(19)日本国特許庁(JP)　(12)特 許 公 報(B2)　(11)特許番号 特許第6176017号 (P6176017)

(45)発行日 平成29年8月9日(2017.8.9)　(24)登録日 平成29年7月21日(2017.7.21)

(51)Int.Cl.		F I		
G06F 17/30	(2006.01)	G06F	17/30	３３０Ｃ
G06F 17/27	(2006.01)	G06F	17/30	１７０Ａ
		G06F	17/30	３２０Ｄ
		G06F	17/30	４１９Ｂ
		G06F	17/27	６８５

請求項の数 6 (全 27 頁)

(21)出願番号 特願2013-191324(P2013-191324)
(22)出願日 平成25年9月17日(2013.9.17)
(65)公開番号 特開2015-60243(P2015-60243A)
(43)公開日 平成27年3月30日(2015.3.30)
審査請求日 平成28年6月6日(2016.6.6)

(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 100074099
弁理士 大菅 義之
(74)代理人 100133570
弁理士 ▲徳▼永 民雄
(72)発明者 大倉 清司
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 潮田 明
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】検索装置、検索方法、およびプログラム

(57)【特許請求の範囲】

【請求項１】

検索文を意味解析してグラフで表現される意味構造を生成する意味解析部と、

前記意味解析部によって生成された意味構造上の２つのノードと該２つのノード間を直接接続するアークの種別とを表すことによって該２つのノード間の意味のつながりを検索対象とした第１のキー、前記意味構造上で複数のアークを介して間接的に接続されている２つの間接接続ノードと該複数のアークの各々についての前記種別である複数の種別とを表すことによって該２つの間接接続ノード間の間接的な意味のつながりを検索対象とした第２のキー、及び、該複数のアークの各々によって接続されている２つの被接続ノードと前記複数の種別とを表すことによって前記２つの間接接続ノード間に存在している該２つの被接続ノード間の意味のつながりを検索対象とした第３のキーを、検索キーとして生成する検索キー生成部と、

前記生成された検索キーにより、検索対象文書を蓄積したデータベース上の検索用インデックスに格納されている当該検索対象文書に対応する意味構造上のノードの組合せとのマッチングを実行することにより、前記検索文にマッチする検索対象文書を検索する検索部と、

を備えることを特徴とする検索装置。

【請求項２】

前記検索キーは、前記第２及び前記第３のキーにおいて表されている前記複数の種別がＯＲ結合されている、

ことを特徴とする請求項１に記載の検索装置。

【請求項３】

前記検索キー生成部は、前記種別が類似する関係を表すアークをグループ化して前記検索キーを生成する、

ことを特徴とする請求項１に記載の検索装置。

【請求項４】

前記意味解析部、前記検索キー生成部、前記検索部、前記検索文を入力する検索クライアント部、前記データベースを作成するデータベース作成部、または該データベース作成部に前記検索対象文書を入力するデータベース作成クライアント部を、ネットワークを介して接続される複数のサーバコンピュータに備える、

ことを特徴とする請求項１ないし３のいずれかに記載の検索装置。

【請求項５】

検索装置が行う検索方法であって、

前記検索装置が備えている意味解析部が、検索文を意味解析してグラフで表現される意味構造を生成し、

前記検索装置が備えている検索キー生成部が、前記意味解析によって生成された意味構造上の２つのノードと該２つのノード間を直接接続するアークの種別とを表すことによって該２つのノード間の意味のつながりを検索対象とした第１のキー、前記意味構造上で複数のアークを介して間接的に接続されている２つの間接接続ノードと該複数のアークの各々についての前記種別である複数の種別とを表すことによって該２つの間接接続ノード間の間接的な意味のつながりを検索対象とした第２のキー、及び、該複数のアークの各々によって接続されている２つの被接続ノードと前記複数の種別とを表すことによって前記２つの間接接続ノード間に存在している該２つの被接続ノード間の意味のつながりを検索対象とした第３のキーを、検索キーとして生成し、

前記検索装置が備えている検索部が、前記生成された検索キーにより、検索対象文書を蓄積したデータベース上の検索用インデックスに格納されている当該検索対象文書に対応する意味構造上のノードの組合せとのマッチングを実行することにより、前記検索文にマッチする検索対象文書を検索する、

ことを特徴とする検索方法。

【請求項６】

検索文を意味解析してグラフで表現される意味構造を生成するステップと、

前記意味解析によって生成された意味構造上の２つのノードと該２つのノード間を直接接続するアークの種別とを表すことによって該２つのノード間の意味のつながりを検索対象とした第１のキー、前記意味構造上で複数のアークを介して間接的に接続されている２つの間接接続ノードと該複数のアークの各々についての前記種別である複数の種別とを表すことによって該２つの間接接続ノード間の間接的な意味のつながりを検索対象とした第２のキー、及び、該複数のアークの各々によって接続されている２つの被接続ノードと前記複数の種別とを表すことによって前記２つの間接接続ノード間に存在している該２つの被接続ノード間の意味のつながりを検索対象とした第３のキーを、検索キーとして生成するステップと、

前記生成された検索キーにより、検索対象文書を蓄積したデータベース上の検索用インデックスに格納されている当該検索対象文書に対応する意味構造上のノードの組合せとのマッチングを実行することにより、前記検索文にマッチする検索対象文書を検索するステップと、

をコンピュータに実行させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、情報を検索するための検索装置、検索方法、およびプログラムに関する。

【背景技術】

【０００２】
　例えば論文や特許の検索、あるいは一般のＷｅｂページの検索などに適用できるキーワードによる検索では、キーワード単位の検索クエリを用いるため、キーワード間の関係を検索条件に含めることができない。例えば、ユーザが「熱の放出を低減することにより、環境負荷を減らす技術」を検索したい場合に、キーワード「熱の放出、環境負荷、低減」を入力する。しかし、このキーワードの検索クエリには、あいまい性が含まれている。この場合、「熱の放出を低減する」と「環境負荷を低減する」の両方の意味を含み、一方の意味に限定していない。また、熱の放出と環境負荷との関係も特定していない。
【０００３】
　このように検索クエリにあいまい性が含まれているキーワード検索には、以下の問題がある。
（１）ユーザが意図した通りに検索されない。
（２）キーワードを含んでいても意図しない文書が検索される。
【０００４】
　特に、（２）の問題は、抽出された文書中から、ユーザが意図した部分を見つけ出す手間が煩雑であるという欠点を含んでいる。例えば、３つのキーワードで検索してすべてのキーワードを含む文が検出された場合でも、その文は意図しない関係を含んでいるかもしれない。つまり、キーワードにヒットした部分を出力しても、必ずしもユーザが欲しい情報ではないことがある。従って、有用な文を抽出する判断に、ユーザは時間を費やすことになる。
【０００５】
　検索精度を向上させる一従来技術として、次のようなものが知られている（例えば特許文献１に記載の技術）。言語解析手段は入力された検索要求を言語解析して検索条件の要素となる単語及び複数の単語より構成された名詞句を抽出する。検索条件生成手段は、抽出した単語及び名詞句を所定の演算子により結合し、かつ単語及び名詞句の各々に所定の重み付けを施して検索条件を生成する。文書検索手段は検索条件生成手段により生成した検索条件に合致した文書を検索対象文書から抽出する。このように、自然言語で入力された検索要求から抽出された単語及び名詞句に基づいて適切な検索条件を設定することにより、精度の高い文書検索結果を得ることを目的とする。
【０００６】
　この技術は、入力された検索要求を言語解析して検索条件の要素となる単語及び複数の単語より構成された名詞句を検索のためのキーワードとするものである。しかしながら、この従来技術は依然として、前述した（２）の問題を解決することはできない。
【０００７】
　検索精度を向上させる他の従来技術として、次のようなものが知られている（例えば特許文献２に記載の技術）。検索を要求する式または文である検索要求情報から検索用語となりうるすべてのキーワードを抽出するキーワード抽出部を備える。抽出したキーワードの類義語を類義語辞書から取得する類義語取得部を備える。キーワード抽出部が抽出したキーワードと、類義語取得部が取得した類義語とを指定して、検索対象の文書を記憶する検索対象文書ＤＢから文書を検索する文書検索部を備える。そして、文書検索部が検索した文書のうち、少なくとも検索要求情報に含まれる係り受け関係にある複数の単語の対と、文書検索部が検索した文書に含まれる係り受け関係にある複数の単語の対との関連する度合いを示す係り受け類似度が大きい文書ほど文書の優先順位を高く評価する検索結果評価部を備える。これにより、ユーザにとって有用な情報を高精度で検索する情報検索装置を提供する。
【０００８】
　この技術は、複数の単語の対の係り受け関係の類似度を検索に導入することで、類義語を使用し検索条件を広げて文書を検索した場合でも、不適切な検索候補文書を排除し、ユーザが頻繁に使用するキーワードに関連した有用な情報を高精度で取得するものである。しかしながら、この従来技術では、係り受け関係になくても意味的に似た文に含まれる複

数の単語については、検索漏れが発生し、検索精度を向上させることはできないという問題点を有していた。
【先行技術文献】
【特許文献】
【０００９】
【特許文献１】特開２００１－１４２８９７号公報
【特許文献２】特開２００６－２１５７１７号公報
【発明の概要】
【発明が解決しようとする課題】
【００１０】
本発明は、意図しない文書の検索や検索漏れの発生を排除することを目的とする。
【課題を解決するための手段】
【００１１】
態様の一例では、検索文を意味解析してグラフで表現される意味構造を生成する意味解析部と、意味解析部によって生成された意味構造上の２つのノードと該２つのノード間を直接接続するアークの種別とを表すことによって該２つのノード間の意味のつながりを検索対象とした第１のキー、意味構造上で複数のアークを介して間接的に接続されている２つの間接接続ノードと該複数のアークの各々についての種別である複数の種別とを表すことによって該２つの間接接続ノード間の間接的な意味のつながりを検索対象とした第２のキー、及び、該複数のアークの各々によって接続されている２つの被接続ノードと該複数の種別とを表すことによって該２つの間接接続ノード間に存在している当該２つの被接続ノード間の意味のつながりを検索対象とした第３のキーを生成する検索キー生成部と、生成された検索キーにより、検索対象文書を蓄積したデータベース上の検索用インデックスに格納されている検索対象文書に対応する意味構造上のノードの組合せとのマッチングを実行することにより、検索文にマッチする検索対象文書を検索する検索部とを備える。
【発明の効果】
【００１２】
意図しない文書の検索や検索漏れの発生を排除することが可能となる。
【図面の簡単な説明】
【００１３】
【図１】本実施形態がベースとする意味検索技術の説明図である。
【図２】自然文から意味最小単位を計算する方法の説明図である。
【図３】基本的な意味検索技術を用いた検索装置の構成図である。
【図４】基本的な意味検索技術を用いた検索装置におけるデータベース作成処理を示すフローチャートである。
【図５】データベース作成の例を示す図である。
【図６】検索用インデックスに格納される意味最小単位の例を示す図である。
【図７】評価値テーブルの例を示す図である。
【図８】基本的な意味検索技術を用いた検索装置における検索処理を示すフローチャートである。
【図９】検索クエリからの意味最小単位の生成を説明する図である。
【図１０】基本的な意味検索技術を用いた検索装置における検索漏れの例を示す図である。
【図１１】検索漏れを防止する方式の説明図である。
【図１２】本実施形態の説明図である。
【図１３】本実施形態による検索装置の構成図である。
【図１４】本実施形態による検索キー生成のための制御処理を示すフローチャートである。
【図１５】ノードリストの作成処理の説明図である。
【図１６】本実施形態による検索キー生成処理を示すフローチャートである。

【図１７】アーク名とグループ名との対応関係の例を示す図である。
【図１８】間接接続の説明図である。
【図１９】間接接続の他の制御方法の説明図である。
【図２０】クラウド構成に対応した実施形態の構成例を示す図である。
【図２１】本実施形態の機能を搭載したソフトウェアプログラムを実行可能なコンピュータのハードウェア構成例を示す図である。
【発明を実施するための形態】
【００１４】
以下、本発明を実施するための形態について図面を参照しながら詳細に説明する。
本実施形態では、意味検索といって、文の意味をグラフで表現し、そのグラフ表現を使って検索することにより、従来のキーワードベースよりも精度が高い検索装置を実現する。機械翻訳は、言葉を処理する上での要素技術が全て入っている。機械翻訳は公知であるが、出願人らは、日英翻訳を例にあげると、日本語の単語を抽出して構文解析を行っている。その次に、それに基づいて、意味構造といって、文の意味を表す構造に変換して、その意味構造はどの言語にも依存しないように設計されているため、その意味構造に落としてから例えば英語の文を生成するという処理を行っている。本実施形態の意味検索では、日本語の文を解析して生成される意味構造を用いた検索を行う。
【００１５】
図１は、本実施形態がベースとする意味検索技術の説明図である。この意味検索技術では、まず、データベース１０４にはあらかじめ、検索対象文書１０６が文ごとに意味解析され、その意味最小単位１０５が格納されている。検索時には、検索クエリ１０１の自然言語文（検索文）１０２に相当する意味最小単位（後述する）１０３が検索キーとして特定される。そして、検索キーの意味最小単位１０３と一致する意味最小単位１０５を含むデータベース１０４内の検索対象文１０６を特定することにより、検索者が意図する単語同士の意味関係等に応じた検索が行われる。
【００１６】
本実施形態の構成について説明する前に、本実施形態がベースとする基本的な意味検索技術について以下に説明する。基本的な意味検索技術では、検索対象を限定しすぎるため意味が非常に似ている文でも検索もれがおこる場合があるが、本実施形態を実現するための重要な技術であるため、まずこの技術について以下に説明する。
【００１７】
図２は、自然文から意味最小単位を計算する方法の説明図である。例えば原文２０１である「太郎は花子に本をあげた。」を意味解析すると、有向グラフ２０２で示される構造になる。これを意味構造と呼ぶ。「あげる」というのがこの文の中心の概念になる。誰が「あげる」という行動をしたかを示す「動作主」が「太郎」で、それを誰に対してかを示す「目標」が「花子」である。そして、何をあげたかを示す「対象」が「本」である。このように、有向グラフ２０２では、依存関係のようなものがグラフで表現されている。これらの「あげる」「太郎」「花子」「本」は文中の単語である。それぞれの単語に対して「GIVE」「HANAKO」「TARO」「BOOK」で示される意味記号が付与されている。例えば「ユーザ」と「利用者」のように、違う単語だが同じ意味を表すものは一つの意味記号にまとまる。また、「あげる」には、「手をあげる」、「棚にあげる」などの様々な意味があるが、その場合は意味に応じて意味記号も変わる。１つの単語に１つの意味記号が対応するとは限らず、様々な可能性がある。意味構造である有向グラフ２０２は、意味記号における記号同士の関係を表したものである。
【００１８】
長い文でも同様の構造になる。グラフ同士をマッチングするのは処理も重くなる。そこで、意味最小単位といって、グラフ表現を分解して単純な構造にする。
【００１９】
ここで、２つの関係に着目する。図２の有向グラフ２０２では、「GIVE　あげる」から「HANAKO　花子」に「目標」という矢印（以下「アーク」と呼ぶ）がつながっている。ま

た、「GIVE　あげる」から「TARO　太郎」に「動作主」というアークがつながっている。「GIVE　あげる」から「BOOK　本」に「対象」というアークがつながっている。ほかに「述語」や「過去」等の接続していないアークもあるが、これらは「GIVE　あげる」の属性を表している。ここで、接続していないことを「NIL」という記号で表す。

【００２０】

意味最小単位の出力においては、有向グラフ（例えば図２の２０２）からアークが抽出され、以下の処理が実行される。

（１）アークが2つのノードをつないでいる場合：
　　それぞれのアークに対して、
　　　（アークが出るノード,アークが向かうノード,アークの名前）
　　が出力される。
（２）アークが出るノードがない場合：
　　それぞれのアークに対して、
　　　（"NIL",アークが向かうノード,アークの名前）
　　が出力される。
（３）アークが向かうノードがない場合：
　　それぞれのアークに対して、
　　　（アークが出るノード,"NIL",アークの名前）
　　が出力される。

【００２１】

上述の（１）～（３）の処理を行った結果出力されたものを「意味最小単位」（例えば図２の２０３）と定義する。

【００２２】

図２の６個に分解された部分２０３を、意味最小単位という。

このように、意味最小単位は、文中の2つの概念間の関係または概念の役割を表すものであり、これをキーにしてデータベースを検索することにより、意味を踏まえた検索が可能となる。

【００２３】

意味最小単位は有向グラフの部分構造を表したものであり、意味最小単位でマッチングすることにより、有向グラフをマッチングさせるより柔軟な検索が可能になる。また、意味最小単位の後述するidf値（意味最小単位を含む文書数が少ないほど値が高くなる）を計算しておけば、マッチした意味最小単位のidf値を使って評価値を計算することにより、検索結果に対してランキングを行うことができる。

【００２４】

図３は、本実施形態がベースとする基本的な意味検索技術を用いた検索装置の構成図である。この検索装置は、データベース作成用として、検索対象テキスト入力部３０１、意味解析部３０２、意味最小単位保存部３０３、データベース作成部３０４、検索用インデックス３０５の記憶部、評価値計算部３０６、評価値テーブル３０７の記憶部を備える。また、検索実行用として、入力部３０８、上述の意味解析部３０２と意味最小単位保存部３０３、検索部３０９、評価値計算部３１０、ランキング部３１１、および検索結果表示部３１２を備える。検索対象テキスト入力部３０１には、検索対象文書３１３が入力される。入力部３０８には、検索クエリ３１４の自然文が入力される。

【００２５】

まず、データベースの作成時には、検索対象テキスト入力部３０１は、検索対象文書３１３の入力を受け付ける。

【００２６】

意味解析部３０２は、検索対象テキスト入力部３０１により受け付けられた検索対象文書３１３を意味解析する。意味解析とは、図２に示されるように、入力文（例えば図２の２０１）を解析し、その中の概念（単語の意味）間の関係を有向グラフ（例えば図２の２０２）により表現することである。意味解析部３０２は、この有向グラフに基づいて、意

味最小単位を生成する。
【００２７】
意味最小単位保存部３０３は、意味解析部３０２が生成した意味最小単位を記憶する。
データベース作成部３０４は、意味最小単位保存部３０３に保存された検索対象文書３１３に含まれる各意味最小単位に基づいて、検索用インデックス（データベース）３０５を作成する。この検索用インデックス３０５には、検索対象文書３１３ごとに、複数の意味最小単位が含まれる。
【００２８】
評価値計算部３０６は、検索用インデックス３０５中の各意味最小単位についての文書頻度（いくつの文書にその意味最小単位が出現するか）を計算しておき、そこから各意味最小単位のidf値を計算し、評価値テーブル３０７に格納しておく。idf値は以下のようにして算出される。

idf＝log(全文書数／df)
df（document frequency）：意味最小単位が出現する文書数
・・・（１）

【００２９】
例えば、意味最小単位が１０万文書中１０万文書に出現していれば、idfの値は１になる。出現文書数が少なければidfの値は大きくなる。例えば１文書にしか現れない非常に珍しい意味最小単位ならidfの値はlog(１０万)になる。　意味最小単位が検索用インデックス３０５上でマッチしたときに、その意味最小単位のidfを評価値として計算することにより、めずらしいような意味構造がヒットしたときはそれが検索上位に現れるようにしている。
【００３０】
なお、評価値テーブル３０７はここではidf値を例にとって説明しているが、他の値でもよい。
【００３１】
以上のようにして、データベースとしての検索用インデックス３０５および評価値テーブル３０７が予め作成された後に、実際の検索が実行される。
【００３２】
入力部３０８は、ユーザから検索クエリ３１４の自然文を受け付ける。
入力部３０８により受け付けられた検索クエリ３１４は、前述した意味解析部３０２により、意味解析されて検索クエリ３１４に含まれる複数の意味最小単位が生成され、それらの意味最小単位が意味最小単位保存部３０３に保存される。
【００３３】
検索部３０９は、検索クエリ３１４から生成され意味最小単位保存部３０３に保存された意味最小単位によって、データベースである検索用インデックス３０５を検索する。
【００３４】
評価値計算部３１０は、評価値テーブル３０７と検索用インデックス３０５を参照して、検索部３０９でマッチした意味最小単位から、評価値を計算する。
【００３５】
ランキング部３１１は、評価値計算部３１０により計算された検索結果の各文書の評価値をソートし、ランキングを行う。ランキングされた結果は、検索結果表示部３１２によりユーザに提示される。
【００３６】
図４は、基本的な意味検索技術を用いた図３の構成を有する検索装置におけるデータベース作成処理を示すフローチャートである。
【００３７】
まず、ステップＳ４０１とＳ４０６の制御によるループ処理によって検索対象文書３１３（図３）のそれぞれにつき、以下のステップＳ４０２からＳ４０５までの処理が実行される。

【００３８】
　まず、検索対象文書３１３の１つが文に分割される（ステップＳ４０２）。
　次に、ステップＳ４０３とＳ４０５の制御によるループ処理によって分割で得られたそれぞれの文につき、意味（意味解析）処理が実行されて意味最小単位が生成され、検索用インデックス３０５（図３）に追加される（ステップＳ４０４）。
【００３９】
　以上の処理は、図３の検索対象テキスト入力部３０１、意味解析部３０２、意味最小単位保存部３０３、およびデータベース作成部３０４の機能に対応するデータベース作成処理である。
【００４０】
　図５は、データベース作成の例を示す図である。例えば、文書Ａの中に文m「本発明による機械翻訳システムは，翻訳家が修正した単語あるいは修正または作成した翻訳文をそれぞれ前記辞書および例文データベースに登録する登録手段を持つ。」があるとする。この文を意味解析し、意味最小単位を生成すると、図５の右側のようになる。理解を容易にするために、意味最小単位が表す意味も右に付記しておく。生成された意味最小単位は文m、文書Ａとひもづけておく。この情報は検索時の評価値計算に使う。同様に他の文に関しても同じように処理を行い、意味最小単位を生成する。
【００４１】
　図６は、検索用インデックス３０５（図３）に格納される意味最小単位の例を示す図である。検索用インデックス３０５は、意味最小単位をキーとして、その意味最小単位の文書ID、文IDや、出現位置などを検索するためのインデックステーブルである。出現位置の情報としては、アークが出発するノードであるノード１の位置と、そのノード１の文字列長、およびアークが到達するノードであるノード２の位置と、そのノード２の文字列長が格納される。
【００４２】
　図４のステップＳ４０３からＳ４０５の各文に対応するループ処理、およびステップＳ４０１からＳ４０６の各検索対象文書３１３に対するループ処理が終了し、検索用インデックス３０５が作成されたら、以下のステップＳ４０６の処理が実行される。検索用インデックス３０５に格納されたそれぞれの意味最小単位について、idf値が計算され、評価値テーブル３０７（図３）に格納される。
【００４３】
　図７は、評価値テーブル３０７の例を示す図である。検索用インデックス３０５中に現れる意味最小単位（図６参照）ごとに、前述した（１）式に基づいてidf値が計算される。図７において、意味最小単位「(SEE,NIL,述語)」は、「見る」という意味の意味最小単位の例であるが、「見る」という語句は一般に多くの文書中に出現するため、（１）式において、df値が大きくなり、これを分母として計算されるidf値は小さい値となる。これに対して、「(SEE,GIRAFFE,目的語)」は、「キリンを見る」という意味の意味最小単位の例であるが、「キリンを見る」という語句はそれほど多くの文書中には出現しないため、（１）式において、df値が小さくなり、これを分母として計算されるidf値は大きい値となる。
【００４４】
　このようにして、評価値テーブル３０７に得られる各意味最小単位ごとのidf値は、その意味最小単位が珍しいものであるほど高い値になる。
【００４５】
　以上の評価値テーブル３０７の作成処理は、図３の評価値計算部３０６の機能に対応する。
【００４６】
　図８は、基本的な意味検索技術を用いた図３の構成を有する検索装置における検索処理を示すフローチャートである。
【００４７】

まず、検索クエリ３１４（図３）は自然言語文で入力される。

次に、検索クエリ３１４が文に分割される（ステップＳ８０１）。

【００４８】

次に、ステップＳ８０２とＳ８０４の制御によるループ処理によって分割で得られたそれぞれの文につき、図４のステップＳ４０４と同様の意味（意味解析）処理が実行されて検索キーのキーワードに追加される（ステップＳ８０３）。

【００４９】

以上の処理は、図３の入力部３０８、意味解析部３０２、および意味最小単位保存部３０３の機能に対応する。

【００５０】

図９は、検索クエリ３１４の例「翻訳家によって修正された単語を辞書に登録する手段をもつ。」から意味最小単位を生成する例を示す図である。図５の１つの文書に対応する検索クエリ３１４の文書から、図５の右側に相当する複数の意味最小単位が生成される。

【００５１】

以上のステップＳ８０２からＳ８０４のループ処理によって、検索クエリ３１４に対応する意味最小単位が自動生成され、この集合が検索キー８０１とされる。

【００５２】

次に、図８のステップＳ８０５とＳ８０７の制御によるループ処理によって検索キー８０１に含まれるそれぞれの意味最小単位（これを意味最小単位（Ａ）と呼ぶ）につき、以下のステップＳ８０６の処理が実行される。図６に例示される構成を有する図３の検索用インデックス３０５から、処理対象の意味最小単位（Ａ）を含むレコード（図６の行に対応する）の文書IDおよび文IDが取得され、記憶される。この処理により、処理対象の意味最小単位（Ａ）がどの文書の何番目の文にマッチするかがわかる。

【００５３】

ステップＳ８０５からＳ８０７のループ処理により、検索キー８０１に含まれる全ての意味最小単位に対して、その意味最小単位を含む文書IDおよび文IDが取得される。

【００５４】

以上の処理は、図３の検索部３０９の機能に対応する。

続いて、データベース中の全文書の評価値が０とされた後（ステップＳ８０８）、ステップＳ８０９とＳ８１２の制御によるループ処理により、上述の処理で検索キー８０１中のいずれかの意味最小単位がマッチした文ごとに、以下のように評価値が算出される。

【００５５】

いま例として、図５に示した文m,n,p,qの４つの文が検索キー８０１にマッチしたとする。ステップＳ８０９とＳ８１２の制御によるループ処理により、それぞれの文に対して評価値が計算される。

【００５６】

まず、処理対象の文（以下、処理対象の文を「文ｉ」とする）に対する評価値が算出される（ステップＳ８１０）。文ｉに対する評価値（Si）は、次式で表される。

文ｉの評価値(Si)＝[検索クエリの意味最小単位の集合(K1,K2,…Kj,…)のうち、
(文ｉに出現するKjのidf値×文ｉにおけるKjの出現回数)
の総和]×M^2 ・・・（２）

【００５７】

上記（２）式において、Mは文ｉに同時に出現するKjの数である。この数Mが増えれば増えるほど、その文iは検索クエリ３１４により良く合致していると考えることができる。そして、Mの数が２個、３個と増えるに従って、その合致度は大幅に向上すると考えられる。そこで、Mの数に従って評価値を大幅に上げるためにMの二乗が乗算される。

【００５８】

いま、理解を容易にするために、検索キー８０１に含まれる全ての意味最小単位のidf値を2.0と仮定する。図５に示される文mについて、その右側の意味最小単位を見てみると、文mは検索クエリ３１４から生成された意味最小単位の６つにマッチしている。意味最

小単位はそれぞれ１回しか文mに出現していない。また、文mに出現する意味最小単位の数は6である。以上から、文mの評価値は（２）式より、

$(2.0*1+2.0*1+2.0*1+2.0*1+2.0*1+2.0*1)\times 6^2$ =432.0　　　（３）

となる。図５の文n,p,qに関しても同様に計算が行われる。

【００５９】

次に、文書の評価値が計算される（ステップＳ８１１）。文書の評価値は、次式により計算される。

【００６０】

文書の評価値(D)＝文nの評価値(Sn)の総計　　　・・・（４）

すなわち、文書の評価値は、その文書に含まれる全ての文の評価値の総和である。具体的な処理としては、ステップＳ８１０で評価値(Si)が算出された処理対象の文iについて、その文iが含まれる文書の評価値に、その文iの評価値(Si)が加算される（ステップＳ８１１）。

【００６１】

以上のステップＳ８０９からＳ８１２の処理が、検索キー８０１内のいずれかの意味最小単位がマッチした文iごとに繰返し実行されることにより、各文iおよびそれが含まれる文書の評価値が算出される。

以上の処理は、図３の評価値計算部３０６の機能に対応する。

【００６２】

最後に、上述のようにして算出された評価値に基づいて文書が例えば降順にソート、すなわちランキングされて、以下の結果が得られる。

【００６３】

いま例えば、図５に示される文書Ａが、文mに加えて特には図示しない文rの２文からなるとし、文rの評価値が18.0だとすると、文書Ａの評価値は、この値と上記（３）式の計算結果の値が加算されることにより、18.0+432.0=450.0と算出される。同様にして文書Ｂ，Ｃ，Ｄに関しても評価値が計算され、例えばそれぞれ、106.0，253.0，90.0となったとする。この結果、ランキングにより、以下の検索結果が、図３の検索結果表示部３１２に表示される。

検索結果１位：　文書Ａ（評価値＝450.0）
検索結果２位：　文書Ｃ（評価値＝253.0）
検索結果３位：　文書Ｂ（評価値＝106.0）
検索結果４位：　文書Ｄ（評価値＝90.0）

【００６４】

また、この例では、以下のように、表現が違ったり、単語間に別の表現が入り込んでいても、意味でマッチングできている。

文m：「翻訳家が修正」が(修正,翻訳家,動作主)にマッチ（検索クエリ３１４では「翻訳家によって修正」）

文m：「単語あるいは修正または…をそれぞれ前記辞書および例文データベースに登録」が(登録,単語,目的語)にマッチ

文q：「単語は辞書３に登録してもよい」が(登録,単語,目的語)にマッチ（「単語」は「登録」の主語ではなく、目的語として解釈される）

【００６５】

ユーザが、例えば文書Ａを選択して表示させたとき、文ごとに評価値が計算されているので、評価値の高い文をハイライト表示するなどができる。

【００６６】

以上の処理は、図３のランキング部３１１および検索結果表示部３１２の機能に対応する。

【００６７】

以上説明した基本的な意味検索技術を用いた検索装置では、検索クエリ３１４から特定された検索キー８０１となる意味最小単位が、検索用インデックス３０５（データベース

）中の意味最小単位（例えば図６参照）と完全一致による検索を行っている。このため、検索対象を限定しすぎて本来マッチすべき意味最小単位がマッチせず、意味が非常に似ている文でも検索漏れがおこることがある。

【００６８】

図１０は、基本的な意味検索技術を用いた検索装置における検索漏れの例を示す図である。

【００６９】

検索クエリ１０００として「肝臓癌に関して、治療成績が向上した年は。」という自然文を入力して、データベース中の近い文書を検索することを考える。検索クエリ１０００を意味解析して意味構造１００１を得、さらに意味最小単位を計算すると、次のような意味最小単位１００２が得られる。

(IMPROVE, YEAR, TIME)

(IMPROVE, CANCER, KANSURU)

(IMPROVE, ABCXYZ, OBJ)

(CANCER, LIVER, MODIFY)

【００７０】

一方で、検索用インデックス３０５（データベース）（図３）中の文書に、以下の文１００３を含む文書があったとする。

「…癌の治療成績について…」

【００７１】

この文１００３は、癌の治療成績について記載されており、検索クエリ１０００と近い意味であるが、この部分を意味解析して得た意味構造１００４から得られる意味最小単位１００５は、次のようになる。

(CANCER, ABCXYZ, MODIFY)

この意味最小単位１００５は、検索キー中の意味最小単位１００２には、1つもマッチしない。意味構造１００１と１００４は、意味が近く、本当はマッチさせたい意味構造である。しかし、検索クエリ１０００から生成される意味構造１００１中では、“CANCER”と”ABCXYZ”は直接つながっていないので、これらの２つのノードを同時に含む意味最小単位は生成されない。この結果、検索漏れが発生する。

【００７２】

この検索漏れを防止するために、意味構造１００１および１００４中のアークをいったん全て切ってしまい、ノード全てに関して組合せを生成し、アークの種類は問わずにマッチさせるという方式が考えられる。図１１は、そのような方式の説明図である。

【００７３】

図１１（ａ）は、単語数が例えば４単語というように少ない検索クエリの文「太郎は花子に本をあげた。」から検索キーを抽出する場合の上記方式の処理例である。この文の意味構造からノードを抽出すると、例えば「GIVE TARO BOOK HANAKO」などのようになる。これから組合せで意味最小単位に展開を行うと、図１１（ａ）の右側に表示されるように、組合せ数として４単語×３＝１２通りの意味最小単位が生成される。

【００７４】

図１１（ｂ）は、検索クエリの文中のキーワードが、例えば１０、２０、３０、４０、ｎ個というように、増えた場合の組合せ数を算出した結果である。このように、上記方式だと、検索クエリの文中のキーワード数が増えると、その組合せ数が膨大となる。ｎ個の場合はｎ×（ｎ－１）通りの組合せとなる。

【００７５】

従って、この方式では、検索クエリの文が長くなると、検索時間が膨大になってしまうという問題点を有している。

【００７６】

さらにこの問題のみならず、ノイズ、すなわち意図しない検索結果となる組合せも多数生成されてしまうため、ユーザが検索に有効な組合せのみを生成する必要がある。

【００７７】
　このような問題点が起こらないようにするための本実施形態について、以下に詳細に説明する。
【００７８】
　図１２は、本実施形態の説明図である。検索クエリ１２００を意味解析し、意味構造１２０１を計算するところまでは、上述した基本的な意味検索技術と同様である。本実施形態では、意味最小単位の集合として検索キーが生成されるのではなく、意味構造による制約１２０２がかけられて検索キーが生成される。これにより、検索漏れを削減し、組合せ数が膨大にならず、検索ノイズが抑制された検索キー１２０３が生成される。
【００７９】
　図１３は、本実施形態による検索装置の構成図である。図３に示した基本的な意味検索技術を用いた検索装置の構成に対して、同じ機能を有する部分には同じ番号を付してある。図１３の構成が図３の構成と異なる部分は、意味解析結果保存部１３０１と、検索キー生成部１３０２である。意味解析結果保存部１３０１は、意味解析部３０２における意味解析結果である意味構造を記憶する。本実施形態では、検索時に、図３の構成のように、検索部３０９は、意味解析部３０２が生成した意味最小単位保存部３０３に保存された意味最小単位の集合として検索キーを使用するのではない。本実施形態では、意味解析結果保存部１３０１に記憶された意味構造１２０１（図１２）を入力とし、検索キー１２０３（図１２）を生成する検索キー生成部１３０２を備える。検索部３０９は、この検索キー生成部１３０２が生成した検索キー１２０３を使って検索処理を実行する。
【００８０】
　図１３の検索装置において、データベース作成時には、図３と同様に、検索対象テキスト入力部３０１、意味解析部３０２、意味最小単位保存部３０３、データベース作成部３０４、および評価値計算部３０６が動作する。本実施形態における図１３の検索装置におけるデータベース作成動作は、図３から図７を用いて前述したものと同様である。
【００８１】
　すなわち、検索対象テキスト入力部３０１は、検索対象文書３１３をそれぞれ文ごとに分割し、文ごとに意味解析部３０２により意味解析を行って意味最小単位を生成し、意味最小単位保存部３０３に保存する。データベース作成部３０４は、意味最小単位保存部３０３に保存された意味最小単位に基づいて、検索用インデックス３０５を作成する。またあらかじめ、評価値計算部３０６が、検索用インデックス３０５に登録されたそれぞれの意味最小単位の出現頻度の評価値を計算し、評価値テーブル３０７に格納する。
【００８２】
　図１４は、本実施形態における検索キー生成のための制御処理を示すフローチャートである。
【００８３】
　まず、ユーザにより入力された検索クエリ３１４を対象に意味解析処理が実行され、意味構造１２０１（図１２参照）が生成される（ステップＳ１４０１）。
【００８４】
　次に、ステップＳ１４０１の意味解析処理で得られた意味構造１２０１が記憶される（ステップＳ１４０２）。
【００８５】
　以上のステップＳ１４０１とＳ１４０２の処理は、図１３の入力部３０８、意味解析部３０２、および意味解析結果保存部１３０１の機能に対応する。
【００８６】
　次に、意味解析結果保存部１３０１に記憶された意味構造１２０１からノードリストが作成される（ステップＳ１４０３）。図１５は、ノードリストの作成処理の説明図である。ここでは、意味構造１２０１に含まれるノードのリスト（ノードリスト）１５０１が作成される。図１５で示すと、ユーザにより入力された検索クエリ１２００の自然文、例えば「翻訳家によって修正された単語を辞書に登録する手段をもつ。」が意味解析されて、

意味構造１２０１が計算される。この意味構造１２０１に含まれるノードを列挙すると、次の通りとなる。

POSSESS
MEANS
REGISTER
WORD
DICTIONARY
CORRECT
TRANSLATOR

【００８７】

次に、ステップＳ１４０４とＳ１４０６の制御によるループ処理によって、ノードリスト１５０１中のそれぞれのノード（以後これを「ノードNi」と呼ぶ）に関して、検索キー生成処理が実行される（ステップＳ１４０５）。

【００８８】

ステップＳ１４０４からＳ１４０６のループ処理によって、全てのノードに対して検索キー１２０３（図１２参照）が作成されたら、生成された検索キー１２０３が出力される（ステップＳ１４０７）。

【００８９】

図１６は、図１４のステップＳ１４０５の検索キー生成処理を示すフローチャートである。

【００９０】

以降、図１５の例で生成されたノードリスト１５０１中のノード"MEANS"を例にとり説明する（ノードNi="MEANS")。

【００９１】

意味解析結果保存部１３０１に保存された意味構造１２０１が参照されることにより、ノードNiを起点として直結するノード（D_Ni_1,D_Ni_2,・・・）が列挙される（ステップ（Ｓ１６０１）。図１５の例において、ノードNi(="MEANS")を起点として直結するノードを列挙すると、1つあり、

D_Ni_1="REGISTER"

となる。なお、ノード"POSSESS"は、"POSSESS"を起点として"MEANS"につながるため対象外となる。

【００９２】

ステップＳ１６０２とＳ１６０５の制御によるループ処理によって、それぞれの直結ノードD_Ni_n（ｎ＝１，２，・・・）に関して、以下の処理が実行される。

【００９３】

まず、ノードNiと直結ノードD_Ni_nをつなぐアーク（これを「アーク（Ａ）」とする）がグループ化される。アーク（Ａ）に対してグループ化されたアークをA_gとする（ステップＳ１６０３）。グループ化については後述する。

【００９４】

次に、検索キー１２０３（図１２）として、「(Ni,D_Ni_A_g)」が生成されて記憶される（ステップＳ１６０４）。

【００９５】

ステップＳ１６０３とＳ１６０４のループ処理の具体例を以下に示す。

＜ループ1回目：D_Ni_1＞

いま、ノードNi(="MEANS")とノードD_Ni_1をつなぐアークは、図１５の意味構造１２０１より、"PP"である。そこで、ステップＳ１６０３で、"PP"がグループ化されて"PP_GROUP"となる。図１７は、アーク名とグループ名との対応関係の例を示す図である。この図に、"PP"がグループ化されて"PP_GROUP"になったことが示されている。

