



(12) 发明专利

(10) 授权公告号 CN 102467490 B

(45) 授权公告日 2014. 01. 01

(21) 申请号 201010533928. 8

CN 101655835 A, 2010. 02. 24, 全文 .

(22) 申请日 2010. 11. 01

CN 1129823 A, 1996. 08. 28, 全文 .

(73) 专利权人 北大方正集团有限公司

CN 1219698 A, 1999. 06. 16, 全文 .

地址 100871 北京市海淀区成府路 298 号中
关村方正大厦 5 层

US 7380217 B2, 2008. 05. 27, 全文 .

专利权人 北京北大方正技术研究院有限公
司

审查员 朱殿尧

(72) 发明人 丁力 张磊 曹佳胤

(74) 专利代理机构 北京天昊联合知识产权代理
有限公司 11112

代理人 陈源 罗建民

(51) Int. Cl.

G06F 17/22 (2006. 01)

(56) 对比文件

CN 2779484 Y, 2006. 05. 10,

CN 101261825 A, 2008. 09. 10, 摘要 .

CN 2779484 Y, 2006. 05. 10,

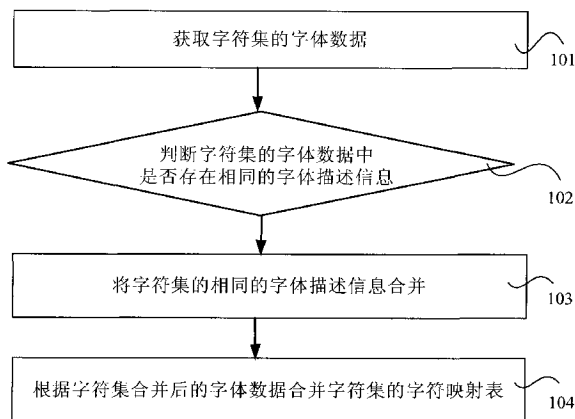
权利要求书1页 说明书6页 附图4页

(54) 发明名称

一种字体数据的处理方法及装置

(57) 摘要

本发明提供一种字体数据的处理方法及装置, 其中, 所述方法包括: 获取字符集的字体数据; 判断所述字符集的字体数据中是否存在相同的字体描述信息; 将所述字符集的字体数据中相同的字体描述信息合并; 根据所述字符集合并后的字体数据合并所述字符集的字符映射表。本实施例中, 通过将字符集的相同字体描述信息合并, 减少了电子文档中字体描述信息和字体名称的种类和数量, 相应减少了字符映射表的数量和复杂度, 同时降低了电子文档的大小, 从而减少了解析电子文档时发生的 I/O 操作次数, 提高了电子文档的解析速度。



1. 一种字体数据的处理方法,其特征在于,包括:
获取电子文档中字符集的字体数据;
判断所述字符集的字体数据中是否存在相同的字体描述信息;
将所述字符集的字体数据中相同的字体描述信息合并;
根据所述字符集合并后的字体数据合并所述字符集的字符映射表。
2. 根据权利要求1所述的字体数据的处理方法,其特征在于,在所述获取字符集的字体数据之前还包括:
获取所述字符集的编码信息以得到所述字符集。
3. 根据权利要求1所述的字体数据的处理方法,其特征在于,所述获取字符集的字体数据还包括:
获取所述字符集的字符映射表,所述字符映射表包括字符编码和字体数据索引。
4. 根据权利要求1所述的字体数据的处理方法,其特征在于,所述根据所述字符集合并后的字体数据合并所述字符集的字符映射表之后包括:
将所述字符集合并后的字体数据和字符映射表更新存储到所述字符集的电子文档中。
5. 根据权利要求1所述的字体数据的处理方法,其特征在于,还包括:
将所述字符集合并后的字体数据存储在一个字体数据文件中。
6. 一种字体数据的处理装置,其特征在于,包括:
获取单元,用于获取电子文档中字符集的字体数据;
判断单元,用于判断所述字符集中各字符的字体描述信息是否相同;
合并单元,用于将所述字符集的字体数据中相同的字体描述信息合并,以及根据所述字符集合并后的字体数据合并所述字符集的字符映射表。
7. 根据权利要求6所述的字体数据的处理装置,其特征在于:
所述获取单元还用于获取所述字符集的编码信息以得到所述字符集。
8. 根据权利要求6所述的字体数据的处理装置,其特征在于:
所述获取单元还用于获取所述字符集的字体数据的字符映射表,所述字符映射表包括字符编码和字体数据索引。
9. 根据权利要求6所述的字体数据的处理装置,其特征在于,还包括:更新单元;
所述更新单元用于将所述字符集合并后的字体数据和字符映射表更新存储到所述字符集的电子文档中。
10. 根据权利要求6所述的字体数据的处理装置,其特征在于:
所述合并单元还用于将所述字符集合并后的字体数据存储在一个字体数据文件中。

一种字体数据的处理方法及装置

技术领域

[0001] 本发明涉及字体技术领域,具体地,涉及一种字体数据的处理方法及装置。

背景技术

[0002] 电子文档在存储和展示的时候,包含图形、图片、表格、公式、多种语言文字的字符等,而文字字符通常是电子文档内容最主要的元素,在电子文档中占的比重也最大,字符的字体数据作为一种资源,存储了电子文档的字符集中各个字符的字体描述信息,在展示电子文档的时候,根据电子文档中的字符集及其对应字体数据中的字体描述信息渲染出字符集的图像,然后显示到计算机屏幕或者输出到打印机上。

[0003] 要真正原版原式的展示电子文档中字符集的内容,必须存储用户设置的字符集颜色、字体、字号大小等属性信息,并将字符集与字体数据整合为一个整体,以保证同一电子文档在任何显示终端输出同样的结果。现有技术中的处理方式为:将字体数据进行处理,以去掉部分冗余的信息,例如去掉 OpenType 字体规范格式中的 glyf 中对字形的描述,但是由于对字体数据中的其它信息没有做处理,所以去掉部分冗余的信息的字体与完整字体的使用方式没有区别。

[0004] 目前,电子文档使用频率越来越多,一个电子文档中往往会存储几十甚至成百上千种的字体描述信息,过多的字体数据导致电子文档很大,同时会使解析电子文档时的 IO 操作过多,从而使电子文档的解析速度很慢。

发明内容

[0005] 为解决上述问题,本发明的实施例提供一种字体数据的处理方法及装置,用于解决现有技术中过多的字体数据导致电子文档的大小很大,解析电子文档时的 IO 操作过多的问题。

[0006] 为此,本发明提供一种字体数据的处理方法,其中,包括:

[0007] 获取字符集的字体数据;

[0008] 判断所述字符集的字体数据中是否存在相同的字体描述信息;

[0009] 将所述字符集的字体数据中相同的字体描述信息合并;

[0010] 根据所述字符集合并后的字体数据合并所述字符集的字符映射表。

[0011] 其中,在所述获取字符集的字体数据之前还包括:

[0012] 获取所述字符集的编码信息以得到所述字符集。

[0013] 其中,所述获取字符集的字体数据还包括:

[0014] 获取所述字符集的字符映射表,所述字符映射表包括字符编码和字体数据索引。

[0015] 其中,所述根据所述字符集合并后的字体数据合并所述字符集的字符映射表之后包括:

[0016] 将所述字符集合并后的字体数据和字符映射表更新存储到所述字符集的电子文档中。

- [0017] 其中,还包括将所述字符集合并后的字体数据存储在一个字体数据文件中。
- [0018] 本发明还提供了一种字体数据的处理装置,其中,包括:
- [0019] 获取单元,用于获取字符集的字体数据;
- [0020] 判断单元,用于判断所述字符集中各字符的字体描述信息是否相同;
- [0021] 合并单元,用于将所述字符集的字体数据中相同的字体描述信息合并,以及根据所述字符集合并后的字体数据合并所述字符集的字符映射表。
- [0022] 其中,所述获取单元还用于获取所述字符集的编码信息以得到所述字符集。
- [0023] 其中,所述获取单元还用于获取所述字符集的字体数据的字符映射表,所述字符映射表包括字符编码和字体数据索引。
- [0024] 其中,还包括:更新单元;
- [0025] 所述更新单元用于将所述字符集合并后的字体数据和字符映射表更新存储到所述字符集的电子文档中。
- [0026] 其中,所述合并单元还用于将所述字符集合并后的字体数据存储在一个字体数据文件中。
- [0027] 本发明具有下述有益效果:
- [0028] 本发明提供的字体数据的处理方法,本实施例中,通过将字符集的相同字体描述信息合并,减少了电子文档中字体描述信息和字体名称的种类和数量,相应减少了字符映射表的数量和复杂度,同时降低了电子文档的大小,从而减少了解析电子文档时发生的 I/O 操作次数,提高了电子文档的解析速度。
- [0029] 本发明提供的字体数据的处理装置,通过获取单元获取电子文档中字符集的字体数据,再通过合并单元将字符集的字体数据中相同的字体描述信息合并为一项,从而减少了电子文档中字体描述信息以及字体名称的种类和数量,并根据字符集合并后的字体数据合并字符集的字符映射表,减少了字符映射表的数量,降低了电子文档的大小,并减少了解析电子文档时发生的 I/O 操作次数,提高了电子文档的解析速度。
- [0030] 附图说明
- [0031] 图 1 为本发明提供的字体数据的处理方法第一实施例的流程图;
- [0032] 图 2 为本发明提供的字体数据的处理方法第二实施例的流程图;
- [0033] 图 3 为本发明提供的字体数据的处理方法第二实施例中电子文档字符集;
- [0034] 图 4 为图 3 中电子文档字符集的字体数据;
- [0035] 图 5 为本发明提供的字体数据的处理装置第一实施例的结构示意图;
- [0036] 图 6 为本发明提供的字体数据的处理装置第二实施例的结构示意图。
- [0037] 为使本领域的技术人员更好地理解本发明的技术方案,下面结合附图对本发明提供的字体数据的处理方法及装置进行详细描述。
- [0038] 图 1 为本发明提供的字体数据的处理方法第一实施例的流程图。
- [0039] 如图 1 所示,本实施例字体数据的处理方法的具体流程包括如下步骤:
- [0040] 步骤 101、获取字符集的字体数据。
- [0041] 在利用计算机终端等智能终端打开一份电子文档时,首先获取计算机终端上的字体数据、该电子文档上字符集的编码信息以及描述该字符集的字体数据,描述该字符集的字体数据包括描述电子文档字符集中每个字符的字体描述信息等,字体数据包括字符的

号、字体和颜色等信息, 字号例如小四或 13 号等, 字体例如楷体 (Type2)、隶书 (TrueType) 和行楷 (OpenType) 等, 其中, Type2、TrueType 和 OpenType 为字体格式规范, 描述该字符集的字体数据通常按照一定的标准存储在不同的字体数据文件中, 在本实施例中, 可以按照字体格式规范的标准将同一种字体格式规范的字体数据存储在同一个字体数据文件中; 获取字符集的字体数据后, 进入步骤 102。

[0042] 步骤 102、判断字符集的字体数据中是否存在相同的字体描述信息。

[0043] 字体数据包括各字符的字体描述信息, 判断字符集的字体数据中是否存在相同的字体描述信息, 本实施例中, 字体数据可以包括分别关于字号、字体、颜色等的字体描述信息, 例如对于字符“中”和“国”, 如果判断出“中”和“国”的字号都为 12 号则说明关于“中”和“国”的字号的字体描述信息相同, 如果判断出“中”和“国”的字号的颜色都为蓝色, 则说明关于“中”和“国”的颜色的字体描述信息相同, 如果字符“中”的字体描述信息中的字体为楷体 (Type2)、字符“国”的字体描述信息中的字体为楷体 (TrueType), 由于字符“中”和“国”得字体的名称都是楷体, 同样也表明关于“中”和“国”的字体描述信息是相同。判断字符集的字体数据中存在相同的字体描述信息之后, 进入步骤 103。

[0044] 步骤 103、将字符集的相同的字体描述信息合并。

[0045] 在本实施例中, 可以将字符“中”的楷体 (Type2) 的字体描述信息和“国”的楷体 (TrueType) 的字体描述信息合并为一项字体描述信息, “中”和“国”合并后的字体为楷体, 同时, 将“中”和“国”的关于字号的字体描述信息合并为一项, 合并后的字号为 12 号, 将“中”和“国”的关于颜色的字体描述信息合并为一项, 合并后的字号为蓝色。得到描述该电子文档字符集的合并后的字体描述信息之后, 进入步骤 104。

[0046] 步骤 104、根据字符集合并后的字体数据合并字符集的字符映射表。

[0047] 各项相同的字体描述信息合并后, 得到描述该电子文档字符集合并后的字体数据, 优选地, 将合并后的字体数据存储在一个字体数据文件中, 然后根据合并后的字体数据文件相应修改字符集的字符映射表, 字符映射表用于表示各字符对应的字体描述信息在字体数据文件中的位置信息, 字体数据合并后, 字符集的字体描述信息的数量和种类明显减少, 使字符集对应的字符映射表的数量明显减少, 结构也更简单明了。

[0048] 本实施例中, 通过将字符集的相同字体描述信息合并, 减少了电子文档中字体描述信息和字体名称的种类和数量, 相应减少了字符映射表的数量和复杂度, 同时降低了电子文档的大小, 从而减少了解析电子文档时发生的 I/O 操作次数, 提高了电子文档的解析速度。

[0049] 图 2 为本发明提供的字体数据的处理方法第二实施例的流程图。如图 2 所示, 本实施例字体数据的处理方法的具体流程包括如下步骤:

[0050] 步骤 201、读取字符集的编码信息以得到字符集。

[0051] 在需要查阅电子文档上存储的字符集时, 首先要获取关于该电子文档的所有字符的编码信息, 其中, 字符的编码方式例如 Unicode、GBK 等, Unicode 或 GBK 等编码为每种语言中的每个字符设定了统一的二进制编码, 智能终端是通过读取电子文档中字符集的 Unicode 或 GBK 等编码来得到字符集。得到电子文档中的字符集之后, 进入步骤 202。

[0052] 步骤 202、获取字符集的字体数据。

[0053] 字符集的字体数据中包括该电子文档中所有字符的字体描述信息。本实施例以

PDF 文档转化为 CEBX 文档时为例来介绍技术方案,图 3 为本发明提供的字体数据的处理方法第二实施例中电子文档字符集,图 4 为图 3 中电子文档字符集的字体数据。获取图 3 所示的 PDF 文档中的字符集的字体数据包括图 4 所示的 6 种字体描述信息,字符集的 6 种字体描述信息包括数学公式字符集和文字字符集,其中,数学公式字符集的两种字体分别为 Cambria Math 和 Calibri,文字字符集的四钟字体分别为楷体 (Type2)、楷体 (TrueType)、微软雅黑 (Type2) 和微软雅黑 (TrueType),在本实施例中,微软雅黑 (Type2) 对应的字体数据存储在第一字体数据文件中,微软雅黑 (TrueType) 对应的字体数据存储在第一字体数据文件中,字体数据文件嵌入存储在电子文档中。

[0054] 步骤 203、获取字符集中各个字符的字符映射表。

[0055] 以图 3 所示的“勾股定理”为例,其中,“勾”和“股”的字体为微软雅黑 (Type2) 字体,其字体数据存储在第一字体数据文件中,字符编码与字体数据索引之间关系的字符映射表如表 1 所示,其中,字体数据索引中 0001 和 0002 用于表示字符对应的字体描述信息在第一字体数据文件中的位置。

[0056]

字符的编码信息 (十六进制)	字体数据索引
52FE	0001
80A1	0002

[0057] 表 1

[0058] “定”和“理”的字体为微软雅黑 (TrueType) 字体,其字体数据存储在第一字体数据文件中,其字符映射表如表 2 所示,其中,字体数据索引中 0001 和 0002 用于表示字符对应的字体描述信息在第一字体数据文件中的位置。

[0059]

字符的编码信息 (十六进制)	字体数据索引
5B9A	0001
7406	0002

[0060] 表 2

[0061] 在获取字符集中的各种字体数据的字体数据文件以及字符映射表之后,进入步骤 204。

[0062] 步骤 204、判断字符集的字体数据中是否存在相同的字体描述信息。

[0063] 根据字符集中各字符的字体描述信息,判断描述字符字号的各字体描述信息是否相同、判断描述字符颜色的字体描述信息是否相同。本实施例中是将 PDF 文档转换为 CEBX 文档,在 CEBX 文档中,数学公式字符集的字体 Cambria Math 和 Calibri 具有相同的字体描述信息,楷体 (Type2) 和楷体 (TrueType) 具有相同的字体描述信息,微软雅黑 (Type2) 和微软雅黑 (TrueType) 具有相同的字体描述信息。在字符集的字体数据中存在相同的字体描述信息时,进入步骤 205。

[0064] 步骤 205、将字符集的相同字体描述信息合并,以得到字符集合并后的字体数据。

[0065] 按照 TrueType 和 Type2 这两种字体格式规范,将图 3 字符集的相同字体描述信息合并,合并前后的字体如表 3 所示。

[0066]

字符集	合并前的字体	合并后的字体
公式字符集	Cambria Math Calibri	Cambria Math
楷体	楷体(Type2) 楷体(TrueType)	楷体
微软雅黑字符集	微软雅黑(Type2) 微软雅黑(TrueType)	微软雅黑

[0067] 表 3

[0068] 步骤 206、根据字符集合并后的字体数据合并字符集的字符映射表。

[0069] 在本实施例中,字体数据合并之前“勾”和“股”的字符映射表如表 1 所示,字体数据合并之前“定”和“理”的字符映射表如表 2 所示,由步骤 204 可知,微软雅黑 (Type2) 字体和微软雅黑 (TrueType) 字体的字体描述信息相同,可以合并为一项字体描述信息,从而得到合并后的字体数据,优选地,合并后的字体数据存储在一个字体数据文件中,根据合并后的字体数据相应修改映射字符集与字符编码之间关系的字符映射表,从而得到合并后的字体数据的字符映射表如表 4 所示。

[0070]

字符编码 (十六进制)	字体数据索引
52FE	0001
80A1	0002
5B9A	0003
7406	0004

[0071] 表 4

[0072] 步骤 207、将字符集合并后的字体数据和字符映射表更新存储在字符集的电子文档中。

[0073] 将合并后的字体数据和字符映射表嵌入到该电子文档中保存,也就是将存储合并后的字体数据的字体数据文件及其字符映射表嵌入到该电子文档中保存,由于字符集的字体数据的总的大小得到缩小,并且字符映射表的数量得到减少,从而缩小了存储有合并后的字体数据和字符映射表的大小,进一步减少了电子文档的大小,并在再次打开电子文档时,相应减少解析电子文档时发生的 I/O 操作次数。

[0074] 本发明实施例中,通过将字符的相同字体描述信息合并,减少了电子文档中字体描述信息及字体名称的种类和数量,缩小了字符集的字体数据的大小,并简化了字符映射表,降低了电子文档的大小,从而减少了解析电子文档时发生的 I/O 操作次数,提高了电子文档的解析速度。

[0075] 图 5 为本发明提供的字体数据的处理装置第一实施例的结构示意图。如图 5 所示,本实施例提供的字体数据的处理装置包括:获取单元 501、判断单元 502 和合并单元 503,其中,获取单元 501 用于获取电子文档中所有字符集的字体数据,判断单元 502 用于判断字符集中各字符的字体描述信息是否相同,如果判断各字符的字体描述信息相同,合并单元

503 将字符集的字体数据中相同字体描述信息合并为一项字体描述信息,以减少字体描述信息的数量,以及根据字符集合并后的字体数据合并字符集的字符映射表,本装置可以应用于计算机、打印机等电子文档的显示装置上。

[0076] 本实施例中,通过获取单元获取电子文档中字符集的字体数据,再通过合并单元将字符集的字体数据中相同的字体描述信息合并为一项,从而减少了电子文档中字体描述信息以及字体名称的种类和数量,并根据字符集合并后的字体数据合并字符集的字符映射表,减少了字符映射表的数量,降低了电子文档的大小,并减少了解析电子文档时发生的 I/O 操作次数,提高了电子文档的解析速度。

[0077] 进一步的,获取单元 501 还可以用于获取字符集的编码信息以获取字符集,以及还可以获取字符集的字符映射表,字符映射表包括字符编码和字体数据索引。

[0078] 图 6 为本发明提供的字体数据的处理装置第二实施例的结构示意图。如图 6 所示,本实施例提供的字体数据的处理装置还包括更新单元 504,更新单元 504 将字符集合并后的字体数据和字符映射表更新存储到字符集的电子文档中,也就是将存储合并后的字体数据的字体数据文件及其字符映射表嵌入到该电子文档中保存,使字符集的字体数据的总的大小被缩小,并且使字符映射表的数量相应减少,从而缩小电子文档的大小,并在再次打开电子文档时,相应减少解析电子文档时发生的 I/O 操作次数。

[0079] 进一步的,合并单元 503 还用于将字符集合并后的字体数据存储在一个字体数据文件中,更新单元 504 将该字体数据文件更新存储到字符集的电子文档中。

[0080] 本实施例中,通过获取单元获取电子文档中所有字符集的字体数据,通过判断单元判断各字符的字体描述信息是否相同,合并单元将相同的字体描述信息合并,减少电子文档中字体描述信息以及字体名称的种类和数量,从而减少了字体数据的总的大小,合并单元还将根据字符集合并后的字体数据合并字符集的字符映射表,使字符映射表的数量也相应减少,从而降低电子文档的大小,并减少解析电子文档时发生的 I/O 操作次数,提高了电子文档的解析速度。

[0081] 可以理解的是,以上实施方式仅仅是为了说明本发明的原理而采用的示例性实施方式,然而本发明并不局限于此。对于本领域内的普通技术人员而言,在不脱离本发明的精神和实质的情况下,可以做出各种变型和改进,这些变型和改进也视为本发明的保护范围。

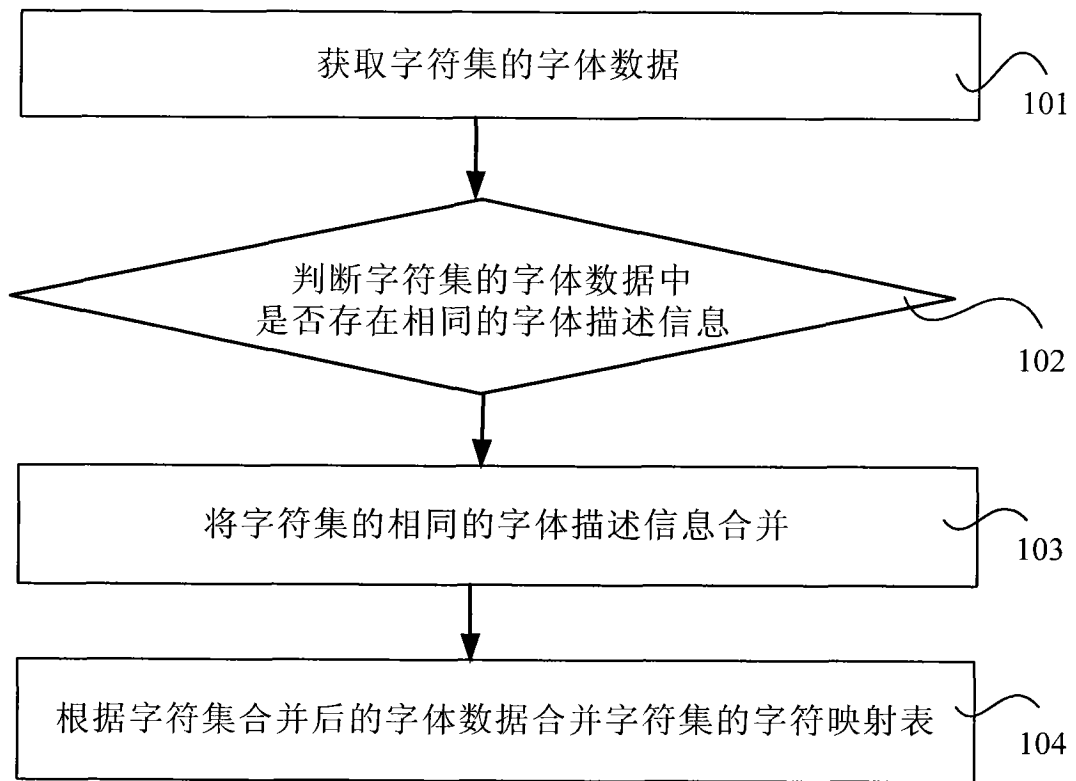


图 1

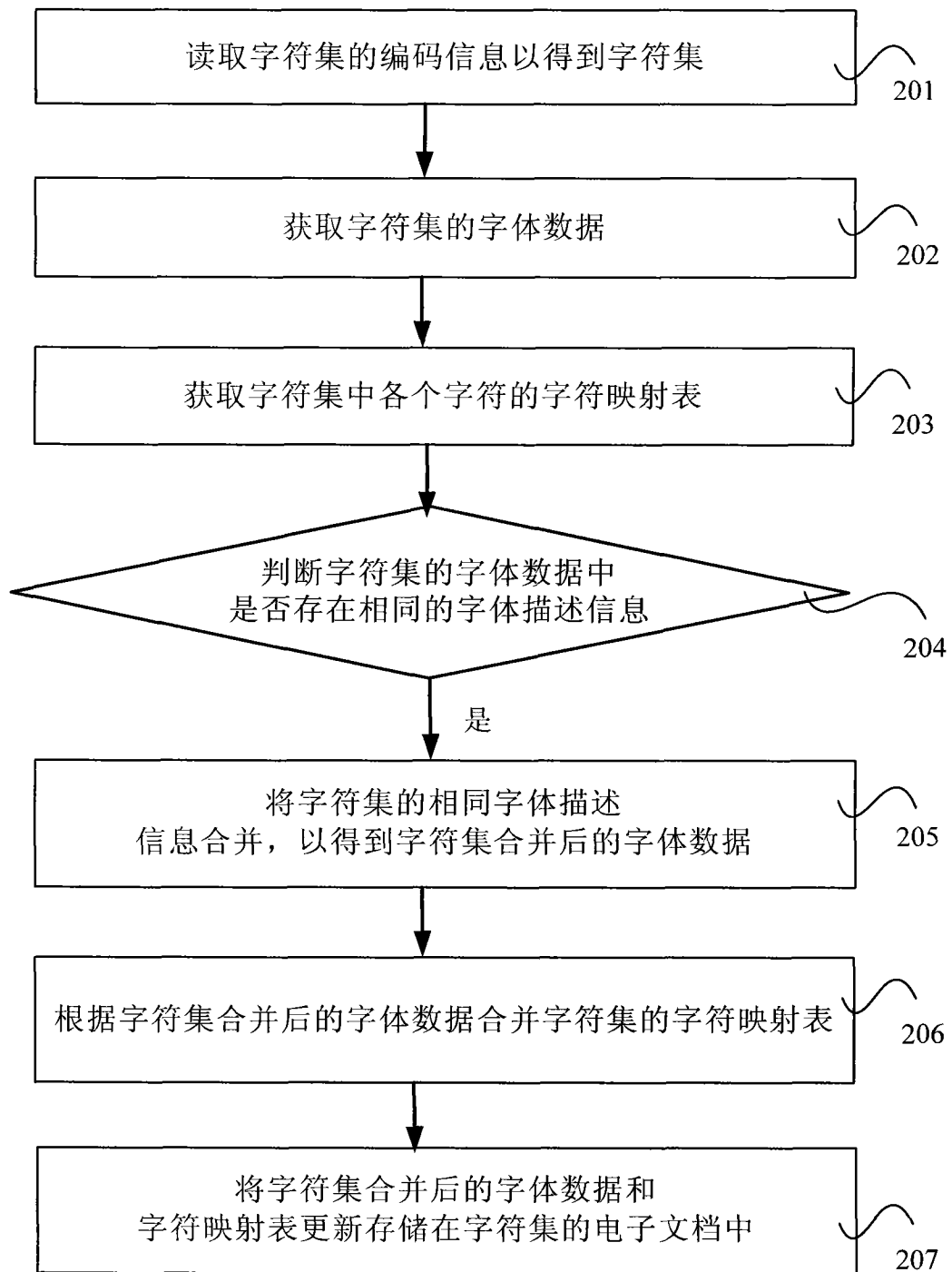


图 2

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right)$$

傅里叶级数

