ISO/IEC JTC 1/SC 34

Date: 2003-10-22

ISO/IEC 13250-4

ISO/IEC JTC 1/SC 34/WG 3

Secretariat: ANSI

Topic Maps — Part 4: Canonicalization

Document type: International Standard
Document subtype:
Document stage:
Document language: E

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.

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 *Topic Maps*:

— *Part 1: Overview and Basic Concepts*

— *Part 2: Data Model*

— *Part 3: XML Syntax*

— *Part 4: Canonicalization*

Introduction

Topic maps are abstract structures that can encode knowledge and connect this encoded knowledge to relevant information resources. Topic maps are organized around topics, which represent subjects of discourse; associations, representing relationships between the subjects; and occurrences, which connect the subjects to pertinent information resources.

Topic maps may be represented in many ways: using topic map syntaxes in files, inside databases, as internal data structures in running programs, and even mentally in the minds of humans. All these forms are different ways of representing the same abstract structure, the Topic Maps Data Model defined in Part 2 of this standard.

Canonicalization is the process of serializing a data structure in such a way that two data structures considered to be the same result in the same serialization and two data structures not considered to be the same result in two different serializations. A canonical form enables direct comparison of two data model instances to determine equality by comparison of their canonical serialization.

This part of ISO/IEC 13250 defines a canonical sort order for any set of of information items from the Topic Maps Data Model and a transformation of an instance of the Topic Maps Data Model to an instance of the XML Infoset model. The canonical sort order defined here can be applied not only to the set properties defined by the Topic Maps Data Model but also to other sets of topic map information items such as those generated as the result of processing a query.

This part of ISO/IEC 13250 also defines a transformation from the Topic Maps Data Model to the XML Infoset. Applications which serialize the XML Infoset model created by applying the transformation defined in this part of ISO/IEC 13250 must do so according to [XML-C14N]. When this serialization is performed, the resulting output string is a canonical representation of the Topic Maps Data Model instance.

Topic Maps — Part 4: Canonicalization

1   Scope

This part of ISO/IEC 13250 specifies an algorithm for the canonicalization of an instance of the Topic Maps Data Model. It defines a canonical ordering for every information item defined by the Topic Maps Data Model and an XML serialisation of the information items and all of their properties. When the XML is serialized in accordance with [XML-C14N], the serialized file is the canonical representation of the Topic Maps Data Model instance.

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.

7. Association role

8. Locator

9. String

10. Set

11. NULL

3.3 Comparison Of String Property Values

Ed. Note.
A lexical comparison. Need to be specific about this.

3.4 Comparison Of Set Property Values

1. Sets sort in order of the number of elements in the collection.

2. For sets of equal size, first sort the items of each set into their canonical ordering. Starting with the lowest item

Variant information items are compared by comparing their properties in the following order.

1. [value]

2. [resource]

3. [scope]

4. [parent]

3.9   Canonical Sort Order For Occurrence Information Items

Issue (prop-parent)
SAM issue prop-parent

Occurrence information items are compared by comparing their properties in the following order.

1. [value]

2. [resource]

3. [type]

4. [scope]

5. [parent]

3.10 Canonical Sort Order For Association Information Items

Association information items are compared by comparing their properties in the following order.

1. [type]

2. [roles]

3. [scope]

3.11 Canonical Sort Order For Association Role Information Items

AssociationRole information items are compared by comparing their properties in the following order.

1. [role playing topic]

2. [type]

3. [scope]

4. [parent]

4   Transformation Of Topic Map Data Model To CXTM XML Infoset

4.1   Introduction

The transformation process creates an instance of the XML Infoset data model from an instance of the Topic Maps data model. The XML Infoset data model created is rooted with a CXTM document information item.

Throughout this clause the value of the [parent] property of element information items and attribute information items is not specified. The [parent] property of an element information item must always be set to the element or

document information item of which the element information item is a direct child. The [parent] property of an attribute information item must be set to the element information item of which the attribute is a child.

The names of properties in the Topic Maps Data Model and in the XML Infoset data model are written in square brackets: [property name].

4.2 CXTM Document Information Item

There is exactly one CXTM document information item in the XML Infoset generated by the canonicalization of the Topic Maps Data Model.

The CXTM document information item has the following properties:

1. [children] A list containing only the representation of the topic map information item in the Topic Maps Data Model instance.

2. [document element] The element information item that represents the topic map information item in the Topic Maps Data Model instance.

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.

4.3 Constructing a representation of a topic map information item

A topic map information item in the Topic Maps Data Model 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 each topic information item in the [topics] property of the topic map information item in canonical sort order.

    2. A representation of each association information item in the [associations] property of the topic map information item in canonical sort order.

3. [attributes] If the value of the [reifier] property is not null, then a representation of the [reifier] property otherwise an empty list.

4. [namespace attributes] The empty set.

5. [in-scope namespaces] The empty set.

6. [base URI] No value.

4.4 Constructing a representation of a topic information item

A topic information item is represented by an element information item in the XML Infoset. The element information item has the following properties.

