

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2020-501267  
(P2020-501267A)

(43) 公表日 令和2年1月16日(2020.1.16)

(51) Int.Cl.  
G16H 10/60 (2018.01)

F1  
G16H 10/60

テーマコード(参考)  
5L099

審査請求 未請求 予備審査請求 未請求 (全 30 頁)

(21) 出願番号 特願2019-530142 (P2019-530142)  
 (86) (22) 出願日 平成29年11月3日(2017.11.3)  
 (85) 翻訳文提出日 令和1年6月5日(2019.6.5)  
 (86) 国際出願番号 PCT/CN2017/109286  
 (87) 国際公開番号 WO2018/103479  
 (87) 国際公開日 平成30年6月14日(2018.6.14)  
 (31) 優先権主張番号 15/373,551  
 (32) 優先日 平成28年12月9日(2016.12.9)  
 (33) 優先権主張国・地域又は機関  
 米国(US)

(71) 出願人 390009531  
 インターナショナル・ビジネス・マシー  
 ズ・コーポレーション  
 INTERNATIONAL BUSIN  
 ESS MACHINES CORPOR  
 ATION  
 アメリカ合衆国10504 ニューヨーク  
 州 アーモンク ニュー オーチャード  
 ロード  
 New Orchard Road, A  
 rmonk, New York 105  
 04, United States o  
 f America  
 (74) 代理人 100108501  
 弁理士 上野 剛史

最終頁に続く

(54) 【発明の名称】 電子医療レコードからの自動的な知識ベースの特徴抽出

(57) 【要約】

【課題】電子医療レコード(EMR)データセットのスキーマにおける相違の問題がある。

【解決手段】臨床知識データのセットにしたがって、知識ツリーが構築される。EMRテーブルのセットに対応するEMRグラフが取得される。EMRグラフは、テーブル・ノードのセットと、属性ノードのセットとを含んでいる。テーブル・ノードのセットと、属性ノードのセットとは、EMRテーブルのセットにおける各EMRテーブルの構造と、EMRテーブルのセットの属性間の参照関係とを表す。知識ツリーおよびEMRグラフに基づいて、複数のサブクエリが生成される。知識ツリーにしたがって、複数のサブクエリを組み合わせることによって、少なくとも1つのクエリが生成される。

【選択図】図3

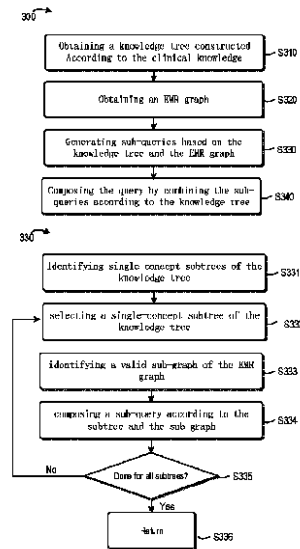


Fig.3

**【特許請求の範囲】****【請求項 1】**

臨床知識に基づいて、臨床的特徴を、電子医療レコード（EMR）テーブルのセットに抽出するためのクエリを生成するための、コンピュータによって実施される方法であって、

臨床知識データのセットにしたがって構築された知識ツリーを取得することと、

EMRテーブルのセットに対応するEMRグラフを取得することであって、前記EMRグラフは、テーブル・ノードのセットと、属性ノードのセットとを備え、前記テーブル・ノードのセットと、前記属性ノードのセットとは、前記EMRテーブルのセットにおける各EMRテーブルの構造と、EMRテーブルのセットの属性間の参照関係とを表す、前記取得することと、

前記知識ツリーおよび前記EMRグラフに基づいて複数のサブクエリを生成することと、

前記知識ツリーにしたがって前記複数のサブクエリを組み合わせることによって、少なくとも1つのクエリを構成することを含む、方法。

**【請求項 2】**

前記知識ツリーおよび前記EMRグラフに基づいて前記複数のサブクエリを生成することは、

前記知識ツリーの複数の単一概念のサブツリーを識別することと、

前記複数の単一概念のサブツリーの各単一概念のサブツリーごとに、

前記知識ツリーの単一概念のサブツリーを選択し、

前記単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、前記EMRグラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることによって、前記EMRグラフの有効なサブグラフを識別し、

前記単一概念のサブツリーおよび前記有効なサブグラフにしたがってサブクエリを構成することを含む、請求項1に記載の方法。

**【請求項 3】**

前記知識ツリーの前記単一概念のサブツリーは、概念ノードから始まりターゲット・ノードまでのパス内のすべてのノードと、概念ノードを備えていないノードの各々の任意の分岐とを備えたサブツリーである、請求項2に記載の方法。

**【請求項 4】**

前記EMRグラフの前記有効なサブグラフは、前記EMRグラフのサブグラフであり、前記知識ツリーのターゲット・ノードに一致するテーブル・ノードが、前記単一概念のサブツリーに存在し、前記単一概念のサブツリーにおける各属性ノードから、前記テーブル・ノードへの有向パスが存在する、請求項2に記載の方法。

**【請求項 5】**

前記単一概念のサブツリーの前記ターゲット・ノードおよび前記概念ノードをそれぞれ、前記EMRグラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることは、

属性によるマッピング、

属性値によるマッピング、および、

少なくとも1つの追加のフィルタを適用することによるマッピングのうちの少なくとも1つを含む、請求項2に記載の方法。

**【請求項 6】**

前記EMRテーブルのセットに対応する前記EMRグラフを取得することは、前記EMRテーブルのセットを、等価なグラフ表現に変換することを含む、請求項1に記載の方法。

**【請求項 7】**

前記少なくとも1つのクエリを使用して、前記EMRテーブルのセットから、1つまたは複数の臨床的特徴を抽出することをさらに含む、請求項1に記載の方法。

10

20

30

40

50

**【請求項 8】**

臨床知識に基づいて、臨床的特徴を、電子医療レコード（EMR）テーブルのセットに抽出するためのクエリを生成するためのデバイスであって、

少なくとも1つのプロセッサと、

前記少なくとも1つのプロセッサに動作可能に結合されたメモリと、

前記メモリに記憶され、

臨床知識データのセットにしたがって構築された知識ツリーを取得することと、

EMRテーブルのセットに対応するEMRグラフを取得することであって、前記EMRグラフは、テーブル・ノードのセットと、属性ノードのセットとを備え、前記テーブル・ノードのセットと、前記属性ノードのセットとは、前記EMRテーブルのセットにおける各EMRテーブルの構造と、EMRテーブルのセットの属性間の参照関係とを表す、前記取得することと、

前記知識ツリーおよび前記EMRグラフに基づいて複数のサブクエリを生成することと、

前記知識ツリーにしたがって前記複数のサブクエリを組み合わせることによって、少なくとも1つのクエリを構成することと

からなる動作を実行するために、前記少なくとも1つのプロセッサによって実行されるコンピュータ・プログラム命令のセットとを備えた、デバイス。

**【請求項 9】**

前記知識ツリーおよび前記EMRグラフに基づいて前記複数のサブクエリを生成することは、

前記知識ツリーの複数の単一概念のサブツリーを識別することと、

前記複数の単一概念のサブツリーの各単一概念のサブツリーごとに、

前記知識ツリーの単一概念のサブツリーを選択し、

前記単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、前記EMRグラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることによって、前記EMRグラフの有効なサブグラフを識別し、

前記単一概念のサブツリーおよび前記有効なサブグラフにしたがってサブクエリを構成することを含む、請求項8に記載のデバイス。

**【請求項 10】**

前記知識ツリーの前記単一概念のサブツリーは、概念ノードから始まりターゲット・ノードまでのパス内のすべてのノードと、概念ノードを備えていないノードの各々の任意の分岐とを備えたサブツリーである、請求項9に記載のデバイス。

**【請求項 11】**

前記EMRグラフの前記有効なサブグラフは、前記EMRグラフのサブグラフであり、前記知識ツリーのターゲット・ノードに一致するテーブル・ノードが、前記単一概念のサブツリーに存在し、前記単一概念のサブツリーにおける各属性ノードから、前記テーブル・ノードへの有向パスが存在する、請求項9に記載のデバイス。

**【請求項 12】**

前記単一概念のサブツリーの前記ターゲット・ノードおよび前記概念ノードをそれぞれ、前記EMRグラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることは、

属性によるマッピング、

属性値によるマッピング、および、

少なくとも1つの追加のフィルタを適用することによるマッピングのうちの少なくとも1つを含む、請求項9に記載のデバイス。

**【請求項 13】**

前記EMRテーブルのセットに対応する前記EMRグラフを取得することは、前記EMRテーブルのセットを、等価なグラフ表現に変換することを含む、請求項8に記載のデバイス。

10

20

30

40

50

## 【請求項 14】

前記動作はさらに、前記少なくとも1つのクエリを使用して、前記EMRテーブルのセットから、1つまたは複数の臨床的特徴を抽出することを含む、請求項8に記載のデバイス。

## 【請求項 15】

臨床知識に基づいて、臨床的特徴を、電子医療レコード(EMR)テーブルのセットに抽出するためのクエリを生成するためのコンピュータ・プログラム製品であって、前記コンピュータ・プログラム製品は、具現化されたプログラム命令を有するコンピュータ可読記憶媒体を備え、前記プログラム命令は、

臨床知識データのセットにしたがって構築された知識ツリーを取得することと、

EMRテーブルのセットに対応するEMRグラフを取得することであって、前記EMRグラフは、テーブル・ノードのセットと、属性ノードのセットとを備え、前記テーブル・ノードのセットと、前記属性ノードのセットとは、前記EMRテーブルのセットにおける各EMRテーブルの構造と、EMRテーブルのセットの属性間の参照関係とを表す、前記取得することと、

前記知識ツリーおよび前記EMRグラフに基づいて複数のサブクエリを生成することと

、  
前記知識ツリーにしたがって前記複数のサブクエリを組み合わせることによって、少なくとも1つのクエリを構成することを含む方法を少なくとも1つのプロセッサに実行させるために、前記少なくとも1つのプロセッサによって実行可能である、コンピュータ・プログラム製品。

## 【請求項 16】

前記知識ツリーおよび前記EMRグラフに基づいて前記複数のサブクエリを生成することは、

前記知識ツリーの複数の単一概念のサブツリーを識別することと、

前記複数の単一概念のサブツリーの各単一概念のサブツリーごとに、

前記知識ツリーの単一概念のサブツリーを選択し、

前記単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、前記EMRグラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることによって、前記EMRグラフの有効なサブグラフを識別し、

前記単一概念のサブツリーおよび前記有効なサブグラフにしたがってサブクエリを構成することを含む、請求項15に記載のコンピュータ・プログラム製品。

## 【請求項 17】

前記知識ツリーの前記単一概念のサブツリーは、概念ノードから始まりターゲット・ノードまでのパス内のすべてのノードと、概念ノードを備えていないノードの各々の任意の分岐とを備えたサブツリーである、請求項16に記載のコンピュータ・プログラム製品。

## 【請求項 18】

前記EMRグラフの前記有効なサブグラフは、前記EMRグラフのサブグラフであり、前記知識ツリーのターゲット・ノードに一致するテーブル・ノードが、前記単一概念のサブツリーに存在し、前記単一概念のサブツリーにおける各属性ノードから、前記テーブル・ノードへの有向パスが存在する、請求項16に記載のコンピュータ・プログラム製品。

## 【請求項 19】

前記単一概念のサブツリーの前記ターゲット・ノードおよび前記概念ノードをそれぞれ、前記EMRグラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることは、

属性によるマッピング、

属性値によるマッピング、および、

少なくとも1つの追加のフィルタを適用することによるマッピングのうちの少なくとも1つを含む、請求項16に記載のコンピュータ・プログラム製品。

## 【請求項 20】

10

20

30

40

50

前記 E M R テーブルのセットに対応する前記 E M R グラフを取得することは、前記 E M R テーブルのセットを、等価なグラフ表現に変換することを含む、請求項 1 5 に記載のコンピュータ・プログラム製品。

【請求項 2 1】

前記方法はさらに、前記少なくとも 1 つのクエリを使用して、前記 E M R テーブルのセットから、1 つまたは複数の臨床的特徴を抽出することを含む、請求項 1 5 に記載のコンピュータ・プログラム製品。

【請求項 2 2】

臨床知識に基づいて、臨床的特徴を、電子医療レコード ( E M R ) テーブルのセットに抽出するためのクエリを生成するための、コンピュータによって実施される方法であって、

10

臨床知識データのセットにしたがって構築された知識ツリーを取得することと、

E M R テーブルのセットを、等価なグラフ表現に変換する、前記 E M R テーブルのセットに対応する E M R グラフを取得することであって、前記 E M R グラフは、テーブル・ノードのセットと、属性ノードのセットとを備え、前記テーブル・ノードのセットと、前記属性ノードのセットとは、前記 E M R テーブルのセットにおける各 E M R テーブルの構造と、E M R テーブルのセットの属性間の参照関係とを表す、前記取得することと、

前記知識ツリーの複数の単一概念のサブツリーを識別することに基づいて、前記知識ツリーおよび前記 E M R グラフに基づいて複数のサブクエリを生成することと、

前記知識ツリーにしたがって前記複数のサブクエリを組み合わせることによって、少なくとも 1 つのクエリを構成することと、

20

前記 E M R テーブルのセットを、等価なグラフ表現へ変換することを含む、方法。

【請求項 2 3】

前記知識ツリーおよび前記 E M R グラフに基づいて前記複数のサブクエリを生成することはさらに、

前記複数の単一概念のサブツリーの各単一概念のサブツリーごとに、

前記知識ツリーの単一概念のサブツリーを選択し、

前記単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、前記 E M R グラフのテーブル・ノードおよび少なくとも 1 つの属性ノードにマッピングすることによって、前記 E M R グラフの有効なサブグラフを識別し、

30

前記単一概念のサブツリーおよび前記有効なサブグラフにしたがってサブクエリを構成することを含む、請求項 2 2 に記載の方法。

【請求項 2 4】

臨床知識に基づいて、臨床的特徴を、電子医療レコード ( E M R ) テーブルのセットに抽出するためのクエリを生成するためのコンピュータ・プログラム製品であって、前記コンピュータ・プログラム製品は、具現化されたプログラム命令を有するコンピュータ可読記憶媒体を備え、前記プログラム命令は、

臨床知識データのセットにしたがって構築された知識ツリーを取得することと、

E M R テーブルのセットを、等価なグラフ表現に変換する、前記 E M R テーブルのセットに対応する E M R グラフを取得することであって、前記 E M R グラフは、テーブル・ノードのセットと、属性ノードのセットとを備え、前記テーブル・ノードのセットと、前記属性ノードのセットとは、前記 E M R テーブルのセットにおける各 E M R テーブルの構造と、E M R テーブルのセットの属性間の参照関係とを表す、前記取得することと、

40

前記知識ツリーの複数の単一概念のサブツリーを識別することに基づいて、前記知識ツリーおよび前記 E M R グラフに基づいて複数のサブクエリを生成することと、

前記知識ツリーにしたがって前記複数のサブクエリを組み合わせることによって、少なくとも 1 つのクエリを構成することと、

前記 E M R テーブルのセットを、等価なグラフ表現へ変換することを含む方法を少なくとも 1 つのプロセッサに実行させるために、前記少なくとも 1 つのプロセッサによって実行可能である、コンピュータ・プログラム製品。

50

**【請求項 25】**

前記知識ツリーおよび前記 E M R グラフに基づいて前記複数のサブクエリを生成することはさらに、

前記複数の単一概念のサブツリーの各単一概念のサブツリーごとに、

前記知識ツリーの単一概念のサブツリーを選択し、

前記単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、前記 E M R グラフのテーブル・ノードおよび少なくとも 1 つの属性ノードにマッピングすることによって、前記 E M R グラフの有効なサブグラフを識別し、

前記単一概念のサブツリーおよび前記有効なサブグラフにしたがってサブクエリを構成することを含む、請求項 24 に記載のコンピュータ・プログラム製品。

10

**【発明の詳細な説明】****【技術分野】****【0001】**

本開示は、一般にデータベースの分野に関し、より詳細には、臨床的特徴を抽出するために電子医療レコードに問い合わせるための問合せメッセージを生成するための方法、システム、およびコンピュータ・プログラムに関する。

**【背景技術】****【0002】**

今日、電子医療レコード ( E M R ) システムにおける臨床情報の量は急速に増加している。臨床情報は通常、リレーショナル・データセットの形式で記憶される。異なる病院または医療機関は、異なるスキーマを使用して、 E M R データを E M R データセットまたは E M R テーブルに記憶し得る。 E M R データからの特徴抽出 / 導出には普遍的な要件がある。臨床データセット、特徴抽出、および特徴構築に関する記述的および予測的分析は、通常、労働集約的である。したがって、医療専門家は一般に、 E M R データを使用して、臨床知識を効率的に抽出または評価することができない。異なるソースに対する E M R データセットのスキーマは変動し、検索システムにおいて情報を取得するために、 E M R データセットに問い合わせるために問合せメッセージが生成され得るように、統一されたスキーマに変更される必要がある。これは労働集約的な処理であり、ロバストではない。異なるスキーマのこれらの E M R データセットを統合し、これらを、臨床的特徴を検索するために使用することが望ましいであろう。異なるスキーマのこれらの E M R データセットを統合し、これらを、臨床的特徴を検索するために使用することが望ましいであろう。

20

30

**【発明の概要】****【発明が解決しようとする課題】****【0003】**

E M R データセットのスキーマにおける相違の問題がある。

**【課題を解決するための手段】****【0004】**

E M R データセットのスキーマにおける相違の問題に対処するために、1つまたは複数の実施形態は、任意のリレーショナル・スキーマ内の E M R データセットから臨床知識を自動的に抽出する、または臨床規則を評価する、あるいはその両方のための技術を提案する。1つまたは複数の実施形態は、 E M R データセットから臨床的特徴を抽出するためのクエリを生成するための方法、システム、およびコンピュータ・プログラムを提案する。

40

**【0005】**

1つの実施形態では、臨床知識に基づいて、臨床的特徴を、電子医療レコード ( E M R ) テーブルのセットに抽出するためのクエリを生成するための方法が提供される。方法は、臨床知識データのセットにしたがって構築された知識ツリーを取得することを含んでいる。 E M R テーブルのセットに対応する E M R グラフが取得される。 E M R グラフは、テーブル・ノードのセットと、属性ノードのセットとを備えている。テーブル・ノードのセットと、属性ノードのセットとは、 E M R テーブルのセットにおける各 E M R テーブルの構造と、 E M R テーブルのセットの属性間の参照関係とを表す。知識ツリーおよび E M

50

R グラフに基づいて、複数のサブクエリが生成される。知識ツリーにしたがって複数のサブクエリを組み合わせることによって、少なくとも1つのクエリが生成される。

【0006】

別の実施形態では、臨床知識に基づいて、臨床的特徴を電子医療レコード（EMR）テーブルのセットに抽出するためのクエリを生成するためのデバイスが提供される。デバイスは、少なくとも1つのプロセッサと、少なくとも1つのプロセッサに動作可能に結合されたメモリとを備えている。コンピュータ・プログラム命令のセットがメモリに記憶され、複数の動作を実行するために、少なくとも1つのプロセッサによって実行される。複数の動作は、臨床知識データのセットにしたがって構築された知識ツリーを取得することを備えている。EMRテーブルのセットに対応するEMRグラフが取得される。EMRグラフは、テーブル・ノードのセットと、属性ノードのセットとを備えている。テーブル・ノードのセットおよび属性ノードのセットは、EMRテーブルのセットにおける各EMRテーブルの構造、およびEMRテーブルのセットの属性間の参照関係を表す。知識ツリーおよびEMRグラフに基づいて、複数のサブクエリが生成される。知識ツリーにしたがって複数のサブクエリを組み合わせることによって、少なくとも1つのクエリが生成される。

10

【0007】

さらなる実施形態では、臨床知識に基づいて、臨床的特徴を、電子医療レコード（EMR）テーブルのセットに抽出するためのクエリを生成するためのコンピュータ・プログラム製品が提供される。コンピュータ・プログラム製品は、具現化されたプログラム命令を有するコンピュータ可読記憶媒体を備えている。プログラム命令は、少なくとも1つのプロセッサに対して、方法を実行させるために、少なくとも1つのプロセッサによって実行可能である。方法は、臨床知識データのセットにしたがって構築された知識ツリーを取得することを含んでいる。EMRテーブルのセットに対応するEMRグラフが取得される。EMRグラフは、テーブル・ノードのセットと、属性ノードのセットとを備えている。テーブル・ノードのセットと、属性ノードのセットとは、EMRテーブルのセットにおける各EMRテーブルの構造と、EMRテーブルのセットの属性間の参照関係とを表す。知識ツリーおよびEMRグラフに基づいて、複数のサブクエリが生成される。知識ツリーにしたがって複数のサブクエリを組み合わせることによって、少なくとも1つのクエリが生成される。

20

【0008】

1つまたは複数の実施形態は、任意のリレーショナル・スキーマにおけるEMRデータセットからの自動的な知識ベースの特徴抽出のための斬新な解決策を提供する。臨床データ分析の分野における既存のアプローチと比較して、実施形態は、知識ベースの特徴エンジニアリングに必要な人間の労力を著しく低減させ得、そのため、文献およびヘテロジニアスなEMRデータセットにおける臨床知識におけるデータ・マイニングのために非常に役立つ。

30

【図面の簡単な説明】

【0009】

【図1】本開示の1つの実施形態にしたがう例示的なコンピュータ・システムを図示する図である。

40

【図2】本開示の1つの実施形態にしたがうテーブル形式のEMRデータセットの例を例示的に図示する図である。

【図3】本開示の1つの実施形態にしたがってEMRデータセットにクエリを生成するためのコンピュータで実施される方法を例示するフローチャートである。

【図4】本開示の1つの実施形態にしたがう知識ツリーを例示的に図示する図である。

【図5】本開示の1つの実施形態にしたがう知識ツリーの3つの単一概念のサブツリーを例示的に図示する図である。

【図6】本開示の1つの実施形態にしたがって図2のEMRデータセットから変換されたEMRグラフを例示的に図示する図である。

【図7】本開示の1つの実施形態にしたがって、知識ツリーの要素をEMRグラフの要素

50

にマッピングすることによって、サブクエリを生成し、最終クエリを構成する処理を例示的に図示する図である。

【図 8】本開示の 1 つの実施形態にしたがって、知識ツリーの要素を E M R グラフの要素にマッピングすることによって、サブクエリを生成し、最終クエリを構成する処理を例示的に図示する図である。

【図 9】本開示の 1 つの実施形態にしたがって、知識ツリーの要素を E M R グラフの要素にマッピングすることによって、サブクエリを生成し、最終クエリを構成する処理を例示的に図示する図である。

【発明を実施するための形態】

【 0 0 1 0 】

10

必要に応じて、詳細な実施形態が本明細書に開示されているが、開示された実施形態は単なる例であり、以下に説明されるシステムおよび方法は、様々な形式で具現化され得ることが理解されるべきである。したがって、本明細書に開示された特定の構造的および機能的な詳細は、限定としてではなく、特許請求の範囲の基礎として、および実質的に適切に詳細な構造および機能において本主題を様々な適用するように当業者に教示するための代表的な基礎として解釈されるべきである。さらに、本明細書で使用される用語およびフレーズは、限定することではなくむしろ、概念の理解可能な説明を提供することが意図されている。

【 0 0 1 1 】

20

ここで図 1 を参照して示すように、1 つまたは複数の実施形態に適用可能なコンピュータ・システム / サーバ 1 2 の一例が図示されている。コンピュータ・システム / サーバ 1 2 は単に例示的なものであり、本明細書で論じられる様々な実施形態の使用または機能の範囲に関していかなる限定を示唆することも意図されていない。少なくともいくつかの事例では、1 つまたは複数の実施形態が、コンピュータ・システム / サーバ 1 2 内で実施される。図 1 は、汎用コンピューティング・デバイスの形式のコンピュータ・システム / サーバ 1 2 を図示している。コンピュータ・システム / サーバ 1 2 の構成要素は、1 つまたは複数のプロセッサまたは処理ユニット 1 6、システム・メモリ 2 8、およびシステム・メモリ 2 8 を含む様々なシステム構成要素をプロセッサ 1 6 に結合するバス 1 8 を含む得るが、これらに限定されない。バス 1 8 は、メモリ・バスまたはメモリ・コントローラ、周辺バス、加速グラフィック・ポート、および様々なバス・アーキテクチャのいずれかを使用するプロセッサまたはローカル・バスを含む、いくつかのタイプのバス構造のうちのいずれかの 1 つまたは複数を表す。限定ではなく例として、このようなアーキテクチャは、業界標準アーキテクチャ ( I S A ) バス、Micro Channel Architecture ( M C A ) バス、Enhanced ISA ( E I S A ) バス、Video Electronics Standards Association ( V E S A ) ローカル・バス、および Peripheral Component Interconnect ( P C I ) バスを含んでいる。

30

【 0 0 1 2 】

コンピュータ・システム / サーバ 1 2 は通常、様々なコンピュータ・システム可読媒体を含んでいる。このような媒体は、コンピュータ・システム / サーバ 1 2 によってアクセス可能な任意の利用可能な媒体であり得、揮発性と不揮発性の媒体、リムーバブルと非リムーバブルの媒体の両方を含んでいる。システム・メモリ 2 8 は、ランダム・アクセス・メモリ ( R A M ) 3 0、またはキャッシュ・メモリ 3 2、あるいはその両方のような揮発性メモリの形態のコンピュータ・システム可読媒体を含み得る。コンピュータ・システム / サーバ 1 2 はさらに、他のリムーバブル / 非リムーバブル、揮発性 / 不揮発性のコンピュータ・システム記憶媒体を含み得る。単なる例として、記憶システム 3 4 は、非リムーバブルな不揮発性の磁気媒体 ( 図示せず、通常「ハードドライブ」と呼ばれる ) との間での読み書きのために提供され得る。図示されていないが、リムーバブルな不揮発性の磁気ディスク ( たとえば「フロッピー ( R ) ディスク」 ) との間での読み書きのための磁気ディスク・ドライブ、および C D - R O M、D V D - R O M、または他の光学媒体のような

40

50



リムーバブルな不揮発性の光ディスクとの間での読み書きのための光ディスク・ドライブが提供され得る。このような事例では、各々は、1つまたは複数のデータ媒体インターフェースによってバス18に接続され得る。以下にさらに描写され論じられるように、メモリ28は、1つまたは複数の実施形態の機能を実行するように構成されたプログラム・モジュールのセット（たとえば、少なくとも1つ）を有する少なくとも1つのプログラム製品を含み得る。

【0013】

プログラム・モジュール42のセット（少なくとも1つ）を有するプログラム/ユーティリティ40は、限定ではなく例として、オペレーティング・システム、1つまたは複数のアプリケーション・プログラム、他のプログラム・モジュール、およびプログラム・データと同様に、メモリ28に記憶され得る。オペレーティング・システム、1つまたは複数のアプリケーション・プログラム、他のプログラム・モジュール、およびプログラム・データ、あるいはそれらの何らかの組合せの各々は、ネットワーキング環境の実施を含み得る。プログラム・モジュール42は、一般に、本明細書で論じた1つまたは複数の実施形態の機能または方法、あるいはその両方を実行する。

10

【0014】

コンピュータ・システム/サーバ12はまた、キーボード、ポインティング・デバイス、ディスプレイ24等のような1つまたは複数の外部デバイス14、ユーザがコンピュータ・システム/サーバ12とインタラクトすることを可能にする1つまたは複数のデバイス、またはコンピュータ・システム/サーバ12が、1つまたは複数の他のコンピューティング・デバイスと通信することを可能にする任意のデバイス（たとえば、ネットワーク・カード、モデム等）、あるいはその組合せと通信し得る。このような通信は、入力/出力(I/O)インターフェース22を介して生じ得る。さらにまた、コンピュータ・システム/サーバ12は、ネットワーク・アダプタ20を介して、ローカル・エリア・ネットワーク(LAN)、一般的な広域ネットワーク(WAN)、または公衆ネットワーク（たとえば、インターネット）、あるいはその組合せのような1つまたは複数のネットワークと通信し得る。描写されるように、ネットワーク・アダプタ20は、バス18を介してコンピュータ・システム/サーバ12の他の構成要素と通信する。図示されていないが、他のハードウェア構成要素、またはソフトウェア構成要素、あるいはその両方が、コンピュータ・システム/サーバ12と連携して使用され得ることが理解されるべきである。例として、マイクロコード、デバイス・ドライバ、冗長処理ユニット、外部ディスク・ドライブ・アレイ、RAIDシステム、テープ・ドライブ、およびデータ・アーカイブ記憶システム等を含むが、これらに限定されない。

20

30

【0015】

本明細書で論じられる様々な実施形態をよりよく理解するために、何らかの関連知識を紹介し、この開示の文脈で電子医療レコードおよび臨床知識の紹介を提供することが有利であろう。1つの実施形態では、クエリは、臨床知識に基づいて、任意のフォーマルなスキーマで、電子医療レコードに自動的に構成される。次いで、クエリは、電子医療レコードから臨床的特徴を抽出するために使用され得る。

【0016】

I. 電子医療レコード

40

【0017】

電子医療レコード(EMR)は、患者および住民の病歴、または健康情報、あるいはその両方のデジタル・レコードである。1つの実施形態では、電子医療レコード(EMR)は、形式化されたEMRデータセットに記憶される。EMRデータセットのスキーマは、異なるデータベース・システム、または異なる病院/医療機関、あるいはその両方によって変わり得る。図2は、EMRデータセットの一例を図示している。データセットは、フォーマルなテーブルを含んでいる。簡略のために、図2には、4つのテーブルT1、T2、T3、およびT4しか例示されていない。

【0018】

50

テーブル T 1 は、患者テーブルと題され、患者の基本データを含んでいる。テーブル T 1 の第 1 行は属性（「属性名」とも呼ばれる）を含み、第 2 行と他の行（図示せず）は、属性の値（「属性値」とも呼ばれる）を含んでいる。テーブル T 1 において、属性「PID」は、患者を一意に識別する患者識別子（ID）を意味する。「PID」は、テーブル T 1 の主キーとして機能することを示すために、「PID」に記号「+」が付されている。属性「性別」は、患者の性別を意味する。属性「年齢」は、患者の年齢を意味する。第 2 行は、患者のデータ・レコードを表し、3つの属性「PID」、「性別」、および「年齢」にそれぞれ対応する属性値「P1」、「M」および「67」を備えている。値「P1」は患者の ID であり、「M」は患者が男性であることを意味し、「67」は患者の年齢である。テーブル T 1 は、異なる患者のための基本データを含み得る。図 2 は、単に 3 列 2 行を例示的に図示している。実際の実施では、列数および行数はこのように限定されない。

10

#### 【0019】

テーブル T 2 は、エンカウンター・テーブルと呼ばれ、患者による通院のためのデータを記憶する。テーブル T 2 の第 1 行は属性を含み、他の行は属性値を含んでいる。第 1 行に、4つの属性が図示されている。属性「EID」は、患者による通院を一意に識別するエンカウンター識別子（ID）を意味する。「EID」は、テーブル T 2 の主キーとして機能することを示すために、「EID」に記号「+」が付されている。属性「診断」は、通院の診断結果を意味する。属性「部門」は、病院の診療部門を意味する。また、「PID」は、T2 の外部キーとして機能することを示すために記号「\*」が付されていることを除いて、T 1 とおける場合と同じものを意味する。外部キー「PID」を用いて、テーブル T 1 に対する参照がなされ得る。第 2 行は、患者による通院のデータ・レコードを表す。第 2 行における属性値「E1」、「DM」、「D1」、および「P1」はそれぞれ、4つの属性「EID」、「診断」、「部門」、および「PID」に対応する。「E1」はエンカウンター ID であり、「P1」は患者 ID であり、「DM」は患者「P1」の診断結果であり、「糖尿病」の略語であり、「D1」は患者「P1」によって訪問された診療部門の ID である。テーブル T 2 は、異なる患者についての情報を含み得、そして特定の患者による異なる通院についての多数の行を含み得る。図 2 では、4列 2行のみを例示的に図示している。実際の実施では、列数および行数はこのように限定されない。

20

#### 【0020】

テーブル T 3 は、ラボ・テスト・テーブルと呼ばれ、1回または複数回の通院において患者に対して行われたラボ・テストに関するデータを記憶する。テーブル T 3 の第 1 行は属性を含み、他の行は値を含んでいる。第 1 行の第 4 列に 4つの属性が図示されている。属性「LID」は、ラボ・テストを一意に識別するラボ・テスト識別子（ID）を意味する。「LID」が T 3 の主キーとして機能することを示すために、「LID」に記号「+」が付されている。「EID」は、テーブル T 2 におけるものと同じものを意味し、記号「\*」によって描写されるように、T 3 の外部キーとして機能し得る。属性「日付」は、ラボ・テストの日付を意味する。属性「技術者」は、ラボ・テストの技術者を意味する。第 2 行は、それぞれ 4つの属性「LID」、「EID」、「日付」、および「技術者」に対応する値「L1」、「E1」、「15-10-11」、および「技術者1」を備えるラボ・テスト用のデータ・レコードを表す。「L1」はラボ・テスト ID であり、「E1」はエンカウンター ID であり、「15-10-11」はラボ・テスト「L1」の日付であり、「技術者1」はラボ・テスト「L1」を担当する技術者の ID である。テーブル T 3 は、異なる患者についての情報を含み得、そして単一の通院中にいくつかのラボ・テストを受け得る特定の患者についての多数の行を含み得る。繰り返すが、テーブル T 3 に含まれ得る列と行の数に制限はない。

30

40

#### 【0021】

テーブル T 4 は、ラボ・テスト指数テーブルと呼ばれ、ラボ・テストのラボ・テスト指数に関するデータを記憶する。テーブル T 4 の第 1 行は属性を含み、他の行は属性値を含んでいる。第 1 行における 3 列に、3つの属性が図示されている。属性「LName」は

50

、ラボ・テスト指数を一意に識別するラボ・テスト指数名を意味する。「L V a l u e」は、ラボ・テスト指数の値を意味する。さらに、「L I D」はテーブルT3におけるものと同じであり、付された記号「\*」によって描写されるように、T4の外部キーとして機能し得る。第2行は、ラボ・テスト指数のデータ・レコードを表す。値「H b A 1 C」、「10.5」、「L1」は、「H b A 1 C」という名称のラボ・テスト指数が、ラボ・テストL1で行われ、ラボ・テスト指数「H b A 1 C」の結果値が「10.5」であることを意味する。テーブルT4は、異なるラボ・テストおよび異なる患者に関する情報を含み得、単一のラボ・テストについて多数の行を含み得る。繰り返すが、テーブルT4に含まれ得る列と行の数に制限はない。

【0022】

II. 臨床知識

【0023】

臨床知識とは、疾患、機序および病因、治療法および相互作用、ならびにラボ・テストの解釈に関する情報の体系を称し、これは、多数の患者および公衆衛生政策に関する決定に広く適用可能である。臨床知識の体系は、臨床規則のセットに形式化され得る。このような臨床規則のセットの例は、以下の通りである。

R1：血糖コントロールが満たされている：1か月における血糖値の80%が満たされている（空腹時血糖値 $< 7.5 \text{ mmol/L}$ または2時間血糖値 $< 10 \text{ mmol/L}$ ）。

R2：血糖値が高いままである：3か月における血糖値の80%が高い（空腹時血糖値 $9 \text{ mmol/L}$ または2時間血糖値 $13 \text{ mmol/L}$ ）。

R3：低血糖値：直近の血糖値が低い（血糖値 $< 3.9 \text{ mmol/L}$ ）。

R4：高リスク患者：年齢が65歳を超えて糖尿病を患い、過去3か月間のH b A 1 c  $> 9.0$ である。

【0024】

1つの実施形態では、臨床知識は、上記で図示したもののようなフォーマルな臨床規則である。臨床知識をフォーマルな臨床規則に形式化する既存の技術が存在する。様々な実施形態の説明を曖昧にすることを避けるために、臨床規則の生成に関する詳細は本明細書では省略される。

【0025】

EMRデータセットを考えると、既知の臨床知識に基づいてEMRデータセットから臨床的特徴を抽出することが望ましいであろう。たとえば、

「高リスク患者は65歳超である人々であり、糖尿病を患っており、テスト指数「H b A 1 c」は、3カ月以内のラボ・テストにおいて9.0を超える」

のように解釈され得る臨床規則R4に基づいて、図2に図示されるようなテーブルから、高リスクの患者を探索することが望ましいであろう。

【0026】

臨床的特徴を自動的に抽出するために、まず、クエリが構成される。ここで、図3を参照して示すように、このようなクエリを構成するための1つまたは複数の実施形態が説明される。図3は、1つの実施形態にしたがって、臨床知識に基づいて、臨床的特徴をEMRテーブルのセットに抽出するためのクエリを生成するためのコンピュータで実施される方法300を例示するフローチャートである。一般に、方法300は以下の動作、すなわち、

S310に描写されるように、臨床知識にしたがって構築された知識ツリーを取得することと、

S320によって描写されるように、EMRテーブルのセットに対応するEMRグラフを取得することとあって、EMRグラフは、EMRテーブルの各々の構造と、EMRテーブルの属性間の参照関係とを表すために、テーブル・ノードおよび属性ノードから構成される、取得することと、

S330によって描写されるように、知識ツリーおよびEMRグラフに基づいてサブクエリを生成することと、

10

20

30

40

50

S 3 4 0 によって描写されるように、知識ツリーにしたがってサブクエリを組み合わせることによってクエリを構成することを含んでいる。

【 0 0 2 7 】

以上の動作を詳細に説明する。図 4 は、1 つの実施形態にしたがう知識ツリー 4 0 0 を例示的に図示している。1 つの実施形態では、知識ツリーは、知識の体系、または規則、あるいはその両方をツリー構造で表現するために使用される。方法 3 0 0 が、臨床規則 R 4 の臨床知識に関連していると仮定する。臨床規則 R 4 によれば、6 5 歳を超える患者が糖尿病を患っており、彼の「H b A 1 c」指数が過去 3 か月以内のラボ・テストにおいて 9 . 0 を超えている場合、その患者は「高リスク患者」のカテゴリに分類される。したがって、臨床知識にしたがって構築された知識ツリーを取得するステップ S 3 1 0 は、図 4

10

【 0 0 2 8 】

図 4 に図示されるように、知識ツリー 4 0 0 は、異なる形状のノードから構成される。五角形のノードは、ノード t - 1 のようなターゲット・ノードを表す。楕円形のノードは演算子ノード o p 2、o p 3、o p 4、o p 8、および o p 1 1 のような算術 / 論理演算子を表す。長方形のノードは、概念ノード c 5、c 9、および c 1 3 のような臨床概念を表す。菱形のノードは、集約ノード a g 7 および A g 1 2 のような集約演算子を表す。平行四辺形のノードは、フィルタ・ノード f 1 4 のような概念フィルタを表す。最後に、三角形のノードは、値ノード v 6、v 1 0、および v 1 5 のような定数値を表す。

20

【 0 0 2 9 】

図 4 に図示するように、ターゲット・ノード t - 1 は、知識ツリー 4 0 0 のルートである。1 つの実施形態では、ターゲット・ノードは、臨床的特徴抽出のターゲットを表す。この例では、ターゲット・ノードは「patient」である。一般に、ターゲット・ノードは、規則または知識ツリーの主題としてみなされ得る。たとえば、「患者」は、「A B C の条件を有する患者は、高リスク患者である」という規則のターゲットであり、構築されるべき特徴は、各「患者」について「患者が、高リスクであるか否か」のようなブール型のものとなる。「エンカウンター」は「X Y Z の条件を有するエンカウンターは、外来のフォローアップ・エンカウンターである」という規則のターゲットであり、構築されるべき特徴は、各「エンカウンター」について、「エンカウンターが、外来のフォローアップ・エンカウンターであるか否か」というブール型のものとなる。

30

【 0 0 3 0 】

演算子ノード o p 2 および o p 3 は、「&&」または「AND」という論理演算子を表す。演算子ノード o p 4 および o p 1 1 は、「>」または「より大きい」という論理演算子を表す。演算子ノード o p 8 は、「=」または「に等しい」という論理演算子を表す。概念ノード c 5、c 9、および c 1 3 はそれぞれ、臨床概念「年齢」、「糖尿病」、および「H b A 1 c」を表す。集約ノード a g 7 および a g 1 2 は、それぞれ集約演算子「任意」および集約演算子「最大」を表す。値ノード v 1 0 は、定数値「Y」を表し、この例では「はい」または論理的「真」を意味する。値ノード v 1 5 は、定数値「9 . 0」を表し、この例では、テスト指標「H b A 1 c」に対する測定値である。

40

【 0 0 3 1 】

当業者に理解されるように、集約演算子は、同じ性質の要素の集合に対して演算を実行するために使用される。上述したように、「最大」および「任意」に加えて、集約演算子の別の例は「平均」であり、これはデジタル要素の集合の平均を計算するために使用される。一般に、集約ノードは、特徴構築における 1 対多または多対多の関係をどのように処理するのかを示す。たとえば、ターゲットが「患者」であり、「患者が今まで北京に住んでいたか否か」という特徴が構築されるべきである。エンカウンターごとに、患者の住所記録が存在し得る。したがって、この 1 対多の関係は、集約演算子「任意」を使用して処理される。同様に、特徴が「過去における患者の最大体重」に変化した場合、「最大」集約演算子が、多数の体重測定記録に適用されねばならない。通常、集約演算子は、「ターゲット」がリレーショナル・テーブルにおける一意のキーではない場合、または「ターゲ

50

ット」属性を有するテーブルが、1対多または多対多の関係を有する別のテーブルと結合する場合に適用される。

#### 【0032】

フィルタ・ノード f 1 4 は、概念フィルタ「過去3カ月」を表す。本明細書で使用されるように、フィルタ・ノードは、その親ノードの関連付けられた属性について満たされねばならない追加条件を表す。この例では、「過去3カ月」を表すフィルタ・ノード f 1 4 は、その親ノード c 1 3 が、日付タイプの何らかの関連付けられた属性を必要とするという制限を指定し、属性値は、過去3カ月以内でなければならない。

#### 【0033】

知識表現ツリーに精通している人は、知識ツリー 4 0 0 が3つの条件、「年齢 > 6 5」（ノード c 5、o p 4、v 6 によって表される）、「糖尿病を患っている」（ノード c 9、a g 7、o p 8、v 1 0 によって表される）、および「過去3か月において H b A 1 c > 9 . 0」（他のノード（o p 2 および o p 3）によって組み合わせられたノード c 1 3、o p 1 1、a g 1 2、v 1 5、f 1 4 によって表される）を含んでいると理解するものとする。異なる形状の使用は、異なる種類のノードを視覚的に区別することであることに留意されたい。当業者は、異なる種類のノードを表す他の類似または同等の手法があり得ることを理解するであろう。

10

#### 【0034】

様々な分野において、ドメイン知識を表すために、知識ツリーを構築し得る既存の技術が存在する。上記の説明に基づいて、当業者は、様々な手法で、臨床知識にしたがって、知識ツリーをどのように構築するのかを容易に知るであろう。1つまたは複数の実施形態は、ステップ S 3 1 0 を実施するために、すなわち、臨床知識にしたがって構築された知識ツリーを取得するために、臨床知識を表すための知識ツリーの構築のための技術を使用し得る。

20

#### 【0035】

図3に戻って頂きたい。ステップ S 3 1 0 について説明したので、次に方法 3 0 0 のステップ 3 2 0 が説明される。ステップ S 3 2 0 は、EMR テーブルのセットに対応する EMR グラフを取得することを含み、EMR グラフは、EMR テーブルの各々の構造および EMR テーブルの属性間の参照関係を表すためにテーブル・ノードおよび属性ノードから構成される。1つの実施形態によれば、EMR テーブルのセットに対応する EMR グラフを取得するステップ S 3 2 0 は、EMR テーブルのセットをその等価のグラフ表現に変換することを含んでいる。

30

#### 【0036】

図6は、1つの実施形態にしたがって、図2のEMR テーブルのセットから変換されたEMR グラフ 5 0 0 を例示的に図示している。EMR グラフ 5 0 0 は、4つのテーブル・ノード t 1、t 2、t 3、および t 4、ならびにそれらの関連付けられたテーブル・ノードに双方向に接続された属性ノードを含んでいる。グラフ 5 0 0 は、EMR テーブルの各々構造、および図2における4つのリレーショナル・テーブル T 1、T 2、T 3、および T 4 の属性間の参照関係を表す。

#### 【0037】

EMR グラフ 5 0 0 に図示されるように、テーブル・ノード t 1 は、テーブル T 1 の主キー、すなわち患者 ID 属性「PID」を表す。したがって、テーブル・ノードは、「キー・ノード」とも呼ばれる。テーブル・ノード t 1 は、2つのプレーン・エッジを介して3つの属性ノード「PID」、「性別」、および「年齢」に双方向に接続されている。3つの属性ノードは、テーブル T 1 の3つの属性「PID」、「性別」、および「年齢」に対応している。同様に、テーブル・ノード t 2 は、テーブル T 2 の主キー、すなわち「EID」を表す。テーブル・ノード t 2 は、図6において双方向の実線で示すように、2つのプレーン・エッジを介して4つの属性ノード「EID」、「診断」、「部門」、および「PID」に双方向に接続されている。4つの属性ノードは、テーブル T 2 の4つの属性「EID」、「診断」、「部門」、および「PID」に対応している。テーブル・ノード

40

50

t 3 は、テーブル T 3 の主キー、すなわち「L I D」を表す。テーブル・ノード t 3 は、2つのプレーン・エッジを介して4つの属性ノード「L I D」、「E I D」、「日付」、および「技術者」に双方向に接続されている。4つの属性ノードは、テーブル T 3 の4つの属性「L I D」、「E I D」、「日付」、および「技術者」に対応する。テーブル・ノード t 4 は、テーブル T 4 の主キー、すなわち「L N a m e」を表す。テーブル・ノード t 4 は、2つのプレーン・エッジを介して2つの属性ノード「L N a m e」および「L V a l u e」に双方向に接続されている。2つの属性ノードは、テーブル T 4 の2つの属性「L N a m e」および「L V a l u e」に対応している。

**【0038】**

テーブル・ノード t 1、t 2、t 3、および t 4 は、以下の原理にしたがって接続される。テーブル t A とテーブル t B との間に外部キー参照が存在する場合、テーブル・ノード t B からテーブル・ノード t A へプレーン・エッジが存在する。テーブル・ノード t A からテーブル・ノード t B への、EMR グラフ 500 において矢印点線で描写されている集約エッジもある。1つの実施形態では、集約エッジは、到来ノードと発信ノードのテーブルが、1対多または多対多の関係を有することを意味し、したがって、発信ノードの属性が、(到来ノードと結合された)特徴構築において使用される必要がある場合は、「集約ノード」が必要とされる。集約エッジは推移的である。言い換えれば、テーブル・ノード t Y がテーブル・ノード t X に直接的に接続されていないが、集約エッジを介したトラバースによってテーブル・ノード t X に到達可能である場合、テーブル・ノード t X からテーブル・ノード t Y への集約エッジが存在する。たとえば、図 6 では、テーブル T 2 と

10

20

**【0039】**

同様に、テーブル T 3 とテーブル T 2 との間には外部キー「E I D」参照があるので、テーブル・ノード t 2 からテーブル・ノード t 3 へのプレーン・エッジと、テーブル・ノード t 3 からテーブル・ノード t 2 への集約エッジとが存在する。テーブル T 4 とテーブル T 3 との間には外部キー「L I D」参照もあるため、テーブル・ノード t 3 からテーブル・ノード t 4 へのプレーン・エッジと、テーブル・ノード t 4 からテーブル・ノード t 3 への集約エッジとが存在する。さらに、テーブル・ノード t 3 は t 1 に直接的に接続されてい

30

**【0040】**

上記の段落では、EMR データセットの関係は、図 2 におけるテーブル T 1 から T 4、および図 6 における EMR グラフ 500 によって表されるとして論じられた。一般的に言って、EMR グラフは、リレーショナル EMR テーブルのセットの属性関係を表す。EMR グラフは、テーブル・ノードと属性ノードとから構成される。テーブル・ノードはそれぞれ、EMR テーブルに対応する。各テーブル・ノードは、テーブル・ノードに対応する EMR テーブルの属性にそれぞれ対応するすべての属性ノードと双方向に接続されている。テーブル・ノードは、対応する EMR テーブルの参照関係に基づいて、プレーン・エッジおよび集約エッジによって互いに接続される。

40

**【0041】**

当業者は、関係にしたがって適切にプログラミングすることによって、テーブル T 1 から T 4 によって表される EMR データセットが、EMR グラフ 500 に自動的に変換され得ることを認識するものとする。さらに、任意のリレーショナル・スキーマにおける EMR データセットは、図 6 に図示されるものと同様に EMR グラフに自動的に変換され得る。このようにして、図 2 に図示されるような EMR テーブルのセットに対応する EMR グラフ 500 が、ステップ S 320 において取得され得る。

50

## 【 0 0 4 2 】

図 3 に戻って示すように、ステップ S 3 1 0 および S 3 2 0 の後、次のステップ S 3 3 0 は、知識ツリーおよび E M R グラフに基づいてサブクエリを生成している。このステップでは、E M R グラフ 5 0 0 からの知識ツリー 4 0 0 の概念ノードの各々に対応して、サブクエリが生成される。1つの実施形態では、知識ツリーおよび E M R グラフに基づいてサブクエリを生成するステップ S 3 3 0 は、

S 3 3 1 に描写されるように、知識ツリーの単一概念のサブツリーを識別することと、単一概念のサブツリーの各単一概念のサブツリーごとに、

S 3 3 2 に描写されるように、知識ツリーの単一概念のサブツリーを選択し、

S 3 3 3 に描写されるように、単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、E M R グラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることによって、E M R グラフの有効なサブグラフを識別し、

S 3 3 4 に描写されるように、単一概念のサブツリーおよび有効なサブグラフにしたがってサブクエリを構成することとを備えている。

## 【 0 0 4 3 】

言い換えれば、1つの実施形態では、ステップ S 3 3 0 は、図 3 に図示されるように、ステップ S 3 3 1 から S 3 3 6 を含む処理で実行される。まず、ステップ S 3 3 1 が実行され、知識ツリーの単一概念のサブツリーを識別する。1つの実施形態では、単一概念のサブツリーは、知識ツリーのサブツリーを称する。単一概念のサブツリーは、概念ノードから始まりターゲット・ノード、つまりルート・ノードまでのパスにおけるすべてのノードを備えている。さらに、単一概念のサブツリーは、他の概念ノードを備えていないパス内のノードの各々の任意の分岐をも備えている。言い換えれば、単一概念のサブツリーは、以下の基準を満たす。(1) 単一概念のサブツリーには、1つの概念ノードしか存在しない。(2) 単一概念のサブツリーは、1つの概念ノードから始まりターゲット・ノードまでのパス内のすべてのノードを備えている。(3) 単一概念のサブツリーは、子に概念ノードが存在しない限り、パス内のノードの任意の子を備えている。

## 【 0 0 4 4 】

図 5 は、知識ツリー 4 0 0 の3つの単一概念のサブツリー 4 0 1、4 0 2、および 4 0 3 を図示している。図 5 に図示するように、単一概念のサブツリー 4 0 1 は、ノード { t - 1、o p 2、o p 3、o p 4、c 5、v 6 } を備えている。単一概念のサブツリー 4 0 2 は、ノード { t - 1、o p 2、o p 3、a g 7、o p 8、c 9、v 1 0 } を備えている。単一概念のサブツリー 4 0 3 は、ノード { t - 1、o p 2、o p 1 1、a g 1 2、c 1 3、f 1 4、v 1 5 } を備えている。1つの実施形態では、知識ツリーの単一概念のサブツリーを識別するステップ S 3 3 1 は、以下のように、知識ツリー 4 0 0 を分解することによって実行される。概念ノード c 5 から始まりターゲット・ノード t - 1 までのパスが、各概念ノード c 5 について発見される。このパスは、分解された1つの接続された構成要素(または、サブツリー)のバックボーンを形成する。このパスにおける各ノードについて、そのすべての子をチェックする。子がいずれの概念ノードも含んでいない場合、子は、接続された構成要素(またはサブツリー)に追加される。パス内のすべてのノードがチェックされると、サブツリーは、知識ツリー全体の分解された構成要素の1つとなる。このようにして、単一概念のサブツリー(以下、「サブツリー」とも称される)が識別され得る。k 個の概念ノードを有する知識ツリーについて、k 個のサブツリーが識別され得ることに留意されたい。たとえば、知識ツリー 4 0 0 に対して、3つのサブツリー 4 0 1、4 0 2、および 4 0 3 が識別され得る。

## 【 0 0 4 5 】

ステップ S 3 3 1 の後、すべてのサブツリー 4 0 1、4 0 2、および 4 0 3 が処理されるまで、ステップ S 3 3 2、S 3 3 3、および S 3 3 4 が、個々のサブツリーについて反復的に実行される。ステップ S 3 3 2 において、サブツリー 4 0 1 が選択される。次に、ステップ S 3 3 3 において、E M R グラフ 5 0 0 の有効なサブグラフが識別される。1つの実施形態では、これは、サブツリー 4 0 1 のターゲット・ノードおよび概念ノードを、

EMRグラフ500のテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすること(以下、「概念マッピング」とも称される)によって実行される。

【0046】

1つの実施形態では、EMRグラフの有効なサブグラフは、以下の基準を満たすEMRグラフのサブグラフである。(1)ターゲット・ノードに一致するテーブル・ノードがサブグラフに存在する。(2)サブグラフ内の各属性ノードは、直接的にまたは有向パスを介して間接的にテーブル・ノードに到達し得る。1つの実施形態では、概念マッピングは、以下のうちの任意の1つまたは任意の組合せを備えている。(1)属性によるマッピング、(2)属性値によるマッピング、および(3)追加フィルタの適用によるマッピング。第1のタイプでは、概念ノードは、概念ノードの概念に一致する(同じである、または同じとみなされる)1つまたは複数の属性ノードにマッピングされる。第2のタイプでは、概念ノードは、1つまたは複数の属性の、ある特定の値にマップされ、概念データ・タイプと、マップされた属性のデータ・タイプとに基づいて、追加の属性に関連付けられる。第3のタイプでは、概念ノードは、第1または第2のタイプのマッピングとしてマッピングされるが、フィルタの定義に基づいて、1つまたは複数の属性との追加の関連付けを有する。

10

【0047】

図7を参照して示すように、ステップS333において、サブツリー401に対する概念マッピングが、以下のように実行される。まず、矢印線601によって図示されるように、サブツリー401のターゲット・ノードt-1が、EMRグラフ500のテーブル・ノード「t1」にマッピングされる。次に、テーブルT1の属性「年齢」が、サブツリー401の概念ノードc5の概念「年齢」と同じであるので、矢印線602によって描写されるように、概念ノードc5が、EMRグラフ500の属性ノード「年齢」にマッピングされる。属性ノード「年齢」は、ターゲット・ノードt-1に一致するテーブル・ノード「t1」に直接到達し得るので、属性ノード「年齢」とテーブル・ノード「t1」を含むサブグラフは、有効なサブグラフ(以下「サブグラフ」とも称される)であると識別される。有効なサブグラフは、図7で強調表示されているように、{t1、年齢}として示され得る。

20

【0048】

ステップS334において、サブツリー401および有効なサブグラフ{t1、年齢}にしたがって、サブクエリが構成される。この場合、サブツリー401の構造に基づいて、サブクエリが構成され得る。具体的には、サブツリー401の構造は「年齢>65」の条件を表しているので、サブクエリは、以下のように、サブクエリ(1)になるように導出される。

30

「t1 . 年齢 > 65」 (1)

【0049】

1つの実施形態では、サブクエリは、サブツリーの構造と、概念ノードの属性ノードへのマッピングとにしたがって導出される。まず、サブクエリの主題が生成される。マッピングされた属性ノード(たとえば、概念ノードc5に一致する属性ノード「年齢」)が、ターゲット・ノード(たとえば、テーブル・ノードt1にマッピングされたターゲット・ノードt-1)に直接接続されている場合、その属性ノードに対応する属性が、主題である。上記の例では、主題は「年齢」である。

40

【0050】

図7に図示される処理は、第1の種類の概念マッピングの例である。この例では、概念ノードc5が、1つの属性ノードにマッピングされる。しかしながら、概念ノードは、複数の属性ノードにマッピングされ得る。いくつかの属性ノードがマッピングされ得る場合、各マッピングは、一致したスコアに基づいて列挙される。たとえば、概念「年齢」が、属性「年齢」のみならず「初妊娠年齢」にマッピングされ得る場合、(より高いスコアを有する)第1の属性へのマッピングが最初に評価される。第2の属性へのマッピングは、より多くのクエリ候補が必要な場合に評価される。

50



## 【0051】

概念マッピングでは、属性ノードが、多数のテーブル・ノードを横切る1つまたは複数のプレーン・エッジによってターゲット・ノードに接続されている場合、1つまたは複数の「結合」が、プレーン・エッジに基づいて、関連するテーブル・ノードに適用され、結合されたテーブルにおいてマッピングされている属性が、主題である。接続されたサブグラフに含まれる集約エッジがある場合は、対応する集約演算子が、テーブル・ノード結合において使用される。

## 【0052】

サブツリー401についてサブクエリを生成する処理が完了した後、ブロックS335によって描写されるように、処理が、サブツリーのすべてについて行われたか否かが判定される。行われたと判定された場合、S330の処理は終了し、ブロックS336によって描写されるように、制御は戻され、その結果、ステップS340が実行される。行われたと判定されなかった場合、処理はS332に進み、別のサブツリー、この場合はサブツリー402を選択する。

## 【0053】

図8を参照されたい。ステップS333において、サブツリー402のターゲット・ノードt-1が、矢印線611によって描写されるように、EMRグラフ500のテーブル・ノード「t1」にマッピングされる。そして、テーブルT2内の属性「診断」に対応する属性値「DM」は、サブツリー402の概念ノードc9に対する概念「糖尿病」と同じであるとみなされ、EMRグラフ500の属性ノード「診断」は、属性「診断」に対応している。概念ノードc9は、テーブル・ノードt2に接続されている矢印線612によって描写されるように、属性ノード「診断」にマッピングされる。この時点において、マッピングされた2つのマップされたノードである「診断」および「t1」は接続されていない。しかしながら、サブツリー402には集約ノードag7が存在し、矢印線613によって描写されているように、テーブル・ノード「t2」から「t1」に集約エッジ<t2、t1>をアクティブにするために適用される。これはテーブル「t2」が、まさに「t1」に到達し得ることを意味する。この時点において、属性ノード「診断」は、プレーン・エッジ<診断、t2>およびアクティブにされた集約エッジ<t2、t1>を介して間接的にテーブル・ノード「t1」に到達するため、テーブル・ノード「t1」およびテーブル・ノード「t2」ならびにノード「診断」は、EMR500の連結されたサブグラフを構成する。これは、図8で強調表示されているように、有効なサブグラフ{t1、t2、診断}であるとして識別され得る。

## 【0054】

次に、ステップS334において、サブツリー402およびサブグラフ{t1、t2、診断}にしたがって、サブクエリが構成される。特に、サブツリー402の構造は、「糖尿病を患っている任意の患者」という条件を表しているため、サブクエリは、以下のようにサブクエリ(2)となるように導出される。

「t1 join t2 on PID、 any ( t2 . 診断 = ' DM ' ) group by PID」 (2)

## 【0055】

サブクエリ(2)の意味は、以下のように解釈され得る。ある患者の診断が「DM」(糖尿病)に等しい場合、新たな(プール型)属性を生成しながら、テーブルT1は、属性「PID」を使用してテーブルT2と結合される(したがって、一意のIDを有する各患者は、新たなテーブルの別の行になる)。

## 【0056】

上記で説明された概念マッピング処理では、集約エッジ<t2、t1>をアクティブにすることによって、マッピングされたテーブル・ノード「t2」と「t1」とを接続するために、集約ノードag7が使用される。たとえば、サブツリーにk個の集約ノードが存在する場合、EMRグラフ500内のマッピングされたノードを接続し、これらを接続状態にするために、k個の集約エッジが使用され得る。図8に図示される処理は、第2のタ

10

20

30

40

50

イブの概念マッピングの例である。第2のタイプの概念マッピングに関して、サブツリーの概念ノードは、1つまたは複数の属性のいくつかの特定の値にマッピングされ、概念データ・タイプおよびマッピングされた属性のデータ・タイプに基づいて、追加の属性と関連付けられ得る。

【0057】

図8の例では、サブグラフ{t1、t2、診断}は、接続されたサブグラフを構成し、以下の基準を満たすと、有効なサブグラフ{t1、t2、診断}であると識別され得る。  
 (1) ターゲット・ノードに一致するテーブル・ノードがサブグラフに存在する。  
 (2) サブグラフ内の各属性ノードは、直接的にまたは有向パスを介して間接的に、テーブル・ノードに到達し得る。1つの実施形態では、「接続されたサブグラフ」という用語も、有効なサブグラフを表すために使用される。一般に、マッピングされたノードからターゲット・ノードへの有向パスが存在する場合、マッピングされたノードとターゲット・ノードが接続されているか、またはマッピングされたノードが、ターゲット・ノードに到達し得る。多数のノード(すなわち、1つのキー・ノードと、1つまたは複数の属性ノード)がマッピングされている場合、サブグラフは、(1) キー・ノードからターゲット・ノードへの有向パスが存在する場合、および(2) 各属性ノードについて、キー・ノードへの有向パスが存在する場合、接続される。ここで、ターゲット・ノード(たとえば、t-1)は、ターゲット・ノードのマッピングされたノード(たとえば、t1)と等価であるとみなされる。

【0058】

図9を参照して示すように、サブクエリ(2)が生成された後、ステップS332において、サブツリー403が選択される。次に、ステップS333において、サブツリー403に対する概念マッピングが、以下のように実行される。最初に、サブツリー403のターゲット・ノードt-1が、矢印線621によって描写されるように、EMRグラフ500のテーブル・ノード「t1」にマッピングされる。そして、テーブルT4の属性「LName」に対応する属性値「HbA1c」が、サブツリー403の概念ノードc13の概念「HbA1c」と同じであるので、分割矢印線622によって描写されるように、概念ノードc13は、EMRグラフ500のノード「LName」およびノード「LValue」にマッピングされる。さらに、概念ノードc13に関連付けられたフィルタ・ノードf14が、時間関連条件を定義するので、概念ノードc13は、矢印線623によって描写されるように、日付タイプの属性ノード「日付」にマッピングされる。次いで、サブツリー403内の集約ノードag12が、矢印線624によって描写されるように、テーブル・ノード「t4」からテーブル・ノード「t1」への集約エッジにマッピングされる。したがって、集約エッジ<t4、t1>がアクティブにされる。この時点において、属性ノード「日付」は、プレーン・エッジ<日付、t3>、プレーン・エッジ<t3、t4>、および集約エッジ<t4、t1>を介してテーブル・ノード「t1」に到達し得、「LName」は、有向パス「LName」「t4」「t1」を介してテーブル・ノード「t1」に到達し得、「LValue」は、有向パス「LValue」「t4」「t1」を介してテーブル・ノード「t1」に到達し得る。したがって、サブツリー403に対する概念マッピングは、図9で強調表示されているように、有効なサブグラフ{t1、t3、t4、日付、LName、LValue}の識別をもたらす。

【0059】

ステップS334では、サブツリー403およびサブグラフ{t1、t3、t4、日付、LName、LValue}にしたがって、サブクエリが構成される。特に、サブツリー403の構造によって表される条件にしたがって、サブクエリは、以下のように、サブクエリ(3)となるように導出される。

```
「t3 join t4 on LID, t4.LName = 'HbA1c' and
t3.日付 > '2015-07-01' as t5;
t1 join t5, max(t5.LValue) > 9.0 group by
PID」 (3)
```

10

20

30

40

50

これは、最近の最大の H b A 1 C 検査結果が、9.0 より大きいすべての患者を照会する。

【0060】

図9に図示される処理は、第3のタイプの概念マッピングの例である。第3のタイプの概念マッピングのための1つの実施形態では、概念ノードは、第1または第2のタイプのマッピングとしてマッピングされるが、フィルタ定義に基づいて1つまたは複数の属性とのさらなる関連付けを有する。

【0061】

図3に戻って示すように、ステップS334が実行された後、処理は、ステップS340に進む。ステップS340では、知識ツリー400にしたがって、サブクエリ(1)、(2)および(3)を組み合わせることによって、クエリが構成される。知識ツリー400内のノードop2のための演算子「&&」と、ノードop3のための演算子「&&」から、クエリは、3つのサブクエリの交わり、すなわち、サブクエリ(1)&&サブクエリ(2)&&サブクエリ(3)、すなわち、「t1.年齢>65」&&「t1 join t2 on PID, any(t2.診断='DM') group by PID」&&「t3 join t4 on LID, t4.LName='HbA1c' and t3.日付>'2015-07-01' as t5; t1 join t5, max(t5.LValue)>9.0 group by PID」である。

【0062】

1つまたは複数の実施形態はまた、対応するデバイスを提供する。1つの実施形態では、臨床知識に基づいて、臨床的特徴をEMRテーブルのセットに抽出するためのクエリを生成するためのデバイスが提供される。デバイスは、1つまたは複数のプロセッサと、プロセッサのうちの少なくとも1つに結合されたメモリと、メモリに記憶されたコンピュータ・プログラム命令のセットとを備えている。コンピュータ・プログラム命令のセットは、以下の動作、すなわち、臨床知識にしたがって構築された知識ツリーを取得することと、EMRテーブルのセットに対応するEMRグラフを取得することとあって、EMRグラフは、EMRテーブルの各々の構造と、EMRテーブルの属性間の参照関係とを表すためのテーブル・ノードおよび属性ノードから構成される、取得することと、知識ツリーおよびEMRグラフに基づいてサブクエリを生成することと、知識ツリーにしたがってサブクエリを組み合わせることによってクエリを構成することとを実行するために、プロセッサのうちの少なくとも1つによって実行される。

【0063】

1つの実施形態によれば、知識ツリーおよびEMRグラフに基づいてサブクエリを生成することは、

- 知識ツリーの単一概念のサブツリーを識別することと、
- 単一概念のサブツリーの各単一概念のサブツリーごとに、
- 知識ツリーの単一概念のサブツリーを選択し、

単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、EMRグラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることによって、EMRグラフの有効なサブグラフを識別し、

単一概念のサブツリーおよび有効なサブグラフにしたがってサブクエリを構成することとを含んでいる。

【0064】

1つの実施形態によれば、EMRテーブルのセットに対応するEMRグラフを取得することは、EMRテーブルのセットを、その等価なグラフ表現に変換することを含んでいる。1つの実施形態では、知識ツリーの単一概念のサブツリーは、概念ノードから始まりターゲット・ノードまでのパス内のすべてのノードと、概念ノードを備えていないノードの各々の分岐とを備えたサブツリーである。

【0065】

1つの実施形態によれば、EMRグラフの有効なサブグラフは、以下の基準を満たすE

MR グラフのサブグラフである。(1) ターゲット・ノードに一致するテーブル・ノードがサブグラフに存在する。(2) サブグラフ内の各属性ノードは、直接的にまたは有向パスを介して間接的に、テーブル・ノードに到達し得る。

【0066】

1つの実施形態によれば、単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、EMR グラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることは、属性によるマッピング、属性値によるマッピング、および、追加のフィルタを適用することによるマッピングのうちの任意の1つまたは任意の組合せを含んでいる。

【0067】

1つの実施形態によれば、臨床知識に基づいて、臨床的特徴を、EMR テーブルのセットに抽出するためのクエリを生成するためのコンピュータ・プログラム製品が提供される。コンピュータ・プログラム製品は、非一時的なコンピュータ可読記憶媒体を備えている。プログラム命令は、非一時的なコンピュータ可読記憶媒体に記憶される。プログラム命令は、少なくとも1つのプロセッサに、

臨床知識にしたがって構築された知識ツリーを取得することと、

EMR テーブルのセットに対応するEMR グラフを取得することであって、EMR グラフは、EMR テーブルの各々の構造と、EMR テーブルの属性間の参照関係とを表すためのテーブル・ノードおよび属性ノードから構成される、取得することと、

知識ツリーおよびEMR グラフに基づいてサブクエリを生成することと、

知識ツリーにしたがってサブクエリを組み合わせることによってクエリを構成することを含む方法を実行させるために、少なくとも1つのプロセッサによって実行可能である。

【0068】

1つの実施形態によれば、知識ツリーおよびEMR グラフに基づいてサブクエリを生成することは、

知識ツリーの単一概念のサブツリーを識別することと、

単一概念のサブツリーの各単一概念のサブツリーごとに、

知識ツリーの単一概念のサブツリーを選択し、

単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、EMR グラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることによって、EMR グラフの有効なサブグラフを識別し、

単一概念のサブツリーおよび有効なサブグラフにしたがってサブクエリを構成することを含んでいる。

【0069】

1つの実施形態によれば、EMR テーブルのセットに対応するEMR グラフを取得することは、EMR テーブルのセットをその等価なグラフ表現に変換することを含んでいる。

【0070】

1つの実施形態によれば、知識ツリーの単一概念のサブツリーは、概念ノードから始まりターゲット・ノードまでのパス内のすべてのノードと、概念ノードを備えていないノードの各々の任意の分岐とを備えたサブツリーである。

【0071】

1つの実施形態によれば、EMR グラフの有効なサブグラフは、以下の基準を満たすEMR グラフのサブグラフである。(1) ターゲット・ノードに一致するテーブル・ノードがサブグラフに存在する。(2) サブグラフ内の各属性ノードは、直接的にまたは有向パスを介して間接的に、テーブル・ノードに到達し得る。

