

## UTF-8 encoded sample plain-text file

Markus Kuhn ['ma'kʊs ku:n] <<http://www.cl.cam.ac.uk/~mgk25/>> – 2002-07-25

The ASCII compatible UTF-8 encoding used in this plain-text file is defined in Unicode, ISO 10646-1, and RFC 2279.

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

Mathematics and sciences:

$\oint E \cdot da = Q, \quad n \rightarrow \infty, \quad \sum f(i) = \prod g(i),$   
 $\forall x \in \mathbb{R}: [x] = -[-x], \quad \alpha \wedge \neg \beta = \neg(\neg \alpha \vee \beta),$   
 $\mathbb{N} \subseteq \mathbb{N}_0 \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C},$   
 $\perp < a \neq b \equiv c \leq d \ll \top \Rightarrow ([A] \Leftrightarrow \langle B \rangle),$   
 $2\text{H}_2 + \text{O}_2 \rightleftharpoons 2\text{H}_2\text{O}, \quad R = 4.7 \text{ k}\Omega, \quad \varnothing 200 \text{ mm}$

$$\left\{ \left( \frac{a^2 + b^3}{c_8} \right)^{\infty} \sum_{i=1}^{\infty} a^i - b^i \right\}$$

Linguistics and dictionaries:

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

APL:

$((V \vee V) = \vee V) / V \leftarrow V \quad \boxed{\leftarrow \vee \rightarrow \rho \Delta \nabla \triangleright \overline{\text{⌈⌋}} \text{⌈⌋}}$

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 €

Combining characters:

STARGA    Ê SG-1, a = v    ´ = r    ;    ā ⊥ b̄

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.

Ἄπ' τὰ κόκκαλα βγαλμένη  
τῶν Ἑλλήνων τὰ ἱερά  
καὶ σὺν πρῶτα ἀνδρειωμένη  
χαῖρε, ὦ χαῖρε, Ἐλευθεριά!

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

Οὐχὶ ταῦτ' ἀρίστα μοι γινώσκειν, ὦ ἄνδρες Ἀθηναῖοι,  
ὅταν τ' εἰς τὰ πράγματα ἀποβλέψω καὶ ὅταν πρὸς τοὺς  
λόγους οὖς ἀκούω· τοὺς μὲν γὰρ λόγους περὶ τοῦ  
τιμωρήσασθαι Φίλιππον ὁρῶ γιγνομένους, τὰ δὲ πράγματ'  
εἰς τοῦτο προήκοντα, ὥσθ' ὅπως μὴ πεισόμεθ' αὐτοῖς  
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' ἐξῆν τῇ πόλει καὶ τὰ αὐτῆς ἔχειν ἀσφαλῶς καὶ Φίλιππον  
τιμωρήσασθαι, καὶ μάλ' ἀκριβῶς οἶδα· ἐπ' ἐμοῦ γάρ, οὐ πάσαι  
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 ὁρῶνς ὑποθέσθαι, μάταιον ἡγοῦμαι περὶ τῆς  
τελευτῆς ὄντινοῦν ποιεῖσθαι λόγον.

Δημοσθένους, Γ' Ὀλυνθιακὸς

Georgian:

From a Unicode conference invitation:

გთხოვთ ახლაცე 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'):

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

๑ แผนดินอันเสื่อมโทรมแสนสังเวช พระปลูกศรของบุกรุ่นใหม่  
สิบสองกษัตริย์ก่อนหน้าแลถัดไป สององค์ซจรเงลาใบปญญา  
ทรงนับถือขึ้นที่เป็นที่พึ่ง บ้านเมืองจึงวิปริตเป็นนั๊กหนา  
โฮจันเรียกทัพทั่วหัวเมืองมา ทูมายจะสามดชั่วตัวสำคัญ  
เหมือนขับไล่เลือดจากเคหา รับหมาป่าเข้ามาเลยอาสัญ  
ฝ่ายออฉนยุแกลโหแตกกัน ไชยสถานเป็นชนวนขึ้นชวนใจ  
พลันลิ้นขลุ่ยก็กลับก่อเหตุ ชางอแปะจริงหนาฟารองไ  
ต้องรบราฆ่าฟันจนบรรลัย ฤทธาใครคำชุกุบบรรลึง 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 እይታረስ ንጉሥ እይከሰስ።  
 ብላ ካለኝ እንደአባቴ በቁመጠኝ።  
 ጌጥ ያለቤቴ ቁምጥና ነው።  
 ደህ በሕልሙ ቅቤ ባይጠግ ንግት በገደለው።  
 የአፍ ወለምታ በቅቤ እይታሽም።  
 እይጥ በበላ ዳዋ ተመታ።  
 ሲተረጉሙ ይደረግሙ።  
 ቀስ በቀስ፥ ዕንቁላል በእግሩ ይሄዳል።  
 ድር ቢያብር እንበሳ ያስር።  
 ሰው እንደቤቴ እንጅ እንደ ጉረቤቴ እይተዳደርም።  
 እግዜር የከፈተውን ጉሮሮ ሳይዘጋው እይድርም።  
 የጎረቤት ሊባ፥ ቢያዩት ይስቅ ባያዩት ያጠልቅ።  
 ሥራ ከመፍታት ልጄን ላፋታት።  
 ዓባይ ማደረያ የለው፥ ግንድ ይዞ ይዞራል።  
 የእስላም አገሩ መካ የአሞራ አገሩ ዋርካ።  
 ተንጋሎ ቢተፋ ተመልስ ባፋ።  
 ውዳጅህ ማር ቢሆን ጨርስህ አትላሰው።  
 እግርህን በፍራሽህ ልክ ዘርጋ።

## Runes :

(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:

תאריך: 15.05.2018  
 שם: ד"ר דוד גורן  
 תפקיד: מנהל מחלקת המחקר  
 מוסד: אוניברסיטת תל אביב  
 נושא: תוכנית מחקר להקמת מרכז מחקר  
 לבינה שיתופית  
 מטרה: להקים מרכז מחקר  
 לבינה שיתופית, המיועד  
 לסייע למגזר העסקי  
 ולממשלה, במטרה  
 להפחית את עלויות  
 המחקר, ולשפר את  
 היעילות של המחקר  
 המבוצע.

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

[illegible]

Hello world, Καλημέρα κόσμε, コンニチハ

x drawing alignment tests:

