



(12) 发明专利申请

(10) 申请公布号 CN 102713902 A

(43) 申请公布日 2012. 10. 03

(21) 申请号 201080054826. 6

(51) Int. Cl.

(22) 申请日 2010. 09. 08

G06F 17/30(2006. 01)

(30) 优先权数据

0958596 2009. 12. 02 FR

(85) PCT申请进入国家阶段日

2012. 06. 01

(86) PCT申请的申请数据

PCT/EP2010/063197 2010. 09. 08

(87) PCT申请的公布数据

W02011/067002 FR 2011. 06. 09

(71) 申请人 萨基姆通讯宽带公司

地址 法国吕埃马迈松

(72) 发明人 J-P·莫拉尔

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 王茂华

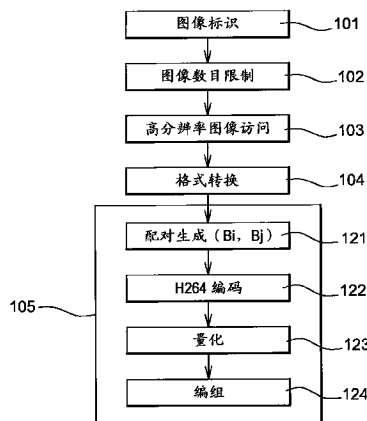
权利要求书 1 页 说明书 4 页 附图 2 页

(54) 发明名称

用于生成使用搜索引擎执行的搜索的结果的方法

(57) 摘要

本发明涉及一种用于生成使用搜索引擎执行的搜索的结果的方法,其中在所述搜索操作之后建立结果列表,其特征在于,所述方法包括如下不同步骤:从所述结果列表中标识(101)与所示结果相关联的缩略图图像(301-308);执行对所述图像的分析操作(105),包括成对比较所标识的缩略图图像;将通过比较所述缩略图图像的操作预先确定的共享相似度的缩略图图像编组(124)成图像族(311;312);以及基于编组成图像族来生成所述搜索的结果。



1. 一种用于生成使用搜索引擎执行的搜索的结果的方法,其中在所述搜索操作之后建立结果列表,其特征在于,所述方法包括如下不同步骤:

- 从所述结果列表中标识 (101) 与所示结果相关联的缩略图类型图像 (301-308);
- 执行对所述图像的分析操作 (105),包括成对比较所标识的缩略图类型图像;
- 将通过比较所述缩略图类型图像的操作确定的共享相似度的所述缩略图类型图像编组 (124) 成图像族 (311 ;312);
- 根据编组成图像族来生成所述搜索的结果。

2. 根据先前权利要求所述的方法,其特征在于,所述比较操作通过编码器执行。

3. 根据先前权利要求所述的方法,其特征在于,所述编码器是 H264 类型。

4. 根据权利要求 2 或 3 中的任意一项所述的方法,其特征在于,通过所述编码器执行的所述比较操作包括如下不同步骤:

- 从所标识的缩略图类型图像构成 (121) 所有可能的缩略图类型图像对 (Bi, Bj);
- 针对每对组成的图像,通过所述编码器执行 (122) 对所述图像对进行编码的操作;
- 从所述编码操作确定所述图像对中所述两个图像之间的差别;
- 将所考虑的所述图像对中的所述两个图像之间的差别量化 (123),以获得量化值;
- 如果所考虑的所述图像对的所述量化值小于先前确定的差别阈值,则将所述图像对中的所述两个图像编组 (124) 到相同的图像族中。

5. 根据先前权利要求中任意一项所述的方法,其特征在于,在所述图像分析步骤之前,所述方法包括以下步骤:访问 (103) 与每个标识的缩略图相关联的高分辨率图像,然后在所述高分辨率图像上执行所述图像分析步骤。

6. 根据先前权利要求中任意一项所述的方法,其特征在于,在所述图像分析步骤之前,所述方法包括以下步骤:将所标识的缩略图类型图像转换 (104) 成相同的图像格式。

