



(12)发明专利

(10)授权公告号 CN 104090915 B

(45)授权公告日 2017.02.15

(21)申请号 201410260916.0

审查员 田竞

(22)申请日 2014.06.12

(65)同一申请的已公布的文献号

申请公布号 CN 104090915 A

(43)申请公布日 2014.10.08

(73)专利权人 小米科技有限责任公司

地址 100085 北京市海淀区清河中街68号  
华润五彩城购物中心二期13层

(72)发明人 张磊

(74)专利代理机构 北京三高永信知识产权代理  
有限责任公司 11138

代理人 刘映东

(51)Int.Cl.

G06F 17/30(2006.01)

权利要求书3页 说明书16页 附图7页

(54)发明名称

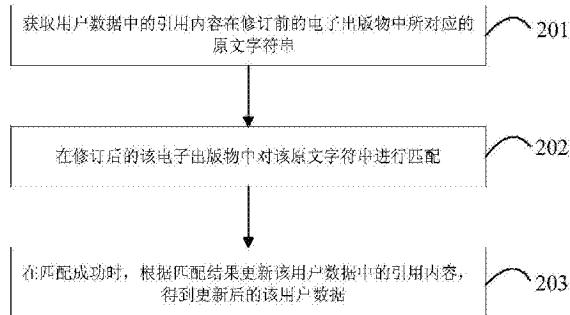
用户数据更新方法及装置

(57)摘要

本公开是关于一种用户数据更新方法及装置，属于电子出版物领域。所述方法包括：获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串；在修订后的该电子出版物中对该原文字符串进行匹配；在匹配成功时，根据匹配结果更新该用户数据中的引用内容，得到更新后的该用户数据。所述装置包括：获取模块、匹配模块和更新模块。本公开通过获取用户数据引用的原文字符串，在修订后的电子出版物只能够对该原文字符串进行匹配，根据匹配结果更新用户数据中引用内容。节省了生成Patch文件的计算代价，且本公开实施例提供的用户数据更新方法无需单独维护，使用范围较广。

B

CN 104090915 B



1. 一种用户数据更新方法,其特征在于,所述方法包括:

获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串;

在修订后的所述电子出版物中对所述原文字符串进行匹配;

在匹配成功时,根据匹配结果更新所述用户数据中的引用内容,得到更新后的所述用户数据;

所述在修订后的所述电子出版物中对所述原文字符串进行匹配,包括:

对所述原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的所述电子出版物中按序对所述n个句子进行匹配;

或,

在修订后的所述电子出版物中对所述原文字符串直接进行匹配;在直接匹配不成功时,对所述原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的所述电子出版物中按序对所述n个句子进行匹配;

所述在修订后的所述电子出版物中按序对所述n个句子进行匹配,包括:

从搜索起始位置对第i个句子进行匹配,1≤i≤n;当i=1时,所述搜索起始位置为所述原文字符串所在章节的起始位置;

在所述第i个句子匹配成功时,将所述搜索起始位置更新为所述第i个句子的匹配内容的结束位置,并从更新后的所述搜索起始位置对第i+1个句子进行匹配,i+1≤n;

在所述第i个句子匹配不成功时,从所述第i个句子所对应的搜索位置对第i+1个句子进行匹配;

在所述n个句子全部匹配后,若存在与所述n个句子完全匹配的匹配内容或存在与所述n个句子中的部分个句子匹配,但匹配程度属于预定匹配范围时,确定分句匹配成功,所述匹配程度为匹配到的字符串的长度与所述原文字符串的长度的比值。

2. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

若所述n个句子中匹配成功的最先一个句子不是第1个句子,则将所述第1个句子至所述匹配成功的最先一个句子之前的所有句子进行分词,得到按序排列的m个分词,m为正整数;

从所述匹配成功的最先一个句子的匹配内容的起始位置开始,逆序对所述m个分词进行匹配,所述从所述匹配成功的最先一个句子的匹配内容的起始位置开始,逆序对所述m个分词进行匹配包括:从所述匹配成功的最先一个句子的匹配内容的起始位置开始向第1个句子的起始位置匹配,在匹配的过程中,从所述m个分词中逆序地取词进行匹配;

将匹配成功的最后一个分词的匹配内容的起始位置作为所述原文字符串在修订后的所述电子出版物中所对应的起始位置。

3. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

若所述n个句子中匹配成功的最后一个句子不是第n个句子,则将所述匹配成功的最后一个句子之后至所述第n个句子之间的所有句子进行分词,得到按序排列的q个分词,q为正整数;

从所述匹配成功的最后一个句子的匹配内容的结束位置开始,顺序对所述q个分词进行匹配;

将匹配成功的最后一个分词的匹配内容的结束位置作为所述原文字符串在修订后的

所述电子出版物中所对应的结束位置。

4. 一种用户数据更新装置,其特征在于,所述装置包括:

获取模块,被配置为获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串;

匹配模块,被配置为在修订后的所述电子出版物中对所述原文字符串进行匹配;

更新模块,被配置为在匹配成功时,根据匹配结果更新所述用户数据中的引用内容,得到更新后的所述用户数据;

所述匹配模块,被配置为对所述原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的所述电子出版物中按序对所述n个句子进行匹配;

或,

所述匹配模块,被配置为在修订后的所述电子出版物中对所述原文字符串直接进行匹配;在直接匹配不成功时,对所述原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的所述电子出版物中按序对所述n个句子进行匹配;

所述匹配模块,包括:

匹配单元,被配置为从搜索起始位置对第i个句子进行匹配,1≤i≤n;当i=1时,所述搜索起始位置为所述原文字符串所在章节的起始位置;

位置更新单元,被配置为在所述第i个句子匹配成功时,将所述搜索起始位置更新为所述第i个句子的匹配内容的结束位置,并从更新后的所述搜索起始位置对第i+1个句子进行匹配,i+1≤n;

继续匹配单元,被配置为在所述第i个句子匹配不成功时,从所述第i个句子所对应的搜索位置对第i+1个句子进行匹配;

确定单元,被配置为在所述n个句子全部匹配后,若存在与所述n个句子完全匹配的匹配内容或存在与所述n个句子中的部分个句子匹配,但匹配程度属于预定匹配范围时,确定分句匹配成功,所述匹配程度为匹配到的字符串的长度与所述原文字符串的长度的比值。

5. 根据权利要求4所述的装置,其特征在于,所述装置还包括:

第一分词单元,被配置为在所述n个句子中匹配成功的最先一个句子不是第1个句子时,将所述第1个句子至所述匹配成功的最先一个句子之前的所有句子进行分词,得到按序排列的m个分词,m为正整数;

分词匹配单元,被配置为从所述匹配成功的最先一个句子的匹配内容的起始位置开始,逆序对所述m个分词进行匹配,所述从所述匹配成功的最先一个句子的匹配内容的起始位置开始,逆序对所述m个分词进行匹配包括:从所述匹配成功的最先一个句子的匹配内容的起始位置开始向第1个句子的起始位置匹配,在匹配的过程中,从所述m个分词中逆序地取词进行匹配;

起始确定单元,被配置为将匹配成功的最后一个分词的匹配内容的起始位置作为所述原文字符串在修订后的所述电子出版物中所对应的起始位置。

6. 根据权利要求5所述的装置,其特征在于,

第二分词单元,被配置为在所述n个句子中匹配成功的最后一个句子不是第n个句子,则将所述匹配成功的最后一个句子之后至所述第n个句子之间的所有句子进行分词,得到按序排列的q个分词,q为正整数;

分词匹配单元,被配置为从所述匹配成功的最后一个句子的匹配内容的结束位置开始,顺序对所述q个分词进行匹配;

结束确定单元,被配置为将匹配成功的最后一个分词的匹配内容的结束位置作为所述原文字符串在修订后的所述电子出版物中所对应的结束位置。

7.一种用户数据更新装置,其特征在于,包括:

处理器;

用于存储所述处理器的可执行指令的存储器;

其中,所述处理器被配置为:

获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串;

在修订后的所述电子出版物中对所述原文字符串进行匹配;

在匹配成功时,根据匹配结果更新所述用户数据中的引用内容,得到更新后的所述用户数据;

所述在修订后的所述电子出版物中对所述原文字符串进行匹配,包括:

对所述原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的所述电子出版物中按序对所述n个句子进行匹配;

或,

在修订后的所述电子出版物中对所述原文字符串直接进行匹配;在直接匹配不成功时,对所述原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的所述电子出版物中按序对所述n个句子进行匹配;

所述在修订后的所述电子出版物中按序对所述n个句子进行匹配,包括:

从搜索起始位置对第i个句子进行匹配, $1 \leq i \leq n$ ;当*i*=1时,所述搜索起始位置为所述原文字符串所在章节的起始位置;

在所述第*i*个句子匹配成功时,将所述搜索起始位置更新为所述第*i*个句子的匹配内容的结束位置,并从更新后的所述搜索起始位置对第*i*+1个句子进行匹配, $i+1 \leq n$ ;

在所述第*i*个句子匹配不成功时,从所述第*i*个句子所对应的搜索位置对第*i*+1个句子进行匹配;

在所述n个句子全部匹配后,若存在与所述n个句子完全匹配的匹配内容或存在与所述n个句子中的部分个句子匹配,但匹配程度属于预定匹配范围时,确定分句匹配成功,所述匹配程度为匹配到的字符串的长度与所述原文字符串的长度的比值。

## 用户数据更新方法及装置

### 技术领域

[0001] 本公开涉及电子出版物领域,特别涉及一种用户数据更新方法及装置。

### 背景技术

[0002] 电子出版物以其修订周期短、修订代价低的优势,越来越受人们的喜爱。在阅读电子出版物时,用户可以生成诸如书签、批注和书摘等用户数据。这些用户数据包括:引用内容和针对该引用内容的用户阅读信息。

[0003] 用户数据中的引用内容一般利用物理偏移量来表示。也即,一个引用内容可以用(位置,长度)来表示,其中‘位置’表示了引用内容在整本电子出版物中的文本偏移位置,‘长度’表示了引用内容的文本长度。比如,一个批注可以表示为“引用内容(第120305字符,共32个字符),批注内容(这段文字写的真好)”。但是由于电子出版物可能会发生多次修订,在电子出版物中的内容发生增删之后,利用物理偏移量表示的引用内容很可能发生改变。为此,相关技术提供了一种用户数据更新方法:在电子出版物发生修订时,采用类似Patch(补丁)文件的方式,记录每个版本详细的修订记录,通过该修订记录来重新计算用户数据中的引用内容。

[0004] 公开人在实现本公开的过程中,发现上述方式至少存在如下缺陷:由于Patch文件在生成时计算代价非常高,且需要单独维护,所以上述用户数据更新方法不利于大范围使用。

### 发明内容

[0005] 为了解决相关技术中Patch文件在生成时计算代价非常高,且需要单独维护,上述用户数据更新方法不利于大范围使用的问题,本公开实施例提供了一种用户数据更新方法及装置,所述技术方案如下:

- [0006] 根据本公开实施例的第一方面,提供一种用户数据更新方法,该方法包括:
- [0007] 获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串;
- [0008] 在修订后的该电子出版物中对该原文字符串进行匹配;
- [0009] 在匹配成功时,根据匹配结果更新该用户数据中的引用内容,得到更新后的该用户数据。
- [0010] 一个实施例,该在修订后的该电子出版物中对该原文字符串进行匹配,包括:
- [0011] 在修订后的该电子出版物中对该原文字符串直接进行匹配;
- [0012] 或,
- [0013] 对该原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的该电子出版物中按序对该n个句子进行匹配;
- [0014] 或,
- [0015] 在修订后的该电子出版物中对该原文字符串直接进行匹配;在直接匹配不成功时,对该原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的该电子出

版物中按序对该n个句子进行匹配。

[0016] 一个实施例,该在修订后的该电子出版物中按序对该n个句子进行匹配,包括:

[0017] 从搜索起始位置对第i个句子进行匹配, $1 \leq i \leq n$ ;当*i*=1时,该搜索起始位置为该原文字符串所在章节的起始位置;

[0018] 在该第*i*个句子匹配成功时,将该搜索起始位置更新为该第*i*个句子的匹配内容的结束位置,并从更新后的该搜索起始位置对第*i*+1个句子进行匹配, $i+1 \leq n$ ;

[0019] 在该第*i*个句子匹配不成功时,从该第*i*个句子所对应的搜索位置对第*i*+1个句子进行匹配;

[0020] 在该n个句子全部匹配后,若存在与该n个句子完全匹配的匹配内容或存在与该n个句子部分匹配但匹配程度属于预定匹配范围的匹配内容时,确定分句匹配成功。

[0021] 一个实施例,该方法还包括:

[0022] 若该n个句子中匹配成功的最先一个句子不是第1个句子,则将该第1个句子至该匹配成功的最先一个句子之前的所有句子进行分词,得到按序排列的m个分词,m为正整数;

[0023] 从该匹配成功的最先一个句子的匹配内容的起始位置开始,逆序对该m个分词进行匹配;

[0024] 将匹配成功的最后一个分词的匹配内容的起始位置作为该原文字符串在修订后的该电子出版物中所对应的起始位置。

[0025] 一个实施例,该方法还包括:

[0026] 若该n个句子中匹配成功的最后一个句子不是第n个句子,则将该匹配成功的最后一个句子之后至该第n个句子之间的所有句子进行分词,得到按序排列的q个分词,q为正整数;

[0027] 从该匹配成功的最后一个句子的匹配内容的结束位置开始,顺序对该q个分词进行匹配;

[0028] 将匹配成功的最后一个分词的匹配内容的结束位置作为该原文字符串在修订后的该电子出版物中所对应的结束位置。

[0029] 根据本公开实施例的第二方面,提供一种用户数据更新装置,该装置包括:

[0030] 获取模块,被配置为获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串;

[0031] 匹配模块,被配置为在修订后的该电子出版物中对该原文字符串进行匹配;

[0032] 更新模块,被配置为在匹配成功时,根据匹配结果更新该用户数据中的引用内容,得到更新后的该用户数据。

[0033] 一个实施例,该匹配模块,被配置为在修订后的该电子出版物中对该原文字符串直接进行匹配;

[0034] 或,

[0035] 该匹配模块,被配置为对该原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的该电子出版物中按序对该n个句子进行匹配;

[0036] 或,

[0037] 该匹配模块,被配置为在修订后的该电子出版物中对该原文字符串直接进行匹配;在直接匹配不成功时,对该原文字符串进行分句,得到按序排列的n个句子,n为正整数;

在修订后的该电子出版物中按序对该n个句子进行匹配。

[0038] 一个实施例，该匹配模块，包括：

[0039] 匹配单元，被配置为从搜索起始位置对第i个句子进行匹配， $1 \leq i \leq n$ ；当*i*=1时，该搜索起始位置为该原文字符串所在章节的起始位置；

[0040] 位置更新单元，被配置为在该第*i*个句子匹配成功时，将该搜索起始位置更新为该第*i*个句子的匹配内容的结束位置，并从更新后的该搜索起始位置对第*i*+1个句子进行匹配， $i+1 \leq n$ ；

[0041] 继续匹配单元，被配置为在该第*i*个句子匹配不成功时，从该第*i*个句子所对应的搜索位置对第*i*+1个句子进行匹配；

[0042] 确定单元，被配置为在该n个句子全部匹配后，若存在与该n个句子完全匹配的匹配内容或存在与该n个句子部分匹配但匹配程度属于预定匹配范围的匹配内容时，确定分句匹配成功。

[0043] 一个实施例，该装置还包括：

[0044] 第一分词单元，被配置为在该n个句子中匹配成功的最先一个句子不是第1个句子时，将该第1个句子至该匹配成功的最先一个句子之前的所有句子进行分词，得到按序排列的m个分词，m为正整数；

[0045] 分词匹配单元，被配置为从该匹配成功的最先一个句子的匹配内容的起始位置开始，逆序对该m个分词进行匹配；

[0046] 起始确定单元，被配置为将匹配成功的最后一个分词的匹配内容的起始位置作为该原文字符串在修订后的该电子出版物中所对应的起始位置。

[0047] 一个实施例，

[0048] 第二分词单元，被配置为在该n个句子中匹配成功的最后一个句子不是第n个句子，则将该匹配成功的最后一个句子之后至该第n个句子之间的所有句子进行分词，得到按序排列的q个分词，q为正整数；

[0049] 分词匹配单元，被配置为从该匹配成功的最后一个句子的匹配内容的结束位置开始，顺序对该q个分词进行匹配；

[0050] 结束确定单元，被配置为将匹配成功的最后一个分词的匹配内容的结束位置作为该原文字符串在修订后的该电子出版物中所对应的结束位置。

[0051] 根据本公开实施例的第三方面，提供一种用户数据更新装置，该装置包括：

[0052] 处理器；

[0053] 用于存储该处理器的可执行指令的存储器；

[0054] 其中，该处理器被配置为：

[0055] 获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串；

[0056] 在修订后的该电子出版物中对该原文字符串进行匹配；

[0057] 在匹配成功时，根据匹配结果更新该用户数据中的引用内容，得到更新后的该用户数据。

[0058] 本公开实施例提供的技术方案可以包括以下有益效果：

[0059] 通过获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串；在修订后的该电子出版物中对该原文字符串进行匹配；在匹配成功时，根据匹配结果更新

该用户数据中的引用内容,得到更新后的该用户数据。采用本公开实施例提供的用户数据更新方法,节省了生成Patch文件的计算代价,且本公开实施例提供的用户数据更新方法无需单独维护,使用范围较广。

[0060] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

## 附图说明

[0061] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理。

[0062] 图1是本公开各个实施例提供的用户数据更新方法所涉及的一种实施环境的结构示意图;

[0063] 图2是根据一示例性实施例示出的一种用户数据更新方法的方法流程图;

[0064] 图3a是根据另一示例性实施例示出的一种用户数据更新方法的方法流程图;

[0065] 图3b是本发明实施例提供的一种修订前的电子出版物的某一章节与修订后的电子出版物中的对应章节的示意图;

[0066] 图4a是根据再一示例性实施例示出的一种用户数据更新方法的方法流程图;

[0067] 图4b是本发明实施例提供的另一种修订前的电子出版物的某一章节与修订后的电子出版物中的对应章节的示意图;

[0068] 图5a是根据还一示例性实施例示出的一种用户数据更新方法的方法流程图;

[0069] 图5b是本发明实施例提供的还一种修订前的电子出版物的某一章节与修订后的电子出版物中的对应章节的示意图;

[0070] 图6是根据一示例性实施例示出的一种用户数据更新装置的框图;

[0071] 图7是根据一示例性实施例示出的一种用于用户数据更新的装置的框图;

[0072] 图8是根据另一示例性实施例示出的一种用于用户数据更新的装置的框图。

[0073] 通过上述附图,已示出本公开明确的实施例,后文中将有更详细的描述。这些附图和文字描述并不是为了通过任何方式限制本公开构思的范围,而是通过参考特定实施例为本领域技术人员说明本公开的概念。

## 具体实施方式

[0074] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本公开相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本公开的一些方面相一致的装置和方法的例子。

[0075] 在对本公开的实施例做详细地解释说明之前,先对本公开的所涉及的用户数据说明。在本公开实施例中,用户数据包括引用内容和针对该引用内容的用户阅读信息。该用户数据可以为书签、书摘或批注中的任意一种。

[0076] 比如,书签可以为“第120305个字符,第3个书签”,引用内容“第120305个字符”用于表示用户上次阅读到第120305个字符的位置,用户阅读信息“第3个书签”用于表示这是本书中的第3个书签。

[0077] 又比如,书摘可以为“第120305个字符,共32个字符,第1个书摘”,引用内容“第120305个字符,共32个字符”用于表示该用户数据从第120305个字符的位置开始,一共有32个字符的长度,而用户阅读信息“第1个书摘”用于表示这是本书中的第1个书摘;或,引用内容还可以为“第120305个字符,第120337个字符”,表示该用户数据从第120305个字符的位置开始,到第120337个字符的位置结束。

[0078] 再比如,批注可以为“第120305字符,共32个字符,这段文字写的真好”,引用内容“第120305个字符,共32个字符”用于表示该批注的引用内容从第120305字符的位置开始,一共有32个字符的长度,用户阅读信息“这段文字写的真好”是用户对该引用内容的批注内容;或,引用内容还可以为“第120305个字符,第120337个字符”,表示该批注的引用内容从第120305字符的位置开始,到第120337个字符的位置结束。

[0079] 图1是本公开各个实施例提供的用户数据更新方法所涉及的一种实施环境的结构示意图。该实施环境包括至少一个终端120和至少一个服务器140。

[0080] 终端120,可以是具有阅读功能的电子设备,该电子设备可以是智能手机、平板电脑、智能电视或者电子书阅读器等等。

[0081] 终端120与服务器140之间可以通过无线网络相连。

[0082] 服务器140,可以是一台服务器,或者由若干台服务器组成的服务器集群,或者是一个云计算服务中心。该服务器140是可以向终端120提供更新用户数据服务的服务器。

[0083] 图2是根据一示例性实施例示出的一种用户数据更新方法的流程图。参见图2,本实施例以该方法应用于如图1所示的终端120中来举例说明。该方法流程包括:

[0084] 在步骤201中,获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串。

[0085] 在步骤202中,在修订后的该电子出版物中对该原文字符串进行匹配。

[0086] 在步骤203中,在匹配成功时,根据匹配结果更新该用户数据中的引用内容,得到更新后的该用户数据。

[0087] 本公开实施例提供的用户数据更新方法,通过获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串;在修订后的该电子出版物中对该原文字符串进行匹配;在匹配成功时,根据匹配结果更新该用户数据中的引用内容,得到更新后的该用户数据。采用本公开实施例提供的用户数据更新方法,节省了生成Patch文件的计算代价,且本公开实施例提供的用户数据更新方法无需单独维护,使用范围较广。

[0088] 由于上述步骤202有3种实现方式:

[0089] 第一,直接匹配方式;

[0090] 在修订后的电子出版物中对原文字符串直接进行匹配;

[0091] 第二,分词匹配方式;

[0092] 对原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的电子出版物中按序对n个句子进行匹配;

[0093] 第三,直接匹配方式与分词匹配方式结合;

[0094] 在修订后的电子出版物中对原文字符串直接进行匹配;在直接匹配不成功时,对原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的电子出版物中按序对n个句子进行匹配。

[0095] 下面,采用三个不同的实施例对上述三种方式分别进行阐述:

[0096] 图3a是根据另一示例性实施例示出的一种用户数据更新方法的流程图。参见图3a,本实施例以该方法应用于如图1所示的终端120中,且终端120采用直接匹配方式来举例说明。该方法流程包括:

[0097] 在步骤3a01中,获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串。

[0098] 在本实施例中,终端获取的用户数据中的引用内容对应的原文字符串可以为书签、书摘或批注中的任意一种所对应的原文字符串。

[0099] 在本实施例中,终端获取该用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串时,有以下两种获取方式:

[0100] 第一种方式,终端根据引用内容的起始位置和长度来获取该原文字符串。

[0101] 比如,当引用内容为:“第120305个字符,共32个字符”时,终端从当前章节的第120305个字符的位置开始,获取该第120305个字符之后的32个字符的长度的字符串,实现对该引用内容所对应的原文字符串的获取。

[0102] 第二种方式,终端根据引用内容的起始位置和结束位置来获取该原文字符串。

[0103] 比如,当引用内容为:“第120305个字符,第120337个字符”时,终端获取当前章节的第120305个字符的位置到第120337个字符的位置之间的字符串,实现对该引用内容所对应的原文字符串的获取。

[0104] 图3b是本发明实施例提供的一种修订前的电子出版物的某一章节与修订后的电子出版物中的对应章节的示意图。参见图3b,则终端获取到的用户数据中的引用内容对应的原文字符串为“周末七国纷争,并入于秦;及秦灭之后,楚、汉分争,又并入于汉;汉朝自高祖斩白蛇而起义,一统天下,。”。

[0105] 在步骤3a02中,在修订后的该电子出版物中对该原文字符串直接进行匹配。

[0106] 当终端获取到原文字符串时,终端在修订后的电子出版物中,从与修订前的电子出版物的某一章节的对应章节的起始位置开始,以获取到的原文字符串的长度为搜索长度,在与该特定章节对应的章节中对该原文字符串进行搜索,实现对该原文字符串的匹配。

[0107] 参见图3b,则终端从修订后的对应章节的起始位置“话”开始,以获取到的原文字符串的长度为搜索长度,在该对应章节中,对该原文字符串进行搜索,终端可以搜索到字符串:“周末七国纷争,并入于秦;及秦灭之后,楚、汉分争,又并入于汉;汉朝自高祖斩白蛇而起义,一统天下,”,从而终端对该原文字符串的匹配成功。

[0108] 在步骤3a03中,在匹配成功时,根据匹配结果更新该用户数据中的引用内容,得到更新后的该用户数据。

[0109] 当终端对原文字符串匹配成功时,终端根据匹配到的字符串的位置信息,对该字符串对应的用户数据的引用内容进行更新,得到更新后的用户数据,其中,该位置信息可以为匹配到的字符串在修订后的电子出版物中的起始位置和匹配到的字符串在修订后电子出版物中的结束位置,或,该位置信息还可以为匹配到的字符串在修订后的电子出版物中的起始位置和该匹配到的字符串在修订后的电子出版物中的文本长度。

[0110] 如,当用户数据为书签时,终端根据该书签的位置,在修订后的电子出版物的对应位置添加书签;当用户数据为书摘时,终端根据该书摘的起始位置和该书摘对应的引用内

容,在修订后的电子出版物的对应位置添加书摘;当用户数据为批注时,终端根据该批注的起始位置、引用内容和该批注的内容,在修订后的电子出版物的对应位置添加批注。

[0111] 以上述步骤3a02中的例子为例,当终端对原文字符串“周末七国纷争,并入于秦;及秦灭之后,楚、汉分争,又并入于汉;汉朝自高祖斩白蛇而起义,一统天下,”匹配成功,且该原文字符串所对应的用户数据为:“第120305个字符,共32个字符,第1个书摘”时,终端根据该字符串的起始位置和该字符串的内容,在修订后的电子出版物中对该字符串对应的用户数据的引用内容进行更新,得到该更新后的用户数据为:“第120305个字符,共32个字符,第1个书摘”

[0112] 本公开实施例提供的用户数据更新方法,通过获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串;在修订后的该电子出版物中对该原文字符串直接进行匹配;并在匹配成功时,根据匹配结果更新该用户数据中的引用内容,得到更新后的该用户数据。采用本公开实施例提供的用户数据更新方法,节省了生成Patch文件的计算代价,且本公开实施例提供的用户数据更新方法无需单独维护,使用范围较广。

[0113] 由于直接匹配方式仅适用于原文字符串未修订的情况,对于原文字符串进行了修订的情况,无法匹配成功。为此,提出如下实施例:

[0114] 图4a是根据再一示例性实施例示出的一种用户数据更新方法的流程图。参见图4a,本实施例以该方法应用于如图1所示的终端120中,且终端120采用分句匹配方式来举例说明。该方法流程包括:

[0115] 在步骤4a01中,获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串。

[0116] 该步骤4a01与上述图3a所示实施例中的步骤3a01相同或相似,本实施例对此不再赘述。

[0117] 在步骤4a02中,对该原文字符串进行分句,得到按序排列的n个句子,n为正整数。

[0118] 图4b是本发明实施例提供的另一种修订前的电子出版物的某一章节与修订后的电子出版物中的对应章节的示意图。以图4b中的修订前的某一章节的原文字符串为例,终端获取到的原文字符串为:“周末七国纷争,并入于秦;及秦灭之后,楚、汉分争,又并入于汉;汉朝自高祖斩白蛇而起义,一统天下,”,终端对该原文字符串进行分句,可以得到以下7个句子:(1)“周末七国纷争,”,(2)“并入于秦;”,(3)“及秦灭之后,”,(4)“楚、汉分争,”,(5)“又并入于汉;”,(6)“汉朝自高祖斩白蛇而起义,”,(7)“一统天下,”。

[0119] 在步骤4a03中,在修订后的该电子出版物中,按序对该n个句子进行匹配。

[0120] 终端在修订后的电子出版物中从搜索起始位置对第i个句子进行匹配, $1 \leq i \leq n$ ;当*i*=1时,该搜索起始位置为该原文字符串所在章节的起始位置。当对第*i*个句子匹配成功时,终端将该搜索起始位置更新为该第*i*个句子的匹配内容的结束位置,并从更新后的该搜索起始位置对第*i*+1个句子进行匹配, $i+1 \leq n$ ;当对第*i*个句子匹配不成功时,从该第*i*个句子所对应的搜索位置对第*i*+1个句子进行匹配。

[0121] 需要说明的是,在对该n个句子全部匹配后,若存在与该n个句子完全匹配的匹配内容或存在与该n个句子部分匹配但匹配程度属于预定匹配范围的匹配内容时,确定句子匹配成功。其中,该预定匹配范围衡量匹配到的句子的正确性,其目的在于防止匹配到不全的雷同字符串,同时防止匹配到过长的雷同字符串。该预定匹配范围可以用公式:80%Len1

$\leqslant \text{Len2} \leqslant 120\% \text{Len1}$  来表示, 其中, Len1为原文字符串的文本长度, Len2为匹配到的字符串的长度。

