

# The search for scientific information

Anastássios Perdicóulis

Assistant Professor, ECT, UTAD (<http://www.tasso.utad.pt>)

Senior Researcher, CITTA, FEUP (<http://www.fe.up.pt/~tasso>)

Visiting Researcher, Oxford Institute for Sustainable Development, OBU, UK

## Abstract

Money aside, finding particular scientific information requires a methodic search with formal and controllable *queries*, as well as access to comprehensive and authoritative *registers* in relevant domains of specialisation.

## 1 Introduction

Scientific information suggests credibility (Perdicóulis, 2014f, 2015b) on the merit of meticulous research and writing (Perdicóulis, 2014e, 2012a). As an integral part of research, the *search* for scientific information (e.g. books, articles) inherits the obligation to be methodic (Perdicóulis, 2014d) — for instance, to be *thorough* (i.e. look in ‘all the right places’) and *critical* (e.g. judge for credibility, relevance, and significance). It is, therefore, a sensible concern to wonder just how ‘correct’ one can be when searching for scientific information.

## 2 Researchers, authors, and registers

In the simplest of cases (Perdicóulis, 2014a, § 2), the researcher places a direct query to the author of the information — Figure 1. This individualised search relies on personal communication skills and presupposes that the researcher already knows the author, that the author will reply thoroughly and promptly, and requires many similar contacts to make the search exhaustive.

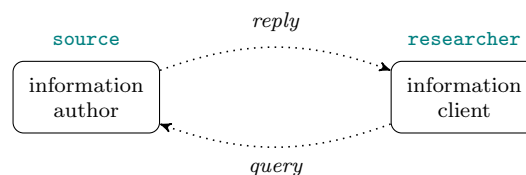


FIGURE 1 Direct search

At a higher order of organisation (Perdicoulis, 2014a, §§ 3–4), scientific information is published and subsequently registered with ‘curators’ or ‘brokers’ — for instance, publishers, libraries, and/or abstracting and indexing services (Perdicoulis, 2014b, 2015a). Despite the good intent and effort of the ancillary professionals, the items in their records are not to be unconditionally trusted for ‘scientific quality’, as this is not their assigned role — responsibility always lies with the researchers, whether authors or readers (Perdicoulis, 2013a, 2014b). Moreover, as registers shift from analogue to digital, researchers also need to update their query skills (Perdicoulis, 2014c,d) — Figure 2.

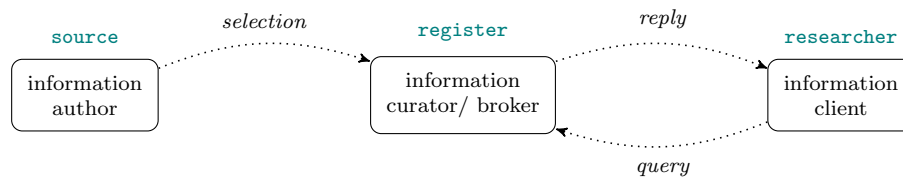


FIGURE 2 Search through a register

### 3 Queries and follow-up

Following the legacy of the analogue library registers (Perdicoulis, 2014c), queries are often based on ‘keywords’. Advanced digital registers, featuring rich meta-information (Perdicoulis, 2015c), allow for queries with specifiers between the relevant terms, such as Boolean operators or proximity indicators (Perdicoulis, 2014d) — Figure 3. Even more advanced queries use preferences<sup>1</sup> for the final sorting of the results, such as indications from the user’s search history or citation counts of the matching articles.

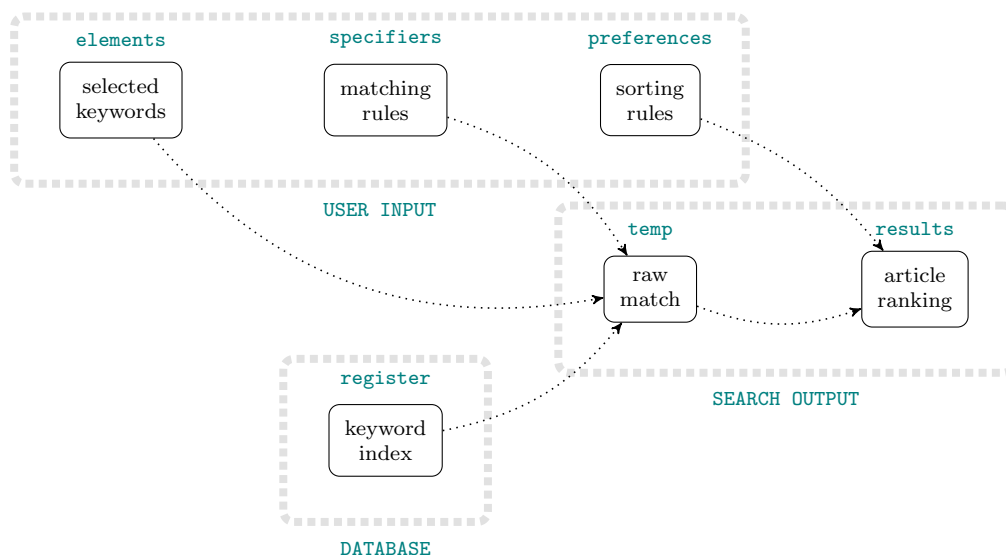


FIGURE 3 A keyword-based search in an indexed database

<sup>1</sup>Albeit typically undisclosed and/ or non user-controlled — for instance, as in the case of AI-based searches.

The sorted results of a search are practically ‘ready to use’ — for instance, proceed to the acquisition of the indicated articles in the order of their appearance, possibly conditioned by the premium associated with their access (Perdicoulis, 2014a, 2015a, 2018). Hence, the sorting rules are pivotal for the selection of the scientific information, despite the fact that they cannot easily incorporate the crucial issue of ‘scientific quality’ (Perdicoulis, 2013a, 2015b), and there lies grand responsibility — namely, dissociated from ‘scientific quality’, the apparent authoritativeness of the search results is liable to jeopardising scientific practice. Figure 4 illustrates how citation-sorting can lead to a ‘vicious cycle’ by accommodating false associations (---->) and user assumptions (Perdicoulis, 2012b, 2015a,b).

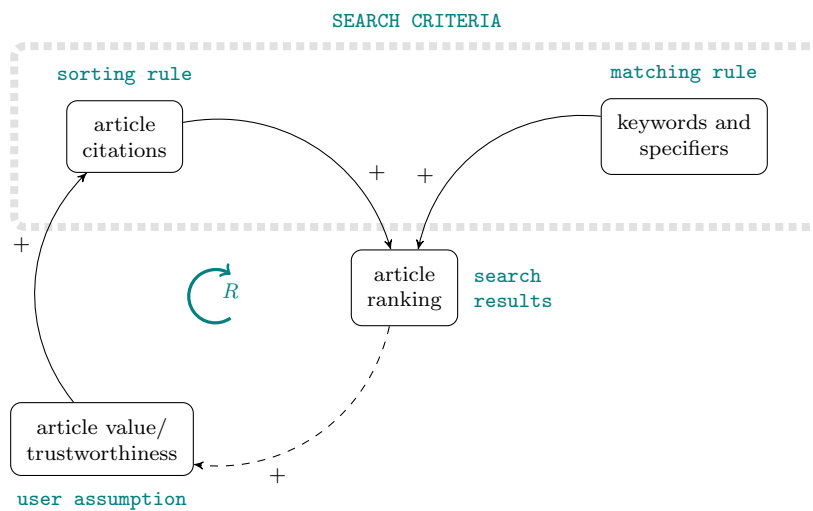


FIGURE 4 Vitiating search, primed by the sorting rule

Although the reasoning illustrated in Figure 4 is rooted in questionable scientific practice (i.e. the predominance of citations — Perdicoulis, 2012b, 2015b), the situation cannot be perceived or documented and the argument cannot be challenged (or defended) without *transparency* in the use of the sorting rules in the query. Furthermore, the practice itself cannot be changed without *user control* in the selection of sorting rules.