7. 根据先前权利要求所述的方法,其特征在于,所标识的缩略图类型图像转换成的所述图像格式是 YUV 格式。

8. 根据先前权利要求中任意一项所述的方法,其特征在于,包括从所述结果列表标识与所示结果相关联的所述缩略图类型图像的步骤在所述结果列表中先前确定数目的结果上执行 (102)。

用于生成使用搜索引擎执行的搜索的结果的方法

技术领域

[0001] 本发明的目标是用于生成使用搜索引擎执行的搜索的结果的方法。本质上,本发明的目标是改进尤其是在因特网网络上的通过搜索引擎发起的计算机搜索的结果的可靠性。

[0002] 本发明的领域总体上是通信网络上的信息搜索的领域。本发明可以在使用搜索引擎发送尤其是可视数据的任意通信网络上实现;这里,所考虑的可视数据是缩略图(thumbnail)类型的可视数据。本发明将在所考虑的通信网络是因特网网络的情况下进行描述,所述因特网网络是用户通过因特网浏览器访问的网络;在该网络上,用户可以通过搜索引擎执行搜索。然而,本发明还可以在其他类型的网络(例如公司内部网网络)上实现。

背景技术

[0003] 现今,越来越多的文档被数字化并且通过搜索引擎可访问。被数字化并且在线可访问的文档数量上的增长的直接后果是当例如通过关键字由因特网浏览器执行搜索时获得的结果数量上的增长。如果因此获得的结果呈现了相对深入的特性,则对于用户而言增加了从搜索结果列表中找到期望信息的难度。某些搜索引擎设法对找到的结果的相关性划分优先级,但是手动步骤仍然存在,用户必须执行该手动步骤以从获得的结果列表中选择用户真正感兴趣的结果。

[0004] 为了促进该手动步骤,通常以与 URL 相关联的文本行的形式返回结果列表的某些浏览器利用缩略图类型图像来补充搜索结果。缩略图是指如下尺寸上减少的图像,其允许与该图像相关联的结果被视觉上标识。例如,缩略图可以由唱片套或书籍的封面的图像构成。缩略图的尺寸可以被自愿减少以便能够例如和与搜索结果相关联的文本协调进行放置,并且能够维持在屏幕上可视地同时显示大量搜索结果的能力。在某些情况下,缩略图本身构成所执行搜索的结果行,而不与文本行相关联。因此,在搜索之后,缩略图可以直接出现或者与 URL 地址类型信息相关联地出现。

[0005] 然而,仍然存在困难:尽管使用了缩略图,但是由搜索引擎执行的搜索的结果可能以相对无组织方式呈现,并且可能对于用户而言难于利用。因此,在大多数情况下,这些结果包括结果族(family),每个结果族对应于指定类似元素的结果。但在实践中,这些结果族没有被编组在一起。

[0006] 因此,例如如果用户搜索给定艺术家的特定专辑(他忘记了该专辑的名称),则该用户会在其搜索引擎的适当窗口中录入该艺术家的名字。然后,不同的专辑将会出现在结果列表中;但是,针对所考虑的艺术家的给定专辑的不同结果不一定编组在一起,例如编组在搜索结果的连续行中。实际上,相同专辑可以由不同 URL 引用,这些 URL 在结果列表中不一定是连续的。

[0007] 另外,相同专辑中的微小差别可能存在于唱片套中;例如,专辑的“简装”版与“收藏”版之间的微小差别就属于上述情况;事实上,相同专辑的不同版本对应于艺术家的相同作品,并且因此可能看起来在由搜索引擎产生的结果列表中将上述版本编组在一起以便促

进用户的搜索是令人感兴趣的。在本发明的上下文中,已经利用艺术家专辑描述的示例可容易地转变成其他项目,诸如书籍、DVD 或其他项目。

发明内容

[0008] 根据本发明的方法提出了一种对上述问题的解决方案。本发明提出了一种解决方案,从而使得当通过搜索引擎发起搜索时生成的缩略图在由搜索引擎产生的结果列表中通过条目族编组在一起。条目族是指如下多个条目,其相关联的缩略图呈现相似性,从而使得这些条目被认为指代相同的对象。为此,提出了执行以下操作以便使得搜索结果对于用户而言更加可读:将由搜索产生的不同缩略图进行比较,以及将呈现强相似性的类似缩略图编组在一起。

