

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3934325号
(P3934325)

(45) 発行日 平成19年6月20日(2007.6.20)

(24) 登録日 平成19年3月30日(2007.3.30)

(51) Int. Cl.

G06F 17/30 (2006.01)

F I

G06F 17/30 340B

請求項の数 9 (全 18 頁)

(21) 出願番号	特願2000-331817 (P2000-331817)	(73) 特許権者	000005108
(22) 出願日	平成12年10月31日(2000.10.31)		株式会社日立製作所
(65) 公開番号	特開2002-140361 (P2002-140361A)		東京都千代田区丸の内一丁目6番6号
(43) 公開日	平成14年5月17日(2002.5.17)	(74) 代理人	110000350
審査請求日	平成15年3月18日(2003.3.18)		ポレール特許業務法人
		(74) 代理人	100068504
			弁理士 小川 勝男
		(74) 代理人	100086656
			弁理士 田中 恭助
		(74) 代理人	100094352
			弁理士 佐々木 孝
		(72) 発明者	稲場 靖彦
			神奈川県川崎市幸区鹿島田890番地 株式会社日立製作所 ビジネスソリューション開発本部内
			最終頁に続く

(54) 【発明の名称】 文書検索方法、文書検索装置及び文書検索プログラムの記憶媒体

(57) 【特許請求の範囲】

【請求項1】

指定された検索条件に基づいて登録された文書を検索する文書検索方法において、
文書検索装置が、

前記検索条件から重み付けされた検索語のリストを含む第1の検索プロファイルを記憶手段上に生成し、ユーザの所望する検索結果が得られるまで、検索された文書に対する評価を受け付け、該評価に基づいて前記第1の検索プロファイルの前記検索語の重みを修正して、文書のデータベースを検索し、

前記ユーザの所望する検索結果が得られたとき、評価された文書の識別子を、所望の検索結果が得られたときの前記第1の検索プロファイルおよび前記ユーザの識別子と関連づけて検索履歴ファイルに記憶し、

入力された検索条件に基づいて作成された第2の検索プロファイルに基づき、前記文書のデータベースから類似度により類似する文書を検索し、その文書識別子を取得し、

蓄積された検索プロファイルから類似度により前記第2の検索プロファイルに類似する検索プロファイルを検索し、その検索プロファイル識別子を取得し、

前記検索履歴ファイルを検索し、取得された前記文書識別子をもつ文書を所望と評価した他のユーザの識別子、又は前記検索プロファイル識別子をもつ他のユーザの識別子を取得して表示することを特徴とする文書検索方法。

【請求項2】

前記検索条件は、種文書であることを特徴とする請求項1記載の文書検索方法。

10

20

【請求項 3】

前記検索語の重みの修正は、所望すると評価された文書の該当する検索語の重みを増し、所望でないと評価された文書の該当する検索語の重みを減じるように前記検索語の重みを修正することを特徴とする請求項 1 記載の文書検索方法。

【請求項 4】

指定された検索条件に基づいて登録された文書を検索する文書検索装置において、前記検索条件から重み付けされた検索語のリストを含む第 1 の検索プロファイルを記憶手段上に生成し、ユーザの所望する検索結果が得られるまで、検索された文書に対する評価を受け付け、該評価に基づいて前記第 1 の検索プロファイルの前記検索語の重みを修正して、文書のデータベースを検索する手段と、

10

前記ユーザの所望する検索結果が得られたとき、評価された文書の識別子を、所望の検索結果が得られたときの前記第 1 の検索プロファイルおよび前記ユーザの識別子と関連づけて検索履歴ファイルに記憶する手段と、

入力された検索条件に基づいて作成された第 2 の検索プロファイルに基づき、前記文書のデータベースから類似度により類似する文書を検索し、その文書識別子を取得する手段と、

蓄積された検索プロファイルから類似度により前記第 2 の検索プロファイルに類似する検索プロファイルを検索し、その検索プロファイル識別子を取得する手段と、

前記検索履歴ファイルを検索し、取得された前記文書識別子をもつ文書を所望と評価した他のユーザの識別子、又は前記検索プロファイル識別子をもつ他のユーザの識別子を取

20

得して表示する手段とを有することを特徴とする文書検索装置。

【請求項 5】

前記検索条件は、種文書であることを特徴とする請求項 4 記載の文書検索装置。

【請求項 6】

前記データベースを検索する手段は、所望すると評価された文書の該当する検索語の重みを増し、所望でないと評価された文書の該当する検索語の重みを減じるように前記検索語の重みを修正することを特徴とする請求項 4 記載の文書検索装置。

【請求項 7】

コンピュータに、指定された検索条件に基づいて登録された文書を検索する手順を実行させるためのプログラムを記録したコンピュータ読み取り可能な記憶媒体において、前記

30

コンピュータに、前記検索条件から重み付けされた検索語のリストを含む第 1 の検索プロファイルを記憶手段上に生成し、ユーザの所望する検索結果が得られるまで、検索された文書に対する評価を受け付け、該評価に基づいて前記第 1 の検索プロファイルの前記検索語の重みを修正して、文書のデータベースを検索する手順、

前記ユーザの所望する検索結果が得られたとき、評価された文書の識別子を、所望の検索結果が得られたときの前記第 1 の検索プロファイルおよび前記ユーザの識別子と関連づけて検索履歴ファイルに記憶する手順、

入力された検索条件に基づいて作成された第 2 の検索プロファイルに基づき、前記文書のデータベースから類似度により類似する文書を検索し、その文書識別子を取得する手順

40

、蓄積された検索プロファイルから類似度により前記第 2 の検索プロファイルに類似する検索プロファイルを検索し、その検索プロファイル識別子を取得する手順、および

前記検索履歴ファイルを検索し、取得された前記文書識別子をもつ文書を所望と評価した他のユーザの識別子、又は前記検索プロファイル識別子をもつ他のユーザの識別子を取得して表示する手順を、実行させるためのプログラムを記録したコンピュータ読み取り可能な記憶媒体。

【請求項 8】

前記検索条件は、種文書であることを特徴とする請求項 7 記載のプログラムを記録したコンピュータ読み取り可能な記憶媒体。

50

【請求項 9】

前記データベースを検索する手順は、所望すると評価された文書の該当する検索語の重みを増し、所望でないと評価された文書の該当する検索語の重みを減じるように前記検索語の重みを修正することを特徴とする請求項 7 記載のプログラムを記録したコンピュータ読み取り可能な記憶媒体。

【発明の詳細な説明】**【0001】****【発明の属する技術分野】**

本発明は、ユーザが指定した検索条件をもとに文書を検索する方法および装置に係わり、特に各ユーザのうち誰がどの情報を知っているか、あるいは興味を持っているかという情報を検索して提示する方法及び装置に関する。

10

【0002】**【従来技術】**

近年、激化する企業間や組織間の競争において、作業効率を上げるために各人が持っている知識やノウハウを共有し、再利用することが必要となっている。

【0003】

このため企業や組織の中で、「この情報を知っているのは誰であるか」といったことを知る方法（以下「KnowWho検索」と呼ぶ）に対する要求が高まってきている。たとえば、「A社に対する営業戦略を知っている人物は誰であるか」といった検索を行なう必要がある。このようなKnowWho検索を実現する技術として、以下に示す二つの技術がある。

20

【0004】

まず第一の技術は、「特開平10-63685号公報」に示されるように、各ユーザが過去に文書検索システムを用いて検索を行なった際の検索条件式を保存しておき、他のユーザがその保存された検索条件式を検索することで、目的の情報を知っているユーザを検索する、という技術である（以下、従来技術1と呼ぶ）。

【0005】

この技術による処理の流れを図2のPAD(Problem Analysis Diagram)図を用いて説明する。この処理ではステップ301において、ステップ302～ステップ309の処理をシステム終了まで繰り返す。まずステップ302において、ユーザからの入力を受け付ける。次にステップ303において文書検索要求が入力されたと判定された場合には、以下のステップ304～ステップ306の処理を実行する。まず、ステップ304において、ユーザからの検索条件式を受け付ける。次に、ステップ305において、ステップ304で取得した検索条件式にもとづいて文書を検索する。次に、ステップ306において、ステップ304で取得した検索条件式を保存する。

30

【0006】

また、ステップ307において、検索条件式の検索が要求されたと判定された場合には、以下のステップ308～ステップ309の処理を実行する。まず、ステップ308において、ユーザからの検索条件式を受け付ける。次に、ステップ309において、ステップ306で保存した検索条件式のうち、ステップ308で取得した検索条件式に類似した検索条件式を利用したユーザ名を提示する。

40

【0007】

たとえば、「A社」「営業」「戦略」の3語を含む文書を検索する」という検索条件を入力すると、過去に類似した検索条件式を入力したユーザの名前を検索結果として提示することが可能となる。このことにより、「A社に対する営業戦略」を知っていると思われる人物を推測することが可能となる。

【0008】

また第二の技術は、「特開平10-83386号公報」に示されるように、各ユーザが自分の知っているノウハウを表すデータ（以下、個人プロフィールと呼ぶ）を予め登録しておき、他のユーザが個人プロフィールを検索することにより、目的とする情報を知っている人物が誰であるかを提示する、という技術である（以下、従来技術2と呼ぶ）。この技

50

術による処理の流れを図3のPAD図を用いて説明する。

【0009】

この処理ではステップ401において、ステップ402～ステップ408の処理をシステム終了まで繰り返す。まずステップ402において、ユーザからの入力を受け付ける。次にステップ403において、ノウハウの登録が要求されたと判定された場合には、以下のステップ404～ステップ405の処理を実行する。まずステップ404において、ユーザからそのユーザの持っているノウハウの入力を受け付ける。次にステップ405において、ステップ404で受け付けたノウハウを保存する。

【0010】

また、ステップ406において、ノウハウの検索が要求されたと判定された場合には、以下のステップ407～ステップ408の処理を実行する。まずステップ407において、ユーザからの検索条件式を受け付ける。次にステップ408において、ステップ407で取得した検索条件にもとづき、ステップ405で保存したノウハウを検索し、検索されたノウハウを登録したユーザ名を提示する。

10

【0011】

たとえば、「A社に対する営業戦略」を知っているユーザは、予めそのユーザの個人プロフィールに「A社」「営業」といった文字列を登録しておく。そして他のユーザは、「A社」「営業」などの検索条件でこの個人プロフィールを検索し、登録したユーザ名を取得することで、この個人プロフィールを登録した人物が「A社に対する営業戦略」を知っていると推測することが可能となる。

20

【0012】

【発明が解決しようとする課題】

しかしながら上述した従来技術1では、過去に利用された検索条件式に類似しているものがあれば、目的とする検索結果が得られたか否かにかかわらず、これを検索結果として提示してしまう。そのため、所望の文書をまったく検索できないような不適切な検索条件式を入力した人物も提示されてしまうという問題があった。

【0013】

たとえば「A社」「営業」「戦略」の3語を含む文書を検索する"という検索条件に似た検索条件式として、「B社」「営業」「戦略」の3語を含む文書を検索する"という検索条件式が、検索条件中の2語が一致しているため検索される場合がある。このように、「A社に対する営業戦略」ではなく「B社に対する営業戦略」を知っていると思われるユーザが検索されてしまう、といった問題があった。

30

【0014】

また従来技術2では、事前に登録者が個人プロフィールを用意するという手間がかかり、また用意されたノウハウしか利用できないという問題があった。また、各ユーザは必ずしも適切な個人プロフィールを用意できないという問題があった。

【0015】

たとえば、各ユーザは「A社に対する営業戦略」などのノウハウについて社内データベースを検索するなどの手段で知るたびに、「A社」「営業」といった文字列を、自分のプロフィールに登録しなければならないという手間がかかる。また、「A社」「営業」という文字列だけでは、「A社」に何の案件で営業するノウハウなのかなどが表現できないなど、必ずしも適切な個人プロフィールを登録することができないという問題があった。

40

【0016】

本発明の目的は、以上のような問題を解決し、「各ユーザが何を知っているか」という情報を各ユーザの負担なく適切に登録することを可能にし、また「この情報を知っているのは誰か」という情報を適切に検索することを可能にする手段を提供することにある。

【0017】

【課題を解決するための手段】

本発明は、指定された検索条件に基づいて登録された文書を検索する文書検索方法において、文書検索の結果として得られた少なくとも1つの文書についてユーザにより所望する

50

か否かの評価の入力を受け付けるステップと、所望すると評価された文書の該当する検索語の重みを増し、所望でないと評価された文書の該当する検索語の重みを減じるようにその検索条件を変更するステップと、変更された検索条件に基づいて上記の文書検索と評価の入力を繰り返すステップと、最終的にユーザにより所望すると評価された文書の識別子をユーザの識別子とともに検索履歴として記憶手段に格納するステップとを有する文書検索方法を特徴とする。またこの検索履歴に最終的に得られた検索条件を格納するステップを有する文書検索方法を特徴とする。また本発明は、このような機能を備えた文書検索装置を特徴とする。

【0018】

また本発明は、少なくとも1つの文書の識別子をキーとしてこの検索履歴を検索し、合致する文書の識別子を有する検索履歴のレコードを取得して表示する文書検索方法を特徴とする。

10

【0019】

また本発明は、指定された検索条件に基づいてこの検索履歴を検索し、指定された検索条件に類似する検索条件を有する検索履歴のレコードを取得して表示する文書検索方法を特徴とする。

【0020】

【発明の実施の形態】

以下、本発明の実施の形態について説明する。

【0021】

まず、本発明の実施形態のシステム構成を図1に示す。本実施形態におけるシステムは、ディスプレイ114、キーボード115、中央演算処理装置(CPU)116、磁気ディスク装置117、フロッピディスクドライブ(FDD)118、主メモリ121およびこれらを結ぶバス120から構成される計算機である。

