巻頭言

リストラクチャ（再構築）

藤原 譲†

バブルがはじけて景気の低迷が長引いているが、計算機や通信システムのダウンサイジングの物理的要因もあって、組織や設備の再構築が“リストラ”の掛声がよく聞かれるようになって来た。またリクルート、共和、佐川、ゼネコンとたて続けに話題を賑わし、結局長期間にわたり政権を維持してきた自民党が野に下り、替わって革新系大合同の細川現体制が出来上がったので、これは政界のリストラである。

悪いことばかりがきっかけになるわけではなく、Jリーグの誕生で、サッカーは大きなブームを惹き起こし、10月28日対イラク戦で引き分けとなり勝てなかったので、世界選手権出場のチャンスは逃がしたものの、宿敵韓国に勝ったことも含め、見違えるような成長を示した。外人選手の導入も含め、リストラの効果といえる。

ところで、最近のJCICSに出たMarsの論文で、化学情報を共用し、再利用するために情報の蓄積、検索、操作の問題点を解決する方法として、概念の構造化が報告されている。これは純物質から位相、状態までもできるだけ基本的な概念で記述する方式をとっている。純物質そのものは元素と分子の部分構造によって表現する。

これは一般の概念を基本となる意味素の組み合わせで記述しようとする考えと類似のもので、基本となる概念の適切な集合をどのように設定するかで対象や利用法が限定されてしまうので、当初狙っていた標準とか、共用のためには不満足なものとなる。またこのような考えは数学的に扱い易く、システム構築は容易であるが、対象領域の僅かな変化や技術の進歩に対して不安定となるなどの欠点もある。したがって化学情報のモデルとしてはまだ大いに研究の余地がある。

いずれにしても、化合物の命名や化合物の特性の個別な記述だけでなく、化学情報全体の構造化を行うことによって化学情報の共用、活用を図ろうとすることは、これまで本格的に検討されていなかったので一石を投じた論文といえる。当研究会の反応分科会で行っている自己組織型の情報ベースは名称の通り、有機合成研究に必要な情報を自動的に構造化して、高度利用が可能となるシステムであり、その方向では我々の研究の方が一歩進んでいる。このような研究やシステム化が進み、化学情報のリストラとその成果を期する次第である。

†筑波大

講演論文

情報知識学のフロンティア

藤原 譲†

1. 序

情報知識学または情報学については種々の意見があるが、その目的は、人間の知的活動の基本である情報に関する理論や知識を体系化すること、及びそれの応用として知的生産性の向上を図ることであるという立場からの考えを述べる。先ず知的生産性を向上させる為に何故情報知識学かということになるが、知的活動の機能について考えることから始める。人間は頭の中で思考活動を行っているけれども、その活動は各種の観点から見ることができる。優れた思考活動ができる頭脳の表現として博識という言葉がある。これは物理的に言えば情報の量が多いということに相当する。また頭脳明晰であるとか頭の回転が速いということも重要な機能であり、これは応答の速度が早く、しかもその結果の精度が高いということになる。これらは一応定量化ができる機能であるが、これだけでは高度な知的活動に対応しているという感じがしない。発想、創造というレベルになると本当の意味で価値を生ずる機能といえる。このことは知的活動が人間の活動の中で一番重要なものであるといわれる由縁である。それにはどのような処理が対応しているかというと、情報を解析したり、帰納推論、類推、仮説推論、連想や評価をすること、さらにその延長上に発想、創造などがある。人間の場合でも計算機の場合でも情報量を増やすというのは比較的易しい。人間なら当然新しい知識をえるための勉強をするし、計算機であれば情報を入力すればよい。次のどう処理するかということについては、簡単ではないが方法が解っていればそれを指示すればできる。これらは人間の場合には学校での主たる教育内容であり、計算機としてもやり易いが、それを超える機能については事情が異なる。最近試験では良い成績を取る学生が研究になると予想されたような成果が挙げられなくて、試験の成績に比べてたいへん見劣りがすることがよくある。会社に入っても、模範社員ではあるが会社を背負って立つ人にはならないことがあるけれども、これらは暗記学習を超える機能の問題が関わっていて、計算機の場合にも人間の場合と同様難しいことになる。結局このような問題を考えると知的な活動の対象でもあり成果でもある情報と知識はとにかく入れることはできるし、また方式の決まっている処理をすることもできるが、それ以上の類推、発想のレベルの処理を行わせるのは容易ではない。人間なら「もう少し考えろ」、「もっと良い方法はないか」というように刺激をすればできることもあるが、機械には叱咤激励は通じない。しかし基本的には計算機に実行させるにしても人間を教育するにしても同じことになるが、情報とはそもそも如何なるものであるかを考える必要がある。

歴史的にみて「思考支援の方式」としてこれまでどのような手段を持っていたのかということを考えると、簡単なそろばんや計算の手助けになる道具は人間の歴史と同じくらい長い歴史があり、基本は加算であるけれども、四則演算ができる道具として各種の形のものがある。そのことを計算機ではスィッチングの機能で行わせる。もともと四則演算が中心であるけれども高速大量の処理ができることでスイッチ演算を通し論理演算もできるということである。最近だとコネクションマシン、ニューロ計算機といったようなもので分類や学習が有る程度できるようになった。(表1参照) もう少し高度な思考機能的処理をすることになると、意味の処理、内容を把握する必要があり、単純な数値や符号の処理だけでは済まなくなる。もしこういうことができれば機能として人工頭脳ができるようになる。

何れにしても情報知識を有効に活用するためには、情報や知識がどういうものであるかを知る必要

†筑波大学

表 1 思考支援方式の発展

方式	技術	思考機能
加算	そろばん 計算尺	四則計算
スイッチング	計算機	数値演算 論理演算
神経回路	コネクションマシン ニューロコンピュータ	分類 学習
意味処理	人工頭脳	問題解決 意志決定

がある。それが情報の学問であるから、情報知識学の目的は情報の特性と理論の体系化、情報に関する技術、手法の開発及びそれを具体的に各分野の情報、思考活動への応用することになる。

情報知識学で問題になるのは、そもそも情報、知識、データなどは何であるかということである。まず情報とは一番広く考えると「認知とか思考の対象となる実体についての認識内容」であり、普通の意味で言われる情報は全てこれに入る。知識とは一般的には情報と同じに使われることもあるし、情報処理、特に人工知能の分野では一定の形式化された知識を指すので、具体的にはプロダクションルールとか1階述語論理で表現されたもの、またはその延長上にあるものということになる。ここでは一番広義の知識と情報処理で使うものの中間になるが、知識とは「体系化された情報」という意味で使うことにする。次に情報はいろいろな形で記述され、表現されるがそれの「最小単位」をデータという。またその「集合」もデータという。以下はこれらの定義に従うことにする。

情報処理に関して、大きな特徴の一つは計算機は生まれたときからその速度が人間のそれに比べて速いことである。その速さは益々速くなっており、初期の装置でも処理の速さは1秒間に千回くらいの演算であったものが、今は百万回から10億回くらいまでに達している。この面では計算機は始めから素晴らしいものであったが、記憶装置の点でいうとだいぶ違っていて、主記憶の容量は初期がキロバイトから現在でメガバイト、補助記憶装置がキロバイト、メガバイトからギガバイト、やがてテラバイトであるが、こういう変化は情報の高度な処理に対しては非常に大きな影響がある。つまりメガバイト以下だと我々が必要とする情報の全部を収録できないので、計算機を使うのは有る特定な目的の為に特定の情報に対して処理をすることであった。ところが、専門家が持っている情報は大体ギガバイトあれば充分であることから、かなり高度な情報の処理を考えたときにも、現在では容量的には必要な全てを対象にした情報の蓄積ができる。ただ人間の脳の容量は10の14乗ないし15乗バイトといわれているので、計算機よりはるかに大きい。とにかく現在の計算機は容量も速度も専門家の要求に充分対応できる段階にあるので、ソフトウェアの機能も専門家が思考の中で使っていることの全てを対象として考える必要がでてくる。したがって通信やメディアの問題も従来とは大きく変わらざるを得なくなる。

2. 情報の基本課題

つまり思考の2大要素である記憶容量と処理速度はどちらも非常に進歩しているので、高度な思考支援ができてもよさそうだが、実際はそうなっていないので、いくつかの解決しなければならない具体的な課題について述べる。

2.1 情報資源化方式の課題

情報を何らかの目的のために利用しようというときに、まず関連情報をデータベースに入れるということが考えられる。これが情報の資源化の第一階段である。資源化は情報の管理、利用に直結するので、そのために単に入力するだけでは充分ではない。例えば特許の文献を検索することはオンラインシステムでできるが、特許の内容検索をしようとすると、総称表現の問題がでてくる。即ち特許では出願する発明について最大限の権利を請求したいので、どうしても個別的表現ではなく、できるだけ包括的総称的な表現を用いることになる。このような総称表現をうまく管理したり表現するのは現在の技術では非常に難しい。同じ様に材料の表現にも総称の問題があるが、そもそも材料には属性が極めて多いため良い記述方式が確立されていない。有機合成も総称表現をよく使う分野である。混合物、複合物は材料の記述と同じような問題がある。また研究開発の情報も一般的に言って複合的な情報が多く、不確かな情報に対して多値論理、例えば様相論理的な問題も出てくる。このように専門的な思考活動の対象となる情報は、現在の技術では適切な取扱いは困難である。それは複合情報は定形化ができないという問題であるとか、たくさん情報を集めると全部の属性の値は集められないので欠落値、即ち空値 (null value) の

問題が出てくる。つまり情報が無いということも意味があるので、意味の処理が必要となる。存在する情報の意味処理も難問だが無い情報の意味はもっと難しい。一方単にデータが多いだけでもその処理は困るわけで、全文検索が従来のキーワード検索方式では精度、速度の問題を生ずるのはこのためである。このように単にデータを入れるということでもいろいろな問題を抱えている。これらの問題の研究はもちろん世界中で精力的に研究されているが、今の所満足すべき結果が得られていない。

2.2 分類、アクセスの課題

次に情報の管理利用の面からも重要な分類は非常に古くから用いられている情報に対するアクセス手法でもあるが、図書であれ特許であれ、また生物、鉱物の分類であれ、問題のない分類というものがない。例えば図書館の場合だと多種類の情報が多量にある。そしてそれぞれの図書館の独自性もあるので標準的な分類ができない。また特許の場合は多量であるうえに変化が激しく、生物の場合には生物としての分類基準のほかに利用目的による分類などが入って来るのでまた適切な分類が困難になる。結局分類には一意性がないことから生ずる問題であって、一意表現としての分類方法は後で説明するように一般的には不可能である。端的に言えば分類の基準に一定の順序関係が決められないので一意分類ができないということである。それなら約束により標準化をすれば良いということになるが、その場合にも分類の基準の多様性と分類の基準が変化したり、新しい分類基準が必要となることから、標準も簡単には決められないし、一度決めても頻繁に変化することになる。

分類手法としては国際的なUDCや、日本のNDCなど十進分類法が確立されており、標準化も進んでいるが、設定しても、できた途端実状とのずれを生じて、それぞれの役割を反映して別の目的で追加、変更をせざるを得ない。分類の非一意性とともに、分類しようとする手法が分類される対象に適合しているのか、分類が可能であるのかという基本的な問題がある。結局UDCは維持をするだけでも大変であるし、それも大きな図書館では独自の分類を行っているので使われないのが実状で、日本もそうだがアメリカもイギリスも国立図書館ではUDCは使っていないことが問題の根深さを示している。

2.3 制約の課題

その次の課題として計算機を用いて情報を処理しようというときに多くの制約がある。それはデータベースでも知識ベースでも、ある一定の形式のもとで統一して管理をし、処理をしたいので、形式を整えるとそのために制約が出てくるということである。データベースの場合、管理やアクセスのために識別子としてキーを設定することが多い。それは簡単であると考えられているが、先ほどの分類と似ていて、可算的でない情報に対しては基本的に不可能な方式である。それを敢えて実行するとすれば、データに制約を加えて入れることになり、結局収録する情報が制限されることになる。蛇足であるがキーとして識別番号ではなく適切な名前や符号を与えることもよく使われる方法であるが、名前や符号を付けることと番号を付けることは基本的には同じ制約になる。

また知識ベースでは通常の表現方式として、if—、then—の形の生成規則 (Production Rule) がある。もちろん直接この形を処理することもできるが、計算機の中で効率よく使うために手続きによらない形としては一階述語論理の表現がある。これらは非常に固い表現なので適用できる情報が著しく制限される。また (1) 式、(2) 式、(3) 式のように書ける。

The Issues of Knowledge Representation

Production Rule :

$$\text{If } P \text{ then } Q \qquad (1)$$

Predicate Calculus :

$$P \rightarrow Q \qquad (2)$$

$$\neg P \vee Q \qquad (3)$$

$$\neg (P \wedge \neg Q) \qquad (3)$$

, where

$$X = P \rightarrow X \neq (\neg P) \qquad (4)$$

これらは1階述語論理の枠の中では等価の表現である。先ほどの同定と識別とがちょうど対偶の関係にあるけれども、それと同じことが (2) 式と (3) 式の関係に当たる。つまり「P ならば Q である」ということは、「P でないかまたは Q である」ということに等しいし、又そのことは「P であってかつ Q でないということはない」ということになるわけだが、これらが成立するのは先ほどの対偶が成立したのと同じであり、(3) 式の下に示したように2値論

理が前提である。ところが使われる情報は2値論理とは限らない。一般には多値論理つまり「そうである」か「そうでないか」のどちらかに割り切れる場合だけでなく、「そうかもしれない」し「そうでないかもしれない」というような場合も含めた論理である。そういう情報に対しては2値論理の手法は使えない、つまり演繹推論であるとか数値計算であるとか符号の照合というのは計算機むきの良い方法ではあるが、それが使えない情報も多いということである。

結局階層型、網型、関係型などの古典的なデータのモデルのみならず、実体-関係型、意味データ型や最近のオブジェクト指向型、ハイパメディア型、および知識表現用などのパラダイムもあるが、大量情報の適切な管理は未解決の課題である。しかもデータベースや知識ベースの持つ制約は思ったよりはるかに厳しく、全体として扱えるものより除外するものの方が著しく多いのが現状である。もう少し柔軟性のある容れ物が必要ということになる。

2.4 識別の課題

識別するということは“AというものはAである(A=A)”、“AはAでないものとは異なる(A≠∼A)”ということだが、こういうことが何時も成立すれば問題はないが、例えば同意語があると、その同意語は別の表現をしているので、意味が一緒であっても表現形式が違うので

$$A = A$$

が成立しない。それから多義性があると表現は同じだが内容が違うので、また

$$A \neq\sim A$$

が成立しないということである。同意語と多義性は言葉の特性の一つであるので、同定とか識別ということが普通の言葉を使う限り非常に難しいことになる。また言葉にはもう一つ問題すなはち概念の多重階層や部分的重なりがある。例えば、中村一郎という日本人がシンガポールに住んでいるとする。中村一郎、日本人、熱い国に住んでいる人、男性、等の概念がこの人に該当する。この時概念の重りはあるが、階層関係のみでは無い。また人間、日本人、情報知識学会の会員などの集合の概念は階層性を持っているので、これらは包含関係にあるということになる。後で説明するように概念の部分共有や階層関係の相対性が現在の技術では適切に扱えないことも基本課題の一つである。したがって利用者の思う情報が単なる符号の照合では出てこないということになる。

識別するためにはキーがあり、従ってキーに必要な性質は、1) ユニーク(一項的)であること、つまり一つの概念や一つの実体にたいして表現が一つであること、2) 一義的、曖昧性がない、つまり一つの表現にたいして二つ以上の実体概念が対応しない、3) 正準化できることの3つである。最後の要件はパフォーマンスのことを考えると重要である。また一項的でかつ一義的ということは1対1対応ということ、それから正準化できること、例えばソートできるということ、これらは全て簡単なことのようだが、実際には非常に厳しい要件である。

分類や識別に関連したことで大変誤解され易い情報の性質がある。それは情報の管理の拠り所として、もし識別子としてのキーがあれば好都合であることは上で述べた通りであるが、それは予想されるよりずっと困難なことである。すなわちそれを番号付けの問題と考えると、番号がきちんと付けられる集合としては有限集合がある。それよりすこし大きな概念で可算集合というのがあり、数えられるけれども有限とは限らない集合である。さらにそれより上位の集合として線形集合というのがあり、それの上位集合が半順序集合で、通常の文字その他で表現されている情報はこれに対応する。要は数字か名前でキーが付けられるのは有限集合か可算集合までである。ところが計算機の中に入っている情報というのはたかだか有限であるから、可算的だと考えることである。しかし情報がIDナンバーであるとかネーミングで管理ができるというのは、各要素がキーによって互いに識別または同定可能であるという前提によっている。この中に新たな情報が加えられた時にはそれが有限個であっても管理機能が変わり得るということである。つまり有限集合は常に不適切なものを削除したり、修正することを含めて維持を怠らなければ有る程度管理できる。それは先ほどの分類も同じことであり、管理とか分類とかは適切に維持すれば、それなりに使える筈であるが、それは実際には極めて困難である。

2.5 複合情報の課題

複合値や抽象型データ(ADT)の問題は、図1に示すようにデータは別の所にある情報を間接的に手

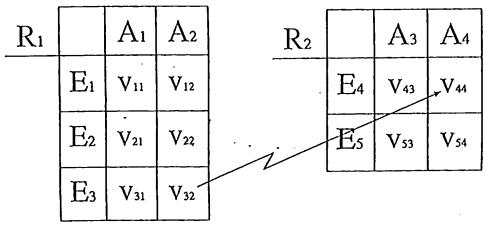


図 1 An Example of Abstract Data Type

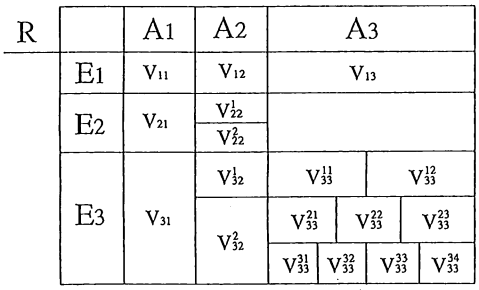


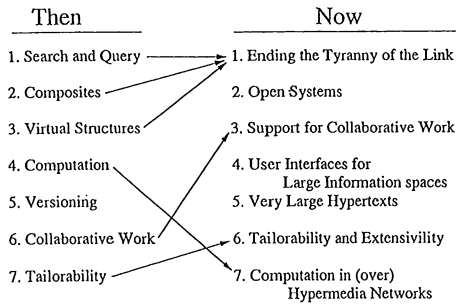
図 2 An Example of Composite Values

続きを経て指すので、実際にその場にデータがあるのではなくて抽象化された表現が有るという場合である。これも計算機のなかで使いたい表現であるが、そうすると識別の問題は一層複雑なものになる。つまり抽象データ型では直接的な表現は内容とは関係の無いものになっているからである。

次に複合値に関する課題を説明する。図 2 に具体的な高分子の例を示す。一つの属性であったものが分割されるとか、一つの実体であったものがいくつかに別れるとか、どこかの値が欠けているということがよくある。それぞれに対して各種の対処の仕方が考えられてはいるが、関係型データベースでいえば、そのような場合にたいして属性の集約 (Aggregation) とかタップルの汎化 (Generarization) とか各種の対策があり、両方合わせて抽象化 (Abstraction) というが、それを実行すると、システムの持っている管理機能が損なわれるということで利用が限定される。つまり柔軟性を増せば管理機能が失われ、管理機能を維持しようとすれば柔軟性が不足してデータ収録を制限せざるをえない。このように情報のいろいろな問題点というのは情報の基本的な性質に深くかかわっている。

ゼロックスが提供していたハイパーメディアシステム NoteCards の経験から、次世代のハイパーメディアに展開するために、解決すべき問題として

表 2 Seven Issues by Halasz: Renewed



Halasz が 7 つの課題 (Seven Issues) を ACM に 87 年に発表した。表 2 の左側に書いてあるのはそれである。その後これらの課題を踏まえて新しいシステム (Aquanet) を開発しているが、その過程でまた新しい開発問題がでてきたので、再整理したものが 91 年の暮れに報告され、表 2 の右側に示してある。始め入っていなかった問題の一つは大量のハイパーテキストを作るのが難しいことである。つまり小さなスケールの情報は扱えるが、本格的に使うとなると大規模な情報を入れなければいけない。大規模な情報を入力し、構造化し、使える段階に資源化し、適切に管理することが困難であるということである。この問題を解決しない限り大型ハイパーメディアは実用的なものにならない。このことに関連しては複合情報や構造の問題に関連があるといっており、柔軟性があり何でもできそうなハイパーメディアも構造化と管理ということが大きな制約になっていることが示されている。

2.6 利用機能の課題

今までの課題をもし解決したとして、最後に利用機能の問題がある。現在の計算機では四則演算や符号照合の処理、即ち数値解析、検索、演繹推論などは高速かつ高精度で処理される。より高度な予測や推定になると、完全ではないが種々の手法があり、実際に使われている。更に高度な機能になると、まだ研究の段階であり、ましてそういうものを複合して最終的な問題解決、意志決定、評価などをすることは更に難しい問題になる。これらが解決されない限り本当に思考支援という段階には至らない。こういう問題に対しては、先ほど述べたような制約を考えると、ハイパーメディアとかオブジェクト指向的な方法がかなり使えるのではないかと期待されている。

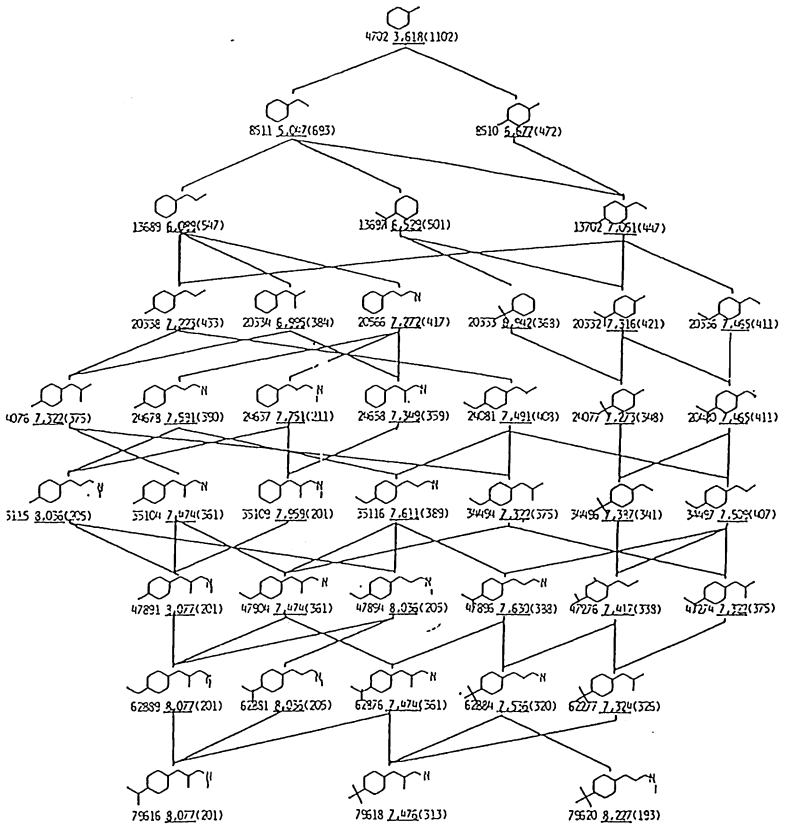


図 3 KOWIST: tree of substructures

以上述べてきた問題をまとめると、基本的には大量の情報を利用しようと思っても適切な形で提供されていない。データベースには沢山の情報が入り、知識ベースでも入れられることにはなっているけれども、知識ベースでは知識の表現の制約から知識の獲得が困難であり、データベースの方は管理、とくに識別、同定からの制約でいずれにしても入れられるものが限られる。つまり全体から見ると現在の技術で扱える情報に比べて積み残した情報の方が圧倒的に多い。それは管理システムの基礎となるモデルと実現方式の柔軟性と管理機能が不足していることに起因する。

2.7 アクセス手法の展開

少し個別な問題でアクセスの問題をもう少し考えてみると上で述べた分類を用いること、キーによって情報を識別することに基づいてアクセスをすること、及びキーワードの索引が今までの代表的なものである。しかしこれらは先ほど述べたようにいろいろな問題点を持っており、それが本質的に情報に付随する問題点であるから簡単に解決できることでは無い。それから新しい方法の全文データベース用シグネチャーファイル方式やマルチメディア用の変換コード、それから従来のネットワーク型データベース管理システムのように情報の構造を直接利用する方法も考えられる。

2.8 概念構造の例

実際の例でいうと、図 3 は化合物の部分構造に関する包含関係である。これはごく一部を取り出したものだが、各種の包含関係があり、構造表現に多様性があることを示している。これは一般の概念の場合も同じで、高分子材料のデータでも製造、加工することに関連したいろいろな原料、材料特性、装置、操作などの概念が多重の入れ子関係を含み複雑な関係になる。このような情報の記述、表現の多様性の説明のため、単純な場合で包含関係だけがあったとして次に図解する。4 つの特性で記述されるべ

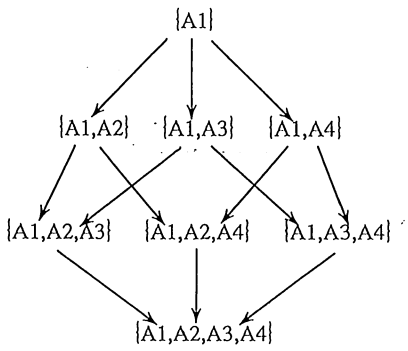


図 4 The Lattice Structure of Generic Hierarchy

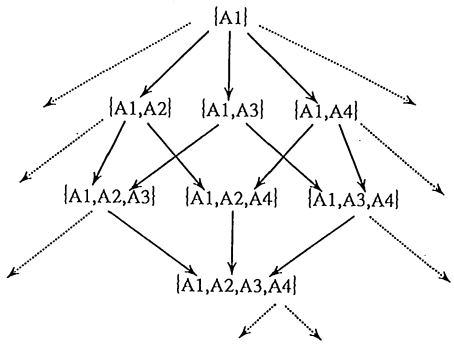


図 5 The Lattice Structure of Generic Hierarchy

き対象があったとして、その世界はこの 4 つの属性の全てを正確に記述するレベルとそれより少ない 3 つ、2 つ、または 1 つの属性で記述する、4 段階がある。実際にはさらにこれらの中間もあるが複雑になり過ぎるのでその議論はここでは省略する。先ず A1 という概念で記述し、その次に A2 で記述し、更に A3、A4 で記述する仕方がある。図 4 の上から下への別のルートがそれぞれ別の記述法に対応している。このように属性が 4 つあるだけでも記述の仕方は 24 通りになる。それから分類も同じ構造で説明すると、分類属性のレベルと組み合わせ方法に依存するので、この概念構造からは少なくとも 24 通りの分類より、詳細には 2 の 15 乗すなわち 32768 通りの分類の仕方があることが示される。記述、表現の多様性は有っても適当な基準で標準化することも考えられる。しかしそれが難しいのは利用目的から記述属性が変更されることも大きな理由であるが、技術、目的の変化により新しい属性を付加しなければならないことにもよる。例えば今 4 つの属性で記述していたところに新しい概念が入ってきて、5 つ目の属性でも記述しなければならなくなったというときに、例えば A4 の属性を 2 つに分けて 5 つの属性にすれば済むのかというと、じつはそれだけではなくて、図 5 に示すように新しい属性が加えられるということは全体の構造がもう一段深くなることであるし、さらに上部の各段階が広がることになる。一つの属性が増えるということは構造全体が変化することで、属性の数が大きくなればなるほどその影響が組み合わせ的に大きくなる。従って分類の方法も、表現の多様性も情報の記述、目的、内容に応じて大きく変化する。分類や表現の多様性は情報の表現の本質的な性質であるということは充分に留意すべき点である。

2.9 意味論的課題

これまで何回か意味の関わる問題ということを述べたけれども、もう少しふえんすると意味の関わる問題の一つは個別実体 (Distinct Entities) の集合を対象としていることである。つまり我々の考えている対象領域では、ある概念の表現と別の概念の表現との間に重なりがなく、別々のものであるというのがデータベースでも知識ベースでも基本的な考えである。実際は先ほどの化合物でも特許や法律でも、概念というのは非常に多くの重なりがあり、それを考慮しないで処理することは無理であり、例えば総称表現が旨く処理できない。また類似性というのも重なりがある概念の関係であるから旨く扱えない。それから空値問題も値が無いから意味処理が困難で厄介なものである。それから実体と実体の間にある関係は意味の表現に直結するものであるが、システムによってはそもそもこのような関係についての表現を持たないものがあり、その典型的なものは関係型データベースモデルで、PC やワークステーションから大型用のデータベース管理システムとして普及しているが、実体間および関係間の関係を扱う機能がない。モデルによっては実体一関係型 (E-R) のように関係を直接扱えるものもある。ただし E-R モデルでは実体と関係それぞれが固定されているので、関係自体を実体としても扱いたいときまたはその逆に実体を関係として扱いたいときにそれができないという問題などが残っている。

既に述べたように、意味には相互に重なりを持った階層がある、したがって表現としては再帰的または差分的な表現になるというのが第 1 の特徴であ

る。次に概念には相対性がある。つまり上位と下位が絶対的ではなく、下位の概念の下にさらに下位の概念ができるということで、上位と下位は、状況により変化する相対的なものである。相対性としては上位、下位以外にも実体と属性、例えば女性とか男性とかは人間の属性になるけれども、見方によってはそれ自身で実体になるというような相対性、それから関係と実体も固定的ではない。例えば私が車を持っている、私と車の関係は所有するとか所有されるという関係であるが、所有という概念は関係としてだけではなく実体にもなり得る。それから先ほどの類似性のような部分的重複も表現が難しい。意味表現の問題は外延 (extension) に基づく既存の情報技術では適切に扱えない。

2.10 情報の特徴と課題のまとめ

情報が持っている特徴と、それらに関連する課題をまとめると表3のようになる。既に述べたように情報というものは数えられるものとして扱われることが多いが、本来可算集合ではないということ、それから計算機では2値論理が処理し易いが、計算機からみても理論的に見ても厄介な多値論理が情報の本質である。それから様相ということを少し述べたが、1階の述語論理の範囲に留まらないのも情報の特徴であるということ、それから表現に多様性がある、つまり多数の同意語のあるのが用語の基本的な特徴である、それから表現されたものには多義性があって曖昧である。辞書を引くと、例えば一つの英語に対して日本語が一つだけ対応しているという言葉は殆ど無い。通常非常に沢山の種類の訳語が書いてある。意味の表現の他に表現の意味解釈の問題があり、それは計算機では情報そのものを扱っているのではなく、媒体上に記述表現されているものを対象としているということから生じている。以上情報の特徴と課題についていろいろ研究もされているし、又解釈しなければならないことが情報には沢山あるということを述べたわけである。

3. 情報知識学の専門領域

情報知識学というものは全体としてはどういう領域に対応する専門分野になるかを自然科学の側から整理してみると、情報知識には理論的な側面と実験的な側面および応用的な側面がある。

表3 情報の特徴と課題

特徴	課題
非可算性	分類
多値論理	大量情報管理 (識別、同定)
高階論理	演繹知識獲得
表現多様性	帰納
表現曖昧性	仮説生成
意味表現	類推
表現解釈	発想

3.1 理論情報知識学の構成

理論的な分野は先ほど述べたことから、まず情報の解析特に構造の解析方式であり、次に分類するにしても情報の表現を考えるにしても、構造を意味に対応させるにもモデルが必要になる。また分類の可能性と分類の手法、それから媒体に関連して記述と表現、記述表現の多様性、その取扱い方、情報の時間的、空間的、意味的変化、管理の可能性、限界などが情報知識学の基本的な面である。

3.2 実験情報知識学の構成

実験的な分野としては、実際の情報を対象として情報の特性、情報の量、情報の質、情報のキャラクタリゼーション、情報の資源化、管理および典型的な情報媒体の要素であるターミノロジやシソーラス、辞書、日本語、マルチメディア、それらの構築、特性、操作処理などが実験的な情報知識学の領域である。

3.3 応用情報知識学の構成

理論や実験が進めば応用も具体的に展開できるわけであり、情報検索の手法は確立されており、数多くのシステムが開発、提供されている。もう少し情報を高度に加工して付加価値を図ること、情報の伝達法として従来からの印刷物とオンラインデータベース、バッチ型データベースや知識ベースなどの位置づけと展開、それから学習、類推、発想も実現し、さらに最終の目的である問題解決、意志決定、評価、人工頭脳までも一応応用情報知識学の対象に含まれる。

3.4 情報学研究事例

情報知識学の研究にはいろいろな側面があるが、情報活動は結局情報の生産技術、提供、利用の技術

にかかわっている。生産では資源化と管理、提供では管理と伝達、利用では情報を伝達と処理をする具体的な技術ないしはそれの理論が関係するわけである。それぞれは特別の専門との境界領域でもあるので、関連分野との学際的視野での活動となる。このような機能が要求されているわけだが、実際にはそれぞれの領域の情報の生産者、出版社、通信や計算機の研究者、技術者がお互いに近寄って初めてこういう技術や学問が確立する。だいぶ一般的な話をしたが、これまで充分な研究が行われておらず、情報知識学会として重要な研究テーマについてすこし具体的に説明をする。

用語の意味関係：

用語の間には様々な関係があるが、我々は以前からいろいろな用語データベースを作って、用語間の関係を、例えば同意語、多義語、階層関係、部分全体関係などを抽出して用語の間の関係を扱えるようにしたシソーラスを自動的に作っている。これは検索のみならず意味処理にも有効である。

因果関係の種類：

同じような積み上げ方式によって概念構造を表すシソーラスだけではなくて論理関係とくに因果関係も自動的に収集構造化することができる。因果関係にも各種のものがあるけれども、自然科学で重要なのは直接結果に結びつく原因結果関係と、いくつかの要因があって結果に結びつく要因結果の関係及び必然性が充分ではないけれども何らかの理由で結果につながる理由結果などの種類がある。これらは構造化すれば演繹推論は単なるナビゲーションとして実現でき、シソーラスと併用して類推も実現できる。

構造化情報の持つ意義：

これらの関係情報を抽出すると、シソーラスとして概念間の構造が組織化されるので、それには先ほどの各種の関係が含まれるわけであるが、例えば類似関係というようなことが直接扱えるようになり、情報の利用に関して非常に重要になる。また論理関係はタキソノミーとして構造化される。更に元の情報が持っている書誌的な情報、つまり物理的な構造などはシステム的に扱い易い基礎的構造である。つまり情報が持ついろいろな意味を構造化することによって、今までに述べた範囲内ではあるけれども計算機で意味が扱えるということである。

同値関係と同義語集合自動抽出：

情報構造の実現方法を簡単に述べると、例えば日本語と英語の対訳用語集には英語にたいして日本語が対応関係が示してある。基本的には用語の訳は同値関係になるが、実際には人間の考えとしては同値関係の場合に上下関係も入れることが多い。それを全てが同値関係だけだとすれば、推移則が成立するので推移閉包をとり、単に全部の同値な用語を結んで同意語集合が得られる。例えばこれは JIS の用語集だが難燃性と同じ意味の表現が “燃える” という表現に対して “炎” と “火” もあって、“難” には “耐” があって、性質を表すのに “性” と “度” がある。このように考えられる組み合わせがほとんど全て使われている。JIS は勿論標準化の為に作るので用語も標準化されているが、それは専門分野別に行われるので全体としては標準化にはほど遠いということであり、これが先ほどから述べている言葉というものの多様性の典型的な例である。これは学術用語でも同じであり、学術分野毎に用語も標準化されているが、標準化されたものが全分野に共通になっているのではなく、広く使われる概念であればあるほど多様な表現が使われている。

シソーラスの自動編集システム：

いろいろなことばについて各種の抽出の仕方があるが、先ほどの上下関係や入れ子構造になる再帰関係がある場合には多義性によるノイズが拡大されるので、上位概念を抽出して推移閉包を求め、その結果を上位概念に結合することにより同意語集合の精度を上げることと、抽出された上位概念はそれを利用して階層関係も構造化できるということで割合簡単な方式でシソーラスができる。それから他の論理関係などについても類似の方法で構造化ができる。

自己組織型情報ベース：

上で述べたような情報の構造化を行って実際の研究開発に役に立つような応用システムの構築の例を示す。そのシステムは Information-Sase System-swith Self Organizing Receptor Interconnections, IBS:SORITES と名前付けられている。要点のみを述べると、情報の持つ階層性、相対性および部分重複などの基本特性は従来のグラフ構造型のモデルでは扱えないので、多項関係を扱えるハイパーグラフに内部構造や意味関係表現のラベル付けなどを拡張した新しいモデルを構築し、それに基づいてシステム開発を行っている。IBS のモデルはハイパーグラフを階層化、ラベル付け、および方向付けの点で拡張した新規のものである。それに基づき検索や演

