

(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
2014 年 7 月 3 日 (03.07.2014)



(10) 国际公布号
WO 2014/101504 A1

- (51) 国际专利分类号:
G06F 11/36 (2006.01) G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2013/084284
- (22) 国际申请日: 2013 年 9 月 26 日 (26.09.2013)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201210572578.5 2012 年 12 月 25 日 (25.12.2012) CN
- (71) 申请人: 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 518000 (CN)。
- (72) 发明人: 丁如敏 (DING, Rumin); 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 518000 (CN)。 霍举振 (HUO, Juzhen); 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 518000 (CN)。 高苡新 (GAO, Yixin); 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 518000 (CN)。

- (74) 代理人: 北京东方亿思知识产权代理有限公司 (BEIJING EAST IP LTD.); 中国北京市东城区东长安街 1 号东方广场东方经贸城东 2 座 1601 室, Beijing 100738 (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, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, 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, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

[见续页]

(54) Title: METHOD AND DEVICE FOR DETECTING WORDS IN APPLICATION PROGRAM

(54) 发明名称: 应用程序中的文字检测方法和装置

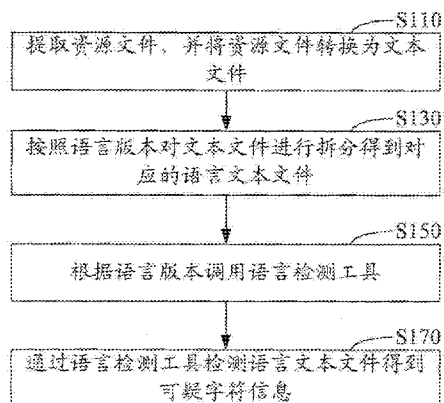


图 1 / Fig. 1

S110 EXTRACTING A RESOURCE FILE, AND CONVERTING THE RESOURCE FILE INTO A TEXT FILE
S130 ACCORDING TO A LANGUAGE VERSION, SPLITTING THE TEXT FILE TO OBTAIN A CORRESPONDING LANGUAGE TEXT FILE
S150 ACCORDING TO THE LANGUAGE VERSION, CALLING ON A LANGUAGE DETECTION TOOL
S170 DETECTING THE LANGUAGE TEXT FILE VIA THE LANGUAGE DETECTION TOOL TO OBTAIN SUSPICIOUS CHARACTER INFORMATION

(57) Abstract: Provided are a method and device for detecting words in an application program. The method comprises: extracting a resource file from a multi-language application program installation package, and converting the resource file into a text file; according to a language version, splitting the text file to obtain a corresponding language text file; according to the language version, calling on a language detection tool; and detecting the language text file via the language detection tool to obtain suspicious character information. The device comprises: a file processing module for extracting a resource file from a multi-language application program installation package, and converting the resource file into a text file; a splitting module for splitting the text file to obtain a corresponding language text file according to a language version; a tool invoking module for calling on a language detection tool according to the language version; and a text file detection module for detecting the language text file via the language detection tool to obtain suspicious character information. Adoption of the present invention can improve the detection efficiency of words in an application program.

(57) 摘要:

[见续页]

WO 2014/101504 A1

**本国际公布:**

— 包括国际检索报告(条约第 21 条(3))。

本发明实施例提供了一种应用程序中的文字检测方法和装置。所述方法包括：从多语言应用程序安装包中提取资源文件，并将所述资源文件转换为文本文件；按照语言版本对文本文件进行拆分得到对应的语言文本文件；根据语言版本调用语言检测工具；通过语言检测工具检测语言文本文件得到可疑字符信息。所述装置包括：文件处理模块，用于从多语言应用程序安装包中提取资源文件，并将所述资源文件转换为文本文件；拆分模块，用于按照语言版本对文本文件进行拆分得到对应的语言文本文件；工具调用模块，用于根据语言版本调用语言检测工具；文本检测模块，用于通过语言检测工具检测语言文本文件得到可疑字符信息。采用本发明能提高应用程序中文字检测的效率。

应用程序中的文字检测方法和装置

技术领域

5 本发明涉及软件调试技术，特别是涉及一种应用程序中的文字检测方法和装置。

背景技术

随着各种应用程序的普遍使用，每一应用程序所面向的用户类型也越来越多，应用程序用户使用的语言并不仅限于单一种类的语言文字，例如中文，而是越来越多的用户使用英文、马来西亚文等各种不同的语言文字，因此，为适应各种不同用户的需求，多语言应用程序根据用户的选择提供了各种不同语言版本。

相应的，在应用程序中的文字检测中也应当对交互界面的语言文字进行检测，应用程序中的文字检测方式包括切换不同的语言版本进行检测和提取多语言应用程序的资源文件进行人工检测。其中，在切换不同的语言版本进行检测的过程中，首先将多语言应用程序安装于终端设备，并运行，切换各种不同的语言版本，检查该语言版本下所有交互界面的语言文字，但是，这一检测方式难以确保遍历检查到每一语言版本下所有的交互界面，并且在检测过程中还需配合多语言应用程序的操作进行，效率较低。

20 而提取资源文件进行人工检测的方式需要交由不同语言的专业人士负责，例如，英文版本的应用程序需由英语专业人员进行检测，印度语版本的应用程序需由印度语专业人员进行检测，效率更为低下。

发明内容

25 基于此，有必要针对检测效率低的问题，提供一种能提高效率的应用程序中的文字检测方法。

此外，还有必要提供一种能提高效率的应用程序中的文字检测装置。

一种应用程序中的文字检测方法，包括如下步骤：

从多语言应用程序安装包中提取资源文件，并将所述资源文件转换为文本文件；

按照语言版本对所述文本文件进行拆分得到对应的语言文本文件；

根据所述语言版本调用语言检测工具；

5 通过所述语言检测工具检测所述语言文本文件得到可疑字符信息。

一种应用程序中的文字检测装置，包括：

文件处理模块，用于从多语言应用程序安装包中提取资源文件，并将所述资源文件转换为文本文件；

10 拆分模块，用于按照语言版本对所述文本文件进行拆分得到对应的语言文本文件；

工具调用模块，用于根据所述语言版本调用语言检测工具；

15 文本检测模块，用于通过所述语言检测工具检测所述语言文本文件得到可疑字符信息。上述应用程序中的文字检测方法和装置，提取资源文件，并自动将提取的资源文件转换为适于进行检测的文本文件，进而根据不同的语言版本将文本文件拆分为相应的语言文本文件，逐一对每一语言版本所对应的语言文本文件进行正确性检测，不需要配合多语言应用程序的操作以及专业人员的人工检测即可完成，大大地提高了效率。

20 附图说明

图 1 为一个实施例中应用程序中的文字检测方法的流程图；

图 2 为图 1 中按照语言版本对文本文件进行拆分得到对应的语言文本文件的方法流程图；

图 3 为另一个实施例中应用程序中的文字检测方法的流程图；

25 图 4 为另一个实施例中应用程序中的文字检测方法的流程图；

图 5 为另一个实施例中检测语言文本文件的方法流程图；

图 6 为一个实施例中应用程序中的文字检测装置的结构示意图；

图 7 为图 7 中拆分模块的结构示意图；

图 8 为另一个实施例中应用程序中的文字检测装置的结构示意图;

图 9 为另一个实施例中检测模块的结构示意图。

具体实施方式

5 如图 1 所示, 在一个实施例中, 一种应用程序中的文字检测方法, 包括如下步骤:

步骤 S110, 从多语言应用程序安装包中提取资源文件, 并将资源文件转换为文本文件。

10 本实施例中, 资源文件用于在应用程序中不需要重新编辑代码即可为用户提供各种交互界面以及交互界面中的文字、图像。资源文件包括了交互界面所对应的位图、字符以及其它数据, 每一应用程序对应了一个资源文件, 并根据提供的语言版本在资源文件中存储了相应语言版本的字符。

15 从应用程序的多语言应用程序安装包中提取资源文件, 将提取到的资源文件转换为文本格式, 得到该资源文件对应的文本文件, 以适用于资源文件中字符的检测。

具体的, 应用程序的多语言应用程序安装包为运行于终端设备, 并安装应用程序的可执行文件, 例如, 在移动终端的 Android 操作系统中, 应用程序的多语言应用程序安装包是后缀名为.apk 的文件。

步骤 S130, 按照语言版本对文本文件进行拆分得到对应的语言文本文件。

20 本实施例中, 资源文件为运行于终端设备中的应用程序提供了至少一个语言版本的交互界面, 并且资源文件的文本文件中, 字符是根据相应的语言版本进行有序存储的, 例如, 文本文件的每一列对应了一个语言版本。按照语言版本对文本文件进行拆分得到每一语言版本对应的语言文本文件, 将包含了多语言字符的文本文件分成多个单语言字符的文本文件, 即每一语言版本所对应的
25 语言文本文件。

如图 2 所示, 在一个实施例中, 上述步骤 S130 的具体过程为:

步骤 S131, 根据语言版本从文本文件中提取对应的字符。

本实施例中, 根据语言版本逐一在文本文件中提取该语言版本所对应的字

符。

步骤 S133, 将字符保存为语言文本文件, 并根据语言版本命名保存的语言文本文件。

5 本实施例中, 将根据语言版本提取到的字符保存为文本格式得到该语言版本所对应的语言文本文件, 并将该语言版本对应的语言名称用作语言文本文件的文件名。

步骤 S150, 根据语言版本调用语言检测工具。

10 本实施例中, 为测试应用程序交互界面中文字是否正确, 需逐一对每一语言版本的语言文本文件中的字符进行检测, 以判断语言文本文件中用于形成交互界面文字的字符正确性。

语言检测工具用于进行文字检测, 不同的语言检测工具对应不同的语言版本, 并对文本格式的文件进行检测。根据语言版本得到语言文本文件的语种, 进而调用该语种所对应的语言检测工具。语言检测工具所进行的文字检测包括拼写检测以及语法检测, 但并不仅限于此。

15 在一个示例中, 语言检测工具附带有记录了正确的字符的相应词库, 即, 字典。语言检测工具在语言文本文件和字典的比对中若语言文本文件的字符与字典记录的字符相一致, 则判定语言文本文件的字符为正确的, 若不相一致, 则判定语言文本文件的字符为可疑字符。

步骤 S170, 通过语言检测工具检测语言文本文件得到可疑字符信息。

20 本实施例中, 应用语言检测工具对语言文本文件中的字符进行检测得到可疑字符信息, 可疑字符信息包括了可疑字符、所在位置以及修改建议等信息, 语言文本文件中得到的可疑字符信息以列表的形式显示。

如图 3 所示, 在另一个实施例中, 上述应用程序中的文字检测方法还包括如下步骤:

25 步骤 S210, 从可疑字符信息中提取出误判的可疑字符信息。

本实施例中, 由语言检测工具进行检测得到的可疑字符信息可能包含了实质为正确的可疑字符信息, 即误判的字符信息, 因此还需对得到的可疑字符信息进行筛选以提取出误判的可疑字符信息。

如图 4 所示，在一个实施例中，上述步骤 S210 之前还包括如下步骤：

步骤 S310，判断可疑字符信息中是否存在误判的可疑字符信息，若是，则进入上述步骤 S210，若否，则进入步骤 S330。

5 本实施例中，获取人工筛选得到的误判的可疑字符信息，并以此为依据判断语言检测工具检测得到的可疑字符信息是否被误判，若是，则提取误判的可疑字符信息，若否，则直接输出语言检测工具检测得到的可疑字符信息，以方便修正语言文本文件中的字符。

步骤 S330，输出可疑字符信息。

步骤 S230，记录误判的可疑字符信息。

10 在一个实施例中，上述应用程序中的文字检测方法还包括：根据记录的误判的可疑字符信息去除检测得到的可疑字符信息中误判的可疑字符信息的步骤。

本实施例中，在对语言文本文件进行检测，并记录了语言文本文件中误判的可疑字符信息之后，还将根据记录的误判的可疑字符再次进行检测，去除可
15 疑字符信息中误判的可疑字符信息，重新得到更为准确的文字检测结果。

例如，在采用字典的示例中，将误判的可疑字符信息记录到字典。然后，利用更新后的字典再次进行检测，以去除可疑字符信息中误判的可疑字符信息，重新得到更为准确的文字检测结果。

如图 5 所示，在另一个实施例中，上述应用程序中的文字检测方法还包括：

20 步骤 S410，获取终端设备的分辨率信息。

本实施例中，获取运行应用程序的终端设备屏幕的分辨率信息。分辨率信息用于通过水平和垂直像素来衡量终端设备屏幕每一行或列中显示文字的数量。

步骤 S430，判断语言文本文件中字符的断行是否与分辨率信息相匹配，若
25 否，则进入步骤 S450，若是，则结束。

本实施例中，上述判断语言文本文件中字符的断行是否与分辨率信息相匹配的具体过程为：根据语言文本文件得到对应的语种并获取应用程序交互界面显示的字号，根据语种和显示的字号计算得到字符所占据的宽度，该宽度是通

过像素表示的，根据分辨率信息计算终端设备屏幕得到每行所包含的像素点数量，进而将每行所包含的像素点数量该宽度相除得到每行容纳字数，根据每行容纳字数判断语言文本文件中字符的断行是否正确，若不正确，则生成相应的可疑断行字符信息，以通过可疑断行字符信息标识断行错误的字符。

- 5 此外，上述步骤 S430 也可通过判断语言文本文件中同一行的字符长度所占据的像素值是否与屏幕中每行所包含的像素点数量相匹配来识别断行错误的字符，判断到同一行的字符长度所占据的像素值超出了屏幕中每行所包含的像素点数量，则说明出现了断行错误。

步骤 S450，生成字符对应的可疑断行字符信息。

- 10 本实施例中，该可疑断行字符信息包括了断行出错的字符、所处位置以及修改建议等信息。

如图 6 所示，在一个实施例中，一种应用程序中的文字检测装置，包括文件处理模块 110、拆分模块 130、工具调用模块 150 和文本检测模块 170。

- 15 文件处理模块 110，用于从多语言应用程序安装包中提取资源文件，并将资源文件转换为文本文件。

- 本实施例中，资源文件用于在应用程序中不需要重新编辑代码即可为用户提供各种交互界面以及交互界面中的文字、图像。资源文件包括了交互界面所对应的位图、字符以及其它数据，每一应用程序对应了一个资源文件，并根据
20 提供的语言版本在资源文件中存储了相应语言版本的字符。

文件处理模块 110 从应用程序的多语言应用程序安装包中提取资源文件，将提取到的资源文件转换为文本格式，得到该资源文件对应的文本文件，以适用于资源文件中字符的检测。

- 具体的，应用程序的多语言应用程序安装包为运行于终端设备，并安装应
25 用程序的可执行文件，例如，在移动终端的 Android 操作系统中，应用程序的多语言应用程序安装包是后缀名为.apk 的文件。

拆分模块 130，用于按照语言版本对文本文件进行拆分得到对应的语言文本文件。

本实施例中，资源文件为运行于终端设备中的应用程序提供了至少一个语言版本的交互界面，并且资源文件的文本文件中，字符是根据相应的语言版本进行有序存储的，例如，文本文件的每一列对应了一个语言版本。拆分模块 130 按照语言版本对文本文件进行拆分得到每一语言版本对应的语言文本文件，将
5 包含了多语言字符的文本文件分成多个单语言字符的文本文件，即每一语言版本所对应的语言文本文件。

如图 7 所示，在一个实施例中，拆分模块 130 包括字符提取单元 131 以及存储单元 133。

字符提取单元 131，用于根据语言版本从文本文件中提取对应的字符。

10 本实施例中，字符提取单元 131 根据语言版本逐一在文本文件中提取该语言版本所对应的字符。

存储单元 133，用于将字符保存为语言文本文件，并根据语言版本命名保存的语言文本文件。

本实施例中，存储单元 133 将根据语言版本提取到的字符保存为文本格式
15 得到该语言版本所对应的语言文本文件，并将该语言版本对应的语言名称用作语言文本文件的文件名。

工具调用模块 150，用于根据语言版本调用语言检测工具。

本实施例中，为测试应用程序交互界面中文字是否正确，需逐一对每一语言版本的语言文本文件中的字符进行检测，以判断语言文本文件中用于形成交互界面文字的字符正确性。
20

语言检测工具用于进行文字检测，不同的语言检测工具对应了不同的语言版本，并对文本格式的文件进行检测。工具调用模块 150 根据语言版本得到语言文本文件的语种，进而调用该语种所对应的语言检测工具。语言检测工具所进行的文字检测包括拼写检测以及语法检测，但并不仅限于此。

25 文本检测模块 170，用于通过语言检测工具检测语言文本文件得到可疑字符信息。

本实施例中，文本检测模块 170 应用语言检测工具对语言文本文件中的字符进行检测得到可疑字符信息，可疑字符信息包括了可疑字符、所在位置以及

修改建议等信息，语言文本文件中得到的可疑字符信息以列表的形式显示。

在一个示例中，语言检测工具附带有记录了正确的字符的相应词库，即，字典。语言检测工具在语言文本文件和字典的比对中若语言文本文件的字符与字典记录的字符相一致，则判定语言文本文件的字符为正确的，若不相一致，
5 则判定语言文本文件的字符为可疑字符。

如图 8 所示，在另一个实施例中，上述应用程序中的文字检测装置还包括误判信息提取模块 210 以及记录模块 230。

误判信息提取模块 210，用于从可疑字符信息提取出误判的可疑字符信息。

本实施例中，由语言检测工具进行检测得到的可疑字符信息可能包含了实质为正确的可疑字符信息，即误判的字符信息，因此还需误判信息提取模块 210
10 对得到的可疑字符信息进行筛选以提取出误判的可疑字符信息。

在另一个实施例中，上述应用程序中文字检测装置还包括误判信息识别模块，该误判信息识别模块用于判断可疑字符信息中是否存在误判的可疑字符信息，若是，则通知上述误判信息提取模块 210，若否，则输出可疑字符信息。

本实施例中，误判信息识别模块获取人工筛选得到的误判的可疑字符信息，并以此为依据判断语言检测工具检测得到的可疑字符信息是否被误判，若是，
15 则提取误判的可疑字符信息，若否，则直接输出语言检测工具检测得到的可疑字符信息，以方便修正语言文本文件中的字符。

记录模块 230，用于记录误判的可疑字符信息。

上述文本检测模块 170 还用于根据记录的误判的可疑字符信息去除检测得到的可疑字符信息中误判的可疑字符信息。
20

本实施例中，在对语言文本文件进行检测，并记录了语言文本文件中误判的可疑字符信息之后，文本检测模块 170 还将根据记录的误判的可疑字符再次进行检测，去除可疑字符信息中误判的可疑字符信息，重新得到更为准确的文字检测结果。
25

例如，在采用字典的示例中，将误判的可疑字符信息记录到字典。然后，利用更新后的字典再次进行检测，以去除可疑字符信息中误判的可疑字符信息，重新得到更为准确的文字检测结果。

如图 9 所示, 在另一个实施例中, 上述应用程序中的文字检测装置还包括获取模块 310 以及断行检测模块 330。

获取模块 310, 用于获取终端设备的分辨率信息。

5 本实施例中, 获取模块 310 获取运行应用程序的终端设备屏幕的分辨率信息。分辨率信息用于通过水平和垂直像素来衡量终端设备屏幕每一行或列中显示文字的数量。

断行检测模块 330, 用于判断语言文本文件中字符的断行是否与分辨率信息相匹配, 若否, 则生成该字符对应的可疑断行字符信息, 若是, 则停止执行。

10 本实施例中, 断行检测模块 330 根据语言文本文件得到对应的语种并获取应用程序交互界面显示的字号, 根据语种和显示的字号计算得到字符所占据的宽度, 该宽度是通过像素表示的, 根据分辨率信息计算终端设备屏幕得到每行所包含的像素点数量, 进而将每行所包含的像素点数量和该宽度相除得到每行容纳字数, 根据每行容纳字数判断语言文本文件中字符的断行是否正确, 若不正确, 则生成相应的可疑断行字符信息, 以通过可疑断行字符信息标识断行错误的字符。
15

此外, 断行检测模块 330 也可通过判断语言文本文件中同一行的字符长度所占据的像素值是否与屏幕中每行所包含的像素点数量相匹配来识别断行错误的字符, 判断到同一行的字符长度所占据的像素值超出了屏幕中每行所包含的像素点数量, 则说明出现了断行错误。可疑断行字符信息包括了断行出错的字符、所处位置以及修改建议等信息。
20

上述应用程序中的文字检测方法和装置, 提取资源文件, 并自动将提取的资源文件转换为适于进行检测的文本文件, 进而根据不同的语言版本将文本文件拆分为相应的语言文本文件, 逐一对每一语言版本所对应的语言文本文件进行正确性检测, 不需要配合多语言应用程序的操作以及专业人员的人工检测即可完成, 大大地提高了效率。
25

以上所述实施例仅表达了本发明的几种实施方式, 其描述较为具体和详细, 但并不能因此而理解为对本发明专利范围的限制。应当指出的是, 对于本领域

的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干变形和改进，这些都属于本发明的保护范围。因此，本发明专利的保护范围应以所附权利要求为准。

1、一种应用程序中的文字检测方法，包括如下步骤：

从多语言应用程序安装包中提取资源文件，并将所述资源文件转换为文本文件；

按照语言版本对所述文本文件进行拆分得到对应的语言文本文件；

5 根据所述语言版本调用语言检测工具；以及

通过所述语言检测工具检测所述语言文本文件得到可疑字符信息。

2、根据权利要求1所述的应用程序中的文字检测方法，其特征在于，所述按照语言版本对所述文本文件进行拆分得到对应的语言文本文件的步骤为：

根据语言版本从所述文本文件中提取对应的字符；

10 将所述字符保存为语言文本文件，并根据所述语言版本命名所述保存的语言文本文件。

3、根据权利要求1所述的应用程序中的文字检测方法，其特征在于，还包括：

获取终端设备的分辨率信息；

15 判断所述语言文本文件中字符的断行是否与所述分辨率信息相匹配，若否，则生成所述字符对应的可疑断行字符信息。

4、根据权利要求1所述的应用程序中的文字检测方法，其特征在于，还包括：

从所述可疑字符信息提取出误判的可疑字符信息；

20 记录所述误判的可疑字符信息；

根据所述记录的误判的可疑字符信息去除所述检测得到的可疑字符信息中误判的可疑字符信息。

5、根据权利要求4所述的应用程序中的文字检测方法，其特征在于，所述从所述可疑字符信息提取出误判的可疑字符信息的步骤之前还包括：

25 判断所述可疑字符信息中是否存在误判的可疑字符信息，若是，则进入所述从所述可疑字符信息提取出误判的可疑字符信息的步骤；

若否，则输出所述可疑字符信息。

6、一种应用程序中的文字检测装置，其特征在于，包括：

文件处理模块，用于从多语言应用程序安装包中提取资源文件，并将所述资源文件转换为文本文件；

5 拆分模块，用于按照语言版本对所述文本文件进行拆分得到对应的语言文本文件；

工具调用模块，用于根据所述语言版本调用语言检测工具；以及

文本检测模块，用于通过所述语言检测工具检测所述语言文本文件得到可疑字符信息。

10 7、根据权利要求6所述的应用程序中的文字检测装置，其特征在于，所述拆分模块包括：

字符提取单元，用于根据语言版本从所述文本文件中提取对应的字符；

存储单元，用于将所述字符保存为语言文本文件，并根据所述语言版本命名所述保存的语言文本文件。

15 8、根据权利要求6所述的应用程序中的文字检测装置，其特征在于，还包括：

获取模块，用于获取终端设备的分辨率信息；

断行检测模块，用于判断所述语言文本文件中字符的断行是否与所述分辨率信息相匹配，若否，则生成所述字符对应的可疑断行字符信息。

20 9、根据权利要求6所述的应用程序中的文字检测装置，其特征在于，还包括：

误判信息提取模块，用于从所述可疑字符信息提取出误判的可疑字符信息；

记录模块，用于记录所述误判的可疑字符信息；

所述文本检测模块还用于根据所述记录的误判的可疑字符信息去除所述检测得到的可疑字符信息中误判的可疑字符信息。

25 10、根据权利要求9所述的应用程序中的文字检测装置，其特征在于，还包括：

误判信息识别模块，用于判断所述可疑字符信息中是否存在误判的可疑字符信息，若是，则通知所述误判信息提取模块，若否，则输出所述可疑字符信息。

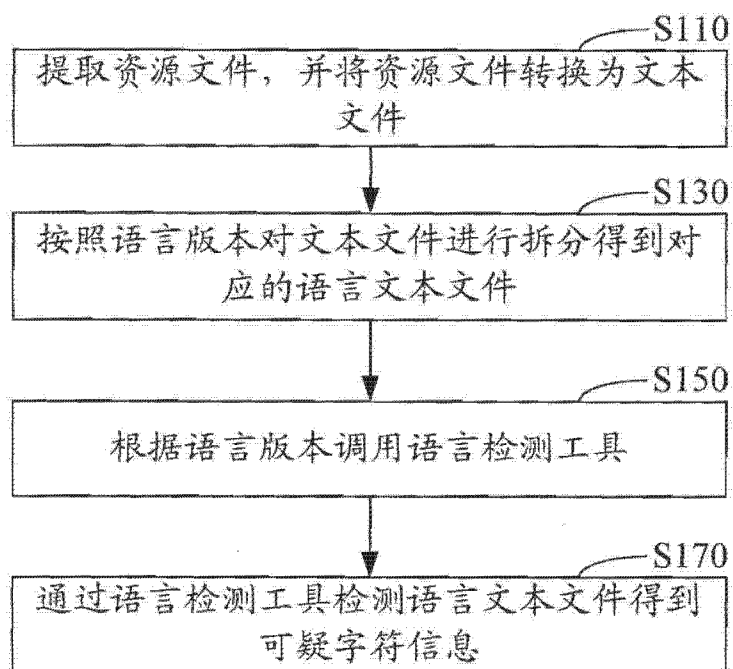


图 1

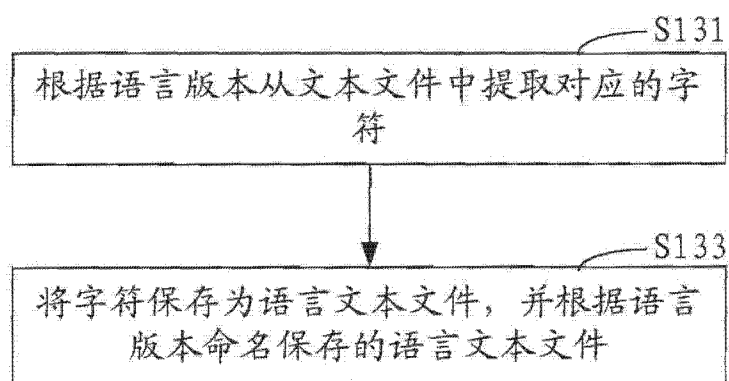


图 2

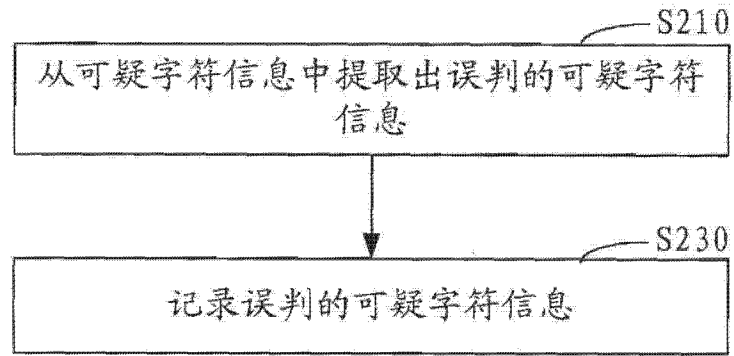


图 3

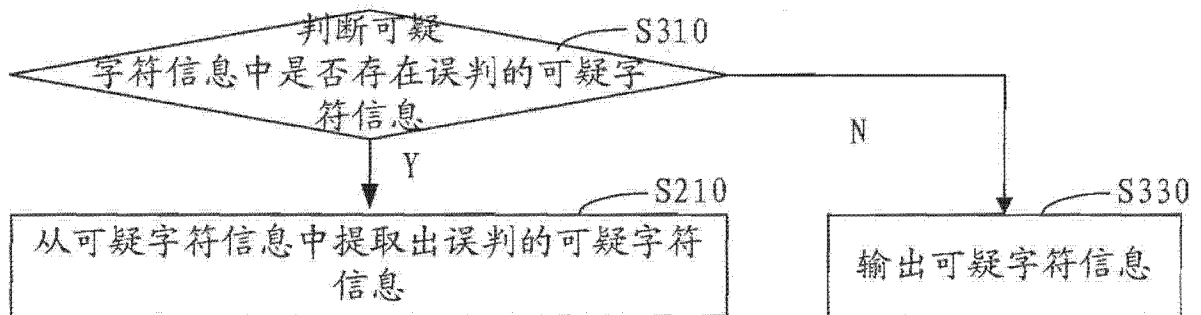


图 4

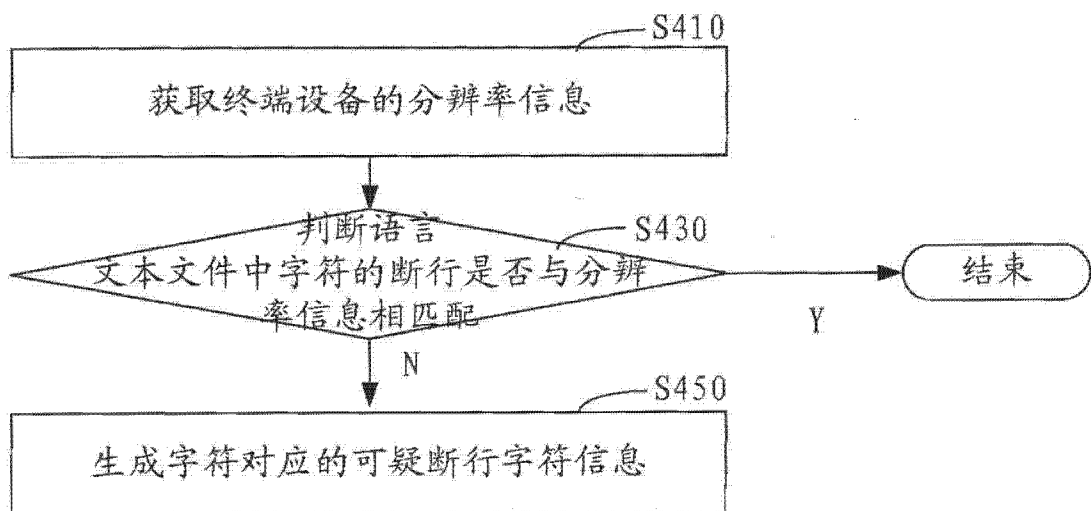


图 5

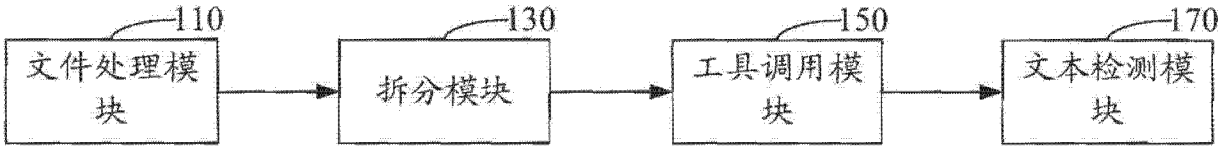


图 6

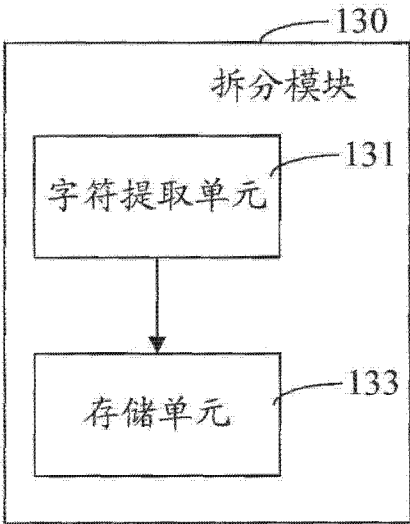


图 7

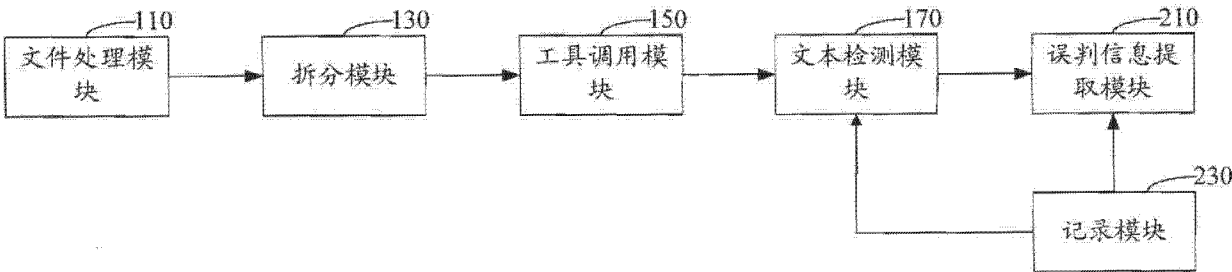


图 8

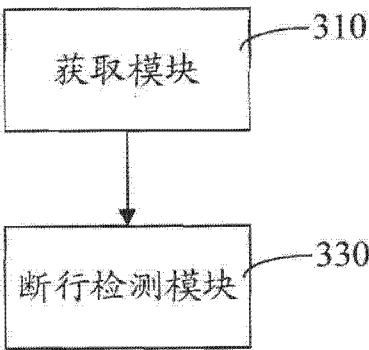


图 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/084284

A. CLASSIFICATION OF SUBJECT MATTER

See the extra sheet

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

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)

CNKI, CNTXT, EPODOC, WPI: resource file, text file, multi-language, suspicious character, language detection, check, line break, resolution, pixel, character, width, resource, text, detect+, char

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101140590 A (ZTE CORPORATION) 12 March 2008 (12.03.2008) description, page 5, lines 6-13, page 7, line 5 from the bottom to page 8, line 4	1, 2, 4-7, 9, 10
A	CN 102541847 A (KINGDEE SOFTWARE (CHINA) CO., LTD.) 04 July 2012 (04.07.2012) the whole document	1-10
A	US 5787452 A (SYBASE, INC.) 28 July 1998 (28.07.1998) the whole document	1-10

☐ 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 09 December 2013 (09.12.2013)	Date of mailing of the international search report 02 January 2014 (02.01.2014)
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer DONG, Gang Telephone No. (86-10) 62413649

