SLOVENSKI STANDARD
SIST-TS ISO/TS 24617-5:2018

01-september-2018

Upravljanje z jezikovnimi viri - Ogrodje za semantično označevanje (SemAF) - 5. del: Struktura diskurza (SemAF-DS)

Language resource management -- Semantic annotation framework (SemAF) -- Part 5: Discourse structure (SemAF-DS)

Gestion de ressources langagières -- Cadre d'annotation sémantique (SemAF) -- Partie 5: Structures de discours (SemAF-DS)

Ta slovenski standard je istoveten z: **ISO/TS 24617-5:2014**

ICS:

01.020	Terminologija (načela in koordinacija)	Terminology (principles and coordination)
35.060	Jeziki, ki se uporabljajo v informacijski tehniki in tehnologiji	Languages used in information technology

SIST-TS ISO/TS 24617-5:2018 **en,fr,de**

TECHNICAL SPECIFICATION

ISO/TS 24617-5

First edition
2014-03-01

Language resource management — Semantic annotation framework (SemAF) —

Part 5: Discourse structure (SemAF-DS)

Gestion de ressources langagières — Cadre d'annotation sémantique (SemAF) —

Partie 5: Structures de discours (SemAF-DS)

Reference number
ISO/TS 24617-5:2014(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

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 meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 04, *Language resource management*.

ISO 24617 consists of the following parts, under the general title *Language resource management — Semantic annotation framework*:

— *Part 1: Time and events (SemAF-Time, ISO-TimeML)*

— *Part 2: Dialogue acts (SemAF-DA)*

— *Part 4: Semantic roles (SemAF-SR)*

— *Part 5: Discourse structures (SemAF-DS)*

— *Part 6: Principles of semantic annotation (SemAF-Basics)*

— *Part 7: Spatial information (ISO-Space)*

— *Part 8: Semantic relations in discourse (SemAF-DRel)*

Introduction

Discourse structures play an essential role in the production and analysis of the syntactic, semantic, and pragmatic features of text, speech, and other types of discourse. This Technical Specification is a basis both for the annotation, generation and translation (among other processes) of these types of discourses and of the syntactic, semantic, and pragmatic features derived from them. Note that discourse structures underlie not only verbal communication (whether spoken, written, or signed) but also non-verbal discourse (such as a silent video).

The annotation scheme provided here specifies discourse structures that consist of segment structures and content structures. It also specifies the mappings between these two structures; the mappings are described by the annotations of discourse segments in texts or some other modalities. In this context, on the one hand, segment structures are spatiotemporal relations that hold between surface segments (such as words, phrases, clauses, sentences, and video scenes) and, on the other hand, content structures are discourse relations that are established between semantic and pragmatic items. Both of these structures can be represented by means of labelled directed graphs or sometimes simply by trees, as standardized by LAF (ISO 24612:2012) and SynAF (ISO 24615:2010).

This scheme also provides a common, language-neutral pivot for the interoperation among diverse formats of discourse structures of various types of document, and can be applied to the generation of linguistic and non-linguistic expressions. For example, if the discourse structures of speech and other linguistic data contained in motion pictures are fitted to this scheme, multilingual subtitles for these pictures can be generated at a reduced cost by means of a standardized tool for multilingual translation. By the same token, this scheme can facilitate interoperability among various discourse corpora and collaboration among researchers who use them.

Language resource management — Semantic annotation framework (SemAF) —

Part 5: Discourse structure (SemAF-DS)

1 Scope

A discourse is a process of communication. This Technical Specification addresses how a discourse is structured in terms of its realization/presentation and content, and shows how its dual structure can be represented in a graph. The current specification focuses on the annotation of discourse structures in text only, but it can be extended to discourses in other modalities.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 15938-5:2003/Amd.1:2004, *Information technology. Multimedia content description interface. Part 5: Multimedia description schemes AMENDMENT 1: Multimedia description schemes extensions (MPEG-7 MDS AMD1)*

ISO 24612:2012, *Language resource management — Linguistic annotation framework (LAF)*

ISO 24615:2010, *Language resource management — Syntactic annotation framework (SynAF)*

ISO 24617-1:2012, *Language resource management — Semantic annotation framework — Part 1: Time and events (SemAF-Time, ISO-TimeML)*

ISO 24617-2:2012, Language resource management — *Semantic annotation framework — Part 2: Dialogue acts*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1
circumstance
entity which is an event (including dialogue act), state, process, relation, proposition, or set of these

3.2
class
unary predicate, which is a set of entities

3.3
discourse
process of communication, consisting of one or more sentences or sentence fragments

Note 1 to entry: From an abstract viewpoint, data (e.g. words, phrases, sentences, and paragraphs) representing a communication process is regarded as a discourse. A discourse can be encoded in various media such as text, hypertext, audio, video, and their possible combinations.

3.4
discourse relation
semantic/pragmatic relation that holds among two or more circumstances

Note 1 to entry: Some discourse relations, such as *example* and *part*, can also hold between objects. In this document, semantic/pragmatic relations (including discourse relations) are given in italics in the text and with a gray background in the Figures (e.g. *agent*, *inference*, and *purpose*).

3.5
discourse structure
structure of discourse, comprising segment structure, content structure, and possibly other types of structure

3.6
entity
semantic/pragmatic entity referenced in discourse, including circumstances, and objects

Note 1 to entry: An entity is represented by a node in a content structure.

3.7
object
semantic entity other than circumstance

Note 1 to entry: Objects include people, buildings, machines, ideas, and rules.

3.8
relational class
class whose instances are circumstances equivalent to relations

3.9
segment
word, phrase, clause, sentence, paragraph, section, chapter, or other partial realization of discourse

Note 1 to entry: A synonym is a 'discourse segment'. A segment references a semantic and/or pragmatic entity, which can be a semantic/pragmatic relation. Intrasentential segments are syntactic constituents such as words, phrases, and clauses. Segments might or might not be continuous: this is discussed in the definition of connectives.

4 Overview

A discourse structure consists of two types of structure: segment structure and content structure. A segment structure (extending intrasentential syntax) is a data structure that describes how a discourse has been organized from a formal syntactic perspective. It consists of

a) a set of segments (some partial realizations of discourse), and

b) the syntactic relations holding among them.

A content structure (extending intrasentential semantics) is a data structure that describes from a logical point of view how a discourse has been organized. It consists of

a) the set of semantic and pragmatic components referred to by the segments of a segment structure (that is, by some segments of some discourse), and

b) the logical relations established between these semantic representations. These two structures organize the whole structure of each discourse.

Both types of structure and content structures in particular, can be represented by means of a labelled directed graph. Various syntactic relations in a segment structure can, for instance, be captured by a tree (single-rooted graph). Discourse relations in a content structure can also be captured by a more general graph: The nodes in the graph stand for semantic and pragmatic components and the edges formalize the relations holding among them. In one way, a segment structure is to a discourse (or part of

