



(12) 发明专利申请

(10) 申请公布号 CN 102081600 A

(43) 申请公布日 2011.06.01

(21) 申请号 201110026157.8

(22) 申请日 2011.01.25

(71) 申请人 珠海全志科技有限公司

地址 519080 广东省珠海市唐家软件园路 1
号生产加工中心 4 号楼 4 层 1 单元

(72) 发明人 吴浪

(74) 专利代理机构 珠海智专专利商标代理有限
公司 44262

代理人 李谨 张中

(51) Int. Cl.

G06F 17/25 (2006.01)

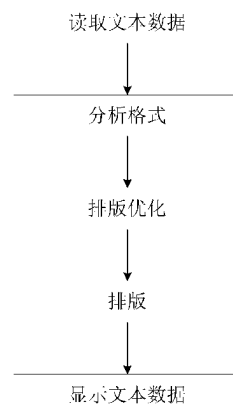
权利要求书 2 页 说明书 5 页 附图 7 页

(54) 发明名称

电子书排版方法及其系统

(57) 摘要

本发明提供一种用于在不同屏幕大小的电子书阅读设备上对电子书进行排版的电子书排版方法,包括:读取电子书中的文本数据;然后分析文本数据的格式;再结合电子书阅读设备的环境参数对文本数据格式进行包括:换行符分析、前后文语境检测及长段落分析的排版优化处理;接着根据排版优化处理的格式对文本数据进行排版;最后显示排版后的电子书文本数据;本发明还提供了一种电子书排版系统;采用本发明,可实现优化的排版功能,用户在不同屏幕大小或参数的电子书阅读设备上均可获得良好的阅读效果。



1. 电子书排版方法,用于在不同屏幕大小的电子书阅读设备上对电子书进行排版,其特征在于:

该方法包括:

读取电子书中的文本数据;

分析所述文本数据的格式;

结合所述电子书阅读设备的环境参数对所述文本数据格式进行包括:换行符分析、前后文语境检测及长段落分析的排版优化处理;

根据所述排版优化处理的格式对所述文本数据进行排版;

显示排版后的电子书文本数据。

2. 根据权利要求 1 所述的电子书排版方法,其特征在于:

所述换行符分析的方法是对当前字符的前一字符和后一字符进行检测,以判断当前换行字符是否有效的方法,具体步骤为:

sa1:读取当前字符;

sa2:判断所述当前字符是否为换行符,若是则进入下一步骤 sa3,否则进入步骤 sa8;

sa3:读取所述当前字符的前一字符;

sa4:判断所述前一字符是否属于:问号、省略号、后书名号、后引号、感叹号等句结束符中的其中之一,若是则初步判断当前换行符有效并进入下一步骤 sa5,否则进入步骤 sa8;

sa5:读取所述当前字符的后一字符;

sa6:判断所述后一字符是否属于项目符号或其他开始符号的其中之一,若是则再次判断当前换行符有效并进入下一步骤 sa7,否则进入步骤 sa9;

sa7:前后文语境分析过程,进入步骤 sa9;

sa8:判定换行符无效,进入长段落分析过程;

sa9:结束。

3. 根据权利要求 1 或 2 所述的电子书排版方法,其特征在于:

所述前后文语境检测的方法是根据所述文本数据的语境类别,通过分析所述文本数据中的:开始标点、首字符、中间字符、尾标点及尾字符来判断句子是整句还是半句,从而判断所述当前换行字符是否有效的方法。

4. 根据权利要求 1 或 2 所述的电子书排版方法,其特征在于:

所述长段落分析的具体过程为:

sb1:获取屏幕的大小 w 和设置显示的字号 f ;

sb2:计算每个字所占的像素面积: $x = (\text{字宽} + \text{字间距}) * (\text{字高} + \text{行间距})$;

sb3:计算分段的字符数: $m = (w * 1/3) / x$;

sb4:根据分段的字符数对最近的句子进行查找,当最近句子中的字符个数大于等于分段的字符数时,进入步骤 sb5;

sb5:进行整句句查找;

sb6:在句子结束处插入回车换行符;

sb7:结束。

5. 根据权利要求 3 所述的电子书排版方法,其特征在于:

所述长段落分析的具体过程为：

sb1 :获取屏幕的大小 w 和设置显示的字号 f ；

sb2 :计算每个字所占的像素面积： $x = (\text{字宽} + \text{字间距}) * (\text{字高} + \text{行间距})$ ；

sb3 :计算分段的字符数： $m = (w * 1/3) / x$ ；

sb4 :根据分段的字符数对最近的句子进行查找,当最近句子中的字符个数大于等于分段的字符数时,进入步骤 sb5 ；

sb5 :进行整句句查找 ；

sb6 :在句子结束处插入回车换行符 ；

sb7 :结束。

6. 根据权利要求 1 或 2 所述的电子书排版方法,其特征在于：

所述分析文本数据格式具体包括：

对所述文本数据进行文本符号分析和段落分析。

7. 电子书排版系统,用于在不同屏幕大小的电子书阅读设备上对电子书进行排版,其特征在于：

该系统包括顺序连接的读取模块、格式分析模块、排版模块和显示模块；

读取模块用于读取电子书的文本数据；

格式分析模块分析所述文本数据的格式；

排版模块结合所述电子书阅读设备的环境参数对所述文本数据格式进行排版优化处理并排版；所述排版模块包括对所述文本数据的换行符分析部分、前后文语境检测部分及长段落分析部分；

显示模块用于显示排版后的电子书文本数据。

8. 根据权利要求 7 所述的电子书排版系统,其特征在于：

所述格式分析模块包括对所述文本数据的文本符号分析部分和段落分析部分。

9. 根据权利要求 7 或 8 所述的电子书排版系统,其特征在于：

所述换行符分析部分对当前字符的前一字符和后一字符进行检测,以判断当前换行字符是否有效；

所述前后文语境检测部分是根据所述文本数据的语境类别,分析所述文本数据中的：开始标点、首字符、中间字符、尾标点及尾字符来判断句子是整句还是半句,从而判断所述当前换行字符是否有效；

所述长段落分析部分是根据屏幕大小分析分段字符数,并结合文字数据类型来判断是否插入回车符。

电子书排版方法及其系统

技术领域

[0001] 本发明涉及一种电子书排版方法及其系统,更具体的是,涉及一种能适应不同屏幕大小电子书阅读设备的电子书排版方法及其系统。

背景技术

[0002] 一般来说,电子书是指电子版图书,又称 e-book,是传统印刷书籍的电子版本,用户可以使用个人计算机或电子书阅读器等下载电子书进行离线阅读。由于电子书可以从网上自动下载,用户可用各种电子书设备,如:计算机、电子书阅读器或手机进行海量阅读,还可通过查找、搜索、画线、注释和超链接等实现类似纸质书籍上的功能,随着网络的发展,电子书的应用也日益广泛。由于携带的便利性,手机和电子书阅读器常常作为电子书的阅读设备,然而,不同的电子书阅读器或手机具有各自不同的屏幕大小或文本格式,大部分电子书阅读设备都没有自动排版功能,不能根据屏幕的不同大小(自动)调整文字的排版格式,而且,对于用户从网络下载的电子书,如:文本电子书、手机电子书等,由于其格式是针对特定的电子阅读设备的,采用其他阅读设备下载后,往往不能正确识别或显示其中的段落、换行和空白行等文本符号,因此,排版功能很差,大大影响了使用者的阅读效果。

发明内容

[0003] 本发明的主要目的是提供一种能适用于不同屏幕大小的电子阅读设备、便于使用者阅读的电子书排版方法。

[0004] 为实现上述目的,本发明提供的电子书排版方法用于在不同屏幕大小的电子书阅读设备上对电子书进行排版,包括:

读取电子书中的文本数据;

分析文本数据的格式;

结合所述电子书阅读设备的环境参数对所述文本数据格式进行包括:换行符分析、前后文语境检测及长段落分析的排版优化处理;

根据排版优化处理的格式对文本数据进行排版;

显示排版后的电子书文本数据。

[0005] 由以上方案可见,本发明结合了电子书阅读设备的环境参数对读取的电子书文本数据格式进行排版,删除其中多余的换行和空白部分,分析长段落并结合电子书阅读设备的环境参数进行段落分割,使得读者在不同屏幕大小的电子书阅读设备上均能具有良好的阅读效果,避免了传统方法中使用不同电子书阅读设备而出现的多余回车换行符号、多余的空白行及没有回车换行长篇文本的情况,大大优化的排版功能。

[0006] 本发明的另一目的是提供一种采用上述方法对电子书阅读设备排版的系统。

[0007] 为实现上述目的,本发明提供的电子书排版系统,用于在不同屏幕大小的电子书阅读设备上对电子书进行排版,包括顺序连接的读取模块、格式分析模块、排版模块和显示模块;

读取模块用于读取电子书的文本数据；

格式分析模块分析文本数据的格式；

排版模块结合电子书阅读设备的环境参数对文本数据格式进行排版优化处理并排版；排版模块包括对文本数据的换行符分析部分、前后文语境检测部分及长段落分析部分；

显示模块用于显示排版后的电子书文本数据。

[0008] 采用本发明的系统能够根据不同屏幕大小的电子书阅读设备的环境参数来对文本数据格式进行排版处理,当排版模块查找到换行符时会对其前后字符进行分析以判断当前换行符的有效性,并根据语境类型判断长段落的分割点,有效避免了在不同电子书阅读设备阅读时,由于屏幕大小、像素等环境参数不同而产生的不正确换行、空白字符和长篇文本的情况,因此,使用本发明的系统优化了排版功能、大大提高了阅读效果,适合不同屏幕大小的电子阅读设备,此外,本系统成本小、安装简单,具有广泛的应用前景。

附图说明

[0009] 图 1 是本发明之电子书排版方法的总流程图。

[0010] 图 2 是本发明之电子书排版方法的换行符分析流程图。

[0011] 图 3 是本发明之电子书排版方法的当前字符的前一字符分析排版的示例图。

[0012] 图 4 是本发明之电子书排版方法的当前字符的后一字符分析排版的示例图。

[0013] 图 5 是本发明之电子书排版方法的前后文语境分析示例图。

[0014] 图 6 是本发明之电子书排版方法的长段落分析流程图。

[0015] 图 7 是本发明之电子书排版方法的长段落分析判断条件示意图。

[0016] 图 8 是本发明之电子书排版系统的系统结构示意图。

[0017] 以下结合各实施例及其附图对本发明作进一步说明。

具体实施方式

[0018] 电子书排版方法实施例

一种用于在不同屏幕大小的电子书阅读设备上对电子书进行排版的电子书排版方法,参见图 1,具体包括如下步骤:首先读取电子书中的文本数据;然后分析文本数据的格式;接着结合电子书阅读设备的环境参数对文本数据格式进行排版优化处理;再根据排版优化处理的格式对文本数据进行排版;最后显示排版后的电子书文本数据。其中,对文本数据格式进行排版优化处理的过程包括:换行符分析、前后文语境检测及长段落分析。

[0019] 分析文本数据格式具体包括:对文本数据进行文本符号分析和段落分析,不同的电子书格式对应不同的语法分析器(parser),语法分析器完成文本数据格式分析后,进入下一步骤对换行符进行分析。

[0020] 换行符分析的方法具体是对当前字符的前一字符和后一字符进行检测,以判断当前换行字符是否有效的方法,通过换行符的分析,可以正确的显示文本数据中的换行符,将文本数据多余换行和空白行删除,参见图 2,其具体步骤如下:

sa1:读取当前字符;

sa2:判断所述当前字符是否为换行符,若是则进入下一步骤 sa3,否则进入步骤 sa8;判断读取的当前字符是否为换行符,若是换行符则对当前字符的前后字符进一步判断,若

不是换行符则进入 sa8 的长段落分析步骤；

sa3 :读取所述当前字符的前一字符；

sa4 :判断所述前一字符是否属于：问号、省略号、后书名号、后引号、感叹号等句结束符中的其中之一，若是则初步判断当前换行符有效并进入下一步骤 sa5，否则进入步骤 sa8；当前字符为换行符，此时判断当前字符的前一字符是否属于句结束符，若属于句结束符则判断当前回车字符有效，否则进入步骤 sa8 的长段落分析过程，同时判断当前回车字符无效；句结束符包括问号(?)、省略号(……)、后书名号(》)、感叹号(!)、后引号(”)等，非句结束符则包括：逗号(,)、前引号(“)、冒号(:)、顿号(、)等，图 3 是对当前字符的前一字符进行判定而进行排版的示例，其中由于读取的当前字符为换行符，则读取前一字符进行分析，而前一字符为逗号，因此，判断当前换行字符无效。

[0021] sa5 :读取所述当前字符的后一字符；

sa6 :判断所述后一字符是否属于项目符号或其他开始符号的其中之一，若是则再次判断当前换行字符有效并进入下一步骤 sa7，否则进入步骤 sa8；当前一字符属于结束符时，进一步判断当前字符的后一字符是否属于项目符号或其他开始符号，如：●、★、◆等，若属于项目符号或其他开始符号，则判断当前换行字符有效，若不属于上述项目符号或开始符号，则进入长段落分析过程，并判断当前换行字符无效；图 4 是对后一字符判断并排版的示例，当前字符为换行符，由于前一字符为文字，因此进入后一字符的判断，后一字符为省略号，省略号不属于项目符号或开始符号，因此，判断当前换行字符无效。

[0022] sa7 :前后文语境分析过程，进入步骤 sa9；通过对当前换行字符的前后字符进行判断，当前一字符属于句结束符、后一字符属于项目符号或开始符号时，当前换行字符才是有效的，才能进入前后文语境分析过程；

其中，前后文语境分析的方法是根据文本数据的语境类别，如：对话、章节、句中、标题或人物介绍等，然后通过分析所述文本数据中的：开始标点、首字符、中间字符、尾标点及尾字符来判断句子是整句还是半句，从而判断所述当前换行字符是否有效的方法。图 5 是前后文语境分析的几个典型示例，其中通过语境判断、开始和尾标点、首尾字符及中间字符的综合判断，得出当前换行字符是否有效的判断结果以进行排版。

[0023] sa8 :判定换行符无效，进入长段落分析过程；当前字符不属于换行符，或其前一字符不属于句结束符，或后一字符不属于项目符号或开始符号时，判断当前字符不是换行符，此时则进入长段落分析的过程；

其中，长段落分析的具体过程如下，其流程参见图 6：

sb1 :获取屏幕的大小 w 和设置显示的字号 f；

sb2 :计算每个字所占的像素面积： $x = (\text{字宽} + \text{字间距}) * (\text{字高} + \text{行间距})$ ；

sb3 :计算分段的字符数： $m = (w * 1/3) / x$ ；

sb4 :根据分段的字符数对最近的句子进行查找，当最近句子中的字符个数大于等于分段的字符数时，进入步骤 sb5；

sb5 :进行整句句子查找；该步骤主要是判断文字数据的类型；判断条件具体包括：是否属于对话、如包含“”、“”等字符，此时判断上述字符之后为句子结束位置；或者是否达到最大段落门限，如果达到则判断句末为句子结束位置；或者判断是否属于项目符号，如：●、★、◆等，若属于则判断在上述项目符号之前为句子结束位置；或者判断是否属于章

节标题,若属于则第二个空格前为句子结束位置;具体判断条件及结果列表参见图7;

sb6:插入回车换行符;根据判断结果在句子结束位置插入回车换行符,从而根据屏幕参数对长段文字进行正确分割;

sb7:结束。结束长段落分析过程,准备进入下一步骤。

[0024] sa9:结束。结束换行符分析过程,准备进入下一步骤。

[0025] 完成上述分析后,根据优化排版处理结果,对电子书的文本数据进行排版后,将结果在电子书阅读设备上显示,此时,文本数据中多余的空白行、换行及没有“回车换行”的长段文字已经正确处理,用户在不同屏幕大小的电子阅读设备上均可获得良好的阅读效果、实现了良好的排版功能。

[0026] 本发明的电子书排版方法可以适用于各种电子书阅读设备,其中的文本适用于各种电子书阅读设备所采用的电子书格式,如:UMD、PDB、TXT等。

[0027] 作为本发明电子书排版方法实施例的一种变换,本发明可以适用于英文文本的排版,同样也可实现本发明的目的,对英文电子书数据进行排版的方法与上述方法所类似,不同之处在于对于标点的判断稍少于中文的标点判断,此处不再赘述。

[0028] 电子书排版系统实施例

用于在不同屏幕大小的电子书阅读设备上对电子书进行排版的电子书排版系统,包括:顺序连接的读取模块、格式分析模块、排版模块和显示模块;其系统结构图参见图8,读取模块用于读取电子书的文本数据;格式分析模块分析文本数据的格式;排版模块结合电子书阅读设备的环境参数对文本数据格式进行排版优化处理并排版;显示模块用于显示排版后的电子书文本数据。

[0029] 其中,格式分析模块包括对文本数据的文本符号分析部分和段落分析部分;格式分析模块可采用语法分析器来实现,针对不同电子书阅读设备,其对应的语法分析器各不相同。

[0030] 排版模块包括对文本数据的换行符分析部分、前后文语境检测部分及长段落分析部分;换行符分析部分对当前字符的前一字符和后一字符进行检测,以判断当前换行字符是否有效;前后文语境检测部分是根据文本数据的语境类别,分析所述文本数据中的:开始标点、首字符、中间字符、尾标点及尾字符来判断句子是整句还是半句,从而判断当前换行字符是否有效;长段落分析部分是根据屏幕大小分析分段字符数,并结合文字数据类型来判断是否插入回车符。

[0031] 在换行符分析部分中,通过换行符的分析,可以正确的显示文本数据中的换行符,将文本数据多余换行和空白行删除,首先读取当前字符并判断当前字符是否为换行符,如果是换行符则读取当前字符的前一字符并判断前一字符是否属于:问号、省略号、后书名号、后引号、感叹号等句结束符中的其中之一;当前一字符属于句结束符,则

读取当前字符的后一字符并判断后一字符是否属于项目符号或其他开始符号的其中之一,项目符号或开始符号包括如:●、★、◆等字符,若后一字符属于项目符号或开始符号,则判断当前换行符有效并由前后文语境分析部分开始分析;前述判断为非的情况都判断当前换行符无效并由长段落分析部分进行分析。

[0032] 前后文语境分析部分中,系统根据文本数据的语境类别,如:对话、章节、句中、标题或人物介绍等,然后通过分析文本数据中的:开始标点、首字符、中间字符、尾标点及尾字

符来判断句子是整句还是半句,从而判断所述当前换行字符是否有效,其具体分析方法和依据参见电子书排版方法实施例中的前后文语境分析方法。

[0033] 长段落分析部分中,系统首先获取屏幕的大小 w 和设置显示的字号 f ;然后计算每个字所占的像素面积 x 和分段的字符数 m : $x = (\text{字宽} + \text{字间距}) * (\text{字高} + \text{行间距})$ 、 $m = (w * 1/3) / x$;根据计算的结果对最近的句子进行查找并判断文字数据的类型;判断条件具体包括:是否属于对话、如包含“”、“”等字符,此时判断上述字符之后为句子结束位置;或者是否达到最大段落门限,如果达到则判断句末为句子结束位置;或者判断是否属于项目符号,如:●、★、◆等,若属于则判断在上述项目符号之前为句子结束位置;或者判断是否属于章节标题,若属于则第二个空格前为句子结束位置;具体判断条件及结果列表参见图7;判断结束后,在句子结束位置插入回车换行符,从而根据屏幕参数对长段文字进行正确分割。

[0034] 排版模块将文本数据中多余的空白行、换行以及未正确出现的换行符进行处理并排版后,通过显示模块将结果显示在电子书阅读设备上。

[0035] 本发明之电子书排版系统实施例中,可通过MCU实现系统的功能,也可以采用其他微控制器或可编程控制器来实现上述功能,同样都可实现本发明的目的。

[0036] 本发明的电子书排版系统可以适用于各种电子书阅读设备,其中的文本适用于各种电子书阅读设备所采用的电子书格式,如:UMD、PDB、TXT等。

[0037] 作为本发明电子书排版系统实施例的一种变换,本发明可以适用于英文文本的排版,同样也可实现本发明的目的,对英文电子书数据进行排版系统结构与上述结构所类似,此处不再赘述。

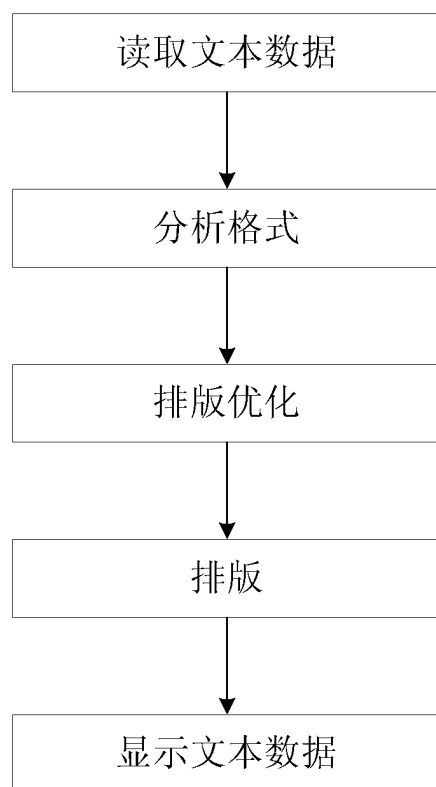


图 1

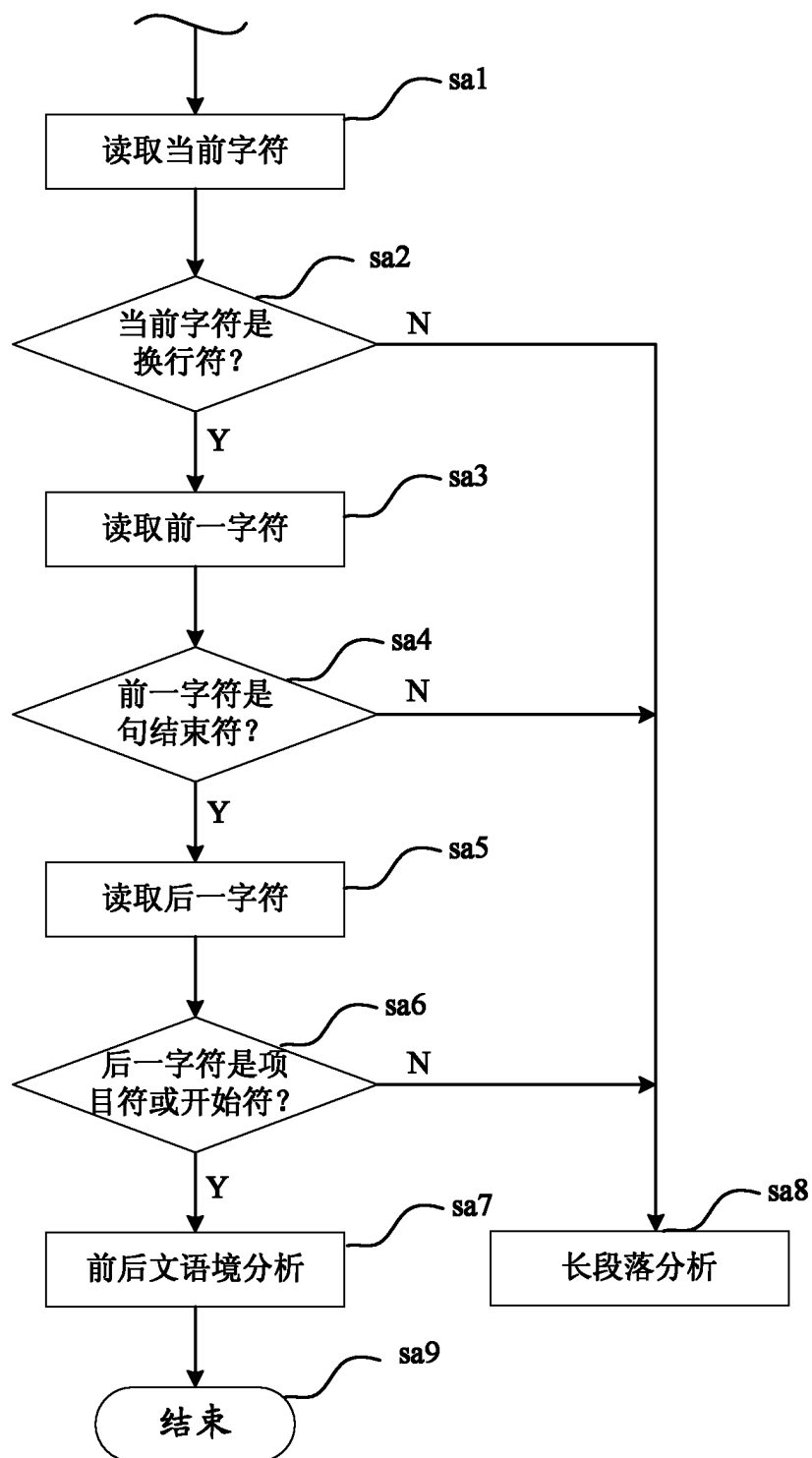


图2

A.判定当前换行符无效的示例

前一字符分析排版前：

这时嘴边已全是白沫，猛地里前腿一软，
跪倒在地。

前一字符分析排版后：

这时嘴边已全是白沫，猛地里前腿一软，跪倒在地。

B.判定当前换行符有效的示例

前一字符分析排版前：

可是谁来照料前面的娇妻幼女？
在身后，凶悍毒辣的敌人正在紧紧追踪。

前一字符分析排版后：

可是谁来照料前面的娇妻幼女？
在身后，凶悍毒辣的敌人正在紧紧追踪。

图 3

A.判定当前换行符无效的示例

后一字符分析排版前：

大哥，把地图给了他们，咱们认输便是。你
……你的身子要紧。

后一字符分析排版后：

大哥，把地图给了他们，咱们认输便是。你……你的身子要紧。

B.判定当前换行符有效的示例

后一字符分析排版前：

北京有几个著名旅游景点：

- ◆ 长城
- ◆ 香山公园

后一字符分析排版后：

北京有几个著名旅游景点：

- ◆ 长城
- ◆ 香山公园

图 4

Item	语境	开始标点	首字符	中间字符	尾标点	尾字符	回车换行符判定
1	人物对话	“「等	X		”」等		无效
		排版前： 那小姑娘也失声惊叫起来：「参， 参，你背上有箭！」那汉子					
		排版后： 那小姑娘也失声惊叫起来：「参，参，你背上有箭！」那汉子					
2	句中	，：	X		。！？		无效
		排版前： 一共六十三人，却带了一百九十多匹健 马，只要马力稍乏，就换一匹马乘坐。					
		排版后： 一共六十三人，却带了一百九十多匹健马，只要马力稍乏，就换一匹马乘坐。					
3	章节		第	Xxxx（数字）		章、节	无效
		排版前： 第五 章 挖出个未来一					
		排版后： 第五章 挖出个未来一					
4	标题						有效
		排版前： 第四章 发笔小财 丛林渐渐被黑幕所遮蔽，银白……					
		排版后： 第四章 发笔小财 丛林渐渐被黑幕所遮蔽，银白……					
5	人物介绍						有效
		排版前： 玄蜂 其实就是巨蜂，腹大如壶，蜇人，有毒，能杀人。					
		排版后： 玄蜂 其实就是巨蜂，腹大如壶，蜇人，有毒，能杀人。					

