

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4179773号
(P4179773)

(45) 発行日 平成20年11月12日(2008.11.12)

(24) 登録日 平成20年9月5日(2008.9.5)

(51) Int.Cl. F 1
G 0 6 F 17/30 (2006.01)
 G 0 6 F 17/30 3 5 0 C
 G 0 6 F 17/30 3 4 0 B

請求項の数 7 (全 18 頁)

(21) 出願番号	特願2001-354310 (P2001-354310)	(73) 特許権者	000004329
(22) 出願日	平成13年11月20日(2001.11.20)		日本ビクター株式会社
(65) 公開番号	特開2003-157285 (P2003-157285A)		神奈川県横浜市神奈川区守屋町3丁目12番地
(43) 公開日	平成15年5月30日(2003.5.30)	(72) 発明者	穴戸 一郎
審査請求日	平成16年3月30日(2004.3.30)		神奈川県横浜市神奈川区守屋町3丁目12番地
審判番号	不服2006-13993 (P2006-13993/J1)	(72) 発明者	清水 徹
審判請求日	平成18年7月3日(2006.7.3)		神奈川県横浜市神奈川区守屋町3丁目12番地
		(72) 発明者	荒蔭 和美
			神奈川県横浜市神奈川区守屋町3丁目12番地
			日本ビクター株式会社内

最終頁に続く

(54) 【発明の名称】 情報検索装置、情報検索プログラム、及び情報検索方法

(57) 【特許請求の範囲】

【請求項1】

コンテンツが有する複数の特徴の度合いを数値化した特徴量を各次元要素とする多次元特徴量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索する情報検索装置であって、

前記複数のコンテンツを格納するコンテンツ格納手段と、

前記コンテンツ毎の前記多次元特徴量を格納する特徴量格納手段と、

前記コンテンツのうちの利用されたコンテンツの利用履歴を格納する利用履歴格納手段と、

前記コンテンツ格納手段に格納された複数のコンテンツの中から検索基準となる任意の1つのコンテンツを検索基準のコンテンツとして指定させるための条件入力手段と、

前記特徴量格納手段に格納されているコンテンツの多次元特徴量の各次元毎の平均値である第1の平均値を計算するとともに、前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前記利用履歴格納手段に格納されている前記利用されたコンテンツに対応する多次元特徴量の各次元毎の平均値である第2の平均値を計算し、これら計算した前記第1の平均値と前記第2の平均値との差分値が大きい次元ほど値が大きくなるような重み付けパラメータを各次元毎に計算し、

前記特徴量格納手段に格納された多次元特徴量の各次元の特徴量に対して、前記重み付けパラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及

10

20

び類似性のうちの少なくとも一方からなる関連度を計算する計算手段と、

前記計算手段によって計算された前記検索基準のコンテンツと前記他のコンテンツとの関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、前記コンテンツ格納手段に格納された前記他のコンテンツから選択する選択手段と、

を有する情報検索装置。

【請求項 2】

コンテンツが有する複数の特徴の度合いを数値化した特徴量を各次元要素とする多次元特徴量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索する情報検索装置であって、

前記複数のコンテンツを格納するコンテンツ格納手段と、

前記コンテンツ毎の前記多次元特徴量を格納する特徴量格納手段と、

外部から供給される前記コンテンツに対する利用者の肯定的な評価、及び否定的な評価に対応する、前記肯定的な評価を示す数値、及び前記否定的な評価を示す数値を少なくとも格納する評価値格納手段と、

前記コンテンツ格納手段に格納された複数のコンテンツの中から検索基準となる任意の1つのコンテンツを検索基準のコンテンツとして指定させるための条件入力手段と、

前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前記評価値格納手段に格納されている前記肯定的な評価を示す数値のコンテンツに対応する多次元特徴量の各次元毎の平均値である第1の平均値を計算するとともに、前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前記評価値格納手段に格納されている前記否

定的な評価を示す数値のコンテンツに対応する多次元特徴量の各次元毎の平均値である第2の平均値を計算し、これら計算した前記第1の平均値と前記第2の平均値との差分値が大きい次元ほど値が大きくなるような重み付けパラメータを各次元毎に計算し、前記特徴量格納手段に格納された多次元特徴量の各次元の特徴量に対して、前記重み付けパラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及び類似性のうちの少なくとも一方からなる関連度を計算する計算手段と、

前記計算手段によって計算された前記検索基準のコンテンツと前記他のコンテンツとの関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、前記コンテンツ格納手段に格納された前記他のコンテンツから選択する選択手段と、

を有する情報検索装置。

【請求項 3】

前記利用履歴格納手段は、前記利用者がコンテンツを利用した利用日時を格納し、

前記計算手段は、前記利用日時を用いて前記各重み付けパラメータを計算することを特徴とする請求項 1 に記載の情報検索装置。

【請求項 4】

コンテンツが有する複数の特徴の度合いを数値化した特徴量を各次元要素とする多次元特徴量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索する情報検索処理をコンピュータに実行させるプログラムであって、前記コンピュータを、

前記複数のコンテンツをコンテンツ格納手段に格納させる手段と、

前記コンテンツ毎の前記多次元特徴量を特徴量格納手段に格納させる手段と、

前記コンテンツのうちの利用されたコンテンツの利用履歴を利用履歴格納手段に格納させる手段と、

前記コンテンツ格納手段に格納された複数のコンテンツの中から検索基準となる任意の1つのコンテンツを検索基準のコンテンツとして指定させるための条件入力手段と、

前記特徴量格納手段に格納されているコンテンツの多次元特徴量の各次元毎の平均値である第1の平均値を計算するとともに、前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前記利用履歴格納手段に格納されている前記利用されたコンテンツに対応する多次元特徴量の各次元毎の平均値である第2の平均値を計算し、これら計算した前記第1の平均値と前記第2の平均値との差分値が大きい次元ほど値が大きくなるよう

10

20

30

40

50

な重み付けパラメータを各次元毎に計算し、
前記特徴量格納手段に格納された多次元特徴量の各次元の特徴量に対して、前記重み付け
パラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検
索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及
び類似性のうちの少なくとも一方からなる関連度を計算する計算手段と、

前記計算手段によって計算された前記検索基準のコンテンツと前記他のコンテンツとの
関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、前記コンテン
ツ格納手段に格納された前記他のコンテンツから選択する選択手段、
として機能させるための情報検索プログラム。

【請求項 5】

コンテンツが有する複数の特徴の度合いを数値化した特徴量を各次元要素とする多次元
特徴量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索す
る情報検索処理をコンピュータに実行させるプログラムであって、前記コンピュータを、
前記複数のコンテンツをコンテンツ格納手段に格納させる手段と、

前記コンテンツ毎の前記多次元特徴量を特徴量格納手段に格納させる手段と、

外部から供給される前記コンテンツに対する利用者の肯定的な評価、及び否定的な評価
に対応する、前記肯定的な評価を示す数値、及び前記否定的な評価を示す数値を少なくと
も評価値格納手段に格納させる手段と、

前記コンテンツ格納手段に格納された複数のコンテンツの中から検索基準となる任意の
1つのコンテンツを検索基準のコンテンツとして指定させるための条件入力手段と、

前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前記評価値格納
手段に格納されている前記肯定的な評価を示す数値のコンテンツに対応する多次元特徴量
の各次元毎の平均値である第1の平均値を計算するとともに、前記特徴量格納手段に格納
されているコンテンツの多次元特徴量のうち前記評価値格納手段に格納されている前記否
定的な評価を示す数値のコンテンツに対応する多次元特徴量の各次元毎の平均値である第
2の平均値を計算し、これら計算した前記第1の平均値と前記第2の平均値との差分値が
大きい次元ほど値が大きくなるような重み付けパラメータを各次元毎に計算し、

前記特徴量格納手段に格納された多次元特徴量の各次元の特徴量に対して、前記重み付け
パラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検
索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及
び類似性のうちの少なくとも一方からなる関連度を計算する計算手段と、

前記計算手段によって計算された前記検索基準のコンテンツと前記他のコンテンツとの
関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、前記コンテン
ツ格納手段に格納された前記他のコンテンツから選択する選択手段、
として機能させるための情報検索プログラム。

【請求項 6】

コンテンツが有する複数の特徴の度合いを数値化した特徴量を各次元要素とする多次元
特徴量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索す
る情報検索装置における情報検索方法であって、

前記情報検索装置は、

コンテンツ格納手段が、前記複数のコンテンツを格納するステップと、

特徴量格納手段が、前記コンテンツ毎の前記多次元特徴量を格納するステップと、

利用履歴格納手段が、前記コンテンツのうちの利用されたコンテンツの利用履歴を格納
するステップと、

条件入力手段が、前記コンテンツ格納手段に格納された複数のコンテンツの中から検索
基準となる任意の1つのコンテンツを検索基準のコンテンツとして指定させるステップと
、

計算手段が、前記特徴量格納手段に格納されているコンテンツの多次元特徴量の各次元
毎の平均値である第1の平均値を計算するとともに、前記特徴量格納手段に格納されてい
るコンテンツの多次元特徴量のうち前記利用履歴格納手段に格納されている前記利用され

10

20

30

40

50

たコンテンツに対応する多次元特徴量の各次元毎の平均値である第2の平均値を計算し、
これら計算した前記第1の平均値と前記第2の平均値との差分値が大きい次元ほど値が大きくなるような重み付けパラメータを各次元毎に計算し、
前記特徴量格納手段に格納された多次元特徴量の各次元の特徴量に対して、前記重み付け
パラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検
索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及
び類似性のうちの少なくとも一方からなる関連度を計算するステップと、
選択手段が、前記計算手段によって計算された前記検索基準のコンテンツと前記他のコ
ンテンツとの関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、
前記コンテンツ格納手段に格納された前記他のコンテンツから選択するステップと、
を有することを特徴とする情報検索方法。

10

【請求項7】

コンテンツが有する複数の特徴の度合いを数値化した特徴量を各次元要素とする多次元
特徴量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索す
る情報検索装置における情報検索方法であって、
前記情報検索装置は、
コンテンツ格納手段が、前記複数のコンテンツを格納するステップと、
特徴量格納手段が、前記コンテンツ毎の前記多次元特徴量を格納するステップと、
評価値格納手段が、外部から供給される前記コンテンツに対する利用者の肯定的な評価
、及び否定的な評価に対応する、前記肯定的な評価を示す数値、及び前記否定的な評価を
示す数値を少なくとも格納するステップと、
条件入力手段が、前記コンテンツ格納手段に格納された複数のコンテンツの中から検索
基準となる任意の1つのコンテンツを検索基準のコンテンツとして指定させるステップと

20

計算手段が、前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前
記評価値格納手段に格納されている前記肯定的な評価を示す数値のコンテンツに対応する
多次元特徴量の各次元毎の平均値である第1の平均値を計算するとともに、前記特徴量格
納手段に格納されているコンテンツの多次元特徴量のうち前記評価値格納手段に格納され
ている前記否定的な評価を示す数値のコンテンツに対応する多次元特徴量の各次元毎の平
均値である第2の平均値を計算し、これら計算した前記第1の平均値と前記第2の平均値
との差分値が大きい次元ほど値が大きくなるような重み付けパラメータを各次元毎に計算
し、
前記特徴量格納手段に格納された多次元特徴量の各次元の特徴量に対して、前記重み付け
パラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検
索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及
び類似性のうちの少なくとも一方からなる関連度を計算するステップと、
選択手段が、前記計算手段によって計算された前記検索基準のコンテンツと前記他のコ
ンテンツとの関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、
前記コンテンツ格納手段に格納された前記他のコンテンツから選択するステップと、
を有することを特徴とする情報検索方法。

30

40

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、複数のマルチメディア・コンテンツの中から所定の条件に基づいて所望のマルチメディア・コンテンツを検索するための検索装置、検索プログラム、及び検索方法に関し、特に利用者毎に最適な検索結果を提供する技術に関するものである。

【0002】

【従来の技術】

近年、コンピュータ技術やインターネット技術の発展により、非常に大量のコンテンツが作成され、保存されるようになってきている。これに伴い必要なコンテンツだけを取り出す検

50

索技術の重要性が高まっている。

【0003】

テキスト、静止画、音楽、映像といったマルチメディア・コンテンツを検索するために、コンテンツを多次元のベクトルで表現し、多次元空間内でのベクトル間の距離や角度を計算する方法が一般的に使われている。例えば、利用者があるコンテンツを指定し、それと類似するコンテンツを検索する場合、2つのコンテンツの違いを(1)式のように2つのベクトルの距離で表わし、距離の近いコンテンツを検索結果とすることが行われている。

【0004】

【数1】

$$D = \sqrt{\sum_i (X_{1i} - X_{2i})^2} \quad (1) \text{ 式}$$

10

【0005】

しかしながら、このような検索に使用される計算式は、予め決められている場合が多く、検索利用者には変更できない場合が大半である。このため、検索利用者の意図や好みを検索結果にうまく反映されないことが多かった。

【0006】

一方で、コンテンツ間の類似性や距離を表わす計算式を検索利用者が変更可能なシステムも提案されている。例えば、特開平11-175535号には、類似する画像を検索する装置に関する技術が開示されており、この技術内容によれば、利用者対話的に類似性や距離の計算式で使われる重み係数を設定し、より利用者に適切な検索結果が得られるようになっている。

20

【0007】

【発明が解決しようとする課題】

しかしながら、従来の技術においては、利用者が計算式の重み係数を全て手動で設定する必要があり、利用者にとって負担であった。この重み係数を設定するためには、検索対象となるコンテンツを構成要素(色特徴量、微分特徴量など)に分解して考える必要があり、かなりの技術的な知識と熟練が必要であった。また、コンテンツを2次元グラフ上に配置して、コンテンツの特徴を直感的に把握できるようにしているが、座標軸の意味が直感的に分かりにくく、この表示方法は重み係数の設定に必ずしも適していなかった。

30

【0008】

本発明は、検索利用者の負担が少なく、なおかつ利用者の意図や好みに合った、より適切な検索を提供することを目的とする。

【0009】

【問題を解決するための手段】

そこで、本発明は、上記課題を解決するために以下の装置、プログラム、及び方法を提供するものである。

40

(1) コンテンツが有する複数の特徴の度合いを数値化した特徴量を各次元要素とする多次元特徴量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索する情報検索装置であって、

前記複数のコンテンツを格納するコンテンツ格納手段と、

前記コンテンツ毎の前記多次元特徴量を格納する特徴量格納手段と、

前記コンテンツのうちの利用されたコンテンツの利用履歴を格納する利用履歴格納手段と、

前記コンテンツ格納手段に格納された複数のコンテンツの中から検索基準となる任意の1つのコンテンツを検索基準のコンテンツとして指定させるための条件入力手段と、

前記特徴量格納手段に格納されているコンテンツの多次元特徴量の各次元毎の平均値で

50

ある第1の平均値を計算するとともに、前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前記利用履歴格納手段に格納されている前記利用されたコンテンツに対応する多次元特徴量の各次元毎の平均値である第2の平均値を計算し、これら計算した前記第1の平均値と前記第2の平均値との差分値が大きい次元ほど値が大きくなるような重み付けパラメータを各次元毎に計算し、

前記特徴量格納手段に格納された多次元特徴量の各次元の特徴量に対して、前記重み付けパラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及び類似性のうちの少なくとも一方からなる関連度を計算する計算手段と、

前記計算手段によって計算された前記検索基準のコンテンツと前記他のコンテンツとの関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、前記コンテンツ格納手段に格納された前記他のコンテンツから選択する選択手段と、

を有する情報検索装置。

(2) コンテンツが有する複数の特徴の度合いを数値化した特徴量を各次元要素とする多次元特徴量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索する情報検索装置であって、

前記複数のコンテンツを格納するコンテンツ格納手段と、

前記コンテンツ毎の前記多次元特徴量を格納する特徴量格納手段と、

外部から供給される前記コンテンツに対する利用者の肯定的な評価、及び否定的な評価に対応する、前記肯定的な評価を示す数値、及び前記否定的な評価を示す数値を少なくとも

も格納する評価値格納手段と、

前記コンテンツ格納手段に格納された複数のコンテンツの中から検索基準となる任意の1つのコンテンツを検索基準のコンテンツとして指定させるための条件入力手段と、

前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前記評価値格納手段に格納されている前記肯定的な評価を示す数値のコンテンツに対応する多次元特徴量の各次元毎の平均値である第1の平均値を計算するとともに、前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前記評価値格納手段に格納されている前記否定的な評価を示す数値のコンテンツに対応する多次元特徴量の各次元毎の平均値である第2の平均値を計算し、これら計算した前記第1の平均値と前記第2の平均値との差分値が大きい次元ほど値が大きくなるような重み付けパラメータを各次元毎に計算し、

前記特徴量格納手段に格納された多次元特徴量の各次元の特徴量に対して、前記重み付けパラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及び類似性のうちの少なくとも一方からなる関連度を計算する計算手段と、

前記計算手段によって計算された前記検索基準のコンテンツと前記他のコンテンツとの関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、前記コンテンツ格納手段に格納された前記他のコンテンツから選択する選択手段と、

を有する情報検索装置。

(3) 前記利用履歴格納手段は、前記利用者がコンテンツを利用した利用日時を格納し、

前記計算手段は、前記利用日時を用いて前記各重み付けパラメータを計算することを特徴とする(1)に記載の情報検索装置。

(4) コンテンツが有する複数の特徴の度合いを数値化した特徴量を各次元要素とする多次元特徴量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索する情報検索処理をコンピュータに実行させるプログラムであって、前記コンピュータを、

前記複数のコンテンツをコンテンツ格納手段に格納させる手段と、

前記コンテンツ毎の前記多次元特徴量を特徴量格納手段に格納させる手段と、

前記コンテンツのうちの利用されたコンテンツの利用履歴を利用履歴格納手段に格納させる手段と、

前記コンテンツ格納手段に格納された複数のコンテンツの中から検索基準となる任意の

10

20

30

40

50

1つのコンテンツを検索基準のコンテンツとして指定させるための条件入力手段と、

前記特微量格納手段に格納されているコンテンツの多次元特微量の各次元毎の平均値である第1の平均値を計算するとともに、前記特微量格納手段に格納されているコンテンツの多次元特微量のうち前記利用履歴格納手段に格納されている前記利用されたコンテンツに対応する多次元特微量の各次元毎の平均値である第2の平均値を計算し、これら計算した前記第1の平均値と前記第2の平均値との差分値が大きい次元ほど値が大きくなるような重み付けパラメータを各次元毎に計算し、

前記特微量格納手段に格納された多次元特微量の各次元の特微量に対して、前記重み付けパラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及び類似性のうちの少なくとも一方からなる関連度を計算する計算手段と、

前記計算手段によって計算された前記検索基準のコンテンツと前記他のコンテンツとの関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、前記コンテンツ格納手段に格納された前記他のコンテンツから選択する選択手段、

として機能させるための情報検索プログラム。

(5) コンテンツが有する複数の特徴の度合いを数値化した特微量を各次元要素とする多次元特微量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索する情報検索処理をコンピュータに実行させるプログラムであって、前記コンピュータを、

前記複数のコンテンツをコンテンツ格納手段に格納させる手段と、

前記コンテンツ毎の前記多次元特微量を特微量格納手段に格納させる手段と、

外部から供給される前記コンテンツに対する利用者の肯定的な評価、及び否定的な評価に対応する、前記肯定的な評価を示す数値、及び前記否定的な評価を示す数値を少なくとも評価値格納手段に格納させる手段と、

前記コンテンツ格納手段に格納された複数のコンテンツの中から検索基準となる任意の1つのコンテンツを検索基準のコンテンツとして指定させるための条件入力手段と、

前記特微量格納手段に格納されているコンテンツの多次元特微量のうち前記評価値格納手段に格納されている前記肯定的な評価を示す数値のコンテンツに対応する多次元特微量の各次元毎の平均値である第1の平均値を計算するとともに、前記特微量格納手段に格納されているコンテンツの多次元特微量のうち前記評価値格納手段に格納されている前記否定的な評価を示す数値のコンテンツに対応する多次元特微量の各次元毎の平均値である第2の平均値を計算し、これら計算した前記第1の平均値と前記第2の平均値との差分値が大きい次元ほど値が大きくなるような重み付けパラメータを各次元毎に計算し、

前記特微量格納手段に格納された多次元特微量の各次元の特微量に対して、前記重み付けパラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及び類似性のうちの少なくとも一方からなる関連度を計算する計算手段と、

前記計算手段によって計算された前記検索基準のコンテンツと前記他のコンテンツとの関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、前記コンテンツ格納手段に格納された前記他のコンテンツから選択する選択手段、

として機能させるための情報検索プログラム。

(6) コンテンツが有する複数の特徴の度合いを数値化した特微量を各次元要素とする多次元特微量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索する情報検索装置における情報検索方法であって、

前記情報検索装置は、

コンテンツ格納手段が、前記複数のコンテンツを格納するステップと、

特微量格納手段が、前記コンテンツ毎の前記多次元特微量を格納するステップと、

利用履歴格納手段が、前記コンテンツのうちの利用されたコンテンツの利用履歴を格納するステップと、

条件入力手段が、前記コンテンツ格納手段に格納された複数のコンテンツの中から検索

10

20

30

40

50

基準となる任意の1つのコンテンツを検索基準のコンテンツとして指定させるステップと

、
 計算手段が、前記特徴量格納手段に格納されているコンテンツの多次元特徴量の各次元毎の平均値である第1の平均値を計算するとともに、前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前記利用履歴格納手段に格納されている前記利用されたコンテンツに対応する多次元特徴量の各次元毎の平均値である第2の平均値を計算し、これら計算した前記第1の平均値と前記第2の平均値との差分値が大きい次元ほど値が大きくなるような重み付けパラメータを各次元毎に計算し、

前記特徴量格納手段に格納された多次元特徴量の各次元の特徴量に対して、前記重み付けパラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及び類似性のうちの少なくとも一方からなる関連度を計算するステップと、

選択手段が、前記計算手段によって計算された前記検索基準のコンテンツと前記他のコンテンツとの関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、前記コンテンツ格納手段に格納された前記他のコンテンツから選択するステップと、
 を有することを特徴とする情報検索方法。

(7) コンテンツが有する複数の特徴の度合いを数値化した特徴量を各次元要素とする多次元特徴量がそれぞれ与えられている複数のコンテンツのうちから所望のコンテンツを検索する情報検索装置における情報検索方法であって、

前記情報検索装置は、
 コンテンツ格納手段が、前記複数のコンテンツを格納するステップと、
 特徴量格納手段が、前記コンテンツ毎の前記多次元特徴量を格納するステップと、
 評価値格納手段が、外部から供給される前記コンテンツに対する利用者の肯定的な評価、及び否定的な評価に対応する、前記肯定的な評価を示す数値、及び前記否定的な評価を示す数値を少なくとも格納するステップと、

条件入力手段が、前記コンテンツ格納手段に格納された複数のコンテンツの中から検索基準となる任意の1つのコンテンツを検索基準のコンテンツとして指定させるステップと

、
 計算手段が、前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前記評価値格納手段に格納されている前記肯定的な評価を示す数値のコンテンツに対応する多次元特徴量の各次元毎の平均値である第1の平均値を計算するとともに、前記特徴量格納手段に格納されているコンテンツの多次元特徴量のうち前記評価値格納手段に格納されている前記否定的な評価を示す数値のコンテンツに対応する多次元特徴量の各次元毎の平均値である第2の平均値を計算し、これら計算した前記第1の平均値と前記第2の平均値との差分値が大きい次元ほど値が大きくなるような重み付けパラメータを各次元毎に計算し、

