NAIST-IS-MT9851087

修士論文

文書情報の可視化による検索絞り込み支援

林 一成

2000年2月14日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士 (工学) 授与の要件として提出した修士論文である。

林 一成

審査委員：　横矢 直和 教授
湊 小太郎 教授
竹村 治雄 助教授

文書情報の可視化による検索絞り込み支援*

林 一成

内容梗概

近年，大量の文書データベース中からの文書検索のための研究が数多く行われている．最も古典的な方法であるキーワード型検索方法は，与えられたキーワードに対して，そのキーワードを含む文書の一覧を表示する検索方法である．ユーザは欲しい文書集合を得るために論理式を書く必要があるが，検索結果の文書集合が膨大になり不要な文書が多数含まれたり，絞り込みすぎて欲しい文書を見つけられないことがある．検索精度のよい論理式を書くためには，検索対象に対する専門的な知識や検索式を書く技術が必要となり，一般のユーザにとっては難しい．結果として，ユーザが大きな文書集合から個々に内容を確認していくことが必要となり，非常に手間がかかる．

そこで，本研究では文書情報の可視化による絞り込み支援法を提案する．特定の文書を特徴づける単語をキーワード候補語を提示し，その候補語による検索結果の件数分布の全体図を可視化する．この可視化結果から，ユーザは自分の興味ある文書やキーワード候補語からの検索が容易になり，かつその検索結果の件数が分かる．これにより，キーワード検索で検索上位にあがらなったユーザが欲しい文書を発見することが可能になる．またキーワード候補語を用いることにより，ユーザの検索概念に即しているが思いつかなかったキーワードを提示できる．これにより，絞り込みに掛かる手間を減少させることが可能になる．

キーワード

文書検索, 絞り込み, 可視化, キーワード, 類似度

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT9851087, 2000年2月14日.

Document Retrieval Support Based on Visualization of Documents Information*

Kazushige Hayashi

Abstract

It is important to search desired documents in the large document database. Most of existing methods use words as a key for document retrievals. However, it is very difficult to search key words that appropriately distinguish desired documents from other non-desired documents. The task of refining query to reduce the search results is the most important step of the process of using search engine. This paper proposes a method for supporting the document retrieval process by visualizing a document information. This information simplifiess the task of refining a query, and helps user to undarestand relation between documents and words.

Keywords:

document retrievals, search, information visualization, keyword, similarity

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9851087, February 14, 2000.

目 次

図 目 次

表 目 次

1．はじめに

近年のコンピュータ技術の発達によるパーソナルコンピュータやインターネット，イントラネットの普及に伴い，WWW 上の文書から個人のメールボックスまで，ユーザが膨大な情報にアクセスできる環境が提供されてきている．例えば，WWW 化による部門内の情報共有やインターネット上での安定した情報提供サービスの出現など，有効に利用できれば業務の効率化だけでなく，業務の質を格段に向上できる状況になりつつある．それに伴い，膨大な文書データベースから，欲しい文書情報を取り出せる文書検索システムに対する要求が高まっている．ところが，自分のデスクトップやメールボックス内ですら必要な情報を発掘するのに窮するほどに情報量が増え，一方で有効な検索システムがないために，必要な情報が発見できないのが実状となっている．

このような状況に対して，大量の情報源からの検索機能が重要かつ有用であるとの認識から，WWW 上では Altavista[39] や goo[57] などの検索サービスが提供されている．また，個人環境でも過去のメールが検索できる++Mail[23] などが発表されている．これらはいずれも，最も古典的な文書検索方法であるキーワード検索システムであり，入力されたキーワードの有無により検索を行う．キーワード検索では，ユーザは欲しい文書に関連すると思われるキーワードを入力し，キーワードを含む文書をデータベースから取り出す．しかし，入力が単語に限定されるため，適切な単語をキーワードとして入力するのが難しく，また適切な単語をキーワードととして入力しても思ったような検索結果が得られない場合がしばしばある．

キーワード検索でのこのような問題点に対して，単語ではなく文書をキーとした検索を行う検索方法がある．本研究では，巨大な文書データベースからの文書集合の検索を目的とする．そのため，検索課程で自分の欲しい文書を一つ得られたとき，その文書をキーとして似た文書の検索ができれば，ユーザの検索に対する負担が減ると考えられる．

指定した文書に関連する文書を検索する類似検索では，キーとした文書に対して似ている割合を類似度という数値で表現し，ユーザには類似度の高い文書が提示される．類似検索には様々なバリエーションがあるため，類似度の計算方法も

様々あるが，キーとして指定した文書内の全ての単語の出現頻度から計算する方法が一般的である．

類似検索の利点は，ユーザが検索対象に対して特に深い知識がなくても，キーとして指定した文書に似た文書が得られる点にある．しかし，キーとした文書に対する類似度が他のどの文書も低い場合や，類似度の高い文書にユーザが想定した内容ではない文書が多数含まれてしまうといった場合，類似検索はうまく働かない．このようなときユーザは，キーワード検索にを利用して文書集合を絞り込む，もしくは検索結果を個々に確認するなどの手段を用いて，欲しい文書集合を見つけることに必要になるが，いずれの方法も試行錯誤を繰り返す必要があり面倒な作業である．

そこで本研究では，類似検索のこのような問題点を解決し，かつユーザが得た一つの欲しい文書の情報をうまく利用した検索方法を提案する．類似検索の問題点である指定文書との関連を可視化により効率的に見せることにより類似検索の問題点を解決し，かつ指定した文書中の単語の情報を利用したキーワード検索を行えるための支援を行う．これらの検索支援により，ユーザは自分の検索意図にあった文書集合の検索を効果的に行える．

以下 2 章では，文書検索の関連研究と本研究の位置付けについて述べる．次に 3 章では本研究で可視化する文書情報について説明し，4 章では提案手法を実装したシステムの詳細を説明する．そして，5 章では作成したシステムを用いた文書に関する評価実験と考察を行う．

2. 文書検索の関連研究と位置付け

文書検索とは，文書集合の中からユーザが欲しい文書を探す作業である．最も原始的であると同時に最も確実な方法として，個々の文書の内容を人間が確認して，自分の欲しい文書を探すという手法が考えられるが，文書集合が大量になると大変手間が掛かる方法である．そこで，文書検索に計算機を利用する様々な手法が提案されている．それらの手法では，ユーザは欲しい文書を求めるために，計算機に自らの検索意図を適切に伝えることが必要になると同時に，検索の結果がユーザに分かりやすく提示される必要がある．

このような視点から本章では，まず文書検索と文書集合の可視化に関する研究分野を概観した後に，本研究で提案する文書情報の可視化による検索支援の概要を述べる．

2.1 文書検索

ここでは，文書検索の際にユーザの検索意図を表現する方法によって，従来の文書検索手法を，キーワード型検索システム，ディレクトリ型検索システム，ならびに類似検索システムの 3 種類に分類し，各々の概要と問題点を説明する．

- **キーワード型検索システム**
  キーワード型検索システムは，与えられたキーワードに対して，そのキーワードを含む文書をデータベースから取り出し，その一覧を提示するシステムである．例として，WWW における検索エンジンの Altavista[39] や goo[57] などがある．キーワード検索の流れを図 1に示す．ユーザは欲しい文書に必ず含まれると思われるキーワードを入力して，文書数を絞り込み，その中から欲しい文書を得る．ユーザにとって，自分の欲しい文書に必ず含まれると思われるキーワードを見つけることは比較的容易であるため，再現率 (正解総数に対する検索結果として取り出した正解の割合) の高い検索システムであるといえる．しかし，容易に思いつくような単語をキーワードとすると，検索結果の文書集合が非常に大きくなり，不要な文書が多く含まれる場合がある．このとき適合率 (検索結果の総数に対する正解数の割

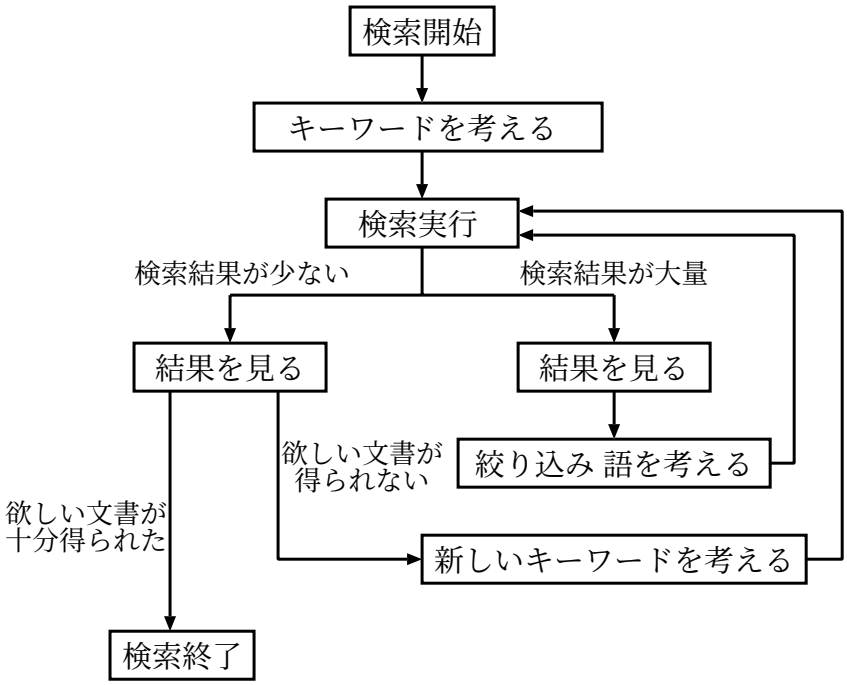


図 1 **キーワード検索を行う手順**

合) が低くなるという問題が生じる．逆に，特異な単語をキーワードとすると，検索結果の文書集合が非常に小さくなり，欲しい文書が検索結果から漏れる場合があるが，この場合には再現率が低くなるという問題が生じる．そのため，ユーザは再現率を維持したまま適合率を高くするために検索式を新たなキーワードを使って作成する必要がある．しかし，そのような検索式を書くためには，検索対象に対する専門的な知識や検索式を書く技術が必要になり，一般のユーザにとっては難しい作業である．結果として，ユーザが大きな文書集合の中から文書の内容を個々に確認していく作業が必要となり，非常に手間が掛かる．

- **ディレクトリ型検索システム**

  単語のみに頼らない例として，ディレクトリ型検索がある．ディレクトリ型検索システムは，データベース内の文書をあらかじめ「芸術」「ビジネス」「教育」といった分野別に整理し，分類を提示するシステムである．例として，Yahoo[67] や Lycos[65] などがある．分類はカテゴリの階層構造となっ

ていて，ユーザは欲しい文書に関連があると思われる，分類されたカテゴリを選択して分類をたどっていくことによって，自分の欲しい文書への絞り込みを行う．キーワード検索と違って，ユーザは適当なキーワードを見つける必要がなく，提示されたカテゴリを選択するだけで絞り込みが行えるため，検索対象に対する専門的な知識を必要とせず，ユーザの負担が少ない検索システムである．しかし，分類の方法が固定的であるため，ユーザの全ての検索要求に対処できない場合がある．また，分類のためのカテゴリ作成は手作業で行われているために，非常に手間が掛かるという問題点がある．さらに，カテゴリ構造が巨大であるために見通しが悪く，迷いや誤解を招く場合がある．

- **類似検索システム**
  ユーザが指定した文書に対して，文書内に含まれる単語の出現頻度が類似した文書を検索するシステムである．例として，ConceptBase[70] や Solution Wave[52] などがある．単語ではなく，文書そのものを検索キーとした検索を行う点が特徴である．文書内の全ての単語の出現頻度から類似度の値が計算されるため，単語の出現頻度が似ている文書の検索に有効な検索方法である．しかし，結果の提示手法は類似度の高いものを上位にランク付けするだけで「どのような計算の結果として類似度が計算されたか」はユーザに提示されない．そのため，ユーザは上位にあるものが,「自分の検索目的に沿った意味で，類似した文書であるかどうか」を個々に読んで確認する必要がある．

2.2 文書情報の可視化

前節で紹介した文書検索システムを利用する際には，検索結果をユーザに分かりやすく提示し，検索意図に沿った結果が得られたか否かを把握させ，得られなかった場合には次にどのような操作を行えばよいかの指標を与えられる必要があると考えられる．しかし，既存の文書検索システムの多くは，URL や内容の要約といった情報を検索結果として検索キーとの関連度の順に並べて，20 件程度提示するといった提示方法に留まっている．そのため，ユーザはこれらの情報から

各文書の内容を推測して，あるいは実際にその文書を読んで内容を確認して，自分の検索意図に沿った文書であるかを判断する必要がある．しかし，関連度の付け方に誰もが納得する一般的な規則があるわけではなく，またその関連度の順にユーザの検索意図の文書が並んでいるわけではない．また，ユーザはどこまで見れば，欲しい文書を見終わったことを確認できるのかは不明である．そして何より，個々の文書を読んで内容を確認する作業は効率的であるとは言い難い．

そこで近年，個々の文書を見ることなしに，文書集合の特徴を把握することを目的として，大量で複雑な文書情報をユーザにわかりやすく表示するための可視化技術に関して，数多くの研究が行われている．以下では，可視化手法の研究分野を概観し，それらの手法と問題点を説明する．

- **文書を 2 次元平面上に点として配置する手法**
  文書間の類似度に注目して，個々の文書間の距離を類似度の値を利用して 2 次元上に配置するシステム [1] である．例として，WEBSOM[71]，galaxy[8] などがある．自分の欲しい文書が全体の中でどういう位置付けにあるのか，また自分の欲しい文書の周りに他にどのような文書があるのかといったことが直感的に分かるため，優れた表現力を持つ可視化手法であると言える．しかし，文書集合が大量になると類似度行列が巨大になり，それを 2 次元平面上で表現するのは非常に難しい作業となる．そのため，大量の文書に対する検索には向いていない．また，似ている欲しい文書間の距離が遠くなることもしばしばあり，ユーザの検索意図とは関係なく，文書間の距離が決められてしまう場合がある．

- **Venn 図を用いた手法**
  複数の入力キーワードから，そのキーワードの組合せから得られる全ての検索結果の件数が分かる検索システムである．例として，InfoCrystal[20] がある．複数の検索語から作られる検索式全てを Venn 図を用いて可視化することにより，検索結果の件数が容易に理解でき，ユーザの検索式をに対して一つの指標が得られる．しかし，検索語が 4 つまではうまく可視化がされているが，5 つ以上になると包含関係の複雑さが増大し，うまく可視化されない検索式が現れてくるという問題点がある．

- **マトリクス形式の文献空間可視化手法**
  単語と文書についてそれぞれ次元と考え，出現頻度をベクトルで表す検索支援方法である．例として，検索結果絞り込み用インターフェース [5] がある．この方法では，単語や文書の空間的な位置を提示するのではなく，単語と文書を行と列に持つマトリクスを直接ユーザに表の形で提示し，検索操作を行や列に対する操作として実現している．大量の文書を扱うときにはマトリクスが巨大になってしまうため，得られた文書集合の一部をサンプリングする．サンプリングした文書集合に対して絞り込みに適したと思われる単語を選びだし，その単語のサンプリングされた文書集合内での出現状況を表で可視化する．サンプリングされた文書内での単語の出現状況から，ユーザはその単語で検索を行ったときの予想結果件数がある程度予想できるため，絞り込みを行う際の一つの指標が与えられる．しかし，サンプリングの方法によりキーワード候補語が変化する場合があり，また，提示されたキーワード候補語の中にユーザの目的の検索語が含まれていない場合があるという問題点がある．

2.3 本研究の位置付けと方針

本研究では，文書検索を効率的に行う手法について提案する．具体的には，

- キーワード検索と類似検索の両方を利用した検索手法
- 文書集合間の関連がユーザに分かりやすい形で提示される可視化手法
- 文書中に含まれる検索に有益であると思われる単語情報の可視化手法

のそれぞれに関して提案する．以下では，この 3 点について本研究ではどのように対処するかを説明する．

単語情報と文書情報を用いた手法

ユーザがある検索意図を持って検索を行うとき，主に使用するのは欲しい文書内に記述されていると考える単語，あるいはその概念を説明する単語であると考

えられる．その単語を含んだ文書集合を探す検索方法であるキーワード検索は，この考え方に沿った手法であり，一般に普及している．しかし，すでに述べたようにキーワード検索では，適切なキーワードの発見と検索式の作成という問題がある．

そこで，この問題点を解決するため，本研究では，ユーザがキーワードを入力する以外の方法で，検索概念を計算機に伝えることが必要であると考え，ユーザが見つけた一つの正解文書中の文書情報を利用した類似検索手法を用いる．類似検索は，文書内の単語の出現頻度が似た文書を類似文書として提示する手法である．ユーザ欲しい文書集合は，正解文書と似た単語の出現頻度である可能性が高いと考えられるため，検索意図に沿った検索が可能になる．また，キーワード検索において，一つの正解文書を見つけるのは，キーワード検索のみで大量の文書中から欲しい文書集合を取り出す作業と比べて，比較的容易な作業であるため，効率的な検索が実現できる．

文書間の関連の可視化

検索を行った結果，ユーザは得られた文書集合を見て，個々の文書を確認するか，検索をさらに続けるかやり直すかという判断を行う必要がある．その判断を支援をための手法として，可視化手法に関する研究が数多く行われているが，本研究ではマトリクス形式の文献可視化手法を採用する．この手法は，単語と文書を行と列に持つマトリクスを直接ユーザに提示することで，どの文書にどの単語が含まれているかの判断が可能になると同時に，文書間の関連の把握が容易になる．

単語情報の可視化

検索を行った結果，ユーザは得られた文書が検索意図に沿ったものであるかどうか，個々に確認することが必要になる．従来の手法では，この確認の作業がテキストを読んで確認するしかなく，原始的な方法で手間が掛かった．また，文書中に含まれる単語がどのような特徴を持った単語であるのかという判断を，ユーザ自身が行うしかなかった．そのため，ユーザはキーワードを予想して検索をを行うことが必要になり，従来のキーワード検索と同様の問題点が残ってしまう．

そこで本研究では，現在ユーザが注目している文書中の単語のうちキーワード候補語として適当であると考えられる単語を抽出し，その単語で絞り込んだ結果件数などの情報を可視化する手法を提案する．単語情報を可視化することで，ユーザは文書の内容を単語から推測できると共に，キーワード候補語の選択に対する判断材料が与えられるため，次の検索の指標が与えられる．

3. 文書情報の可視化による検索絞り込み支援

本章では，検索に有効であると思われる文書情報を，可視化を用いてユーザに把握させ，それにより検索を支援する手法を提案する．

本研究では，文書データベースからユーザが得た一つの正解文書の文書情報を利用し，その情報の可視化により検索の支援を行う．以下で，用いる文書情報について説明し，その情報を用いた，文書間の関連と単語情報についての可視化手法についてそれぞれ説明を行う．

3.1 用いる文書情報

本節では，可視化による検索支援を行うときに用いられる文書情報が，どのように計算されるかを説明する．

3.1.1 単語の切り出し

本研究では，文書を単語の集合として扱う．どのような単語が含まれているかにより，個々の文書の内容はある程度推測できると考えられるため，各文書に含まれる単語の出現頻度を求める．そのため，全文書に含まれる全単語を切り出すために形態素解析を行う．形態素解析には本学において開発された日本語形態素解析システム茶筅 [3] を用いる．全文書に対して形態素解析を行い，その結果得られた全名詞のうち，キーワード検索においてキーワードとして用いられる可能性があると考えられる，一般名詞，固有名詞，副詞可能，サ変接続，形容動詞語幹について，それぞれの出現頻度を計算する．

3.1.2 文書間の類似度

次に，文書を文書ベクトルとして表現し，文書間の類似度を計算する方法について説明する．まず対象になる文書集合に含まれる全単語数を N とするとき，文書 D_i を文書ベクトル d_i として表現する [17][18][19]．

$$d_i = (w(i,1), w(i,2), \ldots, w(i,N))$$

ここで，$w(i,k)$ は文書 D_i に対する単語 W_k の重みであり，次式で定義される．

$$w(i,k) = tf(i,k) \cdot idf(k)$$

ここで用いた $tf \cdot idf$ 値は文書に対する単語の特徴量を表した値であり，次式で定義される．

$$tf(i,k) = (\text{文書 } D_i \text{における単語 } W_k \text{の出現頻度})$$

$$idf(k) = \log\left(\frac{\text{全文書数}}{\text{単語 } W_k \text{が現われる文書数}}\right) + 1$$

tf 値は，文書内でその単語の出現回数が多いと高い値を，出現回数が少ないと低い値をとる．また，idf 値は，少ない数の文書でしかその単語が使われていないと高い値を，多い数の文書でその単語が使われていると低い値をとる．$tf \cdot idf$ 値は tf 値と idf 値の積で求められるため，特定の文書中に集中して出現する単語に対しては大きい値を，逆に多くの文書にまんべんなく出現する単語に対しては低い値を取り，その文書を特徴づける単語の選択に有効な指標である．

文書 D_i と D_j の類似度 $sim(D_i, D_j)$ は，文書ベクトル間のなす角度として次式で定義される．単語の出現頻度が似た文書間は類似度の値が高く，大きく異なる文書間は類似度の値が低くなる．

$$sim(D_i, D_j) = \frac{\sum_{k=1}^{N} w(i,k) \cdot w(j,k)}{\sqrt{\sum_{k=1}^{N} w(i,k)^2 \cdot \sum_{k=1}^{N} w(j,k)^2}}$$

これらの文書情報を検索を通してどのように用いるかは 3.2 節，3.3 節で述べる．

3.2 文書間の関連の可視化に基づく検索支援

2 章で説明した文書間の関連を，マトリクス形式を用いてどのように可視化するかについて説明を行う．マトリクス形式の可視化手法は，文書ベクトルから成るマトリクスの一部を，ユーザに分かりやすい形で見せる可視化手法である．

文書ベクトルの手法を用いると，全単語数を N，全文書数を n，文書 D_i に対する単語 W_k の出現頻度を $tf(i,k)$ とすると，文書 D_i は文書ベクトル d_i として表現される．

$$d_i = (tf(i,1), tf(i,2), \ldots, tf(i,N))$$

全ての文書の全ての単語に対する出現頻度は n 行 N 列の行列で表現される．そのため，本研究でのユーザが行う文書検索は全て，この行列に対する操作により行われる．キーワード検索において，単語 W_k での検索は行列中から $tf(i,k) > 0$ となる文書 D_i に対応する行 i を集める操作である．類似検索において，正解文書 D_i に対する検索は d_i に対する内積の値の順に行列を並べ替える操作である．

そのため，ユーザが検索を行うとき，この行列情報を把握できればスムーズな検索が可能になる．しかし，この行列は文書数に応じて巨大になりユーザに全てを把握させるのは難しい．本学の論文の概要 500 件に対して検索を行うとき，総単語数は 4339 語になり，行列の大きさは 500×4339 にもなる．

そこで本研究では，一つの正解文書を特徴的に表していると考えられるキーワード候補語の列を抽出して，行列を小さくして正解文書の特徴単語の他の類似文書における出現状況を見やすく表示する．これにより，正解文書と類似文書の関連が強調され，検索の支援になる．

図 2は No.2 の文書が正解文書であったときの行列の縮小化の例である．No.2 の文書に対する類似度の高い文書の順に行列が並び替えられ，同時に行列の列の数が減少している．No.2 の文書は総単語数の大きさの文書ベクトルで表現されるが，そのうち No.2 の文書において特徴的に出現している単語に関する列にだけを抽出し，全体の行列の縮小化を実現している．縮小化の際，No.2 の文書の特徴ができるだけ失われないようにするため，No.2 文書の特徴をよく表していると考えられるキーワード候補語を用いる．

