Приложение 2

ОПРЕДЕЛЕНИЕ ТИПА ДОКУМЕНТА ДЛЯ ПЕРЕЧНЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ (DTD)

Версия 1.3
Пересмотр одобрен Комитетом по Стандартам ВОИС (КСВ) на своей одиннадцатой сессии 8 декабря 2023

```
<?xml version="1.0" encoding="UTF-8"?>
<!--Annex II of ST.26, Document Type Definition (DTD) for Sequence Listing

This entity may be identified by the PUBLIC identifier:
*****************************************************************************
****
PUBLIC "-//WIPO//DTD SEQUENCE LISTING 1.3//EN" "ST26SequenceListing_V1_3.dtd"
*****************************************************************************
****
* PUBLIC DTD URL

* http://www.wipo.int/standards/DTD/ST26SequenceListing_V1_3.dtd
*****************************************************************************

Revision of Annex II to WIPO Standard ST.26 is submitted for approval by the Committee
on WIPO Standards (CWS) at its tenth session.

*****************************************************************************
* CONTACTS
*****************************************************************************
xml.standards@wipo.int


*****************************************************************************
* NOTES
*****************************************************************************
The sequence data part is a subset of the complete INSDC DTD V.1.5 that only covers
the requirements of WIPO Standard ST.26.

*****************************************************************************
* REVISION HISTORY
*****************************************************************************
2022-11-25: Comment related to filename approved at CWS/10 (no update to version
number)
2021-11-05: Revised Version 1.3 approved at CWS/9 (small edits to the comments)
2020-05-20: Version 1.3 approved at CWS/8.
Changes:
- Optional originalFreeTextLanguageCode attribute added to to allow applicants to
indicate the language of the free text in the original sequence listing.
- Optional nonEnglishFreeTextLanguageCode attribute added to to allow applicants to
indicate the language of the free text provided in the element .
- Optional id attribute added to INSDQualifier to facilitate comparison of language -
dependent qualifier values between sequence listings.
- Optional element added to element to allow applicants to type language-dependent
qualifiers in a non-English Language with the characters set forth in paragraph 40(a)
of the ST.26 main body document.

2018-10-19: Version 1.2 approved at CWS/6.
Changes:
<INDSQualifier*> changed to <INDSQualifier+> for alignment with business eneds and
advice from NCBI (an INSDFeature quals element (if present) should have one or more
INDSQualifier elements).

2017-06-02: Version 1.1 approved at the CWS/5
```

```
Changes:
Comments added to <INSDSeq_length>, <INSDSeq_division> and <INSDSeq_sequence> to
clarify the reason of the differences between the INSDC DTD v.1.5 and ST26 Sequence
Listing DTD V1_1.
*********************************************************************************
2016-03-24: Version 1.0 adopted by the CWS/4Bis
2014-03-11: Final draft for adoption.
*********************************************************************************

ST26SequenceListing
*********************************************************************************
* ROOT ELEMENT
*********************************************************************************
-->
<!ELEMENT ST26SequenceListing  ((ApplicantFileReference |
(ApplicationIdentification,ApplicantFileReference?)),
EarliestPriorityApplicationIdentification?,(ApplicantName,
ApplicantNameLatin?)?,(InventorName,InventorNameLatin?)?,InventionTitle+,SequenceTotal
Quantity,SequenceData+) >
<!--The elements ApplicantName and InventorName are optional in this DTD to facilitate
the conversion between various encoding schemes-->
<!--originalFreeTextLanguageCode:
The language code (see reference in paragraph 9 to ISO 639-1:2002) for the single
original language in which the language-dependent free text qualifiers
(NonEnglishQualifier_value) were prepared.
-->
<!--nonEnglishFreeTextLanguageCode:
The language code (see reference in paragraph 9 to ISO 639-1:2002) for the language in
which the language-dependent free text qualifiers (NonEnglishQualifier_value)
currently correspond.
-->
<!--fileName:
By default the file name 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
--->
<!ATTLIST ST26SequenceListing
             dtdVersion  CDATA #REQUIRED
             fileName  CDATA #IMPLIED
             softwareName  CDATA #IMPLIED
             softwareVersion  CDATA #IMPLIED
             productionDate  CDATA #IMPLIED  >
             originalFreeTextLanguageCode CDATA #IMPLIED
             nonEnglishFreeTextLanguageCode CDATA #IMPLIED
>
<!--ApplicantFileReference
Applicant's or agent's file reference, mandatory if application identification not
provided.
-->
<!ELEMENT ApplicantFileReference  (#PCDATA) >

<!--ApplicationIdentification
Application identification for which the sequence listing is submitted, when
available.
-->
<!ELEMENT ApplicationIdentification  (IPOfficeCode?,ApplicationNumberText,
             FilingDate?) >

<!--EarliestPriorityApplicationIdentification
Identification of the earliest priority application, which Contains IPOfficeCode,
ApplicationNumberText and FilingDate elements.
-->
<!ELEMENT EarliestPriorityApplicationIdentification  (IPOfficeCode,
ApplicationNumberText,FilingDate?) >

<!—ApplicantName
```

```
The name of the first mentioned applicant in characters set forth in paragraph 40 a)
of the ST.26 main body document.
-->
<!--languageCode: Appropriate language code from ISO 639-1 – Codes for the
representation of names of languages - Part 1: Alpha-2
-->
<!ELEMENT ApplicantName  (#PCDATA) >
<!ATTLIST ApplicantName
                languageCode  CDATA #REQUIRED
>
<!—ApplicantNameLatin
Where ApplicantName is typed in characters other than those as set forth in paragraph
40 b), a translation or transliteration of the name of the first mentioned applicant
must also be typed in characters as set forth in paragraph 40 b).
-->
<!ELEMENT ApplicantNameLatin  (#PCDATA) >

<!--InventorName
Name of the first mentioned inventor typed in the characters as set forth in paragraph
40 a).-->
<!--languageCode: Appropriate language code from ISO 639-1 – Codes for the
representation of names of languages - Part 1: Alpha-2
-->
<!ELEMENT InventorName  (#PCDATA) >
<!ATTLIST InventorName
                languageCode  CDATA #REQUIRED
>
<!--InventorNameLatin
Where InventorName is typed in characters other than those as set forth in paragraph
40 b), a translation or transliteration of the first mentioned inventor may also be
typed in characters as set forth in paragraph 40 b).
-->
<!ELEMENT InventorNameLatin  (#PCDATA) >

<!--InventionTitle
Title of the invention typed in the characters as set forth in paragraph 40 a) in the
language of filing. A translation of the title of the invention into additional
languages may be typed in the characters as set forth in paragraph 40 a) using
additional InventionTitle elements. The title of invention should be between two to
seven words.
-->
<!--languageCode: Appropriate language code from ISO 639-1 - Codes
for the representation of names of languages - Part 1: Alpha-2
-->
<!ELEMENT InventionTitle  (#PCDATA) >
<!ATTLIST InventionTitle
                languageCode  CDATA    #REQUIRED
>
<!--SequenceTotalQuantity
Indicates the total number of sequences in the document.
Its purpose is to be quickly accessible for automatic processing.
-->
<!ELEMENT SequenceTotalQuantity  (#PCDATA) >

<!--SequenceData
Data for individual Sequence.
For intentionally skipped sequences see the ST.26 main body document.
-->
<!ELEMENT SequenceData  (INSDSeq) >
<!ATTLIST SequenceData
                sequenceIDNumber  CDATA    #REQUIRED  >

<!--IPOfficeCode
ST.3 code. For example, if the application identification is PCT/IB2013/099999, then
IPOfficeCode value will be IB.
-->
<!ELEMENT IPOfficeCode  (#PCDATA) >

<!--ApplicationNumberText
The application identification as provided by the office of filing (e.g.
PCT/IB2013/099999)
```

```
-->
<!ELEMENT ApplicationNumberText  (#PCDATA)
>
<!--FilingDate
The date of filing of the patent application for which the sequence listing is
submitted ST.2 format (paragraphs 7 (a) and 11) "CCYY-MM-DD", using a 4-digit calendar
year, a 2-digit calendar month and a 2-digit day within the calendar month, e.g.,
2015-01-31
-->
<!ELEMENT FilingDate  (#PCDATA) >
<!--****************************************************************************
* INSD Part
******************************************************************************

The purpose of the INSD part of this DTD is to define a customized DTD for sequence
listings to support the work of IP offices while facilitating the data exchange with
the public repositories.

The INSD part is subset of the INSD DTD v1.5 and as such can only be used to generate
an XML instance as it will not support the complete INSD structure.

This part is based on:

The International Nucleotide Sequence Database (INSD) collaboration.

INSDSeq provides the elements of a sequence as presented in the GenBank/EMBL/DDBJ-
style flatfile formats. Not all elements are used here.
-->
<!--INSDSeq
Sequence data. Changed INSD V1.5 DTD elements, INSDSeq_division and INSDSeq_sequence
from optional to mandatory per business requirements.
-->
<!ELEMENT INSDSeq  (INSDSeq_length,INSDSeq_moltype,INSDSeq_division,
             INSDSeq_other-seqids?,INSDSeq_feature-table?,INSDSeq_sequence) >

<!--INSDSeq_length
The length of the sequence.INSDSeq_length allows only integer.
-->
<!ELEMENT INSDSeq_length  (#PCDATA) >

<!--INSDSeq_moltype
Admissible values: DNA, RNA, AA
-->
<!ELEMENT INSDSeq_moltype  (#PCDATA) >

<!--INSDSeq_division
Indication that a sequence is related to a patent application. Must be populated with
the value PAT.
-->
<!ELEMENT INSDSeq_division  (#PCDATA) >

<!--INSDSeq_other-seqids
In the context of data exchange with database providers, the IPOs should populate for
each sequence the element INSDSeq_other-seqids with one INSDSeqid containing a
reference to the corresponding published patent and the sequence identification.
-->
<!ELEMENT INSDSeq_other-seqids  (INSDSeqid?) >

<!--INSDSeq_feature-table
Information on the location and roles of various regions within a particular sequence.
Whenever the element INSDSeq_feature-table is used, it must contain at least one
feature.
-->
<!ELEMENT INSDSeq_feature-table  (INSDFeature+) >

<!--INSDSeq_sequence
The residues of the sequence. The sequence must not contain numbers, punctuation or
whitespace characters.
-->
<!ELEMENT INSDSeq_sequence  (#PCDATA) >
```

```
<!--INSDSeqid
Intended for the use of IPOs in data exchange only.

Format:
pat|{office code}|{publication number}|{document kind code}|{Sequence identification
number}
where office code is the code of the IP office publishing the patent document,
publication number is the publication number of the application or patent, document
kind code is the letter codes to distinguish patent documents as defined in ST.16 and
Sequence identification number is the number of the sequence in that application or
patent

Example:
pat|WO|2013999999|A1|123456

This represents the 123456th sequence from WO patent publication No. 2013999999 (A1)
-->
<!ELEMENT INSDSeqid  (#PCDATA) >

<!--INSDFeature
Description of one feature.
-->
<!ELEMENT INSDFeature  (INSDFeature_key,INSDFeature_location,INSDFeature_quals?) >

<!--INSDFeature_key
A word or abbreviation indicating a feature.
-->
<!ELEMENT INSDFeature_key  (#PCDATA) >

<!--INSDFeature_location
Region of the presented sequence which corresponds to the feature.
-->
<!ELEMENT INSDFeature_location  (#PCDATA) >

<!--INSDFeature_quals
List of qualifiers containing auxiliary information about a feature.
-->
<!ELEMENT INSDFeature_quals  (INSDQualifier+) >
<!--INSDQualifier
Additional information about a feature.
For coding sequences and variants see the ST.26 main body document.
-->
<!--id
Unique identifier for the INSDQualifier to facilitate comparison of versions of a
sequence listing specifically having language-dependent qualifier values in different
languages.
-->
<!ELEMENT INSDQualifier  (INSDQualifier_name,INSDQualifier_value?,
NonEnglishQualifier_value?)>
<!ATTLIST INSDQualifier
        id ID #IMPLIED
>
<!--INSDQualifier_name
Name of the qualifier.
-->
<!ELEMENT INSDQualifier_name  (#PCDATA) >
<!--INSDQualifier_value
Value of the qualifier. Where the qualifier is language-dependent its value must be in
the English language and typed with the characters set forth in paragraph 40 (b).
-->
<!ELEMENT INSDQualifier_value  (#PCDATA) >
<!--NonEnglishQualifier_value
Value of a language-dependent qualifier in a language that is not English and typed
with the characters set forth in paragraph 40 (a). The language is indicated with the
attribute nonEnglishFreeTextLanguageCode.
-->
<!ELEMENT NonEnglishQualifier_value (#PCDATA)>
```

[Приложение 3 следует]