【００９６】

ステップＳ１６０４で、検索キー(MEANS,REGISTER,PP_GROUP) が生成されて記憶される

。
＜ループ終わり＞
次に、意味解析結果保存部１３０１に保存された意味構造１２０１が参照されることにより、ノードNiに間接接続されるノード（I_Ni_1,I_Ni_2,・・・）が列挙される（ステップ（Ｓ１６０６）。間接接続の定義は、直接つながってはいないが、n個のアークによりつながっていることをいう。nはシステムで設定できるが、2以上の数で、［n=2か3か4］のように、複数設定してもよい。またアークの向きは無視するものとする。ここではn=2として説明する。
【００９７】
次に、ステップＳ１６０７とＳ１６１１の制御によるループ処理により、それぞれの間接接続ノードI_Ni_n（ｎ＝１，２，・・・）に関して、以下の処理が実行される。
【００９８】
まず、ノードNiと間接接続ノードI_Ni_nをつなぐ複数のアーク（これを「アークA1」,「アークA2」,・・・とする）がグループ化される。これらのグループ化されたアークをA1_g,A2_gとする（ステップＳ１６０８）。グループ化については後述する。
【００９９】
次に、検索キー１２０３（図１２）として、「(Ni,I_Ni_[A1_g|A2_g|...])」が生成されて記憶される（ステップＳ１６０９）。
【０１００】
さらに、検索キー１２０３（図１２）として、以下のものが生成されて記憶される（ステップＳ１６１０）。
(Ni,*1, [A1_g|A2_g|...]) AND (*1,I_Ni_n,[A1_g|A2_g|...])
【０１０１】
ステップＳ１６０８からＳ１６１０のループ処理の具体例を以下に示す。
まず、図１５に例示される意味構造１２０１で、ノードNi(="MEANS")に間接接続するノードが列挙される（ステップＳ１６０６）。図１５で、"MEANS"に間接接続するノードは、"MEANS"からアーク"PP","OBJ"でつながっているノード"WORD"と、アーク"PP","LOC"でつながっているノード"DICTIONARY"である。
よって、
I_Ni_1="WORD"
I_Ni_2="DICTIONARY"
となる。
【０１０２】
それぞれのI_Ni_nに関して、ステップＳ１６０７からＳ１６１１のループ処理が実行される。
＜ループ1回目：I_Ni_1＞
ノードNi(="MEANS")と間接接続ノードI_Ni_1をつなぐアーク"PP","OBJ"がそれぞれグループ化され、"PP_GROUP","OBJ_GROUP"とされる（ステップＳ１６０８）。図１７にその様子が示されている。
【０１０３】
検索キー１２０３として、「(MEANS,WORD,[PP_GROUP|OBJ_GROUP])」が生成されて記憶される（ステップＳ１６０９）。
【０１０４】
さらに、検索キー１２０３として、以下の検索キーが生成されて記憶される（ステップＳ１６１０）。
(MEANS,*1,[PP_GROUP|OBJ_GROUP])AND(*1,WORD,[PP_GROUP|OBJ_GROUP])
【０１０５】
この検索式の意味するところは、例えば図１８（ａ）に示されるように、ある同じノードを1つ介してMEANSとWORDがつながっている構造を探す、ということである。このとき、アークの向きを指定してもしなくてもよい。アークの向きを指定しない場合は図１８（ｂ

）のようになる。
<ループ2回目：I_Ni_2>
ノードNi(="MEANS")と間接接続ノードI_Ni_2をつなぐアーク"PP","LOC"がそれぞれグループ化され、"PP_GROUP","PLACE_GROUP"とされる（ステップS１６０８）。図１７にその様子が示されている。
【０１０６】
検索キー１２０３として、「(MEANS,DICTIONARY,[PP_GROUP|PLACE_GROUP])」が生成されて記憶される（ステップS１６０９）。
【０１０７】
さらに、検索キー１２０３として、以下の検索キーが生成されて記憶される（ステップＳ１６１０）。

(MEANS,*1,[PP_GROUP|PLACE_GROUP])
AND
(*1,DICTIONARY,[PP_GROUP|PLACE_GROUP])

<ループ終わり>
以上により、１つの処理対象ノードNiに対する、図１６のフローチャートで示される図１４のステップＳ１４０５の検索キー生成処理が終了する。
【０１０８】
上述のようにしてステップＳ１４０４からＳ１４０７までのループ処理によりノードNiごとに生成された検索キーが、図１４のステップＳ１４０７で、図１３の検索部３０９に出力されて、検索キー生成のための制御処理が終了する。
【０１０９】
以上の説明において、ステップＳ１６０３またはステップＳ１６０８において、アークが図１７に示されるようにグループ化されている。アークは文の意味を表すときの役割を示すものであり、例えばOBJやOBJ_Aは目的語を表すが、意味解析処理上かなり細かい分類がされている。本実施形態では、検索のヒット率を向上させるために、細かく分類されているアークのうち同じような関係を表すアークが、例えば図１７に示されるようにグループ化される。
【０１１０】
以上のステップＳ１４０３からＳ１４０７までの処理は、図１３の検索キー生成部１３０２の機能に対応する。
【０１１１】
例えば図１５に例示されるノードNi="MEANS"の場合の検索キー１２０３の出力は、以下の５種類となる。

１．(MEANS,REGISTER,PP_GROUP)
２．(MEANS,WORD,[PP_GROUP|OBJ_GROUP])
３．(MEANS,*1,[PP_GROUP|OBJ_GROUP])
AND
(*1,WORD,[PP_GROUP|OBJ_GROUP])
４．(MEANS,DICTIONARY,[PP_GROUP|PLACE_GROUP])
５．(MEANS,*1,[PP_GROUP|PLACE_GROUP])
AND
(*1,DICTIONARY,[PP_GROUP|PLACE_GROUP])

【０１１２】
上述の検索キー１２０３の出力を受けた図１３の検索部３０９では、検索キー１２０３に含まれる各意味最小単位相当の検索要素において、グループ化されたアークの部分については、例えば次のように処理してよい。検索部３０９は、例えば図１７に示される表を検索キー生成部１３０２から受け取る。検索部３０９は、その表を参照して、検索キー１２０３中でＯＲ結合されている各部分のグループ化されているアークを、そのグループに属する個々のアークに分解する。検索部３０９は、この分解した個々のアークとアーク部

分以外の部分とでそれぞれ意味最小単位を生成し、検索用インデックス３０５の意味最小単位とのマッチングをとる。このとき、検索部３０９は、アスタリスク「*」で示されているノード部分については、ワイルドカード検索を行う。ワイルドカード検索とは、その部分のノードは任意のノードでよいという条件にした検索をいう。あるいは、データベースの作成時に、検索用インデックス３０５内の各意味最小単位のアークの部分をグループ化されたアークに置き換えて、検索キー１２０３中の各グループ化されたアークとマッチングをとるようにしてもよい。

【０１１３】

以上説明した本実施形態によれば、ゴミとなる検索キーの数を抑えて、類似した意味合いの文章まで検索可能として、マッチさせたい文書にマッチする検索キーを生成することにより、検索性能を向上させることが可能となる。

【０１１４】

例えば、図１２に示した「翻訳家によって修正された単語を辞書に登録する手段を持つ」というクエリに対して、任意の２つのノードの組み合わせを生成すると、マッチさせたくない文書にマッチしてしまう。以下は、そのような不要な検索キーである：

(CORRECT,DICTIONARY,*) 「辞書を修正する」「辞書の修正」「修正された辞書」「修正辞書」…

(MEANS,POSSESS,*) 「所有手段」「所有する手段」…

(POSSESS,DICTIONARY,*) 「辞書を持つ」「辞書の所有」…

(REGISTER,TRANSLATOR,*) 「翻訳家が登録する」「翻訳家を登録する」…

【０１１５】

本実施形態によれば、このような不要な検索キーが生成されないように制御することが可能となる。

【０１１６】

なお、以上の実施形態の説明では、間接接続の定義として、アークn個まで(nは固定）で接続するノードのつながりとしたが、図１３の意味解析部３０２での意味解析の結果に応じて、検索キー生成部１３０２が、間接接続の検索単位を定義することもできる。

【０１１７】

図１９は、間接接続の他の制御方法の説明図であり、図１９（ａ）は、ノードの品詞とアークを組み合わせた場合、何段階まで展開して検索キーとするかを制御するためのテーブルの構成例を示す図である。ここで「ノード２」とは、アークが到達するノードを示す。ノード２から別のアークが出発する場合、そのアークのグループにより、展開できるかどうかが決定される。

【０１１８】

例えば、図１９（ｂ）で、"MEANS"ノードを起点とする場合、ノード２は"REGISTER"であるが、このノードは動詞の属性を持つ。"REGISTER"を起点としてのアークはPP_GROUPである。図１９（ａ）の制御テーブルによると、動詞属性でPP_GROUPは「可」となっているので、アークの先の"WORD"は、"MEANS"に対して間接接続をしていることとする。

【０１１９】

続いて、"WORD"が間接接続により到達されたノードとなったので、これをまた「ノード２」とすると、"WORD"は名詞属性を持ち、"WORD"を起点として"PP_GROUP"のアークが出ている。図１９（ａ）によると、名詞属性で"PP_GROUP"は「否」とあるので、その先のノード"XYZXYZ"のノードは最初の起点"MEANS"とは間接接続ではない。

【０１２０】

このように、アーク到達ノードの属性と、アーク到達ノードを起点とするアークの種類によりどこまで間接接続かを決定することが可能となる。

【０１２１】

図２０は、クラウド構成に対応した実施形態の構成例を示す図である。

データベース作成クライアント２０００、データベース作成サーバ２０１０、ファイルサーバ２０２０、検索クライアント２０３０、検索サーバ２０４０、解析サーバ２０５０

、およびこれらの機器をつなぐネットワーク２０７０から構成される。図２０の構成は、図１３に示される実施形態の各構成が、上記各サーバに分散されて設置されている構成を有する。図２０において、図１３の場合と同じ機能を有する部分には、同じ番号を付してある。

【０１２２】

データベース作成時は、データベース作成クライアント２０００のユーザがデータベースに格納したい全ての検索対象文書３１３を入力すると、図１３と同様の検索対象テキスト入力部３０１がそれを受け付ける。そして、データベース作成クライアント２０００内のデータベース作成サーバ接続部２００１が、データベース作成サーバ２０１０に接続する。データベース作成サーバ２０１０では、検索対象テキスト受付部２０１１がデータベース作成クライアント２０００からの検索対象文書３１３を受け付ける。データベース作成サーバ２０１０では、解析サーバ接続部２０１２が解析サーバ２０５０に接続する。この結果、解析サーバ２０５０内の意味解析部３０２が、データベース作成サーバ２０１０が受け付けた検索対象文書３１３に対して意味解析処理を実行する。この意味解析処理の結果である意味最小単位のデータは、解析サーバ２０５０からデータベース作成サーバ２０１０に送られる。データベース作成サーバ２０１０では、解析サーバ接続部２０１２が、意味最小単位の各データを受信して、意味最小単位保存部３０３に保存する。データベース作成サーバ２０１０において、データベース作成部３０４は、意味最小単位保存部３０３に保存されている各検索対象文書３１３ごとの意味最小単位に基づいて、検索用インデックス３０５を作成する。また、評価値計算部３０６が、検索用インデックス３０５に登録された各意味最小単位ごとの出現頻度の評価値を計算し、評価値テーブル３０７を作成する。データベース転送部２０１３は、上述の検索用インデックス３０５および評価値テーブル３０７を、ファイルサーバ２０２０に転送する。

