

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6136702号
(P6136702)

(45) 発行日 平成29年5月31日(2017.5.31)

(24) 登録日 平成29年5月12日(2017.5.12)

(51) Int.Cl.

F 1

G06F 17/30 (2006.01)

G06F 17/30 170A

G06F 17/27 (2006.01)

G06F 17/30 210A

G06F 17/30 380D

G06F 17/27 665

請求項の数 6 (全 19 頁)

(21) 出願番号

特願2013-153939 (P2013-153939)

(22) 出願日

平成25年7月24日(2013.7.24)

(65) 公開番号

特開2015-26135 (P2015-26135A)

(43) 公開日

平成27年2月5日(2015.2.5)

審査請求日

平成28年4月5日(2016.4.5)

(73) 特許権者 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(74) 代理人 100105142

弁理士 下田 憲次

(72) 発明者 岡嶋 成司

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(72) 発明者 岩倉 友哉

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(72) 発明者 光石 豊

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】場所推定方法、場所推定装置および場所推定プログラム

(57) 【特許請求の範囲】

【請求項 1】

場所推定装置によって実行される場所推定方法であって、

ユーザの発信情報から第1の単語を抽出し、

前記第1の単語に対応する住所に関する情報を記憶部から抽出し、

前記第1の単語と前記住所に関する情報を含む第1のキーワードに対応する第1の指標値を場所毎に算出し、

前記第1の指標値に基づいて、前記発信情報が示す場所を推定する、
ことを特徴とする場所推定方法。

【請求項 2】

前記場所を推定する処理は、

前記第1の指標値のうち、所定の閾値よりも大きい指標値に対応する場所を、前記発信情報が示す場所として推定し、

前記第1の指標値の中に前記所定の閾値よりも大きい指標値が存在しない場合、前記発信情報が示す場所が存在しないと推定する、

処理を含むことを特徴とする請求項1記載の場所推定方法。

【請求項 3】

場所の情報が付与された文書情報から抽出された第2の単語と、前記第2の単語に対応する住所に関する情報を含む第2のキーワードを用いて、前記第2のキーワードと、前記場所との関連性の強さを示す第2の指標値を算出する処理を更に有し、

前記第1の指標値は、前記第2の指標値に基づいて算出することを特徴とする請求項1又は2に記載の場所推定方法。

【請求項4】

前記記憶部から抽出する処理は、前記第1の単語に対応する場所に関する複数の候補を抽出することを含み、

前記発信情報が示す場所を推定する処理は、場所毎に算出した前記第1の指標値に基づいて、前記複数の候補を絞り込むことを含む、
ことを特徴とする請求項1又は2に記載の場所推定方法。

【請求項5】

ユーザの発信情報から第1の単語を抽出するキーワード抽出部と、

10

前記第1の単語に対応する住所に関する情報を記憶部から抽出する辞書検索部と、

前記第1の単語と前記住所に関する情報を含む第1のキーワードに対応する第1の指標値を場所毎に算出するスコア算出部と、

前記第1の指標値に基づいて、前記発信情報が示す場所を推定する推定部と、
を有することを特徴とする場所推定装置。

【請求項6】

場所推定装置に、

ユーザの発信情報から第1の単語を抽出する処理と、

前記第1の単語に対応する住所に関する情報を記憶部から抽出する処理と、

20

前記第1の単語と前記住所に関する情報を含む第1のキーワードに対応する第1の指標値を場所毎に算出する処理と、

前記第1の指標値に基づいて、前記発信情報が示す場所を推定する処理と、
を実行させるための場所推定プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、場所推定方法、場所推定装置および場所推定プログラムに関する。

【背景技術】

【0002】

30

携帯電話機やスマートフォン等に代表される端末には、GPS(Global Positioning System)のような測位機能が一般的に搭載されてきている。端末のユーザは、端末を用いて現在の場所を測位することができる。また、端末のユーザは、例えばソーシャルメディアを利用して端末から文書を送信する際に、現在の場所の情報を文書に付与することができる。

【0003】

近年では、端末から送信される文書の中で言及されている場所を推定する方法として、辞書ベースの場所推定手法および機械学習ベースの場所推定手法が開示されている。辞書ベースの場所推定手法は、予め用意された、キーワードと当該キーワードに対応する場所との組み合わせの情報を示す辞書を利用することにより、文書の中で言及されている場所を推定する方法である。一方、機械学習ベースの場所推定手法は、文書中に現れるキーワードの傾向を学習し、場所に間接的に対応したキーワード(地域関連語)から、言及されている場所を推定する方法である。

40

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2010-128898号公報

【特許文献2】特表2012-529717号公報

【発明の概要】

【発明が解決しようとする課題】

50

【0005】

辞書ベースの場所推定手法によれば、場所を表すキーワードが辞書に登録されており、当該キーワードに対する場所を特定できた場合は、場所を正しく推定することができる。しかしながら、例えば同一の地名が複数の地域で用いられている場合には、場所を一意に特定できない曖昧性が問題となる。例えば、文書から「港区」が抽出されたとしても、「大阪市港区」、「名古屋市港区」、「東京都港区」のいずれを意味するのかを特定するのには困難である。

【0006】

また、機械学習ベースの場所推定手法によれば、おおよその場所までは特定可能であるが、詳細な場所を特定できないことがある。例えば大阪府までは特定できても、大阪府の中のどの市町であるのかまでは特定できないことがある。このように、機械学習ベースの場所推定手法は、推定の粒度が辞書ベースの場所推定手法に比べて大きいという問題を有している。

10

【0007】

本発明の1つの側面では、ユーザの端末から発信される文書の中で言及されている場所を推定する精度を向上させることが可能な場所推定方法、場所推定装置および場所推定プログラム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】**【0008】**

発明の一観点によれば、場所推定装置によって実行される場所推定方法であって、ユーザの発信情報から第1の単語を抽出し、前記第1の単語に対応する住所に関する情報を記憶部から抽出し、前記第1の単語と前記住所に関する情報を含む第1のキーワードに対応する第1の指標値を場所毎に算出し、前記第1の指標値に基づいて、前記発信情報が示す場所を推定する場所推定方法が提供される。

20

【発明の効果】**【0009】**

一実施態様によれば、ユーザの端末から発信される文書の中で言及されている場所を推定する精度を向上させることが可能な場所推定方法、場所推定装置および場所推定プログラムを提供することができる。

