



# (12) 发明专利申请

(10) 申请公布号 CN 101957831 A

(43) 申请公布日 2011. 01. 26

(21) 申请号 200910210462. 5

(22) 申请日 2009. 11. 03

(66) 本国优先权数据

200910158038. 0 2009. 07. 17 CN

(71) 申请人 刘二中

地址 100089 北京市海淀区北洼西里中海雅园丽晶阁 2D

(72) 发明人 刘二中

(51) Int. Cl.

G06F 17/30 (2006. 01)

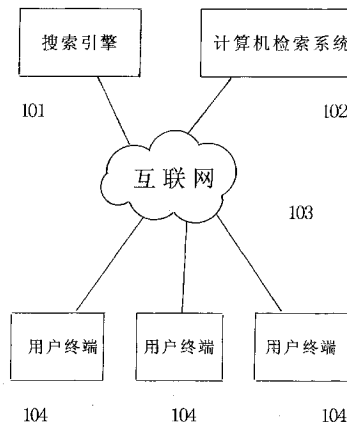
权利要求书 2 页 说明书 11 页 附图 5 页

## (54) 发明名称

文件内容的特征词的输入和处理方法

## (57) 摘要

本发明为一种计算机执行的关于网络终端用户对文件特征判定信息的输入和处理方法包括：操作 A：计算机检索系统根据终端用户提出的查询要求，向用户终端提供多个网站来源的包含符合该查询要求的文件形成的题录序列；操作 B：计算机系统根据在所述题录序列所在页面或该页面直接链接页面上的规定操作方式确定所输入的特征词；操作 C：计算机系统根据在所述题录序列所在页面或该页面直接链接页面上的规定方式，确定与操作 B 所述输入的特征词相对应的题录或文件。本发明可以在终端用户的帮助下进而形成精确的网页的特征词数据库或分类索引，进而形成更好的搜索引擎并且更为便捷地提供更为精炼的搜索结果。



1. 本发明为一种计算机执行的关于网络终端用户对文件特征判定信息的输入和处理方法,包括:

操作 A:计算机检索系统根据终端用户提出的查询要求,向用户终端提供多个网站来源的包含符合该查询要求的文件形成的题录的题录序列;

操作 B:计算机系统根据在所述题录序列所在页面或该页面直接链接页面上的规定操作方式确定所输入的特征词;

其中所述的规定操作方式为下列操作方式之一:

操作方式一:将操作 A 所述题录或其所属文件内容中受到光标选取点击的字词作为所输入的特征词;

操作方式二:将操作 A 所述题录序列所在页面或该页面直接链接的页面上呈现的或者直接链接的待选特征词目录中受到光标选取点击的字词,作为所输入的特征词;

操作方式三:在操作 A 所述题录序列所在页面或该页面直接链接的页面上设置有特征词输入栏,计算机系统根据该输入栏内的输入内容确定所输入的特征词;

本发明的输入和处理方法还包括:

操作 C:计算机系统根据在所述题录序列所在页面或该页面直接链接页面上的规定方式,确定与操作 B 所述输入的特征词相对应的题录或文件;

其中所述规定方式为下列方式之一:

方式 I:将操作 B 中操作方式一所述受到光标选取点击的字词所在的题录或文件,确定为与输入的特征词相对应的题录或文件;

方式 II:将受到点击的题录或文件确定为与输入的特征词相对应的题录或文件;

方式 III:将受到点击的特征词判定操作标示附近的题录或文件,确定为与输入的特征词相对应的题录或文件;

方式 IV:将特征词输入栏所在的页面上距离该输入栏最近的或位于该输入栏规定方位的题录或文件,确定为与所输入的特征词相对应的题录或文件;

方式 V:将特征词输入栏所在的页面上唯一的题录或文件,确定为与输入的特征词相对应的题录或文件;

方式 VI:将操作 B 所述操作方式二所述待选特征词目录所在的页面上距离该目录最近的或位于该目录规定方位的题录或文件,确定为与所输入的特征词相对应的题录或文件;

方式 VII:将操作 B 所述操作方式二所述待选特征词目录所在的页面上唯一的题录或文件,确定为与输入的特征词相对应的题录或文件。

2. 按照权利要求 1 所述的输入和处理方法,其中:所述的特征词是反映相对应题录或文件内容特点的关键词。

3. 按照权利要求 1 所述的输入和处理方法,其中:所述的特征词是在相对应题录或文件中出现的字词。

4. 按照权利要求 1 所述的输入和处理方法,其中:所述的特征词是反映相对应题录或文件内容类别的类别词。

5. 按照权利要求 1 所述的输入和处理方法,还包括:所述计算机系统在根据终端用户输入意见确定与任一题录或文件相对应的特征词时,需要遵循的原则至少考虑以下因素中的一个或多个:

(1) 作出判定的用户的名称或其网站的网址与该文件的提供者的名称或其网址或者文件链接网址的相似程度；

(2) 作出同样判定的用户的数量；

(3) 作出某种判定的时间早晚；

(4) 作出判定的用户或出自同一网址以往点击选择的准确率或评分；

(5) 该种特征词的选择与其它人工选择方法或计算机选择方法或选择系统结果的一致程度；

(6) 是否为检索系统操作人员或工作人员所作判定或与其相似。

(7) 作出判定的用户或终端是否在进行特征词判定或选择的相关网站或网页登记。

6. 按照权利要求 1 所述的输入和处理方法, 还包含：

操作 D1 :检索系统根据所述方法确定的与多个题录或文件相对应的特征词的数据, 生成包含多个文件或题录的特征词内容的数据库。

7. 按照权利要求 1 或 5 或 6 所述的输入和处理方法, 还包含：

操作 D2 :检索系统根据所述方法确定的多个文件或题录的特征词的数据或者操作 D1 生成的包含多个文件或题录的特征词内容的数据库, 生成多个文件或题录的特征词索引。

8. 按照权利要求 1 所述的输入和处理方法, 还包括：

操作 E :计算机检索系统在提供搜索服务时, 根据网络查询用户提出的查询要求, 向用户终端提供的多个文件的题录的序列 ;在所述的题录序列部分或全部各个题录附近, 分别具有各个题录或其所属文件所属的一个或多个特征词的提示。

9. 按照权利要求 1 所述的输入和处理方法, 还包括：

操作 F :令操作 E 所述的题录附近的各个特征词提示, 分别与另外多个文件题录的序列链接 ;所述链接的另外多个文件题录的序列中的部分或全部题录或者题录所属的文件, 至少各有一个所属的特征词, 与该序列链接的原来的特征词提示中的特征词相同。

10. 按照权利要求 1 所述的输入和处理方法, 还包括：

操作 G :在计算机检索系统根据网络查询用户提出的查询要求向用户终端提供的多个文件题录序列附近, 具有多个特征词提示组成的导航目录, 其中各个特征词提示分别与各个不同的包含多个文件题录的序列链接, 所述链接的另外多个文件题录的序列中的部分或全部题录或者题录所属的文件, 至少各有一个所属的特征词, 与该序列链接的原来的特征词提示中的特征词相同。

## 文件内容的特征词的输入和处理方法

### 技术领域

[0001] 本技术属于计算机检索技术或搜索引擎技术。

### 背景技术

[0002] 多年来,计算机数据库检索技术有了极大发展,特别是网络技术的进展,使得人们可以共享的数据库的规模达到了天文数字。这也给人们查找所需信息带来了很大困难。

[0003] 以查询词搜索为核心的搜索引擎技术为用户带来了便利。该系统可以通过客户机上的交互界面以及通讯网络得到查询者的关键词查询请求,在文本索引库或文本库中进行查询,并进行关键词请求与文本的相关性分析,得到相关结果并排序,再经由通讯网络或线路提供到交互界面。这种搜索系统使用起来十分便利迅速,但返还结果包含的索引总数仍然十分庞大,难于逐一查阅。

[0004] 为了能将潜在的对查询者最有价值的查询结果尽量排在前面以方便查询者,第6,285,999号美国专利提出了基于网页超级链接结构分析(佩奇链接)来进行搜索结果排序的技术,超过了其他排序技术,获得空前成功。

[0005] 然而,该技术以及其他各种排序技术,仅仅是在统计学意义上提高了关键词搜索的效率,并不能保证每个人希望的查询结果都能排在庞大索引表的前面。我们在读到期望的信息之前,却无奈地读到种种主要内容一再重复的无关信息。

[0006] 为了方便用户找到所需信息或文件,人们还求助于垂直分类技术和以该技术为基础的目录检索系统。为了给海量信息分类或确定特征,出现了各种计算机文本分类方法。然而,由机器判断某一页面或文本属于某关键词的哪一条或哪几条的语义或特征或类别是十分困难的,其可靠性和准确率不高,特别是在多级分类中,错误率高得令人无法忍受。因此,计算机分类仅仅用于最简单的粗略分类,例如根据若干组词出现的频度或格式特征,判定网上文件是“网页”还是“地图”或是“MP3”等等。

[0007] 目前,准确率较高的垂直分类技术还离不开人工参与。例如20世纪90年代雅虎等网站的人工信息分类系统,只能承担极少一部分网上信息的分类加工。其它如“百度百科”、“维基百科”、“淘宝网”、“阿里巴巴”的各种数量十分有限的专业分类信息,都是通过各自专门的数据库平台,由注册会员或注册用户或网站工作人员针对特定范围的词条,按照特殊的编写规则汇集而成的,其中包含的附带分类内容也只能是针对本数据库的词条或文本的。可以说,对于非特定网上数据库内容的用户来说,在搜索时得到的帮助十分有限。

[0008] 因此,国内外广大网民迫切需要一种新技术,使得计算机检索系统或搜索引擎系统不仅能向查询者提供数万不同网站的亿万网页题录信息,还能确定众多不同网站来源网页的准确特征或类别或者多级特征或类别,并按查询者的查询关键词要求和对网页特征或类别的要求,提供准确率和集中度大为改善的搜索结果。为此,急需一种便于汇集和处理关于网页特征判断意见的便捷技术。

## 发明内容

[0009] 本发明的目的在于提供一种适于计算机检索系统或搜索引擎系统使用的方法,使其在向网上终端或用户提供关于查询词搜索结果的题录序列的同时,可以允许便捷地输入用户或工作人员为不同网站来源的相关文件确定的特征词,并对输入信息进行处理,以便产生便于用户利用的包含不同特征词或分类结果的检索工具,大大提高检索或搜索的效率。

[0010] 本发明为一种计算机执行的关于网络终端用户对文件特征判定信息的输入和处理方法,包括:

[0011] 操作 A:计算机检索系统根据终端用户提出的查询要求,向用户终端提供多个网站来源的包含符合该查询要求的文件形成的题录的题录序列;

[0012] 操作 B:计算机系统根据在所述题录序列所在页面或该页面直接链接页面上的规定操作方式确定所输入的特征词;

[0013] 其中所述的规定操作方式为下列操作方式之一:

[0014] 操作方式一:将操作 A 所述题录或其所属文件内容中受到光标选取点击的字词作为所输入的特征词;

[0015] 操作方式二:将操作 A 所述题录序列所在页面或该页面直接链接的页面上呈现的或者直接链接的待选特征词目录中受到光标选取点击的字词,作为所输入的特征词;

[0016] 操作方式三:在操作 A 所述题录序列所在页面或该页面直接链接的页面上设置有特征词输入栏,计算机系统根据该输入栏内的输入内容确定所输入的特征词。

[0017] 其中,特征词输入栏的输入内容可以来自键盘,也可以来自对操作 A 所述题录所在页面或该页面直接链接页面或特征词输入栏所在页面上部分字词的粘贴。

[0018] 需要时,为了操作简捷,可以限定特征词输入栏仅仅出现在所述题录序列所在页面上。

[0019] 其中,所述特征词可以为终端用户所选择的认为可以反映相应题录或文件内容特点的一个或多个字词或词组。所述字词可以是字符或符号或音符或图形。

[0020] 所述输入栏是指终端页面上输入或填写字词的空间或位置。

[0021] 所述的计算机检索系统可以是搜索引擎系统。所述计算机系统或检索系统可以为计算机检索系统的组成部分。

[0022] 所述终端用户可以是网民或网页的作者或网页的提供者,或者网络或检索系统工作人员。

[0023] 所述文件可以是网页或者网页的部分内容或者是检索系统或其它计算机系统转存内容(如网页快照),可以是或包含文字内容,也可以是或包含图象内容或音频内容或视频内容。

[0024] 所述题录可以是文件的标题或摘要或标题加摘要,可以是或包含图象内容或音频内容或视频内容。

[0025] 本发明的输入和处理方法还包括:

[0026] 操作 C:计算机系统根据在所述题录序列所在页面或该页面直接链接页面上的规定方式,确定与操作 B 所述输入的特征词相对应的题录或文件;

[0027] 其中所述规定方式为下列方式之一:

[0028] 方式 I:将操作 B 中操作方式一所述受到光标选取点击的字词所在的题录或文件,

确定为与输入的特征词相对应的题录或文件；

[0029] 方式 II :将受到点击的题录或文件确定为与输入的特征词相对应的题录或文件；

[0030] 方式 III :将受到点击的特征词判定操作标示附近的题录或文件,确定为与输入的特征词相对应的题录或文件；

[0031] 方式 IV :将特征词输入栏所在的页面上距离该输入栏最近的或位于该输入栏规定方位的题录或文件,确定为与所输入的特征词相对应的题录或文件；

[0032] 方式 V :将特征词输入栏所在的页面上唯一的题录或文件,确定为与输入的特征词相对应的题录或文件；

[0033] 方式 VI :将操作 B 所述操作方式二所述待选特征词目录所在的页面上距离该目录最近的或位于该目录规定方位的题录或文件,确定为与所输入的特征词相对应的题录或文件；

[0034] 方式 VII :将操作 B 所述操作方式二所述待选特征词目录所在的页面上唯一的题录或文件,确定为与输入的特征词相对应的题录或文件。

[0035] 可以根据需要规定操作 B 与操作 C 的先后次序。

[0036] 我们可以将与某题录或某文件相对应的特征词,称为属于该题录或该文件的特征词,或者称为该题录或该文件对应的特征词,或者称为该题录或该文件的特征词。

[0037] 在以上所述方法中,允许同一文件或其题录可以同时具有多个不同的类别词,一种特征词可以同时属于多个不同题录或文件。

[0038] 一般可以认为,一个题录的特征词与该题录所属的文件的特征词相同。

[0039] 所述的特征词可以是反映相对应题录或文件内容特点的关键词,也可以是在相对应题录或文件中出现的词或关键词,所述的特征词输入栏也可以是关键词输入栏。

[0040] 所述的特征词可以是反映相对应题录或文件内容类别的类别词,或是反映其在多级分类系统中不同级别的类别词,所述的特征词输入栏也可以是类别词输入栏。

[0041] 在所述的输入和处理方法中,在所述题录序列所在页面或该页面直接链接页面上,设置附加的待选特征词目录。

[0042] 该待选特征词目录可以是包含多个不同类别词的类别目录。所述待选特征词或类别目录可以是一级目录或多级目录或树状目录。

[0043] 可以安排:在所述类别目录中,上一级类别条目受到点击之前或之后可自动显示其所属下一级类别条目。

[0044] 在本方法中,可以允许通过在所述设置的所述目录中光标点击或刷取所需字词的方式向所述输入栏输入特征词。

[0045] 显然,本方法所述输入的特征词,就是终端点击操作者输入的对相关题录或文件特征的判定信息或与其相对应的特征词。

[0046] 本方法还可以包括:相关计算机系统在其数据库里,可以接受或参考或处理或拒绝终端用户所输入的特征判定意见或特征词或类别词。

[0047] 本发明的输入和处理方法,还可以包括:所述计算机系统或数据库在根据终端用户输入意见确定或输入与任一题录或文件相对应的特征词或类别词时,需要遵循的原则可以至少考虑以下因素中的一个或多个:

[0048] (1) 作出判定的用户的名称或其网站的网址与该文件的提供者的名称或其网址或

者文件链接网址的相似程度；

[0049] (2) 作出同样判定的用户的数量；

[0050] (3) 作出某种判定的时间早晚；

[0051] (4) 作出判定的用户或出自同一网址以往点击选择的准确率或评分；

[0052] (5) 该种特征词的选择与其它人工选择方法或计算机选择方法或选择系统结果的一致程度；

[0053] (6) 是否为检索系统操作人员或工作人员所作判定或与其相似。

[0054] (7) 作出判定的用户或终端是否在进行特征词判定或选择的相关网站或网页登记。

[0055] 本发明的方法还可以包含操作 D 1:检索系统全部或部分根据所述方法确定的与多个题录或文件相对应的特征词的数据,生成包含多个文件或题录的特征词内容或者以其特征词或类别词的异同分类的数据库。

[0056] 本发明的方法还可以包含操作 D2:检索系统全部或部分根据所述方法确定的与多个文件或题录相对应的特征词的数据或者操作 D1 生成的包含多个文件或题录的特征词或类别词内容的数据库,生成多个文件或题录的特征词索引或类别词索引或分类索引。

[0057] 所述特征词索引可以是指,利用该索引可以根据选择的任一特征词去检索或访问或链接对应于该特征词的文件或其题录或其地址或其相关信息。

[0058] 所述分类索引可以是指,利用该索引可以根据选择的任一类别词去检索或访问或链接对应于该类别词的文件或其题录或其地址或其相关信息。

[0059] 本发明的输入和处理方法,还可以包括:利用该方法对多个文件的特征词分类或分类索引替代或修改其它原有的对多个文件的分类或分类索引。

[0060] 本发明的方法,还可以包括:在接受查询时,检索系统利用所述特征词索引或分类索引,提供符合所需特征词或类别词要求的检索或搜索结果。该结果可以包括题录或题录序列或目录或树状目录。

[0061] 本发明的输入和处理方法,还可以包括:在接受查询时,检索系统利用所述特征词索引或分类索引以及计算机检索系统处理终端用户提出的查询要求时所利用的查询词索引或关键词索引,得到或提供既符合所需特征词要求又符合所需查询要求的检索或搜索结果。该结果可以包括题录或题录序列或目录。

[0062] 本发明的输入和处理方法,还可以包括:

[0063] 操作 E:计算机检索系统在提供搜索服务时,根据网络查询用户提出的查询要求,向用户终端提供的多个文件的题录的序列;在所述的题录序列部分或全部各个题录附近,可以分别具有各个题录或其所属文件所属的一个或多个特征词的提示。

[0064] 所述特征词提示可以是该特征词或者是包含该特征词的提示。

[0065] 本方法允许根据终端用户的操作,增加或减少或替换所述特征词提示。

[0066] 所述的特征词提示可以是反映相对应题录或文件内容特点的关键词的提示,也可以是在相对应题录或文件中出现的关键词的提示,需要时,允许关键词的提示出现在操作 E 所述题录的字里行间。

[0067] 所述的特征词提示也可以是类别词提示,可以是单级或多级分类系统的类别提示。

[0068] 本发明的输入和处理方法,还可以包括:

[0069] 操作 F:可以令操作 E 所述的题录附近的各个特征词提示,分别可以与另外多个文件题录的序列链接;所述链接的另外多个文件题录的序列中的部分或全部题录或者题录所属的文件,至少各有一个所属的特征词,与该序列链接的原来的特征词提示中的特征词相同。

[0070] 需要时,可以要求操作 F 所述链接的另外多个文件题录的序列中的部分或全部题录或者题录所属的文件,还要符合操作 E 所述用户原来提出的查询要求。

[0071] 本发明的输入和处理方法,还可以包括操作 G:在计算机检索系统根据网络查询用户提出的查询要求向用户终端提供的多个文件题录序列附近,具有多个特征词提示组成的导航目录,各个特征词提示可以分别与各个不同的包含多个文件题录的序列链接,所述链接的另外多个文件题录的序列中的部分或全部题录或者题录所属的文件,至少各有一个所属的特征词,与该序列链接的原来的特征词提示中的特征词相同。

[0072] 需要时,可以要求操作 G 所述链接的另外多个文件题录的序列中的部分或全部题录或者题录所属的文件,还要符合所述查询用户原来提出的查询要求。

[0073] 所述导航目录的特征词提示可以是反映相对应题录或文件内容特点的关键词的提示,也可以是在相对应题录或文件中出现的关键词的提示,也可以是类别词提示。

[0074] 所述的导航目录可以是一级目录或多级目录。可以允许该目录的上一级特征词选择确定后再自动显示下一级待选的多个特征词提示。

[0075] 本方法允许根据终端用户的操作,增加或减少或替换该目录特征词提示。

[0076] 本方法还允许在提供操作 F 和操作 G 所链接或展示的题录附近或题录序列附近,具有特征词提示或导航目录,以链接或供点击展示更新的题录序列结果。

[0077] 本发明的方法,为搜索引擎系统所能汇集的来自千百万不同网站的亿万网页题录的特征词确定问题,提供了一个可以根本性解决手段。任何网民甚至包括网络系统工作人员,特别是网页的提供者或作者或推广者,在搜索引擎的关键词搜索结果的题录序列中看到与自己的利益或兴趣相关的文件题录时,利用本发明的技术,可以十分便捷地确定或输入该文件的特征词或关键词或类别词。具有多个准确的特征词的网页更容易被优先搜索到,这样,有价值的网页多数会有相关的内行人士为其确定特征词。本发明的方法还可以保证文件相关人的输入意见会得到优先采纳。在本发明的基础上,搜索引擎系统可以为相当比例的优质网页提供高质量的特征词检索服务,甚至多级分类检索服务,得到高度集中或高度浓缩的搜索结果,大大提高广大网民网上搜索的效率,解决多年困扰网民的难题,因而本技术具有突出的实用价值和效果。

#### 附图说明

[0078] 图 1 为本发明的实施例的适用环境示意图。

[0079] 图 2 为本发明的一个实施例的在题录序列页面输入特征词的说明示意图。

[0080] 图 3 为本发明的一个实施例的用户查询搜索结果的题录序列页面的题录或其所属文件附带的对应特征词提示(关键词提示)和导航目录的示意图。

[0081] 图 4 为本发明的另外一个实施例的用户查询搜索结果的题录序列页面的题录或所属文件附带的特征词提示(多级类别词提示)和导航目录的示意图。



[0082] 图 5 为本发明的一个实施例的实现方法的流程示意框图。

[0083] 具体的实施方式

[0084] 以下结合附图说明具体的实施方法。其中,搜索引擎系统 101 为计算机检索系统 102 的一种专门类型。它们通过互联网 103 与用户终端 104 联系(参见图 1)。

[0085] 在图 2、图 3、图 4 的实施例中,201 为输入查询词的查询栏,202 为题录,203 为特征词输入栏,204 为特征词判定操作标示,205 为鼠标装置操作的光标,206 为参考目录,208 为所述方式一所述受到光标选取点击的字词;301 为特征词提示(关键词提示),302 为选定操作标示,303 为增项操作标示,304 为导航目录;401 为特征词提示(类别词提示)。

[0086] 举例来说,实施本方法(参见图 5),应该从操作 A 开始,首先需要相关检索系统或搜索引擎系统接受网络查询用户或终端用户在查询栏 201 输入查询要求(流程 501),向用户终端提供查询搜索服务,即向其提供多个网站来源的符合查询要求的多个文件形成的题录 202 组成的或参与组成的序列(流程 502)。

[0087] 所述文件可以是网页,可以包含文字内容,也可以包含图象内容或音频内容或视频内容。

[0088] 所述题录可以是文件的标题或摘要或标题加摘要或部分内容或转存的内容,如网页快照、缓存网页等。

[0089] 所述文件的题录也可以包含各类内容,例如图象的缩略内容,音节或乐谱片段,或音频或视频的片段或缩略内容,或者截屏或截屏局部画面。

[0090] 本发明的方法对图象内容或音频内容或视频内容的网页或文件的分类或建立分类索引,更具有特别重要的意义。

[0091] 本方法还需要进行操作 B:计算机系统确定终端用户所输入的特征词(流程 503)。

[0092] 所述特征词由终端用户认定的或输入的字词,可以反映相应题录或文件特点,如关键词或类别词,可以是字符或符号或音符或图形或图形标记,需要时,例如可以是与音频文件或视频文件有关的音节或乐谱片段。

[0093] 输入特征词(或关键词或类别词)的具体操作方式主要有 3 种。其一是将操作 A 所述题录或其直接链接的文件内容中受到光标选取点击的字词(208)作为所输入的特征词。

[0094] 所谓选取点击可以是令点击状态的光标在相关字词上滑移,也可以是约定的其它操作方式。具体实施时最好在此前或此后配合点击特征词判定操作标示 204,或者以其它方式使终端页面处于特征词操作状态,以利于计算机识别。

[0095] 所述特征词判定操作标示(简称操作标示),是指用以接受点击进入特征词操作状态或者用以标示特征词判定对应的题录或文件或者用以链接待选特征词目录或其它相关操作的字符或标示或图形或图形键。例如图 2 的“设定特征词”204 或“链接特征词目录”或“划类操作标示”或“参与分类”的字样等等。

[0096] 另一种输入特征词的方式,是将操作 A 所述题录序列所在页面上或该页面直接链接的页面上呈现的或者直接链接的待选特征词目录 206(如图 2 中“参考目录”206)中受到光标选取点击的字词,作为所输入的特征词。

[0097] 页面直接链接的页面或目录,是指题录序列所在页面上的题录或特征词判定操作标示或目录标题或提示或其它词项或内容所链接的页面或目录。

[0098] 需要时,可以令待选特征词目录在终端页面处于特征词操作状态时或其它需要的时候在页面出现。

[0099] 第三种操作方式是在操作 A 所述题录序列所在页面或该页面直接链接的页面上设置有特征词输入栏 203 或输入框,计算机系统根据该输入栏内的输入内容确定所输入的特征词。

[0100] 在本方法中,特征词输入栏的输入内容可以来自键盘,也可以来自对操作 A 所述题录所在页面或该页面直接链接页面上部分字词的粘贴,或者可以允许通过在所述设置的所述题录或文件或待选特征词目录中光标点击或刷取所需字词的方式向所述输入栏 203 输入特征词。

[0101] 实际上,特征词输入栏可以是输入框,或者是在所述页面上相应标记或提示字词(例如“特征词输入”或“特征词”或“关键词”或“类别”)附近的局部空间。

[0102] 需要令终端页面处于特征词操作状态,可以通过查询系统预先设置,或由终端用户点击选择。需要时,也可以规定,当页面上的特征词判定操作标示 204 或待选特征词目录 206 或特征词输入栏 203 受到点击后或具有输入内容时,终端页面进入或处于特征词操作状态。

[0103] 本方法还需要计算机系统根据在所述题录序列所在页面或该页面直接链接页面上的下列方式之一,确定与操作 B 所述输入的特征词相对应的题录或文件(操作 C)(流程 503)。

[0104] 具体来说,可以使用方式 I:将操作 B 中操作方式一所述受到光标选取点击的字词 208 所在的题录 202 或文件,确定为与输入的特征词相对应的题录或文件。此时终端页面应该已处于特征词操作状态,以利于避免与其它链接操作混淆。

[0105] 或者方式 II:将受到点击的题录或文件确定为与输入的特征词相对应的题录 202 或文件。此时终端页面最好已处于特征词操作状态。

[0106] 或者方式 III:将受到点击的特征词判定操作标示 204 附近的题录或文件,确定为与输入的特征词相对应的题录或文件。此时题录或文件最好与特征判定操作标示 204 一一对应。

[0107] 或方式 IV:将特征词输入栏所在的页面上距离该输入栏最近的或位于该输入栏规定方位(例如上方或下方)的题录或文件,确定为与所输入的特征词相对应的题录或文件。

[0108] 或者方式 V:将特征词输入栏所在的页面上唯一的题录或文件,确定为与输入的特征词相对应的题录或文件。

[0109] 或者方式 VI:将操作 B 所述操作方式二所述待选特征词目录所在的页面上距离该目录最近的或位于该目录规定方位(例如左方或右方)的题录或文件,确定为与所输入的特征词相对应的题录或文件。

[0110] 或方式 VII:将操作 B 所述操作方式二所述待选特征词目录所在的页面上唯一的或题录或文件,确定为与输入的特征词相对应的题录或文件。

[0111] 实际上,可以根据需要安排操作 B 与操作 C 的先后次序以及终端用户的操作规则。

[0112] 所述待选特征词目录可以由终端用户在输入特征词时参考或选用的多个字词组成。根据待选特征词的不同,该目录可以具有“参考目录”或“类别目录”或“关键词建议”

之类的标题。

[0113] 在一种实施例中,我们可以在每一条题录下侧设置“推荐关键词:”或者“选定类别词;”的字样,形成输入栏,以便于用户输入。为了避免误操作,输入栏后侧还可以有“选毕”字样,供点击确认。这样,用户只需要在相应题录的输入栏内输入或“粘贴”入关键词或类别词,再点击“选毕”,就完成了对该文件特征词的确定工作。该实施例利用了所述操作方式三以及方式 IV。

[0114] 在本方法的另一个实施例中,在每一条题录的下侧或末尾都具有“分类”的字样(特征词判定操作标示)。用户点击该字样后,页面一侧会出现一个待选类别词目录的最高级的多个类别词。用户点击其中的类别词后,该目录中会出现属于该类别的下一级多个类别词,供用户选择点击。依此类推,用户选择完毕后点击“选定”字样,系统将自动把该题录的多级分类的各个类别词输入进去。该实施例利用了所述操作方式二以及方式 III。