【０１２３】

データベースの検索時には、検索クライアント２０３０のユーザが、検索クエリ１２００（図１２参照）を指定する。この検索クエリ１２００は、検索クライアント２０３０内の入力部３０８で受け付けられる。検索クライアント２０３０では、検索サーバ接続部２０３１が検索サーバ２０４０に接続し、受け付けた検索クエリ１２００を検索サーバ２０４０に送信する。検索サーバ２０４０では、検索クエリ受付部２０４１が検索クライアント２０３０からの検索クエリ１２００を受け付ける。検索サーバ２０４０では、解析サーバ接続部２０４２が解析サーバ２０５０に接続する。この結果、解析サーバ２０５０内の意味解析部３０２が、検索サーバ２０４０が受け付けた検索クエリ１２００に対して意味解析処理を実行する。この意味解析処理の結果である意味構造１２０１（図１２参照）のデータは、解析サーバ２０５０から検索サーバ２０４０に送られる。検索サーバ２０４０では、解析サーバ接続部２０４２が、意味構造１２０１のデータを受信して、意味解析結果保存部１３０１に保存する。検索サーバ２０４０において、検索キー生成部１３０２は、意味解析結果保存部１３０１に保存された意味構造１２０１に基づいて、検索キー１２０３（図１２参照）を生成し、検索部３０９に出力する。検索部３０９は、検索キー生成部１３０２からの検索キー１２０３に基づいて、ファイルサーバ２０２０内の検索用インデックス３０５を検索し、検索結果を受け取る。さらに、検索サーバ２０４０内の評価値計算部３０６が、ファイルサーバ２０２０内の評価値テーブル３０７を参照することにより、検索部３０９で検索された文書ごとの評価値を計算する。検索サーバ２０４０内のランキング部３１１は、評価値計算部３０６での評価値の計算結果に基づいて、検索結果の文書をランキングする。検索部３０９は、ランキングされた検索結果の文書を、検索クライアント２０３０に返信する。検索クライアント２０３０では、検索サーバ接続部２０３１がランキングされた検索結果の文書を受信し、検索結果表示部３１２に表示する。

【０１２４】

以上のクラウド構成により、データベースの作成、検索、記憶、およびそれに伴う意味解析処理を、例えばインターネットであるネットワーク２０６０を介して、分散処理によって効率的に実行することができる。

【０１２５】
　図２１は、上記システムの全部または一部をソフトウェア処理として実現できるコンピュータのハードウェア構成の一例を示す図である。
【０１２６】
　図２１に示されるコンピュータは、ＣＰＵ２１０１、メモリ２１０２、入力装置２１０３、出力装置２１０４、外部記憶装置２１０５、可搬記録媒体２１０９が挿入される可搬記録媒体駆動装置２１０６、及び通信インタフェース２１０７を有し、これらがバス２１０８によって相互に接続された構成を有する。同図に示される構成は上記システムを実現できるコンピュータの一例であり、そのようなコンピュータはこの構成に限定されるものではない。
【０１２７】
　ＣＰＵ２１０１は、当該コンピュータ全体の制御を行う。メモリ２１０２は、プログラムの実行、データ更新等の際に、外部記憶装置２１０５（或いは可搬記録媒体２１０９）に記憶されているプログラム又はデータを一時的に格納するＲＡＭ等のメモリである。ＣＰＵ２１０１は、プログラムをメモリ２１０２に読み出して実行することにより、全体の制御を行う。
【０１２８】
　入出力装置２１０３は、ユーザによるキーボードやマウス等による入力操作を検出し、その検出結果をＣＰＵ２１０１に通知し、ＣＰＵ２１０１の制御によって送られてくるデータを表示装置や印刷装置に出力する。
【０１２９】
　外部記憶装置２１０５は、例えばハードディスク記憶装置である。主に各種データやプログラムの保存に用いられる。
【０１３０】
　可搬記録媒体駆動装置２１０６は、光ディスクやＳＤＲＡＭ、コンパクトフラッシュ（登録商標）等の可搬記録媒体２１０９を収容するもので、外部記憶装置２１０５の補助の役割を有する。
【０１３１】
　通信インタフェース２１０７は、例えばＬＡＮ（ローカルエリアネットワーク）又はＷＡＮ（ワイドエリアネットワーク）の通信回線を接続するための装置である。
【０１３２】
　本実施形態によるシステムは、図１４および図１６のフローチャート等で実現される機能を搭載したプログラムをＣＰＵ２１０１が実行することで実現される。そのプログラムは、例えば外部記憶装置２１０５や可搬記録媒体２１０９に記録して配布してもよく、或いはネットワーク接続装置２１０７によりネットワークから取得できるようにしてもよい。
【０１３３】
　以上の実施形態に関して、更に以下の付記を開示する。
（付記１）
　検索文を意味解析してグラフで表現される意味構造を生成する意味解析部と、
　前記意味解析部によって生成された意味構造上のノード間の意味のつながりの最小単位に対応する検索キーに加えて、前記意味構造上で直接的または間接的に接続するノードの組合せに対応する検索キーを生成する検索キー生成部と、
　前記生成された検索キーにより、検索対象文書を蓄積したデータベース上の検索用インデックスに格納されている当該検索対象文書に対応する意味構造上のノードの組合せとのマッチングを実行することにより、前記検索文にマッチする検索対象文書を検索する検索部と、
　を備えることを特徴とする検索装置。
（付記２）
　前記検索キー生成部は、前記間接的に接続するノードの組合せ間に存在するアークがＯ

Ｒ結合された検索キーを生成する、
　ことを特徴とする付記１に記載の検索装置。
（付記３）
　前記検索部は、前記検索キー中の前記ノードの組合せに対応する前記ＯＲ結合された各アークを有する前記検索用インデックス内の前記ノードの組合せとの間でマッチングを実行する、
　ことを特徴とする付記２に記載の検索装置。
（付記４）
　前記検索キー生成部は、前記間接的に接続するノードの組合せの一方のノードと該組合せ間に存在するノードがワイルドカードで条件指定された検索キーを生成する、
　ことを特徴とする付記１ないし３のいずれかに記載の検索装置。
（付記５）
　前記検索キー生成部は、前記検索キー中の前記ノードの組合せに対応するアークのうち類似する関係を表すアークをグループ化して前記検索キーを生成する、
　ことを特徴とする付記１ないし４のいずれかに記載の検索装置。
（付記６）
　前記検索部は、前記検索キー中の前記ノードの組合せに対応する前記グループ化されたアークに属する各アークを有する前記検索用インデックス内の前記ノードの組合せとの間でマッチングを実行する、
　ことを特徴とする付記５に記載の検索装置。
（付記７）
　前記検索用インデックス内の前記ノードの組合せは前記グループ化されたアークで関係付けられ、
　前記検索部は、前記検索キー中の前記ノードの組合せに対応する前記グループ化されたアークを有する前記検索用インデックス内の前記ノードの組合せとの間でマッチングを実行する、
　ことを特徴とする付記５に記載の検索装置。
（付記８）
　前記検索キー生成部は、前記ノードの属性と当該ノードを起点とするアークの属性により間接接続を定義する、
　ことを特徴とする付記１ないし７のいずれかに記載の検索装置。
（付記９）
　前記意味解析部、前記検索キー生成部、前記検索部、前記検索文を入力する検索クライアント部、前記データベースを作成するデータベース作成部、または該データベース作成部に前記検索対象文書を入力するデータベース作成クライアント部を、ネットワークを介して接続される複数のサーバコンピュータに備える、
　ことを特徴とする付記１ないし８のいずれかに記載の検索装置。
（付記１０）
　検索文を意味解析してグラフで表現される意味構造を生成し、
　前記意味解析によって生成された意味構造上のノード間の意味のつながりの最小単位に対応する検索キーに加えて、前記意味構造上で直接的または間接的に接続するノードの組合せに対応する検索キーを生成し、
　前記生成された検索キーにより、検索対象文書を蓄積したデータベース上の検索用インデックスに格納されている当該検索対象文書に対応する意味構造上のノードの組合せとのマッチングを実行することにより、前記検索文にマッチする検索対象文書を検索する、
　ことを特徴とする検索方法。
（付記１１）
　検索文を意味解析してグラフで表現される意味構造を生成するステップと、
　前記意味解析によって生成された意味構造上のノード間の意味のつながりの最小単位に対応する検索キーに加えて、前記意味構造上で直接的または間接的に接続するノードの組

合せに対応する検索キーを生成するステップと、

前記生成された検索キーにより、検索対象文書を蓄積したデータベース上の検索用インデックスに格納されている当該検索対象文書に対応する意味構造上のノードの組合せとのマッチングを実行することにより、前記検索文にマッチする検索対象文書を検索するステップと、

をコンピュータに実行させるためのプログラム。

【符号の説明】

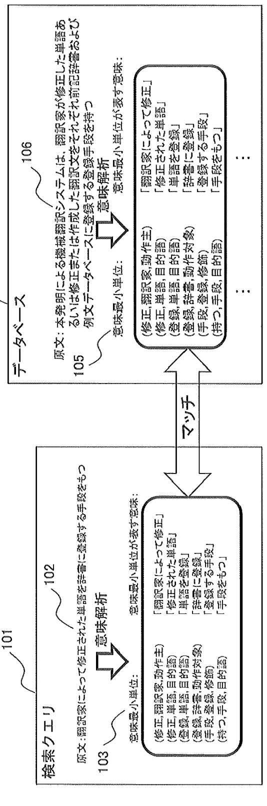
【０１３４】

１０１、１０００、１２００　検索クエリ
１０２　自然言語文
１０３、１０５、２０３、１００２、１００５　意味最小単位
１０４　データベース
１０６　検索対象文
２０１　原文
２０２　有向グラフ
３０１　検索対象テキスト入力部
３０２　意味解析部
３０３　意味最小単位保存部
３０４　データベース作成部
３０５　検索用インデックス
３０６　評価値計算部
３０７　評価値テーブル
３０８　入力部
３０９　検索部
３１０　評価値計算部
３１１　ランキング部
３１２　検索結果表示部
１００１、１００４、１２０１　意味構造
１００３　データベース中の文書の文
１２０２　意味構造による制約
１２０３　検索キー
１３０１　意味解析結果保存部
１３０２　検索キー生成部
１５０１　ノードリスト
２０００　データベース作成クライアント
２００１　データベース作成サーバ接続部
２０１０　データベース作成サーバ
２０１１　検索対象テキスト受付部
２０１２　解析サーバ接続部
２０１３　データベース転送部
２０２０　ファイルサーバ
２０３０　検索クライアント
２０３１　検索サーバ接続部
２０４０　検索サーバ
２０４１　検索クエリ受付部
２０４２　解析サーバ接続部
２０５０　解析サーバ
２０６０　ネットワーク
２１０１　ＣＰＵ
２１０２　メモリ

