確実に検索を進められる．

3. ナビゲーションキーワード提示方式

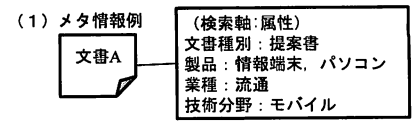
2.3 節で述べた要件を満たすナビゲーションキーワード提示方式を提案する．本方式は，検索対象の特徴を表す情報としてナビゲーションキーワードを抽出し，提示する．ユーザは，ナビゲーションキーワードを一覧することで，検索対象を概観でき，内容を把握できる．また，提案方式では，ナビゲーションキーワードによる検索を可能にする．ユーザはナビゲーションキーワードの中から目的に合致するキーワードを選択するだけで検索を進められる．

3.1. ナビゲーションキーワード

我々は，検索対象の特徴を表す情報として，検索対象に含まれる文書のメタ情報が有効であると考えた．メタ情報は，文書の特徴を表す情報であり，検索対象の文書に含まれるメタ情報をユーザに提示することによって，検索対象の理解を支援できると考える．

メタ情報は，検索軸と属性から構成される．検索軸は，属性の観点を表すものであり，属性は，その観点から文書の特徴を表す語である．図2（1）にメタ情報の一例を示す．例えば，文書Aが，“流通業に関連する文書で，情報端末やパソコン，モバイル機器を提案する提案書”である場合，メタ情報として「文書種別：提案書」「製品：情報端末，パソコン」「業種：流通」「技術分野：モバイル」が付与される．図2（2）に他の検索軸，属性の例を示す．

本稿では，ユーザに提示するメタ情報をナビゲーションキーワードと呼ぶ．ユーザは，提示されたナビゲーションキーワードを一覧することで，検索対象の文書の内容を見なくても，どのような文書が含まれているかが分かる．



(2)メタ情報の種類

検索軸	属性
文書種別	提案書，紹介資料，事例，・・・
製品	ネットワーク機器、サービス基盤、情報端末・・・
業種	製造，流通，サービス，・・・
技術分野	モバイル，セキュリティ，ネットワーク，・・・
・・・	・・・

図 2　文書のメタ情報

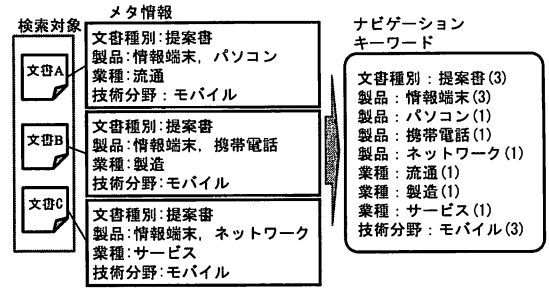


図 3　ナビゲーションキーワード例

3.2. ナビゲーションキーワード分布の分析

ナビゲーションキーワードは検索対象の特徴を表している．例えば，図3に示すように，文書A，B，Cが検索対象の場合，ナビゲーションキーワードを一覧することで，検索対象には「提案書」「情報端末」「モバイル」に関する文書が多いことが分かる．しかし，企業内の文書を検索対象とする場合，検索対象には様々な業種や製品などに関する文書が含まれるため，多くのナビゲーションキーワードが提示されてしまう．その結果，ナビゲーションキーワードを一覧しただけでは検索対象の特徴を捉えにくくなるという問題がある．

そこで我々は，検索対象を検索軸ごとに分析し，文書集合の特徴を示している度合い（注目度）を求め，高い注目度のナビゲーションキーワードを提示すれば検索対象を概観しやすくなると考えた．

具体的には，検索軸毎にナビゲーションキーワードの分布の偏り度合いを調べ，分布の偏りが大きい検索軸と，その検索軸のキーワードを検索対象の特徴として提示する．ユーザは，提示された検索軸，キーワードを見ることで，検索対象の文書集合が，どのような観点（例えば技術分野）の，どのような内容（例えばモバイル）に注目しているかが分かり，効率よく検索対象を概観できる．

図4を例に，検索軸毎の分析方法について，具体的に説明する．

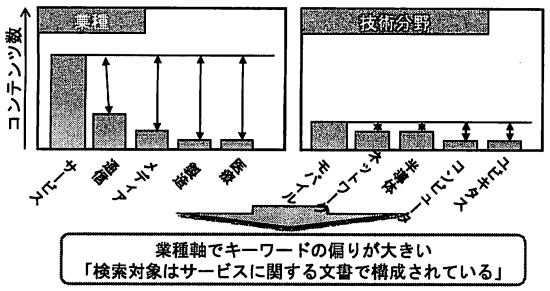


図 4 検索軸ランキング

図 4 は，ある検索対象の業種軸と技術分野軸のナビゲーションキーワードの分布を示している．この検索対象に含まれる文書は「サービス業」に注目した文書が多いといえる．つまり，「業種軸」という観点からみた「サービス」が検索結果の特徴を表しており，検索対象がサービス業に関する文書で構成されていることが分かる．一方，技術分野軸では，どのナビゲーションキーワードにも偏りがなく，技術分野という観点では特別に注目すべき事柄のない文書集合であると考えられる．以上の分析結果から，業種軸のほうが，注目度合いが高いことが分かり，業種軸のナビゲーションキーワードを優先してユーザに提示する．

検索軸毎のナビゲーションキーワードの偏りは次式で求める．本稿では検索軸の偏り具合を注目度と呼ぶ．具体的には，コンテンツ件数の最も多い属性とそれ以外の属性とのコンテンツ数の差の平均値を注目度とした．

$$C_i = \sum_j^{m_i} \left(N_{i,\max} - N_{i,j}\right)^2 \Big/ m_i$$

$$N_{i,\max} = \max\left(N_{i,j}\right)$$

ここで Ci は注目度，mi は，検索対象コンテンツについて検索軸 i に属する属性数，Ni,j は，検索軸 i における属性 j に関連するコンテンツ数を，検索軸 i の全属性のコンテンツ数の総和で正規化したものである．各検索軸の注目度を求め，注目度の大きい検索軸のナビゲーションキーワードを優先してユーザに提示する．

3.3. ナビゲーションキーワードによる検索

ナビゲーションキーワードは，検索対象に含まれる文書のメタ情報であるため，ナビゲーションキーワードを検索キーワードとして利用すれば，検索対象を確実に絞り込める．

図 5 にナビゲーションキーワードによる検索処理の流れを示す．ユーザがナビゲーションキーワードを選択すると，そのキーワードをメタ情報に持つ文書を検索する．そして，検索結果に対して，ナビゲーションキーワードを抽出，検索軸のランキングを行い，ユーザに提示する．ユーザは，提示されたナビゲーションキーワードを選択すると，検索キーワードに追加して，検索結果をさらに絞り込む．

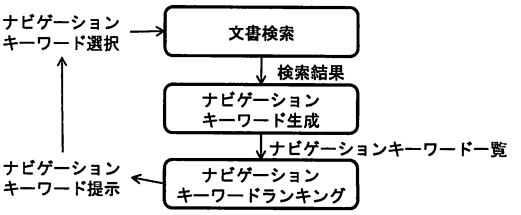


図 5 ナビゲーションキーワードによる検索処理

このように，ユーザは提示されたナビゲーションキーワードを選択するだけで，確実に検索を進められる．

3.4. 従来方式との比較

絞り込み結果が 0 件にならない検索キーワードを提示し，検索支援を行う方式として，ファセットナビゲーションがある[1][2]．ファセットナビゲーションでは，検索条件を ABC 順などで一覧している．これに対し，提案方式では，検索対象の文書分布に応じたランキングによって注目すべき検索軸を示し，ユーザの検索対象の理解を支援している点がファセットナビゲーションに対し，優位である．

