

1. यूनिकोड क्या है?

सर्वप्रथम यह समझना आवश्यक है कि यूनिकोड क्या है? क्या यूनिकोड कोई फॉन्ट है? क्या यूनिकोड कोई टंकण का टूल है? या यूनिकोड कोई हिंदी या भारतीय भाषाओं में टंकण करने का तरीका है?

यूनिकोड एक टेक्नोलॉजी मानक है। यूनिकोड मानक में विश्वस्तर पर एवं प्रचलित सभी लिपियों के वर्णमाला के प्रत्येक अक्षर के लिए यूनिक कोड प्रदान किया गया है।

यूनिकोड (Unicode) प्रत्येक अक्षर के लिए एक विशेष संख्या प्रदान करता है, चाहे कोई भी कम्प्यूटर प्लेटफॉर्म, प्रोग्राम अथवा कोई भी भाषा हो। यूनिकोड मानक को एपल, एच.पी., आई.बी.एम., माइक्रोसॉफ्ट, ऑरेकल, सैप, सन, यूनिसिस जैसी उद्योग की प्रमुख कम्पनियों और कई अन्य ने अपनाया है। यूनिकोड आई.एस.ओ/आई.ई.सी. 10646 (ISO/IEC 10646) एक अंतर्राष्ट्रीय मानक है। यह कई संचालन प्रणालियों, सभी आधुनिक ब्राउजरों और कई अन्य उत्पादों में उपलब्ध है।

यूनिकोड 10.0 वर्जन में कुल 136,690 वर्णों को जोड़ा गया है, कुल 139 स्क्रिप्ट।

भारतीय भाषाओं के लिए यूनिकोड एनकोडिंग के लिए UTF-8 का प्रयोग होता है।

2. यूनिकोड क्यों?

यूनिकोड मानक सार्विक कैरेक्टर इनकोडिंग मानक है जिसका प्रयोग कंप्यूटर प्रोसेसिंग के लिए टेक्स्ट के निरूपण के लिए किया जाता है। कंप्यूटर पर एकरूपता के लिए एकमात्र विकल्प कैरेक्टर इनकोडिंग के लिए यूनिकोड है। इससे हिंदी तथा अन्य भारतीय भाषाओं में कंप्यूटर पर अंग्रेजी की तरह ही सरलता से 100% कार्य किया जा सकता है, कंप्यूटर पर हिंदी में सभी कार्य जैसे – वर्ड प्रोसेसिंग, डाटा प्रोसेसिंग, ई-मेल, वैबसाइट निर्माण आदि किए जा सकते हैं, हिंदी में बनी फाइलों का आसानी से आदान-प्रदान तथा हिंदी की-वर्ड पर गुगल या किसी अन्य सर्च इंजन पर सर्च कर सकते हैं।

राजभाषा विभाग ने अंतरराष्ट्रीय स्तर पर एनकोडिंग की एकरूपता को ध्यान में रखते हुए सभी केंद्रीय कार्यालय को कंप्यूटरों में यूनिकोड एनकोडिंग प्रणाली अथवा यूनिकोड समर्थित ओपन टाइप फॉन्ट का ही प्रयोग करने का निदेश दिया है। परंतु, कंप्यूटर परिचालन से संबंधित छोटी छोटी जानकारी के अभाव में कई केंद्रीय कार्यालय इस निःशुल्क सुविधा की जगह विभिन्न प्रकार के फॉन्ट और बहुभाषी सॉफ्टवेयरों का प्रयोग कर रहे हैं, जिससे सूचना हस्तांतरण में तकनीकी कठिनाइयों का सामना करना पड़ता है। इस कारण हिंदी की फाइलों

को, अंग्रेजी की तरह आसानी से एक कंप्यूटर से दूसरे कंप्यूटर पर, आदान-प्रदान (transfer) नहीं कर पाते हैं। हिंदी पाठ (text) को दूसरे सॉफ्टवेयर में जोड़ने (paste) में भी समस्या आती है। अतः सभी मंत्रालय एवं अधीनस्थ कार्यालय/उपक्रम/सरकारी बैंक केवल यूनिकोड समर्थित फॉन्ट एवं यूनिकोड एनकोडिंग के अनुरूप सॉफ्टवेयर का ही प्रयोग करें। यूनिकोड एनकोडिंग को install/use करना बहुत आसान है। इसकी जानकारी राजभाषा विभाग की साइट (<http://hinditools.nic.in>) पर भी उपलब्ध है।

3. यूनिकोड का महत्व तथा लाभ

- एक ही दस्तावेज में अनेकों भाषाओं के text लिखे जा सकते हैं।
- किसी सॉफ्टवेयर-उत्पाद का एक ही संस्करण पूरे विश्व में चलाया जा सकता है। क्षेत्रीय बाजारों के लिए अलग से संस्करण निकालने की जरूरत नहीं पड़ती।

4. देवनागरी यूनिकोड

- देवनागरी यूनिकोड की (रेंज) 0900 से 097F तक है। (दोनों संख्याएं षोडषाधारी हैं)
- क्ष, त्र एवं ज्ञ के लिये अलग से कोड नहीं है। इन्हें संयुक्त वर्ण मानकर अन्य संयुक्त वर्णों की भांति इनका अलग से कोड नहीं दिया गया है।
- इस रेंज में बहुत से ऐसे वर्णों के लिये भी कोड दिये गये हैं जो सामान्यतः हिन्दी में व्यवहृत नहीं होते। किन्तु मराठी, सिन्धी, मलयालम आदि को देवनागरी में सम्यक ढंग से लिखने के लिये आवश्यक हैं।
- नुक्ता के लिये भी अलग से एक कोड दे दिया गया है। अतः नुक्तायुक्त अक्षर यूनिकोड की दृष्टि से दो प्रकार से लिखे जा सकते हैं - एक बाइट यूनिकोड के रूप में या दो बाइट यूनिकोड के रूप में। उदाहरण के लिए ज़ को 'ज' के बाद नुक्ता (◌̣) टाइप करके भी लिखा जा सकता है।

UTF-8, UTF-16, UTF-32 क्या है?

- यूनिकोड का मतलब है सभी लिपि चिह्नों की आवश्यकता की पूर्ति करने में सक्षम 'एक समान मानकीकृत कोड'।
- पहले सोचा गया था कि केवल 16 बिट के माध्यम से ही दुनिया के सभी लिपि चिह्नों के लिये अलग-अलग कोड प्रदान किये जा सकेंगे। बाद में पता चला कि यह कम है। फिर इसे

32 बिट कर दिया गया। अर्थात इस समय दुनिया का कोई संकेत नहीं है जिसे 32 बिट के कोड में कहीं न कहीं जगह न मिल गयी हो।

- यूनिकोड के तीन रूप प्रचलित हैं - UTF-8, UTF-16 और UTF-32

- इनमें अन्तर क्या है? मान लीजिये आपके पास दस पेज का कोई टेक्स्ट है जिसमें रोमन, देवनागरी, अरबी, गणित के चिन्ह आदि बहुत कुछ हैं। इन चिन्हों के यूनिकोड कोड अलग-अलग होंगे। यहां ध्यान देने योग्य बात है कि कुछ संकेतों के 32 बिट के यूनिकोड में शुरू में शून्य ही शून्य हैं (जैसे अंग्रेजी के संकेतों के लिये)। यदि शुरुआती शून्यों को हटा दिया जाय तो इन्हें केवल 8 बिट के द्वारा भी निरूपित किया जा सकता है और कहीं कोई भ्रम या कांफ्लिक्ट नहीं होगा। इसी तरह रूसी, अरबी, हिब्रू आदि के यूनिकोड ऐसे हैं कि शून्य को छोड़ देने पर उन्हें प्रायः 16 बिट = 2 बाइट से निरूपित किया जा सकता है। देवनागरी, जापानी, चीनी आदि को आरम्भिक शून्य हटाने के बाद प्रायः 24 बिट = तीन बाइट से निरूपित किया जा सकता है। किन्तु बहुत से संकेत होंगे जिनमें आरम्भिक शून्य नहीं होंगे और उन्हें निरूपित करने के लिये चार बाइट ही लगेंगे।

- लगभग स्पष्ट है कि प्रायः UTF-8 में इनकोडिंग करने से UTF-16 की अपेक्षा कम बिट्स लगेंगे।

- इसके अलावा बहुत से पुराने सिस्टम 16 बिट को हैंडिल करने में अक्षम थे। वे एकबार में केवल 8-बिट ही के साथ काम कर सकते थे। इस कारण भी UTF-8 को अधिक अपनाया गया। यह अधिक प्रयोग में आता है।

- UTF-16 और UTF-32 के पक्ष में अच्छाई यह है कि अब कम्प्यूटरों का हार्डवेयर 32 बिट या 64 बिट का हो गया है। इस कारण UTF-8 की फाइलों को 'प्रोसेस' करने में UTF-16, UTF-32 वाली फाइलों की अपेक्षा अधिक समय लगेगा।

कंप्यूटरों में हिंदी में कार्य करने के लिए 3 की-बोर्ड विकल्प हैं:-

- इंस्क्रिप्ट
- रेमिंगटन
- फोनेटिक