SLOVENSKI STANDARD
SIST ISO 24615:2013

01-julij-2013

Upravljanje z jezikovnimi viri - Ogrodje za skladenjsko označevanje (SynAF)

Language resource management -- Syntactic annotation framework (SynAF)

Gestion de ressources langagières -- Cadre d'annotation syntaxique (SynAF)

Ta slovenski standard je istoveten z: ISO 24615:2010

ICS:

01.020	Terminologija (načela in koordinacija)	Terminology (principles and coordination)

SIST ISO 24615:2013   **en,fr,de**

iTeh STANDARD PREVIEW
(standards.iteh.ai)

INTERNATIONAL STANDARD

**ISO
24615**

First edition
2010-10-15

Language resource management — Syntactic annotation framework (SynAF)

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

ISO 24615:2010(E)

iTeh STANDARD PREVIEW

(standards.iteh.ai)

COPYRIGHT PROTECTED DOCUMENT

ISO 24615:2010(E)

Contents

Page

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 24615:2010(E)

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 24615 was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 4, *Language resource management*, in collaboration with the European eContent Project "LIRICS" (**L**inguistic **I**nfrastructure for Inter**o**perable **R**esources and **S**ystems), under the contract e-Content-22236-LIRICS.

ISO 24615 is designed to coordinate closely with ISO 24612, *Language resource management — Linguistic annotation framework (LAF)*, ISO 24613:2008, *Language resource management — Lexical markup framework (LMF)*, and ISO 24611, *Language resource management — Morpho-syntactic annotation framework*.

Introduction

This International Standard is based on numerous projects and pre-standardisation activities that have taken place in the last few years (see Abeillé, 2001[9]), to provide reference models and formats for the representation of syntactic information, whether as the output of a syntactic parser, or as annotations of language resources (treebanks). For several years, the Penn Treebank initiative has served as a *de facto* standard for treebanking, but more recent works e.g. the Negra/Tiger initiative (see: http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/) in Germany or the ISST initiative in Italy [see Montemagni (2003)[18]] demonstrate the viability of a more coherent framework that can account for both (hierarchical) constituency and dependency phenomena in syntactic annotation.

The eContent project "LIRICS", has been seminal in gathering a group of experts, who initiated the ISO 24615 (SynAF) project. While preparing SynAF, this group confirmed that existing initiatives indeed share a common data model that offers a good basis for the SynAF metamodel (see the study made in Deliverable D.3.1 "Evaluation of initiatives for morpho-syntactic and syntactic annotation" of the EU project LIRICS, available at http://lirics.loria.fr/doc_pub/Del3_1_V2.pdf).

This International Standard proposes a metamodel for syntactic annotation together with a list of relevant data categories for syntactic annotation. The data categories are available on the ISOCat server (http://www.isocat.org/) in the syntax profile (as defined in ISO 12620:2009).

iTeh STANDARD PREVIEW

(standards.iteh.ai)

INTERNATIONAL STANDARD   **ISO 24615:2010(E)**

Language resource management — Syntactic annotation framework (SynAF)

1 Scope

This International Standard describes the syntactic annotation framework (SynAF), a high level model for representing the syntactic annotation of linguistic data, with the objective of supporting interoperability across language resources or language processing components. This International Standard is complementary and closely related to ISO 24611 (MAF, morpho-syntactic annotation framework) and provides a metamodel for syntactic representations as well as reference data categories for representing both constituency and dependency information in sentences or other comparable utterances and segments.

2 Normative references

The following referenced documents are indispensable for the application 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 1087-1:2000, *Terminology work — Vocabulary — Part 1: Theory and application*

ISO 1087-2:2000, *Terminology work — Vocabulary — Part 2: Computer applications*

ISO 12620:2009, *Terminology and other language and content resources — Specification of data categories and management of a Data Category Registry for language resources*

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

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 1087-1, ISO 1087-2, ISO 12620:2009, ISO 24611 and the following apply.

3.1
adjunct
non-essential element associated with a verb as opposed to **syntactic arguments** (3.19)

NOTE   Adverbs are possible adjuncts for a sentence.

3.2
chunk
non-recursive **constituent** (3.4)

3.3
clause
group of **phrases** (3.14), usually containing a predicate

NOTE   A clause can be either a **main clause** (3.10) or a **subordinate clause** (3.17). In languages distinguishing finiteness, clauses whose predicate is a verb can be either finite or non-finite, depending on the form of the verb. A main clause alone can build a complete **sentence** (3.15). In the SynAF model, a clause is a special case of a **constituent** (3.4).

ISO 24615:2010(E)

3.4
constituent
syntactic grouping of words [into **phrases** (3.14)], phrases [into **clauses** (3.3) or **other phrases**] or clauses [into a **sentence** (3.15)] on the base of structural (or hierarchical) properties

3.5
dependency
dependency relation
syntactic relation between **word forms** (3.24) or **constituents** (3.4) on the basis of the **grammatical functions** (3.7) that **constituents** play in relation to each other

3.6
syntactic edge
edge
triplet with a source **node** (3.12), a target **node**, and optional **annotations** (3.9)

NOTE **Non-terminal nodes** (3.13) have an outgoing constituency syntactic edge.

3.7
grammatical function
grammatical role of a **word form** (3.24) or **constituent** (3.4) within its embedding syntactic environment

NOTE For example, a noun phrase (NP) can act as a subject within a **sentence** (3.15), or a noun may act as a subject dependent of a verb in a dependency graph. There is a grammatical relation between the subject – NP and the main verb in a sentence. All grammatical relations (subject – predicate, head – modifier, etc.) are subsumed under the concept of **dependency relations** (3.5), whether between terminal or non-terminal nodes.

3.8
syntactic head
head
part of a **constituent** (3.4) which determines its distribution (the syntactic environments in which the constituent may appear) and its grammatical properties (e.g. if the grammatical gender of the head is feminine, then the gender of the entire constituent will be feminine)

NOTE The head of a constituent usually cannot be left out.

3.9
linguistic annotation
annotation
feature-value pair denoting a linguistic property of a linguistic segment

3.10
main clause
clause (3.3), which can act on its own as a complete **sentence** (3.15)

NOTE In languages distinguishing finiteness, the main clause is usually finite. Example: *The train is late*.

3.11
modifier
part of a **constituent** (3.4) which ascribes a property to the **head** (3.8) of the **constituent**

NOTE A modifier can be placed before or after the head of the **phrase** (3.14) (pre-modifier or post-modifier). Modifiers are optional in a constituent.

3.12
node
syntactic node
word form (3.24) or **constituent** (3.4) seen as an elementary syntactic component of a syntactic analysis

3.13
non-terminal node
syntactic node (3.12) which is not a **word form** (3.24)

NOTE   A non-terminal node has an outgoing constituency **edge** (3.6).

3.14
phrase
group of **word forms** (3.24) (usually containing one or more words) which can fulfill a **grammatical function** (3.7), e.g. in a **clause** (3.3)

NOTE   Empty phrases are permitted (being non-realised pronouns, sometimes marked as "pro", and having the role of subjects in clauses). A phrase is typically named after its **head** (3.8), for example noun phrases, verb phrases, adjective phrases, adverbial phrases and prepositional phrases. Phrases have been informally described as "bloated words", in that the parts of the phrase added to the head elaborate and specify the reference of the head. In our model, a phrase is a special case of a **constituent** (3.4)**.**

3.15
sentence
related group of **word forms** (3.24) containing a predication, usually expressing a complete thought and forming the basic unit of discourse structure

NOTE   A sentence consists of one or more **clauses** (3.3). When describing speech, it is common to talk about "utterances" rather than sentences.

3.16
span
pair of points (p1, p2), where p1 $\leqslant$ p2, identifying the segment of the document to which an **annotation** (3.9) is applied

NOTE   A multiple span is a sequence of spans where the ending point of each span is less than or equal to the starting point of the subsequent span.

3.17
subordinate clause
clause which fulfils a **grammatical function** (3.7) in a **phrase** (3.14) [for example a relative **clause** (3.3) modifying the **head** (3.8) noun of a nominal phrase] or in another clause

NOTE   A subordinate clause usually does not act on its own as a sentence, but is part of a larger sentence**.**

3.18
subcategorization frame
set of restrictions indicating the properties of the **syntactic arguments** (3.19) that can or must occur with a verb

EXAMPLE   Alfred (/syntacticArgument/) reads a book (/syntacticArgument/) today (/adjunct/).

NOTE   The subject, indirect object and direct object are subcategorized **grammatical functions** (3.7) within a sentence; they are dependents of the verb (i.e. they can appear in subcategorization frames).

3.19
syntactic argument
functionally essential element that is required and given its interpretation by the head of its **phrase** (3.14) or the **node** (3.12) of which it is a dependent (e.g. the nominal argument of a prepositional phrase or verb)

NOTE   For verbs and verbal phrases, arguments identify the participants in the process referred to by the verb. In some frameworks, syntactic arguments are called complements.

3