

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7626924号
(P7626924)

(45)発行日 令和7年2月5日(2025.2.5)

(24)登録日 令和7年1月28日(2025.1.28)

(51)国際特許分類

F I

G 0 6 F 16/38 (2019.01)

G 0 6 F 16/31 (2019.01)

G 0 6 F 16/38

G 0 6 F 16/31

請求項の数 6 (全19頁)

(21)出願番号	特願2020-218462(P2020-218462)	(73)特許権者	390002761 キャノンマーケティングジャパン株式会社 東京都港区港南2丁目16番6号
(22)出願日	令和2年12月28日(2020.12.28)	(73)特許権者	592135203 キャノンITソリューションズ株式会社 東京都港区港南2丁目16番6号
(65)公開番号	特開2022-103684(P2022-103684 A)	(74)代理人	100189751 弁理士 木村 友輔
(43)公開日	令和4年7月8日(2022.7.8)	(72)発明者	田中 靖大 東京都港区港南2丁目16番6号 キャ ノンITソリューションズ株式会社内
審査請求日	令和5年12月27日(2023.12.27)	審査官	甲斐 哲雄

最終頁に続く

(54)【発明の名称】 情報処理システム、情報処理方法、プログラム

(57)【特許請求の範囲】

【請求項1】

文書に、当該文書が検索された際の検索クエリを補完情報として対応付けて記憶し、前記検索クエリが存在しない文書については、当該文書の内容に基づく情報を補完情報として対応付けて記憶する記憶手段と、
ユーザから受け付けた検索クエリを用いて、前記記憶手段に記憶された文書の内容と、当該文書に対応付けて記憶された補完情報とを検索対象として検索を行う検索手段と、
を備えることを特徴とする情報処理システム。

【請求項2】

前記記憶手段は、前記検索クエリが存在しない文書について、当該文書における複数の項目に係る内容を結合した情報を補完情報として、当該文書に対応付けて記憶することを特徴とする請求項1に記載の情報処理システム。

【請求項3】

前記記憶手段は、前記検索クエリが存在しない文書について、当該文書から抽出した特徴語を補完情報として、当該文書に対応付けて記憶することを特徴とする請求項1または2に記載の情報処理システム。

【請求項4】

前記記憶手段は、前記文書が検索され閲覧された場合に、当該文書が検索された際の検索クエリを補完情報として対応付けて記憶することを特徴とする請求項1乃至3のいずれか1項に記載の情報処理システム。

10

20

【請求項 5】

情報処理システムの記憶手段が、文書に、当該文書が検索された際の検索クエリを補完情報として対応付けて記憶し、前記検索クエリが存在しない文書については、当該文書の内容に基づく情報を補完情報として対応付けて記憶する記憶工程と、

前記情報処理システムの検索手段が、ユーザから受け付けた検索クエリを用いて、前記記憶手段に記憶された文書の内容と、当該文書に対応付けて記憶された補完情報とを検索対象として検索を行う検索工程と、

を備えることを特徴とする情報処理方法。

【請求項 6】

コンピュータを、請求項 1 乃至 4 のいずれか 1 項に記載の各手段として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理システム、情報処理方法、プログラムに関する。

【背景技術】

【0002】

企業において、様々な電子データが作成され蓄積が進み、蓄積した電子データの有効活用が望まれるようになってきている。しかしながら、蓄積されたデータは膨大であり、有用なデータ、欲しいデータを見つけ出すことに多くの時間が費やされることが大きな課題となっている。

【0003】

近年、エンタープライズサーチに代表されるような全文検索システムがデータを効率よく見つけ出すという課題に対する解決手段の一つとして大きな関心が向けられるようになってきている。

【0004】

全文検索においては、一般的に検索結果を提示する際の順位を決めるためにスコアリングが行われている。スコアリングの手法については $tf \cdot idf$ や BM25 などに基づいた様々な提案がなされているが、ユーザの使用語彙や検索意図・観点によって適したスコアリング手法は異なる。

【0005】

ユーザが検索時に自分に適したスコアリング手法を選択することは現実的ではない。また検索システムを利用するユーザ全体に対して、最もバランスのよいスコアリング手法を事前に選択することも困難である。

【0006】

ランキング学習は、クリックログ（閲覧履歴）または、検索クエリに対するアノテーション（検索者による検索結果に対する明示的評価）を、検索クエリに対する適合文書と定義し、学習データとして利用する。

【0007】

そして、検索クエリに対するスコアリング手法によるスコアを 1 つの素性（ランキング素性）と考え、複数のスコアリング手法のスコアの組み合わせを特徴と捉えて、学習データにおける適合文書が上位になるように学習し、ユーザのクリックログやアノテーションにおける適合文書を検索結果の上位に提示することが可能となる。

【0008】

特許文献 1 には、学習データから過去に使用した検索クエリまたは検索クエリから抽出した言語的特徴を抽出し、登録文書に対する新たなフィールド（以降、「検索クエリフィールド」と呼ぶ）に保存し、ランキング素性として利用することで、過去の検索クエリに対しクリック（またはアノテート）された文書を上位に提示する精度の向上を可能とする技術が記載されている。

【先行技術文献】

10

20

30

40

50

【特許文献】

【0009】

【文献】特開2020-71678号公報

【発明の開示】

【発明が解決しようとする課題】

【0010】

しかし、検索対象として登録されている全ての文書に対して、学習データ（クリックログまたはアノテーション）が存在することは稀である。

【0011】

特許文献1においては、検索クエリフィールドに値が存在しない登録文書が存在することになり、検索クエリフィールドに対するスコアが0となってしまう。逆に、検索クエリフィールドに値があれば、僅かな言語的特徴の一致があればスコアを得ることになり、関係の薄い文書でも値のない文書よりスコアが高くなり、上位になってしまう可能性が著しく高くなってしまう。

【0012】

クリックログやアノテーションは検索結果から選択されるため、上位に提示された文書ほど学習データが集まりやすい傾向がある。結果として学習初期に上位に出現した文書が、その後も上位に出現し続けることになり、学習データの偏りが助長され、使用語彙や検索意図の変化に対応できない状況が発生するという課題がある。

【0013】

そのため、適切な検索結果を得るためには、登録文書に対する学習データの有無が、検索結果における順位に与える影響を少なくすることが望まれる。

【0014】

そこで、本発明は、適切な検索結果を得られる仕組みを提供することを目的とする。

【課題を解決するための手段】

【0015】

本発明の情報処理システムは、文書に、当該文書を検索した際の検索クエリを補完情報として対応付けて記憶し、前記検索クエリが存在しない文書については、当該文書の構成要素を補完情報として対応付けて記憶する記憶手段と、ユーザから受け付けた検索クエリを用いて、前記記憶手段に記憶された文書と、当該文書に対応付けて記憶された補完情報とを検索対象として検索を行う検索手段と、を備えることを特徴とする。

【発明の効果】

【0016】

本発明によれば、適切な検索結果を得ることが可能となる。

【図面の簡単な説明】

【0017】

【図1】本発明の実施形態における、全文検索装置のシステム構成の一例を示す図である。

【図2】本発明の実施形態における、全文検索装置、クライアント端末のハードウェア構成の一例を示すブロック図である。

【図3】本発明の実施形態における、全文検索装置の機能構成の一例を示す図である。

【図4】本発明の実施形態における、全文検索装置における学習処理の一例を示すフローチャートである。

【図5】本発明の実施形態における、学習データ保存領域の構成の一例を示す図である。

【図6】本発明の実施形態における、検索対象保存領域の構成の一例を示す図である。

【図7】本発明の実施形態における、検索クエリフィールド追加の処理の一例を示すフローチャートである。

【図8】本発明の実施形態における、学習データ生成処理の一例を示すフローチャートである。

【図9】本発明の実施形態における、検索クエリに対する特徴ベクトル生成処理の一例を示すフローチャートである。

10

20

30

40

50

【図 1 0】本発明の実施形態における、閲覧履歴の一例を示す図である。

【図 1 1】本発明の実施形態における、文書保存テーブルの一例を示す図である。

【図 1 2】本発明の実施形態における、検索クエリを集約した閲覧履歴の一例を示す図である。

【図 1 3】本発明の実施形態における、検索クエリフィールド補完後の文書保存テーブルの一例を示す図である。

【図 1 4】本発明の実施形態における、検索クエリに対する特徴ベクトルの一例を示す図である。

【図 1 5】本発明の実施形態における、ランキング学習に対する学習データの一例を示す図である。

10

【図 1 6】本発明の実施形態における、ランキング学習の結果取得した学習モデルの一例を示す図である。

【図 1 7】本発明の実施形態における、特徴ベクトル生成処理の概略を示す図である。

【図 1 8】本発明の実施形態における、検索処理の一例を示すフローチャートである。

【図 1 9】本発明の実施形態における、検索クエリに対する特徴ベクトルの一例を示す図である。

【図 2 0】本発明の実施形態における、検索結果の一例を示す図である。

【発明を実施するための形態】

【0018】

以下、図面を参照して、本発明の実施形態を詳細に説明する。

20

【0019】

図 1 は、本発明における情報処理システムの構成の一例を示す図である。

【0020】

図 1 は、本発明の実施形態における全文検索装置のシステム構成の一例を示す図である。

【0021】

全文検索装置 100 は、クライアント端末 110 が LAN 120 を介して接続される構成となっている。

【0022】

全文検索装置 100 は、一般的な全文検索システムの機能を有し、事前に登録されている文書に対し、検索要求に対して適合する文書の一覧を検索結果として返し、検索結果のうち任意の文書の内容を確認する機能を提供する。

30

【0023】

さらに、全文検索装置 100 は、検索者による閲覧の履歴または検索結果の評価を取得して検索クエリと関連付けて保存する。

【0024】

クライアント端末 110 は、ブラウザを介して全文検索装置 100 の機能をユーザに提供する。

【0025】

図 2 は、本発明の情報処理装置のハードウェア構成の一例を示すブロック図である。

【0026】

40

図 2 に示すように、情報処理装置は、システムバス 200 を介して CPU (Central Processing Unit) 201、ROM (Read Only Memory) 202、RAM (Random Access Memory) 203、記憶装置 204、入力コントローラ 205、音声コントローラ 206、ビデオコントローラ 207、メモリコントローラ 208、よび通信 I/F コントローラ 209 が接続される。

【0027】

CPU 201 は、システムバス 200 に接続される各デバイスやコントローラを統括的に制御する。

【0028】

ROM 202 あるいは外部メモリ 213 は、CPU 201 が実行する制御プログラムで

50

あるBIOS (Basic Input/Output System) やOS (Operating System) や、本情報処理方法を実現するためのコンピュータ読み取り実行可能なプログラムおよび必要な各種データ (データテーブルを含む) を保持している。

【0029】

RAM 203は、CPU 201の主メモリ、ワークエリア等として機能する。CPU 201は、処理の実行に際して必要なプログラム等をROM 202あるいは外部メモリ 213からRAM 203にロードし、ロードしたプログラムを実行することで各種動作を実現する。

【0030】

入力コントローラ 205は、キーボード 210や不図示のマウス等のポインティングデバイス等の入力装置からの入力を制御する。入力装置がタッチパネルの場合、ユーザがタッチパネルに表示されたアイコンやカーソルやボタンに合わせて押下 (指等でタッチ) することにより、各種の指示を行うことができることとする。

10

【0031】

また、タッチパネルは、マルチタッチスクリーンなどの、複数の指でタッチされた位置を検出することが可能なタッチパネルであってもよい。

【0032】

ビデオコントローラ 207は、ディスプレイ 212などの外部出力装置への表示を制御する。ディスプレイは本体と一体になったノート型パソコンのディスプレイも含まれるものとする。なお、外部出力装置はディスプレイに限ったものはなく、例えばプロジェクタ

20

【0033】

なおビデオコントローラ 207は、表示制御を行うためのビデオメモリ (VRAM) を制御することが可能で、ビデオメモリ領域としてRAM 203の一部を利用することもできるし、別途専用のビデオメモリを設けることも可能である。

【0034】

メモリコントローラ 208は、外部メモリ 213へのアクセスを制御する。外部メモリとしては、ブートプログラム、各種アプリケーション、フォントデータ、ユーザファイル、編集ファイル、および各種データ等を記憶する外部記憶装置 (ハードディスク)、フレキシブルディスク (FD)、或いはPCMCIAカードスロットにアダプタを介して接続されるコンパクトフラッシュ (登録商標) メモリ等を利用可能である。

30

【0035】

通信I/Fコントローラ 209は、ネットワークを介して外部機器と接続・通信するものであり、ネットワークでの通信制御処理を実行する。例えば、TCP/IPを用いた通信やISDNなどの電話回線、および携帯電話の4G回線、5G回線等を用いた通信が可能である。

【0036】

尚、CPU 201は、例えばRAM 203内の表示情報用領域へアウトラインフォントの展開 (ラスタライズ) 処理を実行することにより、ディスプレイ 212上での表示を可能としている。また、CPU 201は、ディスプレイ 212上の不図示のマウスカーソル等でのユーザ指示を可能とする。

40

【0037】

次に図3を用いて、本実施例における情報処理システムの機能を説明する。

【0038】

クライアント端末 110はWebブラウザ 310を介して全文検索装置 100の機能をユーザに提供するものである。Webブラウザ 310は、ユーザから検索クエリの入力を受け付ける画面や、全文検索装置 100により検索された結果を表示するための画面を表示する機能を備える。

【0039】

50

全文検索装置 100 の学習処理部 301 は、学習データ保存領域 302 や検索対象保存領域 303 に記憶されたデータを用いて、学習モデルを生成する機能を備える。

【0040】

学習データ保存領域 302 は、ユーザにより閲覧された文書と、当該閲覧をされた際の検索クエリとを対応付けた閲覧履歴が記録された閲覧履歴保存テーブルや、学習処理部 301 における学習結果である学習モデル等を記憶している。

【0041】

検索対象保存領域 303 は、検索対象となる文書のインデックスが登録された文書保存テーブルを記憶している。

【0042】

検索処理部 304 は、学習データ保存領域 302 に記憶された閲覧履歴の検索クエリに対する特徴ベクトルの生成や、ユーザから Web ブラウザ 310 を介して受け付けた検索クエリに基づき、検索対象保存領域 303 に記憶された文書保存テーブルを検索し、検索結果を Web ブラウザ 310 に返す機能を備える。

【0043】

(学習処理)

次に図 4 のフローチャートを用いて、本発明の実施形態における全文検索装置が実行する学習処理について説明する。

【0044】

図 4 は、全文検索装置 100 の CPU 201 が所定の制御プログラムを読み出して実行する処理であり、所定の期間ごとに実行される過去の閲覧履歴を学習する処理を示すフローチャートである。

【0045】

ステップ S401 では、学習処理部 301 は、図 5 に示す学習データ保存領域 302 の閲覧履歴保存テーブル 501 から過去の閲覧履歴を取得する。

【0046】

ステップ S402 では、学習処理部 301 は、図 6 に示す検索対象保存領域 303 の文書保存テーブル 601 に対して、ステップ S401 で取得した閲覧履歴を検索クエリフィールドに追加する。詳細については後述する。

【0047】

文書保存テーブル 601 は、一般的な全文検索システムと同等の管理構造を有し、文書の各部を格納する複数のフィールドと、フィールドを高速に全文検索するためのインデックスを有する。

また各フィールドは、検索語や登録文書の単語の識別方法（形態素解析、n-gram など）やスコアリングの方法（tf・idf や BM25 など）に応じた複数のインデックスを持つことを可能とする。

【0048】

ステップ S403 では、学習処理部 301 は、ステップ S401 で取得した閲覧履歴から学習データを生成する。詳細については後述する。

【0049】

ステップ S404 では、学習処理部 301 は、ステップ S403 で生成した学習データに基づき、ランキング学習を行い、学習結果として学習モデルを取得する。

ステップ S405 では、学習処理部 301 は、ステップ S404 の学習結果である学習モデルを学習データ保存領域 302 における学習モデル保存テーブル 502 に保存する。

【0050】

(検索クエリフィールド追加)

次に図 7 のフローチャートを用いて、前述したステップ S402 の検索クエリフィールド追加処理について説明する。

【0051】

ステップ S701 では、学習処理部 301 は、文書保存テーブル 601 に検索クエリフ

10

20

30

40

50

フィールドを追加する。

【0052】

ステップS702では、学習処理部301は、処理対象の閲覧履歴を関連文書の文書IDごとに検索クエリを1つのテキストに集約する。

【0053】

ステップS703では、学習処理部301は、集約した関連文書ごとに、ステップS705までの繰り返し処理を開始する。

【0054】

ステップS704では、学習処理部301は、処理対象の集約した関連文書に対し、検索クエリを関連文書の文書IDと関連付けて、文書保存テーブル601の検索クエリフィールドに追加する。

10

【0055】

ステップS705では、学習処理部301は、処理対象となる集約した関連文書がまだあれば、ステップS703の繰り返し処理を実施する。処理対象となる集約した関連文書がなければ、ステップS706に処理を移す。

【0056】

ステップS706では、学習処理部301は、文書保存テーブル601の登録文書ごとに、ステップS708までの繰り返し処理を開始する。

【0057】

ステップS707では、学習処理部301は、処理対象の登録文書に対する検索クエリフィールドが空の場合、ステップS708に処理を移す。登録文書に対する検索クエリフィールドが空でなければ、ステップS709に処理を移す。

20

【0058】

ステップS708では、学習処理部301は、処理対象の登録文書の別フィールドの値に基づいた補完文字列を検索クエリフィールドに代入する。本実施の例においては、「タイトル」および「本文」を結合した文字列を用いる。

