

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7251625号
(P7251625)

(45)発行日 令和5年4月4日(2023.4.4)

(24)登録日 令和5年3月27日(2023.3.27)

(51)国際特許分類 F I
G 0 6 F 16/338(2019.01) G 0 6 F 16/338

請求項の数 10 (全18頁)

(21)出願番号	特願2021-528777(P2021-528777)	(73)特許権者	000001993 株式会社島津製作所 京都府京都市中京区西ノ京桑原町1番地
(86)(22)出願日	令和1年6月27日(2019.6.27)	(74)代理人	110001195 弁理士法人深見特許事務所
(86)国際出願番号	PCT/JP2019/025571	(72)発明者	石川 勇樹 京都府京都市中京区西ノ京桑原町1番地 株式会社島津製作所内
(87)国際公開番号	WO2020/261479	審査官	松尾 真人
(87)国際公開日	令和2年12月30日(2020.12.30)		
審査請求日	令和3年11月2日(2021.11.2)		

最終頁に続く

(54)【発明の名称】 関連文書を検索して表示する方法およびシステム

(57)【特許請求の範囲】

【請求項1】

複数の文書を含むデータベースから検索クエリに関連する少なくとも1つの関連文書を検索して表示する、計算機によって実行される方法であって、

任意の言語単位をベクトル表現に変換するベクトル空間を用いて前記複数の文書の各々についてのベクトル表現と前記検索クエリのベクトル表現との距離を算出し、前記距離に応じて前記少なくとも1つの関連文書を検索する検索工程と、

前記少なくとも1つの関連文書の各々を表示する表示工程とを含み、
前記表示工程は、

当該関連文書に含まれる複数の言語単位の各々のベクトル表現と前記検索クエリのベクトル表現とに基づく当該言語単位と

前記検索クエリと

の関連性の大きさに応じた表示態様で、当該言語単位を表示する工程を含み、

前記関連性の大きさは、前記複数の言語単位の各々の寄与度であり、

前記寄与度は、前記少なくとも1つの関連文書の各々から当該言語単位を除いたベクトル表現と前記検索クエリのベクトル表現との距離から、当該関連文書のベクトル表現と前記検索クエリのベクトル表現との距離を減算した値である、方法。

【請求項2】

前記表示態様は、当該言語単位の周辺領域の色を含む、請求項1に記載の方法。

【請求項3】

10

20

前記関連性の大きさは、前記複数の言語単位の各々のベクトル表現と前記検索クエリのベクトル表現との距離である、請求項 1 または 2 に記載の方法。

【請求項 4】

前記検索工程は、前記検索クエリのベクトル表現と前記少なくとも 1 つの関連文書の各々についてのベクトル表現との距離が短いほど、当該関連文書を上位に順位付ける工程を含み、

前記表示工程は、前記検索工程による順位付けに従って前記少なくとも 1 つの関連文書を表示する工程を含む、請求項 1 に記載の方法。

【請求項 5】

前記複数の言語単位の種類は、文字、形態素、単語、文章、段落、およびこれらの任意の組み合わせからなる群から選択される、請求項 1 に記載の方法。

10

【請求項 6】

前記方法は、クライアントサーバシステムにおいて行われ、

前記表示工程は、サーバサイドにおいて行われる、請求項 1 に記載の方法。

【請求項 7】

前記方法は、クライアントサーバシステムにおいて行われ、

前記表示工程は、クライアントサイドにおいて行われる、請求項 1 に記載の方法。

【請求項 8】

前記検索工程は、表示されている前記複数の言語単位に含まれる或る言語単位がユーザによって選択された場合、当該言語単位を前記検索クエリとして前記少なくとも 1 つの関連文書を検索する工程を含む、請求項 1 に記載の方法。

20

【請求項 9】

前記ベクトル空間は、前記データベースを含むコーパスに対して自然言語処理を行う機械学習によって生成される、請求項 1 に記載の方法。

【請求項 10】

複数の文書から検索クエリに関連する少なくとも 1 つの関連文書を検索して表示するシステムであって、

前記複数の文書を含むデータベースと、

任意の言語単位をベクトル表現に変換するベクトル空間を用いて前記複数の文書の各々についてのベクトル表現と前記検索クエリのベクトル表現との距離を算出し、前記距離に応じて前記少なくとも 1 つの関連文書を検索し、前記少なくとも 1 つの関連文書の各々を表示する検索処理部とを備え、

30

前記検索処理部は、

当該関連文書に含まれる複数の言語単位の各々のベクトル表現と前記検索クエリのベクトル表現とに基づく当該言語単位と

前記検索クエリと

の関連性の大きさに応じた表示態様で、当該言語単位を表示し、

前記関連性の大きさは、前記複数の言語単位の各々の寄与度であり、

前記寄与度は、前記少なくとも 1 つの関連文書の各々から当該言語単位を除いたベクトル表現と前記検索クエリのベクトル表現との距離から、当該関連文書のベクトル表現と前記検索クエリのベクトル表現との距離を減算した値である、システム。

40

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、関連文書を検索して表示する方法およびシステムに関する。

【背景技術】

【0002】

従来、検索クエリに関連する文書（関連文書）を検索して表示する方法が知られている。たとえば、特開 2018 - 10482 号公報（特許文献 1）には、検索クエリと、検索クエリに概念的に適合する検索対象文書である正解文書の集合との組の集合（正解情報）

50

を用いて、ユーザが入力した検索クエリに概念的に適合する検索対象文書を検索する文書概念検索方法が開示されている。当該文書概念検索方法によれば、正解情報を用いることにより検索精度を向上させることができる。

【先行技術文献】

【特許文献】

【0003】

【文献】特開2018-10482号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

人工知能分野の自然言語処理では機械学習を応用して、単語の意味をベクトル表現（分散表現）に変換する技術が用いられている。具体的には、或る単語の意味は当該単語の周囲に出現する単語の分布から決定されるという分布仮説に基づいて、大量の文章情報をニューラルネットワークで学習し、各単語の意味を表すベクトル空間を生成する。

【0005】

互いに類似する意味を有する2つの言語単位のベクトル間距離は、相対的に短くなる。ベクトル表現に基づく文書検索方法によると、検索クエリに近い意味を有する言語単位をより多く含む文書が上位に順位付けされる。そのため、当該文書検索方法によると、検索クエリと一致する文字列が全く含まない文書であっても検索結果の上位に順位付けされる場合があり得る。そのような場合、当該文書が検索された根拠をユーザが理解することができない可能性がある。

【0006】

特許文献1に開示されている文書概念検索方法においては、検索クエリ概念ベクトルと検索対象文書の概念ベクトルとの類似度を降順にランキングした検索対象文書を検索結果として表示する。しかし、検索結果の根拠をユーザに示すことについて考慮されていない。

【0007】

本発明は、このような課題を解決するためになされたものであって、その目的は、複数の文書が格納されたデータベースから検索クエリに関連する少なくとも1つの関連文書を検索して表示する方法において、検索結果の根拠をユーザに示すことである。

【課題を解決するための手段】

【0008】

本発明の第1の態様に係る関連文書を検索して表示する方法は、複数の文書を含むデータベースから検索クエリに関連する少なくとも1つの関連文書を検索して表示する。当該方法は、検索工程と、表示工程とを含む。検索工程は、任意の言語単位をベクトル表現に変換するベクトル空間を用いて複数の文書の各々についてのベクトル表現と検索クエリのベクトル表現との距離を算出し、距離に応じて少なくとも1つの関連文書を検索する。表示工程は、少なくとも1つの関連文書の各々を表示する。表示工程は、当該関連文書に含まれる複数の言語単位の各々のベクトル表現と検索クエリのベクトル表現との関連性の大きさに応じた表示態様で、当該言語単位を表示する工程を含む。

【0009】

本発明の第2の態様に係る関連文書を検索して表示するシステムは、複数の文書から検索クエリに関連する少なくとも1つの関連文書を検索して表示する。当該システムは、データベースと、検索処理部とを備える。データベースには、複数の文書が格納されている。検索処理部は、任意の言語単位をベクトル表現に変換するベクトル空間を用いて複数の文書の各々についてのベクトル表現と検索クエリのベクトル表現との距離を算出し、距離に応じて少なくとも1つの関連文書を検索する。検索処理部は、少なくとも1つの関連文書の各々を表示する。検索処理部は、当該関連文書に含まれる複数の言語単位の各々のベクトル表現と検索クエリのベクトル表現との関連性の大きさに応じた表示態様で、当該言語単位を表示する。

【発明の効果】

10

20

30

40

50

【 0 0 1 0 】

本発明に係る関連文書を検索して表示する方法およびシステムによれば、関連文書に含まれる複数の言語単位の各々のベクトル表現と検索クエリのベクトル表現とに基づく当該言語単位と検索クエリとの関連性の大きさに応じた表示態様で、当該言語単位を表示することにより、検索結果の根拠をユーザに示すことができる。

【 図面の簡単な説明 】

【 0 0 1 1 】

【 図 1 】 実施の形態に係る関連文書を検索して表示するシステムの一例である分析事例検索システムの外觀図である。

【 図 2 】 図 1 の分析事例検索システムの構成を示す機能ブロック図である。

【 図 3 】 分析レポートの一例を示す図である。

【 図 4 】 図 2 の学習処理部によって行われる学習処理の流れを説明するためのフローチャートである。

【 図 5 】 図 2 の検索処理部によって行われる検索処理の流れを説明するためのフローチャートである。

【 図 6 】 図 2 の表示制御部によってディスプレイに表示される検索結果ウィンドウの一例を示す図である。

【 図 7 】 図 6 の関連文書が選択された場合に検索結果ウィンドウに表示される関連文書の内容を示す図である。

【 図 8 】 関連文書に含まれる単語と検索クエリとの関連性として当該単語の寄与度を用いて、関連文書に含まれる複数の単語の各々をハイライトした様子を示す図である。

【 図 9 】 ハイライトされる言語単位が文章である場合の検索結果ウィンドウに表示される関連文書の内容を示す図である。

【 図 1 0 】 ハイライトされる言語単位が段落である場合の検索結果ウィンドウに表示される関連文書の内容を示す図である。

【 図 1 1 】 実施の形態に係る関連文書を検索して表示するシステムの一例である分析事例検索システムがネットワークを介して複数のクライアント端末と接続された場合の、分析事例検索システムの構成を示す機能ブロック図である。

【 発明を実施するための形態 】

【 0 0 1 2 】

以下に、本発明の実施の形態について図面を参照して詳細に説明する。なお、以下では図中の同一または相当部分には同一符号を付してその説明は原則的に繰返さない。

【 0 0 1 3 】

図 1 は、実施の形態に係る関連文書を検索して表示するシステムの一例である分析事例検索システム 1 0 0 の外觀図である。図 1 に示されるように、分析事例検索システム 1 0 0 は、計算機 1 0 と、ディスプレイ 6 0 と、キーボード K B 1 と、マウス M S 1 とを備える。ディスプレイ 6 0 と、キーボード K B 1 と、マウス M S 1 とは、計算機 1 0 に接続されている。

【 0 0 1 4 】

ディスプレイ 6 0 には、検索ウィンドウ W n 1 およびカーソル C r が表示されている。ユーザは、マウス M S 1 を操作することによりカーソルを操作する。ユーザは、キーボード K B 1 を操作することにより検索クエリを検索ウィンドウ W n 1 に入力する。図 1 には、臭素酸の分析に適した分析装置、分析方法、および分析条件等が記載された文書を検索するため、検索クエリとして「臭素酸」という文字列をユーザが検索ウィンドウ W n 1 に入力した場合が示されている。

【 0 0 1 5 】

(分析事例検索システム 1 0 0 の構成)

図 2 は、図 1 の分析事例検索システム 1 0 0 の構成を示す機能ブロック図である。図 2 に示されるように、分析事例検索システム 1 0 0 は、学習処理部 2 0 と、検索処理部 4 0 とを備える。

10

20

30

40

50

【 0 0 1 6 】

分析事例検索システム 1 0 0 は、ユーザが入力した検索クエリに基づいて、データベース 3 0 に含まれる複数の文書を検索し、検索クエリに関連する関連文書を検索する。データベース 3 0 には、分析事例の分野の文書データが含まれている。分析事例の分野の文書データには、たとえば、分析レポート、分析関連論文、および分析関連特許文献が含まれる。分析レポートは、たとえば図 3 に示されるような化合物分析に関する文書であり、分析対象の化合物の分析方法に関する情報、分析装置に関する情報、および分析条件に関する情報等を含む。

【 0 0 1 7 】

(学習処理部 2 0 の構成)

