BS ISO 24613-1:2024

BSI Standards Publication

Language resource management — Lexical markup framework (LMF)

Part 1: Core model

Documents
Document Sharing Hub

Full document details: BS ISO 24613-1:2024

National foreword

This British Standard is the UK implementation of ISO 24613-1:2024. It supersedes BS ISO 24613-1:2019, which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee TS/1, Terminology.

A list of organizations represented on this committee can be obtained on request to its committee manager.

Contractual and legal considerations

This publication has been prepared in good faith, however no representation, warranty, assurance or undertaking (express or implied) is or will be made, and no responsibility or liability is or will be accepted by BSI in relation to the adequacy, accuracy, completeness or reasonableness of this publication. All and any such responsibility and liability is expressly disclaimed to the full extent permitted by the law.

This publication is provided as is, and is to be used at the recipient's own risk.

The recipient is advised to consider seeking professional guidance with respect to its use of this publication.

This publication is not intended to constitute a contract. Users are responsible for its correct application.

© The British Standards Institution 2024
Published by BSI Standards Limited 2024

ISBN 978 0 539 21347 8

ICS 01.020; 35.240.30

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 January 2024.

Amendments/corrigenda issued since publication

Date	Text affected

Documents
Document Sharing Hub

Full document details: BS ISO 24613-1:2024

International Standard

ISO 24613-1

Second edition
2024-01

Language resource management — Lexical markup framework (LMF) —

Part 1: Core model

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

Partie 1: Modèle de base

Reference number
ISO 24613-1:2024(en)

© ISO 2024

Documents
Document Sharing Hub

Full document details: BS ISO 24613-1:2024

COPYRIGHT PROTECTED DOCUMENT

© ISO 2024

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

Documents
Document Sharing Hub

Full document details: BS ISO 24613-1:2024

Contents

Page

Foreword ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Key standards used by LMF ... **3**
4.1 Unicode ... 3
4.2 Language coding ... 3
4.3 Script coding ... 3
4.4 Unified modelling language ... 3
5 The LMF model ... **3**
5.1 General ... 3
5.2 Class inheritance and data category selection procedures ... 4
5.2.1 Class inheritance ... 4
5.2.2 LMF attributes ... 4
5.2.3 Data category selection (DCS) ... 4
5.2.4 User-defined data categories ... 4
5.3 LMF core package ... 4
5.3.1 General ... 4
5.3.2 LexicalResource class ... 5
5.3.3 GlobalInformation class ... 5
5.3.4 Lexicon class ... 6
5.3.5 LexiconInformation class ... 6
5.3.6 LexicalEntry class ... 6
5.3.7 Form class ... 6
5.3.8 OrthographicRepresentation class ... 6
5.3.9 GrammaticalInformation class ... 6
5.3.10 Sense class ... 6
5.3.11 Definition class ... 7
5.4 Cross reference (CrossREF) model ... 7
5.4.1 General ... 7
5.4.2 CrossREF class ... 7
5.4.3 CrossREFConstraint class ... 7
5.5 Methods for data category selection and subclass creation ... 7
5.5.1 General ... 7
5.5.2 Generalization ... 7
5.5.3 Object instantiation ... 8
5.5.4 Design choices ... 8
5.5.5 Data categories for orthographic representation ... 8
5.5.6 Principles for model simplification ... 9
5.6 LMF extension use ... 9
5.6.1 General ... 9
5.6.2 Lexicon comparison ... 10
Annex A (informative) **Data category examples** ... **11**
Bibliography ... **14**

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24613-1:2024

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 document 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

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 second edition cancels and replaces the first edition (ISO 24613-1:2019), which has been technically revised.

The main changes are as follows:

— several changes have been made to Figure 1 “LMF core package”, as follows:

  — the OrthographicRepresentation class associations with the Form and Definition classes previously had a cardinality of 1 to 1, which did not correctly represent the intent of the UML model; the revision of the cardinality to 1 to 0..* in each case now provides a correct model;

  — the type: intern/extern attribute-value pair is no longer included in the CrossREF class since it described linking processes relevant for implementations, not associations relevant for a metamodel;

  — the full names relationship values in the CrossREF class, “synonym/composition” replace the abbreviations, “syn/compo”;

  — the class names in Figure 1 are now harmonized with the LMF style;

— relevant information has been moved from the tables in ISO 24613-2:2020 to Table A.1, meaning that the latter now contains more complete examples of values and attributes allocated to classes first introduced in this document.

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 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24613-1:2024

Introduction

Optimizing the production, maintenance and extension of electronic lexical resources is one of the crucial aspects impacting human language technologies (HLTs) in general and natural language processing (NLP) in particular, as well as human-oriented translation technologies. A second crucial aspect involves optimizing the process leading to their integration in applications. Lexical markup framework (LMF) is an abstract metamodel that provides a common, standardized framework for the construction of computational lexicons. LMF ensures the encoding of linguistic information in a way that enables reusability in different applications and for different tasks. LMF provides a common, shared representation of lexical instances, including morphological, syntactic and semantic aspects.

The goals of LMF are:

— to provide a common model for the creation and use of electronic lexical resources ranging from small to large in scale;

— to manage the exchange of data between and among these resources; and

— to facilitate the merging of large numbers of different individual electronic resources to form extensive global electronic resources.

The ultimate goal of LMF is to create a modular structure that will facilitate true content interoperability across all aspects of electronic lexical resources.

LMF supports existing lexical resource models such as Genelex,[5] the EAGLES International Standard for Language Engineering (ISLE),[6] Multilingual ISLE Lexical Entry (MILE) models,[12] Text Encoding Initiative (TEI) guidelines,[10] Ontolex[9] and the Language Base Exchange (LBX) serialization together with the US Government Wordscape On-Line Dictionary system[7].

LMF uses unified modelling language (UML) modelling processes.[11] The LMF core package describes the basic hierarchy of information of a lexical entry, including information on the word form. The core package is supplemented by various resources that are part of the definition of LMF. These resources include:

— specific data categories used by the variety of resource types associated with LMF (both those data categories relevant to the metamodel itself, and those associated with the extensions to the core package in additional LMF parts. See Annex A for data category examples);

— the constraints governing the relationship of these data categories to the metamodel and to its extensions;

— standard procedures for expressing these categories and thus for anchoring them on the structural skeleton of LMF and relating them to the respective extension models;

— the vocabularies used by LMF that describe how to extend LMF through linkage to a variety of specific resources (extensions) and methods for analysing and designing such linked systems.

LMF parts are expressed in a framework that describes the reuse of the LMF core components (such as structures, data categories and vocabularies) in conjunction with the additional components required for a specific resource.

The ISO 24613 series is designed to coordinate closely with ISO 16642.

© ISO 2024 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24613-1:2024

Documents
Document Sharing Hub