

Patent Cooperation Treaty (PCT) Working Group

Fourteenth Session
Geneva, June 14 to 17, 2021

PROCESSING INTERNATIONAL APPLICATIONS IN FULL-TEXT FORMAT

Document prepared by the International Bureau

SUMMARY

1. This document sets out a proposal to move towards full-text processing of international applications. The International Bureau intends to revise Annex F of the Administrative Instructions to allow the international phase processing of the application body in the XML full-text format as filed by the applicant or processed by Offices, which in turn will enable the transition to full-text international publication. To encourage more filing of full-text applications, the International Bureau will make its DOCX converter and full-text comparison tools available to all IP Offices through web services, which will serve as a reference for implementation of processing applications filed in DOCX format and ensure consistency in the conversion from DOCX into XML.

BACKGROUND

2. In recent years, the PCT System has transitioned from paper-based to electronic filing and processing of international applications. Paper filings have steadily decreased, with electronic filings now representing more than 98 per cent of international applications. This transition has allowed for new opportunities for Offices and online services for applicants. It has also resulted in greater efficiency and improved timeliness of Offices in performing required actions on applications during the international phase. However, despite processing taking place largely electronically, many Offices process and transfer data in the form of images equivalent to paper sheets rather than as readable text that could be directly usable by the International Bureau and national Offices in their different capacities. Moreover, where applications are filed in full-text, page images of the application are often processed by Offices.

3. The technical (IT) environment is one of the four main areas of work in terms of future development of the PCT System outlined in the Memorandum by the Director General “The PCT System – Overview and Possible Future Directions and Priorities”, which the Working Group discussed at its eleventh session in June 2018 (see document PCT/WG/11/5). This underlined the need for PCT application data to be transferred in consistent and effective formats between the applicant, the receiving Office, International Authorities and the International Bureau, and then onto the designated Offices. The Working Group, at its twelfth session in June 2019, discussed a document by the International Bureau on PCT online services (see document PCT/WG/12/10). The document includes an update on the use of XML in the filing and processing of international applications (see paragraphs 14 to 22 of the document) needed to transfer data as readable text. Paragraphs 17 to 38 of the Report of the session (document PCT/WG/12/25) provide details of these discussions.

4. This document makes proposals for moving towards the processing of full-text processing of international applications. These include coordinated actions by the International Bureau and Offices to ensure consistency in the handling of international applications filed or processed in full-text format. The proposals are described in terms of the application as filed and subsequent changes made to the application, but would apply equally to translations provided for the purposes of international publication and for international search and preliminary examination.

BENEFITS OF FULL-TEXT PROCESSING

5. The processing of application bodies in full-text format has many advantages over the exchange of image files, such as:

- easier and more accurate text search of international applications after publication;
- increased overall efficiency of international application processing;
- search engines can include the original text of the application as published, avoiding the need for optical character recognition (OCR) to produce a full-text version of a document, and thereby eliminating errors that can result from the OCR process;
- pre-classification of a patent application can be automated by extraction of technical terms from the text;
- machine translations of the filed application can be provided for search engines and, where necessary, document parts can be provided more efficiently to translation assistance software to assist human translation;
- checking for formal defects in an application can be automated, removing the need for checks for image reproduction such as margins and font sizes, thereby reducing the number of tasks to perform before international publication;
- amendments, corrections and rectifications to the application during processing can be shown in granular mark-up;
- full-text format can support color drawings filed with an application; and
- designated and elected Offices can extract the full-text of the international publication for national phase processing.

6. In order to benefit fully from these advantages, an application must be filed in full-text format, with the end-to-end processing based on the full-text being performed consistently across all Offices and the International Bureau.

CURRENT STATUS

FILING INTERNATIONAL APPLICATIONS IN FULL-TEXT FORMAT

7. In 2020, 98.3 per cent of international applications were filed electronically, compared to 93.6 per cent of full electronic filings in 2015. However, in percentage terms, XML filings over this period have remained at a similar level: 27.9 per cent in 2020, 28.2 per cent in 2015. Increasing the percentage of international applications filed in XML is therefore crucial to a transition to full-text processing.

<i>Receiving Office</i>	<i>Full-text XML applications</i>	<i>PDF applications</i>	<i>Paper applications</i>	<i>Total applications</i>
JP	49,005	0	308	49,313
KR	19,525	0	150	19,675
CN	7,651	64,336	354	72,341
IB	293	12,963	176	13,432
EP	138	37,929	806	38,873
US	0	55,665	108	55,773
Other ROs	160	22,341	2,796	25,297
Total	76,772	193,234	4,698	274,704

Table 1: Filing Format of International Applications for 2020 by Receiving Office

8. Table 1 shows that approximately 89 per cent of international applications filed in XML are received at either the Japan Patent Office (JPO) and the Korean Intellectual Property Office (KIPO); these two Offices receive 99.4 per cent and 99.2 per cent of their international applications in XML, respectively. By contrast, while the International Bureau as receiving Office provides the possibility of submission of a DOCX document and conversion to XML format in accordance with WIPO Standard ST.36, the take up of this option is low. It appears that most applicants remain concerned over the perceived risks of processing a full-text application and continue filing PDF despite the additional 100 Swiss franc reduction for filing in XML.

9. Several other IP Offices are working towards increasing XML filings. The China National Intellectual Property Administration (CNIPA) and the European Patent Office (EPO) are both running projects to facilitate the filing and processing of international applications in full-text. Regarding national and regional systems, the EPO, the National Institute of Industrial Property (INPI) France, and the United States Patent and Trademark Office (USPTO) are implementing systems to process full-text applications.

CURRENT PROCESSING OF INTERNATIONAL APPLICATIONS IN XML FORMAT

10. Where applications are filed in full-text format, much of the subsequent processing is based on image files that have been rendered from the XML, including transmission of image files between Offices. Amendments, corrections and rectifications are handled as replacement sheets, losing the advantages of processing the international application in full-text.

PROPOSAL

11. The International Bureau believes that the way forward to increase the filing and processing of international applications in full-text is to implement a mechanism for applicants to file applications using DOCX with confidence that the application will be treated consistently in the conversion into XML and rendering of the pages for international publication. Annex F of

the Administrative Instructions allows for the filing of the application body in XML format based on the conversion of a DOCX format file that is retained as a pre-conversion reference document. The XML file constitutes the legal copy of the international application, with the applicant reviewing a proof copy rendered from the XML prior to filing.

12. The International Bureau will make its DOCX converter and comparison tools available to all IP Offices by web services. These tools will serve as a reference for implementation of processing applications filed in full-text format. Offices using their own tools will be required to match the reference implementation of these tools to ensure consistency in the conversion from DOCX into XML.

13. In addition to updating the DTDs to enable the transmission of full-text content, the International Bureau proposes to update Annex F to require an application filed in full-text to be processed in XML during the international phase. The changes to the DTDs will also support the markup of changes to applications using common application format (CAF)-compliant markup and will provide structures to support processing of these changes.

EFFECT OF PROPOSED CHANGES ON THE PROCESSING OF INTERNATIONAL APPLICATIONS

Filing of an International Application

14. To file in full-text format, applicants need to submit a DOCX or an XML file. The filing software will provide a tool to convert a DOCX into XML and re-render a proof copy for the applicant to review before filing. The filing package will contain the DOCX as the pre-conversion reference as a backup in case of conversion software defects, along with the XML filing for processing the full-text application. The receiving Office will use the same software to review the international application, and will forward the record copy to the International Bureau as a full-text XML file.

15. In relation to drawings and other parts of an application such as mathematical equations and chemical structures that cannot be handled in text format, the relevant information will be included in the XML application package as embedded images. The International Bureau's conversion tool handles these automatically. Images are automatically converted to black and white TIFF images, though a key goal of developing the processing and publication systems is to permit the use of color images.

Modifications to the International Application

16. Where an applicant needs to file subsequent documents to make changes to an international application filed in XML format, the process will be similar to the initial filing. As amendments, corrections and rectifications cannot be directly inserted into the full-text version of the application body, the applicant will edit the source DOCX or XML file of the application, which would (where appropriate) be converted into XML for comparison. A comparison tool made available in ePCT, or similar solution provided by a receiving Office, will be used to markup changes, convert the file into a markup application body in XML, and generate a request for the changes. Before submitting the changes to the relevant Office, the applicant will be able to review the changes to the full-text application.

17. Initial implementations are expected to require that Offices provide the International Bureau with revised versions of the XML application body without change markup. The International Bureau will then use comparison tools to add change markup. Subsequently, as Office implementations adapt, Offices and the International Bureau will use the same tool as applicants to markup changes and indicate their status in the application body. The full-text application body file would include the latest version of the application with a separate section to hold the change history for viewing earlier versions and modifications, if needed. Examiners considering the proposed changes would be able to refer to an automatically generated view

showing the application as proposed to be changed, a difference file and the explanatory letter required to accompany changes such as corrections under Rule 26.4 or amendments under Rule 66.8.

International Publication

18. The International Bureau will publish applications as full-text if filed or processed in this way, with the XML file constituting the legal copy. For the time being, the International Bureau will also publish these applications in TIFF format as page image documents in a format common with the format for applications that are filed on paper or electronically in PDF.

ISSUES FOR APPLICANTS

19. If the process is sufficiently well designed, it ought to be advantageous to applicants to deliver full-text documents for use in the patent application and publication process. Many national Offices typeset their published applications and granted patents and so need to convert paper or PDF application bodies to a full-text format, requiring careful checking by the applicant before publication. Moreover, a typeset document will always have a different layout from the original document with more limited formatting options than available in most word processors. Consequently, any information in a PDF document that relies on specific layout options is likely to be lost in creating the typeset publication and may be difficult to replicate. It should be easier for Offices to receive full-text documents from the outset and to conduct search and examination on these full-text versions. For applicants, this should limit problems with presentation or conversion and, where problems occur, increase the chances of identifying and resolving them at an early stage.

