

## 世界知识产权组织标准委员会(CWS)

### 第二届会议

2012年4月30日至5月4日，日内瓦

### 关于 WIPO 标准 ST.22 执行和推广情况的调查

秘书处编拟

1. 原标准与文献工作组（SDWG）在 2009 年 10 月举行的第十一届会议上，批准了一份关于进行一次“WIPO 标准 ST.22 执行和推广情况调查”的问卷（见文件 SCIT/SDWG/11/14 的第 49 段和第 50 段以及文件 CWS/1/10 Prov.第 37 号任务第 52 段。）。
2. 根据 SDWG 的决定（见文件 SCIT/SDWG/11/14 第 51 段），调查于 2011 年进行。国际局开发了调查问卷的网上版本，在 WIPO 标准管理数据库（WIPOSTAD）中提供，并于 2011 年 6 月 29 日用 C.CWS 15 号通函通知了各局。
3. 三十个工业产权局参加了本次调查，对问卷提供了答复。收到的答复由秘书处归纳为调查报告的形式，现转录于本文件的附件。
4. 与本次调查有关的资料：问卷、各局答复以及综合结果可见 WIPOSTAD：[http://www.wipo.int/wipostad/en/surveys/ipo\\_practice/](http://www.wipo.int/wipostad/en/surveys/ipo_practice/)，在“OCR practices”项目下。
5. 请标准委员会：
  - (a) 注意附件中所载的信息；
  - (b) 审议并批准在 WIPOSTAD 中发布调查报告。

[后接附件]

## 调查结果

### 导 言

1. 本调查是在 2008 年 11 月通过的 WIPO 标准 ST.22 修订框架内，根据 2009 年 10 月标准与文献工作组（SDWG）批准的问卷于 2011 年进行的（见文件 CWS/1/9 附件第 37 号任务和 CWS/1/10 Prov.第 52 段）。

2. 问卷中的问题涉及 WIPO 标准 ST.22（关于以便于光学字符识别（OCR）的方式制作专利申请的建议）和用纸件提交或者用电子方式提交但申请中文本部分采用图像形式（如 PDF 或 TIFF 图像）的专利申请，以及工业产权局 OCR 做法问题。

3. 下列三十个国家局参与了本次调查：

AT	奥地利	IS	冰岛
AU	澳大利亚	IT	意大利
BR	巴西	JP	日本
BY	白俄罗斯	KR	大韩民国
CN	中国	KZ	哈萨克斯坦
CR	哥斯达黎加	LT	立陶宛
CZ	捷克共和国	MD	摩尔多瓦共和国
DE	德国	PL	波兰
DK	丹麦	RU	俄罗斯联邦
ES	西班牙	SA	沙特阿拉伯
GB	联合王国	SE	瑞典
HR	克罗地亚	SK	斯洛伐克
HU	匈牙利	UA	乌克兰
IE	爱尔兰	US	美利坚合众国
IL	以色列	WO	世界知识产权组织（WIPO） （国际局）

4. 本报告按照问卷的各个部分，分别对答复做了汇总。整个文件中，为缩略、澄清和/或协调起见，改写了原各项答复中的某些评论意见。任何对原意见的偏离都不是有意为之。

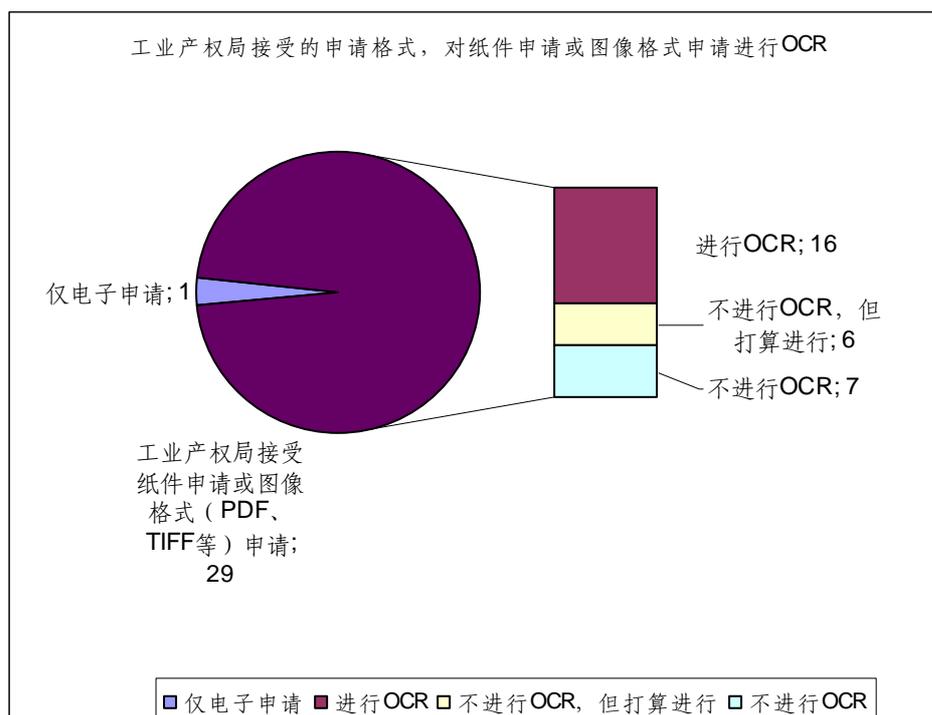
5. 各工业产权局的答复均以其原来的语言（答复所用语言）与自动汇总的结果一同在 WIPOSTAD 中公布。

## 专利申请

6. 前三个问题构成问卷的第一部分；这些问题的重点是工业产权局接受专利申请的程序、统计数据 and 格式。对这些问题的答复率是 100%（30 个局）。

7. 绝大多数工业产权局（30 个局中的 29 个）报告说接受用纸件提交或者用电子方式提交但申请中文本部分采用图像形式提交的专利申请；16 个局对专利申请进行 OCR，6 个局打算将来这样做。一个局（KR）报告说，不接受纸件申请或者图像形式的申请，但在答复中表示本局也进行 OCR。

8. 下图显示了接受纸件和图像形式专利申请的工业产权局关于 OCR 做法的答复分布情况。



选项	答复
进行 OCR	AT、AU、BY、CN、CZ、DE、ES、GB、HR、HU、KZ、PL、RU、SE、UA、WO（16）
不进行 OCR，但打算进行	BR、CR、DK、IL、LT、SK（6）
不进行 OCR	IE、IS、IT、JP、MD、SA、US（7）
仅电子申请	KR（1）

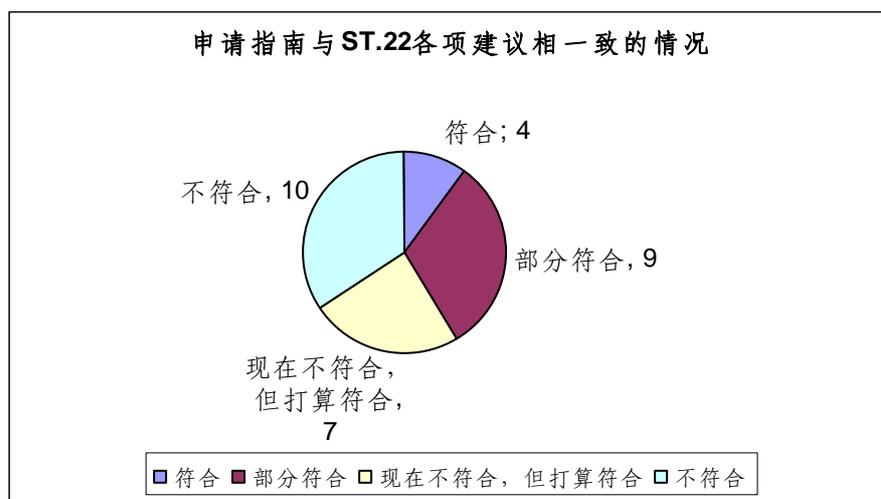
9. 按各局的报告，纸件申请的比例分布很大，从 2.4%（JP）到 100%（LT、BR、MD、BY、SA 和 CR）不等，“图像格式”申请的比例从 0.01%（UA）到 90%（US、DK）不等。同时，应当指出，对大多数做出答复的局（30 个局中的 23 个，77%），这种申请（纸件申请或者图像格式申请）的总比例是 100%，这意味着对于这些局，接受的所有申请均可被认为是可能的 OCR 对象。

10. 根据报告，57%的答复者（30 个局中的 17 个）对专利申请进行 OCR，其他一些（30 个局中的 6 个）打算将来这样做，其余 7 个局（23%）不进行 OCR，也没有这样做的计划。

### WIPO标准ST.22 的推广和使用

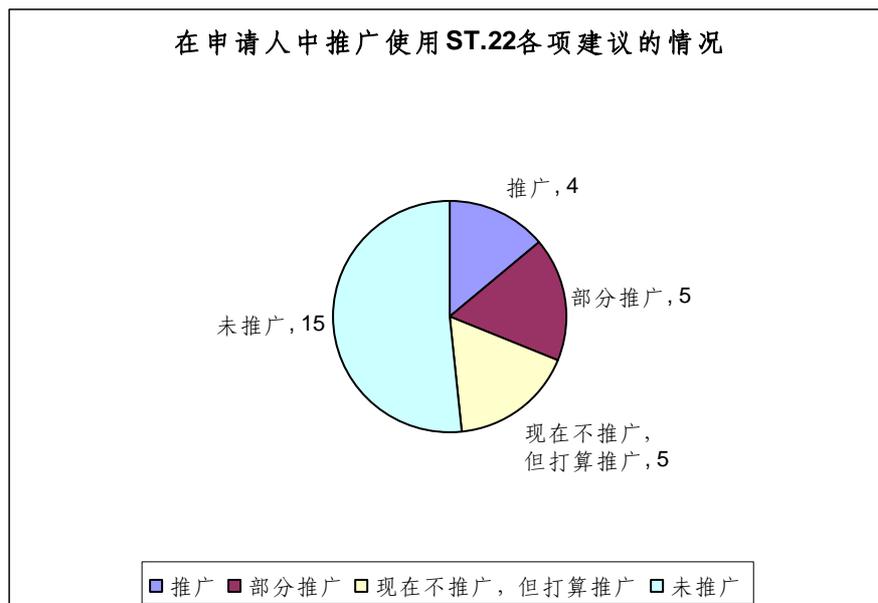
11. 接下来四个问题（问题 4 至 7）构成的部分涉及提高公众对 WIPO 标准 ST.22 各项建议的认识以及鼓励申请人遵守这些建议。这些问题的回答率为 100%（30 个局）