[0009] 本发明本质上涉及一种用于生成使用搜索引擎执行的搜索的结果的方法,其中在所述搜索操作之后建立结果列表,其特征在于,所述方法包括以下不同步骤:

- [0010] - 从结果列表中标识与所示结果相关联的缩略图类型图像;
- [0011] - 执行对图像的分析操作,包括成对比较所标识的缩略图类型图像;
- [0012] - 将通过比较缩略图类型图像的操作确定的共享相似度的缩略图类型图像编组成图像族;
- [0013] - 根据编组成图像族来生成搜索的结果。

[0014] 实践中,如果两个图像实际上表示相同对象,则所述两个图像被认为呈现彼此之间的相似度,所述图像可以按照细节的顺序在它们之间呈现差异。通常,如果在以相同格式并且以相同尺寸返回的情况下构成两个图像的像素的小于 20% 呈现不同的分量值,则可以认为这两个图像是彼此相似的。

[0015] 除了先前段落中已经提到的主要步骤之外,根据本发明的方法可以包括以下中的一个或多个附加特性:

- [0016] - 比较操作通过编码器执行;
- [0017] - 编码器为 H264 类型;
- [0018] - 通过编码器执行的比较操作包括如下不同步骤:
- [0019] - 从标识的缩略图类型图像组成所有可能的缩略图类型图像对;
- [0020] - 针对每对组成的图像,通过编码器执行对所述图像对进行编码的操作;
- [0021] - 从编码操作确定图像对中两个图像之间的差别;
- [0022] - 将所考虑的图像对中的两个图像之间的差别量化,以获得量化值;
- [0023] - 如果所考虑的图像对的量化值小于先前确定的差别阈值,则将图像对中的两个图像编组到相同的图像族中;
- [0024] - 在图像分析步骤之前,所述方法包括以下步骤:访问与每个标识的缩略图相关联的高分辨率图像,然后在该高分辨率图像上执行图像分析步骤;
- [0025] - 在图像分析步骤之前,所述方法包括以下步骤:将标识的缩略图类型图像转换成相同的图像格式;
- [0026] - 所标识的缩略图类型图像转换成的图像格式是 YUV 格式;
- [0027] - 包括从结果列表标识与所示结果相关联的缩略图类型图像的步骤在所述结果列表中先前确定数目的结果上执行。

[0028] 根据本发明的方法的不同附加特性在它们不相互排斥的情况下,根据所有组合可能性进行组合,以产生本发明的实施方式的不同示例。

[0029] 在阅读以下描述并且查看附图之后,将更好地理解本发明及其各种应用。

附图说明

[0030] 附图是出于指示目的而呈现并且不以任何方式限制本发明。其中:

[0031] 图 1 是示出了根据本发明的方法的实现的示例的示图;

[0032] 图 2 是根据本发明生成通过搜索引擎执行的搜索的结果的示意表示;以及

[0033] 图 3 是图示了通过图像族对图像进行编组的值表的表示。

具体实施方式

[0034] 除非另外说明,出现在不同附图中的元素将保持相同的附图标记。

[0035] 图 1 表示了图示了根据本发明的方法的实现示例的示图;图 2 中图示了组织从搜索生成结果的方法的所述程序的不同步骤。

[0036] 因此,将同时对图 1 和图 2 进行描述。

[0037] 本发明在由用户在因特网网络上通过搜索引擎执行搜索的上下文中进行描述。

[0038] 图 2 表示了分别表示根据本发明的方法的实现之前以及根据本发明方法的实现之后的搜索结果的第一屏幕 200 和第二屏幕 300。在所表示的示例中,用户通过在搜索引擎的区域 202 中键入一组关键字 201(这里是“艺术家 X”)来发起搜索。继而,用户获得对应于唱片套、根据特定于每个搜索引擎的方法而分层级地组织的结果列表 203。在所示示例中,为了简单化,结果列表 203 限于缩略图,省略了实践中可能会呈现的与 URL 地址相关联的文本。因此,这里考虑结果列表排他地由所示结果组成,即,由与缩略图相关联的结果组成。仍然是为了简单化,结果列表 203 限于八个结果的列表,按其呈现在屏幕上的顺序引用为 301 至 308。实践中,结果的数目经常大得多,并且对缩略图的使用必须成为用于在用户感兴趣的的结果的选择方面进行辨别的工具。

[0039] 在表示的示例中,针对艺术家 X 列出了三个专辑,第一专辑具有描述“DISC 1”(缩略图 301、304、306 和 308),第二专辑具有描述“DISC 2”(缩略图 302、305 和 307),而第三专辑具有描述“DISC 3”(缩略图 303)。如可以看出的,不同的专辑在结果列表中是以无组织的方式出现:所考虑的专辑的不同情况在结果列表中没有被编组在一起,这对用户而言是不实际的,尤其是在结果列表 203 包括大量结果时。

[0040] 如图 1 中所示,为了根据按本发明的方法的实施方式的示例生成搜索结果,在第一步骤 101 中,本发明执行操作以在结果列表 203 中标识缩略图类型图像。这种操作可以通过传统的图像处理操作或者通过标识与缩略图相关联的 URL 来执行。在本发明的特定实施方式中,在之后的步骤 102 中,将缩略图的标识限制于预先确定的数目(例如五十),以便限制随后介入的计算量,并且通过考虑这种限制将促进用户选择其实际上感兴趣的的结果的任务。在所考虑的示例中,缩略图的数目被限制在先前提到的八个缩略图。

[0041] 有利地,在之后的步骤 103 中,执行操作以访问与所考虑的缩略图相关联的高分辨率图像。为此,可以使用例如与所考虑的缩略图相关联的 URL。

[0042] 在之后的步骤 104 中,执行操作以将缩略图类型图像(如果有必要,以其高分辨率

版本) 从其原始格式 (jpeg、gif 等) 转换成由编码器类型设备直接可用的格式。这种格式例如是 YUV 4:2:0 格式或 RGB 格式。

[0043] 在之后的步骤 105 中, 执行基于经转换的图像来分析图像的步骤。在所考虑的示例中, 该步骤包括以下不同操作:

[0044] - 操作 121, 通过根据所有可能性组合所有标识的图像来构成图像对; 因此, 在所考虑的示例中, 从八个标识的缩略图中获得 28 对两个图形 (B_i, B_j), 其中 i 和 j 属于集合 {1-8}。

[0045] - 操作 122, 在操作 122 期间, 对 (B_i, B_j) 由 H264 类型解码器编码; 在实施方式的其他示例中, 编码器可以是 VC1 或 Dirac 类型。因此在编码器的输出, 针对每个编码对 (B_i, B_j), 以如下数据序列的形式获得信息集: $\{R_i, P(i, 1), P(i, 2) \dots P(i, m)\}$, 其中 R_i 是图像 B_i 的压缩格式, 并且数据 $P(i, m)$ 是指示在图像 B_j 与图像 B_i 之间观察到的改变的数据, 值 m 对应于在该处已经观察到图像 B_i 与图像 B_j 之间的差别的图像的区域 (通常为矩形) 的数目。

[0046] - 操作 123, 在该操作 123 期间, 针对已经经历过编码操作的每对图像 (B_i, B_j) 对图像 B_i 与图像 B_j 之间观察到的差异进行量化; 为此, 考虑信息集 $\{R_i, P(i, 1), P(i, 2) \dots P(i, m)\}$, 并且针对每个编码对 (B_i, B_j), 将所有数据 $P(i, m)$ 相加以获得被称为差异值的值。该值越低, 则图像 B_i 与图像 B_j 之间的相似性就越高。