【図面の簡単な説明】

30

【0010】

【図1】図1は、場所推定システムの構成の一例を示す図である。

【図2】図2は、場所推定装置のハードウェア構成の一例を示す図である。

【図3】図3は、場所推定システムによる場所推定方法の一例を示すフローチャートである。

【図4】図4は、S101における、学習フェーズの処理の一例を示すフローチャートである。

【図5】図5は、地域情報が付与された文書の情報の一例を示す図である。

【図6】図6は、地域情報が付与された文書から抽出された単語のリストの一例を示す図である。

40

【図7】図7は、S203における、住所表現辞書から住所に関する情報を抽出する処理の一例を示すフローチャートである。

【図8】図8は、学習フェーズにおける、住所表現のリストの一例を示す図である。

【図9】図9は、住所表現辞書の一例を示す図である。

【図10】図10は、住所表現辞書から抽出された、住所に関する情報のリストの一例である。

【図11】図11は、学習フェーズにおける、キーワード群の一例を示す図である。

【図12】図12は、地域分類規則の一例を示す図である。

【図13】図13は、S102における、推定フェーズの処理の一例を示すフローチャートである。

50

【図14】図14は、対象文書の情報の一例を示す図である。

【図15】図15は、S403における、住所表現辞書から住所を構成する文字列を抽出する処理の一例を示すフローチャートである。

【図16】図16は、推定フェーズにおける、住所表現のリストの一例を示す図である。

【図17】図17は、住所表現辞書から抽出された住所および座標の情報のリストの一例である。

【図18】図18は、推定フェーズにおける、キーワード群の一例を示す図である。

【図19】図19は、所属スコアの算出結果の一例である。

【図20】図20は、所属スコアが所定の閾値を超えている地域のリストの一例である。

【図21】図21は、推定結果の一例を示す図である。

【図22】図22は、対象文書から、対象文書の中で言及されている場所を推定する別の一例を示す図である。

【図23】図23は、学習フェーズの変形例を示すフローチャートである。

【図24】図24は、学習フェーズにおける、単語リストの一例を示す図である。

【発明を実施するための形態】

【0011】

以下、本発明の実施形態について、図1から図24を参照して具体的に説明する。

【0012】

図1は、場所推定システムの一例を示す図である。図1に示すように、場所推定システムは、場所推定装置10と端末装置30とを有している。場所推定装置10と端末装置30とは、ネットワーク50を介して相互に通信可能に接続されている。

【0013】

場所推定装置10は、端末装置30から送信される文書を受信し、当該文書の中で言及されている場所を推定する装置である。場所推定装置10は、後述するように、例えば辞書ベースの場所推定手法と機械学習ベースの場所推定手法とを組み合わせて推定を行う。場所推定装置10は、例えばパーソナルコンピュータ(PC)、またはサーバ等によって実現される。場所推定装置10が実行する処理の方法については後述する。

【0014】

端末装置30は、例えば文書を発信するユーザまたは場所推定装置10を利用するユーザが所有する端末であり、例えばスマートフォン、携帯電話、ノートPC(Personal Computer)、デスクトップPC、またはタブレット端末等である。ここで、文書とは、例えばニュース記事や、Twitter(登録商標)、ブログなどのソーシャルメディアに書き込まれた文章である。なお、ユーザには、後述する学習フェーズにおいて、地域分類規則を取得するために地域の付与された文書を発信するユーザも含まれる。

【0015】

次に、場所推定装置10のハードウェア構成について説明する。

【0016】

図2は、場所推定装置10のハードウェア構成の一例を示す図である。図2に示すように、場所推定装置10は、CPU(Central Processing Unit)61、ROM(Read Only Memory)62、RAM(Random Access Memory)63、ストレージ装置64、ネットワークインターフェース65、及び可搬型記憶媒体用ドライブ66等を備えている。

【0017】

場所推定装置10の構成各部は、バス67に接続されている。ストレージ装置64は、例えばHDD(Hard Disk Drive)である。場所推定装置10では、ROM62あるいはストレージ装置64に格納されているプログラム(場所推定プログラムを含む)、あるいは可搬型記憶媒体用ドライブ66が可搬型記憶媒体68から読み取ったプログラム(場所推定プログラムを含む)をCPU61等のプロセッサが実行することにより、場所推定装置10の機能が実現される。

【0018】

以下、場所推定装置10を構成する各部の機能について説明する。

10

20

30

40

50

【0019】

図1に示すように、場所推定装置10は、第1記憶部11と、第2記憶部12と、受信部13と、キーワード抽出部14と、辞書検索部15と、スコア算出部16と、推定部17と、出力部18と、送信部19とを備えている。

【0020】

第1記憶部11は、例えば図2のROM62、ストレージ装置64、可搬型記憶媒体用ドライブ66あるいは可搬型記憶媒体68に対応する。第1記憶部11は、端末装置30から送信される文書の中で言及されている場所を推定するための場所推定プログラムを記憶することができる。

【0021】

第2記憶部12は、例えば図2のROM62、RAM63、ストレージ装置64、可搬型記憶媒体用ドライブ66あるいは可搬型記憶媒体68に対応する。第2記憶部12は、場所推定装置10内で行われる各処理に用いる各種情報を記憶するためのデータベース(DB; Data Base)として用いられる。

【0022】

受信部13は、端末装置30から地域の付与された文書を受信する。受信部13は、端末装置30と相互に通信可能に接続されており、例えば図2のネットワークインターフェース65によって実現される。

【0023】

キーワード抽出部14は、受信部13が受信した文書から、単語を抽出することができる。また、キーワード抽出部14は、抽出した単語から住所表現を抽出することができる。ここで、住所表現とは、住所を有する、建造物、公園、企業、学校、または店舗等を示す名称、または住所そのものである。

【0024】

辞書検索部15は、第2記憶部12に格納されている住所表現辞書を検索して、キーワード抽出部14が抽出した各単語に対応する住所に関する情報をそれぞれ抽出する。位置情報抽出部15は、例えば図2のCPU61あるいはMPU等のプロセッサによって実現される。

【0025】

スコア算出部16は、後述する学習フェーズにおいて、キーワード抽出部14によって抽出された単語と、辞書検索部15によって抽出された住所表現のそれぞれについて、各地域との関連性の程度を示す地域関連スコアを地域毎に算出する。算出した地域関連スコアは、地域分類規則のデータとして用いられる。なお、地域関連スコアは、第2の指標値の一例である。

【0026】

また、スコア算出部16は、後述する推定フェーズにおいて、キーワード抽出部14によって対象文書から抽出されたキーワード群が、どの地域に所属する(関連する)可能性が高いかを示す指標である所属スコアを地域毎に算出する。そして、地域毎に算出した所属スコアを比較することにより、文書の中で言及されている場所がどの地域に所属する可能性が高いのかを推定することができる。ここで、単語だけでなく住所を構成する文字列についても地域毎の所属スコアを算出する対象としている点が、本実施形態の特徴の一つである。スコア算出部16は、例えば図2のCPU61あるいはMPU等のプロセッサによって実現される。なお、所属スコアは、第1の指標値の一例である。

【0027】

推定部17は、スコア算出部16によって算出された所属スコアを所定の閾値と比較することによって、入力文書の中で言及されている場所(地域)を推定する。推定部17は、例えば図1のCPU61あるいはMPU等のプロセッサによって実現される。

【0028】

出力部18は、推定部17によって実行された、文書の中で言及されている場所の推定結果を出力する。出力部18は、例えば液晶ディスプレイ、プラズマディスプレイまたは

有機 E L ディスプレイ等の表示装置である。

【 0 0 2 9 】

送信部 1 9 は、推定部 1 7 によって実行された、文書の中で言及されている場所の推定結果をネットワーク 5 0 に向けて送信する機能を有している。例えば、場所推定装置 1 0 を利用するユーザの端末装置 3 0 は、ネットワーク 5 0 を介して場所推定装置 1 0 から送信された推定結果を受信することができる。

【 0 0 3 0 】

次に、場所推定装置 1 0 による場所推定方法について説明する。

【 0 0 3 1 】

図 3 は、場所推定システムによる場所推定方法の一例を示すフローチャートである。

10

【 0 0 3 2 】

図 3 に示すように、場所推定装置 1 0 は、まず場所の情報の一例である地域の情報（地域情報）が付与された文書をもとに学習を行い、地域分類規則を取得する学習フェーズを実行する（S 1 0 1）。続いて、場所推定装置 1 0 は、対象文書を取得し、地域分類規則をもとに、対象文書の中で言及されている場所の推定を行う推定フェーズを実行する（S 1 0 2）。以降では、学習フェーズおよび推定フェーズの詳細について説明する。

【 0 0 3 3 】

まず、S 1 0 1 における学習フェーズの処理の詳細について説明する。

【 0 0 3 4 】

図 4 は、S 1 0 1 における、学習フェーズの処理の一例を示すフローチャートである。

20

【 0 0 3 5 】

まず、受信部 1 3 は、端末装置 3 0 から地域情報が付与された文書の情報を受信する（S 2 0 1）。受信部 1 3 は、受信した文書の情報を第 2 記憶部 1 2 に格納する。

【 0 0 3 6 】

図 5 は、地域情報が付与された文書の情報の一例を示す図である。図 5 に示すように、地域情報が付与された文書の情報は、地域と文書とが対応付けられた構成を有している。地域は、例えば都道府県名である。図 5 の例では、文書「タワーに行って、帰りに名物のウナギを食べた」には、地域「県」が地域情報として付与されている。また、文書「と言えば、やはり 寺でしょう」には、地域「県」が地域情報として付与されている。

30

【 0 0 3 7 】

図 4 に戻り、キーワード抽出部 1 4 は、地域情報が付与された文書から、単語を抽出する（S 2 0 2）。S 2 0 2 において、キーワード抽出部 1 4 は、第 2 記憶部 1 2 に格納されている地域情報が付与された文書を読み出す。そして、キーワード抽出部 1 4 は、地域情報が付与された文書の中に含まれている単語を、地域情報と対応付けながら抽出する。ここで抽出される単語としては、名詞および複合名詞が好ましい。

【 0 0 3 8 】

図 6 は、地域情報が付与された文書から抽出された単語のリストの一例を示す図である。図 6 の例では、地域「県」に対応する単語として、「タワー」、「名物」および「ウナギ」が抽出されている。また、地域「県」に対応する単語として、「」および「寺」が抽出されている。

40

【 0 0 3 9 】

図 4 に戻り、辞書検索部 1 5 は、第 2 記憶部 1 2 に格納されている住所表現辞書を検索し、住所表現に対応する住所を構成する文字列を抽出する（S 2 0 3）。以下、S 2 0 3 の処理の詳細について説明する。

【 0 0 4 0 】

図 7 は、S 2 0 3 における、住所表現辞書から住所に関する情報を抽出する処理の一例を示すフローチャートである。

【 0 0 4 1 】

50

まず、キーワード抽出部14は、地域情報が付与された文書から住所表現を抽出し、住所表現のリストを作成する(S301)。

【0042】

図8は、学習フェーズにおける、住所表現のリストの一例を示す図である。図8に示すように、住所表現のリストには、文書「タワーに行って、帰りに名物のウナギを食べた」から抽出された「タワー」、および文書「と言えば、やはり寺でしょう」から抽出された「寺」が登録されている。S301の処理によれば、住所表現の抽出は固有表現抽出技術等を用いて行われる。固有表現抽出技術については、公知の任意の手法を用いることができる。

【0043】

図7に戻り、S301の処理の後、辞書検索部15は、住所表現のリストに住所表現が登録されているか否かを判定する(S302)。住所表現のリストに情報が登録されないと判定された場合(S302否定)、辞書検索部15は、住所表現のリストから住所表現を1個抽出する。そして、キーワード抽出部14は、抽出した住所表現を住所表現のリストから削除する(S303)。

10

【0044】

例えば、図8に示すように「タワー」および「寺」が住所表現のリストに登録されている状態で、辞書検索部15が「タワー」を抽出したとする。このとき、辞書検索部15は、抽出した「タワー」を住所表現のリストから削除する。その結果、「寺」のみが住所表現のリストに残ることとなる。

20

【0045】

続いて、辞書検索部15は、住所表現辞書を検索し、S303で抽出された住所表現が住所表現辞書に登録されているか否かを判定する(S304)。

【0046】

図9は、住所表現辞書の一例を示す図である。住所表現辞書は、住所表現と、住所表現に対応する場所の情報とが対応付けられたリストである。住所表現辞書は、例えば図9に示すように、「住所表現」、「住所」および「座標(緯度、経度)」の欄を有している。「座標(緯度、経度)」の欄に登録されている情報は、推定フェーズにおいて使用される情報である。辞書検索部15は、抽出した住所表現が「住所表現」の欄に登録されているか否かを検索することにより、図7におけるS304の判定を行う。

30

【0047】

なお、住所表現を検索する際、辞書検索部15によって抽出された住所表現の文字列と住所表現辞書中の住所表現の文字列とが完全一致である必要はない。例えば近似マッチングによる検索手法や、抽出された文字列を正規化後に完全一致検索する手法等の、公知の任意の手法を用いることができる。

【0048】

S303で抽出された住所表現が住所表現辞書に含まれていると判定された場合(S304肯定)、辞書検索部15は、住所表現辞書から、当該住所表現に対応する住所を構成する文字列を抽出する。そして、辞書検索部15は、抽出した結果を第2記憶部12に格納する(S305)。

40

【0049】

例えば、図9の住所表現辞書を参照すると、「住所表現」の欄に「タワー」が登録されていることがわかる。そこで、辞書検索部15は、「タワー」に対応する住所の欄を参照し、住所「県、市、町」を構成する文字列として、「県」、「市」および「町」を抽出する。S305の処理の後、S302に移り、S302以降の処理を再び実行する。

【0050】

一方、抽出した住所表現が住所表現辞書に登録されていないと判定された場合(S304否定)、S302に移り、S302以降の処理を再び実行する。例えば2回目のS302の処理で、「寺」が住所表現のリストに登録されている場合、再びS302肯定と判

50

定される。そして、S303の処理を経てS304に移る。S304の処理において、辞書検索部15は、「寺」に対応する住所の欄を参照し、住所「県、市、町」を構成する文字列として「県」、「市」および「町」をそれぞれ抽出する。

【0051】

S302において、住所表現のリストに住所表現が登録されていないと判定された場合(S302否定)、S204に移る。S302では、例えば「タワー」および「寺」が住所表現のリストから削除されると、住所表現のリストに登録されている住所表現がなくなるため、この場合はS302で否定判定される。そして、住所を構成する文字列を抽出する一連の処理が終了する。

【0052】

図10は、住所表現辞書から抽出された、住所を構成する文字列のリストの一例である。

【0053】

図10に示すように住所を構成する文字列は、地域と対応付けられて登録されている。このように、地域と対応付けて登録しておくことにより、後の処理で単語と住所を構成する文字列とをマージする作業が容易になる。

【0054】

以上のようにして、場所推定装置10は、住所表現辞書から住所を構成する文字列を取得することができる。

【0055】

図4に戻り、S203の処理の後、スコア算出部16は、S202で抽出された単語と、S203で抽出された住所を構成する文字列とをマージすることにより、キーワード群を構築する(S204)。

【0056】

図11は、学習フェーズにおける、キーワード群の一例を示す図である。図6に示す単語のリストと図10に示す住所を構成する文字列のリストとをマージすると、図11に示すキーワード群を得ることができる。図11に示すように、キーワード群のリストには、地域の情報と、地域に対応する単語または住所に関する情報とが対応付けられて登録されている。例えば、地域「県」には、対応する単語である「タワー」、「名物」および「ウナギ」と、対応する住所である「県」、「市」および「町」とが登録されている。また、地域「県」には、対応する単語である「」および「寺」と、対応する住所を構成する文字列である「県」、「市」および「町」とが登録されている。

【0057】

続いて、スコア算出部16は、抽出された単語および住所を構成する文字列の各々について、地域毎の地域関連スコアを算出する(S205)。地域関連スコアおよび後述する所属スコアの算出方法としては、例えばサポートベクターマシン(Support Vector Machine, SVM)によって学習された線形分類器を用いる方法等、機械学習における公知の種々のスコア算出法を用いることができる。地域関連スコアを算出することにより、地域分類規則を取得することができる。

【0058】

図12は、地域分類規則の一例を示す図である。図12に示すように、単語の各々について、地域毎の地域関連スコアが登録されている。図12の例では、地域を都道府県名で表している。地域関連スコアの値が大きいほど、当該地域との関連性が高いことを示している。図12によれば、例えば単語「タワー」の場合、「県」のスコアは0.75、「県」のスコアは0.03、「×県」のスコアは0.04である。よって、3つの地域の中では県との関連性が最も高いことがわかる。取得された地域分類規則は第2記憶部12に格納され、以降の推定フェーズにおいて利用される。

【0059】

以上のようにして、場所推定装置10は、地域分類規則を取得する学習フェーズを実行する。

10

20

30

40

50

【0060】

次に、S102における、推定フェーズの処理の詳細について説明する。

【0061】

図13は、S102における、推定フェーズの処理の一例を示すフローチャートである。

【0062】

まず、受信部13は、場所の推定を行う対象文書を端末装置30から受信する(S401)。受信部13は、受信した対象文書の情報を第2記憶部12に格納する。

【0063】

図14は、対象文書の情報の一例を示す図である。図14の例では、「休暇を取って寺や神社を見てきた」が対象文書である。図5に示す学習用の文書とは異なり、対象文書の情報には地域情報は含まれていない。そのため、対象文書と地域との対応関係が不明である。場所推定装置10は、対象文書の中で言及されている場所を推定するため、S401以降の各処理を実行する。

【0064】

図13に戻り、S401の処理の後、キーワード抽出部14は、対象文書から単語を抽出する(S402)。S402において、キーワード抽出部14は、第2記憶部12に格納されている対象文書を読み出す。そして、キーワード抽出部14は、対象文書に含まれている単語を抽出する。ここでは、対象文書「休暇を取って寺や神社を見てきた」から単語として「休暇」、「寺」および「神社」が抽出される。

【0065】

続いて、キーワード抽出部14は、対象文書から住所表現を抽出する。そして、辞書検索部15は、第2記憶部12に格納されている住所表現辞書を検索して、住所表現に対応する住所を構成する文字列を抽出する(S403)。

【0066】

図15は、S403における、住所表現辞書から住所を構成する文字列を抽出する処理の一例を示すフローチャートである。

【0067】

まず、キーワード抽出部14は、対象文書から住所表現を抽出し、住所表現のリストを作成する(S501)。

【0068】

図16は、推定フェーズにおける、住所表現のリストの一例を示す図である。図16に示すように、キーワード抽出部14は、抽出された単語である「休暇」、「寺」および「神社」の中から、住所表現として「寺」および「神社」を抽出する。

【0069】

続いて、辞書検索部15は、住所表現のリストに情報が登録されているか否かを判定する(S502)。住所表現のリストに情報が登録されていると判定された場合(S502肯定)、辞書検索部15は、住所表現のリストから住所表現を1個抽出する。そして、辞書検索部15は、抽出した住所表現を住所表現のリストから削除する(S503)。一方、住所表現のリストに情報が登録されていないと判定された場合(S502否定)、S404に移る。

【0070】

例えば、図16に示すように「寺」および「神社」が住所表現のリストに登録されているときに、辞書検索部15が「寺」を住所表現のリストから抽出したとする。このとき、辞書検索部15は、「寺」を抽出した後、「寺」を住所表現のリストから削除する。その結果、住所表現のリストには「神社」のみが残ることとなる。

【0071】

続いて、辞書検索部15は、S503で抽出された住所表現が、住所表現辞書に登録されているか否かを判定する(S504)。S503で抽出された住所表現が住所表現辞書に登録されていると判定された場合(S504肯定)、辞書検索部15は、住所表現辞書

10

20

30

40

50

から、抽出した住所表現に対応する場所の情報として住所および座標の情報を抽出する（S505）。

【0072】

図17は、住所表現辞書から抽出された住所および座標の情報のリストの一例である。図17に示すように、住所表現毎に、対応する住所および座標の情報が抽出される。

【0073】

例えば、S504において住所表現「寺」が住所表現辞書に登録されていると判定された場合、S505において辞書検索部15は、住所表現辞書から住所および座標の情報として「県、市、町 34.××××, 138.××××」を抽出する。S505の処理の後、S502に戻り、S502以降の処理を再び実行する。

10

【0074】

再び実行するS502の処理では、「神社」が住所表現のリストに登録されている。この場合、再びS502肯定と判定される。そして、S503の処理を経てS504に移る。S504の処理において辞書検索部15は、「神社」に対応する住所および座標の情報として、2種類の地域の情報、すなわち「県、市、町 34.××××, 137.××××」および「県、市、町 37.××××, 131.××××」を抽出する。このように、2種類の地域の情報が抽出されたのは、「神社」が複数の地域に存在し、住所表現辞書に2つの地域の情報が登録されているためである。S505の処理の後、S502に戻る。

【0075】

3回目に実行するS502の処理では、住所表現の抽出は全て終わっているため、住所表現のリストに登録されている住所表現は存在しない。よって、辞書検索部15は、住所表現のリストに情報が登録されていないと判定し（S502否定）、S404に移る。

20

【0076】

図13に戻り、キーワード抽出部14と、スコア算出部16は、S402で抽出された単語と、S403で抽出された住所を構成する文字列とをマージすることにより、キーワード群を抽出する（S404）。S404において、キーワード抽出部14は、S505で住所表現辞書から抽出された、住所および座標の情報のリストから、住所を構成する文字列を抽出する。例えば、その結果、「県」、「市」、「町」、「県」、「市」、および「町」が抽出される。これらの文字列と、S402で抽出された単語である「休暇」、「寺」および「神社」とをマージすると、キーワード群を得ることができる。

30

【0077】

図18は、推定フェーズにおける、キーワード群の一例を示す図である。以上のようにして、場所の推定に用いるキーワード群を抽出する。

【0078】

図13に戻り、S404の処理の後、スコア算出部16は、キーワード群がどの地域に所属する（関連する）可能性が高いかを推定するため、所属スコアを地域毎に算出する（S405）。スコア算出部16は、キーワード群およびS101で取得した地域分類規則に基づいて、地域毎の所属スコアを算出する。

40

【0079】

図19は、所属スコアの算出結果の一例である。図19の例においても、図12と同様に地域を都道府県名で表している。また、所属スコアの値が大きいほど、当該地域との関連性が強いことを示している。

【0080】

図13に戻り、S405の処理の後、推定部17は、所属スコアが所定の閾値を超えている地域のリストを作成する（S406）。具体的には、推定部17は、S405で算出した地域毎の所属スコアを参照し、所定の閾値を超えている地域を抽出する。例えば閾値を0として図17を参照すると、1.88の所属スコアを有する「県」のみが抽出される。そこで、推定部17は、「県」を載せた地域のリストを作成する。なお、閾値は、

50

既に言及している場所が明らかになっている文書を用いて推定精度を測定する実験を繰り返しながら、最も推定精度が高くなるように設定することが好ましい。この方法により、推定精度の向上を図ることができる。

【0081】

図20は、所属スコアが所定の閾値を超えている地域のリストの一例である。なお、図20の例では、地域として「県」のみが地域のリストに登録されているが、閾値を超えた地域が複数存在する場合は、複数の地域が登録されることとなる。

【0082】

図13に戻り、S405の処理の後、推定部17は、地域のリストに地域の情報が登録されているか否かを判定する(S407)。

10

【0083】

地域のリストに地域の情報が登録されていないと判定された場合(S407否定)、推定フェーズの処理を終了する。

【0084】

一方、地域のリストに地域の情報が登録されていると判定された場合(S407肯定)、推定部17は、地域のリストから地域を一つ抽出する。そして、推定部17は、地域を抽出した後、当該地域を地域のリストから削除する(S408)。

【0085】

例えば、推定部17は、地域の情報である「県」のみが地域のリストに登録されている状態で、「県」を抽出する。その後、辞書検索部15は、地域のリストから「県」を削除する。その結果、地域のリストには、地域の情報が登録されていない状態となる。

20

【0086】

続いて、推定部17は、抽出した地域に関連する、住所に関する情報を抽出する。そして、出力部18は、抽出結果を出力する(S409)。例えば、推定部17は、図17に示す住所および座標の情報のリストから、「県」が含まれる一連の情報として「寺県、市、町 34.××××, 138.××××」および「神社 県、市、町 34.××××, 137.××××」を抽出する。そして、出力部18は、抽出されたこれらの情報を、推定結果として出力する。

【0087】

図21は、推定結果の一例を示す図である。図21に示すように、対象文書に含まれる住所表現の各々について、場所の推定結果を出力することができる。

30

【0088】

図13に戻り、S409の処理の後、S407に戻る。そして、S407以降の処理を再び実行する。本実施例では、地域のリストには「県」しか登録されていなかったため、「県」が削除された後に実行する2回目のS407の処理では、S407否定と判定される。これにより、推定フェーズが終了となる。

【0089】

以上のようにして、場所推定装置10は、受信した文書の中で言及されている場所を推定することができる。

【0090】

40

以下、本実施形態による効果について説明する。

【0091】

図22は、対象文書から、対象文書の中で言及されている場所を推定する別の一例を示す図である。図22(a)は対象文書を示している。図22(b)は、対象文書中の住所表現に対応する、住所表現辞書から抽出された住所および座標の情報のリストを示している。図22(c)は、対象文書中の住所表現をもとに算出した、各地域に関する所属スコアの算出結果を示している。

【0092】

図22(a)に示すように、推定フェーズにおいて、対象文書が「昨日、武蔵小杉駅を利用した」である場合には、住所表現として「武蔵小杉駅」が抽出される。そして、図2

50

2 (b) を参照すると、住所表現辞書には「武藏小杉駅」が登録されているため、住所表現辞書から、「武藏小杉駅」に対応する住所に関する情報「神奈川県、川崎市、中原区」および座標の情報「35.57666667, 139.65944444」を抽出することができる。

【0093】

ところが、学習フェーズにおいて、地域分類規則を作成する際に用いるキーワード群は、辞書中の全ての住所表現を網羅しているとは限らない。例えば学習フェーズにおいて、地域関連スコアを算出する際に用いられるキーワード群に「武藏小杉駅」が含まれていなかつた場合は、「武藏小杉駅」が神奈川県にあることが学習されない。このため、図22 (c) の例に示すように、神奈川県に関する所属スコアを算出した場合に、所属スコアが閾値 (0.3とする) よりも低くなることがあり得る。その結果、辞書による推定において住所表現辞書により正しい推定がなされたにも関わらず、機械学習による推定において辞書による推定結果が却下され、対象文書が場所について言及していないと判定されてしまうことがあり得る。

10

【0094】

一方、本実施形態によれば、学習フェーズおよび推定フェーズにおいて、対象文書から抽出される単語だけでなく、住所表現辞書から抽出された住所を構成する文字列も加えてキーワード群を構築している。これにより、地域関連スコアおよび所属スコアをより正確に算出することができるため、上述のように、住所表現辞書によって正しい推定がされたにも関わらず、機械学習による推定に基づいて、推定結果が却下されることを防ぐことができる。

20

【0095】

単語だけでなく住所を構成する文字列も加えてキーワード群を構築する処理は、推定フェーズだけで行った場合においても相応の効果が期待できる。しかし、上述の処理を学習フェーズおよび推定フェーズの両方において行う方が好ましい。

【0096】

(変形例)

次に、本実施形態における変形例について説明する。なお、本変形例を実現するための情報処理システムは、図1および図2に例示されている場所推定システムの構成を用いることができるため、重複部分についての説明を省略する。

30

【0097】

図23に示す実施形態では、学習フェーズにおいて、住所表現辞書から住所を構成する文字列を抽出する際に、文書から住所表現を抽出してから、各々の住所表現に対応する住所に関する情報を抽出している。これに対して本変形例では、住所表現を抽出せずに、単語に対応する住所を構成する文字列を、住所表現辞書を検索して抽出することを特徴としている。

【0098】

図23は、学習フェーズの変形例を示すフローチャートである。なお、学習フェーズの処理を開始してからS202までの処理は、これまで説明した処理と同様であるので、説明は省略する。

40

【0099】

S202の処理の後、キーワード抽出部14は、S202の処理で抽出された単語をもとに、単語リストを作成する (S301a)。

【0100】

図24は、学習フェーズにおける、単語リストの一例を示す図である。図24に示すように、単語リストには、文書「タワーに行って帰りに名物のウナギを食べた」から抽出された「タワー」、「名物」、「ウナギ」および文書「と言えば、やはり 寺でしょう」から抽出された「」、「寺」が登録されている。

【0101】

図23に戻り、S301aの処理の後、辞書検索部15は、単語リストに登録されてい

50

る単語について、住所表現辞書に対応する住所に関する情報があるか否かを一語ずつ検索していく。

【0102】

まず、辞書検索部15は、単語リストに単語が登録されているか否かを判定する(S302a)。単語リストに単語が登録されていると判定された場合(S302a肯定)、辞書検索部15は、単語リストから単語を1個抽出する。そして、キーワード抽出部14は、抽出した単語を単語リストから削除する(S303a)。

【0103】

続いて、辞書検索部15は、住所表現辞書を検索し、S303aで抽出した単語が住所表現辞書に登録されているか否かを判定する(S304a)。

10

【0104】

抽出した単語が住所表現辞書に含まれていると判定された場合(S304a肯定)、辞書検索部15は、住所表現辞書から、抽出した単語に対応する住所を構成する文字列を抽出する。そして、辞書検索部15は、抽出した結果を第2記憶部12に格納する(S305a)。

【0105】

一方、抽出した単語が住所表現辞書に登録されていないと判定された場合(S304a否定)、S302aに移り、S302a以降の処理を実行する。この処理を繰り返し行っていくと、やがて単語リストに登録されている単語がなくなる。登録されている単語がなくなると、S302aにおいて、住所表現のリストに住所表現が登録されていないと判定され(S302a否定)、S204に移る。これにより、住所を構成する文字列を抽出する一連の処理が終了する。S204以降の処理は、これまで説明した処理と同様であるので、説明は省略する。

20

【0106】

本変形例によれば、単語から住所表現を抽出する処理を省くことができるため、処理を簡略化することができる。また、単語からの住所表現の抽出漏れが発生する懸念を払拭することができる。

【0107】

以上、本発明の好ましい実施例について詳述したが、本発明は特定の実施例に限定されるものではなく、種々の変形や変更が可能である。例えば、これまで説明した場所推定装置10の処理は、辞書ベースの場所推定手法と機械学習ベースの場所推定手法とを組み合わせて推定を行うだけでなく、辞書ベースの場所推定手法のみ、あるいは機械学習ベースの場所推定手法のみのケースにおいても適用可能である。例えば機械学習ベースの場所推定手法のみを用いて推定を行う場合、抽出された単語に対応する住所に関する情報を抽出したあとに、住所に関する情報に含まれる場所の頻度をカウントし、カウント数を場所毎に比較することにより、文書が言及している場所を推定することも可能である。ここで、カウント数は、第1の指標値の別の一例である。

