

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2010年2月4日(04.02.2010)

PCT

(10) 国際公開番号
WO 2010/013472 A1

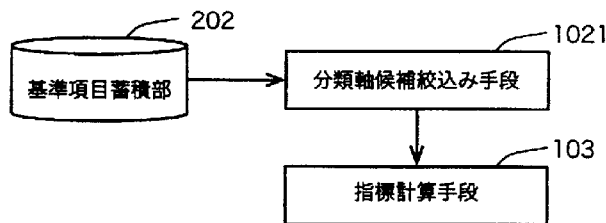
- (51) 国際特許分類:
G06F 17/30 (2006.01)
- (21) 国際出願番号: PCT/JP2009/003601
- (22) 国際出願日: 2009年7月29日(29.07.2009)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2008-195895 2008年7月30日(30.07.2008) JP
- (71) 出願人 (米国を除く全ての指定国について): 日本電気株式会社 (NEC Corporation) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 水口弘紀 (MIZUGUCHI, Hironori) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 立石健二 (TATEISHI, Kenji) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 細見格 (HOSOMI, Itaru) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 久寿居大
- (54) Title: DATA CLASSIFICATION SYSTEM, DATA CLASSIFICATION METHOD, AND DATA CLASSIFICATION PROGRAM
- (54) 発明の名称: データ分類システム、データ分類方法、及びデータ分類プログラム
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ,

[続葉有]



WO 2010/013472 A1

[図24]



- 103 Index Calculation Means
- 202 Criteria Accumulating Unit
- 1021 Category Axis Candidate Limiting Means

(57) Abstract: A data classification system selects, on the basis of hierarchical classification items and data categories that correspond to said classification items, a multiplicity of classification items corresponding to said data categories and outputs the same as a category axis. The data classification system is equipped with a criteria accumulation means, a category axis limiting means, and a priority calculation means. The criteria accumulation means accumulates classification item categories in advance that will become the criteria for selecting classification items. The category axis limiting means selects a multiplicity of classification items from the classification items descendant from the criteria as category axis candidates, and then limits the category axis candidates that will be subject to calculation on the basis of the data quantity corresponding to the classification item and the hierarchical distance between the classification items. The priority calculation means calculates the priority for displaying the category axis candidates limited by the category axis limiting means.

(57) 要約:

[続葉有]

CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, 添付公開書類:
TD, TG).

— 国際調査報告 (条約第 21 条(3))

本発明に係るデータ分類システムは、階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、当該データ群に対応する分類項目を複数選択して分類軸として出力する。該データ分類システムは、基準項目蓄積手段と、分類軸絞込手段と、優先度計算手段と、を備える。基準項目蓄積手段は、分類項目を選択するための基準項目となる分類項目群を予め蓄積する。分類軸絞込手段は、基準項目の子孫の分類項目から複数の分類項目を選択して分類軸候補とし、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む。優先度計算手段は、前記分類軸絞込手段が絞り込んだ分類軸候補について、当該分類軸候補を表示する優先度を計算する。

明 細 書

発明の名称：

データ分類システム、データ分類方法、及びデータ分類プログラム

技術分野

[0001] 本発明は、データ分類システム、データ分類方法、及びデータ分類プログラムに関する。

本願は、2008年7月30日に、日本に出願された特願2008-195895号に基づき優先権を主張し、その内容をここに援用する。

背景技術

[0002] データ整理を行う場合、大量のデータを整理するために階層的な分類項目を付与することが多い。例えば、非特許文献1に記載されているデータベースでは、各文献にMeSHタームと呼ばれる階層的な分類項目が付与されている。また、例えば、特許庁が運営する特許データベースにおいても、各特許文献にFターム等の階層的な分類項目を複数付与している。以下、このようなデータに付与される階層的な分類項目を分類階層という。

[0003] また、非特許文献2には、上記のような階層的な分類項目を付与した文書群を閲覧できる関連技術が記載されている。非特許文献2には、OLAPと呼ばれる多面的なデータ表示方法を拡張し、階層が非常に大きな場合でも高速に処理できるようにしたシステムが記載されている。非特許文献2に記載された関連技術では、ユーザが分類項目を選択することにより、その分類項目に含まれるデータ件数等を高速に表示することができる。また、非特許文献2に記載された関連技術では、縦軸の分類項目と横軸の分類項目とをユーザが選択し、その結果を表形式で表示する。

[0004] 上記のようにすることにより、非特許文献2に記載された関連技術では、複数の分類項目が付与されている文書を一覧表示することができる。以下、データを表示する際に用いる分類項目群を分類軸という。

[0005] しかし、分類階層が非常に大きい場合、ユーザがどの分類項目を選んだら

よいのか判断するのは難しい。例えば、非特許文献2に記載されたシステムで用いている文書群では、文書数が約50万件存在し、分類項目数が約34万分類存在する。従って、この分類階層中から表示したい分類項目をユーザが選択することは非常に困難である。

[0006] また、非特許文献3には、関連する分類項目選択方法が記載されている。非特許文献3には、文書検索において検索結果文書リストとともに、それら検索結果文書リストに関連する分類軸を表示する方法が示されている。非特許文献3に記載された方法では、キーワードを入力して文書を検索し、検索結果文書を表示するとともに、検索結果文書群に予め付与された分類項目を複数表示し分類軸とする。また、分類軸を表示する際に、表示領域が限られているため分類項目を選択する。

[0007] また、非特許文献3には、分類項目のうち、検索結果文書が多い分類項目から順番に決められた上限まで選択する方法や、検索結果文書群を最も多く表示できる分類項目の組合せを選択する方法が記載されている。更に、非特許文献3には、全てのコンテンツを表示するクリック回数等のコストを最小にする方法が示されている。

先行技術文献

非特許文献

[0008] 非特許文献1：“PubMed”、National Center for Biotechnology Information、[平成20年7月4日検索]、インターネット<URL：<http://www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed>>

非特許文献2：猪口、武田、“テキスト分析のためのOLAPシステム”、情報処理学会論文誌、vol. 48、No. SIG11(TOD34)、p. 58-p. 68

非特許文献3：Wisam Dakka, Panagiotis G. Ipeirotis, Kenneth R. Wood, “Automatic Construction of Multifaceted Browsing Interfaces”, Proc. of CIKM' 05, p. 768-p. 775.

発明の概要

発明が解決しようとする課題

- [0009] しかし、非特許文献1～非特許文献3に記載された関連技術を用いたとしても、分類軸の優先度を計算する際に、ユーザにわかりやすい分類軸を選択するためには、多くの計算量を必要としている。その理由は、非特許文献1～非特許文献3に記載された関連技術を用いたとしても、計算対象となる分類軸を効果的に絞り込むことができていないためである。
- [0010] また、非特許文献3に記載された分類項目の選択方法を用いれば、計算時間を短縮するため、対応するデータ量が多い分類項目から順番に分類項目を選ぶことができる。又は、2つ目以降の分類項目を選択する際に、それまでに選択した分類項目と対応づけられていない分類項目のうちデータ量の多いものを選ぶことができる。
- [0011] しかし、非特許文献3に記載された方法では、単にデータ量の多い分類項目を選んでいくにすぎず、分類項目の意味的な関連性を考慮して分類項目を選択することはできない。そのため、非特許文献3に記載された方法を用いたとしても、分類項目を効果的に絞り込むことはできない。従って、分類項目間の意味的な関連性を考慮して分類項目を選択できるようにすることが望ましい。
- [0012] そこで、本発明が解決しようとする課題は、分類階層とその分類項目とに対応するデータ群が与えられたときに、分類軸の優先度の計算時間を短縮することのできるデータ分類システム、データ分類方法、及びデータ分類プログラムを提供することである。

課題を解決するための手段

- [0013] 本発明は、上述の課題を解決するために為されたものであり、本発明に係るデータ分類システムは、階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、前記データ群に対応する分類項目を複数選択して分類軸として出力するデータ分類システムであって、前記分類項目を選択するための基準項目となる分類項目群を予め蓄積する基準項目蓄積手段と、前記基準項目の子孫の分類項目から複数の分類項目を選択して分類軸候補とし、分

類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込手段と、前記分類軸絞込手段が絞り込んだ分類軸候補について、当該分類軸候補を表示する優先度を計算する優先度計算手段と、を備える。

[0014] 本発明に係るデータ分類システムの他の態様は、階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、前記データ群に対応する分類項目を複数選択して分類軸を作成し、複数の分類軸の組合せを多次元分類軸として出力するデータ分類システムであって、前記分類項目を選択するための基準項目となる分類項目群を予め蓄積する基準項目蓄積手段と、前記基準項目の子孫の分類項目においてデータに少なくとも1つ対応する分類項目の組合せに基づいて分類軸を作成する際に、分類項目に対応するデータ量と、分類項目間の階層的な距離と、に基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込手段と、前記分類軸絞込手段が絞り込んだ分類軸候補を組合せることによって、多次元の分類軸候補を作成する多次元分類軸作成手段と、分類階層における分類項目間の階層的な距離に基づいて、多次元分類軸候補の優先度を計算する多次元優先度計算手段と、を備えるものとしてもよい。

[0015] 本発明に係るデータ分類方法は、階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、前記データ群に対応する分類項目を複数選択して分類軸として出力するデータ分類方法であって、前記分類項目を選択するための基準項目となる分類項目群を予めデータベースに蓄積し、前記基準項目の子孫の分類項目から複数の分類項目を選択して分類軸候補とし、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込行程と、絞り込んだ前記分類軸候補について、当該分類軸候補を表示する優先度を計算する優先度計算行程と、を含む。

[0016] 本発明に係るデータ分類方法の他の態様は、階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、前記データ群に対応する分類項目

を複数選択して分類軸を作成し、複数の分類軸の組合せを多次元分類軸として出力するデータ分類方法であって、前記分類項目を選択するための基準項目となる分類項目群を予めデータベースに蓄積し、前記基準項目の子孫の分類項目においてデータに少なくとも1つ対応する分類項目の組合せに基づいて分類軸を作成する際に、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込行程と、絞り込んだ前記分類軸候補を組合せることによって、多次元の分類軸候補を作成する多次元分類軸作成行程と、分類階層における分類項目間の階層的な距離に基づいて、多次元分類軸候補の優先度を計算する多次元優先度計算行程と、を含むものとしてもよい。

[0017] 本発明に係るデータ分類プログラムは、階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、前記データ群に対応する分類項目を複数選択して分類軸として出力するためのデータ分類プログラムであって、前記分類項目を選択するための基準項目となる分類項目群を予め蓄積する基準項目蓄積手段を備えたコンピュータに、前記基準項目の子孫の分類項目から複数の分類項目を選択して分類軸候補とし、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込処理と、絞り込んだ前記分類軸候補について、当該分類軸候補を表示する優先度を計算する優先度計算処理と、を実行させる。

[0018] 本発明に係るデータ分類プログラムの他の態様は、階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、前記データ群に対応する分類項目を複数選択して分類軸を作成し、複数の分類軸の組合せを多次元分類軸として出力するためのデータ分類プログラムであって、前記分類項目を選択するための基準項目となる分類項目群を予め蓄積する基準項目蓄積手段を備えたコンピュータに、前記基準項目の子孫の分類項目においてデータに少なくとも1つ対応する分類項目の組合せに基づいて分類軸を作成する際に、分類項目に対応するデータ量と、分類項目間の階層的な距離と、に基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込処理と、絞り込んだ前記

分類軸候補を組合せることによって、多次元の分類軸候補を作成する多次元分類軸作成処理と、分類階層における分類項目間の階層的な距離に基づいて、多次元分類軸候補の優先度を計算する多次元優先度計算処理と、を実行させてもよい。

発明の効果

[0019] 本発明によれば、分類階層とその分類項目とに対応するデータ群が与えられたときに、分類軸の優先度の計算時間を短縮することができる。

図面の簡単な説明

[0020] [図1]本発明に係るデータ分類システムの構成の一例を示すブロック図である。

[図2]分類階層蓄積部が記憶する情報の一例を示す図である。

[図3]基準項目蓄積部が記憶する情報の一例を示す図である。

[図4]データ蓄積部が記憶する情報の一例を示す図である。

[図5]データ分類システムが行うデータ分類処理の一例を示す流れ図である。

[図6]図6（A）は、分類軸ID、基準項目、分類項目群及びスコアを対応付けたレコードを含むテーブルの例を示す図である。図6（B）は、分類軸ID、分類項目及びデータID群を対応付けたレコードを含むテーブルの例を示す図である。

[図7]第2の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。

[図8]分類階層の例を示す図である。

[図9]第2の実施形態におけるデータ分類システムが行うデータ分類処理の一例を示す流れ図である。

[図10]図10（A）は、分類項目別データ数表の例を示す図である。図10（B）は、データ別分類項目数表の例を示す図である。

[図11]第3の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。

[図12]表示手段が一覧形式で表示する情報の例を示す図である。

[図13]表示手段が表形式で表示する情報の例を示す図である。

[図14]第4の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。

[図15]第4の実施形態におけるデータ分類システムが行うデータ分類処理の一例を示す流れ図である。

[図16]図16(A)は、次元IDと分類軸ID群とスコアとを対応づけたレコードを含むテーブルの例を示す図である。図16(B)は、分類軸IDと基準項目と分類項目群とを対応付けたレコードを含むテーブルの例を示す図である。図16(C)は、分類軸IDと分類項目とデータID群とを対応付けたレコードを含むテーブルの例を示す図である。

[図17]第5の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。

[図18]図18(A)は、分類項目別データ数表の例を示す図である。図18(B)は、データ別分類項目数表の例を示す図である。

[図19]第6の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。

[図20]多次元表示手段が一覧形式で表示する情報の例を示す図である。

[図21]多次元表示手段が表形式で表示する情報の例を示す図である。

[図22]第7の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。

[図23]第8の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。

[図24]データ分類システムの最小の構成例を示すブロック図である。

発明を実施するための形態

[0021] 以下、具体的な実施形態を参照しながら、本発明について説明する。当業者であれば、本発明の記載を基に、多様な異なる実施形態を採り得るであろうし、本発明は、説明の目的で図示された実施形態に限定されるものではない。

- [0022] 以下、本発明の実施形態について説明する。まず、本発明によるデータ分類システムの概要について説明する。データ分類を行う場合、ユーザが分類軸を見る際に、データ量と分類項目との意味的な独立性が保たれていることが必要であると考えられる。また、データ量については、できるだけデータ量が多いことがよいと考えられる。例えば、分類軸を用いて概要把握を行なおうとする場合、分類軸に対応するデータが多くなければ、データ群の概要を把握することはできない。
- [0023] 意味的な独立性については、例えば、分類軸を用いて概要把握や絞込みを行なう場合、類似する分類項目ばかりでは、その他に関連する分類項目があったか否かわからない。そのため、ユーザによる選択操作が繰り返し行われてしまうことになる。また、類似する分類項目は分類階層の兄弟となることが多いので、本発明では、意味的な独立性を評価するために分類階層での分類項目間の距離を用いる。
- [0024] 本発明によるデータ分類システムは、分類軸候補絞込み手段と、指標計算手段とを備える。分類階層とその分類項目に対応するデータ群とが与えられたときに、分類軸候補絞込み手段は、分類項目に対応するデータ量と分類項目間の階層的な距離とが特定の条件を満たすように分類軸候補を作成する。また、指標計算手段は、分類階層における分類項目間の階層的な距離に基づいてその独立性を考慮して、分類軸候補の優先度を決定する。
- [0025] また、本発明によるデータ分類システムは、分類軸候補絞込み手段と、第2指標計算手段とを備える。分類階層とその分類項目に対応するデータ群が与えられたときに、分類軸候補絞込み手段は、分類項目に対応するデータ量と分類項目間の階層的な距離とが特定の条件を満たすように分類軸候補を作成する。また、第2指標計算手段は、分類階層における分類項目間の階層的な距離に基づいてその独立性を考慮し、分類項目の分類階層での深さに基づいてその具体性を考慮し、分類項目に対応するデータ量に基づいて網羅性を考慮し、データの重複度合いに基づいて一意性を考慮して、分類軸候補の優先度を決定する。

- [0026] また、本発明によるデータ分類システムは、分類軸候補絞り込み手段と、第2指標計算手段と、表示手段とを備える。分類階層とその分類項目に対応するデータ群が与えられたときに、分類軸候補絞り込み手段は、分類項目に対応するデータ量と分類項目間の階層的な距離とが特定の条件を満たすように分類軸候補を作成する。また、第2指標計算手段は、分類階層における分類項目間の階層的な距離に基づいてその独立性を考慮し、分類項目の分類階層での深さに基づいてその具体性を考慮し、分類項目に対応するデータ量に基づいて網羅性を考慮し、データの重複度合いに基づいて一意性を考慮して、分類軸候補の優先度を決定する。また、表示手段は、分類軸候補のうち優先度が高いものを分類軸として用いて、データ一覧と分類軸内の分類項目群とを表示する。
- [0027] また、本発明によるデータ分類システムは、多次元分類軸候補絞り込み手段と、多次元分類軸指標計算手段とを備える。分類階層とその分類項目に対応するデータ群が与えられたときに、多次元分類軸候補絞り込み手段は、分類項目に対応するデータ量と分類項目間の階層的な距離の条件とを満たす分類項目に基づいて分類軸を作成し、分類軸の組合せによって多次元分類軸を作成する。また、多次元分類指標計算手段は、各次元の分類軸の分類項目間の階層的な距離を用いて、多次元分類軸候補の優先度を計算する。
- [0028] また、本発明によるデータ分類システムは、多次元分類軸候補絞り込み手段と、第2の多次元分類軸指標計算手段とを備える。分類階層とその分類項目に対応するデータ群が与えられたときに、多次元分類軸候補絞り込み手段は、分類項目に対応するデータ量と分類項目間の階層的な距離の条件とを満たす分類項目に基づいて分類軸を作成し、分類軸の組合せによって多次元分類軸を作成する。また、第2の多次元分類指標計算手段は、各次元の分類軸の分類項目間の階層的な距離だけでなく、各次元の分類軸の分類候補の分類階層での深さ、対応するデータ量、又は対応するデータの重複度合いのうちの少なくともいずれか1つを用いて、多次元分類軸候補の優先度を決定する。
- [0029] また、本発明によるデータ分類システムは、多次元分類軸候補絞り込み手段

と、第2の多次元分類軸指標計算手段と、多次元表示手段とを備える。分類階層とその分類項目に対応するデータ群が与えられたときに、多次元分類軸候補絞り込み手段は、分類項目に対応するデータ量と分類項目間の階層的な距離の条件とを満たす分類項目に基づいて分類軸を作成し、分類軸の組合せによって多次元分類軸を作成する。また、第2の多次元分類指標計算手段は、各次元の分類軸の分類項目間の階層的な距離だけでなく、各次元の分類軸の分類候補の分類階層での深さ、対応するデータ量、又は対応するデータの重複度合いのうちの少なくともいずれか1つを用いて、多次元分類軸候補の優先度を決定する。また、多次元表示手段は、各次元の分類軸とその分類項目とを表示し、分類項目の1つ又は複数を選択することによってデータ群を絞り込み、データ一覧を表示する。

[0030] また、本発明によるデータ分類システムは、データ検索手段と、多次元分類軸候補絞り込み手段と、第2の多次元分類軸指標計算手段と、多次元表示手段とを備える。分類階層とその分類項目に対応するデータ群が与えられたときに、データ検索手段は、データを検索し検索結果データ群を絞り込む。また、多次元分類軸候補絞り込み手段は、分類項目に対応する検索結果データ量と分類項目間の階層的な距離の条件とを満たす分類項目に基づいて分類軸を作成し、分類軸の組合せによって多次元分類軸を作成する。また、第2の多次元分類指標計算手段は、各次元の分類軸の分類項目間の階層的な距離だけでなく、各次元の分類軸の分類候補の分類階層での深さ、対応する検索結果データ量、又は対応する検索結果データの重複度合いのうちの少なくともいずれか1つを用いて、多次元分類軸候補の優先度を決定する。また、多次元表示手段は、各次元の分類軸とその分類項目とを表示し、分類項目の1つ又は複数を選択することによってデータ群を絞り込み、検索結果データ一覧を表示する。

[0031] また、本発明によるデータ分類システムは、データ対応付け手段と、データ検索手段と、多次元分類軸候補絞り込み手段と、第2の多次元分類軸指標計算手段と、多次元表示手段とを備える。分類階層とその分類項目に対応する

データ群が与えられたときに、データ対応付け手段は、分類項目とデータとの対応付けがないデータ、又は対応付けが不十分なデータに対して対応付けを行う。また、データ検索手段は、データを検索し検索結果データ群を絞り込む。また、多次元分類軸候補絞り込み手段は、分類項目に対応する検索結果データ量と分類項目間の階層的な距離の条件とを満たす分類項目に基づいて分類軸を作成し、分類軸の組合せによって多次元分類軸を作成する。また、第2の多次元分類指標計算手段は、各次元の分類軸の分類項目間の階層的な距離だけでなく、各次元の分類軸の分類候補の分類階層での深さ、対応する検索結果データ量、又は対応する検索結果データの重複度合いのうちの少なくともいずれか1つを用いて、多次元分類軸候補の優先度を決定する。また、多次元表示手段は、各次元の分類軸とその分類項目とを表示し、分類項目の1つ又は複数を選択することによってデータ群を絞り込み、検索結果データ一覧を表示する。

[0032] (第1の実施形態)

次に、本発明の第1の実施形態について図面を参照して説明する。図1は、本発明に係るデータ分類システムの構成の一例を示すブロック図である。図1に示すように、データ分類システムは、入力手段101と、分類軸候補絞り込み手段102と、指標計算手段103と、出力手段104と、分類階層蓄積部201と、基準項目蓄積部202と、データ蓄積部203とを含む。

[0033] 本実施形態において、データ分類システムは、具体的には、プログラムに従って動作するパーソナルコンピュータ等の情報処理装置によって実現される。なお、データ分類システムは、1つの情報処理装置にかぎらず、例えば、複数の情報処理装置を用いて実現されてもよい。

[0034] 分類階層蓄積部201は、具体的には、磁気ディスク装置や光ディスク装置等のデータベース装置によって実現される。分類階層蓄積部201は、分類階層とその階層関係とを示す情報を予め蓄積する。図2は、分類階層蓄積部201が記憶する情報の一例を示す図である。図2に示すように、分類階層蓄積部201は、各レコードを親子関係とし、親分類項目と子分類項目群

とを対応付けて含むテーブルを記憶する。

- [0035] なお、図2に示すテーブルは、例えば、データ整理のために、システム管理者等によって予め作成され、分類階層蓄積部201に蓄積される。また、例えば、データ分類システムは、文書データベース等に蓄積するデータに基づいて分類項目を自動抽出してテーブルを作成し、分類階層蓄積部201に記憶させるようにしてもよい。
- [0036] 図2に示す例において、親分類項目は、親子関係の親となる分類項目を表す。また、子分類項目群は、「、」を区切り文字として子の分類項目群を示す。なお、図中の「・・・」は、記載の省略を表している。
- [0037] なお、図2に示す記憶方法は一例であり、分類階層蓄積部201は、例えば、子の分類項目をレコード毎に分割して記憶してもよいし、階層構造データとして記憶してもよい。本実施形態では、表現を簡単にするために、図2に示すように表現している。
- [0038] 基準項目蓄積部202は、具体的には、磁気ディスク装置や光ディスク装置等のデータベース装置によって実現される。基準項目蓄積部202は、分類軸の基準となる分類項目名を予め蓄積する。図3は、基準項目蓄積部202が記憶する情報の一例を示す図である。図3に示すように、基準項目蓄積部202は、基準項目のリストを記憶する。
- [0039] なお、基準項目蓄積部202に記憶させる基準項目は、分類階層蓄積部201が記憶する分類項目から予め選択する。なお、この場合、例えば、システム管理者等によって予め選択され基準項目蓄積部202に蓄積されてもよいし、データ分類システムが分類階層蓄積部201から自動抽出して基準項目蓄積部202に記憶させてもよい。
- [0040] データ蓄積部203は、具体的には、磁気ディスク装置や光ディスク装置等のデータベース装置によって実現される。データ蓄積部203は、分類項目とデータの対応関係とを予め蓄積する。また、データ蓄積部203は、データのその他の属性である作成日等の属性情報やコンテンツ実体等を蓄積してもよい。図4は、データ蓄積部203が記憶する情報の一例を示す図であ

る。図4に示すように、データ蓄積部203は、データID、コンテンツ、及び対応分類項目を対応づけて含むレコードを記憶するデータベースである。

- [0041] なお、図4に示す例において、図中の「・・・」は、記載の省略を表している。また、図4に示す例において、データIDは、データを識別するための識別子である。また、対応分類項目は、データIDで識別されるデータに対応する分類項目を、「、」区切りを用いて示している。なお、図4に示す記憶方法は一例であり、データ蓄積部203は、コンテンツ以外の作成日等の属性を示す属性情報を含むレコードを記憶してもよい。
- [0042] なお、上記に示すデータは、例えば、システム管理者等によって予め収集され、データ蓄積部203に蓄積される。また、例えば、データ分類システムは、ネットワークを介して文書データベース等に蓄積されるデータを収集し、データ蓄積部203に記憶させるようにしてもよい。
- [0043] 入力手段101は、具体的には、プログラムに従って動作する情報処理装置のCPU、キーボードやマウス等の入力デバイス、及び入出力インタフェース部によって実現される。入力手段101は、ユーザ操作に従って、各種情報を入力する機能を備える。又は、入力手段101は、他のシステムからの入力情報を受け取る（入力する）機能を備える。なお、本実施形態では、データ分類システムにおいて、入力手段101は、ユーザ操作に従って、分類項目数Nを受け取る（入力する）ものとする。
- [0044] 分類軸候補絞り込み手段1021は、具体的には、プログラムに従って動作する情報処理装置のCPUによって実現される。分類軸候補絞り込み手段1021は、分類階層蓄積部201が記憶する分類項目と、基準項目蓄積部202が記憶する基準項目とに基づいて、分類項目数分の分類項目を組合せて分類軸候補を作成する機能を備える。
- [0045] 本実施形態では、分類軸候補絞り込み手段1021は、分類軸候補を作成する際に、基準項目の全ての子孫の分類項目を組合せるのではなく、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づき作成する分類

軸の絞込みを行う。そのようにすることにより、分類軸候補絞込み手段 1021 は、優先度を計算する対象となる分類軸を減らすことができ、計算を高速化できる。

[0046] なお、「分類項目に対応するデータの量」とは、分類項目に対応づけられているデータ数である。また、「分類項目に対応するデータ」とは、分類項目に直接対応しているデータ、又は分類項目とその子孫の分類項目に対応しているデータのことである。なお、本実施の形態では、「分類項目に対応するデータ」は、分類項目とその子孫の分類項目に対応しているデータであるとする。また、データの量が多い場合、網羅性が高いと考えられる。そのため、このような分類項目をもつ分類軸は、データ群を良く表しているため、データの概要把握に役立つと考えられる。

[0047] また、「分類項目間の階層的な距離」とは、共通する祖先までの最短のパス長や最長のパス長、又は共通の子孫までの最短のパス長や最長のパス長等である。この場合、分類項目間の階層的な距離が遠いほど、意味的に独立している分類項目であることがわかる。

[0048] 分類軸候補絞込み手段 1021 は、以上に基づいた分類項目の選択を行う場合、分類項目に対応するデータ量が一定数以上で、データ量順で上位から一定割合の分類項目の条件を満たし、かつ分類項目間の階層的な距離が特定の一定数以上若しくは一定の範囲内にある分類項目、又はその子孫の分類項目を含む分類軸を選択する。

[0049] 指標計算手段 103 は、具体的には、プログラムに従って動作する情報処理装置の CPU によって実現される。指標計算手段 103 は、分類軸候補絞込み手段 1021 から分類軸候補を受け取り（入力し）、分類階層蓄積部 201 が記憶する情報と、データ蓄積部 203 が記憶する情報とを参照して、分類軸候補の優先度を計算する機能を備える。

[0050] 指標計算手段 103 は、優先度を、分類項目の分類階層での階層的な距離に基づいて計算する。ここで、「分類項目間の階層的な距離」とは、共通する祖先までの最短のパス長や最長のパス長、又は共通の子孫までの最短のパス

ス長や最長のパス長等である。また、指標計算手段103は、優先度として、分類軸の各分類項目間の階層的な距離の平均や、最大、最小等の値を求める。

[0051] 本実施形態では、指標計算手段103は、分類項目間の階層的な距離として、共通する祖先までの最短パス長を求め、優先度として、階層的な距離の平均の値を求めるものとする。そのようにすることで、階層的な距離が長い方の分類軸の方が、意味的に独立しているということができる。

[0052] 出力手段104は、具体的には、プログラムに従って動作する情報処理装置のCPU、ディスプレイ装置等の表示装置、及び入出力インタフェース部によって実現される。出力手段104は、指標計算手段103から、分類軸候補と優先度とのペアを受け取る（入力する）機能を備える。また、出力手段104は、入力した分類軸候補と優先度とのペアを、分類項目と対応するデータとともに出力する機能を備える。なお、出力手段104が出力する分類軸数は予め決められていてもよい。また、出力手段104は、出力方法として、ディスプレイ装置等の表示装置に表示してもよいし、記憶媒体（例えば、CD-ROM）や別なプログラムにファイル出力するようにしてもよい。

[0053] なお、本実施形態において、データ分類システムを実現する情報処理装置の記憶装置（図示せず）は、データ分類を行うための各種プログラムを記憶している。例えば、データ分類システムを実現する情報処理装置の記憶装置は、コンピュータに、基準項目の子孫の分類項目から複数の分類項目を選択して分類軸候補とし、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む分類軸絞り込み処理と、絞り込んだ分類軸候補について、当該分類軸候補を表示する優先度を計算する優先度計算処理とを実行させる、データ分類プログラムを記憶している。

[0054] 次に、動作について説明する。図5は、データ分類システムが行うデータ分類処理の一例を示す流れ図である。

- [0055] まず、データ分類システムの入力手段101は、ユーザの操作に従って、分類項目数Nを受け取る（ステップS1）。例えば、入力手段101は、分類項目数としてN=3を入力する。次に、分類軸候補絞り込み手段1021は、基準項目蓄積部202から、分類軸の基準項目を取得（抽出）する（ステップS2）。例えば、図3に示す例において、分類軸候補絞り込み手段1021は、基準項目蓄積部202から、「社会」、「自然」及び「文化」等の情報を抽出する。
- [0056] 次に、分類軸候補絞り込み手段1021は、分類項目に対応するデータ量と、分類項目間の階層的な距離都に基づき作成する分類軸の絞り込みを行う（ステップS31）。
- [0057] 次に、分類軸候補絞り込み手段1021は、分類階層蓄積部201が記憶する情報を参照し、分類項目間の階層的な距離に基づいて分類項目を組合せて分類軸を作成する（ステップS32）。本実施形態では、分類軸候補絞り込み手段1021は、分類項目間の階層的な距離として、分類項目間の共通する先祖の分類項目までの長さを用いる。また、分類軸候補絞り込み手段1021は、ステップS31で絞り込んだ分類項目とその子孫の分類項目とに基づいて、分類項目間の階層的な距離が一定数以上である分類項目対を1つ以上含む分類軸を作成する。但し、分類軸中には、先祖と子孫との関係になる分類項目同士はないものとする。
- [0058] 例えば、分類軸候補絞り込み手段1021は、先に絞り込んだ分類項目「生活」、「医療」、「家族」、「家庭」、「医学」、「移植」及び「健康」の中で、分類項目間の階層的な距離が3以上である分類項目を用いて分類軸を作成する。本実施形態では、例えば、分類項目「生活」と階層的な距離が3以上である分類項目に「医学」がある。よって、分類軸候補絞り込み手段1021は、分類項目「生活」と「医学」とその他の先に絞り込んだ分類項目のうちいずれかを用いて分類軸を作成する。この場合、分類軸候補絞り込み手段1021は、この「生活」及び「医学」の2つの分類項目と、先祖と子孫との関係にある分類項目ではない分類項目の中から処理対象の分類項目を選択

する。本例では、分類軸候補絞込み手段 1021 は、分類軸（社会：生活、医学、移植）を作成する。

[0059] また、同様に、分類項目「医療」に対して距離が 3 以上である分類項目は、「家庭」、「家族」及び「健康」である。従って、分類軸候補絞込み手段 1021 は、分類軸（社会：医療、家庭、家族）、（社会：医療、家庭、健康）、及び（社会：医療、家族、健康）を作成する。

[0060] なお、本例では、分類軸候補絞込み手段 1021 は、分類軸内の分類項目間のいずれか 1 つが階層的な距離の条件を満たすように分類項目を作成しているが、分類項目内の全ての分類項目間が階層的な距離の条件を満たすように分類項目を作成してもよい。

[0061] 以上のように、ステップ S 32 の処理が実行されることにより、意味的な独立性を保っていない分類項目が除外される。上記に示す例では、分類軸候補絞込み手段 1021 は、分析軸（社会：家族、家庭、健康）の分類軸を除外するように分類軸を作成する。

[0062] 次に、指標計算手段 103 は、分類軸候補絞込み手段 1021 から分類軸候補を取得（入力）し、分類階層蓄積部 201 が記憶する情報を参照して、各分類軸の優先度を計算する（ステップ S 4）。なお、優先度の計算を行う場合、本実施形態では、指標計算手段 103 は、分類項目の意味的な独立性を図るため、分類項目間の階層的な距離の平均値を計算する。ここで、本実施形態では、指標計算手段 103 は、分類項目間の階層的な距離として、分類項目間で共通する先祖分類項目までの最短パスを求めるものとする。

[0063] 具体的には、指標計算手段 103 は、優先度を、以下の式（1）を用いて求める。

[0064] 優先度（X：C）

$$= 1 / \text{Max} (X) \times 1 / (2 \times \text{組合せ数}) \times \sum \text{ComDist} (c_i, c_j)$$

・・・式（1）

[0065] ここで、式（1）において、X は基準項目を示し、C は分類項目群を示し

ている。また、 c_i 、 c_j は、それぞれ分類項目群内の i 番目の分類項目及び j 番目の分類項目を示している。また、 $Max(X)$ は、基準項目 X の子孫の分類項目で最も深い分類項目までの深さを示している。また、 $ComDist(c_i, c_j)$ は、分類項目 c_i と c_j との距離を示している。更に、組合せ数は、分類項目群 C から2つを選択する組合せ回数である。なお、式(1)において、平均値を $Max(X)$ で除算している理由は、基準項目によって深さが異なるためである。

[0066] 指標計算手段103は、式(1)を用いて、分類軸(社会:家庭、健康、医療)の優先度を、次の式(2)のように求める。

[0067] 優先度(社会:家庭、健康、医療)

$$= 1/2 \times 1/6 \times (ComDist(\text{家庭}, \text{健康}) + ComDist(\text{家庭}, \text{医療}) + ComDist(\text{健康}, \text{医療}))$$
 ...式(2)

[0068] 指標計算手段103は、分類項目「家族」と「健康」との共通先祖が「生活」であるので、 $ComDist(\text{家庭}, \text{健康}) = 2$ と求める。また、指標計算手段103は、「家庭」と「医療」との共通先祖が「社会」であるので、 $ComDist(\text{家庭}, \text{医療}) = 3$ 及び $ComDist(\text{健康}, \text{医療}) = 3$ と求める。従って、指標計算手段103は、分類軸(社会:家庭、健康、医療)の優先度を、次の式(3)のように求める。

[0069] 優先度(社会:家庭、健康、医療)

$$= 1/2 \times 1/6 \times (ComDist(\text{家庭}, \text{健康}) + ComDist(\text{家庭}, \text{医療}) + ComDist(\text{健康}, \text{医療}))$$

$$= 1/2 \times 1/6 \times (2 + 3 + 3)$$

$$= 0.67$$
 ...式(3)

[0070] 次に、出力手段104は、指標計算手段103の計算結果に基づいて、分類軸と、優先度と対応するデータとを出力する(ステップS5)。図6(A)及び図6(B)は、出力手段104が出力する情報の例を示す図である。

図6 (A) 及び図6 (B) に示す例では、出力手段104は、2つのテーブルを出力している。

[0071] 例えば、出力手段104は、図6 (A) に示すように、分類軸ID、基準項目、分類項目群及びスコアを対応付けたレコードを含むテーブルを出力する。なお、図6 (A) において、1つの行で示されるレコードが、それぞれ1つの分類軸を表している。分類軸IDは、分類軸の候補を識別するためのIDである。分類項目群は、「、」で分類項目毎に分けられた複数の分類項目を含む。

[0072] また、例えば、出力手段104は、図6 (B) に示すように、分類軸ID、分類項目及びデータID群を対応付けたレコードを含むテーブルを出力する。なお、図6 (B) において、1つのレコードが、それぞれ各分類軸の分類項目に対応している。データID群は、データIDを「、」で区切って複数のデータIDを含む。また、図中の「・・・」は、記載の省略を表している。

[0073] なお、図6 (A) 及び図6 (B) に示す出力方法は一例であり、出力手段104は、例えば、2つのテーブルを1つにまとめたテーブルを出力してもよいし、各データの属性情報を含んだテーブルを新たに追加して出力してもよい。

[0074] 上記のような構成とすることで、分類項目の意味的な独立性を用いて分類軸を選択することができる。また、そのようにすることにより、ユーザがわかりやすい分類軸を選択することができる。

[0075] 以上に説明したように、本実施形態によれば、基準項目の子孫の分類項目から複数の分類項目を選択して分類軸候補とする。また、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む。そして、絞り込んだ分類軸候補について、その分類軸候補を表示する優先度を計算する。そのため、絞り込んだ分類軸候補のみについて優先度計算を行うようにすることによって、分類軸の優先度の計算時間を短縮することができる。従って、分類階層とその分類項目とに対応する

データ群が与えられたときに、分類軸の優先度の計算時間を短縮することができる。

[0076] また、本実施形態によれば、分類階層が非常に大きな場合であっても、分類項目の階層的な距離を用いて効果的に分類項目を絞り込むことによって、高速に分類軸の優先度を計算することができる。

[0077] (第2の実施形態)

次に、本発明の第2の実施形態について説明する。図7は、第2の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。図7に示すように、本実施形態では、データ分類システムが、図1に示した指標計算手段103に代えて、第2の指標計算手段1031を含む点で、第1の実施形態と異なる。

[0078] 第2の指標計算手段1031は、具体的には、プログラムに従って動作する情報処理装置のCPUによって実現される。第2の指標計算手段1031は、分類軸候補絞り込み手段1021から分類軸候補群を受け取り(入力)、分類軸の優先度を計算する機能を備える。本実施形態では、第2の指標計算手段1031は、分類項目間の階層的な距離だけでなく、分類項目の階層的な深さ、分類項目に対応するデータの量、分類項目に対応するデータの重複度合い、又はこれらの組合せに基づいて、優先度を計算する。

[0079] 「分類項目の階層的な距離」とは、分類階層での分類項目間の距離のことであり、第1の実施形態で示したものと同様である。なお、本実施形態では、この「分類項目の階層的な距離」という指標を、以下、独立性の指標という。

[0080] また、「分類項目の階層的な深さ」とは、基準項目又は分類階層のルートとなる分類項目から分類項目までの最短又は最長のパス長である。階層的な深さが深い場合、分類項目名は、意味的により具体的なものであると考えられる。

[0081] 例えば、図8に示すような分類階層を例に考える。図8に示す分類階層の場合、最も深い分類項目「家族法」や「親族」は、分類項目「社会」よりも

具体的な分類項目であるといえる。具体性のある分類項目は、ユーザにとって理解しやすく概要把握に役立つと考えられる。そこで、本実施形態では、第2の指標計算手段1031は、分類項目の階層的な深さを基準項目から分類項目までの最短パス長とし、分類項目の階層的な深さが深いほど大きな値となるように優先度を計算する。なお、本実施形態では、この「分類項目の階層的な深さ」という指標を、以下、具体性の指標という。

[0082] また、「分類項目に対応するデータの量」とは、分類項目に対応づけられているデータ数である。また、「分類項目に対応するデータ」とは、分類項目に直接対応しているデータ、又は分類項目とその子孫の分類項目に対応しているデータのことである。本実施形態では、第2の指標計算手段1031は、「分類項目に対応するデータ」として、分類項目とその子孫の分類項目に対応しているデータを用いるものとする。この場合、データの量が多い場合、網羅性が高いと考えられる。そのため、網羅性が高い分類項目を用いて作成した分類軸は、データ蓄積部203が蓄積するデータを良く表しているため、概要把握に役立つと考えられる。本実施形態では、第2の指標計算手段1031は、データ量が大きいほど大きな値となるように優先度を計算する。なお、本実施形態では、この「分類項目に対応するデータの量」という指標を、以下、網羅性の指標という。

[0083] また、「分類項目に対応するデータの重複度合い」とは、分類軸内の各分類項目でデータがどの程度同じであることを示す値である。重複度合いが少ない場合、データ蓄積部203が蓄積するデータには一意性があり、概要把握しやすい。もし、一意性がなく、重複するデータばかりであれば、分類軸の各分類項目のデータ数を表示する際に、そのデータ数が重複ばかりであり、良い分類軸とはいえない。

[0084] 例えば、データ蓄積部203が図4に示すデータを蓄積している場合、分類項目「家族」に対応するデータは、データIDが「d1」、「d2」及び「d3」であるデータである。また、分類項目「家庭」に対応するデータも、同じくデータIDが「d1」、「d2」及び「d3」であるデータである

。この場合、表示方法として、分類軸の各分類項目のデータ数のみを表示した場合、分類項目「家族」と分類項目「家庭」とは、データ数が3であるが、実際の中身は同じであるので、この分類軸にはあまり情報量がない。この場合、第2の指標計算手段1031は、重複度合いが高いほど優先度が低い値となるように優先度を計算する。重複度合いとして、各分類項目に対応するデータ数の合計を重複なしのデータ数で除算した値を用いてもよく、データの出現確率から情報量（エントロピー）を計算してもよい。なお、本実施形態では、この「分類項目に対応するデータの重複度合い」という指標を、以下、一意性の指標という。

[0085] 本実施形態では、第2の指標計算手段1031は、最終的には、上記に示した各指標を用いて総合的に優先度を計算する。

[0086] なお、本実施形態において、第2の指標計算手段1031以外の構成要素の機能は、第1の実施形態で示したそれらの機能と同様である。

[0087] 次に、動作について説明する。図9は、第2の実施形態におけるデータ分類システムが行うデータ分類処理の一例を示す流れ図である。図9に示すように、本実施形態では、分類項目の階層的な距離に基づいて分類軸候補の優先度を計算するステップS4の処理に代えて、計算用テーブルを作成するステップS41、及び分類軸候補の優先度を計算するステップS42の処理を実行する点で、第1の実施形態と異なる。以下、主として、第1の実施形態と異なる処理について説明する。

[0088] まず、第1の実施形態で示した処理と同様に、データ分類システムの入力手段101は、ユーザの操作に従って、分類項目数Nを受け取る（ステップS1）。例えば、入力手段101は、分類項目数としてN=3を入力する。次に、第1の実施形態で示した処理と同様に、分類軸候補絞り込み手段1021は、基準項目蓄積部202から、分類軸の基準項目を取得（抽出）する（ステップS2）。例えば、図3に示す例において、分類軸候補絞り込み手段1021は、基準項目蓄積部202から、「社会」、「自然」及び「文化」等の情報を抽出する。

- [0089] 次に、第1の実施形態で示した処理と同様に、分類軸候補絞り込み手段1021は、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づき作成する分類軸の絞り込みを行う（ステップS31）。
- [0090] 次に、第1の実施形態で示した処理と同様に、分類軸候補絞り込み手段1021は、分類階層蓄積部201が記憶する情報とデータ蓄積部203が記憶する情報とを参照し、基準項目の子孫で特定個数以上のデータと対応づけられている分類項目を取得（抽出）する。そして、分類軸候補絞り込み手段1021は、抽出した分類項目数分の組合せを用いて分類軸候補を作成する（ステップS32）。但し、分類軸候補絞り込み手段1021は、先祖子孫関係の分類項目を含む場合には分類軸候補としないようにする。
- [0091] 次に、第2の指標計算手段1031は、分類軸候補絞り込み手段1021から分類軸候補を取得（入力）し、分類階層蓄積部201が記憶する情報とデータ蓄積部203が記憶する情報とを参照し、計算用テーブルを作成する（ステップS41）。ここで、計算用テーブルとは、指標計算のために作成される一時的なテーブルで、第2の指標計算手段1031は、分類項目別データ数表とデータ別分類項目数表との2つのテーブルを作成する。
- [0092] 分類項目別データ数表とは、各分類項目に対応しているデータ数を集計したテーブルであり、分類項目、データ数及び分類項目の深さを対応付けたレコードを含む。また、データ別分類項目数表とは、各データに対応している分類軸内の分類項目の数を集計したテーブルであり、データID及び分類項目数を含む。なお、第2の指標計算手段1031は、作成した計算用の一時的なテーブルを、メモリ上に展開しておくことが望ましい。
- [0093] 図10（A）及び図10（B）は、分類項目別データ数表とデータ別分類項目数表との例を示す図である。このうち、図10（A）は、分類項目別データ数表の例を示している。また、図10（B）は、データ別分類項目数表の例を示している。なお、図10（A）及び図10（B）に示す例では、第2の指標計算手段1031が、分類軸（社会：家族、外交、医療）、図2に示した分類階層蓄積部201が記憶する情報、及び図4に示したデータ蓄積

部 203 が記憶する情報に基づいて、分類項目別データ数表とデータ別分類項目数表とを作成した場合が示されている。

[0094] 図 10 (A) に示す例では、分類項目別データ数表は、分類項目、データ数、及び分類項目の深さを対応付けた各レコードを含む表である。本実施形態では、前述したように、分類項目の深さは、基準項目から各分類項目までのパス長を表している。例えば、分類項目「家族」は、図 4 に示したデータ蓄積部 203 が記憶する情報を参照すると「d1」、「d2」及び「d3」であるので、図 10 (A) に示すように、データ数が 3 である。また、図 2 に示した分類蓄積部 201 が記憶する情報を参照すると、基準項目「社会」から分類項目「家庭」までは、「生活」を介して深さが 2 である。

[0095] また、本実施形態では、前述したように、分類項目とデータの対応は、分類項目とその子孫の分類項目とに直接対応しているデータとしている。例えば、分類項目「医療」に直接対応しているデータはないが、その子孫の分類項目に直接対応するデータを見る。ここで、子孫の分類項目「医学」又は「健康」に対応するデータ ID は、「d2」、「d4」及び「d6」である。よって、図 10 (A) に示すように、データ数を 3 としている。

[0096] データ別分類項目数表は、データ ID と分類項目数とを対応付けたレコードを含む表である。本実施形態では、データ別分類項目数表は、図 10 (B) に示すように、各データ ID に対して、分類軸（社会：家族、外交、医療）と対応する分類項目数を含む。例えば、データ ID 「d1」は、図 4 に示したデータ蓄積部 203 が記憶する情報を参照すると、分類軸内の分類項目「家族」に対応しているので、図 10 (B) に示すように、分類項目数が 1 である。また、データ ID 「d6」は、分類項目「医療」の子孫の分類項目に対応しているので、図 10 (B) に示すように、分類項目数が 1 である。

[0097] 次に、第 2 の指標計算手段 1031 は、計算用テーブルを用いて、分類軸の優先度を計算する（ステップ S42）。本実施形態では、第 2 の指標計算手段 1031 は、前述したように、独立性、具体性、網羅性及び一意性の指標の値を計算し、これら指標の重み付き線形和を求めることによって、式（

4) を用いて総合的な優先度を計算する。

[0098] 優先度 (X : C)

$$= W1 \times \text{独立性 (X : C)} + W2 \times \text{具体性 (X : C)} + W3 \times \text{網羅性 (X : C)} + W4 \times \text{一意性 (X : C)}$$

・・・式 (4)

[0099] ここで、式 (4) において、X は基準項目であり、C は分類項目群である。また、W1、W2、W3 及び W4 は、それぞれの指標の重み係数である。なお、これらの重み係数は、予めシステムがもっていてもよいし (例えば、予めメモリ等の記憶部に設定値として記憶していてもよい)、ユーザによって設定されてもよい。本実施形態では、これらの重み係数は、システムに予め設定されているものとする。

[0100] なお、本実施形態では、独立性の指標値は、第1の実施形態と同様であり、第2の指標計算手段 1031 は、式 (5) を用いて求める。

[0101] 独立性 (X : C)

$$= 1 / \text{Max (X)} \times 1 / (2 \times \text{組合せ数}) \times \sum \text{ComDist (c}_i, \text{c}_j)$$

・・・式 (5)

[0102] ここで、式 (5) において、X、C、Max (X)、組合せ数、及び ComDist (c_i, c_j) は、第1の実施形態で示したものと同様である。

[0103] また、第2の指標計算手段 1031 は、具体性の指標値を、以下の式 (6) を用いて計算する。ここでは、具体性の指標値は、基準項目から分類軸の各分類項目までのパス長の平均の値である。

[0104] 具体性 (X : C)

$$= 1 / \text{Max (X)} \times 1 / N \times \sum \text{Depth (X, c}_i)$$

・・・式 (6)

[0105] 式 (6) において、Max (X) は、基準項目 X の子孫の分類項目で最大の深さである。また、N は、入力手段 101 から与えられた (入力した) 分類項目数である。また、Depth (X, c_i) は、基準項目 X から分類項

目 c_i までの最短パス長である。ここで、パス長の平均を $Max(x)$ で除算しているのは、基準項目によって、子孫の分類項目の深さが異なるためである。第2の指標計算手段1031は、この具体性の指標値を、分類項目別データ数表を用いて、以下の式(7)のように計算することができる。

[0106] 具体性 (X : C)

$$= 1 / Max(X) \times 1 / N \times \sum Depth(X, c_i)$$

$$= 1 / Max(X) \times 1 / N \times \sum (\text{分類項目別データ数表の分類項目の深さ})$$

・・・式(7)

[0107] 式(7)に示すように、分類項目の深さが深いほど、具体性の指標の値が大きくなるのがわかる。

[0108] また、第2の指標計算手段1031は、網羅性の指標値を、以下の式(8)を用いて計算する。ここで、網羅性の指標値は、分類項目に対応するデータの全データに対するカバー率である。

[0109] 網羅性 (X : C)

$$= 1 / DataNum \times |U\text{データ}(c_i)|$$

・・・式(8)

[0110] 式(8)において、 $DataNum$ は、分類対象データの数である。また、 $U\text{データ}(c_i)$ は、分類項目 c_i に対応するデータ集合である。また、「 $U\text{データ}(c_i)$ 」は、分類軸内の分類項目 c_1 から c_N までのデータの和集合である。また、「 $|U\text{データ}(c_i)|$ 」は、分類軸内の分類項目 c_1 から c_N までのデータ集合の要素数である。つまり、「 $|U\text{データ}(c_i)|$ 」は、分類項目に対応するデータ数を表す。第2の指標計算手段1031は、この網羅性の指標値を、先に作成したデータ別分類項目表を用いて、次の式(9)のように計算することができる。

[0111] 網羅性 (X : C)

$$= 1 / DataNum \times |U\text{データ}(c_i)|$$

$$= 1 / DataNum \times RecNum (\text{データ別項目分類表, 分類項目数})$$

0)

・・・式(9)

[0112] 式(9)において、RecNum(データ別分類項目数表, 分類項目数>0)は、データ別分類項目数表の分類項目数が0よりも大きいレコード数である。このRecNum(データ別分類項目数表, 分類項目数>0)は、分類項目に対応しているデータ数に等しい。よって、上記の式(9)のように書き換えることができる。

[0113] また、第2の指標計算手段1031は、一意性の指標値を、以下の式(10)を用いて計算する。ここで、データの重複度合いは、各分類項目に対応するデータ数の合計を重複なしのデータ数の合計で除算した値とする。なお、一意性の指標値は、このデータの重複度合いの逆数で表す。

[0114] 一意性(X:C)

$$= 1 / (1 / |Uデータ(c_i)| \times \sum CatNum(c_i))$$

・・・式(10)

[0115] ここで、式(10)において、|Uデータ(c_i)|は、分類項目に対応する重複なしのデータ数である。また、CatNum(c_i)は、c_iに対応するデータ数である。また、 $\sum CatNum(c_i)$ は、分類軸内の分類項目c_1からc_Nそれぞれに対応するデータ数の合計を表す。第2の指標計算手段1031は、この一意性の指標値を、先に作成した分類項目別データ数表を用いて、次の式(11)のように計算することができる。

[0116] 一意性(X:C)

$$= 1 / (1 / |Uデータ(c_i)| \times \sum CatNum(c_i))$$

$$= 1 / (RecNum(データ別項目分類表, 分類項目数>0) \times \sum (データ別項目分類表のデータ数))$$

・・・式(11)

[0117] 例えば、第2の指標計算手段1031は、分類軸(社会:家族、外交、医療)の場合、図10(A)及び図10(B)に示すテーブルと、図2に示す分類階層蓄積部201が記憶する情報とを参照し、上記の各指標値を、以下

の式(12)～式(15)のように計算する。

[0118] 独立性 (X : C)

$$= 1 / \text{Max}(X) \times 1 / (2 \times \text{組合せ数}) \times \sum \text{ComDist}(C1, C2)$$

$$= 1 / \text{Max}(\text{社会}) \times 1 / (2 \times 3) \times (\text{ComDist}(\text{家族}, \text{外交}) + \text{ComDist}(\text{家族}, \text{医療}) + \text{ComDist}(\text{外交}, \text{医療}))$$

$$= 1 / 2 \times 1 / 6 \times (4 + 4 + 4)$$

$$= 1$$

・・・式(12)

[0119] 具体性 (X : C)

$$= 1 / \text{Max}(X) \times 1 / N \times \sum (\text{分類項目別データ数表の分類項目の深さ})$$

$$= 1 / \text{Max}(\text{社会}) \times 1 / 3 \times (2 + 2 + 1)$$

$$= 1 / 2 \times 1 / 3 \times (2 + 2 + 1)$$

$$= 0.833$$

・・・式(13)

[0120] 網羅性 (X : C)

$$= 1 / \text{DataNum} \times \text{RecNum} (\text{データ別項目分類表, 分類項目数} > 0)$$

$$= 1 / 6 \times 6$$

$$= 1$$

・・・式(14)

[0121] 一意性 (X : C)

$$= 1 / (\text{RecNum} (\text{データ別項目分類表, 分類項目数} > 0) \times \sum (\text{データ別項目分類表のデータ数}))$$

$$= 1 / (1 / 6 \times (3 + 2 + 3))$$

$$= 6 / 8$$

$$= 0.75$$

・・・式（15）

[0122] ここで、重み係数を全て0.25とすると、第2の指標計算手段1031は、次の式（16）を用いて優先度を求めることができる。

[0123] 優先度（X：C）

$$=W1 \times \text{独立性（X：C）} + W2 \times \text{具体性（X：C）} + W3 \times \text{網羅性（X：C）} + W4 \times \text{一意性（X：C）}$$

$$=0.25 \times 1 + 0.25 \times 0.833 + 0.25 \times 1 + 0.25 \times 0.75$$

$$=0.895$$

$$=0.90$$

・・・式（16）

[0124] 次に、第1の実施形態で示した処理と同様に、出力手段104は、分類軸、優先度及びデータを出力する（ステップS5）。

[0125] 以上に説明したように、本実施形態によれば、分類項目間の階層的な距離だけでなく、分類項目の階層的な深さ、分類項目に対応するデータの量、分類項目に対応するデータの重複度合い、又はこれらの組合せに基づいて、優先度を計算する。そのため、分類項目間の階層的な距離に加えて、分類項目の階層的な深さや、分類項目に対応するデータの量、分類項目に対応するデータの重複度合いを考慮して、より効果的に分類軸の優先度の計算時間を短縮することができる。

[0126] （第3の実施形態）

次に、本発明の第3の実施形態について説明する。図11は、第3の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。図11に示すように、本実施形態では、データ分類システムが、図7に示した構成要素に加えて、表示手段105を含む点で、第2の実施形態と異なる。

[0127] 表示手段105は、具体的には、プログラムに従って動作する情報処理装置のCPU及びディスプレイ装置等の表示装置によって実現される。表示手段105は、第2の指標計算手段1031が求めた分類軸、優先度及びデー

タをディスプレイ装置等の表示装置に出力（表示）する機能を備える。例えば、表示手段105は、分類軸の各分類項目と対応するデータ数、及びデータとその属性を、一覧形式又は表形式で出力（表示）する。

[0128] まず、表示手段105が一覧形式で情報を表示する場合を説明する。図12は、表示手段105が一覧形式で表示する情報の例を示す図である。図12に示すように、表示手段105は、分類軸、分類軸候補一覧、及びデータ一覧のそれぞれの表示部分を含む表示画面を表示する。表示手段105は、分類軸には、優先度が最も高かった分類軸、又は分類軸候補一覧から選択した分析軸を表示する。また、表示手段105は、基準項目と分類項目とを表示し、その後ろにデータ件数を表示する。

[0129] 図12に示す例では、分類軸（社会：家族、健康、移植）を用いる場合が示されている。この場合、表示手段105は、それぞれの項目と対応するデータ件数をデータ蓄積部203から取得（抽出）し表示する。

[0130] また、図12に示す例では、表示手段105が表示する表示画面には、項目「その他」の欄が新たに加えられている。ここで、「社会」の下の「その他」の欄は、基準項目「社会」に関連しているが、分類項目「家族」や「健康」、「移植」ではないことを示す分類項目である。また、基準項目「社会」と並列に並ぶ「その他」の欄は、基準項目「社会」とは関係のない分類項目を示す。なお、表示手段105は、これら「その他」の欄のデータ数を、データ蓄積部203を参照することによって得ることができる。また、本実施形態では、データ件数には子孫の分類項目に対応するデータ件数も含まれる。

[0131] また、図12に示す例では、表示手段105は、分類軸候補一覧を、優先度を計算した分類軸を優先度順に表示する。表示手段105は、これらの分類軸を、第2の指標計算手段1031が計算した結果から得ることができる。

[0132] また、図12に示す例では、表示手段105は、データ一覧を、各データを一覧で表示する。この場合、表示手段105は、各データを、データID

、コンテンツ、及び対応する分類項目を対応付けて表示する。なお、表示手段105は、これらの情報を、データ蓄積部203が記憶する情報を参照することで得ることができる。

[0133] なお、本実施形態において、表示手段105以外の構成要素の機能は、第2の実施形態で示したそれらの機能と同様である。

[0134] 次に、一覧形式で情報を表示する場合の表示手段105の動作について説明する。まず、表示手段105は、初期表示として、優先度の最も高い分類軸を分類軸表示部分に表示する。また、表示手段105は、その他の分類軸候補を分類軸候補一覧に優先度順に表示する。また、表示手段105は、データ一覧には、データ蓄積部203が蓄積する全てのデータを表示する。

[0135] 次に、表示手段105は、分類軸表示部分に表示する分類軸の中で、いずれかの分類項目又は基準項目が選択操作された場合、これに対応するデータをデータ一覧に表示する。また、表示手段105は、分類軸候補一覧表示部分に表示する分類軸候補の中のいずれかの分類軸候補が選択操作されると、分類軸表示部分の内容をこの選択した分類軸で置き換えて表示する。

[0136] 次に、表示手段105が表形式で情報を表示する場合を説明する。図13は、表示手段105が表形式で表示する情報の例を示す図である。図13に示すように、表示手段105は、分類表、データ一覧、及び分類軸候補一覧の表示部分を含む表示画面を表示する。

[0137] 表示手段105は、分類表には、横軸に、分類軸候補のうち最も優先度の高い分析軸を表示する。また、表示手段105は、縦軸に、関連する属性を表示する。本実施の形態では、表示手段105は、属性として分類項目を表示している。なお、これは一例であり、表示手段105は、データに作成者等があればこれを表示してもよいし、属性が複数あればユーザ操作に従って選択して表示してもよい。また、表示手段105は、表の各セルには、どのようなデータが存在するかを示す情報を表示する。本実施形態では、表示手段105は、データID群とその数を表示している。

[0138] 次に、横軸の「その他」について説明する。図13に示す分類表において

、基準項目「社会」の下の「その他」は、基準項目以下で分類軸内の分類項目に対応しないデータ群を表す分類項目である。また、図13において、最も右側に示す「その他」は、基準項目「社会」にも対応しないデータ群を示す分類項目である。また、縦軸に示す「その他」は、関連する属性のうち、表示しているものに対応しないデータ群を表す項目である。

[0139] 以下、縦軸に関連する属性を表示する手順を示す。まず、表示手段105は、分類階層蓄積部201が記憶する情報とデータ蓄積部203が記憶する情報とを参照し、分類軸に対応するデータ群を取得（抽出）する。次いで、表示手段105は、取得（抽出）したデータ群の属性を参照し、属性値毎に対応するデータ数を調べる（求める）。そして、表示手段105は、データ数のうち、多いものから順番に縦軸に分類軸の分類項目数分だけ表示する。

[0140] 本実施形態では、表示手段105は、基準項目より下の分類項目を属性値とし、これらの分類項目に対応するデータ数を取得（算出）する。また、表示手段105は、既に分類軸として表示しているもの以外の情報を表示する。具体的には、分析軸（社会：家族、健康、移植）を用いる場合、対応するデータは「d1」、「d2」、「d3」、「d4」及び「d6」である。また、これらのデータに対応する分類項目でデータ数が多いもので分類軸にないものは、分類項目「生活」が4件（「d1」、「d2」、「d3」及び「d4」）であり、分類項目「家庭」が3件（「d1」、「d2」及び「d3」）であり、分類項目「医療」が3件（「d2」、「d4」及び「d6」）であり、分類項目「医学」が3件（「d2」、「d4」及び「d6」）である。

[0141] 上記の場合、表示手段105は、これらの分類項目の中から分類項目数3個をデータ数の多い順に選ぶ。なお、表示手段105は、データ数が同じである場合には、いずれか一方を選択して表示する。なお、本例は一例であり、表示手段105は、属性として、分類項目だけでなく他の情報を選択して表示してもよい。例えば、表示手段105は、データに付属する属性であれば、属性をユーザ操作に従って選択して表示してもよい。また、表示手段1

05は、属性値を、上記のように自動的に選択して決めてもよいし、ユーザ操作に従って選択してもよい。また、図13において、縦軸に表示する属性値の数も、分類軸と同じでなくてもよい。

[0142] また、表示手段105は、分類表のいずれかのセルを選択した際に、その選択したセルに対応するデータ一覧を表示する。本実施形態では、表示手段105は、データID、コンテンツ及び分類項目を表示する。なお、表示手段105は、これらの情報を、データ蓄積部203が記憶する情報を参照して表示する。

[0143] また、表示手段105は、分類軸候補一覧において、優先度を計算した分類軸を優先度順に表示する。なお、表示手段105は、これらの情報を、第2の指標計算手段1031が計算した結果から得ることができる。

[0144] 次に、表形式で情報を表示する場合の表示手段105の動作について説明する。まず、表示手段105は、初期表示として、分類表の横軸に、優先度の最も高い分類軸を表示する。この場合、表示手段105は、前述した方法に従って縦軸となる関連する属性も表示する。なお、表示手段105は、データ一覧には、まだ何も表示しない。

[0145] 次に、表示手段105は、分類表のいずれかのセルが選択操作された場合、この選択操作されたセルに対応するデータをデータ一覧に表示する。

[0146] 次に、表示手段105は、分類軸候補一覧からいずれかの分類軸が選択操作された場合、この選択操作された分類軸を分類表の横軸として表示し直す。この場合、表示手段105は、分類表の縦軸の関連する属性も新たに表示し直す。

[0147] 以上に説明したように、本実施形態によれば、第2の指標計算手段1031が求めた分類軸、優先度及びデータを、一覧形式又は表形式で表示する。そのため、ユーザに対して、分類軸の選択状況や優先度、データを視覚的に認識させることができる。

[0148] (第4の実施形態)

次に、本発明の第4の実施形態について説明する。図14は、第4の実施

形態におけるデータ分類システムの構成の一例を示すブロック図である。図 14 に示すように、本実施の形態では、データ分類システムが、図 1 に示した構成要素に加えて、多次元分類軸作成手段 1023 を含む点で、第 1 の実施形態と異なる。また、本実施形態では、データ分類システムが、指標計算手段 103 に代えて、多次元指標計算手段 1032 を含む点で、第 1 の実施形態と異なる。

[0149] 本実施形態において、分類軸候補絞り込み手段 1021 は、第 1 の実施形態と同様の処理に従って、基準項目の全ての子孫の分類項目を組合せるのではなく、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づき作成する分類軸の絞り込みを行う。このようにすることにより、分類軸候補絞り込み手段 1021 は、優先度を計算する対象となる分類軸を減らすことができ、計算を高速化できる。

[0150] 多次元分類候補作成手段 1023 は、具体的には、プログラムに従って動作する情報処理装置の CPU によって実現される。多次元分類軸候補作成手段 1023 は、分類軸候補絞り込み手段 1021 から分類軸候補を受け取り（入力し）、これら分類軸候補を複数組合せることによって、多次元分類軸を作成する機能を備える。なお、作成する多次元分類軸の次元数は、システムが予めもっていてもよいし（例えば、予めメモリ等の記憶部に設定値として記憶していてもよい）、ユーザ操作に従って入力してもよい。また、多次元分類候補作成手段 1023 は、作成した多次元分類軸を多次元指標計算手段 1032 に渡す（出力する）機能を備える。

[0151] 例えば、多次元分類候補作成手段 1023 は、次元数が 2 である場合には、分類軸を 2 つ組み合わせた多次元分類軸候補を作成する。この場合、多次元分類候補作成手段 1023 は、例えば、多次元分類軸（社会：家庭、家族、健康）－（社会：外交、医学、移植）等を作成する。

[0152] 以下、多次元分類候補作成手段 1023 が作成する多次元分類軸を（基準項目：分類項目 N 個）－（基準項目：分類項目 N 個）と表記する。なお、3 次元以上の多次元分類軸であれば、更にその後「－」記号を用いて新たな

分類軸を追加することによって、多次元分類軸を示す。この場合、「－」記号を用いて繋がれた各分類軸は、それぞれの次元における分類軸を表している。例えば、多次元分類軸（社会：家庭、家族、健康）－（社会：外交、医学、移植）である場合には、1つ目の分類軸（社会：家庭、家族、健康）を第1次元分類軸といい、2つ目の分類軸（社会：外交、医学、移植）を第2次元分類軸という。

[0153] 多次元指標計算手段1032は、具体的には、プログラムに従って動作する情報処理装置のCPUによって実現される。多次元指標計算手段1032は、多次元分類軸候補作成手段1023から多次元分類軸候補を受け取り（入力し）、分類階層蓄積部201が記憶する情報とデータ蓄積部203が記憶する情報とを参照し、分類軸候補の優先度を計算する機能を備える。この場合、多次元指標計算手段1032は、優先度を、分類項目の分類階層での階層的な距離に基づいて計算する。

[0154] 「分類項目間の階層的な距離」とは、共通する祖先までの最短のパス長や最長のパス長、又は共通の子孫までの最短パス長や最長パス長等である。また、多次元指標計算手段1032は、優先度として、分類軸の各分類項目間の階層的な距離の平均や最大、最小の値等を求める。

[0155] 本実施形態では、多次元指標計算手段1032は、「分類項目間の階層的な距離」として共通する祖先までの最短パス長を用いるものとし、優先度として階層的な距離の平均の値を求めるものとする。そのようにするのは、距離が長い方が意味的に独立しているといえるためである。更に、多次元指標計算手段1032は、分類軸内の分類項目だけでなく、分類軸の基準項目間の階層的な距離も用いて優先度を計算する。

[0156] なお、本実施形態において、多次元分類候補作成手段1023及び多次元指標計算手段1032以外の構成要素の機能は、第1の実施形態で示したそれらの機能と同様である。

[0157] 次に、動作について説明する。図15は、第4の実施形態におけるデータ分類システムが行うデータ分類処理の一例を示す流れ図である。

- [0158] まず、第1の実施形態で示した処理と同様に、データ分類システムの入力手段101は、ユーザの操作に従って、分類項目数Nを受け取る（ステップS1）。例えば、入力手段101は、分類項目数としてN=3を入力する。次に、第1の実施形態で示した処理と同様に、分類軸候補絞り込み手段1021は、基準項目蓄積部202から、分類軸の基準項目を取得（抽出）する（ステップS2）。例えば、図3に示す例において、分類軸候補絞り込み手段1021は、基準項目蓄積部202から、「社会」、「自然」及び「文化」等の情報を抽出する。
- [0159] 次に、第1の実施形態で示した処理と同様に、分類軸候補絞り込み手段1021は、分類階層蓄積部201が記憶する情報とデータ蓄積部202が記憶する情報とを参照し、分類項目に対応するデータ量に基づいて分類項目を絞り込む（ステップS31）。
- [0160] 次に、第1の実施形態で示した処理と同様に、分類軸候補絞り込み手段1021は、分類階層蓄積部201が記憶する情報を参照し、分類項目間の階層的な距離に基づいて分類項目を組合せて分類軸を作成する（ステップS32）。
- [0161] 次に、多次元分類軸候補作成手段1023は、作成した分類軸候補を次元数分組み合わせて、多次元分類軸を作成する（ステップS321）。なお、作成する多次元分類軸の次元数は予めシステムがもっていてもよいし（例えば、予めメモリ等の記憶部に設定値として記憶していてもよい）、ユーザ操作に従って入力してもよい。例えば、多次元分類軸候補作成手段1023は、次元数が2である場合には、多次元分類軸（社会：家庭、家族、健康）－（社会：外交、医学、移植）等を作成する。
- [0162] 次に、多次元指標計算手段1032は、多次元分類軸候補作成手段1023から多次元分類軸候補を取得（入力）し、分類階層蓄積部201が記憶する情報を参照して、各分類軸の優先度を計算する（ステップS421）。
- [0163] 多次元指標計算手段1032は、優先度の計算を行う場合、本実施形態では、分類項目の意味的な独立性を図るため、分類項目間の階層的な距離の平

均値と、基準項目間の階層的な距離の平均値とを計算する。ここで、「分類項目間の階層的な距離」又は「基準項目間の階層的な距離」とは、分類項目間で共通する先祖分類項目までの最短パスである。また、多次元指標計算手段1032は、優先度を、以下の式(17)、(18)を用いて求める。

$$\begin{aligned}
 [0164] \quad & \text{多次元優先度} \left((X_1 : C_1) - (X_2 : C_2) - \dots \right) \\
 & = 1 / \text{次元数} \times \sum \text{独立性} (X_i : C_i) + 1 / (2 \times \text{次元数}) \times \sum \text{ComDist} (X_i, X_j) \\
 & \dots \text{式}(17)
 \end{aligned}$$

$$\begin{aligned}
 [0165] \quad & \text{独立性} (X : C) \\
 & = 1 / \text{Max} (X) \times 1 / (2 \times \text{組合せ数}) \times \sum \text{ComDist} (c_i, c_j) \\
 & \dots \text{式}(18)
 \end{aligned}$$

[0166] ここで、式(17)において、 X_1, X_2, \dots, X_i は、第*i*次元の基準項目を示す。また、 C_1, C_2, \dots, C_i は、第*i*次元の分類項目群を示す。なお、 $\text{Max} (X)$ 及び $\text{ComDist} (c_i, c_j)$ は、第1の実施形態で示したものと同様である。式(17)に示すように、多次元指標計算手段1032は、第1項目で次元毎に計算した独立性(分類項目間の階層的な距離)を次元数で除算することで平均値を求める。また、多次元指標計算手段1032は、第2項で基準項目間の階層的な距離の平均値を求める。

[0167] 例えば、多次元指標計算手段1032は、多次元分類軸(社会:家庭、家族、健康) - (社会:外交、医学、移植)の優先度を、次の式(19)のように計算する。この場合、分類項目数 $N=3$ であるので、組合せ数を3とし、次元数を2とする。基準項目「社会」の子孫の分類項目で最も深い分類項目は、図2に示した分類階層の場合には2である。

$$\begin{aligned}
 [0168] \quad & \text{優先度} \left((\text{社会} : \text{家庭、家族、健康}) - (\text{社会} : \text{外交、医学、移植}) \right) \\
 & = 1 / 2 \left(1 / 2 \times 1 / 6 \times (\text{ComDist} (\text{家庭}, \text{家族}) + \text{ComDist} (\text{家庭}, \text{健康}) + \text{ComDist} (\text{家族}, \text{健康})) + (1 / 2 \times 1 / 6 \right.
 \end{aligned}$$

$$\begin{aligned}
 & \times (\text{ComDist}(\text{外交}, \text{医学}) + \text{ComDist}(\text{外交}, \text{移植}) + \text{ComDist}(\text{医学}, \text{移植})) + 1/4 \times (\text{ComDist}(\text{社会}, \text{社会})) \\
 &) \\
 & = 1/2 (1/2 \times 1/6 \times (2+2+2) + (1/2 \times 1/6 \times (4+4+2))) + 1/4 \times (0) \\
 & = 0.67 \\
 & \dots \text{式}(19)
 \end{aligned}$$

[0169] なお、3次元以上の多次元分類軸を用いる場合であっても、多次元指標計算手段1032は、同様の計算を行うことによって、複数次元の優先度を計算できる。

[0170] 上記のように計算することによって、類似した分類項目ではなく、意味的に独立した分類項目を含む分類軸に高い優先度を与えることができる。更に、多次元の分類軸に対応することができる。

[0171] 次に、出力手段104は、多次元指標計算手段1032の計算結果に基づいて、分類軸と、優先度と対応するデータとを出力する（ステップS5）。図16（A）、図16（B）、及び図16（C）は、第4の実施形態における出力手段104が出力する情報の例を示す図である。図16（A）、図16（B）、及び図16（C）に示す例では、出力手段104は、3つのテーブルを出力している。なお、図中の「・・・」は、記載の省略を表している。

[0172] 例えば、出力手段104は、図16（A）に示すように、次元IDと、分類軸ID群と、スコアとを対応づけたレコードを含むテーブルを出力する。すなわち、図16（A）のテーブルには、多次元分類軸候補毎の分類軸とそのスコアとが表されている。図16（A）に示す例では、分類軸ID群を、分類軸IDを「,」で区切って表している。本実施形態では、次元数が2であるので、分類軸ID群は、2つの分類軸IDを含む。なお、3次元以上の多次元分類軸である場合には、この分類軸IDを増やすことによって多次元に対応することができる。

- [0173] また、出力手段104は、図16(B)に示すように、分類軸ID、基準項目及び分類項目群を対応付けたレコードを含むテーブルを出力する。すなわち、図16(B)に示す例では、1つの行が1つの分類軸を表している。
- [0174] また、出力手段104は、図16(C)に示すように、分類軸ID、分類項目及びデータID群を対応付けたレコードを含むテーブルを出力する。すなわち、図16(C)に示す例では、1つのレコードが各分類軸の分類項目に対応している。また、図16(C)に示す例では、データID群を、データIDを「,」で区切って表している。また、図中の「・・・」は、記載の省略を表している。
- [0175] なお、図16(A)、図16(B)、及び図16(C)に示す出力方法は一例であり、出力手段104は、例えば、2つのテーブルを1つにまとめて出力してもよいし、各データの属性情報を含んだテーブルを新たに追加して出力してもよい。
- [0176] 上記のような構成を備えることによって、分類項目の意味的な独立性を用いて分類軸を選択することができる。これにより、ユーザにわかりやすい分類軸を選択することができる。
- [0177] 以上に説明したように、本実施形態によれば、基準項目の子孫の分類項目においてデータに少なくとも1つ対応する分類項目の組合せに基づいて分類軸を作成する際に、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む。また、絞り込んだ分類軸候補を組合せることによって、多次元の分類軸候補を作成する。そして、分類階層における分類項目間の階層的な距離に基づいて、多次元分類軸候補の優先度を計算する。従って、分類階層とその分類項目とに対応するデータ群が与えられたときに、多次元の分類軸についても優先度の計算時間を短縮することができる。
- [0178] (第5の実施形態)

次に、本発明の第5の実施形態について説明する。図17は、第5の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。図

17に示すように、本実施形態では、データ分類システムが、図14に示した多次元指標計算手段1032に代えて、第2の多次元指標計算手段1033を含む点で、第4の実施形態と異なる。

[0179] 第2の多次元指標計算手段1033は、具体的には、プログラムに従って動作する情報処理装置のCPUによって実現される。第2の多次元指標計算手段1033は、多次元分類軸候補作成手段1022から分類軸候補群を受け取り（入力し）、分類軸の優先度を計算する機能を備える。この場合、第2の多次元指標計算手段1033は、分類項目間の階層的な距離だけでなく、分類項目の階層的な深さ、分類項目に対応するデータの量、分類項目に対応するデータの重複度合い、又はこれらの組合せに基づいて優先度を計算する。

[0180] なお、第2の多次元指標計算手段1033は、これらの優先度の計算方法として、例えば、第2の実施形態で示した優先度計算方法を多次元に拡張した方法を用いて、優先度を計算する。

[0181] なお、本実施形態において、第2の多次元指標計算手段1033以外の構成要素の機能は、第4の実施形態で示したそれらの機能と同様である。

[0182] 次に、第2の多次元指標計算手段1033が優先度を計算する場合の優先度の計算方法について説明する。第2の実施形態と同様の処理に従って、第2の多次元指標計算手段1033は、多次元分類軸候補作成手段1023から多次元分類軸候補を取得（入力）し、分類階層蓄積部201が記憶する情報とデータ蓄積部203が記憶する情報とを参照して、計算用テーブルを作成する。但し、本実施形態では、第2の多次元指標計算手段1033は、計算用テーブルを、複数の次元にまたがって作成する点で、第2の実施形態と異なる。

[0183] 第2の多次元指標計算手段1033は、計算用テーブルとして、分類項目別データ数表とデータ別分類項目数表との2つのテーブルを作成する。

[0184] 分類項目別データ数表とは、異なる次元における分類項目の組合せに対応しているデータ数を集計したテーブルである。分類項目別データ数表は、分

類項目組合せ、データ数及び分類項目の深さを対応付けたレコードを含む。また、データ別分類項目数表とは、各データに対応している分類軸内の分類項目の数を集計したテーブルであり、データID及び分類項目組合せ数を含む。なお、第2の多次元指標計算手段1033は、作成した計算用の一時的なテーブルを、メモリ上に展開しておくことが望ましい。

[0185] 図18(A)及び図18(B)は、分類項目別データ数表とデータ別分類項目数表との例を示す図である。このうち、図18(A)は、分類項目別データ数表の例を示している。また、図18(B)は、データ別分類項目数表の例を示している。なお、図18(A)及び図18(B)に示す例では、第2の多次元指標計算手段1033が、多次元分類軸(社会:家庭、家族、健康) - (社会:外交、医学、移植)、図2に示した分類階層蓄積部201が記憶する情報、及び図4に示したデータ蓄積部203が記憶する情報に基づいて、分類項目別データ数表とデータ別分類項目数表とを作成した場合が示されている。

[0186] 図10(A)に示す例では、分類項目別データ数表は、分類項目組合せ、データ数、及び分類項目の深さを対応付けた各レコードを含む表である。本実施形態では、第2の多次元指標計算手段1033は、分類項目の深さを、以降の計算を簡単にするため、次の式(20)のように計算する。

[0187] 深さ (c_{ij} , c_{kl} , ...)

$$= 1 / \text{次元数} \times \sum (1 / \text{Max}(X_i) \times \text{Depth}(X_i, c_{ij}))$$
 ...式(20)

[0188] ここで、式(20)において、 c_{ij} は第*i*次元の*j*番目の分類項目であり、 c_{kl} は第*k*次元の*l*番目の分類項目である。但し、*i*と*k*とは、値が異なるものとする。また、 X_i は、第*i*次元の基準項目である。また、 $\text{Max}(X_i)$ は、基準項目 X_i の子孫の分類項目で最も深い分類項目までの深さを示している。また、 $\text{Depth}(X_i, c_{ij})$ は、第*i*次元の基準項目 X_i からその分類項目 c_{ij} までの最短パス長である。また、 \sum 記号は、異なる次元における分類項目の組合せで合計することを表している。すなわ

ち、式（20）では、分類項目の深さを、分類項目組合せ内での分類項目の深さの平均値で示している。

[0189] 図18（A）において、例えば、1レコード目は、第1次元の分類項目「家庭」と第2次元の分類項目「外交」とにおけるデータを表したものである。図4に示すデータ蓄積部203が記憶する情報を参照すると、本例では、データ蓄積部203は、この2つの分類項目に対応するデータを記憶していないので、図18（A）に示すように、データ数が0である。また、分類項目の深さの部分については、社会から家族までが2であり、社会から医学までが2であり、 $\text{Max}(X_i)$ が2であり、次元数が2であるので、図18（A）に示すように、分類項目の深さが1となる。

[0190] データ別分類項目数表とは、データIDと分類項目組合せ数とを対応付けたレコードを含む表である。図18（B）に示す例では、各データIDに対して、多次元分類軸（社会：家庭、家族、健康）－（社会：外交、医学、移植）と対応する分類項目数を含むデータ別分類項目数表が示されている。図18（B）に示す例において、例えば、データID「d2」は、図4に示すデータ蓄積部203が記憶する情報を参照すると、第1次元の分類項目「健康」と第2次元の分類項目「医学」とに対応しているので、分類項目数が1である。

[0191] 次に、第2の多次元指標計算手段1033は、計算用テーブルを用いて、分類軸の優先度を計算する。本実施形態では、第2の多次元指標計算手段1033は、前述した独立性、具体性、網羅性及び一意性の指標の値を計算し、これらの指標値の重み付き線形和を求めることによって、式（21）を用いて、総合的な優先度を計算する。

[0192] 多次元優先度（ $(X_1 : C_1) - (X_2 : C_2) - \dots$ ）
 $= W_1 \times \text{多次元独立性}（(X_1 : C_1) - (X_2 : C_2) - \dots） + W_2$
 $\times \text{多次元具体性}（(X_1 : C_1) - (X_2 : C_2) - \dots） + W_3 \times \text{多次元網羅性}（(X_1 : C_1) - (X_2 : C_2) - \dots） + W_4 \times \text{多次元一意性}（(X_1 : C_1) - (X_2 : C_2) - \dots）$

・・・式(21)

[0193] ここで、式(21)において、Xは基準項目であり、Cは分類項目群である。また、W1、W2、W3、W4は、それぞれ指標の重み係数である。なお、これらの重み係数は、予めシステムがもっていてもよいし(例えば、予めメモリ等の記憶部に設定値として記憶していてもよい)、ユーザによって設定されてもよい。本実施形態では、これらの重み係数は、システムに予め設定されているものとする。

[0194] なお、本実施形態では、独立性の指標値は、第5の実施形態と同様であり、第2の多次元指標計算手段1033は、式(22)、(23)を用いて多次元独立性の指標値を求める。

[0195] 多次元独立性 ((X1:C1) - (X2:C2) - ...)

$$= 1 / \text{次元数} \times \sum \text{独立性}(X_i : C_i) + 1 / (2 \times \text{次元数}) \times \sum \text{ComDist}(X_i, X_j)$$

・・・式(22)

[0196] 独立性(X:C)

$$= 1 / \text{Max}(X) \times 1 / (2 \times \text{組合せ数}) \times \sum \text{ComDist}(C1, C2)$$

・・・式(23)

[0197] 式(22)、(23)において、X1、X2、Xi、C1、C2、Ci、Max(X)、ComDist、組合せ数及び次元数は、第4の実施形態で示したものと同様である。

[0198] また、第2の多次元指標計算手段1033は、具体性の指標値を、以下のように計算する。本実施形態において、具体性の指標は、分類軸毎の基準項目から各分類項目までのパス長の平均の値である。第2の多次元指標計算手段1033は、具体性の指標値を、分類項目別データ数表を用いて、以下の式(24)、(25)のように計算できる。

[0199] 多次元具体性 ((X1:C1) - (X2:C2) - ...)

$$= 1 / \text{次元数} \times \sum \text{具体性}(X_i : C_i)$$

・・・式(24)

[0200] 具体性 (X : C)

$$= 1 / \text{Max}(X) \times 1 / N \times \sum \text{Depth}(X, c_j)$$

・・・式(25)

[0201] 式(24)、(25)において、 $\text{Max}(X)$ 、 N 、 $\text{Depth}(X, c_j)$ は、第2の実施形態で示したものと同様である。ここで、図18(A)に示すように、分類項目別データ数表における分類項目の深さ指標については、 $1 / \text{次元数} \times \sum 1 / \text{Max}(X_i) \times \text{Depth}(X_i, c_{ij})$ として計算済みであるので、多次元具体性の指標を、以下の式(26)のように計算できる。

