

How to start a search

Mònica Bonich
Albert Cervera
Gema Santos

PID_00169182



Universitat Oberta
de Catalunya

www.uoc.edu



These contents are published under a Creative Commons Attribution-NonCommercial -ShareAlike (BY-NC-SA) Spain licence which allows them to be copied, distributed, publicly, broadcast or used to produce derivative works for non-commercial purposes, provided the author and source are cited. If these contents are transformed, the resulting work may only be distributed under the same or similar licence to this one.

Index

Introduction	5
1. The search process	7
1.1. Defining the consultation accurately	7
1.2. Selecting the most suitable information sources	8
1.3. Preparing and running searches	8
1.4. Assessing the relevance, quality and reliability of the information obtained	10
Bibliography	13

Introduction

The internet is often defined as the source of sources. It is, without a doubt, the most important information source. The internet provides access to an almost limitless quantity of data and documents that grows exponentially by the day. However, this never-ending quantity of information is changing, it is not arranged in any specific way, nor does it follow any kind of structure, and that makes its retrieval particularly difficult. This is why the data or information that we need is not always visible and easy to find immediately, and the process involved in searching for and finding information on the internet almost always requires the use of automated information search tools, such as the so-called **search engines**, database management systems, catalogues, etc.

It is often said that the internet answers practically all our information needs, but we need to ask it the question in the right way. Knowledge of the different sources that may provide us with information and their use and specialisation will also prove extremely useful during a search process.

In any case, the first step in an internet information search process should always be to define clearly the purpose of our specific search. In other words, what the answer is that we are searching for and the fact that it has to meet our information needs. Therefore, the time we invest in clearly defining and delimiting the information that we need, why we want it and where we might find it is time invested well and which very often saves time during the process of searching for and finding the information.



We also need to remember that there may be many types of information searches and that each one of these search types requires a different strategy.

For example, in the search for a known document, such as when we are searching for a specific book in a library catalogue or a bookshop catalogue, or when we are searching a database for a specific article, about which we have sufficient bibliographical details to help us find them (title, author, etc.), the search will be quick and straightforward, and the search strategy will be simple, however, we will need to know the source or sources in which to find them and we will require a basic knowledge about how to put questions to a catalogue or database search engine.

We find ourselves in this very situation when we want to find or confirm a specific piece of data. This is called a **factual search**.

The following provide a few examples of this type of information search: what is the capital of Somalia, what is the website address of Girona City Council, or what is the most popular children's name in 2010 in Catalonia.

In this type of search, the most important factor is to find the answer and not so much the search strategy; the key, therefore, is to find the most suitable information source in each case.

These two search types also have something in common, we know when the search process is complete, i.e. when we have found the document or data that we were looking for. However, this is not the case with another type of search, which we call a **themed search**, where we need to decide when the search process has ended.

In themed searches, the aim is to find information about a question or subject. We do not know beforehand what volume of information there is in the different information sources that we will be consulting or whether we will find all the relevant information or not. These types of searches are ones that require a more elaborate search strategy and that certain knowledge about the search process be put into practice, which we will explain later.

In any event, in themed searches it is the user who has to decide that the search is complete when the results obtained have satisfied their information needs.

1. The search process

As a general rule, any internet information search process must observe the following steps to ensure that it is efficient:

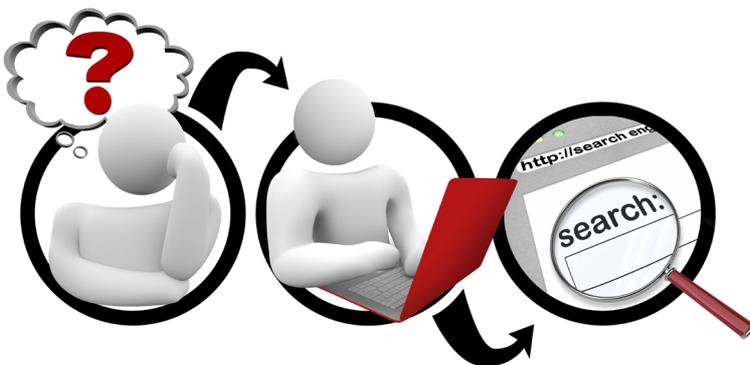
- 1) defining the consultation accurately
- 2) selecting the most suitable information sources
- 3) preparing the search and running it
- 4) assessing the relevance, quality and reliability of the information obtained

Obviously, the time and effort invested in each of these phases will depend on the type of search to be undertaken, and not all information searches require that all these steps be followed thoroughly. However, following this procedure as a matter of course when making searches on the Net tends to be a highly profitable way of investing time, while at the same time obtaining satisfactory results and also in terms of the speed of finding the necessary information.

