Any string may contain multibyte characters, although the length of such a string (as determined by the `strlen` function) is the number of bytes in the string, not the number of characters. In particular, the format strings in calls of the ...`printf` and ...`scanf` functions may contain multibyte characters. As a result, the C99 standard defines the term ***multibyte string*** to be a synonym for *string*.

## Wide Characters

The other way to encode an extended character set is to use wide characters. A ***wide character*** is an integer whose value represents a character. Unlike multibyte characters, which may vary in length, all wide characters supported by a particular implementation require the same number of bytes. A ***wide string*** is a string consisting of wide characters, with a null wide character at the end. (A ***null wide character*** is a wide character whose numerical value is zero.)

Wide characters have the type `wchar_t` (declared in `<stddef.h>` and certain other headers), which must be an integer type able to represent the largest extended character set for any supported locale. For example, if two bytes are enough to represent any extended character set, then `wchar_t` could be defined as `unsigned short int`.

C supports both wide character constants and wide string literals. Wide character constants resemble ordinary character constants but are prefixed by the letter `L`:

```
L'a'
```

Wide string literals are also prefixed by `L`:

```
L"abc"
```

This string represents an array containing the wide characters `L'a'`, `L'b'`, and `L'c'`, followed by a null wide character.

## Unicode and the Universal Character Set

The differences between multibyte characters and wide characters become apparent when discussing ***Unicode***. Unicode is an enormous character set developed by the Unicode Consortium, an organization established by a group of computer manufacturers to create an international character set for computer use. The first 256 characters of Unicode are identical to Latin-1 (and therefore the first 128 characters of Unicode match the ASCII character set). However, Unicode goes far beyond Latin-1, providing the characters needed for nearly all modern and ancient languages. Unicode also includes a number of specialized symbols, such as those used in mathematics and music. The Unicode standard was first published in 1991.

Unicode is closely related to international standard ISO/IEC 10646, which defines a character encoding known as the ***Universal Character Set (UCS)***. UCS was developed by the International Organization for Standardization (ISO), starting at about the same time that Unicode was initially defined. Although UCS originally differed from Unicode, the two character sets were later unified. ISO now