２１０３　入力装置
２１０４　出力装置
２１０５　外部記憶装置
２１０６　可搬記録媒体駆動装置
２１０７　通信インタフェース
２１０８　バス
２１０９　可搬記録媒体

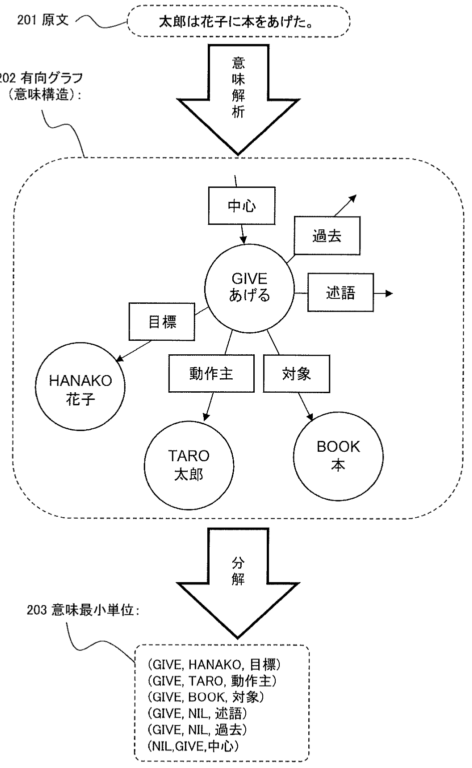
【図１】

本実施形態がベースとする意味検索技術の説明図



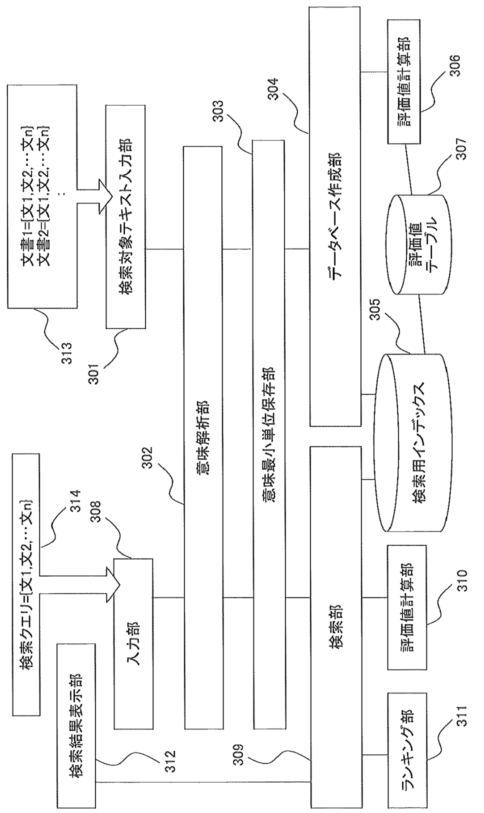
【図２】

自然文から意味最小単位を計算する方法の説明図



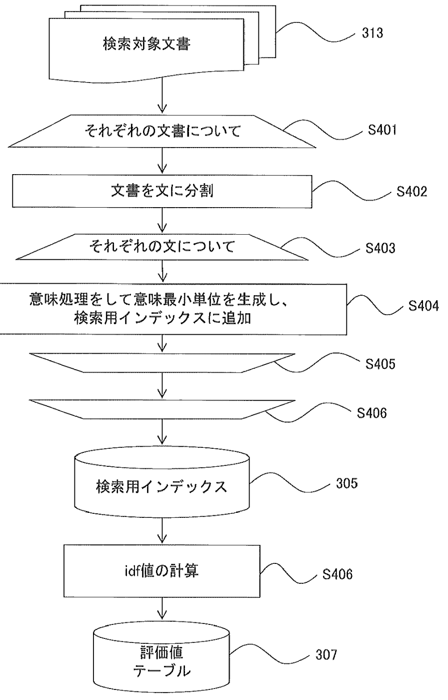
【図３】

基本的な意味検索技術を用いた検索装置の構成図



【図４】

基本的な意味検索技術を用いた検索装置における
データベース作成処理を示すフローチャート

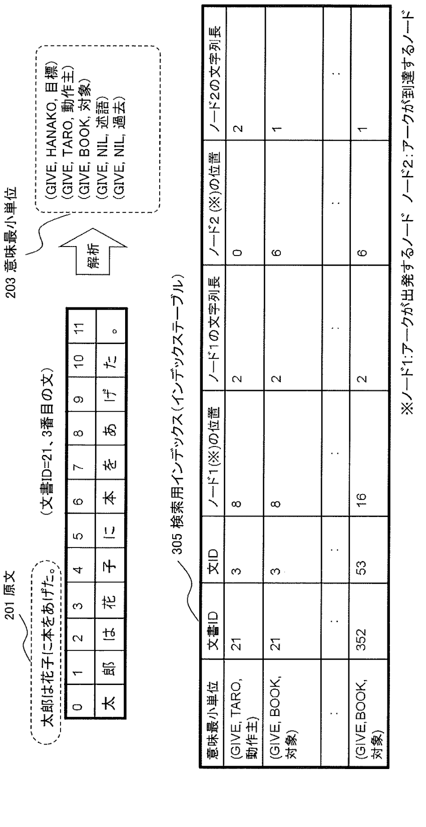


【図５】

データベース作成の例を示す図

【図６】

検索用インデックスに格納される意味最小単位の例を示す図



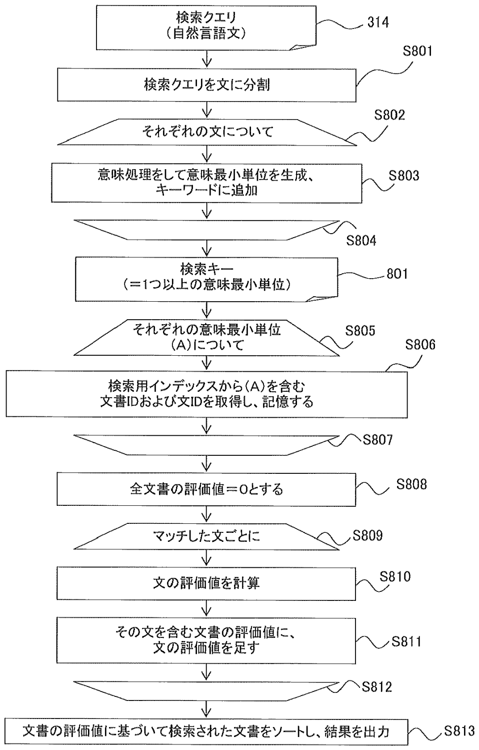
【図７】

評価値テーブルの例を示す図

意味最小単位	idf 値
(NIL,SEE,中心)	4.634235
(SEE,POSSIBLE,ｱｽﾍﾟｸﾄ)	5.325623
(SEE,GIRAFFE,目的語)	6.235621
(SEE,NIL,述語)	3.235321

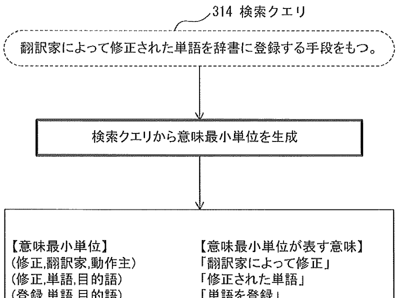
【図８】

基本的な意味検索技術を用いた検索装置における検索処理を示すフローチャート



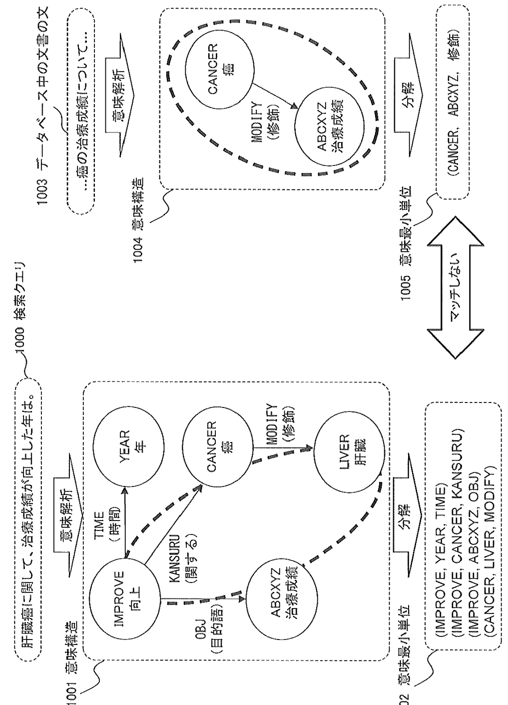
【図９】

検索クエリからの意味最小単位の生成を説明する図



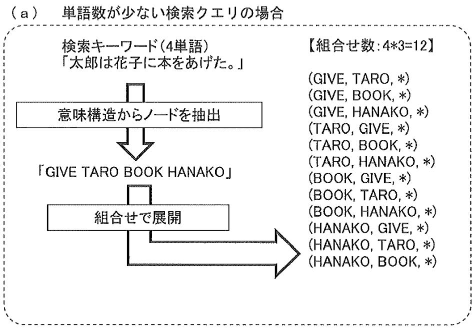
【図１０】

基本的な意味検索技術を用いた検索装置における検索漏れの例を示す図



【図１１】

検索漏れを防止する方式の説明図

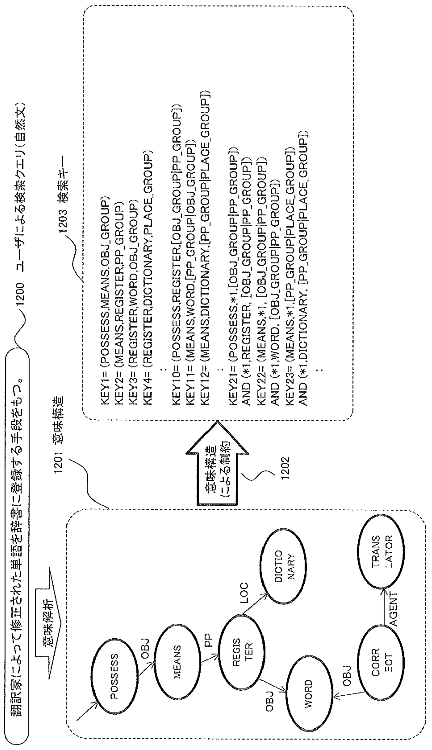


(b) 単語数が多い検索クエリの場合

検索キーワード数	組み合わせ数
10	10*9=90
20	20*19=380
30	30*29=870
40	40*39=1560
n	n*(n-1)

