



सत्यमेव जयते

जुलाई, 2017

कंप्यूटर पर हिंदी के प्रयोग के लिए उपलब्ध तकनीकी सुविधाएं

राजभाषा विभाग
गृह मंत्रालय
भारत सरकार

राजभाषा हिंदी के कार्यान्वयन में हिंदी ई-टूल्स की सार्थकता

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 ब्राउजरो और कई अन्य उत्पादों में उपलब्ध है।

- यूनिकोड 10.0 वर्जन में कुल 136,690 वर्णों को जोड़ा गया है तथा कुल 139 स्क्रिप्ट हैं।
- 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 गूगल या किसी अन्य सर्च इंजन पर सर्च कर सकते हैं।

राजभाषा विभाग ने अंतर्राष्ट्रीय स्तर पर एनकोडिंग की एकरूपता को ध्यान में रखते हुए सभी केंद्रीय कार्यालय को कंप्यूटर 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।

The latest electronic version of the Unicode Standard can be found at यूनिकोड साइट www.unicode.org, यूनिकोड कंसोर्सियम के प्रकाशनों में यूनिकोड मानक के साथ इसके अनुलग्नक और वर्ण शामिल हैं <http://www.unicode.org/ucd/>

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 हिंदी में व्यवहृत नहीं होते। किन्तु मराठी, सिंधी, 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।

5. UTF-8, UTF-16, UTF-32 क्या हैं?

- यूनिकोड का मतलब है सभी लिपि चिहनों 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 चिहनों 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 चिहनों 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 उन्हें 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 के ही 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।

6. कंप्यूटर में हिंदी में कार्य करने के लिए 3 की-बोर्ड विकल्प हैं:-

- इंस्क्रिप्ट
- रेमिंगटन
- फोनेटिक

7. फॉन्ट व टाइपिंग संबंधी समस्याएं एवं उनका निवारण

कंप्यूटर पर हिंदी में कार्य करने के लिए तीन की-बोर्ड विकल्प हैं-रेमिंगटन, इंस्क्रिप्ट तथा फोनेटिक। हालांकि अभी तक रेमिंगटन की-बोर्ड, पूर्व से प्रचलित होने के कारण (popular) है, तुलना में इंस्क्रिप्ट में हिंदी टंकण सीखना बहुत आसान है। इंस्क्रिप्ट ले-आउट भारत सरकार का मानक होने की वजह से सभी ऑपरेटिंग सिस्टम में डिफाल्ट में, यानि पहले से मौजूद रहता है। साथ ही किसी भी एक भाषा में इंस्क्रिप्ट की-बोर्ड सीखने पर, सभी भारतीय भाषाओं में आसानी से टंकण कर सकते हैं। (<http://ildc.in>) पर इंस्क्रिप्ट सीखने के लिए ट्यूटोर भी उपलब्ध है।

उल्लेखनीय है कि फोनेटिक की-बोर्ड में, हिंदी टंकण से अनभिज्ञ अधिकारी, रोमन स्क्रिप्ट का उपयोग करते हुए, हिंदी में आसानी से टंकण कर सकते हैं।

सभी ऑपरेटिंग सिस्टम में हिंदी का डिफॉल्ट फॉन्ट मंगल हैं। इसके साथ-साथ Arial Unicode MS, Kokila, Aparajita, Utsah, NirmalaUI फॉन्ट का भी प्रयोग कर सकते हैं।

<http://ildc.in> साइट पर 160 ओपन टाइप फॉन्ट उपलब्ध कराए गए हैं, इनका भी प्रयोग किया जा सकता है।

8. नॉन-यूनिकोड हिंदी सामग्री को यूनिकोड सामग्री में परिवर्तित करना

यह टूल एक फॉन्ट में लिखे गए डाटा को दूसरे फॉन्ट में बदलता है। यह कई तरह की फाइलों पर इस्तेमाल किया जा सकता है जैसे-माइक्रोसॉफ्ट वर्ड, माइक्रोसॉफ्ट एक्सेल, टैक्सट फाइल इत्यादि।

सभी प्रकार के स्टोरेज एवं फॉन्ट कन्वर्टर डाउनलोड करने के लिए <http://bhashaindia.com> से TBIL Converter 32-bit 4.1 तथा TBIL Converter 64-bit 4.1 डाउनलोड कर सकते हैं। इस पैकेज के माध्यम से भी नॉन-यूनिकोड सामग्री को यूनिकोड आधारित मंगल फॉन्ट में बदला जा सकता है।

9. गूगल वाइस टाइपिंग

अपनी आवाज के साथ टाइप करें, जिसे आसान तरीके से दस्तावेज़ में अपनी आवाज के साथ टाइप कर सकते हैं।

फिलहाल, यह सुविधा "क्रोम ब्राउज़र" में ही उपलब्ध है।

- सबसे पहले यह सुनिश्चित करें कि आपके कंप्यूटर से एक माइक्रोफोन जुड़ा हुआ है और वह काम करता है तथा एक जी-मेल का यूजर आईडी-पासवर्ड होना जरूरी है।
- Chrome ब्राउज़र में <http://docs.google.com> ओपन करें। जी-मेल आईडी से लॉग इन करें
- गूगल डॉक्स में एक नया दस्तावेज़ खोलें।
- उपकरण (Tools) मेनू > वाइस टाइपिंग (Voice Typing) पर क्लिक करें। पॉप-अप माइक्रोफोन बॉक्स से भाषा (हिंदी) का चयन करें।
- आप पढ़कर बोलने के लिए तैयार हैं, तो माइक्रोफोन बॉक्स पर क्लिक करें।
- सामान्य गति और वोल्यूम से स्पष्ट रूप से अपना पाठ बोलें।
- रोकने के लिए माइक्रोफोन पर पुनः क्लिक करें

वाइस टाइपिंग की गलतियों में सुधार

आवाज के साथ टाइप करते हुए अगर गलती हो जाए तो गलती पर कर्सर ले जाकर और माइक्रोफोन से पुनः बोल कर ठीक कर सकते हैं। गलती सुधारने के बाद, यदि आप आवाज टाइपिंग जारी रखना चाहते हैं तो वहां कर्सर वापस ले जाएं।

गूगल वाइस टाइपिंग वीडियो - <http://hinditools.nic.in> में गूगल वाइस टाइपिंग के अंतर्गत देखा जा सकता है।

10. हिंदी स्वयं शिक्षण - लीला-प्रबोध, प्रवीण तथा प्राज्ञ पाठ्यक्रम

- लीला (LILA-Learn Indian Languages through Artificial intelligence) स्वयं शिक्षण मल्टीमीडिया पैकेज है।
- <http://hinditools.nic.in> पर निःशुल्क उपलब्ध
- प्रबोध, प्रवीण और प्राज्ञ स्तर के पाठ्यक्रम (अंग्रेजी, कन्नड़, मलयालम, तमिल, तेलगू, बंगला, असमी, उड़िया, मणिपुरी, मराठी, पंजाबी, कश्मीरी, गुजराती, नेपाली तथा बोडो माध्यम से)
- ऑनलाइन अभ्यास, उच्चारण सुधार, स्वमूल्यांकन आदि की सुविधा।

11. मशीन अनुवाद

क. भारत सरकार के इलेक्ट्रॉनिकी और सूचना प्रौद्योगिकी मंत्रालय के प्रावधान के अंतर्गत Machine Assisted Translation Tool (Tourism, Health & Agriculture domain) www.tdil-dc.gov.in

ख. मंत्र-राजभाषा एक मशीन-साधित अनुवाद सिस्टम है, जो राजभाषा के प्रशासनिक, वित्तीय, कृषि, लघु उद्योग, सूचना प्रौद्योगिकी, स्वास्थ्य रक्षा, शिक्षा एवं बैंकिंग क्षेत्रों के दस्तावेजों का अंग्रेजी से हिंदी में अनुवाद करता है। मंत्र टैक्नालाजी पर आधारित यह सिस्टम राजभाषा विभाग द्वारा सी-डैक, पुणे के एप्लाइड आर्टिफिशियल इंटेलिजेंस ग्रुप की मदद से <https://mantra-rajbhasha.rb-aai.in> तथा <http://hinditools.nic.in> के माध्यम से विकसित कराया गया है।

ग. गूगल अनुवाद

www.translate.google.com गूगल अनुवाद fast एवं General Purpose है।

www.translate.google.com/toolkit गूगल में अकाउंट बनाकर अनुवाद करने पर, मेमोरी में ले लेता है जिससे भविष्य में similar text आने पर सही अनुवाद करता है।

12. ई-महाशब्दकोश

ई-महाशब्दकोश (द्विभाषी एवं द्विआयामी उच्चारण सहित शब्दकोश निःशुल्क <http://hinditools.nic.in> पर उपलब्ध है।

13. मोबाइल एप्प

Android based मोबाइल तथा टैबलेट में अनुवाद सॉफ्टवेयर

- Play Store में जाकर Google Translate डाउनलोड करके इंस्टाल करें।
- इससे आप अंग्रेजी से हिंदी अनुवाद कर सकते हैं तथा अंग्रेजी में प्राप्त SMS को हिंदी में भी पढ़ सकते हैं।
- JPG Format picture में यदि कोई अंग्रेजी टैक्सट हो तो उसे Capture करके हिंदी अनुवाद कर सकते हैं।
- किसी अंग्रेजी में छपे हुए टैक्सट को text में कनवर्ट करके हिंदी में अनुवाद कर सकते हैं।

Google Docs पर कार्य करना

Play Store से Google Docs डाउनलोड करके इंस्टाल करें। मोबाइल फोन में Google Docs पर कार्य तथा कंप्यूटर में एक की गूगल account होने पर Google Docs में फाइलें एक समान ही रहेंगी। फाइलों पर कार्य आप मोबाइल तथा कंप्यूटर पर आसानी से कर सकेंगे।

मोबाइल फोन पर हिंदी टाइपिंग के लिए

1. Play Store में जाकर Google Indic Keyboard डाउनलोड करके इंस्टाल करें।
2. Setting >> Language and Input में जाकर Keyboard & Input Method में से Current Keyboard or Default Keyboard में से English & Indic Languages Google Indic Keyboard को चुनें।

मोबाइल फोन पर गूगल वाइस टाइपिंग के लिए

1. Setting >> Language and Input Option में से Google Voice Typing विकल्प का चयन करें।
2. Language: हिंदी (भारत) का चयन करें। एक समय पर एक ही भाषा का चयन कर सकते हैं।
3. Offline Speech Recognition का भी चयन करें All पर क्लिक करके हिंदी (भारत) को Download करें।

यदि मोबाइल फोन में Android 7.0 OS है तो Setting >> Language and Input >> Virtual Keyboard के अंतर्गत Google Voice Typing विकल्प को चुनें।

4. जब भी टाइप करना हो की-बोर्ड पर उपलब्ध माइक्रोफोन के बटन पर क्लिक करें और सामान्य गति और वोल्यूम से स्पष्ट रूप से अपना पाठ बोलें, इससे आप SMS, WhatsApp, E-mail, Google Docs आदि पर वाइस टाइपिंग कर सकेंगे।

**राजभाषा विभाग
गृह मंत्रालय
भारत सरकार**