

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

UTF-8 encoded sample plain-text file

Markus Kuhn ['markos 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:

□ $E \cdot da = 0$, $n \rightarrow \infty$, $\sum f(i) = \prod g(i)$, $\forall x \in \mathbb{R}$: $[x] = -[-x]$, $\alpha \wedge \neg\beta = \neg(\neg\alpha \vee \beta)$,

$\mathbb{N} \subseteq \mathbb{N}_0 \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$, $\perp < a \neq b \equiv c \leq d$ □ $T \Rightarrow (A \Leftrightarrow B)$,

$2H_2 + O_2 \Rightarrow 2H_2O$, $R = 4.7 \text{ k}\Omega$, $\emptyset 200 \text{ mm}$

Linguistics and dictionaries:

ði intə'næʃənəl fə'nɛtɪk ə'səʊsi'eɪʃn
Y ['Ypsilon], Yen [jεn], Yoga ['jo:gə]

APL:

((ViV)=1ρV)/V←,V ⌈←1→ρΔ∇∘⌈ΠΓ

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 ὅψη
 ποὺ μὲ βία μετράει τὴ γῆ.

Ἄπ' τὰ κόκκαλα βγαλμένη
 τῶν Ἑλλήνων τὰ ιερά¹
 καὶ σὰν πρῶτα ἀνδρειωμένη
 χαῖρε, ὦ χαῖρε, Ἐλευθεριά!

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

Οὐχὶ ταῦτὰ παρίσταται μοι γιγνώσκειν, ὡς ἄνδρες Ἀθηναῖοι,
 ὅταν τ’ εἰς τὰ πράγματα ἀποβλέψω καὶ ὅταν πρὸς τοὺς
 λόγους οὓς ἀκούω· τοὺς μὲν γὰρ λόγους περὶ τοῦ
 τιμωρήσασθαι φίλιππον ὁρῶ γιγνομένους, τὰ δὲ πράγματ’
 εἰς τοῦτο προήκοντα, ὥσθ’ ὅπως μὴ πεισόμεθ’ αὐτοὶ
 πρότερον κακῶς σκέψασθαι δέον. οὐδέν οὖν ἄλλο μοι δοκοῦσιν
 οἱ τὰ τοιαῦτα λέγοντες ἢ τὴν ὑπόθεσιν, περὶ ᾧς βουλεύεσθαι,
 οὐχὶ τὴν οὖσαν παριστάντες ὑμῖν ἀμαρτάνειν. ἐγὼ δέ, ὅτι μέν
 ποτ’ ἔξην τῇ πόλει καὶ τὰ αὐτῆς ἔχειν ἀσφαλῶς καὶ φίλιππον
 τιμωρήσασθαι, καὶ μάλ’ ἀκριβῶς οἶδα· ἐπ’ ἔμοι γάρ, οὐ πάλαι
 γέγονεν ταῦτ’ ἀμφότερα· νῦν μέντοι πέπεισμαι τοῦθ’ ἵκανὸν
 προλαβεῖν ἡμῖν εἶναι τὴν πρώτην, ὅπως τοὺς συμμάχους
 σώσομεν. ἐὰν γὰρ τοῦτο βεβαίως ὑπάρξῃ, τότε καὶ περὶ τοῦ
 τίνα τιμωρήσεται τις καὶ δὸν τρόπον ἔξεσται σκοπεῖν· πρὶν δὲ
 τὴν ἀρχὴν ὀρθῶς ὑποθέσθαι, μάταιον ἥγοῦμαι περὶ τῆς
 τελευτῆς ὃντινοῦν ποιεῖσθαι λόγον.

Δημοσθένους, Γ΄ Ὁλυνθιακὸς

Georgian:

From a Unicode conference invitation:

გთხოვთ ახდავე გაიაროთ რეგისტრაცია **Unicode**-ის მეათე საერთაშორისო კონფერენციაზე დასასწრებად, რომელიც გაიმართება 10-12 მარტს, ქ. მაინცში, გერმანიაში. კონფერენცია შეჰქრებს ერთად მსოფლიოს ექსპერტებს ისეთ დარგებში როგორიცაა ინტერნეტი და **Unicode**-ი, ინტერნაციონალური და კონსოლიდირებული კონფერენცია, როგორიცაა **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.

Thai (UCS Level 2):

Excerpt from a poetry on The Romance of The Three Kingdoms (a Chinese classic 'San Guo'):

The diagram illustrates two groups of horizontal bars, likely representing data blocks or memory locations. The left group contains 10 bars, and the right group contains 11 bars. Each bar is composed of a small square at its left end followed by a series of vertical segments. The bars are arranged in two columns, separated by a vertical line.

(The above is a two-column text. If combining characters are handled correctly, the lines of the second column should be aligned with the | character above.)

Ethiopian:

Proverbs in the Amharic language:

The image shows a rectangular grid of 100 empty square boxes. The grid is organized into 10 horizontal rows and 10 vertical columns. Each box is a simple black outline on a white background. The entire grid occupies most of the page, with some space at the top and bottom.

Runes:

A horizontal row of 12 vertical black bars of equal height, representing binary digits. The first bar is at the far left, and the last bar is at the far right. There are gaps between the bars.

(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:

The image shows a rectangular grid of 100 empty square boxes. The grid is organized into 10 horizontal rows and 10 vertical columns. Each box is a simple black outline on a white background. The entire grid is centered on the page.

The image shows a large grid of 100 empty square boxes arranged in 10 rows and 10 columns. The boxes are white with black outlines, and they are evenly spaced both horizontally and vertically. This grid represents a 10x10 matrix or a 10x10 table.

The diagram consists of two rows of rectangles. The top row contains seven groups of three rectangles each. The bottom row contains five groups of three rectangles each.

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

Compact font selection example text:

Greetings in various languages:

Hello world, Καλημέρα κόσμε, ☀️☀️☀️

Box drawing alignment tests:

