

ISO/IEC JTC 1/SC 34

Date: 2005-01-10

ISO/IEC FCD 13250-2

ISO/IEC JTC 1/SC 34/WG 3

Secretariat: SCC

Topic Maps — Part 2: Data Model

Warning

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this document are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

Copyright notice

This ISO document is a Draft International Standard and is copyright-protected by ISO. Except as permitted under the applicable laws of the user's country, neither this ISO draft nor any extract from it may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, photocopying, recording or otherwise, without prior written permission being secured.

Requests for permission to reproduce should be addressed to 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.ch
 Web www.iso.ch

Reproduction may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

Contents

Page

Foreword.....	iv
Introduction.....	v
1 Scope.....	1
2 Normative references.....	1
3 Terms and definitions.....	1
4 The metamodel.....	4
4.1 Introduction.....	4
4.2 The fundamental types.....	4
4.3 Locators.....	5
4.4 Datatypes.....	5
4.5 Constraints.....	6
5 The data model.....	6
5.1 Source locators.....	6
5.2 The topic map item.....	6
5.3 Topic items.....	7
5.3.1 Subjects and topics.....	7
5.3.2 Identifying subjects.....	8
5.3.3 Topic characteristics.....	8
5.3.4 Scope.....	8
5.3.5 Reification.....	9
5.3.6 Properties.....	9
5.4 Topic name items.....	11
5.5 Variant items.....	12
5.6 Occurrence items.....	13
5.7 Association items.....	14
5.8 Association role items.....	15
6 Merging.....	16
6.1 General.....	16
6.2 Merging topic items.....	16
6.3 Merging topic name items.....	16
6.4 Merging variant items.....	16
6.5 Merging occurrence items.....	17

6.6	Merging association items.....	17
6.7	Merging association role items.....	17
7	Published subjects.....	18
7.1	General.....	18
7.2	The type-instance relationship.....	18
7.3	The supertype-subtype relationship.....	18
7.4	Variant name scopes.....	19
7.5	Topic characteristic types.....	19
7.6	Topic map constructs.....	20
Annex A (informative) Reference Model Application.....		21
Bibliography.....		22

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 13250-2 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. It is that structure which this part of ISO/IEC 13250 defines, in the form of a data model.

Topic Maps — Part 2: Data Model

1 Scope

NOTE This clause defines the scope of this part of ISO/IEC 13250. It should not be confused with the concept of "scope" defined in 5.3.4, which only applies in the context of topic maps.

This part of ISO/IEC 13250 specifies a data model for topic maps. It defines the abstract structure of topic maps, using the information set formalism, and to some extent their interpretation, using prose. The rules for merging in topic maps are also defined, as are some fundamental published subjects.

The purpose of the data model is to define the interpretation of the topic map interchange syntaxes, and to serve as a foundation for the definition of supporting standards for canonicalization, querying, constraints, and so on. All of these standards fall outside the scope of this part of ISO/IEC 13250, however.

NOTE This part of ISO/IEC 13250 does not have a conformance section since it is only a data model, and as such it has no boundary with the outside world in terms of which conformance can be specified.

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.

Unicode, *The Unicode Standard, Version 4.0*, The Unicode Consortium, Reading, Massachusetts, USA, Addison-Wesley, 2003, ISBN 0-321-18578-1

IETF RFC 2396, *Uniform Resource Identifiers (URI): Generic Syntax*, Internet Standards Track Specification, August 1998, available at <<http://www.ietf.org/rfc/rfc2396.txt>>

IETF RFC 2732, *Format for Literal IPv6 Addresses in URLs*, Internet Standards Track Specification, December 1999, available at <<http://www.ietf.org/rfc/rfc2732.txt>>

XML Infoset, *XML Information Set*, World Wide Web Consortium, 24 October 2001, available at <<http://www.w3.org/TR/2001/REC-xml-infoset-20011024>>

ISO10646-1, *ISO 10646-1:2000: Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*, ISO, 2000

ISO10646-2, *ISO 10646-2:2001: Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 2: Supplementary Planes*, ISO, 2001

W3C XML, *Extensible Markup Language (XML) 1.0 (Second Edition)*, W3C Recommendation, 6 October 2000, available at <<http://www.w3.org/TR/2000/REC-xml-20001006>>

W3C XML Schema, part 2, *XML Schema Part 2: Datatypes Second Edition*, W3C Recommendation, 28 October 2004, available at <<http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/>>

3 Terms and definitions

For the purposes of this part of ISO/IEC 13250, the following terms and definitions apply.

3.1 association

a representation of a relationship between one or more subjects

3.2 association role player

a topic participating in an association

3.3 association role

a representation of the involvement of a subject in a relationship represented by an association

3.4 association role type

a topic defining the nature of the participation of an association role player in an association

3.5 association type

a topic representing a subject which describes the nature of the relationship represented by the association

3.6 base name

a name or label for a subject, expressed as a string

3.7 information item

abstract representations of topic map constructs

3.8 information resource

a resource that can be represented as a sequence of bytes, and thus could potentially be retrieved over a network

3.9 instance

Any subject that belongs to a particular type

3.10 locator

a string conforming to some locator notation that references one or more information resources

3.11 merging

a process applied to topic maps in order to reduce the number of redundant topic map constructs

3.12 occurrence

a representation of relationship between a subject and an information resource

3.13 occurrence type

a topic attached to an occurrence to describe the nature of the relationship between the subject and the information resource linked by that occurrence

3.14 published subject

any subject for which there exists at least one published subject indicator

3.15 published subject identifier

the subject identifier of a published subject indicator

3.16 published subject indicator

a subject indicator that is published and maintained at an advertised location for the purposes of supporting topic map interchange and mergeability

3.17 reification

making a topic represent the subject of another topic map construct

3.18 scope

the context within which a topic characteristic assignment is valid

3.19 source locator

a locator assigned to a topic map construct in order to allow it to be referred to

3.20 statement

the assignment of a topic characteristic to a topic

3.21 subject

anything whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever. In particular, it is anything on which the creator of a topic map chooses to discourse.

3.22 subject identifier

a locator that refers to a subject indicator

3.23 subject indicator

an information resource that is referred to from a topic map in an attempt to unambiguously identify the subject of a topic to a human being. Any information resource can become a subject indicator by being referred to as such from within some topic map, whether or not it was intended by its publisher to be a subject indicator.

3.24 subject locator

a locator that refers to the information resource that *is* the subject of a topic. The topic thus represents that particular information resource; i.e. the information resource is the subject of the topic.

3.25 supertype-subtype relationship

the relationship between a more general type (the supertype) and a specialization of that type (the subtype)

3.26 topic

a symbol used within a topic map to represent some subject, about which the creator of the topic map wishes to make statements

3.27 topic characteristic

a topic name, occurrence, or association role belonging to some topic

3.28 topic characteristic assignment

a statement that a certain topic characteristic belongs to a certain topic

3.29 topic map

a set of topics and associations

3.30 topic name

a name for a topic, consisting of the base form, known as the base name, and variants of that base form, known as variant names

3.31 type

an abstraction that captures characteristics common to a set of subjects

3.32 unconstrained scope

the scope used to indicate that a topic characteristic assignment is considered to have unlimited validity

3.33 variant name

an alternative form of a topic name that may be more suitable in certain contexts than the base name of the topic map

4 The metamodel

4.1 Introduction

The metamodel used in this document is the same as that used by the XML Information Set [XML Infoset]. An instance of this data model consists of a number of *information items*, which are abstract representations of topic map constructs. Every information item is an instance of some information item type, which specifies a number of named properties which the information item must have. Throughout this part of ISO/IEC 13250 the term "information item" refers to the information item types defined in this model, while information items of particular types are referred to as "topic items", "topic name items", and so on.

The names of these properties are written in square brackets: [property name], following the convention used in [XML Infoset]. Every property has an associated type that constrains what values it may have. Properties are not allowed to have null as their value unless this is explicitly stated in the definition of the property.

Certain properties in the model are specified as computed properties, which means that they are specified in terms of how their values may be produced from other properties in the model. These properties are specified for reasons of convenience or to better reflect the conceptual model but are strictly speaking redundant.

All types defined in this part of ISO/IEC 13250 have a well-defined test of equality. This equality test is used to avoid duplicate values in properties whose values are of type set. Information items have identity, independent of their values, so items can be compared both by identity and by value. Equality throughout this part of ISO/IEC 13250 should be taken to mean equality according to the rules defined for the types of the values being compared.

UML diagrams [UML][2] are used in addition to the infoset formalism for purposes of illustration. These diagrams are purely informative, and in cases of discrepancy between the diagrams and normative prose, the prose is definitive.

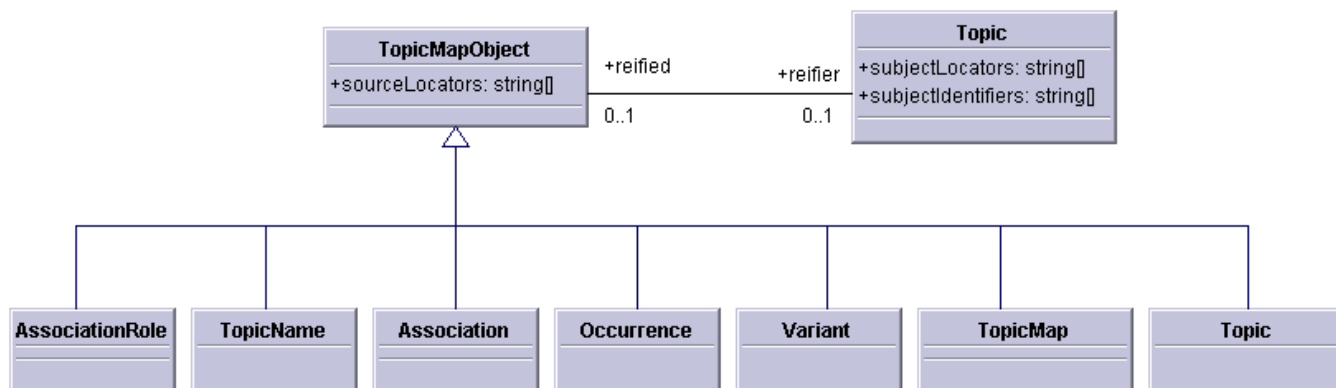


Figure 1 — The class hierarchy

NOTE TopicMapObject is the abstract superclass of all classes used in these UML diagrams. It is used here to simplify the UML diagrams using inheritance, and because the value of the [reifier] property of topic items can be any topic map construct.

4.2 The fundamental types

The values of information item properties may be either other information items, or values of one of the types defined below:

String

Strings are sequences of Unicode scalar values.

Strings are equal if they consist of the exact same sequence of Unicode scalar values. This implies that all comparisons are case-sensitive.

Set

Sets are collections of zero or more unordered elements that contain no elements that are equal to each other. In this data model, the elements of a set are always information items.

Two sets are equal unless there exists an element in one set for which no equal element can be found in the other.

Null

Null is used to indicate that properties have no value; it does not necessarily indicate that the value of the property is unknown. In this model null can never be contained in a set.

Null is distinct from all other values (including the empty set and the empty string); it is only equal to itself.

4.3 Locators

An *information resource* is a resource that can be represented as a sequence of bytes, and thus could potentially be retrieved over a network. Topic maps can refer to information resources external to themselves in order to make statements about them. These information resources are *not* part of the topic map; they are only referenced from it.

A *locator* is a string conforming to some locator notation that references one or more information resources. Locators are always expressed in some locator notation, which is a definition of the formal syntax and interpretation of a class of locators. The definition of locator notations is outside the scope of this part of ISO/IEC 13250. All locators in this model use the notation defined by [IETF RFC 2396].

NOTE This part of ISO/IEC 13250 does not require normalization to be applied to the syntactical expressions of locators in order to detect that syntactically different but logically equivalent locators are in fact equivalent. The application of such logic is encouraged, however. As the it cannot be guaranteed that normalization will be performed reliance on normalization is strongly discouraged.

4.4 Datatypes

The only atomic fundamental types defined in this part of ISO/IEC 13250 are strings and null, but through the concept of datatypes data of any type can be represented in this model. All datatypes used must have a string representation of their value space, and this string representation is what is stored in the topic map. The information about which datatype the value belongs to is stored separately, in the form of a locator identifying the datatype.

This part of ISO/IEC 13250 defines only the following three datatypes, but other datatypes may also be used. These datatypes are all defined by [W3C XML Schema, part 2], and the syntax of the XML datatype is defined by [W3C XML].

String

This is the string datatype, as defined in 4.2. The identifier of this datatype is <http://www.w3.org/2001/XMLSchema#string>.

URI

This is the datatype of locators using the URI notation; the URIs must be absolute. The identifier of this datatype is <http://www.w3.org/2001/XMLSchema#anyURI>.

XML

This is the XML datatype, which represents XML document fragments. The identifier of this datatype is

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

NOTE The datatype of a string value may change its interpretation. For example, the string value "AT&T" means precisely what it says if the datatype is string, but means "AT&T" if the datatype is XML.

4.5 Constraints

The model defined in this part of ISO/IEC 13250 contains not only fundamental types and information item types with named properties, but also constraints on the allowed instances of the model. The purpose of these constraints is to prevent inconsistencies in instances of the data model.

5 The data model

5.1 Source locators

A *source locator* is a locator assigned to a topic map construct in order to allow it to be referred to. It is not specified how and when source locators are assigned to information items, and source locators may be freely assigned to information items.

NOTE In a sense source locators are identifiers for topic map constructs devoid of any specified semantics, and these may be automatically assigned to information items to provide them with an identifier or to identify the origin of the item.

When an instance of the data model is created through deserialization from some topic map syntax, source locators are created that point back to the syntactical constructs that gave rise to the information items in the data model instance. In these cases the source locators will point to the minimal syntactical construct of origin, which means that for topic items created from the XTM syntax, for example, the source locator will point to the originating topic element, rather than the containing topicMap element.

Topic map constructs may have any number source locators since when duplicate information items are merged the resulting information item inherits all the source locators of the original information items.

Constraint: Duplicate source locators

It is an error for two different information items to have string that are equal in their [source locators] properties, unless they are topic items. If they are topic items they must be merged according to the procedure in 6.2.

5.2 The topic map item

A *topic map* is a set of topics and associations. Its purpose is to convey information about subjects through the assignment of characteristics to topics representing those subjects. The topic map itself has no meaning or significance beyond its use as a container for the information about those subjects; in particular, the topic map does not represent anything but itself.

NOTE However, while the topic map does not represent anything it may be reified in order to make statements about the topic map (that is, the collection of topics and associations) as a whole. These statements may for example provide traditional metadata such as author, version, copyright, or they may reference system metadata such as a schema for the topic map, external documentation of it, and so on.

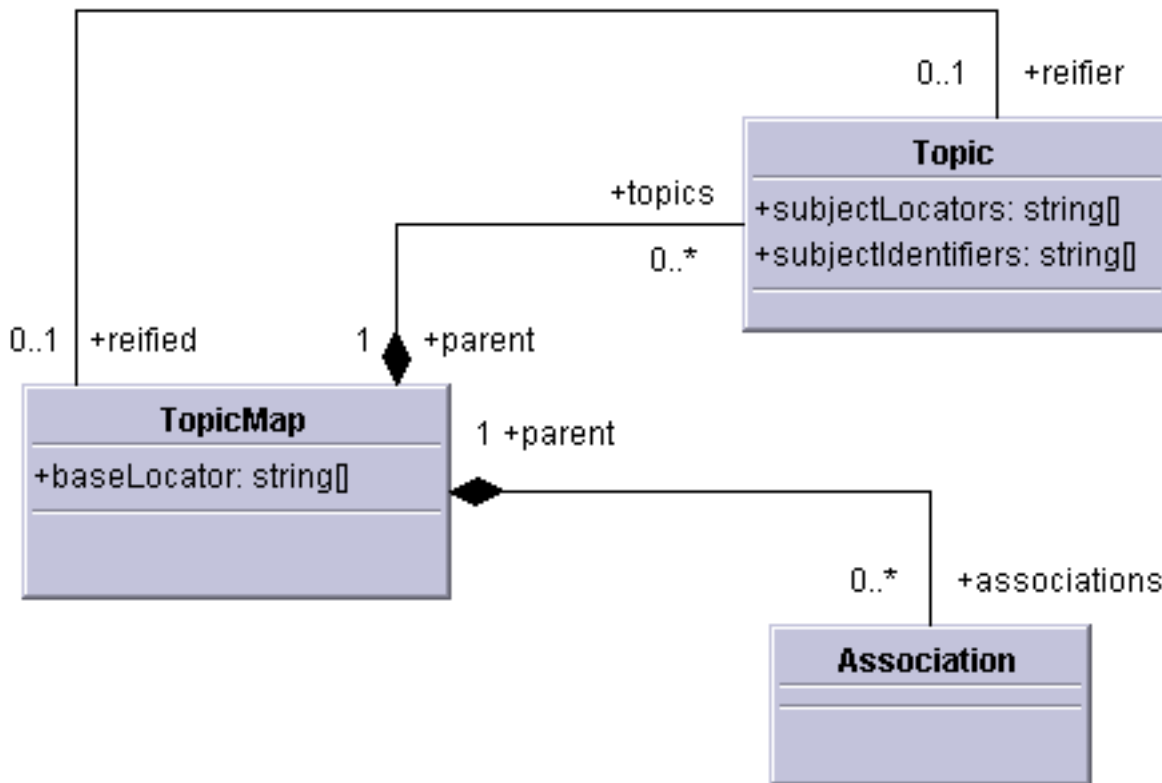


Figure 2 — The topic map item

The topic map item represents the topic map. Topic map items have the following properties:

- 1 **[topics]**: A set of topic items. All the topics in the topic map.
- 2 **[associations]**: A set of association items. All the associations in the topic map.
- 3 **[reifier]**: A topic item, or null. If present, the topic that reifies this topic map construct.
Computed value: if there exists a topic item in whose [subject identifiers] property can be found a string equal to one in the [source locators] property of this information item that topic item is the value of the [reifier] property. If not, its value is null.
- 4 **[base locator]**: A string, or null. A reference to the location where the topic map is stored.
- 5 **[source locators]**: A set of strings. The source locators of the topic map.

5.3 Topic items

5.3.1 Subjects and topics

A *subject* can be anything whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever. In particular, it is anything on which the creator of a topic map chooses to discourse.

EXAMPLE Examples of subjects for which topics may be created are:

- The moon.
- The Soviet Union. This subject no longer exists as an organizational unit, but the idea still exists, and so is still a subject.
- The letters 'A', 'B', 'C', and 'D'. This is a single subject, a set with four elements.

- Plato's notion of the good. This subject is different from, but related to, "good" in the abstract, and John Stuart Mill's notion of "good".

A *topic* is a symbol used within a topic map to represent some subject, about which the creator of the topic map wishes to make statements. Topics are used to represent subjects in order to allow statements to be made about the subjects through the assignment of characteristics to the topics that represent them. A *statement* is the assignment of a topic characteristic to a topic.

Every topic represents one, and only one, subject. The process of merging ensures that whenever two topics are known to represent the same subject they are merged. It may well be, however, that two topics may represent the same subject without this being detectable by the rules of this part of ISO/IEC 13250. Merging beyond the minimal merging required by the rules of Clause 6 is freely allowed. Most commonly this will be done by inferring the subject of the topics from their characteristics.

5.3.2 Identifying subjects

Formal identification of subjects with locators allows topic maps to be merged safely and precisely, and also allows the definition of subjects with semantics that can be implemented in topic map systems. Examples of such subjects can be found in Clause 7.

A *subject indicator* is an information resource that is referred to from a topic map in an attempt to unambiguously identify the subject of a topic to a human being. Any information resource can become a subject indicator by being referred to as such from within some topic map, whether or not it was intended by its publisher to be a subject indicator.

A *subject identifier* is a locator that refers to a subject indicator. Topic maps contain only subject identifiers (and not the corresponding subject indicators), and consequently it is the subject identifier that is the basis for merging; the subject indicator is ignored during merging.

A *subject locator* is a locator that refers to the information resource that *is* the subject of a topic. The topic thus represents that particular information resource; i.e. the information resource is the subject of the topic.

NOTE If a topic has multiple subject locators these must all refer to the same information resource. This of course raises the question of when two information resources can be considered to be the same. This part of ISO/IEC 13250 makes no attempt to clarify this and leaves it for individual locator notations to define.

EXAMPLE Consider the URI <http://www.topicmaps.org>. If given as the subject locator of topic A this would mean that topic A represents the information resource identified by this URI. However, using it as the subject identifier of topic B would mean that B represents what is described in that information resource. At the time of writing this would seem to be the organization known as TopicMaps.Org. (Note: the *organization*; the real-world institution known by that name. This is different from the subject of A, which is the web page itself.)

Note the uncertainty in the last sentence above. The information resource in question is a subject indicator for topic B, but it was not written to be a subject indicator (that is, it is not a *published* subject indicator), and so is not entirely unambiguous with respect to what subject it indicates. Nor is it guaranteed to be stable, so at the time of reading it may indicate some other subject, or it may no longer exist.

5.3.3 Topic characteristics

A *topic characteristic* is a topic name, occurrence, or association role belonging to some topic. A *topic characteristic assignment* is a statement that a certain topic characteristic belongs to a certain topic. In the data model this is represented by the inclusion of an information item representing a topic characteristic in the value of a property of a topic item. Any topic characteristic assignment constitutes a statement about the subject represented by the topic.

The properties of topic items that do not represent topic characteristics are not statements about the subject; they are statements about the topic. As such they are part of the topic map machinery, rather than statements about the subject represented in the topic map.

