



(12) 发明专利申请

(10) 申请公布号 CN 116710926 A

(43) 申请公布日 2023. 09. 05

(21) 申请号 202280009077.8

(22) 申请日 2022.01.18

(30) 优先权数据

2021-038519 2021.03.10 JP

(85) PCT国际申请进入国家阶段日

2023.07.05

(86) PCT国际申请的申请数据

PCT/JP2022/001580 2022.01.18

(87) PCT国际申请的公布数据

W02022/190646 JA 2022.09.15

(71) 申请人 聪丰罗株式会社

地址 日本大阪府

(72) 发明人 白川基光

(74) 专利代理机构 北京三友知识产权代理有限公司 11127

专利代理师 黄纶伟

(51) Int.Cl.

G06N 3/02 (2006.01)

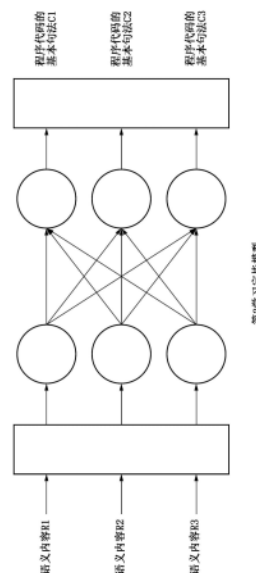
权利要求书2页 说明书10页 附图11页

(54) 发明名称

程序代码自动生成装置以及程序

(57) 摘要

[课题]在生成用于执行业务的程序代码时，能够通过极其容易且不经人手的方式自动地生成。[解决手段]使计算机执行以下步骤：文本数据提取步骤，从文档中提取出文本数据；语义内容搜索单元，参照文本数据与其语义内容以相关度关联起来的第1学习完毕模型，搜索与在所述文本数据提取步骤中提取出的文本数据关联性高的语义内容；以及代码提取步骤，参照语义内容与程序代码的基本句法以相关度关联起来的第2学习完毕模型，基于在所述语义内容搜索步骤中搜索出的语义内容来提取关联性高的程序代码的基本句法。



1. 一种程序代码自动生成装置,其特征在于,具备:

文本数据提取单元,其从文档中提取作为文章的文本数据;

语义内容搜索单元,其参照第1相关性,搜索与由所述文本数据提取单元提取出的文本数据关联性高的语义内容,在该第1相关性中,通过进行词素解析而提取出了包含动词、名词以及格助词在内的文章的各个构成要素的文本数据与其语义内容相互关联;以及

代码提取单元,其参照第2相关性,基于由所述语义内容搜索单元搜索出的语义内容来提取关联性高的程序代码的基本句法,在该第2相关性中,语义内容与程序代码的基本句法相互关联。

2. 根据权利要求1所述的程序代码自动生成装置,其特征在于,

所述语义内容搜索单元参照文本数据与其语义内容之间以3等级以上的相关度关联起来的所述第1相关性,

所述代码提取单元参照语义内容与程序代码的基本句法之间以3等级以上的相关度关联起来的所述第2相关性。

3. 根据权利要求2所述的程序代码自动生成装置,其特征在于,

所述语义内容搜索单元和所述代码提取单元使用与人工智能中的神经网络的节点的各输出的加权系数对应的所述相关度。

4. 根据权利要求1至3中任一项所述的程序代码自动生成装置,其特征在于,

所述程序代码自动生成装置还具备更新单元,该更新单元基于对所述文本数据中包含的各文章、各符号预先分配了语义内容的数据集来更新所述第1相关性,

所述文本数据提取单元提取文本数据中包含的各文章、各符号,

所述语义内容搜索单元参照由所述更新单元更新后的第1相关性,搜索与由所述文本数据提取单元提取出的文本数据中包含的各文章、各符号关联性高的语义内容。

5. 根据权利要求1至4中任一项所述的程序代码自动生成装置,其特征在于,

所述程序代码自动生成装置具备代码生成单元,该代码生成单元通过将所述文本数据提取单元从受理的文本数据中提取出的名词或名词短语代入到由所述代码提取单元提取出的程序代码的基本句法中,来生成程序代码。

6. 一种程序代码自动生成程序,其特征在于,使计算机执行以下步骤:

文本数据提取步骤,从文档中提取作为文章的文本数据;

语义内容搜索步骤,参照第1相关性,搜索与在所述文本数据提取步骤中提取出的文本数据关联性高的语义内容,在该第1相关性中,通过进行词素解析而提取出了包含动词、名词以及格助词在内的文章的各个构成要素的文本数据与其语义内容相互关联;以及

代码提取步骤,参照第2相关性,基于在所述语义内容搜索步骤中搜索出的语义内容来提取关联性高的程序代码的基本句法,在该第2相关性中,语义内容与程序代码的基本句法相互关联。

7. 根据权利要求6所述的程序代码自动生成程序,其特征在于,

在所述语义内容搜索步骤中,参照文本数据与其语义内容之间以3等级以上的相关度关联起来的所述第1相关性,

在所述代码提取步骤中,参照语义内容与程序代码的基本句法之间以3等级以上的相关度关联起来的所述第2相关性。

8. 根据权利要求7所述的程序代码自动生成程序,其特征在于,
在所述语义内容搜索步骤及所述代码提取步骤中,使用与人工智能中的神经网络的节点的各输出的加权系数对应的所述相关度。
9. 根据权利要求6至8中任一项所述的程序代码自动生成程序,其特征在于,
所述程序代码自动生成程序还使计算机执行更新步骤,在该更新步骤中,基于对所述文本数据中包含的各文章、各符号预先分配了语义内容的数据集来更新所述第1学习模型,
在所述文本数据提取步骤中,提取文本数据中包含的各文章、各符号,
在所述语义内容搜索步骤中,参照通过所述更新步骤更新后的第1学习模型,搜索与在所述文本数据提取步骤中提取出的文本数据中包含的各文章、各符号关联性高的语义内容。
10. 根据权利要求6至9中任一项所述的程序代码自动生成程序,其特征在于,
所述程序代码自动生成程序还使计算机执行代码生成步骤,在该代码生成步骤中,通过将所述文本数据提取步骤中从受理的文本数据中提取出的名词或名词短语代入到在所述代码提取步骤中提取出的程序代码的基本句法中,来生成程序代码。

程序代码自动生成装置以及程序

技术领域

[0001] 本发明涉及适于自动生成基于文档中包含的文本数据的语义内容的程序代码的程序代码自动生成装置以及程序。

背景技术

[0002] 在利用程序自动地执行新业务的情况下,需要进行创建程序的作业。关于程序的创建,以往是进行该新业务的要件定义,进行系统设计,进而进行程序代码的开发,之后产生对其进行测试、验证的过程。这样的程序代码通常在每次产生新业务时通过人工来进行编码。

[0003] 然而,随着近年来IT化的快速发展,新业务涉及多方面,新产生这些业务的频率也增加。

[0004] 另外,对于公司内部的一个简单的作业而言,存在必须要根据状况随时实施变更的情况。例如,即便将“将员工○○的本月的加班时间通知给上司”这样的公司内部业务程序代码化并且能够自动地进行,每当“员工○○”、“上司”因调动等而发生变更时,必须要改写该程序代码。

[0005] 这样,会存在如下问题:每当新业务增大和业务内容变更时,如果通过人工来生成程序代码,则作业量将变得庞大,不仅作业者的作业负担增大,而且如果产生作业的延迟,则有可能会阻碍业务的流程。

[0006] 因此,以往,为了在计算机侧的系统中自动地执行业务,提出了如下的程序代码自动生成装置:在生成用于执行该业务的程序代码时,能够极其容易且不经过人手而自动地生成(例如,参照专利文献1)。

[0007] 现有技术文献

[0008] 专利文献

[0009] 专利文献1:日本专利6753598号公报

发明内容

[0010] 发明要解决的课题

[0011] 然而,上述的专利文献1的公开技术只是受理对话语句,并根据该对话语句所表达出的意图来搜索程序代码的基本句法。即,该技术专用于自动生成与用语言发出的一个短语对应的程序代码。因此,根据该专利文献1的公开技术,存在无法将以设计书、手册、规格书、各种说明书、计划书等为代表的各种文档所记载的数千、数万个语句自动地程序代码化的问题。

[0012] 如果能够自动地生成与这样的文档中记载的各语句对应的程序代码,则还能够将迄今为止依赖于人工的操作全部自动化。因此,近年来,对通过仅读入文档来自动且准确地生成与其语义内容相应的程序代码的技术的需求不断提高,但现状是尚未提出能够满足该需求的技术。

[0013] 因此,本发明是鉴于上述问题而提出的,目的在于提供一种程序代码自动生成装置以及程序,能够通过仅读入文档来自动且准确地生成与其语义内容相应的程序代码。

[0014] 用于解决课题的手段

[0015] 第1发明的特征在于,具备:文本数据提取单元,其从文档中提取作为文章的文本数据;语义内容搜索单元,其参照第1相关性,搜索与由所述文本数据提取单元提取出的文本数据关联性高的语义内容,在该第1相关性中,通过进行词素解析而提取出了包含动词、名词以及格助词在内的文章的各个构成要素的文本数据与其语义内容相互关联;以及代码提取单元,其参照第2相关性,基于由所述语义内容搜索单元搜索出的语义内容来提取关联性高的程序代码的基本句法,在该第2相关性中,语义内容与程序代码的基本句法相互关联。

[0016] 第2发明的特征在于,在第1发明中,所述语义内容搜索单元参照文本数据与其语义内容之间以3等级以上的相关度关联起来的所述第1相关性,所述代码提取单元参照语义内容与程序代码的基本句法之间以3等级以上的相关度关联起来的所述第2相关性。

[0017] 第3发明的特征在于,在第2发明中,所述语义内容搜索单元和所述代码提取单元使用与人工智能中的神经网络的节点的各输出的加权系数对应的所述相关度。

[0018] 第4发明的特征在于,在第1发明~第3发明的任一之中,还具备更新单元,该更新单元基于对所述文本数据中包含的各文章、各符号预先分配了语义内容的数据集来更新所述第1相关性,所述文本数据提取单元提取文本数据中包含的各文章、各符号,所述语义内容搜索单元参照由所述更新单元更新后的第1相关性,搜索与由所述文本数据提取单元提取出的文本数据中包含的各文章、各符号关联性高的语义内容。

[0019] 第5发明的特征在于,在第1发明~第4发明的任一之中,具备代码生成单元,该代码生成单元通过将所述文本数据提取单元从受理的文本数据中提取出的名词或名词短语代入到由所述代码提取单元提取出的程序代码的基本句法中,来生成程序代码。

[0020] 第6发明的特征在于,使计算机执行以下步骤:文本数据提取步骤,从文档中提取作为文章的文本数据;语义内容搜索步骤,参照第1相关性,搜索与在所述文本数据提取步骤中提取出的文本数据关联性高的语义内容,在该第1相关性中,通过进行词素解析而提取出了包含动词、名词以及格助词在内的文章的各个构成要素的文本数据与其语义内容相互关联;以及代码提取步骤,参照第2相关性,基于在所述语义内容搜索步骤中搜索出的语义内容来提取关联性高的程序代码的基本句法,在该第2相关性中,语义内容与程序代码的基本句法相互关联。

[0021] 第7发明的特征在于,在第6发明中,在所述语义内容搜索步骤中,参照文本数据与其语义内容之间以3等级以上的相关度关联起来的所述第1相关性,在所述代码提取步骤中,参照语义内容与程序代码的基本句法之间以3等级以上的相关度关联起来的所述第2相关性。

[0022] 第8发明的特征在于,在第7发明中,在所述语义内容搜索步骤及所述代码提取步骤中,使用与人工智能中的神经网络的节点的各输出的加权系数对应的所述相关度。

[0023] 第9发明的特征在于,在第6发明~第8发明中,还使计算机执行更新步骤,在该更新步骤中,基于对所述文本数据中包含的各文章、各符号预先分配了语义内容的数据集来更新所述第1学习模型,在所述文本数据提取步骤中,提取文本数据中包含的各文章、各符

号,在所述语义内容搜索步骤中,参照通过所述更新步骤更新后的第1学习模型,搜索与在所述文本数据提取步骤中提取出的文本数据中包含的各文章、各符号关联性高的语义内容。

[0024] 第10发明的特征在于,在第6发明~第9发明的任一之中,还使计算机执行代码生成步骤,在该代码生成步骤中,通过将在所述文本数据提取步骤中从受理的文本数据中提取出的名词或名词短语代入到在所述代码提取步骤中提取出的程序代码的基本句法中,来生成程序代码。

[0025] 发明的效果

[0026] 根据上述的发明,能够通过极其容易且不经人手的方式将以设计书、手册、规格书、各种说明书、计划书等为代表的各种文档所记载的数千、数万个语句自动地程序代码化。

附图说明

[0027] 图1是实施方式中的程序代码自动生成系统的框图。

[0028] 图2的(a)和图2的(b)是示出程序代码自动生成装置1的结构的一例的示意图。

[0029] 图3是示出第1学习完毕模型的例子的图。

[0030] 图4是示出第1学习完毕模型采用基于人工智能的机器学习的代码表的例子的图。

[0031] 图5是示出作为输入数据而输入文本数据、作为输出数据而输出语义内容的模型的图。

[0032] 图6是示出第2学习完毕模型的例子的图。

[0033] 图7是示出第2学习完毕模型采用基于人工智能的机器学习的例子的图。

[0034] 图8是示出作为输入数据而输入语义内容、作为输出数据而输出程序代码的模型的图。

[0035] 图9是用于说明应用了本发明的程序代码自动生成系统的动作的流程图。

[0036] 图10是用于说明第1相关性、第2相关性的图。

[0037] 图11是示出从文档(设计书、文件)中提取文本数据的例子的图。

具体实施方式

[0038] 以下,参照附图对本发明的实施方式中的程序代码自动生成系统的一例进行说明。

[0039] (实施方式:程序代码自动生成系统100)

[0040] 参照图1~图2,对本实施方式中的程序代码自动生成系统100的结构的一例进行说明。图1是示出本实施方式中的程序代码自动生成系统100的整体结构的示意图。

[0041] 程序代码自动生成系统100主要用于生成用于实现定型作业等这样的业务的辅助(例如业务的自动化处理)的程序代码。程序代码自动生成系统100通过自动生成用于执行业务的程序代码,能够在计算机上自动地进行企业内的各业务(例如,手册中记载的业务流程的执行、作业者的进展状况的收集、任务管理等)。程序代码自动生成系统100特别是能够基于文本数据来设定该程序代码的自动生成,即便是系统管理者等这样的不具有专业知识用户(例如使用程序代码自动生成系统100来管理业务的使用者等),也能够容易地实现

用于使计算机自动地进行各文档中记载的业务流程的程序代码的自动生成。

[0042] 例如如图1所示,程序代码自动生成系统100具备程序代码自动生成装置1,用户可以使用程序代码自动生成装置1。程序代码自动生成系统100例如具备经由通信网4与程序代码自动生成装置1连接的终端2,用户可以经由终端2来使用程序代码自动生成装置1。程序代码自动生成系统100例如具备经由通信网4与程序代码自动生成装置1连接的服务器3,用户可以经由程序代码自动生成装置1或终端2与服务器3进行各种信息的收发来实现各单元。

[0043] <程序代码自动生成装置1>

[0044] 图2的(a)是示出程序代码自动生成装置1的结构的一例的示意图。作为程序代码自动生成装置1,例如使用个人计算机(PC)、智能手机、平板终端等公知的电子设备。程序代码自动生成装置1例如具备壳体10、CPU(Central Processing Unit:中央处理单元)101、ROM(Read Only Memory:只读存储器)102、RAM(Random Access Memory:随机存取存储器)103、保存部104、I/F 105~107、输入部108以及通知部109。各结构101~107通过内部总线110连接。

[0045] CPU 101控制整个程序代码自动生成装置1。ROM 102存储CPU 101的动作代码。RAM 103是在CPU 101进行动作时使用的作业区域。保存部104保存处理用数据等各种信息。作为保存部104,例如使用HDD(Hard Disk Drive:硬盘驱动器)、SSD(Solid State Drive:固态硬盘)等。

[0046] I/F 105是用于与终端2、服务器3、通信网4等进行各种信息的收发的接口。I/F106是用于与输入部108进行各种信息的收发的接口。I/F 107是用于与通知部109进行各种信息的收发的接口。

[0047] 作为输入部108,除了可以使用键盘以外,还可以使用摄像头、扫描仪等装置。使用程序代码自动生成装置1的用户例如经由输入部108读取各种文档中记载的文本数据。在此所说的文档是以设计书、手册、规格书、各种说明书、计划书等为代表的文档,但并不限于于此,包括由个人或在企业内以文件形式创建出的任何文档。另外,不仅包括实际上处于可供非特定多数的人阅读的状态的刊物,也包括只能由特定的人阅读的文档。另外,文档也包括以手写形式书写的笔记。这些文档并不限于通过印刷在纸介质上的印刷物来提供,也可以作为电子数据来提供。

[0048] 输入部108由供这样的文档所记载的数据输入的任何设备构成。若文档被当作印刷在纸介质上的印刷物来提供,则输入部108由能够读取该印刷物上的文本数据的扫描仪、OCR软件构成。另外,若文档由电子数据构成,则输入部108也可以由能够读取记载于该电子数据中的文本数据的OCR软件构成。

[0049] 通知部109显示保存在保存部104中的显示用数据等各种信息、程序代码自动生成装置1的处理状况等。作为通知部109,除了可以使用显示器以外,例如也可以使用扬声器。

[0050] 另外,例如可以使用相同的I/F作为I/F 105~I/F 107,例如也可以分别使用多个I/F作为各I/F 105~I/F 107。另外,在使用触摸面板式的显示器作为通知部109的情况下,通知部109也可以构成为包括输入部108。

[0051] 图2的(b)是示出程序代码自动生成装置1的功能的一例的示意图。程序代码自动生成装置1具备取得部11和运算部12,例如也可以具备执行部13、存储部14、输出部15以及

意图存储部16。此外,图2的(b)所示的各功能通过CPU 101将RAM 103作为作业区域而执行存储在保存部104等中的程序来实现。另外,各功能的一部分也可以由人工智能来控制。在此,“人工智能”可以基于任何公知的人工智能技术。

[0052] <取得部11>

[0053] 取得部11取得记载于文档中的文本数据。取得部11例如经由终端2或输入部108取得输入自文档的文本数据。例如在经由终端2或者输入部108从文档中提取出文本数据的情况下,取得部11使用公知的OCR技术识别文本数据的字符。此外,字符识别技术例如也可以经由通信网4使用云型的字符识别技术。

[0054] <运算部12>

[0055] 运算部12参照数据库,基于取得的文本数据来执行各种处理动作、运算。运算部12通过对所受理的文本数据进行词素解析来提取出以动词、名词、格助词等为代表的句子的各个构成要素。运算部12参照存储部14,提取与文本数据对应的程序代码的基本句法。另外,运算部12通过将从构成文本数据的字符串中提取出的名词或名词短语代入到提取出的程序代码的基本句法中,来生成程序代码。

[0056] <执行部13>

[0057] 执行部13基于在运算部12中生成的程序代码来执行业务处理。作为业务处理,例如可举出如下的定型作业:基于任务的内容、期限向负责人发送邮件;管理工作;以及对任务进行历史进行更新。使用能够将业务处理信息作为程序来使计算机执行的内容。

[0058] <存储部14>

[0059] 存储部14暂时保存由取得部11取得的文本数据。存储在该存储部14中的文本数据有时会基于运算部12、执行部13等的控制而被读出或更新。存储部14保存有第1学习完毕模型和第2学习完毕模型这至少两个学习完毕模型。

[0060] 图3示出了第1学习完毕模型的例子。第1学习完毕模型是从文档中提取出的文本数据与其语义内容之间以3等级以上的相关度被关联起来的学习完毕模型。在该第1学习完毕模型中,文本数据为输入,语义内容为输出。作为文本数据的种类,由文章构成,或由文章与符号的组合构成,或仅由符号构成。

[0061] 例如,在作为输入侧的文本数据而存在“将‘A’文件作为‘B’文件配置在‘C’文件夹中”的情况下,把作为输出侧的“将‘A’文件变更为‘B’文件名并复制到‘C’文件夹中”这样的语义内容以最高的相关度关联起来。

[0062] 在第1学习完毕模型采用基于人工智能的机器学习、深度学习的情况下,例如如图4所示,前提是预先设定了文本数据与语义内容之间的3等级以上的相关度。例如将文本数据P01~P03作为输入数据。例如,该文本数据P01是“将‘A’文件作为‘B’文件配置在‘C’文件夹中”,P02是“将‘B’文件合并到‘A’文件中”,P03是“提取出‘A’文件和‘B’文件”等。这样的作为输入数据的文本数据P01~P03分别与作为输出的语义内容R1~R4连结。

[0063] 另外,语义内容并不限定于如上述那样由实际上人能够阅读并解释的字符串构成的情况,也可以用表示语义内容的符号来进行表达,还可以用参数等进行表达。

[0064] 文本数据与作为其输出解的语义内容(例如,作为语义内容R1的“将‘A’文件变更为‘B’文件名并复制到‘C’文件夹中”等)通过3等级以上的相关度相互关联。文本数据经由该相关度排列在左侧,各语义内容经由相关度排列在右侧。相关度表示排列在左侧的文本

数据与某一语义内容之间的关联性高的程度。换言之,该相关度是表示各文本数据与什么样的语义内容相关联的可能性高的指标,表示根据文本数据选择最可能的语义内容时的准确性。在图4的例子中,作为相关度,示出了w13~w19。w13~w19如下方的表1所示那样用10等级进行了示出,越接近10,表示作为中间节点的组合与作为输出的语义内容相互关联程度越高,相反,越接近1,表示作为中间节点的组合与作为输出的语义内容相互关联程度越低。

[0065] [表1]

符号	相关度
w13	7
w14	2
w15	9
w16	5
w17	2
w18	1
w19	8
w20	6
w21	10
w22	3

[0067] 预先取得这样的图4所示的3等级以上的相关度w13~w19。即,在进行实际的搜索解的判别时,预先累积对文本数据P01~P03和语义内容R1~R4中的任意文本数据和语义内容进行了采用及评价的过去的数据集,通过对它们进行分析、解析来预先生成图4所示的相关度。

[0068] 例如,假设在过去判断并评价为语义内容R1与文本数据P01的适合性最高。通过收集并分析这样的数据集,与该语义内容的相关度变强。

[0069] 该分析、解析也可以通过人工智能进行。在该情况下,例如在是文本数据P01的情况下,在语义内容R1的事例多的情况下,将与该语义内容R1相关的相关度设定得更高,在语义内容R2的事例多的情况下,将与该语义内容R2相关的相关度设定得更高。例如在文本数据P01的例子中,与语义内容R1和语义内容R2链接,但根据以前的事例,将与语义内容R1相关的w13的相关度设定为7,将与语义内容R2相关的w14的相关度设定为2。

[0070] 此外,文本数据由符号构成的情况也同样,通过过去的数据集来学习各符号被解释成了什么样的语义内容。由此,通过参照该第1学习完毕模型,能够根据符号搜索出语义内容。

[0071] 此外,该图4所示的相关度也可以由人工智能中的神经网络的节点构成。即,该神经网络的节点针对输出的加权系数与上述相关度对应。另外,不限于神经网络,也可以由构成人工智能的任何意思决定因子构成。

[0072] 在该情况下,也可以如图5所示那样,作为输入数据而输入文本数据,作为输出数据而输出语义内容,在输入节点与输出节点之间设置至少1个以上的隐藏层,从而进行机器学习。在输入节点或隐藏层节点中的任意一方或双方设定上述相关度,该相关度成为各节点的加权,基于该加权进行输出的选择。并且,也可以在该相关度超过某个阈值的情况下选

择该输出。

[0073] 这样的相关度成为第1学习完毕模型。在生成了这样的第1学习完毕模型之后,能够实际根据文本数据进行语义内容的搜索。

[0074] 图6示出了第2学习完毕模型的例子。第2学习完毕模型是语义内容与程序代码的基本句法以3等级以上的相关度关联起来的学习完毕模型。在该第2学习完毕模型中,语义内容为输入,程序代码的基本句法为输出。在该第2学习完毕模型中,输入侧的语义内容对应于第1学习完毕模型的输出侧。

[0075] 例如,在作为输入侧的语义内容是“将‘A’文件变更为‘B’文件名并复制到‘C’文件夹中”的情况下,作为输出侧的程序代码的基本句法,“cpA./C/B(copy Ato文件夹/B)”以最高的相关度被关联起来。

[0076] 在第2学习完毕模型采用基于人工智能的机器学习、深度学习的情况下,例如如图7所示,前提是预先设定了语义内容与程序代码的基本句法之间的3等级以上的相关度。例如将语义内容R01~R03作为输入数据。

[0077] 即,语义内容R1~R3与作为输出解的程序代码的基本句法C1~C4通过3等级以上的相关度相互关联。语义内容R1~R3经由该相关度排列在左侧,各程序代码的基本句法C1~C4经由相关度排列在右侧。相关度表示排列在左侧的语义内容R1~R3与某一程序代码的基本句法C1~C4关联性高的程度。换言之,该相关度是表示各语义内容R1~R3与什么样的程序代码的基本句法C1~C4相关联的可能性高的指标,表示根据语义内容选择最可能的程序代码的基本句法时的准确性。在图7的例子中,作为相关度的例子示出了w13~w19。

[0078] 预先取得这样的图7所示的3等级以上的相关度w13~w19。即,在进行实际的搜索解的判别时,预先累积对语义内容R1~R3和程序代码的基本句法C1~C4中的任意语义内容和基本句法进行了采用及评价的过去的数据集,通过对它们进行分析、解析来预先生成图7所示的相关度。

[0079] 例如,假设在过去判断并评价为程序代码的基本句法C3与语义内容R02的适合性最高。通过收集并分析这样的数据集,与该语义内容的相关度变强。

[0080] 该分析、解析也可以通过人工智能进行。在该情况下,例如在是语义内容R02的情况下,在程序代码C2的事例多的情况下,将与该程序代码C2相关的相关度设定得更高,在程序代码C3的事例多的情况下,将与该程序代码C3相关的相关度设定得更高。

[0081] 此外,该图7所示的相关度也可以由人工智能中的神经网络的节点构成。在该情况下,也可以如图8所示,作为输入数据而输入语义内容,作为输出数据而输出程序代码,在输入节点与输出节点之间设置至少1个以上的隐藏层,从而进行机器学习。

[0082] 这样的相关度成为第2学习完毕模型。在生成了这样的第2学习完毕模型之后,能够实际根据语义内容进行程序代码的基本句法的搜索。

[0083] 通过将这样的第1学习完毕模型和第2学习完毕模型存储于存储部14,能够在运算部12进行运算的过程中读出并参照该第1学习完毕模型和第2学习完毕模型。

[0084] <输出部15>

[0085] 输出部15输出与通过程序代码执行的动作相关的各种信息。显示用数据经由通知部109或者终端2等以能被用户识别的方式通知给用户。输出部15经由I/F 105向终端2等输出显示用数据等,经由I/F 107向通知部109输出显示用数据等。

[0086] <意图存储部16>

[0087] 意图存储部16存储1个或2个以上的意图。意图也可以与用于确定业务处理的信息对应地存储在该意图存储部16中。另外,用于确定业务处理的信息通常是后述的动作名,但其形式并不限于于此。并且,上述的对应例如也包括意图具有用于确定业务处理的信息的情况。

[0088] <终端2>

[0089] 作为终端2,例如使用个人计算机、智能手机、平板终端等公知的电子设备。终端2例如也可以具备与上述的程序代码自动生成装置1相同的结构及功能的至少一部分。例如可以具备多个终端2,各终端2也可以分别经由通信网4与程序代码自动生成装置1连接。

[0090] <服务器3>

[0091] 在服务器3中例如存储上述的各种信息。在服务器3中累积有例如经由通信网4从程序代码自动生成装置1等发送来的各种信息。也可以在服务器3中存储例如与保存部104相同的信息,服务器3也可以经由通信网4与程序代码自动生成装置1等进行各种信息的收发。即,在程序代码自动生成系统100中,也可以使用服务器3来代替程序代码自动生成装置1或程序代码自动生成装置1的保存部104、存储部14。

[0092] <通信网4>

[0093] 通信网4是供程序代码自动生成装置1经由通信电路进行连接的因特网等。通信网4也可以由所谓的光纤通信网构成。另外,通信网4除了可以通过有线通信网实现之外,也可以通过无线通信网等公知的通信网来实现。

[0094] 接着,对应用了本发明的程序代码自动生成系统100的动作进行说明。

[0095] 如图9所示,在步骤S11中从文档中提取文本数据。具体而言,通过构成输入部108的摄像头或扫描仪等来从文档中取得字符串作为电子数据。另外,在使用扫描仪等的情况下,利用OCR技术进行文字识别来取得文本数据。另外,在取得部11取得了电子数据化后的文本数据的情况下,直接使用该文本数据。以这样的方式取得的文本数据被暂时存储在存储部14中。

[0096] 接着,转移到步骤S12,将在S11中取得并暂时存储在存储部14中的文本数据读出,进行语义内容的关联解析。运算部12读出存储于存储部14的第1学习完毕模型,通过参照该第1学习完毕模型,搜索与文本数据相关度高的语义内容。在该情况下,例如如图4所示,在新取得的文本数据与P02相同或与P02类似的情况下,经由相关度 w_{15} 与语义内容R2关联起来,经由相关度 w_{16} 与语义内容R3关联起来。在该情况下,选择相关度最高的语义内容R2作为最优解。

[0097] 接着,转移到步骤S13,进行与程序代码的基本句法之间的关联解析。在该情况下,进行在步骤S12中搜索出的语义内容与关联性最高的程序代码的基本句法之间的关联解析。在该情况下,例如如图7所示,在新取得的语义内容与R02相同或与R02类似的情况下,经由相关度 w_{15} 与程序代码的基本句法C2关联起来,经由相关度 w_{16} 与程序代码的基本句法C3关联起来。在该情况下,选择相关度最高的程序代码的基本句法C2作为最优解。

[0098] 经过该步骤S12、S13,能够搜索出与从文档中提取出的文本数据关联性最高的语义内容,并且得到与搜索出的该语义内容关联性最高的程序代码的基本句法作为最优解。如果从文档中提取出文本数据,则之后能够自动地得到程序代码的基本句法的最优解。然

后,能够对提取出的各个文本数据分配搜索出的该程序代码的基本句法。

[0099] 接着,转移到步骤S14,进行程序代码的生成。在步骤S13中,如上述那样仅提取出了程序代码的基本句法,通过代入对实际的处理动作的对象、完成处理动作所需的各条件进行了规定的名词或名词短语,来完成程序代码。因此,在步骤S14中,进行如下的处理动作:将对实际的处理动作的对象、完成处理动作所需的各条件进行了规定的名词或名词短语代入到提取出的程序代码的基本句法中。

[0100] 在该情况下,对文本数据进行词素解析,提取出对实际的处理动作的对象、完成处理动作所需的各条件进行了规定的名词或名词短语。词素解析主要由运算部12进行。词素解析技术可以利用公知的任何词素解析技术。

[0101] 例如,在“登记A5-7853K”这样的文本数据中,假设在步骤S14中提取出了“INSERT INTO商品主文件(商品名) VALUES({param1})”作为程序代码的基本句法。此时,从通过词素解析而得到的命令语句中挑出应填充到{param1}处的实际的商品名。其结果是,作为商品名挑出了“A5-7853K”,并将其代入到基本句法中,由此能够完成程序代码。

[0102] 同样地,在“发送员工的本月的加班时间”中,在步骤S14中提取出“SELECT时间 FROM加班数据WHERE日期={param1}AND员工={param2}”作为程序代码的基本句法,从通过词素解析而得到的命令语句中挑出应填充到日期{param1}处的“本月”,并且挑出应填充到员工{param2}处的各员工名(例如“山田太郎”等),将它们代入到基本句法中,由此能够完成程序代码。

[0103] 在该步骤S11~S14的工序中,能够基于在步骤S11中受理的文本数据所记载的各动作的意图来自动生成程序。

[0104] 在以这样的方式完成程序代码之后,可以将其提供给用户,或者也可以使通知部109进行显示,还可以使执行部13执行完成的该程序代码。即,根据本发明,能够直接执行自动生成的该程序代码。因此,在包含从步骤S11开始的工序的情况下,通过从文档中提取出文本数据,能够自动生成加入了该意图的程序代码,并且能够直接执行生成的该程序代码。

[0105] 因此,根据本发明,通过仅读入以设计书、手册、规格书、各种说明书、计划书等为代表的各种文档中记载的数千、数万个语句,能够自动且准确地进行程序代码化。通过能够自动地生成与这样的文档中记载的各个语句对应的程序代码,还能够使迄今为止依赖于人工的操作全部自动化。

[0106] 另外,本发明并不限定于上述的实施方式。例如,如以下的图11所示,也可以应用第1相关性来代替第1学习完毕模型,应用第2相关性来代替第2学习完毕模型。

[0107] 该第1相关性由上述的文本数据与语义内容以相互一对一地对应的方式关联起来的表构成。另外,第2相关性由上述的语义内容与程序代码以相互一对一地对应的方式关联起来的表构成。

[0108] 预先创建这样的第1相关性和第2相关性。并且,在实际自动生成程序代码时,首先参照第1相关性,提取与从文档中提取出的文本数据相同或类似的文本数据所关联的语义内容。接着,参照第2相关性,确定与提取出的该语义内容相关联的程序代码。确定出程序代码后的程序代码的自动生成的过程与上述相同。

[0109] 在应用第1相关性来代替第1学习完毕模型、应用第2相关性来代替第2学习完毕模型的情况下也同样地,通过仅读入各种文档中记载的数千、数万个语句,能够自动且准确地

进行程序代码化。

[0110] 另外,如图10的(a)所示,在该第1相关性、第2相关性中,输入与输出也可以通过相互一对一的关系关联起来,但并不限于于此。也可以如图10的(b)所示那样,对一个输入关联多个输出,也可以对一个输出关联多个输入。

[0111] 实施例1

[0112] 图11示出了在步骤S11中从文档(设计书、文件)中提取出文本数据的例子。也可以提取出“创建(配置位置)/zip_new文件夹”、“将ken_all.zip解压缩并打开”等字符串本身作为记述在文档中的文本数据。另外,在存在“批处理检查一览实施No.1。”、“批处理检查一览实施No.1。”等这样的与引用关系有关的记述的情况下,提取该引用源的字符串作为文本数据。

[0113] 另外,在步骤S14中,在进行将对完成处理动作所需的各条件进行了规定的名词或名词短语代入的处理动作时,在图10的例子中,提取文件夹名“zip_new”、解压缩对象“ken_all.zip”等作为该名词或名词短语。然后,挑出提取出的该名词或名词短语,将其代入到在步骤S13中导出的基本句法中,由此能够完成程序代码。

[0114] 标号说明

[0115] 1 程序代码自动生成装置

[0116] 2 终端

[0117] 3 服务器

[0118] 4 通信网

[0119] 10 壳体

[0120] 11 取得部

[0121] 12 运算部

[0122] 13 执行部

[0123] 14 存储部

[0124] 15 输出部

[0125] 16 意图存储部

[0126] 100 程序代码自动生成系统

[0127] 101CPU

[0128] 102ROM

[0129] 103RAM

[0130] 104保存部

[0131] 105~107I/F

[0132] 108 输入部

[0133] 109 通知部

[0134] 110 内部总线

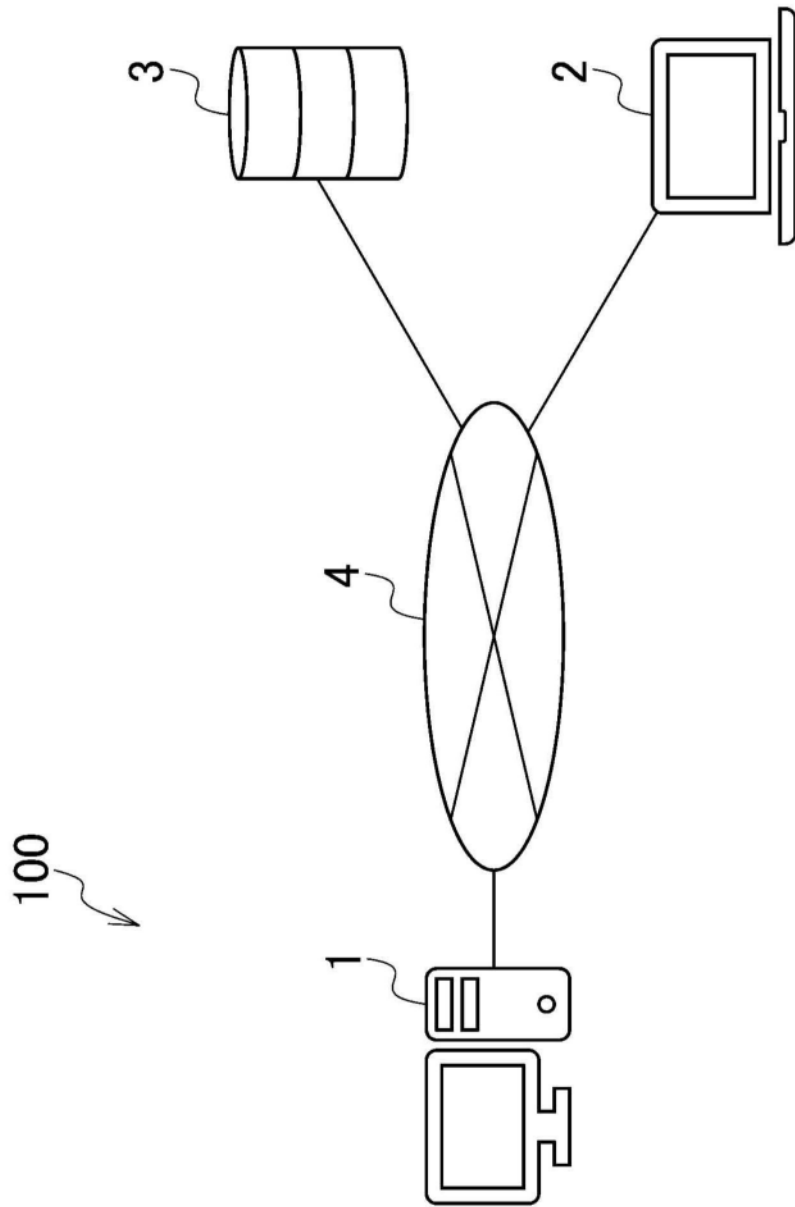
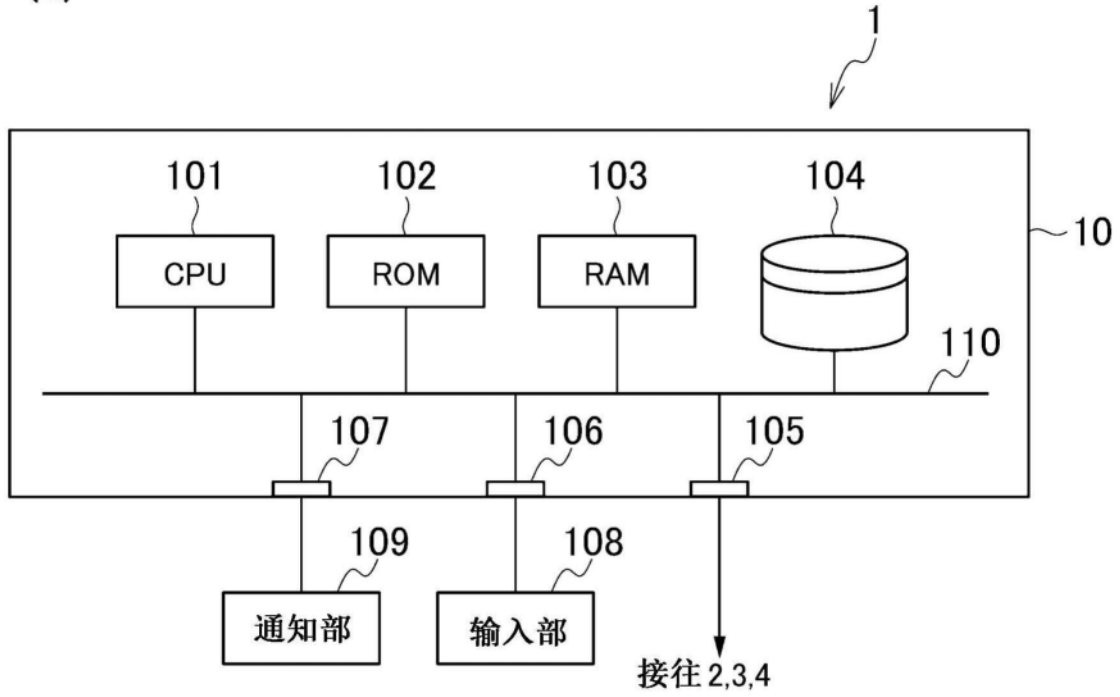


图1

(a)



(b)

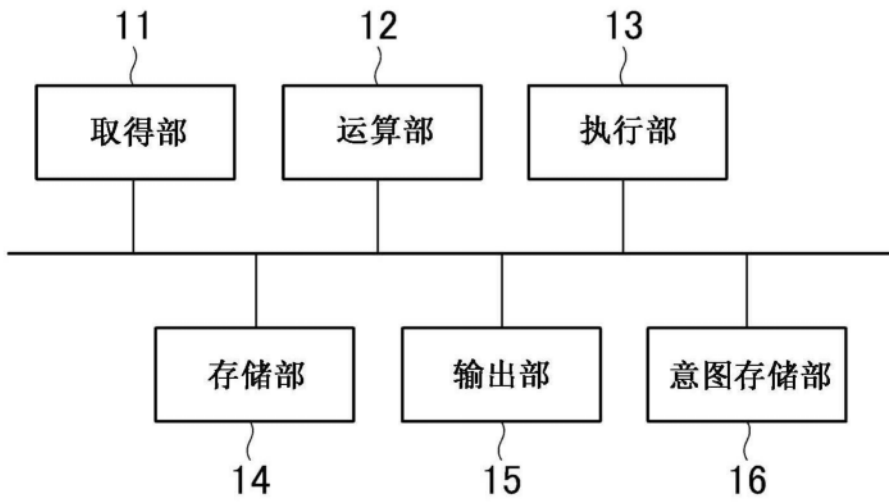
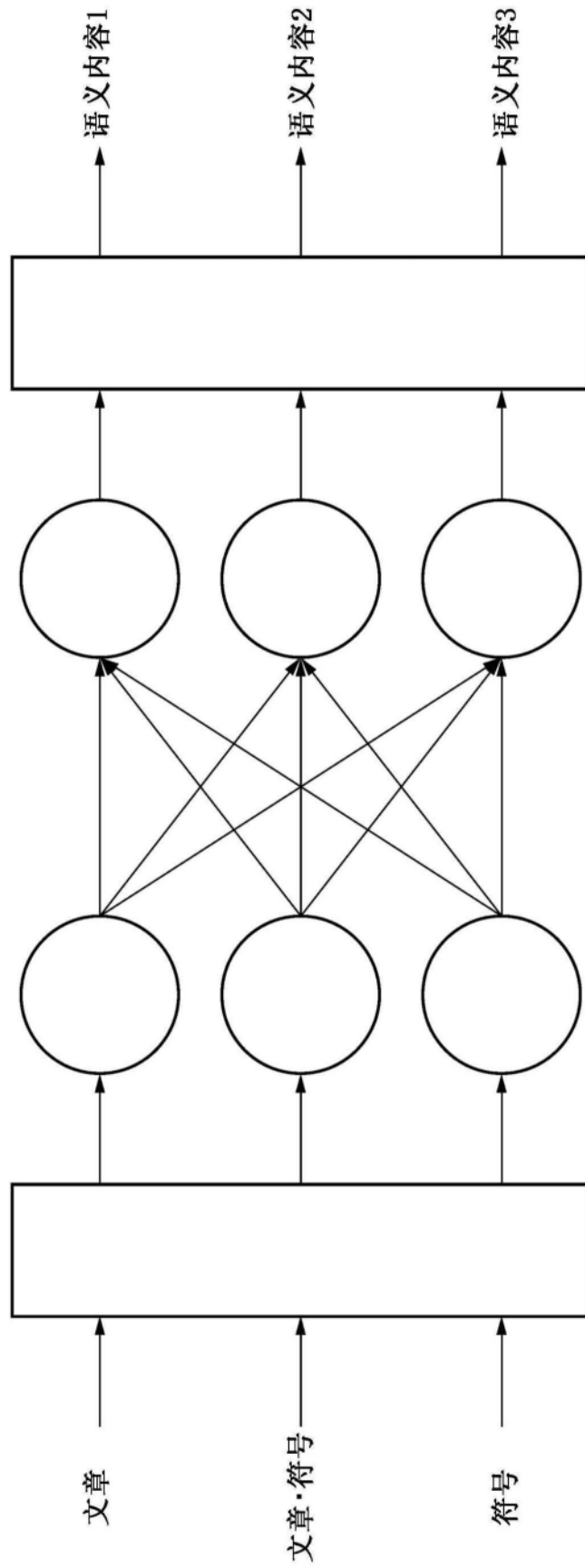
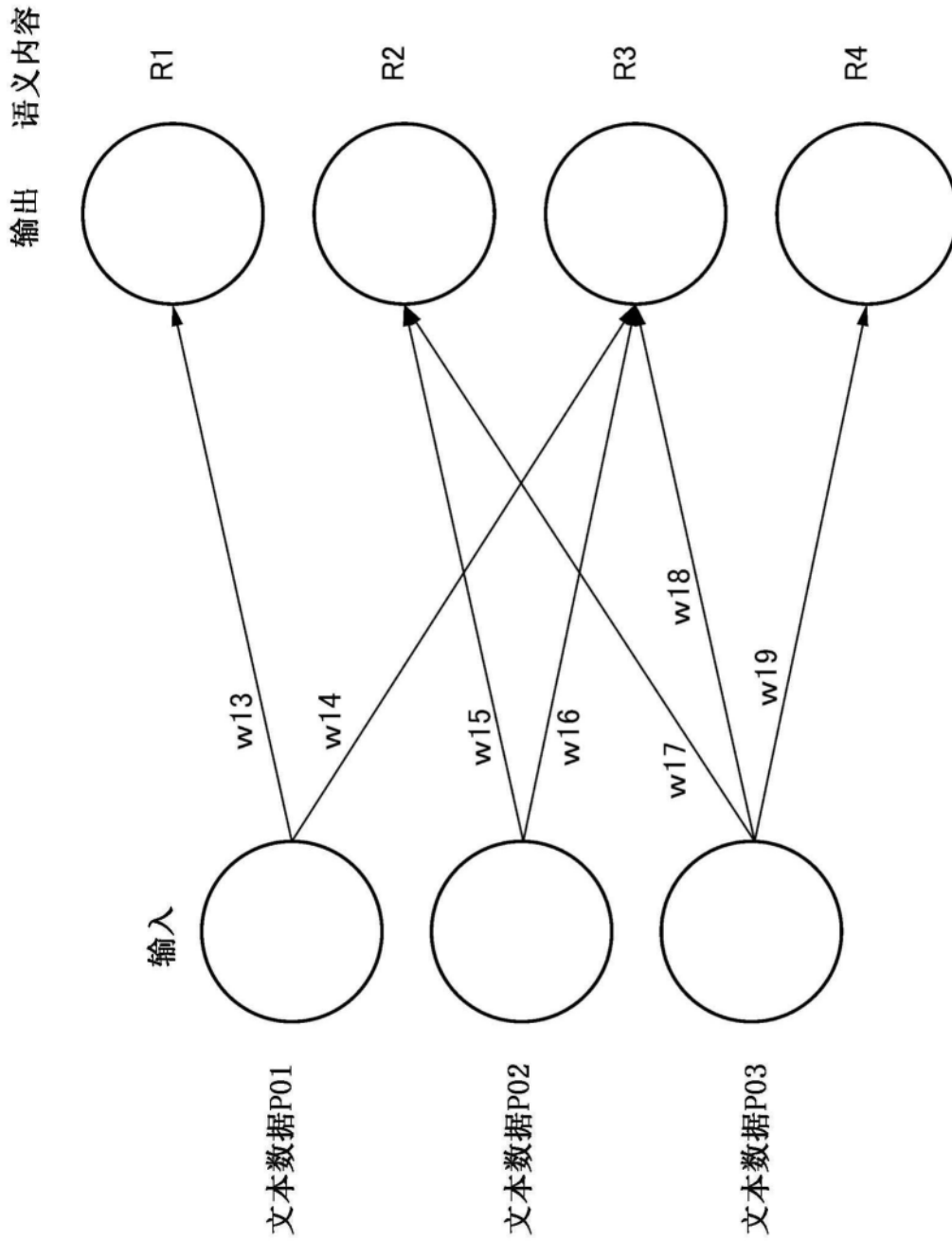


