

【特許請求の範囲】**【請求項 1】**

所定のデータベースを参照する参照部と、

参照されたデータベースに記述された情報を n 次元閉セルとして扱い、いま注目しようとする属性の組からなる情報を i 次元閉セルとして扱い、前記 n 次元閉セルから前記 i 次元閉セルを抽出する抽出部と、

前記注目しようとする属性の組の決定手順、および前記 i 次元閉セルの抽出の手順をホモトピーとして記録する記録部と、

を備え、前記抽出部は、前記 i 次元閉セルを前記データベースが物理的に記憶されている領域とは別の記憶領域へ保存することを特徴とする情報管理装置。 10

【請求項 2】

前記抽出部は、前記 n 次元セルの境界を $(n - i)$ 回取ることにより前記 i 次元セルを抽出する請求項 1 に記載の装置。

【請求項 3】

前記記録部は、前記ホモトピーと対応づけて前記 i 次元セルを記録する請求項 1 に記載の装置。

【請求項 4】

新たに注目しようとする属性の組が決定されたとき、この組に対応する j 次元セルがすでに前記記録部に記録されているか否かを判定する判定部をさらに含む請求項 3 に記載の装置。 20

【請求項 5】

前記参照部は、前記所定のデータベースのほかに第 2 のデータベースを参照し、前記抽出部は、これらふたつのデータベースからそれぞれ i 次元セルを抽出し、

当該情報管理装置はさらに、抽出されたふたつの i 次元セル間で所期の処理を実行する演算部を備える請求項 1 に記載の装置。

【請求項 6】

前記記録部は、前記処理の結果を当該処理の手順とともに記録する請求項 5 に記載の装置。

【請求項 7】

第 1 および第 2 のデータベースを参照する参照部と、

前記第 1 および第 2 のデータベースにおいて注目しようとする属性の組を指定する指定部と、

前記第 1 および第 2 のデータベースに記述された情報をそれぞれ第 1 および第 2 の多次元閉セルとして扱い、前記注目しようとする属性の組からなる情報を目的の閉セルとして扱い、前記第 1 および第 2 の多次元閉セルからそれぞれ目的の閉セルを抽出する抽出部と、抽出された目的の閉セルどうしの間で所期の処理を実施する演算部と、

前記目的の閉セルの抽出手順および前記処理をホモトピーとして記録し、かつ前記処理が実施された前記目的の閉セルを前記ホモトピーと関連づけて記録する記録部と、

前記指定部にて、注目しようとする属性の組が新たに指定されたとき、この組に対応する目的の閉セルがすでに前記記録部に記録されているか否かを判定する判定部と、

を備え、目的の閉セルが前記記録部に記録されていないとき前記抽出部による抽出が行われ、目的の閉セルが前記記録部に記録されているときは前記抽出部による抽出はスキップされ、前記記録部から当該目的の閉セルが読み出されて利用されることを特徴とする情報管理装置。 40

【請求項 8】

前記抽出部は、前記第 1 および第 2 のデータベース自体の書換を行わない方法にて前記目的の閉セルを抽出し、

前記演算部は、前記処理が実施された目的の閉セルを前記第 1 および第 2 のデータベースが物理的に記憶されている領域とは別の記憶領域へ保存する請求項 7 に記載の装置。

【請求項 9】

所定のデータベースを参照する参照部と、
参照されたデータベースに記述された情報を付加空間モデルにおけるセルとして扱い、いま注目すべき情報を前記セルからセル分解操作によって抽出する抽出部と、
前記注目すべき情報の決定および抽出の手順をホモトピーとして記録し、かつ前記注目すべき情報をそのホモトピーに関連づけて記録する記録部と、
新たに注目すべき情報が指定されたとき、この情報が前記記録部にすでに記録されているか否かを判定する判定部と、
を備えることを特徴とする情報管理装置。

【請求項 10】

第1および第2のデータベースを参照する参照部と、
前記第1および第2のデータベースにおいて注目すべき情報を指定する指定部と、
前記第1および第2のデータベースに記述された情報をそれぞれ付加空間モデルにおける第1および第2のセルとして扱い、前記注目すべき情報を前記第1および第2のセルからそれぞれセル分解操作によって目的セルとして抽出する抽出部と、
抽出された目的セルどうしの間で所期の処理を実施する演算部と、
前記目的セルの抽出手順および前記処理をホモトピーとして記録し、かつ前記処理が実施された前記目的セルを前記ホモトピーと関連づけて記録する記録部と、
前記指定部にて、注目すべき情報が新たに指定されたとき、この注目すべき情報に対応する目的セルがすでに前記記録部に記録されているか否かを判定する判定部と、
を備え、目的セルが前記記録部に記録されていないとき前記抽出部による抽出が行われ、目的セルが前記記録部に記録されているときは前記抽出部による抽出はスキップされ、前記記録部から当該目的セルが読み出されて利用されることを特徴とする情報管理装置。 20

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は情報管理技術、とくに、データベースを利用する情報管理装置に関する。

【0002】

【従来の技術】

現代のグローバル化した社会において、ウェブ情報管理システムが非常に重要な役割を果たすようになった。現実に、この社会はウェブ上に構築されるサイバーワールドにドライブされている。サイバーワールドは、設計意図のあるなしに係わらずウェブ上に創造された情報の世界である。「e ファイナンス」と「e マニュファクチャリング」は、サイバーワールドの主要なプレイヤーであり、日々、GDP 同等の商取引がなされ、かつては産業の現場でなされた製造が、ウェブ上で部品を購入し、そのアセンブリのための工場をも購入することに置き換わっている。サイバーワールドのこうした複雑さと成長の速度は、必要に応じて即座にサーバーワールドを処理するウェブ情報管理システムのみを媒介として扱いが可能となる。 30

【0003】

ウェブ情報管理システムは、ウェブ上に提示される極めて多量かつ多様なデータから、目的のデータを的確、迅速に探索する能力をもたなければならない。そのために、最適なデータベースモデルをも併せて提示するものでなければならない。 40

【0004】

【発明が解決しようとする課題】

従来、データベースモデルとしては、リレーションナルデータベースモデルが広く利用されてきた。しかし、リレーションナルデータベースは、データの依存関係を認知している情報管理者の存在を前提とする「ワールドモデル」であり、データの関連性は、情報管理者によって予め固定的に規定される。個人または企業などの閉じた集団内でデータを管理する場合には、静的かつ固定的なデータの依存関係でも利用可能であるが、サイバーワールドでは、独立に目まぐるしく活動を続けるウェブサイトが、WWW を通して有機的にリンクしている。それらの急速で複雑な情報の変化は、明らかに、リレーションナルデータベース 50