5.3.4 Scope

All topic characteristic assignments have a scope. The *scope* represents the context within which a topic characteristic assignment is valid. Outside the context represented by the scope the assignment is not known to be

valid. Formally, a scope is composed of a set of topics that together define the context. That is, the topic characteristic is known to be valid only in contexts where *all* the subjects in the scope apply.

NOTE This is the first topic map standard to define scope as being "all subjects", hence older topic map data may not conform to this. Further, it is possible that not all users will adhere strictly to this interpretation of scope.

The *unconstrained scope* is the scope used to indicate that a topic characteristic assignment is considered to have unlimited validity. In the model this is represented by the empty set.

Precisely how a subject, or a set of subjects, defines a context is not defined by this part of ISO/IEC 13250, but left for those creating topic maps to define as part of the definition of their subjects.

EXAMPLE Examples of the use of scope are given below:

- "Suomi" is the name of the country Finland in Finnish. This corresponds to assigning the topic name "Suomi" to a topic representing Finland, and giving it as scope a topic representing Finnish.
- According to Norman Davies World War II started on June 6, 1937 ([Davies][1]). This corresponds to creating a topic representing WWII, and assigning to it the string "June 6, 1937" as an occurrence of type "start date", and giving this occurrence as scope a topic representing the person Norman Davies.
- According to Peter T. Daniels, the Devanagari script is an instance of the script type "abugida," whereas according to William Bright it is an "alphasyllabary". This corresponds to having two "type-instance" associations, each scoped with a topic representing the relevant authority.

5.3.5 Reification

The act of *reification* is the act of making a topic represent the subject of another topic map construct. For example, creating a topic that represents the relationship an association represents is reification.

NOTE 1 Note that the use of the term 'reification' in this part of ISO/IEC 13250 is not to be confused with its use in philosophy.

In many cases it is desirable to be able to attach additional information to topic map constructs such as topic names or associations. One may want to give an association occurrences, or to give an occurrence a name. The basic topic map model does not allow this, but through reification this can be done by creating a topic that reifies the topic map construct. The necessary information can then be attached to the reifying topic, and the reification relationship is present in a structured form that can reliably be detected by implementations.

Reification is achieved by giving the reifying topic a subject identifier that refers to the topic map construct that is being reified. In model terms, this means that if an information item has a value in its [source locators] property that is equal to one of the values in the [subject identifiers] property of a topic, that topic item reifies the information item.

NOTE 2 One topic cannot reify another. A topic reifying a topic map construct in reality represents the real-world thing that that topic map construct represents. A topic reifying an association really represents the relationship represented by that association, and so if one topic were to reify another that would mean that the topic represents the subject of the other, and so the two would have to merge, since they would have the same subject.

5.3.6 Properties

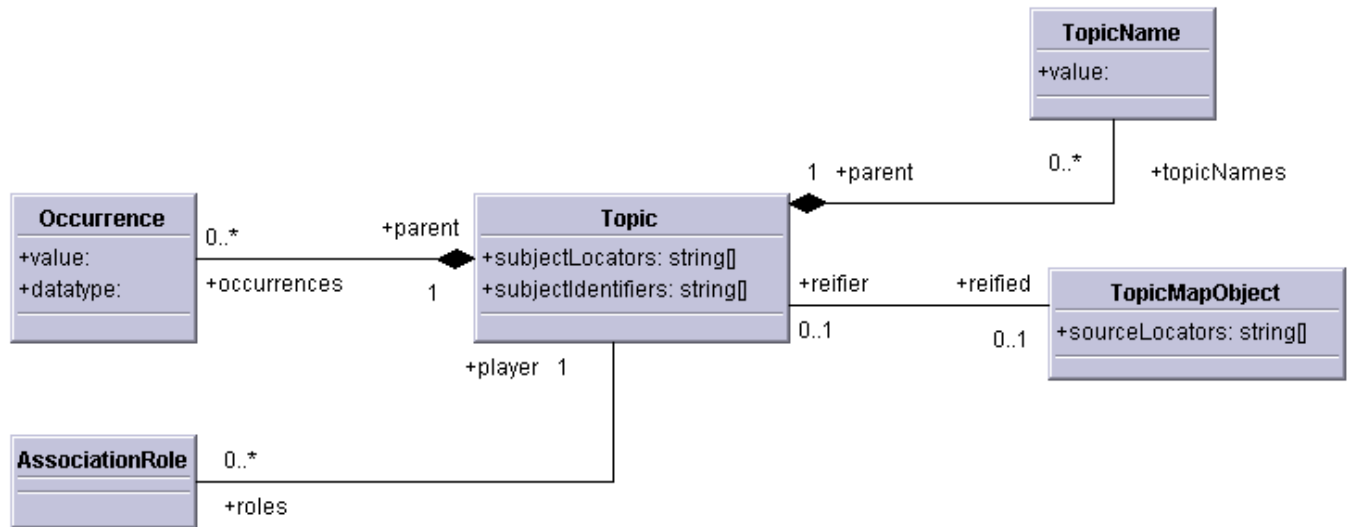


Figure 3 — The topic item

Topic items represent topics. Topic items have the following properties:

- 1 **[topic names]:** A set of topic name items. This is the set of topic names assigned to this topic.
- 2 **[occurrences]:** A set of occurrence items. This is the set of occurrences assigned to this topic.
- 3 **[roles played]:** A set of association role items. This is the set of association roles played by this topic.
Computed value: the set of all association role items whose [player] property value is this topic item.
- 4 **[subject identifiers]:** A set of strings. The locators referring to the subject indicators of this topic.
- 5 **[subject locators]:** A set of strings. The locators referring to the information resource that is the subject of this topic.
- 6 **[reified]:** An information item, or null. If given, the topic map construct that is the subject of this topic.
Computed value: if any information item contains in its [source locators] property a string equal to one in the [subject identifiers] property of this topic item, that information item is the value of the [reified] property. If no such information item is found the value is null.
- 7 **[source locators]:** A set of strings. The source locators of the topic.
- 8 **[parent]:** An information item. The topic map containing the topic.
Computed value: the topic map item whose [topics] property contains this topic item.

Equality rule: Two topic items are equal if they have:

- at least one equal string in their [subject identifiers] properties,
- at least one equal string in their [source locators] properties,
- at least one equal string in their [subject locators] properties,
- an equal string in the [subject identifiers] property of the one topic item and the [source locators] property of the other, or
- the same information item in their [reified] properties.

Constraint: Topic identity required

All topic items must have a value for at least one of the [subject identifiers], [subject locators], and [source locators] properties that is neither the empty set nor null.

NOTE Locators which refer directly to subjects which are not information resources must be used with caution. They should not be used in the [subject locators] property, as this is intended only for references to information resources. Rather, they should be placed in the [subject identifiers] property.

The isbn URN scheme used to identify books ([IETF RFC 2288][3]), for example, does not reference information resources, and so should not be put in the [subject locators] property, but instead in the [subject identifiers] property.

5.4 Topic name items

A *topic name* is a name for a topic, consisting of the base form, known as the base name, and variants of that base form, known as variant names. It is the topic name which is a topic characteristic; the base name and variant names are only parts of the topic name characteristic.

Topic names may have a type, which defines what kind of name the topic name represents. They always have a scope, which defines in what contexts the topic name is an appropriate label for the subject. A topic may have any number of topic names.

