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

特許第4637113号
(P4637113)

(45)発行日　平成23年2月23日(2011.2.23)　(24)登録日　平成22年12月3日(2010.12.3)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 340Z
G06F 17/30 419A
G06F 17/30 330B

請求項の数 8　(全 59 頁)

(21)出願番号　特願2006-540102(P2006-540102)
(86)(22)出願日　平成16年11月26日(2004.11.26)
(65)公表番号　特表2007-519086(P2007-519086A)
(43)公表日　平成19年7月12日(2007.7.12)
(86)国際出願番号　PCT/AU2004/001676
(87)国際公開番号　WO2005/052810
(87)国際公開日　平成17年6月9日(2005.6.9)
審査請求日　平成19年11月26日(2007.11.26)
(31)優先権主張番号　2003906611
(32)優先日　平成15年11月28日(2003.11.28)
(33)優先権主張国　オーストラリア(AU)

(73)特許権者　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100076428
弁理士　大塚　康徳
(74)代理人　100112508
弁理士　高柳　司郎
(74)代理人　100115071
弁理士　大塚　康弘
(74)代理人　100116894
弁理士　木村　秀二
(74)代理人　100130409
弁理士　下山　治
(74)代理人　100134175
弁理士　永川　行光

最終頁に続く

(54)【発明の名称】階層データの好ましいビューを構築するための方法

(57)【特許請求の範囲】
【請求項１】
保持手段と、構築手段と、第一の生成手段と、入力手段と、第一の特定手段と、第二の特定手段と、第二の生成手段とを備える装置における方法であって、
前記保持手段が、複数のノードから構成される第一のツリーをデータベースに保持する保持工程と、
前記構築手段が、第一のコマンドに基づいて前記第一のツリーから第二のツリーを構築する構築工程と、
前記第一の生成手段が、前記第二のツリーに基づいて、（１）前記第二のツリーの各ノードの生起する頻度を示す生起頻度テーブルと、（２）前記第二のツリーの複数のノードの同時に生起する頻度を示す共起頻度テーブルとを生成する第一の生成工程と、
前記入力手段が、検索キーワードを入力する入力工程と、
前記第一の特定手段が、前記第一のツリーのノードの中から、前記入力工程で入力された前記検索キーワードに対応する複数のノードをヒットノードとして特定する第一の特定工程と、
前記第二の特定手段が、前記第一のツリーのノードの中から、前記第一の特定工程により特定された前記複数のヒットノードに共通する祖先ノードを特定する第二の特定工程と、
前記第二の生成手段が、前記生起頻度テーブルと、前記共起頻度テーブルと、前記祖先ノードとに基づいて、（ａ）前記検索キーワードに対する検索結果として前記第一のツリ

ーの一部を提示するための、（ｂ）前記第二のツリーと共通のノードを含み、且つ、（ｃ）前記複数のヒットノードに共通する祖先ノードをルートノードとする、コンテキストツリーを生成する第二の生成工程と、
を有することを特徴とする方法。
【請求項２】
前記第一のコマンドは複数存在し、
前記構築工程では、前記複数の第一のコマンドに基づいて、前記第一のツリーから複数の第二のツリーを構築することを特徴とする請求項１に記載の方法。
【請求項３】
前記共起頻度テーブルは、前記第二のツリーの複数のノードが同時に生起し、かつ子ノードを有しない葉ノードである頻度を示すテーブルを含むことを特徴とする請求項１又は２に記載の方法。
【請求項４】
前記共起頻度テーブルは、前記第二のツリーの複数のノードが同時に生起し、かつ子ノードが一つである頻度を示すテーブルを含むことを特徴とする請求項１又は２に記載の方法。
【請求項５】
前記装置は形成手段をさらに備え、
前記形成手段が、前記コンテキストツリーに基づいて、前記検索キーワードに対する検索結果として前記第一のツリーの一部を提示するために、第二のコマンドを形成する形成工程をさらに有することを特徴とする請求項１乃至４のいずれか１項に記載の方法。
【請求項６】
請求項１乃至５の何れか１項に記載の方法の各工程をコンピュータに実行させるためのプログラム。
【請求項７】
請求項６に記載のプログラムを格納したことを特徴とするコンピュータが読取可能な記憶媒体。
【請求項８】
複数のノードから構成される第一のツリーをデータベースに保持する保持手段と、
第一のコマンドに基づいて前記第一のツリーから第二のツリーを構築する構築手段と、
前記第二のツリーに基づいて、（１）前記第二のツリーの各ノードの生起する頻度を示す生起頻度テーブルと、（２）前記第二のツリーの複数のノードの同時に生起する頻度を示す共起頻度テーブルと、を生成する第一の生成手段と、
検索キーワードを入力する入力手段と、
前記第一のツリーのノードの中から、前記入力手段により入力された前記検索キーワードに対応する複数のノードをヒットノードとして特定する第一の特定手段と、
前記第一のツリーのノードの中から、前記第一の特定手段により特定された前記複数のヒットノードに共通する祖先ノードを特定する第二の特定手段と、
前記生起頻度テーブルと、前記共起頻度テーブルと、前記祖先ノードとに基づいて、（ａ）前記検索キーワードに対する検索結果として前記第一のツリーの一部を提示するための、（ｂ）前記第二のツリーと共通のノードを含み、且つ、（ｃ）前記複数のヒットノードに共通する祖先ノードをルートノードとする、コンテキストツリーを生成する第二の生成手段と、
を備えることを特徴とする装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、情報検索（ｒｅｔｒｉｅｖａｌ）の一般分野に関し、特に、関連性の高いデータを大規模な階層データの情報源から自動的に識別および検索することに関する。
【背景技術】

【０００２】
ＸＭＬ（Ｅｘｔｅｎｓｉｂｌｅ Ｍａｒｋｕｐ Ｌａｎｇｕａｇｅ、拡張可能なマーク付け言語）は、情報を保存および交換するための一般的な階層フォーマットにますますなりつつある。ＸＭＬの階層的な性質により、ＸＭＬは、データ・オブジェクト間の関係を捉えるための優れた手段となる一方で、キーワード検索をより困難にもしている。
【０００３】
キーワード検索は、ＸＭＬなどの構造化されたデータ・フォーマットを扱うときには特に重要であるが、その理由は、これによってユーザが特定のキーワードを、データの内部構造を知ることを必要とせずに迅速に見つけられるようになることにある。これはＸＭＬで作業を行うときに難問となるが、それは、キーワード検索の結果を提示するための最適のまたは明らかに好ましい方法がないためである。伝統的な構造化されていないテキストの環境では、データ・システムから、通常、ユーザに対して、見つかったキーワードをその付近にある他のテキストと一緒に提示される。「ヒット」が１つより多い場合、隣接するテキストにより、ヒット同士を区別するための有用なコンテキストを提供され、それにより、ユーザは、関連性の最も高いヒットの選択をユーザの必要に応じて行えるようになる。
【０００４】
構造化されているＸＭＬの環境では、互いに関係するデータはＸＭＬ文書内部のいくつかのばらばらな（ｄｉｓｊｏｉｎｔ）場所にある可能性があるため、「隣接するデータ」（ｎｅｉｇｈｂｏｕｒｉｎｇ ｄａｔａ）のはっきりした概念がない。このため、キーワード検索においてヒットに適したコンテキストを識別または構築することが難しい。したがって、既存のたいていのＸＭＬベースのシステムでは、ある文書の内部でキーワード・ヒットが生じた場合、（ＸＭＬ文書からなるある集まりの中から）単にそのＸＭＬ文書全体を返し、実質的にその文書全体にそのヒットに対するコンテキストの役割をさせている。これは、文書が大きく、ユーザがその内容のすべてを見ることに興味のないときには、望ましくない。
【０００５】
現実的なデータソース、特にデータベースは、しばしば、ユーザがとにかく一度で見たいと思うよりもずっと多くのデータを含んでいる。たとえば、メール注文の店にあるデータベースには、その製品、顧客、納入業者、運送業者、ならびに過去および未発送の注文のすべてについての詳細が含まれているかもしれない。店員は、特定の製品の現在の在庫量を見たいと希望することもあろうし、ある顧客に対する注文のステータスを確認したいと思うこともあろう。他方、店のマネージャーが、何ヶ月にもわたって特定の製品の総売り上げの変動を見たいと希望することがあるかもしれない。こうした場合のそれぞれで、さらに別の関連性のないデータが雪崩のように提示されるとすれば、それはあまりにもユーザを惑わせることになろう。さらに、ユーザは、そのデータベースの構造に詳しくない限り、通常は、自分の興味がある情報を特定することができないはずである。
【０００６】
関連性の高いデータだけを提供するための伝統的な方法は、システム管理者など、データソースの構造に詳しい誰かの準備した、あらかじめ作成済みの「ビュー」の使用によるものである。各ビューは、そのデータソースの何らかのサブセットを引き出してまとめるものであり、はっきりした目的に合わせて作られている。先に挙げた例では、店員は「在庫量」ビューまたは「注文ステータス」ビューを調べるであろうし、マネージャーは「売上」ビューを持ち出すことになろう。
【０００７】
このあらかじめ作成済みのビューを使用するというアプローチは、ありそうな使用法の想定状況をすべて見越しておけるときには十分であるが、キーワード検索には適していない。キーワード検索の操作では、ユーザは、１つまたは複数のキーワードを入力し、システムは、（ＡＮＤのブール値キーワード検索操作を仮定すると）そのキーワードすべての出現を含むデータ・セットまたはビューによって応答する。ＸＭＬなどの階層的な環境で

は、キーワード・ヒットは、階層内の異なる場所にあるいくつかのデータ項目の中に生じる可能性がある。ユーザの与える可能性のある可能なキーワードの組合せすべてを見越しておくことは実現可能でない以上、階層内のどこでヒットが生じるかをあらかじめ判定することは可能でない。したがって、検索の想定状況すべてをまかなうようなあらかじめ作成済みのビューを提供することは可能ではない。

【０００８】

これに類したキーワード検索の問題点は、関係データベースの環境でも存在する。関係データベースは、そのプライマリおよびフォーリンのキーによって併合されたテーブルを含み、ここで各テーブルは、あるエンティティに対する属性の値からなるｎ－組（ｎ－ｔｕｐｌｅ）をそれぞれが表す複数の行を含む。Ｈｒｉｓｔｉｄｉｓ，Ｖ．ａｎｄ Ｐａｐａｋｏｎｓｔａｎｔｉｎｏｕ，Ｙ．、「ＤＩＳＣＯＶＥＲ：Ｋｅｙｗｏｒｄ Ｓｅａｒｃｈ ｉｎ Ｒｅｌａｔｉｏｎａｌ Ｄａｔａｂａｓｅｓ」、Ｐｒｏｃｅｅｄｉｎｇｓ ｏｆ ｔｈｅ ２８ｔｈ ＶＬＤＢ Ｃｏｎｆｅｒｅｎｃｅ、２００２年で記述されている、関係データベースにおけるキーワード検索に対する伝統的な解決策は、キーワード・ヒットすべてを含む併合済みテーブル全体にわたって併合される行の最小のネットワークである、最小併合ネットワーク（ｍｉｎｉｍａｌ ｊｏｉｎｉｎｇ ｎｅｔｗｏｒｋ）を返すことである。このアプローチの問題点は、キーワード・ヒットがデータベース・テーブルのある行の中のどこかで生じた場合、その行全体がそのヒットに対するコンテキストとして返されるという点で、行が最小のデータ「チャンク」として効果的に扱われることである。このことから、極端な量のデータがユーザに提示されることになる可能性があるが、これは、典型的な関係データベースのテーブルが、しばしば、ユーザには普通興味のない多くの列を含んでいるためである。

【０００９】

さらに、上の技法を、ＸＭＬなどの階層的なデータ構造に適合させると、コンテキストの情報が不十分になる可能性がある。階層的な環境では、関係するデータは、階層内の異なるレベルで格納されている可能性があり、このためしばしば、親または祖先のノードあるいはその子からは、キーワード検索に対する非常に有用なコンテキストが、たとえ最小データ・セットには含まれないかもしれない場合でも、提供される可能性がある。

【００１０】

階層データにおけるキーワード検索の問題に取り組むために、いくつかの試みがなされてきている。Ｆｌｏｒｅｓｃｕ，Ｄ．「Ｉｎｔｅｇｒａｔｉｎｇ Ｋｅｙｗｏｒｄ Ｓｅａｒｃｈ ｉｎｔｏ ＸＭＬ Ｑｕｅｒｙ Ｐｒｏｃｅｓｓｉｎｇ」、Ｎｉｎｔｈ Ｉｎｔｅｒｎａｔｉｏｎａｌ Ｗｏｒｌｄ Ｗｉｄｅ Ｗｅｂ Ｃｏｎｆｅｒｅｎｃｅ、２０００年５月では、構造化問い合わせ言語をキーワード検索演算子ｃｏｎｔａｉｎｓで拡張する方法が開示されている。この演算子では、指定したサブツリーがいくつかの指定したキーワードを含む場合に、評価が真となる。ユーザは、この演算子を、問い合わせを構築するときに使用して、望ましくないデータをふるい除けることができる。この有用な機能では、ユーザは、所与のサブツリー内部のヒットしたキーワードの正確な場所を指定する必要はないが、返されるデータの正確なフォーマットをやはり検索問い合わせの中で指定する必要があり、このため、データソースの構造に詳しいことが必要であるため、十分ではない。言い換えると、フリー・テキストのキーワード検索は、ユーザがデータソース全体を検索の結果として受け入れる気がない限り、依然として可能ではないのである。

【００１１】

ＸＭＬデータソースにおけるキーワード検索に対する既存の別のアプローチでは、ユーザが、スキーマ要素からなる所与のリストから、返されるデータのルート・ノードに相当する要素を選択する必要がある。キーワード・ヒットが生じるのが、選択したスキーマ要素の表すデータ要素の子孫ノード内である場合、ヒットしたキーワードを含む、そのデータ要素の下のサブツリー全体がユーザへと返される。このアプローチは、ユーザの介入を必要とするため、扱いにくい。さらに、ユーザは、サブツリー全体を、たとえそのユーザに興味のないデータを含むかもしれない場合でも、受け入れるよう強いられる。

【００１２】
したがって、階層的なデータの環境における１組の関連性の高いデータの判定を、ユーザの介入またはユーザの事前の階層データの構造の知識を必要としない、キーワードの任意の組合せを含むキーワード検索操作に応答して行うための方法が必要とされている。
【非特許文献１】Ｈｒｉｓｔｉｄｉｓ，Ｖ．ａｎｄ Ｐａｐａｋｏｎｓｔａｎｔｉｎｏｕ，Ｙ．、「ＤＩＳＣＯＶＥＲ：Ｋｅｙｗｏｒｄ Ｓｅａｒｃｈ ｉｎ Ｒｅｌａｔｉｏｎａｌ Ｄａｔａｂａｓｅｓ」、Ｐｒｏｃｅｅｄｉｎｇｓ ｏｆ ｔｈｅ ２８ｔｈ ＶＬＤＢ Ｃｏｎｆｅｒｅｎｃｅ、２００２年
【非特許文献２】Ｆｌｏｒｅｓｃｕ，Ｄ.「Ｉｎｔｅｇｒａｔｉｎｇ Ｋｅｙｗｏｒｄ Ｓｅａｒｃｈ ｉｎｔｏ ＸＭＬ Ｑｕｅｒｙ Ｐｒｏｃｅｓｓｉｎｇ」、Ｎｉｎｔｈ Ｉｎｔｅｒｎａｔｉｏｎａｌ Ｗｏｒｌｄ Ｗｉｄｅ Ｗｅｂ Ｃｏｎｆｅｒｅｎｃｅ、２０００年５月
【特許文献１】オーストラリア特許出願第２００３２０４８２４号
【特許文献２】米国特許出願第１０／４６５２２２号
【発明の開示】
【発明が解決しようとする課題】
【００１３】
既存の諸方法の１つまたは複数の不利な点を克服または少なくとも改善することが、本発明の一目的である。
【課題を解決するための手段】
【００１４】
上記の目的を達成する本発明に係る方法は、
保持手段と、構築手段と、第一の生成手段と、入力手段と、第一の特定手段と、第二の特定手段と、第二の生成手段とを備える装置における方法であって、
前記保持手段が、複数のノードから構成される第一のツリーをデータベースに保持する保持工程と、
前記構築手段が、第一のコマンドに基づいて前記第一のツリーから第二のツリーを構築する構築工程と、
前記第一の生成手段が、前記第二のツリーに基づいて、（１）前記第二のツリーの各ノードの生起する頻度を示す生起頻度テーブルと、（２）前記第二のツリーの複数のノードの同時に生起する頻度を示す共起頻度テーブルとを生成する第一の生成工程と、
前記入力手段が、検索キーワードを入力する入力工程と、
前記第一の特定手段が、前記第一のツリーのノードの中から、前記入力工程で入力された前記検索キーワードに対応する複数のノードをヒットノードとして特定する第一の特定工程と、
前記第二の特定手段が、前記第一のツリーのノードの中から、前記第一の特定工程により特定された前記複数のヒットノードに共通する祖先ノードを特定する第二の特定工程と、
前記第二の生成手段が、前記生起頻度テーブルと、前記共起頻度テーブルと、前記祖先ノードとに基づいて、（ａ）前記検索キーワードに対する検索結果として前記第一のツリーの一部を提示するための、（ｂ）前記第二のツリーと共通のノードを含み、且つ、（ｃ）前記複数のヒットノードに共通する祖先ノードをルートノードとする、コンテキストツリーを生成する第二の生成工程と、
を有することを特徴とする。
【００２１】
ここで、本発明の少なくとも１つの実施形態を、図面を参照して説明することにする。
【発明を実施するための最良の形態】
【００２２】
以下に、図面を参照して、この発明の好適な実施の形態を例示的に詳しく説明する。ただし、この実施の形態に記載されている構成要素はあくまで例示であり、この発明の範囲

をそれらのみに限定する趣旨のものではない。

【００２３】

本開示により、階層的なデータの環境における１組の関連性の高いデータの判定を、１つまたは複数のキーワードを含むキーワード検索操作に応答して行うための方法が提供される。好ましい一実施形態は、階層的なデータ構造におけるデータの好ましいビューの構築を、データが過去のエピソードでどのようにアクセスされたかに基づいて行うベイズ確率に基づく方法を含む。より具体的には、この方法では、データ項目の対とベクトルの間の過去の同時生起の頻度を利用して、あるデータ項目がいくつかの他の義務的データ項目にとって関連性が高いことの確率を計算する。通常、義務的データ項目は、キーワード・ヒットを含むものであり、このため、ユーザに対してキーワード検索結果に入れて返さなければならない。義務的でないデータ項目は、義務的データ項目にとって関連性が高いことの確率が高い場合には、検索結果に入れて返されてキーワード・ヒットに対するコンテキストとして役立つことが見込まれる。

【００２４】

ここに開示する配置の、伝統的なあらかじめ作成されたビューに基づくアプローチに対する顕著な特徴は、前者が、単に既存の保存されたビューを返すのではなく、新しいビューを合成できることである。このため、そのような配置は、任意のキーワード組合せを含むキーワード検索操作を扱うことができ、またビューは、動的に生成されるため、あらかじめ作成されたビューの固定のプールから得られるものよりも、個々の操作によりよく合わせることができる。

【００２５】

ここに開示する方法では、通常、いくつもの代替のビューを構築し、そのビューがユーザにとってどのくらい興味があるものかを示すスコアを、各ビューに対して割り当てる。ある実装形態では、構築したもののうちでスコアの最も高い単一のビューがユーザに返される。ある代替実施形態では、最高から最低へとそのスコアに従ってソートされたビューのリストが返される。

【００２６】

ここに開示する方法は、キーワード検索がその第一の動機ではあるが、ともに「Ｍｅｔｈｏｄｓ ｆｏｒ Ｉｎｔｅｒａｃｔｉｖｅｌｙ Ｄｅｆｉｎｉｎｇ Ｔｒａｎｓｆｏｒｍｓ ａｎｄ ｆｏｒ Ｇｅｎｅｒａｔｉｎｇ Ｑｕｅｒｉｅｓ ｂｙ Ｍａｎｉｐｕｌａｔｉｎｇ Ｅｘｉｓｔｉｎｇ Ｑｕｅｒｙ Ｄａｔａ」という名称の２００３年６月１９日に出願されたオーストラリア特許出願第２００３２０４８２４号、およびこれに対応する、２００３年６月２０日に出願された米国特許出願第１０／４６５２２２号に記載のものなどの、階層データの提示の方法を向上させるのに使用することもできる。その公開（ｐｕｂｌｉｃａｔｉｏｎ）では、最も適切な提示タイプ（テーブル、グラフ、プロット、ツリーなど）の選択を階層的なデータソースの構造および内容に基づいて行うための方法が開示されている。この方法は、提示タイプの選択の前に、本開示の好ましい一実施形態を、表示に最も好ましいデータソースのサブセットを自動的に選択する手段として組み込むことによって向上させることができる。データの好ましいサブセットだけをこのようにして表示することがしばしば有用であるが、これは、階層的なデータソースが、しばしば、ユーザに普通興味のあるはずのものよりも多くの情報を含むためであり、このため、本発明の好ましい実施形態などの「興味のない」データをふるい除ける方法は、ユーザの経験をより満足の行く生産的なものにするのに役立つ可能性がある。

【００２７】

以下にある説明のいくつかの部分は、明示的にまたは暗黙に、コンピュータ・メモリ内部のデータに対する操作のアルゴリズムおよび記号表現によって提示してある。こうしたアルゴリズムの説明および表現は、データ処理の当業者がその作業の実質を他の当業者に最も効率よく伝えるのに使用する手段である。アルゴリズムは、ここでは、また一般的には、所望の結果をもたらす自己整合的な一連のステップと考えられている。ステップは、物理量の物理的操作を必要とするものである。こうした量は、必ずではないが、普通、保

存、転送、組合せ、比較、および他の方法で操作するのに堪える電気または磁気の信号の形式をとる。時折、主に一般的に使用されているという理由で、こうした信号をビット、値、要素、記号、キャラクタ、項、数などと呼ぶと好都合であることがわかっている。

【００２８】

しかし、上のおよび類似の用語は、適切な物理量に関連づけられるべきであり、そうした量に適用される便宜的なラベルにすぎないことを銘記すべきである。そうでないと特に述べない限り、また以下から明らかなように、本明細書を通して、「スキャンする」、「計算する」、「判定する」、「置き換える」、「生成する」、「初期化する」、「出力する」などの用語を使用する議論が指すのは、コンピュータ・システムのレジスタおよびメモリ内部の物理（電子的）量として表現されたデータを、コンピュータ・システムのメモリまたはレジスタあるいはそのような他の情報の保存、伝達、または表示の装置内部の物理量として同様に表現される他のデータへと操作または変換するコンピュータ・システムまたは類似の電子的装置の動作およびプロセスであることが理解されよう。

【００２９】

また、本明細書では、この方法の操作を行うための機器の開示を行う。そのような機器は、必要な目的向けに特別に構築することができ、またはコンピュータ内に格納したコンピュータ・プログラムによって選択的に活性化または再構成される汎用コンピュータまたは他の装置を含んでもよい。本明細書で提示するアルゴリズムおよび表示は、特定のどのようなコンピュータまたは他の機器とも固有の関係をもたない。様々な汎用の機械を、本明細書中の教示によるプログラムとともに使用することができる。あるいは、必要な方法ステップを行うためにより特化した機器を構築することが適切であることもある。従来の汎用コンピュータの構造は、下の説明の中で行うことにする。

【００３０】

さらにまた、本明細書では、この方法の操作を行うためのコンピュータ・プログラムを含むコンピュータ可読媒体の開示を行う。コンピュータ可読媒体は、本明細書では、送信元とあて先の間でコンピュータ・プログラムを通信で送るためのどのような伝送媒体も含むと解釈する。伝送媒体は、磁気もしくは光ディスク、メモリ・チップなどのストレージ装置、または汎用コンピュータとのインターフェースを行うのに適した他のストレージ装置を含むことができる。また、伝送媒体は、インターネット・システムの形で例示されるような配線接続（ｈａｒｄ－ｗｉｒｅｄ）の媒体、またはＧＳＭ移動体電話システムの形で例示されるような無線媒体を含んでもよい。コンピュータ・プログラムは、特定のどのようなプログラミング言語およびその実装形態にも限定されるものではない。様々なプログラミング言語およびそのコーディングは、本明細書に含まれる開示の教示を実装するのに使用できることが理解されよう。

【００３１】

添付の図面の１つまたは複数のいずれかにおいて、符号（ｒｅｆｅｒｅｎｃｅ　ｎｕｍｅｒａｌｓ）の同じステップおよび／または特徴への参照を行っている場合、これらのステップおよび／または特徴は、この説明の目的にとっては、それに反する意図を示していない限り、機能または操作は同じである。

【００３２】

この一般にはキーワード検索、特に階層的なデータ構造の構築の諸方法の実施は、好ましくは、図１から３６のプロセスを、コンピュータ・システム３７００内部で実行されるアプリケーション・プログラムなどのソフトウェアとして実装できる、図３７に示すものなどの汎用コンピュータ・システム３７００を用いて行う。特に、キーワード検索の諸ステップは、コンピュータの実行するソフトウェアの中の命令によってもたらされる。この命令は、それぞれが１つまたは複数の特定のタスクを行うための、１つまたは複数のコード・モジュールとして形成することができる。また、このソフトウェアは、第１の部分が検索の方法を行い、第２の部分が第１の部分とユーザとの間のユーザ・インターフェースを管理する、２つの別々の部分に分割することができる。次いで、このソフトウェアは、たとえば、下で説明するストレージ装置を含むコンピュータ可読媒体に格納することがで

きる。このソフトウェアは、そのコンピュータ可読媒体からコンピュータへとロードされ、次いで、そのコンピュータによって実行される。そのようなソフトウェアまたはコンピュータ・プログラムがその上に記録してあるコンピュータ可読媒体が、コンピュータ・プログラム製品である。そのコンピュータ内のコンピュータ・プログラム製品の使用により、好ましくは、キーワード検索および階層的なデータ構造の構築のための有利な機器がもたらされる。

【００３３】

コンピュータ・システム３７００は、コンピュータ・モジュール３７０１、キーボード３７０２やマウス３７０３などの入力装置、プリンタ３７１５、ディスプレイ装置３７１４、およびラウドスピーカ３７１７を含む出力装置によって形成される。モデム（Ｍｏｄｅｍ、Ｍｏｄｕｌａｔｏｒ－Ｄｅｍｏｄｕｌａｔｏｒ）トランシーバ装置３７１６は、コンピュータ・モジュール３７０１で使用して、たとえば、電話回線３７２１または他の機能媒体（ｆｕｎｃｔｉｏｎａｌ　ｍｅｄｉｕｍ）を介して接続可能な、通信ネットワーク３７２０との間の通信が行われる。モデム３７１６は、インターネット、および、ＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ、構内通信網）やＷＡＮ（Ｗｉｄｅ　Ａｒｅａ　Ｎｅｔｗｏｒｋ、広域通信網）など、他のネットワーク・システムへのアクセスを得るのに使用することができ、一部の実装形態では、コンピュータ・モジュール３７０１へと組み込むこともできる。

【００３４】

コンピュータ・モジュール３７０１は、通常、少なくとも１つのプロセッサ・ユニット３７０５と、たとえば半導体ＲＡＭ（ｒａｎｄｏｍ　ａｃｃｅｓｓ　ｍｅｍｏｒｙ）およびＲＯＭ（ｒｅａｄ　ｏｎｌｙ　ｍｅｍｏｒｙ）から形成される、メモリ・ユニット３７０６とを含む。また、モジュール３７０１は、ビデオ・ディスプレイ３７１４およびラウドスピーカ３７１７へと結合されるオーディオ－ビデオ・インターフェース３７０７、キーボード３７０２およびマウス３７０３、および自由選択でジョイスティック（図示せず）に対する入出力インターフェース３７１３、ならびにモデム３７１６およびプリンタ３７１５に対するインターフェース３７０８を含むいくつもの入出力（Ｉ／Ｏ）インターフェースを含む。一部の実施形態では、モデム３７１６を、コンピュータ・モジュール３７０１内部、たとえばインターフェース３７０８内部に組み込んでもよい。ストレージ装置３７０９が提供されており、通常、ハード・ディスク・ドライブ３７１０およびフロッピー・ディスク・ドライブ３７１１を含む。また、磁気テープ・ドライブ（図示せず）を使用してもよい。ＣＤ－ＲＯＭドライブ３７１２は、通常、データの非揮発性の情報源として提供されている。コンピュータ・モジュール３７０１のコンポーネント３７０５から３７１３は、通常、通信を、相互接続されたバス３７０４を介して、また当業者に知られているコンピュータ・システム３７００の従来の利用モード（ｍｏｄｅ　ｏｆ　ｏｐｅｒａｔｉｏｎ）になるような仕方で行う。ここで説明する配置を実施できるコンピュータの例としては、ＩＢＭ－ＰＣおよびその互換機、Ｓｕｎ　Ｓｐａｒｃｓｔａｔｉｏｎｓ、またはそれから発展した類似のコンピュータ・システムがある。

【００３５】

通常、アプリケーション・プログラムは、ハード・ディスク・ドライブ３７１０上にあり、プロセッサ３７０５により、その実行中に読み込みおよび制御が行われる。このプログラムおよびネットワーク３７２０からフェッチされるどのようなデータの中間ストレージも、半導体メモリ３７０６を用いて、おそらくハード・ディスク・ドライブ３７１０と協力して達成される。一部の場合には、このアプリケーション・プログラムは、ＣＤ－ＲＯＭまたはフロッピー・ディスク上にエンコードされてユーザへと供給され、対応するドライブ３７１２または３７１１を介して読み込まれるが、または代替方法として、ユーザがネットワーク３７２０からモデム装置３７１６を介して読み込むのでもよい。またさらに、このソフトウェアのコンピュータ・システム３７００へのロードは、他のコンピュータ可読媒体から行うことも可能である。本明細書で使用する「コンピュータ可読媒体」という用語は、命令および／またはデータを実行および／または処理のためにコンピュータ

・システム３７００へと提供するのにあずかるどのようなストレージ媒体または伝送媒体も指す。ストレージ媒体の例としては、フロッピー・ディスク、磁気テープ、ＣＤ－ＲＯＭ、ハード・ディスク・ドライブ、ＲＯＭまたは集積回路、光磁気（ｍａｇｎｅｔｏ－ｏｐｔｉｃａｌ）ディスク、またはＰＣＭＣＩＡカードなどのコンピュータ可読カードなどがあり、そのような装置がコンピュータ・モジュール３７０１の内部または外部にあることを問わない。伝送媒体の例としては、無線または赤外線の伝送チャネルならびに別のコンピュータまたはネットワーク化装置へのネットワーク接続、電子メール転送およびＷｅｂサイト上に記録された情報を含むインターネットまたはイントラネットなどがある。

【００３６】

階層的な環境におけるキーワード検索は、その１つまたは複数のキーワードが生起する階層的なデータ構造内のノードまたは要素を識別し、次いで、他のどのような要素がそのキーワードにとって関連性が高いかを判定することを含む。典型的なキーワード検索の想定状況では、ユーザへと提示される結果のデータは、第１のデータ構造から抽出され、検索キーワードのすべてまたは一部およびこうしたキーワードにとって関連性が高いと考えられる他のデータを含む第２の階層的なデータ構造である。そのような、キーワード検索操作の結果としてユーザへと提示される階層的なデータ構造を、コンテキストツリー（ｃｏｎｔｅｘｔ　ｔｒｅｅ）と呼ぶ。

【００３７】

ＸＭＬではしばしばあることであるが、検索されている階層的なデータに支配的な（ｇｏｖｅｒｎｉｎｇ）スキーマがあるとき、一般に有利なのは、そのスキーマのレベルで動作する関連性の高いデータを識別するための方法を使用することである。すなわち、スキーマ表現内部の要素が、解析されて、検索キーワードにとって関連性が高いかどうかが判定される。次いで、関連性の高いスキーマ要素によって総体的に表現されるデータソース内のデータ項目のすべてのインスタンスが、キーワード検索操作の結果としてユーザへと返される。ＸＭＬでは、支配的なスキーマは、それ自体が別の階層的なデータ構造である、ＸＭＬ　Ｓｃｈｅｍａの形をとることが可能である。ＸＭＬ　Ｓｃｈｅｍａでは、関連するＸＭＬデータの構造、そのＸＭＬデータ内の要素および属性のリスト、およびそれらの親子関係が指定される。ＸＭＬ　Ｓｃｈｅｍａ内の要素または属性のそれぞれは、通常、ＸＭＬデータ内の要素および属性の多くのインスタンスを表現しているため、ＸＭＬ　Ｓｃｈｅｍａは、潜在的にずっと小さなデータ構造であり、このため、より効率よく解析することができる。

【００３８】

しばしば、キーワードの検索を、１つより多くの階層的なデータソースの中で行うことが望ましい。階層的なデータソースそれぞれは、それ自体はツリー構造をしているが、複数のデータソースを考慮すると、結果となるデータ構造はより一般的な形式をとる可能性がある。データベースの環境で不可避的に生じるそのようなある形式を図１に示している。この構造は、本質的に、共有ノードのあるいくつものツリーを含み、ここで、ツリーそれぞれは、別々の階層的なデータソースのスキーマを表現しており、共有ノードは、内容が複数のデータソースにわたるデータ・ビューの結果である。具体的には、図１の点線の矩形１００５および１０１０は、第１および第２のデータソースのスキーマをそれぞれ表し、ノード１０１５は、第１のデータソースからのノード１０２０および１０２５と第２のデータソースからのノード１０３０を一緒にするデータ・ビューのルート・ノードである。図１の複数の共有ツリーの構造は、本明細書ではこれをスキーマ・グラフ（ｓｃｈｅｍａ　ｇｒａｐｈ）と呼ぶが、有向非巡回グラフの特別な形式であり、どの２つのノード間にも多くとも１つの有向パスがあるという重要な特徴をもつ。たとえば、ノード１０１５からノード１０３５へは有向パスが１つしかなく、このパスはノード１０２０を通過する。

【００３９】

スキーマ・グラフは、好ましくは、キーワード検索操作の前に構築され、階層的なデータソースの初めは互いに素な個々のツリー構造をしたスキーマからなる。次いで、こうし

たスキーマツリーは、１つより多くのデータソースにわたるデータ・ビューが作成されると、併合される。データ・ビューは、通常、（ＸＭＬの環境におけるＸＱｕｅｒｙなどの）問い合わせを含み、データベース管理者またはユーザが作成することができる。どちらの場合でも、データベース・システムは、好ましくは、こうした問い合わせをログまたは記録をそのストレージ装置で行う。スキーマ・グラフの構築中、ログ済みの問い合わせそれぞれのスキーマ表現が作成され、そのスキーマ・グラフへと挿入される。こうすると、２つ以上の別々のスキーマツリーの併合が、挿入されるスキーマが図１に示すこれらのツリーからのノードを含む場合には、行われることになる。スキーマ・グラフへと挿入されるスキーマが、ただ１つのデータソースからのノードしか含まないこともあり得るのであり、その場合には、別々のスキーマツリーの併合は生じない。その代わりに、挿入操作によって、単に、新しいノードがスキーマ・グラフ内の単一のスキーマツリーからの既存のノードへと追加およびリンクされることになる。

【００４０】

スキーマ・グラフの更新は、絶えず、新しい問い合わせのログのたびに行うこともでき、または新しい問い合わせおよびデータ・ビューがある期間にわたって集められた後、１つまたは複数の機会に行うこともできる。スキーマ・グラフの更新の頻度にかかわらず、キーワード検索操作が開始されると、その操作の時点で現在のものである（ｃｕｒｒｅｎｔ）スキーマ・グラフは、ユーザへと返されるデータ・ビューを判定するのに使用されるものである。この文書の残りの部分では、「スキーマ・グラフ」という用語は、キーワード検索操作が行われる時点で現在のものである（ｃｕｒｒｅｎｔ）スキーマ・グラフを指す。

【００４１】

単一のデータソースの場合でのように、複数の階層的なデータソースの内部でのキーワード検索は、まず、「ヒット」ノードと呼ばれる、検索キーワードの見つかったスキーマ・グラフの内部のノードを識別し、次いで、「コンテキスト」ノードと呼ばれる、そのヒット・ノードにとって関連性の高いノードを識別することを含む。次いで、ヒット・ノードおよびコンテキスト・ノードを含むデータ構造が構築され、ユーザへと提示される。ヒット・ノードは１つより多くのデータソースにあり得るため、ユーザへと提示される結果のデータ構造は、複数のデータソースにわたる可能性がある。また、結果のデータ構造は、その意図している適用例は階層的なデータの環境においてであるため、好ましくは、ツリー構造をしている。

【００４２】

図４に、キーワード検索操作の好ましい構成４０００および一般化された利用モード（ｍｏｄｅ　ｏｆ　ｏｐｅｒａｔｉｏｎ）を示している。構成４０００は、ネットワーク内で一緒に接続された、ＰＣクライアント４００５、データ・サーバ４０１０、データベース４０１５、キーワード検索クライアント４０２５、およびインデックス・サーバ４０３０を含む。

【００４３】

装置４００５、４０１０、４０２５、および４０３０のそれぞれは、通常、システム３７００などの対応する汎用コンピュータ・システムによって形成され、それぞれはネットワーク３７２０によってリンクされるが、これを図４にもっぱら概念的に示している。この概念的な図示を使用して、様々な装置４００５、４０１０、４０２５、および４０３０の間の、ネットワーク３７２０の両側で生じるデータ・フローの煩雑でない表現を提供している。様々な装置４００５、４０１０、４０２５、および４０３０は、必要、適当、または好都合なときには、数のより少ない別個のコンピュータ・システム３７００へとまとめることができる。たとえば、一部の実装形態では、サーバ４０１０と４０３０を組み合わせて１つのコンピュータ・システム３７００にし、クライアント４００５と４０２５を組み合わせて別のコンピュータ・システム３７００にし、それらのシステム３７００はネットワーク３７２０でリンクされるようにすることが好都合である可能性がある。

【００４４】

データベース４０１５に格納されているデータへのアクセスは、通常、ＰＣクライアント４００５でブラウズを行うユーザによって行われる。クライアント４００５内で動作するブラウズ用アプリケーションからは、コマンドの発行が、好ましくは、ＸＱｕｅｒｙ４００６の形式で行われ、次いで、これがデータ・サーバ４０１０へと送信される。ＸＱｕｅｒｙ４００６は、それぞれ、ログ４０２０に記録され、データ・サーバ４０１０によって解析され、その後、要求されたデータ４００７がデータベース４０１５からフェッチされ、ＰＣクライアントのユーザ４００５へと送付される。ある時点、好ましくは十分な量のＸＱｕｅｒｙ４００６のログが行われた後で、インデックス・サーバ４０３０が活性化され、ログ済みのＸＱｕｅｒｙ４０２０が、インデックス・テーブル４０３５を作るために解析される。このプロセスは、データベース４０１５内に格納されたデータのスキーマ・グラフ表現、および、ログ済みのＸＱｕｅｒｙ４０２０によって表現されるその既存のビューを構築し、これらのビューに関連する様々な頻度テーブルを作り、データベース内の検索可能なキーワードを識別し、１つまたは複数のコンテキストツリーの決定およびコンテキストツリーそれぞれに対応するＸＱｕｅｒｙを構築し、最後に、これらのキーワードおよびＸＱｕｅｒｙのインデックス・テーブル４０３５への記録を、後で迅速に取り出せるように行うことを含む。

【００４５】

インデックス・テーブル４０３５の作成プロセスが完了した後は、システム４０００では、キーワード検索クライアント４０２５で呼び出されたキーワード検索操作を行う準備ができている。ユーザの入力した検索キーワード４０２６は、インデックス・サーバ４０３０へと送信され、ここで、そのルック・アップがインデックス・テーブル４０３５と突き合わせて行われ、１つまたは複数のＸＱｕｅｒｙ４０３１が取り出され、検索キーワードにとっての関連性の高さに従って適切にランク付けされて、ユーザへと提示される。ユーザがＸＱｕｅｒｙ４０２７をリストから選択すると、そのＸＱｕｅｒｙ４０２７は、キーワード検索クライアント４０２５によってデータ・サーバ４０１０へと送信され、これは適切なデータ４０１１で応答する。

【００４６】

１つまたは複数の階層的なデータソースを含むキーワード検索の方法２０００を、図２のフローチャートで要約している。方法２０００は、好ましくは、インデックス・サーバ４０３０のコンピュータ上で実行される。方法２０００は、ステップ２００５で始まり、ここでスキーマ・グラフ内のヒット・ノードが識別される。ＸＭＬの環境では、ヒット・ノードがスキーマ・グラフ内で生じる可能性のある仕方が２つある。すなわち、（ｉ）その要素名が検索キーワードのうちの１つを含む可能性があり、または（ｉｉ）それが表現する１つまたは複数のＸＭＬノードが検索キーワードのうちの１つを含む可能性がある。ステップ２００５に続いて、ステップ２０１０では、スキーマ・グラフ内のヒット・ノードおよびコンテキスト・ノードで表現されるデータソース内のノードをそれぞれが含む、スキーマ・グラフ内のコンテキストツリーが識別される。最後にステップ２０１５で、識別されたコンテキストツリーがＸＱｕｅｒｙへと変換され、ランク付けされたリストとしてユーザに提示される。

【００４７】

図２のステップ２０１０の表すコンテキストツリーを識別するための方法の説明を、ここで詳細に行う。まず、スキーマ・グラフ内に単一のヒット・ノードがある特別な場合に対する方法を提示し、その後、１つより多くのヒット・ノードを含む場合を扱えるより一般的な方法が続く。どちらの方法も、２フェーズで動作する。その第１は、スキーマ・グラフのヒット・ノードからのボトム・アップ走査（ｔｒａｖｅｒｓａｌ）であって、その親および祖先のうちのどれがコンテキスト・ノードであるかを判定し、第２のフェーズは、これからトップ・ダウン的に進んで、その子孫のうちのどれが同様にコンテキスト・ノードであるかを判定する。次いで、ヒット・ノードの、コンテキスト・ノードと判定された最上位の祖先は、キーワード検索操作の結果としてユーザへと提示されるコンテキストツリーのルート・ノードに相当する。スキーマ・グラフ内のノードがコンテキスト・ノー

ドであるかどうかを判定する目的で、好ましくは、少なくとも生起頻度テーブルおよび共起頻度テーブルが維持される。前者では、スキーマ・グラフ内の各ノードがログ済みの問い合わせまたはデータ・ビューの中で生起する頻度が記録されるのに対し、後者では、スキーマ・グラフ内のノードの対が、同じログ済みの問い合わせまたはデータ・ビューの中で共起する頻度が記録される。スキーマ・グラフが、新しいノードを含む新しい問い合わせまたはデータ・ビューで更新されると、この新しいノードに相当する新しいエントリが生起頻度テーブルに追加され、それぞれには、そのノードが新しく、それまでに観測されたことがないことを示す頻度の初期値１が与えられる。同様にして、２つの新しいノードまたは新しいノードと既存のノードを含む、新しい問い合わせからのノード対それぞれについては、新しいエントリが共起頻度テーブルに追加され、頻度の初期値１が与えられるのに対して、新しい問い合わせの中にない、新しいノードと既存のノードを含むノード対それぞれについては、新しいエントリは共起頻度テーブルに追加されるが、頻度の初期値０が与えられる。

【００４８】

スキーマ・グラフが走査される際、ヒット・ノードの生起数のもとでの、ノードごとの生起確率が計算される。こうした条件付き確率の値の計算または近似は、それまでにログの行われた問い合わせまたはデータ・ビューが原因で頻度テーブルに保存された値から行われ、あるノードがコンテキスト・ノードであるかどうかの判定に使用される。

【００４９】

以下は、単一のヒット・ノードがある特別な場合に対する第１の方法の説明である。このヒット・ノードをＸで表すことにする。

【００５０】

第１のフェーズでは、スキーマ・グラフを通してのボトム・アップ走査はノードＸを始めにして行われる。Ｘの祖先Y_iのそれぞれを順番に考え、そのＸの生起のもとでの生起確率を計算する。

【００５１】

【数１】

$$\Pr[Y_i \mid X] = \frac{\Pr[Y_i \wedge X]}{\Pr[X]} \approx \frac{freq(Y_i, X)}{freq(X)} \qquad \text{式1}$$

【００５２】

ここで、確率の値は、生起頻度ｆｒｅｑ（Ｘ）および共起頻度ｆｒｅｑ（Y_i、Ｘ）で近似されている。後者は、ＸとY_iが共起する頻度を表し、ここでY_iはＸの祖先である。どちらも、前に述べた生起頻度テーブルおよび共起頻度テーブルから直接に得られる。Ｘの祖先ノードに関して計算されたこれらの確率の値から、Ｘのもとで、特定の祖先Y_iがルート・ノードである確率を決定することが可能である。$Z_1, \ldots Z_n$がY_iの親ノードを表すとすると、

$$\Pr[Y_i \text{ root} \mid X] = \Pr[\neg Z_1 \wedge \ldots \wedge \neg Z_n \wedge Y_i \mid X] \qquad \text{式2}$$

となる。

【００５３】

すなわち、Ｘのもとで、Y_iがルートである確率は、Y_iが存在しその親のどれもが存在しない確率である。式２の右辺を展開すると、

$$\Pr[Y_i \text{ root} \mid X] = \Pr[\neg Z_1 \wedge \ldots \wedge \neg Z_n \mid Y_i \wedge X] \Pr[Y_i \wedge X]$$

$$= (1 - \Pr[Z_1 \vee \ldots \vee Z_n \mid Y_i \wedge X]) \Pr$$

［Ｙ ｉ ∧Ｘ］ 式3

が得られる。

【００５４】

Ｚ １，．．Ｚ ｎは、Ｙ ｉが与えられると互いに排他的であることから（Ｙ ｉは、実際のどのような階層的なデータ構造でも、多くとも１つの親をもつことができる）、

【００５５】

【数２】

$$\Pr[Y_i\ root \mid X] = \left(1 - \sum_j \Pr[Z_j \mid Y_i \wedge X]\right)\Pr[Y_i \mid X]$$

$$= \Pr[Y_i \mid X] - \sum_j \Pr[Z_j \mid Y_i \wedge X]\Pr[Y_i \mid X]$$

$$= \Pr[Y_i \mid X] - \sum_j \Pr[Z_j \wedge Y_i \mid X] \qquad \text{式 4}$$

【００５６】

しかし、Ｚ ｊとＸの間には多くとも１つの有向パスしかない（スキーマ・グラフの１つの特徴）ことから、このパスはＹ ｉを含まなければならないことになり、これより

Ｐｒ［Ｚ ｊ∧Ｙ ｉ∧Ｘ］＝Ｐｒ［Ｚ ｊ∧Ｘ］ 式５

⇔ Ｐｒ［Ｚ ｊ∧Ｙ ｉ｜Ｘ］＝Ｐｒ［Ｚ ｊ｜Ｘ］ 式６

【００５７】

【数３】

$$\Leftrightarrow \quad \Pr[Y_i\ root \mid X] = \Pr[Y_i \mid X] - \sum_j \Pr[Z_j \mid X] \qquad \text{式 7}$$

【００５８】

となる。

【００５９】

好ましい一実装形態では、いくつもの代替のコンテキストツリーがキーワード検索操作の結果としてユーザへと、関連づけられている確率Ｐｒ［Ｙ ｉ　ｒｏｏｔ｜Ｘ］が０より大きいＸの祖先ノードＹ ｉごとに１つ、返される。これらの代替のコンテキストツリーは、それぞれ、関連づけられている確率Ｐｒ［Ｙ ｉ　ｒｏｏｔ｜Ｘ］が割り当てられ、このスコアに従って、最高から最低へとソートされている。スコアがより高いコンテキストツリーは、スコアがより低いものよりもユーザにとって興味があると考えられる。一代替実施形態では、スコアが最高のコンテキストツリーだけが、キーワード検索操作の結果としてユーザへと返される。

【００６０】

コンテキストツリーのルート・ノードの役割のできる（すなわち、そのＰｒ［Ｙ ｉ　ｒｏｏｔ｜Ｘ］＞０）祖先ノードＹ ｉそれぞれについて、第２のフェーズ、すなわちＹ ｉからのトップ・ダウン走査が、その子孫（ヒット・ノードＸを除く）のうちのどれがコンテキスト・ノードであるかを判定するために行われる。このフェーズ中に訪問した親ノードＰ ｊそれぞれについて、解析が、その子のうちのどれがコンテキスト・ノードであるかを判定するために行われる。コンテキスト・ノードであると判定された子ノードそれぞれについて、その子が順番にトップ・ダウン的に解析されて、そのうちのコンテキスト・ノードが識別される。

【００６１】

親ノードＰ ｊの解析には、図３Ａおよび３Ｂに示すように、２つの別々の想定状況がある。第１のものは、図３Ａに示す特別な場合であり、ここでは、Ｐ ｊはルート・ノードＹ

$_i$　３００５からヒット・ノードＸ　３０２０へのパスに沿って存在する。これには、Ｐ$_j$＝Ｙ$_i$の場合が含まれるが、Ｐ$_j$＝Ｘの場合は除外される。この想定状況では、Ｐ$_j$からＸへのパスに沿って存在する、Ｐ$_j$　３０１０の少なくとも１つの子ノード、すなわちこの場合では子ノードＣ$_i$　３０１５が、コンテキスト・ノードとして識別されなければならない。より一般的な、第２の想定状況では、図３に示すように残りの場合すべてを包摂しているが、親ノードＰ$_j$　３０３０は、ルート・ノードＹ$_i$からヒット・ノードＸ　３０３５への有向パスに沿っては存在せず、このため、Ｐ$_j$のどれかの子ノードをコンテキスト・ノードとして識別することは義務的ではない。この第２の想定状況を扱うためのアルゴリズムを最初に提示することにする。

【００６２】

所与のヒット・ノードＸ、および、コンテキストツリーのルート・ノードの役割をする特定のノードＹ$_i$に関して、親ノードＰ$_j$のある子ノードＣ$_k$をコンテキスト・ノードとして識別すべきかどうかを選ぶことは、一般に、Ｙ$_i$からＸへの有向パスに沿ったノードすべての存在のもとで、Ｃ$_k$が生起する確率の関数

Ｐｒ［Ｃ$_k$｜Ｘ∧．．．∧Ｙ$_i$　ｒｏｏｔ∧．．．∧Ｐ$_j$］

である。

【００６３】

この確率の評価または推定は、前にふれた生起および共起の頻度テーブルだけでは可能ではないため、何らかの形の単純化または近似が必要である。採用することが好ましいそのような単純化の１つは、上の確率の式の中のＹ$_i$からＣ$_k$までのノード以外のノードすべての影響を無視することであり、次の式

Ｐｒ［Ｃ$_k$｜Ｙ$_i$　ｒｏｏｔ∧．．．∧Ｐ$_j$］

ここで

【００６４】

【数４】

$$\Pr[C_k \mid Y_i\, root \wedge \cdots \wedge P_j] = \Pr[C_k \mid Y_i\, root \wedge P_j]$$

$$= \frac{\Pr[C_k \wedge Y_i\, root \wedge P_j]}{\Pr[Y_i\, root \wedge P_j]}$$

$$= \frac{\Pr[C_k \wedge Y_i\, root]}{\Pr[Y_i\, root \wedge P_j]} \qquad \text{式}8$$

【００６５】

Ｚ$_l$、ｌ＝１，．．．，ｐがＹ$_i$の親ノードを表すとすると、式８の右辺は、

【００６６】

【数５】

$$\frac{\Pr[C_k \wedge Y_i] - \sum_{l=1}^{p} \Pr[C_k \wedge Z_l]}{\Pr[P_j \wedge Y_i] - \sum_{l=1}^{p} \Pr[P_j \wedge Z_l]} \approx \frac{freq(Y_i, C_k) - \sum_{l=1}^{p} freq(Z_l, C_k)}{freq(Y_i, P_j) - \sum_{l=1}^{p} freq(Z_l, P_j)} \qquad \text{式}9$$

