

[12] 发明专利申请公开说明书

[21] 申请号 99813093.1

[43] 公开日 2001 年 12 月 5 日

[11] 公开号 CN 1325527A

[22] 申请日 1999.9.8 [21] 申请号 99813093.1

[30] 优先权

[32] 1998.9.9 [33] US [31] 09/150,459

[86] 国际申请 PCT/US99/20445 1999.9.8

[87] 国际公布 WO00/14727 英 2000.3.16

[85] 进入国家阶段日期 2001.5.9

[71] 申请人 单一声音技术公司

地址 美国加利福尼亚州

[72] 发明人 D·C·韦博

[74] 专利代理机构 中国专利代理(香港)有限公司

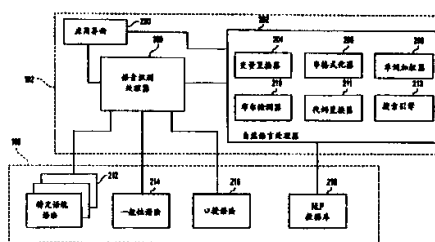
代理人 王 勇 陈景峻

权利要求书 3 页 说明书 13 页 附图页数 7 页

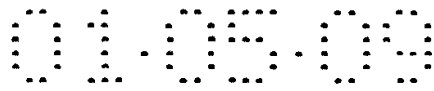
[54] 发明名称 采用语音识别和自然语言处理的交互式用户界面

[57] 摘要

一种用于用话语、语音处理和自然语言处理与计算机交互的系统和方法。该系统包含一个语音处理器,用于在第一语法文件检索话语的匹配短语,并用于如果在第一语法文件找不到匹配短语时,就在第二语法文件检索匹配短语。该系统也包含一个自然语言处理器,用于在一个数据库检索匹配短语的匹配条目;以及一个应用界面,用于如果在数据库找到匹配条目时,就执行一个与匹配条目相关的行动。该系统采用特定语境语法,由此增强语音识别和自然语言处理的效率。该系统另外还能自适应地以交互方式“学习”单词和短语以及它们的相关意思。



ISSN 1008-4274



权 利 要 求 书

1. 一种用话语与计算机交互的方法，该方法包含下列步骤：
在第一语法文件检索所述话语的匹配短语；
如果在所述第一语法文件找不到所述匹配短语，就在第二语法文件检索所述匹配短语；
5 在一个数据库检索所述匹配短语的匹配条目；和
如果在所述数据库找到所述匹配条目，就执行一个与所述匹配条目相关的行动。
2. 权利要求 1 的方法，进一步包含的步骤是，如果在所述数据库找不到所述匹配条目，就用所述匹配短语至少更新所述数据库、所述第一语法文件和所述第二语法文件的其中之一。
3. 权利要求 2 的方法，其中，所述第一语法文件是一个特定语境语法文件。
4. 权利要求 3 的方法，其中，所述第二语法文件是一个一般性
15 语法文件。
5. 权利要求 4 的方法，进一步包括的步骤是，如果在所述一般性语法文件中找不到所述匹配短语，就在一个口授语法中搜索所述匹配短语。
6. 权利要求 1 的方法，进一步包含的步骤是，在所述搜索所述
20 数据库的步骤之前替换所述匹配短语中的至少一个单词。
7. 权利要求 6 的方法，其中，替换所述至少一个单词的步骤包含用一个通配符替换所述至少一个单词。
8. 权利要求 6 的方法，其中，替换所述至少一个单词的步骤包含用一个专用名称替换所述至少一个单词。
- 25 9. 权利要求 1 的方法，进一步包含的步骤是，在所述搜索所述数据库的步骤之前对所述匹配短语进行文本格式化的步骤。
10. 权利要求 1 的方法，进一步包括的步骤是，在所述搜索所述数据库的步骤之前按各个单词的相对重要性加权所述匹配短语中的所述各个单词。
- 30 11. 权利要求 1 的方法，进一步包括的步骤是，为所述匹配条目生成一个可信度值。
12. 权利要求 11 的方法，进一步包括的步骤是，将所述可信度

值与一个阈值比较。

13. 权利要求 12 的方法，进一步包括的步骤是，确定所述匹配条目中是否存在预定数量的匹配短语中的单词。

14. 权利要求 13 的方法，进一步包括的步骤是，如果所述匹配条目中不存在所述预定数量的所述匹配短语中的单词，则询问用户是否所述匹配条目是所述话语的正确解释。

15. 权利要求 14 的方法，进一步包括的步骤是，如果所述匹配条目不是所述话语的正确解释，就提示用户提供另外的信息。

16. 权利要求 15 的方法，进一步包括的步骤是，用所述另外的信息更新所述数据库、所述第一语法文件和所述第二语法文件的至少其中之一。

17. 权利要求 2 的方法，进一步包括的步骤是，在一个与所述匹配条目相关联的特定语境语法检索随后一个话语的随后一个匹配短语。

18. 一种用话语与计算机交互的系统，该系统包括：

一个语音处理器，用于在第一语法文件检索所述话语的匹配短语，并用于如果在所述第一语法文件找不到所述匹配短语时，就在第二语法文件检索所述匹配短语；

一个自然语言处理器，用于在一个数据库检索所述匹配短语的匹配条目；和

一个应用界面，用于如果在所述数据库找到所述匹配条目时，就执行一个与所述匹配条目相关的行动。

19. 权利要求 18 的系统，其中，如果在所述数据库找不到所述匹配条目，所述自然语言处理器就用所述匹配短语至少更新所述数据库、所述第一语法文件和所述第二语法文件的其中之一。

20. 权利要求 19 的系统，其中，所述第一语法文件是一个特定语境语法文件。

21. 权利要求 20 的系统，其中，所述第二语法文件是一个一般性语法文件。

22. 权利要求 21 的系统，其中，如果在所述一般性语法文件中找不到所述匹配短语，所述语音处理器就在一个口授语法中搜索所述匹配短语。

23. 权利要求 18 的系统, 其中, 所述自然语言处理器在搜索所述数据库之前替换所述匹配短语中的至少一个单词。

24. 权利要求 23 的系统, 在所述自然语言处理器中进一步包含一个变量置换器, 用于用一个通配符替换所述匹配短语中的至少一个
5 单词。

25. 权利要求 23 的系统, 在所述自然语言处理器中进一步包含一个代词置换器, 用于用一个专有名称替换所述匹配短语中的至少一个单词。

26. 权利要求 18 的系统, 进一步包含一个串格式化器, 用于在
10 搜索所述数据库之前对所述匹配短语进行文本格式化。

27. 权利要求 18 的系统, 进一步包括一个单词加权器, 用于在搜索所述数据库之前按各个单词的相对重要性加权所述匹配短语中的所述各个单词。

28. 权利要求 18 的系统, 在所述自然语言处理器中进一步包含
15 一个搜索引擎, 用于为所述匹配条目生成一个可信度值。

29. 权利要求 28 的系统, 其中, 所述自然语言处理器将所述可信度值与一个阈值比较。

30. 权利要求 29 的系统, 进一步包含一个布尔检测器, 用于确定所述匹配条目中是否存在预定数量的匹配短语中的单词。

20 31. 权利要求 30 的系统, 其中, 如果所述匹配条目中不存在所述预定数量的所述匹配短语中的单词, 所述自然语言处理器询问用户所述匹配条目是否是所述话语的正确解释。

32. 权利要求 30 的系统, 其中, 如果所述匹配条目不是所述话语的正确解释, 所述自然语言处理器提示用户提供另外的信息。

25 33. 权利要求 32 的系统, 其中, 所述自然语言处理器用所述另外的信息更新所述数据库、所述第一语法文件和所述第二语法文件的至少其中之一。

30 34. 权利要求 19 的系统, 其中, 所述语音处理器在一个与所述匹配条目相关联的特定语境语法检索随后一个话语的随后一个匹配短语。

说 明 书

采用语音识别和自然语言处理的交互式用户界面

5 本发明涉及用于计算机用户界面的语音识别。更具体来说，本发明涉及一种用语音识别和自然语言处理与计算机进行用户界面交互的新颖方法和系统。

10 随着计算机在社会中变得更加盛行，明显有许多人在掌握计算机知识和与计算机交流等方面困难很大。用户要操作计算机，就必须经常学习各种已不通用的命令和非直观的过程。例如，大多数计算机都使用基本上是菜单驱动的基于视窗的操作系统。这要求用户学习什么菜单命令或命令序列会产生预定结果。

15 此外，传统的与计算机的交互的速度经常因诸如键盘或鼠标的手工输入设备而减慢。许多计算机用户都不是打字快手。结果，许多时间都花费在通过这些手工输入设备向计算机发送命令和单词上。已经明显需要有一种更容易、更快速和更直观的与计算机沟通的方法。

20 有人提出的计算机交互的方法是语音识别。语音识别涉及用软件和硬件协作监测听得见的人类语音并把所监测的语音转换成单词串。如现有技术中已知的那样，语音识别的工作方式是把硬件监测到的语音分解成不可分的更小单位的声音—称作音素(phonemes)。音素是不同的声音单元。例如，单词“those”由三个音素组成；第一个音素是“th”音，第二个是“o”音，第三个是“s”音。语音识别软件试图将所监测的音素与一个存储的字典中的已知单词比较。美国专利 4,783,803 号中给出了一个语音识别系统的例子，该专利名称为“Speech Recognition Apparatus and Method”(语音识别装置和方法) (1998年11月8日授权，受让给 Dragon Systems 公司)，这里引用作为参考。目前，市场上有许多语音识别软件包，出自 Dragon Systems 公司和 IBM 公司这样的公司。

30 这些语音识别软件包的一个缺陷是，它们通常只执行命令和控制功能或者口授指令功能。所以，用户仍然需要学习命令语汇才能操作计算机。

有人提出的这些语音识别系统的增强系统，是用自然语言处理系统处理所监测的单词。自然语言处理一般涉及通过分析所监测的单词

的语法关系和相对语境，确定它们的概念上的“意思”（例如说话人想要传达什么意思）。例如美国 4,887,212 号专利叙述了一种用单词隔离、语形学分析 (morphological & analysis)、字典查找和语法分析来分析一个输入单词流的方法。该专利名称为 “Parser for
5 Natural Language Text”（自然语言文本分析器）（1989 年 12 月 12 日授权，受让给 IBM 公司），这里引用作为参考。

与语音识别一起采用的自然语言处理提供一种用口语而不是诸如键盘或鼠标的手工输入操作计算机的强大工具。然而，常规的自然语言处理系统的一个缺点是，它可能无法确定由语音识别系统所监测
10 的单词的准确“意思”。在这样的情况下，用户通常需要重新组合或重新陈述短语 (phrase)，希望自然语言处理系统将能在后面的尝试中确定准确“意思”。显然，这会导致严重的延迟，因为用户需要重新陈述整个句子或命令。

常规系统的另一个缺点是，语音识别所需的时间可能会过分长。
15 这主要由于与大量要处理的信息相比，处理资源的速度有限。例如，在许多常规语音识别程序中，识别话语 (utterance) 所需的时间很长，因为所搜索的字典文件很大。

常规语音识别和自然语言处理系统的再一个缺点是，它们不是交互式的。结果，就不可能与计算机进行对话方式的交互。用户需要向
20 计算机传达完整的概念。用户不能说句子片断，因为这些（与前面的话语有关的）句子片断的意思将丢失。

采用语音识别和自然语言处理的计算机所需要的是避免上述缺点的交互式用户界面。

本发明是一种用于与采用话语、语音处理和自然语言处理的计算机进行交互的新颖和改进的系统和方法。一般来说，该系统包含一个
25 语音处理器 (speech processor)，用于在一个第一语法文件中搜索话语的一个匹配短语，并用于如果在第一语法文件中找不到匹配短语时，在一个第二语法文件中搜索匹配短语。该系统也包括一个用于在一个数据库中搜索匹配短语的对应匹配条目 (matching entry) 的自然语言处理器，以及一个如果在数据库中找到匹配条目就执行一个
30 与匹配条目相关联的行动 (action) 的应用界面 (application interface)。

在最佳实施例中，如果在数据库中找到匹配条目，自然语言处理器就用匹配短语更新数据库、第一语法文件和第二语法文件的至少其中之一。

第一语法文件是一个特定语境语法文件。特定语境语法文件是一种含有与特定主题高度相关的单词和短语的语法文件。第二语法文件是一个一般性语法文件。一般性语法文件是一种含有不必按语境解释的单词和短语的语法文件。就是说，一般性语法文件中的单词和短语不属于任何母体语境 (parent context)。本发明在搜索一般性语法文件之前搜索特定语境语法文件，从而允许用户用更口语化的方式与计算机交流，其中，所说的单词如果在特定语境语法文件中被找到，就被参照最近讨论过的主题进行解释。

在本发明的另外一个方面中，如果在一般性语法文件中找不到匹配短语，语音处理器就在一个口授语法 (dictation grammar) 中搜索匹配短语。口授语法是个一般性单词和短语的大单词表。通过首先搜索特定语境语法和一般性语法，预计将大大缩短语音识别时间 - 因为特定语境语法和一般性语法是物理上比口授语法更小的文件。

在本发明的另外一个方面中，自然语言处理器在搜索数据库之前替换匹配短语中的至少一个单词。这可由自然语言处理器中的用于以通配符替换匹配短语中的至少一个单词的变量置换器 (variable replacer) 完成。通过以通配符替换短语中的某些单词 (称作“单词变量”)，可以大大地减少数据库中条目的数量。此外，自然语言处理器中的代词替换器 (pronoun substituter) 可以以适当名称替换匹配短语中的代词，便于特定于用户的事实在数据库中的存储。

在本发明的另外一个方面中，一个串格式化器 (string formatter) 在搜索数据库之前对匹配短语进行文本格式化。一个单词加权器 (word weighter) 也在搜索数据库之前按各个单词的相对重要性加权匹配短语中的各个单词。这些步骤便于更快、更准确地搜索数据库。

自然语言处理器中的一个搜索引擎为匹配条目生成一个可信度值 (confidence value)。自然语言处理器将该可信度值与一个阈值比较。一个布尔检测器 (boolean tester) 确定匹配条目中是否存在所需数量的匹配短语中的单词。该布尔检测的作用是验证由搜索引擎

返回的结果。

为了消除歧义，自然语言处理器在匹配条目中不存在所需数量的匹配短语中的单词时，提示用户匹配条目是否是话语的正确解释。如果匹配条目不是话语的正确解释，自然语言处理器也提示用户提供另外的信息。用该另外的信息更新数据库、第一语法文件和第二语法文件的至少其中之一。这样，本发明适应性地“学习”更多的话语的意思，由此增强用户界面的效率。

最后，语音处理器将启用并检索一个与匹配条目相关联的特定语境语法，用于其后一个话语的其后一个匹配短语。这保证首先将检索到最相关的单词和短语，由此减少语音识别的次数。

根据以下结合附图所作的详细说明，本发明的特点、目的和优点将变得更加明显。附图中相同的标注符自始至终标识相对应的对象，其中：

图 1 是一个用于本发明的示意性计算机系统的功能框图；

图 2 是本发明的图 1 的计算机系统的 CPU102 和存储介质 108 的扩展的功能框图；

图 3A - 3D 是本发明的方法的流程图；

图 4 是本发明的示意性自然语言处理 (NLP) 数据库的所选择的列的示意图。

现在将参照图 1 的典型计算机系统 100 的功能框图叙述本发明。图 1 中，计算机系统 100 包含一个中央处理单元 (CPU) 102。CPU 102 可以是本领域中已知的、经过适当编程而能执行本文结合图 3A - 3D 所述方法的任何通用微处理器或微控制器。例如，CPU 102 可以是一个诸如由英特尔公司制造的奔腾 II 处理器之类的普通微处理器。

CPU 102 与包括显示器 104、手工输入装置 106、存储介质 108、麦克风 110、扬声器 112 和数据输入端口 114 在内的多个外围设备通信。显示器 104 可以是个直观显示设备 - 如 CRT、LCD 屏幕、触感屏幕，或者是本领域中已知的用于直观地向用户显示图象和文字的其它监视器。手工输入装置 106 可以是一个普通键盘、按键盘、鼠标、跟踪球或者本领域中已知的用于手工输入数据的其它输入设备。存储介质 108 可以是个普通的读/写存储器 - 如磁盘驱动器、软盘驱动器、CDROM 驱动器、硅存储器或本领域中已知的用于存储和检索数据的其

它存储设备。重要的是，存储介质 108 可以远离 CPU 102，通过诸如局域网（LAN）或宽域网（WAN）的网络或者因特网与 CPU 102 相连。麦克风 110 可以是本领域中已知的任何适合用于向 CPU 102 提供音频信号的麦克风。扬声器 112 可以是本领域中已知的任何适合用于重现来自 CPU 102 的音频信号的扬声器。不言自明，麦克风 110 和扬声器 112 可以包含相应的数-模和模-数转换电路。数据输入端口 114 可以是本领域中已知的任何用于用诸如 RS-232、通用串行总线之类的数据协议与外部附件连接的数据端口。

所以，图 1 展示了计算机系统 100 的各功能单元。计算机系统 100 的每个单元可以是如上所述的合适的成品部件。本发明提供一种用于用语音与计算机系统 100 接口的方法和系统。

现在转至图 2，图 2 表示 CPU 102 和存储介质 108 的一个扩展的功能框图。CPU 102 中包括语音识别处理器 200、自然语言处理器 202 和应用界面 220。自然语言处理器 202 进一步包含变量置换器 204、串格式化器 206、单词加权器 208、布尔检测器 210、代词置换器 211 和搜索引擎 213。存储介质 108 中含有多个特定语境语法文件 212、一般性语法文件 214、口授语法 216 和自然语言处理器（NLP）数据库 218。在最佳实施例中，语法文件 212、214 和 216 是巴科斯-诺尔（Bakus-Naur）形式（BNF）的文件，描述由用户说出的语言的结构。BNF 文件是本领域中众所周知的用于描述语言结构的文件，因此这里将不讨论 BNF 文件的详细内容。BNF 文件的一个优点是，可以用分层的树型结构来描述短语或单词序列，而无需显式地列举这些单词序列的全部组合。所以，BNF 文件在最佳实施例中的使用，使存储介质 108 中的语法文件 212、214 和 216 的物理大小最小化，提高了下文所述方式启用和搜索这些文件的速度。不过，在另外的实施例中可以使用其它文件结构。

现在将结合图 3A-3D 的流程图，说明图 2 的这些功能单元的操作和交互。在图 3A 中，流程始于功能框 300，向语音处理器 200 提供一个话语。话语是有开头和结尾的一系列的语音，可包含一个或多个口语单词。捕获口语单词的麦克风 110 可以执行功能框 300 的步骤。或者，可以经过输入端口 144 或者从存储介质 108 向语音处理器 200 提供话语。话语的格式最好是数字格式，如有名的“.wav”音频文件

格式。

5 流程前进到判断框 302，语音处理器 200 在此判断是否特定语境语法 212 的其中之一已经被启用。在最佳实施例中，特定语境语法 212 是 BNF 文件，含有与某个母体语境有关的单词和短语。一般来说，语境就是一个主题领域。例如，在本发明的一个可应用于个人电脑的实施例中，语境的例子可以是“新闻”、“天气”或“股票”。在这种情况下，特定语境语法 212 就各自含有对应于这些语境中的不同语境的命令、控制单词、描述符(descriptor)、限定符(qualifier)或参数。语境的使用提供一种表示各种信息类型的分层结构。下文将结合
10 NLP 数据库 218 进一步对各语境及其应用作出说明。

如果已经有特定语境语法 212 被启用，就搜索该特定语境语法 212，查找在功能框 300 所提供话语的匹配项(match)。然而，如果没有特定语境语法 212 被启用，流程就前进到功能框 308，在此启用一般性语法 214。

15 在最佳实施例中，一般性语法 214 是一个 BNF 文件，文件中含有的单词和短语本身并不属于某个母体语境，但是可能有某特定语境语法文件 212 为此存在的一个相关语境。换言之，一般性语法 214 中的单词和短语可能位于分层语境结构的根部。例如，在一个可应用于个人电脑的实施例中，一般性语法 214 就含有命令和控制短语。

20 在功能框 310 中，搜索一般性语法 214，查找在功能框 300 所提供话语的匹配单词或短语。如果找不到匹配项，就启用口授语法 216。在最佳实施例中，口授语法 216 是一个 BNF 文件，文件中含有的单词列表本身既无某个母体语境也无相关语境。例如，在一个可应用于个人电脑的实施例中，口授语法 216 就含有一个类似于通用词典的较大的
25 的通用单词列表。

在功能框 316 中，搜索口授语法，查找在功能框 300 所提供话语的每个单词的匹配单词。如果找不到匹配单词，可以选择在显示器 104 显示一个可视的出错讯息或者选择通过扬声器 112 重播一个有声的出错讯息，然后过程结束，直到在功能框 300 有另一个话语被提供给语
30 音处理器 200。

所以，如上述说明可见，当某个话语被提供给语音处理器 200 时，首先搜索启用的特定语境语法 212 (如果有的话)。如果所启用的特

定语境语法 212 中没有匹配项，则启用并搜索一般性语法 214。如果一般性语法 214 中没有匹配项，则启用并搜索口授语法 216。

5 在最佳实施例中，当语音识别处理器 200 在搜索特定语境语法 212 或一般性语法 214 时，称其处于“命令和控制”方式。在该方式中，语音识别处理器 200 把整个话语作为整体与语法中的条目作比较。相比之下，当语音识别处理器 200 在搜索口授语法时，称其处于“口授”方式。在该方式中，语音识别处理器 200 每次一个单词地把话语与口授语法 216 中的条目作比较。可以预计，在命令和控制方式中检索整个一个话语的匹配项的速度，一般会比在口授方式中每次搜索一个单词的速度快。

10 进一步可以预计，任何单个的特定语境语法 212 在规模上都小于一般性语法 214（即总的单词和短语更少），后者在规模上又小于口授语法 216。所以，通过首先搜索任何启用的特定语境语法 212 的方法，有可能更快地找到匹配项（如果有的话），这至少是部分地因为文件更小。同样，通过在搜索口授语法 216 之前搜索一般性语法 214 的方法，有可能更快地找到匹配项。

20 此外，正如下文对本发明能自适应地对特定语境语法 212 和一般性语法 214 添加的能力进一步解释的那样，它们将含有最常用的话语。因此可以预计，在特定语境语法 212 或一般性语法 214 中找到匹配项的速度更有可能比在口授语法 216 中的快。

最后，正如下文将进一步解释的那样，在启用的特定语境语法 212 中的单词和短语，是用户更有可能说出的，因为它们是与用户最近说到的主题高度相关的单词。这也使用户能用语句片断以更口语化的风格说话，其单词的意思被按照所启用的特定语境语法 212 解释。

25 通过按上述顺序进行搜索，本发明的搜索的效率，高于在含有所有预期单词和短语的一个大型列表中每次一个条目地进行搜索的效率。

30 返回到判断框 306，如果在特定语境语法 212 中找到匹配项，则有图 3A 中所显示的三种可能的下一步步骤。对于所启用的特定语境语法 212 中的每个匹配条目，可能都有一个相关的要由语音识别处理器 200 采取的行动。功能框 322 表示，一个行动可能是指示应用界面 220 对某个独立软件应用或实体采取某种行动。例如，应用界面 220

可以用微软公司的语音应用编程接口 (SAPI) 标准与某外部应用通信。可以指示该外部应用例如访问特定的因特网网站 URL 或者通过正文向语音转换的方式说出特定的短语。可以采取如下文将结合图 4 的 NLP 数据库 218 所进一步讨论的那样的其它行动。

- 5 功能框 324 表示, 另一个行动可以是直接访问自然语言处理 (NLP) 数据库 218 中的某一行, 由此绕过下文进一步作说明的自然语言处理步骤。功能框 326 表示, 另一个行动可以是把用于所启用语境的单词或短语前置 (prepend) 到在功能框 306 在特定语境语法中找到的匹配单词或短语。例如, 如果所启用的语境是 “movies” (电影),
- 10 匹配的话语是 “8 o'clock” (8 点钟), 则把单词 “movies” 前置到短语 “8 o'clock”, 组成短语 “movies at 8 o'clock” (8 点钟的电影)。

同样, 如果在一般性语法 214 中找到匹配项, 流程就前进到功能框 322 - 在此指示应用界面 220 采取一个如上所述的行动, 或者前进

15 到功能框 324, 直接访问 NLP 数据库中的某一行。然而, 如果在一般性语法 214 中找到匹配项, 则不进行对语境的前置, 因为如上所述, 一般性语法 214 中的条目本身并没有母体语境。

最后参看图 3A, 可以通过手工输入装置 106 输入单词。在这种情况下, 不需要进行语音识别, 不过仍然需要对所输入单词进行自然语

20 言处理。然后, 流程前进到图 3B。

图 3B 中, 在功能框 328, 自然语言处理器 202 为进行自然语言处理分析而对短语进行格式化。这种格式化是用串格式化器 206 完成的, 可包括去除单词之间的多余空格、把所有字母变成小写 (或大写)、把缩略部分完全写出 (例如把 “it's” 变成 “it is”) 之类的

25 文字处理。这个格式化步骤的目的是对短语作分析的准备。

流程继续到功能框 330, 在此由变量置换器 204 用相关的通配符函数 (wildcard function) 置换某些 “单词变量”, 为访问 NLP 数据库 218 作准备。这里所用的术语 “单词变量” 指的是表示数量、日期、时间、货币之类的单词或短语。例如, 在一个实施例中, 短语 “what

30 movies are playing at 8 o'clock” (8 点钟在放什么电影) 要在功能框 330 被转换成 “what movies are playing at \$time” (在 \$time 时在放什么电影), 其中 “\$time” 是一个用来代表任一时间值

的通配符函数。作为另一个例子，在一个实施例中，短语“sell IBM stock at 100 dollars”（按 100 美元的价格卖出 IBM 股票）要在功能框 330 被转换成“sell IBM stock at \$dollars”（按 \$dollar 的价格卖出 IBM 股票），其中“\$dollars”是一个用来代表任一美元值的通配符函数。这个步骤可以由一个简单循环来完成，该循环搜索短语中诸如单词“dollar”或“o'clock”的关键记号并用规定的通配符函数置换这些单词变量。为了跟踪在短语中作过替换的位置，可以使用一个数组。这样能在搜索过 NLP 数据库 218 后将原始的单词变量重新替换回该短语中的某个位置。

10 在功能框 330 用相关通配符函数置换单词变量的目的是减少必须在 NLP 数据库 218 中出现的条目的数量。例如，NLP 数据库 218 会只含有“what movies are playing at \$time”（在 \$time 时在放什么电影），而不是有一个个单独的对应于 8 点钟、9 点钟、10 点钟等等的条目。下文将对 NLP 数据库 218 作进一步说明。

15 在功能框 332，由代词置换器 211 用专有名称置换短语中的代词。例如，在一个实施例中，要把代词“I”（我）、“my”或“mine”（我的）替换成说话人的姓名。这个步骤的目的是能在 NLP 数据库 218 中存储和访问特定于用户的事实。例如，要把语句“who are my children”（谁是我的孩子）转换成“who are Dean's children”
20 （谁是 Dean 的孩子），其中 Dean 是说话人的姓名。这个步骤还是可以用一个简单循环来完成—该循环搜索短语中的代词并把找到的代词替换成专有名称。为了跟踪在短语中作过替换的位置，可以使用一个数组。

在功能框 334 中，由单词加权器 208 对短语中的各个单词按它们
25 对于短语的总体意思的相对“重要性”进行加权。例如，在一个实施例中，分配三个加权因数。将最低的加权因数分配给诸如“a”、“an”、“the”和其它冠词的单词。将最高的加权因数赋予可能与短语的意思有重要关系的单词。例如，这些单词可包括 NLP 数据库 218 中的所有动词、名词、形容词、和专有名称（proper name）。将中间的加
30 权因数赋予短语中的所有其它单词。这种加权的目的是便于增强 NLP 数据库 218 的搜索功能。

图 4 中表示一个实施例的 NLP 数据库 218 的几个选定列的例子。

NLP 数据库 218 包含多个列 400 - 410 和多个行 412A - 412N。列 400 中的各条目代表对 NLP 数据库来说是“已知的”短语。列 402 中显示的是列 400 中每个条目的必需单词的数目。列 404 中显示的是列 400 中每个条目的相关语境或子语境 (sub context)。列 408 和 410 中显示的是列 400 中每个条目的一个或多个相关行动。应当注意，图 4 中所示的 NLP 数据库 218 只是一个用于叙述本发明的简化的例子。其它实施例可以有更多或更少的具有不同条目的列。

返回到图 3B，在功能框 336，搜索 NLP 数据库 218，根据 NLP 数据库 218 的列 400 中的条目是否含有短语中的任何单词 (或单词的同义词) 和这些单词的相对权重，查找短语的可能匹配项。在功能框 338，根据短语中每一个单词的出现次数和它们的相对权重，为每一个可能的匹配条目生成一个可信度值。数据库的加权单词检索在本领域中是众所周知的，可以用市场销售的搜索引擎 (如位于美国弗吉尼亚州阿灵顿的 DT 软件公司生产的“dtsearch”) 执行。同样，同义词检索在本领域中是众所周知的，可以用诸如由位于美国新泽西普林斯顿的普林斯顿大学认识科学实验室开发的“WordNet”这样的可公开获得的工具执行。搜索引擎可以是自然语言处理器 202 的一个组成部分。

在判断框 340，自然语言处理器 202 判断是否可能匹配条目中的任一条目有大于或等于某个预定最小阈值 T 的可信度值。阈值 T 代表据其可判定短语是否与 NLP 数据库中任何条目匹配的可接受的最低可信度值。如果没有可信度值大于或等于 T 的可能匹配条目，则流程就前进到功能框 342，在此将一个可选的出错讯息或者通过显示器 104 向用户显示，或者通过扬声器 112 重播。在一个实施例中，向用户显示的任何出错讯息的类型 (如果有的话)，可以取决于在最高可信度的 NLP 数据库条目中找到的“命中”有多少 (即短语中有多少匹配单词)。有零个或一个命中与有两个或更多的命中相比，会生成不同类型的出错讯息。

然而，如果 NLP 数据库 218 中至少有一个条目的可信度值大于或等于 T，则流程前进到功能框 344，在此从短语中剔除“噪声”单词。“噪声”单词包括相对于短语中的其它单词来说对短语的总体意思没有重要贡献的单词。这些单词可包括冠词、代词、连词和类似性质的

单词。“非噪声”单词则包括对短语的总体意思有重要贡献的单词。

“非噪声”单词包括动词、名词、形容词、专有名称和类似的单词。

流程前进到图 3C，在功能框 346 从 NLP 数据库 218 的列 402 中提取最高可信度匹配条目的非噪声单词要求。例如，如果最高可信度
5 匹配短语是行 412A 中的条目（例如，“what movies are playing at \$time”（在 \$time 时在放什么电影）），则所要求的非噪声单词的数目是 3。

在判断框 348，进行一次检测，确定该短语所要求的非噪声单词的数目实际上出现在从 NLP 数据库 218 提取的最高可信度匹配条目
10 中。这个检测是对在功能框 336 中进行的相关风格搜索（relevance-style search）的精确性的验证，不言自明，一个可以生成大于最小阈值 T 的可信度值的条目，未必是短语的一个可接受的匹配项。

在判断框 348 进行的检测的性质是一个由布尔检测器 210 进行的
15 布尔“AND”检测。该检测确定短语中每一个非噪声单词（或其同义词）是否实际上出现在最高可信度条目中。如果有足够数量的实际上出现在最高可信度条目中的必要单词，流程就前进到功能框 350，自然语言处理器 202 在此指示应用界面 220 采取列 408 或 410 中的一个相关行动。不言自明，也可以有另外的行动列。

例如，如果最高可信度条目是行 412A 中的条目，并且判断框 348
20 的布尔检测确定列 400 中的条目中的短语中实际上有 3 个非噪声单词，则要采取列 408 中的相关行动（例如访问电影网站）。NLP 数据库中的其它条目有其它的相关行动。例如，如果最高可信度条目是行 412E 中的条目（例如，“what time is it”（现在几点钟）），则相关行动可能是，自然语言处理器 202 指示一个文字转换成语音应用程序（未予示出）通过扬声器 112 向用户报出当前时间。另一个例子是，如果最高可信度条目是行 412N 中的条目（例如，“show me the news”（给我显示新闻）），则第一个相关行动可能是访问因特网上预定的新闻网站，第二个相关行动可能是指示一个图象显示应用程序（未予示
25 出）显示新闻的相关图象。也可以采取不同的或另外的行动。
30

同样，对于短语，如果最高可信度条目含有如在判断框 348 确定的预定数量的非噪声单词，自然语言处理器 202 就指示语音识别处理

器 200 启用用于列 404 的相关语境的特定语境语法 212。所以，对于行 412A，要启用用于语境“movies”（电影）的特定语境语法 212。所以，当在图 3A 的功能框 300 向语音识别处理器 200 提供下一个话语时，语音识别处理器在搜索一般性语法 214 之前要在所启用的特定语境语法 212 中检索“movies”（电影）。如上所述，启用恰当的特定语境语法 212，大大增加快速、成功的语音识别的可能性，并且增强了用户以口语化风格与计算机沟通的能力。

然而 - 回到判断框 348，如果从 NLP 数据库 218 提取的最高可信度条目中实际上没有短语中预定数量的非噪声单词，流程就前进到功能框 354，通过显示器 104 或扬声器 112 提示用户回答是否指的是最高可信度条目。例如，如果用户说“how much is IBM stock selling for today”（IBM 股票今日卖多少），则 NLP 数据库 218 中的最高可信度条目可能是行 412B 中的条目。在这种情况下，尽管相关因数可能很高，必要单词（或其同义词）的数量却可能不够。所以，要在功能框 354 提示用户回答他是否指的是“what is the price of IBM stock on August 28, 1998”（1998 年 8 月 28 日 IBM 股票的价格是多少）。

用户可以作肯定回答或否定回答。如果在判断框 356 确定用户作了肯定回答，则在功能框 350 采取与该最高可信度条目相关的行动，然后在功能框 352 启用相关的特定语境语法 212。

然而，如果在判断框 356 确定用户作了否定回答，流程就前进到图 3D，从 NLP 数据库 218 的列 404 提取用于该最高可信度条目的相关语境，并在功能框 360 提示用户用基于语境的交互对话提供信息。例如，如果用户说“what is the price of XICOR stock today”（XICOR 股票今日价格是多少），并且 NLP 数据库 218 中的最高可信度条目是行 412B（例如“what is the price of IBM stock on \$date”（日期\$的 IBM 股票的价格是多少）），则在功能框 354 提示用户回答他是否指的就是这个意思。

如果用户回答是否定的，则在功能框 358 从列 404 提取语境“stock”（股票），并通过显示器 104 或扬声器 112 把用于该股票语境的基于语境的交互对话向用户表示。这种基于语境的交互对话的作用是提示用户提供 XICOR 股票的名称和股票报价符。用户可以口头

回答所需信息。对于每一个可能的语境，可以采用不同的基于语境的交互对话。例如，基于语境“weather”（天气）的交互对话的作用是提示用户提供希望了解其天气情况的地点和名称（例如城市）。同样，基于语境“news”（新闻）的交互对话的作用是提示用户提供文章的类型、新闻来源、新闻网站的因特网 URL 或其它有关信息。

5 基于语境的交互对话结束后，更新 NLP 数据库 218、一般性语法 214 和特定语境语法 212，以包括新的信息。这样，下一次用户询问该信息时，就会找到合适的匹配项，无需提示用户提供更多信息就能采取适当的行动。所以，本发明自适应地“学习”识别用户说出的短

10 语。

在本发明一个实施例中，NLP 数据库 218、特定语境语法 212、一般性语法 214 和口授语法 216 的一个或多个也含有与每个条目相关联的时间戳值（未予示出）。每次使用一个匹配条目时，就更新与该条目相关联的时间戳值。每隔一定时间，或者由用户决定，把时间戳值

15 在一定日期或时间之前的条目从它们各自的数据库/语法中删除。这样，通过“清除”旧的或过时的条目，可以使数据库/语法的大小保持在高效的程度。这也有助于避免错误的匹配。

总之，本发明提供一种用于计算机的交互式用户界面的方法和系统。通过使用特定语境语法，本发明减少语音识别时间，提高用户以

20 口语化风格与计算机沟通的能力。此外，通过采用对各种语法和 NLP 数据库的适应性更新，本发明进一步提高交互的效率。

前面的最佳实施例说明，使本领域任何熟练人员都能制造或使用本发明。这些实施例的各种修改，对本领域的熟练人员来说显而易见，无需发明性技能，就能将本文中所界定的一般原理应用于其它实