图 5

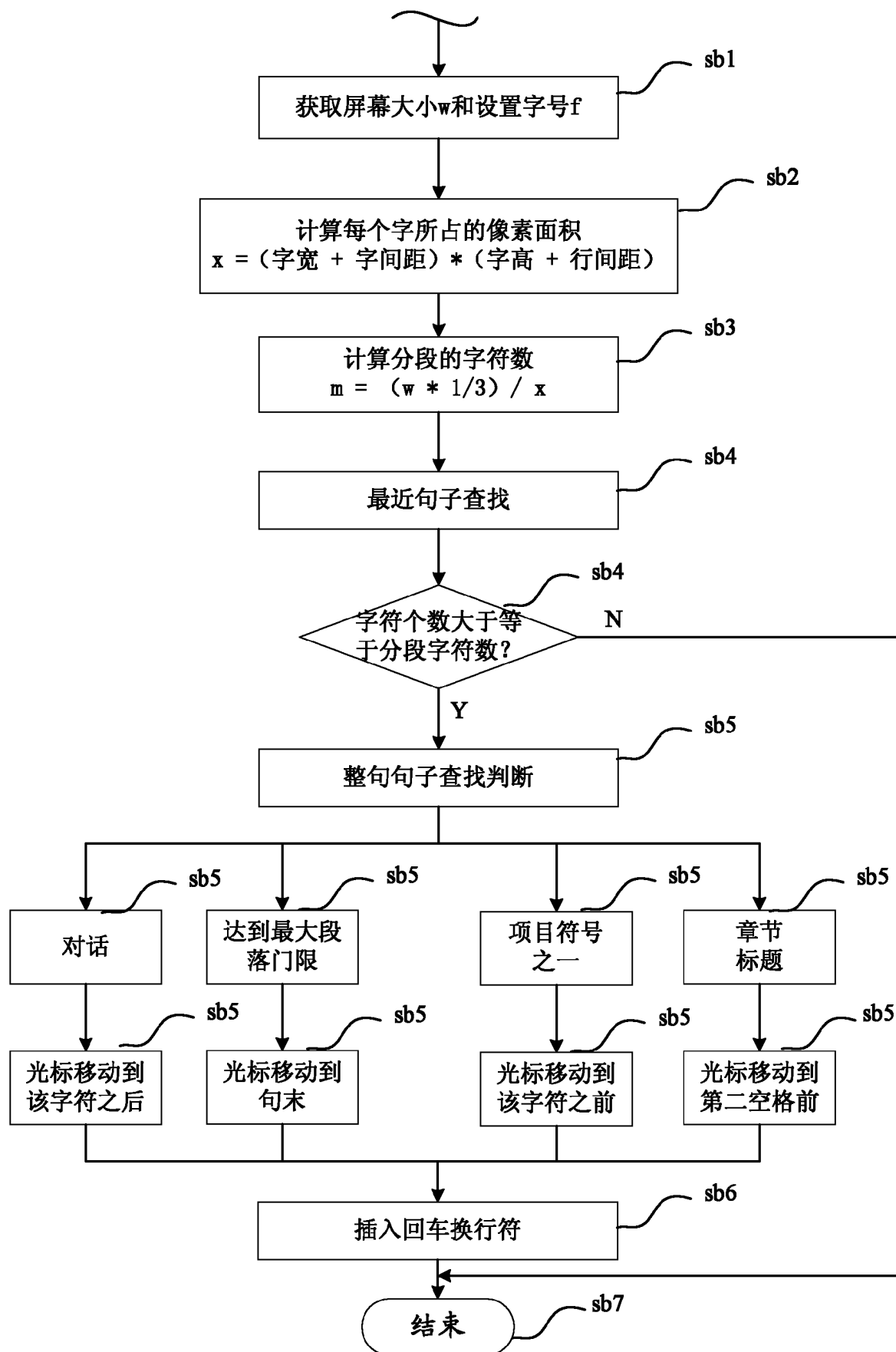


图 6

Item	判断条件	回车换行符插入位置	备注
1	对话，如包含“ ”、[]等	在“/”之后位置	
2	最大段落门限	在句末插入（如。！？等符号后）	
3	在项目符号（如：●★◆）	在符号前插入	
4	章节标题（如第10章xxxx）	在第二个空格处插入	

图 7

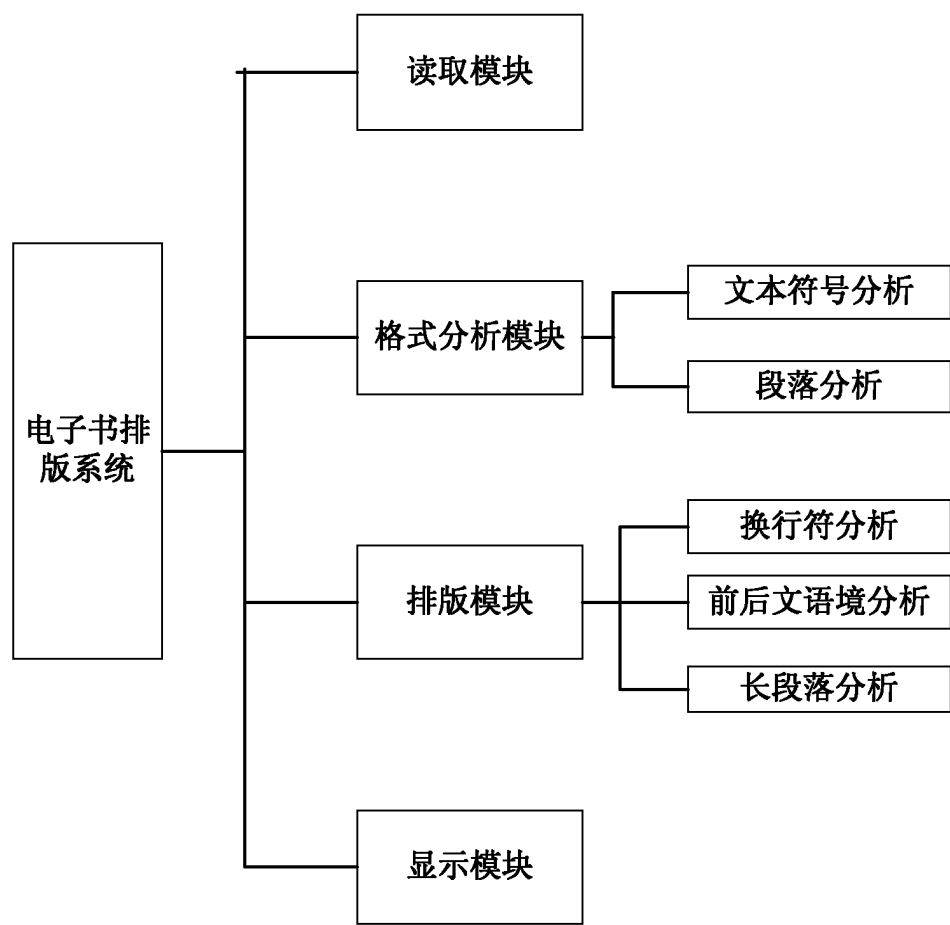


图 8