(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103699558 A
(43)申请公布日 2014.04.02

(21)申请号 201310445223.4

(22)申请日 2013.09.26

(30)优先权数据
13/628,967 2012.09.27 US
13/689,157 2012.11.29 US

(71)申请人 国际商业机器公司
地址 美国纽约

(72)发明人 J·R·克泽罗斯基
C·A·皮茨克维尔 J·M·维本
周如洪

(74)专利代理机构 北京市中咨律师事务所 11247
代理人 于静 张亚非

(51)Int.Cl.
G06F 17/30(2006.01)

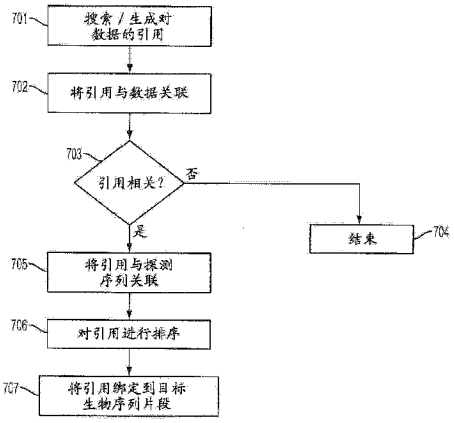
权利要求书4页 说明书11页 附图7页

(54)发明名称

用于将数据与生物序列关联的方法和装置

(57)摘要

本发明涉及一种用于将数据与生物序列关联的方法和装置。所述方法包括:在网络中标识对数据的一个或多个引用,所述数据具有大于预定阈值的相关性级别;将所述一个或多个引用与一个或多个探测序列关联,所述一个或多个探测序列对应于与所述数据有关的生物序列的片段;根据一个或多个准则对所述一个或多个探测序列进行排序;以及至少根据每个引用的排序,为所述一个或多个探测序列分配与所述目标生物序列的片段的亲和性级别。



CN 103699558 A

1. 一种用于将数据与目标生物序列关联的方法，包括：

在网络中标识对数据的一个或多个引用，所述数据具有大于预定阈值的相关性级别；

由处理器将所述一个或多个引用与一个或多个探测序列关联，所述一个或多个探测序列对应于与所述数据有关的生物序列的片段；

由所述处理器根据一个或多个准则对所述一个或多个探测序列进行排序；以及

由所述处理器至少根据每个引用的排序，为所述一个或多个探测序列分配与所述目标生物序列的片段的亲和性级别。

2. 根据权利要求 1 的方法，其中所述引用是统一资源定位符 URL。

3. 根据权利要求 1 的方法，其中标识具有大于预定阈值的相关性级别的所述一个或多个引用包括：分析所述一个或多个引用，以便检测所述一个或多个引用的关键字、词组、符号和源中的一个或多个的存在。

4. 根据权利要求 1 的方法，其中将每个引用与对应于生物序列的片段的探测序列关联包括：将每个引用关联于与所述数据有关的所述生物序列的所述片段互补的生物序列。

5. 根据权利要求 1 的方法，其中对所述一个或多个探测序列进行排序包括：确定每个探测序列和与所述引用有关的所述生物序列之间的相似性，以及

对所述一个或多个探测序列进行排序进一步包括以下操作中的至少一个：确定每个引用的源的重要性、确定每个引用的普及性，以及确定每个引用对与所述引用有关的所述生物序列的历史适用性。

6. 根据权利要求 5 的方法，其中确定每个引用和与所述引用有关的所述生物序列之间的相似性包括：确定每个引用和所述生物序列的补充物之间的匹配。

7. 根据权利要求 6 的方法，其中所述数据包括文档、分析工具和传记信息中的至少一个，所述方法还包括将所述一个或多个探测序列与所述文档、所述分析工具和所述传记信息中的至少一个相关联。

8. 根据权利要求 7 的方法，其中确定每个引用的源的重要性包括以下操作中的至少一个：确定所述文档的作者的引述数量、确定所述文档的作者所属的组织，以及确定由所述分析工具执行的分析类型，

确定每个引用的普及性包括以下操作中的至少一个：确定所述文档的引述数量和所述分析工具的使用频率，以及

确定每个引用对与所述引用有关的所述生物序列的历史适用性包括以下操作中的至少一个：确定结合所述生物序列而引述所述文档的频率，以及确定使用所述分析工具来分析所述生物序列的频率。

9. 根据权利要求 6 的方法，其中所述至少一个探测序列包括对应于所述目标生物序列的同一片段的两个或更多个探测序列，以及

为所述两个或更多个探测序列分配与所述目标生物序列的所述同一片段的亲和性级别包括竞争性地比较所述两个或更多个探测序列，以便为具有较高排序的探测序列分配高于具有较低排序的探测序列的亲和性级别。

10. 根据权利要求 1 的方法，还包括在显示器上显示所述目标生物序列的所述片段的图形表示；以及

通过基于每个探测序列与所述目标生物序列的所述片段的亲和性级别调整该探测序

列的图形表示与所述片段的所述图形表示的物理距离而在所述显示器上显示所述亲和性级别的图形表示。

11. 根据权利要求 1 的方法，还包括：

根据与所述生物序列的所述片段的所述亲和性级别，将所述两个或更多个引用中的第一引用与第一探测序列相关联；以及

根据所确定的表示所述网络中的数据的第二探测序列与所述第一探测序列的亲和性，模拟所述第二探测序列到所述第一探测序列的退火。

12. 一种用于对生物序列模拟退火的方法，包括：

标识目标生物序列；

根据对数据的多个引用具有高于阈值相关性级别的相关性级别的判定，在网络中将所述多个引用标识为相关引用；

将所述相关引用与所述目标生物序列的至少一个片段关联；以及

根据预定准则对所述相关引用进行排序，以便确定所述相关引用与所述目标生物序列的所述至少一个片段的亲和性级别。

13. 根据权利要求 12 的方法，其中标识所述目标生物序列包括：在连接到所述网络的主计算机中标识用户所选择的生物序列。

14. 根据权利要求 13 的方法，其中通过在显示设备上显示所述生物序列的图形表示而选择所述生物序列。

15. 根据权利要求 12 的方法，其中将所述对数据的多个引用标识为相关引用包括：检测所述多个引用参考的所述数据的关键字、词组、符号和源中的一个或多个的存在。

16. 根据权利要求 12 的方法，其中所述多个引用是统一资源定位符 URL。

17. 根据权利要求 12 的方法，其中将所述相关引用与所述目标生物序列的至少一个片段关联包括：将所述相关引用与对应于所述目标生物序列的所述至少一个片段的探测序列关联。

18. 根据权利要求 17 的方法，其中所述探测序列包括与所述目标生物序列的所述至少一个片段的核苷酸序列互补的核苷酸序列。

19. 根据权利要求 12 的方法，其中根据所述预定准则对所述相关引用进行排序包括：根据所述相关引用和所述目标生物序列的所述至少一个片段之间的对应性对所述相关引用进行排序，以及根据以下项中的至少一个对所述相关引用进行排序：所述相关引用参考的数据的源的重要性、所述相关引用参考的所述数据的普及性，以及所述相关引用参考的所述数据对所述目标生物序列的所述至少一个片段的历史适用性。

20. 根据权利要求 19 的方法，其中所述数据包括与所述生物序列的所述片段相关的分析工具，

所述相关引用参考的所述数据源的重要性基于以下项中的至少一个：所述分析工具的源和所述分析工具执行的分析类型，

所述相关引用参考的数据的普及性基于所述分析工具的使用频率，以及

所述相关引用参考的数据的历史适用性基于使用分析工具来分析所述目标生物序列的所述片段的频率。

21. 根据权利要求 19 的方法，其中所述数据包括与所述目标生物序列的所述至少一个

片段相关的文档，

所述相关引用参考的所述数据源的重要性基于以下项中的至少一个：所述文档的作者的引述数量和所述文档的所述作者所关联的组织，

所述相关引用参考的数据的普及性基于对所述文档的引述数量，以及

所述相关引用参考的数据的历史适用性基于结合所述生物序列的所述片段引述所述文档的频率。

22. 根据权利要求 12 的方法，还包括：

在显示器上显示所述目标生物序列的所述至少一个片段的图形表示，以及通过基于所述相关引用与所述目标生物序列的所述至少一个片段的亲和性级别调整所述相关引用的图形表示与所述片段的所述图形表示的物理距离而在所述显示器上显示所述亲和性级别的图形表示。

23. 根据权利要求 22 的方法，其中所述目标生物序列的所述至少一个片段显示为螺旋线，并且根据相应相关引用的不同亲和性级别，将所述相关引用的所述图形表示定位在距所述螺旋线的不同距离处。

24. 根据权利要求 22 的方法，其中显示所述目标生物序列的所述至少一个片段的所述图形表示包括：显示所述目标生物序列的多个连续片段；根据所述多个片段中的相应片段和相应相关引用之间的关联，显示位于沿着所述多个片段的所述图形表示的各位置处的相关引用的多个图形表示；以及根据所确定的所述相应相关引用与所述目标生物序列的相应多个片段的亲和性，显示位于距所述目标生物序列的所述多个片段的所述图形表示不同距离处的相关引用的多个图形表示。

25. 一种用于对生物序列模拟退火的方法，包括：

由处理器在网络中将对数据的引用标识为与生物序列相关的相关引用；

由所述处理器将所述相关引用与所述生物序列的片段关联；以及

由所述处理器根据预定准则对所述相关引用进行排序，以便确定所述相关引用与所述生物序列的所述片段的亲和性级别。

26. 根据权利要求 25 的方法，其中所述方法包括将所述相关引用与一个或多个探测序列关联，所述一个或多个探测序列对应于所述生物序列的一个或多个相应片段；以及

主计算机被配置为竞争性地对与所述生物序列的同一片段对应的所述一个或多个探测序列进行排序，以便具有较高排序的探测序列与所述生物序列的所述片段的亲和性级别高于具有较低排序的探测序列。

27. 根据权利要求 25 的方法，其中对所述相关引用进行排序包括：根据所述一个或多个探测序列和所述生物序列的所述片段之间的对应性，对所述一个或多个探测序列进行排序，以及

对所述一个或多个探测序列进行排序进一步包括根据以下项中的至少一个对所述一个或多个探测序列进行排序：与所述一个或多个探测序列关联的数据的源的重要性、与所述一个或多个探测序列关联的所述数据的普及性，以及与所述一个或多个探测序列关联的所述数据对所述生物序列的所述片段的历史适用性。

28. 根据权利要求 25 的方法，其中所述数据包括文档、分析工具和个人的传记信息中的至少一个，

确定每个引用的源的重要性包括以下操作中的至少一个：确定所述文档的作者的引述数量、确定所述文档的作者所属的组织，以及确定所述分析工具执行的分析类型，

确定每个引用的普及性包括以下操作中的至少一个：确定所述文档的引述数量和所述分析工具的使用频率，以及

确定每个引用对所述片段的历史适用性包括以下操作中的至少一个：确定结合所述生物序列的所述片段引述所述文档的频率，以及确定使用所述分析工具来分析所述片段的频率。

29. 根据权利要求 25 的方法，其中所述引用对应于所述生物序列的同一片段，所述方法还包括：

确定所述相关引用与所述生物序列的所述片段的亲和性级别包括竞争性地比较所述相关引用，以便为具有较高排序的引用分配高于具有较低排序的引用的亲和性级别。

30. 根据权利要求 25 的方法，所述方法还包括：

显示所述生物序列的所述片段的图形表示，以及通过基于每个引用与所述生物序列的所述片段的亲和性级别调整所述引用的图形表示与所述片段的所述图形表示的物理距离而在显示器上显示所述亲和性级别的图形表示。

31. 一种装置，其用于将数据与目标生物序列关联，以便执行权利要求 1-11 中的任一权利要求的方法步骤。

32. 一种装置，其用于对生物序列模拟退火，以便执行权利要求 12-24 中的任一权利要求的方法步骤。

33. 一种装置，其用于对生物序列模拟退火，以便执行权利要求 25-30 中的任一权利要求的方法步骤。

用于将数据与生物序列关联的方法和装置

[0001] 相关申请的交叉引用

[0002] 本申请是2012年9月27日提交的第13/628,967号美国专利申请的延续，后者的公开内容在此全部引入作为参考。

技术领域

[0003] 本公开涉及数据到生物序列的模拟绑定，更具体地说，涉及标识与生物序列相关的数据、根据其重要性对数据进行排序，以及根据排序为用户提供数据。

背景技术

[0004] 生物数据(包括生物序列)的分析可能需要存储在不同计算机上的大量数据，以便执行分析。研究的生物数据可以通过程序注释，以便参考与生物数据相关的数据(例如研究出版物)。这允许研究者查看与生物数据的当前研究相关的其它数据。尽管研究文章和其它出版物对分析生物数据有用，但其它资源也有用，例如分析工具和软件程序。此外，随着时间的流逝，有关作为资源提供的生物数据的信息量在增长。当注释生物数据以便参考相关出版物时，注释也有所增加，这可能使得研究者更加难以标识与当前研究相关的重要信息。

发明内容

[0005] 示例性实施例包括一种用于将数据与目标生物序列关联的方法。所述方法包括在网络中标识对数据的一个或多个引用，所述数据具有大于预定阈值的相关性级别。所述方法包括将所述一个或多个引用与一个或多个探测序列关联，所述一个或多个探测序列对应于与所述数据有关的生物序列的片段。根据一个或多个准则对所述一个或多个探测序列进行排序，并且至少根据每个引用的排序，为所述一个或多个探测序列分配与所述目标生物序列的片段的亲和性级别。

[0006] 实施例还包括一种用于对生物序列模拟退火的方法。所述方法包括标识目标生物序列，并根据对数据的多个引用具有高于阈值相关性级别的相关性级别的判定，在网络中将所述多个引用标识为相关引用。所述方法还包括将所述相关引用与所述目标生物序列的至少一个片段关联。所述方法还包括根据预定准则对所述相关引用进行排序，以便确定所述相关引用与所述目标生物序列的所述至少一个片段的亲和性级别。

[0007] 实施例还包括一种用于对生物序列模拟退火的方法。所述方法包括：由处理器在网络中将对数据的引用标识为与生物序列相关的相关引用；由所述处理器将所述相关引用与所述生物序列的片段关联；以及由所述处理器根据预定准则对所述相关引用进行排序，以便确定所述相关引用与所述生物序列的所述片段的亲和性级别。

[0008] 通过实现本公开的实施例获得其它特性和优点。在此详细描述本公开的其它实施例和方面，并且这些实施例和方面被视为要求保护的本发明的一部分。为了更好地理解实施例(包括优点和其它特性)，请参考说明书和附图。

附图说明

[0009] 在说明书结尾处的权利要求中具体指出并明确要求保护了被视为本公开的实施例的主题。从下面结合附图的详细描述，实施例的上述和其它特性和优点将显而易见，这些附图是：

[0010] 图 1 示出根据本公开的实施例的网络系统；

[0011] 图 2 示出根据实施例的模拟退火模块；

[0012] 图 3 示出根据本公开的实施例的用户定制显示；

[0013] 图 4A 示出根据本公开的一个实施例的退火显示；

[0014] 图 4B 示出根据本公开的一个实施例的退火显示；

[0015] 图 5 示出根据本公开的另一个实施例的退火显示；

[0016] 图 6 示出根据本公开的实施例的表；

[0017] 图 7 示出根据本公开的实施例的方法的流程图；

[0018] 图 8 示出根据本公开的实施例的计算机系统；以及

[0019] 图 9 示出根据本公开的实施例的计算机程序产品。

具体实施方式

[0020] 可以注释到生物序列的大量数据可能使得研究者难以标识重要数据。本公开的实施例涉及显示数据和引用到生物序列的模拟退火，以便允许研究者快速标识重要信息。

[0021] 图 1 示出根据本公开的实施例的网络系统 100。系统 100 包括主计算机 110，其包括模拟退火模块 111、生物序列 112（也称为目标生物序列 112）和数据 113。模拟退火模块 111 被配置为分析数据和对数据的引用，以便确定哪些数据是生物序列 112 的相关数据 114，并且根据预定排序准则，确定相关数据 114 与生物序列 112 的亲和性级别。主计算机 110 可以通过以下操作显示确定的亲和性级别：显示生物序列 112，显示相关数据 114、表示相关数据 114 的符号或者表示对相关数据 114 的引用的符号，并且根据相关数据 114 针对生物序列 112 的亲和性级别调整相关数据 114（或者对应的符号）与生物序列 112 的距离。

[0022] 主计算机 110 可以连接到网络 120。网络 120 可以与一个或多个网络计算机 130 通信，网络计算机 130 在本说明书和权利要求中指连接到网络 120 以便通过网络 120 通信的计算机。网络计算机 130 可以包括数据 131，例如文档 132、分析工具 133 和传记数据（biographical data）134。尽管出于描述目的仅示出几种类型的数据，但网络计算机 130 可以存储任何类型的数据。模拟退火模块 111 可以访问数据 131，以便确定哪些数据 131 与生物序列 112 相关。模拟退火模块 111 可以根据预定排序准则对相关数据 131 进行排序，以便确定数据 131 与生物序列 112 的亲和性级别。

[0023] 主计算机 110 还可以连接到一个或多个存储器件 140，并且存储器件可以存储指向数据 142 的一个或多个引用 141 以及一个或多个生物序列 143，生物序列 143 可以是与数据比较以便确定亲和性级别的目标生物序列或生物序列。例如，存储装置 140 可以包含生物序列 143 的数据库，并且主计算机 110 的用户可以将生物序列 143 从存储装置 140 上传到主计算机 110，以便允许模拟退火模块 111 针对生物序列 143 执行数据（例如数据 113、数据 131 和数据 142）的分析，以便确定数据与生物序列 143 的亲和性级别。

[0024] 此外，网络 120 可以通过连接到因特网 160 的服务器 150，访问存储器 170 和网络

计算机 180 的一个或多个。备选地，主计算机 110 可以直接连接到因特网 160。存储器 170 和网络计算机 180 可以包括数据、引用和生物序列，它们可由主计算机 110 访问以便执行分析。

[0025] 在本公开的实施例中，生物序列 112 或 143 可以包括任何类型的生物序列，包括蛋白质的脱氧核糖核酸（DNA）、核糖核酸（RNA）、氨基酸序列，或者任何其它生物序列。数据包括文档、文件、存储的个人传记信息或组织信息、存储的出版物、有关模拟退火模块 111 或其它系统的多个查询的数据、有关针对生物序列 112 执行的先前分析的数据、分析工具、算法或程序、与生物序列 112 关联的医学治疗、有关出版物或工具的评价或评论的数据，或者任何其它数据。引用包括任何指针或地址，其指示数据位置或提供有关数据的其它信息。实例包括统一资源定位符（URL）、统一资源名称（URN）、超链接、指向数据的 javascript 指针、指向数据的 XML 指针，或者对数据的任何其它类型引用。

[0026] 下面将参考图 1 和 2 描述包括模拟退火模块 111 的主计算机 110 的操作。模块退火模块 111 可以包括生物序列标识符 206 以便标识目标生物序列 112。例如，访问主计算机 110 的用户可以在显示设备上显示生物序列 112 或者对应于生物序列 112 的数据。备选地，模拟退火模块 111 可以自动或根据预定命令来标识预定生物序列或预定生物序列的类或组，搜索主计算机 110、存储装置 140 和 170 以及网络计算机 130 和 180 的一个或多个，以便标识要成为目标生物序列 112 的生物序列。在本说明书和权利要求中，“目标生物序列”被定义为如下生物序列：用户或程序选择用于对相关数据进行排序，并且在某些实施例中进行模拟退火，如本公开的实施例中描述的那样。

[0027] 模拟退火模块 111 可以包括引用标识符 201、引用生成器 202、相关性标识符 203 和引用 / 数据关联器 204。引用标识符 201 可以搜索以下各项的存储器：设备（例如主计算机 110）、连接的存储器件 140、连接到网络 120 的设备 130，或者连接到因特网 160 的设备 170 和 180，以便获得对数据的引用，例如参考特定位置中的数据的 URL。此外，在其中数据未对应于引用，或者引用未采用模拟退火模块 111 可用格式的情况下，引用生成器 202 可以搜索以下各项的存储器：设备（例如主计算机 110）、连接的存储器件 140、连接到网络 120 的设备 130，或者连接到因特网 160 的设备 170 和 180，以便获得数据，例如文档、传记数据、与分析工具相关的数据以及任何其它数据。引用生成器 202 然后可以生成引用，例如指向数据位置的 URL。

[0028] 相关性标识符 203 分析数据（例如搜索的引用指向的数据或者引用生成器 202 标识的数据），以便确定数据是否满足相关性阈值级别。相关性阈值级别可以基于预定准则，例如数据与目标生物序列的相似性、数据源（例如提供数据的组织，例如大学、公司等）、数据作者、数据发布者，以及数据执行所执行的操作类型（例如在分析工具分析生物序列的情况下）。相关性阈值级别还可以基于访问或引用数据的频率、预定阶层（例如研究者、科学家、专业组织等）访问或引用数据的频率，或者数据与目标生物序列关联的频率。换言之，相关性阈值级别可以与目标序列相关，或者可以包括与目标序列无关的准则。相关性阈值级别可以基于数据的内容（例如作为传记信息数据的对象的个人或组织的标识），或者文档或文件的内容。此外，阈值级别相关性可以基于数据的使用，例如多久访问或引用一次数据，或者何人访问或引用数据。

[0029] 根据确定数据满足相关性阈值级别，引用 / 数据关联器 204 将引用（标识的或生成

的）与数据关联。例如，可以将引用和数据或者标识数据的信息存储在引用表 205 中。探测生成器 207 可以生成探测或探测序列，并且探测 / 引用关联器 208 可以例如通过将探测序列添加到引用表 205，将探测序列与引用关联。在一个实施例中，探测表示引用或者与引用关联的数据对应于与数据有关的特定生物序列片段的程度。与数据有关的生物序列片段可以是小于整个生物序列的部分，但在某些实例中，片段可以对应于整个生物序列。在一个实施例中，探测由序列标识，该序列与数据有关的生物序列片段的序列互补。例如，如果与数据有关的生物序列片段具有配置“GGGGAAAATT”，则探测可以对应于互补的探测序列或者“CCCCTTTTAA”。因此，主计算机 110 可以根据探测序列指示的序列，将引用和数据与有关每个引用的生物序列部分匹配。在其它实施例中，探测在空间上、数量上或者图形上标识与其有关的生物序列部分。

[0030] 排序计算器 209 可以计算每个探测序列或每个引用的排序，或者计算对应于每个引用和探测序列的数据的排序。排序可以基于一个或多个准则，并且可以对准则进行加权以便不同准则比其它准则更多地影响排序。例如，用户可以在主计算机 110 或模拟退火模块 111 中建立简档，并且用户可以指示用户更希望为哪个标记提供最大权重。在一个实施例中，加权准则类似于生物退火过程中的探测和生物序列的生物特性。

[0031] 一个准则（可以类似于生物互补要求）是确定数据或探测与目标生物序列片段的相似性、类似性或重叠。例如，文档可以显式描述目标生物序列片段，或者分析工具可以用于分析目标生物序列片段。备选地，文档可以描述相似但不相同的生物序列，这可以导致较低排序。

[0032] 另一个准则（可以类似于生物退火中的探测的结合亲和性）是确定探测、数据或引用的重要性或威望。可以根据有关不一定与目标传记信息相关的数据的信息，确定数据的重要性。例如，数据的重要性可以基于以下项中的一个或多个：文档的作者接收的引述数量、引用文档的个人或组织的标识、生成数据的大学或组织（例如进行研究的地方）、在文档中执行的分析类型、文档作者的姓名，或者可以在领域中提供有关数据威望或重要性的信息的任何其它因素。当数据与分析工具相关时，例如可以根据以下各项确定数据的重要性：工具执行的分析类型、开发分析工具的大学或组织、分析工具的创建者，或者可以提供有关分析工具威望或重要性的信息的任何其它因素。

[0033] 另一个准则（可以类似于生物退火中的探测移动性）是确定数据的普及性。该确定可以考虑访问或引用数据的频率，例如在其它出版物中引用文档的频率或者分析工具用于某领域的频率。在其中数据是传记信息（例如有关研究者的信息）的实施例中，该确定可以考虑引用研究者的频率。换言之，尽管确定数据的重要性或威望涉及不直接与数据关联的因素（例如数据源的威望），但数据的普及性可以与数据本身相关。

[0034] 另一个准则（可以类似于生物退火中的占用约束）是确定数据对目标生物序列、探测或其它探测序列的历史适用性。例如，该确定可以基于分析工具用于分析生物序列的目标片段的频率，或者在对生物序列的目标片段的引用中引用文档的频率。

[0035] 尽管提供了几个排序准则实例，但本公开的实施例包括任何排序准则，包括工具的早先使用、文档的早先引述、对数据的引述的质量、数据作者的从属、工具的易于使用、实现分析工具的成本、软件或工具引述的数量、对数据的引述或参考的日期、众包对数据重要性的投票或其它确定、有关数据的用户评价的内容等。此外，在一个实施例中，可以在排序

中引入随机元素，以便允许优化远离局部极小值。

[0036] 在一个实施例中，探测序列包括或附加有与排序准则关联的数据。当标识目标生物序列时，可以将其与所有的可用探测序列比较，并且可以根据与目标生物序列片段的相似性对可用探测序列进行排序。因此，在一个实施例中，不直接将在系统中标识的数据和引用与目标生物序列关联的数据比较。相反，将存储在其中、附加到其中或本身包括排序准则数据的探测序列与目标生物序列片段比较。

[0037] 探测序列可以由生成数据和 / 或引用的系统或用户生成，或者探测序列可以由标识目标生物序列的系统或用户生成。例如，在一个实施例中，标识目标生物序列的系统搜索网络以便获得先前生成的探测序列，并且执行排序操作。在另一个实施例中，标识目标生物序列的系统可以搜索网络，标识相关数据，并且生成对应于相关数据的探测序列。然后可以将数据、引用或生成的探测序列与预定准则和目标生物序列比较。在另一个实施例中，系统可以根据网络中新标识的数据或引用，将预先生成的探测序列的分析与新探测序列的生成相组合。

[0038] 排序计算器 209 对数据、引用或探测序列进行排序之后，退火显示生成器 210 生成排序的图形显示。图形显示可以显示以下各项的图标或其它表示 :数据、引用或探测以及目标生物序列，或者生物序列的一个或多个目标片段。退火显示生成器可以通过以下操作显示排序 :将与具有较高排序的数据关联的图标显示为更靠近目标生物序列的对应片段，而将与具有较低排序的数据关联的图标远离目标生物序列片段。

[0039] 在本公开的实施例中，模拟退火模块 111 可以分析数据以便确定相关数据，并且通过分析以下项中的一个或多个对数据进行排序 :数据中的关键字、关键字的频率、关键字组、关键字组的频率、与数据关联的元数据，或者任何其它数据内容或与数据相关的内容。

[0040] 根据本公开的实施例，可以竞争性地对数据、引用和探测序列进行排序，以便确保被确定为用户最感兴趣的数据、引用和探测序列与用户分析的目标生物序列更密切关联。

[0041] 在一个实施例中，除了将一个或多个探测序列退火到目标生物序列之外，模拟退火模块 111 还可以将一个或多个探测序列退火到一个或多个其它探测序列。例如，如果一个探测序列表示分析工具，则表示用于提高分析工具效率的程序或工具的另一个探测序列可以被模拟为退火到第一探测序列。在另一个实例中，如果第一探测序列表示软件应用，则表示包括使用软件应用的公式或分析的期刊引述的一个或多个探测序列可以被模拟为退火到第一探测序列。

[0042] 在本公开的实施例中，用户可以确定设置，或者可以生成简档，以便调整或改变数据、引用和探测序列的排序和显示。图 3 示出显示 300 或图形用户界面（GUI）300，其可以在电子显示设备（例如计算机显示器）上显示，以便允许用户设置优选权重。显示 300 包括排序准则 301a、301b、301c 和 301d。在图 3 中，排序准则包括“与目标序列的相似性”301a、“引用的重要性”301b、“引用的普及性”301c 和“对目标序列的历史适用性”301d。但是，本公开的实施例包括任何准则，包括预先设置的准则或用户生成的准则。

[0043] 图 3 进一步示出子排序图标 302a 至 302d，它们可以允许用户进一步指定排序偏好。例如，在子排序图标 302b 下，用户可以指定在确定引用的重要性时，作者或创建者从属的组织更重要，并且接收高于对作者的引述总数的权重。用户还可以设置最低准则，例如数据或引用为获得任何排序而需要的对数据或引用的最小引述数量。

[0044] 显示300可以还包括字段303a至303d，它们能够由用户修改以便调整用户所需的加权。在一个实施例中，排序计算器209使用算法以便组合用户结合包含在相关数据或引用中的信息或者与数据或引用关联的元数据选择的一个或多个准则的权重，计算数据、引用或探测序列的最终排序。

[0045] 图4A和4B示出图标402、403和404的显示400a和400b，这些图标表示根据数据、引用或探测序列的排序，退火到目标生物序列401的数据、引用或探测序列。备选地，除了退火到目标生物序列401，还可以将图标402、403和404的一个或多个退火到402、403和404中的其它图标，这些其它图标表示将探测序列退火到另一个探测序列，该另一个探测序列退火到目标生物序列401。参考图4A，具有不同视觉特性(例如不同剖面线、不同颜色、不同形状或不同图形表示)的图标402、403和404可以对应于不同类型的数据或引用。例如，具有第一类型剖面线的图标402可以对应于文档，具有第二类型剖面线的图标403可以对应于应用，具有第三类型剖面线的图标404可以对应于有关个人(例如研究者)的传记信息。可以由图标402、403和404表示的其它类型数据包括有关组织(例如公司或大学)的数据、计算机程序信息、有关研究项目的项目信息、有关可以包含相关信息的网页的信息等。

[0046] 根据与图标402、403和404表示的数据和引用关联的探测序列，将图标显示为与目标生物序列401的片段垂直(在图4A和4B中)对齐。此外，根据与图标关联的数据或引用的排序，将图标显示为与目标生物序列401的片段具有某一距离(沿着图4A和4B中的水平方向)。例如，标记为402和404的图标可以与相同或非常相似的目标序列片段关联，如图标402和404的紧密垂直对齐所指示的那样。但是，图标404可以与具有高于图标402的排序的数据关联，如沿着水平方向图标404比图标402更靠近目标生物序列所示出的那样。

[0047] 在一个实施例中，用户可以检索图标402、403和404表示的数据，或者可以通过使用光标、触摸或任何其它用户界面选择图标402、403和404，为用户提供有关数据位于哪里的信息。在一个实施例中，不同的排序特性可以更改图标402、403和404的外观。例如，表示经常引用的数据的图标可以比很少引用的数据具有更大的形状。表示通过单击图标可用的数据的图标具有的轮廓可以不同于不能通过单击图标可用的数据。表示个人的图标可以具有个人的图像。表示产品(例如分析工具或程序)的图标可以具有与工具或程序关联的图标或图像(例如商标)。

[0048] 在一个实施例中，如果目标生物序列401的片段或相邻片段包括相对大量的图标，则显示400a可以生成团405。当用户在团405上移动光标或者执行任何其它操作以便选择团405时，可以显示和选择单独的图标。

[0049] 图4B示出显示400b，其与图4A具有相同的目标生物序列401，但排序偏好不同，例如对应于不同用户选择的偏好。因此，图标402、403和404可以以不同方式布置，并且可以具有不同于图4A中的数量。在本公开的实施例中，用户可以修改用户认为重要的信息的偏好，以便个性化向用户显示的与目标生物序列401相关的信息。

[0050] 图5示出根据本公开的另一个实施例的显示500。在图5中，目标生物序列501通过例如表示核苷酸的字母显示，并且对应的探测序列502通过互补字母或者可以与目标生物序列501的核苷酸结合的核苷酸表示。图标503、504、505和506表示不同类型的数据，例如出版物、分析工具、传记信息和网页信息。根据与图标503、504、505和506表示的数据最密切相关的目标生物序列501的片段，可以(在图5中)沿着水平方向定位图标503、504、

505 和 506。图标 503、504、505 和 506 可以与目标生物序列 501 相距某一距离，该距离通过与图标 503、504、505 和 506 关联的数据或引用的排序确定。

[0051] 如图 5 中所示，图标 503 至 506 可以包含与图标表示的数据相关的信息。例如，图标可以包含数值以便指示特定源对数据的引述数量。尽管出于描述目的提供了显示实例，但实施例包括任何类型的显示，其中用户可以根据表示数据的图标与目标生物序列的距离，查看数据相对于目标生物序列的重要性。在一个实施例中，探测序列可以进一步与一个或多个其它探测序列结合。例如，用户可以在图标上移动光标或者选择图标，并且可以显示一个或多个其它链接图标，它们对应于与选定图标表示的数据相关的其它数据。在一个实施例中，探测序列可以被视为目标生物序列，并且可以参考探测序列，采用与原始生物序列相同的方式，对数据进行分析和排序。

[0052] 图 6 示出根据本公开的实施例的表 600 的一个实例。表 600 例如可以对应于图 2 的引用表 205。表 600 将引用（例如 URL、URN，或者其它地址、链接或定位符）与相关数据和探测序列关联。前面讨论了相关数据的实例，并且在图 6 中，探测序列对应于与目标生物序列片段互补的生物序列。表 600 可以还包括用于显示表示数据或引用的图标的图标信息，或者与相关数据和引用关联的任何其它信息。尽管图 2 和 6 示出用于关联数据、对数据的引用和探测序列的表，但本公开的实施例包括用于关联数据（例如数组、指针）的任何数据结构，或者个人或系统可以用于将数据与对数据的引用和探测序列关联的任何其它类型数据结构，

[0053] 图 7 示出根据本公开的实施例的方法的流程图。在方框 701，可以通过搜索计算机中的存储器、搜索存储器件、搜索连接到主机设备（其连接到网络，例如因特网）的设备等，发现对数据的引用（例如地址或指针）。此外，当未发现先前引用时，或者当需要特定类型或格式的引用时，可以生成在一个或多个设备中发现的对数据的引用。

[0054] 在方框 702，将数据与引用关联。例如，可以在表中形成条目，或者可以形成另一个数据结构以便将引用与引用指向的数据关联。在方框 703，可以确定数据是否相关。换言之，可以进行数据相关性的阈值确定。相关性阈值级别可以基于预定准则，例如数据与目标生物序列的相似性、数据源（例如提供数据的组织，例如大学、公司等）、数据作者、数据发布者，以及数据执行所执行的操作类型（例如在分析工具分析生物序列的情况下）。相关性阈值级别还可以基于访问或引用数据的频率、预定阶层（例如研究者、科学家、专业组织等）访问或引用数据的频率，或者数据与目标生物序列关联的频率。换言之，相关性阈值级别可以与目标序列相关，或者可以包括与目标序列无关的准则。相关性阈值级别可以基于数据的内容（例如基于存储的传记信息的个人或组织的标识），或者文档或文件的内容。此外，阈值级别相关性可以基于数据的使用，例如多久访问或引用一次数据，或者何人访问或引用数据。

[0055] 如果在方框 703 确定数据不满足相关性阈值级别，则针对该数据的过程在方框 704 结束。另一方面，如果确定数据充分相关，则在方框 705，可以将数据和引用与探测序列关联。探测序列可以标识与数据有关的目标生物序列的至少一部分。在一个实施例中，探测序列通过目标生物序列的对应片段的互补序列表示。例如，如果生物序列是核苷酸系列，则探测序列可以是互补的核苷酸系列。在另一个实施例中，探测序列仅是标识与数据最密切有关的目标生物序列部分的数据。在本公开的实施例中，数据可以与目标生物序列的一

个片段或多个片段有关。

[0056] 在方框 706，根据预定准则对数据进行排序，以便确定数据、引用或探测序列和目标生物序列片段之间的亲和性或结合。排序可以基于一个或多个准则，并且可以对准则进行加权以便不同准则比其它准则更多地影响排序。排序准则的实例包括数据与目标生物序列的相似性、数据内容与目标生物序列的相关性、数据或引用的重要性或威望、数据或引用的普及性，以及数据或引用对目标生物序列的历史适用性。

[0057] 在一个实施例中，用户可以添加准则，删除准则，并且调整用于对相关数据、对数据的引用和与数据关联的探测序列进行排序的准则的权重。在本公开的实施例中，不同用户可以生成不同简档，或者可以另外指示不同偏好以便对与目标生物序列相关的信息进行排序。在方框 707，可以根据排序，将相关数据绑定到目标生物序列或目标生物序列片段。具体地说，可以显示目标生物序列，并且可以将图标与表示数据、引用和探测序列的目标生物序列一起显示。

[0058] 图形显示可以显示以下各项的图标或其它表示：数据、引用或探测以及目标生物序列，或者生物序列的一个或多个目标片段。可以通过以下操作显示数据、引用或探测序列的排序：将与具有较高排序的数据关联的图标显示为更靠近目标生物序列的对应片段，而将表示具有较低排序的数据的图标远离目标生物序列片段。

[0059] 图 8 示出根据本公开的另一个实施例的计算机系统 800 的框图。计算机 800 例如可以对应于图 1 的主计算机 110。在此描述的方法可以以硬件、软件（例如，固件）或它们的组合实现。在示例性实施例中，在此描述的方法以硬件实现，该硬件作为专用或通用数字计算机（例如个人计算机、工作站、小型计算机或大型计算机）的微处理器的一部分。因此，系统 800 可以包括通用计算机或大型机 801，其能够通过随时间逐渐增加基本程序的工作负载，测试基本程序的可靠性。

[0060] 在示例性实施例中，在硬件体系架构方面，如图 8 中所示，计算机 801 包括一个或多个处理器 805、连接到存储控制器 815 的存储器 810，以及一个或多个输入和 / 或输出（I/O）设备 840、845（或外围设备），它们以通信方式通过本地输入 / 输出控制器 835 连接。输入 / 输出控制器 835 例如可以是一条或多条总线或者其它有线或无线连接，如所属技术领域公知的那样。输入 / 输出控制器 835 可以具有其它元件（在描述中为简单起见而被省略），例如控制器、缓冲器（高速缓存）、驱动器、中继器和接收器，以实现通信。进一步，本地接口可以包括地址、控制和 / 或数据连接以便在上述组件之间实现适当的通信。输入 / 输出控制器 835 可以包括被配置为访问输出设备 840 和 845 的多个子通道。子通道例如可以包括光纤通信端口。

[0061] 处理器 805 是用于执行软件（具体地说，存储在存储装置 820（例如高速缓存存储装置）或存储器 810 中的软件）的硬件设备。处理器 805 可以是任何定制或商用处理器、中央处理单元（CPU）、与计算机 801 关联的多个处理器之间的辅助处理器、基于半导体的微处理器（以微芯片或芯片组的形式）、宏处理器，或者通常用于执行指令的任何设备。

[0062] 存储器 810 可以包括以下各项的任何一个或组合：易失性存储元件（例如，随机存取存储器（RAM，例如 DRAM、SRAM、SDRAM 等））和非易失性存储元件（例如，ROM、可擦式可编程只读存储器（EPROM）、电可擦式可编程只读存储器（EEPROM）、可编程只读存储器（PROM）、磁带、紧凑盘只读存储器（CD-ROM）、磁盘、软盘、盒带、卡带等）。此外，存储器 810 可以包括电、

磁、光和 / 或其它类型的存储介质。要注意的是，存储器 810 可以具有分布式体系架构，其中各种组件可以彼此远离，但可以由处理器 805 访问。

[0063] 存储器 810 中的指令可以包括一个或多个单独程序，每个程序包括用于实现逻辑功能的可执行指令的有序列表。在图 8 的实例中，存储器 810 中的指令包括合适的操作系统（O/S）811。操作系统 811 基本上控制其它计算机程序的执行，并且提供调度、输入输出控制、文件和数据管理、存储管理以及通信控制和相关服务。

[0064] 在示例性实施例中，常规键盘 850 和鼠标 855 可以连接到输入 / 输出控制器 835。其它输出设备（例如 I/O 设备 840、845）可以包括输出设备，例如—但不限于—打印机、扫描仪、麦克风等。最后，I/O 设备 840、845 可以还包括传送输入和输出的设备，例如—但不限于—网络接口卡（NIC）或调制器 / 解调器（用于访问其它文件、设备、系统或网络）、射频（RF）或其它收发器、电话接口、桥接器、路由器等。系统 800 可以还包括连接到显示器 830 的显示控制器 825。在示例性实施例中，系统 800 可以还包括网络接口 860 以便连接到网络 865。网络 865 可以是基于 IP 的网络，以便通过宽带连接在计算机 801 和任何外部服务器、客户机等之间通信。网络 865 在计算机 801 和外部系统之间发送和接收数据。在示例性实施例中，网络 865 可以是由服务提供商管理的受管 IP 网络。网络 865 可以以无线方式实现，例如使用无线协议和技术（例如 WiFi、WiMax 等）。网络 865 还可以是分组交换网络，例如局域网、广域网、城域网、因特网或其它类似类型的网络环境。网络 865 可以是固定无线网络、无线局域网（LAN）、无线广域网（WAN）、个人区域网络（PAN）、虚拟专用网络（VPN）、内联网或其它合适的网络系统，并且包括用于接收和发送信号的设备。

[0065] 当计算机 801 操作时，处理器 805 被配置为执行存储在存储器 810 中的指令，以便往返于存储器 810 而传送数据，并且通常根据指令控制计算机 801 的操作。

[0066] 在示例性实施例中，在此描述的方法可以使用以下技术的任何一种或组合（每种技术都是所属技术领域公知的）实现：具有用于针对数据信号实现逻辑功能的逻辑门的离散逻辑电路（多个）、具有适当的组合逻辑门的专用集成电路（ASIC）、可编程门阵列（多个）（PGA）、现场可编程门阵列（FPGA）等。

[0067] 在本公开的实施例中，模拟退火模块 111 可以包括程序代码，其存储在存储器 810 中并由处理器 805 执行。数据和指向数据的引用可以存储在计算机 801 中，或者可以存储在通过网络连接到计算机 801 的其它计算机、服务器、数据库或其它网络设备中。模拟退火模块 111 可以还包括硬件组件，例如处理器、存储器和逻辑芯片或结构，以便实现模拟退火。

[0068] 如上所述，可以以用于实现这些过程的计算机实现的过程和装置的形式包含实施例。实施例可以包括计算机程序产品 900，如图 9 中所示，在具有计算机程序代码逻辑 904 的计算机可读 / 可用介质 902 中包含指令，这些指令作为制造品包含在有形介质中。计算机可读 / 可用介质 902 的示例性制造品可以包括软盘、CD-ROM、硬盘驱动器、通用串行总线（USB）闪存驱动器或任何其它计算机可读存储介质，其中当计算机程序代码逻辑 904 被加载到计算机并由计算机执行时，该计算机变成用于实现实施例的装置。实施例包括计算机程序代码逻辑 904，例如无论存储在存储介质中，加载到计算机和 / 或由计算机执行，还是通过某种传输介质传输（例如通过电线或电缆、通过光纤或通过电磁辐射），其中当计算机程序代码逻辑 904 被加载到计算机并由计算机执行时，该计算机变成用于实现实施例的装置。当在通用微处理器上实现时，计算机程序代码逻辑 904 各段配置微处理器以产生特定

逻辑电路。

[0069] 所属技术领域的技术人员知道，本公开的各个方面可以实现为系统、方法或计算机程序产品。因此，本公开的各个方面可以具体实现为以下形式，即：完全的硬件实施方式、完全的软件实施方式（包括固件、驻留软件、微代码等），或硬件和软件方面结合的实施方式，这里可以统称为“电路”、“模块”或“系统”。此外，本公开的各个方面还可以实现为在一个或多个计算机可读介质中的计算机程序产品的形式，该计算机可读介质中包含计算机可读的程序代码。

[0070] 可以采用一个或多个计算机可读介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是—但不限于—电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者上述的任意合适的组合。计算机可读存储介质的更具体的例子（非穷举的列表）包括：具有一个或多个导线的电连接、便携式计算机盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0071] 计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括—但不限于—无线、有线、光缆、RF 等等，或者上述的任意合适的组合。

[0072] 可以以一种或多种程序设计语言的任意组合来编写用于执行本公开的各个方面的操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如 Java、Smalltalk、C++ 等，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络—包括局域网（LAN）或广域网（WAN）—连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

[0073] 上面参照根据本公开实施例的方法、装置（系统）和计算机程序产品的流程图和/或框图描述了本公开的各个方面。应当理解，流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合，都可以由计算机程序指令实现。这些计算机程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器，从而生产出一种机器，使得这些指令在通过计算机或其它可编程数据处理装置的处理器执行时，产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。

[0074] 也可以把这些计算机程序指令存储在计算机可读介质中，这些指令使得计算机、其它可编程数据处理装置、或其它设备以特定方式工作，从而，存储在计算机可读介质中的指令就产生出包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的指令的制造品（article of manufacture）。

[0075] 也可以把计算机程序指令加载到计算机、其它可编程数据处理装置、或其它设备上，使得在计算机、其它可编程装置或其它设备上执行一系列操作步骤，以产生计算机实现的过程，从而使得在计算机或其它可编程装置上执行的指令提供实现流程图和/或框图中的一个或多个方框中规定的功能/动作的过程。

[0076] 附图中的流程图和框图显示了根据本公开的不同实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分，所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能可以以不同于附图中所标注的顺序发生。例如，两个连续的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的是，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或动作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

[0077] 在此使用的术语只是为了描述特定的实施例并且并非旨在将本发明限于所述的特定实施例。如在此所使用的，单数形式“一”、“一个”和“该”旨在同样包括复数形式，除非上下文明确地另有所指。还将理解，当在此说明书中使用时，术语“包括”和/或“包含”指定了声明的特性、整数、步骤、操作、元素和/或组件的存在，但是并不排除一个或多个其它特性、整数、步骤、操作、元素、组件和/或其组的存在或增加。

[0078] 下面权利要求中的对应结构、材料、操作以及所有装置或步骤和功能元件的等同替换，旨在包括任何用于与在权利要求中具体指出的其它元件相组合地执行该功能的结构、材料或操作。出于示例和说明目的给出了对公开的描述，但所述描述并非旨在是穷举的或是限于所公开的实施例。在不偏离本公开的实施例的范围和精神的情况下，对于所属技术领域的普通技术人员来说许多修改和变化都将是显而易见的。

[0079] 尽管上面描述了本公开的实施例，但所属技术领域的技术人员将理解的是，可以在现在和将来进行各种属于下面权利要求范围的改进和增强。

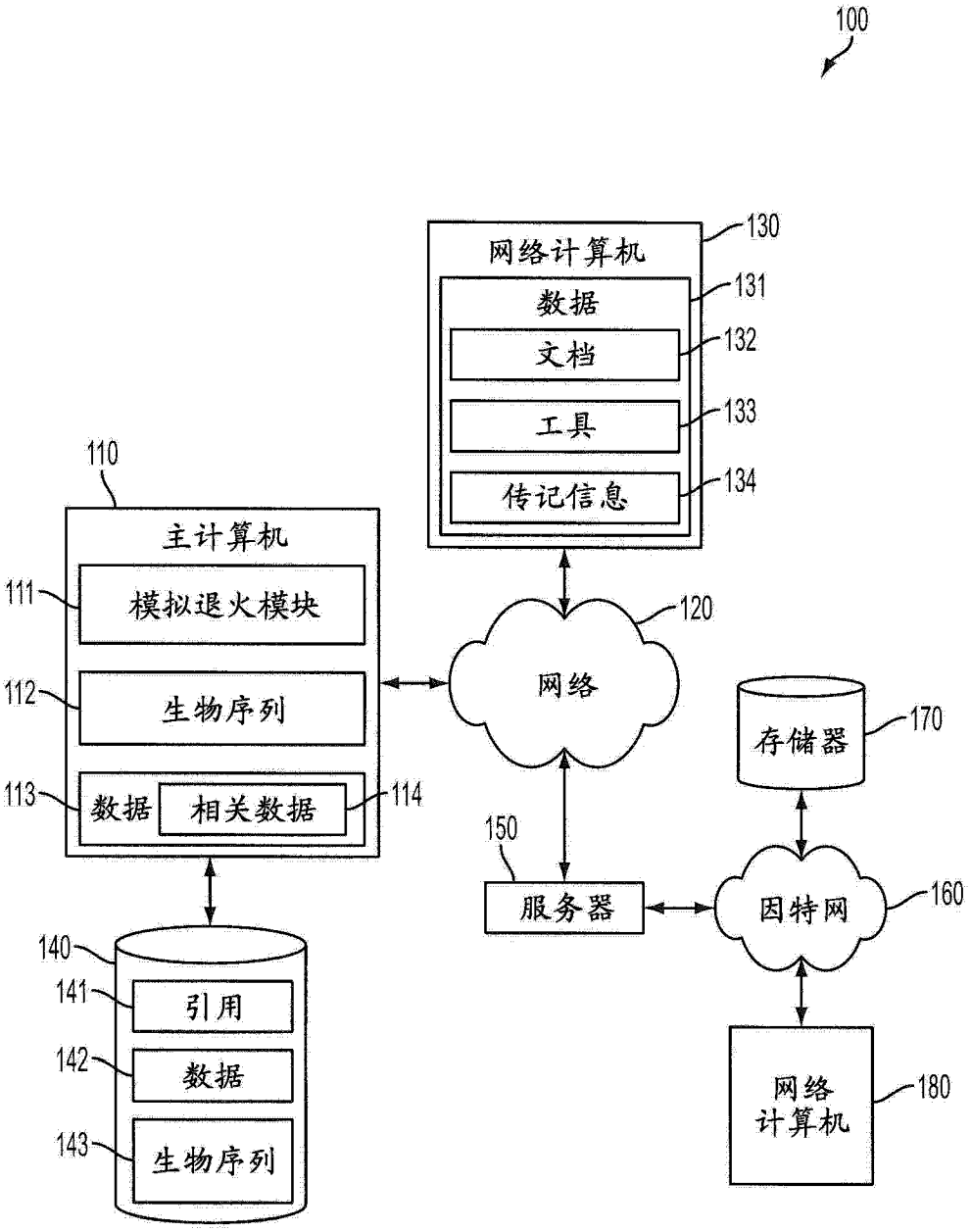

100
网络计算机
130
数据
131
文档
132
工具
133
传记信息
134
110
主计算机
111
模拟退火模块
112
生物序列
113
数据
相关数据
114
120
网络
存储器
170
150
服务器
因特网
160
140
引用
141
数据
142
生物序列
143
网络
计算机
180

图 1

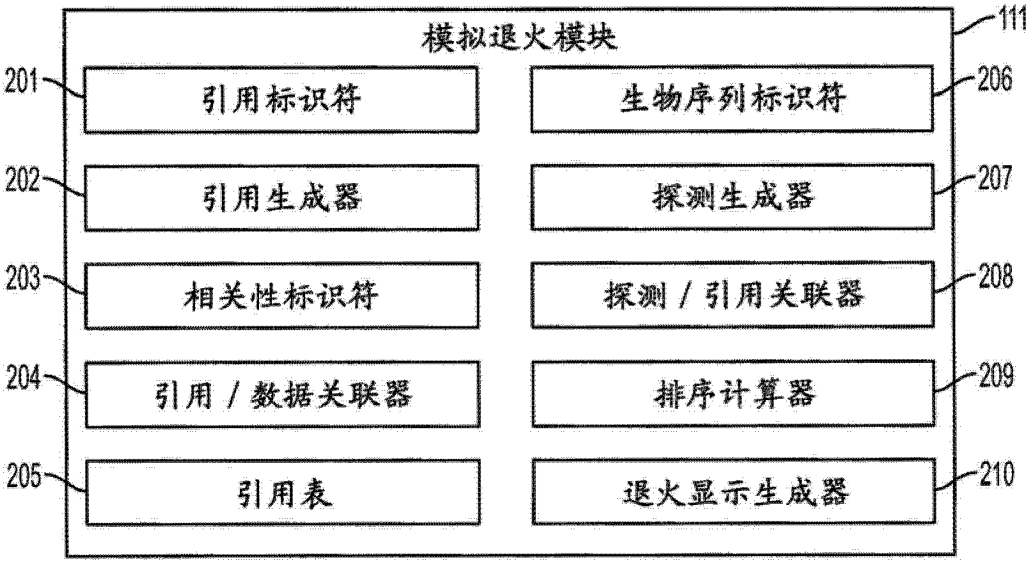

模拟退火模块
111
201
引用标识符
206
生物序列标识符
202
引用生成器
207
探测生成器
203
相关性标识符
208
探测 / 引用关联器
204
引用 / 数据关联器
209
排序计算器
205
引用表
210
退火显示生成器

图 2

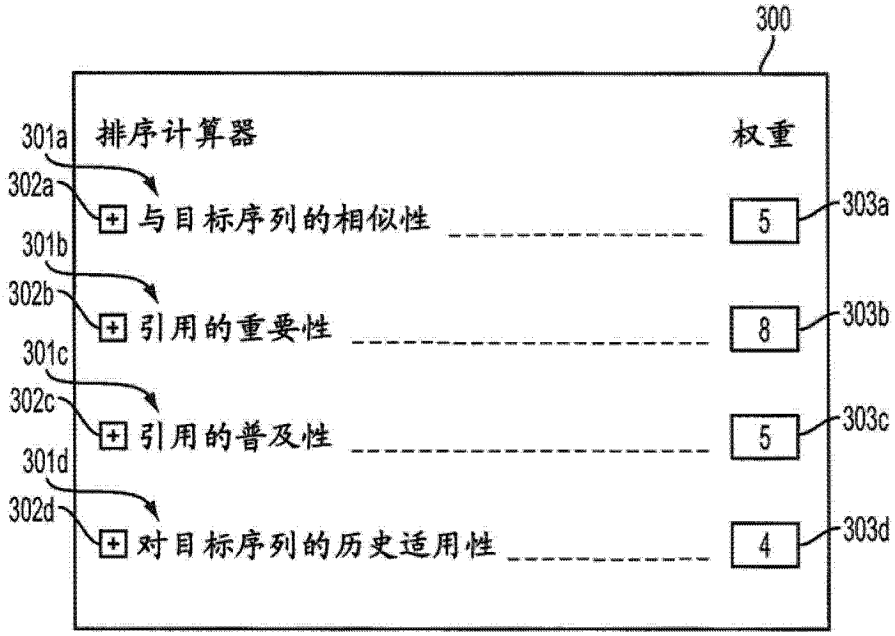

300
301a
排序计算器
权重
302a
与目标序列的相似性
5
303a
301b
302b
引用的重要性
8
303b
301c
302c
引用的普及性
5
303c
301d
302d
对目标序列的历史适用性
4
303d

图 3

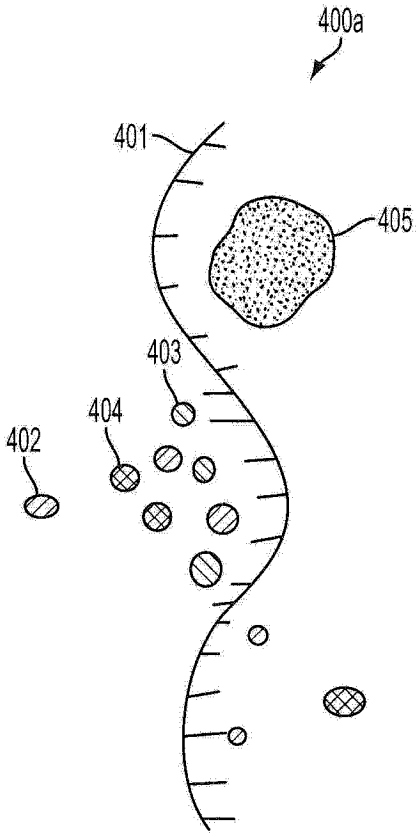

400a
401
405
403
404
402

图 4A

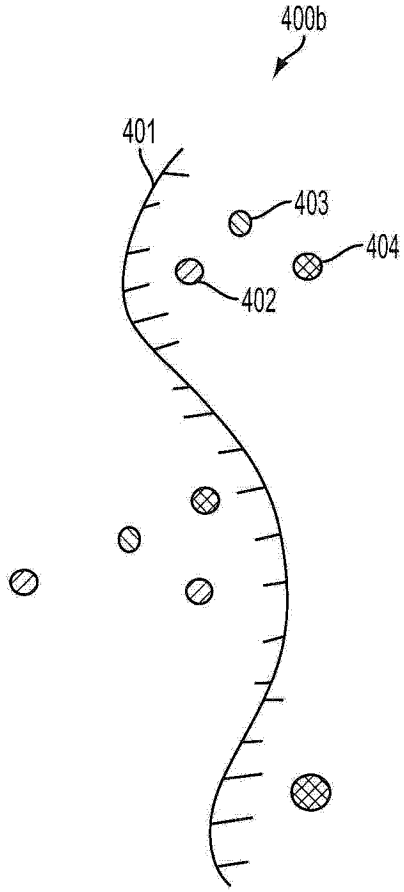

400b
401
403
404
402

图 4B

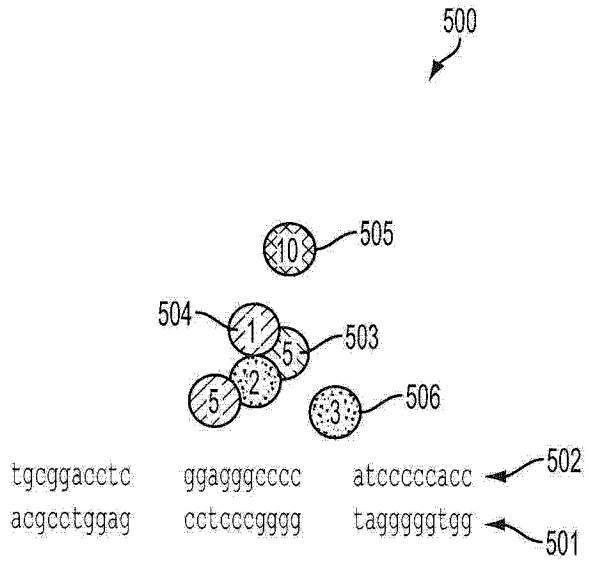

500
10
505
504
1
5
503
5
2
3
506
tgcggacctc ggagggcccc atcccccacc
502
acgcctggag cctcccgggg taggggtgg
501

图 5

表		
引用	数据	探测
URL1	文档	tgccggaatc
URL2	工具	gatgggcccc
URN1	传记信息	gtcgggaac
引用1	组织	tgccggaatc