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2013/084284

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101140590 A	12.03.2008	None	
CN 102541847 A	04.07.2012	None	
US 5787452 A	28.07.1998	None	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2013/084284

Continuation of: **CLASSIFICATION OF SUBJECT MATTER**

G06F 11/36 (2006.01) i

G06F 17/30 (2006.01) i

国际检索报告

国际申请号

PCT/CN2013/084284

A. 主题的分类

见附加页

按照国际专利分类(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

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))

CNKI, CNTXT, EPODOC, WPI; 资源文件, 文本文件, 多语言, 可疑字符, 语言检测, 检查, 断行, 分辨率, 像素, 字符, 宽度, width, resource, text, character, detect+, resolution, pixel, char, check

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	CN1011140590A (中兴通讯股份有限公司) 12.3 月 2008 (12.03.2008) 说明书第 5 页第 6-13 行, 第 7 页倒数第 5 行至第 8 页第 4 行	1-2,4-7,9-10
A	CN102541847A (金蝶软件(中国)有限公司) 04.7 月 2012 (04.07.2012) 全文	1-10
A	US5787452A (SYBASE, INC.) 28.7 月 1998 (28.07.1998) 全文	1-10



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(如具体说明的)

“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

09.12 月 2013 (09.12.2013)

国际检索报告邮寄日期

02.1 月 2014 (02.01.2014)

ISA/CN 的名称和邮寄地址:

中华人民共和国国家知识产权局

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

董刚

电话号码: (86-10) 62413649

关于同族专利的信息

PCT/CN2013/084284

PCT/ISA/210 表(同族专利附件)(2009 年 7 月)

续：主题的分类

G06F 11/36 (2006.01) i

G06F 17/30 (2006.01) i