INTERNATIONAL STANDARD

ISO/IEC 13250-4

First edition
2009-03-01

Information technology — Topic Maps —

Part 4: Canonicalization

Technologies de l'information — Plans relatifs à des sujets —

Partie 4: Canonicalisation

Reference number
ISO/IEC 13250-4:2009(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) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

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

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national 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 and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 13250-4 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 34, *Document description and processing languages*.

ISO/IEC 13250 consists of the following parts, under the general title *Information technology — Topic Maps*:

— *Part 2: Data model*

— *Part 3: XML syntax*

— *Part 4: Canonicalization*

The following parts are under preparation.

— *Part 1: Overview and basic concepts*

— *Part 5: Reference model*

— *Part 6: Compact syntax*

— *Part 7: Graphical notation*

Introduction

This part of ISO/IEC 13250 defines a format known as Canonical XTM, or CXTM for short. The format is an XML format, and has the property that it guarantees that two equivalent Topic Maps Data Model instances (ISO/IEC 13250-2) will always produce byte-by-byte identical serializations, and that non-equivalent instances will always produce different serializations. CXTM thus enables direct comparison of two topic maps to determine equality by comparison of their canonical serializations.

The purpose of CXTM is to allow the creation of test suites for various Topic Maps-related technologies that are easily portable between different Topic Maps implementations, so long as these support CXTM.

CXTM is *not* intended to be used for the interchange of topic maps, although this is possible. The standard format for interchange of topic maps is XTM (ISO/IEC 13250-3).

Information technology — Topic Maps —
Part 4: Canonicalization

1 Scope

This part of ISO/IEC 13250 defines the CXTM format, and specifies how CXTM files are produced from topic maps by means of a transformation from the Topic Maps Data Model (ISO/IEC 13250-2) to the XML Infoset *[XML Infoset]*.

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.

NOTE Each of the following documents has a unique identifier that is used to cite the document in the text. The unique identifier consists of the part of the reference up to the first comma.

ISO/IEC 10646, *Information technology — Universal Multiple-Octet Coded Character Set (UCS)*

Unicode, *The Unicode Standard, Version 5.0.0*, The Unicode Consortium, Reading, Massachusetts, USA, Addison-Wesley Developer's Press, 2007, ISBN 0-321-48091-0, http://www.unicode.org/versions/Unicode5.0.0/

RFC 3986, *Uniform Resource Identifier (URI): Generic Syntax*, Internet Standards Track Specification, January 2005, http://www.ietf.org/rfc/rfc3986.txt

XML-C14N, *Canonical XML, Version 1.0*, World Wide Web Consortium, 15 March 2001, available at <http:// www.w3.org/TR/2001/REC-xml-c14n-20010315>

XML Infoset, *XML Information Set (Second Edition)*, World Wide Web Consortium, 4 February 2004, available at <http://www.w3.org/TR/2004/REC-xml-infoset-20040204>

ISO/IEC 13250-2, *Information technology — Topic Maps — Part 2: Data model*

XMLSCHEMA-2, *XML Schema Part 2: Datatypes Second Edition*, World Wide Web Consortium, 28 October 2004, available at <http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/>

3 Canonicalization

3.1 Introduction

The canonicalization process takes two parameters: a topic map item (that is, an instance of the Topic Maps Data Model, defined in ISO/IEC 13250-2) and a base locator. The process produces a canonicalization of the topic map, with all locators in the topic map rewritten to be relative to the given base locator. The purpose of the base locator is to allow references to the local filesystem to be stripped out, thus making CXTM test cases portable between different systems.

Canonicalization is performed in three steps:

1. A document information item representing the CXTM document is produced from the topic map item as described in 3.3.
2. For each element information item that is a descendant of the document information item from the previous step, the following operations are performed:

    — A character information item is added to the [[children]] property of the information item in the element's [[parent]] property immediately after the element itself. The character information item's [[character code]] property is set to #x0A.

— If the element's [[local name]] property is set to "topicMap", "topic", "name", "variant", "occurrence", "association", "role", "scope", "itemIdentifiers", "subjectLocators", or "subjectIdentifiers", a character information item is added to the [[children]] property of the element as the first element. The character information item's [[character code]] property is set to #x0A.

3. The document information item is serialized to a Canonical XML representation as described in *[XML-C14N]*.

3.2 Notational conventions

Information item properties from [W3C XML-Infoset] are referred to using [[property name]], in order to distinguish them from properties from ISO/IEC 13250-2.

3.3 CXTM document information item

There is exactly one CXTM document information item in the XML Infoset generated by the canonicalization of the topic map item.

The CXTM document information item has the following properties:

1. [[children]] A list containing only the representation of the topic map item
2. [[document element]] The element information item that represents the topic map item
3. [[notations]] The empty set
4. [[unparsed entities]] The empty set
5. [[base URI]] No value
6. [[standalone]] No value
7. [[version]] No value
8. [[all declarations processed]] False

3.4 Constructing a representation of a topic map item

A topic map item is represented by an element information item with the following properties:

1. [[local name]] The string "topicMap"
2. [[children]] A list of element information items in the following order:

    1. A representation of the [item identifiers] property, if any
    2. A representation of each topic item in the [topics] property of the topic map item in canonical sort order
    3. A representation of each association item in the [associations] property of the topic map item in canonical sort order
3. [[attributes]] A representation of the [reifier] property

3.5 Constructing a representation of a topic item

A topic item is represented by an element information item with the following properties:

1. [[local name]] The string "topic"
2. [[children]] A list of element information items in the following order:

    1. If the value of [subject identifiers] property of the topic item is not the empty set, then an element information item with the following properties:

        1. [[local name]] The string "subjectIdentifiers"
        2. [[children]] A representation of each locator in the [subject identifiers] property in canonical sort order
        3. [[attributes]] The empty set
    2. If the value of the [subject locators] property of the topic item is not the empty set, then an element information item with the following properties:

        1. [[local name]] The string "subjectLocators"
        2. [[children]] A representation of each locator in the [subject locators] property in canonical sort order
        3. [[attributes]] The empty set

3. A representation of the [item identifiers] property, if any
4. A representation of each of the topic name items of the [topic names] property in canonical sort order
5. A representation of each of the occurrence items of the [occurrences] property in canonical sort order
6. For each of the association role items of the [roles played] property in canonical sort order, an element information item with the following properties

    1. [[local name]] set to the string "rolePlayed"
    2. [[children]] An empty list
    3. [[attributes]] A set containing one attribute information item as follows:

        1. [[local name]] set to the string "ref"
        2. [[normalized value]] A sequence of character information items representing a string value constructed by the concatenation of:

            1. The string "association."
            2. The position of the association item which is the value of the [parent] property of the association role item, in the canonically sorted [associations] property of the parent topic map item
            3. The string ".role."
            4. The position of the association role item in the canonically sorted [roles] property of the parent association item

3. [[attributes]] A set containing the number attribute for this information item.

3.6 Constructing a representation of the topic name item

Each topic name item is represented by an element information item with the following properties:

1. [[local name]] The string "name"
2. [[children]] A list of element information items in the following order:

    1. A representation of the [value] property
    2. A representation of the [type] property
    3. A representation of the [scope] property
    4. A representation of each of the variant items in the [variants] property in canonical sort order
    5. A representation of the [item identifiers] property, if any

3. [[attributes]] The union of:

    — A representation of the [reifier] property
    — The number attribute for this information item

3.7 Constructing a representation of a variant item

A variant item is represented by an element information item with the following properties:

1. [[local name]] The string "variant"
2. [[children]] A list of element information items in the following order:

    1. A representation of the [value] property
    2. A representation of the [datatype] property
    3. A representation of the [scope] property
    4. A representation of the [item identifiers] property, if any

3. [[attributes]] The union of:

    — A representation of the [reifier] property
    — The number attribute for this information item

3.8 Constructing a representation of an occurrence item

An occurrence item is represented by an element information item with the following properties:

1. [[local name]] The string "occurrence"

2. [[children]] A list of element information items in the following order:

    1. A representation of the [value] property
    2. A representation of the the [datatype] property
    3. A representation of the [type] property
    4. A representation of the [scope] property
    5. A representation of the [item identifiers] property, if any
3. [[attributes]] The union of:

    — A representation of the [reifier] property
    — The number attribute for this information item

3.9 Constructing a representation of an association item

An association item is represented by an element information item with the following properties:

1. [[local name]] The string "association"
2. [[children]] A list of element information items in the following order:

    1. A representation of the [type] property
    2. A representation of each of the items of the [roles] property in canonical sort order
    3. A representation of the [scope] property
    4. A representation of the [item identifiers] property, if any
3. [[attributes]] The union of:

    — A representation of the [reifier] property
    — The number attribute for this information item

3.10 Constructing a representation of the association role item

An association role item is represented by an element information item with the following properties:

1. [[local name]] The string "role"
2. [[children]] A list of element information items in the following order:

    1. An element information item with the following properties:

        1. [[local name]] The string "player"
        2. [[children]] The empty list
        3. [[attributes]] A set of one attribute information item with the following properties:

            1. [[local name]] The string "topicref"
            2. [[normalized value]] The position of the topic item in the [player] property within the canonically sorted [topics] property of the parent topic map item
    2. A representation of the [type] property
    3. A representation of the [item identifiers] property, if any
3. [[attributes]] The union of:

    — A representation of the [reifier] property
    — The number attribute for this information item

3.11 Constructing a representation of the [reifier] property

If the [reifier] property of an information item is null it is represented by the empty set. Otherwise it is represented as a set containing an attribute information item with the following properties:

1. [[local name]] The string "reifier"
2. [[normalized value]] The position of the topic item that is the value of the [reifier] property in the canonically sorted list of all topic items

3.12 Constructing a representation of the [scope] property

If the [scope] property of an information item is the empty set, then it has no representation. Otherwise it is represented by an element information item with the following properties:

1. [[local name]] The string "scope"
2. [[children]] A list of one element information item for each topic item in the value of the [scope] property in canonical sort order. Each element information item has the following properties:
    1. [[local name]] The string "scopingTopic"
    2. [[children]] An empty list
    3. [[attributes]] A list containing a single attribute information item with the following properties:
        1. [[local name]] The string "topicref"
        2. [[normalized value]] The position of the topic item within the canonically sorted list of all topic items in the topic map item being canonicalized
3. [[attributes]] The empty set

3.13 Constructing a representation of the [item identifiers] property

If the [item identifiers] property of an information item is the empty set it has no representation. Otherwise it is represented by an element information item with the following properties:

1. [[local name]] The string "itemIdentifiers"
2. [[children]] A representation of each locator in the [item identifiers] property in canonical sort order
3. [[attributes]] The empty set

3.14 Constructing a representation of the [datatype] property

The [datatype] property of an information item is represented by an element information item with the following properties:

1. [[local name]] The string "datatype"
2. [[children]] A sequence of character information items representing the string value of the normalized locator in the [datatype] property
3. [[attributes]] The empty set

3.15 Constructing a representation of the [type] property

The [type] property of an information item is represented by an element information item with the following properties:

1. [[local name]] The string "type"
2. [[children]] An empty list
3. [[attributes]] A set containing an attribute information item with the following properties:
    1. [[local name]] The string "topicref"
    2. [[normalized value]] The position of the topic item that is the value of the [type] property within the canonically sorted list of all topic items in the Topic Maps Data Model being encoded.

3.16 Constructing a representation of the [value] property

A [value] property of an information item is represented by an element information item with the following properties:

1. [[local name]] The string "value"
2. [[children]] A sequence of character information items corresponding to the string representation of the [value] property, as defined below.
3. [[attributes]] The empty set

The string representation of the [value] property depends on the [datatype] property of the same information item. The representation is produced by following the procedure under the appropriate heading below. If the information item has no [datatype] property the procedure under the heading "Other" is to be used.