再び図 2 を参照して、学習処理部 2 0 は、形態素解析部 2 1、ベクトル生成部 2 3、関連度学習部 2 5、コーパス 2 7、およびデータベース 3 0 を含む。コーパス 2 7 は、自然言語処理を用いる機械学習のために、分析事例の分野に関する大量の文書データが体系化されて蓄積された言語資料である。

【 0 0 1 8 】

形態素解析部 2 1 は、コーパス 2 7 に蓄積されているすべての文書データを、意味を有する最小の言語単位 (形態素あるいは単語) に形態素解析によって分割することによって細分化する。

【 0 0 1 9 】

ベクトル生成部 2 3 は、形態素解析部 2 1 による形態素解析の結果に対する機械学習によって、単語をベクトル表現に変換するベクトル空間を生成する。当該ベクトル空間は、ニューラルネットワークによって形成されるモデルに対する機械学習の過程で生成される。当該モデルとしては、たとえば、周辺の単語 (コンテキスト) から中央の単語 (ターゲット) を推論する C B O W (Continuous Bag-of-Words) モデル、あるいは中央の単語から周辺の単語を推論する s k i p - g r a m モデルを挙げることができる。

【 0 0 2 0 】

ベクトル生成部 2 3 は、単語の意味表現である単語ベクトルを生成する。ベクトル生成部 2 3 は、文章の特徴を表す文章ベクトル、および文書の特徴を表す文書ベクトルを文書中に含まれている単語の特徴量の総和等から生成する。なお、文書データの細分化は、任意の言語単位で行われてもよい。任意の言語単位の種類としては、文字、形態素、単語、文章、あるいは段落を挙げることができる。文章とは意味のある文の最小単位であり、文書は複数の文章で構成される。日本語の文章は句点で区切られ、英語等の言語で作成された文章はピリオド (終止符) で区切られる。上記の形態素解析によって文書を句点、あるいはピリオドごとに文に分割して、当該形態素解析の結果に対する上記の機械学習をもとに文章ベクトルを生成する。文書を段落ごとに分割して、段落単位に文章ベクトルを生成してもよい。

【 0 0 2 1 】

ベクトル生成部 2 3 で生成された単語ベクトル、文章ベクトル、および文書ベクトルは、関連度学習部 2 5 に送信される。関連度学習部 2 5 は、単語ベクトル学習部 2 5 a、単語 - 文章間学習部 2 5 b、および単語 - 文書間学習部 2 5 c を含む。

【 0 0 2 2 】

単語ベクトル学習部 2 5 a は、単語のベクトル空間におけるベクトル同士の距離が単語間の意味的な類似度を表すことから、ベクトル空間における単語間の意味的な関連度、および単語間のベクトル距離を計算する。同様に単語 - 文章間学習部 2 5 b は、ベクトル空間における単語と文章間の関連度、およびベクトル距離を計算し、単語 - 文書間学習部 2 5 c は、ベクトル空間における単語と文書間の関連度、およびベクトル距離を計算する。なお、2 つのベクトル表現の間の距離としては、たとえば、コサイン距離 (コサイン類似度) を挙げることができる。2 つのベクトル表現の間のコサイン距離が小さいほど、2 つのベクトルによってそれぞれ表現される 2 つの言語単位の意味は近い。

【 0 0 2 3 】

10

20

30

40

50

単語ベクトル学習部 25 a、単語 - 文章間学習部 25 b、および単語 - 文書間学習部 25 c での計算結果は、単語、文章、および文書各々を座標軸とする多次元ベクトル空間を有するデータベース 30 に、学習データとともに、単語ベクトル、単語 - 文章ベクトル、および単語 - 文書ベクトルとして格納される。

【0024】

なお、コーパス 27 において、たとえば分析事例のような特定分野を中心としたデータを蓄積することで、その分野により合致したベクトル表現が得られる。その結果、特定分野の文章等の検索精度を向上させることができる。また、コーパス 27 は、社内報告書、技報、アプリケーションニュース等を格納した内部コーパスと、ウィキペディア（登録商標）等のウェブ上で外部に公開されているデータ等を集めた外部コーパスとから構成されてもよい。外部コーパスはベクトル表現の学習の向上を図るものであるため、外部コーパスを検索対象から除外することにより、検索速度の低下を防止することができる。

10

【0025】

図 4 は、図 2 の学習処理部 20 によって行われる学習処理の流れを説明するためのフローチャートである。以下ではステップを単に S と記載する。図 4 に示されるように、形態素解析部 21 は、S 11 において、コーパス 27 に格納されている文書データ（学習データ）を既存の辞書を使用した形態素解析により複数の単語に分割する。

【0026】

ベクトル生成部 23 は、S 11 に続く S 13 において、S 11 における形態素解析の結果をもとに、単語の意味表現である単語ベクトルを機械学習によって生成するとともに、文章の特徴を表す文章ベクトル、および文書の特徴ベクトルである文書ベクトルを文書中に含まれている単語の特徴量の総和等から生成する。

20

【0027】

単語 - 文章間学習部 25 b は、S 13 に続く S 15 において、ベクトル空間における単語と文章との間の関連度、およびベクトル距離を計算する。単語 - 文書間学習部 25 c は、S 15 に続く S 17 において、ベクトル空間における単語と文書間の関連度、およびベクトル距離を計算する。

【0028】

関連度学習部 25 は、S 17 に続く S 19 において、学習データとして使用したコーパス 27 の文書データとともに、S 13、S 15、S 17 での計算結果を、単語ベクトル、単語 - 文章ベクトル、および単語 - 文書ベクトルとしてデータベース 30 に格納する。

30

【0029】

（検索処理部 40 の構成）

再び図 2 を参照して、検索処理部 40 は、入力部 1、解析部 11、特徴抽出部 13、検索部 15、表示制御部 17、および出力部 5 を含む。検索処理部 40 において、入力部 1 にはユーザによって検索クエリが入力される。検索クエリは、たとえば、分析関連検索キーワード、分析関連化合物名、分析関連分析対象物名を含む。入力部 1 は、キーボード KB 1、およびマウス MS 1 を含む。出力部 5 は、ディスプレイ 60 を含む。

【0030】

解析部 11 は、入力部 1 に入力された検索クエリに対して、予め定義された検索辞書をもとに形態素解析を行って、検索クエリを単語に分割する。特徴抽出部 13 は、学習処理部 20 によって生成されたベクトル空間を用いて、検索クエリのベクトル表現を算出する。

40

【0031】

検索部 15 は、特徴抽出部 13 から取得した検索クエリのベクトル表現を用いて、データベース 30 から検索クエリに関連する関連文書を検索する。検索部 15 は、データベース 30 から検索クエリのベクトル表現との距離が閾値よりも小さい関連文書を検索する。検索部 15 は、検索クエリのベクトル表現と、検索された複数の関連文書の各々についてのベクトル表現との距離が短いほど当該関連文書を上位に順位付けた検索結果を表示制御部 17 に出力する。表示制御部 17 は、検索部 15 による順位付けの順に関連文書が表示されるように出力部 5 を制御する。出力部 5 は、表示制御部 17 による制御結果に応じた

50

情報をディスプレイ 60 に表示する。

【0032】

図5は、図2の検索処理部40によって行われる検索処理の流れを説明するためのフローチャートである。図5に示されるように、入力部1は、S21において、ユーザによって入力されたより検索クエリを受け付ける。解析部11は、S21に続くS23において、検索クエリに対して形態素解析を行って、検索クエリを最小単位の形態素(単語)に分割する。特徴抽出部13は、S23に続くS25において、検索クエリの形態素解析の結果および学習処理部によって生成されたベクトル空間を用いて、検索クエリのベクトル表現を算出する。

【0033】

検索部15は、S25に続くS27において、コーパス27に対する学習によってベクトル化された学習データ等が蓄積されたデータベース30から、検索クエリと関連がある関連文書を検索する。S27においては、検索結果として、検索クエリに関連のある、あるいは関連性の高い文書が検索される。検索クエリに関連の高い文書とは、あらかじめ単語と文書間の関連を計算して得た、ベクトル空間における単語と文書間の関連度が高く、ベクトル距離が短い文書である。検索部15は、検索された複数の関連文書をベクトル距離が小さい順に順位付けする。

【0034】

表示制御部17は、S27に続くS29において、検索された複数の関連文書を検索部15の順位付けに基づいて出力部5に表示する。ユーザは、検索された少なくとも1つの関連文書の各々と検索クエリとの関連性の順位に従って閲覧する関連文書を決定することができる。

【0035】

図6は、図2の表示制御部17によってディスプレイ60に表示される検索結果ウィンドウWn2の一例を示す図である。図6に示されるように、検索結果ウィンドウWn2には、関連文書D1~D4が順位1~4とともに順番にそれぞれ表示されている。関連文書D1~D4の各々にはハイパーリンクが設定されている。図6において、データベース30から検索された複数の関連文書のうち、「臭素酸」という検索クエリに最も関連のある関連文書はD1である。

【0036】

再び図5も参照して、図6においてユーザがカーソルCrを操作して関連文書D1のハイパーリンクを選択した場合、表示制御部17は、S29に続くS31において、選択されたハイパーリンクに対応する関連文書の内容を表示する。図7は、図6の関連文書D1が選択された場合に検索結果ウィンドウWn2に表示される関連文書D1の内容を示す図である。

【0037】

図7に示されるように、表示制御部17は、検索クエリのベクトル表現と、関連文書D1に含まれる複数の単語の各々について、当該単語のベクトル表現と検索クエリのベクトル表現との距離に応じて、検索結果ウィンドウWn2における当該単語の周辺領域の色を変更することにより、当該単語をハイライトする。関連文書D1に含まれる単語のベクトル表現と検索クエリのベクトル表現との距離は、距離Ds3より大きい範囲R1、距離Ds2(<Ds3)より大きく距離Ds3以下の範囲R2、距離Ds1(<Ds2)より大きく距離Ds2以下の範囲R3、距離Ds1以下の範囲R4の4段階に分けられている。カラーマップCM1において、範囲R1~R4にそれぞれ互いに異なる色CL1~CL4が割り当てられている。なお、各単語と検索クエリとの関連性は、カラーマップCM1において連続的な色変化(グラデーション)として表示されてもよい。

【0038】

色CL2でハイライトされた単語WD5と検索クエリとの関連性は、色CL1でハイライトされた単語(不図示)と検索クエリとの関連性よりも大きい。色CL3でハイライトされた単語WD2, WD4, WD7と検索クエリとの関連性は、単語WD5と検索クエリ

10

20

30

40

50

との関連性よりも大きい。色 C L 4 でハイライトされた単語 W D 1 , W D 3 , W D 6 , W D 8 と検索クエリとの関連性は、単語 W D 2 , W D 4 , W D 7 と検索クエリとの関連性よりも大きい。

【 0 0 3 9 】

再び図 5 も参照して、表示制御部 1 7 は、S 3 1 に続く S 3 3 において、検索結果ウィンドウにおいてハイライトされている言語単位に対して選択操作（たとえばマウス操作によるダブルクリック）がされたか否かを判定する。ハイライトされている言語単位に対して選択操作がされた場合（S 3 3 において Y E S ）、表示制御部 1 7 は、選択された言語単位を S 3 5 において検索クエリに設定して、処理を S 2 3 に戻す。たとえば、図 7 の単語 W D 2 の周辺領域にカーソル C r が重なっている状態において、ユーザがマウス M S 1 をダブルクリックした場合、単語 W D 2 が検索クエリに設定されて、図 5 の S 2 3 からの検索処理が開始される。ユーザは今回の検索クエリとの関連性に着目して、新たな検索クエリを決定することができる。

10

【 0 0 4 0 】

ハイライトされている言語単位に対する選択操作がされていない場合（S 3 3 において N O ）、表示制御部 1 7 は、S 3 7 において検索結果ウィンドウの閉止操作がされたか否かを判定する。検索結果ウィンドウの閉止操作（たとえば図 7 のボタン B n 3 の押下）がされていない場合（S 3 7 において N O ）、表示制御部 1 7 は、処理を S 3 3 に戻す。検索結果ウィンドウの閉止操作がされた場合（S 3 7 において Y E S ）、表示制御部 1 7 は、処理を終了する。

20

【 0 0 4 1 】

図 7 に示される検索結果ウィンドウ W n 2 において、ユーザは、関連文書 D 1 に含まれる複数の単語の各々について、検索クエリとの関連性をハイライト色の違いとして確認することができる。関連文書 D 1 に検索クエリと一致する文字列が含まれていない場合でも、ユーザは、関連文書 D 1 が検索された根拠を視覚的に把握することができる。また、複数の単語の各々のベクトル表現と検索クエリのベクトル表現との距離に応じて当該単語のハイライト色を変更することにより、検索された関連文書に含まれる複数の単語の各々と検索クエリとの関連性を、当該単語以外の他の単語に依存しない直接的な関連性としてユーザに示すことができる。

【 0 0 4 2 】

図 7 においては、関連文書に含まれる複数の単語の各々について、当該単語と検索クエリとの関連性を、当該単語のベクトル表現と検索クエリのベクトル表現との距離として表現する場合について説明した。当該単語と検索クエリとの関連性は、当該関連文書のベクトル表現と検索クエリのベクトル表現との距離への当該単語の寄与度として表現されてもよい。

30

【 0 0 4 3 】

データベース 3 0 に格納されている文書 W_i は、以下の式 (1) のように単語 $w d_{i,k}$ の集合として表現される。なお、自然数 i は、1 から自然数 $D (> 1)$ までのいずれかの自然数である。自然数 k, t の各々は、1 から自然数 $N (> 1)$ までのいずれかの自然数である。

40

【 0 0 4 4 】

【 数 1 】

数 1

$$W_i = \{w d_{i,1}, w d_{i,2}, \dots, w d_{i,t}, \dots, w d_{i,N}\} \quad \dots (1)$$

【 0 0 4 5 】

データベース 3 0 に格納されている複数の文書は、以下の式 (2) のように文章集合 W として表現される。

【 0 0 4 6 】

50

【数 2】

数 2

$$W = \{W_1, W_2, \dots, W_i, \dots, W_D\} \quad \dots (2)$$

【0047】

文書 W_i と検索クエリ Q との距離 Ds_i は、以下の式 (3) のように表現される。

【0048】

【数 3】

数 3

$$Ds_i = f(W_i) - f(Q) \quad \dots (3)$$

10

【0049】

式 (3) における関数 f は、引数のベクトル表現を返す関数である。関数 f としては、たとえば、Doc2Vec、K-hot ベクトル、One-hot ベクトルの線形結合、単語の数え上げによるベクトル表現、およびトピックモデルを挙げることができる。

【0050】

式 (1) の文書 W_i に含まれる単語 $w_{d_i, t}$ の距離 Ds_i への寄与度を求めるために、文書 W_i から単語 $w_{d_i, t}$ を削除した文書 $W_{i, /t}$ を以下の式 (4) のように定義する。

20

【0051】

【数 4】

数 4

$$W_{i, /t} = \{wd_{i,1}, wd_{i,2}, \dots, wd_{i,N}\} \quad \dots (4)$$

【0052】

文書 $W_{i, /t}$ と検索クエリ Q との距離 $Ds_{i, /t}$ は、以下の式 (5) のように表現される。

【0053】

【数 5】

数 5

$$Ds_{i, /t} = f(W_{i, /t}) - f(Q) \quad \dots (5)$$

30

【0054】

検索クエリ Q と文書 W_i との関連性への単語 $w_{d_i, t}$ の寄与度が大きい程、検索クエリ Q と文書 $W_{i, /t}$ との関連性は小さくなる。すなわち、検索クエリ Q と文書 W_i との関連性への単語 $w_{d_i, t}$ の寄与度が大きい程、距離 $Ds_{i, /t}$ が大きくなる。その結果、距離 $Ds_{i, /t}$ と距離 Ds_i との差が大きくなる。単語 $w_{d_i, t}$ の寄与度 $Cn_{i, n}$ は、以下の式 (6) のように表される。

40

【0055】

【数 6】

数 6

$$Cn_{i, t} = Ds_{i, /t} - Ds_i \quad \dots (6)$$

【0056】

図 8 は、関連文書 D_1 に含まれる単語と検索クエリとの関連性として当該単語の寄与度を用いて、関連文書 D_1 に含まれる複数の単語の各々をハイライトした様子を示す図であ

50

る。図 8 の検索結果ウィンドウ W n 2 の内容は、図 7 の検索結果ウィンドウ W n 2 のカラーマップ C M 1 がカラーマップ C M 2 に置き換えられた内容である。これ以外は同様であるため、説明を繰り返さない。

【 0 0 5 7 】

図 8 に示されるように、表示制御部 1 7 は、関連文書 D 1 に含まれる複数の単語の各々について、当該単語の寄与度に応じて、検索結果ウィンドウ W n 2 における当該単語の周辺領域の色を変更し、当該単語をハイライトする。関連文書 D 1 に含まれる単語の寄与度は、寄与度 C n 1 より小さい範囲 R 1 1、寄与度 C n 1 以上であって寄与度 C n 2 (> C n 1) より小さい範囲 R 1 2、寄与度 C n 2 以上であって寄与度 C n 3 (> C n 2) より小さい範囲 R 1 3、および寄与度 C n 3 以上の範囲 R 1 4 の 4 段階に分けられている。カラーマップ C M 2 において、範囲 R 1 1 ~ R 1 4 にそれぞれ互いに異なる色 C L 1 ~ C L 4 が割り当てられている。なお、単語の寄与度は、カラーマップ C M 2 において連続的な色変化 (グラデーション) として表示されてもよい。

10

【 0 0 5 8 】

検索された関連文書に含まれる複数の単語の各々の寄与度に応じて当該単語のハイライト色を変更することにより、当該関連文書に含まれる複数の言語単位の各々と検索クエリとの関連性を、当該言語単位以外の他の言語単位と検索クエリとの関連性が反映された総合的な関連性としてユーザに示すことができる。

【 0 0 5 9 】

図 7 および図 8 においては、検索結果ウィンドウにおいてハイライトされる言語単位が単語である場合について説明した。ハイライトされる言語単位は、単語以外であってもよい。言語単位の種類を変化させることにより、複数の観点から関連文書が検索された根拠をユーザに示すことができる。

20

【 0 0 6 0 】

図 9 は、ハイライトされる言語単位が文章である場合の検索結果ウィンドウ W n 2 に表示される関連文書 D 1 の内容を示す図である。図 9 に示されるカラーマップ C M 2 は、文章の寄与度の分布を示す。

【 0 0 6 1 】

図 9 に示されるように、色 C L 2 でハイライトされた文章 S T 4 と検索クエリとの関連性は、色 C L 1 でハイライトされた文章 (不図示) と検索クエリとの関連性よりも大きい。色 C L 3 でハイライトされた文章 S T 3 , S T 6 と検索クエリとの関連性は、文章 S T 4 と検索クエリとの関連性よりも大きい。色 C L 4 でハイライトされた文章 S T 1 , S T 2 , S T 5 , S T 7 と検索クエリとの関連性は、文章 S T 3 , S T 6 と検索クエリとの関連性よりも大きい。

