

(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
2020年1月9日(09.01.2020)



(10) 国際公開番号

WO 2020/009208 A1

- (51) 国際特許分類:
G06F 16/36 (2019.01) *G06F 17/22* (2006.01)
- (21) 国際出願番号: PCT/JP2019/026736
- (22) 国際出願日: 2019年7月4日(04.07.2019)
- (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 2018-128341 2018年7月5日(05.07.2018) JP
- (71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 山 嵯 磨 与 (YAMASAKI, Mayo); 〒1808585 東京都武蔵野市緑町3丁目9-

1 1 N T T 知的財産センタ内 Tokyo (JP).
川邊 秀樹(KAWABE, Hideki); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T
知的財産センタ内 Tokyo (JP).

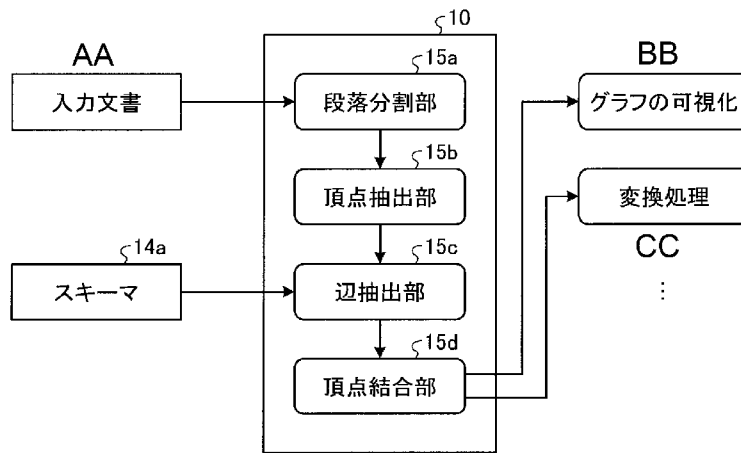
(74) 代理人: 特許業務法人酒井国際特許事務所 (SAKAI INTERNATIONAL PATENT OFFICE); 〒1000013 東京都千代田区霞が関3丁目8番1号 虎の門三井ビルディング Tokyo (JP).

(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, 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, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,

(54) Title: EXTRACTION METHOD, EXTRACTION DEVICE, AND EXTRACTION PROGRAM

(54) 発明の名称: 抽出方法、抽出装置および抽出プログラム

[図4]



- 14a Schema
15a Paragraph dividing unit
15b Vertex extraction unit
15c Edge extraction unit
15d Vertex joining unit
AA Input document
BB Graph visibility
CC Conversion processing

(57) Abstract: A paragraph dividing unit (15a) divides a document having markup information applied thereto that indicates a vertex of a graph structure, dividing same into paragraphs. A vertex extraction unit (15b) extracts a vertex for each paragraph. An edge extraction unit (15c) extracts edges between extracted vertices, for each paragraph. A storage unit (14) stores a schema (14a) that specifies edges between the attributes of vertices in the graph structure. The paragraph dividing unit (15a) divides into paragraphs a document having marking information applied thereto that indicates vertices in the graph structure and attributes of the vertices. The vertex extraction unit (15b) extracts the vertices and the attributes of the vertices, for each paragraph. The edge extraction unit (15c) can extract an edge specified in the schema (14a), for each paragraph. The vertex joining unit (15d) can join extracted vertices that are duplicated between paragraphs.

[続葉有]

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, 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, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (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), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類：

- 一 国際調査報告 (条約第21条(3))

(57) 要約：段落分割部 (15 a) が、グラフ構造の頂点を示すマークアップ情報が付与された文書を、段落ごとのまとまりに分割し、頂点抽出部 (15 b) が、段落ごとに、頂点を抽出し、辺抽出部 (15 c) が、段落ごとに、抽出された頂点間の辺を抽出する。記憶部 (14) が、グラフ構造の頂点の属性間の辺を特定するスキーマ (14 a) を記憶し、段落分割部 (15 a) が、グラフ構造の頂点と該頂点の属性とを示すマークアップ情報が付与された文書を、段落ごとのまとまりに分割し、頂点抽出部 (15 b) が、段落ごとに、頂点と該頂点の属性とを抽出し、辺抽出部 (15 c) が、段落ごとに、スキーマ (14 a) において特定されている辺を抽出してもよい。頂点結合部 (15 d) が、段落間で重複して抽出された頂点を結合してもよい。

明 細 書

発明の名称：抽出方法、抽出装置および抽出プログラム

技術分野

[0001] 本発明は、抽出方法、抽出装置および抽出プログラムに関する。

背景技術

[0002] 近年、情報セキュリティコミュニティにおいて、情報共有の必要性が広く認識されている。そして情報共有のため、頂点群と頂点間の連結関係を表す辺群とで構成されるグラフ構造を再利用可能に表現できる構造化形式が提案されている。一方、多くの場合、情報共有は、文書の形式でなされている。

[0003] 従来、グラフ構造を入力として受け付けて、解析や表示等の処理を行う装置が知られている。例えば、DOT言語と呼ばれる独自のDSL (Domain-Specific Language) を入力として受け付ける、Graphvizと呼ばれる装置が知られている（非特許文献1参照）。また、XML (eXtensible Markup Language) を用いたGXL (Graph eXchange Language) を入力として受け付ける装置（非特許文献2参照）やGraphMLを入力として受け付ける装置（非特許文献3参照）等が知られている。これらの装置では、自然言語ではなく、独自の方式で記述されたグラフ構造を入力として受け付ける。

