

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2019-164409
(P2019-164409A)

(43) 公開日 令和1年9月26日 (2019.9.26)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 16/00 (2019.01)	G06F 17/30	350C
G06F 16/30 (2019.01)	G06F 17/30	170A
	G06F 17/30	340B

審査請求 未請求 請求項の数 8 O L (全 15 頁)

(21) 出願番号	特願2018-50504 (P2018-50504)	(71) 出願人	000233055 株式会社日立ソリューションズ 東京都品川区東品川四丁目12番7号
(22) 出願日	平成30年3月19日 (2018.3.19)	(74) 代理人	110000279 特許業務法人ウィルフォート国際特許事務所
		(72) 発明者	井上 直樹 東京都品川区東品川四丁目12番7号 株式会社日立ソリューションズ内
		(72) 発明者	大谷 聡 東京都品川区東品川四丁目12番7号 株式会社日立ソリューションズ内

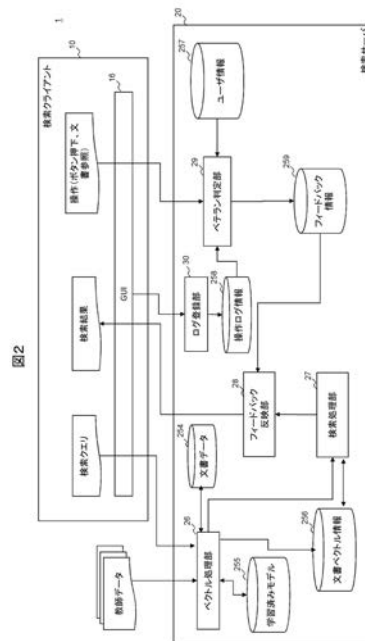
(54) 【発明の名称】 文書検索装置、文書検索方法、及び文書検索プログラム

(57) 【要約】

【課題】 利用者の要求に合致する文書を容易かつ適切に検索できるようにする。

【解決手段】 検索サーバ20において、検索クエリを受け付け、受け付けた検索クエリをベクトルに変換するベクトル処理部26と、検索クエリのベクトルと、検索対象文書のベクトルとに基づいて、検索クエリと検索対象文書との類似度を決定する検索処理部27と、フィードバック情報259中に検索クエリのベクトルと同一又は類似のベクトルを有する検索クエリに対応するフィードバック情報が存在する場合に、検索処理部27により決定された類似度を、フィードバック情報の評価値情報に基づいて調整して調整後類似度を決定し、調整後類似度に基づいて、検索クエリと類似する文書の検索結果を作成し、検索結果を含む検索結果画面を所定の装置に表示させるフィードバック反映部28と、を備えるように構成する。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

検索クエリを受け付けて、検索クエリに基づいて、複数の検索対象文書の中から類似する文書を検索する文書検索装置であって、

複数の前記検索対象文書の文章に対応するベクトルを記憶するベクトル情報記憶部と、過去に利用した検索クエリのベクトルと、前記検索クエリと同一又は類似の検索クエリにより検索された文書を特定する情報と、前記検索された文書に対する利用者の操作に基づく評価値情報とを含むフィードバック情報を記憶するフィードバック情報記憶部と、前記検索クエリを受け付ける受付部と、

前記受け付けた検索クエリをベクトルに変換するベクトル変換部と、

10

前記検索クエリのベクトルと、前記検索対象文書のベクトルとに基づいて、前記検索クエリと前記検索対象文書との類似度を決定する類似度決定部と、

前記フィードバック情報記憶部に前記検索クエリのベクトルと同一又は類似のベクトルを有する検索クエリに対応するフィードバック情報が存在する場合に、前記類似度決定部により決定された前記類似度を、前記フィードバック情報の前記評価値情報に基づいて調整して調整後類似度を決定する調整処理部と、

前記調整後類似度に基づいて、前記検索クエリと類似する文書の検索結果を作成する検索結果作成部と、

前記検索結果を含む検索結果画面を所定の装置に表示させる表示制御部と、を備える文書検索装置。

20

【請求項 2】

前記評価値情報は、利用者が検索された文書を参照する操作、利用者が検索された文書に対して役に立つことを表明する操作、又は利用者が検索された文書に対して役に立たないことを表明する操作の少なくともいずれか一つに対応する評価値情報である請求項 1 に記載の文書検索装置。

【請求項 3】

前記表示制御部は、前記検索結果画面に、前記検索結果の文書の情報に対応して、前記文書に対する利用者の操作を受け付けるための操作受付領域を表示させ、

前記操作受付領域に対する前記利用者の操作を受け付けて、前記フィードバック情報における前記操作受付領域に対応する文書についての評価値情報を更新する操作反映評価値更新部をさらに有する

30

請求項 1 又は請求項 2 に記載の文書検索装置。

【請求項 4】

前記利用者の属性情報を記憶する属性情報記憶部と、

前記利用者の属性情報に基づいて、所定の属性を持つ利用者を特定し、特定した利用者による検索によって参照された文書の調整後類似度が高くなるように前記文書に対応する評価値情報を更新する属性反映評価値更新部と、をさらに有する

請求項 1 から請求項 3 のいずれか一項に記載の文書検索装置。

【請求項 5】

複数の利用者の検索時における操作ログを取得する操作ログ取得部と、

40

前記取得した操作ログに基づいて、所定の行動特性を有する利用者を特定し、特定した利用者による検索によって参照された文書の調整後類似度が高くなるように前記文書に対応する評価値情報を更新する行動特性反映評価値更新部と、をさらに有する

請求項 1 から請求項 4 のいずれか一項に記載の文書検索装置。

【請求項 6】

前記所定の行動特性とは、一回の検索当たりの参照した文書の数が所定の閾値よりも少ない特性である

請求項 5 に記載の文書検索装置。

【請求項 7】

検索クエリを受け付けて、検索クエリに基づいて、複数の検索対象文書の中から類似す

50

る文書を検索する文書検索装置による文書検索方法であって、

前記文書検索装置には、複数の前記検索対象文書の文章に対応するベクトルを記憶するベクトル情報記憶部と、過去に利用した検索クエリのベクトルと、前記検索クエリと同一又は類似の検索クエリにより検索された文書を特定する情報と、前記検索された文書に対する利用者の操作に基づく評価値情報とを含むフィードバック情報を記憶するフィードバック情報記憶部と、が備えられており、

前記検索クエリを受け付けるステップと、

前記受け付けた検索クエリをベクトルに変換する変換ステップと、

前記検索クエリのベクトルと、前記検索対象文書のベクトルとに基づいて、前記検索クエリと前記検索対象文書との類似度を決定する決定ステップと、

10

前記フィードバック情報記憶部に前記検索クエリのベクトルと同一又は類似のベクトルを有する検索クエリに対応するフィードバック情報が存在する場合に、決定ステップで決定された前記類似度を、前記フィードバック情報の前記評価値情報に基づいて調整して調整後類似度を決定する調整ステップと、

前記調整後類似度に基づいて、前記検索クエリと類似する文書の検索結果を作成する検索結果作成ステップと、

前記検索結果を含む検索結果画面を所定の装置に表示させる表示制御ステップと、を備える文書検索方法。

【請求項 8】

検索クエリを受け付けて、検索クエリに基づいて、複数の検索対象文書の中から類似する文書を検索する文書検索装置を構成するコンピュータに実行させるための文書検索プログラムであって、

20

前記コンピュータを、

前記検索クエリを受け付ける受付部と、

前記受け付けた検索クエリをベクトルに変換するベクトル変換部と、

前記検索クエリのベクトルと、前記検索対象文書のベクトルとに基づいて、前記検索クエリと前記検索対象文書との類似度を決定する類似度決定部と、

過去に利用した検索クエリのベクトルと、前記検索クエリと同一又は類似の検索クエリにより検索された文書を特定する情報と、前記検索された文書に対する利用者の操作に基づく評価値情報とを含むフィードバック情報を記憶するフィードバック情報記憶部に、前記検索クエリのベクトルと同一又は類似のベクトルを有する検索クエリに対応するフィードバック情報が存在する場合に、前記類似度決定部により決定された前記類似度を、前記フィードバック情報の前記評価値情報に基づいて調整して調整後類似度を決定する調整処理部と、

30

前記調整後類似度に基づいて、前記検索クエリと類似する文書の検索結果を作成する検索結果作成部と、

前記検索結果を含む検索結果画面を所定の装置に表示させる表示制御部と、

して機能させる

文書検索プログラム。

40

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、検索クエリを受け付けて、検索クエリに基づいて、複数の検索対象文書の中から類似する文書を検索する文書検索装置等に関する。

【背景技術】

50

【0002】

機械学習・AI (Artificial Intelligence) 技術を用いた類似検索を行う検索システムにおいては、全文検索とは異なり、文書中の文章の内容の類似性を用いて検索を行うことができるので、検索語が検索対象の文書に含まれていなくても、関連する文書として検出することができる。

【0003】

しかしながら、このような機械学習・AI 技術を用いた検索システムであっても、利用者が求めているものと異なる検索結果となってしまう場合がある。

【0004】

例えば、AI を用いた技術として、SDI 検索の精度を高めるために、検索に用いる学習モデルをユーザが送信した教師データで構築する技術が知られている（例えば、特許文献1参照）。

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開2018-18167号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

検索システムにおいては、いかに利用者が求めている検索結果を高精度に返すことが要請されている。

【0007】

本発明は、上記事情に鑑みなされたものであり、その目的は、利用者の要求に合致する文書を容易かつ適切に検索できるようにする技術を提供することにある。

【課題を解決するための手段】

【0008】

上記目的を達成するため、一観点に係る文書検索装置は、検索クエリを受け付けて、検索クエリに基づいて、複数の検索対象文書の中から類似する文書を検索する文書検索装置であって、複数の検索対象文書の文章に対応するベクトルを記憶するベクトル情報記憶部と、過去に利用した検索クエリのベクトルと、検索クエリと同一又は類似の検索クエリにより検索された文書を特定する情報と、検索された文書に対する利用者の操作に基づく評価情報とを含むフィードバック情報を記憶するフィードバック情報記憶部と、検索クエリを受け付ける受付部と、受け付けた検索クエリをベクトルに変換するベクトル変換部と、検索クエリのベクトルと、検索対象文書のベクトルとに基づいて、検索クエリと検索対象文書との類似度を決定する類似度決定部と、フィードバック情報記憶部に検索クエリのベクトルと同一又は類似のベクトルを有する検索クエリに対応するフィードバック情報が存在する場合に、類似度決定部により決定された類似度を、フィードバック情報の評価情報に基づいて調整して調整後類似度を決定する調整処理部と、調整後類似度に基づいて、検索クエリと類似する文書の検索結果を作成する検索結果作成部と、検索結果を含む検索結果画面を所定の装置に表示させる表示制御部と、を備える。

