

ISO/IEC JTC 1/SC 22 N

Date: 1999-10-22

ISO/IEC TR 10176:1999/PDAM 1

ISO/IEC JTC 1/SC 22/WG 20

Secretariat: ANSI

Information technology — Guidelines for the preparation of programming language standards — Amendment #1

Élément introductif — Élément central — Élément complémentaire

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.

Document type: Technical Report
Document subtype: Amendment
Document stage: (30) Committee
Document language: E

Copyright notice

This ISO document is a working draft or committee draft and is copyright-protected by ISO. While the reproduction of working drafts or committee drafts in any form for use by participants in the ISO standards development process is permitted without prior permission from ISO, neither this document nor any extract from it may be reproduced, stored or transmitted in any form for any other purpose without prior written permission from ISO.

Requests for permission to reproduce this document for the purpose of selling it should be addressed as shown below or to ISO's member body in the country of the requester:

*[Indicate :
the full address
telephone number
fax number
telex number
and electronic mail address*

as appropriate, of the Copyright Manager of the ISO member body responsible for the secretariat of the TC or SC within the framework of which the draft has been prepared]

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

Violators may be prosecuted.

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.

The main task of technical committees is to prepare International Standards, but in exceptional circumstances a technical committee may propose the publication of a Technical Report of one of the following types:

- type 1, when the required support cannot be obtained for the publication of an International Standard, despite repeated efforts;
- type 2, when the subject is still under technical development or where for any other reason there is the future but not immediate possibility of an agreement on an International Standard;
- type 3, when a technical committee has collected data of different kind from that which is normally published as an International Standard ("state of the art", for example).

Technical Reports of types 1 and 2 are subject to review within three years of publication, to decide whether they can be transformed into International Standards. Technical Reports of type 3 do not necessarily have to be reviewed until the data they provide are considered to be no longer valid or useful.

Technical Reports are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

Attention is drawn to the possibility that some of the elements of this Amendment may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Amendment 1 to Technical Report ISO/IEC TR 10176:1999 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 22, *Programming languages, their environments and system software interfaces*.

Introduction

This amendment corrects errors in the TR 10176, Annex A.

The table of characters to be used for identifiers in programming languages is based on ISO/IEC 10646-1:1993 including amendments 1 through 7.

When the second edition of ISO/IEC 10646-1 will be issued, Annex A of TR 10176 will have to be amended again to allow the newly coded characters, that meet the requirements, to be used for identifiers in programming languages.

Information technology — Guidelines for the preparation of programming language standards — Amendment #1

Replace the text of Annex A with the following text.

Annex A

Recommended extended repertoire for user-defined identifiers

The recommended extended repertoire consists of those characters which collectively can be used to generate word-like identifiers for most natural languages of the world. This list comprises the letters (combining or not), syllables, and ideographs from ISO/IEC 10646-1, together with the modifier letters and marks conventionally used as parts of words. The list excludes punctuation and symbols not generally included in words or considered appropriate for use in identifiers. Also excluded are most presentation forms of letters and a number of compatibility characters. The inclusion of combining characters corresponds to those allowed under a level 2 implementation of ISO/IEC 10646-1. These are the minimum required to do a reasonable job of representing word-like identifiers in Hebrew, Arabic, and scripts of South and Southeast Asia, which make general use of combining marks. However, combining marks for level 3 implementations of ISO/IEC 10646-1 are not included in the list, so as to avoid the problem of alternative representations of identifiers.

Attention is drawn to the fact that using the extended repertoire for identifiers may impact source code portability, since the presence of these characters in program text may not be supported on systems that implement less than the full repertoire of ISO/IEC 10646-1.

The character repertoire listed in this annex is based on the ISO/IEC 10646-1:1993 with its COR 1 and AMD 1 through 9. It is subject to expansion in the future, to track future amendments to the standard. However, characters currently listed in this Annex will not be removed from the recommended extended repertoire in future revisions.

The character repertoire listed in this annex should be conceived of as a recommendation for the minimum extended repertoire for use in user-defined identifiers. Each programming language standard or implementation of the standard can extend the repertoire at the adaptation, in accordance with established practice of identifier usage for the language and any additional user requirements that may be present. For example, the C language should allow U003F LOW LINE in addition to the character repertoire listed below; COBOL should allow U002D HYPHEN-MINUS as well; Java allows a rather large extension to support a level 3 implementation of 10646-1. Some programming language standards may allow half- or full-width compatibility characters from ISO/IEC 10646-1, and some of the standards, e.g. COBOL, may recognize these characters in a width-insensitive manner.

Programming language standards generally have restrictions on what characters may be allowed as the first character of an identifier. For example, digits are often constrained from appearing as the first character of an identifier. To assist in their identification, the decimal digits in ISO/IEC 10646-1 are separately noted in the list below. In addition, combining characters should not appear as the first character of an identifier. To maximize the chances of interoperability between programming languages (as for example, when linking compiled objects between languages), programming language standards and their implementations should follow these restrictions when making use of the extended repertoire for user-defined identifiers.

The recommended characters consist of the following characters of ISO/IEC 10646-1, using their code values in hexadecimal form. Combining characters for scripts are separated out and marked with a "C" following the respective script entries.

ISO/IEC TR 10176:1999/PDAM 1

Latin:	0041-005A, 0061-007A, 00AA, 00BA, 00C0-00D6, 00D8-00F6, 00F8-01F5, 01FA-0217, 0250-02A8, 1E00-1E9B, 1EA0-1EF9, 207F
Greek:	0386, 0388-038A, 038C, 038E-03A1, 03A3-03CE, 03D0-03D6, 03DA, 03DC, 03DE, 03E0, 03E2-03F3, 1F00-1F15, 1F18-1F1D, 1F20-1F45, 1F48-1F4D, 1F50-1F57, 1F59, 1F5B, 1F5D, 1F5F-1F7D, 1F80-1FB4, 1FB6-1FBC, 1FC2-1FC4, 1FC6-1FCC, 1FD0-1FD3, 1FD6-1FDB, 1FE0-1FEC, 1FF2-1FF4, 1FF6-1FFC
Cyrillic:	0401-040C, 040E-044F, 0451-045C, 045E-0481, 0490-04C4, 04C7-04C8, 04CB-04CC, 04D0-04EB, 04EE-04F5, 04F8-04F9
Armenian:	0531-0556, 0561-0587
Hebrew:	05D0-05EA, 05F0-05F2
Hebrew (C):	05B0-05B9, 05BB-05BD, 05BF, 05C1-05C2
Arabic:	0621-063A, 0640-064A, 0671-06B7, 06BA-06BE, 06C0-06CE, 06D0-06D3, 06D5, 06E5-06E6
Arabic (C):	064B-0652, 0670, 06D6-06DC, 06E7-06E8, 06EA-06ED
Devanagari:	0905-0939, 0950, 0958-0961
Devanagari (C):	0901-0903, 093E-094D, 0951-0952, 0962-0963
Bengali:	0985-098C, 098F-0990, 0993-09A8, 09AA-09B0, 09B2, 09B6-09B9, 09DC-09DD, 09DF-09E1, 09F0-09F1
Bengali (C):	0981-0983, 09BE-09C4, 09C7-09C8, 09CB-09CD, 09E2-09E3
Gurmukhi:	0A05-0A0A, 0A0F-0A10, 0A13-0A28, 0A2A-0A30, 0A32-0A33, 0A35-0A36, 0A38-0A39, 0A59-0A5C, 0A5E, 0A74
Gurmukhi (C):	0A02, 0A3E-0A42, 0A47-0A48, 0A4B-0A4D
Gujarati:	0A85-0A8B, 0A8D, 0A8F-0A91, 0A93-0AA8, 0AAA-0AB0, 0AB2-0AB3, 0AB5-0AB9, 0ABD, 0AD0, 0AE0
Gujarati (C):	0A81-0A83, 0ABE-0AC5, 0AC7-0AC9, 0ACB-0ACD
Oriya:	0B05-0B0C, 0B0F-0B10, 0B13-0B28, 0B2A-0B30, 0B32-0B33, 0B36-0B39, 0B5C-0B5D, 0B5F-0B61
Oriya (C):	0B01-0B03, 0B3E-0B43, 0B47-0B48, 0B4B-0B4D
Tamil:	0B85-0B8A, 0B8E-0B90, 0B92-0B95, 0B99-0B9A, 0B9C, 0B9E-0B9F, 0BA3-0BA4, 0BA8-0BAA, 0BAE-0BB5, 0BB7-0BB9
Tamil (C):	0B82-0B83, 0BBE-0BC2, 0BC6-0BC8, 0BCA-0BCD
Telugu:	0C05-0C0C, 0C0E-0C10, 0C12-0C28, 0C2A-0C33, 0C35-0C39, 0C60-0C61
Telugu (C):	0C01-0C03, 0C3E-0C44, 0C46-0C48, 0C4A-0C4D