前記特徴量格納手段に格納された多次元特徴量の各次元の特徴量に対して、前記重み付けパラメータを用いて各次元毎に重み付けを行って、前記条件入力手段で指定された前記検索基準のコンテンツと、前記コンテンツ格納手段に格納された他のコンテンツとの距離及び類似性のうちの少なくとも一方からなる関連度を計算するステップと、

選択手段が、前記計算手段によって計算された前記検索基準のコンテンツと前記他のコンテンツとの関連度をもとに、前記検索基準のコンテンツと関連性の高いコンテンツを、前記コンテンツ格納手段に格納された前記他のコンテンツから選択するステップと、
 を有することを特徴とする情報検索方法。

【0013】

【発明の実施の形態】

・第1の実施例

本発明に係る情報検索装置の第1の実施例の構成を図1に示す。コンテンツを一意に識別するコンテンツID、コンテンツの名称や制作者等の属性情報、及びコンテンツ本体を格納するコンテンツ格納手段1と、コンテンツに対応した多次元の特徴量を格納する特徴量

格納手段 4 と、多次元特徴量を使ってコンテンツ間の距離あるいは類似性を計算する距離計算手段 2 と、距離の小さなコンテンツあるいは類似性の大きなコンテンツを選択する選択手段 5 と、利用者の利用履歴を格納する利用履歴格納手段 7 と、検索条件を指定する条件入力手段 3 と、ディスプレイ等の表示手段 6 によって構成されている。なお、このような情報検索装置は、通常のコンピュータを使ってソフトウェア的に構成することが可能である。

【 0 0 1 4 】

次に、本発明に係る情報検索装置の第 1 の実施例の処理フローを図 2 に示す。同図によれば、検索利用者は、検索を行うにあたり、パスワードの入力等により個人認証を行う（ステップ S 1）。なお、検索利用者にはあらかじめ利用者を一意に識別するユーザ ID が付与されている。

10

【 0 0 1 5 】

次に、検索利用者は、条件入力手段 3 を使って、コンテンツ格納手段 1 に格納されている任意のコンテンツを 1 つ指定する（以下、例示コンテンツとする）（ステップ S 2）。例示コンテンツを選ぶには、例えば、コンテンツ名称、制作者などの属性情報を指定し、これらの文字列照合によりコンテンツを絞り込めばよい。以下の処理により例示コンテンツと類似したコンテンツとを検索する。

【 0 0 1 6 】

コンテンツ格納手段 1 では、図 3 に示すような形式で、コンテンツを一意に識別するコンテンツ ID、コンテンツの名称や制作者やジャンルなどの属性情報、コンテンツ本体などのデータが格納されている。以下では格納されているコンテンツの総数を N とする。また、特徴量格納手段 4 においては、図 4 に示すような形式で、N 個のコンテンツの M 次元の特徴量が格納されている。

20

【 0 0 1 7 】

次に、距離計算手段 2 において、距離計算パラメータを計算する（ステップ S 3）が、これについては後述する。

【 0 0 1 8 】

次に、距離計算手段 2 において、例示コンテンツと他の全てのコンテンツとの重み付き距離 $D[i]$ (i) を (2) 式に従って計算する（ステップ S 4）。

【 0 0 1 9 】

【 数 2 】

30

$$D[i] = \sqrt{\sum_{j=1}^M W[j](P[i][j] - P_s[j])^2} \quad (2) \text{ 式}$$

【 0 0 2 0 】

ここで、 $P_s[j]$ ($j = 1 \sim M$) は例示コンテンツの特徴量であり、 \mathcal{C} は、例示コンテンツ以外の全てのコンテンツの集合であり、 $P[i][j]$ は \mathcal{C} に属するコンテンツの特徴量である。また、 $W[j]$ は重み係数である。選択手段 5 においては、この距離 $D[i]$ が小さい順に一定数のコンテンツを選択し、検索結果とする（ステップ S 5）。検索利用者が検索結果の中から利用するコンテンツを選択して利用すると、利用履歴格納手段 7 に図 5 に示すような形式で利用履歴が格納される（ステップ S 6）。

40

【 0 0 2 1 】

なお、本実施例における距離計算パラメータは (2) 式の重み係数 $W[j]$ であり、以下の方法で算出する。利用履歴格納手段 7 を調べ、検索利用者が過去に利用したコンテンツの集合 \mathcal{C}_1 を求め、 \mathcal{C}_1 の要素数を N_1 とする。次に、集合 \mathcal{C}_1 を対象にして、M 次元特徴量の平均ベクトル $P_{a1}[j]$ ($j = 1 \sim M$) を (3) 式に従って計算する。

【 0 0 2 2 】

50

【数 3】

$$Pa[j] = \frac{\sum_{i \in \phi} P[i][j]}{N1} \quad (3) \text{ 式}$$

【0023】

あるいは(3)式の代わりに(4)式を使って平均ベクトルPaを計算しても良い。ここで、Tcは現在日時、T[i]は利用履歴格納手段7に格納されているコンテンツを利用した日時、G(x)は図6に示すような単調減少関数である。(4)式を使うことにより、利用者の最新の利用傾向をより検索結果に反映させることができる。 10

【0024】

【数 4】

$$Pa[j] = \frac{\sum_{i \in \phi} G(Tc - T[i])P[i][j]}{\sum_{i \in \phi} G(Tc - T[i])} \quad (4) \text{ 式}$$

【0025】

次に(5)式に従って、標準化特徴量Pb[j](j=1~M)を求める。 20

【0026】

【数 5】

$$Pb[j] = \frac{Pa[j] - Pm[j]}{S[j]} \quad (5) \text{ 式}$$

【0027】

ここで、Pm[j]は、全てのコンテンツの平均値であり、S[j]は、全てのコンテンツの標準偏差である。標準化特徴量Pb[j]は、「0」に近いほど平均的であり、その絶対値が大きい程平均からずれていることを示す。また利用者が過去に利用したコンテンツは、その利用者の好きなコンテンツである確率が高い。すなわち、標準化特徴量が「0」から離れている特徴量ほど、検索利用者の好みが明確に表われていると考えられる。そこで、標準化特徴量が「0」に近い次元の重みW[j]を小さく、標準化特徴量の絶対値が大きな次元の重み係数W[j]を大きくするような処理を行う。例えば、(6)式に従って重み係数を決める。ここで、F(x)は図7に示すような特性の単調増加関数である。なお、重み係数W[j]を検索の都度計算するのではなく、適当なタイミングで計算した結果を格納し利用するようにしても良い。 30

【0028】

【数 6】

$$W[j] = F(|Pb[j]|) \quad (6) \text{ 式}$$

【0029】

以上の処理により、検索利用者の好みをよりの確に反映している特徴量の重みを大きくして距離計算を行うことができるので、検索結果に検索者の好みを反映することができる。

【0030】

・第2の実施例

次に本発明に係る情報検索装置の第2の実施例の構成を図8に示す。なお、第1の実施例の構成と同じ構成要素については同一の符号を付す。コンテンツを一意に識別するコンテンツID及びコンテンツの名称や制作者等の属性情報及びコンテンツ本体を格納するコンテンツ格納手段1と、コンテンツに対応した多次元の特徴量を格納する特徴量格納手段4と、多次元特徴量を使ってコンテンツ間の距離あるいは類似性を計算する距離計算手段2と、距離の小さなコンテンツあるいは類似性の大きなコンテンツを選択する選択手段5と、利用者のコンテンツに対する評価値を格納する評価値格納手段8と、検索条件を指定する条件入力手段3と、ディスプレイ等の表示手段6で構成されている。なお、このような情報検索装置は、通常のコンピュータを使ってソフトウェア的に構成することが可能である。

10

【0031】

次に第2の実施例の処理フローを図9に示す。まず、検索利用者は、検索を行うにあたり、パスワード入力により個人認証を行う(ステップS11)。なお、検索利用者にはあらかじめ利用者を一意に識別するユーザIDが付与されている。

【0032】

検索に先立ち検索利用者は、少なくとも一部のコンテンツについて評価しておく(ステップS12)。評価方法として、いくつかの方法が考えられるが、例えば、「好き」、「どちらでもない」、「きらい」の3段階で評価する。評価値格納手段8には、図10に示す形式で、ユーザID、コンテンツID、評価値が格納されている。評価値は例えば、「好き」に対して「+1」、「どちらでもない」に対して「0」、「きらい」に対して「-1」というように数値化されている。次に検索利用者は、条件入力手段3を使って、コンテンツ格納手段1に格納されている任意のコンテンツを1つ指定する(例示コンテンツ)(ステップS13)。

20

【0033】

また、コンテンツ格納手段1では、図3に示すような形式で、コンテンツを一意に識別するコンテンツID、コンテンツの名称や制作者やジャンルなどの属性情報、コンテンツ本体などのデータが格納されている。以下では格納されているコンテンツの総数をNとする。

【0034】

特徴量格納手段4においては、図4に示すような形式で、N個のコンテンツのM次元の特徴量が格納されている。

30

【0035】

次に、距離計算手段2において、距離計算パラメータを計算する(ステップS14)が、これについては後述する。

【0036】

次に、距離計算手段2において、例示コンテンツと他の全てのコンテンツとの重み付き距離 $D[i]$ (i)を上述した(2)式に従って計算する(ステップS15)。ここで、 $P_s[j]$ ($j = 1 \sim M$)は例示コンテンツの特徴量であり、 P は、例示コンテンツ以外の全てのコンテンツの集合である。また、 $P[i][j]$ は i に属するコンテンツの特徴量であり、 $W[j]$ は重み係数である。選択手段5では、この距離が小さい順に一定数のコンテンツを選択し、検索結果とする(ステップS16)。

40

【0037】

本実施例における距離計算パラメータは(2)式の重み係数 $W[j]$ であり、以下の方法で算出する。検索利用者が、「好き」と評価したコンテンツの集合を A 、「嫌い」と評価したコンテンツの集合を B とする。各次元ごとに、(7)式に従って、集合 A 内の分散 $V_A[j]$ 、集合 B 内の分散 $V_B[j]$ ($j = 1 \sim M$)を計算する。ここで、 N_A は集合 A に属するコンテンツ数、 N_B は集合 B に属するコンテンツ数、 $P_A[j]$ は集合 A のコンテンツの平均値、 $P_B[j]$ は集合 B のコンテンツの平均値を表わす。

【0038】

【数7】

50

$$V\alpha[j] = \frac{\sum_{i \in \alpha} (P[i][j] - P\alpha[j])^2}{N\alpha}$$

$$V\beta[j] = \frac{\sum_{i \in \beta} (P[i][j] - P\beta[j])^2}{N\beta}$$

(7) 式

【 0 0 3 9 】

更に、(8)式に従って指標 $Z[j]$ ($j = 1 \sim M$) を計算する。指標 $Z[j]$ は、集合 α 、
集合 β 内の分散が小さく、集合 α と集合 β の平均値の差が大きいく程大きな値となる。すな
わち、指標 $Z[j]$ が大きな特徴量は、検索利用者のコンテンツに対する好き嫌いをより反
映したものと考えることができる。

10

【 0 0 4 0 】

【 数 8 】

$$Z[j] = \frac{(P\alpha[j] - P\beta[j])^2}{\frac{N\alpha \times V\alpha[j]}{N\alpha + N\beta} + \frac{N\beta \times V\beta[j]}{N\alpha + N\beta}}$$

(8) 式

20

【 0 0 4 1 】

従って、例えば(9)式に従って、指標 $Z[j]$ が大きくなる程、距離計算手段 2 の重み係
数 $W[j]$ を大きくなるように設定する。ここで $F(x)$ は図 7 に示すような特性の単調増
加関数である。

【 0 0 4 2 】

【 数 9 】

$$W[j] = F(Z[j])$$

(9) 式

30

【 0 0 4 3 】

なお、重み係数 $W[j]$ を検索の都度計算するのではなく、適当なタイミング計算した結果
を格納し利用するようにしても良い。

【 0 0 4 4 】

以上の処理により、検索利用者の好みをよりの確に反映している特徴量の重みを大きくし
て距離計算を行うことができるので、検索結果に検索者の好みを反映することができる。

【 0 0 4 5 】

・ 第 3 の実施例

40

次に、本発明に係る情報検索装置の第 3 の実施例の構成を図 11 に示す。なお、上述した
実施例と同じ構成要素については同一の符号を付している。コンテンツを一意に識別する
コンテンツ ID 及びコンテンツの名称や制作者等の属性情報及びコンテンツ本体を格納す
るコンテンツ格納手段 1 と、コンテンツに対応した多次元の特徴量を格納する特徴量格納
手段 4 と、多次元特徴量を使ってコンテンツ間の距離あるいは類似性を計算する距離計算
手段 2 と、距離の小さなコンテンツあるいは類似性の大きなコンテンツを選択する選択手
段 5 と、特徴量の大きいコンテンツ及び小さいコンテンツの属性情報を表示し、距離計算
のパラメータを指定する距離パラメータ指定手段 9 と、設定した距離パラメータを格納す
る距離パラメータ格納手段 10 と、検索条件を指定する条件入力手段 3 と、ディスプレイ
等の表示手段 3 で構成されている。なお、このような情報検索装置は、通常のコンピュー

50

タを使ってソフトウェア的に構成することが可能である。

【0046】

第3の実施例の処理フローを図12に示す。まず検索利用者は、検索を行うにあたり、パスワード入力により個人認証を行う(ステップS21)。検索利用者にはあらかじめ利用者を一意に識別するユーザIDが付与されている。

【0047】

検索利用者は、条件入力手段3を使って、コンテンツ格納手段1に格納されている任意のコンテンツを1つ指定する(例示コンテンツ)(ステップS22)。

【0048】

コンテンツ格納手段1では、図3に示すような形式で、コンテンツを一意に識別するコンテンツID、コンテンツの名称や制作者やジャンルなどの属性情報、コンテンツ本体などのデータが格納されている。以下では格納されているコンテンツの総数をNとする。

【0049】

特徴量格納手段4においては、図4に示すような形式で、N個のコンテンツのM次元の特徴量が格納されている。

【0050】

距離パラメータ指定手段9では、図13に示すように各特徴量ごとに、特徴量の大きなコンテンツの属性情報を数個、特徴量の小さなコンテンツの属性情報を数個、表示手段6に表示する。本実施例における距離計算パラメータは重み係数 $W[j]$ であり、検索利用者は、図13の下方にあるスライダー部を使用して距離計算に使う重み係数 $W[j]$ ($j = 1 \sim M$)を設定する(ステップS23)。スライダー部のスライダーの初期値は前回その利用者が設定した値が再現されるようになっている。検索利用者は特徴量ごとに、特徴量の大きなコンテンツと小さなコンテンツを見比べて、両者の違いが大きいと判断した場合は、その特徴量の重み係数を大きく設定する。逆に違いが小さいと感じた場合には、その特徴量の重み係数を小さく設定する。このように設定された重み係数 $W[j]$ は、図14に示す形式で検索利用者毎に距離パラメータ格納手段10に格納される(ステップS24)。そして、格納された値は次回その検索利用者が検索する場合の初期値として利用される。

【0051】

距離計算手段2においては、例示コンテンツと他の全てのコンテンツとの重み付き距離 $D[i]$ (i)を上述した(2)式に従って計算する(ステップS25)。

【0052】

ここで、 $P_s[j]$ ($j = 1 \sim M$)は例示コンテンツの特徴量であり、 P は、例示コンテンツ以外の全てのコンテンツの集合である。また、 $P[i][j]$ は P に属するコンテンツの特徴量であり、 $W[j]$ は利用者が設定した重み係数である。選択手段においては、この距離 $D[i]$ が小さい順に一定数のコンテンツを選択し、検索結果とする(ステップS26)。

【0053】

なお、上記各実施例においては、多次元特徴量の距離を用いて、利用者が所望するコンテンツの選択を行ったが、距離の代わりに類似性を用いて選択を行っても良い。

【0054】

【発明の効果】

以上、詳述したように、本発明に係る情報検索装置、情報検索プログラム、及び情報検索方法によれば、利用者にほとんど負担をかけずに、利用者の好みを反映したマルチメディア・コンテンツの類似検索が可能になるという効果を奏する。

【0055】

また、請求項1、請求項4、及び請求項6に記載の発明によれば、利用者が全く入力作業をすることなく、距離計算のパラメータを利用者ごとに自動的に設定することができ、利用者の好みを反映した検索結果を得ることができるという効果を奏する。

【0056】

更に、請求項2、請求項5、及び請求項7に記載の発明によれば、利用者の好みを更に

10

20

30

40

50

精度良く反映して検索することが可能になるという効果を奏する。

【0059】

また、請求項3に記載の発明によれば、検索利用者が検索を行った日付を使って計算パラメータを決定することにより、利用者の最新の好みを反映した検索が可能になるという効果を奏するものである。

【図面の簡単な説明】

【図1】本発明に係る情報検索装置の第1の実施例の構成を示す図である。

【図2】本発明に係る情報検索装置の第1の実施例の処理フローを示す図である。

【図3】本発明に係る情報検索装置のコンテンツ格納手段のデータ形式を示す図である。

【図4】本発明に係る情報検索装置の特徴量格納手段のデータ形式を示す図である。

10

【図5】本発明に係る情報検索装置の利用履歴格納手段のデータ形式を示す図である。

【図6】本発明に係る情報検索装置の距離計算手段で使用する関数Gの特性を示す図である。

【図7】本発明に係る情報検索装置の距離計算手段で使用する関数Fの特性を示す図である。

【図8】本発明に係る情報検索装置の第2の実施例の構成を示す図である。

【図9】本発明に係る情報検索装置の第2の実施例の処理フローを示す図である。

【図10】本発明に係る情報検索装置の評価値格納手段のデータ形式を示す図である。

【図11】本発明に係る情報検索装置の第3の実施例の構成を示す図である。

【図12】本発明に係る情報検索装置の第3の実施例の処理フローを示す図である。

20

【図13】本発明に係る情報検索装置の第3の実施例における操作画面の一例を示す図である。

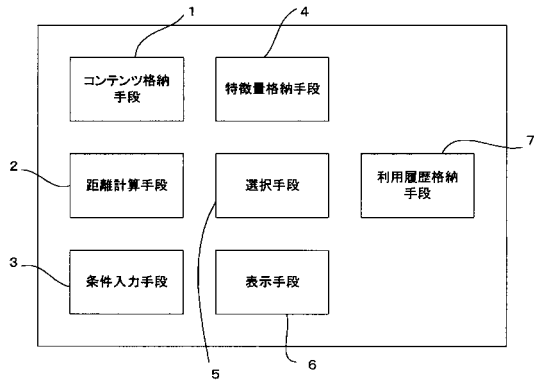
【図14】本発明に係る情報検索装置の距離パラメータ格納手段のデータ形式を示す図である。

【符号の説明】

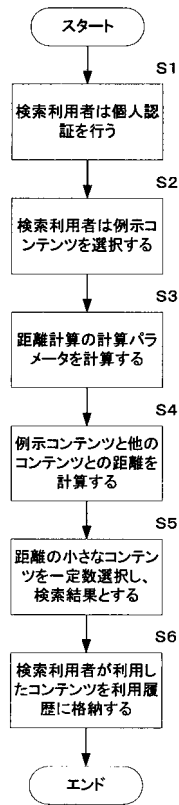
- 1 コンテンツ格納手段
- 2 距離計算手段
- 3 条件入力手段
- 4 特徴量格納手段
- 5 選択手段
- 6 表示手段
- 7 利用履歴格納手段
- 8 評価値格納手段
- 9 距離パラメータ指定手段
- 10 距離パラメータ格納手段

30

【図1】



【図2】



【図3】

コンテンツID	名称	制作者	ジャンル	コンテンツ本体
ID-1					
ID-2					
.....					
ID-N					

図3

【図6】

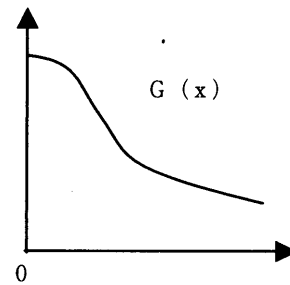


図6

【図4】

コンテンツID	特徴量1	特徴量2	特徴量M
ID-1				
ID-2				
.....				
ID-N				

図4

【図5】

ユーザID	コンテンツID	利用日時
XXXX	OOO	2001 01/01 13:55

図5

【図7】

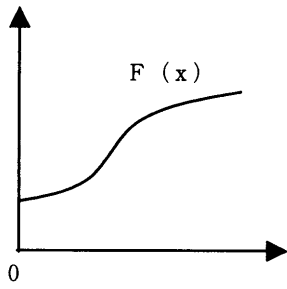
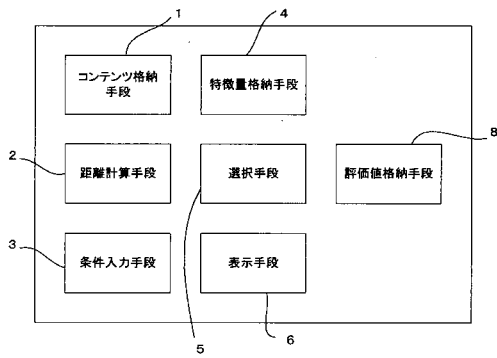


図7

【図8】

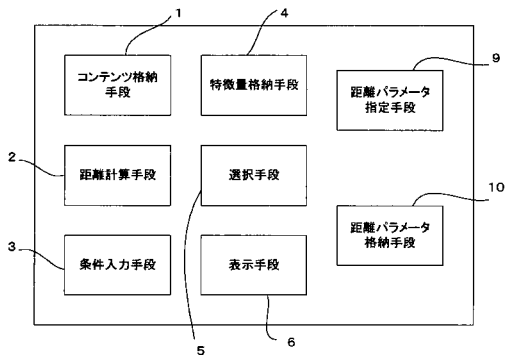


【図10】

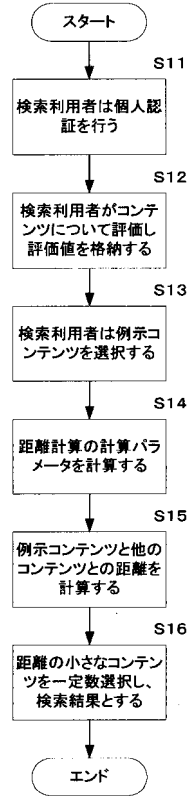
ユーザID	コンテンツID	評価値
ユーザ1	コンテンツID-1	+1
ユーザ1	コンテンツID-2	-1
ユーザ2	コンテンツID-1	0
.....

図10

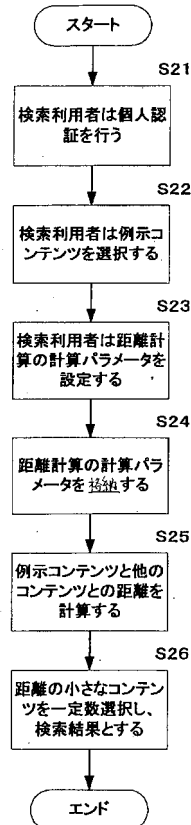
【図11】



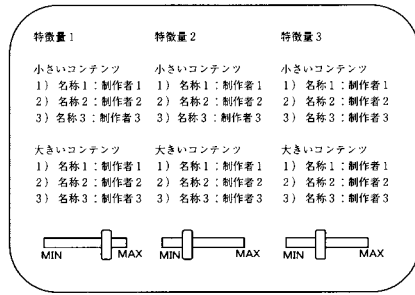
【図9】



【図12】



【図 13】



【図 14】

ユーザID	W[1]	W[2]	W[M]
ユーザ 1				
ユーザ 2				
.....				
ユーザ K				

図 14

フロントページの続き

合議体

審判長 田口 英雄

審判官 和田 財太

審判官 飯田 清司

(56)参考文献 特開2000-35974(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F17/30

G06T 7/00