INTERNATIONAL STANDARD

ISO 24613-4

First edition
2021-01

Language resource management — Lexical markup framework (LMF) —

Part 4: TEI serialization

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

Partie 4: Sérialisation TEI

Reference number
ISO 24613-4:2021(E)

COPYRIGHT PROTECTED DOCUMENT

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

Contents

Page

Foreword ..... **v**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 General principles **1**
5 Serialization of the LMF core model (ISO 24613-1) **2**
5.1 Implementing the LexicalResource class 2
5.2 Implementing the GlobalInformation class 2
5.3 Implementing the Lexicon class 3
5.4 Implementing the LexiconInformation class 4
5.5 Implementing the LexicalEntry class 4
5.6 Implementing the Form and Lemma classes 5
5.6.1 Form class 5
5.6.2 Lemma class 5
5.7 Implementing the GrammaticalInformation class 5
5.8 Implementing the Sense class 7
5.9 Implementing the Definition class 7
5.10 Implementing the OrthographicRepresentation class 8
5.11 Implementing the CrossREF class 8
5.12 Data category selection 9
6 Serialization of the MRD model (ISO 24613-2) **9**
6.1 Implementing the form representations for the Form class 9
6.2 Implementing classes derived from the Form class 11
6.2.1 General principles 11
6.2.2 Implementing the WordForm class 11
6.2.3 Implementing the Stem class 11
6.2.4 Implementing the WordPart class 11
6.2.5 Implementing the RelatedForm class 12
6.3 Implementing the Bibliography class 12
6.4 Implementing the CrossREF mechanism to refer to external media files 12
6.5 Implementing the Example class 12
6.6 Implementing the Translation class 13
6.7 Implementing the TextRepresentation class 13
6.8 Implementing the SubjectField class 14
6.9 Implementing the CrossREF mechanism to represent related entries 14
7 Implementing the classes from the etymological extension (ISO 24613-3) **15**
7.1 Implementing the Etymology class 15
7.2 Implementing the Etymon class 15
7.2.1 Referencing forms in an etymon 16
7.2.2 Representing the meaning of an etymon 16
7.2.3 Representing the language of an etymon 16
7.2.4 Associating grammatical information to an etymon 16
7.2.5 Dating an etymon 16
7.2.6 Providing sources associated with an etymon 17
7.2.7 Further prose component in an etymological description 17
7.3 Implementing the EtyLink class 17
7.4 Implementing the CognateSet class 17
7.5 Implementing the Cognate class 17
7.6 Implementing the Gloss class 18
8 Additional mechanisms **18**
8.1 Overview 18
8.2 Representing punctuation marks with <pc> 18

8.3 Representing various labels with <lbl> 18
8.4 Encoding simple separating characters with <metamark> 19
8.5 Providing rendition information with the @rend attribute 19
8.6 Usage example of the additional features introduced in Clause 8 19
Bibliography **20**

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 4, *Language resource management*.

This first edition of ISO 24613-4, together with ISO 24613-1:2019, ISO 24613-2:2020, ISO 24613-3:—[1)] and ISO 24613-5:—[2)], 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.

1) Under preparation. Stage at the time of publication: ISO/FDIS 24613-3:2020.

2) Under preparation. Stage at the time of publication: ISO/DIS 24613-5:2020.

Language resource management — Lexical markup framework (LMF) —

Part 4: TEI serialization

1 Scope

This document describes the serialization of the lexical markup framework (LMF) model defined as an XML model compliant with the Text Encoding Initiative (TEI) Guidelines. This serialization covers the classes of ISO 24613-1 (the LMF core model) as well as classes provided by ISO 24613-2 (the machine readable dictionary, MRD, model) and ISO 24613-3 (the etymological extension).

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 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*

ISO 24613-3, *Language resource management — Lexical markup framework (LMF) — Part 3: Etymological extension*

IETF BCP 47, *Tags for Identifying Languages*. (ed A. Phillips, M. Davis). September 2009. Best Current Practice. https://tools.ietf.org/html/bcp47

W3C XML Recommendation, *Extensible Markup Language (XML) 1.0* (Fifth Edition), 26 November 2008, http://www.w3.org/TR/xml/

TEI P5, *Guidelines for Electronic Text Encoding and Interchange*. [Version number: 4.1.0]. [Last modified date: 2020-08-19]. TEI Consortium. http://www.tei-c.org/Guidelines/P5/

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 24613-1 and in ISO 24613-3 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/

4 General principles

This document aims at providing mostly univocal constructs for each LMF class from the core model (ISO 24613-1), the MRD extension (ISO 24613-2) and the etymological extension (ISO 24613-3), thus making the original Text Encoding Initiative (TEI) recommendations from the dictionary chapter