Information Retrieval in Food Science Research: A Bibliographic Database Analysis

Tuba Karaarslan Urhan, Hannah Gascho Rempel, Lisbeth Meunier-Goddik, and Michael H. Penner

New Horizons in Food Research

Abstract: The aim of the present research was to ascertain the importance of electronic bibliographic database selection and multiple database usage during the information retrieval phase of research in the food sciences. Six commonly recommended databases were compared with respect to overall journal coverage and journal overlap. Databases were also evaluated with respect to coverage of food science-based journals and the extent of article coverage therein. A case study approach, focused on bile acid/dietary fiber interactions, was used to illustrate the ramifications of database selection/usage when dealing with specific research topics. Databases differed with respect to the breadth of disciplines covered, the total number of journals indexed, the number of food science discipline-specific journals indexed, and the number of articles included per indexed journal. All of the databases contained citations that were unique to the given database. The data resulting from the case study provide an example of the extent to which relevant information may be missed if pertinent databases are not mined. In the present case, over half of the articles retrieved on the focus research topic were unique to a single database. The combined data from this study point to the importance of thoughtful database selection and multiple database usage when comprehensively assessing knowledge in the food sciences.

Keywords: food sciences, research, databases, information retrieval, dietary fiber

Practical Application: This paper provides insights into article database usage for food science-relevant information retrieval. Online information retrieval is an efficient way to assess current knowledge in any of the food science disciplines. Acquired knowledge in turn is the underpinning of effective problem solving; whether it be private sector- or academic/government-based research.

Introduction

Information retrieval is a necessary and critical component of modern research. Online information is available through many platforms, including postings on private/public web pages, social media, blogs, and bibliographic databases (Stanbury & Selman, 2008). Researchers typically prefer to work with peer-reviewed publications and thus their primary sources for information gathering are electronic databases that index such documents. Online database usage is thus integral to the research process; it is essentially impossible to stay abreast of relevant literature manually. Information retrieval, via "literature searches" done toward the beginning of a research project, is typically aimed at getting an overview of current knowledge, generating problem solving ideas, determining the novelty of experimental approaches, or identifying experts in a given research field. As a research project progresses, literature searches tend to focus more on specific aspects of a project (for example, alternative analytical methods) or to obtain peripheral supporting information (Hart, 2001).

In most cases researchers aim, at least in the beginning, to perform an overall assessment of the literature pertaining to a particular topic. Literature searches with this objective can be daunting due to the overwhelming amount of information available. A 2016 online ranking of journals covering most academic disciplines listed 23,226 journals (see Scimago; www.scimagojr.com). Those journals combined contained over 2.3 million documents in 2016 alone. When the subject area was narrowed to "Food Science," the same survey listed 260 journals; those journals included 32,755 documents in 2016. The amount of published information pertaining to the food sciences is obviously staggering, particularly if one includes the more fundamental subject areas upon which the food sciences are based. Harnessing this information requires thorough and efficient information gathering, which is enabled by electronic bibliographic databases. These databases are valuable links between authors, who have generated knowledge, and information seekers who are trying to access such knowledge. Effective use of these databases requires a general understanding of database construction, familiarity with the mechanics of searching databases, and an awareness of approaches for achieving maximum benefit from online database searching (Booth, 2016; Jensen, Saric, & Bork, 2006).

Primary questions related to information retrieval are "What database(s) is best for finding information pertaining to particular subject areas?" and "To what extent is it necessary to access multiple databases when assessing current knowledge in a given field?" The trivial answer to the latter would be to search all databases, but searching databases is time consuming and, at some point, further searches bring diminishing returns (Stevinson & Lawlor, 2004). Taking both the time commitment and the anticipated diminishing returns into account, it is reasonable to assume that researchers will access multiple databases to the extent they feel it productive. Our informal surveys aimed at verifying this assumption suggest that many doing research in the food sciences tend to rely on a single preferred database,

JFDS-2018-0210 Submitted 2/6/2018, Accepted 10/10/2018. *Authors Urhan, Meunier-Goddik, and Penner are with Dept. of Food Science and Technology, Oregon State Univ., Corvallis, OR 97331, U.S.A. Author Rempel is with Valley Library, Oregon State Univ., Corvallis, OR 97331, U.S.A. Direct inquiries to author Penner (E-mail: mike.penner@oregonstate.edu).*

doi: 10.1111/1750-3841.14388

Journal of Food Science
A Publication of the Institute of Food Technologists

without giving much consideration to the benefits of accessing additional databases; and that the preferred databases differ among food science professionals. Recognition of these inconsistencies led to the work outlined in this manuscript. Herein we address the relevance of database selection and multiple database usage by considering the retrieval of food science-related publications in general. Search concepts are then demonstrated through a case study of information retrieval for a specific, currently pertinent, research topic: "in vitro bile acid binding properties of dietary fibers." The focus throughout the manuscript is on databases commonly recommended for research in the food sciences, and results are presented in the context of common database usage.

Background

The nature of electronic bibliographic databases

A database is a collection of information typically organized for efficient utilization. Electronic databases present data in digital/electronic formats, such that the information can be accessed and manipulated using the appropriate computer software. A bibliographic database contains bibliographic information; for a journal article, this commonly includes the paper's title, authors and year of publication, the title of the journal in which it was published, volume and issue of the journal, and page numbers. Bibliographic databases may also include the article's abstract; such databases are formally referred to as abstracting databases. Yet other databases may contain an electronic version of the full text of the article; such databases are formally referred to as full-text databases. Bibliographic databases may also be referred to as indexes. The term index is useful in that it points to the role of these databases, which is to help the user find information that may be distributed in many different sources such as journals, theses, government documents, or conference proceedings—just as a book's index helps the reader find information distributed throughout that book. Bibliographic databases or indexes are thus essential tools for those trying to assess current knowledge in any given subject area.

Accessing bibliographic databases

Bibliographic databases can be accessed through online commercial vendors, through major academic, public or private libraries, or directly via the database's website. Some databases are proprietary and thus require a licensing agreement for access. Licensing agreements may be directly between a researcher and the database provider or, more commonly, researchers take advantage of licensing agreements between their library and database providers. Some databases are in the public domain and are thus open access, requiring no license. Database access, through commercial vendors or through libraries, may be granted for the use of single or multiple databases. Web of Science, for example, can be accessed through the provider's website or via any one of many academic libraries. Once Web of Science is accessed, one can search the databases included in the individual or institutional licensing agreement, which at a minimum, will include the "Core Collection." The Web of Science Core Collection itself includes multiple databases; the most relevant of which to the food sciences is likely Science Citation Index Expanded. EBSCOhost is another online information retrieval system similar to Web of Science in the sense that it can serve as a gateway to many proprietary databases, some of which are compiled by EBSCO itself (for example, Academic Search Premier). But EBSCO also provides access to content from other database vendors via the EBSCO platform (for example, Food Science & Technology Abstracts). A food science-relevant example of an open access portal to powerful databases is PubMed; which is developed and maintained by the National Center for Biotechnology Information (NCBI) at the US National Library of Medicine (NLM), which is itself a part of the National Institutes of Health (NIH). Within the PubMed umbrella, MEDLINE (NLM's citation/abstract bibliographic database) is the most relevant to the subject matter of this paper.

One can often access impressive numbers of distinct databases when working through academic institutions. Our recent survey of representative US academic libraries indicated Oregon State Univ. (OSU) provides access to 316 bibliographic databases; Cornell Univ., 733; Univ. of California at Davis, 1,249; and Harvard Univ., 5,392. These databases cover a wide range of subjects, ranging from the sciences, engineering, and technology to the humanities, business, and law. When considering academic institution-accessible databases, it is important to recognize that some of the databases may be accessible to the general public, either onsite or via the internet, while others will be accessible only to those formally associated with the institution (one typically needs to "login" to access these). For example, of the 316 databases accessible via the OSU Libraries only 83 are "open access" (available to the general public without a subscription fee) and 233 are "OSU restricted." In the latter case, a user must login using an OSU ID to obtain the right to access the databases to which the library has paid subscription fees.