【0059】

閲覧履歴に代入する値は、特定のフィールドだけに限定してもよいし、文字列の長さが近いなどの統計的特徴により、登録文書ごとに選択してもよいし、更には、他の文書との比較において、特徴的な単語（特徴語）または文を抽出して用いてもよい。

30

【0060】

ステップS709では、学習処理部301は、処理対象となる登録文書がまだあれば、ステップS706からの処理を実施する。処理対象となる登録文書がなければ、ステップS710に処理を移す。

【0061】

ステップS710では、学習処理部301は、検索クエリフィールドに対して、インデックスを作成する。本実施の例においては、「形態素解析」と「n-gram」の2つの単語の抽出方法に対して、BM25によるスコアリングを行うものとする。

【0062】

このように、過去に文書が閲覧された際の検索クエリを当該文書と対応付けて記憶し、当該検索クエリについても検索対象として以降の検索処理を実行することで、ユーザの求める文書を適切に検索することが可能となる。また、過去に文書が閲覧された際の検索クエリが存在しない文書については、当該文書に関する要素を検索クエリとみなして対応付けて記憶することで、過去に閲覧履歴がない文書などについても、適切な検索スコアを算出でき、適切なランキングで検索結果を表示することが可能となる。

40

（学習データ生成処理）

次に図8のフローチャートを用いて、前述したステップS403の学習データ生成処理について説明する。

【0063】

ステップS801では、学習処理部301は、一時領域にある学習データを初期化する。

50

【 0 0 6 4 】

ステップ S 8 0 2 では、学習処理部 3 0 1 は、ステップ S 4 0 1 で取得した閲覧履歴ごとに、ステップ S 8 0 6 までの繰り返し処理を開始する。

【 0 0 6 5 】

ステップ S 8 0 3 では、学習処理部 3 0 1 は、処理対象の閲覧履歴の検索クエリに対して、検索処理部 3 0 4 に特徴ベクトルの生成を要求し、取得する。検索処理部 3 0 4 における特徴ベクトルの生成については後述する。

【 0 0 6 6 】

ステップ S 8 0 4 では、学習処理部 3 0 1 は、ステップ S 8 0 3 で取得した特徴ベクトルが、処理対象の閲覧履歴における文書を含んでいる場合、処理をステップ S 8 0 5 に移す。含んでいない場合、処理をステップ S 8 0 6 に移す。

10

【 0 0 6 7 】

ステップ S 8 0 5 では、学習処理部 3 0 1 は、ステップ S 8 0 3 で取得した検索クエリに対する特徴ベクトルの中で、処理対象の閲覧履歴における文書に対応する特徴ベクトルを、文書と関連付けて、一時領域の学習データに追加する。

【 0 0 6 8 】

ステップ S 8 0 6 では、学習処理部 3 0 1 は、処理対象となる閲覧履歴がまだあれば、ステップ S 8 0 2 からの繰り返し処理を実施する。処理対象となる閲覧履歴がなければ処理を終了する。

【 0 0 6 9 】

20

(検索クエリに対する特徴ベクトル生成処理)

次に図 9 のフローチャートを用いて、前述したステップ S 8 0 3 における検索処理部 3 0 4 の検索クエリに対する特徴ベクトル生成処理について説明する。

【 0 0 7 0 】

ステップ S 9 0 1 では、検索処理部 3 0 4 は、処理対象である検索クエリに対する特徴ベクトルを一時領域に作成し初期化する。

【 0 0 7 1 】

ステップ S 9 0 2 では、検索処理部 3 0 4 は、検索対象保存領域 3 0 3 の文書保存テーブル 6 0 1 において、検索クエリフィールドを含めて登録文書に対して指定された全文検索対象のフィールドごとに、ステップ S 9 1 1 までの繰り返し処理を開始する。

30

【 0 0 7 2 】

ステップ S 9 0 3 では、検索処理部 3 0 4 は、処理対象のフィールドに対し付与されたインデックスごとに、ステップ S 9 1 0 までの繰り返し処理を開始する。

【 0 0 7 3 】

ステップ S 9 0 4 では、検索処理部 3 0 4 は、処理対象の検索クエリを用いて、処理対象のインデックスを検索し、検索結果として文書とスコアを取得する。

【 0 0 7 4 】

ステップ S 9 0 5 では、検索処理部 3 0 4 は、ステップ S 9 0 4 で取得した検索結果の文書ごとに、ステップ S 9 0 9 までの繰り返し処理を開始する。

【 0 0 7 5 】

40

ステップ S 9 0 6 では、検索処理部 3 0 4 は、文書に対する特徴ベクトルが、一時領域に存在すれば、処理をステップ S 9 0 8 に移す。文書に対する特徴ベクトルが一時領域になければ、処理をステップ S 9 0 7 に移す。

【 0 0 7 6 】

ステップ S 9 0 7 では、検索処理部 3 0 4 は、初期化した特徴ベクトルを一時領域に文書と関連付けて追加する。文書に対する特徴ベクトルは、各検索フィールドの各インデックスに対するスコアを要素として 0 で初期化される。

【 0 0 7 7 】

ステップ S 9 0 8 では、検索処理部 3 0 4 は、文書に対する特徴ベクトルにおける、処理対象のフィールドとインデックスに対する要素として、検索結果における文書のスコア

50

を代入する。

【 0 0 7 8 】

ステップ S 9 0 9 では、検索処理部 3 0 4 は、処理対象となる文書が検索結果にまだあれば、ステップ S 9 0 5 からの繰り返し処理を実施する。処理対象の文書がなければ、ステップ S 9 1 0 に処理を移す。

【 0 0 7 9 】

ステップ S 9 1 0 では、検索処理部 3 0 4 は、処理対象となるインデックスがまだあれば、ステップ S 9 0 3 からの繰り返し処理を実施する。処理対象のインデックスがなければ、ステップ S 9 1 1 に処理を移す。

