मानकः पथप्रदर्शकः

COMPENDIUM OF INDIAN STANDARDS ON INDIAN LANGUAGE TECHNOLOGIES

PREFACE

Standardization in Indian language technologies is a significant area of work overseen by Bureau of Indian Standards (BIS), majorly through Indian Language Technologies and Products Sectional Committee (LITD 20). These efforts aim to facilitate the processing and interchange of information in various Indian languages. Standards have been developed for crucial aspects including coding of machine-readable characters like MICR and OCR, as well as character sets for information interchange, progressing through 7-bit and 8-bit coded character sets. These Indian Standards are identical are aligned with International Organization for Standardization (ISO) standards, with specific conventions fitting the requirements of Indian Languages.

The common phonetic structure and shared origin from the ancient Brahmi script of many Indian languages have played a key role in enabling the development of common character sets like ISCII and later aligning with international standards like IS 16350 (Based on ISO/IEC 10646, Universal Coded Character Set). Beyond coding, standards also cover implementation details such as usage on magnetic tape and, notably, keyboard layouts, including the In-script and Enhanced In script layouts, designed to support multiple Indian scripts, often based on language-specific keyboards while retaining phonetic parallelism.

Furthermore, standards are being developed for linguistic resources, such as Part-of-Speech (PoS) tag sets for annotation, critical for Natural Language Processing (NLP) research and application development. These standards are **code of practices** and designed with principles like extensibility, layered approaches, and compatibility to handle the diversity and complexity of Indian languages.

IS 10315:1997 - 7-BIT CODED CHARACTER SET FOR INFORMATION INTERCHANGE (First Revision)

This Indian Standard specifies a 7-bit coded character set consisting of 128 combinations, which include control characters and graphic characters. It is primarily intended for use in information processing systems, associated equipment, and data communication systems. The standard also provides guidance on exercising options for specific national and application-oriented versions and identifies the International Reference Version (IRV). It allows for code extension techniques when the character set is insufficient for particular applications.

IS 11407-1988 - IMPLEMENTATION OF THE 7-BIT CODED CHARACTER SET AND ITS 7-BIT AND 8-BIT EXTENSIONS ON 9-TRACK 12.7MM MAGNETIC TAPE

This International Standard (adopted as an Indian Standard) specifies the implementation. It details the implementation of the 7-bit coded character set and its 7-bit and 8-bit extensions. The implementation is specifically defined for use on 9-track 12.7 mm (0.5 in) magnetic tape.

IS 12326 : 2005 ISO 2022- INFORMATION TECHNOLOGY - CHARACTER CODE STRUCTURE AND EXTENSION TECHNIQUES (FIRST REVISION)

This Indian Standard, identical to ISO/IEC 2022:1994, specifies the structure of 8-bit codes and 7-bit codes used for the coding of character sets in information technology. It details various techniques for extending the capabilities of these elementary codes, including means to increase the number of available control functions and graphic characters. The standard provides techniques to construct and formalize specific code definitions and identify their structure and constituent elements. These codes are designed for data processed sequentially in a forward direction.

IS 12877 : 1989 ISO 3275:1974- IMPLEMENTATION OF THE 7-BIT CODED CHARACTER SET AND ITS 7-BIT AND 8-BIT EXTENSIONS ON 3.81 MM MAGNETIC TAPE CASSETTE FOR DATA INTERCHANGE

This Indian Standard, specifies the implementation of the 7-bit coded character set and of its 7-bit and 8-bit extensions. This implementation is specifically for the interchange of data on 3.81 mm magnetic tape cassette. It references ISO 646 for the 7-bit set and ISO 2022 for code extension techniques.

IS 11203 : 2024 ISO 2033 : 1983 - CODING OF MACHINE READABLE CHARACTERS (MICR AND OCR) FOR INFORMATION PROCESSING

(FIRST REVISION)

This standard serves as a specification for the coded representation of characters recognizable by machine reading equipment such as MICR and OCR, detailing the characters, their coding, and specific fonts like E13B and CMC7. It is an Indian Standard identical to ISO 2033:1983, adopted without deviations for encoding machine-readable characters. The standard is applicable to information processing systems and equipment, making it relevant to stakeholders involved in MICR and OCR technology.

IS 10315 : 1997 ISO/IEC 646 : 1991 - 7 BIT CODED CHARACTER SET FOR INFORMATION INTERCHANGE (FIRST REVISION)

This standard defines a 7-bit coded character set comprising 128 characters for use in information interchange within information processing systems and data communication. It serves as a foundational standard for encoding machine-readable characters and including specific fonts like E13B and CMC7.

IS 10401 : 1998 ISO/IEC 4873 : 1991 - 8-BIT CODED CHARACTER SET FOR INFORMATION INTERCHANGE (First Revision)

This standard specifies an 8-bit coded character set utilizing code extension techniques defined in ISO 2022, for information interchange and systems. It is applicable to information interchange and systems that require an 8-bit character set compatible with the 7-bit ISO/IEC 646 standard and utilizing ISO 2022 code extension techniques.

IS 13194 : 1991 - INDIAN SCRIPT CODE FOR INFORMATION INTERCHANGE — ISCII

The standard specifies the Indian Script Code for Information Interchange (ISCII), a character encoding system enabling simultaneous use of English and Indian scripts in 7- or 8-bit environments. It supports ten Brahmi-based scripts, maintains compatibility with ASCII/ISO standards, and defines a unified keyboard layout (Inscript). Key features include code tables, phonetic-based encoding, support for complex characters and script selection attributes.

IS 16350 : 2016 - ENHANCED INSCRIPT KEYBOARD LAYOUTS

This standard defines character code-sets and Inscript keyboard layouts for representing Indian languages in digital environments. It builds upon ISCII-based layouts and aligns with ISO/IEC 10646 to support Brahmi-

based scripts across 22 official Indian languages. The standard ensures compatibility, usability, phonetic consistency, and bi-lingual support across devices and platforms. It also provides developers with guidance on text input, display, storage, and processing complexities.

IS 17627 : 2021 - LINGUISTIC RESOURCES — POS TAG SET FOR INDIAN LANGUAGES — GUIDELINES FOR DESIGNING TAGSETS AND SPECIFICATION

This standard provides a framework for designing and using Part-of-Speech (POS) tagsets for Indian languages to support linguistic annotation in NLP applications. It defines a hierarchically structured, extensible, and user-friendly superset of POS tags applicable to major Indian languages. Intended for linguists, developers, and researchers, it emphasizes clarity, compatibility, and a layered approach to linguistic data. It also complements morphological analysis and enables deeper text-level processing.

IS 16333 : Part 3 : 2022: INDIAN LANGUAGE SUPPORT FOR MOBILE PHONE HANDSETS

This standard specifies requirements for mobile handset keyboards to support input in Indian languages. It ensures that virtual keyboards provide consistent layouts, character sets, and usability for 22 scheduled Indian Languages. The standard promotes linguistic inclusivity and enhances user experience by enabling seamless text input in regional languages on mobile devices.

DIGITAL PLATFORMS OF BIS

BIS Website: www.bis.gov.in Manak

Online: www.manakonline.in

BIS Care App: https://play.google.com/store/apps/details?id=com.bis.bisapp

Facebook: https://www.facebook.com/IndianStandards/

Instagram: https://www.instagram.com/indianstandards/

Youtube: http://bit.ly/BISYouTubeOfficial

Linkedin: http://bit.ly/BISLinkedInOfficial

Twitter: https://twitter.com/IndianStandards

CONTACT DETAILS

Shri Devansh Deolekar, Sc-D (LITD)

Email Id: litd@bis.gov.in

BUREAU OF INDIAN STANDARDS

MANAK BHAWAN G, BAHADUR SHAH ZAFAR MARG, NEW DELHI 110002

Scan Code