

複合階層索引を用いたデータ検索システムの 利用者インタフェース

河野弘史 古川哲也

九州大学経済学部

本稿ではデータ検索の形態を分類し、複合階層索引を用いたデータ検索システムの利用者インタフェースについて議論する。複合階層索引は、オブジェクト集合を性質ごとに分類する階層を作り、それを合成して、さまざまな階層を提供する。従来のキーワード索引やキーワードパターン照合を用いたデータ検索システムでは、蓄えられたデータに対してある程度の知識が必要である。しかし複合階層索引を用いたデータ検索システムでは、利用者が求めるデータを的確に効率よく指示でき、あらかじめ蓄えられているデータに関する知識を必要としないという特徴がある。

User Interface of an Information Retrieval System with Complex Hierarchical Indices

Hirofumi Kawano and Tetsuya Furukawa

Dept. Economic Eng., Kyushu Univ.

In this paper, we discuss a user interface of an information retrieval system with complex hierarchical indices by classifying styles of information retrievals. Complex hierarchical indices, which consists of hierarchies classifying objects for every characteristics, prepare various hierarchies by composing them. Information retrieval systems by keyword indices or keyword pattern matching require us some knowledge on the stored data. On the other hand, in information retrieval systems with complex hierarchical indices, users can indicate target data precisely and efficiently, and need not have knowledge on the data.

1. はじめに

計算機技術の発展に伴い、システムが大容量化し、またネットワーク技術の発展により情報の流通量も増大している。それによって、様々な分野で大量の情報を蓄え、それを利用するシステムが必要となっている。収集したデータを利用できるようにするためには、必要なデータを取り出せるようにシステムを構築しなければならない。

そのような利用を実現するために、情報検索の分野などで多くの研究が行われている。従来のキーワード索引を用いた情報検索システムに加えて、キーワードパターン照合による情報検索システムも数多く利用されている。また、検索方法に関するものには、データの意味的な同一性から検索するもの^[3]や多次元の空間にデータを配置するものなど^[4]がある。これらは利用者のデータについての知識に対する要求を緩和するもの^[4]であると考えられる。データに対する索引に対する索引の方式についてもシグネチャファイルなどの研究がある^[2]。近年では、インターネットを介した情報検索におけるデータマイニングなどの研究がなされている。データマイニングとは、蓄積されたデータの内容全体を分析し、ある法則性を導き出し、データ検索システムの利用者へ有益な情報を提供する手法である。これはデータを分析することに主体を置いている。

収集したデータを整理し、利用できるようにする方法として、複合階層索引を用いてデータの分類をするものがある^[1]。複合階層索引はデータを複数の性質で分類し整理する。本稿では、複合階層索引を用いて構成されたデータに対する検索システムの利用者インタフェースについて検討する。

データ検索システムの利用を考えたとき、利用者にとって利用しやすく、求めるデータを的確に、また効率よく検索できるシステムを提供するためには、以下のことを考えたサービスを提供する必要がある。

- 検索の目的は何か。
利用者は検索システムに対してどのような検索結果の出力を期待しているのか。
- 目的を実現するために必要な操作は何か。
利用者が期待する検索結果を出力するためには、蓄積されたデータに対して、どのような操作を行えなければならないのか。

- 必要な操作をいかに分かりやすく、利用者に提供するか。

期待する検索結果を出力するために必要な操作をどのような形で提供するのかが。

本稿では、これらの事項を整理することによって、複合階層索引を用いたデータ検索システムの利用者インタフェースがどのようなものであるべきかを検討する。2節では、データ検索システムの利用者がどのような検索を期待しているかについて議論する。3節では、データに対して複合階層索引を構成する方式について述べる。それを用いたデータ検索システムでどのような操作が必要になるかを4節で示す。5節では、そのような操作を実現する利用者インタフェースについて検討する。6節は全体のまとめと今後の研究課題である。

2. データ検索の形態

データ検索とは、後に利用することを想定して蓄積されたデータの中から、ある特定の属性をもつものを選択する行為である。検索の対象となる個々のデータをオブジェクトとする。利用しやすいデータ検索システムを構築するためには、どのような検索結果が得られれば利用者は満足するのか、また、そのような検索結果を得るためにはどのような操作が必要かを考えなければならない。しかし、利用者によって要求はそれぞれ異なるので、どのような目的で利用者が検索を行っているのかを分類して考えることが必要である。本節では、データ検索の形態を分類し、その特徴を明らかにするとともに、データ検索を実現しているシステムとの整合性について考察する。

