

专利合作条约（PCT）工作组

第十五届会议

2022年10月3日至7日，日内瓦

以全文本格式处理国际申请

国际局编拟的文件

概 述

1. 本文件提供了 PCT 工作组第十四届会议上提出的转为全文本处理国际申请的提案（见文件 PCT/WG/14/8）的最新情况。特别是，本文件讨论了全文本处理的原则和要求，涵盖 DOCX 格式国际申请的准备，国际申请后续修改的提交、处理和标注，以及原始文件的存储以作为出现任何后续转换错误时的保障。

背 景

XML 申请

2. 首次颁布的 PCT 行政规程附件 F（见 2001 年 12 月 27 日的 PCT 公报特刊 S 04/2001）中体现了对 PCT 电子申请的最初计划，其中设想申请通常将以 XML 格式提交和处理。这曾是（并仍然是）“基本通用标准”的要求，目的是使所有受理局都能接受。其他格式（如 PDF）基本上可以作为过渡性安排，以实现结构良好的全文本文件处理。

3. 然而，尽管提供了费用优惠，但在不允许使用 PDF 提交申请的日本（RO/JP）和大韩民国（RO/KR）受理局，XML 申请只占国际申请的大多数。在其他受理局，大多数申请人认为 PDF 是比 XML 更为方便可靠的提交媒介。尽管对于需要对公布进行排版的主管局而言（申请量最大的所有指定局），有必要对进入国家阶段的任何申请的 PDF 文件进行光学字符识别处理，并由于排版格式中没有的布局或其他排版选项，而存在出现错误和信息丢失的风险。

4. 使用 XML 向中国受理局 (RO/CN) 提交的申请从未超过向该局提交申请总数的 20%，但该主管局最近决定利用 PCT 行政规程新的第 705 条之三，将所有 PDF 和纸质申请转换为 XML 格式进行处理。

5. ePCT 申请支持 XML 申请，既可以直接上传使用外部转换程序准备的产权组织标准 ST.36 XML 和参考图像文件，也可以上传 DOCX 文件，然后将其转换为产权组织标准 ST.36 XML 并在申请提交前向申请人显示结果。在上传 DOCX 文件进行转换的所有情况下，以及在大多数使用外部转换器的情况下，原始 DOCX 文件将作为行政规程第 706 条规定的转换前文件提供。然而，除 RO/KR（不允许使用 PDF 提交）以外，使用 ePCT 提交的申请中只有 2% 是 XML 申请。

6. 此外，多个主管局以本国名义已开始提倡或要求使用 XML 提交申请，典型的做法是申请人以 Office Open XML (DOCX) 格式提交申请体文件，然后将其转换为 XML 格式，无论是产权组织标准 ST.36 还是 ST.96。

XML 处理

7. 自 2003 年以来，可以用 XML 格式提交国际申请，但没有就后续处理活动达成一致标准。如果需要修改¹，通常会像 PDF 申请一样进行公布，并根据修改的位置视需要创建替换页。

8. 之前的 PCT 会议上已经对文本处理问题进行了讨论。特别是在 2009 年，第十六届 PCT 国际单位会议上讨论了日本特许厅 (JPO) 关于 PCT 申请段落修正的提案 (文件 PCT/MIA/16/14)。该届会议报告 (文件 PCT/MIA/16/15) 第 98 至 101 段总结了这些讨论。报告第 100 段指出，“各单位承认有必要建立一个段落替换系统，但尚未就所需的细节达成一致。特别是，有一个单位对与编号系统有关的若干细节表示关切，以及除一对一替换以外应如何进行修改的安排表示关切……”。由于没有达成共识，该提案没有进一步阐述。

9. 2017 年，欧洲专利局提交了一份修改提案 (PCT/EF/PFC 17/005)，建议增加 DOCX 作为 PCT 行政规程附件 F 中规定的可用于申请体的提交格式，并增加 PNG 作为可接受的图像文件格式。这在原则上得到了很大程度的支持，但对于 DOCX 文件本身的法律地位和处理安排，以及由其产生的申请体的任何视图 (无论是直接产生还是通过转换为 ST.36 XML)，都没有达成一致。因此，该提案仍然待定。

10. 其他若干会议也对这些问题进行了一般性的讨论。最近，在 2021 年 6 月召开的工作组第十四届会议上讨论了关于以全文格式处理国际申请的文件 (见文件 PCT/WG/14/8)。该文件包括近年来 PCT 体系从纸质申请向电子申请过渡的概要，阐述了全文处理的好处，并对当时以全文格式提交国际申请的情况提供了概述。

11. 在该文件中，国际局建议修订行政规程附件 F，以便在国际阶段处理由申请人提交或由主管局处理的 XML 全文格式的申请体文件，而这又可以推动向全文国际公布的过渡。此外，为鼓励更多地提交全文申请，国际局提出通过演示网页和网络服务，向所有知识产权局提供其 DOCX 转换器和全文比对工具，这将作为处理以 DOCX 格式提交的申请的参考实现，并确保从 DOCX 转换为全文 XML 的一致性。

12. 一般来说，国际申请的电子提交在很大程度上实现了将申请表著录项目信息作为登记本的一部分以 XML 格式传输。因此，本文件在提到以 XML 格式处理时，涉及的是以 XML 格式处理说明、说明书、权利要求书、摘要和附图。这将使国际公布进一步自动化，并改善所提供的全文内容的质量。

¹ 在大多数行政管理方面，改正、援引加入、更正和修正基本上是相同的程序。本文件将此类行为统称为“修改”。

13. 日本特许厅和韩国特许厅已经表明，申请人对于使用 XML 提交申请通常没有什么困难。但是，没有主管局能够通过提供选择或小额费用优惠来说服申请人以 XML 格式提交申请。在允许的情况下，总是首选 PDF。除其他好处外，PDF 通常允许在 PDF 文件被“打印”之前，可以看到在文字处理器中基于页面的精准呈现²。此外，如果多年后发生诉讼，所提交申请的确切外观很可能是可见的。相比之下，DOCX 没有标准的呈现或产权组织标准 XML，申请体通常呈现为页面视图，与原始 DOCX 中的视图无关。

14. 各主管局对 DOCX 申请和转换有若干不同的安排。这些安排：

- (a) 使用不同的文件转换工具；
- (b) 转换过程的可见度不同；
- (c) DOCX 文件和转换后的 XML 文件具有不同的法律地位；
- (d) 识别和纠正任何存在的转换错误的条件和时限不同。

15. 这种差异有可能造成混淆，并降低对全文本处理的目标和安排的信心，特别是在 PCT 领域，在国际阶段申请文件可能由多个主管局处理，之后在不要求翻译的情况下，可能构成其他主管局进行国家阶段处理的基础。

全文本处理的现状

16. 截至目前，申请人一直在使用若干工具和系统来创建 XML 申请体，以提交国际申请。虽然一些国家主管局多年来一直使用产权组织标准 ST. 36 格式处理这些申请，但最近，另一些国家局开始使用产权组织标准 ST. 96 处理国家申请。此外，已经实施了一些 DOCX 到 XML 的转换器，并且致力于使这些转换器协调一致。然而，协调一致是国际阶段处理的一个持续挑战，因为任何不兼容都会给申请人和主管局带来问题。此外，对于处理全文本格式的国际申请的修改，只有有限的规定。

以全文本格式提交国际申请

17. 2021 年，98.7%的国际申请以电子方式提交，而 2015 年全电子申请的比例为 93.6%。然而，按百分比计算，这一时期的 XML 申请基本保持不变：2021 年为 27.2%，2015 年为 28.2%。因此，提高用 XML 提交的国际申请的百分比对过渡到全文本处理至关重要。

受理局	全文本 XML 申请	PDF 申请	纸质申请	申请总数
JP	48,720	0	320	49,040
KR	20,399	0	126	20,525
CN	5,730	67,480	246	73,456
IB	283	13,122	101	13,506
EP	192	37,472	658	38,322
US	0	56,367	61	56,428
其他受理局	163	23,570	2,106	25,839
总计	75,487	198,011	3,618	277,116

² 基于页面的精准呈现是无法保证的——即使 PDF 是一个文档呈现系统，但基于查看器功能的不同也会存在差异，即使是文本，但特别是当呈现依赖于重叠的图形或透明度时。

表 1: 按受理局开列的 2021 年国际申请提交格式

18. 表 1 显示, 日本特许厅或韩国特许厅收到的以 XML 提交的国际申请约占 91.6%; 这两个主管局收到的 XML 国际申请分别占 99.3%和 99.4%。相比之下, 虽然国际局作为受理局允许提交 DOCX 文件并根据产权组织标准 ST.36 转换为 XML 格式, 但选择这么做的申请人很少。大多数申请人似乎仍然担心处理全文本申请存在风险, 尽管用 XML 提交申请可以额外减费 100 瑞士法郎, 但仍然继续提交 PDF 文件。与 2020 年相比, 申请格式的概况没有变化 (见文件 PCT/WG/14/8 表 1)。

19. 其他一些知识产权局正在继续努力增加 XML 申请提交量。中国国家知识产权局 (国知局) 和欧洲专利局 (欧专局) 都在开展项目来为国际申请的全文本提交和处理提供便利。关于国家和地区系统, 欧专局、法国国家工业产权局 (INPI) 和美国专利商标局 (美国专商局) 已经或正在实施全文本申请的处理系统, 作为打算用于除纸质和 PDF 以外的主要处理途径。

全文本申请的传送

20. 国际局从一些受理局系统中收到页面图像格式的国际申请, 作为登记本的一部分, 尽管申请是以 XML 格式提交的。这在目前是必要的, 因为本地系统是在页面图像处理的时代开发的。国际局为以 XML 格式提交的申请人提供折扣, 但整个 PCT 体系目前并未从 XML 中受益。同样, 有一些主管局接受电子申请, 但仍然只接受基于图像的申请。为了能够以全文本格式处理这些申请, 国际局要求尚未过渡到以 XML 格式传送登记本的受理局, 在收到该格式的申请后, 尽早进行过渡。

目前对全文本格式国际申请的处理

21. 对于以全文本格式提交的申请, 大部分后续处理是基于根据 XML 制作的图像文件, 包括在主管局之间传送图像文件。修正、改正和更正作为替换页处理。在国际局, 图像申请体和 XML 申请体“双重工作”并行处理, 从而在进行包含替换页的传统公布的同时也保留全文本。

22. ePCT 已经接受依据第 19 条, 通过提交 DOCX 文件或基于文本的 PDF 文件进行的修改。然而, 这是最简单的修改情况, 因为修改后的权利要求是在最初提交的权利要求之外显示的, 而不是替换它们。此外, 这些提交的文件需要用户提供描述修改的单独信息, 如哪些权利要求已被修改。

基于全文本的国际公布的临时处理

23. 迄今为止, 国际局已将图像格式和 XML 格式 (对于 XML 申请) 的改正和更正, 作为一项并行业务活动进行处理; 这种重复工作有可能造成错误和不一致。作为消除这种重复的第一步, 国际局正准备接受来自国知局的包含 XML 格式的改正和更正的申请体, 国际局打算将其输入 XML 处理中, 以生成这些国际申请公布的页面图像文件。在这一临时处理安排中, 国际局不会为修改内容生成真正的“替换页”, 而是使用“分支编号” (“1.”、“1.1”、“1.1.1”) 和空白段落生成新的 XML, 以确保段落和图的编号的一致性。修改后的版本将标记出修改后的段落, “替换页”的页脚将出现在任何含有与原来不同的段落或图的页面底部。然而, 分页符通常与提交的国际申请中的分页符不匹配。

全文本处理的原则

24. 国际局继续支持转为全文本处理申请。目的是:

- (a) 申请人应该能够在提交的国际申请中提交全文本格式的申请体, 如果有任何修改, 在使用在线申请系统时, 应能立即看到在已公布申请中如何呈现, 或作为国际专利性初步报告所附的修正 (第二章);

(b) 全文本申请产生的视图在国际阶段处理的各个阶段都可以同等使用；

(c) 指定局和第三方应能查看申请的全文本版本，显示所有已接受的改正、更正和修正，清楚地标记以区分原始内容和修改内容，类似于现行制度中的“替换页”标记，但这是在标题、段落、权利要求和图的层面，而不是页面层面。

25. 国际局正在努力设计一个简单、一致的程序，以满足申请人、主管局和第三方的要求。下文列出了一些需要考虑的问题，并指出在分布式处理环境中准确识别国际申请版本的重要性，在这种环境下可能很难在后期看到所有文件，从而了解原来的背景和时间。

准备 DOCX 申请的形式要求

26. 目前，细则第 11 条是针对纸质文件的准备工作的。许多要求与 XML 文件无关，因为 XML 文件没有边距、字体大小或页面布局的概念。另一方面，为了确保申请人知道如何准备 DOCX 申请，以便毫无困难地转换为处理和准备所需的 XML 格式，还需要其他规定。细则第 11 条应规定（直接或通过参考 PCT 行政规程）明确的起草要求。不应强制要求申请人使用特定模板，但应了解可以和不使用的功能，以及特定类型的特殊内容将如何被导入和使用。此类要求的例子如下：

(a) 段落文本必须以单一字体大小的黑色文本呈现，仅使用支持的效果（粗体、下划线、斜体、上标、下标、小写）。任何其他版式设计将被忽略（转换软件会发出警告）。为了有效公开而需要的任何其他版式设计特征都应该以图像的形式呈现（除了可支持数学和化学公式的标记）。

(b) 所有文本都必须以 Unicode 字体呈现，但某些通常由文字处理器功能使用 Symbol 字体插入的字符除外，这些字符将被映射为 XML 中 Unicode 的适当同等字体（转换软件会发出警告）。

(c) 表格、化学和数学公式可以包括在内，但版式设计的选择可能有限（有待确定）和/或被转换为由 XML 引用的位图图像。

(d) 附图应作为简单的位图图像包含在内，或明确支持复杂的嵌入内容类型，也可以转换为由 XML 引用的位图图像。

(e) 不应使用行号，如果包含行号将被忽略。

(f) 不得使用标注修改模式和评论（ePCT 目前将拒绝使用这两种功能的 DOCX 文件的上传）。

标注修改的粒度

27. 虽然可以在多个层面标注全文本处理文件的修改，例如在段落或字符层面，但由于需要用相关的单位、规定和日期来标记修改，因此除非在段落、权利要求或图的层面上进行，否则很难显示。名称和标题也可能被修改。所有这些要素都在申请体 DTD 中规定。因此，本文件提出了在段落、权利要求或图、名称和标题层面上标注修改的原则。在这种情况下，“标题”包括任何一般标题和任何 IP5 通用申请格式（CAF）的标题（见下文第**错误! 未找到引用源。**段）。

28. 在这一层面标注修改时，我们进一步建议遵循保留文件中所有原始内容的原则。通过将原始内容移到带有原始标识符的“修改”部分，同时将新内容插入到带有新的唯一标识符的文件中，这种模式的目的是使文件的接收者既能使用现有的样式表逻辑制作公布内容，又能查看被取代的内容。如果图、化学公式、数学方程等图像内容有修改，建议将原始图像内容保留在 XML 文件包中，并在“修改”部分的内容中予以提及。

通用申请格式的标题

29. 国际局目前维护的产权组织 DOCX 转换器在处理识别为通用申请格式 (CAF) 的标题时, 创建了 CAF 规范中定义的容器元素, 将与每一部分相关的段落分组。国际局注意到, 这些标题元素增加了全文本国际申请的制作和标注修改的复杂性。除非有任何主管局正在利用 XML 的这种额外结构, 否则国际局建议更新其 DOCX 转换器, 以处理所有的标题, 将其简单识别为标题和段落, 而不产生各部分的容器元素。这不需要对样式表进行修改。

替换页

30. 目前基于页面的公布系统依赖于申请体的修改, 这些修改被记录在申请体中插入和移除的页上。目前, 国际局处理替换页的方式与处理 PDF 和纸质申请的方式相同。为了提供这些申请的全文本, 国际局目前正在做“双重工作”, 即对这些国际申请的全文本和图像副本都进行替换。如上述第**错误! 未找到引用源。**段所述, 国际局正在制定临时解决方案, 以避免双重工作, 但这不足以支持完全自动化处理和向指定局和专利信息用户提供高质量文本。

