

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5266761号
(P5266761)

(45) 発行日 平成25年8月21日(2013.8.21)

(24) 登録日 平成25年5月17日(2013.5.17)

(51) Int.Cl.

F 1

G 0 6 F 17/30 (2006.01)

G 0 6 F 17/30 3 1 0 Z

G 0 6 F 17/30 1 7 0 C

請求項の数 21 (全 22 頁)

(21) 出願番号 特願2008-3576 (P2008-3576)
 (22) 出願日 平成20年1月10日 (2008.1.10)
 (65) 公開番号 特開2009-169470 (P2009-169470A)
 (43) 公開日 平成21年7月30日 (2009.7.30)
 審査請求日 平成22年12月22日 (2010.12.22)

(73) 特許権者 000003997
 日産自動車株式会社
 神奈川県横浜市神奈川区宝町2番地
 (74) 代理人 110000486
 とこしえ特許業務法人
 (72) 発明者 斎藤 大介
 神奈川県横浜市神奈川区宝町2番地 日産
 自動車株式会社内
 (72) 発明者 富樫 実
 神奈川県横浜市神奈川区宝町2番地 日産
 自動車株式会社内
 (72) 発明者 大野 健
 神奈川県横浜市神奈川区宝町2番地 日産
 自動車株式会社内

最終頁に続く

(54) 【発明の名称】 情報案内システムおよびその認識辞書データベース更新方法

(57) 【特許請求の範囲】

【請求項1】

音声信号を取得する音声入力部と、
ナビゲーションによる経路案内が可能な施設名称である初期キーワードをあらかじめ記憶させた認識辞書データベースと、
 音声入力部で取得された音声信号より、認識辞書データベースに記憶されたキーワードを認識する音声検索部と、
 音声検索部による認識結果に基づいて案内情報を生成する案内情報生成手段と、
 案内情報を提示する出力部とを備える情報案内システムにおいて、
 前記初期キーワードに意味的な関連性をもつ語彙を関連キーワードとしてWEB情報から収集する関連語彙収集部と、
 該関連語彙収集部が収集した前記関連キーワードを所定のタイミングで前記認識辞書データベースに追加する辞書更新部とを有して、
 前記音声検索部は前記初期キーワードと関連キーワードとを認識対象のキーワードとし

、
前記関連語彙収集部は、前記初期キーワードを用いてWEB検索を実行し、検索されたWEBページに含まれるテキスト情報を解析して、前記初期キーワードの別称あるいは前記初期キーワードを代表する事象を表現するための特定の言い回しを伴って出現頻度が所定値以上の語彙を前記関連キーワードとして収集することを特徴とする情報案内システム

10

20

【請求項 2】

前記案内情報生成手段は、前記案内情報として、前記初期キーワードに対応する施設名称で特定される所定の地点までの経路案内の情報を生成することを特徴とする請求項 1 に記載の情報案内システム。

【請求項 3】

前記関連語彙収集部は、前記出現頻度に加えて、さらに語彙毎に重み付けを施した関連度を算出し、該関連度の積算値が所定値以上の語彙に絞って前記関連キーワードを抽出することを特徴とする請求項 1 または 2 に記載の情報案内システム。

【請求項 4】

前記関連度を算出する重み付けは、語彙毎にその表示位置および表示形式にもとづいて設定されることを特徴とする請求項 3 に記載の情報案内システム。

10

【請求項 5】

前記関連度を算出する重み付けは、語彙の表示位置がWEB ページ内のタイトル、見出し、およびページ先頭に近い位置にある程大きい値とすることを特徴とする請求項 4 に記載の情報案内システム。

【請求項 6】

前記関連度を算出する重み付けは、語彙の表示位置が前記初期キーワードの出現位置からの距離が近いほど大きい値とすることを特徴とする請求項 4 または 5 に記載の情報案内システム。

【請求項 7】

20

前記出現位置からの距離は、少なくとも前記初期キーワードから対象の語彙までに含まれる文字数、形態素数、句数、文数、段落数の何れかで測ることを特徴とする請求項 6 に記載の情報案内システム。

【請求項 8】

前記関連度を算出する重み付けは、語彙の文字サイズ、下線の有無および書体のいずれかで表わされる書式がWEB ページ内で最も多く使われる通常書式と異なる場合に、通常書式よりも大きい値とすることを特徴とする請求項 4 から 7 のいずれか 1 に記載の情報案内システム。

【請求項 9】

前記関連度を算出する重み付けは、前記初期キーワードの別称あるいは初期キーワードを代表する事象を表現するための特定の言い回しに伴って出現した語彙について、大きい値とすることを特徴とする請求項 3 に記載の情報案内システム。

30

【請求項 10】

前記特定の言い回しには、少なくとも、「～と言えば」、「～で有名」、「～が有名」、「～として有名」、「～で知られる」、「～が知られる」、「～として知られる」、「～で著名」、「～が著名」、「～として著名」、「～で名高い」、「～が名高い」、「～として名高い」、「～の代名詞」、「～で親しまれる」、「～として親しまれる」、「～と呼ばれる」、「～とも呼ばれる」、「別称～」、「別称～」、「愛称～」、「略称～」、「通称～」、および「略して～」のいずれかが含まれることを特徴とする請求項 1 から 9 のいずれか 1 に記載の情報案内システム。

40

【請求項 11】

前記関連語彙収集部は、前記関連キーワードを収集する際に、前記初期キーワードとしての施設名称 A を包含する別の施設名称 B あるいは施設名称 A を包含する地域名称 C が存在する場合に、該施設名称 B または地域名称 C を用いてWEB 検索を実行して検索されたWEB ページを解析対象から除外することを特徴とする請求項 1 から 10 のいずれか 1 に記載の情報案内システム。

【請求項 12】

前記認識辞書データベースは、前記初期キーワードとしての施設名称とその施設種別を対応させて記憶し、

前記関連語彙収集部は、前記関連キーワードを収集する際に、前記施設名称に対応する

50

施設種別以外の施設種別名の語彙が所定値以上の頻度で含まれるWEBページを解析対象から除外することを特徴とする請求項 1 から 1 1 のいずれか 1 に記載の情報案内システム。

【請求項 1 3】

前記辞書更新部は、起動時あるいは現在地が所定の領域を超えた時に、前記関連語彙収集部に対して前記初期キーワードについて関連キーワードの収集を要求し、該関連語彙収集部が収集し返信した関連キーワードを用いて、前記認識辞書データベースを更新することを特徴とする請求項 1 から 1 2 のいずれか 1 に記載の情報案内システム。

【請求項 1 4】

前記関連語彙収集部は、前記認識辞書データベースに記憶されている関連キーワードと、関連語彙収集部が収集した関連キーワードとに差異がある場合に、収集した関連キーワードのうち差異がある関連キーワードを前記辞書更新部へ返信することを特徴とする請求項 1 3 に記載の情報案内システム。

10

【請求項 1 5】

前記音声検索部は、施設検索利用履歴を監視し、地域別に利用頻度を記録する機能を備え、

前記辞書更新部は、前記利用頻度が所定値以上の地域について、前記認識辞書データベースを更新することを特徴とする請求項 1 3 または 1 4 に記載の情報案内システム。

【請求項 1 6】

前記音声検索部は、施設検索利用履歴を監視し、地域別に利用頻度を記録する機能を備え、

20

前記辞書更新部は、前記利用頻度が所定値より低い地域について、前記認識辞書データベースの関連キーワードを消去することを特徴とする請求項 1 3 から 1 5 のいずれか 1 に記載の情報案内システム。

【請求項 1 7】

前記音声検索部は、認識した関連キーワードについて重複する複数の初期キーワードがある場合、前記出力部を通じて該複数の初期キーワードを選択可能にユーザへ提示することを特徴とする請求項 1 から 1 6 のいずれか 1 に記載の情報案内システム。

【請求項 1 8】

前記認識辞書データベース、音声入力部、音声検索部、辞書更新部および出力部をクライアント側に配置し、前記関連語彙収集部をサーバ側に配置して、クライアント・サーバ型に構成したことを特徴とする請求項 1 から 1 7 のいずれか 1 に記載の情報案内システム。

30

【請求項 1 9】

前記認識辞書データベース、音声入力部、音声検索部、辞書更新部、関連語彙収集部および出力部を統合してスタンドアロン型に構成したことを特徴とする請求項 1 から 1 8 のいずれか 1 に記載の情報案内システム。

【請求項 2 0】

認識手段が、音声認識により認識辞書データベースに記憶されたキーワードを認識して、提供手段が、案内情報を提供する情報案内システムにおいて、

40

収集手段が、認識辞書データベースにあらかじめ記憶されたナビゲーションによる経路案内が可能な施設名称である初期キーワードに意味的な関連性をもつ関連キーワードをWEB情報から収集し、

更新手段が、収集した関連キーワードを前記認識辞書データベースに追加して、前記キーワードを更新する認識辞書データベース更新方法であって、

前記収集手段は、前記初期キーワードを用いてWEB検索を実行し、検索されたWEBページに含まれるテキスト情報を解析して、前記初期キーワードの別称あるいは前記初期キーワードを代表する事象を表現するための特定の言い回しを伴って出現頻度が所定値以上の語彙を前記関連キーワードとして収集することを特徴とする認識辞書データベース更新方法。

50

【請求項 2 1】

前記認識辞書データベースおよび前記更新手段はクライアント側に配置され、前記収集手段はサーバ側に配置され、

前記収集手段による前記関連キーワードの収集はサーバ側で任意のタイミングで実行され、

前記更新手段による収集した関連キーワードの前記認識辞書データベースへの追加は、クライアント側の起動時またはクライアント側の現在地が所定の領域を超えた時に実行されることを特徴とする請求項 2 0 に記載の認識辞書データベース更新方法。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、音声認識により認識辞書データベースに記憶されたキーワードを認識して、案内情報を提供する情報案内システムおよびこれに用いる認識辞書データベースの更新方法に関する。

【背景技術】

【0002】

従来から、ユーザの発話を音声認識技術によって認識し、認識結果に応じて情報提供や機器操作を行うような音声インタフェースが知られており、このような音声インタフェースは、例えば自動車のナビゲーションシステムや電話応答システムである I V R (I n t e r a c t i v e V o i c e R e s p o n s e : 音声自動応答装置)等に適用されている。

20

こうしたシステムでよく利用されるサービスに、施設の検索がある。

これは、ユーザが施設名称、あるいは施設のある都道府県、施設の種別などを発話し、これをシステムが認識・検索することで、所望の施設の場所へのナビゲーション、電話接続などのサービスを提供するものである。

こうした施設の検索では、各施設に関連する情報が多く、また更新頻度も早いことから、情報の鮮度が保てないという課題がある。

【0003】

そこで、例えば特開 2 0 0 1 - 2 8 9 6 6 1 号公報には、常に最新の情報で施設を検索できるように、インターネットの情報を伴う音声検索によって、最新の情報で施設を検索するようにしたものが提案されている。

30

これは、ユーザが検索したい施設の条件を指定すると、ホームページなどインターネットコンテンツからの施設のイベントや料金、天候情報などを加味して、合致する施設を抽出して案内するものである。

これにより、所望の目的はおおよそ決まっているが、施設名称は確定していない場合に、適切な施設が推薦されるので、利便性が高まる。

【特許文献 1】特開 2 0 0 1 - 2 8 9 6 6 1 号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

40

一方、検索したい施設がユーザの中では既に確定しており、このユーザが欲する施設を検索する場合には、認識辞書データベースに記載された認識語彙、すなわち施設名称を発話することが前提となる。

ところがこの認識辞書データベースは、各施設主が公的機関等に申請した名称を使用して作成されている。したがってユーザが所望の検索結果を得るためには、認識辞書データベース上に記録された施設名称をユーザが正しく発話する必要がある。

【0005】

このため、施設の略称やその施設の特徴的なキーワードをユーザが覚えていたとしても、正確な語彙(正しい施設名称)を覚えていない場合には、ユーザが所望する施設の検索を行うことができない。

50

たとえば鎌倉にある寺社「名月院」は、「あじさいで有名な寺」として広く知られており、別名「あじさい寺」とも呼ばれているので、ユーザが「鎌倉にあじさいで有名な寺」があることは知っているても、その寺社の正式名が「名月院」であることを知らない場合には、施設の検索ができないこととなる。

【0006】

したがって本発明は、上記の問題点に鑑み、認識辞書データベースに記録された名称を知らない場合でも、効率よく認識辞書データベースの更新を行って関連する語彙を蓄積し、ユーザが所望する検索を行うことができる情報案内システムおよびその認識辞書データベースの更新方法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】

【0007】

本発明は、認識辞書データベースにあらかじめ記憶されたナビゲーションによる経路案内が可能な施設名称である初期キーワードに意味的な関連性をもつ関連キーワードをWEB情報から収集し、収集した関連キーワードを所定のタイミングで認識辞書データベースに追加して、音声認識対象を増すものとした。

また、本発明においては、関連キーワードをWEB情報から収集する際に、初期キーワードを用いてWEB検索を実行し、検索されたWEBページに含まれるテキスト情報を解析して、初期キーワードの別称あるいは初期キーワードを代表する事象を表現するための特定の言い回しを伴って出現頻度が所定値以上の語彙を関連キーワードとして収集するようにした。

【発明の効果】

【0008】

本発明によれば、正確な語彙を知らないユーザであっても、認識辞書データベースが初期キーワードに関連する関連キーワードで自動的に更新されるから、認識辞書データベースに含まれる初期キーワードおよび関連キーワードを認識対象として、ユーザが所望する目的の情報検索が可能となる。

【発明を実施するための最良の形態】

【0009】

次に本発明の実施の形態を実施例により説明する。

図1は本発明をナビゲーションシステムに適用した実施例の全体構成を示す。

不図示の車両に搭載されたクライアント側のナビゲーション装置1と、ナビゲーション装置1からアクセス可能な情報センター2と、情報センター2が通信を行いWEB情報(インターネット上のドキュメント情報)の取得を行うためのインターネット3とよりナビゲーションシステム5が構成される。

本実施例ではナビゲーション装置1によりナビゲーション情報として施設の案内を行う場合を例に説明する。

【0010】

まず、ナビゲーション装置1は、マイクロフォン10と、AD変換部11と、演算部12と、認識辞書データベース(認識辞書DB)13と、通信部14と、DA変換部15と、スピーカ・ディスプレイ16と、現在位置取得部50とより構成されている。

マイクロフォン10は、ユーザが検索のために発話した音声を取得するためのものである。

現在位置取得部50は、不図示のGPS(Global Positioning System)信号受信アンテナによって受信したGPS信号などより、自車両の現在位置を取得するものである。

【0011】

AD変換部11は、マイクロフォン10で取得された信号をアナログ信号からデジタル信号に変換し、演算部12に出力する。

なお、マイクロフォン10はユーザからの音声信号の他、車両内外で発生している雑音やオーディオ、ナビゲーションシステム等から出力される音信号も拾う可能性があるので

10

20

30

40

50

、雑音等を排して音声信号を精度よく取得するための手段を併設するのが好ましい。

【0012】

その例としては、(1)マイクロフォン10とAD変換部11、もしくはAD変換部11と演算部12との間に、音信号中の非目的成分を弱めるためのフィルタを設ける、(2)ゲイン調整機構を設けて入力信号のパワー(ゲイン)が適切となるよう増幅量を調整する、(3)入力信号のパワー変化等に基づき入力信号中で音声の含まれる区間を正確に切り出す音声抽出(VAD: Voice Activity Detection)機構を設ける、(4)オーディオシステムやナビゲーションシステムから出力される音信号を打ち消すエコーキャンセリング機構を設ける、(5)指向性の強いマイクロフォンを用いたりマイクロフォンを複数用いたマイクロフォンアレーなどを用いて、音信号の種類や方向
10

【0013】

演算部12は、CPU、MPU、DSP、FPGA等の一般的な動作回路を組み合わせ、音声検索部17と、辞書更新部18と、ナビゲーション機能部19とより構成されている。

音声検索部17は、マイクロフォン10を介してユーザから取得した発話音声の一部、あるいは全てについて、認識辞書データベース13に記憶された語彙(キーワード)との比較・照合を行い、一致度の高い尤もらしい語彙の一つあるいは複数を認識結果として取得する。
20

【0014】

なお、音声検索部17による、音声認識処理における発話音声と認識辞書データベース13に記憶された情報との比較には、特徴量を用いることができる。特徴量としては、LPCケプストラム、LPCデルタケプストラム、メルケプストラム、対数パワー等を組み合わせた時系列ベクトルデータが知られている。

音声検索部17はまた、ユーザの音声および手操作による施設検索利用履歴を監視し、地域別に利用頻度を記録する機能を備えている。

【0015】

辞書更新部18は、認識辞書データベース13に記憶された関連キーワード(詳しくは後述する)の更新タイミングの判断、更新内容・更新範囲の制御、通信部14を用いた更新情報の取得、そして認識辞書データベース13の更新処理を行うものである。
30

通信部14が扱う通信網としては、携帯電話網、UWB、DSRC、無線LAN、地上波データ放送、衛星放送波等があり、通信を行うために適切なものを選択すればよい。

【0016】

ナビゲーション機能部19は、基本機能として、走行履歴等を蓄積・解析する機能を有する。そして、音声検索部17によって認識された音声を目的地として受け付け、さらに目的地の履歴を記憶したり、例えば経由地として入力された登録位置を受け付け、また認識辞書データベース13に記憶された施設名などを目的地や登録位置として受け付けて、現在位置取得部50によって取得された自車両の現在位置などにもとづいて、自車両の経路を案内する案内情報を生成する。
40

【0017】

認識辞書データベース13は、音声検索部17による音声認識処理において、認識対象となる初期キーワードおよび該初期キーワードと関連性の高い語彙である関連キーワードを、その関連度と共に記憶するものである。

認識辞書データベース13としては、キャッシュメモリ、メインメモリ、HDD、CD、MD、DVD、光ディスク、FDD等、一般的な記憶媒体が使用される。

【0018】

認識辞書データベース13には、図2に示すように、初期キーワードとしての施設名称に付随する形で、その施設種別、経度・緯度の位置情報、郵便番号、住所、電話番号が記録され、更に、関連キーワードが記録される。
50

なお、図2の関連キーワードの欄は、追加されていない状態を示してあり、全てNULLとなっている。

また本実施例において、認識辞書データベース13は関連キーワードが更新されるたびに記憶内容が変更されるため、更新毎にデータベースのバージョンを示す辞書バージョンIDが記録される。

【0019】

音声認識処理では、音響辞書と言語辞書とから構成される認識辞書を用いて音声の比較・照合処理が行われる。

音響辞書は、音響モデルとも呼ばれ、言語の音響的特徴量を、例えば隠れマルコフモデル(HMM: Hidden Markov Model)等の形式で記録したものである。

一方の言語辞書は、言語モデルとも呼ばれ、認識させるキーワードを、単語の列記、あるいは単語の接続の定義として記録したものである。

【0020】

さらに言語辞書としては、認識させるキーワードおよびその接続関係をネットワークとして明示的に記述するネットワーク型言語モデルと、個々の単語の出現確率および単語同士の接続確率を膨大なテキストデータ(テキストコーパス)の解析から自動的に決定する統計確率型言語モデル(統計モデル、n-gram)が知られている。

本実施例では、言語モデルの形式に制限されるものではないが、認識させたいキーワードは案内可能な施設に限定するため、ネットワーク型言語モデルが比較的に適している。ネットワーク型言語モデルに用いられる認識語彙は、目的地を含む地域名あるいは目的施設のジャンル等から最終の目的地に至る階層構造で表現される。

【0021】

ここで、言語モデルの一例を図3に示す。

この例では、ネットワーク型の言語モデルとして、図2の認識辞書データベースのうち、住所の都道府県、市区町村、初期キーワード、関連キーワードの項目を自由に組み合わせた発話が認識可能なように構成されている。なお、ガベージモデル(Garbage)と組み合わせられて、発話中の間投詞や言い直しなどはスキップされる。

本言語モデルによって、例えば、「神奈川県」のように、複合検索条件を認識することが可能であり、その認識結果を「検索キー」として、認識辞書データベースから候補を抽出することで、所望の検索結果すなわち施設名称を得ることが可能である。

なお、同様にして、電話番号や郵便番号を組み合わせた発話を認識する言語モデルを生成することも可能である。

【0022】

また本実施例では、関連キーワードを更新することを特徴とし、具体的には図2の関連キーワード(図2では全てNULLとなっている)を更新することに相当する。また、認識辞書データベースが更新を受けた場合には、図3に示した言語モデルの関連キーワード(図3では全て空値(NULL)となっている)も自動的に更新される。

【0023】

スピーカ・ディスプレイ16は、音声検索部17による検索結果をユーザに対し音声または映像によって提示するものである。

また検索結果の提示後は、ナビゲーション機能部19は提示した内容に応じてさらにユーザからの入力を促し、入力に応じた制御に遷移できる構成となっている。

例えば音声検索部17による検索の結果、複数の施設が候補として得られ、該候補をスピーカ・ディスプレイ16を通じて提示してユーザに選択を促したあと、ユーザの選択によって一意の施設に絞られた場合には、ナビゲーション機能部19は選択された施設の詳細説明情報の提示、地図の表示、ルート案内、電話機能などの制御を行う。

【0024】

情報センター2は、通信部20と、関連語彙収集部21と、関連キーワードデータベース22とを備える。

10

20

30

40

50

通信部 20 は、ナビゲーション装置 1 の通信部 14 と通信を行うためのものである。

関連語彙収集部 21 は、インターネット 3 と接続可能であり、インターネット 3 上に存在する WEB 情報 30 を取得可能となっている。

具体的には関連語彙収集部 21 は、WEB 情報 30 から、初期キーワードをキーとしたテキスト解析を行い、初期キーワードに関連性の高い関連キーワードを抽出する。

【0025】

また関連語彙収集部 21 は、ナビゲーション装置 1 から送信された辞書更新部 18 からの更新要求を通信部 20 を通じて受け、更新要求に対して抽出した関連キーワードを通信部 20、通信部 14 を通じて辞書更新部 18 へ返信する。

なお、関連語彙収集部 21 は、情報センターにおいて手入力によりインターネット検索が行われた際の検索キーワードと到達した WEB ページの記録を蓄積するようになっている。

この関連語彙収集部 21 の詳細な関連語彙収集方法については後述する。

【0026】

関連キーワードデータベース 22 は、初期キーワードと、それに付随する関連キーワードなどが記録される。この初期キーワードは、上述の認識辞書データベース 13 が保持する初期キーワードと同一である。

関連キーワードデータベース 22 が記録するデータ的具体例を図 4 に示す。図 4 の例では、初期キーワード、関連キーワードのほか、施設種別および住所、関連キーワードの関連度、バージョン ID などが記録されている。

これらのデータは関連キーワードの新たな収集および更新処理を効率的に行うために用いることができる。

【0027】

次に、関連語彙収集部 21 における具体的な関連キーワード収集処理として、WEB ページそのものに含まれるテキスト情報を用いて収集する例について説明する。

ここでは、図 4 に示した関連キーワードデータベース 22 に登録された初期キーワードを検索キーとして、インターネット 3 上の WEB 情報 30 の検索を行う。

検索結果として取得された WEB 情報 30 の WEB ページのうち、上位所定のページ（例えば上位 20 ページなど）を解析対象とし、WEB ページ内のテキスト情報の解析処理を行う。