【発明の効果】

【0009】

本発明によれば、利用者の要求に合致する文書を容易かつ適切に検索することができる。

【図面の簡単な説明】

【0010】

【図1】図1は、一実施形態に係る検索システムの全体構成図である。

【図2】図2は、一実施形態に係る検索システムの機能構成図である。

【図3】図3は、一実施形態に係る文書ベクトル情報の構成図である。

【図4】図4は、一実施形態に係るユーザ情報の構成図である。

10

20

30

40

50

【図5】図5は、一実施形態に係る操作ログ情報の構成図である。

【図6】図6は、一実施形態に係るフィードバック情報の構成図である。

【図7】図7は、一実施形態に係る検索処理画面を示す図である。

【図8】図8は、一実施形態に係る検索処理のフローチャートである。

【図9】図9は、一実施形態に係るフィードバック情報更新処理のフローチャートである。

。

【発明を実施するための形態】

【0011】

実施形態について、図面を参照して説明する。なお、以下に説明する実施形態は特許請求の範囲に係る発明を限定するものではなく、また実施形態の中で説明されている諸要素及びその組み合わせの全てが発明の解決手段に必須であるとは限らない。

10

【0012】

以下の説明では、各種情報をテーブル構造として説明するが、各種情報は、どのようなデータ構造で表現されていてもよい。

【0013】

図1は、一実施形態に係る検索システムの全体構成図である。

【0014】

検索システム1は、1以上の検索クライアント10と、文書検索装置の一例としての検索サーバ20とを備えている。検索クライアント10と、検索サーバ20とは、ネットワーク40を介して接続されている。ネットワーク40は、例えば、有線LAN(Local Area Network)や無線LANなどのネットワークである。

20

【0015】

検索クライアント10は、例えば、PC(Personal Computer)によって構成されている。検索クライアント10は、プロセッサの一例としてのCPU(Central Processing Unit)11と、メモリ12と、入出力デバイス13と、通信インターフェース(I/F)14と、記憶デバイス15と、を備えている。

【0016】

通信I/F14は、例えば、有線LANカードや無線LANカードなどのインターフェースであり、ネットワーク40を介して他の装置(例えば、検索サーバ20)と通信する。

30

【0017】

入出力デバイス13は、例えば、マウス、キーボード等のユーザによる操作入力を受け付ける入力装置と、ディスプレイ等の各種情報を表示出力する表示装置とを含む。

【0018】

CPU11は、メモリ12及び/又は記憶デバイス15に格納されているプログラムに従って各種処理を実行する。

【0019】

メモリ12は、例えば、RAM(Random Access Memory)であり、CPU11で実行されるプログラムや、必要な情報を記憶する。

【0020】

記憶デバイス15は、例えば、ハードディスクドライブやSSD(Solid State Drive)などであり、CPU11で実行されるプログラムや、CPU11に利用されるデータを記憶する。なお、記憶デバイス15に格納されているプログラムは、メモリ12に呼び出されてCPU11に実行される。本実施形態では、記憶デバイス15は、プログラムとして検索性プログラム151を記憶する。

40

【0021】

検索性プログラム151は、GUI(Graphical User Interface)として後述する検索処理画面50(図7参照)を入出力デバイス13に表示させ、利用者(ユーザ)から検索に用いる検索クエリ(検索語、又は検索文)を受け付けて、検索サーバ20に送信し、検索サーバ20から検索結果を受信して検索処理画面50(検索

50

結果画面の一例)に表示させるためのプログラムである。

【0022】

検索サーバ20は、例えば、PCによって構成されている。検索サーバ20は、プロセッサの一例としてのCPU21と、メモリ22と、記録媒体ドライブ23と、通信I/F24と、記憶デバイス25と、を備えている。

【0023】

通信I/F24は、例えば、有線LANカードや無線LANカードなどのインターフェースであり、ネットワーク40を介して他の装置(例えば、検索クライアント10)と通信する。

【0024】

記録媒体ドライブ23は、例えば、DVD-ROM等の不揮発性記録媒体41に格納されたプログラムやデータを読み込むドライブである。

【0025】

CPU21は、メモリ22及び/又は記憶デバイス25に格納されているプログラムに従って各種処理を実行する。

【0026】

メモリ22は、例えば、RAMであり、CPU21で実行されるプログラムや、必要な情報を記憶する。

【0027】

記憶デバイス25は、例えば、ハードディスクドライブやSSDなどであり、CPU21で実行されるプログラムや、CPU21に利用されるデータを記憶する。なお、記憶デバイス25に格納されているプログラムは、メモリ22に呼び出されてCPU21に実行される。本実施形態では、記憶デバイス25は、プログラムとして、ベクトル処理プログラム251と、検索処理プログラム252と、ベテラン判定プログラム253とを記憶し、データとして、文書データ254と、学習済みモデル255と、文書ベクトル情報256と、ユーザ情報257と、操作ログ情報258と、フィードバック情報259とを記憶する。ここで、ベクトル処理プログラム251、検索処理プログラム252、及びベテラン判定プログラム253が、文書検索プログラムの一例である。また、記憶デバイス25は、ベクトル情報記憶部、フィードバック情報記憶部、及び属性情報記憶部の一例である。

【0028】

ベクトル処理プログラム251は、学習済みモデル255を用いて、入力された文(文章)をベクトルに変換する(ベクトル化する)処理を実行する。例えば、ベクトル処理プログラム251は、検索クライアント10から送信された検索クエリをベクトル化する。また、ベクトル処理プログラム251は、学習用の文書データを教師データとして、モデルを学習することができる。ここで、モデルとしては、例えば、Doc2Vecのモデルを使用することができる。

【0029】

検索処理プログラム252は、検索クライアント10から受信した検索クエリに対応するベクトルと、検索対象の文書データのベクトルとの類似度を算出し、算出した類似度を、フィードバック情報259の情報に基づいて調整し、調整した類似度(調整後類似度)に基づいて、検索クエリに対応する検索結果とする文書を特定し、特定した文書を含む検索結果を検索クライアント10に送信する。

【0030】

ベテラン判定プログラム253は、操作ログ情報258に基づいて、所定の行動特性を有する利用者をベテランと判定し、その利用者が参照している文書が検索結果として選択されやすいようにフィードバック情報259を更新する処理を実行する。

