

## 情報の”三つ組”モデルに基づく情報蓄積・検索ツール:情報箱

松永義文

富士ゼロックス(株) システム・コミュニケーション研究所

情報の”三つ組”(情報本体,名前,諸属性)モデルを採用した情報蓄積・検索ツール”情報箱”を実現した。三つ組モデルは「すべての情報には名前とその他の諸属性がある」というシンプルな原理に基づいている。三つ組モデルをユーザインタフェースの設計にも反映した結果、情報箱は、個人が、テキスト情報を、とにかく「どんどん蓄積しさと取り出すこと」を可能とするツールとなった。各三つ組は相互に独立であるが、個人毎に独自の意味や関連を与えることができるために、情報箱に蓄積された情報からは、ユーザの様々な固有性を引き出せる可能性がある。

### ***Joohoobako* and its information model**

Yoshifumi MATSUNAGA

Systems & Communications Lab., FUJI XEROX Co., Ltd.

430 Sakai, Nakai-machi, Ashigarakami-gun, Kanagawa 259-01, Japan

We developed a software tool, named *Joohoobako*, that enabled us to accumulate and retrieve electronic texts quickly. We adopted, in *Joohoobako*, the NEA data model; Name-Entity-Attributes. NEA data model is based on a very simple principle that every information has its name and attributes. All NEAs are mutually independent, but the users of *Joohoobako* can give their original meanings or relations to NEAs. We can expect, therefore, that a set of NEAs gives us some characteristics of its owner.

## 1.はじめに

個人にとっての情報の収集・整理・活用に関する方法は種々提案されている[1][2]が、我々は、整理や活用まで行かなくとも「どんどんため込んで、さっと取り出す」ことができるだけでも十分に役立つものになるという確信の下、個人用の情報蓄積・検索ツール“情報箱”を実現した[3]。

情報箱では、特に個人が、テキスト情報を、その内容によらず一元的に蓄積・検索することができる。課題は、次の個人向けの六点であった。

- (1) 容易さ: 入力(蓄積)も検索も容易
- (2) 頑強さ: どんな情報でも扱える
- (3) 多数性: 細かいデータを何件でも扱える
- (4) 迅速さ: レスポンスが速い
- (5) 軽快さ: ランタイムリソースが少なく済む
- (6) 永続性: 流通力をもち環境の変化に強い

これらの課題を解決するために、情報箱では、概念レベルのデータモデルとして、[4]でも紹介している情報の“三つ組”(情報本体, 名前, 諸属性)モデルを採用した。情報本体とは蓄積の対象であるテキストそれ自体である(以降、特に情報箱に関連した説明の中ではテキスト本体と書く)。名前とは情報本体につけられた文字どおり名前である。諸属性とは情報本体の様々な属性であり、情報箱では、検索する際に使用するキーワードとなる(以降、特に情報箱に関連した説明の中ではキーワードと書く)。情報箱のユーザは、蓄積しておきたいテキストごとにこの三つ組を作っては情報箱に格納し、名前かいずれかのキーワードをキーとして利用して検索する。

三つ組モデルの採用により、情報箱は、(1)から(6)の全ての課題をほぼ解決している。これは、1991年からの5年間の実使用に基づくユーザ評価による。

従来、電子テキストの蓄積・検索には、データベースシステム、ファイリング/全文検索システム、ハイパーテキストシステムが使われている。これらを、特に個人が(1)~(6)の要求を持って使用する場合に、我々は以下のような評価を与えている。

データベースシステムは、スキーマを定義することにより緻密な検索が可能であるが、(1)容易さ(5)軽快さ(6)永続性に課題が残る。ファイリング/全文検索システムは、蓄積されている大量のテキスト情報の全域を、記号列ないしは言語表現として網羅的に検索することが可能であるが、少なく

とも(3)多数性に課題が残る。ハイパーテキストシステムは、テキスト間の関係も含めて管理するが、(3)多数性(4)迅速性(5)軽快さに課題が残る。

情報箱は、上記のいずれのシステムにもおさまりに切らない、個人向けの新しいツールである。

本稿は、情報箱に以上の特徴を与えた、情報の三つ組モデルの概要を紹介することを主目的とする。第2章でまず三つ組モデルの思想、さらにそれが蓄積や検索のメディアとして“テキスト”を誘引する理由を紹介する。第3章で、情報箱ユーザインタフェースの概要紹介を通じて、実際にユーザにとってどのように三つ組がとらえられるかを示す。さらに第4章では、情報箱を実際に使った結果を踏まえて行った三つ組についての考察を記す。最後に第5章として、今後の課題についてまとめる。

## 2.情報の三つ組モデルとテキスト

### 2.1.情報の三つ組モデル

情報の生成とそれを取り出すための管理とを同時に果たすための、情報の最少構成モデル、それが情報箱で採用した、情報の三つ組モデルである。三つ組モデルのよって立つ原理は、次のように表現することができる。

#### すべての情報には名前とその他の諸属性がある

三つ組モデルは、情報の存在論的見方に立脚している。ある情報が取り出されるためには、その情報に、発見されるための手掛かりが与えられている必要がある。これがその情報の“諸属性”である。また、多数の情報の中にあつて、その情報を区別的に扱うには、手掛かりの中でも最もその情報に固有な識別子が与えられているべきである。これがその情報の“名前”である。

以上のように考えるなら、三つ組モデルは、情報の、存在および被操作(検索や選択などの操作に答える)のための最も基本的で自然な構造の一つとしてとらえられるのではない。

三つ組モデルに形態的にもっとも近い構造を提起しているものに、オグデンとリチャーズの三角形がある[5]。彼らは、事物とその象徴は思想あるいは指示を媒介して間接的に結合していると言った。この考え方は、事物と言語(象徴や思想あるいは指示)との関係をとらえているところが重要である。例えば、犬という実体に、“犬”という象徴が与えられているのは、二つの間に存在してい

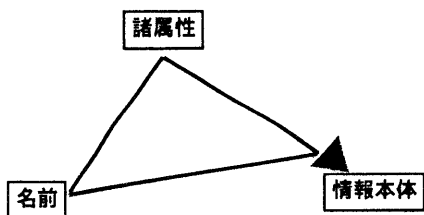


図1 オグデンとリチャーズの三角形  
に対応させた三つ組モデル

る思想や指示を媒介にしているということである。

情報もまた事物であると見立てれば、図1に示した関係が成り立つ。図1では、諸属性が思想あるいは指示に、名前が象徴に該当する。三つ組モデルが、個性の強い情報を対象にした、その存在と被操作のためのモデルであるのに対して、オグデンとリチャーズの三角形は、言語一般を事物との関係で説明するものである。したがって、オグデンとリチャーズの三角形では、名前と情報本体とは、直接の有縁性を持たないとして破線で結ばれることになるが、逆に情報箱の場合には、全く個人的に付与され得るため、強い関連が認められる場合が多いと予測される(そこで図1ではこの間を実線で結んでいる)。このことは、実際の使用結果を踏まえて第4章で確認する。

しかしながら、本節でオグデンとリチャーズの三角形を援用した理由は、事物であれ情報であれ、それが言語との関連で存在し操作されるという必然性を裏付けているからである。

## 2.2. 情報の表現/検索メディアとしてのテキスト

情報箱では、情報本体の表現メディアの種類をテキストに限定した。この限定は、様々な情報が言語(テキスト)で記述されるという考え方からは、図2に示すように、単なる限定ではなく、一つの集約となっている。この関係が、多種多様な情報を一元的に蓄積・検索するための重要な構図を表している。以下、本説では、三つ組モデルを情報箱で採用してゆく際に重要となるテキストについて、その位置づけを説明する。

著者は、言語について次のような立場をとっている。

「情報は言語で表現される。画像も音も、それは一体何を表現したものであるかを、言語なし

に説明することはできない。音声言語はもとより、身振りも無声アニメーションも、およそ考えられる手段はすべて、世界が言語によって分節されていることを前提にしている。」

たとえば、ある棚の上に置かれたビデオテープに、「情報箱プロモーションビデオ ver.1(VHS)。'91.4.17製作。製作者は萩原/宮内/中山/松永。時間4分05秒。製作目的は社外へトセトラ研究コンセプトのアピール(特にリクルートの学生向け)。保管場所は、本社リクルート担当者に聞けば分かる。」という言語による説明を与え得ることは、容易に理解されよう。この説明文という新しい実体を、ももとの実体である「ある棚の上に置かれたビデオテープ」の代理体とし、文字言語表現、つまりテキストとして表現する。このように考えるならば、全ての事物をもまた、それを言語によって記述することによってテキストの世界に写像できる(図2)。

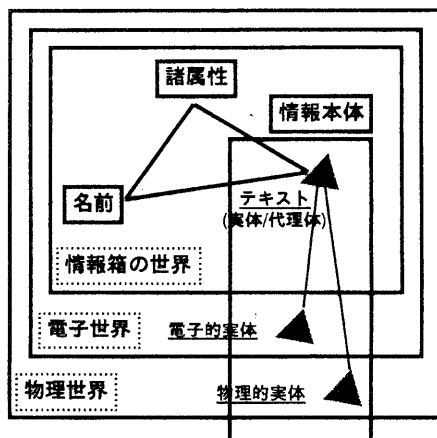


図2 情報箱は電子テキストの世界を扱う

今聞いている音声、あるいは頭にあるものを、直接テキストとして入力した場合、そのテキストそのものがオリジナルな情報本体となる。書籍のある部分を抜き書き的に入力した場合も、そのテキストは代理ではない情報本体である。

情報本体とその代理体の境界であるが、ここでは、情報本体に到達するための媒介の機能を果たし、情報本体の存在とともに意味をなすものを代理体と呼ぶことにする。本来代理体だったものが、それだけで独立の情報として機能するようになれば、それは情報本体と見なすことにする。索

引などが二次情報として独立の著作物として認められるのに似ている。

情報本体、代理体の手続きにより、結果として、全ての情報が、なんらかの言語記述として、すなわちテキストとして表現される。このように、テキスト世界に限定することによって、以下の効用もまた期待され得る。

- (1) 情報箱が管理する情報のサイズも小さくなり、電子世界における流通力が高まる。
- (2) キーワード検索技術や全文検索技術等も含めて、言語処理の技術を適用することが可能になるので、ユーザインタフェースを万人が自然に使用可能な自然言語をベースに設計でき、使いやすかつ普遍性のある道具作りが可能となる。

### 3.情報箱のユーザインタフェース

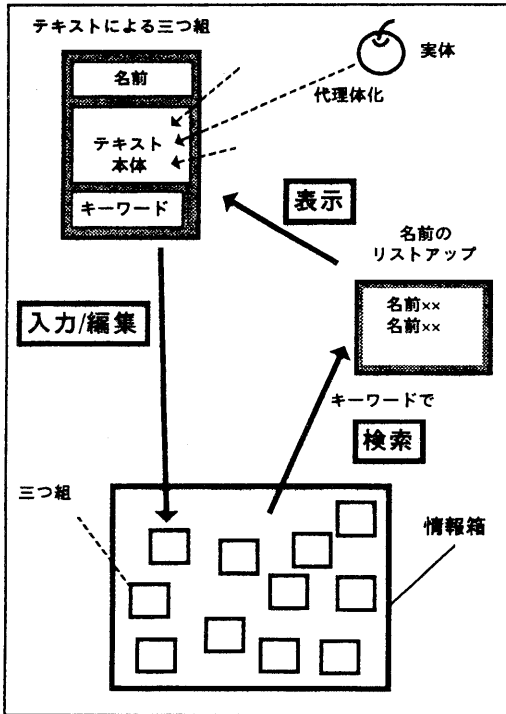


図3 三つ組群を保持する情報箱の入出力

情報箱では、既に述べたように、三つ組の要素を、テキスト本体、名前、キーワードとして、全てテキストとして表現する。情報箱に、三つ組を入力し、それをキーワード(または名前)で検索して名前のリストを得、その名前の中から一つを指定

して、テキスト本体または三つ組の内容を表示する(図3)。

本章では、実際の情報箱において、以上の手続きが、ユーザインタフェースとしてどのように実現されているかを、具体例を用いながら紹介する。

### 3.1.入力

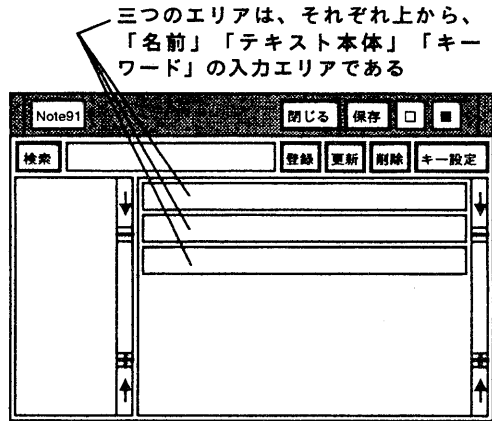


図4 情報箱の入力インタフェース初期画面

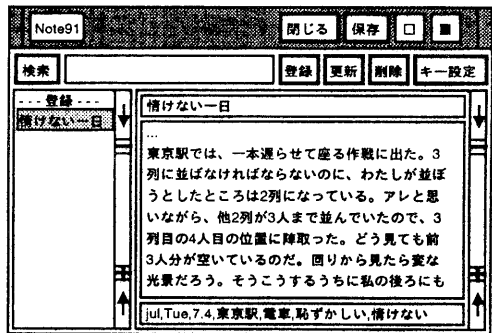


図5 情報箱への情報入力例

図4は入力の初期画面を示している。データ表示エリアに、三つ組に対応して三つのテキスト入力用エリアがある。上から順に名前、テキスト本体、キーワードを入力するためのエリアである。

この三つのエリアに、ある年の7月4日の日記を入力、登録、選択表示したところを図5に示した。この場合、『情けない一日』というのは、日記の

本体に与えられた名前であり、『jul,Tue,7.4,東京駅,電車,恥ずかしい,情けない』は、カンマで区切られた7つのキーワードである。

このように、テキスト本体の内容はまったく自由であり、名前の長さや、キーワードの数についても、十分な大きさを許容する。

### 3.2.検索

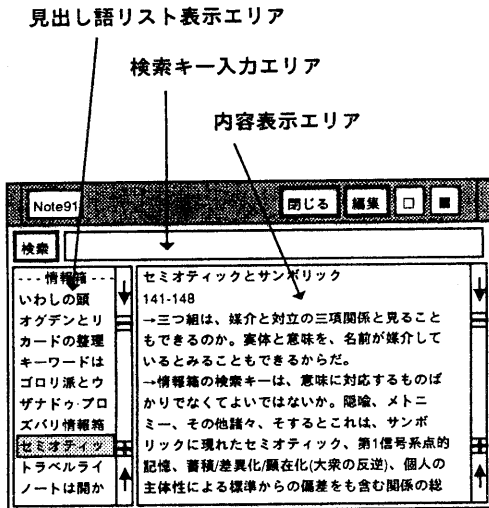


図6 情報箱の検索例

蓄積された三つ組は、キーワードかあるいは名前を指定するだけで、容易に検索可能である。

図6に、Note91という名前の情報箱を開いて、検索キー入力エリアから'情報箱'というキーワードを入力して検索ボタンを押し、見出し語リスト表示エリアにリストアップされた検索結果の名前から、『セミオティックとサンボリック』という名前を選択して、そのテキスト本体の内容が内容表示エリアに表示されている様子を示した。

このように、検索時のユーザインタフェースを、三つ組の各要素に対応して三つのエリアから構成している。

### 4.情報の三つ組モデルの考察

1991年から約5年間、実際に情報箱を個人の情報蓄積・検索ツールとして試用した結果得られた考察を記す。この考察の目的は、次の二つである。

- (1) 三つ組モデルの実地検証

- (2) 個人の蓄積結果や蓄積プロセスを分析して、三つ組から抽出可能なその個人の固有性のポイントを探ること

まず4.1.で、三つ組の各二つの関係を考察し、4.2.で、ユーザとの関係において三つ組内部に張り巡らされている暗黙のリンクについて述べる。

#### 4.1.三つ組の各二項の関係について

個人的に使う範囲では、三つ組の各項目は、極めて関係が深いものになる。オグデンとリチャーズの三角形との対応では、少なくとも名前は情報の本体とは直接の有縁性はなかったわけであるが、そのケースに該当するのはわずかであった(4.1.1.)。以下、各二項の関係について、経験に基づく考察を記す。

##### 4.1.1.名前とテキスト本体

情報に名前を付けるわけであるから、テキスト本体と名前の関係は、通常は、「名付け」の問題として考えられる。

名前の付けられ方は、実際に蓄積した三つ組を分析した結果、次の六つに分類できた。このうち、「代表」以外は全て、テキスト本体との関係において付けられるものである。

**要約:** テキストをひとことで言うところなる

**核心:** テキストのポイントはこちら

**感想:** テキストに対する感想

**識別:** とにかく他との区別ができればいい

**代表:** 色々の検索キーの代表(テキスト本体との関係ではない)

**実体:** 名前=テキスト本体である場合

「要約」と「核心」の違いは、「核心」のほうが、より主観的である点にある。「核心」の場合、気の利いた名前、しゃれた名前、暗示的な名前、そういう意識とともに、テキストの文面そのものからは離れて、おおよそ比喩的に核心を突こうとする。その分、主観性の度合いが強くなり、他人には、どうしてこの名前がこのテキストに対して付けられているのか理解できないこともある。

「感想」は、たとえば「そういえばそうですね」というように、さらにテキストとの個人的関係が強くなったものである。

「識別」とは、テキストの最初の5文字とか、最初の文節とか、最初の文とか、が名前として便宜上

与えられる場合、あるいは、A003とかの厳密な識別子が名前になる場合を指している。オグデンとリチャーズの三角形でいうところの象徴(シンボル)に相当すると考えられる。確かに、この場合は、テキスト本体と名前との間に直接的な有縁性はなさそうである。

「代表」とは、これはテキストとの関係というよりは、ひとかたまりの中から、より代表性あるいは識別性の強い要素をまさしく代表としてとりあげる場合である。たとえば、会社において、使われているパソコンを設備として管理しようとするとき、機種名、仕様タイプ、メーカー名、購入価格、購入年月日、主使用者(チーム)名、管理部課名、保守履歴、などなど様々の要素が一体となってテキスト本体を構成する。このとき何を「名前」として一つ取り上げるかは、通常は管理の都合から代表としてピックアップされると考えられる。

「実体」とは、テキスト本体が一つの語や語句であり、テキスト本体をそのまま名前としても登録してしまう場合である。たとえば'万般'という語句だけというようなときである。この場合'万般'はテキスト本体の全体でありかつ名前そのものである。なお、'万般'については、ことばの意味は理解しているが、ことば遣いそのものに関心があり、出典情報や'用語'などのキーワードとともに登録していたケースである。

さて、これら六つのうち、実際の検索の時に、検索キーとして利用されやすいケースというのは、「核心」、「代表」、「実体」のいずれかの場合である。「要約」の場合は、一語である場合より、句や文になる場合が多いので、検索キーとしてユーザが丸ごとを覚えている可能性は非常に小さいと考えられる。また「識別」の場合も、同様の事情があるか、もしくは識別番号的で非日常的である場合が多いだろう。また「感想」は、その文句をきちんと覚えている可能性は小さい。したがって、通常はこれらのキーでは検索がしにくい。

#### 4.1.2. テキスト本体とキーワード

キーワード自体、テキスト本体の内部から拾ってくる場合も多く、明らかに、キーワードは、テキスト本体を指し示すものとして期待される。しかしながら、テキスト本体とキーワードの関係については、次の二つの点に配慮が払われるべきである。

#### (1) テキスト本体に文字列あるいは単語として含まれないキーワード

全てのキーワードが、文字列あるいは単語としてテキスト本体に含まれているならば、それは、キーワードをわざわざ与えなくても、全文検索機能を活用すれば、漏れなく検索できることになる。実際にデータを分析したところでは、テキスト本体に文字列あるいは単語として含まれないキーワードの率は、平均的に30%~40%程度あるという結果が出ている。テキスト本体の出典など文脈的なキーワードがこれにあたる。このあたりが、キーワードを付与する意義につながっていると考えられる。

#### (2) 短いキーワードへの分解

キーワードにどの程度の複雑さを許すかは、ユーザが自ら検索する際にその検索キーを想起できる可能性がどれくらいあると感じるかによって決まる。たとえば、そのテキスト本体を指し示すのに'比較生産優位説'というキーワードが妥当であるとしても、実際の検索時のことを考えて、'比較生産'、'生産'、'優位'などの要素をもまたキーワードとして付与することをよくする。これは、キーワードとテキスト本体の関係よりも、キーワードと検索主体としてのユーザとの関係がより強くなるときがあることを物語っていると考えられる。

#### 4.1.3. キーワードと名前

「すべての情報には名前とその他の諸属性がある」という原理において、名前もまた諸属性の一つであるという考え方ができる。アリストテレスは、属性を自体的と付帯的に分けて、自体的な属性をそのものの本質であるとした[6]。キーワードが付帯的属性であり、名前が自体的属性であると考えれば対応がつく。しかしながら、名付けの分類の一つに取り上げた「代表」(4.1.1.)の例を見るならば、そのものの本質と言っても、利用目的というユーザ側の文脈に依存してしまう場合も考えられる。

逆に、検索の都合で、名前の部分文字列や言い換えをキーワードにするという関係が存在することもある。たとえば名前を文で表した場合等は、そのまるごとの文で検索することは困難なので、文のより小さな構成素のうちのいくつかをキーワードとして設定する場合である。また、一つのことを言うのに、'思考'でもいいが'考える'でもいいなとか、'思考性'や'思考的'、'思考的な'でも検索

できるべきだとか、一旦決めた名前の、言い換え表現をあるだけキーワードにしてしまう場合もある。

#### 4.2.三つ組間の関係について

情報箱においては、各三つ組は相互に全く独立である。しかしながら、ユーザが検索キーを用いて検索をする場合に、同じキーワードを持つ三つ組は、暗黙のリンクで結ばれているとみることができる。暗黙のリンクについて考究した結果、次の三種類のリンクが存在していることが分かった。図7はリンクの様子を示している。

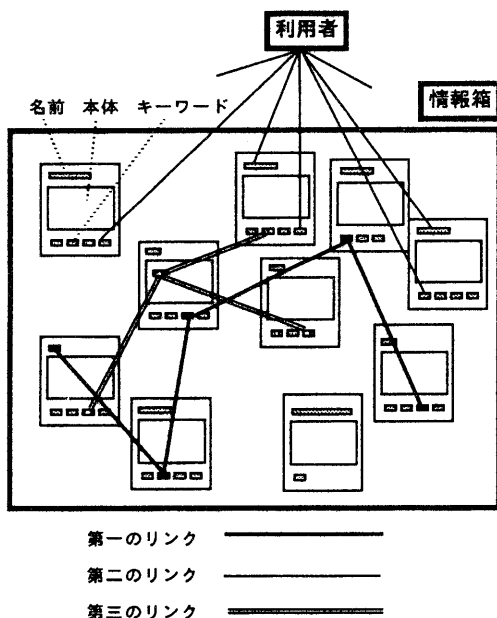


図7 三種類の暗黙のリンクのイメージ

##### (1) キーワード間リンク

第一の暗黙のリンクは、キーワードとキーワードの間に張られて三つ組間をつなぐリンクである。どの三つ組とどの三つ組が同じキーワードを持っているかを正確に把握しないと、検索要求に答えられないので、情報箱としても、このリンクに関しては明確なメカニズムを実現している。三つ組間で同じキーワード間に張られるリンクの数は、三つ組の総数と、キーワードの重複率に比例して大きくなる。キーワードの重複率とは、全三つ組にわたって数え上げた延べキーワード数を、異なりキーワード数で割った

