INTERNATIONAL STANDARD

ISO 26162-2

First edition
2019-11

Management of terminology resources — Terminology databases —

Part 2: **Software**

Gestion des resources terminologiques — Bases de données terminologiques —

Partie 2: Logiciels

Reference number
ISO 26162-2:2019(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 3, *Management of terminology resources*.

This first edition of ISO 26162-2, together with ISO 26162-1, cancels and replaces ISO 26162:2012, which has been technically revised.

The main changes compared to the previous edition are as follows:

— the document has been split into parts. The first part is focusing on the design of terminology database design, the second part on the development of terminology management systems;

— all references to generic software design principles and specific use cases have been removed.

A list of all parts of the ISO 26162 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Terminologies are the totality of concepts in given subject fields represented by terms and other designations and described by using additional terminological data. In general, these data are organized in structured terminology databases and are usually manipulated in specific software applications called terminology management systems. Terminology databases usually vary with regard to their underlying data model and consist of different sets of data categories, while terminology management systems generally differ depending on their functionality and the platform they are designed for.

The ISO 26162 series gives guidance on designing terminology databases and on essential terminology management system features. The series can also be used to evaluate the conformance and suitability of terminology databases and terminology management systems.

Management of terminology resources — Terminology databases —

Part 2:
Software

1 Scope

This document specifies essential features of terminology management systems, regardless of specific software engineering paradigms, user interface and user assistance design principles, and specific data models. These features enable maximum efficiency and quality in terminology work and, thus, support creating, processing, and using high quality terminology. The intended audiences of this document are software engineers/developers as well as terminologists, technical communicators, translators, interpreters, language planners, and subject field experts.

This document describes all features needed for recording, editing, maintaining, exchanging, and presenting terminological data. Term extraction features used to identify new terms are out of the scope of this document.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 704, *Terminology work — Principles and methods*

ISO 1087, *Terminology work — Vocabulary*

ISO 8601 (all parts), *Date and time — Representations for information interchange*

ISO 12620, *Management of terminology resources — Data category specifications*

ISO 16642, *Computer applications in terminology — Terminological markup framework*

ISO 23185, *Assessment and benchmarking of terminological resources — General concepts, principles and requirements*

ISO 26162-1:2019, *Management of terminology resources — Terminology databases — Part 1: Design*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 1087, ISO 26162-1 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
terminology management system
TMS
software tool specifically designed with a metadata structure for collecting, maintaining, and accessing terminological data

[SOURCE: ISO 1087:2019, 3.6.13]

3.2
concept entry structure
part of a terminology database that defines the structure for concept entries

3.3
input template
template that is created based on a *concept entry structure* (3.2) (or a subset of it) for the purpose of data input

Note 1 to entry: Depending on the degree of differentiation of user roles, more than one input template can be necessary.

3.4
layout template
template that defines how concept entries are displayed in the user interface

Note 1 to entry: Depending on the degree of differentiation of user roles, more than one layout template can be necessary.

3.5
terminology database definition
database information that defines the data categories and the *concept entry structure* (3.2) of a terminology database

Note 1 to entry: Depending on the degree of management requirements, more than one terminology database definition can be necessary.

4 Terminology management systems

4.1 General

The design of terminology management systems requires a deep understanding of terminology theory and terminology work. In this sense, and in order to achieve high quality results, the following shall be used:

— established terms and definitions as defined in ISO 1087;

— principles and methods as defined in ISO 704;

— data modeling criteria as defined in ISO 16642 and ISO 12620;

— usability metrics as defined in ISO 23185.

Terminology databases are used for collecting terminologies. Terminologies can achieve a high level of structural complexity and can be created and used by different user groups, each of them with their own specific needs. Consequently, terminology management systems shall be designed to respond to a wide range of potential use cases that can differ from those of conventional database management systems, particularly with regards to human-machine interaction and the interaction with other terminology-processing systems, such as translation memory systems, localization tools or authoring tools.

4.2 Terminological metamodel

Every terminology database shall comply with the terminological metamodel defined in ISO 16642 (for more information see also ISO 26162-1). Thus, terminology management systems shall allow for the instantiation of every level of the terminological metamodel, from high-level containers (global and complementary information) down to the term component section.

4.3 Creating terminology databases

4.3.1 Terminology database definition

When creating terminology databases, the terminology management system shall support the following essential features to allow users to perform the following actions:

— assign a name to a terminology database;

— describe a terminology database (content, rights, copyright, etc.);

— provide a core concept entry structure based on the terminological metamodel (see 4.2) including the option to add/remove the term component section;

— use predefined concept entry structures;

— set up customizable concept entry structures;

— aggregate data categories at the core concept entry structure and further nest these data categories;

— explicitly aggregate data categories[1)], such as /definition/, /source/ or /note/ at the language level;

— replicate part of the concept entry structure when setting up a terminology database (for example, copy a set of data categories from the French language section for creating an identical data category set for the Spanish language section);

— repeat data categories (see ISO 26162-1:2019, 4.4.3);

— define the use (e.g. mandatory, see ISO 26162-1:2019, 4.3) and the cardinality of data categories (see 4.3.3);

— preview terminology database definitions;

— extend, condense and/or otherwise change terminology database definitions.

4.3.2 Data categories

Terminological data in specific data fields are instances of terminological data categories that have been previously defined during the terminology database setup and then have been associated with the corresponding levels of the terminological metamodel (for data categories see ISO 26162-1:2019, 4.3). The terminology management system shall also allow parent-child relationships between data categories, such as the relationship between /definition/ and /source/ illustrated in the following example:

EXAMPLE

Concept entry identifier: 3 (for concept *circuit-breaker*)

Concept position: 1.2

1) In this document, data category names used in running text are indicated between forward slashes (for example, /definition/).

Definition: mechanical switching device, capable of making, carrying and breaking currents under normal circuit conditions and also making, carrying for a specified duration and breaking currents under specified abnormal circuit conditions such as those of short circuit[2)]

Source (of definition): Electropedia

During terminology database setup, terminology management systems shall provide a predefined set of data categories and picklist values from established repositories and standards in order to support a high grade of interoperability between terminology databases (and other applications and resources). For example, common terminological data categories are specified in DatCatInfo (see Reference [10]), language codes are defined in ISO 639 (all parts), country codes are defined in ISO 3166 (all parts), and subject field codes can be taken from universal classification systems, such as the Universal Decimal Classification. However, terminology management systems shall also allow for the creation of customizable data categories and picklist values if appropriate data categories or picklist values are not available. If predefined data categories and picklist values are implemented for use in multilingual environments, these data categories and picklist values shall be made available in the desired user interface language.

4.3.3 Data types and cardinalities

In terminology, free text strings (such as terms, definitions or contexts), picklists, binaries (such as figures, audio or video), cross-references (such as between concept entries or links to external resources), and Boolean values are the most common data types. However, in order to cover any potential use case, terminology management systems should support other data types like those defined for XML Schema, for example, or other schema or programming languages.

Furthermore, terminology management systems shall allow users to configure cardinalities, i.e. to set a specific number of occurrences for any data category, for instance, to allow for only one definition at the concept entry level.

4.3.4 Roles and rights

Distributed terminology work usually requires the assignment of different roles and rights (see also ISO 26162-1:2019, 4.3.2.3 and 4.5). Specific terminology database objects (such as terminology database definitions, input templates, export/import profiles, filters) or instances (such as concept entries, certain language and/or certain term sections) require specific roles and rights. Thus, appropriate roles and rights can be granted depending on:

— authorization types, such as supervisor, creator, proposer, end user;

— work profiles, such as author, translator, service and marketing staff.

4.4 Data input and editing

In the context of data input and editing, terminology management systems shall support fundamental features to allow users to perfom the following actions:

— use predefined input templates;

— set up customizable input templates;

— copy parts of input templates for reuse;

— preview input templates during setup;

— copy concept entries (or parts of concept entries) to be used as a basis for the creation of new concept entries (or for the modification of concept entries, for example, by inserting a new language section copied from an adjacent concept entry);

2) Definition taken from Electropedia, http://www.electropedia.org

— automatically generate administrative data, such as authorship for creating and updating a concept entry or an instance of it, as well as concept entry identifiers and time/date stamps, according to ISO 8601 (all parts);

— set certain values for certain data fields when creating an input template. For example, /part of speech/ could have the default value "noun" if it is predetermined that most terms in the terminology database will be nouns. This can increase productivity by not requiring users to select the value for /part of speech/ for every term. However, users shall have the option to change these values at any time. To prevent unintentional contamination of data, it is recommended to prompt users to review their default values when starting a new session;

— automatically propagate changes to data elements throughout the whole terminology database or instances of it. For example, data fields can contain misspelled words or inconsistent spelling that need to be tackled by search and replace features. The need for search and replace can also apply to picklist values once they have been changed in the terminology database definition;

— add links to other concept entries, such as a link to a concept entry for an entailed term that is used in a definition or other text field, even if the term in the text does not have the same surface form as the term that designates a concept in another concept entry. For example, it shall be possible to link the plural form "circuit-breakers" in a definition to the concept entry that contains the term "circuit-breaker" as a designation of the corresponding concept. Terminology management systems shall offer the option to establish links between concept entries using terms or concept entry identifiers;

— add links to external resources (multimedia files, web pages, etc.);

— define as many languages as needed (multilingual environments);

— save a search term as candidate for a new concept entry when a search has returned no results;

— alert the user when trying to introduce a term that already exists. However, it shall be possible to enter duplicate terms in order to accommodate the existence of homographs;

— save partially completed concept entries in draft mode, and then allow additional information to be added as it becomes available;

— perform bulk editing and deletion of concept entries.

4.5 Searching and filtering

Most typical users are interested in finding information about a term or other kind of terminological information within concept entries or throughout the whole terminology database. In particular, terminology management systems shall allow users to:

— look up the definition of a concept;

— search for a target language equivalent;

— confirm a term's spelling or check the contexts in which it is used;

— look up whether a term has abbreviations or synonyms or look up the full form of an abbreviation;

— display all terms representing a concept including their status;

— find information about a concept and its relationships to other concepts;

— use all languages as search languages;

— apply predefined search and filter criteria;

— carry out complex searching or filtering for data management purposes, for example, to find concept entries that have been updated on a specific date or by a specific person, to find duplicate concept entries or to find concept entries where specific fields are empty.

Most searches are carried out on term fields. However, search and filter mechanisms shall be available for all data fields for all languages defined in the terminology database definition. The following list illustrates some search and search supporting mechanisms:

— **exact match**: the system displays all terms that exactly match the term being searched;

— **autocomplete**: the user starts typing a search term and the system suggests a complete term (or a list of complete terms);

— **automatic truncation**: the user types the beginning of a search term and the system displays all terms that start with the typed string;

— **explicit truncation**: the user includes one or more wildcard symbols such as the asterisk ("*", which represents zero or more characters) or the question mark ("?", which represents one character) in the search term to specify the position where truncation is to be applied. Wildcards can occur at the beginning (*circuit-breaker, searching for all terms that end with "circuit-breaker"), in the middle (locali?ation, searching for "localization" and "localisation") or at the end of the search string (vacuum*, searching for all terms starting with "vacuum");

— **regular expression**: users can create search patterns such as "m(asc)?\." to catch disallowed gender forms (in the present example "masc." and "m.") and replace them with a permitted value ("masculine", for example);

— **fuzzy match**: the system displays all terms that are similar to the search term. This mechanism allows the retrieval of orthographic variants, inflected or derived forms of terms;

— **single-word search and permutation**: the system displays all terms that contain single parts (words) of a multiword search term or all parts of the search term, possibly in a different word order;

— **ignoring special characters and capitalization**: the system displays matching terms, ignoring differences in case, or the presence of hyphens, blank characters or diacritics.

The search function shall be capable of retrieving all types of terms: the full form, abbreviated forms of the term, orthographic, stylistic and regional variants.

Terminology management systems in which concept relations or concept systems are recorded shall feature non-alphabetical systematic search features. Systematic search includes searching by the position of the concept within a concept system, or by one of its characteristics, or by navigating through a graphical representation of the concept system.

Searching by using the concept entry identifier is helpful. The concept entry identifier is often recorded in reports that are generated to check the consistency of terminology databases or when data are imported or exported. Some users need to search terminology databases by concept entry identifiers in order to address issues identified by these reports. For this purpose, concept entry identifiers shall be unique and stable in a given terminology database.

Filtering can be used to specify terminology database subsets. For instance, it is possible that users are only interested in:

— concepts from a given subject field;

— English terms that are only used in Canada;

— acronyms that are preferred by a specific company;

— concept entries that are not older than three years;

— a term with a definition taken from a specific document.

Terminology management systems shall allow for filtering and accessing any data field in a terminology database, and for combining query conditions with logical operators, such as AND, OR and NOT. For example, users can be interested in searching for:

— concept entries that have a source term, but NOT terms in one or more target languages;

— concept entries without a definition AND status approved.

Furthermore, it shall be possible to expand Boolean operators by additional operators, such as "greater than", "less than" or other expressions such as "contains", "starts with", "ends with", and so forth. For example, it can be useful to filter all concept entries that were created after or before a certain date, or all concept entries that do or do not include a definition.

Terminology management systems shall allow users to browse through the terminology database without searching for anything in particular (thus without specifying any search criteria) or allow users to specify and refine search criteria. Hence, terminology management systems shall allow users to:

— browse the whole terminology database by hand;

— browse a list of search results;

— refine a search/filter on a given result.

Depending on the data model and the terminological data stored in each concept entry, browsing can be alphabetical or systematic using the underlying ontology.

If searching does not provide the desired results, a free text search can help. Therefore, the system shall allow for free text search in all data fields at a time (definitions, contexts, notes, even in data fields with picklist values).

4.6 Sorting

Terminology management systems shall allow users to sort the results in a given order and according to the requirements of the languages used in the terminology database. It shall also be possible to sort according to several criteria, for example first on the basis of subject field and then alphabetically within each of these subject fields. In terminology databases that contain concept relations, it shall be possible to sort the output both alphabetically and systematically, i.e. according to the position of the concepts in the concept system.

4.7 Data validation

Data validation shall verify that each data field contains correct and error free data. Terminology management systems shall support data validation features, such as:

— check for duplicates to determine whether a term is stored more than once;

— check for forbidden terms to ensure the use of approved terms in /term/ and other open data categories;

— spell check and grammar check;

— character check to verify the character types used in an input field, for example Arabic terms using Arabic characters vs. Arabic terms using Latin characters;

— completeness check to ensure that data are present where required;

— format check to determine that data conform to a specified layout;

— plausibility check to determine that a value conforms to specified criteria, for example minimum length or standard formulations in definitions;

— link check to prevent invalid internal or external links;

— checking for unnecessary whitespace, for example hard line breaks, multiple whitespace, tabs, etc.

4.8 Data representation

Terminological data shall be displayed in accordance with user needs. In this sense, terminology management systems shall support essential features to:

— display characters and text correctly for all languages used in the terminology management system;

— provide predefined layout templates;

— allow users to create customized layout templates and to preview them during setup;

— display search and filter results in a user-friendly way. For example, when searching in several terminology databases, the search results shall be displayed in such a way that the provenance of the term can be easily identified. It is also required that the result list displays terms from adjacent concept entries and/or a list of alphabetically adjacent terms. It shall also be possible to open the concept entry for a given term from the search result list. If the search does not produce any results, the terminology management system should display alphabetically adjacent terms along with a message of the type "Term not found". All search filters should be accompanied by a number indicating the numeric results of the filter;

— allow the user to customize the display of search and filter results;

— allow users to expand and collapse data fields (such as "Figure", "Definition", etc.);

— allow users to open a target concept entry directly from a text field (for instance from a definition of a related concept);

— provide distinctly visible edit and view modes;

— provide for the visualization of concept relations and allow for navigation through the visualized concept system, even display full concept entries from the visual concept system.

4.9 Data exchange

Situations that can require data exchange range from sharing and reusing terminology to data maintenance, tool change, mergers and acquisitions or creation of backups.

In general, terminology management systems shall support:

— common predefined and customizable exchange and print formats (TBX, XML, CSV, XLSX, HTML, DOCX, PDF, etc.);

— filters for exporting and importing a selection of concept entries;

— filters for exporting and importing concept entries that contain a subset of the available data categories;

— customizable data mapping features;

— preview of export and import results.

For import, terminology management systems shall ensure the integrity of both the concept entries in the terminology database and those in the import file. A terminology management system shall typically offer the following import options:

1) add new concept entries to the terminology database;

2) add an import file concept entry to a terminology database concept entry;

NOTE 1 In options 1 and 2, a conflict can occur if a concept entry in the import file contains a term with the same surface form as a term in the terminology database. In that case user interaction can be needed.

3) replace a terminology database concept entry with an import file concept entry;

NOTE 2 If a concept entry in the import file contains a term that has the same surface form as a term in the terminology database, the concept entry in the terminology database is overwritten. This action also assumes that the two concept entries describe the same concept.

4) ignore the import file concept entry and keep the terminology database concept entry intact;

NOTE 3 If a concept entry in the import file contains a term that has the same surface form as a term in the terminology database, the concept entry in the import file is ignored and the terminology database concept entry remains unchanged. Usually, import file concept entries are ignored if two concept entries describe the same concept (but the information in the terminology database is more accurate) or if two concept entries describe different concepts. In these cases, the corresponding import file concept entries can be added as separate concept entries in subsequent import procedures.

5) replace data from the terminology database with imported data and synchronize on concept entry identifiers and perhaps other data categories such as /subject field/.

NOTE 4 The import file contains concept entry identifiers that correspond to the concept entry identifiers in the terminology database. This option is typically used to import data that have been previously exported from the terminology management system. For example, export and reimport are important issues if global changes or batch edits have to be made to a large numbers of concept entries by editing an exported file rather than attempting to do this directly in the terminology database. The imported concept entries replace the concept entries in the terminology database, but the system checks first to make sure that the concept entry identifiers, and any other specified data categories, are the same before making the replacement.

Terminology management systems shall provide facilities for interactive import, allowing the user to decide whether to import a complete concept entry or a subset of data categories (i.e., terms only or preferred terms together with definitions and the corresponding sources, for example). Such an interactive feature shall also identify possible duplicates during import and support the option to flag imported concept entries with status attributes (such as "proposed", "under review", approved", "deprecated").

4.10 Analysis and statistics

The analysis and statistics features of terminology management systems shall allow users to:

— version particular states of content of the whole terminology database or components thereof (concept entries, terms, definitions and so on together with authorship and time/date stamp);

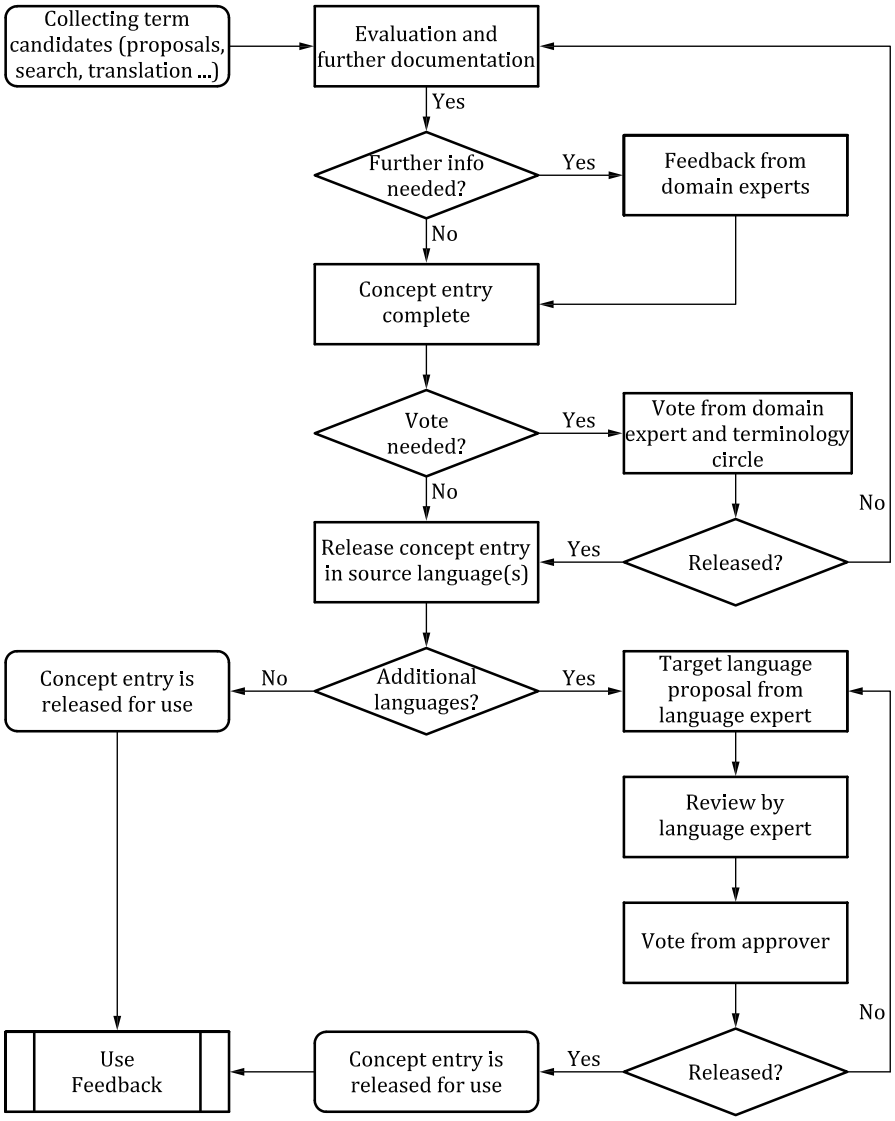
— create log files for all important actions such as concept entry creation and update, data validation, export/import (successfully exported/imported concept entries, ignored concept entries, synchronized concept entries) or deletion;

— record successful and unsuccessful queries (the latter can be used to add new terms to the terminology database, for example);

— report on the total number of concept entries, concept entries with definition(s), terms per language, abbreviations, authors, changes, search results, etc.

4.11 Workflow and object management

Distributed terminology management requires a sophisticated degree of labor division. Efficient management of data flow avoids data conflicts and helps to make an optimal use of resources and to provide high quality terminology. Data flow shall be expressed in data flowcharts that show the transitions among the different procedures. A typical terminology workflow is illustrated in Figure 1.



SOURCE: DTT 2014, M5-15. The original figure has been adapted with the permission of the authors.

Figure 1 — Typical terminology workflow

As terminology management systems usually interact with other applications, interfaces shall be implemented to ensure the use of terminology in integrated environments (such as content management systems, authoring tools, translation memory systems, localization tools, term extraction tools or term checkers). These interfaces shall allow interactions (such as search and paste terms into the aforementioned application or add terms to the terminology database from these systems) without having to leave these applications.

Another important aspect of terminology management system design includes the implementation of a comprehensive object management architecture. In this sense, terminology management systems shall support features that allow users to save, export and import all kinds of templates and results for reuse and archiving, such as:

— terminology database definitions;

— input and layout templates;

— validation procedures;

— search and filter criteria and results;

— customized mapping and import/export routines.

Furthermore, terminology management systems shall allow the user to open and use more than one terminology database at a time.

Bibliography

[1] ISO 639 (all parts), *Codes for the representation of names of languages*

[2] ISO 1087:2019, *Terminology work — Vocabulary*

[3] ISO 3166 (all parts), *Codes for the representation of names of countries and their subdivisions*

[4] ISO/IEC 10646, *Information technology — Universal Coded Character Set (UCS)*

[5] ISO 10241-1, *Terminological entries in standards — Part 1: General requirements and examples of presentation*

[6] ISO 12616, *Translation-oriented terminography*

[7] ISO 30042, *Management of terminology resources — TermBase eXchange (TBX)*

[8] Deutscher Terminologie-Tag e.V. (DTT) (2014): Terminologiearbeit – Best Practices 2.0. Cologne: SDK.

[9] European Union: EuroVoc. http://eurovoc.europa.eu.

[10] LTAC/TerminOrgs. DatCatInfo (DCR). http://www.datcatinfo.net.

[11] Introduction to TermBase eXchange (TBX). http://www.tbxinfo.net.

[12] Consortium U.D.C. Universal Decimal Classification (UDC). http://www.udcc.org.

ICS 01.020; 35.240.30

Price based on 12 pages