

NORME INTERNATIONALE

ISO 24615-1

Première édition
2014-02

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

Partie 1: Modèle syntaxique

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

Part 1: Syntactic model

Sample Document

get full document from standards.iteh.ai



Numéro de référence
ISO 24615-1:2014(F)

© ISO 2014

Sample Document

get full document from standards.iteh.ai



DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO 2014

Tous droits réservés. Sauf prescription différente ou nécessité dans le contexte de sa mise en œuvre, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie, ou la diffusion sur l'internet ou sur un intranet, sans autorisation écrite préalable. Une autorisation peut être demandée à l'ISO à l'adresse ci-après ou au comité membre de l'ISO dans le pays du demandeur.

ISO copyright office
Case postale 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Genève
Tél.: +41 22 749 01 11
Fax: +41 22 749 09 47
E-mail: copyright@iso.org
Web: www.iso.org

Publié en Suisse

Sommaire

Page

Avant-propos	iv
Introduction	v
1 Domaine d'application	1
2 Références normatives	1
3 Termes et définitions	1
4 Métamodèle SynAF	4
4.1 Introduction	4
4.2 Métamodèle SynAF.....	5
4.2.1 Vue d'ensemble	5
4.2.2 Classe SyntacticNode.....	5
4.2.3 Classe T_Node.....	5
4.2.4 Classe NT_Node.....	5
4.2.5 Classe SyntacticEdge	5
4.2.6 Classe Annotation	5
Annexe A (normative) Catégories de données pour SynAF	7
Annexe B (informative) Relation avec le cadre d'annotation linguistique	17
Bibliographie	19

Sample Document

get full document from standards.iteh.ai

Avant-propos

L'ISO (Organisation internationale de normalisation) est une fédération mondiale d'organismes nationaux de normalisation (comités membres de l'ISO). L'élaboration des Normes internationales est en général confiée aux comités techniques de l'ISO. Chaque comité membre intéressé par une étude a le droit de faire partie du comité technique créé à cet effet. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO participent également aux travaux. L'ISO collabore étroitement avec la Commission électrotechnique internationale (IEC) en ce qui concerne la normalisation électrotechnique.

Les procédures utilisées pour élaborer le présent document et celles destinées à sa mise à jour sont décrites dans les Directives ISO/IEC, Partie 1. Il convient, en particulier de prendre note des différents critères d'approbation requis pour les différents types de documents ISO. Le présent document a été rédigé conformément aux règles de rédaction données dans les Directives ISO/IEC, Partie 2 (voir www.iso.org/directives).

L'attention est appelée sur le fait que certains des éléments du présent document peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO ne saurait être tenue pour responsable de ne pas avoir identifié de tels droits de propriété et averti de leur existence. Les détails concernant les références aux droits de propriété intellectuelle ou autres droits analogues identifiés lors de l'élaboration du document sont indiqués dans l'Introduction et/ou dans la liste des déclarations de brevets reçues par l'ISO (voir www.iso.org/brevets).

Les appellations commerciales éventuellement mentionnées dans le présent document sont données pour information, par souci de commodité, à l'intention des utilisateurs et ne sauraient constituer un engagement.

Pour une explication de la signification des termes et expressions spécifiques de l'ISO liés à l'évaluation de la conformité, ou pour toute information au sujet de l'adhésion de l'ISO aux principes de l'Organisation mondiale du commerce (OMC) concernant les obstacles techniques au commerce (OTC), voir le lien suivant: www.iso.org/iso/fr/avant-propos.html.

Le comité chargé de l'élaboration du présent document est l'ISO/TC 37, *Langage et terminologie*, sous-comité SC 4, *Gestion des ressources linguistiques*.

Cette première édition de l'ISO 24615-1 annule et remplace l'ISO 24615:2010 dont elle constitue une révision mineure.

L'ISO 24615, dans toutes ses parties, est conçue de manière coordonnée avec l'ISO 24612, *Gestion des ressources linguistiques – Cadre d'annotation linguistique* (LAF), l'ISO 24613:2008, *Gestion des ressources linguistiques – Cadre d'annotation lexicale* (LMF), et l'ISO 24611, *Gestion des ressources langagières – Cadre d'annotation morpho-syntaxique*.

