

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

UTF-8 encoded sample plain-text file

Markus Kuhn ['maɪkʊs 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 ['ɪpsɪlɔn], Yen [jɛn], Yoga ['jo:ɡɑ]

APL:

$((V \uparrow V) = \uparrow \rho V) / V \leftarrow, V \quad \square \leftarrow \uparrow \rightarrow \rho \Delta \nabla \triangleright \bar{\uparrow} \nabla \uparrow$

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.

Ἐπ’ 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 οὖς 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 οὖσαν παριστάντες ὑμῖν ἀμαρτάνειν. ἐγὼ δέ, ὅτι μὲν

ποτ' ἔξῃν τῆ 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' ἰκανὸν 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' Ὀλυμπιακὸς

Georgian:

From a Unicode conference invitation:

გთხოვთ ახლავებ გაიაროთ რეგისტრაცია Unicode-ის მეთექვსმეტე საერთაშორისო კონფერენციისაზე 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'):