なお初回の関連語彙収集時点では、図 4 における関連キーワード、関連度、およびバージョン ID（タイムスタンプ）のデータは NULL となっている。

【0028】

ここで、WEB 情報 30 の WEB ページから解析したいものは、初期キーワードである施設名称についての関連キーワードである。

したがって、解析すべき WEB ページは、該施設について言及しているページであることが望ましい。

しかしながら、検索結果の WEB 情報中には対象とすべきではない WEB ページが含まれていることがある。

例えば、「A 寺」という寺社についての関連キーワードを抽出するために、「A 寺」をキーとして WEB 検索を行った場合、「A 寺駅」、「A 寺高校」のような、初期キーワードを包含する別の施設の WEB ページが該当してしまうことがある。また、「A 寺」と同名の地域名称「X 県 Y 市 A 寺」が存在する場合も考えられる。

【0029】

こうした WEB ページを「A 寺」用の関連キーワード解析対象としてしまうと、「A 寺」にはふさわしくない関連キーワードが抽出されてしまう可能性がある。

そこでこれを避けるために関連語彙収集部 21 は、施設種別や住所といった情報を付随して収集し、これらの施設種別や住所などの情報を用いて、検索で得られた WEB ページについてフィルタリング（ふさわしくない WEB ページを解析対象から除外する処理）を施す。

【 0 0 3 0 】

具体的には「A寺」の例で言えば、「寺」以外の施設種別である「学校」や「駅」といったテキストが所定値以上の頻度で出現するWEBページは、「A寺」について言及するページではない可能性が高いため、解析から排除する。

同様に、住所情報でチェックすることにより、「X県Y市A寺」のようなテキストが所定値以上の頻度で出現するページは「A寺」について言及するページではない可能性が高いと見て排除する。

【 0 0 3 1 】

上記フィルタリング処理により、解析対象となるWEBページが決定したら、関連語彙収集部21はさらに、該WEBページ内に含まれるテキスト情報の解析処理を行って、関連キーワードを抽出する。解析方法としては、一般的な形態素解析、構文解析で用いられる方法が適用できる。

関連キーワードの抽出にはいくつかの方法がある。

第1の方法は、テキスト解析で得た単語の出現頻度によるものである。

すなわち、単語分割処理により、文章を単語に分割、品詞を振り分け、間投詞、助詞、接続詞のような不要語を消去して、名詞、動詞、形容詞など施設の別名や施設へのイメージに繋がるキーワードとなりうる品詞のみを抽出し、単語(語彙)毎の頻度を算出する。

そして、解析対象となる全てのWEBページについて、単語の頻度を積算した結果、頻度が所定以上となった単語、あるいは全てのページのテキスト量から相対値として見た出現頻度が所定以上となった単語を関連キーワードとして抽出する。

ここで、検索キーとして用いた初期キーワードはもちろん除外される。

【 0 0 3 2 】

なお、単純に出現頻度をカウントすると、時として関連キーワードとしてはふさわしくない単語が高頻度に出現していることが考えられる。このため、以下の種々の処理を組み合わせ、抽出精度を高めることができる。

例えば、施設の営業時間などがWEBページに記載されていることが多く、「営業時間」、「拝観時間」、「時～時」といったキーワードが高頻度に計上され、関連キーワードになってしまう場合がある。

そこで、こうした例外語彙を予め用意しておき、定期的に検査して例外語彙をメンテナンスすることで、これらを排除するようにすることが好ましい。

【 0 0 3 3 】

第2の方法としては、抽出精度を高めるため、初めから「特定の言い回しを伴って出現した単語」のみをカウントするものである。

特定の言い回しとしては、「～と云えば」、「～で(が)(として)有名」、「～で(が)(として)著名」、「～で(が)(として)名高い」、「～の代名詞」、「～で(として)親しまれる」、「～で(として)知られる」、「～と(とも)呼ばれる」、「別名～」、「別称～」、「愛称～」、「略称～」、「通称～」、「略して～」等が考えられる。

これらの言い回しと共に出現した単語(上記の“～”に相当する単語)のみをカウントすることで、別名や施設のイメージとなる関連キーワードを精度よく絞ることができる。

【 0 0 3 4 】

さらに、第3の方法として、初期キーワードに対する関連キーワードの関連性を「関連度」として算出し、その積算値によって関連キーワードの選択を行うことができる。

関連度は、「ある施設を探す時に連想する言葉」を反映した数値であり、関連度が高い単語とは「検索時に発話しやすいキーワード」と言える。この関連度は、上記の出現頻度に加え、上述の特定の言い回しを伴っての出現や、後述する各種情報を加味して算出される値であり、抽出された関連キーワードの優先順位を決める「重み係数」と捉えることができる。

例えば関連キーワードデータベース22の記憶容量は有限のため、関連キーワードを無数に登録することはできない。そこで、関連度が高いものから採用(記憶)することによ

10

20

30

40

50

って、ユーザの利便性と記憶容量の両立が可能となる。

【0035】

関連度の算出の具体例としては、(1)WEB情報30上で、特定の言い回しを伴い出現したキーワードを、それ以外のキーワードより関連度を高く設定する、(2)WEBページ中の出現位置の情報を加味し、ページ上段に位置するキーワードほど関連度を高く設定する、(3)WEBページ全体の文章レイアウト、段落構成から見出し部分を判別し関連度を高く設定するなどがある。

さらに、(4)テキスト中に対象の初期キーワードそのものが存在した場合に、該初期キーワードからの距離、すなわち、初期キーワードと対象キーワードとの間に含まれる単語数、あるいは文字数が少ない程、関連度を高く設定する、(5)大きなフォントサイズ、ボールド体フォント、下線、文字色が周囲と異なる等、強調表現がなされているキーワードに対して、関連度を高く設定するなどもある。

10

【0036】

上記の関連度設定方法を反映した関連度の算出式を以下に示す。

キーワード“KW”のカウント値をCount(KW)、i番目に現れたキーワード“KW”をKW(i)で表すと、

$$\text{Count}(KW) = \sum_{i=1}^n (1 + \text{IF}(KW(i)) * \dots + \text{IF}(KW(i)) * \dots + \text{IF}(KW(i)) * \dots)$$

となる。

ただし、IF(KW(i))、IF(KW(i))、IF(KW(i))・・・は、上述したキーワード(KW(i))について関連度を高く設定する際の条件とする。

20

関連度を高く設定する際の条件が成立している場合には、成立している条件(IF(KW(i))、IF(KW(i))、IF(KW(i))・・・)を1とする。

したがって、ある条件が成立した場合には、そのキーワード(KW)のカウント値(Count(KW))にはボーナス、・・・が加算される。

一方、どの条件にも当てはまらず、単純にキーワード“KW”が出現していた時には、ボーナス部分は全て0となるため、1がカウント値(Count(KW))として算出される。

【0037】

こうして得られたキーワード“KW”のカウント値(Count(KW))を、次式のように全てのキーワード(KW=1~n)に対するカウント値で割ることで、キーワード“KW”の関連度を算出する。

30

$$\text{Cor}(KW) = \text{Count}(KW) / \sum_{KW=1}^n (\text{Count}(KW))$$

なお、ボーナス、・・・はWEBテキスト解析を行いながら実験的に算出することで、条件(関連度を高く設定するための条件)毎に調整するのが望ましい。

図4および後述する図5の関連キーワードデータベース22のデータは上記の関連度によって抽出した例を示している。

関連度はまた、音声検索部17がマイクロフォン10を通じて入力されたユーザの音声を認識する際に、この関連度を加味し、関連度の高い関連キーワードが、関連度の低いキーワードよりも認識されやすくするといった利用も可能である。

40

【0038】

次に、関連語彙収集部21が、関連キーワードを収集するタイミングについて説明する。

本実施例においては、関連語彙収集部21が情報センター2に配置され、車載のナビゲーション装置1とは分離された構成であるため、関連キーワードの収集処理を常時実行することが可能であり、また任意のタイミングでも可能である。

【0039】

一例として、関連語彙収集部21は、上述したWEB情報の検索結果自体に変化がないか、そしてWEBページのテキスト情報に変化がないかを、例えば1日一回等、定期的に行い監視するようにしてもよい。そして、変化があった場合には、変化のあったWEB

50

ページについてのみ、テキスト解析を行う。

前回の検索で抽出されたWEBページのアドレスおよびそのページ内容をキャッシュしておいて、新規検索の結果と前回までの検索結果の差分をとることにより、差分となったWEBページについて解析処理を行い、新規収集の関連キーワードとして更新することができる。

