

明 細 書

発明の名称：情報検索装置および情報検索方法

技術分野

[0001] 本発明は、単語を用いて情報を検索する機能を備えた情報検索装置および情報検索方法に関する。

背景技術

[0002] 情報検索において用いる単語は、例えばキーボードから入力してもよいし、マイクを介して音声で入力してもよい。例えば下記の特許文献1には、入力された音声に対して音声認識処理を行い、認識結果の単語を用いて情報を検索する技術が開示されている。

先行技術文献

特許文献

[0003] 特許文献1：特開平10-21254号公報

発明の概要

発明が解決しようとする課題

[0004] 実際に情報検索を行う場合、一旦、単語を入力した後に、新たな単語を入力（例えば音声入力）することで、(a)新たな単語を追加して検索を行うケースおよび(b)先に入力された単語を新たな単語に置換して検索を行うケースが想定される。

[0005] このとき、新たな単語を追加すべきか、新たな単語を先に入力された単語と置換すべきかは、一意に定まらない。人間であれば、新たな単語を追加した場合の意味と、新たな単語を先に入力された単語と置換した場合の意味とを比較することで、追加と置換のうちどちらが妥当かは分かる。ところが、人手を介さずに上記のような単語の設定処理を簡易且つ適切に行うためには、意味を考えたルールを予め設定しておくことが必須となり、膨大な労力を必要とする。

課題を解決するための手段

[0006] 本発明は、上記の課題を解決するために成されたものであり、意味を考えたルール of の事前設定を必須とすることなく、新たな単語の入力時にクエリ (query) として妥当な単語を簡易且つ適切に設定することができる情報検索装置および情報検索方法を提供することを目的とする。

[0007] 本発明に係る情報検索装置は、単語に関連する情報を検索する情報検索装置であって、単語を入力するための入力部と、所定数の単語が入力された後、新たな単語が入力されたときに、前記新たな単語を先に入力された単語に追加した場合および前記新たな単語を先に入力された単語と置換した場合の単語群パターンを生成するパターン生成部と、生成された単語群パターンそれぞれについて、当該単語群パターンが出現する確率に対応する出現情報を導出する出現情報導出部と、導出された出現情報に基づいて、新たな検索に用いるべき単語群パターンを決定する決定部と、を備えることを特徴とする。なお、決定された「新たな検索に用いるべき単語群パターン」が新たな検索のクエリとなる。

[0008] このような構成の情報検索装置では、所定数の単語が入力された後、新たな単語が入力されたとき、パターン生成部が、新たな単語を先に入力された単語に追加した場合および新たな単語を先に入力された単語と置換した場合の単語群パターンを生成し、出現情報導出部が、生成された単語群パターンそれぞれについての出現情報を導出し、そして、決定部が、導出された出現情報に基づいて、新たな検索に用いるべき単語群パターンを決定する。これにより、さまざまな単語群パターンそれぞれの出現情報に応じて、新たな検索に用いるべき単語群パターンを簡易且つ適切に決定することができる。即ち、意味を考えたルール of の事前設定を必須とすることなく、新たな単語の入力時にクエリとして妥当な単語を簡易且つ適切に設定することができる。

[0009] 上記のような情報検索装置は、さまざまな単語についてのクラス情報を記憶したクラス情報データベースをさらに備え、パターン生成部は、先に入力された単語および新たな単語それぞれのクラス情報を前記クラス情報データベースから読み出し、「新たな単語のクラス情報とは異なるクラス情報の先

に入力された単語」と「当該新たな単語」とを置換した場合を除外して、単語群パターンを生成してもよい。これにより、互いにク拉斯情報の異なる単語同士を置換した単語群パターンを回避でき、見当ちがいな単語群パターンを排除することができる。

[0010] なお、決定部は、導出された出現情報が高い順に単語群パターンを、選択可能なランキン形式で出力し、ユーザにより選択された単語群パターンを、新たな検索に用いるべき単語群パターンとして決定してもよい。この場合、ユーザからみて最適な単語群パターンを選択可能となる。また、決定部は、出現情報が最大である単語群パターンを、新たな検索に用いるべき単語群パターンとして決定してもよい。この場合、ユーザの介入は不要となり、自動的に単語群パターンを決定できる。

[0011] また、単語を入力するための入力部は、さまざまな形態を採用できる。例えば、入力部は、音声を入力するための音声入力部と、入力された音声から単語を認識する認識部と、を含んで構成してもよい。このように単語を音声入力する形態以外に、キーボード等のさまざまな情報入力手段により単語を入力する形態を採用してもよい。また、出現情報導出部は、さまざまな単語群パターンの出現情報を記憶した出現情報データベースを含み、生成された単語群パターンそれぞれの出現情報を出現情報データベースから読み出すことで、当該単語群パターンそれぞれの出現情報を導出してもよい。

[0012] なお、上記の「単語群パターンの出現情報」は、例えば(1)当該単語群パターンを構成する個々の単語の出現確率、(2)当該単語群パターンを構成する複数の単語の共起確率および共起確率から所定手順で導出される情報(例えば、共起確率を個々の単語の出現確率によって順に割り算して得られた値など)、(4)入手でルール化した情報などを含む。また、上述した「所定数の単語」は、1つの単語でもよいし、複数の単語でもよい。そのため、「所定数の単語」が1つの単語のとき、新たな単語を先に入力された単語(1つの単語)と置換した場合の単語群パターンは、「新たな単語のみ」となるが、本発明における単語群パターンにはこのようなケースも含まれ

る。一方、「所定数の単語」を複数の単語としてもよく、この場合、出現情報導出部は、出現情報として、共起情報を導出する。この「共起情報」は、例えば、上述した(2)当該単語群パターンを構成する複数の単語の共起確率、(3)上記出現確率および共起確率から所定手順で導出される情報（例えば、共起確率を個々の単語の出現確率によって順に割り算して得られた値など）、(4)人手でルール化した情報などを含む。

[0013] また、本発明では、「先に入力された単語への新たな単語の追加」は、新たな単語を別の単語として追加する場合と、新たな単語を先に入力された一の単語の前へ結合する場合と、新たな単語を先に入力された一の単語の後へ結合する場合と、を含んでもよい。一方、「先に入力された単語への新たな単語の追加」は、上記のうち、新たな単語を別の単語として追加する場合のみを含んでもよい。

[0014] 上記のような情報検索装置に係る本発明は、情報検索方法に係る発明として捉えることもでき、以下のように記述することができる。即ち、本発明に係る情報検索方法は、情報検索装置により実行され、単語に関連する情報を検索する情報検索方法であって、単語を入力するための入力ステップと、所定数の単語が入力された後、新たな単語が入力されたときに、前記新たな単語を先に入力された単語に追加した場合および前記新たな単語を先に入力された単語と置換した場合の単語群パターンを生成するパターン生成ステップと、生成された単語群パターンそれぞれについて、当該単語群パターンが出現する確率に対応する出現情報を導出する出現情報導出ステップと、導出された出現情報に基づいて、新たな検索に用いるべき単語群パターンを決定する決定ステップと、を備えることを特徴とする。

発明の効果

[0015] 本発明によれば、意味を考えたルールの事前設定を必須とすることなく、新たな単語の入力時にクエリとして妥当な単語を簡易且つ適切に設定することができる。

図面の簡単な説明

[0016] [図1]第1実施形態の情報検索装置の機能的構成を示すブロック図である。

[図2]情報検索装置のハードウェア構成を示す図である。

[図3]第1実施形態の情報検索処理を示すフローチャートである。

[図4]第2実施形態の情報検索装置の機能的構成を示すブロック図である。

[図5]第2実施形態の情報検索処理を示すフローチャートである。

[図6]第3実施形態の情報検索装置の機能的構成を示すブロック図である。

[図7]第3実施形態の情報検索処理を示すフローチャートである。

[図8]共起確率データベースに記憶されたデータ例を示す図である。

[図9]クラス情報データベースに記憶されたデータ例を示す図である。

[図10]単語群パターンをランキング形式で出力した出力例を示す図である。

[図11]単語群パターンの生成処理を説明するための図である。

[図12]本発明に係る作用・効果を実現するために必要な構成要件を備えた情報検索装置の構成図である。

[図13]図12の情報検索装置により実行される情報検索処理を示すフローチャートである。

[図14]1つの単語が入力された後、新たな単語が入力された場合の単語群パターンの生成処理を説明するための図である。

発明を実施するための形態

[0017] 本発明に係る情報検索装置および情報検索方法のさまざまな実施形態について図面を参照して説明する。以下の第1実施形態では、所定数の単語（一例として複数の単語）が入力された後、新たな単語が入力されたときの単語群パターンの生成、各単語群パターンの出現情報（一例として共起確率）の導出、および出現情報に基づく単語群パターンの決定を行う実施形態を説明し、第2実施形態では、単語のクラス情報に応じて単語群パターンを生成する実施形態を説明し、第3実施形態では、共起情報が高い順に単語群パターンをランキング形式で出力し、ユーザ選択可能とする実施形態を説明する。なお、可能な場合には、同一の部分には同一の符号を付して、重複する説明を省略する。

[0018] [第1実施形態]

図1は、情報検索装置1の機能的構成を示すブロック図である。情報検索装置1は、単語をキーにした情報検索機能（即ち、単語を入力して当該単語に関連のある情報を検索する機能）を備えた情報検索装置である。

[0019] 図1に示すように、情報検索装置1は、機能的には、入力部11、検索部12、パターン生成部13、共起確率導出部14（出現情報導出部）、決定部15、および検索結果出力部16を備える。なお、第1～第3実施形態では、最初に単語が入力され当該単語を用いた情報検索が行われた後、新たな単語が入力されたときの適正な単語群パターンの決定について説明するが、最初に単語が入力された直後に当該単語を用いた情報検索を行うことは必須ではない。即ち、最初に単語が入力され当該単語を用いた情報検索を行う前に、新たな単語が入力されたときにも、本発明は適用可能である。これについては、図12と図13を用いて後述する。

[0020] 図2は、情報検索装置1のハードウェア構成図である。情報検索装置1は、物理的には、図2に示すように、CPU101、主記憶装置であるRAM102およびROM103、データ送受信デバイスである通信モジュール104、ハードディスク、フラッシュメモリ等の補助記憶装置105、入力デバイスであるキーボード等の入力装置106、ディスプレイ等の出力装置107などを含むコンピュータシステムとして構成されている。図1に示した各機能は、図2に示すCPU101、RAM102等のハードウェア上に所定のコンピュータソフトウェアを読み込ませることにより、CPU101の制御のもとで通信モジュール104、入力装置106、出力装置107を動作させるとともに、RAM102および補助記憶装置105におけるデータの読み出しおよび書き込みを行うことで実現される。再び、図1を参照し、情報検索装置1の各機能部について説明する。

[0021] 入力部11は、単語を入力するための部分であって、例えばマイクといった音声入力部11Aと、入力された音声から単語を認識する認識部11Bとを含んで構成される。認識部11Bは、周知の音声認識装置により構成でき

、例えば上記入力された音声を複数の区間に区切り、内蔵した辞書データベース（不図示）を参照して複数の区間毎に音声認識処理を実施する。なお、認識部 11B は、図 1 に示すように情報検索装置 1 に内蔵された認識部によって構成する以外に、外部の音声認識サーバによって構成してもよい。かかる場合は、情報検索装置 1 に外部の音声認識サーバを加えた構成が、本発明に係る情報処理装置に相当する。

[0022] 検索部 12 は、入力された単語に関連のある情報を検索する部分であって、周知の検索装置により構成できる。例えば、ユーザから音声入力部 11A に「渋谷でラーメン」という音声が入力された場合、「渋谷」と「ラーメン」の 2 つの単語が認識部 11B によって認識され、これらの単語に関連のある情報が検索部 12 によって検索される。その後、ユーザから音声入力部 11A に「ハンバーグだとう」という音声が入力された場合、「ハンバーグ」という単語が認識部 11B によって認識され、後述するパターン生成部 13、共起確率導出部 14 および決定部 15 によって、検索済みの「渋谷」と「ラーメン」の 2 つの単語に新たな単語「ハンバーグ」を追加した場合の単語群パターン、および、新たな単語「ハンバーグ」を先に入力された単語「渋谷」又は「ラーメン」と置換した場合の単語群パターンの中から、新たな検索に用いるべき単語群パターンが決定される。なお、検索部 12 における検索条件は任意に設定可能であり、ここではいわゆる AND 条件の検索を想定しているが、これに限定されるものではなく、いわゆる OR 条件の検索であってもよい。この点は以下の実施形態でも同様である。

[0023] パターン生成部 13 は、情報の検索が行われた後、新たな単語が入力されたときに、新たな単語を先に入力された単語に追加した場合および新たな単語を先に入力された単語と置換した場合の単語群パターンを生成する部分であり、その処理の詳細は後述する。

[0024] 共起確率導出部 14 は、生成された単語群パターンそれぞれの共起確率を導出する部分であり、具体的には、さまざまな単語群パターンの共起確率を記憶した共起確率データベース 14A（出現情報データベース）を含み、生

成された単語群パターンそれぞれの共起確率を共起確率データベース 14 A から読み出すことで、当該単語群パターンそれぞれの共起確率を導出する。共起確率データベース 14 A には、例えば図 8 に示すように、さまざまな単語群パターンの共起確率が、単語群パターンに対応付けて記憶されている。なお、本実施形態のように「複数の単語」が入力された後に、新たな単語が入力されたときに単語群パターンを生成する場合、生成される単語群パターンは複数の単語を含むため、ここでは「単語群パターンが出現する確率に対応する出現情報」として、複数の単語に関する共起情報を用い、その共起情報の一例として共起確率を用いた例を説明する。ただし、用いられる共起情報としては、共起確率以外にも、個々の単語の出現確率および共起確率から所定手順で導出される情報（例えば、共起確率を個々の単語の出現確率によって順に割り算して得られた値など）および人手でルール化した情報などが適用できる。

[0025] 決定部 15 は、導出された共起確率に基づいて、新たな検索に用いるべき単語群パターンを決定する部分であり、本実施形態において決定部 15 は、共起確率が最大である単語群パターンを、新たな検索に用いるべき単語群パターンとして決定する。

[0026] 検索結果出力部 16 は、検索部 12 による検索処理の結果を出力する部分である。出力の形式は、例えば表示出力でもよいし、音声出力でもよく、特定の出力形式に限定されるものではない。

[0027] 以下、図 3 のフローチャートに沿って、第 1 実施形態の情報検索装置 1 により実行される処理（本発明の情報検索方法に係る処理）を説明する。

[0028] 入力部 11 を介して複数の単語が入力され（図 3 のステップ S1：入力ステップ）、検索部 12 は当該入力された複数の単語を用いて情報検索を実行し、検索結果が検索結果出力部 16 により出力される（ステップ S2：検索ステップ）。そして、新たな単語が入力されると、パターン生成部 13 は、新たな単語を先に入力された単語に追加した場合および新たな単語を先に入力された単語それぞれと置換した場合の単語群パターンを生成する（ステッ

プS 3：パターン生成ステップ）。

[0029] ここでの単語群パターンの生成処理について、図11を参照して説明する。図11(a)に示すように、複数の単語「渋谷 ラーメン」が入力された後、新たな単語「ハンバーグ」が入力された場合を想定する。ここで、パターン生成部13は、新たな単語を先に入力された単語それぞれと置換した場合の単語群パターンとして、単語群パターンB「渋谷 ハンバーグ」と単語群パターンC「ハンバーグ ラーメン」とを生成する。

[0030] 一方、新たな単語の追加について、本発明では、(1)新たな単語を別の単語として追加する場合と、(2)新たな単語を先に入力された一の単語の前へ結合する場合と、(3)新たな単語を先に入力された一の単語の後へ結合する場合とを広く含む。そのため、新たな単語を追加した場合の単語群パターンとしては、図11(a)に示す単語群パターンA「渋谷 ラーメン ハンバーグ」、単語群パターンD「ハンバーグ渋谷 ラーメン」、単語群パターンE「渋谷ハンバーグ ラーメン」、単語群パターンF「渋谷 ハンバーグラーメン」、単語群パターンG「渋谷 ラーメンハンバーグ」が考えられる。ただし、実際には、上記(2)新たな単語を先に入力された一の単語の前へ結合する場合と、(3)新たな単語を先に入力された一の単語の後へ結合する場合については、一般に存在しないような単語（例えばハンバーグ渋谷など）までも処理対象になってしまう可能性が高いため、上記(2)新たな単語を先に入力された一の単語の前へ結合する場合と、(3)新たな単語を先に入力された一の単語の後へ結合する場合は除外し、上記(1)新たな単語を別の単語として追加する場合のみを含んでもよい。そのため、以下の実施形態では、新たな単語の追加については、(1)新たな単語を別の単語として追加する場合のみを想定して説明する。

[0031] 以上により、本実施形態のパターン生成部13は、図11(b)に示すように、新たな単語を先に入力された単語に追加した場合の単語群パターンとして、単語群パターンA「渋谷 ラーメン ハンバーグ」を生成し、新たな単語を先に入力された単語それぞれと置換した場合の単語群パターンとして

、単語群パターンB「渋谷 ハンバーグ」と単語群パターンC「ハンバーグ ラーメン」とを生成する。

[0032] 次に、共起確率導出部14は、ステップS3で生成された単語群パターンそれぞれの共起確率を共起確率データベース14Aから読み出すことで、当該単語群パターンそれぞれの共起確率を導出する（ステップS4：出現情報導出ステップ）。例えば図8の例では、共起確率導出部14は、単語群パターンA「渋谷 ラーメン ハンバーグ」の共起確率0.3、単語群パターンB「渋谷 ハンバーグ」の共起確率0.9、および単語群パターンC「ハンバーグ ラーメン」の共起確率0.6を導出する。

[0033] そして、決定部15は、導出された共起確率が最大である単語群パターンである単語群パターンB「渋谷 ハンバーグ」を、新たな検索に用いるべき単語群パターンとして決定する（ステップS5：決定ステップ）。

[0034] その後、検索部12は、ステップS5で決定された単語群パターンB「渋谷 ハンバーグ」を用いて情報検索を実行し（ステップS6）、新たな検索結果が検索結果出力部16により出力される（ステップS7）。

[0035] 以上のような処理により、さまざまな単語群パターンそれぞれの共起情報（ここでは共起確率）に応じて、新たな検索に用いるべき単語群パターンを簡易且つ適切に決定することができる。

[0036] [第2実施形態]

第2実施形態では、単語のクラス情報に応じて単語群パターンを生成する実施形態を説明する。図4に示すように、第2実施形態の情報検索装置1は、図1の第1実施形態の情報検索装置に対し、クラス情報データベース17をさらに追加した構成とされている。クラス情報データベース17には、例えば図9に示すように、さまざまな単語についてのクラス情報が単語に対応付けて記憶されている。また、第2実施形態におけるパターン生成部13は、先に入力された単語および新たな単語それぞれのクラス情報をクラス情報データベース17から読み出し、クラス情報に応じた単語群パターンの生成を行う機能を持つ。具体的にパターン生成部13は、「新たな単語のクラス

情報とは異なるクラス情報の先に入力された単語」と「当該新たな単語」とを置換した場合を除外して、単語群パターンを生成する。

[0037] 図5には、第2実施形態の情報検索装置1により実行される処理を示す。第2実施形態における処理は、ステップS3の単語群パターン生成処理が図3の第1実施形態の処理と異なるため、ステップS3について説明する。

[0038] 図5に示すステップS3にて、パターン生成部13は、先に入力された単語および新たな単語それぞれのクラス情報をクラス情報データベース17から読み出す（ステップS3A）。例えば、第1実施形態と同様、複数の単語「渋谷 ラーメン」により検索処理が実行された後、新たな単語「ハンバーグ」が入力された場合を想定すると、パターン生成部13は、図9のクラス情報データベース17から、「渋谷」のクラス情報として「地名」を、「ラーメン」のクラス情報として「食品」を、「ハンバーグ」のクラス情報として「食品」を、それぞれ読み出す。

[0039] そして、パターン生成部13は、「新たな単語のクラス情報とは異なるクラス情報の先に入力された単語」と「当該新たな単語」とを置換した場合を除外して、単語群パターンを生成する。上記例では、新たな単語「ハンバーグ」のクラス情報は「食品」であり、「ラーメン」のクラス情報と同じだが、「渋谷」のクラス情報「地名」とは異なるため、新たな単語「ハンバーグ」のクラス情報とは異なるクラス情報の先に入力された単語は、「渋谷」となる。そのため、パターン生成部13は、「渋谷」と新たな単語「ハンバーグ」とを置換した場合（図11（b）の単語群パターンC「ハンバーグ ラーメン」）を除外して、単語群パターンを生成する。よって、パターン生成部13は、図11（c）に示す単語群パターンA「渋谷 ラーメン ハンバーグ」および単語群パターンB「渋谷 ハンバーグ」を生成する。

[0040] 以後は、上記単語群パターンA、BのみについてステップS4以降の処理が実行される。

[0041] 以上のような第2実施形態によれば、互いにクラス情報の異なる単語同士を置換した単語群パターンを回避することで、見当ちがいな単語群パターン

を排除することができる。

[0042] また、第1、第2実施形態のように、決定部により共起情報（ここでは共起確率）が最大である単語群パターンを、新たな検索に用いるべき単語群パターンとして決定することで、ユーザの介入は不要となり自動的に単語群パターンを決定することができる。

[0043] [第3実施形態]

第3実施形態では、共起情報（ここでは共起確率）が高い順に単語群パターンをランキング形式で出力し、ユーザ選択可能とする実施形態を説明する。図6に示すように、第3実施形態の情報検索装置1は、図4の第2実施形態の情報検索装置において決定部15の構成を変更したものである。決定部15は、共起確率導出部14により導出された共起確率が高い順に単語群パターンを、選択可能なランキング形式で出力するランキング出力部15Aと、ユーザにより選択された単語群パターンを受け付け、当該選択された単語群パターンを新たな検索に用いるべき単語群パターンとして決定する受付部15Bとを含んでいる。ランキング出力部15Aによる出力の形式は、例えば表示出力でもよいし、音声出力でもよく、特定の出力形式に限定されるものではない。ここでは、一例としてランキング出力部15Aは、図10（a）、（b）のようなランキング形式の単語群パターンを示す画面を表示出力する。

[0044] 図7には、第3実施形態の情報検索装置1により実行される処理を示す。第3実施形態における処理は、ステップS5の単語群パターン生成処理が図5の第2実施形態の処理と異なるため、ステップS5について説明する。

[0045] 図7に示すステップS5にて、ランキング出力部15Aは、共起確率導出部14により導出された共起確率が高い順に単語群パターンを、選択可能なランキング形式で出力する（ステップS5A）。例えばパターン生成部13により、図11（c）に示す単語群パターンA「渋谷 ラーメン ハンバーグ」および単語群パターンB「渋谷 ハンバーグ」が生成された場合、ステップS4で単語群パターンA、Bそれぞれの共起確率が導出され、ステップ

S 5 Aでは、図 1 0 (a) に示すように単語群パターン A、B が、共起確率の高い順に、選択可能なランキング形式で出力される。これにより、ユーザは、例えば図 1 0 (a) の選択欄を用いて、新たな検索に用いるべき単語群パターンをランキング形式の出力画面から選択することができる。なお、図 1 0 (a) の出力例では、各単語群パターンの共起確率の情報を出力しているが、共起確率の情報を出力することは必須ではない。

[0046] ユーザにより、新たな検索に用いるべき単語群パターンが選択されると、受付部 1 5 B は、当該選択された単語群パターンを受け付け、当該単語群パターンを新たな検索に用いるべき単語群パターンとして決定する（ステップ S 5 B）。以後は、決定された単語群パターンについてステップ S 6 以降の処理が実行される。

[0047] 以上のような第 3 実施形態によれば、ユーザからみて、新たな検索に用いるべき単語群パターンをランキング形式の出力画面から選択可能となる。

[0048] なお、第 3 実施形態では、図 4 の第 2 実施形態の情報検索装置において決定部 1 5 の構成を変更した例を説明したが、図 1 の第 1 実施形態の情報検索装置において決定部 1 5 の構成を変更してもよい。その場合、例えば図 3 のステップ S 4 でパターン生成部 1 3 により、図 1 1 (b) に示す単語群パターン A「渋谷 ラーメン ハンバーグ」、単語群パターン B「渋谷 ハンバーグ」および単語群パターン C「ハンバーグ ラーメン」が生成され、ステップ S 4 で単語群パターン A～C それぞれの共起確率が導出され、ステップ S 5 A では、図 1 0 (b) に示すように単語群パターン A～C が、共起確率の高い順に、選択可能なランキング形式で出力される。これにより、ユーザは、例えば図 1 0 (b) の選択欄を用いて、新たな検索に用いるべき単語群パターンをランキング形式の出力画面から選択することができる。

[0049] 上記第 1、第 2 実施形態では、決定部 1 5 が、共起情報（一例として共起確率）が最大である単語群パターンを、新たな検索に用いるべき単語群パターンとして決定する例を説明したが、共起情報が最大の単語群パターンが複数存在するケースおよび共起情報の高い順で第 1 位の共起情報と第 2 位の共

起情報とで差が殆ど無いケースなどが想定される。そのため、変形例として、共起情報の高い順で第1位の共起情報との差が所定値以下である共起情報、共起情報がある場合には、該当する共起情報の単語群パターンと第1位の共起情報の単語群パターンとを、上記第3実施形態のようにユーザ向けに出力して(例えばランキン形式で出力して)、ユーザにより、新たな検索に用いるべき単語群パターンを選択させてもよい。