[0004] また、文書中に簡易なグラフ構造を付与することができる、Markdown、reStructuredText等の形式で記述された簡易なグラフ構造を入力として受け付ける装置が知られている（非特許文献4参照）。

先行技術文献

非特許文献

[0005] 非特許文献1: John Ellson, Emden R. Gansner, Eleftherios Koutsofi os, Stephen C. North, Gordon Woodhull, “Graphviz and Dynagraph - Static and Dynamic Graph Drawing Tools”, In Graph drawing software, Springer, 2004年, p.127-148

非特許文献2: Richard C. Holt, Andreas Winter, Andy Schurr, “Gxl: Toward a Standard Eexchange Format”, In ReverseEngineering, Seventh Working Conference, IEEE, 2000年, p.162-171

非特許文献3: Ulrik Brandes, Markus Eiglsperger, Ivan Herman, Michael Himsolt, M. Scott Marshall, “GraphML Progress Report Structural Layer Proposal”, In International Symposium on Graph Drawing, Springer, 2001年, p.501-512

非特許文献4: John Gruber, Aaron Swartz, et al., “Markdown”, [online], 2004年, [2018年5月18日検索], インターネット<URL:https://daringfireball.net/projects/markdown/>

発明の概要

発明が解決しようとする課題

[0006] しかしながら、従来の技術では、文書と文書中に含まれるグラフ構造とを同時に入力として受け付けて処理することは困難であった。つまり、従来の技術では、文書中に複雑なグラフ構造のマークアップ情報を記述することは困難であった。したがって、ある文書と、文書から抽出できる知識等を表すグラフ構造との処理を行う場合には、文書と、文書から別途抽出したグラフ構造のマークアップ情報とを別々に入力する必要があった。

[0007] そのため、装置による再利用性の高いグラフ構造を、別途文書から抽出したり文書変更に伴って更新したりするための人的コストが高くなっていた。その結果、再利用性の高いグラフ構造の情報の普及が妨げられていた。

[0008] 本発明は、上記に鑑みてなされたものであって、文書と文書に含まれるグラフ構造とを同時に入力として受け付けて処理することを目的とする。

課題を解決するための手段

[0009] 上述した課題を解決し、目的を達成するために、本発明に係る抽出方法は、抽出装置で実行される抽出方法であって、グラフ構造の頂点を示すマークアップ情報が付与された文書を、段落ごとのまとまりに分割する段落分割工程と、前記段落ごとに、頂点と該頂点の属性とを抽出する頂点抽出工程と、

前記段落ごとに、抽出された前記頂点間の辺を抽出する辺抽出工程と、を含んだことを特徴とする。

発明の効果

[0010] 本発明によれば、文書と文書に含まれるグラフ構造とを同時に入力として受け付けて処理することができる。

図面の簡単な説明

[0011] [図1]図1は、抽出装置の概略構成を例示する模式図である。

[図2]図2は、頂点抽出部の処理を説明するための説明図である。

[図3]図3は、辺抽出部の処理を説明するための説明図である。

[図4]図4は、抽出装置の処理を説明するための説明図である。

[図5]図5は、抽出処理手順を示すフローチャートである。

[図6]図6は、抽出処理の効果を説明するための説明図である。

[図7]図7は、抽出プログラムを実行するコンピュータの一例を示す図である。

発明を実施するための形態

[0012] 以下、図面を参照して、本発明の一実施形態を詳細に説明する。なお、この実施形態により本発明が限定されるものではない。また、図面の記載において、同一部分には同一の符号を付して示している。

[0013] [抽出装置の構成]

図1は、抽出装置の概略構成を例示する模式図である。図1に例示するように、抽出装置10は、パソコン等の汎用コンピュータで実現され、入力部11、出力部12、通信制御部13、記憶部14、および制御部15を備える。

[0014] 入力部11は、キーボードやマウス等の入力デバイスを用いて実現され、操作者による入力操作に対応して、制御部15に対して処理開始などの各種指示情報を入力する。出力部12は、液晶ディスプレイなどの表示装置、プリンター等の印刷装置等によって実現される。

[0015] 通信制御部13は、NIC (Network Interface Card) 等で実現され、

L A N (Local Area Network) やインターネットなどの電気通信回線を介した外部の装置と制御部 1 5 との通信を制御する。

[0016] 記憶部 1 4 は、R A M (Random Access Memory)、フラッシュメモリ (Flash Memory) 等の半導体メモリ素子、または、ハードディスク、光ディスク等の記憶装置によって実現され、後述する抽出処理により学習されたデータの生成モデルのパラメータ等が記憶される。なお、記憶部 1 4 は、通信制御部 1 3 を介して制御部 1 5 と通信する構成でもよい。本実施形態において、記憶部 1 4 には、スキーマ 1 4 a が記憶されている。

[0017] ここで、スキーマ 1 4 a とは、グラフ構造の頂点の属性間の辺を特定する情報であり、グラフ構造のひな形に該当する。例えば、スキーマ 1 4 a では、属性「s u b」の頂点から属性「s u p e r」の頂点への方向の辺は、辺「i s - a」というように特定されている。

[0018] 制御部 1 5 は、C P U (Central Processing Unit) 等を用いて実現され、メモリに記憶された処理プログラムを実行する。これにより、制御部 1 5 は、図 1 に例示するように、段落分割部 1 5 a、頂点抽出部 1 5 b、辺抽出部 1 5 c および頂点結合部 1 5 d として機能する。なお、これらの機能部は、それぞれあるいは一部が異なるハードウェアに実装されてもよい。

[0019] 段落分割部 1 5 a は、グラフ構造の頂点を示すマークアップ情報が付与された文書を、段落ごとのまとまりに分割する。具体的には、段落分割部 1 5 a は、グラフ構造の頂点と該頂点の属性とを示すマークアップ情報が付与された文書の入力を受け付けて、段落ごとのまとまりに分割する。

[0020] 頂点抽出部 1 5 b は、段落ごとに、頂点を抽出する。具体的には、頂点抽出部は、段落ごとに、頂点と該頂点の属性とを抽出する。

[0021] ここで、図 2 は、頂点抽出部 1 5 b の処理を説明するための説明図である。図 2 に例示するように、入力された文書には、グラフ構造の頂点と該頂点の属性とを示す情報として、例えば「頂点名」{頂点属性}の形式で表されるマークアップ情報が付与されている。図 2 に例示する文書には、例えば、頂点名「A p p l e」、その属性「s u b」である頂点を示すマークアップ

情報が含まれている。

[0022] 頂点抽出部15bは、図2に例示した文書の第一文からは、頂点名「apple」、頂点属性「sub」である頂点と、頂点名「orange」、頂点属性「sub」である頂点と、頂点名「fruit」、頂点属性「super」である頂点とを抽出する。また、頂点抽出部15bは、図2に例示した文書の第二文からは、頂点名「human」、頂点属性「sub」である頂点と、頂点名「animal」、頂点属性「super」である頂点とを抽出する。

[0023] 辺抽出部15cは、段落ごとに、抽出された頂点間の辺を抽出する。具体的には、辺抽出部15cは、段落ごとに、スキーマ14aにおいて特定されている辺を抽出する。

[0024] ここで、図3は、辺抽出部15cの処理を説明するための説明図である。図3に示すように、辺抽出部15cは、例えば、文書の段落Aからは、属性「sub」、頂点名「apple」である頂点から、属性「super」、頂点名「fruit」である頂点の方向の辺「is-a」を抽出する。また、辺抽出部15cは、属性「sub」、頂点名「orange」である頂点から、属性「super」、頂点名「fruit」である頂点の方向の辺「is-a」を抽出する。このようにして、文書の段落Aからは、3つの頂点と2つの辺で構成されるグラフ構造が抽出される。

[0025] 同様に、辺抽出部15cは、文書の段落Bからは、属性「sub」、頂点名「human」である頂点から、属性「super」、頂点名「animal」である頂点の方向の辺「is-a」を抽出する。このようにして、文書の段落Bからは、2つの頂点と1つの辺で構成されるグラフ構造が抽出される。

[0026] このように、変更頻度の少ないスキーマ14aを予め記憶しておくことにより、同様のグラフ構造が含まれる文書から同一のスキーマ14aを参照して、効率よく辺を抽出することができる。

[0027] なお、スキーマ14aが記憶部14に記憶されていない場合には、辺抽出

部 15 c は、段落ごとに、抽出された全ての頂点の間に辺が張られるものとして、全ての頂点間の辺を抽出すればよい。これにより、辺の記述を省略することができる。

[0028] 図 1 の説明に戻る。頂点結合部 15 d は、段落間で重複して抽出された頂点を結合する。例えば、図 3 に示した段落 A の頂点と段落 B の頂点とで、頂点名および頂点属性が同一の頂点がある場合には、その頂点を介して段落 A のグラフ構造と段落 B のグラフ構造とを結合する。これにより、文書から、より再利用性の高いグラフ構造を抽出できる。

[0029] 図 4 は、抽出装置 10 の処理を説明するための説明図である。図 4 に示すように、抽出装置 10 は、グラフ構造の頂点と該頂点の属性とを示す情報が付与された文書の入力を受け付ける。段落分割部 15 a は、入力された文書を段落ごとに分割する。頂点抽出部 15 b は、段落から頂点と頂点属性とを抽出する。

[0030] また、抽出装置 10 は、グラフ構造のひな形であるスキーマ 14 a の入力を受け付けて、記憶部 14 に記憶する。辺抽出部 15 c は、スキーマ 14 a を参照し、抽出された頂点間から、スキーマ 14 a で特定されている辺を抽出する。また、頂点結合部 15 d は、段落間で重複して抽出された頂点を結合する。

[0031] 抽出装置 10 は、このようにして抽出したグラフ構造を用いて、グラフの可視化や、他のデータ形式への変換等の処理を行う。このように、抽出装置 10 では、グラフ構造を含んだ文書の入力を受け付けて、グラフ構造を抽出し、文書と、抽出したグラフ構造とに対する処理を実行することができる。

[0032] [抽出処理]

次に、図 5 を参照して、本実施形態に係る抽出装置 10 による抽出処理について説明する。図 5 は、抽出処理手順を示すフローチャートである。図 5 のフローチャートは、例えば、ユーザが抽出処理の開始を指示する操作入力を行ったタイミングで開始される。

[0033] まず、段落分割部 15 a が、処理対象であるグラフ構造を含む文書の入力

を受け付ける（ステップS 1）。そして、段落分割部 1 5 a が入力された文書を段落ごとのまとまりに分割する（ステップS 2）。

[0034] 次に、頂点抽出部 1 5 b が、段落ごとに、頂点と頂点属性とを抽出する（ステップS 3）。また、辺抽出部 1 5 c が、段落ごとに、抽出された頂点間から、スキーマ 1 4 a で特定されている辺を抽出する（ステップS 4）。これにより、文書からグラフ構造が抽出される。

[0035] また、頂点結合部 1 5 d が、段落間で重複して抽出された頂点があれば、その頂点を介してグラフ構造を結合する（ステップS 5）。制御部 1 5 は、このようにして文書から抽出したグラフ構造を用いて、グラフの可視化やデータ形式の変換処理等のグラフ構造に対する処理を実行する（ステップS 6）。また、制御部 1 5 は、グラフ構造を抽出した文書に対する処理を実行する。これにより、一連の抽出処理が終了する。

[0036] 以上、説明したように、本実施形態の抽出装置 1 0 において、段落分割部 1 5 a が、グラフ構造の頂点を示すマークアップ情報が付与された文書を、段落ごとのまとまりに分割する。また、頂点抽出部 1 5 b が、段落ごとに、頂点を抽出する。また、辺抽出部 1 5 c が、段落ごとに、抽出された頂点間の辺を抽出する。

[0037] これにより、煩雑に辺の記述がされていなくても、グラフ構造の頂点を示す軽量のマークアップ情報を用いて、文書に含まれるグラフ構造を抽出することができる。このように、文書と文書に含まれるグラフ構造とを同時に入力として受け付けて処理することが可能となる。

[0038] また、記憶部 1 4 が、グラフ構造の頂点の属性間の辺を特定するスキーマ 1 4 a を記憶してもよい。その場合には、段落分割部 1 5 a が、グラフ構造の頂点と該頂点の属性とを示すマークアップ情報が付与された文書を、段落ごとのまとまりに分割し、頂点抽出部 1 5 b が、段落ごとに、頂点と該頂点の属性とを抽出し、辺抽出部 1 5 c が、段落ごとに、スキーマ 1 4 a において特定されている辺を抽出する。これにより、文書に煩雑な辺の記述がなくても、変更頻度の少ないグラフ構造のひな形であるスキーマ 1 4 a を用いて

、効率よくグラフ構造の辺を抽出することができる。

[0039] また、頂点結合部 15d が、段落間で重複して抽出された頂点を結合する。これにより、文書から、より規模の大きなグラフ構造を抽出することが可能となる。

[0040] ここで、図 6 は、抽出処理の効果を説明するための説明図である。図 6 に示すように、従来、文書と文書に含まれるグラフ構造に対して解析を行って任意の処理を行う場合には、文書と文書から抽出されたグラフ構造のマークアップ情報とを別々に装置に入力して処理を行わせる必要があった。

[0041] これに対し、本実施形態の抽出装置 10 による抽出処理では、グラフ構造のマークアップ情報が記述された文書をそのまま入力し、文書とグラフ構造とに対する任意の処理を行わせることができる。そのため、別途文書からグラフ構造を抽出して記述する作成コストや、作成のための訓練や文書が変更された場合にグラフ構造を抽出し直す更新等の保守コスト等の人的コストを削減することができる。

[0042] したがって、再利用性の高い構造化された情報を軽量の形式で文書に記述して、効率よく共有化することが可能となる。また、視覚的にも、文書とグラフ構造との対応関係を把握しやすく、意味の乖離が生じる恐れを抑止できる。これにより、グラフ構造として記述できる様々な知見や知識の共有を促進することが可能となる。

[0043] [プログラム]

上記実施形態に係る抽出装置 10 が実行する処理をコンピュータが実行可能な言語で記述したプログラムを作成することもできる。一実施形態として、抽出装置 10 は、パッケージソフトウェアやオンラインソフトウェアとして上記の抽出処理を実行する抽出プログラムを所望のコンピュータにインストールさせることによって実装できる。例えば、上記の抽出プログラムを情報処理装置に実行させることにより、情報処理装置を抽出装置 10 として機能させることができる。ここで言う情報処理装置には、デスクトップ型またはノート型のパーソナルコンピュータが含まれる。また、その他にも、情報

処理装置にはスマートフォン、携帯電話機やP H S (Personal Handyphone System)などの移動体通信端末、さらには、P D A (Personal Digital Assistants)などのスレート端末などがその範疇に含まれる。

[0044] また、抽出装置10は、ユーザが使用する端末装置をクライアントとし、当該クライアントに上記の抽出処理に関するサービスを提供するサーバ装置として実装することもできる。例えば、抽出装置10は、グラフ構造を示す情報が付与された文書を入力とし、文書およびグラフ構造に対する処理結果を出力する抽出処理サービスを提供するサーバ装置として実装される。この場合、抽出装置10は、Webサーバとして実装することとしてもよいし、アウトソーシングによって上記の抽出処理に関するサービスを提供するクラウドとして実装することとしてもかまわない。以下に、抽出装置10と同様の機能を実現する抽出プログラムを実行するコンピュータの一例を説明する。

[0045] 図7は、抽出プログラムを実行するコンピュータの一例を示す図である。コンピュータ1000は、例えば、メモリ1010と、CPU1020と、ハードディスクドライバインタフェース1030と、ディスクドライバインタフェース1040と、シリアルポートインタフェース1050と、ビデオアダプタ1060と、ネットワークインタフェース1070とを有する。これらの各部は、バス1080によって接続される。

[0046] メモリ1010は、ROM (Read Only Memory) 1011およびRAM 1012を含む。ROM 1011は、例えば、BIOS (Basic Input Output System) 等のブートプログラムを記憶する。ハードディスクドライバインタフェース1030は、ハードディスクドライブ1031に接続される。ディスクドライバインタフェース1040は、ディスクドライブ1041に接続される。ディスクドライブ1041には、例えば、磁気ディスクや光ディスク等の着脱可能な記憶媒体が挿入される。シリアルポートインタフェース1050には、例えば、マウス1051およびキーボード1052が接続される。ビデオアダプタ1060には、例えば、ディスプレイ1061が接