[0115] 在具体的实施过程中,还可以利用操作方式一以及方式 I,

[0116] 或利用操作方式二以及方式 II,

[0117] 或利用操作方式二以及方式 VI,

[0118] 或利用操作方式二以及方式 VII

[0119] 或利用操作方式三以及方式 III,

[0120] 或利用操作方式三以及方式 V,

[0121] 或利用操作方式三以及方式 VI,

[0122] 或利用操作方式三以及或方式 VII,为题录序列中相应的题录或相应文件确定其特征词或关键词或类别词。

[0123] 本方法允许提供一种或多种或多套特选特征词目录或关键词目录或待选类别词目录或多级类别词目录,以供用户终端选用。

[0124] 在一般情况下,可以认为,某个题录的特征词与该题录所属文件的特征词相同或相似,可以直接根据某个题录的特征词得到该题录所属文件的特征词,或者反过来进行判定。

[0125] 显然,本方法所述输入的特征词,就是终端点击操作者输入的对相关题录或文件特征的判定信息或与其相对应的特征词。

[0126] 本方法还可以包括相关计算机系统在其数据库里,可以接受或参考或处理或拒绝终端用户所输入的特征判定意见或特征词或类别词。

[0127] 这样,根据操作 A、B、C,就输入了终端点击操作者对相关题录或文件的特征词或关键词或类别词信息,即 504 “按输入确定题录文件特征词”。计算机系统或检索系统可以直接利用这些信息,但有可能还需要对输入的划类信息进行处理。

[0128] 显然,检索系统根据互联网用户的点击选择来确定每个文件的特征词还存在一个问题:如果多个用户或终端操作者作出了不同选择,应该怎么办?这就是图 5 的流程 505 “不同输入意见处理”要解决的问题。

[0129] 检索系统面对用户或终端操作者输入的可能的矛盾意见,确定或输入与任一题录或文件相对应的特征词或类别词时,需要遵循的原则可以至少考虑以下因素中的一个或多个:

[0130] (1) 作出判定的用户的名称或其网站的网址与该文件的提供者的名称或其网址或

者文件链接网址的相似程度；

[0131] 越相似,类别选择的用户与原来文件的提供者一致的可能性就越大。

[0132] (2) 作出同样判定的用户的数量；

[0133] 同样意见的用户数越多,意见越可靠。

[0134] (3) 作出某种判定的时间早晚；

[0135] 为了尽快形成分类索引,不能等待太久;但是后来的修改意见可能更为中肯。

[0136] (4) 作出判定的用户或出自同一网址以往点击选择的准确率或评分；

[0137] 应该更重视高水平用户的意见。

[0138] (5) 该种特征词的选择与其它人工选择方法或计算机选择方法或选择系统结果的一致程度；

[0139] 这样既可以参考已有成果,也可以避免改变过多。

[0140] (6) 是否为检索系统操作人员或工作人员所作判定或与其相似。

[0141] (7) 作出判定的用户或终端是否在进行特征词判定或选择的相关网站或网页登记。

[0142] 实际上,需要时可以优先考虑(1)或(6)或(7),再考虑其它因素。

[0143] 也可以编写某种目标函数的代数式,该函数式的变量至少包括上述7种因素的一个或多个。可以按照目标函数值的大小确定不同分类的优先顺序。

[0144] 针对任一题录或文件的特征词(特别是关键词)可能数量很多,可以参考以上因素安排优先顺序,并适度限制保留或提供的最高数量。

[0145] 实际上,针对任一题录或文件的同一级类别词或分类选择并不一定只有一个,可以有两个或更多,可以有优先顺序。可以限定对应于任一题录或文件的每一级类别词的数量,例如2或3种。

[0146] 本发明的方法还可以包含506“形成文件特征词数据、索引”:检索系统全部或部分根据所述方法确定的与多个题录或文件相对应的特征词的数据,生成包含多个文件或题录的特征词内容或者以其特征词或类别词的异同分类的数据库;以及生成多个文件或题录的特征词索引或类别词索引或分类索引或关键词索引,或者人们熟知的特征词或关键词或类别词倒排索引,或者倒排题录索引。

[0147] 所述特征词索引可以是指,利用该索引可以根据选择的任一特征词去检索或访问或链接对应于该特征词的文件或其题录或其地址或其相关信息。

[0148] 利用分类索引可以根据选择的任一类别词去检索或访问或链接对应于该类别词的文件或其题录或其地址或其相关信息。

[0149] 需要时,还可以按照各个题录或文件的特征词或关键词或类别词的不同,生成包含多个不同子集或多级子集的分类文件或题录数据库或多级分类索引。

[0150] 本发明的方法,还可以包括:利用该方法对多个文件的特征词分类或分类索引替代或修改其它原有的对多个文件的分类或分类索引。

[0151] 本方法还可以包括:在接受查询时,检索系统利用所述特征词索引或分类索引,提供符合所需特征词或类别词要求的检索或搜索结果。该结果可以包括题录或题录序列或目录或树状目录。

[0152] 在流程504或流程506,如果终端用户希望开始对其它题录或文件进行特征词判

定,可以返回(流程 510)至流程 501。

[0153] 显然,本发明的目的不仅仅是建立有关文件的特征词数据库或特征词倒排索引。本发明的方法还包括任何利用这些索引或数据进行题录搜索。

[0154] 因此,本方法还可以包括:

[0155] 计算机检索系统根据网络查询用户提出的查询要求,向用户终端提供的多个文件的题录序列;在所述的题录序列部分或全部各个题录附近,可以分别具有各个题录或其所属文件所属的特征词提示(301 或 401)(流程 507)。

[0156] 所述特征词也可以是与该题录或其所属文件相关的关键词提示 301(图 3)。

[0157] 各个题录或其所属文件所属的特征词提示可以是单级或多级类别提示 401(参见图 4)。

[0158] 所谓所属的多级类别词提示,就是显示适合于该题录或其所属文件的多个分别属于不同级别分类的类别词或类别名称或类别条目或者代表类别的符号或图形键。

[0159] 显然,所谓所属的多级类别词提示的各个类别词,无论类别大小,都是该题录或其所属文件所属类别词。这样比起显示一般的树状目录或泛泛的导航目录来说,不仅大大减少了占用空间,还对相关题录具有直接针对性或可类比性或提示性。

[0160] 例如,某一文件或题录属于“学知”这一大类类别词中的下一级子类类别词“学术”里面的“物理”这一更下一级子类类别词,那么,在该题录的附近例如将出现“学知;学术;物理”的字样 401,作为该题录的多级类别词提示。

[0161] 所述与该题录或其所属文件相关的关键词提示,可以不是指原来的查询关键词,最好是能反映该题录或其所属文件区别于原序列许多其它题录的特点或内容的意义重大的关键词。

[0162] 实现在题录附近增加或显示该题录或其所属文件所属的多级特征词或关键词或类别词提示,可以有多种方法。一种是利用该题录附带的其所属文件的地址或网址访问该文件,进而获得该文件的特征词或关键词或类别词信息(利用 506 的结果),再添加到原题录附近。另一种方法是在生成带有自身特征词信息文件的关键词或查询词倒排题录索引时,直接令各个题录附带原文件的多级特征词或关键词或类别词信息,与各个题录一起显示。或者利用其它方法。

[0163] 我们可以令所述的题录附近的特征词提示,分别可以与另外多个文件题录的序列链接 509;所述链接的另外多个文件题录的序列中的部分或全部题录或者其所属的文件,其特征词(或关键词或类别词与该序列链接的原来的提示的特征词(或关键词或类别词)相同,并且可以符合或不符合原来用户提出的查询要求。