【００６７】

と展開することができる。

【００６８】

しかし、上の式は、個々の子ノードＣ$_k$それぞれを孤立して扱っているのにすぎない。Ｃ$_k$が（Ｙ$_i$　ｒｏｏｔのもとで）互いに独立ない限りは、それらの同時確率を考える必

要がある。しかし、これには、ノードからなる大量の組合せの同時生起頻度を格納する頻度テーブルを維持することが必要であるが、そのうちの多くは、ほとんど観測されず、そのため、その同時確率の信頼できる推定をその同時生起頻度から行うことは可能ではないはずである。他方、（Y_i ｒｏｏｔのもとで）C_kの間の独立を仮定すると、子ノードC_kが、その個々の（Y_i ｒｏｏｔのもとでの）生起確率が低い場合に、どれも選択されないなどの望ましくない結果がもたらされる。

【００６９】

C_kの間の独立の過程の望ましくない影響を避ける一方で、同時に、大量の同時生起頻度の値を維持する必要性を避けるために、図５のフローチャートで示すヒューリスティックな方法５０００を、子ノードをコンテキスト・ノードとして選択するのに使用することができる。方法５０００は、好ましくは、サーバ４０３０上の方法２０００内部のサブプログラムとして動作する。

【００７０】

方法５０００は、ステップ５００５で始まり、ここで、$Q_k = Pr[C_k | Y_i \text{ root} \wedge \ldots \wedge P_j]$で表される、ルート・ノード$Y_i$のもとでの子ノード$C_k$それぞれの生起確率の計算が、式８および式９を用いて行われる。次のステップ５０１０で、確率Q_kが子ノードC_kすべてにわたって総計されるが、その総和はＴで表される。方法５０００は継続され、ステップ５０１５で確率の値が最も大きいノードC_kがコンテキスト・ノードとして選択される。確率の値が同じで最大である子ノードが１つより多くある場合は、そのようなノードはすべてコンテキスト・ノードとして選択される。次いで、それまでにコンテキスト・ノードとして選択された子ノードすべての確率の総和Ｓの計算が、ステップ５０２０で行われる。次いで、実行は判断ステップ５０２５へと進み、このポイントで、子ノードC_kがすべてコンテキスト・ノードとして選択されている場合には、方法はステップ５０４０で終了する。しかし、まだコンテキスト・ノードとして選択されていない１つまたは複数の子ノードC_kがある場合は、方法５０００はもう１つの判断ステップ５０３０へと続く。ステップ５０３０で、$S \geqq T/2$かどうかを確かめるために検査が行われ、そうである場合、方法はここでもステップ５０４０で終了する。$S < T/2$である場合、実行はステップ５０３５へと進む。ここでは、現在まだコンテキスト・ノードとして選択されていない子ノードC_kのリストが検査されて、そのうちで確率が最大のものが識別される。それがコンテキスト・ノードとして選択され、方法５０００は、ステップ５０２０へと戻って、上で論じたような処理をさらに行う。

【００７１】

方法５０００には、いくつもの望ましい特徴がある。すなわち、

・ログ済みの問い合わせまたはデータ・ビューに共通部分があって頻度が十分に高い（すなわち、比較的多数の子ノードが共通にある）場合、方法５０００はその共通部分をコンテキスト・ノードとして返す傾向がある。これからは、十分に大きな共通部分は（ヒット・ノードに関する）十分なコンテキスト情報を担う傾向があるため、容認できる結果がもたらされそうである。

【００７２】

・ログ済みの問い合わせまたはデータ・ビューに、共通に比較的少数の子ノードしかなかった場合、結果となるコンテキスト・ノードの組は、その共通部分だけでなくさらに別のノードも含む傾向がある。発明者の行った実験では、結果となるコンテキストの子ノードの組は、頻度が最も高いログ済みの問い合わせまたはデータ・ビューのそれを反映する傾向がある。このことは、共通部分だけでは十分なコンテキスト情報を含むことはありそうにないであろうから、重要である。

【００７３】

・確率の値が等しく最大である子ノードをまとめて含めることが原因で、方法５０００は、ノードのコンテキスト・ノードとしての識別を、少なくではなく多く行う方へのバイアスをもつ。ログ済みの問い合わせの中にある子ノードの組が互いに排他的であり等しい頻度で生じる場合には、この方法では、子ノードすべてがコンテキスト・ノードとして識

別される。

【００７４】

方法５０００では、親ノードがコンテキスト・ノードと同じであるときには、その子のうちの１つまたは複数も、必ずコンテキスト・ノードとして識別される。このことは、親ノードが、まったく子がない状態で生じる（すなわち、葉ノードとして生じる）ログ済みの問い合わせまたはデータ・ビューが数多くある場合には、望ましくない。直観的には、これが十分しばしば生じる場合には、その親ノードだけを、まったく子がないコンテキスト・ノードとして識別して、頻繁に観察される振る舞いを反映させるべきである。

【００７５】

この問題点を改善するために、好ましい一実装形態では、インデックス・サーバ４０３０の生成および保存する追加の葉共起頻度テーブルを利用する。このテーブルでは、ノードP_jが、葉ノードとして過去のログ済みの問い合わせおよびデータ・ビュー内でその祖先Y_iと共起している頻度を、スキーマ・グラフ内に子のないノードP_jを除き、そのようなノードP_jとY_iのあらゆる可能な対に対して保存する。次いで、この新しい頻度テーブルを使用して、P_jおよびあるルート・ノードY_iのもとで、ノードP_jが葉ノードとして生起する確率

【００７６】

【数６】

$$\Pr[P_j\ leaf \mid Y_i\ root \wedge P_j] = \frac{\Pr[P_j\ leaf \wedge Y_i\ root]}{\Pr[Y_i\ root \wedge P_j]}$$

$$\approx \frac{freq(Y_i, P_j\ leaf) - \sum_{l=1}^{p} freq(Z_l, P_j\ leaf)}{freq(Y_i, P_j) - \sum_{l=1}^{p} freq(Z_l, P_j)} \qquad \text{式10}$$

【００７７】

が推定される。ここで、Z_l、ｌ＝１，．．．，ｐは、前に定義したようにY_iの親ノードを表し、ｆｒｅｑ（Y_i，P_j ｌｅａｆ）およびｆｒｅｑ（Z_j，P_j ｌｅａｆ）は、新しい葉共起頻度テーブルから得られる共起頻度の値である。

【００７８】

確率Ｐｒ［P_j ｌｅａｆ｜Y_i ｒｏｏｔ∧P_j］は、好ましくは、図５に与える方法５０００に先立つ追加の判断ステップで決定され、P_jのどの子ノードがコンテキスト・ノードであるかが識別される。Ｐｒ［P_j ｌｅａｆ｜Y_i ｒｏｏｔ∧P_j］が０．５より小さい場合は、P_jの子ノードはコンテキスト・ノードとしては選択されず、そうでない場合は、方法５０００が行われて、どの子ノードがコンテキスト・ノードであるかが識別される。

【００７９】

代替実施形態も可能であり、そこでは、フローチャートを図６に与えている代替の方法６０００を使用して、１組の子ノードC_k、ｋ＝１，．．．，ｍのうちでコンテキスト・ノードが選択される。方法６０００は、これもインデックス・サーバ４０３０で行われるが、ステップ６００１で始まり、そこでは、概念上、架空の子ノードC_0が作成され、実際の子ノードC_1，．．．，C_mからなるリストへと追加され、確率値Ｐｒ［P_j ｌｅａｆ｜Y_i ｒｏｏｔ∧P_j］が式１０を用いて割り当てられる。次のステップ６００５で、実際の子ノードC_kには、その通常の確率値Q_k＝Ｐｒ［C_k｜Y_i ｒｏｏｔ∧P_j］が式８および式９を用いて割り当てられる。次いで、方法６０００は継続され、ステップ６００６で、方法５０００をステップ５０１０で呼び出して（ステップ５００５を飛ばして）子ノードC_0，．．．，C_mのうちで１組のコンテキスト・ノードを選択する。方法５０００を抜けると、方法６０００は再開され、判断ステップ６０１０で検査が行わ

れて、架空の子ノードC_0がコンテキスト・ノードとして選択されたかどうかが判定される。そうである場合、実行は継続され、ステップ６０２０で、C_0がコンテキスト・ノードとして除外される。方法６０００は、その後、ステップ６０１５で終了する。検査が６０１０で失敗した場合、方法６０００は、直接に終了ステップ６０１５へと進む。

【００８０】

P_jの子ノードのうちのどれもコンテキスト・ノードではないという可能性を組み込むための代替の方法６０００の背後にある考え方は、第１の方法と本質的に同じである。すなわち、Ｐｒ［P_j ｌｅａｆ｜Y_i ｒｏｏｔ∧P_j］が十分に大きいときである。しかし、コンテキスト・ノードの最終的なセットに対するＰｒ［P_j ｌｅａｆ｜Y_i ｒｏｏｔ∧P_j］の影響は、この代替のアプローチでは、よりゆるやかであり、一般に、第１のアプローチでの急激なオン・オフの切り替えよりも好適である。

【００８１】

親ノードP_jがルート・ノードY_iからヒット・ノードＸへの有向パスに沿って存在する特別な想定状況では、P_jからＸへのパスに沿って存在するP_jの子ノードがコンテキスト・ノードとして識別されるように、特別な考慮がなされなければならない。一般性を失うことなく、この子ノードを、図３に項目３０１５として示すC_1とする。前に提示した一般的な想定状況用の方法５０００は、（たとえば、C_1の生起確率を他の子ノードすべてのそれを上回るように、ステップ５０１５の前に、吊り上げておくことによって）変更が可能であるが、そのようなアプローチは正しい結果を生じない可能性がある。その理由は、説明した方法５０００は、ルートY_iおよび親P_jのもとでコンテキスト・ノードとして最も頻繁に生起する子ノードの組を選択するように工夫されているためである。この組が自然にC_1を含まない場合、それは基本的に、C_1がその組の中のノードと関係がないことを意味している。C_1を無理に含めても、ただ、共通するところが少ししかない子ノードの組ができてしまい、C_1に対する（またその後Ｘに対する）コンテキストは少ししか得られないはずである。

【００８２】

方法５０００を変更する代わりに、好ましくは、異なるが手続き的にはいくぶん似ている、図７に示す方法７０００が、別の実装形態では採用される。新しい７０００と以前の５０００の間の違いは、使用される確率の独立性の仮定にある。P_jがY_iからＸへの有向パスに沿って存在しない、一般的な場合に行った第１の単純化が、

Ｐｒ［C_k｜Ｘ∧．．．∧Y_i ｒｏｏｔ∧．．．∧P_j］

は、Y_iからP_jまでのノード以外のノードとは独立であるという仮定であったことを想起されたい。P_jの１つの子ノードC_1がＸからP_jへのパスに沿って存在するという現在の想定状況では、C_kが（C_1を含めて）P_jからＸまでのノードとは独立であると仮定することは、これらがヒット・キーワードＸをC_kにリンクするＸの必要な祖先であることから、賢明ではあるまい。問題を扱いやすくしておくには単純化がいくつか必要であるため、C_kとそのP_jより上位のルート・ノードY_iに向かう祖先の間の確率の独立性を仮定することはよい選択肢ということになる。この仮定のもとで、興味の対象となる確率は、

Ｐｒ［C_k｜Ｘ∧．．．∧C_1∧P_j］　　$k \neq 1$

である。

【００８３】

再び、ＸとP_jをリンクする有向パスが多くとも１つあるため、上の式は

【００８４】

【数７】

$$\Pr[C_k \mid X \wedge P_j] = \frac{\Pr[C_k \wedge X \wedge P_j]}{\Pr[X \wedge P_j]} \qquad \text{式 11}$$

【００８５】

に等しい。

【００８６】
式１１の右辺の分子は、２つではなく３つのノードを含んでいる以上、これまでにふれた生起および共起の頻度テーブルからは得ることができない。したがって、ノードの３つ組（３－ｔｕｐｌｅ）の間の、追加の同時生起頻度テーブルが必要である。さいわい、こうした３つ組は、（ノードの任意のどの対もではなく）親子ノードC_kとP_jからなる対を含んでおり、各ノードC_kには、実際には、少数の親しかないため、新しい同時生起頻度テーブルは、ノードの対を含む共起頻度テーブルよりも、わずかに大きいのにすぎないはずである。
【００８７】
新しい同時生起頻度テーブルにより、$Pr[C_k \mid X \wedge P_j]$は
【００８８】
【数８】

$$\Pr[C_k \mid X \wedge P_j] \approx \frac{freq(C_k, P_j, X)}{freq(P_j, X)} \qquad \text{式12}$$

【００８９】
と推定することができ、ここで、ｆｒｅｑ（C_k，P_j，Ｘ）は、ノードC_k、P_j、およびＸの間の同時生起頻度を表し、P_jはC_kの親でＸの祖先であり、C_kはＸでもＸの祖先でもない。
【００９０】
C_1とともにコンテキスト・ノードとして含めるべきC_1の兄弟（ｓｉｂｌｉｎｇｓ）からなる組を決定するための方法７０００は、すでに説明した方法５０００に非常に似ている。方法７０００は、ステップ７００１で始まり、ここでは、$Q_k = Pr[C_k \mid X \wedge P_j]$で表される、親ノード$P_j$およびヒット・ノードＸのもとでの各子ノード$C_k \neq C_1$の生起確率の計算が式１２を用いて行われる。次のステップ７００５で、確率Q_kが子ノード$C_k \neq C_1$すべてにわたって総計されるが、その総和はＴで表される。方法７０００は継続され、ステップ７０１０で、ノードC_1がコンテキスト・ノードとして選択され、次いでその後ステップ７０１５で、確率の値が最大の$C_k \neq C_1$もコンテキスト・ノードとして選択される。確率の値が最大で同じ子ノードが１つより多く存在する場合は、そのようなノードはすべてコンテキスト・ノードとして選択される。次いで、C_1を除きそれまでにコンテキスト・ノードとして選択した子ノードすべての確率の総和の計算を、ステップ７０２０で行い、その総和はＳで表される。次いで、実行は、判断ステップ７０２５へと進み、このポイントで、子ノードC_kがすべてコンテキスト・ノードとして選択されている場合、方法７０００は、ステップ７０４０で終了する。しかし、１つまたは複数の子ノードC_kがまだコンテキスト・ノードとして選択されていない場合は、この方法は、別の判断ステップ７０３０へと続く。ステップ７０３０で、検査を行ってＳ≧Ｔ／２であるかどうかを確かめ、そうである場合は、方法７０００は、ここでも、ステップ７０４０で終了する。Ｓ＜Ｔ／２である場合は、実行はステップ７０３５へと進む。ここで、現在まだコンテキスト・ノードとして選択されていない子ノード$C_k \neq C_1$からなるリストが検査されて、それらのうちで確率の値が最大のものが識別される。これらは、コンテキスト・ノードとして選択され、方法は、処理をさらに行うためにステップ７０２０へと戻る。
【００９１】
C_1の兄弟が解に含まれていない場合を見越して、方法７０００にいくつかの変更が必要である。これは、ノードP_jが、過去のログ済みの問い合わせおよびデータ・ビューにP_jの子ノードが１つしかない（P_jからＸへのパスに沿うC_1）ようなその子孫Ｘのうちの１つと共起する頻度の格納されている単独子共起頻度テーブルを導入することによって達成される。次いで、この頻度テーブルを使用して、親P_jおよびヒット・ノードＸのもとでC_1に兄弟がない確率
【００９２】

【数９】

$$\Pr[C_1\ no\ sibling \mid P_j \wedge X] = \Pr[C_1 \wedge \neg C_k\ \forall k \neq 1 \mid P_j \wedge X]$$

$$= \Pr[P_j\ has\ 1\ child \mid P_j \wedge X]$$

$$= \frac{\Pr[P_j\ has\ 1\ child \wedge P_j \wedge X]}{\Pr[P_j \wedge X]}$$

$$= \frac{\Pr[P_j\ has\ 1\ child \wedge X]}{\Pr[P_j \wedge X]}$$

$$\approx \frac{freq(P_j\ has\ 1\ child, X)}{freq(P_j, X)} \qquad \text{式13}$$

【００９３】

が推定され、ここで、ｆｒｅｑ（Ｐ$_j$　ｈａｓ　１　ｃｈｉｌｄ，Ｘ）は、Ｐ$_j$がその子孫Ｘと共起し、Ｐ$_j$に単一の子ノード（Ｃ$_1$）がある頻度を表しており、新しい頻度テーブルから得られる。

【００９４】

ある実装形態では、確率Ｐｒ［Ｃ$_1$　ｎｏ　ｓｉｂｌｉｎｇ｜Ｐ$_j$∧Ｘ］を、図７に与えている方法７０００に先立つ追加の判断ステップで使用して、Ｐ$_j$のどの子ノードがコンテキスト・ノードであるかが識別される。Ｐｒ［Ｃ$_1$　ｎｏ　ｓｉｂｌｉｎｇ｜Ｐ$_j$∧Ｘ］が０．５より小さい場合は、Ｃ$_1$以外のＰ$_j$の子ノードはコンテキスト・ノードとして選択されず、そうでない場合は、方法７０００が行われてどの子ノードがコンテキスト・ノードであるかが識別される。

【００９５】

代替実施形態も可能である。ここでは、フローチャートを図８に与えている代替の方法８０００を使用して、１組の子ノードＣ$_k$、ｋ＝１，．．．，ｍのうちでコンテキスト・ノードが選択される。方法８０００は、ステップ８００１で始まり、そこでは、概念上、架空の子ノードＣ$_0$が作成され、実際の子ノードＣ$_1$，．．．，Ｃ$_m$からなるリストへと追加され、確率値Ｑ$_0$＝Ｐｒ［Ｃ$_1$　ｎｏ　ｓｉｂｌｉｎｇ｜Ｐ$_j$∧Ｘ］が式１３を用いて割り当てられる。ステップ８００５で、Ｃ$_1$を除く実際の子ノードＣ$_k$には、その通常の確率値Ｑ$_k$＝Ｐｒ［Ｃ$_k$｜Ｘ∧Ｐ$_j$］が式１１を用いて割り当てられる。次いで、方法８０００は継続され、ステップ８００６で、方法７０００をステップ７００５で呼び出して（ステップ７００１を飛ばして）子ノードＣ$_0$，．．．，Ｃ$_m$のうちで１組のコンテキスト・ノードを選択する。方法７０００を抜けると、方法８０００は再開され、判断ステップ８０１０で検査が行われて、架空の子ノードＣ$_0$がコンテキスト・ノードとして選択されたかどうかが判定される。そうである場合、実行は継続され、ステップ８０２０で、Ｃ$_0$がコンテキスト・ノードとして除外される。方法８０００は、その後、ステップ８０１５で終了する。検査が８０１０で失敗した場合、方法８０００は、直接に終了ステップ８０１５へと進む。

【００９６】

これまでの議論で、コンテキスト・ノードである１組の子ノードから決定するための２つの異なる方法６０００および８０００を説明している。好ましくは、後者は、親ノードＰ$_j$がルート・ノードＹ$_i$からヒット要素Ｘへの有向パスに沿って存在する想定状況で適用されるのに対し、前者は他の親ノードすべてに対して使用される。一代替実装形態では、第１の方法６０００は、Ｐ$_j$がＹ$_i$からＸへのパスに沿って存在する場合でも使用される。これからＣ$_1$を含む１組のコンテキストの子ノードができる場合は、その組が採用さ

れるが、そうでない場合は、その組は捨てられ、第２の方法８０００が、コンテキストの子ノードからなる新しい組を決定するために適用される。この第１の方法の有利さの背後にある理論的根拠は、そこで計算される確率の値が条件としているのが、ルート要素Y_iであり、ヒット・ノードＸではないことである。本発明者の行った試験は、データ・ビューのルート要素が、どのようなノードがそのビュー内にあるかのよりよい指標であることを示唆するように思われる。

【００９７】

本明細書で開示するキーワード検索システム４０００は、学習システムの一形式である。ログ済みの問い合わせおよび既存のデータ・ビューの組、これはトレーニングの例に似ているが、これからシステムは、データの新しいビューを合成することができる。ログ済みの問い合わせまたはデータ・ビューの中にパターンが存在する場合、そのパターンは頻度テーブルの中に反映され、これが今度はシステム４０００の振る舞いに影響することになる。どのような学習システムにとっても望ましい特徴は、その学習システムが、関係があるがまだ見たことのない問題を扱うときに、１組の問題から学習したパターンを使用してそのパフォーマンスを改善することができる何らかの形の一般化を行う能力である。階層的な環境で重要な一般化の１つの側面は、データの特定の下位構造の生起パターンを観察し、それを他の類似または同一の下位構造へと一般化する能力である。

【００９８】

図９に示すデータ構造９０００を考えるが、この中には、２つの同じ従業員（Ｅｍｐｌｏｙｅｅ）下位構造９０１０および９０３０が（点線の曲線で囲まれている）、１つは管理職（Ｍａｎａｇｅｒ）９００５の下に、もう１つはプロジェクト・メンバー（Ｐｒｏｊｅｃｔ　Ｍｅｍｂｅｒ）９０２５の下にある。ログ済みの問い合わせまたはデータ・ビューすべての中で、第１のＥｍｐｌｏｙｅｅサブツリー内の下位要素の名（ＦｉｒｓｔＮａｍｅ）９０１５および姓（ＬａｓｔＮａｍｅ）９０２０が、ともに現れることがすでに観察されているのに対して、第２の従業員（Ｅｍｐｌｏｙｅｅ）サブツリー９０３０を含む問い合わせまたはデータ・ビューはまだ何も観察されていないとする。さらに、従業員の名前に対するキーワード検索操作が呼び出されて、「ヒット」が第２の従業員（Ｅｍｐｌｏｙｅｅ）サブツリー９０３０の名（ＦｉｒｓｔＮａｍｅ）下位要素９０３５の中で見つかり、９０３５がヒット・ノードになっているとする。たとえ、この下位要素が存在する問い合わせまたはビューの例に出遭っていなくても、第１の従業員（Ｅｍｐｌｏｙｅｅ）サブツリー９０１０に関して観察された生起パターンから、第２の従業員（Ｅｍｐｌｏｙｅｅ）サブツリー９０３０内の下位要素の姓（ＬａｓｔＮａｍｅ）９０４０をコンテキスト・ノードとして識別すべきであることは、直観的に明らかである。

【００９９】

そのような一般化の能力は、ＸＭＬデータの作業を行うときには、同一のデータ下位構造が、しばしば、データ階層内のいくつかの場所に存在するため（たとえば、参照されるスキーマ要素を使用する結果として）、特に重要である。そのようなものは、確率平均化（ｐｒｏｂａｂｉｉｔｙ　ａｖｅｒａｇｉｎｇ）を通して実現することができる。確率平均化の動作は、スキーマ・グラフ内で名前またはＩＤまたはラベルが同一のノードの生起確率を適切に平均することによって行われる。ここで、確率平均化の適用の説明を、まずコンテキストツリーの構築の第１のトップ・ダウンのフェーズについて、そうしてその後で第２のボトム・アップのフェーズについて行う。

【０１００】

第１のフェーズの動作は、確率の値$Pr[Y_i|X]$を利用していることを想起されたい。ここで、Y_iはヒット・ノードＸの祖先である。確率平均化を容易にするために、$Pr[Y_i|X]$を、好ましくは、次のように漸化式（ｉｎｃｒｅｍｅｎｔａｌ　ｆｏｒｍ）へと定式化し直す。すなわち、Ｗを、Y_iからＸへの唯一の有向パスに沿って存在する、Y_iの子とする。すると、$Pr[Y_i|X]$を

【０１０１】

【数１０】

$$\Pr[Y_i \mid X] = \frac{\Pr[Y_i \wedge X]}{\Pr[X]}$$

$$= \frac{\Pr[Y_i \wedge W \wedge X]}{\Pr[X]} \qquad (\mathrm{Y_i}\text{から}\ \mathrm{X}\ \text{へのパスは}\ \mathrm{W}\ \text{を含む})$$

$$= \frac{\Pr[Y_i \mid W \wedge X]\Pr[W \wedge X]}{\Pr[X]}$$

$$= \Pr[Y_i \mid W \wedge X]\Pr[W \mid X] \qquad \text{式}14$$

【０１０２】

のように書き直すことができる。

【０１０３】

すなわち、Ｐｒ［Ｙ$_{i}$｜Ｘ］は、その子ノードＷの確率の値、すなわちＰｒ［Ｗ｜Ｘ］から漸進的に得ることができる。考え方は、手続きをヒット・ノードＸで始め、上の式を利用して、次々に上位の祖先ノードに対する確率の値を得ることである。各ステップで、そのときに確率平均化の方法が、式１４の右辺の第１項に適用される。たとえば、Ｐｒ’［Ｂ｜Ｘ］が、あるノードＢの確率平均化の結果としての変更後の確率の値を表すとすると、Ｐｒ’［Ｙ$_{i}$｜Ｘ］を、次の再帰式（ｒｅｃｕｒｓｉｖｅ　ｆｏｒｍｕｌａｅ）で定義することができる。すなわち

Ｐｒ’［Ｘ｜Ｘ］＝１　式１５

【０１０４】

【数１１】

$$\Pr'[Y_i \mid X] = \begin{cases} 0 & \textit{if } \Pr'[W \mid X] = 0 \\ \Pr_{mean}[Y_i \mid W \wedge X]\Pr'[W \mid X] & \textit{otherwise} \end{cases} \qquad \text{式}16$$

【０１０５】

ここで

【０１０６】

【数１２】

$$\Pr_{mean}[Y_i \mid W \wedge X] = \frac{\sum_k \Pr[Y_{ik} \mid W_k \wedge X_k]\Pr[W_k \wedge X_k]}{\sum_k \Pr[W_k \wedge X_k]}$$

$$= \frac{\sum_k \Pr[Y_{ik} \wedge W_k \wedge X_k]}{\sum_k \Pr[W_k \wedge X_k]}$$

$$= \frac{\sum_k \Pr[Y_{ik} \wedge X_k]}{\sum_k \Pr[W_k \wedge X_k]} \qquad (\mathrm{Y_{ik}}\text{から}\ \mathrm{X_k}\ \text{までのパスは}\ \mathrm{W_{ik}}\text{を含む})$$

$$\approx \frac{\sum_k freq(Y_{ik}, X_k)}{\sum_k freq(W_k, X_k)} \qquad \text{式}17$$

【０１０７】
であり、これが表すのは、（Y_i，Ｘ）に等しい（ｋのある値に対する）ノードの対（Y_{ik}，X_k）すべてにわたって、X_0およびY_{i0}（すなわち、ｋ＝０）をそれぞれＸおよびY_iの別名（ａｌｉａｓｅｓ）として計算した、ＷおよびＸのもとでのY_iの重み付け平均（ａｖｅｒａｇｅ　ｏｒ　ｍｅａｎ）確率である。これらの等しい対（Y_{ik}，X_k）のそれぞれに対して、総和の中の項W_kは、Y_{ik}からX_kへの有向パスに沿って存在する、Y_{ik}のすぐ下の（ｉｍｍｅｄｉａｔｅ）子を表す。
【０１０８】
ノード対（Y_{ik}，X_k）がノード対（Y_i，Ｘ）に等しいというのは、
（ｉ）Y_{ik}の名前またはラベルまたはＩＤがY_iと同じであり、X_kの名前、ラベル、またはＩＤがＸと同じである、
（ｉｉ）Y_{ik}とX_kの間に直接の祖先－子孫関係があり、Y_iとＸの間でも同様である、
（ｉｉｉ）Y_{ik}からX_kへの有向パスに沿うノードW_kそれぞれに対して、（W_k，Ｘ）が（Ｗ，Ｘ）に等しく、（Y_i，W_k）が（Y_i，Ｗ）に等しいようなY_iからＸへの有向パスに沿う対応するノードＷがなければならない、
（ｉｖ）Y_iとY_{ik}にちょうど同じ数の親があり、Y_iの親Z_jそれぞれに対して、（Z_{kj}，Y_{ik}）および（Z_j，Y_i）が上の条件（ｉ）から（ｉｉｉ）を満たすようなY_{ik}の親Z_{kj}がある
場合である。
【０１０９】
すると、確率平均化が原因で変更された、ＸのもとでY_iがルートである確率は、
【０１１０】
【数１３】

$$\Pr'[Y_i\ root \mid X] = \Pr'[Y_i \mid X] - \sum_j \Pr'[Z_j \mid X] \qquad \text{式18}$$

【０１１１】
ここで、式１６から、Y_iをZ_jで、ＷをY_iで置き換えることによって得られるように、

$$\Pr'[Z_j \mid X] = \Pr_{mean}[Z_j \mid Y_i \wedge X]\ \Pr'[Y_i \mid X] \qquad \text{式19}$$

である、
式１７の右辺の分母が０である場合、これはログ済みの問い合わせおよびデータ・ビューでノード対（W_k，X_k）のうちのどれも観察されていないことを示すが、式１７、したがって式１９および式１８は不定（ｕｎｄｅｆｉｎｅｄ）となり、したがって、コンテキスト・ノードを識別するための何らかの代替方法が必要になる。好ましいアプローチは、$\Pr_{mean}[Z_j \mid Y_i \wedge X]$の代替の定義を、次のように、$Z_j$のヒット・ノードＸからの距離によって行うことである。すなわち
【０１１２】

【数１４】

$$\Pr_{mean}[Z_j \mid Y_i \wedge X] =$$

$$\begin{cases} \dfrac{\sum_k freq(Z_{jk}, X_k)}{\sum_k freq(Y_{ik}, X_k)} & if \sum_k freq(Y_{ik}, X_k) \neq 0 \\ 1 & if \sum_k freq(Y_{ik}, X_k) = 0, \quad dist(Z_j, X) \leq d_{\max} \\ 0 & if \sum_k freq(Y_{ik}, X_k) = 0, \quad dist(Z_j, X) > d_{\max} \end{cases} \qquad \text{式 20}$$

【０１１３】

ここで、d_{max}はある閾値の定数であり、ｄｉｓｔ（Ａ，Ｂ）は、２つのノードＡとＢの間のスキーマ・グラフ内での距離であり、ＡとＢの間のパスに沿うリンクの数として定義される。関連性の高い過去のログ済みの問い合わせおよびデータ・ビューがない場合は、２つのノードの間の距離によって、それらが互いにどう関係しているかのよい表示が与えられるが、これは、実際には、関係のあるデータは、普通、互いに近くに保存されているためである。

【０１１４】

ヒット・ノードＸの祖先ノード$Ｙ_i$がコンテキストツリーのルート・ノードである確率の計算を、確率平均化によって、祖先ノード$Ｙ_i$すべてについて行うための方法１０００００のフローチャートを、図１０に示している。方法１００００は、ステップ１０００１で、$Ｙ_i$＝Ｘ、したがってＰｒ’［$Ｙ_i$｜Ｘ］＝１として始まる。次のステップ１０００５で、式１９および式２０を使用して、Ｐｒ’［$Ｚ_j$｜Ｘ］の計算が$Ｙ_i$の親ノード$Ｚ_j$それぞれに対して行われる。ステップ１０００５の後、ステップ１００１０では、Ｐｒ’［$Ｙ_i$　ｒｏｏｔ｜Ｘ］の計算が式１８に従って行われる。次いで、ステップ１００２５では、検査が、$Ｙ_i$の親ノードすべてが処理されているかどうかを判定するために行われる。そうでない場合、方法１００００は、ステップ１００１５へと進み、ここで、$Ｙ_i$の親ノード$Ｚ_j$が選択される。ステップ１００２０に到達すると、方法１００００は、ステップ１０００５で（ステップ１０００１を飛ばして）再帰的に、ただし選択した親ノード$Ｚ_j$が$Ｙ_i$の役割を果たす状態で呼び出される。この呼び出しが戻ると、方法１００００の現在の実行は再開され、ステップ１００２５へと戻って、親ノードがまだあるかの検査が行われる。親ノードがすべて処理されてしまうと、方法１００００は、ステップ１００３０で終わる。

【０１１５】

確率平均化は、第２のトップ・ダウン走査のフェーズにも適用される。このフェーズで、親ノード$Ｐ_j$がルート・ノード$Ｙ_i$からヒット・ノードＸへの有向パスに沿って存在しない一般の場合には、確率平均化を第１のフェーズで使用したのと同じ方法で適用することができる。キーワード検索の結果にコンテキスト・ノードとして含めるための$Ｙ_i$の子ノード$Ｃ_k$の選択は、確率

Ｐｒ［$Ｃ_k$｜$Ｙ_i$　ｒｏｏｔ∧$Ｐ_j$］

に基づく。確率平均化によって、上の式は、平均確率

【０１１６】

【数１５】

$$\Pr_{mean}[C_k \mid Y_i\ root \wedge P_j] = \frac{\sum_h \Pr[C_{kh} \wedge Y_{ih}\ root]}{\sum_h \Pr[Y_{ih}\ root \wedge P_{jh}]} \qquad \text{式21}$$

【０１１７】

によって置き換えられるが、ここで、（Ｙｉｈ，Ｃｋｈ）は（Ｙｉ，Ｃｋ）に等しく、（Ｐｊｈ，Ｃｋｈ）は（Ｐｊ，Ｃｋ）に等しく、Ｙｉ０、Ｃｋ０、およびＰｊ０（すなわち、ｈ＝０）はそれぞれ、Ｙｉ、Ｃｋ、およびＰｊの別名とする。ＺｊをＹｉの親、同様にＺｊｈをＹｉｈの対応する親とする。上の式を展開して

【０１１８】

【数１６】

$$\Pr_{mean}[C_k \mid Y_i\ root \wedge P_j] = \frac{\sum_h \left\{ \Pr[C_{kh} \wedge Y_{ih}] - \sum_j \Pr[C_{kh} \wedge Z_{jh}] \right\}}{\sum_h \left\{ \Pr[P_{jh} \wedge Y_{ih}] - \sum_j \Pr[P_{jh} \wedge Z_{jh}] \right\}}$$

$$\approx \frac{\sum_h \left\{ freq(Y_{ih}, C_{kh}) - \sum_j freq(Z_{jh}, C_{kh}) \right\}}{\sum_h \left\{ freq(Y_{ih}, P_{jh}) - \sum_j freq(Z_{jh}, P_{jh}) \right\}} \qquad \text{式22}$$

【０１１９】

にすることができる。

【０１２０】

上の式が平均確率Ｐｒｍｅａｎ［Ｃｋ｜Ｙｉ　ｒｏｏｔ∧Ｐｊ］の正確な近似であるためには、右辺の分母が十分に大きくなければならない（たとえば、＞ある正定数ｆｍｉｎ）。そうでないときには、好ましい一実装形態において採用された好ましい改善方法を使用して、まずＰｒｍｅａｎ［Ｃｋ｜Ｙｉ　ｒｏｏｔ∧Ｐｊ］が、Ｐｒｍｅａｎ［Ｃｋ｜Ｙｉ∧Ｐｊ］で近似され、ここで、確率は、Ｙｉ　ｒｏｏｔ∧ＰｊでなくＹｉ∧Ｐｊを条件としている。こうして

【０１２１】

【数１７】

$$\Pr_{mean}[C_k \mid Y_i\ root \wedge P_j] \approx \Pr_{mean}[C_k \mid Y_i \wedge P_j]$$

$$\approx \frac{\sum_h \Pr[C_{kh} \wedge Y_{ih}]}{\sum_h \Pr[Y_{ih} \wedge P_{jh}]}$$

$$\approx \frac{\sum_h freq(Y_{ih}, C_{kh})}{\sum_h freq(Y_{ih}, P_{jh})} \qquad \text{式23}$$

【０１２２】

式２３の右辺の分母がやはり十分に大きくない場合は、Ｐｒ$_{mean}$［Ｃ$_k$｜Ｙ$_i$∧Ｐ$_j$］はさらに、Ｙ$_i$ではなくＷを条件とする確率で近似され、ここで、ＷはＹ$_i$のすぐ下の子であり、Ｃ$_k$の祖先である。すなわち

【０１２３】

【数１８】

$$\Pr_{mean}[C_k \mid Y_i\ root \wedge P_j] \approx \Pr_{mean}[C_k \mid W \wedge P_j]$$

$$= \frac{\sum_h \Pr[C_{kh} \wedge W_h]}{\sum_h \Pr[W_h \wedge P_{jh}]}$$

$$\approx \frac{\sum_h freq(W_h, C_{kh})}{\sum_h freq(W_h, P_{jh})} \qquad \text{式 24}$$

【０１２４】

この方法は繰り返して、さらに、右辺の分母に十分に大きな値が得られるまで、そうでない場合は、ＷがＣ$_k$の親を表すまで行われる。後者の場合、Ｐｒ$_{mean}$［Ｃ$_k$｜Ｙ$_i$　ｒｏｏｔ∧Ｐ$_j$］への値の割り当ては、Ｃ$_k$とＹ$_i$の間の距離に基づいて

【０１２５】

【数１９】

$$\Pr_{mean}[C_k \mid Y_i\ root \wedge P_j] \approx \begin{cases} 1 & if\ dist(C_k, Y_i) \le d_{\max} \\ 0 & otherwise \end{cases} \qquad \text{式 25}$$

【０１２６】

または、Ｃ$_k$とヒット・ノードＸの間の距離に基づいて

【０１２７】

【数２０】

$$\Pr_{mean}[C_k \mid Y_i\ root] \approx \begin{cases} 1 & if\ dist(C_k, X) \le d_{\max} \\ 0 & otherwise \end{cases} \qquad \text{式 26}$$

【０１２８】

のように行われる。

【０１２９】

Ｐｒ$_{mean}$［Ｃ$_k$｜Ｙ$_i$　ｒｏｏｔ∧Ｐ$_j$］の近似が、最終的に式２２、式２３、式２４、式２５、または式２６で行われたのに応じて、ルート・ノードＹ$_i$のもとで、親ノードＰ$_j$にコンテキストの子ノードがない平均確率の計算は、それぞれ、式２７、式２８、式２９、式３０、または式３１を用いて行われる。

【０１３０】

【数２１】

$$\Pr_{mean}[P_j\ leaf \mid Y_i\ root \wedge P_j] = \frac{\sum_h \Pr[P_{jh}\ leaf \wedge Y_{ih}\ root]}{\sum_h \Pr[Y_{ih}\ root \wedge P_{jh}]}$$

$$\approx \frac{\sum_h \left\{ freq(Y_{ih}, P_{jh}\ leaf) - \sum_j freq(Z_{jh}, P_{jh}\ leaf) \right\}}{\sum_h \left\{ freq(Y_{ih}, P_{jh}) - \sum_j freq(Z_{jh}, P_{jh}) \right\}} \quad \text{式27}$$

【０１３１】

【数２２】

$$\Pr_{mean}[P_j\ leaf \mid Y_i\ root \wedge P_j] \approx \Pr_{mean}[P_j\ leaf \mid Y_i \wedge P_j]$$

$$= \frac{\sum_h \Pr[P_{jh}\ leaf \wedge Y_{ih}]}{\sum_h \Pr[Y_{ih} \wedge P_{jh}]}$$

$$\approx \frac{\sum_h freq(Y_{ih}, P_{jh}\ leaf)}{\sum_h freq(Y_{ih}, P_{jh})} \quad \text{式28}$$

【０１３２】

【数２３】

$$\Pr_{mean}[P_j\ leaf \mid Y_i\ root \wedge P_j] \approx \Pr_{mean}[P_j\ leaf \mid W \wedge P_j]$$

$$= \frac{\sum_h \Pr[P_{jh}\ leaf \wedge W_h]}{\sum_h \Pr[W_h \wedge P_{jh}]}$$

$$\approx \frac{\sum_h freq(W_h, P_{jh}\ leaf)}{\sum_h freq(W_h, P_{jh})} \quad \text{式29}$$

【０１３３】

【数２４】

$$\Pr_{mean}[P_j\ leaf \mid Y_i\ root \wedge P_j] \approx \begin{cases} 0 & if\ dist(P, Y_i) + 1 \le d_{\max} \\ 1 & otherwise \end{cases} \quad \text{式30}$$

【０１３４】

【数２５】

$$\Pr_{mean}[P_j\ leaf \mid Y_i\ root \wedge P_j] \approx \begin{cases} 0 & if\ dist(P, X) + 1 \le d_{\max} \\ 1 & otherwise \end{cases} \quad \text{式31}$$

【０１３５】

ルート要素Y_iのもとでの親ノードP_jに対するコンテキストの子ノードの決定を、P_jがY_iからヒット・ノードXへの有向パスに沿って存在しない一般の場合について、確率平均化によって行うための好ましい手続きは、図６に示すものに類似しており、これを図１３に示している。方法１３０００は、ステップ１３００１で始まり、そこでは、概念上、架空の子ノードC_0が作成され、実際の子ノードC_1，．．．，C_mからなるリストへと追加され、確率値$Q_0 = Pr_{mean}[P_j \ leaf | Y_i \ root \wedge P_j]$が、式２７、式２８、式２９、式３０、または式３１を用いて計算されて割り当てられ、次のステップ１３００５で、実際の子ノードC_kには、それに対応して、確率値$Q_k = Pr_{mean}[C_k | Y_i \ root \wedge P_j]$が、それぞれ、式２２、式２３、式２４、式２５、または式２６を用いて計算されて割り当てられる。どの場合でも、ステップ１３００６がステップ１３００５に続き、方法５０００をステップ５０１０で呼び出して（ステップ５００５を飛ばして）、子ノードC_0，．．．，C_mのうちで１組のコンテキスト・ノードを選択する。方法５０００を抜けると、方法１３０００は再開され、判断ステップ１３０１０で検査が行われて、架空の子ノードC_0がコンテキスト・ノードとして選択されたかどうかが判定される。そうである場合、実行は継続され、ステップ１３０２０で、C_0がコンテキスト・ノードとして除外される。方法１３０００は、その後、ステップ１３０１５で終了する。検査が１３０１０で失敗した場合、この方法は、直接に終了ステップ１３０１５へと進む。

【０１３６】

方法１３０００および６０００は、そのキーワード検索での用途とは別に、階層的なデータの選択的な提示の手段としても使用することができる。すでに論じたように、実際の階層的なデータソースは、通常、ユーザが任意の所与の時点で見たいと希望するよりもずっと多くのデータを含んでいる。ユーザが、階層的なデータソースの閲覧を、そのデータ構造内部のノードを選択することによって行うとき、通常、提示アプリケーションは、選択したノードの下のサブツリーの中のデータ項目すべてを表示し、そのうちの一部は、しばしば、ユーザにとって興味のないものである可能性がある。その提示アプリケーションが、興味のないデータのふるい除けを、ユーザの何らかのそれまでに観察された閲覧パターンに基づいて行えるのであれば、非常に望ましいはずである。説明した方法１３０００および６０００は、この作業に非常に適している。Y_i＝ユーザが閲覧のために選択したルート・ノードと設定することにより、その方法によって識別されたコンテキスト・ノードの組は、ユーザにとって興味があり、表示することが好ましいそうなノードを構成し、一方、コンテキスト・ノードとして識別されていない残りのノードは、好ましくはふるい除けられる。

【０１３７】

親ノードP_jがルート・ノードY_iからヒット・ノードXへの有向パス上に存在する特殊な場合に関しては、コンテキスト・ノードとして含めるためのY_iの子ノードC_kの選択が、確率

【０１３８】

【数２６】

$$\Pr[C_k \mid X \wedge P_j] = \frac{\Pr[C_k \wedge X \wedge P_j]}{\Pr[X \wedge P_j]} \qquad \text{式32}$$

【０１３９】

に基づくことを想起されたい。確率平均化によって、これらは、平均確率

【０１４０】

【数２７】

$$\Pr_{mean}[C_k \mid X \wedge P_j] = \frac{\sum_h \Pr[C_{kh} \wedge X_h \wedge P_{jh}]}{\sum_h \Pr[X_h \wedge P_{jh}]}$$

$$\approx \frac{\sum_h freq(C_{kh}, P_{jh}, X_h)}{\sum_h freq(P_{jh}, X_h)} \qquad \text{式33}$$

【０１４１】

によって置き換えられるが、ここで、（Ｐ$_{jh}$，Ｃ$_{kh}$）は（Ｐ$_{j}$，Ｃ$_{k}$）に等しく、（Ｐ$_{jh}$，Ｘ$_{h}$）は（Ｐ$_{j}$，Ｘ）に等しく、Ｐ$_{j0}$、Ｃ$_{k0}$、およびＸ$_{0}$（すなわち、ｈ＝０）はそれぞれ、Ｐ$_{j}$、Ｃ$_{k}$、およびＸの別名とする。

【０１４２】

上の式が平均確率Ｐｒ$_{mean}$［Ｃ$_{k}$｜Ｘ∧Ｐ$_{j}$］の正確な近似であるためには、式３３の右辺の分母が十分に大きくなければならない（たとえば、＞ｆ$_{min}$）。そうでないときには、使用することのできる別の好ましい改善方法は、Ｐｒ$_{mean}$［Ｃ$_{k}$｜Ｘ∧Ｐ$_{j}$］の近似を、Ｐｒ$_{mean}$［Ｃ$_{k}$｜Ｘ’∧Ｐ$_{j}$］、すなわち、Ｘではなく Ｘ’に対する条件付き確率で行うことであり、ここでＸ’はＹ$_{i}$からＸへの有向パス上に存在するＸのすぐ上の親である。

【０１４３】

Ｐｒ$_{mean}$［Ｃ$_{k}$｜Ｘ∧Ｐ$_{j}$］に対する近似の決定に使用するノードＸ’を識別するための方法２２０００のフローチャートを、図２２に示している。方法２２０００は、ステップ２２００５で始まり、ここで、Ｘ’はまずＸに初期化される。次のステップ２２０１０で、総和

【０１４４】

【数２８】

$$\sum_h freq(P_{jh}, X'_h)$$

【０１４５】

が計算され、Ｄに割り当てられるが、ここでノード対（Ｐ$_{jh}$，Ｘ’$_{h}$）は（Ｐ$_{j}$，Ｘ’）に等しい。次いで、判断ステップ２２０１５が続き、Ｄがある正の閾値定数ｆ$_{min}$以上であるかどうかの検査が行われる。そうである場合は、方法２２０００は、ステップ２２０２５で抜けて成功となる。判断ステップ２２０１５が失敗した場合は、実行は別の判断ステップ２２０３０へと進み、ここで検査が行われて、Ｘ’がＰ$_{j}$のすぐ下の子であるかどうかが判定される。そうである場合は、この方法はステップ２２０３５で抜けて失敗となり、そうでない場合は、方法は継続して、ステップ２２０４０で、Ｘ’が、Ｐ$_{j}$からＸへの有向パスに沿って存在するその親で置き換えられる。次いで、方法２２０００は、ループしてステップ２２０１０へと戻る。

【０１４６】

方法２２０００が成功して、ノードＸ’および対応する値Ｄが得られた場合、Ｐｒ$_{mean}$［Ｃ$_{k}$｜Ｘ∧Ｐ$_{j}$］には、値

【０１４７】

【数２９】

$$\Pr_{mean}[C_k \mid X \wedge P_j] \approx \frac{\sum_h freq(C_{kh}, P_{jh}, X'_h)}{D} \qquad \text{式34}$$

【０１４８】
が割り当てられる。
【０１４９】
方法２２０００を抜けて失敗になった場合には、Ｐｒｍｅａｎ［Ｃｋ｜Ｘ∧Ｐｊ］への値の割り当ては、ＣｋとＹｉの間の距離に基づいて
【０１５０】
【数３０】

$$\Pr_{mean}[C_k \mid X \wedge P_j] \approx \begin{cases} 1 & if\ dist(C_k, Y_i) \le d_{\max} \\ 0 & otheriwse \end{cases} \qquad 式35$$

【０１５１】
または、ＣｋとヒットＸの間の距離に基づいて
【０１５２】
【数３１】

$$\Pr_{mean}[C_k \mid X \wedge P_j] \approx \begin{cases} 1 & if\ dist(C_k, X) \le d_{\max} \\ 0 & otheriwse \end{cases} \qquad 式36$$

【０１５３】
行われる。
【０１５４】
Ｐｒｍｅａｎ［Ｃｋ｜Ｘ∧Ｐｊ］の近似が、最終的に式３４、式３５、または式３６で行われたのに応じて、親ノードＰｊに、コンテキストの子ノードが、ＰｊからＸへの有向パス上に存在する子ノードＣ１以外ないことの、Ｐｊおよびヒット・ノードＸのもとでの平均確率の計算は、それぞれ、式３７、式３８、または式３９を用いて行われる。すなわち
【０１５５】
【数３２】

$$\Pr_{mean}[C_1\ no\ sibling \mid P_j \wedge X] \approx \frac{\sum_h freq(P_{jh}\ has\ 1\ child, X'_h)}{D} \qquad 式37$$

【０１５６】
【数３３】

$$\Pr_{mean}[C_1\ no\ sibling \mid P_j \wedge X] \approx \begin{cases} 0 & if\ dist(P, Y_i)+1 \le d_{\max} \\ 1 & otherwise \end{cases} \qquad 式38$$

【０１５７】
【数３４】

$$\Pr_{mean}[C_1\ no\ sibling \mid P_j \wedge X] \approx \begin{cases} 0 & if\ dist(P, X)+1 \le d_{\max} \\ 1 & otherwise \end{cases} \qquad 式39$$

【０１５８】
ここで、（Ｐｊｈ，Ｘ’ｈ）は（Ｐｊ，Ｘ’）に等しく、Ｘ’およびＤは方法２２００によって得られる。
【０１５９】
親ノードＰｊに対するコンテキストの子ノードの決定を、ＰｊがＹｉからヒット・ノードＸへの有向パスに沿って存在する特別な場合について、確率平均化によって行うための好ましい手続きは、図８に示すものに非常に類似しており、これを図１４に示している。
【０１６０】

図１４に示す方法１４０００は、ステップ１４００１で始まり、そこでは、概念上、架空の子ノードC_0が作成され、実際の子ノードC_1，．．．，C_mからなるリストへと追加され、確率値$Q_0 = Pr_{mean}[C_1 \text{ no sibling} | P_j \wedge X]$が、式３７、式３８、または式３９を用いて計算されて割り当てられ、次のステップ１４００５で、C_1を除く実際の子ノードC_kには、それに対応して、確率値$Q_k = Pr_{mean}[C_k | X \wedge P_j]$が、それぞれ、式３４、式３５、または式３６を用いて計算されて割り当てられる。どの場合でも、ステップ１４００６がステップ１４００５に続き、方法７０００をステップ７００５で呼び出して（ステップ７００１を飛ばして）、子ノードC_0，．．．，C_mのうちで１組のコンテキスト・ノードを選択する。方法７０００を抜けると、方法１４０００は再開され、判断ステップ１４０１０で検査が行われて、架空の子ノードC_0がコンテキスト・ノードとして選択されたかどうかが判定される。そうである場合、実行は継続され、ステップ１４０２０で、C_0がコンテキスト・ノードとして除外される。方法１４０００は、その後、ステップ１４０１５で終了する。検査が１４０１０で失敗した場合、この方法は、直接に終了ステップ１４０１５へと進む。

【０１６１】

前述の議論では、スキーマ・グラフ内に多くとも単一のヒット・ノードがある特別な場合においてコンテキスト・ノードを識別するための方法を説明している。これは、ユーザが単一の検索キーワードだけを入力するときの普通の想定状況である。そのキーワードがスキーマ・グラフ内の複数の場所に現れる場合には、これは１つより多くのヒットがあることを意味するが、ヒットそれぞれを別々に扱うことが好ましい。すなわち、説明した方法がスキーマ・グラフ内の第１のヒット・ノードに対して適用され、複数のコンテキストツリーがそのヒット・ノードに対して決定される。次いでその後で、同じ方法が残りのヒット・ノードのそれぞれに対して適用されて、新しい複数のコンテキストツリーが得られる、などである。ヒット・ノードすべてが処理されてしまうと、生成されたコンテキストツリーは、複数のヒット・ノードを包摂するとわかった場合には、スコアが付け直され、さらに重複しているコンテキストツリーは取り除かれる。次いで、残りのコンテキストツリーのリストは、（もしあれば）その新しいスコアに従って整列し直され、キーワード検索操作の結果としてユーザへと返される。

