(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116783658 A
(43)申请公布日 2023.09.19

(21)申请号 202180089273.6
(22)申请日 2021.12.07
(30)优先权数据
2021-001610 2021.01.07 JP
(85)PCT国际申请进入国家阶段日
2023.07.04
(86)PCT国际申请的申请数据
PCT/JP2021/044992 2021.12.07
(87)PCT国际申请的公布数据
WO2022/149394 JA 2022.07.14
(71)申请人 富士胶片株式会社
地址 日本国东京都
(72)发明人 鑓水大和 疋田泰士
(74)专利代理机构 中科专利商标代理有限责任公司 11021
专利代理师 高颖
(51)Int.Cl.
G16C 20/40(2006.01)

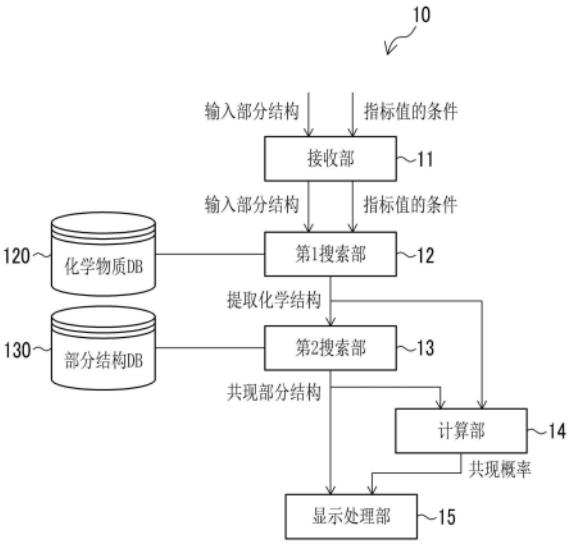
权利要求书2页 说明书6页 附图7页

(54)发明名称

信息处理装置、信息处理方法及信息处理程序

(57)摘要

一种信息处理装置具备至少一个处理器。处理器进行如下处理:接收表示化学物质的部分结构的部分结构数据及与表示化学物质的性能的指标值相关的条件的输入,针对多个已知化学物质中的每一个,从记录有表示该化学物质结构的结构数据的数据库中提取已知化学物质,所述已知化学物质包含由部分结构数据表示的部分结构即输入部分结构且满足所述条件,提取所提取的已知化学物质的结构中包含的除输入部分结构以外的部分结构作为共现部分结构,显示所提取的所述共现部分结构。



CN 116783658 A

1.一种信息处理装置,其具备至少一个处理器,所述信息处理装置中,

所述处理器进行如下处理:

接收表示化学物质的部分结构的部分结构数据及与表示化学物质的性能的指标值相关的条件的输入,

从针对多个已知化学物质中的每一个记录有表示该化学物质结构的结构数据的数据库中,提取包含由所述部分结构数据表示的部分结构即输入部分结构且满足所述条件的已知化学物质,

提取所提取的已知化学物质的结构中包含的除所述输入部分结构以外的部分结构作为共现部分结构,

显示所提取的所述共现部分结构。

2.根据权利要求1所述的信息处理装置,其中,

所述处理器进行示出所提取的共现部分结构与所述输入部分结构的连接关系并显示所提取的共现部分结构的处理。

3.根据权利要求1或2所述的信息处理装置,其中,

所述处理器进行如下处理:

针对所提取的所述共现部分结构,计算与所述输入部分结构一起出现的概率即共现概率,

在提取出结构相互不同的多个所述共现部分结构的情况下,按照所述共现概率的顺序排列显示所提取的多个所述共现部分结构。

4.根据权利要求3所述的信息处理装置,其中,

所述处理器进行将所提取的所述共现部分结构与针对该共现部分结构计算出的所述共现概率一起显示的处理。

5.一种信息处理方法,其中,由信息处理装置所具备的处理器执行如下处理:

接收表示化学物质的部分结构的部分结构数据及与表示化学物质的性能的指标值相关的条件的输入,

从针对多个已知化学物质中的每一个记录有表示该化学物质结构的结构数据的数据库中,提取包含由所述部分结构数据表示的部分结构即输入部分结构且满足所述条件的已知化学物质,

提取所提取的已知化学物质的结构中包含的除所述输入部分结构以外的部分结构作为共现部分结构,

显示所提取的所述共现部分结构。

6.一种信息处理程序,其用于使信息处理装置所具备的处理器执行如下处理:

接收表示化学物质的部分结构的部分结构数据及与表示化学物质的性能的指标值相关的条件的输入,

从针对多个已知化学物质中的每一个记录有表示该化学物质结构的结构数据的数据库中,提取包含由所述部分结构数据表示的部分结构即输入部分结构且满足所述条件的已知化学物质,

提取所提取的已知化学物质的结构中包含的除所述输入部分结构以外的部分结构作为共现部分结构,

显示所提取的所述共现部分结构。

信息处理装置、信息处理方法及信息处理程序

技术领域

[0001] 发明的技术涉及一种信息处理装置、信息处理方法及信息处理程序。

背景技术

[0002] 作为与化学物质的设计支持相关的技术，已知有以下技术。例如，日本特开2019-40422号公报记载了一种化合物设计装置，其具备：部分结构选定单元，选定化合物的任意部分结构；部分结构检测单元，在化合物数据中检测相对于由部分结构选定单元选定的部分结构共现的部分结构；以及共现性量化单元，将由部分结构检测单元检测出的多个部分结构作为对象，对共现的程度进行量化。

发明内容

[0003] 发明要解决的技术课题

[0004] 构成化学物质的一部分的部分结构有时会对该化学物质的性能产生很大影响。因此，以制造为目的来设计发挥期望的性能的化学物质的结构时，需要掌握该化学物质应该具备的部分结构是什么。但是，化学物质的性能不仅根据部分结构的存在发生变化，而且还根据该部分结构与该化学物质中包含的其他结构之间的位置关系和组合等而发生变化。由于影响化学物质的性能的因素很复杂，因此在完全掌握这样的因素的基础上进行化学物质的结构设计并不容易。

[0005] 发明的技术是鉴于上述问题而完成的，其目的在于支持发挥期望的性能的化学物质的结构设计。

[0006] 用于解决技术课题的手段

[0007] 发明的技术所涉及的信息处理装置具备至少一个处理器，该信息处理装置中，处理器进行如下处理：接收表示化学物质的部分结构的部分结构数据及与表示化学物质的性能的指标值相关的条件的输入，从针对多个已知化学物质中的每一个记录有表示该化学物质结构的结构数据的数据库中，提取已知化学物质，该已知化学物质包含由部分结构数据表示的部分结构即输入部分结构且满足条件，提取所提取的已知化学物质的结构中包含的除输入部分结构以外的部分结构作为共现部分结构，显示所提取的共现部分结构。

[0008] 处理器可以进行一边示出与输入部分结构的连接关系一边显示所提取的共现部分结构的处理。针对所提取的共现部分结构，处理器可以计算与输入部分结构一起出现的概率即共现概率。在提取出结构相互不同的多个共现部分结构的情况下，也可以进行按照共现概率的顺序排列显示所提取的多个共现部分结构的处理。处理器也可以进行将所提取的共现部分结构与针对该共现部分结构计算出的共现概率一起显示的处理。

[0009] 发明的技术所涉及的信息处理方法是由信息处理装置所具备的处理器执行如下处理的方法：接收表示化学物质的部分结构的部分结构数据及与表示化学物质的性能的指标值相关的条件的输入，从针对多个已知化学物质中的每一个记录有表示该化学物质结构的结构数据的数据库中，提取已知化学物质，该已知化学物质包含由部分结构数据表示的

部分结构即输入部分结构且满足条件，提取所提取的已知化学物质的结构中包含的除输入部分结构以外的部分结构作为共现部分结构，显示所提取的共现部分结构。

[0010] 发明的技术所涉及的信息处理程序是用于使信息处理装置所具备的处理器执行如下处理的程序：接收表示化学物质的部分结构的部分结构数据及与表示化学物质的性能的指标值相关的条件的输入，从针对多个已知化学物质中的每一个记录有表示该化学物质结构的结构数据的数据库中，提取已知化学物质，该已知化学物质包含由部分结构数据表示的部分结构即输入部分结构且满足条件，提取所提取的已知化学物质的结构中包含的除输入部分结构以外的部分结构作为共现部分结构，显示所提取的共现部分结构。

[0011] 发明效果

[0012] 根据发明的技术，能够支持发挥期望的性能的化学物质的结构设计。

附图说明

[0013] 图1是表示发明技术的实施方式所涉及的信息处理装置的硬件结构的一例的图。

[0014] 图2是表示以图表形式表达的化学物质的结构数据的一例的图。

[0015] 图3是表示发明技术的实施方式所涉及的化学物质数据库的一例的图。

[0016] 图4是表示发明技术的实施方式所涉及的部分结构数据库的一例的图。

[0017] 图5是表示发明技术的实施方式所涉及的信息处理装置的功能结构的一例的功能框图。

[0018] 图6是表示发明技术的实施方式所涉及的输入部分结构的一例的图。

[0019] 图7是表示发明技术的实施方式所涉及的共现部分结构的显示方式的一例的图。

[0020] 图8是表示发明技术的实施方式所涉及的显示处理的流程的一例的流程图。

[0021] 图9是表示羧酸的结构的图。

[0022] 图10是表示马来酸的结构的图。

[0023] 图11是表示输入部分结构的一例即羧基的结构的图。

具体实施方式

[0024] 以下，参考附图对发明技术实施方式的一例进行说明。另外，对各附图中相同或等价的构成要素和部分标注相同的参考符号，并适当省略重复说明。

[0025] 图1是表示发明技术的实施方式所涉及的信息处理装置10的硬件结构的一例的图。信息处理装置10包括CPU(Central Processing Unit:中央处理器)101、作为临时存储区域的存储器102以及存储部103。并且，信息处理装置10包括液晶显示器等显示部104、包含键盘和鼠标等输入设备而构成的输入部105以及与网络连接的网络I/F(InterFace:接口)106。CPU101、存储器102、存储部103、显示部104、输入部105及网络I/F106分别与总线108连接。

[0026] 存储部103由HDD(Hard Disk Drive:硬盘驱动器)、SSD(Solid State Drive:固态驱动器)或闪存等非易失性存储介质来实现。存储部103中存储有信息处理程序110、化学物质数据库120及部分结构数据库130。CPU101从存储部103读出信息处理程序110后，在存储器102中展开并执行。另外，作为信息处理装置10的例子，可以举出服务器计算机等。CPU101是发明技术中的处理器的一例。

