



СТАНДАРТ ST.36

Версия 1.2

РЕКОМЕНДАЦИИ ПО ОБРАБОТКЕ ПАТЕНТНОЙ ИНФОРМАЦИИ С ИСПОЛЬЗОВАНИЕМ XML (РАСШИРЯЕМОГО ЯЗЫКА РАЗМЕТКИ)

Изменения одобрены рабочей группой SDWG ST.36 23 ноября 2007 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	2
ОПРЕДЕЛЕНИЯ.....	3
ОБЛАСТЬ ДЕЙСТВИЯ СТАНДАРТА.....	3
ТРЕБОВАНИЯ СТАНДАРТА.....	4
Основные.....	4
Символы.....	6
Присвоение имен международным общим элементам.....	6
Присвоение имен ведомственным элементам.....	7
Атрибуты.....	7
Добавление, отмена или изменение элементов.....	8
Правила для элементов и атрибутов.....	8
Правила для DTD.....	9
Правила для записи документа.....	12
Внешние объекты.....	14
TIFF.....	14
JPEG.....	14
Стандарт ВОИС ST.33.....	15
Стандарт ВОИС ST.35.....	15
PDF.....	15
МЕГА СОДЕРЖАНИЕ.....	15
Стандартные DTD.....	16
Модельные DTD для патентных публикаций.....	16
ССЫЛКИ.....	16

ПРИЛОЖЕНИЯ

Приложения доступны по адресу: http://www.wipo.int/standards/en/xml_material/st36/

ПРИЛОЖЕНИЕ А: Модельные DTD (xx-patent-document.dtd)

ПРИЛОЖЕНИЕ В: Пример записи документа XML

ПРИЛОЖЕНИЕ С: Международные общие элементы (ICEs)



СТАНДАРТ ST.36

Версия 1.2

РЕКОМЕНДАЦИИ ПО ОБРАБОТКЕ ПАТЕНТНОЙ ИНФОРМАЦИИ С ИСПОЛЬЗОВАНИЕМ XML (РАСШИРЯЕМОГО ЯЗЫКА РАЗМЕТКИ)

Изменения одобрены рабочей группой SDWG ST.36 23 ноября 2007 г.

ВВЕДЕНИЕ

1. Данный Стандарт рекомендует средства XML (eXtensible Markup Language, расширяемого языка разметки), используемые для подачи, обработки, публикации и обмена всеми видами патентной информации. Он основан в большей части на Административных инструкциях к Соглашению о патентной кооперации (PCT), части 7, Приложении F, Приложении I (Patent Cooperation Treaty, Administrative Instructions, Part 7, Annex F, Appendix I) (далее – Приложение F). Термин «ресурсы XML» относится к любым компонентам, используемым для создания и реализации XML. Хотя ресурсы XML обычно включают таблицы стилей, схемы W3C и другие объекты, данный Стандарт в настоящее время включает только определения типа документа (DTD), модели содержания, элементы и небольшое число символов. Более подробно информацию о W3C (World Wide Web Consortium, Консорциум всемирной компьютерной сети) см. на <http://www.w3c.org/>.
2. Данный Стандарт относится к применению расширяемого языка разметки (XML) 1.1.
См. <http://www.w3.org/TR/2004/REC-xml11-20040204/>:

“Расширяемый язык разметки (XML) является сокращенным вариантом SGML (т.е., стандартного обобщенного языка разметки) и полностью описан в данном документе. Его задача – возможность использования, получения и обработки общего SGML в Интернете тем же способом, каким в настоящее время это происходит с HTML. XML был разработан для облегчения реализации и для взаимодействия сетей как с SGML, так и с HTML.”

3. В соответствии с данным Стандартом, включенная в запись документа XML разметка является примером представления содержания документа с использованием XML, при помощи которого “документы составляются из единиц памяти, называемых объектами, содержащих как проанализированные, так и не проанализированные данные. Проанализированные данные состоят из знаков, некоторые из которых образуют данные, а некоторые – разметку. Разметка кодирует описание формата хранения документа и логическую структуру. XML предоставляет механизм наложения ограничений на формат хранения и логическую структуру” (W3C).
4. XML не может быть использован сам по себе в качестве основы для обработки патентных документов – “Данные технические условия не накладывают ограничения на семантику, использование, или (не считая синтаксиса) названия типов элементов и атрибутов...” (W3C).
5. Таким образом, данный Стандарт определяет элементы и их родовые идентификаторы, или «тэги», и атрибуты для разметки патентных документов. То есть, данный Стандарт описывает в некоторой степени семантику (значения), использование и наименования элементов и атрибутов, входящих в состав различных типов рассматриваемых документов.

«[Определение: Каждый документ XML включает один или более элементов, границы которых определены либо начинаяющими и заканчивающими тэгами, либо для незанятых элементов – тэгами незанятых элементов. Каждый элемент относится к типу, определяемому его именем, иногда называемым «родовым идентификатором» (GI), и может иметь набор описаний атрибутов.] Каждое описание атрибута имеет имя и значение” (W3C)

Примечание: Подробное описание и определения см. спецификацию XML на <http://www.w3.org/TR/2004/REC-xml11-20040204/>.

6. Целью введения данного Стандарта является описание логических, независимых от системы структур для обработки патентных документов, как для текста, так и для изображений. Это значит, что данный Стандарт может быть использован вместо Стандартов ВОИС [ST.30](#), [ST.32](#), [ST.33](#) и [ST.35](#), касающихся подачи, обработки, публикации и обмена библиографическими данными, рефератами или полными текстами любых типов патентных документов. Стандарт описывает ресурсы XML для следующих данных:



- (a) Полный или частичный текст патентных документов, включая библиографические данные, записанные как закодированные символами данные.
- (b) Страницы документов целиком, представленные как одно изображение (изображение страницы), вне зависимости от их содержания (библиографические данные, текст или изображения).
- (c) Данные в рамках полнотекстовых документов, которые не могут быть записаны как закодированные символами данные, такие как рисунки, химические формулы и особенно сложные таблицы (так называемые встроенные изображения).
7. Согласно данному Стандарту записи XML должны быть правильно построенными XML-документами и должны соответствовать одному из Описаний Типа Документа (DTD), содержащемуся в приложении F, или ведомственному DTD, которое, в свою очередь, соответствует данному Стандарту. DTD, соответствующее данному Стандарту, должно быть построено из элементов в соответствии с рекомендациями данного Стандарта. DTD, содержащиеся в Приложении F, опубликованы на http://www.wipo.int/pct-safe/epct/xml_canon.htm, где DTD будут обновляться по мере принятия нового варианта. Как только обновленное DTD публикуется на веб-сайте, оно становится доступно для официального использования.

