

(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 9 月 20 日 (20.09.2012)



(10) 国际公布号
WO 2012/122934 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2012/072285
- (22) 国际申请日: 2012 年 3 月 13 日 (13.03.2012)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201110060342.9 2011 年 3 月 14 日 (14.03.2011) CN
- (71) 申请人 (对除美国外的所有指定国): 北京小米科技有限责任公司 (BEIJING XIAOMI TECHNOLOGY CO., LTD) [CN/CN]; 中国北京市朝阳区望京西路甲 50 号-1 卷石天地大厦 A 座 12 层, Beijing 100102 (CN)。
- (72) 发明人; 及
- (75) 发明人/申请人 (仅对美国): 汪轩然 (WANG, Xuan-ran) [CN/CN]; 中国北京市朝阳区望京西路甲 50 号-1 卷石天地大厦 A 座 12 层, Beijing 100102 (CN)。范典 (FAN, Dian) [CN/CN]; 中国北京市朝阳区望京西路甲 50 号-1 卷石天地大厦 A 座 12 层, Beijing 100102 (CN)。屈恒 (QU, Heng) [CN/CN]; 中国北京

市朝阳区望京西路甲 50 号-1 卷石天地大厦 A 座 12 层, Beijing 100102 (CN)。洪锋 (HONG, Feng) [CN/CN]; 中国北京市朝阳区望京西路甲 50 号-1 卷石天地大厦 A 座 12 层, Beijing 100102 (CN)。黄江吉 (HUANG, Jiangji) [CN/CN]; 中国北京市朝阳区望京西路甲 50 号-1 卷石天地大厦 A 座 12 层, Beijing 100102 (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,

[见续页]

(54) Title: METHOD FOR REARRANGING WEB PAGE

(54) 发明名称: 一种网页重排版的方法

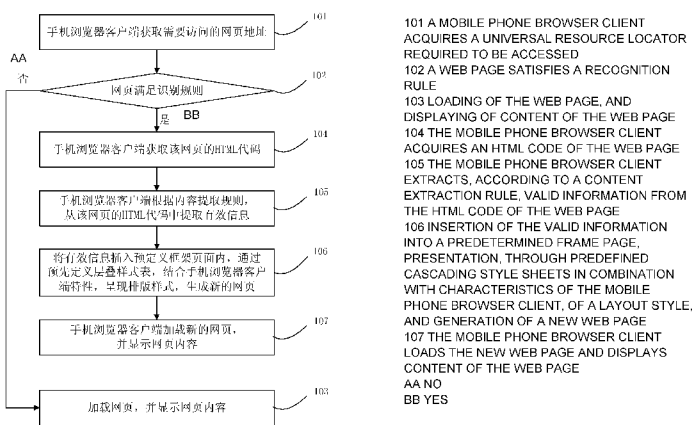


图1 / Fig.1

(57) Abstract: Disclosed is a method for rearranging a web page. The method comprises: a mobile phone browser client acquiring a universal resource locator; the mobile phone browser client determining whether a web page corresponding to the universal resource locator satisfies a recognition rule; if the recognition rule is not satisfied, loading the web page, and displaying content of the web page; and if the recognition rule is satisfied, the mobile phone browser client acquiring an HTML code of the web page (104); the mobile phone browser client extracting, according to a content extraction rule, an element comprising valid information from the HTML code of the web page, and extracting the valid information from the element; and the mobile phone browser client inserting the valid information of the web page into a predefined frame page, generating and loading a new web page, and displaying content of the web page. The foregoing method is completely adapted to the screen resolution of a device, maintains information and interaction of the original web page to the greatest extent, improves the speed for loading the web page, and saves network bandwidth.

[见续页]



WO 2012/122934 A1



RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, **本国际公布:**

CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, — 包括国际检索报告(条约第 21 条(3))。
TG)。

(57) 摘要:

公开了一种网页重排版的方法,该方法包括:手机浏览器客户端获取网页地址;手机浏览器客户端判断网页地址对应的网页是否满足识别规则,如果不满足,则加载网页,并显示网页内容,如果满足,手机浏览器客户端获取网页的 HTML 代码(104);手机浏览器客户端根据内容提取规则,从网页的 HTML 代码中提取包含有效信息的元素,再从元素中提取有效信息;手机浏览器客户端将网页的有效信息插入预定义的框架页面内,生成加载新的网页,并显示网页内容。采用了上述方法,能够完全适应设备的屏幕分辨率,最大程度的保留原网页的信息和交互,并且提高网页的加载速度,节省网络带宽。

说明书

一种网页重排版的方法

技术领域

5 本发明涉及移动互联网技术领域，尤其涉及一种网页重排版的方法。

背景技术

在移动互联网领域，如何将传统互联网上丰富的内容在移动设备上以用户友好的方式呈现是一个被广泛研究的课题。在这一课题中，如何让传统互联网
10 中大量为高分辨率屏幕设计的网页能显示在移动设备相对较低分辨率的屏幕上，并且不失原网页方便的浏览和交互，是其中非常重要的一个方向。

目前市面上的主流移动设备的浏览器都在这一方向上进行过一些尝试，微软公司在其早期的 Windows Mobile 系统中的 IE Mobile 系列浏览器，使用的策略是将网页中的所有元素纵向排列，以方便用户的阅读。谷歌在其 Android
15 系统的浏览器中，采用了文字自动换行技术，即在网页缩放时，重新对网页中的文字段落进行重排版，让文字根据当前缩放比例和屏幕的关系自动换行，保证在用户阅读时，不需要对网页进行左右滚屏操作。而苹果公司在 iPhone 系列手机的浏览器中，以及微软公司在其 Windows Phone 7 系统的浏览器中，则采取了 Text Scaling 技术，通过在网页首次渲染的时候对网页不同容器中的
20 字体大小进行不同的调整，保证在该容器被缩放到屏幕正中时，容器中的文字大小适合用户阅读并且不需要左右滚屏，这一技术成功避免了每次缩放都对网页进行重排版。

但这种技术的一个最大的缺点在于仅仅是提升了网页中的大段文字在移动设备上的阅读体验，而对于更多的网页元素如图片，视频，就无能为力了。
25 并且，这些技术在局部上破坏了网页的布局，容易造成网页的排版混乱，内容

重叠或大块空白等问题。

另一个研究方向则是引入服务器重排版技术，该技术最具代表性的成果是由优视动景公司（UCWEB 浏览器）发明的服务器缓存加速技术。这一技术通过在服务器对网页进行重新排版，使得网页的字体和宽度使用移动设备较低的分
5 辨率，并通过缓存重排后的网页，减少了和网站服务器的连接次数。

但是由于移动设备的多样性，不同的设备有着不同的分辨率，导致在服务器端重排的网页，对用户的设备屏幕的适配程度未能达到最优化。

对于一些涉及到用户隐私的网站（如电子商务网站，网络论坛），由于服务器重排技术要求客户端仅建立和重排缓存服务器的连接，因此用户的隐私内
10 容都需要保存到重排缓存服务器中，增加了隐私泄漏的风险。

由于网站的多样性，其重排的结果很难保证其版式的易用性和美观。

服务器重排技术需要投入大量的服务器资源，成本较高。

由于需要缓存重排后的网页，因此对于一些实时性要求很高的网页（如网络直播），处理的会比较滞后，丧失了网站本来实时刷新的特点。

15

发明内容

本发明的目的在于提出一种网页重排版的方法，能够完全适应设备的屏幕分辨率，带给用户非常好的浏览体验，同时最大程度的保留原网页的信息和交互，并且有效地过滤掉网页中的无关元素，提高网页加载速度，节省网络带宽。

20 为达此目的，本发明采用以下技术方案：

一种网页重排版的方法，包括以下步骤：

A、手机浏览器客户端获取网页地址；

B、手机浏览器客户端判断所述网页地址对应的网页是否满足识别规则，如果满足，则转至步骤 C，如果不满足，则加载所述网页，并显示网页内容；

25 C、手机浏览器客户端获取所述网页的 HTML 代码；

D、手机浏览器客户端根据内容提取规则，从所述网页的 HTML 代码中提取包含有效信息的元素，再从元素中提取有效信息；

E、手机浏览器客户端将网页的有效信息插入预定义的框架页面内，生成新的网页；

5 F、手机浏览器客户端加载新的网页，并显示网页内容。

所述识别规则包括网址规则、特殊元素规则和网页格式规则，其中网址规则通过正则表达式实现，特殊元素规则是通过查找网页中的符合条件的元素来判断当前网页是否符合要求，网页格式规则是根据网页元素的整体层次结构来判断是否符合要求。

10 特殊元素规则包括判断网页中的 body 元素的 id 是否为一个特殊字符；网页格式规则包括判断网页的 body 是否包含两个 div 元素。

所述内容提取规则通过 XPath 技术来实现。

所述内容提取规则包括新闻内容网页内容提取规则、小说阅读网页内容提取规则和论坛帖子列表网页内容提取规则。

15 有效信息包括内部 HTML 代码和超链接信息。

步骤 E 进一步包括以下步骤：

手机浏览器客户端将网页的有效信息插入预定义的框架页面内；

在框架页面中，通过预先定义层叠样式表，结合手机浏览器客户端的特性，呈现排版样式，生成新的网页。

20 手机浏览器客户端的特性包括分辨率和显示特征。

采用了本发明的技术方案，能够产生以下技术效果：

1、由于只针对某些符合特定规律的网站进行重排版，保证了排版后的网页有更好的浏览和交互体验。

25 2、通过对网页内容的重新排版，可以过滤掉原网站中的很多不相关内容

(如广告)，提升了阅读体验，并节省了网络带宽。

3、针对服务端缓存重排涉及到的隐私问题，采用的是完全的客户端重排，所有的数据交互只发生在客户端和网站服务器，不加入任何第三方服务器的干预，保护了用户的隐私。

5 4、针对网站服务器实时返回的内容进行重排版，保证呈现在用户设备中的内容是网站的实时内容。

附图说明

图1是本发明具体实施方式中网页重排版的流程图。

10

具体实施方式

下面结合附图并通过具体实施方式来进一步说明本发明的技术方案。

图 1 是本发明具体实施方式中网页重排版的流程图。如图 1 所示，该网页重排版的流程包括以下步骤：

15 步骤 101、手机浏览器客户端获取需要访问的网页地址。

步骤 102、手机浏览器客户端判断该网页地址对应的网页是否满足识别规则，如果满足，则转至步骤 104，如果不满足，则转至步骤 103。

该识别规则存储在手机浏览器客户端，包括网址规则、特殊元素规则和网页格式规则。

20 其中网址规则通过正则表达式实现。

特殊元素规则是通过查找网页中的符合条件的元素来判断当前网页是否符合要求，例如包括判断网页中的 body 元素的 id 是否为一个特殊字符。

网页格式规则是根据网页元素的整体层次结构来判断是否符合要求，例如包括判断网页的 body 是否包含两个 div 元素。

25 步骤 103、加载网页，并显示网页内容。

步骤 104、手机浏览器客户端获取该网页的 HTML 代码。

步骤 105、手机浏览器客户端根据内容提取规则，从该网页的 HTML 代码中提取包含有效信息的元素，再从元素中提取有效信息，该有效信息包括内部 HTML 代码和超链接信息。

5 该内容提取规则存储在手机浏览器客户端，包括新闻内容网页内容提取规则、小说阅读网页内容提取规则和论坛帖子列表网页内容提取规则，针对符合不同类型的网页，需要定义不同的规则。该内容提取规则描述的是某一个 HTML 元素或一组 HTML 元素的集合，一般通过 XPath 技术来实现。

10 步骤 106、手机浏览器客户端将网页的有效信息插入预定义的框架页面内，在框架页面中，通过预先定义层叠样式表，结合手机浏览器客户端的特性，呈现排版样式，生成新的网页。手机浏览器客户端的特性包括分辨率和显示特征等。

15 步骤 107、手机浏览器客户端加载新的网页，并显示网页内容。对于不同的网页类型，采取的框架页面和页面中包含的样式定义各不相同，而对于同一类型的网页，采用相同的框架页面和样式，因此重排版后的页面效果是一致的。

20 以上所述，仅为本发明较佳的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉该技术的人在本发明所揭露的技术范围内，可轻易想到的变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应该以权利要求的保护范围为准。

