

第 4 章

意味の数学モデルのマルチデータベースシステムへの 応用

4.1 マルチデータベースシステムとは

マルチデータベース環境とは、様々なデータベースシステムを利用できる環境のことである。しかし、各々のデータベースシステムは、別々の環境で作成され、ポリシーも異なる。この様な環境では、利用者は、各データベースの操作方法を個別に習得せねばならず、また、探したいデータのキーワードもデータベースごとに異なるので、様々なキーワードを試して、各々のデータベースからデータを集めなくてはならない。

このような状況を打破するためには、

- (1) 全てのデータベースを統一した環境に再構築する。
- (2) 利用者にデータベースが統一されているようにみせる。

の方法が考えられる。

しかし、(1)の方法は何十年もかけて構築された、世界中のデータベースを再構築するには、かなりの手間がかかる一方、データベース群を、どのように再構築するかを決定せねばならない。(2)の方法は、今までのデータベースシステムには一切手を加えず、システム内でデータベースシステムごとの違いを吸収し、あたかも、一つのデータベースに見せる方法である。この場合、世界中のデータベースに手を加えることなく、利用者に、データベース群が統合された環境を提供することが可能である。

4.2 意味の数学モデルのマルチデータベースシステムへの応用

マルチデータベース環境で意味による検索を行うことは、重要な課題となっている [2, 29, 36, 35]. 本意味的連想検索方式を、複数のデータベース・システムの支援システムに適用することにより、マルチデータベース・システムの検索系の主要な機能を実現することが可能になる。本意味的連想検索方式を適用したマルチデータベース・システムでは、利用者の問い合わせ中のキーワードを、各々のデータベース中に存在する、意味の近いキーワードに変換することができる。通常のデータベース・システムでは、利用者が発行した検索キーワードと同じ表記のデータが存在しなければ、問い合わせは失敗してしまう (図 4.1)。しかし、本方式を適用し、利用者の問い合わせ中のキーワードを、データベース中に存在する意味の近いキーワードに変換することにより、表記を意識すること無く、意味の近いデータを引き出すことが可能になる。つまり、利用者は、自分の求めるデータが、各々のデータベース中で、どのような表現で存在しているかを意識せずに、意味的連想検索を行うことが可能になる (図 4.2)。この方式により、問い合わせ中のキーワードの変換を自動的に行うことができるので、マルチデータベース・システムが、異なるデータベース群の各々に問い合わせを発行することにより、利用者は、それらのデータベース中において、検索対象データがどのような表記で存在するかを意識することなく、情報を引き出すことが可能になる。