## 4 Discussion

The effectiveness of a search for scientific information starts with the good preparation of the researchers — for instance, to know (or at least to explore and understand) what they are looking for (e.g. mental model, objective, end result), and then express it formally in the terms required for each type of query, register, or search engine.

On the other side, registers (as well as publishers and authors) must prepare the scientific information in a way that it can be *searched* — for instance, by a *keyword search*, in the traditional way, or by a *semantic search* (e.g. through concept maps or concept association diagrams, indicating the contextual meaning of terms — Perdicoulis, 2014g). In addition, the correct *identification* of the information sources (e.g. articles, journals, books, websites, blogs) requires appropriate handling of *meta-data* by all parties involved, including the researchers, librarians, publishers, and registers (Perdicoulis, 2013b, 2014a, 2016, 2015c). The selection of the scientific sources to be included in the

registers must be done responsibly — for instance, registers must indicate their comprehensiveness and objectivity or partiality (e.g. associated publishers, excluded sources), as well as their position to ‘scientific quality’ (e.g. examination procedures, selection criteria, quality standards).

Correspondingly, queries must be at least *transparent* regarding the *coverage of sources* — for instance, mainstream or independent publishers; commercial or non-profit journals; automated or curated registers. For an even better conduct of a scientific search, queries must allow for *user control* in the rules to be used — namely, (a) the *matching rules* (e.g. keyword or semantic search, type or scope of sources), and (b) *sorting rules* (e.g. user profile or search history, relating to selection bias; keyword density or frequency, subject to author and/ or publisher manipulation; popularity such as view or citation counts, relating to social bias; randomisation, introducing entropy).

Finally, owing to budget restrictions, potential financial impediments to legally accessing scientific information may act as early-stage deterrents to researchers, rather than final-stage per-item filtering factors (Perdicoulis, 2018) — for instance, by excluding entire registers that are known to contain ‘expensive’ articles. In most cases, premature preclusion of sources is an excessive restriction to a thorough scientific search.

## 5 Challenges

Searching for scientific information requires the diligent effort of multiple parties: researchers must know well and express formally what is being sought, registers must be comprehensive, objective, and transparent in their selection of scientific sources, and search engines must have widely controllable search rules. The follow-up challenge to a good search lies in the step from ‘finding’ to ‘accessing’ the actual information, placed and confronted in the broader context of public-domain scientific research.

## References

- Perdicoulis, A. (2018) The public science paradox. *oestros*, **26**.
- Perdicoulis, A. (2016) *Documents*. Perdicoulis Publishing: Folio Division, Technical Collection.
- Perdicoulis, A. (2015c) *Metadata*. Perdicoulis Publishing: Folio Division, Technical Collection.
- Perdicoulis, A. (2015b) Scientific credibility. *oestros*, **21**.
- Perdicoulis, A. (2015a) The science marketplace. *oestros*, **19**.
- Perdicoulis, A. (2014g) *Language*. Perdicoulis Publishing: Folio Division, Technical Collection.
- Perdicoulis, A. (2014f) *Science*. Perdicoulis Publishing: Folio Division, Technical Collection.
- Perdicoulis, A. (2014e) *Research*. Perdicoulis Publishing: Folio Division, Technical Collection.
- Perdicoulis, A. (2014d) *Search*. Perdicoulis Publishing: Folio Division, Technical Collection.
- Perdicoulis, A. (2014c) *Cataloguing*. Perdicoulis Publishing: Folio Division, Technical Collection.
- Perdicoulis, A. (2014b) Curators of scientific publications. *oestros*, **18**.
- Perdicoulis, A. (2014a) Sharing the science. *oestros*, **16**.

Perdicoulis, A. (2013b) Shadow measurements. *oestros*, **9**.

Perdicoulis, A. (2013a) Educated readership. *oestros*, **7**.

Perdicoulis, A. (2012b) Recreating established systems. *Systems Planner*, **10**.

Perdicoulis, A. (2012a) Scientific writing. *oestros*, **5**.

