(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 102708115 B
(45)授权公告日 2015.12.09

(21)申请号 201210031669.8
(22)申请日 2005.08.08
(30)优先权数据
10/914,722 2004.08.09 US
(62)分案原申请数据
200580026931.8 2005.08.08
(73)专利权人 亚马逊技术股份有限公司
地址 美国内华达州
(72)发明人 N·B·肖尔 A·W·德纽
(74)专利代理机构 上海专利商标事务所有限公司 31100
代理人 张政权
(51)Int.Cl.
G06F 17/30(2006.01)
G06Q 30/02(2012.01)
(56)对比文件
CN 1269897 A,2000.10.11, 全文.
CN 1306258 A,2001.08.01, 全文.
US 2004/0103070 A1,2004.05.27, 全文.
US 5845278 A,1998.12.01, 全文.
审查员 李迪

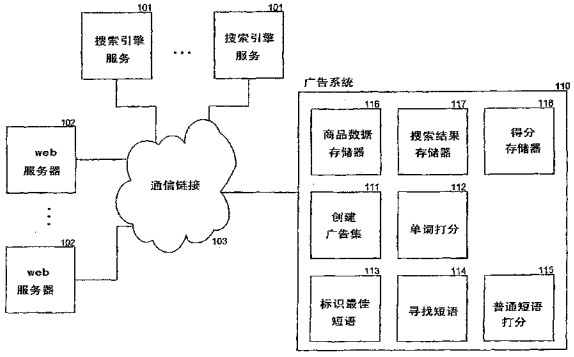
权利要求书2页 说明书5页 附图6页

(54)发明名称

标识用于放置关键词目标广告的关键词的方法和系统

(57)摘要

提供了标识用于放置关键词目标广告的关键词的方法和系统。广告系统选择要广告商品的描述。广告系统然后检索与该选定描述相匹配的文档。广告系统对所检索到的文档的每个词打分，表示该词与要广告商品的相关性。在对词打分之后，广告系统标识文档内与商品相关的词。广告系统然后根据所标识短语生成该要广告商品的搜索词。广告系统向搜索引擎服务提交搜索词和广告，用于放置该商品的付费广告。



CN 102708115 B

1. 一种在计算机系统中用于从在搜索引擎的搜索结果中标识为与商品相关的文档中标识与所述商品相关的短语的方法，所述方法包括：

对所述搜索结果中的所述文档的词产生得分，所述得分表示所述词与所述商品的相关性，该对所述搜索结果中的所述文档的词产生得分的步骤包括：

计算在所述文档内所述词的平均频率，所述平均频率表示该词对每个所述文档的频率的平均值；

检索所述词的普通频率，所述普通频率表示该词在所述文档的文档集中的平均频率；

根据所述词的频率得分和包含得分的线性组合产生所述词的得分，其中所述频率得分提供所述平均频率相对所述普通频率的得分，所述包含得分表示搜索结果的文档中包含选定词的比例；

选择具有最高得分的词；

定位所述文档内的每个选定词，作为短语的锚词；

对表示所述短语与所述商品的相关性的每个短语产生得分，其中所述短语的得分是至少部分地基于所述短语中的各词的相应词得分来确定的；

当扩展后的短语的得分高于扩展前的短语的得分时，将每个短语扩展到包含接近所述短语的词；以及

将扩展后的短语提交给所述搜索引擎作为附加搜索请求来提供所述搜索结果的附加信息源。

2. 如权利要求 1 所述的方法，其特征在于还包括在对所述文档的词产生得分之前：

选择所述商品的描述；以及

选择与所述商品的选定描述相匹配的、相关于所述商品的文档。

3. 如权利要求 2 所述的方法，其特征在于，选择与所述商品相关的文档包括向搜索引擎服务提交所述商品的选定描述，并且其中所选择的文档基于由所述搜索引擎服务提供的搜索结果来检索。

4. 如权利要求 2 所述的方法，其特征在于，选择所述商品的描述包括从商品目录检索所述描述。

5. 如权利要求 4 所述的方法，其特征在于，所述商品的描述是存储在所述商品目录中的商品的名称。

6. 如权利要求 1 所述的方法，其特征在于，短语仅扩展到具有表示与所述商品的相关性的得分的词。

7. 如权利要求 1 所述的方法，其特征在于，对表示所述短语与所述商品的相关性的每个短语产生得分的步骤包括确定所述短语在所述文档内出现的次数。

8. 如权利要求 1 所述的方法，其特征在于，与短语中的另一个词相似的词不被添加到所述短语中。

9. 如权利要求 1 所述的方法，其特征在于，当遇到与已在短语中的词相似的词时，终止对所述短语的扩展。

10. 如权利要求 1 所述的方法，其特征在于，忽略噪声字。

11. 如权利要求 1 所述的方法，其特征在于，忽略通常在一般文档集中得分高的词。

12. 如权利要求 1 所述的方法，其特征在于在对所述文档的词产生得分之前，忽略与其

它检索到的文档相似的文档。

13.一种用于从与商品相关的信息源中标识与所述商品相关的短语的计算系统，包括：

第一打分子系统，它对所述信息源的词产生得分，每个得分表示所述词与所述商品的相关性，该对所述信息源的词产生得分包括：

计算在所述信息源内所述词的平均频率，所述平均频率表示该词对每个所述信息源的频率的平均值；

检索所述词的普通频率，所述普通频率表示该词在所述信息源的信息源集中的平均频率；

根据所述词的频率得分和包含得分的线性组合产生所述词的得分，其中所述频率得分提供所述平均频率相对所述普通频率的得分，所述包含得分表示搜索结果的信息源中包含选定词的比例；

定位子系统，它定位所述信息源内得分最高的词，作为短语的锚词；

第二打分子系统，它对表示所述短语与所述商品的相关性的每个短语产生得分，其中所述短语的得分是至少部分地基于所述短语中的各词的相应词得分来确定的；以及

短语扩展子系统，当扩展后的短语的得分高于扩展前的短语的得分时，将每个短语扩展到包含接近所述短语的词，其中扩展后的短语被提交给搜索引擎作为附加搜索请求来提供搜索结果的附加信息源。