20. However, patent attorneys in many PCT Contracting States have been wary of proposals in this area. Previous attempts to move to full-text processing have only been successful in countries where other available filing options are significantly more onerous for the applicant. This is practical in national systems or where applicants usually use the same Office as both receiving Office and International Searching and (where relevant) Preliminary Examining Authority. However, where different Offices are involved, it requires coordination and consistency to ensure benefits for both applicants and Offices. In turn, this will need attorneys to assist in the process to ensure that their key needs are met and to adapt to new ways of working that – if done properly – will benefit their clients and the system as a whole.

21. A traditional concern relates to potential errors in conversion from word processing formats with a control on the layout of text and graphics to the simplified XML formats underlying document publication and exchange processes. In any system it is essential to ensure that:

- (a) conversion of application bodies from page-based layouts to text-based XML content with appropriate references to images is highly reliable and does not affect the substantive meaning of the description, claims or drawings being converted; and
- (b) in any rare cases where conversion errors occur, the process of search and examination helps to identify those errors at an early stage and the applicant is able to correct the errors whenever they come to light.

22. In addition, for international processing, it is important that conversions are consistent, giving the same results irrespective of which Office a document is submitted to, so that an initial filing at one Office can be properly combined and compared with amendments, corrections and rectifications submitted to a different Office. Consistency in conversion results between national Offices is also more generally desirable to reduce costs and avoid confusion by needing to understand different requirements for national and international processing.

23. Given the issues of typesetting an application or granted patent with specific and limited formatting, drafting an international application should already involve ensuring that the application does not convey information using formatting that cannot be reliably reproduced in the publication. A full-text filing solution is essentially the same issue. A well-designed converter will retain all the substantive meaning of the content, provided that the original document uses a Unicode font and discloses the invention using only the formatting options permitted in the converted format.

24. Issues remain to be addressed – notably the effective presentation of tables, where applicants frequently use formatting “tricks” such as font sizes, manual line breaks, text alignment and manual column sizing to achieve a particular visual arrangement, which will frequently fail to be reproduced in a converted table. Concern has also often been expressed over the use of symbols, which may be displayed in a word processor in a manner decided by the software, beyond the control of the applicant. However, this can be detected and allowed for, provided that other characters in the font directly selected by the applicant adhere to the Unicode standard. Similarly, items such as mathematical formulae in an editable (non-image) format that is not supported by WIPO Standard ST.36 can be detected and image versions can be extracted for inclusion within the converted document where necessary.

25. Full color drawing support is already a key goal of the associated work on development of publication systems. For a successful move to full-text processing, additional input from attorneys is important on the aspects of disclosure that may be difficult to represent or convert adequately within the options currently available in WIPO Standard ST.36. Such issues might be addressed either by expanding the options available directly within the XML Standard (for example, by more complete support for mathematical and chemical formula standards), or (as with color drawings) by extending the range of formats permitted to be included as referenced documents.

TRANSITION TO FULL-TEXT PROCESSING FOR ALL INTERNATIONAL APPLICATIONS

26. In the long term, the International Bureau would like to transition to full-text processing and publication for all international applications. After this transition, any application filed on paper or in image format will be converted into XML, with the converted XML file being the legal copy (subject to the same possibilities of returning to the paper or image original to correct errors). This could be realistic when, for example, more than 98 per cent of international applications are filed in full-text. At current filing levels, this would require conversion of page images for about 7,000 applications. When this threshold has been reached, the International Bureau would continue to produce a TIFF image file of international applications for at least five more years and an assessment would be made of longer-term needs.

IMPLEMENTATION

27. While the International Bureau and other IP Offices will make tools available for applicants to file international applications in full-text, implementation of the proposals in this document will require both the International Bureau and Offices to work together to support applicants in the transition to processing text rather than page images, delivering training and providing assistance when needed. The International Bureau will also provide guidance, in particular, to receiving Offices for processing applications in full-text, as required when the proposals in paragraph 13, above, enter into force.

CONSIDERATION BY THE WORKING GROUP

28. The Working Group is invited to consider the proposals in paragraph 13, above, and the consequent changes to the filing and processing of international applications in paragraphs 14 to 27, above. In particular, the Working Group is invited to agree on the following principles for development of software and processes:

- in the international phase, where international applications are filed in full-text format, PCT Offices and Authorities will process them in full-text, rather than as page-based images;
- PCT Offices and Authorities will process full-text applications using the XML copy of the application as the official copy – the DOCX (or any equivalent format that might be supported) will, however, be recognized as an official view of the application as filed, in a manner that permits correction of any conversion errors discovered during the international phase (and, if possible, through the national phase as well);
- PCT Offices and Authorities should use the reference DOCX converter made available by the International Bureau or another converter that is compatible with the International Bureau converter so that differences between the version as filed and any subsequently submitted amendments, corrections and rectifications can be accurately identified and marked up;
- for the time being, the International Bureau will publish full-text applications in both full-text and image format, and other applications in image format; and
- the International Bureau will eventually move to full-text processing and publication of all applications when a sufficiently large proportion of international applications are filed in full-text format.

29. The Working Group is invited to identify issues that should be addressed in developing tools for conversion of word processor document formats to simplified XML formats, as well as the procedures for using and comparing application bodies so converted.

30. The Working Group is invited to comment on the proposals in the present document.

[End of document]