

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

UTF-8 encoded sample plain-text file

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

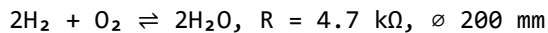
The ASCII compatible UTF-8 encoding of ISO 10646 and Unicode plain-text files is defined in RFC 2279 and in ISO 10646-1 Annex R.

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

Mathematics and Sciences:

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

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



Linguistics and dictionaries:

`ði` `intə'næʃənəl` `fə'netɪk` `əsousi'eɪʃn`  
`Y` [`'ypsɪlɔn`], `Yen` [`jɛn`], `Yoga` [`'jo:ɡɑ`]

APL:

$$((\vee iV)=i\rho V)/V\leftarrow, V \quad \boxed{\leftarrow i \rightarrow \rho \Delta \nabla \supset \underline{\phi} \overline{\phi}}$$

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: 11I|, 00D, 8B
- the euro symbol: 

14.95 €
---------

Greek (in Polytonic):

The Greek anthem:

Σὲ γνωρίζω ἀπὸ τὴν κόψη  
τοῦ σπαθοῦ 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 ταῦτὰ παρίσταται μοι γινώσκειν, ὦ ἄνδρες Ἀθηναῖοι, ὅταν τ' εἰς τὰ πράγματα ἀποβλέψω καὶ ὅταν πρὸς τοὺς λόγους οὐδ' ἀκούω· τοὺς μὲν γὰρ λόγους περὶ τοῦ τιμωρῆσασθαι Φίλιππον ὁρῶ γιγνομένους, τὰ δὲ πράγματ' εἰς τοῦτο προήκοντα, ὥσθ' ὅπως μὴ πεισοόμεθ' αὐτοῖ πρότερον κακῶς σκέψασθαι δέον. οὐδέν 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 τιμωρῆσασθαι, καὶ μάλ' ἀκριβῶς οἶδα· ἐπ' ἐμοῦ γάρ, οὐ πάλα γέγονεν ταῦτ' ἀμφοτέρα· νῦν μέντοι ἐπέπεισμα τοῦθ' ἱκανὸν προλαβεῖν ἡμῖν εἶναι τὴν πρώτην, ὅπως τοὺς συμμάχους σώσομεν. ἐὰν γὰρ τοῦτο βεβαίως ὑπάρξῃ, τότε καὶ περὶ τοῦ τίνα τιμωρῆσεται τις καὶ ὅν τρόπον ἐξέσται σκοπεῖν· πρὶν δὲ τὴν ἀρχὴν ὀρθῶς ὑποθέσθαι, μάταιον ἡγοῦμαι περὶ τῆς τελευτῆς ὄντινον ποιεῖσθαι λόγον.

Δημοσθένους, Γ' Ὀλυμπιακός

Georgian:

From a Unicode conference invitation:

გთხოვთ ახლავე 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 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'):

[----- -----]
๐ แผ่นดินชั้นเสื่อมโทรมแสนสังเวช พระปกเกล้าก่องนักรู้ขึ้นใหม่
สิบสองกษัตริย์ก่อนหน้าแลถัดไป สององค์ไซ้โรงเขลาเบาปัญญา
ทรงนับถือขั้นที่เป็นที่พึ่ง บ้านเมืองจึงวิปริตเป็นนักรหนา
โสรจันเรียกทัพทั่วหัวเมืองมา หมายถึงข้ามดชั่วตัวสำคัญ
เหมือนขับไล่ไล่เสื่อจากเคหา รับหมาป่าเข้ามาเลยอาสัญ
ฝ่ายอ้องอุ้นยุแยกให้แตกกัน ไซ้สาวนั้นเป็นชนวนขึ้นชวนใจ
พลันลี้ฉุยก็กลับก่อเหตุ ข้างอาเพศจริงหนาฟ้าร้องไห้
ต้องรบราฆ่าฟันจนบรรลัย ฤหาใครคำชู้กับบรรลึงก์ ฯ

(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:

ሰማይ አይታረስ ንጉሥ አይከሰስ።
ብላ ካለኝ እንደአባቴ በቆመጠኝ።
ጌጥ ያለቤቴ ቁምጥና ነው።
ደሀ በሕልሙ ቅቤ ባይጠጣ ንጣት በገደለው።
የአፍ ወለምታ በቅቤ አይታሸም።
አይጥ በበለ ዳዋ ተመታ።
ሲተረገሙ ይደረግሙ።
ቀስ በቀስ፣ ዕንቁላል በእግሩ ይሄዳል።
ድር ቢያብር አንበሳ ያሰር።
ሰው እንደቤቴ እንጅ እንደ ጉረቤቴ አይተዳደርም።
እግዜር የከፈተውን ጉሮሮ ሳይዘጋው አይድርም።
የጎረቤት ሌባ፣ ቢያዩት ይስቅ ባያዩት ያጠልቅ።
ሥራ ከመፍታት ልጄን ላፋታት።
ዓባይ ማደሪያ የለው፣ ግንድ ይዞ ይዘራል።
የእስላም አገሩ መካ የአሞራ አገሩ ዋርካ።
ተንጋሎ ቢተፋ ተመልሶ ባፋ።
ወዳጅህ ማር ቢሆን ጨርስህ አትላሰው።
እግርህን በፍራሽህ ልክ ዘርጋ።

Runes:

H M K P F B D F T H M B N M M F t D F M F t F M M t F R D P M F R M N M P I D D F P M M F

(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:

Braille representation of the text above, using compact font selection.

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

Compact font selection example text:

ABCDEFGHIJKLMNOPQRSTUVWXYZ /0123456789  
 abcdefghijklmnopqrstuvwxyz €@µÀÆÖþRéöÿ  
 --'“”„†•…‰™œŠŸž€ ΑΒΓΔΩαβγδω ΑΒΓΔαβγδ  
 √∂€℞Λ∪≅∞ ↑↖↗↘ ↓↙↘ ↵ ↶ ↷ ↸ ↹ ↺ ↻ ↼ ↽ ↾ ↿ ↽ ↾ ↿ ↽ ↾ ↿

Greetings in various languages:

Hello world, Καλημέρα κόσμε, コンニチハ

Box drawing alignment tests:

