九州大学学術情報リポジトリ
Kyushu University Institutional Repository

# 多次元マトリックス法による電子ジャーナルZoological Scienceの検索と分析

関, 隆宏
九州大学大学評価情報室

安元, 裕司
九州大学大学院システム情報科学府

和多, 太樹
九州大学大学院システム情報科学府

伊藤, 希
筑波大学大学院生命環境科学研究科

他

https://hdl.handle.net/2324/1526187

出版情報：Digital Libraries. 30, pp.72-80, 2006-03. ディジタル図書館編集委員会
バージョン：
権利関係：

# 多次元マトリックス法による電子ジャーナル Zoological Science の検索と分析

関　隆宏 *1　　安元　裕司 *2　　和多　太樹 *2　　伊藤　希 *3　　廣川　佐千男 *4
1 九州大学大学評価情報室　　*2 九州大学大学院システム情報科学府
3 筑波大学大学院生命環境科学研究科　　*4 九州大学情報基盤センター
1,*2,*4 〒 812-8581 福岡県福岡市東区箱崎 6-10-1
3 〒 305-8572 茨城県つくば市天王台 1-1-1
1 tsekiuoc@mbox.nc.kyushu-u.ac.jp　　*2 {y-yasu, t-wada}@i.kyushu-u.ac.jp
3 nozomi@biol.tsukuba.ac.jp　　*4 hirokawa@cc.kyushu-u.ac.jp

## 概要

大量の文書を扱う検索システムでは，分かりやすい検索結果提示と効率的な絞り込みが大きな課題となっている．XML に代表される半構造化文書では構造的部分文書ごとに記述される属性が異なるので，検索結果の文書群を複数の観点で同時に比較しなければならない．多次元マトリックス法は文書を複数の観点から検索し，検索結果をマトリックス状に表示することにより分析する手法である．本研究では，日本動物学会が発行する電子ジャーナル Zoological Science について著者，タイトル，出版年，キーワードなどの複数の観点から多面的に検索・分析できるシステムを開発し，その定性的評価を行った．

キーワード
多次元マトリックス法

# A Retrieval and Analysis of the Electronic Journal "Zoological Science" by Multiple Semantic Matrix Method

Takahiro Seki *1　　Yuji Yasumoto *2　　Taiki Wada *2　　Nozomi Ytow *3　　Sachio Hirokawa *4
1 Office for Information of University Evaluation, Kyushu University
2 Graduate School of Information Science and Electrical Engineering, Kyushu University
3 Graduate School of Life and Environmental Sciences, University of Tsukuba
4 Computing and Communications Center, Kyushu University
1,*2,*4 6-10-1 Hakozaki, Higashi-ku, Fukuoka, Japan, 812-8581
3 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan, 305-8572
1 tsekiuoc@mbox.nc.kyushu-u.ac.jp　　*2 {y-yasu, t-wada}@i.kyushu-u.ac.jp
3 nozomi@biol.tsukuba.ac.jp　　*4 hirokawa@cc.kyushu-u.ac.jp

## Abstract

Comprehensive visualization with focusing enhances search engines for a huge number of documents. Simultaneously multiple faceted display improves narrowing efficiency especially for semistructured document, including XML, with different attributes. The multiple semantic matrix provides a visualisation tool for retrieving and analyzing documents from two viewpoints by displaying clustered documents in matrices. This research implements the multiple semantic matrix method for an electronic journal "Zoological Science" and evaluates the implementation qualitatively.

## Keywords

Multiple semantic matrix method

## 1．はじめに

文書群の増加は，検索場面における従来の意味での「検索」ばかりでなく「分析」の重要性を高めている．すなわち，検索において，数件の文書を探すばかりでなく，あるテーマについての調査・分析も同時に行うことが求められている．そのためには，検索結果に現れる重要な事柄を発見し，検索結果全体を概観できることが望まれる．また，検索結果を見て新たなキーワードを学習し，新たなキーワードや新たな観点での再検索を連続的かつ効率的に行わなければならない．

図書データや事務的文書は構造化文書の形式をとることが多いが，構造化文書の検索においては検索結果を多面的に見ることが求められる．部分構造ごとの検索結果が一覧できると検索効率が上がると考えられる．通常の検索システムでは検索結果をリスティング，ランキング，ディレクトリあるいはクラスタリング形式で表示するが，これらは検索結果を一面的に表示しているにすぎない．半構造化文書の検索例であるXML検索では，構造と内容の両方の観点からの検索方式，あるいはランキング技術との融合について研究されている [3,5]．また，複数の項目から絞り込みを行うために，複数の独立したカテゴリ体系を設定するファセット分類が近年注目されており，たとえば検索結果のナビゲーションに利用されている [1,4]．

多次元マトリックス法は，検索結果をユーザが任意に選ぶ２つの観点でクラスタリングし，その分布を２次元マトリックス表示する手法である．２次元表示の例として，関連性のある検索結果のページ群を近くにまとめて可視化するシステム KartOO [2] があるが，そのまとまりや位置関係の解釈は十分に与えられていない．これに対し，本手法では各クラスタの特徴語を自動抽出するため，検索結果の全体像を視覚的にも意味的にも概観できる．そして，通常の検索エンジンでは検索過程で発見した新たなキーワードを使って絞り込みを行うのに対し，本手法では注目するセルに対する行と列の特徴語群を見るだけで絞り込みを実現する．

すでに筆者らは，本手法を大学の教員データに対して実装し [9]，病院の評判情報における品詞の分析に利用している [10]．本稿では，電子ジャーナルである日本動物学会学会誌

Zoological Science [6] のアブストラクトのページ（全部で 1065 件）にある複数項目を対象とする多次元マトリックス法の実装とそれを用いた検索とその分析を通じた定性的評価について述べる．

## 2．多次元マトリックス法

多次元マトリックス法は検索結果を２つの項目に基づいて２次元マトリックス表示する手法で，ファセット分類の考え方を利用したものである．これを「多次元」と呼んでいるのは，項目を変えながら検索することで多面性を実現しているからである．本手法では検索対象項目と検索結果項目を自由に設定するため，インデックス作成にあたり，各項目についてインデックスを作成する多次元インデックス法と呼ばれる技術を採用する．各文書に対し，通常の検索技術では１つのインデックスを割り当てるのに対し，本システムでは複数のインデックスを割り当てる．

本稿でいうマトリックス表示は，検索結果をユーザが任意に選択した２つの項目からクラスタリングを行い，その分布状況を文書リストあるいは文書数として表示することを指す．表示にあたって検索結果の意味的な構造を認識できるようにするため，クラスタリングの際に各クラスタに属する文書群から特徴語を抽出する．これにより各クラスタの意味内容が与えられ，ユーザは各クラスタの特徴語を見て所望のセルを選択し，その結果を見ることになる．さらに，そのセルに属する数が多い場合，ズーミングによりさらなる絞り込みを行う．

また，これらのことは複数の観点からの分析も可能にする．２つの項目に出現する特徴語の比較を通じて，特徴語が２つの項目に共通して現れるのでそのセルは○○について強い関わりがありそうであるとか，この項目から見るとあるセルはちょっと異質であるといった分析ができる．あるいは，当初予想していなかっ

た特徴語が示されることで新たな発見があるかもしれない．また，クラスタリングで生成されるクラスタ数 (以下，分割数) を変えたり，ズーミングしたりすることによって，あるセル (あるいはクラスタ) をより詳細に分析できる．例えば，特徴語を見ると明らかに異なるカテゴリに属するものが混在しているが，これらの操作を通じて混在していたものが別々のカテゴリに分離できると考えられる．

## 3．多次元マトリックス法の実装

多次元マトリックス法の実装にあたって，ベクトル空間モデルに基づくクラスタリング計算ができるエンジンの存在を仮定する．今回開発したシステム (以下，本システム) は，Perl で記述した CGI プログラムで実装されており，ベクトル空間モデル用のインデックスとして国立情報学研究所で開発された汎用連想計算エンジン（GETA）[8]，クラスタリング計算のライブラリとして CPAN の perl モジュール Algorithm::Cluster を利用している．また，扱うデータが英語であることから学名以外の語についてステミング処理を行う．

データは多次元インデックス法により作成される．今回扱う Zoological Science の場合，各論文からなるファイル集合を用意し，それらのデータについて作成したいインデックスの項目ごとに切り分ける．各論文はすべて同じ構造を持っているが，今回はタイトル，概要，著者，参考文献，発行年，全項目などの複数項目からなる多次元インデックスを作成する．そしてクラスタリング計算が行えるように，項目ごとに各文書に各単語が何回出現するか記した頻度ファイルを作成する (図１).

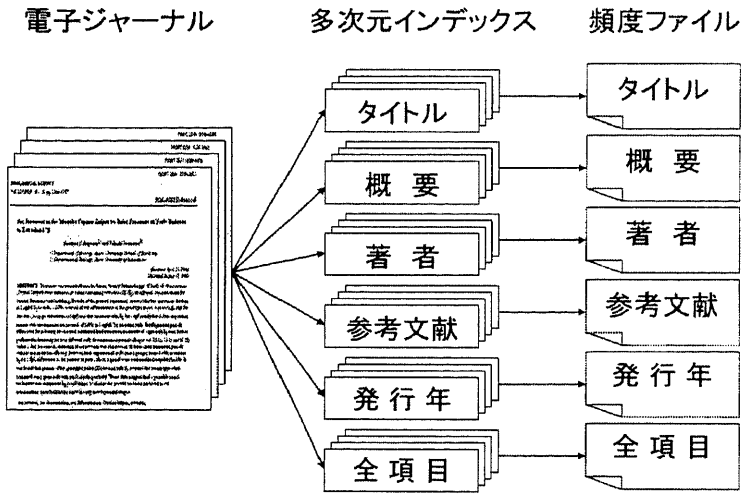


図1　データの作成

本システムは次のように利用される．ユーザは検索要求を入力し，検索対象項目そして得られた検索結果をクラスタリングする２つの項目とそれぞれの分割数を選択する (図２)．そして，これらの情報を送信するとマトリックスが生成される (図３).

マトリックス生成の際，内部では以下の処理を行っている (図４)．まず，検索要求からユーザが選択した項目で検索し，検索結果である文書リストを得る．次に，この文書リストを，ユーザが選択した２つの項目でそれぞれ指定した数にクラスタリングし，あわせて特徴語を抽出する．なお，本システムでは完全リンク法，単一リンク法，群平均法，重心法の４つのクラスタリング計算方式から選択し，特徴語はイエーツの補正公式，カイ２乗値，コサイン，ダイス係数，自己相互情報量，対数尤度比，頻度，補完類似度の８つの抽出方式 [7] から選択した方式に基づいて抽出された 10 語を表示する．ある文書のクラスタリング

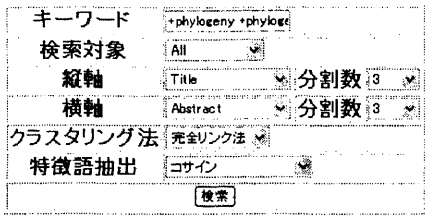


図2　入力画面

| | | | | | |
|---|---|---|---|---|---|
| Title | development(81), speci(76), japanese(67), mitochondrial(49), ascidian(45), genu(40), relationship(38), molecular(38), phylogenetic(34), phylogeny(16) | 177 | 44 | 1 | 222 |
| | gene(83), mitochondrial(49), sequence(40), phylogenetic(34), inferred(22), cytochrome(20), control(19), region(14), rrna(7), crane(1) | 12 | 11 | 0 | 23 |
| | expression(83), medaka(33), oryzias(23), latipes(21), cyclase(8), guanylyl(8), geographic(8), membrane(8), soluble(6), genomic(6) | 5 | 1 | 2 | 8 |
| | 総数 | 194 | 56 | 3 | 253 |
| | | cell(302), speci(292), found(236), gene(207), region(168), sequence(162), female(153), protein(141), population(114), ascidian(48) | result(432), speci(292), gene(207), group(126), population(114), relationship(114), phylogenetic(102), mitochondrial(87), genetic(85), mustela(5) | clade(27), pcr-rflp(4), 1141-bp(2), mitotype(2), cobitis(2), subclade(2), reaction-restriction(2), takatsuensis(1), kanto(1), chugoku-kyushu(1) | 総数 |
| | | Abstract | | | |

図3　出力画面

結果がそれぞれクラスタ i, j に属するとき，その文書はマトリックスの i 行 j 列セルの要素となる．これを検索された文書リストすべてに対して行うと，セル内にそのセルに含まれる文書リストならびに文書数を記した２つのマトリックスが出力される．そして所望のセルをクリックすると，そのセルに含まれる文書リストについて同じ条件でマトリックスを生成するズーミングを行う．

## ４．定性的評価実験

PCR (DNA ポリメラーゼ連鎖反応) の実用化により，特定遺伝子の塩基配列決定が広く行われるようになった．また，そうして得られた配列情報の蓄積と計算速度の向上とが相まって，塩基配列間の類似性から類縁関係を計算機上で推定する分子系統と呼ばれる手法が一般に用いられている．本来この手法は配列情報を手がかりに多様な生物種間の類縁関係 ( 平たく言えば進化) を推定するために用いられていたが，塩基配列そのものについての関係からその配列の機能を推定する手法としても用いられるようになった．分子系統はこのどちらのタイプの論文にもなり得るし，場合によっては２報にもなるが，その結果，系統

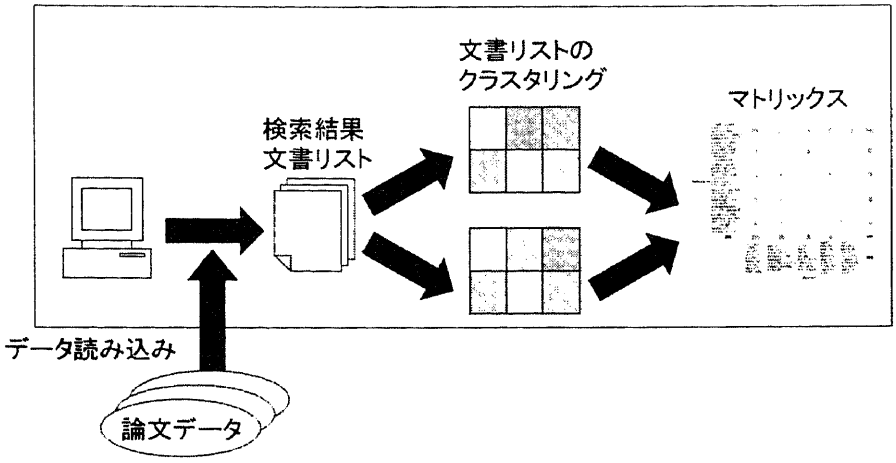


図4　システムの構成

(phylogeny, phylogenetic) という語では論文の主題が生物多様性なのか，生物に共通する機能なのか判断できない状況が生じている．一方，系統関係は本質的に生物の多様性を対象としているため，生物名や遺伝子名を指定してしまうと対象となる論文が著しく制約されてしまう．生物種の類縁関係を調べる際に広く用いられる遺伝子はリボソーム遺伝子やミトコンドリア遺伝子などある程度は決まってはいるが，それらの遺伝子機能の多様性に関する研究も排除できない．そこで，これらの判別を本手法によりどの程度支援できるかについて検討した．

検索語としては phylogeny と phylogenetic が考えられるが，全 1065 論文中 OR 検索で 253 件，AND 検索で 135 件，phylogeny のみで 49 件，phylogenetic のみで 69 件であった．本稿では少なくとも一方を含む論文 253 件を対象とした．このうち，広く用いられる遺伝子との関係ではミトコンドリア遺伝子への言及があるもの 108 件，リボソーム遺伝子への言及があるもの 61 件であった．仮にこれらが生物種の類縁関係を主題とする論文であったとすると，系統学的な論文の 43 % (ミトコンドリア遺伝子のみ) ないし 67 % (ミトコンドリア遺伝子とリボゾーム遺伝子) がこれに相当するが，種間関係の推定に複数遺伝子を用いることはしばしば行われるので 67 %ということはないと考えられる．一方，生物個体群 (population) に言及している論文は 102 件であるので，これもあわせて考えれば概ね半分程度が生物多様性を主題とする論文ではないかと予想された．ここまでは通常の検索表示でも得られる情報である．

本システムで phylogeny OR phylogenetic を検索要求とし，タイトルと概要で 2 × 2 表示 (クラスタリングは完全リンク法，特徴語抽出方法はコサイン) した結果を表 5 (括弧内の数字は小計を表す．以下同様) に示す．2 行 2 列セルは 0 である (以下，このようなセルを「空セル」と呼ぶ) から，2 行および 2 列の解釈は容易である．2 行の特徴語，すなわちタイトルの特徴語に 16 という数字があり，これは 16S リボゾーム RNA 遺伝子であることがわかる．この行の特徴語としてミトコンドリア，チトクロムも挙げられており，これらのことから生物種の系統を主題とする性粒多様性関連の論文であろうと推定された．実際にこれら 29 論文のタイトルを調べると，種群内でのミトコンドリア遺伝子導入という極めてユニークな現象に関する論文 (DOI:10.2108/zsj.21.795)1 報以外は全て生物種の系統に関する論文であった．一方，2 列の特徴語にある clade は系統的に単一の分岐に帰属される生物群を表す語であり，PCR-RFLP という多様性解析に用いられる手法とあわせ，このクラスタも生物多様性に関する論文と推測され，また実際その通りであった．ここまでで，合計 31 文献が生物種の系統に関する論文であることがわかった．上記のユニークな論文は分子機能に関するものでもないので，生物多様性，生物機能に加え「その他」というカ

| | | |
|---|---|---|
| 221 | 3 | (224) |
| 29 | 0 | ( 29) |
| (250) | ( 3) | (253) |

| | |
|---|---|
| 1行 | gene, expression, speci, japanese, ascidian, genu, molecular, relationship, population, phylogenetic |
| 2行 | gene, mitochondrial, sequence, inferred, cytochrome, control, based, region, 16, crane |
| 1列 | result, cel, speci, gene, sequence, group, population, relationship, phylogenetic, genetic |
| 2列 | clade, pcr-rflp, 1141-bp, mitotype, cobitis, subclade, reaction-restriction, takatsuensis, kanto, chugoku-kyushu |

表5　「系統」を検索要求とする2×2表示

テゴリが必要である．

残る1行1列セルの221件を分析しようとすると，従来の手法では他のクラスタを排除する検索条件を考え出す必要があった．しかし，本手法ではそのような手間をかけることなく，分割数の増加による細粒化，221件を含むセル，行，列に対するズーミングという複数手法による分析が可能である．特に，ワンクリックでのズーミングは直観的でわかりやすい．以下，これらの手法を比較していく．

分割数の増加による細粒化は，特に本システムが階層的クラスタリングを採用していることから，細粒化前後の解釈がスムーズに接続できるという特徴がある．しかし，2次元での分割数の変更に伴うクラスタの分割は必ずしも直観的に分かるものではない．1次元での細分はクラスタの分割を意味し，分割数を1増やせばどれか1つのクラスタが2つに分割されるだけなので，分割の追跡と解釈は容易である．ところが，2次元での細分を行うと1つのセルが最大4つのセルに分割される．一般に $m \times n$ 表示を $(m+1) \times (n+1)$ 表示にしたとき，前者のうち $(m+n-1)$ 個のセルが $2(m+n)$ 個のセルに分割されることになるので，この多重分割の解釈は容易でない．また，この分割は既に分析したクラスタについても当然起こり得るので，分析済みクラスタの追跡は必須であるが，これは分割数の増加による細粒化の最大の難点である．

分割数の増加による細粒化について具体的に見ていく．表5で示した2×2表示を3×3表示に細粒化した結果を表6に示す．表6の小計欄を見ると，1行の210と3行の14の和が224であり，これは表5の1行の小計欄の値と一致するので，表5の1行が表6の1行と3行に分割されたと判断できる．同様にして表5の1列が表6の1列と2列に分割されたことが分かる．しかし，このような判断は一般に容易ではなく，特に分割数が多い場合には困難である．さて，表5の1行1列セルの221件が表6で細分されてできた1行1列，1行2列，3行1列，3行2列の4セルについて分析する．3行1列セルの6件はすべて生物多様性に関する論文であったが，3行2列セルの6件のうち3件が生物多様性に関する論文で，残り3件が機能発現に関する論文であった．1行2列の40件について詳細に見ると，生物機能に関する論文が23件，生物多様性に関するものが12件，その他が5件であった．さらに細分化していくことはもちろん可能であるが，それは必ずしも適当ではない．特に，件数の多いクラスタが頑強な場合は，分割数を増やしてもごく少数あるいは空のセルがいたずらに増えるばかりである．一つの目安としては，セルの半分近くが空セルになった時点で分割数の増加以外の方法を考えるべきである．

一般的に本手法でのズーミングは，ユーザが興味の対象を選んでクリックするだけであり，その解釈はユー

| | | | |
|---|---|---|---|
| 169 | 40 | 1 | (210) |
| 20 | 9 | 0 | ( 29) |
| 6 | 6 | 2 | ( 14) |
| (195) | ( 55) | ( 3) | (253) |

表6　「系統」を検索要求とする3×3表示

ザ自身が直観的に行うことになる．ズーミングは一つのセル，一つの行または一つの列のいずれに対しても可能であり，どれが良いかは状況に依存する．ズーミングにおける分割数は以前の値を引き継ぐので，たとえば表5をズーミングする場合は得られる結果もまた2×2表示されることになる．

ではズーミングについて具体的に見ていく．表5の1行1列セルの 221 件，1行の 224 件および1列の 250 件をズーミングした結果をそれぞれ図7 (a), (b), (c) に示す．この例では列に対するズーミング (c) が最も密かつ均質であり有望であるが，分割数を増加させてみると5×5表示程度がマトリックス利用効率上の限界であり，4×4表示でセル群が二つのクラスタに分かれていることが見てとれる．幸いこの場合は一方のクラスタが1列だけでできているので列ズーミングを繰り返してある程度の詳細化は可能であるが，効率がよいとはいえなかった．セルに対するズーミング結果 (a) では，2列の特徴語として酵素名の一部，基質や遺伝子名と思われる語が挙がっていた．タイトルを見るとこれらは全て遺伝子発現 (遺伝子が実際に機能すること) に関する論文であり，カテゴリとしては生物の機能に関する論文である．これら 4 論文のタイトルには expression (発現) という語が共通して現れていた．さらに同じ行にある 27 件について詳細を調べると，カエルの生態に関する 1 報 (DOI:10.2108/zsj.18.605)，ギボシムシ (DOI:10.2108/zsj.18.57) およびナマコ (DOI:10.2108/zsj.19.1113) での形態形成にかかわる遺伝子発現についての論文以外は生物種の系統に関するものであった．なお，このうちの 1 報 (DOI:10.2108/zsj.21.473) はエゾジカの個体群自体に関する論文であり，その他に分類できなくもない．また，上記ギボシムシ及びナマコについての論文は進化的意義に着目して研究材料を選択したものでもある．カテゴライズの際にはこれらの扱いにやや恣意性があり得る．タイトルの特徴語には expression があったが，実際には 27 論文中 2 論文で主題であったのみであり，注意が必要であるし，普通の検索では見落とすかもしれない．ただ，これは先述のように2列にある 4 件の論文のタイトルに expression が現れていたことが影響していると考えられる．

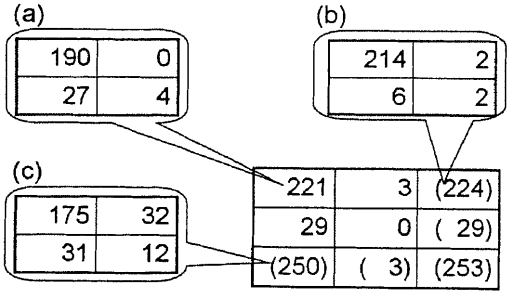


図7　表5のズーミング

そこで，ズーミング結果図7 (a) の1列，すなわち上記 4 件を除いた残り 217 件について再度ズーミングを行うと，2×2表示では全てのセルに文献が割り当てられる．分析を簡単にするため分割数を増やした．

4 分割以上では空セルが増えるのみなので，空セルが全セルの半分以下となるように３×３表示としたが，その結果を図８に示す．１行３列セルの５件のうち１件は先述のエゾジカに関するものであり，他は生物多様性を主題とする論文であった．３行１列セルの３件及び２行２列セルの１件は生物多様性に関するものであり，２行１列セルの４件中１件は系統に関する論文，残り３件は新種記載であり，いずれも生物多様性を主題としていた．適当な分割数は対象によって異なるが，適応的な分割は可能であろう．また，クラスタリング手法についても検討を加えることで，より分析しやすい分割が可能かもしれない．分割とクラスタリングについては更なる検討が必要である．

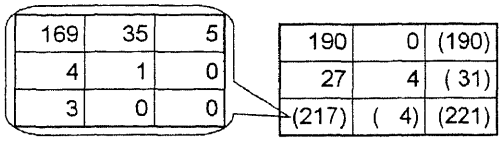


図8　図7(a)の1列のズーミング

一方，分割数増加による細粒化やズーミングを繰り返すよりは，検索式を洗練した方が効率がよい場合もある．特に，大きめの頑強なクラスタがある場合にはその方がよい．検索式の洗練にあたっては排除すべき特徴語を発見するのは一つの手段であるが，実は先述の分割数変更とズーミングを併用するとこのような語を比較的容易に発見できる．例えば，先述の expression などは生物多様性に注目するならば排除してよい単語だと推測されるが，事前に思い付く特徴語ではなく，マトリックスで示されてはじめて「なるほど」と思う類の語である．実際，系統と遺伝子発現の組合せはたとえば「○○という生物は系統上ユニークな位置を占めており，その形態形成メカニズムを知ることは重要である．そこで，本研究では○○の形態形成で発現している遺伝子について調べた」といった内容の概要にしばしば現れていた．こうした弱い関係にある特徴語をクラスタリングによってあぶり出せる本手法は，必ずしも自明ではない隠された特徴語を示唆する有効な手法であると考えられる．

さて，実際に (phylogeny OR phylogenetic) NOT expression で検索すると 194 件の検索結果が得られる．これらについて分割数を変えたり，ズーミングしたりして調べると，全てではないがほとんどが生物多様性にかかわる論文であった．また，expression を含むものについて同様の検索を行うと，生物多様性にかかわる論文は皆無ではないがほとんど含まれなかった．この事実を踏まえると，phylogeny, phylogenetic, expression の OR 検索を行ったときに，マトリックスで対角的分離が可能なように思われる．実際にタイトル，アブストラクト，全項目を対象とする２×２表示を調べてみると，それぞれ表９ (a), (b), (c) のようになり，明らかにタイトルを対象とした検索が対角的であることが分かる．さらに，phylogeny と expression の中間的なクラスタが形成されることを期待して３×３表示も試みたが，いずれにおいても第三対角成分は 0 であった．このことから，分離にはタイトルを対象とする検索が有効であること，phylogeny と expression の組合せでは中間的な論文はほとんど存在しないことが示唆された．

| | |
|---|---|
| 70 | 13 |
| 3 | 31 |

(a) タイトル

| | |
|---|---|
| 159 | 36 |
| 1 | 32 |

(b) アブストラクト

| | |
|---|---|
| 421 | 80 |
| 6 | 0 |

(c) 全項目

表9　検索対象を変えたときの結果の違い

大まかな見積もりとして 253 件中 194 件が生物多様性に関する論文であるとすると，phylogeny または

phylogenetic という語を含む論文の 77%に相当する．これはミトコンドリアやリボソームという語からの見積もりよりも多いが，系統関係は遺伝子以外に形態等によっても推定され得ることから説明可能である．

## 5．まとめと今後の課題

本稿では，電子ジャーナル Zoological Science のアブストラクトのページのタイトル，概要，著者，参考文献，発行年，全項目などからなる複数項目に関して，任意の観点について検索を行い，任意に選んだ2つの観点から結果をマトリックス表示するシステムを開発し，その定性的評価実験として系統 (phylogeny, phylogenetic) という語に着目し，本システムの検索結果に関して専門家による分析を行った．その結果，本システムが分析をある程度支援できるものであることが分かった．

各種クラスタリング方法や適切な特徴語抽出方法に関する検討が評価実験以前から予想していた課題であったが，新たに分析しやすい分割数や分析のための細粒化やズーミングの手法に関する検討も課題であることが分かった．その上で提案手法による検索の効率についての定量的な評価実験を行いたい．また，他の電子ジャーナルや図書館のメタデータを利用した実験も検討している．

## 謝辞

本研究の一部は日本学術振興会平成 17 年度科学研究費補助金 (研究成果公開促進費) 学術誌データベース (課題番号 179013)，日本学術振興会科学研究費補助金基盤研究 (B) (課題番号 17300071) により行われた．

## 参考文献

[1] V.Christophides, D.Plexousakis, M.Scholl, S.Tourtounis. "On Labeling Schemes for the Semantic Web", WWW2003, pp.544-555, 2003

[2] KartOO. http://www.kartoo.com/

[3] L.Guo, F.Shao, C.Botev, J.Shanmugasundaram. "XRANK: Ranked Keyword Search over XML Documents", SIGMOD2003, 2003

[4] Y.Tzitzikas, N.Spyratos, P.Constantopoulos, A.Analyti. "Extended Faceted Taxonomies for Web Catalogs", WISE'2002, 2002

[5] C.Yu, H.Qi, H.V.Jagadish. "Integration of IR into an XML Database", INEX Workshop 2002, 2002

[6] Zoological Science. http://wwwsoc.nii.ac.jp/zsj/zs/zs_index.html

[7] 中條清美，内山将夫，長谷川修治．統計的指標を利用した時事英語資料の特徴語選定に関する研究，英語コーパス研究第 12 号，pp. 19-35, 2005

[8] 汎用連想計算エンジン (GETA). http://geta.ex.nii.ac.jp/

[9] 廣川佐千男，関 隆宏，安元裕司，山田泰寛．教員データに対する多面的検索システム，情報処理学会研究報告 2005-DBS-137，pp.665-672, 2005

[10] 安元裕司，和多太樹，関 隆宏，廣川佐千男．病院評判情報の多面的解析，人工知能学会研究会資料 SIG-KBS-A501，pp.1-4，2005