【0031】

文書データ254は、検索対象となる複数の文書データである。学習済みモデル255は、ベクトル処理プログラム251によって使用される学習済みモデルである。この学習

10

20

30

40

50

済みモデル 255 は、他の装置において学習された学習済みモデルを取得してもよく、ベクトル処理プログラム 251 が教師データを用いることにより構築されてもよい。なお、教師データとしては、検索対象の文書に含まれている文書データを利用するようにしてもよく、全く別の文書データを利用するようにしてもよい。

【0032】

文書ベクトル情報 256 は、検索対象の文書データに対応するベクトルである。ユーザ情報 257 は、ユーザの属性に関する情報である。操作ログ情報 258 は、ユーザによる検索処理における操作に関するログの集合である。フィードバック情報 259 は、検索結果に対するユーザのフィードバックに関する情報である。

【0033】

図 2 は、一実施形態に係る検索システムの機能構成図である。

【0034】

検索クライアント 10 は、CPU 11 が検索用プログラム 151 を実行することにより提供される GUI 16 を有する。検索クライアント 10 のユーザは、入出力デバイス 13 の入力装置を操作することにより、GUI 16 に対応する検索処理画面 50 (図 7 参照) を介して、検索クエリや、操作を行うことができるとともに、検索処理画面 50 により検索結果を確認することができる。

【0035】

検索サーバ 20 は、ベクトル処理部 26 と、検索処理部 27 と、フィードバック反映部 28 と、ベテラン判定部 29 と、ログ登録部 30 とを有する。ここで、ベクトル処理部 26 は、受付部及びベクトル変換部の一例である。また、検索処理部 27 は、類似度決定部の一例である。また、フィードバック反映部 28 は、調整処理部、検索結果作成部、及び表示制御部の一例である。また、ベテラン判定部 29 は、操作反映評価値更新部、属性反映評価値更新部、及び行動特定反映評価値更新部の一例である。また、ログ登録部 30 は、操作ログ取得部の一例である。

【0036】

ベクトル処理部 26 は、CPU 21 がベクトル処理プログラム 251 を実行することにより構成される。ベクトル処理部 26 は、教師データを入力して、学習済みモデル 255 を構築する。また、ベクトル処理部 26 は、文書データ 254 の各文書をベクトル化して文書ベクトル情報 256 を作成する。ベクトル処理部 26 は、検索クライアント 10 から受信した検索クエリをベクトル化して、検索処理部 27 に渡す。

【0037】

検索処理部 27 と、フィードバック反映部 28 とは、CPU 21 が検索処理プログラム 252 を実行することにより構成される。検索処理部 27 は、検索クエリのベクトルと、文書ベクトル情報 256 の各文書のベクトルとを用いて、検索クエリと各文書との類似度を算出し、算出結果をフィードバック反映部 28 に渡す。フィードバック反映部 28 は、検索クエリと各文書との類似度を、フィードバック情報 259 に基づいて調整し、調整した類似度 (調整類似度) に基づいて、検索結果を決定して検索クライアント 10 に送信する。フィードバック反映部 28 は、検索結果の各文書に対して、文書を参照するための記述 55 (文書へのリンク) と、文書が役に立ったことを明示するための役立つボタン 56 と、文書が役に立たなかったことを明示するための役に立たないボタン 57 とを検索処理画面 50 に表示させる。

【0038】

ベテラン判定部 29 と、ログ登録部 30 とは、CPU 21 がベテラン判定プログラム 253 を実行することにより構成される。ベテラン判定部 29 は、操作ログ情報 258 と、ユーザ情報 257 と、フィードバック情報 259 とを用いて、ベテランである利用者を特定し、ベテランが参照した文書が検索結果として選択されやすくするためにフィードバック情報 259 を更新する処理を実行する。ログ登録部 30 は、検索クライアント 10 からの操作のログを操作ログ情報 258 に登録する。

【0039】

10

20

30

40

50

次に、文書ベクトル情報 2 5 6 の詳細な構成について説明する。

【 0 0 4 0 】

図 3 は、一実施形態に係る文書ベクトル情報の構成図である。

【 0 0 4 1 】

文書ベクトル情報 2 5 6 は、検索対象の文書のベクトルを管理する情報であり、検索対象の文書のそれぞれに対応するエントリを格納する。文書ベクトル情報 2 5 6 のエントリは、文書 ID 2 5 6 a と、文書ベクトル 2 5 6 b とのフィールドを格納する。文書 ID 2 5 6 a には、エントリに対応する文書の ID (識別情報) が格納される。文書ベクトル 2 5 6 b には、エントリに対応する文書の文章についてのベクトル (ベクトルの各成分の数値) が格納される。本実施形態では、文書ベクトル 2 5 6 b には、学習済みモデル 2 5 5 を使って文書データをベクトル化した場合におけるベクトルが格納されている。

10

【 0 0 4 2 】

次に、ユーザ情報 2 5 7 の詳細な構成について説明する。

【 0 0 4 3 】

図 4 は、一実施形態に係るユーザ情報の構成図である。

【 0 0 4 4 】

ユーザ情報 2 5 7 は、ユーザの属性を管理する情報であり、検索クライアント 1 0 を利用するユーザに対応するエントリを格納する。ユーザ情報 2 5 7 のエントリは、ユーザ ID 2 5 7 a と、職位 2 5 7 b と、褒賞・資格 2 5 7 c とのフィールドを含む。ユーザ ID 2 5 7 a には、エントリに対応するユーザの ID が格納される。職位 2 5 7 b には、エントリに対応するユーザの職位の名称が格納される。褒賞・資格 2 5 7 c には、エントリに対応するユーザに与えられた褒賞や、資格の名称が格納される。