【0040】

なお、WEBページのテキスト内容自体をキャッシュして比較するには大きな記憶容量が必要となるが、インターネットで活用されているWEBページの更新情報の表現形式であるRSS(Rich Site Summary)等を使えば、当該WEBページの更新部分のみを知ることが可能である。

10

【0041】

さらに他の収集タイミングとしては、情報センター2の十分に計算速度の速いサーバ機能および十分に早い通信環境を前提として、ナビゲーション装置1側から辞書更新命令を受けたタイミングで関連語彙収集部21がWEB検索を行って関連キーワード抽出処理までを実行し、その結果をナビゲーション装置1側に返答するようにしてもよい。

【0042】

次に、ナビゲーション装置1の辞書更新部18における辞書更新方法について説明する。

ここでは、更新する情報の受け渡しを通信部14、20を介して行うので、通信量削減の観点から、差分更新を行う。

20

図4に示す関連キーワードデータベース22に記憶された情報では、関連キーワードごとに、バージョンIDとしてタイムスタンプ情報が記録されており、図2に示す認識辞書データベース13には現在の辞書バージョンIDが記録されているので、関連キーワードのそれぞれのバージョンを比較することで、差分のみを通信することができる。

【0043】

このため、辞書更新部18は、辞書更新の命令とともに、認識辞書データベース13の辞書バージョンIDを情報センター2の関連語彙収集部21へ送信する。

関連語彙収集部21では、関連キーワードデータベース22を参照し、受信した認識辞書データベース13の辞書バージョンIDと、関連キーワードデータベース22の各関連キーワードのバージョンIDとを比較し、関連キーワードデータベース22内の認識辞書データベース13の辞書バージョンIDよりも新規に更新された関連キーワードを抽出する。

30

【0044】

図2に示した認識辞書データベース13の例では辞書バージョンIDが20071201000000であるため、図4に示した関連キーワードデータベース22からはバージョンIDが斜線を付した20071201000000以上の関連キーワードを更新対象として抽出する。

対象となる関連キーワードを抽出すると、関連語彙収集部21は当該関連キーワードを関連語彙更新情報としてナビゲーション装置1の辞書更新部18へ送信する。

辞書更新部18は、受信した関連キーワードを認識辞書データベース13に登録する。

40

【0045】

この結果、例えば図5に示すように、辞書バージョンIDが最新となり、新たな関連キーワード(図5中で太字で示す関連キーワード)が追加される。

追加された関連キーワードを用いた言語モデルの例を図6に示す。図3の言語モデルではNULLであった関連キーワードの登録語彙に、新たな関連キーワード(あじさい寺、横浜マリノス・・・)が追加されており、該キーワードを認識語彙とした音声検索部17による音声検索が可能になる。

【0046】

なお辞書更新の頻度、あるいはタイミングであるが、全国の施設は無数にあるため、全ての施設に対して関連キーワードを取得するためには、記憶容量、通信速度、データ更新

50

処理速度等、全ての条件が高次元で達成されていなければならない。

しかしながら、カーナビゲーション装置での利用の場合には、ある特定のユーザによって全国の施設が目的地に設定される可能性は高くない。

また、解析された関連キーワードも、1日のうちに何度も更新する必要性はない。

そこで、住む地域やユーザの個人性によって、検索される地域をある程度限定でき、更新の頻度も、ナビゲーション装置1が利用される1トリップについて1回程度で十分と考えられる。

【0047】

したがって、車両のエンジン始動のためイグニションスイッチがONされ、ナビゲーション装置1が起動した時点を通常の更新タイミングとする。

なお、情報センター2は常時作動しているため、ナビゲーション装置1の起動がナビゲーションシステムとしての起動となる。

1日のうちに複数回イグニションスイッチのON、OFFを繰り返すケースでは、日付が同じであれば更新しないようにしてもよい。

【0048】

また、更新地域を、現在認識辞書データベース13に保持している関連キーワードの地域範囲の一部であるユーザの住む地域やよく利用する地域に限定し、該地域を越えた時を更新タイミングとする。

例えば、図7に示すように、エンジンが始動されてナビゲーション装置1が起動した地点(あるいは自宅として登録してある地点)をP00、過去に検索した施設あるいは登録してある地点をP01、P02、P03であるとするとき、起動時点P00では、各地点P00~P03から所定距離の地域を更新範囲A00、A01、A02、A03とする。

さらには、地点P00から地点P01、P02、P03までのルートをR01、R02、R03とし、この地域についても更新範囲としてもよい。

【0049】

このように、ナビゲーション装置起動時に辞書更新部18は、情報センター2の関連語彙収集部21と通信を行うことによって、更新範囲A00、A01、A02、A03やルートR01、R02、R03を対象範囲として、認識辞書データベース13における関連キーワードの更新を行う。

なお図7の例では、各地点P00~P03から所定距離地域、すなわち円状の領域を更新範囲A00~A03としているが、各地点が所属する都道府県やその近隣県といった行政区画も更新範囲としてもよい。

そのほか、地点が登録されていなくとも、走行履歴にもとづいて利用地域を解析し、該地域を更新対象地域とすることも可能である。

【0050】

次に、辞書更新について、とくに辞書更新部18と関連語彙収集部21間における具体的な処理を図8のフローチャートを用いて説明する。

ここでは、現在位置、登録位置、履歴位置情報に基づいて、関連キーワードの更新処理を行うものとする。

前述のように、この更新処理はイグニションスイッチがONされてナビゲーション装置1が起動すると開始される。

まずステップ100において、辞書更新部18は、現在位置取得部50より自車両の現在位置を取得する。

【0051】

ステップ101において辞書更新部18は、起動後にすでに辞書更新を実行したかどうかをチェックする。これは辞書更新の有無を示すフラグを用いるなどにより判断できる。今回が初めての辞書更新処理であれば、ステップ103へ進み、すでに辞書更新処理を1回行っていればステップ102へ進む。

ステップ103では、辞書更新部18は、ナビゲーション機能部19より登録位置、目的地の履歴位置情報を取得する。

10

20

30

40

50

そしてステップ104において、辞書更新部18は、取得した登録位置、履歴位置を、音声検索部17が施設検索利用履歴に基づいて記録している地域別の利用頻度を参照して、利用頻度が所定値以上のものに絞る。

【0052】

ステップ105において辞書更新部18は、認識辞書データベース13に記録された辞書バージョンIDを取得する。

このあと、ステップ106において、辞書更新部18は、取得した現在位置、認識辞書データベース13の辞書バージョンID、および利用頻度が所定値以上のものに絞った登録位置、履歴位置情報の各データを情報センター2の関連語彙収集部21へ送信する。

【0053】

ステップ107において関連語彙収集部21は、辞書更新部18から受信した現在位置、登録位置、履歴位置情報にもとづき、関連キーワードデータベース22を参照して更新対象となるデータ範囲を選択する。

例えば、送信された情報に含まれた施設名称、緯度・経度情報や都道府県情報などの初期キーワードが、関連キーワードデータベース22に記憶された情報に含まれていれば、当該初期キーワードを含む情報を更新対象データ範囲として選択する。

ステップ108において関連語彙収集部21は、ステップ107で選択した更新対象データ範囲に該当する関連キーワードのバージョンIDを読み出す。

【0054】

ステップ109において関連語彙収集部21は、関連キーワードのバージョンIDと、受信した認識辞書データベース13の辞書バージョンIDとを比較し、差分データが存在するかどうかを判断する。

具体的には、認識辞書データベース13の辞書バージョンIDよりも新しい関連キーワードのバージョンIDが存在する場合にはステップ110へ進み、存在しないときはステップ100へ戻る。

ステップ110において関連語彙収集部21は、認識辞書データベース13の辞書バージョンIDよりも新しいバージョンIDを有する関連キーワードを更新対象の関連キーワードとして抽出する。

【0055】

ステップ111において関連語彙収集部21は、ステップ110で抽出した関連キーワードをナビゲーション装置1の辞書更新部18へ送信する。

ステップ112において辞書更新部18は、関連語彙収集部21から受信した更新対象の関連キーワードを認識辞書データベース13に新規の関連キーワードとして追加する。同時に、認識辞書データベース13の辞書バージョンIDを、受信した関連キーワードの最新のバージョンIDで更新する。

ここで、ナビゲーション装置1は音声検索部17による音声検索の待機状態となる。

ステップ112のあと、ステップ100へ戻る。

【0056】

次のフローにおいては、すでに辞書更新を実行しているので、ステップ101からステップ102へ進む。

ステップ102において、辞書更新部18は、ステップ100で取得した自車両の現在位置が、予め設定してある所定の領域を超えたかどうかをチェックする。

自車両の現在位置が当該所定の領域を超えていないときはステップ100へ戻り、所定の領域を超えたときはステップ103へ進む。

以上の流れにより、ナビゲーション装置1が起動するとまず1回だけ認識辞書データベース13における関連キーワードの更新が実行され、その後ナビゲーション装置1が作動中の間は自車両の現在位置が所定の領域を超えたときにあらためて関連キーワードの更新が実行されることになる。

【0057】

なお、関連キーワードが更新され増加するに伴い、ある関連キーワードに対して複数の

10

20

30

40

50

初期キーワードすなわち施設名称が該当する、いわゆる関連キーワードの重複が起こる可能性が出てくる。

この場合には、音声検索部 17 は認識された関連キーワードでデータベースを検索した際に、重複した初期キーワード、すなわち複数の施設名称をスピーカ・ディスプレイ 16 により提示して、対話形式でユーザが選択できるようにする。

【0058】

本実施例においては、マイクロフォン 10 および A/D 変換部 11 が発明における音声入力部を構成し、ナビゲーション機能部 19 が案内情報生成手段に該当し、D/A 変換部 15 およびスピーカ・ディスプレイ 16 が出力部を構成している。

【0059】

本実施例は以上のように構成され、音声認識を行なってユーザの求める情報を提供するナビゲーションシステムにおいて、WEB 情報 30 から初期キーワードに関連する関連キーワードを収集する関連語彙収集部 21 と、その収集結果を認識辞書データベース 13 に反映する辞書更新部 18 を持つ構成となっているため、音声検索部 17 によって認識辞書データベース 13 に記憶された初期キーワードおよび関連キーワードを用いた音声検索が可能になり、正式な語彙を知らないユーザであっても、所望の検索結果を得ることができる。

また、あらかじめ関連キーワードを登録する必要が無く、ユーザにとって必要度合いが高いと思われるものに関して、自動で認識辞書データベースの生成、更新を行うことができる。

初期キーワードが施設の名称であるときに、ユーザがその正式な名称を知らない場合であっても、所望の施設を検索することができ、当該施設への経路案内にとくに有効である。

【0060】

とくに、関連語彙収集部 21 が関連キーワードを収集する際に、WEB 検索において出現頻度が所定以上のものを抽出し、または特定の言い回しを伴って出現したキーワードを抽出することにより信頼性の高い関連キーワードを収集することができる。

【0061】

また、初期キーワードの施設名称 A に対して、該施設名称 A を包含する別の施設名称 B あるいは該施設名称 A を包含する地域名称 C が存在する場合には、施設名称 B あるいは地域名称 C を用いて検索された WEB ページを解析対象から除外するものとしたので、施設名が類似の別施設の WEB ページや、例えば「国分寺」という寺社に対する「国分寺市」のように施設名と同名の地域の WEB ページが存在する場合に、本来の施設の WEB ページ以外のページが分析対象外となり、より信頼性の高い関連キーワードを収集することができる。

【0062】

さらに、初期キーワードの施設名称に対応する施設種別以外の施設種別名が所定頻度以上含まれるページを解析対象から除外するものとしたので、施設種別が異なる同名の施設名称が存在しても、それぞれ該当する施設種別以外の WEB ページが分析対象外となり、信頼性を向上させることができる。

【0063】

また、関連語彙収集部 21 が、関連キーワードを収集する際に、その表示位置および表示形式に基づく重み付けを施して算出された関連度を用いることにより、より信頼性の高い関連キーワードを収集することができる。

より具体的には、関連度算出時の重み付けの値について、出現する語彙の表示位置がページ内のタイトル、見出し、およびページ先頭に近い位置にある程大きい値とすることにより、WEB ページに含まれる情報のうち、重要な情報を優先的に抽出ことができ、関連キーワードの信頼性を向上させることができる。

【0064】

また、関連度算出時の重み付けの値について、出現する語彙の表示位置が初期キーワー

10

20

30

40

50

ドの出現位置からの距離が小さいほど大きい値とすることにより、初期キーワードに近接する関連キーワード程優先的に収集され、関連キーワードの信頼性を向上させることができる。

出現位置からの距離として、少なくとも初期キーワードから対象の語彙までに含まれる文字数、形態素数、句数、文数、段落数のいずれかとすることにより、テキスト情報中において、初期キーワードに近接する関連キーワード程優先的に収集することができ、関連キーワードの信頼性を向上させることができる。

【0065】

また、出現する語彙の書式、すなわち文字サイズ、下線の有無、書体等が通常と異なる強調書式であるとき、重み付けを大きく設定することにより、テキスト情報中において、強調された関連キーワードを優先的に収集でき、関連キーワードの信頼性を向上させることができる。

10

【0066】

さらに、他の関連度の重み付けとして、特定の言い回しに伴って出現した語彙について重み付けを大きく設定することにより、特定の言い回しに伴って出現するような、定型表現的な関連キーワードを効率よく収集することができる。

とくに、特定の言い回しとして、「～と言えば」、「～で(が)(として)有名」・・・などは、施設名称等の初期キーワードの別称等を述べるのに典型的に用いられる表現形式であることから、こうした定型句として現れる語彙の重み付けを大きくすることにより、関連キーワードを効率よく収集することができる。

20

【0067】

また辞書更新部18は、ナビゲーション装置1の起動時あるいはナビゲーション装置1を搭載した車両の現在地が所定の領域を超えた時に、認識辞書データベース13を更新するものとしたので、認識辞書データベースのメモリ容量を抑え、更新回数の不要な増大を抑えながら、最新の関連キーワードを認識辞書データベース13に追加することができる。

【0068】

また、認識辞書データベース13の更新は、認識辞書データベース13に記憶された関連キーワードと関連語彙収集部21が収集した関連キーワードを比較し、変化があった場合に関連語彙収集部21が収集した関連キーワードを用いて認識辞書データベース13を更新するものとしたので、最新の関連キーワードがある場合にのみ更新すれば済む。

30

とくに、辞書更新部18からは認識辞書データベース13の辞書バージョンIDを関連語彙収集部21へ送信し、関連語彙収集部21は受信した辞書バージョンIDと関連語彙収集部21が収集した関連キーワードのバージョンIDを比較して、辞書バージョンIDより新しいバージョンIDの関連キーワードを抽出して辞書更新部18へ送信するので、辞書更新部18と関連語彙収集部21間の通信量も低減することができる。

【0069】

さらに、更新のため比較する関連キーワードは、登録されまたは過去に検索した地点から所定範囲内に存在する施設名称とすることにより、ユーザが検索する可能性の大きい地域の関連キーワードに限定して認識辞書データベース13を更新することができ、最新の関連キーワードによる音声検索を、低い更新コストで提供することが可能になる。

40

【0070】

また音声検索部17がユーザの施設検索利用履歴を監視して地域別の利用頻度を記録しているため、辞書更新部18は利用頻度の高い地域についてのみ関連キーワードを更新することにより、ユーザの施設検索の利用実態に合わせ、最新の関連キーワードによる音声検索を提供することができる。

【0071】

また、音声検索部17による音声検索時に、同一の関連キーワードを持つ初期キーワードが存在する場合に、該複数の初期キーワードを選択提示するものとしているので、ユーザは関連キーワードで検索した結果が複数ある場合でも、提示された選択肢から選択する

50

ことで、所望の検索結果を得ることができる。

【0072】

また、関連語彙収集部21をナビゲーション装置1から離隔した情報センター2に置くことにより、関連語彙収集部21では常時など任意のタイミングで最新の関連キーワードを収集、保持しつつ、クライアントのナビゲーション装置1は利用開始時あるいは必要時のみ認識辞書データベース13更新処理を行うので、低い更新コストで、最新の音声検索用の認識辞書データベースを保持することができる。

【0073】

なお、実施例ではナビゲーション装置1側からの要求で更新処理を開始するが、例えば情報センター2側の関連語彙収集部21から、放送波等に更新情報を含ませるなどして、任意のタイミングでナビゲーション装置1へ関連キーワードの更新情報を送信する構成としてもよい。

10

【0074】

次に変形例について説明する。

上記実施例は、クライアント側のナビゲーション装置1に音声検索機能を持たせ、情報センター2に関連キーワードの収集および蓄積処理を持たせたクライアントサーバ型の構成としたが、この変形例は、情報センターの機能をナビゲーション装置に統合してスタンドアロン型としたものである。

図9は変形例にかかるナビゲーション装置の全体構成を示す。

図1の実施例に対して、関連キーワードデータベース22を削除し、ナビゲーション装置1Aに関連語彙収集部21を持たせて、ナビゲーション装置1Aでナビゲーションシステムを構成している。

20

【0075】

辞書更新時には、ナビゲーション装置1A自身の関連語彙収集部21がインターネット3のWEB情報30に接続して関連キーワードを収集し、その結果を用いて直接認識辞書データベース13を更新して音声認識を行うように構成される。

ナビゲーション装置と情報センター間の通信も不要なため、通信部14、20も削除される。

その他の構成は実施例と同じであり、図1と同じ構成については、同一番号を付して説明を省略する。図8のフローチャートのステップ106、111における送信は、通信部14、20を介さない直接の送信と読み替える。

30

【0076】

本変形例においても、実施例と同様に、音声検索部17によって認識辞書データベース13に記憶された初期キーワードとその関連キーワードを用いた音声検索が可能になり、正式な語彙を知らないユーザであっても、所望の検索結果を得ることができる。

その他、実施例と同じく、段落[0059]～[0071]に記載した効果を奏する。

また、本変形例はスタンドアロン型の構成のため、収集した関連キーワード等の通信に伴う制約から解放されるという利点を有している。

【0077】

なお、上記実施例および変形例における図8のフローチャートでは、辞書更新のタイミングをナビゲーション装置1の起動時あるいは認識辞書データベース13が保持している関連キーワードの地域範囲の一部に設定した所定の領域を自車両が越えた時としたが、それ以外の更新タイミングとして、認識辞書データベース13が現在保持している関連キーワードの地域範囲を越えた時、音声検索が起動された時、音声検索の結果に十分な認識スコア(一般的な音声認識で用いられる、尤度、対数尤度、事後確率、信頼度といった数値)を持つ候補がなかった時、音声検索の結果がユーザから訂正、否定された時、ユーザから明示的な更新要求があった場合、所定時間経過時(定期更新)など、適宜に設定することができる。

40

【0078】

とくに、ナビゲーション装置が交通情報や、差分地図情報、各種インターネット情報等

50

を適宜ダウンロード可能で、さらには自車両の走行履歴を定期的にアップロード可能となっている場合には、そのネットワーク接続のタイミングを利用して、例えば通信量の剰余分を認識辞書データベースの更新に利用すれば、認識辞書データベース13内の情報をより効率的に最新の状態に保つことができる。

【0079】

また実施例および変形例では、初期キーワード自体の更新については説明していないが、一般のナビゲーション装置の地図情報更新の方法に準じて初期キーワードも更新可能である。

ナビゲーション装置では地点データ、施設データ、道路データを差分更新するが、認識辞書データベースに関しても、同様に差分更新することができる。

これにより、新規施設等が音声で検索できるようになり、ナビゲーション装置が案内できる施設が増えることになる。

【0080】

上記実施例および変形例においては、辞書更新部18が現在位置、登録位置、履歴位置情報等で特定される利用地域に存在する施設について関連キーワードの更新処理を行うものとしたが、一旦関連キーワードが更新された場合でも、音声検索部17が監視する施設検索利用履歴を参照して、所定期間利用されない地域が存在した場合には、辞書更新部18は当該関連キーワードを初期化することが望ましい。

これにより、利用されない地域の関連キーワードを検索対象から除外することができるので、利用実態の変化に合わせて認識辞書データベース13を更新しつつ、音声検索部17による認識精度が向上するとともに、キーワードの照合に必要なメモリ領域を節約することが可能である。

【0081】

なお、実施例および変形例の関連語彙収集部21における関連キーワード収集処理では、検索結果上位のWEBページ内に含まれているテキストを解析対象としたが、そのほか、第2の方法として、ユーザがインターネットの検索サイトでWEBページを検索する際には、複数の検索キーを組み合わせて検索することも多いので、WEBページを検索する際のユーザの検索キーワードの情報を解析してもよい。

例えば、あじさいで有名な「A公園」があった場合に、検索キーとして、「A公園 あじさい」で検索したとすると、この情報はユーザが「A公園」と「あじさい」を関連するキーワードとして認識していることの表れであり、関連キーワードを抽出するのに非常に適した情報といえる。

【0082】

そこで、初期キーワード（この場合は「A公園」）と共に利用された検索キー（共起キーワード）を蓄積し、その頻度上位を関連キーワードとして抽出するものとする。抽出基準の頻度値は検索キーの数に応じたパーセンテージなどで設定しておく。

これにより、ユーザが特定の施設を調べようとしたときに、同時に入力されたキーワードのうち、高頻度のものを音声検索に利用できるため、ユーザにとって親和性の高い音声検索機能を提供することができる。

【0083】

関連語彙収集部21における関連キーワード収集処理のさらに第3の方法として、第1の実施例のように初期キーワードを検索キーとしてWEB検索を実行し、検索結果の上位所定数のWEBページに対して、上記第2の方法に準じてインターネットの検索サイトから到達した際に用いた検索キーを蓄積して、その頻度の高いものを上記初期キーワードに対する関連キーワードとしてもよい。

前述の第2の方法では、初期キーワードとともに用いられたキーワードが関連キーワードの候補となったが、第3の方法では、単独の検索キーであっても、当該WEBページに到達する際に用いられた検索キーは全て関連キーワードの候補となる。

【0084】

例えば「A公園」のページに到達する際に用いた検索キーが「あじさい」であれば、こ

10

20

30

40

50

れが関連キーワード候補となる。候補となった検索キーの頻度を解析し、その上位を関連キーワードとして抽出することが可能となる。

これにより、ユーザが初期キーワードのWEBページすなわち所望の施設のWEBページに到達する際に検索エンジンで使したキーワードのうち、高頻度のものを音声検索に利用できるため、ユーザにとって親和性の高い音声検索機能を提供することができる。

【0085】

なお上記実施例および変形例では、認識辞書データベース13の関連キーワード更新に際して、処理量や通信量削減の観点から差分更新を行うものとしたが、これらに制限がなければ認識辞書データベース13内の全データを上書き更新してもよいことはもちろんである。

10

【0086】

また、実施例および変形例では、ナビゲーション装置における認識辞書データベースの更新について説明したが、ナビゲーション装置に限定されず、本発明は例えば電話回線を用いた施設案内や、WEBにおける音声検索のための認識辞書データベースの更新等にも適用可能である。

【図面の簡単な説明】

【0087】

【図1】実施例の全体構成を示すブロック図である。

【図2】認識辞書データベースの例を示す図である。

【図3】言語モデルの例を示す図である。

20

【図4】関連キーワードデータベースの例を示す図である。

【図5】認識辞書データベースの例を示す図である。

【図6】言語モデルの例を示す図である。

【図7】辞書更新対象範囲の例を示す図である。

【図8】実施例における辞書更新処理の流れを示す図である。

【図9】変形例の全体構成を示すブロック図である。

【符号の説明】

【0088】

1、1A ナビゲーション装置

2 情報センター

30

3 インターネット

5 ナビゲーションシステム

10 マイクロフォン

11 AD変換部

12 演算部

13 認識辞書データベース

14、20 通信部

15 DA変換部

16 スピーカ・ディスプレイ

17 音声検索部

40

18 辞書更新部

19 ナビゲーション機能部(案内情報生成手段)