【 0 0 8 0 】

ステップ S 9 1 1 では、検索処理部 3 0 4 は、処理対象となるフィールドがまだあれば、ステップ S 9 0 2 からの繰り返し処理を実施する。処理対象のフィールドがなければ、処理を終了する。図 1 7 は、特徴ベクトルの生成処理を概念的に示した図である。

【 0 0 8 1 】

(学習処理の具体例)

次に、学習処理の具体例として、図 1 0 に示す閲覧履歴保存テーブル 5 0 1 および図 1 1 に示す文書保存テーブル 6 0 1 に対して、図 4 に示す学習処理が実施された場合について説明する。

【 0 0 8 2 】

ステップ S 4 0 1 では、学習処理部 3 0 1 は、図 1 0 に示す閲覧履歴保存テーブル 5 0 1 から過去の閲覧履歴を取得する。

【 0 0 8 3 】

図 1 0 に示すとおり、閲覧履歴保存テーブル 5 0 1 には、閲覧履歴として、閲覧された文書を一意に特定するための文書 ID と、当該文書を閲覧した際の検索クエリとが対応付けて保存されている。例えば、図 1 0 の 1 0 0 1 のレコードは、「スクリプトを実行するには」という検索クエリによって検索処理が実行され、文書 ID 「 d 3 」の文書が閲覧されたことを示している。

【 0 0 8 4 】

ステップ S 4 0 2 では、学習処理部 3 0 1 は、図 1 1 に示す文書保存テーブル 6 0 1 に対して、ステップ S 4 0 1 で取得した閲覧履歴を追加する図 7 のフローチャートで示した検索クエリフィールドの追加処理を開始する。

ステップ S 7 0 1 では、学習処理部 3 0 1 は、文書保存テーブル 6 0 1 に検索クエリフィールド 1 1 0 1 を追加する。

【 0 0 8 5 】

ステップ S 7 0 2 では、学習処理部 3 0 1 は、図 1 0 の閲覧履歴テーブル 5 0 1 に対して、図 1 2 に示す関連文書の文書 ID で集約された閲覧履歴 1 2 0 1 を取得する。

【 0 0 8 6 】

図 1 2 に示す通り、ステップ S 7 0 2 の処理によって、同一の文書 ID と対応付けられた検索クエリが集約される。

【 0 0 8 7 】

ステップ S 7 0 3 では、学習処理部 3 0 1 は、関連文書に対して、ステップ S 7 0 5 までの繰り返し処理を開始する。

【 0 0 8 8 】

ステップ S 7 0 4 では、学習処理部 3 0 1 は、集約された検索クエリ 1 2 0 2 を、文書保存テーブル 6 0 1 における同じ文書 ID (d 1) の登録文書 1 1 0 2 の検索クエリフィールドに追加する。

【 0 0 8 9 】

以下同様に、学習処理部 3 0 1 は、集約した関連文書全てに対して、ステップ S 7 0 3 からステップ S 7 0 5 の繰り返し処理を実施し、文書保存テーブル 6 0 1 は図 1 1 に示すような状態となる。

10

20

30

40

50

【0090】

図11に示すとおり、文書保存テーブル601には、文書を一意に特定するための文書ID、文書のタイトル、文書の内容である本文、集約された検索クエリから構成される。

ステップS706では、学習処理部301は、文書保存テーブル601の登録文書1102に対して、ステップS708までの繰り返し処理を開始する。

【0091】

ステップS707では、学習処理部301は、登録文書1102に対する検索クエリフィールドが空ではないので、ステップS709に処理を移す。

【0092】

ステップS709では、学習処理部301は、処理対象となる登録文書1103がまだあるので、ステップS706からの処理を実施する。

10

【0093】

ステップS706では、学習処理部301は、文書保存テーブル601の登録文書1103に対して、ステップS708までの繰り返し処理を開始する。

【0094】

ステップS707では、学習処理部301は、処理対象の登録文書1103に対する検索クエリフィールドが空なので、ステップS708に処理を移す。

【0095】

ステップS708では、学習処理部301は、処理対象の登録文書1103の「タイトル」および「本文」フィールドの値を結合して登録文書1103の検索クエリフィールドに代入する。

20

【0096】

以下同様に、学習処理部301は、集約した関連文書全てに対して、ステップS706からステップS709の繰り返し処理を実施し、文書保存テーブル601は図13に示すような状態となる。

【0097】

図13に示す通り、ステップS705までの処理によって作成された文書保存テーブル(図11)において検索クエリ欄が空であったレコードについて、検索クエリの内容(本実施例においてはタイトル及び本文を結合した値)が追加されている。

【0098】

ステップS710では、学習処理部301は、検索クエリフィールド1101に対して、インデックスを作成する。

30

【0099】

ステップS403では、学習処理部301は、閲覧履歴テーブルの501の閲覧履歴に対し、図8に示す学習データ生成処理を実施する。

【0100】

ステップS801では、学習処理部301は、一時領域にある学習データを初期化する。

【0101】

ステップS802では、学習処理部301は、閲覧履歴テーブルの501の閲覧履歴1001の検索クエリに対して、ステップS806までの繰り返し処理を開始する。

40

【0102】

ステップS803では、学習処理部301は、処理対象の閲覧履歴1001の検索クエリに対する図9に示す特徴ベクトル生成処理を検索処理部304に要求して、図14に示すような特徴ベクトル1401を取得する。

【0103】

特徴ベクトルの生成については、ここでは図17に概略図を示すにとどめ具体的な説明は省略する。

【0104】

ステップS804では、学習処理部301は、ステップS803で生成した特徴ベクトル1401が処理対象の閲覧履歴における文書(ID=d3)を含んでいるので、処理を

50

ステップ S 8 0 5 に移す。

【 0 1 0 5 】

ステップ S 8 0 5 では、学習処理部 3 0 1 は、処理対象の閲覧履歴 1 0 0 1 における文書と対応する特徴ベクトル 1 4 0 2 を関連付けて、一時領域の学習データ 1 5 0 1 に追加する (1 5 0 2)。