【０１６２】

しかし、ユーザが、ブール値のＡＮＤ操作で結合された複数の検索キーワードを含む「すべて見つける」（ｆｉｎｄ　ａｌｌ）キーワード検索操作を開始した場合は、キーワード・ヒットは、潜在的に、スキーマ・グラフ内の２つ以上のヒット・ノードに現れる可能性がある。コンテキストツリーを決定するためのより一般的な方法の説明を、ここで、そのような想定状況を扱うために行う。

【０１６３】

図１１に、内部に複数のヒット・ノード１１０１０、１１０２０、および１１０２５のある、スキーマ・グラフの例１１０００を示している。これらのヒット・ノードをX_1，．．．，X_nで表すことにする。当然、コンテキストツリーがヒット・ノードすべてを含むためには、Ａで表している（ノード１１００５）、ヒット・ノードすべてを含む最も小さいサブツリーのルート・ノードが、コンテキスト・ノードとして、Ａからヒット・ノードそれぞれへの有向パスに沿って存在するノードすべてとともに返されなければならない。したがって、ノード１１０１５は、ＡからX_2へ（またＡからX_3へ）の有向パスに沿って存在するため、コンテキスト・ノードでなければならない。

【０１６４】

コンテキストツリー決定の方法の第１の、ボトム・アップのフェーズは、ノードＡで始まり、走査は上向きに行われる。Y_iをＡまたはＡの祖先とするが、ヒット・ノードのもとでのそれの確率、すなわち$Pr[Y_i | X_1 \wedge . . . \wedge X_n]$を評価することが、コンテキストツリーの可能なルート・ノードを決定するためには必要である。これを数学的に表現すると

【０１６５】

【数３５】

$$\Pr[Y_i \mid X_1 \wedge \cdots \wedge X_n] = \frac{\Pr[Y_i \wedge X_1 \wedge \cdots \wedge X_n]}{\Pr[X_1 \wedge \cdots \wedge X_n]} \qquad \text{式40}$$

【０１６６】

となる。

【０１６７】

この時点で、何らかの確率の独立性の仮定が必要であるが、理由は、式４０の右辺の分子と分母はどちらも、ｎの一般の値について既存の頻度テーブルから直接に求めまたは推定することが（ｎ≦２のときの分母を除いて）できないためである。妥当な１つの仮定は、X_1の組が、共通の祖先Y_iのもとで、互いに独立というものである。言い換えれば、

$$\Pr[X_1 \wedge \ldots \wedge X_n \mid Y_i] = \Pr[X_1 \mid Y_i] \ldots \Pr[X_n \mid Y_i] \qquad \text{式41}$$

したがって、

【０１６８】

【数３６】

$$\begin{aligned}\Pr[Y_i \wedge X_1 \wedge \cdots \wedge X_n] &= \Pr[X_1 \wedge \cdots \wedge X_n \mid Y_i]\Pr[Y_i] \\ &= \Pr[X_1 \mid Y_i] \cdots \Pr[X_n \mid Y_i]\Pr[Y_i] \\ &= \frac{\Pr[X_1 \wedge Y_i] \cdots \Pr[X_n \wedge Y_i]}{\Pr[Y_i]^{n-1}}\end{aligned} \qquad \text{式42}$$

【０１６９】

$\Pr[Y_i] = 0$のときの特異性を取り除くために、$\Pr[Y_i \wedge X_1 \wedge \ldots \wedge X_n]$は

【０１７０】

【数３７】

$$\Pr[Y_i \wedge X_1 \wedge \cdots \wedge X_n] = \begin{cases} 0 & \text{if } \Pr[Y_i] = 0 \\ \dfrac{\Pr[X_1 \wedge Y_i] \cdots \Pr[X_n \wedge Y_i]}{\Pr[Y_i]^{n-1}} & \text{otherwise} \end{cases} \qquad \text{式43}$$

【０１７１】

同様に

【０１７２】

【数３８】

$$\begin{aligned}\Pr[X_1 \wedge \cdots \wedge X_n] &= \Pr[A \wedge X_1 \wedge \cdots \wedge X_n] \\ &= \Pr[X_1 \wedge \cdots \wedge X_n \mid A]\Pr[A] \\ &= \Pr[X_1 \mid A] \cdots \Pr[X_n \mid A]\Pr[A] \\ &= \frac{\Pr[X_1 \wedge A] \cdots \Pr[X_n \wedge A]}{\Pr[A]^{n-1}}\end{aligned} \qquad \text{式44}$$

【０１７３】

単一のヒット・ノードしかない場合のように、ヒット・ノードすべてのもとでのY_iの生起確率の表現は、好ましくは、そのすぐ下の子ノードの確率によって漸進的に、確率平

均化が容易になるように、次のように行われる。すなわち、Wが、Y_iからAへの有向パスに沿うY_iのすぐ下の子ノードを表すとすると、

【０１７４】

【数３９】

$$\Pr'[Y_i \mid X_1 \wedge \cdots \wedge X_n] = \begin{cases} 1 & Y_i = A \\ 0 & Y_i \neq A, \Pr'[W \mid X_1 \wedge \cdots \wedge X_n] = 0 \\ \Pr_{mean}[Y_i \mid W \wedge X_1 \wedge \cdots \wedge X_n] \bullet \Pr'[W \mid X_1 \wedge \cdots \wedge X_n] & otherwise \end{cases} \quad \text{式45}$$

【０１７５】

ここで

【０１７６】

【数４０】

$$\Pr_{mean}[Y_i \mid W \wedge X_1 \wedge \cdots \wedge X_n] = \frac{\sum_h \Pr[Y_{ih} \wedge W_h \wedge X_{1h} \wedge \cdots \wedge X_{nh}]}{\sum_h \Pr[W_h \wedge X_{1h} \wedge \cdots \wedge X_{nh}]} \quad \text{式46}$$

【０１７７】

ここで、対（Y_{ih}，W_h）は（Y_i，W）に等しく、l＝１，．．．，nに対して（W_h，X_{lh}）は（W，X_l）に等しく、Y_{i0}、W_0、およびX_{l0}（h＝０）はそれぞれ、Y_i、W、およびX_lの別名とする。分子の総和の内側の項は、式４３で置き換えることができる。分母の総和の内側の項も、WにY_iの役割をさせることにより、式４３で置き換えることができ、こうして

【０１７８】

【数４１】

$$\Pr_{mean}[Y_i \mid W \wedge X_1 \wedge \cdots \wedge X_n] = \sum_h N_h \Big/ \sum_h D_h \quad \text{式47}$$

【０１７９】

が得られ、ここで

【０１８０】

【数４２】

$$N_h = \begin{cases} 0 & if\ \Pr[Y_{ih}] = 0 \\ \dfrac{\Pr[Y_{ih} \wedge X_{1h}] \cdots \Pr[Y_{ih} \wedge X_{nh}]}{\Pr[Y_{ih}]^{n-1}} & otherwise \end{cases}$$

$$\approx \begin{cases} 0 & if\ freq(Y_{ih}) = 0 \\ \dfrac{freq(Y_{ih}, X_{1h}) \cdots freq(Y_{ih}, X_{nh})}{freq(Y_{ih})^{n-1}} & otherwise \end{cases} \quad \text{式48}$$

【０１８１】

【数４３】

$$D_h = \begin{cases} 0 & if\ \Pr[W_h] = 0 \\ \dfrac{\Pr[W_h \wedge X_{1h}] \cdots \Pr[W_h \wedge X_{nh}]}{\Pr[W_h]^{n-1}} & otherwise \end{cases}$$

$$\approx \begin{cases} 0 & if\ freq(W_h) = 0 \\ \dfrac{freq(W_h, X_{1h}) \cdots freq(W_h, X_{nh})}{freq(W_h)^{n-1}} & otherwise \end{cases} \qquad \text{式 49}$$

【０１８２】

$\Pr_{mean}[Y_i \mid W \wedge X_1 \wedge \ldots \wedge X_n]$は、

【０１８３】

【数４４】

$$\sum_h D_h = 0.$$

【０１８４】

の場合には不定である。これが起きるとき、$\Pr_{mean}[Y_i \mid W \wedge X_1 \wedge \ldots \wedge X_n]$への値の割り当ては、好ましくは、$Y_i$からヒット・ノード$X_1, \ldots, X_n$への距離に基づいて次のように行われる。すなわち

【０１８５】

【数４５】

$$\Pr_{mean}[Y_i \mid W \wedge X_1 \wedge \cdots \wedge X_n] = \begin{cases} 1 & if \min_{i=1,\ldots,n} dist(Y_i, X_i) \le d_{\max} \\ 0 & otherwise \end{cases} \qquad \text{式 50}$$

【０１８６】

ノードY_iが、ヒット・ノードすべてを含むコンテキストツリーのルート・ノードである確率の計算を、Y_iとして選んだものすべてについて行うための方法１２０００のフローチャートを、図１２に示している。方法１２０００は、ステップ１２００１で始まり、ここでは、ヒット・ノードすべてを含む、スキーマ・グラフ内の最小のサブツリーのルート・ノードが識別され、Ａで表される。次いで、実行は継続して、ステップ１２００２で、Y_iがＡに初期化され、したがって$\Pr'[Y_i \mid X_1 \wedge \ldots \wedge X_n] = 1$となる。次のステップ１２００５で、式４５、式４７を、式４８および式４９、またはその代わりに式５０とともに使用して、$\Pr'[Z_j \mid X_1 \wedge \ldots \wedge X_n]$の計算が、$Y_i$の親ノード$Z_j$それぞれに対して行われる。ステップ１２００５の後、ステップ１２０１０では、$\Pr'[Y_i \ root \mid X_1 \wedge \ldots \wedge X_n]$の計算が、式

【０１８７】

【数４６】

$$\Pr'[Y_i\ root \mid X_1 \wedge \cdots \wedge X_n] = \Pr'[Y_i \mid X_1 \wedge \cdots \wedge X_n] - \sum_j \Pr'[Z_j \mid X_1 \wedge \cdots \wedge X_n] \qquad \text{式 51}$$

【０１８８】

に従って行われる。

【０１８９】

次いで、方法１２０００はステップ１２０１５へと進み、ここで、Y_iの親ノードZ_jが選択される。ステップ１２０２０に到達すると、方法１２０００は、ステップ１２００５で（ステップ１２００１および１２００２を飛ばして）再帰的に、ただし選択した親ノードZ_jがY_iの役割を果たす状態で呼び出される。この再帰呼び出しが戻ると、実行が判断ステップ１２０２５で再開され、ここで検査が行われて、Y_iの親ノードすべてが処理されているかどうかの判定が行われる。そうである場合は、方法はステップ１２０３０で終わり、そうでない場合は、方法は継続して、ステップ１２０１５でY_iの別の親ノードZ_jが、処理のために選択される。

【０１９０】

第２のトップ・ダウン走査のフェーズで、親ノードP_jがルート・ノードY_iからどのヒット・ノードへの有向パスに沿っても存在しない一般の場合には、P_jの子ノードがコンテキスト・ノードであるかどうかを判定するための方法は、これまでに単一のヒット・ノードしかない場合のために使用した方法１３０００とそのまま変わらない。

【０１９１】

親ノードP_jがY_iから１つまたは複数のヒット・ノードへの有向パスに沿って存在する特別な場合には、単一のヒット・ノードの場合のために使用した方法を変更して、１つより多くのP_jの子ノードをコンテキスト・ノードとして含めなければならない可能性を見越しておく必要がある。１つのヒット・ノードXだけしか含まない場合には、子ノードC_kがコンテキスト・ノードであるかどうかの判定は、確率の値

$$Pr[C_k \mid X \wedge P_j]$$

に基づいて行われることを想起されたい。ここで、XはP_jの子孫であるが、C_kまたはC_kの子孫ではない。１つより多くのヒット・ノードがあるときの拡張は、子ノードC_k、$k=1,\ldots,m$の選択プロセスを、親ノードP_jおよびP_jの子孫のであるヒット・ノードX_lすべてのもとでのC_kの確率に基づくようにし、一方でP_jの子孫ではないヒット・ノードの影響を無視することである。当然、それ自身がヒット・ノードであり、または１つまたは複数のヒット・ノードの祖先である子ノードC_kすべては、コンテキスト・ノードでなければならない。一般性を失うことなく、これらの子ノードを、$C_1,\ldots,C_r$とし、ここで$1 \leqq r \leqq m$とする。同様に、$C_1,\ldots,C_r$の子孫であるヒット・ノードの組を$X_1,\ldots,X_s$とし、ここで$r \leqq s \leqq n$とする。$s=1$（したがって$r=1$）の場合、この想定状況は、単一のヒット・ノードしかない場合に等しく、したがって、この場合に関して説明した方法１４０００を使用することができる。$s>1$の場合に関して一般化するための好ましい実装形態で採用する方法は、項$Pr[C_k \mid X \wedge P_j]$を、式

【０１９２】

【数４７】

$$\sum_{l=1}^{s} \Pr[C_k \mid X_l \wedge P_j]$$

【０１９３】

で置き換えることであり、これは、確率平均化の後、

【０１９４】

【数４８】

$$Q_k = \sum_{l=1}^{s} \Pr_{mean}[C_k \mid X_l \wedge P_j] \qquad \text{式 52}$$

【０１９５】

となり、ここで$Pr_{mean}[C_k \mid X \wedge P_j]$は、式３３で定義した通りである。$Q_k$は、任意の$X_l$、$l=1,\ldots,s$に対して$freq_{mean}(P_j, X_l)=0$の場合には、不定となり、ここで

【０１９６】

【数４９】

$$freq_{mean}(P_j, X_l) = \sum_h freq(P_{jh}, X_{lh})$$

【０１９７】

であり、（Ｐ$_{jh}$，Ｘ$_{lh}$）は（Ｐ$_{j}$，Ｘ$_{l}$）に等しい。ｆｒｅｑ$_{mean}$（Ｐ$_{j}$，Ｘ$_{l}$）が非零ではあるが小さな数である（たとえば、＜ｆ$_{min}$）のときでも、Ｑ$_{k}$の推定は、頻度テーブルからは十分な精度で行うことができない。単一のヒット・ノードを含む場合のように、この問題の克服は、ヒット・ノードＸ$_{1}$，．．．，Ｘ$_{s}$を、そのヒット・ノードの一部またはすべてをその祖先で置き換えた新しいノードの組Ｓで置き換え、その後でＱ$_{k}$をＳの要素によって定義し直すことによって行うことができる。

【０１９８】

図２１のフローチャートで示している方法２１０００が、好ましくは、ヒット・ノードＸ$_{1}$，．．．，Ｘ$_{s}$を置き換えるこのノードの新しい組Ｓを決定するのに使用される。方法２１０００は、ステップ２１００５で始まり、ここで、ヒット・ノードの初期の組Ｘ$_{1}$，．．．，Ｘ$_{s}$はＳで表される。次のステップ２１０１０で、Ｓのうちの処理済みでない要素Ｘ$_{p}$が選択される。次いで、判断ステップ２１０１５が続き、ここでは検査が行われて、ｆｒｅｑ$_{mean}$（Ｐ$_{j}$，Ｘ$_{p}$）が、ある閾値定数ｆ$_{min}$以上であるかが判定される。そうである場合、Ｘ$_{p}$は組Ｓの中に保持され、方法２１０００は判断ステップ２１０２０へと続き、ここで検査が行われてＳの中の要素すべてが処理されているかが判定される。１つまたは複数の処理済みでない要素が残っている場合は、実行はステップ２１０１０へと戻って、Ｓの別の要素が処理のために選択される。他方、要素すべてが処理されている場合は、方法はステップ２１０２５で終わって成功になる。

【０１９９】

ここで判断ステップ２１０１５へと戻る。検査の条件が失敗した場合、別の判断ステップ２１０３０が続き、これにより、選択したノードＸ$_{p}$がＰ$_{j}$の子ノードであるかが検査される。そうである場合は、方法２１０００はステップ２１０４０で終わって失敗となり、そうでない場合は、ステップ２１０３５が続く。ステップ２１０３５で、Ｓの中の要素Ｘ$_{p}$は、Ｐ$_{j}$からＸ$_{p}$への有向パスに沿って存在するその親Ｘ’$_{p}$で置き換えられ、Ｘ’$_{p}$の子孫すべてがＳから取り除かれる。次いで、実行はステップ２１０２０へと進む。

【０２００】

上で説明した方法２１０００が成功で戻った場合は、結果となる組Ｓの中の要素を使用して、値Ｑ$_{k}$の計算が子ノードＣ$_{1}$，．．．，Ｃ$_{r}$それぞれに対して行われる。すなわち

【０２０１】

【数５０】

$$Q_k = \sum_{X \in S} \Pr_{mean}[C_k \mid X \wedge P_j] \qquad \text{式53}$$

【０２０２】

しかし、方法２１０００が失敗で戻った場合は、値Ｑ$_{k}$の決定は、好ましくは、Ｃ$_{k}$とルート・ノードＹ$_{i}$の間の距離から、すなわち

【０２０３】

【数５１】

$$Q_k = \begin{cases} 1 & dist(C_k, Y_i) \le d_{\max} \\ 0 & otherwise \end{cases} \qquad \text{式54}$$

【０２０４】

またはこの代わりに、Ｃ$_{k}$とヒット・ノードＸ$_{1}$，．．．，Ｘ$_{n}$の間の距離から、すなわち

【０２０５】

【数５２】

$$Q_k = \begin{cases} 1 & \min_{i=1,\ldots,n} dist(C_k, X_i) \le d_{\max} \\ 0 & otherwise \end{cases} \qquad \text{式 55}$$

【０２０６】

のように行われる。

【０２０７】

また、単一のヒット・ノードＸがある場合、親ノードＰ$_j$およびヒット・ノードＸのもとで、Ｐ$_j$の１つの子ノードＣ$_1$だけが生起する確率

Ｐｒ［Ｃ$_1$　ｎｏ　ｓｉｂｌｉｎｇ｜Ｘ∧Ｐ$_j$］

の評価が必要であることを想起されたい。ここで、Ｃ$_1$はＸであり、またはＸの祖先である。

【０２０８】

この量を現在の想定状況へと一般化する際、２つの可能性、すなわち、ｒ＝１である特別な場合、およびｒ＞１であるより一般的な場合が生じる可能性がある。前者の例を図１５に示しているが、そこには、ノード１５００５をルートとするサブツリーの内部にある３つのヒット・ノード１５０３０、１５０３５、および１５０４０がある。しかし、３つのヒット・ノードはすべて、ノード１５００５の単一の子ノード１５０１０の下にある。

【０２０９】

この特別な場合を扱うための１つのアプローチは、項Ｐｒ［Ｃ$_1$　ｎｏ　ｓｉｂｌｉｎｇ｜Ｘ∧Ｐ$_j$］の式

【０２１０】

【数５３】

$$Q_0 = \sum_{l=1}^{s} \Pr_{mean}[C_1 \, no \, sibling \mid X_l \wedge P_j] \qquad \text{式 56}$$

【０２１１】

による置き換えを、量Ｑ$_k$を式５２で使用したのと類似のしかたで行うことである。上のＱ$_k$の場合でと同様に、式５６は、任意のＸ$_1$、ｌ＝１，．．．，ｓに対してｆｒｅｑ$_{mean}$（Ｐ$_j$，Ｘ$_l$）＝０の場合には、不定となる可能性がある。したがって、Ｑ$_0$に割り当てられる実際の値は、方法２１０００が成功で戻った場合は、方法２１０００から得られた組Ｓに基づく。すなわち

【０２１２】

【数５４】

$$Q_0 = \sum_{X \in S} \Pr_{mean}[C_1 \, no \, sibling \mid X \wedge P_j] \qquad \text{式 57}$$

【０２１３】

そうでなく方法２１０００が失敗であった場合は、Ｑ$_0$への値の割り当ては、Ｐ$_j$のルート・ノードＹ$_i$からの距離に基づいて

【０２１４】

【数５５】

$$Q_0 = \begin{cases} 0 & dist(P_j, Y_i) + 1 \le d_{\max} \\ 1 & otherwise \end{cases} \qquad \text{式 58}$$

【０２１５】

またはこの代わりに、Ｐ$_j$とヒット・ノードＸ$_1$，．．．，Ｘ$_n$の間の距離に基づいて

【０２１６】

【数５６】

$$Q_0 = \begin{cases} 0 & \min_{l=1,\ldots,n} dist(P_j, X_l) + 1 \le d_{\max} \\ 1 & otherwise \end{cases} \qquad \text{式59}$$

【０２１７】

行われる。

【０２１８】

親ノードＰ$_j$の子ノードＣ$_k$の組のうちでコンテキスト・ノードの識別をｒ＝１、ｓ＞１の場合に関して行うための、図１６のフローチャートに示す方法１６０００は、単一のヒット・ノードの場合のための方法１４０００に非常に似ている。方法１６０００は、ステップ１６００１で始まり、そこでは、概念上、架空の子ノードＣ$_0$が作成され、実際の子ノードＣ$_1$，．．．，Ｃ$_m$からなるリストへと追加され、式５７、式５８、または式５９の中で定義された値Ｑ$_0$が割り当てられ、次のステップ１６００５で、Ｃ$_1$を除く実際の子ノードＣ$_k$には、それに対応して、それぞれ、式５３、式５４、または式５５の中で定義された値Ｑ$_k$が割り当てられる。ステップ１６００６はステップ１６００５に続き、方法７０００をステップ７００５で呼び出して（ステップ７００１を飛ばして）、子ノードＣ$_0$，．．．，Ｃ$_m$のうちで１組のコンテキスト・ノードを選択する。方法７０００を抜けると、方法１６０００は再開され、判断ステップ１６０１０で検査が行われて、架空の子ノードＣ$_0$がコンテキスト・ノードとして選択されたかどうかが判定される。そうである場合、実行は継続され、ステップ１６０２０で、Ｃ$_0$がコンテキスト・ノードとして除外される。方法１６０００は、その後、ステップ１６０１５で終了する。検査が１６０１０で失敗した場合、方法１６０００は、直接に終了ステップ１６０１５へと進む。

【０２１９】

ｒ＞１（したがってｓ＞１）である一般の場合では、ｒ＝１の場合に使用したＰｒ［Ｃ$_1$　ｎｏ　ｓｉｂｌｉｎｇ｜Ｘ$_1$∧．．．∧Ｘ$_s$∧Ｐ$_j$］に類似した量は

【０２２０】

【数５７】

$$\sum_{X \in S} \Pr[C_1 \wedge \cdots \wedge C_r \wedge \neg C_{r+1} \wedge \cdots \wedge \neg C_m \mid X \wedge P_j]$$

【０２２１】

であり、ここで、Ｓは、方法２１０００が成功で抜けた場合に、これによって返される組である。残念ながら、総和の中にある確率の推定は、既存の頻度テーブルからは容易に行うことができない。したがって、わずかに異なる式がその代わりに使用される。子ノードＣ$_k$、１≦ｋ≦ｒそれぞれにルートがあるサブツリーの中にない組Ｓの要素を１≦ｌ≦ｓ$_k$に対するＨ$_{kl}$で表すとし、ここでｓ$_k$≦｜Ｓ｜である。子ノードＣ$_k$、１≦ｋ≦ｒそれぞれに対し、次のものが計算される。すなわち

【０２２２】

【数５８】

$$Q_k = \sum_{l=1}^{s_k} \Pr_{mean}[C_k \mid H_{kl} \wedge P_j] \qquad \text{式60}$$

【０２２３】

上の式の量の背後にある理論的根拠は、Ｃ$_k$、１≦ｋ≦ｒすべてにわたって足し合わせると、子ノードＣ$_1$，．．．，Ｃ$_r$がともに生起する確率を近似する量が得られるというものである（が、＞１である値をとることがあるため、真の確率ではない）。ｒ＝１の場合でのように、方法２１０００が失敗で戻った場合、Ｑ$_k$は、Ｐ$_j$のルート・ノードＹ$_i$からの距離から、１≦ｋ≦ｒに対して

【０２２４】

【数５９】

$$Q_k = \begin{cases} 0 & dist(P_j, Y_i) + 1 \le d_{\max} \\ 1 & otherwise \end{cases} \qquad 式61$$

【０２２５】

またはこの代わりに、Ｐ$_j$とヒット・ノードＸ$_1$，．．．，Ｘ$_n$の間の距離から

【０２２６】

【数６０】

$$Q_k = \begin{cases} 0 & \min_{i=1,\ldots,n.} dist(P_j, X_i) + 1 \le d_{\max} \\ 1 & otherwise \end{cases} \qquad 式62$$

【０２２７】

のように得られる。

【０２２８】

子ノードの組Ｃ$_1$，．．．，Ｃ$_m$のうちでコンテキスト・ノードを選択するための方法１７０００を、ここで、ｒ＞１、ｓ＞１の一般の場合に関して、図１７のフローチャートを参照して説明する。方法１７０００は、ステップ１７００１で始まり、ここで、子ノードＣ$_k$、１≦ｋ≦ｒそれぞれには、式６０、式６１、または式６２を用いて計算された値Ｑ$_k$が割り当てられる。ステップ１７００５がこれに続き、ここでは、残りの子ノードには、それに対応して、式５３、式５４、または式５５を用いて計算された値Ｑ$_k$が、それぞれ、割り当てられる。次のステップ１７０１０で、値Ｑ$_k$が、子ノードすべてにわたって総計され、Ｔと表される。方法１７０００は継続して、ステップ１７０１５で、そのサブツリーにヒット・ノードを含む子ノードすべて、すなわち、Ｃ$_k$、１≦ｋ≦ｒがコンテキスト・ノードとして選択される。次のステップ１７０２０で、割り当てられた値が残りの子ノードのうちで最高のノードＣ$_k$もコンテキスト・ノードとして選択される。値が最大で同じ子ノードが１つより多く存在する場合は、そのようなノードはすべて、コンテキスト・ノードとして選択される。次いで、それまでにコンテキスト・ノードとして選択された子ノードすべての割り当てられた値の総和が、ステップ１７０２５で計算され、Ｓと表される。次いで、実行は判断ステップ１７０３０へと進み、このポイントで、子ノードＣ$_k$すべてがコンテキスト・ノードとして選択されている場合、方法１７０００は、ステップ１７０４０で終了する。しかし、コンテキスト・ノードとしてまだ選択されていない１つまたは複数の子ノードＣ$_k$がある場合は、方法１７０００は、別の判断ステップ１７０３５へと続く。ステップ１７０３５で、検査が行われて、Ｓ≧Ｔ／２であるかどうかが確かめられ、そうである場合、方法１７０００はここでもステップ１７０４０で終了する。Ｓ＜Ｔ／２である場合は、実行はステップ１７０２０へと戻り、ここでノードのコンテキスト・ノードとしての選択がさらに行われる。

