

【特許請求の範囲】**【請求項 1】**

対象種に係わる論文に関する論文情報を記憶する対象種別論文記憶手段と、
前記対象種に係わるタンパク質のタンパク質名を少なくとも含むタンパク質情報を記憶する対象種別タンパク質記憶手段と、
抗体の製造者名を少なくとも含む製造者情報を記憶する製造者記憶手段と、
前記対象種別論文記憶手段に記憶された前記論文情報に基づいて、前記抗体について書かれている前記論文中の部分を特定する特定手段と、
前記論文中の部分に対して、前記対象種別タンパク質記憶手段に記憶された前記タンパク質名、および、前記製造者記憶手段に記憶された前記製造者名と照合する照合手段と、
前記照合手段により前記論文中の部分に、前記タンパク質名および前記製造者名と一致するものがあると照合された場合、当該タンパク質名および当該製造者名と、照合対象の前記論文に関する前記論文情報と、を対応付けた抗体情報を抗体情報記憶手段に登録する抗体情報登録手段と、
前記抗体情報記憶手段に登録された前記抗体情報に基づいて、任意の抗原に対して作製されている前記抗体が記載されている前記論文の数をランキング表示させるランキング表示手段と、
を備えたことを特徴とする、抗体情報取得装置。

10

【請求項 2】

前記照合手段により前記論文中の部分に、前記タンパク質名および/または前記製造者名と一致するものがないと照合された場合、当該タンパク質名の候補となる候補タンパク質名および/または当該製造者名の候補となる候補製造者名を抽出する候補抽出手段と、
前記候補抽出手段により抽出された前記候補タンパク質名および/または前記候補製造者名と、照合対象の前記論文に関する前記論文情報と、を対応付けた仮登録データを仮登録データ記憶手段に登録する仮登録データ登録手段と、
前記仮登録データに含まれる前記候補タンパク質名および/または前記候補製造者名が、所定数以上の前記論文で使用されている場合、当該候補タンパク質名を前記タンパク質名として前記対象種別タンパク質記憶手段に登録、および/または、当該候補製造者名を前記製造者名として前記製造者記憶手段に登録する再登録手段と、
を更に備えたことを特徴とする、請求項 1 に記載の抗体情報取得装置。

20

30

【請求項 3】

前記照合手段は、
基準となる基準文字列の位置に基づいて、前記論文中の部分の所定範囲に対して、前記対象種別タンパク質記憶手段に記憶された前記タンパク質名、および、前記製造者記憶手段に記憶された前記製造者名と照合することを特徴とする、請求項 1 または 2 に記載の抗体情報取得装置。

【請求項 4】

前記特定手段により特定された前記論文中の部分に対して、空白、カンマ、ピリオド、および、改行コードを含む制御文字を削除し、大文字に変換した空白等削除文字列に加工する文字列加工手段、
を更に備え、
前記照合手段は、
前記空白等削除文字列に対して、前記対象種別タンパク質記憶手段に記憶された前記タンパク質名、および、前記製造者記憶手段に記憶された前記製造者名と照合することを特徴とする、請求項 1 乃至 3 のいずれか一つに記載の抗体情報取得装置。

40

【請求項 5】

前記抗体情報登録手段は、
更に、前記抗体の区分情報、および/または、種情報を含む前記抗体情報を前記抗体情報記憶手段に登録することを特徴とする、請求項 1 乃至 4 のいずれか一つに記載の抗体情報取得装置。

50

【請求項 6】

前記照合手段は、

更に、基準となる基準文字列の位置に基づいて、前記論文中の部分の所定範囲に対して、前記抗体の区分情報、および/または、種情報を取得することを特徴とする、請求項 1 または 2 に記載の抗体情報取得装置。

【請求項 7】

前記論文中の部分は、

material and method が記載されている部分であることを特徴とする、請求項 1 乃至 6 のいずれか一つに記載の抗体情報取得装置。

【請求項 8】

抗体情報取得装置において実行される抗体情報取得方法であって、

対象種に係わる論文に関する論文情報を記憶する対象種別論文記憶手段に記憶された前記論文情報に基づいて、抗体について書かれている前記論文中の部分を特定する特定ステップと、

前記論文中の部分に対して、前記対象種に係わるタンパク質のタンパク質名を少なくとも含むタンパク質情報を記憶する対象種別タンパク質記憶手段に記憶された前記タンパク質名、および、前記抗体の製造者名を少なくとも含む製造者情報を記憶する製造者記憶手段に記憶された前記製造者名と照合する照合ステップと、

前記照合ステップにて前記論文中の部分に、前記タンパク質名および前記製造者名と一致するものがあると照合された場合、当該タンパク質名および当該製造者名と、照合対象の前記論文に関する前記論文情報と、を対応付けた抗体情報を抗体情報記憶手段に登録する抗体情報登録ステップと、

前記抗体情報記憶手段に登録された前記抗体情報に基づいて、任意の抗原に対して作製されている前記抗体が記載されている前記論文の数をランキング表示させるランキング表示ステップと、

を含むことを特徴とする、抗体情報取得方法。

【請求項 9】

抗体情報取得装置に実行させるためのプログラムであって、

対象種に係わる論文に関する論文情報を記憶する対象種別論文記憶手段に記憶された前記論文情報に基づいて、抗体について書かれている前記論文中の部分を特定する特定ステップと、

前記論文中の部分に対して、前記対象種に係わるタンパク質のタンパク質名を少なくとも含むタンパク質情報を記憶する対象種別タンパク質記憶手段に記憶された前記タンパク質名、および、前記抗体の製造者名を少なくとも含む製造者情報を記憶する製造者記憶手段に記憶された前記製造者名と照合する照合ステップと、

前記照合ステップにて前記論文中の部分に、前記タンパク質名および前記製造者名と一致するものがあると照合された場合、当該タンパク質名および当該製造者名と、照合対象の前記論文に関する前記論文情報と、を対応付けた抗体情報を抗体情報記憶手段に登録する抗体情報登録ステップと、

前記抗体情報記憶手段に登録された前記抗体情報に基づいて、任意の抗原に対して作製されている前記抗体が記載されている前記論文の数をランキング表示させるランキング表示ステップと、

を実行させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、抗体情報取得装置、抗体情報取得方法、および、プログラムに関する。

【背景技術】

【0002】

従来から、抗体情報を取得する技術が開示されている。

10

20

30

40

50

【0003】

非特許文献1乃至3に記載の手法においては、Antibodypedia (<http://www.antibodypedia.com/>)を用いて抗体を検索することが開示されている。

【0004】

また、非特許文献4に記載の手法においては、Antibody Registry (<http://antibodyregistry.org/>)を用いて抗体を同定することが開示されている。

【先行技術文献】

【非特許文献】

10

【0005】

【非特許文献1】A chromosome-centric analysis of antibodies directed toward the human proteome using Antibodypedia. Alm T, von Feilitzen K, Lundberg E, Sivertsson A, Uhlen M. J Proteome Res. 2014 Mar 7;13(3):1669-76. doi: 10.1021/pr4011525. Epub 2014 Feb 17.

【非特許文献2】Antibodypedia seeks to answer the question: how good is that antibody? Cottingham K. J Proteome Res. 2008 Oct;7(10):4213. doi: 10.1021/pr800626d. Epub 2008 Sep 4.

20

【非特許文献3】Antibodypedia, a portal for sharing antibody and antigen validation data. Bjorling E, Uhlen M. Mol Cell Proteomics. 2008 Oct;7(10):2028-37. doi: 10.1074/mcp.M800264-MCP200. Epub 2008 Jul 29.

【非特許文献4】On the reproducibility of science: unique identification of research resources in the biomedical literature. Vasilovsky NA, Brush MH, Paddock H, Ponting L, Tripathy SJ, LaRocca GM, Haendel MA. PeerJ. 2013 Sep 5;1:e148. doi: 10.7717/peerj.148. eCollection 2013.

30

【発明の概要】

【発明が解決しようとする課題】

【0006】

ここで、医学・生物学などの分野の当業者は、抗体を利用することが必要となることが多いが、従来手法では、同じ抗原に対して抗体が多くの製造会社で製造されているため、当業者が、高価なものが多い抗体について、どの抗体を取得すべきかについての選択に困ることが多く、選択に間違ふことも多いという問題点を有していた。

40

【0007】

本発明は、上記問題点を鑑みてなされたもので、医学・生物学系分野の当業者が、それぞれの作業で必要な抗体が既にあるか、あるとしたらどの抗体が最も自分の作業に合致したものを簡単に知る事が出来る抗体情報取得装置、抗体情報取得方法、および、プログラムを提供することを目的とする。

【課題を解決するための手段】

【0008】

このような目的を達成するため、本発明の抗体情報取得装置は、対象種に係わる論文に

50

関する論文情報を記憶する対象種別論文記憶手段と、前記対象種に係わるタンパク質のタンパク質名を少なくとも含むタンパク質情報を記憶する対象種別タンパク質記憶手段と、抗体の製造者名を少なくとも含む製造者情報を記憶する製造者記憶手段と、前記対象種別論文記憶手段に記憶された前記論文情報に基づいて、前記抗体について書かれている前記論文の部分を特定する特定手段と、前記論文の部分に対して、前記対象種別タンパク質記憶手段に記憶された前記タンパク質名、および、前記製造者記憶手段に記憶された前記製造者名と照合する照合手段と、前記照合手段により前記論文の部分に、前記タンパク質名および前記製造者名と一致するものがあると照合された場合、当該タンパク質名および当該製造者名と、照合対象の前記論文に関する前記論文情報と、を対応付けた抗体情報を抗体情報記憶手段に登録する抗体情報登録手段と、前記抗体情報記憶手段に登録された前記抗体情報に基づいて、任意の抗原に対して作製されている前記抗体が記載されている前記論文の数をランキング表示させるランキング表示手段と、を備えたことを特徴とする。

10

【0009】

また、本発明の抗体情報取得装置は、上記記載の抗体情報取得装置において、前記照合手段により前記論文の部分に、前記タンパク質名および/または前記製造者名と一致するものがないと照合された場合、当該タンパク質名の候補となる候補タンパク質名および/または当該製造者名の候補となる候補製造者名を抽出する候補抽出手段と、前記候補抽出手段により抽出された前記候補タンパク質名および/または前記候補製造者名と、照合対象の前記論文に関する前記論文情報と、を対応付けた仮登録データを仮登録データ記憶手段に登録する仮登録データ登録手段と、前記仮登録データに含まれる前記候補タンパク質名および/または前記候補製造者名が、所定数以上の前記論文で使用されている場合、当該候補タンパク質名を前記タンパク質名として前記対象種別タンパク質記憶手段に登録し、および/または、当該候補製造者名を前記製造者名として前記製造者記憶手段に登録する再登録手段と、を更に備えたことを特徴とする。

20

【0010】

また、本発明の抗体情報取得装置は、上記記載の抗体情報取得装置において、前記照合手段は、基準となる基準文字列の位置に基づいて、前記論文の部分の所定範囲に対して、前記対象種別タンパク質記憶手段に記憶された前記タンパク質名、および、前記製造者記憶手段に記憶された前記製造者名と照合することを特徴とする。

30

【0011】

また、本発明の抗体情報取得装置は、上記記載の抗体情報取得装置において、前記特定手段により特定された前記論文の部分に対して、空白、カンマ、ピリオド、および、改行コードを含む制御文字を削除し、大文字に変換した空白等削除文字列に加工する文字列加工手段、を更に備え、前記照合手段は、前記空白等削除文字列に対して、前記対象種別タンパク質記憶手段に記憶された前記タンパク質名、および、前記製造者記憶手段に記憶された前記製造者名と照合することを特徴とする。

【0012】

また、本発明の抗体情報取得装置は、上記記載の抗体情報取得装置において、前記抗体情報登録手段は、更に、前記抗体の区分情報、および/または、種情報を含む前記抗体情報を前記抗体情報記憶手段に登録することを特徴とする。

40

【0013】

また、本発明の抗体情報取得装置は、上記記載の抗体情報取得装置において、前記照合手段は、更に、基準となる基準文字列の位置に基づいて、前記論文の部分の所定範囲に対して、前記抗体の区分情報、および/または、種情報を取得することを特徴とする。

【0014】

また、本発明の抗体情報取得装置は、上記記載の抗体情報取得装置において、前記論文の部分は、material and methodが記載されている部分であることを特徴とする。

【0015】

50

また、本発明の抗体情報取得方法は、対象種に係わる論文に関する論文情報を記憶する対象種別論文記憶手段に記憶された前記論文情報に基づいて、抗体について書かれている前記論文中の部分を特定する特定ステップと、前記論文中の部分に対して、前記対象種に係わるタンパク質のタンパク質名を少なくとも含むタンパク質情報を記憶する対象種別タンパク質記憶手段に記憶された前記タンパク質名、および、前記抗体の製造者名を少なくとも含む製造者情報を記憶する製造者記憶手段に記憶された前記製造者名と照合する照合ステップと、前記照合ステップにて前記論文中の部分に、前記タンパク質名および前記製造者名と一致するものがあると照合された場合、当該タンパク質名および当該製造者名と、照合対象の前記論文に関する前記論文情報と、を対応付けた抗体情報を抗体情報記憶手段に登録する抗体情報登録ステップと、前記抗体情報記憶手段に登録された前記抗体情報に基づいて、任意の抗原に対して作製されている前記抗体が記載されている前記論文の数をランキング表示させるランキング表示ステップと、を含むことを特徴とする。

10

【0016】

また、本発明のプログラムは、対象種に係わる論文に関する論文情報を記憶する対象種別論文記憶手段に記憶された前記論文情報に基づいて、抗体について書かれている前記論文中の部分を特定する特定ステップと、前記論文中の部分に対して、前記対象種に係わるタンパク質のタンパク質名を少なくとも含むタンパク質情報を記憶する対象種別タンパク質記憶手段に記憶された前記タンパク質名、および、前記抗体の製造者名を少なくとも含む製造者情報を記憶する製造者記憶手段に記憶された前記製造者名と照合する照合ステップと、前記照合ステップにて前記論文中の部分に、前記タンパク質名および前記製造者名と一致するものがあると照合された場合、当該タンパク質名および当該製造者名と、照合対象の前記論文に関する前記論文情報と、を対応付けた抗体情報を抗体情報記憶手段に登録する抗体情報登録ステップと、前記抗体情報記憶手段に登録された前記抗体情報に基づいて、任意の抗原に対して作製されている前記抗体が記載されている前記論文の数をランキング表示させるランキング表示ステップと、を抗体情報取得装置に実行させることを特徴とする。

20

【発明の効果】

【0017】

この発明によれば、論文情報に基づいて、抗体について書かれている論文中の部分を特定し、論文中の部分に対して、タンパク質名、および、製造者名と照合し、論文中の部分に、タンパク質名および製造者名と一致するものがあると照合された場合、当該タンパク質名および当該製造者名と、照合対象の論文に関する論文情報と、を対応付けた抗体情報を登録し、登録された抗体情報に基づいて、任意の抗原に対して作製されている抗体が記載されている論文の数をランキング表示させるので、ユーザは使いたい抗体の有無、製造、または、利用情報を簡単に閲覧でき、ある抗原に対して作られている抗体のうち、ユーザコミュニティで最も多く使っている抗体を簡単に知り、抗体の取得が容易になるという効果を奏する。

30

【0018】

この発明によれば、論文中の部分に、タンパク質名および/または製造者名と一致するものがないと照合された場合、当該タンパク質名の候補となる候補タンパク質名および/または当該製造者名の候補となる候補製造者名を抽出し、抽出された候補タンパク質名および/または候補製造者名と、照合対象の論文に関する論文情報と、を対応付けた仮登録データを登録し、仮登録データに含まれる候補タンパク質名および/または候補製造者名が、所定数以上の論文で使用されている場合、当該候補タンパク質名をタンパク質名として登録、および/または、当該候補製造者名を製造者名として登録するので、既存のデータベースに登録されていないタンパク質情報、および、会社情報を取得できるという効果を奏する。また、この発明によれば、論文内の抗体情報が記載されている周辺の文字列から、タンパク質名および製造会社名を推測し、分析後にデータベースに新規登録することで、抗体情報検索の精度と速度を高めることが可能となるという効果を奏する。

40

【0019】

50

この発明によれば、基準となる基準文字列の位置に基づいて、論文中の部分の所定範囲に対して、記憶されたタンパク質名、および、記憶された製造者名と照合するので、抗体情報検索の速度を高めることが可能となるという効果を奏する。

