The World of Process Engineering and Biotechnology Information: Seven Points for Reflecting on Your Information Behavior

Thomas Hapke
University Library
Hamburg University of Technology (TUHH)
hapke@tuhh.de
https://www.tub.tuhh.de/en/thomas-hapke/
ORCID https://orcid.org/0000-0002-5135-2693

December 2017

Updated and modified version of a text which appeared as “Appendix I: The World of Biotechnology Information: Seven Points for Reflecting on Your Information Behavior” in (Buchholz, Kasche, and Bornscheuer 2012, 553–64). A former version appeared as “Appendix I: The World of Biotechnology Information: Eight Points for Reflecting on Your Information Behavior” in (Buchholz, Kasche, and Bornscheuer 2005, 419–26)

“It is crucial that biotechnology [and engineering] students are able to access the relevant information for their studies and can critically evaluate information and its sources. Information literacy is part of lifelong learning and prepares biotechnology [and engineering] graduates for their careers.” (Ward and Hockey 2007, 374)

1. Thinking About Your Information Behavior

Information literacy is a crucial key skill for self-directed learning in scholarly and professional everyday life. In addition to efficient retrieval and navigation strategies, it includes – above all – the creativity to organize and shape one’s own information process in a conscious and demand-oriented way. For the searcher, it is no longer questionable to find some information, but rather to filter reliable information from many similar offers. In a time, “where information and data are cheap, proliferating through digital environments and always at the end of a search engine query,” thinking critically about information includes “understanding the process through which truth become authenticated, and the underlying assumptions, values, biases, presuppositions and belief systems which inform that process” (Tredinnick 2008, 114).
Like every subject, process engineering and biotechnology have their own special information media, in addition to particular retrieval strategies, to meet the subject-related information needs. Which of the available databases match your specific needs and are reliable? It is generally important to consider, and focus on information under the aspect of reliability. There is a considerable range of publications, especially peer-reviewed journals, available that underwent a critical review process which should provide information of adequate quality and relevance. The so-called “invisible web” or “deep web” contains information sources that in some cases are not collected by most search engines such as Google™ – that is, it includes the content of special databases, websites secured by password access or only available in an intranet, and script-based websites.

There is a whole range of reasons for reading and informing for research: To provide you with ideas and enhance your creativity; to understand and be able to effectively criticize what other researchers have done in your subject; to broaden your perspective and view your work in context to others (direct personal experience is never enough); to legitimize your arguments; to avoid double efforts in research; “to learn more about research methods and their application in practice”; and to find new areas for research (Blaxter, Hughes, and Tight 2010, 100). Before beginning to search information, first reflect on your topic and specific information needs, gather background information – and then focus your research.

2. Subject Gateways, Tutorials and Literature Guides

Subject gateways on the net, tutorials and literature guides help to inform yourself about searching information. So-called “subject gateways” are good starting points for relevant web sites containing collections of subject-specific links. Two examples are the web site of the “Technische Informationsbibliothek (TIB) – German National Library of Science and Technology – Leibniz Information Centre for Science and Technology and University Library” at https://www.tib.eu/en/ and the web site of the “U.S. National Center for Biotechnology Information” (https://www.ncbi.nlm.nih.gov).

Web-based learning tutorials for information literacy in engineering can be used independent of time and space. One example of an online tutorial is “TUlib - Finding, Using and Judging Information” provided by the university library of the Dutch Delft University of Technology (http://ocw.tudelft.nl/courses/information-skills/tulib/). So-called short LibGuides as this one about “Chemical and Biomolecular Engineering Research Resources” of the University Library of the University of Illinois at Urbana-Champaign (http://guides.library.illinois.edu/chbe) as well as this handout and poster “Process Engineering Information @ TUHH” (Hapke 2017) give short overviews for using engineering information.
Printed literature guides still provide a comprehensive overview about all forms of primary sources like journal articles, reports, dissertations, patents or preprints and secondary literature (textbooks, monographs, encyclopedias, reviews, abstracting services etc.) of the treated subject. Guides like MacLeod and Corlett’s *Information Sources in Engineering* (MacLeod and Corlett 2005) and Osif’s *Using the engineering literature* (Osif 2012) contain chapters of specific value for the process and biochemical engineer. A guide for the life sciences comes from Schmidt (2014), for chemistry from Currano and Roth (2014). Overviews on research in a lot of disciplines including engineering and the physical sciences can be found in Keeran and Levine-Clark (2014).

### 3. Orientation Through Using Encyclopedias and Your Local Library


The easily accessible internet source Wikipedia covers a broad range of articles related to process engineering as well as biotechnology. The users should be aware that the information does not come from validated scientific sources. Hence, extensive coverage of a specific field, misleading or even wrong information might be included, which in addition can change rapidly. The reliability of the information found in Wikipedia can be evaluated by the literature it cites. The reliability should increase with the number of citations from peer-reviewed journals.

Even in the Internet age, a visit to the local university library can ease information retrieval. If they do not possess the item, you are interested in, library union catalogs offer a wide range of library materials that can be ordered through interlibrary loan or document delivery. In many countries special libraries function as a National Library for Science and Technology, (e.g., in Germany the TIB in Hannover; [https://www.tib.eu](https://www.tib.eu)). Databases available in the local intranet of universities and other institutions also provide references to further information.
(e.g. journal articles) not necessarily housed by the library itself. In addition subject librarians can provide information consulting.

4. **Playing with Databases and Search Terms**

When searching a database it is important to use appropriate key words that allow retrieval of the desired information. Too general key words lead to too many hits from which often only a fraction is useful; when using too specific key words, important information might not be found. It is also recommended to use logical, so-called Boolean operators (AND, OR, NOT) to link search terms and to use wildcard (joker) symbols (‘*’ or ‘?’ or ‘$’, which one depends on the search interface). For example, searching with ‘biodegr*’ retrieves documents containing ‘biodegradation or biodegradable or biodegraded or biodegradability or…’.

Today’s user interfaces allow so-called “facettet search” or “drilldowns” to reduce the number of results after searching. In addition, search results are arranged as default setting according relevance – in the past often ordered by descending date. Here you are faced with the difficulty that it is unknown how relevance is determined by the search engine or database.

A search term worksheet can help to structure your query and to find additional search terms like synonyms. For this, the topic must be divided into components and key words chosen for every component. For searching, terms in each of the worksheet’s columns have to be combined with the “OR”, the resulting sets with “AND”:

<table>
<thead>
<tr>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>microbi* degrad*</td>
<td>aromat*</td>
<td>soil*</td>
</tr>
<tr>
<td>biodegrad*</td>
<td>polyaromat*</td>
<td>clay*</td>
</tr>
<tr>
<td>bioremed*</td>
<td>benzene</td>
<td>compost*</td>
</tr>
<tr>
<td>microbe* decompos*</td>
<td>PAH</td>
<td>sediment*</td>
</tr>
</tbody>
</table>

So a search in a specific database drawn from the search term diagram above can look like: (“microbi* degrad*” OR biodegrad* OR bioremed*) AND (aromat* OR polyaromat* OR benzene OR PAH) AND (soil* OR clay* OR compost* OR sediment*)

“Playing” with a wide range of databases as well as with search terms and their synonyms in such a way still today may be essential to be more sure not to miss something important which has been published somewhere.
5. Searching Journal Articles, Patents, and Data

Today, most recent research results are published in scientific journals, seldom in subject-specific textbooks. More and more of these are integrated with databases and other digital media which become more accessible, sociable, and personalized (Hull et al. 2008). The difference between primary and secondary sources vanishes. However, it should be emphasized that sources that appeared in peer-reviewed journals should be preferred. A literature search (see above) provides rapid identification of specific journal articles, reviews and recent books. The original publications can then either be downloaded directly from publishers’ websites, for example also through link resolver, in addition through the university library homepage as gateway or are available as printed versions in the library (see also the “German Electronic Journals Library” at http://www.bibliothek.uni-regensburg.de/ezeit/ or the “DOAJ Directory of Open Access Journals” at http://www.doaj.org ). In addition, patents are an important and often less frequently used source by academia.