12. 三分之二的局答复说，它们已经修改申请指南，完全符合（4 个局）或者部分符合（9 个局）WIPO 标准 ST.22 修订后的各项建议，或者计划在将来这样做（7 个局）。答复的统计和分布见下列图表：



选项	答复
符合	AU、BY、GB、WO（4）
部分符合	AT、DE、ES、HU、KR、LT、PL、RU、SA（9）
不符合, 但打算符合	BR、HR、IL、IS、IT、SK、UA（7）
不符合	CN、CR、CZ、DK、IE、JP、KZ、MD、SE、US（10）

13. 50%的局（30 个局中的 15 个）反映说，它们已经在申请人中全部（5 个局）或部分（5 个局）或者将来计划（5 个局）推广使用 ST.22 各项建议。另一半（30 个局中的 15 个）无此计划。答复的统计和分布见下列图表。



选项	答复
推广	AT、AU、BY、GB、WO (5)
部分推广	ES、KR、LT、PL、RU (5)
现在不推广, 但打算推广	DK、HR、IL、SK、UA (5)
未推广	BR、CN、CR、CZ、DE、HU、IE、IS、IT、JP、KZ、MD、SA、SE、US (15)

14. 一些局 (AT、AU、BY、GB、PL、RU 和 SK) 提供了 ST.22 修订通知的 URL 例子。

15. 除了修改申请指南或相应的规章并予以发布外, 各局还使用下列手段推广 ST.22: ST.22 网站链接、将标准或者标准的一部分译成本国语言、信息通知、在知识产权杂志上发布、培训课程、咨询服务等。

#### WIPO标准ST.22 的执行情况

16. 接下来六个问题 (问题 8 至 13) 涉及工业产权局执行 ST.22 的做法。这些问题的对象是那些表示已完全或部分向申请人推广使用 ST.22 的答复者 (10 个局, 33%)。6 个工业产权局答复了这些问题, 即 AT、BY、GB、ES、RU 和 WO。AU 局认为现在评估进展为时尚早, 因为相应的指南仅施行一个月 (到 2011 年 7 月)。

17. 多数答复者注意到根据 ST.22 建议提交的申请在形式表现和布局方面的质量有所提高。因此, BY、ES 和 RU 指出了明显提高。RU 局认为提高的原因是普遍的申请人计算机化, 不是新规章的施行。AT 和 GB 没注意到大的提高。AT 局指出, 新规章让申请人不能为申请的文本部分使用太小的字体; WO 报告说未注意到提高, 因为未进行衡量。

18. 在因申请人对 WIPO 标准 ST.22 的认识而导致 OCR 质量和费用方面出现变化的问题上, 不能得出任何相关结论。一个答复者 (BY) 指出 OCR 质量明显提高, 而两个答复者 (GB 和 ES)

没注意到大的提高，做出答复的局中有一半（AT、RU 和 WO）没有注意到任何提高。但是，RU 局和 WO 局表示未收集相应统计数据，AT 局表示有关申请尚未进行 OCR。BY 局指出工作时间方面的节约使 OCR 费用明显减少，ES 局没有注意到 OCR 费用减少很多。其他四个局（AT、GB、RU 和 WO）报告说这方面没有注意到减少。

19. 由于申请不符合 ST.22 而要求提交替换页的做法存在不同：一些答复者（7 个局中的 4 个，57%）表示有这样的要求，其他表示没有。根据收到的意见，一般做法是根据国家或者 PCT 规章要求替换页，而 ST.22 各项建议被全部或部分收入这些规章。

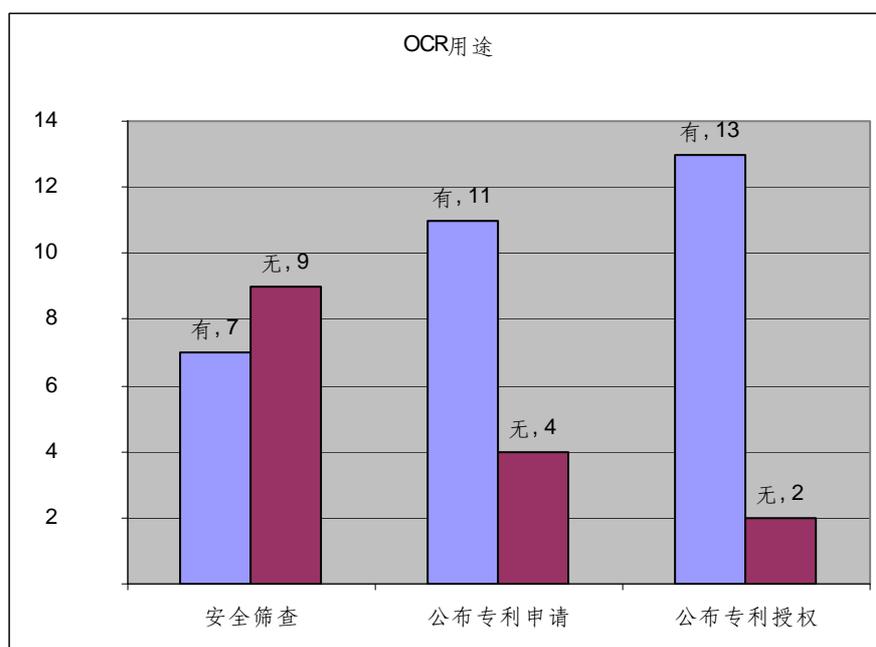
20. 三个局通报了被要求提交替换页的申请所占的大概比例。BY 局对 15%的专利申请要求替换页，GB 局的这一比例为 10-15%，RU 局在 2010 年下半年和 2011 年上半年为 3-5%。

21. 没有局表示打算根据所提交申请符合 ST.22 的程度来计算费用；但 GB 局认为该设想令人感兴趣。

### 工业产权局的OCR做法

22. 后面五个问题（问题 14 至 18）涉及工业产权局采用的 OCR 做法的一般方面，如 OCR 的各种用途、外包、准确率要求和质量检查手段。

23. 16 个局回答了关于 OCR 用途的问题或者其中一些次级问题，答复的汇总见下列图表。



选项	安全筛查	公布专利申请	公布专利授权
有	BY、CN、DE、HU、KR、RU、UA (7)	AT、AU、CN、GB、HU、HR、KR、RU、SE、UA、WO (11)	AT、AU、BY、CN、CZ、HR、HU、KR、KZ、PL、RU、SE、UA (13)
无	AT、AU、GB、CZ、HR、KZ、PL、SE、WO (9)	BY、CZ、KZ、PL (4)	GB、WO (2)

24. 一些局通报了对不同 OCR 用途适用的准确率要求。这方面，在专利申请的安全筛查方面，CN 局的准确率为 99.99%。用于公布专利申请时，从 99% (AU) 到 99.99% (CN) 不等。GB 局抓取摘要的准确率为 99.85%，100% 经过人工检查。WO 局表示，现阶段确保 OCR 的检索质量，即 99.5%，WO 局的公布质量为 99.95%。AT 和 RU 局现阶段没有准确率检查。AU 局对专利授权的准确率要求为 99%。在 AT 局和 PL 局，100% 的文件经过人工检查，RU 局没有规定的准确率要求。

25. 关于 OCR 的其他用途，答复者指出了开发检索数据库 (BY、RU)，以及将摘要文本载入内部信件撰写系统，在公布和传送至 EPO 之前由技术专利审查员进行编辑 (GB)。在 SA 局，OCR 用于专利申请的内部注释。

26. 超过 50% 的答复者 (28 个局中的 16 个) 表示，没有采用质量检查措施来控制经过 OCR 的专利文件的质量。同时，相当数量的答复者 (28 个局中的 12 个) 报告说有这样的措施。这些措施包括比照电子卷宗中的图像或纸件对 OCR 文本进行人工检查 (ES、GB、PL、RU、SE 和 UA)，用于检查和设定统一格式的 MS Word 宏，拼写检查，或者当混合排列时检查拉丁字符和西里尔字符的错误识别 (RU)，随机抽取文件进行横向和纵向文字检查、文本检查和标签检查 (CN)，软件报告的 OCR 信任度 (WO)。

27. CN 局采用的“半自动化”质量检查程序在下节“OCR 用软件和硬件”中概述 (见第 32 段)。

28. 下列 8 个局 (28 个局答复了该问题) 对外文专利文件进行了 OCR:

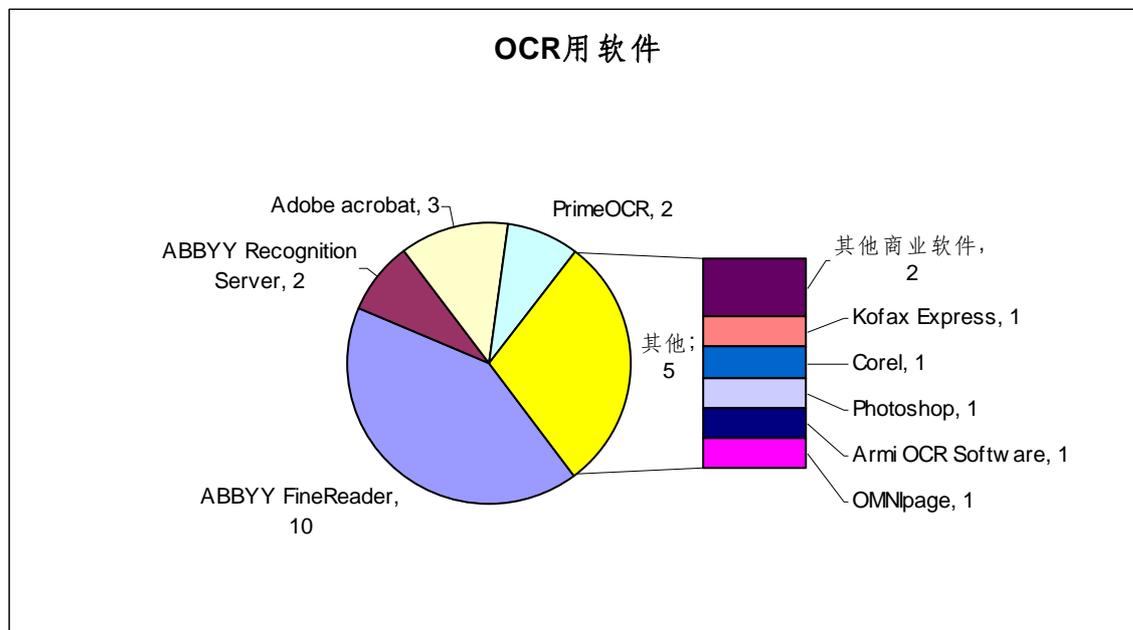
局	OCR 的外文文件
DE	未说明具体语言
HU	英语
KR	英语
RU	英语
SE	英语、法语和德语的在瑞典有效力的欧洲专利文件
UA	英语、德语、法语、西班牙语、意大利语、希腊语
US	韩语、瑞典语、德语、中文、法语、意大利语、西班牙语、日语和葡萄牙语
WO	英语、法语、德语、西班牙语、葡萄牙语、韩语、中文、日语和俄语

29. 超过 70% 的答复者 (28 个局的 20 个, 71%) 不外包专利文件的 OCR，其他局在程序的不同阶段进行 OCR: 文件一经收到 (CN、JP)，授权前阶段 (CZ、US) 以及授权阶段 (US)，公布前 (ES) 等。

#### OCR用软件和硬件

30. 接下来三个问题 (问题 19 至 21) 涉及工业产权局或其承包商在执行专利文件 OCR 时使用的设备和软件。这些问题的答复率为 83% (25 个局)

31. 如下图所示，ABBYY FineReader、Adobe acrobat、ABBYY Recognition server 和 PrimeOCR 是答复问卷的各局使用最多的软件（进一步详情见汇总结果）。四个局表示，对所用的软件开发了具体扩展，用于文件自动化处理（UA）、提高用户友好性（JP）以及执行 XML 输出（WO）。



32. CN 局通报说，对使用不同软件的 OCR 结果进行比较，如果发现任何差异，文件将转入人工检查。

33. GB 局报告说，使用 Madras-Phoenix（电子案件管理系统）的打印功能获得 PDF 格式的摘要图像，然后用 OMNIpage 打开并保存，作为文本载入内部专用信件撰写系统，由内部技术审查员酌情修改。该文本之后用于公布程序。

34. SA 局不对专利文件进行 OCR，因为申请人随所提交的申请附上申请的软拷贝。该局表示，计划未来转至电子申请。

35. 工业产权局及其承包商在对专利文件进行 OCR 时使用的硬件见下表：

局	硬 件
AT	Windows XP 客户端
RU	PC HP 3GHz 用柯达 i620 扫描仪和富士通 fi-5750C 扫描的文件进入 OCR 程序。
CN	国知局在专利文件的 OCR 中使用标准 PC。
UA	富士通图像扫描仪（如 5530C）
US	美国专商局硬件
JP	Windows PC
SE	Windows server、Windows Vista

局	硬 件
HU	富士通 FI-6670/6770A 扫描仪
KZ	扫描仪
HR	PC、扫描仪
WO	Linux PC 服务器
AU	扫描仪
BY	惠普 scanjet 5590、惠普 scanjet 自动文件送入器
PL	扫描仪：Microtec S400、富士通 fi-6230、富士通 fi-5900C、Microtec I900
SA	该局拥有专业扫描仪，可用于执行专利文件的 OCR。这些扫描仪被用于把所有申请扫描为 TIFF 格式，用于公布。
KR	惠普 DL580 G5 (P4 Xeon)
ES	申请服务器

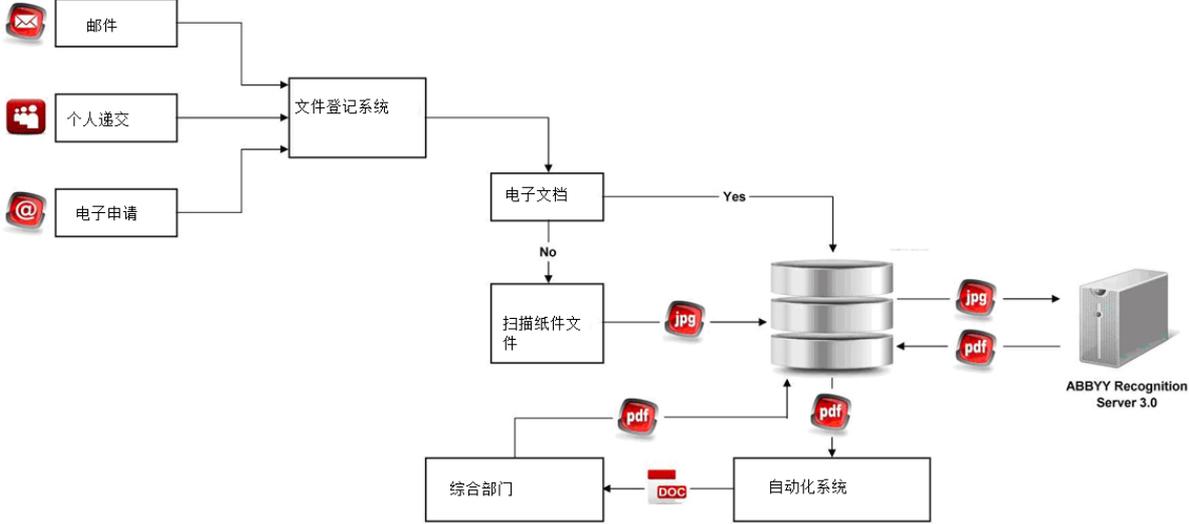
#### 工作流程

36. 后八个问题（问题 22 至 29）涉及专利文件 OCR 的流程、存储、后续更正、OCR 与其他文件处理组件的关联、其他局对已 OCR 文件的使用以及客户对已 OCR 文件的用途。这些问题的答复率为 60%（18 个局）。

37. 15 个局对专利文件 OCR 使用的流程提供了说明（见下表）。

局	OCR 工作流程
AT	<ol style="list-style-type: none"> <li>1. 审查员确定要公布的页。</li> <li>2. 对指定的页进行扫描。</li> <li>3. OCR+拼写检查+MS Word 人工调整格式。</li> <li>4. 格式调整后的文本与原纸页比较。</li> <li>5. 修改错误。</li> <li>6. MS-Word 版本转成 PDF。</li> <li>7. 生成的 PDF 与首页 PDF（已另外做好）和（针对实用新型）检索报告的 PDF 合并。</li> </ol>
BR	公布和编制索引后，2006 年 8 月 1 日以来公布的所有专利文件传送至 WIPO 的 PATENTSCOPE，协议涉及这种文献的 OCR，即将开始实施。
BY	初步审查结束后对权利要求进行扫描；著录项目数据不扫描；对专利申请的其他部分进行扫描，用于官方公布。
CN	OCR 工作流程包括八个程序：扫描，识别，垂直文字检查，水平文字检查，文本检查，标签，标签检查和质量检查。

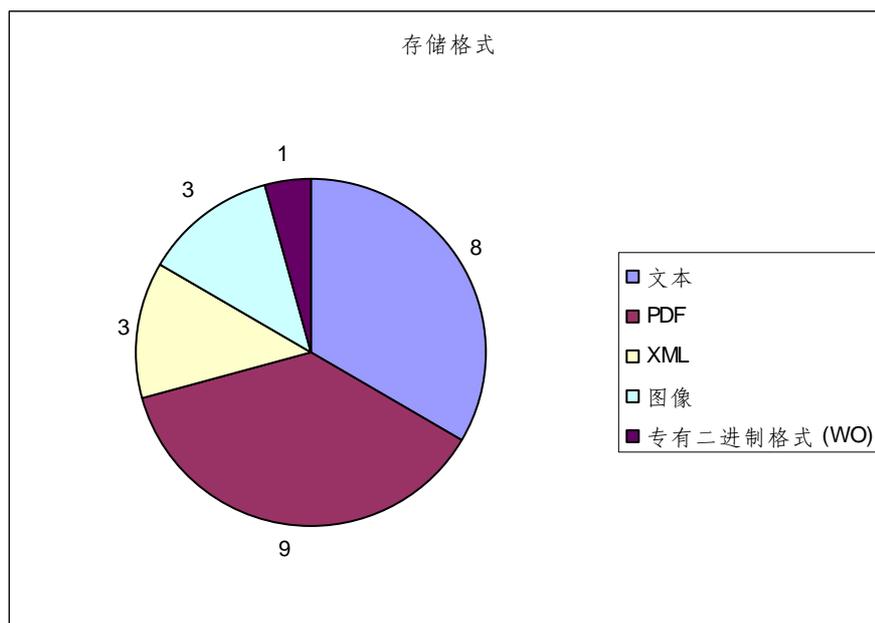
局	OCR 工作流程
ES	该局扫描以纸件提交的文件并通过 FTP 以电子格式传输给承包商进行 OCR。
GB	<p>申请人要求检索时（申请之日起 12 个月内）IPO OCR（获取和转化）摘要文本。如果申请进行到公布阶段，该摘要文本（经过技术审查员修改）用于公布程序（装入 EPOQUE）。</p> <p>公布后：欧洲专利局（申请日起 19-20 个月）通过第三方协议将 GB 全文（说明书和权利要求）装入 EPOQUE 和 Worldwide Esp@cenet 数据库。</p>
HR	<ol style="list-style-type: none"> <li>1. 扫描文件。</li> <li>2. 生成 PDF 文件。</li> <li>3. 将 PDF 文件输入 OCR 软件（FineReader）。</li> <li>4. 对文件的所有部分进行标记（文本、表格和图像）。</li> <li>5. 阅读已标记的块。</li> <li>6. 保存 doc 文档。</li> <li>7. 在 MS Word 中检查。</li> </ol>
HU	扫描文件、储存文件、每晚自动批量 OCR。
JP	通过扫描仪将申请文件转为图像格式，通过 OCR 软件转成文本数据。
KR	文件经过扫描和 OCR，对转化后的文本进行修改。
PL	<ol style="list-style-type: none"> <li>1. 对原文件进行扫描和 OCR（所用软件：ABBYY Fine Reader 9.0 Professional Edition）。</li> <li>2. 通过本局工作人员进行 OCR 文本的初步检查。</li> <li>3. 使用规定模板，基于 OCR 文本、著录项目数据、图像和权利要求创建文件（所用软件：MS Word）。</li> <li>4. 保存为 DOC 格式。</li> <li>5. 专利局工作人员对 DOC 文件和纸件文件进行比较——DOC 文件最终修订。</li> <li>6. 用预定义脚本将 DOC 文件转成可检索的 PDF（所用软件：Microsoft Word、Adobe Acrobat Professional）。</li> <li>7. 在专利局官网数据库和公布服务器上公布。</li> </ol>
RU	<p>用软件 ABBYY Fine Reader 9.0 进行 OCR。</p> <ol style="list-style-type: none"> <li>1. OCR 程序开始前，将扫描文件分成块：文本、表格或图像。</li> <li>2. 操作员校对 OCR 文本。</li> <li>3. 用 MS Word 将页保存为 RTF 格式。</li> <li>4. 根据申请组成部分所属的类型对每个文件进行命名——摘要、说明书、权利要求书。</li> <li>5. 文本用 MS Word 格式化。</li> </ol> <p>数学和物理公式作为 Microsoft 公式编辑器图形对象放入文本。化学公式作为 ISIS Draw 软件对象放入。</p>

局	OCR 工作流程
SE	每天对新的专利摘要进行 OCR，随后进行人工检查。每周一次对已公布专利申请和专利文件进行自动 OCR。
UA	<p>UA 局的专利文件 OCR 工作流程</p> 
WO	<ol style="list-style-type: none"> <li>1. 批量自动 OCR。</li> <li>2. 人工对 OCR 软件字符识别信任水平确认为较差的文本进行校对。</li> <li>3. 将 OCR 以 XML 和 HTML 输出。</li> </ol>

38. 60%的局（25 个局中的 15 个）答复说对已 OCR 的专利文件进行质量检查。检查由局雇员人工进行，要么为全面检查（AT、GB、PL、RU 和 UA），要么为部分文件检查（SE 和 WO）。

39. 大多数局通报说，文件公布后发现有误时，它们公布更正或者再次公布该文件，80%（20 个局中的 16 个）在答复中明确说明这一点。其中 6 个（AT、BY、CN、GB、HR 和 SK）表示根据相应 WIPO 标准（ST.50、ST.16 和 ST.9）的建议进行处理。5 个局（CN、CR、HU、KR 和 US）报告说根据申请人请求进行更正。GB 局说，OCR 获取程序中出现的错误报告给欧洲专利局，以在 EPO 的数据库中进行更正。CN 局通报说，发现有误的文件发回承包方重新处理，WO 局表示用外部承包商的服务改进往期已公布文件的 OCR。

40. 本节中的一个问题，即问题 25，具体涉及 OCR 后专利文件的储存问题。25 局描述了它们的做法，回答了该问题或者该问题的一些次级问题。21 个工业产权局提供了关于 OCR 文件存储格式的信息（见下列图表）。调查显示，使用最多的两种格式是 PDF 和文本（主要是 MS Word），使用的其他格式是图像（主要是 TIFF）和 XML。



局	使用的格式		
AT	文本 (MS Word)	PDF (带书签)	
AU		PDF	
BY	文本 (MS Word (rtf))	PDF	图像 (TIFF)
CN	XML		
CR			图像
CZ	文本 (MS Word (doc))		
ES		PDF	
GB	文本 (MS Word、HTML)		
HR	文本 (MS Word (doc))		
HU		PDF	
JP	XML (符合 ST.36)		
KZ		PDF	
LT		PDF	
MD	文本 (MS Word)		
PL		PDF	
RU	文本 (MS Word (doc、rtf))		图像 (TIFF)
SE	文本		

局	使用的格式
UA	PDF
US	XML
WO	WIPO 使用专有二进制格式，包含 OCR 程序产生的所有信息。

41. 超过 50%的答复者（22 个局中的 13 个）称它们使用的存储格式允许以后对已 OCR 专利文件进行质量改善。进一步细节见 AT、GB 和 JP 的单独回复。

42. 就能否快速鉴别有 OCR 缺陷的专利文件而言，67%的答复者（21 个局中的 14 个）回答说，所使用的存储格式无法进行这种识别。三分之一的答复者（BY、GB、HU、KR、KZ、RU 和 WO）说实践中存在这种可能性（进一步信息见 GB 和 HU 的单独回复）。JP 局称，特许厅的做法能保证防止出现有缺陷的专利文件。

43. 在答复能否以不同的转换形式浏览或交换专利文件时，60%的局（20 个局中的 12 个）报告说使用的存储格式允许如此。

44. 22 个局中的 9 个（41%）回复说它们保存从 OCR 流程中获得的详细、完整的原始信息；其他 59%的局对该问题作了否定回答。在 AT 局，“Fine reader document”被保存一段时间，以检查拼写或者再次进行 OCR（见单独回复）。CN 局保存复杂工作单元，例如数学和化学公式，在页面上的位置信息。

45. 关于将表格内容和数学、化学公式获取为文本格式的问题，18 个局做了回答。其中 5 个局（BY、KR、RU、UA、US）作了肯定回答，8 个局（AU、CN、CR、ES、HU、JP、PL、WO）作了否定回答，5 个局（AT、BR、HR、SA、SE）没有提供明确答案。因此，由于工业产权局的做法不同，从调查结果中不能对表格内容和数学、化学公式获取为文本格式确定总体趋势。此外，对于一个局，该问题根据不同的条件可能有不同的回答，例如 OCR 对象的复杂性（AT），或者能否以文本格式获取上下文（SE）。

46. 一半的答复者（24 个局中的 12 个）表示专利文件的 OCR 提高了工作效率。答复指出，OCR 可以便于安全筛查、编写准备和审查员重写摘要（CN），专利检索（US），将可检索的专利文件迅速装入局数据库（PL）和快速查阅这些文件（IL），审查流程（BY、SE），也有助于翻译（摘要和报告）并用于完成全文本公布（WO）。

47. 26 个局中的 11 个回复说，不仅是专利文件，它们也 OCR 其他文件，例如保护不同工业产权权利类型（实用新型、商标的商品和服务单、显著性标志等）的文件，异议的无效文本（AT），内部用非专利文献（RU、SE），修改文件和申请人的意见（CN），技术转让合同（BR），来往信件（MD）。

48. 根据收到的回复，OCR 专利文件可以由工业产权局审查员和社会公众用于专利检索。主要公布方式是主管局的网站和公布服务器，局方制作的电子产品，专利信息提供方（如回复中提到的 EPOQUE 和 Espacenet）的数据库，或者 CD-ROM（MIMOSA）上。关于各局经验的详细信息可见汇总报告。

49. 26 个答复局中的 11 个局（42%）使用其他局提供的 OCR 专利文件。最“流行”的资源是 Patentscope 和欧洲专利局数据库，另外提到的有：进入国家阶段的 PCT 文件（AU），谷歌专利和 DEPATISNet（PL），ES 局提供的文件（CR）。WO 局进行“反向流程”：国际局对墨西哥、南非、摩洛哥、以色列、巴西、巴拿马、古巴、西班牙（非常老的文件）、多米尼加共和国、非洲地区知识产权组织和肯尼亚的文件进行 OCR。

## 评论和结论

50. 调查确认，全球各工业产权局接受的文件中，大多数以纸件或图像格式提交，同时对公布专利文件和专利信息数据库实行的质量要求非常高。因此，OCR 作为把收到的申请人“原始”信息转化为主管局或其他专利信息提供方发布的格式中主要步骤，其质量仍是一个非常重要的问题，这种重要性将来也不会降低。

51. WIPO 标准 ST.22 的目标是在自动读取专利申请文本时实现尽可能低的错误率，同时仍然允许人对文件的高效阅读，这项工具旨在帮助把专利申请准备为打印形式，适于后续制作专利申请的电子数字化记录。调查显示，ST.22 的各项建议直接或间接得到大多数工业产权局的采用，或者计划在未来采用，以确保纸件申请或者以图像格式提交的申请的质量足以在以后用 OCR 技术转换成存储和公布格式。

52. 综上所述，可以得出结论，现行 WIPO 标准 ST.22 实现了其目的，为制作适于后续 OCR 的专利申请提供必要的指导。因此，该标准目前无需修订。

53. 请各工业产权局把本次调查的结果作为了解其他局采用的 OCR 做法的信息资源。为确保专利信息服务的必要质量，不妨考虑各局在所用存储格式以及与申请人和外部承包商的互动方面采取的不同方法。

[ 附件和文件完 ]