

## UTF-8 encoded sample plain-text file

---

Markus Kuhn ['markun] <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:

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

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

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

Linguistics and dictionaries:

ði intə'næʃənəl fə'nɛtɪk əsəʊsi'eɪʃn  
Y ['Ypsɪlən], Yen [jɛn], Yoga ['jo:gə]

APL:

$((V \downarrow V) = \iota \rho V) / V \leftarrow, V \quad \Box \leftarrow \iota \rightarrow \rho \Delta \nabla \triangleright \Box \Box \Gamma$

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.

’Απ’ τὰ κόκκαλα βγαλμένη  
 τῶν ’Ελλήνων τὰ ιερά  
 καὶ σὰν πρῶτα ἀνδρειωμένη  
 χαῖρε, ὦ χαῖρε, ’Ελευθεριά!

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

Οὐχὶ ταῦτὰ παρίσταται μοι γιγνώσκειν, ὡς ἄνδρες ’Αθηναῖοι,  
 ὅταν τ’ εἰς τὰ πράγματα ἀποβλέψω καὶ ὅταν πρὸς τοὺς  
 λόγους οὖς ἀκούω· τοὺς μὲν γὰρ λόγους περὶ τοῦ  
 τιμωρήσασθαι φίλιππον ὄρῳ γιγνομένους, τὰ δὲ πράγματ’  
 εἰς τοῦτο προήκοντα, ὃσθ’ 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 οὗσαν παριστάντες ὑμῖν ἀμαρτάνειν. ἐγὼ δέ, ὅτι μέν

ποτ' ἔξῆν τῇ πόλει καὶ τὰ αὐτῆς ἔχειν ἀσφαλῶς καὶ φίλιππον τιμωρήσασθαι, καὶ μάλ' ἀκριβῶς οἶδα· ἐπ' ἔμοῦ γάρ, οὐ πάλαι γέγονεν ταῦτ' ἀμφότερα· νῦν μέντοι πέπεισμαι τοῦθ' ἰκανὸν προλαβεῖν ἡμῖν εἶναι τὴν πρώτην, ὅπως τοὺς συμμάχους σώσομεν. ἐὰν γὰρ τοῦτο βεβαίως ὑπάρξῃ, τότε καὶ περὶ τοῦ τίνα τιμωρήσεται τις καὶ ὃν τρόπον ἔξέσται σκοπεῖν· πρὶν δὲ τὴν ἀρχὴν ὀρθῶς ὑποθέσθαι, μάταιον ἥγοῦμαι περὶ τῆς τελευτῆς ὄντινοῦν ποιεῖσθαι λόγον.

Δημοσθένους, Γ' Ὀλυνθιακὸς

Georgian:

From a Unicode conference invitation:

გთხოვთ ახდავე გაიაროთ რეგისტრაცია Unicode-ის მეათე საერთაშორისო კონფერენციაზე დასასწრებად, რომელიც გაიმართება 10-12 მარტს, ქ. მაინცი, გერმანიაში. კონფერენცია შეჰვრებს ერთად მსოფლიოს ექსპერტებს ისეთ დარგებში როგორიცაა ინტერნეტი და Unicode-ი, ინტერნაციონალიზაცია და ღოვანიზაცია, Unicode-ის გამოყენება თვერაციულ 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 Gua'):

[-----|-----]

    □ የዚህንናንፃዕስ በዚህንናንፃዕስ  
    ዚህንናንፃዕስ                         ዚህንናንፃዕስ  
    ዚህንናንፃዕስ                         ዚህንናንፃዕስ  
    ዚህንናንፃዕስ                         ዚህንናንፃዕስ  
    ዚህንናንፃዕስ                         ዚህንናንፃዕስ  
    ዚህንናንፃዕስ                         ዚህንናንፃዕስ  
    ዚህንናንፃዕስ                         ዚህንናንፃዕስ  
    ዚህንናንፃዕስ                         ዚህንናንፃዕስ

(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:

    በዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ  
    በዚህንናንፃዕስ  
    በዚህንናንፃዕስ እንደዚህንናንፃዕስ

Runes:

ᚠ ᚠᚠᚠ ᚠᚠᚠ ᚠ ᚠᚠᚠ ᚠ ᚠᚠᚠ ᚠ ᚠᚠᚠ

(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:

ᚠᚠᚠᚠᚠᚠ

ᚠᚠᚠᚠᚠᚠ  
ᚠᚠᚠᚠᚠᚠ  
ᚠᚠᚠᚠᚠ  
ᚠᚠᚠ  
ᚠᚠᚠ  
ᚠᚠᚠ  
ᚠᚠᚠ  
ᚠ

ᚠᚠᚠᚠᚠᚠ

ᚠᚠᚠ  
ᚠᚠᚠ  
ᚠ  
ᚠ  
ᚠ  
ᚠ  
ᚠ  
ᚠ  
ᚠ  
ᚠ  
ᚠ  
ᚠ  
ᚠ  
ᚠ  
ᚠ  
ᚠ

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

Compact font selection example text:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z / 0 1 2 3 4 5 6 7 8 9  
a b c d e f g h i j k l m n o p q r s t u v w x y z € © μ Α Ε Ο Ρ Β ε ö ÿ  
— „ “ „ , † • ... % ™ œ Š Ÿ ž € А В Г Δ Ω α β γ δ ω А Б В Г Д а б в г д

forall i < len  
 if arr[i] == target:  
 return i  
return -1

Greetings in various languages:

Hello world, Καλημέρα κόσμε, こんにちは

Box drawing alignment tests:

