SLOVENSKI STANDARD
oSIST ISO/FDIS 24613-3:2021

01-marec-2021

Upravljanje jezikovnih virov - Ogrodje za označevanje leksikonov (LMF) - 3. del: Etimološka razširitev

Language resource management -- Lexical markup framework (LMF) - Part 3: Etymological extension

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Gestion des ressources linguistiques -- Cadre de balisage lexical (LMF) - Partie 3: Extension étymologique

oSIST ISO/FDIS 24613-3:2021
https://standards.iteh.ai/catalog/standards/sist/06cdd71b-2cfd-4a36-b407-8fe021caaed9/osist-iso-fdis-24613-3-2021

Ta slovenski standard je istoveten z: ISO/FDIS 24613-3

ICS:

01.140.20	Informacijske vede	Information sciences
35.240.30	Uporabniške rešitve IT v informatiki, dokumentiranju in založništvu	IT applications in information, documentation and publishing

oSIST ISO/FDIS 24613-3:2021 **en,fr**

2003-01.Slovenski inštitut za standardizacijo. Razmnoževanje celote ali delov tega standarda ni dovoljeno.

oSIST ISO/FDIS 24613-3:2021

iTeh STANDARD PREVIEW
(standards.iteh.ai)

oSIST ISO/FDIS 24613-3:2021
https://standards.iteh.ai/catalog/standards/sist/06cdd71b-2efd-4a36-b407-8fe021caaed9/osist-iso-fdis-24613-3-2021

oSIST ISO/FDIS 24613-3:2021

FINAL
DRAFT

INTERNATIONAL
STANDARD

**ISO/FDIS
24613-3**

ISO/TC **37**/SC **4**

Secretariat: KATS

Voting begins on:
2020-12-24

Voting terminates on:
2021-02-18

Language resource management — Lexical markup framework (LMF) —

Part 3: Etymological extension

Gestion des ressources linguistiques — Cadre de balisage lexical (LMF) —

Partie 3: Extension étymologique

iTeh STANDARD PREVIEW
(standards.iteh.ai)
oSIST ISO/FDIS 24613-3:2021
https://standards.iteh.ai/catalog/standards/sist/06cdd71b-2efd-4a36-b407-8fe021caaed9/osist-iso-fdis-24613-3-2021

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

Reference number
ISO/FDIS 24613-3:2020(E)

© ISO 2020

oSIST ISO/FDIS 24613-3:2021

iTeh STANDARD PREVIEW
(standards.iteh.ai)

oSIST ISO/FDIS 24613-3:2021
https://standards.iteh.ai/catalog/standards/sist/06cdd71b-2efd-4a36-b407-8fe021caaed9/osist-iso-fdis-24613-3-2021

COPYRIGHT PROTECTED DOCUMENT

© ISO 2020

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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

Published in Switzerland

© ISO 2020 – All rights reserved

oSIST ISO/FDIS 24613-3:2021

Contents

Page

Foreword ..... **iv**

1 Scope **1**

2 Normative references **1**

3 Terms and definitions **1**

4 The LMF etymology extension **2**
4.1 The Cognate class and the Etymon class 2
4.2 The Etymologizable class 2
4.3 The Etymology class and the EtyLink class 3
4.4 The CognateSet class 4
4.5 The Date class 4
4.6 The Gloss class 4

Annex A (informative) **Examples of possible etymological typologies** **6**

Annex B (normative) **Data categories for etymology description** **15**

Bibliography **22**

iTeh STANDARD PREVIEW
(standards.iteh.ai)

oSIST ISO/FDIS 24613-3:2021
https://standards.iteh.ai/catalog/standards/sist/06cdd71b-2efd-4a36-b407-8fe021caaed9/osist-iso-fdis-24613-3-2021

© ISO 2020 – All rights reserved

oSIST ISO/FDIS 24613-3:2021

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.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

This document was prepared by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 4, *Language resource management*.

oSIST ISO/FDIS 24613-3:2021
https://standards.iteh.ai/catalog/standards/sist/06cdd71b-2efd-4a36-b407-

This first edition of ISO 24613-3, together with ISO 24613-1, ISO 24613-2, ISO 24613-4 and ISO 24613-5, cancels and replaces ISO 24613:2008, which has been technically revised.

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

— entire revision of the content and its subdivision into several parts.

A list of all parts in the ISO 24613 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.

 © ISO 2020 – All rights reserved

oSIST ISO/FDIS 24613-3:2021

Language resource management — Lexical markup framework (LMF) —

Part 3: Etymological extension

1 Scope

This document describes an extension to ISO 24613-1 and ISO 24613-2 to support the development of detailed descriptions of common etymological phenomena and/or diachronic information with respect to lexical entries in born-digital and/or retro-digitized lexicons. It provides both a meta-model for such an extension as well as the relevant data categories.

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.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 8601-1, *Date and time — Representations for information interchange — Part 1: Basic rules*

ISO 8601-2, *Date and time — Representations for information interchange — Part 2: Extensions*

oSIST ISO/FDIS 24613-3:2021
https://standards.iteh.ai/catalog/standards/sist/06cdd71b-2eld-4a36-b407-8fe021caaed9/osist-iso-fdis-24613-3-2021

ISO 24613-1, *Language resource management — Lexical markup framework (LMF) — Part 1: Core model*

ISO 24613-2, *Language resource management — Lexical markup framework (LMF) — Part 2: Machine-readable dictionary (MRD) model*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 24613-1 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
cognate
form in a related language which shares a common etymological origin as a form in the language of the lexicon

3.2
etymologizable
meeting the conditions for having an *etymology* (3.3)

Note 1 to entry: "Etymologizable" is a category of lexical elements and usages (encompassing for instance lexical entries, senses, word forms).

3.3
etymology
origin and historical development of any aspect of a given lexical item

© ISO 2020 – All rights reserved

oSIST ISO/FDIS 24613-3:2021

3.4
etymon
lexical entry from which another lexical entry is derived

Note 1 to entry: An etymon can also be simply an earlier stage of a lexical item.

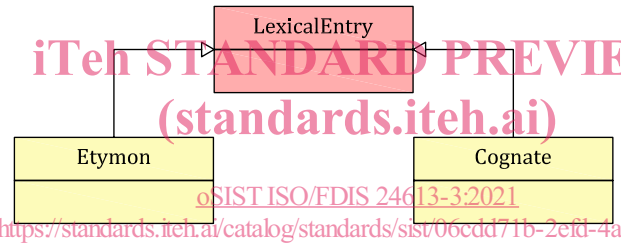
3.5
onomasiology
approach to the investigation of word meaning which takes a given concept as a starting point and studies the different lexical items in a language or languages that are used to refer to it

4 The LMF etymology extension

NOTE See Annex A for examples of possible etymological typologies.

4.1 The Cognate class and the Etymon class

Cognate and Etymon are defined as subclasses of the LexicalEntry class from the LMF core module (see Figure 1)[1)]. Both classes define lexical entries which have been added to a lexical resource with the purpose of describing the etymologies of one or more other lexical entries. Instances of either Etymon or Cognate can be assigned a language which is different from the language of the lexicon as a whole (this is specified in the LexiconInformation class as described in ISO 24613-1).



iTeh STANDARD PREVIEW
(standards.iteh.ai)
oSIST ISO/FDIS 24613-3:2021
https://standards.iteh.ai/catalog/standards/sist/06cdd71b-2efd-4a36-b407-8fe021caaed9/osist-iso-fdis-24613-3-2021

Figure 1 — Cognate and Etymon as subclasses of LexicalEntry

Individuals of both the Etymon and the Cognate classes shall be in an aggregation relationship with at least one individual of type EtyLink (see 4.3). When describing etymologies, there are cases in which it is necessary to deal with instances of LexicalEntry (and hence also by the subclass relation instances of Etymon and Cognate) which are roots, and in particular reconstructed roots. In these cases, the fact of being a root and the type of the root in question shall be specified using the attribute rootType. In the case of reconstructed roots or other word forms, the attribute status serves to associate the element with a written description of the likelihood of its having been in use (see the example in A.8). See Table 1 for a list of attributes to be used with these two classes.

4.2 The Etymologizable class

The Etymologizable class provides a means of referring to the set of linguistic elements that can have etymologies. By defining a single class encompassing all such 'etymologizable' elements, the classes of elements which can have etymologies can be easily extended in the future wherever the necessity arises. The following classes are subtypes of the Etymologizable class (see Figure 2): LexicalEntry, Sense, Form and CognateSet (see 4.4).

1) In this document, the following colour scheme is used in diagrams: classes in yellow are introduced in this document, and classes in pink have been previously introduced in ISO 24613-1 and ISO 24613-2.

 © ISO 2020 – All rights reserved

oSIST ISO/FDIS 24613-3:2021

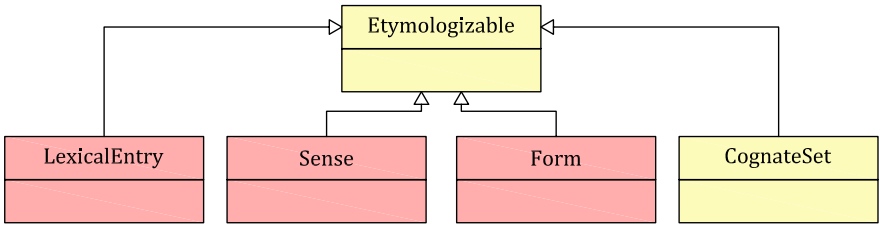
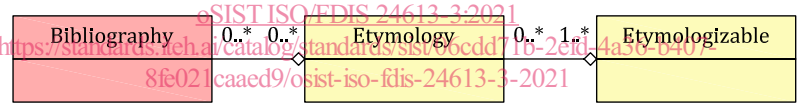


Figure 2 — The Etymologizable class and its subclasses

4.3 The Etymology class and the EtyLink class

The Etymology class allows for the description of the etymology of a linguistic element. More specifically, it allows for the description of those linguistic elements that are subclasses of the Etymologizable class. The type or types of etymological process involved in a given etymology can be specified using the type attribute, and also potentially the subtype attribute (in the case when the type of the etymology can be further specified). Possible values for type and subtype can vary according to the theoretical approach adopted by the compiler of a resource and/or the linguistic or editorial focus of the resource. The use of nested Etymology instances allows a combination of etymological processes to be described. Examples of etymological processes that shall be used as values for type/subtype: *borrowing, inheritance; word formation: compounding, derivation; sense shifts: narrowing, widening, amelioration, pejoration, metaphor, metonymy; phonetic/phonological processes: place assimilation, dissimilation, epenthesis, metathesis, hardening, weakening, etc*. The list of data categories provided in Annex B shall be used in complement to the appropriate classes. Individual links between two elements in an etymology can also be given a type, see the description of EtyLink below. Given that an Etymology instance can be taken from an external source, it can be associated with a Bibliography instance, which shall be defined as per ISO 24613-2 (see Figure 3).

iTeh STANDARD PREVIEW
(standards.iteh.ai)



oSIST ISO/FDIS 24613-3:2021
https://standards.iteh.ai/catalog/standards/sist/86cdd71b-2ef1-4a56-b407-8fe021caaed9/osist-iso-fdis-24613-3-2021

Figure 3 — The Etymology class

Instances of Etymology are associated with one or more EtyLink instances, each of which represents a single stage or step in the etymology of a given lexical item (see Figure 4). EtyLink serves to link together individuals belonging to the subclasses of Etymologizable. EtyLink is a subclass of the CrossREF class as defined in ISO 24613-1. The use of CrossREF requires that the target objects representing the given lexical content be given id attributes. The use of the id attribute on an individual of the Etymologizable class as a target allows for the modelling of a generic sequential temporal ordering of multiple elements, using the attributes prev and next. Instances of the EtyLink class can further specify additional temporal relationships using various temporal attributes associated with the source and target of each EtyLink instance.

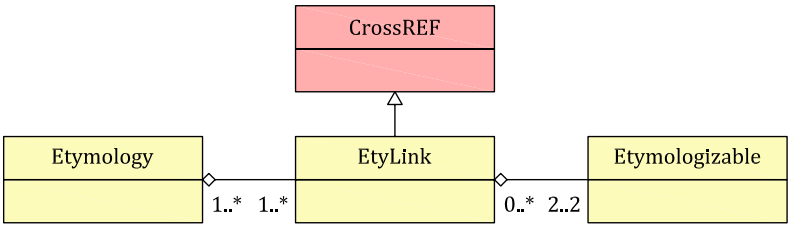


Figure 4 — EtyLink

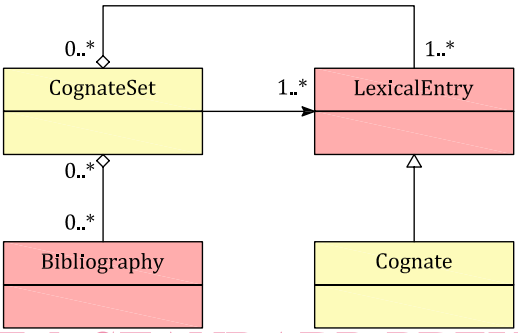
© ISO 2020 – All rights reserved

oSIST ISO/FDIS 24613-3:2021

Individuals of the Etymon and Cognate classes (subtypes of LexicalEntry) shall be associated with at least one individual of EtyLink. See Table 1 for a list of attributes to be used with the Etymology and EtyLink classes.

4.4 The CognateSet class

The CognateSet class (see Figure 5) is a container for sets of one or more Cognate items and zero or more Bibliography items (see ISO 24613-2). The CognateSet is a construct related to onomasiology. Its contents are items from languages related to that of a given LexicalEntry (and therefore by the subclass relation of any given Etymon or Cognate) and which have been gathered together with the purpose of demonstrating linguistic similarities or dissimilarities of salient kinds. The use of CognateSet implies that the LexicalEntry (and therefore Etymon and Cognate) instances which it contains share an etymological source.



iTeh STANDARD PREVIEW

Figure 5 — CognateSet

(standards.iteh.ai)

4.5 The Date class

oSIST ISO/FDIS 24613-3:2021
https://standards.iteh.ai/catalog/standards/sist/06cdd71b-2efd-4a36-b407-

The components of a LexicalEntry and its subclasses shall be associated with a specific date by making use of the Date class. Furthermore, Date allows the specification of a number of degrees of precision. A precise year, and potentially month and day, shall be stated using the date attribute and a rough date with the attribute circa. Within a span of time with different levels of specificity, there is the possibility of using one or more dating attributes. Where a span of time is known (or asserted), the lower and upper ends of the span can be specified using notBefore, notAfter respectively. For date and time formats, ISO 8601-1 and ISO 8601-2 shall be used.

4.6 The Gloss class

The Gloss class (see Figure 6) represents a textual description of the meaning of a word or a phrase that is intended for human consumption. Individuals of the class can either represent paraphrases or synonyms and these may be in the language of the entry or in another language. See Table 1 for a list of attributes to be used with this class.

© ISO 2020 – All rights reserved

oSIST ISO/FDIS 24613-3:2021

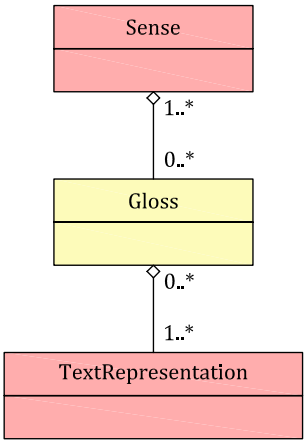


Figure 6 — Gloss

Table 1 — Example of class adornment

Class name	Example of attributes
Etymon	xml:lang, gloss, rootType, status
Etymology	type, subtype
EtyLink	type, prev, next
CognateSet	
Cognate	xml:lang, gloss, rootType, status
Date	notBefore, notAfter, circa, date
Gloss	xml:lang

iTeh STANDARD PREVIEW
(standards.iteh.ai)
oSIST ISO/FDIS 24613-3:2021
https://standards.iteh.ai/catalog/standards/sist/06cdd71b-2efd-4a36-b407-8fe021caaed9/osist-iso-fdis-24613-3-2021

© ISO 2020 – All rights reserved