30

【 0 0 6 2 】

図 1 0 は、ハイライトされる言語単位が段落である場合の検索結果ウィンドウ W n 2 に表示される関連文書 D 1 の内容を示す図である。図 1 0 に示されるカラーマップ C M 2 は、段落の寄与度の分布を示す。

【 0 0 6 3 】

図 1 0 に示されるように、色 C L 2 でハイライトされた段落 P R 3 と検索クエリとの関連性は、色 C L 1 でハイライトされた段落 (不図示) と検索クエリとの関連性よりも大きい。色 C L 3 でハイライトされた段落 P R 1 と検索クエリとの関連性は、段落 P R 3 と検索クエリとの関連性よりも大きい。色 C L 4 でハイライトされた段落 P R 2 と検索クエリとの関連性は、段落 P R 1 と検索クエリとの関連性よりも大きい。

40

【 0 0 6 4 】

なお、検索結果ウィンドウ W n 2 においてハイライトされる言語単位は 1 種類である必要はなく、文字、形態素、単語、文章、段落、およびこれらの任意の組み合わせからなる群から選択されてもよい。たとえば、文字および形態素がハイライトされる言語単位とされてもよいし、単語、文章、および段落がハイライトされる言語単位とされてもよい。

【 0 0 6 5 】

50

図 1 1 は、実施の形態に係る関連文書を検索して表示するシステムの一例である分析事例検索システム 1 0 0 A がネットワーク 7 0 を介して複数のクライアント端末 8 0 a ~ 8 0 n に接続された場合の、分析事例検索システム 1 0 0 A の構成を示す機能ブロック図である。図 1 1 に示されるように、分析事例検索システム 1 0 0 A は、たとえばインターネット環境において、ユーザ等からの要求に応じて分析事例の検索を可能にし、検索結果をユーザ等に提供する。

【 0 0 6 6 】

分析事例検索システム 1 0 0 A と、複数のクライアント端末 8 0 a ~ 8 0 n とが、インターネット等の情報通信のネットワーク 7 0 を介して通信可能に接続されている。分析事例検索システム 1 0 0 A および複数のクライアント端末 8 0 a ~ 8 0 n は、クライアントサーバシステム 1 0 0 0 を構成する。分析事例検索システム 1 0 0 A 内の通信部 6 1 は、ネットワーク 7 0 とのインターフェイスである。制御部 6 5 は、CPU (Central Processing Unit) を含み、学習処理部 2 0 および検索処理部 4 0 を含む分析事例検索システム 1 0 0 A の全体の制御を司る。

10

【 0 0 6 7 】

メモリ 6 7 には、上述した学習処理部 2 0 における学習処理プログラム、および検索処理部 4 0 による検索処理プログラム等が格納されている。制御部 6 5 は、メモリ 6 7 からこれらのプログラムを読み出して、図 4 および図 5 に示される所定の処理等を実行する。なお、分析事例検索システム 1 0 0 A を、ネットワーク 7 0 に接続されたサーバ装置と位置づけることもできる。すなわち、分析事例検索システム 1 0 0 A においては、検索処理部 4 0 による表示処理 (表示工程) がサーバサイドにおいて行われる。既存のクライアント端末をサーバ装置に接続することにより、当該クライアント端末を介して関連文書が検索された根拠をユーザに示すことができる。

20

【 0 0 6 8 】

検索処理部 4 0 による表示処理は、複数のクライアント端末 8 0 a ~ 8 0 n (クライアントサイド) において行われてもよい。クライアント端末を既存のサーバ装置に接続することにより、当該クライアント端末を介して関連文書が検索された根拠をユーザに示すことができる。

【 0 0 6 9 】

以上、実施の形態に係る関連文書を検索して表示する方法およびシステムによれば、検索結果の根拠をユーザに示すことができる。

30

【 0 0 7 0 】

[態 様]

上述した複数の例示的な実施の形態は、以下の態様の具体例であることが当業者により理解される。

【 0 0 7 1 】

(第 1 項) 一態様に係る関連文書を検索して表示する方法は、複数の文書を含むデータベースから検索クエリに関連する少なくとも 1 つの関連文書を検索して表示する。当該方法は、検索工程と、表示工程とを含む。検索工程は、任意の言語単位をベクトル表現に変換するベクトル空間を用いて複数の文書の各々についてのベクトル表現と検索クエリのベクトル表現との距離を算出し、距離に応じて少なくとも 1 つの関連文書を検索する。表示工程は、少なくとも 1 つの関連文書の各々を表示する。表示工程は、当該関連文書に含まれる複数の言語単位の各々のベクトル表現と検索クエリのベクトル表現とに基づく当該言語単位と検索クエリとの関連性の大きさに応じた表示態様で、当該言語単位を表示する工程を含む。

40

【 0 0 7 2 】

第 1 項に記載の方法によれば、当該関連文書に含まれる複数の言語単位の各々のベクトル表現と検索クエリのベクトル表現とに基づく当該言語単位と検索クエリとの関連性の大きさに応じた表示態様で、当該言語単位を表示することにより、検索結果の根拠をユーザに示すことができる。

50

【 0 0 7 3 】

(第2項)第1項に記載の方法において、当該言語単位の表示態様は、当該言語単位の周辺領域の色を含む。

【 0 0 7 4 】

第2項に記載の方法によれば、検索された関連文書に含まれる複数の言語単位の各々と検索クエリとの関連性が、当該言語単位の周辺領域の色の違いとして視覚的に把握することができる。

【 0 0 7 5 】

(第3項)第1項または第2項に記載の方法において、当該言語単位と検索クエリとの関連性の大きさは、複数の言語単位の各々のベクトル表現と検索クエリのベクトル表現との距離である。

10

【 0 0 7 6 】

第3項に記載の方法によれば、検索された関連文書に含まれる複数の言語単位の各々と検索クエリとの関連性を、当該言語単位以外の他の言語単位に依存しない直接的な関連性としてユーザに示すことができる。

【 0 0 7 7 】

(第4項)第1項または第2項に記載の方法において、当該言語単位と検索クエリとの関連性の大きさは、複数の言語単位の各々の寄与度である。寄与度は、少なくとも1つの関連文書の各々から当該言語単位を除いたベクトル表現と検索クエリのベクトル表現との距離から、当該関連文書のベクトル表現と検索クエリのベクトル表現との距離を減算した値である。