[0202] 多次元具体性 ((X1 : C1) - (X2 : C2) - ...)

$$= 1 / \text{次元数} \times 1 / N \times \sum \sum (1 / \text{Max}(X_i) \times \text{Depth}(X_i, c_{ij}))$$

$$= 1 / (N \text{次元数}) \times \sum (\text{分類項目別データ数表の深さ指標})$$

・・・式(26)

[0203] また、第2の多次元指標計算手段1033は、網羅性の指標を、以下のよう
に計算する。本実施形態において、網羅性の指標は、各次元の分類項目の
組合せに対応するデータの全データに対するカバー率である。第2の多次元
指標計算手段1033は、網羅性の指標を、先に作成したデータ別分類項目
表を用いて、次の式(27)のように求めることができる。

[0204] 多次元網羅性 ((X1 : C1) - (X2 : C2) - ...)

$$= 1 / \text{DataNum} \times | \text{Uデータ}(c_{ij}, c_{kl}, \dots) |$$

$$= 1 / \text{DataNum} \times \text{RecNum}(\text{データ別項目分類表}, \text{分類項目数} > 0)$$

・・・式(27)

[0205] ここで、式(27)において、「データ(c_{ij} , c_{kl} , ...)」は、第*i*次元の*j*番目の分類項目 c_{ij} 、第*k*次元の*l*番目の分類項目 c_{kl} 、及びその他の次元の分類項目の全てに対応するデータ集合である。また、

Data Numは、データ集合の数である。Rec Num（データ別分類項目数表，分類項目数>0）は、データ別分類項目数表の分類項目数が0よりも大きいレコード数である。このRec Num（データ別分類項目数表，分類項目数>0）は、分類項目の組合せに対応しているデータ数に等しい。よって、上記の式（27）のように書き換えることができる。

[0206] また、第2の多次元指標計算手段1033は、一意性の指標を、以下のよう
に計算する。本実施形態において、一意性の指標は、データの重複度合い
の逆数で表される。ここで、データの重複度合いは、各分類項目の組合せに
対応するデータ数の合計値を重複なしのデータ数の合計値で除算した値とす
る。第2の多次元指標計算手段1033は、一意性の指標を、先に作成した
分類項目別データ数表を用いて、次の式（28）のように求めることができ
る。

[0207] 多次元一意性（（X1：C1）－（X2：C2）－・・・）

$$= | \cup \text{データ}(c_{ij}, c_{kl}, \dots) | / \sum \text{Cat Num}(c_{ij}, c_{kl}, \dots)$$

$$= \text{Rec Num}(\text{データ別項目分類表, 分類項目数} > 0) / \sum (\text{データ別項目分類表のデータ数})$$
 ・・・式（28）

[0208] 例えば、多次元分類軸（社会：家庭、家族、健康）－（社会：外交、医学、移植）である場合、第2の多次元指標計算手段1033は、図10（A）
に示した分類項目別データ数表、及び図2に示す分類階層蓄積部201が記憶する情報を参照して、上記の各指標値を、以下の式（29）～式（32）
のように計算する。

[0209] 多次元独立性（（X1：C1）－（X2：C2）－・・・）

$$= 1 / \text{次元数} \times \sum \text{独立性}(X_i : C_i) + 1 / (2 \times \text{次元数}) \times \sum \text{ComDist}(X_i, X_j)$$

$$= 1 / 2 (1 / 2 \times 1 / 6 \times (\text{ComDist}(\text{家庭, 家族}) + \text{ComDist}(\text{家庭, 健康}) + \text{ComDist}(\text{家庭, 健康})) + (1 / 2 \times 1 / 6$$

$$\begin{aligned}
& \times (\text{ComDist}(\text{外交}, \text{医学}) + \text{comDist}(\text{外交}, \text{移植}) + \text{ComDist}(\text{医学}, \text{移植})) + 1/4 \times (\text{ComDist}(\text{社会}, \text{社会})) \\
& = 1/2 (1/2 \times 1/6 \times (2+2+2) + (1/2 \times 1/6 \times (4+4+2))) + 1/4 \times (0) \\
& = 0.667 \\
& \dots \text{式}(29)
\end{aligned}$$

[0210] 多次元具体性 ((X1 : C1) - (X2 : C2) - ...)

$$\begin{aligned}
& = 1 / (\text{N次元数}) \times \Sigma (\text{分類項目別データ数表の深さ指標}) \\
& = 1 / 9 \times (1+1+1+1+1+1+1+1+1) \\
& = 1 \\
& \dots \text{式}(30)
\end{aligned}$$

[0211] 多次元網羅性 ((X1 : C1) - (X2 : C2) - ...)

$$\begin{aligned}
& = 1 / \text{DataNum} \times \text{RecNum} (\text{データ別項目分類表}, \text{分類項目数} > 0) \\
& = 1 / 6 \times 2 \\
& = 0.333 \\
& \dots \text{式}(31)
\end{aligned}$$

[0212] 多次元一意性 ((X1 : C1) - (X2 : C2) - ...)

$$\begin{aligned}
& = \text{RecNum} (\text{データ別項目分類表}, \text{分類項目数} > 0) / \Sigma (\text{データ別項目分類表のデータ数}) \\
& = 2 / (0+1+0+2+0+0) \\
& = 2 / 3 \\
& = 0.667 \\
& \dots \text{式}(32)
\end{aligned}$$

[0213] ここで、重み係数を全て0.25とすると、第2の多次元指標計算手段1033は、次の式(33)を用いて優先度を求めることができる。

[0214] 優先度 (X : C)

$$\begin{aligned}
 &= W1 \times \text{独立性} (X : C) + W2 \times \text{具体性} (X : C) + W3 \times \text{網羅性} (X : C) \\
 &+ W4 \times \text{一意性} (X : C) \\
 &= 0.25 \times 0.667 + 0.25 \times 1 + 0.25 \times 0.333 + 0.25 \\
 &\times 0.667 \\
 &= 0.67 \\
 &\dots \text{式} (33)
 \end{aligned}$$

[0215] 以上に説明したように、本実施形態によれば、分類項目間の階層的な距離だけでなく、分類項目の階層的な深さ、分類項目に対応するデータの量、分類項目に対応するデータの重複度合い、又はこれらの組合せに基づいて、優先度を計算する。そのため、分類項目間の階層的な距離に加えて、分類項目の階層的な深さや、分類項目に対応するデータの量、分類項目に対応するデータの重複度合いを考慮して、より効果的に多次元の分類軸の優先度の計算時間を短縮することができる。

[0216] (第6の実施形態)

次に、本発明の第6の実施形態について説明する。図19は、第6の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。図19に示すように、本実施形態では、データ分類システムが、図17に示した構成要素に加えて、多次元表示手段1051を含む点で、第5の実施形態と異なる。

[0217] 多次元表示手段1051は、具体的には、プログラムに従って動作する情報処理装置のCPU及びディスプレイ装置等の表示装置によって実現される。多次元表示手段1051は、第2の多次元指標計算手段1033が求めた分類軸、優先度及びデータをディスプレイ装置等の表示装置に出力(表示)する機能を備える。例えば、多次元表示手段1051は、各次元の分類軸の分類項目と対応するデータ数、及びデータとその属性を、一覧形式又は表形式で出力(表示)する。

[0218] まず、多次元表示手段1051が一覧形式で情報を表示する場合を説明する。図20は、多次元表示手段1051が一覧形式で表示する情報の例を示

す図である。図20に示すように、多次元表示手段1051は、多次元分類軸、多次元分類軸候補一覧、及びデータ一覧のそれぞれの表示部分を含む表示画面を表示する。多次元表示手段1051は、多次元分類軸には、優先度が最も高かった多次元分類軸、又は多次元分類軸候補一覧から選択した分析軸を表示する。また、多次元表示手段1051は、各次元に基準項目と分類項目とを表示し、その後ろにデータ件数を表示する。

[0219] 図20に示す例では、多次元分類軸（社会：家族、健康、移植）－（社会：家庭、外交、医療）を用いる場合が示されている。この場合、多次元表示手段1051は、それぞれの項目と対応するデータ件数をデータ蓄積部203から取得（抽出）し表示する。

[0220] また、図20に示す例では、多次元表示手段1051が表示する表示画面には、項目「その他」の欄が新たに加えられている。ここで、「社会」の下の「その他」の欄は、基準項目「社会」に関連しているが、分類軸内の分類項目に対応しないデータを分類する項目である。また、全データの子にある「その他」の欄は、全ての分類軸とは関係のない分類項目を示す。なお、多次元表示手段1051は、これら「その他」の欄のデータ数を、データ蓄積部203が記憶する情報を参照することによって得ることができる。また、本実施形態では、データ件数には子孫の分類項目に対応するデータ件数も含まれる。

[0221] また、図20に示す例では、多次元表示手段1051は、多次元分類軸候補一覧を、優先度を計算した分類軸を優先度順に表示する。多次元表示手段1051は、これらの多次元分類軸を、第2の多次元指標計算手段1033が計算した結果から得ることができる。

[0222] また、図20に示す例では、多次元表示手段1051は、データ一覧を、各データを一覧で表示する。この場合、多次元表示手段1051は、各データを、データID、コンテンツ、及び対応する分類項目を対応付けて表示する。なお、多次元表示手段1051は、これらの情報を、データ蓄積部203が記憶する情報を参照することで得ることができる。

- [0223] なお、本実施形態において、多次元表示手段1051以外の構成要素の機能は、第5の実施形態で示したそれらの機能と同様である。
- [0224] 次に、一覧形式で情報を表示する場合の多次元表示手段1051の動作について説明する。まず、多次元表示手段1051は、初期表示として、優先度の最も高い多次元分類軸を多次元分類軸表示部分に表示する。また、多次元表示手段1051は、その他の多次元分類軸候補を多次元分類軸候補一覧に優先度順に表示する。また、多次元表示手段1051は、データ一覧には、データ蓄積部203が蓄積する全てのデータを表示する。
- [0225] なお、多次元表示手段1051は、多次元分類軸表示部分に表示する多次元分類軸において、各次元内の分類項目を1つ又は複数選択することによって、それら全てに対応するデータのみをデータ一覧に表示する。
- [0226] 次に、多次元表示手段1051は、多次元分類軸候補一覧表示部分に表示する多次元分類軸の中のいずれかの分類軸が選択操作されると、多次元分類軸表示部分の内容をこの選択した多次元分類軸で置き換えて表示する。
- [0227] なお、本実施形態では、次元数が2次元である場合を示したが、多次元表示手段1051は、3次元以上の次元数の多次元分類軸についても、同様の処理に従って表示することができる。この場合、多次元表示手段1051は、多次元分類軸表示部分に3次元以降の分類軸を追加して表示することによって、データの一覧表示を行う。
- [0228] 次に、多次元表示手段1051が表形式で情報を表示する場合を説明する。図21は、多次元表示手段1051が表形式で表示する情報の例を示す図である。図21に示すように、多次元表示手段1051は、多次元分類表、データ一覧、及び多次元分類軸候補一覧の表示部分を含む表示画面を表示する。
- [0229] 多次元表示手段1051は、多次元分類表には、多次元分類軸候補のうち最も優先度の高い多次元分析軸を用いて、横軸に第1次元の情報を表示し、縦軸に第2次元の情報を表示する。なお、3次元以上の多次元分類表である場合には、多次元表示手段1051は、縦軸や横軸に更にもう1次元分の情

報を追加して表示する。例えば、多次元分類表が3次元である場合、多次元表示手段1051は、横軸に第1次元の情報を表示し、その隣に第3次元の情報を表示し、縦軸に第2次元の情報を配置して表示する。また、多次元表示手段1051は、表の各セルには、どのようなデータが存在するかを示す情報を表示する。本実施形態では、多次元表示手段1051は、データID群とその数を表示している。

[0230] また、多次元表示手段1051は、多次元分類表のいずれかのセルを選択した際に、その選択したセルに対応するデータ一覧を表示する。本実施形態では、多次元表示手段1051は、データID、コンテンツ及び分類項目を表示する。なお、多次元表示手段1051は、これらの情報を、データ蓄積部203が記憶する情報を参照して表示する。

[0231] また、多次元表示手段1051は、多次元分類軸候補一覧において、優先度を計算した多次元分類軸を優先度順に表示する。なお、多次元表示手段1051は、これらの情報を、第2の多次元指標計算手段1032が計算した結果から得ることができる。

[0232] 次に、表形式で情報を表示する場合の多次元表示手段1051の動作について説明する。まず、多次元表示手段1051は、初期表示として、多次元分類表に、優先度の最も高い多次元分類軸を表示する。この場合、多次元表示手段1051は、前述した方法に従って横軸と縦軸とを表示する。なお、多次元表示手段1051は、データ一覧には、まだ何も表示しない。

[0233] 次に、多次元表示手段1051は、多次元分類表のいずれかのセルが選択操作された場合、この選択操作されたセルに対応するデータをデータ一覧に表示する。

[0234] 次に、多次元表示手段1051は、多次元分類軸候補一覧からいずれかの多次元分類軸が選択操作された場合、この選択操作された分類軸を多次元分類表の各軸として表示し直す。

[0235] 以上に説明したように、本実施形態によれば、第2の多次元指標計算手段1033が求めた多次元の分類軸、優先度及びデータを、一覧形式又は表形

式で表示する。そのため、ユーザに対して、多次元の分類軸の選択状況や優先度、データを視覚的に認識させることができる。

[0236] (第7の実施形態)

次に、本発明の第7の実施形態について説明する。図22は、第7の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。図22に示すように、本実施形態では、データ分類システムが、図19に示した構成要素に加えて、検索手段106を含む点で、第6の実施形態と異なる。

[0237] 検索手段106は、具体的には、プログラムに従って動作する情報処理装置のCPUによって実現される。検索手段106は、入力手段101から検索キーワードと分類項目数とを受け取り（入力し）、データ蓄積部203が記憶するコンテンツやその他の属性部分の情報を検索する機能を備える。また、検索手段106は、検索結果データID群を取得（抽出）し、分類軸候補絞り込み手段1021に渡す（出力する）機能を備える。なお、検索手段106は、検索処理を実行する場合に、既存の全文検索エンジンやリレーショナルデータベースの手法を用いて、コンテンツや属性部分の情報を検索する。

[0238] 本実施形態では、データ分類システムが、データ蓄積部203から検索結果のデータID群を選択したデータベースに対して処理を行う点が、第6の実施形態と異なる。なお、データ分類システムが実行するその他の処理は、第6の実施形態で示した処理と同様である。

[0239] なお、本実施形態において、データ分類システムは、第2の多次元指標計算手段1033に代えて、多次元指標計算手段1032を用いて処理を実行してもよい。更に、データ分類システムは、多次元分類軸作成手段1023を含まず、第2の多次元指標計算手段1033に代えて、指標計算手段103又は第2の指標計算手段1031を用いて処理を実行してもよい。また、データ分類システムは、多次元表示手段1051に代えて表示手段105を用いて処理を実行してもよい。

[0240] 上記の構成によって、ユーザの検索結果に対してのみ多次元分類軸、又は分類軸を表示することができる。

[0241] 以上に説明したように、本実施形態によれば、データ蓄積部203が記憶するコンテンツやその他の属性部分の情報を検索し、検索した情報についてのみ分類軸候補の絞り込みを行う。従って、分類階層とその分類項目とに対応するデータ群が与えられたときに、より効率的に分類軸の優先度の計算時間を短縮することができる。

[0242] (第8の実施形態)

次に、本発明の第8の実施形態について説明する。図23は、第8の実施形態におけるデータ分類システムの構成の一例を示すブロック図である。図23に示すように、本実施形態では、データ分類システムが、図22に示した構成要素に加えて、データ対応付け手段107を含む点で、第7の実施形態と異なる。

[0243] データ対応付け手段107は、具体的には、プログラムに従って動作する情報処理装置のCPUによって実現される。データ対応付け手段107は、分類階層蓄積部201が記憶する情報とデータ蓄積部203が記憶する情報とを参照し、データと分類項目との対応付けを行う機能を備える。なお、データ対応付け手段107は、対応付け方法として、分類項目名のデータのコンテンツ中の出現による方法や、分類項目名とデータとのコンテンツのコサイン類似度を測る方法等の既存の手法を用いて、対応付けを行うことができる。

[0244] なお、データ対応付け手段107は、多次元分類候補作成手段1023が分類軸候補を作成する前に、対応付けを行うことが望ましい。

[0245] なお、本実施形態において、データ分類システムは、第2の多次元指標計算手段1033に代えて、多次元指標計算手段1032を用いて処理を実行してもよい。更に、データ分類システムは、多次元分類軸作成手段1023を含まず、第2の多次元指標計算手段1033に代えて、指標計算手段103又は第2の指標計算手段1031を用いて処理を実行してもよい。また、

データ分類システムは、多次元表示手段 1051 に代えて、表示手段 105 を用いて処理を実行してもよい。

[0246] 以上に説明したように、本実施形態によれば、分類階層蓄積部 201 が記憶する情報とデータ蓄積部 203 が記憶する情報とを参照し、データと分類項目との対応付けを行った上で、分類軸候補の絞り込みを行う。従って、分類階層とその分類項目とに対応するデータ群が与えられたときに、よりの確に分類軸の優先度の計算時間を短縮することができる。

[0247] 次に、データ分類システムの最小構成について説明する。図 24 は、データ分類システムの最小の構成例を示すブロック図である。図 24 に示すように、データ分類システムは、最小の構成要素として、基準項目蓄積部 202、分類軸候補絞り込み手段 1021、及び指標計算手段 103 を含む。

[0248] 基準項目蓄積部 202 は、分類項目を選択するための基準項目となる分類項目群を予め蓄積する。また、分類軸候補絞り込み手段 1021 は、基準項目の子孫の分類項目から複数の分類項目を選択して分類軸候補とし、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む機能を備える。また、指標計算手段 103 は、分類軸候補絞り込み手段 1021 が絞り込んだ分類軸候補について、当該分類軸候補を表示する優先度を計算する機能を備える。

[0249] 図 24 に示す最小構成のデータ分類システムによれば、分類階層とその分類項目とに対応するデータ群が与えられたときに、分類軸の優先度の計算時間を短縮することができる。

[0250] なお、上記の各実施形態では、以下の (1) ~ (14) に示すようなデータ分類システムの特徴的構成が示されている。

[0251] (1) データ分類システムは、階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、データ群に対応する分類項目を複数選択して分類軸として出力するデータ分類システムであって、分類項目を選択するための基準項目となる分類項目群を予め蓄積する基準項目蓄積手段（例えば、基準項目蓄積部 202 によって実現される）と、基準項目の子孫の分類項目が

ら複数の分類項目を選択して分類軸候補とし、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込手段（例えば、分類軸候補絞込み手段1021によって実現される）と、分類軸絞込手段が絞り込んだ分類軸候補について、当該分類軸候補を表示する優先度を計算する優先度計算手段（例えば、指標計算手段103によって実現される）とを備える。

[0252] （2）データ分類システムにおいて、分類軸絞込手段は、分類項目に対応するデータ量が所定数以上であること、又は分類項目に対応するデータ量が上位所定割合以内であることのうちの少なくともいずれか一方の条件と、分類項目間の共通祖先までの長さが特定の範囲内であることという条件とを満たす分類軸を選択するように構成されていてもよい。

[0253] （3）データ分類システムにおいて、優先度計算手段は、分類階層における分類項目間の共通の先祖までの長さに応じて、優先度を決定するように構成されていてもよい。

[0254] （4）データ分類システムにおいて、優先度計算手段は、分類階層における分類項目間の階層的な距離を計算し、更に、分類項目の分類階層での深さ、分類項目群に対応するデータ量、又は分類項目群に対応するデータの重複度合いのうちの少なくともいずれか1つに基づいて、分類軸候補の優先度を決定するように構成されていてもよい。

[0255] （5）データ分類システムは、分類軸絞込手段が絞り込んだ分析軸候補と、優先度計算手段が計算した優先度とを読み込み、データ群とともに表示制御する表示制御手段（例えば、表示手段105によって実現される）を備え、表示制御手段は、分類軸候補を優先度順に表示し、分類軸候補を選択することによって分類軸を切り替えて表示し、分類軸内の分類項目を選択することによって、データ群を選択又は絞り込み可能であるように構成されていてもよい。

[0256] （6）データ分類システムは、データ群を検索キーワードに基づいて検索し、データ群を検索した結果を分類軸候補絞込手段に出力するデータ検索手

段（例えば、検索手段 106 によって実現される）を備え、分類軸絞込手段は、データ検索手段の検索結果に基づいて分類軸候補を絞り込み、優先度計算手段は、データ検索手段が検索したデータ群に対応する分類軸の優先度を計算するように構成されていてもよい。

[0257] （7）データ分類システムは、階層的な分類項目とデータ群とを入力し、入力した分類項目とデータとの対応付けを行うデータ対応付け手段（例えば、データ対応付け手段 107 によって実現される）を備えるように構成されていてもよい。

