PD ISO/TS 24617-5:2014

BSI Standards Publication

Language resource management — Semantic annotation framework (SemAF)

Part 5: Discourse structure (SemAF-DS)

bsi.

...making excellence a habit.™

Document Sharing Hub

Full document details: PD ISO/TS 24617-5:2014

National foreword

This Published Document is the UK implementation of ISO/TS 24617-5:2014.

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 secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2014. Published by BSI Standards Limited 2014

ISBN 978 0 580 84556 7

ICS 01.020

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

This Published Document was published under the authority of the Standards Policy and Strategy Committee on 31 March 2014.

Amendments issued since publication

Date	Text affected

Documents
Document Sharing Hub

Full document details: PD ISO/TS 24617-5:2014

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)

Documents
Document Sharing Hub
© ISO 2014

Full document details: PD ISO/TS 24617-5:2014

COPYRIGHT PROTECTED DOCUMENT

© ISO 2014

All rights reserved. Unless otherwise specified, 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
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

© ISO 2014 – All rights reserved

Documents
Document Sharing Hub

Full document details: PD ISO/TS 24617-5:2014

Contents

Page

Foreword ..... **iv**
Introduction ..... **v**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Overview **2**
5 Segment structure **3**
6 Content structure **4**
7 Mapping between segment and content structures **7**
8 Concluding remarks **16**
Bibliography **17**

© ISO 2014 – All rights reserved

Documents
Document Sharing Hub
Full document details: PD ISO/TS 24617-5:2014

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)*

© ISO 2014 – All rights reserved
Documents
Document Sharing Hub
Full document details: PD ISO/TS 24617-5:2014

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.

© ISO 2014 – All rights reserved

Documents
Document Sharing Hub
Full document details: PD ISO/TS 24617-5:2014

Documents
Document Sharing Hub

Full document details: PD ISO/TS 24617-5:2014

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.

© ISO 2014 – All rights reserved

Documents
Document Sharing Hub

Full document details: PD ISO/TS 24617-5:2014