标识用于放置关键词目标广告的关键词的方法和系统

[0001] 本申请是国际申请日为 2005 年 8 月 8 日，发明名称为“标识用于放置关键词目标广告的关键词的方法和系统”的第 200580026931.8 号中国专利申请的分案申请。

技术领域

[0002] 所述技术一般涉及与商品相关的词，尤其涉及用于放置商品广告的搜索词。

背景技术

[0003] 诸如 Google 和 Overture 的许多搜索引擎服务，提供对经由因特网可访问的信息的搜索。这些搜索引擎服务使用户能搜索用户感兴趣的网页和其它因特网可访问的资源。在用户提交包括搜索词的搜索请求之后，搜索引擎服务标识可能与那些搜索词相关的网页。为了快速标识相关网页，搜索引擎服务可保持关键词与网页的映射。该映射可通过“爬寻”web（即万维网）来生成，以标识各网页的关键词。为了爬寻 web，搜索引擎服务可使用根网页列表来标识可通过那些根网页访问的所有网页。任何特定网页的关键词可使用各种公知信息检索技术来标识，诸如标识标题行的词、在网页的元数据中提供的词、高亮的词等等。一些搜索引擎服务甚至可搜索不可经由因特网访问的信息源。例如，图书出版者可将其图书的内容提供给搜索引擎服务。该搜索引擎可生成关键词和图书之间的映射。当搜索引擎服务接收到包括一个或多个搜索词的搜索请求时，它使用其映射来标识其关键词与搜索词最接近匹配的那些信息源（例如网页或图书）。与搜索词最接近匹配的信息源的集合被称为“搜索结果”。然后该搜索引擎服务基于各匹配的接近性、网页的流行性（例如 Google 的页面排序）等来排列搜索结果的信息源。然后搜索引擎服务按基于其排序的顺序向用户显示与那些信息源的链接。

[0004] 一些搜索引擎服务不为了在搜索结果中包含与其网页的链接而向网页的提供者收费。相反，搜索引擎服务通过将广告和搜索结果放置在一起来获得收入。为广告付款的那些通常称为“广告链接”、“广告匹配”、或“付款搜索结果”。想要将商品广告与某些搜索结果放置在一起的广告商向搜索引擎服务提供广告和搜索词。当接收到搜索请求时，搜索引擎服务标识其搜索词与搜索请求的搜索词最接近地相匹配的广告。该搜索引擎服务可对将广告与搜索结果放置在一起收费（即按印象收费），或者仅在用户实际选择与广告相关联的链接时收费（即按点击收费）。

[0005] 广告商想要使用于支付与搜索结果放置在一起的广告的广告费的效用最大。那些广告商尝试标识导致广告商最高利益（例如：最高利润）的广告商品的搜索词。需要具有通过标识更针对或相关于广告商品的搜索词而使广告商将其广告费的效用最大化的技术。

附图说明

[0006] 图 1 是示出一实施例中广告系统的各个部分的框图。

[0007] 图 2 是示出一实施例中创建广告集部分的处理的流程图。

[0008] 图 3 是示出一实施例中单词打分部分的处理的流程图。

[0009] 图 4 是示出一实施例中标识最佳短语部分的处理的流程图。

[0010] 图 5 是示出一实施例中寻找短语部分的处理的流程图。

[0011] 图 6 是示出一实施例中通用短语打分部分的处理的流程图。

具体实施方式

[0012] 提供了用于标识将广告与搜索结果放置在一起的搜索词的一种方法和系统。在一实施例中，广告系统选择要广告商品的描述。例如，如果商品是一本书，则描述可以是该书的标题；或者如果商品是一电器，则描述可以是该电器的概述。然后广告系统检索与从信息源全集中选出的描述相匹配（例如，最接近地相关于其主题）的文档或其它信息源。例如，广告系统可将选定的描述提交给搜索引擎服务，其中搜索结果的网页为检索文档。然后该广告系统对检索文档的每个词打分，指示该词与要广告商品的相关性。在一实施例中，广告系统可对在检索文档中比在信息源全集中频繁得多地使用的词打高分。例如，如果商品是哈利波特丛书，则诸如“Hogwarts”、“Fluffy”、“three-headed”、“dog”、“Hermione”和“Granger”的单词会有相对较高的得分，因为这些单词在哈利波特的描述中比无关描述出现得更为频繁。在对词打分后，广告系统标识文档内可能与商品相关的词的短语。例如，广告系统可标识短语“Fluffy the three-headed dog”和“Hermione Granger”可能与该书相关。然后广告系统根据所标识短语来生成要广告商品的搜索词。该广告系统向搜索引擎服务提交搜索词和广告，用于放置该商品的付款广告。例如，广告系统可将哈利波特丛书的广告与搜索词“Hermione Granger”放置在一起。当某人将“Hermione Granger”的搜索请求提交给搜索引擎服务时，它使该广告与搜索结果一起显示。这样，广告系统可基于信息源中使用的已知相关于要广告商品的短语来标识搜索词。

[0013] 在一实施例中，广告系统标识可能与要广告商品相关的短语。因为当 n 是文档内单词的数量时文档内短语的数量为 $O(n^2)$，并且文档集中可能短语的数量为 k^1，其中 k 为不同单词的数量而 1 是短语的长度，所以计算和跟踪所有可能的短语在计算上极为昂贵。为了减少所估算短语的数量，广告系统对彼此接近的词的组合打高分。广告系统开始时对文档内与商品相关的词打分。该分数指示该词与商品相关的可能性。广告系统然后可标识高度相关词和相关词。高度相关词具有诸如最高 10％的分数的极高分，并且相关词具有诸如最高 25％的分数的高分。广告系统在文档中搜索高度相关词。文档内的每个高度相关词被视为短语的“锚词”。广告系统尝试扩展短语使其包括附近的相关词。在一实施例中，广告系统可通过跟在锚词后面的任何相邻相关词来扩展该短语。例如，如果“Hermoine”是高度相关词，而“Granger”是相关词，则短语“Hermione Granger”在“Hermoine”于文档中跟在“Granger”后面时将被标识为一短语。或者，广告系统可将短语扩展成还包括锚词之前的词。例如，如果“Granger”是高度相关词而“Hermoine”仅仅是相关词，则仍将标识出短语“Hermione Granger”。广告系统可计算短语得分，并且只要经扩展短语的得分变高就继续扩展短语，而不管该短语的所有词是否是相关词。本领域技术人员将理解，用于标识这些短语的技术可用于除生成广告的搜索词之外的环境中。例如，搜索引擎服务可将在搜索结果中标识的短语用作搜索请求，用于定位要提供给用户的附加相关信息源。或者，广告系统可从附加相关信息源中标识更多短语。更一般地，给定一信息源集，用于标识短语的技术可用来标识信息源的主题。例如，如果信息源是聊天讨论，则所标识的短语可代表聊天讨论的

最流行话题。

[0014] 图 1 是示出一实施例中广告系统的各个部分的框图。广告系统 110 经由通信链接 103 与搜索引擎服务计算机系统 101 和 web 服务器计算机系统 102 相连。广告系统将商品的描述提交给搜索引擎服务计算机系统，并接收由 web 服务器计算机系统提供的匹配网页链接。然后广告系统从 web 服务器计算机系统中检索匹配网页。广告系统从那些匹配网页中标识短语，并从所标识短语中得到搜索词。然后广告系统向搜索引擎服务提交搜索词以及商品的广告。搜索引擎服务对匹配搜索词的搜索查询显示广告以及搜索结果。

[0015] 广告系统包括创建广告集部分 111、单词打分部分 112、标识最佳短语部分 113、寻找短语部分 114、通用短语打分部分 115、商品数据存储器 116、搜索结果存储器 117 和得分存储器 118。商品数据存储器包含每个要广告商品的标识符（例如 SKU）以及商品的描述。例如，商品数据存储器可以是要广告图书的电子目录。各目录条目可包括商品标识符、标题、作者名字、概述等等。搜索结果存储器包含搜索词所标识的商品的匹配网页。得分存储器包含搜索结果存储器的单词和短语的得分。创建广告集部分拥有商品标识符，并标识在广告该商品时要使用的搜索词（例如关键词）。创建广告集部分请求搜索引擎服务提供搜索结果，检索那些搜索结果的网页，调用单词打分部分和标识最佳短语部分，然后生成广告集。单词打分部分对搜索结果的每个词打分，指示该词与该商品相关的可能性。标识最佳短语部分调用寻找短语部分和通用短语打分部分，以标识可能与该商品相关的短语。

[0016] 广告系统可在包括中央处理单元、存储器、输入设备（例如键盘和定位设备）、输出设备（例如显示设备）和存储设备（例如盘驱动器）的计算机系统和服务器上实现。存储器和存储设备是可包含实现广告系统的指令的计算机可读介质。此外，数据结构和消息结构可经由数据传输介质，诸如通信链接上的信号存储或传送。可使用各种通信链接，诸如因特网、局域网、广域网、或点对点拨号连接。

[0017] 图 2 是示出在一实施例中创建广告集部分的处理的流程图。该部分得到所传递的商品标识符，并返回带有从可能与商品相关的短语中导出的搜索词的广告集。在框 201，商品检索该商品的描述。例如，描述可以是书名或组合有制造商名称的商品名（例如索尼 DVD 播放器）。在框 202，该部分请求搜索引擎服务将检索到的描述用作搜索请求来执行搜索。该部分接收搜索结果。如果搜索结果是诸如网页的 URL 的链接，则该部分检索所链接的网页并将它们存储在搜索结果存储器中。该部分可仅存储和使用搜索结果的最佳匹配网页（例如，最前面的 15 个）。在框 203，该部分调用单词打分部分来对搜索结果中的每个词打分。被调用的部分将得分存储在得分存储器中。在框 204，该部分调用标识最佳短语部分来标识与商品最为高度相关的短语。被调用的部分将短语得分存储在得分存储器中。在框 205，该部分使用最佳短语生成该商品的广告集。然后该部分完成。然后这些广告集可被提交给一个或多个搜索引擎服务。

[0018] 图 3 是示出一实施例中单词打分部分的处理的流程图。单词打分部分对存储在搜索结果存储器的网页中的每个词打分。该部分将得分存储在得分存储器中。在框 301-308，该部分循环选择搜索结果中的每个词，并计算其得分。在框 301，该部分选择搜索结果中的下一个词。在判定框 302，如果已选择了搜索结果中的所有词，则该部分返回，否则该部分在框 303 继续。本领域技术人员将理解，该部分将跳过噪声字（例如“of”、“a”、“the”等等）。在框 303，组件计算在搜索结果的文档（例如网页）内选定词的平均频率。词的“频

率”是文档（例如网页）内该词的出现次数除以各词在该文档出现的总次数。例如，如果一个词在包含200个词的文档内出现了10次，则其频率为0.05（即10/200），这表示它占文档中词的5％。搜索结果内词的“平均频率”是该词对每个文档的频率的平均值。例如，如果在具有4个文档的搜索结果中一词的频率为0.05、0.04、0.02和0.01，则该词的平均频率为0.03（例如 (0.05+0.04+0.02+0.01)/4）。平均频率由以下方程来表示：

[0019] $$\overline{f}=\frac{\sum_{i=1}^{n} f_i}{n} \qquad (1)$$

[0020] 其中$\overline{f}$是词的平均频率，f_i是该词在文档i中的频率，并且n是文档的数量。在框304，该部分检索该词的“普通频率”。普通频率表示该词在诸如全部网页的极大文档集中的平均频率。在框305，该部分计算选定词的“频率得分”。如果选定词的平均频率比选定词的普通频率高得多，则该词可与商品高度相关。频率得分提供平均频率相对普通频率的得分。频率得分可由以下方程来表示：

[0021] $$S_f=0.5+\frac{a\tan\left(\frac{\overline{f}=\widetilde{f}}{10*\widetilde{f}}\right)}{\pi} \qquad (2)$$

[0022] 其中S_f是该词的频率得分，$\widetilde{f}$是该词的普通频率，而atan是反正切函数。本领域技术人员将理解：该方程仅仅是可用来生成频率得分的许多方程之一。所使用的特定方程可基于给予词的平均频率和普通频率之差的权重来选择。在框306，该部分计算包含选定词的搜索结果的文档的数量。在框307，该部分计算表示搜索结果的文档中包含选定词的比例的“包含得分”。该包含得分可由以下方程来表达：

[0023] $$S_c=\frac{n'}{n} \qquad (3)$$

[0024] 其中S_c是包含得分，并且n’是搜索结果的包含选定字的文档的数量。在框308，该部分计算选定字的得分。在一实施例中，单词得分是频率得分和包含得分的线性组合。频率得分和包含得分的权重可被设置成反映是频率得分、还是包含得分被视为该词与商品相关的可能性的更准确表示。单词得分可由以下方程表示：

[0025] $S = \alpha * S_f + (1-\propto) * S_c$ (4)

[0026] 其中S是单词得分而∝从0到1地变化，并表示给予频率得分的权重。该部分然后回到框301，以选择搜索结果中的下一个词。

[0027] 图4是示出一实施例中标识最佳短语部分的处理的流程图。在框401，该部分选择搜索结果的高度相关词。这些高度相关词可以是其得分为最高15％的那些词。最高相关词被用作该短语的锚词。在框402，该部分选择搜索结果的相关词。相关词可以是其得分为最高40％的那些词。相关词包括高度相关词。短语可被扩展成包括靠近锚词的相关词。本领域技术人员将理解：可使用各种标准来选择高度相关词和相关词。例如，高度相关词可以是具有最高得分的10个词，而相关词可以是具有最高得分的50个词。此外，高度相关词和相关词可以是相同的词集（例如具有最高得分的20个词）。在框403-405，该部分循环选择搜索结果中的文档，并在那些文档内寻找短语。在框403，该部分选择搜索结果中的下一个文档。在判定框404中，如果已选择了搜索结果中的所有文档，则该部分在框406继续，

否则该部分在框 405 继续。在框 405，该部分调用寻找短语部分来在选定文档中寻找短语。然后该部分循环至框 403 以选择下一文档。在框 406，当已在所有文档中找到短语之后，该部分选择通用短语，即在文档内频繁出现的短语。例如，通用短语可以是在文档内出现 5 次以上、或在文档中以一定百分比出现的短语。在框 407，该部分调用普通短语打分部分来对每个通用短语生成短语得分。然后该部分返回。广告系统从通用短语中得到搜索词。

[0028] 图 5 是示出一实施例中寻找短语部分的处理的流程图。该部分得到所传递的一个文档，并在该文档中标识短语。在框 501-509，该部分循环标识文档内具有作为锚词的高度相关词的短语。在框 501，该部分在文档内选择高度相关词。在判定框 502，如果已经选择了文档的全部相关词，则该部分完成，否则该部分在框 503 继续。在框 503，该部分将具有高度相关词的短语初始化为锚词。在框 504-509，该部分循环扩展短语以使其包括附近的相关词。在框 504，该部分选择该文档内的下一个词。在判定框 505，如果选定词是一相关词，则该部分在框 506 继续，否则该部分终止短语的扩展，并循环至框 501 以标识该文档内的下一短语。在判定框 506，如果选定词与已在短语中的词相似，则该部分终止短语的扩展，并循环至框 501 以标识下一短语，否则该部分在框 507 继续。在判定框 507，如果选定词将提高短语得分，则该部分在框 509 继续，否则该部分在框 508 继续。在判定框 508，如果选定词和选定词之后的下一个词将提高短语得分，则该部分在框 509 继续，否则该部分终止短语的扩展，并循环至框 501 以标识下一短语。在框 509，该部分将选定词添加到短语中，并循环至框 504 以选择用于扩展该短语的下一词。

[0029] 图 6 是示出在一实施例中普通短语打分部分的处理的流程图。该部分计算通用短语的短语得分。或者，当标识每个普通短语时，可计算该短语得分。在框 601，该部分选择下一普通短语。在判定框 602，如果已经选择了全部普通短语，则该部分返回，否则该部分在框 603 继续。在框 603，该部分初始化选定普通短语的短语得分。在框 604-607，该部分循环将普通短语的词的单词得分合成 (factor in) 为短语得分。在框 604，该部分选择选定普通短语的下一个词。在判定框 605，如果已经选择了选定普通短语的所有词，则该部分在框 607 继续，否则该部分在框 606 继续。在框 606，该部分将选定词的单词得分加到短语得分中，然后循环至框 604 以选择选定普通短语的下一个词。本领域技术人员将理解 :可使用许多不同技术来计算短语得分。例如，高度相关词的单词得分的两倍可被添加到短语得分以强调高度相关字的重要性，可使用单词得分的非线性组合，等等。在框 607，该部分将短语得分乘以选定普通短语在搜索结果内的出现次数，然后该部分循环至框 601 以选择下一普通短语。

[0030] 本领域技术人员将理解 :尽管为了作出说明已在本文中描述了广告系统的具体实施例，但可作各种更改而不背离本发明的精神和范围。术语“商品”可包括任何可广告的产品、服务、或观念。例如，政党可放置有关特定候选人或目标的广告。此外，广告集可能没有与之相关联的链接。广告商可能仅仅想要向使用某搜索词提交请求的用户显示广告的信息。例如，候选人可能想要在用户提交其对手的名字为搜索词的搜索请求时显示广告。本领域技术人员将理解 :可使用用于计算得分的各种方程和技术。此外，如果搜索结果包含复制品（或极为相似）的文档，则广告系统可忽略复制文档。广告系统可维持不应添加到短语中的单词列表，诸如所有网页上非常常见的词（例如，“下一页”、“保密策略”）。因此，除了所附权利要求之外，本发明不受到其它限制。

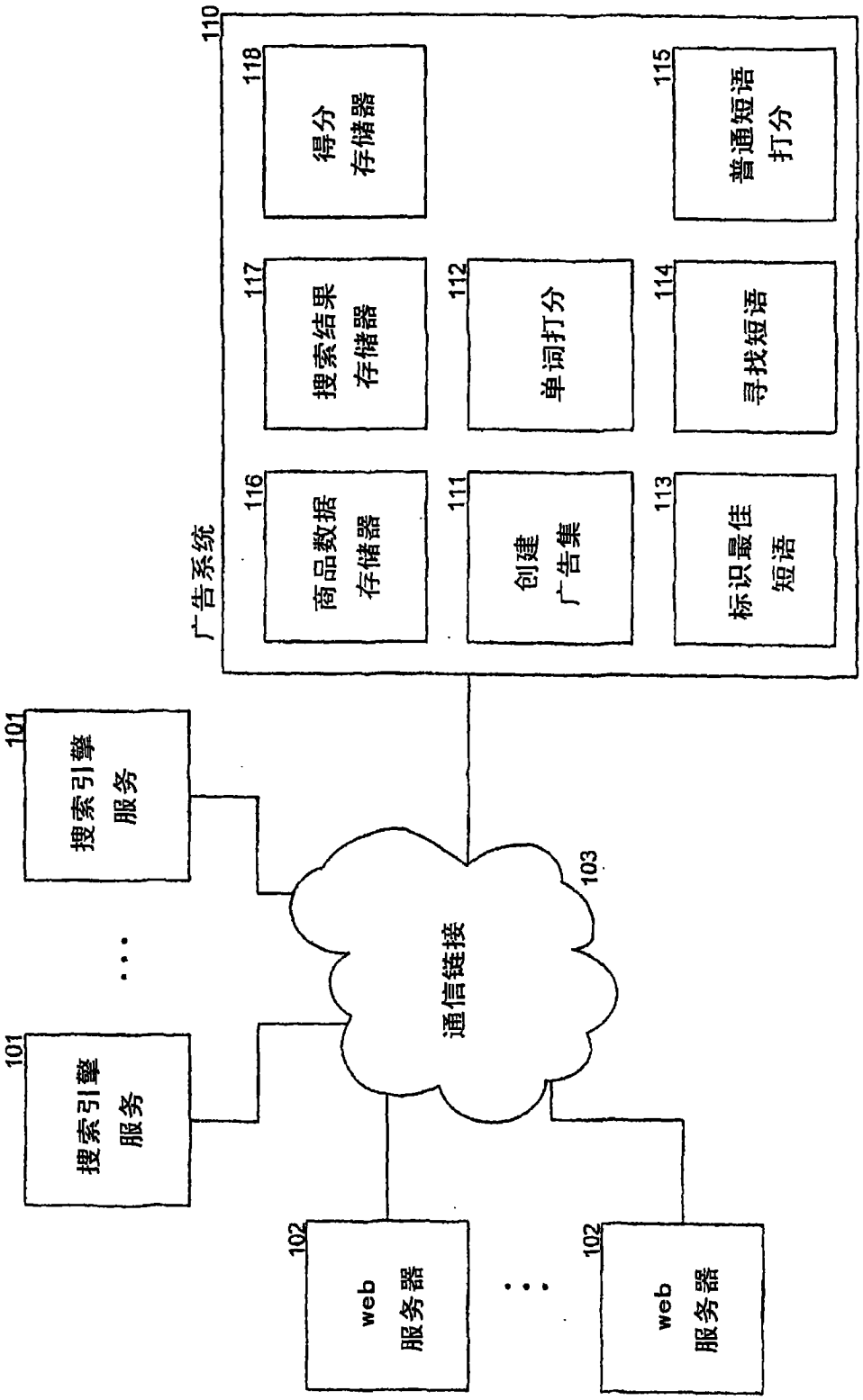

110
广告系统
118
得分
存储器
115
普通短语
打分
117
搜索结果
存储器
112
单词打分
114
寻找短语
116
商品数据
存储器
111
创建
广告集
113
标识最佳
短语
101
搜索引擎
服务
101
搜索引擎
服务
103
通信链接
102
web
服务器
102
web
服务器

图 1

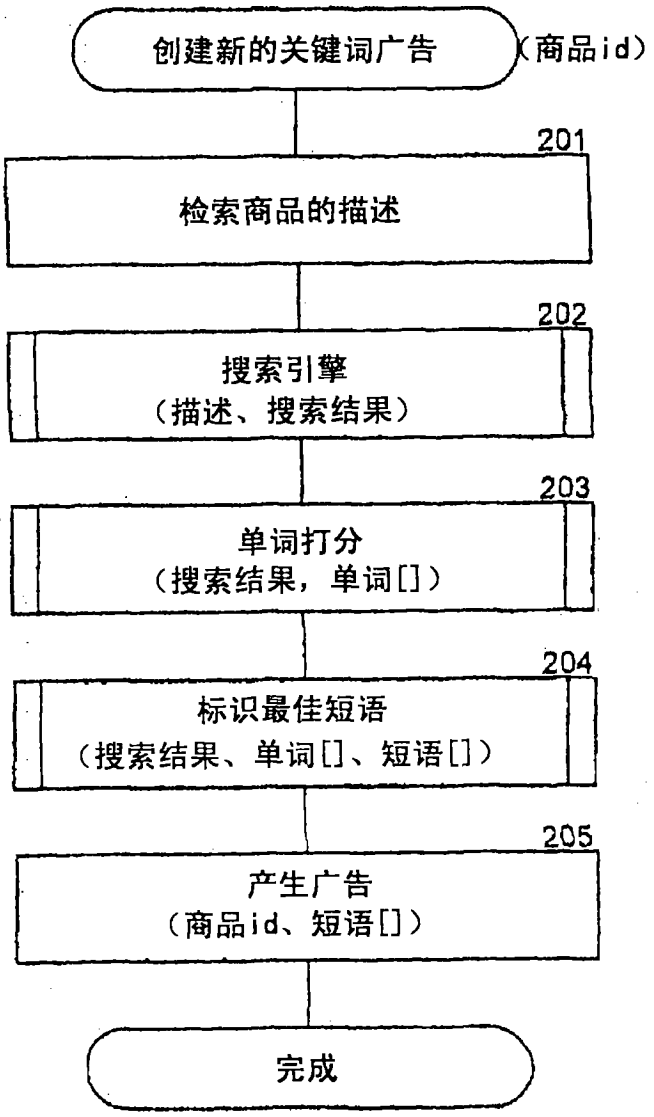

创建新的关键词广告
（商品id）
201
检索商品的描述
202
搜索引擎
（描述、搜索结果）
203
单词打分
（搜索结果，单词[]）
204
标识最佳短语
（搜索结果、单词[]、短语[]）
205
产生广告
（商品id、短语[]）
完成

图 2

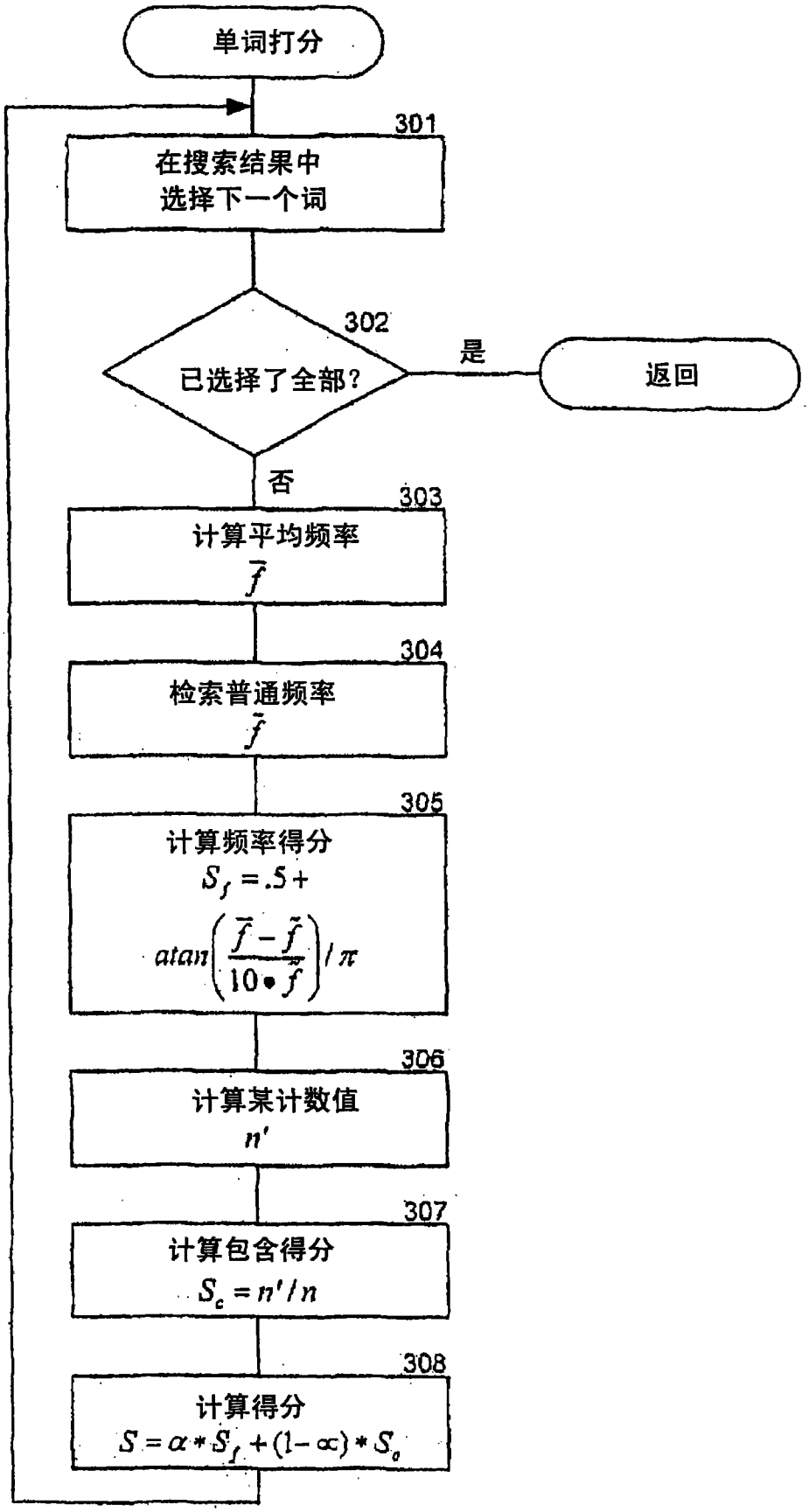

单词打分
301
在搜索结果中
选择下一个词
302
已选择了全部?
是
返回
否
303
计算平均频率
$\bar{f}$
304
检索普通频率
$\tilde{f}$
305
计算频率得分
$S_f = .5 + atan\left(\frac{\bar{f} - \tilde{f}}{10 * \tilde{f}}\right) / \pi$
306
计算某计数值
n'
307
计算包含得分
$S_c = n'/n$
308
计算得分
$S = \alpha * S_f + (1-\alpha) * S_c$

