情報検索における概念の問題

橋　本　　　寛

1. はじめに

概念の問題は，情報検索において，基本的問題の1つであるばかりでなく，広く種々の情報処理とくに自然言語の機械処理においても，重要な位置をしめており，これまで多くの議論がなされている［1,2,3,4］。

文献検索システムを例にとって考えると，概念は文献および検索要求を記述するために用いられ,一般にシソーラスの形で整理されている［5,6,7］。シソーラスでは各概念について上位概念，下位概念，同義語、同形異義語，関連語などが示されているが，さらに高度の処理をおこなおうとするときには，これ以上の情報が必要となる。

従来の概念の研究は，実用上の観点から，複雑な実際的問題に対して考察が注がれているが，自然言語の困難性のため，議論の見通しがわるくなり，またあいまいになる場合もある。たとえば，一般的に議論しているかぎり，新しい概念を合成するための概念間の演算すなわち概念算としては，通常の集合算の範囲をこえることは困難であるし，概念を分割していくさいに，その分割が主観的におこなわれるおそれもある。

本論文においては，上記のような概念に関するいくつかの問題について考察をおこない，その接近法に対する若干の手がかりを与える。また，ここでは，議論を具体的にするため，概念の領域を特定の集合に限定する。そうすることにより，概念間の演算が明確になり，概念間の包含関係も決定しやす

くなる。さらに，演算の意味を明確にすることにより，新しい概念算を導入することも可能となる。

そういう領域として，数学における行列の集合をとりあげよう。行列の集合において，正則行列，正規行列，直交行列，エルミット行列など種々の部分集合が定義されている。数学における用語であるから，定義を明確にすることにより，あいまいさをとりのぞくことが可能である。

2．概念間の包含関係

概念の定義としては種々のものがあるが，ここでは一定の全体集合の中で定義された部分集合を概念とよぶことにしよう。すなわち，一定の集合を伴うものを概念とよぶのである。通常はもっと広義に解されているようだが，本論文ではこのように限定する。

すでに述べたように，ここでは行列の集合を全体集合として考え，その中での部分集合，たとえば対称行列，対角行列，直交行列などをとりあつかうことにしよう。

したがって，２つの概念が同等かどうかは，それら２つが集合として相等しいかどうかを判定してきめることになる。それゆえ，意味の問題として議論されているもの，たとえば２つの単語が意味的に等しいかどうかという問題の一部は，その単語の指定する概念すなわち部分集合を比較することにより処理できることになる。しかし処理できるのは，あくまでも，意味のもつ側面の一部である。このような規準で２つの概念を比較すれば客観的判定が可能である。

以下，概念に関する基本的項目に対して考察をおこなっていく。

(1)　同義語

与えられた２つの単語が同義であるかどうかは，ふつう言語学では文章の中で２つの単語の置換が可能であるかどうかによって決定される。しかしこ

こではすでに述べたように，２つの単語の指定する概念が，同じ部分集合を示すとき，２つの概念の名称は同義語ということにしよう。

（例）

順列行列と置換行列

正則行列と可逆行列

ただし，このように集合として等しいものは同義であるとすると，通常同義語といわれているものよりも広くなってしまう。この問題については概念の定義に関して４節でも述べる。

(2)　条件付き同義語

ある２つの概念が，与えられた別の概念すなわち部分集合の上で等しくなることがある。たとえば，対称行列とエルミット行列は複素行列の範囲では等しくないが，実行列の範囲では等しくなる。すなわち

（対称行列）∩（実行列）＝（エルミット行列）∩（実行列）

である。

このような関係にあるものを，その基礎となる部分集合の上での条件付き同義語ということにする。このような例は他にもいくつかある。たとえば，複素直交行列とユニタリ行列の関係である。

(3)　同形異義語

同じ名称が別々の部分集合に与えられているときに生じる。異なる専門分野で，同じ用語が別々の意味に使用されることはしばしばみられることであるが，行列の名称に関してもそのようなものがみられる。たとえば，つぎのようなものがあげられる。

[ユニモジュラー行列]

（定義１）行列 U_1 が正則で，U_1，U_1^{-1} がともに整数行列であるとき，U_1 をユニモジュラー行列という。

（定義２）実行列 U_2 において，どの正方部分行列の行列式も１，-1または０のいずれかであるとき，U_2 をユニモジュラー行列という。

[アダマール行列]

（定義１）正方行列 $H_1 = [h_{ij}^{(1)}]$ において

$$|h_{jj}^{(1)}| > \sum_{i \neq j} |h_{ij}^{(1)}| \quad \text{（すべての j に対して）}$$

が成立するとき H_1 をアダマール行列という。

（定義２）$n \times n$ 正方行列 H_2 があって，その要素は $+1$ か -1 であって

$$H_2'H_2 = H_2H_2' = nI$$

であるならば，H_2 をアダマール行列という。

上記ユニモジュラー行列の２つの定義は関連はありそうだが，定義としては異なるものである。たとえば，（定義１）を満足する行列として

$$U_1 = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \quad U_1^{-1} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

をあげることができるが，（定義２）によれば行列 U_2 の要素は１，-1 または０でなければならないから，この例の行列は（定義２）を満足しない。（定義１）と（定義２）の根本的な違いは，（定義１）の行列は正方行列であるが，（定義２）の行列は正方行列とはかぎらぬことである。

アダマール行列（定義１）と（定義２）の間にはほとんど関連はない。（定義１）は狭義優対角行列とよばれることがある。（定義２）のアダマール行列は実験計画法などで現れるものであり，別の表現で定義されることもある。この２つの行列は応用上よく用いられるものであるが，使用される分野が一般にことなっているので，混乱はほとんどない。しかし，議論する場合やシソーラス等の作成においては形式的に整理せずに，定義を確認する必要がある。

その他にも，非負行列と非負定値行列，正則行列と正規行列など混同されている例がある。

(4)　関連語

関連の定め方によって，いろいろな場合があるが，ここでは，２つの概念の重なりの度合の大きいものを関連が深いと考えることにしよう。ただし，関連の尺度は問題によって適当に選ぶものとする。たとえば，表示行列［４

］によっても測ることが可能であるが，集合の面積のようなものを定義できる場合には，それによってもよい。

(5)　反意語

1つの定め方として，その集合の補集合に対応する概念を反意語とすることが考えられる。たとえば，正方行列の中では，正則行列と特異行列は互いに反意語であると考えることができる。正則行列は行列式が非零であり，特異行列の行列式は0であるので，互に排他的である。

しかし，3角行列における上3角行列と下3角行列は，上記のように考えるとき，反意語とはいえない。なぜなら，通常の定義では上3角行列と下3角行列は排他的ではなく，共通部分が存在し，それは対角行列に対応しているからである。

このように，形式的には反意語であるようにみえても，そうでない場合がある。このような問題点が明確にできることは，具体例によって議論をおこなっている場合の有利な点であろう。

(6)　上位概念と下位概念

ある概念の指示する集合が他の概念の指示する集合に包含されるとき，包含する方を上位概念とよび，他を下位概念という。たとえば，正規行列とユニタリ行列について考えると，正規行列は，ユニタリ行列の上位概念であり，ユニタリ行列は正規行列の下位概念である。

概念間の包含関係は文献検索においても重要な役割を演じる。たとえば，上の場合で，ユニタリ行列に関する数値計算法を検索するとき，もしそれに関する文献が存在しなくても，正規行列に関する数値計算法があれば，一般にはそれで間に合わせることができる場合もある。とくに，定理の検索を考える場合には，このことはもっと明確になるし，また他の情報検索の場合にも同様の事情に出会う。

(7)　概念の合成

いくつかの概念を組み合わせることにより，新しい概念を合成していくことができる。たとえば，エルミット行列と冪等行列の共通部を射影行列とい

う。すなわち

射影行列 ≡ エルミット行列 ∩ 冪等行列

である。概念を合成するさいの基本となる概念は基底概念とよばれることがある［2］。

概念の合成は新しい概念を生成していく点からも興味があり，また検索の過程においても重要である。なぜなら，検索の要求者が目的とする概念の名称を知らない場合，既知の概念を組み合わせて，検索することが可能となるからである。もし，上記の例であれば，射影行列という名称を知らなくても，エルミット行列と冪等行列を知っていれば検索できるからである。概念の検索は，文献検索の前の段階として，無視できないところである。

