

## Where can I find a text?

The web makes texts available, especially ones which are no longer in copyright, either in the form of webpages, or in repositories, or by providing access to facsimiles of original works or editions. A major example of a large-scale national repository is [Gallica](#), which has been developed by the [Bibliothèque nationale de France](#). These texts exist in many different forms and environments and make different kinds of work possible.

One option is to use a generic search engine and to enter a search string:

dickinson chariot

Another option is to use a major text repository, like [Literature Online \(LION\)](#), and to enter the same search string. The outcomes of these two searches are not identical: what are the salient differences, in your view?

A further option is to search for even more specialized resources:

emily dickinson archive

Dickinson's manuscripts are available in a number of major archives and through these one can have direct access to a facsimile of '[Because I could not stop for Death](#)'.

## How does search work?

These experiments prompt the initial question: what kinds of resource does a generic **search engine** like **Google** or **DuckDuckGo** locate? A further question is this: how does a search engine determine the order in which it presents its results? Are the outcomes of searches using different engines comparable?

How, in turn, are we equipped to evaluate the resources to which a search engine directs us? Will these engines provide direct access to the text of 'The Chariot' as it is presented in **LION**? LION has its own search facility: does this operate in the same way as a generic search engine?

In brief, generic search engines refer us to webpages, on the basis of indexes that they compile. But other important resources are not indexed by these search engines and are located in what is sometimes termed the **deep web**.

The **Emily Dickinson Archive** has its own dedicated search facilities. Are these more comparable to those of LION or a generic search engine? What kinds of work do they make possible? One of the options in this archive is to search within metadata associated with a given item. Are these **metadata** comparable to those found in ordinary webpages?

## Can a machine read?

Dickinson's poem opens with these two lines:

Because I could not stop for Death —  
He kindly stopped for me —

Is the sense of the word 'stop' identical in these two occurrences? By what means can we resolve this question?

A resource like the Macmillan Dictionary allows you to verify the sense of the word in **American** and **British** usage. How does search operate in a resource like this?

This and other major contemporary dictionaries are **derived from corpora** and depend in turn on specialized tools that enable you to derive very precise information from very large datasets. In the dictionary itself, the resulting information comes to be encoded in ways that facilitate retrieval.

But not all interactions with natural language data demand that examples be tagged, as the **Stanford CoreNLP software** shows. See what happens if you enter the two lines above from Dickinson's poem.

## How does a machine aid reading?

The examples that we have looked at allow us to answer the question ‘Can a machine read?’ in a variety of ways. A browser is equipped to ‘read’ HTML and to make the content available as (more or less) structured text. A generic search engine ‘reads’ data (and metadata) contained in webpages. A dedicated search engine or search facility, by contrast, ‘reads’ data according to pre-defined categories by using specific filters: the **UCC phone and email directory** is a case in point. In other words, different search engines allow *you* to access the data that a machine can read in controlled ways.

While traditional reading of print is linear, books too make use of links, e.g. in cross-references and indexes. But the scale of linking in hypertext is vastly greater, prompting the question: do we read differently online? This and related questions are explored in **an essay** by N. Katherine Hayles which we’ll discuss further in class.

How we answer the question above will depend also on how we conceptualize reading. Is there only kind of reading of print media? How might different ways of reading books relate to the kinds of reading that devices and hypertext make possible?