[0164] 例如需要时,搜索用户在点击供选择的多个提示中的某个特征词提示时,就会得到新的属于该特征词的并且符合原来用户提出的查询要求的文件题录序列(流程 509),这样可以大大缩小或自由调控搜索范围,得到查询结果及所需文件。

[0165] 显然,在所述得到的新的符合该特征词的文件题录序列中的题录附近,也可以同时具有该题录或其所属文件所属的多个不同特征词或类别词提示或关键词提示;也可以令其中题录所属文件所属的多个不同特征词提示或类别词提示或者关键词提示,分别与另外的与这些提示相关的多个文件题录的序列链接。并可以依此类推。

[0166] 在已有检索技术中有时也会提供在特定的范围内的多级分类目录(如专利文献

国际分类目录),但是非专业的普通用户往往不能准确把握每个类别词的含意或确切覆盖范围,常常错误地选择类别,严重影响了检索速度。

[0167] 有些搜索引擎系统在搜索结果的题录末尾提供“类似网页”或“相同网站”之类的提示或链接,但得到的结果过于笼统或杂乱,用处十分有限。

[0168] 而本发明的在查询时提供的题录附近同时显示的多级特征词提示的方法,会为查询者带来极大便利。用户看到有兴趣的题录时,如果希望得到与原题录大类(较高级类别)相同的题录序列,可以点击提示中较高级的特征词或类别词(例如前述的“学知”);如果希望得到与原题录细类(较低级类别)相同的题录序列,可以直接点击提示中较低级的特征词(例如前述的“物理”)。这样,可以同时保持查询者点击选择的准确性和灵活性,大大提高了查询的效率,并改善了用户的查询体验。

[0169] 本发明所述类别词提示或关键词提示与新的题录序列的链接,可以是直接链接或者间接链接 509。

[0170] 所述提示可以首先链接到在原查询的基础上增加了提示中特征词或相应类别词或关键词的查询搜索,从而得到所需的题录序列。

[0171] 所述提示也可以首先链接到在原查询的搜索题录序列结果基础上进一步所作的以提示中的特征词或相应类别词或关键词为查询逻辑要求的查询搜索,从而得到所需的题录序列。

[0172] 需要时,也可以将原查询结果序列中未出现在所述新的题录序列的题录,安排在所述新的题录序列的后面或者适当后移。

[0173] 需要时可以安排:在计算机检索系统根据网络查询用户提出的查询要求向用户终端提供的题录序列附近,具有多个提示组成的导航目录(流程 508),各个特征词提示可以分别与各个不同的包含多个文件题录的序列链接。也就是说,如果用户在搜索时点击该目录中的某一特征词(也可以安排需要再点击一下“搜索”或“确认”或其它名称的操作键),就会得到新的符合该特征词的文件题录序列 509,该序列中的题录所属的文件的特征词,与该序列链接的原来的提示中(被点击的)的特征词相同,并且可以仍然符合或不符合原来用户提出的查询要求。

[0174] 所述的导航目录也可以是一级目录或多级目录。可以允许该目录的上一级类别选择确定前或确定后自动显示下一级待选类别。

[0175] 所述的导航目录的特征词提示可以是类别词提示也可以是关键词提示。

[0176] 所述导航目录的特征词提示与新的题录序列的链接,可以是直接链接或者间接链接。所述提示可以首先链接到在原查询词的基础上增加了提示中关键词提示要求的查询搜索,从而得到所需的题录序列。所述提示也可以首先链接到在原查询要求的搜索题录序列结果基础上进一步所作的以提示中特征词为查询逻辑要求的查询搜索,从而得到所需的新的题录序列。需要时,也可以将原查询结果序列中,未出现在所述新的题录序列的题录,安排在所述新的题录序列的后面或适当后移。需要时,可在流程 509 的题录序列上重复流程 507 或 508,令其具有相应的特征词提示或导航目录,以链接或供点击展现更新的题录序列结果。

[0177] 搜索完毕后,搜索者可以返回(流程 510),重新开始操作。

[0178] 以上内容为本发明方法的示例性说明,不得以此来限制本发明的权利范围。

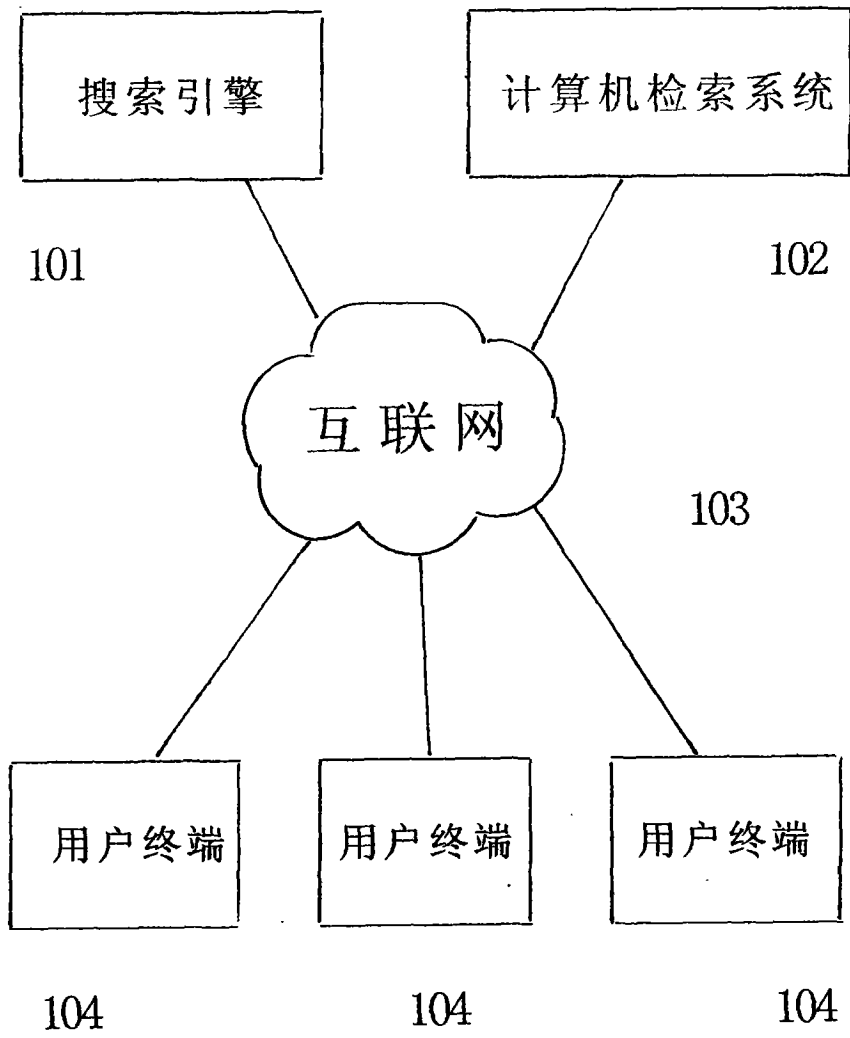


图 1

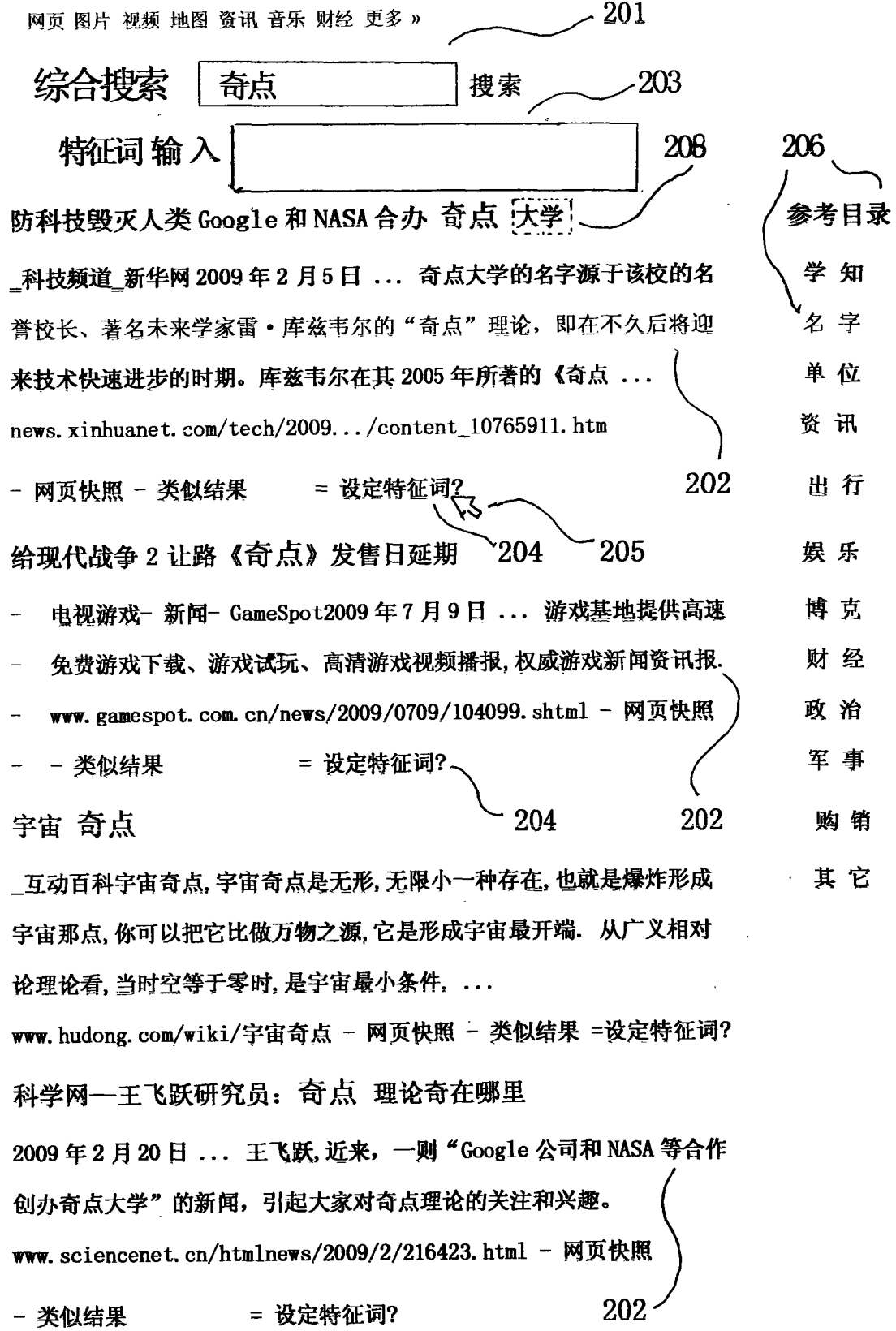