行列の縮小化の結果，正解文書が No.73 の文書とは「検索」と「類似」という単語に関して関連があり，No.18 の文書とは「類似」と「画像」という単語に関して関連があることが分かる．これにより，類似度だけからでは分からない正解文書と類似文書の繋がりが容易に分かる．

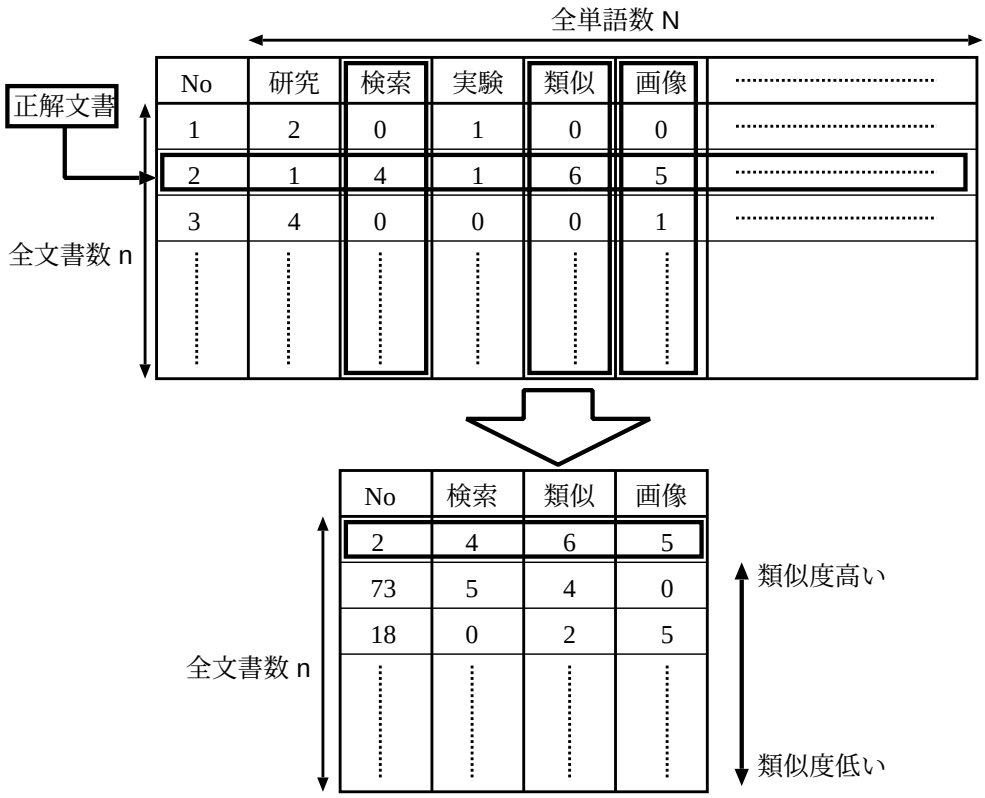

全単語数 N
正解文書
全文書数 n
No
研究
検索
実験
類似
画像
1 2 0 1 0 0
2 1 4 1 6 5
3 4 0 0 0 1
No
検索
類似
画像
2 4 6 5
73 5 4 0
18 0 2 5
全文書数 n
類似度高い
類似度低い

図 2 **行列の縮小化**

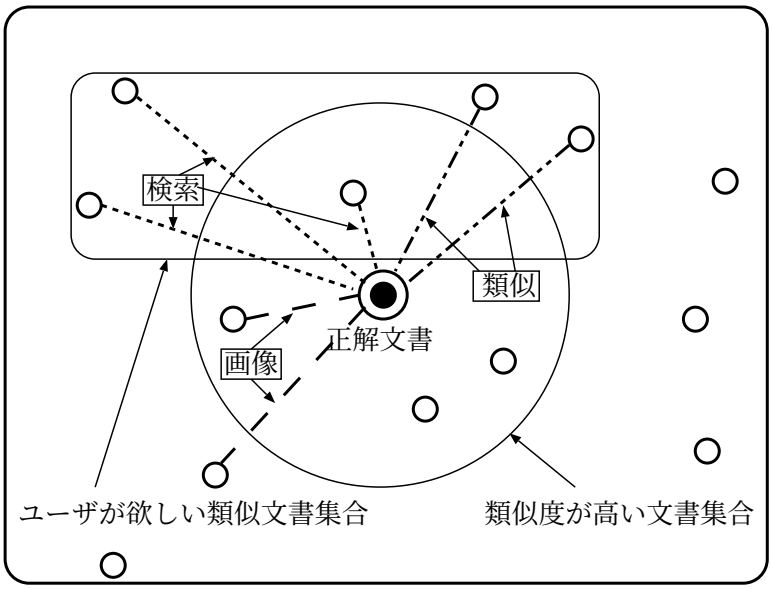


図 3 **正解文書との関連**

図 3は，ユーザが「検索」と「類似」という語に関心があるときの例である．このとき，可視化により正解文書とユーザが欲しい類似文書の単語での繋がりから文書の内容把握が容易になる．

3.3 単語情報の可視化に基づく追加キーワードの決定

正解文書中のキーワード候補語をどのように可視化するか説明する．本研究ではキーワード候補語の判断基準を以下に基づいて決定した．

- 正解文書中の出現頻度
- 他の文書中での出現状況

正解文書中の出現頻度と他の文書での出現状況を考慮した結果，tf 値と idf 値の積である $tf \cdot idf$ 値の大きな単語をキーワード候補語とふさわしいと考えた．

この結果より，キーワード候補語は数値的にキーワードとして有用であると判断された単語である．キーワードとして用いれば文書数を減少させ，かつ正解文書を特徴的に表現した単語である．しかし，これらの単語が本当に検索者の意図

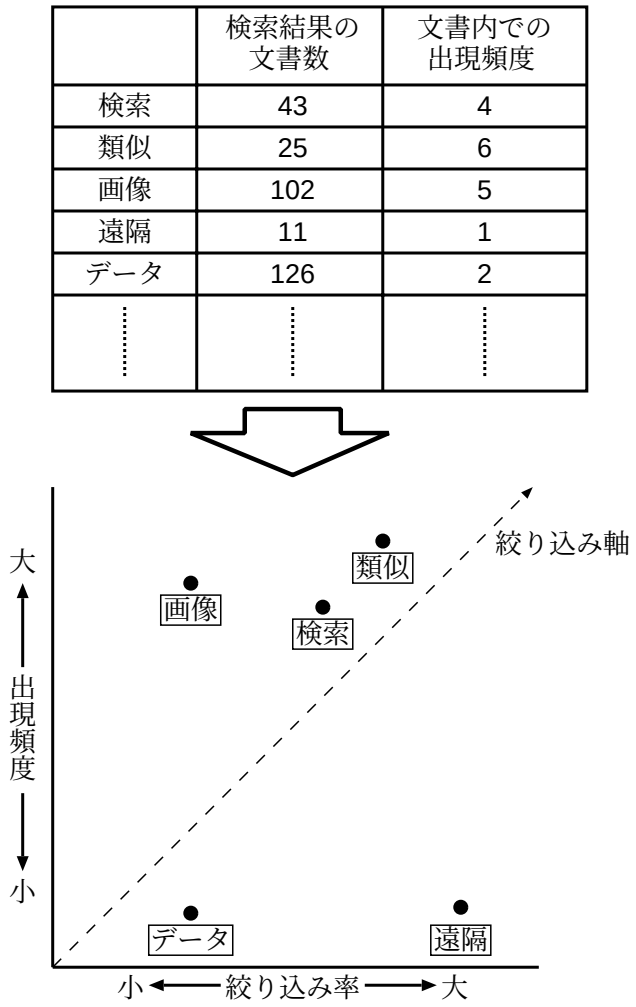


図 4 **キーワード候補語情報の可視化**

に沿った単語かどうか，この単語で絞り込みを行った結果ユーザの意図する文書集合が得られるかどうかは，ユーザが実際に検索を行って結果を確認する必要がある．そのため,「どの単語を用いて絞り込みを行えば，どのような検索結果が得られるか」という情報を，計算機が効果的にユーザに提示することが必要になる．

そこで，絞り込みの際の結果の評価基準として,「結果の件数」と「結果の内容」という二つの基準が考えられる．

「結果の内容」については，検索結果の文書集合がユーザの欲しい文書集合であるかを判断することは難しく結果を見たユーザにしか分からない．そのため，

計算機は「結果の内容」に関して，絞り込みの支援となる情報を提示できない．

「結果の件数」については以下のことを考える．最初の検索により，得られた検索結果が 10000 件であったとする．絞りこんだ結果が 0 件になる単語は明らかに絞り込みに適してない単語であると言える．また，ある単語はその文書集合を 5000 件に絞り込むが，別の単語では 100 件に絞り込むとき，絞り込みに用いられた単語が共にユーザの検索概念に沿った単語であったなら，後者は前者よりユーザにとってより好ましい結果だと考える．そのため,「結果の件数」については客観的な情報であるため，絞り込みの件数の提示が絞り込みの支援となると考えられる．

そこで，その単語で絞り込んだ結果と得られる文書数の情報をユーザに理解させるため，キーワード 候補語の「結果の件数」の情報をを可視化する．また，正解文書中で出現頻度が高い単語はユーザの検索意図に沿っている可能性が高いと考え，キーワード 候補語の正解文書内での出現頻度も同時に可視化する．これにより，正解文書の特徴の把握が容易になると考えられる．そのため，キーワード 候補語の中にユーザの検索概念に近い単語が存在したとき，その語による絞り込みの件数が分かり，検索の支援になる．

図 4はこれらを考慮した結果，正解文書中のキーワード 候補語の特徴を可視化した結果の例である.「画像」という単語は，正解文書中に多く含まれているため，文書を特徴づけるのに大きな影響を与えているが，絞り込みに用いると大量の結果が得られることが分かる.「遠隔」という単語は，正解文書中にあまり含まれていないため，文書の特徴づけにはあまり影響を与えていないが，絞り込みに用いると少ない文書が得られることが分かる．

ユーザはこれらの情報と自分の検索概念とを照らし合わせて検索を行うことができる．検索式の作成は，検索結果を絞り込みたいときは AND 検索，幅広い内容の文書が得たいときには OR 検索を用いて検索することができる．これらの検索を，単語の意味と検索結果を理解した状態で行えば，ユーザはより効率的な検索が可能になる．

4. 文書情報の可視化に基づく検索システムの構築

本章では，3 章で提案した可視化手法を実装した提案手法による検索手順について説明する．そして，作成した手法を用いて，検索を行った例を示す．

4.1 検索手順とユーザインターフェース

4.1.1 提案手法による検索手順

図 5に本手法の構成を示す．本手法は大きく分けて二つの部分化から構成される．一つの正解文書を見つけるまでの検索部分と，正解文書を見つけた後の検索部分である．

正解文書の検索

一つの正解文書の検索には，キーワード検索を用いる．ユーザは,「キーワードを考える」「検索実行」「結果を見る」「新しいキーワードを考える」のプロセスを繰り返しを行い，正解文書を探す．

文書集合の検索

正解文書を発見した後は，その正解文書に関して類似検索を行う．類似検索により，正解文書と出現頻度の似た文書が他数提示される．この提示された類似文書集合の中から，ユーザは「類似文書を見る」ことにより，自分が欲しい文書集合を取り出す．このとき，本研究では文書間の関連の可視化を行い，他の類似文書が正解文書とどのようなつながりがあるのかを把握させる．類似文書が大量であったときは，キーワード候補語を用いて絞り込みを行う．

図中の「類似文書を見る」と「キーワードの可視化結果を見る」部分に対して，本手法がどのような可視化を行い，どのような検索支援が行われるかを説明する．

4.1.2 ユーザインターフェース

提案した手法を実装したシステムのユーザインターフェースを図 6に示す．検索手順は結果表示部分と単語情報表示部分の 2 つの部分から構成されている．

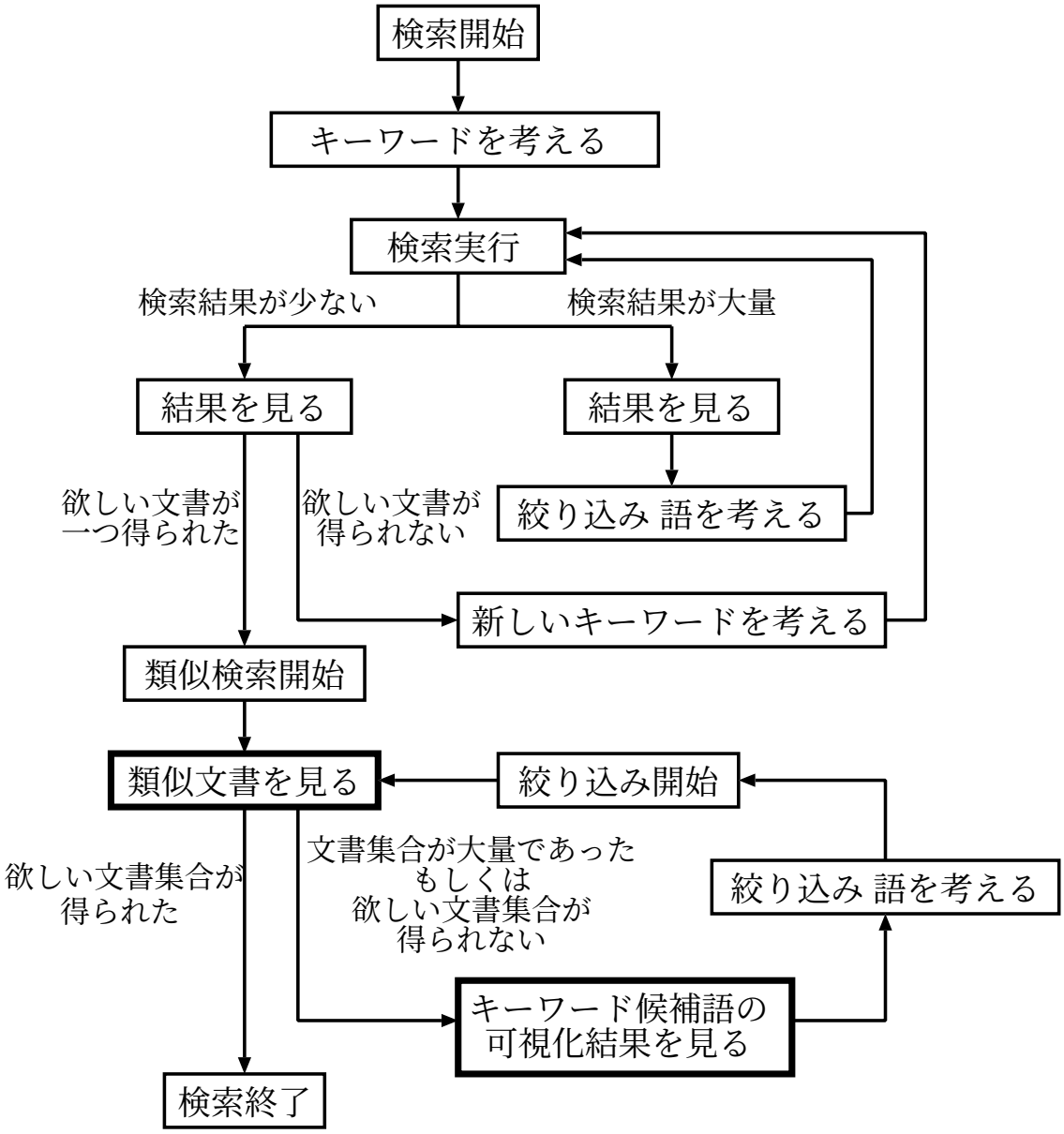


図 5 **提案手法による検索の手順**

結果表示部分

図 6(a) に示すように，検索の結果得られた文書集合を表示する部分である．検索ウインドウでは常に一つの注目文書に対する表示を行う．注目文書とは列の一番上に注目された文書のことをいう．

キーワード検索を行ったときには，キーワード語を最も含んだ文書が注目文書となる．ユーザが見つけた文書について，類似検索を行ったときには，その文書が注目文書となる．

以下で各オブジェクトの説明を行う．

キーワード ボックス： 入力キーワードを入れるためのテキストボックスである．最初にキーワードを入力するとき，絞り込みを行うときに用いる．

検索ボタン： 検索の開始をシステムに伝えるためのボタンである．

リセット ボタン： 新しく検索を行うためのボタンである．現在の検索を終了するときに用いる．

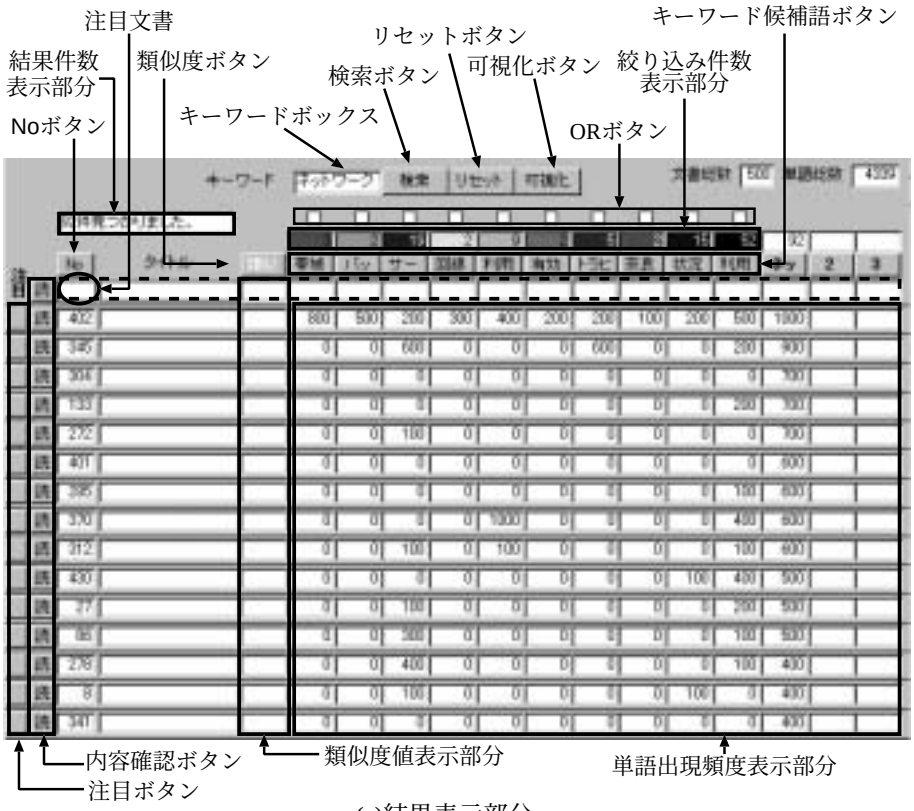
可視化ボタン： 現在のキーワード候補語に対する可視化ウインドウを表示させるためのボタンである．可視化結果から注目文書中に含まれる単語の特徴を把握するときに用いる．

OR ボタン： クリックすることにより単語にマークをつけ，さらにクリックすることでマークを外すことができるボタンである．検索ボタンをクリックすると，マークがついた全ての単語に関する OR 検索が実行できる．ユーザの検索意図に沿った複数の単語について，同時に検索を行うときに用いる．

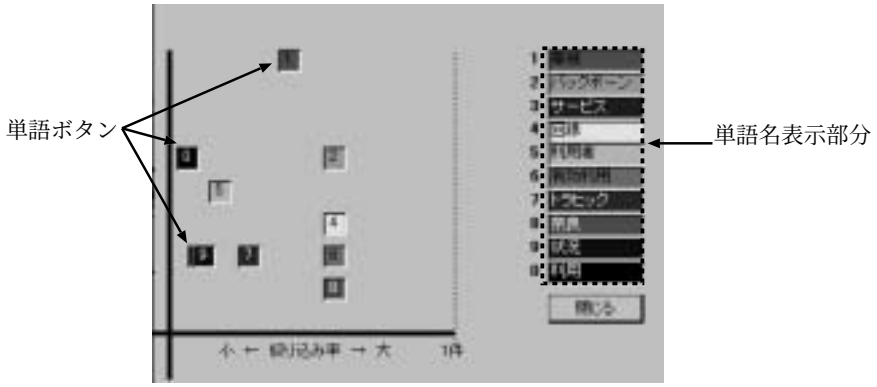
No ボタン： No 順に文書を並べ替えるためのボタンである．

類似度ボタン： 現在の注目文書に対する類似度順に文書集合を並べ替えるためのボタンである．

キーワード 候補語ボタン： ラベルに書かれているキーワード候補語の名前を表示し，その名前の単語で絞り込み検索を行うためのボタンである．ユーザの



(a)結果表示部分



(b)単語情報表示部分

図 6 **実装システムのユーザインターフェース**

検索意図に沿った単語が提示されていたとき，キーワードを入力することなく，このボタンをクリックすることで検索を行うことができる．

注目ボタン： ユーザが指定した文書を注目文書とするためのボタンである．このボタンを押すことで，注目文書が変化すると同時に，現在の文書集合は新しい注目文書に対する類似度順に並べ替えられ，キーワード候補語も新しい注目文書のものに変化して表示される．

内容確認ボタン： ユーザが指定した文書の内容を実際に読んで確認するためのボタンである．ユーザが内容を確認するという判断が行われたときに用いる．

絞り込み件数表示部分： 下に書かれた単語で現在の文書集合を絞り込みを行ったときの結果の件数を表示している．

結果件数表示部分： 現在の文書の件数を表示している．

類似度値表示部分： 現在の注目文書に対する，他の文書の類似度を表示している．0 から 100 までの数字で表され，大きいほど注目文書に対する類似度が高い．

単語出現頻度表示部分： 現在の注目文書のキーワード候補語の他の文書内での出現頻度を表示している．表示されている数字は，その文書内の単語の出現頻度に 100 を掛けた数値である．

単語情報表示部分

図 6(b) に示すように正解文書中のキーワード候補語の情報を表示する部分である．

単語名表示部分： キーワード候補語の単語名を表示している．

単語ボタン： 正解文書から抽出されたキーワード候補語で現在の文書集合を絞り込んだ件数を x 軸，文書内で何回用いられているかを y 軸とし，2 次元平面上の位置で単語の情報を表示している．ボタンを押すことで，その単語による絞り込み検索が行える．

4.2 システム動作例

本節では，作成したシステムを用いて実際に文書検索を行った結果を示す．ここでは，奈良先端科学技術大学大学院大学の過去の修士論文の内容梗概 500 件を対象として，画像に関する論文を検索する状況を想定する．

キーワードの入力

最初に，ユーザは欲しい文書集合を得るためにキーワードボックスにキーワードを入力する．

図 7は,「画像」というキーワードを入力し，検索を行った結果を表している．「画像」という単語を含む文書を検索し，103 件の文書が得られ，出現頻度の高いものから順にリスト表示した結果を表示している．No.257 の文書が,「画像」という単語の出現頻度が最も高かったため，一番上の列に表示されており，注目文書となっている．この注目文書に対する 10 個のキーワード候補語がボタンのラベルとして表示され，同時にその上に絞り込み結果件数の情報も表示している．ユーザは，この結果を参考にして検索を開始する．

正解文書の検索

次に，図 7のように得られた文書集合の中から，ユーザは自分の検索概念に合う文書を一つ探す．現在注目文書となっている No.257 の文書がユーザの検索概念に合う文書でなかったとする．このとき，行列情報を有効的に利用して正解文書の検索を行う．例えば，キーワード候補語である「検索」「データベース」という単語を共に含んだ No.315 と No.71 の文書は，注目文書と似ている内容の文書である可能性が高く，ユーザの検索概念に合わない文書の可能性が高いという判断ができる．このような可視化結果からの判断により，効率的な正解文書の検索が可能になる．

正解文書からの検索

正解文書が見つかったとき，ユーザはその正解文書の情報を用いて検索を行う．

図 8は，No.498 の文書「ビデオシースルー型拡張現実感のためのステレオ画像合成」がユーザの欲しい文書であり，その文書を注目文書として注目ボタンを押

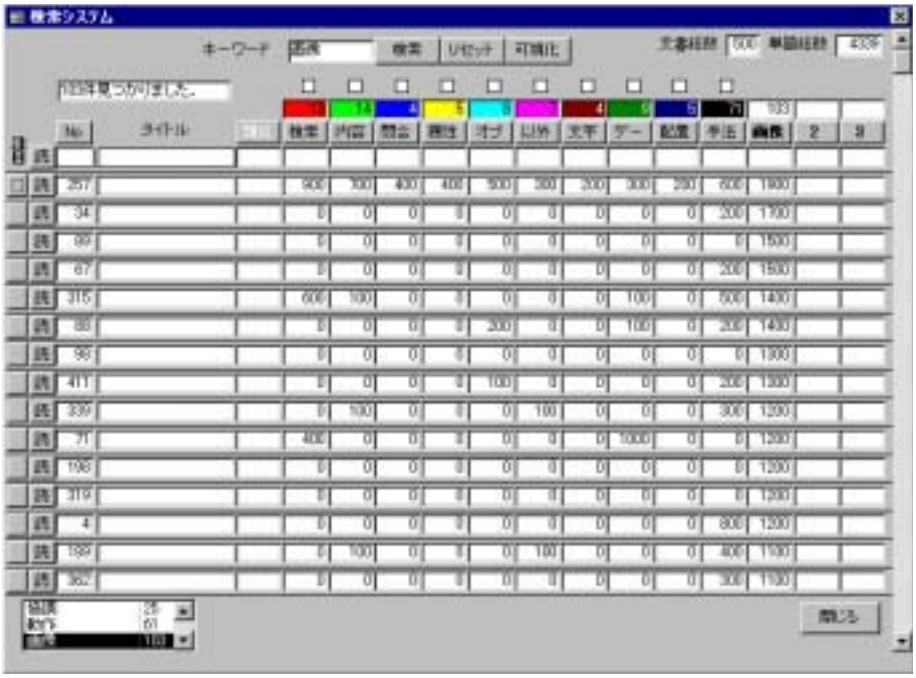


図 7 **画像というキーワードで検索を行った結果**

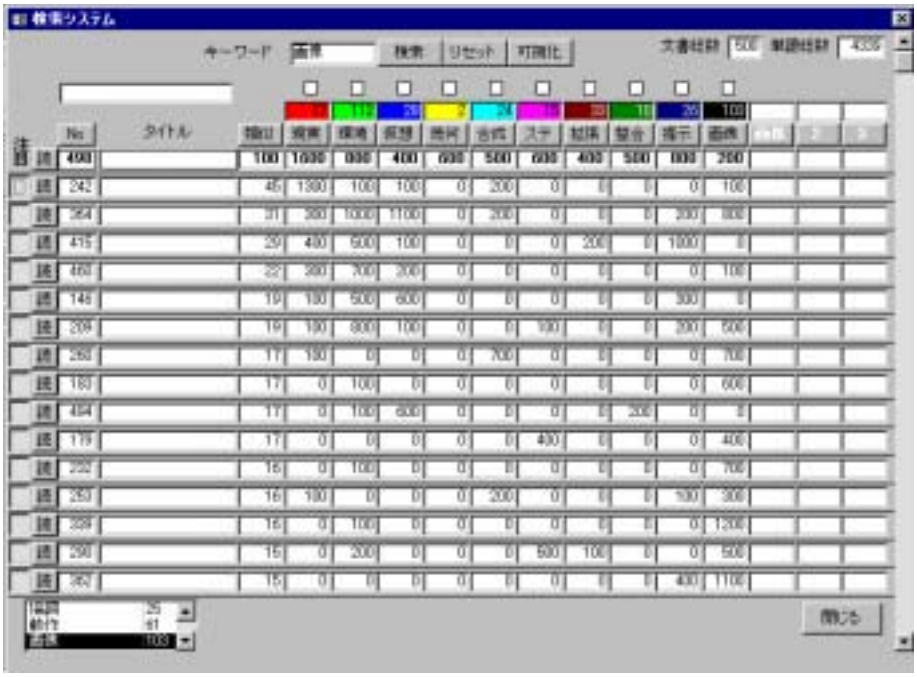


図 8 **正解文書に対する結果表示**

した結果である．No.498 の文書が正解文書に指定されるのに伴い，文書集合が正解文書に対する類似度順に並べ替えられ，キーワード候補語も変化して表示している．

このとき，得られた正解文書のキーワード候補語を可視化した結果が図 9である.「現実」「環境」「仮想」といった単語の正解文書内での出現頻度が高く，正解文書の特徴をよく表した単語であることが分かる．また同時に,「環境」という語は絞り込み率が低いため，得られる検索結果が多いということが分かる．

ユーザはこの可視化情報から検索を行う．この可視化結果から検索を行う．図 10は,「画像」と「仮想」という二つのキーワードで AND 検索を行った結果を表している．その結果，得られた 7 件の文書は以下の通りである．

242： 強調現実感による卓上作業の視覚的支援に関する研究」

364： カラー画像と距離画像の融合によるパノラマ仮想空間の構築

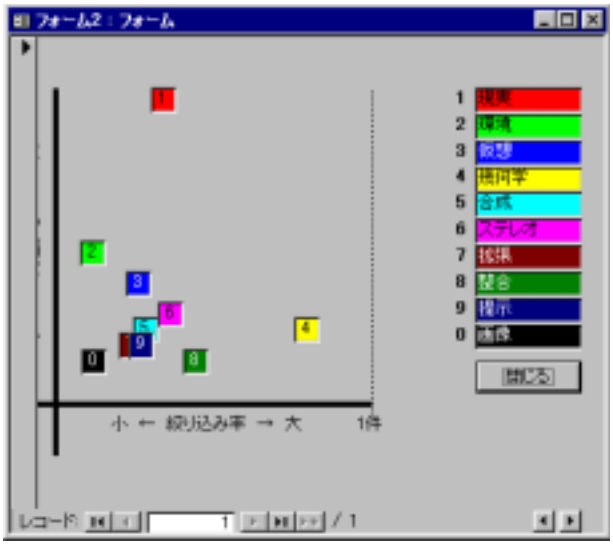


図 9 **正解文書の文書情報の可視化表示**

460： 強調現実感技術を用いた電子部品検査の支援環境に関する研究

209： 光学式位置センサと大型ディスプレイを用いた VR 環境

466： 多地点全方位画像を利用した仮想環境の構築

119： スネークスによる輪郭抽出アルゴリズムへのニューラルネットワークの適用と学習用データ作成の手法

正解文書「ビデオシースルー型拡張現実感のためのステレオ画像合成」に内容的に近い文書が類似度順に並べられ，なおかつ「仮想」という単語が含まれている文書のみが提示された．ここで，得られた 466 の文書は正解文書と内容が近い文書であったにも関わらず，類似検索では類似度値が低いため，仮想という語によるキーワードが思いつかないと発見が難しいと思われる文書であった．この検索手法の結果，欲しい文書集合が容易に獲得できた．

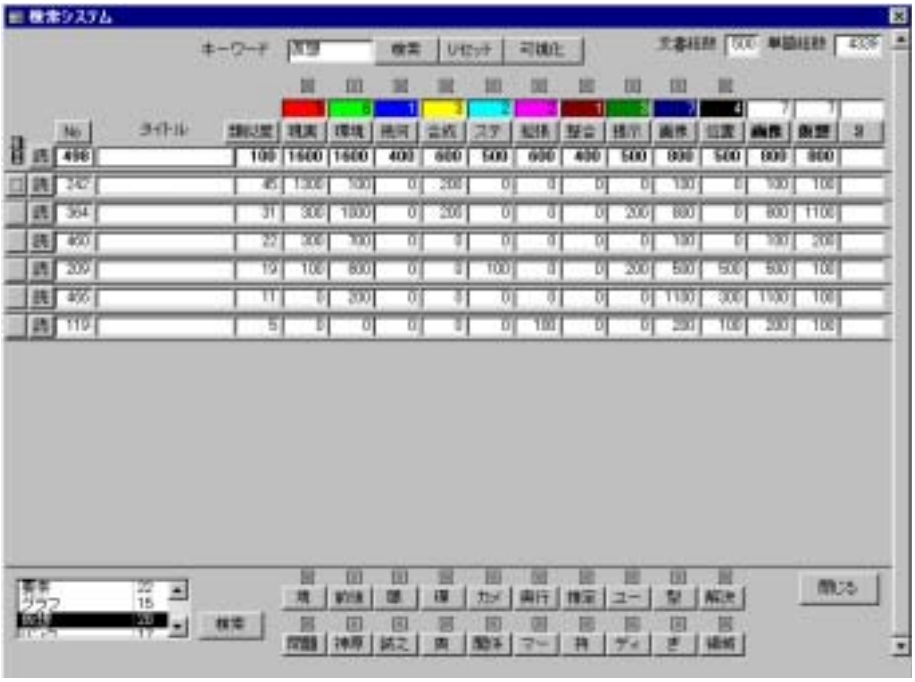


図 10 **仮想というキーワード 候補語で検索を行った結果**

5. 実験

5.1 実験課題

本研究では，提案手法を評価するために実験を行った．情報検索システムの評価は，精度のよい結果が得られるかどうかという基準で行う必要がある．精度がいい結果が得られたかの性能評価には，再現率と適合率が用いられることが多い．しかし，同じ検索目的に対しても正解と判断される文書集合はユーザごとに異なるため，正解に関する客観的な基準は存在しないため判断ができない．

そこで，提案手法を用いたとき，どのような結果が得られるのかを実際に使ってみて，その結果どのような支援が得られたのかを検証していく．

5.1.1 実験の手順

以下を調査者が行う．

- 文書検索に関する論文の検索
- ロボットに関する論文の検索
- 暗号に関する論文の検索

奈良先端科学技術大学大学院大学の過去の修士論文の内容梗概 500 件を対象として上に挙げた検索を行い，その結果から本手法を評価する．提案手法を用いることで，正解文書の情報を利用した検索支援が行えると考えられる．

5.1.2 実験結果

文書検索に関する論文の検索

文書検索に関する論文の検索を行うため，まずキーワードに文書を入力した結果を図 11に示す．その後に絞り込みのため，検索というキーワードを入力した結果を図 12に示す．それぞれの検索結果は 17 件，10 件となり，絞り込みが行われ件数が減少している．10 件に減少したとき，その中から欲しい文書があるかの確

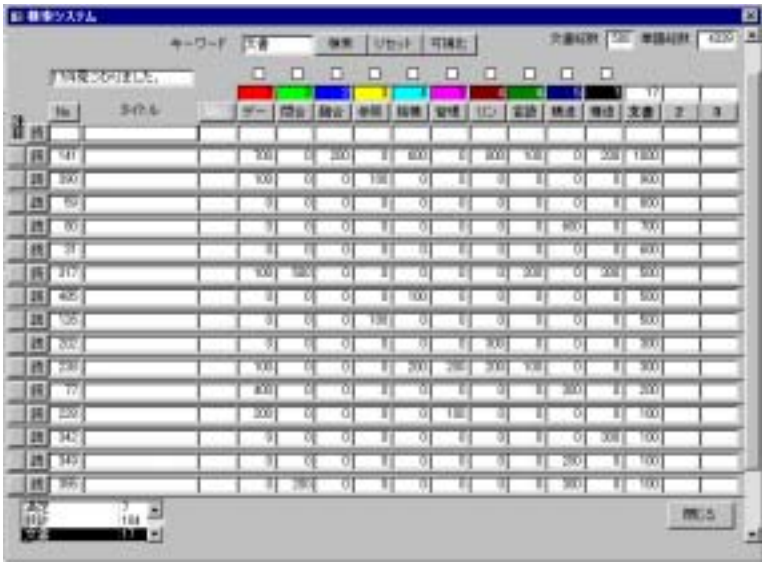


図 11 **文書というキーワードで検索を行った結果**

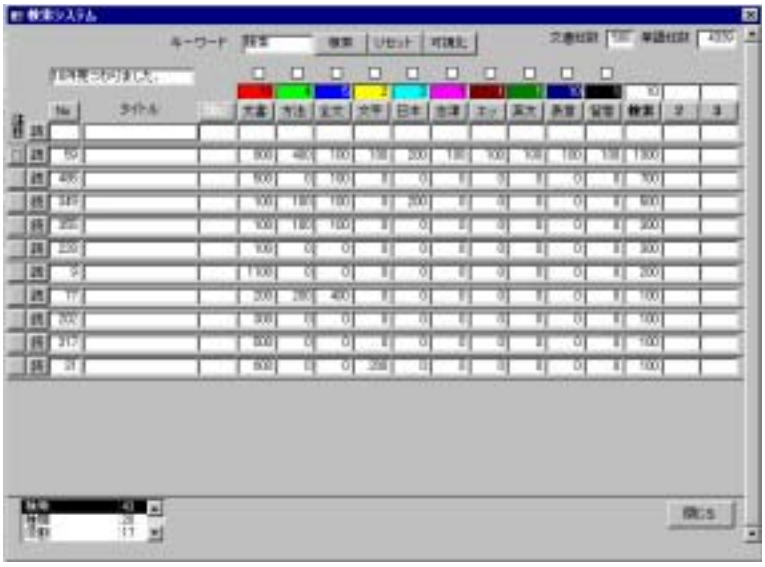


図 12 **検索というキーワードで絞り込みを行った結果**

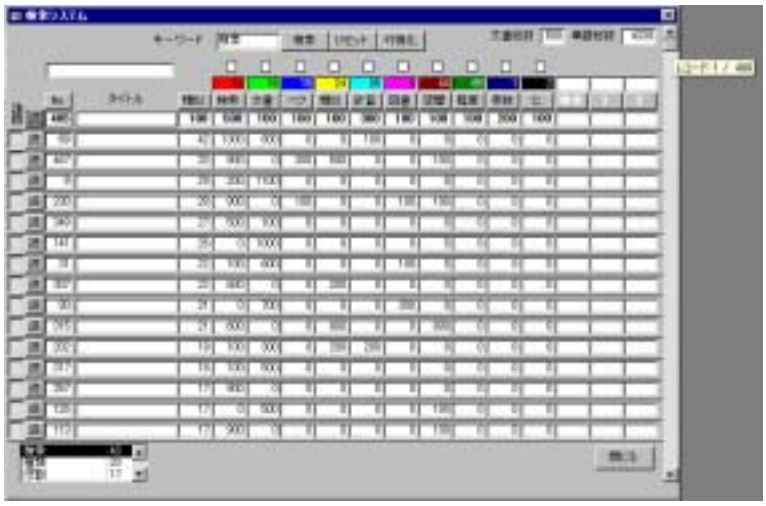


図 13　No.485 の文書を正解文書とした結果 1

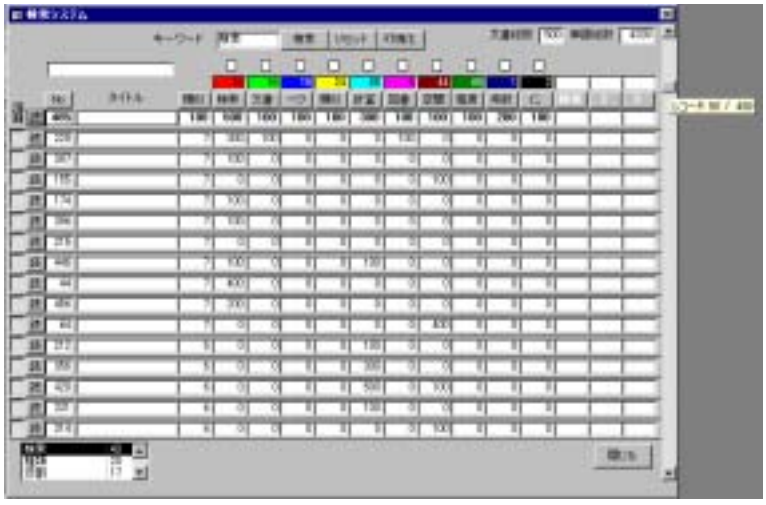


図 14　No.485 の文書を正解文書とした結果 2

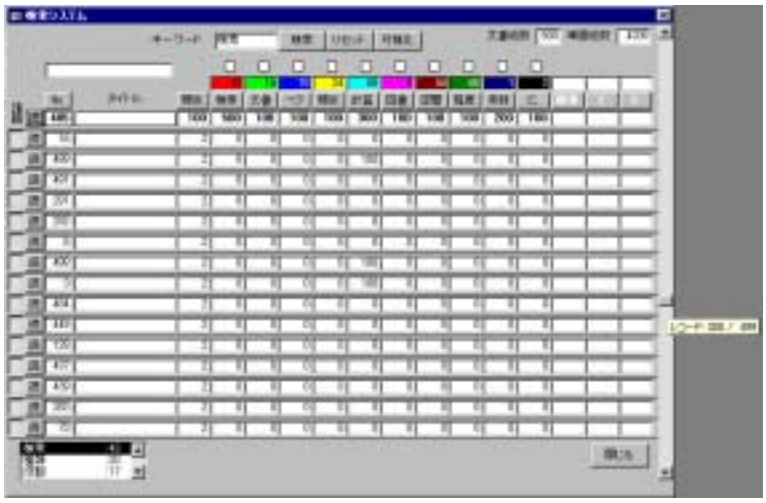


図 15　No.485 **の文書を正解文書とした結果** 3

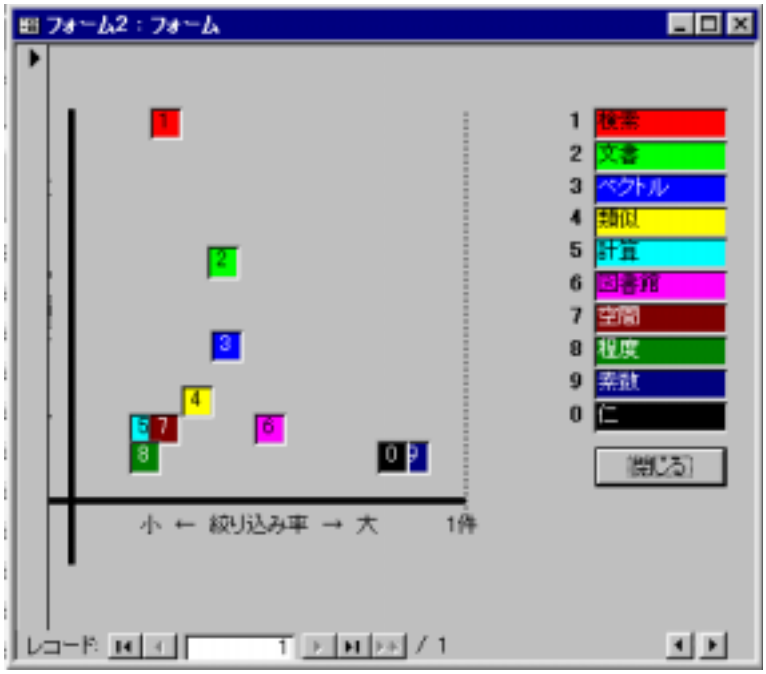


図 16　**正解文書における単語情報の可視化結果**

認を個々に行い欲しい文書を得た．No.485 の文書である「ベクトル空間モデルを用いた論文検索におけるベクトル化法および評価」というタイトルの論文が検索意図と合致した論文であった．

そこで，この論文を正解文書として指定する．指定した後の結果を図 13に示す．全ての文書が，No.485 の文書に対する類似度順に並べ替えられている．また，No.485 の文書に対するキーワード候補語 10 個が提示されている．この結果を図 16に示す．No.485 のキーワード候補語は,「検索」「文書」「ベクトル」「類似」「計算」「図書館」「空間」「程度」「素数」「仁」であった．

ここで，図 14，図 15に正解文書に対する類似度値が 50 番目付近の文書，300 番目付近の文書のキーワード候補語の出現状況を示す．これらは共にキーワード候補語の出現頻度が低く，検索意図とは大きく離れた文書である可能性が高い．そのため，類似度が上位のものを見ることで，検索が効率的に行えることが可視化結果から分かる．

そこで，図 13付近に注目して類似文書の検索を行う．類似度が高いが，文書という単語を含んでいない No.467 の文書は，検索意図に沿った文書ではないという推測ができた．この可視化結果に基づく検索から,「文書」と「検索」を両方含んだ，もしくはいずれかを含んだ文書が分かり，検索意図に沿った文書 No.59 と No.9 を発見した．

ロボットに関する論文の検索

ロボットに関する論文の検索を行うため，まずキーワードにロボットを入力した．その結果をそれぞれ図 17に示す．32 件の検索結果が得られたことが分かる．ここで欲しい文書があるかの確認を行い，欲しい文書を得た．No.291 の文書である「遠隔超音波画像診断におけるプローブ操作用ロボット」というタイトルの論文が検索意図と合致した論文であった．

そこで，この論文を正解文書として指定する．指定した後の結果を図 18に示す．全ての文書が，No.291 の文書に対する類似度順に並べ替えられている．また，No.291 の文書に対するキーワード候補語 10 個が提示されている．この結果を図 19に示す．No.291 のキーワード候補語は,「ロボット」「遅延」「操作」「画像」「遠隔」「超音波」「衛星通信」「衛星」「医療」「回線」であった．

