COLLEGE OF THE NORTH ATLANTIC

ASCII Codes

American Standard Code for Information Interchange

By Daryle Niedermayer, I.S.P.

COLLEGE OF THE NORTH ATLANTIC QATAR

How Humans Count

- As humans, we use base 10 to count:
  - We start at 0, and count to 9, using ten different digits (because we have ten fingers)
  - Then we need to expand our count by one column and start over again: 10, 11, 12, 13…19
  - When we do this ten times, we need to move to a second new column representing hundreds (10x10)

COLLEGE OF THE NORTH ATLANTIC QATAR

The Size of a Number

- As humans, if we have a limited number of digits we can have in a number, we can only count to 1 less than that number:
  - Eg. 3 digits allows for a maximum size of 10x10x10 (10^3) or 1,000. Therefore, we can only count to 1,000-1 or 999.
  - If we have 6 digits available, we can count to 999,999 (or 10^6-1)
  - Remember Y2K?

COLLEGE OF THE NORTH ATLANTIC QATAR

Counting in Base 10

100's (10^2)	10's (10^1)	1's (10^0)
		0
		1
		2
		…
		9
	1	0
	1	1
	1	2
		…
	1	9

100's (10^2)	10's (10^1)	1's (10^0)
	2	0
		…
	9	9
1	0	0
1	0	1
1	0	2
		…
9	9	8
9	9	9
		…

COLLEGE OF THE NORTH ATLANTIC QATAR

How Computers Count

- Computers use base 2 to count:
  - They start at 0 and then count to 1 before they have to expand by another column and start over: 0b, 1b, 10b, 11b
  - We use “b” to denote that these are binary or base 2 numbers and not decimal numbers (11b = 3 in base 10)
  - Once we reach 4 in decimal, we already need 3 columns…

COLLEGE OF THE NORTH ATLANTIC QATAR

Counting in Base 2

Decimal	16's (2^1)	8's (2^0)	4's (2^2)	2's (2^1)	1's (2^0)
1					1
2				1	0
3				1	1
4			1	0	0
5			1	0	1
6			1	1	0
7			1	1	1
8		1	0	0	0

COLLEGE OF THE NORTH ATLANTIC QATAR

Base 8 and 16

- Binary digits get big in a hurry.
- To make things easier, we often "chunk" them into 3 or 4 bit segments, representing base 8 or 16 numbers.

COLLEGE OF THE NORTH ATLANTIC QATAR

- Eg. 101101111001b

- is 5 5 7 1 in Base 8 (octal)
- Eg. 101101111001b

- Or 11 7 9 in Base 16 (hex)

COLLEGE OF THE NORTH ATLANTIC QATAR

Abbreviations for Base 8 & 16

- We usually call Base 8 numbers “Octal” from the Greek word for “eight” (eg. Octopus, Octogon)
- We call Base 16 numbers “Hexadecimal” because they use a decimal base (Base 10) with another 6 numbers (“Hex” from the Greek word for 6; eg. Hexagon).
- To avoid confusion, we often refer to Octal numbers with a small “o” in the front (eg. o452) and Hexadecimal numbers with a small “x” in the front and an “h” behind (eg. x4525h).

COLLEGE OF THE NORTH ATLANTIC QATAR

- Because we only have ten digits in our vocabulary but we need 16 for Hexadecimal, we use the letters A-F to represent the Hexadecimal digits ten, eleven, twelve, thirteen, fourteen and fifteen.
- In this way, the number x11-7-9 from the previous slide becomes xB79

10	11	12	13	14	15
A	B	C	D	E	F

COLLEGE OF THE NORTH ATLANTIC QATAR

Where Else are Base 2, 8 or 16 Used?

- Internet routing uses Base 2 numbers for IP numbers, subnetting and routing tables
- Base 8 numbers are used to define file permissions on UNIX operating systems
- Base 16 is used for defining colour codes in Web pages (HTML).

COLLEGE OF THE NORTH ATLANTIC QATAR

Data Representation

- Each binary digit is called a “bit” (**BI**nary digi**T**).
- A collection of 8 bits is called a “byte”.
  - A byte can normally represent 2^8 or 256 different values.
  - However, early computers saved one bit for error checking or “parity”, so they could only store 2^7 or 128 different values.

COLLEGE OF THE NORTH ATLANTIC QATAR

Parity Checking

- Early computers were prone to errors so one bit out of every byte was reserved as a parity check.
- If a computer was to have “even” parity, then there would have to be an even number of “1” bits in the byte. “Odd” parity meant an odd number of bits in each byte.
- If a computer set for even parity discovered that a byte had an odd number of “1”s, it would report an error.

COLLEGE OF THE NORTH ATLANTIC QATAR

ASCII and Parity

- With only 128 characters available there was just enough bits to represent
  - All 26 English characters in uppercase and lowercase (52 in all)
  - The digits 0-9
  - Punctuation marks
  - The control characters need to represent tabs, return and control keys

COLLEGE OF THE NORTH ATLANTIC QATAR

ASCII without Parity

- With more reliable computers, it seemed silly to waste 1/8 (12.5%) of all our storage on error checking.
- What to do with the extra 128 options?
  - Other language sets (ANSI)
  - “Graphic” and other European language sets

COLLEGE OF THE NORTH ATLANTIC QATAR

Features of ASCII Sets

- Numeric values represent alphabetic order. The letter “A” is 65d, the letter “B” is 66d; thus A<B.
- To make an upper case letter lower case, add 32d to it’s value:
  - Eg. “A” is 65d and “a” is 97d so to capitalize the letter “a”, subtract 32d.

COLLEGE OF THE NORTH ATLANTIC QATAR

Line Feeds

- Different Operating Systems use different ASCII codes to represent a newline:
  - UNIX/Linux use a Line Feed (10d)
  - MS-DOS requires both Line Feed and Carriage Return characters (10d, 13d)
  - Apple Macintosh (prior to OS/X) used a Carriage Return (13d).
  - This is why Notepad often garbles line endings

COLLEGE OF THE NORTH ATLANTIC QATAR

Limitations to a 1 Byte word

- Many languages cannot be properly represented with only 256 characters.
- Even those that can must have their Operating System set to the correct language base
  - What about international business where a document is written in one language and then displayed on a machine set to another language?

COLLEGE OF THE NORTH ATLANTIC QATAR

Unicode

- With Java came “Unicode”, initially a 2-byte representation of each character but can now support up to 32-bits or 4 bytes per character.
- 2-bytes allows 2^{16} or 65,536 different characters.
- Unicode is backwards compatible with ASCII and ANSI and uses Hexadecimal notation for its character sets.

COLLEGE OF THE NORTH ATLANTIC QATAR

References

- http://en.wikipedia.org/wiki/Newline