ものであり、この値が大きいと、三つ組は、相互により密な関係を持っていることになる。

##### (2) ユーザリンク

第二の暗黙のリンクは、ユーザと三つ組をつなぐリンクである。これはユーザが、検索行為をするときに動的にイメージされるものである。キーワード間リンクが、ユーザの検索行為に関係なく決定されるのに対し、ユーザリンクは、ユーザが様々な文脈で、どのような頻度でどの検索キーを与えるかを問題にしている。これは、ユーザの検索行為を分析しなければその様子は分からないし、またユーザ毎に大きな特徴が現れる点であろうと思われる。

##### (3) テキスト/キーワード間リンク

第三の暗黙のリンクは、テキスト本体に埋め込まれていることばとキーワードとを三つ組間でつなぐリンクである。情報箱の検索では、たまたま検索して表示しているそのテキスト本体の部分文字列を選択して、それを次の検索のキーワードにするということがよく起こる。このことを配慮したキーワード付けがなされていると、連続的な検索がスムーズに行われると考えられる。テキスト本体の(それに付けられた)キーワード含有率(4.1.2(2))を大きくすることに相当すると考えられるが、ここでは、各キーワードは、それらを付与されているテキスト本体だけでなく、他の三つ組のテキスト本体の部分文字列を構成していても構わない。

以上の暗黙のリンクは、システムが直接規定しているものではなく、本来全く独立に存在している三つ組が、言語というメカニズムと、ユーザという利用主体とによって、自動的に張り巡らされていくものである。その張られ方にユーザの固有性が現れるものと考えられる。

また、このリンクの状況を把握し制御できれば、たとえば複数の「三つ組集合」間の演算が可能となり、活用の幅を広げることになると考える。

#### 5.今後の課題

- (1) 三つ組モデルを、既存のデータモデルとの比較で評価したい。
- (2) 三つ組の考察を、より定量的に行いたい。たとえば、名前とテキスト本体の関係の6分類においては、量的にはどの分類が最も多いかなどである。

- (3) 三つ組の情報本体は、情報箱の場合、テキストに限定しているが、当然、他のメディアで実体が形成されている場合も、本稿で提案した代理体の考え方を使って、実際の守備範囲に入れていきたい。なお、一般に様々な電子的実体を保持可能にすることは、マルチメディア情報の保持の問題を避けて通れないと考えられるので、その場合には、マルチメディア情報空間についての先行研究([7]など)との関係も明らかにしていきたい。
- (4) 情報箱は、個人使用目的だけでなく組織の情報の共有化の場面でも活用が実践されている[8]。今後とも、有効な活用法を探り、そこからさらにモデル論議にフィードバックをかけていきたい。

## 6.おわりに

とにかく「どんだんため込んでさっと取り出す」ことを主目的とした、情報蓄積・検索ツール情報箱を実現した。情報箱で採用した情報の三つ組(情報本体, 名前, 諸属性)モデルは、情報の存在および被操作のためのもっとも基本的な構造であった。

情報箱は、既に多数のユーザによって活用されており、これまでの蓄積・検索ツールの枠に収まらないとの感想が寄せられている。もとより情報箱は実用を目的としたものであるが、そこで蓄積される三つ組は、そこから、ユーザの固有性を抽出するための格好の素材になる可能性を秘めている。

今後、5章で示した課題への取組を通じて、情報箱および「三つ組」モデルについて、さらに考察を深めてゆきたい。

## 参考文献

- [1] 梅棹忠夫: "知的生産の技術", 1969, 岩波書店  
[2] 外山滋比古: "思考の整理学", 1986, 筑摩書房  
[3] "VP情報箱操作説明書", 富士ゼロックス(株), 1992  
[4] "富士ゼロックス、新モデルで知識ベース開発", 日経インテリジェントシステム, 1994.5.16, 日経BP社, p13  
[5] オグデン/リチャーズ共著(石橋幸太郎訳): "意味の意味", 1967, 新泉社  
[6] アリストテレス(出隆訳): "形而上学(上)", 1959, 岩波書店

- [7] 小澤英昭他: "情報空間のモデルに基づく情報検索システム", 情報処理学会情報メディア研究会 7-7, 1992.7.10  
[8] 田丸恵理子他: "オフィスにおけるコンピュータ利用者の適応活動の分析とモデル化に関する事例研究", 計測自動制御学会第9回ヒューマンインタフェースシンポジウム, 1993.10.18-20, 神戸