权 利 要 求 书

1、一种网页重排版的方法，其特征在于，包括以下步骤：

A、手机浏览器客户端获取网页地址；

5 B、手机浏览器客户端判断所述网页地址对应的网页是否满足识别规则，如果满足，则转至步骤 C，如果不满足，则加载所述网页，并显示网页内容；

C、手机浏览器客户端获取所述网页的 HTML 代码；

D、手机浏览器客户端根据内容提取规则，从所述网页的 HTML 代码中提取包含有效信息的元素，再从元素中提取有效信息；

10 E、手机浏览器客户端将网页的有效信息插入预定义的框架页面内，生成新的网页；

F、手机浏览器客户端加载新的网页，并显示网页内容。

2、根据权利要求 1 所述的一种网页重排版的方法，其特征在于，所述识别规则包括网址规则、特殊元素规则和网页格式规则，其中网址规则通过正则
15 表达式实现，特殊元素规则是通过查找网页中的符合条件的元素来判断当前网页是否符合要求，网页格式规则是根据网页元素的整体层次结构来判断是否符合要求。

3、根据权利要求 2 所述的一种网页重排版的方法，其特征在于，特殊元素规则包括判断网页中的 body 元素的 id 是否为一个特殊字符；网页格式规则
20 包括判断网页的 body 是否包含两个 div 元素。

4、根据权利要求 1 所述的一种网页重排版的方法，其特征在于，所述内容提取规则通过 XPath 技术来实现。

5、根据权利要求 4 所述的一种网页重排版的方法，其特征在于，所述内容提取规则包括新闻内容网页内容提取规则、小说阅读网页内容提取规则和论坛帖子列表网页内容提取规则。
25

6、根据权利要求 1 所述的一种网页重排版的方法，其特征在于，有效信息包括内部 HTML 代码和超链接信息。

7、根据权利要求 1 所述的一种网页重排版的方法，其特征在于，步骤 E 进一步包括以下步骤：

5 手机浏览器客户端将网页的有效信息插入预定义的框架页面内；

在框架页面中，通过预先定义层叠样式表，结合手机浏览器客户端的特性，呈现排版样式，生成新的网页。

8、根据权利要求 7 所述的一种网页重排版的方法，其特征在于，手机浏览器客户端的特性包括分辨率和显示特征。

说明书附图

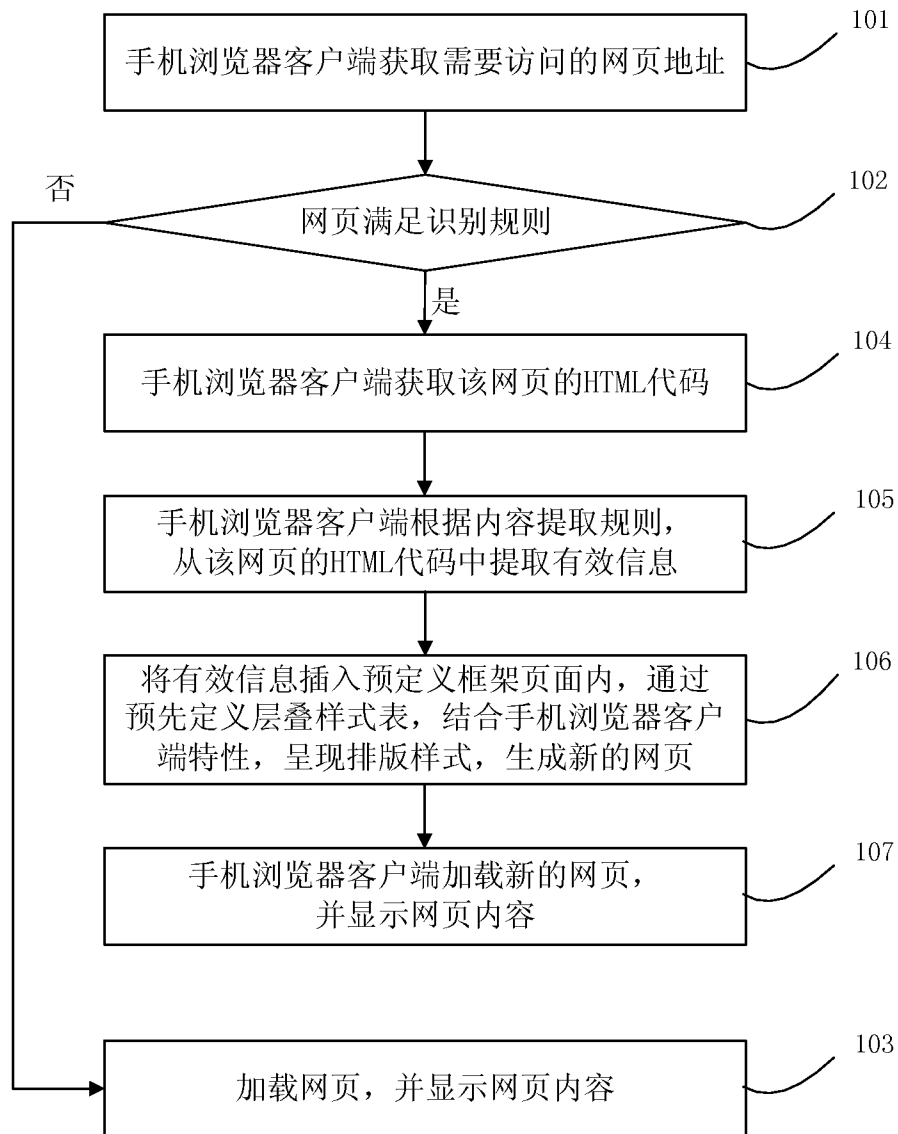


图 1

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/072285

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (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, H04L

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, CNPAT, VEN: browser, mobile phone, terminal, typesetting, rearrangement, display, block diagram, format, hypertext, hyperlink, element, preset, rule, HTML, page, webpage, mobile, phone, client, portable, frame, block, layout

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 101202748 A (YOUSHI DONGJING (BEIJING) TECHNOLOGY SERVICE CO., LTD.), 18 June 2008 (18.06.2008), claims 1-11, and description, page 7, line 17 to page 8	1-8
Y	CN 101859322 A (ASPIRE TECHNOLOGIES (SHENZHEN) CO., LTD.), 13 October 2010 (13.10.2010), claims 1-10, and description, paragraphs 20-78	1-8
A	CN 101894168 A (UC MOBILE LTD.), 24 November 2010 (24.11.2010), the whole document	1-8
A	CN 101146128 A (YANG, Jinyu), 19 March 2008 (19.03.2008), the whole document	1-8
A	CN 100392641 C (BEIJING FOUNDER ELECTRONICS CO., LTD. et al.), 04 June 2008 (04.06.2008), the whole document	1-8

☐ 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 18 May 2012 (18.05.2012)	Date of mailing of the international search report 14 June 2012 (14.06.2012)
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 LIU, Li Telephone No.: (86-10) 62411687

International application No.
PCT/CN2012/072285

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号
PCT/CN2012/072285

A. 主题的分类

G06F 17/30 (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, H04L

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, CNPAT, VEN: 网页, 页面, 浏览器, 手机, 终端, 客户端, 排版, 重排, 布局, 显示, 版面, 框架, 框图, 格式, 超文本, 超链接, 元素, 预置, 预设, 预定, 规则, HTML, page, webpage, mobile, phone, client, portable, frame, block, layout

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	CN 101202748 A (优视动景(北京)技术服务有限公司) 18.6 月 2008(18.06.2008) 权利要求 1-11, 说明书第 7 页第 17 行-第 8 页	1-8
Y	CN 101859322 A (卓望数码技术(深圳)有限公司) 13.10 月 2010(13.10.2010) 权利要求 1-10, 说明书第 20 段-第 78 段	1-8
A	CN 101894168 A (优视科技有限公司) 24.11 月 2010(24.11.2010) 全文	1-8
A	CN 101146128 A (杨金钰) 19.3 月 2008(19.03.2008) 全文	1-8
A	CN 100392641 C (北京北大方正电子有限公司等) 04.6 月 2008(04.06.2008) 全文	1-8

☐ 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
18.5 月 2012(18.05.2012)

国际检索报告邮寄日期
14.6 月 2012 (14.06.2012)

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

受权官员

刘力
电话号码: (86-10) **62411687**

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2012/072285

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN 101202748 A	18.06.2008	无	
CN 101859322 A	13.10.2010	无	
CN 101894168 A	24.11.2010	无	
CN 101146128 A	19.03.2008	无	
CN 100392641 C	04.06.2008	CN 1908930 A	07.02.2007