

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5837850号
(P5837850)

(45) 発行日 平成27年12月24日(2015.12.24)

(24) 登録日 平成27年11月13日(2015.11.13)

(51) Int.Cl.

G06F 17/30 (2006.01)

F 1

G06F 17/30 340Z
G06F 17/30 210A
G06F 17/30 380D

請求項の数 5 (全 12 頁)

(21) 出願番号 特願2012-68155 (P2012-68155)
 (22) 出願日 平成24年3月23日 (2012.3.23)
 (65) 公開番号 特開2012-226738 (P2012-226738A)
 (43) 公開日 平成24年11月15日 (2012.11.15)
 審査請求日 平成27年3月20日 (2015.3.20)
 (31) 優先権主張番号 13/089, 222
 (32) 優先日 平成23年4月18日 (2011.4.18)
 (33) 優先権主張国 米国(US)

早期審査対象出願

(73) 特許権者 502096543
 パロ・アルト・リサーチ・センター・イン
 コーポレーテッド
 Palo Alto Research
 Center Incorporated
 アメリカ合衆国、カリフォルニア州 94
 304、パロ・アルト、コヨーテ・ヒル・
 ロード 3333
 (74) 代理人 100079049
 弁理士 中島 淳
 (74) 代理人 100084995
 弁理士 加藤 和詳

最終頁に続く

(54) 【発明の名称】意味のある実体に基づき、導かれる、関連文書の検索方法

(57) 【特許請求の範囲】

【請求項 1】

受信機構及び検索機構を備える関連文書検索システムによって実行可能な、関連文書を検索する方法であって、

前記受信機構が、ソース文書を受信するステップと、

前記検索機構が、前記ソース文書から抽出された意味論的実体に基づいて、前記関連文書を検索するステップと、

前記受信機構が、前記ソース文書から抽出される意味論的実体の数を指示しつつユーザによって調整された実体抽出感知を含むと共に前記ユーザによって調整された検索基準を前記ユーザから受信するステップと、

前記関連文書検索システムが、前記検索基準に基づいて、前記関連文書の検索結果を更新するステップと、

を含む、方法。

【請求項 2】

前記検索基準は、

実体が前記関連文書の検索に含まれるか、あるいは、前記実体が前記関連文書の検索から除外されるかを指示する、前記実体に関する選択指示と、

前記関連文書の検索用の前記実体の重要性を指示する、前記実体に関する重みと、

前記検索結果に含まれる関連文書の数に応じた検索感知指示であって、より高く感知することによって前記検索結果により多くの関連文書が得られ、より低く感知することによ

つて前記検索結果により少ない関連文書が得られる前記検索感知指示と、
の少なくとも 1 つを含む、請求項 1 に記載の方法。

【請求項 3】

前記受信機構が、追加された 1 つ以上のソース文書を受信するステップと、
前記関連文書検索システムが、前記追加された 1 つ以上のソース文書から抽出された意味論的実体に基づいて、前記関連文書の検索結果を更新するステップと、
をさらに含む、請求項 1 に記載の方法。

【請求項 4】

前記関連文書検索システムが、前記ユーザによる 1 つ以上の付加的な意味論的実体の指定を可能にするステップと、 10

前記関連文書検索システムが、前記ユーザによって前記指定された前記意味論的実体に基づいて、前記関連文書の検索結果を更新するステップと、
をさらに含む、請求項 1 に記載の方法。

【請求項 5】

前記ユーザによって前記指定された前記意味論的実体は、一群の知られている実体から選択された関連する実体である、請求項 4 に記載の方法。

【発明の詳細な説明】

【発明の概要】

【0001】

一実施形態は、関連文書を検索するためのシステムを提供する。動作中に、システムはソース文書を受信する。システムは、その後、ソース文書から抽出された意味論的実体に基づいて、関連文書を検索する。次に、システムは、ユーザによって調整された検索基準を受信し、それに基づいて関連文書の検索結果を更新する。 20

【0002】

本実施形態の一変形例では、検索基準は、ソース文書から抽出された実体の数に応じた実体抽出の感知と、実体が関連文書の検索に含まれているか、あるいは、関連文書の検索から除外されているかを指示する実体に関する選択指示と、関連文書の検索用の実体の重要性を指示する実体に関する重みと、検索結果における関連文書の数に応じた検索感知指示であって、より高く感知することによって検索結果におけるより多くの関連文書が得られ、より低く感知することによって検索結果におけるより少ない関連文書が得られる検索感知指示との少なくとも 1 つを含んでいる。 30

【0003】

本実施形態の一変形例では、システムは、1 つ以上の追加されたソース文書を受信し、ソース文書から抽出された意味論的実体に基づいて関連文書の検索結果を更新する。

【0004】

本実施形態の一変形例では、システムは、ユーザによる 1 つ以上の意味論的実体の指定を可能にし、ユーザによって指定された意味論的実体に基づいて関連文書の検索結果を更新する。

【0005】

さらなる一変形例では、ユーザによって指定された意味論的実体は、一群の知られている実体から選択された適切な実体である。 40

【0006】

本実施形態の一変形例では、検索基準は、構成ファイルに保管される。

【0007】

本実施形態の一変形例では、システムは、ソース文書に関して他の文書の類似度レベルを計算することによって、関連文書を検索する。

【図面の簡単な説明】

【0008】

【図 1】 本発明の一実施形態による関連文書を検索するための例示的なグラフィカル・ユーザ・インターフェースを示す図である。 50

【図2】本発明の一実施形態による関連文書検索システムを示す図である。

【図3】本発明の一実施形態による関連文書の検索プロセスを示すフローチャートである。

【図4】本発明の一実施形態による関連文書の検索結果の更新プロセスを示すフローチャートである。

【図5】本発明の一実施形態による関連文書を検索するための例示的なコンピュータシステムを示す。図では、参照符号の同じものは、図の同じ要素を示している。

【発明を実施するための形態】

【0009】

本発明の実施形態は、ソース文書に関連した文書から抽出された意味論的実体に基づいて、対話型および反復型の検索の問題を解決する。動作中に、システムはソース文書を受信する。システムは、その後、ソース文書から抽出された意味論的実体に基づいて、関連文書を検索する。システムは、ユーザによる検索基準の調整を可能にし、ユーザによって調整された検索基準に基づいて関連文書の検索結果を更新する。

10

【0010】

関連したメッセージまたは会話を検索するために、本発明の実施形態は、文書における意味のある語または「実体」の生起を比較する計算方法を利用する。実体は、類似度計算用のソース文書から抽出される。ソース文書および実体の追加／削除と、抽出された実体の数の変更と、検索用の実体の個々の重みの変更とを含む検索基準を対話型および反復型に調整および改良するために、柔軟な枠組みがユーザに提供される。

20

【0011】

図1は、本発明の一実施形態による関連文書を検索するための例示的なグラフィカル・ユーザ・インターフェースを示す図である。この例では、この文書タイプは、電子メールメッセージである。GUI100は、ソース文書入力バー102と、ソース文書リスト104と、実体リスト106と、実体感知スライダ108と、検索感知スライダ110と、実体入力バー112と、関連文書リスト114、116とを含む。検索を開始するために、ユーザは、GUI100の上端に位置するソース文書入力バー102に1つ以上の文書を入力またはドラッグおよびドロップできる。ソース文書リスト104は、ユーザによって追加された現在のソース文書（2つの電子メール）を示している。

30

【0012】

意味論的実体は、その後、ソース文書から抽出されて、これらの実体の全てまたは多数の部分集合を含む関連文書の検索基盤を形成する。ソース文書から抽出された実体は、実体リスト106において視覚化される。実体を、原文どおり（例えば、人または会社の名称）または図形によって（例えば、人または会社ロゴの写真）のどちらか一方で示すことができる。個々の実体を、実体に関連する名称または写真を強調表示することによって、検索に含むことができるか、あるいは、検索から除外できる。ソース文書から抽出された実体の数を、ユーザが実体感知スライダ108を用いて制御できる。例えば、実体感知スライダ108を右にスライドされることにより、ユーザは、ソース文書からより多くの実体を抽出する。ユーザはまた、実体リスト106における実体名称または写真をサイズ変更することによって、検索用の個々の実体の重みを調整できる。

40

【0013】

付加的な実体の場合、ソース文書から抽出された、追加される必要のある実体に加えて、ユーザは、実体名称をタイプしてもよいし、あるいは、知られている実体の広域データベースから実体入力バー112に実体をドラッグおよびドロップしてもよい。ユーザによって追加された付加的な実体を、実体リスト106における抽出された実体と共に示し、続いて、検索に含むことができるか、あるいは、検索から除外できる。検索感知スライダ110は、ユーザが実体に基づいた関連文書の検索の感知を調整することを可能にする。より高く感知することによって、検索結果においてより多くの関連文書が得られ、より低く感知することによって、検索結果においてより少ない関連文書が得られる。GUI100では、検索結果は、関連文書リスト114、116に示される。リストは、関連した電

50

子メールメッセージがソース電子メールメッセージと同じ会話に属しているかに応じて分けられる。

【0014】

図2は、本発明の一実施形態による関連文書検索システムを示す図である。関連文書検索システム200は、受信機構202と、ユーザによって調整可能な多数の検索基準204～210と、検索機構212とを含む。動作中に、受信機構202は、関連文書を検索するためのソース文書220を受信する。ソース文書220は、テキスト、記号、図、および／または、それらの任意の組み合わせを含む、任意のタイプの文書であってもよい。ソース文書220はまた、平文、マイクロソフト（登録商標）社のワード、ハイパーテキストマークアップ言語（H T M L）、ポータブル・ドキュメント・フォーマット（P D F）、または、電子メールメッセージといった、任意のタイプのファイル形式であってもよい。10

【0015】

次に、ソース文書220に多数の検索基準204～210を適用し、意味論的実体を抽出する。意味論的実体は、語、語の組み合わせ、または、特定の意味をもつシーケンスであってもよい。語または語のシーケンスは、人々の名称、会社の名称、日付および時間、所在地住所、業界特有の用語、電子メールアドレス、統一資源位置指定子（U R L）、および、電話番号といった、特定の語グループに属していれば、意味論的実体と見なされる。実体感知204は、ソース文書から抽出された実体の数を制御する。実体が抽出されると、ユーザは、実体選択206を調整することによって実体を検索に含めるか、あるいは、検索から除外するかを決定できる。もう1つの実施形態では、ユーザによって指定された実体を、広域実体データベースといった外部ソースから追加してもよい。実体の重み208を、ユーザが調整でき、これにより、関連文書を検索するために各実体の重要性が示される。ユーザはまた、検索結果において生成された関連文書の数に相当する検索感知210を調整してもよい。ここで、より高く感知することによって、より多くの関連文書が得られ、より低く感知することによって、より少ない関連文書が得られる。20

【0016】

ソース文書220および検索基準204～210から抽出された実体に基づいて、検索機構212は、検索を実施し、関連文書の検索結果230を出力する。関連文書を識別するために、検索機構212は、実体および実体の重みの生起性および共起性に基づいて、様々な基準を用いて目標文書とソース文書との類似度を計算する。例えば、所定の閾値を越える類似度値は、目標文書がソース文書に関連していることを示している。30

【0017】

図3は、本発明の一実施形態による実体の抽出プロセスを示すフローチャートである。動作中に、システムは、ユーザからソース文書を受信する（動作302）。システムは、その後、ソース文書から抽出された意味論的実体に基づいて関連文書を検索する（動作304）。次に、システムは、ユーザによって調整された検索基準を受信する（動作306）。続いて、システムは、ユーザによって調整された検索基準に基づいて、関連文書の検索結果を更新する（動作308）。システムは、適宜、今後使用するために検索基準を構成ファイルに保管する（動作310）。40

【0018】

ユーザによって調整可能な検索基準は、ソース文書から抽出された実体の数に応じた実体抽出の感知と、実体を関連文書の検索に含むか、あるいは、検索から除外するかを指示する実体に関する選択指示と、関連文書を検索するための実体の重要性を指示する実体に関する重みと、検索結果における関連文書の数に応じた検索感知指示とを含んでいてもよい。より高く感知することによって、検索結果においてより多くの関連文書が得られ、より低く感知することによって、検索結果においてより少ない関連文書が得られる。

【0019】

さらに、関連文書検索システムは、ユーザによる1つ以上のソース文書の追加または削除を可能にする。関連文書の検索結果は、更新されたソース文書から抽出された意味論的50

実体に基づいて、更新される。システムはまた、ユーザによる1つ以上の意味論的実体の指定を可能にし、ユーザによって指定された意味論的実体に基づいて関連文書の検索結果を更新する。ユーザによって指定された意味論的実体は、必ずしもソース文書から抽出されるわけではない。つまり、実体は、知られている実体の広域データベースから選択された適切な実体であってもよい。実体構成の各修正、または、ソースメッセージ入力は、これらの新たに修正されたパラメータによって、新しい実体に基づいた類似検索をトリガする。検索結果は、その後、結果を絶え間なく改良できるユーザに示される。

【0020】

図4は、本発明の一実施形態による関連文書の検索結果の更新プロセスを示すフローチャートである。動作中に、システムは、1つ以上の追加されたソース文書を受信する（動作402）。次に、システムは、ユーザによって指定された1つ以上の意味論的実体を受信する（動作404）。システムは、その後、追加されたソース文書およびユーザによって指定された意味論的実体に基づいて、関連文書の検索結果を更新する（動作406）。

10

【0021】

システムはまた、ユーザが、知られている実体の広域データベースから付加的な意味論的実体を追加する必要があるとき、抽出された実体に最も適した実体をユーザに提案してもよい。ユーザが実体を選択し、および／または、検索するための実体の個々の重みを調整した後、システムは、適切な実体のリストを更新でき、ユーザによって指定された実体に対する実体の適合性に基づいて、実体をソートする。方法を用いてユーザによって細かく調整された文書検索を、その後使用するために構成ファイルに保管できる。保管された構成ファイルは、実体抽出感知、個々の実体の重み、および、検索感知といった、ユーザによってカスタマイズされた検索基準のペクトルと共に、ソース文書および実体のリストを含む。

20

【0022】

図5は、本発明の一実施形態による関連文書を検索するための例示的なコンピュータシステムを示している。一実施形態では、コンピュータおよび通信システム500は、プロセッサ502、メモリ504、および、記憶装置506を含む。記憶装置506は、文書検索アプリケーション508、および、アプリケーション510、512といった他のアプリケーションを記憶する。動作中に、文書検索アプリケーション508を、記憶装置506からメモリ504に読み込み、次に、プロセッサ502が実行する。プログラムを実行している間、プロセッサ502は、機能を実施する。コンピュータおよび通信システム500は、任意の表示装置514、キーボード516、および、ポインティングデバイス518に連結されている。

30

【0023】

本詳細な説明に記載したデータ構造およびコードは、通常、コンピュータ読取り可能記憶媒体に格納される。媒体は、コンピュータシステムによって用いられるためのコードおよび／またはデータを記憶できる任意のデバイスまたは媒体であってもよい。コンピュータ読取り可能記憶媒体は、揮発性メモリ、不揮発性メモリ、ディスクドライブといった磁気記憶装置および光学記憶装置、磁気テープ、CD（コンパクトディスク）、DVD（デジタル多用途ディスクまたはデジタルビデオディスク）、または、現在知られている、あるいは、後に開発されるコンピュータ読取り可能媒体を記憶できる他の媒体を含むが、それらに限定されるものではない。

40

【0024】

本詳細な説明部分に記載した方法およびプロセスを、上述したようなコンピュータ読取り可能記憶媒体に格納されうるコードおよび／またはデータとして具体化できる。コンピュータシステムが、コンピュータ読取り可能記憶媒体に格納されたコードおよび／またはデータを読み取り、実行するとき、コンピュータシステムは、データ構造およびコードとして具体化されてコンピュータ読取り可能記憶媒体内に格納された方法およびプロセスを実施する。

【0025】

50

さらに、本明細書において説明した方法およびプロセスを、ハードウェアモジュールまたはハードウェア機器に含むことができる。これらのモジュールまたは機器は、特定用途向け I C (A S I C) チップ、書替え可能ゲートアレイ (F P G A) 、特定の時間において特定のソフトウェアモジュールまたはコードの一部を実行する、専用プロセッサまたは共用プロセッサ、および / または、現在知られている、あるいは、後に開発される他のプログラマブルロジックデバイスを含んでいてもよいが、それらに限定されるものではない。ハードウェアモジュールまたはハードウェア機器をアクティブにするとき、モジュールまたは機器は、それらに含まれる方法およびプロセスを実施する。

【図1】

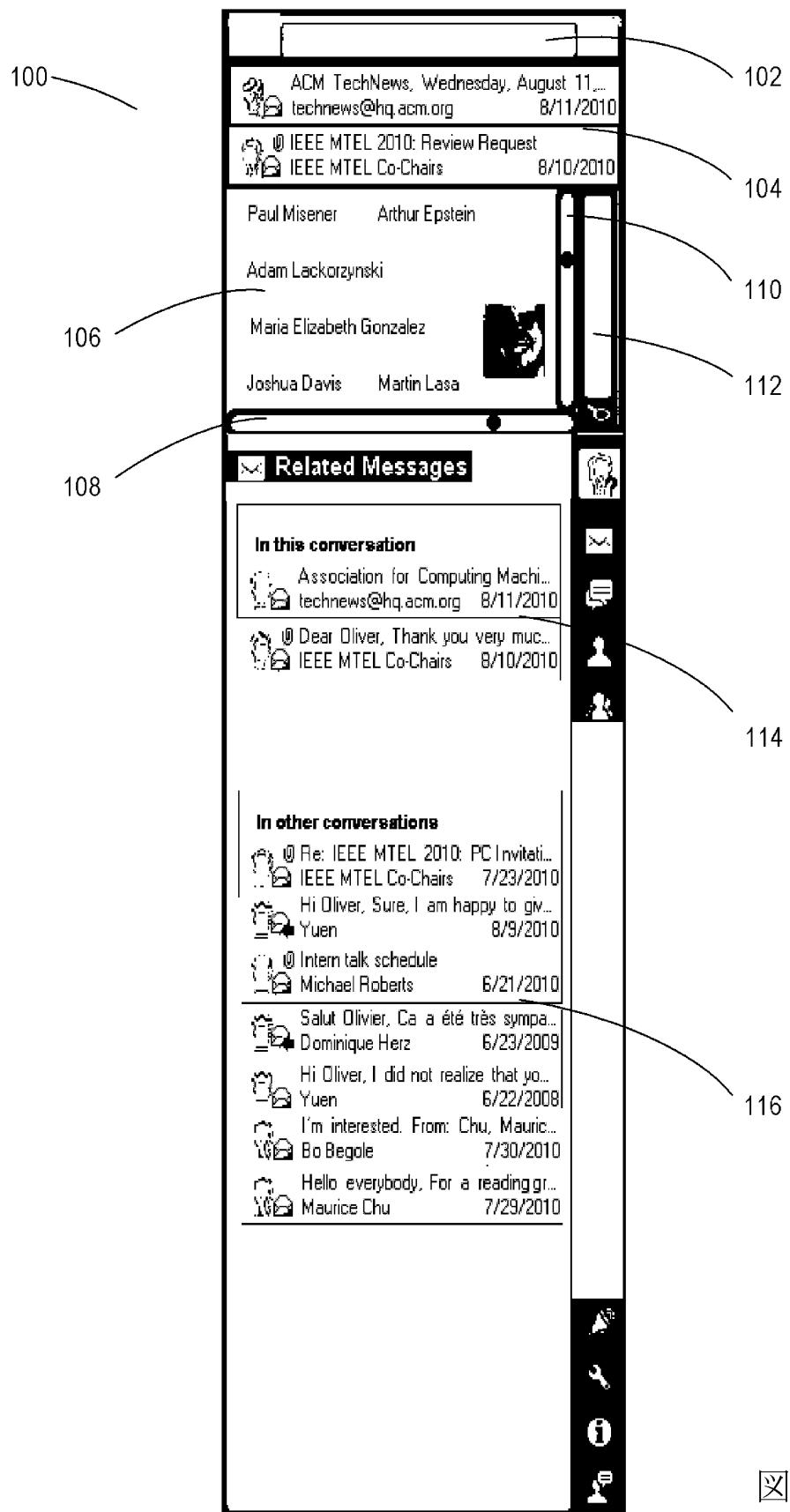


図1

【図2】

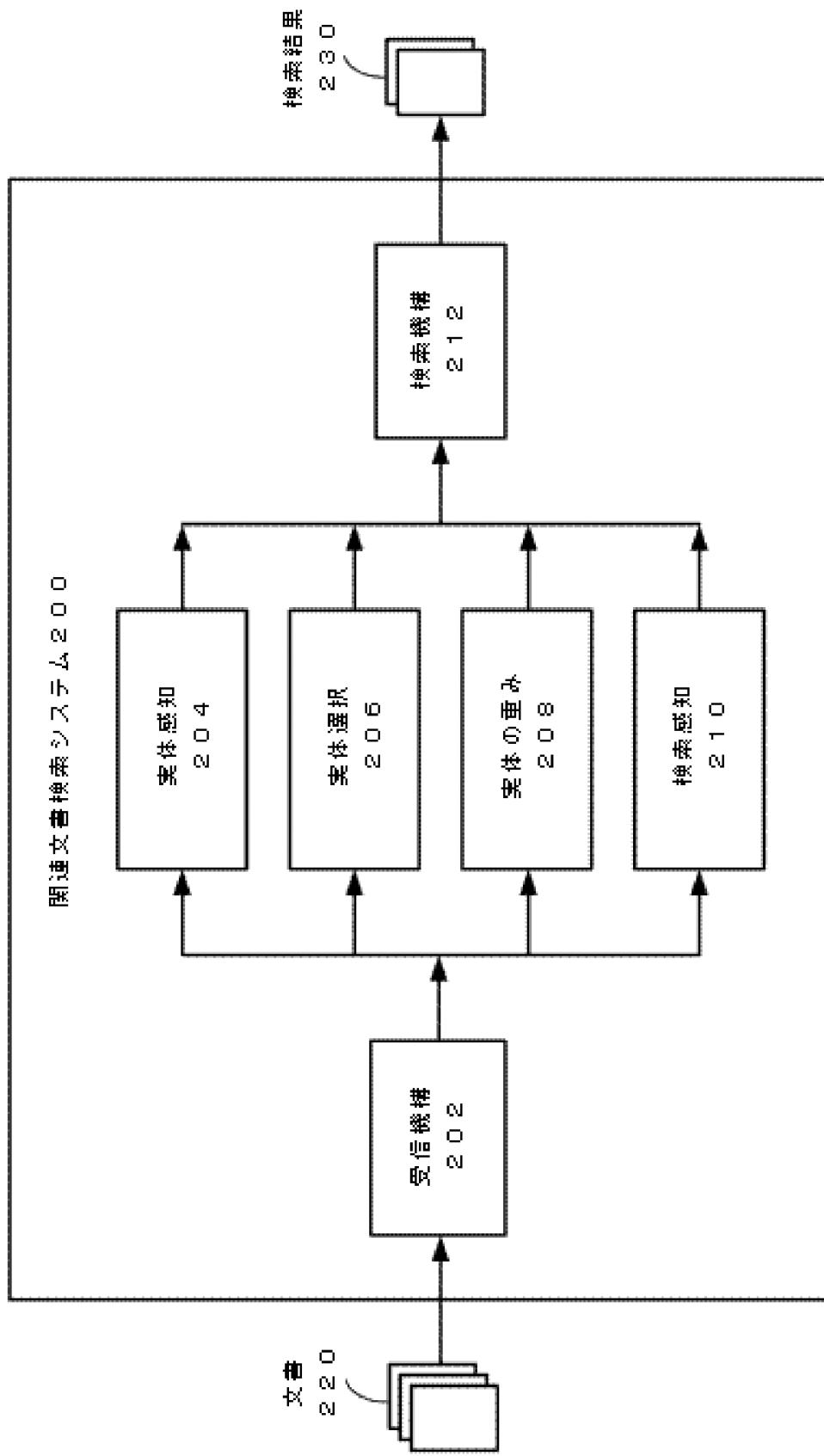


図2

【図3】

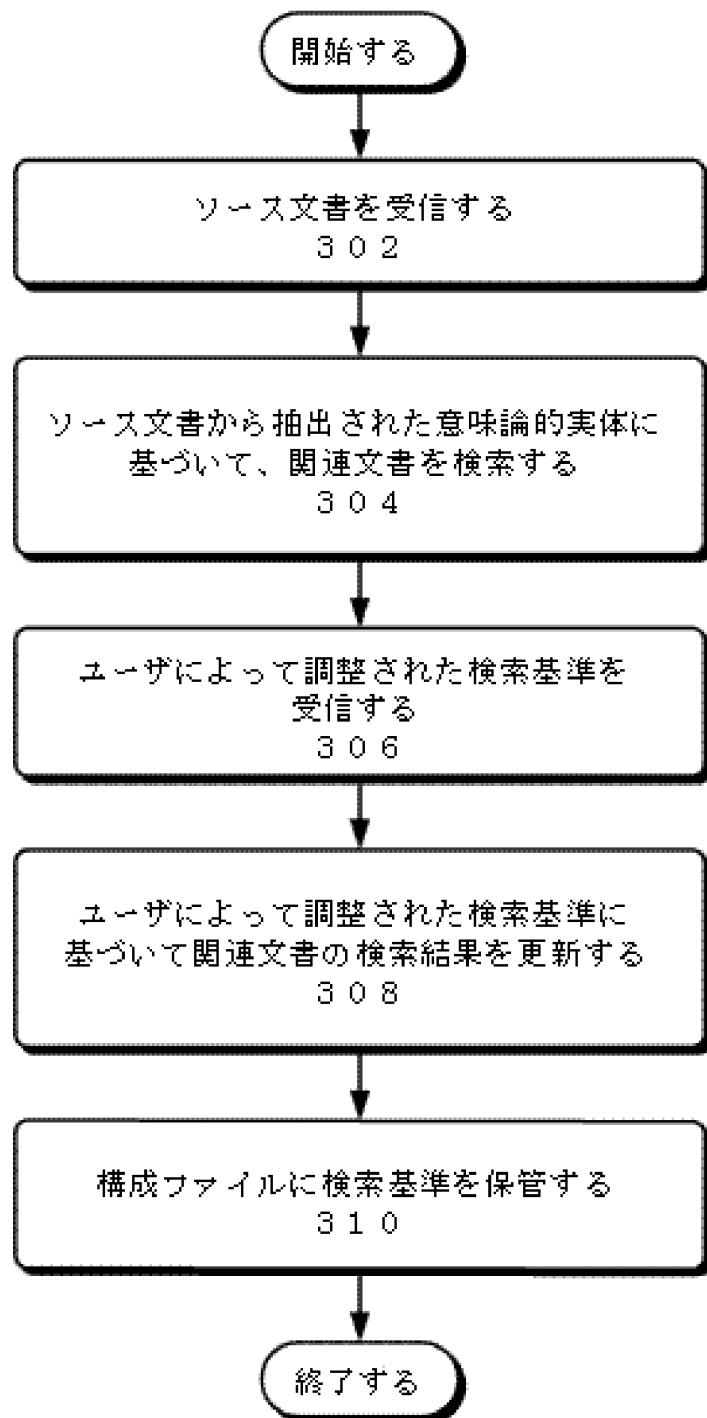


図3

【図4】

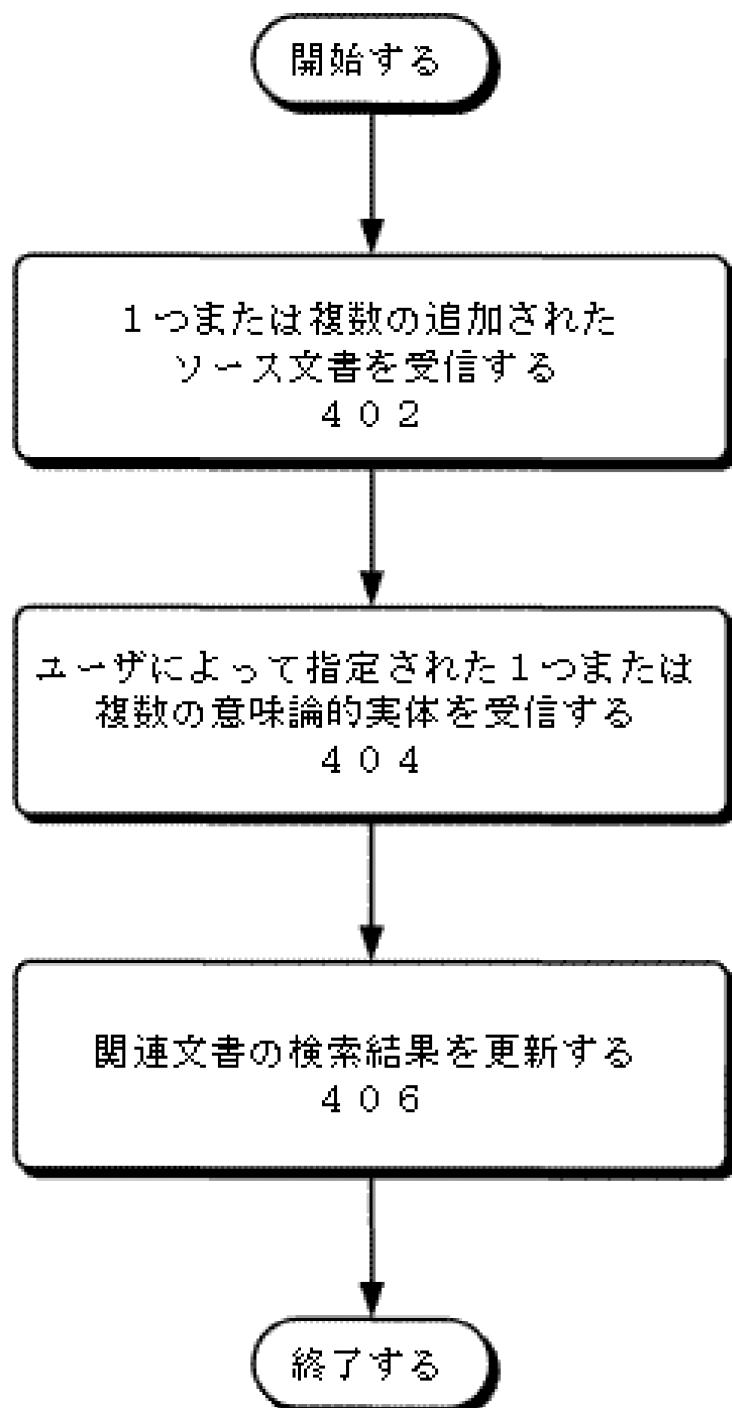


