

---

---

**Information technology — Topic Maps —  
Part 6:  
Compact syntax**

*Technologies de l'information — Plans relatifs à des sujets —  
Partie 6: Syntaxe compacte*

**iTeh STANDARD PREVIEW  
(standards.iteh.ai)**

ISO/IEC 13250-6:2010

<https://standards.iteh.ai/catalog/standards/sist/cf5f762f-0fe4-4783-87a1-f2510b838b76/iso-iec-13250-6-2010>

**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.

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

[ISO/IEC 13250-6:2010](https://standards.iteh.ai/catalog/standards/sist/cf5f762f-0fe4-4783-87a1-f2510b838b76/iso-iec-13250-6-2010)

<https://standards.iteh.ai/catalog/standards/sist/cf5f762f-0fe4-4783-87a1-f2510b838b76/iso-iec-13250-6-2010>



**COPYRIGHT PROTECTED DOCUMENT**

© ISO/IEC 2010

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](mailto:copyright@iso.org)  
Web [www.iso.org](http://www.iso.org)

Published in Switzerland

# Contents

Page

Foreword.....	iv
Introduction.....	v
1 Scope.....	1
2 Normative references.....	1
3 Syntax description.....	1
3.1 About the syntax.....	1
3.2 Deserialization.....	1
3.3 Common syntactical constructs.....	2
3.3.1 Whitespace.....	2
3.3.2 Comments.....	2
3.3.3 Creating IRIs from strings.....	2
3.3.4 Creating IRIs from QNames.....	3
3.3.5 IRI References.....	3
3.3.6 Topic Identity.....	3
3.3.7 Topic References.....	4
3.3.8 Creating locators from wildcards.....	4
3.3.9 Scope.....	4
3.3.10 Reifier.....	4
3.3.11 Type.....	4
3.4 Literals.....	4
3.4.1 General.....	4
3.4.2 String Escape Sequences.....	5
3.5 Topic Map.....	6
3.6 Encoding Directive.....	6
3.7 Version Directive.....	6
3.8 Topics.....	6
3.9 Topic Tail.....	7
3.10 Occurrences.....	7
3.11 Names.....	7
3.12 Variants.....	7
3.13 Associations.....	8
3.14 Templates.....	8
3.15 Template Invocation.....	9
3.16 Directives.....	10
3.16.1 Prefix Directive.....	10
3.16.2 Include Directive.....	10
3.16.3 Mergemap Directive.....	10
Annex A (informative) CTM integer.....	12
Annex B (informative) Syntax.....	13
Bibliography.....	14

## 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 13250-6 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 1: Overview and basic concepts*

— *Part 2: Data model*

— *Part 3: XML syntax*

— *Part 4: Canonicalization* <https://standards.iteh.ai/catalog/standards/sist/cf5f762f-0fe4-4783-87a1-f2510b838b76/iso-iec-13250-6-2010>

— *Part 5: Reference model*

— *Part 6: Compact syntax*

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

[ISO/IEC 13250-6:2010](https://standards.iteh.ai/catalog/standards/sist/cf5f762f-0fe4-4783-87a1-f2510b838b76/iso-iec-13250-6-2010)

<https://standards.iteh.ai/catalog/standards/sist/cf5f762f-0fe4-4783-87a1-f2510b838b76/iso-iec-13250-6-2010>

## Introduction

CTM (Compact Topic Maps) is a text-based notation for representing topic maps. It provides a simple, lightweight notation that complements the existing XML-based interchange syntax defined in ISO/IEC 13250-3:2007 and can be used for

- manually authoring topic maps;
- providing human-readable examples in documents;
- serving as a common syntactic basis for TMCL and TMQL.

The principal design criteria of CTM are compactness, ease of human authoring, and maximum readability. CTM supports all constructs of ISO/IEC 13250-2, except item identifiers on constructs that are not topics.

This part of ISO/IEC 13250 should be read in conjunction with ISO/IEC 13250-2 since the interpretation of the CTM syntax is defined through a mapping from the syntax to the data model there defined.

## iTeh STANDARD PREVIEW (standards.iteh.ai)

[ISO/IEC 13250-6:2010](https://standards.iteh.ai/catalog/standards/sist/cf5f762f-0fe4-4783-87a1-f2510b838b76/iso-iec-13250-6-2010)

<https://standards.iteh.ai/catalog/standards/sist/cf5f762f-0fe4-4783-87a1-f2510b838b76/iso-iec-13250-6-2010>

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

ISO/IEC 13250-6:2010

<https://standards.iteh.ai/catalog/standards/sist/cf5f762f-0fe4-4783-87a1-f2510b838b76/iso-iec-13250-6-2010>

# Information Technology — Topic Maps —

## Part 6:

# Compact syntax

## 1 Scope

This part of ISO/IEC 13250 defines a text-based notation for representing instances of the data model defined in ISO/IEC 13250-2. It also defines a mapping from this notation to the data model. The syntax is defined through an Extended Backus-Naur Form (EBNF) grammar.

## 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.

IANA-CHARSETS, *CHARACTER SETS*, Internet Assigned Numbers Authority, 14 May 2007, available at <<http://www.iana.org/assignments/character-sets>>

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

XSDT, *XML Schema Part 2: Datatypes Second Edition*, W3C Recommendation, 28 October 2004, available at <<http://www.w3.org/TR/xmlschema-2/>>

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

IETF RFC 3987, *Internationalized Resource Identifiers (IRIs)*, Internet Standards Track Specification, January 2005, available at <<http://www.ietf.org/rfc/rfc3987.txt>>

## 3 Syntax description

### 3.1 About the syntax

The acronym *CTM* is often used to refer to the syntax defined in this part of ISO/IEC 13250. Its full name is Compact Topic Maps Syntax.

This clause defines the syntax of CTM using an EBNF grammar based on the notation described in XML 1.0. It uses prose to define the mapping from CTM to ISO/IEC 13250-2. The full EBNF can be found in Annex A.

### 3.2 Deserialization

This clause defines how instances of the CTM syntax are deserialized into instances of the data model defined in ISO/IEC 13250-2. Serialization is only implicitly defined, but implementations should produce CTM serializations that, when deserialized to a new data model instance, produce a data model instance that has the same canonicalization as the original data model instance, according to ISO/IEC 13250-4: 2009.

Each CTM instance shall produce a valid data model instance according to ISO/IEC 13250-2.

The input to the deserialization process is:

- An optional character set name according to IANA-CHARSETS which specifies the encoding of the CTM instance.
- A byte stream which is converted into a character stream by the following steps:
  - If the optional input character set name is provided, the encoding is set to the provided value.
  - If the byte stream starts with EF BB BF (UTF-8 byte order mark), the encoding is set to "UTF-8".

It is an error if the encoding is already set to another value due to the optional character set name.

- If the next bytes in the stream contain the sequence 25 65 6E 63 6F 64 69 6E 67 (%encoding in ASCII), all following 09 or 20 bytes are skipped until the very first 22 byte is read. The following byte sequence until the next 22 (exclusive) byte is interpreted as character set name according to IANA-CHARSETS.

It is an error if the encoding is already set to another value.

- If the previous steps did not produce an encoding, the encoding shall be set to "UTF-8"
- The byte stream is converted according to the encoding into Unicode (c.f. ), where the optional byte sequence EF BB BF is removed previously from the start of the byte stream.

This character stream is processed according to the grammar specified by this part of ISO/IEC 13250.

- An absolute IRI. This is the IRI from which the byte stream was retrieved, known as the document IRI. This IRI shall always be provided, as it is necessary in order to assign the item identifiers of the topic items created during deserialization. If the CTM instance was not read from any particular IRI the application is responsible for providing an IRI considered suitable.
- An absolute IRI which will be used to resolve wildcards against 3.3.8; this IRI is called "wildcard-iri". If the wildcard-iri is not provided its value is set to the document IRI.
- A non-negative integer, called "wildcard-counter". If the wildcard-counter is not provided, the wildcard-counter is initialized with 0.

STANDARD PREVIEW  
(standards.iteh.ai)

Deserialization is performed by processing each component of the CTM source in document order. Components are defined in terms of text that matches a syntactic variable of the EBNF. For each component encountered the operations specified in the clause for the corresponding syntactic variable are performed.

2510b838b76/iso-iec-13250-6-2010

Whenever a new information item is created, those of its properties which have set values are initialized to the empty set; all other properties are initialized to null.

### 3.3 Common syntactical constructs

#### 3.3.1 Whitespace

*Whitespace* consists of one or more space (#x20) characters, carriage returns, line feeds, or tabs.

Whitespace character are allowed everywhere to separate tokens (terminals and non-terminals).

#### 3.3.2 Comments

*Comments* are fragments of the character stream which are ignored by a CTM processor. Comments are allowed where whitespace characters are allowed.

Multiline comments are delimited with #( and )# and may be nested.

Single line comments are introduced by a hash (#) and continue until the end of the current line, or until the end of the character stream, whichever comes first.

#### 3.3.3 Creating IRIs from strings

The delimiters < and > are removed from iri-delimited; the resulting string may represent either an absolute or a relative IRI which shall meet the requirements of IETF RFC 3986 and IETF RFC 3987.



To create an IRI from a string, unescape the string by replacing %HH escape sequences with the characters they represent, and decode the resulting character sequence from UTF-8 to a sequence of abstract Unicode characters. The resulting string is turned into an absolute IRI by resolving it against the document IRI.

### 3.3.4 Creating IRIs from QNames

QNames are used to abbreviate IRIs. The syntax of QNames is as follows:

A QName causes a locator to be created. During deserialization, the IRI to which the prefix is bound is concatenated with the local part. The result of such a process is always an absolute IRI.

It is an error if the prefix has not been bound to an IRI as specified in 3.16.1.

```
%prefix isbn urn:isbn:
```

```
isbn:3-7026-4850-X isa book;
- "Das kleine Ich bin Ich".
```

### 3.3.5 IRI References

IRI references are either QNames or IRIs. They are interchangeable: Everywhere an IRI can be used a QName can also be used (provided the prefix is defined at that point).

### 3.3.6 Topic Identity

iTeh STANDARD PREVIEW

(standards.iteh.ai)

Topics are referenced by an item identifier, a subject identifier, or a subject locator.

During deserialization, one topic item is created for each topic-identity.

If the topic-identity is an identifier, a locator is created by concatenating the document IRI, a # character, and the value of the identifier. The locator is added to the [item identifiers] property of the topic item.

If the topic-identity is specified by a subject identifier, a locator is created and added to the [subject identifiers] property of the topic item.

If the topic-identity is specified by a subject locator, a locator is created (the leading = is not part of the locator) and added to the [subject locators] property of the topic item.

If the topic-identity is specified by an item identifier, a locator is created (the leading ^ is not part of the locator) and added to the [item identifiers] property of the topic item.

If the topic-identity is specified by a wildcard, the locator created in accordance with the procedure described in 3.3.8 is added to the [item identifiers] property of the topic item.

If the topic item created through deserialization of a topic-identity is equal to another topic item (c.f. ISO/IEC 13250-2 5.3); the two topic items are merged according to the procedure given in ISO/IEC 13250-2.

Variables shall occur only within templates.

```
# A topic referenced by the subject locator "http://www.isotopicmaps.org/"
= http://www.isotopicmaps.org/ .
```

```
# A topic referenced by a subject identifier
http://psi.example.org/John_Lennon .
```

```
# A topic with a unique item identifier. Within the CTM
```

# document it is not possible to reference this topic by item identifier.  
? - "A topic".

# A topic with a unique item identifier which may be  
# referenced again within the topic map.  
?foo - "A new, unique topic".

# Another reference to the same topic later in the document  
?foo isa subject .

### 3.3.7 Topic References

If the topic-ref is specified by a topic-identity, the topic item resolved by 3.3.6 is used.

If the topic-ref is an embedded-topic, a topic item is created. A locator is created according to the procedure described in 3.3.8, where the [ is hereby handled like an anonymous-wildcard. The locator is added to the [item identifiers] property of the topic item. The topic item is used as input for the topic-tail (3.9) procedure.

### 3.3.8 Creating locators from wildcards

If a wildcard is specified by an anonymous-wildcard, the wildcard-counter's value is incremented by 1. A locator is created by concatenating the wildcard-iri, the string #\$\_ and the value of the wildcard counter.

If a wildcard is specified by a named-wildcard, and that named-wildcard has not been processed before, the wildcard-counter's value is incremented by 1. A locator is created by concatenating the wildcard-iri, the string #\$\_, the value of the wildcard-counter, the . character and the identifier part of the named-wildcard.

If the named-wildcard has been processed before, the locator produced by its first occurrence is reused.

### 3.3.9 Scope

The scope construct is used to assign a scope to an information item.

During deserialization, each topic-ref element is processed according to the procedure described in 3.3.7. These topic items are gathered into a set that is assigned as the value of the [scope] property of the statement being processed.

### 3.3.10 Reifier

The reifier construct is used to refer from the statement on which it appears to the topic reifying that construct. The reference is a topic-ref as described in 3.3.7

During deserialization the topic-ref is resolved into a topic item following the procedure in 3.3.7. The topic item is set as the value of the [reifier] property of the statement being processed.

### 3.3.11 Type

The type construct is used to assign a type to the information item in which it occurs. The type is always a topic, indicated by topic-ref.

During deserialization the topic-ref produces a topic item following the procedure in 3.3.7, which is set as the value of the [type] property of the information item being processed.

## 3.4 Literals

### 3.4.1 General

A literal is a string value with an optional datatype.

For convenience, this part of ISO/IEC 13250 supports integers, decimals, IRIs, dates and dateTime values, which shall be valid according to XSDT, directly.

The following implicit datatypes are associated with the above mentioned literals:

**iri**  
<http://www.w3.org/2001/XMLSchema#anyURI>

**integer**  
<http://www.w3.org/2001/XMLSchema#integer>

**decimal**  
<http://www.w3.org/2001/XMLSchema#decimal>

**ctm-integer**  
<http://psi.topicmaps.org/iso13250/ctm-integer>

**date**  
<http://www.w3.org/2001/XMLSchema#date>

**date-time**  
<http://www.w3.org/2001/XMLSchema#dateTime>

**string**  
<http://www.w3.org/2001/XMLSchema#string>

Any literal can be expressed by representing the value as a string and appending the datatype qualifier (^) and a iri-ref which indicates the datatype. (standards.iteh.ai)

If the literal occurs inside an occurrence (3.10), the occurrence [value] property is set to the value of the literal and the occurrence [datatype] property is set to the datatype of the literal.

If the literal occurs inside a variant (3.12), the variant [value] property is set to the value of the literal and the variant [datatype] property is set to the datatype of the literal.

Variables shall occur only within templates.

42 # equivalent to "42"^^xsd:integer

"09-19"^^xsd:gMonthDay

### 3.4.2 String Escape Sequences

String escape sequences can be used to either escape characters which have special meaning in the grammar (i.e. ") or to enter characters which are not available on the keyboard by providing their Unicode codepoint.

The CTM processor shall replace binary sequences which start with the backslash character (\) with the Unicode code point equivalent: \\ becomes U+005C, \" becomes U+0022, \n becomes U+000A, \r becomes U+000D, and \t is replaced with U+0009.

Sequences which are introduced by \u and \U represent a Unicode code point as hexadecimal number, where the hexdigit part refers to the Unicode character encoding.

The hexdigit part shall be decoded into the Unicode character they represent and the \u and \U prefixes shall be removed from the byte stream.