

(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
2015年12月23日 (23.12.2015) WIPO | PCT

(10) 国际公布号
WO 2015/192567 A1

- (51) 国际专利分类号:
G06F 17/21 (2006.01)
- (21) 国际申请号: PCT/CN2014/089181
- (22) 国际申请日: 2014年10月22日 (22.10.2014)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201410270720.X 2014年6月17日 (17.06.2014) CN
- (71) 申请人: 中兴通讯股份有限公司 (ZTE CORPORATION) [CN/CN]; 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。
- (72) 发明人: 傅尚斌 (FU, Shangbin); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。 周侗 (ZHOU, Ti); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。
- (74) 代理人: 北京派特恩知识产权代理有限公司 (CHINA PAT INTELLECTUAL PROPERTY OFFICE); 中国北京市海淀区海淀南路21号中关村知识产权大厦B座2层, Beijing 100080 (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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, 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, ST, 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)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

(54) Title: METHOD AND DEVICE FOR EXTRACTING CHINESE LUNAR TIME FROM TEXT, AND COMPUTER STORAGE MEDIUM

(54) 发明名称: 从文本中提取农历时间的方法、装置及计算机存储介质

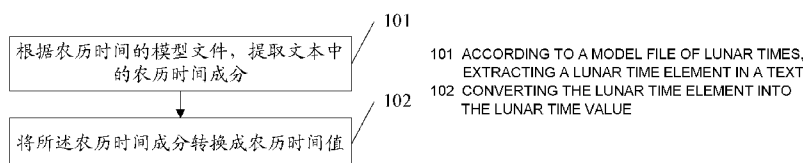


图 1 / Fig. 1

(57) Abstract: Disclosed in an embodiment of the present invention is a method for extracting Chinese lunar time from a text, comprising: according to a model file of lunar times, extracting a lunar time element in a text (101); and converting the lunar time element into the lunar time value (102). Also disclosed in the embodiment of the present invention are a device for extracting the lunar time from the text and a computer storage medium.

(57) 摘要: 本发明实施例公开一种从文本中提取农历时间的方法, 包括: 根据农历时间的模型文件, 提取文本中的农历时间成分 (101); 将所述农历时间成分转换成农历时间值 (102)。本发明实施例同时还公开了一种从文本中提取农历时间的装置和计算机存储介质。



WO 2015/192567 A1

从文本中提取农历时间的方法、装置及计算机存储介质

技术领域

本发明涉及终端领域中的数据处理技术，尤其涉及一种从文本中提取农历时间的方法、装置及计算机存储介质。

5 背景技术

在自然语言的发展过程中，人们走过一条从规则到统计的认识过程。在特定场景下，利用统计的方式，可以归纳出描述时间的种种方式，比如：描述具体日期的有XXXX年XX月XX日、几月几号；描述星期的有星期三、下周一等等。

10 在描述时间的过程中，农历与阳历存在一定的区别，例如：在描述日期时，从一号到十号，农历的表达方式是在数字前面加上“初”字，所以当出现几月初几的时候即可确定是用农历描述的；从十一号开始，农历与阳历的描述方式又是相同的，这需要用户根据生活环境及表达习惯等其他知识来判别。

15 目前，随着科学技术的发展，手机、平板电脑等便携式智能终端设备的普及率越来越高，短信、email等通讯软件在人们日常生活中发挥的作用也越来越大。在实际应用中，将会议及日常事务的行程安排等通过短信、email等方式传递给终端设备，在终端设备接收到信息后，若能自动地提取信息中与日程相关的农历时间信息，将更加方便人们的生活。以上也是相
20 关技术中需要解决的技术问题。

发明内容

有鉴于此，本发明实施例期望提供一种从文本中提取农历时间的方法、装置及计算机存储介质，能够自动地将农历时间从文本中提取出来，方便

用户进行日程管理。

为达到上述目的，本发明实施例的技术方案是这样实现的：

本发明实施例提供一种从文本中提取农历时间的方法，所述方法包括：

根据农历时间的模型文件，提取文本中的农历时间成分；将所述农历
5 时间成分转换成农历时间值。

上述方案中，该方法还包括：生成农历时间的模型文件；

所述生成农历时间的模型文件包括：统计农历时间的表达规则；根据
所述规则生成农历时间的正则表达式；将所述正则表达式封装成农历时间
的模型文件。

10 上述方案中，所述生成农历时间的正则表达式，包括：将表示农历时
间的确切时间设置于表示农历时间的相对时间之前。

上述方案中，所述根据农历时间的模型文件，提取文本中的农历时间
成分，包括：将文本中的内容与所述模型文件中的正则表达式进行遍历匹
配；对匹配得到的农历时间成分进行提取。

15 上述方案中，所述方法还包括：构建时间对象；所述将农历时间成分
转换成农历时间值，包括：将所述提取的农历时间成分赋给所述时间对象
的控制字段，获取农历时间值；所述时间对象的控制字段，依次包括：年、
月、日、时、分、秒。

20 上述方案中，所述方法还包括：当所述农历时间成分中含有相对时间
时，对获取的农历时间值进行相对时间计算。

本发明实施例还提供一种从文本中提取农历时间的装置，所述装置包
括：提取模块、以及转换模块；其中，所述提取模块，配置为根据农历时
间的模型文件，提取文本中的农历时间成分；所述转换模块，配置为将所
述农历时间成分转换成农历时间值。

25 上述方案中，所述装置还包括生成模块，所述生成模块，配置为生成
农历时间的模型文件，包括：统计农历时间的表达规则，根据所述规则生

成农历时间的正则表达式，将所述正则表达式封装成农历时间的模型文件。

上述方案中，所述生成模块，配置为将表示农历时间的确切时间设置于表示农历时间的相对时间之前。

上述方案中，所述提取模块，配置为将文本中的内容与所述模型文件
5 中的正则表达式进行遍历匹配；对匹配得到的农历时间成分进行提取。

上述方案中，所述装置还包括构建模块，所述构建模块，配置为构建时间对象；所述转换模块，配置为将所述提取的农历时间成分赋给所述时间对象的控制字段，获取农历时间值。

上述方案中，所述转换模块，还配置为当所述农历时间成分中含有相
10 对时间时，对获取的农历时间值进行相对时间计算。

本发明实施例还提供一种计算机存储介质，所述存储介质包括一组计算机可执行指令，所述指令用于执行本发明实施例所述的从文本中提取农历时间的方法。

本发明实施例提供的从文本中提取农历时间的方法、装置及计算机存
15 储介质，根据农历时间的模型文件，提取文本中的农历时间成分；将所述农历时间成分转换成农历时间值。如此，能够基于农历时间的模型文件，自动地从文本中提取农历时间成分，并将所提取的农历时间成分转换成具体的农历时间值，这样，实现了农历时间的自动提取和转换，能够有效地方便用户对生活及工作进行日程管理。

20 附图说明

图 1 为本发明实施例从文本中提取农历时间的方法实现流程示意图；

图 2 为本发明实施例生成农历时间的模型文件的实现流程示意图；

图 3 为本发明实施例提取及转换农历时间成分的实现流程示意图；

图 4 为本发明实施例从文本中提取农历时间的装置组成结构示意图。

具体实施方式

本发明实施例中，根据农历时间的模型文件，提取文本中的农历时间成分；将所述农历时间成分转换成农历时间值。

本发明实施例提供一种从文本中提取农历时间的方法，如图 1 所示，
5 图 1 为本发明实施例从文本中提取农历时间的方法实现流程示意图，所述方法包括以下步骤：

步骤 101：根据农历时间的模型文件，提取文本中的农历时间成分；

本发明实施例所述的从文本中提取农历时间的方法可以分为两部分：
农历时间模型文件的生成、以及农历时间成分的提取和转换。

10 本步骤中，在提取文本中的农历时间成分之前，首先生成农历时间的模型文件，也即本发明实施例所述方法的第一部分：农历时间模型文件的生成；图 2 为本发明实施例生成农历时间的模型文件的实现流程示意图，如图 2 所示，所述生成农历时间的模型文件包括以下步骤：

