

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(43) 国际公布日
2007 年 12 月 27 日 (27.12.2007)

(10) 国际公布号
WO 2007/147359 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2007/070114

(22) 国际申请日: 2007 年 6 月 14 日 (14.06.2007)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
200610061188.6
2006 年 6 月 15 日 (15.06.2006) CN

(71) 申请人 (对除美国外的所有指定国): 腾讯科技 (深圳) 有限公司(TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。

(72) 发明人: 及

(75) 发明人/申请人 (仅对美国): 余祥鑫(YU, Xiangxin) [CN/CN]; 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。熊应(XIONG, Ying) [CN/CN]; 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。刘致远(LIU, Zhiyuan) [CN/CN]; 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518044 (CN)。

(74) 代理人: 北京德琦知识产权代理有限公司(DEQI INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区知春路 1 号学院国际大厦 7 层, Beijing 100083 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

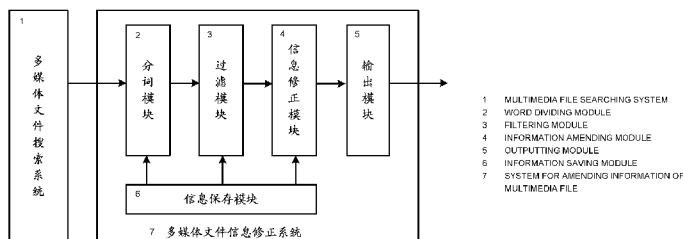
本国际公布:

— 包括国际检索报告。

所引用双字母代码及其它缩写符号, 请参考刊登在每期 PCT 公报期刊起始的“代码及缩写符号简要说明”。

(54) Title: SYSTEM AND METHOD FOR AMENDING INFORMATION OF MULTIMEDIA FILE

(54) 发明名称: 多媒体文件信息修正系统及方法



(57) Abstract: A system and method for amending the information of the multimedia file, the system comprising an information saving module and an information amending module, wherein the information saving module is used for saving the determined information of the multimedia file; the information amending module is used for looking up the determined information related to the multimedia file to be amended in the information saving module, and replacing the information of the multimedia file with the determined information which has been looked up. A method for amending the information of the multimedia file, the method comprising: A) looking up the determined information related to the information of the multimedia file to be amended in the determined information of the multimedia file base on the information of the multimedia file to be amended; B) replacing the information of the multimedia file to be amended with the determined information which has been looked up.

[见续页]

WO 2007/147359 A1



(57) 摘要:

一种多媒体文件信息修正系统，该系统包括信息保存模块和信息修正模块，其中：所述信息保存模块用于保存多媒体文件的已确定信息；所述信息修正模块用于在所述信息保存模块中查找与待修正多媒体文件相关的已确定信息，并用查找到的已确定信息替换待修正多媒体文件信息。一种多媒体文件信息修正方法，该方法包括：
A) 根据待修正多媒体文件的信息在多媒体文件的已确定信息中查找与待修正多媒体文件信息相关的已确定信息；
B) 用查找到的已确定信息替换待修正多媒体文件信息。

多媒体文件信息修正系统及方法

技术领域

本发明涉及对多媒体文件信息进行修正的技术领域，尤其是多媒体文件信息修正系统和多媒体文件信息修正方法。

发明背景

目前的音乐文件信息采集主要有以下两种方式：使用爬虫对整个互联网进行爬行和对专门的音乐网站进行爬行。

通过使用爬虫对整个互联网进行爬行，并从中提取出与音乐文件有关的链接，保存对应的锚文本、标题等文本信息，并对这部分文本建立索引，用户通过输入关键字在其中进行检索，然后通过搜索结果输出系统展现给用户。目前，网络上大多数的搜索引擎都是采用这种方式进行处理。

通过使用爬虫对整个互联网进行爬行来采集音乐文件信息，虽然得到的可搜索的链接数量较多，并且需要较少的人工干预，但是这种方式得到信息的准确率较低，常出现以下的错误：信息不完整，如缺少歌曲名、歌手名、专辑名等中的一个或多个；信息的内容不准确，比如出现错别字、错误写法等；歌曲描述信息为无意义的广告或乱码；歌曲描述信息与真实内容不符；故意堆砌大量广告关键字或热门歌曲名进行欺骗。

图 1 所示为利用上述现有技术进行搜索后得到的结果示意图，如图 1 所示，存在如下问题：

第 2 条搜索信息中，被列在歌曲名中的“舞娘”应该是专辑名，而不是歌曲名的一部分；

第 10 条搜索信息中，歌曲名中的“爱娱乐”是提供该歌曲的网站名，不是歌曲名，属于不需要的广告文字；

第 11 条搜索信息中，歌曲名写法不完全正确，并且此条目缺少歌手名与专辑名；

第 16 条，搜索信息中，歌曲名中的最后一个字是乱码，且缺少歌手名与专辑名。

另一种方法是利用对专门的音乐网站进行爬行获取音乐文件信息，然后利用人工维护的模板来提取页面上的音乐文件信息，虽然该方法可以获得比较准确的信息，但是由于该方法使用人工维护的模板来提取页面上的音乐文件信息，而针对不同的网站通常需要维护不同的模板，工作量非常大，这就导致了效率较低，只能采集较少网站上的音乐文件信息。

发明内容

本发明实施例提出了一种多媒体文件信息修正系统，其目的在于有效地提供较为准确的多媒体文件信息。本发明实施例还提出了一种多媒体文件信息修正方法，用来有效地提供较为准确的多媒体文件信息。

为实现上述目的，本发明实施例提供了一种多媒体文件信息修正系统，该系统包括信息保存模块和信息修正模块，其中：

所述信息保存模块用于保存多媒体文件的已确定信息；

所述信息修正模块用于在所述信息保存模块中查找与待修正多媒体文件相关的已确定信息，并用查找到的已确定信息替换待修正多媒体文件信息。

本发明实施例还提供了一种多媒体文件信息修正方法，该方法包括：

A. 根据待修正多媒体文件的信息在多媒体文件的已确定信息中查

找与待修正多媒体文件信息相关的已确定信息；

B. 用查找到的已确定信息替换待修正多媒体文件信息。

从上述技术方案可以看出，由于本发明利用多媒体文件的已确定信息，根据待修正多媒体文件在其中查找到相关的已确定信息，然后用所查找到的已确定信息替换待修正多媒体文件原来的信息，从而提供了较为准确的多媒体文件信息，并且上述技术方案的实施不依赖于人工维护的模板，因此本发明实施例的技术方案还具有较高的效率。

另外，本发明实施例的技术方案可以应用于对搜索引擎的音乐文件信息的修正、对本地计算机等设备中的多媒体文件信息进行修正等多个方面。

附图简要说明

图 1 为利用现有技术进行搜索得到的音乐文件信息的示意图；

图 2 为本发明实施例的多媒体文件信息修正系统的结构示意图；

图 3 为本发明实施例的多媒体文件信息修正方法的流程示意图。

实施本发明的方式

为使本发明的目的、技术方案和优点更加清楚，以下举实施例对本发明进一步详细说明。

图 2 为本发明实施例的多媒体文件信息修正系统的结构示意图，如图 2 所示，其至少包括信息保存模块和信息修正模块，还可以进一步包括分词模块、过滤模块、输出模块等。

本发明实施例的多媒体文件信息修正系统可以用在多个方面，例如对现有搜索引擎搜索到的音乐文件等多媒体文件的信息进行修正、对本地计算机等设备中的多媒体文件的信息进行修正以及对其它场合的多

媒体文件的信息进行修正。在下面的描述中，以修正搜索引擎搜索得到的音乐文件的信息为例进行说明，本领域技术人员只要将音乐文件替换为其它多媒体文件、将音乐文件信息替换为其它多媒体文件对应的信息即可。

由于要对搜索引擎搜索得到的音乐文件的信息进行修正，如图 2 所示，多媒体文件信息修正系统进一步与多媒体文件搜索系统相连接，该多媒体文件搜索系统利用现有的搜索技术从互联网设备和/或本地设备获取多媒体文件信息，例如利用爬虫对互联网爬行，从中提取与音乐文件相关的链接，保存对应的锚文本、标题、标签（Tag）等文本信息，并将这些信息作为待修正的音乐文件信息，提供给多媒体文件信息修正系统。在这里，所述的音乐文件信息包括歌曲名、歌手名、专辑名等。另外，由于多媒体文件搜索系统为现有技术，在此不再赘述。

下面结合附图对本发明的多媒体文件信息修正系统进行详细描述。

如图 2 所示，信息保存模块中保存有音乐文件的已确定信息，即：已经得到确定的歌曲名、歌曲对应的歌手名及专辑名等音乐文件信息。表 1 给出了信息保存模块中音乐文件信息的一种实现方式：

表 1

序号	歌曲名	歌手名	专辑名
1	A1	A2	A3
2	B1	B21	B31
		B22	B32
3	C1	C2	C31
			C32
...	...	...	...

在表 1 中描述了多种情况:

- 1、歌曲 A1, 唯一由歌手 A2 演唱, 其对应的专辑为 A3;
- 2、歌曲 B1, 歌手 B21 和 B22 都演唱过, 其对应的专辑分别为 B31 和 B32;
- 3、歌曲 C1, 唯一由歌手 C2 演唱, 但该歌在两张专辑 C31 和 C32 中出现。

当然上述表 1 只是一种示范, 一首歌也可以由两位以上的歌手演唱, 也可以出现在 2 张以上的专辑中。

在本发明实施例中, 也可以利用另外的方式实现音乐信息的保存, 例如: 存储所有专辑的信息, 包括专辑中的歌曲名和对应的歌手名; 或者存储所有歌手的信息, 包括歌手名、演唱过的歌曲名和歌曲对应的专辑名。也可以不保存上述歌曲名、歌手名以及专辑名之间的对应关系, 只是保存若干个歌曲名、歌手名、专辑名。

该信息保存模块可以是一个单独的模块, 也可以由多个保存模块组成。即, 歌曲信息可以保存在一个保存模块中, 也可以分多个模块进行保存。

图 2 中的分词模块用于对多媒体文件搜索系统搜索得到的音乐文件信息进行分词处理, 即对以文本形式保存的锚文本、网页标题和 Tag 内容进行分词处理。

上述分词模块的分词处理分为两种情况:

第一种情况是锚文本、网页标题和 Tag 内容中存在空格和/或标点符号。此时, 分词模块以空格和/或标点符号作为分隔符对文本进行分词, 如“周杰伦-夜曲”会被切分为“周杰伦”和“夜曲”两个词。

第二种情况是锚文本、网页标题和 Tag 内容中没有分隔符隔开, 如“周杰伦夜曲”。因为其中不带有分隔符, 因此使用上述的利用分隔符

分词的方法无法实现分词，在此，本发明实施例使用信息保存模块中的歌曲名、歌手名及专辑名作为词典，利用诸如逆向最大匹配法、正向最大匹配法、基于统计的分词方法等分词方法进行分词，当然也可以多种分词方法结合起来实现分词，如将正向最大匹配法和逆向最大匹配法结合起来构成双向匹配法实现分词。

一般来说，歌手名、歌曲名、专辑名等均为专有名词，较少产生歧义，使用上述方法进行分词即可达到较好的效果，例如完全可以将“周杰伦夜曲”正确地分为“周杰伦”和“夜曲”两个词。

对文本形式保存的锚文本、网页标题和 Tag 内容进行分词处理后，待修正的音乐文件信息中的锚文本、网页标题和 Tag 内容在形式上都转化为一个或一组词语，如上面的例子中就被转化为“周杰伦”和“夜曲”两个词语。

如图 2 所示的过滤模块用于对待修正的音乐文件中进行过滤处理，去除不需要的信息，例如广告、欺骗信息等对用户来说没有任何意义的信息。过滤模块可以直接对未经分词处理的待修正多媒体文件信息进行过滤，也可以是对经过分词处理后的待修正多媒体文件信息进行过滤，下面以对分词处理后的待修正多媒体文件信息进行过滤为例进行说明。

过滤模块中保存有过滤条件，过滤模块根据过滤条件对广告或欺骗信息等不需要的信息进行过滤，如果发现符合过滤条件，则将待修正多媒体文件信息中的相关条目去除。

具体来说，过滤条件主要包括两部分：动作和对比值，其中动作可以是包括、相同、长度大于、长度小于等，对比值可以是各种信息。本实施例中的过滤条件例如为“包括 WWW”、“包括 XX”（XX 为某些色情词汇）。如果待修正多媒体文件信息中的一个或多个信息满足过滤条件，则过滤模块将待修正多媒体文件信息中的相关信息去除。如，待修

正多媒体文件信息中包括“WWW”或“XX”等，则过滤模块将相应的信息去除。

另外，上述的过滤条件可随时进行修改、删除、增加等操作。

同时，在待修正多媒体文件信息中，还有可能出现以下的情况：由于音乐网站会在标题中罗列大量的热门歌曲名来提高自己的排名，但并不提供真实的内容。对于这种情况，本发明中的过滤模块还用于根据另外的过滤条件过滤这种信息，处理流程如下：

首先，统计待修正多媒体文件信息中来自同一网站的搜索条目的总数；

其次，统计分词后各信息在该网站的信息中出现的次数；

最后，判断二者之间的比例，如果超过一定的阈值（如 50%）时，即可判定为欺骗信息，将待修正多媒体文件信息中的相应条目去除。

如从某网站下共采集了 1000 条搜索条目，其中“七里香”在搜索条目的标题中出现了超过 500 次，则可判断这个词语被该网站用来当作欺骗信息，因此删除相关条目的信息。

通过上述的处理后，基本过滤了欺骗信息、广告信息以及其他一些不是用户所需要的信息，得到了一组过滤后的信息。

结合图 2 所示，上述的对过滤模块的描述中，过滤条件保存在过滤模块中，当然，也可以将过滤条件保存在信息保存模块中，并将过滤模块与信息保存模块连接，在进行过滤处理的时候，由过滤模块调用过滤条件进行过滤即可实现。

在过滤模块对分词后的锚文本、网页标题和 Tag 内容的词组过滤掉部分搜索条目的信息后，得到了一组过滤后的信息。

如图 2 所示的信息修正模块，用于在信息保存模块中查找与待修正多媒体文件相关的已确定信息，并用查找到的已确定信息替换待修正多

媒体文件的信息。这里的待修正多媒体文件信息可以是未经分词、过滤的原始多媒体文件信息，可以是只经过分词或只经过过滤的多媒体信息，可以是经过分词和过滤的多媒体文件信息。下面以经过分词和过滤的多媒体文件信息为例进行说明。

现在以修正音乐文件的歌曲名、歌手名、专辑名为例对信息修正模块的处理进行详细描述，其包括歌曲名确定步骤、歌手名确定步骤和专辑名确定步骤，其中：

歌曲名确定步骤，将分词、过滤后的信息按锚文本、标题、Tag 的顺序排序，然后依次与信息保存模块中的歌曲名进行匹配查找，看是否有完全匹配的歌曲名，如果有，将第一个匹配的歌曲名作为该音乐文件的歌曲名，否则将相似度评判标准之上的相似度最高的歌曲名作为该音乐文件的歌曲名。也就是说，本发明实施例中的查找到的歌曲名、歌手名、专辑名可以是精确匹配的，也可以是相似度最高的。

在此，相似度的定义为：两个信息 S1 与 S2 相同的字符数与 S1 和 S2 的平均长度的比值，其中，S1 与 S2 为比对的两个信息。例如，“ABC”与“BCA”的相似度为 100%，而“ABC”与“BCD”的相似度为 67%，而“ABC”与“BA”的相似度为 80%。

在此，相似度的评判标准应该设置一个合适的值，如 70%，如果相似度低于 70% 则不能将其作为歌曲名。

歌手名确定步骤，在确定歌曲名之后，如果音乐信息保存模块中该歌曲名对应的歌手名是唯一的，则可同时确定歌手名。如果对应多个歌手名，则说明这是一首曾被多人翻唱过的同名歌曲，因此，将分词、过滤后的信息按锚文本、标题、Tag 的顺序依次与歌曲名对应的歌手名进行匹配查找，看是否有完全匹配的歌手名，如果有，将第一个匹配的歌手名作为该音乐文件的歌手名，否则将评判标准之上的相似度最高的歌

手名作为该音乐文件的歌手名，如果找不到，则将歌手名这一项留空。

专辑名确定步骤，在确定了歌曲名和歌手名之后，如果音乐信息保存模块中该歌曲名对应的专辑名是唯一的，则可确定专辑名。如果对应多个专辑名，则将分词、过滤后的词语按锚文本、标题、Tag 的顺序依次与歌曲名对应的专辑名进行匹配查找，看是否有完全匹配的专辑名，如果有，将第一个匹配的专辑名作为该音乐文件的专辑名，否则将评判标准之上的相似度最高的专辑名作为该音乐文件的专辑名，如果找不到，则将专辑名这一项留空。

在每个音乐文件的歌曲名、歌手名和专辑名确定后，将待修正多媒体文件信息中的对应信息进行替换。

如图 2 所示的输出模块，用于输出信息修正模块修正之后的多媒体文件信息。

另一方面，信息保存模块中保存的也可以是多媒体文件的哈希码与该多媒体文件的已确定信息之间的对应关系。那么信息修正模块可以根据待修正多媒体文件的哈希码查找对应的已确定信息作为与待修正多媒体文件相关的已确定信息，并利用已确定信息替换待修正信息，从而实现多媒体文件信息的修正。

信息保存模块中的数据保存格式可以如表 2 所示：

表 2

文件哈希码	歌曲名	歌手名	专辑名
0x1234ABCD	A1	B1	C1
0x5678CDEF	A2	B2	C2
...	...	...	...