続される。

[0047] ここで、ハードディスクドライブ1031は、例えば、OS1091、アプリケーションプログラム1092、プログラムモジュール1093およびプログラムデータ1094を記憶する。上記実施形態で説明した各情報は、例えばハードディスクドライブ1031やメモリ1010に記憶される。

[0048] また、抽出プログラムは、例えば、コンピュータ1000によって実行される指令が記述されたプログラムモジュール1093として、ハードディスクドライブ1031に記憶される。具体的には、上記実施形態で説明した抽出装置10が実行する各処理が記述されたプログラムモジュール1093が、ハードディスクドライブ1031に記憶される。

[0049] また、抽出プログラムによる情報処理に用いられるデータは、プログラムデータ1094として、例えば、ハードディスクドライブ1031に記憶される。そして、CPU1020が、ハードディスクドライブ1031に記憶されたプログラムモジュール1093やプログラムデータ1094を必要に応じてRAM1012に読み出して、上述した各手順を実行する。

[0050] なお、抽出プログラムに係るプログラムモジュール1093やプログラムデータ1094は、ハードディスクドライブ1031に記憶される場合に限られず、例えば、着脱可能な記憶媒体に記憶されて、ディスクドライブ1041等を介してCPU1020によって読み出されてもよい。あるいは、抽出プログラムに係るプログラムモジュール1093やプログラムデータ1094は、LANやWAN (Wide Area Network) 等のネットワークを介して接続された他のコンピュータに記憶され、ネットワークインタフェース1070を介してCPU1020によって読み出されてもよい。

[0051] 以上、本発明者によってなされた発明を適用した実施形態について説明したが、本実施形態による本発明の開示の一部をなす記述および図面により本発明は限定されることはない。すなわち、本実施形態に基づいて当業者等によりなされる他の実施形態、実施例および運用技術等は全て本発明の範疇に含まれる。

符号の説明

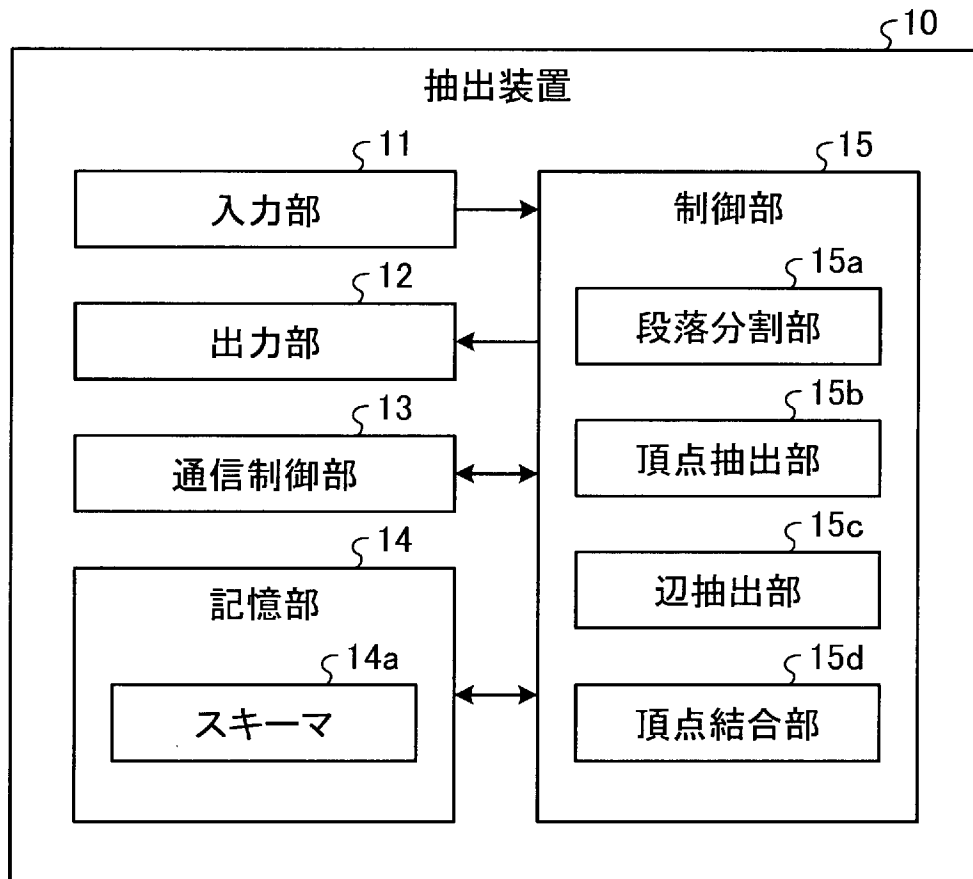
- [0052] 1 0 抽出装置
- 1 1 入力部
- 1 2 出力部
- 1 3 通信制御部
- 1 4 記憶部
- 1 4 a スキーマ
- 1 5 制御部
- 1 5 a 段落分割部
- 1 5 b 頂点抽出部
- 1 5 c 辺抽出部
- 1 5 d 頂点結合部

