

(just a page useful to test UTF-8. Please ignore)

UTF-8 SAMPLER

¥ · £ · € · \$ · ¢ · Ø · Ğ · F · £ · ḡ · 𐀀 · Pts · Rs · ₩ · 𐀀 · 𐀀 · K · ₪ · ৳ · ₹

Frank da Cruz

The Kermit Project - Columbia University

New York City

fdc@columbia.edu

Last update: Sat Apr 30 14:22:05 2011

[PEACE Poetry I Can Eat Glass Pangrams HTML Features Credits, Tools, Commentary](#)

UTF-8 is an ASCII-preserving encoding method for Unicode (ISO 10646), the Universal Character Set (UCS). The UCS encodes most of the world's writing systems in a single character set, allowing you to mix languages and scripts within a document without needing any tricks for switching character sets. This web page is encoded directly in UTF-8.

As shown [HERE](#), Columbia University's Kermit 95 terminal emulation software can display UTF-8 plain text in Windows 95, 98, ME, NT, XP, Vista, or Windows 7 when using a monospace Unicode font like Andale Mono WT J or Everson Mono Terminal, or the lesser populated Courier New, Lucida Console, or Andale Mono. C-Kermit can handle it too, if you have a Unicode display. As many languages as are representable in your font can be seen on the screen at the same time.

This, however, is a Web page, which started out as a kind of stress test for UTF-8 support in Web browsers, which was spotty when this page was first created but which has become standard in all modern browsers. The problem now is mainly the fonts and the browser's (or font's) support for the nonzero Unicode planes (as in, e.g., the Braille and Gothic examples below). And to some extent the rendition of combining sequences, right-to-left rendition (Arabic, Hebrew), and so on. [CLICK HERE](#) for a survey of Unicode fonts for Windows.

The subtitle above shows currency symbols of many lands. If they don't appear as blobs, we're off to a good start! (The one on the end is the new Indian Rupee sign which won't show up in fonts for a while.)

Poetry

From the Anglo-Saxon Rune Poem (Rune version):

𐀀𐀁𐀂𐀃𐀄𐀅𐀆𐀇𐀈𐀉𐀊𐀋𐀌𐀍

𐀎𐀏𐀐𐀑𐀒𐀓𐀔𐀕𐀖𐀗𐀘𐀙𐀚𐀛𐀜𐀝

𐀞𐀟𐀠𐀡𐀢𐀣𐀤𐀥𐀦𐀧𐀨𐀩𐀪𐀫𐀬

From [Lazamon's Brut](#) (The Chronicles of England, Middle English, West Midlands):

An preost wes on leoden, Lazamon was ihoten
He wes Leovenaðes sone — liðe him be Drihten.
He wonede at Ernleze at æðelen are chirechen,

Uppen Sevarne stape, sel þar him þuhte,
Onfest Radestone, þer he bock radde.

(The third letter in the author's name is Yogh, missing from many fonts; [CLICK HERE](#) for another Middle English sample with some explanation of letters and encoding).

From the Tagelied of Wolfram von Eschenbach (Middle High German):

Sîne klâwen durh die wolken sint geslagen,
er stîget ûf mit grôzer kraft,
ich sih in grâwen tâgelîch als er wil tagen,
den tac, der im geselleschaft
erwenden wil, dem werden man,
den ich mit sorgen în verliez.
ich bringe in hinnen, ob ich kan.
sîn vil manegiu tugent michz leisten hiez.

Some lines of Odysseus Elytis (Greek):

Monotonic:

Τη γλώσσα μου έδωσαν ελληνική
το σπίτι φτωχικό στις αμμουδιές του Ομήρου.
Μονάχη έγνοια η γλώσσα μου στις αμμουδιές του Ομήρου.

από το Άξιον Εστί
του Οδυσσέα Ελύτη

Polytonic:

Τῆ γλῶσσα μου ἔδωσαν ἑλληνικὴ
τὸ σπίτι φτωχικὸ στὶς ἀμμουδιές τοῦ Ὁμήρου.
Μονάχη ἔγνοια ἡ γλῶσσα μου στὶς ἀμμουδιές τοῦ Ὁμήρου.

ἀπὸ τὸ Ἄξιον ἔστί
τοῦ Ὁδυσσέα Ἐλύτη

The first stanza of Pushkin's Bronze Horseman (Russian):

На берегу пустынных волн
Стоял он, дум великих полн,
И вдаль глядел. Пред ним широко
Река неслася; бедный чёлн
По ней стремился одиноко.
По мшистым, топким берегам
Чернели избы здесь и там,
Приют убогого чухонца;

И лес, неведомый лучам
В тумане спрятанного солнца,
Кругом шумел.

Šota Rustaveli's Vep'xis T'qaosani, The Knight in the Tiger's Skin (Georgian):

ვეპ'ხის ტყაოსანი შოთა რუსთაველი

ღმერთისი შეშვედრე, ნუთუ კვლა დამხნას სოფლისა შრომასა, ცუცხლს, წყალსა და მიწასა, ჰაერთა
თანა მრომასა; მომცნეს ფრთენი და აღფრინდრე, მივჰხვდე მას ჩემსა ნდომასა, დღისით და ღამით
ვჰხვდვოდრე მზისა ელვათა კრთომასა.

Tamil poetry of Subramaniya Bharathiyar: சுப்ரமணிய பாரதியார் (1882-1921):

யாமறிந்த மொழிகளிலே தமிழ்மொழி போல் இனிதாவது எங்கும் காணோம்,
பாமரராய் விலங்குகளாய், உலகனைத்தும் இகழ்ச்சிசொலப் பான்மை கெட்டு,
நாமமது தமிழரெனக் கொண்டு இங்கு வாழ்ந்திடுதல் நன்றோ? சொல்லீர்!
தேமதூரத் தமிழோசை உலகமெலாம் பரவும்வகை செய்தல் வேண்டும்.

Kannada poetry by Kuvempu — ಬಾ ಇಲ್ಲಿ ಸಂಭವಿಸು

ಬಾ ಇಲ್ಲಿ ಸಂಭವಿಸು ಇಂದೆನ್ನ ಹೃದಯದಲಿ
ನಿತ್ಯವೂ ಅವತರಿವ ಸತ್ಯವತಾರ

ಮಣ್ಣಾಗಿ ಮರವಾಗಿ ಮಿಗವಾಗಿ ಕಗವಾಗಿ...
ಮಣ್ಣಾಗಿ ಮರವಾಗಿ ಮಿಗವಾಗಿ ಕಗವಾಗಿ
ಭವ ಭವದಿ ಭತಿಸಿಹೇ ಭವತಿ ದೂರ

ನಿತ್ಯವೂ ಅವತರಿವ ಸತ್ಯವತಾರ

ಬಾ ಇಲ್ಲಿ

I Can Eat Glass

And from the sublime to the ridiculous, here is a certain phrase¹ in an assortment of languages:

Sanskrit: काचं शक्नोम्यत्तुम् । नोपहिनस्ति माम् ॥

Sanskrit (standard transcription): kācaṃ śaknomyattum; nopahinasti mām.

Classical Greek: ὕαλον φαγεῖν δύναμαι· τοῦτο οὐ με βλάπτει.

Greek (monotonic): Μπορώ να φάω σπασμένα γυαλιά χωρίς να πάθω τίποτα.

Greek (polytonic): Μπορώ νὰ φάω σπασμένα γυαλιὰ χωρίς νὰ πάθω τίποτα.

Etruscan: (NEEDED)

Latin: Vitrum edere possum; mihi non nocet.

Old French: Je puis mangier del voirre. Ne me nuit.

French: Je peux manger du verre, ça ne me fait pas mal.

Provençal / Occitan: Pòdi manjar de veire, me nafrariá pas.