图2



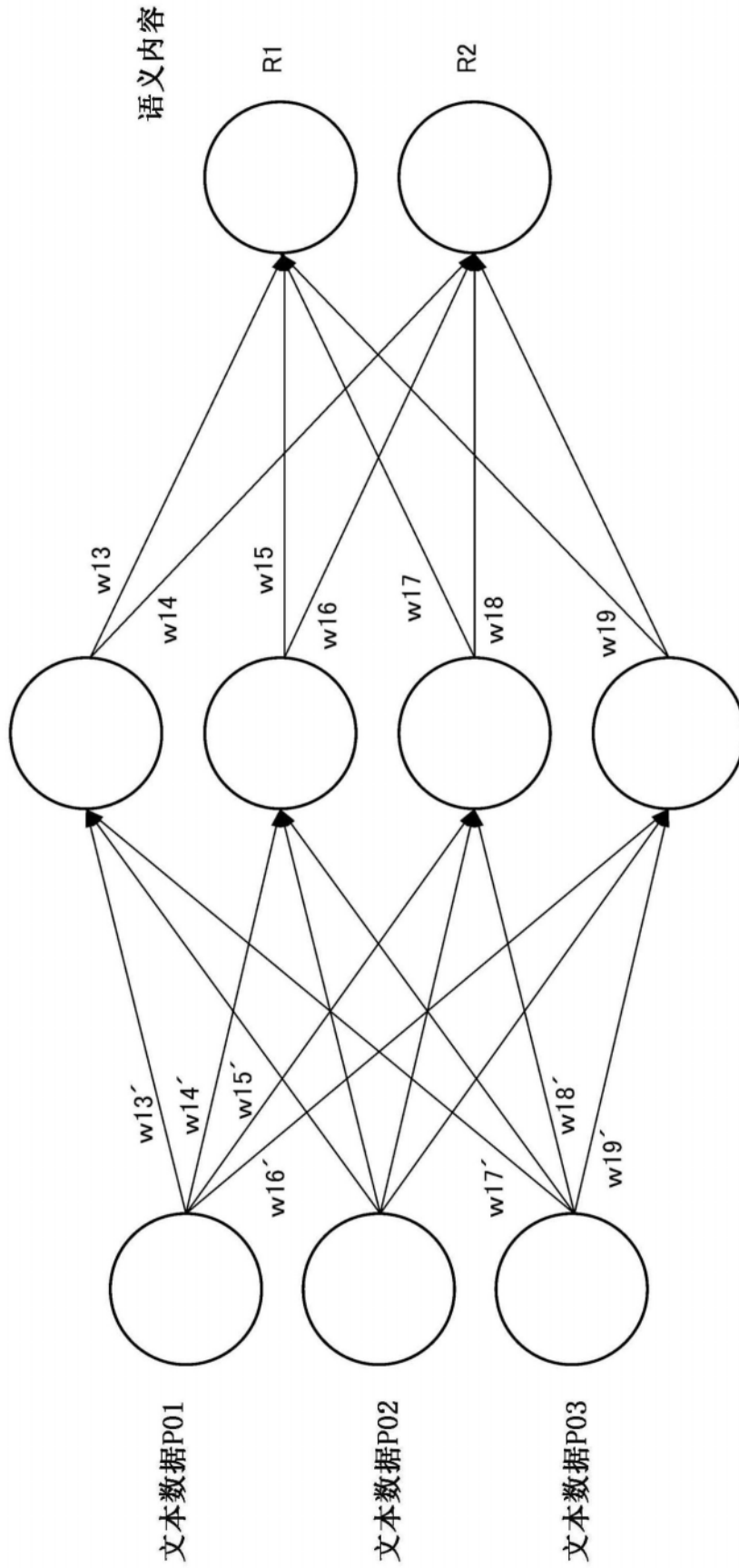
第1学习完毕模型

图3



第1学习完毕模型

图4



第1学习完毕模型

图5

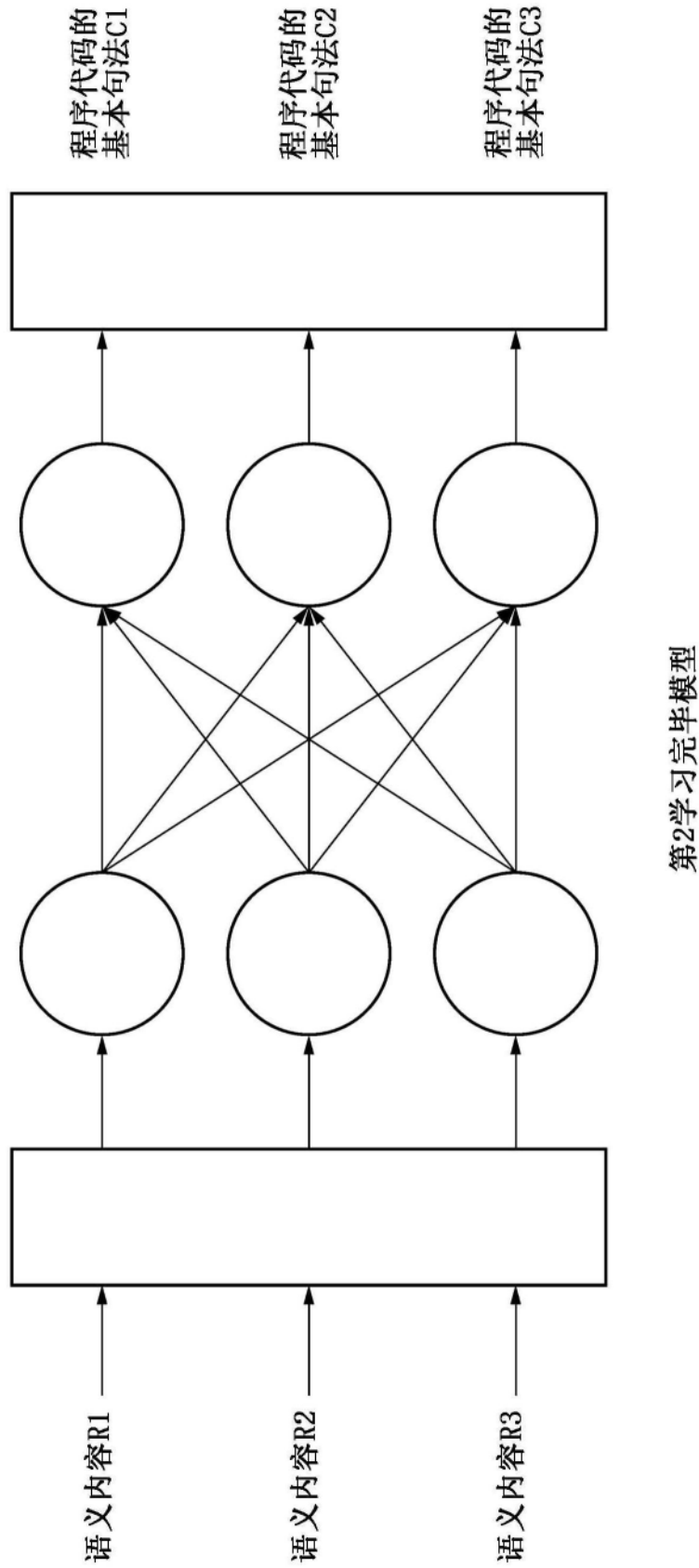


图6

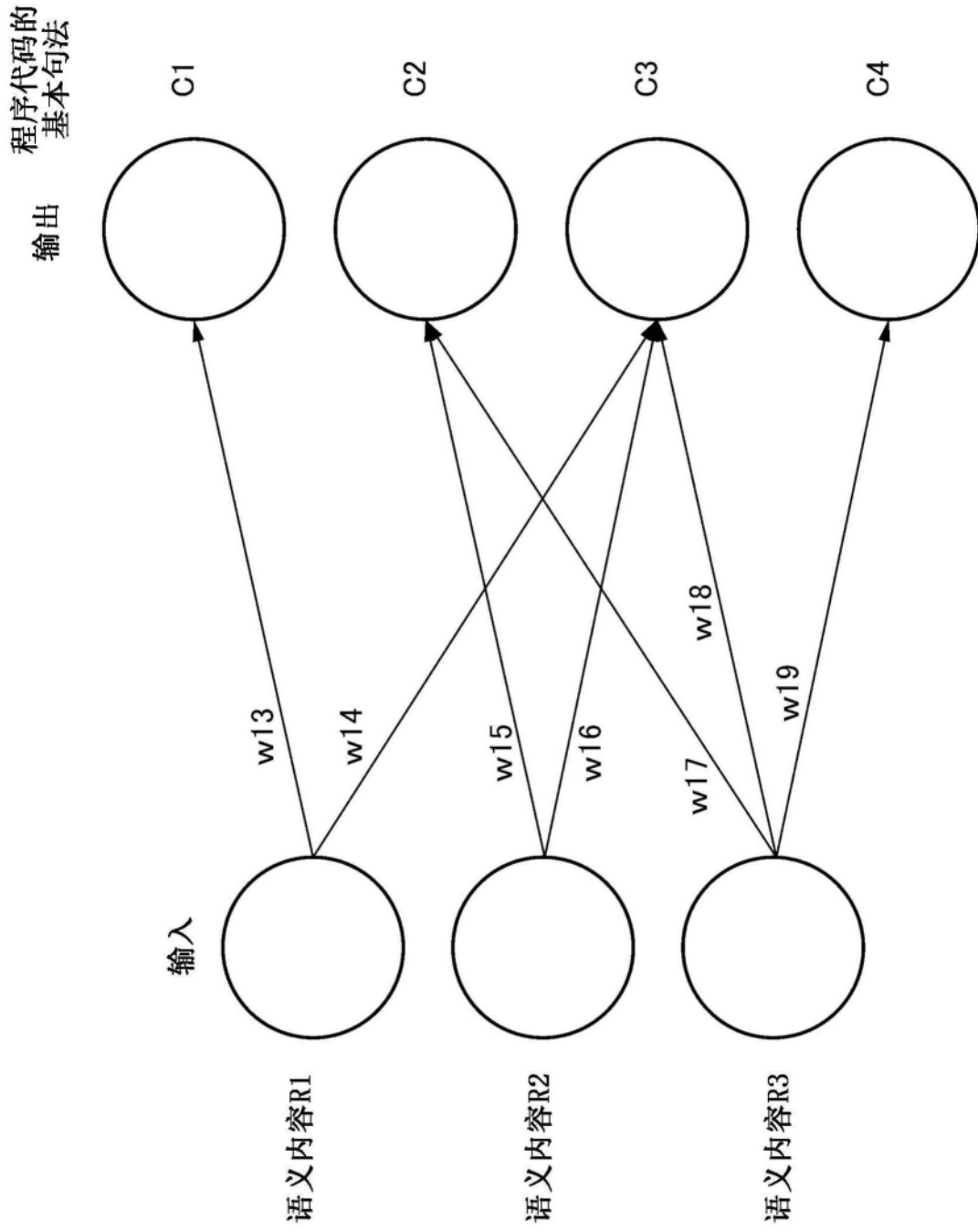


图7

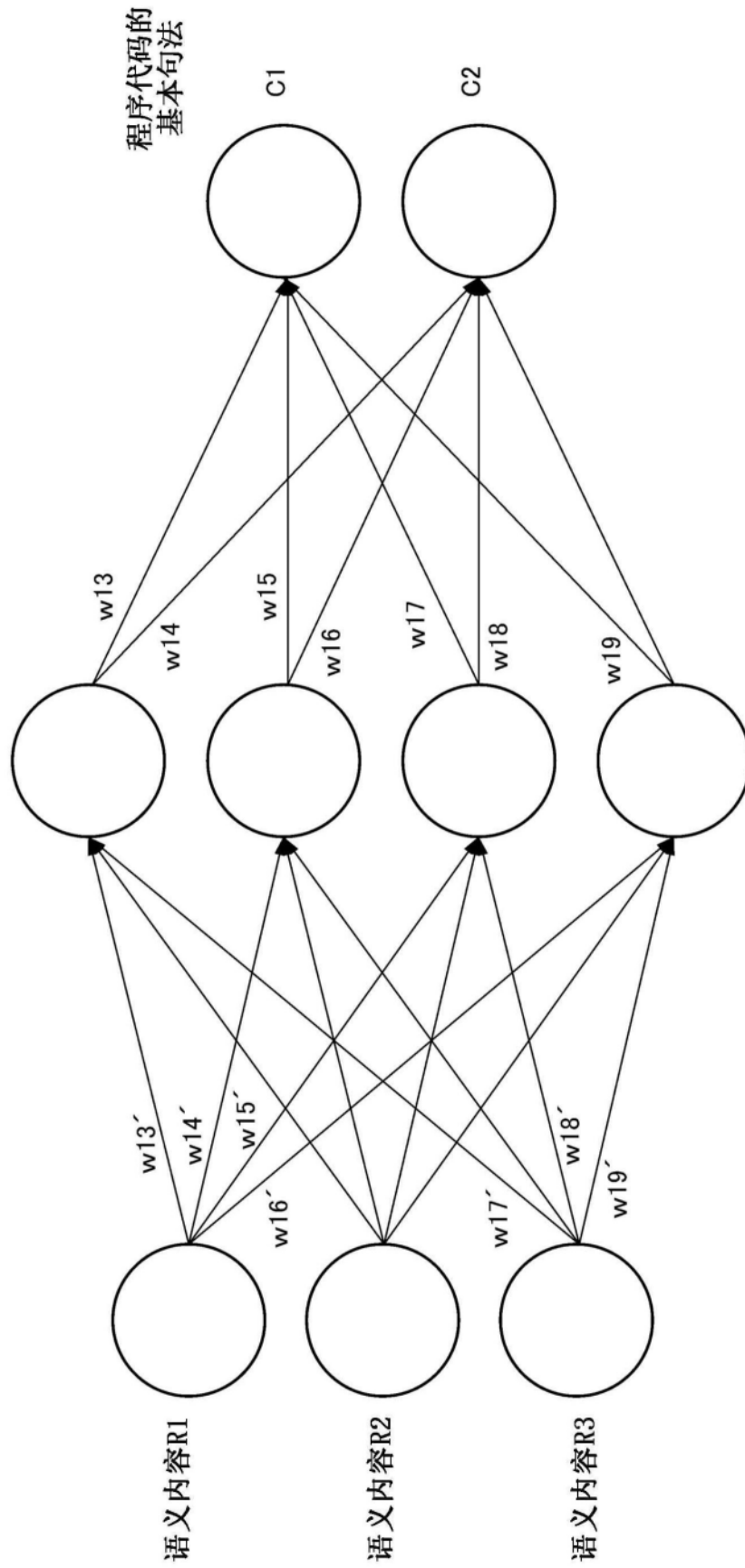


图8

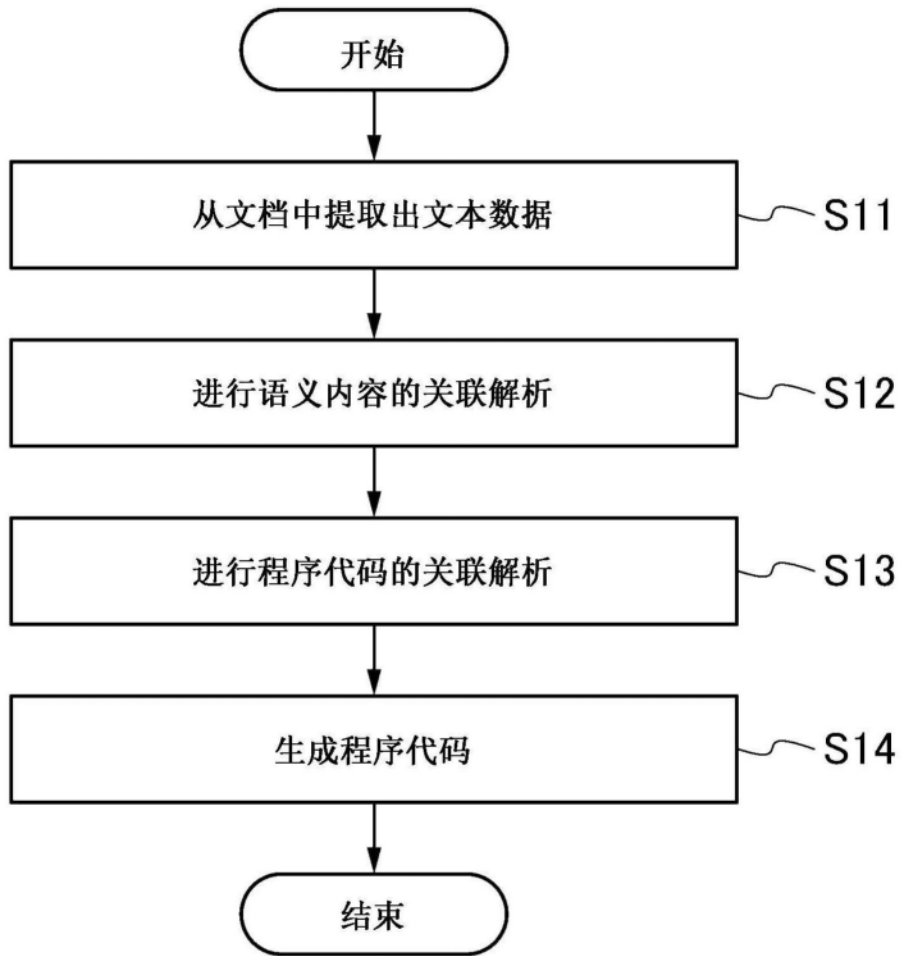


图9

第1相关性

文本数据P01	语义内容R1
文本数据P02	语义内容R2
文本数据P03	语义内容R3
文本数据P04	语义内容R4
文本数据P05	语义内容R5
文本数据P06	语义内容R6
文本数据P07	语义内容R7
.....
.....

第1相关性

文本数据P01	语义内容R1
文本数据P01	语义内容R2
文本数据P01	语义内容R3
文本数据P02	语义内容R4
文本数据P03	语义内容R4
.....
.....

第2相关性

语义内容R1	程序代码的基本句法C1
语义内容R2	程序代码的基本句法C2
语义内容R3	程序代码的基本句法C3
语义内容R4	程序代码的基本句法C4
语义内容R5	程序代码的基本句法C5
语义内容R6	程序代码的基本句法C6
语义内容R7	程序代码的基本句法C7
.....
.....

第2相关性

语义内容R1	程序代码的基本句法C1
语义内容R1	程序代码的基本句法C2
语义内容R1	程序代码的基本句法C3
语义内容R2	程序代码的基本句法C4
语义内容R3	程序代码的基本句法C4
.....
.....

(a)

(b)

图10

