８.構造操作　：記憶構造，直接アクセス，構造経由アクセス，同型性，準同型性，拡張関係型操作－関係グラフ，抽象化（汎化，集約化）

９.意味処理　：内容検索，演繹推論

類似性処理：共有概念，類似度－帰納，類推，仮説推論，連想，発想など

１０.応用システムおよび展望：自己組織型情報ベースシステム，人工頭脳など

（２．４）情報の特性と課題の例

まず情報の基本的特性を挙げると次に示すようなものがある.

a.媒体依存性

b.記述，表現の多様性

c.様相性（Modality）

d.非加算性

e.階層性（入れ子構造）

f.相対性，双対

などがあり以下に簡単に説明する.

（ａ）媒体依存性

情報はそれ自身で実在することは少なく通常なんらかの媒体上に記述，表現されるので必然的に記述および表現の形式が媒体に依存することになる．例えば風景を表現するのに写真を用いるか文章を用いるかを比較してみれば違いは説明するまでもない.

媒体として見ると文字に比し画像や音声は抽象化の水準が低いが，情報量が多く理解も容易である．これが先に述べたマルチメディアへの期待につながっている.

（ｂ）記述，表現の多様性

情報の媒体が多様であるので記述，表現の多様性があるのは，避けられないことであるが，同じ媒体であっても想像以上に様々な形態をとり得る．典型的な例は言葉でいえば同意語である．一般的に情報の記述，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１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

図１ 概念階層の束構造

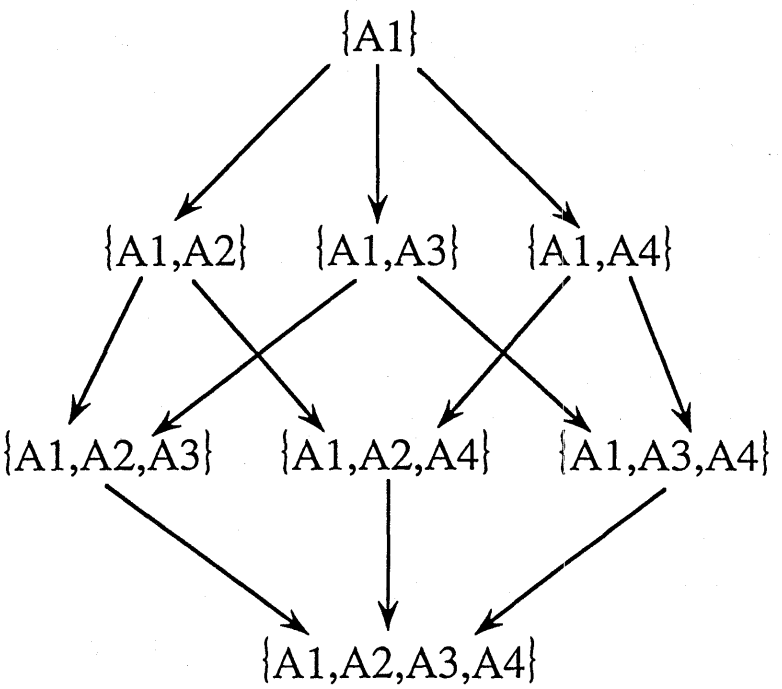
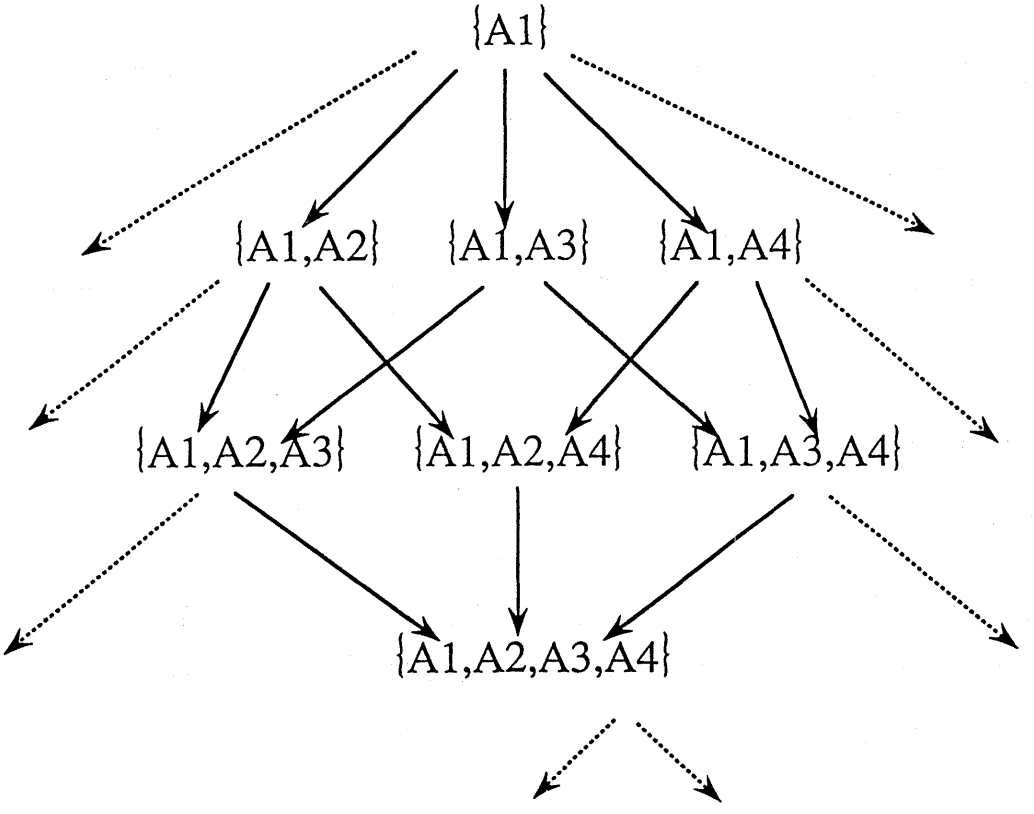


図２ 属性附加による概念階層の束構造の変化



でも記述の仕方は１６通りになる．さらに属性が一つ増えると図２に示すように階層も深くなり，かくレベルのノードも増加するだけでなく全体の構造も変化し記述法も３２通りとなる．なお分類の多様性も同じ構造で説明出来て，それぞれ２４通り，１２０通りとなる．一般にｎヶの属性に対して記述法は２のｎ乗通り，分類法はｎ！通りとなる．

（ｃ）様相性

検索やＡＩで符号照合のとき，

Ａ＝Ａ

と

Ａ≠（～Ａ）

は対偶であるから同じことを意味するとしたり，一階述語論理で，「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，２値論理が前提である．ところが使われる情報は２値論理型とは限らない．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．

（ｄ）非加算性

意味の関わる問題の一つは個別実体(Distinct Entities)の集合を通常仮定することである．順序関係の成立する外延(Extension)として概念を取り扱うことは対象を著しく制限することになる．

（ｅ）階層性

情報，概念の間には抽象化や総称表現に基づく包含関係などのため階層関係があり，とくに技術の進歩や生活様式の変化による新しい概念の生まれることが多く，入れ子型の構造になる．

（ｆ）相対性（双対）

実体と実体の間にある関係はそれぞれが固定されているのではなく，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があり，これを双対(dual)ちう．また実体と属性，階層関係における上下関係なども状況に応じて変化するので相対的である．これも従来型のシステムでは扱えない問題

である.

（２．５）情報の資源化

最近マルチメディアが注目されているが，ハイパメディアはマルチメディアの有力な利用形態のひとつである．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Halazsが８７年と９１年に改訂しCACMに発表した問題の一つは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．このことは，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課題になっていることを示している.

少し個別な問題でアクセスの問題を考えてみる．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ったが，マルチメディア情報は上で述べたように本質的な問題点を持っており，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.

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，現在では最も柔軟なシステムと考えられているハイパーメディアとオブジェクト指向的ＤＢＳも基本的には満足できるものではないことになる．一般的に従来型の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，前者は管理，とくに識別，同定からの制約のため，後者では知識の表現の制約から知識の獲得が困難であり，いずれにしても入力できるものが限られる．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.

結局，基本的には利用者向きで情報媒体の面からも望ましい大量のマルチメディア情報の利用のためには情報の特性に即した新しいモデルに基づくシステムの開発が必要である.

（２．６）利用機能

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も，完全ではないが種々の手法があり，実際に使われている．

更に高度な機能として，類推，帰納推論，仮説推論，発想，連想などと，それらを複合して問題解決，設計，意志決定，評価などが要求されている．

このような高度な機能実現のためには意味とくに類似性，関連性の処理が重要であるが，情報が媒体経由の間接表現のため困難な問題である．しかし意味の関係を概念間の関係として構造の形で組織化ができれば，意味処理に道が開けることになる．大量の情報の構造化は人手で行うことは極めて困難なことであるから，システム的に，即ち自己組織的に行わなければならないし，そのような試みがなされているので以下に一つの例を示す．

（２．７）新しい情報システムの展望

（ａ） 意味関係の構造化

実体や概念の間の様々な関係は主として用語の間の関係としてあつかうことができる．専門用語のデータベースを作って，用語間の関係，例えば同意語，多義語，階層関係，部分全体関係などをC-TRAN(Constrained Transitive Closure)およびSS-KWIC(Semantically Structured Key Word Element in Terminological Context)などを用いて抽出して用語の間の関係を扱えるようにしてシソーラスを自動的に作ることができる．

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対して日本語の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，実際には用語の使われかたとして同値関係の場合に上下関係も入ることが多い．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，全部の同値な用語を結んだ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．
いろいろな用語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．
自動作成されたシソーラスは概念構造を表し，情報の構造化による意味処理のみならず内容検索にも有効である．

ＳＳ－ＫＷＩＣは専門用語が主として複合語であり，構成要素間に造語規則が存在することを利用して階層関係や関連関係を抽出する方法である．

同じような積み上げ方式によっ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．これには SS-SANS (Semantically Specified Syntactic Analysis of Sentences) および SANS (Semantic Analysis of Sentences)を用いる．前者は先ず特定用語中心とする一定の構文を利用して，概念間の関係を抽出する．次にその結果を用いて新しい特定用語と構文を得る．これを再帰的に繰り返す方法である．この方法は構文の不明確な文章や，構文のない用語の集合例えばキーワード集合の間の関係も抽出できる方法である．概念間の論理関係として，因果関係にも各種のものがあるが，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．これらを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，概念構造を表すシソーラスと併用して類推も実現できる．

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と索引など，物理的構造は基礎的構造である．
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．

（ｂ）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

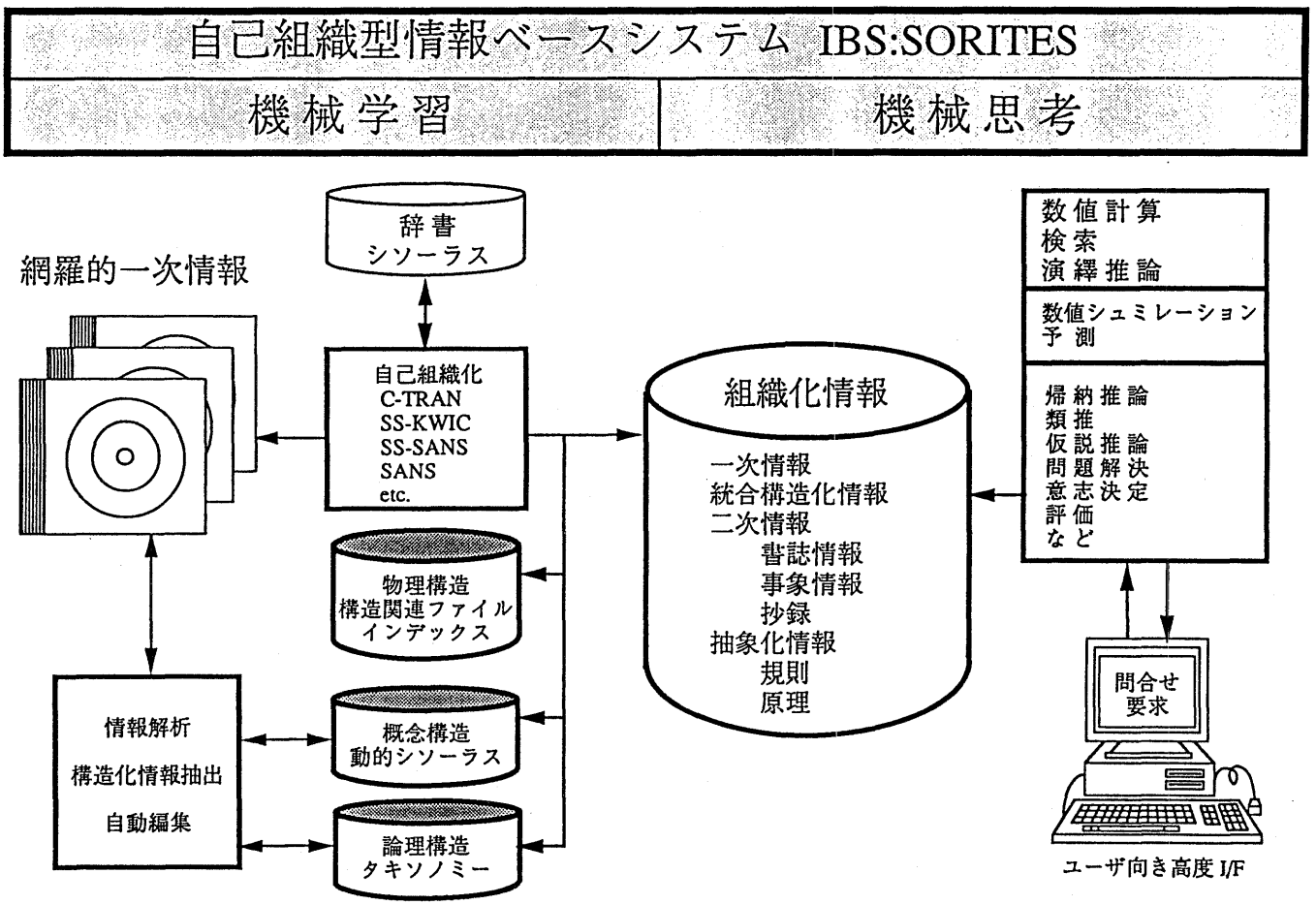
システムの構築の例を示す．そのシステムはInformation-Base Systems with Self Organizing Receptor Interconnections，IBS:SORITES と名付けられている．

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，および役割を示す方向などを導入して拡張した新しいモデルＳＳＲ(Structured Semantic Relationship)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３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などである．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ーの形で抽出し，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研究用マルチメディア型情報に応用され高分子，ＮＭＲ，有機合成，半導体，超伝導，非線形光学材料，常温核融合等が対象となっている．

図3　IBS：SORITESのシステム構成



2．情報学研究の現状と展望

2．1　言語情報

（1）言語技術の情報学における位置づけ

情報を担っているものは言語だけでなく，図面，写真，映像，その他種々のものがある．それらは表現しようとする情報内容によって使い分けられる．情報はまたいかに客観的に相手に伝えられるかという立場からその媒体を考えることもできる．機械の設計図面などは世界中で共通的に理解されうるものである．それでは言語はどのような情報の表現に適しているのだろうか．あるいは情報学における言語の位置付けはどうであろうか．それは次のように考えられる．

（i）言葉は思想を表現するための最適の媒体である．

（ii）言葉は誰にでも理解でき，人による理解の相違を最も小さくし，正確な情報を伝達することのできる媒体である．

（iii）歴史的に見て，人類の知的財産が言語によって最も多く表現され，また蓄積され今日に伝えられている．

（iv）今日の情報技術においては，言語が最も安価に，最も容易に扱える媒体である．

このような理由から，言語技術は情報学の中で重要な位置を占めていることが分かるだろう．

（2）言語技術の現状

（i）字づら処理

言語を扱う場合には，それを構成する基本である文字，単語が明確である必要がある．文字としては欧米諸国言語のアルファベットのように100文字以下の場合と，日本語，中国語のように数千～数方文字の場合がある．数千あるとされる世界の言語のうち文字が確定している言語はそれほど多くないとしても，やはり膨大で，これを計算機でユニフォームに扱うために，1文字を何バイトでどのように表現するのがよいかが現在真剣に検討されている．文を構成する単語を確定する形態素解折は言葉の持つ多義性の問題からユニークには決定できない．現在日本語では単語単位で99．9%程度まであげる努力をする必要があろう．

（ii）文解析

文中の各単語の文法的役割を明確にすることであり，多くの場合各単語のもつ

意味にまで踏み込んで扱わねば正確な解析は出来ない．文解析の中心的役割を占めるものは文法であり，過去30年の間に種々の文法形式が提案され，それらのいくつかは計算機による文の自動解析に用いられた．しかし，いまだに種々の複雑な文を90%以上の正確さで解析することの出来る十分精密な文法は作られていない．最近は従来の文法の概念でなくニューロネットワークや部分的な文の類似性などを発見する方法など，種々のヒューリスティック手法が試みられている．一般的には非常に困難であった長い文の解析もかなりできるようになり，照応の問題，省略語句の推定，文脈関係の把握といった問題が研究されている．

（iii）文生成

最近になってようやく研究が盛んになって来たが，何から出発して文を作るかが明確でなく，研究方向が定まっていない．質問に対する応答や定まったパターンの文を作り出す程度のことは出来るが，話者の聴者との関係，聴者に対する心的態度，話すべき内容をどのような方略でどのような文形式でどのような順序に従って読みやすい文脈的表現にしながらまとまった1つの文章として作り出すかは未だに全くといってよいほど未解決である．

（iv）翻訳

言語の機械翻訳は不完全ながら実用されている．欧米では英語，フランス語，ドイツ語などを中心として使用されているが，日本においてはほとんどが日本語と英語との間の翻訳である．１文ごとの翻訳しかできないこともあって，人手による後修正が必要となる．ただどのような内容の文章であったかを把握するためであれば後修正なしに使うこともできる．ほとんど後修正を必要としない質の高い翻訳システムは21世紀の初頭まで待たねばならないだろう

（v）テキスト情報の圧縮と分類

与えられたテキストからキーワードを自動抽出したり，自動要約をしたり，またそのテキストがどのような分野のものであるかを自動判定したりする研究が盛んになって来た．現在まだ実用になるようなしっかりした方法は確立されていないが，社会的な要求も強く，研究も進んでいるので近い将来実用となる方法が出てくるものと考えられる．

（３）これからの課題

（i）言語技術基盤の確立

言語技術を発展させるためには研究開発のための環境整備が必要である．それは，膨大なテキストデータの蓄積（特に多言語翻訳対で），これらのテキストデ

ータに出来るだけ豊富な言語情報を付与したテキストデータベースの整備，膨大な多言語の単語・フレーズ辞書の整備，形態素解析，構文解析，その他のソフトウェアの整備，これらの全ての研究者への公開ないし低価格による配布，などであり，誰もが自分の目的とする言語処理をすぐ行えるような環境整備が重要である．米国を中心としてこのような環境整備の動きがあり，日本でも早急に検討しなければならない．

（ii）言語理解のための知識辞書の作成

我々人間が言葉を理解できるのは，文法を知っていたり，単語の意味を知っているというだけでなく，言葉によって語られている外界・対象に関して種々の知識を持っているからである．機械に人間と同じように言葉を理解させ，適切に応答させるためには人間の持つ世界に関する知識を機械が取り扱える形に整備しなければならない．これがこれから挑戦すべき最大の問題である．

（iii）電子図書館システム

今後ほとんどの出版が電子的に行われるようになり，またワープロによって文書が作成され，電子メールシステムで流される時代になるが，その時の図書館はこれらの活動にマッチした形の電子図書館となるだろう．そして世界中の電子図書館がネットワーク接続され，多様なユーザの要求に対応しなければならなくなってくると，以上に述べたあらゆる言語技術が必須ものとなる．すなわち情報学における言語技術は，十分な意味における電子図書館の実現と言いかえてもよいのである．

２．２　情報標準化

二つの流れ

情報に関する標準化には，科学技術会議が総理大臣に答申した「科学技術情報流通に関するる基本政策」に従った情報流通技術の標標準化（ＳＩＳＴ）と，従来展開されてきた日本工業規格（ＪＩＳ）での情報処理技術の標準化の二つの流れがある．いずれもそれぞれの自的に沿って，情報の流通・処理における技術の整合性を高めることに主眼を置いている．この経緯には，科学技術振興における支援活動として早くから進められてきた前者と，機器の進展に比して産業化の遅れた情報活動に対する後者の認識の差が認められる．経緯はともあれ，両者とも国際標準への整合を掲げており，その点で共通の基盤をもっているということができる．強いていえば，後者が工業技術の延長上で「情報」を眺めているに対し，

前者は科学技術が内包する「情報」に視点を置いているところに相違がある．また，前者が20年の実績もっているに対し，後者は国際標準化機構への国家対応の公式規格であることに，それぞれの特徴を有している．制度的には，前者が科学技術庁科学技術振興局の発行文書であるに対し，後者は工業標準化法にもとづく法律の認知を受けている．それだけに憲法が保証する表現の自由の限界の内にあることの認識が重要である．内容的には前者が書誌情報および情報生産を対象としているに対し，後者は用語・略語・記号・符号などか法律にその対象として例示されている．両者の境界については微妙なところがないでもないが，ＳＩＳＴの一部　ＪＩＳ化も実現しており，この面の発展のため今後の協力が期待されている．

ＳＩＳＴ

現在までに制定されたＳＩＳＴ（科学技術情報流通技術基準）は１３にのぼり，大別すると，抄録作成・参照文献記述・レコードフォーマット形式など書誌情報に関するものと，学術雑誌構成・学術論文構成・科学技術レポート様式・予稿集様式などの情報生産に関するものの２種となる．作業部会が作成する基準案を，学協会・大学・研究機関・関係省庁からの研究者・情報専門家からなる諮問機関（科学技術情報流通技術基準検討会）が審議する手順をとっており，適用者が情報作業を実施する際の高質化，効率化への奇与を図っている．「基準」のそれぞれの特性や適用者の事情が勘案されて，適用および記述のレベルにバラツキのあるのは発展期においてやむをえないのであろう．普及については，ハンドブックの発行や説明会の開催などの実績を積んでいるが，全般的な利用に一層の努力が必要である．

内容では，電算機利用が大きな主眼になって，多くは機械可読情報を対象としているが，電算機の使用を前提においているのだから，印刷形式の基準文献を発行するばかりでなく，基準の電子形式の発行，さらに一歩進んで，適用例のパッケージを作成提供するにとによって，適用者の便宜を図るなどの基準開発が課題であろう．

ＪＩＳ

ISO/TC 46（情報ドキュメンテーション）の制定した国際規格の中には，日本国内の規格として必要とみなされていたものが散見されていた．その内のいくつかが，1988年，JISとして制定された．それが，

JIS X 0304-1988 国名コード

JIS X 0305-1988 ＩＳＢＮ

JIS X 0306-1988 ＩＳＳＮ

である．その後，これらの規格に加え，用語規格として JIS X 0701，0702，0705，0706 のドキュメンテーション用語が出版された．現在は，電子出版やオフィス・ドキメュンテーション分野で注目されている

ＳＧＭＬ（Standard Generalized Markup Language）のＤＴＤ（Document Type Definition），各種情報検索サービス間のコマンドの差異を乗り越えようとするコモン・コマンド言語等の新たなＪＩＳ原案の作成作業や，JIS 国名コードの改正原案の作成作業など，情報の標準化に関して，広範囲な活動が継続されている．

ＩＳＯ（国際標準化情報）への対応

標準化の国際専門活動として，ＩＳＯ（国際標準化機構）があり，情報学分野に関連の深い下部組織として，TC 37（ターミノロジー技術委貫会）と TC 46（情報ドキュメンテーション技術委員会）がある．ＩＳＯへの日本の対応は工業技術院が窓口であり，活動全般に対しては日本工業標準調査会（ＪＩＳＣ）が事務局を担当し，国際規格の原案作成から制定までの各段階で，内容の審議を各国内対策委員会が行っている．

ＩＳＯ・ＩＥＣ指針が示すように，ＩＳＯの各ＴＣ（技術委員会）は TC 37 が制定した用語原則に従うよう示唆されており，TC 37 の活動は全標準化作業の基本になるものであるが，日本では国際活動への寄与へ踏み出したところであり，また，国内での標準整備が現在の課題となっている．一方，TC 46 分野では，対応のための国内体制はほぼ整っており，積極的な提案にいたるまでの活発な活動を目指している．

２．３　全文データベース

全文データベースの発達　－－　いわゆるデータベース・サービスは，文献の書誌的データと要旨を収録した二次文献データベースを中心に発達してきた．これには，従来の抄録誌の編集作業が電算化され，そこで得られる電子化ファイルが，当初は副産物的位置付けにあったが，次第に主製品としてのデータベースに転化してきたという背景がある．一方，数値情報系のデータベースは，株価の即

時配信システムという，データベース的ではないオンラインシステムからはじまったが，その後，データの蓄積機能を取り入れてデータベースを構築するようになった．また統計情報関係では，統計データのデータベース的整備と，これに対する分析プログラムの組合せでサービスするシステムが早期に事業化されている．全文データベースは，上述の二次文献データベースの進化の延長上にあるものである．

全文データベースの発達は，図書・雑誌・新聞等，一般の出版物の原稿作成や編集におけるコンピュータ利用の浸透，また電算写植機による印刷の普及を背景にしたもので，この点，抄録誌のデータベース化と同様の経過をたどっている．こうして，現今ではオンラインでサービスされるデータベースの半数以上が全文データベースになっている．また，パッケージ型データベースと称される半数以上が全文データベースになっている．また，パッケージ型データベースと称されるＣＤ－ＲＯＭによるデータベース出版物は，近年わが国でも普及が進んでいるが，これらの多くが内容的には全文データベースであると考えられる．

オンライン系全文データベースでは，通信量と通信料金の制約，また標準化の問題から，図表の類を除外した，本文だけの全文データベースがほとんどであるのに対して，こうした制約の小さいＣＤ－ＲＯＭ版データベースでは，むしろ画像・音声に主力をおいたマルチメディアデータベースが盛んになっている．もっとも，高速・大容量・低価格のインターネットの普及に伴って，オンライン系でも，画像・音声を含むハイパーテキスト仕立ての全文データベースの構成方式ＷＷＷが普及のきざしをみせている．また，既存出版物の各頁を画像として蓄積・配信することも，インターネットの普及により現実的になっており，このような全文データベースの集積と配信を総合したシステムを「電子図書館」と称して，その実用化にむけた開発計画が米国ではいくつ試みられている．

文字・文章・文書情報の基礎的研究　－－　こうした状況をうけて，情報技術の研究開発では，マルチメディア，端的には画像・音声に関係したものが多くとり上げられるようになっている．これに比べて，テキストデータ自体に関する研究は地味な印象を与え，またビジネス機会の観点からも興味薄とみられるせいか，報道されることも少ない．しかし，オックスフォード大のテキストアーカイブズのように，全文データベースを広範に蓄積するという試みも着実に進捗している．こうした全文データの蓄積は，単に検索・参照の用に供するのみならず，むしろこれを実験試料として利用した様々の研究が広い学問分野にわたって可能になる

という点で，情報資源というにふさわしいものであると考えられる．

人間活動の多くが言語に依存し，またその記録や伝達の多くが，これを文字で表わした文書に依存していることは明らかである．このことは，画像・音声を別にした，せまい意味での全文データベースについて，その構築，蓄積，検索，表示などの基本的機能に関わる研究が，すべての学問分野の発展に寄与する基盤的な研究であることを示唆している．

ワープロの普及に象徴されるように，文書の電子化は社会のすみずみまで浸透しつつある．こうしたＯＡ機器・システムの普及は，文書の作成を効率化させるから，文字情報の絶対量も増大させるであろう．これを，文字情報の氾濫というような混沌事態に陥らせないために，文字情報の処理に関する基礎的・統合的な研究が必要である．例えば平文ファイルに対する高速走査検索手法は，全文データベース向きのシステムではすでに実用されているが，この種の研究もさらに推進されるべきである．また，テキストデータに多様な切口を与えるＳＧＭＬはすでに国際規格として成立しており，上記のテキストアーカイブズでも，この方式に依拠した全文データベースの蓄積が推進されている．ＳＧＭＬは，文字情報の氾濫に対する，文書の作成段階での対応法のひとつとみなされる．しかし，わが国では大方の支援を欠くためか，この方面での動きは鈍い．

上記のような視点に立つとき，現在，全文データベースに関連して展開されている様々な研究は，本来，文字・文章・文書情報に関する研究として，組織化・統合化されるべきものと考えられる．そのためには，これまで例えばデータベース，テキスト処理，機械翻訳，電子図書館など，個別の応用を目的に進められてきた諸研究を，適切に位置付けて統合化するような学問体系の構築が必要である．そして，この役割はまさに情報学が担うべきものであると考えられる．

２．４　マルチメディア

我々の日常はマルチメディアの世界である．文字，画像，音響，などさまざまな情報メディアに囲まれて生活してる．しかしメディアの多様性ということをほとんど意識していなかった．ところが１９８０年代に入ってマルチメディアという言葉が新聞や雑誌広告などにしばしば見られるようになった．これはコンピュータや通信の世界において単に文字や数値だけではなく，静止画のみならず，動画や音響も扱える様になりつつあろことを反映して特に強調されて出てきた表現である．

画像や音響データはそれをデジタル化すると膨大なデータ量になる．一枚の写真も文庫本１０冊分ぐらいの量になる．それを数千枚数万枚を扱うとなると大変な量である．従来のコンピュータではシステムが大きくなり過ぎた．ましてやパーソナルコンピュータでは不可能であった．ところが大量のデータを蓄積することが出来る媒体，光ディスクの出現によって事情が一変した．数百メガバイトのデータを蓄積することができる光磁気ディスクやＣＤ－ＲＯＭをコンピュータに接続することによって従来の計算機のメディアの世界が飛躍的に広がって，我々の日常的なメディアの世界に近づいてきたのである．

また写真や書類をコピーするような感覚でスキャナーから入力することが出来るようになった．更に動画を高速にデジタル化するＩＣの開発，画像データを品質を落とさずに圧縮する技術（ＪＰＥＧ，ＭＰＥＧ）の開発，高品質の画像表示装置，などの出現が大きな役割を演じている．コンピュータのディスプレイ上でビデオや音も出すことが出来るようになった．異なる媒体の情報をすべてデジタルにしてコンピュータ・システムだけで多様なメディアを扱うことが出来るのである．また大量のデータも光ケーブルや衛星通信を介して高速に伝送することが出来るようになって，コンピュータとネットワークの接続で利用方法が広がった．

画像や音などがコンピュータで扱える様になると，その応用範囲は急激に広がる．従来の知的生産の技術ではカードに文字を書いて情報を整理し活用する工夫がなされており，それのコンピュータ化はかなり進んでいるが，写真や音はまた別の媒体として扱わねばならなかった．それが文字も画像も音も同じレベルで扱うことが出来るようになって知的生産の技術が一段と向上したことになる．これは人文学の分野にとっても有効な方法を提供することになる．

このマルチメディア・システムがゲームや家庭内の情報機器として利用され出すと新しい応用が広がるというので話題になっている．

マルチメディアを扱うシステムの問題点はいかにしてデータの間に関係をつけるかということである．異なる情報媒体をどのような情報によって結び合わせるかということが最も重要である．単純なやり方は関係する対象をあらかじめ指定し，それをリンクする情報をデータとして持つ方法であろ．しかしこれは拡張性がない．新規に追加した場合またリンクをつけ直すというのでは大変である．それぞれが独立していてしかも関係をつける工夫がいる．

それにはそれぞれのメディア情報に対して自然言語による記述データをつけておき，その自然言語を介してリンクをつけるという方法がある．この時問題になることは自然言語の持つ表現の多様性である．同じような事柄に対しても異なっ

た表現をする場合がある．そこで類語集（シソーラス）を用意し，単語の表現を変換して一致を取ることにすれば柔軟なシステムになる．

２．５　情報自己組織化

現在の計算機では四則演算や符号処理，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．更に高度な機能になると，学習，類推，帰納，仮説推論やそれらを複合して問題解決，発想，意志決定，評価などをすることになる．これらが実現されれば本当に思考支援であり，人工頭脳的機能が実現できることになる．

このような高度な思考機能に対応する情報処理をするには情報の意味処理をする必要がある．ところが意味の関わる問題の多くは未解決である．例えばデータベースや知識ベースでは個別実体(Distinct Entities)の集合としてデータや知識を対象としていることである．つまり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基本的な考えである．実際に使われる情報では特許や法律でも化合物でも，概念に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，例えば総称表現が適切に処理できない．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同様である．あるべき情報が欠落している空値問題はさらに意味処理が困難である．

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も普及しているが，実体間および関係間の関係を扱う機能がない．一方実体一関係型（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．これは意味ネット，ハイパーメディアやオブジェクト指向システムにも共通する問題点である．

また意味の相互重なりに対する記述表現としては再帰的または差分的な表現の問題がある．次に概念には相対性があり，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あり得るので，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．例えばある人が車を持っている，人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ので双対関係である．それから先ほどの類似性のような部分的重複も記述表現が難しい．一般に意味の記述表現の問題は外延(extension)に基づく既存の情報技術では適切に扱えなく，情報の管理や，知識の獲得の困難さの原因となっている．

実際の例でいうと，図４に示すように，化合物の部分構造に関する包含関係のごく一部を取り出したものであるが，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，たとえば製品の情報でも製造場所，製造日時，原料材料とその特性，加工法，装置，条件などの多くの概念が多重の入れ子関係を含み複雑な関係になる．

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ので意味の処理が困難になるのである．この様な問題の解決策として脳における機能と同じように情報の意味的関係を自動的に構造化する自己組織化の研究が注目されている．

自己組織化方式

多様，複雑，かつ大量の情報を収録，管理し，それらの高度利用のため学習，類推，仮説生成，発想などが可能なシステムを実現するための基礎研究とプロトタイプシステムの例として「自己組織化機能を持つ情報ベースシステム」についてのべる．

脳における学習に対応して，概念や情報間の意味的関係を抽出して情報の組織化を行う．概念に内在する関係は概念構造としてシソーラスを構築する．論理関係は，原因ー結果，理由ー結果，要因ー結果な主としてタキソノミーの形で論理構造を組織化する．原情報の所在，書誌情報，アクセス情報などは物理構造として構築する．メディア変換，意味関係抽出等によりマルチメディア型原情報の概念構造，論理構造，物理構造などの自己組織化を行ない演繹推論，帰納推論，類推などの可能な人工頭脳型システムを設計する．研究に必要な情報の動的構造の記述操作のためには新しい型の情報構造を持つモデルＳＳＲ(Structured Semantic Relationship)の開発がなされている．．そのためデータベース，知識ベースおよびハイパーメディアなどの要素技術を大幅に拡張し，新しいモデルで統合的に記述表現，操作し，思考支援できる機能を持つ情報ベースシステムの設計と，

基礎となる理論の研究が行なわれている.

自己組織化は，自動生成されるシソーラスおよびタキソノミーなどを用いる概念構造化により実現する.

概念構造及び論理構造にもとづく自己組織化には次に示すような方法がある.

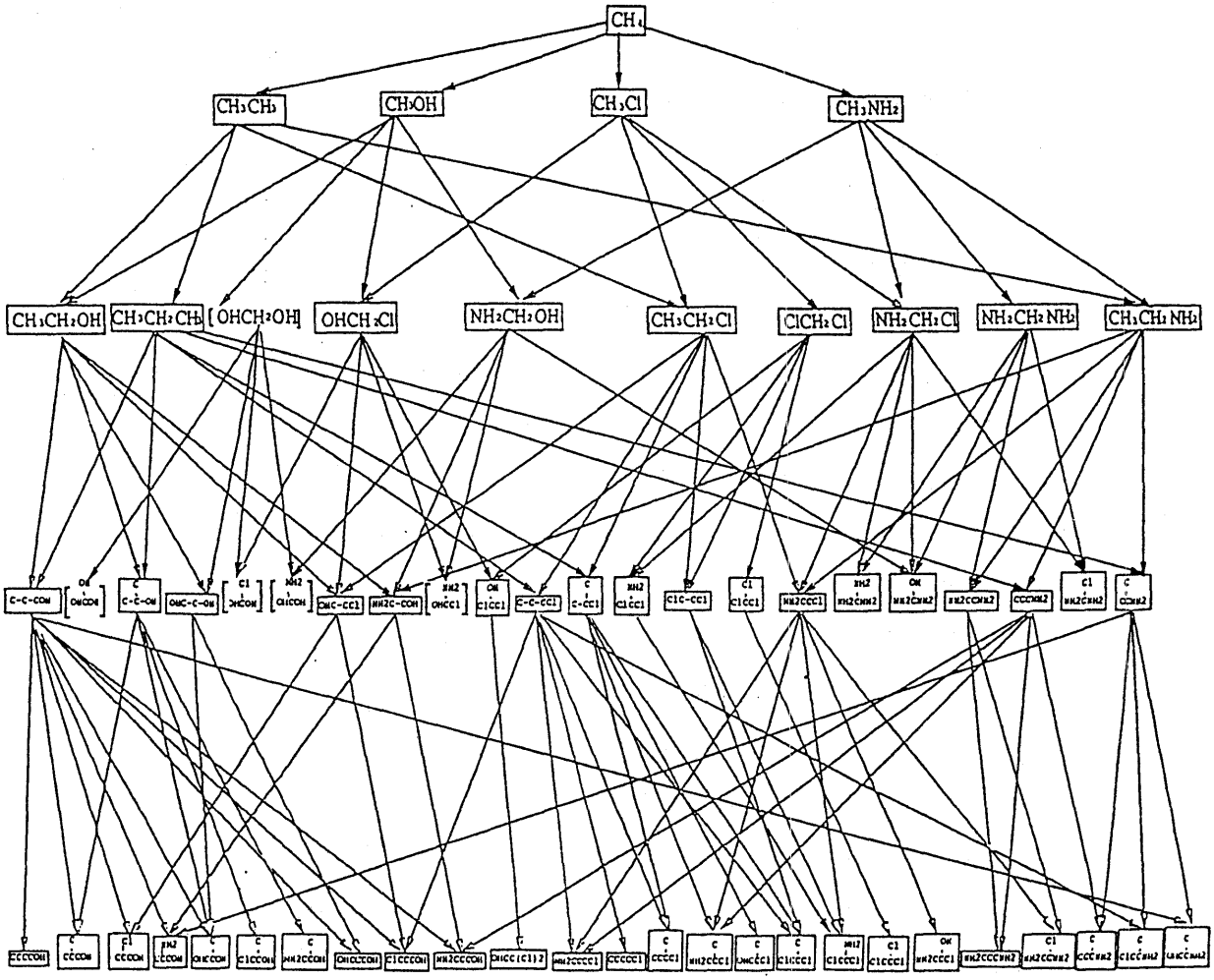
表現の多様性のために存在する同意語は，原情報に内在する概念の同値関係や，対訳用語集における対訳関係の推移閉包を求めるＣ－ＴＲＡＮ(Constrained Transitive Closure)で同値関係と副次的に階層関係が得られる．また複合語の造語規則を用いて階層関係と関連関係を求めるのがＳＳ－ＫＷＩＣ(Semantically Structured Key Word Element in Terminological Context)である．また意味解析のためＳＳ－ＳＡＮＳ(Semantically Specified Syntactic Analysis of Sentences)およびＳＡＮＳ(Semantic Analysis of Sentences)などを用いて，動的概念関係シソーラスや論理関係タキソノミー構築できる．これらに基づき類推，帰納や仮説生成による学習，問題解決，発想支援を実現する.

これに関して現在のハイパーメディアは小規模情報には有効であるが，リンクの生成，管理の限界およびグラフ構造の制約から大規模化が困難であり，柔軟性にも欠ける．　また類推，帰納推論などの研究も人工知能分野で広く行なわれているが，方法として確立されていない.

以上の方式は具体的対象として機能性材料，および有機合成などの研究用情報を取り上げプロトタイプシステムの構築と，そのテスト評価を行なわれている.

原情報の収集，入力提供には光学材料，および有機合成それぞれの専門家多数の協力によっている.

図４　化合物の概念構造


CH4
CH3CH3
CH3OH
CH3Cl
CH3NH2
CH3CH2OH
CH3CH2CH3
OHCH2OH
OHCH2Cl
NH2CH2OH
CH3CH2Cl
ClCH2Cl
NH2CH2Cl
NH2CH2NH2
CH3CH2NH2