20

【 0 0 4 5 】

次に、操作ログ情報 2 5 8 の詳細な構成について説明する。

【 0 0 4 6 】

図 5 は、一実施形態に係る操作ログ情報の構成図である。

【 0 0 4 7 】

操作ログ情報 2 5 8 は、検索クライアント 1 0 を利用する複数のユーザによる操作のログを管理する情報であり、各ユーザの各操作に対応するエントリ (ログ) を有する。操作ログ情報 2 5 8 のエントリは、ユーザ ID 2 5 8 a と、操作内容 2 5 8 b と、日時 2 5 8 c とのフィールドを含む。ユーザ ID 2 5 8 a には、エントリに対応するユーザの ID が格納される。操作内容 2 5 8 b には、エントリに対応する操作内容が格納される。例えば、検索クエリにより検索をした場合には、操作内容には、検索を開始したことを示す情報と、検索クエリの内容とが含まれる。また、検索結果の文書を参照した場合には、操作内容には、文書を参照したことを示す情報と、参照した文書の ID とが含まれる。日時 2 5 8 c には、エントリに対応する操作を実行した日時が格納される。

30

【 0 0 4 8 】

次に、フィードバック情報 2 5 9 の詳細な構成について説明する。

【 0 0 4 9 】

図 6 は、一実施形態に係るフィードバック情報の構成図である。

40

【 0 0 5 0 】

フィードバック情報 2 5 9 は、検索クエリ情報 2 6 0 と、検索結果情報 2 6 1 とを含む。検索クエリ情報 2 6 0 は、検索に使用した検索クエリを管理するエントリを有する。本実施形態では、ベクトル同士が類似している (所定の類似度よりも高い) 複数の検索クエリを一つのエントリとして管理している。これにより、フィードバック情報 2 5 9 として必要な記憶容量を、異なる検索クエリをそれぞれ別のエントリとして管理する場合よりも低減することができる。なお、記憶容量に問題がない場合には、異なる検索クエリをそれぞれ別のエントリとして管理してもよい。

【 0 0 5 1 】

検索クエリ情報 2 6 0 は、検索 ID 2 6 0 a と、検索クエリ 2 6 0 b と、検索ベクトル

50

260cとのフィールドを含む。検索ID260aは、エントリに対応する検索クエリによる検索を識別するID(検索ID)が格納される。検索クエリ260bには、エントリに対応する検索クエリが格納される。検索ベクトル260cには、エントリに対応する検索クエリのベクトルが格納される。

【0052】

検索結果情報261は、検索毎の検索結果として選択された文書毎のエントリを有する。検索結果情報261のエントリは、検索ID261aと、文書ID261bと、A評価値261cと、B評価値261dと、C評価値261eとのフィールドを含む。検索ID261aには、エントリに対応する文書が検索された検索に対応する検索IDが格納される。文書ID261bには、エントリに対応する文書のIDが格納される。A評価値261cには、エントリに対応する検索における検索結果の文書についての、役に立つことに対する評価値が格納される。本実施形態では、A評価値261cには、例えば、エントリに対応する検索における検索結果の文書に対して、検索処理画面50における役に立つボタン56が押下された回数の総計が格納されている。B評価値261dには、エントリに対応する検索における検索結果の文書についての、役に立たないことに対する評価値が格納される。本実施形態では、B評価値261dには、例えば、エントリに対応する検索における検索結果の文書に対して、検索処理画面50における役に立たないボタンが押下された回数の総計が格納されている。C評価値261eには、エントリに対応する検索における検索結果の文書について、文書が参照されたことに対する評価値が格納される。本実施形態では、C評価値261eには、例えば、エントリに対応する検索における検索結果の文書について、検索処理画面50における文書を参照するためのリンクがクリックされた回数の総計が格納されている。

10

20

【0053】

フィードバック情報259によると、或る検索クエリと同一又は類似する検索クエリによって検索された文書について、ユーザが役に立つとした評価値、ユーザが役に立たないとした評価値、ユーザにより参照されたことに対する評価値を適切に把握することができる。

【0054】

次に、検索サーバ20によって検索クライアント10に表示される検索処理画面50について説明する。

30

【0055】

図7は、一実施形態に係る検索処理画面を示す図である。

【0056】

検索処理画面50は、例えば、検索クライアント10において、検索用プログラム151が実行された場合に表示される画面であり、検索クエリを入力するための検索クエリ入力領域51と、検索クエリ入力領域51に入力した検索クエリにより検索を実行させるための実行ボタン52と、検索結果の一覧が表示される検索結果一覧領域53とを含む。

【0057】

検索結果一覧領域53には、検索サーバ20による検索によって得られた各文書に関する情報が表示される文書毎領域54が表示される。文書毎領域54には、文書の名前と、文書の文章(全部またはその一部)が表示されるとともに、その文書を参照するためのリンクが張られた文書リンク領域55と、対応する文書が役に立ったことを検索サーバ20に知らせるために押下する役に立つボタン56と、対応する文書が役に立たなかったことを検索サーバ20に知らせるために押下する役に立たないボタン57と、が表示される。

40

【0058】

この検索処理画面50によると、或る検索クエリにより検索された検索結果の各文書に対しての、ユーザの操作に基づく評価の情報を適切に検索サーバ20に通知することができる。

【0059】

次に、検索システム1における処理動作について説明する。

50

【 0 0 6 0 】

図 8 は、一実施形態に係る検索処理のフローチャートである。

【 0 0 6 1 】

この検索処理は、検索サーバ 2 0 において実行される処理である。

【 0 0 6 2 】

検索クライアント 1 0 において、検索処理画面 5 0 の検索クエリ入力領域 5 1 に検索クエリ（単語、文、文章でもよい）が入力され、実行ボタン 5 2 が押下されて、検索クエリ入力領域 5 1 に入力された検索クエリが検索サーバ 2 0 に送信されると、ログ登録部 3 0 は、検索クエリを送信したユーザのユーザ ID と、検索開始との情報と、検索クエリと、受信した日時とを含む新たなエントリ（操作ログ）を操作ログ情報 2 5 8 に記録する（ステップ S 1 1 ）。