なお，このような形で新しく概念を合成した場合，たとえば概念 C_3 を，2つの概念 C_1，C_2 によって

$$C_3 \equiv C_1 \cap C_2$$

で合成したとき，つぎの包含関係が自動的に成立することに注意すべきである。

$$C_1 \supseteqq C_3,\ C_2 \supseteqq C_3$$

同様に，$C_4 \equiv C_1 \cup C_2$ であれば

$$C_4 \supseteqq C_1,\ C_4 \supseteqq C_2$$

となる。

概念の合成の問題は概念間の演算と密接な関連があるので，本節(9)の概念算の拡張のところでも，これについてふれる。

(8)　fuzzy 集合

fuzzy 集合［8］は境界の明確でない集合である。このような集合が，一般の場合より比較的厳格であるとおもわれる行列の名称に関してもあらわれる。たとえば，零要素の多い行列であるスパース行列，また行列式の絶対値が十分小さい行列すなわち特異行列に近い行列などがそれである。このような定義はあいまいまたは不正確であって，集合が明確ではない。それにもかかわらず，これらのあいまいな行列は数値計算等においてしばしばあらわれ

る。

もちろん，定義を明確にすれば，集合が確定する。仮に，零要素が半数以上であればスパース行列という，というように定めれば確定する。しかし，実用上はあいまいのままでもほとんど問題はないようである。

(9)　概念算の拡張

新しく概念を合成していくさい，すでに(7)で述べた通常の概念算すなわち集合算だけによっていては，新しく合成できる概念はかぎられてしまう。たとえば，1つの概念 C から概念算によって生成できる概念はせいぜいつぎの4個である。

$$\mathrm{C},\ \mathrm{C}^c,\ \phi,\ \Omega$$

ただし，C^c は C の補集合，ϕ は空集合，Ω は全体集合である。また，2つの概念 C_1, C_2 から合成できる概念はつぎの4個の基本積

$$\mathrm{C}_1^c \cap \mathrm{C}_2^c,\ \mathrm{C}_1^c \cap \mathrm{C}_2,\ \mathrm{C}_1 \cap \mathrm{C}_2^c,\ \mathrm{C}_1 \cap \mathrm{C}_2$$

を組み合わせて得られる　$2^4=16$ 個である。概念 C_1 と C_2 の相互の関係によって，合成される相ことなる概念は実質的には16個よりすくなくなることもある。

たとえば，C_1, C_2 に関して，もし $\mathrm{C}_1 \supseteqq \mathrm{C}_2$ であれば，

$$\mathrm{C}_1 \cap \mathrm{C}_2 = \mathrm{C}_2$$

であるから，共通集合 $\mathrm{C}_1 \cap \mathrm{C}_2$ によって新しい部分集合がつくられるわけではない。また，このとき

$$\mathrm{C}_1^c \cap \mathrm{C}_2 = \phi$$

であって，$\mathrm{C}_1^c \cap \mathrm{C}_2$ に属する要素は存在しない。

現実の自然語の場合には，2つの単語 W_1 と W_2 を，そのまままたは連結詞を介して連結して得られる新しい概念

$$\mathrm{W}_1\mathrm{W}_2$$

について考えてみると，この結果が W_1 と W_2 の集合算だけで表現できるとはとてもいいきれない。むしろ，$\mathrm{W}_1\mathrm{W}_2$ が集合算では表現できないような集合を指定している場合も多いと考えられる。しかしながら，一般的に議論し

ているかぎり集合算以外の演算を導入することは困難であるし，たとえ導入したとしてもどれほどの意義があるか疑問である。ところが，いまは議論の基礎として行列の集合を考えているので，集合算以外の演算の導入が可能であるし，その演算の具体的意味も明確になり，また自然語研究のモデルとしても意義がある。