請求の範囲

- [請求項1] 抽出装置で実行される抽出方法であって、
グラフ構造の頂点を示すマークアップ情報が付与された文書を、段落ごとのまとまりに分割する段落分割工程と、
前記段落ごとに、頂点を抽出する頂点抽出工程と、
前記段落ごとに、抽出された前記頂点間の辺を抽出する辺抽出工程と、
を含んだことを特徴とする抽出方法。
- [請求項2] 前記抽出装置は、グラフ構造の頂点の属性間の辺を特定するスキーマを記憶する記憶部を備え、
前記段落分割工程において、グラフ構造の頂点と該頂点の属性とを示すマークアップ情報が付与された文書を、段落ごとのまとまりに分割し、
前記頂点抽出工程において、前記段落ごとに、前記頂点と該頂点の属性とを抽出し、
前記辺抽出工程において、前記段落ごとに、前記スキーマにおいて特定されている辺を抽出することを特徴とする請求項1に記載の抽出方法。
- [請求項3] 前記段落間で重複して抽出された前記頂点を結合する頂点結合工程を、さらに含んだことを特徴とする請求項1または2に記載の抽出方法。
- [請求項4] グラフ構造の頂点を示すマークアップ情報が付与された文書を、段落ごとのまとまりに分割する段落分割部と、
前記段落ごとに、前記頂点を抽出する頂点抽出部と、
前記段落ごとに、抽出された前記頂点間の辺を抽出する辺抽出部と、
を備えることを特徴とする抽出装置。
- [請求項5] グラフ構造の頂点を示すマークアップ情報が付与された文書を、段

落ごとのまとまりに分割する段落分割ステップと、
前記段落ごとに、前記頂点を抽出する頂点抽出ステップと、
前記段落ごとに、抽出された前記頂点間の辺を抽出する辺抽出ステップと、
をコンピュータに実行させるための抽出プログラム。

[図1]

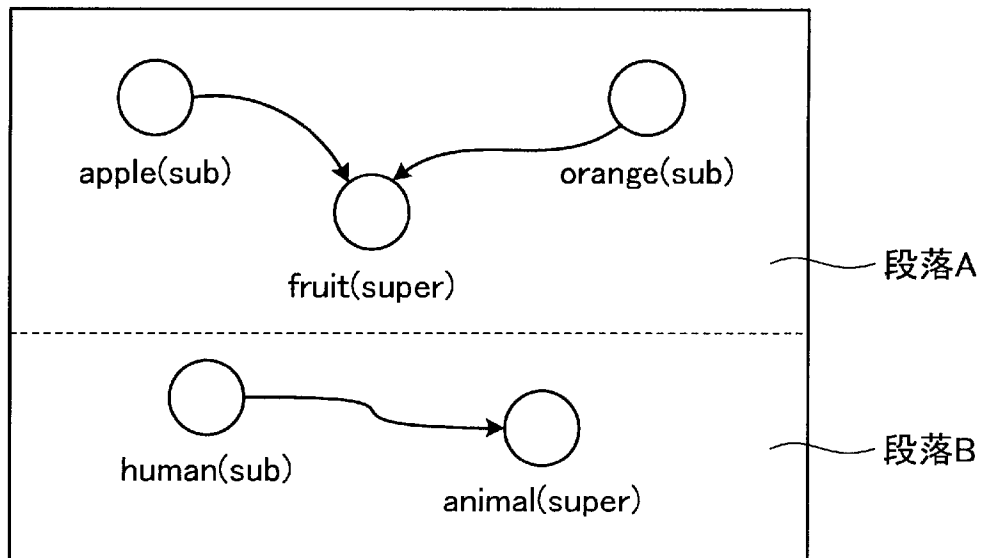


[図2]

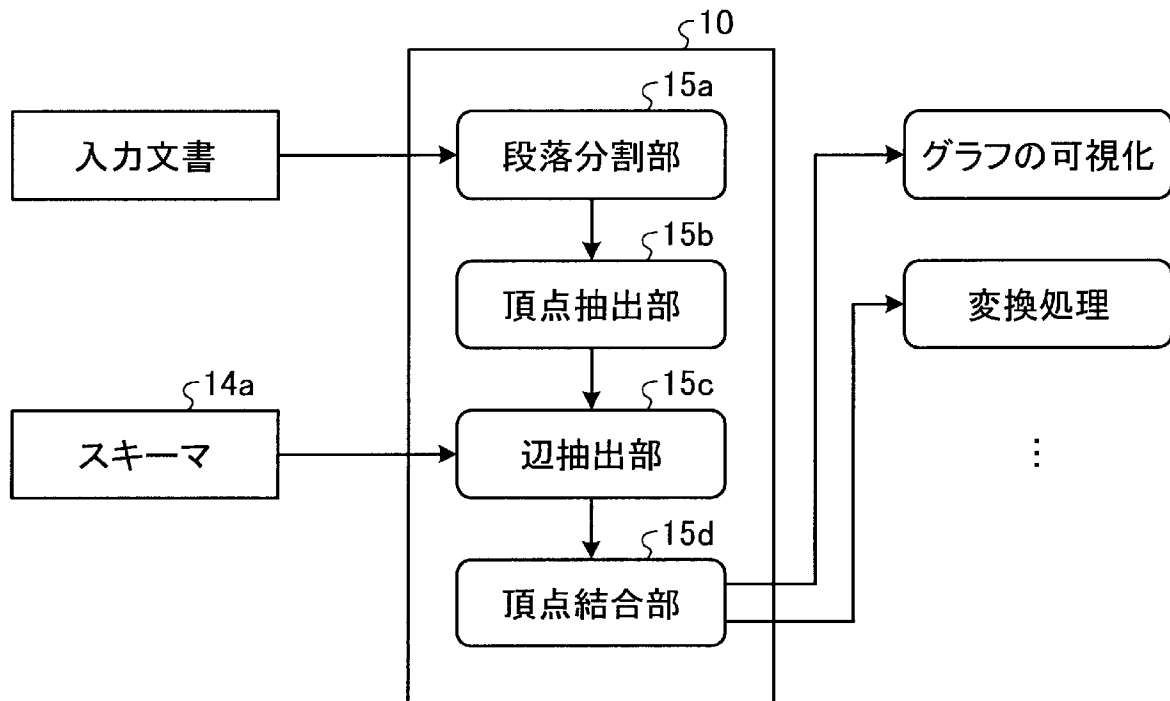
[Apple]{sub} and [orange]{sub} are [fruit]{super}.

