

Original by Markus Kuhn, adapted for HTML by Martin Dürst.

UTF-8 encoded sample plain-text file

Markus Kuhn [`'ma'ku:s ku:n`] <mkuhn@acm.org> – 1999-08-20

The ASCII compatible UTF-8 encoding of ISO 10646 and Unicode plain-text files is defined in RFC 2279 and in ISO 10646-1 Annex R.

Using Unicode/UTF-8, you can write in emails and source code things such as
Mathematics and Sciences:

$\oint E \cdot da = Q, \quad n \rightarrow \infty, \quad \sum f(i) = \prod g(i), \quad \forall x \in \mathbb{R}: \lceil x \rceil = -\lfloor -x \rfloor, \quad \alpha \wedge \neg \beta = \neg(\neg \alpha \vee \beta),$
 $\mathbb{N} \subseteq \mathbb{N}_0 \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}, \quad \perp < a \neq b \equiv c \leq d \ll T \Rightarrow (A \Leftrightarrow B),$
 $2\text{H}_2 + \text{O}_2 \rightleftharpoons 2\text{H}_2\text{O}, \quad R = 4.7 \text{ k}\Omega, \quad \varnothing 200 \text{ mm}$

Linguistics and dictionaries:

`ði intə'næfənəl fə'nɛtɪk əsoʊsi'eɪʃn`
`Y ['ypsɪlɒn], Yen [jɛn], Yoga ['jo:gɑ]`

APL:

`((V∨V)=ιρV)/V←,V ⌈←L←ρΔ∇▷⌊⌊⌊⌈`

Nicer typography in plain text files:

- ‘single’ and “double” quotes
- Curly apostrophes: “We’ve been here”
- Latin-1 apostrophe and accents: ‘`
- ,deutsche‘ „Anführungszeichen“
- †, ‡, ‰, •, 3-4, −, −5/+5, ™, ...
- ASCII safety test: 1lI|, 00D, 8B
- the euro symbol:

14.95 €

Greek (in Polytonic):

The Greek anthem:

Σὲ γνωρίζω ἀπὸ τὴν κόψη
τοῦ σπαθιοῦ τὴν τρομερή,
σὲ γνωρίζω ἀπὸ τὴν ὄψη
ποὺ μὲ βία μετράει τὴ γῆ.

Ἀπ’ τὰ κόκκαλα βγαλμένα
τῶν Ἑλλήνων τὰ ἱερά
καὶ σὰν πρῶτα ἀνδρειωμένα

χαῖρε, ὦ χαῖρε, Ἐλευθεριά!

From a speech of Demosthenes in the 4th century BC:

Οὐχὶ ταῦτ᾽ ἀρίστα μοι γινώσκεις, ὦ ἄνδρες Ἀθηναῖοι, ὅταν τ' εἰς τὰ πράγματα ἀποβλέψω καὶ ὅταν πρὸς τοὺς λόγους οὐκ ἀκούω· τοὺς μὲν γὰρ λόγους περὶ τοῦ τιμωρῆσθαι Φίλιππον ὁρῶ γιννομένους, τὰ δὲ πράγματ' εἰς τοῦτο προήκοντα, ὥσθ' ὅπως μὴ πεισόμεθ' αὐτοῖς πρότερον κακῶς σκέψασθαι δέον. οὐδὲν οὖν ἄλλο μοι δοκοῦσιν οἱ τὰ τοιαῦτα λέγοντες ἢ τὴν ὑπόθεσιν, περὶ ἧς βουλευέσθαι, οὐχὶ τὴν οὖσαν παριστάντες ὑμῖν ἀμαρτάνειν. ἐγὼ δέ, ὅτι μὲν ποτ' ἐξῆν τῇ πόλει καὶ τὰ αὐτῆς ἔχειν ἀσφαλῶς καὶ Φίλιππον τιμωρῆσθαι, καὶ μάλ' ἀκριβῶς οἶδα· ἐπ' ἐμοῦ γάρ, οὐ πάλα γέγονεν ταῦτ' ἀμφοτέρω· νῦν μέντοι πέπεισμαι τοῦθ' ἱκανὸν προλαβεῖν ἡμῖν εἶναι τὴν πρώτην, ὅπως τοὺς συμμάχους σώσομεν. ἐὰν γὰρ τοῦτο βεβαίως ὑπάρξῃ, τότε καὶ περὶ τοῦ τίνα τιμωρήσεται τις καὶ ὃν τρόπον ἐξέσται σκοπεῖν· πρὶν δὲ τὴν ἀρχὴν ὀρθῶς ὑποθέσθαι, μάταιον ἡγοῦμαι περὶ τῆς τελευτῆς ὀντινοῦν ποιεῖσθαι λόγον.

Δημοσθένους, Γ' Ὀλυνθιακός

Georgian:

From a Unicode conference invitation:

გთხოვთ ახლავე გაიართო რეგისტრაცია Unicode-ის მეათე საერთაშორისო კონფერენციაზე დასასწრებად, რომელიც გაიმართება 10-12 მარტს, ქ. მაინცში, გერმანიაში. კონფერენცია შეჰკრებს ერთად მსოფლიოს ექსპერტებს ისეთ დარგებში როგორიცაა ინტერნეტი და Unicode-ი, ინტერნაციონალიზაცია და ლოკალიზაცია, Unicode-ის გამოყენება ოპერაციულ სისტემებსა, და გამოყენებით პროგრამებში, შრიფტებში, ტექსტების დამუშავებასა და მრავალენოვან კომპიუტერულ სისტემებში.

Russian:

From a Unicode conference invitation:

Зарегистрируйтесь сейчас на Десятую Международную Конференцию по Unicode, которая состоится 10-12 марта 1997 года в Майнце в Германии. Конференция соберет широкий круг экспертов по вопросам глобального Интернета и Unicode, локализации и интернационализации, воплощению и применению Unicode в различных операционных системах и программных приложениях, шрифтах, верстке и многоязычных компьютерных системах.

Ethiopian:

Proverbs in the Amharic language:

ሰማይ አይታረስ ንጉሥ አይከሰስ።
ብላ ካለኝ አንደአባቴ በቆመጠኝ።
ጌጥ ያለቤቴ ቁምጥና ነው።
ደህ በሕልሙ ቅቤ ባይጣጣ ንግት በገደለው።
የአፍ ወለምታ በቅቤ አይታሽም።
አይጥ በበላ ዳዋ ተመታ።
ሲተረጉሙ ይደረግሙ።
ቀስ በቀስ፥ ዕንቁላል በእግሩ ይሄዳል።
ድር ቢያብር አንበሳ ያስር።
ሰው አንደቤቴ አንጅ አንደ ጉረቤቴ አይተዳደርም።
እግዚር የከፈተውን ጉሮሮ ሳይዘጋው አይድርም።
የጎረቤት ሊባ፥ ቢያዩት ይስቅ ባያዩት ያጠልቅ።
ሥራ ከመፍታት ልጄን ላፋታት።

ግብጽ ማደሪያ የለው፥ ግንድ ይዞ ይዞራል።
 የእስላም እገሩ መካ የአሞራ እገሩ ዋርካ።
 ተጋጋሎ ቢተፋ ተመልሶ ባፋ።
 ወዳጅህ ማር ቢሆን ጨርስህ አትላሰው።
 እግርህን በፍራሽህ ልክ ዘርጋ።

Runes:

𐌆𐌆 𐌱𐌱𐌲𐌲 𐌲𐌴𐌰 𐌆𐌆 𐌲𐌆𐌆𐌆 𐌴𐌰 𐌲𐌴𐌆 𐌴𐌴𐌲𐌆𐌆𐌆 𐌲𐌴𐌲𐌴𐌆𐌆𐌆𐌆𐌆𐌆𐌆 𐌲𐌴𐌲 𐌲𐌴𐌲 𐌲𐌆𐌆𐌆

(Old English, which transcribed into Latin reads 'He cwaeth that he bude thaem lande northweardum with tha Westsae.' and means 'He said that he lived in the northern land near the Western Sea.')

Braille:

[illegible][illegible]

המחיר הממוצע של המכונית הוא 10,000 ש"ח.

[illegible]

(The first couple of paragraphs of "A Christmas Carol" by Dickens)

Compact font selection example text:

ABCDEFGHIJKLMNOPQRSTUVWXYZ /0123456789
abcdefghijklmnopqrstuvwxyz £©µÀÆÖßéöÿ
—‘’”„†•…™œŠŽ€ ΑΒΓΔΩαβγδω ΑΒΓΔαβγδ
∀∃∈ℝΛ∪≡∞ ↑↗↑↓ ↕⌈⋈⌋Ⓜ▶☺ fi◊ꞤñBīwē:

Greetings in various languages:

Hello world, Καλημέρα κόσμε, コンニチハ