の管理の限界を超えていた。

【0005】

また、リレーションナルデータベースは閉じた空間しか表現できないため、データベースごとにデータ管理の方法に違いができる結果、異なるデータベース間でデータ管理、データ操作が不可能になる。リレーションナルデータベースモデルでは、状況変化が起こるたびにデータベースの再正規化という再設計を行わねばならず、変化の激しいサイバーワールドのモデル化にはまったく不向きになってしまった。

本発明はこうした認識からなされてものであり、その目的は、サイバーワールドのような、オープンで変化が大きな情報世界をも的確に扱うことのできる情報管理技術を提供することにある。

10

【0006】

【課題を解決するための手段】

本発明の情報管理装置は、所定のデータベースを参照する参照部と、参照されたデータベースに記述された情報を n 次元閉セルとして扱い、いま注目しようとする属性の組からなる情報を i 次元閉セルとして扱い、 n 次元閉セルから i 次元閉セルを抽出する抽出部と、注目しようとする属性の組の決定手順、および i 次元閉セルの抽出の手順をホモトピーとして記録する記録部とを備える。抽出部は、 i 次元閉セルを前記のデータベースが物理的に記憶されている領域とは別の記憶領域へ保存する。

【0007】

この装置によれば、 i 次元閉セルが最初のデータベースとは別に記録されるので、必要に応じて再利用できる。記録部が前記のホモトピーと対応づけて i 次元セルを記録していれば、再利用の確実性が担保される。そのため、いちいちもとのデータベースへ戻る回数が減り、操作量、制御量が減る。

20

【0008】

抽出部は、前記の n 次元セルの境界を $(n - i)$ 回取ることにより i 次元セルを抽出してもよい。これは付加空間モデルにおけるセル分解のプロセスと考えてもよい。 $(n - i)$ 回境界を取る動作は、一時になされてもよい。

【0009】

この装置は、新たに注目しようとする属性の組が決定されたとき、この組に対応する j 次元セルがすでに記録部に記録されているか否かを判定する判定部をさらに含んでもよい。

30

【0010】

参照部は、前記のデータベースのほかに第2のデータベースを参照し、抽出部は、これらふたつのデータベースからそれぞれ i 次元セルを抽出し、この装置はさらに、抽出されたふたつの i 次元セル間で所期の処理を実行する演算部を備えてもよい。演算の例は、リレーションナルデータベースモデルにおけるジョインのほか、N O P (ノーオペレーション) を含む任意の論理演算、算術演算であってよい。この様によれば、ふたつのデータベースの構造が違っても、それらのデータベースから必要な情報を生成して利用できる。リレーションナルデータベースモデルでは、構造が異なるデータベース間でデータ操作ができないが、付加空間モデルに基づく本装置ではこれが可能になる。

40

【0011】

本発明の別の様も情報管理装置に関する。この装置は、第1および第2のデータベースを参照する参照部と、第1および第2のデータベースにおいて注目しようとする属性の組を指定する指定部と、第1および第2のデータベースに記述された情報をそれぞれ第1および第2の多次元閉セルとして扱い、注目しようとする属性の組からなる情報を目的の閉セルとして扱い、第1および第2の多次元閉セルからそれぞれ目的の閉セルを抽出する抽出部と、抽出された目的の閉セルどうしの間で所期の処理を実施する演算部と、目的の閉セルの抽出手順および処理をホモトピーとして記録し、かつ処理が実施された目的の閉セルをホモトピーと関連づけて記録する記録部と、指定部にて、注目しようとする属性の組が新たに指定されたとき、この組に対応する目的の閉セルがすでに前記記録部に記録されているか否かを判定する判定部とを備え、目的の閉セルが記録部に記録されていないとき

50

抽出部による抽出が行われ、目的の閉セルが記録部に記録されているときは抽出部による抽出はスキップされ、記録部から当該目的の閉セルが読み出されて利用される。

【0012】

抽出部は、第1および第2のデータベース自体の書換を行わない方法にて目的の閉セルを抽出し、演算部は、処理が実施された目的の閉セルを第1および第2のデータベースが物理的に記憶されている領域とは別の記憶領域へ保存してもよい。

【0013】

本発明のさらに別の態様も情報管理装置に関し、第1および第2のデータベースを参照する参照部と、第1および第2のデータベースにおいて注目すべき情報を指定する指定部と、第1および第2のデータベースに記述された情報をそれぞれ付加空間モデルにおける第1および第2のセルとして扱い、注目すべき情報を第1および第2のセルからそれぞれセル分解操作によって目的セルとして抽出する抽出部と、抽出された目的セルどうしの間で所期の処理を実施する演算部と、目的セルの抽出手順および処理をホモトピーとして記録し、かつ処理が実施された目的セルをホモトピーと関連づけて記録する記録部と、指定部にて、注目すべき情報が新たに指定されたとき、この注目すべき情報に対応する目的セルがすでに記録部に記録されているか否かを判定する判定部とを備え、目的セルが記録部に記録されていないとき抽出部による抽出が行われ、目的セルが記録部に記録されているときは抽出部による抽出はスキップされ、記録部から当該目的セルが読み出されて利用される。

【0014】

なお、以上の構成要素の任意の組合せ、本発明の表現を方法、装置、システム、記録媒体、コンピュータプログラムなどの間で変換したものもまた、本発明の態様として有効である。

【0015】

【発明の実施の形態】

以下本発明を好適な実施の形態をもとに説明する。まず、実施の形態の基礎として本発明者が提唱する付加空間モデル(Adjunction Space Model)およびセル構造空間モデル(以下単にセル情報モデルともいう)の基礎理論を前提技術として述べ、しかるのち具体的な実施の形態を説明する。なお、セル情報モデルは、付加空間モデルにセルの次元という概念を加えた下位モデルである。

【0016】

【前提技術】

セル情報モデルは、情報モデルとしては不規則データモデルの分野に適用できるもので、時空における諸元を状況(situation)というかたちでとらえる。数学的にいえば、セル情報モデルはホモトピーに関する理論的なフレームワークに位置づけられたセル空間構造理論にもとづき、グラフ理論の拡張理論にあたる。[1]サイバーワールドのモデリング