[Human]{sub} is a [animal]{super}.

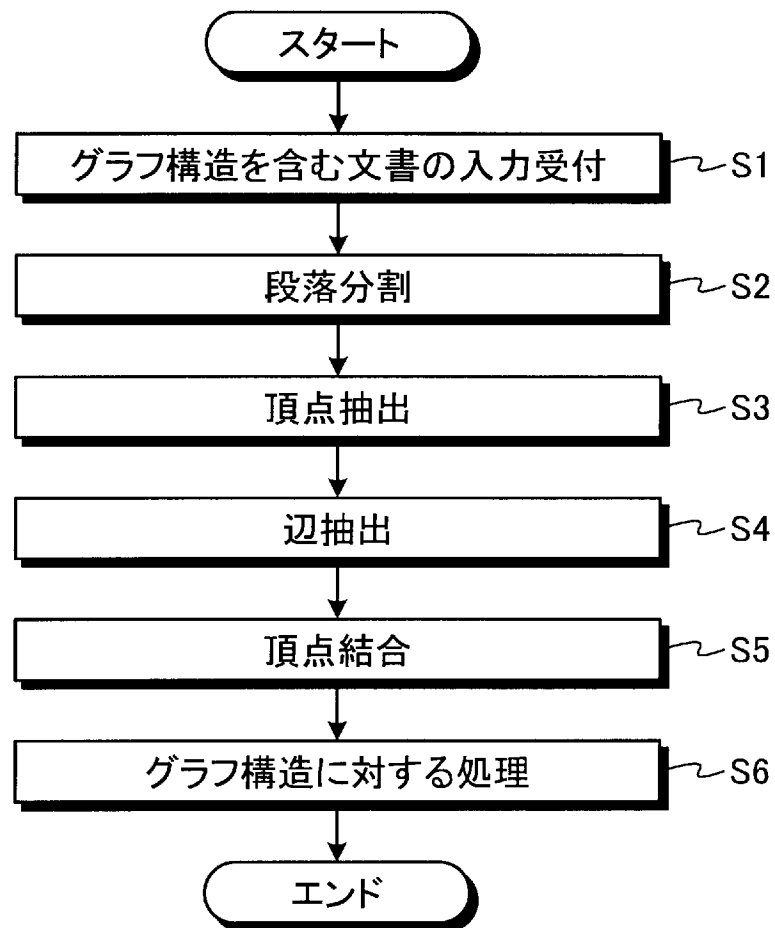
[図3]



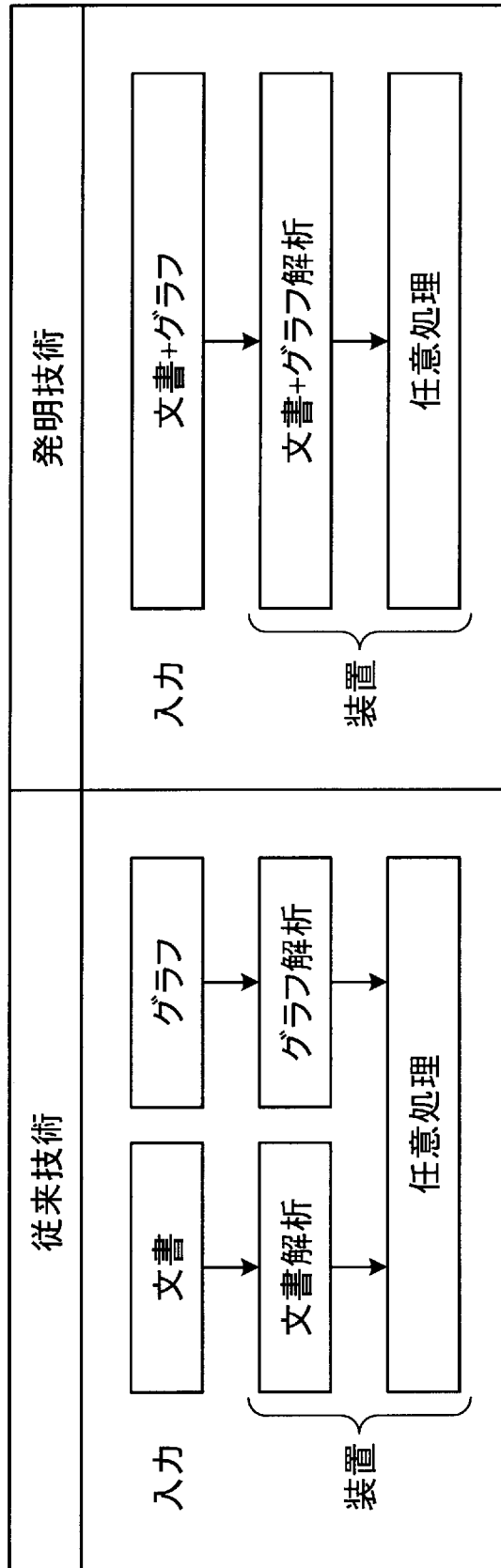
[図4]