For searching specific papers, the net offers a huge diversity of databases. Free databases like Google™ Scholar often lead you to fulltexts. In case you are asked for a login or for your credit card number, remember that perhaps the library offers the print version of the article or a further e-version which is only available from another source through the local intranet.

Publishers’ portals like ScienceDirect (Elsevier), SpringerLink or the Wiley Online Library – only to name the three biggest publishers in science and engineering - offer fulltext searching for their own e-books and e-journals, a feature reference databases like Web of Science or INSPEC don’t do. The searchability of information depends on choices made by authors, publishers and database providers (Falciola 2009). Be prepared for change: Information sources on the net and their user interfaces are updated and enhanced constantly.

A range of specialized databases for information retrieval in chemistry, biotechnology and related fields, which are available for free either only within the intranet of universities or companies, is listed for example in a German-based “database of databases” called DBIS (Datenbank-Infosystem) at


Some databases worth mentioning are included in the following lists:
Reference databases for searching journal articles

- Web of Science with Science Citation Index at [http://wokinfo.com/](http://wokinfo.com/)
  This database, hosted by Clarivate Analytics, indexes leading science and technology journals as well as journals from social sciences and humanities. As interdisciplinary citation database you can search here with documents as “search terms”. It answers questions such as: Who have cited a specific document or a specific author? How often is a document cited? Coverage goes back to 1945.

- Scopus of the publisher Elsevier at [http://www.scopus.com](http://www.scopus.com)
  Interdisciplinary citation database, compared to Web of Knowledge it contained also proceedings and its contents as well as more publications in other languages than English.

  Interdisciplinary for medicine, also of great importance for biotechnology as it allows search of nucleotide or protein sequences, genome data and as it is linked to enzyme structure databases. Free access

- SciFinder at [http://www.cas.org/products/scifinder](http://www.cas.org/products/scifinder)
  This resource is the online version of Chemical Abstracts from the American Chemical Society (ACS), which indexes chemical research reported in journal articles, patents, symposia, conference proceedings, dissertations, technical reports, and new books. For all areas of chemistry, and related sciences like the materials sciences and the environmental sciences. Coverage goes back to 1907.

- TEMA® Technology and Management (wti Frankfurt) at [https://www.wti-frankfurt.de/en/](https://www.wti-frankfurt.de/en/)
  This database contains references with abstracts, keywords and descriptors on technology and management. It provides information from German and international scientific and practical technical literature like journals, conference proceedings, reports, dissertations, as well as non-conventional, so-called “grey” literature. Coverage goes back to 1968.

  Of importance because information technology plays a considerable role in all areas of engineering today.
Substance property databases
(more at https://www.tub.tuhh.de/en/substance-property-data/)

- ChemSpider at https://www.chemspider.com/
  ChemSpider extracts a wide range of information about chemicals and materials by crawling the Internet. Included are basic property information and a wide range of spectra, like IR, HNMR, mass spectra etc., a rich set of physical, biological, safety, and environmental data. ChemSpider can be searched by names, substructures, CAS registry numbers, and other identifiers.

  Created by US National Institutes of Health (NIH) to provide information about small organic molecules.

- NIST Chemistry Webbook at https://webbook.nist.gov/chemistry/
  Contains high quality, critically reviewed data compiled by the National Institute of Standards and Technology (NIST). Included are thermochemical, thermophysical, and ion energetics properties as well as reaction thermochemistry data and a variety of spectra, when available. The database is predominantly organic with a few small inorganic compounds.

- Physical Properties Sources Index (PPSI) at http://www.eqi.ethz.ch/en/
  Lists recommended databases, handbooks, and websites (data, definition, measurement) for physico-chemical and other material properties.

Patent databases
(more at https://www.tub.tuhh.de/en/patents/)

About 60 % of the information in patents is not published elsewhere. When searching in patent databases using notations of the International Patent Classification is a good tool to find subject specific patents that often are described in the title or the patent itself very generally. Some databases for searching patents as well as getting the full texts:

- DEPATISnet-the German patent information system at https://www.dpma.de/recherche/depatisnet/index.html.
  This German patent information system contains the full text of every German and American patent in pdf format, also patents from other countries. You have to know the exact patent number. Searching in other database fields – e.g., title, patent inventor, or abstract field – is possible from a distinct year. So you can search for
German patents in the title or inventor field from the year 1981. Free access.

