



(12) 发明专利申请

(10) 申请公布号 CN 102622358 A

(43) 申请公布日 2012. 08. 01

(21) 申请号 201110029775. 8

(22) 申请日 2011. 01. 27

(71) 申请人 天脉聚源(北京)传媒科技有限公司

地址 100007 北京市东城区安定门东大街  
28号2号楼816室

(72) 发明人 伍昕 吴鹏 高晓光

(51) Int. Cl.

G06F 17/30(2006. 01)

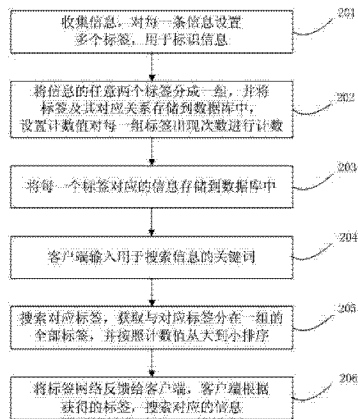
权利要求书 1 页 说明书 3 页 附图 2 页

(54) 发明名称

一种搜索信息的方法和系统

(57) 摘要

本发明公开了一种搜索信息的方法和系统,首先收集信息,对每一条信息设置不少于1个标签;将每一条信息的任意两个标签分成一组,并将每一组的两个标签及其之间的对应关系存储到数据库中,并设置计数值对一组标签出现次数进行计数;将每一个标签对应的信息存储到数据库中;客户端输入用于搜索信息的关键词;搜索数据库中的对应标签,获取与对应标签分在一组的全部标签,并按照计数值从大到小排序;将对应标签和第一级标签反馈给客户端,客户端根据获得的标签,搜索对应的信息。采用了本发明的技术方案,能够节省获取信息的时间成本,加强从不同角度对信息的认识。



1. 一种搜索信息的方法,其特征在于,包括以下步骤:
  - A、收集信息,对每一条信息设置不少于 1 个标签,用于标识信息;
  - B、将每一条信息的任意两个标签分成一组,并将每一组的两个标签及其之间的对应关系存储到数据库中,并设置计数值对数据库中每一组标签出现次数进行计数;
  - C、将每一个标签对应的信息存储到数据库中;
  - D、客户端输入用于搜索信息的关键词;
  - E、根据关键词搜索数据库中的对应标签,获取与对应标签分在一组的全部标签,作为对应标签的第一级标签,并将第一级标签按照计数值从大到小排序;
  - F、将对应标签和第一级标签反馈给客户端,客户端根据获得的标签,搜索对应的信息。
2. 根据权利要求 1 所述的一种搜索信息的方法,其特征在于,步骤 E 中,对于全部第一级标签,获取与每一个第一级标签分在一组的全部标签,作为对应标签的第二级标签,并将第二级标签按照计数值从大到小排序。
3. 根据权利要求 2 所述的一种搜索信息的方法,其特征在于,客户端预设数值,作为获取对应标签的标签级数。
4. 根据权利要求 2 或者 3 所述的一种搜索信息的方法,其特征在于,将获取的全部标签以对应标签为中心,形成标签网络,反馈给客户端。
5. 根据权利要求 1-3 中任一权利要求所述的一种搜索信息的方法,其特征在于,客户端选择一个标签,重复步骤 E 和步骤 F。
6. 一种搜索信息的系统,其特征在于,包括标签索引单元、数据库、标签挖掘单元和客户端,数据库分别与标签索引单元和标签挖掘单元连接,客户端与标签挖掘单元连接,其中,标签索引单元用于收集信息的标签;数据库用于存储标签组、标签对应的信息和标签组出现的次数;标签挖掘单元用于根据客户端输入的关键词从数据库获取对应的标签,并组织成标签网络反馈给客户端;客户端用于输入关键词、选择关键词,并接收标签挖掘单元反馈的标签网络。

## 一种搜索信息的方法和系统

### 技术领域

[0001] 本发明涉及海量信息检索技术领域,尤其涉及一种搜索信息的方法和系统。

### 背景技术

[0002] 借助互联网技术的发展,每天在我们所生活在这个世界出现了大量的信息,信息的增长速度绝对是一件近乎恐怖的事情。海量信息会让我们感到烦躁,因为面对大量的信息,我们有时会不知从何开始从何结束,当这些信息被阅读完毕,新的信息又来了。我们同时还要花费更大的精力来辨别信息内在联系,对时间成本消耗也是不容忽视的。

[0003] 在海量信息的系统上,传统的搜索引擎解决了人们在海量信息包含有特定关键词的单向信息查找方式。搜索引擎整理信息的过程称为“建立索引”。搜索引擎不仅要保存搜集起来的信息,还要将它们按照一定的规则进行编排。这样,搜索引擎根本不用重新翻查它所有保存的信息而迅速找到所要的资料。用户向搜索引擎发出查询,搜索引擎接受查询并向用户返回资料。搜索引擎每时每刻都要接到来自大量用户的几乎是同时发出的查询,它按照每个用户的要求检查自己的索引,在极短时间内找到用户需要的资料,并返回给用户。

[0004] 上述技术方案对知识探索是建立在使用系统的人本身已经形成的知识点网络之上,使用系统的人通过输入一些特定的标签或者关键词来搜索相关知识信息时,当他想获得从其他角度看待该知识点时,由于已经受限于自己知识网络,他所能得到的信息一定是不全面的。当他尝试用其它构想的关键词时,又面临搜索系统反馈回大量的无用信息造成信息过载,也就无法系统形成自己的知识网络。

### 发明内容

[0005] 本发明的目的在于提出一种搜索信息的方法和系统,能够节省获取信息的时间成本,加强从不同角度对信息的认识。

[0006] 为达此目的,本发明采用以下技术方案:

一种搜索信息的方法,包括以下步骤:

A、收集信息,对每一条信息设置不少于 1 个标签,用于标识信息;

B、将每一条信息的任意两个标签分成一组,并将每一组的两个标签及其之间的对应关系存储到数据库中,并设置计数值对数据库中每一组标签出现次数进行计数;

C、将每一个标签对应的信息存储到数据库中;

D、客户端输入用于搜索信息的关键词;

E、根据关键词搜索数据库中的对应标签,获取与对应标签分在一组的全部标签,作为对应标签的第一级标签,并将第一级标签按照计数值从大到小排序;

F、将对应标签和第一级标签反馈给客户端,客户端根据获得的标签,搜索对应的信息。

[0007] 步骤 E 中,对于全部第一级标签,获取与每一个第一级标签分在一组的全部标签,作为对应标签的第二级标签,并将第二级标签按照计数值从大到小排序。

[0008] 客户端预设数值,作为获取对应标签的标签级数。

[0009] 将获取的全部标签以对应标签为中心,形成标签网络,反馈给客户端。

[0010] 另外,客户端选择一个标签,重复步骤 E 和步骤 F。

[0011] 一种搜索信息的系统,包括标签索引单元、数据库、标签挖掘单元和客户端,数据库分别与标签索引单元和标签挖掘单元连接,客户端与标签挖掘单元连接,其中,标签索引单元用于收集信息的标签;数据库用于存储标签组、标签对应的信息和标签组出现的次数;标签挖掘单元用于根据客户端输入的关键词从数据库获取对应的标签,并组织成标签网络反馈给客户端;客户端用于输入关键词、选择关键词,并接收标签挖掘单元反馈的标签网络。

[0012] 采用了本发明的技术方案,把原来相互独立分散毫不关联的信息,通过对信息标签内在关系的建立,从而在大的信息层面把信息重新有意义的组织起来,当人们在系统上输入关键词时,会自动调出与该关键词相关的知识网络,每个与之相关的知识点都是从海量信息里挖掘出的关系,信息量越大关系也就越准确,每个知识点都可以无限探索下去,在为人们建立自己的更全面的知识网络提供更有价值的参考同时,也使获得相关信息的准确性更高。

