

Content Authoring with HTML

Components of an HTML Document

Components of an HTML Document

- Document structure elements – these are inferred
- Document type declaration
- Document title
- Character set declaration
- Content – plain text and content elements

Document Type Declaration

- Must be the first thing to appear in an HTML document
- Identifies the document as an HTML document
- Indicates which version of the HTML standard the browser should use when interpreting the document
- Omission will result in minor display anomalies
- Declared as part of the `<!DOCTYPE>` tag
- HTML5 is identified using the “html” argument like so:

```
<!DOCTYPE html>
```

Document Title

- A title identifies an HTML document “by name”
- It is displayed in the browser heading, as the page name when bookmarking, and on search engine results
- It is declared using the <title> element
- A document must contain exactly one title

```
<title>Janelle Monáe</title>
```

Character Set Declaration

- Computers store data in a binary format – ones and zeros
- In order for a computer to display glyphs – letters, numbers, etc. – it must contain software which can interpret sets of binary and translate them into the appropriate glyph
- There are many different glyph “dictionaries” available, some vary in size and number of glyphs covered

Character Set Declaration

- In early computing ASCII was used – a 128 glyph character set – thus, it only covered a small subset of the Latin alphabet
- The International Standards Organization established standard character sets for use in different language environments – all the ISO sets include the ASCII set within, with an additional 128 characters, depending on the set, for a total of 256 glyphs
- Since the ISO sets did not suffice for multilingual environments, the Unicode Consortium established its own character set, that includes “all the characters, punctuations, and symbols in the world.”

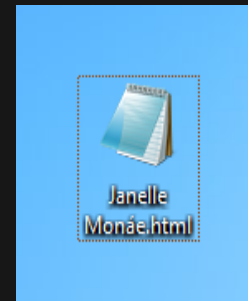
Character Set Declaration

- The character set declaration is made as part of the <meta> element
- It is made using the “charset” attribute like so:

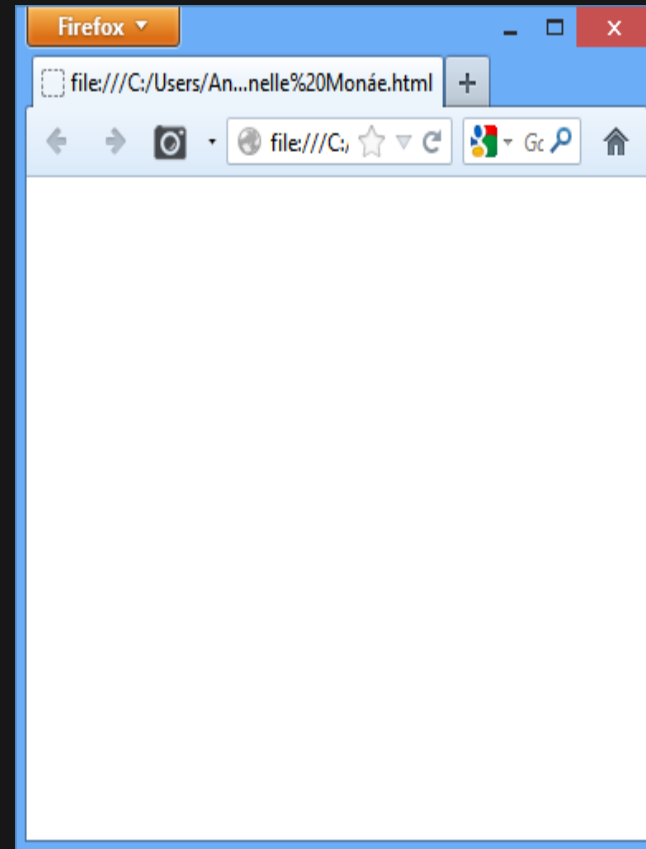
```
<meta charset="utf-8">
```

- Omission will result in the browser using it's default character set – usually ISO-8859-1, Latin alphabet part 1 – which will result in display anomalies under certain circumstances

<!DOCTYPE html>

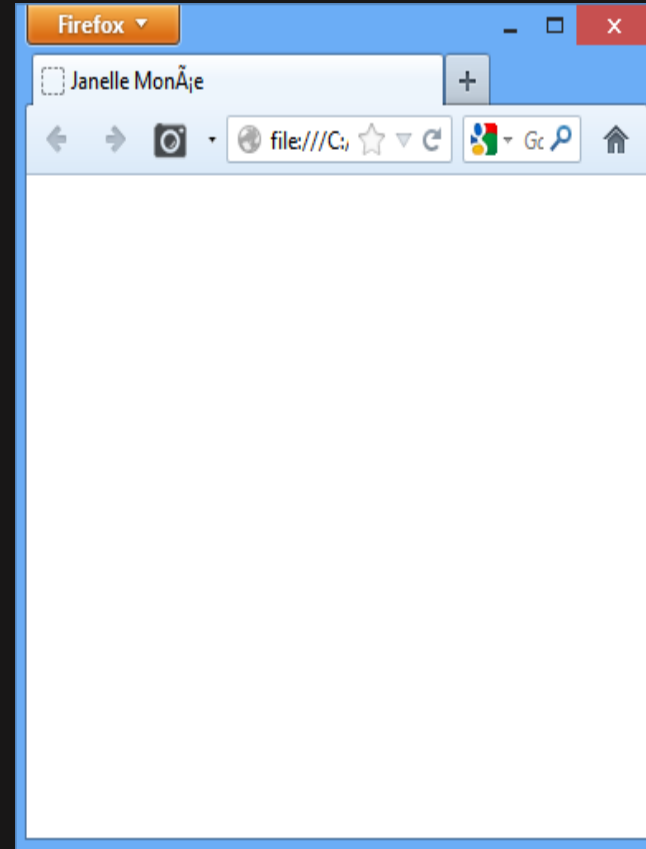


<!DOCTYPE html>



```
<!DOCTYPE html>
```

```
<title>Janelle Monáe</title>
```



```
<!DOCTYPE html>
```

```
<title>Janelle Monáe</title>
```

```
<meta charset="UTF-8">
```