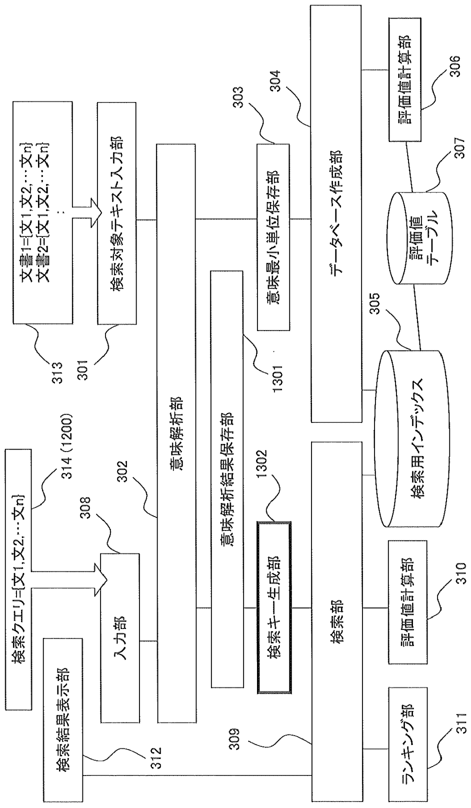
【図１２】

本実施形態の説明図



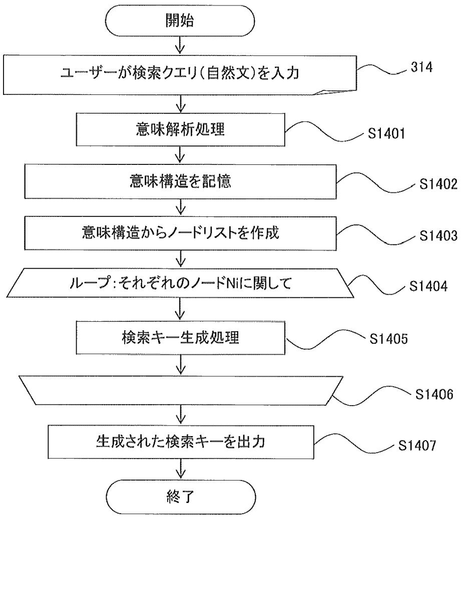
【図１３】

本実施形態による検索装置の構成図



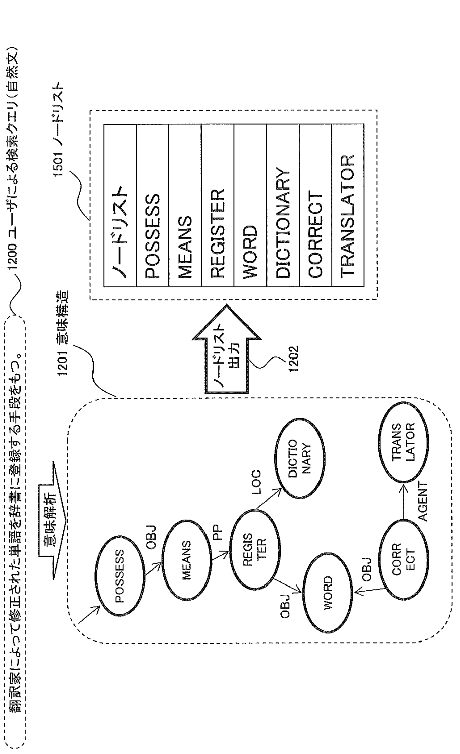
【図１４】

本実施形態による検索キー生成のための制御処理を示すフローチャート



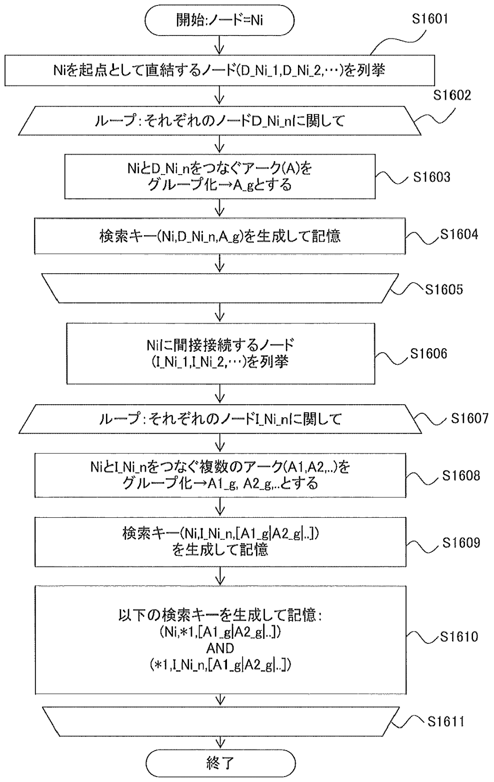
【図１５】

ノードリストの作成処理の説明図



【図１６】

本実施形態による検索キー生成処理を示すフローチャート



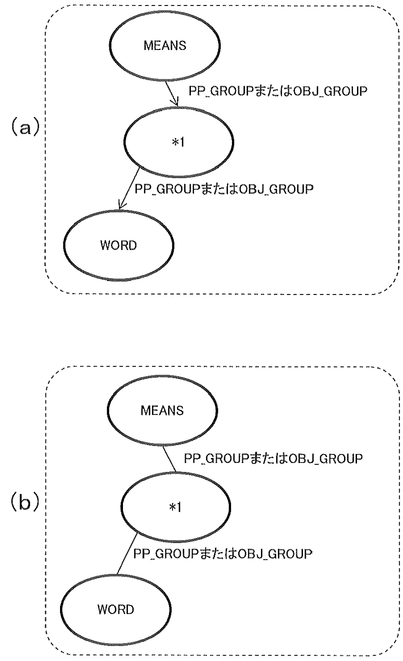
【図１７】

アーク名とグループ名との対応関係の例を示す図

アーク名	グループ名
OBJ	OBJ_GROUP
OBJ_A	OBJ_GROUP
PP	PP_GROUP
LOC	PLACE_GROUP
PLACE	PLACE_GROUP
PLACE_A	PLACE_GROUP
:	:

【図１８】

間接接続の説明図



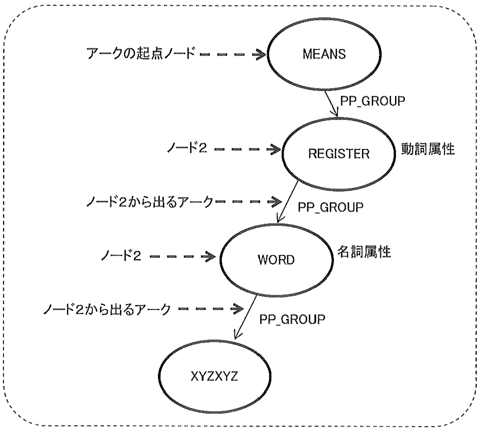
【図１９】

間接接続の他の制御方法の説明図

(a)

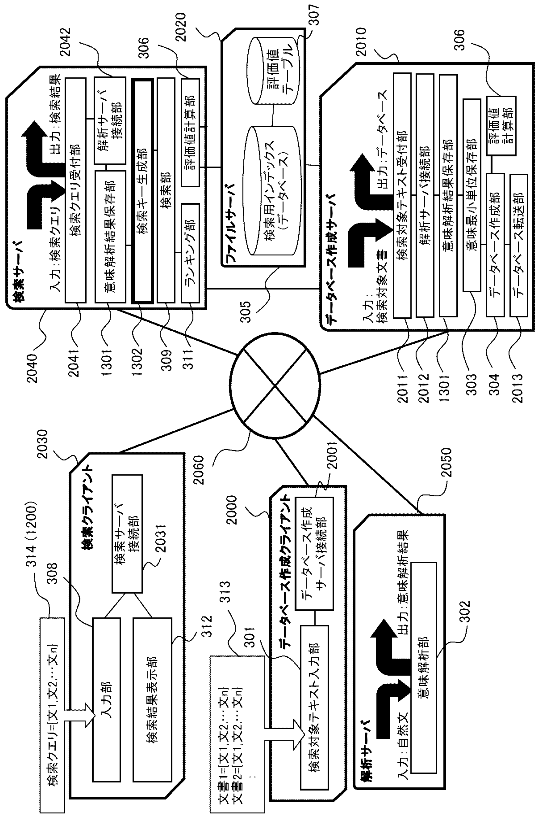
ノード2の属性	ノード2から出るアークの種類	拡張の可否
名詞属性	PP_GROUP	否
動詞属性	PP_GROUP	可
:	:	:

(b)



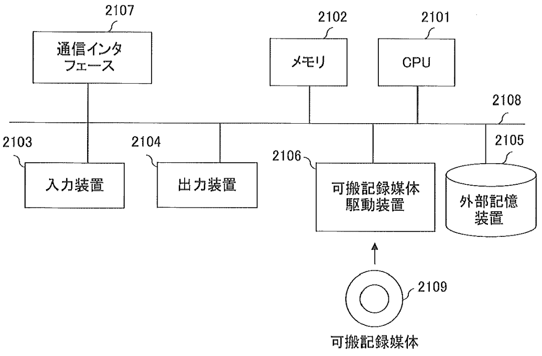
【図２０】

クラウド構成に対応した実施形態の構成例を示す図



【図２１】

本実施形態の機能を搭載した
ソフトウェアプログラムを実行可能な
コンピュータのハードウェア構成例を示す図



フロントページの続き

審査官　齊藤　貴孝

(56)参考文献　特開２０１０－０７９９１５（ＪＰ，Ａ）
特開２００１－０８４２５０（ＪＰ，Ａ）
特開２００８－１５８５６４（ＪＰ，Ａ）
特開２００４－３１８３８１（ＪＰ，Ａ）
特開平０８－００６９５７（ＪＰ，Ａ）
米国特許出願公開第２０１３／００８０４７６（ＵＳ，Ａ１）
特表２０１０－５３８３７５（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　　１７／３０
Ｇ０６Ｆ　　１７／２７