#### 附图说明

[0013] 图 1 是本发明具体实施方式中搜索信息的系统结构示意图。

[0014] 图 2 是本发明具体实施方式中搜索信息的流程图。

#### 具体实施方式

[0015] 下面结合附图并通过具体实施方式来进一步说明本发明的技术方案。

[0016] 本发明技术方案的主要思想是通过对每一个信息进行标签化处理,把该信息最精华的部分通过多个简短短语进行唯一标识,这样当海量信息经过处理后就会形成一个基于标签的巨型知识网络。网络里每一个节点就是一个标签,每对标签之间存在一个权重的关系来判断它们之间相似性。每个标签也包含关联的具体信息,这样形成了一个标签与标签、标签与信息、信息与信息相紧密联系的立体知识网络。这个巨型网络同时也是一个动态网络,随着每个新标签加入,节点之间的关系也会相应进行调整,网络自生长并且充满活力。

[0017] 图 1 是本发明具体实施方式中搜索信息的系统结构示意图。如图 1 所示,该搜索信息的系统包括标签索引单元 101、数据库 102、标签挖掘单元 103 和客户端 104,数据库分别与标签索引单元和标签挖掘单元连接,客户端与标签挖掘单元连接。

[0018] 其中,标签索引单元收集信息的标签,数据库存储标签组、标签对应的信息和标签组出现的次数,标签挖掘单元根据客户端输入的关键词从数据库获取对应的标签,并组织成标签网络反馈给客户端,客户端输入关键词、选择关键词,并接收标签挖掘单元反馈的标签网络。

[0019] 图 2 是本发明具体实施方式中搜索信息的流程图。如图 2 所示,该搜索信息的流程包括以下步骤:

步骤 201、收集海量的信息,对每一条信息设置多个标签,用于标识该条信息。

[0020] 步骤 202、将每一条信息的任意两个标签分成一组,并将每一组的两个标签及其之

间形成的对应关系存储到数据库中,并设置计数值对数据库中每一组标签出现次数进行计数,即出现一次,计数值加 1。

[0021] 步骤 203、将每一个标签对应的信息也存储到数据库中。

[0022] 步骤 204、客户端输入用于搜索信息的关键词。

[0023] 步骤 205、根据关键词搜索数据库中的对应标签,获取与对应标签分在一组的全部标签,作为对应标签的第一级标签,并将第一级标签按照计数值从大到小排序,计数值越大,表示两个标签之间的关系越密切。

[0024] 客户端可以预设一个数值,作为获取对应标签的标签级数。例如这个数值为 2,那么对于全部第一级标签,再获取与每一个第一级标签分在一组的全部标签,作为对应标签的第二级标签,并将第二级标签按照计数值从大到小排序。

[0025] 如果这个数值为 3,还可以继续对第二级标签去获取分在一组的全部标签,作为对应标签的第三级标签,并将第三级标签按照计数值从大到小排序。

[0026] 步骤 206、将对应标签和第一级标签反馈给客户端,或者将获取的全部标签以对应标签为中心,形成标签网络,反馈给客户端,客户端根据获得的标签,搜索对应的信息。

[0027] 另外,也可以通过客户端去选择一个标签,对这个标签重复步骤 205 和步骤 206,这样只要信息足够多,就可以对每一个标签无限地探索下去,以获取与该关键词相关的知识网络。

[0028] 以上所述,仅为本发明较佳的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉该技术的人在本发明所揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应该以权利要求的保护范围为准。

