

Meeting of International Authorities under the Patent Cooperation Treaty (PCT)

Thirty-First Session
Beijing, October 16 and 17, 2024

PCT TEXT PROCESSING TASK FORCE

Document prepared by the International Bureau

SUMMARY

1. International Authorities are invited to comment on requirements, priorities and goals for work related to processing international applications in full text format to assist the quality and efficiency of international search and preliminary examination.

BACKGROUND

2. The first session of the PCT Text Processing Task Force was held from January 29 to 31, 2024 as virtual meeting, attended by delegations from 23 Contracting States, two intergovernmental organizations and seven non-governmental organizations.

3. The session highlighted the desire in principle of many national Offices to move towards full text processing of patent applications in order to improve the quality of processing and of published patent information, allowing full access to the content of applications without the errors involved in optical character recognition. It also highlighted some of the challenges involved, including the wide differences in the preparation and processing of applications across applicants and Offices, the risk of introducing errors during conversions of format, and the difficulty in establishing trusted records based on file formats that do not establish a unique view of an application equivalent to printed pages, including aspects of layout as well as content.

4. It is intended to continue the series of task force meetings, but further analysis is needed of a variety of technical and legal issues in order to present suitable concrete discussion points.

VISION

5. The vision of the International Bureau is that international applications should be filed and processed in full text formats, with corrections, rectifications and amendments normally being proposed in a self-service arrangement by applicants (subject to acceptance or otherwise by the relevant competent authority). A history of the application would be maintained, with easy comparison of the latest version of the application with the previous or as-filed versions, as well as comparison with earlier applications from which priority is claimed.
6. Such a system should allow use of full color drawings (or photographs) and be designed to allow for extension to permit reference to other types of embedded material, such as 3D models or video, should a need be established and standards agreed.
7. The working format for the application body should be display independent. That is, it should be possible to take the content and display it automatically in a variety of different formats, according to the needs of different users. This could include web pages (free-flowing text with no fixed width and text resizable for accessibility), side-by-side version comparisons or page-based PDFs in single column format as currently used in PCT XML filing or dual column format as used in the typeset publications of various national Offices.
8. This is broadly unchanged from the vision set out in the original version of Annex F of the PCT Administrative Instructions in 2002, but to make it effective will require applicants and Offices to embrace a common vision and common technical approaches to the processing of data.

ISSUES

9. Around 28 per cent of international applications are filed in XML format, a figure that has changed only slightly in the past ten years. Recently, the China National Intellectual Property Administration in its role as receiving Office has begun to convert all international applications filed in PDF and paper format to XML, which is then used for processing, including international publication. As a result, over half of international applications are now published based on XML versions of the international application. It is desired to build on this and to deliver ever higher quality and more useful data to designated Offices and patent information users.
10. Some of the key issues that need to be addressed and balanced include:
 - (a) Legal certainty – It is essential that what was filed can be clearly determined and understood if necessary in national phase litigation many years later. Paper and (suitable variants of) PDF have a definitive page-based appearance that conveys meaning through a combination of content and layout. DOCX and the WIPO Standard ST.36 and ST.96 XML formats may be represented in different ways depending on the tool used to view the files.
 - (b) Identification and handling of problematic content – Most word processors have an enormous range of features. The use of special fonts whose characters may not align with the Unicode standard, the inclusion of embedded content from further sources or even simply the dependency on colors and sizes of text or specific spacing arrangements to place text at a particular point on a page to convey information can lead to applications that are difficult to convert to other formats, without the applicant necessarily being aware that they are causing conversion problems. This has always been an issue for Offices creating typeset publications, requiring applicants to check the proposed publication, but becomes a particular issue at the time of filing if the intended content is either entirely incompatible with conversions performed at the time of filing, or else is accepted and converted, but with results that change or lose the substantive meaning without warning to the applicant.

(c) Clarity and consistency – The filing system must be simple, well understood and consistent. Differences between national and PCT document requirements at a receiving Office are likely to cause confusion and errors. Attorneys expecting a conversion tool to work in a particular way may be surprised by a different Office's tool delivering a different result. Offices may be surprised by apparent differences between earlier and later applications even though they might have been created from the same DOCX file. Differences between receiving Offices in document requirements or ways of rendering views of the international applications are likely to cause processing difficulties in the PCT, particularly in cases where International Authorities are competent to act for applications from receiving Offices with different requirements.

(d) Legal basis – PCT Rule 11 and related provisions should properly reflect the requirements of filing and processing international applications in an electronic environment.

(e) Accessibility – The system must work for all parties, including applicants from different countries, filing in different languages, and using different word processors, as well as for Offices with limited automation capacity.

(f) Transitional arrangements – It is necessary to be able to process applications filed in older formats (including paper and PDF) both in the international and national phases.

11. Many national Offices are aiming to update their national processing systems in similar ways. They face similar problems but are approaching them in different ways. This is likely to lead to inconsistencies and the risk of different requirements between Offices for the PCT international phase and for national processing, resulting in confusion for applicants.

EFFECTS ON INTERNATIONAL AUTHORITIES

12. The issues are primarily matters for receiving Offices and designated Offices, but it is important that systems be designed to assist the work of International Authorities and that International Authorities in turn arrange their work to ensure that they assist the applicant as greatly as possible. Some possible issues include:

(a) Feedback would be useful on how International Authorities might benefit most from the use of full text applications in the process of search and examination, including questions of text formats, integration with search systems (notably including artificial intelligence-based systems), markup of changes (which may be particularly significant for international preliminary examination and national phase processing as designated Offices).

(b) Where an international application is originally filed in DOCX format, which is subsequently converted to simpler XML according to WIPO Standard ST.36 (or potentially ST.96 in the future) as a working format, it is highly desirable that the International Authority conduct its search and examination on the basis of either PDF or HTML generated from the working format, rather than views generated directly from the DOCX version. This maximizes the chances that any loss or change in meaning as a result of the conversion process will be spotted at a time when it can be easily corrected.

(c) The systems will need to ensure that International Authorities are easily and accurately able to handle rectifications and amendments in full text formats, deciding on whether they are acceptable and ensuring that the changes are properly incorporated into the working document, irrespective of the receiving Office, noting that where the receiving Office is different from the Office acting as International Searching or Preliminary Examining Authority, there is a risk that a different document conversion tool might have been used. In an ideal system, this would normally be done by the applicant submitting

the proposed changes in full text format such that the examiner can see the proposed changes side by side with the original and simply accept or reject them, with the necessary markup then being applied automatically according to a common standard.

13. To assist the next stages of analysis, the International Bureau would welcome feedback from International Authorities, particularly around the questions of how full text processing could improve and assist international search and preliminary examination, but also with respect to any other related issues.

14. The Meeting is invited to comment on the requirements for the development of systems for the filing and processing of international applications in full text format.

[End of document]