图 3

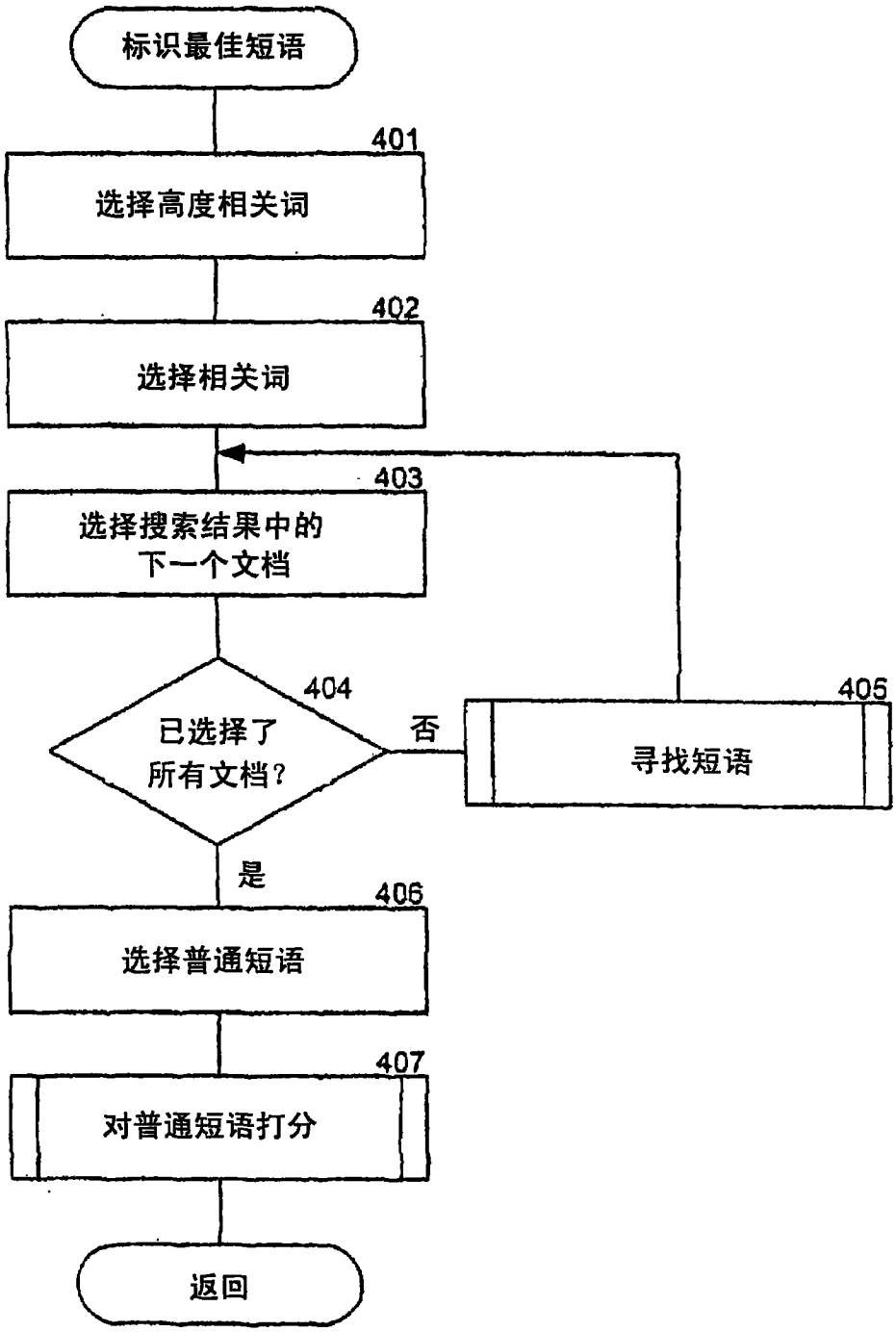

标识最佳短语
401
选择高度相关词
402
选择相关词
403
选择搜索结果中的
下一个文档
404
已选择了
所有文档？
否
405
寻找短语
是
406
选择普通短语
407
对普通短语打分
返回