30

【0108】

また、例えば、上述の変形例では、学習フェーズにおいて、単語に対応する住所を構成する文字列を、住所表現辞書を検索して抽出する処理について説明した。一方、当該処理は、推定フェーズにおいて、住所表現辞書を用いて、単語に対応する住所および座標の情報を抽出する処理に適用することも可能である。

40

【符号の説明】

【0109】

10：場所推定装置

11：第1記憶部

12：第2記憶部

13：受信部

14：キーワード抽出部

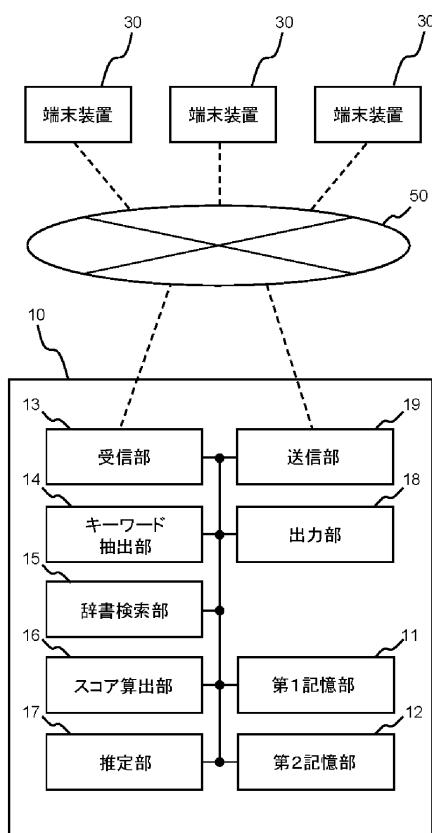
15：辞書検索部

50

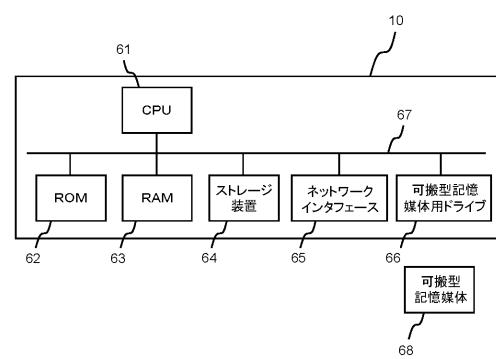
- 1 6 : スコア算出部
 1 7 : 推定部
 1 8 : 出力部
 1 9 : 送信部
 3 0 : 端末装置
 5 0 : ネットワーク
 6 1 : C P U
 6 2 : R O M
 6 3 : R A M
 6 4 : ストレージ装置
 6 5 : ネットワークインターフェース
 6 6 : 可搬型記憶媒体用ドライブ
 6 7 : バス
 6 8 : 可搬型記憶媒体

10

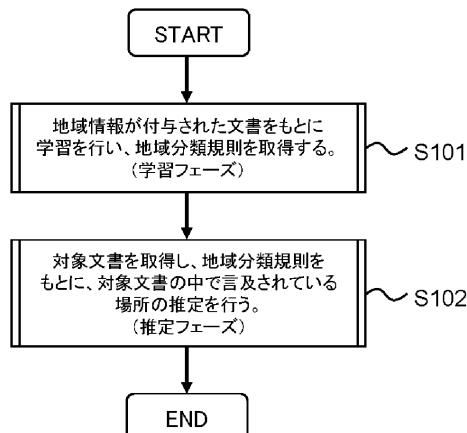
【図1】



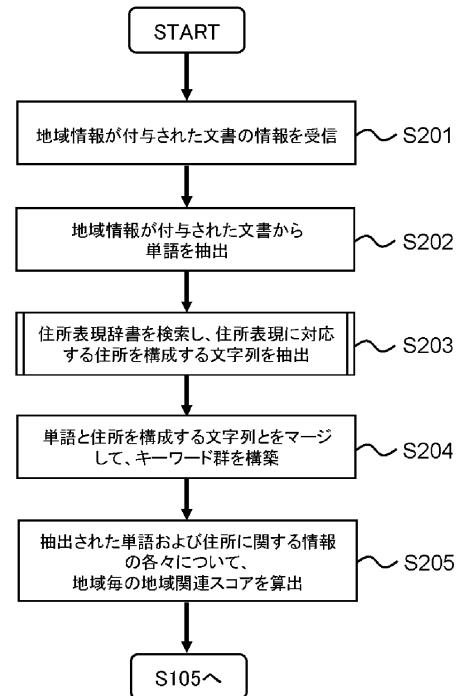
【図2】



【図3】



【図4】



【図5】

地域の付与された文書

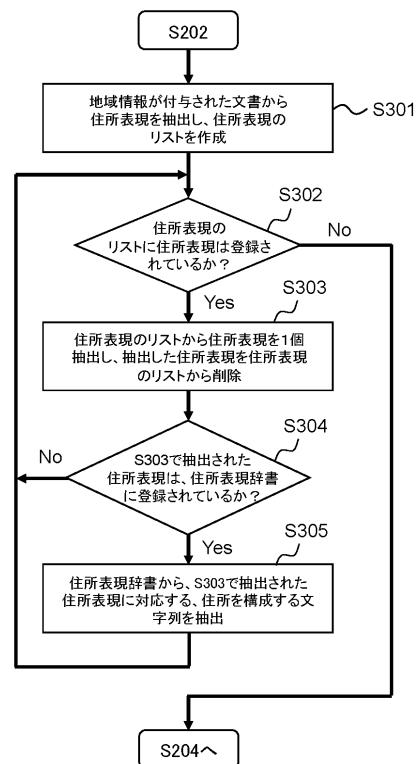
地域	文書
○県	○タワーに行って、帰りに名物のウナギを食べた。
△県	△と言えば、やはり△寺でしょう。
...	...
...	...

【図6】

抽出された単語

地域	単語
○県	○タワー, 名物, ウナギ
△県	△, △寺
...	...
...	...

【図7】



【図 8】

抽出された住所表現

住所表現
○タワー, △寺

【図 9】

住所表現辞書

住所表現	住 所	座標(緯度・経度)
...
○タワー	○県, ○市, ○町	35.xxxx, 139.xxxx
△寺	△県, △市, △町	34.xxxx, 138.xxxx
△神社	△県, △市, ▲町	34.xxxx, 137.xxxx
△神社	□県, □市, □町	37.xxxx, 131.xxxx
×県, ×市, ×町	×県, ×市, ×町	34.xxxx, 139.xxxx
...