【０２２９】

以上の説明では、階層的なデータ構造内のキーワード検索を行うときに出遭う異なる段階および動作の想定状況を扱うための様々な方法を提示している。こうした方法は、単一の全体的な手続き１８０００へと組み込まれるが、これは、図２のステップ２０１０を詳細にしたものであり、図１９および図２０にそれぞれ示す下位の手続き１９０００および２００００を含む図１８のフローチャートで示している。方法１８０００は、判断ステップ１８００５で始まり、ここで、検査が行われて、スキーマ・グラフ内に複数のヒット・ノードがあるかどうかの判定が行われる。そうである場合、実行はステップ１８０１５へと進み、ここで方法２００００が呼び出され、そうでない場合は、実行はステップ１８０１０へと進み、ここで方法１９０００が呼び出される。どちらの場合でも、方法２００００または１９０００は、コンテキストツリーからなるリストとともに戻り、それぞれには、スコアが関連づけられている。以下は、方法１９０００の詳細な説明であり、これに方法２００００のそれが続く。

【０２３０】
方法１９０００は、ステップ１９００１で始まり、ここで、方法１００００が呼び出されて、ヒット・ノードＸの祖先ノードである可能なルート・ノード$Ｙ_{i}$からなるリストが決定される。$Ｙ_{i}$それぞれは、可能なコンテキストツリーのルート・ノードである。また、方法１００００では、値$Ｓ_{i}$＝Ｐｒ’［$Ｙ_{i}$｜Ｘ］の計算がノード$Ｙ_{i}$それぞれに対して行われる。次いで、方法１９０００は継続して、ステップ１９００５で、その前のステップで決定されたノード$Ｙ_{i}$が処理のために選択される。次のステップ１９０１０で、方法１９０００内部にある下位プロセスである方法３８０００が呼び出されて、ノード$Ｙ_{i}$をルートとするサブツリーの中にあるコンテキスト・ノードが識別される。次いで、方法１９０００は継続して、ステップ１９０３０で、識別されたコンテキスト・ノードすべてを含み、$Ｙ_{i}$をルート・ノードとするコンテキストツリーが構築される。このツリーには、ステップ１９００１で計算された$Ｓ_{i}$のスコアが割り当てられる。次いで、方法１９０００は、判断ステップ１９０３５へと進む。ステップ１９００１で得られたノード$Ｙ_{i}$すべてが処理されている場合は、この方法はステップ１９０４０で終わり、そうでない場合は、この方法はステップ１９００５へと戻って、別のノード$Ｙ_{i}$の処理が行われる。

【０２３１】
方法１９０００の内部で呼び出される方法３８０００は、ステップ３８０１０で始まり、ここで、ノード$Ｙ_{i}$が、まず$Ｐ_{j}$に割り当てられる。実行は、判断ステップ３８０１５へ、次いで、$Ｐ_{j}$が$Ｙ_{i}$からヒット・ノードＸへの有向パス上に存在しない場合は、３８０２０へと進む。ステップ３８０２０で、方法１３０００が呼び出されて、$Ｐ_{j}$の子ノードのうちで１組のコンテキスト・ノードが選択される。続くステップ３８０２５では、方法３８０００の再帰的な呼び出しが、ステップ３８０２０で（ステップ３８０１０および３８０１５を飛ばして）、コンテキスト・ノードとして選択された葉でない子ノード$Ｃ_{k}$それぞれに対して行われ、$Ｃ_{k}$に$Ｐ_{j}$の役割をさせることによって、さらに別のコンテキスト・ノードがその子孫のうちで識別される。そのような子ノードすべてに対する呼び出しが戻ると、方法３８０００は、ステップ３８０４０で終了する。また、方法３８０００は、$Ｐ_{j}$に子ノードがない場合、またはその葉でない子ノードのうちのどれもが、ステップ３８０２０でコンテキスト・ノードとして選択されていない場合には、直接に終了ステップ３８０４０へと進む。

【０２３２】
判断ステップ３８０１５は、$Ｐ_{j}$が$Ｙ_{i}$からＸへの有向パス上に存在する場合に成功し、その場合、実行はステップ３８０４５へと進む。ここで、方法１４０００が呼び出されて、$Ｐ_{j}$の子ノードのうちで１組のコンテキスト・ノードが選択されるが、$Ｃ_{1}$は、$Ｐ_{j}$からＸへの有向パス上に存在する子ノードを表す。続くステップ３８０５０で、方法３８０００の再帰的な呼び出しが、ステップ３８０１５で（ステップ３８０１０を飛ばして）、コンテキスト・ノードとして選択された葉でない子ノード$Ｃ_{k}$それぞれに対して行われ、$Ｃ_{k}$に$Ｐ_{j}$の役割をさせることによって、さらに別のコンテキスト・ノードがその子孫のうちで識別される。そのような子ノードすべてに対する呼び出しが戻ると、方法３８０００は、ステップ３８０４０で終了する。

【０２３３】
方法２００００は、ステップ２０００１で始まり、ここで、方法１２０００が呼び出されて、ヒット・ノード$Ｘ_{1}$，．．．，$Ｘ_{n}$の祖先ノードである可能なルート・ノード$Ｙ_{i}$からなるリストが決定される。$Ｙ_{i}$それぞれは、可能なコンテキストツリーのルート・ノードである。また、方法１２０００では、値$Ｓ_{i}$＝Ｐｒ’［$Ｙ_{i}$｜$Ｘ_{1}$∧．．．∧$Ｘ_{n}$］の計算がノード$Ｙ_{i}$それぞれに対して行われる。次いで、方法２００００は継続して、ステップ２０００５で、その前のステップで決定されたノード$Ｙ_{i}$が処理のために選択される。次のステップ２００１０で、方法２００００内部にある下位プロセスである方法３９０００が呼び出されて、ノード$Ｙ_{i}$をルートとするサブツリーの中にあるコンテキスト・ノードが識別される。次いで、方法２００００は継続して、ステップ２００６０で、識別されたコンテキスト・ノードすべてを含み、$Ｙ_{i}$をルート・ノードとするコンテキストツ

リーが構築される。このツリーには、ステップ２０００１で計算されたＳ$_i$のスコアが割り当てられる。次いで、方法２００００は、判断ステップ２００６５へと進む。ステップ２０００１で得られたノードＹ$_i$すべてが処理されている場合は、この方法はステップ２００７０で終わり、そうでない場合は、この方法はステップ２００２５へと戻って、別のノードＹ$_i$の処理が行われる。

【０２３４】

方法１９０００の内部で呼び出される方法３９０００は、ステップ３９０１０で始まり、ここで、ノードＹ$_i$が、まずＰ$_j$に割り当てられる。実行は、判断ステップ３９０１５へ、次いで、Ｐ$_j$のところのサブツリーのルートにおいてヒット・ノードがない場合は、ステップ３９０２０へと進む。ステップ３９０２０で、方法１３０００が呼び出されて、Ｐ$_j$の子ノードのうちで１組のコンテキスト・ノードが選択される。続くステップ３９０２５では、方法３９０００の再帰的な呼び出しが、ステップ３９０２０で（ステップ３９０１０および３９０１５を飛ばして）、コンテキスト・ノードとして選択された葉でない子ノードＣ$_k$それぞれに対して行われ、Ｃ$_k$にＰ$_j$の役割をさせることによって、さらに別のコンテキスト・ノードがその子孫のうちで識別される。そのような子ノードすべてに対する呼び出しが戻ると、方法３９０００は、ステップ３９０６０で終了する。また、方法３９０００は、Ｐ$_j$に子ノードがない場合、またはその葉でない子ノードのうちのどれもが、ステップ３９０２０でコンテキスト・ノードとして選択されていない場合には、直接に終了ステップ３９０６０へと進む。

【０２３５】

判断ステップ３９０１５は、Ｐ$_j$をルートとするサブツリーの内部に１つまたは複数のヒット・ノードがある場合に成功し、その場合、実行は別のステップ３９０３０へと進む。Ｐ$_j$の下のサブツリーの中に単一のヒット・ノードしかない場合は、この判断ステップは失敗し、実行はステップ３９０３５へと進み、そうでない場合は、実行はまた別の判断ステップ３９０４０へと進む。判断ステップ３９０４０で、検査が行われて、Ｐ$_j$の下のヒット・ノードすべてが、その子ノードのうちの１つの下にしかないかどうかが判定される。そうである場合は、実行はステップ３９０４５へと進み、そうでない場合は、実行はステップ３９０５０へと進む。ステップ３９０５０では、Ｃ$_1$，．．．，Ｃ$_r$で１つまたは複数のヒット・ノードが下にあるＰ$_j$の子ノードを表すとして、方法１７０００が呼び出されて、Ｐ$_j$の子ノードのうちで１組のコンテキスト・ノードが選択される。しかし、判断ステップ３９０４０からステップ３９０４５へと導かれた場合は、方法１６０００が呼び出されて、Ｐ$_j$の子ノードのうちで１組のコンテキスト・ノードを選択することが、Ｃ$_1$をそのサブツリーの中にヒット・ノードを含む、Ｐ$_j$の唯一の子ノードとして行われる。

【０２３６】

ここで３９０３５へと戻り、Ｐ$_j$からその唯一の子孫のヒット・ノードへのパスが、その子ノードＣ$_1$を通過するとする。方法１４０００が呼び出されて、Ｐ$_j$の子ノードのうちで１組のコンテキスト・ノードが選択される。

【０２３７】

ステップ３９０３５、３９０４５、および３９０５０のそれぞれが完了すると、方法３９０００のそれ自体の再帰的な呼び出しが、ステップ３９０１５で（ステップ３９０１０を飛ばして）、コンテキスト・ノードとして選択された葉でない子ノードＣ$_k$それぞれに対して行われ、Ｃ$_k$にＰ$_j$の役割をさせることによって、さらに別のコンテキスト・ノードがその子孫のうちで識別される。そのような子ノードすべてに対する呼び出しが戻ると、方法３９０００は、ステップ３９０６０で終了する。

【０２３８】

説明のための例

ここで、好ましい実装形態の動作の説明を、下の階層的なＸＭＬデータソースの例によって行う。このＸＭＬソースは、「ＸＹＺ」という名称の企業に関する、Ｗｅｂアドレス、支社の名称および場所、各支社での販売製品の範囲などのデータを含んでいる。このＸ

ＭＬデータのスキーマ・グラフ表現を、図２３に示している。

【０２３９】

ＸＭＬソース

```
＜ｃｏｍｐａｎｙ＞ ＜！－－企業－－＞
 ＜ｎａｍｅ＞ＸＹＺ＜／ｎａｍｅ＞ ＜！－－名称－－＞
 ＜ｗｅｂ＞ｈｔｔｐ：／／ｗｗｗ．ｘｙｚ．ｃｏｍ＜／ｗｅｂ＞ ＜！－－ Ｗｅｂアドレス－－＞
 ＜ｄｅｓｃｒｉｐｔｉｏｎ＞ ＜！－－説明－－＞
  Ｃｏｍｐａｎｙ ｆｏｕｎｄｅｄ ｉｎ １９９９ ｓｐｅｃｉａｌｉｓｉｎｇ ｉｎ ｈｉ－ｔｅｃｈ ｃｏｎｓｕｍｅｒｓ ｅｌｅｃｔｒｏｎｉｃｓ ＜！－－１９９９年設立のハイテク家電製品に特化した企業－－＞
 ＜／ｄｅｓｃｒｉｐｔｉｏｎ＞
 ＜ｂｒａｎｃｈ＞ ＜！－－支社－－＞
  ＜ｎａｍｅ＞Ｎｏｒｔｈ Ｒｙｄｅ＜／ｎａｍｅ＞ ＜！－－名称－－＞
  ＜ｐｈｏｎｅ＞０２９１２３０００＜／ｐｈｏｎｅ＞ ＜！－－電話番号－－＞
  ＜ａｄｄｒｅｓｓ＞ ＜！－－住所－－＞
   ＜ｎｕｍｂｅｒ＞１＜／ｎｕｍｂｅｒ＞ ＜！－－番地－－＞
   ＜ｓｔｒｅｅｔ＞Ｌａｎｅ Ｃｏｖｅ＜／ｓｔｒｅｅｔ＞ ＜！－－街区 Ｃｏｖｅ通り－－＞
   ＜ｃｉｔｙ＞Ｓｙｄｎｅｙ＜／ｃｉｔｙ＞ ＜！－－都市 シドニー－－＞
   ＜ｃｏｕｎｔｒｙ＞Ａｕｓｔｒａｌｉａ＜／ｃｏｕｎｔｒｙ＞ ＜！－－国 オーストラリア－－＞
  ＜／ａｄｄｒｅｓｓ＞
  ＜ｍａｎａｇｅｒ＞ ＜！－－支社長－－＞
   ＜ｆｉｒｓｔＮａｍｅ＞Ｊｉｍ＜／ｆｉｒｓｔＮａｍｅ＞ ＜！－－名－－＞
   ＜ｌａｓｔＮａｍｅ＞Ｓｍｉｔｈ＜／ｌａｓｔＮａｍｅ＞ ＜！－－姓－－＞
   ＜ｅｍａｉｌ＞ｊｓｍｉｔｈ@ｘｙｚ．ｃｏｍ＜／ｅｍａｉｌ＞ ＜！@電子メール－－＞
  ＜／ｍａｎａｇｅｒ＞
  ＜ｐｒｏｄｕｃｔ＞ ＜！－－製品－－＞
   ＜ｉｄ＞１＜／ｉｄ＞ ＜！－－ＩＤ－－＞
   ＜ｎａｍｅ＞Ｐｌａｓｍａ ＴＶ＜／ｎａｍｅ＞ ＜！－－名称プラズマＴＶ－－＞
   ＜ｐｒｉｃｅ＞＄１０００＜／ｐｒｉｃｅ＞ ＜！－－価格－－＞
   ＜ｓｕｐｐｌｉｅｒ＞ＪＥＣ＜／ｓｕｐｐｌｉｅｒ＞ ＜！－－メーカー－－＞
   ＜ｓｔｏｃｋ＞１０＜／ｓｔｏｃｋ＞ ＜！－－在庫数－－＞
  ＜／ｐｒｏｄｕｃｔ＞
  ＜ｐｒｏｄｕｃｔ＞ ＜！－－製品－－＞
   ＜ｉｄ＞２＜／ｉｄ＞ ＜！－－ＩＤ－－＞
   ＜ｎａｍｅ＞Ｍｐ３ ｐｌａｙｅｒ＜／ｎａｍｅ＞ ＜！－－名称 ＭＰ３プレイヤー－－＞
   ＜ｐｒｉｃｅ＞＄５００ ＜／ｐｒｉｃｅ＞ ＜！－－価格－－＞
   ＜ｓｕｐｐｌｉｅｒ＞ＨＧ＜／ｓｕｐｐｌｉｅｒ＞ ＜！－－メーカー－－＞
   ＜ｓｔｏｃｋ＞２０＜／ｓｔｏｃｋ＞ ＜！－－在庫数－－＞
  ＜／ｐｒｏｄｕｃｔ＞
 ＜／ｂｒａｎｃｈ＞
 ＜ｂｒａｎｃｈ＞ ＜！－－支社－－＞
  ＜ｎａｍｅ＞Ｍｏｒｌｅｙ＜／ｎａｍｅ＞ ＜！－－名称－－＞
```

```
      <phone>08912300000</phone> <!--電話番号-->
      <address> <!--住所-->
        <number>1</number> <!--番地-->
        <street>Russel</street> <!--街区 Russel-->
        <city>Perth</city> <!--都市 パース-->
        <country>Australia</country> <!--国 オーストラリア-->
      </address>
      <manager> <!--支社長-->
        <firstName>Ted</firstName> <!--名-->
        <lastName>White</lastName> <!--姓-->
        <email>twhite@xyz.com</email> <!@電子メール-->
      </manager>
      <product> <!--製品-->
        <id>3</id> <!--ID-->
        <name>Video phone</name> <!--名称 ビデオ電話-->
        <price>$2000</price> <!--価格-->
        <supplier>NVC<supplier> <!---メーカー--->
        <stock>15</stock> <!--在庫数-->
      </product>
      <product> <!--製品-->
        <id>4</id> <!--ID-->
        <name>PDA</name> <!--名称 PDA-->
        <price>$1000</price> <!--価格-->
        <supplier>LP</supplier> <!---メーカー--->
        <stock>50</stock> <!--在庫数-->
      </product>
    </branch>
  </company>
```

図２３で、各ノードの次に示す整数は、そのノードに割り当てられている一意のＩＤである。このデータソースの既存のビューが３つあるとする。第１のものは、この企業の名称、説明、およびＷｅｂアドレスを表示しているビューである。第２のものは、この企業の支社およびその場所のリストであり、最後に第３のビューには、各支社での製品の範囲（ｌｉｎｅ）がリストされている。これらのビューのスキーマ・グラフ表現を、図２４、図２５、および図２６にそれぞれ示してある。これらのビューの結果として、生起２７０００、共起２８０００、葉共起２９０００、および単独子共起３００００の頻度テーブルを、図２７、図２８、図２９、および図３０にそれぞれ示してある。同時生起頻度テーブルは、３次元であるが、５つの別々の２次元テーブル３１０００、３２０００、３３０００、３４０００、および３５０００によって示してある。図３１は、P_j＝ノード１であるテーブル内のエントリｆｒｅｑ（C_k，P_j，Ｘ）を含んでいる。同様に、図３２、図３３、図３４、および図３５それぞれは、P_j＝ノード３、ノード８、ノード９、およびノード１０であるエントリをそれぞれ含む。ここに示す頻度テーブルすべてにおいて、図２８に見られるような、項目２８００５などの空のセルは、それに関連する頻度を保存する必要のない無効なノードの組合せを表す。

【０２４０】

ユーザが、自分のいる市の中で特定の製品を見つけることを希望しているとする。ユー

ザは、その製品の名前、「ＭＰ３プレイヤー」および市の名前「シドニー」を入力し、両方の名前に対するキーワード検索を行う。図２３からわかるように、これから、２つのヒット・ノードX_1＝ノード１９およびX_2＝ノード１３が得られる。このキーワード検索操作に対する可能なコンテキストツリーを決定するために、システム４０００では、図１８の方法１８０００を呼び出す。ヒット・ノードが１つより多くあることから、方法１８０００は、その後、方法２００００の呼び出しをステップ１８０１５で行う。方法２００００は、今度は、方法１２０００の呼び出しをステップ２００１で行って、結果となるコンテキストツリーのルート・ノードの役をするノードY_iのリストが得られる。

【０２４１】

方法１２０００では、まず、ステップ１２００１で、ノード３を、ヒット・ノードX_1とX_2をどちらも含む最小のサブツリーのルート・ノードとして識別する。したがって、Ａ＝ノード３である。次いで、方法１２０００は、再帰手続きを始めて、このヒット・ノードのもとで、Ａおよびその祖先のそれぞれに対する生起確率の値を計算する。ノードＡでは

$Pr'[A \mid X_1 \wedge X_2] = 1$

である。Y_1＝ノード１、すなわちノードＡの親では、式４７、式４８、および式４９を用いて

【０２４２】

【数６１】

$$Pr_{mean}[node\,1 \mid A \wedge X_1 \wedge X_2] = \frac{freq(node\,1, node\,19)\, freq(node\,1, node\,13)\, freq(node\,3)}{freq(node\,3, node\,19)\, freq(A, node\,13)\, freq(node\,1)}$$

【０２４３】

したがって

$Pr'[node\ 1 \mid X_1 \wedge X_2] = 0$

であり

$Pr'[A\ root \mid X_1 \wedge X_2] = 1$

である。

【０２４４】

このため、方法１２０００は、ノードＡをコンテキストツリーの単一のルート・ノード候補として抜ける。このコンテキストツリーには１というスコアが割り当てられる。方法１２０００の完了後、方法２００００では、第２の、トップ・ダウン走査のフェーズが続けられ、そこでは、ルート・ノードY_1＝Ａの子孫が処理されて、それらのうちでコンテキスト・ノードが識別される。このフェーズはステップ２００１０で始まり、ここで、P_jがまずノード３になるように設定される。このノードは、２つの異なる子ノードの下にある、ヒット・ノードX_1およびX_2の祖先であることから、実行は、結局、方法３９０００のステップ３９０５０へと進み、ここで方法１７０００が呼び出されて、コンテキスト・ノードがその子のうちで決定される。方法１７０００の結果、ノード３の子ノード６～１０それぞれに割り当てられる値$Q_1, \ldots, Q_5$は、次の通りである。

【０２４５】

【数６２】

$$Q_1 = \Pr_{mean}[node\ 6 \mid X_1 \wedge P_j] + \Pr_{mean}[node\ 6 \mid X_2 \wedge P_j]$$

$$= \frac{freq(node\ 6, node\ 3, node\ 19)}{freq(node\ 3, node\ 19)} + \frac{freq(node\ 6, node\ 3, node\ 13)}{freq(node\ 3, node\ 13)}$$

$$= \frac{1}{1} + \frac{1}{1}$$

$$= 2$$

$$Q_2 = \Pr_{mean}[node\ 7 \mid X_1 \wedge P_j] + \Pr_{mean}[node\ 7 \mid X_2 \wedge P_j]$$

$$= \frac{freq(node\ 7, node\ 3, node\ 19)}{freq(node\ 3, node\ 19)} + \frac{freq(node\ 7, node\ 3, node\ 13)}{freq(node\ 3, node\ 13)}$$

$$= 2$$

$$Q_3 = \Pr_{mean}[node\ 8 \mid X_1 \wedge P_j]$$

$$= \frac{freq(node\ 8, node\ 3, node\ 19)}{freq(node\ 3, node\ 19)}$$

$$= 0$$

$$Q_4 = \Pr_{mean}[node\ 9 \mid X_1 \wedge P_j] + \Pr_{mean}[node\ 9 \mid X_2 \wedge P_j]$$

$$= \frac{freq(node\ 9, node\ 3, node\ 19)}{freq(node\ 3, node\ 19)} + \frac{freq(node\ 9, node\ 3, node\ 13)}{freq(node\ 3, node\ 13)}$$

$$= 1$$

$$Q_5 = \Pr_{mean}[node\ 10 \mid X_2 \wedge P_j]$$

$$= \frac{freq(node\ 10, node\ 3, node\ 13)}{freq(node\ 3, node\ 13)}$$

$$= 0$$

【０２４６】

したがって、降順でソートされた組Ｑ$_1$，．．．，Ｑ$_5$は、｛Ｑ$_1$，Ｑ$_2$，Ｑ$_4$，Ｑ$_3$，Ｑ$_5$｝であり、その総計はＴ＝５となる。方法１７０００によって選択されたコンテキスト・ノードの組は、こうして、ノード６、ノード７（Ｑ$_1$＋Ｑ$_2$＞Ｔ／２であるため）、およびノード８、ノード１０（ヒット・ノードの祖先であるため）を含む。次いで、方法３９０００は、ステップ３９０５５で再開して、それ自体を再帰的に呼び出して、子のある選択されたノードそれぞれに対してコンテキストの子ノードの識別を行う。

【０２４７】

Ｐ$_j$＝ノード８の場合、実行は、ノード８に単一の子孫のヒット・ノード（ノード１３

）があることから、ステップ３９０３５へと進み、このポイントで方法１４０００が呼び出されて、コンテキスト・ノードが子ノードの組１１～１４のうちで識別される。方法１４０００の結果、P_jの子ノード１１、１２、および１４にそれぞれ割り当てられる確率の値Q_1、Q_2、Q_4は次の通りである。

【０２４８】

【数６３】

$$Q_1 = \Pr_{mean}[node\,11 \mid X_2 \wedge P_j]$$

$$= \frac{freq(node\,11, node\,8, node\,13)}{freq(node\,8, node\,13)}$$

$$= 1$$

$$Q_2 = \Pr_{mean}[node\,12 \mid X_2 \wedge P_j]$$

$$= \frac{freq(node\,12, node\,8, node\,13)}{freq(node\,8, node\,13)}$$

$$= 1$$

$$Q_4 = \Pr_{mean}[node\,14 \mid X_2 \wedge P_j]$$

$$= \frac{freq(node\,14, node\,8, node\,13)}{freq(node\,8, node\,13)}$$

$$= 1$$

【０２４９】

さらに、方法１４０００では、架空の子ノードC_0に対する値Q_0の計算も行われる。

【０２５０】

【数６４】

$$Q_0 = \Pr_{mean}[node\,13\ no\ sibling \mid X_2 \wedge P_j]$$

$$= \frac{freq(node\,8\ has\ 1\ child, node\,13)}{freq(node\,8, node\,13)}$$

$$= 0$$

【０２５１】

したがって、降順でソートされた確率の値の組は、｛Q_1，Q_2，Q_4，Q_0｝であり、その総計はＴ＝３となる。方法１４０００によって選択されたコンテキスト・ノードの組は、こうして、ノード１１、ノード１２、ノード１４（$Q_1+Q_2+Q_4>T/2$かつ$Q_1=Q_2=Q_4$であるため）、およびノード１３（ヒット・ノードの祖先であるため）を含む。

【０２５２】

これに似た実行パスを、P_j＝ノード１０の場合にはたどることになり、似た結果が得られる。ノード１０のコンテキストの子ノードの組は、ノード１８～２２である。したが

って、このコンテキストツリーのスキーマ・グラフ３６００は、図３６に示すようになり、ヒット・ノード１９および１３、ならびにコンテキスト・ノード３、６～８、１０～１４、１８～２２を含む。これらのノードの表現するデータ項目を含む、ユーザへと返される実際のコンテキストツリーは、次のようなものである。