【0020】

この発明によれば、特定された論文中の部分に対して、空白、カンマ、ピリオド、および、改行コードを含む制御文字を削除し、大文字に変換した空白等削除文字列に加工し、空白等削除文字列に対して、記憶されたタンパク質名、および、記憶された製造者名と照合するので、文字列1つに対し検索を行うことによって、ワード（語）単位の検索に比べ高速化が実現できるという効果を奏する。

【0021】

この発明によれば、更に、抗体の区分情報、および/または、種情報を含む抗体情報を登録するので、より詳細な抗体情報を取得できるという効果を奏する。

【0022】

この発明によれば、更に、基準となる基準文字列の位置に基づいて、論文中の部分の所定範囲に対して、抗体の区分情報、および/または、種情報を取得するので、抗体情報検索の速度を高めることが可能となるという効果を奏する。

【0023】

この発明によれば、論文中の部分は、material and methodが記載されている部分であるので、抗体情報検索の精度と速度を高めることが可能となるという効果を奏する。

【図面の簡単な説明】

【0024】

【図1】図1は、本実施の形態の基本原理を示すフローチャートである。

【図2】図2は、本実施の形態における抗体情報取得装置の構成の一例を示すブロック図である。

【図3】図3は、本実施の形態の抗体情報取得装置の処理の一例を示すフローチャートである。

【図4】図4は、本実施の形態における表示画面の一例を示す図である。

【図5】図5は、本実施の形態における表示画面の一例を示す図である。

【図6】図6は、本実施の形態の抗体情報取得装置の処理の一例を示すフローチャートである。

【発明を実施するための形態】

【0025】

以下に、本発明にかかる抗体情報取得装置、抗体情報取得方法、および、プログラムの実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態によりこの発明が限定されるものではない。

【0026】

[本発明の実施の形態の概要]

以下、本発明の実施の形態の概要について図1を参照して説明し、その後、本実施の形態の構成および処理等について詳細に説明する。図1は、本実施の形態の基本原理を示すフローチャートである。本実施の形態は、概略的に、以下の基本的特徴を有する。

【0027】

すなわち、本実施の形態の抗体情報取得装置は、図1に示すように、記憶された論文情報に基づいて、抗体について書かれている論文中の部分を特定する（ステップSA-1）。

【0028】

そして、抗体情報取得装置は、論文中の部分に対して、記憶されたタンパク質名、および、記憶された製造者名（製造会社名）と照合する（ステップSA-2）。

【0029】

そして、抗体情報取得装置は、論文中の部分に、記憶されたタンパク質名および製造者

10

20

30

40

50

名（製造会社名）と一致するものがあるか否か判定する（ステップ S A - 3）。

【 0 0 3 0 】

そして、抗体情報取得装置は、ステップ S A - 3 にて論文中の部分に、タンパク質名および製造者名（製造会社名）と一致するものがあると照合された場合（ステップ S A - 3 : Y e s ）、当該タンパク質名および当該製造者名（製造会社名）と、照合対象の論文に関する論文情報と、を対応付けた抗体情報を登録する（ステップ S A - 4 ）。

【 0 0 3 1 】

そして、抗体情報取得装置は、登録された抗体情報に基づいて、任意の抗原に対して作製されている抗体が記載されている論文の数をランキング表示させ（ステップ S A - 5 ）、処理を終了する。

10

【 0 0 3 2 】

一方、抗体情報取得装置は、ステップ S A - 3 にて論文中の部分に、記憶されたタンパク質名および / または製造者名（製造会社名）と一致するものがないと照合された場合（ステップ S A - 3 : N o ）、当該タンパク質名および / または当該製造者名（製造会社名）の候補となる候補タンパク質名および / または候補製造者名を抽出する（ステップ S A - 6 ）。

【 0 0 3 3 】

そして、抗体情報取得装置は、抽出された候補タンパク質名および / または候補製造者名と、照合対象の論文に関する論文情報と、を対応付けた仮登録データを登録する（ステップ S A - 7 ）。

20

【 0 0 3 4 】

そして、抗体情報取得装置は、仮登録データに含まれる候補タンパク質名および / または候補製造者名が、所定数以上の論文で使用されている場合、当該候補タンパク質名をタンパク質名として対象種別タンパク質記憶手段に登録、および / または、当該候補製造者名を製造者名として製造者記憶手段に登録し（ステップ S A - 8 ）、処理をステップ S A - 2 に移行させる。

【 0 0 3 5 】

このように、本実施の形態における抗体情報の取得においては、タンパク質データベースおよび製造会社（製造者）データベースの情報と完全一致した抗体情報を取得し、抗体情報データベースへ登録する登録機能と、タンパク質データベースおよび / または製造会社（製造者）データベースの情報と一致しなかったが、抗体情報の可能性がある文字列を取得し、検証後、確定したデータを抗体情報データベースに登録する再登録機能に機能分割している。

30

【 0 0 3 6 】

それにより、本実施の形態においては、内容が一般公開されている論文の中から抗体名、製造会社（製造者）および抗体利用結果情報などの抗体情報を収集し、ある抗原に対して作製されている多種類の抗体を分別し、それぞれの抗体が使われている論文の数が多いものからランキング表示させ、関連文献の抗体利用結果情報を閲覧できるシステムを提供している。

40

【 0 0 3 7 】

以上で、本実施の形態の概要の説明を終える。

【 0 0 3 8 】

[抗体情報取得装置 1 0 0 の構成]

次に、本実施の形態における抗体情報取得装置 1 0 0 の構成の詳細について、図 2 を参照して以下に説明する。図 2 は、本実施の形態における抗体情報取得装置 1 0 0 の構成の一例を示すブロック図であり、該構成のうち本発明に係る部分のみを概念的に示している。ここで、本実施の形態における抗体情報取得装置 1 0 0 においては、各構成が一筐体内に全て備えられ、単独で処理を行うもの（スタンドアローン型）を、抗体情報取得装置 1 0 0 として説明するが、当該実施例に限らず、各構成が分離した筐体内に備えられ、ネットワーク 3 0 0 等を介して接続されて一つの概念としての装置を構成するもの（例え

50

ば、クラウドコンピューティング等)であってもよい。

【0039】

図2において、外部システム200は、ネットワーク300を介して、抗体情報取得装置100と相互に接続され、論文情報、タンパク質情報、および/もしくは、製造者(製造会社)情報等に関する外部データベース、ならびに/または、ユーザインターフェース等を実行するウェブサイトを提供する機能等を有していてもよい。

【0040】

ここで、外部システム200は、WEBサーバやASPサーバ等として構成していてもよい。また、外部システム200のハードウェア構成は、一般に市販されるワークステーション、パーソナルコンピュータ等の情報処理装置およびその付属装置により構成していてもよい。また、外部システム200の各機能は、外部システム200のハードウェア構成中のCPU、ディスク装置、メモリ装置、入力装置、出力装置、通信制御装置等およびそれらを制御するプログラム等により実現されてもよい。

【0041】

また、ネットワーク300は、抗体情報取得装置100と外部システム200等とを相互に接続する機能を有し、例えば、インターネット等である。

【0042】

また、抗体情報取得装置100は、概略的に、制御部102と通信制御インターフェース部104と記憶部106と入出力制御インターフェース部108とを備える。ここで、抗体情報取得装置100は、更に、表示部112を少なくとも含む出力部、および、入力部114を備えていてもよい。また、出力部は、更に、音声出力部、および、印刷出力部等を含んでいてもよい。ここで、制御部102は、抗体情報取得装置100の全体を統括的に制御するCPU等である。

【0043】

また、通信制御インターフェース部104は、通信回線等に接続されるルータ等の通信装置(図示せず)に接続されるインターフェースであり、入出力制御インターフェース部108は、出力部、および、入力部114に接続されるインターフェースである。また、記憶部106は、各種のデータベースやテーブルなどを格納する装置である。これら抗体情報取得装置100の各部は任意の通信路を介して通信可能に接続されている。更に、この抗体情報取得装置100は、ルータ等の通信装置および専用線等の有線または無線の通信回線を介して、ネットワーク300に通信可能に接続されている。

【0044】

記憶部106に格納される各種のデータベースやテーブル(対象種別論文データベース106a、対象種別タンパク質データベース106b、製造者データベース106c、抗体情報データベース106d、仮登録データデータベース106e、および、仮登録抗体情報データベース106f)は、固定ディスク装置等のストレージ手段である。例えば、記憶部106は、各種処理に用いる各種のプログラム、テーブル、ファイル、データベース、および、ウェブページ等を格納する。

【0045】

これら記憶部106の各構成要素のうち、対象種別論文データベース106aは、対象種に係わる論文に関する論文情報を記憶する対象種別論文記憶手段である。ここで、対象種別論文データベース106aは、対象種に係わる公開論文に関する論文情報が格納されたデータベースであってもよい。また、論文情報は、論文番号(一意)、文章(全文)、使用画像、および/または、掲載年月日(公開年月日)に関する情報を含んでいてもよい。また、論文情報は、検索処理に関わる部分のみ抜粋されたものであってもよい。

【0046】

これら論文情報は、対象種別論文データベース106aに予め記憶されており、抗体情報取得装置100の制御部102は、定期的に、および/または、制御部102による処理に応じてネットワーク300を介して最新のデータを外部システム200(例えば、PubMed Central(登録商標)(<http://www.ncbi.nlm>

10

20

30

40

50

n i h . g o v / p m c /) 等) からダウンロードして対象種別論文データベース106aに記憶された論文情報をアップデートしてもよい。

【0047】

また、対象種別タンパク質データベース106bは、対象種に係わるタンパク質のタンパク質名を少なくとも含むタンパク質情報を記憶する対象種別タンパク質記憶手段である。ここで、タンパク質情報は、タンパク質番号(一意)、タンパク質名(Protein Name)、遺伝子名(Gene Name)、および/または、シノニム名(別名)に関する情報(項目)を含んでいてもよい。また、タンパク質情報は、検索処理に関わる部分のみ抜粋されたものであってもよい。

【0048】

これらタンパク質情報は、対象種別タンパク質データベース106bに予め記憶されており、抗体情報取得装置100の制御部102は、定期的に、および/または、制御部102による処理に応じてネットワーク300を介して最新のデータを外部システム200(例えば、UniProt(<http://www.uniprot.org/>)等)からダウンロードして対象種別タンパク質データベース106bに記憶されたタンパク質情報をアップデートしてもよい。

【0049】

また、タンパク質については論文内で複数の記載方法が存在し得るため、対象種別タンパク質データベース106bの中に、マスタデータベースのほか、名称管理データベースを含んでいてもよい。ここで、名称管理データベースは、確定データ登録処理により確定したデータを追加登録していく成長型データベースであってもよい。ここで、名称管理データベースに記憶されたタンパク質情報は、タンパク質番号(外部キー:マスタデータとの関連付け)、および/または、名称(タンパク質名、遺伝子名、および/または、シノニム名)に関する情報(項目)を含んでいてもよい。また、二つの項目の組み合わせが一意(主キー)であってもよい。また、名称管理データベースに記憶されたタンパク質情報は、検索処理に関わる部分のみ抜粋されたものであってもよい。

【0050】

ここで、製造者(製造会社)データベース106cは、抗体の製造者名(製造会社名)を少なくとも含む製造者情報を記憶する製造者記憶手段である。これら製造者名(製造会社名)は、製造者(製造会社)データベース106cに予め記憶されており、抗体情報取得装置100の制御部102は、定期的に、および/または、制御部102による処理に応じてネットワーク300を介して最新のデータを外部システム200からダウンロードして製造者(製造会社)データベース106cに記憶された製造者名(製造会社名)をアップデートしてもよい。

【0051】

また、抗体情報データベース106dは、対象種別論文データベース106aに記憶された論文情報に基づく論文中の部分に一致するものがある、対象種別タンパク質データベース106bに記憶されたタンパク質名、および、製造者(製造会社)データベース106cに記憶された製造者名と、照合対象の論文に関する論文情報と、を対応付けた抗体情報を記憶する抗体情報記憶手段である。

【0052】

また、仮登録データ(仮登録ビッグデータ)データベース106eは、仮登録データ(仮登録ビッグデータ)を記憶する仮登録データ記憶手段である。ここで、仮登録データ(仮登録ビッグデータ)データベース106eは、対象種別論文データベース106aに記憶された論文情報に基づく論文中の部分に、対象種別タンパク質データベース106bに記憶されたタンパク質名と一致するものがない候補タンパク質名、または、製造者(製造会社)データベース106cに記憶された製造者名と一致するものがない候補製造者名と、照合対象の論文に関する論文情報と、を対応付けた仮登録データを記憶していてもよい。

【0053】

10

20

30

40

50

また、仮登録抗体情報データベース106fは、仮登録抗体情報を記憶する仮登録抗体情報記憶手段である。ここで、仮登録抗体情報データベース106fは、候補タンパク質名および候補製造者名、対象種別論文データベース106aに記憶された論文情報に基づく論文中の部分に一致するものがある対象種別タンパク質データベース106bに記憶されたタンパク質名および候補製造者名、または、候補タンパク質名および対象種別論文データベース106aに記憶された論文情報に基づく論文中の部分に一致するものがある製造者データベース106cに記憶された製造者名と、照合対象の論文に関する論文情報と、を対応付けた仮登録抗体情報を記憶していてもよい。

【0054】

また、図2において、通信制御インターフェース部104は、抗体情報取得装置100とネットワーク300（またはルータ等の通信装置）との間における通信制御を行う。すなわち、通信制御インターフェース部104は、外部システム200、および、他の端末等と通信回線を介してデータを通信する機能を有する。

10

【0055】

また、入出力制御インターフェース部108は、出力部（表示部112）、および、入力部114の制御を行う。

【0056】

ここで、表示部112としては、アプリケーション等の表示画面を表示する表示手段（例えば、液晶または有機EL等から構成されるディスプレイ、モニタ、または、タッチパネル等）であってもよい。また、入力部114は、例えば、キー入力部、タッチパネル、コントロールパッド（例えば、タッチパッド、および、ゲームパッド等）、マウス、キーボード、または、マイク等であってもよい。また、音声出力部としては、例えば、スピーカ等であってもよい。また、印刷出力部としては、例えば、プリンタ等であってもよい。

20

【0057】

また、図2において、制御部102は、OS（Operating System）等の制御プログラムや、各種の処理手順等を規定したプログラム、および、所要データを格納するための内部メモリを有する。そして、制御部102は、これらのプログラム等により、種々の処理を実行するための情報処理を行う。制御部102は、機能概念的に、特定部102a、文字列加工部102b、照合部102c、抗体情報登録部102d、ランキング表示部102e、候補抽出部102f、仮登録データ登録部102g、仮登録抗体情報登録部102h、および、再登録部102iを備える。

30

【0058】

このうち、特定部102aは、対象種別論文データベース106aに記憶された論文情報に基づいて、抗体について書かれている論文中の部分を特定する特定手段である。ここで、論文中の部分は、material and methodが記載されている部分であってもよい。

【0059】

また、文字列加工部102bは、特定部102aにより特定された論文中の部分に対して、空白、カンマ、ピリオド、および、改行コードを含む制御文字を削除し、大文字に変換した空白等削除文字列に加工する文字列加工手段である。

40

【0060】

また、照合部102cは、特定部102aにより特定された論文中の部分に対して、対象種別タンパク質データベース106bに記憶されたタンパク質名、および、製造者データベース106cに記憶された製造者名と照合する照合手段である。ここで、照合部102cは、基準となる基準文字列の位置に基づいて、特定部102aにより特定された論文中の部分の所定範囲に対して、対象種別タンパク質データベース106bに記憶されたタンパク質名、および、製造者データベース106cに記憶された製造者名を照合してもよい。また、照合部102cは、文字列加工部102bにより加工された空白等削除文字列に対して、対象種別タンパク質データベース106bに記憶されたタンパク質名、および、製造者データベース106cに記憶された製造者名を照合してもよい。また、照合部1

50

02cは、更に、基準となる基準文字列の位置に基づいて、特定部102aにより特定された論文中の部分の所定範囲に対して、抗体の区分情報、および/または、種情報を取得してもよい。

【0061】

また、抗体情報登録部102dは、照合部102cにより論文中の部分に、タンパク質名および製造者名と一致するものがあると照合された場合、当該タンパク質名および当該製造者名と、照合対象の論文に関する論文情報と、を対応付けた抗体情報を抗体情報データベース106dに登録(格納)する抗体情報登録手段である。ここで、抗体情報登録部102dは、更に、抗体の区分情報、および/または、種情報を含む抗体情報を抗体情報データベース106dに登録してもよい。

10

【0062】

また、ランキング表示部102eは、抗体情報データベース106dに登録された抗体情報に基づいて、任意の抗原に対して作製されている抗体が記載されている論文の数をランキング表示させるランキング表示手段である。ここで、抗体情報データベース106dに登録された抗体情報に基づいて、任意の抗原に対して作製されている抗体が記載されている論文の数を表示部112にランキング表示させてもよい。また、ランキング表示部102eは、更に、抗体情報データベース106dに登録された抗体情報に基づいて、任意の抗原に対して作製されている抗体が記載されている論文に関する論文情報を表示させてもよい。また、ランキング表示部102eは、更に、抗体情報データベース106dに登録された抗体情報に基づいて、任意の抗原に関するタンパク質情報を表示させてもよい。また、任意の抗原は、ユーザにより入力部114を介して入力された抗原名に基づくものであってもよい。

20

【0063】

また、候補抽出部102fは、照合部102cにより論文中の部分に、タンパク質名および/または製造者名と一致するものがないと照合された場合、当該タンパク質名の候補となる候補タンパク質名および/または当該製造者名の候補となる候補製造者名を抽出する候補抽出手段である。

【0064】

また、仮登録データ登録部102gは、仮登録データを仮登録データデータベース106eに登録(格納)する仮登録データ登録手段である。ここで、仮登録データ登録部102gは、候補抽出部102fにより抽出された候補タンパク質名および/または候補製造者名と、照合対象の論文に関する論文情報と、を対応付けた仮登録データを仮登録データデータベース106eに登録してもよい。また、仮登録データ登録部102gは、特定部102aにより特定された論文中の部分に関する論文情報を仮登録データデータベース106eに登録してもよい。

30

【0065】

また、仮登録抗体情報登録部102hは、仮登録抗体情報を仮登録抗体情報データベース106fに登録(格納)する仮登録抗体情報登録手段である。ここで、仮登録抗体情報登録部102hは、候補抽出部102fにより抽出された候補タンパク質名および候補抽出部102fにより抽出された候補製造者名、特定部102aにより特定された論文中の部分に一致するものがある対象種別タンパク質データベース106bに記憶されたタンパク質名および候補抽出部102fにより抽出された候補製造者名、または、候補抽出部102fにより抽出された候補タンパク質名および特定部102aにより特定された論文中の部分に一致するものがある製造者データベース106cに記憶された製造者名と、照合対象の論文に関する論文情報と、を対応付けた仮登録抗体情報を仮登録抗体情報データベース106fに登録してもよい。

40

【0066】

また、再登録部102iは、仮登録データデータベース106eに登録された仮登録データに含まれる候補タンパク質名および/または候補製造者名が、所定数以上の論文で使用されている場合、当該候補タンパク質名をタンパク質名として対象種別タンパク質デー

50

データベース106bに登録(格納)、および/または、当該候補製造者名を製造者名として製造者データベース106cに登録(格納)する再登録手段である。

【0067】

以上で、本実施の形態における抗体情報取得装置100の構成の一例の説明を終える。

【0068】

[抗体情報取得装置100の処理]

次に、このように構成された本実施の形態における抗体情報取得装置100の処理の詳細について、以下に図3乃至図6を参照して詳細に説明する。

【0069】

[登録プロセス]

まず、図3乃至図5を参照して、本実施の形態における登録プロセスについて説明する。図3は、本実施の形態における抗体情報取得装置100の処理の一例を示すフローチャートである。

【0070】

まず、図3に示すように、特定部102aは、対象種別論文データベース106a(D1)に記憶された論文情報を取得する(ステップM-R1)。

【0071】

そして、特定部102aは、対象種別論文データベース106a(D1)に記憶された論文情報に基づく対象種に係わる論文(M-R1)内から実験プロトコルが記載されているMaterial and Methodが書かれている文章(M-R2)を抽出する(ステップM-P1)。すなわち、このプロセスでは、論文内から抗体情報が書かれている場所を特定している。

【0072】

ここで、本実施の形態においては、対象種別タンパク質データベース106bには大量のデータがあり(例えば、ヒトで約2万種類6万語)、すべてのデータと比較するには時間がかかるという問題があるため、Material and Methodが書かれている章(M-R2)の中から候補となりうるタンパク質名をすべて取得することで、絞り込みを行っている。

【0073】

そして、照合部102cは、特定部102aにより特定された論文中の部分から「anti」という文字列が含まれている部分を探索する(ステップM-P3)。

【0074】

そして、照合部102cは、「anti」という文字列(基準文字列)が含まれている部分が見つかった場合、発見場所から±200文字分の長いセンテンス(Anti文字列群)を抽出する(ステップM-R4)。

【0075】

そして、照合部102cは、当該長いセンテンス(Anti文字列群)から、対象種別タンパク質データベース106b(D2)に記憶されたタンパク質名、および、製造会社データベース106c(D3)に記憶された製造会社名(製造者名)を登録データ群(M-R7)として抽出する(ステップM-P5)。すなわち、このプロセスでは、データベースとのマッチング処理を行うことにより、プロテイン名称、および、会社名のデータベースと照合することで抗体情報を確定させている。

【0076】

そして、抗体情報登録部102dは、照合部102cにより抽出された登録データ群(M-R7)と、照合対象の論文に関する論文情報と、を対応付けた抗体情報を抗体情報データベース106d(D4)に登録する(ステップM-P7)。抗体情報は、更に、抗体の区分情報、および/または、種情報を含む抗体情報を含んでいてもよい。

【0077】

そして、ユーザにより入力部114を介して抗原名が指定された場合、ランキング表示部102eは、抗体情報データベース106dに登録された抗体情報に基づいて、当該抗

10

20

30

40

50

原名に基づく抗原に対して作製されている抗体が記載されている論文の数を表示部 1 1 2 にランキング表示させ (ステップ M - R 8)、処理を終了する。ここで、ランキング表示部 1 0 2 e は、更に、抗体情報データベース 1 0 6 d に登録された抗体情報に基づいて、ユーザにより指定された抗原名に基づく抗原に対して作製されている抗体が記載されている論文に関する論文情報、および / または、ユーザにより指定された抗原名に基づく抗原に関するタンパク質情報を表示部 1 1 2 に表示させてもよい。

【 0 0 7 8 】

ここで、図 4 および図 5 を参照して、本実施の形態におけるランキング表示の一例について説明する。図 4 および図 5 は、本実施の形態における表示画面の一例を示す図である。

10

【 0 0 7 9 】

本実施の形態においては、図 4 に示す検索キーワード (Search Word) 入力画面において、ユーザにより入力部 1 1 4 を介して検索キーワードが入力された場合、図 5 に示すように、検索キーワードを含む抗原名の抗原に対して作製されている抗体が記載されている論文の数を表示部 1 1 2 にランキング表示させてもよい。さらに、本実施の形態においては、図 5 に示すように、検索結果として、ランキング表示部 1 0 2 e は、更に、抗体情報データベース 1 0 6 d に登録された抗体情報に基づいて、ユーザにより指定された検索キーワードを含む抗原名のタンパク質の候補リストと、当該タンパク質が記載された論文中の図表の画像と、を対応付けて表示部 1 1 2 に表示させてもよい。

20

【 0 0 8 0 】

図 3 に戻り、上記の処理の高速化を図る場合、文字列加工部 1 0 2 b は、ステップ M - P 1 にて特定部 1 0 2 a により抽出された Material and Method が書かれている文章 (M - R 2) に対して、空白、カンマ、ピリオド、および、改行コードを含む制御文字を削除し、大文字に変換した一つの長い文字列である空白等削除文字列 (M - R 3) に文字列加工する (ステップ M - P 2)。

【 0 0 8 1 】

そして、照合部 1 0 2 c は、文字列加工部 1 0 2 b により加工された空白等削除文字列 (M - R 3) に対して、対象種別タンパク質データベース 1 0 6 b (D 2) に記憶されたタンパク質名、および、製造会社データベース 1 0 6 c (D 3) に記憶された製造会社名 (製造者名) と照合することにより候補データ群 (M - R 5) として抽出し (ステップ M - P 4)、処理をステップ M - P 5 に移行させる。