検索対象の特徴を表すキーワードを提示し，検索を支援する方式には，様々な方式が提案されている．効率のよい絞り込みを支援する方式として，戸田ら[3]や西森ら[4]の方式がある．戸田ら[3]は，絞り込める文書集合間の差が大きく，文書数のばらつきが小さいキーワードをもつカテゴリを優先的に提示する．西森ら[4]も絞り込める文書数が近いキーワードを優先的に提示する．我々は，街中でのレストラン検索を対象に，戸田らと同様のランキング方式と，提案方式の両者を用いたシステムを構築し，実験を行った[5]．実験では，提案方式の方が，所望のコンテンツを見つけるまでの検索キーワードの設定回数は多かったが，設定した検索キーワードを修正する回数が少なかった．戸田らの方式では，ユーザの目的に合致したキーワードが提示されれば，効率よく絞り込むことができるが，企業内の文書検索のように，検索結果を見ながら徐々に検索を進めていくことが多いシステムでは，やり直しの少ない提案方式が有効である．

4. 試作システム

本節では，3 節で提案した方式を実装したシステムについて述べる．

図 6 に試作システムの機能構成を示す．試作システムはサーバ・クライアントで構成される．サーバはクライアントで設定された検索キーワードをもとに文書を検索し，文書のメタ情報からナビゲーションキーワードを生成する．そして検索軸ごとに注目度を求め，注目度に基づいて検索軸をランキングする．そしてクライアントで表示する画面を生成し，クライアントに

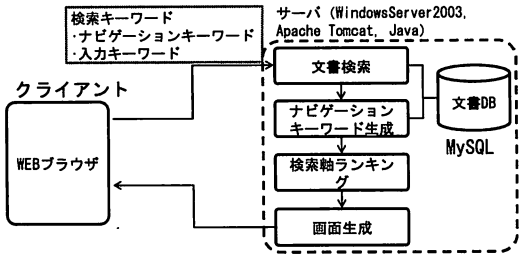


図 6　試作システム

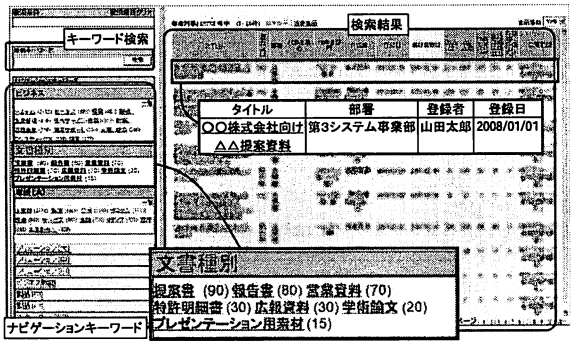


図 7　クライアント画面例

送信する．本システムではキーワード検索機能も実装し，入力されたキーワードも検索キーワードとして利用できるようにする．

図 7 にクライアントの画面例を示す．クライアントは 2 ペインで構成され，左ペインにナビゲーションキーワードの一覧とキーワード検索入力欄とを表示し，右ペインに検索結果を一覧表示する．

ナビゲーションキーワードは，検索軸名とその検索軸のナビゲーションキーワードを，注目度順に列挙する．ただし，画面の制約上，すべてのナビゲーションキーワードを表示できない．そのため，注目度が低い検索軸は，画面ロード時には検索軸名だけを表示しておき，ユーザが検索軸を選択したときに，ナビゲーションキーワードを表示するようにした．ユーザがナビゲーションキーワードを選択すると，選択されたキーワードを検索キーワードとしてサーバに送信する．そしてサーバが検索結果や検索結果から生成したナビゲーションキーワードをもとに画面を生成し，クライアントで表示する．また，キーワード検索入力欄で，任意のキーワードを入力すると，検索キーワードとしてサーバに送信し，サーバで生成した検索結果やナビゲーションキーワードを表示する．

検索結果一覧には，文書のタイトルや作成者，所属，登録日などが表示される．ユーザがさらに文書の詳細をみたい場合は，文書のタイトルリンクをクリックすると，概要を表示したり，文書ファイルをダウンロードしたりできる．

サーバは，WindowsServer2003，Apache Tomcat 上で動作する Java サーブレットで構築した．文書データ（メタ情報）は，MySQL データベースに格納した．

5. 評価

本節では，試作システムを用いて行った評価実験について述べる．実験内容，および実験結果について述べ，提案方式の有効性を評価する．

5.1. 実験方法

社内には，全社的に文書を共有できる文書管理システムがあり，社員が作成した提案資料や商談資料など，様々な文書を共有し，再利用できるようになっている．

本実験では，社内の文書管理システムを頻繁に利用する社員 25 名を被験者とし，普段利用しているシステムの代わりに，3 週間試作システムを利用して文書を検索してもらった．また，社内の文書管理システムと同じ文書データを利用し，約 11000 件の文書（検索軸は 14 個，1 文書あたり平均 12 個のメタ情報）を検索対象とした．

評価は，被験者に対する検索ログとアンケートから行った．検索ログは，ユーザの操作を記録したものであり，選択したキーワードや閲覧したコンテンツなどを記録した．検索ログからは，提案方式によって検索対象の概観を把握しながら文書を検索できたかどうかについて評価した．アンケートは設問に答える部分と自由記入部分から構成され，ナビゲーションキーワードによる検索キーワードの設定について主観的な評価を行った．

5.2. 実験結果・考察

実験実施状況

表 1 に本システムの利用状況を示す．ここでセクションとは，1 回の検索行動のことを指す．具体的には，検索を開始してナビゲーションキーワードを選択したり，コンテンツ一覧を操作したりするなど，様々な操作を行い，検索を終了するまでの検索行動をセクションと呼ぶ．本実験では，本システムのトップページへ戻るリンクを押すか，Apache Tomcat のセッションが切れた場合に，検索を終了したと判定した．表 1 から 3 週間の実験期間中に，1 人あたり約 5 回の検索行動を行ったことが分かる．

ログ分析による評価

提案方式によって，検索対象の概観を把握しながら文書を検索できたかどうかを評価した．

検索対象の概観把握がうまく言っている場合は，検索行動の中で，検索結果の文書を個別に確認する行動が少ないと考える．つまり，セクションあたりの文書閲覧数は少ない．一方絞り込みがうまく行われない場合は，何度も検索を行い，その都度，検索結果の文書を閲覧しているため，セクションあたりの文書閲覧数

表 1 利用状況

被験者25名　　単位：回

	合計	平均一人当たり
セクション数	127	5.08

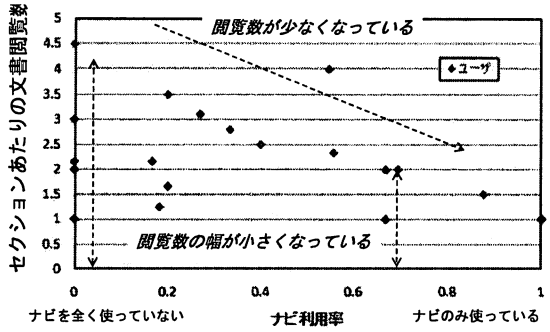


図 8 ナビ利用率と文書閲覧数の関係

は多いと考える．

そこで，目的の文書を見つけたセクションについて，セクションあたりの文書閲覧数を用いて，検索効率を評価する．

本実験では，検索結果のタイトルリンクをクリックし，文書の詳細を 1 度でも閲覧したセクションを，目的の文書を閲覧したセクションと判断した．

また，ナビゲーションキーワードの利用頻度として各ユーザのナビ率を定義する．ナビ率とは，全検索操作回数（キーワード検索回数とナビゲーションキーワード選択回数）に対するナビゲーションキーワードの選択回数の比率である．ナビ率が 1 に近いユーザは，ナビゲーションキーワードを多く検索に利用したユーザであることを示す．