10

【 0 0 6 3 】

次いで、ベクトル処理部 2 6 は、学習済みモデル 2 5 5 を用いて受信した検索クエリをベクトル化することによりベクトル（検索ベクトル）を算出し、検索ベクトルを検索処理部 2 7 に通知する（ステップ S 1 2 ）。

【 0 0 6 4 】

次いで、検索処理部 2 7 は、文書ベクトル情報 2 5 6 を参照して、検索ベクトルと類似するベクトルを有する文書を検索し、検索結果の文書をフィードバック反映部 2 8 に通知する（ステップ S 1 3 ）。具体的には、検索処理部 2 7 は、検索ベクトルと、各文書の文章のベクトル（文書ベクトル）との類似度（具体的には、ベクトル同士のコサイン類似度）を算出し、類似度が所定の値以上となる文書を検索結果の文書として選択し、フィードバック反映部 2 8 に通知する。

20

【 0 0 6 5 】

次いで、フィードバック反映部 2 8 は、フィードバック情報 2 5 9 を参照し、フィードバック情報 2 5 9 に、検索ベクトルと類似する検索ベクトルに対応するエントリが存在する場合には、検索結果の文書の類似度を調整し、調整後の類似度（調整後類似度）により検索結果とする文書を決定するとともに、検索結果における文書の配置の順序を決定する（ステップ S 1 4 ）。

【 0 0 6 6 】

ここで、文書の類似度の調整方法の一例について説明する。

30

【 0 0 6 7 】

フィードバック反映部 2 8 は、以下の式（ 1 ）に従って調整後類似度を算出する。

調整後類似度 = 調整前の類似度 + f_1 （ A 評価値 ） - f_2 （ B 評価値 ） + f_3 （ C 評価値 ）・・・（ 1 ）

ここで、 f_1 、 f_2 、 f_3 は、関数である。例えば、各関数を次のようにしてもよい、 $f_1(x) = 0.1x$ 、 $f_2(x) = 0.1(x)$ 、 $f_3(x) = 0.05(x)$ 。

【 0 0 6 8 】

例えば、ユーザの各々が独立して単純な意思決定を行っていても、それが積み重なったとき、その集合全体が知性を持ってふるまっているように見えることがある。なお、このようにして生まれる知性は集合知と呼ばれる。 f_1 、 f_2 、 f_3 を、上記したような関数とすると、間違っただけに基づく評価値同士が打ち消しあって、全体数が多くなると全体として正しい結果を示すこととなる。なお、本実施形態では、文書の参照を行う際のクリック数による評価値（ C 評価値 ）によっても調整後類似度が上昇するようになっているので、役立つボタン 5 6 や役に立たないボタン 5 7 をユーザが明示的に押下しない場合であっても、参照回数が多く役に立っている可能性が高い文書の類似度を上昇させることができ、この文書を検索結果のより上位に表示させるようにすることができる。なお、 f_1 、 f_2 、 f_3 は、例えば、値域が $-1 \sim 1$ となる双曲線関数を用いて、 $f_1(x) = \tanh(x)$ 、 $f_2(x) = \tanh(x)$ 、 $f_3(x) = \tanh(0.5x)$ としてもよい。このようにすると、評価値自体が大きな値をとるような場合であっても、類似度の調整量にキャップを持たせることができる。

40

50

【0069】

次いで、フィードバック反映部28は、決定した検索結果を検索クライアント10の検索処理画面50に表示させる(ステップS15)。

【0070】

これにより、検索クライアント10においては、検索処理画面50の検索結果一覧領域53には、調整後類似度に応じた順番で検索された文書の情報が表示されることとなる。このため、ユーザは、複数のユーザによる操作に基づく評価が反映された順番で、検索結果の文書を把握することができる。これにより、ユーザは効率よく所望する文書に到達する可能性が高くなる。

【0071】

この後、検索クライアント10において、検索処理画面50に対するユーザの操作があった場合には、操作内容がログ登録部30に通知され、ログ登録部30は、操作内容のログを操作ログ情報258に記録する(ステップS16)。

【0072】

次いで、操作内容が検索結果一覧領域53に表示されている文書の参照(文書参照)である場合には、ベテラン判定部29は、フィードバック情報259中に、参照対象の文書を検索した検索クエリの検索ベクトルと類似のベクトルの情報が格納されているか否かを判定し(ステップS17)、類似のベクトルの情報が格納されている場合(ステップS17:あり)には、フィードバック情報259を更新する(ステップS18)。具体的には、ベテラン判定部29は、この類似ベクトルの検索IDに対応し、且つ参照対象の文書の文書IDに対応する、検索結果情報261のエントリのC評価値261eの値に1を加算する。この後、処理は、ステップS16の直前に進められる。

【0073】

一方、類似のベクトルの情報が格納されていない場合(ステップS17:なし)には、フィードバック情報259に新たなエントリを追加する(ステップS19)。具体的には、ベテラン判定部29は、検索クエリ情報260に、検索クエリに対応するエントリを追加するとともに、検索結果情報261に、この類似ベクトルの検索IDに対応し、且つ参照対象の文書の文書IDに対応するエントリを追加し、このエントリのC評価値261eの値を1に設定する。この後、処理はステップS16の直前に進められる。

【0074】

一方、操作内容が検索結果一覧領域53に表示されている文書に対するボタン(役に立つボタン56又は役に立たないボタン57)の押下である場合には、ベテラン判定部29は、フィードバック情報259中に、参照対象の文書を検索した検索クエリの検索ベクトルと類似のベクトルの情報が格納されているか否かを判定し(ステップS20)、類似のベクトルの情報が格納されている場合(ステップS20:あり)には、フィードバック情報259を更新する(ステップS21)。具体的には、ベテラン判定部29は、この類似ベクトルの検索IDに対応し、且つ参照対象の文書の文書IDに対応する、検索結果情報261のエントリの押下されたボタンに対応するフィールド(役に立つボタン56の場合はA評価値261c、役に立たないボタン57の場合にB評価値261d)の値に1を加算する。この後、処理は、ステップS16の直前に進められる。

【0075】

一方、類似のベクトルの情報が格納されていない場合(ステップS20:なし)には、フィードバック情報259に新たなエントリを追加する(ステップS22)。具体的には、ベテラン判定部29は、検索クエリ情報260に、検索クエリに対応するエントリを追加するとともに、検索結果情報261に、この類似ベクトルの検索IDに対応し、且つ参照対象の文書の文書IDに対応するエントリを追加し、このエントリの押下されたボタンに対応するフィールドの値を1に設定する。この後、処理はステップS16の直前に進められる。

【0076】

この検索処理によると、過去のユーザによる評価が反映された検索結果が表示されるよ

10

20

30

40

50

うになるとともに、その検索結果に対するユーザの操作に応じた評価がフィードバック情報 259 に反映されることとなる。

【0077】

次に、フィードバック情報更新処理について説明する。

【0078】

図9は、一実施形態に係るフィードバック情報更新処理のフローチャートである。

【0079】

フィードバック情報更新処理は、任意の時点、例えば、所定の時間ごとに実行される。

【0080】

ベテラン判定部 29 は、操作ログ情報 258 を参照し、検索結果からファイルを開いた（参照した）回数が所定数よりも少ない検索クエリと、その検索クエリによる検索結果において参照された文書を抽出し、メモリ 22 に抽出情報 262 として記憶する（ステップ S31）。ここで、検索結果からファイルを開いた回数は、一のユーザにより検索開始があった後、このユーザにより次の検索開始があるまでの間における文書参照のエントリの数により特定することができる。このステップによると、検索クエリに対して適切であると考えられる文書を抽出することができる。

10

【0081】

次いで、ベテラン判定部 29 は、ユーザ情報 257 と、操作ログ情報 258 とを参照して、所定の職位以上のユーザ（ベテランの候補）による検索クエリと、その検索クエリによる検索結果において参照された文書を抽出し、メモリ 22 に抽出情報 262 として記憶する（ステップ S32）。このステップによると、比較的経験を有しており検索に対してベテランと考えられるユーザにより参照された文書を抽出することができる。

20

【0082】

次いで、ベテラン判定部 29 は、操作ログ情報 258 を参照し、所定の行動特性のユーザ（例えば、1回の検索あたりの参照した文章の数が少ない（所定数以下）ユーザ）をベテランと判定し、このベテランのユーザによって参照された文書を抽出情報 262 から特定する（ステップ S33）。ここで、ユーザについての1回の検索あたりの参照した文章の数については、同一ユーザによる各検索開始後から次の検索開始までの間の参照された文書の数を、検索回数で除算することにより把握することができる。

【0083】

次いで、ベテラン判定部 29 は、ベテランのユーザによって参照された文書として抽出された文書が、類似の検索クエリによる以降の検索においてランクアップするようにフィードバック情報 259 を更新する（ステップ S34）。具体的には、ベテラン判定部 29 は、特定された文書の検索に使用した検索クエリに対応し、且つ特定された文書に対応する、フィードバック情報 259 のエントリにおける、類似度の増加に寄与する評価値（A 評価値 261c 又は C 評価値 261e の少なくとも一方）を加算する。このステップによって、検索クエリの検索結果としてより適していると考えられる、ベテランにより参照された文書が検索結果として選択され、検索結果のより上位に配置されるように調整することができる。

30

【0084】

なお、本発明は、上述の実施形態に限定されるものではなく、本発明の趣旨を逸脱しない範囲で、適宜変形して実施することが可能である。

40

【0085】

例えば、上記実施形態では、文書を参照するクリック、役立つボタンの押下、又は、役に立たないボタンの1回の押下について、すべてのユーザについて、同じ評価値を加算するようにしていたが、本発明はこれに限られず、ユーザによって、評価値を異ならせるようにしてもよい。

【0086】

また、上記実施形態において、ユーザ情報として、ユーザの得意分野等の情報を含ませるようにしておき、検索クエリとユーザの得意分野との類似度に応じて、このユーザの操

50

作に対する評価値を変えるようにしてもよい。例えば、検索クエリとユーザの得意分野との類似度が高い場合には、このユーザの1回の操作に対する評価値を高くするようにしてもよい。

【0087】

また、上記実施形態において、ユーザ情報の職位を用いて、ベテランの候補のユーザを決定し、そのユーザが検索時に参照した文書を抽出するようにしていたが、例えば、褒賞や資格などを条件としてベテランの候補を決定してもよい。

【0088】

また、上記実施形態では、検索結果からファイルを開いた回数が少ない場合の文書（参照回数に基づく文書）と、一定以上の職位のユーザにより検索された文書（職位に基づく文書）とを抽出し、それら文書の中からランクアップする文書を決定するようにしていたが、本発明はこれに限られず、例えば、参照回数に基づく文書と、職位に基づく文書のいずれか一方の中からランクアップする文書を特定するようにしてもよい。

10

【0089】

また、上記実施形態において、CPUがプログラムを実行することにより行っていた処理の一部又は全部を、ハードウェア回路で行うようにしてもよい。また、上記実施形態におけるプログラムは、プログラムソースからインストールされてよい。プログラムソースは、プログラム配布サーバ又は記録媒体（例えば、不揮発性の可搬型の記録媒体）であってもよい。