利用者が検索システムの出力に対して期待していることに注目すると、検索の形態は以下のように分類できる。

(1) 条件検索

検索条件を与え、その条件を満たすデータが検索結果となる検索。他の検索結果と独立に必要なデータかどうかが決まる存在検索と他の検索に依存して必要なデータかどうかが決まる候補検索に分けることができる。

● 存在検索

条件を満たすオブジェクトが存在するかどうかを目的とする検索。例えば、文献検索においてある内容の論文があるかどうかという場合は、検

検索結果の有無や件数のみを問題とする。そのような文献が無くても結果の件数が多くてもそれで終了する。検索結果の件数の多寡は問題ではない。

● 候補検索

情報を得るための候補の検索。条件を満たすオブジェクトが存在するときに、その中からいくつかのオブジェクトを選択する、すなわち、ある条件で検索し、検索結果が多ければ、その結果に対してさらに条件を加えて検索結果を絞り込む。逆に検索結果が少ないときには、条件を弛めて、検索結果を増やす。検索結果は適当な件数でなければならない。また、検索結果の優先順位も重要になる。例えば、ある研究分野のことを知りたいときに、その分野に関する情報を得るために文献を検索する場合などである。候補検索では検索結果には適当な件数が期待されている。

(2) 特定検索

あらかじめ存在が分かっているオブジェクトの情報を得るための検索。例えば、ある文献が出版されていることは分かっているが、どの雑誌あるいは論文誌に掲載されているか分からないときに、その文献に関する情報を得るために検索する。

このようにデータ検索の目的はさまざまであり、それぞれに対して最適な検索システムを構成する必要がある。

データ検索機能を有するシステムにはデータベースシステムや情報検索システムなどがある。これらのシステムで、データ検索のための操作として実現されているものには以下のようなものがある。

[データベースの問い合わせ記述]

SQLなどの問い合わせ言語を用いて、必要なデータを記述する。データベースがデータベース管理システムを用いて構築されているシステムで利用できる。

[キーワード検索]

利用者がキーワードを入力することによって、キーワードに関連するデータを検索する。検索の方法として、索引を用いるものとパターン照合を用いるものがある。

● キーワード索引

オブジェクトにはあらかじめキーワードが定義されており、それがオブジェクトに対する索引となっている。検索結果は入力されたキーワードが

定義されているオブジェクトとなる。

● キーワードパターン照合

データとしてテキストを持つオブジェクトに対し、入力されたキーワードをテキスト中に持つオブジェクトが検索結果となる。一般には、キーワードは正規表現などで表され、テキストにそのパターンを含むものを検索する。

以上のようなデータ検索システムを利用する上で、次のような問題がある。

[データに対する知識]

蓄えられているデータに対して、ある程度の知識がなければ適切な検索ができない。

- SQLなどの問い合わせ言語を用いる場合、データベースのスキーマが分かっていると質問を記述できない。
- キーワード検索では、適切なキーワードを指定できなければ、検索することはできない。

あるデータが検索結果に含まれるべきものであったとしても、利用者が入力したキーワードとそのデータに対して定義されたキーワードが違っていたり、蓄積情報の中にキーワードが含まれていなかったりして、検索されないことがある。

キーワード索引を用いた検索システムでは、キーワードに対するシソーラスが作成されていれば検索の手助けとなる。シソーラスは、索引作成者および検索者の用いる索引語および検索語として統制された語を集め、個々の語の概念間の関係を分かりやすく体系化したものであり、それを用いれば適切なキーワードが選択できる。しかし、データベースが対象としている分野ごとにシソーラスを作るには多大な労力を要する。

また、キーワードパターン照合による検索システムは、テキスト中の任意の文字列がキーワードとなり得るため、シソーラスの作成はほとんど不可能である。

[検索結果の絞り込み]

候補検索や特定検索では、利用者は結果の質を考えないというわけではないが、検索結果の質よりも、どちらかといえばその量に対して注目がちである。検索結果が非常に多ければ、さらに条件を加えて結果を絞り込む必要があるが、どのような条件を加えるかが問題となる。加える条件がなければ、それ以上の絞り込みはできない。