ユーザ毎のナビ率とセクションあたりの平均文書閲覧数とを図 8 に示す．本結果から，ナビ率が高い，すなわち，ナビゲーションキーワードをよく利用するユーザの方が，1 セクション内の文書閲覧数が少ないことが分かる．また，ナビゲーションキーワードを利用する被験者のほうが，文書閲覧数の分散が小さい．

本結果から，キーワード検索では，1 度で適切に絞り込むことができれば文書閲覧数は少なくて済むが，そうでない場合は，目的の文書が見つかるまで，何度も文書を閲覧しながらキーワードの修正候補を検討していると考えられる．これに対し，提案方式では，検索結果を概観しながら確実に絞り込むことができるので，逐次コンテンツを閲覧する手間を軽減できたと考える．

アンケートによる主観評価

表 2 にアンケート結果を示す．本結果では，ナビゲーションキーワードによる検索機能が，簡単でわかりやすく，目的の文書を見つけるために有用な機能であるという評価を得ている．自由記入では，「検索キーワードが思いつかないときに利用できる」「検索上のヒントになる」「該当するコンテンツの件数が表示されるのは便利」というコメントがあり，ナビゲーションキーワードによって検索結果の特徴を把握でき，簡単に検索キーワードを設定できたと考えられる．

表 2 アンケート結果

単位：人

項目	簡単・分かりやすい	難しい・分かりにくい	その他
システム全般	13	1	2
ナビゲーション	12	3	1

単位：人

辿り着いた文書の内容		
	目的以上の情報	0
	目的通り	11
	目的以下	2
	その他	2

自由記入

検索キーワードが思いつかない場合、利用できると思いました。
検索上のヒントになると感じました（つまりどのようなカテゴリというかキーワードがあるのかが分かり、それを条件に組み込める為）。また自分でも気付かないキーワードもあり有用と感じます
・検索条件を選択できること・途中の検索条件を消せること機能自体は非常に便利だと感じました
見やすく、検索のし直しも容易であった。これまでは、キーワード検索を多用していたが、ナビゲーションが使いよかった。
個人的には直感的に利用できた感じがします。
業務分類が分かりやすかった
ナビゲーションキーワードは良いと思います。
該当するコンテンツの件数が表示されるのは便利でよいです。
アクセス人数や更新日時による絞込みができれば、なお良いと思う。ソートだけでは絞り込みづらい。
このナビゲーションについて事前知識のない利用者には動きがわかりにくいかもしれません。

6. おわりに

本稿では，検索対象の特徴を表すナビゲーションキーワードを用いて，検索対象の概観を提示し，検索キーワード設定を支援するナビゲーション方式を提案した．試作システムによる評価実験の結果から，提案方式により，検索対象を概観しながら，簡単に検索キーワードを設定して検索を進められることを示した．今後は，ユーザの興味や業務内容を用いたナビゲーションキーワードのパーソナライズを行い，文書検索の効率的なナビゲーションを目指す．

文　献

[1] MARTI A. HEARST, CLUSTERING VERSUS FACETED CATEGORIES FOR INFORMATION EXPLORATION, COMMUNICATION OF THE ACM, Vol. 49, No. 4, pp. 59-61, April 2006.

[2] Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst, Faceted Metadata for Image Search and Browsing, Proceedings of ACM CHI 2003.

[3] 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，Vol. 46，No. SIG13(TOD27)，pp. 40-52，2005.

[4] 西森崇，前田茂則，小島良宏，検索効率向上のためのコンテンツ均等分類手法の提案および評価，情報処理学会研究報告，No. 105(HI-120), pp. 67-73, 2006.

[5] 河野泉，宮崎陽司，原雅樹，池上輝哉，状況に応じた着目点提示によるモバイル向け対話型情報ナビゲーション，情報処理学会論文誌，Vol. 48, No. 3，pp. 1186-1196，2007.