L'ISO 24615 comporte les parties suivantes, sous le titre général *Gestion des ressources linguistiques – Cadre d'annotation syntaxique* (SynAF):

— *Partie 1: Modèle syntaxique*

La partie suivante est en cours d'élaboration:

— *Part 2: Sérialisation XML (<Tiger2/>)*

Introduction

L'ISO 24615 se fonde sur de nombreux projets et activités de pré-normalisation qui ont été menés au cours des dernières années (voir Abeillé, 2001^[9]) en vue de constituer des modèles de référence et des formats pour la représentation de l'information syntaxique, soit sous forme de sortie d'analyseur syntaxique, soit sous forme d'annotations de ressources linguistiques (treebanks). Depuis plusieurs années l'initiative Penn Treebank constitue le standard *de facto* pour l'annotation syntaxique, mais des travaux plus récents, par exemple l'initiative Negra/Tiger (voir <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/>) en Allemagne ou l'initiative ISST en Italie (voir Montemagni (2003)^[18]) démontre la viabilité d'un cadre plus cohérent qui tient compte à la fois des constituants hiérarchiques et des phénomènes de dépendance syntaxique dans l'annotation syntaxique.

Le projet eContent "LIRICS" est un projet majeur car il a rassemblé un groupe d'experts, initiateur du projet ISO 24615 (SynAF). Lors de la préparation de SynAF, ce groupe a confirmé que les initiatives actuelles partagent effectivement un modèle de données commun qui offre une bonne base de départ pour le métamodèle SynAF (voir l'étude "Evaluation of initiatives for morpho-syntactic and syntactic annotation" du projet européen LIRICS, disponible à l'adresse http://lirics.loria.fr/doc/pub/Del3_1_V2.pdf).

Cette partie de l'ISO 24615 propose un métamodèle pour l'annotation syntaxique ainsi qu'une liste de catégories de données pertinente pour l'annotation syntaxique. Les catégories de données sont disponibles sur le serveur ISOCat (<http://www.isocat.org/>) au sein du profil syntaxe (tel que défini dans l'ISO 12620:2009).

Sample Document

get full document from standards.iteh.ai

Sample Document

get full document from standards.iteh.ai

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

Partie 1: Modèle syntaxique

1 Domaine d'application

Cette partie de l'ISO 24615 décrit le cadre d'annotation syntaxique (SynAF), modèle de haut niveau pour représenter l'annotation syntaxique des données linguistiques, dans le but d'offrir l'interopérabilité entre les ressources linguistiques ou les composants du traitement de la langue. La présente partie de l'ISO 24615 est complémentaire par rapport à l'ISO 24611 (MAF, cadre d'annotation morpho-syntaxique) et y est étroitement liée. Elle fournit un métamodèle pour les représentations syntaxiques, avec les catégories de données de référence destinées à représenter tout aussi bien des informations de constituance que des informations de dépendance, dans des phrases ou autres énonciations et segments du même ordre.

2 Références normatives

Les documents suivants sont indispensables à l'application de ce document. Pour les références datées, seule l'édition citée s'applique. Pour les références non datées, la dernière édition du document référencé s'applique (incluant ses éventuels amendements).

ISO 1087-1:2000, *Travaux terminologiques — Vocabulaire — Partie 1: Théorie et application*

ISO 12620:2009, *Terminologie et autres ressources langagières et ressources de contenu — Spécification de catégories de données et gestion d'un registre de catégories de données pour les ressources langagières*

ISO 24611:2012, *Gestion des ressources langagières — Cadre d'annotation morphosyntaxique (MAF)*

3 Termes et définitions

Pour les besoins du présent document, les termes et définitions figurant dans l'ISO 1087-1:2000, l'ISO 12620:2009, l'ISO 24611:2012 ainsi que ce les suivants s'appliquent.

3.1 circonstant

élément non essentiel associé à un verbe en opposition avec les *arguments syntaxiques* (3.19)

Note 1 à l'article: Les adverbes sont des circonstants possibles pour une phrase.

3.2 bloc bloc syntaxique

constituant non récursif (3.4)

Note 1 à l'article: Le terme «bloc» est la traduction en français du terme anglais «chunk»