種々の演算を考えるために，まず記法を定めよう。一般に行列をAで示し，行列の集合を $\mathbf{A}$ で示す。

$$\mathbf{A} = \{A\}$$

また，行列の集合に対する単項演算 $f(\mathbf{A})$ を，行列に対する演算結果 $f(A)$ の集合として定める。

$$f(\mathbf{A}) = \{f(A)\}$$

行列の場合には，たとえばつぎのような単項演算を定めることができる。

A'（転置），A^{-1}（逆行列）

$\bar{A}$（共役），A^*（共役転置）

$\hat{A}$（非対角要素の符号を変える）

さらに，行列からスカラーへの演算としてつぎの演算がある。

det(A)（行列式）

rank(A)（階数）

tr(A)（跡）

2項演算として，以下のようなものを導入できる。

$$\mathbf{A} + \mathbf{B} = \{A + B\} \quad (\text{和})$$

$$\mathbf{A} \circ \mathbf{B} = \{A \circ B\} \quad (\text{積})$$

$$\mathbf{A} \dotplus \mathbf{B} = \{A \dotplus B\} \quad (\text{直和})$$

ただし，$\mathbf{A} + \mathbf{B}$ においては A とBは同じ大きさであるとし，$\mathbf{A} \circ \mathbf{B}$ においてはAの列数とBの行数は等しいものとする。AまたはBがスカラー（1次行列）のときも，この規約に従うものとする。$\mathbf{A} \dotplus \mathbf{B}$ のときはAとBは正方行

列であればよく，次数はことなっていてよい。さらに，行列の集合の間で演算をおこなう場合は，上記のような制約のもとで，すべての可能な行列の間で演算をおこなうものとする。

このような演算を用いて新しい概念を定義していくことができる。

〔例〕

（1）**A**：正則行列

このとき$\mathbf{A}^c$：特異行列

ただし，正方行列の範囲で考える。

（2）**A**：上3角行列

このとき$\mathbf{A}'$：下3角行列

（3）**A**：ユニタリ行列，**B**：上3角行列

$\mathbf{A}\circ\mathbf{B}$（正則行列を含む）

⑽　関係による概念の定義

与えられた集合の中で，一定の条件式を満たす要素すなわち一定の性質を有する要素の集合として，新しい集合を定義することができる。ここでは単項演算と2項関係によって構成される条件式について考えよう。

その例として，まず逆行列を定義すなわち条件式の中に含むものをいくつか示す。下記の例において，i は虚数単位である。

$\{A \mid A^{-1}=A\}$　（対合行列）

（例）　$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

$\{A \mid A^{-1}=A'\}$（複素直交行列）

（例）　$\begin{bmatrix} i & \sqrt{2} \\ -\sqrt{2} & i \end{bmatrix}$

$\{A \mid A^{-1}=A^*\}$　（ユニタリ行列）

（例）　$\begin{bmatrix} 0 & i \\ 1 & 0 \end{bmatrix}$

$\{A \mid A^{-1}=\bar{A}\}$

（例）　$\begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}$

$\{A \mid A^{-1} = \hat{A}\}$

（例）　$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$

これらは正則行列の中の特殊な行列であって，それらの逆行列はそれぞれの定義中の単項演算によって容易に求められる。$\hat{A}$ はすでに述べたように，A の非対角要素の符号を変えたものである。したがって，行列

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

の逆行列は

$$A^{-1} = \hat{A} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

となる。

単項演算と２項関係による概念の定義としてはこの他にもつぎのようなものもある。

$\{A \mid A = \bar{A}\}$　（実行列）

$\{A \mid A = -A'\}$（反対称行列）

$\{A \mid A = -A^*\}$（反エルミット行列）

$\{A \mid A = A^*\}$　（エルミット行列）

$\{A \mid A > O\}$　（正行列）

$\{A \mid A \geq O\}$

$\{A \mid A \geqq O\}$　（非負行列）

$\{A \mid A \geqq I\}$

$\{A \mid A \leqq E\}$

ただし，O は零行列，I は単位行列，E はすべての要素が 1 の行列である。

また$A \geq O$はAの全要素が非負で，すくなくとも1つの要素は正であることを示す。

さらに行列の持つ特定の値によって，行列の集合を定めることができる。

$\{A \mid \det(A)=0\}$　(特異行列)

$\{A \mid \det(A)=1\}$

$\{A \mid \det(A)=-1\}$

$\{A \mid \mathrm{tr}(A)=0\}$

演算，関係としては上記の他にも種々のものがある。これらを組み合わせていくと，多数の新しい概念が定義でき，また，これらの概念間の包含関係として，行列に関する定理が表現できる。これに関してはつぎの3節で考察する。

(11)　概念空間

概念について議論をしているとき，概念の定義域すなわち概念空間を明確にしておく必要がある。換言すれば，基礎となる全体集合をはっきりさせ，議論はすべてその上でおこない，演算結果がその集合に含まれない演算は考えないようにする。とくに，補集合の場合は，はじめの全体集合を与えないことには定義できない。