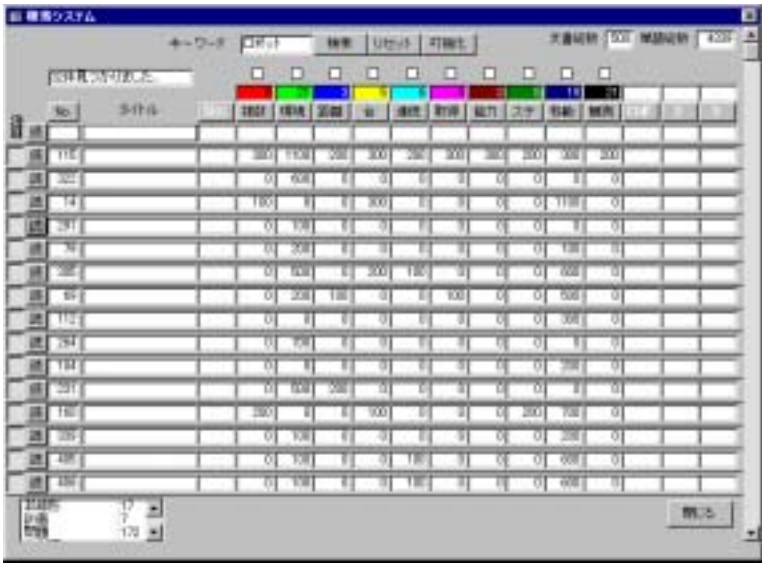


図 17 **ロボットというキーワードで検索を行った結果**

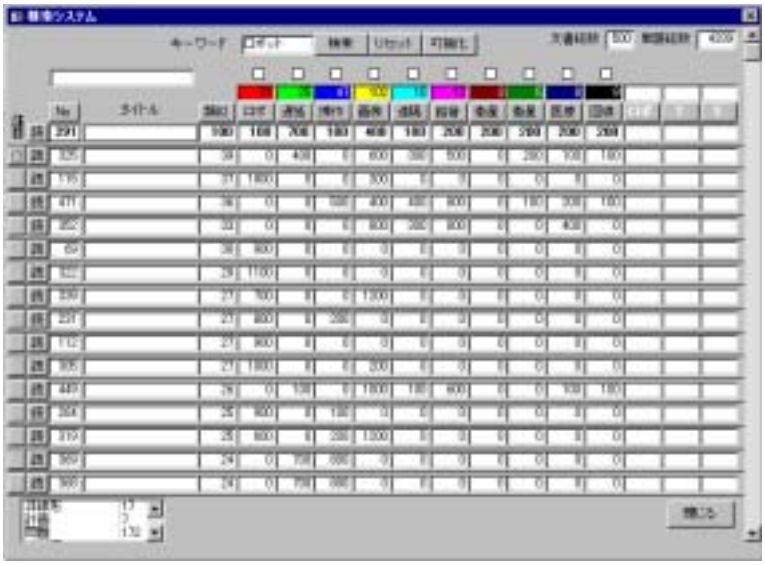


図 18 No.291 **の文書を正解文書とした結果**

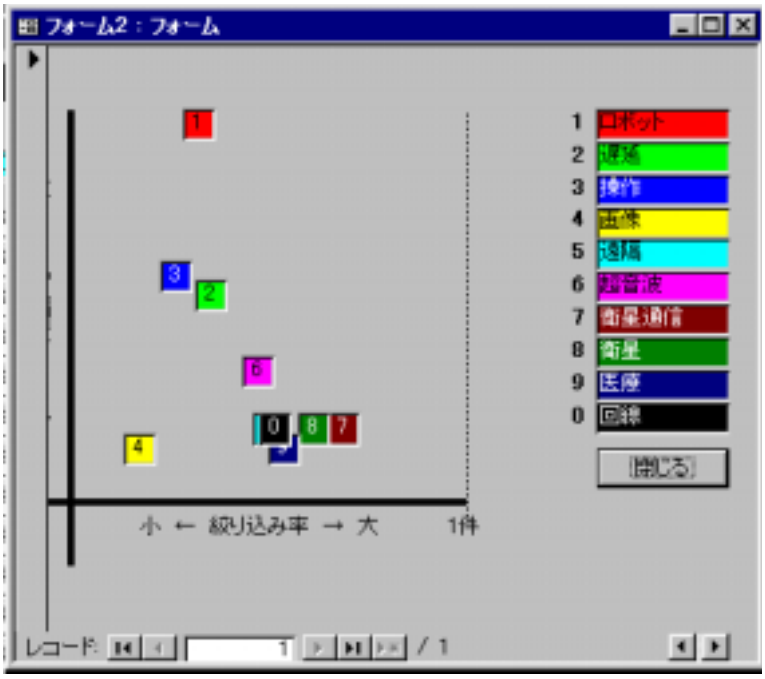


図 19　**正解文書における単語情報の可視化結果**

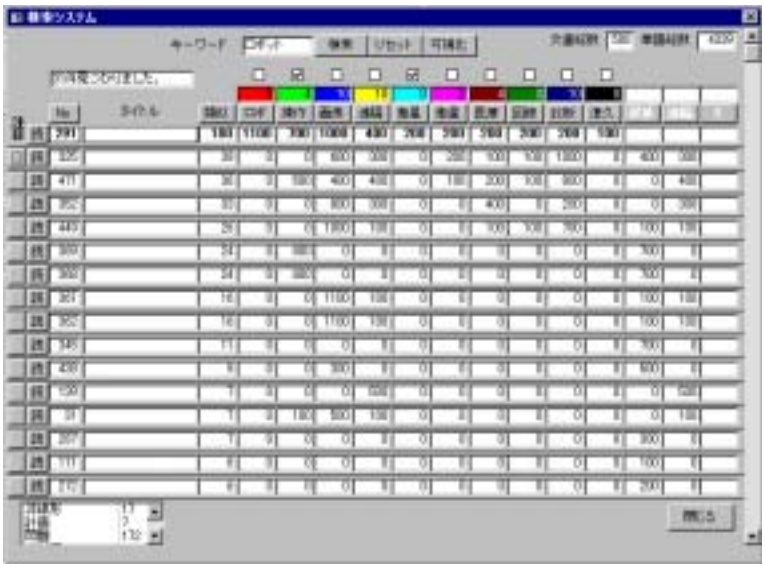


図 20　**遠隔 OR 操作に関するキーワードで検索を行った結果**

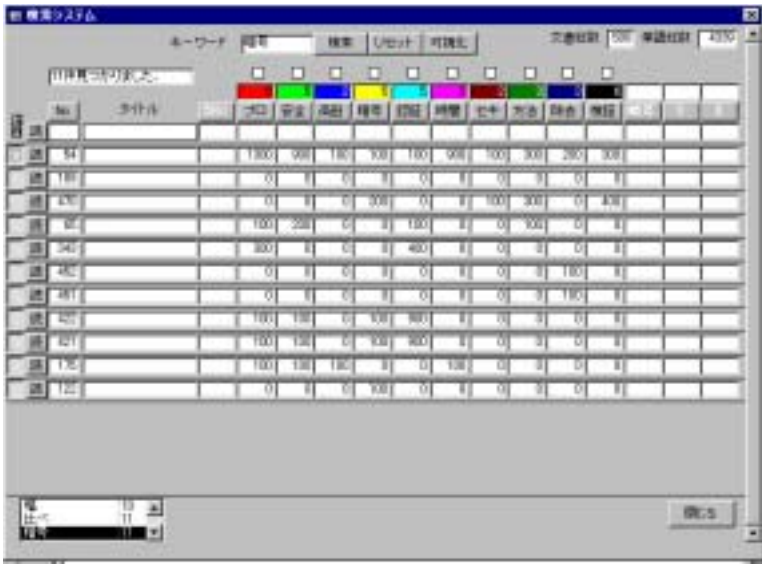


図 21 **暗号というキーワードで検索を行った結果**

ここで絞り込みに適していると考えられるのは「遠隔」と「操作」の単語であると思われたため,「遠隔」と「操作」のキーワードに関する OR を取る検索を行った．その結果を図 20に示す．検索上位に欲しい文書を発見することができた．

暗号に関する論文の検索

暗号に関する論文の検索を行うため，まずキーワードに暗号を入力した．その結果をそれぞれ図 21に示す．10 件の検索結果が得られたことが分かる．ここで欲しい文書があるかの確認を行い，欲しい文書を得た．No.54 の文書である「時間の概念を含む暗号プロトコルの公理的安全性判定法について」というタイトルの論文が検索意図と合致した論文であった．

そこで，この論文を正解文書として指定する．指定した後の結果を図 22に示す．全ての文書が，No.54 の文書に対する類似度順に並べ替えられている．また，No.54 の文書に対するキーワード候補語 10 個が提示されている．この結果を図 23に示す．No.54 のキーワード候補語は,「プロトコル」「安全性」「公理」「概念」「暗号」「公理系」「時間」「検証」「記述」「除去」であった．

ここでは絞り込みに適していると考えられるキーワード候補語がなく，また検

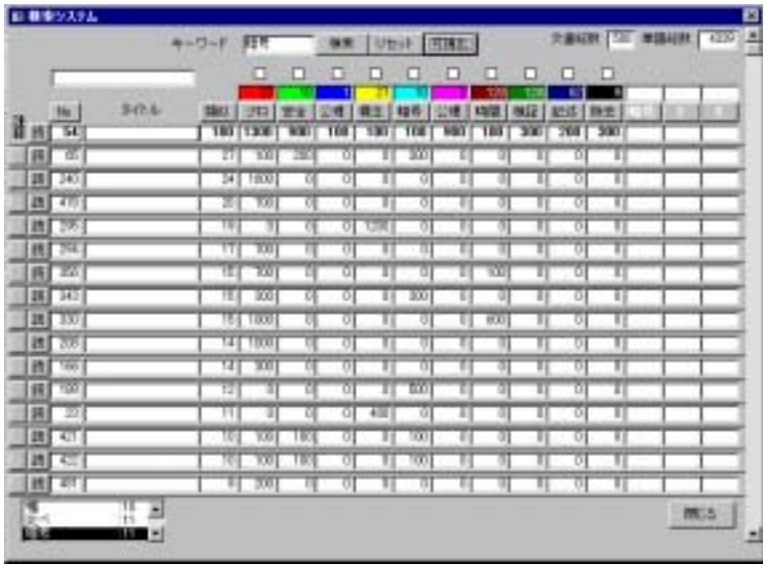


図 22　No.291 の文書を正解文書とした結果

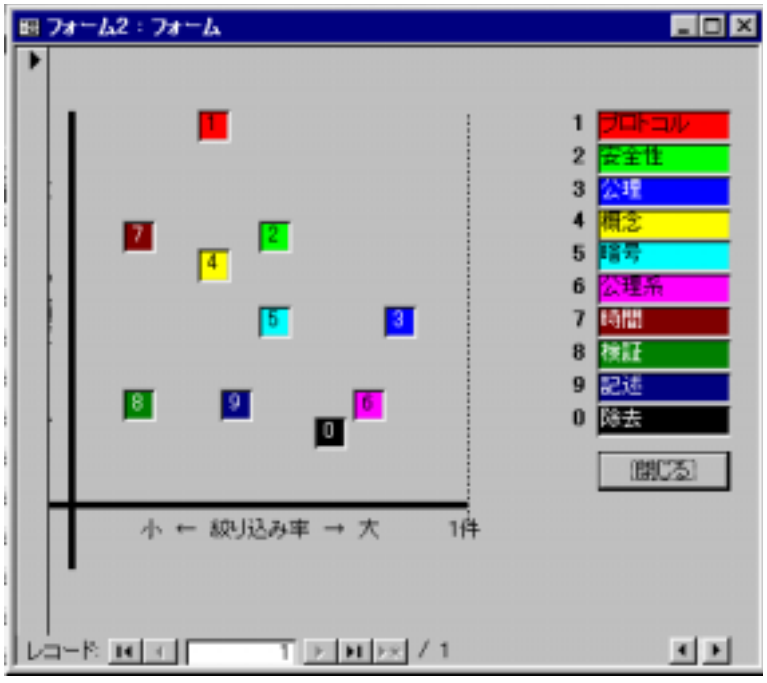


図 23　正解文書における単語情報の可視化結果

索要求を拡大できるような単語もなかったために，類似度の高い文書を個々に確認するして欲しい文書を探す必要が生じた．その結果，最初のキーワード検索で得られた文書集合しか得られなかった．

5.2 考察

以上の実験を元に，本研究で提案手法を用いた可視化による文書検索支援について，考察を行う．

類似度検索の上位には，検索意図に合致したものが多く，一つの文書からの類似度検索は有効であることが分かった．それと同時に，提案した可視化手法を用いることで，類似度の高い文書中に含まれている単語からの内容判断が可能になることが分かった．

検索意図と全く関係ないと思われるような単語も，候補語のなかに出ることがあったが，基本的には，キーワード候補語は比較的文書の内容をよく反映したものが多かった．そのため，検索意図に沿ってないような単語が提示されることもあったが，直観的にはキーワード候補語上位には妥当と思える単語が多く提示されていた．

また，本手法では一つの正解文書を得られたことを前提としたが，正解文書を得るのが難しいときや，キーワード検索で欲しい文書集合が十分得られたときには，提案手法の有効性が言えなかった．

6．むすび

本論文では，大量の文書集合から，文書情報の可視化により検索絞り込みの支援を行う文書検索システムを提案した．提案手法は，ユーザが発見した一つの文書の情報を利用した検索を行い，計算機がその情報を可視化することにより，効率的な検索を実現しようとした点に特徴がある．

この実現のために本論文では，

- キーワード検索と類似検索を用いた手法
- 文書間の関連を可視化する手法
- 文書中の単語情報を可視化する手法

を提案した．

提案手法では，計算機が処理可能な文書中の単語の出現頻度などの情報を可視化することにより，これらの情報をユーザにうまく把握させることにより，検索意図に沿った検索が可能になると考えられる．本論文では，その提案手法について基本方針と実装に関する詳細を述べ，本検索手順を用いた実験と考察を行った．

実験において本手法はキーワード候補語に検索意図に沿った単語が選択されたときには，検索絞り込みの支援が行われ，提案手法の有効性が認められた．

今後は，提案手法をインターフェースとしての評価実験を行い，インターフェースのそれぞれについて検討していく必要があると考えられる．また，キーワード候補語の選択について，統計的な結果や辞書的な意味などを利用して，有用だと思われる単語を提示する必要があると考えられる．そして，それらの結果をもとに改良したシステムを，大規模で様々な文書データベースに適応し，そこから得られた知識を利用して精度のよい検索が行えたかの評価を行い，本研究の実用性に関する評価を行う必要がある．

謝辞

本研究を進めるにあたり，終始暖かい御指導を頂いた ソフトウェア基礎講座 横矢 直和 教授 に厚く御礼申し上げます.

副指導教官として御助言を頂いた 像情報処理講座 湊 小太郎 教授，並びにソフトウェア基礎講座 竹村 治雄 助教授 に深く感謝致します.

本研究を進める上にあたり厳しくも暖かい御指導を頂いたソフトウェア基礎講座 岩佐 英彦 助手に深く感謝致します.

本研究を進める上で貴重な御助言を頂いたソフトウェア基礎講座 山澤 一誠 助手に深く感謝致します.

本研究に関する貴重な助言や御指導を頂いた, 神原 誠之 氏 , 松宮 雅俊 氏, 及び佐藤 哲 氏に深く感謝致します.

日々の研究室活動を支えていただいた, 福永 博美 女史に深く感謝致します.

この 2 年間苦楽を共にし, 互いに励まし合い, 有意義な研究生活を共に過ごすことのできたソフトウェア基礎講座の M2 諸氏に深く感謝致します.

本研究を進めるにあたり，多大なる御協力頂いたソフトウェア基礎講座の諸氏に深く感謝いたします.

参考文献

[1] 林一成, 岩佐秀彦, 竹村治雄，横矢直和: 文書間の相関の可視化による文書検索支援, 情報処理学会第 59 回全国大会, Vol.3, pp.237-238, 1999.

[2] 多変量解析入門 II: 河口至商，森北出版株式会社

[3] 松本裕治, 北内啓, 山下達雄, 平野善隆, 今一修, 今村友明: 日本語形態素解析システム『茶筌』version 2.0 使用説明書, 1999-12, http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html.

[4] 坂倉健太郎, 吉川耕平, 西村英樹，稗田薫: 情報検索システム「Datahunter」におけるビジュアライズ機能, 情報処理学会第 59 回全国大会, Vol.3, pp.99-100, 1999.

[5] 早川和宏, 大久保正且, 田中一男: 検索結果絞り込み用インターフェースの開発, 情報処理学会研究報告, HI76, pp.25-30, 1998-1.

[6] 松田勝志, 福島俊一: 文書タイプ分類による問題解決向き WWW 検索システムの開発と評価, 情報処理学会研究報告,FI-53-2, pp. 9-16, 1999-3.

[7] 山田洋志, 福島俊一: 数値情報を用いたテキスト検索の提案と評価, 情報処理学会研究報告,FI-53-3, pp.17-24, 1999-3.

[8] 高橋裕信, 新田義貴, 岡隆一: 非線形クラスタリングによるパターンの分類, 信学技法, PRMU98-13, pp.1-7, 1998-5.

[9] 舘村純一: DocSpace：文献空間のインタラクティブ視覚化, インタラクティブシステムとソフトウエア IV, 近代科学社, 1996.

[10] 中島浩之, 木谷強: 単語の文書頻度を利用した決定木学習アルゴリズムによる relevance feedback の高精度化, 情報処理学会研究報告,FI-45-2, pp.7-12, 1997-5.

[11] 熊本睦, 島田茂夫, 加藤恒昭: 概念ベースの情報検索への適用－概念ベースを用いた検索の特性評価－, 情報処理学会研究報告,ICS-115-2, pp.9-16, 1999-1.

[12] 福原知宏, 武田英明, 西田豊明: 統計情報と概念知識を用いたテキスト間の話題特定, 情報処理学会研究報告,ICS-115-2, pp.1-8, 1999-1.

[13] 塩澤秀和, 相馬隆宏, 野田純也, 松下温: 切り取り操作による柔軟な情報選択ができるWWW視覚化, 情報処理学会研究報告, 97-HI-72, pp.61-66, 1997-5.

[14] 柿元俊博, 上原祐介: 3 次元情報検索インターフェース, 情報処理学会研究報告,FI-41-6, pp.45-52, 1996-4.

[15] 江口浩二, 伊藤秀隆, 隈元昭: ユーザへの適応性を考慮した WWW 情報検索における漸次的なクエリの拡張, 情報処理学会研究報告, NL121, pp.135-142, 1997-9.

[16] 木谷強, 高木徹, 木原誠, 関根道隆: フルテキストと抽出キーワードを利用した情報検索情報処理学会研究報告, NL115-18, pp. 71-76, 1996-9.

[17] Salton,G. and Allen,J.: Text Retrieval Using the Vector Processing Model, *Proc 3rd Annual Symposium on Document Analysis and Information Retrieval*, 1994.

[18] Salton, G., Singhal, A., Buckley, C. and Mitra, M.: Automatic Text Decomposition Using Text Segments and Text Themes, *Hypertext '96 Proc.*, pp.53-65, 1996.

[19] Salton, G.: *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley, Reading, MA, 1989.

[20] Spoerri,A: Tools for Information Retrieval, *Proceedings IEEE Symposium on Visual Languages*, pp.160-168, 1993.

[21] Card,S.K.: Visualizing Retrieved Information, *A survey, IEEE Computer Graphics and Application*, Vol.16, No.2, pp.63-67, 1996.

[22] James A. Wise, et. al. Visualizing the Non-Visual: Spatial analysis and interaction with information from text documents. *Proc. of IEEE Information Visualization '95*, pp.239-244, 1994.

[23] ++Mail : ソフトフロント社.

[24] 多変量解析入門 II : 河口至商, 森北出版株式会社

[25] 早川和宏, 福永博信, 鈴木達郎: ユーザの利用履歴に基づくWWWサーバの地図型ディレクトリ, 情報処理学会研究報告, HI70, pp.17-24, 1997-1.

[26] 河野浩之, 長谷川利治: WWW データ資源に対する重み付き相関ルール導出アルゴリズムの適用, 重点領域研究「高度データベース」松江ワークショップ, 1, pp.90-99, 1996-9.

[27] 仲川こころ, 高田喜朗, 関浩之: 検索目的を反映したカテゴリ構造に基づく WWW 検索支援, 情報処理学会研究報告, HI82, pp.59-64, 1999-1.

[28] 木谷強: 日本語情報検索システム評価用テストコレクション BMIR-J2, 情報処理学会研究報告, DBS114-3, pp.15-22, 1998-1.

[29] Dreilinger, D. and Howe, A. E.: Experiences with Selecting Search Engines Using Metasearch, *ACM Trans. Information Systems*, Vol.15, No.3, pp.195-222, 1997-7.

[30] Fishkin, K. and Stone, M. C.: Enhanced Dynamic Queries via Movable Filters, *Proc.CHI 95*, pp.415-420, 1995-5.

[31] Golovchinsky, G.: Queries? Links? Is there are difference? , *Proc.CHI 97*, pp.407-414, 1997-3.

[32] 原田昌紀: Freya version 0.92, 1998-6, `http://odin.ingrid.org/freya/`.

[33] Lawrence, S. and Gilesd, C. L.: *Searching the World Wide Web*, Vol.280, No.5360, Issue 3, pp.98-100, 1998-4.

[34] Lokuge, I., Gilbert, S. A. and Richards, W. : Structuring Information With Mental Models: A Tour of Boston, *Proc.CHI 96*, pp.413-419, 1996-4.

[35] Pinkerton, B.: Finding What People Want: Experiences with the WebCrawler, *Electronic Proc. 2nd World Wide Web Conf.*, 1994, `http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/pinkerton/WebCrawler.html`.

[36] Pirolli, P., Shank, P., Hearst, M. and Diehl, C.: Scatter/Gather Browsing Communicates the Topic Structure of a Very Large Text Collection, *Proc. CHI 96*, pp.213-220, April 1996.

[37] Robertson, G. G., Card, S. K. and Mackinlay, J. D.: Information Visualization using 3D Interactive Animation, *Comm.ACM*, Vol.36, No.4, PP.57-71, 1993-4.

[38] Selberg, E. and Etzioni, O.: Multi-Service Search and Comparison Using the MetaCrawler, *Fourth International WWW Conf.*, 1995, `http://www.w3.org/Conferences/WWW4/Papers/169/`.

参照 URL

[39] Digital Equipment Corporation: AltaVista, `http://altavista.digital.com/`.

[40] Excite Inc.: Excite, `http://www.excite.com/`.

[41] Excite Inc.: Excite(日本版), `http://jp.excite.com/`.

[42] Excite Inc.: WebCrawler, `http://www.webcrawler.com/`.

[43] go2net Inc.: MetaCrawler,
http://www.metacrawler.com/.

[44] 原田昌紀: ODIN,
http://odin.ingrid.org/.

[45] Infoseek Corporation: Infoseek,
http://www.infoseek.com/.

[46] Infoseek Corporation: Infoseek Japan,
http://japan.infoseek.com/.

[47] 木原英人: Meta-Search JP,
http://www.shiratori.riec.tohoku.ac.jp/~kihara/metasearch.html.

[48] McKinley Group Inc.: Magellan,
http://www.mckinley.com/.

[49] 富士通 (株): InfoNavigator,
http://infonavi.infoweb.or.jp/.

[50] 東芝 (株): WebSearch,
http://search.softpark.jplaza.com/.

[51] 日本電気 (株): NETPLAZA,
http://netplaza.biglobe.or.jp/keyword.html/.

[52] 日本電気 (株): SolutionWave,
http://www.sw.nec.co.jp/search/.

[53] Netcraft Web Server Surbey,
http://www.netcraft.co.uk/survey/.

[54] Network Wizards Internet Domain Survey,
http://www.nw.com/zone/WWW/.

[55] NTT: NTT DIRECTORY,
http://navi.ntt.co.jp/.

[56] NTT: TITAN,
http://titan.navi.ntt.co.jp/.

[57] NTT アド: Goo,
http://www.goo.ne.jp/.

[58] RCAAU Mo-n-do-u(問答),
http://www.kuamp.kyoto-u.ac.jp/labs/infocom/mondou/.

[59] リクルート (株): ACARA NAVI,
http://www.acara.com/.

[60] Savvy Search Limited.: SavvySearch,
http://www.savvysearch.com/.

[61] 清水奨: 日本の Search Engine のリスト,
http://www.ingrid.org/w3conf-bof/search.html.

[62] 田村建人: Senrigan(千里眼),
http://senrigan.ascii.co.jp/.

[63] Web spawns 1.5 million pages daily according to findings from Alexa Internet,
Alexa Internet Press Release, 1998-8,
http://www.alexa.com/company/inthenews/webfacts.html.

[64] Wired Digital Inc.: HotBot,
http://www.hotbot.com/.

[65] Lycos Inc.: Lycos,
http://www.lycos.com/.

[66] Yahoo! Inc.: Yahoo!,
http://www.yahoo.com/.

[67] Yahoo! Japan Corporation: Yahoo! JAPAN,
http://www.yahoo.co.jp/.

[68] 山名早人: WWW 情報検索サービスの動向,
http://www.etl.go.jp/~yamana/Research/WWW/survey.html.

[69] 吉川耕平，西村英樹，稗田薫，宇都宮速人: ネットワーク上の情報検索とブラウジング, http://narsgw.sharp.co.jp/softcenter/DataHunter/Tech/Ss98/SS98.html.

[70] Justsystem:ConceptBase Search20,
http://www.justsystem.co.jp/product/applicat/cb20/index.html.

[71] WEBSOM: http://www.websom.hut.fi/websom/.