25 施例。所以，本发明不限于本文所示的实施例，而应涵盖与本文所披露的原理和新颖特点相符的最宽的范围。

说明书附图

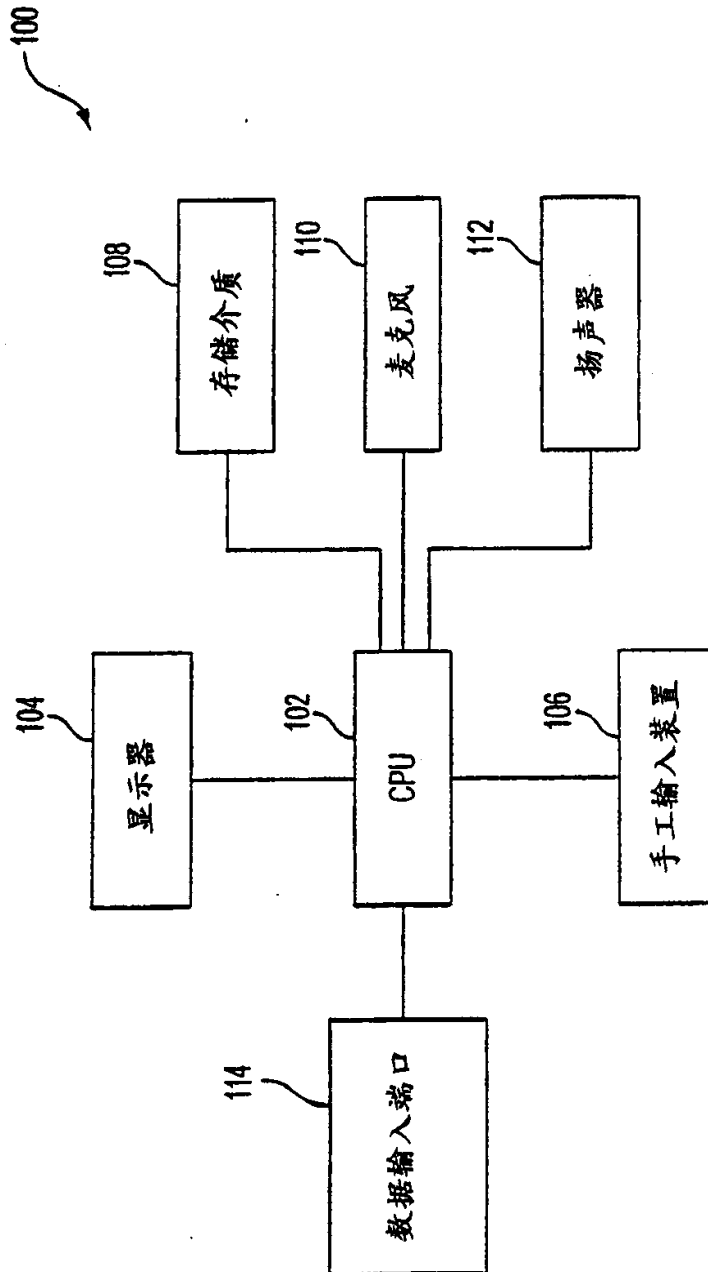


图 1

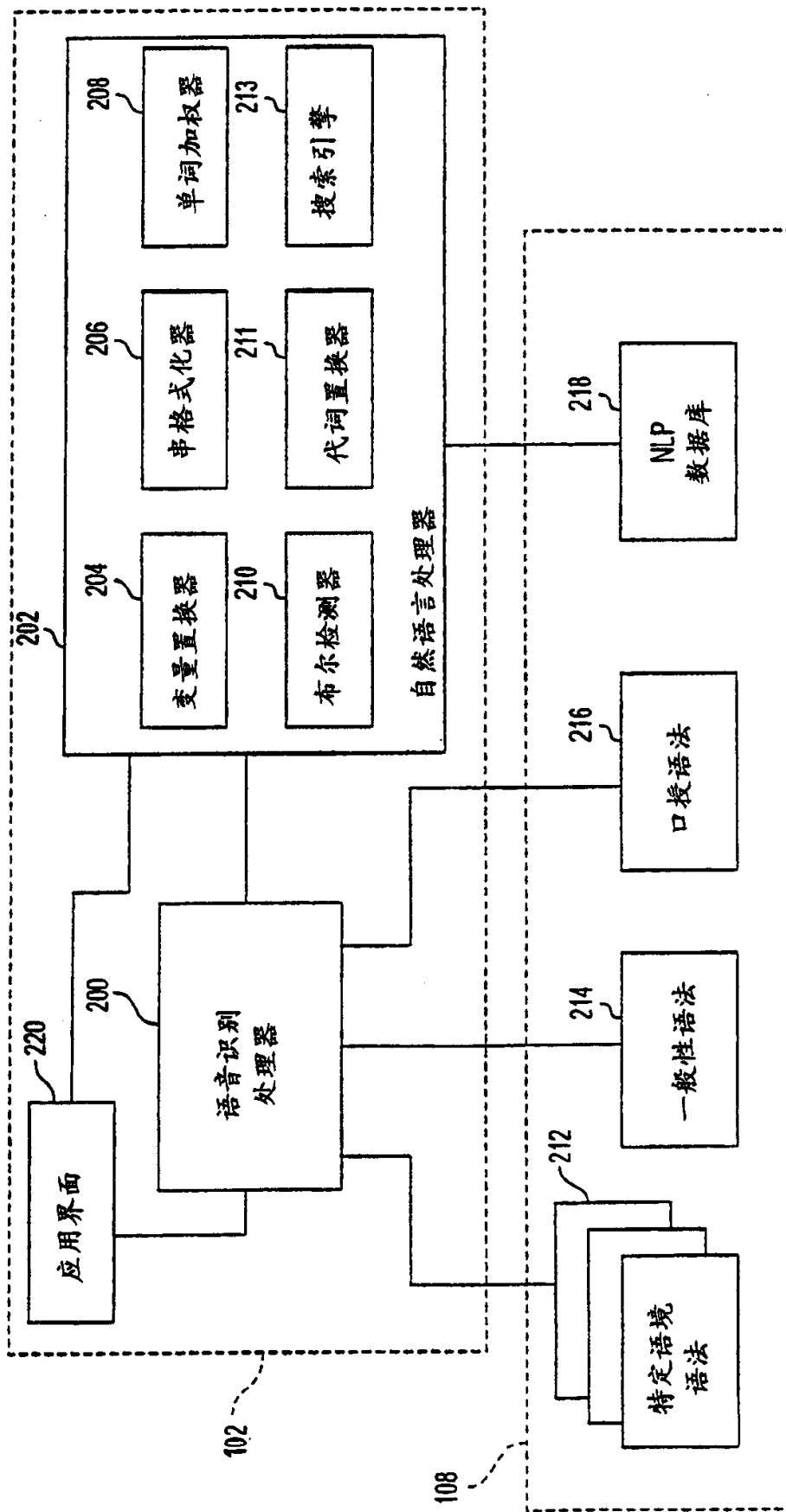


图 2

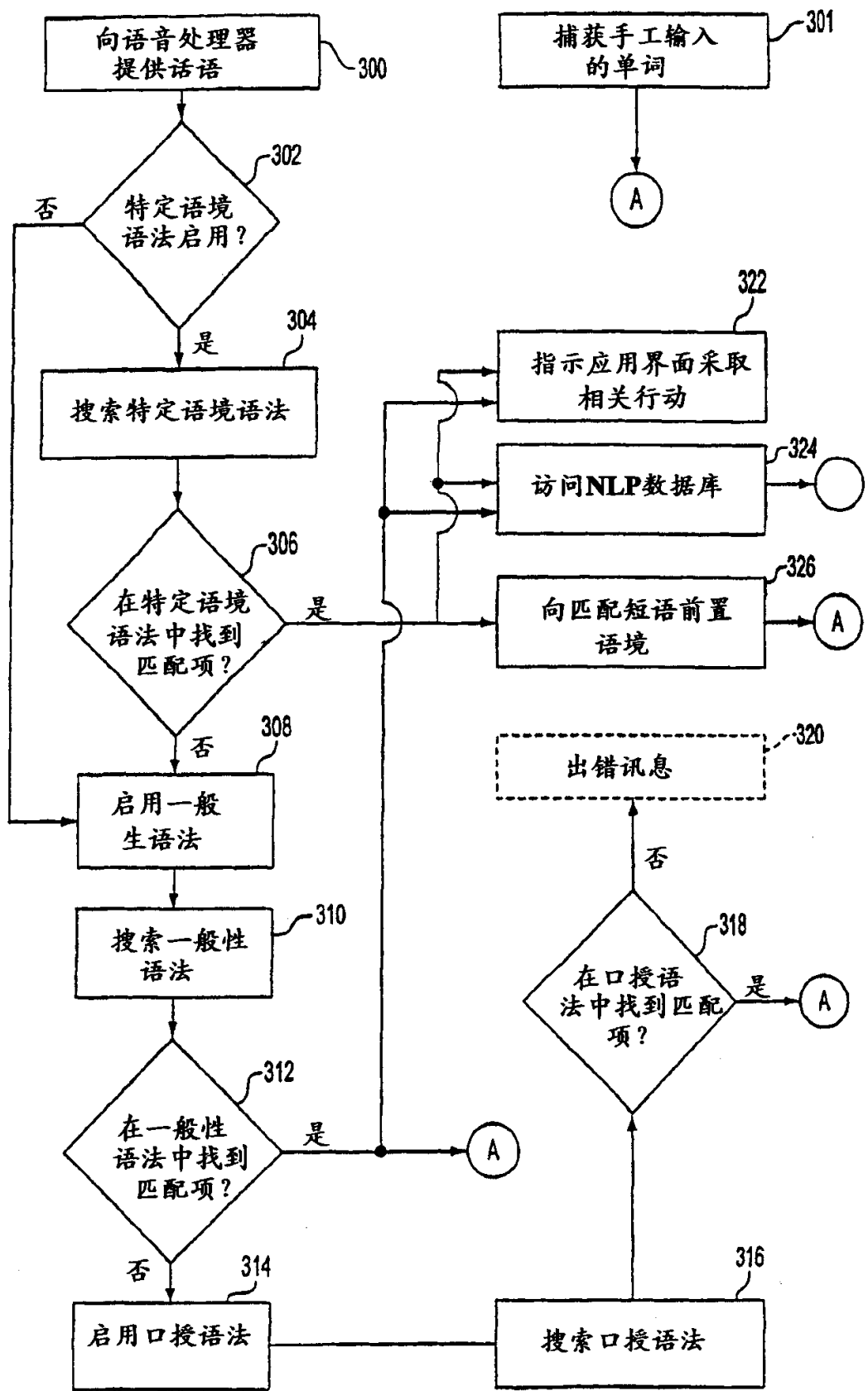


图 3A

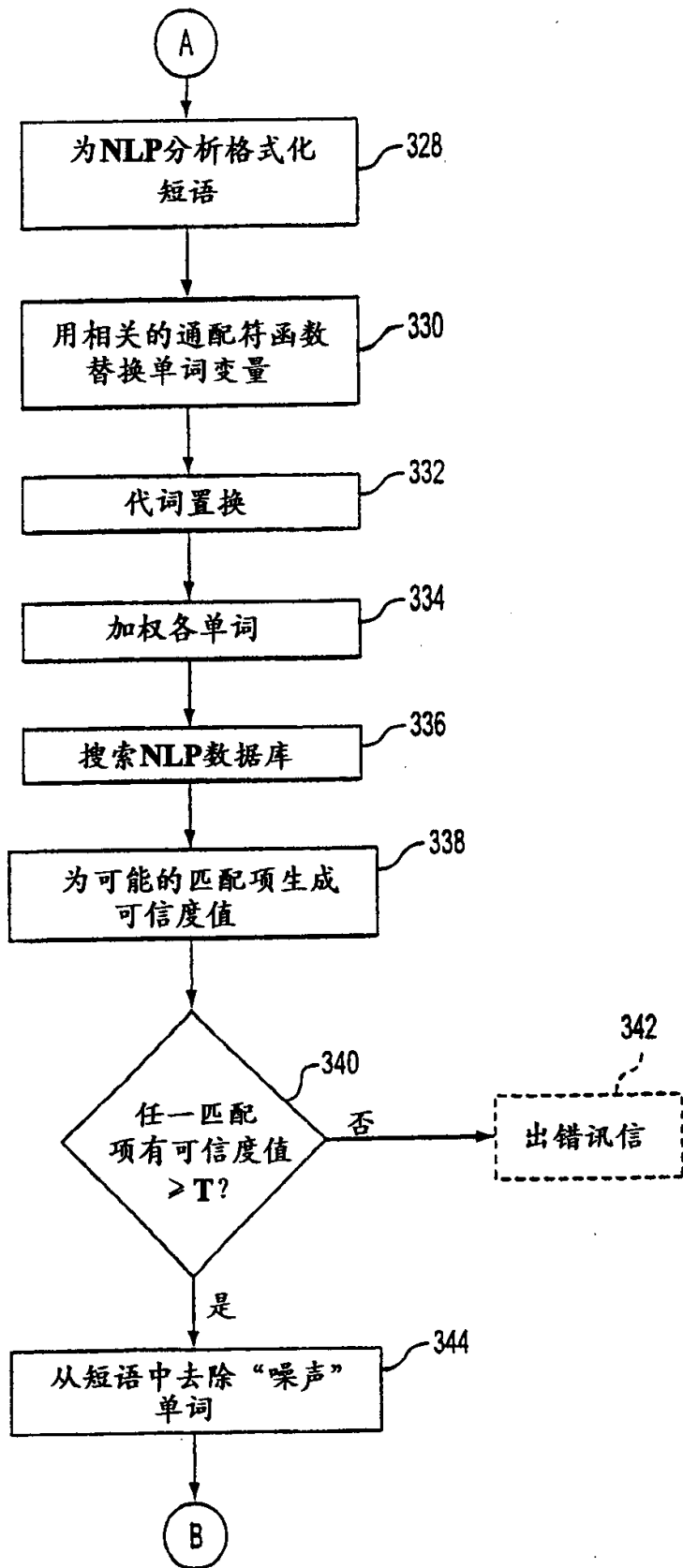


图 3B

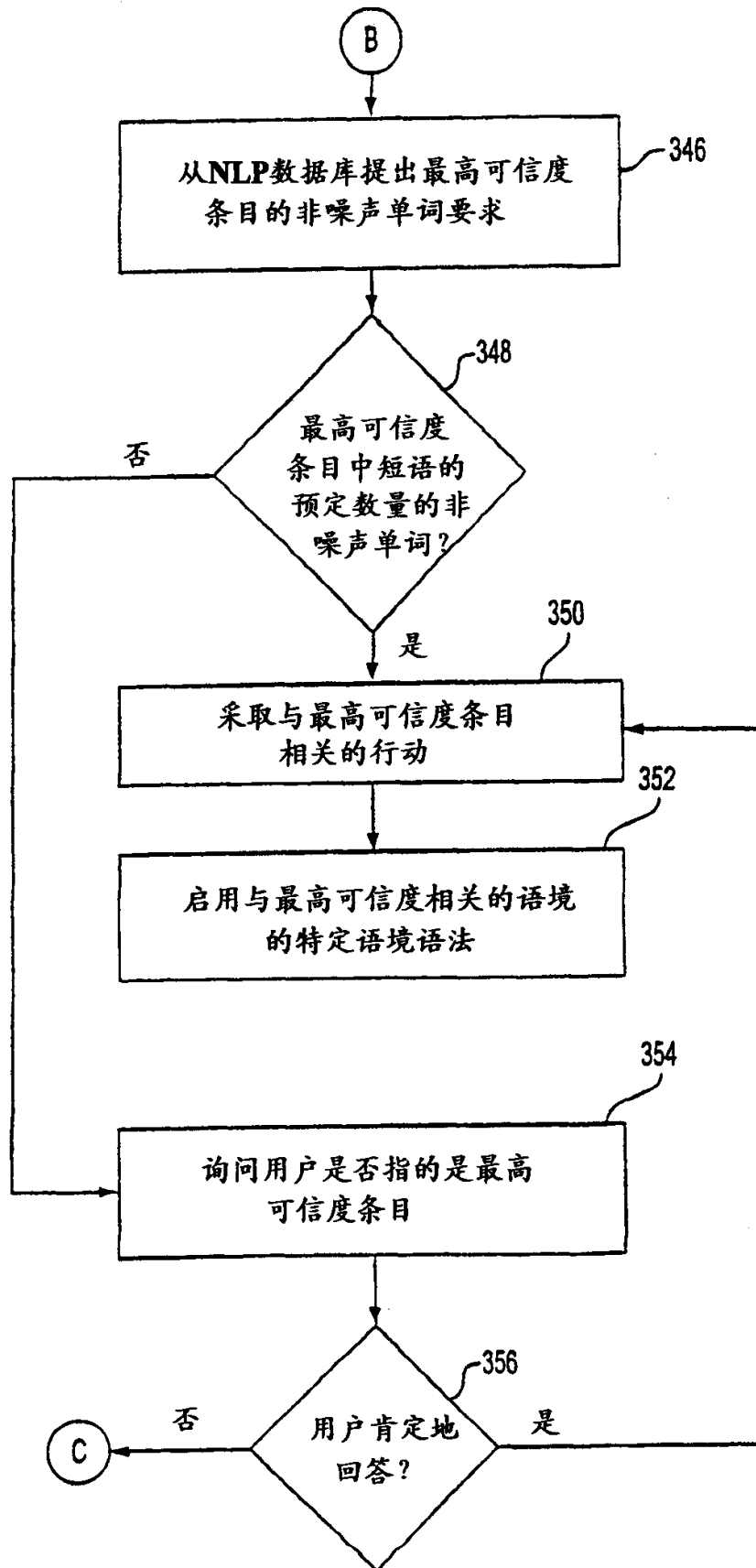


图 3C

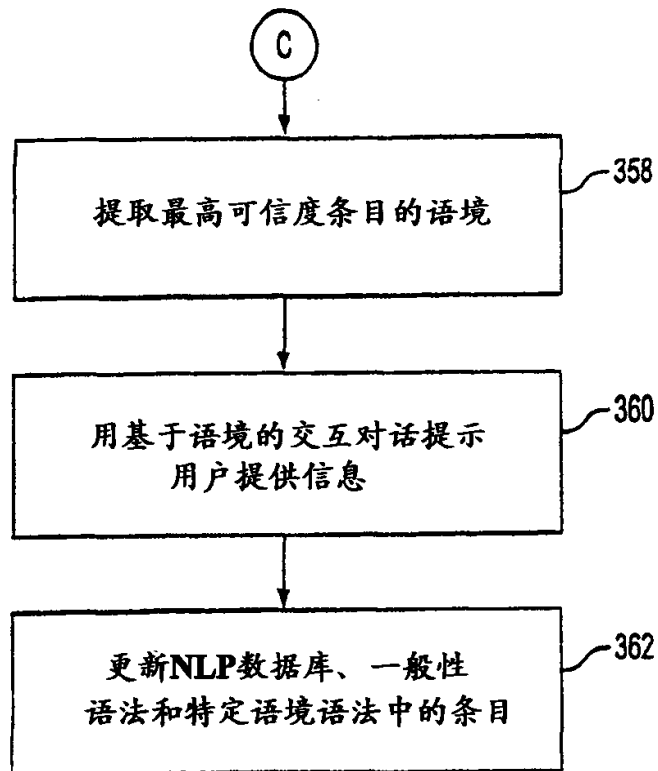


图 3D

218	400	402	404	408	410
	短语	预定 单词数	语境/ 子语境	行动1	行动2
412A	WHAT MOVIES ARE PLAYING AT \$ TIME	3	电影	访问电影网站	无
412B	WHAT IS THE PRICE OF IBM STOCK ON \$ DATE	4	股票	访问股票价格 网站	无
412C	SELL IBM STOCK AT \$ DOLLARS	4	股票	访问股票交易 网站	无
412D	WHAT IS THE WEATHER IN \$ LOCATION	3	天气	访问天气网站	无
412E	WHAT TIME IS IT	2	时间	无	时间的 正文到 言语转换
412N	SHOW ME THE NEWS	2	新闻	访问新闻 网站	显示 图象

图 4