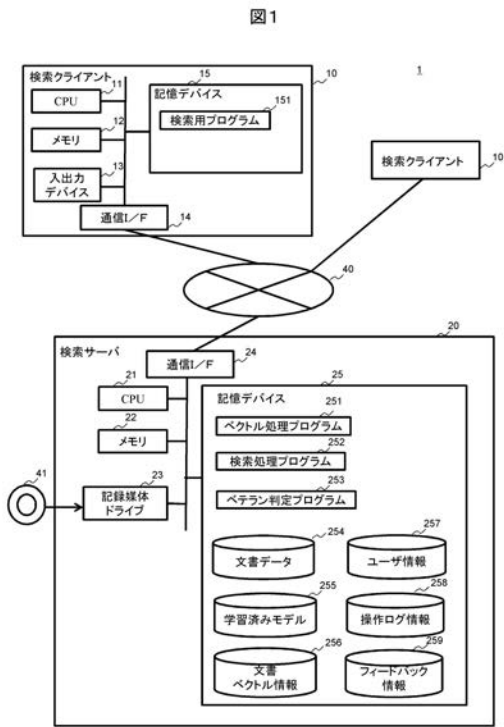
20

【符号の説明】

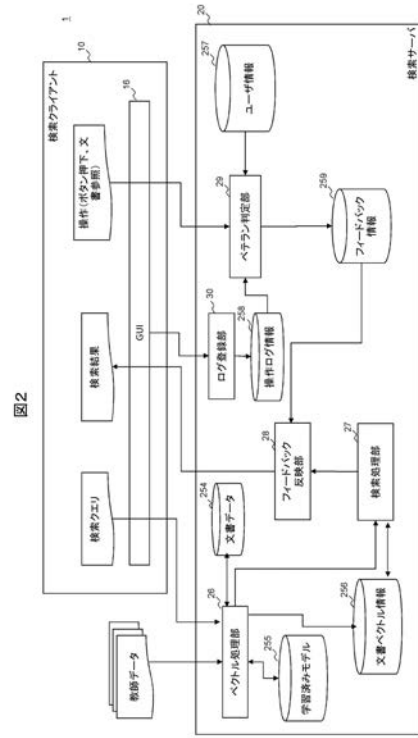
【0090】

1 ... 検索システム、10 ... 検索クライアント、11 ... CPU、12 ... メモリ、13 ... 入出力デバイス、14 ... 通信I/F、15 ... 記憶デバイス、20 ... 検索サーバ、21 ... CPU、22 ... メモリ、23 ... 記録媒体ドライブ、24 ... 通信I/F、25 ... 記憶デバイス

【 図 1 】



【 図 2 】



【 図 3 】

図3

256a	256b	256
文書ID	文書ベクトル	
D001	999,999,999,...	
D002	999,999,999,...	
⋮	⋮	

【 図 6 】

図6

260a	260b	260c	260
検索ID	検索クエリ	検索ベクトル	
1	⋯⋯	999,999,999,...	
2	⋯⋯	999,999,999,...	
⋮	⋮	⋮	

261a	261b	261c	261d	261e	261
検索ID	文書ID	A評価値	B評価値	C評価値	
1	D0055	20	5	100	
1	D0105	30	40	60	
⋮	⋮	⋮	⋮	⋮	

【 図 4 】

図4

257a	257b	257c	257
ユーザID	職位	褒賞・資格	
U001	部長		
U002	課長	開発技術賞	
⋮	⋮	⋮	

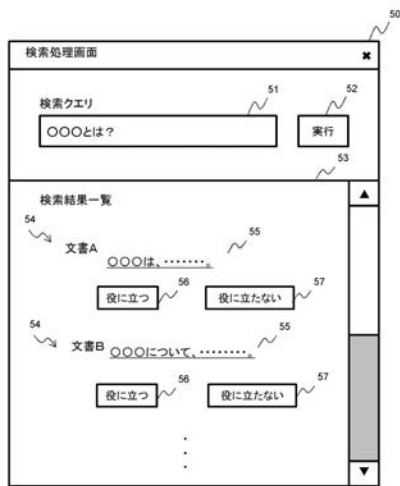
【 図 5 】

図5

258a	258b	258c	258
ユーザID	操作内容	日時	
U005	検索開始、検索クエリ=⋯⋯	2018/3/19:00:00	
U005	文書参照、文書=D003	2018/3/19:02:10	
⋮	⋮	⋮	

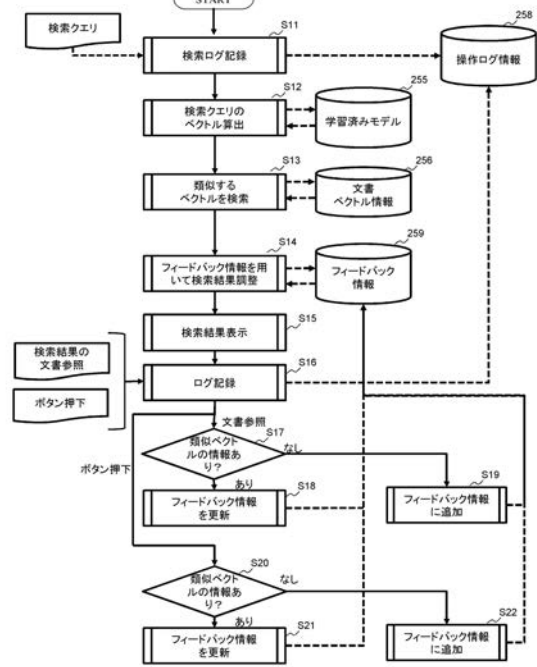
【 図 7 】

図7



【 図 8 】

図8



【 図 9 】

図9