20

【0022】

磁気ディスク装置117は二次記憶装置の一つであり、検索プロファイル格納ファイル109、登録文書格納ファイル110、フィードバック履歴検索ファイル111、検索シーケンス格納ファイル112、およびプロファイル履歴格納ファイル113が格納される。FDD118を介してフロッピディスク119に格納されている情報が、主メモリ121あるいは磁気ディスク装置117へ読み込まれる。

30

【0023】

主メモリ121には、情報検索プログラム101および履歴検索プログラム201が格納される。情報検索プログラム101は、検索プロファイル生成プログラム102、類似文書検索プログラム103、検索プロファイル修正プログラム104、フィードバック履歴保存プログラム105、および履歴保存プログラム106で構成される。ここで履歴保存プログラム106は、検索シーケンス保存プログラム107およびプロファイル保存プログラム108で構成される。

【0024】

履歴検索プログラム201は、検索プロファイル生成プログラム102、類似文書検索プログラム103、類似プロファイル検索プログラム203、および類似検索シーケンス検索プログラム204で構成される。ここで、検索プロファイル生成プログラム102および類似文書検索プログラム103は、情報検索プログラム101を構成する検索プロファイル生成プログラム102、類似文書検索プログラム103とそれぞれ同一のものである。

40

【0025】

ディスプレイ114は、検索条件を入力するための案内画面及び検索結果を表示するための表示装置である。キーボード115は、検索条件やその他の指示を入力するための入力装置である。図1のシステム構成は、いわば検索サーバにディスプレイ114及びキーボード115が接続され、ユーザがこのディスプレイ114及びキーボード115を介して検索サーバにアクセスする構成であるが、ディスプレイ114及びキーボード115を設

50

けるクライアント端末とネットワークを介して検索サーバにアクセスする、いわゆるクライアント - サーバ型のシステムであってもよい。

【0026】

なおフロッピディスク119など記憶媒体上に情報検索プログラム101および履歴検索プログラム201を格納し、FDD118などの駆動装置を介して当該計算機の主メモリ121に読み込み、CPU116によって実行することが可能である。

【0027】

次に本実施形態における全体的な処理の流れを図4のPAD図を用いて説明する。

【0028】

本発明の実施形態では、図2に示した従来技術におけるステップ304～ステップ306の代わりに、ステップ1004において情報検索プログラム101を起動する。また、ステップ308～ステップ309の代わりに、ステップ1006において履歴検索プログラム201を起動する。

10

【0029】

まずユーザは最初に図5に示す画面900において、情報検索プログラム101を用いて文書情報の検索を行なうのか、または過去の検索シーケンスの履歴を検索するのを選択する。テキストボックス901に種文書を入力して、「類似文書を検索する」ボタン902を押下すると、情報検索プログラム101が起動されて、種文書と類似した文書を検索する。同様に「類似ユーザ、類似プロファイルを検索する」ボタン903を押下すると、履歴検索プログラム201が起動されて、類似検索シーケンスを検索することになる。

20

【0030】

本実施形態における処理では、ステップ1001において、以下に示す1002～1006の処理をシステム終了まで繰り返す。まずステップ1002において、ユーザからの入力を受け付ける。次にステップ1003において、ユーザからの入力類似文書の検索要求であると判定された場合、ステップ1004で情報検索プログラム101を起動する。次にステップ1005において、ユーザからの入力類似する検索シーケンスの検索要求であると判定された場合、ステップ1006で履歴検索プログラム201を起動する。以上が本実施形態におけるおおまかな処理の流れである。

【0031】

以下、ステップ1004で起動される情報検索プログラム101を説明する。

30

【0032】

情報検索プログラム101は、ユーザが入力した文書（以下、種文書と呼ぶ）と内容の類似する文書を登録文書から選択し、ユーザに提示する類似文書検索プログラム103を有する。また、検索結果をユーザが評価してそれをもとに検索をやり直す、いわゆるレリバンсフィードバック機能を備えている。ここで、ユーザが最初に種文書を入力し、レリバンсフィードバックを繰り返して、検索を終了するまでの一連の動作を検索シーケンスと呼ぶことにする。

【0033】

まず以下に、本発明における情報検索プログラム101について図6を用いて説明する。

図6は本発明における情報検索プログラム101の構成を示す図である。

40

【0034】

本発明における情報検索プログラム101は、検索プロファイル生成プログラム102と、類似文書検索プログラム103と、検索プロファイル修正プログラム104と、フィードバック履歴保存プログラム105と、履歴保存プログラム106から構成される。またアクセスされるファイルは、検索プロファイル格納ファイル109、登録文書格納ファイル110、フィードバック履歴格納ファイル111、検索シーケンス履歴格納ファイル112及びプロファイル履歴格納ファイル113である。ここで履歴保存プログラム106は、検索シーケンス保存プログラム107と、プロファイル保存プログラム108から構成される。以下、それぞれのプログラムについて説明する。

【0035】

50

まず、検索プロファイル生成プログラム102は、ユーザ100から検索条件を受け付け、ユーザ100の検索したい内容を要素として持つデータ（以下、検索プロファイルと呼ぶ）を生成するプログラムである。

【0036】

本プログラムは、ユーザ100から入力される種文書を受け付ける。そしてその種文書から検索プロファイルを生成して、検索プロファイル格納ファイル109に書き込む。この検索プロファイルが検索条件となる。

【0037】

ここで検索プロファイルの内容と生成方法の例を図7を用いて説明する。本図の例では、ユーザ100が種文書501として、「サッカーに続き高校野球が開幕・・・」という文書を入力した。このとき検索プロファイル生成プログラム102は、種文書501から自立語の可能性のある検索語の文字列（以下、特徴文字列とよぶ）を1つ以上抽出し、各特徴文字列について類似文書検索のときに用いる重み値として種文書501内での出現回数を算出する。そして抽出した特徴文字列とその重みを検索プロファイル503に書き込む。以上が、検索プロファイル生成プログラム102の説明である。

【0038】

次に、類似文書検索プログラム103は、登録文書格納ファイル110に登録された各文書に対し、指定された種文書に概念的な類似度に応じた得点を付与し、類似度の高い文書を検索結果としてユーザ100に表示するプログラムである。

【0039】

本プログラムは、検索プロファイル格納ファイル109に格納された検索プロファイルに基づき、登録文書格納ファイル110に格納された各文書に対して類似度を算出し、類似度の高い少なくとも1つの文書をユーザ100に表示する。ここで類似度の算出方法としては、以下の式1のようなものを用いる。

【0040】

【数1】

数 1

$$S(D) = \sum_i^N \{ Frq(i) \times w(i) \} \quad \dots(式1)$$

この式で、 $S(D)$ は登録文書格納ファイル110内の文書番号Dの文書の類似度であり、 $Frq(i)$ は種文書に出現する単語iの文書番号Dの文書における出現回数であり、 $w(i)$ は検索プロファイル格納ファイル109内の単語iの重み、すなわち種文書における単語iの出現回数である。ここで類似度算出式としてはこれ以外のものを用いてもよい。以上が、類似文書検索プログラム103の説明である。

【0041】

次に、検索プロファイル修正プログラム104は、検索結果文書に対してユーザ100が、「所望の文書である」或いは「所望の文書ではない」といった評価を入力した場合、その評価に基づいて検索プロファイルの内容を修正するプログラムである。本プログラムは、ユーザ100から、ユーザ100が評価した文書と、その評価を受け付ける。そしてその評価に基づいて検索プロファイル格納ファイル109に格納されている検索プロファイルの内容を修正する。

【0042】

ここで検索プロファイルの修正方法の例を図7を用いて説明する。本図の例では、「サッカーの試合の結果は・・・」という文書504が検索結果としてユーザ100に提示され

10

20

30

40

50

、その文書に対してユーザ100が「所望でない」という評価を与えたときの例である。このとき、本プログラムにおいて、検索プロファイル生成プログラム102と同様に、文書504から特徴文字列をその出現回数505を抽出する。そして、生成されている検索プロファイル503の該当する文字列の重みをその出現回数505によって減算する。その結果、検索プロファイル503の内容は修正後の検索プロファイル506の内容に修正される。一致する文字列の出現回数を減算するのは、ユーザが所望しない文書であるので、文字列が一致してもたまたま一致しただけであり、ユーザには興味のない文字列であろうと推定して重みを減じるのである。なお減算の結果として重みが負になってもよいし、0で止まりとしてもよい。

【0043】

一方、「所望である」という評価がされた文書については、生成されている検索プロファイル503の該当する文字列の重みにその出現回数505を加算する。以上が、検索プロファイル修正プログラム104の説明である。

【0044】

ユーザは修正された検索プロファイルに基づいて類似文書検索プログラム103を実行させ、登録文書格納ファイル110を再検索させることができる。そして再検索の結果を上記手順に従って再評価し、再び検索プロファイル修正プログラム104を実行させて検索プロファイル格納ファイル109を修正することができる。このようにしてレリバンスフィールドバックを繰り返すことにより、ユーザの希望する文書に接近することができる。

【0045】

次に、フィールドバック履歴保存プログラム105は、ユーザ100がレリバンスフィールドバックを行なった際にその履歴を保存しておくプログラムである。本プログラムは、ユーザ100が評価した文書の識別子とその評価を、フィールドバック履歴格納ファイル111に追加書き込みする。ユーザ100がレリバンスフィールドバックを繰り返して、次々と検索を行なうたびに、その際に評価した文書の識別子とその評価が追加書き込みされることになる。

【0046】

ここでフィールドバック履歴格納ファイル111に格納される情報の例を図8に示す。本図の例では、ユーザ「Matsui」が最初に「005.txt」というファイル名の文書を種文書として入力し、その後検索シーケンスを通じて得られた文書65と文書50を「所望である」と評価し、文書35を「所望でない」と評価した、というレリバンスフィールドバックを行なった場合である。そしてこのとき、「0005.pfl」というファイル名で保存されている検索プロファイルが検索に用いられており、レリバンスフィールドバックのたびにこの検索プロファイルの内容が修正されている。以上が、フィールドバック履歴保存プログラム105の説明である。

【0047】

次に、履歴保存プログラム106を構成する検索シーケンス保存プログラム107およびプロファイル保存プログラム108について説明する。まず、検索シーケンス保存プログラム107は、ユーザ100がレリバンスフィールドバックを繰り返して、検索を終了する際に、一連の評価した文書とそれぞれの評価を、検索シーケンス履歴格納ファイル112に追加保存するプログラムである。

【0048】

本プログラムは、ユーザ100からログアウト指示などの検索シーケンス終了宣言を受け付けたときに起動され、フィールドバック履歴格納ファイル111に格納された情報を読み込み、その内容を検索シーケンス履歴格納ファイル112に追加書き込みする。このときの例を図9に示す。本図の例は、過去に検索シーケンスID(識別子)1~4を付与された検索シーケンスが行われていたが、図の最下行の部分が新たに追加書き込みされ、検索シーケンスIDとして5が付与されたという例である。検索シーケンス履歴格納ファイル112の各行が各々検索シーケンスによって得られた最終結果のレコードを示している。以上が検索シーケンス保存プログラム106の説明である。

10

20

30

40

50

【0049】

次に、プロファイル保存プログラム108は、ユーザ100がレリバンスフィールドバックを繰り返した結果、内容が修正され洗練されてきた検索プロファイルを保存するプログラムである。

【0050】

本プログラムは、ユーザ100から検索シーケンス終了宣言を受け付けたときに起動され、検索プロファイル格納ファイル109に格納された情報を読み込み、その内容をプロファイル履歴格納ファイル113に保存する。プロファイル履歴格納ファイル113の内容は、複数のユーザについて各ユーザの検索プロファイル格納ファイル109、すなわち修正後の検索プロファイル506を蓄積したものである。以上がプロファイル保存プログラム108の説明である。

10

【0051】

以上の検索シーケンス履歴格納ファイル112やプロファイル履歴格納ファイル113は、後述するように、履歴検索プログラム201により後の検索者が、「以前に同様の検索を行なった人物はいないだろうか」といった検索をする際の検索対象として用いられる。

【0052】

以上が本実施形態における情報検索プログラム101を構成する各プログラムの説明である。

【0053】

ここで、検索プロファイル生成プログラム102において、検索プロファイル格納ファイル109に書き込む文字列は、抽出した特徴文字列全てを書き込む方法であっても良いし、抽出した特徴文字列のうち種文書内での出現回数が上位のものから順に所定数に限定しても良い。また特徴文字列を抽出する方法は、「特開平11-143902号公報」に開示されている方法を用いても良いし、形態素解析を用いて各文書中の単語を抽出する方法や、その他の方法であっても良い。

20

【0054】

次に、ステップ1006において起動される履歴検索プログラム201について図10を用いて説明する。図10は本実施形態における履歴検索プログラム201の構成を示す図である。

【0055】

本発明における履歴検索プログラム201は、過去に情報検索プログラム101を用いて行なわれた検索の履歴を検索する。これにより、類似した観点を持つユーザが誰であるかといった情報を取得したり、情報検索プログラム101に入力する検索条件として過去に用いられたものを取得し再利用することを可能にするものである。

30

【0056】

本実施形態における履歴検索プログラム201は、検索プロファイル生成プログラム102と、類似文書検索プログラム103と、類似プロファイル検索プログラム203と、類似検索シーケンス検索プログラム204とから構成される。またアクセスされるファイルは、検索プロファイル格納ファイル109、登録文書格納ファイル110、プロファイル履歴格納ファイル113および検索シーケンス履歴格納ファイル112である。

40

【0057】

ここで、検索プロファイル生成プログラム102と、類似文書検索プログラム103と、類似プロファイル格納ファイル109と、類似文書登録文書格納ファイル110と、プロファイル履歴格納ファイル113と、検索シーケンス履歴格納ファイル112は、上述した情報検索プログラム101に関して説明したものと同一のものである。すなわち、これらのプログラム及びファイルは、情報検索プログラム101と、履歴検索プログラム201とで共有されている。以下、それぞれのプログラムについて説明する。

【0058】

まず、検索プロファイル生成プログラム102は、ユーザ200から入力される種文書から、履歴検用の検索プロファイルを作成するためのプログラムである。本プログラムの

50

処理内容は、上述したのでここでは説明を省く。

【 0 0 5 9 】

次に、類似文書検索プログラム 1 0 3 は、登録された各文書に対し、指定された種文書に概念的な類似度に応じた得点を付与し、類似度の高い文書を取得するプログラムである。ここで取得された文書は、後述するように過去に検索を行なったユーザのレリバンスフィードバックの履歴として記録された文書と照合するために用いられる。たとえば、ここで取得された文書を過去に「所望である」と評価したユーザは、このとき履歴検索を行なっているユーザと検索の観点

(目的) が似ていると判断できるからである。

【 0 0 6 0 】

本プログラムは、検索プロファイル格納ファイル 1 0 9 に格納された検索プロファイルに基づき、登録文書格納ファイル 1 1 0 に格納された各文書に対して類似度を算出する。そして、類似度が一定の閾値を超える文書の識別子を検索結果として取得する。

【 0 0 6 1 】

次に、類似プロファイル検索プログラム 2 0 3 は、ユーザ 2 0 0 によって入力された種文書に基づいて生成された検索プロファイルと類似した、過去に用いられた検索プロファイルを検索するプログラムである。

【 0 0 6 2 】

本プログラムは、検索プロファイル格納ファイル 1 0 9 に格納された検索プロファイルに基づき、プロファイル履歴格納ファイル 1 1 3 に格納された各検索プロファイルに対して類似度を算出する。そして、類似度が一定の閾値を超える文書の識別子を検索結果として取得する。ここで類似度の算出方法としては類似文書検索プログラム 1 0 3 と同様の方法を用いても良いし、別の方法を用いても良い。以上が、類似プロファイル検索プログラム 2 0 3 の説明である。

【 0 0 6 3 】

次に、類似検索シーケンス検索プログラム 2 0 4 は、上記の類似文書検索プログラム 1 0 3 および類似プロファイル検索プログラム 2 0 3 の検索結果にもとづき、過去に行なわれた検索シーケンスのうち、検索の観点

(目的) が類似するものを抽出するプログラムである。

【 0 0 6 4 】

本プログラムでは、まず類似文書検索プログラム 1 0 3 および類似プロファイル検索プログラム 2 0 3 の検索結果を取得する。そして、検索シーケンス履歴格納ファイル 1 1 2 に格納された過去の各検索シーケンスと照合して、類似検索シーケンスを照合結果としてユーザ 2 0 0 に提示する。

【 0 0 6 5 】

ここで照合処理の例を図 1 1 に示す。本図の例では、類似文書検索プログラム 1 0 3 の検索結果 8 0 1 として文書 8 と文書 3 という 2 つの文書が取得された。また、類似プロファイル検索プログラム 2 0 3 の検索結果 8 0 2 として 0 0 0 2 .pfl と 0 0 0 3 .pfl という 2 つのプロファイルが取得された。このとき類似検索シーケンス検索プログラム 2 0 4 は、検索シーケンス履歴格納ファイル 1 1 2 に格納された各検索シーケンスのうち、「所望と評価した文書」の項目に検索結果 8 0 1、「検索プロファイル」の項目に検索結果 8 0 2 の内容を含むものを抽出する。ここでは、検索シーケンス ID : 3 の「所望と評価した文書」の中に文書 3 と文書 8 が含まれ、検索シーケンス ID : 2 の「検索プロファイル」の項目に 0 0 0 2 .pfl が含まれる。そこで、検索シーケンス ID : 3 と 2 が類似した検索シーケンスであるとして、該当シーケンスを実行したユーザ名や、利用されたプロファイルを履歴検索を行なっているユーザ 2 0 0 に提示する。

【 0 0 6 6 】

このとき、以上の例では検索結果 8 0 1 や検索結果 8 0 2 の内容を含んでいるだけで類似シーケンスであると判定するとしたが、所定数以上含んでいないと類似シーケンスと判定しないなどの方法や、所定のプログラムで類似度を算出して、類似度が所定の閾値を超え

10

20

30

40

50

たものだけを類似シーケンスと判定するなどの方法をとっても良い。

【0067】

また、履歴検索を行なっているユーザ200の所望しない概念を表すプロファイルを作成し、「不要と評価した文書」の項目に格納された文書との類似度を用いるなどの方法を同時に利用しても良い。また検索条件として少なくとも1つの文書の識別子を入力し、この文書識別子をキーとして検索シーケンス履歴格納ファイル112を検索し、指定された文書識別子を「所望と評価した文書」に含むレコードを抽出して表示してもよい。以上が、類似検索シーケンス検索プログラム204の説明である。また以上が本発明における履歴検索プログラム201を構成する各プログラムの説明である。

【0068】

次に、ステップ1004において起動される情報検索プログラム101の処理の流れを図12のPAD図を用いて説明する。まず、ステップ1101において、種文書が入力されたと判定された場合、次のステップ1102を実行する。ステップ1102では、検索プロファイル生成プログラム102により、検索プロファイルを生成し、検索プロファイル格納ファイル109に保存する。次に、ステップ1103において、検索結果に対する評価が入力されたと判定された場合、次のステップ1104、ステップ1105を実行する。ステップ1104では、検索結果に対する評価に基づき、検索プロファイル修正プログラム104により、検索プロファイル格納ファイル109に格納されている検索プロファイルの内容を修正する。ステップ1105では、ステップ1103で入力された評価をフィードバック履歴保存プログラム105により、フィードバック履歴格納ファイル111

10

20

【0069】

次に、ステップ1106において、検索プロファイル格納ファイル109に格納された検索プロファイルの内容に基づき、類似文書検索プログラム103により類似文書を検索する。次に、ステップ1107において、検索シーケンスの終了要求が入力された、またはタイムアウトが発生したと判定された場合、次のステップ1108とステップ1109を実行する。ステップ1108では、検索シーケンス保存プログラム107により、フィードバック履歴格納ファイル111内の情報を検索シーケンス履歴格納ファイル112に追加書き込みする。ステップ1109では、プロファイル保存プログラム108により、検索プロファイル格納ファイル109内の情報をプロファイル履歴格納ファイル113に保

30

【0070】

以上がステップ1004において起動される情報検索プログラム101の処理の流れである。以上の処理により、ユーザはレリバンスフィードバックを用いた類似文書検索が可能となる。また、検索シーケンス終了時にはどのようなレリバンスフィードバックを行なったかという情報と、レリバンスフィードバックによって洗練された検索プロファイルを保存することが可能となる。

【0071】

次に、ステップ1006において起動される履歴検索プログラム201の処理の流れを図13のPAD図を用いて説明する。まず、ステップ1201において、検索プロファイル生成プログラム102により、種文書から検索プロファイルを作成し、検索プロファイル格納ファイル109に書き込む。次に、ステップ1202において、類似文書検索プログラム103により、ステップ1201で生成した検索プロファイルに基づき類似文書を検索する。

40

【0072】

次に、ステップ1203において、類似プロファイル検索プログラム203により、ステップ1201で生成した検索プロファイルに基づきプロファイル履歴格納ファイル113に格納された類似プロファイルを検索する。次に、ステップ1204において、類似検索

50

シーケンス検索プログラム204により、ステップ1202およびステップ1203の検索結果をもとに、検索シーケンス履歴格納ファイル112内の情報のうち類似検索シーケンスを抽出して、ユーザに提示する。ここで、類似検索シーケンスの抽出方法は、類似検索シーケンス検索プログラム204の説明で上述したとおりである。

【0073】

このときのユーザへ提示する情報の例を図14に示す。図14は、図5のテキストボックス901に種文書を入力して「類似ユーザ、類似プロファイルを検索する」ボタン903を押下し、履歴検索プログラム201が起動された結果として表示される画面である。本図の例では、表示エリア1301に、「Etohさん」や「Shimizuさん」といった過去に類似観点で検索を行なったユーザの名前が提示される。また、表示エリア1302に、それらのユーザが行なった検索において生成・修正された検索プロファイルの内容が提示される。

10

【0074】

これらの情報をもとにユーザは、自分の知りたい情報について「Etohさん」や「Shimizuさん」が知っている可能性があることがわかり、彼らにノウハウなどを聞けば良いということがわかる。

【0075】

また、検索プロファイルが提示されるため、それを入力にして検索を行ったり、提示された検索プロファイルに含まれる文字列を、自分の種文書や検索プロファイルに追加して検索するなどといった応用が可能となる。

20

【0076】

以上がステップ1006において起動される履歴検索プログラム201の処理の流れである。以上の処理により、どのユーザがどのようなノウハウを持っているかを検索して知ることができ、かつレリバンスフィールドバックによって洗練された検索プロファイルといった検索ノウハウを取得して再利用することができる。

【0077】

以上が本発明による実施の形態である。なお、本発明の実施形態では、ユーザが検索条件として種文書を入力し類似文書を検索するものとしたが、検索タームを羅列したり、それらのブーリアン条件で検索するものとしても良い。

【0078】

また、本発明の実施形態では、ユーザが検索シーケンスを終了した時点で、その履歴を検索シーケンス履歴格納ファイル112に保存するものとしたが、検索シーケンスの途中でユーザが評価を入力するたびに、検索シーケンス格納ファイル112の内容を書き換える形態にしても良い。その場合、実行中の検索シーケンスも、類似検索シーケンス検索プログラム204の検索対象となることになる。

30

【0079】

また、検索条件を登録しておき、条件に合致する文書情報が発生した際にユーザに該文書情報を配信する、文書情報配信システムにおいても適用できる。すなわち、文書情報配信システムにおいて各ユーザが登録している検索プロファイルそれぞれに対して、レリバンスフィールドバックの履歴を保存しておき、その履歴を類似検索シーケンス検索プログラム204を用いて検索することにより、類似した観点をもつユーザと、その検索プロファイルが抽出可能となる。

40

【0080】

また、検索結果に対する評価の入力方法は、ひとつの検索結果文書に対しユーザが評価を入力するものとしても良いし、複数の検索結果文書に対してそれぞれ評価を一度に入力するものとしても良い。また本実施例では、評価の種類は「所望である」と「所望でない」の2種類としたが、「やや所望である」「やや所望ではない」といったものを用いて多段階の評価ができるようにしても良い。

【0081】

【発明の効果】

50

以上述べたように本発明によれば、各ユーザが行なったレリバンスフィードバックの履歴と、レリバンスフィードバックによって修正された検索条件を対応づけ、それぞれの検索シーケンスを単位に履歴管理する。

【0082】

このことにより、ユーザが行なった検索の試行錯誤の履歴と、その結果修正・洗練された検索条件が個人プロファイルとして自動的に蓄積される。そのため「各ユーザが何を知っているか」という情報を、各ユーザの負担なく登録することが可能となる。

【0083】

また、蓄積された検索の試行錯誤の履歴や検索条件を、他のユーザが検索することができるため、「この情報を知っているのは誰か」という情報を適切に検索することが可能となる。

10

【図面の簡単な説明】

【図1】本発明の実施形態の構成を示す図である。

【図2】従来技術1による処理の流れを示すPAD図である。

【図3】従来技術2による処理の流れを示すPAD図である。

【図4】本発明の実施形態における全体的な処理の流れを示す図である。

【図5】本発明の実施形態におけるディスプレイ900に表示される初期画面の例を示す図である。

【図6】本発明の実施形態における情報検索プログラム101の構成を示す図である。

【図7】本発明の実施形態における検索プロファイルの内容と生成方法の例を示す図である。

20

【図8】本発明の実施形態におけるフィードバック履歴格納ファイル111に格納される情報の例を示す図である。

【図9】本発明の実施形態における検索シーケンス保存プログラム107が、検索シーケンス履歴格納ファイル112に情報を追加書込みする例を示す図である。

【図10】本発明の実施形態における履歴検索プログラム201の構成を示す図である。

【図11】本発明の実施形態における類似検索シーケンス検索プログラム204が、保存された検索シーケンスの履歴を検索条件と照合する例を示す図である。

【図12】本発明の実施形態における情報検索プログラム101の処理の流れを示すPAD図である。

30

【図13】本発明の実施形態における履歴検索プログラム201の処理の流れを示すPAD図である。

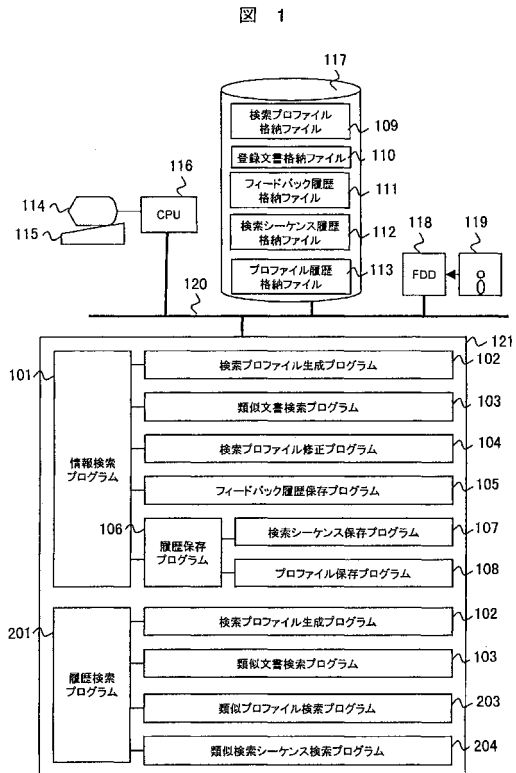
【図14】本発明の実施形態における類似検索シーケンス検索プログラム204による検索結果として、ユーザに提示される画面の例である。