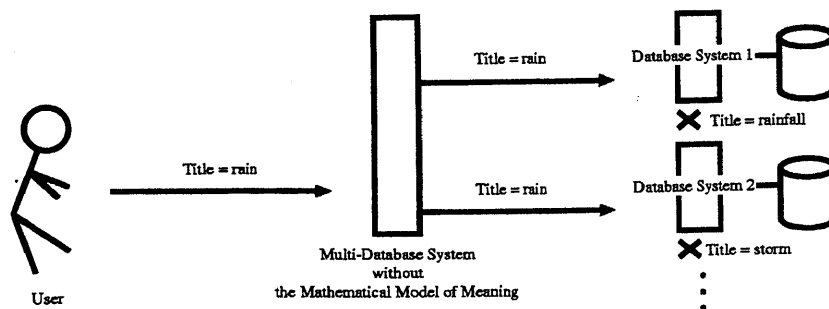


図 4.1: 単純なマルチデータベース環境

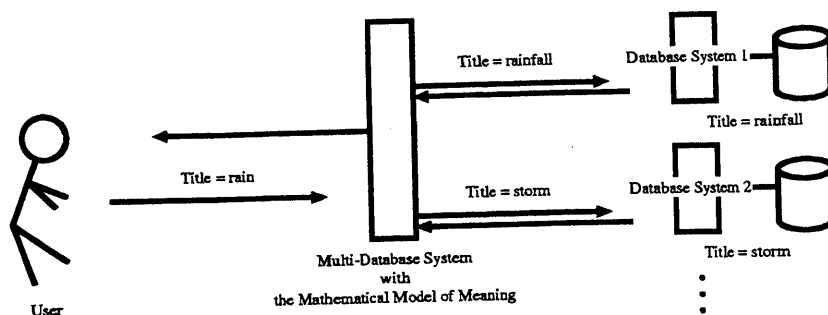


図 4.2: 意味の数学モデルのマルチデータベースシステムへの適用

4.3 実験

4.3.1 実験環境

本実験では、対象データベース・システムとして、筑波大学内にある UTOPIA データベース群を利用した [41]。UTOPIA データベース・システムは、筑波大学情報処理センター内に存在する様々なデータベースを利用するためのシステムである。扱っているデータベースは、主に図書についてのもので、多くの収録件数がある。

マルチデータベース・システムへの適用時には、問い合わせ中の属性名に関する変換への適用と、問い合わせ中の属性値に関する変換への適用が考えられるが、本実験では、後者への適用を行った。

また、本実験では、Longman Dictionary of Contemporary English [22] において基本語とされている英単語 2328 種類を、The General Basic English Dictionary [24] を用いて定義し、 2328×874 の行列を作成し、イメージ空間を構成した。

実験に使用した計算機は Sun4/ELC。OS は SunOS 4.1.4 である。

4.3.2 実験方法

属性値の実験では、UTOPIA 内から利用できる ENV と POLL (以下 POL とする) という 2 つのデータベースを対象とした。ENV は環境破壊についてのデータベースであり、POL は環境汚染についてのデータベースである。この 2 つのデータベースのうち、検索は、本のタイトル中に“TSUKUBA”または“KASUMIGAURA”を含むもの、という条件で予め検索をかけ、その結果を実験対象とした。その結果、ENV データベースからは 45 の属性値 (表 4.1)、POL データベースからは 49 の属性値 (表 4.2) を得た。

表 4.1: ENV データベースの属性値と定義

academic	[a.] [sp.] , to do with teaching
area	[n.] stretch of land [etc.]
basin	[n.] wide , round , open vessel used for liquids , [etc.]
bedrock	[n.] hard stone under looser material
bloom	[n.] flower
building	[n.] house or house-like structure
case	[n.] example of a thing's *taking *place
chemical	[a.] of or made by chemistry
chemistry	[n.] science having to do with substances , their behaviour and effects
composition	[n.] COMPOSING , [sp.] art of writing or writing music
concentrations	[n.] [sp.] , power of concentrating attention
concept	[n.] idea of a quality common to group of things
construction	[n.] [sp.] , thing made
crystalline	[a.] made of crystal
discharge	[v.t.i.] send away from -work , -position
dissolved	[v.t.i.] (of liquid) take (solid) completely into itself
distributions	[n.] distribution give out
drainage	[n.] draining
dynamic	[a.] of physical power , forces producing motion
dynamics	[n.] branch of physics to do with acting of forces
effect	[n.] reaction , outcome , that produced by a cause
element	[n.] a simple part of something complex
events	[n.] thing , [sp.] important thing , *taking *place
influence	[n.] acting of something in unseen ways
lake	[n.] great stretch of inland water
model	[n.] copy of a thing made to scale
new	[a.] first made , designed , seen , experienced , used , at the time in question , not -long made , used , in existence
oxygen	[n.] gas without -colour , -taste , or -smell , present in air and necessary to existence of animals and plants
particles	[n.] very small bit
quality	[n.] (high) degree in which a thing is good
rainfall	[n.] amount of rain falling in some place in given time , measured in inches
resources	[n.] thing , way of acting , which gives , is turned to for , help when in need , trouble
seasonal	[a.] dependent on or changing with the seasons
selected	[v.t.] take (some things) from a number , [sp.] as being the best , most desired , right for some purpose
size	[n.] size1 measure of a thing
streams	[n.] natural body of water moving in one direction , river , current
studies	[v.t.i.] give time and thought to getting knowledge of , [sp.] , from books
succession	[n.] a coming one after another in time or order
temporal	[a.] of , in , time
town	[n.] place where there are a great number of buildings grouped in streets [opp.] -country
trace	[n.] mark , sign , of something as having been present , *taken *place , in the past
tunnel	[v.t.i.n.] (make) way through the earth , [sp.] , for road , train [etc.]
urban	[a.] of , living in , a town
variations	[n.] [sp.] , (degree of) being variant
water	[n.] liquid without -colour , -smell , or -taste , formed of HYDROGEN and OXYGEN , changed by heat into steam and by cold into ice , [sp.] , as forming rain , sea , rivers [etc.]

表 4.2: POL データベースの属性値と定義

accessory	[n.a.] thing going with , helping , another thing , but not a -necessary part of it
alga	[n.] water plant of very simple structure
application	[n.] [sp.] something applied
aquatic	[a.] (of plants , animals) living in , near , water
atmospheric	[a.] body of gases about the earth or another star
bay	[n.] wide-mouthed inlet of sea [etc.]
behaviour	[n.] way of behaving
bloom	[n.] flower
bog	[n.] wet sponge-like land
carbon	[n.] a chemically simple substance , not a -metal , forming chief part of coal
case	[n.] example of a thing's *taking *place
chemical	[a.] of or made by chemistry
city	[n.] great town
community	[n.] group of persons living in same place under same government , political or other
composition	[n.] COMPOSING , [sp.] art of writing or writing music
cost	[v.t.] have as price
damage	[n.] loss of -value , wound , undergone
debris	[n.] broken bits
different	[a.] not the -same
distributions	[n.] distribution give out
dynamics	[n.] branch of physics to do with acting of forces
effect	[n.] reaction , outcome , that produced by a cause
estimating	[v.t.] come to opinion as to (amount , value)
impact	[n.] coming together or with force
indicated	[v.t.] get pointed out , make clear
lake	[n.] great stretch of inland water
load	[n.] what is to be lifted , transported
losses	[n.] losing
model	[n.] copy of a thing made to scale
models	[n.] copy of a thing made to scale
musty	[a.] covered with MOULD , old
nutrient	[a.] acting as food
organic	[a.] of , to do with , organs of body
particles	[n.] very small bit
plant	[n.] living structure which is not an -animal , [freq.] with leaves , getting its food from air , water , earth
population	[n.] (number of) persons living in a place
rapid	[a.] very quick
ratio	[n.] relation between 2 amounts based on division of the one by the other
role	[n.] actor's part in play
seasonal	[a.] dependent on or changing with the seasons
secondary	[a.] coming after , dependent on , less -important than , what is first or chief
size	[n.] size measure of a thing
species	[n.] group of things having common qualities
storm	[n.] violent weather conditions with thunder , strong wind , much rain or snow [etc.]
succession	[n.] a coming one after another in time or order
treatment	[n.] way of treating person or thing
uptake	[n.] power to see point
variations	[n.] [sp.] , (degree of) being variant
welfare	[n.] well-being

検索の対象となる単語群は、The General Basic English Dictionary の定義を用いて定義した。また、利用者が使用可能なキーワードは、Longman Dictionary of Contemporary English において基本語とされている 2328 単語を、The General Basic English Dictionary の定義を用いて定義し、同じ見出し語を持つ単語群の定義を合成し、2115 単語を用意した。

実験では、利用者の問い合わせ中のキーワードを、各々のデータベース中に存在するキーワードに変換する時に、高速な意味的連想検索のアルゴリズムが、どの程度有効であったかを調べた。

4.3.3 実験結果

実験結果を表 4.3 に示した。実験結果 1 は、利用者が検索キーワードとして rain を与え、その時の文脈として weather を与えた時の結果である。ENV データベース中には rain という表記のデータは存在していないので、通常のデータベース検索では、この問い合わせは失敗してしまう。しかし、意味の数学モデルによって、ENV データベース中に存在するキーワードの中で、rain に最も意味の近いキーワードは rainfall であるという結果が抽出されている。そこで、rain というキーワードの代わりに rainfall というキーワードを用いて問い合わせを発行することにより、関連する情報を引き出すことが可能となる。POL データベースにおいても、同様のことを行うことができる。POL データベース中にも rain という表記のデータは存在していない、しかし、意味の数学モデルによって、POL データベース中に存在するキーワードの中で、rain に最も意味の近いキーワードは storm であるという結果が抽出されている。利用者は、rain というキーワードの代わりに storm というキーワードを使用して問い合わせを発行すれば、その関連情報を引き出すことができる。

本実験システムにおいて、変換にかかった距離計算回数に注目すると、ENV データベースでは 45 個の検索対象の属性値について、わずか 3 回の距離計算で、3 つの解が求まっている。POL データベースにおいても、49 個の検索対象属性値の内、6 回の距離計算で、3 つの解を求めることが可能であった。

実験結果 2 と 3 は、意味的な関係が文脈に応じて変化する例である。実験結果 2、実験結果 3 と同様に、キーワードは water で同一であるが、POL データベースでは、文脈が weather の時は storm が選択され、文脈が pool の時は lake が選択されている。ENV データベースには、water というデータが存在するので、それが選択されて

実験結果 1

keyword : rain
context : weather

ENV database				POL database			
No.	Selected Word	Distance	R	No.	Selected Word	Distance	R
1	rainfall	0.521618	3/45	1	storm	0.494476	1/49
2	streams	0.566394	0/45	2	atmospheric	0.695877	3/49
3	oxygen	0.595867	0/45	3	plant	0.761912	2/49

R = 距離計算を行った単語数 / 全検索対象単語数

実験結果 2

keyword : water
context : weather

ENV database				POL database			
No.	Selected Word	Distance	R	No.	Selected Word	Distance	R
1	water	0.733407	1/45	1	storm	0.974418	1/49
2	oxygen	1.218185	3/45	2	atmospheric	1.443099	3/49
3	rainfall	1.277171	0/45	3	plant	1.512229	1/49

R = 距離計算を行った単語数 / 全検索対象単語数

実験結果 3

keyword : water
context : pool

ENV database				POL database			
No.	Selected Word	Distance	R	No.	Selected Word	Distance	R
1	water	1.438020	45/45	1	lake	1.971830	49/49
2	streams	1.887675	0/45	2	aquatic	2.112745	0/49
3	lake	1.971830	0/45	3	alga	2.136467	0/49

R = 距離計算を行った単語数 / 全検索対象単語数

実験結果 4

keyword : fish
context : water

ENV database				POL database			
No.	Selected Word	Distance	R	No.	Selected Word	Distance	R
1	streams	0.296871	8/45	1	lake	0.320872	8/49
2	lake	0.320872	2/45	2	aquatic	0.333773	3/49
3	oxygen	0.338915	1/45	3	plant	0.335020	0/49

R = 距離計算を行った単語数 / 全検索対象単語数

表 4.3: 高速な意味的連想検索のマルチデータベースへの適用実験

いる。もし、ENV データベースを対象として、water 以外のキーワードを使用する場合には、3 番目の検索結果である rainfall や lake を使用する。実験結果 3 は、高速な意味的連想検索が有効に働かなかった場合である。この場合でも、距離計算回数が、本提案アルゴリズムを使用しない場合を超えることはない。

実験結果 4 は、利用者が与えたキーワードをについて、検索対象語の中から、最も意味の近い単語が選ばれている一般的な例である。この結果でも、高速な意味的連想検索のアルゴリズムが有効であることがわかる。