

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)公開番号
特開2024-123763
(P2024-123763A)

(43)公開日 令和6年9月12日(2024.9.12)

(51)国際特許分類
G 0 6 F 16/31 (2019.01)

F I
G 0 6 F 16/31

テーマコード (参考)
5 B 1 7 5

審査請求 未請求 請求項の数 6 O L (全16頁)

(21)出願番号	特願2023-31419(P2023-31419)	(71)出願人	515277757 株式会社A I S a m u r a i 東京都千代田区大手町一丁目6番1号 大手町ビル4階
(22)出願日	令和5年3月1日(2023.3.1)	(72)発明者	淵 武志 東京都千代田区大手町一丁目6番1号 大手町ビル4階 株式会社A I S a m u r a i 内
		Fターム (参考)	5B175 DA01 KA08

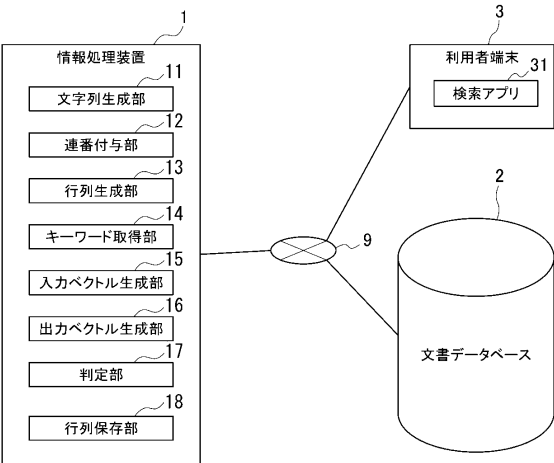
(54)【発明の名称】 情報処理装置、情報処理方法および情報処理プログラム

(57)【要約】 (修正有)

【課題】 検索の事前準備のための時間を短縮する情報処理装置、情報処理方法および情報処理プログラムを提供する。

【解決手段】 情報処理装置は、文書に含まれる文字をスキャンして出現する文字が重複しない文字列を生成する文字列生成部と、文書が有する文章に含まれる文字が生成された文字列に含まれるか否かを示す数字列を文章毎に並べた行列を生成する行列生成部と、キーワードを取得するキーワード取得部と、取得したキーワードに含まれる文字が生成された文字列に含まれるときは要素を正值とする、文字列の大きさを有する入力ベクトルを生成する入力ベクトル生成部と、生成された入力ベクトルと生成された行列との積を演算して出力ベクトルを生成する出力ベクトル生成部と、生成した出力ベクトルの要素の正值の数を判定する判定部と、を備える。

【選択図】 図6



【特許請求の範囲】

【請求項 1】

検索対象のキーワードに基づきデータベースに記憶された複数の被検索対象の文書を検索する情報処理装置であって、

前記文書に含まれる文字をスキャンして出現する文字が重複しない文字列を生成する文字列生成部と、

前記文書が有する文章に含まれる文字が前記文字列生成部において生成された前記文字列に含まれるか否かを示す数字列を前記文章毎に並べた行列を生成する行列生成部と、

前記キーワードを取得するキーワード取得部と、

前記キーワード取得部において取得された前記キーワードに含まれる文字が前記文字列生成部において生成された前記文字列に含まれるときは要素を真値とする、前記文字列の大きさを有する入力ベクトルを生成する入力ベクトル生成部と、

前記入力ベクトル生成部で生成された前記入力ベクトルと前記行列生成部において生成された前記行列との積を演算して出力ベクトルを生成する出力ベクトル生成部と、

前記出力ベクトル生成部において生成された前記出力ベクトルの要素の真値の数を判定する判定部と

を備える情報処理装置。

【請求項 2】

前記文書に含まれる文字に対して文章毎に連番を付与する連番付与部をさらに備え、

前記行列生成部は、前記行列を、前記連番付与部において付与された前記連番を用いた前記数字列を前記文章毎に並べて生成する、請求項 1 に記載の情報処理装置。

【請求項 3】

前記判定部は、前記真値の数が前記キーワードの文字数であって、かつ、前記真値の順序が前記行列生成部において生成された前記行列における要素の順である場合に、前記文書が前記キーワードにおいてヒットしたと判定する、請求項 2 に記載の情報処理装置。

【請求項 4】

前記判定部は、前記真値の数が前記キーワードの文字数である場合、前記文書の全文を前記キーワードにおける再検索の対象として判定する、請求項 1 または 2 に記載の情報処理装置。

【請求項 5】

コンピュータが、

検索対象のキーワードに基づきデータベースに記憶された複数の被検索対象の文書を検索する情報処理方法であって、

前記文書に含まれる文字をスキャンして出現する文字が重複しない文字列を生成する文字列生成ステップと、

前記文書が有する文章に含まれる文字が前記文字列生成ステップにおいて生成された前記文字列に含まれるか否かを示す数字列を前記文章毎に並べた行列を生成する行列生成ステップと、

前記キーワードを取得するキーワード取得ステップと、

前記キーワード取得ステップにおいて取得された前記キーワードに含まれる文字が前記文字列生成ステップにおいて生成された前記文字列に含まれるときは要素を真値とする、前記文字列の大きさを有する入力ベクトルを生成する入力ベクトル生成ステップと、

前記入力ベクトル生成ステップで生成された前記入力ベクトルと前記行列生成ステップにおいて生成された前記行列との積を演算して出力ベクトルを生成する出力ベクトル生成ステップと、

前記出力ベクトル生成ステップにおいて生成された前記出力ベクトルの要素の真値の数を判定する判定ステップと

を実行する情報処理方法。

【請求項 6】

コンピュータに、

検索対象のキーワードに基づきデータベースに記憶された複数の被検索対象の文書を検索させる情報処理プログラムであって、

前記文書に含まれる文字をスキャンして出現する文字が重複しない文字列を生成する文字列生成機能と、

前記文書が有する文章に含まれる文字が前記文字列生成機能において生成された前記文字列に含まれるか否かを示す数字列を前記文章毎に並べた行列を生成する行列生成機能と、

前記キーワードを取得するキーワード取得機能と、

前記キーワード取得機能において取得された前記キーワードに含まれる文字が前記文字列生成機能において生成された前記文字列に含まれるときは要素を真値とする、前記文字列の大きさを有する入力ベクトルを生成する入力ベクトル生成機能と、

前記入力ベクトル生成機能で生成された前記入力ベクトルと前記行列生成機能において生成された前記行列との積を演算して出力ベクトルを生成する出力ベクトル生成機能と、

前記出力ベクトル生成機能において生成された前記出力ベクトルの要素の真値の数を判定する判定機能と

を実現させる情報処理プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置、情報処理方法および情報処理プログラムに関する。

【背景技術】

【0002】

従来から、ユーザの発話又は操作等に基づいて文章を入力して対話を行う対話システムが提案されている。対話システムに用いられているものとして、入力した文章に自動応答を行うためのルールを記述するAIML (Artificial Intelligence Markup Language) が知られている。

【0003】

特許文献1には、文章を入力する入力部と、前記入力部によって入力された文章と、予め設定されたマッチング用文章とのマッチングを判断するマッチング判断部と、を備え、前記マッチング用文章は、複数の文字列の単位に分割された構成となっており、当該単位の少なくとも一つは、更に複数の文字列に分割された複数の候補を含んでおり、前記マッチング判断部は、前記入力部によって入力された文章を複数の文字列の単位に分割されたものとして、マッチングを判断する、文章マッチングシステムが記載されている。

【0004】

また、従来から、検索対象の全ての文書について転置インデックスを予め作成しておき、検索用語と転置インデックスとを用いて文書検索を行う検索方法が知られている。

【0005】

特許文献2には、情報処理装置の抽出手段は、文書と該文書の特徴を対応付けた転置インデックスから、選択された文書の特徴を抽出し、検索手段は、前記抽出手段によって抽出された特徴のいずれかを有する文書を、前記転置インデックスによって検索し、選択手段は、前記検索手段による複数の検索結果に対応する特徴のうち、該検索結果の絞り込みの効果の高い特徴を選択し、提示手段は、前記選択手段によって選択された特徴の組み合わせを検索条件として提示する情報処理装置が記載されている。

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2019-192195号公報

【特許文献2】特開2016-091354号公報

【発明の概要】

【発明が解決しようとする課題】

10

20

30

40

50

【0007】

しかし、特許文献1に記載の技術においては、入力した文章とマッチングを行うためのマッチング用文章を、予め用意しておく必要がある。従来のA I M Lでは、入力した文章に対して適切な自動応答を行うためには、一部が異なるだけの数多くのマッチング用文章を用意しておかなければならず、検索の事前準備としてのマッチング用文章の作成に時間がかかっていた。

【0008】

また、特許文献2に記載の技術においては、予め転置インデックスを記録しておく必要があり、検索の事前準備としての転置インデックスの作成に時間がかかっていた。

【0009】

本発明は上記事情に鑑みてなされたものであり、検索の事前準備のための時間を短縮することを一つの目的とする。

【課題を解決するための手段】

【0010】

上記の課題を解決するための一つの実施形態として、情報処理装置は、検索対象のキーワードに基づきデータベースに記憶された複数の被検索対象の文書を検索する情報処理装置であって、文書に含まれる文字をスキャンして出現する文字が重複しない文字列を生成する文字列生成部と、文書が有する文章に含まれる文字が文字列生成部において生成された文字列に含まれるか否かを示す数字列を文章毎に並べた行列を生成する行列生成部と、キーワードを取得するキーワード取得部と、キーワード取得部において取得されたキーワードに含まれる文字が文字列生成部において生成された前記文字列に含まれるときは要素を正值とする、文字列の大きさを有する入力ベクトルを生成する入力ベクトル生成部と、入力ベクトル生成部で生成された入力ベクトルと行列生成部において生成された行列との積を演算して出力ベクトルを生成する出力ベクトル生成部と、出力ベクトル生成部において生成された出力ベクトルの要素の正值の数を判定する判定部と、を備える。

【発明の効果】

【0011】

本発明の一つの実施形態によれば、情報処理装置は、文書に含まれる文字をスキャンして出現する文字が重複しない文字列を生成し、文書が有する文章に含まれる文字が生成された文字列に含まれるか否かを示す数字列を文章毎に並べた行列を生成し、キーワードを取得し、取得されたキーワードに含まれる文字が生成された文字列に含まれるときは要素を正值とする、文字列の大きさを有する入力ベクトルを生成し、生成された入力ベクトルと生成された行列との積を演算して出力ベクトルを生成し、生成された出力ベクトルの要素の正值の数を判定することにより、検索の事前準備のための時間を短縮することができる。