【 0 1 0 6 】

ステップ S 8 0 6 では、学習処理部 3 0 1 は、処理対象となる閲覧履歴 1 0 0 2 がまだあるので、ステップ S 8 0 3 からの繰り返し処理を実施する。

【 0 1 0 7 】

以下、同様の処理を繰り返し、学習データを生成する。

【 0 1 0 8 】

ステップ S 4 0 4 では、学習処理部 3 0 1 は、ステップ S 4 0 3 で生成した学習データに基づき、ランキング学習を行い、図 1 6 に示すような学習モデルを獲得する。

【 0 1 0 9 】

ステップ S 4 0 5 では、学習処理部 3 0 1 は、ステップ S 4 0 4 の学習結果である学習モデルを学習データ保存領域 3 0 2 における学習モデル保存テーブル 5 0 2 に保存する。

【 0 1 1 0 】

(検索処理)

次に図 1 8 のフローチャートを用いて、ステップ S 4 0 4 で獲得した学習モデルを用いた検索処理について説明する。

【 0 1 1 1 】

図 1 8 は、全文検索装置 1 0 0 の CPU 2 0 1 が所定の制御プログラムを読み出して実行する処理であり、クライアント端末 1 1 0 のブラウザ 3 1 0 を介したユーザからの検索要求に対して登録文書の中から適合する文書を検索する処理を示すフローチャートである。

【 0 1 1 2 】

ステップ S 1 8 0 1 では、検索処理部 3 0 4 は、ユーザからの検索要求として、検索クエリを取得する。

【 0 1 1 3 】

ステップ S 1 8 0 2 では、検索処理部 3 0 4 は、図 9 に示した処理により、ステップ S 1 8 0 1 で取得した検索クエリに対する特徴ベクトルを生成する。

【 0 1 1 4 】

ステップ S 1 8 0 3 では、検索処理部 3 0 4 は、取得した検索クエリに対する特徴ベクトルに含まれる特徴ベクトルごとに、ステップ S 1 8 0 5 までの繰り返し処理を開始する。

【 0 1 1 5 】

ステップ S 1 8 0 4 では、検索処理部 3 0 4 は、処理対象の特徴ベクトルに対して、学習データ保存領域 3 0 2 における学習モデル保存テーブル 5 0 2 における学習モデルを参照して、処理対象の特徴ベクトルから関連付けられた文書に対するスコアを算出する。

【 0 1 1 6 】

ステップ S 1 8 0 5 では、検索処理部 3 0 4 は、処理対象となる特徴ベクトルがまだあれば、ステップ S 1 8 0 3 からの繰り返し処理を実施する。処理対象となる特徴ベクトルがなければステップ S 1 8 0 6 に処理を移す。

【 0 1 1 7 】

ステップ S 1 8 0 6 では、検索処理部 3 0 4 は、S 1 8 0 4 で求めたスコアに応じて特徴ベクトルに関連付けられた文書を並び替えて、検索結果とする。

【 0 1 1 8 】

(検索処理具体例)

次に、検索処理の具体例として、閲覧履歴に含まれない文書 d 2 を対象とした検索クエリ「スクリプトの編集について」に対して、図 1 6 に示した学習モデルを用い、図 1 8 に示す検索処理が実施した場合について説明する。

【 0 1 1 9 】

10

20

30

40

50

ステップ S 1 8 0 1 では、検索処理部 3 0 4 は、ユーザからの検索要求として、検索クエリ「スクリプトを編集について」を取得する。

【 0 1 2 0 】

ステップ S 1 8 0 2 では、検索処理部 3 0 4 は、図 9 に示した処理により、ステップ S 1 8 0 1 で取得した検索クエリ「スクリプトの編集について」に対する特徴ベクトルとして、図 1 9 に示す特徴ベクトル 1 9 0 1 を生成する。

【 0 1 2 1 】

ステップ S 1 8 0 3 では、検索処理部 3 0 4 は、取得した検索クエリ「スクリプトの編集について」に対する特徴ベクトル 1 9 0 1 に含まれる特徴ベクトル 1 9 0 2 に対して、ステップ S 1 8 0 5 までの繰り返し処理を開始する。

【 0 1 2 2 】

ステップ S 1 8 0 4 では、検索処理部 3 0 4 は、処理対象の特徴ベクトル 1 9 0 2 に対して、図 1 6 に示す学習モデルを参照して、特徴ベクトル 1 9 0 2 に関連付けられた文書 d 2 に対するスコアとして 1 6 7 . 7 0 5 を算出する。

【 0 1 2 3 】

ステップ S 1 8 0 5 では、検索処理部 3 0 4 は、処理対象となる特徴ベクトル 1 9 0 3 がまだあるので、ステップ S 1 8 0 3 からの繰り返し処理を実施する。

【 0 1 2 4 】

以下同様の処理を繰り返して、図 2 0 に示す計算結果 2 0 0 1 を取得する。

【 0 1 2 5 】

ステップ S 1 8 0 6 では、検索処理部 3 0 4 は、S 1 8 0 4 で求めたスコアに応じて特徴ベクトルを並び替えて、検索結果とする。

【 0 1 2 6 】

文書 d 2 は閲覧履歴に含まれていない。しかしながら検索クエリフィールドに対する補完が行われているので、特徴ベクトルにおける検索クエリに対する値 (2 0 0 4 、 2 0 0 5) は 0 とならず、文書 d 2 のスコアは文書 d 1 および文書 d 7 より上位となる。

【 0 1 2 7 】

このように、検索クエリフィールドを補完することで、閲覧履歴に含まれない文書に対しても、適切なスコアを得ることができ、表示順位における悪影響を抑えることが可能となる。