1. [local name] The string "topic"

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

    1. A representation of each of the topic name information items of the [topic names] property in canonical sort order.

    2. A representation of each of the occurrence information items of the [occurrences] property in canonical sort order.

    3. 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 list of locator element information items encoding the value of the [subject identifiers] property.

        3. [attributes] An empty list.

        4. [namespace attributes] The empty set.

        5. [in-scope namespaces] The empty set.

        6. [base URI] No value.

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

        1. [local name] The string "subjectAddress"

        2. [children] A single locator element information item encoding the value of the [subject address] property.

        3. [attributes] An empty list.

        4. [namespace attributes] The empty set.

        5. [in-scope namespaces] The empty set.

        6. [base URI] No value.

3. [attributes] If the value of the [reifier] property is not null, then a representation of the [reifier] property, otherwise an empty list.

4. [namespace attributes] The empty set.

5. [in-scope namespaces] The empty set.

6. [base URI] No value.

4.5   Constructing a representation of the topic name information item

Each topic name information item in the Topic Maps Data Model is represented by an element information item in the XML infoset. The element information item has the following properties.

1. [local name] The string "topicName"

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 "value"

    2.   [children] A list of character information items. The list contains one character information item for each character in the string value of the [value] property of the topic name information item. For each character information item, the [character code] property must be the ISO 10646 character code of the character.

    3.   [attributes] An empty list.

    4.   [namespace attributes] The empty set.

    5.   [in-scope namespaces] The empty set.

    6.   [base URI] No value.

  2.   If the value of the [type] property is not null, a representation of the [type] property.

  3.   If the value of the [scope] property is not null, a representation of the [scope] property.

  4.   A representation of each of the items of the [variants] property in canonical sort order.

3.   [attributes] If the value of the [reifier] property is not null, then a representation of the [reifier] property otherwise an empty list.

4.   [namespace attributes] The empty set.

5.   [in-scope namespaces] The empty set.

6.   [base URI] No value.

4.6 Constructing a representation of a variant information item

A variant information item in the Topic Maps Data Model is represnted 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.   If the value of the [value] property is not empty or null, an element information item with the following properties:

    1.   [local name] The string "value"

    2.   [children] A list of character information items. The list contains one character information item for each character in the string value of the [value] property of the variant information item. For each character information item, the [character code] property must be the ISO 10646 character code of the character and the [parent] property must be set to this value element information item.

    3.   [attributes] An empty list.

    4.   [namespace attributes] The empty set.

    5.   [in-scope namespaces] The empty set.

    6.   [base URI] No value.

  2.   If the value of the [resource] property is not null, an element information item representing the value of the [resource] property.

  3.   If the value of the [scope] property is not the empty set, an element information item representing the

[scope] property.

3. [attributes] If the value of the [reifier] property is not null, then a representation of the [reifier] property otherwise an empty list.

4. [namespace attributes] The empty set.

5. [in-scope namespaces] The empty set.

6. [base URI] No value.

4.7   Constructing a representation of an occurrence information item

An occurrence information item in the Topic Maps Data Model is represented by an element information item in the CXTM XML Infoset.

1. [local name] The string "occurrence"

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

    1. If the value of the [value] property is not null, then an element information item with the following properties:

        1. [local name] The string "value"

        2. [children] A list of character information items. The list contains one character information item for each character in the string value from the Topic Maps Data Model being encoded in this value element information item. For each character information item, the [character code] property must be the ISO 10646 character code of the character and the [parent] property must be set to this value element information item.

        3. [attributes] An empty list.

        4. [namespace attributes] The empty set.

        5. [in-scope namespaces] The empty set.

        6. [base URI] No value.

    2. If the value of the [resource] property is not null, then an element information item representing the value of the [resource] property.

    3. If the value of the [scope] property is not the empty set, then an element information item representing the [scope] property.

    4. If the value of the [type] property is not null, then an element information item representing the [type] property.

3. [attributes] If the value of the [reifier] property is not null, then a representation of the [reifier] property otherwise an empty list.

4. [namespace attributes] The empty set.

5. [in-scope namespaces] The empty set.

6. [base URI] No value.

4.8   Constructing a representation of an association information item

An association information item in the Topic Maps Data Model is represented by an element information item in the CXTM XML Infoset with the following properties:

1. [local name] The string "association"

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

    1. If the value of the [scope] property is not the empty set, then an element information item representing the [scope] property of the association information item.

    2. If the value of the [type] property of the association information item is not null, then an element information item representing the [type] property of the association information item.

    3. A representation of each of the items of the [roles] property of the association information item in canonical sort order.

3. [attributes] If the value of the [reifier] property is not null, then a representation of the [reifier] property otherwise an empty list.

4. [namespace attributes] The empty set.

5. [in-scope namespaces] The empty set.

6. [base URI] No value.

4.9   Constructing a representation of the association role information item

An association role information item in the Topic Maps Data Model 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. If the value of the [role playing topic] property is not null, then an element information item with the following properties:

        1. [local name] The string "rolePlayer"

        2. [children] The empty list

        3. [attributes] A list containing one attribute information item with the following properties:

            1. [local name] The string "topicref"

            2. [normalized valued] The string representation of the position of the topic information item that is the value of the encoded property within the canonically sorted list of all topic information items in the Topic Maps Data Model being encoded. The canonically sorted list is numbered from 1 for the first topic information item.

            3. [specified] The boolean value true

            4. [owner element] The element information item that this attribute information item is a child of.

        4. [namespace attributes] The empty set.

        5. [in-scope namespaces] The empty set.

        6. [base URI] No value.

    2. If the value of the [type] property is not null, then an element information item representing the [type] property.

3. [attributes] If the value of the [reifier] property is not null, then a representation of the [reifier] property

otherwise an empty list.

4. [namespace attributes] The empty set.

5. [in-scope namespaces] The empty set.

6. [base URI] No value.

4.10 Constructing a representation of a locator information item

A locator information item in the Topic Maps Data Model is represented by an element information item with the following properties:

1. [local name] The string "locator"

2. [children] An empty list.

3. [attributes] A list containing the following attribute information items:

    1. An attribute information item with [local name] set to "notation", [normalized value] set to the value of the [notation] property of the locator information item, [specified] set to the boolean value true and [owner element] set to the element information item that represents the locator information item.

    2. An attribute information item with [local name] set to "address", normalized value] set to the value of the [address] property of the locator information item, [specified] set to the boolean value true and [owner element] set to the element information item that represresents the locator information item.

4. [namespace attributes] The empty set.

5. [in-scope namespaces] The empty set.

6. [base URI] No value.

4.11 Constructing a representation of the [reifier] property

The [reifier] property of a topic map information item, topic information item, topic name information item, variant information item, occurrence information item, association information item or association role information item is represented as an attribute information item with the following properties:

1. [local name] The string "reifier"

2. [normalized valued] The string representation of the position of the topic information item that is the value of the [reifier] property within the canonically sorted list of all topic information items in the Topic Maps Data Model being encoded. The canonically sorted list is numbered from 1 for the first topic information item.

3. [specified] The boolean value true

4. [owner element] The element information item that this attribute information item is a child of.

4.12 Constructing a representation of the [resource] property

The [resource] property of a variant information item or occurrence information item is represented by an element information item with the following properties:

1. [local name] The string "resource"

2. [children] An empty list.

3. [attributes] A list containing the following attribute information items:

1. An attribute information item with [local name] set to "notation", [normalized value] set to the value of the [notation] property of the locator information item that is the value of the [resource] property, [specified] set to the boolean value true and [owner element] set to the element information item that represents the variant information item or occurrence information item to which this [resource] property belongs.

2. An attribute information item with [local name] set to "address", [normalized value] set to the value of the [address] property of the locator information item that is the value of the [resource] property, [specified] set to the boolean value true and [owner element] set to the element information item that reprepresents the variant information item or occurrence information item to which this [resource] property belongs.

4. [namespace attributes] The empty set.

5. [in-scope namespaces] The empty set.

6. [base URI] No value.

4.13 Constructing a representation of the [scope] property

The [scope] property of a topic name information item, variant information item, occurrence information item or association information item 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 information 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 valued] The string representation of the position of the topic information item that is the value of the [type] property within the canonically sorted list of all topic information items in the Topic Maps Data Model being encoded. The canonically sorted list is numbered from 1 for the first topic information item.

      3. [specified] The boolean value true

      4. [owner element] The element information item that this attribute information item is a child of.

   4. [namespace attributes] The empty set.

   5. [in-scope namespaces] The empty set.

   6. [base URI] No value.

3. [attributes] An empty list.

4. [namespace attributes] The empty set.

5. [in-scope namespaces] The empty set.

6. [base URI] No value.

4.14 Constructing a representation of the [type] property

The [type] property of a topic name information item, occurrence information item, association information item or association role 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 list containing an attribute information item with the following properties:

    1. [local name] The string "topicref"

    2. [normalized valued] The string representation of the position of the topic information item that is the value of the [type] property within the canonically sorted list of all topic information items in the Topic Maps Data Model being encoded. The canonically sorted list is numbered from 1 for the first topic information item.

    3. [specified] The boolean value true

    4. [owner element] The element information item that this attribute information item is a child of.

4. [namespace attributes] The empty set.

5. [in-scope namespaces] The empty set.

6. [base URI] No value.

4.15 Constructing a representation of the [value] property

Ed. Note.
This clause is currently not referenced from anywhere, but is kept here for now because it might be a better way to handle the [value] property which recurs three times.

A [value] property in the Topic Maps Data Model is represented by an element information item with the following properties:

1. [local name] The string "value"

2. [children] A list of character information items. The list contains one character information item for each character in the string value from the Topic Maps Data Model being encoded in this value element information item. For each character information item, the [character code] property must be the ISO 10646 character code of the character and the [parent] property must be set to this value element information item.

3. [attributes] An empty list.

4. [namespace attributes] The empty set.

5. [in-scope namespaces] The empty set.

6. [base URI] No value.