【図面の簡単な説明】

【0012】

【図1】実施形態の情報処理方法における処理の概要の一例を示す図である。

【図2】実施形態の情報処理装置における第1の処理方法による検索前処理一例を示す図である。

【図3】実施形態の情報処理装置における第1の処理方法による検索処理一例を示す図である。

【図4】実施形態の情報処理装置における第2の処理方法による検索前処理一例を示す図である。

【図5】実施形態の情報処理装置における第2の処理方法による検索処理一例を示す図である。

【図6】実施形態の情報処理装置におけるソフトウェア構成の一例を示すブロック図である。

【図7】実施形態の情報処理装置におけるハードウェア構成の一例を示すブロック図である。

10

20

30

40

50

【発明を実施するための形態】

【0013】

以下、図面を参照して本発明の一実施形態における情報処理装置、情報処理方法および情報処理プログラムについて詳細に説明する。

【0014】

先ず、図1を用いて、情報処理方法の概要を説明する。図1は、実施形態の情報処理方法における処理の概要の一例を示す図である。

【0015】

図1において、情報処理方法は、検索前処理S1と検索処理S2を含む。検索前処理S1は、検索処理S2が実施される前に実施される処理である。本実施例における情報処理方法は、検索前処理S1を高速化することにより、検索の事前準備のための時間を短縮することができる。

10

【0016】

本実施形態における検索処理S2は、検索対象のキーワードに基づきデータベースに記憶された複数の被検索対象の文書を検索する処理である。検索対象のキーワードとは、実施形態の情報処理方法を利用して検索を実行する利用者によって入力されるキーワードである。キーワードは、1または複数の文字からなる文字列である。キーワードには、複数の文字の種類を含めることができる。文字の種類とは、例えば、ひらがな、漢字、英字、数字、句読点、記号、絵文字等である。

20

【0017】

検索処理S2は、データベースに記憶された複数の文書を被検索対象として実施される。文書は、1または複数の文章から構成される。文章は、1または複数の文字から構成される。本実施形態における文章は、例えば、句点、ピリオド、またはスペース等によって区切られる文字列である。

【0018】

検索前処理S1は、文字列生成処理S11、連番付与処理S12、行列生成処理S13および行列保存処理S14を含む。

【0019】

文字列生成処理S11は、被検索対象の文書に含まれる文字をスキャンして出現する文字が重複しない文字列を生成する処理である。ここで、図2および図3を用いて文字列生成処理S11の第1の処理方法を説明する。図2は、実施形態の情報処理装置における第1の処理方法による検索前処理S11の一例を示す図である。また、図3は、実施形態の情報処理装置における第1の処理方法による検索処理一例を示す図である。

30

【0020】

図2において、文字列生成処理S11は、先頭文字抽出処理S111および先頭文字列生成処理S112を含む。先頭文字抽出処理S111は、文書の先頭から文字をスキャンしていき、文書において最初に出現する文字を抽出する。図2は、被検索対象の文書が、「今日は晴れていて気持ちの良い日です。昨日は全日曇りで、とても寒かったです。明日は雨の予報なので、心配です。」の3つの文章を有する場合を示す。なお、「今日は晴れていて気持ちの良い日です。」を第1文章とする。「昨日は全日曇りで、とても寒かったです。」を第2文章とする。また、「明日は雨の予報なので、心配です。」を第3文章とする。先頭文字抽出処理S111において文書において最初に出現する文字として抽出された文字を塗りつぶしで示す。先頭文字抽出処理S111において抽出された文字は、「今日は晴れてい気持ちの良です。昨日全曇り、とも寒かった明雨予報な心配」となる。

40

【0021】

先頭文字列生成処理S112は、先頭文字抽出処理S111において抽出された文字を文字の種類毎に文字列として生成する。図2においては、文字の種類が、ひらがな、句読点および漢字の3種類である場合を例示している。先頭文字列生成処理S112において、先頭文字抽出処理S111において抽出された文字である、「今日は晴れてい気持ちの良です。昨日全曇り、とも寒かった明雨予報な心配」をひらがな、句読点および漢字の順

50

の文字列として、「はれていちのですりともかったな、。今日晴気持良昨全曇寒明雨予報心配」の文字列が生成される。

【0022】

なお、文字列生成処理S11においては、被検索対象の文書に含まれる文字をスキャンして出現する文字が重複しない文字列を生成する処理として、文書の先頭からスキャンして最初に出現する文字を文字の種類毎に文字列として生成する処理（先頭文字抽出処理S111および先頭文字列生成処理S112）を例示した。しかし、文字列生成処理S11において生成される文字列は文字が重複しないものであればよく、上記実施例に限定されるものではない。文字列生成処理S11は、例えば、文字が出現する順序に無関係に並べられた文字列を生成するものであってもよい。

10

【0023】

また、文字列生成処理S11は、文字の種類とは無関係に並べられた文字列を生成するものであってもよい。

【0024】

また、文字列生成処理S11は、文字列は予め定められた固定の文字列（以下、「固定文字列」という。）を生成（選択）するものであってもよい。固定文字列におけるそれぞれの文字は重複しないものであれば任意の文字列を用いることができる。例えば、固定文字列は被検索対象の文書とは無関係な文字列としてもよい。この場合、固定文字列の文字は被検索対象の文書に含まれていない可能性があり、後述する判定処理においてヒットと判定されても原文の再検索において却下と判定される可能性が高くなるおそれがある。しかし、固定文字列とすることにより、被検索対象の文書が変更または追加された場合であっても文字列生成処理S11を再度行う必要がなくなり、処理の高速化が可能となる。

20

【0025】

図1に戻り、検索前処理S1は、文字列生成処理S11を実行した後に、連番付与処理S12を実行する。連番付与処理S12は、被検索対象の文書に含まれる文字に対して文章毎に連番を付与する処理である。図2を用いて連番付与処理S12の一例を説明する。