不变量を根拠に理論構築を行う。サイバーワールドを非可逆な空間としての時間を含む空間の1タイプと考え、その表現に、自由度としての次元と、異なる次元の空間がいかに接続されているかを示す接続性とを含む不变量が適切であることを示す。

【0017】

一般論として、サイバーワールドをモデリングするためには以下の4つのステップが必要である。

第1に、サイバーワールドと現実世界との相違点および共通点を明確化しなければならない。もっとも明白な違いは成長速度にあり、したがってその複雑さにある。ローカルな世界を同時に世界規模のウェブの世界にリンクする性質はきわめて特殊であり、かつそのスピードも光速に匹敵する。現実に、人類の歴史において、ウェブ上に実現される高速性は、かつてもったことのない能力を人類に与えた。ウェブ上で働くすべての人はサイバーワールドの構築と破壊を同時に行っているのである。

【0018】

10

20

30

40

50

第2に、明確化された相違点と共通点を特徴づけるための適切なモデリングの方法論を見いだす必要がある。極端な複雑度および変化の高速性により、モデリングの規模を最小化するための方法は、階層化された概念のうえに構築されるべきである。さらにその階層は、絶え間なく変化するサイバーワールドの中から普遍的な特性を特定するために、不变量の階層であり、後に概念の付加がモジュラーな形でつぎつぎに可能な形式にすべきである。

【0019】

第3に、そうして構築されたモデリングの方法論を、現実の設計 (design) に落とし込む必要がある。一般に設計は、不变量の適切な選択と特定の情報構造および演算を必要とする。たとえば不变量の概念階層は、不变量を相続する階層として設計される。これまでの研究により、ふたつの不变量、すなわち自由度としての次元およびそれらの接続性の重要性を認識した。すなわち、情報構造としてセル空間構造を考え、演算としてセルの構築 (コンポジション) および分解 (デコンポジション) を考えるに至った。

【0020】

第4に、そうして得られた設計をセル情報モデルと名付けた情報モデルとして実装する。セル情報モデルは、既存の種々のデータモデルの能力を強化するものであり、セルの境界、セルの次元およびセルの接続性のすべてを保証することができる。セル情報モデルは、サイバーワールドを一貫性のある形で表現でき、その正当性を証明できる。

[2] 不变量の概念階層

科学的な研究において、モデリングは非常に重要なステップである。とくに自然科学では、現実世界をモデリングするために、不变量の概念を中心として理論構築がなされる。オブジェクトおよび現象は不变量をもとに分類され、モデリングされる。物理では、相対性理論が発表されるまでエネルギーと質量は不变量であった。数学では、オブジェクトをモデリングするために以下の手順がとられる。すなわち、同値関係により、数学的なオブジェクトをその部分集合の排他的論理和として表現できる同値類へ分類する。同値関係にもとづく概念階層のレベルは以下のとおりである。

【0021】

1. 拡張されたレベル、その特別な場合としてホモトピーレベル
2. 集合レベル
3. 位相幾何学レベル、その特別な場合としてグラフ理論上のレベル
4. 付加空間レベル
5. セル構造空間レベル
6. 表現 (プレゼンテーション) レベル
7. 可視化レベル

これらの階層は、モジュラー、かつのちにインクリメンタルに付加していくことが可能な設計、すなわちサイバーワールドの不变量の相続的な階層を実現するために有力である。

[3] セル情報モデル

サイバーワールドをモデリングするためには、CW空間などのセル空間構造にもとづくアプローチが、グラフ理論に基づくそれに比べ、はるかに適している。セル構造空間レベルによれば、オブジェクトを境界の存在する、または存在しないセルとして、認識可能かつ計算可能な空間内に位置づけることができるためである。境界をもつセルは「閉」 (closed) であり、境界のないセルは「開」 (open) である。n次元のセル、すなわち「nセル」は、n次元球と同相の空間である (nは整数)。ここで、オープンなnセルを e^n と表記し、クローズなnセルを

【数1】

10

20

30

40

$$\mathcal{B}^n$$

と表記する。また、クローズな n セルの内部を
【数 2】

$$Int \mathcal{B}^n = B^n$$

10

と表記する。したがって、
【数 3】

$$\partial \mathcal{B}^n = \mathcal{B}^n - B^n = S^{n-1}$$

20

はクローズな n セルの境界にあたり、これは $(n - 1)$ 次元の球 S^{n-1} に等しい。セルモデリングによれば、セルの構築と分解はセルの次元と接続性を不变量に保ったまま実現できる。したがって、オブジェクトのアイデンティフィケーションは、アイデンティフィケーションのための写像をとおして体系的に実施される。後述するように、データベーススキーマの構築およびスキーマの分解は、セル構築およびセル分解の特別な場合に相当する。

【0022】

ここで次元の例を述べる。たとえばサイバーワールドにおいて、ひとつの属性をもつオブジェクトは、ひとつの属性から他の属性へ移行することができないため、その自由度は 0 であり、したがってその次元も 0 である。そのため、これを表現レベルでは「点」で表すことができる。属性とは、オブジェクトが本来有する特質や特徴を同定するための互いに独立な集合をいう。属性をふたつ有するオブジェクトでは、一方の属性から他方の属性への移行が可能なため、その自由度も次元も「1」である。したがって、表現レベルにおいて直線として表すことができる。同様に、属性が 3 および 4 のオブジェクトは、それぞれ 2 次元および 3 次元に相当し、曲面および球として表現できる。一般に、 n 個の属性を有するオブジェクトは $(n - 1)$ の自由度を有し、その次元は $n - 1$ である。これは、 $(n - 1)$ 次元の球として表現できる。リレーションナルモデルでは、 n 個の属性をもつオブジェクトはリレーションナルスキーマとして表され、 n 列のテーブルとして実体化される。リレーションナルモデルは集合の直積にもとづき、したがってそれは集合論レベルにおける表現といえる。

【0023】

一方、接続性は連続かつ全射な写像である接着写像（アスタッチングマップ）によって定義される。ある写像 $f : X \rightarrow Y$ が全射であるとは、

【数 4】

$$(y \in Y) \wedge (x \in X) \wedge [f(x) = y]$$

を意味する。「写像 $f : X \rightarrow Y$ が連続である」とは、「 $\{f^{-1}(y) \mid y \in A\}$ が X においてオープンであり、かつその場合にかぎり、 Y の部分集合である A が Y においてオープンになる」ことを意味する。

