

Journal of Japan Society of Information and Knowledge

情報知識学会誌

Vol.16 No.1 (Jan. 2006)

~~~~~ 目次 ~~~~~

|      |                                           |                                 |    |
|------|-------------------------------------------|---------------------------------|----|
| 論文   | 不適合情報を利用した情報検索手法の評価                       | 松村 敦, 宇陀 則彦                     | 1  |
| 論文   | 主成分分析と3次元スキャナによる指文字認識                     | 王 宇, 板井 聖治, 小野 智司, 中山 茂         | 15 |
| 論文   | イネのトランスポゾンディスプレイ解析のためのデータベースシステム          | 井上 悦子, 吉廣 卓哉, 川路 英哉, 堀端 章, 中川 優 | 28 |
| 事例報告 | 宇部興産(株)におけるエンドユーザー教育—営業部門に対する知的財産情報教育の試み— | 出口 昌信, 岡本 和彦                    | 39 |
| 事例報告 | 山口 TLO に対する知的財産情報教育支援—産学連携における試み—         | 岡本 和彦, 出口 昌信                    | 44 |
| 事例報告 | 生物遺伝資源総合データベース構築における XML 技術適用について         | 山川 武廣, 山崎 由紀子                   | 52 |
| 論 談  | 我が国における情報学の始まり                            | 藤原 鎮男                           | 60 |
| お知らせ | 情報知識学会第14回(2006年度)研究報告会                   | 田窪 直規                           | 68 |
|      | 特集「情報倫理(情報倫理学・技術倫理/工業倫理等)」の論文募集           | 二階堂 善弘, 岡本 由起子                  | 70 |
|      | 情報知識関係新刊図書一覧                              | 平田 周                            | 71 |
|      | 平成18年度役員選出について                            |                                 | 72 |
|      | 投稿規定と執筆要領                                 |                                 | 74 |



情報知識学会

<http://www.jsik.jp/>

# 不適合情報を利用した情報検索手法の評価

## Evaluation of an Information Retrieval Method using Non-Relevancy Information

松村 敦<sup>\*,†</sup> 宇陀 則彦<sup>\*,‡</sup>

Atsushi MATSUMURA and Norihiko UDA

利用者が検索システムに投入する問合せの背後には複雑な検索要求が存在する。この情報を利用できないことが検索精度の低下を招く原因の1つである。本研究では、自然文で表現された検索要求のうち不適合条件に着目し、これを係受け解析といくつかのルールを用いて構造化し、この構造を検索に反映させる手法を提案する。情報検索システム評価用テストコレクション NTCIR-2 を用いて評価実験を行なったところ、初期検索が単語1, 2個程度の比較的短い検索質問の場合には、本手法の検索精度は Rocchio 型の適合フィードバック手法と同等であることが示された。

Complex information needs are hidden in every query that is used to retrieve information requested by users. The non-availability of these complex information needs decreases the accuracy of information retrieval systems. Our proposed retrieval method uses non-relevancy information to clarify users' complex information needs. In our method, the non-relevancy information, which is represented by natural language sentences, is first structured using dependency analysis and heuristics. Next, the structured non-relevancy information is translated into a structured query and then used to score and rank the documents retrieved. Retrieval experiments using the NTCIR-2 test collection for information retrieval systems have shown that our method is equivalent to Rocchio's automatic relevance feedback method, when the initial query consists of only a few words.

キーワード：情報検索, 不適合情報, 係受け解析, NTCIR

Information Retrieval, Non-Relevancy Information, Dependency Analysis, NTCIR

### 1 はじめに

利用者が検索システムに投入する問合せの裏側には、その表現からは推測しきれない複雑な検索要求がある。これらの情報を利用できないことが、検索の精度を落す要因の1つである。しかし、一般的な利用者は検索システムに対して数語程度のキーワード入力しか行なわない<sup>[1]</sup>。利用者がこのような行動をとる背景には、複雑な検索要求を容易に入力するインタフェースが存在しない、複雑な検索要求を十分に検索結果に反映させる手法がな

い、といった問題がある。いずれにしても、情報検索システムには数語程度のキーワード入力から精度の高い検索結果を出せるだけではなく、より複雑な検索要求を理解しそれを検索結果に反映させることができるような仕組みが求められる。

従来から問合せの情報不足を補う手法としては、シソーラスや疑似適合フィードバックを利用した質問拡張が提案され盛んに研究されている<sup>[2]</sup>。しかしながらこれらの手法は、あらかじめ与えられた共通の知識あるいは検索された文書集合によって自動的に情報を補っているため、個々の利用者にとって本当に最適な手法であるとは言い難い。

このような背景から、本研究では利用者の持つ複雑な検索要求のうち、特に不適合条件に着目し、これを利用する検索手法を開発す

\* 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

† matsumur@slis.tsukuba.ac.jp

‡ uda@slis.tsukuba.ac.jp

る。これによって、個々の利用者により適切な検索結果を提供することを目的とする。

ここで、本研究では利用者が複雑な検索要求を持つ（あるいは持つようになる）場合を想定するが、この複雑な検索要求の具体例として、情報検索システム評価用テストコレクション NTCIR-1<sup>[3]</sup>および NTCIR-2<sup>[4]</sup>の検索課題を利用した。

以下、2節では NTCIR テストコレクションの検索課題に現れる検索要求の構造の分析を行なった結果を述べる。これをもとに3節ではその中から抽出した「不適合条件」の構造化とこれを利用した検索手法について示す。4節では評価のための検索実験の結果とその考察を述べ、5節で結論と今後の課題について述べる。

## 2 検索要求の構造

情報検索システム評価用のテストコレクションは、文献データ集合、検索課題集合、各検索課題に対する文献の適合不適合判定からなるもので、同一のテストコレクションを利用することにより共通の基準で情報検索システムを評価することができるようにしたものである。日本での代表的なテストコレクションとして国立情報学研究所の提供する NTCIR がある。

本研究では、これらのテストコレクションから NTCIR-1 および NTCIR-2 を選び、そこに含まれる検索課題集合を利用者の検索要求の一例として取り上げ利用する。

図1に NTCIR-2 の検索課題の一例を示す。ここで示されているように、検索課題には TITLE, DESCRIPTION, NARRATIVE というフィールドがある。これらは検索要求をそれぞれ1語程度、1文程度、複数文の詳しい文章、という形式で検索要求を持つ本人によって記述されたものである。本研究では、複数文の詳しい文章で記述された NARRATIVE の情報を分析対象とする。

これまでの NTCIR を利用した研究を見ると、TITLE や DESCRIPTION だけでなく

```

<TOPIC q=0101>
<TITLE>
B型肝炎
</TITLE>
<DESCRIPTION>
遺伝子工学的手法によるB型肝炎ワクチンの開発について論じている文献
</DESCRIPTION>
<NARRATIVE>
肝炎などのウイルス性疾患に対する安全かつ有効な予防法の確立は21世紀に向けての医療分野での重要な課題である。そのため、遺伝子工学的手法によるB型肝炎ワクチンの開発について論じていれば検索要求を満たす。開発されたB型肝炎ワクチンの物理化学的特性を論じているものやその免疫力増強に有効な免疫アジュバントについて論じているものも検索要求を満たす。しかし、遺伝子工学的手法に触れていない論文は不可。また、B型肝炎以外のワクチンも不可。
</NARRATIVE>
<CONCEPT>
a. B型肝炎,
b. 遺伝子工学的手法,
c. ワクチン, 予防接種
</CONCEPT>
<FIELD>
7. 医学・歯学
</FIELD>
</TOPIC>

```

図1 NTCIR-2 の検索課題の例

NARRATIVE も利用した方が検索精度が高くなるという結果が出ている。これは、主に関連するキーワードが増えることにより検索要求がより明確になるということによるものである。しかしながら、NARRATIVE に書かれているような複雑で詳しい説明文を検索システムに入力するよう求めることは利用者にとって大きな負担となる。したがって、NARRATIVE の中の重要な要素のみを利用して利用者の負担を押えつつ、検索精度の向上を狙う必要がある。このような観点から NTCIR-1 の検索課題 83 件と NTCIR-2 の検索課題 49 件の合計 132 件を対象に分析を行なった。

分析の結果、NARRATIVEは主に「背景」「適合条件」「不適合条件」の3つの内容に分けることができることが分かった<sup>1</sup>。「背景」は検索要求が生じた理由や、その分野の歴史などである。「適合条件」は、適合であるための追加の情報や詳しい条件である。「不適合条件」は適合でないものに対する条件を提示しているものである。

一般に、「背景」や「適合条件」は検索要求の曖昧性を解消し検索結果を絞り込む効果と、逆に、関連する情報によって検索の範囲を広げる効果の両方を持つ。例えば、図1の検索要求では、「背景」として「ウイルス性疾患に対する安全かつ有効な予防法の確立」とあるが、これはB型肝炎よりも広い概念によって、検索される文書の範囲を広げる働きがある。一方で、「適合条件」の「免疫アジュバントについて論じているもの」という記述は、より限定的な表現で検索結果を絞り込む働きを持つ。これに対して、「不適合条件」の場合は主に検索結果を絞り込む効果を持つ。これは否定表現を含むことによる大きな特徴である。例えば、同じ図1の検索要求では、不適合条件に「遺伝子工学的手法に触れていない論文は不可」とあるが、これはB型肝炎に関する文献のうち、遺伝子工学的手法に触れているものに絞り込む効果がある。

そこで、本研究では利用者の検索要求の特徴の1つとして不適合条件に着目し、これを係受け解析と不適合の内容を表す特徴語を利用して構造化し検索に利用する方策を検討した。解析対象の不適合条件文は、文末表現を手がかりとして検索課題のNARRATIVEから人手で抽出した。これらの総数は110文(NTCIR-1に68文、NTCIR-2に42文)である。

### 3 不適合条件を利用した検索手法

利用者の入力した不適合条件を検索に利用するためには、不適合を表す内容(不適合内

<sup>1</sup> NTCIR-4からは、同様の分類をタグで指定し、役割を明示しているが、適合情報と不適合情報の区別はない<sup>[5]</sup>。

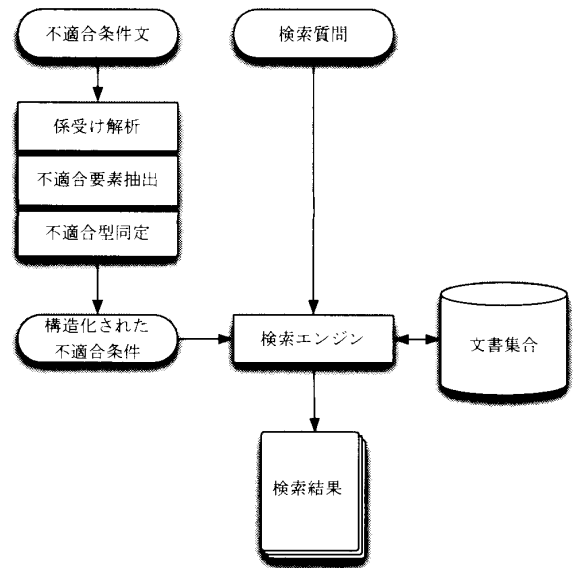


図2 不適合条件を利用した検索手法概要

容)を正確に表現し、これを検索に反映させる必要がある。本手法では、NTCIRの検索課題のうちTITLEやDESCRIPTIONのようなキーワードや自然言語で表現された短い検索質問と、NARRATIVEに含まれる自然言語で書かれた不適合条件文を利用する。手法の概要を図2に示した。不適合条件文は係受け解析といくつかのルールを用いて構造化する。この構造化された不適合条件と検索質問とを同時に利用することによって不適合条件を考慮した検索手法を実現する。以下では、最初に不適合条件を構造化するプロセスについて説明し、次に、これらを用いた文書の得点付けとランキングについて説明する。

#### 3.1 不適合条件の構造化

はじめに、不適合条件文の係受け関係を得るために、係受け解析を行なう。本研究では、係受け解析器としてCaboCha<sup>[6]</sup>を利用した。(なお、CaboChaは形態素解析部分にChaSen<sup>[7]</sup>を利用している。)係受け解析の結果から、不適合条件文を「は不可とする。」や「は適合しない。」のような不適合文末表現と不適合の条件を表す不適合内容(名詞句)とに分割する。不適合文末表現はNTCIR-1およ

表 1 不適合表現の分類

| 不適合型              | 不適合条件文の例                                                           |
|-------------------|--------------------------------------------------------------------|
| 肯定                | インスリン注入型人工膵移植 [は不可.]                                               |
| 否定                | 特異点の必然性について言及していないもの [は不可.]                                        |
| 以外 + 肯定<br>のみ     | B 型肝炎以外のワクチン [も不可.]<br>宇宙定数の測定方法や測定計画についてのみ述べたもの<br>[は, 要求を満たさない.] |
| 以外 + 肯定 + 以外 + のみ | 小細胞癌以外の組織型や肺以外の部位に関してのみ論じているもの [は不可.]                              |

び NTCIR-2 の検索質問に含まれる 110 文の不適合条件文を分析して収集した。これを正規表現による不適合文末規則として作成し利用した。

次に、「は不可とする.」のような文末表現を削除して残った不適合内容を以下で定義する 4 種類の不適合要素に分解した。

否定 否定表現になっている部分

以外 「以外」が含まれる部分

のみ 「のみ」が含まれる部分

肯定 上記 3 種類以外で肯定表現の部分

各不適合内容は上記 4 種類の不適合要素のいずれか、またはその組合せによって表現される。この組合せを不適合型と定義する。

例えば、「B 型肝炎以外のワクチンも不可.」の場合、最初に、「も不可.」という不適合表現を削除し、残った「B 型肝炎以外のワクチン」を不適合内容とする。次に、不適合要素を特徴付ける「以外」という語があるため、「B 型肝炎以外」を「以外」要素、「ワクチン」を「肯定」要素として抽出する。その結果、この不適合条件文は「以外」+「肯定」型と決定する。

各型に分類された不適合条件文の例を表 1 に示した。表中の例では、不適合文末表現は括弧で括り、分類の指標として利用した特徴語をゴシック体で示している。これらの型は検索要求と不適合内容との意味関係を反映したものである。3.2 節に述べるように、検索における文書の得点付けのためのパラメータは、これらの意味関係を反映するように設定されることになる。

最後に、不適合要素を特定する「ない」、「以外」、「のみ」等の特徴語に係る単語の集合を係受け解析の結果を用いて同定する。その結果、不適合要素を [不適合要素名] { 単語集合 } で表現し、これらの組合せで不適合条件を表現する。例えば、「B 型肝炎以外のワクチンも不可.」は以下のように表現される。

[以外]{B, 型, 肝炎}[肯定]{ワクチン}

### 3.2 文書の得点付けとランキング

文書の検索と得点付けは、検索質問と不適合条件の 2 つを利用して行なう。まず、検索質問を形態素解析システム ChaSen<sup>[7]</sup>で解析し、キーワードを切り出す。次に切り出したキーワードを利用して、BM25<sup>[8]</sup>により文書  $D$  の得点  $SI_D$  を計算する。BM25 は、確率型モデルを用いた検索システム Okapi で採用されている文書の得点計算式である。現時点で高精度な検索結果を得ることができると実証されており、多くのシステムで採用されている。

一方、前節で表現した不適合条件を検索に利用する手順は以下の通りである。

1. 検索質問  $Q$  で検索した文書に対して、不適合条件  $NR$  を構成する各不適合要素  $E$  に含まれる単語による文書得点 (不適合要素得点)  $SN_{D,E}$  を求める。
2.  $SI_D$  と不適合得点を不適合表現の型に応じて組み合わせて総得点  $S_D$  を求める。
3. 総得点で文書をランキングし出力する。

手順 1 の計算は検索質問に対する計算方式

と同様に BM25 を利用する。また、手順 2 における 2 種類の得点の組合せは、式 (1) で示す線形結合の式を利用する。

$$S_D = SI_D + \sum_{E \in NR} \alpha_E \times SN_{D,E} \quad (1)$$

ここで、 $\alpha_E$  は不適合要素  $E$  に応じて得点を調整するためのパラメタであり、以下の式 (2) のように定義する。

$$\alpha_E = \begin{cases} -0.5 & (\text{肯定/のみ要素の場合}) \\ +0.5 & (\text{否定/以外要素の場合}) \end{cases} \quad (2)$$

基本的に、不適合内容が肯定されるか否定されるかを考慮し、肯定的な要素に対しては負の重みを、逆に否定的な要素に対しては正の重みを与えている。要素の組合せや列挙があった場合にはこれらを加算することで複数の要素の効果を計算する。ただし、不適合型が「否定(以外)+肯定(のみ)」の場合だけ例外的に、どちらも  $\alpha_E = +0.5$  とする。これらの値は絶対値を 0.5 に固定して予備実験を行ない符号を決定した。

また、 $SN_{D,E}$  の計算の際には、検索質問  $Q$  に含まれる単語を省いて計算する。これらの単語は  $SI_D$  の計算で既に考慮されているため、再度、得点要素として加算することはその語の持つ重みを過剰に利用することになり、適切な文書得点を与えることができなくなるからである。さらに、不適合要素  $E$  に含まれる単語のうち、「研究」「論文」「もの」といった一般的な語は得点付けの計算から省いた。このような一般的な語はノイズを多く拾い、不適合条件の焦点をぼかしてしまうからである。

#### 4 検索実験と評価

本手法による不適合条件の利用の効果を検証するために NTCIR-2 を利用して検索実験を行なった。NTCIR-2 の検索対象文書は約 73 万件の学術文書 (学会発表データベース著者抄録と研究費成果報告書) であり、検索課題は 49 件である。この中から今回は不適合条件文を含む 30 件の検索課題を利用して評価を行

なった。適合判定は、高適合 (S)、適合 (A)、部分的適合 (B)、不適合 (C) の 4 段階判定で行なわれているが今回の実験では S と A を正解とした。初期検索は短い検索質問を想定するため、検索課題の TITLE のみを用いた場合と DESCRIPTION のみを用いた場合について行なった。なお、不適合条件文を含む 30 件の検索課題の TITLE および DESCRIPTION の平均単語数は 3.2 個と 7.3 個であった。

基本となる検索システムには情報検索パッケージ<sup>[9]</sup>を利用した。このパッケージは BM25 を採用しており、NTCIR-2 テストコレクションを利用したワークショップに参加したシステムのうち上位のシステムと同等の検索精度を持つ。したがって baseline としては十分なシステムと考えて良い。また、同パッケージに含まれる Rocchio 型の疑似適合フィードバック<sup>[10]</sup>(以下、Rocchio 型)を利用した検索も行ない比較の対象とした。

評価指標には、各検索課題毎の平均精度 (Average Precision, AP) と平均精度の平均 (Mean Average Precision, MAP) を利用した。検索課題  $i$  に対する平均精度  $AP_i$  は以下の式 (3) によって求められる。

$$AP_i = \frac{1}{11} \sum_{i=0}^{10} P_i \quad (3)$$

ここで、 $P_i$  ( $i = 0, 1, \dots, 10$ ) は再現率が 0 から 1 までの 0.1 刻みの値 (0, 0.1, ..., 1.0) での精度であり、 $AP_i$  は再現率全域に渡っての精度を考慮に入れた指標となっている。また、平均精度の平均は、全ての検索課題に対して平均精度を平均したものであり、以下の式 (4) によって求められる。

$$MAP = \frac{1}{qnum} \sum_i AP_i \quad (4)$$

ただし、

$qnum$  は検索課題の数 (今の場合は 30) である。

$AP_i$  は検索課題  $i$  の平均精度である。

表2 平均精度の平均 (TITLE のみの場合)

| 手法              | MAP    | 向上率   |
|-----------------|--------|-------|
| baseline (BM25) | 0.1661 | —     |
| Rocchio 型       | 0.2035 | 22.5% |
| 本手法             | 0.2031 | 22.3% |

#### 4.1 実験結果

はじめに、TITLEのみを利用した場合の3つの手法の平均精度の平均を表2に示す。この結果から分かるように、baselineに比べてRocchio型は22.5%、本手法は22.3%の精度向上を達成している。わずかにRocchio型の方が精度が高いがほぼ同等であるといえる。

しかしながら、検索課題毎の平均精度はRocchio型と本手法では大きく異なっている。図3に問合せ毎の平均精度のbaselineとの差をプロットしたものを示す。横軸は本実験における検索課題の番号(1~30)で、縦軸は検索課題毎の各手法の平均精度とbaselineの平均精度との差である。グラフが上に伸びているほど精度は向上しており、逆に下に伸びているほど精度は低下していることを示している。この図から、非常に多くの問合せで両手法の精度が大きく異なっていることが分かる。具体的には半数以上の16件の問合せで両手法の精度は0.1以上の開きがある。これは、Rocchio型と本手法が別々の効果を持っていることを示している。

一方、初期検索にDESCRIPTIONを用いた場合についてのMAPを表3に示す。この場合、Rocchio型は5.2%の精度向上を達成しているのに対して、本手法はbaselineよりも精度が落ちている。検索課題毎にbaselineからの差を示したグラフ(図4)を見ても分かるように、この場合にはRocchio型の手法の精度向上が大きい。DESCRIPTION程度(平均7単語程度)の長さの検索質問を使う場合には、本手法に比べてRocchio型が有効であるといえる。

表3 平均精度の平均 (DESCRIPTION のみの場合)

| 手法              | MAP    | 向上率   |
|-----------------|--------|-------|
| baseline (BM25) | 0.3156 | —     |
| Rocchio 型       | 0.3321 | 5.2%  |
| 本手法             | 0.3118 | -1.2% |

#### 4.2 考察

ここでは検索課題を個別に考察する。表4、表5にそれぞれTITLEを利用した場合、DESCRIPTIONを利用した場合の本手法の効果を検索課題毎にまとめた。一番左の列qidの欄に検索課題番号、次のRoの欄にはRocchio型の平均精度のbaselineからの向上の様子、OMの欄には本手法の平均精度のbaselineからの向上の様子を記号で表した。記号~はbaselineとの差が0.1未満の場合、↑(↓)は精度の向上(低下)が0.1以上0.2未満の場合、↑(↓)は精度の向上(低下)が0.2以上の場合を示している。

##### 4.2.1 TITLEを利用した場合の効果(表4)

**不適合条件の適切な利用** 本手法の精度向上に大きく影響している要因の1つは、不適合条件から否定要素、以外要素の重要な単語を抽出し、質問拡張を適切に行なえていることである。0.2以上の精度向上があった検索課題0107, 0113, 0123, 0.1以上の精度向上があった検索課題0101, 0104, 0105が該当する。例えば、検索課題0105では、TITLEが「新規キノロン剤」であるのに対して、不適合条件は「ウサギ、ラット以外の動物を用いたものは不可。」である。この場合は本手法によって「ウサギ」、「ラット」を含む文書の得点を上げることに成功し、精度も向上している。

これに対して、検索課題0122は逆の効果で精度を上げている。この場合はTITLEは「リテラシーと教育」であるが、不適合条件は肯定要素の「情報リテラシー、メディアリテラシー、コンピュタリテラシーなどは含まない。」となっている。本手法は、TITLEの「リテラシー」という単語で検索されてしまうこ

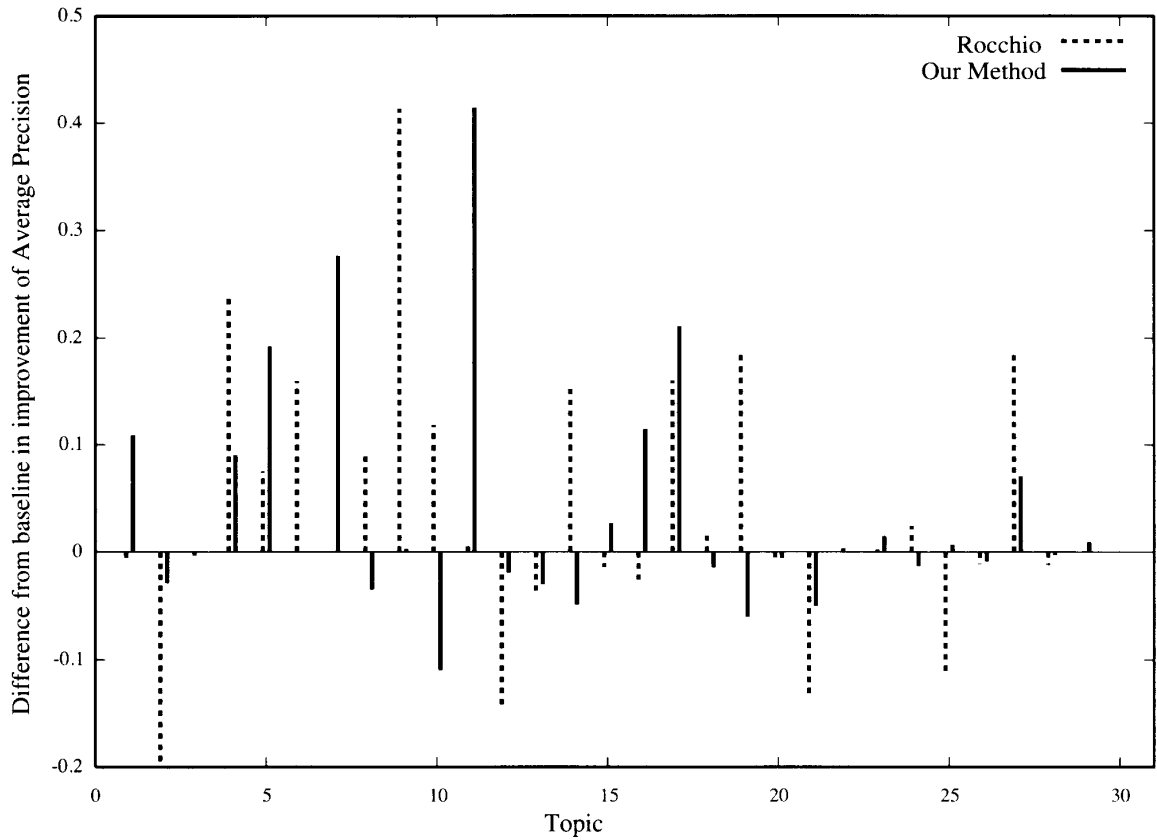


図3 検索課題毎の平均精度の baseline との差 (TITLE のみの場合)

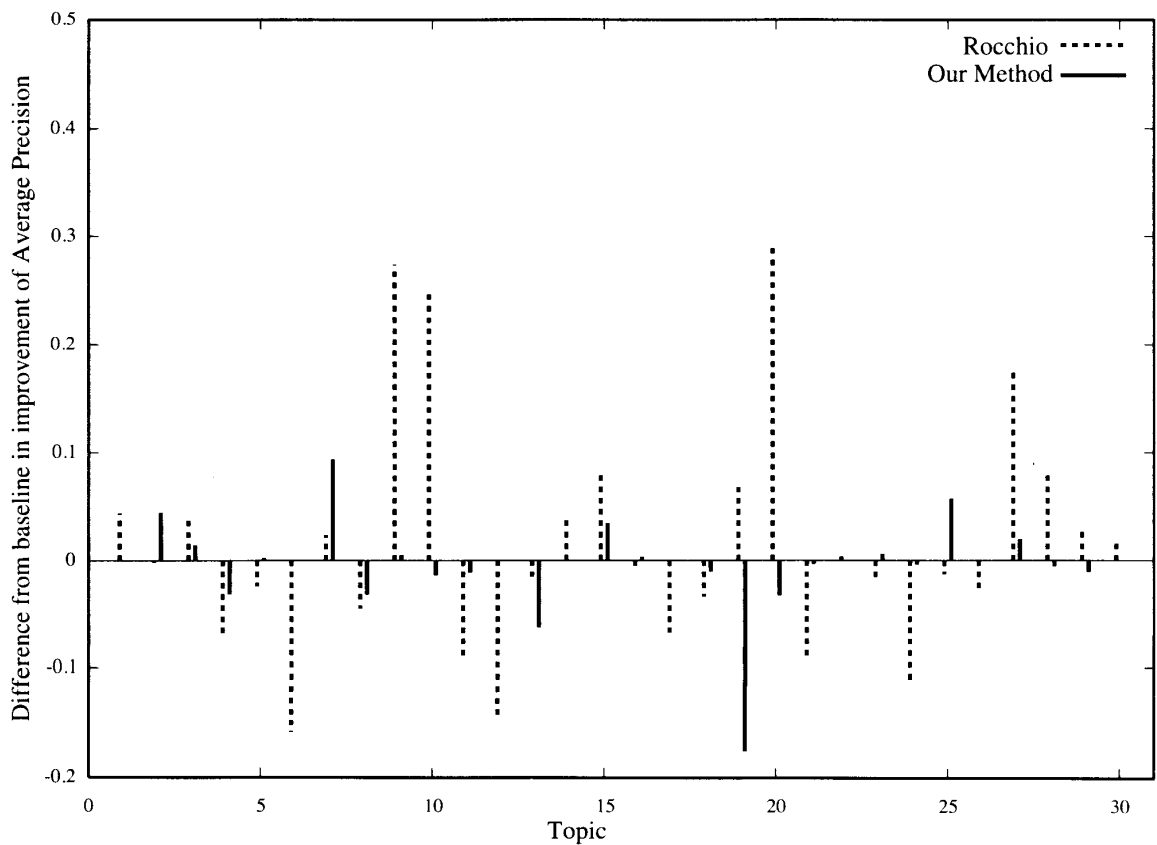


図4 検索課題毎の平均精度の baseline との差 (DESCRIPTION の場合)



表 4 検索課題毎の本手法の効果 (TITLE の場合)

| qid  | Ro | OM | 本手法の効果                                                                                         |
|------|----|----|------------------------------------------------------------------------------------------------|
| 0101 | ～  | ↑  | 「遺伝子工学的手法」は有効に機能する。一方、「ワクチン」は独立に扱われるため精度低下を招く。                                                 |
| 0102 | ↓  | ～  | 「インスリン」は独立に扱われるため精度低下を招く。                                                                      |
| 0103 | ～  | ～  | 「国内」の判断ができないため精度低下を招く。                                                                         |
| 0104 | ↑  | ↑  | 「モノクローナル抗体」は有効に機能する。                                                                           |
| 0105 | ～  | ↑  | 「ウサギ」, 「ラット」は有効に機能する。                                                                          |
| 0106 | ↑  | ～  | インデックスで数字を省略しているため「 $\beta$ 」と「 $\beta 3$ 」の区別がなく, 不適合条件を正しく利用できないことにより精度が低下する。                |
| 0107 | ～  | ↑  | 「リピオドール」, 「肝臓癌」は有効に機能する。                                                                       |
| 0109 | ～  | ～  | 「速度」, 「限界」が独立に扱われることで精度が低下する。また, 現状の得点付式では「のみ要素」を正確に活用できない。                                    |
| 0111 | ↑  | ～  | 係受け解析の失敗で本来は否定要素である「交通システム」が肯定要素となり, 精度が低下する。                                                  |
| 0112 | ↑  | ↓  | 「レーザ」が独立に扱われることで精度が低下する。                                                                       |
| 0113 | ～  | ↑  | 「レーザー干渉計」は有効に機能する。「重力波」の効果もあるが意味を正しくとらえていない。                                                   |
| 0114 | ↓  | ～  | TITLE の言い換えで効果なし。                                                                              |
| 0116 | ～  | ～  | 接尾辞「性」により精度低下。また, 「意識」「調査」を独立に扱うため精度が低下する。                                                     |
| 0117 | ↑  | ～  | 不適合条件の「インターネット」と正解文書中の「インターネット」は役割が違っていても単語として同等に扱ってしまうため精度が低下する。                              |
| 0119 | ～  | ～  | 「調査に基づいているかどうか」の判断ができないため効果なし。                                                                 |
| 0122 | ～  | ↑  | 「メディアリテラシー」「コンピュータリテラシー」といった同表記の関連語に関する文書を的確に排除する。ただし, 「計算機リテラシー」は排除できない。                      |
| 0123 | ↑  | ↑  | 「感染症」が有効に機能する。「形成」も有効だが, 正しい意味をとらえていない。                                                        |
| 0124 | ～  | ～  | 「MRSA」, 「感染」, 「防止」, 「対策」, 「マニュアル」を独立に扱うため効果が相殺される。                                             |
| 0125 | ↑  | ～  | 「次亜塩素水」や「消毒剤」が適合文書中で背景として使われていることを考慮できないため精度が低下する。                                             |
| 0127 | ～  | ～  | 「発症」, 「機序」を独立に扱っているため精度が低下する。                                                                  |
| 0129 | ↓  | ～  | 「超」や「的」のような接頭・接尾辞が独立に扱われ精度が低下する。                                                               |
| 0130 | ～  | ～  | 「～の解析性」と「～を使った解析」が区別できないため効果が得られない。                                                            |
| 0132 | ～  | ～  | 「物理的対象物を明示」しているかどうかを区別できないため効果が得られない。                                                          |
| 0136 | ～  | ～  | 「脳」, 「硬塞」を独立に扱うので精度が低下する。                                                                      |
| 0137 | ↓  | ～  | 「この技術を方法としてのみ」の意味を適切に扱えないため, 「技術」や「方法」といった曖昧な言葉では結果を絞り込めずに効果が得られない。                            |
| 0141 | ～  | ～  | 「半」, 「経験」, 「的」, 「分子」, 「軌道」, 「法」のように複合名詞の構成名詞を独立に扱うため精度が低下する。                                   |
| 0143 | ↑  | ～  | 係受け解析の失敗で「情報検索」が否定要素ではなく肯定要素となったため精度が低下する。                                                     |
| 0147 | ～  | ～  | 「電磁波」を「応用」しているかどうか判定できないため, 精度が低下する。「シミュレーション」が精度に影響する可能性があるが, 「のみ要素」として正確に活用できていないため効果が得られない。 |
| 0148 | ～  | ～  | 「プレストレストコンクリート」は有効だが, 効果はほとんどない。                                                               |
| 0149 | ～  | ～  | 不適合条件の情報は検索された文書にほとんど含まれないため効果はない。                                                             |

これらの概念を含む文書を的確に排除することに成功している。

ただし、検索課題 0113 の「重力波」は精度向上に寄与しているが、不適合条件で記述されている「重力波の検出を目標としていない～」という文の意味を正確に反映しているわけではなく、単語の出現のみで文書をとらえている。したがって、さらに精度向上を目指す場合には、後に述べるようなより深い言語処理を行なう必要がある。検索課題 0123 の「形成」も同様な特徴を持つ。

**単語を独立して扱う影響** 一方、精度が低下する場合、および効果が得られない場合の原因の多くは、不適合要素を単語集合で捉え、検索の際に独立に利用していることにある。検索課題の 0102, 0109, 0112, 0116, 0124, 0127, 0129, 0136, 0141 が該当する。これらは、複数の単語を組み合わせて不適合要素を形成しているが、一部の単語の効果だけが突出するような場合に大きく精度を落してしまう。例えば、検索課題 0112 は TITLE が「放電誘導」であり、不適合条件が「放電現象のみを扱ったものや、レーザによってストリーマを誘導する実験についての論文は要求を満たさない。」である。肯定要素の「レーザ」は「ストリーマ」と「誘導」を伴って適切な不適合条件となるため、「レーザ」を独立に不適合条件として利用する本手法の枠組では効果は得られない。この問題を解決するには、不適合要素も係受け構造を維持したまま利用するような方策が必要となってくる。このように不適合条件という限定的な表現の効果を得るには、言語をより厳密に扱う必要がある。

**自然言語処理の問題** 自然言語処理の設定やエラーから起こる問題もいくつか見られた。検索課題の 0106, 0111, 0143 が該当する。検索課題 0106 の場合に問題となるのは数字の扱いである。本システムの検索では数字を全て省くため「β3」も「β」もすべて「β」と処理される。この課題の TITLE は「β3 アドレナリン受容体遺伝子変異」で、不適合条件

は「β アドレナリン受容体としか表記していないものは不可。」である。にもかかわらず、どちらも「β」として扱われるため、「β」と「β3」を区別するはずの不適合条件は全く効果がなくなってしまう。

また、検索課題 0111, 0143 は係受け解析に失敗している。例えば、検索課題 0111 の不適合条件は「画像と関係のない ITS 関連のものや、交通システムとは関係のない動画認識に関するものは要求を満たさない。」である。本来、「交通システムとは」は「関係のない」に係り、否定要素となるべきだが、「満たさない」に係ってしまい肯定要素として機能している。そのため、検索結果に不適合要素の意図とは全く逆の影響を与えてしまっている。

このような自然言語処理のエラーやシステムの設定の不整合を改善することは比較的簡単に行なえる。システムの設定をチェックして、より精度の良い検索を実現することが可能である。

**深い自然言語処理の必要性** 一方、いくつかの検索課題で検索精度を上げるには、より深い自然言語処理を必要とする。例えば、検索課題 0103 の場合、不適合条件「日本国内での調査のみ論じているものは不可。」の「国内」を判断することができず、検索精度を大きく落している。調査の場所が外国であるということ判断するには辞書等の知識をうまく活用する必要があるが簡単ではない。

また、単語レベルで同等でも係受け関係が異なることで大きく意味が変わる場合もある。例えば、検索課題 0130 の場合、TITLE の「S 行列の解析性」と不適合条件にある「S 行列を使った解析」を区別しなければならない。

さらに、単語に対して文書中での役割を考慮しなければならない場合もある。例えば、検索課題 0125 の場合には、TITLE が「電解水」、不適合条件が「機能性水と表記されている中で、次亜塩素水や消毒剤の水溶液など電解による生成水でないものは除く。」である。不適合条件に含まれる「次亜塩素水」や「消毒剤」は、不適合を意味するキーワードであって

も、正解文書内で背景情報を述べるために使われていることが多い。検索課題 0117 の「インターネット」という単語も同様の特徴を持つが、このような文書内での役割を同定できないために検索精度の低下が起こっている。

このような問題よりもさらに難しい問題に、文書中でどのように現れるかが自明でない不適合条件がある。例えば、検索課題 0113 の「重力波の検出を目標とする」、0119 の「調査に基づかない」、0132 の「物理的対象物を明示」、0147 の「電磁波を応用した」が該当する。他にも検索課題 0137 の「この技術を方法としてのみ使用した」は指示代名詞を含み、より複雑な意味を表している。また、「のみ」という表現を的確に活用する枠組ができていない。キーワードによって文書得点を下げるだけでは「のみ」という条件は表現できないため、なんらかの定式化が必要となる。このような表現を適切に理解し文書の中から該当するものを取り出すのは容易ではない。これらの問題を解決するには、少なくとも検索質問だけでなく文書についてもより深い自然言語処理が必要となる。松村らは検索質問および検索対象文書に対して、係受け解析を行ない、キーワード間の関係を用いて検索を行なう手法の評価を行なっている<sup>[11]</sup>。このような手法を参考に今後より深い自然言語処理の適用可能性を探る必要がある。

**不適合条件が TITLE の言い換えの場合** 不適合条件が TITLE の単なる言い換えの場合には得られる情報はないため、基本的に精度向上の効果は得られない。これには検索課題 0114 が該当する。このような場合には、本手法は無力であるが、適合フィードバックに適切に切替える判断は難しくない。そのようにすれば、検索精度を落すことなく高精度な検索システムを実現できる可能性がある。

#### 4.2.2 DESCRIPTION を利用した場合の効果 (表 5)

**不適合条件の適切な利用** DESCRIPTION を利用した場合も、TITLE と同様の効果が

見られる。不適合条件の適切な利用による精度向上は、検索課題 0102, 0107, 0122, 0143 で見られた。ただし、初期検索で DESCRIPTION を利用することによって、不適合条件で効果的に利用できる単語の数が減るため、適切な利用によって精度が向上する検索課題数も減少した。

#### DESCRIPTION との重複語を省く影響

これとは逆に、単語を独立して扱う影響による精度低下が多くなった。これは、DESCRIPTION と重複する語を不適合条件から省くという処理とも関連している。すなわち、不適合条件と DESCRIPTION との重なりが増えるため、これらを省いたあとに実質的に意味のある単語や単語の集合が残らない場合が多くなるのである。例えば、検索課題 0141 では DESCRIPTION に「非経験的分子軌道法」とあり、不適合条件に「半経験的分子軌道法」とある。これらの重複を省くことで、不適合条件は「半」のみになってしまう。このような中途半端に不適合条件を考慮することになる検索課題は、他にも 0105, 0112, 0113, 0125, 0129, 0132, 0136, 0141 と多くある。

#### 自然言語処理に関する問題

単純な自然言語処理の問題は TITLE の場合と同様に残っている (0106, 0111)。一方、意味を考慮するような深い自然言語処理の必要性も同様に残っている (0103, 0109, 0113, 0117, 0119, 0130, 0132, 0137, 0147)。新しく、検索課題 0104 では文書中の否定表現の扱いが必要であることが示されている。さらに、キーワードの集合で不適合内容を表現することの限界を示した例もある。検索課題 0117 の DESCRIPTION は「歴史史料を電子化し、データベースとしてインターネット上で利用できるようにしたものはないか。」であり、一方、不適合条件は「インターネット上で利用できる史料を使って行なった歴史研究などは含めない。」である。両者はキーワード集合で見ればほぼ同等であり、本手法の効果はほとんど無いが、実際には文の意味は全く異なる。このような場合に

表5 検索課題毎の本手法の効果 (DESCRIPTION の場合)

| qid  | Ro | OM | 本手法の効果                                                                                          |
|------|----|----|-------------------------------------------------------------------------------------------------|
| 0101 | ～  | ～  | DESCRIPTION の言い換えで効果なし。                                                                         |
| 0102 | ～  | ～  | 「インスリン」による効果はあるが、一部の適合文書の排除も行ってしまう。                                                             |
| 0103 | ～  | ～  | 「国内」の判断ができないため精度低下を招く。                                                                          |
| 0104 | ～  | ～  | 文書に含まれる否定表現「肺非小細胞」を正しく扱えないために精度が低下する。                                                           |
| 0105 | ～  | ～  | 「ウサギ」, 「ラット」を利用していないという条件を文書の適合判定に利用できないため効果が得られない。                                             |
| 0106 | ↓  | ～  | インデックスで数字を省略しているため「β」と「β3」の区別がなく、不適合条件を正しく利用できないことにより精度が低下する。                                   |
| 0107 | ～  | ～  | 「リピオドール」が有効に機能する。                                                                               |
| 0109 | ～  | ～  | 「速度」, 「限界」が独立に扱われることで精度が低下する。また、現状の得点付式では「のみ要素」を正確に活用できない。                                      |
| 0111 | ↑  | ～  | 係受け解析の失敗で本来は否定要素である「交通システム」が肯定要素となり、精度が低下する。                                                    |
| 0112 | ↑  | ～  | 「実験」のような語のみが利用され、精度が低下する。                                                                       |
| 0113 | ～  | ～  | DESCRIPTION との重複を省くため、不適合条件が「技術」と「目標」となり精度の低下を招く。また、「～を目標としていない」を正しく活用する枠組になっていないことも精度低下の原因である。 |
| 0114 | ↓  | ～  | DESCRIPTION の言い換えで効果は得られない。                                                                     |
| 0116 | ～  | ～  | 「意識」, 「調査」が独立に扱われるため精度が低下する。                                                                    |
| 0117 | ～  | ～  | 不適合要素の単語集合と文書の単語集合が同一でも、係受けが異なるため精度が低下する。                                                       |
| 0119 | ～  | ～  | 「調査に基づいているかどうか」の判断ができないため効果が得られない。                                                              |
| 0122 | ～  | ～  | 「メディアリテラシー」, 「コンピュータリテラシー」といった同表記の異義語に関する文書を的確に排除するが、効果は小さい。                                    |
| 0123 | ～  | ～  | DESCRIPTION とほぼ同等であるため効果なし。                                                                     |
| 0124 | ～  | ～  | 「感染」, 「防止」, 「対策」, 「マニュアル」が独立に扱われるため精度が下がる。                                                      |
| 0125 | ～  | ↓  | 「電解」, 「生成」, 「水」が独立に扱われるため精度が下がる。特に「生成」のような語の影響が大きい。                                             |
| 0127 | ↑  | ～  | 「発症」, 「機序」を独立に扱っているため精度低下。「発症調査」「抑制機序」等が検索されてしまう。                                               |
| 0129 | ～  | ～  | 「方法」や「測定」といった語が独立に扱われて精度が下がる。                                                                   |
| 0130 | ～  | ～  | 「～の解析性」と「～を使った解析」のような係受けの違いを考慮に入れていないため精度が低下する。                                                 |
| 0132 | ～  | ～  | 「性質」や「性」などが独立して扱われるため精度低下が起こる。また、「物理的対象物を明示」しているかどうかを区別できないため効果が得られない。                          |
| 0136 | ↓  | ～  | 「疾患」や接尾辞「性」を独立に扱うことが精度低下の原因。                                                                    |
| 0137 | ～  | ～  | 「この技術を方法としてのみ」の意味を適切に扱えず、「技術」や「方法」といった曖昧な言葉では結果を絞り込めないため効果が得られない。                               |
| 0141 | ～  | ～  | DESCRIPTION の単語との重なりを省くことで、中途半端な不適合条件の考慮になって、精度の低下を招いている。例えば、「半経験的分子軌道法」では「半」だけが不適合条件となっている。    |
| 0143 | ↑  | ～  | 「インタフェース」は有効に機能する一方で、DESCRIPTION との重複を省いた「一般」や、接尾辞の「的」や「用」が独立して扱われ精度の低下が起こる。                    |
| 0147 | ～  | ～  | 「報告」, 「製作」, 「素材」が独立して扱われるため精度の低下が起こる。また、「シミュレーション」が精度に影響するが、「のみ要素」として正確に活用できていないため効果が得られない。     |
| 0148 | ～  | ～  | DESCRIPTION とほとんど単語が重なるため効果が得られない。                                                              |
| 0149 | ～  | ～  | 不適合条件の情報は検索された文書にほとんど含まれないため効果が得られない。                                                           |

は、より詳細に不適合条件の内容を獲得する必要があり、やはり、不適合内容をキーワード集合ではなく、キーワード間の関係も考慮に入れたもの、例えば係受け関係とする方法などを検討することが考えられる。

なお、不適合条件が DESCRIPTION の言い換えであるものも増えている (0101, 0114, 0123, 0148)。

**情報量増加の影響** 以上のように、TITLE を利用した場合に比べて、DESCRIPTION を利用した場合には検索精度の向上が見られなくなる。その原因の1つは、個々の検索結果の分析から明らかなように、不適合条件で利用できる単語としての情報量が相対的に減少するということである。確かに検索課題や不適合条件、文書を単語の集合で扱う場合には、情報量の大小が検索の精度に大きく影響してくる。しかしながら、不適合条件で扱う情報は本質的に検索質問とは違った内容を持っており、同じ単語の集合でもその意味は大きく異なっている場合があることも今回の分析で明らかになってきている。このような情報をどのように利用すれば効率良く検索精度を上げることができるかということも、個々の検索結果の分析から明らかになってきている。

その方策は、1つにはより厳密な自然言語処理を行なうことであり、もう1つは得点付け手法の改良を行なうことである。より厳密に不適合要素及び文書を扱うには、キーワード集合ではなく、キーワード間の関係も考慮に入れたもの、例えば係受け関係とする方法などを検討することが考えられる。このような手法を利用してより厳密に意味を捉える試みを行なう必要がある。一方、得点付けの方は、現在は不適合要素が複数ある場合には単純に加減で処理しているが、不適合要素の粒度も不均一であるため、より精密な得点付けが必要である。Dkakiらは、適用先はパッセージ検索ではあるが、本研究と同様に複雑な検索要求の構造化を試みている<sup>[12]</sup>。得点付け手法の改良にはこの手法を参考にできると考えられる。

#### 4.2.3 利用者の利便性

本手法の特徴は利用者自身が不適合条件を自然言語で入力できるところにある。このことはインタフェースの面と検索精度の両面において貢献できる可能性がある。

**インタフェースとしての利便性** 本手法に比較的類似した手法に、ブール型検索の not 演算子による検索がある。これは不適合な単語を指定し検索する手法であるが、利用者が not 演算子を利用して検索を行なうことは難しいのが現状である。その理由として考えられるのは、利用者の検索意図に含まれる不適合条件を、not 演算子を利用して論理式として組み立てることが難しいということである。これに対して、本手法を用いれば、利用者は自分の言葉で不適合条件を自然に表現することができるようになる。本手法は、Rocchio 型のように自動で再検索を行なう手法よりは入力の手間がかかるのは確かであるが、論理式を組み立てる難しさに比べれば、はるかに利用しやすいインタフェースを提供することが可能となる。

**検索精度に関する有効性** 4.2.1 節、4.2.2 節で述べてきたように、検索精度に関してはまだ改善の余地がある。しかし、例えば検索課題 0122 のようにいくつかの同表記の異義語がある場合には、それらを不適合条件として入力すれば、関連する不必要な文書を検索結果からの確に排除することが可能となる。このことは、利用者が自分の言葉で不適合条件を入力することが、適切な単語による再検索に結び付くことを示している。Rocchio 型のように検索結果文書から自動的に単語を抽出して再検索する手法では、それぞれの人に最適なフィードバックが行なえとは限らない。

また、本手法は、不適合条件の係受け解析を行ない、複雑な利用者の意図をできる限り正確に把握して利用することを目指している。Rocchio 型のように単語の頻度情報を利用するだけの手法では、利用者の意図を厳密に理解することはできないため、検索精度にも限

界がある。本手法を発展させ、利用者の意図を詳細に捉え、検索に反映させることがより正確に行なえるようになれば、高精度な検索結果をもたらす手法を実現することが可能となる。

## 5 おわりに

利用者の検索要求の1つとして不適合条件に着目し、これを構造化し検索へ反映させる手法を提案した。NTCIR-2 テストコレクションを利用して評価実験を行なった結果、初期検索を TITLE のみで行なった場合の検索精度は Rocchio 型適合フィードバック手法と同等であった。これは、初期検索に入力する情報が非常に少ない状態では不適合情報の利用が検索精度の向上に十分寄与することを示している。

一方で、DESCRIPTION のように7語程度の長さの検索質問を与える場合には、BM25 による重み付けと Rocchio 型適合フィードバック手法は適切に機能するが、不適合情報を利用する本手法は十分な精度をあげられていない。その大きな原因の1つには、DESCRIPTION を利用すると、不適合情報を利用する本手法との情報量の差がなくなることがある。このことは今回の分析からも明らかになった。しかし、その一方で、より深い自然言語処理を行ない、検索質問や文書、不適合情報を厳密に扱うことによって、本手法が効果を発揮する可能性があることも明らかとなった。

今後は、結果の詳細な分析を行なうとともに、得点付け手法の改良や適合情報、背景情報といった検索要求も利用する枠組の検討が課題である。また、本手法と適合フィードバックが異なる効果を示しているが、2つの手法の使い分けの指針は明らかになっていない。これらを明らかにすることも今後の課題となる。今回は、NTCIR の NARRATIVE を利用者の検索要求として扱ったが、現実の検索場面で利用者がこのような形式で要求を持つかどうかは明らかではない。実際に利用者が自分の言葉で不適合情報を入力でき、検索結果

を緻密に調整できるところが本手法の利点である。そのような特色を活かすためにも、今後は、インタラクティブな検索も視野に入れて利用者の検索要求についての調査も行なう必要がある。

## 謝辞

本研究では、国立情報学研究所の提供する情報検索システム評価用テストコレクション NTCIR-1 (本格版) および NTCIR-2 (本格版) を利用した。

## 参考文献

- [1] Jansen, B. J.; Spink, A.; Bateman, J.; Saracevic, T.: "Real life information retrieval: A study of user queries on the web", SIGIR Forum, 32(1), pp.5-17, 1998.
- [2] Sakai, T.; Koyama, M.; Kumano, A.; Manabe, T.: "Toshiba BRIDJE at NTCIR-4 CLIR: Monolingual/Bilingual IR and Flexible Feedback", Proceedings of the Fourth NTCIR Workshop on Research in Information Access Technologies, 2005. <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings4/CLIR/NTCIR4-CLIR-SakaiT.pdf> (2005年12月5日参照)
- [3] Kando, N.; Nozue, T.: editors: NTCIR Workshop 1: Proceedings of the First NTCIR Workshop on Research in Japanese Text Retrieval and Term Recognition, 1999. <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings/index.html> (2005年12月5日参照)
- [4] Eguchi, K.; Kando, N.; Adachi, J.: editors: NTCIR Workshop 2: Proceedings of the Second NTCIR Workshop on Research in Chinese & Japanese Text Retrieval and Text Summarization, 2001, <http://research.nii.ac.jp/ntcir/>

- workshop/OnlineProceedings2/index.html (2005年12月5日参照)
- [5] Kishida, K.; Chen, Hua, K.; Lee, S.; Kuriyama, K.; Kando, N.; Chen, H.-H.; Myaeng, S. H.; Eguchi, K.: “Overview of CLIR Task at the Fourth NTCIR Workshop”, In Kando, N.; Ishikawa, H.; editors: Proceedings of the Fourth NTCIR Workshop on Research in Information Access Technologies, 2005, <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings4/CLIR/NTCIR4-CLIR-TaskOverview.pdf> (2005年12月5日参照)
- [6] 工藤拓; 松本裕治: 「チャンキングの段階適用による日本語係り受け解析」, 情報処理学会論文誌, Vol.43, No.6, pp.1834-1842, 2002.
- [7] 松本裕治; 北内啓; 山下達雄; 平野善隆; 松田寛; 高岡一馬; 浅原正幸: 形態素解析システム『茶釜』version 2.2.9 使用説明書, 2002, <http://chasen.aist-nara.ac.jp/chasen/doc/chasen-2.2.9-j.pdf> (2005年12月5日参照)
- [8] Robertson, S. E.; Walker, S.: “Okapi/Keenbow at TREC 8”, In Proceedings of TREC 8, pp.151-162, 2000.
- [9] 内山将夫; 井佐原均: 「情報検索パッケージの実装」, 情報処理学会研究報告, 2001-FI-63, Vol.2001, No.74, pp.57-64, 2001.
- [10] Baeza-Yates, R.; Ribeiro-Neto, B.: Modern Information Retrieval, Addison-Wesley, 1999.
- [11] 松村敦; 高須淳宏; 安達淳: 「情報検索における単語間の関係の効果」, 情報処理学会研究報告, 2001-DBS-125(1), Vol.2001, No.70, pp.257-264, 2001.
- [12] Dkaki, T.; Mothe, J.: “Combining Positive and Negative Query Feedback in Passage Retrieval”, In Proceedings of RIAO2004, pp.661-672, 2004.

(2005年10月4日受付)

(2005年12月15日採択)

## 主成分分析と3次元スキャナによる指文字認識

Human Recognition with Ear Image  
by Principal Component Analysis王宇<sup>\*,†</sup> 板井聖治<sup>\*,‡</sup> 小野智司<sup>\*,§</sup> 中山茂<sup>\*,¶</sup>Yu WANG Seiji ITAI  
Satoshi ONO and Shigeru NAKAYAMA

3次元スキャナを用いて日本語五十音指文字を認識する方式を提案する。データグローブなどのデバイスの装着を必要としないジェスチャや指文字の認識方式は快適で負担が少ないものの、カメラから得られる2次元画像のみでは、類似した形状の指文字の識別が困難である。本稿で提案する方式は、精度と速度を両立した認識を行うために、3次元スキャナから得られる距離画像に対して主成分分析を行う。10名の被験者による評価実験を行い、提案する方式が、従来の方式と同程度の精度を保ちつつ、認識速度を大幅に改善し、10ミリ秒程度で指文字を認識できることを確認した。

This paper proposes a method for Japanese finger character recognition using 3-dimensional (3D) scanner. Although methods for gesture or finger character recognition without wearable devices, like data gloves, are convenient and stressless, Japanese finger characters involve characters which have similar shapes and cannot be distinguished by them, because most of them recognizes the characters by 2-dimensional images. The proposed method uses distance images obtained by using 3D scanner and principal component analysis in order to recognize the input character accurately and efficiently. Experimental results with 10 testees have showed that the proposed method recognized finger characters within about 10 milliseconds, a significantly improvement in comparison with an existing method, while the recognition accuracy was kept almost equivalent to the existing method.

キーワード：指文字認識, 主成分分析, マハラノビス距離, 3次元スキャナ

Finger character recognition, Principal component analysis, Mahalanobis' distance, 3-dimensional scanner

## 1 はじめに

近年、コンピュータと人間の自然な対話の実現を目指し、人間の動作や表情を検出する研究が盛んに行われている<sup>[1][2]</sup>。その中で、手指は多くの関節を含むため、ジェスチャや指文字など様々な形状を表現できるが、その一

方で、形状の認識、モデル化が難しい部位である<sup>[3-6]</sup>。表現力の高い手指動作は、自然かつ手軽なコンピュータへの入力手段であり、仮想空間制御や3次元形状モデリングに利用されている<sup>[3][4][7-9]</sup>。

キーボードやマイクの利用が困難である、または憚られる状況におけるコンピュータへの自然な入力の一手段として、日本語五十音指文字が着目されている。指文字は手話と同様、聴覚障害者の主要なコミュニケーション手段である。手話が存在しない言葉や固有名詞を相手に伝える場合、手話の単語を説明する場合などに用いられる。手話は、大まかな手形状とその動きに主眼を置いて認識が行わ

\* 鹿児島大学工学部情報工学科

Department of Information and Computer Science, Faculty of Engineering, Kagoshima University

† sc098011@ics.kagoshima-u.ac.jp

‡ sc101013@ics.kagoshima-u.ac.jp

§ ono@ics.kagoshima-u.ac.jp

¶ shignaka@ics.kagoshima-u.ac.jp



れるのに対し、指文字認識は指の開き具合、曲げ具合に着目して手形状の詳細な認識を行う必要がある。

手形状の情報を得る方法は、接触型センサを用いる方法と非接触型のセンサを用いる方法とに大別される。データグローブ等の接触型センサを用いると、各関節の角度を取得することができ<sup>[8-10]</sup>、指文字認識を容易に行うことができる。しかし、装置の脱着が必要となるため、被観測者に負担を与えるだけでなく、ケーブルによって動作範囲が制限される。このため、近年では、非接触型センサを用いた方式が注目されている<sup>[3][4][11-14]</sup>。カメラを用いて手形状の情報を得る方式は、被観測者の負担を軽減できるものの、正確な形状データを取得することが難しい、背景が単純でなければならないなどの問題がある。また、日本語五十音指文字は、よく似た形状を持つ文字が複数個存在しており、カメラを用いた2次元画像処理では正確な識別が難しい<sup>[15][16]</sup>。複数台のカメラを利用する方式<sup>[6][17]</sup>は、形状情報を取得することができるが、手を囲むように、またはカメラ間の距離、角度を一定に保つなど、カメラの設置に関する制限によって、利用できる状況が限られてしまう。

近年、3次元スキャナを用いて身体形状の一部を取得し、認証、採寸などに用いる研究が行われている<sup>[18][19]</sup>。カメラのかわりに3次元スキャナを用いて手形状を取得することにより、背景に依存せずに手形状のみを取得できる。また、立体形状データを取得できるため、従来のカメラを用いた方式では判別が難しい、類似した指文字を正しく識別することができる<sup>[20]</sup>と考える。

著者らは、3次元スキャナとカメラを併用し、手話、指文字の双方を認識する実用システムの開発を目指しており、これまでに、3次元スキャナと3次元テンプレートマッチング(3-Dimensional Template Matching: 3D-TM)により指文字の認識を行う方式を提案している<sup>[20]</sup>。3D-TMは、高精度で指文字の認識を行えるが、1文字あたり10秒以上の認識時間がかかるという問題があった。

本稿では、3次元形状データに対して主成分分析<sup>[21-24]</sup>による次元圧縮を行った特徴ベクトルにより、指文字の認識を行う方式を提案し、その有効性を検証する。提案する指文字認識方式は、3次元スキャナを用いて距離画像を作成し、形状を直接比較するのではなく、次元圧縮した特徴ベクトルの比較により指文字を認識することで、3D-TMと同等の精度を保ちつつ、より高速に認識を行うことができる。

なお、現状では3次元スキャナのデータの取得に0.6秒の時間が必要となるため、動作を伴う手形状の認識は困難であり、実用的な手話、指文字認識を開発するためには、3次元スキャナとカメラの併用が必要である。3次元スキャナとカメラの併用方法については、別の論文で検討するものとし、本稿では、3次元スキャナを用いた指文字認識方式についてのみ述べる。

以下、2章では、指文字の概要および従来方式の問題点について述べ、3章で本方式の方針、処理手順について説明する。4章では、10人の被験者から採取した指文字データをもとに実験を行い、本方式の有効性を検証する。

## 2 日本語指文字と従来の認識方式

### 2.1 日本の指文字

指文字とは、片手の指の曲げ伸ばしにより文字や数字を表す体系であり、アルファベット26文字を表す指文字体系や、日本語かな五十音を表す体系がある。五十音を表す指文字は、一般的に大曾根式指文字が利用されることが多い。指文字は、手話単語が確定していない固有名詞や外来語を表す際に用いられる。漢字に対するふりがなのように、手話に対する読み方の説明に用いられることや、手話と組み合わせられて使われることもある。

五十音指文字の一覧を図1に示す。図1の指文字は、相手から見た手指の形を示している。「の」「も」「り」「を」「ん」の5文字および濁音などは動きを伴うものであり、その他



### 3 主成分分析による指文字認識

#### 3.1 基本方針

本稿で提案する指文字認識方式の基本方針と特徴を以下に示す。

**方針 1: 3次元スキャナを用いて手形状を入力する。** グローブ型デバイスやマーカなどをユーザが装着する必要がないため、ユーザの動作が制限されない、手軽に利用できるなどの利点がある。また、3次元スキャナを利用することで、複雑な背景や動的に変化する領域を含む背景であっても、手形状のみを容易に取得できる。

本方式では、単一方向からのみ形状の取得を行う3次元スキャナを利用する。このため、通常のカメラと同様の設置、運用が可能であり、ロボットなどへの実装も可能である。

**方針 2: 主成分分析を用いて形状特徴データの次元圧縮を行う。** 3D-TM は高い精度で認識を行えるものの、学習データ数に比例して実行速度が低下してしまう。本稿で提案する方式は、主成分分析を用いて次元圧縮を行うことにより、高速に認識を行える。また、学習データ数が増加した場合に、実行速度が低下することなく、より高精度で認識を行うことができる。

#### 3.2 システムの構成と処理手順

本方式の構成及び処理手順を図2に示す。

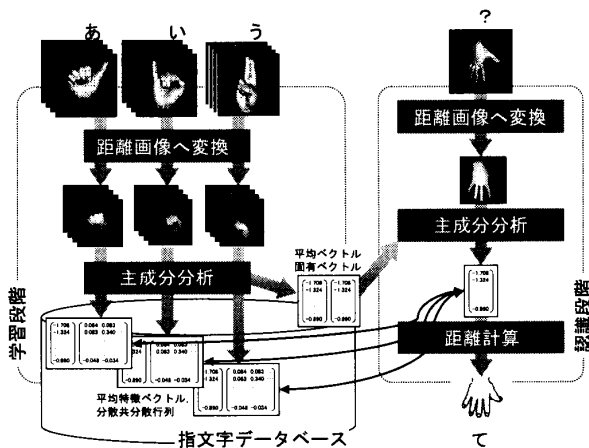


図2 処理手順

本方式は、指文字形状をデータベースに登録する学習段階、および、入力された指文字形状から指文字を認識する認識(テスト)段階からなる。

学習段階では、まず、取得した3次元手形状データを距離画像へと変換する。このとき、3次元手形状データに回転を施すことで、派生画像を生成する。次に、距離画像に対してモザイク処理を行う。モザイク処理を行った画像に対して主成分分析を行い、次元圧縮に用いる固有ベクトルを得る。得られた固有ベクトルを用いて、各指文字ごとの平均特徴ベクトルおよび分散・共分散行列を求める。固有ベクトル、各指文字の平均特徴ベクトル、および分散・共分散行列からなるデータベースを指文字データベースと呼ぶ。

認識段階では、まず、学習段階と同様に、次元圧縮された特徴ベクトルを得る。次に、固有指文字データベースに登録されている各指文字とのマハラノビス距離を求め、距離が最短となる指文字を認識結果として出力する。

#### 3.3 3次元スキャナによる手形状の取得

本研究における手形状の取得環境を図3に示す。3次元スキャナと手の距離は60~70cmとする。被観測者の手を固定せずに、3次元スキャナに接続したモニタを参照し、モニタ内に手が全て映るよう、大まかに手の位置を調整する。3次元スキャナの両脇には、照明を設置する。また、本稿の実験では、手首の写り具合のばらつきを抑えるため、黒い布で

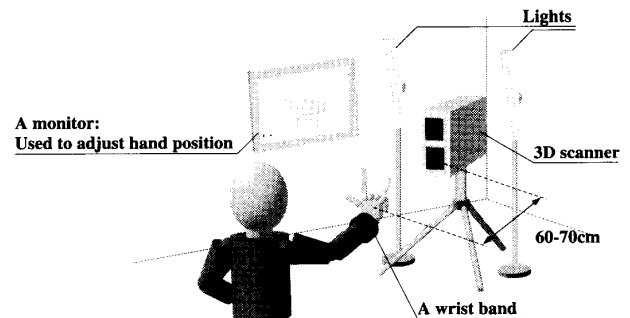


図3 手形状の取得環境

作成したリストバンドを装着する。手首位置の推定を行うことで<sup>[15][16]</sup>，将来的にはリストバンドの装着は不要とする。

図4(a)に3次元スキャナのモニタ出力を，図4(b)に3次元スキャナで取得した手形状の例を示す。被験者以外の人物を含み，肌色領域が重なっているような，複雑な背景を含むにも関わらず，距離差を利用して手形状のみを取得できていることがわかる。

なお，本稿では， $x$ ， $y$ ，および $z$ 軸を図5のように設定する。

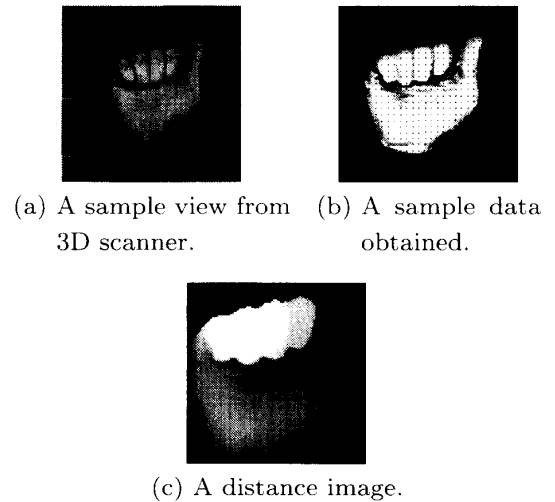


図4 3次元スキャナの出力

### 3.4 距離画像への変換

本研究で用いる3次元スキャナは，単一方向からデータの取得を行うため，取得した手形状データを，奥行き情報を明暗で表す距離画像へと変換する。3次元手形状データはポリゴンの集合として表現されている。距離画像は，重心を中心として最大 $100 \times 100$ ピクセル内に収まるように調整し，奥行き，すなわち $z$ 軸の座標値を階調値とする。

本方式では，手を固定せずに撮影を行うため，位置ずれや角度ずれに対応する必要がある。このため，学習用の手形状データ1個から得られる距離画像(元距離画像)に加え，手形状データを $x$ - $y$ 平面内で平行移動することで得られる画像，および， $z$ 軸に平行で重心を通る軸を中心とし，回転を行うことで得られる画像を，派生距離画像として学習に用いる。すなわち，1枚の元距離画像と派生距離画像 $N_D$ 枚を学習に用いることとなり，各指文字について3枚の手形状データを取得し， $C$ 文字の指文字について学習する場合は， $3 \times C \times (1 + N_D)$ 枚の距離画像を学習に用いることとなる。

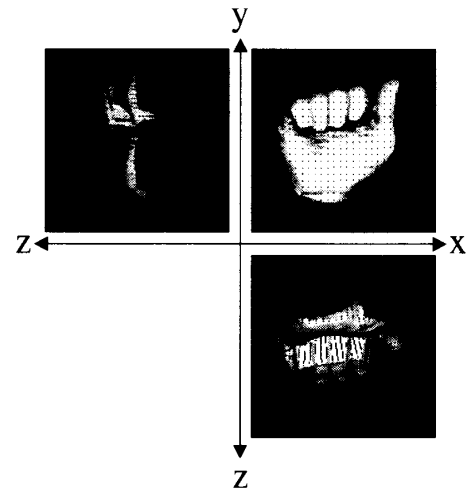


図5 軸設定

### 3.5 モザイク処理

前節までの処理によって得られる距離画像を直接，主成分分析に用いると，逆行列の計算に膨大な時間が必要となるため，モザイク処理を行う(図6)。横 $5 \times$ 縦 $5$ 画素ごとに距離の平均を求め，その平均値で代用すると，

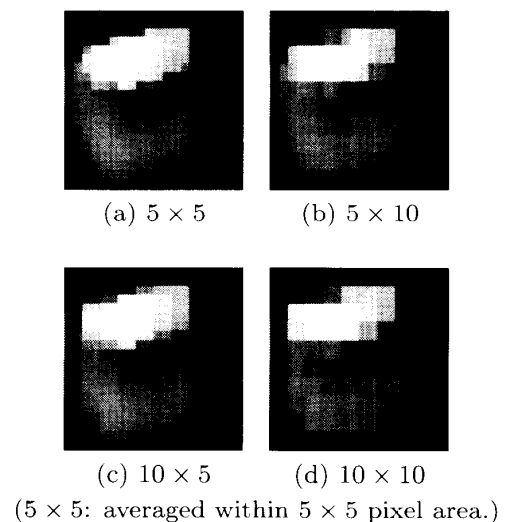


図6 モザイク処理

横 20 × 縦 20画素のモザイク距離画像が作成でき、画素数に対応するベクトルは、10,000次元(横 100 × 縦 100 画素) から 400 次元まで圧縮できる。

### 3.6 主成分分析による次元圧縮

主成分分析<sup>[27-29]</sup>は、画像の輝度値や形状、オプティカルフローなどの特徴ベクトルの集合から、射影成分の分散が大きい射影軸をあらかじめ求めておき、特徴ベクトルをその射影軸への射影成分(主成分)で表す手法である。多変量の計測値から変量間の相関を除去し、より低次元の変量によって元の計測値の特性を記述できる。

主成分分析による次元圧縮の手順を以下に示す。

[Step 1] 学習用距離画像  $N$  枚に対し、各画素値を並べた  $d$  次元のベクトル  $x^{(n)}$  ( $n = 1, 2, \dots, N$ ) を作成し、平均ベクトル  $\mu$ 、分散・共分散行列  $R$  を求める。

[Step 2] 次式 (1), (2) を満たす、行列  $R$  の固有値  $\lambda_i$  ( $i = 1, 2, \dots, d$ ) の上位  $k$  個、および対応する固有ベクトルを  $\psi_j$  ( $j = 1, 2, \dots, k$ ) を求める。

$$\psi_j^t \psi_{j'} = \begin{cases} 1 & (j = j') \\ 0 & (j \neq j') \end{cases} \quad (1)$$

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d \quad (2)$$

[Step 3] 次式 (3) に従い、 $\mu$  および  $\psi_j$  を用いて、各学習用画像  $x^{(n)}$  を次元圧縮した特徴ベクトル  $\hat{x}^{(n)} = (\hat{x}_1^{(n)}, \hat{x}_2^{(n)}, \dots, \hat{x}_k^{(n)})$  を得る。

$$\hat{x}_j^{(n)} = (x^{(n)} - \mu)^t \psi_j \quad (3)$$

### 3.7 特徴ベクトルの比較

本方式では、分散を考慮して入力と各分布の平均との距離を計算するために、マハラノビス距離を用いる。マハラノビス距離は、多変量正規分布のもとで等確率密度と等距離を対応させたものである。  $k$  次元に圧縮された特徴ベクトル  $M_c$  個 が属するクラス  $c$  と、入力

ベクトル  $y$  を 3.6 節に従って次元圧縮した特徴ベクトル  $\hat{y}$  とのマハラノビス距離  $D(\hat{y}, m_c)$  を次式 (4) で計算する。

$$\{D(\hat{y}, m_c)\}^2 = (\hat{y} - m_c)^t \Sigma_c^{-1} (\hat{y} - m_c) \quad (4)$$

ここで、 $m_c$  はクラス  $c$  の平均特徴ベクトルを、 $\Sigma_c$  は分散・共分散行列を表す。

学習段階では、各クラスの  $m_c$ 、 $\Sigma_c$ 、および 3.6 節の  $\mu$ 、 $\psi_j$  を算出し、固有指文字データベースに格納する。認識段階では、入力された手形状に対し、マハラノビス距離が最小となるクラスを認識結果として出力する。

## 4 評価実験

### 4.1 実験準備

本方式の有効性を評価するために、10 人の被験者 ( $P_1, P_2, \dots, P_{10}$ ) から手指形状を取得し、学習用派生画像種の違い (4.2 節)、学習データ数 (4.3 節)、被験者ごとの認識率 (4.4 節)、および、モザイクサイズ (4.5 節) による認識率の変化について実験を行った。また、先行研究との比較検討を行った。本節では、上記の実験に共通する条件について述べる。

本実験では、3 次元スキャナとしてミノルタ製 3 次元デジタイザ VIVID 300 を用い、PC/AT 互換機 (CPU: 3.6 GHz, メモリ: 2 GByte) を用いて計算を行った。VIVID 300 は 1 回の測定に 0.6 秒の時間を要するため、先行研究<sup>[16][20]</sup>と同様に、日本語指文字 46 文字のうち動きを伴わない 41 文字を用いた。被験者 1 人につき、4 セットの指文字データを取得した。実験は被験者ごとに行うものとし、Leave-one-out 法により実験を行った。すなわち、4 セットの指文字データのうち 1 セットを評価用、残り 3 セットを学習用とする実験を 1 回とし、評価用データを変更して 4 回実験を行うものとした。本実験では、指文字の使用経験がない、または少ない被験者が、図 1 を参照しながら指文字を提示してもらうものとした。これは、指文字の使用経験がない被験者は同じ指文字での手指形状のばらつきが大きく、使

用経験がない被験者の指文字を認識できれば、指文字常用者の指文字を容易に認識できるためである。5人の被験者  $P_1, P_2, \dots, P_5$  は1日で4セットの学習データを取得し、他の5人の被験者  $P_6, P_7, \dots, P_{10}$  は2日以上に分けて4セットの学習データを取得した。2日以上に分けてデータを取得する場合は、同じ被験者による同じ指文字であっても、指の曲げ具合などのばらつきが大きくなるを考える。

本稿におけるすべての実験は被験者ごとに行うものとした。すなわち、学習用データは、被験者10人から取得した全ての学習データを用いるのではなく、評価に用いるデータと同じ被験者の学習用データのみを用いることとした。41文字の指文字のうち正しく認識した割合を、全被験者の全試行において平均したものを認識率とした。すなわち、被験者  $p$  の試行  $t$  において、正しく認識を行った指文字の数を  $c_{p,t}$  とすると、認識率は以下の式で表せる。

$$\frac{1}{10 \times 4} \sum_{p=1}^{10} \sum_{t=1}^4 \frac{c_{p,t}}{41}$$

認識時間は指文字1文字の認識に必要なとする平均処理時間とし、学習時間は指文字データベースの作成に要する平均処理時間とした。

## 4.2 学習用派生画像種

取得した形状データに対して回転および平行移動操作を施すことによって得られる派生距離画像によって、認識性能がどの程度変化するかを評価するために、以下の5種類の方式を比較する。モザイクサイズは  $5 \times 5$ 、学習データセット数は3とし、以下の各設定について、1名の被験者あたり4回の実験を10名の被験者に対して行った。

$M_s$ : 1つの形状データを、 $x-y$ 平面上で8方向へ平行移動することで得られる画像8枚を含む9枚の距離画像を生成する。

$M_{r1}$ : 1つの形状データを、 $z$ 軸を中心として  $-10$ 度から  $+10$ 度まで、1度単位

で回転することで得られる画像20枚を含む21枚の距離画像を生成する。

$M_{r2}$ : 1つの形状データを、 $z$ 軸を中心として  $-20$ 度から  $+20$ 度まで、2度単位で回転することで得られる画像20枚を含む21枚の距離画像を生成する。

$M_{r3}$ : 1つの形状データを、 $z$ 軸を中心として  $-30$ 度から  $+30$ 度まで、3度単位で回転することで得られる画像20枚を含む21枚の距離画像を生成する。

$M_{r+s}$ :  $M_s$ と同様の平行移動により得られる画像8枚、および、 $M_{r2}$ と同様の回転により得られる画像20枚を含む29枚の距離画像を生成する。

利用する特徴ベクトルの次元数を変化させた場合の、認識率の変化を図7に、処理時間の変化を図8に示す。本稿では、次元数が学習画像数を上回るために、分散・共分散行列  $\Sigma_i$  の正則性を保てない次元においては、実験を行わないものとした。例えば、本節の実験の

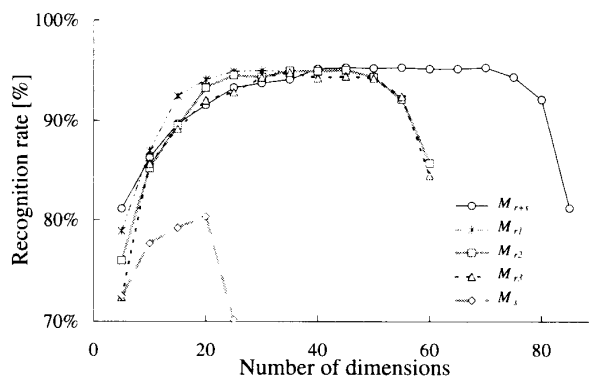


図7 学習用派生画像種による認識率の変化

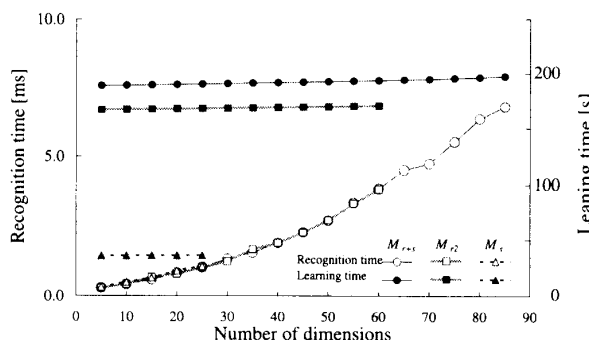


図8 学習用派生画像種による処理時間の変化

$M_s$ では、3セットの学習データを用い、各形状データから9枚の距離画像を生成するため、指文字1文字あたり $3 \times 9 = 27$ 枚の画像を学習に利用する。このため、27次元以上では $\Sigma$ が正則性を満たさない。また、 $M_{r1}$ 、 $M_{r2}$ および $M_{r3}$ の認識時間および学習時間は、同程度であったため、図8には $M_{r2}$ のみの処理時間を示す。

図7より、回転による派生画像の生成は認識率の改善に大きな効果があることがわかる。 $M_{r1}$ 、 $M_{r2}$ 、および $M_{r3}$ を比較すると $M_{r2}$ の認識率が若干ではあるがもっとも高い。平行移動による派生画像の生成は、先行研究<sup>[24]</sup>では認識率が向上していたが、本方式では、手形状の重心を中心として位置の初期化を行っているため、認識率の改善に寄与していない。

図8より、認識時間は特徴ベクトルの次元数の増加に応じて、学習時間は派生画像の枚数に応じて、それぞれ増加することがわかる。

本稿における以後の実験では、 $M_{r2}$ の方法で学習用派生画像種を作成する。

### 4.3 学習データ数

学習に用いるデータセット数を1セットから3セットへ変化させた場合の認識性能の変化を比較する。モザイクサイズは $5 \times 5$ とした。

それぞれの学習用データセット数の場合において、学習用データとテスト用データの可能な組合せ全てについて試行を行うものとした。すなわち、3セットのデータを学習に用いる場合は、各テストデータにつき1回の試行を行うため、合計4回の試行を行う。1セットまたは2セットのデータを学習に用いる場合は、各テストデータにつき学習データを変えて3回の試行を行い、合計12回の試行を行う。

学習データセット数を1から3まで変更した場合の認識率を図9に、平均認識時間および平均学習時間を図10に示す。図9より、より多くの学習用データセットを用いることで、より多次元の特徴ベクトルを利用できる

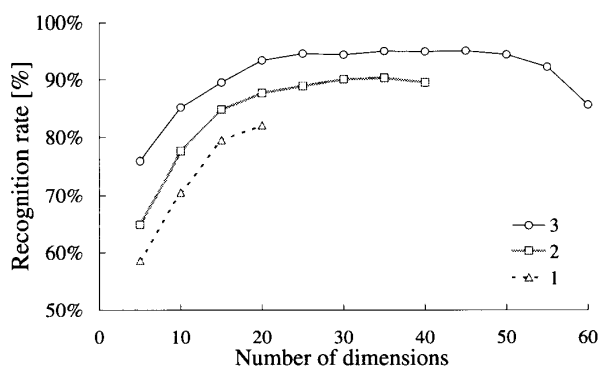


図9 学習データセット数による認識率の変化

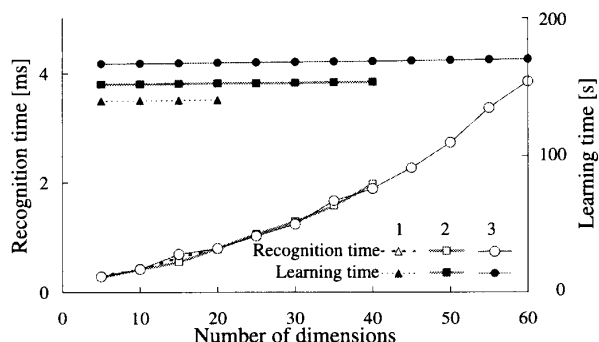


図10 学習データセット数による処理時間の変化

こと、および、認識率が改善されることがわかる。また、図10より、平均認識時間は学習データセット数に依存せず、特徴ベクトルの次元数にのみ依存することがわかる。学習データセット数が増加するにつれて、平均学習時間は長くなるが、3セットの学習データを用いる場合でも3分程度で学習が終了することがわかる。

### 4.4 被験者ごとの認識率の違い

被験者から手形状データを取得する際の方法の違いによる認識率の差を調べた。モザイクサイズは $5 \times 5$ とし、3セットのデータを用いて学習を行った。

被験者 $P_1, P_2, \dots, P_5$ の平均認識率と、被験者 $P_6, P_7, \dots, P_{10}$ の平均認識率を図11に示す。図11より、1日で4セットの手形状データを取得した被験者の方が認識率が高いことがわかる。これは、被験者が指文字の使用経験を持たなかったため数日にわけて4セット

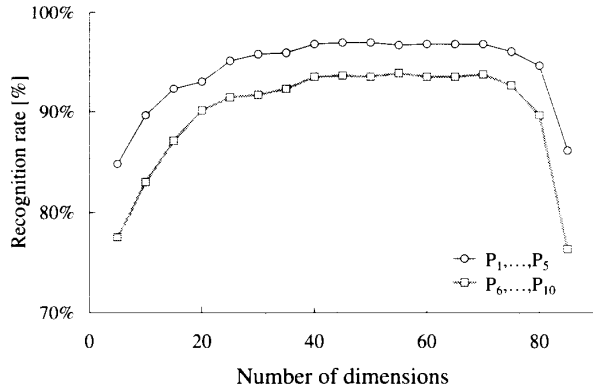


図 11 データ取得方法の違いによる認識率の変化

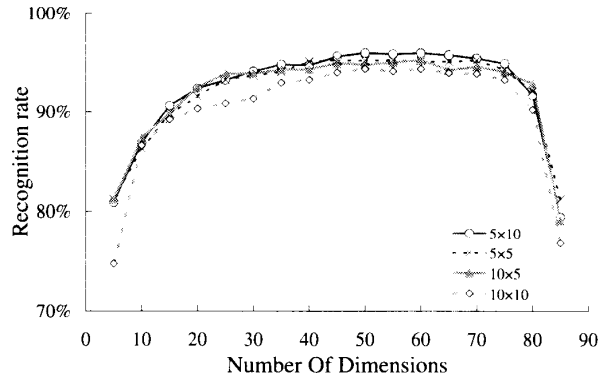
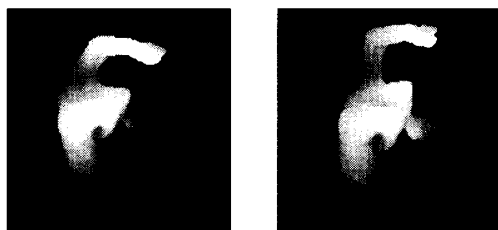


図 13 モザイクサイズによる認識率の変化



(a) Example characters representing “ろ” obtained by examinee P<sub>6</sub>.



(b) Example characters representing “へ” obtained by examinee P<sub>8</sub>.

図 12 誤認識した指文字の例

のデータを取得した被験者の場合、同じ被験者による同じ指文字であっても手形状や角度が大きく異なる指文字が存在したためである。上記の理由で認識に失敗した例を図 12 に示す。図 12 の例では、「ろ」の指文字は「せ」や「ひ」と間違えて認識され、「へ」の指文字は「あ」と間違えて認識された。同じ指文字であっても、手首や指の曲げ具合が大きく異なっていることがわかる。学習データ取得時に、手首、指の曲げ具合を変えながらより多くのデータを取得することで、認識率の改善を行えると考えられる。

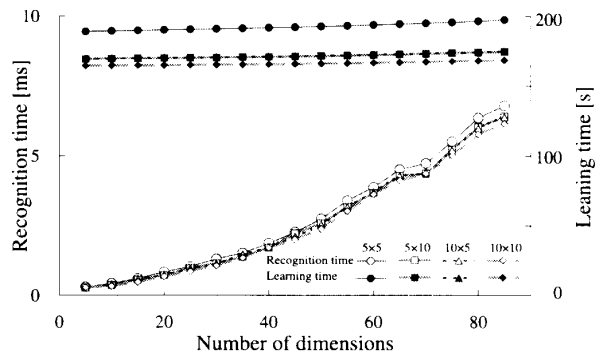


図 14 モザイクサイズによる処理時間の変化

## 4.5 モザイクサイズ

距離画像に施すモザイクのサイズを、横 5 ピクセル × 縦 5 ピクセル (5×5), 5×10, 10×5, 10×10 と設定して実験を行った。学習データセット数は 3 とした。

実験結果を図 13 および図 14 に示す。図 13 から、10×10 を除く 3 種類のモザイクサイズの差は小さく、若干ではあるが、次元数が 30, モザイクサイズが 5×10 サイズの場合に、最も認識率が高くなることがわかる。これは、高解像度であれば認識率が高くなるのではなく、認識対象に応じた最適な解像度が存在すること [30][31], および、モザイク処理後、主成分分析による次元圧縮を行って得られた特徴ベクトルの次元数が同じであれば、認識率に大きな差はないことを示している。同様の理由により、認識時間もモザイクサイズに依存せず、特徴ベクトルの次元数に依存していると考えられる。また、図 14 から、モザイクサイズが大きいほど学習時間が短縮されることがわかる。



#### 4.6 カメラを利用する先行研究との比較

2.2節で概説した先行研究のうち、センサ、マーカやグローブの装着が不要な方式と本方式を比較する。提案する方式は、4.3節と同様の条件とした。

齋藤らの Perceptual Glove<sup>[17]</sup>と比較すると、Perceptual Glove が判別を行う 6 種のジェスチャ “Zero”, “One”, “Two”, “Five”, “OK”, “Point” はそれぞれ、日本語五十音指文字の「さ」、「ひ」、「か」、「て」、「め」、「れ」に相当する。Perceptual Glove は未登録者を含む実験であるが、6 種類のみジェスチャを 85% の精度で識別するのに対し、本稿における実験では、未登録者を含まないものの、データベースに 41 文字の指文字が登録された状態で、上記 6 種の指文字の認識率は 95% となる。よって、本方式と Perceptual Glove の直接的な比較は困難であるものの、本方式は Perceptual Glove と同等以上の識別率で、Perceptual Glove の 6 種類のジェスチャ以外にも多くの指文字を認識できることがわかる。

広瀬の細線化に基づく方式<sup>[16]</sup>と比較する。細線化に基づく方式は、ビデオ「百万人の手話」(ダイナミックセラーズ) から取得した指文字画像を用いて実験を行っているのに対し、本稿では、指文字の知識を持たない 10 人の被験者からデータを取得したため、指の開き具合等のばらつきが大きいデータを多く含む。細線化に基づく方式で認識を行えなかった指文字「い」と「ち」、「え」と「さ」、「せ」と「た」と「ひ」と「ら」、および、「ぬ」と「ろ」の、本方式における認識率は 93% であり、細線化に基づく方式では識別が困難な指文字を、本方式は 3D-TM 同様、距離情報を利用することで<sup>[20]</sup>高い精度で認識できることがわかる。

Perceptual Glove および細線化に基づく方式との比較から、本方式は多くの指文字を高い精度で認識できることがわかる。また、上記の 2 方式は、背景が均一な画像でなければならない、または肌色を含んではならないなどの制約を持つのに対し、本方式は複雑な背

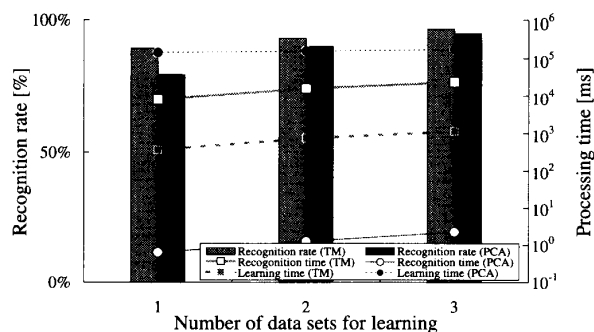


図 15 主成分分析とテンプレートマッチングの比較

景でも手領域を正確に抽出できる。

#### 4.7 3次元テンプレートマッチングとの比較

提案する主成分分析に基づく方式 (PCA) と、従来方式である 3D-TM による指文字認識方式との比較を行った。学習に用いるデータセット数が 1, 2, および 3 の場合における平均認識率、平均認識時間、および平均学習時間を図 15 に示す。なお、提案する方式は 4.3 節と同様の条件で実験を行い、次元数はデータセット数が 1, 2, および 3 の場合にそれぞれ、15 次元、30 次元、45 次元の特徴ベクトルを用いるものとした。3D-TM では、2 セット以上の学習データを用いる場合、全てのデータをデータベースに保持するものとした。

図 15 から、認識率は全般的に 3D-TM の方が高いものの、学習データセット数が増加するにつれて、PCA の認識率が 3D-TM の認識率に近づくことがわかる。また、3D-TM は、認識時間が 10 秒以上であるのに対し、PCA の認識時間は 10ms 以下と非常に高速であることがわかる。学習時間は 3D-TM が高速であることから、学習データ数が少ない場合や、逐次的なデータ取得による学習が必要な場合には 3D-TM が適し、高速な認識を行う必要がある場合には PCA が適することがわかる。

表 1 に、学習データセット数を 3 とした場合における、PCA および 3D-TM を用いた方式の、指文字ごとの平均認識率を示す。「い」

表 1 指文字毎の認識率

|   | 3D-TM | PCA  |   | 3D-TM | PCA  |      | 3D-TM | PCA  |
|---|-------|------|---|-------|------|------|-------|------|
| あ | 98%   | 93%  | そ | 98%   | 90%  | ほ    | 98%   | 100% |
| い | 95%   | 88%  | た | 100%  | 98%  | ま    | 90%   | 95%  |
| う | 95%   | 90%  | ち | 100%  | 93%  | み    | 98%   | 95%  |
| え | 100%  | 93%  | つ | 93%   | 98%  | む    | 100%  | 88%  |
| お | 95%   | 98%  | て | 100%  | 98%  | め    | 98%   | 100% |
| か | 90%   | 93%  | と | 100%  | 95%  | や    | 98%   | 98%  |
| き | 100%  | 98%  | な | 95%   | 93%  | ゆ    | 100%  | 100% |
| く | 98%   | 98%  | に | 100%  | 98%  | よ    | 95%   | 95%  |
| け | 93%   | 98%  | ぬ | 95%   | 100% | ら    | 95%   | 90%  |
| こ | 100%  | 100% | ね | 95%   | 100% | る    | 90%   | 93%  |
| さ | 100%  | 90%  | は | 100%  | 100% | れ    | 95%   | 95%  |
| し | 95%   | 95%  | ひ | 95%   | 98%  | ろ    | 98%   | 93%  |
| す | 95%   | 98%  | ふ | 98%   | 90%  | わ    | 100%  | 98%  |
| せ | 98%   | 85%  | へ | 98%   | 98%  | Avg. | 97%   | 95%  |

「さ」「せ」および「ち」など、類似した指文字では、3D-TMの認識率が若干高いものの、PCAと3D-TMの認識率が同程度である指文字が多いことがわかる。

#### 4.8 考察

本稿の実験では、モザイクサイズが $5 \times 10$ 、学習データセット数が3、学習用派生画像種が $M_{r,2}$ 、利用する特徴ベクトルの次元数が60の場合に認識率が最も高くなり、95.9%であった。また、上記の場合における認識時間は、3.7msであった。本稿の実験により、主成分分析に基づく指文字認識は極めて高速であることがわかり、3次元スキャナの入力速度が改善されることで、キーボード等の入力機器の利用が困難な状況における、計算機との対話を実現する一助となると考える。

カメラを利用した従来の手話やジェスチャ認識方式に、3次元スキャナを利用する本方式を援用することで、複雑で動的に変化する背景のもとでの正確な手形状の取得および認識が可能となる。上記の特徴を活かした一つの応用例として、自律移動ロボットへの実装がある。ロボットと共に移動を行いながらの認識や、人混みの中など複雑かつ動的な背景を含む環境でも、ジェスチャや指文字によって

ロボットとの対話を実現できる本方式は、聴覚障害者だけでなく、人間とロボットの自然なコミュニケーションを実現する一方式として有用であると考えられる。

#### 5 おわりに

本稿では、3次元スキャナを用いて手形状を取得し、主成分分析によって指文字を認識する方式を提案した。また、実験により、モザイクサイズ、学習用派生画像生成方法などについて検討を行った。本方式の特徴を以下に示す。

- センサやマーカの装着が不要である。
- 複雑で肌色を含む背景や動きを伴う背景であっても安定した認識を行うことができる。
- 本研究で用いる3次元スキャナは、単一方向からのみデータの取得を行うものであるため、1台のカメラと同様の設置が可能である。
- 3次元テンプレートマッチングと同等の精度で、かつ、3次元テンプレートマッチングよりも高速に指文字の認識を行うことができる。
- 学習データを増加させることで、より高

い精度で認識を行え、認識速度も低下しない。

本稿は、3次元スキャナを用いて取得した手形状データをもとに、指文字の識別を行う方式の基礎的な有効性を示すものである。次元圧縮された特徴情報を用いた認識方式は、情報知識学の様々な応用分野で有用であると考えられる。今後、判別分析<sup>[32]</sup>や制約相互部分空間法<sup>[33]</sup>などの利用による認識率の改善、未登録者による指文字の認識性能の検証、および、カメラとの併用による実用的なインタフェースの実現について検討する。

## 参考文献

- [1] 本郷仁志;大矢光範;渡辺博己;安本護;山本和彦:「注意機構を備えた顔・手認識による視覚インタフェースの開発」, 電気学会論文誌, Vol.122-c, No.4, pp.624-629, 2002.
- [2] 黒澤篤志;岡本教佳:「重なりを考慮した顔領域抽出法」, 電気学会論文誌, Vol.122-C, No.4, pp.662-668, 2002.
- [3] 島田伸敬;白井良明;久野義徳:「確率に基づく探索と照合を用いた画像からの手指の3次元姿勢推定」, 電子情報通信学会論文誌 D-II, Vol.J79-D-II, No.7, pp.1210-1217, 1997.
- [4] 岩井儀雄;八木康史;谷内田正彦:「単眼動画からの手の3次元運動と位置の推定」, 電子情報通信学会論文誌 D-II, Vol.J80-D-II, No.1 pp.44-55, 1997.
- [5] 長田礼子;尾崎哲;青木輝勝;安田浩:「手指動からの特徴抽出によるリアルタイム個人認証」, 電子情報通信学会論文誌 D-II, Vol.J84-D-II, No.2, pp.258-265, 2001.
- [6] 上田悦子;松本吉央;今井正和;小笠原司:「多視点シルエット画像を用いた手の形状推定」, 情報処理学会研究報告, 2001-CVIM-128, pp.25-31, 2001.
- [7] 舟橋健司;安田孝美;横井茂樹;鳥脇純一郎:「3次元仮想空間における仮想手による物体操作モデルと一実現法」, 電子情報通信学会論文誌 D-II, Vol.J81-D-II, No.5, pp.822-831, 1998.
- [8] 西野浩明;凍田和美;宇津宮孝一:「オンライン学習機能を備えた対話型両手ジェスチャインタフェース」, 電子情報通信学会論文誌 D-II, Vol.J81-D-II, No.5, pp.897-905, 1998.
- [9] 西野浩明;凍田和美;宇津宮孝一:「両手ジェスチャで変形可能な3次元形状表現法」, 情報処理学会論文誌, Vol.40, No.2, pp.698-701, 1999.
- [10] Lee, Christopher; Xu, Yangsheng: "Online, Interactive Learning of Gesture for Human/Robot Interfaces", IEEE International Conference on Robotics and Automation, pp.2982-2987, 1996.
- [11] 西村拓一;向井理朗;岡隆一:「白黒動画像からの形状特徴を用いたジェスチャのスポットニング認識システム」, 電子情報通信学会論文誌 D-II, Vol.J81-D-II, No.8, pp.1812-1821, 1998.
- [12] 吉野和芳;川嶋稔夫;青木由直:「色の組み合わせによるジェスチャの直接的推定」, 電子情報通信学会論文誌 A, Vol.J79-A, No.2, pp.424-431, 1996.
- [13] 吉野和芳:「カラーパッチ手袋による動作ゆらぎを考慮した指文字推定の検討」, 情報処理学会研究報告, 99-CVIM-116, pp.25-30, 1999.
- [14] 渡辺賢;岩井儀雄;八木康史;谷内田正彦:「カラークロブを用いた指文字の認識」, 電子情報通信学会論文誌 D-II, Vol.J80-D-II, No.10, pp.2713-2722, 1997.
- [15] 長嶋祐二;小野寺卓;長嶋秀世;寺内美奈;大和玄一:「指文字認識に関する基礎的検討」, 電子情報通信学会技術研究報告, Vol.SP92-64, pp.23-30, 1992.
- [16] 広瀬健一:「細線化画像を用いた指文字認識」, 情報処理学会研究報告 コンピュータビジョン, Vol.93, No.84, pp.1-6, 1993.

- [17] 齋藤真希子；佐藤洋一；小池英樹：「Perceptual Glove: 多視点画像に基づく手形状・姿勢の実時間入力とその応用」, 情報処理学会論文誌, Vol.43, No.1, pp.185-194, 2002.
- [18] 矢原弘樹；日隈直紀；福井幸男；西原清一；持丸正明；河内まき子：「FFDを用いた3次元足部モデルの解剖学的特徴点抽出」, 電子情報通信学会論文誌 D-II, Vol.J87-D-II, No.4, pp.967-977, 2004.
- [19] 王宇；小野智司；中山茂：「3次元耳画像を用いた個人認識」, 情報処理学会火の国情報シンポジウム 2005, B-4-4, 2005.
- [20] 東山和弘；小野智司；王宇；中山茂：「3次元テンプレートマッチングによる指文字認識」, 電気学会論文誌 C, Vol.125, No.9, pp.1444-1454, 2005.
- [21] Sirovich, L.; Kirby, M.: "Low dimensional procedure for the characterization of human faces", *Journal of Optical Society of America*, Vol.4, No.3, pp.519-524, 1987.
- [22] Turk, M.A; Pentland, A.P.: "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71-86, 1991.
- [23] Turk, M.A; Pentland, A.P.: "Face recognition using eigenfaces", *Proc.of IEEE Conf. on Computer Vision and Pattern Recognition*, pp.586-591, 1991.
- [24] 王宇；小野智司；武田和大；佐藤公則；中山茂：「主成分分析による耳画像を用いた個人認識」, 情報知識学会誌, Vol.14, No.1, pp.1-10, 2004.
- [25] Gent, I.P.; Walsh, T.: "Towards an understanding of hill-climbing procedures for SAT", *AAAI'93*, pp.28-33, 1993.
- [26] Yokoo, M.: "Why Adding More Constraints Makes a Problem Easier for Hill-Climbing Algorithms: Analyzing Landscape of CSPs", *3rd Int'l Conf. on Principles and Practice of Constraint Programming*, pp. 356-370, 1997.
- [27] 赤松茂：「コンピュータによる顔の認識 — サーベイ —」, 電子情報通信学会論文誌 Vol.J80-D-II, No.8, pp.2031-2046, 1997.
- [28] 石井健一郎；上田修功；前田英作；村瀬洋：「パターン認識」, オーム社, 1998.
- [29] 奥野恵一；久米均；芳賀敏郎；吉澤正：「多変量解析法」, 日科技連, 1971.
- [30] 多田昌裕；加藤俊一：「階層的な判別分析を利用した感性的な類似画像検索システム」, *日本データベース学会 Letters*, Vol.1, No.2, pp.20-23, 2003.
- [31] 鈴木一史；加藤俊一；大津展之：「同値類に基づく回転不変特徴量を用いた3次元物体モデルの類似検索」, 電子情報通信学会論文誌 D-II, Vol.J86-D-II, No.8, pp.1234-1243, 2003.
- [32] Belhumeur, P.N.; Hespanha, J.; Kriegman, D.J.: "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", in *Proceedings of 1996 European Conference on Computer Vision*, No.1, pp.45-58, 1996.
- [33] 小坂谷達夫；山口修；福井和広：「制約相互部分空間法を用いた顔認識システムの開発と評価」, 情報処理学会論文誌, Vol.45, No.3, pp.951-959, 2004.

(2005年10月11日受付)

(2005年11月17日採択)

# イネのトランスポゾンディスプレイ解析のための データベースシステム

## A Database System for High-Throughput Transposon Display Analyses of Rice

井上 悦子\* 吉廣 卓哉† 川路 英哉‡ 堀端 章§ 中川 優†

Etsuko INOUE Takuya YOSHIHIRO Hideya KAWAJI  
Akira HORIBATA and Masaru NAKAGAWA

本研究では、トランスポゾンを用いた遺伝子型と表現型との相関解析を支援するデータベースシステムを構築し、イネのトランスポゾン *mPing* を利用した実験に適用し評価した。我々は、*mPing* を利用して大規模なイネの変異体シリーズを作成し、遺伝子型と表現型との関係を解析することで遺伝子機能の推定を目指している。しかし、実験結果から変異の発見作業に多大な労力がかかることが、実験規模を拡大する上でのボトルネックの一つであった。本システムでは、実験で得られる全データをデータベースに格納し、実験結果から変異を発見後、表現型との相関を解析するまでの効率的なインタフェースを提供する。本システムを *mPing* を利用した実験に適用した結果、波形データの重ね合わせや変異箇所候補の自動検出などの効果で、作業時間を従来の数分の一程度に短縮できた。本システムにより、ハイスループットに大規模な実験や解析を行うことが可能となる。

We developed a database system to support whole experiments which analyze correlation between genotypes and phenotypes by using transposons. We also evaluated the system by applying to our experiments with transposon *mPing* in rice. Our research group grow large-scale mutant series of rice by taking advantage of *mPings*, and study correlation between genotypes and phenotypes for predictions of gene functions. However, the analytical phase, in which we find mutation spots from waveform data, involves several problems from a viewpoint of labor amount, and it becomes one of bottlenecks in large-scale experiments. As a solution, our database system manages all the analytical data throughout the experiments, and provides interfaces to perform overall analyses by detecting mutation spots and analyzing correlation between the spots and traits. In order to evaluate the system, we apply to experiments with *mPing* in rice, and confirmed that our system works effective to reduce the analytical labor by several times. This system realizes high-throughput transposon display analyses in large scale.

キーワード：データベース, トランスポゾンディスプレイ解析, 波形比較, 相関分析

Database, Transposon Display Analysis, Comparison of Waveform Data, Correlation Analysis

### 1 はじめに

近年、国際的な研究協力のもとで、ヒトや多数のモデル生物の全ゲノム配列が次々に解読され、インターネット上で公開されている。しかし、いずれの生物も個々の遺伝子の塩基配列がわかっても、その遺伝子の持つ機能を

\* 和歌山大学大学院システム工学研究科  
Graduate School of Systems Engineering,  
Wakayama University  
s031007@sys.wakayama-u.ac.jp

† 和歌山大学システム工学部  
Faculty of Systems Engineering, Wakayama University

‡ NTT ソフトウェア  
NTT Software Corporation

§ 近畿大学生物理工学部  
Faculty of Biology-Oriented Science and Technology, Kinki University

解明することはできていない。そこで、現在はポストゲノム研究として、遺伝子や遺伝子産物である mRNA やタンパク質などの機能や相互作用、パスウェイなどの解析を目的に、様々な研究が行われている。

遺伝子解析のアプローチの一つとして、トランスポゾンによる突然変異体を利用する方法がある。トランスポゾンは、通常の遺伝子とは異なりゲノム上を動くことのできる遺伝子であり、転移によって別の遺伝子に挿入すると遺伝子を破壊することがある。破壊された遺伝子は正常に機能しなくなり、表現型が変化することがある。これを利用し、破壊された遺伝子と表現型との関係を解析することで、遺伝子の機能を明らかにしようというアプローチである。

我々の研究グループでは、イネを対象にトランスポゾン *mPing*<sup>[1]</sup>を利用して変異体シリーズを大規模に作成し、遺伝子型と表現型との共分離の度合いなどを解析することで遺伝子機能の推定を行っている。しかし、このトランスポゾンディスプレイ解析<sup>[2]</sup>を大規模に行う場合には、栽培や *in vitro* な実験作業と同様に、変異を特定する解析作業に多大な時間と労力がかかることが、実験全体を大規模に進める上でのボトルネックの1つとなっていた。

そこで、これを解決するために、トランスポゾンによる変異体を用いて、変異による遺伝子型と表現型との相関解析を支援するデータベースシステムの構築を行い、*mPing* を用いた実験に適用し評価した。本システムでは、トランスポゾンディスプレイ解析の負担を軽

減するために、波形データをシステム上で重ね合わせ、さらに変異箇所の候補を自動検出することで、変異発見にかかる労力を低減し、作業を効率化する。また、変異情報だけでなく表現型情報もデータベース化し、遺伝子型と表現型の相関解析に至るまでの全作業をシステム上で行う web インタフェースを提供することで、解析作業をスムーズに行えるようにするだけでなく、データ管理の効率化も行う。

本論文の構成は以下の通りである。2章でトランスポゾン *mPing* と変異を調べるためのトランスポゾンディスプレイ解析とその問題点を説明し、3章~5章ではそれぞれ、本システムの三つの機能「系統情報と形質情報の管理」「トランスポゾンディスプレイ解析の支援」「遺伝子型と表現型の相関解析」について説明する。最後に、6章でシステムの評価を、7章でまとめを述べる。

## 2 トランスポゾン *mPing* と変異体の解析方法

### 2.1 *mPing* による変異体シリーズの作成

*mPing* は、2003年に発見されたイネゲノム上に存在する MITE 型トランスポゾンの一つである。*mPing* には二つの特徴がある。

- (1) イネゲノム上に 100 コピー程度存在し、自然環境下での適度な転移活性を持つ
- (2) 外来遺伝子ではないため、圃場での大量栽培が可能である

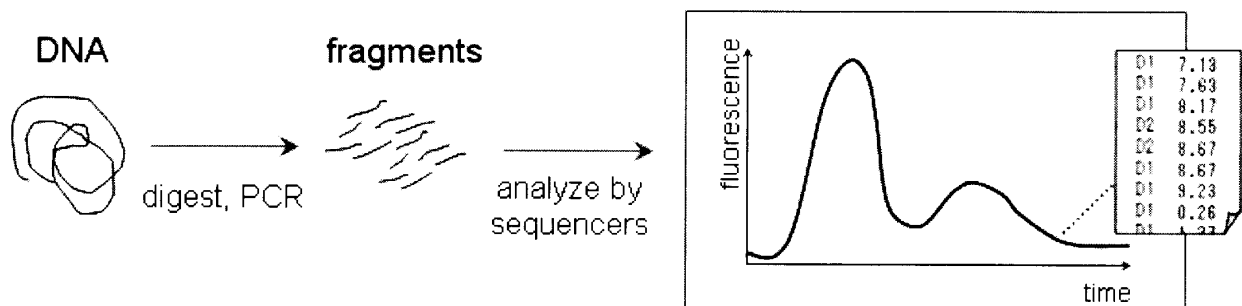


図1 トランスポゾンディスプレイ解析の流れ (波形データを得るまで)

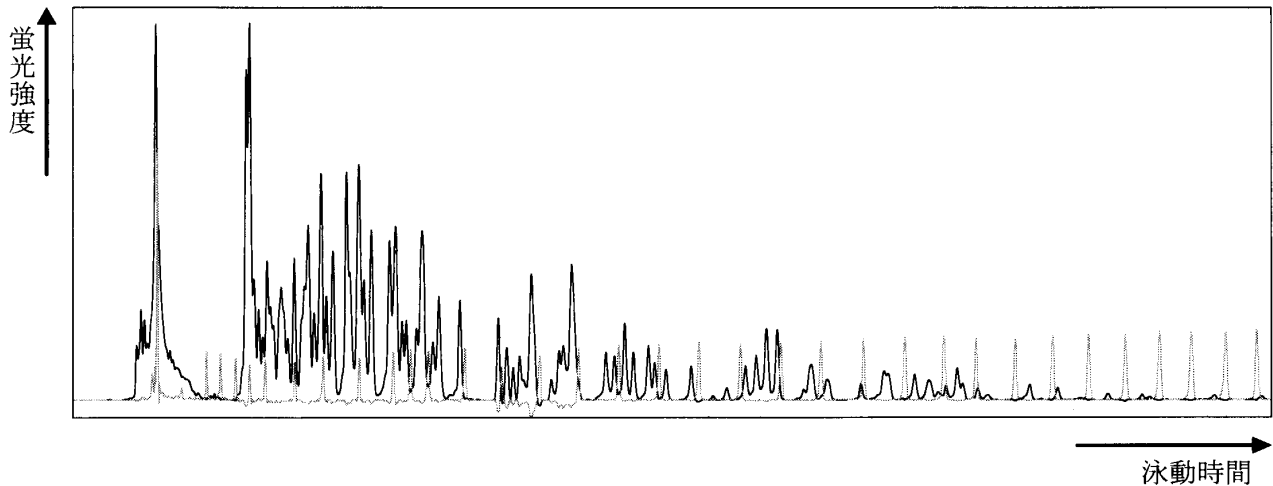


図2 波形データの例 (黒線がサンプル, 灰色線がサイズマーカーの波形)

これらの特徴から, 転移活性のある *mPing* を保持する実験システムを用いることで, 大規模な変異体シリーズを効率よく作成できる. *mPing* は次代交配時に転移を起こすと考えられている.

我々の実験では, 実験対象とするシステムを選別して1システムにつき50~100個体程度を圃場で栽培し, 各個体の出穂日や穂長など20以上の形質項目についてデータを収集する. この実験では, 各個体の系統情報や形質情報と, 2.2節で説明するトランスポゾンディスプレイ解析による変異情報をもとに, 遺伝子の機能推定を行う.

## 2.2 トランスポゾンディスプレイ解析

トランスポゾンの転移によって表現型に変化のあった個体(突然変異体)は, 遺伝子多型を調べるためにトランスポゾンディスプレイ法<sup>[2]</sup>による解析を行い, ゲノム上の変異箇所を特定する. トランスポゾンディスプレイ法は, AFLP法<sup>[3]</sup>を基に開発された手法の一つである.

我々の実験手順を簡単に述べる. まず, 個体ごとに電気泳動結果である波形データを得る. 図1はこの工程を示している. まず, サンプルから抽出したDNAを制限酵素で断片化し, トランスポゾン配列に特異的なプライ

マーによるPCR法を行ってトランスポゾンに隣接する断片のみを選択的に増幅する. 得られたPCR産物をDNAシーケンサーで電気泳動し, 波形データを得る. 次に, 得られた波形データの例を図2に示す. 波形データの縦軸は蛍光強度を, 横軸は電気泳動の経過時間(断片長が短いものから長いものへ順に計測される)を表している. この波形データはイメージファイルではなく, 一定のサンプリング時間ごとに計測した蛍光強度の値のリストファイルであり, 各サンプリング点の蛍光強度はそこに含まれるおおよその断片量を表す. つまり, この波形データにより, どの長さのDNA断片がどのくらいの量含まれているかを知ることができる. なお, 図2には, 決まった間隔ごとにピークが現れるサイズマーカーが同時に泳動されているが, これについては後述する.

次に, 比較対象となる2個体の波形データを比較して変異箇所を検出する. トランスポゾンの転移によってDNA配列が変化している場合には, 転移前後のトランスポゾンの周辺位置の断片の長さが変化するため, 波形データが変化する. よって, 複数の波形データを比較して断片量が異なる箇所を探すことで, 変異箇所を特定できる.

## 2.3 変異と形質の相関解析

トランスポゾンディスプレイ解析による変異情報と形質情報から、遺伝子型と表現型との相関を解析する方法を簡単に説明する。それぞれの形質の変化に関与している変異を推定するためには、各変異の有無によって個体を2群に分割し、形質の測定値に有意差があるかどうかを調べる方法がある。これには、一般的な統計解析の手法である  $t$  検定や  $\chi^2$  検定を用いることができる。

この相関解析は、ある特定の系統や親子・兄弟関係にある個体など、比較的遺伝子型の近い個体グループを単位として行う必要がある。これは次の理由による。トランスポゾンディスプレイ解析で得られる変異情報は、遺伝子そのものではなく、変異した遺伝子周辺にあった断片の長さである。そのため、遺伝子を特定する場合には、その塩基数をもとに該当する断片を抽出し、その塩基配列をシーケンサーで解読して DDBJ<sup>[4]</sup>などのゲノム DB に問い合わせる必要がある。しかし、比較個体の系統が大きく異なる場合には、DNA 配列の相違点が多くなり、制限酵素で断片化すると同一の遺伝子が異なる断片長として得られる可能性が高くなるなど、解析作業が困難となる。したがって、同一系統など DNA 配列がよく似た個体のグループで、トランスポゾンディスプレイ解析および表現型との相関解析を行うことが重要である。

## 2.4 従来手法の問題点

トランスポゾンディスプレイ法は、トランスポゾン配列に隣接する断片に注目できるので多型検出の効率がよい。しかし、波形データを比較して変異箇所を特定する作業に多大な時間と労力がかかる。ところが、この比較作業に適したアプリケーションソフトウェアがないため、従来は、波形データを紙にプリントアウトしたものなどを2つ並べて目視によって比較を行うなど、原始的な方法で変異箇所を調べていた。しかし、波形データ

は図2のように、形状が微細で複雑である上、時間経過とともに蛍光強度の測定値を表す縦軸スケールが大きく減衰するという特徴がある。また、実験時の室温などの影響を受けると、電気泳動の流速が微妙に変化する場合もあり、時間軸のスケールのずれによって比較作業がさらに困難となる。

したがって、我々の研究グループの実験の場合は、*mPing* を利用することで大規模な変異体シリーズは効率よく作成できるが、変異体ごとに変異箇所を調べるトランスポゾンディスプレイ解析での波形データの比較作業にかかる時間や労力が問題となり、このことが実験を大規模に進める上でボトルネックの一つとなっていた。

## 2.5 実験を支援するデータベースシステムの構築

我々は、このようなトランスポゾンによる変異体を用いた解析の問題点を解決するために、*mPing* によるイネの実験をモデルケースとして、一連の実験や解析作業を効率よく進めるためのデータベースシステムを設計し、構築した。本システムでは、トランスポゾンディスプレイ解析で問題となっている波形データの比較作業を単純化するとともに、すべての実験データをデータベース化し、ネットワークを通じて利用できるインタフェースを提供することで、複数の研究者で一貫したデータの利用を可能にするための一元的なデータ管理を行う。これらを実現することで、トランスポゾンディスプレイ解析にかかる時間を短縮し、大規模な実験を効率よくハイスループットに行うことが可能になり、多くの実験データを集積してより精度の高い遺伝子機能の分析が実現できる。

本システムは、(1) 表現型情報の管理、(2) トランスポゾンディスプレイ解析の支援、(3) 表現型と遺伝子型との相関解析の3つの機能に分けることができる。それぞれについて、次章以降で説明する。なお、本システムは web アプリケーションシステムであり、Apache,



PHP および Perl, PostgreSQL で構築した。

### 3 表現型情報の管理

#### 3.1 概要

イネのトランスポゾン *mPing* による実験では、表現型情報として、圃場で栽培した個体の系統情報と形質情報を収集し、管理する必要がある。ここで系統とは、同一個体を親とする個体群を表す識別情報である。これら2種類のデータは、従来、スプレッドシートなどで個別にファイル管理されていたため、データの管理や利用に手間がかかっていた。また、これらのデータは項目の種類が多く、データの発生時期にばらつきがあるという点が問題であった。

本システムでは、表現型情報も含めて一連の実験データをすべてデータベースで一元管理をする。ただし、システム化によりユーザの負担を増さないように、これらのデータは従来と同様のスプレッドシートを使用してデータの入力や編集を行い(ただし、ファイルフォーマットはCSV形式)、そのファイルをアップロードする操作のみでデータベースにデータを反映する。また、データの発生時期が異なる点には、データ登録をいくつかの段階に分割し、それぞれに対応した専用フォーマットのシートを使用することで対応する。

#### 3.2 系統・形質情報の登録と閲覧

系統情報と形質情報の登録と閲覧の手順を説明する。これらのデータは発生時期が異なるため、それぞれに専用フォーマットのCSVファイルにデータを入力し、そのファイルをデータ管理画面からアップロードしてシステムのデータベースにデータを登録する。登録データは一覧画面で閲覧できる。登録データを修正する場合には、一覧画面からデータをCSVファイルに出力し、これを編集後に再びアップロードしてデータベースに修正を反映する。

系統情報には、系統ID、栽植年度、系統名、系統番号、親個体ID、播種日があり、田植えの時期には全項目のデータが発生する。形質情報には、個体ID、系統ID、20以上の形質項目がある。形質項目には、収穫時期までにデータが得られる項目(出穂日、穂長、粒型など)と、収穫後に個別の実験を行うなどして調査する項目(耐冷性、耐乾燥性など)に分類できる。そこで、系統情報と形質情報をそれぞれ専用フォーマットのCSVファイルで登録する。また、形質情報のうち、収穫時にデータが得られない項目に対応するために、形質項目単位でデータを追加登録できるファイルを提供する。これにより、各データの発生時期にあわせて登録作業が行えるだけでなく、形質項目単位でファイルの入力作業を複数人で分担することも可能となる。

データベースに登録したデータは、それぞれのデータ一覧インタフェースで、栽植年度や系統名などの検索条件を指定し該当データを表形式で閲覧できる。また、必要に応じて検索条件に合致するデータのみをCSVファイルに出力できる。登録データの編集は、この出力ファイルのデータを修正または追加し、これをアップロードすることで行う。この出力ファイルの各レコードには、レコードのデータから計算されたハッシュ値が追加されており、これを用いて不用意な誤修正を防ぐことができる。

### 4 トランスポゾンディスプレイ解析の支援

#### 4.1 概要

本章では、トランスポゾンディスプレイ解析で遺伝子の変異箇所を特定するための、波形データの比較支援インタフェースについて説明する。波形データの比較を効率的に行うためには、2つの波形を的確に重ね合わせることが必要となる。しかし、実験時の室温などの影響で流速が異なる場合には波形がずれるため、そのままでは重ね合わせることがで

きない。

そこで本システムでは、2つの波形データをそれぞれ補正して重ね合わせた画像を生成し、さらに、変異箇所の候補を表示することで、比較作業を支援する。また、波形データの重ね合わせ画像と変異箇所候補のリストをもとに、各候補が実際に変異であるかどうかを専門家が確認し、チェックボックスを用いて選択する簡単な操作のみで、正しい変異情報をデータベースに登録できるインタフェースを提供する。

本章では、トランスポゾンディスプレイ解析支援のための一連のインタフェースと、波形データの重ね合わせおよび変異箇所候補の検出のための2つのアルゴリズムを説明する。

## 4.2 トランスポゾンディスプレイ解析の支援インタフェース

トランスポゾンディスプレイ解析をシステム上で行うインタフェースについて説明する。トランスポゾンディスプレイ法による波形データを本システムのデータベースに登録してから変異情報を比較するまでの流れを4つの段階に分けて説明する。

### i) 波形データの登録

登録方法には2種類ある。ひとつはシステム画面上のフォームからファイルを個別にアップロードする方法、もうひとつはFTPを用いて所定ディレクトリに複数ファイルを一括アップロードする方法である。アップロードされた波形データは、個体IDと関連付けて保存される。登録済みの波形データ情報は、一覧画面で確認できる。

### ii) 波形データの重ね合わせ画像の表示

変異比較インタフェースでは、波形データが登録されている個体IDリストから、比較の基準となる参照個体と比較対象の個体とを指定し、これら2つの波形データを重ね合わせたグラフ画像を表示する(図3)。ここで参照個体とは、比較グループ中の全個体を比較

するための基準となるデータであり、比較グループ中で1個体が選ばれる。図3は、2つの波形データを補正して重ね合わせたグラフであり、参照個体と対象個体の波形データをそれぞれ赤と青で表示する。グラフの縦軸は蛍光強度の測定値であるが、2つのデータで値のスケールが異なるため、それぞれ左端と右端に異なる目盛りを表示することで、スケールを合わせている。横軸は断片長を表すEST Size(bp)である。横軸は元データでは泳動時間であるが、4.3節で説明する補正アルゴリズムによりEST Sizeに変換することで、波形間の横軸のずれを吸収し、重ね合わせを実現している。

### iii) 変異情報の登録

変異箇所候補のそれぞれを重ね合わせた波形上で確認し、正しい変異情報をシステムに登録する。変異箇所の候補は、波形データの重ね合わせ時に自動計算する。図3のグラフ中の2本の縦線は変異箇所候補を表し、傍らの数値はそのEST Sizeである。EST Sizeの後ろの記号は変異の種類を表し、「+」は対象個体のみが存在するピークであり、「-」は参照個体のみが存在するピークであることを意味する。ユーザは、候補に挙がっていない変異箇所を手動で追加でき、また、誤検出された変異を削除することができる。これらは、ii)で表示したグラフ画像と変異箇所候補のリストを利用して変異比較画面で行う。

### iv) 登録した変異情報の調整

登録された変異情報を系統などのまとまりで一覧し、変異箇所のずれを調整する。iii)で特定した変異箇所は、図4のような表で一覧できる。この表は、ある比較グループにおいて、各個体を参照個体と比較したときの、変異箇所の一覧である。表の1列目は各個体のID、1行目の数値は変異箇所として登録されたEST Sizeを示している。表中の「+」または「-」の記号は変異の種類を示しており、意味はiii)で説明した通りである。この一覧表示は図3の比較とは異なり、多くの個体の変異

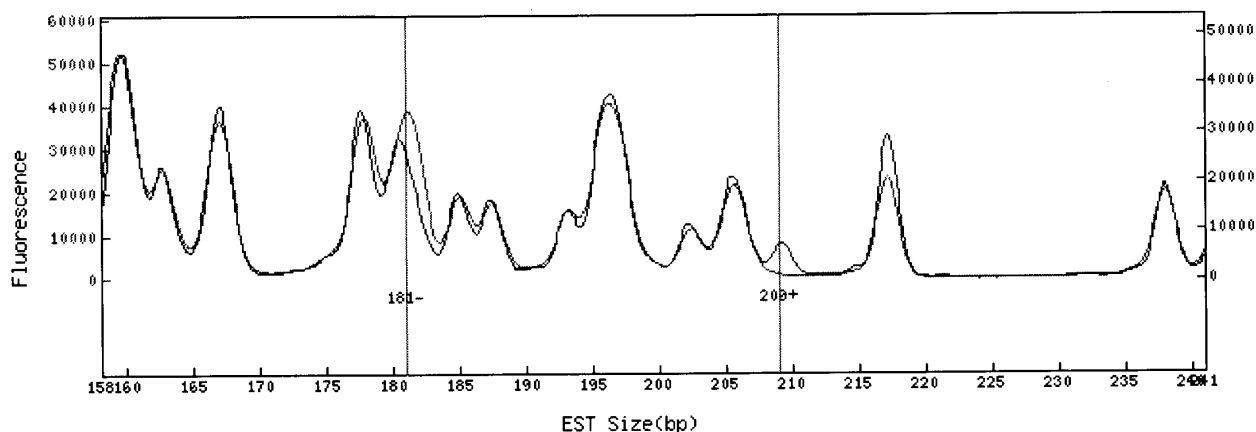


図3 波形データを重ね合わせ、変異箇所候補を表示した例

| individual ID   | 106 | 107 | 117 | 228 | 232 | 285 | 290 | 356 | 362 | 391 | 392 | 413 | 419 | 458 | 478 | 486 | 492 | 502 | 514 | 535 | 558 | 620 | 622 |   |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|
| 2004-MLR009-002 |     |     |     |     |     | -   | +   | +   |     | -   |     |     | -   |     |     |     | +   |     |     | -   |     |     |     |   |
| 2004-MLR009-003 |     |     |     |     |     |     | +   |     |     | -   |     |     | -   |     |     |     |     |     |     |     |     | +   | +   |   |
| 2004-MLR009-004 |     |     |     |     |     |     | +   |     |     | -   |     |     | -   |     |     |     |     |     |     |     |     |     |     | + |
| 2004-MLR009-005 |     |     |     |     | +   |     | +   |     |     | -   |     |     | -   | +   |     |     |     |     |     | -   | -   |     |     |   |
| 2004-MLR009-006 |     | +   | +   | +   |     |     | +   | +   |     |     |     |     | -   |     |     |     |     |     |     | -   | -   |     |     |   |
| 2004-MLR009-007 | -   |     |     |     |     |     | +   |     | -   | -   |     |     | -   |     | -   |     |     |     |     |     |     |     |     |   |
| 2004-MLR009-008 | -   |     |     |     |     |     | +   |     | -   | -   |     |     | -   |     |     |     |     |     |     |     | -   |     |     |   |
| 2004-MLR009-009 | -   |     |     |     |     |     | +   |     | -   | -   |     | +   | -   | +   |     |     |     | +   |     |     |     |     |     |   |
| 2004-MLR009-010 | -   |     |     |     |     |     | +   |     |     | -   |     |     | -   |     | -   | -   |     |     |     |     | -   |     |     |   |

図4 変異による遺伝子多型の登録データ一覧表示

を同時に見比べることができるが、図3では同時に2つの波形しか比較できないため、同じ比較グループ内でも、同一の変異(同じトランスポゾン転移による変化)を別のEST Sizeとして登録してしまうことがある。例えば、300 bp 近辺に1つの変異があった場合に、これを299 bpとして登録した個体と301 bpとして登録した個体があった場合、一覧表においては、これらは別の変異断片として扱われる。これを補正するために、ユーザが同一の変異であると判断した変異を特定のEST Sizeに統合したり、統合した変異を元に戻したりといった微調整が必要となる。図4の例では、グレーになっているEST Size「300」は、前後数 bp の範囲で登録された変異情報を同一のものと判断し、例えば299 bp や302 bp といった変異も表示上は300 bp の変異情報として表示されるように統合したことを表す。

### 4.3 波形の重ね合わせと変異箇所の検出アルゴリズム

本システムでは、波形データの比較作業を簡単化するために、比較する波形データの重ね合わせと変異箇所候補の検出の2つの処理を自動化する。波形データを補正し適切に重ね合わせることで、比較作業の効率が大きく向上するだけでなく、変異検出の精度を上げることができる。また、変異箇所の候補を示すことで変異の検出が容易になり、比較作業に要する時間を短縮できる。

本節では、これら2つの自動化処理の手法を説明する。ここでは紙面の都合上、基本原理のみを述べる。実用上問題のない程度まで精度を向上するために、我々はさらにいくつかの工夫を行っている。

#### 4.3.1 波形データの重ね合わせ

波形データは一定のサンプリング時間ごとに測定した蛍光強度のリストファイルである。しかし、サンプリング時間が一定であっても、電気泳動時の室温などの実験条件による流速の差や泳動後半の流速の減衰などの影響を含むため、生データを時間軸のままそのまま重ね合わせることができない。

そこで、波形データを補正し、サンプリング点の計測時間を流速に無関係な断片長 (EST size, 単位は bp) に変換する。これには、サンプルに混合して電気泳動する DNA サイズマーカーの測定データを利用する。図 2 の灰色の波形で示したものが DNA サイズマーカーのデータであり、一定間隔ごとにピークが出るように調整されている。このマーカーデータから、断片長既知のマーカーに対応するピークのサンプリング点を検出し、各マーカーの区間ごとに試料のデータリストのサンプリング点を断片長に変換する。

まず、マーカーピークのサンプリング点の検出手順を述べる。マーカーのデータリストを調べ、蛍光強度の測定値が閾値以上 (この閾値は実際のデータから計算した値である) である区間を検出し、その区間内で値が最大となるサンプリング点をマーカーピークの検出点とする。マーカーはピークの蛍光強度がほぼ一定となるよう調整されているが、必ずしも理想的なマーカーデータが検出できるわけではない。そこで、マーカーピークの間隔などを用いてマーカーピークの検出漏れや誤検出がないかをチェックし、最終的なマーカーのピークリストを作成する。

次に、マーカーのピークリストから、試料の波形データを補正する手順を述べる。データリストの泳動開始から終了までのサンプリング点を  $i$  ( $i = 1, 2, 3, \dots, n$ ),  $j$  番目のマーカーピークの断片長を  $l_j$  ( $j = 1, 2, 3, \dots, m$ ), そのサンプリング点を  $m_j$  とすると、サンプリング点  $i$  の断片長  $X_i$  は次式によって変換できる。

$$X_i = l_j + \frac{l_{j+1} - l_j}{m_{j+1} - m_j} \times (i - m_j) \quad (1)$$

ただし、 $m_j \leq i < m_{j+1}$  である。

マーカーピークの断片長は全ての波形データで共通であるから、これを利用して変換したデータリストは、横軸を断片長、縦軸を蛍光強度とするグラフとして、重ね合わせて描画できる (図 2)。縦軸の蛍光強度のスケールは波形データごとに大きく異なる場合があるが、これは描画する際に縦軸スケールを波形ごとに別々に設定することで対応する。

#### 4.3.2 変異箇所候補の検出

変異箇所を検出するには、重ね合わせた複雑な波形データを注意深く見比べる必要がある。これを支援するために、重ね合わせた波形から変異箇所の候補を自動計算してハイライト表示する。自動検出した候補が実際に変異箇所であるかどうかは、研究者が確認して判断する必要があるが、この仕組みにより変異箇所の検出作業を効率化できる。

変異箇所候補の検出は、比較する 2 つの波形データからそれぞれピークリストを作成し、これをマッチングすることにより行う。ピークリスト作成の前処理として、サンプリング点を断片長に変換した補正済みの波形データに対し、断片長  $\pm 0.5$  bp の範囲で移動平均をとる。0.5 bp としたのは、断片長は整数値をとることから、ピーク間隔は最小で 1 bp と考えられるためである。これによりノイズ成分による影響を低減する。次に、この波形の傾きが + から - に変化する点をピークとして検出し、ピークリストを作成する。最後に、それぞれの波形データのピークリストをマッチングし、マッチングに漏れたピークを変異箇所候補として検出する。

ピークのマッチングは、次のように行う。2 つのピークリストを A, B とすると、A の各ピークに対して、B のピークリスト中で最も断片長に近いピークをマッチング候補とする。同様に、B の各ピークに対して、A のピークリスト中で最も断片長に近いものをマッチング候補とする。そして、A と B のピークでお互いをマッチング候補としている組を見つけ、これをマッチングペアする。最終的に、ペアに

アグリバイオ >> イネ >> データ解析支援システム
形質・変異検定

形質・変異一覧 **形質・変異検定**

栽培年度: ALL | 系統名: ALL | 系統番号: ALL |  | グループ名: 2004-XXX001

**形質**

検定形質: 脱粒性

粒型: ALL

**変異塩基**

タイプ組合せ:  (0,+)  (0,-)  (+,-)

個体数: 5 個以上の変異タイプのみ検定

検定方法:  | 閾値:  $p \leq 0.10$  |

グループ名: 2004-XXX001  
 基系統: 2004-XXX001  
 参照個体: 2004-XXX001-001  
 総個体数: 40

検定方法:  $\chi^2$ 検定  
 帰無仮説: 変異タイプ間の脱粒性値分布に関連は無い

| 変異塩基 | 個体数 | 脱粒性 |   |   |   |   |   |   | p値▲          | 備考                               |
|------|-----|-----|---|---|---|---|---|---|--------------|----------------------------------|
|      |     | 0   | 1 | 2 | 3 | 4 | 5 |   |              |                                  |
| 270  | 0   | 20  | 4 | 5 | 5 | 5 | 0 | 1 | 0.0093166298 |                                  |
|      | +   | 20  | 0 | 3 | 1 | 6 | 7 | 3 |              |                                  |
| 312  | 0   | 25  | 1 | 5 | 3 | 9 | 4 | 3 | 0.0739701861 | 2004-XXX001-009にて複数変異を統合した±変異を使用 |
|      | +   | 10  | 3 | 2 | 2 | 0 | 3 | 0 |              |                                  |
| 180  | 0   | 19  | 0 | 5 | 4 | 6 | 1 | 3 | 0.0825321793 |                                  |
|      | -   | 21  | 4 | 3 | 2 | 5 | 6 | 1 |              |                                  |

}

解析方法の選択

}

解析結果の表示

図 5 相関解析インタフェース

ならなかったピークを変異箇所の候補とする。

このようにすると、それなりにうまく変異箇所候補の検出ができるが、かなり高さの違うピーク同士をマッチングペアと判定してしまうことがある。ピークの高さが著しく異なる場合には、そのピークに、長さは同じだが異なる断片が混在している可能性があり、その場合は変異として検出すべきである。このため、高さの比がある閾値以上であれば、マッチしたピークであっても変異箇所候補として検出することとした。なお、この閾値は、すべてのマッチングペアについて、Aのピークに対するBのピークの高さの比の平均および標準偏差から、有意水準5%の両側検定により求めた値である。

## 5 表現型と遺伝子型の相関解析

データベースに登録された変異情報(遺伝子型)と形質情報(表現型)を使い、遺伝子の機能推定のための相関解析を行う。本システムの相関解析インタフェース(図5)では、上部で解析対象の変異体グループ、検定する形質項目、検定方法、および有意水準などを指定して「検定結果の表示」ボタンを押すことで、指定された形質と登録されている変異情報の間で相関解析を行い、画面下部に結果を表示する。これにより、特定の形質の発現に関与している遺伝子周辺の断片を簡単にリストアップできる。

本システムの相関解析では、各変異の有無で個体を2群に分けてそれぞれの形質情報との共分離の度合いを評価する。具体的には、

指定されたグループに属する個体を，検定対象となる変異の登録の有無で個体を2群に分け，相関解析の対象である形質項目の測定値の平均値に有意差があるかどうかを検定する．評価方法には，仮説検定 ( $\chi^2$  検定と  $t$  検定) を用いる．検定結果は， $p$  値が小さいものから順にユーザの指定した有意水準を満たすものを出力する．このインタフェースにより，各形質値に相関が認められる変異の Est Size を容易に特定することができる．

登録データの傾向を概観するために，変異情報と形質情報をまとめて一覧するインタフェースも提供する．このインタフェースでは，指定されたグループごとに変異情報と形質情報とをまとめて表示し，複数項目をキーとしての多段階ソートに対応する．また，系統情報や形質情報の一覧インタフェースと同様に，画面に表示したデータを CSV ファイルへ出力することもできる．

## 6 評価

本章では，トランスポゾンディスプレイ解析支援のための「波形データの重ね合わせ」と「変異箇所候補の検出」の2つのアルゴリズムの有効性の検証と，我々の構築したデータベースシステム全体の評価，および本システムを利用して得られた成果についてまとめる．

波形データの重ね合わせアルゴリズムは，実用上問題のない精度で動作している．我々は，約400の波形データを対象に，時系列サンプリング点を断片長 (bp) に補正する実験を行ったところ，約20の少数例を除いて，問題なく動作した．補正に失敗した事例はすべて，実験ミスによりマーカーピークが正しく測定されておらず，マーカーピークのサンプリング点がコンピュータ上で正しく認識できないことが原因であった．しかし，このような事例は，目視でも電気泳動の失敗が明らかであり，再実験によりデータを取り直すことになるため，実用上は問題とならない．なお，実験の失敗率は，研究者の実験スキルの向上により改善されると予想される．

変異箇所候補の検出アルゴリズムの評価には，システムによって自動検出した変異箇所候補のデータと，研究者が実際に変異箇所であると判断したデータとを比較した．125個のデータについて比較した結果，A. 研究者が変異箇所であると判断した断片長を自動検出できる割合は約75%であるが，B. 自動検出した変異箇所候補のうち実際に変異であると判断されたもの (自動検出の正解率) は約23%程度であり，誤検出が多いことがわかった．なお，変異箇所候補を決定するためのピークの断片量に基づく処理を行うことにより，自動検出割合 (A) は大きく変化しなかったものの，自動検出の正解率 (B) は，約23%から約41%への大幅な向上が見られている．

本システムを実際に研究者に使用してもらいアンケート評価したところ，正確な定量は難しいものの，従来のトランスポゾンディスプレイ解析に要した時間の数分の一程度の時間で変異箇所の検出が可能となるとの評価を得た．また，効率化に貢献した機能について，波形データを重ね合わせた画像と変異箇所候補のリストを用いた変異検出が非常に効率的であるとのコメントが得られた．

なお，堀端ら (2005) は，我々のデータベースシステムを利用し，トランスポゾンディスプレイ解析で検出した変異と粒型および脱粒性変異とを分析し，粒型変異に関与していると思われる DNA 断片を特定した<sup>[5][6]</sup>．

## 7 おわりに

我々の構築したデータベースシステムでは，*mPing* により効率よく作成した大規模な変異体シリーズを利用して変異による遺伝子多型と表現型との関係を研究するために，ハイスループットなデータ処理の実現を目的とした．従来の進め方でボトルネックの一つとなっていたトランスポゾンディスプレイ解析にかかる時間と労力を軽減するために，比較する波形データの重ね合わせや多型候補の算出をシステム上で自動化し，これにより解析に要する時間を従来の数分の一に短縮した．実験で

得られるデータを従来の個別ファイルでの管理からシステムでデータベース化することにより、効率のよいデータの利用と複数人の研究者による一貫したデータの管理を実現した。また、実利用評価によりその有効性を検証し、本システムを用いた生物学的成果を得ることができた。今後の課題としては、本システムの他生物種への適用と変異の自動検出アルゴリズムの精度のさらなる向上が挙げられる。

## 謝辞

本研究は科学技術振興機構、和歌山県地域結集型共同研究事業費により実施した。

## 参考文献

- [1] T. Nakazaki; Y. Okumoto; A. Horibata; S. Yamahira; M. Teraishi; H. Nishida; H. Inoue; T. Tanisaka: "Mobilization of Transposon in the rice genome", *Nature*, 421, pp.170-172, 2003.
- [2] S. Ayyadevara; J. J. Thaden; R. J. Shmookler Reis: "Anchor Polymerase Chain Reaction Display: A High-Throughput Method to Resolve, Score, and Isolate Dimorphic Genetic Markers Based on Interspersed Repetitive DNA Elements", *Analytical Biochemistry*, 284, pp.19-28, 2000.
- [3] Vos, P.; R. Hogers; M. Bleeker; M. Reijans; T. Van de Lee; M. Hornes; A. Frijters; J. Pot; J. Peleman; M. Kuiper: "AFLP: a new technique for DNA fingerprinting", *Nucleic Acids Res.*, 23, pp.4407-4414, 1995.
- [4] DNA Data Bank of Japan (DDBJ) <http://www.ddbj.nig.ac.jp/>
- [5] A. Horibata; K. Matsui; E. Inoue; T. Yoshihiro; H. Kawaji; M. Nakagawa; Y. Okumoto; N. Nakazaki; T. Tanisaka: "Spontaneous and Frequent Transposition of a Miniature Inverted-Repeat Transposable Element, *mPing*, in an Experimental Line of Rice (*Oryza sativa* L.)", *The Society for the Advancement of Breeding Researches in Asia and Oceania (SABRAO)*, 2005.
- [6] 堀端章; 松井和幸; 井上悦子; 吉廣卓哉; 川路英哉; 中川優; 奥本裕; 中崎鉄也; 谷坂隆俊: 「誘発突然変異遺伝子と共分散する *mPing* のトランスポゾンディスプレイによる同定」, 日本育種学会第 107・108 回講演会, 2005.
- [7] E. Inoue; T. Yoshihiro; H. Kawaji; A. Horibata; M. Nakagawa: "A Database System for High-Throughput Transposon Display Analyses of Rice," *BIOINFO* 2005, 2005.

(2005 年 11 月 14 日受付)

(2006 年 1 月 6 日採択)

**宇部興産 (株) におけるエンドユーザー教育  
— 営業部門に対する知的財産情報教育の試み —  
End User Education at Ube Industries, Ltd.  
— Trial Education of Intellectual Property Information  
for Business Section —**

出口 昌信<sup>\*,†</sup> 岡本 和彦<sup>\*,‡</sup>

**Masanobu DEGUCHI and Kazuhiko OKAMOTO**

宇部興産 (株) のグループ企業全体に対する知的財産情報教育の一環として、日本特許情報については、エンドユーザー向け特許情報検索システムをグループ企業内で広く普及させ、利用している。これまでの、知的財産情報教育は、研究者、技術者へのエンドユーザー教育を主体に行ってきたが、研究者、技術者の営業部門への異動などに伴い、営業部門でもエンドユーザー向け日本特許情報データベースを利用、活用することが増えている。これらの諸事情を踏まえて、同社で初めての試みとして、東京本社の営業部門において、知的財産情報教育を実施した。具体的には、特許庁 IPDL、エンドユーザー向け特許情報検索システムである NRI サイバーパテントの効果的利用方法を主体に教育を行った。

We propagate use of Japanese patent information retrieval system for an End User to Ube Industries and its group companies. Also, intellectual property information education for End User, such as researcher, engineer has been done. But recently, some parts of such End Users are transferred to the sales department of Tokyo head office. Gradually, the needs of Japanese patent information database are increased in sales department. Thinking of all these things, we carried out intellectual property information education as the first trial or the case study in sales department of the Tokyo head office. We educate effective usage of the NRI cyber patent and IPDL of Japan Patent Office, both of them are Japanese patent information retrieval system that designed for End Users.

キーワード：エンドユーザー教育，知的財産情報教育，営業部門

End User education, Intellectual property information education, Business section

## 1 はじめに

宇部興産 (株) では、グループ企業全体に対する知的財産情報教育の一環として、日本特許情報については、エンドユーザー向け特許情報検索システムをグループ企業内で広く普及させ、利用している。これまでの、知的財産情報教育は、研究者、技術者へのエンドユーザー教育を主体に行ってきたが、研究者、技

術者の営業部門への異動などに伴い、営業部門でもエンドユーザー向け日本特許情報データベースを利用、活用することが増えている。

これらの諸事情を踏まえて、同社で初めての試みとして、2004年10月下旬に、東京本社の営業部門において、知的財産情報教育を実施した。具体的には、特許庁 IPDL、エンドユーザー向け特許情報検索システムである NRI サイバーパテントの効果的利用方法を主体に教育を行った。

本報告では、弊社東京本社の営業部門に対して行った、知的財産情報教育の詳細な内容及び、営業部門の受講生の反応等について報告する。

\* 宇部興産 (株) 研究開発本部知的財産部  
Intellectual Property Department, Corporate Research & Development Ube Industries, Ltd.

† 29740u@ube-ind.co.jp

‡ 27124u@ube-ind.co.jp



## 2 教育内容 (1)・NRI サイバーパテント

### (1) NRI サイバーパテント・番号照会検索の方法

公開番号、特許番号などからの検索により、必要な特許明細書の入手方法を教えた。

### (2) NRI サイバーパテント・概念検索の方法

NRI サイバーパテントの特徴として「概念検索」機能がある。この概念検索とは、検索時に指定された文章と、特許公報の要約もしくは特許請求の範囲の請求項を比較し、文章に記載された内容、意味を比較し、類似性、関連度合いの高い公報から順番にランキングし、表示する検索システムである。

その際、関連度合いの最も高い公報のスコア(最高スコア)と、指定された(例えば100件)件数目の公報のスコア(最低スコア)とが表示される。概念検索による検索結果にはヒット件数という概念がなく、スコア値を参考に一覧表示する行数を指定することになる。

概念検索の対象は要約本文や特許請求の範囲の請求項本文または抄録であるため、出願人等を入力しても検索されない。

(以上 NRI サイバーパテント HP より引用)。

上記概念検索システムは、特許情報検索システムの知識、情報検索の知識、国際特許分類(IPC)などの、特許情報特有のインデックス、検索タームの知識等を知らなくても、特許情報検索を行うことが可能である。知的財産情報の予備知識の無い研究者・技術者であっても気楽に検索することができる。研究開発活動、実験の合間に、アイデア、イメージが湧いたときに、そのアイデア、イメージを検索画面から、数行程度の文章にして入力することで、特許情報検索をすることが可能である。

### (3) NRI サイバーパテント・複合検索

NRI サイバーパテントの複合検索画面を利用し、各検索タームを利用しての検索、検索内容の掛け合わせなどについて、実例を見せながら、説明していった。ここで教育した検索タームとは、以下のものである。

#### ①キーワード

検索対象範囲を特許明細書全文にする場合と、特許請求の範囲に限定する場合の2種類の使い分けについて、実例を挙げて説明した。また、NRI サイバーパテントの場合、キーワード、言葉による検索は完全一致検索である。言語表現が違えば、同義語はヒットしないことも明言した。PATOLIS-IVやJOISのような情報検索システムの場合、シソーラスが完備され、統制語が存在するが、NRI サイバーパテントの場合は違う検索システムであることを理解してもらうよう注力した。

#### ②IPC(国際特許分類)、

#### ③FI(特許庁ファイルインデックス)、

#### ④FT(Fターム)、

#### ⑤出願人・権利者

出願人・権利者の場合、完全一致検索ではなく、前方一致検索で行うことを基本として教えた。検索漏れを少なくするためである。また、企業同士の合併があった場合は、合併前の企業名+合併後の企業名を全て入力し、合算する必要があることも説明した。さらに、公表公報などでみられるように、外国企業の場合はカタカナ表記になるが、日本国内の代理人の翻訳・表記の仕方でヴァリエーションが発生していることを説明した。可能な限りカタカナ表記のヴァリエーションを把握し、入力し、合算することを徹底するよう、強調して説明した。

#### ⑥発明者

発明者からの検索方法も説明した。留意事項としては、ベンチャー企業のような場合、企業名での出願ではなく、社長が個人名で出願している場合があることも説明しておいた。

### (4) NRI サイバーパテント・SDI 検索

研究開発テーマ、営業戦略テーマについて、検索式を設定し、週2回のペースで新着特許情報が自動配信できるが、その設定方法についても説明した。その際、早期審査特許案件、国際出願特許案件が漏れないよう、特許公報には5種類あり、その全ての特許公報をSDI検索の対象とすることの必要性を強調した。

### (5) NRI サイバーパテント・経過情報表示(フロー形式)

NRI サイバーパテントの特徴として、出願から審査・審判・登録に至る経過情報あるいは、法的状況(リーガルステータス)をフロー形式で表示することにより、特許制度に不慣れな人にも理解しやすくした機能がある。この機能についても説明し、容易に経過情報、法的状況を確認することができることを強調した。

### 3 教育内容(2)・特許庁 IPDL

(1) IPDL の効果的な利用方法(どこを使えば便利か?)

NRI サイバーパテントだけではなく、特許庁 IPDL の利用方法も解説した。NRI サイバーパテントに比較し、特許庁 IPDL はキーワードからの検索には不便である。

しかし、IPC(国際特許分類)、FI(特許庁ファイルインデックス)、FT(F ターム)からの検索には有効である。また、パテントマップガイダンスを利用すれば、上記、IPC、FI、FTの内容を調べたい場合には非常に有効である。実習では、NRI サイバーパテントとは別にブラウザを立ち上げ、特許庁 IPDL のパテントマップガイダンス画面から IPC、FI、FT などの各検索タームを解析しながら、NRI サイバーパテントを利用する方法を説明し、実演して見せた。

さらに、IPDL の初心者検索画面での、キーワードからの特許・実用新案検索、具体的商品名、役務名からの簡易商標検索も、それらの利用方法を解説した。

### 4 受講生の反応

紙面の都合や諸事情で詳細な結果を記載できないが、アンケート結果の一部とそれに対する受講生の回答の概要について紹介する。

- Q1. 説明会のどの部分が興味深かったですか。または、さらに詳細な説明を聞きたい項目などありましたらご記入ください。
- A1. 概念検索が利用しやすく、興味深いので、

是非利用したい。

- A2. 審査経過情報表示など、未知の機能がわかりよかった。
- Q2. 本日の説明会についてのご意見、ご感想などお気づきになったことをご記入ください。
- A1. 初めて(知的財産情報教育を)受けたが、是非(営業)業務に利用したい。
- A2. もう少し、上級レベルの講習会も開催して欲しい。
- Q3. 営業部門への知的財産教育、知的財産情報教育についてご意見をお願いします。
- A1. もっと多くの社員に知的財産の重要性が理解できるよう教育してほしい。
- Q4. 知財部に対して今後おこなってほしい教育、説明会等がございましたらご記入ください。
- A1. パテントマップの作成方法、情報解析について教育して欲しい。

## 5 結論

### 5.1 アンケート結果に対する考察

今回、東京本社の営業部門のスタッフに対する知的財産情報教育は、初めての試みである。しかし、上記のようなアンケート結果が得られたということは、以下のことが考えられる。

営業部門のスタッフも、研究所、事業部などから営業部門に異動した技術営業の人間が多く、知的財産制度、情報検索について、既に、ある程度の予備知識を有している者も多い。宇部興産(株)において、エンドユーザー教育を実施した際、情報検索アドバイザー側からの意見として、特許教育、特許明細書作成教育の定期的な開催の要望があった。その理由として、「特許明細書作成のポイントを理解していなければ、特許調査においても、そのポイントが理解できない。目的に合致した的確な特許調査ができないため」、ということがあげられた。

上記内容と同じく、営業部門のスタッフも、

単なる、情報検索のみの教育ではなく、知的財産法も組み合わせた教育の実施を求めていると思われる。

営業部門スタッフの、知的財産法の知識の必要性については、以下の理由がある。

第一に、自社製品及び技術の営業活動を行うにあたり、自社技術が特許によって保護され、信頼性を有しているか、という点について、顧客(ユーザー)から徹底して質問されるからである。

具体的には、自社と競合他社との特許権の関係について、権利侵害を起こしていない、もしくは起こす心配がないことを顧客に対し説明する必要があるからである。これは自社技術の信頼性の問題としてとらえることもできる。

次に、特許権のクロスライセンス、権利譲渡、実施許諾等のライセンス交渉に発展した場合、営業部門だけで判断することなく、必ず、知的財産部、法務部と相談のうえ、経営上の判断をしなくてはならないこともある。

また、知的財産法の正しい知識を営業スタッフが持っていない場合の弊害として以下のリスクがある。

- (1) 自社特許の裏付けのない技術、言い換えれば、自社が特許権を有していない技術範囲をカタログ、社外向けHP等に記載することによって発生する権利侵害行為及び権利侵害訴訟のリスクがある。さらに、
- (2) 他社のカタログデザイン等を模倣する事による著作権法又は不正競争防止法上の違法行為及び訴訟のリスクである。

むしろ、これらのリスク回避のために、営業スタッフも知的財産法を知っておくことが重要である。

さらに、顧客(ユーザー)の出願している用途特許を営業部門スタッフが把握し、理解していることは、顧客(ユーザー)ニーズの早期把握を促進することにつながり、ビジネスチャンスを増やすことになる。加えて、新たな顧客(ユーザー)の開拓もしくは獲得において有効な情報源になることが期待できるのである。

## 5.2 注目点

知財部に対して今後おこなってほしい教育、説明会等として、パテントマップの作成方法、情報解析について教育して欲しい、という意見が出されている。パテントマップによる特許情報の分析、加工、解析は一般的に行なわれていることではあるが、やはり、それを知的財産戦略というレベルで活用することが求められていることの現れであろう。営業部門においては、当然、マーケットリサーチを実施し、商品企画も行うが、その実行可能性を考えれば、知的財産情報解析、知的財産戦略と組み合わせ、実現可能性を高めておくことが必要である。このような認識が具体的な要望として出されていると解される。

## 謝辞

本報告投稿に当り、許可をご快諾いただいた宇部興産(株)の関係諸氏に対し、この紙面を借りて、深く感謝する次第である。

## 参考文献

- [1] 慈道佐代子：「利用者向け研修プロジェクト－企画と実施例－」, 情報の科学と技術, Vol.53, No.5, pp.263-268, 2003.
- [2] 情報科学技術協会セミナー(エンドユーザー検索－企業内教育の取り組み 2003年9月26日開催)資料, 2003.
- [3] 岡本和彦：「宇部興産(株)におけるエンドユーザー教育：研究開発本部における『情報検索アドバイザー制度』」, 第40回情報科学研究集会予稿集, pp.91-94, 2003.
- [4] 岡本和彦：「宇部興産(株)におけるエンドユーザー教育：研究開発本部における『情報検索アドバイザー制度』」, 情報管理, Vol.47, No.1, pp.15-19, 2004.
- [5] 特集：特許検索に必要なスキルと知識, 情報の科学と技術, Vol.54, No.5, 2004.
- [6] 岡本和彦：「宇部興産(株)におけるエンドユーザー教育：研究開発本部における『情

- 報検索アドバイザー制度』, 日本知財学会第二回研究発表会講演要旨集, pp.362-365, 2004.
- [7] 岡本和彦:「全国図書館大会への招待, 第6分科会(専門図書館), 専門図書館の人材育成, 宇部興産(株)におけるエンドユーザー教育:研究開発本部における『情報検索アドバイザー制度』, 図書館雑誌, Vol.98, No.9. pp.645, 2004.
- [8] 岡本和彦;出口昌信:「宇部興産(株)におけるエンドユーザー教育Ⅱ:グループ企業内におけるPATOLIS-Ⅳの導入教育」, 第1回情報プロフェッショナルシンポジウム(INFOPRO 2004)予稿集, pp.99-102, 2004.
- [9] 岡本和彦:「宇部興産(株)におけるエンドユーザー教育:研究開発本部における『情報検索アドバイザー制度』及びインフォプロのスキルアップ(私見)」, 平成16年度(第90回)全国図書館大会要綱, pp.45-46, 2004.
- [10] 岡本和彦;出口昌信:「山口大学に対する知的財産情報教育支援～産学連携における試み～」, 第11回医学図書館研究会・継続教育コース予稿集, 2004.
- [11] 岡本和彦:「宇部興産(株)におけるエンドユーザー教育:研究開発本部における『情報検索アドバイザー制度』, 2005年情報学シンポジウム講演論文集, pp.23-27, 2005.
- [12] 岡本和彦:「宇部興産(株)におけるエンドユーザー教育:研究開発本部における『情報検索アドバイザー制度』及びインフォプロのスキルアップ(私見)」, 平成16年度(第90回)全国図書館大会香川大会記録, pp. 242-251, 2005.
- [13] 岡本和彦;出口昌信:「山口TLOに対する知的財産情報教育支援-産学連携における試みⅡ-」, 日本知財学会第三回学術研究発表会講演要旨集, pp.268-271, 2005.
- [14] 出口昌信;岡本和彦:「宇部興産(株)におけるエンドユーザー教育:営業部門に対する知的財産情報教育の試み」, 日本知財学会第三回学術研究発表会講演要旨集, pp.280-283, 2005.

(2005年10月7日受付)

(2005年12月15日採択)

## 山口 TLO に対する知的財産情報教育支援

## — 産学連携における試み —

Support of Intellectual Property Information Education  
for Yamaguchi TLO

## — Trial for academia industry cooperation —

岡本 和彦<sup>\*,†</sup> 出口 昌信<sup>\*,‡</sup>

Kazuhiko OKAMOTO and Masanobu DEGUCHI

山口大学と宇部興産(株)の包括連携協定締結に基づき、産学連携活動の一環として、山口 TLO に対する知的財産情報教育支援を行った。特許庁 IPDL, NRI サイバーパテントを利用しての情報検索教育に加えて、以下の教育を行った。

1. 知的財産情報入門(特許情報の種類, 特許公報の読み方等の基礎知識)
2. 研究開発における特許調査の考え方(知的財産戦略)
3. 最終日の理解度テスト; 実技試験(これにより, 特許情報検索における, 各受講者の取り組み方, 考え方を見ることができた.)

受講生が理解し難かったところとしては, ①特許制度, ②特許関連専門用語, ③IPC(国際特許分類)の, 3点が挙げられた。OJT 教育受講生はこの後, 調査結果のパテントマップ化, 解析を行った。

Based on Yamaguchi University and the inclusion cooperation agreement conclusion of Ube Industries, Ltd., We helped intellectual property information education for Yamaguchi TLO as a part of academia industry cooperation. As well as information retrieval education that we used Japan Patent Office IPDL, NRI cyber patent, I educated the following.

1. An introduction to intellectual property information (a kind of patent information, basics knowledge such as how to read patent bulletin)
2. A way of thinking (an intellectual property strategy) of patent investigation in research and development
3. An understanding degree test of a last day; A practical skill examination (each this, student attending a lecture in patent information retrieval)

According to the results of questionnaire, it is difficult for students to understand IPC (international patent classification), patent system, technical terms of patent.

キーワード: 山口 TLO, 知的財産情報教育, 産学連携

Yamaguchi TLO, Intellectual property information education, Academia industry cooperation

## 1 はじめに

山口大学と宇部興産(株)の包括連携協定締結に基づき、産学連携活動の一環として、山

口 TLO に対する知的財産情報教育支援を行った。具体的には, 2004 年 9 月に, 1 コース(1 回 2 時間の講義+実習を計 3 回)を, 山口 TLO で OJT トレーニングを受ける, 山口大学工学部の学生(学部生, 大学院生計 8 名)に対して行った。特許庁 IPDL, NRI サイバーパテントを利用しての情報検索教育に加えて, 以下の教育を行った。

\* 宇部興産(株) 研究開発本部知的財産部  
Intellectual Property Department, Corporate Research & Development Ube Industries, Ltd.

† 27124u@ube-ind.co.jp

‡ 29740u@ube-ind.co.jp

1. 知的財産情報入門 (特許情報の種類, 特許公報の読み方等の基礎知識)
2. 研究開発における特許調査の考え方 (知的財産戦略)
3. 最終日の理解度テスト; 実技試験 (これにより, 特許情報検索における, 各受講者の取り組み方, 考え方を見ることができた.)

OJT 教育受講生はこの後, 調査結果の Patent マップ化, 解析を行った。

本論文では, 産学連携の一環として行われた, 山口 TLO に対する, 知的財産情報教育支援の詳細な内容及び, 受講生の反応等について報告する。

## 2 教育内容

教育目的が, 起業家育成訓練の一環としての, 特許調査であり, 調査結果に基づく分析・加工及び Patent マップ化につながる内容を必要としていた。この事情を踏まえ, 単なる, エンドユーザー向け日本特許情報検索システム (NRI サイバーパテント) のユーザートレーニングではなく, 必要な知的財産制度, データベース及びデジタルコンテンツの著作権も含めた内容を教育した。実際には, 前半の 1 時間を講義, 後半の 1 時間を情報検索実習とした。実習の際は, 山口大学知的財産本部スタッフ及び山口大学特許情報検索インストラクターが受講生の理解をサポートした。

### 2.1 第 1 回目

#### 2.1.1 知的財産情報入門 (特許情報の種類, 特許公報の読み方等の基礎知識)

情報学の意義 (情報学と情報工学の差異など), 情報の種類 (1 次情報, 2 次情報, 3 次情報), 上記各段階の情報に相当する知的財産情報の解説, 知的財産情報の種類

特許公報には 5 種類あることを説明し, 公開公報, 公表公報, 再公表公報, 特許公報, 公告公報の見方, 読み方等を説明した。同時に,

これら 5 種類の特許公報が発生する過程を理解してもらうために, 日本特許制度, 国際特許制度, 特に PCT 出願によって発生する公表公報, 再公表公報に関連づけて説明した。

さらに, 特許公報を読むうえで是非とも必要な知識である, IPC (国際特許分類), FI (特許庁ファイルインデックス), FT (F ターム) などについても, 具体的事例, サンプルを利用しながら解説した。これらの, 知的財産情報の基礎知識の部分は, 日常の知的財産部の業務においても, 研究者サイドから一番質問が多い個所である。

#### 2.1.2 特許番号等からの検索方法, 特許公報入手方法

前項で, 特許情報の種類, 特許公報の種類を解説してあることから, 各特許公報番号から, 実際の特許公報を検索する方法を教育した。研究者が特許公報の入手を頼んでくる場合, 出願番号, 公開番号, 特許番号の区別がつかないまま依頼をしてくる事例を多く経験していることからである。番号検索を教えた後は, 特許公報の入手方法も合わせて教育した。

#### 2.1.3 NRI サイバーパテント・概念検索の方法

NRI サイバーパテントの特徴として「概念検索」機能がある。この概念検索とは,

検索時に指定された文章と, 特許公報の要約もしくは特許請求の範囲の請求項を比較し, 文章に記載された内容, 意味を比較し, 類似性, 関連度合いの高い公報から順番にランキングし, 表示する検索システムである (以上 NRI サイバーパテント HP より引用)。

上記概念検索システムは, 特許情報検索システムの知識, 情報検索の知識, 国際特許分類 (IPC) などの, 特許情報特有のインデックス, 検索タームの知識等を知らなくても, 特許情報検索を行うことが可能である。

知的財産情報の予備知識の無い研究者・技術者であっても気楽に検索することができる。研究開発活動, 実験の合間に, アイデア, イメージが湧いたときに, そのアイデア, イメー

ジを検索画面から、数行程度の文章にして入力することで、特許情報検索をすることが可能である。

教育においては、1回目の教育の、後半の1時間を、概念検索の説明及び、端末実習にあて、実際に概念検索実習を体験してもらった。検索課題は任意の想定課題を与えることもしたが、できるかぎり、受講生が現在起業化を考えている研究テーマ、卒業論文、修士論文等の研究テーマでの検索を試みてもらうこととした。

## 2.2 第2回目

### 2.2.1 研究開発における特許調査の考え方 (知的財産戦略)

前半の1時間はPPTを利用した講義を行った。この教材は、本来、宇部興産(株)研究開発本部の研究所のグループリーダークラスを対象とした社内教育用に作成した教材を、日本知的財産協会・研修会・入門コース(関西)・「知的財産情報」用に改訂し、さらに本教育用に改訂したものである。研究開発における、知的財産戦略の考え方を理解してもらうことを目的とした内容となっている。具体的には以下の内容を盛り込んで作成された教材である。

- ① 研究開発における知的財産戦略の考え方、
- ② 研究開発の各ステージに対応した特許調

査及びその考え方、

- ③ 特許審査における、先行技術調査の方法、
- ④ 研究開発において、特許権利侵害を起こさないための注意事項

等である。

また、是非とも強調しておきたい点として、「研究のための調査(特許性の調査)」と、「権利侵害のための調査」とは、調査のための考え方、方法が全く別である、ということがある。

前者の「研究のための調査」はエンドユーザーである研究者・技術者(情報検索、調査のアマチュア)がやってもよいが、後者の「権利侵害のための調査」は必ず、知的財産部に相談し、知的財産情報のインフォプロ(情報専門職)が行うことが必要不可欠である。

### 2.2.2 知的財産関連訴訟の調べ方

知的財産関連訴訟の調べ方についても、簡単に紹介した。具体的には、最高裁判所HPより、知的財産判決速報を、言葉から簡単に検索する方法を例示した。

### 2.2.3 NRI サイバーパテント 複合検索

1回目の教育では、NRI サイバーパテントの概念検索機能を教育したが、2回目では、インフォプロ(情報専門職)が行う特許情報検索に近い内容を教育した。NRI サイバーパテ

## 調査目的の峻別

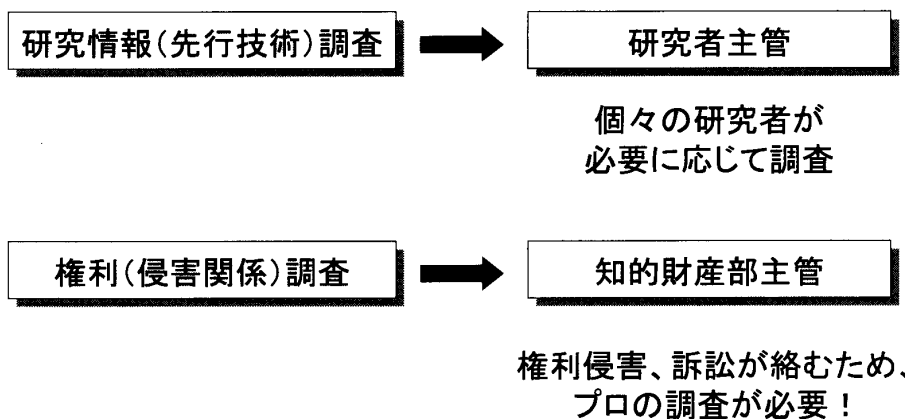


図1 調査目的の峻別

トの複合検索画面を利用し、各検索タームを利用しての検索、検索内容の掛け合わせなどについて、実例を見せながら、各受講者が端末実習をする形式を取った。ここで教育した検索タームとは、以下のものである。

キーワード；検索対象範囲を特許明細書全文にする場合と、特許請求の範囲に限定する場合の2種類の使い分けについて、実例を挙げて説明した。また、NRIサイバーパテントの場合、キーワード、言葉による検索は完全一致検索である。言語表現が違えば、同義語はヒットしないことも明言した。PATOLIS-IVやJOISのような情報検索システムの場合、シソーラスが完備され、統制語が存在するが、NRIサイバーパテントの場合はこれらとは異なる特許情報検索システムであることを理解してもらうよう注力した。

他に、IPC(国際特許分類)、FI(特許庁ファイルインデックス)、FT(Fターム)、出願人・権利者等の検索タームの活用方法についても教育した。

出願人・権利者の場合、完全一致検索ではなく、前方一致検索で行うことを基本として教えた。これは検索漏れを少なくするためである。また、企業同士の合併があった場合は、合併前の企業名+合併後の企業名を全て入力し、合算する必要があることも説明した。さらに、公表公報などでみられるように、外国企業の場合はカタカナ表記になるが、日本国内の代理人の翻訳・表記の仕方でヴァリエーションが発生していることを説明した。可能な限りカタカナ表記のヴァリエーションを把握し、入力し、合算することを徹底するよう、強調して説明した。

発明者からの検索方法も説明した。留意事項としては、ベンチャー企業のような場合、企業名での出願ではなく、社長が個人名で出願している場合があることも説明しておいた。

## 2.2.4 IPDLの効果的な利用方法

(どこを使えば便利か?)

NRIサイバーパテントだけではなく、特許庁IPDLの利用方法も解説した。NRIサイバー

パテントに比較し、特許庁IPDLはキーワードからの検索には不便である。しかし、IPC(国際特許分類)、FI(特許庁ファイルインデックス)、FT(Fターム)からの検索には有効である。また、パテントマップガイダンスを利用すれば、上記IPC、FI、FTの内容を調べたい場合には非常に有効である。実習では、NRIサイバーパテントとは別にブラウザを立ち上げ、特許庁IPDLのパテントマップガイダンス画面からIPC、FI、FTなどの各検索タームを解析しながら、NRIサイバーパテントを利用する方法を説明し、実施させた。

さらに、IPDLの初心者検索画面での、キーワードからの特許・実用新案検索、具体的商品名、役務名からの簡易商標検索も、それらの利用方法を解説した。

## 2.2.5 その他、有効と思われる関連サイトの紹介

さらにレベルアップしたい受講生のために、情報検索、知的財産関係の資格試験関連サイトを紹介した。具体的には以下の試験の関連サイトである。

- ① 弁理士試験(特許庁HP),
- ② 知的財産検定  
(日本知的財産教育協会HP),
- ③ ビジネス著作権検定  
(ビジネス著作権検定専用HP),
- ④ 情報検索基礎能力試験、情報検索応用能力試験2級(情報科学技術協会HP)

## 2.3 第3回目

### 2.3.1 情報検索実習

前半の1時間は、第1日目、第2日目で教育した内容を基に、仮想事例、検索課題を示し、各受講者が端末実習として、情報検索実習を行った。第1回目の実習の時から見られた傾向であるが、情報検索実習を実際にやらせてみると、受講生の反応はさまざまであった。時間にゆとりがあり、いろいろな角度から検索を試みる者から、やり方がわからず立ち往生している者まで大きな個人差があった。



### 2.3.2 最終日の理解度テスト (実技試験)

後半の1時間で、理解度を見るために情報検索の実技試験を行った。内容は、検索課題を10問与え、IPDL, NRI サイバーパテントを利用しての情報検索実技試験を行った。検索課題10問の解答時間は1時間である。解答にあたっては、検索ツール、検索式、検索結果を記載させた。この方法により、特許情報検索における、各受講者の取り組み方、考え方を見ることができた。採点においても、プロではないので、検索結果が正しいかどうか、ということよりも、柔軟に、多角的に情報検索の切り口を考えることができるかどうか、という点を評価のポイントにした。

## 3 受講生の反応 (アンケート結果)

教育終了後、受講生に対してアンケート調査を行った。その結果について述べる。

### 3.1 最も理解しにくかった部分

受講生が理解し難かったところとしては、

- ① 特許制度
- ② 特許関連専門用語
- ③ IPC(国際特許分類)

の、3点が挙げられた。以下のアンケート結果の集計を示す。

### 3.2 アンケート集計結果

(図2)～(図7)にアンケート集計結果を示す。

## 4 結論

### 4.1 アンケート結果に対する考察

3.1でも述べたが、受講生の理解が困難であった箇所として、

- ④ 特許制度
- ⑤ 特許関連専門用語
- ⑥ IPC(国際特許分類)

の、3点が挙げられている。やはり、単なる情報検索教育だけではなく、知的財産制度を平行して教育することが必要と解される。一般的なこととして、知的財産情報のプロを育成する場合、知的財産制度が理解できていなければ、知的財産情報のプロとは言いがたい。知的財産制度を理解していなければ、図書館情報学の専門家になってしまうからである。

宇部興産(株)において、エンドユーザー教育を実施した際、情報検索アドバイザー側からの意見として、特許教育、特許明細書作成教育の定期的な開催の要望があった。その理由として、特許明細書作成のポイントを理解していなければ、特許調査においても、そのポイントが理解できない。目的にあった的確な特許調査ができないため、ということがあげられた。

### 4.2 山口 TLO に対する知的財産情報教育支援の目的

山口 TLO に対する知的財産情報教育支援の目的は、起業家育成のための知的財産情報教育である。単なるインフォプロ(情報専門職)／サーチャー(情報検索担当者)育成の訓練ではない。そのことから、知的財産情報の分析・加工、さらには知的財産戦略立案へとつながる教育を心がけたつもりである。教育においては、単に情報検索の技能を修得するだけではなく、知的財産戦略に対する考え方を強調して教育をおこなった。

### 4.3 起業家訓練を踏まえた知的財産教育での留意点

以上のことから考えると、起業における、知的財産戦略の立案へとつながる、知的財産情報教育である必要がある。同時に、知的財産制度の教育が必要不可欠である。

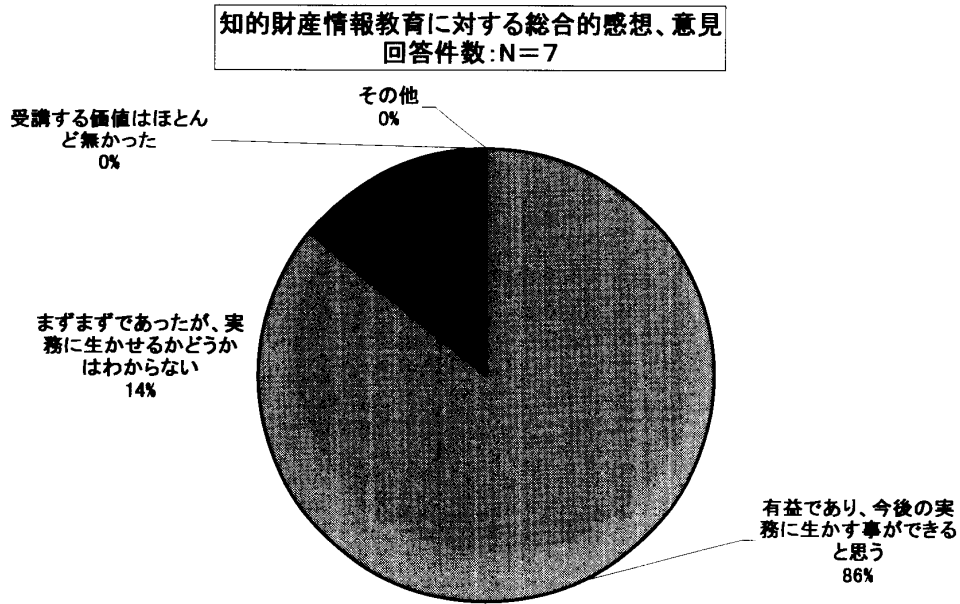


図2 知的財産情報教育に関する総合的感想、意見

**詳細な説明が聞きたい項目**  
回答件数: N=15 (複数回答)

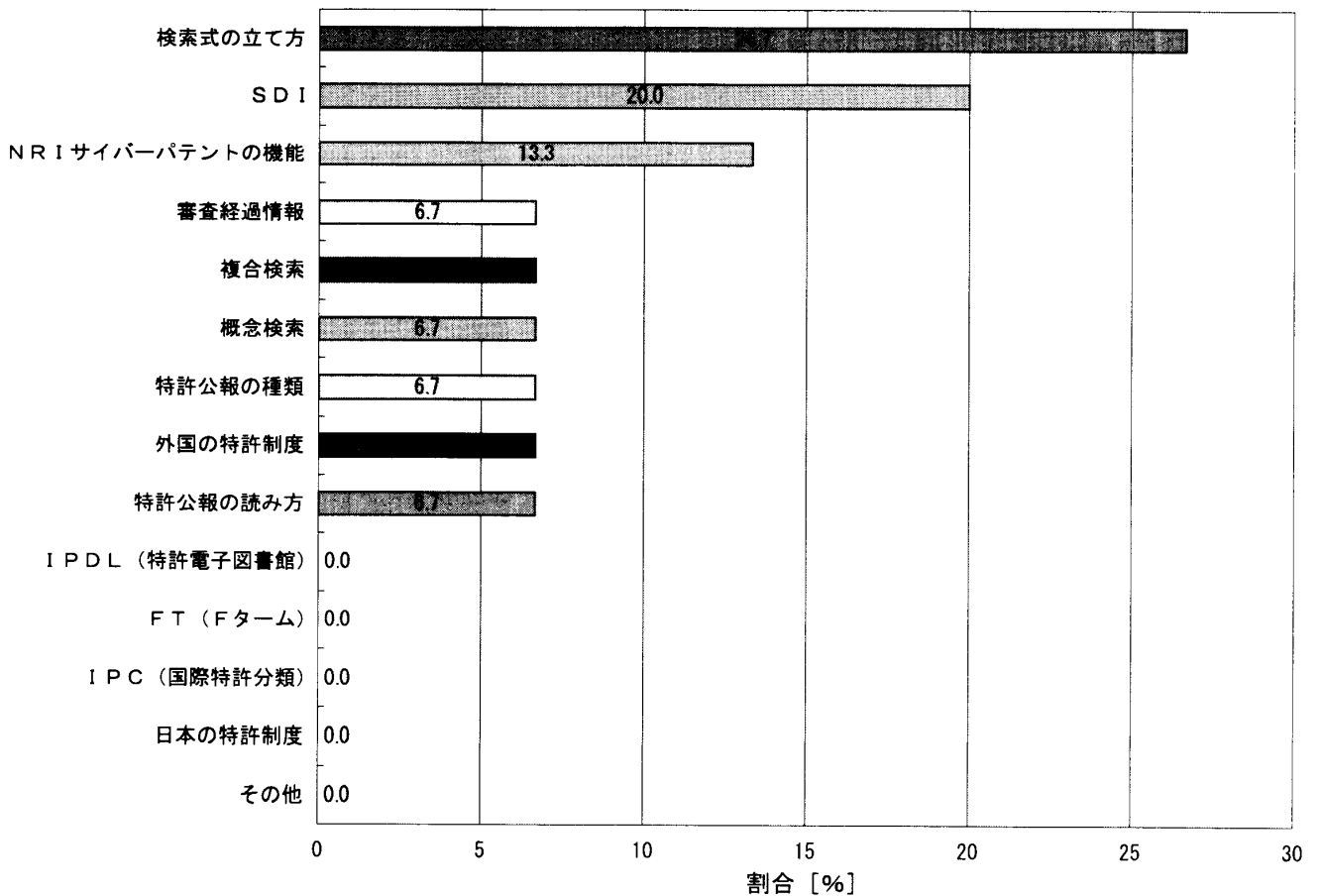
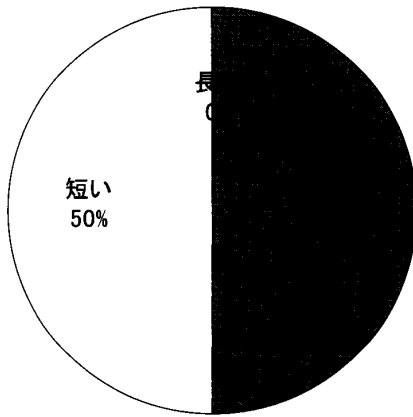


図3 詳細な説明が聞きたい項目

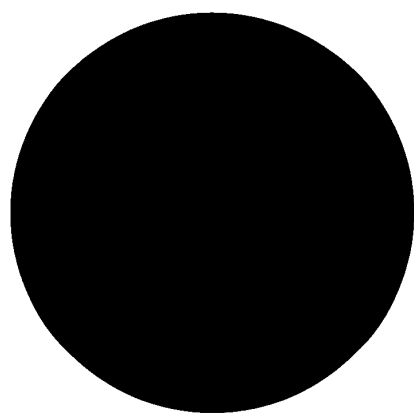
**実習時間**  
回答件数：N=8



■長い ■適当 □短い

図4 実習時間に対する感想

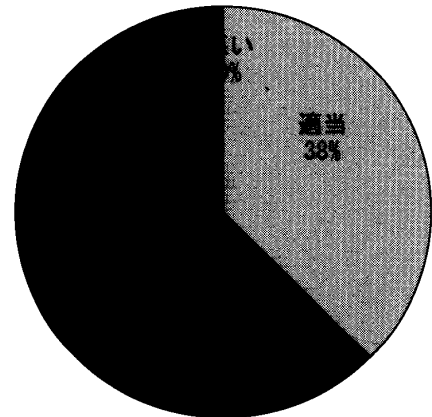
**実習内容**  
回答件数：N=8



■難しい ■適当 □易しい

図5 実習内容に関する感想

**実習中の指導内容**  
回答件数：N=8



■適当 ■良い □悪い

図6 実習中の指導内容に対する感想

**各教育の内容、説明に対しての理解の程度**

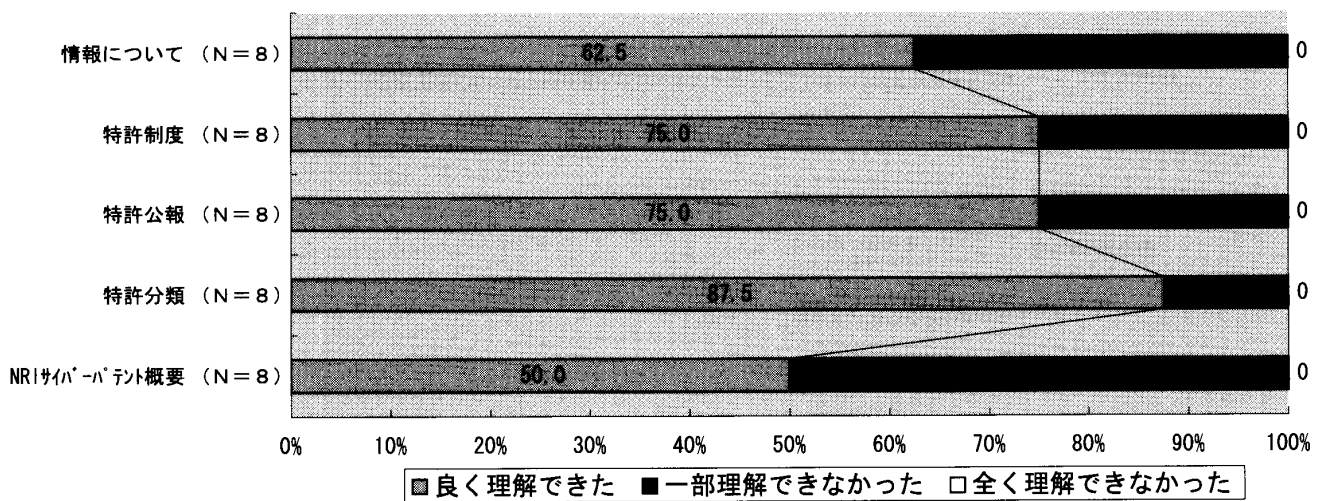


図7 各教育の内容、説明に対する理解度

**謝辞**

本報告投稿に当り、許可をご快諾いただいた山口TLO、山口大学産学連携コーディネーター及び宇部興産(株)の関係諸氏に対し、この紙面を借りて、深く感謝する次第である。

**参考文献**

- [1] 慈道佐代子：「利用者向け研修プロジェクトー企画と実施例ー」, 情報の科学と技術, Vol.53, No.5, pp.263-268, 2003.
- [2] 情報科学技術協会セミナー(エンドユーザー検索ー企業内教育の取り組み 2003年9月26日開催), 資料, 2003.

- [3] 岡本和彦：「宇部興産(株)におけるエンドユーザー教育：研究開発本部における『情報検索アドバイザー制度』」, 第40回情報科学研究集会予稿集, pp.91-94, 2003.
- [4] 岡本和彦：「宇部興産(株)におけるエンドユーザー教育：研究開発本部における『情報検索アドバイザー制度』」, 情報管理, Vol.47, No.1, pp.15-19, 2004.
- [5] 特集：特許検索に必要なスキルと知識, 情報の科学と技術, Vol.54, No.5, 2004.
- [6] 岡本和彦：「宇部興産(株)におけるエンドユーザー教育：研究開発本部における『情報検索アドバイザー制度』」, 日本知財学会第二回研究発表会講演要旨集, pp.362-365, 2004.
- [7] 岡本和彦：「全国図書館大会への招待, 第6分科会(専門図書館), 専門図書館の人材育成, 宇部興産(株)におけるエンドユーザー教育：研究開発本部における『情報検索アドバイザー制度』」, 図書館雑誌, Vol.98, No.9, pp.645, 2004.
- [8] 岡本和彦; 出口昌信：「宇部興産(株)におけるエンドユーザー教育Ⅱ：グループ企業内におけるPATOLIS-Ⅳの導入教育」, 第1回情報プロフェッショナルシンポジウム(INFOPRO2004)予稿集, pp.99-102, 2004.
- [9] 岡本和彦：「宇部興産(株)におけるエンドユーザー教育：研究開発本部における『情報検索アドバイザー制度』及びインフォプロのスキルアップ(私見)」, 平成16年度(第90回)全国図書館大会要綱, pp.45-46, 2004.
- [10] 岡本和彦; 出口昌信：「山口大学に対する知的財産情報教育支援～産学連携における試み～」, 第11回医学図書館研究会・継続教育コース予稿集, 2004.
- [11] 岡本和彦：「宇部興産(株)におけるエンドユーザー教育：研究開発本部における『情報検索アドバイザー制度』」, 2005年情報学シンポジウム講演論文集, pp.23-27, 2005.
- [12] 岡本和彦：「宇部興産(株)におけるエンドユーザー教育：研究開発本部における『情報検索アドバイザー制度』及びインフォプロのスキルアップ(私見)」, 平成16年度(第90回)全国図書館大会香川大会記録, pp.242-251, 2005.
- [13] 岡本和彦; 出口昌信：「山口TLOに対する知的財産情報教育支援-産学連携における試みⅡ-」, 日本知財学会第三回学術研究発表会講演要旨集, pp.268-271, 2005.
- [14] 出口昌信; 岡本和彦：「宇部興産(株)におけるエンドユーザー教育：営業部門に対する知的財産情報教育の試み」, 日本知財学会第三回学術研究発表会講演要旨集, pp.280-283, 2005.

(2005年10月7日受付)

(2006年1月13日採択)

# 生物遺伝資源総合データベース構築における XML 技術適用について The Application of XML Technology in Genetic Resource Database

山川 武廣\* 山崎 由紀子\*,†

Takehiro YAMAKAWA and Yukiko YAMAZAKI

国立遺伝学研究所・生物遺伝資源情報総合センターでは、国内における様々な生物の遺伝資源情報に関してデータベース化を行っており、これまでに動物から植物、微生物に至るおよそ 20 生物種のデータベースを構築している。最近はこうした生物種毎のデータベースを利用するだけでなく、これらの情報を種横断的に利用したいという要望が寄せられるようになった。そこで昨年より生物遺伝資源総合データベース（バイオリソースワールド）の開発を開始し、現在試験運用を行っている。本稿では、「多様性」という特徴をもつバイオ分野の情報を統合的に扱う上で、XML 技術をどのように適用したのか、そしてその適用効果について解説する。

Here at the Center for Genetic Resource Information in the National Institute of Genetics, we have constructed genetic resource databases of approximately 20 species of organisms from animals, plants to microorganisms. Lately, instead of databases designed specifically for one particular organism, there is a demand for integrated databases where data can be utilized regardless of species. Thus, we started to develop an integrated database of genetic resources (BioResource World) which is now released as a test version. This paper explains how XML Technology was applied to handle integrated data in the biotechnology field which is characteristically diversified and the effects of its application.

キーワード：XML, 生物遺伝資源  
XML, Genetic Resource

## 1 はじめに

国立遺伝学研究所・生物遺伝資源情報総合センターでは、国内における様々な生物の遺伝資源<sup>1</sup>情報をデータベース化し、世界中の研究者にインターネットを通して情報発信している。また 2002 年に立ち上がった「ナショナルバイオリソースプロジェクト (以下 NBRP)」においては、情報中核機関としてリソース中

核機関の情報整備を支援している。(図 1)

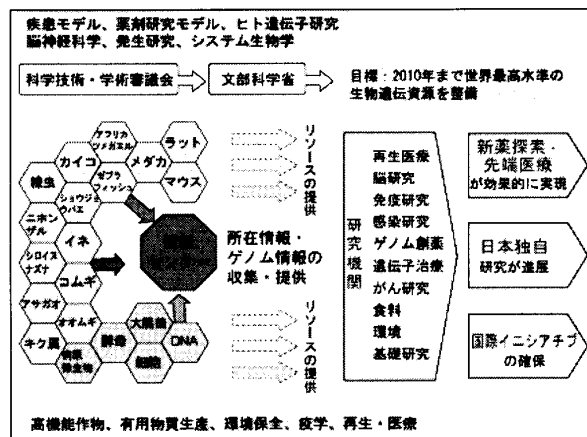


図 1 NBRP 実施体制図

\* 国立遺伝学研究所  
National Institute of Genetics

† yyamazaki@lab.nig.ac.jp

<sup>1</sup> ここで言うところの遺伝資源は、実験研究に使われる生物材料のことで、たとえばマウス、ショウジョウバエなどの個体、植物の種、細胞、DNA など、増殖して配布することのできるものを意味する。

われわれが整備する情報の利用者として想定しているのは、研究目的のために当該リソースを使う研究者が中心ではあるが、研究者を目指す学生や、科学知識に興味のある一般市民も対象として考えており、広い利用者層にわかり易い情報公開を目指している。具体的には、マシンやブラウザに依存しないシステム、日本語と英語の両言語に同時対応する、閲覧・検索・ダウンロードファイルの整備、予備知識がなくてもある程度の情報が入手できる仕組みの実装などを心がけている。また最近ではミニアンケートを掲載することによって、気軽に不満をフィードバックしてもらうなど、利用者のニーズ把握にも積極的に対応している。われわれは情報を利用して研究成果が出た場合に URL を引用してもらう以外に利用に関する制限をいっさい設けていない。[判り易い][速い][新しい][情報を網羅している]ことを常に心がけている。

1995年に資源情報の公開サイトを開設して以来、生物種毎にデータベースの整備を進めており、これまでに動物から植物、微生物に至るおよそ20生物種のデータベースを構築している。利用状況として、平均月間利用者数は2005年10月現在、3万2千人、平均月間利用件数は65万件である。情報整備支援という立場上、情報を保有する機関との連携は必須であり、データベースもそれぞれの運用形態に合わせた設計となる。OSはLinuxまたはUnix、DBMSはOracle, Sybase, MySQL, PostgreSQL, ObjectStoreなど多種類扱っており、運用も当センターのサーバマシンで完全に管理しているものから、外部機関のサーバマシン上のDBを遠隔管理するもの、DBのみ内部管理し、HTMLページのみ外部編集を許すもの、あるいは複数の機関で共同管理するものまで、実に多様である。

最近是这样した生物種毎のデータベースを利用するだけではなく、これらの情報を種横断的に利用したいという要望が寄せられるようになった。そこで昨年より生物遺伝資源総合データベース(バイオリソースワールド以下BRW)の開発を開始し、現在試験運用を行っ

ている。

本稿では、「多様性」という特徴をもつバイオ分野の情報を統合的に扱う上で、XML技術をどのように適用したのか、そしてその適用効果について解説する。

## 2 資源情報の特徴—多様性

遺伝資源情報の1単位は以下の情報から成り立っている。

- (1) 生物種名
- (2) リソース種類
- (3) リソース名
- (4) 分譲の可否
- (5) 保存機関・入手先
- (6) リソースの特性や付随する研究情報

例として、線虫、イネ、藻類、大腸菌のリソース情報を図2、図3、図4、図5に示した。

図2-5の例から、(1)-(5)の情報形式は共通であるが、(6)の特性および付随情報は、生物種毎に大きく異なっていることがわかる。同じ生物種内でもリソースの種類、研究分野によって(6)の内容は異なるし、さらに他の生物種を含めると多様性は増すばかりである。

従って、このような情報を扱うためには多様性に柔軟に対応できるデータベースシステムが必要である。

## 3 生物遺伝資源総合データベース (BRW)

### 3.1 BRWの全体像

BRWには現在24種類の異なる生物種の情報を格納しており、遺伝資源情報および関連情報の閲覧・検索ならびに入手方法に関する情報提供を可能としている。図6、図7、図8に画面例を示す。

トップページでは、生物種を超えた全文検索を行うことが出来る。また、生物種別の登録件数の閲覧、リソース種類別の登録件数を閲覧することが出来る。

検索結果一覧表示画面では、検索結果を一

図2 線虫の遺伝資源情報の例

図4 藻類の遺伝資源情報の例

図3 イネの遺伝資源情報の例

図5 大腸菌の遺伝資源情報の例

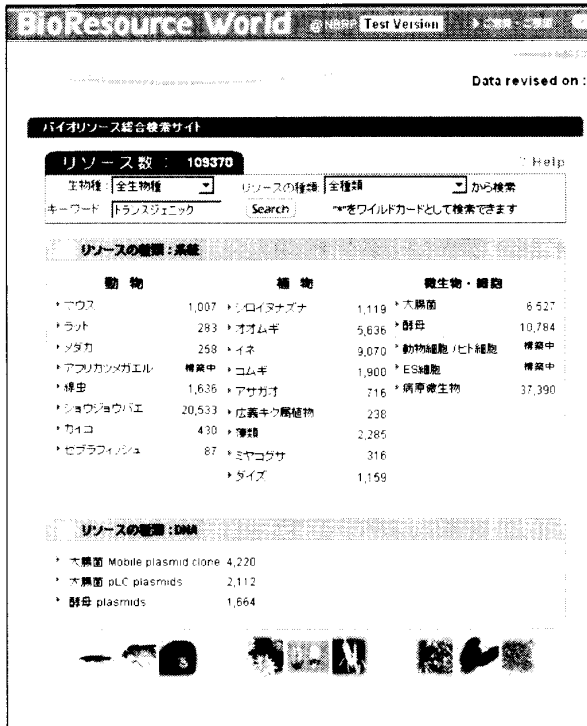


図 6 BRW のトップページ

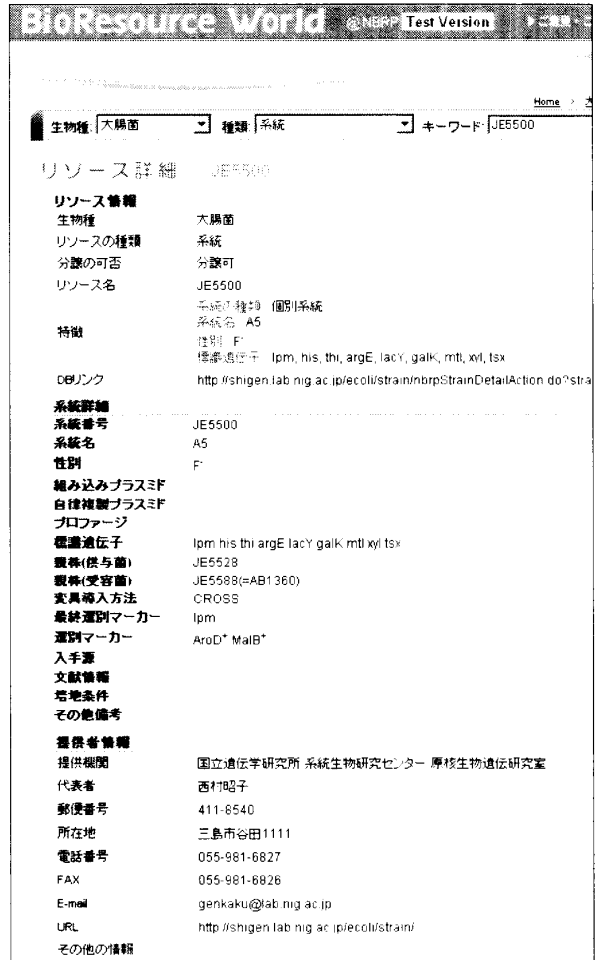


図 8 BRW 遺伝資源情報詳細画面

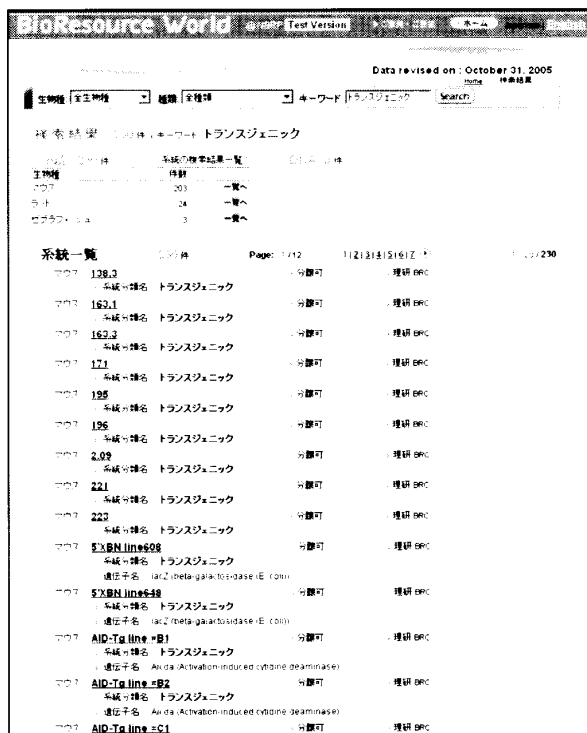


図 7 BRW 検索結果一覧表示画面

覧形式で閲覧出来る。また、生物種別の検索ヒット数の閲覧、検索キーワードのハイライト表示、ソート機能を利用することが出来る。

遺伝資源情報詳細画面では、(1) 生物種名、(2) リソース種類、(3) リソース名、(4) 分譲の可否、(5) 保存機関・入手先、(6) リソースの特性、等の情報を得ることが出来る。また、当該リソースの生物種データベースへのリンクを辿ることで、さらに詳細な研究情報を取得することが出来る。

### 3.2 データベース管理システムの要件と検討

遺伝資源情報を格納するデータベース管理システム (以下 DBMS) は、データの特徴や運用面を考慮して選定する必要がある。以下に考慮した点を挙げる。



### 3.2.1 DBMS の要件

#### (1) 情報量と検索性能

現在 BRW のデータ数は十数万件であるが、今後 DNA などのゲノミックリソースの情報を取り込む予定があり、データ総数は百万のオーダーになる。

これらの膨大な情報から迅速に結果を抽出して、画面に表示する必要がある。しかも

形式の異なる全生物種の特性および付随情報を対象とした全文検索を実装する必要がある。

従って、このような要件を満たすデータベースシステムが必要であった。

#### (2) データ運用・管理の容易さ

遺伝資源は基本的に生物であることからデータ変更の発生頻度が高い。リアルタイムに情報更新が行われないと、分譲中止の遺伝資源に対して依頼手続きが行われしまう等の問題が発生することになる。

また、遺伝資源を保有する機関は、大学の研究室や国の研究機関が多いため、送付されてくる情報のフォーマットは必ずしも統一されていないのが実状である。このような状況下でも情報を迅速に総合データベースに反映する必要がある。

従って、大規模なデータの追加・削除・更新をサポートしているツールが存在する等、運用・管理面においても使いやすいデータベースシステムを適用する必要がある。

### 3.2.2 DBMS の検討

DBMS には幾つかの種類が存在するが、3.2.1 で示した要件を踏まえ、以下の 3 種類についてそれぞれ検討した。

#### (1) リレーショナルデータベース

データの集合を「テーブル」と呼ばれる表で表す方式で、古くから多くのシステムで使われており、DBMS では最も普及しているタイプである。

当システムの適用を考えた場合に、膨大な情報からの検索性能や安定性、運用・管理面でのツールの充実度など、優れた点が多いが、多様な情報を持つ研究情報の格納、検索という点で適用は困難であると判断した。

#### (2) XML データベース

XML の階層構造をそのままデータベースに格納して管理できる。異なるデータが混在する情報を格納するのに適しているが、リレーショナル型に比べて歴史が浅く、管理ツールが少ない等の問題がある。

当システムはデータ運用・管理面、及び、情報量が大きい場合の安定性など不安定な要素が大きいため採用を見送った。

#### (3) XML 対応型リレーショナルデータベース

リレーショナル型と XML 型の双方の特徴を備えたハイブリッド型で、双方のメリットをバランス良く備えている。

基本的にはデータをテーブルとして扱うが、必要に応じてカラムの属性に XML 型を選択することも可能である。

また、検索に関しては、リレーショナル型で使われる SQL 言語をベースに、XML 部分の検索に関しては XQuery 言語等にも対応している。

当システムでは、3.2.1 で述べた 2 点の要件から XML 対応型リレーショナルデータベースを採用した。

## 3.3 システム実装

### 3.3.1 データ作成、格納

XML 対応型リレーショナルデータベースに格納する際に、データ形式が決まっておらず、管理頻度の高い項目を表形式で管理し、データ形式が多様な項目を XML 形式で管理するようにした。

#### (1) 表形式で管理する情報

生物種、リソース名、分譲の可否、保存機関

#### (2) XML 形式で管理する情報 (図 9)

リソースの種類、特性情報および付随情報

### 3.3.2 データの検索機能の実装

表形式部分のデータ検索には一般的な SQL を用い、XML 形式部分に関しては DBMS ベンダーの提供する全文検索関数を使って検索機能を実装した。

```

<?xml version="1.0"?>
<Resource>
  <Resource_types>
    <Resource_type>
      <en>seed</en>
      <ja>種子</ja>
    </Resource_type>
  </Resource_types>
  <Features>
    <Feature>
      <en key="Species">0. ridleyi</en>
      <ja key="Species">0. ridleyi</ja>
    </Feature>
    <Feature>
      <en key="Genome">HHJJ</en>
      <ja key="Genome">HHJJ</ja>
    </Feature>
  </Features>
</Resource>

```

図9 XML形式データ例

### 3.3.3 情報の表示

遺伝資源の特性情報や付随情報などのように、形式が多様であるデータの格納は、XMLを使うことで実現した。しかしそのまま一律に表示すると、データによっては不自然で非常に使い勝手が悪くなる。そこで内容別に細やかな表示設定ができ、より自然な表示を実現するために、XSL(Extensible Stylesheet Language)とXSLT(XSL Transformations)技術を適用した。

図10に表示メカニズムの構造図を示す。

まず、XML対応型リレーショナルデータベースに遺伝資源情報を格納する。基本情報に関してはテーブル形式、特性・付随情報に関してはXML形式で格納する。

次にWebアプリケーションサーバにおいて、利用者が必要とする情報を動的にデータベースから取得し、出力結果をHTMLで生成する。このとき、生物種毎に異なる特性・付随情報に関しては、XSLTを使って、XMLと

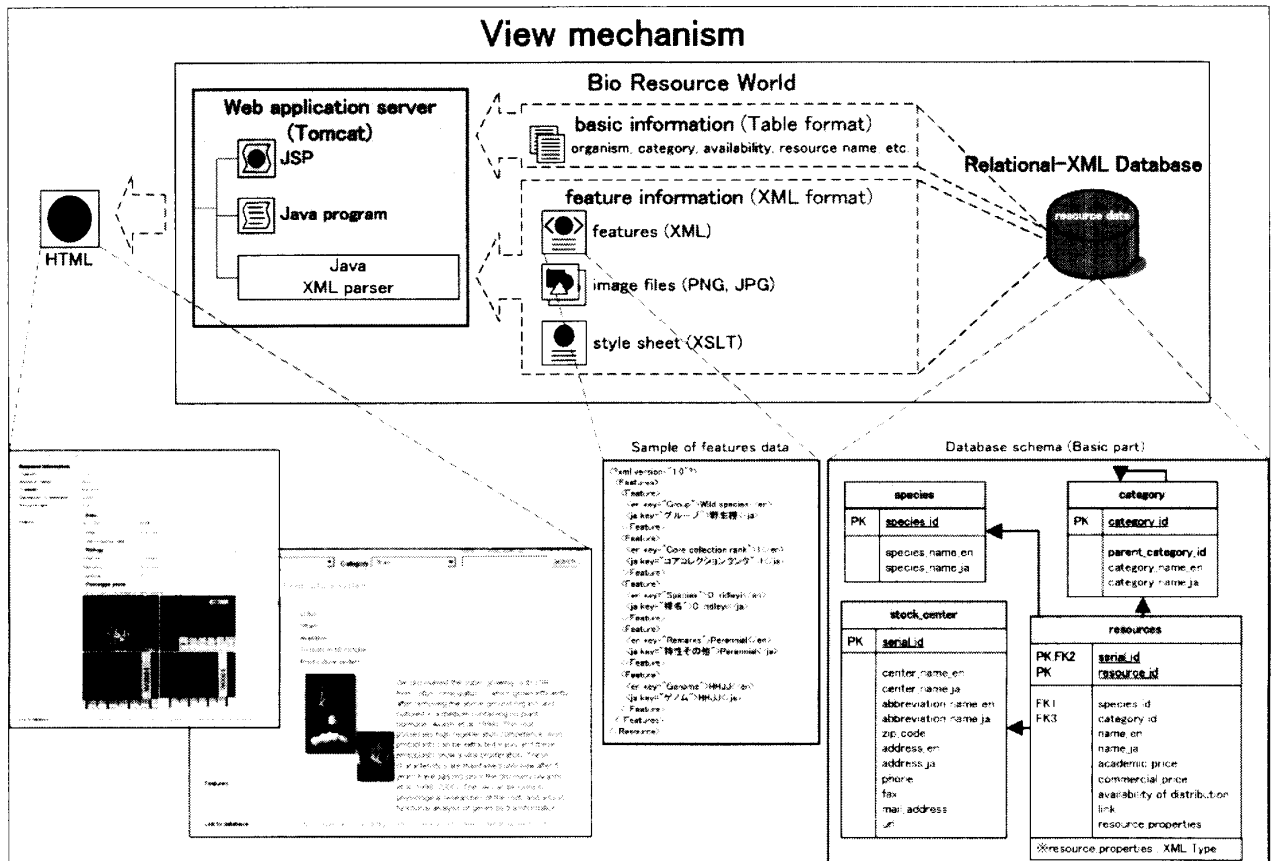


図10 BRW表示メカニズム構造図

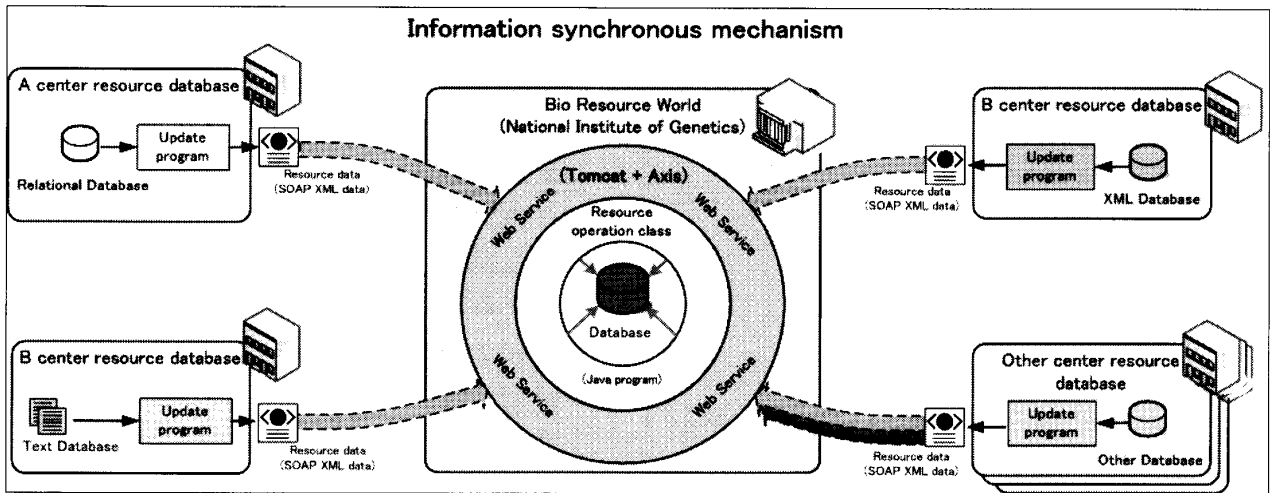


図 11 Web サービス技術を用いた情報同期

XSL, 及び画像情報を合成して利用者に理解しやすい表示形式に変換する。

### 3.4 XML 技術適用の効果

今回のデータベース構築で XML 技術を適用したことにより, 以下の効果を得ることができた。

- (1) データ格納, 運用・管理  
データ形式が固定で, 比較的更新の頻度が高い情報をリレーショナルデータベースに格納し, データ形式が柔軟に変化する研究情報を XML 形式として格納することで, 情報の多様性, 運用・管理の容易さに関する要件を満たすことができた。
- (2) 性能  
XML 対応型リレーショナルデータベースを利用したことで, 100 万件のデータに対する全文検索を数秒で取得することができ, 性能要件を満たすことができた。
- (3) データとレイアウトの分離  
研究データに関して, データ部分を XML, レイアウトを XSL に分離することにより, データのみの修正, レイアウトのみの修正それぞれに対して柔軟に対処することができた。

## 4 今後の展望

情報のデータベース格納時, 及び, 表示時に関しては XML 技術適用で先に述べた効果を得ることができたが, 情報の収集という観点では, 各研究機関から依然として CSV 形式や Excel 形式等, ファイル形式の統一が取れておらず, データ整形に時間がかかってしまっている。

この問題を解消するため, 次期開発フェーズでは, Web サービス技術を用いたデータ更新方法を検討している。(図 11 参照)

BRW 側には, データベースの外側にデータの更新を可能とする Web サービスインタフェースを設置し, 外部からの正確な情報更新を可能とする。

また, 各研究機関には, BRW の Web サービスインタフェースを介して情報更新を行うプログラムを配布し, 機関毎のデータ保存形式の違いによるデータ整形時間を短縮する。さらに, このインタフェースを用いて, 新規に参入する機関が自由に遺伝資源情報を追加・更新する事ができないか検討を進めている。

このように XML 技術を使うことによって, 複数の異なるデータベースを対象とした総合検索を実現することはできたが, さらに高次な情報利用サービスを提供するために, 現在我々は XML 技術の延長上にあるセマンティック

Web の構築を計画している。セマンティック Web とは、Web サイトの意味をコンピュータに理解させることで、より高次の情報収集を可能とする技術である。

情報の意味を示すメタ情報をメタデータ記述用フレームワーク「RDF」(Resource Description Framework) を使って定義し、生物種間の情報の関係を明確にするため、オントロジ技術を適用する。

幸いオントロジという考え方は生物学の分野にも 2000 年に登場し、今では Biological Ontology として多数のオントロジプロジェクトが立ち上がっている。Gene Ontology (遺伝子オントロジ), Anatomical Ontology (解剖学オントロジ), Developmental Stage Ontology (発生オントロジ), Phenotype and Trait Ontology (形質オントロジ) などはモデル生物を中心に実際に利用されはじめており、中でも Gene Ontology は最も普及しているものの 1 つである。図 12 にはこれらのオントロジを利用した種横断的な情報利用のイメージを示した。中央のオントロジは必要に応じて入れ替えることができる。各データベースは、格納されている情報と各種オントロジとの対応情報を持つだけでよく、データベースの内容を変更する必要は全くない。利用者は、

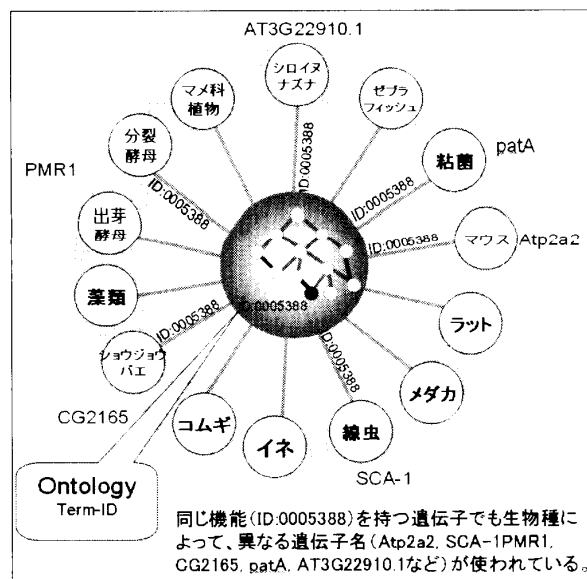


図 12 Biological ontology による種横断的な情報利用の実現イメージ

よく知っているある 1 つのキーワードを頼りに検索を開始するだけで、関連する情報を種横断的かつ網羅的に抽出することができるようになるはずである。

(2005 年 11 月 29 日受付)

(2005 年 12 月 15 日採択)

# 我が国における情報学の始まり

## The Dawn of Informatics in Japan

藤原 鎮男

Shizuo FUJIWARA

### 1 はしがき：本稿の趣旨

学会誌の編集委員会から、学会発足時のことを本誌に寄稿するようお奨めがあった。古い言葉に「温故知新」というのがある。古いものにも現代に有益なものがあるという注意であり、より積極的には、古きものから有用なものを発掘すると良いとの注意でもあろう。ともかく学会発足時の原点を振り返り、今後の活動の参考にしようという趣旨かと思ってお受けした。考えれば、当時の関係者のかなりの方がすでに亡く、現役の会員諸氏とても多くは新世代の方であるので、学会が創立された頃の事の多くは「神代の神話」になろう。そういう意味では往時のことを記録しておくことも有意義かも知れない。ただし、相当前のことであり、筆者の記憶違い、個人的な認識の誤りがあるかも知れない。そういうことでお気づきの点があったら訂正のご指摘をいただきたい。同じ懸念から平成17年3月の本学会月例懇話会で、本稿の内容のかなりの部分を卓話として披瀝した。その後ご注意は無く、むしろ卓話は好評で、初めて知ったことが多いとの反響もあった。それで、卓話の原稿に加筆して本稿を記した。もう一つお断りせねばならぬのは、本稿は「我が国における」と表題に謳うが、筆者自身の直接経験の事項が主となり、当然触れるべき方々のことを失しているおそれがあることである。これもご寛恕を願いたい。

### 2 我が国における情報学の始まり

#### 2.1 淵源

「情報学」を学術の情報・知識を集積・整

理しその流通の方策を攻究する学問であるとすれば、先ず知識の集積ではデータベースの構築があり、その始まりは後で述べる「本草学」(ほんぞうがく)に遡る。近代科学の時代になっては学术论文の抄録事業が第一で、東北大学の真島利行教授の『日本化学総攬』の刊行(1927)に始まり、これは藤瀬新一郎教授に引き継がれ、1963年に日本科学技術情報センター[JICST, 現科学技術振興機構(JST)]の事業に吸収されるまで個人で運営された。英米仏独においても同様な抄録事業はあったが、米国化学会のケミカルアブストラクツ(CAS)に吸収されているから、我が国もこれと同様の経過をたどったといえよう。この抄録事業は化学という専門分野に限定されているので、化学内のこととして情報学とは別に取り扱われる。知識のデータベース構築を情報学の仕事に含めれば、情報学の始まりは、古く律令時代に始まり、今日に至っていると言える。

近代科学の時代にあっても人は案外自分の経歴の古さを誇示する傾向がある。例えば、大学の歴史の話では近代史への登場から数えたがる。「近代」の開扉は自分だという意識からである。我が国の人は素直に西欧を讃えるので、我が国の大学は世界の中では非常に若いほうに入ってしまう。若いのがどうということもないが、公平に見て、足利学校の気象、天文、その他は結構立派なレベルの高等専門教育であったようである。それ故、そこから数えれば、我が国の高等教育の始まりは、世界有数の古さになる。同様の趣旨で、物質の悉皆記録作成を情報学の業績とするならば、本草学にみるとおり、我が国はその先進性を誇ってよい。我々の自尊、自信のこととして、この自覚はあってよい。

筆者は先に「日本の情報化学」の題で我が国の「情報化学」の初期のことをかなり細かく述べた<sup>[1]</sup>。それで、「我が国における情報学の始まり」とする本報に併せて、同報文をも参照願えれば幸いである。

## 2.2 日本学術会議の発足

ともかくここでは我が国における「情報学」の発足は、1949年の日本学術会議発足に基づくとしておこう。学術会議の発足の由来については、産官学の関係機関がそれぞれの立場で意見を持つであろうが、それはそれとして、ここでは一若年研究者としてその発足を眺めていた筆者の情報学関連の学術会議像を述べる。

日本学術会議は、当時の学界の指導者たちが、戦後の日本の再建のために学界、さらに言えば、科学自体がどのように働くべきかを真剣に考えた結果作ったものだと思う。当時の状況でイデオロギーに基づく動きが目立ったのは事実だが、少なくとも科学技術畑のもの的大部分は、素直に亀山直人会長が主導した「科学立国」の旗印に共鳴し努力した。亀山会長は、「皆が自分の持ちものを持ち寄り、国の再建に努力しよう」と科学者に呼びかけた。筆者は東大理学部大学院化学専攻課程に在籍中の学生であったが、指導教授の木村健二郎先生は、亀山先生の言葉を受けて「我々は分析化学の立場」でそれに答えようと言われた。

筆者はそれを受けて「化学探鉱」をした。それは木村先生の指導で、戦争末期に米の生産に必須のリン酸肥料が国に皆無となり、能登半島で地下水中のリン酸イオンの微量分析をして地層中のリン酸資源を探査し、実際にその発見に成功した経験があることによる。戦後、米の増産が国の至上命題になり、化学肥料の硫安(硫酸アンモニウム)の原料である硫化鉱石の増産が求められ、鉱山地域の沢や谷の水の微量分析から鉱床の露頭を探査することになったのであった。筆者は山歩きの経験は皆無だが、地質、鉱物の専門家の支援、協力があり、木村先生のもとで休暇毎に研究室を出て東北や北海道の鉱山地域の山野を巡り、

硫化鉱脈の存否を地表の流水の微量分析から探索した。直接鉱脈の発見につながったのは1例にすぎなかったが、間接的には、我々が化学探鉱で発見した露頭から各鉱山会社が探査を続け、10年もたつて鉱脈の発見に至った例がいくつもある。鉱山関係の企業は化学探鉱の部門を作り、それがその後ウランやゲルマニウム探査に展開した。また、同種の化学探鉱を実は世界各国も実施していたと知ったのは後日だった。さらに近年は環境化学や公害分析のような意外な展開にもなっている。

ともかく、ここで述べたかったことは、学界が真剣に各自の専門で世の中に寄与する努力をしたということであり、情報学という未知の分野に小生が首を突っ込む素地も同根であるかもしれない。より直接の理由は、小生が1951年(昭和26年)に核磁気共鳴(NMR)の装置を手製し、その延長としてスペクトルの解析のために電子計算機との関係が出来、さらにその関連で「情報処理」の研究に入ったことである。

## 2.3 「情報学研連」の発足

「情報学」は、戦後の日本復興の旗手の一つとして学術会議創立時に作られた学術の新分野であった。その推進者は、順不同に挙げれば松前正義、八木秀次、浜田成徳氏ら通信科学界の諸氏、物理の茅誠司氏らであった。西欧の情報学の動きをキャッチし、新時代の科学の動向に沿った学術分野の一つとして、学術会議の発足と同時に「情報学」を冠する「学術研究連絡委員会」(研連)が設立され、その下に「文献情報分科会」と「数値情報分科会」の二つの分科会が発足した。分科会の委員長は、前者が大塚明朗氏、後者が小谷正雄氏であったと思う。

情報学研連のこの構成は、当時の世界の情勢を反映していた。それは戦後の世界の文化活動に大きく関わったユネスコの動向である。彼等は特に電子計算機による学術資源の取り扱を進めようとした。それで世界各国が科学技術情報センターを作ることを唱道し、実際、

1960年代初頭には世界規模でそれが実現した。ただし、それは電子計算機により、学術論文の最新知識を求める医学や化学の現場の研究者に情報を配信することが主内容であった。医学のメドラス (MEDLARS) や化学のケミアブ (Chemical Abstracts) はそれである。1972年ころ筆者は文部省の委嘱で慶應義塾大学の津田良成教授と一緒に英米独をまわって、その状況を調査視察した。当時は幼稚なもので、ワシントンでは計算機からの印刷出力をトラックで配信している所もあった。

しかしながら、学術情報は学会誌の最新号の論文に記載されたものだけを対象にすればよいというわけにはゆかない。当然、過去分を含む学術論文抄録の計算機による集積と配信は、図書館その他の情報蓄積機関との調整が必要である。そのためユネスコ発足時に英米独仏加などの先進国の連携で、情報学関連の国際機関として国際ドキュメンテーション連盟 (FID, Institute International de Documentation から改称された) と CODATA の二つが形成された。これらは一国一メンバーの形の国際機関であり、各国が分担金を拠出する。我が国も学術会議がメンバーとなって分担金を支払うこととし、前記の二研連分科会が情報学の推進を担うことになった。このように、情報学の事業の実態が図書館系と情報センター系の二系統に分かれての進行を余儀なくされたのは、我が国のみのことではない。世界全体がそうであって、これは、「学術知識」の実態によることである。ちなみに、「科学技術」ないし、「科学・技術」と科学と技術を結んだ言葉が現在は定着しているが、これは、原子力や科学技術情報を主管する科学技術庁が新設されたときに作られた用語である。

以上は、情報学のコンテンツで言えばいわば既に生まれた情報に関することである。当然、これらを生む側にも情報学があった。すなわち、元来、数学、物理学、化学、生物学、地学などの伝統的学問の個々の専門分野には、国際ユニオンがあり、国別に当該分野を代表する機関が定まっていた、この機関が集まって国際連盟 (ユニオン) を形成し、その総会は

各国の分担金を決め、メンバーはそれを工面して本部に納入する。一国一機関であるので、この分担金は我が国では国費で支払っている。

情報学についても先述のように、日本学術会議は FID と CODATA の両機関に分担金を支払うため、前者は情報学研連学術文献情報分科会、後者は情報学研連数値情報分科会へ補助金を出し、隔年の総会に代表を送る旅費も支給される習慣になっていた。これは、学術会議の全研連中におけるただ二つの特例であった。戦後生まれた新しい学問分野であるとはいえ、特別の措置であり、情報学研連創立時の上記諸先輩の努力が偲ばれる。分科会のメンバーは、「図書館学関係者」と「学術雑誌」依存の「学術専門分野研究者」の二つに分化していた。このことはその後の「情報学」の動向に大きく影響した。すなわち、図書館にはかれこれ 100 年程前に図書以外の文書まで含めてそれを内容別に「コード化」して整理すること (ドキュメンテーション) が始まり、このコード使用を有料のライセンス制とする UDC システムが普及していた。FID の運営資金は始めはこのシステムからの寄金によっていたが直ぐそれは無くなり、ユネスコからの補助金と、FID に対する各国の分担金が資金源となった。ヨーロッパの図書館でも UDC の利用が顕著であったが、英国図書館が先ずそれをあらため、また米国は元来これになじまないという状況で、結局、FID は UDC を早くに離別したのであるが、その後、活動不振で先年解散となった。ただし、それは後年のことで、英国、カナダ、米国の協力で FID と CODATA はそれぞれ情報処理の方策の確立に努力した。我が国の情報学研連でも、もっと「情報学」に即した活動を活発にすべきだという議論が起こり、1984年に改組し、それと同時に情報知識学会を創立した。

## 2.4 情報知識学会 (JSIK) の発足

情報知識学会の発足には、小谷正雄、大塚明朗、湯川康秀、島内武彦、米田幸夫、藤原譲、それに小生らが情報学研連側から企画に

関わり、米田、藤原譲、月見里礼次郎氏らが当事者としての実務的奔走をして JSIK が発足した。発足時のことはすでに記録されているので、ここには述べない。

### 3 我が国における情報学の展開：その展望

#### 3.1 初期から中期

情報学が元来は「電子計算機の活用」を基本のモチーフにしていることは既に述べた通りである。そしてその展開は既に述べたところから窺われるように、従前からの「知識」関連の3種の機関の流れに沿って進んだ。第一は専門図書館界が捉えた流れに沿った行き方で、FID 協調の路線である。このグループの活動も守旧派と進歩派に別れ、前者は、元来は北欧やドイツに主勢力を持つ UDC の関係者で、後者は、コンピュータの進歩の積極的導入をはかろうとする改革派であり、両者がそれぞれの立場から FID の改革を論じた。しかし 1980 年代には UDC は衰退し、さらには終焉に至った。我が国においても状況はそれと軌を一にする経過を辿った。

第二は科学技術情報センターの系統である。元来この種の機関は産業支援活動を主目的にしており、これは世界各国に共通の性格である。これに対して第三は、学術の専門分野の情報知識に関わる「情報学」関連の学術を推進しようとする流れである。

これらの三つの流れがそれぞれの性格を保ちながら整備を図った時期を「情報学の中期」と言おう。筆者の私見として言うと、この時期には「情報学」はこの三つの流れが混交して進行した。世界的にもそうであった。しかも同時に世界は、計算機利用の情報処理の大勢が大きく変化し続け、通信、ネット化に進んでおり、メディア関連の社会が大変化をしていた。しかも、その当事者は情報学の直接の担当者ではなかった。そして情報学の直接の担当者はこうした社会の大勢に流されて、「情報学」の振興、推進はいかにあるべきかにつ

いての議論や対応が不十分のまま推移したように見える。

FID の内部においては、主として米国の計算機関係研究者が激烈にこの状況を非とし、時代に沿った情報学を振興すべしと主張した。しかしながら、いかんせんこれらの意見は大勢にならなかった。CODATA もまた、化学や生物学の数値情報、化学構造解析、化学反応解析の論議が多く、それは英米加が主力であったが、これに対して、フランスの代表らが数値情報や実験直結の情報処理の本質的考察の重要性を指摘した。結局、英米の数値情報利用派の勢いが強く、こちらも「情報学」としては混迷した。

さらに大きく見ると、CAS を主体とする学術論文の情報処理派と米国標準局 (NBS)などを主体とする実験・数値派の対立は情報学の主要な二面を表し、どちらが情報学の主体になるかについて激越な論争があった。例えば純正応用化学連合 (IUPAC) は、1972 年に各部にまたがる化学情報委員会 Interdivisional Committee on Information Chemistry を設立した。委員長は仏パリ第 7 大学の J.-E. Dubois、副委員長は米国の CAS の Dr. Tate であった。小生は当初からこれに委員として参加したが、抄録事業重視の副委員長と実験データの交換重視の委員長とは終日激論する状況で、それは当のご本人が言った言葉で fight であった。実は小生は、東大大型計算機センターでこの両方の立場を融合する仕事を実現していた。これは米仏両国を始め関係者が刮目し、賞賛した事実であった。ただし、我が国内ではこれはあまり暖かくは迎えてもらえなかった。つまり、計算機を科学計算に利用したい研究者にとってこれは迷惑であったらしく、当事者はまたこの環境で苦勞した。

しかし、当方の関係者は立派な成果を挙げている。その二三を紹介しておきたい。山本毅雄氏は根岸正光氏らの協力者と東大大型計算機センターにあって米国の CAS がデータベース化した化学情報のオンライン処理システム TSIR-1 (Today Scientific Information Retrieval)<sup>[2][3]</sup>を我が国でいち早く実現



し、ついで TOOL-IR(Tokyo University On Line Information Retrieval System)<sup>[4]</sup>を1972年に作った。システムとしては先導の仕事であって、米国はNSF主導で共同セミナーの開催を提案してきた<sup>[5]</sup>。

化学研究への筆者の計算機利用は古い。最初は清水博氏と6スピン系のNMR高分解能スペクトルの解析<sup>[6]</sup>で、これは実に東大のPL-1を借りたものであった。しかもそれは、当時の世界の先端が4スピン系の解析にとどまっていた状況でのことであった。ついでNMRの実測データを大型計算機センターの計算機に結ぶことにまで伸び、分子内回転異性体の分率を計測したり、DEC社の計算機をESR分光計に接続して、スピララベル化合物の分子運動の実測などをしたりした<sup>[7]</sup>。これらは現在日常のことになったが、当時はかなりの刮目事で、とくにフランスのDubois教授は評価してくれた。また、この時期に生まれた米田幸夫氏のEROIKA<sup>[8]</sup>は有機化合物、有機金属化合物の属性を数値情報から推算するシステムであり、我が国で生まれた独創的な情報学の成果として特記さるべきである。

### 3.2 情報学関係行政

1980年代になると、FIDもCODATAも急激な計算機と通信産業の進歩や一般社会の情報関連産業の変化への対応に追われた。その中で、我が国の動向はユニークであった。すなわち、「情報学」への直接的支援としては、図書館情報大学の創立や、全国レベルの「情報学科」の創設、「学術情報センター」の創立があった。それに先だつては、1976年当時東大図書館長をしておられた安藤良雄氏のもとに情報図書館学研究センターが設立され、それを小生が引き継ぎ、根岸正光、井上如、柴田正美氏をメンバーとして「全国大学図書館学術雑誌総合目録作成事業」を開始した。これは画期的な大事業であった。すなわち、機械処理の対象たり得る雑誌目録データベースがこれで実現したのである。その上、その内容を詳細に見ると、我が国特有の綿密な目配り

が見られる。例えば、根岸氏による「学術雑誌誌名変遷マップ」などは、複雑なその委細まで丁寧に追跡した書誌データベースで、世界でも類を見まい。ともかく実にこれで我が国の図書館界は、初めて先進国レベル級の基盤整備をしたものと言える。ただし、当事者の労苦は多大で、傍で見ていて体力の限界かと危ぶませるものであった。このセンターは学術情報センターを経て、国立情報学研究所に引き継がれている。

これらの行政対応は、研究者の養成や専門分科としての情報学の充実を図る我が国の努力を明示し、これは例えば強力な米国の図書館学科例えばシカゴ大学のそれが閉鎖されるような環境のなかで、「情報学推進」における我が国の姿勢を強く印象付けるものであった。しかし、社会全体のその後の実情は変化が急で、情報学の関係者自身「情報利用」の立場に立つ米国流に流され、あるいは通信技術の革新的進歩に幻惑されて、その導入に忙殺された観がある。すなわち、「情報学の確立」の視線が主体になり得なかったように見える。

それはともかくとして、今の状況で思うと、現在の学会がなすべきことは「情報学」の確立であり、それを果たす最善、最速の道は「温故知新」であり、またその実行を学術の先端に直結して行うことであると思う。そこで指摘したいのは我が国が持つ知的資産としての「知識コード」と「分類」に関する知識の活用である。それは实际的であり、手っ取り早い方途であろうと思う。これについて少し説明したい。

## 4 提言「温故知新」

### 4.1 情報のコード

2.3と3.1で述べたFIDやUDCの動向は純学術的に見ればはなはだ不幸である。何故ならば、上記の経過には100年以上に及ぶUDC研究者、学究者による「知識の分類体系の構築」とその「コード化」に関する真摯な活動が存在したのであるが、それが消滅しよう

しており、しかも、その内容こそ、現代、将来の情報学の重要基盤になるものだからである。このことについて、もう少し述べよう。我が国でもっとも古くUDCを「情報知識」のコード化の最適手段であると認識して、それをほとんどマニア的に取り上げたのは、文部省にあって終始UDC事業の育成に尽力した石原紘氏である。ヨーロッパではUDCは使用ライセンス料を徴収するビジネスの一面を持ち、それは、情報学、ドキュメンテーション活動自体を複雑にした。我が国においては、UDCのライセンスビジネスまで含めた事業化を中村幸雄氏が実現した。このUDCが持つ学的資産は継承し活用されるべきである。

## 4.2 学術用語

我が国は、「学術用語」の標準化に世界でもっとも古くかつ内容に富む実績を持つと言える。つまり、蘭学の輸入に際し当時の知識人は、蘭学に記載された文物の知見を国語で表記せねばならなくなった。実物を知らないで、外国語で記載された内容の意味を考えて、当該物の国語の表記を作ってゆかねばならなかったのである。その際幸いだったのは、同時に中国も同様な状況にあり、中国の漢字表記を借りることが出来たことである。しかもこれにはもっと古いところからの由来がある。

元来、物質のデータベースは中国の「本草學」(ほんぞうがく)がはじまりであり、その最古の文献は『周易参同契』(しゅうえきさんどうけい、後漢AD2)とされ、中国も我が国も物質の記録は、その伝承である『抱朴子』(ほうぼくし、AD317)にずっと依存してきた。延喜の時代に我が国の国勢が伸張し、自主、自立の意識が高まり、我が国の国産の物産も記録することになり、深根輔仁(ふかねすけひと)が『本草和名』を著した。つまり、この当時から、国語による「用語」表示の苦心があったことが窺わせる。そしてこれは、日本最古の物質データベースなのである。

近代になって「標準学術用語」の必要が認識され、その作成の機運も高まり、とくにそ

れは化学で顕著で明治20年代には日本化学会の手で実際の仕事が行われた。文部省もそれに呼応し、学会と協力して熱心にこれを推進した。我々の時代でこの面で特記すべきは、大塚明朗氏、また文部省サイドで献身的努力を傾けられた青戸邦夫氏である。特に青戸氏の献身的努力と各学協会の協力で、40に近い学術部門別のそれぞれに標準学術用語集が出来たことは世界に類のない偉業であり、我が国の宝と言える。遺憾ながらこの成果に対して軽率な負の評言をよく聞くが、大局的に見たときの学術的意義を見失っている。

すなわち、よく考えれば用語は一朝一夕に、また一人の人の思いつきで出来たものではない。大げさに言えば、平安時代の文化人が努力して作ったものであり、元は中国に根をおいている。それに明治期の近代科学が接ぎ木されたものなのである。そう理解するならば、「標準学術用語」は現代の目で活かす道を考えるべきなのである。筆者は藤原譲氏と協力してこの意義を活かすために少し努力した。

藤原譲氏と筆者は共同で、FIDの全理事の協力を得て、「多分野共通・多言語表現標準科学技術用語集」の作成を図った。すなわち、我が国が作った25部門の標準学術用語20万語を同一ファイルに集め、この中で2部門共通の用語を採集し、それを両藤原が眺めて重要語5千語を選定した。これで「多分野共通・重要標準学術用語」が選定された。そこでFIDの理事にそれぞれの出身母国の国語訳をこれに付けることを依頼し、それで英独仏西訳がついた。英語訳は元来我が国版にあるが、他国の訳語がついた段階で英国人を見ると、彼我に理解の相違が見出される例があった。その他にも同一語に対するそれぞれの国語訳に意見の相違が出るものがあり、これらは適当に折り合わせた。これは、「機械検索」の現代において初めて可能となる国家民族の特性解析であり、情報学による社会科学的研究になると思った。

この成果は神奈川大学知識情報研究所から研究所報告として出版した。これは当方が予期せぬ働きもした。中曽根元首相が、首相在

任時に欧米に出張された時この報告書を購入し、出張先への土産の文化交流資料として利用して下さったのであった。

### 4.3 数値データの国際コード提案

我が国の情報学の先達である小谷正雄先生は、CODATAの会長ご在任中にこの提案をされた。提案の内容は、数値データの性質を熟慮された結果の分類であり、広く流通するまでには至らなかったが、有益な国際貢献である。

JSIKは上記のような実績の上に立っている。そこで温故知新を学会の活動目標に採ってはどうか。まずは、繰り返すようだが、先人の努力の結晶である標準学術用語の活用である。これは古いと言う人があるが、それは間違いであろう。これは日本のみが持つ世界的財産である。だからこれを生かす道を探るとよい。さらには、「知識」のコード化で実績があり世界遺産であるUDCの活用である。その他温故知新の視点による会員諸賢の活動を期待したい。

#### 補遺付言：

##### A 近代科学の情報学的解析

会員の活動を期待する対象を挙げれば、たとえば米国のNSFは10年前に、今後の科学研究の目標として、「宇宙」、「真実」、「生命」、「こころ」の解明を掲げた。これを起草したのは現在NSFで研究担当長官を勤めるDr. P. J. Werbosである。ちなみに筆者は先年同氏に、10年前のこの目標が現在実現していることを称揚し、現在の氏の心境を問うた。そうしたら、氏は当方の称揚に謝意を述べ、同時にいまのNSFは「この総括期」にあると言われた。そこで思ったのは「これこそ現代科学の総括だ。ならば、「科学の科学」の意識で当学会は、情報学の視点で上記の4テーマをキーワードとして、現代科学技術の用語の総括をしたらどうか。それこそ現代科学の総括になる仕事になろう」ということであった。

## B 情報経済学

情報学は個々の専門分野にまたがる学術分野である。その意味では個々の分野の先端の動向にも注意が必要である。これについて筆者は苦い思い出がある。

東大で筆者のもとで学位論文をまとめたDr. A. Steuartは在学中は電波分光の研究であったが、丁度当方の情報処理システムの研究が進行中で、同氏はそれを吸収し、母国カナダに帰国後は、順次、大学、州政府、中央政府の情報解析部門に進み、結局、OECDの経済統計局長になった。小生がJSIK会長在任中の1992年ころ、「学会活動の一部として、世界経済情報の解析・配信をするOECDとの共同作業をしないか。合意するならOECDは東京部局を設置する」と提案してきた。先方は十分具体化の目算を持ってのことであったようで、来日して自身で外務省、文部省、神奈川大学などを周り、小生も理事会に提議したが、結局実現出来なかった。この動きは経済学への情報学応用とも考えられるブラック・ショウルズの理論(1997年ノーベル経済学賞)を捉えた動きであったかとも思われる。

#### 参考文献

- [1] S. Fujiwara: "Information Chemistry in Japan", J. Chem. Inf. & Comp. Sci., 25(3), pp.207-11, 1985.
- [2] Y. Yamamoto; T. Kumai; K. Nakano; C. Ikeda; T. L. Kunii; H. Takahashi; S. Fujiwara: TSIR-1 I, J. Chem. Doc., 11(4), pp.228-31, 1971.
- [3] T. Yamamoto; M. Ushimaru; T. L. Kunii; H. Takahashi; S. Fujiwara: TSIR II, J. Chem. Doc., 12(2), pp.113-6, 1972.
- [4] T. Yamamoto; M. Negishi; M. Ushimaru; Y. Tozawa; K. Okabe; S. Fujiwara: "TOOL-IR", USA-Japan Computer Conference Proceedings, 2nd., pp.159-65, 1975.

- [5] H. B. Mark, Jr.; S. Fujiwara: "Computer-assisted Chemical Research Design", University of Tokyo Press, 1975 Tokyo; "Second Joint Japan-United States Seminar", J. Chem. Inf. Comput. Sci., 17(4), pp.232-4, 1977.
- [6] S. Fujiwara; H. Shimizu: "Analysis of the Nuclear Magnetic Resonance Spectrum of a Six-Spin System of Monofluorobenzene", J. Chem. Phys. 32(6), pp.1636-1642, 1960.
- [7] T. Watanabe; T. Yahagi; S. Fujiwara: "EPR studies of the molecular motion of nitroxide radicals in poly (vinyl alcohol) gels", J.A.C.S., 102(16), pp.5187-5191, 1980.
- [8] Y. Yoneda: Data for science and technology: Proceedings of the Seventh International CODATA Conference, Kyoto, Japan, 8-11 October 1980: (ed. Phyllis S. Glaeser), Pergamon Press, pp.254, 1981.

(2005年11月29日受付)

(2005年12月26日採択)

## 情報知識学会第 14 回 (2006 年度) 研究報告会 発表論文募集等について

実行委員長 田窪 直規 (近畿大学)

来年度、研究報告会・総会を大阪の近畿大学で開催致します。2日間の日程で、初日は午後から開催し、夕刻には懇親会も設ける予定です。会員が交流できる楽しい会にしたいと念じておりますので、皆様のご発表とご参加を心よりお待ちしております。会場は大阪のターミナルの一つである難波(「ミナミ」)に比較的近いロケーションにあります。なお、できるだけ4月中にプログラムの詳細を決めて、当学会サイト及びMLで、学会員の皆様にお知らせいたします。

予定期日：2006年5月26日(金)、27日(土)

予定会場：近畿大学B館10階MM会議室(会場が変更になりましたのでお気をつけ下さい)

〒577-8502 東大阪市小若江3-4-1

近鉄大阪線長瀬駅下車徒歩約15分(新大阪駅から約50分)

近畿大学ホームページ：<http://www.kindai.ac.jp/>

研究報告会の発表論文の募集要項を以下に記しますので、奮ってご応募下さい。

### 1. 募集分野

- (1) 情報知識の構造解析, モデル化, 可視化, 知識発見
- (2) 情報・知識の表現, 生産, 組織化, 検索, 提供
- (3) 分類, シソーラス, オントロジー, ターミノロジー
- (4) メディアの記述, メタデータ, マークアップ言語
- (5) 電子出版, 電子図書館, 電子文書館, 電子博物館, Web
- (6) マルチメディア
- (7) 知識情報の流通と知的所有権
- (8) 専門分野における品質管理, 標準化
- (9) その他の情報知識学関連分野

### 2. 応募方法

発表論文題目, 著者名(連名の場合, 登壇発表者に○印), 所属, 論文概要(200字以内), 連絡代表者の氏名, 住所, 電話・FAX番号, 電子メールアドレスを明記の上, 下記宛になるべく電子メールにてお申し込みください。

〒577-8502 東大阪市小若江3-4-1 近畿大学 短期大学部

川原亜希世 (第14回(2006年度)研究報告会事務局)

TEL: 06-6721-2332(内5516) FAX: 06-6728-7546 E-mail: akwhr@kindai.ac.jp

応募期限：2006年2月28日(火)

採択可否通知：2006年3月10日(金)

原稿提出期限：2006年4月8日(土)

### 3. 論文執筆・発表について

- (1) 論文執筆要領の詳細については、次ページに掲載しておりますので、ご参照ください。
- (2) 発表時間は、質疑応答を含めて30分を予定していますが、発表本数によって伸び縮みする可能性があります。なお、論文提出がないと発表はできません。
- (3) 登壇発表者は当学会員に限ります。当日入会も可能です。

## 情報知識学会・第14回(2006年度)研究報告会 予稿集原稿執筆要領

1. 原稿はワープロ A4 判で作成してください。そのまま B5 判に縮小撮影し、印刷・製本しますので、図・表等は、あまり小さくしないでください。
2. 頁数は1題に2~6頁で、必ず偶数頁とします。4頁までは無料ですが、それ以上は2頁ごとに¥2,000の追加料を頂きます。
3. A4判1頁に横1段組、1行全角40文字、40行が目安です。余白は天地左右各25ミリを目安とします。頁番号は入れないでください。
4. 原稿1頁目に論文題目、発表者氏名を日本語・英語で併記してください。題目(文字は本文より大きくしたり、ゴシックに変えたりしても構いません)は第1~4行目に日本語で、第5~8行目に英語で記入し、発表者氏名(連名の場合は登壇発表者の前に丸印を付す)は、第9行目から右寄せで記入してください。第10行目から論文概要を記述し、その後本文が続きます。
5. 論文執筆には、日本語または英語を使用してください。主文に日本語を使用する場合、概要は英語(300ワード程度)とし、主文が英語の場合は、概要を日本語(500字程度)とします。また、英語と日本語で同一内容の論文を併記することも可能です。この場合は英文を先にし、図版は英文部分に挿入してください。なお、この場合も、仕上がり原稿の枚数規定に変更はありません。
6. 図表は原稿に直接書き込むか、または鮮明なコピーを張り込んでください。図のキャプションは図の下部に、表のキャプションは表の上部に記入してください。写真も掲載可能ですが、カラー印刷はできません。
7. 提出はいずれかの方法でお願いします。
  - (a) PDF ファイルを電子メールで送る。
  - (b) MS-Word ファイルを電子メールで送り、印刷したのも FAX で送る。  
(MS-Word はバージョンによって体裁が異なることがあるため)
  - (c) 紙に印刷したものを郵送する。(質的にはこれが一番確実です。)

※ L<sup>A</sup>T<sub>E</sub>X 形式などは受け付けません。一旦、PDF 形式に変換して、a の方法でお願いします。
8. 原稿は返却しません。研究報告会終了後、情報知識学会誌またはホームページに題目・発表者・論文概要を掲載する場合がありますので、ご了承ください。予稿集に関し、著作権は原則として本学会に帰属することとします。予稿集に掲載された執筆内容が第三者の著作権を侵害するなどの指摘がなされ、損害を与えた場合は執筆者がその責任を負うことになります。
9. 応募論文で上記の条件に適合しないものは一切受理しません。提出期限は平成18年4月6日(木)です。提出先は情報知識学会事務局ではなく、下記宛です。  
〒577-8502 東大阪市小若江3-4-1 近畿大学 短期大学部  
川原亜希世 (第14回(2006年度)研究報告会事務局)  
TEL: 06-6721-2332(内5516) FAX: 06-6728-7546 E-mail: akwhr@kindai.ac.jp

### その他の発表に関する連絡事項

- ・ 発表者には、会場受付で予稿集(¥3,000)を1部差し上げます。さらに部数が必要な方には発表者、共著者ともに割引価格1部¥300で販売します。10部以上ご希望のかたは、印刷の都合がありますので、4月6日までにjsik@nifty.com(情報知識学会事務局)へお知らせください。別刷りはしません。
- ・ パソコンを使用する発表者のために、会場にプロジェクターを用意しますが、接続するパソコンはご自分でご持参ください(なお、念のために、併せて、プレゼンテーション用のデータをUSBメモリなどに入れてお持ちください)。インターネットに接続される方は、念のためにPHSカードなどをご用意ください。
- ・ プログラムの詳細は、できるだけ4月中にメールでお知らせするとともに、情報知識学会のホームページ(<http://www.jsik.jp/>)に掲載します。

## 特集 「情報倫理 (情報倫理学・技術倫理／工業倫理等)」 の論文募集

Vol.16, No.3 (2006年7月刊行予定)

情報知識学会誌編集委員 二階堂 善弘 (nikaido@ipcku.kansai-u.ac.jp)

岡本 由起子 (okamoto@kasei-gakuin.ac.jp)

### 1. 主旨

近年は「情報倫理」といってもいわゆるネットワークの不正使用や著作権の問題、ハッキング、個人情報保護といった個別の諸問題だけでなく、情報社会における人間のあり方そのものが問われるようになって来ました。情報技術は、技術者とその技術の利用者という2項関係を超えて人間社会の基盤そのものにも浸透して来ました。もはや、情報技術は、作り出され、それ利用するといった単なる道具ではなく、むしろその中に暮らす、また新しい世代はその中に生まれ育つ「環境」の一部となっています。米国では Virtual Environments という言葉が人口に膾炙しています。仮想世界については「代替世界症候群」というものが報告されたこともあります。「工業倫理」においても技術者教育にとどまらず、「科学技術倫理」という広い視野から考えてみようという試みが始まっています。哲学・倫理学の立場からだけでなく、技術者の立場から、あるいは企業家の立場、また社会学的教育学的立場から、具体的な事例に則した報告や原理的考察等、広く「情報倫理」について関心のある方の論考を募ります。

### 2. 公募のテーマ

- (a) 情報倫理 (学)      (b) 技術 (者) 倫理      (c) 「工業倫理」の情報倫理への適用  
(d) ネットワーク倫理      (e) 技術者教育      (f) 仮想世界と子供の教育 等。

### 3. 期限および原稿送付先

投稿原稿締め切り：2006年4月7日(金)

修正原稿締め切り：2006年5月12日(金)

### 4. 原稿送付先

〒110-8560

東京都台東区台東 1-5-1 凸版印刷(株)内 情報知識学会事務局

Tel: 03(3835)5692 FAX: 03(3837)0368 E-mail: jsik@nifty.com

## 情報知識関係新刊図書一覧 2005年7～12月

- 『アメリカー コミュニケーション研究の源流』 E. デニス/E. ウォーテラ編著 春風社 4286円  
『ユビキタス社会のキーテクノロジー』 東倉洋一/山本毅雄著 丸善 760円  
『モバイル社会白書』 NTTドコモ モバイル社会研究所企画・監修 NTT出版 3800円  
『人工知能』 本位田真一監修 松本一教共著 オーム社 2500円  
『検索エンジン戦争』 ジェフ・ルート/佐々木俊尚著 アスペクト 1500円  
『ビジネス情報の法とセキュリティ』 松田貴典著 白桃書房 3900円  
『情報学の楽しさ』 東京情報大学「情報学の楽しさ」編集委員会編 東京農業大学出版会 1600円  
『知識の構造化と知の戦略』 齋藤雄志著 専修大学出版局 2900円  
『新聞がなくなる日』 歌川令三著 草思社 1400円  
『IT・ネット社会データ総覧 2005』 日本能率協会総合研究所編集 生活情報センター 14800円  
『オントロジ技術入門』 AIDOS編著 東京電機大学出版局 2200円  
『ユビキタス社会の情報モラル』 宮田仁監修・編著 尾賀聡一郎ほか著 一橋出版 520円  
『個人情報保護とリスク分析』 畠中伸敏編著 加藤文也著 日本規格協会 2800円  
『意味を生み出す記号システム』 加藤雅人著 世界思想社 1800円  
『図書館・情報学研究入門』 三田図書館・情報学会編 勁草書房 2700円  
『デジタルメディア時代の〈方法序説〉』 船木亨著 ナカニシヤ出版 2500円  
『図書館と著作権』 名和小太郎/山本順一編 日本図書館協会 1700円  
『情報化白書 2005』 日本情報処理開発協会編 コンピュータ・エージ社 6000円  
『デジタル情報クライシス』 中島洋編著 日経BP企画 日経BP出版センター 1600円  
『ザ・サーチ グーグルが世界を変えた』 ジョン・バツテル著 日経PB社 1800円  
『人工知能学事典』 人工知能学会編 共立出版 22000円  
『情報メディア白書 2006』 電通総研編 ダイヤモンド社 15000円  
『2010年のITロードマップ』 野村総合研究所技術調査室著 東洋経済新報社 2200円  
『ケータイ研究の最前線』 日本記号学会編 慶應義塾大学出版会 2800円  
『情報学的転回』 西垣通著 春秋社 1800円  
『インターネットにおける言語と文化受容』 C&C振興財団編 上村圭介著 NTT出版 2800円

(注) 図書館流通センター刊『週刊新刊全点案内』より収録。コンピュータの技術・操作関連のもの、初心者入門的なもの、またビジネス色の強いものは除外した。

制作者：平田 周



## 情報知識学会・平成 18 年度役員選出について

情報知識学会会員 各位

本学会の役員選出規定に基づき、情報知識学会誌 15 巻 4 号誌上にて正会員の皆様に平成 18 年度役員候補者の推薦を募りましたところ、締切日の昨年 11 月末日までに選挙管理委員会へ 10 名の被推薦者名が届きました。

以下の要領で平成 18 年度役員選挙を実施します。正会員は、1 月下旬に郵送でお届けする投票用紙(葉書)へ候補者名をご記入のうえ、平成 18 年 2 月 15 日(水)までに選挙管理委員会へ返送してください。

|               |     |      |
|---------------|-----|------|
| 情報知識学会選挙管理委員会 | 委員長 | 深見拓史 |
|               | 委員  | 孫 媛  |
|               | 委員  | 相田 満 |
|               | 委員  | 相良佳弘 |
|               | 委員  | 高橋仁一 |

1. 正会員には選挙権ならびに被選挙権があります。
2. 投票用紙(葉書)とともにお届けする正会員一覧(約 300 名)、被推薦者一覧(10 名)、現役員一覧(29 名)を参考にし、会長 1 名、副会長 2 名、理事 5 名、監事 1 名を上限として投票用紙へご記入ください。記入名の合計が 9 名に満たない投票も有効とします。
3. 投票は無記名投票です。選挙管理委員会へ本年 2 月 15 日(水)必着でお願いします。
4. 投票結果は 5 月総会のご案内状に同封しお届けします。総会の承認を経て新役員(任期 2 年)が確定します。
5. 定款および役員選出規程の全文は学会誌 15 巻 3 号および本学会のホームページ(<http://www.jsik.jp/>)に掲載してあります。

## 事務局からのお知らせ

### [1] 平成 18 年度年会費の納入をお願いします

平成 18 年度 (2006 年 4 月 1 日～2007 年 3 月末日) の年会費を、郵便局または銀行の下記口座へお振込願います。1 年分の年会費は正会員 8 千円、学生会員 4 千円です。前年度分未納のかたは合計額を納入してください。請求書が必要なかたは、その旨、事務局へお知らせ願います。

#### 1. 振込先 (振込手数料はご本人負担でお願いします)

- a. 郵便振替口座 00150-8-706543 情報知識学会 (代表 細野公男)
- b. 三菱東京 UFJ 銀行 秋葉原駅前支店 普通預金 3586133 情報知識学会 (会長 細野公男)

#### 2. 納入した年月日の確認方法

情報知識学会から郵送された封筒の宛名ラベルをご覧ください。[ ] 内に過去 4 年間、ご自分の納入日が印字されているので確認できます。納入年 (西暦の下 2 桁)、月 (2 桁)、日 (2 桁) の 6 桁です。年会費を滞納している場合は、[未納] と表示してあります。金融機関へ振り込まれてから事務局へ通知が届き、宛名ラベルに印字、発送するまで 10 日ほどかかりますので、ご了承ください。

### [2] 最近 1 ヶ月以内に事務局からのメールを受信しなかったかた

現在、8 割以上の会員がメールアドレスを事務局へ登録されています。各部会の活動や月例懇話会の予告・報告など、頻繁に受発信しており、電子メールは必須の連絡手段となりました。過去に登録されても、最近 1 ヶ月以内に情報知識学会事務局からメールを 1 通も受信しなかったかたは、不達が予想されますので、再度、アドレスを事務局 [jsik@nifty.com](mailto:jsik@nifty.com) へご連絡ください。添付ファイルが開けないかたも、お知らせくださればテキスト文に直して送信します。

### [3] 電話でのお問い合わせ

事務局の業務は土日祝日を除き、月曜から金曜日までの毎日行っています。お問い合わせなどの電話は、できるだけ午後 1 時半から 5 時までをお願いします。連絡には電子メールや FAX も、どうぞご利用ください。

情報知識学会事務局

〒110-8560 東京都台東区台東 1-5 凸版印刷 (株) 内  
TEL: 03-3835-5692 FAX: 03-3837-0368  
E-mail: [jsik@nifty.com](mailto:jsik@nifty.com) URL: <http://www.jsik.jp/>

## 「情報知識学会誌」投稿規定

2002年8月27日 制定

2003年3月19日 一部改訂

0. 情報知識学会誌編集規程による本会機関誌「情報知識学会誌（以下、会誌という）」への投稿に関する事項は、この規定の定めるところによる。

### 1. 投稿資格

投稿者の少なくとも1人は本会員でなければならない。ただし、編集委員会による依頼原稿の場合にはこの限りではない。

### 2. 投稿原稿

2.1 広い意味での情報知識学に関連し、またその発展に貢献するもの（情報／知識の収集、整理、蓄積、検索および各種解析、利用などに関するもの）とする。刊行時において未発表の原著でなければならない。本会誌の記事の種類を以下に示す。

2.2 投稿者は会誌記事の種類を明記して投稿しなければならない。ただし、編集委員会に変更することがある。

(1) 研究論文 (Research Paper)：オリジナルな研究論文で、内容の主要な部分が学術論文として他に公表されていないもの。

(2) 事例／調査報告 (Report)：情報知識学に関連したシステムなどの開発、利用、調査に関するもの。資料も含む。

(3) 解説／展望 (Review)：情報知識に関連した特定分野の論文や学説などを総括、解説、紹介、あるいは技術動向などを展望したもの。技術、研究上の処理、解析方法などに関する解説。

(4) 論談 (Proposal Paper)：情報知識学に関連した新たな意見の表明、提案など。

(5) 討論 (Discussion)：本会誌に掲載された論文についての学術的な討論。

(6) 研究速報 (Notes)：技術、手法、新事実などの簡単な報告。

(7) 講座 (Lecture)：情報知識学の各分野に関する基礎理論、技術の適用などについて、テーマを定めて系統的に説明するもの。

(8) 学会記事 (News)：本会の事業、運営などの報告、記事、資料など。

(9) ニュース、お知らせ (News)：ニュース、お知らせ。最近刊行された単行本やモノグラフの紹介。

(10) 講演 (Lecture)：特別号などにおける講演資料。

(11) その他：編集委員会が適当と判断したもの。

2.3 会誌記事の種類のうち、(1) から (6) までは査読を行う。その他については編集委員会編集を行う。

### 3. 投稿原稿

#### 3.1 原稿の形式

##### (1) 投稿時の原稿

以下の A, B のいずれかの体裁でプリントされたワープロ原稿（横書き）4部の提出とする。その他、執筆に関する詳細は「執筆要領」を参照のこと。

A 刷り上り原稿を想定したレイアウト (A4判, 2段組, 20字×46行×2段)。

図, 表は希望の位置に配置すること。

B ベタ打ち原稿 (A4判, 40文字×40行)。

図, 表は, 1枚ずつ別の用紙に印刷すること。ベタ打ち原稿右余白に図表の挿入位置を朱書きすること。

(2) 採択決定後の原稿

以下の C, D の両方の形式で記録された電子媒体1部の提出とする。

原稿の送付にあたってはフロッピーディスクなどの適当な電子媒体とする。詳細は提出時に事務局に相談のこと。

C Microsoft Word, 一太郎, DVI, PDF などの代表的なフォーマット。

D 図, 表は十分な品質で印刷できる形式 (JPEG, GIF など)。

### 3.2 原稿の制限

(1) 原稿の長さを原則として次のように制限する。

研究論文, 事例/調査報告, 解説/展望, 論談 : 刷り上がり 20ページ以内

討論, 研究速報, 講座 : 刷り上がり 6ページ以内

ニュース他 : 刷り上がり 2ページ以内

(2) 図原稿 (原図) の大きさは A3判を越えないものとする。

(3) 原則として, 図版も含めてモノクロ印刷とする。ただし, カラーでなければならぬ図版を使用する場合は, 別途編集委員会と相談する。なお, カラーページやページを超過する分については, 印刷費を著者の全額負担とする。

(4) 使用言語は日本語または英語とする。

### 4. 原稿の採否

投稿原稿の採否は, 専門家による査読の後, 編集委員会において決定する。不採択となった原稿は, 編集委員長より理由を付して通知する。

### 5. 査読のプロセス

学会員の中から編集委員会が指名した査読者2名によって査読を行う。内容によっては, 編集委員会は著者に照会し, 原稿の修正を求めたうえで, 再査読を行うことがある。

### 6. 校正のプロセス

採択が決定した投稿原稿は, 掲載原稿として著者に校正を依頼する。著者による校正は原則として1回とする。その際, 字句の修正以外は原則として認めない。

### 7. 別刷

別刷 (抜刷) は著者の実費負担とする。希望部数を事務局に申し出ること。

### 8. 投稿の手続き

原稿投稿時には下記の書類を添え, 原稿送付先に郵送する。

#### 8.1 必要書類

(1) 最初の投稿時 (a, b, c の3つが必要です。)

a. 投稿原稿整理カード : ホームページからコピーして, 必要事項を記入し, 印刷したものを1部。掲載原稿整理カードと兼ねるので, コピーを保存しておくこと。

b. 紙媒体の原稿 (図, 表を含む) : 4部。

なお, 投稿者は著者校正用に原稿のコピーを保存しておくこと。

c. E-mailによる連絡票

・論文種別, 標題 (和文, 英文), 著者名 (和文, ローマ字), 所属機関/住所 (和文, 英文), 要旨 (和文, 英文). キーワード (和文, 英文). 刷り上り予定ページ数  
・連絡先: 著者1名の連絡先 (氏名, 所属機関/部局, 所属機関住所, 電話番号, Fax番号, E-mail アドレス).

なお, 投稿後の連絡は主として E-mail で行う.

(2) 採択決定後の投稿

- a. 掲載原稿整理カード: 投稿時のカードに追加事項を記述し, 印刷したものを1部.
- b. 3.1(2)に指定した電子媒体: 1部.
- c. 3.1(2)のCの印刷出力 (プリントアウト): 1部.

8.2 原稿の送付先

〒110-8560 東京都台東区台東 1-5-1 凸版印刷(株)内  
情報知識学会事務局  
電話: 03-3835-5692 Fax: 03-3837-0368  
E-mail: jsik@nifty.com

8.3 原稿の受付

事務局が原稿を受け取った日を受付日とする. 受付の確認を1週間以内に投稿者の連絡先に E-mail で通知する. 不備のある投稿原稿は返送し, 再提出するものとする.

9. 原稿提出期日

投稿は随時とする. ただし, 特集号などは除く.

10. 著作権

- 10.1 機関誌『情報知識学会誌』に掲載された論文 (電子版を含む) の著作権 (著作財産権, copyright) は情報知識学会に帰属する.
- 10.2 掲載論文は冊子による出版の他, 電子的に蓄積し, 本会が行う情報提供サービスなどを通じて公開する.
- 10.3 本学会誌に掲載された執筆内容が第三者の著作権を侵害するなどの指摘がなされた場合には, 執筆者がその責任を負う.

11. 規定の改訂

11.1 本規定の改訂は, 編集委員会の議を経て, 理事会の承認を得なければならない.

12. 施行

- 12.1 本規定は2002年8月27日より施行する.
- 12.2 本規定の施行により, 現行規定 (第4版 (暫定板) 2002年3月) は廃止する.

13. 改訂履歴

2003年3月19日 一部改訂. 「10. 著作権」に, 10.3項を追加.

## 「情報知識学会誌」執筆要領

2002年8月27日 制定

2003年5月2日 一部改訂

### 1. 一般的な事項

本会誌への投稿は、「投稿規定」に従い、投稿原稿は本執筆要領に従って作成されなければならない。

本会誌の投稿原稿の種類には、研究論文、事例／調査報告、解説／展望、論談、討論、研究速報、講座、本会記事、講演、ニュース、その他がある。

### 2. 日本語原稿の構成

#### 2.1 全体構成

##### (1) 第1ページ（査読者には見せない）

- ・ 標題（和文および英文）
- ・ 著者名（和文およびローマ字、ローマ字による著者名は、名、姓の順で、姓は全て大文字を使用する。）
- ・ 所属（和文および英文による所属機関名）
- ・ 住所（和文による所属機関の住所、E-mail、脚注とする。）
- ・ 見出し用原稿（研究論文、事例／調査報告、解説／展望、論談の原稿には、刷り上がりページ上部欄外につける著者名および標題を30字以内で書く。）

##### (2) 第2ページ目以降（査読者に見せる）

- ・ 要旨（研究論文、事例／調査報告、解説／展望、論談の原稿には、和文および英文で要旨をつける。和文要旨の長さは400字以内とする。英文要旨の長さは200語以内とする。要旨中には、図、表、数式などを用いない。本文中の図、表、数式、文献などを番号で引用しない。）
- ・ キーワード（研究論文、事例／調査報告、解説／展望、論談、討論、研究速報、講座にはキーワードをつける。和文および英文でそれぞれ5個程度、和文と英文のキーワードは、対応することが望ましい。キーワードはカンマ（,）で区切る。）
- ・ 本文（和文または英文）
- ・ 文献、付録など（和文または英文）
- ・ その他（とくに長い論文の場合、読者の便宜を考えて内容目次を付してもよい。ただし、章、節の見出し程度とする。）

#### 2.2 本文（Body）

##### (1) 構成

章、節などの構成は、第1レベルは1, 2, …, 第2レベルは1.1, 1.2, …, 第3レベルは1.1.1, 1.1.2, …のようにする。

##### (2) 脚注

脚注はできるだけ避ける。止む無く使用する場合は簡潔な文とする。

##### (3) 図および表

- a. 図、表にはそれぞれ通し番号をつける。図1 (Fig.1), 図2 (Fig. 2), … 表1 (Table 1), 表2 (Table 1), … など。
- b. 通し番号とともに説明文（キャプション）をつける。キャプションの位置は図は下部に、表は上部とする。

- (4) 数式, 化学式
  - a. 数式 (独立式), 化学式は, 段落外で記述されているものも本文中で一回は参照する.
  - b. 数式には, 通し番号を振る.
- (5) リスト (または箇条書き)
  - a. 記号なしリスト.
  - b. 記号つきリスト. リストの記号は, 数字, アルファベット, 記号を用いることができる. ただし, これらの混在した使用は避ける. アルファベットは1論文中では大文字, 小文字の使い分けをしない.
  - c. 複雑化を避け, せいぜい2段 (親子関係) のリストとし, ネストを跨ぐ順序づけを用いない.
- (6) 注記および参考文献  
本文中で少なくとも一回は参照すること. 通し番号で参照し, タイトルなどでの参照は避ける.

### 2.3 後付け (End)

- (1) 謝辞  
本文の最後に続けて記述する. 章番号は用いない. 章題は「謝辞」とする. 最終原稿時に記述することが望ましい.
- (2) 注記および参考文献
  - a. 注記または参考文献には, 参照順に通し番号を付し, 本文の最後に番号順にまとめて記述する. 章番号は持ちない. 章題は「参考文献」とする.
  - b. 1つの番号には1つの注記または参考文献を対応させる.
  - c. 注記中には参考文献を含めない. 注記はできる限り簡潔に表現すること.
  - d. 参考文献の記述形式は, 以下の形式を満たさなければならない.
  - e. URLを参照してもよいが, 移動または削除される可能性があるため, 極力避ける. 原著がURLでのみしか参照できない場合など, やむをえない場合は用いてもよい. その場合, 参照時点でのハードコピーを保管しておくなど, 参考文献へのアクセス手段を確保するよう努力しなければならない.

#### 【参考文献の形式】

- 1. 雑誌中の1論文  
[引用通し番号] 著者名: 論文名, 雑誌名, 巻号, 掲載ページ, 出版年, その他.
- 2. 図書1冊  
[引用通し番号] 著者名: 書名, 版表示, 出版地, 出版社, 総ページ数, 出版年, その他.
- 3. 図書の1部  
[引用通し番号] 著者名: 論文名, 書名, 版表示, 出版地, 出版社, 掲載ページ, 出版年, その他.
- 4. 会議報告  
[引用通し番号] 著者名: 論文名, 書名 (会議名), 版表示, 編集者名, 会議開催地, 会議開催年, 会議開催機関, 出版地, 出版社, 掲載ページ, 出版年, その他.
- 5. インターネット上の論文  
[引用通し番号] 著者名や標題など可能な限り詳細な書誌事項, URL, 参照年月日. (単なるホームページなどは参考文献にしないこと).

【参考文献の記述】

1. 著者名，編集者名の記述

- (1) 個人著者名は，姓，名の順に記述する．欧文著者名は，カンマ（，）で姓，名を区切る．
- (2) 複数著者の場合は，各著者をセミコロン（；）で区切る．
- (3) 翻訳図書などの翻訳者名の場合は，著者名の後に括弧（）に入れて記述する．

2. 論文名，書名の記述

- (1) 論文名，書名は，和文の場合はかぎ括弧（「」），欧文の場合はダブルクォーテーション（“ ”）に入れて記述する．
- (2) 図書中の一部を引用した場合の書名は，和書の場合は二重かぎ括弧（『』）に入れ，欧文の場合はイタリック体で記述する．

3. 掲載ページの記述

- (1) 論文の場合は，開始ページと終了ページを記述する．「pp. 開始ページ-終了ページ」とする．
- (2) 図書の場合は，総ページ数とする．「総ページ数 p.」とする．

【参考文献の記述例】

- [1] 藤原 譲：「情報知識学試論」，情報知識学会，Vol.1, No.1, pp.3-10, 1990.
- [2] 原正一郎；安永尚志：「国文学研究支援のための SGML/XML データシステム」，情報知識学会，Vol.11, No.4, pp.17-35, 2002.
- [3] Fujiwara, Shizuo: “East-West Communication and Information Transfer — Coordination of Specificity”, Journal of Japan Society of Information and Knowledge, Vol.4, No.2, pp.11-18, 1994.
- [4] Ellis, David (細野公男監訳，斎藤泰則，鈴木志元，村上泰子訳)：「情報検索論」，丸善，180p., 1994.
- [5] 根岸正光：「学術情報の流通と利用」，『情報学とは何か』情報学シリーズ 3, 丸善, pp.43-69, 2002.
- [6] 名和 小太郎：「デジタル図書館と著作権」，デジタル図書館，No.4,  
[http://www.dl.ulis.ac.jp/DLjournal/No\\_4/nawa/nawa.html](http://www.dl.ulis.ac.jp/DLjournal/No_4/nawa/nawa.html) (2002年8月27日参照)

3. 文章と文体

- 3.1 文体はひらがなと漢字による口語常態（である調）とし，現代かなづかいを用いる．
- 3.2 漢字は当用漢字とする．ただし，固有名詞や学界で広く用いられている慣用の術語はこの限りではない．
- 3.3 句読点その他には「，」「。」を用いる．
- 3.4 本文中の人名には敬称をつけない．ただし，謝辞の人名はこの限りではない．
- 3.5 数量を表す数字はアラビア数字とする．
- 3.6 数式は印刷に便利のように十分注意して記号を記すこと．原則として数量（変化量）を表す記号はイタリックとする．
- 3.7 ローマ字の人名の姓は大文字体とする．
- 3.8 固有名詞で読み誤るおそれのあるものにはふりがなをつける．
- 3.9 英数字は原則として半角英数文字で記述する．



4. 英文原稿

英文による投稿原稿の場合も、原則として和文による投稿原稿の諸規定に従う。英語圏以外の著者の場合、著者名表記にその国語による表記を認めるが、可能な限り英文表記とする。

4.1 研究論文、事例／調査報告、解説／展望、論談、討論、研究速報などの原稿は英文でもよい。

4.2 英文原稿は語学的に難点の少ないものであることを必要とし、著者の責任において完全を期する。

4.3 英文原稿には、英文による要旨 200 語程度、ならびに日本語による 400 字以内の要旨をつける。ただし、著者が日本語を理解できない場合は日本語要旨を省略できる。

5. その他

原稿は和文または英文によるものとする。文章は語学的に難点の少ないものであることとし、著者の責任において完全を期する。編集委員会は語学的校正を行わない。

6. 要領の改訂

6.1 本要領の改訂は、編集委員会の承認を得なければならない。

7. 施行

7.1 本規定は 2002 年 8 月 27 日より施行する。

8. 改訂履歴

2003 年 5 月 2 日一部改訂。英語要旨の長さを 500 語から 200 語に変更。図、表のキャプション位置を訂正。

様式1 / 様式2

## 「情報知識学会誌」投稿原稿整理カード / 掲載原稿整理カード

1. 論文種別 (投稿規定の2.2からお選びください。)
2. 標題 (和文)  
   標題 (英文)
3. 著者名 (和文, ローマ字)
4. 所属機関名 (和文, 英文)
5. キーワード (和文, 英文)
6. 連絡責任者 (1名)  
   氏名, 所属機関 / 部局, 同住所, 電話番号, Fax 番号, E-mail アドレス
7. 送付投稿原稿  
   テキスト部分の枚数  
   図の枚数  
   表の枚数  
   付録の枚数
8. 図の返却希望 (YES, NO)
9. カラー図の有無 (原則として認めていないが, カラーでなければならない場合など, 全額実費著者負担)
10. 投稿日  
\_\_\_\_\_
11. 登録番号
12. 受付日 (再受付日)
13. 受理日
14. 送付掲載原稿  
   フロッピーディスクなど 枚数  
   プレインテキストプリント 枚数 (刷り上がり見本, 図表なども貼り込んだもの)
15. 別刷り (抜刷り) の希望部数 (全額実費負担)

### 【投稿原稿整理カード】

1. 標題を「様式1 投稿原稿整理カード」とし, 1から10項目 (11項目以降は採択後) をもれなく A4 判横書き 2枚程度に, ワードプロでお作りください。
2. 投稿原稿と一緒に送り下さい。
3. なお, 投稿時には「投稿規定」にある E-mail による連絡票もお忘れなく, お送り下さい。

### 【掲載原稿整理カード】

1. 標題を「様式2 掲載原稿整理カード」とし, 全項目をもれなく A4 判横書き 2枚程度に, ワードプロでお作り下さい。掲載原稿と一緒に送り下さい。

---

---

情報知識学会誌 編集委員会

編集委員長 安永 尚志 国文学研究資料館  
副編集委員長 宇陀 則彦 筑波大学図書館情報学系

編集委員

|      |              |       |             |
|------|--------------|-------|-------------|
| 芦野俊宏 | 東洋大学         | 石塚英弘  | 筑波大学図書館情報学系 |
| 伊藤鉄也 | 国文学研究資料館     | 岩田 覚  | 東京大学        |
| 内田 努 | 北海道大学        | 岡本由起子 | 東京家政学院大人文学部 |
| 神立孝一 | 創価大学経済学部     | 国沢 隆  | 東京理科大学理工学部  |
| 阪口哲男 | 筑波大学図書館情報学系  | 菅原秀明  | 国立遺伝学研究所    |
| 中川 優 | 和歌山大学システム工学部 | 長田孝治  | (株) システムソフト |
| 中山伸一 | 筑波大学         | 二階堂善弘 | 茨城大学人文学部    |
| 西脇二一 | 奈良大学社会学部     | 根岸正光  | 国立情報学研究所    |
| 原田隆史 | 慶應義塾大学文学部    | 藤井賢一  | 産業技術総合研究所   |
| 藤原 譲 | 工業所有権総合情報館   | 細野公男  | 慶應義塾大学文学部   |
| 山本 昭 | 愛知大学文学部      | 山本毅雄  | 国立情報学研究所    |

---

---

■複写をされる方に

本誌に掲載された著作物を複写したい方は、(社)日本複写権センターと包括複写許諾契約を締結されている企業の従業員以外は、著作権者から複写権等の行使の委託を受けている次の団体から許諾を受けて下さい。著作物の転載、翻訳のような複写以外の許諾は、直接本会へご連絡ください。

〒107-0052 東京都港区赤坂 9-6-41 乃木坂ビル 学術著作権協会

TEL: 03-3475-5618 FAX: 03-3475-5619 E-mail: naka-atsu@muji.biglobe.ne.jp

アメリカ合衆国における複写については、次に連絡してください。

Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA. 01923, USA

TEL: 978-750-8400 FAX: 978-750-4744 URL: <http://www.copyright.com/>

情報知識学会誌 Vol.16, No.1 2006年1月19日発行 編集・発行 情報知識学会  
頒布価格 3000円

---

---

情報知識学会 (JSIK: Japan Society of Information and Knowledge)

会長 細野 公男

事務局

〒110-8560 東京都台東区台東 1-5-1 凸版印刷(株)内

TEL: 03(3835)5692 FAX: 03(3837)0368 E-mail: [jsik@nifty.com](mailto:jsik@nifty.com)

URL: <http://www.jsik.jp/>

---

---

# Journal of Japan Society of Information and Knowledge

## Contents

### Research Paper

- Evaluation of an Information Retrieval Method using Non-Relevancy Information  
..... Atsushi MATSUMURA and Norihiko UDA... 1
- Human Recognition with Ear Image by Principal Component Analysis  
..... Yu WANG, Seiji ITAI, Satoshi ONO and Shigeru NAKAYAMA... 15
- A Database System for High-Throughput Transposon Display Analyses of Rice  
..... Etsuko INOUE, Takuya YOSHIHIRO,  
Hideya KAWAJI, Akira HORIBATA and Masaru NAKAGAWA... 28

### Case Study

- End User Education at Ube Industries, Ltd. — Trial Education of Intellectual  
Property Information for Business Section —  
..... Masanobu DEGUCHI and Kazuhiko OKAMOTO... 39
- Support of Intellectual Property Information Education for Yamaguchi TLO — Trial  
for academia industry cooperation —  
..... Kazuhiko OKAMOTO and Masanobu DEGUCHI... 44
- The Application of XML Technology in Genetic Resource Database  
..... Takehiro YAMAKAWA and Yukiko YAMAZAKI... 52

### Proposal Paper

- The Dawn of Informatics in Japan ..... Shizuo FUJIWARA... 60

### Information

- Call for Papers on Conference in 2006 ..... Naoki TAKUBO... 68
- Call for Papers on Special Issue ..... Yoshihiro NIKAIDO and Yukiko OKAMOTO... 70
- Books ..... Shu HIRATA... 71
- Recommendation of Director Candidate ..... 72
- Information for Authors ..... 74

情報知識学会誌 第16巻1号 2006年 1月19日発行

編集兼発行人 情報知識学会 〒110-8560 東京都台東区台東1-5-1 凸版印刷(株)内

TEL: 03(3835)5692 FAX: 03(3837)0368 (振替: 00150-8-706543)

学術刊行物 ISSN 0917-1436