

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3856388号  
(P3856388)

(45) 発行日 平成18年12月13日(2006.12.13)

(24) 登録日 平成18年9月22日(2006.9.22)

(51) Int. Cl. F I  
**G06F 17/30 (2006.01)**  
 G06F 17/30 419B  
 G06F 17/30 330C  
 G06F 17/30 350C

請求項の数 15 (全 18 頁)

|   |  |
|---|--|
| <p>(21) 出願番号 特願2003-110159 (P2003-110159)<br/>                 (22) 出願日 平成15年4月15日(2003.4.15)<br/>                 (65) 公開番号 特開2004-318381 (P2004-318381A)<br/>                 (43) 公開日 平成16年11月11日(2004.11.11)<br/>                 審査請求日 平成15年6月6日(2003.6.6)</p> | <p>(73) 特許権者 301021533<br/>                 独立行政法人産業技術総合研究所<br/>                 東京都千代田区霞が関1-3-1<br/>                 (73) 特許権者 503360115<br/>                 独立行政法人科学技術振興機構<br/>                 埼玉県川口市本町4丁目1番8号<br/>                 (74) 代理人 100107010<br/>                 弁理士 橋爪 健<br/>                 (72) 発明者 橋田 浩一<br/>                 東京都江東区青海2丁目4番6 独立行政法人産業技術総合研究所臨海副都心センター内</p> |
|---|--|

最終頁に続く

(54) 【発明の名称】 類義性計算方法、類義性計算プログラム、類義性計算プログラムを記録したコンピュータ読み取り可能な記録媒体

(57) 【特許請求の範囲】

【請求項1】

情報検索の文脈に応じて類義性を求めるための類義性計算方法であって、  
 処理部は、ノード、ノードのラベル、ノード間のリンク及びリンクのラベルからなるグラフで表現した検索質問Qを記憶部又は入力部から入力し、ラベルJ及びKの間の類義性を示す実数値である類似度T(J, K)を与える部分関数であるシソーラスTを記憶したシソーラス記憶部を参照し、検索質問Qのノード又はリンクxのラベルMとの類似度T(L, M)が大きなラベルLをxに追加して検索質問Q'を作成するステップと、

処理部は、データベースDを参照し、グラフの近似照合により、検索質問Q'のノードとリンクの集合からデータベースDのノードとリンクの集合への部分関数で表現される解候補f'を求めるステップと、

処理部は、解候補f'の解候補としての良さを示す実数値である解候補スコアs(f')を求めるステップと、

処理部は、求めた解候補スコアs(f')とに基づき、検索質問Q、データベースD、及びシソーラスTに応じ、検索質問Qの各ノード及びノ又は各リンクxについて、xの既存のラベルと他のラベルLとの文脈における類義性を示す実数値である文脈類義性S(x, L)を求めるステップと、

処理部は、求めた文脈類義性S(x, L)を、ノード又はリンクxに対応して記憶部に記憶する又は表示部に表示するステップと

を含む前記類義性計算方法。

10

## 【請求項 2】

処理部は、データベースDを参照し、グラフの近似照合により、検索質問Qのノードとリンクの集合からデータベースDのノードとリンクの集合への部分関数で表現される解候補fを求めるステップと

をさらに含む請求項1に記載の類義性計算方法。

## 【請求項 3】

処理部は、入力された検索質問Qを解析してグラフで表現するステップをさらに含む請求項1に記載の類義性計算方法。

## 【請求項 4】

前記グラフで表現するステップは、  
形態素解析によって、文章を語に分割するステップと、  
統語解析によって、分割された語を句又は節に統合するステップと、  
意味解析によって、動作主及び動作対象等の意味構造を求めるステップと、  
照応解析によって、語間の共参照又は照応等の関連を認識するステップと  
を含む請求項3に記載の類義性計算方法。

10

## 【請求項 5】

前記文脈類義性 $S(x, L)$ を求めるステップは、文脈類義性 $S(x, L)$ を、以下の値のいずれか又は複数に関して単調増加として求める請求項1に記載の類義性計算方法。

(1) ノード又はリンクxの各ラベルMに対する類似度 $T(L, M)$

(2) 検索質問Qにおいてノード又はラベルxにラベルLを追加して得られる拡張された検索質問Q'について、検索質問Q'のデータベースDにおける各解候補、つまり、検索質問Q'のノードとリンクの集合からデータベースDのノードとリンクの集合への部分関数、のスコア

20

## 【請求項 6】

検索質問QおよびQ'のノード及びリンクのラベルのスコアは、ラベル毎に記憶部に記憶されており、処理部は、記憶部を参照することでスコアを得る請求項5に記載の類義性計算方法。

## 【請求項 7】

ラベル間の類義性を示すラベルのスコアは、処理部は、類似度 $T(L, M)$ 及びノード又は以前に求めた文脈類義性 $S(x, L)$ に基づき求める請求項5に記載の類義性計算方法。

30

## 【請求項 8】

前記文脈類義性 $S(x, L)$ を求めるステップは、  
シソーラスTから、ノード又はリンクxの各ラベルMについて類似度 $T(L, M)$ を読み出し、類似度 $T(L, M)$ の最大値を求めるステップと、

検索質問Q'のデータベースDにおける各解候補f'に対して解候補スコア $s(f')$ の最大値を求めるステップと、

求めた両方の最大値に従い文脈類義性 $S(x, L)$ を求めるステップとを含む請求項1に記載の類義性計算方法。

## 【請求項 9】

処理部は、文脈類義性 $S(x, L)$ に応じてノード又はリンクxのラベルの候補Lをソートして、各ノード又はリンクxに対応するラベルの候補を表示部に表示する請求項1に記載の類義性計算方法。

40

## 【請求項 10】

処理部は、ノード又はリンクxのラベルの候補Lに対応して文脈類義性 $S(x, L)$ の値を表示部に表示する請求項1に記載の類義性計算方法。

## 【請求項 11】

前記解候補スコア $s(f')$ を求めるステップは、次の量のいずれか又は複数を記憶部から読み出すことにより、解候補スコア $s(f')$ をそれらの量に関する何らかの単調増加関数の値として求める請求項1に記載の類義性計算方法。

(1) ノード $f(u)$ が定義されるノードuの個数と経路 $f(u, v)$ が定義されるリン

50

ク  $u$   $v$  の本数に基づく、 $f$  の定義域の大きさ

(2) 検索質問  $Q$  の各ノード  $u$  について、ノード  $u$  とノード  $f(u)$  との類似性、

(3) 検索質問  $Q$  の各リンク  $u$   $v$  について、リンク  $u$   $v$  と経路  $f(u$   $v)$  との類似性、

(4)  $D$  の経路  $f(u$   $v)$  がノード  $f(u)$  を始点とするような、検索質問  $Q$  のリンク  $u$   $v$  の本数、

(5)  $D$  の経路  $f(u$   $v)$  がノード  $f(v)$  を終点とするような、検索質問  $Q$  のリンク  $u$   $v$  の本数。

【請求項 1 2】

処理部は、入力された検索質問  $Q$  に従い、データベース  $D$  を参照し、検索質問  $Q$  による 10  
データベースの検索結果として解候補集合  $F$  を求める検索ステップと、

処理部は、求めた解候補集合  $F$  を表示部に表示するステップと、

処理部は、解候補集合  $F$  のいくつかの要素が解に該当するかどうかについての入力情報を入力部から入力するステップと、

処理部は、その入力情報に従い、解候補集合  $F$  のいくつかの要素を、解候補集合  $F$  から削除するステップと、

処理部は、入力部からのシソーラス  $T$  及び / 又は検索質問  $Q$  に関する削除、追加又は変更についての入力情報に基づき、シソーラス  $T$  の内容を変更する及び / 又は検索質問  $Q$  のノード、リンク、及び / 又はそれらのラベルを削除又は追加するステップと、

処理部は、利用者から再検索の要求があれば前記検索ステップに戻り処理を繰返し、一 20  
方、その要求がなければ処理を終了するステップと  
を含む請求項 1 に記載の類義性計算方法。

【請求項 1 3】

処理部は、表示部に検索質問  $Q$  を表示するステップと、

処理部は、検索質問  $Q$  の 2 つのノードを結ぶリンクがない場合にそのようなリンクを挿入することを指示する入力情報を入力部から入力するステップと、

処理部は、その入力情報に従い、そのリンクを挿入するステップと、

処理部には、検索質問  $Q$  のリンクを削除することを指示する入力情報を入力部から入力するステップと、

処理部は、その入力情報に従い、そのリンクを削除するステップと、 30

処理部には、検索質問  $Q$  に新たなノードを付加することを指示する入力情報が入力部から入力されるステップと、

処理部は、その入力情報に従い、検索質問  $Q$  にそのノードを付加するステップと、

処理部には、検索質問  $Q$  のノードを削除することを指示する入力情報を入力部から入力するステップと、

処理部は、その入力情報に従い、検索質問  $Q$  からそのノードを削除するステップと、  
をさらに含む請求項 1 に記載の類義性計算方法。

【請求項 1 4】

情報検索の文脈に応じて類義性を求めるための類義性計算プログラムであって、

処理部は、ノード、ノードのラベル、ノード間のリンク及びリンクのラベルからなるグ 40  
ラフで表現した検索質問  $Q$  を記憶部又は入力部から入力し、ラベル  $J$  及び  $K$  の間の類義性を示す実数値である類似度  $T(J, K)$  を与える部分関数であるシソーラス  $T$  を記憶したシソーラス記憶部を参照し、検索質問  $Q$  のノード又はリンクに対してそのラベル  $M$  との類似度  $T(L, M)$  に基づきラベル  $L$  を追加して検索質問  $Q'$  を作成するステップと、

処理部は、データベース  $D$  を参照し、グラフの近似照合により、検索質問  $Q'$  のノードとリンクの集合からデータベース  $D$  のノードとリンクの集合への部分関数で表現される解候補  $f'$  を求めるステップと、

処理部は、解候補  $f'$  の解候補としての良さを示す実数値である解候補スコア  $s(f')$  を求めるステップと、

処理部は、求めた解候補スコア  $s(f')$  に基づき、検索質問  $Q$ 、データベース  $D$ 、及 50

びシソーラスTに応じ、検索質問Qの各ノード及びノ又は各リンクについて、検索質問Qの中のノード又はリンクxの既存のラベルと他のラベルLとの文脈における類義性を示す実数値である文脈類義性 $S(x, L)$ を求めるステップと、

処理部は、求めた文脈類義性 $S(x, L)$ を、ノード又はリンクxに対応して記憶部に記憶する又は表示部に表示するステップと  
をコンピュータに実行させるための類義性計算プログラム。

【請求項15】

情報検索の文脈に応じて類義性を求めるための類義性計算プログラムを記録したコンピュータ読み取り可能な記録媒体であって、

処理部は、ノード、ノードのラベル、ノード間のリンク及びリンクのラベルからなるグラフで表現した検索質問Qを記憶部又は入力部から入力し、ラベルJ及びKの間の類義性を示す実数値である類似度 $T(J, K)$ を与える部分関数であるシソーラスTを記憶したシソーラス記憶部を参照し、検索質問Qのノード又はリンクに対してそのラベルMとの類似度 $T(L, M)$ に基づきラベルLを追加して検索質問Q'を作成するステップと、

処理部は、データベースDを参照し、グラフの近似照合により、検索質問Q'のノードとリンクの集合からデータベースDのノードとリンクの集合への部分関数で表現される解候補 $f'$ を求めるステップと、

処理部は、解候補 $f'$ の解候補としての良さを示す実数値である解候補スコア $s(f')$ を求めるステップと、

処理部は、求めた解候補スコア $s(f')$ に基づき、検索質問Q、データベースD、及びシソーラスTに応じ、検索質問Qの各ノード及びノ又は各リンクについて、検索質問Qの中のノード又はリンクxの既存のラベルと他のラベルLとの文脈における類義性を示す実数値である文脈類義性 $S(x, L)$ を求めるステップと、

処理部は、求めた文脈類義性 $S(x, L)$ を、ノード又はリンクxに対応して記憶部に記憶する又は表示部に表示するステップと  
をコンピュータに実行させるための類義性計算プログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、情報検索の文脈に応じて類義性を求めるための類義性計算方法、類義性計算プログラム、類義性計算プログラムを記録したコンピュータ読み取り可能な記録媒体に係り、特に、情報検索において、ラベル付きグラフに関する個々の検索質問及びそのデータベースとの関係に応じてキーワード間の類義性を動的に求めるための類義性計算方法、類義性計算プログラム、類義性計算プログラムを記録したコンピュータ読み取り可能な記録媒体に関する。

【0002】

【従来の技術】

従来の情報検索は、キーワードやキーワードに対応する識別番号をAND条件やOR条件等の論理式によって組合せたものを検索質問とし、文字列照合と統計的処理によって検索を行なうものであった。利用者とのインタラクションのためには、例えば、解候補集合のいくつかの部分集合について、その各々を特徴付けるキーワード・語句を統計的な方法によって求め、それらのキーワード・語句を検索要求に加えるキーワード・語句の候補として提示して、利用者には選ばせるなどの方法が用いられていた（非特許文献1及び2）。また、従来より、語の間の類義性を語の間の共起関係等に基づいて求める方法は、以下の非特許文献3～5のように、いくつか知られている。

【0003】

【非特許文献1】

林 良彦・小橋 喜嗣（1998） WWW上の検索サービスの技術動向． 情報処理， 39巻9号．

10

20

30

40

50

## 【非特許文献2】

藤田 澄男 (1999) 自然言語処理を利用した情報の検索・分類へのアプローチ、情報処理、40巻4号。

## 【非特許文献3】

Hindle, D. Noun classification from predicate-argument structures. Proceedings of the 28th ACL, pp. 268-275, 1990.

## 【非特許文献4】

Pereira, F., Tishby, N., and Lee, L. Distributional clustering of English words. Proceedings of the 31st ACL, pp.183-190, 1993.

## 【非特許文献5】

Tokunaga, T., Iwayama, M., and Tanaka, H. Automatic thesaurus construction based on grammatical relations. Proceedings of IJCAI '95, pp. 1308-1313, 1995.

## 【0004】

## 【発明が解決しようとする課題】

一般に、語の間の類義性は文脈に依存する。たとえば、「作る」、「書く」、「建てる」という類義語に関して、「家を作る」の「作る」は「書く」よりも「建てる」に似ているが、「答案を作る」の「作る」は「建てる」よりも「書く」に似ている。情報検索で検索質問の中のキーワードを類義語で拡張する際には、一般的なシソーラスを用いるだけではなく、このような文脈依存性を考慮して、たとえば「家を作る」という検索質問においては「作る」の類義語として「建てる」を優先する必要がある。

しかし、通常は、このような文脈の種類は非常に多いので、予め全ての文脈における類義性を求めておくことは事実上不可能である。また、従来、情報検索において個々の検索質問とデータベースによって定まる個別的な文脈に応じて類義性を求める方法はなかった。

## 【0005】

本発明は、情報検索における文脈に応じた類義性を動的に効率良く求めることを目的とする。また、本発明は、情報検索における質問の改訂や、事例ベース推論における事例の類似性の評価に利用することができる類義性計算方法、類義性計算プログラム、類義性計算プログラムを記録したコンピュータ読み取り可能な記録媒体を提供することを目的とする。

## 【0006】

また、本発明の他の目的は、情報検索において適格な情報を利用者を与えることにより、有効なインタラクションを行ない、検索の効率と精度を向上させることにある。

さらに、本発明の他の目的は、検索質問と検索対象とが、自然言語の文章のような不定形な構造を持つグラフとして取り扱われ、その構造を手掛かりとして利用者が検索エンジンと適格なインタラクションを行なうことを可能とし、検索の効率と精度を向上させることにある。

## 【0007】

## 【課題を解決するための手段】

本発明の解決手段によると、

情報検索の文脈に応じて類義性を求めるための類義性計算方法であって、

処理部は、ノード、ノードのラベル、ノード間のリンク及びリンクのラベルからなるグラフで表現した検索質問Qを記憶部又は入力部から入力し、ラベルL及びMの間の類義性を示す実数値である類似度 $T(L, M)$ を与える部分関数であるシソーラスTを記憶したシソーラス記憶部を参照し、検索質問Qのノード又はリンクxに対してラベルLを追加して検索質問Q'を作成するステップと、

処理部は、データベースDを参照し、グラフの近似照合により、検索質問Q'のノードとリンクの集合からデータベースDのノードとリンクの集合への部分関数で表現される解候補f'を求めるステップと、

処理部は、解候補f'の解候補としての良さを示す実数値である解候補スコア $s(f')$ を求めるステップと、

10

20

30

40

50

処理部は、求めた解候補スコア  $s(f')$  及びシソーラス  $T$  に応じ、検索質問  $Q$  の中のノード又はリンク  $x$  の既存のラベルと他のラベル  $L$  との文脈における類義性を示す実数値である文脈類義性  $S(x, L)$  を求めるステップと、

処理部は、求めた文脈類義性  $S(x, L)$  を、ノード又はリンク  $x$  に対応して記憶部に記憶する又は表示部に表示するステップと

を含む前記類義性計算方法、類義性計算プログラム、類義性計算プログラムを記録したコンピュータ読み取り可能な記録媒体が提供される。

【0008】

【発明の実施の形態】

#### 1. 前提の説明

本実施の形態では、文脈の意味構造として、1個以上のノードとそれらのノードを結ぶリンクからなり、各ノードに1個以上のラベルが付いた無向グラフを考える。検索質問  $Q$  及び検索対象であるデータベース  $D$  はいずれもそのようなグラフだとする。また、グラフの間の近似的な照合等に基づいて検索質問  $Q$  や検索範囲のインタラクティブな変更が効果的に行なえるようにする。文書の検索の場合には、たとえば、ノードは語の指示対象を表わし、リンクはそれらの間の意味的な関係を表わし、ラベルは語である。

【0009】

本実施の形態で「検索」とは、検索質問  $Q$  に似たデータベース  $D$  の部分グラフを見付けることである。検索質問  $Q$  のノードのいくつかは、そのような部分グラフのいずれかのノードに対応すると考えられる。その対応関係を検索質問  $Q$  のノードからデータベース  $D$  のノードへの関数によって表わし、その関数を解候補と呼ぶ。また、各解候補のスコア（例えば、類似度、関連度、確率等に関する値）が定義されるとする。スコアの高いいくつかの解候補の集合を解候補集合  $F$  とし、

$F(x) = \{f(x) \mid f \in F\}$  ( $x$  は検索質問  $Q$  のノード、 $f(x)$  はノード  $x$  に対応するデータベース中のノード)

$f(Q) = \{f(x) \mid x \text{ は、検索質問 } Q \text{ のノード}\}$  ( $f \in F$ )

とする。

ここで、検索質問  $Q$ 、解候補集合  $F$  等について具体例で説明する。

【0010】

図1に、ノード、リンク、検索質問  $Q$  についての説明図を示す。

・検索質問  $Q$  のノード  $x$  とそのラベルは、例えば、「関数」、「解析」、「意味」、「自動」である。

・検索質問  $Q$  のリンクは、「関数 - 解析」、「解析 - 意味」、「解析 - 自動」である。

・検索質問  $Q$  は、これらノードとラベルにより構成される、図示のようなものである。

【0011】

図2に、解候補  $f$  において検索質問  $Q$  のノード  $x$  に対応するデータベース中のノード  $f(x)$ 、解候補集合  $F$  においてノード  $x$  に対応するデータベース中のノードの集合  $F(x)$  についての説明図を示す。

・ $f(x)$  は、例えば、ノード（ラベル）「関数」については、 $f(\text{関数})$  と表され（ $f_1(\text{関数})$ 、 $f_2(\text{関数})$ 、...）、「関数」、「プログラム」、「関手」、「関係」、「サブルーチン」、「射影」、「全射」のそれぞれをラベルとするデータベース  $D$  のノードである。

・ $F(x)$  は、例えば、ノード（ラベル）「関数」については、 $F(\text{関数})$  と表され、すべての  $f \in F$  にわたる  $f(\text{関数})$  の全集合 {「関数」、「プログラム」、「関手」、「関係」、「サブルーチン」、「射影」、「全射」} をいう。

【0012】

図3に、検索質問  $Q$  についての解候補  $f$  の値域  $f(Q)$  や解候補集合  $F$  についての説明図を示す。 $f'(Q)$ 、 $f''(Q)$ 、 $f'''(Q)$  は解候補  $f'$ 、 $f''$ 、 $f'''$  の値域である。

・ $f(Q)$  は、「プログラムで...言語を...分析する」、「意図した投資を...表す関数が」、「内容を自動的に...整理したい」、「暗黙の...意思を推測しながら」、「把握できない

10

20

30

40

50

...データの意味を...プログラムに」、「分析に用いた方法を...意味する」のそれぞれに対応する。

・Fはfの集合であり、f(Q)の集合として表示され、{「プログラムで...言語を...分析する」、「意図した投資を...表す関数が」、「内容を自動的に...整理したい」、「暗黙の...意思を推測しながら」、「把握できない...データの意味を...プログラムに」、「分析に用いた方法を...意味する」}をいう。

また、以下に説明する本実施の形態では、シソーラスTとは、例えば、グラフ中のノードのラベルLとラベルMの組から両者の間の類似性の度合いを示す実数値である類似度 $T(L, M)$ への部分関数であり、スコアの計算に用いる。解候補集合Fを求める際には、シソーラス全体Tではなく、シソーラスTの部分集合Rを用いる。この時、例えば、シソーラスTには操作者により、入力部又は記憶部から予め定められた使用可能とされた部分と使用不可とされた部分があり、解候補集合Fを求める際には、T全体ではなく、Tのうち使用可能とされた部分を用いる。スコアの定義、グラフの表現法、及びデータベースDとシソーラスT又はTの部分集合Rと検索質問Qから解候補集合Fを求める方法(後述の図5のフローチャートの「検索実行」及びそのステップS2の説明箇所)には公知のものがいくつかあり、それを適宜用いることができるのでここでは詳細に触れない。

例えば、ラベル「関数」と「解析」との類似度を示すスコアが実数値 $S_c$ (関数、解析)として、シソーラスTにより与えられる。

【0013】

## 2. 類義性計算の前提

以下に類義性を計算するに当たり、その前提となる事項を説明する。なお、以上の説明ではxをノードとしたが、以下、ノード又はリンクをxと拡張する場合がある。データベースD及び検索質問Qはグラフであり、その各ノード及び各リンクは0個以上のラベルを持ち、ノード及びリンクごとに各ラベルは実数値のスコアを持つとする(同一のラベルでも別のノード又は別のリンクにおいては別のスコアをもちうる)。シソーラスTは、ラベルLとラベルMに対し、それらの間の類義性を示す実数値である類似度 $T(L, M)$ を与える部分関数である(部分関数なので、あらゆるLとMに対して類似度 $T(L, M)$ が定義されている必要はない)。なお、スコアは上述のように、類似度、関連度、確率等に関する値であり、類似度 $T(L, M)$ 、拡張前又は以前に求めた文脈類義性 $S(x, L)$ (ノード又はリンクxのラベルとして、ラベルLの類義性を表す実数値であり、詳細は後述する。)等により適宜定めることができる。シソーラスTは、スコア $S_c(L, M)$ をさらに与える部分関数である。

自然言語のデータに関する情報検索の場合、データベースDと検索質問Qは自然言語の表現の意味的な構造を表わすグラフであり、そのノードとリンクのラベルは自然言語の語句であり、シソーラスTは自然言語の語句に関するシソーラスである。

【0014】

図4に、グラフの説明図を示す。

たとえば、「太郎が家を作る。母親がそこに住む。」という文章は、図示のようなグラフで表現できる。ここで、楕円がノードでその中の文字列がそのラベル、矢印付きの線分がリンクでそれに重なる長方形の中の文字列がそのラベルである。この例では、各ノード及び各リンクはちょうどひとつのラベルを持つ。

【0015】

図5に、グラフ作成処理についてのフローチャートを示す。

データベースDに関して検索質問Qに関して、自然言語のデータからこのようなグラフ(意味ネットワーク)を求める作業は、形態素解析、統語解析、及び意味解析等の公知又は周知の技術により自動的に行なうこともできるし、人手で行なうこともできる。たとえば「太郎が家を建てる」という文については、まず形態素解析によって「太郎+が+家+を+建てる」のように語に分割し(S101)、次に統語解析によって「((太郎 が)((家 を) 建てる))」のように分析し(S103)、さらに意味解析によって「太郎」が「建てる」の動作主であり「家」が「建てる」の動作対象であることを求めることが

10

20

30

40

50

できる (S105)。また、「太郎が家を作る。母親がそこに住む。」において、照応解析により、「母親」が太郎の母親であり「そこ」が「家」であることを認識する (S107)。これらの技術は公知又は周知のものであるのでここでは詳述しない。

#### 【0016】

つぎに、図6に、良い解候補  $f$  の性質を表す説明図を示す。

検索のひとつの解候補は、検索質問  $Q$  の一部と対応するデータベース  $D$  の一部であり、検索質問  $Q$  のノードとリンクの集合からデータベース  $D$  のノードとリンクの集合への部分関数  $f$  で表わせる。ここで、検索質問  $Q$  のノード  $u$  に対してノード  $f(u)$  はデータベース  $D$  のノードである。また、検索質問  $Q$  のノード  $u$  からノード  $v$  に達するリンク  $u \rightarrow v$  に対して経路  $f(u \rightarrow v)$  はデータベース  $D$  の経路 (0本以上のリンクの連鎖) である。

部分関数  $f$  の解候補としての良さを表わす解候補スコア  $s(f)$  は次の量のいずれか又は複数に関する単調増加の値として定義できる。たとえば、 $s(f)$  をこれらの量の和とすることが考えられる。

- ・部分関数  $f$  の定義域の大きさ。(つまり、ノード  $f(u)$  が定義されるノード  $u$  の個数と経路  $f(u \rightarrow v)$  が定義されるリンク  $u \rightarrow v$  の本数との和。)

- ・検索質問  $Q$  の各ノード  $u$  について、ノード  $u$  とノード  $f(u)$  との類似性。(たとえば、ノード  $u$  とノード  $f(u)$  とが共有するラベルのスコアの最大値など。)

- ・検索質問  $Q$  の各リンク  $u \rightarrow v$  について、リンク  $u \rightarrow v$  と経路  $f(u \rightarrow v)$  との類似性。(たとえば、リンク  $u \rightarrow v$  と経路  $f(u \rightarrow v)$  とが共有するラベルのスコアの最大値など。ここで、リンク及び経路のラベルは単純なラベル複数個の連鎖でありうると考える。たとえば、リンク  $u \rightarrow v$  がラベル  $L$  を持ち、リンク  $v \rightarrow w$  がラベル  $M$  を持つとき、これらのリンクの連鎖である経路  $u \rightarrow v \rightarrow w$  はラベル  $LM$  を持つ。1本のリンクもこのような複合的なラベルを持ちうるとする。)

- ・図示のようにデータベース  $D$  の経路  $f(u \rightarrow v)$  がノード  $f(u)$  を始点とするような、検索質問  $Q$  のリンク  $u \rightarrow v$  の本数。

- ・データベース  $D$  の経路  $f(u \rightarrow v)$  が  $f(v)$  を終点とするような、検索質問  $Q$  のリンク  $u \rightarrow v$  の本数。

#### 【0017】

つぎに、図7に検索質問  $Q$  とその解候補  $f$  の例についての説明図を示す。

この図は、検索質問  $Q$  (「家を作る」に対応するグラフ) と、図4の内容を持つデータベース  $D$  に関して、解候補  $f$  の具体例を示す。ここで、解候補  $f$  は、ノード  $y$  をラベル「家」を共有するノード  $f(y)$  に対応させ、リンク  $x \rightarrow y$  をラベル「対象」を共有するリンク  $f(x \rightarrow y)$  に対応させ、リンク  $f(x \rightarrow y)$  の終点はノード  $f(y)$  に等しく、ノード  $f(x)$  は未定義である。検索質問  $Q$  及びデータベース  $D$  からこのような解候補を求める問題はグラフの近似照合と呼ばれており、その方法にはいくつか公知又は周知のものがあるので、ここでは詳述しない。

たとえば、自然言語の文書に関する検索においては、最初に与えた検索質問  $Q$  に対していきなり正解が得られないことが多い。そのような場合、検索質問  $Q$  がデータベース  $D$  とより良く照合して良い解候補が生成されるように検索質問  $Q$  を改訂する必要がある。そのためのひとつの方法は、検索質問  $Q$  のノードやリンクにラベルを追加することである。そのラベルは、検索質問  $Q$  の当該のノード又はリンクの既存のラベルと類似していることが望ましい。

#### 【0018】

本実施の形態は、検索質問  $Q$ 、データベース  $D$ 、及びシソーラス  $T$  に応じ、検索質問  $Q$  の各ノード及び各リンクについて、そのラベルと他のラベルとの類義性を求める方法を与える。

$x$  は検索質問  $Q$  の中のノード又はリンクであるとする。ここで、ノード又はリンク  $x$  の既存のラベルと他のラベル  $L$  との文脈類義性  $S(x, L)$  は以下のそれぞれの値 (その値が定義されていれば) に関して単調増加である。

- ・ノード又はリンク  $x$  の各ラベル  $M$  に対し、類似度  $T(L, M)$ 。

10

20

30

40

50



・検索質問Qにおいてノード又はリンクxのラベルにLを追加して得られる検索質問Q'について、

検索質問Q'のデータベースDにおける各解候補(検索質問Q'のノードとリンクの集合からデータベースDのノードとリンクの集合への部分関数)のスコア。(計算のコストを抑制するため、ノード又はリンクxを含むある範囲に検索質問Q'の解候補の定義域を限定しても良い。限定された定義域としては、ノード2個とそれらを結ぶリンク1本からなるものが考えられる。)

#### 【0019】

図8に、検索質問Qの拡張検索質問Q'とその解候補f'についての説明図を示す。

この図は、図7の検索質問Qを拡張して得られる検索質問Q'と、検索質問Q'のデータベースDにおける解候補f'を示す。ここで、検索質問Q'は図7の検索質問Qにおいてノードxのラベルとして、既存のラベル「作る」にL = 「建てる」を加えたものであり、解候補f'は図7の解候補fにおいてノードxを定義域に加え、f'(x)をf(x, y)の始点としたものである。このような解候補f'は解候補スコアが大きいので、ノードxのラベルとしての「建てる」は文脈類義性S(x, L)が大きいことになる。

文脈類義性S(x, L)は、たとえば次式のように定義することができる。ここで解候補f'は検索質問Q'のデータベースDにおける解候補、s(f')はその解候補スコアとする。

$$S(x, L) = \max_M T(L, M) + \max_f s(f')$$

#### 【0020】

### 3. ハードウェア

図9に、類義性計算装置の構成図を示す。

類義性計算装置は、表示部1、入力部2、処理部(CPU)3、主記憶部4、シソーラス記憶部5、データベース(検索対象)6、バス7を備える。

処理部3は、入力部2、表示部1、主記憶部4、シソーラス記憶部5、データベース(検索対象)6とバス7により接続され、各種情報を入出力する。表示部1は、例えば、検索入力、検索出力、検索途中結果等を画面に表示するためのディスプレイ装置である。入力部2は、例えば、検索質問、指示、条件等の検索に必要な各種データ等を入力するための入力手段であり、キーボード、マウス、ポインティングデバイス等の適宜の装置が用いられる。なお、他の装置、記憶媒体等にデータを出力する出力部を備えるようにしてもよい。

主記憶部4には、検索プログラム、初期設定、パラメータ等の各種データや、検索最終結果、中間結果等の検索状況に関するデータが記憶される。主記憶部4は、例えば、次のデータを記憶する。

- ・ノード、ノードのラベル、ノード間のリンク及びリンクのラベルからなるグラフで表現した検索質問Q、検索質問Q'
- ・文脈類義性S(x, L)
- ・解候補f、f'
- ・解候補スコアs(f)、s(f')

シソーラス記憶部5は、検索に必要な各ノードの関係、関連度又は非関連度、類似度又は相違度、確率、確からしさ等を示すデータであるシソーラスTを記憶する。シソーラス記憶部5は、例えば、次のシソーラスTのデータを記憶する。

- ・あるラベルに対する類義語のラベル
- ・ラベルLとMとの間の類似度T(L, M)を与える部分関数

データベース6は、検索対象となるデータ(データベースD)を記憶しており、ノード、ラベル、リンク等が記憶される。データベース6は、例えば、次のデータを記憶する。

- ・ノード、ノードのラベル、ノード間のリンク及びリンクのラベルからなるグラフで表現したデータベースD

#### 【0021】

### 4. 類義性計算のフローチャート

10

20

30

40

50

図10に、類義性計算処理のフローチャートを示す。

CPU(処理部)3は、ノード、ノードのラベル、ノード間のリンク及びリンクのラベルについての情報を含む検索質問Qを主記憶部4又は入力部2から入力する(S201)。つぎに、処理部3は、入力された検索質問Qを解析してグラフで表現するグラフ作成処理を実行し(S203)、検索質問Qを主記憶部4に記憶する(S205)。処理部3は、検索質問Qに対応するデータベースD6を参照し、グラフの近似照合により、検索質問Qのノードとリンクの集合からデータベースDのノードとリンクの集合への部分関数で表現される解候補fを求める(S206)。なお、通常は、fは複数個ある。

処理部3は、ノード、ノードのラベル、ノード間のリンク及びリンクのラベルからなるグラフで表現した検索質問Qを主記憶部4(又は入力部2)から入力し、ラベルL及びMとの間の類義性を示す実数値である類似度 $T(L, M)$ を与える部分関数であるシソーラスTを記憶したシソーラス記憶部5を参照し、検索質問Qのノード又はリンクに対してラベルを追加して検索質問Q'を作成する(S207)。処理部3は、データベースDを参照し、グラフの近似照合により、検索質問Q'のノードとリンクの集合からデータベースDのノードとリンクの集合への部分関数で表現される解候補f'を求める(S209)。なお、通常は、Q'も複数個あり、その各々に対してf'も複数個ある。