【0072】

1つの実施形態によれば、単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、EMR グラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることは、属性によるマッピング、属性値によるマッピング、および、追加の

10

20

30

40

50

フィルタを適用することによるマッピングのうちの任意の1つまたは任意の組合せを含んでいる。

【0073】

本発明は、任意の可能な技術的詳細レベルの統合におけるシステム、方法、またはコンピュータ・プログラム製品、あるいはその組合せであり得る。コンピュータ・プログラム製品は、プロセッサに対して、本発明の態様を実行させるためのコンピュータ可読プログラム命令を有するコンピュータ可読記憶媒体（または複数の媒体）を含み得る。

【0074】

コンピュータ可読記憶媒体は、命令実行デバイスによる使用のための命令を保持および記憶し得る有形のデバイスであり得る。コンピュータ可読記憶媒体は、たとえば、電子記憶デバイス、磁気記憶デバイス、光記憶デバイス、電磁記憶デバイス、半導体記憶デバイス、またはこれらの任意の適切な組合せであり得るが、これらに限定されない。コンピュータ可読記憶媒体のより具体的な例の非網羅的リストは、ポータブル・コンピュータ・ディスクレット、ハード・ディスク、ランダム・アクセス・メモリ（RAM）、読出専用メモリ（ROM）、消去可能プログラマブル読出専用メモリ（EPROMまたはフラッシュ・メモリ）、スタティック・ランダム・アクセス・メモリ（SRAM）、ポータブル・コンパクト・ディスク読出専用メモリ（CD-ROM）、デジタル多用途ディスク（DVD）、メモリ・スティック、フロッピー（R）・ディスク、パンチカードのような機械的に符号化されたデバイス、または命令が記録された溝の隆起構造、および上記の任意の適切な組合せを含んでいる。本明細書で使用されるコンピュータ可読記憶媒体は、電波または他の自由に伝播する電磁波、導波管または他の伝送媒体を伝播する電磁波（たとえば、光ファイバ・ケーブルを通過する光パルス）、または、ワイヤを介して送信される電気信号のような、一時的な信号自体であると解釈されるべきではない。

【0075】

本明細書において説明されたコンピュータ可読プログラム命令は、コンピュータ可読記憶媒体からそれぞれのコンピューティング/処理デバイスに、または、たとえば、インターネット、ローカル・エリア・ネットワーク、広域ネットワーク、もしくは無線ネットワーク、あるいはその組合せのようなネットワークを介して、外部コンピュータもしくは外部記憶装置にダウンロードされ得る。ネットワークは、銅伝送ケーブル、光伝送ファイバ、無線伝送、ルータ、ファイアウォール、スイッチ、ゲートウェイ・コンピュータ、またはエッジ・サーバ、あるいはその組合せを備え得る。各コンピューティング/処理デバイス内のネットワーク・アダプタ・カードまたはネットワーク・インターフェースは、ネットワークからコンピュータ可読プログラム命令を受信し、それぞれのコンピューティング/処理デバイス内のコンピュータ可読記憶媒体に記憶するためにコンピュータ可読プログラム命令を転送する。

【0076】

本発明の動作を実行するためのコンピュータ可読プログラム命令は、アセンブラ命令や、命令セット・アーキテクチャ（ISA）命令や、マシン命令や、マシン依存命令や、マイクロコードや、ファームウェア命令や、状態設定データや、集積回路構成のための構成データや、または、Smalltalk（R）、C++のようなオブジェクト指向プログラミング言語、および「C」プログラミング言語もしくは同様のプログラミング言語のような手続き型プログラミング言語を含む、1つもしくは複数のプログラミング言語の任意の組合せで記述されたソース・コードもしくはオブジェクト・コードのいずれかであり得る。コンピュータ可読プログラム命令は、完全にユーザのコンピュータにおいて、部分的にユーザのコンピュータにおいて、スタンドアロンのソフトウェア・パッケージとして、部分的にユーザのコンピュータおよび部分的にリモート・コンピュータにおいて、または全体的にリモート・コンピュータもしくはサーバにおいて実行し得る。後者のシナリオでは、リモート・コンピュータは、ローカル・エリア・ネットワーク（LAN）または広域ネットワーク（WAN）を含む任意のタイプのネットワークを介してユーザのコンピュータに接続され得るか、または（たとえば、インターネット・サービス・プロバイダを使用

してインターネットを介して)外部コンピュータに接続され得る。いくつかの実施形態において、たとえば、プログラマブル・ロジック回路構成、フィールド・プログラマブル・ゲート・アレイ(FPGA)、またはプログラマブル・ロジック・アレイ(PLA)を含む電子回路構成は、本発明の態様を実行するために、コンピュータ可読プログラム命令の状態情報を利用することによって、コンピュータ可読プログラム命令を実行し、電子回路構成をパーソナル化し得る。

【0077】

本発明の態様は、本発明の実施形態にしたがう方法、装置(システム)、およびコンピュータ・プログラム製品のフローチャート例図、またはブロック図、あるいはその両方を参照して本明細書で説明される。フローチャート例図、またはブロック図、あるいはその両方の各ブロック、ならびにフローチャート例図、またはブロック図、あるいはその両方内のブロックの組合せは、コンピュータ可読プログラム命令によって実施され得ることが理解されよう。

10

【0078】

コンピュータまたは他のプログラマブル・データ処理装置のプロセッサを介して実行する命令が、フローチャート、またはブロック図ブロックもしくは複数のブロック、あるいはその組合せにおいて指定された機能/動作を実施するための手段を生成するべく、これらのコンピュータ可読プログラム命令は、汎用コンピュータ、専用コンピュータ、またはマシンを生成するための他のプログラマブル・データ処理装置のプロセッサに提供され得る。フローチャート、またはブロック図ブロックもしくは複数のブロック、あるいはその組合せにおいて指定された機能/動作の態様を実施する命令を含む製造物品を備えるべく、これらのコンピュータ可読プログラム命令は、特定の手法で機能するように、コンピュータ、プログラム可能データ処理装置、または他のデバイス、あるいはその組合せに対して命令することができるコンピュータ可読記憶媒体に記憶され得る。

20

【0079】

コンピュータ、他のプログラマブル装置、または他のデバイス上で実行する命令が、フローチャート、またはブロック図ブロックもしくは複数のブロック、あるいはその組合せにおいて指定された機能/動作を実施するべく、コンピュータ可読プログラム命令はまた、コンピュータ、他のプログラマブル・データ処理装置、または他のデバイスにロードされ得、コンピュータによって実施される処理を生成するために、一連の動作ステップを、コンピュータ、他のプログラマブル装置、または他のデバイスにおいて実行させる。

30

【0080】

図中のフローチャートおよびブロック図は、本発明の様々な実施形態にしたがってシステム、方法、およびコンピュータ・プログラム製品の可能な実施のアーキテクチャ、機能、および動作を例示している。これに関して、フローチャートまたはブロック図における各ブロックは、指定された論理機能を実施するための1つまたは複数の実行可能命令を含むモジュール、セグメント、または命令の一部を表し得る。いくつかの代替実施では、ブロックに言及されている機能は、図面において言及されている順序とは異なる順序で生じ得る。たとえば、連続して図示されている2つのブロックは、実際には、実質的に同時に実行され得るか、または、これらブロックは、関連する機能に応じて、しばしば逆の順序で実行され得る。ブロック図、またはフローチャート例図、あるいはその両方の各ブロック、およびブロック図、またはフローチャート例図、あるいはその両方におけるブロックの組合せは、指定された機能または動作を実行する、または、専用ハードウェアとコンピュータ命令との組合せを実行する、専用ハードウェア・ベースのシステムによって実施され得ることが注目されるであろう。

40

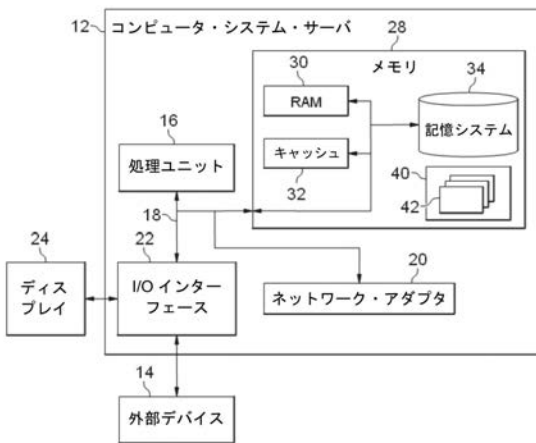
【0081】

本発明の様々な実施形態の説明は、例示の目的のために提示されているが、網羅的であること、または開示された実施形態に限定されることは意図されていない。多くの修正および変形が、説明された実施形態の範囲および思想から逸脱することなく、当業者に明らかになるであろう。本明細書で使用される用語は、実施形態の原理、実用化、または、市

50

場で見られる技術に対する技術的改善を最も良く説明するため、または当業者が本明細書で開示される実施形態を理解できるようにするために選択された。

【 図 1 】



【 図 2 】

**T1: 患者**

PID+	性別	年齢	...
P1	M	67	...
...	...	...	...

**T2: エンカウンター**

EID+	診断	部門	PID*	...
E1	DM	D1	P1	...
...	...	...	...	...

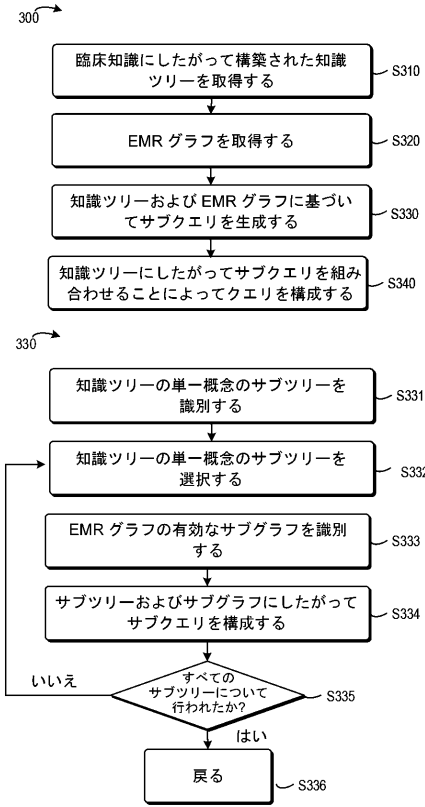
**T3: ラボ・テスト**

LID+	EID*	日付	技術者	...
L1	E1	15-10-11	Tech1	...
...	...	...	...	...

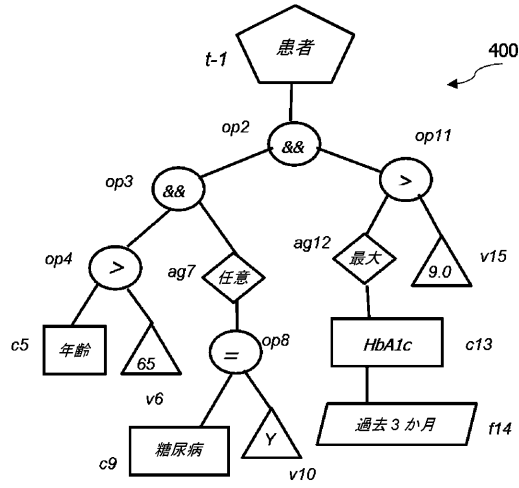
**T4: ラボ・テスト指数**

LName	LValue	LID*	...
HbA1C	10.5	L1	...
...	...	...	...

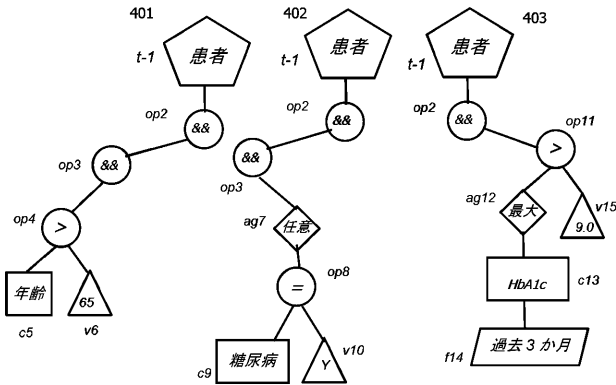
【 図 3 】



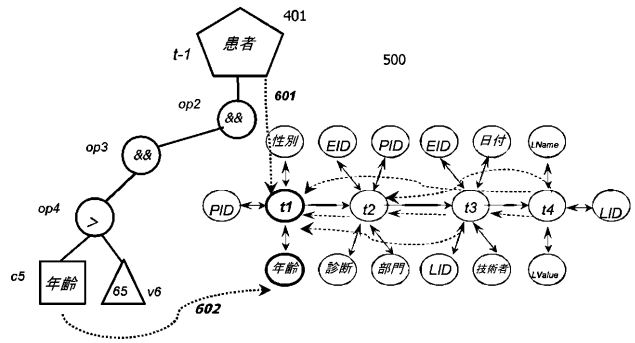
【 図 4 】



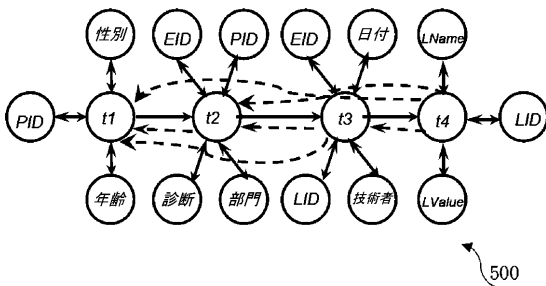
【 図 5 】



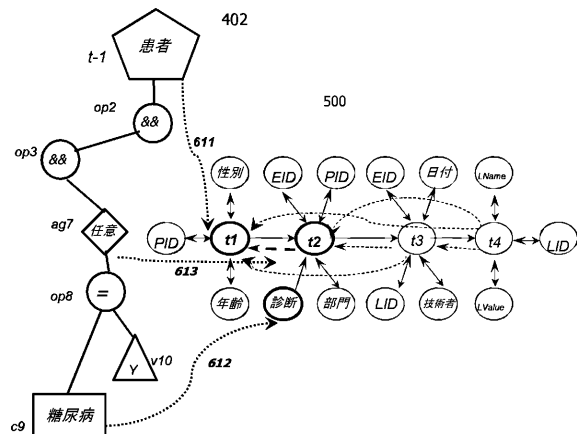
【 図 7 】



【 図 6 】



【 図 8 】







前記複数の単一概念のサブツリーの各単一概念のサブツリーごとに、  
前記知識ツリーの単一概念のサブツリーを選択し、  
前記単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、前記 E M R グラフのテーブル・ノードおよび少なくとも 1 つの属性ノードにマッピングすることによって、前記 E M R グラフの有効なサブグラフを識別し、  
前記単一概念のサブツリーおよび前記有効なサブグラフにしたがってサブクエリを構成することを含む、請求項 1 に記載の方法。

【請求項 3】

前記知識ツリーの前記単一概念のサブツリーは、概念ノードから始まりターゲット・ノードまでのパス内のすべてのノードと、概念ノードを備えていないノードの各々の任意の分岐とを備えたサブツリーである、請求項 2 に記載の方法。

【請求項 4】

前記 E M R グラフの前記有効なサブグラフは、前記 E M R グラフのサブグラフであり、前記知識ツリーのターゲット・ノードに一致するテーブル・ノードが、前記単一概念のサブツリーに存在し、前記単一概念のサブツリーにおける各属性ノードから、前記テーブル・ノードへの有向パスが存在する、請求項 2 に記載の方法。

【請求項 5】

前記単一概念のサブツリーの前記ターゲット・ノードおよび前記概念ノードをそれぞれ、前記 E M R グラフのテーブル・ノードおよび少なくとも 1 つの属性ノードにマッピングすることは、  
属性によるマッピング、  
属性値によるマッピング、および、  
少なくとも 1 つの追加のフィルタを適用することによるマッピングのうちの少なくとも 1 つを含む、請求項 2 に記載の方法。

【請求項 6】

前記 E M R テーブルのセットに対応する前記 E M R グラフを取得することは、前記 E M R テーブルのセットを、等価なグラフ表現に変換することを含む、請求項 1 に記載の方法。

【請求項 7】

前記少なくとも 1 つのクエリを使用して、前記 E M R テーブルのセットから、1 つまたは複数の臨床的特徴を抽出することをさらに含む、請求項 1 に記載の方法。

【請求項 8】

臨床知識に基づいて、臨床的特徴を、電子医療レコード ( E M R ) テーブルのセットから抽出するためのクエリを生成するためのデバイスであって、  
少なくとも 1 つのプロセッサと、  
前記少なくとも 1 つのプロセッサに動作可能に結合されたメモリと、  
前記メモリに記憶され、

臨床知識データのセットにしたがって構築された知識ツリーを取得することと、

前記 E M R テーブルのセットに対応する E M R グラフを取得することであって、前記 E M R グラフは、テーブル・ノードのセットと属性ノードのセットとを備え、前記テーブル・ノードのセットと前記属性ノードのセットとは、前記 E M R テーブルのセットにおける各 E M R テーブルの構造と、前記 E M R テーブルのセットの属性間の参照関係とを表す、前記 E M R グラフを取得することと、

前記知識ツリーおよび前記 E M R グラフに基づいて複数のサブクエリを生成することと、

前記知識ツリーにしたがって前記複数のサブクエリを組み合わせることによって、少なくとも 1 つのクエリを構成することと

を実行するために、前記少なくとも 1 つのプロセッサによって実行されるコンピュータ・プログラム命令のセットとを備えた、デバイス。

【請求項 9】

臨床知識に基づいて、臨床的特徴を、電子医療レコード（EMR）テーブルのセットから抽出するためのクエリを生成するためのコンピュータ・プログラムであって、

プロセッサに、

臨床知識データのセットにしたがって構築された知識ツリーを取得することと、

前記EMRテーブルのセットに対応するEMRグラフを取得することであって、前記EMRグラフは、テーブル・ノードのセットと属性ノードのセットとを備え、前記テーブル・ノードのセットと前記属性ノードのセットとは、前記EMRテーブルのセットにおける各EMRテーブルの構造と、前記EMRテーブルのセットの属性間の参照関係とを表す、前記EMRグラフを取得することと、

前記知識ツリーおよび前記EMRグラフに基づいて複数のサブクエリを生成することと

、

前記知識ツリーにしたがって前記複数のサブクエリを組み合わせることによって、少なくとも1つのクエリを構成することと、

を実行させるためのコンピュータ・プログラム。

【請求項10】

臨床知識に基づいて、臨床的特徴を、電子医療レコード（EMR）テーブルのセットから抽出するためのクエリを生成するための方法であって、

臨床知識データのセットにしたがって構築された知識ツリーを取得することと、

前記EMRテーブルのセットを、等価なグラフ表現に変換する、前記EMRテーブルのセットに対応するEMRグラフを取得することであって、前記EMRグラフは、テーブル・ノードのセットと属性ノードのセットとを備え、前記テーブル・ノードのセットと前記属性ノードのセットとは、前記EMRテーブルのセットにおける各EMRテーブルの構造と、前記EMRテーブルのセットの属性間の参照関係とを表す、前記EMRテーブルを取得することと、

前記知識ツリーの複数の単一概念のサブツリーを識別することに基づいて、前記知識ツリーおよび前記EMRグラフに基づいて複数のサブクエリを生成することと、

前記知識ツリーにしたがって前記複数のサブクエリを組み合わせることによって、少なくとも1つのクエリを構成することと、

前記EMRテーブルのセットを、等価なグラフ表現へ変換することを含む、方法。

【請求項11】

前記知識ツリーおよび前記EMRグラフに基づいて前記複数のサブクエリを生成することはさらに、

前記複数の単一概念のサブツリーの各単一概念のサブツリーごとに、

前記知識ツリーの単一概念のサブツリーを選択し、

前記単一概念のサブツリーのターゲット・ノードおよび概念ノードをそれぞれ、前記EMRグラフのテーブル・ノードおよび少なくとも1つの属性ノードにマッピングすることによって、前記EMRグラフの有効なサブグラフを識別し、

前記単一概念のサブツリーおよび前記有効なサブグラフにしたがってサブクエリを構成することを含む、請求項10に記載の方法。

## 【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International application No. <b>PCT/CN2017/109286</b>
<b>A. CLASSIFICATION OF SUBJECT MATTER</b> G06F 19/00(2018.01)i  According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>  Minimum documentation searched (classification system followed by classification symbols) G06F; G06Q  Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) WPI, EPODOC, CNPAT, CNKI: EMR, record, history, database, graph, tree?, graph?, table?, node?, character, know+, query, inquire		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2016306791 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 20 October 2016 (2016-10-20) description paragraphs [0003]-[0012], paragraphs [0034]-[0099], claims 12, figures 1-12	1,6-8,13-15,20-22,24
A	US 2004243552 A1 (DICTAPHONE CORPORATION) 02 December 2004 (2004-12-02) the whole document	1-25
A	US 2015286788 A1 (ARKOFF, HAROLD) 08 October 2015 (2015-10-08) the whole document	1-25
A	US 2004078236 A1 (MEDTAMIC HOLDINGS) 22 April 2004 (2004-04-22) the whole document	1-25
A	CN 105447609 A (INTERNATIONAL BUSINESS MACHINES CORPORATION) 30 March 2016 (2016-03-30) the whole document	1-25
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search <b>09 January 2018</b>		Date of mailing of the international search report <b>26 January 2018</b>
Name and mailing address of the ISA/CN <b>STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA 6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088 China</b> Facsimile No. (86-10)62019451		Authorized officer  <b>SUN, Juan</b>  Telephone No. (86-10)53961383

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.  
**PCT/CN2017/109286**

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2016306791	A1	20 October 2016	None			
US	2004243552	A1	02 December 2004	None			
US	2015286788	A1	08 October 2015	None			
US	2004078236	A1	22 April 2004	None			
CN	105447609	A	30 March 2016	US	2016063195	A1	03 March 2016

## フロントページの続き

(81) 指定国・地域 AP(BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, RU, TJ, TM), EP(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT

(74) 代理人 100112690

弁理士 太佐 種一

(72) 発明者 ハオ、ピボ

中華人民共和国 1 0 0 1 9 3 ベイジン ドンバイワン ウエスト・ロード 8 ジョングァンツン  
・ソフトウェア・パーク 1 9

(72) 発明者 フー、ガン

中華人民共和国 1 0 0 1 9 3 ベイジン ドンバイワン ウエスト・ロード 8 ジョングァンツン  
・ソフトウェア・パーク 1 9

(72) 発明者 リー、ジン

中華人民共和国 1 0 0 1 9 3 ベイジン ドンバイワン ウエスト・ロード 8 ジョングァンツン  
・ソフトウェア・パーク 1 9

(72) 発明者 スン、ウェン

中華人民共和国 1 0 0 1 9 3 ベイジン ドンバイワン ウエスト・ロード 8 ジョングァンツン  
・ソフトウェア・パーク 1 9

(72) 発明者 シエ、グオトン

中華人民共和国 1 0 0 1 9 3 ベイジン ドンバイワン ウエスト・ロード 8 ジョングァンツン  
・ソフトウェア・パーク 1 9

(72) 発明者 ユイ、イーキン

中華人民共和国 1 0 0 1 9 3 ベイジン ドンバイワン ウエスト・ロード 8 ジョングァンツン  
・ソフトウェア・パーク 1 9

F ターム(参考) 5L099 AA23