また検索もれが起こらないように、必要なデー

タが検索結果に必ず含まれるように検索の条件を与えると、不必要なデータ、すなわちノイズも増加する。条件の否定を記述できればノイズをかなり減らすことができる。すなわち、必要なデータを含まないような条件があり、その否定を検索の条件とすることができれば、その条件を満たすデータは検索結果に含まれない。

図1で、検索結果 S と必要なデータ I に対し、ノイズ N は $S - I$ である。 $I \cap C = \emptyset$ となるデータ集合 C を規定する条件の否定を検索の条件として記述できれば、検索結果は $S - C$ となる。

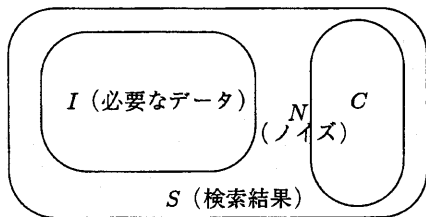


図1 必要なデータとノイズ

[データの種類とシステムの適合性]

どのようなデータを蓄えるかによって、どのシステムを用いるのかを考える必要がある。

- テキストデータは、データベース管理システムの属性値による比較演算には向かない。例えば、関係完備な問い合わせ言語では、属性値に対するキーワードのパターン照合を記述できない。
- キーワードパターン照合を用いる場合、属性という概念が使いにくいいため、属性値に対する比較演算による条件の記述ができない。

キーワードパターン照合は、蓄積されている情報に対して、キーワードによる全文検索を行うもので、データ検索システムの構築が比較的容易である。しかし、利用者の要求を十分に満足するような検索をすることは困難である。

3. 複合階層索引

オブジェクトの集合を分類する。分類によってできたオブジェクトの集合をさらに分類すると、分類の階層ができる。本節では、分類の規準となる複数の性質に対し独立した階層を構成する複合階層索引の概略について述べる。

オブジェクト集合を分類することによって生成された個々の集合をグループとする。グループ g に含まれるオブジェクト集合を $m(g)$ で表す。

[定義1] 階層 H はグループに対応する節点の集合とグループの包含関係を表す枝集合からなる木構造である。木の根グループは、要素を全オブジェクト集合とするグループである。 □

すなわち、階層 H において、グループ g がグループ g' の祖先であれば $g \supset g'$ である。階層の深さは不均一でよく、グループ g の子グループ g, g' に対し $g_1 \cap g_2 = \emptyset$ でなくともよい。

[例1] オブジェクトをある性質に従って分類してできたグループが $\{g_{1,1}, g_{1,2}, g_{1,3}\}$ であったとする。 $g_{1,1}$ はさらに分類され $\{g_{1,4}, g_{1,5}, g_{1,6}\}$ に、 $g_{1,3}$ は $\{g_{1,7}, g_{1,8}\}$ になったとする。この階層 H_1 は根グループを $g_{1,0}$ とすると図2 (a) となる。 □

分類に適用する性質の順序で階層の構造が異なる。そのため性質ごとに分類階層を作り、それを合成することで利用する分類階層を得る。すなわち、階層をオブジェクト集合に対する索引とする。そのような階層の集合を合成して得られた階層索引を複合階層索引という。

本稿では階層の合成には触れず、個々の階層を独立して用いる。したがって、複合階層索引は階層の集合である。

[例2] 図2 (a) の階層中のグループをさらに異なる性質で分類するのではなく、性質ごとに分類の階層を作成する。図2 (b) の H_2 や図2 (c) の H_3 はそのような階層である。複合階層索引は階層の集合 $\{H_1, H_2, H_3\}$ である。 □

4. 複合階層索引におけるデータ操作

複合階層索引を用いて検索システムを構成するために、複合階層索引上での操作について議論する。

データ検索は全オブジェクト集合に条件を与えることによって、必要なデータを得るものとする。個々での条件はデータが属すべきグループを選択することによって与える。各階層について選択されたグループ集合を階層におけるビューとして定義する。

[定義2] 複合階層索引 $IX = \{H_1, H_2, \dots, H_n\}$

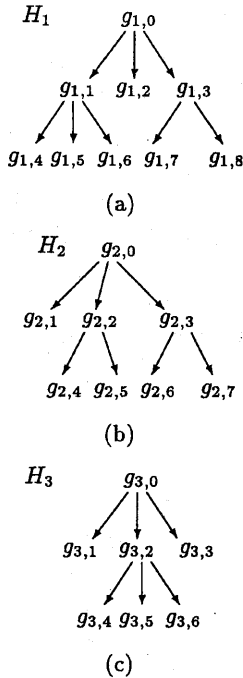


図2 複合階層索引

において、階層 H_i におけるビュー VH_i は H_i 中の包含関係のないグループの集合である。 VH_i 中のグループに含まれるオブジェクトの集合を $m(VH_i) = \bigcup_{g \in VH_i} m(g)$ とする。

すなわち、 VH_i 中の任意のグループ g, g' に対し、 $g \not\subseteq g'$ である。これは H_i 中で g は g' の子孫でないことに対応する。

階層はある1つの性質に注目して、オブジェクトを分類したものである。利用者が自分の求めるオブジェクトを検索する際に、利用する性質は1つの場合もあれば、複数の場合もある。つまり複合階層索引において、利用する階層は1つであるとは限らない。

1つの性質に対する条件が階層におけるビューの選択となり、検索は複数の条件に対して行われるので、検索結果は各階層におけるビューのオブジェクトの共通集合となる。

[定義3] 複合階層索引 $IX = \{H_1, H_2, \dots, H_n\}$ のビューは $VX = \bigcap_{i=1}^n m(VH_i)$ である。

階層のビューはグループの集合であり、複合階層索引のビューはオブジェクト集合である。

複合階層索引を用いたデータ検索は、索引のビューが求めるオブジェクト集合となるように、各階層のビューを変更することである。

VH_i の初期値を H_i の根グループとする。ある性質でデータに条件を加えることは、その性質でさらに分類したもの中から必要なものを選ぶことになる。複合階層索引では、 VH_i 中のグループ g に対し、 g をその子グループのいくつかに置き換えることに対応する。また、条件を厳しくする性質を選ぶことは、階層の集合 $\{H_1, H_2, \dots, H_n\}$ の中から H_i を選択することである。

[例3] 図2の複合階層索引に対し、 H_1, H_2, H_3 のそれぞれの階層のビューの初期値は $VH_1 = \{g_{1,0}\}$, $VH_2 = \{g_{2,0}\}$, $VH_3 = \{g_{3,0}\}$ 、索引のビューの初期値は全オブジェクト集合 $g_{1,0} \cap g_{2,0} \cap g_{3,0} = g_U$ である。

$g_{1,0}$ の子グループ集合 $\{g_{1,1}, g_{1,2}, g_{1,3}\}$ から $g_{1,1}$ と $g_{1,3}$ を選択し、さらに $g_{1,3}$ から $g_{1,8}$ を選択すると、 H_1 のビューは $VH_1 = \{g_{1,1}, g_{1,8}\}$ となる(図3(a))。このとき索引のビューは $(g_{1,1} \cup g_{1,8}) \cap g_{2,0} \cap g_{3,0} = g_{1,1} \cup g_{1,8}$ となる(図3(b))。

次に H_2 についても同様にビューを $VH_2 = \{g_{2,2}, g_{2,7}\}$ にすると、索引のビューは $(g_{1,1} \cup g_{1,8}) \cap (g_{2,2} \cup g_{2,7}) \cap g_{3,0} = (g_{1,1} \cup g_{1,8}) \cap (g_{2,2} \cup g_{2,7})$ となる。

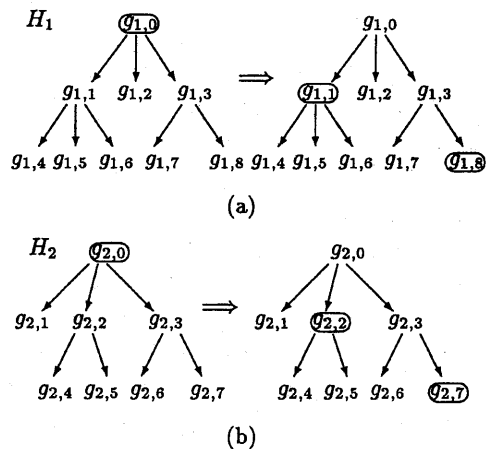


図3 階層のビューの変更

逆に、実際に検索を実行した結果の数によっては、条件を弛めて検索をやり直したい場合もある。これは階層のビューにグループを加える、または

階層ビューに含まれるグループをその親グループに置き換えることに対応する。親グループに置き換える際は、その親グループの子孫のグループは階層のビューから除かれる。

データ検索において、階層索引が作成されているすべての性質を常に利用するわけではない。作成されている階層の数が多いときは、階層の選択時に利用しないものは表示されない方がよい。そのために階層の表示を取りやめる操作を準備する。また検索の結果、利用中の階層では絞り込みができず、表示していない階層の性質を利用することもある。非表示の階層を再び表示させる操作も必要となる。