【 0 1 2 8 】

以上説明した通り、本願発明においては、過去に文書が閲覧された際の検索クエリを当該文書と対応付けて記憶し、当該検索クエリについても検索対象として以降の検索処理を実行することで、ユーザの求める文書を適切に検索することが可能となる。また、過去に文書が閲覧された際の検索クエリが存在しない文書については、当該文書に関する要素を検索クエリとみなして対応付けて記憶することで、過去に閲覧履歴がない文書などについても、適切な検索スコアを算出でき、適切なランキングで検索結果を表示することが可能となる。

本発明は、例えば、システム、装置、方法、プログラムもしくは記録媒体等としての実施態様をとることが可能である。具体的には、複数の機器から構成されるシステムに適用しても良いし、また、一つの機器からなる装置に適用しても良い。

【 0 1 2 9 】

また、本発明におけるプログラムは、図 4 、図 7 ~ 図 9 、図 1 8 に示すフローチャートの処理方法をコンピュータが実行可能なプログラムであり、本発明の記憶媒体は図 4 、図 7 ~ 図 9 、図 1 8 の処理方法をコンピュータが実行可能なプログラムが記憶されている。なお、本発明におけるプログラムは図 4 、図 7 ~ 図 9 、図 1 8 の各装置の処理方法ごとのプログラムであってもよい。

【 0 1 3 0 】

以上のように、前述した実施形態の機能を実現するプログラムを記録した記録媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ (または C P

10

20

30

40

50

UやMPU)が記録媒体に格納されたプログラムを読み出し、実行することによっても本発明の目的が達成されることは言うまでもない。

【0131】

この場合、記録媒体から読み出されたプログラム自体が本発明の新規な機能を実現することになり、そのプログラムを記録した記録媒体は本発明を構成することになる。

【0132】

プログラムを供給するための記録媒体としては、例えば、フレキシブルディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、DVD-ROM、磁気テープ、不揮発性のメモリカード、ROM、EEPROM、シリコンディスク等を用いることが出来る。

【0133】

また、コンピュータが読み出したプログラムを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムの指示に基づき、コンピュータ上で稼働しているOS(オペレーティングシステム)等が実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0134】

さらに、記録媒体から読み出されたプログラムが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPU等が実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0135】

また、本発明は、複数の機器から構成されるシステムに適用しても、ひとつの機器から成る装置に適用しても良い。また、本発明は、システムあるいは装置にプログラムを供給することによって達成される場合にも適用できることは言うまでもない。この場合、本発明を達成するためのプログラムを格納した記録媒体を該システムあるいは装置に読み出すことによって、そのシステムあるいは装置が、本発明の効果を享受することが可能となる。

【0136】

さらに、本発明を達成するためのプログラムをネットワーク上のサーバ、データベース等から通信プログラムによりダウンロードして読み出すことによって、そのシステムあるいは装置が、本発明の効果を享受することが可能となる。なお、上述した各実施形態およびその変形例を組み合わせた構成も全て本発明に含まれるものである。

【符号の説明】

【0137】

- 100 全文検索装置
- 110 クライアント端末
- 120 LAN

10

20

30

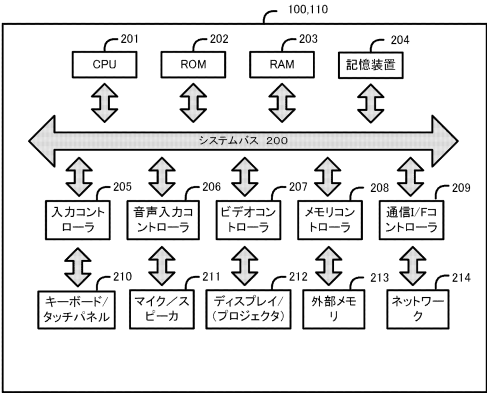
40

50

【図面】
【図 1】



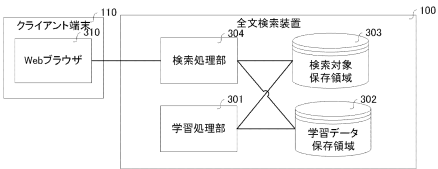
【図 2】



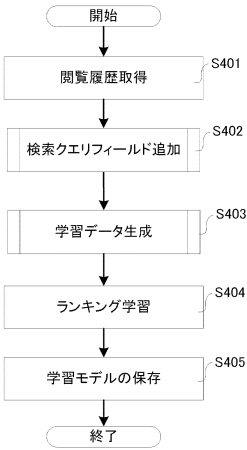
10

20

【図 3】



【図 4】

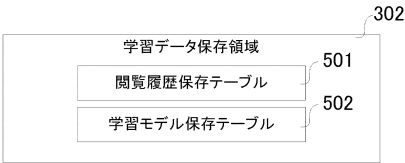


30

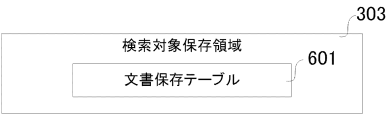
40

50

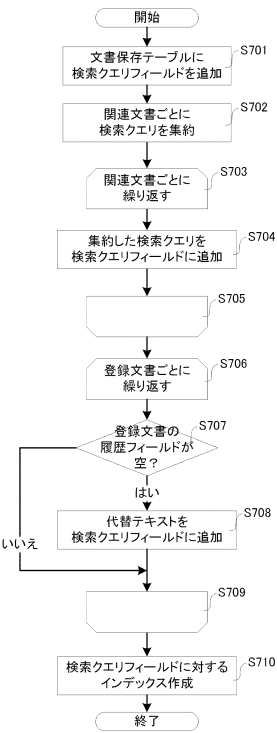
【図 5】



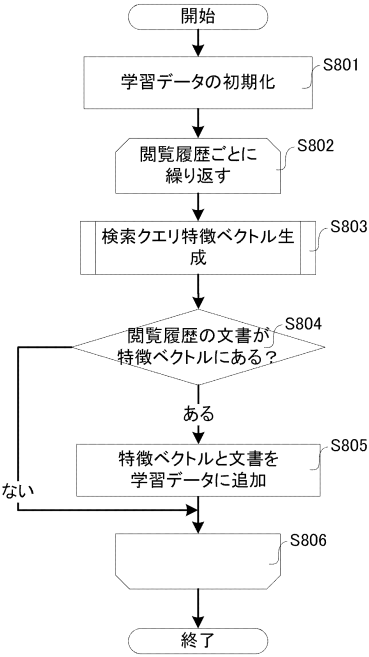
【図 6】



【図 7】



【図 8】



10

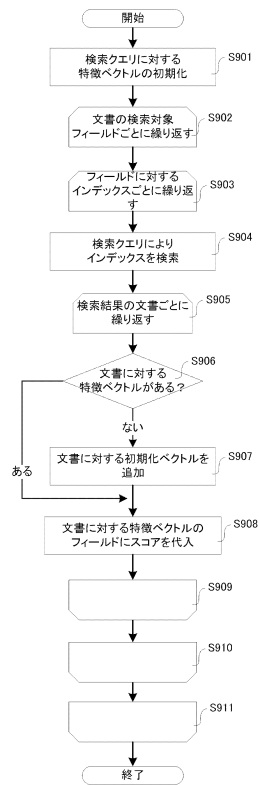
20

30

40

50

【図 9】



【図 10】

文書ID	検索クエリ
d3	スクリプトを実行するには
d3	スクリプトの実行方法について
d1	画面の編集ができない
d3	スクリプトの実行方法について
d1	画面エディットの方法
	:

【図 11】

文書ID	タイトル	本文	検索クエリ
d1	画面デザイン	画面の編集は…	画面の編集ができない 画面エディットの方法
d2	スクリプト編集	編集画面において…	
d3	スクリプト実行	指定したスクリプトは…	スクリプトを実行するには スクリプトの実行方法について
d4	実行エラー	ボタンを押す…	
:	:	:	:

【図 12】

文書ID	検索クエリ
d1	画面の編集ができない 画面エディットの方法
d3	スクリプトを実行するには スクリプトの実行方法について
	:

10

20

30

40

50

【図 1 3】

文書ID	タイトル	本文	検索クエリ
d1	画面デザイン	画面の編集は…	画面の編集ができない 画面エディットの方法
d2	スクリプト編集	編集画面において…	スクリプト編集 編集画面において…
d3	スクリプト実行	指定したスクリプトは…	スクリプトを実行するには スクリプトの実行方法について
d4	実行エラー	ボタンを押す…	実行エラー ボタンを押す…
⋮	⋮	⋮	⋮

【図 1 4】

タイトル形態素解析(1)	タイトルn-gram(2)	本文形態素解析(3)	本文n-gram(4)	検索クエリ形態素解析(3)	検索クエリn-gram(4)	文書ID
5.275	8.303	8.988	1.081	5.247	1.907	d7
5.122	4.979	5.846	2.229	7.612	3.056	d3
4.867	4.723	1.115	5.410	5.751	0.057	d8
2.951	4.762	4.129	0.466	2.556	3.860	d2
⋮	⋮	⋮	⋮	⋮	⋮	⋮

10

20

【図 1 5】

タイトル形態素解析(1)	タイトルn-gram(2)	本文形態素解析(3)	本文n-gram(4)	検索クエリ形態素解析(3)	検索クエリn-gram(4)	文書ID
5.122	4.979	5.846	2.229	7.612	3.056	d3
⋮	⋮	⋮	⋮	⋮	⋮	⋮

【図 1 6】

$$\sum_{i=\text{特徴}} \omega_i \times f_i$$

f_i : i 番目の特徴の特徴量

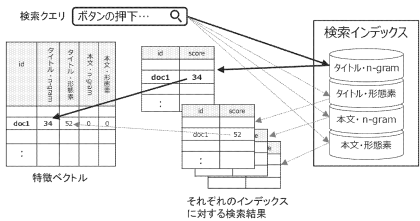
$\omega 1$	$\omega 2$	$\omega 3$	$\omega 4$	$\omega 5$	$\omega 6$
0.634	7.915	2.585	2.604	9.928	10.750

30

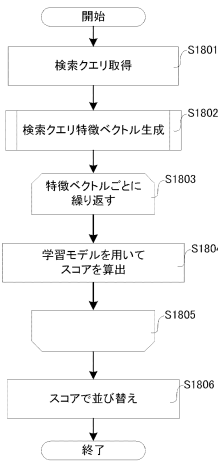
40

50

【図 17】



【図 18】



10

20

【図 19】

文書ID						
検索クエリ n-gram (1)	検索クエリ n-gram (2)	本文形態素解析 (3)	本文 n-gram (4)	検索クエリ形態素解析 (3)	検索クエリ n-gram (4)	
5.275	8.303	8.988	1.081	5.247	1.907	d2
5.122	4.979	5.846	2.229	7.612	3.056	d3
4.867	4.723	1.115	5.410	5.751	0.057	d1
2.951	4.762	4.129	0.466	2.556	3.860	d7
⋮	⋮	⋮	⋮	⋮	⋮	⋮

【図 20】

スコア						文書ID
検索クエリ n-gram (1) × w1	検索クエリ n-gram (2) × w2	本文形態素解析 (3) × w3	本文 n-gram (4) × w4	検索クエリ形態素解析 (3) × w5	検索クエリ n-gram (4) × w6	
3.347	55.720	23.234	2.814	52.091	30.499	d2
3.250	39.410	15.112	5.803	75.570	32.851	d3
3.088	37.384	12.882	14.085	27.094	10.613	d1
1.872	37.692	10.673	1.213	25.375	21.493	d7
⋮	⋮	⋮	⋮	⋮	⋮	⋮

30

40

50

フロントページの続き

- (56)参考文献 特開平 0 9 - 1 9 8 3 9 5 (J P , A)
特開 2 0 1 3 - 0 3 3 4 8 8 (J P , A)
特開 2 0 0 4 - 1 4 5 7 0 6 (J P , A)
- (58)調査した分野 (Int.Cl. , D B 名)
G 0 6 F 1 6 / 0 0 - 1 6 / 9 5 8