

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2010年4月15日(15.04.2010)

PCT

(10) 国際公開番号
WO 2010/041678 A1

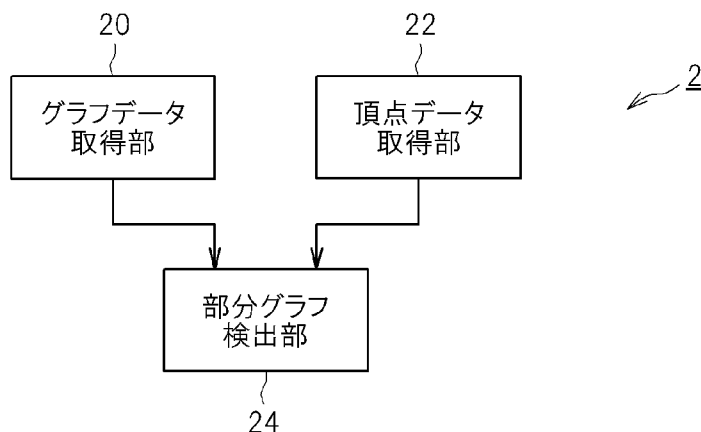
- (51) 国際特許分類:
G06F 17/30 (2006.01) G06F 19/00 (2006.01)
- (21) 国際出願番号: PCT/JP2009/067476
- (22) 国際出願日: 2009年10月7日(07.10.2009)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2008-260827 2008年10月7日(07.10.2008) JP
- (71) 出願人 (米国を除く全ての指定国について): 国立大学法人お茶の水女子大学(OCHANOMIZU UNIVERSITY) [JP/JP]; 〒1128610 東京都文京区大塚2丁目1番1号 Tokyo (JP).
- (72) 発明者: および
- (75) 発明者/出願人 (米国についてのみ): 瀬々 潤 (SESE, Jun) [JP/JP]; 〒1128610 東京都文京区大塚二丁目1番1号 国立大学法人お茶の水女子大学内 Tokyo (JP). 関 美緒 (SEKI, Mio) [JP/JP]; 〒1128610 東京都文京区大塚二丁目1番1号 国立大学法人お茶の水女子大学内 Tokyo (JP).
- (74) 代理人: 特許業務法人はるか国際特許事務所 (HARUKA PATENT & TRADEMARK ATTORNEYS); 〒1510051 東京都渋谷区千駄ヶ谷五丁目27番7号 日本プランズウィックビル9階 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (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, SE, SI, SK, SM, TR), OAPI (BF, BJ,

[続葉有]

(54) Title: PARTIAL GRAPH DETECTION DEVICE, PARTIAL GRAPH DETECTION METHOD, PROGRAM, DATA STRUCTURE OF DATA AND INFORMATION RECORDING MEDIUM

(54) 発明の名称: 部分グラフ検出装置、部分グラフ検出方法、プログラム、データのデータ構造、及び情報記憶媒体

[図6]



- 20 GRAPH DATA-ACQUIRING SECTION
- 22 VERTEX DATA-ACQUIRING SECTION
- 24 PARTIAL GRAPH-DETECTING SECTION

(57) Abstract: Provided is a partial graph detection device which enables, for example, acquisition of a list of genes showing changes in expression in common caused by the administration of specific drug(s) while considering the relationship among the genes (for example, the interacting relationship among the genes), acquisition of a list of proteins showing changes in expression in common caused by the administration of specific drug(s) while considering the relationship among the proteins (for example, the interacting relationship among the proteins), or acquisition of a list of users buying the same commercial product(s) while considering the relationship among the users (for example, friendships). A graph data-acquiring section (20) acquires graph data showing a graph including a plurality of vertexes. A vertex data-acquiring section (22) acquires vertex data wherein the vertexes are associated with information. Based on the graph data and the vertex data, a partial graph-detecting section (24) detects a partial graph which is a partial graph of the aforesaid

graph and in which the sets of information associated with the individual vertexes have a specific relationship.

(57) 要約:

[続葉有]

WO 2010/041678 A1

CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, 添付公開書類:
TD, TG).

— 国際調査報告 (条約第 21 条(3))

例えば、ある医薬又は医薬群の投与により共通して発現が変化する遺伝子のリストを遺伝子間の関係（例えば遺伝子間の相互作用関係）を考慮して取得すること、ある医薬又は医薬群の投与により共通して発現が変化するタンパク質のリストをタンパク質間の関係（例えば、タンパク質間の相互作用関係）を考慮して取得することや、同じ商品又は商品群を購入したユーザのリストをユーザ間の関係（例えば交友関係）を考慮して取得すること等が可能になる部分グラフ検出装置を提供する。グラフデータ取得部（20）は、複数の頂点を含むグラフを示すグラフデータを取得する。頂点データ取得部（22）は、頂点に情報を関連づけてなる頂点データを取得する。部分グラフ検出部（24）は、グラフデータと、頂点データと、に基づいて、グラフの部分グラフであって、該部分グラフの各頂点に関連付けられた情報が所定関係を有する部分グラフを検出する。

明 細 書

発明の名称：

部分グラフ検出装置、部分グラフ検出方法、プログラム、データのデータ構造、及び情報記憶媒体

技術分野

[0001] 本発明は、部分グラフ検出装置、部分グラフ検出方法、プログラム、データのデータ構造、及び情報記憶媒体に関する。

背景技術

[0002] 例えば、複数のユーザの各々に購入商品のリストを対応づけてなるデータに基づいて、ある商品又は商品群を購入したユーザのリストやその商品又は商品群を購入していないユーザのリストを取得することが行われている。また、例えば、複数の文献の各々にキーワードのリストを対応づけてなるデータに基づいて、あるキーワード又はキーワード群が設定された文献のリストを取得することが行われている。

[0003] 例えば、下記特許文献1には、複数のユーザの購入商品のリストを基に、ある商品又は商品群（例えば、購入された頻度、回数の大きい商品又は商品群）を購入していないユーザを特定することが開示されている。

先行技術文献

特許文献

[0004] 特許文献1：特開2007-287139号公報

発明の概要

発明が解決しようとする課題

[0005] 従来、例えば、同じ商品又は商品群を購入したユーザのリストをユーザ間の関係（例えば交友関係）を考慮して取得することはできなかった。また例えば、同じキーワード又はキーワード群が設定された文献のリストを文献間の関係（引用関係）を考慮して取得することはできなかった。

[0006] ところで、例えば、ある医薬又は医薬品群の投与により共通して発現が変化する遺伝子のリストを遺伝子間の関係を考慮して取得することができれば、創薬研究の上で有用な情報を得ることができる。例えば、発現の影響を互いに及ぼし合うような遺伝子間の関係（遺伝子間の相互作用関係）や、一方の遺伝子と他方の遺伝子とが同時に発現するような遺伝子間の関係を考慮して上記のリストを取得することができれば、例えば創薬ターゲットを絞り込む際の有用な情報を得ることができる。しかしながら、ある医薬又は医薬群の投与により共通して発現が変化する遺伝子のリストを遺伝子間の関係を考慮して取得することはできなかった。

[0007] また、例えば、ある医薬又は医薬品群の投与により共通して発現が変化するタンパク質のリストをタンパク質間の関係を考慮して取得することができれば、創薬研究の上で有用な情報を得ることができる。例えば、互いに結合して複合体を形成するようなタンパク質間の関係（タンパク質間の相互作用関係）を考慮して上記のリストを取得することができれば、例えば創薬ターゲットを絞り込む際の有用な情報を得ることができる。しかしながら、ある医薬又は医薬群の投与により共通して発現が変化するタンパク質のリストをタンパク質間の関係を考慮して取得することはできなかった。

[0008] 本発明は、上記課題に鑑みてなされたものであって、例えば、ある医薬又は医薬群の投与により共通して発現が変化する遺伝子のリストを遺伝子間の関係（例えば遺伝子間の相互作用関係）を考慮して取得すること、ある医薬又は医薬群の投与により共通して発現が変化するタンパク質のリストをタンパク質間の関係（例えば、タンパク質間の相互作用関係）を考慮して取得すること、同じ商品又は商品群を購入したユーザのリストをユーザ間の関係（例えば交友関係）を考慮して取得することや、同じキーワード又はキーワード群が設定された文献のリストを文献間の関係（引用関係）を考慮して取得すること等が可能になる部分グラフ検出装置、部分グラフ検出方法、プログラム、データのデータ構造、及び情報記憶媒体の提供をその目的とする。

課題を解決するための手段

- [0009] 上記課題を解決するために、本発明に係る部分グラフ検出装置は、複数の頂点と1又は複数のエッジとを含むグラフを示すグラフデータを取得するグラフデータ取得手段と、前記頂点に情報を関連づけてなる頂点データを取得する頂点データ取得手段と、前記グラフデータと、前記頂点データと、に基づいて、前記グラフの部分グラフであって、該部分グラフの各頂点に関連付けられた前記情報が所定関係を有する部分グラフを検出する部分グラフ検出手段と、を含むことを特徴とする。
- [0010] また、本発明に係る部分グラフ検出方法は、複数の頂点と1又は複数のエッジとを含むグラフを示すグラフデータを取得するグラフデータ取得ステップと、前記頂点に情報を関連づけてなる頂点データを取得する頂点データ取得ステップと、前記グラフデータと、前記頂点データと、に基づいて、前記グラフの部分グラフであって、該部分グラフに含まれる各頂点に関連付けられた前記情報が所定関係を有する部分グラフを検出する部分グラフ検出ステップと、を含むことを特徴とする。
- [0011] また、本発明に係るプログラムは、複数の頂点と1又は複数のエッジとを含むグラフを示すグラフデータを取得するグラフデータ取得手段、前記頂点に情報を関連づけてなる頂点データを取得する頂点データ取得手段、及び、前記グラフデータと、前記頂点データと、に基づいて、前記グラフの部分グラフであって、該部分グラフに含まれる各頂点に関連付けられた前記情報が所定関係を有する部分グラフを検出する部分グラフ検出手段、としてコンピュータを機能させることを特徴とする。
- [0012] また、本発明に係る情報記憶媒体は、上記プログラムを記録したコンピュータ読み取り可能な情報記憶媒体である。
- [0013] また、本発明に係るデータ構造は、複数の頂点と1又は複数のエッジとを含むグラフを示すグラフデータと、前記頂点に情報を関連づけてなる頂点データと、を含むことを特徴とする。また、本発明に係る情報記憶媒体は、複数の頂点と1又は複数のエッジとを含むグラフを示すグラフデータと、前記頂点に情報を関連づけてなる頂点データと、を含むデータを記録したコンピ

ユーザ読み取り可能な情報記憶媒体である。

[0014] 本発明は、複数の頂点と1又は複数のエッジとを含むグラフの部分グラフを検出する技術に関するものである。本発明では、前記グラフを示すグラフデータが取得される。例えば、各頂点が遺伝子一つ一つに対応し各エッジ（辺）が遺伝子間の関係を表すグラフを表すグラフデータが取得される。ここで、遺伝子間の関係とは、例えば、発現の影響を互いに及ぼし合うような遺伝子間の関係（遺伝子間の相互作用関係）や、一方の遺伝子と他方の遺伝子とが同時に発現するような遺伝子間の関係である。また、例えば、各頂点がタンパク質一つ一つに対応し、各エッジ（辺）がタンパク質間の関係を表すグラフデータが取得される。ここで、タンパク質間の関係とは、例えば、互いに結合して複合体を形成するようなタンパク質間の関係（タンパク質間の相互作用関係）である。また、例えば、各頂点がユーザー一人一人に対応し各エッジ（辺）がユーザ間の交友関係を表すグラフを示すグラフデータが取得される。なお、グラフデータは、例えば、隣接リスト、隣接行列として表される。また、本発明では、前記頂点に情報を関連づけてなる頂点データが取得される。そして、グラフデータと、頂点データと、に基づいて、前記部分グラフの各頂点に関連づけられた情報が所定関係を有する部分グラフが検出される。本発明によれば、グラフデータと、頂点データと、に基づいて部分グラフが検出される。その結果、例えば、ある医薬又は医薬群の投与により共通して発現が変化する遺伝子のリストを遺伝子間の関係（例えば、遺伝子間の相互作用関係）を考慮して取得することが可能になる。また、例えば、ある医薬又は医薬群の投与により共通して発現が変化するタンパク質のリストをタンパク質間の関係（例えば、タンパク質間の相互作用関係）を考慮して取得することが可能になる。また、例えば、同じ商品又は商品群を購入したユーザのリストをユーザ間の関係（例えば交友関係）を考慮して取得することが可能になる。

[0015] また、本発明の一態様では、前記情報は、少なくとも一種の要素を含む要素集合を示し、前記部分グラフ検出手段は、前記グラフの部分グラフであっ

て、該部分グラフの各頂点に関連づけられた要素集合の積集合が N ($N \geq 1$) 種以上の要素を含む部分グラフの検出を実行するようにしてもよい。例えば、各頂点が遺伝子一つ一つに対応し、かつ、要素集合が、遺伝子の発現を変化させる医薬又は医薬群に対応する場合、ある医薬又は医薬群の投与により共通して発現が変化する遺伝子のリストを遺伝子間の関係（例えば、遺伝子間の相互作用関係）を考慮して取得することが可能になる。また、例えば、各頂点がタンパク質一つ一つに対応し、かつ、要素集合が、タンパク質の発現を変化させる医薬又は医薬群に対応する場合、ある医薬又は医薬群の投与により共通して発現が変化するタンパク質のリストをタンパク質間の関係（例えば、タンパク質間の相互作用関係）を考慮して取得することが可能になる。また、例えば、各頂点がユーザー一人一人に対応し、かつ、要素集合が、ユーザが購入した商品又は商品群に対応する場合、同じ商品又は商品群を購入したユーザのリストをユーザ間の関係（例えば交友関係）を考慮して取得することが可能になる。

[0016] また、本発明の一態様では、前記部分グラフ検出手段は、「各ノードが前記グラフの頂点の集合を示す探索木であって、かつ、親ノードが示す頂点集合に当該頂点集合のうちの末端頂点に隣接する一つの頂点を新たな末端頂点として追加してなる頂点集合を当該親ノードの子ノードとして有する探索木」の各ノードを、深さ優先探索アルゴリズムに従った探索順序で判定対象とし、前記判定対象のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合が N 種以上の要素を含むか否かの判定を実行する判定手段と、前記判定手段の判定結果に基づいて前記検出を実行する手段と、を含み、前記判定手段は、前記探索木のうちの第1のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合が、「前記第1のノードが示す頂点集合のうちの末端頂点と同じ頂点を末端頂点とする頂点集合を示すノードであって、かつ、前記第1のノードよりも探索順序が前のノードである第2のノード」が示す頂点集合の各頂点に関連づけられた要素集合の積集合に含まれる場合、前記第1のノードの子孫のノードが示す頂点集合を前記判定対象とした前

記判定を実行しないようにしてもよい。

[0017] また、本発明の一態様では、前記部分グラフ検出手段は、前記判定が実行された場合に、前記判定対象の頂点集合の各頂点に関連づけられた要素集合の積集合を、基準集合として、当該判定対象の頂点集合のうちの末端頂点に関連づけて基準集合記憶手段に記憶させる手段を含み、前記判定手段は、前記第1のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合が、前記第1のノードが示す頂点集合のうちの末端頂点に関連づけて前記基準集合記憶手段に記憶される基準集合に含まれる場合、前記第1のノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないようにしてもよい。

[0018] また、本発明の一態様では、前記部分グラフ検出手段は、「各ノードが前記グラフの頂点の集合を示す探索木であって、かつ、親ノードが示す頂点集合に当該頂点集合のうちの末端頂点に隣接する一つの頂点を新たな末端頂点として追加してなる頂点集合を当該親ノードの子ノードとして有する探索木」の各ノードを、深さ優先又は幅優先探索アルゴリズムに従った探索順序で判定対象とし、前記判定対象のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合がN種以上の要素を含むか否かの判定を実行する判定手段と、前記判定手段の判定結果に基づいて前記検出を実行する手段と、を含み、前記判定手段は、前記探索木のうちの第1のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合が、「前記第1のノードが示す頂点集合のうちの末端頂点と同じ頂点を末端頂点とする頂点集合を示すノードであって、かつ、前記第1のノードが示す頂点集合を含む頂点集合を示す第2のノード」が示す頂点集合の各頂点に関連づけられた要素集合の積集合に含まれる場合、前記第1のノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないようにしてもよい。

[0019] また、本発明の一態様では、前記部分グラフ検出手段は、「各ノードが前記グラフの頂点の集合を示す探索木であって、かつ、親ノードが示す頂点集合に当該頂点集合のうちの末端頂点に隣接する一つの頂点を新たな末端頂点

として追加してなる頂点集合を当該親ノードの子ノードとして有する探索木」の各ノードを、深さ優先探索アルゴリズムに従った探索順序で判定対象とし、前記判定対象のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合がN種以上の要素を含むか否かの判定を実行する判定手段と、前記判定手段の判定結果に基づいて前記検出を実行する手段と、を含み、前記判定手段は、前記探索木のうちの第3のノードが示す頂点集合が一つの頂点のみを含み、かつ、当該頂点に関連づけられた要素集合が、「当該頂点を末端頂点とする頂点集合を示すノードであって、かつ、前記第3のノードよりも探索順序が前のノードである第4のノード」が示す頂点集合の各頂点に関連づけられた要素集合の積集合に含まれる場合、前記第3のノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないようにしてもよい。

[0020] また、本発明の一態様では、前記部分グラフ検出手段は、前記判定が実行された場合に、前記判定対象のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合を、基準集合として、前記判定対象のノードが示す頂点集合のうちの末端頂点に関連づけて基準集合記憶手段に記憶させる手段を含み、前記判定手段は、前記第3のノードが示す頂点集合が一つの頂点のみを含み、かつ、当該頂点に関連づけられた要素集合が、当該頂点に関連づけて前記基準集合記憶手段に記憶される基準集合に含まれる場合に、前記第3のノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないようにしてもよい。

[0021] また、本発明の一態様では、前記部分グラフ検出手段は、「各ノードが前記グラフの頂点の集合を示す探索木であって、かつ、親ノードが示す頂点集合に当該頂点集合のうちの末端頂点に隣接する一つの頂点を新たな末端頂点として追加してなる頂点集合を当該親ノードの子ノードとして有する探索木」の各ノードを、深さ優先又は幅優先探索アルゴリズムに従った探索順序で判定対象とし、前記判定対象のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合がN種以上の要素を含むか否かの判定を実行する判定

手段と、前記判定手段の判定結果に基づいて前記検出を実行する手段と、を含み、前記判定手段は、前記探索木のうちの第3のノードが示す頂点集合が一つの頂点のみを含み、当該頂点に関連づけられた要素集合が、当該頂点を末端頂点とする頂点集合を示す第4のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合に含まれる場合、前記第3のノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないようにしてもよい。

- [0022] また、本発明の一態様では、前記部分グラフ検出手段は、「各ノードが前記グラフの頂点の集合を示す探索木であって、かつ、親ノードが示す頂点集合に当該頂点集合の末端頂点に隣接する一つの頂点を新たな末端頂点として追加してなる頂点集合を当該親ノードの子ノードとして有する探索木」の各ノードを、深さ優先探索又は幅優先探索アルゴリズムに従った探索順序で判定対象とし、前記判定対象のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合がN種以上の要素を含むか否かの判定を実行する判定手段と、前記判定手段の判定結果に基づいて前記検出を実行する手段と、を含み、前記判定手段は、前記ノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合がN種以上の要素を含んでいない場合、当該ノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないようにしてもよい。

図面の簡単な説明

- [0023] [図1]本発明の実施形態に係る部分グラフ検出装置を含む部分グラフ検出システムの構成を示す図である。
- [図2]グラフデータが示すグラフの一例を示す図である。
- [図3]頂点データの内容の一例を示す図である。
- [図4]部分グラフCSGの一例を示す図である。
- [図5]頂点データの内容の一例を示す図である。
- [図6]本発明の実施形態に係る部分グラフ検出装置で実現される機能群を示す図である。

[図7] ツリーの一例を示す図である。

[図8] 基準集合テーブルの内容の一例を示す図である。

[図9] 本発明の実施形態に係る部分グラフ検出装置で実行される処理の一例を示すフローチャート図である。

[図10] 本発明の実施形態に係る部分グラフ検出装置で実行される処理の一例を示すフローチャート図である。

[図11] 本発明の実施形態に係る部分グラフ検出装置で実行される処理の一例を示すフローチャート図である。

発明を実施するための形態

[0024] 以下、本発明の実施形態について図面を参照しながら説明する。

[0025] [1. 部分グラフ検出システムの構成]

図1は、本発明の実施形態に係る部分グラフ検出装置2を含む部分グラフ検出システム6の構成を示す図である。同図に示すように、部分グラフ検出システム6は、部分グラフ検出装置2と、データベース4と、を含む。部分グラフ検出装置2と、データベース4と、は、ネットワークを介して相互に通信することが可能になっている。

[0026] [2. 部分グラフ検出装置のハードウェア構成]

部分グラフ検出装置2は、例えば、パーソナルコンピュータである。図1に示すように、部分グラフ検出装置2は、制御部8と、記憶部10と、通信部12と、表示部14と、操作部16と、を含む。各部は、バス18を介して通信可能に接続される。

[0027] 制御部8は、記憶部10に記憶されたプログラムに従って、各種情報処理を実行する。制御部8は、パーソナルコンピュータとしての処理を行うが、以下では本発明に関連する処理について主に説明する。制御部8によって実行される処理の具体的内容については後に述べる。

[0028] 記憶部10はメインメモリと不揮発性メモリとを含む。不揮発性メモリには制御部8によって実行されるプログラムが保持される。このプログラムは、例えばプログラム配信装置からネットワークを介してダウンロードされ、

不揮発性メモリに格納される。また、例えば光ディスク又はメモリカードなどのコンピュータ読み取り可能な情報記憶媒体から複写されて不揮発性メモリに格納される。メインメモリには、不揮発性メモリから読み出されるプログラムや制御部 8 の処理において必要となる種々のデータが必要に応じて書き込まれる。

[0029] 通信部 12 はネットワークインターフェイスカードを含み、例えば、ネットワークを介して到来したデータを制御部 8 に出力する。表示部 14 は例えばモニタを含み、制御部 8 から入力されるデータを表示する。操作部 16 はキーボードやマウスを含み、ユーザの操作内容を示すデータを制御部 8 に出力する。

[0030] [3. データベース]

データベース 4 は、例えばデータベースサーバのハードディスク（情報記憶媒体）に格納される。なお、データベースは部分グラフ検出装置 2 の不揮発性メモリ（情報記憶媒体）に格納されるようにしてもよい。

[0031] データベース 4 は、複数の頂点と、一又は複数の辺（エッジ）と、を含むグラフを示すグラフデータを記憶する。なお、グラフの「頂点」は「ノード」と呼ばれる場合もあるが、ここでは、後述のツリー（図 7 参照）のノードと区別するために「頂点」の語句を用いることとする。

[0032] グラフデータは、例えば、隣接リストや隣接行列を示すデータである。図 2 はグラフデータの一例を示す図である。図 2 に示すグラフ G は、6 つの頂点 v と、頂点 v と他の頂点 v とを連結する辺 e と、を含んでいる。つまり、図 2 に示すグラフ G は、6 つの頂点 $\{v_1, v_2, v_3, v_4, v_5, v_6\}$ と、8 つの辺 $\{e_{12}, e_{13}, e_{14}, e_{16}, e_{23}, e_{34}, e_{45}, e_{46}\}$ と、を含む。

[0033] 本実施形態の場合、図 2 に示すグラフ G において各頂点 v は一つ一つの遺伝子に対応しており、グラフ G は遺伝子間の関係を示している。ここで、遺伝子間の関係とは、例えば、発現の影響を互いに及ぼし合うような遺伝子間の関係（遺伝子間の相互作用関係）や、一方の遺伝子と他方の遺伝子とが同

時に発現するような遺伝子間の関係である。頂点 v_i と頂点 v_j とを連結する辺 e_{ij} は、頂点 v_i に対応する遺伝子と頂点 v_j に対応する遺伝子とが関係を有することを示す。例えば、図2に示すグラフGでは、辺 e_{12} が頂点 v_1 と頂点 v_2 とを連結しており、頂点 v_1 に対応する遺伝子と頂点 v_2 に対応する遺伝子とが関係を有することが表されている。なお、図2に示すグラフGでは、頂点 v_1 は辺 e_{14} によって頂点 v_4 と連結されており、さらに、頂点 v_4 は辺 e_{45} によって頂点 v_5 と連結されている。すなわち、頂点 v_1 に対応する遺伝子と頂点 v_5 に対応する遺伝子とが、頂点 v_4 に対応する遺伝子を介して間接的に関係を有することが表されている。

[0034] また、データベース4は、グラフGの各頂点 v に情報を関連づけてなる頂点データを記憶する。図3は頂点データの一例を示す図である。図3に示す頂点データは、頂点 v と、少なくとも一種のアイテム i (要素) を含むアイテム集合 $I(v)$ と、を関連づけてなるテーブルになっている。本実施形態の場合、アイテム集合 $I(v_i)$ に含まれる各アイテム i_j は、頂点 v_i に対応する遺伝子の発現を変化させる医薬に対応する。図3に示す例では、頂点 v_1 に対応づけられたアイテム集合 $I(v_1)$ は、アイテム i_1 (医薬A) と、アイテム i_2 (医薬B) と、アイテム i_3 (医薬C) と、アイテム i_4 (医薬D) と、を含んでおり、医薬A、医薬B、医薬C、又は医薬Dの投与により頂点 v_1 に対応する遺伝子の発現が変化することを表している。なお例えば、アイテム i を、遺伝子の発現を変化させる、医薬以外の要因に対応させるようにしてもよい。

[0035] ところで、図4に示すように、グラフGのうちの一部である部分グラフCSGでは、該部分グラフCSGに含まれる頂点 v_1 , v_3 , v_4 の各々に関連づけられたアイテム集合 $I(v_1)$, $I(v_3)$, $I(v_4)$ が、それぞれ、共通のアイテム集合 $\{i_1, i_2\}$ を含んでいる。これは、直接的又は間接的に関係を有する3つの遺伝子(頂点 v_1 , v_3 , v_4 に対応する3つの遺伝子)の発現が共通の医薬[医薬A、医薬B]の投与により変化することを表している。この場合、これら3つの遺伝子は、直接的又は間接的に関係を有し、かつ

、共通の医薬の投与により発現が変化するのであるから、これら3つの遺伝子の発現は、同一又は類似の医薬の投与により変化する可能性が高いと推測される。これら3つの遺伝子のような、直接的又は間接的に関係を有し、かつ、共通の医薬の投与により共通して発現が変化する遺伝子群を特定することができれば、例えば、創薬研究の上で有用である。例えば、創薬のターゲットとすべきを遺伝子を絞り込む際に有用である。

[0036] この点、部分グラフ検出装置2では、グラフGの部分グラフであって、該部分グラフの各頂点 v に関連づけられたアイテム集合 $I(v)$ が共通のアイテム集合を含むような部分グラフCSGが検出されるようになっている。その結果、直接的又は間接的に関係を有し、かつ、共通の医薬の投与により共通して発現が変化する遺伝子群が検出されるようになっている。

[0037] なお、グラフGの各頂点 v が一つ一つのタンパク質に対応し、かつ、頂点データにおいて各アイテム i_j が、頂点 v_j に対応するタンパク質の発現を変化させる医薬に対応するようにしてもよい。

[0038] この場合、図2に示すグラフGは、タンパク質間の関係を示し、頂点 v_i と頂点 v_j とを連結する辺 e_{ij} は、頂点 v_i に対応するタンパク質と頂点 v_j に対応するタンパク質とが関係を有することを示すことになる。ここで、タンパク質間の関係とは、例えば、互いに結合して複合体を形成するようなタンパク質間の関係（タンパク質間の相互作用関係）である。例えば、図2に示すグラフGの場合、辺 e_{12} が頂点 v_1 と頂点 v_2 とを連結しているため、頂点 v_1 に対応するタンパク質と頂点 v_2 に対応するタンパク質とが関係を有することが表されていることになる。また、図2に示すグラフGの場合、頂点 v_1 は辺 e_{14} によって頂点 v_4 と連結されており、さらに、頂点 v_4 は辺 e_{45} によって頂点 v_5 と連結されているので、頂点 v_1 に対応するタンパク質と頂点 v_5 に対応するタンパク質とが、頂点 v_4 に対応するタンパク質を介して間接的に関係を有することが表されていることになる。

[0039] また、この場合、データベース4には、図3に示す頂点データと同様の頂点データが記憶される。図3に示す頂点データでは、頂点 v_1 に対応づけられ

たアイテム集合 $I(v_1)$ に、アイテム i_1 (医薬A) と、アイテム i_2 (医薬B) と、アイテム i_3 (医薬C) と、アイテム i_4 (医薬D) と、が含まれているため、医薬A、医薬B、医薬C、又は医薬Dの投与により頂点 v_1 に対応するタンパク質の発現が変化することが表されていることになる。なお例えば、アイテム i を、タンパク質の発現を変化させる、医薬以外の要因 (例えば、ヒートショック、窒素枯渇環境など) に対応させるようにしてもよい。

[0040] 先述のように (図4参照)、グラフGのうちの一部である部分グラフCSGでは、該部分グラフCSGに含まれる頂点 v_1 , v_3 , v_4 の各々に関連づけられたアイテム集合 $I(v_1)$, $I(v_3)$, $I(v_4)$ が、それぞれ、共通のアイテム集合 $\{i_1, i_2\}$ を含んでいる。これは、直接的又は間接的に関係を有する3つのタンパク質 (頂点 v_1 , v_3 , v_4 に対応する3つのタンパク質) の発現が、共通の医薬 [医薬A、医薬B] の投与により変化することを表している。この場合、これら3つのタンパク質は、直接的又は間接的に関係を有し、且つ、共通の医薬の投与により発現が変化するので、これら3つのタンパク質の発現は、同一又は類似の医薬の投与により変化する可能性が高いと推測される。これら3つのタンパク質のような、直接的又は間接的に関係を有し、かつ、共通の医薬の投与により共通して発現が変化するタンパク質群を特定することができれば、創薬研究の上で有用である。例えば、創薬のターゲットとすべきタンパク質を絞り込む際に有用である。

[0041] この点、先述のように、部分グラフ検出装置2では、グラフGの部分グラフであって、該部分グラフの各頂点 v に関連づけられたアイテム集合 $I(v)$ が共通のアイテム集合を含むような部分グラフCSGが検出されるようになっている。その結果、例えば、グラフGの各頂点 v が一つのタンパク質に対応し、かつ、各アイテム i_i が、頂点 v_i に対応するタンパク質の発現を変化させる医薬に対応するような場合には、直接的又は間接的に関係を有し、かつ、共通の医薬の投与により共通して発現が変化するタンパク質群が検出されるようになる。

[0042] また、グラフGの各頂点 v が一人一人のユーザに対応し、頂点データにお

いて各アイテム i_j が、頂点 v_j に対応するユーザが購入した商品に対応するようによい。

[0043] この場合、図2に示すグラフGはユーザ間の交友関係を示し、頂点 v_i と頂点 v_j とを連結する辺 e_{ij} は、頂点 v_i に対応するユーザと頂点 v_j に対応するユーザとが友人関係にあることを示すことになる。例えば、図2に示すグラフGの場合、辺 e_{12} が頂点 v_1 と頂点 v_2 とを連結しているため、頂点 v_1 に対応するユーザと頂点 v_2 に対応するユーザとが友人関係にあることが表されていることになる。また、図2に示すグラフGの場合、頂点 v_1 は辺 e_{14} によって頂点 v_4 と連結され、さらに、頂点 v_4 は辺 e_{45} によって頂点 v_5 と連結されているため、頂点 v_1 に対応するユーザと頂点 v_5 に対応するユーザとが、頂点 v_4 に対応するユーザを介して間接的に交友関係を有することが表されていることになる。

[0044] 図5に、各アイテム i_j を、頂点 v_j に対応するユーザが購入した商品に対応させた場合の頂点データの一例を示す。図5に示す例では、頂点 v_1 に対応づけられたアイテム集合 $I(v_1)$ に、アイテム i_1 (りんご) と、アイテム i_2 (みかん) と、アイテム i_3 (ぶどう) と、アイテム i_4 (ナシ) と、が含まれているため、頂点 v_1 に対応するユーザがりんご、みかん、ぶどう及びナシを購入したことが表されていることになる。なお、ここでは、アイテム $i_1 \sim i_5$ が同種の商品(果物)になっているが、アイテム $i_1 \sim i_5$ は異なる種類の商品であってもよい。

[0045] 先述のように(図4参照)、グラフGのうちの一部である部分グラフCSGでは、該部分グラフCSGに含まれる頂点 v_1, v_3, v_4 の各々に関連づけられたアイテム集合 $I(v_1), I(v_3), I(v_4)$ が、それぞれ、共通のアイテム集合 $\{i_1, i_2\}$ を含んでいる。これは、直接的又は間接的に交友関係を有する3名のユーザ(頂点 v_1, v_3, v_4 に対応する3名のユーザ)が共通の商品{りんご、みかん}を購入していることを表している。この場合、共通の商品{りんご、みかん}又はこれらの商品に関連する他の商品に関する情報が、頂点 v_1, v_3, v_4 に対応する3名のユーザの間で口コミによっ

て伝わる可能性が高いと推測される。これら3名のユーザのような、直接的又は間接的に交友関係を有し、かつ、共通の商品を購入したユーザ群を特定することができれば、例えば、商品の広告宣伝のターゲットとすべきユーザを好適に絞り込むことができるようになり、その結果として、商品の広告宣伝が効率的に行えるようになる。

[0046] この点、部分グラフ検出装置2では、グラフGの部分グラフであって、該部分グラフの各頂点vに関連づけられたアイテム集合I(v)が共通のアイテム集合を含むような部分グラフCSGが検出されるようになっている。その結果、直接的又は間接的に交友関係にあり、かつ、共通の商品を購入したユーザ群が検出されるようになっている。

[0047] 以下、各頂点vに関連づけられたアイテム集合I(v)が共通のアイテム集合を含むような部分グラフCSGを検出するための技術について説明する。なお、以下では、各頂点vに関連づけられたアイテム集合I(v)が共通のアイテム集合を含むような部分グラフCSGのことを、共通部分グラフCSGと記載する。

[0048] [4. 機能ブロック]

図6は、部分グラフ検出装置2で実現される機能のうち、本発明に関連するものを主として示す機能ブロック図である。同図に示すように、部分グラフ検出装置2は、グラフデータ取得部20と、頂点データ取得部22と、部分グラフ検出部24と、を機能的に含む。これらの機能は、記憶部10から読み出された上記プログラムに従って制御部8が動作することによって実現される。

[0049] [4-1. グラフデータ取得部]

グラフデータ取得部20は、例えば制御部8と通信部12とを主として実現される。グラフデータ取得部20は、複数の頂点を含むグラフを示すグラフデータを取得する。本実施形態の場合、グラフデータ取得部20は、例えば図2に示すようなグラフGのグラフデータをデータベース4から読み出して取得する。

[0050] [4-2. 頂点データ取得部]

頂点データ取得部 22 は、例えば制御部 8 と通信部 12 とを主として実現される。頂点データ取得部 22 は、グラフに含まれる複数の頂点の各々に情報を対応づけてなる頂点データを取得する。ここで、「情報」とは、グラフ G の頂点間の接続関係を示す情報以外の情報である。例えば、頂点が遺伝子に対応する場合、上記の「情報」とは、遺伝子に関する情報である。ここで、「遺伝子に関する情報」とは、例えば、遺伝子の発現を変化させるような医薬を示す情報である。また例えば、頂点がタンパク質に対応する場合、上記の「情報」とは、タンパク質に関する情報である。ここで、「タンパク質に関する情報」とは、例えば、タンパク質の発現を変化させるような医薬を示す情報である。さらに例えば、頂点がユーザに対応する場合、上記の「情報」とは、ユーザに関する情報である。ここで、「ユーザに関する情報」とは、例えば、ユーザの属性を示す情報や、ユーザが購入した商品を示す情報である。本実施形態の場合、頂点データ取得部 22 は、例えば図 3 又は図 5 に示すような頂点データをデータベース 4 から読み出して取得する。

[0051] [4-3. 部分グラフ検出部]

部分グラフ検出部 24 は、例えば制御部 8 を主として実現される。部分グラフ検出部 24 は、グラフデータと、頂点データと、に基づいて、グラフの部分グラフであって、該部分グラフの各頂点に関連づけられた情報が所定関係を有するような部分グラフを検出する。本実施形態の場合、部分グラフ検出部 24 は、グラフ G の部分グラフであって、該部分グラフの各頂点 v に関連づけられたアイテム集合 $I(v)$ の積集合が N (N は正の整数) 種以上のアイテム (商品) を含む共通部分グラフ CSG を検出する。具体的には、例えば、「共通の N 種以上の医薬の投与により共通して発現が変化する複数の遺伝子」に対応する複数の頂点 v を含む共通部分グラフ CSG が検出される。また例えば、「共通の N 種以上の医薬の投与により共通して発現が変化する複数のタンパク質」に対応する複数の頂点 v を含む共通部分グラフ CSG が検出される。また例えば、共通の N 種以上の商品を購入した複数のユーザ

に対応する複数の頂点 v を含む共通部分グラフ CSG が検出される。

[0052] なお本実施形態では、特に、閉じている共通部分グラフ CSG （以下、閉共通部分グラフ $CCSG$ と記載する）が検出される。閉共通部分グラフ $CCSG$ とは、共通部分グラフ CSG であって、該共通部分グラフ CSG に含まれる頂点 v に隣接しているどの頂点 v を該共通部分グラフ CSG に追加しても、共通アイテム集合が変わってしまうような共通部分グラフ CSG のことである。

[0053] 以下、閉共通部分グラフ $CCSG$ の具体例について説明する。ここでは、共通アイテム集合に2種以上のアイテムが含まれる部分グラフが共通部分グラフ CSG であることとして説明する。例えば、部分グラフ $\{v_1, v_4\}$ の共通アイテム集合は、 $\{i_1, i_2\}$ であるから（図3参照）、部分グラフ $\{v_1, v_4\}$ は共通部分グラフ CSG である。しかし、部分グラフ $\{v_1, v_4\}$ に頂点 v_4 に隣接する頂点 v_3 （図1、図4参照）を追加しても、共通アイテム集合は $\{i_1, i_2\}$ のままで変化しないから、部分グラフ $\{v_1, v_4\}$ は閉共通部分グラフ $CCSG$ ではない。一方、部分グラフ $\{v_1, v_4\}$ に頂点 v_3 が追加された部分グラフ $\{v_1, v_3, v_4\}$ の共通アイテム集合は $\{i_1, i_2\}$ であるから（図3参照）、部分グラフ $\{v_1, v_3, v_4\}$ は共通部分グラフ CSG である。また、部分グラフ $\{v_1, v_3, v_4\}$ に、頂点 v_1, v_3, v_4 のいずれかに隣接する頂点 v_2, v_5, v_6 のいずれを追加しても、共通アイテム集合は $\{i_1, i_2\}$ から $\{i_1\}$ に変化するので、部分グラフ $\{v_1, v_3, v_4\}$ は、閉共通部分グラフ $CCSG$ である。

[0054] 共通アイテム集合が同じ場合、閉共通部分グラフ $CCSG$ の頂点数は、閉じていない共通部分グラフ CSG の頂点数よりも多くなる。例えば、部分グラフ $\{v_1, v_4\}$ の共通アイテム集合と、部分グラフ $\{v_1, v_3, v_4\}$ の共通アイテム集合とはともに $\{i_1, i_2\}$ であるが、閉共通部分グラフ $CCSG$ である部分グラフ $\{v_1, v_3, v_4\}$ の方が、閉共通部分グラフ $CCSG$ ではない部分グラフ $\{v_1, v_4\}$ よりも頂点の数が多い。

[0055] より多くの頂点を含む共通部分グラフ CSG を検出することができれば、

例えば、直接的又は間接的に関係を有し、かつ、共通の医薬の投与により共通して発現が変化する遺伝子群として、より遺伝子数が多いものが検出されるようになるので、より多くの遺伝子に影響を及ぼす医薬を同定することができ、例えば創薬の際に有用である。また、例えば、直接的又は間接的に関係を有し、かつ、共通の医薬の投与により共通して発現が変化するタンパク質群として、よりタンパク質数が多いものが検出されるようになるので、より多くのタンパク質に影響を及ぼす医薬を同定することができ、例えば創薬の際に有用である。また、例えば、直接的又は間接的に交友関係を有し、かつ、共通の商品を購入したユーザ群として、よりユーザ数が多いものが検出されるようになり、その結果として、商品の広告宣伝をより効率的に行えるようになる。そこで、本実施形態では、より多くの頂点を含む共通部分グラフCSGを得るために、特に閉共通部分グラフCCSGを検出するようにしている。部分グラフ検出部24の動作の詳細については後述する。

[0056] [5. 処理]

部分グラフ検出装置2で実行される、共通部分グラフCSG（特に閉共通部分グラフCCSG）を検出するための処理について説明する。なお、以下では、上記のNの値が2である場合、すなわち、各頂点vに関連づけられたアイテム集合I(v)の積集合が二種以上のアイテムiを含むような共通部分グラフCSGを検出する場合について説明する。

[0057] [5-1. 処理の概要]

まず処理の概要について説明する。図7は処理の概要を説明するための図である。部分グラフ検出装置2では、深さ優先探索アルゴリズムを用いて、図7に示すようなツリーT（探索木）が生成されることによって共通部分グラフCSG（特に閉共通部分グラフCCSG）が検出される。図7に示すツリーTは、グラフGの各頂点 $v_1 \sim v_6$ に対応するルートノード n_r ($n_{r1} \sim n_{r6}$)と、ルートノード $n_{r1} \sim n_{r6}$ よりも下位のノード n_c ($n_{c1} \sim n_{cM}$)と、を含む。また、図に示すツリーTの各ノードnには、そこまでの経路に含まれる頂点vの集合を示す頂点集合V(n)と、頂点集合V(n)のうちの

全ての頂点 v が有する共通のアイテム i の集合を示す共通アイテム集合 $I(n)$ と、が含まれている。図7では、各ノード n の左側領域に頂点集合 $V(n)$ が表され、右側領域に共通アイテム集合 $I(n)$ が表されている。なお、以下、ツリー T のノード n に含まれる頂点集合 $V(n)$ のことを、ツリー T のノード n の頂点集合 $V(n)$ と呼び、ツリー T のノード n に含まれるアイテム集合 $I(n)$ のことを、ツリー T のノード n のアイテム集合 $I(n)$ と呼ぶ。

[0058] 同図に示すツリー T では、各ノード（ノード n_x とする）の頂点集合 $V(n_x)$ に他の頂点 v が一つ追加された頂点集合が、その子のノード n_{x+1} の頂点集合 $V(n_{x+1})$ になっている。具体的には、ノード n_x の頂点集合 $V(n_x)$ に当該頂点集合 $V(n_x)$ のうちの末端頂点に隣接する一つの頂点が新たな末端頂点として追加された頂点集合が、ノード n_{x+1} の頂点集合 $V(n_{x+1})$ になっている。

[0059] 例えば、ノード n_{r1} の頂点集合 $V(n_{r1})$ である $\{v_1\}$ に、末端頂点である頂点 v_1 に隣接する頂点 v_2 が追加された頂点集合 $\{v_1, v_2\}$ が、ノード n_{r1} の子のノード n_{c1} の頂点集合 $V(n_{c1})$ になっている。この場合、追加された頂点 v_2 が頂点集合 $V(n_{c1})$ の末端頂点になっている。また例えば、ノード n_{c1} の頂点集合 $V(n_{c1})$ である $\{v_1, v_2\}$ に、末端頂点である頂点 v_2 に隣接する頂点 v_3 が追加された頂点集合 $\{v_1, v_2, v_3\}$ が、ノード n_{c1} の子のノード n_{c2} の頂点集合 $V(n_{c2})$ になっている。この場合、追加された頂点 v_3 が頂点集合 $V(n_{c2})$ の末端頂点になっている。このように、親のノード n_x の頂点集合 $V(n_x)$ に当該頂点集合 $V(n_x)$ の末端頂点に隣接する頂点が追加された頂点集合が、子のノード n_{x+1} の頂点集合 $V(n_{x+1})$ になっている。

[0060] 各ノード n の頂点集合 $V(n)$ に含まれる各頂点は、当該頂点集合 $V(n)$ に含まれる他のいずれかの頂点と隣接するようになっており、各ノード n の頂点集合 $V(n)$ は、当該頂点集合 $V(n)$ を含む部分グラフを表していることになる。

[0061] 部分グラフ検出部 24 は、深さ優先探索アルゴリズムに従った探索順序で、ツリー T の各ノード n を判定対象とし、判定対象のノード n の頂点集合 $V(n)$ が複数の頂点 v を含む場合に、当該頂点集合 $V(n)$ が示す部分グラフが共通部分グラフ CSG であるか否かの判定を実行する。すなわち、頂点集合 $V(n)$ の各頂点 v に関連付けられたアイテム集合 $I(v)$ の積集合が二種以上のアイテムを含むか否かの判定を実行する。こうして、部分グラフ検出部 24 は、ツリー T に基づいて共通部分グラフ CSG を検出する。なお、ここでは、図 7 に示すようなツリー T を生成するために深さ優先探索アルゴリズムを用いているが、幅優先探索アルゴリズムを用いるようにしてもよい。

[0062] [5-2. 基準集合テーブルについて]

部分グラフ検出装置 2 では、無駄な処理を実行しないようにするためのデータとして、基準集合テーブルが記憶される。ここで、基準集合テーブルについて説明しておく。

[0063] まず、基準集合テーブルを設ける意義について説明しておく。あるノード（ノード n_x とする）と他ノード（ノード n_y とする）との間に以下の 3 つの関係がすべて成立する場合、ノード n_x の子孫のノード（ノード n_{xd} とする）の頂点集合 $V(n_{xd})$ を含む部分グラフは閉部分グラフにはならない。

[0064] （関係 1）ノード n_x の頂点集合 $V(n_x)$ の末端頂点と、ノード n_y の頂点集合 $V(n_y)$ の末端頂点と、が共通すること。

（関係 2）ノード n_y がノード n_x よりも前に判定対象とされること。

（関係 3）ノード n_y の共通アイテム集合 $I(n_y) \supseteq$ ノード n_x の共通アイテム集合 $I(n_x)$ であること。

[0065] ノード n_{xd} が示す頂点集合 $V(n_{xd})$ を含む部分グラフが閉部分グラフにならないということは、ノード n_{xd} が示す頂点集合 $V(n_{xd})$ を含む部分グラフは閉共通部分グラフ CCSG にはなり得ないということである。そこで、本実施形態では、部分グラフ検出部 24 は、あるノード n_x （第 1 のノード）と他ノード n_y （第 2 のノード）との間に関係 1 ~ 3 がすべて成立する場合

、ノード n_x の子孫のノード n_{x_d} の頂点集合 $V(n_{x_d})$ を判定対象として当該頂点集合 $V(n_{x_d})$ が表す部分グラフが共通部分グラフCSGであるか否かを判定しないようになっている。その結果として、ツリー T のすべてのノード n を判定対象として、判定対象のノード n の頂点集合 $V(n)$ が表す部分グラフが共通部分グラフであるか否かの判定を実行しなくても、閉共通部分グラフCCSGが検出可能になり、処理負荷の軽減が図られる。

[0066] ここで、あるノード n_x と他ノード n_y との間に関係1~3がすべて成立する場合にノード n_x の子孫のノード n_{x_d} の頂点集合 $V(n_{x_d})$ が表す部分グラフが閉部分グラフにならない理由を説明する。

[0067] まず、ノード n_y がノード n_x よりも前に生成されるノードである場合（関係2）、ノード n_x の頂点集合 $V(n_x)$ の末端頂点と、ノード n_y の頂点集合 $V(n_y)$ の末端頂点と、が共通していれば（関係1）、「ノード n_y の頂点集合 $V(n_y) \supseteq$ ノード n_x の頂点集合 $V(n_x)$ 」の関係（以下、この関係を「関係4」と記載する。）が成立する。例えば、図7において、ノード $n_{c_{10}}$ と、ノード $n_{c_{10}}$ よりも前に生成されるノード n_{c_7} と、の間では、「ノード n_{c_7} の頂点集合 $V(n_{c_7}) \supseteq$ ノード $n_{c_{10}}$ の頂点集合 $V(n_{c_{10}})$ 」が成立している。

[0068] 従って、ノード n_x の頂点集合 $V(n_x)$ にその末端頂点に隣接する頂点 v が追加された頂点集合を、ノード n_x の子孫のノード n_{x_d} の頂点集合 $V(n_{x_d})$ とし、ノード n_y の頂点集合 $V(n_y)$ に同じ頂点 v が追加された頂点集合を、ノード n_y の子孫のノード n_{y_d} の頂点集合 $V(n_{y_d})$ とした場合、ノード n_{x_d} の頂点集合 $V(n_{x_d})$ と、ノード n_{y_d} の頂点集合 $V(n_{y_d})$ との間には、「ノード n_{y_d} の頂点集合 $V(n_{y_d}) \supseteq$ ノード n_{x_d} の頂点集合 $V(n_{x_d})$ 」の関係（以下、この関係を「関係5」と記載する。）が成立する。例えば、図7において、ノード n_{c_7} の子孫のノード n_{c_8} と、ノード $n_{c_{10}}$ の子孫のノード $n_{c_{11}}$ と、の間には、「ノード n_{c_8} の頂点集合 $V(n_{c_8}) \supseteq$ ノード $n_{c_{11}}$ の頂点集合 $V(n_{c_{11}})$ 」が成立している。

[0069] ここで、上記関係5が成立する場合、ノード n_{y_d} の頂点集合 $V(n_{y_d})$ の

各頂点に関連付けられたアイテム集合の積集合（すなわち、共通アイテム集合 $I(n_{y_d})$ ）に含まれるアイテムの数が、ノード n_{x_d} の頂点集合 $V(n_{x_d})$ の各頂点に関連付けられたアイテム集合の積集合（すなわち、共通アイテム集合 $I(n_{x_d})$ ）に含まれるアイテムの数よりも多くなることはない。頂点集合に含まれる頂点の数が多くなるほど、その頂点集合の各頂点に関連付けられたアイテム集合の積集合に含まれるアイテムの数が減るからである。このため、上記関係 5 が成立する場合、「ノード n_{y_d} の共通アイテム集合 $I(n_{y_d}) \subseteq$ ノード n_{x_d} の共通アイテム集合 $I(n_{x_d})$ 」の関係（以下、この関係を「関係 6」と記載する。）が成立することになる。

[0070] 一方、「ノード n_y の共通アイテム集合 $I(n_y) \supseteq$ ノード n_x の共通アイテム集合 $I(n_x)$ 」である場合（関係 3）、ノード n_{x_d} の頂点集合 $V(n_{x_d})$ と、ノード n_{y_d} の頂点集合 $V(n_{y_d})$ とは、それぞれ、頂点集合 $V(n_{x_d})$ 及び頂点集合 $V(n_{y_d})$ のそれぞれに同じ頂点 v が追加された頂点集合であるため、ノード n_{x_d} の共通アイテム集合 $I(n_{x_d})$ と、ノード n_{y_d} の共通アイテム集合 $I(n_{y_d})$ と、の間には、「ノード n_{y_d} の共通アイテム集合 $I(n_{y_d}) \supseteq$ ノード n_{x_d} の共通アイテム集合 $I(n_{x_d})$ 」の関係（以下、この関係を「関係 7」と記載する。）が成立することになる。

[0071] 以上のように、上記関係 1～3 がすべて成立する場合、上記関係 6, 7 が同時に成立することになる。上記関係 6, 7 が同時に成立する場合とは、ノード n_{x_d} の共通アイテム集合 $I(n_{x_d})$ と、ノード n_{y_d} の共通アイテム集合 $I(n_{y_d})$ とが、「ノード n_{y_d} の共通アイテム集合 $I(n_{y_d}) =$ ノード n_{x_d} の共通アイテム集合 $I(n_{x_d})$ 」の関係を有する場合である。このため、上記関係 1～3 がすべて成立する場合には、ノード n_{y_d} の共通アイテム集合 $I(n_{y_d})$ と、ノード n_{x_d} の共通アイテム集合 $I(n_{x_d})$ と、の間に、「ノード n_{y_d} のアイテム集合 $I(n_{y_d}) =$ ノード n_{x_d} のアイテム集合 $I(n_{x_d})$ 」の関係が成立することになる。これは、ノード n_{x_d} の頂点集合 $V(n_{x_d})$ に、ノード n_{y_d} の頂点集合 $V(n_{y_d})$ に含まれ、かつ、ノード n_{x_d} の頂点集合 $V(n_{x_d})$ に含まれない頂点 v を追加しても、共通アイテム集合が変わらな

いことを示している。つまり、あるノード n_x と他のノード n_y との間に上記関係 1～3 がすべて成立する場合に、ノード n_x の子孫のノード n_{x_d} の頂点集合 $V(n_{x_d})$ が表す部分グラフは閉部分グラフにならないということになる。

[0072] 例えば、図 7 に示す場合、ノード n_{c10} の子のノード n_{c11} の共通アイテム集合 $I(n_{c11})$ と、ノード n_{c7} の子のノード n_{c8} の共通アイテム集合 $I(n_{c8})$ とは、ともに $\{i_1\}$ であり、等しくなっている。そして、ノード n_{c8} の頂点集合 $V(n_{c8})$ に含まれ、かつ、ノード n_{c11} の頂点集合 $V(n_{c11})$ に含まれない頂点 v_3 をノード n_{c11} の頂点集合 $V(n_{c11})$ に追加しても共通アイテム集合は $\{i_1\}$ のままで変化しない。よって、頂点集合 $V(n_{c11})$ が表す部分グラフは閉部分グラフにはならない。

[0073] なお、上記関係 1 と上記関係 2 とが成立する場合、上記関係 4 が成立することは以上で述べた。そのため、「あるノード n_x と他ノード n_y との間に上記関係 1, 3, 4 がすべて成立する場合に、ノード n_x の子のノード n_{c_x} (言い換えれば、ノード n_{c_x} の頂点集合 $V(n_{c_x})$) を判定対象として上記の判定を行わない」とは、「ノード n_x とノード n_y との間に上記関係 1, 3, 4 がすべて成立する場合に、頂点集合 $V(n_{x_d})$ を判定対象として上記の判定を行わない」ともいうことができる。

[0074] ところで、ツリー T の各ノードを幅優先探索アルゴリズムに従った探索順序で判定対象とする場合にも、あるノード n_x (第 1 のノード) と他のノード n_y (第 2 のノード) との間に、上記関係 1、上記関係 3、及び下記関係 2' がすべて成立する場合に、ノード n_x の子孫のノード n_{x_d} の頂点集合 $V(n_{x_d})$ が表す部分グラフが共通部分グラフ CSG であるか否かを判定しないようにすることが可能である。

(関係 2') ノード n_x がノード n_y よりも前に判定対象とされること。

[0075] このようにすれば、ツリー T の各ノードを幅優先探索アルゴリズムに従った探索順序で判定対象とする場合においても、すべての頂点集合 $V(n)$ を判定対象として当該頂点集合 $V(n)$ が表す部分グラフが共通部分グラフ C

SGであるか否かを判定しなくても、閉共通部分グラフCCSGが検出可能になる。なぜなら、ノード n_x がノード n_y よりも前に判定対象にされるノードである場合（関係2'）、幅優先探索アルゴリズムの探索順序を考慮すると、ノード n_x の頂点集合 $V(n_x)$ の末端頂点と、ノード n_y の頂点集合 $V(n_y)$ の末端頂点と、が共通しているのなら（関係1）、上記関係4「ノード n_y の頂点集合 $V(n_y) \supseteq$ ノード n_x の頂点集合 $V(n_x)$ 」が成立する。従って、上記関係1、上記関係3、及び上記関係4が成立するのだから、深さ優先探索アルゴリズムの場合と同様に、上記関係5~7が成立することになる。また、その結果、「ノード n_{y_d} のアイテム集合 $I(n_{y_d}) =$ ノード n_{x_d} のアイテム集合 $I(n_{x_d})$ 」の関係が成立することとなり、ノード n_{x_d} の頂点集合 $V(n_{x_d})$ に、ノード n_{y_d} の頂点集合 $V(n_{y_d})$ に含まれ、かつ、ノード n_{x_d} の頂点集合 $V(n_{x_d})$ に含まれない頂点 v を追加しても、共通アイテム集合が変わらないということになるからである。この場合も、上記関係1、3、4がすべて成立する場合に、頂点集合 $V(n_{x_d})$ を判定対象として上記の判定を行わないようにしているということができる。

[0076] 基準集合テーブルは、あるノード n_x と他ノード n_y との間に上記関係1~3がすべて成立する場合に、ノード n_x の子のノード n_{c_x} の頂点集合 $V(n_{c_x})$ を判定対象として上記の判定を行わないようするために用いられる。図8は、基準集合テーブルの一例を示す図である。基準集合テーブルは、グラフGに含まれる各頂点 v と、該頂点 v を末端頂点とする頂点集合 $V(n)$ の各頂点に関連付けられたアイテム集合の積集合である基準集合（すなわち、アイテム集合 $I(n)$ ）を少なくとも1つ含む基準集合群 $P(v)$ と、を関連付けてなるテーブルである。この基準集合テーブルの役割については後述する。

[0077] [5-3. 処理の詳細]

図9は制御部8が実行する処理を示すフロー図である。以下、図9を参照しながら制御部8が行う処理の詳細について説明する。なお、図9に示す処理を制御部8が実行することによって、グラフデータ取得部20、頂点デー

タ取得部 22、及び部分グラフ検出部 24 が実現される。なお、図 9 の S 101 の処理は、グラフデータ取得部 20 及び頂点データ取得部 22 に対応しており、図 9 の S 102 以降の処理は部分グラフ検出部 24 に対応している。

[0078] まず、制御部 8 は、共通部分グラフ CSG を検出するための前処理 (S 101 ~ S 104) を実行する。

[0079] まず、制御部 8 はグラフデータと頂点データとをデータベース 4 から読み出して取得する (S 101)。そして、制御部 8 は、読み出したグラフデータを加工する。具体的には、二種以上のアイテム i を含まないアイテム集合 $I(v)$ が関連付けられている頂点 v につながる辺 e を削除し、他の頂点 v とのつながりを切断する (S 102)。上述したように、共通部分グラフ CSG は、すべての頂点 v が二種以上の共通のアイテムを有するような部分グラフである。二種以上のアイテム i を含んでいないアイテム集合 $I(v)$ が関連付けられている頂点 v 、すなわち、二種以上のアイテム i を有しない頂点 v を含む部分グラフは共通部分グラフ CSG にはなり得ない。このため、このような頂点 v を S 103 以降の処理で無視するために S 102 の処理が実行される。S 102 の処理が実行されることによって、無駄な処理が実行されなくなり、共通部分グラフ CSG の検出にかかる時間が短縮される。

[0080] その後、制御部 8 は基準集合テーブルを初期化する (S 103)。具体的には、各頂点 v の基準集合群 $P(v)$ が $null$ に初期化される。また、制御部 8 は、グラフ G に含まれる頂点 v の集合を頂点集合 V_0 として取得する (S 104)。

[0081] そして、制御部 8 は、頂点集合 V_0 に含まれる各頂点 v を処理対象の頂点 v_x として順次選択し、S 105 ~ S 108 の処理を実行する。例えば頂点 v_1 が選択された場合には、S 105 ~ S 108 の処理が実行されることによって、図 7 に示すツリー T のうちの、頂点 v_1 に対応するルートノード n_{r_1} よりも下位のノード n_0 が生成され、頂点 v_1 を含むような共通部分グラフ CSG の検出が実行される。

- [0082] まず、制御部 8 は、頂点集合 V_a に含まれる頂点 v の集合を頂点集合 V_b として取得する (S 105)。そして、制御部 8 は、頂点 v_x に対応するルートノード n_{rx} を生成し、記憶部 10 に保存する (S 106)。例えば、頂点 v_x が頂点 v_1 である場合、ルートノード n_{r1} が生成される。ルートノード n_{r1} の頂点集合 $V(n_{r1})$ は $\{v_1\}$ であり、共通アイテム集合 $I(n_{r1})$ は $\{i_1, i_2, i_3, i_4\}$ である。なお、頂点集合 $V(n_{r1})$ の末端頂点は v_1 である。その後、制御部 8 はサブルーチン S を実行する (S 107)。
- [0083] ここで、サブルーチン S について説明する。図 10 及び図 11 はサブルーチン S の処理を示すフロー図である。サブルーチン S は、頂点集合に相当する引数 VP と、一つの頂点 v に対応する引数 vp と、ツリー T の一つのノード n に対応する引数 np と、の 3 つの引数を有する。S 107 の処理では、頂点集合 V_b を引数 VP とし、頂点 v_x を引数 vp とし、ルートノード n_{rx} を引数 np として、サブルーチン S が呼び出される。
- [0084] 図 10 に示すように、サブルーチン S では、まず、頂点集合 VP に含まれる頂点 v のうちの、頂点 vp に隣接する頂点 v の集合が隣接頂点集合 V_n として取得される (S 201)。そして、制御部 8 は、隣接頂点集合 V_n が空集合であるか否かを判定する (S 202)。隣接頂点集合 V_n が空集合である場合 (S 202 の Y)、すなわち、頂点集合 VP に含まれる頂点 v のうちに頂点 vp に隣接する頂点 v が存在しない場合、制御部 8 はサブルーチン S を終了して、図 9 の S 108 の処理を実行する。すなわち、頂点集合 V_a から頂点 v_x が除かれ (S 108)、頂点集合 V_a から他の頂点 v が新たな処理対象の頂点 v_x として選択されて S 105 ~ S 108 の処理が再実行される。
- [0085] 一方、隣接頂点集合 V_n が空集合でない場合 (S 202 の N)、制御部 8 は、隣接頂点集合 V_n に含まれる各頂点 v を処理対象の頂点 v_{nx} として順次選択して、S 203 ~ S 209 の処理を実行する。
- [0086] まず、制御部 8 は頂点データを参照して、頂点 v_{nx} に関連づけられたアイテム集合 $I(v_{nx})$ を取得する。そして、制御部 8 は、アイテム集合 $I(v_{nx})$ と、ツリー T のノード np の共通アイテム集合 $I(np)$ と、の積集合を

共通アイテム集合 I_c として取得する (S203)。なお、ノード n_p の共通アイテム集合 $I(n_p)$ とは、上述のように、ノード n_p の頂点集合 $V(n_p)$ の各頂点 v に関連づけられたアイテム集合 $I(v)$ の積集合である。

[0087] そして、制御部 8 (判定手段) は、共通アイテム集合 I_c に二種以上のアイテム i が含まれるか否かを判定する (S204)。なお、S204 の処理では、ノード n_p の頂点集合 $V(n_p)$ に頂点 v_{nx} を追加してなる頂点集合が示す部分グラフを判定対象として、該部分グラフの全頂点 v が二種以上の共通のアイテムを共有するか否かが判定されていることになる。つまり、ノード n_p の頂点集合 $V(n_p)$ に頂点 v_{nx} を追加してなる部分グラフが共通部分グラフ CSG であるか否かが判定されていることになる。

[0088] なお、上述のように、ノード n_p の頂点集合 $V(n_p)$ に当該頂点集合 $V(n_p)$ のうちの頂点 v_p (末端頂点) に隣接する頂点 v_{nx} を追加してなる頂点集合は、ノード n_p の子のノードに相当する。また、S203 で取得された共通アイテム集合 I_c は、ノード n_p の子のノードが示す頂点集合 (部分グラフ) の全頂点 v の共通アイテム集合に相当する。このため、S204 の処理では、ノード n_p の子のノードが示す頂点集合 (部分グラフ) の全頂点 v が二種以上の共通のアイテムを共有するか否かが判定されているともいうことができる。すなわち、ノード n_p の子のノードが示す頂点集合 (部分グラフ) が共通部分グラフ CSG であるか否かが判定されているともいうことができる。

[0089] 共通アイテム集合 I_c に二種以上のアイテム i が含まれる場合 (S204 の Y)、制御部 8 は、基準集合テーブル (図 8 参照) から、頂点 v_{nx} に対応する基準集合群 $P(v_{nx})$ を読み出し、共通アイテム集合 I_c が、基準集合群 $P(v_{nx})$ のうちのいずれかの基準集合 I_p の部分集合であるか否かを判定する (S205)。この処理の役割については後述する。

[0090] 共通アイテム集合 I_c がいずれの基準集合 I_p の部分集合でもない場合 (S205 の N)、制御部 8 は、記憶部 10 (基準集合記憶手段) に記憶される基準集合テーブル (図 8 参照) において、頂点 v_{nx} に対応する基準集合群 P

(v_{nx})に共通アイテム集合 I_c を基準集合 I_p として追加する(S206)。その後、制御部8は、ツリーTのうちの、ノード n_p の子のノード n_c であって、かつ、頂点 v_{nx} に対応するノード n_{cx} を生成し、記憶部10に保存する(S207)。この場合、ノード n_{cx} の頂点集合 $V(n_{cx})$ は、ノード n_p の頂点集合 $V(n_p)$ に頂点 v_{nx} を追加した頂点集合に設定され、共通アイテム集合 $I(n_{cx})$ は共通アイテム集合 I_c に設定される。また、頂点集合 $V(n_{cx})$ の末端頂点は頂点 v_{nx} になる。

- [0091] ここで、S207で生成されるノード n_{cx} は一つの共通部分グラフCSGを表している。なぜなら、ノード n_{cx} の頂点集合 $V(n_{cx})$ は頂点 v の集合を示し、共通アイテム集合 $I(n_{cx})$ は、頂点集合 $V(n_{cx})$ に含まれる頂点 v が共通に有するアイテム i の集合を示し、かつ、二種以上のアイテム i を含んでいるからである(S204参照)。このため、ノード n_{cx} を生成することは共通部分グラフCSGを検出することに相当する。
- [0092] その後、頂点集合VPから頂点 v_{nx} が除かれ(S208)、サブルーチンSが再帰的に呼び出される(S209)。この場合、サブルーチンSは、頂点集合VPを引数VPとし、頂点 v_{nx} を引数 v_p とし、S207で生成されたノード n_{cx} を引数 n_p として呼び出される。このサブルーチンSの呼び出しによって、図7に示すツリーTのうちのノード n_{cx} より下位のノード n_c が生成される。
- [0093] なお、S204の処理において、共通アイテム集合 I_c に二種以上のアイテム i が含まれないと判定された場合(S204のN)、制御部8(制限手段)はS205~S209の処理を実行しない。この場合、隣接頂点集合 V_n に含まれる他の頂点 v が新たな処理対象の頂点 v_{nx} として選択され、S203の処理が実行される。
- [0094] ここで、S204の処理の役割について説明する。ここでは、図9のS107の処理でサブルーチンSが、ノード n_{r1} を引数 n_p として呼び出され、かつ、S204の処理において、頂点 v_2 が処理対象の頂点 v_{nx} として選択されている場合を想定する。この場合、上述したように、S204の処理では

、ノード n_{r1} の頂点集合 $V(n_{r1})$ である $\{v_1\}$ に頂点 v_2 を追加してなる頂点集合 $\{v_1, v_2\}$ 、言い換えれば、ノード n_{r1} の子のノード n_{c1} に対応する頂点集合 $\{v_1, v_2\}$ が、共通部分グラフ CSG であるか否かが判定されることになる。すなわち部分グラフ $\{v_1, v_2\}$ が共通部分グラフ CSG であるか否かが判定されることになる。そして、頂点 v_1 と頂点 v_2 とが共通に有するアイテム i はアイテム i_1 のみであるため（図3参照）、S204の処理では、共通アイテム集合 I_c に二種以上のアイテム i が含まれないと判定される。すなわち、部分グラフ $\{v_1, v_2\}$ が共通部分グラフ CSG でないと判定される。その結果、S205～S209の処理が実行されなくなる。すなわち、図7に示すノード n_{c1} が生成されなくなる。また、頂点 v_{nx} (v_2) を引数 v_p とし、かつ、ノード n_{c1} を引数 n_p としたようなサブルーチン S の呼び出しも行われなくなり、図7に示すノード n_{c1} よりも下位のノード $n_{c2} \sim n_{c5}$ も生成されなくなる。つまり、ノード n_{c1} の子孫のノード（ノード n_x とする）の頂点集合 $V(n_x)$ を判定対象とした、頂点集合 $V(n_x)$ により表される部分グラフが共通部分グラフ CSG であるか否かの判定が実行されなくなる。

- [0095] 例えば、頂点 v_{nx} (v_2) を引数 v_p とし、かつ、ノード n_{c1} を引数 n_p としたようなサブルーチン S の呼び出しが行われた場合には、S204の処理では、ノード n_{c1} の頂点集合 $V(n_{c1})$ である $\{v_1, v_2\}$ に、頂点 v_{nx} (v_2) に隣接する頂点 v_3 が追加されてなる頂点集合 $\{v_1, v_2, v_3\}$ 、すなわち部分グラフ $\{v_1, v_2, v_3\}$ が共通部分グラフ CSG であるか否かの判定処理が実行されることになる。例えば、部分グラフ $\{v_1, v_2\}$ の各頂点 v が二種以上の共通のアイテム i を有しておらず、部分グラフ $\{v_1, v_2\}$ が共通部分グラフ CSG でない場合、当然のことながら、その部分グラフ $\{v_1, v_2\}$ に頂点 v_3 を追加してなる部分グラフ $\{v_1, v_2, v_3\}$ も共通部分グラフ CSG にはなり得ない。このため、上記のような判定処理は不要である。この点、本実施の形態では、S204の処理において部分グラフ $\{v_1, v_2\}$ が共通部分グラフ CSG でないと判定された場合、S205～S20

9の処理が実行されなくなるため、部分グラフ $\{v_1, v_2, v_3\}$ が共通部分グラフCSGであるか否かの判定も行われなくなる。すなわち、不要な処理が実行されないように図られる。

[0096] また、S205の処理において、共通アイテム集合 I_c が基準集合群 $P(v_{n_x})$ のうちのいずれかの基準集合 I_p の部分集合であると判定された場合、制御部8（制限手段）はS206以降の処理を実行しない。この場合、隣接頂点集合 V_n に含まれる他の頂点 v が新たに頂点 v_{n_x} として選択され、S203の処理が実行される。

[0097] ここで、基準集合テーブル（図8参照）と、S205の処理と、の役割について説明する。ここでは、例えば、図11のS209でサブルーチンSが、頂点 v_3 を引数 v_p とし、かつ、ノード n_{c6} を引数 n_p として呼び出され、頂点 v_4 を処理対象の頂点 v_{n_x} として、S203～S206の処理が実行される場合を想定する。この場合、上述したように、S204の処理では、ノード n_{c6} の頂点集合 $V(n_{c6})$ である $\{v_1, v_3\}$ に、頂点 $v_{n_x}(v_3)$ に隣接する頂点 v_4 が末端頂点として追加されてなる頂点集合 $\{v_1, v_3, v_4\}$ 、言い換えれば、ノード n_{c7} に対応する頂点集合 $\{v_1, v_3, v_4\}$ が共通部分グラフCSGであるか否かが判定される。すなわち部分グラフ $\{v_1, v_3, v_4\}$ が共通部分グラフCSGであるか否かが判定される。そして、頂点 v_1, v_3, v_4 が共通に有するアイテム i はアイテム i_1, i_2 であるため（図3参照）、S204の処理では、共通アイテム集合 $I_c\{i_1, i_2\}$ に二種以上のアイテム i が含まれると判定される。すなわち、部分グラフ $\{v_1, v_3, v_4\}$ が共通部分グラフCSGであると判定される。また、この時点における基準集合テーブル（図8参照）では、頂点 v_4 に対応する基準集合群 $P(v_4)$ が初期状態（nullの状態）になっているため、S205の処理では、共通アイテム集合 $I_c\{i_1, i_2\}$ が基準集合群 $P(v_4)$ のいずれの基準集合 I_p の部分集合でもないと判定される。その結果、S206の処理が実行され、基準集合テーブル（図8参照）において、頂点 v_4 に対応する基準集合群 $P(v_4)$ に共通アイテム集合 $I_c\{i_1, i_2\}$ が基準集合として追加される

。

[0098] さらに、ここで、図9のS107でサブルーチンSが、頂点 v_1 を引数 v_p とし、ノード n_{r1} を引数 n_p として呼び出され、頂点 v_4 を処理対象の頂点 v_{n_x} として、S203~S205の処理が実行される場合を想定する。この場合、上述したように、S204の処理では、ノード n_{r1} の頂点集合 $V(n_{r1})$ である $\{v_1\}$ に、頂点 v_1 に隣接する頂点 v_4 が末端頂点として追加されてなる頂点集合 $\{v_1, v_4\}$ 、言い換えれば、ノード n_{c10} に対応する頂点集合 $\{v_1, v_4\}$ が共通部分グラフCSGであるか否かが判定される。すなわち部分グラフ $\{v_1, v_4\}$ が共通部分グラフCSGであるか否かが判定される。そして、頂点 v_1, v_4 が共通に有するアイテム i はアイテム i_1, i_2 であるため(図3参照)、S204の処理では、共通アイテム集合 $I_c\{i_1, i_2\}$ に二種以上のアイテム i が含まれると判定される。すなわち、部分グラフ $\{v_1, v_4\}$ が共通部分グラフCSGであると判定される。一方、この時点における基準集合テーブル(図8参照)では、頂点 v_4 に対応する基準集合群 $P(v_4)$ に基準集合 $I_p\{i_1, i_2\}$ が含まれている。ここで、基準集合 $I_p\{i_1, i_2\}$ は、既に生成済みのノード n_{c7} の共通アイテム集合 $I(n_{c7})$ である。頂点 v_4 に対応する基準集合群 $P(v_4)$ に基準集合 $I_p\{i_1, i_2\}$ が含まれているため、S205の処理では、部分グラフ $\{v_1, v_4\}$ の共通アイテム集合 $I_c\{i_1, i_2\}$ が基準集合群 $P(v_4)$ のいずれかの基準集合 I_p の部分集合であると判定される。その結果、S206~S208の処理が実行されなくなる。すなわち、S207の処理が実行されず、図7に示すノード n_{c10} が生成されなくなる。また、S208の処理が実行されず、頂点 $v_{n_x}(v_4)$ を引数 v_p とし、かつ、ノード n_{c10} を引数 n_p としたようなサブルーチンSの呼び出しも行われなくなる。その結果、図7に示すノード n_{c10} よりも下位のノード $n_{c11} \sim n_{c12}$ も生成されなくなる。つまり、ノード n_{c10} の子孫のノード(ノード n_x とする)の頂点集合 $V(n_x)$ が判定対象として取得されず、頂点集合 $V(n_x)$ により表される部分グラフが共通部分グラフCSGであるか否かも判定されなくなる。

- [0099] 例えば、頂点 v_{n_x} (v_4) を引数 v_p とし、かつ、ノード $n_{c_{10}}$ を引数 n_p としたようなサブルーチン S の呼び出しが行われた場合には、 $S204$ の処理では、ノード $n_{c_{10}}$ の頂点集合 $V(n_{c_{10}})$ である $\{v_1, v_4\}$ に、頂点 v_{n_x} (v_4) に隣接する頂点 v_5 が追加されてなる頂点集合 $\{v_1, v_4, v_5\}$ 、すなわち部分グラフ $\{v_1, v_4, v_5\}$ が共通部分グラフ CSG であるか否かが判定されることになる。この点、本実施の形態では、 $S205$ の処理において、部分グラフ $\{v_1, v_4\}$ の共通アイテム集合 $I_c \{i_1, i_2\}$ が基準集合群 $P(v_4)$ のいずれかの基準集合 I_p の部分集合であると判定された場合、 $S205 \sim S209$ の処理が実行されなくなるため、ノード $n_{c_{10}}$ の子のノード $n_{c_{11}}$ の頂点集合 $V(n_{c_{11}})$ により表される部分グラフ $\{v_1, v_4, v_5\}$ が共通部分グラフ CSG であるか否かの判定も行われなくなる。
- [0100] $S205$ の処理で部分グラフ $\{v_1, v_4\}$ の共通アイテム集合 $I_c \{i_1, i_2\}$ が基準集合群 $P(v_4)$ のいずれかの基準集合 $I_p \{i_1, i_2\}$ の部分集合であると判定される場合とは、頂点 v_4 を含む部分グラフに関してすでに検出済みの共通アイテム集合に含まれないような共通アイテム集合が検出される可能性がない場合である。
- [0101] また、 $S205$ の処理で部分グラフ $\{v_1, v_4\}$ の共通アイテム集合 $I_c \{i_1, i_2\}$ が基準集合群 $P(v_4)$ のいずれかの基準集合 $I_p \{i_1, i_2\}$ の部分集合であると判定される場合とは、上述のように基準集合 $I_p \{i_1, i_2\}$ は既に生成済みのノード n_{c_7} の共通アイテム集合 $I(n_{c_7})$ に相当するので、既に生成済みのノード n_{c_7} の共通アイテム集合 $I(n_{c_7})$ にノード $n_{c_{10}}$ の共通アイテム集合 $I(n_{c_{10}})$ が含まれる場合である。さらに、ノード n_{c_7} の頂点集合 $V(n_{c_7})$ の末端頂点と、ノード $n_{c_{10}}$ の頂点集合 $V(n_{c_{10}})$ の末端頂点と、はともに頂点 v_4 なのであるから、 $S205$ の処理で部分グラフ $\{v_1, v_4\}$ の共通アイテム集合 $I_c \{i_1, i_2\}$ が基準集合群 $P(v_4)$ のいずれかの基準集合 $I_p \{i_1, i_2\}$ の部分集合であると判定される場合とは、ノード n_{c_7} をノード n_y とし、ノード $n_{c_{10}}$ をノード n_x とした場合に、ノード n_y とノード n_x との間で上記の関係 1～3 がすべて成立する場合であ

る。このような場合には、部分グラフ $\{v_1, v_4\}$ に頂点 v_5 が追加されてなる頂点集合 $\{v_1, v_4, v_5\}$ 、すなわち部分グラフ $\{v_1, v_4, v_5\}$ が閉部分グラフでないのは明らかだから、共通部分グラフCSGであるか否かの判定処理を実行する必要がない。この点、本実施の形態では、このような不必要な判定処理が実行されないようになる。

[0102] 図9に示すように、すべての頂点 v についてS105～S108の処理が実行された場合、制御部8は共通部分グラフCSGの検出結果を出力する（S109）。上述したように、記憶部10に保持されたツリーTのノード n は共通部分グラフCSG（本実施形態の場合、閉共通部分グラフCCSG）を表している。S109の処理では、記憶部10に保持されたツリーTのノード n に基づいて、例えば、頂点 v 又は辺 e の数が最も多いような共通部分グラフCSGが表示出力される。または、頂点 v 又は辺 e の数が所定の基準数よりも多いような共通部分グラフCSGが表示出力される。または、検出されたすべての共通部分グラフCSGが表示出力される。以上が、部分グラフ検出装置2で実行される処理の詳細である。

[0103] [5. まとめ]

上記の部分グラフ検出システム6では、例えば、複数の遺伝子に対応する複数の頂点 v_i を含み、該複数の遺伝子の関係を示すグラフGのグラフデータが記憶される。また、複数の頂点 v_i の各々にアイテム集合 $I(v_i)$ を対応づけてなる頂点データが記憶され、この頂点データは各遺伝子の発現を変化させる医薬を示すデータとなっている。そして、部分グラフ検出装置2では、これらのデータに基づいて、直接的又は間接的に関係を有し、かつ、共通の医薬の投与により共通して発現が変化する遺伝子群を表す共通部分グラフCSGが検出される。この検出結果を用いることによって、例えば、創薬のターゲットとすべき遺伝子を絞り込む上で有用な情報が得られるようになる。

[0104] また例えば、部分グラフ検出システム6では、複数のタンパク質に対応する頂点 v_i を含み、かつ、該複数のタンパク質の関係を示すグラフGのグラフ

データが記憶され、各タンパク質の発現を変化させる医薬を示す頂点データが記憶される。この場合、直接的又は間接的に関係を有し、かつ、共通の医薬の投与により共通して発現が変化するタンパク質群を表す共通部分グラフCSGが検出される。この検出結果を用いることによって、例えば、創薬のターゲットとすべきタンパク質を絞り込む上で有用な情報が得られるようになる。

[0105] また例えば、部分グラフ検出システム6では、複数のユーザに対応する複数の頂点 v_i を含み、かつ、該複数のユーザの交友関係を示すグラフGのグラフデータが記憶され、各ユーザが購入した商品を示す頂点データが記憶される。この場合、直接的又は間接的に交友関係を有し、かつ、共通の商品を購入したユーザ群を表す共通部分グラフCSGが検出される。この検出結果を用いることによって、例えば、上述したように、商品の広告宣伝を効率的に行えるようになる。

[0106] ところで、図10のS204及びS205の少なくとも一方を省略しても、共通部分グラフCSGの検出は可能である。例えば、図10のS204及びS205の少なくとも一方を省略しても、閉共通部分グラフCCSGの検出は可能である。例えば、すべての頂点集合、すなわち、すべての部分グラフを判定対象として上記のような判定処理を行ってすべての共通部分グラフを検出し、これらのうちから閉共通部分グラフを特定することが考えられる。しかしながら、この場合、例えば、グラフGに含まれる頂点 v の数（例えば、10000個）が非常に多い場合には、すべての共通部分グラフCSGの検出に膨大な時間がかかる可能性がある。

[0107] この点、上記の部分グラフ検出装置2では、例えば、図10のS204又はS205の処理が実行されることによって、不必要な処理の実行が抑制される。つまり、上記判定処理が、閉部分グラフでない部分グラフを判定対象としては行われなくなる。このため、共通部分グラフCSG（閉共通部分グラフCCSG）の検出にかかる時間が短縮される。そのため、グラフGに含まれる頂点 v の数が非常に多い場合であっても、共通部分グラフCSG（閉

共通部分グラフCCSG)の検出が比較的短時間でされるようになる。

[0108] [6. 変形例]

なお、本発明は、以上に説明した実施形態に限定されるものではない。

[0109] [6-1. 変形例1]

例えば、図9乃至図11の処理では、頂点集合 $V(n_x)$ が複数の頂点を含むようなノード n_x と、他のノード n_y と、の間で上記関係1~3がすべて成立している場合に、ノード n_x の子孫となるノード n_{xd} の頂点集合 $V(n_{xd})$ を判定対象とした上記判定処理が省略されるようになっている。しかしながら、頂点集合 $V(n_x)$ が一の頂点のみを含むようなノード n_x (第3のノード)と、ノード n_y (第4のノード)と、の間で上記関係1~3(又は上記関係1, 3, 4)がすべて成立している場合に、ノード n_x の子孫となるノード n_{xd} の頂点集合 $V(n_{xd})$ を判定対象とした上記判定処理が省略されるようにしてもよい。

[0110] 例えば、制御部8は、図9のS107の処理を実行する前に、図10のS205と類似の処理を実行するようによい。すなわち、制御部8は、頂点 v_x に対応する基準集合群 $P(v_x)$ を基準集合テーブル(図8参照)から取得し、頂点 v_x のアイテム集合 $I(v_x)$ が基準集合群 $P(v_x)$ のうちのいずれかのアイテム集合 I_p の部分集合であるか否かを判定するようによい。そして、アイテム集合 $I(v_x)$ がいずれのアイテム集合 I_p の部分集合でもない場合にS107の処理を実行し、アイテム集合 $I(v_x)$ がいずれかのアイテム集合 I_p の部分集合である場合には、S107の処理を実行しないで、S108の処理を実行するようによい。このようにしても、共通部分グラフCSG(閉共通部分グラフCCSG)を検出するために不必要な処理の実行が抑制される。

[0111] また、ツリーTの各ノードを幅優先探索アルゴリズムに従った探索順序で判定対象とする場合は、頂点集合 $V(n_x)$ が一の頂点のみを含むようなノード n_x (第3のノード)と、ノード n_y (第4のノード)と、の間で上記関係1, 2', 3(又は上記関係1, 3, 4)がすべて成立している場合に、ノ

ード n_x の子孫となるノード n_{x_d} の頂点集合 $V(n_{x_d})$ を判定対象とした上記判定処理が省略されるようにしてもよい。こうすれば、ツリー T の各ノードを幅優先探索アルゴリズムに従った探索順序で判定対象とする場合も、共通部分グラフ CSG (閉共通部分グラフ $CCSG$) を検出するために不必要な処理の実行が抑制される。

[0112] [6-2. 変形例2]

また、共通部分グラフ CSG を検出するための処理は上記に説明したものに限られない。例えば、まず、 N 種以上のアイテム i を含む基準アイテム集合 I_d が決定されるようにしてもよい。その後、各頂点 v に関連付けられたアイテム集合 $I(v)$ が該基準アイテム集合 I_d を含むような共通部分グラフ CSG が検出されるようにしてもよい。この場合、制御部 8 は、図 10 の S203 の処理において、アイテム集合 I_d が共通アイテム集合 I_c の部分集合であるか否かが判定されるようにすればよい。

[0113] [6-3. 変形例3]

例えば、図 5 に示す頂点データにおけるアイテム i は、ユーザが利用している又は過去に利用したことのあるサービスに対応するようにしてもよい。この場合、頂点データは各ユーザが利用している又は過去に利用したことのあるサービスを示すデータになる。このようにすれば、直接的又は間接的に交友関係にあり、かつ、共通のサービスを利用している又は利用したことのあるユーザ群を検出することが可能になる。そして、この検出結果を用いることによって、例えば、サービスの広告宣伝を効率的に行えるようになる。

[0114] [6-4. 変形例4]

例えば、グラフデータは、複数の文献に対応する複数の頂点 v を含み、該複数の文献の関係 (例えば引用関係) を示すグラフのデータであってもよい。このグラフでは、例えば、第 1 の文献が第 2 の文献を引用している場合、第 1 の文献に対応する頂点 v_1 と第 2 の文献に対応する頂点 v_2 が辺 e_{12} によって連結される。また、この場合、頂点データにおけるアイテム i は、例えば、文献に関連するキーワードに対応するようにしてもよい。この場合、頂

点データは各文献のキーワードを示すデータになる。この変形例 1 によれば、上記のグラフデータ及び頂点データに基づいて、例えば、直接的又は間接的に引用関係にあり、かつ、共通のキーワードが設定された文献群を検出できるようになる。なお例えば、アイテム i は文献の著者に対応させるようにしてもよい。

[0115] [6-5. その他]

また、グラフ G は、遺伝子、タンパク質、ユーザ、文献以外の関係を示すグラフであってもよい。すなわち、グラフ G に含まれる頂点 v は、遺伝子、タンパク質、ユーザ、文献以外に対応する頂点であってもよい。また、頂点データにおいて頂点 v に関連づけられる情報は、頂点間の接続関係を示す情報以外の情報であれば、ユーザに関する情報、キーワード、医薬に限らず、どのような情報であってもよい。

[0116] また、以上では、グラフ G が無向グラフである場合について述べたが、グラフ G は有向グラフであってもよい。この場合、図 7 のツリー T は、頂点間の有向関係を考慮して生成される。例えば、制御部 8 は、図 10 の S 201 のステップにおいて、頂点集合 V_P に含まれる頂点 v_p に隣接する頂点 v のうち、頂点 v_p から伸びる辺に接続された頂点 v の集合を、隣接頂点集合 V_n として取得するようにする。

[0117] [7. 実験]

なお、本実施形態に係る方法 A と他の方法 B とのそれぞれで、「M」本以上の辺を含む閉共通部分グラフ $CCSG$ の集合 C を検出する実験を試みたので、その実験内容及び実験結果について説明する。

[0118] ここでは、以下の条件を全て満足するような閉共通部分グラフ $CCSG$ の集合 C を検出することとする。

- ・ 2 種以上の共通のアイテムが、集合 C のどの閉共通部分グラフ $CCSG$ にも含まれる。（各閉共通部分グラフ $CCSG$ の共通アイテム集合の積集合には 2 種以上のアイテムが含まれる）。

- ・ 集合 C に含まれるすべての閉共通部分グラフ $CCSG$ は、集合 C に含まれ

る他のどの閉共通部分グラフCCSGとも頂点を共通しない。

[0119] また、実験に用いたマシン環境を以下に示すとおりである。

- ・CPU : 3. 2GB AMD Opteron (登録商標)
- ・メモリ容量 : 1GB
- ・OS : Linux (kernel 2. 6)

[0120] (方法A)

フェーズA1 : 図9乃至図11に示す処理を行って、閉共通部分グラフCCSGを特定する。

フェーズA2 : フェーズA1で特定した閉共通部分グラフCCSGのうちから「M」本以上の辺を含む閉共通部分グラフCCSGを特定する。

フェーズA3 : フェーズA2で特定した閉共通部分グラフCCSGの共通アイテム集合に基づき集合Cを検出する。

[0121] (方法B)

フェーズB1 : アプリオリ法により2種以上のアイテムを含む頻出アイテム集合を検出する。ここで、頻出アイテム集合とは、頻出アイテム集合を含むアイテム集合に関連づけられた頂点が「F」個以上存在するアイテム集合である。なお、本実験では、頻出アイテム集合を検出するためにフリーのデータマイニングソフト「Weka」を用いた。

フェーズB2 : 頻出アイテム集合ごとに、該頻出アイテム集合を含むアイテム集合に関連づけられた頂点からなる、辺が「M」本以上の閉共通部分グラフCCSGを特定することにより、集合Cを検出する。

[0122] なお、方法Bにおいて「F」の値を「2」に設定し且つ「M」の値を「1」に設定した場合、方法AのフェーズA1が行われた場合に特定される閉共通部分グラフCCSGと同じ閉共通部分グラフCCSGが特定される。

[0123] まず最初の実験では、「F」の値を「2」に設定し且つ「M」の値を「10」に設定した上で、実験対象グラフから集合Cを検出した。なお、実験対象グラフとしては、頂点が「文献」に対応し、辺が「文献間の引用関係」に対応し、かつ、頂点に関連づけられたアイテムが「文献の著者」に対応する

ようなグラフを用いた。なお、著者の総数は16638人である。また、実験対象グラフの頂点、辺の数は以下のとおりである。

- ・頂点の数：22178個
- ・辺の数：112304本

[0124] 方法Aでは、7.9秒で集合Cの検出が完了した。一方、方法Bでは、実験開始から101秒経過したところでメモリ不足となり、その結果、処理が強制的に停止された。これは、フェーズB1において頻出アイテム集合を検出するために大量の頻出アイテム集合候補が生成されるためであると考えられる。

[0125] また次の実験では、「F」の値を「4」に設定し且つ「M」の値を「7」に設定した上で、実験対象グラフから集合Cを検出した。なお、実験対象グラフとしては、頂点が「遺伝子」に対応し、辺が「遺伝子間の相互作用関係」に対応し、かつ、頂点に関連づけられたアイテムが「遺伝子の発現を変化させる要因」に対応するようなグラフを用いた。なお、上記要因の総数は173種である。また、実験対象グラフの頂点、辺の数は以下のとおりである。