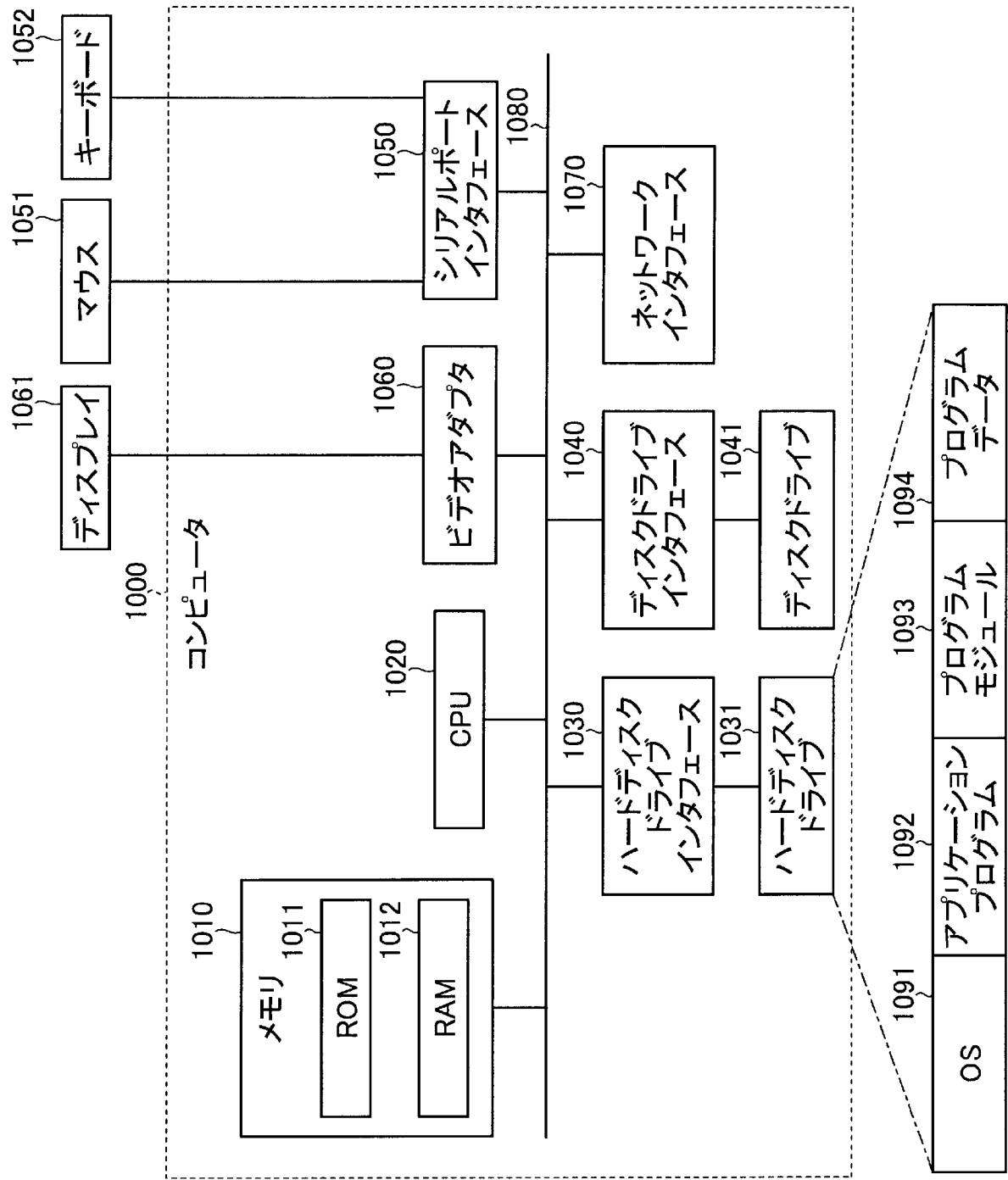
[図5]



[図6]



[図7]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2019/026736

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G06F16/36(2019.01) i, G06F17/22 (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)

Int.Cl. G06F16/36, G06F17/22

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

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2019
Registered utility model specifications of Japan	1996-2019
Published registered utility model applications of Japan	1994-2019

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	山寄麿与、外 1 名、文書からの脅威と脆弱性知識の自動抽出、情報処理学会シンポジウムシリーズ [CD-ROM], 16 October 2017, vol. 2017, no. 2, pp. 393-398, (YAMASAKI, Mayo et al., "Automatic Extraction of Threat and Vulnerability Knowledge from Documents", Symposium series of Information Processing Society of Japan [CD-ROM])	1-5
Y	US 2015/0310096 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 29 October 2015, paragraphs [0030]-[0040], [0059]-[0061] & CN 105095229 A	1-5

☐

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
13 September 2019 (13.09.2019)

Date of mailing of the international search report
24 September 2019 (24.09.2019)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G06F16/36(2019.01)i, G06F17/22(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. G06F16/36, G06F17/22

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	1971-2019年
日本国実用新案登録公報	1996-2019年
日本国登録実用新案公報	1994-2019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）

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
Y	山寄磨与、外1名、文書からの脅威と脆弱性知識の自動抽出、情報処理学会シンポジウムシリーズ [CD-ROM], 2017.10.16, Vol.2017, No.2, pp.393-398	1-5
Y	US 2015/0310096 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2015.10.29, 段落 0030-0040, 0059-0061 & CN 105095229 A	1-5

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

「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
「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
「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）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「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
「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
「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
「&」同一パテントファミリー文献

国際調査を完了した日

13.09.2019

国際調査報告の発送日

24.09.2019

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)

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）

早川 学

5 N

9 6 5 2

電話番号 03-3581-1101 内線 3586