ОПРЕДЕЛЕНИЯ

8. Для целей данного Стандарта даются следующие определения:

- a) Выражение «**патентный документ**» включает патенты на изобретения, патенты на растения, патенты на промышленные образцы, свидетельства о полезности, полезные модели, дополнительные документы к ним, опубликованные заявки и описания, а также документы, относящиеся к ведению дел по заявкам и выданным охранным документам, включая действия после выдачи патента, по поддержанию прав собственности и всю переписку между ведомством и заявителем, а также ведомств между собой.
- b) **Разметка** определяется как текст, добавляемый к содержанию документа и описывающий структуру и другие атрибуты документа способом, не являющимся способом системной спецификации, независимо от какой-либо обработки, которой он может быть подвергнут.
- c). Остальные определения см. описание XML: <http://www.w3c.org/TR/2004/REC-xml11-20040204/>

ОБЛАСТЬ ДЕЙСТВИЯ СТАНДАРТА

9. Несмотря на то, что DTD, описанные в Приложении F, были разработаны для использования в соответствии с Соглашением о Патентной Кооперации, данный Стандарт рекомендует использовать их всеми патентными ведомствами для электронной подачи заявок. Образец DTD, отраженный в Приложении А к данному Стандарту, предназначен служить руководством по использованию международных общих элементов (ICE) для публикации патентных документов. По мере развития Стандарта в список, приведенный ниже, могут быть добавлены другие DTD.

Стандартные DTD, включенные посредством ссылок

mathml2.dtd

soextblx.dtd (может называться также calstblx.dtd)

10. Некоторые DTD, включенные в Приложение F, перечислены ниже вместе с соответствующей стадией делопроизводства по заявке в качестве иллюстрации их предполагаемого использования. Таблица является только руководством к возможному использованию этих DTD в процессе патентного делопроизводства; разные ведомства могут иметь разные нужды.



Руководство по информации и документации в области промышленной собственности

Стандарты – ST.36

Страница: 3.36.4

Название DTD	Производственный процесс						
	Подача	Публикация	Экспертиза	Выдача патента	После выдачи патента	Повторная публикация	Переписка
amendment-request	●					●	
application-body	●	●		●		●	
bio-deposit	●	●		●		●	
Dispatch	●						
Declaration	●						●
Demand	●						●
fee-sheet	●						●
Iper			●				●
package-data	●	●	●	●	●	●	●
Pkgheader	●	●	●	●	●	●	●
power-of-attorney	●		●	●			●
priority-doc	●						●
Request	●						
search-report		●				●	●
sequence-list	●	●		●		●	
table-entity	●	●		●		●	
xmit-receipt	●		●		●		●
xx-patent-document		●		●		●	
xx-bibliographic-data		●		●		●	

ТРЕБОВАНИЯ СТАНДАРТА

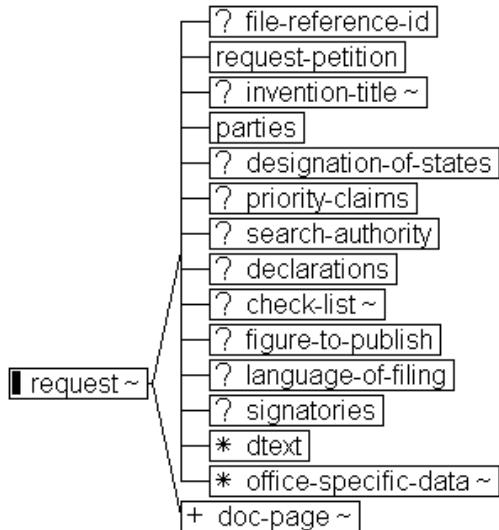
Основные

11. Международные Общие Элементы (ICE) являются основой данного Стандарта. ICE строятся на базе Приложения F, Стандарта ВОИС ST.32 и других источников. См.: http://www.wipo.int/standards/en/xml_material/st36/index.html.

12. ICE должны использоваться так, как определено в данном Стандарте, то есть они должны иметь такие название, содержание, атрибуты и значение, которые указаны в списке ICE. Поскольку данный Стандарт и Приложение F, возможно, не включают все элементы, необходимые для всех патентных ведомств, в таких случаях разрешается использовать специфические для ведомств DTD, как описано ниже.

13. Специфическая для ведомства информация может быть обработана следующим образом.

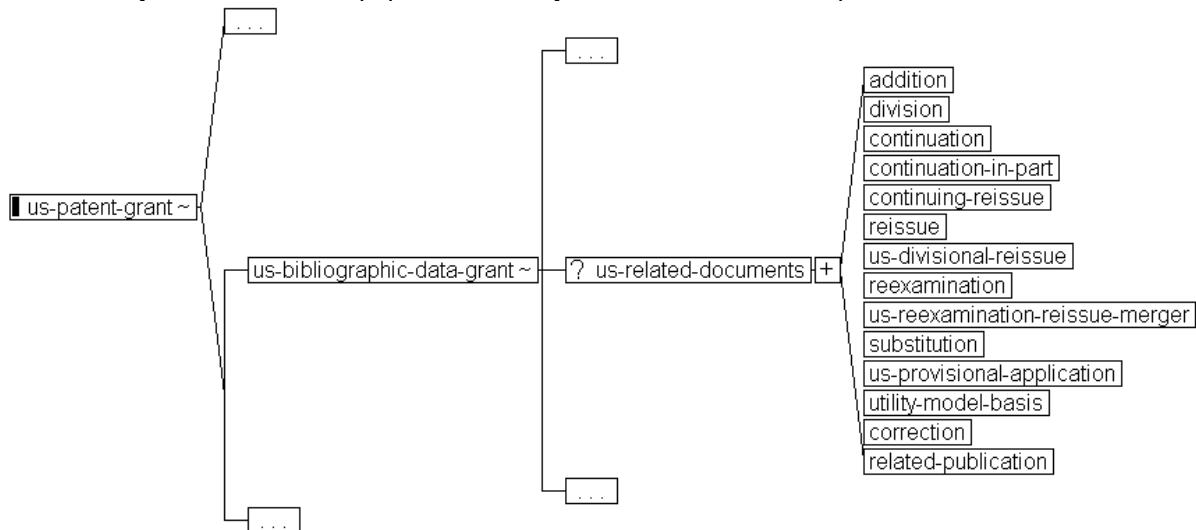
- Выделена ссылкой в отдельное DTD, например, из DTD `request` элементом `office-specific-data` (рекомендуется).



b) Включена напрямую в элемент `office-specific-data`, в этом случае элемент может быть изменен из незаполненного таким образом, чтобы включать #PCDATA или другие необходимые модели содержания; и добавкой префикса двухбуквенного кода страны в `office-specific-data`, например: `wo-office-specific-data`. Модель содержания `office-specific-data` не должна быть изменена без добавления префикса ведомства.

c) Использованием согласования пространства (присваиваемых) имен. Пространства имен XML предусматривают простой способ уточнения имен элемента и атрибута, используемых в документах XML, путем объединения их с пространством имен, определяемым ссылками URL (унифицированный указатель ресурса) (см.: <http://www.w3.org/TR/REC-xml-names/>).

d) Смешиванием специфических для ведомства элементов с международными общими элементами. Например, в фрагменте DTD для публикации, приведенном ниже, специфические для ведомства элементы добавлены в модель содержания элемента `related-documents`. Подробнее см. *Соглашения DTD*, приведенные ниже.