[0047] - 操作 124, 用于将图像编组在一起, 在以上描述的操作之后呈现低于或者低于或等于先前确定的阈值的图像 B_i 与图像 B_j 在操作 124 期间被编组在一起到相同的图像族中。编组到相同图像族中的图像继而被认为共享相似度。

[0048] 图 3 示意性地图示了图像族的构成。在该图中表示了表 320, 该表的每行与该表的每列对应于八个缩略图 301-308 之一。与缩略图相关联的行和与另一缩略图相关联的列的交叉点给出根据以上描述的方法计算的被称为比较值的值。因此, 通过对其比较值小于先前确定的阈值的图像进行编组来执行图像编组操作。

[0049] 在所考虑的示例中, 值“2”被当作阈值。因此, 如图 2 中所示, 在示意性地表示遵循根据本发明的方法的实现而生成的搜索的结果的屏幕 300 处:

[0050] - 缩略图 301、304、306 和 308 被编组在一起到相同行上以形成第一图像族 311; 实际上, 这些缩略图之间仅有的差异包括在单词“DISC”的不同字母之间存在符号, 这些缩略图在所有其他方面都是相同的。

[0051] - 缩略图 302、305 和 307 被编组在一起到相同行上以形成第二图像族 312; 实际上, 这些缩略图之间仅有的差异包括出现在单词“DISC”的不同字母之间存在符号, 这些缩略图在所有其他方面都是相同的。

[0052] 因此, 用户可以更容易地定位所搜索的对象。

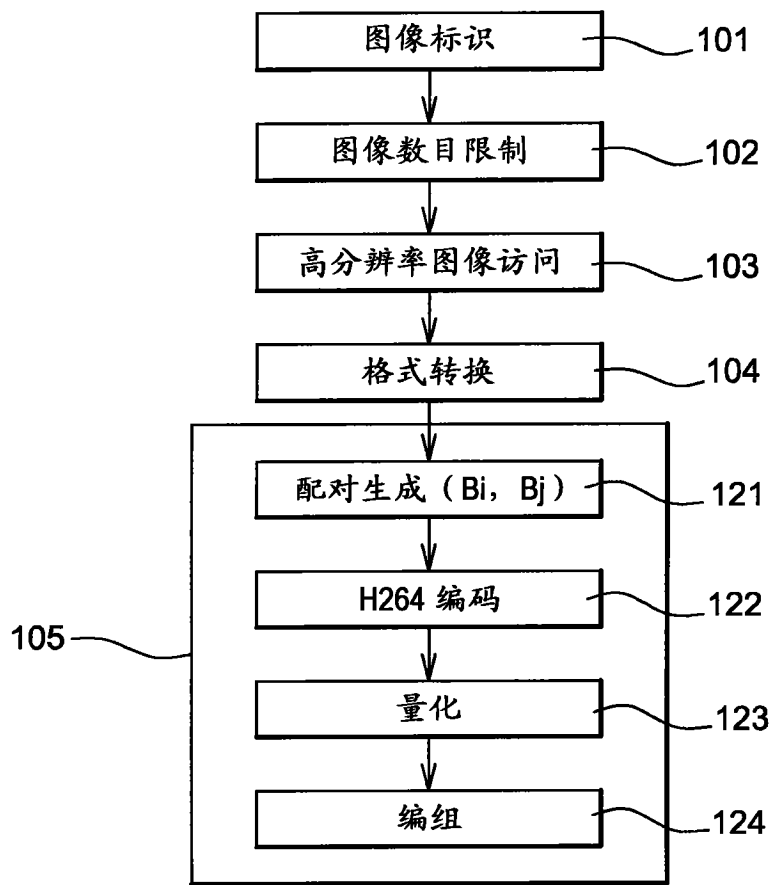


图 1

	-301-	-302-	-303-	-304-	-305-	-306-	-307-	-308-
-301-								
-302-	3							
-303-	5	4						
-304-	①	3	6					
-305-	4	②	7	3				
-306-	②	7	4	6	3			
-307-	6	①	6	5	②	3		
-308-	①	3	4	4	4	4	5	