これまで行列の集合を，とくに正方行列の集合を例にとって考察をおこなってきた。そのとき，暗黙のうちに要素は有限の複素数であり，実行列はその特別の場合とし，また次数も有限であるとしてきた。したがって，正方行列に対してはつねに行列式が計算でき，その行列式の値によって正則行列と特異行列の2つに分割することができた。しかし，もし行列の要素として無限大（∞）を認めると事情がことなってくる。たとえば，∞を要素にもつ行列に関しては，行列式が一般には定められないであろう。

（例）

$$\begin{pmatrix} 1 & \infty \\ 1 & \infty \end{pmatrix}$$

この例のような行列に対しては行列式が定められない。したがって，正則

行列，特異行列に加えて第3の行列が出現する。

同様のことは，行列の次数として無限のものを認めたときにも生ずる。すなわち，有限次の行列で成立していた性質が無限次の行列で成立するとはかぎらぬからである。しかし，このような行列はそのような困難がありながら，実用上はきわめて重要であり，∞を要素としてもつ行列はネットワーク理論において，また無限次の行列は量子力学において用いられている。

したがって，議論の土台となる全体集合を明確にしておく必要がある。そうでなければ，概念の包含関係，存在性の問題は議論できない。以下では，一応有限の複素数を要素とする有限次の正方行列を考えていく。

3．情報空間

従来，情報に関しては種々の定義が与えられている。しかし，情報検索の情報が何かという点については，まだ十分に明確にされているとはいえない。実用的観点からは文献そのものを情報という場合もある。しかし，ここでは以下のような情報の基本的構成要素を考え，その要素の全体集合を情報空間とよぶことにする。

情報空間のモデルを構成するために，再び行列論を例にとって考えよう。行列論では，つぎのような事実がある。

正方行列Aに関して，$A^p = O$ なる自然数pが存在すれば，行列Aは正則ではない。

すなわち

$$\text{冪零行列} \Rightarrow \text{特異行列}$$

である。冪零行列の集合を $\mathbf{C}_1$，特異行列の集合を $\mathbf{C}_2$ とすれば

$$\mathbf{C}_1 \subsetneqq \mathbf{C}_2$$

となる。

ところが，容易にわかるように $\mathbf{C}_1$，$\mathbf{C}_2$ に関しては

$$\mathbf{C}_1 \subsetneqq \mathbf{C}_2$$

である。すなわち，冪零行列でない特異行列が存在する。したがって，$\mathbf{C}_1 \subseteqq \mathbf{C}_2$ よりも $\mathbf{C}_1 \subsetneqq \mathbf{C}_2$ の方がより詳しい事実を示している。いいかえれば，$\mathbf{C}_1 \subsetneqq \mathbf{C}_2$ の方が情報が多いと見ることができる。

このことをより明確に知るために，$\mathbf{C}_1 \subseteqq \mathbf{C}_2$ および $\mathbf{C}_1 \subsetneqq \mathbf{C}_2$ がつぎのように解釈できることに注目する。

$$\mathbf{C}_1 \subseteqq \mathbf{C}_2 \Leftrightarrow \mathbf{C}_1 \cap \mathbf{C}_2^c = \phi$$

$$\mathbf{C}_1 \subsetneqq \mathbf{C}_2 \Leftrightarrow \mathbf{C}_1 \cap \mathbf{C}_2^c = \phi,\ \mathbf{C}_1^c \cap \mathbf{C}_2 \neq \phi$$

すなわち，包含関係で与えられる情報は，基本積の存在性に関する情報であることがわかる。$\mathbf{C}_1$ と $\mathbf{C}_2$ から構成される基本積はつぎの4個である。

$$\mathbf{C}_1^c \cap \mathbf{C}_2^c,\ \mathbf{C}_1^c \cap \mathbf{C}_2,\ \mathbf{C}_1 \cap \mathbf{C}_2^c,\ \mathbf{C}_1 \cap \mathbf{C}_2$$

このうち，$\mathbf{C}_1 \subsetneqq \mathbf{C}_2$ のときは2個について，存在性が与えられていることになる。ある基本積が空でないことを示すには，それを満足する具体的例を示せば十分である。たとえば，冪零行列でない特異行列としてはつぎの例がある。

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

また，2つの概念 $\mathbf{C}_1$，$\mathbf{C}_2$ が集合として等しいということは，基本積に関する情報としては，つぎのようになる。

$$\mathbf{C}_1 = \mathbf{C}_2 \Leftrightarrow \mathbf{C}_1 \cap \mathbf{C}_2^c = \phi,\ \mathbf{C}_1^c \cap \mathbf{C}_2 = \phi$$

したがって，$\mathbf{C}_1 = \mathbf{C}_2$ という関係は2つの基本積に関する存在性の情報を与えている。以下ではこのような情報について考える。

いま，n個の概念 $\mathbf{C}_1, \mathbf{C}_2, \cdots, \mathbf{C}_n$ があり，それらを組み合わせてつくられる 2^n 個の基本積で示される領域の存在性が情報として与えられるものとする。これ以外の情報は考えないことにする。このとき，基本積の集合が情報空間に相当する。

また，文献は基本積に関する存在性の情報のみを有するものと考え，文献をつぎのように記述する。すなわち，文献の有する基本積の和集合をとり，それを簡約化した形で，記述するものとする。同様に検索要求も指定すべき

基本積の和で表わすことにする。

たとえば，３個の概念があって，ある文献がつぎの情報を有しているものとする。　$C_1 \cap C_2 \cap C_3^c \neq \phi$

$C_1 \cap C_2^c \cap C_3^c = \phi$

このとき，文献の記述はつぎのようになる。

$(C_1 \cap C_2 \cap C_3^c) \cup (C_1 \cap C_2^c \cap C_3{}^c) = C_1 \cap C_3^c$

すなわち，$C_1 \cap C_3^c$で与えられる。

また，つぎの基本積の存在性に関する情報をもつ文献を検索する場合について考えよう。

$C_1 \cap C_2 \cap C_3^c$

$C_1 \cap C_2 \cap C_3$

このときの検索要求はつぎのように記述される。

$(C_1 \cap C_2 \cap C_3^c) \cup (C_1 \cap C_2 \cap C_3) = C_1 \cap C_2$

このように，文献と検索要求を記述したとき，つぎに問題となるのは両者の記述の間のマッチングである。このマッチングの１つの方法は，以下のように定めることであろう。

上記の例の文献D＝$C_1 \cap C_3^c$，検索要求Q＝$C_1 \cap C_2$について考えよう。まず，両者に含まれる概念をもとにしてすべての基本積をつくる。この場合，概念がC_1，C_2，C_3であるから，$2^3 = 8$個の基本積が構成できる。つぎに，文献および検索要求に含まれる基本積の集合を｛D｝，｛Q｝で示せば

$\{D\} = \{C_1 \cap C_2^c \cap C_3^c,\ C_1 \cap C_2 \cap C_3^c\}$

$\{Q\} = \{C_1 \cap C_2 \cap C_3^c,\ C_1 \cap C_2 \cap C_3\}$

となる。この２つの集合を比較することにより，文献Dは検索要求の指定する２個の基本積のうち，１個について，その存在性に関する情報を有していることがわかる。したがって，文献Dと検索要求Qのマッチングの度合は½とするのが適当であろう。なお，｛D｝，｛Q｝の要素としての基本積は，記号列として扱い，それの指示する集合ではないとする。たとえば，上記の｛D｝の要素$C_1 \cap C_2^c \cap C_3^c$は'$C_1 \cap C_2^c \cap C_3^c$'とでも書くべきかもしれな

い。しかし，わずらわしいので表記を区別しない。D，Qの記述についても同様である。

一般には，文献Dと検索要求Qとのマッチングの度合r(D, Q)は次式で与えられるであろう。

$$r(D,Q)=\begin{cases}\dfrac{\mu(\{D\}\cap\{Q\})}{\mu(\{Q\})}\cdots\mu(\{Q\})\neq 0\text{のとき}\\ 1\cdots\cdots\cdots\cdots\cdots\cdots\mu(\{Q\})=0\text{のとき}\end{cases}$$

ただし，ある有限集合Sの要素数を μ(S) で示す。

このr(D, Q)は，検索要求の指定する情報を，文献が完全に有しているとき1となり，そうでなければ1より小さくなる。通常は，このr(D, Q)の大きい文献からとり出せばよい。しかし，問題になるのは要求された情報がいくつかの文献に分散しているときである。このとき文献を組み合わせて出力する必要がある。この点については，従来ほとんど議論されていないようにおもわれる。

文献を組み合わせることは，いまは基本積の存在性に関する情報だけを考えているので，比較的容易に実行できる。

4．概念の定義における等価性

すでに，2節の⑽で述べたように，与えられた全体集合の中で，一定の条件を満足する要素の集合として，概念を定義していくことができる。たとえば，エルミット行列は，正方行列の中でつぎのように定められる。

$$\mathbf{C}=\{A\mid A=A^*\}$$

概念をこのように定義したときに，まず問題となるのは，2つの与えられ

た概念が等しいかどうかということである。しかし，ここで考えている概念は，それぞれ集合をともなっているので，2つの集合を比較してみればよい。

このように考えれば，すでに述べたように通常意味とよばれているものの一部は，その基礎になっている集合で説明できるであろう。すなわち，2つの形式的表現があった場合，その2つの意味が等しいかどうかの判定を，それぞれによって定義される集合を見ておこなうことができよう。

しかしながら，これにも問題がないわけではない。それは，2つの概念の定義があまりにも異なっているとき，それによって与えられる2つの集合が等しいからといって，定義が等価であるといってよいかという問題である。

たとえば，正方行列の範囲で，つぎのような定義が考えられる。

$\mathbf{C}_1$：{行列式が0でない行列}

$\mathbf{C}_2$：{固有値に0をもたない行列}

$\mathbf{C}_1$と$\mathbf{C}_2$は，集合としては等しいが，これらの集合が等しいことを知るには，若干の知識が必要であろう。

これに対して，つぎのような場合がある。

$$\{A \mid A + A^* = O\}$$

$$\{A \mid A^* = -A\}$$

$$\{A \mid A = -A^*\}$$

これらが，反エルミット行列の定義として等価であることは，ほとんど自明である。

したがって，2つの定義が合同であるかどうかを判定する規準を設定する必要がある。すなわち，簡単に等価であることが決定できる定義は合同であるとし，そうでないものは合同でないとしなければならないであろう。

これと同様の事情は，一般の定理の検索［9］においても生ずる。すなわち，ある定理から別の定理を導くとき，導出に制限をつけなければ，原理的には，どんな定理でも任意の定理から導出可能となってしまう。

概念の定義において，同じ内容の定義を種々の形で表現できる場合がしばしばあり，表現の一意性の問題は実際の場合にはやっかいである。表現が異

なる定義を同一の定義として扱うか，また別々の定義として扱うかの問題を処理しなければならない。

5．まとめ

情報検索における概念の問題に関して，行列論を題材として，考察をおこなった。具体的例をとりあげたのは，議論および研究を着実なものとするためである。

概念に関するいくつかの話題，とくに包含関係，概念の定義，表現の問題について議論をおこない，若干の着想をのべた。また，従来あいまいであった情報についても，簡単ではあるが，1つの手がかりを与えた。もちろん，すべての情報が基本積のような形で与えられるわけではないので，さらに議論すべき余地はある。

情報空間を用いる検索モデル，包含関係で与えられる定理の情報量，および情報が散在しているとき文献を組み合わせて出力する場合の具体的手順等については，別の機会に報告したい。

なお，ここでの議論を定理の検索に適用することを計画しており，そのとき問題となる概念の合成と検索に関して，考察を進めている。

謝辞

本研究の一部は，名古屋大学福村晃夫教授の助言に負うものである。

文　　献

〔1〕中村：“情報処理Ⅰ—言語と概念—”，共立出版，昭和43年10月.

〔2〕河口：“一つの単語概念空間モデル”，昭和43年度電子通信学会全国大会，講演論文集分冊1，48.

〔3〕栗原・吉田・鶴丸："概念構成に関する一考察", 九大工学集報, 第43巻第3号, p.338-p.343 (昭和45年6月).
〔4〕橋本・福村："二項関係による概念の分割", 電子通信学会オートマトンと言語研究会資料 AL 72-135 (1973. 3).
〔5〕Salton, G.: "*Automatic Information Organization and Retrieval*", McGraw-Hill, 1968.
〔6〕Meadow, C. T.: "*The Analysis of Information Systems*", John Wiley & Sons, Inc., 1967.
〔7〕"シソーラス入門", 日本ドクメンテーション協会, 1970.
〔8〕Zadeh, L. A.: "Fuzzy Sets", *Information and Control* 8, 338-353 (1965).
〔9〕細井："論理的にみた定理検索の可能性", 数理科学 No. 129, MARCH 1974, p.66-p.70.