SLOVENSKI STANDARD
SIST ISO 15919:2005

01-november-2005

±bʐcfa Uʰ[_U΄]b˙Xc_i a YbʰUWʲ΄UˈËˈHfUbg`]hYfUWʲ΄U˙d]gUj Y˙XYj UbU[Uf]˙]b˙Xfi [ ]\ gcfcXbʰ]\ ˙]bX]ʰg_ʰ]\ ˙d]gUj ˙j ˈ΄Uʰ]b] bY˙nbU_Y

Information and documentation -- Transliteration of Devanagari and related Indic scripts into Latin characters

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Information et documentation -- Translittération du Devanagari et des écritures indiennes liées en caractères latins

Ta slovenski standard je istoveten z: **ISO 15919:2001**

<u>ICS:</u>

01.140.10   Úãæ}ʰ/ɟɅ¦^ ¦\[çæ}ʰ   Writing and transliteration

SIST ISO 15919:2005   **en**

iTeh STANDARD PREVIEW

(standards.iteh.ai)

INTERNATIONAL STANDARD

ISO
15919

First edition
2001-10-01

Information and documentation —
Transliteration of Devanagari and related
Indic scripts into Latin characters

*Information et documentation — Translittération du Devanagari et des
écritures indiennes liées en caractères latins*

© ISO 2001

ISO 15919:2001(E)

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Contents

Page

ISO 15919:2001(E)

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

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

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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

International Standard ISO 15919 was prepared by Technical Committee ISO/TC 46, *Information and documentation*, Subcommittee SC 2, *Conversion of written languages*.

Annexes A, B and C form a normative part of this International Standard. Annexes D, E and F are for information only.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Introduction

Script conversion is often required for documents such as historical and literary texts, geographical texts (including maps and atlases), bibliographies, catalogues, lists and passports (and other identification documents).

Text in Devanagari script or other Indic scripts sometimes needs to be shown in Latin script, where users, or equipment that they are using, cannot read or write the text.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

iTeh STANDARD PREVIEW

(standards.iteh.ai)

Information and documentation — Transliteration of Devanagari and related Indic scripts into Latin characters

1 Scope

This International Standard provides tables which enable the transliteration into Latin characters from text in Indic scripts which are largely specified in rows 09 to 0D of UCS (ISO/IEC 10646-1 and Unicode).

The tables provide for the Devanagari, Bengali (including the characters used for writing Assamese), Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Sinhala, Tamil, and Telugu scripts which are used in India, Nepal, Bangladesh and Sri Lanka. The Devanagari, Bengali, Gujarati, Gurmukhi, and Oriya scripts are North Indian scripts, and the Kannada, Malayalam, Tamil, and Telugu scripts are South Indian scripts.

The Burmese, Khmer, Thai, Lao and Tibetan scripts which also share a common origin with the Indic scripts, and which are used predominantly in Myanmar, Cambodia, Thailand, Laos, Bhutan and the Tibetan Autonomous Region within China, are not covered by this International Standard.

This International Standard applies to transliteration of Devanagari, and to Indic scripts related to Devanagari, independent of the period in which it is or was used (i.e. for Devanagari script it can be used for transliterating text in classical Sanskrit, Hindi, Marathi, and the Vedic language, for instance).

Other Indic scripts whose character repertoires are covered by the tables may also be transliterated using this International Standard.

Options in this International Standard are defined in clause 9.

2 Conformance

Text originally in non-Latin script which is converted to a Latin-script representation conforms to this International Standard with or without any of the specific recommendations, if it follows the rules defined in 8.1 and the conversion tables given in clause 7 and normative annexes A and B, with or without following any of the three recommendations given in 8.2 and clause 12, all in accordance with the options defined in clause 9.

A claim of conformance shall specify which options have been chosen, and which recommendations have been followed.

3 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this International Standard. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO/IEC 10646-1, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*

ISO/IEC 646:1991, *Information technology — ISO 7-bit coded character set for information interchange*

ISO 15919:2001(E)

4 Terms and definitions

For the purposes of this International Standard, the following terms and definitions apply.

4.1
conversion
representing graphic characters from a source script by the graphic characters of a target script, most commonly by romanization

NOTE   The two basic methods of conversion of a system of writing are transliteration and transcription. The use of the terms source script and target script in transliteration is analogous to the terms source language and target language in translation.

4.2
script
set of graphic characters used for the written form of one or more languages

4.3
graphic character
character (other than a control character) that has a visual representation, normally handwritten, printed or displayed

NOTE   A graphic character is a single element of a script. Examples are letters, conjunct characters, numerical digits, punctuation marks or diacritical marks.

4.4
reverse transliteration
process whereby the characters of a target script are transliterated into those of the source script

NOTE   This International Standard aims to enable reverse-transliterated text to be identical to the original source text up to equivalent orthography. However, non-reversible transcription-like transliterations are often found to be useful when quoting recent material.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

SIST ISO 15919:2005
https://standards.iteh.ai/catalog/standards/sist/634075a4-42d6-4df0-8a94-
f3b699196f78/sist-iso-15919-2005

4.5
romanization
conversion of non-Latin graphic characters into Latin graphic characters, using either transliteration or transcription

4.6
transcription
representation of the sounds of a source language by graphic characters associated with a target language

4.7
transliteration
representation of the graphic characters of a source script by the graphic characters of a target script

NOTE   In transcription, pronunciation conventions are of primary importance, while in transliteration, writing conventions are of primary importance.

4.8
UCS
Universal Multiple-Octet Coded Character Set (UCS) as defined in ISO/IEC 10646-1

NOTE 1   The Indic scripts listed in ISO/IEC 10646-1:1993 form a subset (with identical codes) of the Indic scripts listed in ISO/IEC 10646-1:2000. Similarly, the Indic scripts listed in the Unicode standard (version 1.0 onwards) form a subset (with identical codes) to the Indic scripts listed in ISO/IEC 10646-1:2000 and the Unicode standard, version 3.0. Any of these standards provide valid character codes for the specific characters concerned.

NOTE 2   ISO/IEC 10646-1 is increasingly used for providing character identifiers in a wide range of International Standards, including some in this International Standard. Use of these identifiers does not impose any requirements to use ISO/IEC 10646-1 or any other character coding standard to represent either the source characters or the target characters in any computer system or in information interchange.

5   Abbreviated terms

— Ben.   Bengali script

— Dev.   Devanagari script

— Guj.   Gujarati script

— Gur.   Gurmukhi script

— Kan.   Kannada script

— Mal.   Malayalam script

— Ori.   Oriya script

— Tam.   Tamil script

— Tel.   Telugu script

— Sin.   Sinhala script

— P-A.   Perso-Arabic script

iTeh STANDARD PREVIEW
(standards.iteh.ai)

6   Characteristics of Indic scripts

Characters in Indic scripts represent vowels, consonants and their combinations; nasalization, breathings, numerals and punctuation.

Each vowel has a full form (occupying a full character space in text, and required when beginning a word or in vowel hiatus) and a combining form (*mātrā*) used when the vowel follows a consonant, except that the short *a* standing at the beginning of Indic alphabets has only a full form, because no mātrā is required (see below).

Consonants include stops, semivowels, spirants, and other speech sounds. Stop consonants are arranged in classes, or *vargas*, according to the point of articulation, and within each class are subdivided into unvoiced or voiced, unaspirated or aspirated consonants, and a nasal consonant.

Characters for consonants are most simply quoted in a form which includes the inherent vowel *a*, as in the first consonant *ka* in Table 1. The inherent vowel is removed by the virāma sign of the relevant script (Dev., Ben., Guj.,

Gur., Ori. ◌੍ , Tam. ◌், Tel. ◌్ , Kan. ◌్, Mal. ◌്, Sin. ◌්. The relevant mātrā is used when any other vowel

follows a consonant. Consonant clusters frequently form conjunct characters. Use of virāma to form consonant clusters is unusual, except in Tamil where it is the normal method. When a mātrā is associated with a consonant, it replaces the inherent vowel. Mātrās have various forms, even in a single script, and details may be found in dictionaries and grammars.

It is important to note that many Indic characters have variant forms. Such differences of orthography are not distinguished in this International Standard.

Devanagari is used for writing various modern languages, such as Hindi, Marathi, Rajasthani and other languages in India, and Nepali in Nepal. Devanagari and most of the other Indic scripts are used for writing classical languages often used in religious texts, such as the Sanskrit and Vedic languages, and Pali. In some cases, text in Indic scripts uses additional characters for writing words in languages which do not normally use these scripts. Thus some Urdu consonants are typically represented by adding a dot (*nuqta*) below certain letters (see Table 1, normative annex C and informative annex D). Two English vowels may also be represented. Devanagari has also been extended to write South Indian languages.

ISO 15919:2001(E)

Sinhala script (used in Sri Lanka) has additional letters, in comparison with the scripts which are used in India, Nepal and Bangladesh. Tamil script (used in South India and also in Sri Lanka) uses fewer characters, in comparison with other scripts which are used in India, Nepal, Bangladesh and Sri Lanka.

When the Bengali script is used to write the Assamese language (in parts of North India), two characters not used in writing Bengali are required. Hence the Assamese script is sometimes regarded as separate from the Bengali script.

7 Transliteration tables

7.1 The transliteration from each Indic script to the Latin script shall be as specified in the Tables 1 to 10 and A.3, subject to the rules specified in 8.1 and the options specified in clause 9.

7.2 The structure of the transliteration tables is explained in the following paragraphs.

The target characters (Latin script) fall within the ranges 0020-01FF and 0300-0332 of ISO/IEC 10646-1:2000.

The repertoires for many of the source characters fall within the following ranges of ISO/IEC 10646-1:2000, for the script concerned:

— 0900-097F   Devanagari

— 0980-09FF   Bengali

— 0A00-0A7F   Gurmukhi

— 0A80-0AFF   Gujarati

— 0B00-0B7F   Oriya

— 0B80-0BFF   Tamil

— 0C00-0C7F   Telugu

— 0C80-0CFF   Kannada

— 0D00-0D7F   Malayalam

— 0D80-0DFF   Sinhala

Some additional Indic scripts whose character repertoires are included in the character repertoires of these scripts are listed in informative annex E.

Consonants are shown with their inherent vowel *a*.

Only a single form of each Indic character is shown, just as in ISO/IEC 10646-1. Specifications of alternative forms of these characters, including shapes when these are included in conjunct forms or in consonant-vowel combinations, are outside the scope of this International Standard.

This clause gives tables for each script, with references to the rules of 8.1. Numerals are shown in Table A.3 of annex A. Tables 1 to 10 are in the order of ISO 10646-1:2000. Vowels are shown in full form followed by a typical form of the corresponding mātrā.

Normative annex A gives tables showing linguistically equivalent characters in each script (except that Gurmukhi Bindi is not exactly equivalent to anusvara in the other scripts). Extended and ancient characters, apart from numerals, are shown in Table A.2 unless an equivalent modern character exists in another script, in which case they are enclosed in round brackets in Table A.1. (See also the requirements in clause 10.) In Tables A.1 to A.3 the scripts are ordered according to similarity of character repertoires.