



Nederlandse norm

NEN-ISO 24615-2

(en)

Language resource management - Syntactic annotation framework (SynAF) - Part 2: XML serialization (Tiger vocabulary) (ISO 24615-2:2018, IDT)

ICS 01.020
maart 2018

Als Nederlandse norm is aanvaard:

- ISO 24615-2:2018, IDT

Normcommissie 400037 'Terminologie'



THIS PUBLICATION IS COPYRIGHT/PROTECTED

DEZE PUBLICATIE IS AUTEURSRECHTELIJK BESCHERMD

Apart from exceptions provided by the law, nothing from this publication may be duplicated and/or published by means of photocopy, microfilm, storage in computer files or otherwise, which also applies to full or partial processing, without the written consent of the Royal Netherlands Standardization Institute.

The Royal Netherlands Standardization Institute shall, with the exclusion of any other beneficiary, collect payments owed by third parties for duplication and/or act in and out of law, where this authority is not transferred or falls by right to the Reproduction Rights Foundation.

Auteursrecht voorbehouden. Behoudens uitzondering door de wet gesteld mag zonder schriftelijke toestemming van het Koninklijk Nederlands Normalisatie-instituut niets uit deze uitgave worden verveelvoudigd en/of openbaar gemaakt door middel van fotokopie, microfilm, opslag in computerbestanden of anderszins, hetgeen ook van toepassing is op gehele of gedeeltelijke bewerking.

Het Koninklijk Nederlands Normalisatie-instituut is met uitsluiting van ieder ander gerechtigd de door derden verschuldigde vergoedingen voor verveelvoudiging te innen en/of daartoe in en buiten rechte op te treden, voor zover deze bevoegdheid niet is overgedragen c.q. rechtens toekomt aan de Stichting Reprorecht.

Although the utmost care has been taken with this publication, errors and omissions cannot be entirely excluded. The Royal Netherlands Standardization Institute and/or the members of the committees therefore accept no liability, not even for direct or indirect damage, occurring due to or in relation with the application of publications issued by the Royal Netherlands Standardization Institute.

Hoewel bij deze uitgave de uiterste zorg is nagestreefd, kunnen fouten en onvolledigheden niet geheel worden uitgesloten. Het Koninklijk Nederlands Normalisatie-instituut en/of de leden van de commissies aanvaarden derhalve geen enkele aansprakelijkheid, ook niet voor directe of indirecte schade, ontstaan door of verband houdend met toepassing van door het Koninklijk Nederlands Normalisatie-instituut gepubliceerde uitgaven.



©2018 Koninklijk Nederlands Normalisatie-instituut
Postbus 5059, 2600 GB Delft
Telefoon (015) 2 690 390, Fax (015) 2 690 190

**Language resource management —
Syntactic annotation framework
(SynAF) —**

**Part 2:
XML serialization (Tiger vocabulary)**

*Gestion de ressources linguistiques — Cadre d'annotation syntaxique
(SynAF) —*

Partie 2: Sérialisation XML (vocabulaire Tiger)

Preview



Copyright
Preview



COPYRIGHT PROTECTED DOCUMENT

© ISO 2018

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
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

	Page
Foreword.....	iv
Introduction.....	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Graph structure and meta-model	2
5 Meta-model objects in XML serialization	3
6 Primary data and terminal representation	4
6.1 General.....	4
6.2 Sequential representation.....	4
6.3 Standoff representation.....	5
6.4 Typing of nodes and edges.....	6
7 Annotations	6
7.1 General.....	6
7.2 Domain declaration.....	7
7.3 Declaring annotations in an external file.....	7
7.4 Feature values.....	8
7.5 Data category references.....	8
7.6 Using the @type attribute.....	9
8 Corpus structure	10
Bibliography	12

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 on 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 the following URL: www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 4, *Language resource management*.

A list of all parts in the ISO 24615 series can be found on the ISO website.

Introduction

The need for standardization of syntactic annotation was recognized and addressed in detail with the publication of ISO 24615-1:2014. As a result of the work on ISO 24615-1:2014, it was anticipated that such a reference model for syntactic annotation should be associated with a concrete XML serialization in order to meet the specific needs of such applications as syntactic parsers or syntactic treebanks, where representations have to be exchanged and reused. Furthermore, such a serialization should be independent from the theoretical orientation and specific details of any specific annotation scheme.

This document answers this need on the basis of the seminal work carried out on the TigerXML format^[3]. This starting point was chosen as a reference because it is widely used as a de facto standard for unrelated XML treebanks, with the advantages in terms of interoperability offered by its XML-based representations, as opposed to other frequently used formats, in particular, the Penn Treebank bracketing format^[5] or the CoNLL format for dependency structures (see Reference^[4]).

The document is designed to complement ISO 24615-1:2014 and to coordinate closely with ISO 24610, ISO 24611, ISO 24612 and ISO 12620.

This document therefore extends ISO 24615-1:2014 with an XML model based upon the Tiger XML vocabulary for the interchange of syntactically annotated data which is both standardized as well as language- and theory-independent. The proposed format directly instantiates all features of the meta-model defined in ISO 24615-1 and defines concrete serialized interfaces to the complementary ISO 24611 and ISO 12620, which provides the background for the DatCatInfo data category registry.

Copyright
Preview

Probleem
Preview

Language resource management — Syntactic annotation framework (SynAF) —

Part 2: XML serialization (Tiger vocabulary)

1 Scope

This document describes an XML-conformant serialization of the ISO 24615-1 meta-model, with the objective of supporting interoperability across language resources or language processing components in the domain of syntactic annotations. As an extension of ISO 24615-1, this document is also coordinated with ISO 24612.

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 12620, *Terminology and other language and content resources — Data category specifications*

ISO 24610 (all parts), *Language resource management — Feature structures*

ISO 24611, *Language resource management — Morpho-syntactic annotation framework (MAF)*

ISO 24612, *Language resource management — Linguistic annotation framework (LAF)*

ISO 24615-1, *Language resource management — Syntactic annotation framework (SynAF) — Part 1: Syntactic model*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 12620, ISO 24610 (all parts), ISO 24611, ISO 24612 and ISO 24615-1 and the following apply.

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

- IEC Electropedia: available at <http://www.electropedia.org/>
- ISO Online browsing platform: available at <http://www.iso.org/obp>

3.1 domain

class of elements to which a certain set of *labels* (3.2) can be assigned

Note 1 to entry: Domains can refer generally to the set of all edges, terminal nodes or non-terminal nodes.

3.2 label

unit of annotation consisting of the name of a feature and a value, which together can be applied to appropriate model elements and add arbitrary feature-value annotations to such elements

Bestelformulier

Stuur naar:

NEN Standards Products & Services
t.a.v. afdeling Klantenservice
Antwoordnummer 10214
2600 WB Delft



NEN Standards Products & Services

Postbus 5059
2600 GB Delft

Vlinderweg 6
2623 AX Delft

T (015) 2 690 390
F (015) 2 690 271

www.nen.nl/normshop

Ja, ik bestel

__ ex. NEN-ISO 24615-2:2018 en

€ 52.53

Wilt u deze norm in PDF-formaat? Deze bestelt u eenvoudig via www.nen.nl/normshop

Gratis e-mailnieuwsbrieven

Wilt u op de hoogte blijven van de laatste ontwikkelingen op het gebied van normen, normalisatie en regelgeving? Neem dan een gratis abonnement op een van onze e-mailnieuwsbrieven. www.nen.nl/nieuwsbrieven

Gegevens

Bedrijf / Instelling

T.a.v. O M O V

E-mail

Klantnummer NEN

Uw ordernummer BTW nummer

Postbus / Adres

Postcode Plaats

Telefoon Fax

Factuuradres (indien dit afwijkt van bovenstaand adres)

Postbus / Adres

Postcode Plaats

Datum Handtekening

Retourneren

Fax: 015 2 690 271

E-mail: klantenservice@nen.nl

Post: NEN Standards Products & Services,

t.a.v. afdeling Klantenservice
Antwoordnummer 10214,
2600 WB Delft

(geen postzegel nodig).

Voorwaarden

- De prijzen zijn geldig tot 31 december 2018, tenzij anders aangegeven.
- Alle prijzen zijn excl. btw, verzend- en handelingskosten en onder voorbehoud bij o.m. ISO- en IEC-normen.
- Bestelt u via de normshop een pdf, dan betaalt u geen handeling en verzendkosten.
- Meer informatie: telefoon 015 2 690 391, dagelijks van 8.30 tot 17.00 uur.
- Wijzigingen en typfouten in teksten en prijsinformatie voorbehouden.
- U kunt onze algemene voorwaarden terugvinden op: www.nen.nl/leveringsvoorwaarden.