31. 从长远来看, 目标是“页”的概念应该消失。当需要修改时, 申请人应该以独立于分页的方式提供适当的替换内容。这应该被存储为 XML, 提供原始文件、提交的修改和修改后的文件的记录, 以便生成视图, 显示在段落、图等层面上的变化。目前的原型设计所设想的是, 修改通常是通过提供一个完整的新元素(说明书、权利要求书、附图或摘要), 并由系统检测出与前一版本的不同之处。可能还需要其他机制, 以便在因需要修改而导致从原始 DOCX 格式转换困难的情况下, 允许改变个别段落或图。然而, 如果该系统能够支持使用全彩图像, 这种情况应该很少见, 因为从彩色到黑白的转换是目前遇到的最大转换困难。

32. 为了支持这一点, 需要对细则 26.4、46.5、48.2、66.8、70.2、70.16、74.1 和相关条款进行修正, 对通过除页以外的形式提交和处理替换内容做出明确规定。

编号做法

33. 在 2009 年第十六届国际单位会议上关于段落修正的讨论中(见上文第**错误! 未找到引用源。**段), 各单位审议了段落编号问题。当时有观点指出, 系统能力和做法是编号系统的一个考虑因素(见文件 PCT/MIA/16/14 和文件 PCT/MIA/16/15 第 98 至 101 段)。应再次审议这一问题, 兼顾法律、技术和行政方面的因素, 以确定一个清晰简单的解决方案。

34. 段落编号不应视为实质内容的一部分, 应允许处理申请的主管局将其作为行政事项进行添加、重新编号或重定格式。说明书中的段落应予编号, 但可以重定格式, 例如使“1.”变成“[0001]”或重新编号以提供连续的编号, 无论是在提交的申请中还是在修改后³。这种重新编号不应视为修改, 其本身不应导致段落被标注为已修改。原则上, 这一点也应适用于权利要求书和附图, 但鉴于其特殊性需要进一步考虑, 包括附图普遍采用子编号(图 1a, 图 1b), 以及权利要求经常被交叉引用的事实, 但根据目前的规则, 如果根据第 34 条(细则 66.8(c))进行修改, 则完全被替换。

35. 因此, 应该避免在说明书中交叉引用段落(已经不常使用)。然而, 需要进一步考虑的是, 如果引用的改变仅仅是由于另一段落的编号发生变化, 软件应该如何处理段落引用并进行标记。

³ 这与第 23 段中提到的临时安排不同, 后者使用分支编号, 以避免后续段落的完全重新编号。

36. 这有可能导致申请人和审查员之间的混淆，因为正在处理的申请中的编号可能与最初提交的 DOCX 文件中的编号不同（ePCT 可以发现这一点，并在发生时提供警告）。为了尽量减少这种风险，编号安排必须明确，以便代理人能够可靠地撰写申请，使编号安排符合转换工具的预期。

原始 DOCX 文件的状态和转换保障

37. 迄今为止，阻碍向全文本处理过渡的关键问题之一是需要对全文本处理充满信心。在将 DOCX 文件转换为 XML 时，申请人应相信申请的实质内容得以保留，而在发生转换错误改变公开内容的极少数情况下，可以在任何阶段纠正问题，包括在国家阶段。因此，如果提交的是 DOCX 文件，应被视为最终的原始申请，即使所有后续处理都是基于简化的 XML 格式进行的。

38. 从技术存储的角度来看，根据《行政规程》第 706 条，DOCX 文件在转换为 ST.36 XML 的同时被自动存储为“转换前文件”的现有安排，似乎符合要求。然而，需要进一步考虑法律问题（目前 ST.36 XML 被认为是“原始的”，尽管有权进行修改）和可靠地确定真正的原始内容的问题，因为 DOCX 这一格式会产生不同的呈现，这取决于所打开的文字处理器。

预期的 ePCT 项目工作

39. 国际局建议继续开展试点实施，使得向接收 ePCT 申请的受理局提交全文本申请并选择在 ePCT 中进行国际检索的国际检索单位的申请人，可以通过上传修改后的 DOCX 提交细则 26 和细则 91.1 请求。这将直接为申请体的拟议修改创建所需的 XML，可供国际局和国际单位以 XML 格式处理。这将为向其他受理局和国际单位公开相同的技术功能作为网络服务，和/或把相关组件打包为可在本地部署、用于国家主管局托管服务的系统做准备。

40. 请工作组就 PCT 全文本处理的原则和要求，特别是就文件 PCT/WG/15/14 第 24 至 39 段所列的问题发表评论意见。

[文件完]