[0122] 以上述步骤4a02中得到的7个句子以及图4b所示的修订后的对应章节为例进行说明。终端将该修订后的对应章节的起始位置“话”作为搜索起始位置, 在该修订后的对应章节中, 按顺序对以上7个句子进行匹配。终端从该起始位置“话”开始, 以该第1个句子的长度为搜索长度, 在该章节中搜索该第1个句子, 则终端可以搜到该第1个句子, 从而终端对该第1个句子的匹配成功, 终端将搜索起始位置更新为该第1个句子的结束位置“,”, 从该第1个句子的结束位置“,”开始, 以该第2个句子的长度为搜索长度, 在该章节中搜索该第2个句子, 终端也可以搜索到该第2个句子, 从而终端对该第2个句子的匹配也成功, 终端将搜索起始位置更新为该第2个句子的结束位置“;”, 进而终端从该第2个句子的结束位置“;”开始, 对第3个句子进行匹配, 终端也可以匹配成功, 终端将搜索起始位置更新为第3个句子的结束位置“,”, 进而终端从该第3个句子的结束位置“,”开始, 以该第4个句子的长度为搜索长度, 在该章节中搜索该第4个句子, 直至搜索到该章节的结束位置“。”, 终端都没有搜索到该第4个句子, 于是, 终端对该第4个句子的匹配失败; 终端不更新搜索起始位置, 继续从该第3个句子的结束位置“,”开始, 以该第5个句子的长度为搜索长度, 在该章节中, 搜索该第5个句子, 则终端可以搜到该第5个句子, 从而终端对该第5个句子的匹配成功; 与此同理, 终端对第6个和第7个句子都匹配成功。当对第7个句子的匹配结束时, 终端对该原文字符串的句子匹配结束。

[0123] 在上述对该7个句子进行匹配的过程中, 终端没有匹配到第4个句子, 该对原文字符串的匹配属于部分匹配, 进而终端计算匹配到的字符串“周末七国纷争, 并入于秦; 及秦灭之后, 又并入于汉; 汉朝自高祖斩白蛇而起义, 一统天下,”的长度, 与原文字符串“周末七国纷争, 并入于秦; 及秦灭之后, 楚、汉分争, 又并入于汉; 汉朝自高祖斩白蛇而起义, 一统天下,”的长度之间的关系, 终端可以计算得原文字符串的文本长度为47, 匹配到的字符串的文本长度为41, 由于原文字符串的文本长度80%为37.5, 原文字符串的文本长度120%为56.4, 且 $37.5 < 47 < 56.4$ , 则终端对该原文字符串的匹配成功。

[0124] 需要说明的是, 在上述举例过程中, 如果终端仅匹配到字符串的文本长度长度大于56.4或小于37.5, 则此时匹配失败, 该匹配的具体过程与上述例子同理, 这里不再赘述。

[0125] 在步骤4a04中, 在匹配成功时, 根据匹配结果更新该用户数据中的引用内容, 得到更新后的该用户数据。

[0126] 该步骤4a04与上述图3a所示实施例中的步骤3a03相同或相似, 本实施例对此不再赘述。

[0127] 本公开实施例提供的用户数据更新方法, 通过获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串; 对该原文字符串进行分句, 得到按序排列的n个句子, n为正整数; 在修订后的该电子出版物中按序对该n个句子进行匹配; 并在匹配成功时, 根据匹配结果更新该用户数据中的引用内容, 得到更新后的该用户数据。采用本公开实施例提供的用户数据更新方法, 节省了生成Patch文件的计算代价, 且本公开实施例提供的用户数据更新方法无需单独维护, 使用范围较广。

[0128] 本公开实施例提供的用户数据更新方法, 还通过分句匹配方式, 实现了即便原文字符串发生句子级别的改变, 也能够完成匹配过程的效果,

[0129] 本公开实施例提供的用户数据更新方法,还通过判断匹配程度与预定匹配范围之间的关系后,根据判断结果对用户数据的引用内容进行更新,提高了更新到的用户数据的正确性。

[0130] 由于分句匹配方式仅适用于原文字符串中发生了句子级别修订的情况,对于原文字符串发生了字符级别修订的情况,无法匹配成功。为此,提出如下实施例:

[0131] 图5a是根据还一示例性实施例示出的一种用户数据更新方法的流程图。参见图5a,本实施例以该方法应用于如图1所示的终端120中,终端采用直接匹配和分句匹配方式两种结合的方式,且分句匹配不成功时还采用分词匹配方式来举例说明。该方法流程包括:

[0132] 在步骤5a01中,获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串。

[0133] 该步骤5a01与上述图3a所示实施例中的步骤3a01相同或相似,本实施例对此不再赘述。

[0134] 在步骤5a02中,在修订后的该电子出版物中对该原文字符串直接进行匹配。

[0135] 图5b是本发明实施例提供的还一种修订前的电子出版物的某一章节与修订后的电子出版物中的对应章节的示意图。参见图5b,则终端获取到的用户数据中的引用内容对应的原文字符串为“周末七国纷争,并入于秦;及秦灭之后,楚、汉分争,又并入于汉;汉朝自高祖斩白蛇而起义,一统天下。”。与图3a所示实施例中的步骤3a02同理,终端从修订后的对应章节的起始位置“话”开始,以获取到的原文字符串的长度为搜索长度,在该修订后的对应章节中,对该原文字符串进行搜索,直至搜索至该章节的结束位置“。”,终端都没有搜索到该原文字符串,终端对该原文字符串的直接匹配失败。

[0136] 在步骤5a03中,在直接匹配不成功时,对该原文字符串进行分句,得到按序排列的n个句子,n为正整数。

[0137] 以图5b中的修订前的某一章节的原文字符串为例,终端对该原文字符串进行分句,可以得到以下7个句子:(1)“周末七国纷争,”,(2)“并入于秦;”,(3)“及秦灭之后,”,(4)“楚、汉分争,”,(5)“又并入于汉;”,(6)“汉朝自高祖斩白蛇而起义,”,(7)“一统天下。”。

[0138] 在步骤5a04中,在修订后的该电子出版物中按序对该n个句子进行匹配。

[0139] 终端从搜索起始位置对第i个句子进行匹配, $1 \leq i \leq n$ ;当*i*=1时,该搜索起始位置为该原文字符串所在章节的起始位置;当对该第*i*个句子匹配成功时,将该搜索起始位置更新为该第*i*个句子的匹配内容的结束位置,并从更新后的该搜索起始位置对第*i*+1个句子进行匹配, $i+1 \leq n$ ;当对该第*i*个句子匹配不成功时,从该第*i*个句子所对应的搜索位置对第*i*+1个句子进行匹配。

[0140] 以上述步骤5a03中得到的7个句子以及图5b所示的修订后的对应章节为例进行说明。终端将该对应章节的起始位置“话”为搜索起始位置,在该对应章节中,按顺序对以上7个句子进行匹配。终端从该起始位置“话”开始,以该第1个句子的长度为搜索长度,在该对应章节中搜索该第1个句子,直至搜索至该对应章节的结束位置“。”,终端都没有搜索到与该第1个句子,于是,终端对该第1个句子的匹配失败;终端不更新搜索起始位置,继续以该对应章节的起始位置“话”为搜索起始位置,以该第2个句子的长度为搜索长度,在该对应章节中搜索该第2个句子,直至搜索至该对应章节的结束位置“。”,终端都没有搜索到与该第2个句子,于是,终端对该第2个句子的匹配也失败;终端不更新搜索起始位置,继续以该对应

章节的起始位置“话”为搜索起始位置,以第3个句子的长度为搜索长度,在该对应章节中搜索该第3个句子,则终端可以搜到该第3个句子,从而终端对该第3个句子的匹配成功,终端将搜索起始位置更新为该第3个句子的结束位置“,”,进而终端从该该第3个句子的结束位置“,”开始,在该章节中搜索第4个句子,并实现对该第4个句子的成功匹配,与此同理,终端对第5个和第6个句子都匹配成功,当终端对该第6个句子匹配成功时,终端将搜索起始位置更新为该第6个句子的结束位置“,”,并从该第6个句子的结束位置“,”开始,以第7个句子的长度为搜索长度,在该对应章节中,搜索该第7个句子,直至搜索至该对应章节的结束位置“。”,终端都没有搜索到与该第7个句子,于是,终端对该第7个句子的匹配也失败。此时,终端对该原文字符串的句子匹配结束。

[0141] 在步骤5a05中,若该n个句子中匹配成功的最先一个句子不是第1个句子,则将该第1个句子至该匹配成功的最先一个句子之前的所有句子进行分词,得到按序排列的m个分词,m为正整数。

[0142] 当终端在对原文字符串的句子进行匹配时,如果该原文字符串的第一个句子没有匹配成功,则终端对该原文字符串中最先一个匹配成功的句子与该原文字符串的第一个句子之间的所有句子进行分词,得到m个分词,该m为正整数。

[0143] 继续以上述步骤5a03中得到的7个句子为例以及上述步骤5a04中对各个分句的匹配结果为例进行说明。在上述步骤5a04中,终端在对各个句子进行匹配时,匹配成功的最先一个句子为原文字符串中的第3句,即,终端最先匹配到的一个句子为:“及秦灭之后,”,也即是,终端没有匹配到该原文字符串的第1句和第2句,于是终端对该最先一个匹配成功的句子“及秦灭之后,”和该原文字符串的第一个句子“周末七国纷争,”之间的所有句子进行分词,即,终端对该原文字符串的第一个和第二个句子进行分词。

[0144] 该原文字符串的第一个句子和第二个句子分别为:(1)“周末七国纷争,”,(2)“并入于秦;”,终端对该两个句子进行分词,可以得到“周末”、“七国”、“纷争”和“并入”、“于”、“秦”按序排列的6个分词。

[0145] 在步骤5a06中,从该匹配成功的最先一个句子的匹配内容的起始位置开始,逆序对该m个分词进行匹配,并将匹配成功的最后一个分词的匹配内容的起始位置作为该原文字符串在修订后的该电子出版物中所对应的起始位置。

[0146] 终端从该匹配成功的最先一个句子的匹配内容的起始位置开始,逆序对该m个分词进行匹配,即终端首先对最后一个分词进行匹配,接下来对倒数第二个分词进行,依次类推,终端最后对第一个分词进行匹配。

[0147] 需要说明的是,终端在对该m个分词进行匹配时,是在该章节中最先一个匹配成功的句子之前的内容中,对该m个分词进行匹配的。

[0148] 还需要说明的是,在对该m个句子全部匹配后,若存在与该m个句子完全匹配的匹配内容或存在与该m个句子部分匹配但匹配程度属于第一预定匹配范围的匹配内容时,确定分词匹配成功,终端将最后一个匹配成功的分词的位置,作为原文字符串在修订后的该电子出版物中所对应的起始位置。当匹配程度不属于第一预定匹配范围的匹配内容时,确定分词匹配失败,终端将第一个匹配成功的句子的位置,作为原文字符串在修订后的电子出版物中所对应的起始位置。其中,该第一预定匹配范围可以为: $m_1 \geq 50\% m$ , $m$ 与步骤5a05中的m相同,即m表示对最先一个匹配成功的句子与该原文字符串的第一个句子之间的所有

句子进行分词,得到的分词个数,m1表示匹配到的分词的个数。

[0149] 以上述步骤5a05中得到的6个分词以及图5b所示的修订后的对应章节为例进行说明。在进行分句匹配时,终端匹配到的最先一个句子为“及秦灭之后”,终端将该“及秦灭之后”的起始位置“及”作为搜索起始位置,在该章节中最先一个匹配成功的句子之前的内容中,逆序对该6个分词进行匹配。

[0150] 终端将“及”作为搜索起始位置,在该对应章节中的该“及”之前的内容中逆向对该最后一个分词“秦”,进行搜索,直至搜索到位置“话”,终端都没有搜索到该分词“秦”,于是终端对该分词“秦”的匹配失败,终端不更新搜索起始位置,继续将该“及”作为搜索起始位置,在该对应章节中的该“及”之前的内容中逆向,对倒数第二个分词“于”进行搜索,终端发现对该分词“于”的匹配也失败,进而终端继续从该“及”的位置开始,逆向对分词“并入”进行匹配,其匹配结果也是失败,终端不更新搜索起始位置,从该位置“及”开始逆向对分词“纷争”进行搜索,终端可以在当前搜索位置在搜索到该“纷争”,终端对该“纷争”的匹配成功,进而终端将搜索起始位置更新为“纷”,从该“纷”开始逆向对分词“七国”进行匹配,匹配成功,终端将搜索起始位置更新为“七”,从该“七”开始逆向对分词“周末”进行匹配,匹配失败,匹配结束。

[0151] 以上在对该6个分词进行匹配时,有部分分词没有匹配成功,因此属于部分匹配,终端计算该匹配程度与第一预定匹配范围之间的关系,终端可以计算得参与匹配的分词有6个,匹配成功的分词有2个,参与匹配的分词的50%为3,且 $2 < 3$ ,因此,分词匹配失败,终端不更新原文字符串在修订后的该电子出版物中所对应的起始位置,即,终端将“及”作为原文字符串在修订后的该电子出版物中所对应的起始位置。

[0152] 需要说明的是,若终端计算得匹配成功的分词的个数大于或等于3,则分词匹配成功,此时,终端需要将原文字符串在修订后的该电子出版物中的起始位置更新为最后一个匹配成功的分词的位置。

[0153] 在步骤5a07中,若该n个句子中匹配成功的最后一个句子不是第n个句子,则将该匹配成功的最后一个句子之后至该第n个句子之间的所有句子进行分词,得到按序排列的q个分词,q为正整数。

[0154] 当终端在对原文字符串的句子进行匹配时,如果该原文字符串的最后一个匹配成功的句子不是第n个句子,则终端将该匹配成功的最后一个句子之后至该第n个句子之间的所有句子进行分词,得到按序排列的q个分词,q为正整数。

[0155] 继续以上述步骤5a03中得到的7个句子为例以及上述步骤5a04中对各个分句的匹配结果为例进行说明。在上述步骤5a04中,终端在对各个句子进行匹配时,匹配成功的最后一个句子为原文字符串中的第6句,即,终端最后匹配到的一个句子为:“汉朝自高祖斩白蛇而起义,”,也即是,终端没有匹配到该原文字符串的第7句,于是终端对该最后一个匹配成功的句子“及秦灭之后,”和该原文字符串的最后一个句子“一统天下,”之间的所有句子进行分词,即,终端对该原文字符串的第7个句子进行分词。

[0156] 该原文字符串的第7个句子为:(7)“一统天下,”,终端对该句子进行分词,可以得到“一”、“统”、和“天下”按序排列的3个分词。

[0157] 在步骤5a08中,从该匹配成功的最后一个句子的匹配内容的结束位置开始,顺序对该q个分词进行匹配,并将匹配成功的最后一个分词的匹配内容的结束位置作为该原文

字符串在修订后的该电子出版物中所对应的结束位置。

[0158] 终端从该匹配成功的最后一个句子的匹配内容的结束位置开始,顺序对该q个分词进行匹配。

[0159] 需要说明的是,终端在对该q个分词进行匹配时,是在该章节中最后一个匹配成功的句子之后的内容中,对该q个分词进行匹配的。

[0160] 还需要说明的是,在对该q个句子全部匹配后,若存在与该q个句子完全匹配的匹配内容或存在与该q个句子部分匹配但匹配程度属于第一预定匹配范围的匹配内容时,确定分词匹配成功,终端将匹配成功的最后一个的分词的位置,作为原文字符串在修订后的该电子出版物中所对应的结束位置,当匹配程度不属于第一预定匹配范围的匹配内容时,确定分词匹配失败,终端将最后一个匹配成功的句子的位置,作为原文字符串在修订后的电子出版物中所对应的结束位置。其中,该第一预定匹配范围可以为: $q_1 \geq 50\% q$ ,其中,q与步骤5a07中的q相同,即q表示对匹配成功的最后一个句子之后至第n个句子之间的所有句子进行分词,得到的分词的个数,q1表示匹配到的分词的个数。

[0161] 以上述步骤5a07中得到的3个分词以及图5b所示的修订后的对应章节为例进行说明。在进行分句匹配时,终端匹配到的最后一个句子为“汉朝自高祖斩白蛇而起义,”,终端将该“汉朝自高祖斩白蛇而起义,的结束位置”,作为搜索起始位置,在该对应章节中最后一个匹配成功的句子之后的内容中,顺序对该3个分词进行匹配。

[0162] 终端将“,”作为搜索起始位置,在该对应章节中的该“,”之后的内容中对该第一个分词“一”,进行搜索,直至搜索到位置“。”,终端都没有搜索到该分词“一”,于是终端对该分词“一”的匹配失败,终端不更新搜索起始位置,继续将该“,”作为搜索起始位置,在该对应章节中的该“,”之后的内容中顺序对第二个分词“统”进行搜索,终端可以在当前搜索位置在搜索到该“统”,终端对该分词“统”的匹配成功,进而终端将搜索起始位置更新为“统”,从该“统”开始顺序对分词“天下”进行匹配,匹配成功,匹配结束。

[0163] 以上在对该3个分词进行匹配时,有部分分词没有匹配成功,因此属于部分匹配,终端计算该匹配程度与第一预定匹配范围,终端可以计算得参与匹配的分词有3个,匹配成功的分词有2个,参与匹配的分词的50%为1.5,且 $2 > 1.5$ ,因此,分词匹配成功,终端将该“天下”的结束位置作为原文字符串在修订后的电子出版物中所对应的结束位置。

[0164] 需要说明的是,若终端计算得匹配成功的分词的个数小于1.5,则分词匹配失败,此时,终端不更新原文字符串在修订后的该电子出版物中所对应的结束位置,即,终端将“,”作为原文字符串在修订后的该电子出版物中所对应的结束位置。

[0165] 在步骤5a09中,计算匹配程度与预定匹配范围之间的关系,当匹配程度属于预定匹配范围时,确定匹配成功。

[0166] 在上述步骤5a01-5a08结束后,终端计算匹配到的程度与预定匹配范围之间的对应关系,当匹配程度属于预定匹配范围时,确定对原文字符串的匹配成功。其中,该预定匹配范围可以用公式: $80\% Len_1 \leq Len_2 \leq 120\% Len_1$ 来表示,其中,Len1为原文字符串的文本长度,Len2为匹配到的字符串的长度。

[0167] 以上述步骤5a01-5a08的原文字符串以及对该原文字符串的匹配结果为例进行说明。该原文字符串为:“周末七国纷争,并入于秦;及秦灭之后,楚、汉分争,又并入于汉;汉朝自高祖斩白蛇而起义,一统天下,”,匹配到的字符串为:“及秦灭之后,楚、汉分争,又并入于

汉;汉朝自高祖斩白蛇而起义,统天下,”,终端可以计算得原文字符串的文本长度为47,匹配到的字符串的文本长度为34,由于原文字符串的文本长度80%为37.5,原文字符串的文本长度120%为56.4,且 $34 < 37.5$ ,则终端对该原文字符串的匹配失败。

[0168] 需要说明的是,在上述举例过程中,如果终端仅匹配到字符串的文本长度长度在37.5和56.4之间,则此时匹配成功,该匹配的具体过程与上述例子同理,这里不再赘述。

[0169] 在步骤5a10中,在匹配成功时,根据匹配结果更新该用户数据中的引用内容,得到更新后的该用户数据。

[0170] 该步骤5a10与上述图3a所示实施例中的步骤3a03相同或相似,本实施例对此不再赘述。

[0171] 本公开实施例提供的用户数据更新方法,通过获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串;在修订后的该电子出版物中对该原文字符串直接进行匹配;在直接匹配不成功时,对该原文字符串进行分句,得到按序排列的n个句子,在修订后的该电子出版物中按序对该n个句子进行匹配,以及当匹配成功的最先一个句子不是第1个句子,和/或匹配成功的最后一个句子不是该原文字符串的最后一个句子时,对该原文字符串的匹配成功的最先一个句子和匹配成功的第一个句子之间的句子,以及最后一个匹配成功的句子和该原文字符串的最后一个句子之间的句子进行分词处理,并在修订后的对应章节中,对该各个分词进行匹配,根据匹配结果对该原文字符串对应的用户数据进行更新,节省了生成Patch文件的计算代价,增加了用户数据更新方法的使用范围较广,进一步地,通过限定分词的匹配程度与第一预定匹配范围的关系,提高了更新到的用户数据的正确性以及保证了该用户数据更新方法的完善性。

[0172] 需要说明的是,以上图2-图5a所示的实施例中,该对用户数据的引用内容进行更新的执行主体是终端,而事实上,在本公开提供的其他实施例中,对该用户数据的引用内容进行更新的执行主体还可以是服务器,终端可以从服务器端下载该更新后的用户数据的引用内容。

[0173] 图6是根据一示例性实施例示出的一种用户数据更新装置的框图。该用户数据更新装置可以通过软件、硬件或者两者的结合实现成为DLNA服务器的部分或者全部。该用户数据更新装置可以包括:获取模块601、匹配模块602和更新模块603。

[0174] 获取模块601,被配置为获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串;该获取模块601与匹配模块602相连,匹配模块602,被配置为在修订后的该电子出版物中对该原文字符串进行匹配;该匹配模块602与更新模块603相连,更新模块603被配置为在匹配成功时,根据匹配结果更新该用户数据中的引用内容,得到更新后的该用户数据。

[0175] 一个实施例,该匹配模块603,被配置为在修订后的该电子出版物中对该原文字符串直接进行匹配;

[0176] 或,

[0177] 该匹配模块603,被配置为对该原文字符串进行分句,得到按序排列的n个句子,n为正整数;在修订后的该电子出版物中按序对该n个句子进行匹配;

[0178] 或,

[0179] 该匹配模块603,被配置为在修订后的该电子出版物中对该原文字符串直接进行

匹配；在直接匹配不成功时，对该原文字符串进行分句，得到按序排列的n个句子，n为正整数；在修订后的该电子出版物中按序对该n个句子进行匹配。

[0180] 一个实施例，该匹配模块603，包括：

[0181] 匹配单元，被配置为从搜索起始位置对第i个句子进行匹配， $1 \leq i \leq n$ ；当*i*=1时，该搜索起始位置为该原文字符串所在章节的起始位置；

[0182] 位置更新单元，被配置为在该第*i*个句子匹配成功时，将该搜索起始位置更新为该第*i*个句子的匹配内容的结束位置，并从更新后的该搜索起始位置对第*i*+1个句子进行匹配， $i+1 \leq n$ ；

[0183] 继续匹配单元，被配置为在该第*i*个句子匹配不成功时，从该第*i*个句子所对应的搜索位置对第*i*+1个句子进行匹配；

[0184] 确定单元，被配置为在该n个句子全部匹配后，若存在与该n个句子完全匹配的匹配内容或存在与该n个句子部分匹配但匹配程度属于预定匹配范围的匹配内容时，确定分句匹配成功。

[0185] 一个实施例，该装置还包括：

[0186] 第一分词单元，被配置为在该n个句子中匹配成功的最先一个句子不是第1个句子时，将该第1个句子至该匹配成功的最先一个句子之前的所有句子进行分词，得到按序排列的m个分词，m为正整数；

[0187] 分词匹配单元，被配置为从该匹配成功的最先一个句子的匹配内容的起始位置开始，逆序对该m个分词进行匹配；

[0188] 起始确定单元，被配置为将匹配成功的最后一个分词的匹配内容的起始位置作为该原文字符串在修订后的该电子出版物中所对应的起始位置。

[0189] 一个实施例，

[0190] 第二分词单元，被配置为在该n个句子中匹配成功的最后一个句子不是第n个句子，则将该匹配成功的最后一个句子之后至该第n个句子之间的所有句子进行分词，得到按序排列的q个分词，q为正整数；

[0191] 分词匹配单元，被配置为从该匹配成功的最后一个句子的匹配内容的结束位置开始，顺序对该q个分词进行匹配；

[0192] 结束确定单元，被配置为将匹配成功的最后一个分词的匹配内容的结束位置作为该原文字符串在修订后的该电子出版物中所对应的结束位置。

[0193] 本公开实施例提供的用户数据更新装置，通过获取用户数据中的引用内容在修订前的电子出版物中所对应的原文字符串；在修订后的该电子出版物中对该原文字符串进行匹配；在匹配成功时，根据匹配结果更新该用户数据中的引用内容，得到更新后的该用户数据。采用本公开实施例提供的用户数据更新方法，节省了生成Patch文件的计算代价，且本公开实施例提供的用户数据更新方法无需单独维护，使用范围较广。

[0194] 关于上述实施例中的装置，其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述，此处将不做详细阐述说明。

[0195] 图7是根据一示例性实施例示出的一种用于用户数据更新的装置700的框图。例如，装置700可以是移动电话，计算机，数字广播终端，消息收发设备，游戏控制台，平板设备，医疗设备，健身设备，个人数字助理等。

[0196] 参照图7,装置700可以包括以下一个或多个组件:处理组件702,存储器704,电源组件706,多媒体组件708,音频组件710,输入/输出(I/O)接口712,传感器组件714,以及通信组件716。

[0197] 处理组件702通常控制装置700的整体操作,诸如与显示,电话呼叫,数据通信,相机操作和记录操作相关联的操作。处理组件702可以包括一个或多个处理器720来执行指令,以完成上述的方法的全部或部分步骤。此外,处理组件702可以包括一个或多个模块,便于处理组件702和其他组件之间的交互。例如,处理组件702可以包括多媒体模块,以方便多媒体组件708和处理组件702之间的交互。

[0198] 存储器704被配置为存储各种类型的数据以支持在装置700的操作。这些数据的示例包括用于在装置700上操作的任何应用程序或方法的指令,联系人数据,电话簿数据,消息,图片,视频等。存储器704可以由任何类型的易失性或非易失性存储设备或者它们的组合实现,如静态随机存取存储器(SRAM),电可擦除可编程只读存储器(EEPROM),可擦除可编程只读存储器(EPROM),可编程只读存储器(PROM),只读存储器(ROM),磁存储器,快闪存储器,磁盘或光盘。

[0199] 电源组件706为装置700的各种组件提供电力。电源组件706可以包括电源管理系统,一个或多个电源,及其他与为装置700生成、管理和分配电力相关联的组件。

[0200] 多媒体组件708包括在装置700和用户之间的提供一个输出接口的屏幕。在一些实施例中,屏幕可以包括液晶显示器(LCD)和触摸面板(TP)。如果屏幕包括触摸面板,屏幕可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中,多媒体组件708包括一个前置摄像头和/或后置摄像头。当装置700处于操作模式,如拍摄模式或视频模式时,前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[0201] 音频组件710被配置为输出和/或输入音频信号。例如,音频组件710包括一个麦克风(MIC),当装置700处于操作模式,如呼叫模式、记录模式和语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器704或经由通信组件716发送。在一些实施例中,音频组件710还包括一个扬声器,用于输出音频信号。

[0202] I/O接口712为处理组件702和外围接口模块之间提供接口,上述外围接口模块可以是键盘,点击轮,按钮等。这些按钮可包括但不限于:主页按钮、音量按钮、启动按钮和锁定按钮。

[0203] 传感器组件714包括一个或多个传感器,用于为装置700提供各个方面状态评估。例如,传感器组件714可以检测到装置700的打开/关闭状态,组件的相对定位,例如所述组件为装置700的显示器和小键盘,传感器组件714还可以检测装置700或装置700一个组件的位置改变,用户与装置700接触的存在或不存在,装置700方位或加速/减速和装置700的温度变化。传感器组件714可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件714还可以包括光传感器,如CMOS或CCD图像传感器,用于在成像应用中使用。在一些实施例中,该传感器组件714还可以包括加速度传感器,陀螺仪传感器,磁传感器,压力传感器或温度传感器。

[0204] 通信组件716被配置为便于装置700和其他设备之间有线或无线方式的通信。装置700可以接入基于通信标准的无线网络,如WiFi,2G或3G,或它们的组合。在一个示例性实施例中,通信组件716经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中,通信组件716还包括近场通信(NFC)模块,以促进短程通信。例如,在NFC模块可基于射频识别(RFID)技术,红外数据协会(IrDA)技术,超宽带(UWB)技术,蓝牙(BT)技术和其他技术来实现。

[0205] 在示例性实施例中,装置700可以被一个或多个应用专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑器件(PLD)、现场可编程门阵列(FPGA)、控制器、微控制器、微处理器或其他电子元件实现,用于执行上述方法。

[0206] 在示例性实施例中,还提供了一种包括指令的非临时性计算机可读存储介质,例如包括指令的存储器704,上述指令可由装置700的处理器720执行以完成上述方法。例如,所述非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0207] 一种非临时性计算机可读存储介质,当所述存储介质中的指令由装置700的处理器执行时,使得装置700能够执行一种上述用户数据更新方法。

[0208] 图8是根据另一示例性实施例示出的一种用于用户数据更新的装置800的框图。例如,装置800可以被提供为一服务器。参照图8,装置800包括处理组件822,其进一步包括一个或多个处理器,以及由存储器832所代表的存储器资源,用于存储可由处理部件822执行的指令,例如应用程序。存储器832中存储的应用程序可以包括一个或一个以上的每一个对应于一组指令的模块。此外,处理组件822被配置为执行指令,以执行上述用户数据更新方法。

[0209] 装置800还可以包括一个电源组件826被配置为执行装置800的电源管理,一个有线或无线网络接口850被配置为将装置800连接到网络,和一个输入输出(I/O)接口858。装置800可以操作基于存储在存储器832的操作系统,例如Windows Server™,Mac OS X™,Unix™,Linux™,FreeBSD™或类似。

[0210] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本公开的其它实施方案。本申请旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的真正范围和精神由下面的权利要求指出。

[0211] 应当理解的是,本公开并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本公开的范围仅由所附的权利要求来限制。

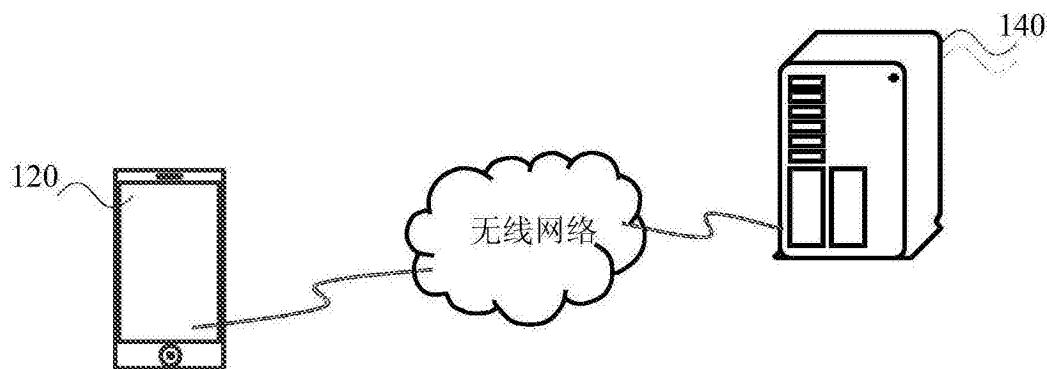


图1

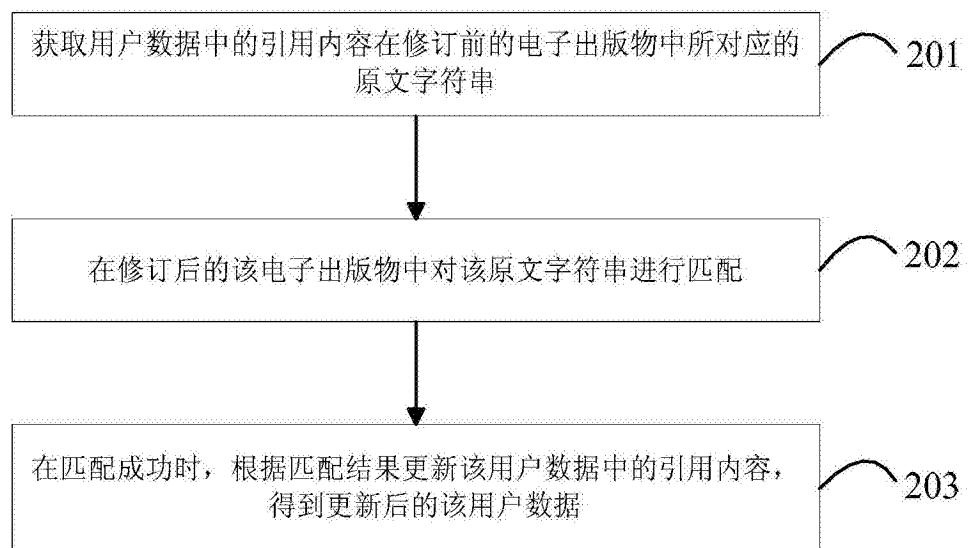


图2

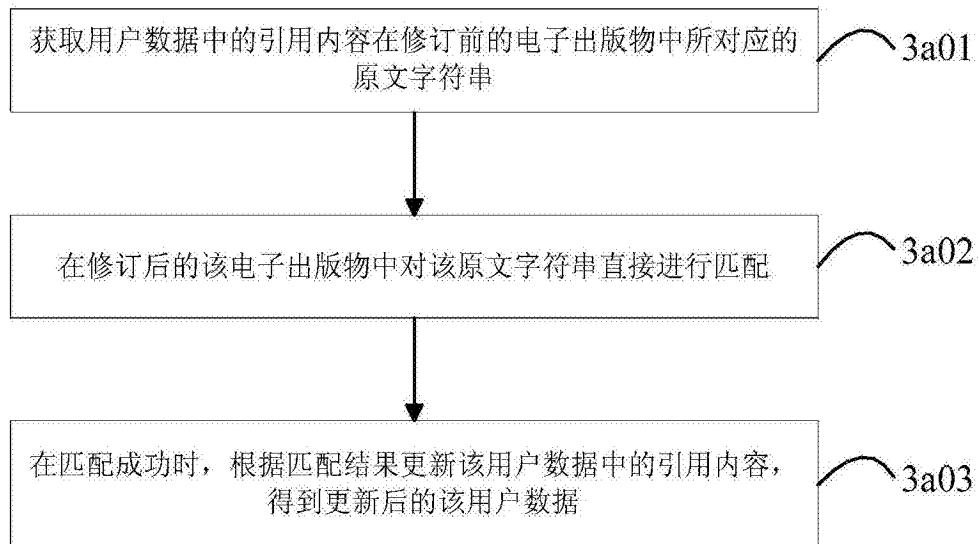


图3a

<p>话说天下大势， 分久必合，合久必分； <u>周末七国纷争，并入于秦；及秦灭之后，楚、汉分争，又并入于汉；汉朝自高祖斩白蛇而起义，一统天下，后来光武中兴，传至献帝，遂分为三国。</u></p>	<p>话说天下大势， 分久必合，合久必分； <u>周末七国纷争，并入于秦；及秦灭之后，楚、汉分争，又并入于汉；汉朝自高祖斩白蛇而起义，一统天下，后来光武中兴，传至献帝，遂分为三国。</u></p>
--	--

修订前的某一章节

修订后的对应章节

图3b

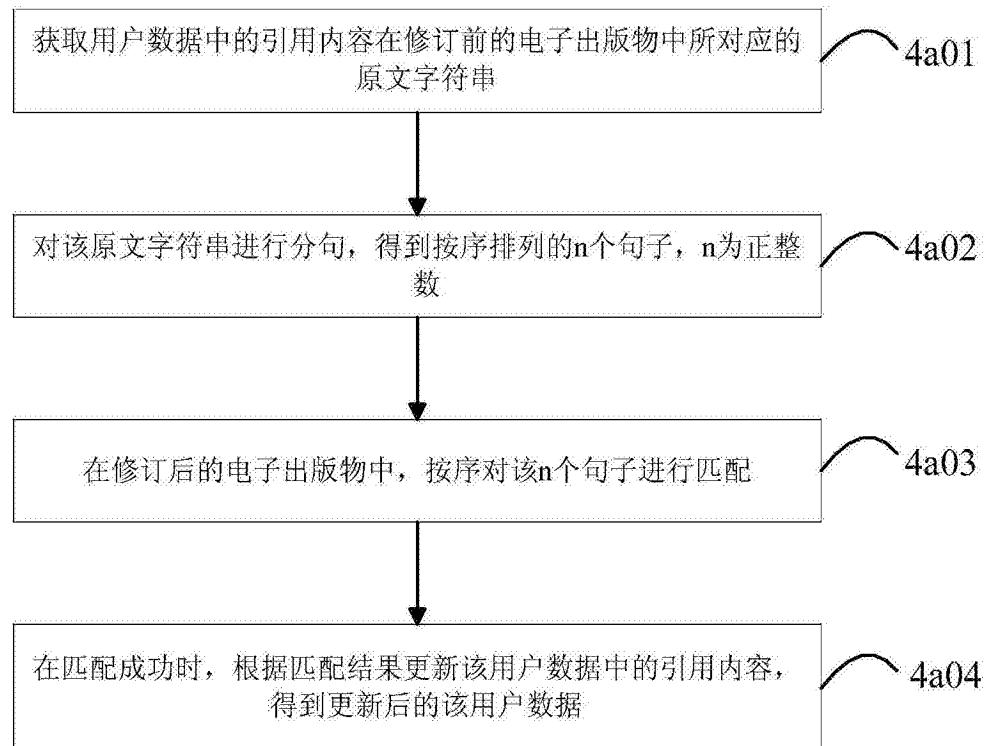


图4a

<p>话说天下大势， 分久必合，合久必分； <u>周末七国纷争，并入于秦；及秦灭之后，</u> <u>楚、汉分争，又并入于汉；汉朝自高祖斩白蛇而起义，一统天下，</u> <u>后来光武中兴，传至献帝，遂分为三国。</u></p>	<p>话说天下大势， 分久必合，合久必分； <u>周末七国纷争，并入于秦；及秦灭之后，</u> <u>又并入于汉；汉朝自高祖斩白蛇而起义，一统天下，</u> <u>后来光武中兴，传至献帝，遂分为三国。</u></p>
--	--

修订前的某一章节

修订后的对应章节

图4b

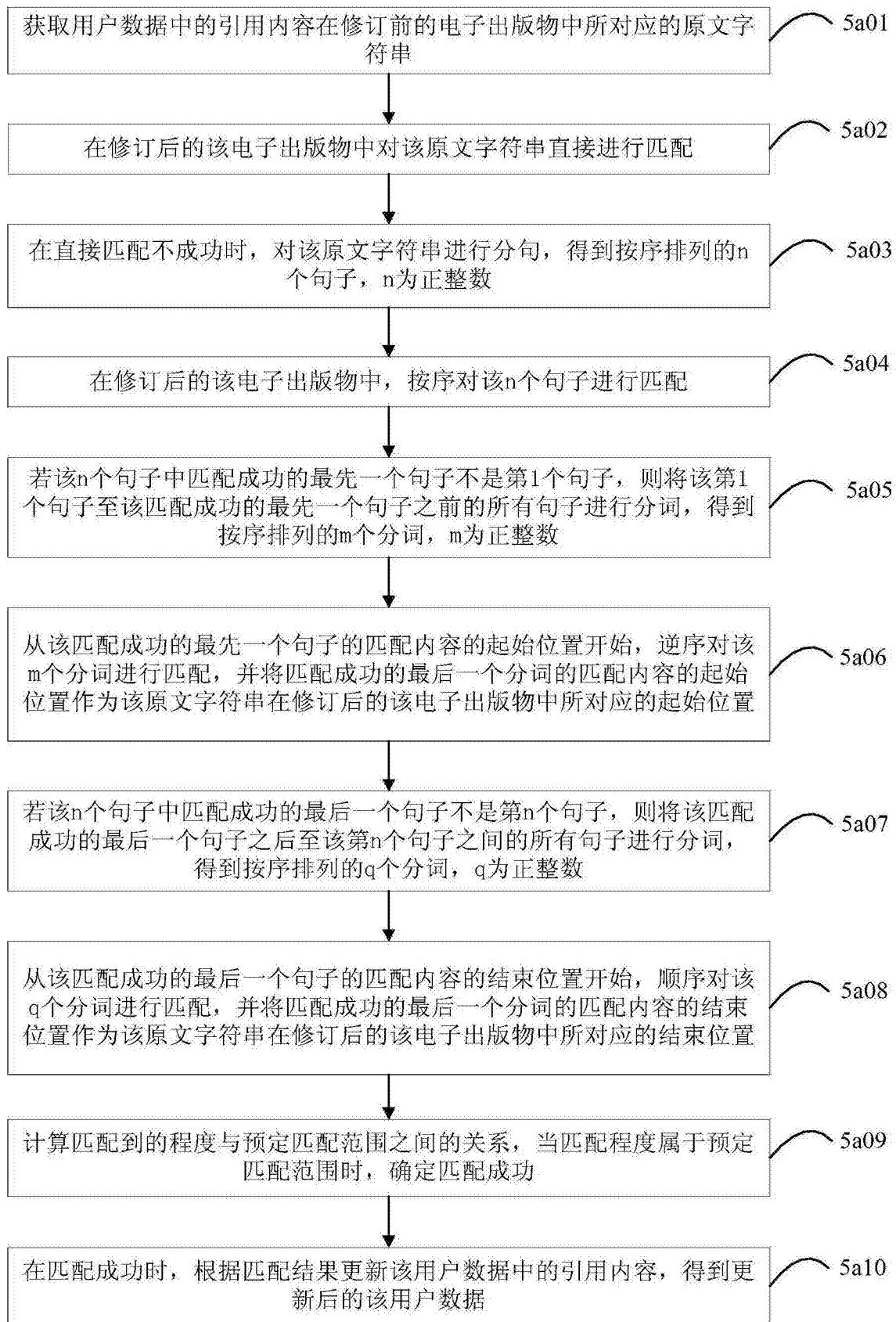
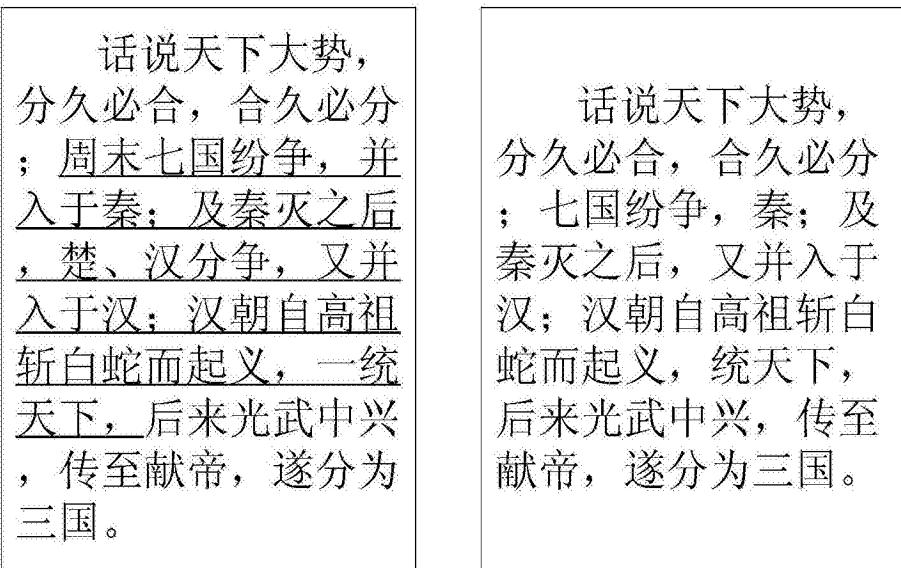


图5a



修订前的某一章节

修订后的对应章节

图5b

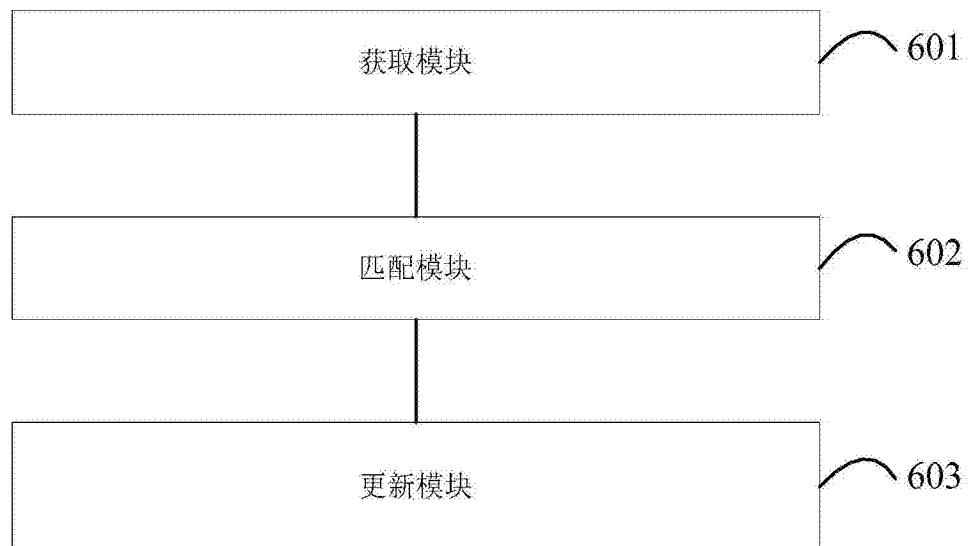


图6

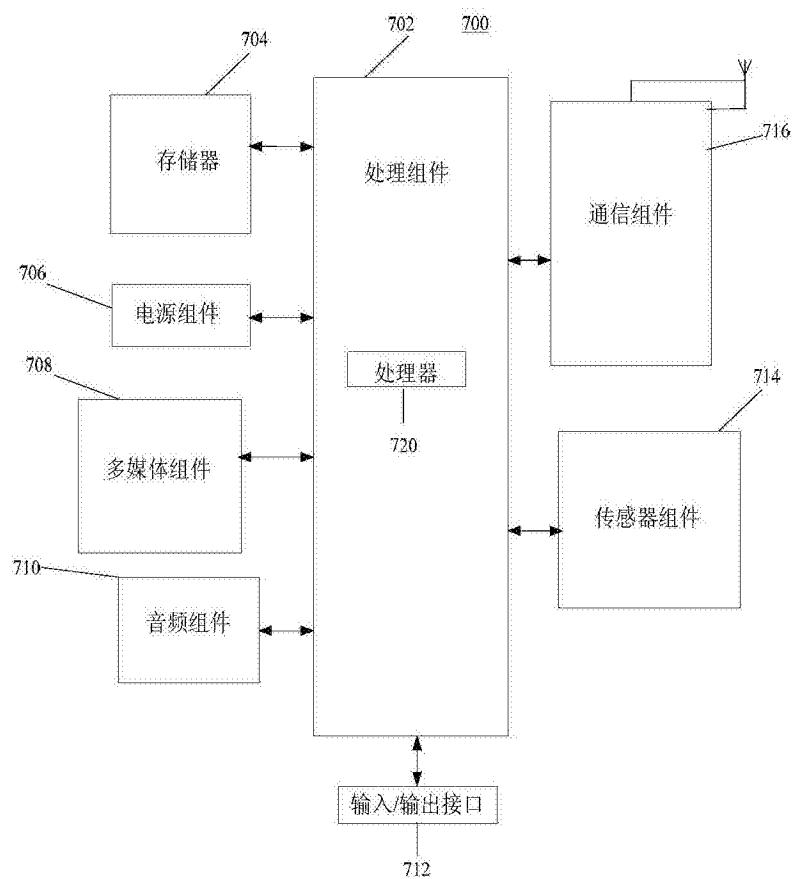


图7

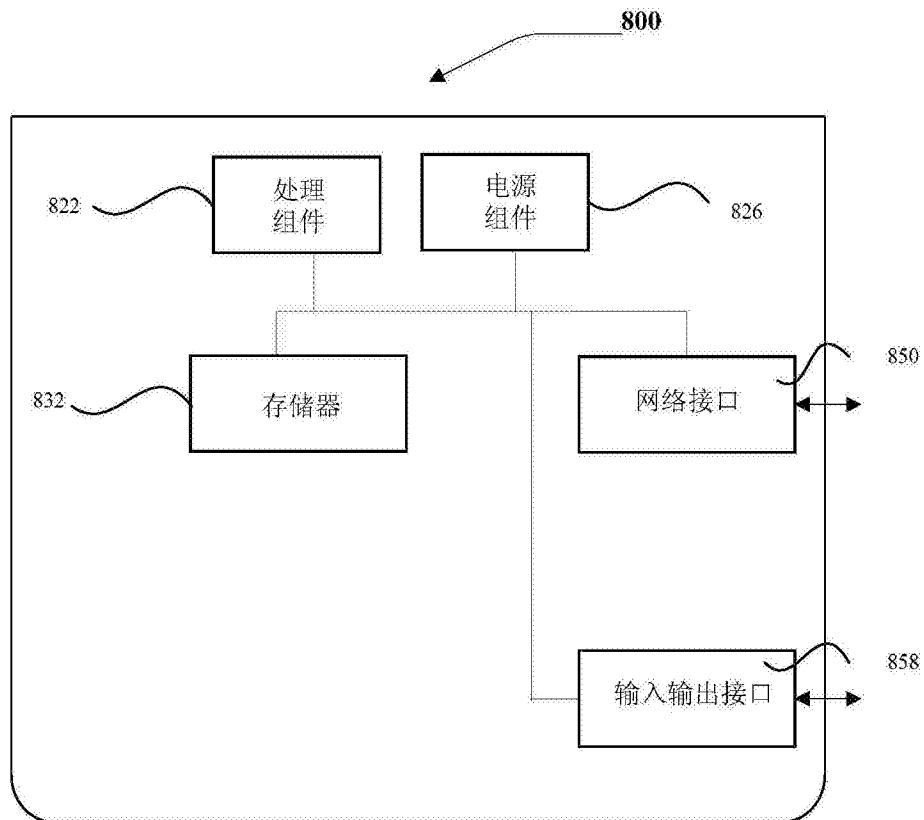


图8