【図 10】

抽出された住所を構成する文字列

地 域	住所を構成する文字列
○県	○県, ○市, ○町
△県	△県, △市, △町
...	...
...	...

【図 11】

学習キーワード群

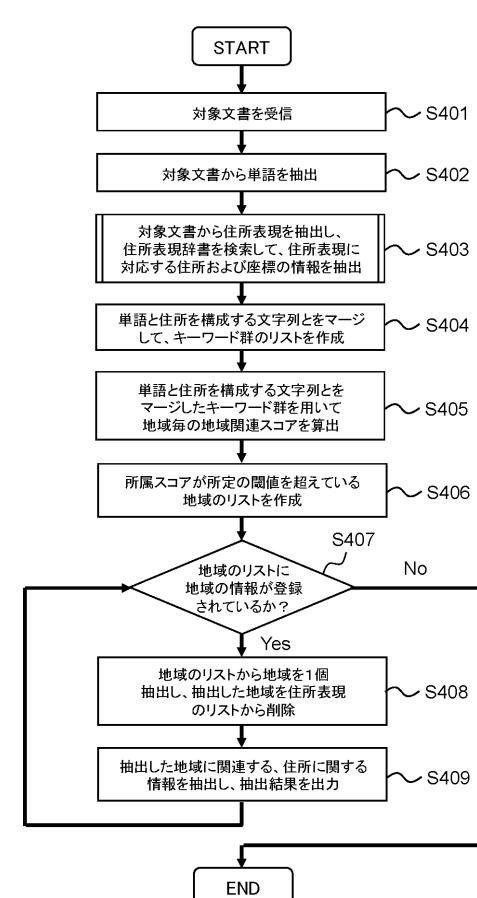
地 域	単語または住所
○県	○タワー, 名物, ウナギ, ○県, ○市, ○町
△県	△, △寺, △県, △市, △町
...	...
...	...

【図 12】

地域分類規則

単 語	地 域			
	○県	△県	×県	...
○タワー	0.75	0.03	0.04	...
名物	-0.01	-0.02	-0.01	...
ウナギ	0.38	-0.07	0.10	...
○県	0.88	-0.23	-0.24	...
○市	0.81	-0.33	-0.13	...
...

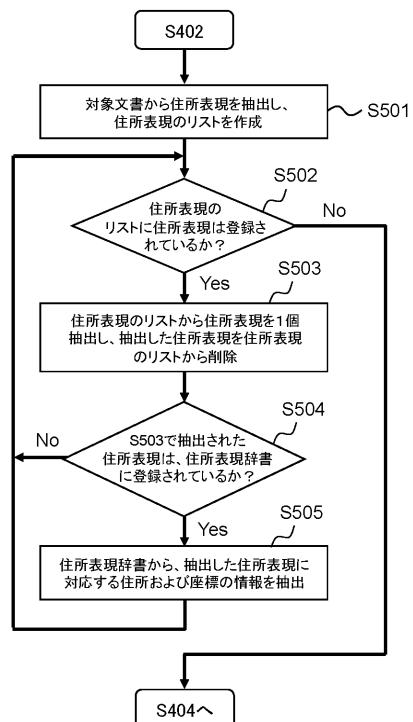
【図 13】



【図14】

対象文書
文書
休暇を取って△寺や△神社を見てきた。

【図15】



【図16】

抽出された住所表現

住所表現
△寺, △神社

【図19】

所属スコアの算出結果

地 域	所属スコア
○県	-0.89
△県	1.88
□県	-0.10
...	...

【図17】

抽出された住所および座標の情報

住所表現	住 所	座標(緯度・経度)
△寺	△県, △市, △町	34.xxxx, 138.xxxx
△神社	△県, △市, ▲町	34.xxxx, 137.xxxx
△神社	□県, □市, □町	37.xxxx, 131.xxxx

【図18】

場所の推定に用いるキーワード群

キーワード群
休暇, △寺, △神社, △県, △市, △町, ▲町, □県, □市, □町

【図20】

所属スコアが閾値を超える地域

地 域
△県

【図21】

推定結果		
住所表現	住所	座標(緯度・経度)
△寺	△県, △市, △町	34.xxxx, 138.xxxx
△神社	△県, △市, ▲町	34.xxxx, 137.xxxx

【図22】

(a) 対象文書

文書
昨日、武蔵小杉駅を利用した。

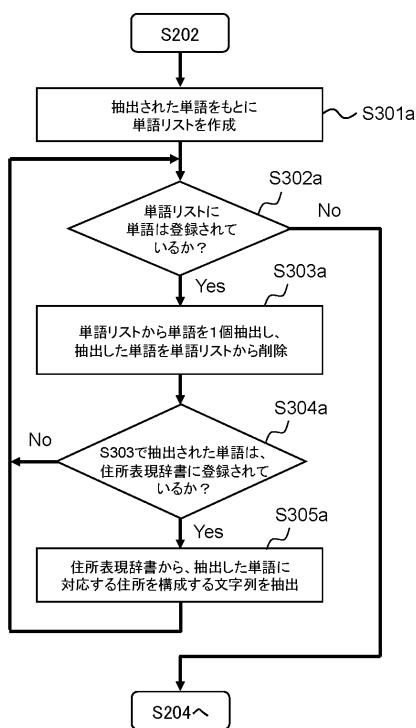
(b) 住所表現辞書

住所表現	住所	座標(緯度・経度)
...
武蔵小杉駅	神奈川県, 川崎市, 中原区	35.5766666667, 139.659444444
...
...

(c) 所属スコアの算出結果

地 域	所属スコア
東京都	-0.89
大阪府	1.88
神奈川県	-0.02
...	...

【図23】



【図24】

単語リスト
単語
○タワー, 名物, ウナギ, △, △寺

フロントページの続き

審査官 小太刀 慶明

(56)参考文献 特開2010-134504 (JP, A)

特開2010-231560 (JP, A)

特開2013-054620 (JP, A)

特開2010-198281 (JP, A)

菊井 玄一郎 他, リッチャノテーション: 固有表現に焦点をあてた知識抽出の試み, 情報処理学会研究報告, 社団法人情報処理学会, 2008年 7月10日, Vol. 2008, No. 67, pp. 73-78

(58)調査した分野(Int.Cl., DB名)

G 06 F 17 / 30

G 06 F 17 / 27