[0027] 信息处理装置10用于化学物质的结构设计，并具有作为分子设计编辑器的功能。表示本实施方式所涉及的信息处理装置10所处理的化学物质的结构的结构数据由图表形式来表达。图2是表示以图表形式表达的化学物质的结构数据200的一例的图。以图表形式表达的结构数据200中，节点201来表达构成该化学物质的原子，由边缘202来表达原子间的结合。另外，信息处理装置10所处理的结构数据的形式并不只限于图表形式，例如也可以是DNA(DeoxyriboNucleic Acid:脱氧核糖核酸)碱基序列这样的字符串形式。

[0028] 图3是表示存储在存储部103中的化学物质数据库120的一例的图。化学物质数据库120针对多个已知化学物质中的每一个，记录了表示该化学物质的整体结构的结构数据。结构数据以图表形式来表达。表示该化学物质的性能的至少一个指标值与结构数据中的每一个建立了关联。作为指标值的例子，可以举出沸点、熔点、玻璃化转变温度、分配系数、密度、粘度、热膨胀率、分子量等。指标值例如可以是通过过去的实验得到的实测值或标称值。

[0029] 图4是表示存储在存储部103中的部分结构数据库130的一例的图。部分结构数据库130针对多个已知部分结构中的每一个，记录了表示该部分结构的部分结构数据。部分结构是指构成化学物质的一部分的结构。作为部分结构，例如可以举出羧基、醛基、羟基等官能团的结构等。部分结构的结构数据以图表形式来表达。表示该部分结构的性能的至少一个指标值与部分结构的结构数据中的每一个建立了关联。作为指标值的例子，可以举出表示有无致癌性、有无毒性、以及在水中容易溶解的程度等。指标值例如可以是通过过去的实验得到的实测值或标称值。

[0030] 图5是表示信息处理装置10的功能结构的一例的功能框图。信息处理装置10包括接收部11、第1搜索部12、第2搜索部13、计算部14及显示处理部15而构成。通过由CPU101执行信息处理程序110，信息处理装置10作为接收部11、第1搜索部12、第2搜索部13、计算部14及显示处理部15发挥作用。

[0031] 使用信息处理装置10进行化学物质的结构设计的用户将可以包含设计对象的化学物质的部分结构输入到信息处理装置10中。以下，将输入到信息处理装置10中的部分结构称为输入部分结构。可以通过操作输入部105来进行输入部分结构向信息处理装置10中的输入。接收部11接收表示由用户输入的输入部分结构的部分结构数据。图6是表示输入部分结构300的一例的图。

[0032] 并且，用户根据设计对象的化学物质应该具备的性能来制定与特定指标值相关的条件，并将所制定的条件输入到信息处理装置10中。例如，在将沸点设计为150℃以下的化学物质的结构时，作为与指标值相关的条件，输入表示“沸点150℃以下”的信息。可以通过操作输入部105来进行与指标值相关的条件向信息处理装置10中的输入。接收部11接收由用户输入的与特定指标值相关的条件。与特定指标值相关的条件可以是两个以上。接收部11将所接收的输入部分结构和与特定指标值相关的条件提供给第1搜索部12。

[0033] 第1搜索部12从化学物质数据库120中搜索并提取包含输入部分结构且满足与指标值相关的条件的化学物质。即，第1搜索部12从化学物质数据库120中包含的多个已知化学物质中提取包含输入部分结构且能够发挥所要求的性能的化学物质。当化学物质数据库120中存在多个包含输入部分结构且满足与指标值相关的条件的化学物质时，第1搜索部12提取相应的全部化学物质。

[0034] 第1搜索部12将表示所提取的化学物质的结构的结构数据提供给第2搜索部13和

计算部14。以下，将由第1搜索部12提取的化学物质的结构称为提取化学结构。

[0035] 第2搜索部13提取提取化学结构中包含的除输入部分结构以外的部分结构作为共现部分结构。即，共现部分结构是在某种化学物质中与输入部分结构一同存在的部分结构。第2搜索部13通过参考部分结构数据库130来提取共现部分结构。即，当在提取化学结构中发现除输入部分结构以外的部分结构、并且与记录在部分结构数据库130中的部分结构一致的部分结构时，第2搜索部13提取该部分结构作为共现部分结构。当在提取化学结构中发现多个与上述相应的部分结构时，第2搜索部13提取相应的全部部分结构作为共现部分结构。第2搜索部13将表示共现部分结构的部分结构数据提供给显示处理部15和计算部14。

[0036] 针对共现部分结构，计算部14计算与输入部分结构一起出现的概率即共现概率。在提取出结构相互不同的多个共现部分结构的情况下，计算部14针对多个共现部分结构分别计算共现概率。共现概率例如也可以由下述式(1)表示。式(1)中，P为共现概率，A为由第2搜索部13提取的某一种共现部分结构的提取数，B为由第1搜索部12提取的化学物质的提取数。另外，B可以是记录在化学物质数据库120中的化学物质的总数。认为共现概率相对较高的共现部分结构与输入部分结构一起存在，因此满足与指标值相关的条件的可能性有相对较高的倾向。