40

50

【0024】

共通部分をもたない位相空間 X および Y について、

【数5】

$$Y \sqcup_f X = Y \sqcup X / \sim$$

は、接着写像 $f : X \cup Y$ によって X を Y に接着することにより得られる付加空間（接着空間ともいう）である。 X と Y の単なる排他的論理和の空間から、接着関数 f によって定まる付加空間への写像 g がアイデンティフィケーション写像であり、この写像は *adjoin space model* において中心的な役割を果たす。付加空間を利用するモデルを付加空間モデルという。付加空間モデルは、接着関数とアイデンティフィケーション写像の組合せで表現される。一方、セル情報モデルは、付加空間モデルにセルの次元という概念を追加したものである。ここで、

【数6】

 \sqcup

10

20

は、排他的論理和を表し、しばしば $+$ であらわされる。 \sim は同値関係を示す。同値関係とは、同一律「 $x \sim x$ 」、対象律「 $x \sim y$ なら $y \sim x$ 」、推移律「 $x \sim y$ かつ $y \sim z$ なら $x \sim z$ 」がすべて成り立つ関係であり、集合論的な同値関係、ホモトピー同値関係、トポロジー同値関係などがある。推移律は、空間を、共通部分をもたない同値類と呼ばれる部分空間へ分割する。

【0025】

サイバースペースをより明確にモデリングするための基礎として、同値関係および同値類について述べる。 $X / \sim = \{ y \in X : x \sim y \}$ によって定義される X の部分集合は、 X の同値類と呼ばれる。ここで、「類（クラス）」は実際には集合のことであるが、昔から類（クラス）と呼ばれているため、その表記にしたがう。すべての同値類からなる集合 X / \sim は、 X の商空間と呼ばれ、以下のように表記される。

【0026】

【数7】

$$X / \sim = \{ x / \sim \in 2^X \mid x \in X \} = 2^X$$

推移律から、 $x \sim X$, x / \sim を満たすそれぞれの x について、以下の式が成り立つ。

【数8】

30

40

$$x \sim y \Leftrightarrow x / \sim = y / \sim, \text{ and}$$

$$x \not\sim y \Leftrightarrow x / \sim \cap y / \sim = \emptyset$$

これは、集合 X が、空ではなく、共通部分をもたない同値類へ分割または分解されたことを意味する。ここで同値類を x / \sim と表記し、これは以下の意味である。

【数9】

$$x / \sim = \{ y \in X \mid x \sim y \}$$

簡単な例で説明する。「濃度（cardinality）」は、集合論上の同値関係で

50

あり、もとの集合を同じ濃度を有する、共通部分をもたない部分集合へ分割する。別な例として、グラフ理論において「同型」は同値関係であり、グラフの集合も、共通部分をもたない同型のグラフの部分集合へ分解できる。

【0027】

ユークリッド幾何学において、「合同」はひとつの同値関係を形成し、すべての図形を互いに合同な図形からなる部分集合へ分解できる。これらの部分集合は共通部分をもたず、その和集合がもとの集合、すなわちすべての図形の集合に一致する。この和集合が商空間に当たる。「相似」もひとつの同値関係である。「合同」および「相似」はともにアフィン変換の例である。一方、「対称」という関係は、群理論における同値関係の例であり、対称な図形からなる互いに共通部分をもたない部分集合の和集合へ分解する。

10

【0028】

以上が接着写像の実例である。ここで、接着写像の一般的な定義に触れる。すべての同値類の集合は X / \sim と表記され、以下の式で示される。

【数10】

$$X / \sim = \{ x / \sim \mid x \in X \} = 2^X$$

これは X の商空間とも呼ばれる。接着写像 f は全射かつ連続な以下の写像である。

【数11】

$$f : X_0 \rightarrow Y (X_0 \rightarrow Y)$$

【数12】

20

$$X \sqcup Y / \sim$$

は商空間であり、以下の関係をもつ。

【数13】

30

$$X \sqcup Y / \sim = X \sqcup Y / (x \sim f(x) \mid \forall x \in X_0) = X \sqcup_f Y$$

ここでは、後述するように、情報スキーマの統合およびウェブ上の情報マイニングによる情報の統合のための特別な場合を考える。いま S^{n-1} は、クローズな n セルの境界であり、

【数14】

40

$$S^{n-1} = \partial B^n = B^n - \text{Int } B^n = B^n - e^n$$

と表記できる。ここで全射かつ連続な接着写像 f を

【数15】

$$f : S^{n-1} \rightarrow X$$

と定義する。このとき、付加空間 Y は以下のように商空間として定義される。

【0029】

50

【数16】

$$Y = X \sqcup_f \mathcal{B}^n = X \sqcup \mathcal{B}^n / \{f(u) \sim u \mid u \in S^{n-1}\}$$

いま、ホモトピックな写像 f および g 、

【数17】

$$f, g : S^{n-1} \rightarrow X$$

を考える。すると、

【数18】

$$X \sqcup_f \mathcal{B}^n \simeq X \sqcup_g \mathcal{B}^n$$

というホモトピー同値関係が生じる。

【0030】

J. H. C. Whitehead の指摘によれば、位相空間として任意のサイバーワールド X が与えられたとき、この X から整数 Z によってインデックスが与えられた X の部分空間である X^p セルの有限または無限の配列を帰納的に構成することができる。すなわちフィルトレーションと呼ばれる空間 $\{X^p \mid X^p \subset X, p \in Z\}$ が以下のように形成できる。ここで X^p は X の被覆とよばれ、以下の関係がなりたつ。

【数19】

$$X = \bigcup_{p \in Z} X^p$$

さらに、 X^{p-1} は X^p の部分空間であり、すなわち、

【数20】

$$X^0 \subset X^1 \subset X^2 \subset \dots \subset X^{p-1} \subset X^p \subset \dots \subset X$$

と表記できる。フィルトレーションはスケルトンとも呼ばれる。最大で p 次元のスケルトンは p -スケルトンと呼ばれる。 $X^0, X^1, X^2, \dots, X^{p-1}$ および X^p はサイバーワールド X の部分サイバーワールドである。フィルトレーションと位相的に同値な空間はフィルトレーション空間と呼ばれる。

【0031】

実用上、重要なセル空間がある。それらは CW 複体および多様体である。フィルトレーション空間が有限であるとき、これは CW 空間と同値である。さらに、CW 空間が可微分性を有するとき、これは多様体と同値である。

[4] セル情報モデルによる情報マイニングをとおしたウェブ情報のモデリング、帰納的なウェブ情報スキーマの統合およびウェブ情報の統合

ウェブ情報をモデリングするための第一歩として、サイバーワールドがいかに出現し、その実体がなにであるかを見きわめるために、ウェブ上の共有情報世界であるサイバーワールド形成の本質の特徴づけを行う。サイバーワールド X は、しばしば多くのウェブサイトにおけるローカルかつ多岐にわたる活動の結果ウェブ上に形成される。企業内の情報とは異なり、開始点となるスキーマの集合を与えてくれる情報管理者の存在を仮定することはできない。情報マイニングのプロセスをとおし、ローカルに存在する複数のウェブサイトにおける特別な情報を発見してサイバースペース X を知ることができる。もちろん情報マイニングは手当たりしだいすべきものではない。ウェブサイトをブラウザでながめたのち

10

20

30

40

50

、複数のウェブサイトに分散して存在する情報およびその統合からなにをマイニングすべきか、およびいかなるものが出現すると予測されるかについてアイデアを抽出しなければならない。この種の情報マイニングは一般に「設計に基づく情報マイニング」と呼ばれる。なぜなら、マイニングすべき対象に関し、「統合指針」として適用すべき所定の規則が存在するためである。この統合指針は、なにをどのように統合するかについて設計指針として働く。

【0032】

上述のWhiteheadの帰納的スキームに基づくウェブ上の情報マイニングによれば、ローカルなウェブの世界の全世界規模のサイバーワールドへの統合は完全な形で実現される。帰納的な統合によってn次元のサイバーワールド X^n が取得される具体的な方法を、以下ウェブ上のサーチおよび統合のプロセスによって説明する。

10

20

【0033】

帰納的な統合はふたつのフェイズからなる。すなわち情報のスキーマ統合フェイズおよび情報統合フェイズである。第1のフェイズである情報スキーマの統合フェイズは以下の手順で進行する。

1. 興味の対象である属性である

【数21】

$$\mathcal{B}_i^0$$

をすべて読み出し、以下の0次元のサイバーワールド X^0 を形成する。

【数22】

$$X^0 = \{ e^0_1, e^0_2, e^0_3, \dots, e^0_j \}$$

2. 1次元のサイバーワールド X^1 を生成するために、ウェブサイトにおける興味の対象であるふたつの属性のすべての組合せ

【数23】

30

$$\mathcal{B}_i^I$$

を読み出す。そののち、それらの共通部分のない和集合、

【数24】

$$\sqcup_i \mathcal{B}_i^I = \mathcal{B}_1^I \sqcup \mathcal{B}_2^I \sqcup \mathcal{B}_3^I \sqcup \dots \sqcup \mathcal{B}_k^I$$

40

|

を接着写像 F によって X^0 へ接着する。こうして、以下の1次元サイバーワールド X^1 を得ることができる。

【数25】

$$X^I = X^0 \sqcup_F (\sqcup_i \mathcal{B}^I{}_i) = X^0 \sqcup (\sqcup_i e^I{}_i)$$

ただしここで、 $i = 1, 2, \dots, k$ であり、接着写像 F は、
【数 2 6】

$$F: \sqcup_i \partial \mathcal{B}^I{}_i \rightarrow X^0$$

10

である。

【0 0 3 4】

3. 属性の読み出しあより統合を繰り返すことにより、情報マイニングを経て ($n - 1$) 次元のサイバーワールド X^{n-1} を構築したとする。ここで X^{n-1} は n 個の属性を有する。 $(n + 1)$ 個の属性を有する n 次元のサイバーワールド X^n を統合的に生成するため、いままでと同様の方法でウェブサイトにおいて興味の対象である $(n + 1)$ 個の属性
【数 2 7】

20

$$\mathcal{B}^n{}_i$$

のすべての組み合わせを読み出す。つづいてそれらの共通部分のない和集合、
【数 2 8】

30

$$\sqcup_i \mathcal{B}^n{}_i = \mathcal{B}^n{}_1 \sqcup \mathcal{B}^n{}_2 \sqcup \mathcal{B}^n{}_3 \sqcup \dots \sqcup \mathcal{B}^n{}_m$$

をすでに構築された $(n - 1)$ 次元のサイバーワールド X^{n-1} へ接着写像 G をとおして接着する。この結果、以下のように n 次元のサイバーワールド X^n を生成することができる。

【数 2 9】

40

$$X^n = X^{n-1} \sqcup_G (\sqcup_i \mathcal{B}^n{}_i) = X^{n-1} \sqcup (\sqcup_i e^n{}_i)$$

ただし、 $i = 1, 2, \dots, k$ であり、接着写像 G は、
【数 3 0】

$$G: \sqcup_i \partial \mathcal{B}^n_i \rightarrow X^{n-1}$$

である。以上のプロセスにより、情報スキーマの統合が完了する。

【0035】

一方、第2のフェイズである情報統合フェイズは、きわめて単純であるが計算量が多い。このフェイズは、設計指針に基づき、セル接着によって生成されるサイバーワールドに含まれるべきインスタンスを判断および決定するために、スキーマ統合の際に行われるセル接着のすべてのステップにおいて、すべてのインスタンスを検査する。

【0036】

Whiteheadの帰納的な方法論に基づいて構築したサイバーワールドは、以下の関係式を満たす。

【数31】

$$X^0 \quad X^1 \quad X^2 \quad \dots \quad X^{n-1} \quad X^n \quad \dots \quad X$$

サイバーワールドの有効性の観点からいえば、この式は任意の有効なサイバーワールドがそれ以下の次元のサイバーワールドを含み、かつそれらのサイバーワールドが有効であることを意味する。

【0037】

上述の例において、アイデンティフィケーションは同値関係に基づく同値類によって行われる。「同値類によるアイデンティフィケーション」はリレーションナルモデルにおける統合(*join*)演算の一般化である。この点は、セル情報モデルの実用上の能力の一部を示している。ウェブ上のサイバーワールドの高度に複雑かつきわめて速い変化に鑑みれば、セル情報モデルのこの統合能力は、ウェブ情報モデルとして真の理論的基礎を提供するものである。