- ・頂点の数：6152個
- ・辺の数：3318本

[0126] 方法Aでは、35.9秒で集合Cの検出が完了した。一方、方法Bでは、実験開始から17302秒経過したところでメモリ不足となり、その結果、処理が強制的に停止された。先述の実験と同様、これは、フェーズB1において頻出アイテム集合を検出するために大量の頻出アイテム集合候補が生成されるためであると考えられる。

[0127] 上記の実験結果より、本実施形態に係る方法Aでは、方法Bに比べて、処理に必要なメモリの量が軽減され、処理時間も短縮される。

請求の範囲

- [請求項1] 複数の頂点と1又は複数のエッジとを含むグラフを示すグラフデータを取得するグラフデータ取得手段と、
前記頂点に情報を関連づけてなる頂点データを取得する頂点データ取得手段と、
前記グラフデータと、前記頂点データと、に基づいて、前記グラフの部分グラフであって、該部分グラフの各頂点に関連付けられた前記情報が所定関係を有する部分グラフを検出する部分グラフ検出手段と、
、
を含むことを特徴とする部分グラフ検出装置。
- [請求項2] 前記情報は、少なくとも一種の要素を含む要素集合を示し、
前記部分グラフ検出手段は、
前記グラフの部分グラフであって、該部分グラフの各頂点に関連づけられた要素集合の積集合が N ($N \geq 1$) 種以上の要素を含む部分グラフの検出を実行すること、
を特徴とする請求項1に記載の部分グラフ検出装置。
- [請求項3] 前記部分グラフ検出手段は、
各ノードが前記グラフの頂点の集合を示す探索木であって、かつ、親ノードが示す頂点集合に当該頂点集合のうちの末端頂点に隣接する一つの頂点を新たな末端頂点として追加してなる頂点集合を当該親ノードの子ノードとして有する探索木の各ノードを、深さ優先探索アルゴリズムに従った探索順序で判定対象とし、前記判定対象のノードが示す頂点集合が複数の頂点を含む場合に該頂点集合の各頂点に関連づけられた要素集合の積集合が N 種以上の要素を含むか否かの判定を実行する判定手段と、
前記判定手段の判定結果に基づいて前記検出を実行する手段と、を含み、
前記判定手段は、

前記探索木のうちの第1のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合が、前記第1のノードが示す頂点集合のうちの末端頂点と同じ頂点を末端頂点とする頂点集合を示すノードであって、かつ、前記第1のノードよりも探索順序が前のノードである第2のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合に含まれる場合、前記第1のノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないこと、

を特徴とする請求項2に記載の部分グラフ検出装置。

[請求項4]

前記部分グラフ検出手段は、

前記判定が実行された場合に、前記判定対象の頂点集合の各頂点に関連づけられた要素集合の積集合を、基準集合として、当該判定対象の頂点集合のうちの末端頂点に関連づけて基準集合記憶手段に記憶させる手段を含み、

前記判定手段は、

前記第1のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合が、前記第1のノードが示す頂点集合のうちの末端頂点に関連づけて前記基準集合記憶手段に記憶される基準集合に含まれる場合、前記第1のノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないこと、

を特徴とする請求項3に記載の部分グラフ検出装置。

[請求項5]

前記部分グラフ検出手段は、

各ノードが前記グラフの頂点の集合を示す探索木であって、かつ、親ノードが示す頂点集合に当該頂点集合のうちの末端頂点に隣接する一つの頂点を新たな末端頂点として追加してなる頂点集合を当該親ノードの子ノードとして有する探索木の各ノードを、深さ優先又は幅優先探索アルゴリズムに従った探索順序で判定対象とし、前記判定対象のノードが示す頂点集合が複数の頂点を含む場合に該頂点集合の各頂点に関連づけられた要素集合の積集合がN種以上の要素を含むか否か

の判定を実行する判定手段と、

前記判定手段の判定結果に基づいて前記検出を実行する手段と、を含み、

前記判定手段は、

前記探索木のうちの第1のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合が、前記第1のノードが示す頂点集合のうちの末端頂点と同じ頂点を末端頂点とする頂点集合を示すノードであって、かつ、前記第1のノードが示す頂点集合を含む頂点集合を示す第2のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合に含まれる場合、前記第1のノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないこと、

を特徴とする請求項2に記載の部分グラフ検出装置。

[請求項6]

前記部分グラフ検出手段は、

各ノードが前記グラフの頂点の集合を示す探索木であって、かつ、親ノードが示す頂点集合に当該頂点集合のうちの末端頂点に隣接する一つの頂点を新たな末端頂点として追加してなる頂点集合を当該親ノードの子ノードとして有する探索木の各ノードを、深さ優先探索アルゴリズムに従った探索順序で判定対象とし、前記判定対象のノードが示す頂点集合が複数の頂点を含む場合に該頂点集合の各頂点に関連づけられた要素集合の積集合がN種以上の要素を含むか否かの判定を実行する判定手段と、

前記判定手段の判定結果に基づいて前記検出を実行する手段と、を含み、

前記判定手段は、

前記探索木のうちの第3のノードが示す頂点集合が一つの頂点のみを含み、かつ、当該頂点に関連づけられた要素集合が、当該頂点を末端頂点とする頂点集合を示すノードであって、かつ、前記第3のノードよりも探索順序が前のノードである第4のノードが示す頂点集合の

各頂点に関連づけられた要素集合の積集合に含まれる場合、前記第3のノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないこと、

を特徴とする請求項2に記載の部分グラフ検出装置。

[請求項7]

前記部分グラフ検出手段は、

前記判定が実行された場合に、前記判定対象のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合を、基準集合として、前記判定対象のノードが示す頂点集合のうちの末端頂点に関連づけて基準集合記憶手段に記憶させる手段を含み、

前記判定手段は、

前記第3のノードが示す頂点集合が一つの頂点のみを含み、かつ、当該頂点に関連づけられた要素集合が、当該頂点に関連づけて前記基準集合記憶手段に記憶される基準集合に含まれる場合に、前記第3のノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないこと、

を特徴とする請求項6に記載の部分グラフ検出装置。

[請求項8]

前記部分グラフ検出手段は、

各ノードが前記グラフの頂点の集合を示す探索木であって、かつ、親ノードが示す頂点集合に当該頂点集合のうちの末端頂点に隣接する一つの頂点を新たな末端頂点として追加してなる頂点集合を当該親ノードの子ノードとして有する探索木の各ノードを、深さ優先又は幅優先探索アルゴリズムに従った探索順序で判定対象とし、前記判定対象のノードが示す頂点集合が複数の頂点を含む場合に該頂点集合の各頂点に関連づけられた要素集合の積集合がN種以上の要素を含むか否かの判定を実行する判定手段と、

前記判定手段の判定結果に基づいて前記検出を実行する手段と、を含み、

前記判定手段は、

前記探索木のうちの第3のノードが示す頂点集合が一つの頂点のみを含み、当該頂点に関連づけられた要素集合が、当該頂点を末端頂点とする頂点集合を示す第4のノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合に含まれる場合、前記第3のノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないこと、

を特徴とする請求項2に記載の部分グラフ検出装置。

[請求項9]

前記部分グラフ検出手段は、

各ノードが前記グラフの頂点の集合を示す探索木であって、かつ、親ノードが示す頂点集合に当該頂点集合の末端頂点に隣接する一つの頂点を新たな末端頂点として追加してなる頂点集合を当該親ノードの子ノードとして有する探索木の各ノードを、深さ優先探索又は幅優先探索アルゴリズムに従った探索順序で判定対象とし、前記判定対象のノードが示す頂点集合が複数の頂点を含む場合に該頂点集合の各頂点に関連づけられた要素集合の積集合がN種以上の要素を含むか否かの判定を実行する判定手段と、

前記判定手段の判定結果に基づいて前記検出を実行する手段と、を含み、

前記判定手段は、

前記ノードが示す頂点集合の各頂点に関連づけられた要素集合の積集合がN種以上の要素を含んでいない場合、当該ノードの子孫のノードが示す頂点集合を前記判定対象とした前記判定を実行しないこと、を特徴とする請求項2に記載の部分グラフ検出装置。

[請求項10]

複数の頂点と1又は複数のエッジとを含むグラフを示すグラフデータを取得するグラフデータ取得ステップと、

前記頂点に情報を関連づけてなる頂点データを取得する頂点データ取得ステップと、

前記グラフデータと、前記頂点データと、に基づいて、前記グラフ

の部分グラフであって、該部分グラフに含まれる各頂点に関連付けられた前記情報が所定関係を有する部分グラフを検出する部分グラフ検出ステップと、

を含むことを特徴とする部分グラフ検出方法。

[請求項11]

複数の頂点と1又は複数のエッジとを含むグラフを示すグラフデータを取得するグラフデータ取得手段、

前記頂点に情報を関連づけてなる頂点データを取得する頂点データ取得手段、及び、

前記グラフデータと、前記頂点データと、に基づいて、前記グラフの部分グラフであって、該部分グラフに含まれる各ノードに関連付けられた前記情報が所定関係を有する部分グラフを検出する部分グラフ検出手段、

としてコンピュータを機能させるプログラム。

[請求項12]

複数の頂点と1又は複数のエッジとを含むグラフを示すグラフデータを取得するグラフデータ取得手段、

前記頂点に情報を関連づけてなる頂点データを取得する頂点データ取得手段、及び、

前記グラフデータと、前記頂点データと、に基づいて、前記グラフの部分グラフであって、該部分グラフに含まれる各ノードに関連付けられた前記情報が所定関係を有する部分グラフを検出する部分グラフ検出手段、

としてコンピュータを機能させるプログラムを記録したコンピュータ読み取り可能な情報記憶媒体。

[請求項13]

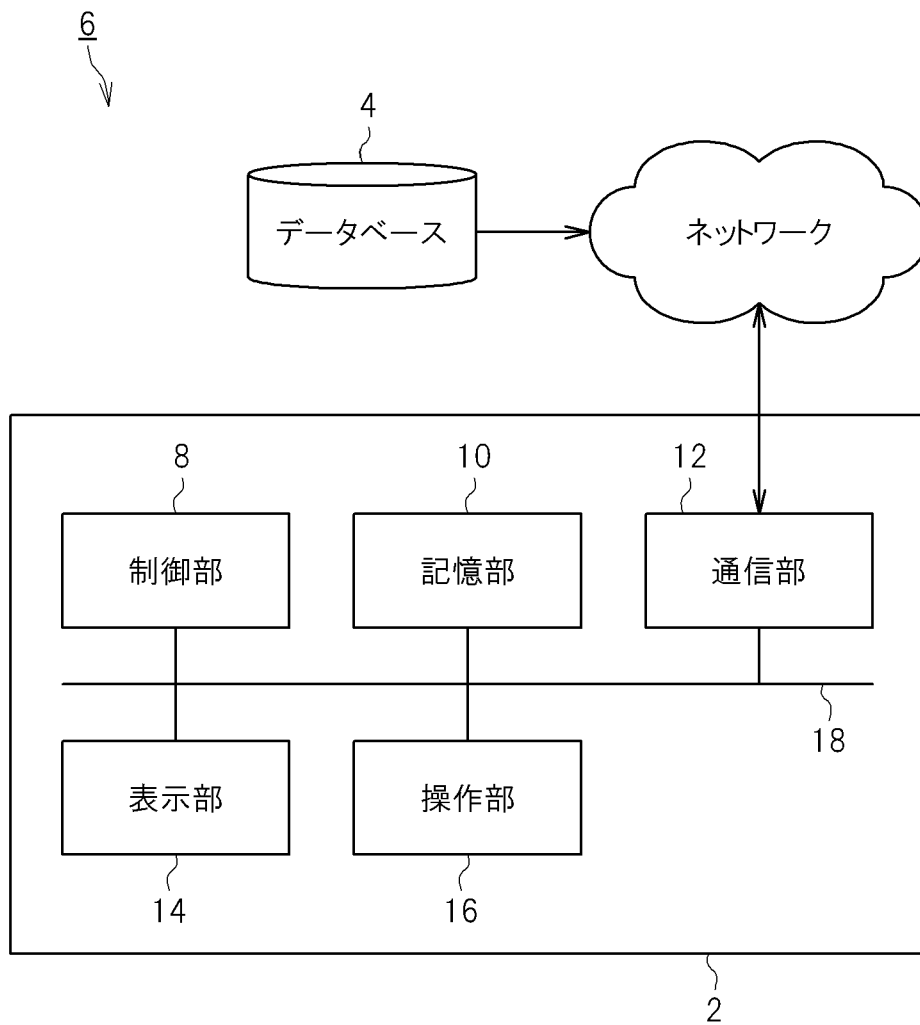
複数の頂点と1又は複数のエッジとを含むグラフを示すグラフデータと、前記頂点に情報を関連づけてなる頂点データと、を含むことを特徴とするデータのデータ構造。

[請求項14]

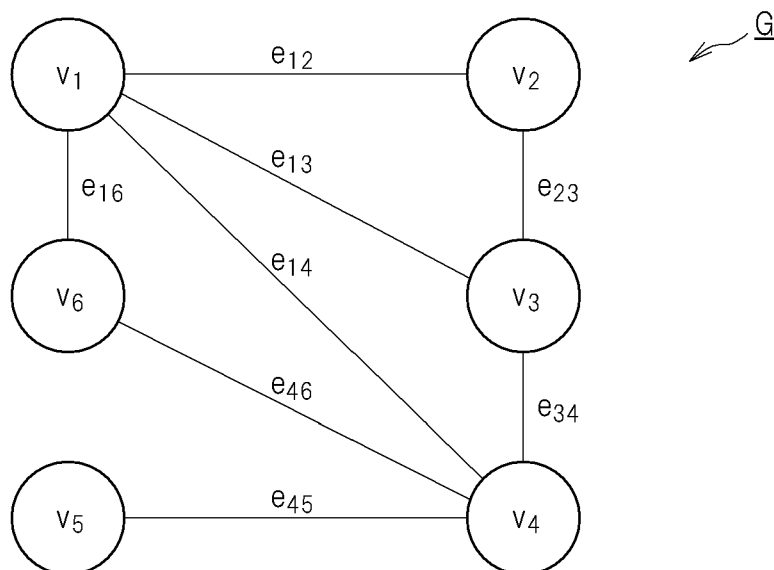
複数の頂点と1又は複数のエッジとを含むグラフを示すグラフデータと、前記頂点に情報を関連づけてなる頂点データと、を含むことを

特徴とするデータを記録したコンピュータ読み取り可能な情報記憶媒体。

[図1]