[例 4] H_3 は利用していないので、表示しないようにする。索引のビューのオブジェクト数が多いとき、適当な数が意味をなす検索では更なる絞り込みが必要となる。そのような場合には非表示とした H_3 を利用することになる。 □

以上のような操作で複合階層索引によるデータ検索を行う。この検索には次の特徴がある。

- データに関する知識を必要としない。
データが階層的に分類されており、その分類に基づいて操作できる。
- 否定と同等の操作ができる。
不必要なオブジェクトからなるグループを階層のビューから削除できる。

実際の利用に際しては、検索によって得られた索引のビューに対して、以下のことができると有用である。

- (1) オブジェクトの数の表示
- (2) オブジェクトの一覧の出力
- (3) 得られたオブジェクトに対する操作

これらは、複合階層索引に対する操作ではないので、本稿では言及しない。特に、得られたオブジェクトに対する操作（パタン照合など）ができれば、利用者にとって、さらに使いやすいシステムになる。しかし、複合階層索引はオブジェクトの階層の集合を用いて必要なオブジェクトを出力することが目的である。複合階層索引がどのようなデータベースに対して作られているか、すなわち得られたオブジェクトに対する操作については複合階層索引の構成やその操作には関係しない。

5. 利用者インタフェース

複合階層索引を用いたデータ検索はデータの絞り込みであり、その操作は階層のビュー VH_i の変更である。本節では絞り込みに必要な操作、すなわち VH_i の変更を行う利用者インタフェースについて議論する。このインタフェースは検索システムの利用者が階層の存在を意識せずに使えるものである。

操作の項目を、階層のビューの変更、利用できる階層の表示からの削除と再表示、階層のビューの表示、索引のビューの出力とする。これをメニューとして準備する（図4）。

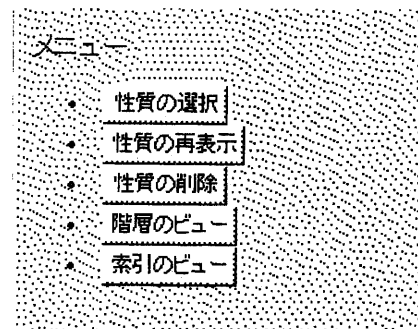


図4 操作のメニュー

階層のビューの変更は、まず変更する階層の選択となる。

[階層の選択]

階層ビューを変更するために、どの階層に対して変更するかを選択する。このための操作は次のものとなる。

- 階層集合の表示
- 表示された階層集合からの階層の選択

階層の選択に対するインタフェースは図5のようになる。図中の表示の性質は階層のことである。表示された階層の一覧から条件を指定する階層を選択し、ビューの変更へ進む。

[ビューの変更]

ある性質に対する条件を加えるために、階層のビューのグループをその子グループで置き換えるには、次の操作が必要となる。

- ビュー VH_i に含まれるグループの表示
- VH_i からのグループの選択
- 選択されたグループの子グループの表示

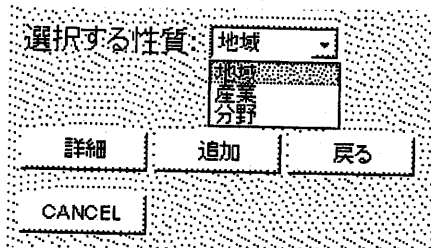


図5 階層の選択

- 表示された子グループ集合からのグループの選択

ある性質で指定した条件を弛めるために、階層ビューのグループを変更するための操作は次のものとなる。

- ビュー VH_i に含まれるグループの表示
- VH_i からのグループの選択
- 選択されたグループのその親グループとの置換、または選択されたグループの兄弟グループの追加。

ビューの変更はグループの選択と、実際の変更の2つの操作からなる。図6 (a) では選択されている階層が表示され、ビューに含まれるグループが選択中の項目として一覧表示されている。グループを選択する際に、それをどのように変更することも選択する。

条件を加える場合は、選択された階層とグループが表示され、そのグループを置き換える候補が表示される。候補のグループ（複数であってもよい）を選択し、それと置き換える。