[0050] また、上記第1～第3実施形態では、単語を入力するための入力部11として、単語を音声入力する形態を例にとりて説明したが、単語を音声入力する形態以外にも、キーボード等のさまざまな情報入力手段により単語を入力する形態を採用してもよく、上記第1～第3実施形態と同様の作用・効果を齎すことができる。

[0051] また、上記第1～第3実施形態では、「新たな単語の追加」について、(1)新たな単語を別の単語として追加する場合、(2)新たな単語を検索済みの一の単語の前へ結合する場合、および(3)新たな単語を検索済みの一の単語の後へ結合する場合のうち、(1)新たな単語を別の単語として追加する場合のみを想定して説明したが、上記(1)～(3)全ての場合を想定してもよいし、上記(1)～(3)のうち何れか1つを想定してもよいし、上記(1)～(3)のうち2つを想定してもよい。いずれの場合も、上記第1～第3実施形態と同様の作用・効果を齎すことができる。

[0052] さて、図12には、本発明に係る作用・効果を実現するために必須な構成要件を備えた情報検索装置の構成例を示す。この図12に示すように、情報検索装置1は、単語を入力するための入力部11と、所定数の単語が入力された後、新たな単語が入力されたときに、新たな単語を先に入力された単語に追加した場合および新たな単語を先に入力された単語と置換した場合の単語群パターンを生成するパターン生成部13と、生成された単語群パターンそれぞれについて、当該単語群パターンが出現する確率に対応する出現情報を出出力する出現情報導出部14と、導出された出現情報に基づいて、新たな検索に用いるべき単語群パターンを決定する決定部15と、を備える。こ

のような情報検索装置 1 では、図 1 3 に示すように、所定数の単語が入力された（ステップ S 1）後、新たな単語が入力されたとき、パターン生成部 1 3 が、新たな単語を先に入力された単語に追加した場合および新たな単語を先に入力された単語と置換した場合の単語群パターンを生成し（ステップ S 3）、出現情報導出部 1 4 X が、生成された単語群パターンそれぞれについての出現情報を導出し（ステップ S 4 X）、そして、決定部 1 5 が、導出された出現情報に基づいて、新たな検索に用いるべき単語群パターンを決定する（ステップ S 5）。これにより、さまざまな単語群パターンそれぞれの出現情報に応じて、新たな検索に用いるべき単語群パターンを簡易且つ適切に決定することができる。即ち、意味を考えたルールの事前設定を必須とすることなく、新たな単語の入力時にクエリとして妥当な単語を簡易且つ適切に設定することができる。

[0053] 上記第 1 ～第 3 実施形態では、「2 つの単語」が入力された後、新たな単語が入力された場合の新たな単語群パターンの決定について説明したが、本発明は、「3 つ以上の単語」が入力された後、新たな単語が入力された場合の新たな単語群パターンの決定についても適用可能であり、同様の作用・効果を得ることができる。

[0054] 同様に、本発明は、「1 つの単語」が入力された後、新たな単語が入力された場合の新たな単語群パターンの決定についても適用可能である。例えば図 1 4 に示すように、1 つの単語「渋谷」が入力された後、新たな単語「ラーメン」が入力されたときに、図 1 2 のパターン生成部 1 3 は、新たな単語を先に入力された単語に追加した場合の単語群パターンとして、単語群パターン X「渋谷 ラーメン」を生成し、新たな単語を先に入力された単語と置換した場合の単語群パターンとして、単語群パターン Y「ラーメン」を生成する。そして、出現情報導出部 1 4 X は、生成された単語群パターン X、Y それぞれについての出現情報を導出する。ここでの出現情報としては、例えば、複数の単語を含む単語群パターン X については、共起確率を個々の単語の出現確率で順に割り算して得られた値を導出し、1 つの単語を含む単語群

パターンYについては、当該単語の出現確率を出現情報としてもよい。さらに、決定部15が、導出された出現情報に基づいて、新たな検索に用いるべき単語群パターンを決定する。以上のようにして、「1つの単語」が入力された後、新たな単語が入力された場合の新たな単語群パターンの決定についても本発明は適用可能であり、同様の作用・効果を得ることができる。なお、パターン生成部13は、第2実施形態のように、各単語のクラス情報を参照することで、先に入力された単語「渋谷」のクラス情報「地名」と新たな単語「ラーメン」のクラス情報「食品」とが異なることを認識し、これらのクラス情報が異なる単語同士の置換の場合を除外して、単語群パターンを生成してもよい。この場合は、図14の下段に示すように、単語群パターンX「渋谷 ラーメン」のみが生成される。

符号の説明

[0055] 1…情報検索装置、11…入力部、11A…音声入力部、11B…認識部、12…検索部、13…パターン生成部、14…共起確率導出部、14A…共起確率データベース、14X…出現情報導出部、15…決定部、15A…ランキング出力部、15B…受付部、16…検索結果出力部、17…クラス情報データベース、101…CPU、102…RAM、103…ROM、104…通信モジュール、105…補助記憶装置、106…入力装置、107…出力装置。

請求の範囲

[請求項1]

単語に関連する情報を検索する情報検索装置であって、
単語を入力するための入力部と、
所定数の単語が入力された後、新たな単語が入力されたときに、前記新たな単語を先に入力された単語に追加した場合および前記新たな単語を先に入力された単語と置換した場合の単語群パターンを生成するパターン生成部と、
生成された単語群パターンそれぞれについて、当該単語群パターンが出現する確率に対応する出現情報を導出する出現情報導出部と、
導出された出現情報に基づいて、新たな検索に用いるべき単語群パターンを決定する決定部と、
を備える情報検索装置。

[請求項2]

前記情報検索装置は、
さまざまな単語についてのクラス情報を記憶したクラス情報データベース、をさらに備え、
前記パターン生成部は、
前記新たな単語および先に入力された単語それぞれのクラス情報を前記クラス情報データベースから読み出し、前記新たな単語のクラス情報とは異なるクラス情報の先に入力された単語と当該新たな単語とを置換した場合を除外して、前記単語群パターンを生成する、
ことを特徴とする請求項1に記載の情報検索装置。

[請求項3]

前記決定部は、導出された出現情報が高い順に前記単語群パターンを、選択可能なランキング形式で出力し、ユーザにより選択された単語群パターンを、新たな検索に用いるべき単語群パターンとして決定することを特徴とする請求項1又は2に記載の情報検索装置。

[請求項4]

前記決定部は、導出された出現情報が最大である単語群パターンを、新たな検索に用いるべき単語群パターンとして決定することを特徴とする請求項1又は2に記載の情報検索装置。

- [請求項5] 前記入力部は、
音声を入力するための音声入力部と、
入力された音声から単語を認識する認識部と、
を含むことを特徴とする請求項 1 ～ 4 の何れか一項に記載の情報検索装置。
- [請求項6] 前記出現情報導出部は、
さまざまな単語群パターンの出現情報を記憶した出現情報データベースを含み、前記生成された単語群パターンそれぞれの出現情報を前記出現情報データベースから読み出すことで、当該単語群パターンそれぞれの出現情報を導出することを特徴とする請求項 1 ～ 5 の何れか一項に記載の情報検索装置。
- [請求項7] 前記所定数の単語は、複数の単語であり、
前記出現情報導出部は、前記出現情報として、共起情報を導出する、
ことを特徴とする請求項 1 ～ 6 の何れか一項に記載の情報検索装置。
- [請求項8] 先に入力された単語への前記新たな単語の追加には、
前記新たな単語を別の単語として追加する場合と、
前記新たな単語を先に入力された一の単語の前へ結合する場合と、
前記新たな単語を先に入力された一の単語の後へ結合する場合と、
が含まれることを特徴とする請求項 1 ～ 7 の何れか一項に記載の情報検索装置。
- [請求項9] 情報検索装置により実行され、単語に関連する情報を検索する情報検索方法であって、
単語を入力するための入力ステップと、
所定数の単語が入力された後、新たな単語が入力されたときに、前記新たな単語を先に入力された単語に追加した場合および前記新たな単語を先に入力された単語と置換した場合の単語群パターンを生成す

るパターン生成ステップと、

生成された単語群パターンそれぞれについて、当該単語群パターン
が出現する確率に対応する出現情報を導出する出現情報導出ステップ
と、

導出された出現情報に基づいて、新たな検索に用いるべき単語群パ
ターンを決定する決定ステップと、

を備える情報検索方法。