

Extracts from ISO/IEC 10646-1 : 1993

**Information technology —
Universal Multiple-Octet
Coded Character Set (UCS)**

**Part 1: Architecture and Basic
Multilingual Plane**

Committees responsible for this British Standard

The preparation of this British Standard was entrusted by the Information Systems Technology Standards Policy Committee (IST/-) to Technical Committee IST/2, upon which the following bodies were represented:

- British Computer Society
- British Federation of Printing Machinery and Supplies Ltd.
- British Telecommunications plc
- Digital Equipment Co. Ltd.
- IBM United Kingdom Ltd.
- Information Systems Committee of the Universities Funding Council
- Institution of Electrical Engineers
- International Computers Limited
- Philips Research Laboratories

This British Standard, having been prepared under the direction of the Information Technology Systems Standards Policy Committee, was published under the authority of the Standards Board and comes into effect on 15 August 1993

© BSI 1993

The following BSI references relate to the work on this standard:
 Committee reference IST/2
 Draft for comment 90/68644 DC

ISBN 0 580 22228 4

Amendments issued since publication

Amd. No.	Date	Text affected

Licensed copy: Techstreet Content, ISO Exchange - Michigan, Version correct as of 09/03/2021.

Contents

	Page
Committees responsible	Inside front cover
National foreword	ii
<hr/>	
1 Scope	1
2 Conformance	1
3 Normative references	2
4 Definitions	2
5 General structure of the UCS	3
6 Basic structure and nomenclature	4
7 Special features of the UCS	7
8 The Basic Multilingual Plane	7
9 Other planes	7
10 The Restricted Use zone	7
11 Private Use groups and planes	8
12 Revision and updating of the UCS	8
13 Subsets	8
14 Coded representation forms of the UCS	8
15 Implementation levels	9
16 Use of control functions with the UCS	9
17 Declaration of identification of features	9
18 Structure of the code tables and lists	11
19 Block names	11
20 Characters in bi-directional context	12
21 Special characters	12
22 Order of characters	12
23 Combining characters	13
24 Hangul syllable composition method	14
25 Code tables and lists of character names	14
<hr/>	
Annexes	
A Collections of graphic characters for subsets	152
B List of combining characters	154
C Mirrored characters in Arabic bi-directional context	157
D Alternate format characters	159
E Alphabetically sorted list of character names	162
F The use of 'signatures' to identify UCS	196
G UCS transformation format (UTF-1)	197
H Recommendation for combined receiving/originating devices with internal storage	199
J Notations of octet value representations	200
K Character naming guidelines	201
L Sources of characters	203
M External references to character repertoires	205
N Scripts under consideration for future editions of ISO/IEC 10646	207
<hr/>	

National foreword

This Published Document has been prepared under the direction of the Information Systems Technology Policy Committee, by technical committee IST/2.

It provides extracts from ISO/IEC 10646-1 : 1993¹⁾ *Information technology — Universal Multiple — Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane* for the convenience of those users who do not need the code tables of unified Han and Korean Hangul characters. We expect that this information will be most frequently required in the UK.

Extracts from the following pages are provided:

- pages 1-129;
- pages 240-261;
- pages 699-754.

Compliance with a British Standard does not of itself confer immunity from legal obligations.

¹⁾ ISO/IEC 10646-1 : 1993 is implemented as BS ISO/IEC 10646-1 : 1993.

Information technology — Universal Multiple-Octet Coded Character Set (UCS) —

Part 1:

Architecture and Basic Multilingual Plane

1 Scope

ISO/IEC 10646 specifies the Universal Multiple-Octet Coded Character Set (UCS). It is applicable to the representation, transmission, interchange, processing, storage, input and presentation of the written form of the languages of the world as well as additional symbols.

This part of ISO/IEC 10646 specifies the overall architecture, and

- defines terms used in ISO/IEC 10646;
- describes the general structure of the coded character set;
- specifies the Basic Multilingual Plane (BMP) of the UCS, and defines a set of graphic characters used in scripts and the written form of languages on a world-wide scale;
- specifies the names for the graphic characters of the BMP, and the coded representations;
- specifies the four-octet (32-bit) canonical form of the UCS: UCS-4;
- specifies a two-octet (16-bit) BMP form of the UCS: UCS-2;
- specifies the coded representations for control functions;
- specifies the management of future additions to this coded character set.

The UCS is a coding system different from that specified in ISO 2022. The method to designate UCS from ISO 2022 is specified in 17.2.

2 Conformance

2.1 General

Whenever Private Use characters are used as specified in ISO/IEC 10646, the characters themselves shall not be covered by these conformance requirements.

2.2 Conformance of Information Interchange

A coded-character-data-element (CC-data-element) within coded information for interchange is in conformance with ISO/IEC 10646 if

- a) all the coded representations of graphic characters within that CC-data-element conform to clauses 6 and 7, to an identified form chosen from clause 14, and to an identified implementation level chosen from clause 15;
- b) all the graphic characters represented within that CC-data-element are taken from those within an identified subset (clause 13);
- c) all the coded representations of control functions within that CC-data-element conform to clause 16.

A claim of conformance shall identify the adopted form, the adopted implementation level and the adopted subset by means of a list of collections and/or characters.

2.3 Conformance of devices

A device is in conformance with ISO/IEC 10646 if it conforms to the requirements of item a) below, and either or both of items b) , and c).

NOTE - The term *device* is defined (in 4.17) as a component of information processing equipment which can transmit and/or receive coded information within CC-data-elements. A device may be a conventional