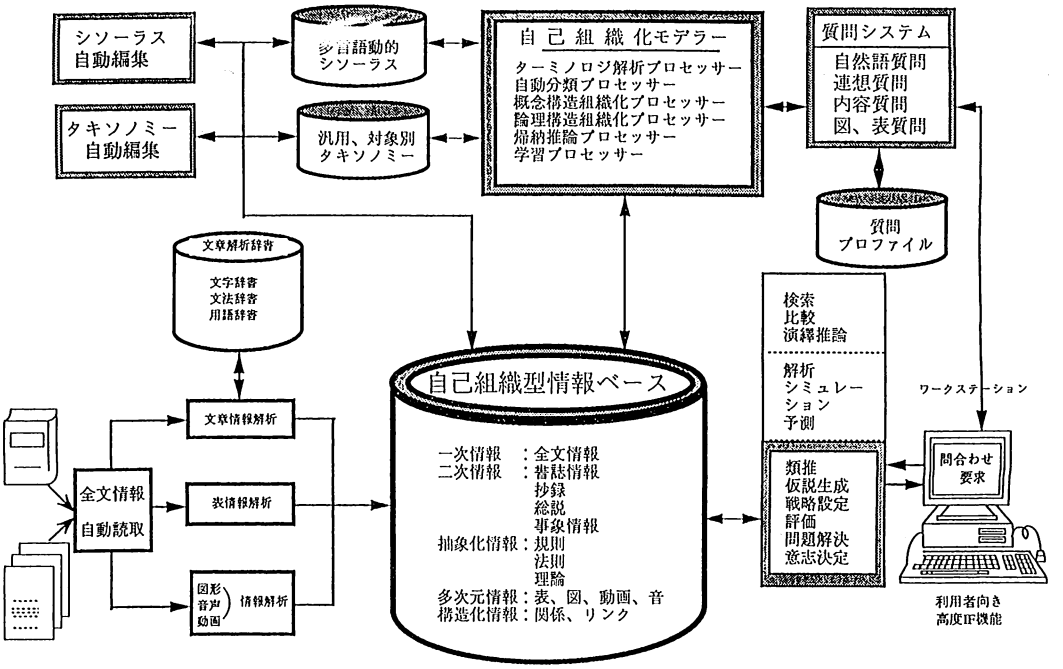


図 6 情報自己組織化研究—概念構造に基づく情報自己組織化の研究

繹推論のみでなく類推や帰納推論が使えるシステムが実現できる。全体構成としては図 6 に示すように、まず一次情報を CD-ROM に入れておく。理由は CD-ROM の記憶容量が大きく、540 メガあるので専門家に必要な情報がほぼ網羅的にこの中に入ることと、読み取り専用記憶装置で書換ができないので管理が非常に簡単になることである。次に一次情報から概念構造をシソーラスとして、論理構造をタキソノミの形で抽出し、それを用いて一次情報を構造化して意味処理に使うという方式である。このシステムは種々の情報に応用されているところで高分子、NMR、有機合成、半導体、超伝導、非線光材料、常温核融合等が対象となっている。

4. むすび

以上述べたことが結論として次のようになる。まず情報知識学というものをどう捉えるかということでは、新しい専門分野であるということは、従来からある分野のものとは異なる固有の活動成果がなければならない。そのためには情報知識の本質的特性や理論および応用に関する専門的研究の論文やレポートを会誌に出すこと、まとめて専門書や入門書をつくること、研究交流のため研究会や年次大会や国際会議を開いて研究の向上をはかること、情報知識学の普及啓蒙のため専門家の養成、講習会、講演会、研修会などを開くことが必要である。また情報知識学の領域の標準原案や専門資格として、国家試験や専門家養成のカリキュラムを作る必要がある。少し積極的に言えば国家的ないし国際的プロジェクトを提案するというようなことがあれば更に望ましい。これらのことを実施するには、専門性の確立と計画的な専門の活動が必要である。それが世の中に認められるには活動の量と範囲が広がる必要があり、そのためにはどうしても会員の増強が重要になる。これらが総合されて学会と情報知識学が普及定着することを期している。

本稿は「情報知識学会」総会において、1992 年 9 月 17 日（木）に行った講演の内容を改訂したものである。

文 献

1) C. Berge: "Hypergraphs" North-Holland (1989).

2) Y. Fujiwara, N. Uda: "Self Organization of Information in Libraries Based on Terminology", Proc. of Int. Conf. on National Libraries Towerds the 21st Century (Taipei)S4 p. 131-137 (1993).

3) E. F. Codd: "Extending the Database Relational Model to Capture More Meaning" ACM Trasactions on Database Systems, 4(4) p. 397-434 (1979).

4) P. P. S. Chen: "The Entity-Relationship Model:Toward a Unified View of Data" ACM Transactions on Database System, 1(1) p. 9-36 (1986).

5) J. Banerbee, W. K., H. J. Kim, and Henry F. Korth: "Semantics and Implementation of Schema Evolution in Object-Oriented Databases" ACM SIGMOD, p. 311-322 (1987).

6) W. Kim, J. Banerjee, H. T. Chou, J. F. Garza, and D. Woelk: "Composite Object Support in an Object-OrientedDatabase System" In OOPSLA '87 Proc. p. 118-125 (Oct.1987).

7) M. Stonebraker, B. Rubenstein, and A. Guttman: "Application of Abstract Data Types and Abstract Indices to CAD Data" In Proc. of Ann. Meeting Database Week, p. 107-115 (Sun Jose, 1983).

8) J. M. Smith and D. C. P. Smith: "Database Abstractions:Aggregation and Generalization" In ACM TODS 2(2), p. 105-133 (1977).

9) D. Maier, J. Stein, A. Otis, and A. Purdy: "Development of an Object-Oriented DBMS" In Proc. of the First ACM Conf. on Object-Orient Programming Systems Languages and Applications, 21(11), p. 472-482 (1986).

10) S. B. Zdonik and D. Maier: "Readings in Object Oriented Database Systems" Morgan Kauffman, (1990).

11) H. Boley: "Directed Recursive Labelnode Hypergraphs: A New Representation Language" Artificial Intelligence, 9(1) p. 49-85 (1977).

12) Y. Fujiwara, Gyoto Chang, Y. Ishikawa: "A Dynamic Thesaurus for Intelligent Access to Research Databases" Proc. 43rd FID Conference (Helsinki,1988) p. 173-181.

13) H. Sano and Y. Fujiwara: "Syntactic and semantic structure analysis of article titles" J. Inf. Sci. Prierciples of Practice Bull(2)to be published (1993).

14) Y. Fujiwara, Zhog Qing Wang: "The Multicategorical Structures of Information for Inferences and Reasoning in the Self-organizing Information-Base System" Proc. of CAMSE 2 to be published (1992).

15) Y. Fujiwara, N. Uda, Wangyu Ree: "Analogical Reasoning in Polymer Information-Base Systems" CADATA Bull 24(2) p. 59-66 (1992).

16) Zhong Qing Wang, Si Qing Zheng, Xu Yu, K. Yamaguchi, H. Kitagaw, N. Ohbo, Y. Fujiwara: "Learning and Anarogical Reasoning in the Information-Base System for Organic Synthesis Research" J. of Japan Soc. of Inf. and Knowledge 2 p. 71-82 (1991).

著 者 紹 介

藤原 譲（正会員）

1957年東京大学工学部応用物理学科卒業．同年（株）クラレ入社．中央研究所，ノースカロライナ大学，スタンフォード大学留学を経て1976年より筑波大学電子情報工学科系教授．基礎情報学，とくに情報構造解析，モデル化，データベース，情報ベースなどに関する研究と応用システムの開発を行っている．電子情報通信学会，人工知能学会，情報科学技術協会，情報知識学会，ACM，IEEE，AAAI，ASIS，ACS，ASTMなど会員．