INTERNATIONAL STANDARD

ISO 24613-5

First edition
2022-01

Language resource management — Lexical markup framework (LMF) —

Part 5: Lexical base exchange (LBX) serialization

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

Partie 5: Sérialisation de l'échange de bases lexicales (LBX)

This document is a preview generated by EVS

Reference number
ISO 24613-5:2022(E)

© ISO 2022

This document is a preview generated by EVS

COPYRIGHT PROTECTED DOCUMENT

© ISO 2022

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 2022 – All rights reserved

Contents

Page

Foreword ..... **v**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 General requirements **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 OrthographicRepresentation class 5
5.7 Implementing the Form class 6
5.7.1 Form class 6
5.7.2 Lemma class 6
5.8 Implementing the GrammaticalInformation class 6
5.9 Implementing the Sense class 7
5.10 Implementing the Definition class 8
5.11 Implementing the CrossREF class 8
6 Serialization of the MRD extension (ISO 24613-2) **10**
6.1 Implementing OrthographicRepresentation subclasses 10
6.2 Implementing the FormRepresentation class 10
6.3 Implementing the Form subclasses 11
6.3.1 General principles 11
6.3.2 Implementing the WordForm class 11
6.3.3 Implementing the Stem class 11
6.3.4 Implementing the WordPart class 11
6.3.5 Implementing the RelatedForm class 12
6.3.6 Implementing the TextRepresentation class 13
6.3.7 Implementing the Translation class 14
6.3.8 Implementing the Example class 14
6.4 Implementing the SubjectField class 15
6.5 Implementing the Bibliography class 15
7 Implementing the CrossREF mechanism to refer to external media files **15**
8 Implementing the classes from the etymological extension (ISO 24613-3) **15**
8.1 Implementing the Etymology class 15
8.2 Implementing the Etymon class 16
8.2.1 General 16
8.2.2 Referencing forms in an etymon 16
8.2.3 Representing the meaning of an etymon 16
8.2.4 Representing the language of an etymon 16
8.2.5 Dating an etymon 17
8.2.6 Providing sources associated with an etymon 17
8.3 Implementing the EtyLink class 17
8.4 Implementing the CognateSet class 17
8.5 Implementing the Cognate class 17
9 Additional mechanisms **18**
9.1 Overview 18
9.2 XML feature structure implementation 18
9.3 Representing various labels with <LBL> 18
9.4 Providing rendering information with the @rend attribute 18

This document is a preview generated by EVS

© ISO 2022 – All rights reserved

Annex A (informative) **LBX data category selection** **19**
Annex B (informative) **LBX feature structure implementation** **24**
Annex C (informative) **LBX examples for applying LBX serialization** **27**
Bibliography **32**

This document is a preview generated by EVS

© ISO 2022 – All rights reserved

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 24613-5, together with ISO 24613-1:2019, ISO 24613-2:2020, ISO 24613-3:2021 and ISO 24613-4:2021, cancels and replaces ISO 24613:2008, which has been technically revised.

The main change compared to the previous edition is as follows:

— entire revision of the content and its subdivisions 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.

This document is a preview generated by EVS

© ISO 2022 – All rights reserved

This document is a preview generated by EVS

Language resource management — Lexical markup framework (LMF) —

Part 5:
Lexical base exchange (LBX) serialization

This document is a preview generated by EVS

1 Scope

This document describes the serialization of the lexical markup framework (LMF) model defined as an extensible markup language (XML) model derived from the language base exchange (LBX) schema and compliant with the W3C XML schema. This serialization covers the classes, data categories, and mechanisms of ISO 24613-1 (core model), ISO 24613-2 (machine-readable dictionary (MRD) model), and ISO 24613-3 (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 15924, *Information and documentation — Codes for the representation of names of scripts*

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. PHILLIPS, A., DAVIS, M. (eds.), September 2009. Best Current Practice. Available from: https://tools.ietf.org/html/bcp47

W3C. Extensible Markup Language (XML) 1.1 (Second Edition). W3C Recommendation 16 August 2006, edited in place 29 September 2006. Available from: https://www.w3.org/TR/2006/REC-xml11-20060816/

3 Terms and definitions

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

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

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

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

4 General requirements

This document aims at providing constructs for each LMF class from ISO 24613-1 (core model), ISO 24613-2 (MRD extension), and ISO 24613-3 (etymological extension). It requires compliance

© ISO 2022 – All rights reserved