【符号の説明】

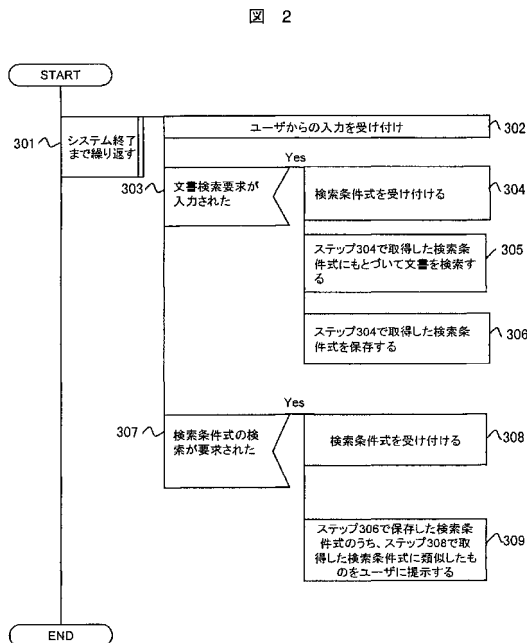
101 情報検索プログラム、102 検索プロファイル生成プログラム、103 類似文書検索プログラム、104 検索プロファイル修正プログラム、105 フィードバック履歴保存プログラム、106 履歴保存プログラム、107 検索シーケンス保存プログラム、108 プロファイル保存プログラム、109 検索プロファイル格納ファイル、110 登録文書格納ファイル、111 フィードバック履歴格納ファイル、112 検索シーケンス履歴格納ファイル、113 プロファイル履歴格納ファイル、201 履歴検索プログラム、203 類似プロファイル検索プログラム、204 類似検索シーケンス検索プログラム