- esp@cenet – European Patent Office at https://worldwide.espacenet.com/

Hazardous substances databases

- GESTIS Substance Database at http://www.dguv.de/ifa/gestis/gestis-stoffdatenbank/index-2.jsp
  Information system on hazardous substances of the German Social Accident Insurance. Free access
  Toxicological Data Network of the U.S. National Library of Medicine for searching information on toxicology, hazardous chemicals, environmental health, and toxic releases. Free access
- International Chemical Safety Cards (ICSCs): International Programme on Chemical Safety at https://www.cdc.gov/niosh/ipcs/default.html
  Available in a lot of languages. Free Access. Also available at https://www.bfr.bund.de/en/databases-573.html via the German Federal Institute for Risk Assessment (BfR)

6. After Searching: Evaluating and Processing Information

After searching successfully, you have to evaluate your search findings with respect to relevance. How to be sure, that all the potentially important documents are included in your resulting set? How to modify your query to reach this goal? But it is also important to evaluate critically the quality of the documents you have found. In case the document is published in a scholarly peer-reviewed journal, independent experts have evaluated the article before acceptance and publication. Who is the author of the text? What is his or her background? Why is the document being provided? How current is it?

The process of information retrieval interweaves more and more – especially in the digital age - with writing and communication processes (Cottrell 2008; Divan 2009; Hofmann 2010; Johnson and Scott 2009; Lebrun 2010; Macgilchrist 2014; Northey, Draper, and Knight 2015; Silyn-Roberts 2013). Networking and collaboration opportunities like weblogs, wikis, and
other tools of the “social web” are central themes today and enhance data sharing and new ways to stay current (Cann, Dimitriou, and Hooley 2011; Lumerman Oliver 2009). Subject-specific tools of the “social web” for the bioengineer include resources like https://openwetware.org, a wiki “for researchers and groups who are working in biology and biological engineering”, https://www.proteopedia.org, a 3D-encyclopedia of proteins and other molecules, or https://www.cazypedia.org, an encyclopedia of carbohydrate-active enzymes. A general engineering example is the Engineering and Technology History Wiki at http://ethw.org.

To keep yourself up-to-date, using RSS feeds (see https://en.wikipedia.org/wiki/RSS), which can be collected in feed readers, replace getting a free table of contents via e-mail from publishers of journals. Reading subject-specific weblogs or following a researcher via Twitter substitute subscribing to subject-specific mailing lists. Social networking site for scientists and researchers like ResearchGate or Academia.edu offer sharing papers, asking and answering questions, and finding collaborators. They more and more are also important to become visible as scientist and scholar. Each academic author should also have an ORCID (Open Researcher and Contributor ID, https://orcid.org/) as a nonproprietary alphanumeric code to be uniquely identified, similar to those created for objects on digital networks by digital object identifiers (DOIs).

Note-taking strategies today occur by using reference management software like Zotero, Jabref, Mendeley, Endnote or Citavi. Such software allows organizing references, quotations as well as full text files. Formatting texts in specific citation styles supports the publishing process.

In addition to impact factors which allow the evaluation and ranking of journals (Journal Citation Reports¹, http://www.eigenfactor.org) there exists further quantitative measures like citation rates and personal impact factors as the h-index to evaluate research through citation analysis (Bellis 2009; Gingras 2016) e.g. within Web of Science or Scopus. Software using Google™ Scholar as source for evaluation is also available (Harzing 2010). But “no matter how considerate and extensive a [bibliometric] evaluation is, it will be implemented only to the extent that it is in consonance with the prevailing power structure” at the commissioning higher education institution (Seglen 2003, 151).

¹ Look at https://clarivate.com/products/journal-citation-reports/
7. Information and the World

What is a publication – what is an author, a document, a journal, a collection, or a library? In the digital world of the net all of these terms have changed their meaning and use. Thinking about information is particularly of interest in biotechnology (Braman 2004). At a time when historians of science describe “… biology’s metamorphosis in an information science” (Lenoir 1999, 43), it is necessary to reflect about information and its communication and use (Feather and Sturges 2003; Bawden and Robinson 2012). Even new uses of the word biotechnology arise as the following citation shows: “I also would treat as biotechnology those affective technologies including so-called new media technologies that have permitted us to rethink the body in terms of digitization” (Clough 2007, 312).

Thinking on the interdependency of information and society today is necessary and well-established (Buckland 2017). Also a historical view can be useful to understand information and its infrastructure today (DeSart et al. 2017; Hapke 2008). In addition: As the citation from Tredinnick in the first part of this paper suggests, reflection about the sciences (Barker and Kitcher 2014; Gauch Jr 2012; Harker 2015; Pruza 2016) - about their characteristics and history, their theories, methods, purposes, visibility and their benefits - is important in a society today, whose present and whose future especially depends on science and technology and which is urged to develop sustainably.

In spite of information overload, only a limited part of information is freely available on the Internet. Access to commercial information sources for scholarly research such as reference databases and the full text of a specific journal is usually subject to a license fee and controlled by password. However, they often are offered within the intranet of universities or companies that have paid for the subscriptions.

Open access activities try to free access to scholarly publications at least for research and educational purposes (Suber 2012). Their aim is to make scholarly information freely accessible to all people worldwide, free of charge and as free as possible of technical and legal barriers. Examples are the journals from PLOS at https://journals.plos.org/ and from BioMed Central at https://www.biomedcentral.com. Research for books can be done in the Directory of Open Access Books at https://doabooks.org. Publishing Open Access occurs also through institutional Open Access repositories, like the one of the Hamburg University of Technology tub.dok at https://tubdok.tub.tuhh.de/?locale=en, or through a research data repository like Zenodo at https://zenodo.org. Beneath university publications like dissertations, Open Access repositories of universities can include also post-review versions of journal articles from most publishers, although they have been published in a subscription-based journal. (“green” Open Access). The likelihood to be read and then to be cited for Open Access publications is higher than for other articles which have not been offered as Open Access (Gargouri et al. 2010).
So called open educational resources (OER) are freely accessible, openly licensed materials - whatever text, media, digital or not - that are useful for teaching, learning, and assessing as well as for research purposes. OER are public domain or carry an open license like Creative Commons (https://creativecommons.org/). This guide of an Australian university library offers an overview for the subject biotechnology https://rmit.libguides.com/c.php?g=405195&p=4089965.

Data are the result of each research process in all subject areas. Primary or raw data can be very different (data of analyzing, results of polls or surveys, diagrams, photos, videos, drawings etc.), but most of them are available in digital format today. After finishing a research project, it seems to be useful to keep such data. They serve as reference for own publications and can be used in other projects. More and more making available research data is an important part of publishing which also brings reputation. Institutions like universities think about the effective publication and use of research data. Similar to journal articles yet today, research data can get so-called DOIs, Digital Object Identifier, to enable easily citing and linking. The TIB in Hannover gives more information about research data at https://www.tib.eu/en/publishing-archiving/research-data/.

Issues in intellectual property and copyright increase in a “cut-and-paste” environment. They are especially part of biotechnology research (Castle 2009). Why is it important to cite sources of information? What is the right way to cite? Questions of information ethics (Quinn 2017) like plagiarism as well as information policy (ownership, access, privacy) become important. Today open science, as "the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society, amateur or professional" (Wikipedia 2017), is propagated more and more (McKiernan et al. 2016; Bartling and Friesike 2014). Does there exist a digital divide? Even think of the preservation and long-term stability of information. What will be happening with digital records or data in 30 or 50 years’ time?

References


Bellis, Nicola de. 2009. *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Lanham, Md. Scarecrow Press.


Harzing, Anne-Wil. 2010. *The Publish or Perish Book: Your Guide to Effective and Responsible Citation Analysis.* Melbourne: Tarma Software Research Pty Ltd.