[0258] （8）データ分類システムは、階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、データ群に対応する分類項目を複数選択して分類軸を作成し、複数の分類軸の組合せを多次元分類軸として出力するデータ分類システムであって、分類項目を選択するための基準項目となる分類項目群を予め蓄積する基準項目蓄積手段（例えば、基準項目蓄積部 202 によって実現される）と、基準項目の子孫の分類項目においてデータに少なくとも 1 つ対応する分類項目の組合せに基づいて分類軸を作成する際に、分類項目に対応するデータ量と、分類項目間の階層的な距離と、に基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込手段（例えば、分類軸候補絞込み手段 1021 によって実現される）と、分類軸絞込手段が絞り込んだ分類軸候補を組合せることによって、多次元の分類軸候補を作成する多次元分類軸作成手段（例えば、多次元分類候補作成手段 1023 によって実現される）と、分類階層における分類項目間の階層的な距離に基づいて、多次元分類軸候補の優先度を計算する多次元優先度計算手段（例えば、多次元指標計算手段 1032 によって実現される）と、を備えるように構成されていてもよい。

[0259] （9）データ分類システムにおいて、分類軸絞込手段は、各次元の分類軸の分類項目に対応するデータ量が所定数以上であるか、又は上位所定割合以内である分類項目を含み、且つ各次元の分類軸の分類項目間の階層的な距離として分類項目間の共通祖先までの長さが特定の範囲内となる分類項目を含

む多次元分類軸を選択するように構成されていてもよい。

- [0260] (10) データ分類システムにおいて、多次元優先度計算手段は、各次元の分類軸の分類階層における分類項目間の共通の先祖までの長さに応じて、多次元分類軸の優先度を変更するように構成されていてもよい。
- [0261] (11) データ分類システムにおいて、多次元優先度計算手段は、各次元の分類軸の分類階層における分類項目間の階層的な距離を計算し、更に、各次元の分類軸の分類項目の分類階層での深さ、分類項目群に対応するデータ量、又は分類項目群に対応するデータの重複度合いのうちの少なくともいずれか1つに基づいて、多次元分類軸候補の優先度を決定するように構成されていてもよい。
- [0262] (12) データ分類システムは、分類軸絞込手段が絞り込んだ多次元分析軸候補と、多次元優先度計算手段が計算した優先度とを読み込み、データ群とともに一覧形式又は表形式で表示制御する多次元表示制御手段（例えば、多次元表示手段1051によって実現される）を備え、多次元表示制御手段は、多次元分析軸候補を選択し、各次元の分類項目を表形式又は一覧形式に配置して表示し、1つ又は複数の分類項目を選択することによって、選択した分類項目に対応するデータの数、データ名、データ属性又は特徴語のうちの少なくともいずれか1つを表示するように構成されていてもよい。
- [0263] (13) データ分類システムは、データ群を検索キーワードに基づいて検索し、データ群を検索した結果を多次元分類軸候補絞込手段に出力するデータ検索手段（例えば、検索手段106によって実現される）を備えるように構成されていてもよい。
- [0264] (14) データ分類システムは、階層的な分類項目とデータ群とを入力し、入力した分類項目とデータとの対応付けを行うデータ対応付け手段（例えば、データ対応付け手段107によって実現される）を備えるように構成されていてもよい。
- [0265] 本発明は、上述の実施形態に限定されず、本発明の主旨を逸脱しない範囲で、適宜修正や変更が可能である。

産業上の利用可能性

[0266] 本発明は、大量文書群の概要把握を支援するための文書分類装置や、このような文書分類装置を実現するためのプログラムの用途に適用できる。また、本発明は、大量の画像を分類して表示する分類表示装置や、かかる分類表示装置を実現するためのプログラムの用途にも適用可能である。

符号の説明

- [0267]
- 101 入力手段
 - 103 指標計算手段
 - 104 出力手段
 - 105 表示手段
 - 106 検索手段
 - 107 データ対応付け手段
 - 201 分類階層蓄積部
 - 202 基準項目蓄積部
 - 203 データ蓄積部
 - 1021 分類軸候補絞込み手段
 - 1023 多次元分類候補作成手段
 - 1031 第2の指標計算手段
 - 1032 多次元指標計算手段
 - 1033 第2の多次元指標計算手段
 - 1051 多次元表示手段

請求の範囲

- [請求項1] 階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、前記データ群に対応する分類項目を複数選択して分類軸として出力するデータ分類システムであって、
- 前記分類項目を選択するための基準項目となる分類項目群を予め蓄積する基準項目蓄積手段と、
- 前記基準項目の子孫の分類項目から複数の分類項目を選択して分類軸候補とし、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込手段と、
- 前記分類軸絞込手段が絞り込んだ分類軸候補について、当該分類軸候補を表示する優先度を計算する優先度計算手段と
- を備える、データ分類システム。
- [請求項2] 請求項1に記載のデータ分類システムであって、
- 分類軸絞込手段は、分類項目に対応するデータ量が所定数以上であること、又は前記分類項目に対応するデータ量が上位所定割合以内であることのうちの少なくともいずれか一方の条件と、分類項目間の共通祖先までの長さが特定の範囲内であることという条件とを満たす分類軸を選択する、データ分類システム。
- [請求項3] 請求項1に記載のデータ分類システムであって、
- 優先度計算手段は、分類階層における分類項目間の共通の先祖までの長さに応じて、優先度を決定する、データ分類システム。
- [請求項4] 請求項1に記載のデータ分類システムであって、
- 優先度計算手段は、分類階層における分類項目間の階層的な距離を計算し、更に、分類項目の分類階層での深さ、分類項目群に対応するデータ量、又は分類項目群に対応するデータの重複度合いのうちの少なくともいずれか1つに基づいて、分類軸候補の優先度を決定する、データ分類システム。

- [請求項5] 請求項1に記載のデータ分類システムであって、
分類軸絞込手段が絞り込んだ分析軸候補と、優先度計算手段が計算した優先度とを読み込み、データ群とともに表示制御する表示制御手段を更に備え、
前記表示制御手段は、前記分類軸候補を優先度順に表示し、前記分類軸候補を選択することによって分類軸を切り替えて表示し、分類軸内の分類項目を選択することによって、データ群を選択又は絞り込み可能である、データ分類システム。
- [請求項6] 請求項1に記載のデータ分類システムであって、
データ群を検索キーワードに基づいて検索し、前記データ群を検索した結果を分類軸候補絞込手段に出力するデータ検索手段を備え、
前記分類軸絞込手段は、前記データ検索手段の検索結果に基づいて分類軸候補を絞り込み、
優先度計算手段は、前記データ検索手段が検索したデータ群に対応する分類軸の優先度を計算する、データ分類システム。
- [請求項7] 請求項1に記載のデータ分類システムであって、
階層的な分類項目とデータ群とを入力し、入力した分類項目とデータとの対応付けを行うデータ対応付け手段を備える、データ分類システム。
- [請求項8] 階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、前記データ群に対応する分類項目を複数選択して分類軸を作成し、複数の分類軸の組合せを多次元分類軸として出力するデータ分類システムであって、
前記分類項目を選択するための基準項目となる分類項目群を予め蓄積する基準項目蓄積手段と、
前記基準項目の子孫の分類項目においてデータに少なくとも1つ対応する分類項目の組合せに基づいて分類軸を作成する際に、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計

算対象となる分類軸候補を絞り込む分類軸絞込手段と、

前記分類軸絞込手段が絞り込んだ分類軸候補を組合せることによって、多次元分類軸候補を作成する多次元分類軸作成手段と、

分類階層における分類項目間の階層的な距離に基づいて、多次元分類軸候補の優先度を計算する多次元優先度計算手段とを備える、データ分類システム。

[請求項9]

請求項8に記載のデータ分類システムであって、

分類軸絞込手段は、各次元の分類軸の分類項目に対応するデータ量が所定数以上であるか、又は上位所定割合以内である分類項目を含み、且つ各次元の分類軸の分類項目間の階層的な距離として分類項目間の共通祖先までの長さが特定の範囲内となる分類項目を含む多次元分類軸を選択する、データ分類システム。

[請求項10]

請求項8に記載のデータ分類システムであって、

多次元優先度計算手段は、各次元の分類軸の分類階層における分類項目間の共通の先祖までの長さに応じて、多次元分類軸の優先度を変更する、データ分類システム。

[請求項11]

請求項8に記載のデータ分類システムであって、

多次元優先度計算手段は、各次元の分類軸の分類階層における分類項目間の階層的な距離を計算し、更に、各次元の分類軸の分類項目の分類階層での深さ、分類項目群に対応するデータ量、又は分類項目群に対応するデータの重複度合いのうちの少なくともいずれか1つに基づいて、多次元分類軸候補の優先度を決定する、データ分類システム。

[請求項12]

請求項8に記載のデータ分類システムであって、

分類軸絞込手段が絞り込んだ多次元分析軸候補と、多次元優先度計算手段が計算した優先度とを読み込み、データ群とともに一覧形式又は表形式で表示制御する多次元表示制御手段を備え、

前記多次元表示制御手段は、前記多次元分析軸候補を選択し、各次

元の分類項目を表形式又は一覧形式に配置して表示し、1つ又は複数の分類項目を選択することによって、選択した分類項目に対応するデータの数、データ名、データ属性又は特徴語のうちの少なくともいずれか1つを表示する、データ分類システム。

[請求項13] 請求項8に記載のデータ分類システムであって、
データ群を検索キーワードに基づいて検索し、前記データ群を検索した結果を多次元分類軸候補絞込手段に出力するデータ検索手段を更に備える、データ分類システム。

[請求項14] 請求項8に記載のデータ分類システムであって、
階層的な分類項目とデータ群とを入力し、入力した分類項目とデータとの対応付けを行うデータ対応付け手段を更に備える、データ分類システム。

[請求項15] 階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、前記データ群に対応する分類項目を複数選択して分類軸として出力するデータ分類方法であって、

前記分類項目を選択するための基準項目となる分類項目群を予めデータベースに蓄積し、前記基準項目の子孫の分類項目から複数の分類項目を選択して分類軸候補とし、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込行程と、

絞り込んだ前記分類軸候補について、当該分類軸候補を表示する優先度を計算する優先度計算行程と
を含む、データ分類方法。

[請求項16] 請求項15に記載のデータ分類方法であって、
分類軸絞込行程は、分類項目に対応するデータ量が所定数以上であること、又は前記分類項目に対応するデータ量が上位所定割合以内であることのうちの少なくともいずれか一方の条件と、分類項目間の共通祖先までの長さが特定の範囲内であることという条件とを満たす分

類軸を選択する、データ分類方法。

[請求項17]

請求項15に記載のデータ分類方法であって、

優先度計算行程は、分類階層における分類項目間の共通の先祖までの長さに応じて、優先度を決定する、データ分類方法。

[請求項18]

請求項15に記載のデータ分類方法であって、

優先度計算行程は、分類階層における分類項目間の階層的な距離を計算し、更に、分類項目の分類階層での深さ、分類項目群に対応するデータ量、又は分類項目群に対応するデータの重複度合いのうちの少なくともいずれか1つに基づいて、分類軸候補の優先度を決定する、データ分類方法。

[請求項19]

階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、前記データ群に対応する分類項目を複数選択して分類軸を作成し、複数の分類軸の組合せを多次元分類軸として出力するデータ分類方法であって、

前記分類項目を選択するための基準項目となる分類項目群を予めデータベースに蓄積し、前記基準項目の子孫の分類項目においてデータに少なくとも1つ対応する分類項目の組合せに基づいて分類軸を作成する際に、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込行程と、

絞り込んだ前記分類軸候補を組合せることによって、多次元の分類軸候補を作成する多次元分類軸作成行程と、

分類階層における分類項目間の階層的な距離に基づいて、多次元分類軸候補の優先度を計算する多次元優先度計算行程とを含む、データ分類方法。

[請求項20]

請求項19に記載のデータ分類方法であって、

分類軸絞込行程は、各次元の分類軸の分類項目に対応するデータ量が所定数以上であるか、又は上位所定割合以内である分類項目を含み

、且つ各次元の分類軸の分類項目間の階層的な距離として分類項目間の共通祖先までの長さが特定の範囲内となる分類項目を含む多次元分類軸を選択する、データ分類方法。

[請求項21]

請求項19に記載のデータ分類方法であって、

多次元優先度計算行程は、各次元の分類軸の分類階層における分類項目間の共通の先祖までの長さに応じて、多次元分類軸の優先度を変更する、データ分類方法。

[請求項22]

請求項19に記載のデータ分類方法であって、

多次元優先度計算行程は、各次元の分類軸の分類階層における分類項目間の階層的な距離を計算し、更に、各次元の分類軸の分類項目の分類階層での深さ、分類項目群に対応するデータ量、又は分類項目群に対応するデータの重複度合いのうちの少なくともいずれか1つに基づいて、多次元分類軸候補の優先度を決定する、データ分類方法。

[請求項23]

階層的な分類項目と、当該分類項目に対応するデータ群とに基づいて、前記データ群に対応する分類項目を複数選択して分類軸として出力するためのデータ分類プログラムであって、

前記分類項目を選択するための基準項目となる分類項目群を予め蓄積する基準項目蓄積手段を備えたコンピュータに、

前記基準項目の子孫の分類項目から複数の分類項目を選択して分類軸候補とし、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込処理と、

絞り込んだ前記分類軸候補について、当該分類軸候補を表示する優先度を計算する優先度計算処理と

を実行させる、データ分類プログラム。

[請求項24]

請求項23に記載のデータ分類プログラムであって、

コンピュータに、

分類軸絞込処理にて、分類項目に対応するデータ量が所定数以上で

あること、又は前記分類項目に対応するデータ量が上位所定割合以内であることのうちの少なくともいずれか一方の条件と、分類項目間の共通祖先までの長さが特定の範囲内であることという条件とを満たす分類軸を選択する処理を実行させる、データ分類プログラム。

[請求項25]

請求項23に記載のデータ分類プログラムであって、
コンピュータに、

分類軸絞込処理で絞り込んだ分析軸候補と、優先度計算処理で計算した優先度とを読み込み、データ群とともに表示制御する表示制御処理を実行させ、

前記表示制御処理にて、前記分類軸候補を優先度順に表示し、前記分類軸候補を選択することによって分類軸を切り替えて表示し、分類軸内の分類項目を選択することによって、データ群を選択又は絞り込み可能に処理を実行させる、データ分類プログラム。

[請求項26]

請求項23に記載のデータ分類プログラムであって、
コンピュータに、

データ群を検索キーワードに基づいて検索し、前記データ群を検索した結果を出力するデータ検索処理を実行させ、

分類軸絞込処理にて、前記データ検索処理の検索結果に基づいて分類軸候補を絞り込む処理を実行させ、

優先度計算処理にて、前記データ検索処理で検索したデータ群に対応する分類軸の優先度を計算する処理を実行させる、データ分類プログラム。

[請求項27]

請求項23に記載のデータ分類プログラムであって、
コンピュータに、

階層的な分類項目とデータ群とを入力し、入力した分類項目とデータとの対応付けを行うデータ対応付け処理を実行させる、データ分類プログラム。

[請求項28]

階層的な分類項目と、当該分類項目に対応するデータ群とに基づい

て、前記データ群に対応する分類項目を複数選択して分類軸を作成し、複数の分類軸の組合せを多次元分類軸として出力するためのデータ分類プログラムであって、

前記分類項目を選択するための基準項目となる分類項目群を予め蓄積する基準項目蓄積手段を備えたコンピュータに、

前記基準項目の子孫の分類項目においてデータに少なくとも1つ対応する分類項目の組合せに基づいて分類軸を作成する際に、分類項目に対応するデータ量と、分類項目間の階層的な距離とに基づいて、計算対象となる分類軸候補を絞り込む分類軸絞込処理と、

絞り込んだ前記分類軸候補を組合せることによって、多次元の分類軸候補を作成する多次元分類軸作成処理と、

分類階層における分類項目間の階層的な距離に基づいて、多次元分類軸候補の優先度を計算する多次元優先度計算処理とを実行させる、データ分類プログラム。

[請求項29]

請求項28に記載のデータ分類プログラムであって、

コンピュータに、

分類軸絞込処理にて、各次元の分類軸の分類項目に対応するデータ量が所定数以上であるか、又は上位所定割合以内である分類項目を含み、且つ各次元の分類軸の分類項目間の階層的な距離として分類項目間の共通祖先までの長さが特定の範囲内となる分類項目を含む多次元分類軸を選択する処理を実行させる、データ分類プログラム。

[請求項30]

請求項28に記載のデータ分類プログラムであって、

コンピュータに、

分類軸絞込処理で絞り込んだ多次元分析軸候補と、多次元優先度計算処理で計算した優先度とを読み込み、データ群とともに一覧形式又は表形式で表示制御する多次元表示制御処理を実行させ、

前記多次元表示制御処理で、前記多次元分析軸候補を選択し、各次元の分類項目を表形式又は一覧形式に配置して表示し、1つ又は複数

の分類項目を選択することによって、選択した分類項目に対応するデータの数、データ名、データ属性又は特徴語のうちの少なくともいずれか1つを表示する処理を実行させる、データ分類プログラム。

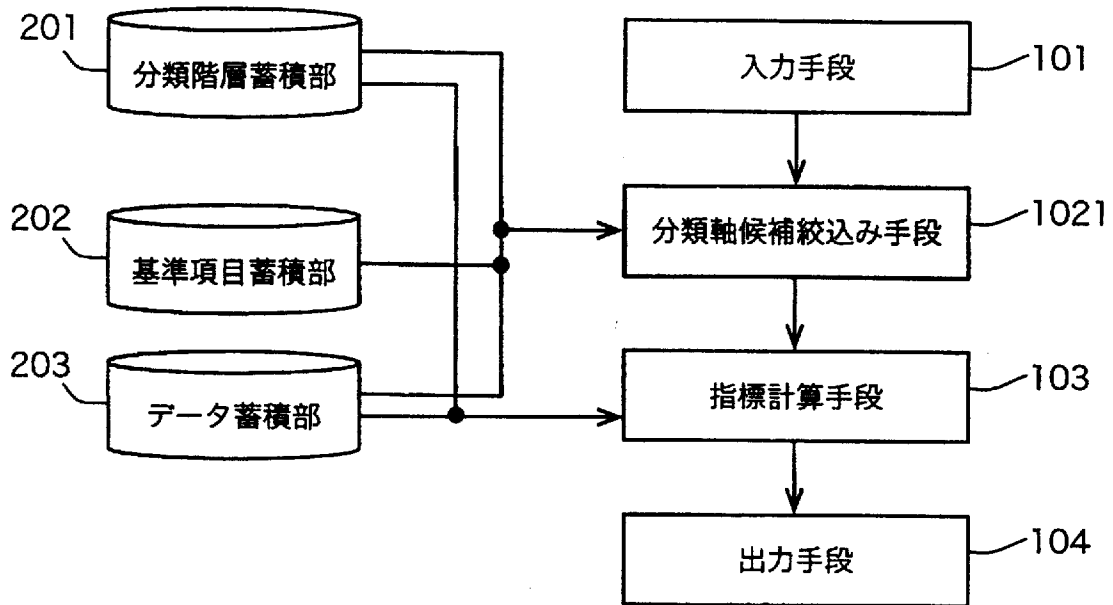
[請求項31]

請求項28に記載のデータ分類プログラムであって、
コンピュータに、
データ群を検索キーワードに基づいて検索し、前記データ群を検索した結果を出力するデータ検索処理を実行させる、データ分類プログラム。

[請求項32]

請求項28に記載のデータ分類プログラムであって、
コンピュータに、
階層的な分類項目とデータ群とを入力し、入力した分類項目とデータとの対応付けを行うデータ対応付け処理を実行させる、データ分類プログラム。

[図1]



[図2]

親分類項目	子分類項目群
主要カテゴリ	社会、自然、...
社会	生活、政治、医療
生活	家庭、家族、健康
政治	外交、行政
医療	医学、移植
...	...

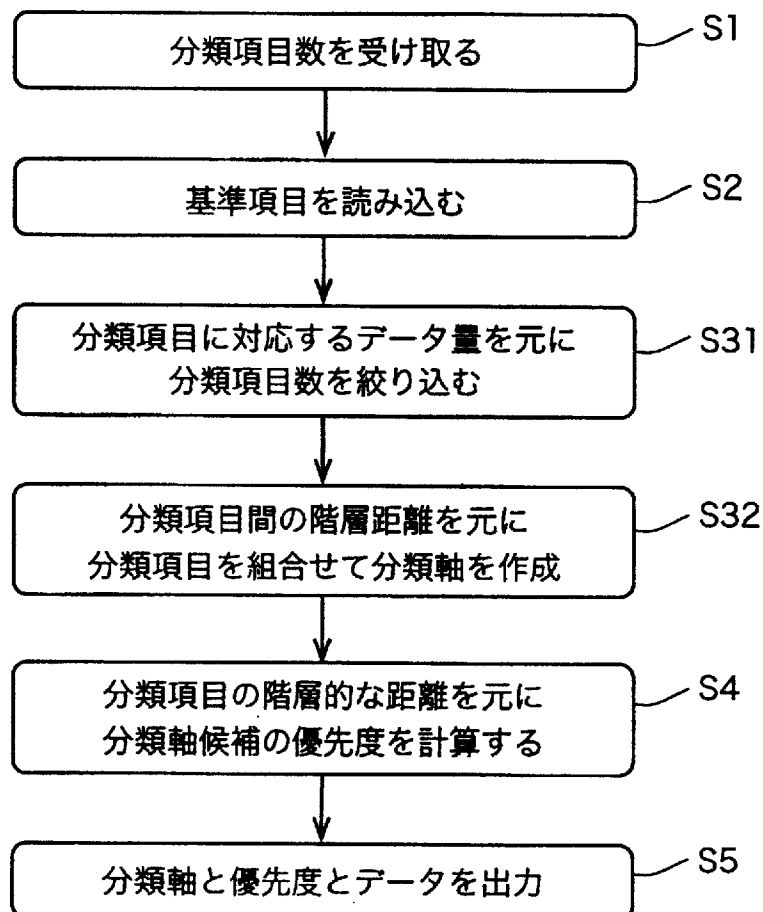
[図3]

基準項目
社会
自然
文化
...