其中，文件哈希码可以是一个 32 位或 64 位的整数，常用的哈希算法有循环冗余校验 32 (CRC32)、报文摘要算法 5 (MD5) 等。如果两个文件 F1 和 F2 用同一算法所计算出来的哈希码相等，则可认为 F1 和 F2 的内容完全相等。

一般来说，互联网上的音乐文件都存在一个文件多份拷贝的情况。假设可以确定某音乐文件 F 的准确信息，则可以将这一信息保存在信息保存模块中。这样，在以后遇到与音乐文件 F 的哈希码相同的文件时，可以直接从信息保存模块中查找到该音乐文件的准确信息。

根据本发明实施例的音乐文件信息修正方法主要包括如下步骤：

步骤 S1，利用现有的搜索技术从互联网获取音乐文件的待修正信息。

步骤 S2，对待修正多媒体文件信息进行分词处理，具体分词处理如前所述，这里不再追述。

步骤 S3，对分词处理后的待修正多媒体文件信息进行过滤处理，去除不需要的信息。

步骤 S4，根据分词、过滤后的待修正多媒体文件信息，查找与待修正多媒体文件相关的已确定信息，并用查找到的已确定信息替换待修正多媒体文件信息。

以音乐文件信息为例，该步骤可以进一步包括：如果查找到已确定歌曲名，则用所查找到的歌曲名替换待修正音乐文件歌曲名；如果所查找到的歌曲名存在唯一对应的歌手名，则用该歌手名替换待修正音乐文件歌手名，否则用所查找到的歌手名替换待修正音乐文件歌手名；如果所查找到的歌曲名和/或歌手名存在唯一对应的专辑名，则用该专辑名替换待修正音乐文件专辑名，否则用所查找到的专辑名替换待修正音乐文件专辑名。这里所述的所查找到的歌曲名、歌手名和专辑名为精确匹配

的歌曲名、歌手名和专辑名或相似度最高的歌曲名、歌手名和专辑名。

如果预先保存有多媒体文件的哈希码与该多媒体文件的已确定信息之间的对应关系，则本步骤中根据待修正多媒体文件的哈希码查找对应的已确定信息作为与待修正多媒体文件相关的已确定信息，并用查找到的已确定信息替换待修正多媒体文件信息。

步骤 S5，输出修正后的多媒体文件信息，并进一步展现给用户。

这样，通过本发明实施例的多媒体文件信息修正系统和方法的使用，可以利用已经确定的多媒体文件信息对各种不准确的多媒体文件信息进行有效的修正，得到多媒体文件的准确信息。由于不需要人工模板等工具，本发明实施例的多媒体文件修正系统和方法还具有很高的效率。

以上所述仅是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以作出若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

权利要求书

1、一种多媒体文件信息修正系统，其特征在于，该系统包括信息保存模块和信息修正模块，其中：

所述信息保存模块用于保存多媒体文件的已确定信息；

所述信息修正模块用于在所述信息保存模块中查找与待修正多媒体文件相关的已确定信息，并用查找到的已确定信息替换待修正多媒体文件信息。

2、根据权利要求1所述的系统，其特征在于，该系统进一步包括分词模块和/或过滤模块，其中：

所述分词模块用于对待修正多媒体文件信息进行分词处理；

所述过滤模块用于对待修正多媒体文件信息进行过滤，删除不需要的信息。

3、根据权利要求1所述的系统，其特征在于，

所述信息保存模块中保存有多媒体文件的哈希码与该多媒体文件的已确定信息之间的对应关系；

所述信息修正模块根据待修正多媒体文件的哈希码查找对应的已确定信息作为与待修正多媒体文件相关的已确定信息。

4、根据权利要求1所述的系统，其特征在于，该系统进一步包括输出模块，所述输出模块用于输出多媒体文件修正后的信息。

5、根据权利要求1所述的系统，其特征在于，所述多媒体文件信息修正系统与多媒体文件搜索系统相连，所述多媒体文件搜索系统用于从互联网设备和/或本地设备获取多媒体文件信息，并提供给多媒体文件信息修正系统作为待修正多媒体文件的信息。

6、一种多媒体文件信息修正方法，其特征在于，该方法包括：

A. 根据待修正多媒体文件的信息在多媒体文件的已确定信息中查找与待修正多媒体文件信息相关的已确定信息;

B. 用查找到的已确定信息替换待修正多媒体文件信息。

7、根据权利要求 6 所述的方法, 其特征在于, 步骤 A 之前进一步包括: 对待修正多媒体文件信息进行分词处理; 和/或,

对待修正多媒体文件信息进行过滤, 删除不需要的信息。

8、根据权利要求 6 所述的方法, 其特征在于, 该方法预先保存有多媒体文件的哈希码与该多媒体文件的已确定信息之间的对应关系;

在步骤 B 中, 根据待修正多媒体文件的哈希码查找对应的已确定信息作为与待修正多媒体文件相关的已确定信息。

9、根据权利要求 6 所述的方法, 其特征在于, 该方法进一步包括: 输出多媒体文件修正后的信息。

10、根据权利要求 6 所述的方法, 其特征在于, 所述待修正多媒体文件信息为从互联网设备和/或本地设备获取的多媒体文件信息。

11、根据权利要求 6 所述的方法, 其特征在于, 所述多媒体文件为音乐文件;

所述多媒体文件的信息包括歌曲名、和/或歌手名、和/或专辑名。

12、根据权利要求 6 所述的方法, 其特征在于, 所述多媒体文件为音乐文件, 所述多媒体文件的信息包括歌曲名、歌手名和专辑名;

步骤 B 包括:

如果查找到已确定歌曲名, 则用所查找到的歌曲名替换待修正音乐文件歌曲名;

如果所查找到的歌曲名存在唯一对应的歌手名, 则用该歌手名替换待修正音乐文件歌手名; 否则用所查找到的歌手名替换待修正音乐文件歌手名;

如果所查找到的歌曲名和/或歌手名存在唯一对应的专辑名，则用该专辑名替换待修正音乐文件专辑名；否则用所查找到的专辑名替换待修正音乐文件专辑名。

13、根据权利要求 12 所述的方法，其特征在于，所查找到的歌曲名为精确匹配的歌曲名或相似度最高的歌曲名；和/或

所查找到的歌手名为精确匹配的歌手名或相似度最高的歌手名；和/或

所查找到的专辑名为精确匹配的专辑名或相似度最高的专辑名。

	歌曲名	歌手名	专辑名	试听	歌词	铃声	大小	格式
1	假装	蔡依林	舞娘	试听	歌词		2.4 M	wma
2	舞娘 假装	蔡依林	舞娘	试听	歌词		2.5 M	wma
3	假装	蔡依林	舞娘	试听	歌词		0.8 M	wma
4	假装	蔡依林	舞娘	试听	歌词		2.4 M	mp3
5	假装	蔡依林	舞娘	试听	歌词		3.0 M	wma
6	假装	蔡依林	舞娘	试听	歌词		2.4 M	wma
7	舞娘 假装	蔡依林	舞娘	试听	歌词		2.1 M	mp3
8	假装	蔡依林	舞娘	试听	歌词		2.4 M	wma
9	假装	蔡依林	舞娘	试听	歌词		0.8 M	wma
10	假装 爱娱乐	蔡依林	舞娘	试听	歌词		2.4 M	wma
11	假装 蔡依林 舞娘			试听			8.3 M	mp3
12	假装	蔡依林	舞娘	试听	歌词		0.2 M	mp3
13	假装	蔡依林	舞娘	试听	歌词		2.4 M	wma
14	假装	蔡依林	舞娘	试听	歌词		1.0 M	wma
15	假装	蔡依林	舞娘	试听	歌词		2.4 M	wma
16	假装 MP3 蔡依林 编			试听	歌词		未知	mp3
17	假装	蔡依林	舞娘	试听	歌词		2.4 M	wma