图 3

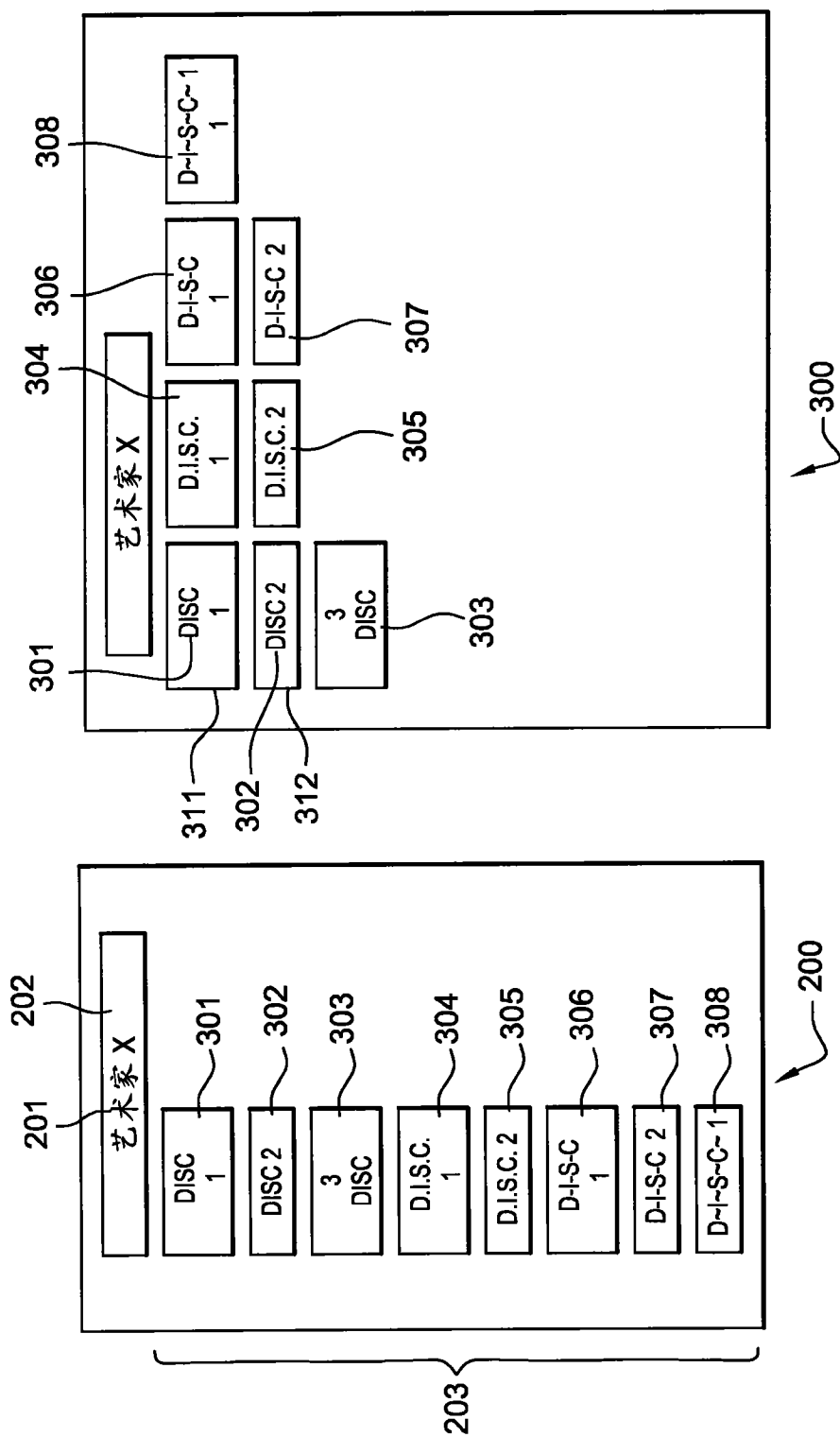


图 2