

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

UTF-8 encoded sample plain-text file

Markus Kuhn ['maʁkʊs 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:

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

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

$2\text{H}_2 + \text{O}_2 \rightleftharpoons 2\text{H}_2\text{O}, \quad R = 4.7 \text{ k}\Omega, \quad \varnothing 200 \text{ mm}$

Linguistics and dictionaries:

ði ɪntə'næʃənəl fə'netɪk əsoʊsi'eɪʃn
Y ['ɪpsɪlɒn], Yen [jɛn], Yoga ['jo:ɡɑ]

APL:

$((V \uparrow V) = \uparrow p V) / V \leftarrow, V \quad \square \leftarrow \uparrow \rightarrow p \Delta \nabla \supset \neg \square \square \uparrow$

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 οὖς ἀκούω· τοὺς μὲν γὰρ λόγους περὶ τοῦ
τιμωρῆσθαι Φίλιππον ὁρῶ γιγνομένους, τὰ δὲ πράγματ’
εἰς τοῦτο προήκοντα, ὥσθ’ ὅπως μὴ πεισόμεθ’ αὐτοῖ
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' ἔξῃν τῇ πόλει καὶ τὰ αὐτῆς ἔχειν ἀσφαλῶς καὶ Φίλιππον
τιμωρήσασθαι, καὶ μάλ' ἀκριβῶς οἶδα· ἐπ' ἐμοῦ γάρ, οὐ πάλαι
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 τιμωρήσεταιί τις καὶ ὃν τρόπον ἐξέσται σκοπεῖν· πρὶν δὲ
τὴν ἀρχὴν ὀρθῶς ὑποθέσθαι, μάταιον ἡγοῦμαι περὶ τῆς
τελευτῆς ὄντινοῦν ποιεῖσθαι λόγον.

Δημοσθένους, Γ' Ὀλυνθιακὸς

Georgian:

From a Unicode conference invitation:

გთხოვთ ახლავთ გაიაროთ რეგისტრაცია Unicode-ის მეათე საერთაშორისო
კონფერენციაზე დასასწრებად, რომელიც გაიმართება 10-12 მარტს,
ქ. მაინცში, გერმანიაში. კონფერენცია შეჰყრებს ერთად მსოფლიოს
ექსპერტებს ისეთ დარგებში როგორიცაა ინტერნეტი და Unicode-ი,
ინტერნაციონალიზაცია და ლოკალიზაცია, Unicode-ის გამოყენება
ოპერაციულ სისტემებში, 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 Gua'):

The diagram consists of two parallel horizontal tracks. Each track contains 10 rectangular blocks of varying lengths. The blocks are arranged in a way that suggests a sequence or process. The left track starts with a small block, followed by a long block, and then several blocks of intermediate length. The right track starts with a long block, followed by a medium block, and then several blocks of intermediate length. The blocks are arranged in a way that suggests a sequence or process.

1111 111111 1111 11111111
 11 111 1111111 11111111
 11 1111 1111 1111
 11 1111 11 1111 111 1111111
 111 1111 111 1111111
 111 111 11 1111
 11111 1111111
 11 1111 111111 1111 11111
 11 1111 1111 1111
 11 11111 111 111 1111 111111111
 1111 1111111 111 11111 1111111
 11111 111 1111 111 1111 11111
 11 11111 111 11111
 111 1111 1111 111 11 111111
 11111 111 11 1111 111 11111
 1111 111 1111 111
 1111 11 111 1111 1111111
 11111 111111 11 11111

Runes:

ᚷ ᚠᚢᚳᚦ ᚱᚢᚦ ᚱᚢ ᚠᚢᚳᚦ ᚱᚢ ᚱᚢᚦ ᚠᚢᚳᚦ ᚠᚢᚳᚦᚠᚢᚳᚦᚠᚢᚳᚦ ᚱᚢᚦ ᚱᚢ ᚠᚢᚳᚦ

(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:

ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ

ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ
ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ
ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠᠨᠠ
ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ
ᠠᠨᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ
ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ

ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠᠨᠠ

ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ
ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ
ᠠᠨᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ
ᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠᠨᠠ
ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ
ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠ
ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ
ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠᠨᠠᠨᠠ ᠠᠨᠠ
ᠠᠨᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠ ᠠᠨᠠᠨᠠᠨᠠᠨᠠ

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

Compact font selection example text:

ABCDEFGHIJKLMNOPQRSTUVWXYZ /0123456789
abcdefghijklmnopqrstuvwxyz £©μÀÆÖÞßéöÿ
—‘”„†•…‰™œŠŸŽ€ ΑΒΓΔΩαβγδω ΑΒΓΔαβγδ

∇∂∈ℝλ∪≡∞
 ↗↖↗↖↗↖↗↖
 7+⌈⌋⌈⌋⌈⌋⌈⌋
 2
 ♀
 fi02ήΒÿŴe:
 ☐☐☐
 ⋔

Greetings in various languages:

Hello world,
 Καλημέρα κόσμε,
 ☐☐☐☐

Box drawing alignment tests:

