INTERNATIONAL STANDARD

ISO 24610-2

First edition
2011-10-01

Language resource management — Feature structures —

Part 2: Feature system declaration

Gestion des ressources langagières — Structures de traits —

Partie 2: Déclaration de système de structures de traits

Reference number
ISO 24610-2:2011(E)

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.

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 24610-2 was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 4, *Language resource management*.

ISO 24610 consists of the following parts, under the general title *Language resource management — Feature structures*:

— *Part 1: Feature structure representation*

— *Part 2: Feature system declaration*

Introduction

ISO 24610 is organized in two separate main parts.

— Part 1, *Feature structure representation*, is dedicated to the description of feature structures, providing an informal and yet explicit outline of their characteristics, as well as an XML-based structured way of representing feature structures in general and typed feature structures in particular. It is designed to lay a basis for constructing an XML-based reference format for exchanging (typed) feature structures between applications.

— Part 2, *Feature system declaration*, will provide an implementation standard for XML-based typed feature structures, first by defining a set of types and their hierarchy, then by formulating type constraints on a set of features and their respective admissible feature values and finally by introducing a set of validity conditions on feature structures for particular applications, especially related to the goal of language resource management.

A feature structure is a general-purpose data structure that identifies and groups together individual features by assigning a particular value to each. Because of the generality of feature structures, they can be used to represent many different kinds of information. Interrelations among various pieces of information and their instantiation in markup provide a meta-language for representing linguistic content. Moreover, this instantiation allows a specification of a set of features and values associated with specific types and their restrictions, by means of feature system declarations, or other XML mechanisms to be discussed in this part of ISO 24610.

Some of the statements here are copied from ISO 24610-1:2006 in order to make this part standalone without referring to part 1.

Language resource management — Feature structures —

Part 2:
Feature system declaration

1 Scope

This part of ISO 24610 provides a format to represent, store or exchange feature structures in natural language applications, for both annotation and production of linguistic data. It is ultimately designed to provide a computer format to define a type hierarchy and to declare the constraints that bear on a set of feature specifications and operations on feature structures, thus offering means to check the conformance of each feature structure with regards to a reference specification. Feature structures are an essential part of many linguistic formalisms as well as an underlying mechanism for representing the information consumed or produced by and for language engineering applications.

A feature system declaration (FSD) is an auxiliary file used in conjunction with a certain type of text that makes use of fs (that is, feature structure) elements. The FSD serves four purposes.

— It provides an encoding by which types and their subtyping and inheritance relationships can be introduced and defined, thus laying the basis for constructing a feature system.

— It provides a mechanism by which the encoder can list all of the feature names and feature values and give a prose description as to what each represents.

— It provides a mechanism by which type constraints can be declared, against which typed feature structures are validated relative to a given theory stated in typed feature logic. These constraints may involve constraints on the range of a feature's value, constraints on which features are permitted within certain types of feature structures, or constraints that prevent the co-occurrence of certain feature-value pairs. The source of these constraints is normally the empirical domain being modelled.

— It provides a mechanism by which the encoder can define the intended interpretation of underspecified feature structures. This involves defining default values (whether literal or computed) for missing features.

The scheme described in this part of ISO 24610 may be used to document any feature system, but is primarily intended for use with the typed feature structure representation defined in ISO 24610-1. The feature structure representations of ISO 24610-1 specify data structures that are subject to the typing conventions and constraints specified using ISO 24610-2. The feature structure representations of ISO 24610-1 are also used within some of the elements defined in ISO 24610-2.

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 24610-1:2006, *Language resource management — Feature structures — Part 1: Feature structure representation*

ISO/IEC 19757-2, *Information technology — Document Schema Definition Language (DSDL) — Part 2: Regular-grammar-based validation — RELAX NG*