



## (U) Ask Raul: Fonts and Encoding

FROM: 'Raul', a DNI Analyst  
Unknown  
Run Date: 08/09/2004

(U//FOUO) Note from SIGINT Communications: This is the first of a monthly series of articles on DNI (Digital Network Intelligence) matters by guest columnist Raul:

 **SERIES:**  
(U) **Ask Raul - Answers to DNI Questions**

- (U) Hey, Raul! What's the difference between a font and a character set encoding? Aren't they the same?

(U) Argh!!! If there is one thing that will drive old Raul crazy it is hearing people -- people who have been through endless NSA C2C/DNI courses -- referring to an encoding as a font or vice versa. So, in order to help keep my blood pressure down, here's how it works:

(U) A font is simply a collection of symbols (glyphs) of a particular style and size. Fonts have names which may or may not be of any use in figuring out what they cover. For example: Times, Times New Roman, Arial, Helvetica, Times New Roman Cyrillic, Ivan's Crazy Cyrillic, etc. For a given language or glyph set, the different fonts are simply different styles of the same thing.

For example, here are the letters 'ABCDEF' in several different fonts:



(U) In this case there isn't much difference to the eye between Helvetica and Times but AvantGarde and Courier are obviously different from one another and the last two.

(U) Computers use fonts as the means to allow them to display things, whether that be on a monitor or printer, to a human being. At the core though, the computer simply sees things as numbers. To give you an idea, here is an entire font, Lucida Sans Typewriter:



(U) In this case, the font name, which is listed at the top of this image, tells us the encoding this font uses is ISO8859-1. This standard says that certain glyphs will be encoded with specific values. Here, 'A', regardless of how fancy it might be, will always be represented by the number 0x41 (decimal 65), 'B' will be 0x42 (66), 'a' will be 0x61 (97), etc.

(U) An encoding is simply a method of representing one thing with something else. In this case it is glyphs with numbers. In a file, on a computer using ISO8859-1 encoding, the phrase,

"The big black bird"

is actually stored as:

0x540x680x650x200x620x690x670x200x620x6c0x610x630x6b0x  
200x620x690x720x640x2e

1. [Ask Raul](#) : Fonts and Encoding
2. [Ask Raul](#) : Dictionary Equations
3. [Ask Raul](#) : HTML Coding and Email
4. [Ask Raul](#) : PDF Files
5. [Ask Raul](#): Damaged Data
6. [Ask Raul](#) : Getting the Most from Metadata

It is only when we send this to a display device that it gets transformed, via a font, into:

### The big black bird

(U) Every foreign language is covered by a multitude of encodings. Some are international standards while others are completely home grown.

(U) In some cases, an encoding will use a single 7 or 8-bit value to represent a single glyph. In others, 16, 24, or even 32 bits can be used. For example, here are a few of the encodings that cover the Cyrillic letter 'A':

Glyph	Encoding	Value
• A	• ISO8859-5	• 0xb0
• A	• KOI-8	• 0xf1
• A	• CP1251	• 0xd0
• A	• JAVA	• \u0410
• A	• UCS-2LE	• 0x100x04
• A	• UCS-2BE	• 0x040x10
• A	• UTF-8	• 0xd00x90

(U) As you can readily see, if you go looking for something and don't allow for the encoding, you are not likely to find what you want.

(U) So, do Raul a favor and keep your fonts and encodings straight.

---

*(U) If you have a DNI-related question for Raul, please use the "comments/suggestions" function at the bottom of this page to pose it. We will make sure Raul gets it.*

**"(U//FOUO) SIDtoday articles may not be republished or reposted outside NSANet without the consent of S0121 ([DL sid comms](#))."**