21 関連語彙収集部

22 関連キーワードデータベース

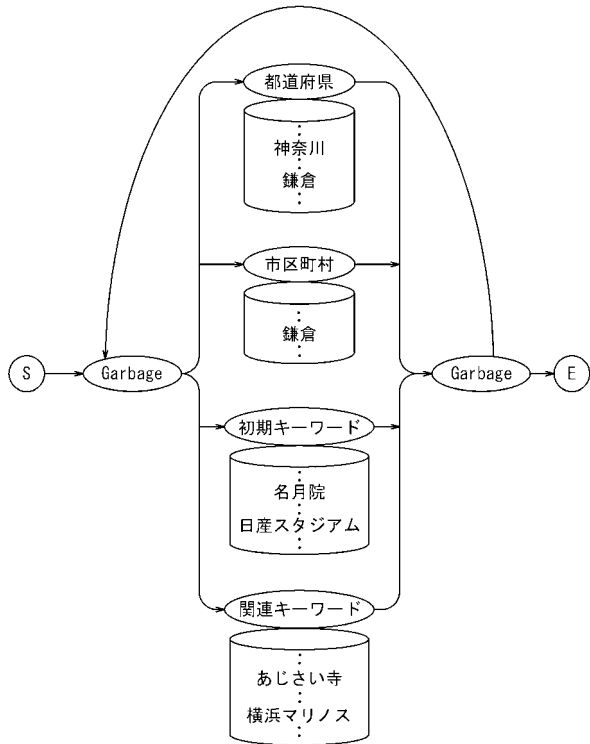
30 WEB情報

50 現在位置取得部

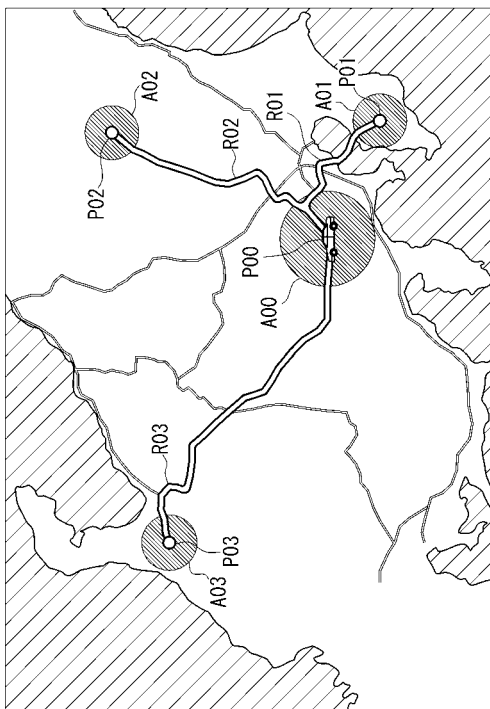
【図5】

辞書バージョンID-20071201020021									
ID	初期キーワード(施設名称)	施設種別	位置情報(経度、緯度)	郵便番号	住所	電話番号	...	関連キーワード	...
1	名月院	寺社仏閣 宗教施設	+139.32.20.09 +035.18.48.76	248-0016	神奈川県 鎌倉市長谷**	046-***-****	...	あじさい(0.02) あじさい寺(0.01)	...
2	日産スタジアム	スタジアム 競技場	+139.36.34.80 +035.30.24.29	222-0036	神奈川県横浜市 港北区小机町**	046-***-****	...	横浜国際競技場(0.01) 横浜マリノス(0.033) サッカー(0.001) Jリーグ(0.005)	...
3	東大寺	寺社仏閣 宗教施設	+135.50.37.95 +034.40.59.61	630-8587	奈良県奈良市 雑司町**	***-***-****	...	大仏(0.008) 大仏殿(0.03)	...
4	平等院	寺社仏閣 宗教施設	+034.53.09.81	606-8225	京都府京都市 伏見区日野西 大連町**	***-***-****	...	鳳凰堂(0.06) 10円玉(0.05) 世界遺産(0.004)	...
5	日本万国博覧会記念公園	公園	+135.32.36.121 +034.48.24.83		大阪府吹田市 千里万博公園	***-***-****	...	万博公園(0.002) 大阪万博(0.1) 太陽の塔(0.1) 岡本太郎(0.016)	...

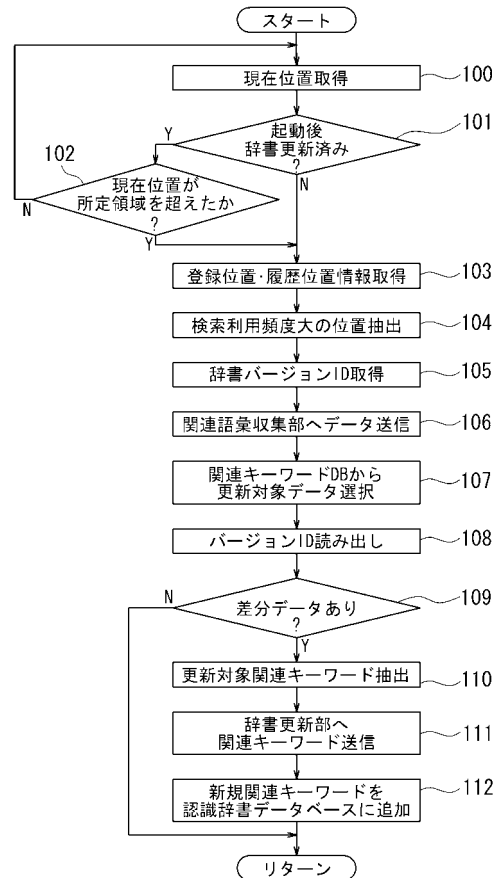
【図6】



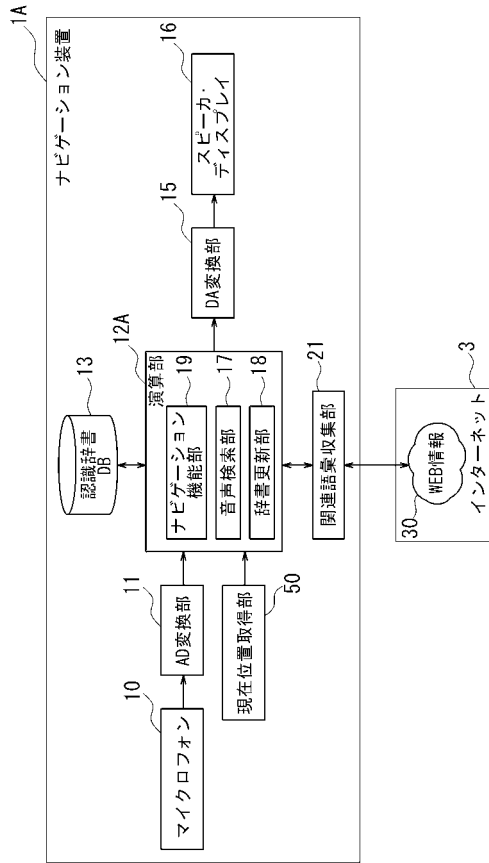
【図7】



【図8】



【図9】



フロントページの続き

(72)発明者 桂川 景子
神奈川県横浜市神奈川区宝町2番地 日産自動車株式会社内

審査官 吉田 誠

(56)参考文献 特開2004-334280(JP,A)
特開2005-338274(JP,A)
特開2006-139203(JP,A)
特開平10-171815(JP,A)
特開2005-327316(JP,A)
特開2001-83983(JP,A)
小原 恭介,ウェブを利用した関連用語収集,FIT2004 第3回情報科学技術フォーラム
一般講演論文集 第2分冊 データベース 自然言語・文書・ゲーム 人工知能 音声・音楽
生体情報科学,社団法人電子情報通信学会,2004年 8月20日,183-184ページ

(58)調査した分野(Int.Cl.,DB名)
G06F 17/30
G08G 1/00-99/00