1.1. Defining the consultation accurately

In the previous section, we said that the time invested in clearly defining and delimiting the information we need, why we need it and where we can find it is time well invested and that, very often, it saves time in the process spent searching for and finding information.

Before starting a search process, what we first need to do is to define our objective as specifically as possible. This means having a clear idea of what our information need is and this will enable us to design an optimum search strategy for each specific case and select the most suitable sources.



In this first phase, we need to ask ourselves questions, such as the amount of information that we would like to find, in which languages we would like to retrieve it, in which formats and, if we feel that our search needs to be mediated by a search engine, decide what the terms should be that we will be using for the search, ensuring that they are sufficiently representative and provide an adequate definition to identify the information we need.

1.2. Selecting the most suitable information sources

In this second stage, we need to ask ourselves what types of documents may contain the information we need to find. To do this, we need to know which information sources may give us access to these documents and, therefore, the information we want to find.

We should avoid making a list of the different information sources, defining them or classifying the information sources by type, as these classifications or definitions are of little help to the user when they are deep within a search process. However, we should once again stress that it is important before embarking unthinkingly on one of the so-called general search engines that users are aware that not all the information is visible through these types of search engines and that not only are these search engines useful in meeting our information needs, but they can also be useful when searching for sources where we can find the information that we need.

A specific example of this might be when we want to find some data and we believe that a statistics source or that of an official body would be a more appropriate and reliable way of finding it. In this case, if we did not know which specific source could provide us with the information, we would need to use a search engine to find and access the source, not to expect to find the specific piece of data.

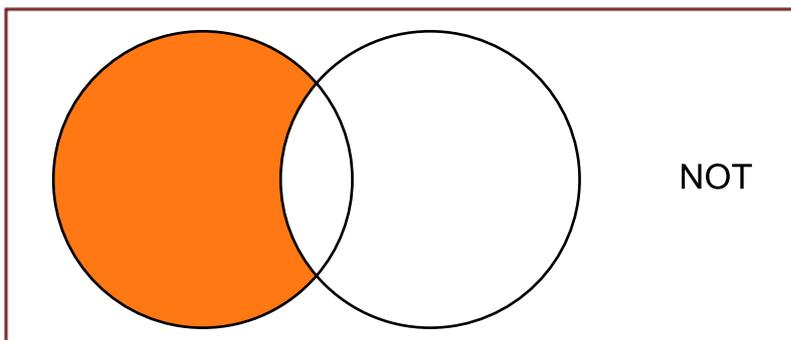
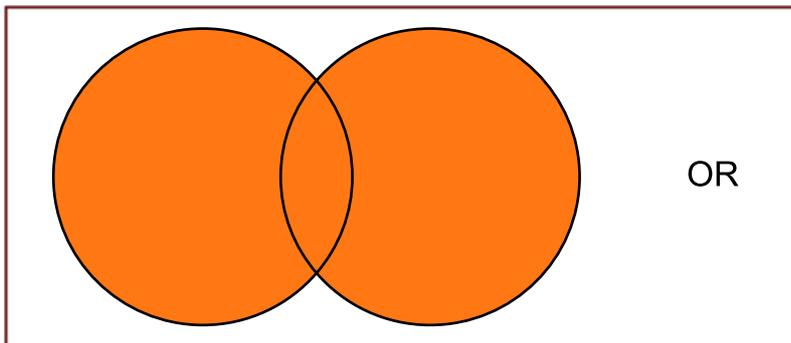
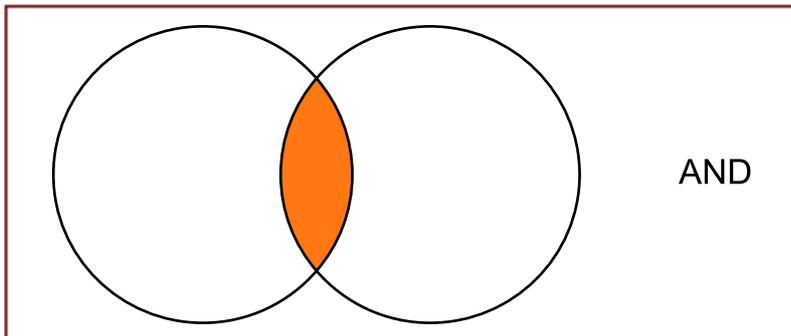
1.3. Preparing and running searches

It is when carrying out **subject searches** that users need to take time to prepare their search. They should consider, before carrying out a search on a source of information, whether they need to use operators, wildcards or masks, and which specific fields should be consulted. In short, the query should be translated into a documentary language and adapted to the source of information in question in order to obtain the best possible results.