【0038】

なお、設計指針を実行するために「興味の対象である属性」というとき、「興味」とは、すくなくとも部分的な意味において、アイデンティフィケーションのための同値関係の選択を意味する。すなわち、「アイデンティフィケーションのための同値関係の選択」は設計指針の主要部分を占めている。ウェブに関連する情報システムにおいて、設計指針は、インターネットまたはコミュニティネットとしてのローカルなサイトを統治するためにローカルに存在するか、または国境のないサイバーワールドにおいて作用すべくグローバルに存在する。設計指針は、ウェブに基づく情報システムにおいて再利用可能なリソースである。

【5】帰納的でない情報スキーマの統合としてのウェブ情報の状況モデリング、およびセル情報モデルに基づく情報の統合

ウェブにおいて、しばしば、任意のサイバーワールドから新たなサイバーワールドを創造する必要が生じる。これは、前章で述べた帰納的手法をつうじた情報マイニングよりも一般的であり、ウェブ上の電子商取引を含むe-ビジネスにおいてよく見られる要請である。たとえば、時空の両面で変化するウェブの状況をモデリングするために電子商取引の状況を考える。電子商取引の情報システムを構築するために、情報スキーマの観点からウェブ上の商取引の構造を見いだすことが一般に必要である。典型的な電子商取引における状況は以下のものを含む。

【0039】

状況1. ある商品を購買するe-カスタマは、その商品をもっとも安い値段で販売するe-ショップを探すためにウェブをブラウズする。

状況2. ウェブ上で商品を販売するe-ショップは、セールスを拡大するためにe-カスタマのリストをブラウズする。

10

20

20

30

40

50

【0040】

この状況において、ウェブ上でわれわれはe - ショップ、e - カスタマおよびe - 商品に関するすべての詳細情報を見いだすことに興味をもつわけではない。ここで、e - ショップ、e - カスタマおよびe - 商品をそれぞれs、cおよびm次元のサイバーワールドとし、したがってそれぞれsセルe^s、cセルe^cおよびmセルe^mと表記する。

【0041】

状況1において、e - カスタマは、あるe - ショップにおいて所望のe - 商品がもっとも安い価格で売られているとき、購入者としての興味をもって、そのe - ショップにおける商品名を特定する。この状況は、セル分解演算およびそのあとに行われるアイデンティファイケーション演算によって特徴づけることができる。セル分解演算は、下に示す写像fであらわされる。この写像fは、接着写像gが保存されるかたちで、任意のn次元セルeⁿをつぎの2つの共通部分をもたないセルの和集合へ射影する。

【数32】

$$e^u \sqcup_g e^v \ (u, v \leq n)$$

【数33】

10

20

$$f: \mathcal{B}^n \rightarrow \mathcal{B}^u \sqcup_g \mathcal{B}^v = e^n \rightarrow e^u \sqcup e^v$$

後述するように、各セル分解において接着写像を保存することにより、セル分解をホモトピックにすることができる。状況1に関する結論は、それを以下の状況モデルで理解することである。

1. セル分解

e - ショップとしてのsセルe^s、e - カスタマとしてのcセルe^cおよびe - 商品としてのmセルe^mをセル分解する。このとき、電子商取引に関連する属性を特定すべく、同値セルe^qをそれ以外の部分から分離する。属性の例として、たとえばe^qをe²へ簡略化して示せば、商品名、商品の識別情報および商品の価格がある。

2. セル接着によるセル構築

同値セルe^qを接着写像によってアイデンティファイする。すなわち、e - 商品としてのmセルe^m、およびe - ショップとしてのsセルe^sをe - カスタマとしてのcセルe^cへ接着する。

【0042】

状況2も同様に、以下の状況モデルとして具体化される。

30

40

1. セル分解

状況1同様である。

2. セル接着によるセル構築

同値セルe^qを接着写像をとおしてアイデンティファイする。すなわち、e - 商品としてのmセル{e^m_i}およびe - カスタマとしてのcセル{e^c_i}をe - ショップとしてのsセルe^sへ接着する。

[6] 空間 / 時間情報および空間 / 時間演算のためのセル情報モデルの理論的フレームワークとしてのホモトピー

新世紀が幕を開けたいま、われわれは、現実の世界に対し、非常に根幹的な方法で影響を与えることができる時代になった。そのような瞬間に立ち会えることはきわめて幸運なこ

50

と言わねばならない。21世紀に大きな役割を果たすと期待されるウェブおよびサイバーワールドに関する科学を構築することは、ウェブを基礎とする情報テクノロジーの構築に最大の貢献をするであろう。サイバーワールドは情報の世界であり、その意味においてウェブおよびサイバーワールドの情報モデルはキーエレメントである。同様に幸運なことに、われわれはセル空間構造として述べた科学を創造するために必要な数学的フレームワークをもっている。以下述べるホモトピー理論もそうである。

【0043】

ホモトピー理論は、セル空間構造の基礎理論として働く。すなわち、サイバーワールドの時間および空間における変化を扱うとき、空間 / 時間情報および空間 / 時間演算を収容するためにホモトピー理論が利用される。いまたとえば、ひとつの位相空間 X から別の位相空間 Y への写像関数 f の変化を考える。変化ののち、 f は別の写像関数 g になる。したがって、以下の f から g への連続変形を設計する。

【数34】

$f, g : X \rightarrow Y$

この変形を正規化された区間 $[0, 1]$ について考える。この区間は時間的、空間的とを問わない。いま、位相空間 X のうち変化のない部分 A を X の部分空間 A として $A \subset X$ と表記する。設計すべきホモトピー H は以下のとおりである。

【0044】

【数35】

$H : X \times I \rightarrow Y$ ただし、

$$(x, 0) (H(x, 0)) = f(x) \text{ and } H(x, 1) = g(x)$$

および

$$(a, t) (H(a, t)) = f(a) = g(a)$$

このとき f は A に関して g とホモトピックとよばれ、以下のように表記される。

【数36】

$$f \simeq g \text{ (rel } A)$$

20

30

ここで新たな設計上の問題が生じる。すなわち、2つの位相空間 X および Y をホモトピイ同値、

【数37】

$$X \simeq Y$$

40

として設計する方法、つまりこれらを同じホモトピー型をもつよう設計する方法である。これは以下の手順で解決される。すなわち、 $f : X \rightarrow Y$ および $h : Y \rightarrow X$ が以下の条件を満たせばよい。

【数38】

$$h \circ f \simeq 1_X \text{ and } f \circ h \simeq 1_Y$$

50

ここで 1_x および 1_y は恒等写像であり、以下の式をみたす。

【数39】

$1_x : X \times X$ and $1_y : Y \times Y$

以上の手法で、セルの次元をホモトピックに変化させることができる。ホモトピー同値はトポロジー同値よりも広い概念である。ホモトピー同値は、変化の前後において、位相幾何学的にはもはや同値といえないサイバーワールドのいかなる変化をもアイデンティファイすることができる。サイバーワールドは種々の演算および処理によって変遷を重ねてゆくが、その変化のプロセスはホモトピーによって特定され、またホモトピー同値によって有効性が保障される。たとえば、それぞれのセル分解を実行する際、なぜ接着写像が保存されるかが理解できる。それはセル分解をホモトピックに保ち、したがってセル分解のプロセスを逆向きにたどることができるためである。

10

【0045】

ホモトピックな情報モデルの研究は、情報モデルの科学を探求するために、今後取り組まれていくべき分野である。いかなる情報演算がホモトピー同値になるかを検討することは非常に興味深い研究テーマになる。

【0046】

【具体例】

以上の前提技術をもとに、以下本発明に係る情報管理装置を説明する。

図1は実施の形態に係る情報管理装置10の構成を示す。情報管理装置10は、データベース参照部12、メモリ14、解析部16、GUI(グラフィカル・ユーザ・インターフェイス)18、整形部24を備える。データベース参照部12はインターネット経由でウェブ上から、またはローカルを含む任意の記憶装置から、ユーザが興味を示すデータベースを探索し、参照する。参照されたデータベースは、ワークエリアとしてのメモリ14で保持される。解析部16はGUI18から与えられる指示をもとにメモリ14に格納されたデータベースを解析し、必要な情報を抽出し、整形部24へ出力する。整形部24は渡された情報を図示しない表示用テンプレートへ流し込み、その他表示に必要な処理を施し、図示しない表示装置へ出力する。ユーザは表示装置で必要な情報を確認する。以下、実施の形態では、第1、第2データベースというふたつの異なるデータベースを読み出して利用するが、これらを併せて単にデータベースとよぶこともある。

20

【0047】

解析部16は、抽出部20、演算部22、手順取得部26、記録部28、判定部30を備える。GUI18は属性指定部32を備える。ユーザは、属性指定部32において、データベース中、注目しようとする情報を指定する。指定は、例えば「ある商品の価格」のように行われ、そこからデータベースの属性「商品」「価格」などが抽出部20へ送られる。以下、これら注目すべき属性の組を単に「注目情報」ともいう。

30

【0048】

抽出部20は第1、第2データベースからそれぞれ注目情報に当たる部分を抽出する。抽出は前提技術でいうセル分解であり、その結果、それぞれのデータベースから第1、第2の注目情報のセルが抽出される。セル分解にはいくつかのアルゴリズムが考えられるが、セルの境界をとることによってセルの次元を落としていく場合、すなわち、 B^n から B^n を生成していく場合、必要な次数になるまで処理を繰り返すため、抽出部20には自身へのフィードバックループが形成されている。

40

【0049】

演算部22は、抽出の結果得られたふたつの注目情報のセル間で所定の演算を施し、その結果を整形部24へ出力する。セル間の演算は、たとえばそれらのセルの統合であり、前提技術でいう接着写像で表現される。演算の結果は記録部28へ記録され、必要に応じて再利用される。

【0050】

手順取得部26は、GUI18から解析部16に与えられた指示、抽出部20による抽出の手順、演算部22による演算の手順をすべてログ的に取得し、一連の操作を示すホモト

50

ピーを記録部 28 へ記録する。その際、演算部 22 の演算の結果と当該ホモトピーを関連づけて保存する。

【0051】

判定部 30 は、ユーザが注目情報を指定したとき、それに対応する目的のセルがすでに記録部 28 に存在するかどうかを記録部 28 内を検索して判定する。存在する場合、判定部 30 は抽出部 20 に対して抽出処理を停止するよう指示し、記録部 28 から目的のセルを読み出してこれを整形部 24 へ提供する。一方、セルが記録部 28 に存在しないとき、判定部 30 は抽出部 20 に対して抽出処理の実行を指示する。

【0052】

図 2 は、抽出部 20 によるセル分解の概念を示す。同図では、データベース全体を n 次元閉セル B^n として表現し、その中のひとつの属性を外した境界を B^n と表示している。ここでは、外した属性は「 a_0 」であるが、いずれの属性を外していくかは、属性指定部 32 から入力された注目情報をもとに判断する。たとえば、注目情報を構成する属性が「 a_3 」と「 a_4 」であれば、図 2 の n 次元閉セルの属性を昇順に走査し、「 a_0 」「 a_1 」「 a_2 」「 a_5 」……のように外していくべき。

【0053】

以上の構成による情報管理装置 10 の動作を具体例をもとに説明する。ここでは、第 1、第 2 データベースとして、ふたつの店舗がそれぞれ構築する販売データベースを考える。これらの店舗は自身で用いるデータベースを最適化している結果、それらのデータベース構造は異なり、リレーションナルデータベースモデルでは両データベースを跨ぐようなデータ操作を行うことができない。ここでは、それらのデータベースから、ある特定の商品の実販累積数を調べたいとする。

【0054】

図 3、図 4 は、それぞれ第 1 データベース 100、第 2 データベース 120 の構成を示す。第 1 データベース 100 は、商品欄 102、価格欄 104、累積欄 106、メーカー欄 108 をもつ。第 2 データベース 120 は、商品欄 102、メーカー欄 108、発売日欄 122、累積欄 106、色欄 124 をもつ。

【0055】

ユーザはまず、GUI 18 の属性指定部 32 により、「商品 $d e f$ および $g h i$ の実販累積数を知りたい」と入力する。ここから、注目情報が「商品」「累積」と判明する。抽出部 20 は第 1 データベース 100 および第 2 データベース 120 から注目情報を抽出する。図 5、図 6 は、それぞれ第 1 データベース 100、第 2 データベース 120 から分解によって抽出された注目情報の閉セルである第 1 注目セル 130 と第 2 注目セル 140 を示す。第 1 注目セル 130、第 2 注目セル 140 は、それぞれ ID 欄 132、商品欄 102、累積欄 106 を有する。ID 欄 132 は、セル自身を特定する ID を格納し、これにより再利用性が確保される。第 1 注目セル 130 と第 2 注目セル 140 は、それぞれ「ID 1」「ID 2」という ID で必要に応じて記録部 28 へ記録される。商品欄 102 には、それぞれ商品「 $d e f$ 」と「 $g h i$ 」のみが挙げられている。