【0026】

図2において、連番付与処理S12は、第1の連番付与処理S121を含む。第1の連番付与処理S121においては、被検索対象の文書を文章ごとに分解して、それぞれの文章毎に連番を付与する。例えば、第1文章である「今日は晴れていて気持ちの良い日です。」は18文字で構成されるため、1～18の連番が付与される。同様に、第2文章である、「昨日は全日曇りで、とても寒かったです。」には、1～19の連番が付与される。また、第3文章である、「明日は雨の予報なので、心配です。」には、1～16の連番が付与される。

30

【0027】

図1に戻り、検索前処理S1は、連番付与処理S12を実行した後に、行列生成処理S13を実行する。行列生成処理S13は、文書が有する文章に含まれる文字が文字列生成処理S11において生成された文字列に含まれるか否かを示す数字列を文章毎に並べた行列を生成する処理である。図2を用いて行列生成処理S13の一例を説明する。

【0028】

図2において、行列生成処理S13は、第1の行列生成処理S131を含む。第1の行列生成処理S131は、文書が有する文章に含まれる文字が先頭文字列生成処理S112において生成された文字列に含まれる正値であるか否かを示す数字列を文章毎に並べた行列（以下、「第1行列」という。）を生成する処理である。生成された文字列に含まれるか否かを示す数字列とは、生成された文字列に含まれる正値であるか、文字列に含まれない負値であるか否かを示す数字を並べたものである。図2は、当該数字列が、第1の連番付与処理S121において付与された連番を用いた数字列（以下、「第1数字列」という。）である場合を示している。

40

【0029】

例えば、第1文章である「今日は晴れていて気持ちの良い日です。」には第1の連番付

50

図３において、入力ベクトル生成処理Ｓ２２は、入力ベクトル生成処理Ｓ２２１を含む。文字列生成処理Ｓ１１において生成された文字列は、上述のように、「はれていちのですりともかったな。今日晴気持良昨全曇寒明雨予報心配」の３３文字である。このため、入力ベクトルの要素の数は３３となる。文字列の１番目の「は」の文字は、キーワードである「良い日」に含まれない。このため、入力ベクトル生成処理Ｓ２２１において、入力ベクトルの１番目の要素は、負値であることを示す「０」とされる。同様に、文字列の２番目の「れ」の文字および３番目の「て」の文字は、「良い日」に含まれないため、入力ベクトルの２番目の要素および３番目の要素は「０」とされる。一方、文字列の４番目の「い」の文字は、「良い日」に含まれるため、入力ベクトルの４番目の要素は、正値であることを示す「１」とされる。同様に、３３個の全ての要素について「１」または「０」のいずれかの値を有する入力ベクトルが生成される。その結果、入力ベクトル生成処理Ｓ２２１において、３３個の要素を有する「０００１０００００００００００００００１０００１００００００００００」の入力ベクトルが生成される。

【 0 0 3 8 】

図 1 に戻り、検索処理 S 2 は、入力ベクトル生成処理 S 2 2 を実行した後に、出力ベクトル生成処理 S 2 3 を実行する。出力ベクトル生成処理 S 2 3 は、入力ベクトル生成処理 S 2 2 で生成された入力ベクトルと行列生成処理 S 1 3 において生成された行列との積を演算して出力ベクトルを生成する。例えば、行列が A、入力ベクトルが x であった場合、出力ベクトル y は、 $y = Ax$ で生成される。

【 0 0 3 9 】

図３において、出力ベクトル生成処理Ｓ２３は、出力ベクトル生成処理Ｓ２３１を含む。出力ベクトル生成処理Ｓ２３１において、入力ベクトル生成処理Ｓ２２１で生成された３３個の要素を有する入力ベクトルと、行列生成処理１３１で生成された３行３３列の第１行列との積が実行される。出力ベクトル生成処理Ｓ２３１においては、入力ベクトルと第１行列の第１行の積、入力ベクトルと第２行の積、および入力ベクトルと第３行の積の３つの積が実行される。入力ベクトルと第１行の積は、「００００１１００００００００
０００００００９１０００００００００００」と演算される。入力ベクトルと第２行の積は、「００００００００００００００００００００００００００００」と演算される。また、入力ベクトルと第３行の積は、「００００００００００００００００００
００００００００００００００００」と演算される。出力ベクトルは上記３つの行を有する３３列のベクトルとなる。

【 0 0 4 0 】

出力ベクトル生成処理 S 2 3 における上記 3 つの積は並列で処理することができるため処理時間を短縮することができる。例えば、出力ベクトル生成処理 S 2 3 1 において並列処理が可能な GPU (Graphics Processing Unit) を用いた場合、3 つの積は並列で処理できる。このため、出力ベクトル生成処理 S 2 3 1 の処理時間は、逐次処理を行う場合に対して約 $1/3$ に短縮することができる。

【 0 0 4 1 】

図 1 に戻り、検索処理 S 2 は、出力ベクトル生成処理 S 2 3 を実行した後に、判定処理 S 2 4 を実行する。判定処理 S 2 4 は、出力ベクトル生成処理 S 2 3 において生成された出力ベクトルの要素の正値の数を判定する処理である。判定処理 S 2 4 の詳細を、図 3 を用いて説明する。

【 0 0 4 2 】

図 3 において、判定処理 S 2 4 は、第 1 判定処理 S 2 4 1 を含む。第 1 判定処理 S 2 4 1 は、出力ベクトル生成処理 S 2 3 1 において生成された 3 行 3 3 列の出力ベクトルのそれぞれの行において、要素が「0」以外の正値の数を数える。判定処理 S 2 4 1 は、正値の数がキーワードの文字数である 3 個であって、かつ、出力ベクトル生成処理 S 2 3 1 において生成された出力ベクトルの真値の順序が行列生成処理 1 3 1 において生成された第 1 行列における要素の順序である場合に、文書がキーワードにおいてヒットしたと判定する。ここで、ヒットの判定は、正値の数がキーワードの文字数である 3 個であって、かつ

示す図である。また、図５は、実施形態の情報処理装置における第２の処理方法による検索処理一例を示す図である。第２の処理方法は、第１の処理方法において実行される連番付与処理Ｓ１２を実行しない処理方法である。

【 0 0 5 0 】

図 4 において、文字列生成処理 S 1 1 は、先頭文字抽出処理 S 1 1 3 および先頭文字列生成処理 S 1 1 4 を含む。先頭文字抽出処理 S 1 1 3 の処理は先頭文字抽出処理 S 1 1 1 の処理と同様である。また、先頭文字列生成処理 S 1 1 4 の処理は先頭文字列生成処理 S 1 1 2 と同様である。このため、先頭文字抽出処理 S 1 1 3 および先頭文字列生成処理 S 1 1 4 の説明は省略する。

【 0 0 5 1 】

図４において、第２の行列生成処理Ｓ１３２は、文書が有する文章に含まれる文字が、先頭文字列生成処理Ｓ１１４において生成された文字列に含まれる正値であるか、または文字列に含まれない負値であるかを示す数字列を文章毎に並べた行列（以下、「第２行列」という。）を生成する処理である。図４は、当該数字列が、正値である場合「１」、負値である場合「０」を用いる場合を示している。第２行列は第１行列に含まれていた連番が含まれない。このため、行列生成処理Ｓ１３２で生成される行列には、先頭文字抽出処理Ｓ１１３で抽出された文字の順序は記録されない。

【 0 0 5 2 】

行列生成処理 S 1 3 2 において、先頭文字列生成処理 S 1 1 4 において生成された文字列「はれていちのですりともかったな。今日晴気持良昨全曇寒明雨予報心配」の１番目の「は」の文字は第 1 文章の３番目に含まれるため、「は」には正值である「1」の数字が割り当てられる。２番目の「れ」の文字は第 1 文章の５番目に含まれるため、「れ」には正值である「1」の数字が割り当てられる。一方、９番目の「り」の文字は第 1 文章には含まれないため、「り」には、文字列に含まれないことを示す負値の「0」の数字が割り当てられる。この処理を先頭文字列生成処理 S 1 1 4 において生成された文字列全てについて行くと、「1、1、1、1、1、1、1、1、0、0、0、0、0、0、0、0、1、1、1、1、1、1、1、0、0、0、0、0、0、0、0、0、0、0、0」の 31 個の数字からなる第 2 数字列が生成される。第 2 の行列生成処理 S 1 3 2 においては、第 2 文章および第 3 文章においても同様の処理を行うことにより、図示する 3 行 33 列の第 1 行列を生成することができる。

【 0 0 5 3 】

図 5 において、取得処理 S 2 1 2 は、取得処理 S 2 1 1 と同様であり、入力ベクトル生成処理 S 2 2 2 は、入力ベクトル生成処理 S 2 2 1 と同様である。このため、取得処理 S 2 1 2 および入力ベクトル生成処理 S 2 2 2 の説明は省略する。

【 0 0 5 4 】

図５において、出力ベクトル生成処理Ｓ２３は、出力ベクトル生成処理Ｓ２３２を含む。出力ベクトル生成処理Ｓ２３２において、行列生成処理１３２で生成された３行３３列の第１行列と、入力ベクトル生成処理Ｓ２２２で生成された３３個の要素を有する入力ベクトルとの積が実行される。出力ベクトル生成処理Ｓ２３２においては、第２行列の第１行と入力ベクトルとの積、第２行と入力ベクトルとの積、および第３行と入力ベクトルとの積の３つの積が実行される。第１行と入力ベクトルとの積は、「００００１１００００
０００００００００００９１００００００００００」 と演算される。第２行と入力ベクトルとの積は、「００００００００００００００００００００００００
００」と演算される。また、第３行と入力ベクトルとの積は、「００００００００００
０００００００００００００００００００００」 と演算される。出力ベクトルは上記３つの行を有する３行３３列のベクトルとなる。

【 0 0 5 5 】

判定処理 S 2 4 は、第 1 判定処理 S 2 4 2 を含む。第 2 判定処理 S 2 4 2 は、出力ベクトル生成処理 S 2 3 2 において生成された 3 行 3 3 列の出力ベクトルのそれぞれの行において、要素が「1」の正値の数を数える。判定処理 S 2 4 2 は、正値の数が 3 個である場

合、文書がキーワードにおいてヒットしたと判定する。一方、判定処理 S 2 4 2 は、正値の数が 3 個未満である場合、文書がキーワードにおいてヒットしていないと判定（却下判定）する。

【0056】

第 2 の処理方法においては、上述の通り、生成される行列に被検索対象の文書における文字（最初に出現する文字）の順序を記録しないため、ヒットした文書においては行列と入力ベクトルとの積により算出される出力ベクトルにもその順序が記録されない。このため、第 2 の処理方法においては、第 1 の処理方法に比べて、再検索により却下判定となる可能性が高くなる。しかし、第 2 の処理方法においては、第 1 の処理方法で実行される連番付与処理を省略できるため、検索前処理 S 1 の負荷がさらに軽減される。