A *base name* is a name or label for a subject, expressed as a string. That is, it is something that identifies the subject (though not necessarily uniquely) and can be used as a label for the subject in user interfaces. The notion of a base name corresponds closely to the common sense notion of a name.

NOTE Suitable base names for people, countries, and organizations are their names, while base names for documents, musical works, and movies might be their titles. Base names may have variant names, which are alternative forms of the base name that may be more appropriate in specific contexts. Essentially, a base name is a specialized kind of occurrence.

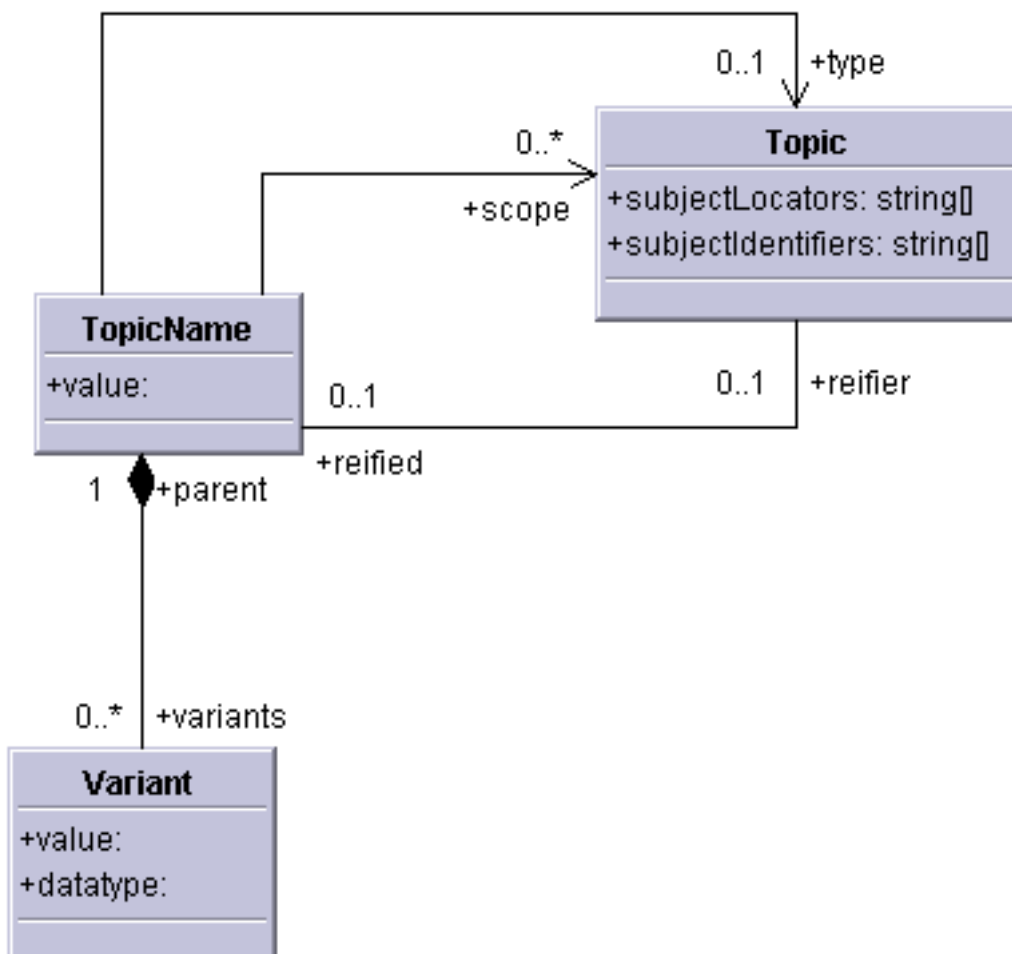


Figure 4 — The topic name item

Topic name items represent topic names. Topic name items have the following properties:

- 1 **[value]**: A string. The base name; the base form of the topic name.
- 2 **[type]**: A topic item, or null. If given, the topic defining what kind of name is represented by this topic name.
- 3 **[scope]**: A set of topic items. The scope that represents the validity of this topic name as a label for the topic.
- 4 **[variants]**: A set of variant name items. The variant names that are alternative forms of the topic name.
- 5 **[reifier]**: A topic item, or null. If present, the topic that reifies this topic map construct.

Computed value: if there exists a topic item in whose [subject identifiers] property can be found a string equal to one in the [source locators] property of this information item that topic item is the value of the [reifier] property. If not, its value is null.

- 6 **[source locators]**: A set of strings. The source locators of this topic name.
- 7 **[parent]**: An information item. The topic containing the topic name.

Computed value: the topic name item whose [topic names] property contains this topic name item.

Equality rule: Topic name items are equal if the values of their [value], [type], [scope], and [parent] properties are equal.

5.5 Variant items

A *variant name* is an alternative form of a topic name that may be more suitable in certain contexts than the base name of the topic map. The scope of the variant name is the only basis for establishing what variant name is most suitable in any given situation. A variant name may be a string, but it may also be any other kind of information resource.

NOTE When choosing a label for a topic, the topic name considered most appropriate should be chosen, and then the form of that topic name best suited for display in that particular context should be chosen, which may be the base name or one of its variants. This part of ISO/IEC 13250 does not constrain the process by which this is done.

7.4 defines some published subjects that may be useful for scoping variant names.

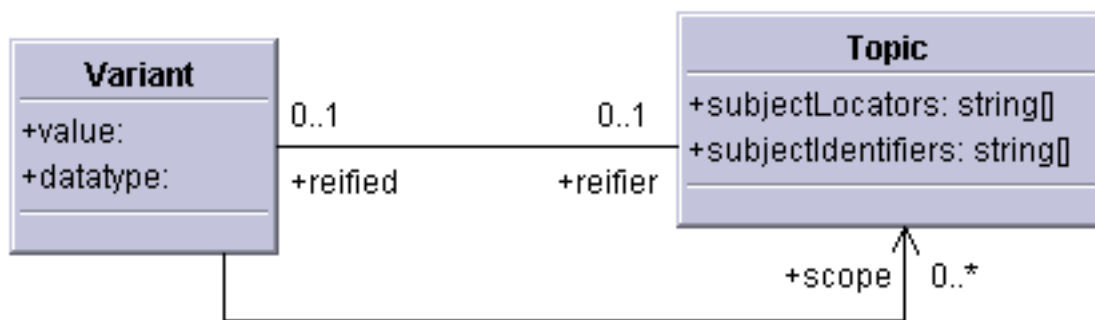


Figure 5 — The variant name item

Variant items represent variant names. Variant items have the following properties:

- 1 **[value]**: A string. If the datatype is URI, a locator referring to the information resource that is the variant name; otherwise the string is the variant name.
- 2 **[datatype]**: A string. A locator identifying the datatype of the variant name value.
- 3 **[scope]**: A non-empty set of topic items. The scope that describes in what context(s) the variant name may be preferred as a label for the topic.
- 4 **[reifier]**: A topic item, or null. If present, the topic that reifies this topic map construct.

Computed value: if there exists a topic item in whose [subject identifiers] property can be found a string equal to one in the [source locators] property of this information item that topic item is the value of the [reifier] property. If not, its value is null.

- 5 **[source locators]**: A set of strings. The source locators of the variant name.
- 6 **[parent]**: An information item. The topic name containing the variant.

Computed value: the topic name item whose [variants] property contains this variant item.

Equality rule: Variant items are equal if the values of their [value], [datatype], [scope], and [parent] properties are equal.

Constraint: Variant scope

The value of the [scope] property of each variant item must be a true superset of the value of the [scope] property of the topic name item in its [parent] property.

5.6 Occurrence items

An *occurrence* is a representation of relationship between a subject and an information resource. The subject in question is that represented by the topic which contains the occurrence. The information resource may either be a value inside the topic map or an external information resource. Occurrences are essentially a specialized kind of association, where one participant in the association must be an information resource. An *occurrence type* is a topic attached to an occurrence to describe the nature of the relationship between the subject and the information resource linked by that occurrence.

All occurrences have a scope, which defines the contexts in which the occurrence relationship between the information resource and the subject is valid.

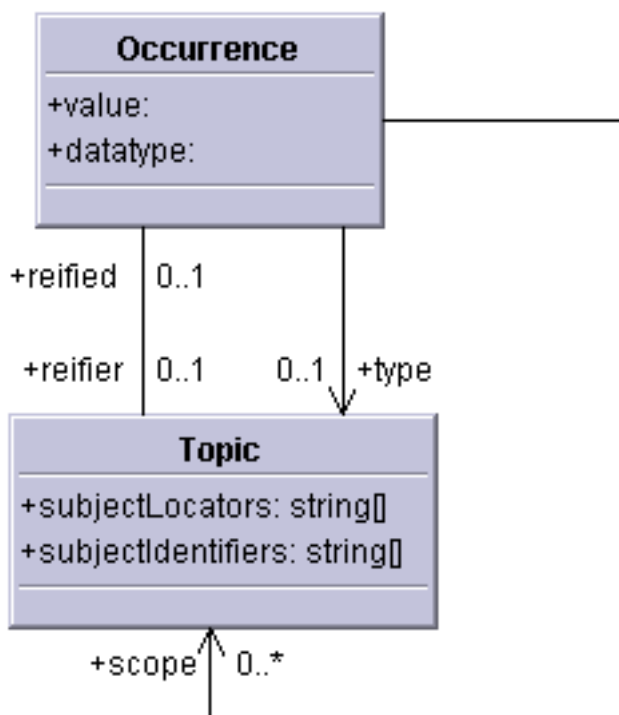


Figure 6 — The occurrence item

Occurrence items represent occurrences. Occurrence items have the following properties:

- 1 **[value]**: A string. If the datatype is URI, a locator referring to the information resource the occurrence connects with the subject; otherwise the string is the information resource.
- 2 **[datatype]**: A string. A locator identifying the datatype of the occurrence value.
- 3 **[scope]**: A set of topic items. The scope that describes in what context the occurrence relationship may be considered valid.

4 **[type]**: A topic item, or null. The topic that defines the nature of the occurrence relationship.

5 **[reifier]**: A topic item, or null. If present, the topic that reifies this topic map construct.

Computed value: if there exists a topic item in whose [subject identifiers] property can be found a string equal to one in the [source locators] property of this information item that topic item is the value of the [reifier] property. If not, its value is null.

6 **[source locators]**: A set of strings. The source locators of the occurrence.

7 **[parent]**: An information item. The topic containing the occurrence.

Computed value: the topic item whose [occurrences] property contains this occurrence item.

Equality rule: Occurrence items are equal if the values of their [value], [datatype], [scope], [type], and [parent] properties are equal.

5.7 Association items

An *association* is a representation of a relationship between one or more subjects. Associations have an *association type*, a topic representing a subject which describes the nature of the relationship represented by the association.

An *association role* is a representation of the involvement of a subject in a relationship represented by an association. An association role connects two pieces of information within an association: the *association role player*, that is, a topic participating in an association, and the *association role type*, that is, a topic defining the nature of the participation of an association role player in an association.

EXAMPLE An example of an association might be the 'authorship' relationship between Henrik Ibsen and the play 'Peer Gynt'. In this relationship there are two roles: Ibsen plays the role of 'author', while 'Peer Gynt' plays the role of 'work'.

Another example might be the 'parenthood' relationship between Hamlet, King Hamlet, and Queen Gertrude. This relationship has three roles: Hamlet plays the role of 'child', the King that of 'father', and the Queen that of 'mother'.

All associations have a scope, which defines the context in which the relationship represented by the association can be considered valid. The scope applies to the assignment of the roles to the topics playing them in the same way as it does to the association as a whole.

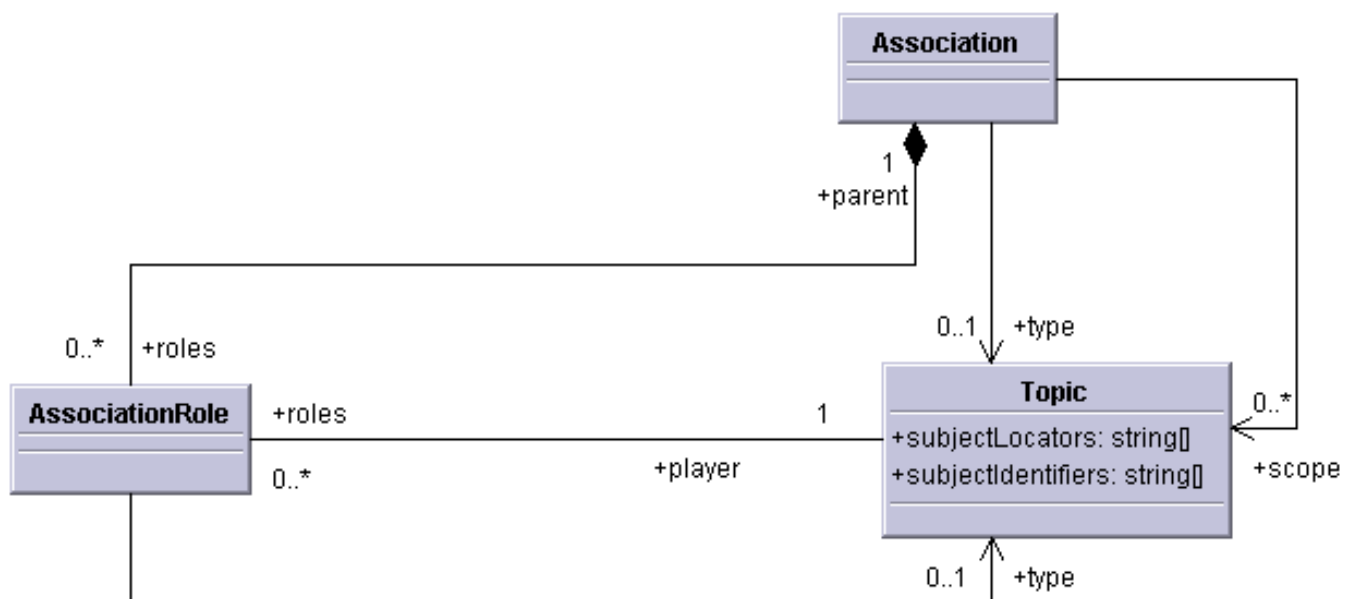


Figure 7 — The association item

Association items represent associations. Association items have the following properties:

- 1 **[type]**: A topic item, or null. If given, the topic that defines the nature of the relationship represented by the association.
- 2 **[scope]**: A set of topic items. The scope that describes in what context the association may be considered valid.
- 3 **[roles]**: A non-empty set of association role items. The association roles for all the topics that participate in this relationship.
- 4 **[reifier]**: A topic item, or null. If present, the topic that reifies this topic map construct.