图 2



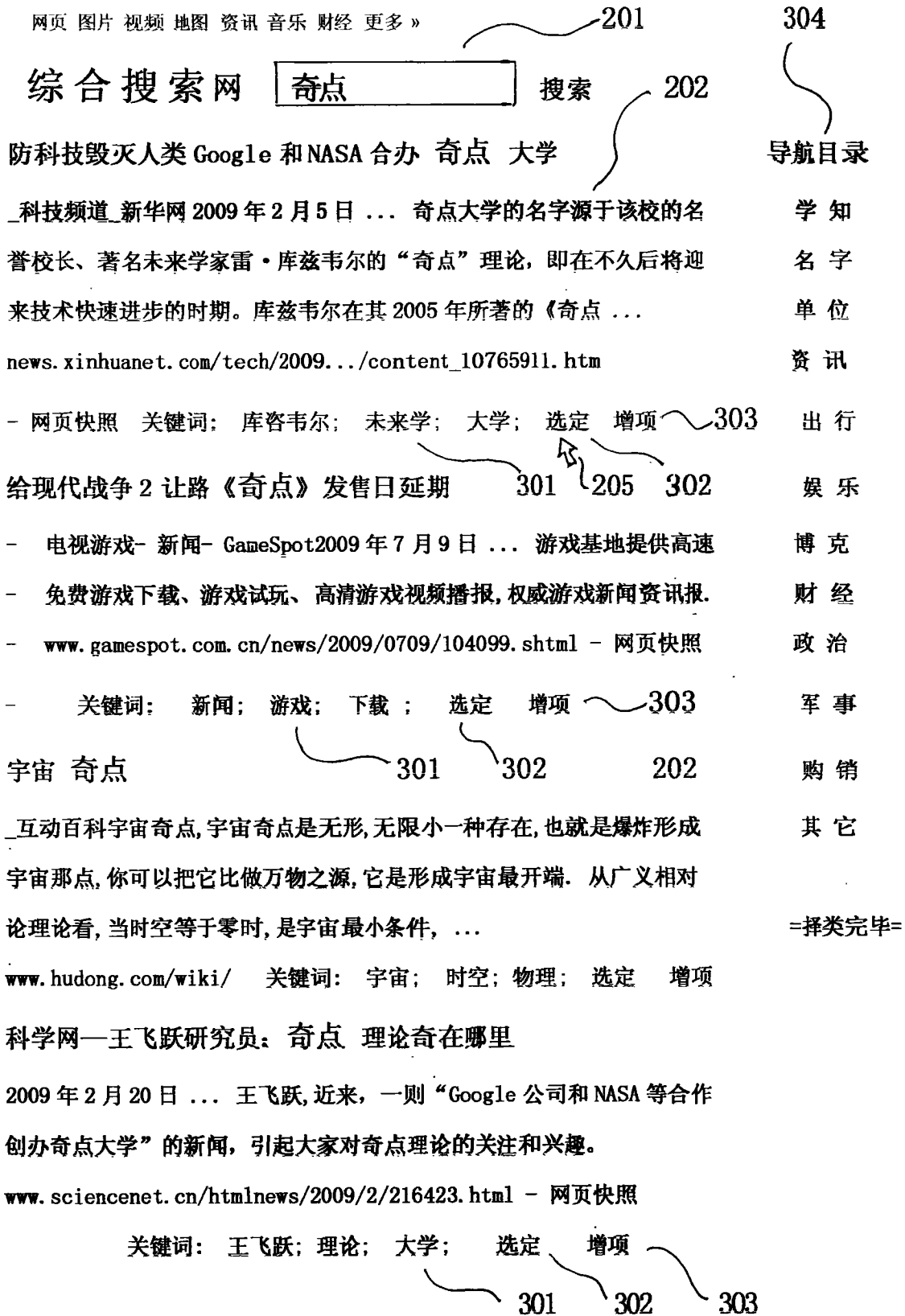


图 3

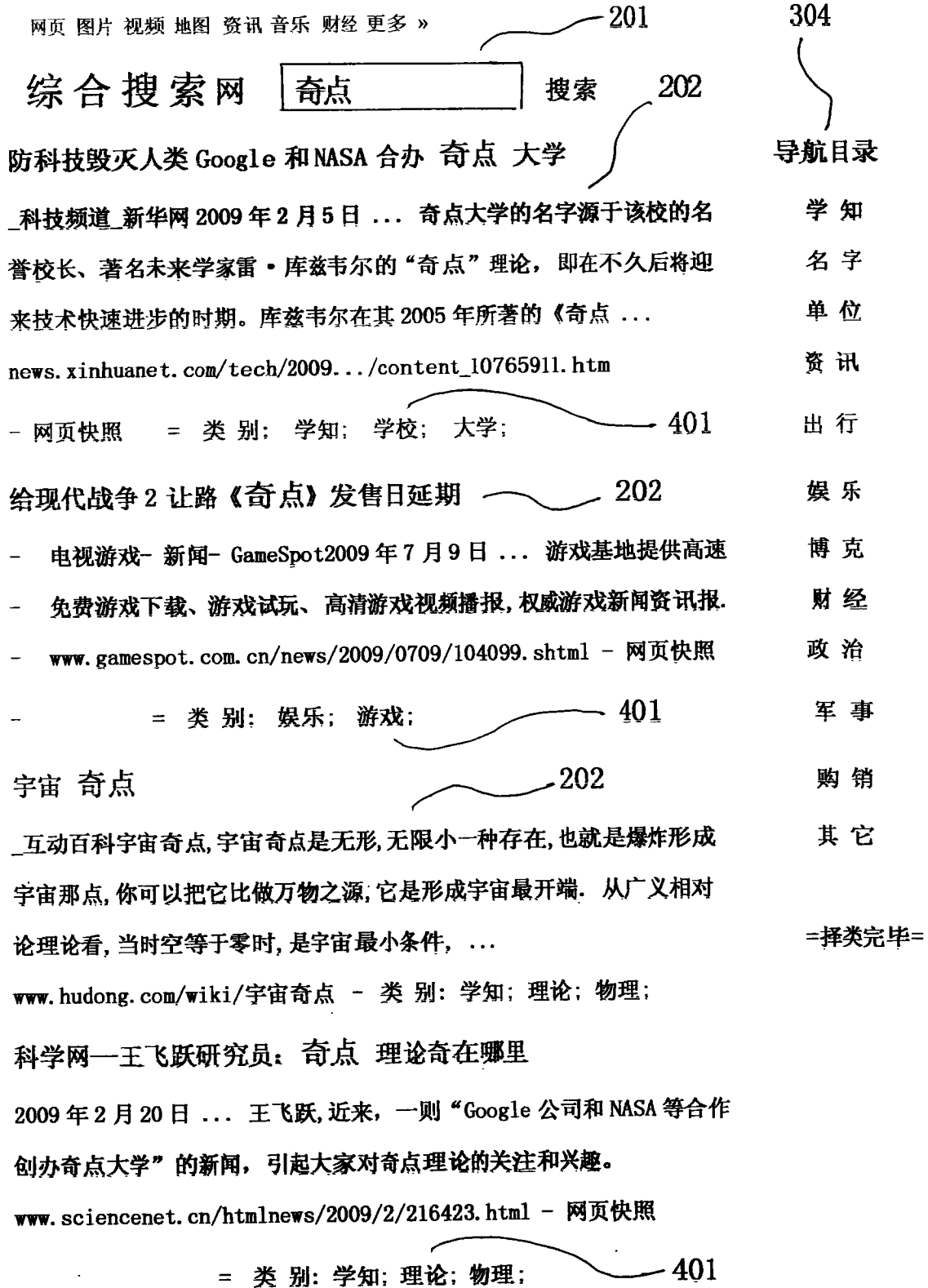


图 4

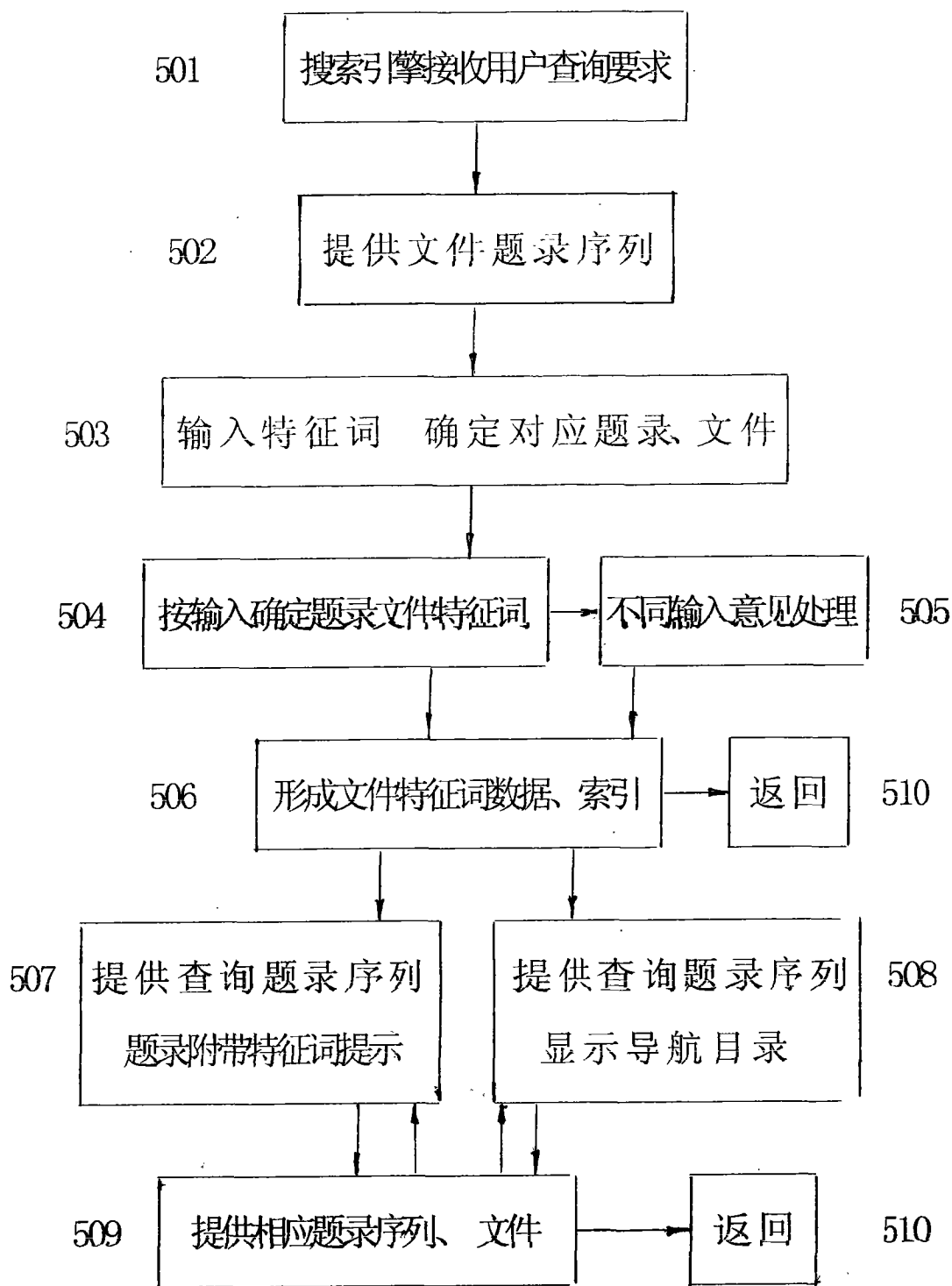


图 5