

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

## UTF-8 encoded sample plain-text file

Markus Kuhn ['ma'kʊs 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:

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

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

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

Linguistics and dictionaries:

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

APL:

$((\forall iV)=\iota\rho V)/V\leftarrow,V \quad \square\leftarrow\iota\rightarrow\rho\Delta\nabla>\overline{\square}\square$

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  
ποῦ μὲ βία μετράει τὴ γῆ.

Ἄπ' τὰ κόκκαλα βγαλμένη  
τῶν Ἑλλήνων τὰ ἱερά  
καὶ σὰν πρῶτα ἀνδρειωμένη  
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 ὀρῶ 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 οὕσαν 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 ὄντινοῦν ποιεῖσθαι λόγον.

Δημοσθένους, Γ' Ὀλυμπιακός

Georgian:

From a Unicode conference invitation:

გთხოვთ ახდევთ გაიაროთ რეგისტრაცია Unicode-ის მეათე საერთაშორისო კონფერენციაზე დასასწრებად, რომელიც გაიმართება 10-12 მარტს, ქ. მაინცში, გერმანიაში. კონფერენცია შეჰკრებს ერთად მსოფლიოს ექსპერტებს ისეთ დარგებში როგორცაა 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 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:

ገበያ ገበያ ገበያ ገበያ  
ገበያ ገበያ ገበያ ገበያ

𐀀 𐀁 𐀂 𐀃 𐀄 𐀅  
 𐀆 𐀇 𐀈 𐀉 𐀊 𐀋 𐀌 𐀍  
 𐀎 𐀏 𐀐 𐀑 𐀒 𐀓 𐀔  
 𐀕 𐀖 𐀗 𐀘 𐀙 𐀚  
 𐀛 𐀜 𐀝 𐀞 𐀟  
 𐀠 𐀡 𐀢 𐀣 𐀤 𐀥 𐀦 𐀧  
 𐀨 𐀩 𐀪 𐀫 𐀬  
 𐀭 𐀮 𐀯 𐀰 𐀱 𐀲 𐀳 𐀴 𐀵  
 𐀶 𐀷 𐀸 𐀹 𐀺 𐀻 𐀼 𐀽  
 𐀾 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽  
 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿  
 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿  
 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿  
 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿  
 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿  
 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿  
 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿  
 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿  
 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿

Runes:

𐀀 𐀁 𐀂 𐀃 𐀄 𐀅 𐀆 𐀇 𐀈 𐀉 𐀊 𐀋 𐀌 𐀍 𐀎 𐀏 𐀐 𐀑 𐀒 𐀓 𐀔

(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:

⠠ 𐀀 𐀁 𐀂 𐀃 𐀄 𐀅 𐀆 𐀇 𐀈 𐀉 𐀊 𐀋 𐀌 𐀍

⠠ 𐀆 𐀇 𐀈 𐀉 𐀊 𐀋 𐀌 𐀍 𐀎 𐀏 𐀐 𐀑 𐀒 𐀓 𐀔 𐀕 𐀖 𐀗 𐀘 𐀙 𐀚  
 ⠠ 𐀛 𐀜 𐀝 𐀞 𐀟 𐀠 𐀡 𐀢 𐀣 𐀤 𐀥 𐀦 𐀧 𐀨 𐀩 𐀪 𐀫 𐀬  
 ⠠ 𐀭 𐀮 𐀯 𐀰 𐀱 𐀲 𐀳 𐀴 𐀵 𐀶 𐀷 𐀸 𐀹 𐀺 𐀻 𐀼 𐀽  
 ⠠ 𐀾 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽  
 ⠠ 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿  
 ⠠ 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿

⠠ 𐀀 𐀁 𐀂 𐀃 𐀄 𐀅 𐀆 𐀇 𐀈 𐀉 𐀊 𐀋 𐀌 𐀍 𐀎 𐀏 𐀐 𐀑 𐀒 𐀓 𐀔

⠠ 𐀆 𐀇 𐀈 𐀉 𐀊 𐀋 𐀌 𐀍 𐀎 𐀏 𐀐 𐀑 𐀒 𐀓 𐀔 𐀕 𐀖 𐀗 𐀘 𐀙 𐀚  
 ⠠ 𐀛 𐀜 𐀝 𐀞 𐀟 𐀠 𐀡 𐀢 𐀣 𐀤 𐀥 𐀦 𐀧 𐀨 𐀩 𐀪 𐀫 𐀬  
 ⠠ 𐀭 𐀮 𐀯 𐀰 𐀱 𐀲 𐀳 𐀴 𐀵 𐀶 𐀷 𐀸 𐀹 𐀺 𐀻 𐀼 𐀽  
 ⠠ 𐀾 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽  
 ⠠ 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿  
 ⠠ 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿  
 ⠠ 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿 𐀽 𐀿

□□□□ □□□□□□ □□□□□ □□ □□ □□□□□□ □□□□□□□□□□□□ □□□  
 □□□□□ □□□ □□ □□□□□ □□ □ □□□□□□□□□□□□□□□□

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

Compact font selection example text:

ABCDEFGHIJKLMNOPQRSTUVWXYZ /0123456789  
 abcdefghijklmnopqrstuvwxyz £©μÀÆÕÞβέöÿ  
 — ‘ ’ “ ” „ † • … ‰ ™ ª Š Ÿ Ž € Α Β Γ Δ Ω α β γ δ ω Α Β Γ Δ α β γ δ  
 √ ∂ ∈ ℝ λ ∪ ≡ ∞ ↑ ↗ ⚡ ∩ ↓ ∩ † † † ⚡ ▶ ◉ ♀ fi 0 □ 2 ħ ß ſ ∞ : □ □ □ ∩

Greetings in various languages:

Hello world, Καλημέρα κόσμε, □□□□

Box drawing alignment tests:

