

(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
2012 年 10 月 26 日 (26.10.2012)



(10) 国际公布号
WO 2012/142962 A1

- (51) 国际专利分类号:
G06F 9/46 (2006.01)
- (21) 国际申请号: PCT/CN2012/074426
- (22) 国际申请日: 2012 年 4 月 20 日 (20.04.2012)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201110101178.1 2011 年 4 月 21 日 (21.04.2011) CN
- (71) 申请人 (对除美国外的所有指定国): **中国银联股份有限公司 (CHINA UNIONPAY CO., LTD.)** [CN/CN]; 中国上海市浦东新区含笑路 36 号银联大厦, Shanghai 200135 (CN)。
- (72) 发明人: 及
- (75) 发明人/申请人 (仅对美国): **陈林 (CHEN, Lin)** [CN/CN]; 中国上海市浦东新区含笑路 36 号银联大厦 7 楼, Shanghai 200135 (CN)。 **鲁志军 (LU, Zhijun)** [CN/CN]; 中国上海市浦东新区含笑路 36 号银联大厦 7 楼, Shanghai 200135 (CN)。 **茅毓铭 (MAO, Yuming)** [CN/CN]; 中国上海市浦东新区含笑路 36 号银联大厦 7 楼, Shanghai 200135 (CN)。 **杨燕明 (YANG, Yanming)** [CN/CN]; 中国上海市浦东新区含笑路 36 号银联大厦 7 楼, Shanghai 200135 (CN)。 **华锦芝 (HUA, Jinzhi)** [CN/CN]; 中国上海市浦东新区含笑路 36 号银联大厦 7 楼, Shanghai 200135 (CN)。

(74) 代理人: 中国专利代理 (香港) 有限公司 (CHINA PATENT AGENT (HK) LTD); 中国香港特别行政区香港湾仔港湾道 23 号鹰君中心 22 字楼, Hong Kong (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, 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, KM, 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, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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。

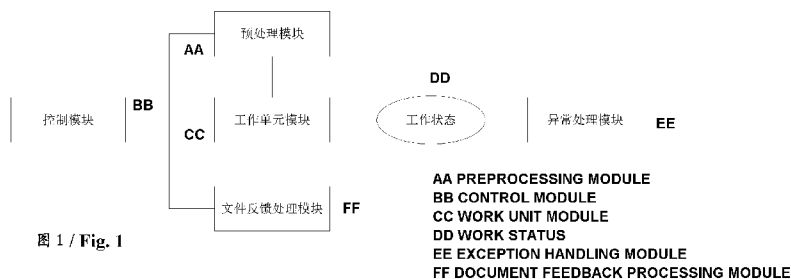
(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, 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 DOCUMENT LOADING

(54) 发明名称: 一种用于文件装载的方法和装置



(57) Abstract: Provided is a method and a device for document loading, the method comprising the following steps: a document is divided into one or more sections according to the record count of the document; two or more work units are utilized to concurrently process said one or more sections; each of the two or more work units updates a work status after processing a preset number of records within the corresponding sections; the work status is used to provide information for exception handling.

(57) 摘要:

提供了一种用于文件装载的方法和装置。该方法包括下列步骤: 根据文件中的记录数, 将文件分成一个或多个段; 利用两个或者更多个工作单元并发处理所述一个或多个段, 其中, 该两个或者更多个工作单元的每一个每次在处理完对应的段中的预定数量的记录后就更新工作状态, 该工作状态用于为异常处理提供信息。

WO 2012/142962 A1

一种用于文件装载的方法和装置

优先权要求

本申请要求了 2011 年 4 月 21 日提交的、申请号为 201110101178.1、
5 发明名称为“一种用于文件装载的方法和装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

10 本发明涉及文件装载的并发处理和断点续装，并且尤其涉及用于文件装载的方法和装置。

背景技术

在金融或电信等系统中，将应用参数录入到系统，即文件装载，是保持业务正确运行的前提，尤其地，在联机系统中快速装入参数是及时准确响应
15 服务的必要前提。现有技术中，常见的参数录入方式有人工逐参数录入或者通过数据库工具导入与数据库所要求的格式匹配的参数文件。现有技术中的这些方式有如下缺点：耗时、对参数录入以及参数文件格式要求较高、出错几率大，以及出错后的修复十分麻烦。

因此，需要一种能够实现文件的快速装载、对文件格式要求较低、出错
20 几率小以及修复简单的方法和装置。

发明内容

为解决现有技术的上述问题和其它问题，本发明的一个目的在于提供一种文件装载方法，包括以下步骤，

25 根据文件中的记录数，将文件分成一个或多个段，

利用两个或者更多个工作单元并发处理一个或多个段，

其中，两个或者更多个工作单元的每一个每次在处理完对应的段中的预定数量的记录后就更新各自的工作状态，工作状态用于为异常处理提供信息。

30 优选地，将对应于一个或多个段的段信息分配给两个或者更多个工作单

元，每个工作单元根据被分配到的段信息处理相应的段。

优选地，段信息包括用于指示相应的段的开始记录和结束记录的指针。

优选地，工作状态包括对预定数量的记录的处理结果。

优选地，工作状态进一步包括发生异常的记录的指针。

- 5 优选地，当工作状态指示处理结果为异常，从发生异常的记录的指针处开始恢复处理。

优选地，采用轮询模式监控工作状态。

优选地，上述方法进一步包括向外部系统发送关于工作状态的信息，并将来自外部系统的响应信息更新到工作状态中。

- 10 优选地，工作状态被存储在文件中，以及两个或者更多个工作单元共同维护一个工作状态文件或者两个或者更多个工作单元分别维护各自的工作状态文件。

优选地，文件是参数文件，记录是参数，处理是将参数文件中的参数录入到系统的数据库中。

- 15 优选地，处理通过将参数直接装载到正式参数表来完成录入。

根据本发明的另一个目的，提供了一种文件装载装置，包括：

预处理模块，用于根据文件中的记录数，将文件分成一个或多个段，

工作单元模块，包括两个或者更多个工作单元，用于并发处理一个或多个段，

- 20 异常处理模块，用于进行异常处理，

其中，两个或者更多个工作单元的每一个每次在处理完对应的段中的预定数量的记录后就更新各自的工作状态，工作状态用于为异常处理模块提供信息。

- 25 优选地，预处理模块将对应于一个或多个段的段信息分配给两个或者更多个工作单元，每个工作单元根据被分配到的段信息处理相应的段。

优选地，段信息包括用于指示相应的段的开始记录和结束记录的指针，工作状态包括对预定数量的记录的处理结果和发生异常情况的记录的指针，

异常处理模块被配置成监控工作单元的工作状态，当工作状态指示处理结果为异常时，从发生异常的记录开始恢复处理。

- 30 优选地，该装置进一步包括文件反馈处理模块，用于向外部系统发送关于工作状态的信息，并将来自外部系统的响应信息更新到工作状态中。

优选地，该装置进一步包括文件控制模块，用于文件的预处理和工作单

元的处理装载进行监控。

本文提出了一种支持并发和断点续装的参数文件装载方法和装置。这种方法针对到文件读写的效率比较低,不能满足大记录数下短时间内解析并实时装载的需求,提出了并发装载,如采用的多线程或者多线程并发的方式。

- 5 对待装载文件首先根据文件大小决定文件并发的数目,对每一个工作单元(相应的线程或者进程)分配不同的文件段,不同线程在各自的段内独立工作互不影响,这种并发装载的设计有效地克服了文件串行处理以及 I/O 上的低效率的缺陷。

- 10 现有技术中的文件装载通常先将参数慢慢装入临时表,处理好后再统一导入正式表。本发明针对大记录数文件装载时间长,并且在期间发生磁盘、数据库、网络、主机掉电等异常时恢复起来麻烦的缺陷,本发明的方法将参数直接装载到正式参数表,提高参数的时效性,并实时记录每个工作单元已处理工作点的方式。因此即使发生了异常也能够从断点处迅速恢复,不会出现大面积重做或者正式表记录已经有前面处理的冗余数据问题。

- 15 本发明的技术方案在应用金融信息等系统的参数文件装载中能够缩短业务数据生效时间、提高服务质量和异常恢复能力。

附图说明

- 20 在参照附图阅读了本发明的具体实施方式以后,本领域技术人员将会更清楚地了解本发明的各个方面。本领域技术人员应当理解的是,这些附图仅仅用于配合具体实施方式说明本发明的技术方案,而并非意在对本发明的保护范围构成限制。其中,