30

【 0 0 8 2 】

このように、本実施の形態においては、論文内から抗体情報を抽出し、登録することができる。また、本実施の形態においては、論文毎に異なる抗体表記法についてパターン解析を行い、抗体情報を特定することができる。また、本実施の形態においては、タンパク質名データ (ヒトの場合: 2 万種類約 6 万語) との比較を高速に行う工夫がなされている。また、本実施の形態において、抗体表記の差異は吸収することができる。

【 0 0 8 3 】

また、論文内で抗体情報について書かれている部分には、パターン A : anti - [タンパク質名] ([製造会社])、または、パターン B : antibody against [タンパク質名] from [製造会社] など様々な表記方法がある。そこで、本実施の形態においては、この抗体表記そのものの差異についてパターン解析を行い、論文ごとに異なる抗体表記そのものについてその差異を吸収し、抗体情報を抽出してもよい。

40

【 0 0 8 4 】

そこで、本実施の形態の登録プロセスにおける、基準文字列に基づく照合処理の一例について説明する。

【 0 0 8 5 】

例えば、本実施の形態においては、基準となる文字列 (基準文字列) を 4 カテゴリー (anti**、antibody against***、antibody to***、および、mab / pab***) 定義する。ここで、抗体情報は基準文字列の前後にお

50

いて、前側にはmonoclonal/polyclonal区分および種の情報が存在していることが多く、後側には抗原のタンパク質情報および製造会社（製造者）情報が存在していることが多い。

【0086】

そこで、本実施の形態においては、基準文字列が含まれているセンテンスの先頭部分から基準文字列までの部分（前側）を網羅的に検索し、monoclonal/polyclonal区分および種（host species、または、reactive species等）の情報を取得してもよい。

【0087】

また、本実施の形態においては、基準文字列から基準文字列が含まれているセンテンスの終わりまでの部分（後側）を網羅的に検索し、タンパク質名および製造会社名（製造者名）を取得してもよい。

10

【0088】

ここで、センテンスの終わりについて、本来ピリオド（.）までをとればいいのだが、センテンス内にピリオドが存在する場合がある（例えば、「Dr.」または「0.025」など）。そこで、本実施の形態においては、後側は、基準文字列から200文字分の文字情報すべてを対象としている。また、本実施の形態において、ピリオドについては、前後が数字のもの（小数点）、敬称のもの（Ph.D.、Dr.、M.D.、Mr.、または、Mrs.）または会社情報（Co.Ltd.、または、Inc.）以外のピリオドをセンテンスの終止符と認識させている。

20

【0089】

また、論文内において、タンパク質情報（抗原）については基準文字列のすぐ後ろにくるパターン、および、anti-[種(reactive species)]の後に続くパターンが存在している。また、タンパク質名は、複数語からなるものが多く存在している。

【0090】

そこで、対象となる部分から、A：1語分（空白まで）、B：区切り分（カンマor括弧のはじまり「（」まで）、C：センテンス分、D：200文字分の最後まで、または、anti-[種]が見つかったパターンではE：「antibody」または「antibodies」が存在するまでの5つの文字列はタンパク質以外の情報も存在している場合がある。そこで、本実施の形態においては、この5つの文字列ワード区切りの文字列パターンに細分化する。

30

【0091】

本実施の形態において、具体的には、「AAAAA BBBBBB CCCCC DDDDD」という文字列において、1.AAAAA、2.BBBBBB、3.CCCCCC、4.DDDDD、5.AAAAA BBBBBB、6.BBBBBB CCCCCC、7.CCCC DD DDD、8.AAAAA BBBBBB CCCCCC、9.BBBBBB CCC CC DD DDD、および、10.AAAAA BBBBBB CCCCC DDDDDに対して、対象種別タンパク質データベース106bと照合（比較）してもよい。

【0092】

また、論文中の記載において、製造会社名（製造者名）は、後述するように、括弧の中に存在するパターン、fromの後に存在するパターン、または、製造者の場合は「gift」文字列がある単語の前後に存在するパターンが存在する。但し、括弧の中は、希釈率またはクローン番号を記載している場合もある。

40

【0093】

具体的には、抗体との関連付けで、anti-[タンパク質]***（[製造者]）、anti-[タンパク質]（[クローン番号]など）***（[製造者]）、anti-[タンパク質]、anti-[タンパク質]、または、anti-[タンパク質]from（[製造者]）などのような記載も存在する。

【0094】

50

そこで、本実施の形態において、これらを網羅するために1. a n t iの後からセンテンス終了まで200文字分検索し、1、2番目に出現する括弧の中、2. a n t iの後から200文字分検索し「f r o m」が出現した後の文字列からセンテンス終了まで、または、3. 「g i f t」が存在する単語の前後5語分、を取得し、データベース内の情報と照合してもよい。

【0095】

以上で、本実施の形態における登録プロセスの一例の説明を終える。

【0096】

[再登録プロセス]

次に、図3および図6を参照して、本実施の形態における再登録プロセスについて説明する。図6は、本実施の形態における抗体情報取得装置100の処理の一例を示すフローチャートである。

10

【0097】

まず、図3に示すように、ステップM-R4にて照合部102cにより抽出された「a n t i」という文字列が含まれているA n t i文字列群と論文番号とを解析用データとして仮登録ビッグデータデータベース106e(D5)に登録する(ステップM-P9)。これは、マスターデータベース(対象種別タンパク質データベース106b)の新規登録が発生した際、再度登録プロセスを行うためのものである。

【0098】

そして、候補抽出部102fは、A n t i文字列群から、タンパク質名の候補となる全ての候補タンパク質名および/または製造者名の候補となる全ての候補製造者名を仮登録データ群(M-R6)として抽出し(ステップM-P6)、仮登録データ群を解析用データとして仮登録ビッグデータデータベース106e(D5)に登録する(ステップM-P9)。

20

【0099】

このように、再登録プロセスを行う準備段階として、登録プロセスにおけるステップM-P5の処理に基づいて、仮登録するデータを抽出する。また、ステップM-P6の処理においては、候補となる論文内文字列をすべて仮データとしてデータベース(仮登録ビッグデータ)に格納している。結果として、データベースにマッチしない候補となる文字列が大量に登録されることになり、複数の論文検索時にこの処理を行うことで、複数論文で使用された文字列が登録情報となり得る可能性が高いということになる。

30

【0100】

また、仮登録抗体情報登録部102hは、A n t i文字列群から、仮登録抗体情報を抽出し、仮登録抗体情報データベース106f(D6)に登録する(ステップM-P8)。

【0101】

そして、図6に示すように、再登録部102iは、仮登録ビッグデータデータベース106e(D5)に登録された仮登録データ群(M-R6)に対応する仮登録文字列情報データ(D5-1)に含まれる候補タンパク質名および/または候補製造者名が、所定数以上の論文で使用されている、すなわち、複数論文で使用されていると判定した場合、登録情報となり得る可能性が高いため、当該候補タンパク質名および/または当該候補製造者名を確定データ(R-R2)として取得する(ステップR-P1)。すなわち、本実施の形態においては、複数論文での出現回数に基づいて、信頼性が高い(出現回数が多い)文字列について、文字列の精査を行っている。

40

【0102】

そして、再登録部102iは、判定で確定した文字列に関する確定データ(R-R2)に含まれる、候補タンパク質名をタンパク質名として対象種別タンパク質データベース106b(マスターデータ)(D2)に登録、および/または、候補製造者名を製造者名として製造会社データベース106c(マスターデータ)(D3)に登録する(ステップR-P5)。

【0103】

50

そして、照合部 102c は、仮登録ビッグデータデータベース 106e (D5) に記憶された Ant i 文字列群データ (D5-2) に基づく Ant i 文字列群に対する再検索を行うことで、対象種別タンパク質データベース 106b (D2) に記憶されたタンパク質名、および、製造会社データベース 106c (D3) に記憶された製造会社名を候補データ群 (R-R3) として抽出する (ステップ R-P3)。すなわち、本実施の形態において、Ant i 文字列群データ (D5-2) は、論文番号と紐付を行っているため、全論文を検索する必要がなく、対象文字列が存在するデータ (論文) のみと比較を行うだけで済むことになる。