[0037] $$P=A/B\times 100[\%]\cdots(1)$$

[0038] 显示处理部15进行在显示部104上显示由第2搜索部13提取的共现部分结构的处理。图7是表示在显示部104的显示画面104A上显示的共现部分结构310的显示方式的一例的图。

[0039] 显示处理部15进行一边示出与输入部分结构300的连接关系一边显示所提取的共现部分结构310的处理。以图表形式显示共现部分结构310和输入部分结构300。例如，在共现部分结构310与输入部分结构300直接连接的情况下，连接部位由边缘表示。显示处理部15也可以进行可识别地显示共现部分结构310和输入部分结构300的处理。例如，可以进行以不同颜色显示构成共现部分结构310的节点和构成输入部分结构300的节点的处理。另外，显示处理部15也可以进行仅显示共现部分结构310和输入部分结构300中的共现部分结构310的处理。并且，显示处理部15也可以进行显示包含输入部分结构300和共现部分结构310双方的整体结构、即由第1搜索部12提取的化学物质的整体结构的处理。

[0040] 在提取出结构相互不同的多个共现部分结构的情况下，显示处理部15进行按照由计算部14计算出的共现概率的顺序排列显示所提取的多个共现部分结构的处理。例如，如图7所示，显示处理部15也可以进行按照共现概率高的顺序从显示画面104A的左侧向右侧依次排列显示所提取的多个共现部分结构310的处理。并且，也可以按照共现概率高的顺序从显示画面104A的上侧向下侧依次排列显示。显示处理部15也可以进行将共现部分结构与共现概率一起显示的处理。

[0041] 图8是表示通过由CPU101执行信息处理程序110来实施的显示处理的流程的一例的流程图。

[0042] 在步骤S1中，接收部11接收表示用户通过操作输入部105来输入的输入部分结构的部分结构数据。在步骤S2中，接收部11接收用户通过操作输入部105来输入的与特定指标值相关的条件。与特定指标值相关的条件根据设计对象的化学物质应该具备的性能来制定。

[0043] 在步骤S3中，第1搜索部12从化学物质数据库120搜索并提取已知化学物质，该已知化学物质包含在步骤S1中接收到的输入部分结构且满足在步骤S2中接收到的与指标值相关的条件。

[0044] 在步骤S4中，第2搜索部13通过参考部分结构数据库130，提取在步骤S3中提取的已知化学物质的结构(即，提取化学结构)中包含的除输入部分结构以外的部分结构作为共现部分结构。

[0045] 在步骤S5中，针对在步骤S4中提取的共现部分结构，计算部14计算与输入部分结构一起出现的概率即共现概率。计算部14例如根据上述式(1)计算共现概率。

[0046] 在步骤S6中，显示处理部15进行在显示部104上显示在步骤S4中提取的共现部分结构的处理。具体而言，显示处理部15进行一边示出与输入部分结构的连接关系一边显示共现部分结构的处理。在提取出结构相互不同的多个共现部分结构的情况下，显示处理部15进行按照共现概率的顺序排列显示所提取的多个共现部分结构的处理。

[0047] 构成化学物质的一部分的部分结构有时会对该化学物质的性能产生很大影响。因此，在设计发挥期望的性能的化学物质时，需要掌握该化学物质应该具备的部分结构是什么。

[0048] 在此，图9所示的羧酸一般通过氢键形成二聚体。如羧酸这样形成氢键的化学物质，其沸点比不形成氢键的分子量相同程度的化学物质高。另一方面，图10所示的马来酸是羧酸之一，但由于两个羧基相邻存在，因此不与其他分子形成氢键。因此，即使马来酸是羧酸，其沸点也比较低。

[0049] 这样，化学物质的性能(例如沸点)不仅根据部分结构的存在发生变化，而且还根据该部分结构与该化学物质中包含的其他结构之间的位置关系和组合等而发生变化。由于影响化学物质的性能的因素很复杂，因此在完全掌握这样的因素的基础上进行化学物质的结构设计并不容易。

[0050] 根据发明技术的实施方式所涉及的信息处理装置10，提取包含输入部分结构且满足与特定指标值相关的条件(即，与性能相关的要求)的已知化学物质，并且提取并显示所提取的化学物质的结构中包含的除输入部分结构以外的部分结构作为共现部分结构。

[0051] 例如，假设向信息处理装置10中输入图11所示的羧基作为输入部分结构，并且输入“沸点150℃以下”作为与特定指标值相关的条件的情况。在这种情况下，提取图10所示的马来酸作为包含输入部分结构且满足条件的已知化学物质。而且，提取并显示马来酸中包含的两个羧基中的与输入部分结构不同的羧基作为共现部分结构。

[0052] 这样，根据信息处理装置10，在与输入部分结构一起存在的情况下，有可能满足与指标值相关的条件的部分结构作为共现部分结构提示给用户。用户可以采用所提示的共现部分结构作为设计对象的化学物质中包含的部分结构的候选。这样，根据发明技术的实施方式所涉及的信息处理装置10，能够支持发挥期望的性能的化学物质的结构设计。

[0053] 并且，根据信息处理装置10，由于显示从化学物质数据库120中提取的已知化学物质的结构中提取的共现部分结构，因此可以提高所显示的共现部分结构对于发挥期望的性能有用的可能性。并且，根据信息处理装置10，由于在示出与输入部分结构的连接关系的状态下显示所提取的共现部分结构，因此可以更有效地进行设计支持。并且，根据信息处理装置10，在提取出结构相互不同的多个共现部分结构的情况下，由于所提取的多个共现部分

结构以共现概率的顺序排列的状态被显示,因此可以明示满足与指标值相关的条件的可能性相对较高的共现部分结构。

[0054] 另外,本实施方式中,例示了在提取化学结构中包含的部分结构中,将与记录在部分结构数据库130中的部分结构一致的部分结构作为共现部分结构提取的情况,但并不限定于该方式。可以从提取化学结构中随机提取或者根据规定的规则选择的部分结构作为共现部分结构。

[0055] 另外,上述实施方式中,例如,作为接收部11、第1搜索部12、第2搜索部13、计算部14及显示处理部15等执行各种处理的处理部(processing unit:处理单元)的硬件结构,可以使用以下所示的各种处理器(processor)。如上所述,除了执行软件(程序)而作为各种处理部发挥作用的通用的处理器即CPU之外,上述各种处理器还包括:FPGA(Field Programmable Gate Array:现场可编程门阵列)等在制造后能够变更电路结构的处理器即可编程逻辑器件(Programmable Logic Device:PLD)、以及ASIC(Application Specific Integrated Circuit:专用集成电路)等具有为了执行特定的处理而专门设计的电路结构的处理器即专用电路等。

[0056] 一个处理部可以由这些各种处理器中的一个构成,也可以由相同或不同种类的两个以上的处理器的组合(例如,多个FPGA的组合或者CPU与FPGA的组合)构成。并且,也可以由一个处理器构成多个处理部。

[0057] 作为由一个处理器构成多个处理部的例子,首先,如客户端和服务器等计算机为代表的那样,有由一个以上的CPU和软件的组合构成一个处理器,并且该处理器作为多个处理部发挥作用的方式。其次,有以片上系统(System OnChip:SoC)等为代表的,使用由一个IC(Integrated Circuit:集成电路)芯片实现包含多个处理部的整个系统的功能的处理器的方式。这样,各种处理部作为硬件结构,使用上述各种处理器中的一个以上构成。

[0058] 而且,作为这些各种处理器的硬件结构,更具体而言,可以使用组合了半导体元件等电路元件而成的电路(circuitry)。

[0059] 并且,在上述实施方式中,说明了信息处理程序110预先存储(安装)在存储部103中的方式,但并不限于此。也可以以记录在CD-ROM(Compact Disc Read Only Memory:光盘只读存储器)、DVD-ROM(Digital Versatile Disc Read Only Memory:数字多功能光盘只读存储器)及USB(Universal Serial Bus:通用串行总线)存储器等记录介质中的方式提供信息处理程序110。并且,信息处理程序110也可以设为经由网络从外部装置下载的方式。

[0060] 另外,2021年1月7日申请的日本专利申请2021-001610号的发明,其全部内容通过参考纳入本说明书中。并且,本说明书中所记载的所有文献、专利申请以及技术标准,以与具体且个别记载了通过参考纳入个别文献、专利申请以及技术标准的情况相同程度地,通过参考纳入本说明书中。

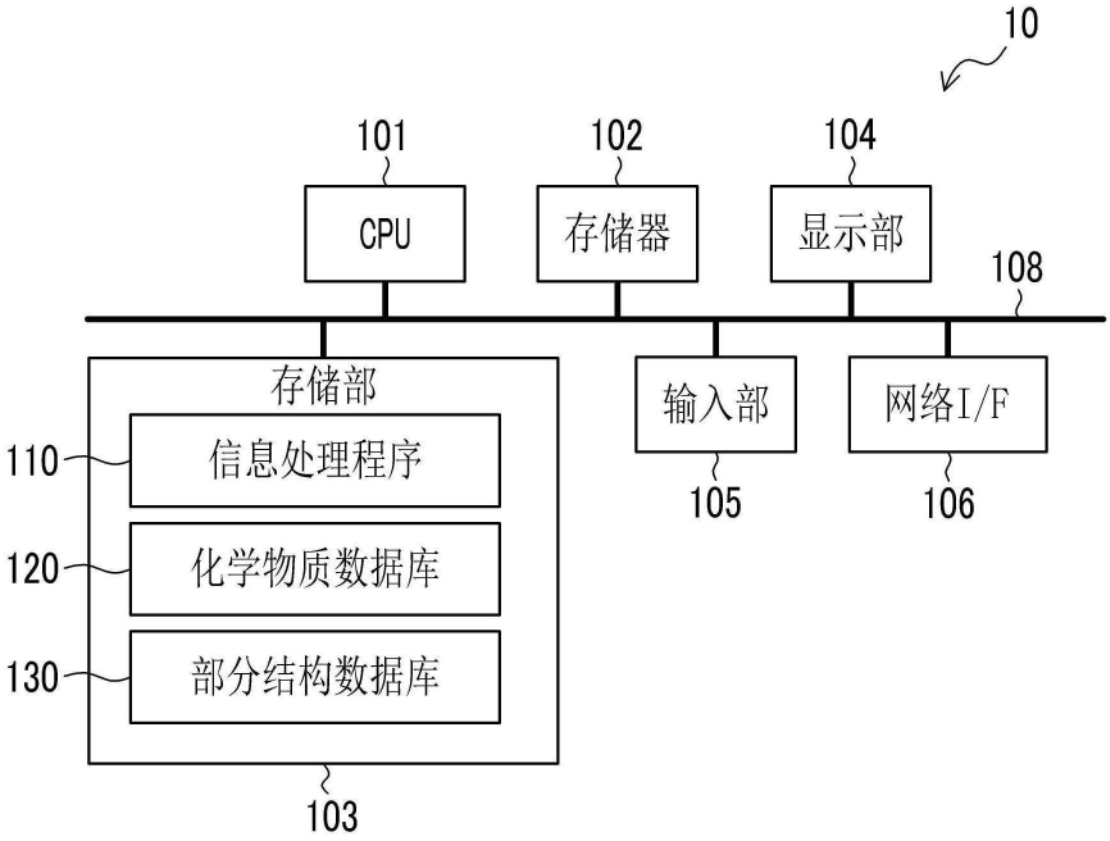

10
101
CPU
102
存储器
104
显示部
108
存储部
110
信息处理程序
120
化学物质数据库
130
部分结构数据库
103
输入部
105
网络I/F
106

图1

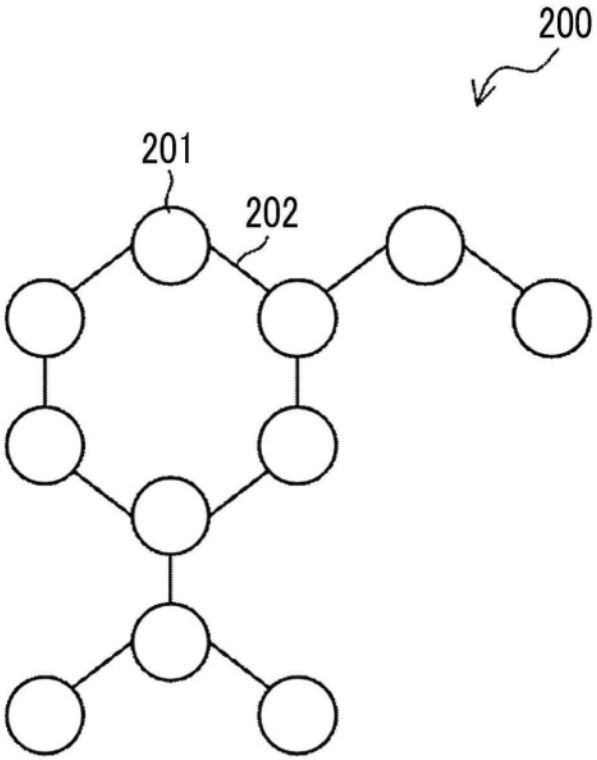
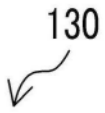