図4

【図5】

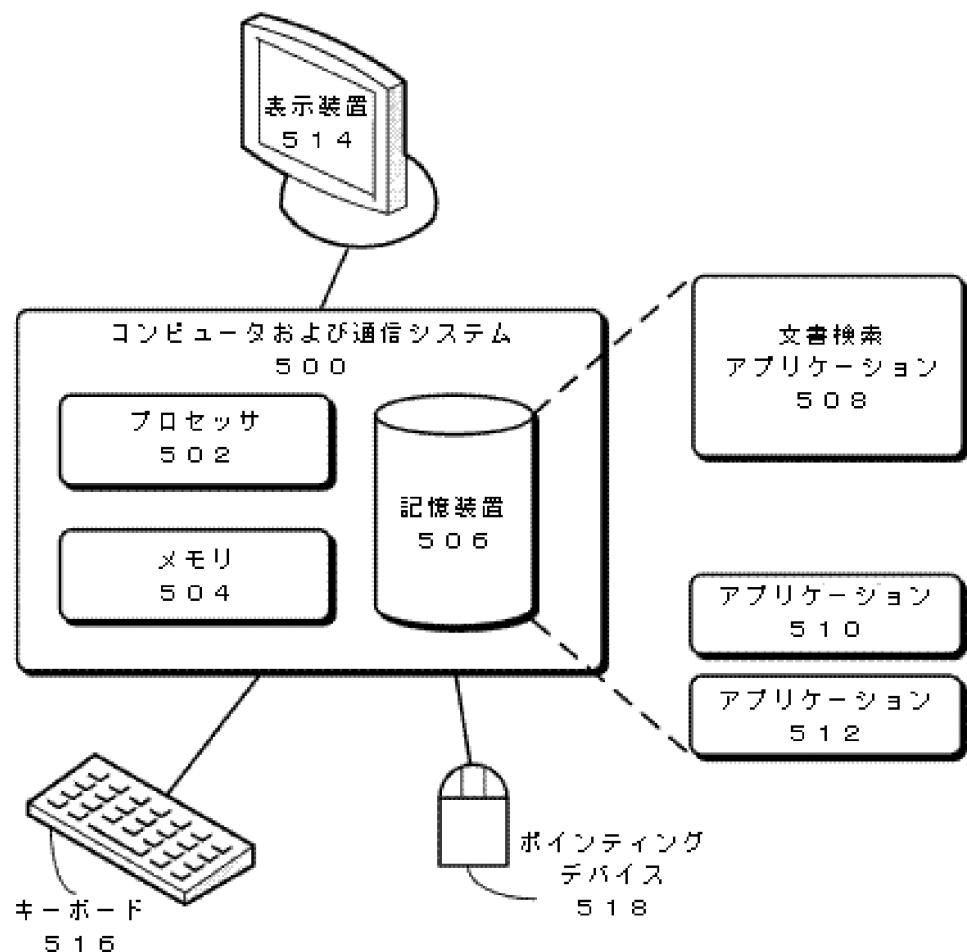


図5

フロントページの続き

(72)発明者 オリヴァー・ブルディスカ

アメリカ合衆国 カリフォルニア州 94304 マウンテン・ビュー オルテガ・アヴェニュー
565 ナンバー11

(72)発明者 ペトロ・イザレフ

アメリカ合衆国 カリフォルニア州 94309 パロ・アルト シエリダン・アヴェニュー 4
10 アパートメント 340

(72)発明者 ク里斯・ホームズ

アメリカ合衆国 カリフォルニア州 94061 レッドウッド・シティ オリバー・ストリート
1165

(72)発明者 アンドレ・モカルスキ

アメリカ合衆国 カリフォルニア州 95126 サンノゼ サドル・ラック・アベニュー 13
10 アパートメント ナンバー443

審査官 野崎 大進

(56)参考文献 国際公開第2011/039848 (WO, A1)

特開2009-048369 (JP, A)

特開2007-065745 (JP, A)

米国特許出願公開第2005/0055365 (US, A1)

(58)調査した分野(Int.Cl., DB名)

G 06 F 17 / 30