【0056】

つづいて、これらふたつの注目セルは演算部 22 へ投入される。演算部 22 は演算として、「ふたつの注目セルの統合」とその際、「累積数の加算」を実行する。

【0057】

図 7 はその結果得られた最終目的セルである第 3 注目セル 150 を示す。第 3 注目セル 150 は、第 1 注目セル 130 等と同じ構造をもつが、累積欄 106 における数字が第 1 注目セル 130 と第 2 注目セル 140 の合計になっている。第 3 注目セル 150 は整形部 24 へ出力され、ユーザへ結果として表示される。また、「ID 3」が付与され、記録部 28 へ格納される。

【0058】

図 8 は、記録部 28 に記録されたログ情報 200 を示す。ログ情報 200 は、一連の処理を時系列に示すホモトピー 202 と、その結果最終的に得られた閉セルである第 3 注目セ

10

20

30

40

50

ル 150へのリンク情報 204を含む。ホモトピー 202として、処理が以下のごとく最初から記述されている。

【0059】

第10行： 第1データベース 100と第2データベース 120の参照

第20行： 商品 d e f と g h i の累積を集計すべき旨の指示の入力

第30行： 第1データベース 100から注目情報である第1注目セル 130 (ID1) を生成

第40行： 第2データベース 120から注目情報である第2注目セル 140 (ID2) を生成

第50行： 注目情報である第1注目セル 130と第2注目セル 140を商品 d e f と g h i について接着写像 f で集計し、第3注目セル 150を生成 10

第60行： 第3注目セル 150 (ID3) を保存。

【0060】

以上、実施の形態を説明した。この実施の形態によれば、以下のような作用効果が得られる。

まず、リレーションナルデータベースモデルと異なり、構造が違うデータベース間で自由なデータの参照と操作が可能になる。また、その操作をホモトピーで残すことにより、最初のデータベースから最後の結果に至る任意の時点のデータを取り出すことができる。リレーションナルデータベースモデルでは、ホモトピーを残す発想がないため、射影（プロジェクト）によって必要な属性を取り出すと、もとのテーブルとの関連性を維持することができず、データの柔軟な活用が困難である。 20

【0061】

なお、手順取得部 26 は、ホモトピーを利用して任意の時点の注目情報を特定する機能、すなわち G U I 18 を介してユーザが求める時点を特定し、ホモトピーを上から、または下から辿ることにより、所期の時点における注目セルを特定する機能を備えてよい。

【0062】

目的セルである第3注目セル 150を最初のデータベースである第1データベース 100 、第2データベース 120 とは別に生成し、別の個所へ記録したため、目的セルの再利用性が確保されている。 30

【0063】

以上、実施の形態は例示であり、さまざまな変形例が可能であり、そうした変形例も本発明に含まれることは当業者に理解されるところである。

【0064】

【発明の効果】

本発明によれば、データベースの情報管理を効果的に行うことができる。

【図面の簡単な説明】

【図1】実施の形態に係る情報管理装置の構成図である。

【図2】実施の形態で注目情報を境界から抽出する概念を示す図である。

【図3】実施の形態で読み込む第1データベースを示す図である。

【図4】実施の形態で読み込む第2データベースを示す図である。 40

【図5】実施の形態で第1データベースから抽出された注目情報に相当する第1注目セルを示す図である。

【図6】実施の形態で第2データベースから抽出された注目情報に相当する第2注目セルを示す図である。

【図7】第1注目セルと第2注目セルから演算の結果求められた目的セルである第3注目セルを示す図である。

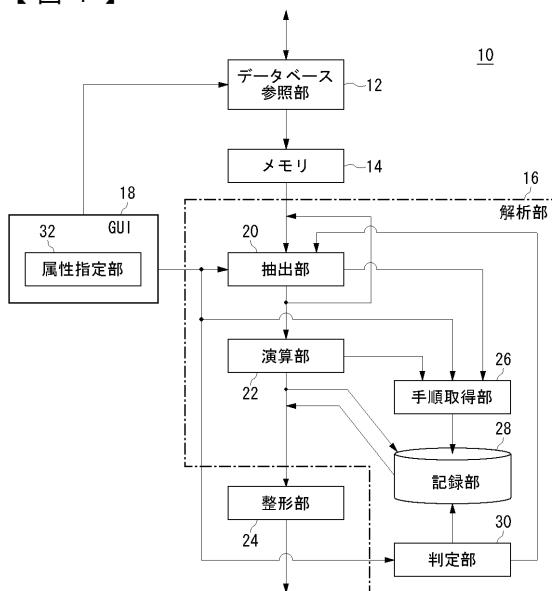
【図8】実施の形態の一連の処理を記述したホモトピーを含むログ情報を示す図である。

【符号の説明】

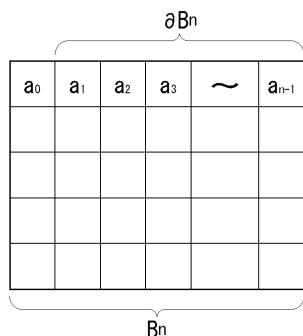
10 情報管理装置、 12 データベース参照部、 20 抽出部、 22 演算部、

26 手順取得部、 28 記録部、 30 判定部、 32 属性指定部。 50

【 図 1 】



【 図 2 】



【 図 3 】

商品	価格	累積	メーカー	
abc	1500	35	X	
def	800	8	Y	
ghi	3000	22	Z	

〔図4〕

商品	メーカー	発売日	累積	色	---
def	Y	Jan1	106	白	---
ghi	Z	May10	83	赤	---
hmn	W	Nov15	46	黄	---

〔 図 5 〕

132	102	106
ID1	商品	累積
	def	8
	ghi	22

〔 义 6 〕

132	102	106
ID2	商品	累積
	def	106
	ghi	83

〔 図 7 〕

132	102	106
ID3	商品	累積
	def	114
	ghi	105

〔 8 〕

10 : 第1データベース、第2データベースの参照	
20 : 商品def、ghiの累積集計指示	
30 : 第1データベースから注目情報(ID1)の生成	202
40 : 第2データベースから注目情報(ID2)の生成	
50 : 注目情報(ID1、ID2)の集計(f)	
60 : 集計結果(ID3)の保存	
LINK : ID3 at xxxxxh	204

140