图 1

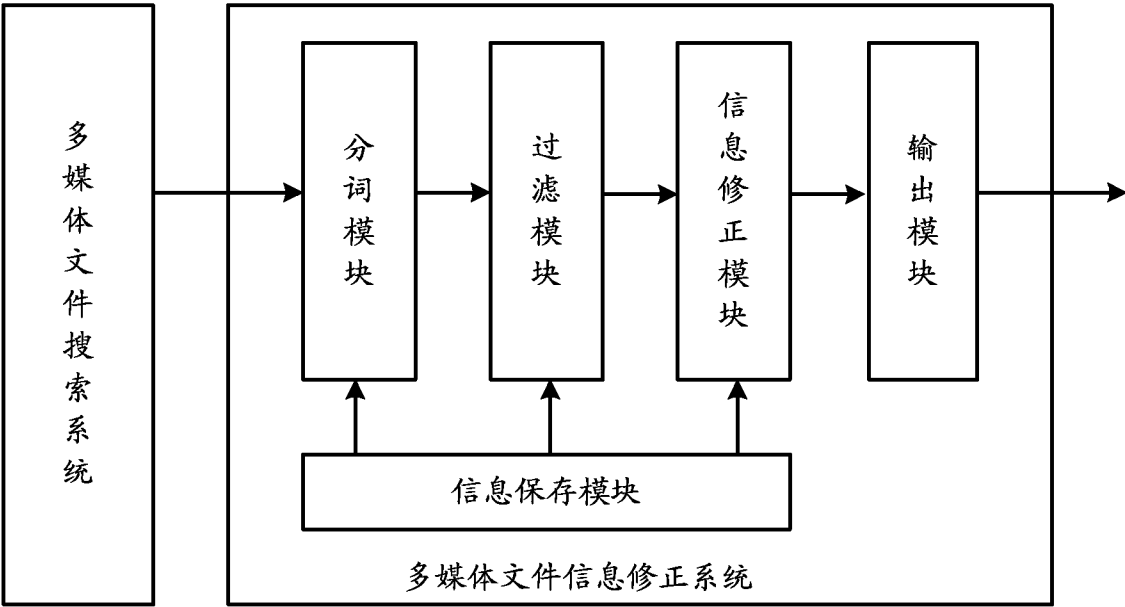


图 2

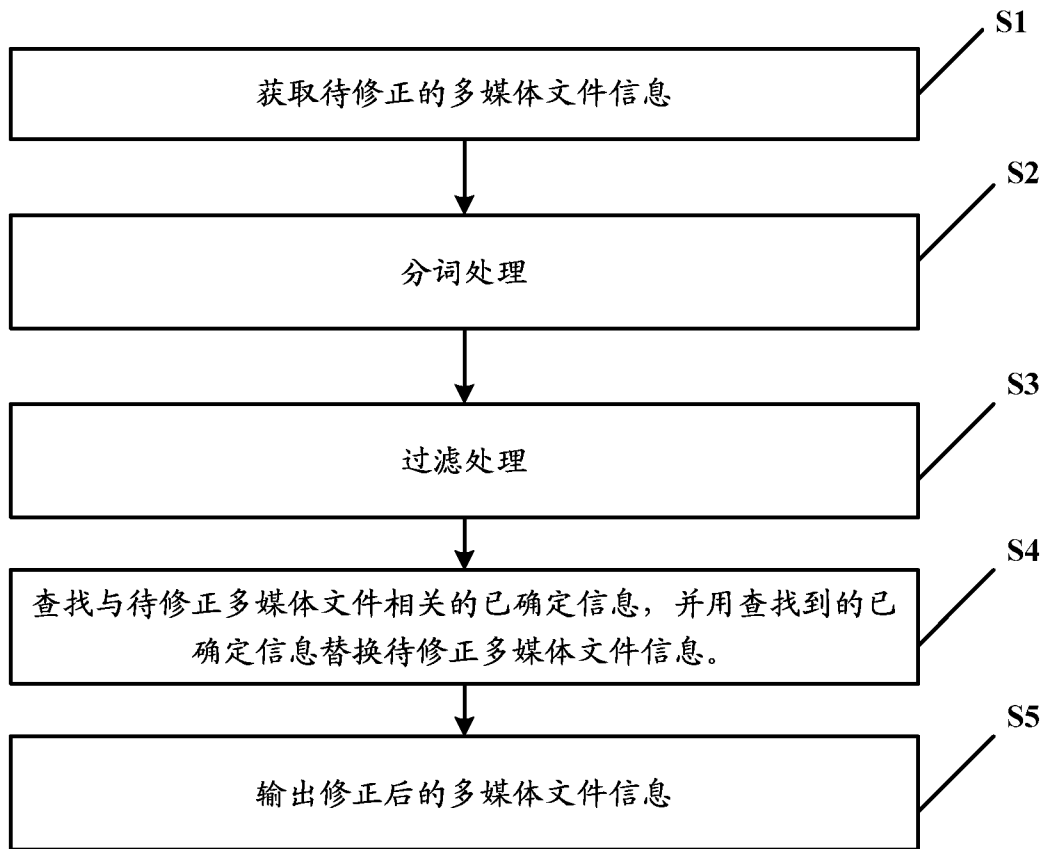


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2007/070114

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC G06F 17/-, G06F 15/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

database: WPI, EPODOC, PAJ, CNPAT, CNKI

search word: file, information, update, amend, change, substitute, multimedia, music, song, audio, video, save, store, determined, accurate, correct

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN1501395A (SONY ELECTRONICS INC) 02 Jun. 2004 (02.06.2004) see claim 1, fig. 2	1,2,4-7,9-13
A		3, 8
A	US20050065912A1 (Cafrelli et al.) 24 Mar. 2005 (24.03.2005) see the whole document	1-13
A	JP2005309712A (SHARP KK) 04 Nov. 2005(04.11.2005) see the whole document	1-13