【０２５３】

```
    <ｂｒａｎｃｈ> <！－－支社－－>
      <ｎａｍｅ>Ｎｏｒｔｈ Ｒｙｄｅ</ｎａｍｅ> <！－－名称－－>
      <ｐｈｏｎｅ>０２９１２３０００</ｐｈｏｎｅ> <！－－電話番号－－
>
      <ａｄｄｒｅｓｓ> <！－－住所－－>
      <ｎｕｍｂｅｒ>１</ｎｕｍｂｅｒ> <！－－番地－－>
      <ｓｔｒｅｅｔ>Ｌａｎｅ Ｃｏｖｅ</ｓｔｒｅｅｔ> <！－－街区 Ｃｏ
ｖｅ通り－－>
      <ｃｉｔｙ>Ｓｙｄｎｅｙ</ｃｉｔｙ> <！－－都市 シドニー－－>
      <ｃｏｕｎｔｒｙ>Ａｕｓｔｒａｌｉａ</ｃｏｕｎｔｒｙ> <！－－国 オ
ーストラリア－－>
      </ａｄｄｒｅｓｓ>
      <ｐｒｏｄｕｃｔ> <！－－製品－－>
      <ｉｄ>１</ｉｄ> <！－－ＩＤ－－>
      <ｎａｍｅ>Ｐｌａｓｍａ ＴＶ</ｎａｍｅ> <！－－名称 プラズマＴＶ
－－>
      <ｐｒｉｃｅ>＄１００００</ｐｒｉｃｅ> <！－－価格－－>
      <ｓｕｐｐｌｉｅｒ>ＪＥＣ</ｓｕｐｐｌｉｅｒ> <！－－メーカー－－>
      <ｓｔｏｃｋ>１０</ｓｔｏｃｋ> <！－－在庫数－－>
      </ｐｒｏｄｕｃｔ>
      <ｐｒｏｄｕｃｔ> <！－－製品－－>
      <ｉｄ>２</ｉｄ> <！－－ＩＤ－－>
      <ｎａｍｅ>Ｍｐ３ ｐｌａｙｅｒ</ｎａｍｅ> <！－－名称 ＭＰ３プレ
イヤー－－>
      <ｐｒｉｃｅ>＄５００</ｐｒｉｃｅ> <！－－価格－－>
      <ｓｕｐｐｌｉｅｒ>ＨＧ</ｓｕｐｐｌｉｅｒ> <！－－メーカー－－>
      <ｓｔｏｃｋ>２０</ｓｔｏｃｋ> <！－－在庫数－－>
      </ｐｒｏｄｕｃｔ>
    </ｂｒａｎｃｈ>
    <ｂｒａｎｃｈ> <！－－支社－－>
      <ｎａｍｅ>Ｍｏｒｌｅｙ</ｎａｍｅ> <！－－名称－－>
      <ｐｈｏｎｅ>０８９１２３０００</ｐｈｏｎｅ> <！－－電話番号－－
>
      <ａｄｄｒｅｓｓ> <！－－住所－－>
      <ｎｕｍｂｅｒ>１</ｎｕｍｂｅｒ> <！－－番地－－>
      <ｓｔｒｅｅｔ>Ｒｕｓｓｅｌ</ｓｔｒｅｅｔ> <！－－街区 Ｒｕｓｓｅ
ｌ－－>
      <ｃｉｔｙ>Ｐｅｒｔｈ</ｃｉｔｙ> <！－－都市 パース－－>
      <ｃｏｕｎｔｒｙ>Ａｕｓｔｒａｌｉａ</ｃｏｕｎｔｒｙ> <！－－国 オ
ーストラリア－－>
      </ａｄｄｒｅｓｓ>
      <ｐｒｏｄｕｃｔ> <！－－製品－－>
      <ｉｄ>３</ｉｄ> <！－－ＩＤ－－>
      <ｎａｍｅ>Ｖｉｄｅｏ ｐｈｏｎｅ</ｎａｍｅ> <！－－名称 ビデオ電
話－－>
```

＜ｐｒｉｃｅ＞＄２０００＜／ｐｒｉｃｅ＞＜！－－価格－－＞
＜ｓｕｐｐｌｉｅｒ＞ＮＶＣ＜ｓｕｐｐｌｉｅｒ＞　＜！－－メーカー－－＞
＜ｓｔｏｃｋ＞１５＜／ｓｔｏｃｋ＞　＜！－－在庫数－－＞
＜／ｐｒｏｄｕｃｔ＞
＜ｐｒｏｄｕｃｔ＞　＜！－－製品－－＞
＜ｉｄ＞４＜／ｉｄ＞　＜！－－ＩＤ－－＞
＜ｎａｍｅ＞ＰＤＡ＜／ｎａｍｅ＞　＜！－－名称　ＰＤＡ－－＞
＜ｐｒｉｃｅ＞＄１０００＜／ｐｒｉｃｅ＞　＜！－－価格－－＞
＜ｓｕｐｐｌｉｅｒ＞ＬＰ＜／ｓｕｐｐｌｉｅｒ＞　＜！－－メーカー－－＞
＜ｓｔｏｃｋ＞５０＜／ｓｔｏｃｋ＞　＜！－－在庫数－－＞
＜／ｐｒｏｄｕｃｔ＞
＜／ｂｒａｎｃｈ＞

産業上の利用可能性

上記から、説明した配置がコンピュータおよびデータ処理の産業に対して、特に複数の検索からの情報を提示することに関して適用可能であることは明らかである。

【０２５４】

上記では、本発明の一部の実施形態だけを説明しており、これへの変更および／または改変は、本発明の範囲および趣旨から逸脱することなく行うことができ、これら実施形態は例示的であり、制限的ではない。

【０２５５】

（オーストラリア限定）この明細書の文脈では、「含む」（ｃｏｍｐｒｉｓｉｎｇ）が意味するのは、「主として、しかし必ずしももっぱらにではなく含む」（ｉｎｃｌｕｄｉｎｇ　ｐｒｉｎｃｉｐａｌｌｙ　ｂｕｔ　ｎｏｔ　ｎｅｃｅｓｓａｒｉｌｙ　ｓｏｌｅｌｙ）または「有する」（ｈａｖｉｎｇ）または「含む」（ｉｎｃｌｕｄｉｎｇ）であって、「だけから構成される」（ｃｏｎｓｉｓｔｉｎｇ　ｏｎｌｙ　ｏｆ）ではない。「ｃｏｍｐｒｉｓｅ」や「ｃｏｍｐｒｉｓｅｓ」など、「ｃｏｍｐｒｉｓｉｎｇ」という語の変種の意味は、これに対応して変わる。

【図面の簡単な説明】

【０２５６】

【図１】スキーマ・グラフの例の図である。
【図２】キーワード検索方法のフローチャートである。
【図３Ａ】スキーマ・グラフ内の親ノードの２つの例を示す図である。
【図３Ｂ】スキーマ・グラフ内の親ノードの２つの例を示す図である。
【図４】サーバおよびクライアントのコンピュータからなるネットワークの図である。
【図５】ルート・ノードからヒット・ノードへの有向パスに沿って存在しない親ノードの１組の子ノードのうちでコンテキスト・ノードを識別するための方法のフローチャートである。
【図６】ルート・ノードからヒット・ノードへの有向パスに沿って存在しない親ノードの１組の子ノードのうちでコンテキスト・ノードを識別するための別の方法のフローチャートである。
【図７】ルート・ノードからヒット・ノードへの有向パスに沿って存在する親ノードの１組の子ノードのうちでコンテキスト・ノードを識別するための方法のフローチャートである。
【図８】ルート・ノードからヒット・ノードへの有向パスに沿って存在する親ノードの１組の子ノードのうちでコンテキスト・ノードを識別するための別の方法のフローチャートである。
【図９】２つの同じサブツリーをスキーマ・グラフの例の図である。
【図１０】確率平均化によるコンテキスト・ノード識別方法の第１の、ボトム・アップ横断フェーズのフローチャートである。
【図１１】複数のヒット・ノードのあるスキーマ・グラフの例の図である。

【図１２】確率平均化による、複数のヒット・ノードを含むコンテキスト・ノード識別方法の第１の、ボトム・アップ横断フェーズのフローチャートである。
【図１３】確率平均化による、ルート・ノードからヒット・ノードへの有向パスに沿って存在しない親ノードの１組の子ノードのうちでコンテキスト・ノードを識別するための方法のフローチャートである。
【図１４】確率平均化による、ルート・ノードからヒット・ノードへの有向パスに沿って存在する親ノードの１組の子ノードのうちでコンテキスト・ノードを識別するための方法のフローチャートである。
【図１５】子孫のヒット・ノードがすべて単一の子ノードの下に位置する親ノードの例の図である。
【図１６】複数のヒット・ノードすべてが単一の子ノードの下に位置する場合に対する、確率平均化による、ルート・ノードからヒット・ノードへの有向パスに沿って存在する親ノードの１組の子ノードのうちでコンテキスト・ノードを識別するための方法のフローチャートである。
【図１７】複数のヒット・ノードが複数の子ノードの下に位置する場合に対する、確率平均化による、ルート・ノードからヒット・ノードへの有向パスに沿って存在する親ノードの１組の子ノードのうちでコンテキスト・ノードを識別するための方法のフローチャートである。
【図１８】１つまたは複数のヒット・ノードが存在する可能性のある、コンテキスト・ノードを識別するための方法のフローチャートである。
【図１９】単一のヒット・ノードを含む場合に対するコンテキストツリーを構築するための方法のフローチャートである。
【図２０】複数のヒット・ノードを含む場合に対するコンテキストツリーを構築するための方法のフローチャートである。
【図２１】観測頻度がヒット・ノードからなる元の組での観測頻度よりも高いヒット・ノードからなる代替の組を構築するための方法のフローチャートである。
【図２２】観測頻度がヒット・ノードからなる組よりも高いそのヒット・ノードからなる組の祖先を選択するための方法のフローチャートである。
【図２３】スキーマ・グラフの例の図である。
【図２４】データ・ビューの例のスキーマ・グラフの図である。
【図２５】データ・ビューの別の例のスキーマ・グラフの図である。
【図２６】データ・ビューのさらに別の例のスキーマ・グラフの図である。
【図２７】図２４、２５、および２６のデータ・ビューから生じる生起頻度テーブルの図である。
【図２８】図２４、２５、および２６のデータ・ビューから生じる共起頻度テーブルの図である。
【図２９】図２４、２５、および２６のデータ・ビューから生じる葉の共起頻度テーブルの図である。
【図３０】図２４、２５、および２６のデータ・ビューから生じる単独子の共起頻度テーブルの図である。
【図３１】図２４、２５、および２６のデータ・ビューから生じる同時生起頻度テーブルのある部分の図である。
【図３２】図２４、２５、および２６のデータ・ビューから生じる同時生起頻度テーブルの別の部分の図である。
【図３３】図２４、２５、および２６のデータ・ビューから生じる同時生起頻度テーブルのさらに別の部分の図である。
【図３４】図２４、２５、および２６のデータ・ビューから生じる同時生起頻度テーブルのさらに別の部分の図である。
【図３５】図２４、２５、および２６のデータ・ビューから生じる同時生起頻度テーブルのさらに別の部分の図である。

【図３６】２つのキーワードを含むキーワード検索操作の結果として返されるコンテキストツリーのスキーマ・グラフの図である。
【図３７】記載した配置を実施できる汎用コンピュータの図式的な構成図である。
【図３８】図１９に示す単一のヒット・ノードを含む場合に対するコンテキストツリーを構築するための方法内部の下位プロセスのフローチャートである。
【図３９】図２０に示す複数のヒット・ノードを含む場合に対するコンテキストツリーを構築するための方法内部の下位プロセスのフローチャートである。

【図１】

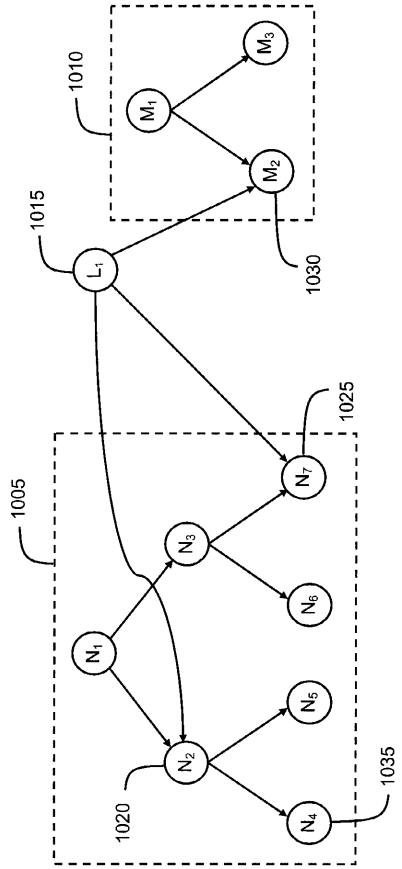

Fig. 1

【図２】

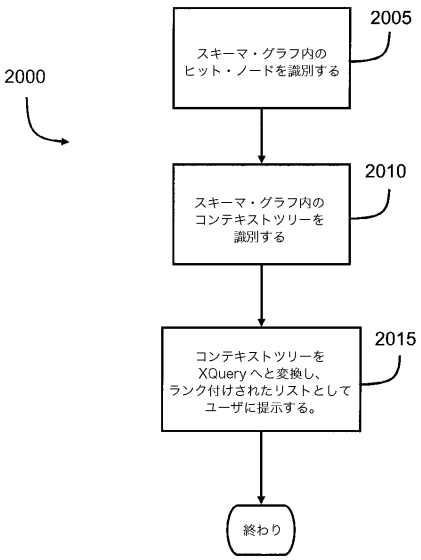

Fig. 2

【図３Ａ】

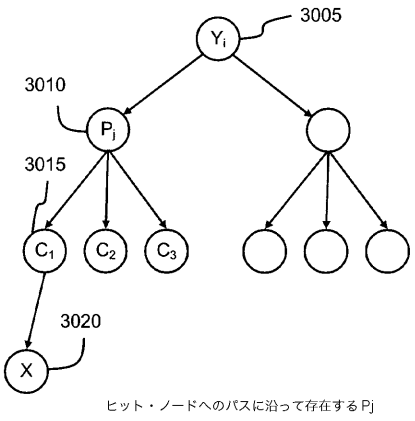

Fig. 3A

【図３Ｂ】

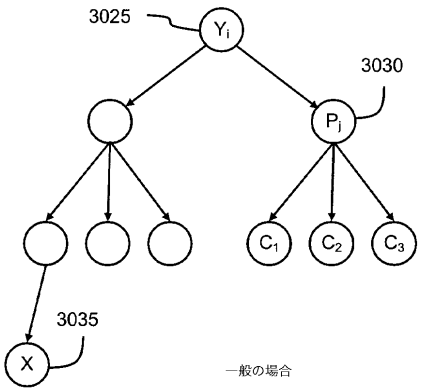

Fig. 3B

【図４】

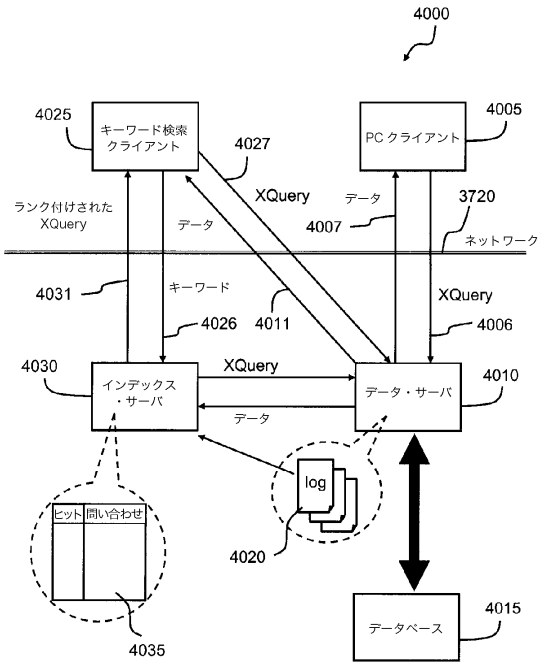

Fig. 4

【図５】

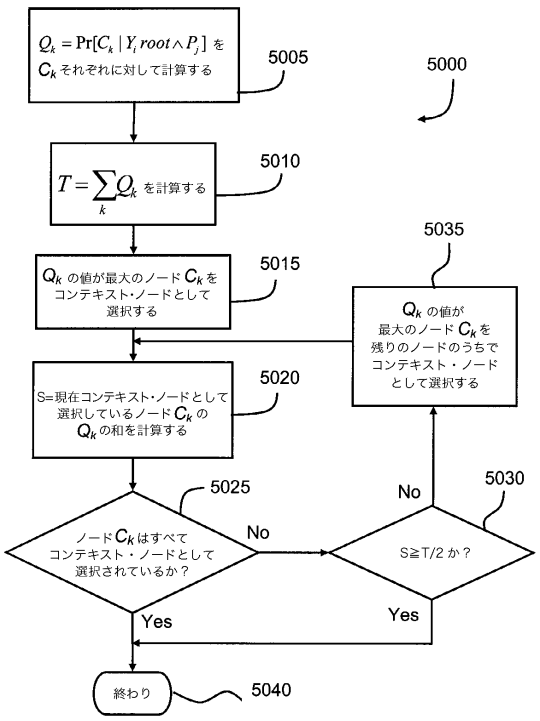

Fig. 5

【図６】

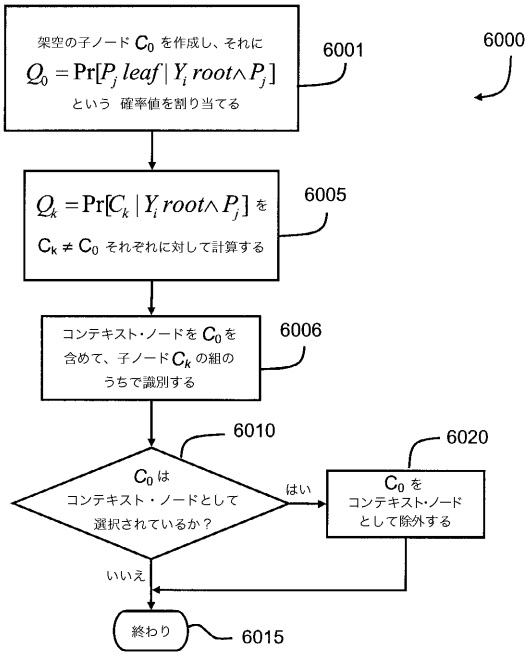

Fig. 6

【図７】

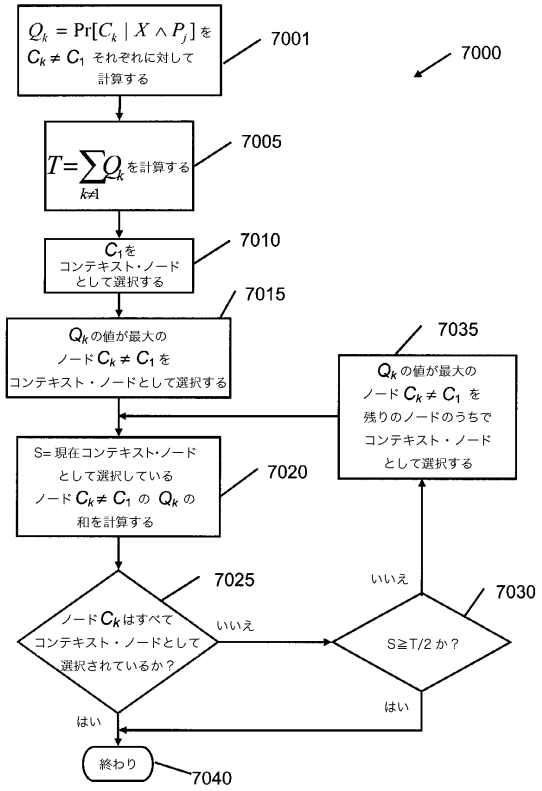

Fig. 7

【図８】

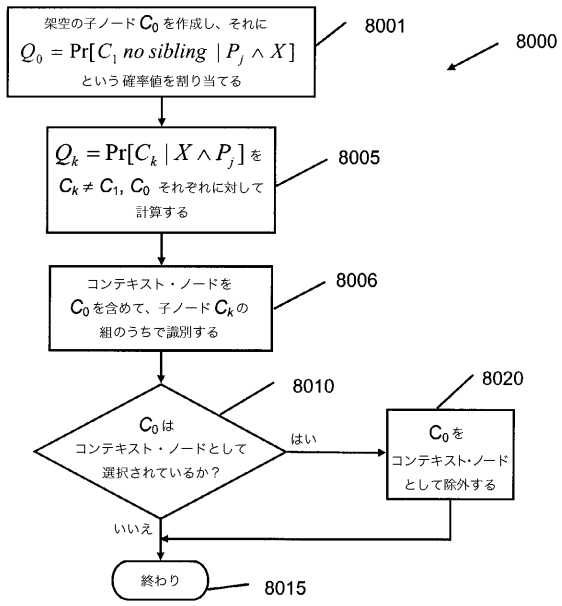

Fig. 8

【図９】

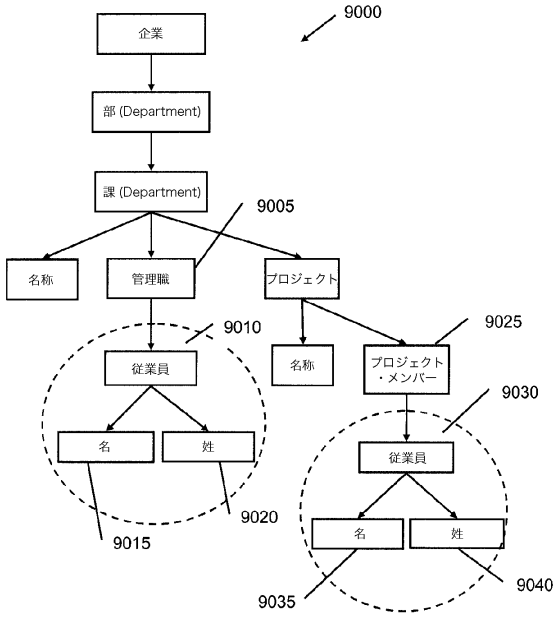

Fig. 9

【図１０】

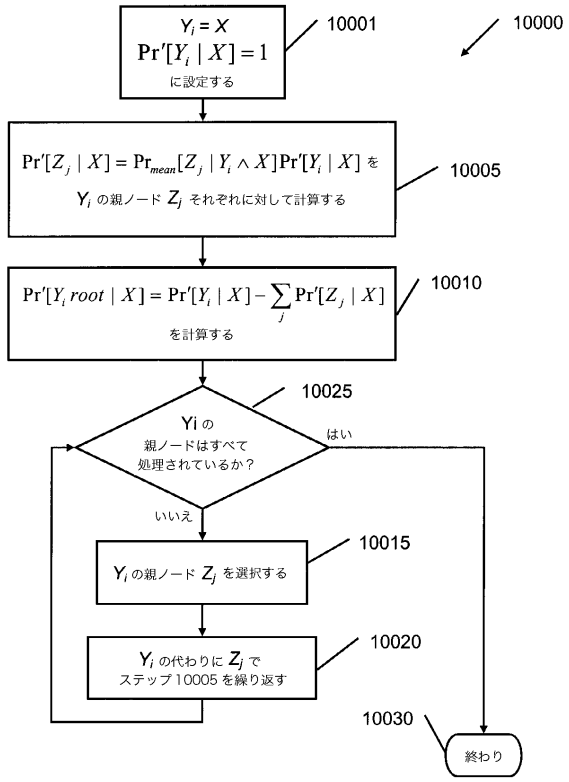

Fig. 10

【図１１】

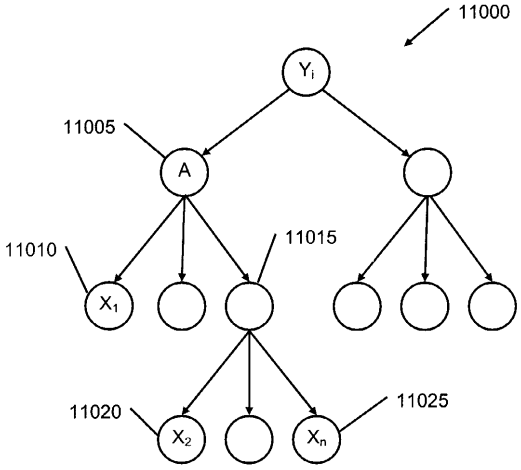

Fig. 11

【図１２】

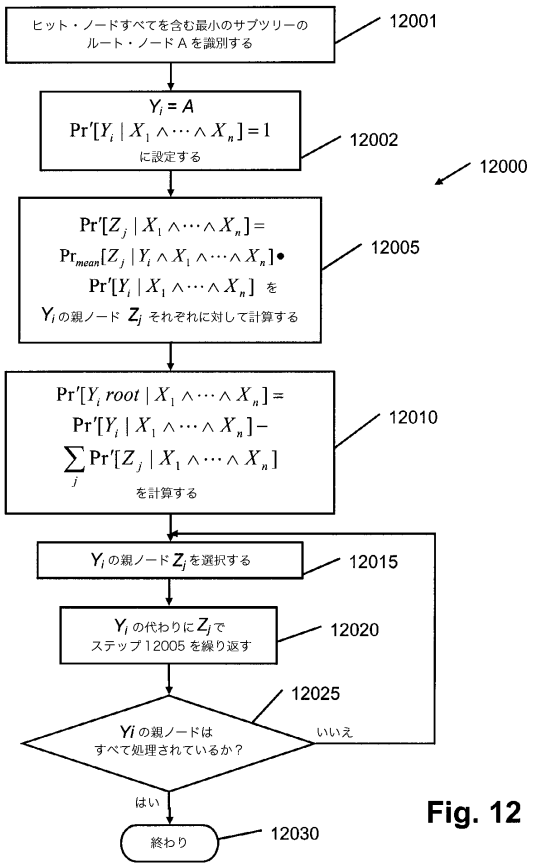

Fig. 12

【図１３】

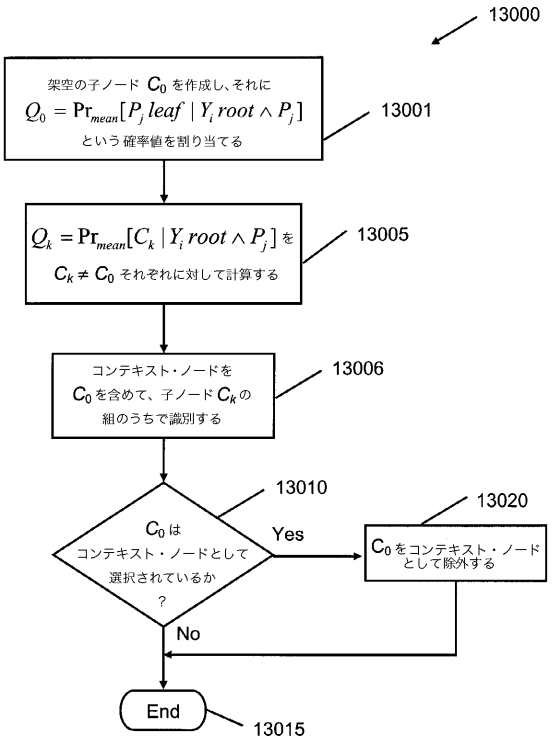

Fig. 13

【図１４】

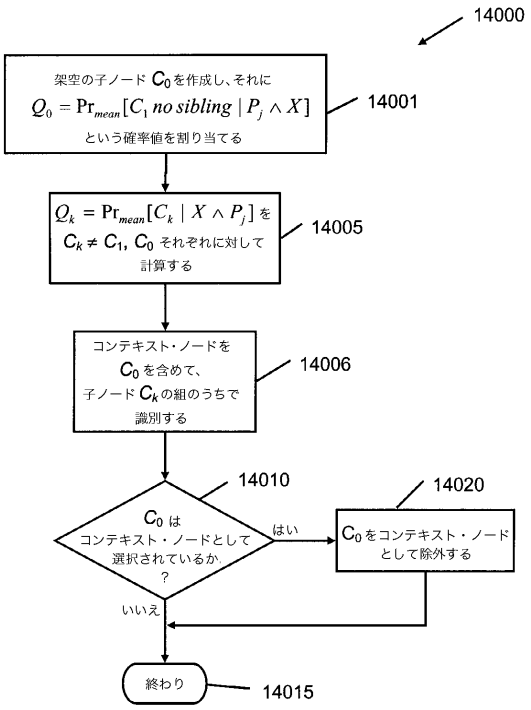

Fig. 14

【図１５】

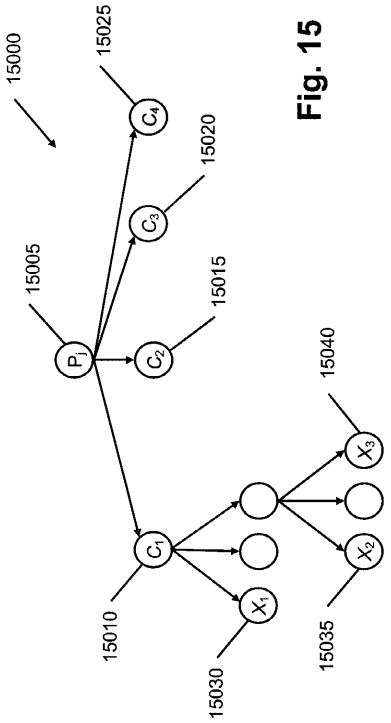

Fig. 15

【図１６】

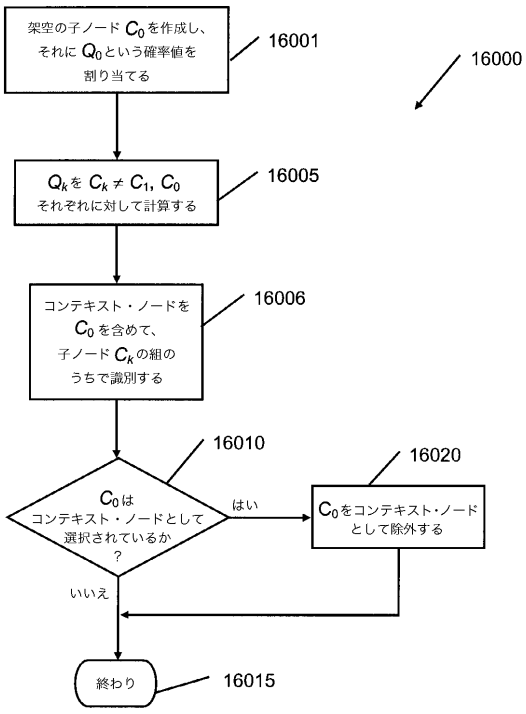

Fig. 16

【図１７】

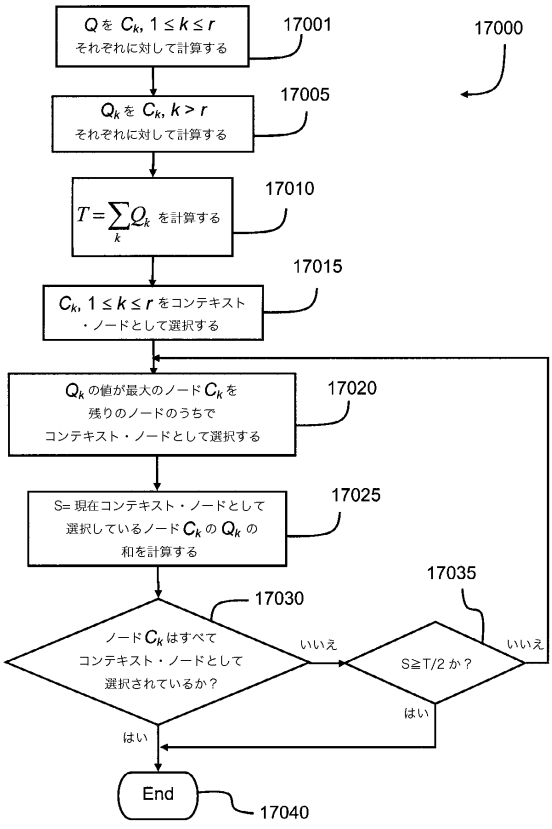

Fig. 17

【図１８】

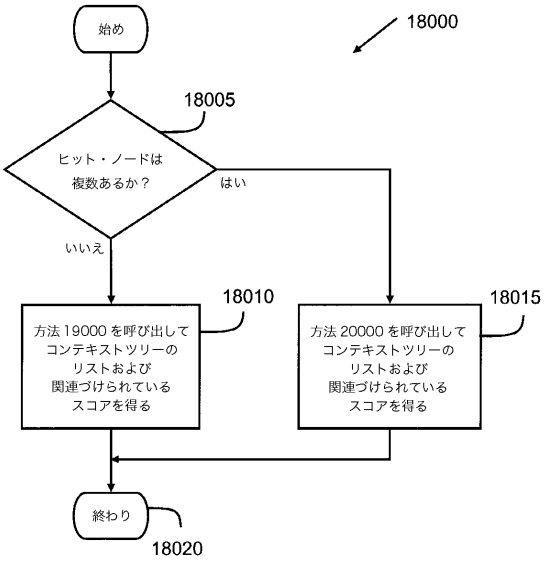

Fig. 18

【図１９】

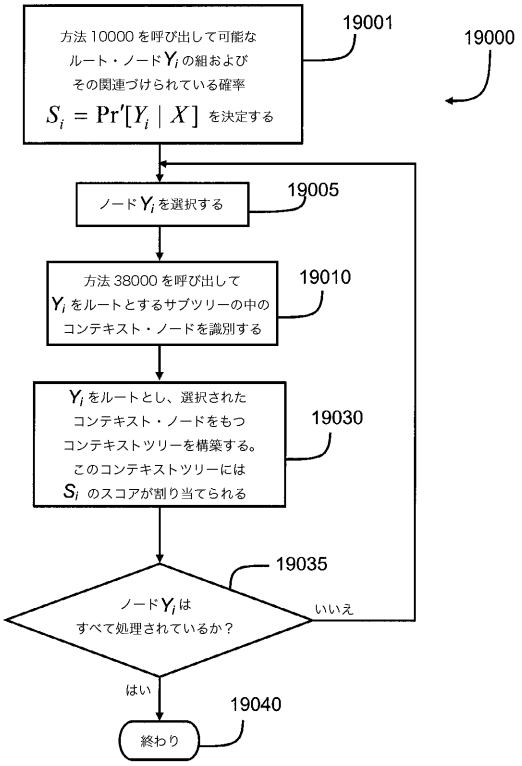

Fig. 19

【図２０】

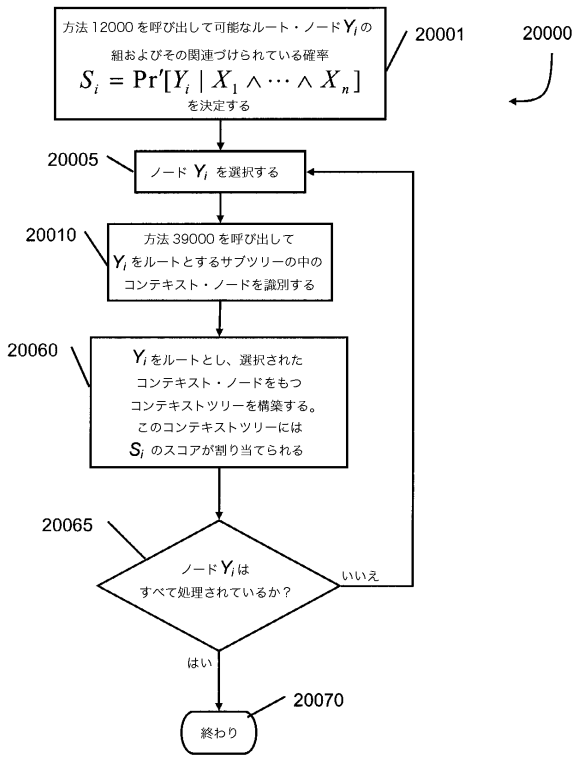

Fig. 20

【図２１】

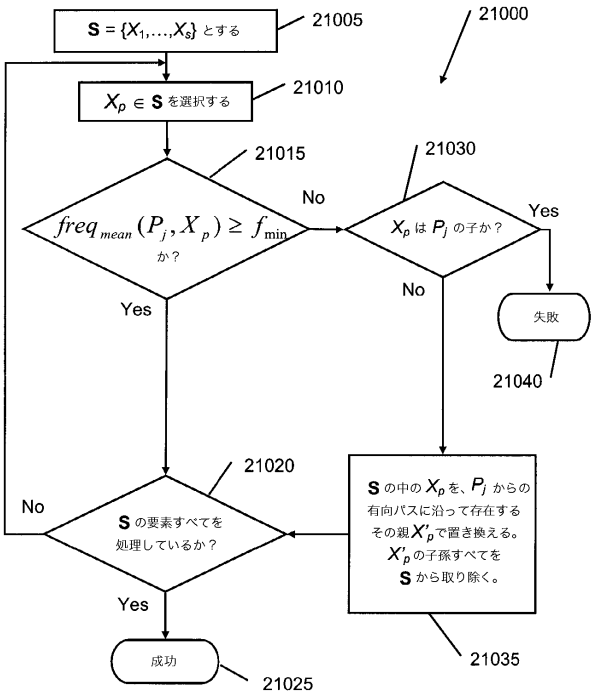

Fig. 21

【図２２】

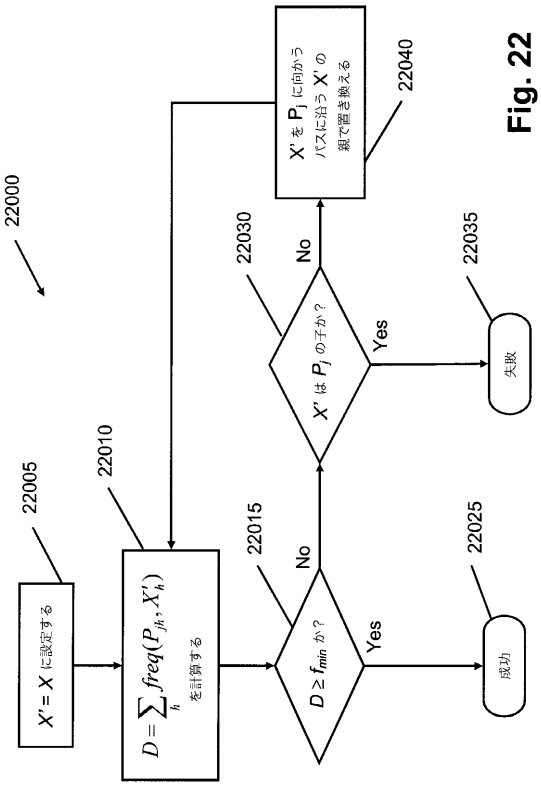

Fig. 22

【図２３】

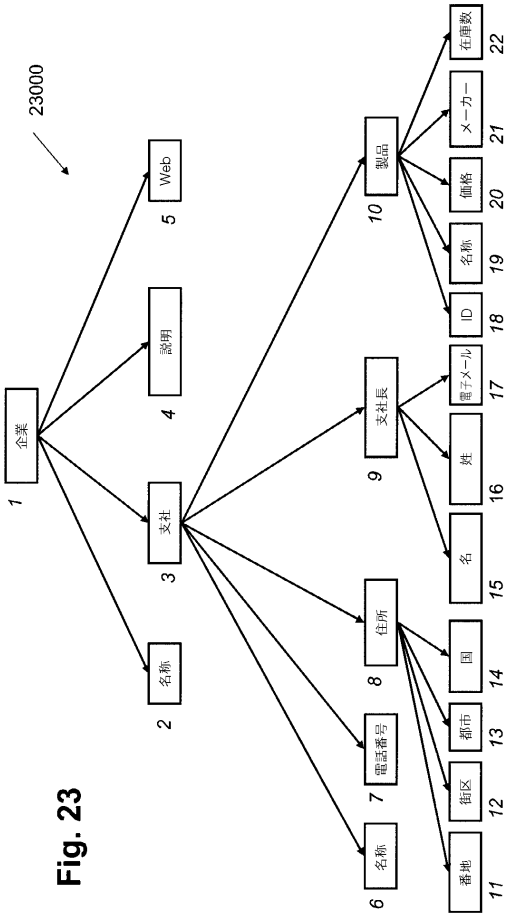

Fig. 23

【図２４】

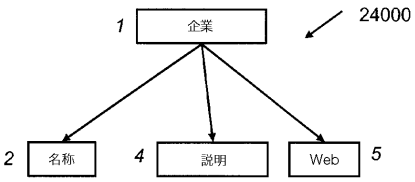

Fig. 24

【図２５】

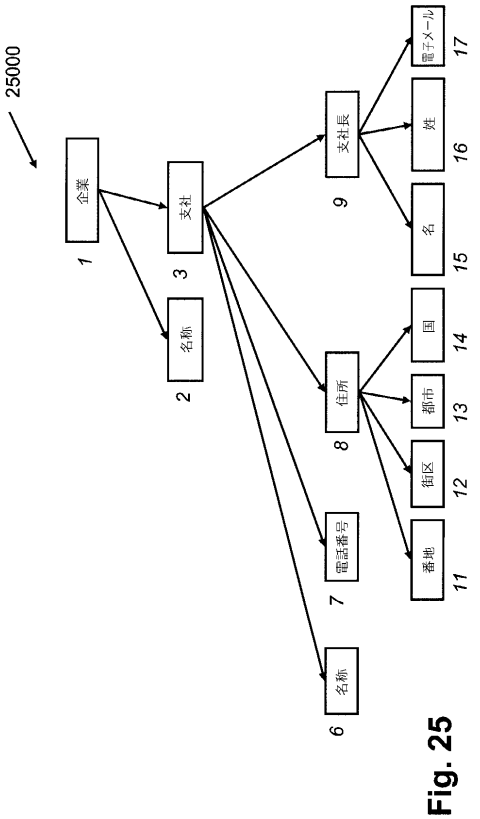

Fig. 25

【図２６】

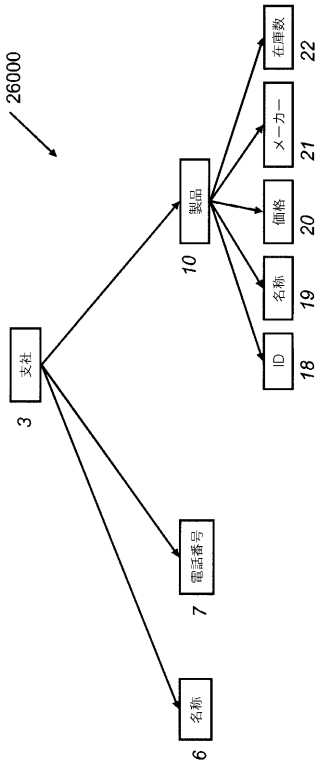

Fig. 26

【図２７】

X	freq(X)
1	2
2	2
3	2
4	1
5	1
6	2
7	2
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1

27000

Fig. 27

【図２８】

freq(Y_i, X)

28000

	Y_i				
X	1	3	8	9	10
2	2				
3	1				
4	1				
5	1				
6	1	2			
7	1	2			
8	1	1			
9	1	1			
10	0	1			
11	1	1	1		
12	1	1	1		
13	1	1	1		
14	1	1	1		
15	1	1		1	
16	1	1		1	
17	1	1		1	
18	0	1			1
19	0	1			1
20	0	1			1
21	0	1			1
22	0	1			1

28005

Fig. 28

【図２９】

freq(Y_i, P_j *leaf*)

	Y_i	
P_j	1	3
3	0	
8	0	0
9	0	0
10	0	0

29000

Fig. 29

【図３０】

freq(P_j *has 1 child*, X)

	P_j				
X	1	3	8	9	10
2	0				
3	0				
4	0				
5	0				
6	0	0			
7	0	0			
8	0	0			
9	0	0			
10	0	0			
11	0	0	0		
12	0	0	0		
13	0	0	0		
14	0	0	0		
15	0	0		0	
16	0	0		0	
17	0	0		0	
18	0	0			0
19	0	0			0
20	0	0			0
21	0	0			0
22	0	0			0

30000

Fig. 30

【図３１】

$freq(C_k, P_j, X)$: P_j = ノード1

X	C_k 2	3	4	5
2		1	1	1
3	1		0	0
4	1	0		1
5	1	0	1	
6	1		0	0
7	1		0	0
8	1		0	0
9	1		0	0
10	0		0	0
11	1		0	0
12	1		0	0
13	1		0	0
14	1		0	0
15	1		0	0
16	1		0	0
17	1		0	0
18	0		0	0
19	0		0	0
20	0		0	0
21	0		0	0
22	0		0	0

31000

Fig. 31

【図３２】

$freq(C_k, P_j, X)$: P_j = ノード3

X	C_k 6	7	8	9	10
6		2	1	1	1
7	2		1	1	1
8	1	1		1	0
9	1	1	1		0
10	1	1	0	0	
11	1	1		1	0
12	1	1		1	0
13	1	1		1	0
14	1	1		1	0
15	1	1	1		0
16	1	1	1		0
17	1	1	1		0
18	1	1	0	0	
19	1	1	0	0	
20	1	1	0	0	
21	1	1	0	0	
22	1	1	0	0	

32000

Fig. 32

【図３３】

$freq(C_k, P_j, X)$: P_j = ノード8

X	C_k 11	12	13	14
11		1	1	1
12	1		1	1
13	1	1		1
14	1	1	1	

33000

Fig. 33

【図３４】

$freq(C_k, P_j, X)$: P_j = ノード9

X	C_k 15	16	17
15		1	1
16	1		1
17	1	1	

34000

Fig. 34

【図３５】

$freq(C_k, P_j, X)$: P_j = ノード10

X	C_k 18	19	20	21	22
18		1	1	1	1
19	1		1	1	1
20	1	1		1	1
21	1	1	1		1
22	1	1	1	1	

35000

Fig. 35

【図３６】

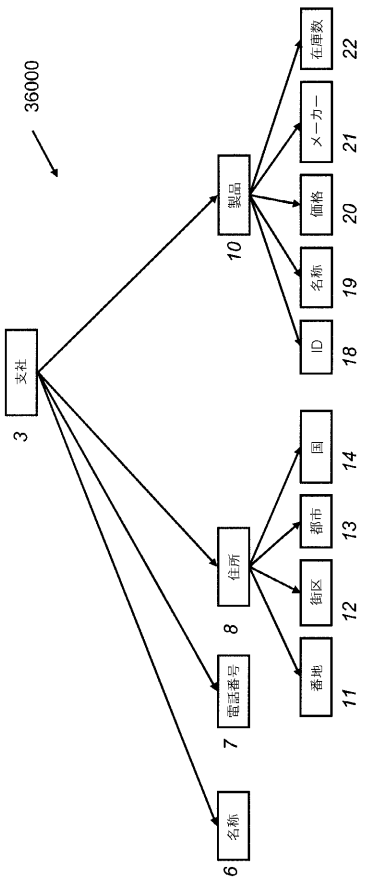

Fig. 36

【図３７】

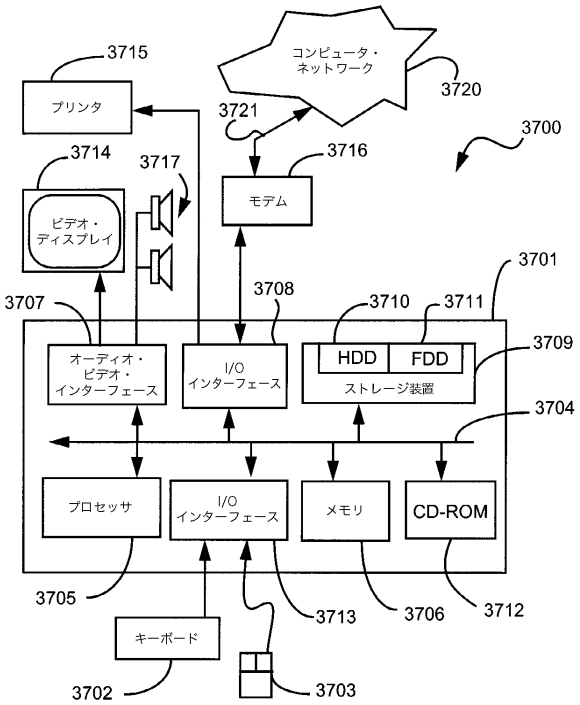

Fig. 37

【図３８】

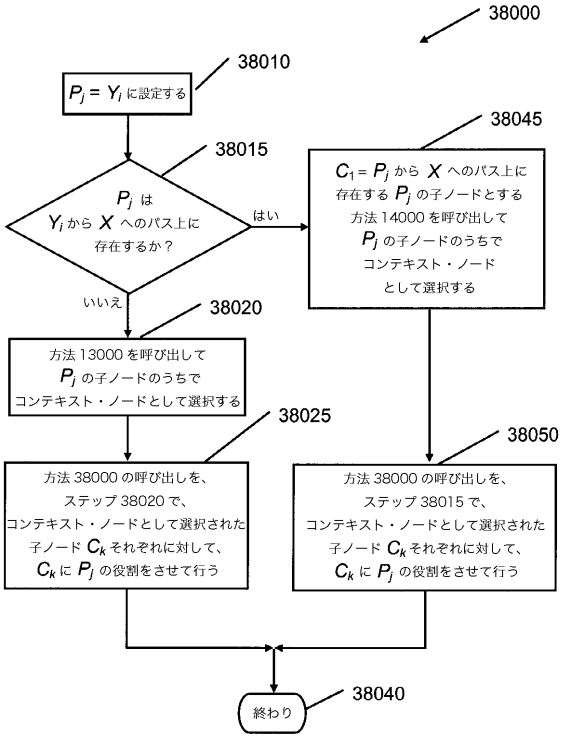

Fig. 38

【図３９】

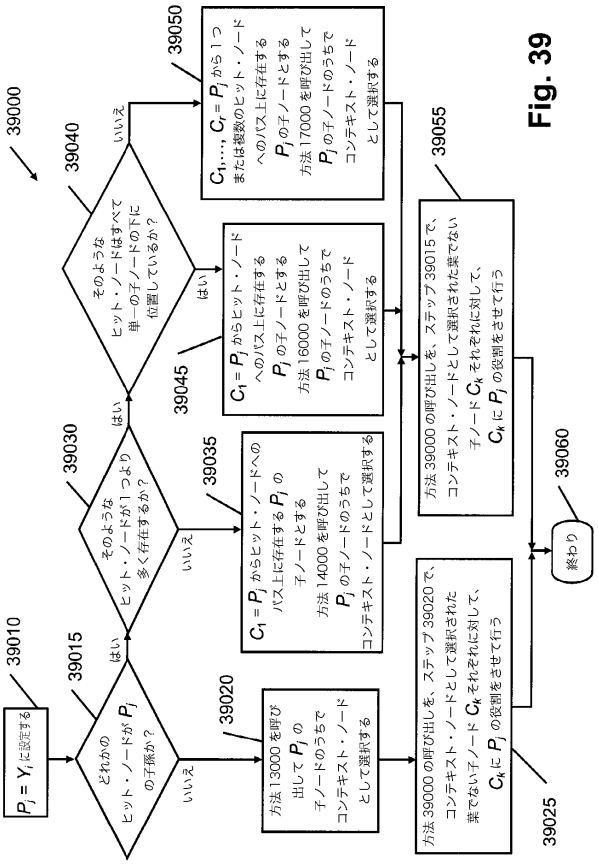

Fig. 39

フロントページの続き

(72)発明者 ドーン，カーン，フィ，ヴァン
オーストラリア国 ２１１３ ニュー サウス ウェールズ州，ノース ライド， トーマス ホルト ドライブ １ キヤノン インフォメーション システムズ リサーチ オーストラリア プロプライエタリー リミテツド 内
(72)発明者 レノン，アリソン，ジョアン
オーストラリア国 ２１１３ ニュー サウス ウェールズ州，ノース ライド， トーマス ホルト ドライブ １ キヤノン インフォメーション システムズ リサーチ オーストラリア プロプライエタリー リミテツド 内

審査官 紀田 馨

(56)参考文献 特開２００３－１６７８７９（ＪＰ，Ａ）
国際公開第０２／０２７５４４（WO，Ａ１）
国際公開第０１／０２４０４５（WO，Ａ１）
池田 ，基礎講座，日経コンピュータ ，日本，日経ＢＰ社 Nikkei Business Publications，Inc.，１９９８年 3月 2日，ｎｏ．４３８，第222頁乃至第227頁

(58)調査した分野(Int.Cl.，ＤＢ名)
JSTPlus/JMEDPlus/JST7580(JDreamII)
G06F 17/30