20

【 0 0 7 8 】

第4項に記載の方法によれば、検索された関連文書に含まれる複数の言語単位の各々と検索クエリとの関連性を、当該言語単位以外の他の言語単位と検索クエリとの関連性が反映された総合的な関連性としてユーザに示すことができる。

【 0 0 7 9 】

(第5項)第1項に記載の方法において、検索工程は、検索クエリのベクトル表現と少なくとも1つの関連文書の各々についてのベクトル表現との距離が短いほど、当該文書を上位に順位付ける工程を含む。表示工程は、検索工程による順位付けに従って少なくとも1つの関連文書を表示する工程を含む。

30

【 0 0 8 0 】

第5項に記載の方法によれば、ユーザは、検索された少なくとも1つの関連文書の各々と検索クエリとの関連性の順位に従って、閲覧する関連文書を決定することができる。

【 0 0 8 1 】

(第6項)第1項に記載の方法において、複数の言語単位の種類は、文字、形態素、単語、文章、段落、およびこれらの任意の組み合わせからなる群から選択される。

【 0 0 8 2 】

第6項に記載の方法によれば、複数の言語単位の種類を変化させることにより、複数の観点から関連文書が検索された根拠をユーザに示すことができる。

【 0 0 8 3 】

(第7項)第1項に記載の方法は、クライアントサーバシステムにおいて行われる。表示工程は、サーバサイドにおいて行われる。

40

【 0 0 8 4 】

第7項に記載の方法によれば、既存のクライアント端末をサーバ装置に接続することにより、当該クライアント端末を介して関連文書が検索された根拠をユーザに示すことができる。

【 0 0 8 5 】

(第8項)第1項に記載の方法は、クライアントサーバシステムにおいて行われる。表示工程は、クライアントサイドにおいて行われる。

【 0 0 8 6 】

50

第 8 項に記載の方法によれば、クライアント端末を既存のサーバ装置に接続することにより、当該クライアント端末を介して関連文書が検索された根拠をユーザに示すことができる。

【 0 0 8 7 】

(第 9 項) 第 1 項に記載の方法において、検索工程は、表示されている複数の言語単位に含まれる或る言語単位がユーザによって選択された場合、当該言語単位を検索クエリとして少なくとも 1 つの関連文書を検索する工程を含む。

【 0 0 8 8 】

第 9 項に記載の方法によれば、ユーザは今回の検索クエリとの関連性に着目して、新たな検索クエリを決定することができる。

【 0 0 8 9 】

(第 10 項) 第 1 項に記載の方法において、ベクトル空間は、データベースを含むコーパスに対して自然言語処理を行う機械学習によって生成される。

【 0 0 9 0 】

第 10 項に記載の方法によれば、コーパスを用いたモデルに対する機械学習の過程で、コーパスの特徴が高精度に反映されたベクトル空間を生成することができる。

【 0 0 9 1 】

(第 11 項) 第 1 項に記載の関連文書を検索して表示するシステムは、複数の文書から検索クエリに関連する少なくとも 1 つの関連文書を検索して表示する。当該システムは、データベースと、検索処理部とを含む。データベースは、複数の文書を含む。検索処理部は、任意の言語単位をベクトル表現に変換するベクトル空間を用いて複数の文書の各々についてのベクトル表現と検索クエリのベクトル表現との距離を算出し、距離に応じて少なくとも 1 つの関連文書を検索する。検索処理部は、少なくとも 1 つの関連文書の各々を表示する。検索処理部は、当該関連文書に含まれる複数の言語単位の各々のベクトル表現と検索クエリのベクトル表現とに基づく当該言語単位と検索クエリとの関連性の大きさに応じた表示態様で、当該言語単位を表示する。

【 0 0 9 2 】

第 11 項に記載のシステムによれば、関連文書に含まれる複数の言語単位の各々のベクトル表現と検索クエリのベクトル表現とに基づく当該言語単位と検索クエリとの関連性の大きさに応じた表示態様で、当該言語単位を表示することにより、検索結果の根拠をユーザに示すことができる。

【 0 0 9 3 】

なお、上述した実施の形態および変更例について、明細書内で言及されていない組み合わせを含めて、不都合または矛盾が生じない範囲内で、実施の形態で説明された構成を適宜組み合わせることは出願当初から予定されている。

【 0 0 9 4 】

今回開示された実施の形態はすべての点で例示であって制限的なものではないと考えられるべきである。本発明の範囲は上記した説明ではなくて請求の範囲によって示され、請求の範囲と均等の意味および範囲内でのすべての変更が含まれることが意図される。

【符号の説明】

【 0 0 9 5 】

1 入力部、5 出力部、10 計算機、11 解析部、13 特徴抽出部、15 検索部、17 表示制御部、20 学習処理部、21 形態素解析部、23 ベクトル生成部、25 関連度学習部、25 a 単語ベクトル学習部、25 b 単語 - 文章間学習部、25 c 単語 - 文書間学習部、27 コーパス、30 データベース、40 検索処理部、60 ディスプレイ、61 通信部、65 制御部、67 メモリ、70 ネットワーク、80 a ~ 80 n クライアント端末、100, 100 A 分析事例検索システム、1000 クライアントサーバシステム、KB1 キーボード、MS1 マウス。