【0104】

そして、抗体情報登録部 102d は、照合部 102c により抽出された候補データ群 (R-R3) と、照合対象の論文に関する論文情報と、を対応付けた抗体情報を抗体情報データベース 106d (D4) に登録する (ステップ R-P4)。

10

【0105】

一方、再登録部 102i は、ステップ R-P1 にて仮登録ビッグデータデータベース 106e (D5) に記憶された仮登録データ群 (M-R6) に対応する仮登録文字列情報データ (D5-1) に含まれる候補タンパク質名および / または候補製造者名が、所定数以上の論文で使用されていないと判定した場合、仮登録データを否定文字列 (R-R1) として仮登録ビッグデータデータベース 106e (D5) から削除するとともに、仮登録抗体情報データベース 106f (D6) に仮登録データを否定リストとして登録する (ステップ R-P2)。なお、本実施の形態においては、否定リスト内文字が登録されないようにすることで、ステップ M-P6 における仮登録データ抽出処理の精度を高めることができる。

20

【0106】

このように、本実施の形態においては、タンパク質名 (抗原) および製造会社 (製造者) の表記差異について吸収することができる。また、本実施の形態においては、データベースに登録されていないタンパク質名および製造会社 (製造者) を推測し、仮登録することができる。また、本実施の形態においては、複数の論文で仮登録を行い、出現回数が高いものを検証・判定後、データベースに登録し、再検索することでタンパク質名 (抗原) および製造会社 (製造者) を特定することで、検索精度を高めることができる。

【0107】

すなわち、本実施の形態においては、正式名称で記載されていないタンパク質名 (抗原)、および / または、データベースに登録されていない製造会社 (製造者) を推測し仮登録しておくことができる。また、本実施の形態においては、この仮登録のデータ (候補となりうる文字列) の判定を行い、特定させることでタンパク質 (抗原) および / または製造会社 (製造者) の表記名を再度データベースに登録し、再検索を行うことで、抗体情報取得の精度を高めることができる。

30

【0108】

以上のように、プロテイン名については、本来、正式な名称で書くのが正しいのですが、論文の中には、Gene Name を使っているもの、または、短縮された通称名などを使用しているものもある。また、会社名については、同じ会社名でも正式会社名、または、一般的な通称など表記の違いがあるほか、当業者による Gift など既存のデータベースに登録されていないものも存在する。

40

【0109】

また、本実施の形態の登録プロセスにおいて取得できる情報は、既存のデータベースに登録されている正しいプロテイン名で記載されているもの、および、データベースに登録されている会社名に限定される。

【0110】

そこで、本実施の形態の再登録プロセスでは、準備段階で、候補となる「文字列」を抽出している。そして、本実施の形態の再登録プロセスでは、抗体情報の可能性がある文章から、プロテイン候補となる「文字列」、および、会社名候補となる「文字列」を抽出し

50

、複数論文で使用されている文字列について信頼度を高めて登録するプロセスを導入している。それにより、本実施の形態の再登録プロセスにおいては、プロテイン名、および、会社名の表記の差異（揺れ）について対応させている。

【0111】

そして、本実施の形態においては、このプロセスにより、既存のデータベースに登録されていないタンパク質名、および、会社名を素早く発見でき、また、その確定データをデータベースに反映させることで、抗体情報取得の精度を格段に高めることができる。すなわち、本実施の形態においては、登録プロセスでデータベースから論文へのアプローチで抗体情報を取得し、再登録プロセスで論文からデータベースへのアプローチでデータ登録をし、両プロセスを繰り返すことで、抗体情報取得精度を高めていることを大きな特徴としている。

10

【0112】

以上のように、本実施の形態においては、ある抗原に対して、抗体が既に作製されているかどうかを探索している。また、本実施の形態においては、多くの抗体が市販されている場合（あるいは個人から提供される場合）、コミュニティが最も多く使っている抗体が最も信頼の高いものとし、その信頼度を抗体が使われた論文の数で評価している。そのため、本実施の形態においては、内容が一般公開されている医学・生物学系の全論文を収集し、その中の抗体情報、抗原情報、抗体製造会社（作成者）情報、および、抗体利用内容（図）などを取得し、データベース化している。そして、本実施の形態においては、データベースを検索し、同一抗原に対する抗体のうち、論文で使われている数の多いものから並べて、ランキングしている。

20

【0113】

したがって、本実施の形態においては、過去に出版され、内容が一般公開されている医学・生物系の全論文を対象に、抗体の対応抗原名、抗体利用結果（図）、および、製造会社情報などを取得し、ある抗原に対する抗体が使われた論文数が多い順に並べて、ユーザにその抗体利用結果（図）情報と共に提供することができる。

【0114】

以上で、本実施の形態における抗体情報取得装置100の処理の一例の説明を終える。

【0115】

[他の実施の形態]

さて、これまで本発明の実施の形態について説明したが、本発明は、上述した実施の形態以外にも、特許請求の範囲に記載した技術的思想の範囲内において種々の異なる実施の形態にて実施されてよいものである。

30

【0116】

例えば、抗体情報取得装置100がスタンドアローンの形態で処理を行う場合を一例に説明したが、抗体情報取得装置100は、クライアント端末（抗体情報取得装置100とは別筐体である）からの要求に応じて処理を行い、その処理結果を当該クライアント端末に返却するようにしてもよい。

【0117】

また、実施の形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行われるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。

40

【0118】

このほか、上記文献中や図面中で示した処理手順、制御手順、具体的名称、各処理の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

【0119】

また、抗体情報取得装置100に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。

【0120】

50

例えば、抗体情報取得装置100の各装置が備える処理機能、特に制御部102にて行われる各処理機能については、その全部または任意の一部を、CPU(Central Processing Unit)および当該CPUにて解釈実行されるプログラムにて実現してもよく、また、ワイヤードロジックによるハードウェアとして実現してもよい。尚、プログラムは、後述する、コンピュータに本発明に係る方法を実行させるためのプログラム化された命令を含む、一時的でないコンピュータ読み取り可能な記録媒体に記録されており、必要に応じて抗体情報取得装置100に機械的に読み取られる。すなわち、ROMまたはHDD(Hard Disk Drive)などの記憶部106などには、OS(Operating System)と協働してCPUに命令を与え、各種処理を行うためのコンピュータプログラムが記録されている。このコンピュータプログラムは、RAMにロードされることによって実行され、CPUと協働して制御部を構成する。

10

【0121】

また、このコンピュータプログラムは、抗体情報取得装置100に対して任意のネットワークを介して接続されたアプリケーションプログラムサーバに記憶されていてもよく、必要に応じてその全部または一部をダウンロードすることも可能である。

【0122】

また、本発明に係るプログラムを、コンピュータ読み取り可能な記録媒体に格納してもよく、また、プログラム製品として構成することもできる。ここで、この「記録媒体」とは、メモリーカード、USBメモリ、SDカード、フレキシブルディスク、光磁気ディスク、ROM、EPROM、EEPROM、CD-ROM、MO、DVD、および、Blu-ray(登録商標) Disc等の任意の「可搬用の物理媒体」を含むものとする。

20

【0123】

また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、OS(Operating System)に代表される別個のプログラムと協働してその機能を達成するものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取るための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。

【0124】

記憶部106に格納される各種のデータベース等(対象種別論文データベース106a、対象種別タンパク質データベース106b、製造者データベース106c、抗体情報データベース106d、仮登録データベース106e、および、仮登録抗体情報データベース106f)は、RAM、ROM等のメモリ装置、ハードディスク等の固定ディスク装置、フレキシブルディスク、および、光ディスク等のストレージ手段であり、各種処理やウェブサイト提供に用いる各種のプログラム、テーブル、データベース、および、ウェブページ用ファイル等を格納する。

30

【0125】

また、抗体情報取得装置100は、既知の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、携帯電話、スマートフォン、PHS、およびPDA等の携帯端末装置、ならびに、ワークステーション等の抗体情報取得装置として構成してもよく、また、該抗体情報取得装置に任意の周辺装置を接続して構成してもよい。また、抗体情報取得装置100は、該抗体情報取得装置に本発明の方法を実現させるソフトウェア(プログラム、データ等を含む)を実装することにより実現してもよい。