图 1 是根据本发明实施例的文件装载装置的结构示意图。

25 具体实施方式

下面参照附图,对本发明的具体实施方式作进一步的详细描述。

如图 1 所示,文件装载装置可以包括预处理模块、工作单元模块、异常处理模块、控制模块和文件反馈处理模块。

- 30 预处理模块模块是文件装载的入口,负责接收来自外部系统的调用指令,并将调用指令进行解析和预处理。其根据调用指令装载文件或参数文件

的记录，并且根据文件中的记录数，将文件分成一个或多个段。

在一个实施例中，预处理模块可以将对应于一个或多个段的段信息分配给两个或者更多个工作单元，每个工作单元根据被分配到的段信息处理相应的段。

- 5 在一个实施例中，段信息可以包括待处理文件的记录的指针，指示相应的段的开始记录和结束记录。也就是说，段信息可以包含待处理参数文件的指针，指示段的始末位置。进一步，可以控制单个文件的分段数量。当可被并发处理的段的数量达到预定上限时，则等待至前面被派发出去的段由工作单元处理完成，即直到有工作单元闲置的时候再继续分配段。后文将详细描述工作单元的处理。

- 10 这里，预处理模块将单个文件拆分成一个或多个段并让多个工作单元并发处理的优点在于，一方面，分段并发处理可以充分利用当前主机系统的并发特性，大幅度地提升装载文件的速度；另一方面，如本文后面将详细描述，分段并发处理能够降低异常处理的颗粒度。当某个段异常时，只需重做该段，从而提升装载文件的效率。

本领域的技术人员可以理解的本发明不限于将参数文件内数据提取到参数数据库，还可以被应用于任何文件的装载或者对于文件的任何处理。

工作单元模块包括两个或者更多个工作单元，用于并发处理所述一个或多个段以完成每个文件段的装置任务。

- 20 在一个实施例中，两个或者更多个工作单元的每一个每次在处理完对应的段中的预定数量的记录后就更新工作状态，工作状态可以被用于为后文将描述的异常处理模块提供信息。例如，工作单元每解析装载段内的预定数量的数据则提交数据库并更新记录工作单元的工作状态。这里，工作状态可以被存储在文件中，工作单元可以共同维护一个工作状态文件或者分别维护各自的工作状态文件。作为文件记录的工作状态可以被用于实现参数文件的断点续传。举例来说，段内具有 1000 条记录，而“预定数量”可以被设定为 10 条，即工作单元处理好 10 条后更新工作状态。如此，更新越频繁后续出现异常时断点续装中重复处理记录的数量越少。

- 25 在一个实施例中，工作状态进可以包括发生异常情况的记录的指针来指示恢复操作的起点。

在一个实施例中，工作状态可以包括对所述预定数量的记录的处理结果，该处理结果可以包括正常情况和异常情况。

异常处理模块，用于进行异常处理，对工作状态为异常的段进行异常排查，待异常处理完毕，将工作状态更新为正常。

在一个实施例中，当工作状态指示处理结果为异常时，异常处理模块从发生异常情况的记录开始恢复处理。例如，工作单元异常终止或者主机断电等应用重启后，均从工作单元状态的文件记录的断点处，即发生异常的记录处开始恢复。如此能够缩减重复处理概率和异常恢复所需要时间，例如，当10000条记录已经处理了9990条记录时，则该重启后根据上次保留的工作状态，需要重做的记录数目是10条，而无需从头文件开始处重新处理。

在一个实施例中，异常处理模块可以采用轮询模式监控工作状态。

可选地，文件装载装置可以包括控制模块，用于对文件装载请求的处理状态以及各个工作单元的处理进行监控。当发现文件分段调度完成并且所有被调度工作单元已完成时，控制模块启动下一步文件处理。本领域的技术人员可以理解的还可以利用其它已知的各种技术来实现对反馈文件生成流程的监控和/或触发响应。

可选地，文件装载装置可以包括文件反馈处理模块，用于向外部系统发送关于工作状态的信息，并将来自外部系统的响应信息更新到工作状态中。反馈处理模块可以作为反馈文件生成过程中的对外“出口”。本领域的技术人员可以理解的还可以利用其它已知的各种技术来实现文件反馈的功能。

本发明的上述的模块都可以设计成多进程并发模式，通过将单一参数文件任务拆分成多段的装载任务、任务并发执行、可单任务重做，配合各个任务在任务状态可实现异常时从断点处重装，即可实现高效、高可用的文件装载。

本发明的支持并发和断点续装的文件装载方法和装置，充分利用了系统资源、进程间通信以及数据库资源，提高了文件装载的效率和异常恢复能力，尤其在大记录数（如十万到千万级的记录）时，实时将参数文件装载直接装载到正式表的性能与异常恢复方面具有很高的价值。正常情况下所用时间近似为非并发下的“并发数分之一”，而无断点续装机制的情形下，异常恢复并完成装载的时间大于正常装载时间，有断点续装机制的情形下异常恢复并完成装载的时间约为正常装载时间减异常前已处理时间。

通过以上实施方式的描述，本领域的技术人员可以清楚地了解到本发明可借助软件结合硬件平台的方式来实现，当然也可以全部通过硬件来实施。本领域中的普通技术人员能够理解，在不偏离本发明的精神和范围的情况

下，还可以对本发明的具体实施方式作各种变更和替换。这些变更和替换都落在本发明权利要求书所限定的范围内。

权 利 要 求

- 1、一种文件装载方法，其特征在于，
根据文件中的记录数，将文件分成一个或多个段，
5 利用两个或者更多个工作单元并发处理所述一个或多个段，
其中，所述两个或者更多个工作单元的每一个每次在处理完对应的段中的
预定数量的记录后就更新各自的工作状态，所述工作状态用于为异常处理
提供信息。
- 10 2、如权利要求 1 所述的方法，其特征在于，将对应于所述一个或多个
段的段信息分配给所述两个或者更多个工作单元，每个工作单元根据被分配
到的段信息处理相应的段。
- 3、如权利要求 2 所述的方法，其特征在于，所述段信息包括用于指示
15 所述相应的段的开始记录和结束记录的指针。
- 4、如权利要求 1 所述的方法，其特征在于，所述工作状态包括对所述
预定数量的记录的处理结果。
- 20 5、如权利要求 4 所述的方法，其特征在于，所述工作状态进一步包括
发生异常的记录的指针。
- 6、如权利要求 5 所述的方法，其特征在于，当工作状态指示处理结果
为异常，从发生异常的记录的指针处开始恢复处理。
- 25 7、如权利要求 6 所述的方法，其特征在于，采用轮询模式监控工作状
态。
- 8、如权利要求 1 所述的方法，其特征在于，进一步包括向外部系统发
30 送关于工作状态的信息，并将来自外部系统的响应信息更新到工作状态中。
- 9、如权利要求 1 所述的方法，其特征在于，所述工作状态被存储在文

件中，以及所述两个或者更多个工作单元共同维护一个工作状态文件或者所述两个或者更多个工作单元分别维护各自的工作状态文件。

10、如权利要求 1 所述的方法，其特征在于，所述文件是参数文件，所述记录是参数，所述处理是将所述参数文件中的参数录入到系统的数据库中。

11、如权利要求 10 所述的方法，其特征在于，所述处理通过将所述参数直接装载到正式参数表来完成录入。

10

12、一种文件装载装置，其特征在于，包括，
预处理模块，用于根据文件中的记录数，将文件分成一个或多个段，
工作单元模块，包括两个或者更多个工作单元，用于并发处理所述一个或多个段，

15

异常处理模块，用于进行异常处理，
其中，所述两个或者更多个工作单元的每一个每次在处理完对应的段中的预定数量的记录后就更新各自的工作状态，所述工作状态用于为异常处理模块提供信息。

20

13、如权利要求 12 所述的装置，其特征在于，所述预处理模块将对应于所述一个或多个段的段信息分配给所述两个或者更多个工作单元，每个工作单元根据被分配到的段信息处理相应的段。

25

14、如权利要求 13 所述的装置，其特征在于，所述段信息包括用于指示所述相应的段的开始记录和结束记录的指针，所述工作状态包括对所述预定数量的记录的处理结果和发生异常情况的记录的指针，

