

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

UTF-8 SAMPLER

¥ · £ · € · \$ · ¢ · ₩ · ₪ · ₧ · ₮ · ₯ · ₣ · ₧ · ₦ · ₩ · ₪ · ₧ · ₩ · ₪ · ₧

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):

Ƿ̄Ƿ·Bƿþ·ƿRFFNꝑ·ƿIRF·XΜHΡѧLNꝑM
ጀLΜFF·ƿMFF·MFF+FF·XΜHΡѧLΓ·MΙLΓN+·HѧL·MFFFF+
XΙF·Hℳ·PIM·PFF·MRIH↑+M·MFFMh·HΓJ↑FF+:

From Laȝamon's Brut (The Chronicles of England, Middle English, West Midlands):

An preost wes on leoden, Laȝamon was ihoten
He wes Leovenaðes sone — liðe him be Drihten.
He wonede at Ernleȝe at æðelen are chirechen,
Uppen Sevarne staþe, 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 tägelîch als er wil tagen,
 den tac, der im geselleschaft
 erwenden wil, dem werden man,
 den ich mit sorgen ïn 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 ή γλώσσα μου στις αμμουδιές του Ομήρου.

από το Άξιον Εστί¹
 του Οδυσσέα Ελύτη

Polytonic:

Τὴ γλῶσσα μοῦ ἔδωσαν ἑλληνικὴ
 τὸ σπίτι φτωχικὸ στὶς ἀμμουδιὲς τοῦ Ὁμήρου.
 Μονάχη ἔγνοια ἡ γλῶσσα μου στὶς ἀμμουδιὲς τοῦ Ὁμήρου.

ἀπὸ τὸ Ἄξιον ἐστί¹
 τοῦ Ὁδυσσέα Ἔλύτη

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 Veřxis Tqaosani, Th, The Knight in the Tiger's Skin (Georgian):

ვეპხის ტყაოსანი შოთა რუსთაველი

ღმერთის შემვედრე, ნუთუ კვლა დამხსნას სოფლისა შრომასა, ცეცხლს, წყალსა და მიწასა, ჰაერთა თანა მრომასა; მომცნეს ფრთენი და აღვფრინდე, მივჰევდე მას ჩემსა ნდომასა, დღისით და ღამით ვჰედვიდე მზისა ელვათა კრთომასა.

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

யாமறிந்த மொழிகளிலே தமிழ்மொழி போல் இனிதாவது எங்கும் காணோம்,

පාමරරාය් ඩිලන්කුකளාය්, ඔලකනෙන්තුම් මූකුස්සිසිසාල්ප පාණ්මේ කෙටු,
න්‍යාමතු තමිලුබරනක් කොන්ගු මූන්ත්තිභෙක් නුත්තෙ? සීල්ව්ර!
තොමතුරත් තමිලුෂාස් ඔලකමෙලාම් පරවුම්වකේ සේය්තල් වෙන්ගුම්.

Kannada poetry by Kuvempu — ಬಾ ಇಲ್ಲಿ ಸಂಭವಿಸು

ಬಾ ಇಲ್ಲಿ ಸಂಭವಿಸು ಇಂದನ್ನು ಹೃದಯದಲ್ಲಿ

ನಿತ್ಯವೂ ಅವಶರಿಪ ಸತ್ಯವತಾರ

ಮಣಾಗಿ ಮರವಾಗಿ ಮಿಗವಾಗಿ ಕಗವಾಗಿ...

ಮಣಾಗಿ ಮರವಾಗಿ ಮಿಗವಾಗಿ ಕಗವಾಗಿ

ಭವ ಭವದಿ ಭತ್ತಿಸಿಹೇ ಭವತಿ ದೂರ

ನಿತ್ಯವೂ ಅವಶರಿಪ 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 शक्नोम्यतुम् । नोपहिनस्ति माम् ॥

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

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

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

Greek (polytonic): Μπορῶ νὰ φάω σπασμένα γυαλιὰ χωρὶς 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.

Québécois: J'peux manger d'la vitre, ça m'fa pas mal.

Walloon: Dji pou magnî do vêre, coula m' freut nén må.

Chamenois: (NEEDED)

Lorrain: (NEEDED)

Picard: Ch'peux mingi du verre, cha m'foé mie n'ma.

Corsican/Corsu: (NEEDED)

Jèrriais: (NEEDED)

Kreyòl Ayisyen (Haiti): Mwen kap manje vè, li pa blese'm.

Basque: Kristala jan dezaket, ez dit minik ematen.

Catalan / Català: Puc menjar vidre, que no em fa mal.

Spanish: Puedo comer vidrio, no me hace daño.

Aragonés: Puedo minchar beire, no me'n fa mal .

Aranés: (NEEDED)

Mallorquín: (NEEDED)

Galician: Eu podo xantar cristais e non cortarme.

European Portuguese: Posso comer vidro, não me faz mal.

Brazilian Portuguese (8): Posso comer vidro, não me machuca.

Caboverdiano/Kabuverdianu (Cape Verde): M' podê cumê vidru, ca ta maguâ-m'.