图 6

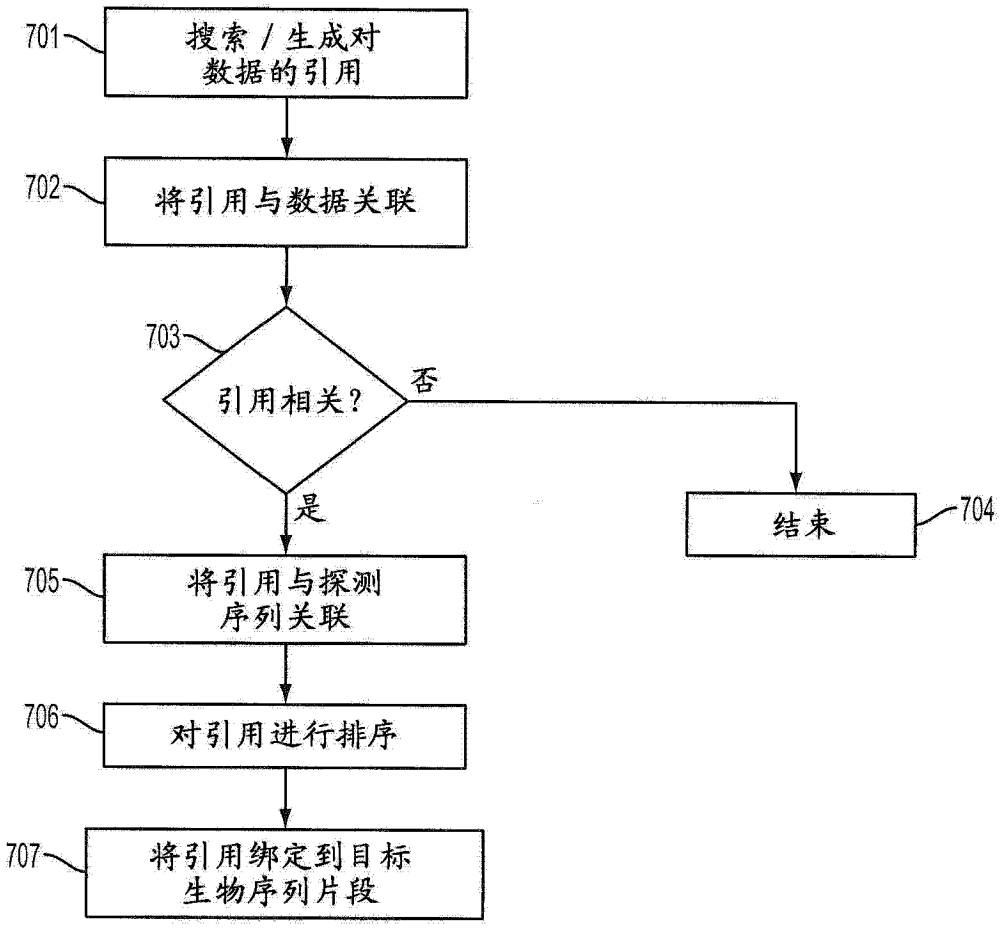

701
搜索 / 生成对数据的引用
702
将引用与数据关联
703
引用相关？
否
是
704
结束
705
将引用与探测序列关联
706
对引用进行排序
707
将引用绑定到目标生物序列片段

图 7

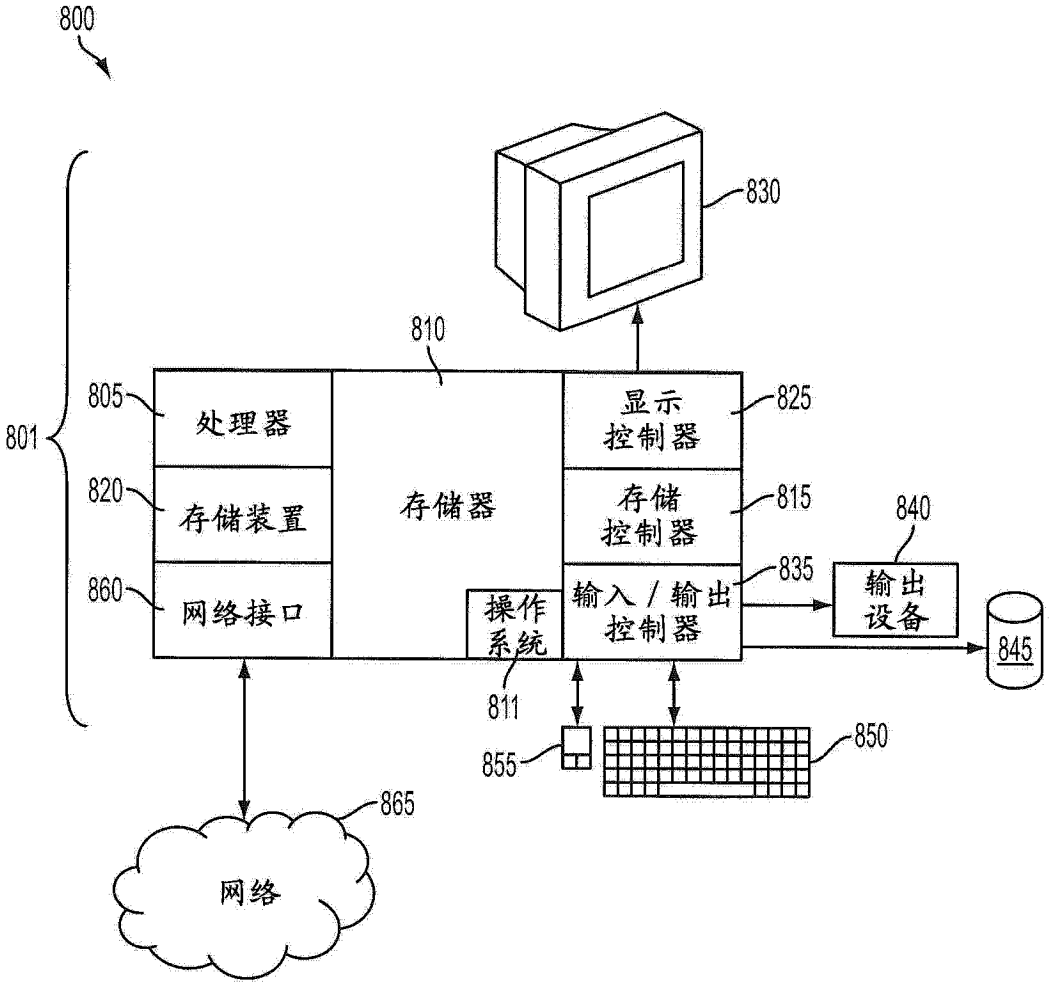

800
830
810
801
805
处理器
825
显示
控制器
820
存储装置
存储器
815
存储
控制器
840
860
网络接口
操作
系统
835
输入 / 输出
控制器
输出
设备
845
811
855
850
865
网络

图 8

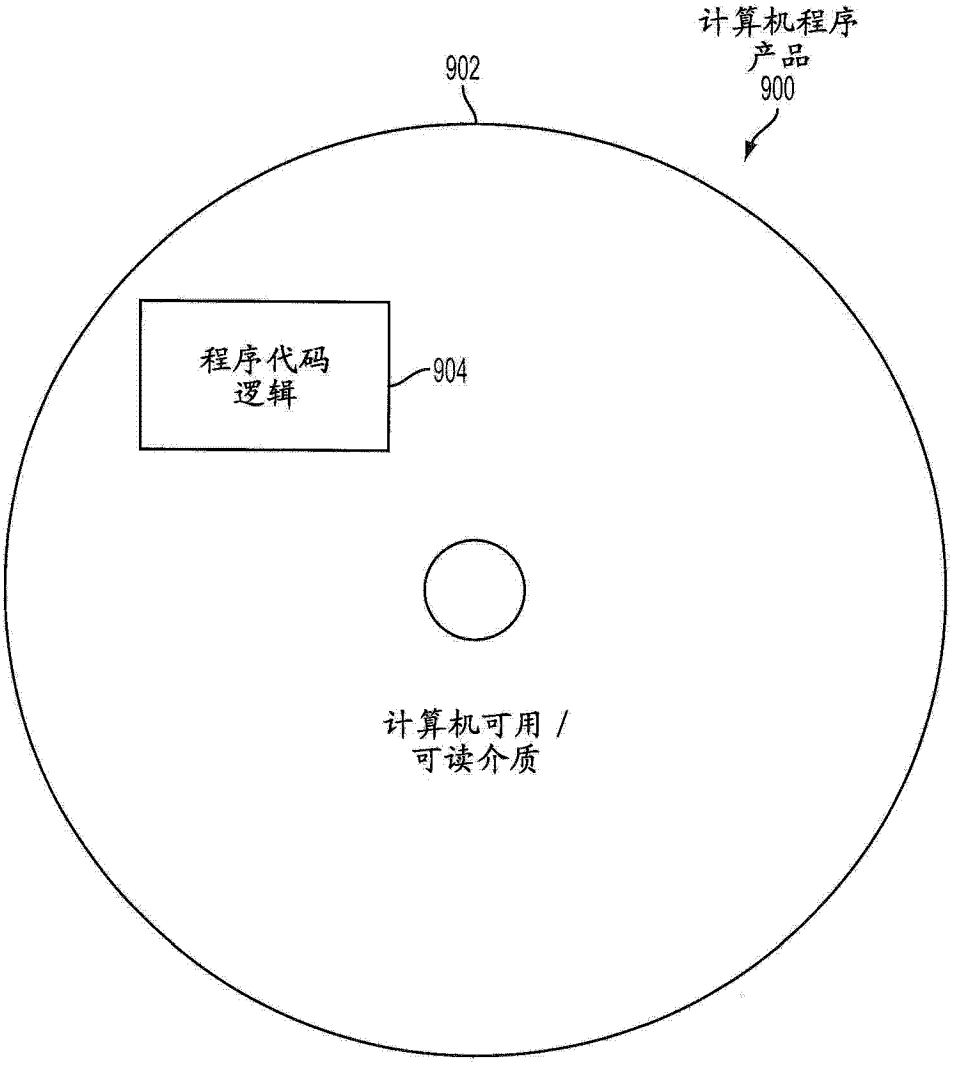


图 9