Kannada:	0C85-0C8C, 0C8E -0C90, 0C92-0CA8, 0CAA-0CB3, 0CB5-0CB9, 0CDE, 0CE0-0CE1
Kannada (C):	0C82-0C83, 0CBE -0CC4, 0CC6-0CC8, 0CCA0CCD
Malayalam:	0D05-0D0C, 0D0E -0D10, 0D12-0D28, 0D2A-0D39, 0D60-0D61
Malayalam (C):	0D02-0D03, 0D3E-0D43, 0D46-0D48, 0D4A-0D4D,
Thai:	0E01-0E30, 0E32-0E33, 0E40-0E46, 0E50-0E59
Thai (C):	0E31, 0E34-0E3A, 0E47-0E4E
Lao:	0E81-0E82, 0E84, 0E87-0E88, 0E8A, 0E8D, 0E94-0E97, 0E99-0E9F, 0EA1-0EA3, 0EA5, 0EA7, 0EAA-0EAB, 0EAD -0EAE, 0EB0, 0EB2-0EB3, 0EBD, 0EC0-0EC4, 0EC6, 0EDC-0EDD
Lao (C):	0EB1, 0EB4-0EB9, 0EBB-0EBC, 0EC8-0ECD,
Tibetan:	0F00, 0F40-0F47, 0F49-0F69, 0F88-0F8B,
Tibetan (C):	0F18-0F19, 0F35, 0F37, 0F39, 0F71-0F84, 0F86-0F87, 0F90-0F95, 0F97, 0F99-0FAD, 0FB1-0FB7, 0FB9
Georgian:	10A0-10C5, 10D0-10F6
Hiragana:	3041-3093
Katakana:	30A1-30F6, 30FB-30FC
Bopomofo:	3105-312C
Hangul:	AC00-D7A3
CJK Unified Ideographs:	4E00-9FA5
Digits:	0030-0039, 0660-0669, 06F0-06F9, 0966-096F, 09E6-09EF, 0A66-0A6F, 0AE6-0AEF, 0B66-0B6F, 0BE7-0BEF, 0C66-0C6F, 0CE6-0CEF, 0D66-0D6F, 0E50-0E59, 0ED0-0ED9, 0F20-0F29
Special characters:	00B5, 02B0-02B8, 02BB, 02BD-02C1, 02D0-02D1, 02E0-02E4, 037A, 0559, 093D, 0B3D, 1FBE, 203F-2040, 2102, 2107, 210A-2113, 2115, 2118-211D, 2124, 2126, 2128, 212A-2131, 2133-2138, 2160-2182, 3005-3007, 3021-3029