WHATS ASCII AND UTF-8 ?

ASCII :

ASCII (American Standard Code for Information Interchange) is the most common format for text files in computers and on the Internet. In an ASCII file, each alphabetic, numeric, or special character is represented with a 7-bit binary number (a string of seven 0s or 1s). 128 possible characters are defined.
ASCII was first developed and published in 1963 by the X3 committee, a part of the American Standards Association (ASA).

The ASCII table is divided into three different sections.

- **Non-printable**, system codes between 0 and 31.
- **Lower ASCII**, between 32 and 127. This table originates from the older, American systems, which worked on 7-bit character tables.
- **Higher ASCII**, between 128 and 255. This portion is programmable; characters are based on the language of your operating system or program you are using. Foreign letters are also placed in this section.

UTF-8 :

UTF-8 is a compromise character encoding that can be as compact as ASCII (if the file is just plain English text) but can also contain any unicode characters (with some increase in file size).

UTF stands for Unicode Transformation Format. The '8' means it uses 8-bit blocks to represent a character. The number of blocks needed to represent a character varies from 1 to 4.

One of the really nice features of UTF-8 is that it is compatible with nul-terminated strings. No character will have a nul (0) byte when encoded. This means that C code that deals with char[] will "just work".