条件を弛める場合は、選択されたグループを親グループに置き換えるので、その操作の指示だけでよい。兄弟グループを追加するには、グループの選択に戻り、グループを再び選択する。

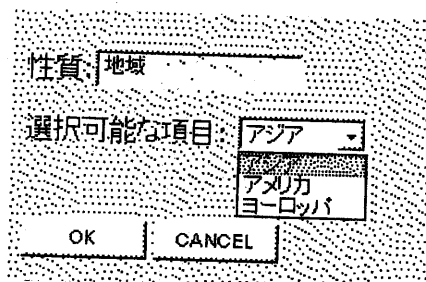
[階層の選択における候補の変更]

階層の選択において候補の削除／再表示を指示する。

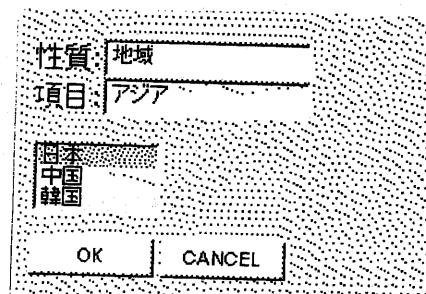
階層の選択の際に利用されない性質を非表示とするために必要となる操作は、次のものとなる。

- 候補となる階層の表示
- 表示された階層からの削除

ある性質に対する条件を加えていった結果、それ以上の条件を加えることができないときに、非



(a)



(b)

図6 ビューの変更

表示とした性質を再び条件として利用する。その際に次の操作が必要となる。

- 階層の選択で表示されない階層の表示
- 表示された階層からの選択

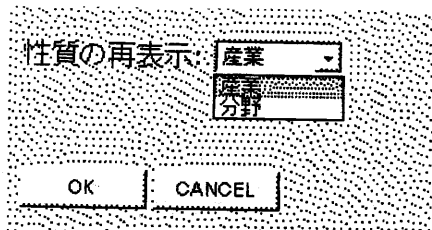
階層の選択における候補の削除／再表示のインタフェースは図7 (a) と図7 (b) のようになる。図7 (a) では利用可能な階層の一覧が表示され、その中から利用しない階層を選択する。図7 (b) では削除された利用されていない階層の一覧が表示され、その中から再利用する階層を選択する。

複合階層索引における操作は以上のようなインタフェースとなる。

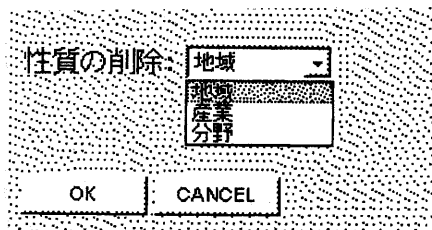
6. おわりに

データ検索の形態を分析し、複合階層索引を用いたデータ検索システムの利用者インタフェースについて検討した。このシステムは、データに対する知識を必要とせず、容易に必要なデータを得ることを可能にする。

本稿で示したインタフェースは複合階層索引の



(a)



(b)

図7 階層の選択候補の変更

柔軟性を活用していない。それを利用すれば、複数の分類に対して異なる階層を適用でき、さらに機能を拡張することができる。

今後は、本稿で検討した利用者インタフェースをWWWを用いて開発するとともに、複合階層索引の能力を活かした機能の拡張について検討する予定である。

謝辞 本研究の一部は文部省科学研究費補助金特定領域研究「高度データベース」の援助を受けている。

参考文献

- [1] 古川哲也, “データの多重分類階層の構成”, 情報処理学会研究報告, 98-DBS-116-35, 平成10年7月.
- [2] Y. Ishikawa, H. Kitagawa, and N. Ohbo, “Evaluation of Signature as Set Access Facilities in OODBs,” *Proc. ACM SIGMOD Int. Conf. on Management of Data*, pp. 247-256, 1993.
- [3] Y. Kiyoki, T. Kitagawa, and T. Tahara, “A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning,” *SIGMOD Record*, Vol. 23-24, pp. 34-41, Dec. 1994.
- [4] K. Parsaye, M. Chignell, S. Khoshafian, and H. Wong, *Intelligent Databases*, John Wiley

& Sons, 1992.

- [5] 渡辺正裕, 石川佳治, 吉川正俊, 植村俊亮, “多次元ベクトル空間の視覚的操作機能機能を有する情報検索”, 情報処理学会研究報告, 96-DBS-109-2, 平成8年7月.