10

【0057】

以上説明した情報処理方法は、以下に説明する情報処理装置において実行することができる。図 6 を用いて、情報処理装置のソフトウェア構成を説明する。図 6 は、実施形態の情報処理装置におけるソフトウェア構成の一例を示すブロック図である。

【0058】

図 6 において、情報処理装置 1 は、ネットワーク 9 を介して、文書データベース 2 および利用者端末 3 と通信可能に接続されている。文書データベース 2 は、被検索対象の文書を複数記憶することができる。文書データベースに記憶された文書は、複数の文章を有する文書であって、例えば特許文献である。利用者端末 3 は、情報処理装置 1 を利用する利用者が操作する端末である。利用者端末 3 は検索アプリ 3 1 を有する。検索アプリ 3 1 は、情報処理装置 1 を介して文書データベース 2 に記憶された文書を検索するアプリであって、上述した検索処理 S 2 の処理を情報処理装置 1 に実行させる。検索アプリ 3 1 は、検索処理 2 の処理を専用に行うアプリであってもよく、また、Web ブラウザ等の汎用のアプリであってもよい。ネットワーク 9 は、有線または無線を介したネットワークであり、例えばインターネットである。

20

【0059】

なお、図 6 は情報処理装置 1 の接続先を例示するものであり、情報処理装置 1 の接続先を限定するものではない。例えば、情報処理装置 1 は、複数の文書データベース 2 または複数の利用者端末 3 と接続されるものであってもよい。また、図 6 は、情報処理装置 1 と文書データベース 2 とがネットワーク 9 を介して接続される場合を示したが、情報処理装置 1 と文書データベース 2 は、ネットワーク 9 を介さない一体の装置またはシステムにおいて構成されるものであってもよい。

30

【0060】

情報処理装置 1 は、文字列生成部 1 1、連番付与部 1 2、行列生成部 1 3、キーワード取得部 1 4、入力ベクトル生成部 1 5、出力ベクトル生成部 1 6、判定部 1 7 および行列保存部 1 8 を有する。なお、情報処理装置 1 は、図 1 等において説明した情報処理方法を実行する装置である。情報処理装置 1 における、文字列生成部 1 1、連番付与部 1 2、行列生成部 1 3、キーワード取得部 1 4、入力ベクトル生成部 1 5、出力ベクトル生成部 1 6、判定部 1 7 および行列保存部 1 8 は、本実施形態における情報処理を実行するための情報処理プログラム（ソフトウェア）によって実現される機能モジュールであるものとして説明する。

40

【0061】

文字列生成部 1 1 は、データベースに記憶された複数の被検索対象の文書に含まれる文字をこの文書の先頭からスキャンして最初に出現する文字を文字の種類毎に文字列として生成する。文字列生成部 1 1 は、図 1 において説明した文字列生成処理 S 1 1 を実行する。

【0062】

連番付与部 1 2 は、データベースに記憶された複数の被検索対象の文書に含まれる文字に対して文章毎に連番を付与する。連番付与部 1 2 は、図 1 において説明した連番付与処理 S 1 2 を実行する。

50

【0063】

行列生成部13は、データベースに記憶された複数の被検索対象の文書が有する文章に含まれる文字が文字列生成部11において生成された文字列に含まれるか否かを示す数字列を被検索対象の文書が有する文章毎に並べた行列を生成する。行列生成部13は、図1において説明した行列生成処理S13を実行する。

【0064】

キーワード取得部14は、検索対象のキーワードを取得する。キーワード取得部14は、図1において説明したキーワード取得処理S21を実行する。

【0065】

入力ベクトル生成部15は、キーワード取得部14において取得されたキーワードに含まれる文字が文字列生成部11において生成された文字列に含まれるときは要素を真値とする、文字列の大きさを有する入力ベクトルを生成する。入力ベクトル生成部15は、図1において説明した入力ベクトル生成処理S22を実行する。

10

【0066】

出力ベクトル生成部16は、入力ベクトル生成部15で生成された入力ベクトルと行列生成部13において生成された行列との積を演算して出力ベクトルを生成する。出力ベクトル生成部16は、図1において説明した出力ベクトル生成処理S23を実行する。

【0067】

判定部17は、出力ベクトル生成部16において生成された出力ベクトルの要素の真値の数を判定する。判定部17は、図1において説明した判定処理S24を実行する。

20

【0068】

行列保存部18は、行列生成部13において生成された行列を保存する。行列保存部18は、図1において説明した行列保存S14を実行する。

【0069】

なお、情報処理装置1が有する上述の各機能部は、情報処理装置1の機能部の一例を示したものであり、情報処理装置1が有する機能を限定したものではない。例えば、情報処理装置1は、上記全ての機能部を有している必要はなく、一部の機能部を有するものであってもよい。また、情報処理装置1は、上記以外の他の機能を有するものであってもよい。例えば、情報処理装置1は、情報を入力するために入力機能、情報を表示するための表示機能、装置の稼働状態を報知する報知機能等を有していてもよい。

30

【0070】

また、情報処理装置1が有する上記各機能部は、上述の通り、ソフトウェアによって実現されるものとして説明した。しかし、情報処理装置1が有する上記機能部の中で少なくとも1つ以上の機能部は、ハードウェアによって実現されるものであってもよい。

