

Define um conjunto de regras para codificar documentos num formato que é legível por humanos e máquinas.

É uma metalinguagem permitindo que qualquer pessoa crie a sua própria linguagem para diferentes tipos de documentos.

Um documento XML é considerado bem formado se:

- contém 1 ou mais elementos.
- tem uma e uma só raíz.
- os seus elementos se aninham adequadamente uns com os outros.

Desde o XML1.1 todos os documentos têm que começar com uma instrução que indica a sua versão, senão é considerado XML1.0. A codificação por defeito é UTF-8.

```
<?xml version="1.1" encoding="utf-8"?>
```

Secções CDATA são usadas para incluir texto que inclui blocos que, noutro caso, seriam lidos como markup. Começam com `<![CDATA` e terminam com `]]>`. Exemplo:

```
<![CDATA <warning>Not markup!</warning>]]>
```

Elementos são definidos com tag de abertura e fecho. Todos têm que ser fechados. Todos os elementos abertos dentro de outro elemento têm que ser fechados antes do pai.

Atributos são usados para associar pares nome-valor a um elemento. Só aparecem na tag inicial.