
**Information technology — Topic Maps —
Part 4:
Canonicalization**

*Technologies de l'information — Plans relatifs à des sujets —
Partie 4: Canonicalisation*

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2009

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing 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

Contents

Page

Foreword.....	iv
Introduction.....	v
1 Scope.....	1
2 Normative references.....	1
3 Canonicalization.....	1
3.1 Introduction.....	1
3.2 Notational conventions.....	2
3.3 CXTM document information item.....	2
3.4 Constructing a representation of a topic map item.....	2
3.5 Constructing a representation of a topic item.....	2
3.6 Constructing a representation of the topic name item.....	3
3.7 Constructing a representation of a variant item.....	3
3.8 Constructing a representation of an occurrence item.....	3
3.9 Constructing a representation of an association item.....	4
3.10 Constructing a representation of the association role item.....	4
3.11 Constructing a representation of the [reifier] property.....	4
3.12 Constructing a representation of the [scope] property.....	5
3.13 Constructing a representation of the [item identifiers] property.....	5
3.14 Constructing a representation of the [datatype] property.....	5
3.15 Constructing a representation of the [type] property.....	5
3.16 Constructing a representation of the [value] property.....	5
3.17 Constructing a representation of locator values.....	6
3.18 Normalizing locator values.....	6
3.19 Constructing the number attribute.....	6
3.20 Encoding of string properties.....	7
3.21 Encoding of positional values.....	7
3.22 Default property values for element information items.....	7
3.23 Default property values for attribute information items.....	7
4 Canonical sort order.....	7
4.1 Introduction.....	7
4.2 Information type and basic type sort order.....	7
4.3 Comparison of strings.....	8
4.4 Comparison of sets.....	8
4.5 Comparison order for locators.....	8
4.6 Canonical sort order for topic items.....	8
4.7 Canonical sort order for topic name items.....	8
4.8 Canonical sort order for variant items.....	8
4.9 Canonical sort order for occurrence items.....	8
4.10 Canonical sort order for association items.....	9
4.11 Canonical sort order for association role items.....	9
Annex A (informative) RELAX-NG schema for CXTM.....	10
Bibliography.....	12

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/>