[図4]

データID	コンテンツ	対応分類項目
d1	...	生活、家族、家庭
d2	...	家族、家庭、健康、医学
d3	...	家庭、家族
d4	...	健康、外交、医学、移植
d5	...	外交
d6	...	医学、移植

[図5]



[図6]

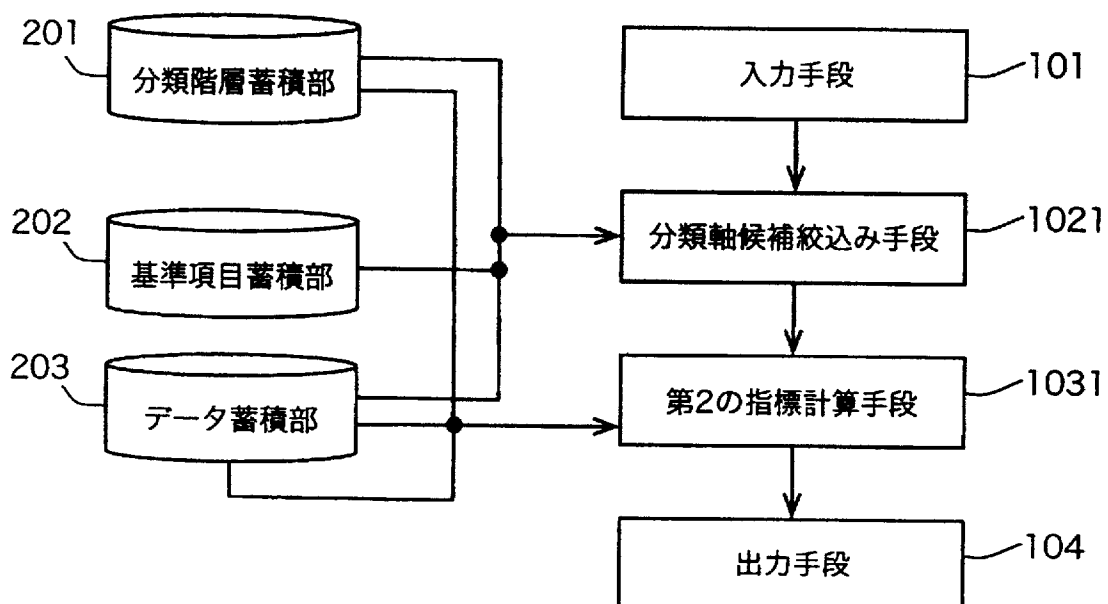
分類軸ID	基準項目	分類項目群	スコア
1	社会	家庭、家族、政治	...
2	社会	家庭、家族、外交	...
3	社会	家庭、家族、医療	0.66
...	

(A)

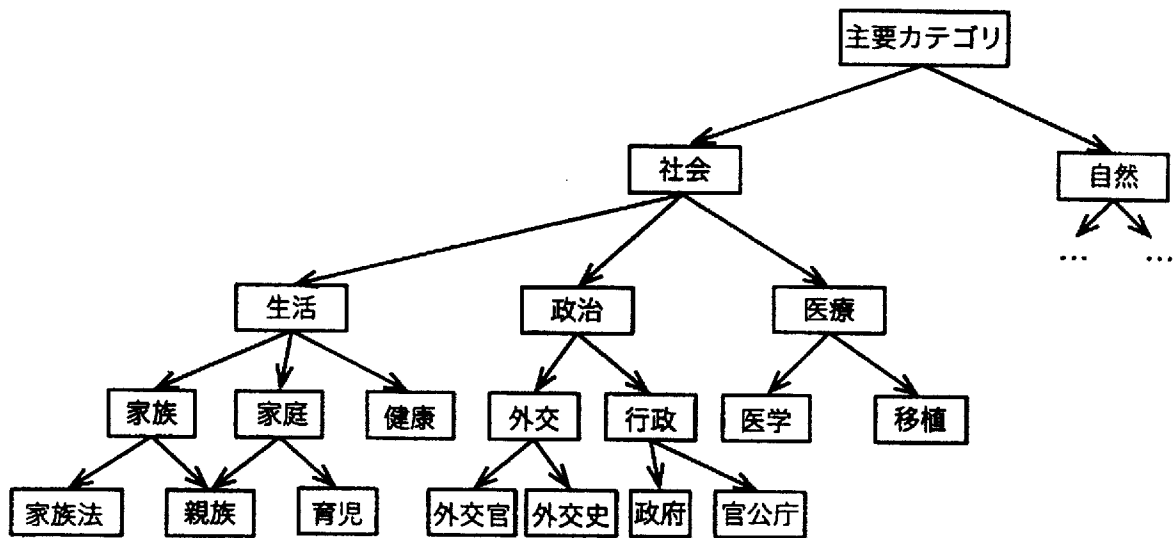
分類軸ID	分類項目	データID群
1	家庭	d1,d2,d3
1	家族	d1,d2,d3
1	政治	d4,d5
1	その他	d6
...

(B)

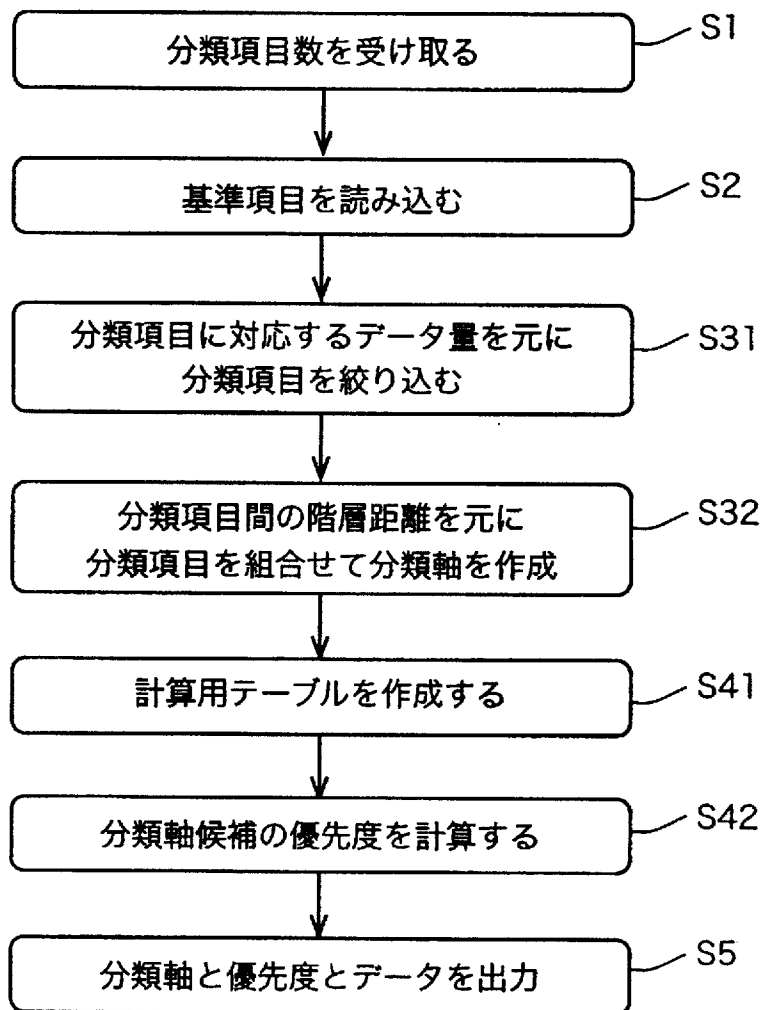
[図7]



[図8]



[図9]



[図10]

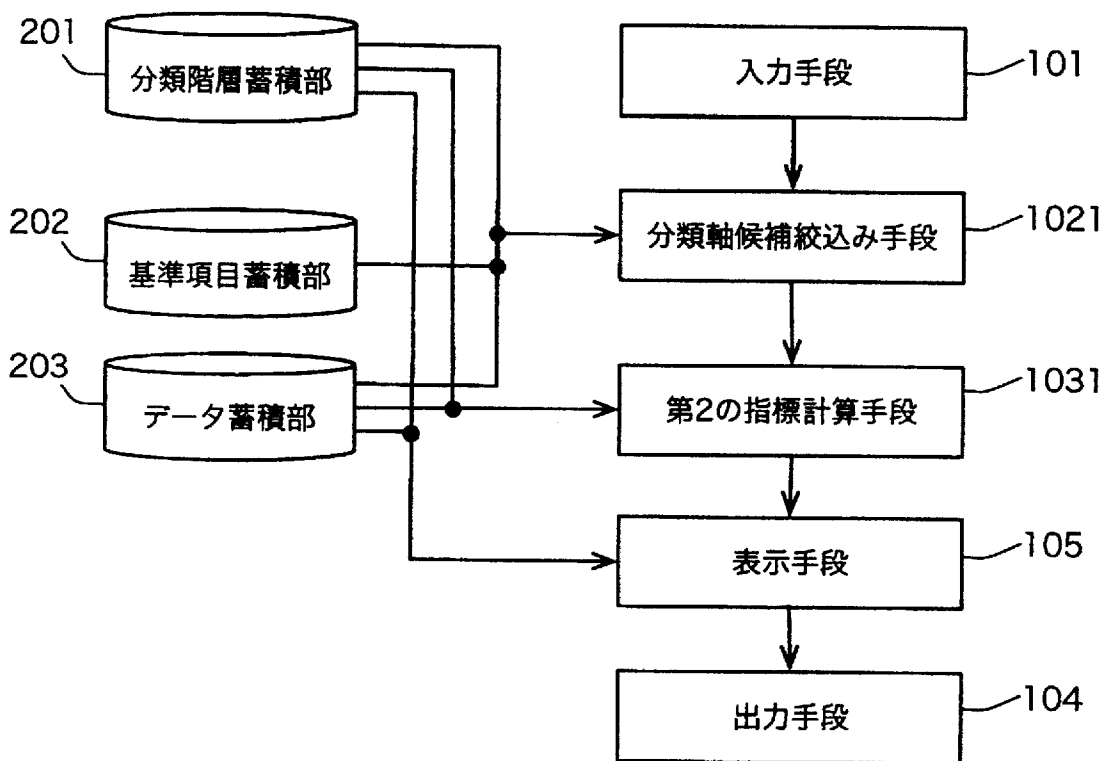
分類項目	データ数	分類項目の深さ
家族	3	2
外交	2	2
医療	3	1

(A)

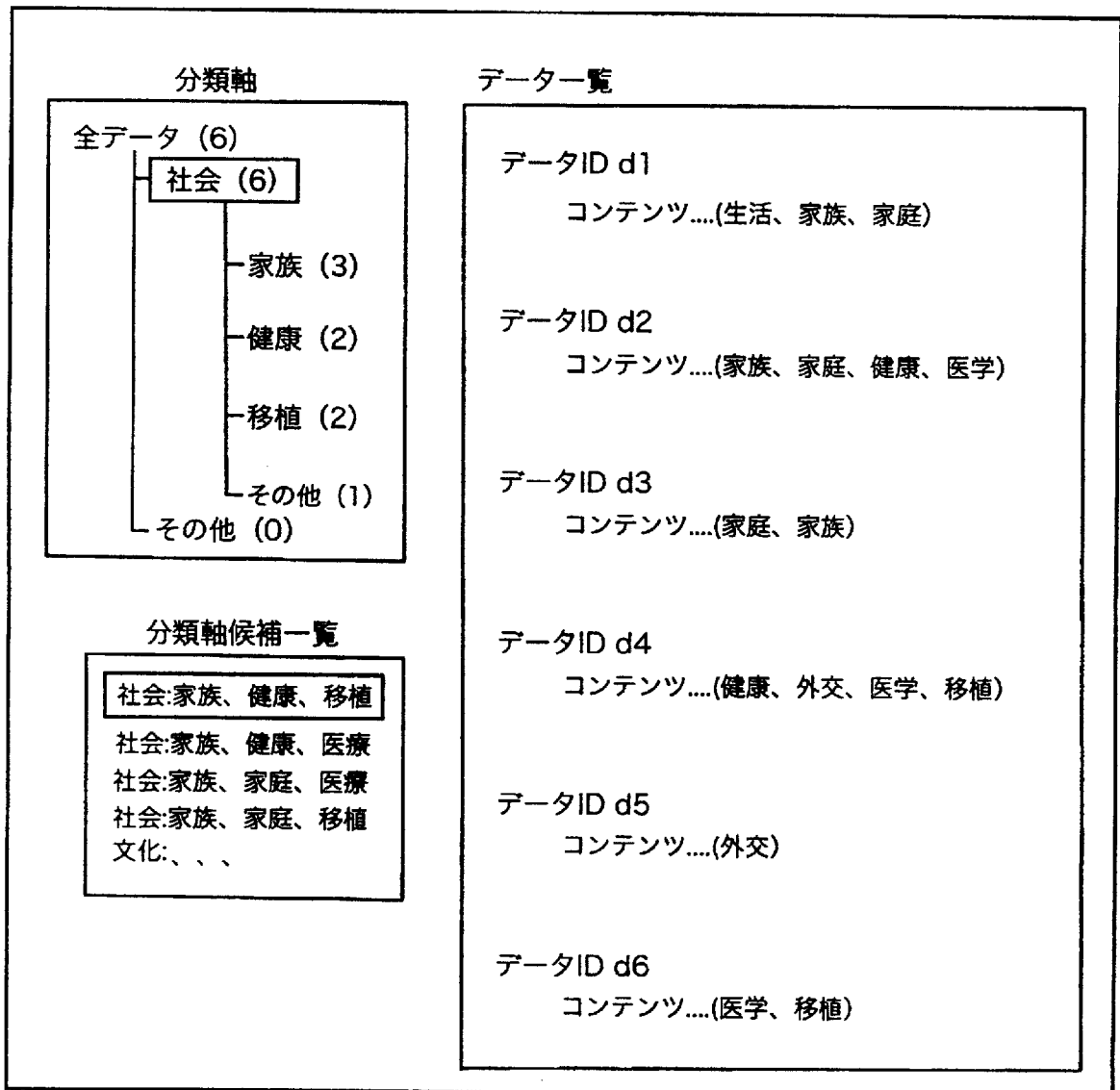
データID	分類項目数
d1	1
d2	2
d3	1
d4	2
d5	1
d6	1

(B)

[図11]



[図12]



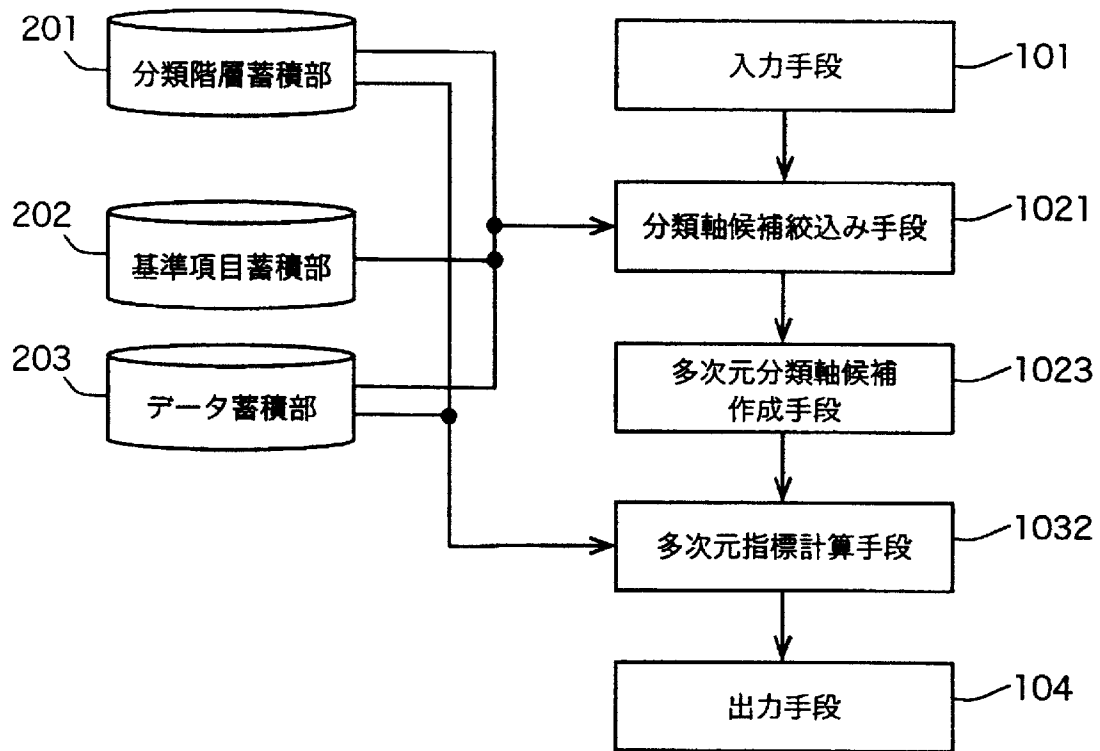
[図13]

分類軸 関連属性		社会(6)				その他(0)
		家族(3)	健康(2)	移植(2)	その他(1)	
生活(4)		d1,d2,d3 3件	d2,d4 2件	d4 1件	0件	0件
家庭(3)		d1,d2,d3 3件	d2 1件	0件	0件	0件
医療(3)		d2 1件	d2 1件	d4,d6 2件	0件	0件
その他(1)		0件	0件	0件	d5 1件	0件

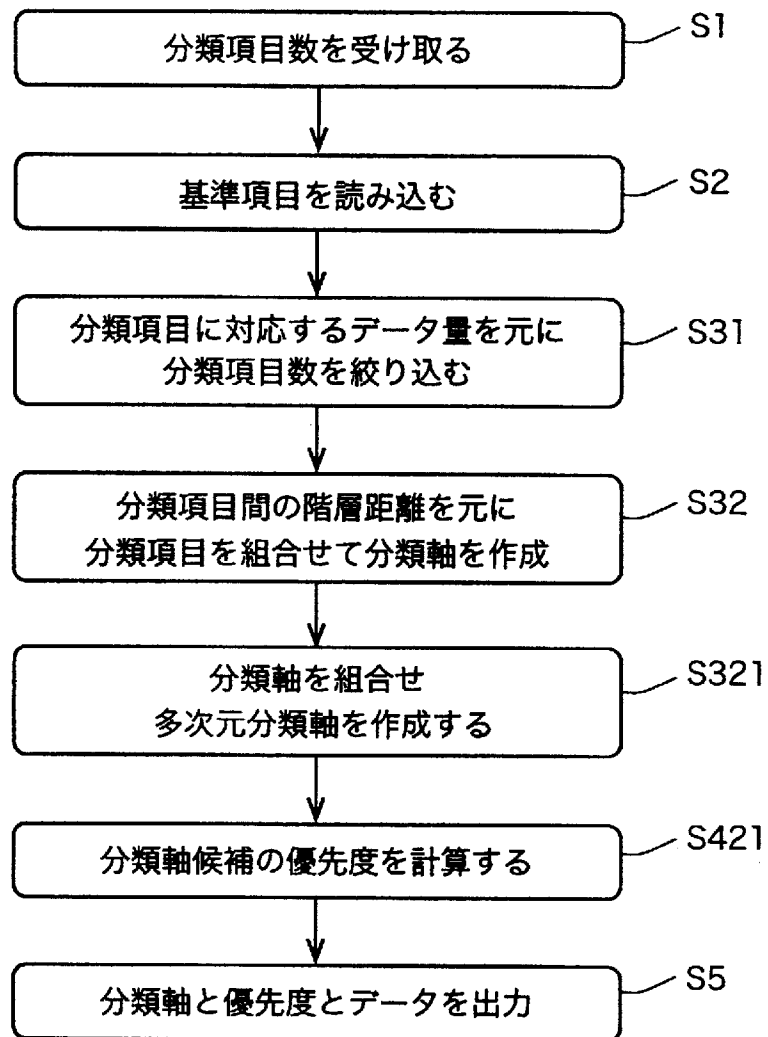
データ一覧	
データID d1	コンテンツ...(生活、家族、家庭)
データID d2	コンテンツ...(家族、家庭、健康、医学)
データID d3	コンテンツ...(家庭、家族)

分類軸候補一覧	
社会:家族、健康、移植	
社会:家族、健康、医療	
社会:家族、家庭、医療	
社会:家族、家庭、移植	
文化:、、、	

[図14]



[図15]



[図16]

(A)

次元ID	分類軸ID群	スコア
1	1,2	0.67
2	3,4	0.5
...