40

【0126】

更に、装置の分散・統合の具体的な形態は図示するものに限られず、その全部または一部を、各種の付加等に応じて、または、機能負荷に応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。すなわち、上述した実施の形態を任意に組み合わせる実施してもよく、実施の形態を選択的に実施してもよい。

【産業上の利用可能性】

50

【 0 1 2 7 】

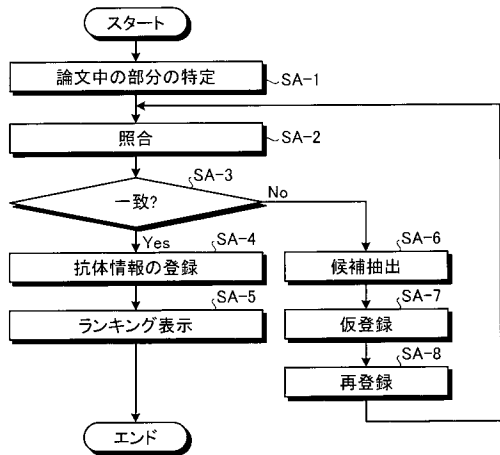
以上詳述に説明したように、本発明によれば、医学・生物学系分野の当業者が、それぞれの作業に必要な抗体が既にあるか、あるとしたらどの抗体が最も自分の作業に合致したものを簡単に知る事が出来る抗体情報取得装置、抗体情報取得方法、および、プログラムを提供することができるので、特に医療、製薬、創薬、および、生物学研究等の様々な分野において極めて有用である。

【 符号の説明 】

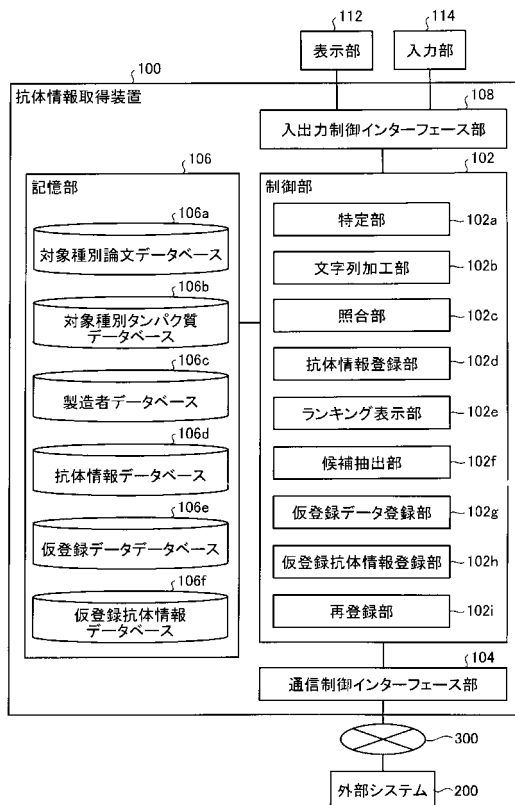
【 0 1 2 8 】

1 0 0	抗体情報取得装置	
1 0 2	制御部	10
1 0 2 a	特定部	
1 0 2 b	文字列加工部	
1 0 2 c	照合部	
1 0 2 d	抗体情報登録部	
1 0 2 e	ランキング表示部	
1 0 2 f	候補抽出部	
1 0 2 g	仮登録データ登録部	
1 0 2 h	仮登録抗体情報登録部	
1 0 2 i	再登録部	
1 0 4	通信制御インターフェース部	20
1 0 6	記憶部	
1 0 6 a	対象種別論文データベース	
1 0 6 b	対象種別タンパク質データベース	
1 0 6 c	製造者データベース	
1 0 6 d	抗体情報データベース	
1 0 6 e	仮登録データデータベース	
1 0 6 f	仮登録抗体情報データベース	
1 0 8	入出力制御インターフェース部	
1 1 2	表示部	
1 1 4	入力部	30

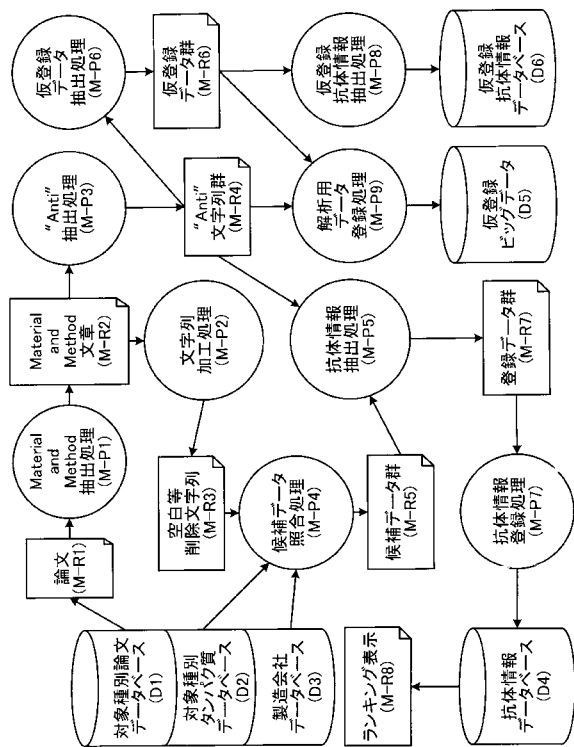
【 図 1 】



【 図 2 】



【 図 3 】



【 図 4 】

Antibodies Information Searching System Ver1.0:update:2012/08/11

Relationship of antibodies, articles, and corporations

Index
 -SearchUnitProtein
 -SearchUnitAntibody
 -ArticleData
 -ProteinData
 -ReportSearchSystem
 -PDFAnalyse
 link
 1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12

Search Word: what's new!!

2012/08/11 ver7.0 Update New Database
 2012/07/31 ver6.1 Delete Duplication Data
 2012/07/15 ver6.0 Import corporation Data
 2012/06/12 ver5.0 Update New version
 2012/05/25 ver5.8 Create the PDF module
 2012/04/18 ver5.7 Changed location
 2012/04/09 ver5.6 Update search engine
 2012/03/18 ver5.5 Check the link file
 2012/03/16 ver5.3 Update main module
 2012/02/06 ver5.2 Update New Database

Accession	Title	Abstract	Link
21088561	Anti-CD133 antibody	CD133 is a marker for stem cells in a number of tissues. It is expressed by hematopoietic stem cells, endothelial cells, and epithelial cells. We have generated a monoclonal antibody against CD133, which recognizes CD133 in various tissues. This antibody is useful for the detection and isolation of CD133-positive cells.	http://www.ncbi.nlm.nih.gov/
21462819	Anti-CD133 antibody	CD133 is a marker for stem cells in a number of tissues. It is expressed by hematopoietic stem cells, endothelial cells, and epithelial cells. We have generated a monoclonal antibody against CD133, which recognizes CD133 in various tissues. This antibody is useful for the detection and isolation of CD133-positive cells.	http://www.ncbi.nlm.nih.gov/
11238251	Anti-CD133 antibody	CD133 is a marker for stem cells in a number of tissues. It is expressed by hematopoietic stem cells, endothelial cells, and epithelial cells. We have generated a monoclonal antibody against CD133, which recognizes CD133 in various tissues. This antibody is useful for the detection and isolation of CD133-positive cells.	http://www.ncbi.nlm.nih.gov/
12289116	Anti-CD133 antibody	CD133 is a marker for stem cells in a number of tissues. It is expressed by hematopoietic stem cells, endothelial cells, and epithelial cells. We have generated a monoclonal antibody against CD133, which recognizes CD133 in various tissues. This antibody is useful for the detection and isolation of CD133-positive cells.	http://www.ncbi.nlm.nih.gov/

【 図 5 】

Rank	Concentration	Number of Article
1	3μg	20
2	3μg	12
3	3μg	9
3	3μg	9
5	3μg	5
6	3μg	3
7	3μg	1
7	3μg	1
7	3μg	1
7	3μg	1
7	3μg	1
7	3μg	1
7	3μg	1

検索結果
65件中1~20件まで表示
2016.02.20日

Ranking		Title		Abstract		Images	
Rank	Concentration	Free-Text	Local-PDF	Free-Text	Local-PDF	Free-Text	Local-PDF
1	3μg	20141103	PMC3013389	21471008	PMC3013389	21608391	PMC3017248
2	3μg	2100391	PMC3018748	2100391	PMC3018748		

Candidate Protein List

Images

【 図 6 】