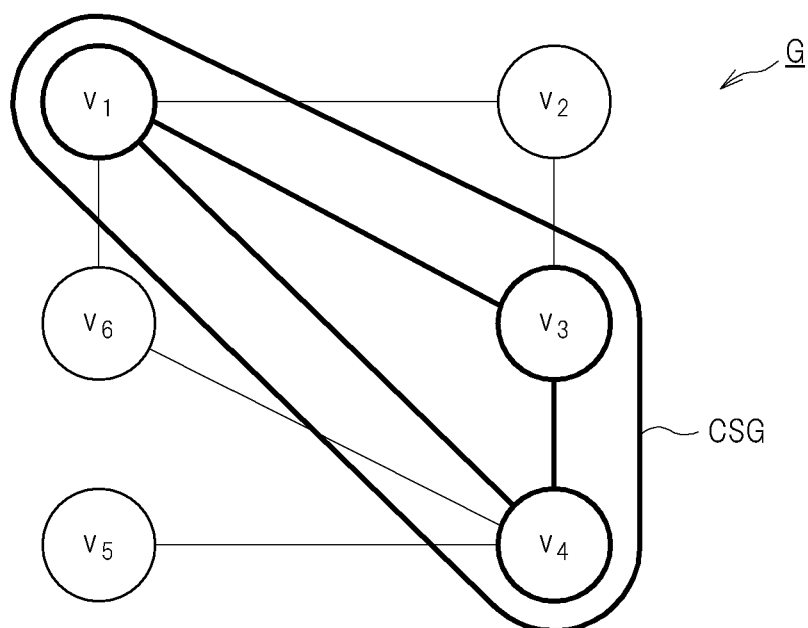
[図2]



[図3]

v	I(v)
v ₁	i ₁ (医薬A), i ₂ (医薬B), i ₃ (医薬C), i ₄ (医薬D)
v ₂	i ₁ , i ₅ (医薬E)
v ₃	i ₁ , i ₂ , i ₄
v ₄	i ₁ , i ₂ , i ₅
v ₅	i ₁ , i ₃ , i ₄
v ₆	i ₁ , i ₃ , i ₅

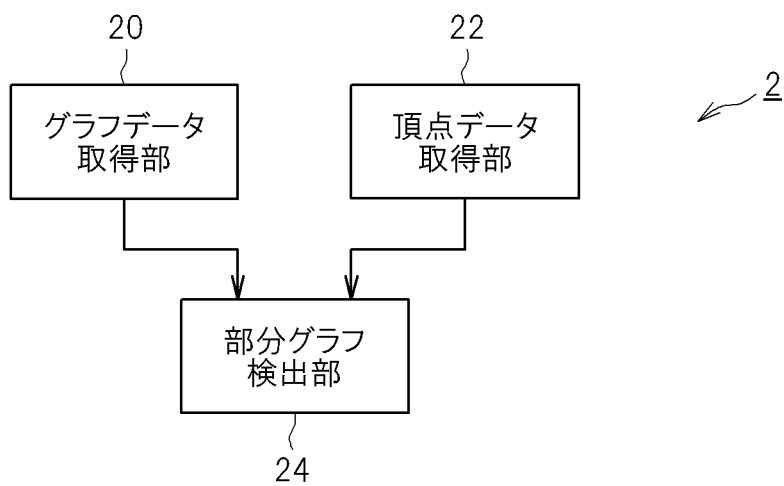
[図4]



[図5]

v	$I(v)$
v_1	i_1 (りんご), i_2 (みかん), i_3 (ぶどう), i_4 (ナシ)
v_2	i_1, i_5 (メロン)
v_3	i_1, i_2, i_4
v_4	i_1, i_2, i_5
v_5	i_1, i_3, i_4
v_6	i_1, i_3, i_5

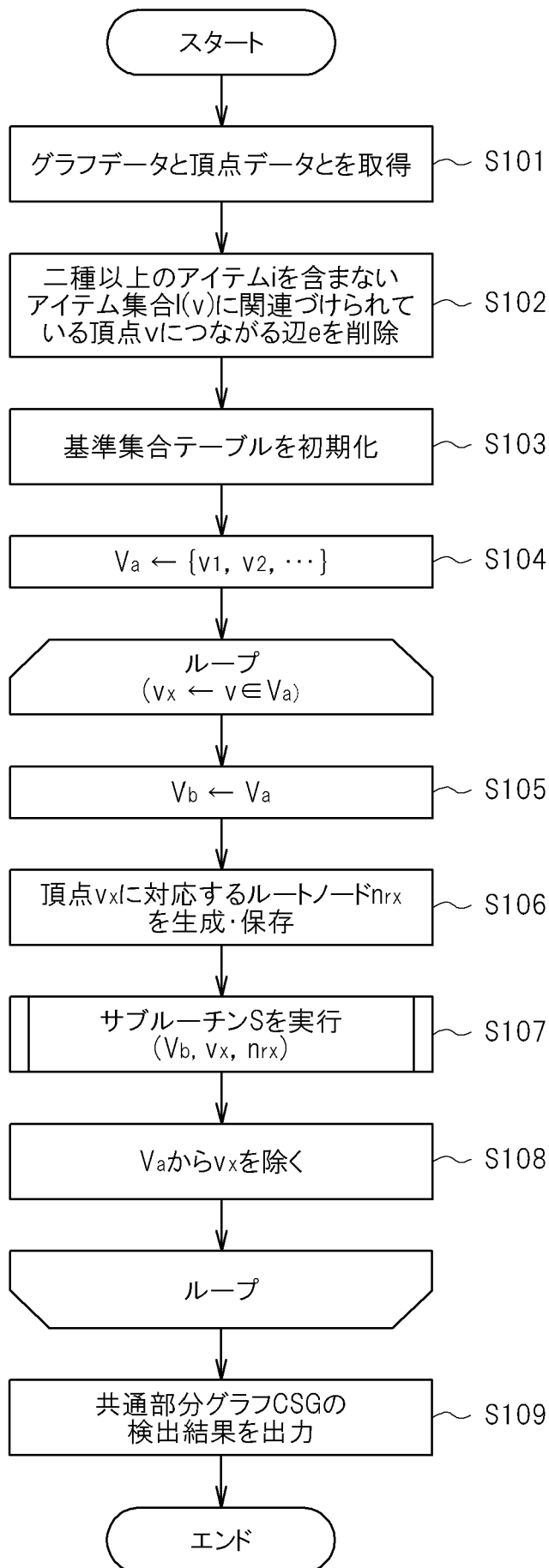
[図6]



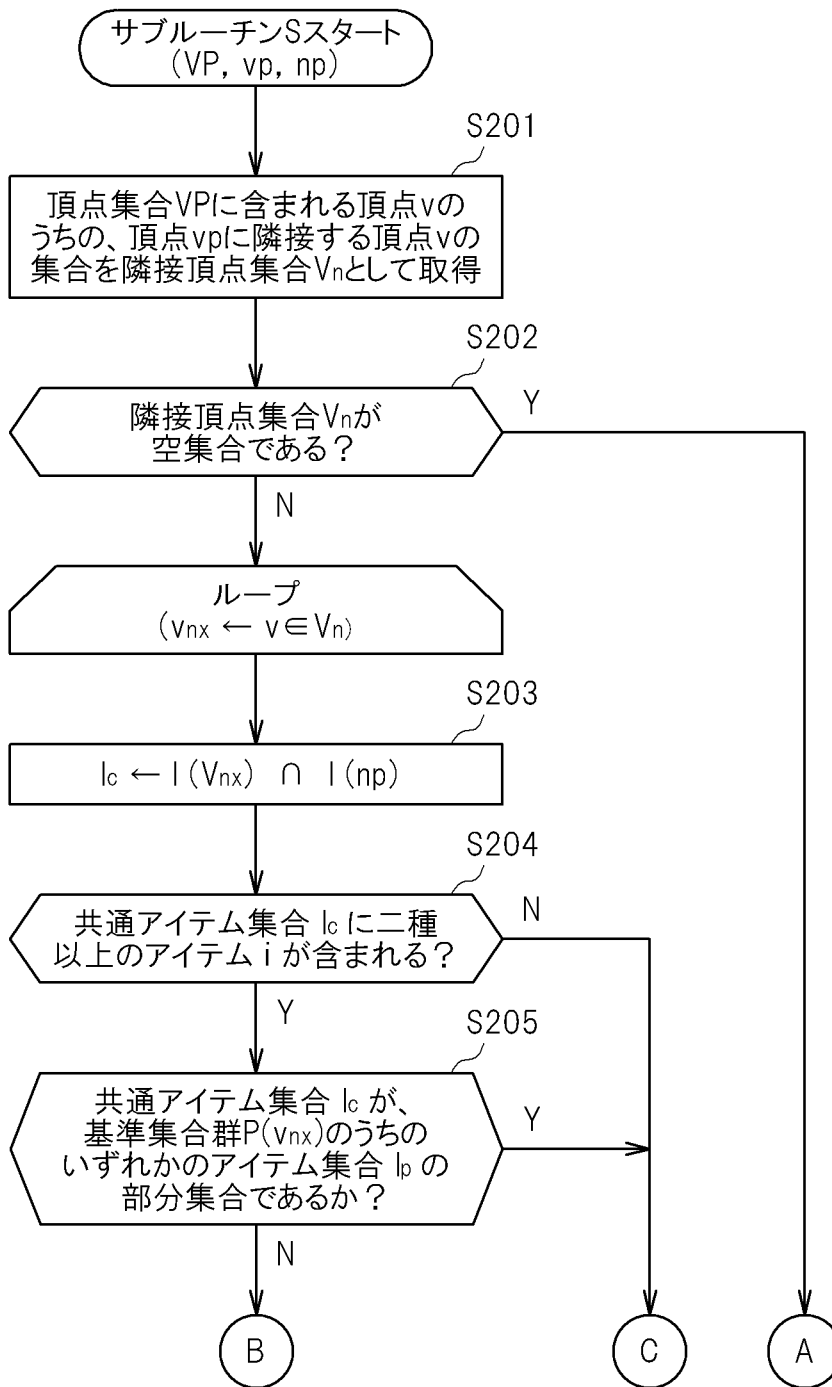
[図8]

v	P (v)
v ₁	{i ₁ , i ₂ , i ₃ , i ₄ }
v ₂	{i ₁ , i ₅ }
v ₃	{i ₁ , i ₂ , i ₄ }
v ₄	{i ₁ , i ₂ }{i ₁ , i ₂ , i ₅ }
v ₅	{i ₁ , i ₃ }
v ₆	{i ₁ , i ₃ }

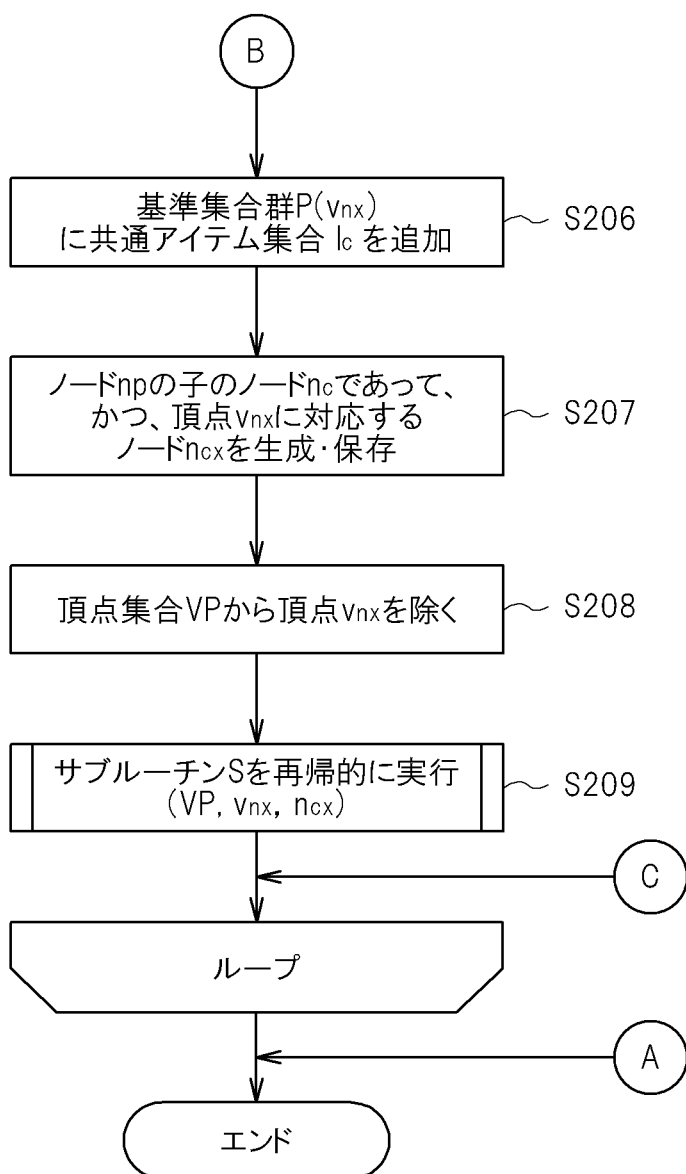
[図9]



[図10]



[図11]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/067476

A. CLASSIFICATION OF SUBJECT MATTER

G06F17/30(2006.01) i, G06F19/00(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F17/30, G06F19/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2009
Kokai Jitsuyo Shinan Koho	1971-2009	Toroku Jitsuyo Shinan Koho	1994-2009

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	JP 2002-091991 A (INTEC Web and Genome Infomatics Corp.), 29 March 2002 (29.03.2002), claim 1; fig. 14 (Family: none)	13-14 1-12
A	US 2007/0100680 A1 (KUMAR S, et al.), 03 May 2007 (03.05.2007), paragraphs [0366] to [0392]; fig. 13 & US 2007/0094066 A1 & US 2007/0094067 A1 & US 2007/0118439 A1 & EP 1949271 A & EP 1862963 A1 & WO 2007/048008 A2 & WO 2007/048008 B	1-14

 Further documents are listed in the continuation of Box C. 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
28 October, 2009 (28.10.09)Date of mailing of the international search report
10 November, 2009 (10.11.09)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2009/067476

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2000-123041 A (Nippon Telegraph And Telephone Corp.), 28 April 2000 (28.04.2000), claims 1 to 5; fig. 4 (Family: none)	1-14
A	JP 2004-220227 A (NEC Corp.), 05 August 2004 (05.08.2004), claims 1 to 7; fig. 2 (Family: none)	1-14

A. 発明の属する分野の分類 (国際特許分類 (IPC))
 Int.Cl. G06F17/30(2006.01)i, G06F19/00(2006.01)i

B. 調査を行った分野
 調査を行った最小限資料 (国際特許分類 (IPC))
 Int.Cl. G06F17/30, G06F19/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2009年
日本国実用新案登録公報	1996-2009年
日本国登録実用新案公報	1994-2009年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)
 WPI

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X A	JP 2002-091991 A (インテック・ウェブ・アンド・ゲノム・インフォマティクス株式会社) 2002.03.29, 請求項 1, 図 14 (ファミリーなし)	13-14 1-12
A	US 2007/0100680 A1 (KUMAR S, et al.) 2007.05.03, [0366]-[0392], Figure 13 & US 2007/0094066 A1 & US 2007/0094067 A1 & US 2007/0118439 A1 & EP 1949271 A & EP 1862963 A1 & WO 2007/048008 A2 & WO 2007/048008 B	1-14

C欄の続きにも文献が列挙されている。 パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー	の日の後に公表された文献
「A」 特に関連のある文献ではなく、一般的技術水準を示すもの	「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)	「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「O」 口頭による開示、使用、展示等に言及する文献	「&」 同一パテントファミリー文献
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願	

国際調査を完了した日 28. 10. 2009	国際調査報告の発送日 10. 11. 2009
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 岩間 直純 電話番号 03-3581-1101 内線 3599

5M 9287

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2000-123041 A (日本電信電話株式会社) 2000.04.28, 請求項 1-5, 図 4 (ファミリーなし)	1-14
A	JP 2004-220227 A (日本電気株式会社) 2004.08.05, 請求項 1-7, 図 2 (ファミリーなし)	1-14