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 07 Sep. 2007(07.09.2007)	Date of mailing of the international search report 20 Sep. 2007 (20.09.2007)
---	--

Name and mailing address of the ISA/CN
The State Intellectual Property Office, the P.R.China
6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China
100088
Facsimile No. 86-10-62019451

Authorized officer
XIE, Zhiyuan
Telephone No. (86-10)62085029

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2007/070114

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN1501395A	02.06.2004	WO9705616A1	13.02.1997
		AU6344896A	26.02.1997
		EP0840927A1	13.05.1998
		US5751672A	12.05.1998
		AU700004B	17.12.1998
		JP11514482T	07.12.1999
		MX9800613A1	01.07.1998
		EP0840927B1	01.05.2000
		KR19990035885A	25.05.1999
		DE69608281E	15.06.2000
		US6128255A	03.10.2000
		US6128625A	03.10.2000
		US6147940A	14.11.2000
		US6247022B1	12.06.2001
		US2001004338A1	21.06.2001
		US6272078B1	07.08.2001
		US6272078B2	07.08.2001
		US2001056424A1	27.12.2001
		US2001056430A1	27.12.2001
		US6388958B1	14.05.2002
		US6388957B2	14.05.2002
		CN1192285A	02.09.1998
		CN1495787A	12.05.2004
		CN1501396A	02.06.2004
		JP2004355627A	16.12.2004
		JP2005339800A	08.12.2005
		JP2006019014A	19.01.2006
		CN1160728C	04.08.2004
		JP2006073183A	16.03.2006
		JP2006073184A	16.03.2006
		JP2006073185A	16.03.2006
		KR100471585B	07.07.2005
		CA2227519C	12.12.2006
US20050065912A1	24.03.2005	none	
JP2005309712A	04.11.2005	none	

A. 主题的分类

G06F 17/30(2006.01)i

按照国际专利分类表(IPC)或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC G06F 17/-, G06F 15/-

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

数据库: WPI, EPODOC, PAJ, CNPAT, CNKI

检索词: 文件, 信息, 更新, 修正, 修改, 替换, 多媒体, 音乐, 歌曲, 音频, 视频, 保存, 存储, 确定, 准确, 精确

file, information, update, amend, change, substitute, multimedia, music, song, audio, video, save, store, determined, accurate, correct

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN1501395A (索尼电子有限公司) 02.6 月 2004 (02.06.2004) 参见权利要求 1, 图 2	1,2,4-7,9-13
A		3,8
A	US20050065912A1 (Cafrilli 等) 24.3 月 2005 (24.03.2005) 参见全文	1-13
A	JP2005309712A (夏普株式会社) 04.11 月 2005 (04.11.2005) 参见全文	1-13

☐ 其余文件在 C 栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

07.9 月 2007 (07.09.2007)

国际检索报告邮寄日期

20.9 月 2007 (20.09.2007)

中华人民共和国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路 6 号 100088

传真号: (86-10)62019451

受权官员

谢志远

电话号码: (86-10) 62085029

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2007/070114

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN1501395A	02.06.2004	WO9705616A1	13.02.1997
		AU6344896A	26.02.1997
		EP0840927A1	13.05.1998
		US5751672A	12.05.1998
		AU700004B	17.12.1998
		JP11514482T	07.12.1999
		MX9800613A1	01.07.1998
		EP0840927B1	01.05.2000
		KR19990035885A	25.05.1999
		DE69608281E	15.06.2000
		US6128255A	03.10.2000
		US6128625A	03.10.2000
		US6147940A	14.11.2000
		US6247022B1	12.06.2001
		US2001004338A1	21.06.2001
		US6272078B1	07.08.2001
		US6272078B2	07.08.2001
		US2001056424A1	27.12.2001
		US2001056430A1	27.12.2001
		US6388958B1	14.05.2002
		US6388957B2	14.05.2002
		CN1192285A	02.09.1998
		CN1495787A	12.05.2004
		CN1501396A	02.06.2004
		JP2004355627A	16.12.2004
		JP2005339800A	08.12.2005
		JP2006019014A	19.01.2006
		CN1160728C	04.08.2004
		JP2006073183A	16.03.2006
		JP2006073184A	16.03.2006
		JP2006073185A	16.03.2006
		KR100471585B	07.07.2005
		CA2227519C	12.12.2006
US20050065912A1	24.03.2005	无	
JP2005309712A	04.11.2005	无	