

What is a document?

A document can be understood as a complete and self-sufficient written or printed artefact that contains information, however organized, or evidence of an event or transaction. Documents can be physical, or digital, or both. In certain cases, they can as a matter of course be reproduced; in others, not. Real-world examples include communications from public authorities; invoices or receipts; a list of items for discussion, e.g. in class (see Gardiner and Musto, *The Digital Humanities*, pp. 37–42).

Documents can include a number of distinct elements, among them the following:

- images
- time-specific data or information
- coded information, e.g. bar codes
- structured elements, e.g. an ordered or unordered list

Documents are historical records, like the **US constitution**, and can generate further documents in turn. Documents can far outlast the context in which they were first produced and today ancient documents can be reproduced in electronic media (e.g. **Irish Scripts on Screen**).

What is a text?

A text is a particular kind of document, often associated with those which are tokens of special authority (e.g. the text of the Bible) and with literary works (thus, the text of a play or poem). Texts like these are as a rule designed for transmission, e.g. through publication, and can accordingly take a variety of forms over time. A poem of Emily Dickinson's was first published with the title 'The Chariot', but in its **manuscript form** did not have a title, and made distinctive use of dashes as a form of punctuation.

We also use the word 'text' to refer to a part or an element of a document, for instance, within **a work** that contains images as well as words.

In information technology, the word 'document' is often used as the synonym of an electronic file, in which information is encoded in **bits** and **bytes**. A text can be reproduced in files of many different formats, some of them binary (e.g. **MS Word documents** or **PDFs**), some of them in 'plain text' (e.g. **.txt files** or pages written in **HTML** for the web). A binary file can itself be made up of several different kinds of information contained in separate documents; in the case of a Word file, these are **separate files** containing information concerning formatting, for instance, and are recorded in **XML**.

What is a domain?

Today, the **World Wide Web** is, among many other things, a major repository of texts. Once a poem like ‘The Chariot’, whether the manuscript or the published version, is encoded in HTML, it can readily be reproduced on webpages or **in large repositories** organized as databases within which a poem is located with reference to its author.

The web forms part of **the internet** (see also Ince, *A Dictionary of the Internet*). A webpage is to be found on what is termed a ‘domain’, in other words a computer or group of computers that are connected through a web **server** to the Internet.

The means by which we can designate a domain or a particular webpage is through its **uniform resource locator**:

<i>Scheme</i>	<i>Server and domain</i>	<i>Path</i>
https://	www.ucc.ie	/french/
https://	www.ucc.ie	/admin/registrar/modules?mod=WL4102

What is distinctive about text on the Web is that it consists of documents interconnected via the Internet through **hyperlinks** — hence ‘hypertext’.

How does a text find its way into a document?

Computers are used today to encode or to mark up texts, as well as to transmit and to access them. Documents stored in this manner can be displayed as text because a **browser** is programmed to format them (in addition, the browser is the program that requests a file from a remote server on your own (e.g. `www.ucc.ie`) or a different domain).

A markup language like HTML contains elements that allow a text to be represented and in turn transmitted as a structured whole. Here is a stanza from the poem by Dickinson:

```
<p>Because I could not stop for Death – <br />  
He kindly stopped for me – <br />  
The Carriage held but just Ourselves – <br />  
And Immortality.</p>
```

Thus, a complete paragraph is contained within a pair of matching tags that designate this as a distinct element of the document:

<i>Opening tag</i>	<i>Content</i>	<i>Closing tag</i>
<code><p></code>	Let's read a poem.	<code></p></code>