HTML5 and character encoding

WHAT IS A CHARACTER ENCODING?

Words and sentences in text are created from characters. Examples of characters include the Latin letter á or the Chinese ideograph 請 or the Devanagari character ह.

Characters are grouped into a *character set* (also called a *repertoire*). This is then called a*coded character set* when each character is assigned a particular number, called a *code point.* These code points will be represented in the computer by one or more bytes.

The *character encoding* is the key that maps code points to bytes in the computer memory, and read the bytes back into codepoints.

DECLARING CHARACTER ENCODINGS IN HTML5

Always declare the encoding of your document using a `meta` element with a `charset`attribute, or using the `http-equiv` and `content` attributes (called a pragma directive). The declaration should fit completely within the first 1024 bytes at the start of the file, so it's best to put it immediately after the opening `head` tag.

```
<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="utf-8"/>
...
```

or

```
<!DOCTYPE html>
<html lang="en">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
```

```
    ...
```

It doesn't matter which you use, but it's easier to type the first one. It also doesn't matter whether you type `UTF-8` or `utf-8`.

Content authors need to find out how to declare the character encoding used for the document format they are working with.