R

WIPO
ВСЕМИРНАЯ ОРГАНИЗАЦИЯ
ИНТЕЛЛЕКТУАЛЬНОЙ
СОБСТВЕННОСТИ

CWS/13/28
ОРИГИНАЛ: АНГЛИЙСКИЙ
ДАТА: 10 ОКТЯБРЯ 2025 ГОДА

Комитет по стандартам ВОИС (КСВ)

Тринадцатая сессия
Женева, 10–14 ноября 2025 года

ПЕРЕВОД В ЦИФРОВОЙ ФОРМАТ ДОКУМЕНТАЦИИ В ОБЛАСТИ ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

Документ подготовлен Международным бюро

РЕЗЮМЕ

1. По отзывам членов Комитета по стандартам ВОИС (КСВ), перевод ведомственной документации в области интеллектуальной собственности (ИС) в цифровой формат является для них одной из главных проблем. На этом фоне патентные ведомства, претендующие на включение своей патентной документации в состав минимума документации Договора о патентной кооперации (PCT), в настоящее время работают над тем, чтобы обеспечить соблюдение соответствующих требований Административной инструкции к PCT. В настоящем документе изложены некоторые из тех проблем, с которыми сталкиваются ведомства, и объясняется необходимость разработки руководящих принципов в области перевода в цифровой формат патентных фондов как приоритетного вида документации.

СПРАВОЧНАЯ ИНФОРМАЦИЯ

2. В 2024 году Секретариат провел обследование с целью понять проблемы, с которыми сталкиваются члены КСВ при обмене данными об ИС, и возможные варианты их решения. Результаты обследования показали, что одной из главных проблем, с которыми сталкиваются ведомства, независимо от размера, является отсутствие материалов бюллетеней ИС в машиночитаемом формате и нехватка ресурсов для поддержки этой деятельности как с точки зрения недостаточной квалификации персонала, так и с точки зрения ИТ-ресурсов (см. пункт 8 документа CWS/12/23).

3. В 2005 году Заседание Международных органов (ЗМО) в рамках PCT создало Целевую группу по минимуму документации PCT («Целевая группа») под руководством Европейского патентного ведомства (ЕПВ) и Ведомства по патентам и товарным знакам США (ВПТЗ США) для проведения всестороннего анализа минимума документации PCT. Однако к полноценной работе этот орган приступил только в начале 2017 года, когда был разработан план действий, ориентированный на достижение четырех следующих целей (см. приложение к документу PCT/MIA/24/4):

(a) Цель A: составить обновленный перечень патентной и непатентной литературы, входящей в состав действующего минимума документации PCT.

(b) Цель B: рекомендовать критерии и стандарты, которым должен соответствовать национальный патентный фонд, включаемый в состав минимума документации PCT.

(c) Цель C: внести предложения относительно необходимости наличия в патентных фондах, входящих в состав минимума документации PCT, четко определенных библиографических и текстовых элементов патентных данных.

(d) Цель D: рекомендовать критерии и стандарты, которые будут использоваться для анализа, пополнения и ведения фондов сведений об уровне техники, основанных на непатентной литературе и традиционных знаниях, и затем, на основе критериев, которые будут выработаны, провести оценку пересмотренного предложения индийских компетентных органов в отношении базы данных Индийской цифровой библиотеки традиционных знаний.

4. По итогам обсуждений, состоявшихся в Целевой группе и Рабочей группе по PCT, в июле 2023 года Ассамблея Союза PCT утвердила поправки к правилам 34, 36 и 63 Инструкции к PCT, которые вступят в силу 1 января 2026 года (документ PCT/A/55/2). Поправки к положениям Административной инструкции к PCT, включая новое приложение H, были обнародованы в июне 2024 года (циркулярное письмо C. PCT 1672). В приложении H конкретизированы технические требования, призванные обеспечить бесплатный доступ любого Международного поискового органа (МПО) ко всем патентным фондам и фондам полезных моделей в составе минимума документации PCT в машиночитаемом формате по крайней мере для публикаций, датированных 1 января 2026 года и позднее, с тем пониманием, что публикации, датированные по крайней мере 1991 годом и позднее, будут доступны в полнотекстовом формате в течение следующих десяти лет.

5. В приложении H указано, что данные с возможностью полнотекстового поиска могут быть представлены в одном из трех форматов: документы в XML, созданные с соблюдением требований стандарта ВОИС ST.36 или стандарта ST.96, или файлы в текстовом формате. Органы могут выбрать оптимальный, по их мнению, формат, при условии, что он позволяет проводить качественный поиск по их патентным фондам. Международное бюро рекомендует использовать формат XML, по крайней мере для представления библиографических данных. Ведомствам, которые сделают выбор в пользу текстового формата, настоятельно рекомендуется рассмотреть возможность перехода на один из стандартов ВОИС в попытке повысить качество своих патентных публикаций и сделать их более удобными для поиска.

ПОДГОТОВКА

6. В порядке поддержки усилий МПО по выполнению технических требований к моменту наступления упомянутого выше крайнего срока (1 января 2026 года) Международное бюро провело в 2025 году серию двусторонних консультаций по

минимуму документации PCT. Ведомства заранее подготовили вопросы для Международного бюро, а оно в свою очередь проследило за тем, чтобы на соответствующих консультациях присутствовал профильный эксперт, готовый ответить на них. В ходе этих мероприятий было отмечено, что в Административной инструкции к PCT нет указаний насчет качества полнотекстовых документов. В ответ на озвученную просьбу Международное бюро предоставило Целевой группе по ведомственному досье свои соображения относительно минимальных требований к тому, что оно считает «приемлемым» документом в машиночитаемом формате.

7. Вниманию ведомств, которые в настоящее время переводят свои патентные документы в цифровой формат, предлагаются следующие рекомендации, призванные помочь привести ведомственные материалы в соответствие с требованиями приложения H:

- документы в XML, соответствующие требованиям стандарта ST.36, должны иметь правильный формат, унифицированный корневой элемент и минимальное число OCR-ошибок. Библиографические данные должны включать по крайней мере реквизиты заявки, реквизиты публикации и название изобретения;

- документы в XML, соответствующие требованиям стандарта ST.96, должны быть подготовлены с соблюдением компонента `PatentPublication` на уровне документа, соответствующего ST.96, однако для них не нужно внедрять схему `MathML` или `OASISTable`; и

- текстовые файлы должны иметь минимальное число OCR-ошибок, текст в них должен быть представлен в виде правильных абзацев, не должен содержать номера строк и страниц и должен позволять читать информацию в верном направлении. Желательно, чтобы такие файлы сопровождались библиографическими данными в формате XML.

8. Целевая группа по минимуму документации PCT согласовала процедуру сертификации в три этапа, которую должны пройти МПО или другие патентные ведомства, претендующие на включение фондов в минимум документации PCT. На втором этапе или этапе верификации проверялось, способно ли ведомство предоставить для загрузки в систему архив данных, содержащий документы в одном из трех форматов, указанных в приложении H, и может ли оно создать ведомственное досье, отвечающее требованиям стандарта ST.37, со всеми необходимыми индикаторами, позволяющими проводить поиск по тексту.

ПРОБЛЕМЫ, С КОТОРЫМИ СТАЛКИВАЮТСЯ ВЕДОМСТВА

9. В ходе консультаций по минимуму документации PCT и на этапе верификации Международное бюро отреагировало на поступившие замечания и вопросы в порядке оказания помощи в подготовке машиночитаемых патентных публикаций и создании ведомственных досье на основе стандарта ST.37. Все проблемы, с которыми сталкиваются ведомства при подготовке машиночитаемых документов, были проанализированы и разбиты на следующие группы:

(a) для всех форматов: низкое качество OCR и обусловленные этим многочисленные ошибки; в результате происходит замусоривание поисковых индексов, особенно применительно к патентным документам, содержащим химические формулы;

(b) для всех форматов: низкое качество распознавания или обработки макета, в результате чего в текст включается информация из верхнего или нижнего колонтитула или в данных на выходе появляются номера строк;

(c) для всех форматов: некачественное деление на сегменты разделов с описанием, формулой изобретения и чертежами, в результате чего, например, в полнотекстовую выдачу формулы изобретения включается текст описания;

(d) для всех форматов: низкое качество разметки (тегирования) языка полнотекстовых разделов, что в дальнейшем приводит к критическим проблемам индексации в поисковых системах;

(e) для документов, соответствующих стандарту ST.96: неверные документы, которые не были проверены на соответствие компоненту `PatentPublication` на уровне документа и непоследовательное использование языковых кодов в иерархии;

(f) для документов, соответствующих стандарту ST.36: непонимание, что можно считать «верным» документом и что является минимальным стандартом при оценке документа в XML на предмет «приемлемости». Например, нужно ли использовать модельное DTD из стандарта ST.36; и

(g) для текстового формата: текст представлен в виде строк, а не в виде правильных абзацев, что затрудняет поиск; библиографические данные представлены в форматах, не соответствующих стандартам ВОИС.

ТЕХНИЧЕСКАЯ ПОМОЩЬ

10. Стремясь помочь в решении хотя бы некоторых из указанных выше проблем, Международное бюро оказывает техническую помощь ведомствам, нуждающимся в поддержке. Инструмент WIPO OCR помогает создавать XML-документы, соответствующие стандарту ST.36, на основе PDF-файлов с изображениями и возможностью поиска по тексту и закладками. Однако до сведения КСВ доводится, что этот инструмент не поддерживает функцию поиска и разметки номера формулы изобретения и воспринимает таблицы и сложные химические и математические формулы как вставленные рисунки.

11. Инструмент WIPO Publish позволяет ведомствам извлекать данные в формате стандарта ST.36 и контролировать их качество, причем уже скоро в результате перехода на новые технологии данные, получаемые на выходе, будут соответствовать стандарту ST.96. Это облегчит обмен материалами бюллетеней для многих ведомств ИС. В 2025 году Международное бюро добавило в архитектуру функции, позволяющие ведомствам, применяющим требования минимума документации PCT, извлекать патентные документы. Пилотный проект был благополучно завершен в Ведомстве интеллектуальной собственности Филиппин. Аналогичная инициатива может быть реализована и в других патентных ведомствах, чьи фонды входят в состав минимальной документации PCT; все заинтересованные ведомства могут обратиться по этому вопросу в Международное бюро.

12. Более того, Международное бюро оказывает непосредственную помощь не менее чем 30 ведомствам, поддерживая процесс перевода их документации в цифровой формат, включая полнотекстовое распознавание документов в случае нескольких ведомств. Эта процедура включает в себя сканирование документов, их перевод в цифровой формат, а затем проверку качества библиографических данных. Во всех случаях эта работа выполняется внешними подрядчиками.

ПРЕДЛАГАЕМЫЕ ДАЛЬНЕЙШИЕ ДЕЙСТВИЯ

13. С учетом изложенных трудностей Международное бюро предлагает подготовить руководящие принципы по переводу в цифровой формат документов в области ИС в порядке помощи в устранении недопонимания, о котором было сказано выше. Для

дальнейшего изучения трудностей, с которыми сталкиваются ведомства, занимающиеся переводом своей документации в области ИС в цифровой формат, Международное бюро предлагает провести обследование и разослать потенциальным респондентам соответствующий вопросник в совместном циркулярном письме КСВ и PCT. Международное бюро намерено подготовить такой вопросник и предложить членам обоих органов принять участие в обследовании в первом квартале 2026 года.

14. Несмотря на то что текущее обсуждение посвящено только оцифровке патентных фондов, упомянутые руководящие принципы вполне могут быть распространены на другие объекты и дополнены инструкциями по переводу в цифровой формат документов, касающихся товарных знаков и промышленных образцов.

15. После того как срок проведения обследования подойдет к концу, Международное бюро намерено организовать в первой половине 2026 года совещание и обсудить подготовку руководящих принципов по переводу ведомственной документации в цифровой формат. Мероприятие будет открыто для всех членов КСВ и PCT; дополнительная информация о нем, включая формат проведения и сроки, будет обнародована позднее. На этом совещании Международное бюро планирует доложить о результатах обследования и обозначить потенциальные сроки подготовки указанных руководящих принципов по переводу в цифровой формат документации в области ИС. Членам КСВ рекомендуется принять участие в обследовании и совещании, поскольку их замечания и комментарии необходимы для разработки упомянутых принципов. Международное бюро также намерено доложить о результатах этого совещания на четырнадцатой сессии КСВ.

16. Наконец, Международное бюро планирует провести практический семинар по стандартам ВОИС ST.36 и ST.96 во втором квартале 2026 года. Приглашения на это мероприятие будут разосланы членам и наблюдателям КСВ, а информация опубликована на веб-сайте ВОИС. Международное бюро также предлагает всем ведомствам, пользующимся стандартами ВОИС, обращаться к нему за технической помощью по мере необходимости.

17. КСВ предлагается:

(a) принять к сведению информацию, изложенную в настоящем документе, в частности проблемы, с которыми сталкиваются ведомства, упомянутые в пункте 9;

(b) одобрить предложение о проведении обследования и пригласить к участию членов КСВ и PCT путем направления совместного циркулярного письма, как указано в пункте 13 выше;

(c) рекомендовать своим членам принять участие в этом обследовании и совещании и принять к сведению, что отчет о результатах совещания будет представлен на четырнадцатой сессии КСВ, как указано в пункте 15 выше; и

(d) рекомендовать своим членам и наблюдателям принять участие в практическом семинаре по стандартам ВОИС ST.36 и ST.96, который состоится во втором квартале 2026 года, как указано в пункте 16 выше.

[Конец документа]