$$a^2 + b^2 = c^2$$

勾股定理

图 3

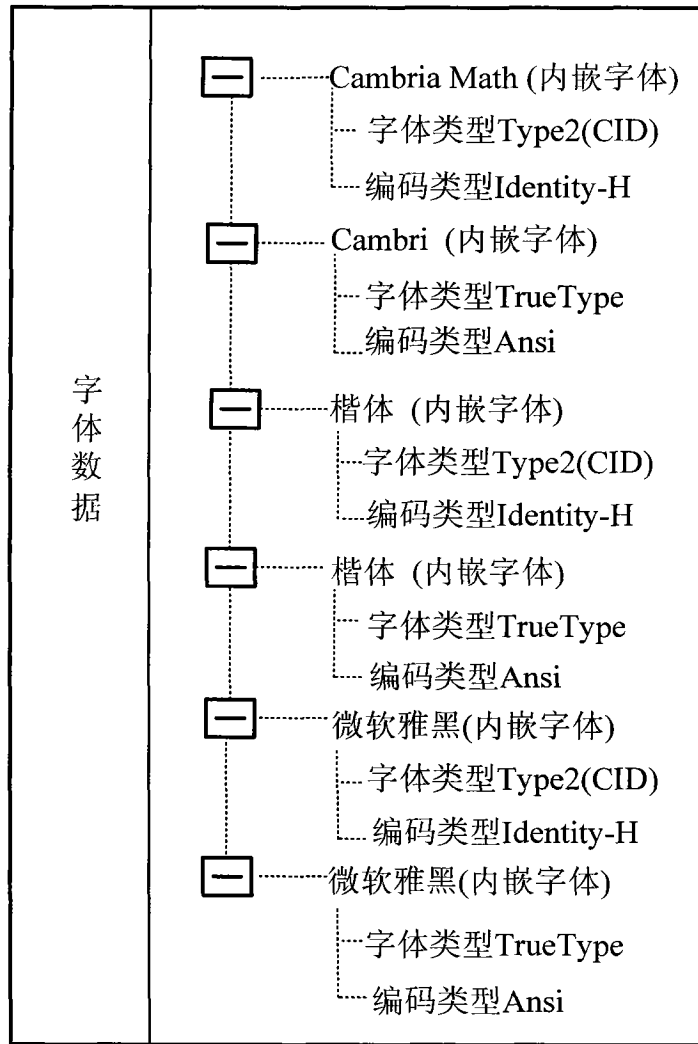


图 4

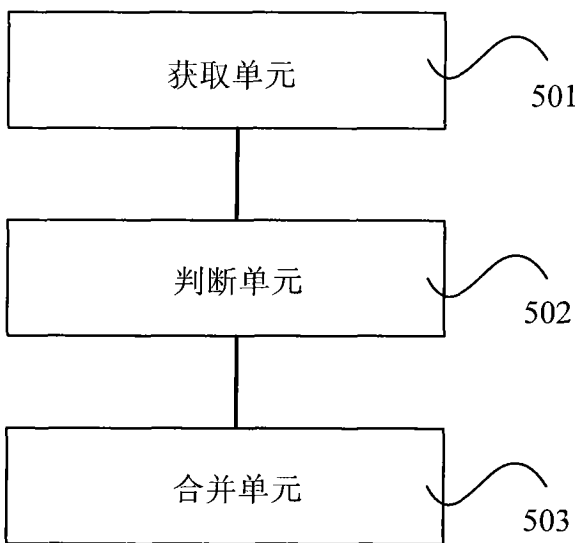


图 5

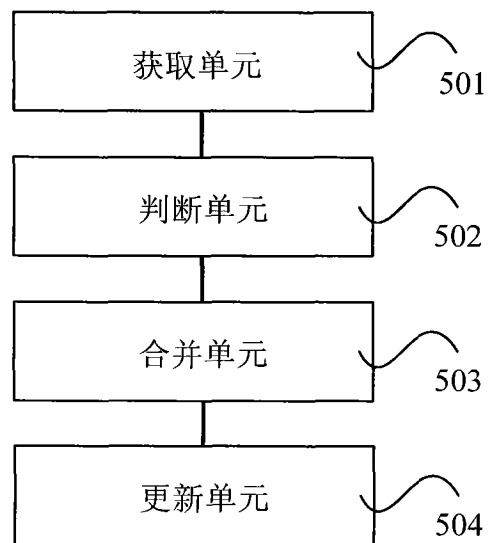


图 6