图2

120

整体结构	指标值1	指标值2	···	指标值n
图表A	10	30	···	50
图表B	15	27	···	80
⋮	⋮	⋮	⋮	⋮

图3



部分结构	指标值1	指标值2	···	指标值n
部分图表A	25	13	···	90
部分图表B	42	78	···	50
⋮	⋮	⋮	⋮	⋮

图4

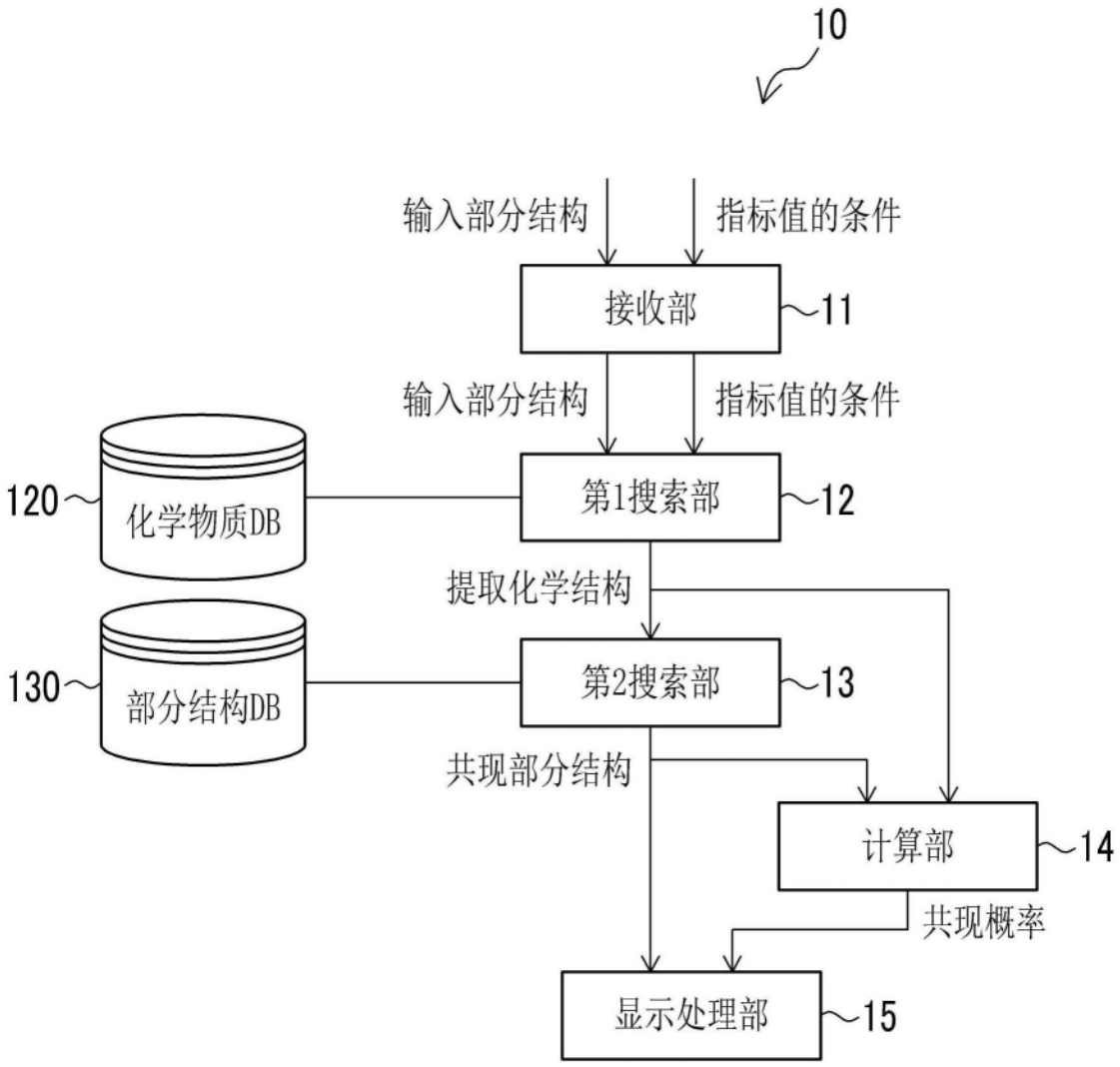

10
输入部分结构
指标值的条件
接收部
11
输入部分结构
指标值的条件
120
化学物质DB
第1搜索部
12
提取化学结构
130
部分结构DB
第2搜索部
13
共现部分结构
计算部
14
共现概率
显示处理部
15