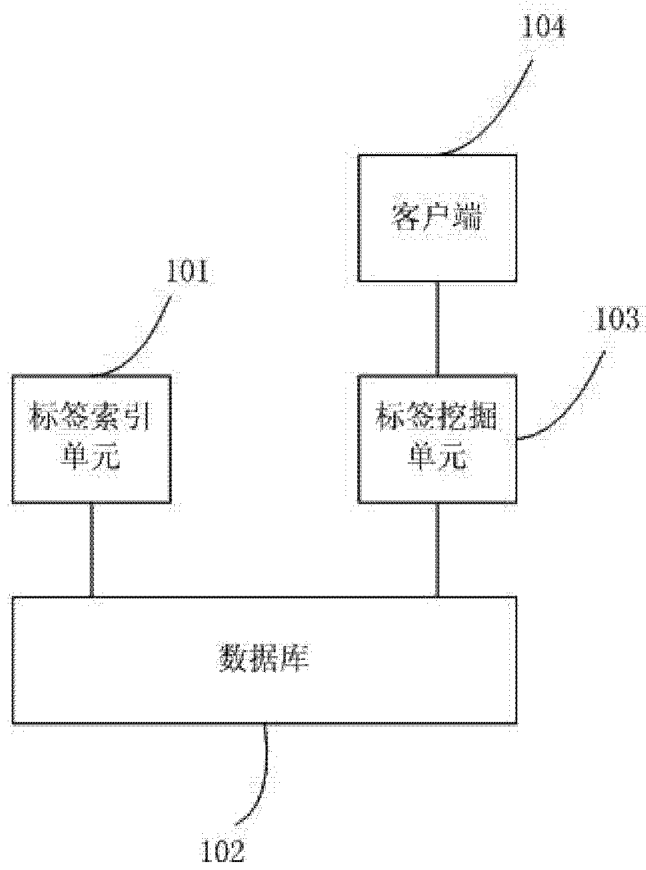


图 1

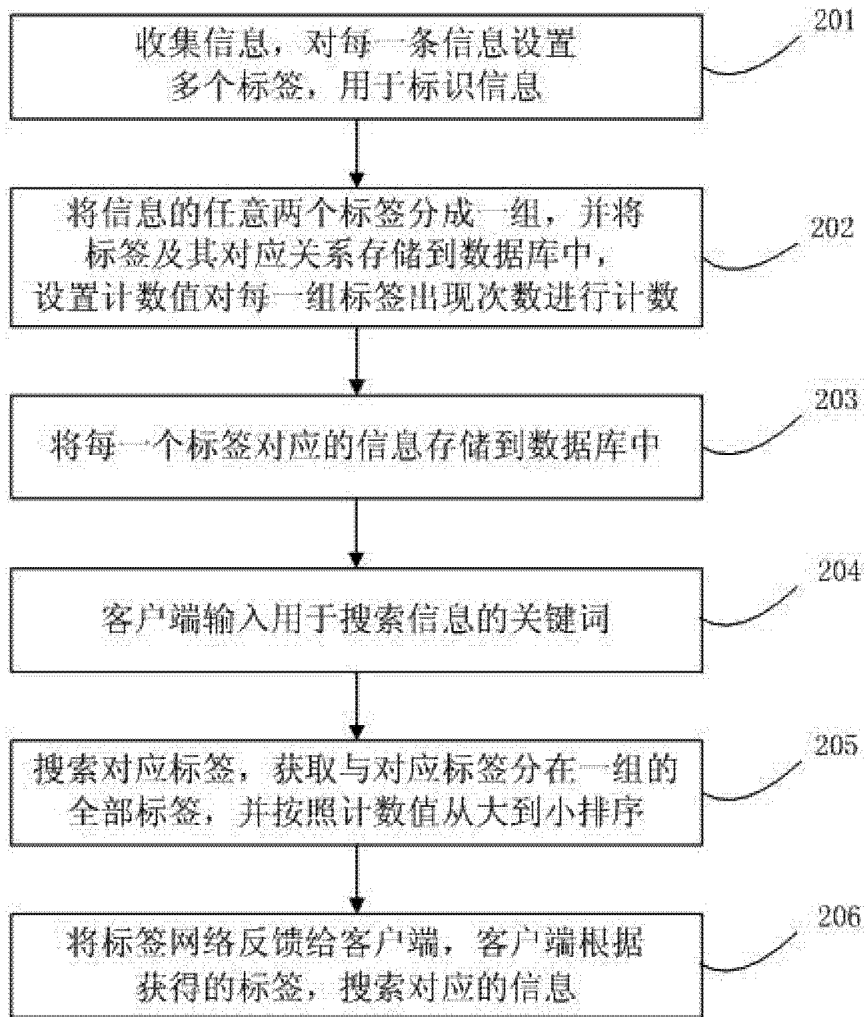


图 2