所述异常处理模块被配置成监控工作单元的工作状态，当工作状态指示处理结果为异常时，从发生异常的记录开始恢复处理。

30

15、如权利要求 12 所述的装置，其特征在于，进一步包括文件反馈处理模块，用于向外部系统发送关于工作状态的信息，并将来自外部系统的响应信息更新到工作状态中。

16、如权利要求 12 所述的装置，其特征在于，进一步包括文件控制模块，用于对文件的预处理和工作单元的处理装载进行监控。

1/1

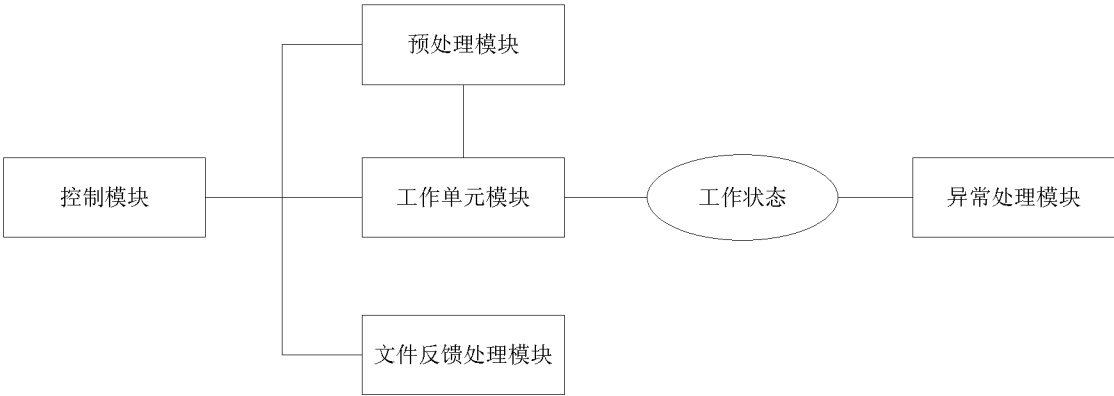


图 1

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/074426**A. CLASSIFICATION OF SUBJECT MATTER**

G06F 9/46 (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)

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)

DWPI, SIPOABS, CNABS, CNKI: file, segment, partition, split, upload, load, error, exception, parameter, database; input, pointer, address

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 101179392 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 14 May 2008 (14.05.2008), description, page 4, line 1 to page 8, line 3, and figures 1-4	1-16
Y	CN 101729586 A (CHINA MOBILE GROUP GUANGDONG CO., LTD. et al.), 09 June 2010 (09.06.2010), description, paragraphs [0030]-[0051], and figures 1 and 2	1-16
A	US 7593922 B1 (AT & T INTELLECTUAL PROPERTY, LL.P.), 22 September 2009 (22.09.2009), the whole document	1-16

☐ 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 06 June 2012 (06.06.2012)	Date of mailing of the international search report 05 July 2012 (05.07.2012)
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 YANG, Peng Telephone No.: (86-10) 62411658

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2012/074426

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101179392 A	14.05.2008	None	
CN 101729586 A	09.06.2010	None	
US 7593922 B1	22.09.2009	None	

国际检索报告

国际申请号

PCT/CN2012/074426

A. 主题的分类

G06F 9/46 (2006.01) i

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))

DWPI, SIPOABS, CNABS, CNKI: file, segment, partition, split, upload, load, error, exception, parameter, database;
文件, 参数, 分割, 分段, 划分, 录入, 装载, 载入, 上传, 数据库, 指针, 地址, 异常

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
Y	CN101179392A(腾讯科技(深圳)有限公司) 14.5 月 2008(14.05.2008) 说明书第 4 页第 11 行至第 8 页第 3 行, 图 1-4	1-16
Y	CN101729586A(中国移动通信集团广东有限公司等) 09.6 月 2010(09.06.2010) 说明书[0030]-[0051]段, 图 1,2	1-16
A	US7593922B1(AT&T INTELLECTUAL PROPERTY, I. L. P.) 22.9 月 2009(22.09.2009) 全文	1-16



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

06.6 月 2012(06.06.2012)

国际检索报告邮寄日期

05.7 月 2012 (05.07.2012)

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

授权官员

杨鹏

电话号码: (86-10) **62411658**

关于同族专利的信息

PCT/CN2012/074426