

Комитет по стандартам ВОИС (КСВ)

Десятая сессия
Женева, 21–25 ноября 2022 года

ПРЕДЛОЖЕНИЕ О ПЕРЕСМОТРЕ СТАНДАРТА ВОИС ST.26

Документ подготовлен Международным бюро

РЕЗЮМЕ

1. Целевая группа по перечням последовательностей (Целевая группа по SEQL) предлагает внести поправки в текущую версию стандарта ВОИС ST.26 с целью исправления редакционных ошибок, выявленных Целевой группой, и добавления трех новых примеров в приложение VI. На девятой сессии Комитета по стандартам ВОИС (КСВ) Международное бюро приняло к сведению обеспокоенность ведомств в связи с внесением дальнейших изменений в стандарт ВОИС ST.26 и в этой связи подготовило рекомендации относительно подхода к внесению таких изменений.

СПРАВОЧНАЯ ИНФОРМАЦИЯ

2. На своей возобновленной четвертой сессии, состоявшейся в марте 2016 года, КСВ принял стандарт ВОИС ST.26 (см. пункт 52 документа CWS/4BIS/16), озаглавленный «Рекомендуемый стандарт представления перечней нуклеотидных и аминокислотных последовательностей с использованием языка XML (расширяемого языка разметки)». Целевая группа по перечням последовательностей ответственна за внесение поправок в стандарт ВОИС ST.26 в рамках задачи №44, которая сформулирована следующим образом:

«Оказать поддержку Международному бюро, направляя ему информацию о запросах и мнениях пользователей о программном средстве для составления и проверки текста заявок в соответствии со стандартом ST.26; оказать поддержку Международному бюро при последующем пересмотре Административной инструкции к РСТ; и подготовить необходимые поправки к стандарту ВОИС ST.26».

3. Поправки в стандарт ВОИС ST.26 были одобрены КСВ на его пятой, шестой, седьмой и восьмой сессиях, в результате чего были подготовлены версии 1.1, 1.2, 1.3 и 1.4 соответственно (см. пункт 41 документа CWS/5/22, пункт 112

документа CWS/6/34, пункт 133 документа CWS/7/29 и пункт 49 документа CWS/8/24). Совсем недавно, на девятой сессии, состоявшейся в ноябре 2021 года, КСВ одобрил ряд редакционных и субстантивных изменений и опубликовал их в качестве последней на данный момент версии стандарта ВОИС ST.26 – версии 1.5 (см. пункт 63 документа CWS/9/25). В том числе были внесены изменения, направленные на сохранение соответствия требованиям консорциума «Международное сотрудничество баз данных о нуклеотидных последовательностях» (INSDC).

4. В ходе девятой сессии делегации выразили обеспокоенность по поводу потенциальных последствий внесения изменений в стандарт ВОИС ST.26 после того, как все государства-члены внедрили данный стандарт, в случае, например, необходимости перенастройки их ИТ-систем (см. пункт 62 документа CWS/9/25). Международное бюро сообщило, что представит рекомендации по данному вопросу на десятой сессии.

5. Ряд редакционных поправок были внесены в приложение VI и дополнение к приложению VII после того, как были выявлены несколько ошибок с отсылками применительно к идентификаторам последовательностей и неверными указаниями местоположения характеристик по итогам пересмотра, проведенного на девятой сессии. Эти поправки были внесены и опубликованы в мае 2022 года.

ПРЕДЛАГАЕМЫЕ ИЗМЕНЕНИЯ В СТАНДАРТ ВОИС ST.26

6. Целевая группа по перечням последовательностей подготовила предложение по пересмотру стандарта ВОИС ST.26 для рассмотрения и, при условии целесообразности, утверждения КСВ. Зачеркнутый текст указывает на изменения, в то время как выделенный жирным шрифтом текст обозначает добавление.

7. Предлагаемые изменения к основной части (текста на английском языке) стандарта ВОИС ST.26 можно резюмировать следующим образом:

- (a) В связи с вступлением в силу новой версии стандарта (версия 1.6) заменить первое предложение редакционного примечания следующим текстом:

«Комитет по стандартам ВОИС на своей десятой сессии постановил, что версия 1.6 стандарта ВОИС ST.26 вступит в силу 1 июля 2023 года. До этого момента следует продолжать использовать версию 1.5 указанного стандарта»;

- (b) Обновить пункт 44, чтобы модифицировать значение имени файла так, чтобы в нем вместо «_» использовался знак «-»: «*filename="US11-405455-SEQL.xml"*», так как некоторые ВИС не допускают использование знака подчеркивания «_» в имени подаваемых электронных файлов;

- (c) Обновить пункт 45, чтобы уточнить определение элемента ApplicationNumberText следующим образом: *The application number as provided by the office of filing (e.g., PCT/IB2013/099999)*. («Номер заявки, представленный ведомством подачи (например, PCT/IB2013/099999)»);

- (d) Обновить пункт 73 для обеспечения его соответствия описанию раздела 6 путем добавления следующей фразы во второй половине первого предложения: *Section 8 provides the exclusive listing of qualifiers, **and their specified value formats, if any**, for each amino acid sequence feature key.* («в разделе 8 содержится единый перечень квалификаторов, **и их указанные форматы значений, если таковые**

имеются, для каждого функционального ключа аминокислотной последовательности»);

- (e) Обновить пункт 76, указав на наличие определения термина *and one optional attribute* («и одного необязательного атрибута»);
- (f) Обновить пункт 76, включив в него новое определение атрибута *id* для зависимых от языка квалификаторов в свободном формате в качестве новой, четвертой строки приводимой таблицы. Описание выглядит следующим образом: *A qualifier with language-dependent free text value may be uniquely identified by using the optional XML attribute «id» in the element INSDQualifier (see paragraph 87(d)). The value of the «id» attribute must start with the letter «q» and continue with any positive integer. The value of an «id» attribute must be unique to one INSDQualifier element, i.e. the attribute value must only be used once in a sequence listing file* («Классификатор с зависящим от языка свободным текстом может быть уникально идентифицирован с помощью необязательного XML-атрибута «id» в элементе INSDQualifier (см. пункт 87(d)). Значение атрибута «id» должно начинаться с буквы «q», за которым должно следовать любое целое положительное число. Значение атрибута «id» должно быть уникальным для каждого элемента INSDQualifier, т.е. значение атрибута не должно использоваться в файле перечня последовательностей более одного раза»);
- (g) Обновить пункт 77, добавив пробел и знак " между *organism* («организм») и *for* («для») в первой строке: *...and "organism" for amino acid sequences...* («...и «organism» для аминокислотных последовательностей...»);
- (h) Обновить пример 2 в пункте 79: исключить пробел перед *mol_type*;
- (i) Обновить пример в пункте 83: исключить пробел перед *mol_type*;
- (j) Обновить пункт 86 следующим образом: *The use of free text must be limited to a few short terms indispensable for the understanding of a characteristic of the sequence. For each qualifier other than the "translation" qualifier, the free text must not exceed 1000 characters* («Использование свободного текста должно быть ограничено несколькими короткими терминами, необходимыми для понимания характеристик последовательности. Для каждого квалификатора, **кроме квалификатора «translation»**, свободный текст не должен превышать 1000 знаков (символов)»);
- (k) Обновить начало пункта 87 (d), изменив его следующим образом: *For qualifiers with a language-dependent free text value, ...* («Для квалификаторов со значением свободного текста, зависящего от языка, ...»);
- (l) Поменять все словосочетания *industrial property* («промышленная собственность») на *intellectual property* («интеллектуальная собственность»), так как на девятой сессии КСВ было одобрено обновление с целью расширения сферы применения Справочника ВОИС; и
- (m) По всему тексту основной части стандарта унифицировать ссылку на международные, национальные или региональные процедуры именно в таком порядке.

8. Предлагаемые изменения в приложениях стандарта ВОИС ST.26 (текст на английском языке) можно резюмировать следующим образом:

- (a) Приложение I, раздел 6: обновить таблицу 5, включив *frequency* («частота») в качестве квалификатора со значением свободного текста, зависящего от языка;
- (b) Приложение I, раздел 6.21: обновить обязательный формат значения, указав, что квалификатор *frequency* («частота») зависит от языка;

- (c) Приложение I, раздел 6: обновить название таблицы 5 следующим образом: *List of qualifiers for nucleotide sequences with language dependent free text values* («Список значений квалификаторов для нуклеотидных последовательностей с зависимыми от языка значениями свободного текста»);
- (d) Приложение I, раздел 6 и 8: обновить примечание, добавив слово *international* («международных»), так как в настоящее время в нем указано только следующее: *national or regional procedures* («национальных или региональных процедур»);
- (e) Приложение I, раздел 6.77: убрать в примере пробелы после значения антикодона, но перед символом «меньше»;
- (f) Приложение I, раздел 8: обновить примечание, чтобы оно соответствовало примечанию в разделе 6. Обновленное примечание будет начинаться следующим образом: *Any qualifier value provided for a qualifier with a language dependent "free text" value format may require translation...* («Любое значение квалификатора, предусмотренное для квалификатора с зависимым от языка значением формата free text («свободный текст»), может потребовать перевода...»);
- (g) Приложение I, раздел 6 и 8: во всех квалификаторах с обязательным форматом значения, в котором указано, что они зависят от языка, отныне указывается, что перевод требуется для «международных/национальных/региональных процедур»; для этого нужно добавить формулировку «международные»;
- (h) Приложение I, раздел 6 и 8: в начале каждого из разделов указать, что английская формулировка *'language dependent "free text"'* изменена следующим образом *'language-dependent free text'* для придания единообразия тексту стандарта;
- (i) Приложение I, раздел 6 и 8: изменить название второго столбца таблицы, заменив слово *value* («значение») словом *qualifier* («квалификатор»), в результате чего соответствующее название будет выглядеть так: «Квалификатор свободного текста, зависящего от языка»;
- (j) Приложение II: включить в определение типа документа (DTD) следующее замечание для разъяснения того, какое значение должно быть представлено в атрибуте «*fileName*»: *By default this will be set to the value provided for the project name in WIPO Sequence. If the value is identical to the actual ST.26 XML filename, it should be noted that Offices may enforce their requirements for the filename used which may restrict which characters are allowable for submitted electronic files. It is also acceptable for the value of the filename attribute and the actual file name to be different. Please refer to the WIPO Sequence and ST.26 Knowledge Base for further details on Offices' naming conventions for electronic files* («По умолчанию это значение будет соответствовать значению названия проекта в WIPO Sequence. Следует отметить, что, если значение идентично фактическому названию файла XML стандарта ST.26, ведомства могут применять свои требования к используемому названию файла, согласно которым могут вводиться ограничения в отношении допустимых для представленных электронных файлов символов. Кроме того, допускаются разные значения атрибута «название файла» и фактического названия файла. Дополнительная информация о правилах ведомств по присвоению имен электронным файлам содержится в Базе знаний по WIPO Sequence и стандарту ST.26»);
- (k) Приложение VI, пример 3(c)-1, вопрос 3: исправить термин «ацилирование» на «ацетилирование»;
- (l) Приложение VI, пример 3(k)-2, вопрос 3: исправить ссылку на SEQ ID NO. 12 в третьем абзаце следующим образом: *the feature key "misc_difference" with feature location "10" should be used together with two "replace" qualifiers where the value for one would be "gk" and the second would be "c"* («...следует использовать

функциональный ключ «misc_difference» с местоположением «10» совместно с двумя квалификаторами «gerplace», где значение для одного будет «gк» и значение для второго будет «с»»);

- (m) Приложение VI, пример 7(b)-2: удалить первое из двух изображений;
- (n) Приложение VI, пример 7(b)-2, вопрос 3: заменить термин «перекрестная связь» термином «разветвленная структура»;
- (o) Приложение VI, пример 28: исправить SEQ ID 46 с тем, чтобы это значение соответствовало остаткам, показанным в соответствующей последовательности в дополнении к приложению VI;
- (p) Приложение VI, пример 7(b)-4 и 7(b)-5: исправить SEQ ID NO. 89, SEQ ID NO.90 и NO. 91 на SEQ ID NO. 90, SEQ ID NO. 91 и SEQ ID NO. 92;
- (q) Приложение VI и дополнение к приложению VI: добавить серию из трех новых примеров: пример 94-2, пример 30-2 и пример 12-1. Новые примеры в полном виде приведены в качестве приложения I к настоящему документу;
- (r) Приложение VII, сценарий 4: заменить первое предложение следующим текстом: *For both nucleotide sequences and amino acid sequences, ST.26 has the mandatory feature key "source" with two mandatory qualifiers* («Как для нуклеотидных последовательностей, так и для аминокислотных последовательностей, стандарт ST.26 предусматривает обязательный ключ характеристики source с двумя обязательными квалификаторами»);
- (s) Annex VII, сценарий 4: заменить текст подписи к таблице «Нуклеотидные последовательности» на следующий: *ST.26 - feature key 5.37 source; mandatory qualifier 6.39 mol_type (see ST.26 paragraph 75)* (ST.26 – функциональный ключ 5.37 «source»; обязательный квалификатор 6.39 «mol_type» (см. ST.26 пункт 75));
- (t) Приложение VII, сценарий 4: заменить текст подписи к таблице «Аминокислотные последовательности» на следующую: *ST.26 - feature key 7.30 source; mandatory qualifier 8.1 mol_type (see ST.26 paragraph 75)* (ST.26 – функциональный ключ 7.30 «source»; обязательный квалификатор 8.1 «mol_type» (см. ST.26 пункт 75);
- (u) Приложение VII, сценарий 8: обновить строку 18 приводимой таблицы, добавив в четвертую колонку следующий текст: **TATA_signal and (if <223> present):»** (**TATA_signal и (если есть <223>):**), а также исключить из примечания следующий текст: «(if <223> present)» ((если есть <223>));
- (v) Приложение VII, сценарий 8: заменить второе предложение в сноске следующим текстом: *In such a case, in order to avoid addition of subject-matter that may lead to partial loss of priority, it is recommended to include the more limited term "TATA_signal" in a "note" qualifier as shown in the above table (item N° 18). If in rare cases the Applicant considers that the use of the "TATA_box" value for the "regulatory_class" qualifier is not appropriate, the value:"other" may be used instead of "TATA_box". In this case, the term "TATA_signal" must be included in a "note" qualifier associated to the "regulatory" feature key* («В таком случае во избежание добавления объекта (признака), которое может привести к частичной потере приоритета, рекомендуется указывать в квалификаторе "note" более узкий термин "TATA_signal", как показано в таблице (графа 18). Если в редких случаях заявитель полагает, что использование значения "TATA_box" для квалификатора "regulatory_class" нежелательно, он может указать вместо этого значение "other". В этом случае термин "TATA_signal" должен быть указан в квалификаторе "note", относящемся к функциональному ключу "regulatory"»);

- (w) Дополнение к приложению VI: любые соответствующие изменения, необходимые в связи с обновлением приложения VI. Дополнение к приложению VI приводится в приложении II к настоящему документу.

РУКОВОДЯЩИЕ ПРИНЦИПЫ ОБНОВЛЕНИЯ ВЕРСИИ СТАНДАРТА ST.26

9. Как указано выше, на девятой сессии КСВ Международное бюро отметило озабоченность относительно регулярных обновлений стандарта ВОИС ST.26 и их потенциального влияния (см. пункт 4 настоящего документа).
10. Следует отметить, что стандарт ВОИС ST.26 будет обновляться не чаще одного раза в год по мере необходимости и с одобрения КСВ. В предлагаемую пересмотренную редакцию, представленную на десятой сессии КСВ, были внесены как редакционные, так и субстантивные поправки. Редакционные изменения в основном тексте не должны оказывать влияния на пользователей, включая ВИС. Субстантивные изменения в приложениях VI и VII также не должны повлиять на ВИС и пользователей.
11. В будущих версиях стандарта ВОИС ST.26 Международное бюро будет указывать соответствующий номер версии, X.Y, что будет информировать пользователей стандарта о том, является ли данная версия обратно совместимой с предыдущей. Незначительные обновления стандарта ВОИС ST.26 будут обозначаться увеличением только второй цифры («Y»), такие версии будут считаться обратно совместимыми. Крупные обновления, обозначаемые увеличением первой цифры («X»), вероятно, будут оказывать влияние на ИТ-системы и могут потребовать их перенастройки. Номер версии определения типа документа (DTD) может отличаться от номера версии основного текста стандарта и других приложений.
12. Поскольку в новой версии никаких изменений к определению типа документа стандарта ВОИС ST.26 предложено не было, за исключением добавления замечания, и изменения, предлагаемые к основному тексту и приложениям стандарта, носят в основном редакционный характер или предназначены для разъяснения, номер версии DTD в данном случае останется прежним – 1.3, а номер версии самого стандарта будет обновлен до версии 1.6.
13. Если предлагаемая версия будет одобрена КСВ, Международное бюро, в сотрудничестве с Целевой группой по перечням последовательностей включит утвержденное содержание в стандарт ВОИС ST.26 и опубликует его в 2022 году.

14. КСВ предлагается:

(a) принять к сведению информацию, содержащуюся в настоящем документе и приложениях к нему; и

(b) рассмотреть и утвердить предлагаемые изменения к стандарту ВОИС ST. 26, которые изложены в пунктах 7 и 8 выше и приложениях I и II к настоящему документу.

[Приложение I (предлагаемые новые
примеры для приложения VI)
следует]

PROPOSED NEW EXAMPLES FOR ANNEX VI OF STANDARD ST.26

Paragraph 94 – Variant sequence disclosed as a single sequence with enumerated alternative residues

Example 94-2 – Representation of single sequence with enumerated alternative amino acids that may be modified amino acids

A patent application describes the following polypeptide:

Leu-Glu-Tyr-Cys-Leu-Lys-Arg-Trp-Xaa-Glu-Thr-Ile-Ser-His-Cys-Ala-Trp

where Xaa can be Ile, Ala, Phe, Tyr, alle, Melle, or Nle.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated peptide provides 16 specifically defined amino acids. Therefore, the sequence must be included in a sequence listing as required by ST.26 paragraph (7)(b).

Question 3: How should the sequence(s) be represented in the sequence listing?

The most restrictive ambiguity symbol that can encompass “Ile, Ala, Phe, Tyr, alle, Melle, or Nle” is “X”. Therefore, the sequence must be included in a sequence listing as:

LEYCLKRWXETISHCAW (SEQ ID NO: xx)

ST.26 paragraph 30 requires that “[a] modified amino acid must be further described in the feature table”. However, paragraph 30 does not require any specific feature key be used to describe modified amino acids. While paragraph 30 describes the use of feature keys “CARBOHYD”, “LIPID”, “MOD_RES”, and “SITE”, these feature keys are more appropriate for scenarios where the modified amino acid is not within a list of alternatives for a specific location. In this example, the feature key “VARIANT” satisfies the requirement of paragraph 30 since it allows for the inclusion of all of the alternatives for the variant site. So, the feature key “VARIANT” with the qualifier “note” “Ile, Ala, Phe, Tyr, alle, Melle, or Nle” as a qualifier value should be used to describe the variant site at position 9. The use of a second feature key such as “SITE” with a qualifier “note” may be used to further identify the modified amino acids found at position 9.

Relevant ST.26 paragraph(s): 3(a), 7(b), 27, 30, **94**, 96, and Annex I, Section 4, Table 4

Paragraph 30 – Annotation of a modified amino acid

Example 30-2 – Post-translationally modified amino acids

A patent application describes the following polypeptide:

Leu-Glu-Tyr-Cys-Leu-Lys-**Arg**-Trp-Glu-Thr-Ile-Ser-His

wherein the Arg at position 7 may be post-translationally deiminated to citrulline.

Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated peptide provides 13 specifically defined amino acids. Therefore, the sequence must be included in a sequence listing as required by ST.26 paragraph (7)(b).

Question 3: How should the sequence(s) be represented in the sequence listing?

According to ST.26 paragraph 29, a modified amino acid should be represented in the sequence as the corresponding unmodified amino acid whenever possible.

Therefore, the sequence should be included in a sequence listing as:

LEYCLKRWETISH (SEQ ID NO: xx)

where the symbol “R” is used to represent the arginine at position 7.

A further description indicating that the arginine at position 7 may be modified to citrulline is required. The modification of arginine to citrulline is a post-translational modification. Therefore, the feature key “MOD_RES” should be used together with the mandatory qualifier “note” to indicate that the arginine may be deiminated to form citrulline. The location descriptor in the feature location element is the residue position number of the modified arginine.

Relevant ST.26 paragraph(s): 3(a), 7(b), **30**, and Annex I, Section 7, Feature Key 7.18

Paragraph 12 – Circular nucleotide sequence

Example 12-1: Circular nucleotide sequence

A patent application contains the following figure, disclosing the DNA sequence of plasmid pCIRC1:



Question 1: Does ST.26 require inclusion of the sequence(s)?

YES

The enumerated nucleotide sequence has more than 10 specifically defined nucleotides. Therefore, the sequence must be included in a sequence listing as required by ST.26 paragraph (7)(a).

Question 3: How should the sequence(s) be represented in the sequence listing?

According to ST.26 paragraph 12, when nucleotide sequences are circular in configuration, the applicant must choose the nucleotide in residue position number 1. For the purposes of this example, the “a” residue identified by the arrow in the figure will be used as position 1. However, any residue may be chosen as position 1. With the residue indicated by the arrow as position 1, the sequence should be included in a sequence listing as:

atggataatgaagaagttaacgaagaatgtatgagattatcccccaagaacgctcgtgcgcatctggat
aaacatctaacaatcaaggttgacatgcatgaaatgaaaatgcatatacagttcagatgcttcctg
gatggaatacatcgcaaactactagggtttctcgaagagctacttttgaaacaagaaaatgtaccac
caatg (SEQ ID NO: xx)

The sequence should be further described using feature key “misc_feature” with a location of “212^1”, which indicates that the last residue in the sequence, position 212, is linked to residue 1. A “note” qualifier must be included with a value indicating that the molecule is circular.

Relevant ST.26 paragraphs: 7(a), 12, and Annex I, Section 5, Feature Key 5.15

[Приложение II (дополнение к приложению VI) следует]