图 4

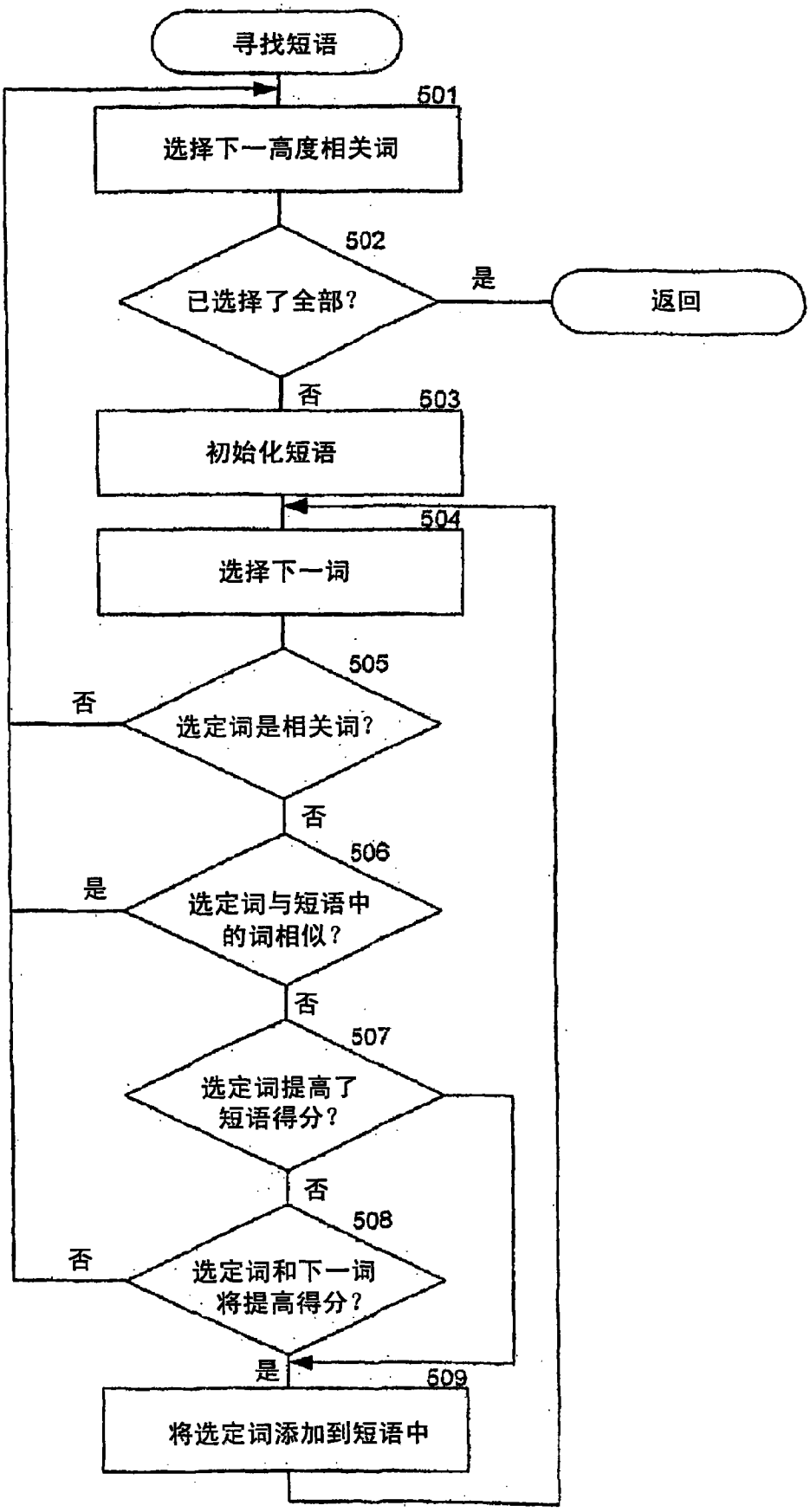

寻找短语
501
选择下一高度相关词
502
已选择了全部?
是
返回
否
503
初始化短语
504
选择下一词
505
选定词是相关词?
否
否
506
选定词与短语中的词相似?
是
否
507
选定词提高了短语得分?
否
508
选定词和下一词将提高得分?
否
是
509
将选定词添加到短语中

图 5

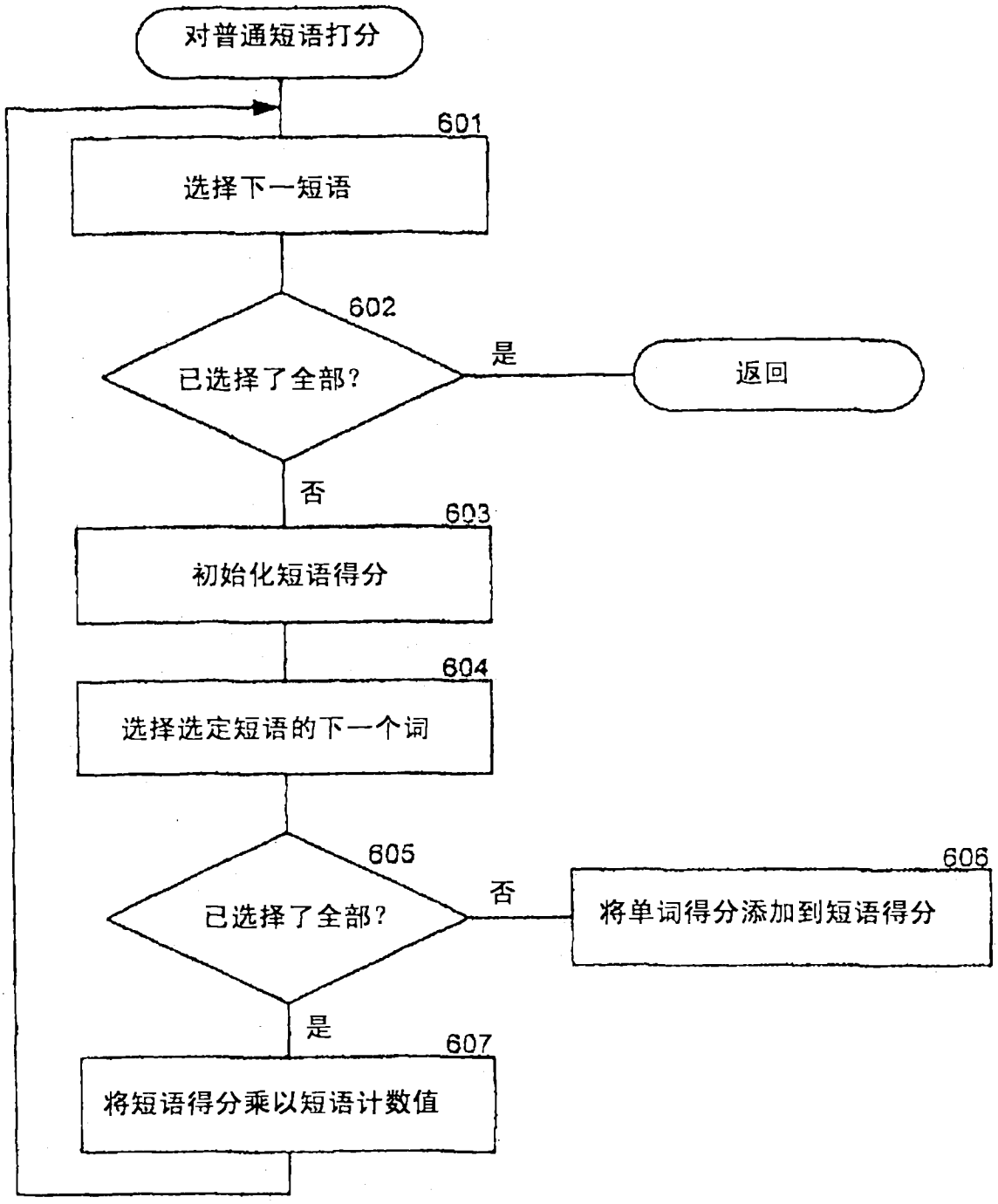

对普通短语打分
601
选择下一短语
602
已选择了全部?
是
返回
否
603
初始化短语得分
604
选择选定短语的下一个词
605
已选择了全部?
否
606
将单词得分添加到短语得分
是
607
将短语得分乘以短语计数值

图 6