40

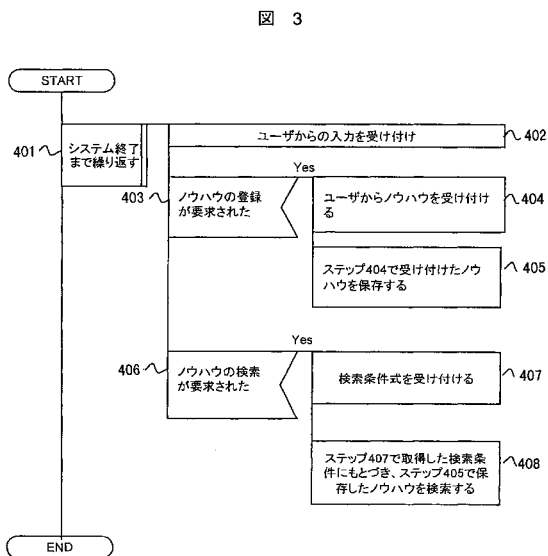
【 図 1 】



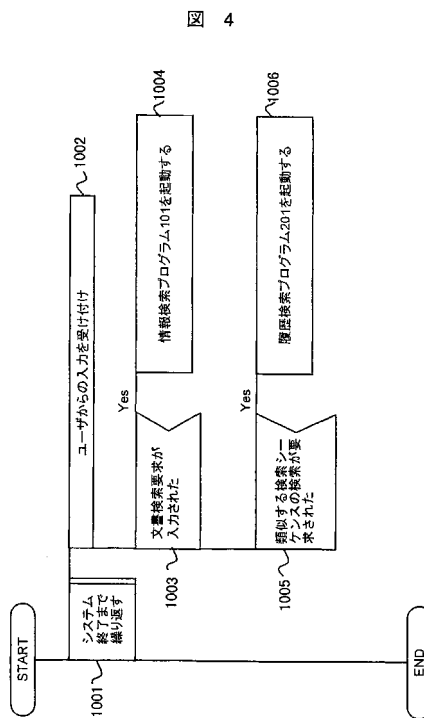
【 図 2 】



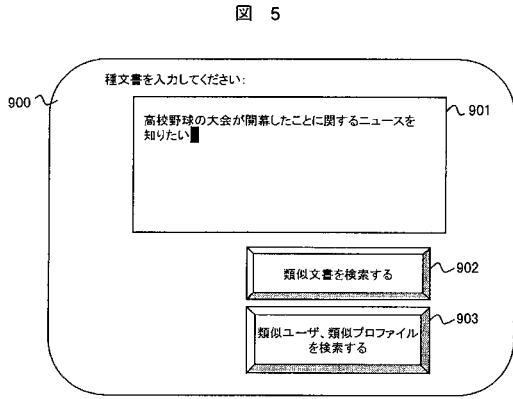
【 図 3 】



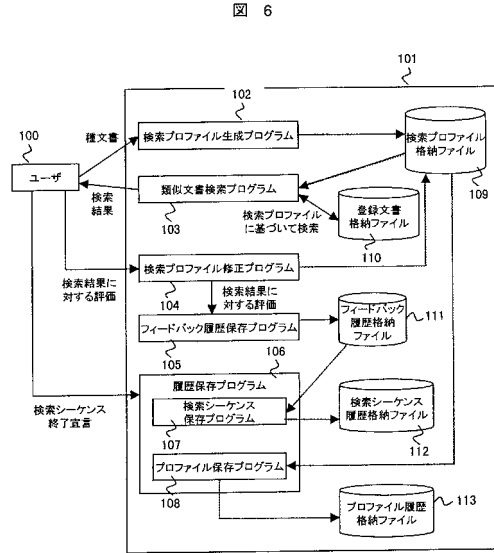
【 図 4 】



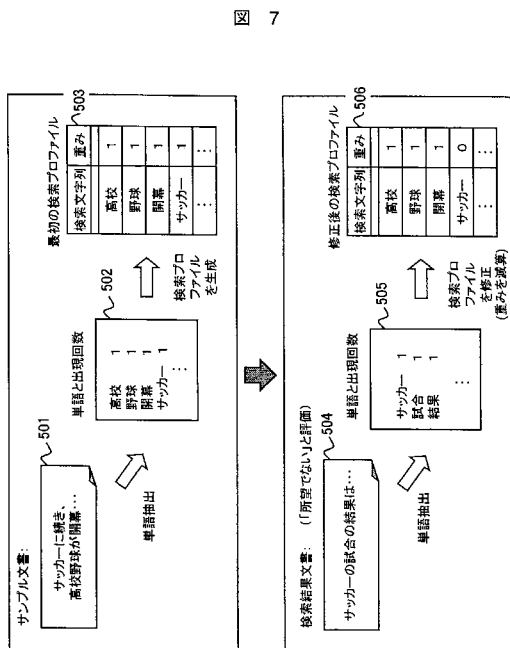
【 図 5 】



【 図 6 】



【 図 7 】



【 図 8 】

図 8

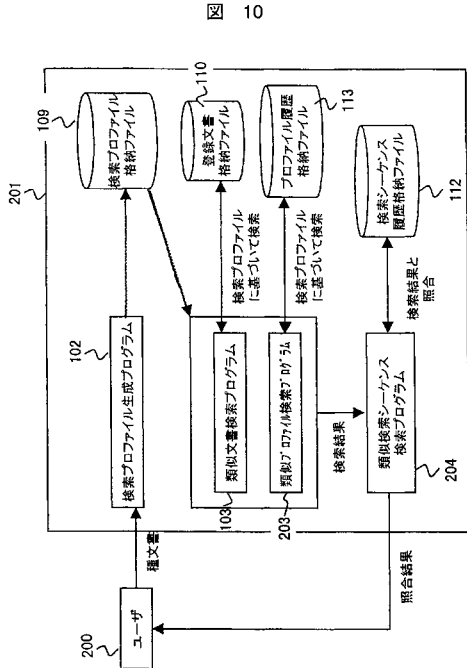
実行中のユーザ	種文書	所望と評価した文書	不要と評価した文書	検索プロフィール
Matsui	005.txt	文書65、文書50	文書35	0005.Pfl

【 図 9 】

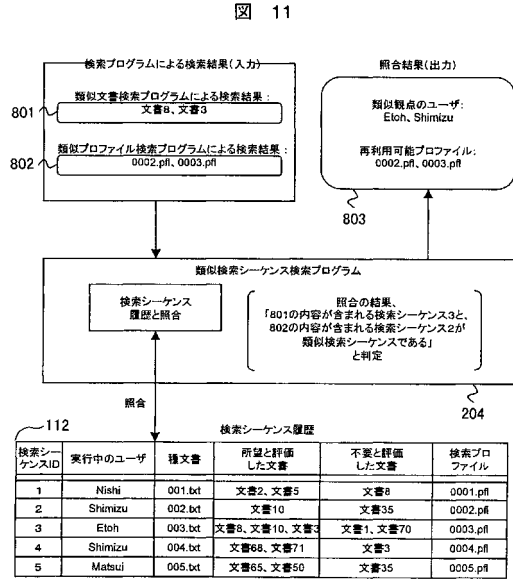
図 9

検索シーケンスID	実行中のユーザ	種文書	所望と評価した文書	不要と評価した文書	検索プロフィール
1	Nishi	001.txt	文書2、文書5	文書8	0001.pfl
2	Shimizu	002.txt	文書10	文書35	0002.pfl
3	Etoh	003.txt	文書9、文書10、文書3	文書1、文書70	0003.pfl
4	Shimizu	004.txt	文書68、文書71	文書3	0004.pfl
5	Matsui	005.txt	文書65、文書50	文書35	0005.pfl

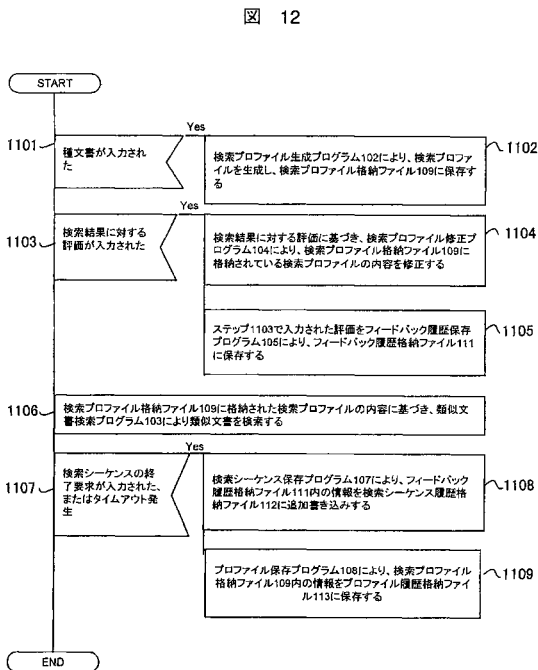
【 図 1 0 】



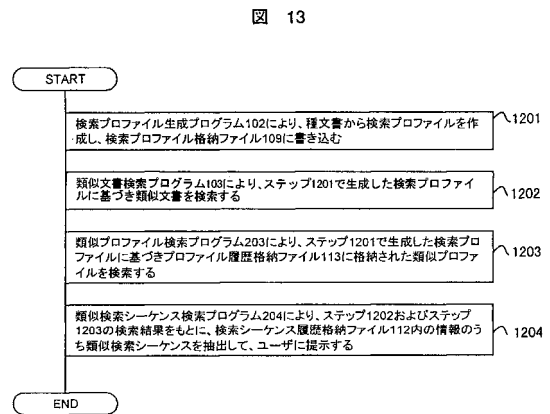
【 図 1 1 】



【 図 1 2 】

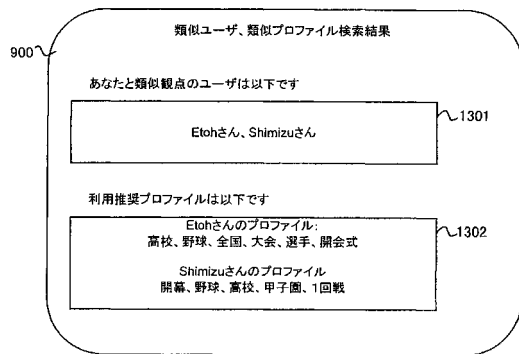


【 図 1 3 】



【 図 1 4 】

図 14



フロントページの続き

- (72)発明者 多田 勝巳
神奈川県川崎市幸区鹿島田890番地 株式会社日立製作所 ビジネスソリューション開発本部内
- (72)発明者 菅谷 奈津子
神奈川県川崎市幸区鹿島田890番地 株式会社日立製作所 ビジネスソリューション開発本部内
- (72)発明者 松林 忠孝
神奈川県川崎市幸区鹿島田890番地 株式会社日立製作所 ビジネスソリューション開発本部内
- (72)発明者 山口 明彦
神奈川県川崎市幸区鹿島田890番地 株式会社日立製作所 ビジネスソリューション開発本部内
- (72)発明者 徳永 幹彦
神奈川県横浜市戸塚区戸塚町5030番地 株式会社日立製作所 ソフトウェア開発本部内

審査官 辻本 泰隆

- (56)参考文献 特開2000-132559(JP,A)
特開平07-134717(JP,A)
特開平10-055369(JP,A)
特開平10-320416(JP,A)
特開平10-063685(JP,A)
神場 知成, ユーザプロフィール管理エージェントの提案と試作, 情報処理学会研究報告 Vol. 97 No. 2, 日本, 社団法人情報処理学会, 1997年 1月16日, 第1-8頁

- (58)調査した分野(Int.Cl., DB名)
G06F 17/30
JSTPlus(JDream2)
NRIサイバーパテント