Computed value: if there exists a topic item in whose [subject identifiers] property can be found a string equal to one in the [source locators] property of this information item that topic item is the value of the [reifier] property. If not, its value is null.

- 5 **[source locators]**: A set of strings. The source locators of the association.
- 6 **[parent]**: An information item. The topic map containing the association.

Computed value: the topic map item whose [associations] property contains this association item.

Equality rule: Association items are equal if the values of their [scope], [type], and [roles] properties are equal.

5.8 Association role items

NOTE See 5.7 for the definition of the term 'association role'.

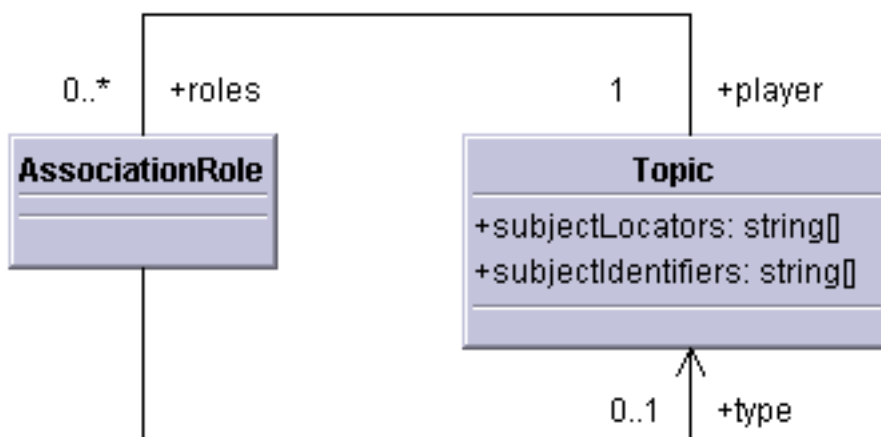


Figure 8 — The association role item

Association role items represent association roles. Association role items have the following properties:

- 1 **[player]**: A topic item. The topic that plays this role in the association.
- 2 **[type]**: A topic item, or null. If given, the topic that represents the nature of the involvement of the association role player in the association.
- 3 **[reifier]**: A topic item, or null. If present, the topic that reifies this topic map construct.

Computed value: if there exists a topic item in whose [subject identifiers] property can be found a string equal to one in the [source locators] property of this information item that topic item is the value of the [reifier] property. If not, its value is null.

- 4 **[source locators]**: A set of strings. The source locators of this association role.
- 5 **[parent]**: An information item. The association containing the association role.

Computed value: the association item whose [roles] property contains this association role item.

Equality rule: Association role items are equal if the values of their [type], [player], and [parent] properties are

equal.

6 Merging

6.1 General

A central operation in topic maps is that of *merging*, a process applied to topic maps in order to reduce the number of redundant topic map constructs. This clause defines the rules that define in which situations merging must occur, but these rules are insufficient to ensure that all redundant information is removed from a topic map.

Any change to a topic map that causes a set to contain two information items equal to each other must be followed by the merging of those two information items according to the rules given below for the type of information item to which the two equal information items belong.

6.2 Merging topic items

The procedure for merging two topic items A and B (whose [parent] properties must contain the same topic map item) is given below.

1. Create a new topic item C.
2. Replace A by C wherever it appears in one of the following properties of some information item: [topics], [scope], [type], and [player].
3. Repeat for B.
4. Set C's [source locators] property to the union of the values of A and B's [source locators] properties.
5. Set C's [subject identifiers] property to the union of the values of A and B's [subject identifiers] properties.
6. Set C's [subject locators] property to the union of the values of A and B's [subject locators] properties.
7. Set C's [topic names] property to the union of the values of A and B's [topic names] properties.
8. Set C's [occurrences] property to the union of the values of A and B's [occurrences] properties.

6.3 Merging topic name items

The procedure for merging two topic name items A and B is given below.

1. Create a new topic name item C.
2. Set C's [source locators] property to the union of the value of the [source locators] properties of A and B.
3. Set C's [value] property to the value of the [value] property of A. B's value is equal that of A, and need therefore not be taken into account.
4. Set C's [scope] property to the value of the [scope] property of A. B's value is equal that of A, and need therefore not be taken into account.
5. Set C's [variants] property to the union of the [variants] properties of A and B.
6. Remove A and B from the [topic names] property of the topic item in their [parent] properties, and add C.

6.4 Merging variant items

The procedure for merging two variant items A and B is given below.

1. Create a new variant item, C.
2. Set C's [source locators] property to the union of the values of A's and B's [source locators] properties.
3. Set C's [value] property to the value of A's [value] property. B's value is equal to that of A, and need therefore not be taken into account.
4. Set C's [datatype] property to the value of A's [datatype] property. B's value is equal to that of A, and need therefore not be taken into account.
5. Set C's [scope] property to the value of A's [scope] property. B's value is equal to that of A, and need therefore not be taken into account.
6. Remove A and B from the [variants] property of the topic name item in their [parent] properties, and add C.

6.5 Merging occurrence items

The procedure for merging two occurrence items A and B is given below.

1. Create a new occurrence item, C.
2. Set C's [source locators] property to the union of the values of A's and B's [source locators] properties.
3. Set C's [value] property to the value of A's [value] property. B's value is equal to that of A, and need therefore not be taken into account.
4. Set C's [datatype] property to the value of A's [datatype] property. B's value is equal to that of A, and need therefore not be taken into account.
5. Set C's [scope] property to the value of A's [scope] property. B's value is equal to that of A, and need therefore not be taken into account.
6. Set C's [type] property to the value of A's [type] property. B's value is equal to that of A, and need therefore not be taken into account.
7. Remove A and B from the [occurrences] property of the topic item in their [parent] properties, and add C.

6.6 Merging association items

The procedure for merging two association items A and B is given below.

1. Create a new association item, C.
2. Set C's [source locators] property to the union of the values of A's and B's [source locators] properties.
3. Set C's [scope] property to the value of A's [scope] property. B's value is equal to that of A, and need therefore not be taken into account.
4. Set C's [type] property to the value of A's [type] property. B's value is equal to that of A, and need therefore not be taken into account.
5. Set C's [roles] property to the value of A's [roles] property. B's value is equal to that of A, and need therefore not be taken into account.
6. Remove A and B from the [associations] property of the topic map item in their [parent] properties, and add C.

6.7 Merging association role items

The procedure for merging two association role items A and B is given below.

1. Create a new association role item, C.
2. Set C's [source locators] property to the union of the values of A's and B's [source locators] properties.
3. Set C's [type] property to the value of A's [type] property. B's value is equal to that of A, and need therefore not be taken into account.
4. Set C's [player] property to the value of A's [player] property. B's value is equal to that of A, and need therefore not be taken into account.
5. Remove A and B from the [roles] property of the association item in their [parent] properties, and add C.

7 Published subjects

7.1 General

A *published subject indicator* is a subject indicator that is published and maintained at an advertised location for the purposes of supporting topic map interchange and mergeability. A *published subject* is any subject for which there exists at least one published subject indicator. A *published subject identifier* is the subject identifier of a published subject indicator.

This clause defines a number of published subjects for core subjects in order to achieve interoperability through consistent behaviour. These subjects are central to this part of ISO/IEC 13250, yet there is no requirement that they be used, and alternative subjects for this functionality may be used instead.

All published subjects defined by this part of ISO/IEC 13250 are distinct.

7.2 The type-instance relationship

A *type* is an abstraction that captures characteristics common to a set of subjects. Any subject that belongs to a particular type is known as an *instance* of that type. A type may itself be an instance of another type, and there is no limit to the number of types a subject may be an instance of.

The type-instance relationship is not transitive. That is, if B is an instance of the type A, and B is a type of which C is an instance, it does not follow that C is an instance of A.

The type-instance relationship between two topics can be asserted using an association item that conforms to the following rules:

- The [type] property must be set to a topic item that has in its [subject identifiers] property the string "http://psi.topicmaps.org/sam/1.0/#type-instance".
- The [roles] property must contain exactly two association role items.
- One of the association items in the [roles] property must have its [type] property set to a topic item whose [subject identifiers] property is set to the string "http://psi.topicmaps.org/sam/1.0/#type". The [role player] property will then contain the topic item representing the type.
- One of the association items in the [roles] property must have its [type] property set to a topic item whose [subject identifiers] property is set to the string "http://psi.topicmaps.org/sam/1.0/#instance". The [role player] property will then contain the topic item representing the instance.

Association items that use one or more of the published subject identifiers defined in this clause, but which do not conform to these structural rules, are not considered to represent type-instance relationships.

Scope applies to this association type in just the same way as it does to any other.

7.3 The supertype-subtype relationship

The *supertype-subtype relationship* is the relationship between a more general type (the supertype) and a specialization of that type (the subtype). If B is the subtype of A, it follows that every instance of B is also an instance of A. The converse is not necessarily true. A type may have any number of subtypes and supertypes.

The supertype-subtype relationship is transitive, which means that if B is a subtype of A, and C a subtype of B, C is also a subtype of A.

NOTE Loops in this relationship are allowed, and should be interpreted to mean that the sets of instances for all types in the loop are the same. This does not, however, necessarily imply that the types are the same.

The supertype-subtype relationship between two types can be asserted using an association item that conforms to the following rules:

- The [type] property must be set to a topic item that has in its [subject identifiers] property the string "http://psi.topicmaps.org/sam/1.0/#supertype-subtype".
- The [roles] property must contain exactly two association role items.
- One of the association items in the [roles] property must have its [type] property set to a topic item whose [subject identifiers] property is set to the string "http://psi.topicmaps.org/sam/1.0/#supertype". The [player] property will then contain the topic item representing the supertype.
- One of the association items in the [roles] property must have its [type] property set to a topic item whose [subject identifiers] property is set to the string "http://psi.topicmaps.org/sam/1.0/#subtype". The [player] property will then contain the topic item representing the subtype.

Association items that use one or more of the published subject identifiers defined in this clause, but which do not conform to these structural rules, are not considered to represent supertype-subtype relationships.

Scope applies to this association type in just the same way as it does to any other.

EXAMPLE This means that if 'a' is an instance of 'b' in scope 'Y' and 'X', and 'b' is a subtype of 'c' in scope 'Y' and 'Z' 'a' is an instance of 'c' *only* in the context where all three topics 'X', 'Y', and 'Z' apply.

7.4 Variant name scopes

Sort names are a form of topic names used to sort topics. Sort names will be sorted in Unicode code point order. To get the desired sort order sort names must be used that, when sorted with this algorithm, give that order.

Sort names can be represented by variant items whose [scope] property contains a topic item whose [subject identifiers] property contains the string "http://psi.topicmaps.org/sam/1.0/#sort".

Display names are a form of topic names intended to be used as the preferred label for the topic.

Display names can be represented by variant items whose [scope] property contains a topic item whose [subject identifiers] property contains the string "http://psi.topicmaps.org/sam/1.0/#display".

7.5 Topic characteristic types

A unique topic characteristic is one where the value of the characteristic effectively identifies the subject of the topic, which means that if two different topics have the same value for a unique characteristic they also represent the same subject, and so must be merged.

A topic name, occurrence, or association item represents a unique topic characteristic if its [type] property contains a topic item which plays the role of instance in a type-instance association conforming to the structural rule in 7.2 where a topic item plays the role of type, and has the string "http://psi.topicmaps.org/sam/1.0/#unique-characteristic" in its [subject identifiers] property.

Topics having unique topic characteristics are to be merged according to the procedure in 6.2 if:

- each has a topic name item representing a unique topic characteristic in its [topic names] property where two

such topic name items have equal values in their [value], [type], and [scope] properties, but not in their [parent] properties,

- each has an occurrence item representing a unique topic characteristic in its [occurrences] property where two such occurrence items have equal values in their [value], [type], [datatype], and [scope] properties, but not in their [parent] properties,
- each has an association role item whose [parent] property contains an association representing a unique topic characteristic where two such association items have equal values in their [type] and [scope] properties and where all association role items in their [roles] properties are equal, except for the two association role items that triggered the merging, which only need to be equal in their [type] properties.

7.6 Topic map constructs

This clause describes published subjects for the main topic map constructs, useful as types for topics reifying topic map constructs of various kinds.

The subject identifier <http://psi.topicmaps.org/sam/1.0/#topic-name> identifies the type of topic names, as described in 5.4.

The subject identifier <http://psi.topicmaps.org/sam/1.0/#variant> identifies the type of variant names, as described in 5.5.

The subject identifier <http://psi.topicmaps.org/sam/1.0/#occurrence> identifies the type of occurrences, as described in 5.6.

The subject identifier <http://psi.topicmaps.org/sam/1.0/#association> identifies the type of associations, as described in 5.7.

The subject identifier <http://psi.topicmaps.org/sam/1.0/#association-role> identifies the type of association roles, as described in 5.8.

Annex A (informative)

Reference Model Application

This annex discloses the data model defined in this part of ISO/IEC 13250 as an application of [ISO 13250-5][4]. The name of the application is "the Topic Maps Data Model".

Every information item in an instance of this data model is a "subject proxy" in [ISO 13250-5][4] terms, and every property of such information items a "property instance". The properties defined in this part of ISO/IEC 13250 are all "OP property classes" in [ISO 13250-5][4] terms. Constraints on the values of these property classes are given above, as are the property instance merging rules, as are the conferral rules for computed properties.

The application defined in this annex has two additional properties classes, and every subject proxy has one instance of each:

itemtype

Instances of this property class contain the name of the information item type of the information item as a string. For the topic map item this will be "topic map", for topic items "topic", for topic name items "topic name", and so on.

tmdm-sip

This property class is a SIP class, and contains a set of (property class name, property instance value) pairs for every property class used by the equality rule of the information item type that the information item is an instance of.

The subject sameness detection rule used by this SIP class is a simple equality test, using the general set equality rule defined above.

Bibliography

- [1] *Europe: A History*, Norman Davies, Oxford University Press, 1996, ISBN 0-19-820171-0
- [2] *Unified Modeling Language (UML)*, Version 1.5, Object Management Group, available at <<http://www.omg.org/technology/documents/formal/uml.htm>>
- [3] *Using Existing Bibliographic Identifiers as Uniform Resource Names*, Informational Memo, February 1998, available at <<http://www.ietf.org/rfc/rfc2288.txt>>
- [4] *ISO 13250-5: Topic Maps — Reference Model*, ISO, 2004-11-07, available at <<http://www.isotopicmaps.org/TMMM/TMMM-latest-clean.html>>