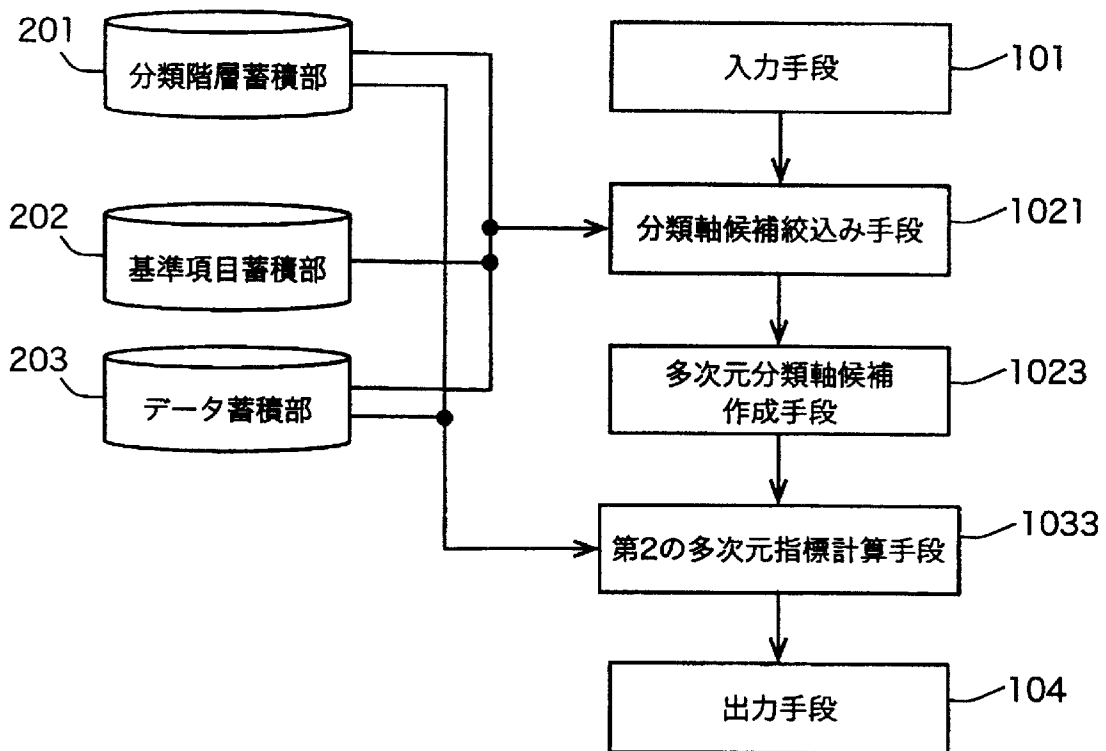
(B)

分類軸ID	基準項目	分類項目群
1	社会	家庭、家族、健康
2	社会	家庭、家族、政治
3	社会	家庭、家族、外交
4	社会	家庭、家族、医療
...

(C)

分類軸ID	分類項目	データID群
1	家庭	d1,d2,d3
1	家族	d1,d2,d3
1	健康	d2,d4
1	その他	d5,d6
...

[図17]



[図18]

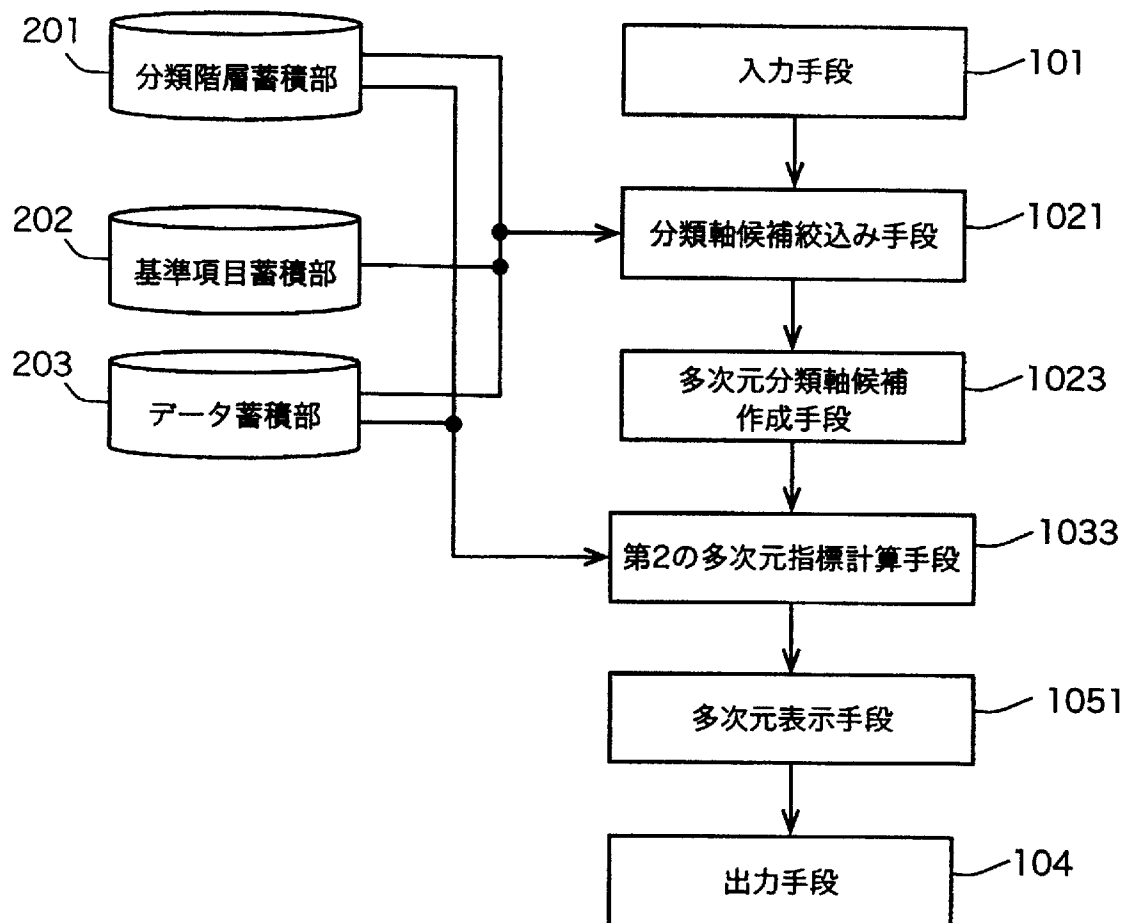
分類項目 組合せ	データ数	分類項目の深さ
家族、外交	0	1
家族、医学	1	1
家族、移植	0	1
家庭、外交	0	1
家庭、医学	1	1
家庭、移植	0	1
健康、外交	1	1
健康、医学	2	1
健康、移植	1	1

データID	分類項目 組合せ数
d1	0
d2	1
d3	0
d4	2
d5	0
d6	0

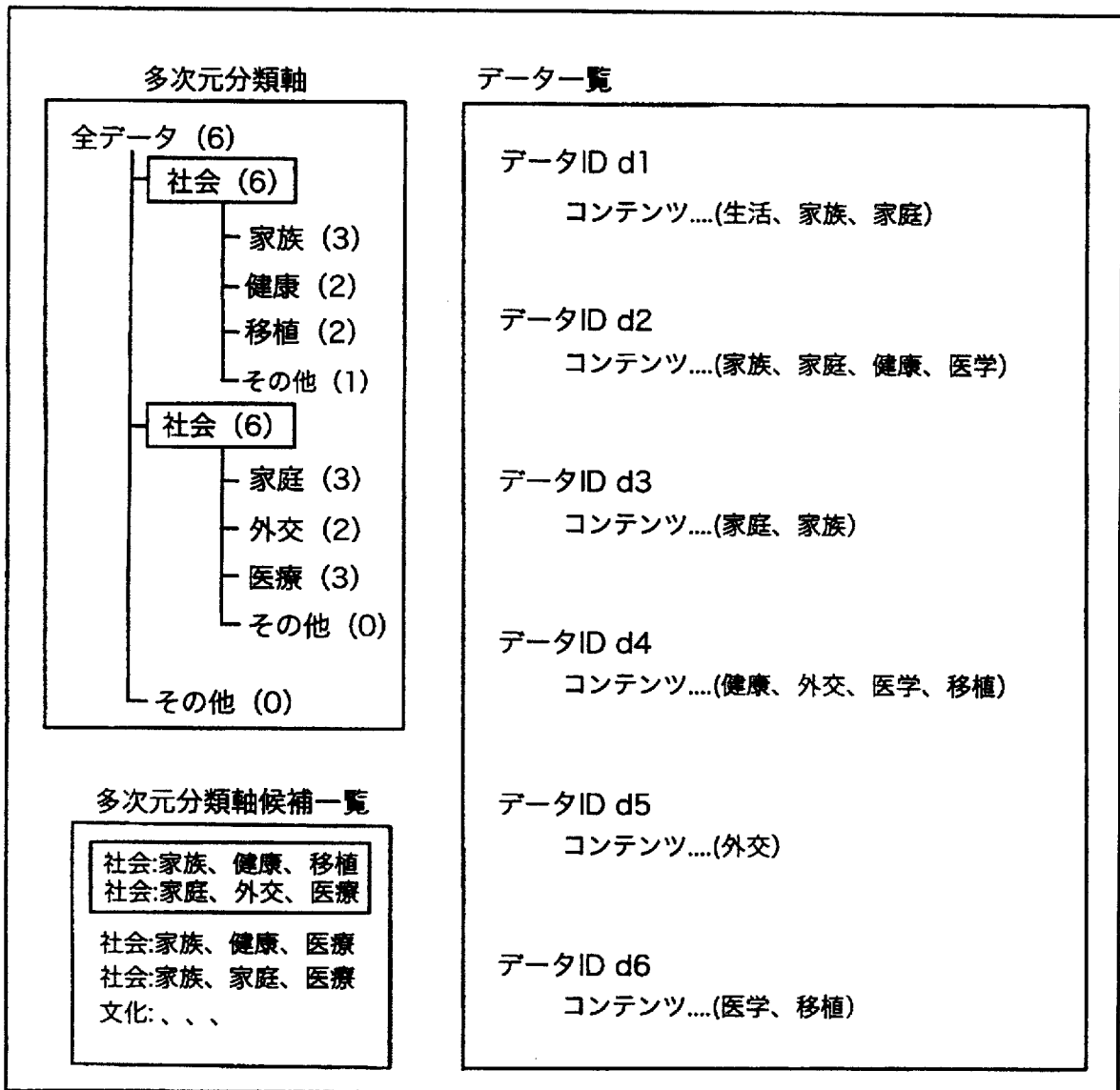
(A)

(B)

[図19]



[図20]



[図21]

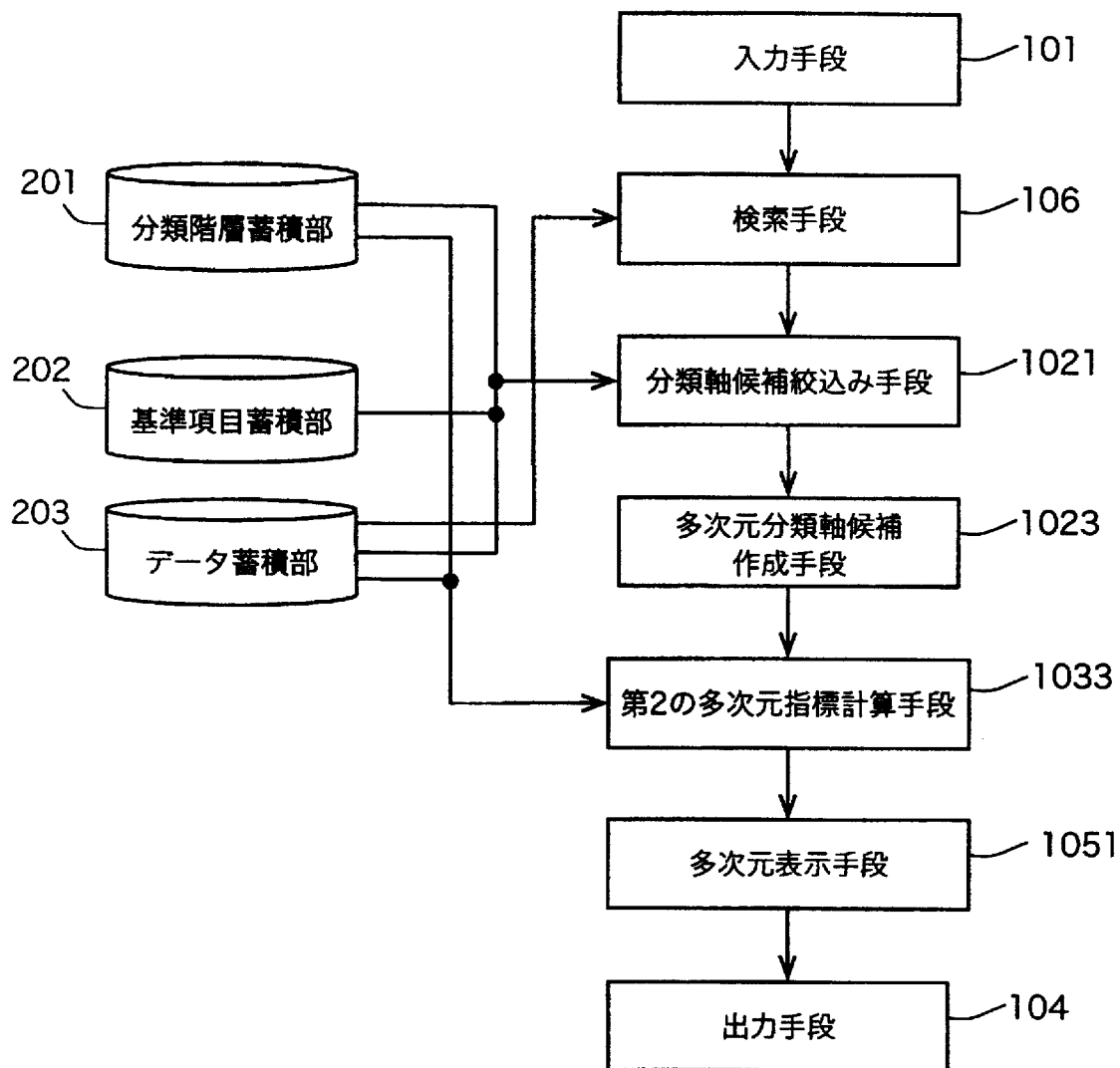
分類軸		社会(6)				その他(0)
		家族(3)	健康(2)	移植(2)	その他(1)	
社会 (6)	家庭(3)	d1,d2,d3 3件	d2 1件	0件	0件	0件
	外交(2)	0件	d4 1件	d4 1件	d5 1件	0件
	医療(3)	d2 1件	d2,d4 2件	d4,d6 2件	0件	0件
	その他(0)	0件	0件	0件	1件	0件
	その他(0)	0件	0件	0件	0件	0件

データ一覧	
データID d1	コンテンツ...(生活、家族、家庭)
データID d2	コンテンツ...(家族、家庭、健康、医学)
データID d3	コンテンツ...(家庭、家族)

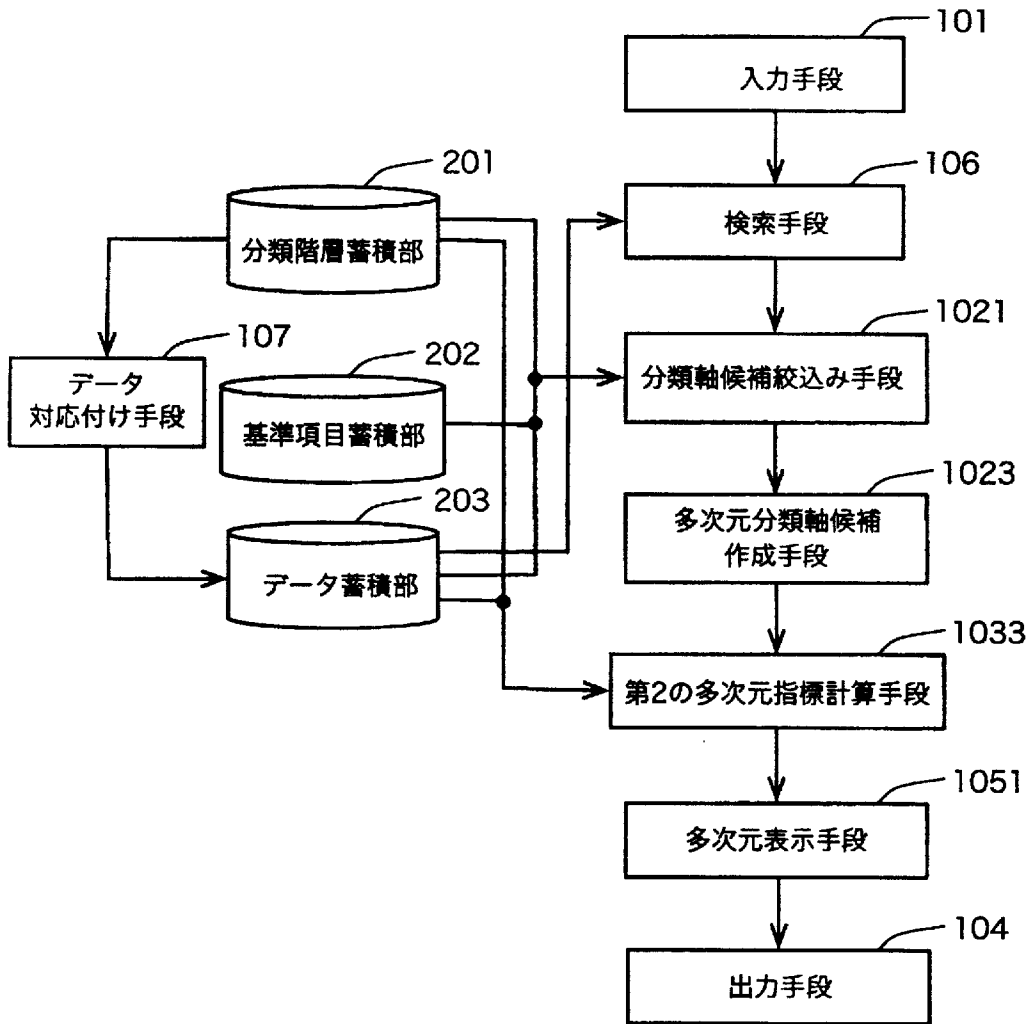
多次元分類軸候補一覧

- 社会:家族、健康、移植
- 社会:家族、外交、医療
- 社会:家族、健康、医療
- 社会:家族、家庭、医療
- 文化:、、、

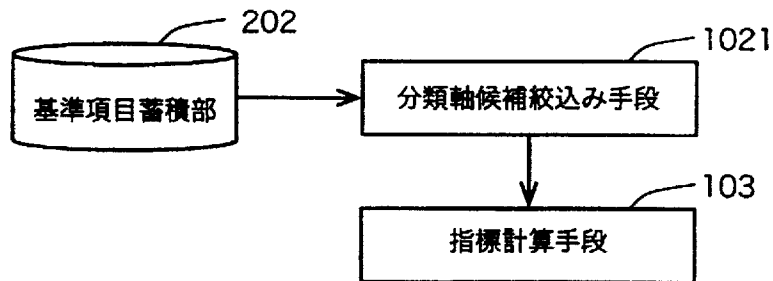
[図22]



[図23]



[図24]



INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2009/003601

A. CLASSIFICATION OF SUBJECT MATTER
G06F17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2009
Kokai Jitsuyo Shinan Koho	1971-2009	Toroku Jitsuyo Shinan Koho	1994-2009

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2005-202535 A (Hitachi, Ltd.), 28 July 2005 (28.07.2005), entire text & US 2005/0165819 A1	1-32
A	JP 2007-102309 A (Mitsubishi Electric Corp.), 19 April 2007 (19.04.2007), entire text (Family: none)	1-32
A	JP 2006-171931 A (Mitsubishi Electric Corp.), 29 June 2006 (29.06.2006), entire text (Family: none)	1-32

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 08 October, 2009 (08.10.09)	Date of mailing of the international search report 20 October, 2009 (20.10.09)
------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------

Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer
Facsimile No.	Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC)) Int.Cl. G06F17/30(2006.01)i										
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) Int.Cl. G06F17/30										
最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922-1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971-2009年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996-2009年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994-2009年</td> </tr> </table>			日本国実用新案公報	1922-1996年	日本国公開実用新案公報	1971-2009年	日本国実用新案登録公報	1996-2009年	日本国登録実用新案公報	1994-2009年
日本国実用新案公報	1922-1996年									
日本国公開実用新案公報	1971-2009年									
日本国実用新案登録公報	1996-2009年									
日本国登録実用新案公報	1994-2009年									
国際調査で使用了電子データベース (データベースの名称、調査に使用した用語)										
C. 関連すると認められる文献										
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号								
A	JP 2005-202535 A (株式会社日立製作所) 2005.07.28, 全文 & US 2005/0165819 A1	1-32								
A	JP 2007-102309 A (三菱電機株式会社) 2007.04.19, 全文 (ファミリーなし)	1-32								
A	JP 2006-171931 A (三菱電機株式会社) 2006.06.29, 全文 (ファミリーなし)	1-32								
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input type="checkbox"/> パテントファミリーに関する別紙を参照。										
* 引用文献のカテゴリー 「A」 特に関連のある文献ではなく、一般的技术水準を示すもの 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す) 「O」 口頭による開示、使用、展示等に言及する文献 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献 「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの 「&」 同一パテントファミリー文献										
国際調査を完了した日 08.10.2009	国際調査報告の発送日 20.10.2009									
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 梅本 達雄 電話番号 03-3581-1101 内線 3599	5M 9648								