つぎに、処理部3は、解候補f'の解候補としての良さを示す実数値である解候補スコア $s(f')$ を求める(S211)。さらに、処理部3は、求めた解候補スコア $s(f')$ と類似度 $T(L, M)$ に基づき、検索質問Q、データベースD、シソーラス及びデータベースDに応じ、検索質問Qの各ノード及び各リンクについて、検索質問Qの中のノード又はリンクxの既存のラベルと他のラベルLとの文脈における類義性を示す実数値である文脈類義性 $S(x, L)$ を求める(S213)。

処理部3は、求めた文脈類義性 $S(x, L)$ を、ノード又はリンクxに対応して記憶部に記憶し(S215)、表示部に表示する(S217)。表示するステップS217では、処理部3は、文脈類義性 $S(x, L)$ に応じてノード又はリンクxのラベルをソートして、各ノード又はリンクxに対応するラベルを表示部に表示することができる。また、処理部3は、ノード又はリンクxのラベルに対応して文脈類義性 $S(x, L)$ の値を表示部に表示してもよい。その表示するタイミングも適宜設定することができる。こうして求められた文脈類義性 $S(x, L)$ を、情報検索における質問の改訂や、事例ベース推論における事例の類似性の評価に利用することができる。また、適宜のステップ又はタイミングで、文脈類義性 $S(x, L)$ を主記憶部4に書き込むこと又はそこから読み出すこと、表示部1に表示することができる。

#### 【0022】

図11に、文脈類義性 $S(x, L)$ を求めるフローチャートを示す。

このフローチャートは上式に基づいて文脈類義性 $S(x, L)$ を求めるひとつの方法を示すものであり、これに限られない。この手順が終了したときの変数Aの値が文脈類義性 $S(x, L)$ である。この手順では、ノード又はリンクxのいずれかのラベルMに対して類似度 $T(L, M)$ が定義されている場合にのみ検索質問Q'の解候補を調べるようになっているが、ラベルの種類が少ない場合等には常に検索質問Q'の解候補を調べても良い。

まず、変数Aに0を代入する(S301)。そして、ノード又はリンクxの各ラベルMについて、類似度 $T(L, M)$ が定義されて変数 $A < \text{類似度 } T(L, M)$ ならば、変数Aに類似度 $T(L, M)$ を代入する(S303)。変数Aが0である場合、このステップは終了するが、変数Aが0でない場合は、変数Bに0を代入する(S307)。検索質問Q'のDにおける各解候補f'について、変数 $B < \text{解候補スコア } s(f')$ ならば変数Bに解候補スコア $s(f')$ を代入する(S309)。そして、変数Aに $A + B$ を代入し(S311)、このステップを終了する。

#### 【0023】

### 5. 情報検索

図12に、情報検索処理のフローチャートを示す。以下に示す情報検索は、一例を示すも

10

20

30

40

50

のであり、これに限定されない。また、検索処理中の適宜のステップ又はタイミングで、文脈類義性  $S(x, L)$  を主記憶部 4 から読み出し、表示又は計算処理等に用いることができる。

まず、初期入力として、データベース D がデータベース記憶部 6 に予め記憶され、シソーラス T 又は T の一部の部分集合 R がシソーラス記憶部 5 に予め記憶されているとする。ステップ S 1 では、CPU 3 は、削除された解候補の集合 G を空に初期設定し、利用者からノード、ノードのラベル、ノード間のリンクに関する情報を含む検索質問 Q の入力を受け付ける。CPU 3 は、検索質問 Q に関するデータを主記憶部 4 等の適宜の記憶部に記憶し、必要に応じてそこから読み出す。

#### 【0024】

ステップ S 2 では、CPU 3 は、表示部 1 に表示された「検索実行ボタン」をクリックすることにより、利用者の要求に応じて検索（又は再検索）を行う。CPU 3 は、入力された検索質問 Q に従いシソーラス記憶部 5 及びデータベース記憶部 6 を参照し、シソーラス T のうち使用可能とされた部分又は T の部分集合 R において定義されるラベル間の類似度を用いて、検索質問 Q によるデータベース D の検索結果として解候補集合 F を求める（上述のようにその方法は公知であるのでここでは述べない）。その際、削除された解候補集合 G の要素である解候補及び削除された解候補集合 G の要素を含む解候補は解候補集合 F に含めない（解候補は関数であり、関数は順序対の集合だから、解候補の間で包含関係が成り立ちうる）。

ステップ S 3 では、CPU 3 は、インタラクションの手掛かりとして以下の (1) ~ (5) の情報を、表示部 1 により利用者に提示する。(2)、(4)、(5) のリストの表示は、たとえば、リストの要素であるラベルを持つノードを含む解候補のスコアの最大値の降順に従う。また、(3) のリストの表示は、たとえば、ノードのラベル L の文脈類義性  $S(x, L)$  の順に従う。なお、文脈類義性  $S(x, L)$  の値をラベル L に対応して表示してもよい。) 利用者は、下記の各情報に応じて箇条書きで記した仕方で解候補集合 F 中の解候補が解かどうかをチェックしたり、解候補集合 F と削除された解候補集合 G とシソーラス T 又は T の部分集合 R と検索質問 Q を変更したりできる。CPU 3 は、それぞれの選択肢についての情報を表示部 1 に表示する。CPU 3 は、利用者から入力部 2 により入力された入力情報に従い、各選択肢の削除、追加又は変更等を行い、主記憶部 4 に記憶し、このデータと関係するシソーラス、検索対象等のデータをシソーラス記憶部 5、データベース 6 から適宜読み取る。

#### 【0025】

図 13 は、表示画面の例を示す説明図である。この図は、自然言語の文書の検索に関して、この手順のステップ S 3 での表示とインタラクションをサポートするインタフェースの例を示す。図の中の (1) ~ (5) は次の (1) ~ (5) と対応する。

##### (1) 解候補集合 F

ここには、解候補スコア  $s(f)$  の高い解候補のリストが表示される。図中、太字は検索質問の中の語のシソーラス拡張にあたる語である。利用者は次のように、この表示に対する操作ができる。

- ・解候補集合 F のいくつかの要素が解かどうかをチェックする。これは、例えば、リストに表示された情報だけで行なえることもあるが、それだけでチェックできない場合には、各解候補をクリックしてその周辺のさらに広い範囲を表示することによって行う。

- ・解候補集合 F のいくつかの要素を解候補集合 F から削除し、削除された解候補集合 G の要素とする。これは図 13 では、F に含まれていた解候補（図では  $f_1$  で示す）を F に含めない（ $f_1$  で示す）ようにすることである。

#### 【0026】

##### (2) 検索質問 Q

ここには、検索質問が表示される。利用者は、次のように、ノードの追加と削除、及びリンクの挿入と削除ができる。

- ・検索質問 Q の 2 つのノードを結ぶリンクがないいくつかの場合にそのようなリンクを挿

10

20

30

40

50

入する。

- ・ 検索質問 Q のいくつかのリンクを削除する。
- ・ 検索質問 Q に新たなノードをいくつか付加する。
- ・ 検索質問 Q のノードをいくつか削除する。

【 0 0 2 7 】

( 3 ) ここには、検索質問 Q に含まれるノードのラベル ( 図 1 3 では「関数」等 ) をシソーラス拡張した結果でスコア又は文脈類義性  $S(x, L)$  の高いものが表示される。なお、文脈類義性  $S(x, L)$  の値をラベル L に対応して表示してもよい。より正確には、このリストは、( 検索質問 Q のノード x ごとに ) ノード x のラベル L についてシソーラス T において  $T(L, M)$  が定義されているようなデータベース D のノードのラベル ( 要素 ) M のリストである。利用者は、次のように、その各要素を検索範囲に含める ( 図 1 3 では示す ) か含めない ( 示す ) かを指定できる。

・ このリストのいくつかの要素 M でシソーラス T の部分集合 R において  $R(L, M)$  が定義されていないものにつき、R の定義を拡張して  $R(L, M) = T(L, M)$  とする。あるいは、このリストのいくつかの要素 M についてシソーラス T において  $T(L, M)$  の定義を使用可能とする。つまり、M を検索範囲に含める。

・ このリストのいくつかの要素 M で  $R(L, M)$  が定義されているものにつき、R の定義を縮小して  $R(L, M)$  を未定義とする。あるいは、このリストのいくつかの要素 M について  $T(L, M)$  の定義を使用不可とする。つまり、M を検索範囲に含めない。

【 0 0 2 8 】

( 4 ) ここでは、検索質問 Q のノード ( 図 1 3 では「関数」等のノード ) に直接つないで検索質問に付加できるノードのラベルが表示される。さらに詳細には、このリストは、( 検索質問 Q のノード x ごとに ) リンク  $y - z$  がデータベース D に含まれて y のラベルが L であるようなノード y とノード z  $F(x)$  が存在するような、ラベル L のリストである。ラベル L に対応するノード y ( リンク  $y - z$  がデータベース D に含まれて y のラベルが L であるようなノード z  $F(x)$  が存在するようなノード y ) が少ない場合には、そのような y ごとに、y の周辺のいくつかのノードのラベルを L に加えたものをリストの要素として表示するようにしてもよい。利用者は、次のように、このリストの各要素によって検索質問 Q を拡張する ( ) かない ( ) かを指定できる。

・ このリストのいくつかの要素 M について、M をラベルとするノード Y とリンク  $x - Y$  とを検索質問 Q に付加する。つまり、M によって検索質問 Q を拡張する。M をリストから選択する代わりに直接入力することも可能である。

【 0 0 2 9 】

( 5 ) ここには、検索質問 Q において 2 つのノード ( 図 1 3 では「関数」と「解析」等 ) の間に入るノードのラベルが表示される。さらに詳細には、このリストは、( 検索質問 Q のリンク  $x - y$  ごとに ) 解候補 f 中のノード  $f(x)$  と  $f(y)$  を結ぶ最短経路がノード z を含み、解候補の値域  $f(Q)$  がノード z を含まないような解候補 f が存在するノード z のラベルのリストである。利用者は、次のように、このリストの要素を検索質問 Q に挿入する ( 図 1 3 では示す ) かない ( 示す ) かを指定できる。

・ このリストの要素をラベルとするノード z とリンク  $x - z$  とリンク  $z - y$  を検索質問 Q に付加する。つまり、この要素を検索質問 Q に挿入する

ステップ S 4 では、利用者から「検索実行ボタン」により再検索の要求があればステップ S 2 に戻る。一方、再検索の要求がなければ処理を終了する。

本発明の情報検索方法又は情報検索装置・システムは、その各手順をコンピュータに実行させるための情報検索処理プログラム、情報検索処理プログラムを記録したコンピュータ読み取り可能な記録媒体、情報検索処理プログラムを含みコンピュータの内部メモリにロード可能なプログラム製品、そのプログラムを含むサーバ等のコンピュータ、等により提供されることができる。

【 0 0 3 0 】

【 発明の効果 】

10

20

30

40

50

本発明によると、以上説明した通り、情報検索における文脈に応じた類義性を動的に効率良く求めることができる。また、本発明によると、情報検索における質問の改訂や、事例ベース推論における事例の類似性の評価に利用することができる類義性計算方法、類義性計算プログラム、類義性計算プログラムを記録したコンピュータ読み取り可能な記録媒体を提供することができる。

【0031】

また、本発明によると、情報検索において適格な情報を利用者を与えることにより、有効なインタラクションを行ない、検索の効率と精度を向上させることができる。

さらに、本発明によると、検索質問と検索対象とが、自然言語の文章のような不定形な構造を持つグラフとして取り扱われ、その構造を手掛かりとして利用者が検索エンジンと適格なインタラクションを行なうことを可能とし、検索の効率と精度を向上させることができる。

10

【図面の簡単な説明】

【図1】ノード、リンク、検索質問Qについての説明図。

【図2】検索質問Qに含まれる各ラベルのシソーラス拡張についての説明図。

【図3】検索質問Qについての解候補及び解候補集合Fについての説明図。

【図4】グラフの説明図。

【図5】グラフ作成処理についてのフローチャート。

【図6】良い解候補fの性質を表す説明図。

【図7】検索質問Qとその解候補fの例についての説明図。

20

【図8】検索質問Qの拡張検索質問Q'とその解候補f'についての説明図。

【図9】類義性計算装置の構成図。

【図10】類義性計算処理のフローチャート。

【図11】文脈類義性 $S(x, L)$ を求めるフローチャート。

【図12】情報検索処理のフローチャート。

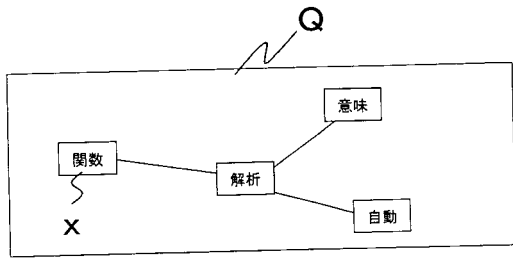
【図13】表示画面の例を示す説明図。

【符号の説明】

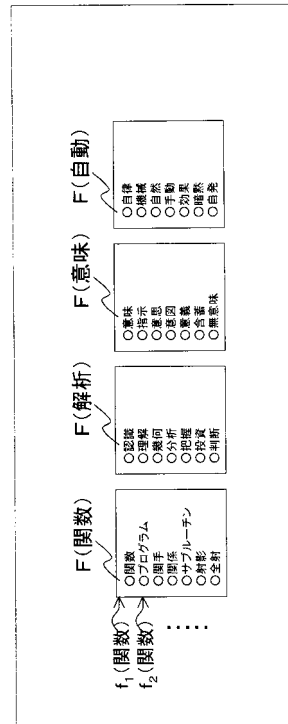
- |   |              |
|---|--------------|
| 1 | 表示部          |
| 2 | 入力部          |
| 3 | CPU          |
| 4 | 主記憶部         |
| 5 | シソーラス記憶部     |
| 6 | データベース(検索対象) |

30

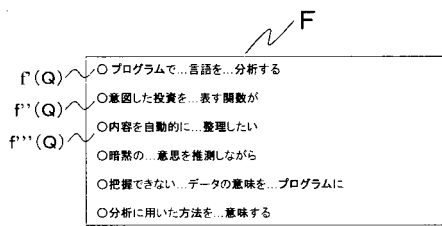
【 図 1 】



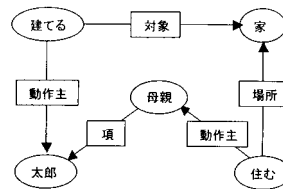
【 図 2 】



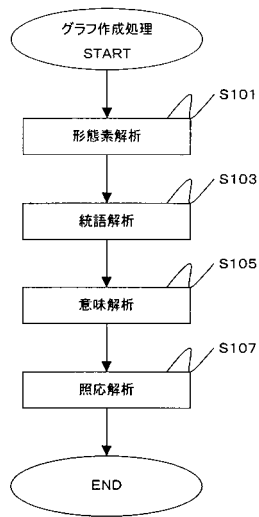
【 図 3 】



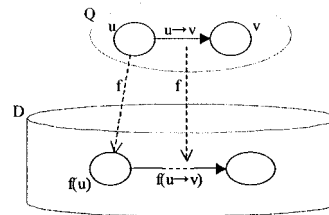
【 図 4 】



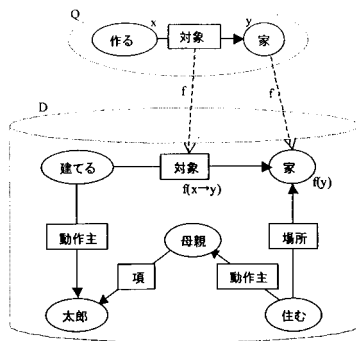
【 図 5 】



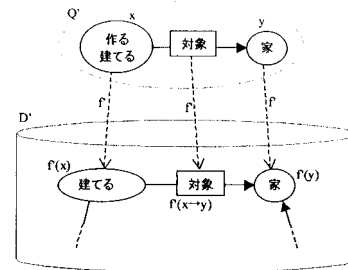
【 図 6 】



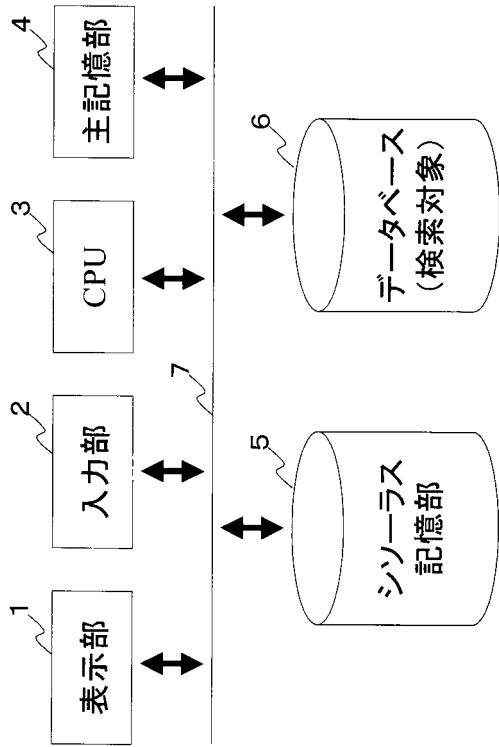
【 図 7 】



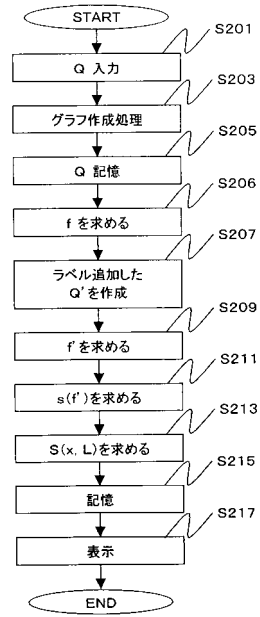
【 図 8 】



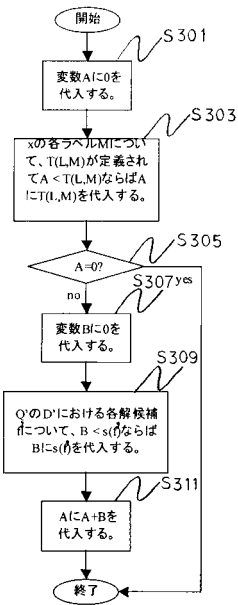
【 図 9 】



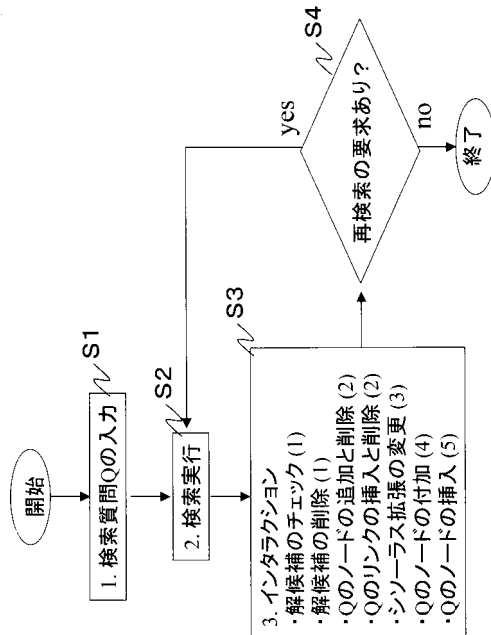
【 図 10 】



【 図 11 】



【 図 12 】







---

フロントページの続き

(72)発明者 宮田 高志

東京都江東区青海2丁目4番6 独立行政法人産業技術総合研究所臨海副都心センター内

審査官 丹治 彰

(56)参考文献 特開平02-058166(JP,A)

特開平04-139580(JP,A)

特開平06-124305(JP,A)

原田実ほか, 意味グラフのマッチングによる事故問い合わせ文からの判例検索システムJCar  
e, 自然言語処理, 日本, 言語処理学会, 2002年 4月10日, 第9巻 第2号, 第3頁~  
第22頁

M. Montes-y-Gomez et al., Flexible Comparison of Conceptual Graphs, Lect Notes Comput  
Sci, ドイツ, 2001年, vol.2113, pp.102--111

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

JSTPlus(JDream2)