图5

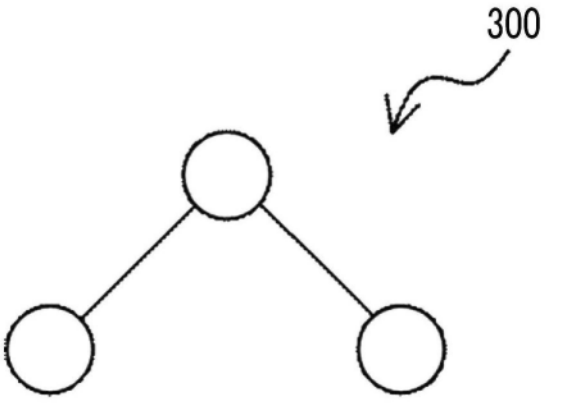

300

图6

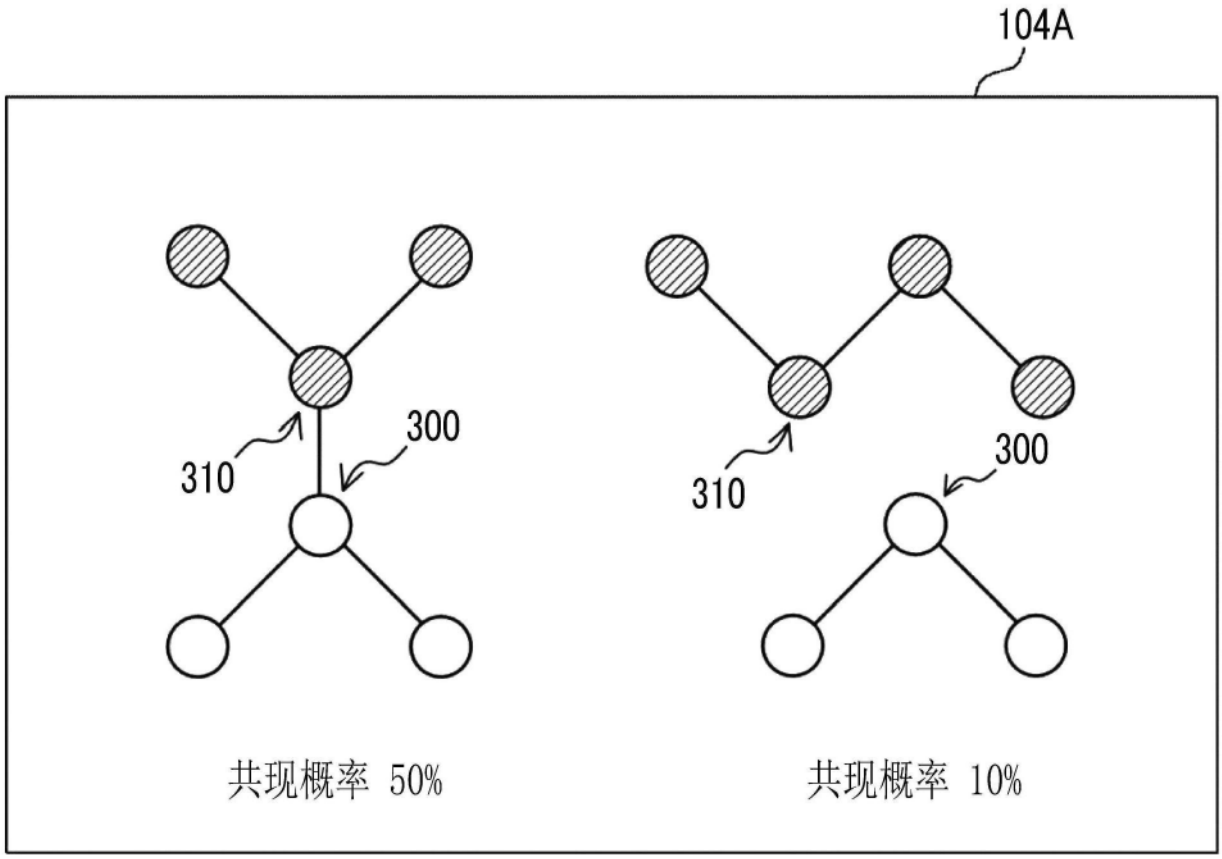


图7

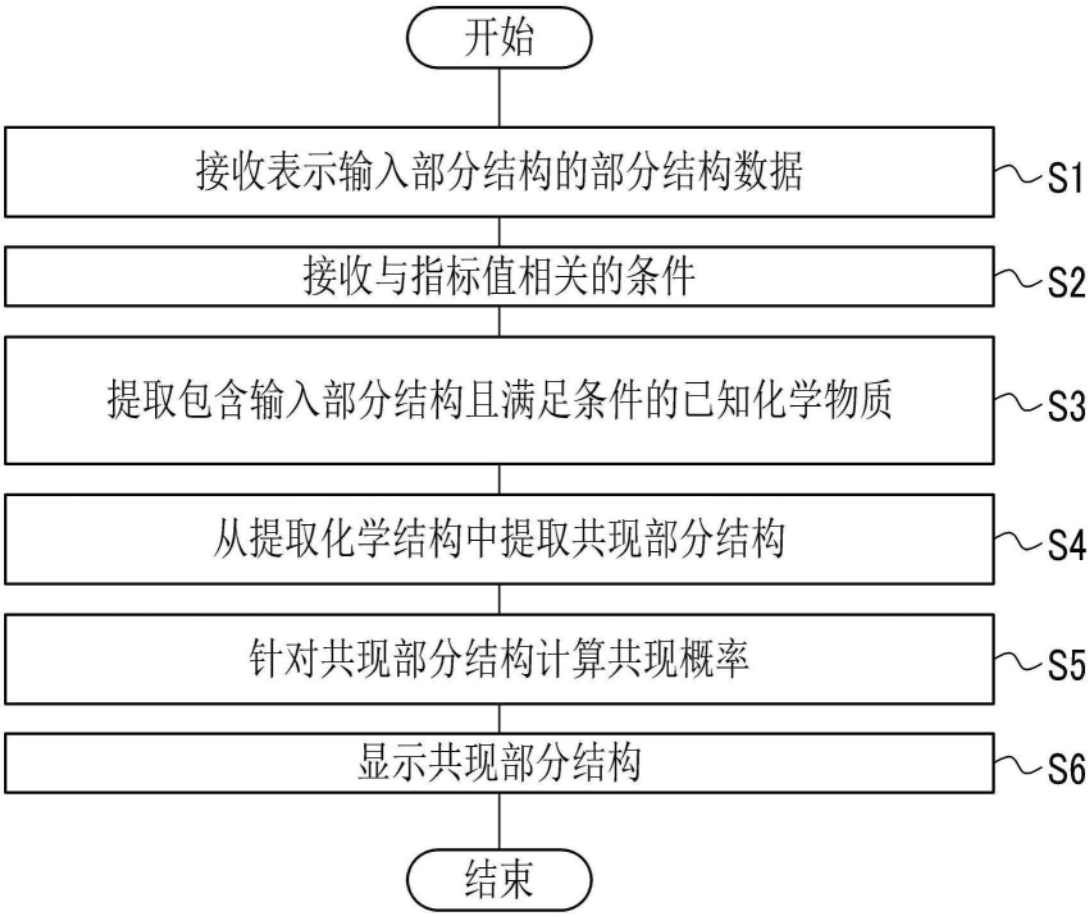

开始
接收表示输入部分结构的部分结构数据
S1
接收与指标值相关的条件
S2
提取包含输入部分结构且满足条件的已知化学物质
S3
从提取化学结构中提取共现部分结构
S4
针对共现部分结构计算共现概率
S5
显示共现部分结构
S6
结束

图8

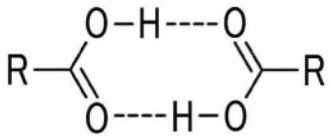

O-H----O
R
R
O----H-O

图9

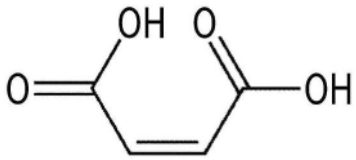

OH
O
O
OH

图10

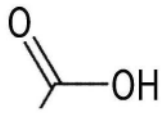


图11