Most sources of information (databases, electronic journals, library catalogues, etc.) use their own query or search software. Despite the differences there may be between them, most allow for searches of specific fields (author, title, subject, etc.), and for the focusing and adjusting of searches to meet specific needs through the combination of terms using Boolean operators (AND, OR and NOT) or truncation (with wildcards such as (*), (\$) or (?)).

Boolean operators are the logical operators AND, OR, and NOT. They are used when creating search strategies and allow for inclusion, exclusion or union of search terms. By adhering to set theory, we can combine different search

terms using three logical operations: conjunction (AND), disjunction (OR) and exclusion (NOT). This allows us to increase (OR) or limit (AND or NOT) the results obtained.



When we talk about **shortening**, we are referring to the use of symbols (*, \$, ?) that we usually put at the end or in the middle of a search term. By using shortening, we are able to search via the roots of words. Shortening means that we do not have to specify the different endings for the same word, which simplifies the search strategy. When we use these symbols in the middle of a word, they are called masks and prove **useful** when we are searching for terms with alternative pronunciations.

Examples of searches with Boolean operators

AND: This retrieves all the documents that contain both terms; for example: democracy and spain: this will return all the documents containing the terms "democracy" and "spain". Both terms must appear in the document.

OR: This retrieves documents which contain at least one of the search terms; for example: democracy or spain: this will return documents with "democracy" or "spain"; i.e. this will include documents with "democracy", or "spain", or both.

NOT: This operator ensures no documents are returned with the term entered; for example: democracy not spain: this will return all those documents containing the term "democracy" but not the term "spain".

Examples of searches with truncation

The most commonly used truncation symbols are the question mark (?), asterisk (*) and dollar sign (\$).

The **asterisk (*)** or **dollar sign (\$)** substitute any word ending, regardless of the number of characters this may represent. For example, if you search for "documen*", the results will include "documentation", "document", "documents", "documentary", etc.

The **question mark (?)** takes the place of any one single character in a given search term. It is very useful for finding terms that include the ñ, for example, when the keyboard being used does not have this character. For example, if you search for "m?cronutrients", the results will include both "macronutrients" and "micronutrients".

The truncation symbols (*) and (?) can be combined if the platform so allows. For example, if you search for "espa?ol*", the results will include "español", "española", "españoles", "españolas", etc.

Although the search interface of the different information sources that we use offers its own way of functioning and features, most systems use very similar techniques, mainly when offering the possibility of conducting what is known as an "advanced search". Knowledge of the characteristics specific to each information source and knowing how to use them is important when it comes to getting the most out of their possibilities and retrieving the most suitable information.

1.4. Assessing the relevance, quality and reliability of the information obtained

When a search is complete, i.e. once we have retrieved the information from the sources consulted, we need to assess the information obtained basically by following the three premises set out in the title of this sub-section: relevance, quality and reliability. In other words, we need to check whether the data or information that we have retrieved is useful or relevant depending on our information needs and whether it meets minimum quality and reliability standards, in short, whether the origin and authorship are reliable.



However, when assessing the search results, and with regard to the quality of the information obtained, we might encounter two basic problems: document silence or document noise.

- We talk of *silence* in a search when we do not retrieve all the relevant documents, i.e. when we have obtained fewer results than we could have. In this case, the search is incomplete. We need to consider whether the search strategy used is correct and whether the resources consulted are the most appropriate.
- We talk of *noise* in a search when we do not manage to eliminate non-relevant documents from the results, i.e. when we have obtained more results than we ought to. In this case, we need to filter the results to reduce their number and eliminate *document noise*. This can be achieved, for example, by expanding the number of search terms, using more precise terms and using operators.

In any event, when assessing the information obtained through a search process, as we have mentioned above, we need to remember that accredited, top-flight scientific documents that are hugely relevant to the research co-exist on the internet with pseudo-science and documents of very little or no informative value and which are not particularly well scrutinised. Consequently, we need to bear in mind that although it is very true that the internet provides free access to a whole wealth of documents, it is also true to a certain extent that the most valuable information in scientific terms is still not free to access and requires payment. This is what is known as the *invisible web*, which includes databases or scientific journals, which are essential to keep up-to-date with the advances made in the different fields of knowledge.

Bibliography

Català, C. (2007). *Introducció a les fonts d'informació*. Barcelona: UOC.

Gallego, J.; Juncà, M. (2009). *Fonts d'informació*. Barcelona: Universitat Oberta de Catalunya.

Ferran, N.; Pérez-Montoro, M. (2009). *Búsqueda y recuperación de la información*. Barcelona: Editorial UOC.

Maldonado, As. (coord.)(2006). *La información especializada en Internet* (2a. ed. corrected & augmented). Madrid: CSIC-CINDOC.

Torres, I. (coord.)(1998). *Las Fuentes de información: estudios teórico-prácticos*. Madrid: Síntesis.

