

# यूनिकोड

<http://hi.wikipedia.org/s/g83>

मुक्त ज्ञानकोश विकिपीडिया से  
यहाँ जाएँ: [भ्रमण](#), [खोज](#)

**यूनिकोड** (Unicode), प्रत्येक अक्षर के लिए एक विशेष संख्या प्रदान करता है, चाहे कोई भी कम्प्यूटर प्लेटफॉर्म, [प्रोग्राम](#) अथवा कोई भी भाषा हो। यूनिकोड स्टैंडर्ड को [एपल](#), [एच.पी.](#), [आई.बी.एम.](#), [जस्ट सिस्टम](#), [माइक्रोसॉफ्ट](#), [ऑरेकल](#), [सैप](#), [सन](#), साईबेस, यूनिसिस जैसी उद्योग की प्रमुख कम्पनियों और कई अन्य ने अपनाया है। यूनिकोड की आवश्यकता आधुनिक मानदंडों, जैसे [एक्स.एम.एल](#), [जावा](#), एकमा स्क्रिप्ट ([जावास्क्रिप्ट](#)), एल.डी.ए.पी., [कोर्बा 3.0](#), [डब्ल्यू.एम.एल](#) के लिए होती है और यह [आई.एस.ओ/आई.ई.सी. 10646](#) को लागू करने का अधिकारिक तरीका है। यह कई संचालन प्रणालियों, सभी आधुनिक [ब्राउजरों](#) और कई अन्य उत्पादों में होता है। यूनिकोड स्टैंडर्ड की उत्पत्ति और इसके सहायक उपकरणों की उपलब्धता, हाल ही के अति महत्वपूर्ण विश्वव्यापी [सॉफ्टवेयर प्रौद्योगिकी](#) रुझानों में से हैं।

यूनिकोड को [ग्राहक-सर्वर](#) अथवा बहु-आयामी उपकरणों और [वेबसाइटों](#) में शामिल करने से, परंपरागत उपकरणों के प्रयोग की अपेक्षा खर्च में अत्यधिक बचत होती है। यूनिकोड से एक ऐसा अकेला सॉफ्टवेयर उत्पाद अथवा अकेला वेबसाइट मिल जाता है, जिसे री-इंजीनियरिंग के बिना विभिन्न प्लेटफॉर्मों, भाषाओं और देशों में उपयोग किया जा सकता है। इससे आँकड़ों को बिना किसी बाधा के विभिन्न प्रणालियों से होकर ले जाया जा सकता है।

## अनुक्रम

[[छुपाएँ](#)]

- [1 यूनिकोड क्या है?](#)
- [2 यूनिकोड की विशेषताएँ](#)
- [3 यूनिकोड का महत्व और लाभ](#)
  - [3.1 हानियाँ](#)
- [4 देवनागरी यूनिकोड](#)

- [5 यूनिकोड कन्सॉर्शियम](#)
- [6 UTF-8, UTF-16 तथा UTF-32](#)
- [7 उपयोगी यूनिकोड औजार \(टूल्स\)](#)
  - [7.1 फॉण्ट परिवर्तक](#)
  - [7.2 जंक \(विकृत\) यूनिकोड को सही करने के उपाय](#)
- [8 संदर्भ](#)
- [9 इन्हें भी देखें](#)
- [10 बाहरी कड़ियाँ](#)
  - [10.1 यूनिकोड उपकरण तथा फॉण्ट](#)

## यूनिकोड क्या है?[संपादित करें]

यूनिकोड प्रत्येक अक्षर के लिए एक विशेष नम्बर प्रदान करता है,

- चाहे कोई भी प्लैटफॉर्म हो,
- चाहे कोई भी प्रोग्राम हो,
- चाहे कोई भी भाषा हो।

कम्प्यूटर, मूल रूप से, नंबरों से सम्बंध रखते हैं। ये प्रत्येक अक्षर और वर्ण के लिए एक नंबर निर्धारित करके अक्षर और वर्ण संग्रहित करते हैं। यूनिकोड का आविष्कार होने से पहले, ऐसे नंबर देने के लिए सैंकड़ों विभिन्न [संकेत लिपि प्रणालियां](#) थीं। किसी एक संकेत लिपि में पर्याप्त अक्षर नहीं हो सकते हैं : उदाहरण के लिए, [यूरोपीय संघ](#) को अकेले ही, अपनी सभी भाषाओं को कवर करने के लिए अनेक विभिन्न संकेत लिपियों की आवश्यकता होती है। अंग्रेजी जैसी भाषा के लिए भी, सभी अक्षरों, विरामचिह्नों और सामान्य प्रयोग के तकनीकी प्रतीकों हेतु एक ही संकेत लिपि पर्याप्त नहीं थी।

ये संकेत लिपि प्रणालियां परस्पर विरोधी भी हैं। इसीलिए, दो संकेत लिपियां दो विभिन्न अक्षरों के लिए, एक ही नंबर प्रयोग कर सकती हैं, अथवा समान अक्षर के लिए विभिन्न नम्बरों का प्रयोग कर सकती हैं। किसी भी कम्प्यूटर (विशेष रूप से सर्वर) को विभिन्न संकेत लिपियां संभालनी पड़ती है; फिर भी जब दो विभिन्न संकेत लिपियों अथवा [प्लैटफॉर्मों](#) के बीच [डाटा](#) भेजा जाता है तो उस डाटा के हमेशा खराब होने का जोखिम रहता है।

**यूनिकोड से यह सब कुछ बदल रहा है।**

## यूनिकोड की विशेषताएँ

- १) यह विश्व की सभी लिपियों से सभी संकेतों के लिए एक अलग कोड बिन्दु प्रदान करता है।
- २) यह वर्णों (कैरेक्टर्स) को एक कोड देता है, न कि ग्लिफ (glyph) को।
- ३) जहाँ भी सम्भव यूनिकोड होता है, यह भाषाओं का एकीकरण करने का प्रयत्न करता है। इसी नीति के तहत सभी पश्चिम यूरोपीय भाषाओं को [लैटिन](#) के अन्तर्गत समाहित किया गया है; सभी स्लाविक भाषाओं को सिरिलिक (Cyrilic) के अन्तर्गत रखा गया है; हिन्दी, संस्कृत, मराठी, नेपाली, सिन्धी, कश्मीरी आदि के लिए 'देवनागरी' नाम से एक ही ब्लॉक दिया गया है; चीनी, जापानी, कोरियाई, वियतनामी भाषाओं को 'युनिहान्' (UniHan) नाम से एक ब्लॉक में रखा गया है; अरबी, फारसी, उर्दू आदि को एक ही ब्लॉक में रखा गया है।
- ४) बाएँ से दाएँ लिखी जाने वाली लिपियों के अतिरिक्त दाएँ-से-बाएँ लिखी जाने वाली लिपियों (अरबी, हिब्रू आदि) को भी इसमें शामिल किया गया है। उपर से नीचे की तरफ लिखी जाने वाली लिपियों का अभी अध्ययन किया जा रहा है।
- ५) यह ध्यान रखना जरूरी है कि यूनिकोड केवल एक कोड-सारणी है। इन लिपियों को लिखने/पढ़ने के लिए [इनपुट मेथड एडिटर](#) और [फॉण्ट-फाइलें](#) जरूरी हैं।
- ५) यूनिकोड १६ बिट्स को एक ईकाई के रूप में लेकर चलता है।

## यूनिकोड का महत्व और लाभ[\[संपादित करें\]](#)

- एक ही दस्तावेज में अनेकों भाषाओं के टेक्स्ट लिखे जा सकते हैं।
- टेक्स्ट को केवल एक निश्चित तरीके से संस्कारित करने की जरूरत पड़ती है जिससे विकास-खर्च एवं अन्य खर्चे कम लगते 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
- किसी भी भाषा का टेक्स्ट पूरे संसार में बिना भ्रष्ट हुए चल जाता है। पहले इस तरह की बहुत समस्याएं आती थीं।

## हानियाँ[संपादित करें]

यूनिकोड, [आस्की](#) तथा अन्य [कैरेक्टर कोडों](#) की अपेक्षा अधिक स्मृति (मेमोरी) लेता है। कितनी अधिक स्मृति लगेगी यह इस बात पर निर्भर करता है कि आप कौन सा यूनिकोड प्रयोग कर रहे हैं। UTF7, UTF8, UTF16 या वास्तविक यूनिकोड - एक अक्षर अलग-अलग बाइट प्रयोग करते हैं।

## देवनागरी यूनिकोड[संपादित करें]

- देवनागरी यूनिकोड की परास (रेंज) 0900 से 097F तक है। (दोनों संख्याएं षोडशाधारी 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 वाले वर्णों (जैसे ज्ञ) के लिये यूनिकोड निर्धारित किया गया है। इसके अलावा नुक्ता के लिये भी अलग से एक यूनिकोड 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 लिये ज्ञ को 'ज ' के बाद नुक्ता ( ं ) टाइप करके भी लिखा जा सकता है।

### देवनागरी के यूनिकोड

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U+090x		ँ	ं	ः	ओ	अ	आ	इ	ई	उ	ऊ	ऋ	ॠ	एँ	ऐ	ए
U+091x	ऐ	ऑ	ओ	औ	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट	
U+092x	ठ	ड	ढ	ण	त	थ	द	ध	न	॒	प	फ	ब	भ	म	य

U+093x	र	र	ल	ळ	ळ	व	श	ष	स	ह			़	s	ा	ि
U+094x	ी	ु	ू	ृ	ृ	ँ	े	ै	ॉ	ो	ो	ौ	्			
U+095x	ॐ	ं	ं	े	े				क	ख	ग	ज	ड	ढ	फ	य
U+096x	ऋ	ॠ	ॡ	ॢ	।	॥	०	१	२	३	४	५	६	७	८	९
U+097x	॰		अँ									ग	ज		ड	ब

## यूनिकोड कन्सॉर्शियम[संपादित करें]

[यूनिकोड कन्सॉर्शियम](#), एक लाभ न कमाने वाला एक संगठन है जिसकी स्थापना [यूनिकोड स्टैंडर्ड](#), जो आधुनिक सॉफ्टवेयर उत्पादों और मानकों में पाठ की प्रस्तुति को निर्दिष्ट करता है, के विकास, विस्तार और इसके प्रयोग को बढ़ावा देने के लिए की गई थी। इस कन्सॉर्शियम के सदस्यों में, कम्प्यूटर और सूचना उद्योग में विभिन्न निगम और संगठन शामिल हैं। इस कन्सॉर्शियम का वित्तपोषण पूर्णतः सदस्यों के शुल्क से किया जाता है। यूनिकोड कन्सॉर्शियम में सदस्यता, विश्व में कहीं भी स्थित उन संगठनों और व्यक्तियों के लिए खुली है जो यूनिकोड का समर्थन करते हैं और जो इसके विस्तार और कार्यान्वयन में सहायता करना चाहते हैं।

## UTF-8, UTF-16 तथा UTF-32[संपादित करें]

1. यूनिकोड का मतलब है सभी लिपिचिह्नों की आवश्यकता की पूर्ति करने में सक्षम 'एकसमान मानकीकृत कोड'।
2. पहले सोचा गया था कि केवल १६ बिट के माध्यम से ही दुनिया के सभी लिपिचिह्नों के लिये अलग-अलग कोड प्रदान किये जा सकेंगे। बाद में पता चला कि यह कम है। फिर इसे ३२ बिट कर दिया गया। अर्थात् इस समय दुनिया का कोई संकेत नहीं है जिसे ३२ बिट के कोड में कहीं न कहीं जगह न मिल गयी हो।
3. ८ बिट के कुल २पर घात ८ = २५६ अलग-अलग बाइनरी संख्याएं बन सकती हैं; १६ बिट से २ पर घात १६ = ६५५३६ और ३२ बिट से ४२९४९६७२९६ भिन्न (distinct) बाइनरी संख्याएं बन सकती हैं।

4. यूनिकोड के तीन रूप प्रचलित हैं। UTF-8, UTF-16 और UTF-32.
5. 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 अलग-अलग होंगे। यहाँ ध्यान देने योग्य बात है कि कुछ संकेतों के ३२ 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 ८ बिट के द्वारा भी निरूपित किया जा सकता है और कहीं कोई भ्रम या कांफ़्लिक्ट नहीं होगा। इसी तरह रूसी, अरबी, हिब्रू आदि के यूनिकोड ऐसे हैं कि शून्य को छोड़ देने पर उन्हें प्रायः १६ बिट = २ बाइट से निरूपित किया जा सकता है। देवनागरी, जापानी, चीनी आदि को आरम्भिक शून्य हटाने के बाद प्रायः २४ 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।
6. बुन्दु (५) में बताए गये काम को UTF-8, UTF-16 और UTF-32 थोड़ा अलग अलग ढंग से करते हैं। उदाहरण के लिये यूटीएफ-८ क्या करता है कि कुछ लिपिचिह्नों के लिये १ बाइट , कुछ के लिये २ बाइट, कुछ के लिये तीन और चार बाइट इस्तेमाल करता है। लेकिन UTF-16 इसी काम के लिये १६ न्यूनतम बिटों का इस्तेमाल करता है। अर्थात जो चीजें UTF-8 में केवल एक बाइट जगह लेती थीं वे अब १६ बिट==२ बाइट के द्वारा निरूपित होंगी। जो UTF-8 में २ बाइट लेती थी यूटीएफ-१६ में भी दो ही लेंगी। किन्तु पहले जो संकेतदि ३ बाइट या चार बाइट में निरूपित होते थे यूटीएफ-१६ में ३२ बिट=४ बाइट के द्वारा निरूपित किये जायेंगे। (आपके पास बड़ी-बड़ी ईंटें हों और उनको बिना तोड़े खम्भा बनाना हो तो खम्भा ज्यादा बड़ा ही बनाया जा सकता है। )
7. लगभग स्पष्ट है कि प्रायः UTF-8 में इनकोडिंग करने से UTF-16 की अपेक्षा कम बिट्स लगेंगे।
8. इसके अलावा बहुत से पुराने सिस्टम १६ बिट को हैंडिल करने में अक्षम थे। वे एकबार में केवल ८-बिट ही के साथ काम कर सकते थे। इस कारण भी UTF-8 को अधिक अपनाया गया। यह अधिक प्रयोग में आता है।
9. UTF-16 और UTF-32 के पक्ष में अच्छाई यह है कि अब कम्प्यूटरों का हार्डवेयर ३२ बिट या ६४ 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।

**उपयोगी यूनिकोड औजार (टूल्स)[[संपादित करें](#)]**