SEQUENCE LISTING XML STANDARD FOR THE PRESENTATION OF NUCLEOTIDE AND AMINO ACID SEQUENCE LISTINGS IN PATENT APPLICATIONS

* CONTACTS

Stephane Nauche (European Patent Office)
Raul Suarez Y Gonzales (European Patent Office)
Bruce B Cox (US Patent & Trademark Office)

Date draft created: 2012-02-16

* NOTES

The DTD for the general information part intentionally departs from WIPO Standards ST.36 and ST.96 to ensure management independently of those standards.

The purpose of the bibliographic information contained in the general information part is solely for association of the sequence listing to the patent application for which the sequence listing is submitted.

The sequence data part is composed of one or more sequence data elements containing information about one sequence. The sequence data elements include various feature keys and subsequent qualifiers agreed upon by the International Nucleotide Sequence Database Collaboration (INSDC) and the UniProt Consortium. The sequence data part is a subset of the complete INSDC DTD that only covers the requirements of WIPO ST.26.

VERSION 1.0

Initial Version of the DTD.

-->

<!-- ST26SequenceListing

* ROOT ELEMENT

-->

<!ELEMENT ST26SequenceListing ((ApplicantFileReference | (ApplicationNumber, ApplicationFilingDate )+, PriorityApplicationNumber?,EarliestPriorityDate?, ApplicantName,ApplicantNameCharacters?, InventorName,InventorNameCharacters?, InventionTitle,InventionTitleCharacters?, SequencesTotalNumber,SequenceData+) >
<!ATTLIST ST26SequenceListing
dtdVersion CDATA #IMPLIED
fileName CDATA #IMPLIED
softwareName CDATA #IMPLIED
softwareVersion CDATA #IMPLIED
productionDate CDATA #IMPLIED>

<!-- ApplicantFileReference
Applicant's or agent's file reference, mandatory if application number not provided.
-->
<!ELEMENT ApplicantFileReference (#PCDATA)>

<!-- ApplicationNumber
Application information, when available.
-->
<!ELEMENT ApplicationNumber (FilingOfficeCode,ApplicationNumberText)>

<!-- FilingOfficeCode
WIPO ST.3 code
-->
<!ELEMENT FilingOfficeCode (#PCDATA)>

<!-- ApplicationNumberText
The application number without filing office code.
-->
<!ELEMENT ApplicationNumberText (#PCDATA)>

<!-- ApplicationFilingDate
The date of filing of the application.
ISO8601 long format:
yyyy-mm-dd
-->
<!ELEMENT ApplicationFilingDate (#PCDATA)>

<!-- PriorityApplicationNumber
Application Number of the earliest claimed priority.
Also contains FilingOfficeCode and ApplicationNumberText elements,
see ApplicationNumber for details.
-->
<!ELEMENT PriorityApplicationNumber (FilingOfficeCode?,ApplicationNumberText)>

<!-- EarliestPriorityDate
Where the international application claims the priority of several earlier applications, the filing date of the earliest application whose priority is claimed must be indicated as the priority date.
ISO8601 long format:
yyyy-mm-dd
-->
<!ELEMENT EarliestPriorityDate (#PCDATA)>

<!-- ApplicantName
The name of the first mentioned applicant in the Latin Alphabet.
-->
<!ELEMENT ApplicantName (#PCDATA)>
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 v1.4 and as such can only be used to generate an XML instance as it will not support the complete INSD structure.

This part 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.

All mandatory elements must be populated.
For intentionally skipped sequences please refer to ST.26.

The following elements are not necessary for ST.26 and were removed:

- INSDSeq_locus
- INSDSeq_strandedness
- INSDSeq_topology
- INSDSeq_update-date
- INSDSeq_create-date
- INSDSeq_update-release
- INSDSeq_create-release
- INSDSeq_definition
- INSDSeq_primary-accession
- INSDSeq_entry-version
- INSDSeq_accession-version
- INSDSeq_secondary-accessions
- INSDSeq_project
- INSDSeq_keywords
- INSDSeq_segment
- INSDSeq_source
- INSDSeq_organism
- INSDSeq_taxonomy
- INSDSeq_comment
- INSDSeq_primary
- INSDSeq_source-db
- INSDSeq_database-reference
- INSDSeq_contig
- INSDSeq_references

INSDSeq_sequence has been made mandatory

<!-- ELEMENT INSDSeq (INSDSeq_length, INSDSeq_moltype, INSDSeq_division,
INSDSeq_other-seqids?, INSDSeq_feature-table?, INSDSeq_sequence) -->

<!-- ELEMENT INSDSeq_length (#PCDATA) -->

<!-- ELEMENT INSDSeq_moltype (#PCDATA) -->

<!-- ELEMENT INSDSeq_division (#PCDATA) -->

<!-- ELEMENT INSDSeq_other-seqids -->
<!ELEMENT INSDSeq_other-seqids (INSDSeqid*) >

<!-- INSDSeqid
Intended for the use of Patent Offices in Data Exchange only.

Format:

PAT][{country code}|{publication number}|{Sequence identification number}

-->  
<!ELEMENT INSDSeqid (#PCDATA) >

<!-- INSD_seq_feature-table
Whenever the element INSDSeq_feature-table is used, it must contain at least one feature.

-->  
<!ELEMENT INSDSeq_feature-table (INSDFeature+) >

<!-- INSDFeature
The following elements were removed from the feature for the purpose of ST26

   INSDFeature_intervals,
   INSDFeature_operator,
   INSDFeature_partial5,
   INSDFeature_partial3

-->  
<!ELEMENT INSDFeature (INSDFeature_key,INSDFeature_location,INSDFeature_quals?) >

<!-- INSDFeature_key

-->  
<!ELEMENT INSDFeature_key (#PCDATA) >

<!-- INSDFeature_location

-->  
<!ELEMENT INSDFeature_location (#PCDATA) >

<!-- INSDFeature_quals

-->  
<!ELEMENT INSDFeature_quals (INSDQualifier*) >

<!-- INSDQualifier

Coding sequences and Variants see ST.26.

-->  
<!ELEMENT INSDQualifier (INSDQualifier_name,INSDQualifier_value?) >

<!ELEMENT INSDQualifier_name (#PCDATA) >

<!ELEMENT INSDQualifier_value (#PCDATA) >

<!-- INSDSeq_sequence

-->  
<!ELEMENT INSDSeq_sequence (#PCDATA) >