【0071】

また、情報処理装置1が有する上記何れかの機能部は、1つの機能部を複数の機能部に分割して実施してもよい。また、情報処理装置1が有する上記何れか2つ以上の機能部を1つの機能部に集約して実施してもよい。情報処理装置1は、情報処理装置1が有するそれぞれの機能をそれぞれの機能ブロックで表現したものであり、プログラムファイル数等の物理的な構成を限定するものではない。例えば、各機能部がそれぞれ別個のプログラムファイル等で構成されていてもよく、また複数の機能部が1つのプログラムファイル等で構成されていてもよい。

40

【0072】

また、情報処理装置1は、1つの筐体によって実現される装置であっても、ネットワーク等を介して接続された複数の装置から実現されるシステムであってもよい。例えば、情報処理装置1は、その機能の一部または全部をクラウドコンピューティングシステムによって提供されるクラウドサービス等、他の仮想的な装置によって実現するものであってもよい。また、情報処理装置1は、上記各機能部のうち、少なくとも1以上の機能部を他の装置において実現するようにしてもよい。また、情報処理装置1は、デスクトップPC等の汎用的なコンピュータであってもよく、機能が限定された専用の装置であってもよい。

50

【0073】

次に、図7を用いて、情報処理装置1のハードウェア構成を説明する。図7は、実施形態における情報処理装置1のハードウェア構成の一例を示すブロック図である。

【0074】

情報処理装置1は、CPU (Central Processing Unit) 101、RAM (Random Access Memory) 102、ROM (Read Only Memory) 103、I/O機器104、および通信I/F (Interface) 105を有する。情報処理装置1は、図1で説明した情報処理プログラムを実行する装置である。

【0075】

CPU 101は、RAM 102またはROM 103に記憶された情報処理プログラムを実行することにより、情報処理装置1の制御を行う。情報処理プログラムは、例えば、プログラムを記録した記録媒体、又はネットワークを介したプログラム配信サーバ等から取得されて、ROM 103にインストールされ、CPU 101から読出されて実行される。

【0076】

I/O機器104は、操作入力機能と表示機能（操作表示機能）を有する。I/O機器104は、例えばタッチパネルである。タッチパネルは、情報処理装置1の利用者に対して指先又はタッチペン等を用いた操作入力を可能にする。本実施形態におけるI/O機器104は、操作表示機能を有するタッチパネルを用いることができるが、I/O機器104は、表示機能を有する表示装置と操作入力機能を有する操作入力装置とを別個有するものであってもよい。その場合、タッチパネルの表示画面は表示装置の表示画面、タッチパネルの操作は操作入力装置の操作として実施することができる。なお、I/O機器104は、ヘッドマウント型、メガネ型、腕時計型のディスプレイ等の種々の形態によって実現されてもよい。

【0077】

通信I/F 105は、通信用のI/Fである。通信I/F 105は、例えば、無線LAN、有線LAN、赤外線等の近距離無線通信を実行する。図は通信用のI/Fとして通信I/F 105のみを図示するが、情報処理装置1は複数の通信方式においてそれぞれの通信用のI/Fを有するものであってもよい。

【0078】

なお、本実施形態で説明した装置を構成する機能を実現するためのプログラムを、コンピュータ読み取り可能な記録媒体に記録して、当該記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することにより、本実施形態の上述した種々の処理を行ってもよい。なお、ここでいう「コンピュータシステム」とは、OSや周辺機器等のハードウェアを含むものであってもよい。また、「コンピュータシステム」は、WWWシステムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、フラッシュメモリ等の書き込み可能な不揮発性メモリ、CD-ROM等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。

【0079】

さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムが送信された場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリ（例えばDRAM (Dynamic Random Access Memory)）のように、一定時間プログラムを保持しているものも含むものとする。また、上記プログラムは、このプログラムを記憶装置等に格納したコンピュータシステムから、伝送媒体を介して、あるいは、伝送媒体中の伝送波により他のコンピュータシステムに伝送されてもよい。ここで、プログラムを伝送する「伝送媒体」は、インターネット等のネットワーク（通信網）や電話回線等の通信回線（通信線）のように情報を伝送する機能を有する媒体のことをいう。また、上記プログラムは、

10

20

30

40

50

前述した機能の一部を実現するためのものであっても良い。さらに、前述した機能をコンピュータシステムにすでに記録されているプログラムとの組合せで実現するもの、いわゆる差分ファイル（差分プログラム）であっても良い。

【0080】

以上、本発明の実施形態について、図面を参照して説明してきたが、具体的な構成はこ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においての種々の変更も含まれる。

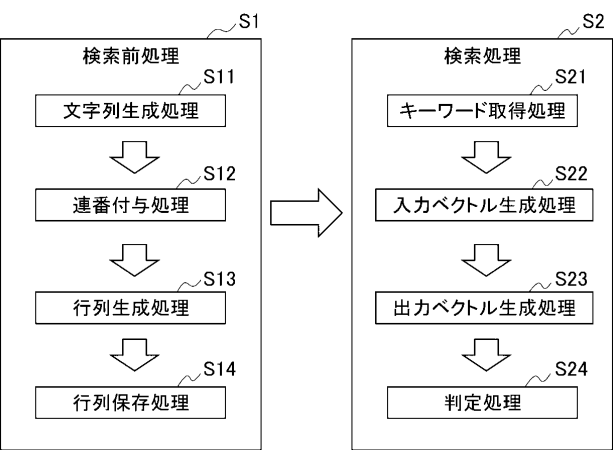
【符号の説明】

【0081】

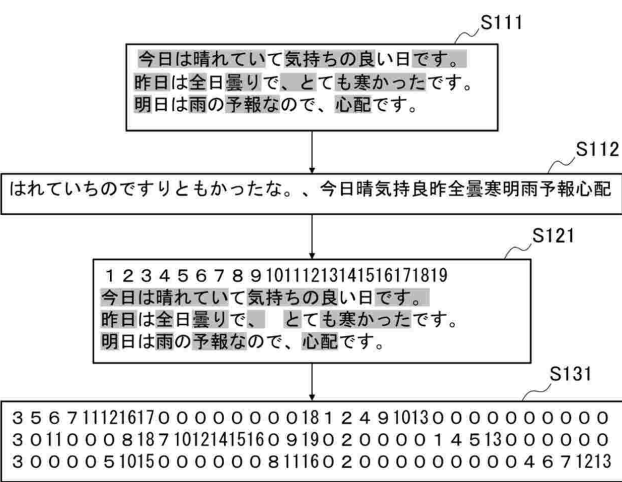
- 1 情報処理装置
- 11 文字列生成部
- 12 連番付与部
- 13 行列生成部
- 14 キーワード取得部
- 15 入力ベクトル生成部
- 16 出力ベクトル生成部
- 17 判定部
- 18 行列保存部
- 2 文書データベース
- 3 利用者端末
- 31 検索アプリ
- 9 ネットワーク
- 101 CPU
- 102 RAM
- 103 ROM
- 104 I/O機器
- 105 通信I/F

【図面】

【図1】



【図2】



10

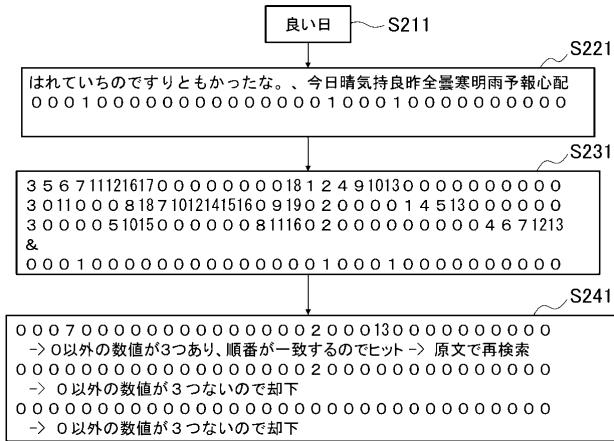
20

30

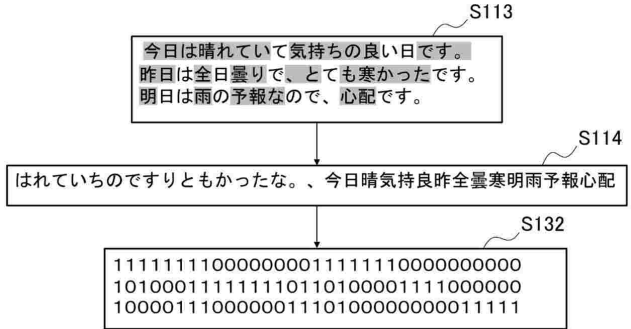
40

50

【図 3】

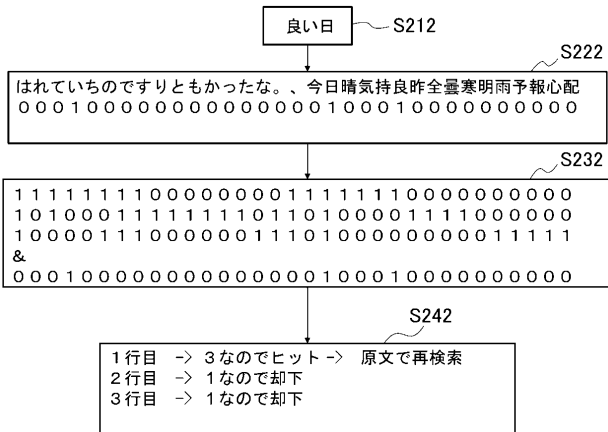


【図 4】

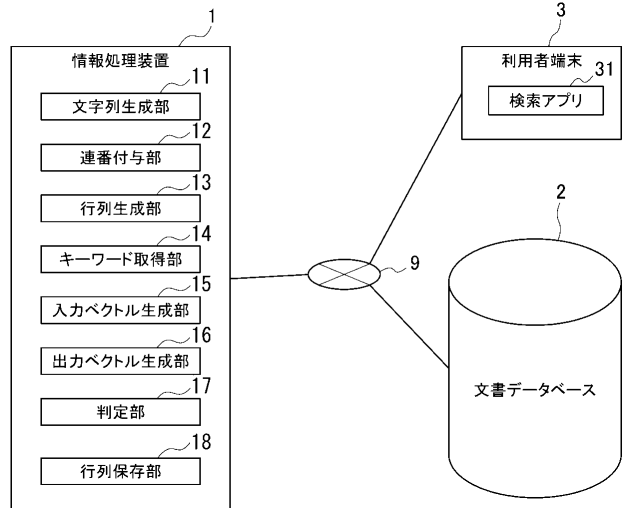


10

【図 5】



【図 6】



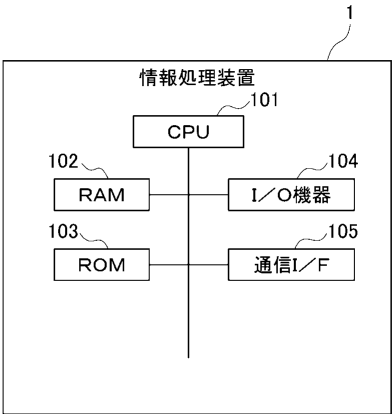
20

30

40

50

【図 7】



10

20

30

40

50