

DRAFT INTERNATIONAL STANDARD

ISO/IEC DIS 19757-3

ISO/IEC JTC 1/SC 34

Secretariat: JISC

Voting begins on:
2014-12-01

Voting terminates on:
2015-03-01

Information technology — Document Schema Definition Languages (DSDL) —

Part 3: Rule-based validation — Schematron

*Technologies de l'information — Langages de définition de schéma de documents (DSDL) —
Partie 3: Validation de règles orientées — Schematron*

ICS: 35.240.30

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
<https://standards.iteh.ai/catalog/standards/sist/e166298b-9fda-46a3-92fb-03c579134a0f/iso-iec-19757-3-2016>

THIS DOCUMENT IS A DRAFT CIRCULATED FOR COMMENT AND APPROVAL. IT IS THEREFORE SUBJECT TO CHANGE AND MAY NOT BE REFERRED TO AS AN INTERNATIONAL STANDARD UNTIL PUBLISHED AS SUCH.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.



Reference number
ISO/IEC DIS 19757-3:2014(E)

© ISO/IEC 2014

ITeH STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
<https://standards.iteh.ai/catalog/standards/sist/e166298b-9fda-46a3-92fb-03c579134a0f/iso-iec-19757-3-2016>

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.org
Web www.iso.org

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

Violators may be prosecuted.

Contents

Page

Foreword	v
Introduction.....	vi
1 Scope.....	1
2 Normative references.....	1
3 Terms and definitions	1
4 Notation	4
5 Syntax.....	4
5.1 Well-formedness.....	4
5.2 Namespace.....	4
5.3 Whitespace.....	4
5.4 Core Elements	4
5.4.1 active element.....	4
5.4.2 assert element.....	4
5.4.3 extends element	5
5.4.4 include element	5
5.4.5 let element	5
5.4.6 name element.....	6
5.4.7 ns element.....	6
5.4.8 param element.....	6
5.4.9 pattern element.....	6
5.4.10 phase element.....	6
5.4.11 report element.....	7
5.4.12 rule element.....	7
5.4.13 schema element.....	7
5.4.14 value-of element	7
5.5 Ancillary Elements and Attributes.....	8
5.5.1 diagnostic element.....	8
5.5.2 diagnostics element.....	8
5.5.3 dir element	8
5.5.4 emph element.....	8
5.5.5 flag attribute	8
5.5.6 icon attribute	8
5.5.7 p element.....	8
5.5.8 properties element	8
5.5.9 property element	9
5.5.10 role attribute	9
5.5.11 see attribute.....	9
5.5.12 span element.....	9
5.5.13 subject attribute.....	9
5.5.14 title element.....	9
6 Semantics.....	10
6.1.1 Validation Function	10
6.2 Minimal Syntax	10
6.3 Abstract pattern processing	11
6.4 Query Language Binding.....	12
6.5 Order and side-effects	13
7 Conformance	14
7.1 Simple Conformance.....	14

- 7.2 Full Conformance14
- Annex A (normative) RELAX NG schema for Schematron15
- Annex B (normative) Schematron Schema for Additional Constraints.....19
- Annex C (normative) Default Query Language Binding.....20
- Annex D (informative) Schematron Validation Report Language21
 - D.1 Description21
 - D.2 RELAX NG Compact Syntax Schema21
 - D.3 Schematron Schema23
- Annex E (informative) Design Requirements27
- Annex F (informative) Use of Schematron as a Vocabulary28
- Annex G (informative) Use of Schematron for Multi-Lingual Schemas.....29
- Annex H (normative) Query Language Binding for XSLT 230
- Annex I (informative) Query Language Binding for XPath 2.....32
- Annex J (informative) Query Language Binding for EXSLT33
- Annex K (informative) Query Language Binding for STX34
- Annex L (informative) Example usage of Schematron Properties35
 - L.1 Running Example.....35
 - L.2 Annotation with controlled vocabulary information35
 - L.3 Annotation with static information35
 - L.4 Annotation with static structured information35
 - L.5 Type annotation with XSD Simple Type36
 - L.6 Type annotation with dynamic information36
 - L.7 Type annotation with dynamic structured information36
 - L.8 Annotation with ISO19757-7 character repertoire reference.....36
 - L.9 SVRL document36
- Bibliography37

Full Text Available Free of Charge
at <https://standards.iso.org/iso/iec/19757-3-2016>
(Standard ID: 166298b-9fda-46a3-92fb-03c5791344d1)

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 19757-3 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 34, *Document description and processing languages*.

This second edition cancels and replaces the first edition (ISO/IEC 19757-3:2006), of which has been technically revised.

ISO/IEC 19757 consists of the following parts, under the general title *Information technology — Document Schema Definition Languages (DSDL)*:

- *Part 2: Regular-grammar-based validation -- RELAX NG*
- *Part 3: Rule-based validation -- Schematron*
- *Part 4: Namespace-based Validation Dispatching Language (NVDL)*
- *Part 5: Extensible Datatypes*
- *Part 7: Character Repertoire Description Language (CREPDL)*
- *Part 8: Document Semantics Renaming Language (DSRL)*
- *Part 9: Namespace and datatype declaration in Document Type Definitions (DTDs)*
- *Part 11: Schema association*

Introduction

ISO/IEC 19757 defines a set of Document Schema Definition Languages (DSDL) that can be used to specify one or more validation processes performed against Extensible Stylesheet Language (XML) or Standard Generalized Markup Language (SGML) documents. (XML is an application profile SGML ISO 8879:1986.)

A document model is an expression of the constraints to be placed on the structure and content of documents to be validated with the model. A number of technologies have been developed through various formal and informal consortia since the development of Document Type Definitions (DTDs) as part of ISO 8879, notably by the World Wide Web Consortium (W3C) and the Organization for the Advancement of Structured Information Standards (OASIS). A number of validation technologies are standardized in DSDL to complement those already available as standards or from industry.

To validate that a structured document conforms to specified constraints in structure and content relieves the potentially many applications acting on the document from having to duplicate the task of confirming that such requirements have been met. Historically, such tasks and expressions have been developed and utilized in isolation, without consideration of how the features and functionality available in other technologies might enhance validation objectives.

The main objective of ISO/IEC 19757 is to bring together different validation-related tasks and expressions to form a single extensible framework that allows technologies to work in series or in parallel to produce a single or a set of validation results. The extensibility of DSDL accommodates validation technologies not yet designed or specified.

In the past, different design and use criteria have led users to choose different validation technologies for different portions of their information. Bringing together information within a single XML document sometimes prevents existing document models from being used to validate sections of data. By providing an integrated suite of constraint description languages that can be applied to different subsets of a single XML document, ISO/IEC 19757 allows different validation technologies to be integrated under a well-defined validation policy.

The structure of this part of ISO/IEC 19757 is as follows. Clause 5 describes the syntax of an ISO Schematron schema. Clause 6 describes the semantics of a correct ISO Schematron schema; the semantics specify when a document is valid with respect to an ISO Schematron schema. Clause 7 describes conformance requirements for implementations of ISO Schematron validators. Annex A is a normative annex providing the ISO/IEC 19757-2 (RELAX NG) schema for ISO Schematron. Annex B is a normative annex providing the ISO Schematron schema for constraints in ISO Schematron that cannot be expressed by the schema of Annex A. Annex C is a normative annex providing the default query language binding to XSLT1. Annex D is a non-normative annex providing a ISO/IEC 19757-2 (RELAX NG compact syntax) schema and corresponding ISO Schematron schema for a simple XML language Schematron Validation Report Language. Annex E is a non-normative annex providing motivating design requirements for ISO Schematron. Annex F is a normative annex allowing certain Schematron elements to be used in external vocabularies. Annex G is a non-normative annex with a simple example of a multi-lingual schema.

This edition is backwards compatible with IS19757-3:2006, supercedes it, and augments it with the following capabilities: patterns may validate different documents, the inclusion mechanism has been supplemented by an enhanced extension mechanism, assertions may have linked properties, and SVRL may take richer text. As well, this edition provides extra query language bindings, in particular for XSLT2.

Considered as a document type, a Schematron schema contains natural-language assertions concerning a set of documents, marked up with various elements and attributes for testing these natural-language assertions, and for simplifying and grouping assertions.

Considered theoretically, a Schematron schema reduces to a non-chaining rule system whose terms are Boolean functions invoking an external query language on the instance and other visible XML documents, with syntactic features to reduce specification size and to allow efficient implementation.

Considered analytically, Schematron has two characteristic high-level abstractions: the pattern and the phase. These allow the representation of non-regular, non-sequential constraints that Part 2 cannot specify, and various dynamic or contingent constraints.

This part of ISO/IEC 19757 is based on the Schematron[2] assertion language. The let element is based on XCSL[4]. Other features arise from the half-dozen early Open Source implementations of Schematron in diverse programming languages and from discussions in electronic forums by Schematron users and implementers

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Full standard:
<https://standards.iteh.ai/catalog/standards/sist/e166298b-9fda-46a3-92fb-03c579134a0f/iso-iec-19757-3-2016>

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Full standard:
<https://standards.iteh.ai/catalog/standards/sist/e166298b-9fda-46a3-92fb-03c579134a0f/iso-iec-19757-3-2016>

Information technology — Document Schema Definition Languages (DSDL) — Part 3: Rule-based Validation - Schematron

1 Scope

This part of ISO/IEC 19757 specifies Schematron, a schema language for XML. This part of ISO/IEC 19757 establishes requirements for Schematron schemas and specifies when an XML document matches the patterns specified by a Schematron schema.

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.

IRI, RFC 3987, IETF Request for Comment, Current version, <http://tools.ietf.org/html/rfc3987>

W3C XML 1.0, Extensible Markup Language (XML) 1.0 (Fourth Edition), W3C Recommendation, 14 June 2006, <http://www.w3.org/TR/2006/PER-xml-20060614/>

XPath, XML Path Language (XPath) Version 1.0, W3C Recommendation, 16 November 1999, <http://www.w3.org/TR/xpath>

XPath2, XML Path Language (XPath) 2.0, W3C Recommendation, 23 January 2007, <http://www.w3.org/TR/xpath20/>

XPath2 Functions, XQuery 1.0 and XPath 2.0 Functions and Operators, W3C Recommendation, 23 January 2007, <http://www.w3.org/TR/xpath-functions/>

XDM, XQuery 1.0 and XPath 2.0 Data Model (XDM), W3C Recommendation, 23 January 2007, <http://www.w3.org/TR/xpath-datamodel/>

XSLT1, XSL Transformations (XSLT) Version 1.0, W3C Recommendation, 16 November 1999, <http://www.w3.org/TR/xslt>

XSLT2, XSL Transformations (XSLT) Version 2.0, W3C Recommendation, 23 January 2007, <http://www.w3.org/TR/xslt20/>

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply:

- 3.1 abstract pattern**
pattern in a rule that has been parameterized to enable reuse
- 3.2 abstract rule**
collection of assertions which can be included in other rules but which does not fire itself
- 3.3 active pattern**
pattern belonging to the active phase
- 3.4 active phase**
one particular phase, whose patterns are used for validation
- 3.5 assertion**
natural-language assertion with corresponding assertion test and ancillary attributes: assertions are marked up with assert and report elements
- 3.6 assertion test**
assertion modelled or implemented by a Boolean query; an assertion test "succeeds" or "fails"
- 3.7 compound document**
a notional instance document which has been divided into the original instance and a number of subordinate well-formed XML documents. These subordinate documents may also contain schema-like information.
- 3.8 correct schema**
schema that satisfies all the requirements of this part of ISO/IEC 19757
- 3.9 diagnostic**
named natural language statements providing information to end-users of validators concerning the expected and actual values together with repair hints
- 3.10 elaborated rule context expression**
single rule context expression which explicitly disallows items selected by lexically previous rule contexts in the same pattern
- 3.11 good schema**
correct schema with queries which terminate and do not add constraints to those of the natural-language assertions. Note: It may not be possible to compute that a schema is good.
- 3.12 implementation**
implementation of a Schematron validator
- 3.13 name**
token with no whitespace characters

3.14**natural-language assertion**

natural-language statement expressing some part of a pattern; a natural-language assertion is "met" or "unmet"

3.15**pattern**

named structure in instances specified in a schema by a lexically-ordered collection of rules

3.16**phase**

named, unordered collection of patterns; patterns may belong to more than one phase; two names, #ALL and #DEFAULT, are reserved with particular meanings

3.17**progressive validation**

the validation of constraints in stages determined or grouped to some extent by the schema author rather than, for example, entirely determined by document order

3.18**property**

named data giving additional metadata on an assertion or report

3.19**query language binding**

named set, specified in a document called a Query Language Binding, of the languages and conventions used for assertion tests, rule-context expressions and so on, by a particular Schematron implementation.

3.20**rule**

unordered collection of assertions with a rule-context expression and ancillary attributes

3.21**rule context**

element or other information item used for assertion tests; a rule is said to fire when an information item matches the rule context

3.22**rule-context expression**

a query to specify subjects; a rule-context is said to match an information item when that information item has not been matched by any lexically-previous rule context expressions in the same pattern and the information item is one of the information items that the query would specify

3.23**schema**

specification of a set of XML documents

3.24**subject**

particular information item which corresponds to the object of interest of the natural-language assertions and typically is matched by the context expression of a rule

3.25**valid with respect to a schema**

member of the set of XML documents described by the schema: an instance document is valid if no assertion tests in fired rules of active patterns fail

**3.26
variable**

constant value, evaluated within the parent schema, phase, pattern or rule and scoped within the parent schema, phase, pattern or rule

4 Notation

This part of ISO/IEC 19757 uses XPath to identify information items in Schematron schemas.

5 Syntax

5.1 Well-formedness

A Schematron schema shall be a well-formed XML document, according to the version of XML used.

5.2 Namespace

All elements shown in the grammar for Schematron are qualified with the namespace URI [IRI]:

<http://purl.oclc.org/dsdl/schematron>

In subsequent clauses, the prefix `sch` is taken as bound to the Schematron namespace URI for exposition purposes. The prefix `sch` is not reserved or required by this part of ISO/IEC 19757. Any element can also have foreign attributes in addition to the attributes shown in the grammar. A foreign attribute is an attribute with a name whose namespace URI is neither the empty string nor the Schematron namespace URI. Any non-empty element may have foreign child elements in addition to the child elements shown in the grammar. A foreign element is an element with a name whose namespace URI is not the Schematron namespace URI. There are no constraints on the relative position of foreign child elements with respect to other child elements.

5.3 Whitespace

Any element can also have as children strings that consist entirely of whitespace characters, where a whitespace character is one of U+0020, U+009, U+00D or U+00A. There are no constraints on the relative position of whitespace string children with respect to child elements.

NOTE Leading and trailing whitespace should be stripped from attributes defined by this part. Whitespace should be collapsed in elements defined by this part that allow text. Whitespace may be stripped from elements defined by this part that do not allow text.

5.4 Core Elements

The grammar for Schematron elements is given in Annex A.

5.4.1 active element

The required `pattern` attribute is a reference to a pattern that is active in the current phase.

5.4.2 assert element

An assertion made about the context nodes. The data content is a natural-language assertion. The required `test` attribute is an assertion test evaluated in the current context. If the test evaluates positive, the report succeeds. The optional `diagnostics` attribute is a reference to further diagnostic information.

The natural-language assertion shall be a positive statement of a constraint.