

ICS 35.040  
CCS L 71



**NATIONAL STANDARD OF THE  
PEOPLE'S REPUBLIC OF CHINA**

**GB/T 13000-2025**

---

**Information technology - Universal coded character set(UCS)  
(English Translation)**

**Issued on: January 24, 2025**

**Implemented on: August 1, 2025**

---

**Issued by: State Administration for Market Regulation, China National  
Standardization Administration**

## Table of Contents

|                                        |
|----------------------------------------|
| Preface                                |
| 1 Scope                                |
| 2 Normative references                 |
| 3 Terms and Definitions                |
| 4 Compliance                           |
| 4.1 General requirements               |
| 4.2 Compliance of information exchange |
| 4.3 Equipment compliance               |
| 5 Electronic Data Annex                |
| 6 Overall structure of UCS             |
| 7 Basic structure and nomenclature     |
| 7.1 Structure                          |
| 7.2 Character Encoding                 |
| 7.3 Code Type                          |
| 7.4 Character Naming                   |
| 7.5 Character Aliases                  |
| 7.6 Code Point Short Identifier (UID)  |
| 7.7 UCS sequence identifier            |
| 7.8 Octet Sequence Identifier          |
| 8 Revision and Update of UCS           |
| 9 Subset                               |
| 9.1 Overview                           |
| 9.2 Finite subsets                     |
| 9.3 Selecting a subset                 |
| 10 UCS encoding form                   |
| 10.1 Overview                          |
| 10.2 UTF-8                             |
| 10.3 UTF-16                            |
| 10.4 UTF-32                            |
| 11 UCS encoding scheme                 |
| 11.1 Overview                          |
| 11.2 UTF-8                             |

|      |                                                   |
|------|---------------------------------------------------|
| 11.3 | UTF-16BE                                          |
| 11.4 | UTF-16LE                                          |
| 11.5 | UTF-16                                            |
| 11.6 | UTF-32BE                                          |
| 11.7 | UTF-32LE                                          |
| 11.8 | UTF-32                                            |
| 12   | Control functions and UCS combined use            |
| 13   | Declaration of identification characteristics     |
| 13.1 | Purpose and reason of labeling                    |
| 13.2 | Identification of the UCS coding scheme           |
| 13.3 | Identification of graphic character subsets       |
| 13.4 | Identification of control function sets           |
| 13.5 | Identification of the ISO/IEC 2022 coding system  |
| 14   | Structure of code tables and character name lists |
| 15   | Block and collection names                        |
| 15.1 | Block Name                                        |
| 15.2 | Collection Name                                   |
| 16   | Mirror Characters in Bidirectional Contexts       |
| 16.1 | Mirror Characters                                 |
| 16.2 | Directionality of Bidirectional Context           |
| 17   | Special characters                                |
| 17.1 | Overview                                          |
| 17.2 | Spacer Characters                                 |
| 17.3 | Currency Symbols                                  |
| 17.4 | Format Characters                                 |
| 17.5 | Ideographic Descriptors                           |
| 17.6 | Variant Selectors and Variant Sequences           |
| 18   | Character appearance                              |
| 19   | Compatible characters                             |
| 20   | Character order                                   |
| 21   | Combining characters                              |
| 21.1 | Sequence of combining characters                  |
| 21.2 | Combination categories and regular order          |
| 21.3 | Display in the code table                         |
| 21.4 | Alternative Code Representations                  |

- 21.5 Multiple Combination Characters
- 21.6 Collections containing combining characters
- 21.7 Combining grapheme connectors
- 22 Normalized form
- 23 Characteristics of individual words and symbol vocabulary
  - 23.1 The syllable compounding of Hangul (Korean)
  - 23.2 Characteristics of the writing system in India and some South Asian countries
  - 23.3 Byzantine musical notation
  - 23.4 Etymological references for graphic symbols
- 24 CJK Chinese character etymology reference
  - 24.1 Etymology Reference List
  - 24.2 CJK Chinese Character Etymology Reference File
  - 24.3 CJK unified Chinese character etymology reference information representation method
  - 24.4 CJK compatible Chinese character etymology reference information representation method
- 25 Etymology of Xixia characters
  - 25.1 Etymology Reference List
  - 25.2 Etymology of Xixia characters Reference file
  - 25.3 The representation of etymological reference information of Xixia characters
- 26 Etymology of Nüshu characters
  - 26.1 Etymology Reference List
  - 26.2 Etymology of Nüshu characters Reference document
- 27 Character names and notes
  - 27.1 Entity Name
  - 27.2 Name structure
  - 27.3 Single name
  - 27.4 Name Invariance
  - 27.5 Name Uniqueness
  - 27.6 Character names of CJK Chinese characters
  - 27.7 Names of Xixia characters
  - 27.8 Names of characters in Nüshu
  - 27.9 Character names of Khitan small characters
  - 27.10 Character names of Hangul (Korean) syllables
- 28 Named UCS sequence identifier
- 29 The structure of the Basic Multilingual Plane (BMP)

|      |                                                                                                                   |
|------|-------------------------------------------------------------------------------------------------------------------|
| 30   | Structure of the Supplementary Multilingual Plane (SMP) for the Coding of Texts and Symbols                       |
| 31   | Structure of the Supplementary Ideographic Plane (SIP)                                                            |
| 32   | Structure of the Third Ideographic Plane (TIP)                                                                    |
| 33   | Auxiliary Special Purpose Plane (SSP) Structure                                                                   |
| 34   | Code table and character name list                                                                                |
| 34.1 | Overview                                                                                                          |
| 34.2 | Code Table                                                                                                        |
| 34.3 | List of character names                                                                                           |
| 34.4 | Standardized Variant Sequence Summary                                                                             |
| 34.5 | Code table and character name list                                                                                |
|      | Appendix A (Informative) Characters used in identifiers                                                           |
|      | Appendix B (Informative) External References to Character Vocabularies                                            |
| B.1  | Character Vocabulary and Its Encoding Method                                                                      |
| B.2  | Identification of ASN.1 Character Abstract Syntax                                                                 |
| B.3  | Identification of ASN.1 Character Transfer Syntax                                                                 |
|      | Appendix C (Informative) Notation for Eight-bit Value Representation                                              |
|      | Appendix D (informative) Recommendation 76 for combined receiving and transmitting equipment with internal memory |
|      | Appendix E (Normative) Collection of graphic characters for subsets                                               |
| E.1  | Collection of coded graphic characters                                                                            |
| E.2  | Block Name List                                                                                                   |
| E.3  | Fixed collections of the entire UCS (except the Unicode collection)                                               |
| E.4  | CJK Collection                                                                                                    |
| E.5  | Other collections                                                                                                 |
| E.6  | Unicode Collection                                                                                                |
|      | Appendix F (informative) Alphabetical list of character names                                                     |
|      | Appendix G (Informative) Format Characters                                                                        |
| G.1  | Overview                                                                                                          |
| G.2  | General format characters                                                                                         |
| G.3  | Format characters applicable to specific text                                                                     |
| G.4  | Interline comment characters                                                                                      |
| G.5  | Reverse format characters                                                                                         |
| G.6  | Shorthand format characters                                                                                       |
| G.7  | Invisible mathematical operators                                                                                  |
| G.8  | Western Music Notation                                                                                            |