



(12) 发明专利申请

(10) 申请公布号 CN 105183704 A

(43) 申请公布日 2015. 12. 23

(21) 申请号 201410270720. X

(22) 申请日 2014. 06. 17

(71) 申请人 中兴通讯股份有限公司

地址 518057 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法务部

(72) 发明人 傅尚斌

(74) 专利代理机构 北京派特恩知识产权代理有限公司 11270

代理人 李睿 姚开丽

(51) Int. Cl.

G06F 17/21(2006. 01)

G06F 17/30(2006. 01)

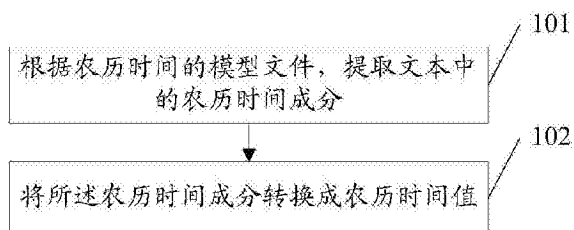
权利要求书2页 说明书5页 附图2页

(54) 发明名称

一种从文本中提取农历时间的方法及装置

(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;将所述农历时间成分转换成农历时间值;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。



1. 一种从文本中提取农历时间的方法,其特征在于,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。
2. 根据权利要求1所述的方法,其特征在于,该方法还包括:生成农历时间的模型文件,具体为:
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。
3. 根据权利要求2所述的方法,其特征在于,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,其特征在于,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。
5. 根据权利要求1至4任一项所述的方法,其特征在于,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。
6. 根据权利要求5所述的方法,其特征在于,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。
7. 一种从文本中提取农历时间的装置,其特征在于,所述装置包括:提取模块、以及转换模块;其中,
所述提取模块,用于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,用于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,其特征在于,所述装置还包括生成模块,
所述生成模块,用于生成农历时间的模型文件,具体为: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。
9. 根据权利要求8所述的装置,其特征在于,
所述生成模块,具体用于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。
10. 根据权利要求8所述的装置,其特征在于,
所述提取模块,具体用于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,其特征在于,所述装置还包括构建模块,
所述构建模块,用于构建时间对象;
所述转换模块,具体用于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,其特征在于,

所述转换模块,还用于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。

一种从文本中提取农历时间的方法及装置

技术领域

[0001] 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及装置。

背景技术

[0002] 在自然语言的发展历程中,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。

[0003] 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。

[0004] 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。

发明内容

[0005] 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及装置,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。

[0006] 为达到上述目的,本发明的技术方案是这样实现的:

[0007] 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:

[0008] 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。

[0009] 较佳的,该方法还包括:生成农历时间的模型文件,具体为: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。

[0010] 较佳的,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。

[0011] 较佳的,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。

[0012] 较佳的,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。

[0013] 较佳的,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。

[0014] 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,用于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,用于将所述农历时间成分转换成农历时间值。

[0015] 较佳的,所述装置还包括生成模块,所述生成模块,用于生成农历时间的模型文件,具体为: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。

[0016] 较佳的,所述生成模块,具体用于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。

[0017] 较佳的,所述提取模块,具体用于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。

[0018] 较佳的,所述装置还包括构建模块,所述构建模块,用于构建时间对象;所述转换模块,具体用于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。

[0019] 较佳的,所述转换模块,还用于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。

[0020] 本发明实施例提供的从文本中提取农历时间的方法及装置,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,能够有效地方使用户对生活及工作进行日程管理。

附图说明

[0021] 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;

[0022] 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;

[0023] 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;

[0024] 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

[0025] 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。

[0026] 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,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:

[0027] 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;

[0028] 可选的,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。

[0029] 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;可选的,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

文件具体包括以下步骤：

[0030] 步骤 201 :统计农历时间的表达规则；

[0031] 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。

[0032] 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。

[0033] 步骤 202 :根据所述规则生成农历时间的正则表达式；

[0034] 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：

[0035] ([元正腊]月初([一十]+|\d+))|([元正腊]月(\d+))|(大年(\d+))|(大年初一([一十]+|\d+))|((([润]?)(\d+))月初([一十]+|\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])。)

[0036] 可选的,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。

[0037] 需要说明的是,有关正则表达式的具体编写及运算过程为现有技术,此处不再赘述。

[0038] 步骤 203 :将所述正则表达式封装成农历时间的模型文件；

[0039] 可选的,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。

[0040] 步骤 102 :将所述农历时间成分转换成农历时间值；

[0041] 可选的,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具体可以包括以下步骤：

[0042] 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；

[0043] 可选的,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。

[0044] 需要注意的是,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。

[0045] 步骤 302:对匹配得到的农历时间成分进行提取;

[0046] 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。

[0047] 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;

[0048] 可选的,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;

[0049] 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,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。

[0050] 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。

[0051] 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;

[0052] 可选的,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。

[0053] 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,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,

[0054] 所述提取模块 42,用于根据农历时间的模型文件,提取文本中的农历时间成分;

[0055] 可选的,所述提取模块 42,具体用于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。

[0056] 可选的,所述装置还可以包括生成模块 41,所述生成模块 41,用于生成农历时间

的模型文件,具体为: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;

[0057] 所述生成模块 41,具体用于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。

[0058] 所述转换模块 43,用于将所述农历时间成分转换成农历时间值;

[0059] 可选的,所述装置还可以包括构建模块 44;其中;所述构建模块 44,用于构建时间对象;

[0060] 所述转换模块 43,具体用于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。

[0061] 所述转换模块 43,还用于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。

[0062] 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。

[0063] 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取存储介质中。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。

[0064] 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、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。

[0065] 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储器中的指令产生包括指令装置的制品,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。

[0066] 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。

[0067] 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。

[0068] 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。

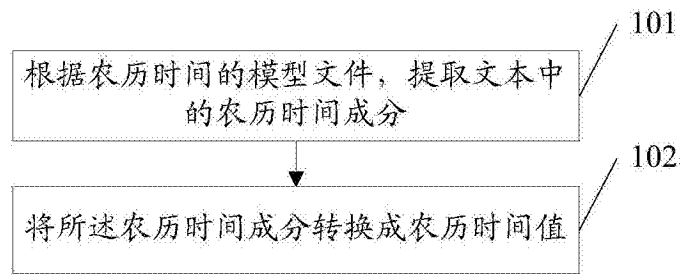


图 1

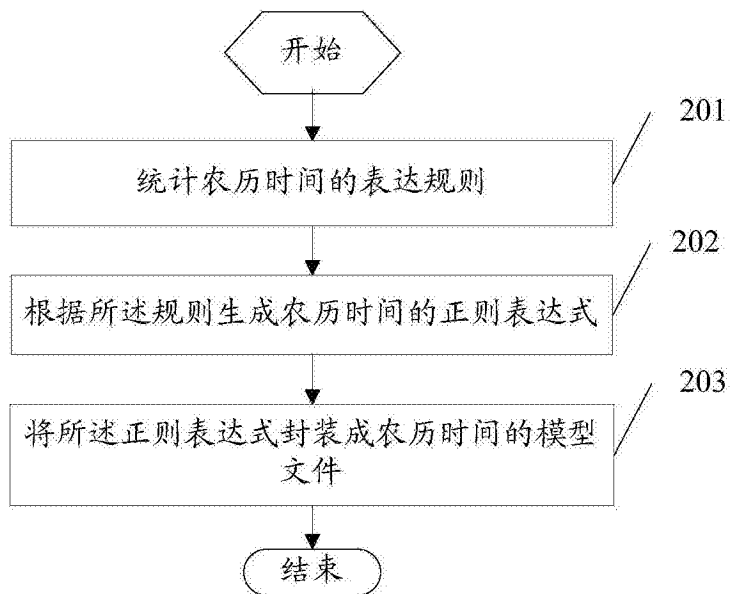


图 2

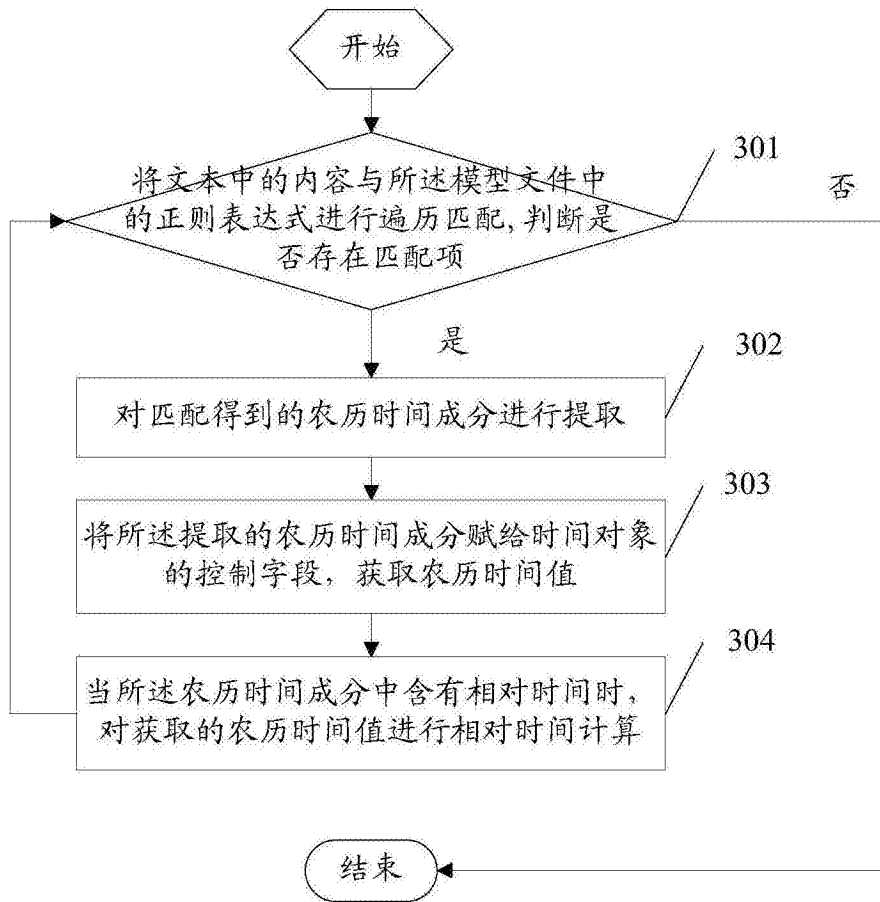


图 3

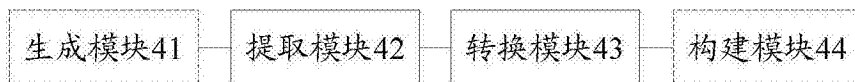


图 4