10

20

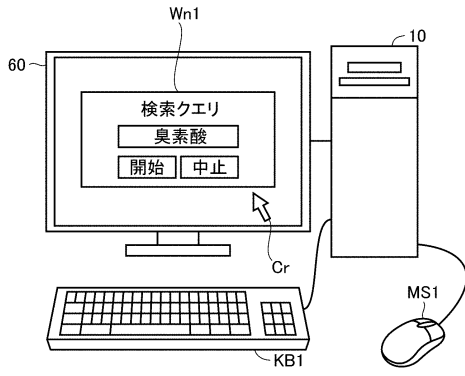
30

40

50

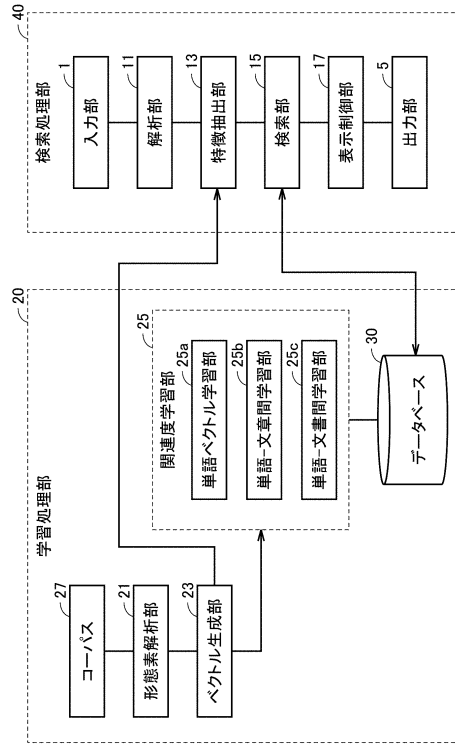
【図面】
【図 1】

FIG.1
100



【図 2】

FIG.2
100



10

20

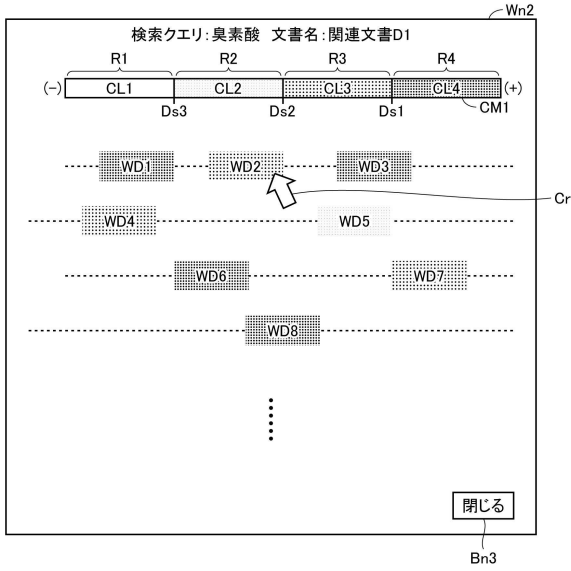
30

40

50

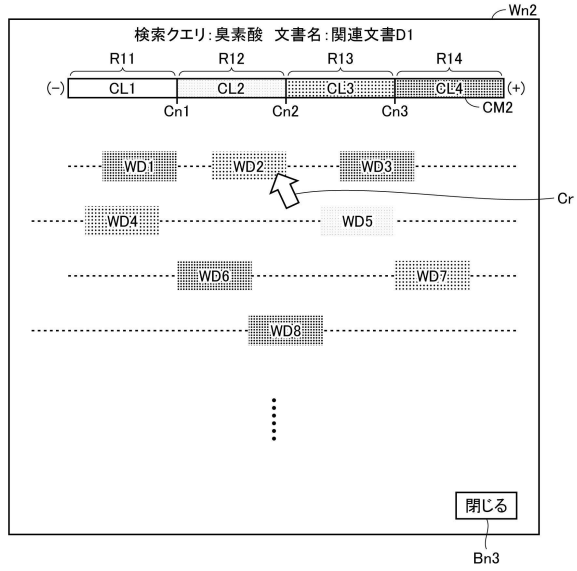
【 図 7 】

FIG.7



【 図 8 】

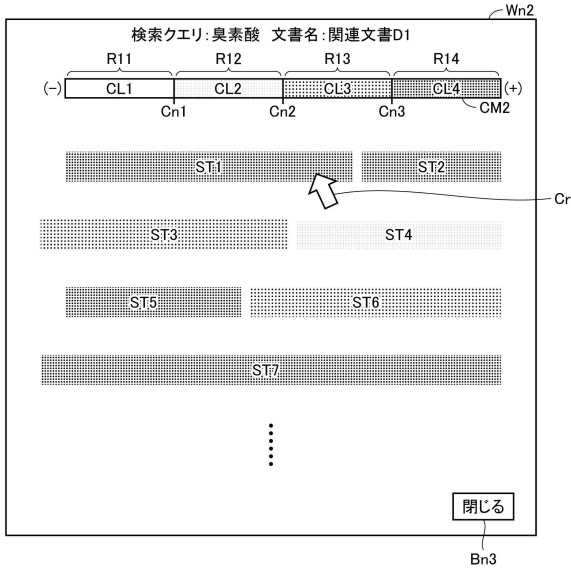
FIG.8



10

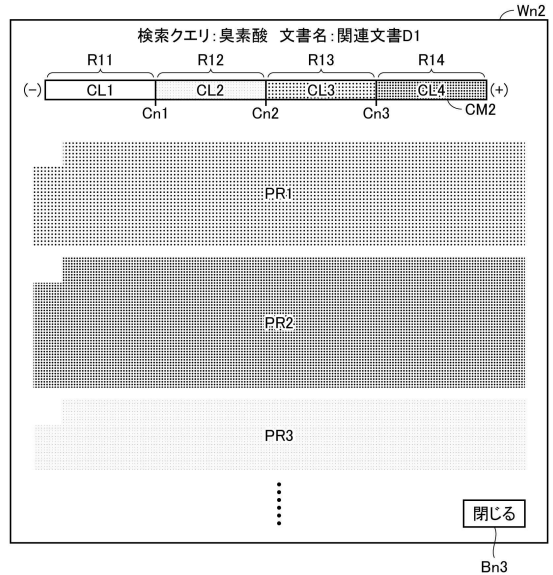
【 図 9 】

FIG.9



【 図 10 】

FIG.10



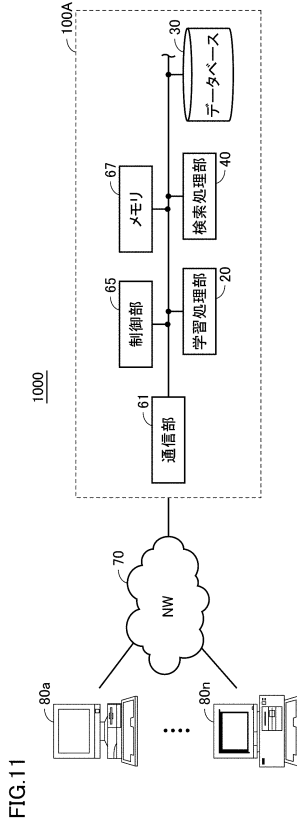
20

30

40

50

【図 11】



10

20

30

40

50

フロントページの続き

- (56)参考文献 特開平 1 1 - 0 4 5 2 5 4 (J P , A)
特開 2 0 1 4 - 2 1 1 8 7 0 (J P , A)
特開 2 0 0 2 - 2 2 2 2 1 0 (J P , A)
特開 2 0 0 5 - 0 9 2 2 5 3 (J P , A)
特開 2 0 0 4 - 3 3 4 3 4 1 (J P , A)
特開 2 0 1 7 - 2 0 1 4 7 8 (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