14. На более высоком уровне эти данные могут быть включены в отдельные документы, упоминаемые в DTD package-data посредством элемента `other-documents`. DTD или тэги, упоминаемые в `office-specific-data` или `other-documents`, находятся полностью под контролем соответствующего ведомства.

15. Имена ведомственных DTD и/или элементов должны начинаться с двухбуквенного кода страны соответствующего ведомства по [ST.3](#), за которым идет разделитель (дефис или двоеточие) и имя объекта. Любые другие имена будут пониматься как международные (родовые) DTD или элементы. Следовательно, рекомендуется ограничить использование имен, начинающихся с двухбуквенного слова, только теми, которые представляют действующий код страны. Например, `request` переходит в `ep-request` для использования в ЕПВ в качестве имени файла DTD или корневого элемента.

16. Для обеспечения взаимодействия между патентными ведомствами в области подачи заявок необходимо использовать следующие DTD, как указано в Приложении F, а не ведомственные альтернативы: `application-body`, `table-external`, `pdoc-certificate`, `package-header`, `package-data` и `xmit-receipt`.

17. При включении ведомственных элементов и/или ссылок на ведомственные DTD выпускающий орган должен предоставить другим ведомствам и пользователям уведомление, содержащее информацию о содержании и значении этих элементов и/или DTD. Такое уведомление должно быть размещено на доступном сайте ведомства или на сайте ВОИС. Уведомление должно включать DTD и полное описание каждого элемента.

Символы

18. Несмотря на то, что XML допускает разные кодировки символов, данный Стандарт рекомендует исключительно Юникод. Может быть полезным добавление объектов символов для символов, не включенных в Юникод, как указано на `wipo.ent` (расположено на http://www.wipo.int/pct-safe/epct/xml_canon.htm). В этот файле объекта перечислены общие имена объектов, которые могут быть использованы на месте точек кода из кодировок, как указано на `wipo.ent`. Использование этих объектов требует создания глифов для представления, которые еще не существуют. См. <http://www.w3.org/XML/Core/2002/10/charents-20021023> для получения дальнейшей информации об объектах символов.

19. Записи документов должны включать описание XML в качестве первой строки файла.

```
<?xml version='1.1' encoding='utf-8' ?>
```

20. В данном Стандарте рекомендуется только UTF-8. Тем не менее, в случае идеографического шрифта Юникод в UTF-8 может создавать исключительно большие файлы, т.к. это кодирование может использовать до четырех байтов на символ. В таких случаях национальные ведомства могут выбрать кодирование, которое



Руководство по информации и документации в области промышленной собственности приведет к файлам управляемых размеров. Ведомства, решившие поступить таким образом, должны быть готовы проводить консультации со своими партнерами по обмену и дать соответствующее публичное извещение.

21. Символы, которые разрешено помещать в документ XML, указаны в *XML 1.1 W3C Recommendation*; они рекомендуются и данным Стандартом за следующим исключением. Символы, использующиеся в именах элемента или атрибута, описанных в данном Стандарте, ограничены следующим набором:

{ abcdefghijklmnopqrstuvwxyz1234567890- } .

22. Ведомствам настоятельно рекомендуется создавать записи документов для публикации и обмена, «нормализованные» в соответствии с *Моделью Символов для Интернета (Character Model for the World Wide Web)* (<http://www.w3c.org/TR/2003/WD-charmod-20030822/>). Синтаксические анализаторы, поддерживающие XML 1.1, могут быть использованы для тестирования нормализации. При этом значительно улучшается согласованность сортировки и операций построчного сравнения при условии, что опции определенных символов кодирования, доступные в Юникоде, будут использоваться в международном патентном сообществе согласованно.

Присвоение имен Международным Общим Элементам

23. Все имена элементов должны быть словами из английского языка.

24. Когда для имени элемента требуется более одного слова, слова должны разделяться дефисом (-).

25. Для имен элементов в данном Стандарте используется только латинский алфавит, ограниченный следующим набором символов: { abcdefghijklmnopqrstuvwxyz1234567890- }. Подчеркнутые символы и символы верхнего регистра не используются. Исторически сложилось, что имена элементов в элементе SDOBI Стандарта [ST.32](#) сохраняют прописное В и другие прописные имена элементов.

26. Имена должны быть описательными, не мнемоническими или сокращенными, насколько это возможно практически. Значение имени элемента должно быть понятно каждому без обращения к какой-либо другой документации, либо при незначительном обращении. Некоторые заметные исключения включают большинство общих элементов, используемых в патентной заявке, например, «*p*» для параграфа, и другие, например, полученные из HTML, а также некоторые другие широко используемые элементы форматирования (например, в *application-body* DTD). Маловероятно, что будут необходимы какие-либо дальнейшие исключения.

27. Должны быть включены одно или два предложения, описывающие значение имени элемента и его предполагаемое содержание. Описание должно упоминать любые применяемые правила или нормы и очень кратко резюмировать их сущность. В DTD описание должно быть включено в примечание, непосредственно предшествующее элементу или атрибуту, к которому оно относится.

28. Исторически сложилось, что некоторые элементы в ICE записаны в Графе имен в [ST.32](#) в формате Bnnn, где n – число. Такие имена элементов являются связанными элементами из Стандарта BOINC ST.32, Рекомендации по Разметке Патентных Документов с использованием SGML (Стандартного Обобщенного Языка Разметки). Со временем данная графа может быть удалена, как только ведомства откажутся от использования имен тэгов в формате Bnnn.

Присвоение имен ведомственным элементам

29. Применимы правила для Международных общих элементов (предыдущий раздел).

30. Все имена элементов должны быть словами английского языка, если это возможно.

31. Имя каждого ведомственного элемента должно начинаться с кода [ST.3](#) того ведомства, которому принадлежит элемент. Код [ST.3](#) должен быть отделен от имени элемента либо дефисом (-), либо двоеточием (:). Например, *jp:fterm*, или *ep:printer-name*. Двоеточие используется только там, где ведомство-владелец применяет пространство имен W3C XML (см. *Пространство имен XML*, <http://www.w3c.org/TR/REC-xml-names/>).

Атрибуты



Руководство по информации и документации в области промышленной собственности

32. Если ведомство хочет добавить или изменить атрибуты для ICE, следует представить запрос на изменение.
33. Ведомственные элементы могут иметь любые нужные атрибуты, при условии, что они не противоречат атрибутам, определенным в данном Стандарте (см. следующие параграфы).
34. Имена атрибутов не должны переопределяться в DTD. То есть, имя должно всегда иметь одно и то же значение, вне зависимости от элемента, к которому оно будет применяться. Примечание должно объяснять возможные значения атрибута, что они означают и, по возможности, как их создавать. Примечание к атрибуту должно быть включено в примечание к элементу, которому принадлежит атрибут.
35. Имена атрибутов должны быть использованы повторно там, где значения допустимого атрибута идентичны. Если значения атрибута не одинаковы, имена должны отличаться. Например, атрибут `file` должен быть использован в любом месте, где значением атрибута является имя компьютерного файла. С этой целью не должно использоваться никакое другое имя атрибута.

Добавление, отмена или изменение элементов

36. Для добавления, отмены или изменения ICE требуется утверждение согласно процедурам, установленным ВОИС.
37. Как только определено, что ICE отменен (он не должен больше использоваться), в базе данных ICE должно быть отмечено изменение статуса, а также дата, после которой элемент не должен использоваться.
38. Для добавления, отмены или изменения ведомственных элементов требуется утверждение только со стороны ведомства по промышленной собственности, создавшего элемент. Тем не менее, ведомствам рекомендуется обеспечить отличие значения и описания ведомственных элементов от ICE.

39. Ведомственный элемент, ставший применимым для других ведомств, может быть переведен в статус общего элемента. Переведение в статус общего элемента обычно состоит из отмены ведомственного элемента и добавления нового ICE, имеющего то же имя, что и ведомственный элемент, за исключением предваряющего кода [ST.3](#).
40. Как только ведомство определило, что специфический для ведомства элемент отменен (оно больше не будет его использовать), ведомство посылает запрос на изменение статуса элемента в базе данных, указав дату, после которой оно больше не будет использовать элемент.
41. Если возможно, запросы на изменение DTD, затронутых добавлением, отменой или изменением общего или ведомственного элемента должны быть представлены одновременно с запросом на создание, отмену или модификацию элемента. Введение в действие модифицированных элементов и DTD должно по возможности происходить одновременно.

Правила для атрибутов и элементов

42. Определенные элементы и атрибуты имеют специфическое значение среди международных общих элементов. В данном разделе описано их предполагаемое применение, поскольку в результате их унифицированного использования ожидается значительная польза для сообщества по промышленной собственности.
43. Элемент `date` широко используется в моделях содержания xx-patent-document для записи дат. Следовательно, он всегда должен быть производным от другого элемента, который устанавливает контекст и поясняет, какому событию соответствует дата. Даже если в исходном элементе присутствует слово «дата», элемент `date`, содержащий действующую дату, все равно должен использоваться. Содержание элемента всегда должно выражаться как ряд чисел, в котором первые четыре позиции представляют год, следующие две позиции – месяц (слева помещается ноль, если необходимо), и последние две позиции представляют день месяца (слева помещается ноль, если необходимо). Например, 20020717 представляет 17 июля 2002 г.
44. Элемент `document-id` используется для определения типа документа промышленной собственности, к которому применяется данный Стандарт. Следовательно, он всегда должен быть производным от другого элемента, который определяет контекст и поясняет, какой документ идентифицируется. Элемент `date` в рамках элемента `document-id` представляет дату публикации для опубликованных документов или дату подачи для неопубликованных документов. Никакую другую дату использовать не следует. Элемент `doc-number` не должен включать код вида документа, дату или страну, они фиксируются отдельно. Если необходимо отобразить



Руководство по информации и документации в области промышленной собственности идентификационные данные документа в соответствии со Стандартом ВОИС [ST.10/C](#) или каким-либо другим Стандартом, необходимо использовать таблицу стилей для систематизации содержания. Элемент name используется для одной или другой стороны, связанной с документом.

45. Поощряется создание ведомствами тэгов для каждого местонахождения любого типа номера документа в опубликованном документе. Польза от этого для поисковых систем будет значительной.

46. Атрибут id не является обязательным, но его использование может значительно добавить ценность документам. Значение, присвоенное атрибуту id, должно быть в соответствии со стандартом XML уникальным для каждого документа. Единственной целью присвоения данного атрибута является однозначная идентификация объекта, к которому он прикреплен в данном документе, он не может быть использован для другой цели. Хотя значением id может быть любая последовательность символов, было бы полезным для облегчения прочтения человеком в данном значении отражать тип элемента, которому он назначен. Например, параграфам в описании могут быть присвоены следующие значения id: "p0001", "p0002" ... "p0101" и т.д.

47. Следует с осторожностью присваивать значения таким способом. Например, если публикуется больше чем одна совокупность пунктов формулы изобретения (например, на нескольких языках), то будет существовать более чем один пункт формулы, пронумерованный «1». Даже в этом случае значение id должно быть уникальным, т.е. на практике описанный выше подход должен быть модифицирован, чтобы обеспечить уникальность, к примеру, путем включения идентификации языка, например: "cl-en-0001", "cl-de-0001", "cl-fr-0001" и т.д.

48. Если исправления будут вноситься автоматически, было бы полезным сделать значение атрибута id уникальным в оболочке файла. В этом случае, когда, например, id назначается параграфу, он никогда не должен изменяться, независимо от того, сколько раз изменялся сам документ. Когда удаляется параграф, значение id должно быть изъято из оболочки файла. Когда добавляется новый параграф, оболочке файла должно быть присвоено новое значение id.

49. Атрибут id часто является целевым объектом атрибута idref, связанного с claim-ref, crossref или figref. Составление пар из значений id и idref поддерживает гипертекстовые ссылки при отображении документа в браузере. Элемент crossref предназначен указывать на любой произвольный объект в записи документа, отличный от пункта формулы изобретения или рисунка. Во избежание противоречий с установленным порядком действия браузеров, важно не использовать эти атрибуты для других целей.

50. Базовые DTD и модели содержания элементов в базовых DTD не могут быть изменены для использования в национальных ведомствах (далее они называются базовыми элементами). Даже если базовый элемент используется в DTD, не являющимся базовым DTD, он не может быть изменен. Если ведомство хочет внедрить в базовый элемент дополнительную информацию, id и idref могут быть использованы без изменения базового элемента. Назначьте id данному элементу. Создайте специфический для ведомства элемент при помощи элемента idref, который указывает на базовый элемент. В ведомственный элемент добавьте любые дополнительные элементы, необходимые для национального использования.

51. Атрибут lang обычно включает двухбуквенный код, основанный на стандартах ISO для языка содержания элемента, к которому он присоединяется. В случаях, когда двухбуквенный код является недостаточным, ведомствам рекомендуется следовать правилам, установленным Рабочей группой проектирования Интернета (the Internet Engineering Task Force), и описанным в "Tags for the Identification of Languages" (Тэгах для идентификации языка) (<http://www.rfc-editor.org/rfc/rfc3066.txt>).

Правила для DTD

52. Имя файла DTD должно указывать его версию. Имя файла с номером версии должно представлять собой последовательность знаков, состоящую из индикации типа документа, дефиса, буквы "v", цифры, обозначающей главную версию, и необязательного раздела, состоящего из дефиса и цифры, обозначающей номер последующего пересмотра. Например: request-v1-2.dtd

53. Каждое DTD должно включать в себя номер версии и дату версии в общедоступном идентификаторе. Примеры:

PUBLIC "-//WIPO//DTD APPLICATION BODY 1.3//EN" "application-body-v1-3.dtd"



Руководство по информации и документации в области промышленной собственности
Reference this DTD as PUBLIC "-//USPTO//DTD us-patent-grant v4.0 2004-07-30//EN"
Alias: Grant Red Book (GRB)

54. История пересмотра в начале DTD должна описывать изменения, сделанные в документе, с указанием даты и описанием каждого изменения, в обратном хронологическом порядке. Примеры:

```
<!--
***** Revision history *****
2004-04-15 H. Li
. Added num attribute to li element.
. Changed content model maths from (img | math) to (img | (math,img?)).
. Added date attribute to element us-patent-application.
. Added math, chemistry, table, program-listing to img-content allowed values
.. and changed dna to DNA.
. Add figref to li, dd, claim-text elements and table cell entry.
. Changed subname? to subname* in serial, article and online elements.
. Renamed table-external element to table-external-ref.
2004-03-04 B. Cox
. Renamed us-continued-prosecution-application to
.. us-issued-on- continued-prosecution-application
.. and renamed attribute from cpa-text to grant-cpa-text
.. and changed attribute value (see below).
***** End Revision History *****
-->
```

55. В вводную часть включаются контактные данные. Примеры:

```
<!--
Contacts:
```

Стандарты – ST.36

Страница: 3.36.10

EPO: Paul Brewin pbrewin-z@not-epo.org
JPO: Shiro Ankyu ankyu-shiro-z@not-jpo.go.jp
USPTO: Bruce B. Cox bruce.cox-z@not-uspto.gov
WIPO: Hideto Tanaka hideto.tanaka-z@not-wipo.int

-->

```
<!--
Contact: Bruce B. Cox
U.S. Patent and Trademark Office
Crystal Park 8, Suite 1032
Washington, DC 89231
+1-783-906-6062
bruce.cox@uspto.gov-->
```

56. DTD, включающие при помощи ссылки стандартные DTD, часто являются нечитаемыми (не могут быть проанализированы), если внешние DTD не доступны на локальном компьютере. Следующий пример представляет способ простого временного скрытия таких внешних ссылок, что позволяет осуществлять редактирование и просмотр DTD там, где доступ к внешним ссылкам не удобен. Такое решение предполагает знание отмеченных разделов DTD. (Более подробно см. <http://etext.lib.virginia.edu/TEI/p4/tei.cgi?div=div2&id=SG17BIS> и аналогичные сайты).

```
<!-- -to include mathm!2.dtd change MATHML2_DTD value to "INCLUDE",
change MATH_PLACEHOLDER value to "IGNORE", the same for the
TABLE_DTD, TABLE_PLACEHOLDER, and WIPO_ENT
INCLUDE
IGNORE
<!ENTITY % UNICODE_PLANE1D_ESCAPE "IGNORE">
<!ENTITY % WIPO_ENT "IGNORE">
<!ENTITY % MATHML2_DTD "IGNORE">
<!ENTITY % TABLE_DTD "IGNORE">
<!ENTITY % MATH_PLACEHOLDER "INCLUDE">
```



Руководство по информации и документации в области промышленной собственности

```
<!ENTITY % TABLE_PLACEHOLDER "INCLUDE">
<![%UNICODE_PLANE1D_ESCAPE; [
<!ENTITY % planeID "&#38;#38;#xE">
<![%WIPO_ENT;[
<!--
    import character entity set. Download from:
    http://pcteasy.wipo.int/efiling\_standards/schemaDocs/wipo.ent Note that nsgmls-based
    parsers (SP, Near & Far Designer, etc.) may not be able to process this file for reasons
    described below in MathML comments.
-->
<!ENTITY % wipo PUBLIC "-//WIPO//ENTITIES WIPO 1.0//EN" "wipo.ent">
%wipo;
]]>
<![%MATHML2_DTD; [
<!-- - DTD MathML2: maintained by W3C. Download from:
    http://www.w3.org/TR/MathML2/DTD-MathML-20010221.zip
    If using nsgmls-based parser (SP, Near & Far Designer, etc.) Uncomment 'mathml-charent-
    module' switch below or replace the Referenced MathML2 DTD with the version
    downloadable from: http://www.w3.org/Math/DTD/dtd-sp.zip
    This notice copied from: http://www.w3.org/Math/DTD/
    "DTD for nsgmls"
    Some systems (including the popular nsgmls parser) may not be able to process files using
    'plane 1' characters which have Unicode numbers higher than #xFFFF. The versions of the
    DTD provided here incorporate the modifications mentioned above, but the high characters
    are replaced by the equivalent mchar construct <mchar name="..." /> this allows the DTD to
    be read and for MathML files to be validated using such systems."-->
```

```
-->
<!--ENTITY % mathml-charent.module "IGNORE" -->
<!ENTITY % MATHML_prefixed "IGNORE">
<!ENTITY % MATHML_xmns "">
<!--import MathML2 dtd -->
<!ENTITY % mathm12 PUBLIC "-//W3C//DTD MathML 2.0//EN" "mathm12.dtd">
%mathm12;
]]>

<![%TABLE_DTD; [
<!-- DTD OASIS Open XML Exchange Table Model.
    Maintained by OASIS; download from:
    http://oasis-open.org/specs/soextblx.dtd
    Note that the FPI in soextblx.dtd refers to itself as 'calstblx'.
    That convention has been followed here.
-->
<!-- create content for title element in table -->
<!ENTITY % title "<!ELEMENT title (#PCDATA | b | i | u | sup | sub | smallcaps)*">">
%title;
<!--override OASIS Exchange <entry> model -->
<!ENTITY % tbl.entry.mdl "#PCDATA | b | i | u | sup | sub | smallcaps | br
    | patcit | nplcit | bio-deposit | crossref | figref | img
    | dl | ul | ol | chemistry | maths)* ">

<!--import OASIS Exchange model - ->
<!ENTITY % calstblx PUBLIC "-//OASIS//DTD XML Exchange Table Model 19990315//EN"
    "soextblx.dtd">
<!ENTITY % yesorno "NMOKEN" >
<!ENTITY % tb.l.table.att      "pgwide    %yesorno; #IMPLIED
                                orient      (port|land) #IMPLIED
                                tabstyle   NMOKEN    #IMPLIED">
```



Руководство по информации и документации в области промышленной собственности

```
%calstblx;  
]]>  
<![%MATH_PLACEHOLDER; [  
<!---(PLACEHOLDER:w3c math dtd)--&gt;<br/><!ELEMENT math (#PCDATA)>  
]]>  
<![%TABLE_PLACEHOLDER; [  
<!---(PLACEHOLDER:cals table dtd)--&gt;<br/><!ELEMENT table (#PCDATA)>  
]]>
```

57. DTD должно, как правило, содержать список элементов, начинающийся с корневого элемента, и отсортированных по ветвям в естественном порядке документов с атрибутами, следующими непосредственно за элементами, к которым они относятся. Тем не менее, допустим любой порядок, помогающий пользователям понять логическую структуру документа.

58. Как указано выше, каждый элемент в DTD включает примечание, стоящее непосредственно перед элементом, из которого можно уяснить его значение или применение. Примечание должно содержать информацию, достаточную для правильного использования элемента или атрибута, и ссылку к более полной документации, если это необходимо. Примеры:

```
<!---The problem the invention purports to solve (Rule 5.1 (a) (iii))-->  
<!ELEMENT tech-problem (heading*, p+)+>  
<!ATTLIST tech-problem id ID #IMPLIED >
```

59. Правила для ICE DTD применимы к ведомственным DTD.

60. По возможности DTD должны использовать ICE, оставляя ведомственные элементы только для тех случаев, когда национальная практика значительно отличается от международной.

61. Ведомственные элементы могут быть включены, куда необходимо, в другом Стандарте ВОИС по DTD. Такое измененное DTD становится ведомственным DTD и должно быть названо соответственно, то есть корневой элемент, имя файла и публичный идентификатор ведомственного DTD должны начинаться с соответствующего кода [ST.3.](#)

Стандарты – ST.36

Страница: 3.36.12

62. Любое DTD, которое включает хотя бы один ведомственный элемент, является по определению ведомственным DTD.

Правила для записи документа

63. Следующие указания следует принимать во внимание при формировании записей документов в соответствии с данным Стандартом.

64. Записи, соответствующие данному Стандарту, должны быть правильно построены и должны соответствовать XML v1.1.

65. Рекомендуется, чтобы при присвоении имен отдельным файлам записи документа соответствовали договоренности по присвоению имен, описанной в Приложении F (опубликовано на сайте ВОИС по адресу <http://www.wipo.int/scit/en/standards/index.htm>) для электронной подачи заявок и их обработки. Например, следующий список файлов соответствует опубликованному выданному патенту США, включающему семь страниц изображений.

USO6282717-20010904.xml
US06282717-20010904-DOOOOO.TIF
US06282717-20010904-D00001.TIF
US06282717-20010904-D00002.TIF
US06282717-20010904-D00003.TIF
US06282717-20010904-D00004.TIF
US06282717-20010904-D00005.TIF
US06282717-20010904-D00006.TIF

66. Для международных заявок каждое получающее ведомство по договору РСТ должно создавать записи документов, полностью соответствующие требованиям Приложения F.



Руководство по информации и документации в области промышленной собственности

67. Записи документов XML должны начинаться с вводной части. Она должна содержать инструкцию по обработке XML, которая указывает версию XML и схему кодирования. Вводная часть также должна содержать описание типа документа, известное также как оператор DOCTYPE. Описание типа документа может необязательно содержать публичный идентификатор (предваряемый ключевым словом PUBLIC), а также должно содержать системный идентификатор (предваряемый ключевым словом SYSTEM, если не использовано ключевое слово PUBLIC), который устанавливает URI (универсальный идентификатор ресурса), ссылаясь на DTD, которому соответствует запись документа. Публичные идентификаторы могут находиться в примечаниях в начале каждого файла DTD.

Пример – вводная часть с публичным идентификатором:

```
<?xml version="1.1" encoding="UTF-8"?>
<!DOCTYPE request PUBLIC "-//WIPO//DTD REQUEST 1.1//EN" "request-vl-1.dtd">
```

Пример – вводная часть с системным идентификатором:

```
<?xml version="1.1" encoding="UTF-8"?>
<!DOCTYPE request SYSTEM "request-vl-1.dtd">
```

68. Каждая запись документа должна однозначно указывать DTD, которому она соответствует, без исключения. Такое указание не оказывает какого-либо воздействия на обработку документа в каком-либо ведомстве.

Примеры корневого элемента с версией атрибута:

```
<request dtd-version="request-vl-1.dtd">
<request dtd-version="1.1">
```

Примеры вводной части с указанием DOCTYPE:

```
<?xml version="1.1" encoding="UTF-8"?>
<!DOCTYPE request SYSTEM "request-vl-1.dtd" >
```

```
<?xml version="1.1" encoding="UTF-8"?>
<!DOCTYPE request PUBLIC "-//WIPO//DTD REQUEST 1.1 2003-06-02//EN"
SYSTEM "request-vl-1.dtd" >
```

69. Необязательно: две инструкции по обработке в каждой записи документа указывают программное обеспечение, использовавшееся для создания документа, одна – для наименования программного обеспечения, вторая – для версии. Такие инструкции по обработке должны находиться во вводной части документа. См. на <http://www.w3.org/TR/2000/REC-xml-20001006#sec-pi> описание синтаксиса для инструкций по обработке.

```
<?software_name [наименование программного обеспечения]?>
<?software_version [версия программного обеспечения]?>
```

70. Если применяется таблица стилей, либо DTD, либо каждый документ, соответствующий DTD, должны указывать таблицу стилей, в соответствии с Рекомендациями по таблицам стилей в документах XML (W3C Recommendation Style Sheets with XML documents Version 1.0), опубликованными на <http://www.w3.org/TR/xml-stylesheet/>

Пример указания таблицы стилей:

```
<?xmlstylesheet href="..\stylesheet_factory\pap.xsl" type="text/xsl"?>
```

71. Формирование пробелов в записях документов может значительно улучшить читаемость необработанных данных. Например, отдельные документы могут включать обрыв строки после закрывающего тэга каждого параграфа для того, чтобы каждый параграф начинался с отдельной строки.

72. Стилевое оформление текста основано на допущении, что содержание находится между начальным и конечным тэгами. Хотя в некоторых случаях пробел уничтожается при обработке XML, другое программное обеспечение, используемое для обработки документов, возможно, не будет этого делать. Нужно избегать добавления лишних пробелов, не являющихся частью содержания. Содержание должно начинаться непосредственно после начального тэга и заканчиваться перед конечным тэгом.



Руководство по информации и документации в области промышленной собственности Например,

<postcode>20231</postcode>
лучше чем
<postcode> 20231 </postcode>.

Внешние объекты

73. Внешний объект – это любой объект, сопровождающий запись документа XML, ссылка на который указывается в документе. Внешние объекты являются неотъемлемой частью патентного документа. Без них документ XML не может быть успешно проанализирован, визуализирован или понят.

74. В случае патентных документов внешними объектами чаще всего являются страницы с рисунками, а также могут включаться встроенные изображения, программное обеспечение, таблицы, перечни последовательностей, неопределенные символы, или символические объекты. В большинстве случаев в качестве внешних объектов используются встроенные изображения, то есть ссылка на внешний файл с изображением вставляется в документ в том месте, где должно находиться изображение при визуализации документа.

75. Изображения в патентных документах могут быть полностью отсканированными страницами или так называемыми встроенными изображениями. Встроенные изображения обычно являются частями документа, которые нельзя закодировать и сохранить при помощи набора символов. Это могут быть рисунки, химические формулы, сложные таблицы, неопределенные символы и т.д. Неопределенные символы – такие символы, которые не определены набором символов или не доступны как ссылка на символический объект (см. выше). В настоящее время существует четыре типа данных, относящихся к изображениям, которые разрешены в рамках Приложения F в элементе ICE : TIFF, JPEG, [ST.33](#) и [ST.35](#); они рассмотрены ниже.

76. Следующие подразделы – руководство по использованию файлов с изображениями в форматах, предусмотренных Приложением F. Если ведомство планирует публиковать документы, в которых есть ссылки на внешние объекты с другим типом кодирования, или если ведомство отказывается от рекомендаций, перечисленных ниже, оно должно предоставить пользователям информацию, необходимую для правильного просмотра.

Стандарты – ST.36

Страница: 3.36.14

TIFF

77. Встроенные изображения должны быть включены в заголовок TIFF (тэговый формат файла с изображением) и помещаться за аннотацией, описанной в Приложении F (опубликовано на сайте ВОИС на www.wipo.int/pct/en/texts/index.htm). См. на <http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf> полное описание TIFF.

78. Рекомендованная схема кодирования данных для изображений в формате TIFF основана на технологии сжатия данных Modified READ II для факсимильного оборудования ITU-T (CCITT) Group 4, как описано в рекомендации ITU-T (CCITT) T.6, известной как Fax Group 4. См. Стандарт [ST.35](#), Приложения 3 и 4, для получения более подробной информации о Fax Group 4 и о возможном содержании информации в заголовке TIFF.

JPEG

79. Встроенные изображения могут быть включены также в заголовок JPEG, и помещаться за аннотацией, описанной в Приложении F (опубликовано на сайте ВОИС на www.wipo.int/pct/en/texts/index.htm). См. на <http://www.jpeg.org/> полное описание JPEG.

80. При использовании JPEG данный формат должен соответствовать ведомственным правилам и нормам. Например, РСТ допускает использование только черно-белых изображений (Правило РСТ 11.13: «Рисунки должны быть выполнены стойкими, черными, достаточно густыми и темными, однородно жирными и четкими линиями и штрихами без окрашивания»). Если ведомство допускает другие опции в JPEG, они должны быть определены и опубликованы.

Стандарт ВОИС ST.33



Руководство по информации и документации в области промышленной собственности

81. При форматировании внешних файлов с изображениями согласно Стандарту [ST.33](#) ведомство должно предоставлять пользователям ссылку на Стандарт ВОИС [ST.33](#) – Рекомендованный Стандартный Формат для Обмена Данными Факсимильной Информации Патентных Документов (*Recommended Standard Format for Data Exchange of Facsimile Information of Patent Documents*), см.: http://www.wipo.int/standards/en/part_03_standards.html.

Стандарт ВОИС ST.35

82. Если ведомство форматирует внешние файлы с изображениями согласно Стандарту [ST.35](#), оно должно предоставлять пользователям ссылку на Стандарт ВОИС [ST.35](#) - Рекомендуемый стандартный формат для обмена данными опубликованной патентной информацией смешанного типа на катушках или картриджах IBM 3480/90 (*Recommended Standard Format for Data Exchange of Mixed-Mode Published Patent Document Information on Reel-To-Reel and IBM 3480/90 Cartridge Tapes*); см. http://www.wipo.int/standards/en/part_03_standards.html

83. Данный стандарт относится только к использованию части ST.35, касающейся данных с изображениями, а не к остальным типам данных, описанным в Стандарте [ST.35](#).

PDF

84. Внешние объекты, являющиеся файлами PDF, должны располагаться за аннотацией, описанной в Приложении F (опубликовано на сайте ВОИС на www.wipo.int/pct/en/texts/index.htm). Для подробной информации о PDF см.: <http://www.adobe.com/products/acrobat/adobepdf.html>.

85. Если ведомство планирует публиковать документы, в которых есть ссылки на внешние объекты, использующие собственные кодировки, ведомство должно, как минимум, предоставить пользователям информацию, необходимую для правильного просмотра.

МЕГА СОДЕРЖАНИЕ

86. Содержание некоторых патентных заявок и публикаций по ним является настолько большим или объемным, что это мешает или серьезно ухудшает их стандартную машинную обработку (далее - " mega содержание"). Один из способов преодоления этого затруднения, используемый некоторыми ведомствами, - обработка mega содержания в качестве внешнего объекта. Данный раздел описывает, как применять такой подход для различных типов mega содержания.

87. Мега содержание в качестве внешнего объекта остается неотъемлемой частью документа (поданной заявки или публикации), но содержится в особом файле, который может быть проигнорирован при определенных типах обработки или обработан отдельно от документа, частью которого он является. Обычно на внешний объект дается ссылка из того места документа, где данный объект должен отображаться во время визуализации, т.е. из

Стандарты – ST.36

Страница: 3.36.15

логической позиции в документе. Данный раздел дает инструкции ведомствам, которые решили обрабатывать mega содержание как внешний объект.

88. Если некоторая часть содержания, например, запись последовательности или таблица, имеет размер больше допустимого предела, установленного ведомством по промышленной собственности, ведомство может обрабатывать mega содержание в качестве внешнего объекта. Например, ведомство может установить, что таблицы, занимающие 300 или более печатных страниц, будут обрабатываться как внешний объект. При публикации таблицы большого размера в качестве внешнего объекта данная таблица может быть свободно исключена из автоматического распечатывания, что помогает избежать ненужного и излишнего расходования бумаги экспертами или другими пользователями. Она может быть также исключена при загрузке, пока не будет специально запрошена заказчиком или экспертом, чтобы избежать лишней нагрузки на сети и серверы.

89. Для записи последовательностей Стандарт ВОИС [ST.25](#) определяет формат внешнего объекта.

90. В случае таблиц, для внешнего объекта следует использовать тот же формат XML, который используется для таблиц в записи документа. Немного измененное стандартное DTD для таблиц в этих целях: table-external.dtd

91. Для записи компьютерной программы форматом внешнего объекта обычно является простой текст ASCII со структурой, определенной синтаксисом языка компьютерной программы. Такой подход можно реализовать, поместив запись программы в одноячеичную таблицу, на которую дается ссылка из table-external-doc в записи table-external.dtd.



Руководство по информации и документации в области промышленной собственности

92. В случае химических структур или математических формул, когда ожидается, что мега содержание будет включать большое количество относительно небольших объектов, содержание следует включить в одну или более таблиц, а полученные таблицы затем обрабатывать как внешние объекты, используя `table-external.dtd`.

Стандартные DTD

93. Если содержание документа не относится только к области промышленной собственности, используют стандартные DTD. Такие DTD не включаются в международные DTD или в ведомственные DTD, они вводятся с помощью ссылки (примеры см. в `application-body.dtd`).

94. Для математических формул следует использовать MathML, версию 2. Полное описание см. на: <http://www.w3c.org/Math>.

95. Для таблиц следует использовать XML таблицу DTD OASIS. См. на <http://www.oasis-open.org/specs/tm9901.html> Технический меморандум OASIS TR 9901:1999 «*XML Exchange Table Model Document Type Definition*»..

96. Если ведомство намерено использовать другие стандартные DTD, должен быть представлен запрос на изменение.

97. Включение стандартных DTD рекомендуется только с помощью ссылки.

98. Такие стандартные DTD не могут быть изменены каким-либо образом при ссылке к ним в ведомственных DTD. Исключение – использование перезаписи для добавления атрибутов в корневой элемент таблицы DTD OASIS в DTD `xx-patent-publication`. Таблица DTD OASIS создана специально для создания локальных определений содержания ячейки. Другие исключения должны быть представлены в запросе на изменение.

Модельные DTD для патентных публикаций

99. Модельное DTD `xx-patent-document.dtd` не предназначено для использования в том виде, в котором оно опубликовано. Оно служит базисом, на основе которого каждое ведомство может с минимальными усилиями построить свое ведомственное DTD, максимально при этом используя общие элементы и логические структуры высокого уровня.

100. При изменении модельного DTD для патентных публикаций (`xx-patent-document.dtd`) следует придерживаться вышеуказанных правил для ведомственных имен. Ведомствам не рекомендуется вносить какие-либо изменения, не являющиеся необходимыми, т.к. это снижает способность к взаимодействию.

101. Ведомственные ограничения, не отраженные в DTD, упомянутых в данном описании, должны быть отражены в других доступных ведомственных описаниях.

ССЫЛКИ

102. Данный Стандарт ссылается на следующие стандарты и документы:

- (a) Стандарт ВОИС [ST.3](#) - Рекомендуемый стандарт на двухбуквенные коды для представления стран, административных единиц и межправительственных организаций.
- (b) Стандарт ВОИС [ST.9](#) – Рекомендации по библиографическим данным в патентных документах и свидетельствах дополнительной охраны (SPC) и относящихся к ним.
- (c) Стандарт ВОИС [ST.16](#) – Рекомендуемые стандартные коды для идентификации различных видов патентных документов.
- (d) Стандарт ВОИС [ST.25](#) – Стандарт по представлению перечней последовательностей нуклеотидов и аминокислот в патентных заявках.
- (e) Стандарт ВОИС [ST.32](#) - Рекомендации по разметке патентных документов с использованием SGML



Руководство по информации и документации в области промышленной собственности
(Standard Generalized Markup Language).

- (f) Стандарт ВОИС [ST.33](#) – Рекомендуемый стандартный формат для обмена факсимильной информацией по патентным документам
- (g) Стандарт ВОИС [ST.35](#) – Рекомендуемый стандартный формат для обмена данными о информации опубликованных патентных документов в смешанной форме на магнитных лентах в виде катушек или катриджей типа IBM 3480/90 (MMMT).
- (h) Административные Инструкции РСТ по Договору о Патентной Кооперации: Часть 7 – Инструкции Относительно Электронной Подачи и Обработки Международных Заявок
- (i) Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler. Extensive Markup Language (XML) 1.0 (Второе Издание), W3C Рекомендация, 4 февраля 2004 г., замененная 15 апреля 2004 г. См.: <http://www.w3.org/TR/2004/REC-xml11-20040204>
- (j) Международный Стандарт ISO/IEC 10646-1:1993(E): Information technology -- Universal Multiple-Octet Coded Character Set (UCS) -- Part 1: Architecture and Basic Multilingual Plane. International Organization for Standardization, Geneva, 1993 (Информационные Технологии – Универсальный Набор Символов, Закодированный Мультибайтовым Способом (UCS) – Часть 1: Структура и Основная Многоязычная Матрица. Международная Организация по Стандартизации, Женева, 1993 г.)
- (k) Международный Стандарт ISO/IEC 10646-2, Information technology -- Universal Multiple-Octet Coded Character Set (UCS) -- Part 2: Supplementary Planes. First edition, International Organization for Standardization, Geneva, 2001. (Информационные Технологии - Универсальный Набор Символов, Закодированный Мультибайтовым Способом (UCS) – Часть 2: Дополнительные Матрицы. Первое издание, Международная Организация по Стандартизации, Женева, 2001.)