步骤 201：统计农历时间的表达规则；

15 本步骤中，所述统计农历时间的表达规则，也就是统计日常生活中有关农历的时间表达方式，这里统计的是日常生活中有关农历时间的一类表达式，而不是具体的农历时间描述。

这里，所述日常生活中的农历时间除了可以为我们经常使用的日期之外，还可以为二十四节气、传统节日、天干地支等。

20 步骤 202：根据所述规则生成农历时间的正则表达式；

本步骤中，针对日常生活中农历时间的表达规则，生成一系列农历时间的正则表达式，不同的正则表达式之间用或运算符进行分割，目前常用的农历时间的正则表达式的种类可以包括但不限于：

25
$$([\text{元正腊}]月初([一三四五六七八九十]+\backslash d+))([\text{元正腊}]月(\backslash d+))((大年(\backslash d+)))(大年初([一三四五六七八九十]+\backslash d+))(((\text{润}?)\backslash d+)月初([一三四$$

五六七八九十]+\d+)))(春节|(元宵|清明|端午|七夕|中秋|重阳|腊八([节]?)))(立春|雨水|惊蛰|春分|清明|谷雨|立夏|小满|芒种|夏至|小暑|大暑|立秋|处暑|白露|秋分|寒露|霜降|立冬|小雪|大雪|冬至|小寒|大寒)|([子丑寅卯辰巳午未申酉戌亥]时)|([寅卯辰巳午未申酉戌亥子丑]月)|((甲子|乙丑|丙寅|丁卯|戊辰|己巳|庚午|辛未|壬申|癸酉|甲戌|乙亥|丙子|丁丑|戊寅|己卯|庚辰|辛巳|壬午|癸未|甲申|乙酉|丙戌|丁亥|戊子|己丑|庚寅|辛卯|壬辰|癸巳|甲午|乙未|丙申|丁酉|戊戌|己亥|庚子|辛丑|壬寅|癸卯|甲辰|乙巳|丙午|丁未|戊申|己酉|庚戌|辛亥|壬子|癸丑|甲寅|乙卯|丙辰|丁巳|戊午|己未|庚申|辛酉|壬戌|癸亥)[年])。)

考虑到人们日常的语言表达习惯，在生成上述农历时间的正则表达式时，需要将表示农历时间的确切时间设置于表示农历时间的相对时间之前，以方便相对时间的计算。

步骤 203：将所述正则表达式封装成农历时间的模型文件；

将上述生成的一系列有关农历时间的正则表达式封装成农历时间的模型文件，所述模型文件相当于有关农历时间模型的数据库。

步骤 102：将所述农历时间成分转换成农历时间值；

上述步骤 101 和步骤 102 中的内容，也就是本发明实施例所述方法的第二部分：农历时间成分的提取和转换；图 3 为本发明实施例提取及转换农历时间成分的实现流程示意图，如图 3 所示，所述提取及换算农历时间成分可以包括以下步骤：

步骤 301：将文本中的内容与所述模型文件中的正则表达式进行遍历匹配，判断是否存在匹配项，若是，则执行步骤 302，若否，则结束流程；

在进行农历时间成分的提取时，需要将文本中的内容与所述模型文件中的各正则表达式进行遍历匹配，若存在匹配项，则执行步骤 302，若不存在匹配项，则结束流程。

需要说明的是，在匹配的过程中，如果文本中的某部分内容为一个完整的时间描述，需要将所述时间描述作为一个整体进行匹配；例如，假如

文本中存在连续的时间描述“大年三十晚上八点”，所述“大年三十”可以作为一个时间描述，所述“晚上八点”也可以作为一个时间描述，但由于两者连在一起表达的是一个完整的时间描述，因此，需要将“大年三十晚上八点”作为一个整体进行匹配。

5 步骤 302: 对匹配得到的农历时间成分进行提取;

需要说明的是，所述提取的农历时间成分，如果只是简单的有关农历时间的描述，则没有任何意义，因此，所述农历时间成分除了可以包括有关农历时间的描述之外，还可以包括但不限于：地点、人物、以及事件等。

10 步骤 303: 将所述提取的农历时间成分赋给时间对象的控制字段，获取农历时间值;

本发明实施例所述的方法需要构建一个时间对象，所述时间对象的控制字段，可以依次包括：年、月、日、时、分、秒；将提取的农历时间成分按照年、月、日、时、分、秒的顺序依次赋给时间对象的控制字段，可以获取农历时间值;

15 例如：假设提取的农历时间成分中有“大年三十晚上六点吃年夜饭”这个时间描述，而所述“大年”表示的是农历十二月还是正月，需要根据其后表示日期的数字而定，若其后的数字是二十以上，则表明所述“大年”表示的是农历十二月，若其后的数字是初几，则表明所述“大年”表示的是正月，因此，上述“大年三十”中所述“大年”表示的是十二月；此时，
20 首先将“大年”转换成“12”，并赋给时间对象控制字段中的“月”，“三十”是月份之后的日期，将所述“三十”转换成“30”，并赋给时间对象控制字段中的“日”；“晚上六点”表达的是日期后的小时，将所述“晚上六点”转换成“18”，并赋给时间对象控制字段中的“时”，以此获取农历时间值。

25 需要说明的是，当所述提取的农历时间成分中不存在与时间对象的某一控制字段相匹配的时间描述时，则跳过该控制字段，直接进行农历时间成分与下一控制字段的匹配过程。

步骤 304: 当所述农历时间成分中含有相对时间时, 对获取的农历时间值进行相对时间计算;

例如: 在所提取的农历时间成分“大年三十晚上六点吃年夜饭”之后, 补充一句“两小时后, 全家收看春晚”, 这里, 所述“两小时后”表示的是
5 相对时间, 即: 在“大年三十晚上六点”的两个小时之后, 因此, 需要对上述获取的农历时间值进行相对时间计算, 即: 将所述农历时间值中表示小时的数值加 2; 之后, 继续执行步骤 301, 直至文本中不存在与模型文件中的正则表达式相匹配的匹配项。

本发明实施例还提供一种从文本中提取农历时间的装置, 如图 4 所示,
10 图 4 为本发明实施例从文本中提取农历时间的装置组成结构示意图, 所述装置包括: 提取模块 42、以及转换模块 43; 其中,

所述提取模块 42, 配置为根据农历时间的模型文件, 提取文本中的农历时间成分;

所述提取模块 42, 可配置为将文本中的内容与所述模型文件中的正则
15 表达式进行遍历匹配; 对匹配得到的农历时间成分进行提取。

所述装置还可以包括生成模块 41, 配置为生成农历时间的模型文件;
所述生成农历时间的模型文件包括: 统计农历时间的表达规则, 根据所述规则生成农历时间的正则表达式, 将所述正则表达式封装成农历时间的模型文件;

20 所述生成模块 41, 配置为将表示农历时间的确切时间设置于表示农历时间的相对时间之前。

所述转换模块 43, 配置为将所述农历时间成分转换成农历时间值;

所述装置还可以包括构建模块 44, 配置为构建时间对象;

所述转换模块 43, 配置为将所述提取的农历时间成分赋给所述时间对
25 象的控制字段, 获取农历时间值。

所述转换模块 43, 还配置为当所述农历时间成分中含有相对时间时,

对获取的农历时间值进行相对时间计算。

在实际应用中，所述生成模块 41、提取模块 42、转换模块 43、以及构建模块 44 可由装置的中央处理器（CPU）、微处理器（MPU）、数字信号处理器（DSP）、或现场可编程门阵列（FPGA）实现。

5 本发明实施例还提供一种计算机存储介质，所述存储介质包括一组计算机可执行指令，所述指令用于执行本发明实施例所述的从文本中提取农历时间的方法。

本发明实施例中所述的从文本中提取农历时间的方法如果以软件功能模块的形式实现并作为独立的产品销售或使用，也可以存储在一个计算机可读存储介质中。基于这样的理解，本领域内的技术人员应明白，本
10 申请的实施例可提供为方法、系统、或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质上实施的计算机程序产品的形式，所述存储介质
15 包括但不限于 U 盘、移动硬盘、只读存储器（ROM, Read-Only Memory）、磁盘存储器、CD-ROM、光学存储器等。

本申请是根据本申请实施例的方法、装置、以及计算机程序产品的流程图和 / 或方框图来描述的。应理解可由计算机程序指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、
20 嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

25 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存

存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

5 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

10 相应的，本发明实施例还提供一种计算机存储介质，其中存储有计算机程序，该计算机程序用于执行本发明方法实施例中所述的从文本中提取农历时间的方法。

以上所述，仅为本发明的较佳实施例而已，并非用于限定本发明的保护范围。

权利要求书

- 1、一种从文本中提取农历时间的方法，所述方法包括：
根据农历时间的模型文件，提取文本中的农历时间成分；
将所述农历时间成分转换成农历时间值。
- 5 2、根据权利要求 1 所述的方法，其中，该方法还包括：生成农历时间的模型文件；
所述生成农历时间的模型文件包括：
统计农历时间的表达规则；
根据所述规则生成农历时间的正则表达式；
10 将所述正则表达式封装成农历时间的模型文件。
- 3、根据权利要求 2 所述的方法，其中，所述生成农历时间的正则表达式，包括：
将表示农历时间的确切时间设置于表示农历时间的相对时间之前。
- 4、根据权利要求 2 所述的方法，其中，所述根据农历时间的模型文件，提取文本中的农历时间成分，包括：
15 将文本中的内容与所述模型文件中的正则表达式进行遍历匹配；
对匹配得到的农历时间成分进行提取。
- 5、根据权利要求 1 至 4 任一项所述的方法，其中，所述方法还包括：
构建时间对象；
20 所述将农历时间成分转换成农历时间值，包括：
将所述提取的农历时间成分赋给所述时间对象的控制字段，获取农历时间值；所述时间对象的控制字段，依次包括：年、月、日、时、分、秒。
- 6、根据权利要求 5 所述的方法，其中，所述方法还包括：
25 当所述农历时间成分中含有相对时间时，对获取的农历时间值进行

相对时间计算。

7、一种从文本中提取农历时间的装置，所述装置包括：提取模块、以及转换模块；其中，

所述提取模块，配置为根据农历时间的模型文件，提取文本中的农
5 历时间成分；

所述转换模块，配置为将所述农历时间成分转换成农历时间值。

8、根据权利要求 7 所述的装置，其中，所述装置还包括生成模块，
所述生成模块，配置为生成农历时间的模型文件；

所述生成农历时间的模型文件包括：统计农历时间的表达规则，根
10 据所述规则生成农历时间的正则表达式，将所述正则表达式封装成农
历时间的模型文件。

9、根据权利要求 8 所述的装置，其中，

所述生成模块，配置为将表示农历时间的确切时间设置于表示农
历时间的相对时间之前。

15 10、根据权利要求 8 所述的装置，其中，

所述提取模块，配置为将文本中的内容与所述模型文件中的正则表
达式进行遍历匹配；对匹配得到的农历时间成分进行提取。

11、根据权利要求 7 至 10 任一项所述的装置，其中，所述装置还包
括构建模块，

20 所述构建模块，配置为构建时间对象；

所述转换模块，配置为将所述提取的农历时间成分赋给所述时间对
象的控制字段，获取农历时间值。

12、根据权利要求 11 所述的装置，其中，

所述转换模块，还配置为当所述农历时间成分中含有相对时间时，
25 对获取的农历时间值进行相对时间计算。

13、一种计算机存储介质，所述存储介质包括一组计算机可执行指

令,所述指令用于执行权利要求 1-6 任一项所述的从文本中提取农历时间的方法。

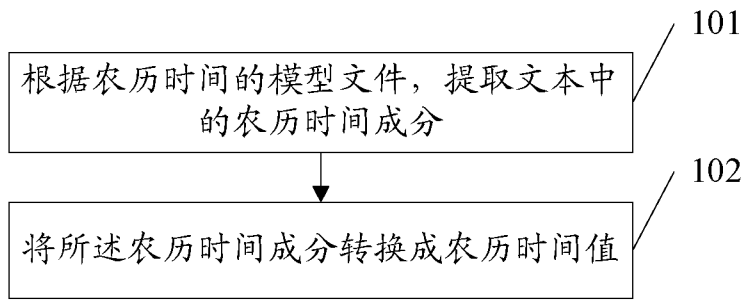


图 1

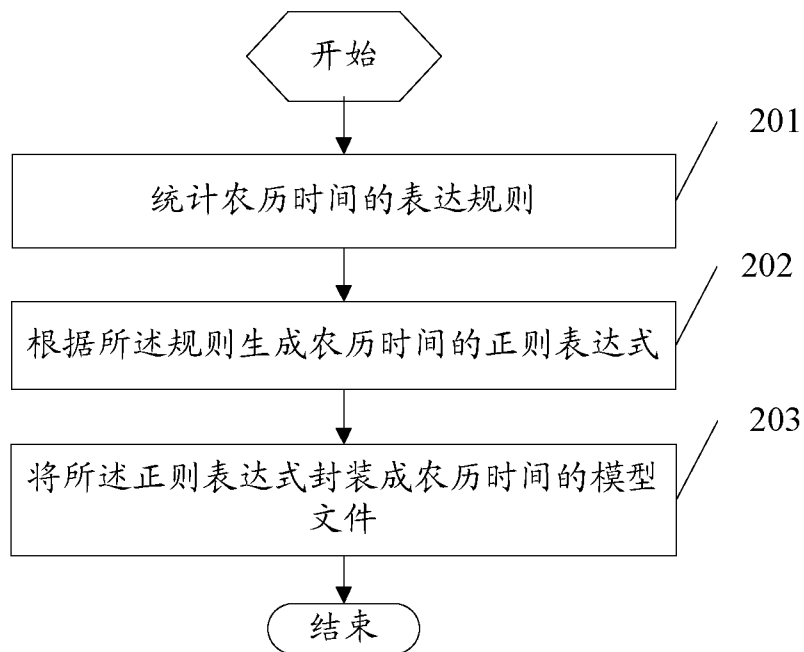


图 2

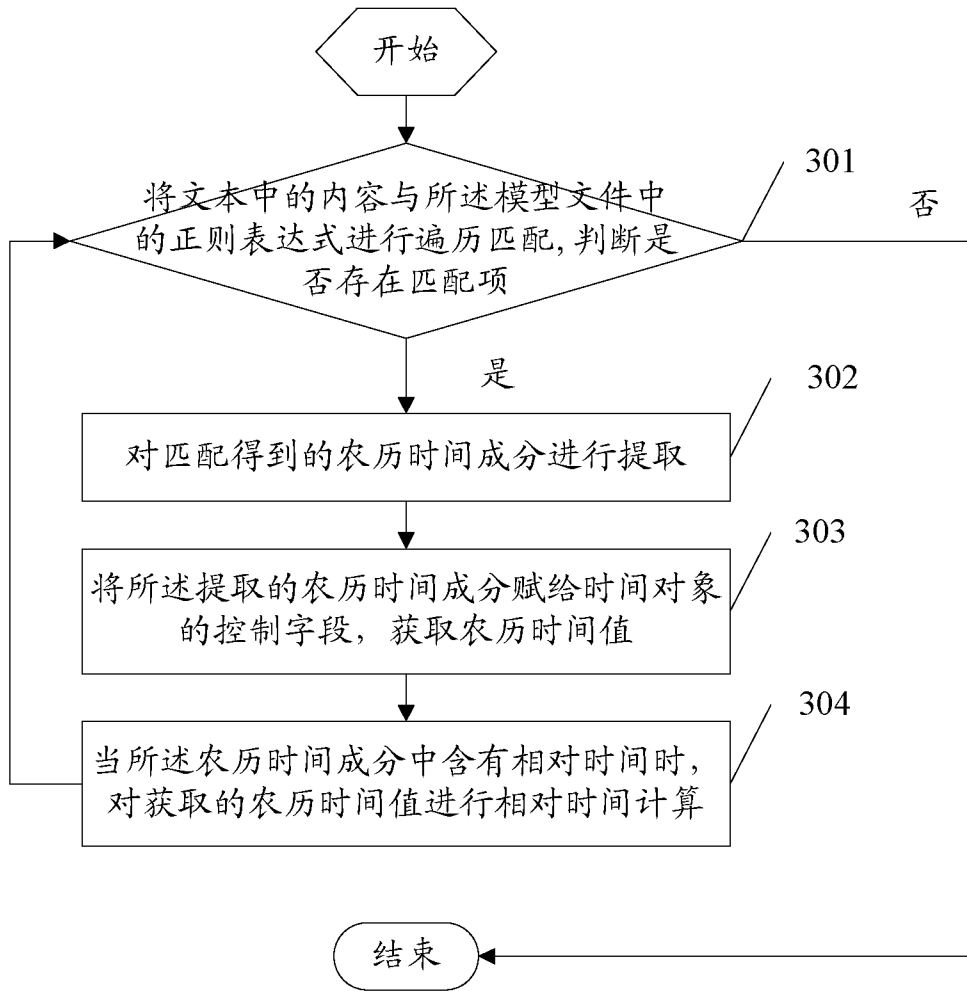


图 3

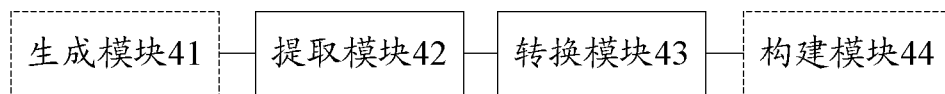


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2014/089181

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/21 (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)

G06F 17/21

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)

CPRSABS; CNTXT; CNABS; DWPI; VEN; CJFD; CNKI: time, word, template, date, lunar, part?, model?, extract+, comput+, express+, generat+, config+, text+, calend+, week?, festival?, solar, phrase, sentence, compos+, pattern?, obtain+, get+, read+, divid+, segment+, transf+, conver+, match+, regular, packag+, traver+, produc+, control+, filed?, file+, paragraph

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LIN, Jing et al., "Automatic TIME X2 Tagging of Chinese Temporal Information", JOURNAL OF TSINGHUA UNIVERSITY (SCIENCE AND TECHNOLOGY), vol. 48, no. 1, 15 January 2008 (15.01.2008), abstract, and the text, sections 1-2	1, 7, 13
Y	LIN, Jing et al., "Automatic TIME X2 Tagging of Chinese Temporal Information", JOURNAL OF TSINGHUA UNIVERSITY (SCIENCE AND TECHNOLOGY), vol. 48, no. 1, 15 January 2008 (15.01.2008), abstract, and the text, sections 1-2	2-6, 8-13
Y	CN 101727438 A (PEKING UNIVERSITY FOUNDER GROUP CORP. et al.), 09 June 2010 (09.06.2010), description, paragraphs 55-85	2-6, 8-13
A	US 2009192956 A1 (XEROX CORP.), 30 July 2009 (30.07.2009), the whole document	1-13

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"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</p> <p>"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</p> <p>"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</p> <p>"&" document member of the same patent family</p>
---	---

Date of the actual completion of the international search
20 December 2014 (20.12.2014)

Date of mailing of the international search report
04 January 2015 (04.01.2015)

Name and mailing address of the ISA/CN:
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
GUO, Yongqiang
Telephone No.: (86-10) **62412084**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2014/089181

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101727438 A	09 June 2010	CN 101727438 B	18 July 2012
US 2009192956 A1	30 July 2009	US 7991709 B2	02 August 2011

国际检索报告

国际申请号

PCT/CN2014/089181

<p>A. 主题的分类</p> <p>G06F 17/21 (2006.01) i</p> <p>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</p>																	
<p>B. 检索领域</p> <p>检索的最低限度文献 (标明分类系统和分类号)</p> <p>G06F 17/21</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CPRSABS; CNTXT; CNABS; DWPI; VEN; CJFD; CNKI: 时间, 农历, 成分, 模型, 提取, 计算, 表达, 生成, 配置, 文本, 日历, 日期, 星期, 节日, 阴历, 阳历, 部分, 词组, 词语, 语句, 文件, 段, 句, 模板, 模式, 获取, 获得, 得到, 读取, 分割, 划分, 转换, 换算, 匹配, 正则, 封装, 产生, 遍历, 控制, 字段, date, lunar, part?, model?, extract+, comput+, express+, generat+, config+, text+, calend+, week?, festival?, solar, phrase, sentence, compos+, pattern?, obtain+, get+, read+, divid+, segment+, transf+, conver+, match+, regular, packag+, traver+, produc+, control+, filed?, file+, paragraph</p>																	
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>林静等. "中文时间信息的TIMEX2自动标注" 清华大学学报(自然科学版), 第48卷, 第1期, 2008年1月15日 (2008-01-15), 摘要、正文第1-2节</td> <td>1, 7, 13</td> </tr> <tr> <td>Y</td> <td>林静等. "中文时间信息的TIMEX2自动标注" 清华大学学报(自然科学版), 第48卷, 第1期, 2008年1月15日 (2008-01-15), 摘要、正文第1-2节</td> <td>2-6, 8-13</td> </tr> <tr> <td>Y</td> <td>CN 101727438 A (北大方正集团有限公司等) 2010年6月09日 (2010-06-09) 说明书第55-85段</td> <td>2-6, 8-13</td> </tr> <tr> <td>A</td> <td>US 2009192956 A1 (XEROX CORP) 2009年7月30日 (2009-07-30) 全文</td> <td>1-13</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	林静等. "中文时间信息的TIMEX2自动标注" 清华大学学报(自然科学版), 第48卷, 第1期, 2008年1月15日 (2008-01-15), 摘要、正文第1-2节	1, 7, 13	Y	林静等. "中文时间信息的TIMEX2自动标注" 清华大学学报(自然科学版), 第48卷, 第1期, 2008年1月15日 (2008-01-15), 摘要、正文第1-2节	2-6, 8-13	Y	CN 101727438 A (北大方正集团有限公司等) 2010年6月09日 (2010-06-09) 说明书第55-85段	2-6, 8-13	A	US 2009192956 A1 (XEROX CORP) 2009年7月30日 (2009-07-30) 全文	1-13
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	林静等. "中文时间信息的TIMEX2自动标注" 清华大学学报(自然科学版), 第48卷, 第1期, 2008年1月15日 (2008-01-15), 摘要、正文第1-2节	1, 7, 13															
Y	林静等. "中文时间信息的TIMEX2自动标注" 清华大学学报(自然科学版), 第48卷, 第1期, 2008年1月15日 (2008-01-15), 摘要、正文第1-2节	2-6, 8-13															
Y	CN 101727438 A (北大方正集团有限公司等) 2010年6月09日 (2010-06-09) 说明书第55-85段	2-6, 8-13															
A	US 2009192956 A1 (XEROX CORP) 2009年7月30日 (2009-07-30) 全文	1-13															
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																	
<p>* 引用文件的具体类型:</p> <p>"A" 认为不特别相关的表示了现有技术一般状态的文件</p> <p>"E" 在国际申请日的当天或之后公布的在先申请或专利</p> <p>"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)</p> <p>"O" 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>"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</p> <p>"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</p> <p>"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</p> <p>"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</p> <p>"&" 同族专利的文件</p>																	
<p>国际检索实际完成的日期</p> <p>2014年 12月 20日</p>	<p>国际检索报告邮寄日期</p> <p>2015年 1月 04日</p>																
<p>ISA/CN的名称和邮寄地址</p> <p>中华人民共和国国家知识产权局 (ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国</p> <p>传真号 (86-10)62019451</p>	<p>受权官员</p> <p>郭永强</p> <p>电话号码 (86-10)62412084</p>																

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2014/089181

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	101727438	A	2010年 6月 09日	CN	101727438	B	2012年 7月 18日
US	2009192956	A1	2009年 7月 30日	US	7991709	B2	2011年 8月 02日

表 PCT/ISA/210 (同族专利附件) (2009年7月)