



BSI Standards Publication

Information technology — Document Schema Definition Languages (DSDL)

Part 7: Character Repertoire Description Language (CREPDL)

National foreword

This British Standard is the UK implementation of [ISO/IEC 19757-7:2020](#). It supersedes [BS ISO/IEC 19757-7:2009](#), which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee IST/41, Document description and processing language.

A list of organizations represented on this committee can be obtained on request to its committee manager.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2020

Published by BSI Standards Limited 2020

ISBN 978 0 580 52075 4

ICS 35.240.30

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 August 2020.

Amendments/corrigenda issued since publication

Date	Text affected
------	---------------

INTERNATIONAL STANDARD

ISO/IEC
19757-7

Second edition
2020-08-05

Information technology — Document Schema Definition Languages (DSDL) —

Part 7: Character Repertoire Description Language (CREPDL)

Technologies de l'information — Langages de définition de schéma de documents (DSDL) —

Partie 7: Langage de description de répertoire de caractères (CREPDL)



Reference number
ISO/IEC 19757-7:2020(E)

AVA **Documents**
Document Sharing Hub
© ISO/IEC 2020



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2020, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Notation	2
5 Overview	3
5.1 Basic constructs and compound constructs	3
5.2 Characters and code points	3
5.3 Grapheme clusters	3
5.4 Kernel and Hull	3
6 Syntax	3
6.1 General	3
6.2 RELAX NG schema	4
6.3 NVDL script	5
6.4 Regular Expressions	5
7 Semantics	5
7.1 General	5
7.2 char 6	
7.3 union	7
7.4 intersection	7
7.5 difference	7
7.6 ref 8	
7.7 repertoire	8
8 Validation	8
Annex A (informative) Differences of conformant processors	10
Annex B (informative) Example CREPDL schemas	11
Bibliography	15

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see <http://patents.iec.ch>).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document 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 19757-7:2009](http://www.iso.org/iso/19757-7:2009)), which has been technically revised. It also incorporates the Technical Corrigendum ISO/IEC 19757-7:2009/Cor 1:2015.

The main changes compared to the previous edition are as follows:

- addition of validation of grapheme clusters such as 'n' followed by COMBINING GRAVE ACCENT (U+0300) and a CJK unified ideograph followed by a variation selector.
- addition of the Unicode Ideographic Variation Database as a registry.

A list of all parts in the [ISO/IEC 19757](http://www.iso.org/iso/19757) series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

[ISO/IEC 19757](#) (all parts) defines a set of Document Schema Definition Languages (DSDL) that can be used to specify one or more validation processes performed against Extensible Markup Language (XML) documents. A number of validation technologies are standardized in DSDL to complement those already available as standards or from industry.

The main objective of [ISO/IEC 19757](#) (all parts) is to bring together different validation-related technologies 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.

This document provides a language for describing character repertoires. Descriptions in this language can be referenced from schemas. Furthermore, they can also be referenced from forms and stylesheets.

Descriptions of character repertoires doesn't need to be exact. Non-exact descriptions are made possible by kernels and hulls, which provide the lower and upper limits, respectively.

The structure of this document is as follows. [Clause 5](#) provides an informal overview of CREPDL. [Clause 6](#) specifies the syntax of CREPDL schemas. [Clause 7](#) specifies the semantics of a correct CREPDL schema; the semantics specify when a code point or code point sequence is in a character repertoire described by a CREPDL schema. [Clause 8](#) defines the behaviour of CREPDL processors. Finally, [Annex A](#) describes differences of conformant CREPDL processors; [Annex B](#) provides examples of CREPDL schemas.

Although the first edition was restricted to the validation of characters, this edition can also enable the validation of grapheme clusters such as 'n' followed by COMBINING GRAVE ACCENT (U+0300) and a CJK unified ideograph followed by a variation selector.

CREPDL schemas conformant to the first edition do not conform to this edition. In particular, this edition changes the namespace name for CREPDL schemas.

This page deliberately left blank