Choosing an appropriate bibliographic database(s)

An early step in the information retrieval process is the selection of appropriate databases. That decision can involve consideration of hundreds, or even thousands, of choices. Most researchers are likely to access databases through academic library "database hubs." Such sites will typically provide guidance on making database choices; for example, they may provide a list of the most commonly used databases, or identify databases as "starting points," or provide an opportunity to search available databases based on subject area. The "most common" and "starting point" databases are often multidisciplinary (also referred to as "encyclopedic"; Gasparyan, Ayvazyan, & Kitas, 2013); more specific databases are sometimes categorized as "specialized" or "narrow-specialized" (Gasparyan et al., 2016).

Typically, a library website will have a "search by subject" link that allows filtering of accessible databases based on the user's subject area or topic of interest. Once filtered in this way, the suggested databases may contain a combination of encyclopedic and specialized databases. For example, the Oregon State Univ. Libraries returns 14 "Best Bet" databases for those seeking information in the area of "Food Science and Technology." That Food Science and Technology-relevant databases include a number of choices is not surprising, since this subject area is defined broadly, including food production, storage, processing, distribution, preparation, consumption patterns, nutrition, chemistry, engineering, microbiology, and so on (Duran & McDonald, 2006). Common library-suggested databases for information retrieval in the "food sciences" subject field are listed in Table 1 (based on our survey of 10 representative academic libraries; these libraries were chosen based on their campuses having widely recognized academic programs in the food sciences). Some of these databases are only accessible to licensed users, others are open access. All provide bibliographic information, abstracts, and links to open access journals; some provide a gateway to full-text articles that are not open access. Seven of the nine databases listed in Table 1 are accessible via the authors' campus library (Oregon State Univ. Valley Library); the characteristics of 6 of those databases

Table 1–Commonly recommended bibliographic databases for information retrieval in the "food sciences" subject area.[1]

Academic Search Premier (E,I,J)[2]
AGRICOLA (A,B,C,D,E,F,H,I,J)[3]
Biological Abstracts (A,B,C,D,F,J)[4]
CAB Direct (A,B,C,D,E,G,H,I,J)[5]
Food Science and Technology Abstracts (A,B,C,D,E,F,H,J)
PubMed (A,B,D,E,F,G,I,J)[6]
SciFinder (A,B,D,E,F,H,I,J)
Scopus (D,I,J)
Web of Science (A,D,E,F,G,I,J)

[1]Web pages from campus libraries at 10 representative US universities with established food science programs were accessed to ascertain databases commonly recommended for information retrieval in the subject field "food sciences." The "food sciences" subject area at different libraries was identified as "food science(s)," "food science and nutrition," "food and nutrition," or "food science and technology."
[2]In all cases, parenthetical letters indicate the universities recommending particular databases; (A) Cornell Univ., (B) Michigan State Univ., (C) North Caroline State Univ., (D) Ohio State Univ., (E) Oregon State Univ., (F) Pennsylvania State Univ., (G) Univ. of Georgia, (H) Univ. of Massachusetts Amherst, (I) Univ. of Minnesota, and (J) Univ. of Wisconsin–Madison. To be included in the list above, a database must have been suggested by a minimum of 3 of the 10 universities.
[3]Acronym for "AGRICultural OnLine Access"
[4]Recommended as "Biological Abstracts" or "BIOSIS Previews;" the latter combines Biological Abstracts and Biological Abstracts/Reports/Reviews/Meetings.
[5]"CAB Direct" is a database platform providing access to CABI (Centre for Agriculture and Biosciences International) database subscriptions, including CAB Abstracts and CAB Abstracts Archives.
[6]Recommended as PubMed or MEDLINE; the former includes MEDLINE, PubMed Central (PMC) and National Center for Biotechnology Information (NCBI) Bookshelf.

are presented in Table 2 (although accessible via the authors' campus library, SciFinder Scholar was omitted from subsequent analyses based on our inability to do refined searches in this database using methods analogous to those used for the other six databases).

Methods

Database journal coverage and extent of journal overlap between databases

The number of journals indexed by specific databases (Table 3) was determined based on the number of unique ISSN numbers cited by the databases; the total number of database-declared journals was obtained via the database's website, by direct communication with database personnel, or from information provided by Oregon State Univ.'s Valley Library. In some cases the database's list of indexed documents included bulletins, magazines, patents, reports, conference proceedings, and so on; in such cases all but peer-reviewed journals were eliminated. Extents of journal overlap between databases was determined by comparison of indexed p-ISSN (print-ISSN) numbers. These comparisons were done within Excel files using Statistical Analysis System (SAS) PROC SORT, with the NODUPKEY option.

Database coverage of selected food science-based journals

Twenty-four well-recognized food science-based journals were included in this phase of the study (journal titles listed in Table 4 and 5). Journals were chosen based on being representative of established journals that cover (a) all aspects of food science, (b) specific academic disciplines within the food sciences, or (c) commodity-specific food science research/technology. Whether or not databases index specific journals was determined by searching each online database using journal titles and pISSNs. To determine the number of articles indexed from a given journal for a given year an initial search of each database was done to retrieve all articles from the journal of interest; the retrieved articles were then filtered by year. This evaluation included all 24 journals for the following years: 1976, 1996, and 2016 (a span of 40 years; see Table 5). PubMed, Web of Science, and CAB Direct were accessed via their individual database platforms; AGRICOLA, FSTA, and Academic Search Premier were accessed via the EBSCOhost platform.

Case study

A case-study approach was used to provide an illustrative example of the importance of using multiple databases for information retrieval. The case-study approach allows the focus to be on a specific research topic; the topic chosen for this phase of the study was "in vitro bile acid binding properties of dietary fibers." The six databases characterized in Table 2 were searched in an analogous manner to determine the number of indexed articles each contained that specifically dealt with this research topic. The keyword search protocol for article retrieval from each database, including Boolean operators, is given in Table 6. Appropriate keywords were chosen based on preliminary reviews of titles, abstracts, and keywords of pertinent papers. Searches were done via Oregon State Univ.'s Valley Library subscription access during the period October 16 through October 20, 2017. Searches were limited to English text, peer-reviewed, academic, journal articles. Recovered articles were processed using the reference/citation management software Zotero. Zotero, as a citation management tool, allows the user to initially catalog citations retrieved from databases and then do keyword searches, title comparisons, and so on of cataloged citations (for reference, alternative citation management tools include EndNote and Mendeley).

Overlap between retrieved articles from different databases was determined by doing title comparisons using Zotero and manual inspection. Pairwise comparisons were made based on traditional overlap (TO) and relative overlap (RO); as defined below (Hood & Wilson, 2003).

$$\mathrm{TO}\ (\%) = 100 \times \left(\frac{|A_{\mathrm{Records}} \cap B_{\mathrm{Records}}|}{|A_{\mathrm{Records}} \cup B_{\mathrm{Records}}|} \right)$$

$$\mathrm{ROA}\ (\%) = 100 \times \left(\frac{|A_{\mathrm{Records}} \cap B_{\mathrm{Records}}|}{|A_{\mathrm{Records}}|} \right)$$

$$\mathrm{ROB}\ (\%) = 100 \times \left(\frac{|A_{\mathrm{Records}} \cap B_{\mathrm{Records}}|}{|B_{\mathrm{Records}}|} \right)$$

where "A_{Records}" and "B_{Records}" are the two data sets being compared (that is, the sets of retrieved articles/records from database A and database B, respectively). Percent TO (%) values indicate the extent of overlap in terms of the combined databases; ROA (%) and ROB (%) values indicate the extent of overlap in terms of ("relative to") the individual databases, either "A" or "B". The symbol "∪" denotes the union of two sets; the symbol "∩" denotes the intersection of two sets.

Results and Discussion

In the first inquiry, six of the databases from Table 1 were searched in an analogous manner to quantify the number of unique academic journals associated with each database and to determine the extent of journal overlap between databases. The search was based on journal ISSN numbers and refined as described in "Methods." The numerical values in Table 3 demonstrate the disparity in the size of the recommended databases and the extent of journal overlap. The values in the main diagonal (M*i*,*i*) of Table 3 indicate

Table 2–Characteristics of selected commonly recommended bibliographic databases for information retrieval in the "food science" subject area.[1]

	Databases					
Attributes	**Academic Search Premier[2] (ASP)**	**AGRICOLA[2]**	**CAB Direct**	**Food Science and Technology Abstracts[2] (FSTA)**	**PubMed**	**Web of Science (WoS)**
Web Address	https://www.ebsco.com/products/research-databases/academic-search-premier	https://agricola.nal.usda.gov	https://www.cabdirect.org	https://www.ifis.Org/fsta	https://www.ncbi.nlm.nih.gov/pubmed	https://webofknowledge.com/
Managing Entity (country)	EBSCO Publishing (US)	National Agricultural Library (NAL), United States Department of Agriculture (US)	The Centre For Agriculture and Biosciences International (CABI) (UK)	International Food Information Service (IFIS) Publishing (UK)	National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), (US)	Clarivate Analytics (US)
Open Access or Subscription Based	Subscription	Open Access	Subscription	Subscription	Open Access	Subscription
Databases Covered	ASP Database	NAL Catalog	CABI Databases[3]	FSTA Database	MEDLINE, OLDMEDLINE, PubMed Central, NCBI Bookshelf	Web of Science Core Collection[4]
Update Frequency	Daily	Daily	Monthly, Random	Weekly	Daily	Daily
Subject Areas	Life Sciences, Physical Sciences, Social Sciences, Humanities, Engineering, Technology	Agriculture and Allied Disciplines	Applied Life Sciences	Food Science and Technology, Human Nutrition	Biomedical and Life-Science Related	Sciences, Technology, Social Sciences, Arts and Humanities
Covered Items Include[5]	BK, CP, GD, J, L, M, MB, N, R, SP, V	B, BK, CP, D, EM, J, M, MA, P, R	BK, CP, D, J, MA, MB, P, R, S	BK, CP, D, J, L, P, R, S	B, BK, CP, CT, D, J, N, P, R	B, BK, CP, DR, EM, J, MA, P, R

[1]Information gathered from database websites (provided above) and links therein, January 2018.
[2]Database was accessed via the EBSCOhost portal
[3]Vendors provide options with respect to CABI databases included in subscriptions
[4]Subscription flexibility allows additional databases to be added to the Web of Science Core Collection. The Web of Science Core Collection per se consists of the following 10 indexes: Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Emerging Sources Citation Index, Conference Proceeding Citation Index-Science, Conference Proceeding Citation Index-Social Sciences and Humanities, Book Citation Index-Science, Book Citation Index-Social Sciences and Humanities, Current Chemical Reactions, and Index Chemicus. Further details of index coverage can be found at the following link: http://images.webofknowledge.com.ezproxy.proxy.library.oregonstate.edu//WOKRS530JR6/help/WOS/hp_database.html).
[5]'B,' Bibliographies; 'BK,' Books; 'CP,' Conference Proceedings; 'CT,' Clinical Trials; 'D,' Dissertations and Theses; 'DR,' Database Reviews; 'EM,' Editorial Materials; 'GD,' Government Documents; 'J,' Journals; 'L,' Legislation/Law; 'M,' Monographs; 'MA,' Meeting Abstracts; 'MB,' Magazines/Bulletins; 'N,' Newspaper Articles; 'P,' Patents; 'R,' Scientific, Technical, Industrial and/or Educational Reports; 'S,' Standards; 'SP,' Speeches; 'SR,' Software Reviews; 'V,' Videos; lists of covered items are based on database declarations and may not be exhaustive.

Table 3–Journal coverage and extent of journal overlap for selected databases.[1]

Databases[2,3] → ↓	WoS	ASP	PubMed	CAB Direct	FSTA	AGRICOLA
WoS	19,312					
ASP	6,880	14,596				
PubMed	5,893	3,094	13,036			
CAB Direct	3,724	2,627	2,252	7,476		
FSTA	673	437	385	762	1,028	
AGRICOLA	633	380	330	576	225	834

[1]Values in the main diagonal cells refer to the total number of journals indexed by the databases corresponding to those cells; these values are based on the number of database-declared journals having unique ISSN (see "Methods"). Numbers in sub-diagonal cells refer to the number of journals that are indexed in both of the databases corresponding to those cells (that is, the number of journals that are common to both databases and thus represent the absolute extent of journal overlap for the 2 databases); sub-diagonal values were obtained by overlap analysis as described in "Methods."

[2]Databases in column one are listed in descending order of predominance by number of journals indexed; acronyms are as described in Table 2.

[3]Numbers of journals were obtained from database sources. The journal list for WoS was obtained via the website of Clarivate Analytics in October 2017 (http://mjl.clarivate.com); that for ASP was downloaded from the website of EBSCO*host* in October 2017 (https://www.ebsco.com/products/research-databases/academic-search-prem%C3%ADer); that for MEDLINE and PMC was obtained by direct communication with U.S National Library of Medicine personnel in October 2017 (numbers reflect current plus former journals; this value can also be retrieved through "journalspmc[All Fields] OR reportedmedline[All Fields]" from NLM Catalog online); that for CAB Direct was obtained via direct communication with CABI personnel in September 2017; that for FSTA was obtained via direct communication with IFIS personnel in October 2016; that for AGRICOLA was obtained from Oregon State University Library records in April 2016. Numbers obtained from database sources were refined by removal of entries for non-journal listings, those without p-ISSN designations, and replicate p-ISSN designations. Thus, values reported in the table are for peer-reviewed journals with p-ISSN designations.

Table 4–Coverage by different databases of selected food science journals.[1,2]

Journal Title[3]	p-ISSN[4]	Databases					
		ASP	AGRICOLA	CAB Direct	FSTA	PubMed	WoS
AM J ENOL VITICULT	0002-9254	**N**	Y	Y	Y	**N**	Y
CEREAL CHEM	0009-0352	**N**	Y	Y	Y	**N**	Y
FOOD CHEM	0308-8146	Y	Y	Y	Y	Y	Y
FOOD HYDROCOLLOID	0268-005X	Y	Y	Y	Y	**N**	Y
FOOD MICROBIOL	0740-0020	Y	Y	Y	Y	Y	Y
FOOD RES INT	0963-9969	Y	Y	Y	Y	Y	Y
FOOD STRUCT	2213-3291	**N**	**N**	**N**	Y	**N**	**N**
HORTSCIENCE	0018-5345	Y	Y	Y	Y	**N**	Y
J AGR FOOD CHEM	0021-8561	**N**	Y	Y	Y	Y	Y
J DAIRY SCI	0022-0302	Y	Y	Y	Y	Y	Y
J FOOD BIOCHEM	0145-8884	**N**	Y	Y	Y	**N**	Y
J FOOD COMPOS ANAL	0889-1575	Y	Y	Y	Y	**N**	Y
J FOOD ENG	0260-8774	Y	Y	Y	Y	**N**	Y
J FOOD PROCESS PRES	0145-8892	**N**	Y	Y	Y	**N**	Y
J FOOD PROTECT	0362-028X	Y	Y	Y	Y	Y	Y
J FOOD SAFETY	0149-6085	Y	Y	Y	Y	**N**	Y
J FOOD SCI	0022-1147	Y	Y	Y	Y	Y	Y
J FOOD QUALITY	0146-9428	**N**	Y	Y	Y	**N**	Y
J AM SOC BREW CHEM	0361-0470	**N**	Y	Y	Y	**N**	Y
J SCI FOOD AGR	0022-5142	Y	Y	Y	Y	Y	Y
LWT-FOOD SCI TECHNOL	0023-6438	Y	Y	Y	Y	**N**	Y
MEAT SCI	0309-1740	Y	Y	Y	Y	Y	Y
POSTHARVEST BIOL TEC	0925-5214	Y	Y	Y	Y	**N**	Y
POTATO RES	0014-3065	Y	Y	Y	Y	**N**	Y

[1]Y = yes, the journal is covered in specified database; N = no, the journal is not covered in specified database; whether or not journals are indexed by given databases is based on journal lists described in Table 3.

[2]In all cases the "N" designation indicates that at the time of this study the indicated database did not declare it indexed the specified journal. However, in some cases articles from "non-indexed" journals can be found in databases due to inclusion for a variety of reasons. For example, NIH-funded studies published in a journal not typically indexed by PubMed may still be found in the PubMed database. This anomaly is relevant to the data presented in Table 5.

[3]Journal abbreviations as indicated at "Web of Science – Journal Title Abbreviations" (https://images.webofknowledge.com/images/help/WOS/A_abrvjt.html)

[4]p-ISSN = print-ISSN.

the number of journals indexed by the different databases. The sub-diagonal values ($M_{i,j}$) correspond to the number of journals common to "*i*" and "*j*" databases (that is, journal overlap). The number of journals covered by WoS approaches 20,000; it was the largest database considered in this study. AGRICOLA was the smallest database considered, at least based on number of journals indexed, with 834. Clearly, the larger databases index many journals not directly dealing with the food sciences. These journals should not be summarily ignored, however, because they may contain articles with information germane to food systems. When considering overlap it is useful to question the merit of using a larger diversified database compared with a smaller specialized database, for example, WoS or CAB Direct compared with FSTA. The data in Table 3 help to answer this question. The overlap in FSTA by WoS is 66% (that is, 66% of the journals indexed in FSTA are also indexed in WoS); the analogous overlap for FSTA by CAB Direct is 74%. The latter example indicates that searching CAB Direct allows access to approximately three-fourths of the journals indexed in FSTA, while simultaneously allowing access to a large number of additional journals that possibly contain articles of interest. Questions like "How important is it to access the journals unique to FSTA?" and "What is the likelihood of missing important papers by searching the specialized database (FSTA) rather than the diversified database (CAB Direct)?" become

Table 5–Number of articles indexed from selected journals by different databases.[1,2]

Journal Titles[3]	Total Citations[4,5] ASP	AGRI COLA	CAB Direct	FSTA	Pub-Med	WoS	1976 Citations[4,6] ASP	AGRI COLA	CAB Direct	FSTA	Pub-Med	WoS	1996 Citations[4,6] ASP	AGRI COLA	CAB Direct	FSTA	Pub-Med	WoS	2016 Citations[4,6] ASP	AGRI COLA	CAB Direct	FSTA	Pub-Med	WoS
AM J ENOL VITICULT	NA	1,786	1,702	1,934	1	2,551	NA	–	21	30	–	35	NA	62	43	43	–	64	NA	–	53	38	–	50
CEREAL CHEM	NA	4,467	4,390	5,561	NA	5,733	NA	–	76	118	NA	117	NA	142	59	138	NA	139	NA	90	93	89	NA	92
FOOD CHEM	19,014	19,659	20,188	21,307	10,152	21,454	–	–	5	11	–	–	–	87	128	199	–	208	2,032	2,055	1,919	1,905	1,922	1,895
FOOD HYDROCOL-LOID	3,915	4,028	3,666	4,358	13	4,408	–	–	–	–	–	–	–	57	11	52	–	57	541	520	504	502	2	494
FOOD MICROBIOL	2,564	2,898	2,691	3,133	1,968	2,847	–	–	–	–	–	–	–	56	32	49	–	56	198	168	167	167	147	161
FOOD RES INT	5,343	3,856	5,030	5,751	784	5,264	–	–	–	–	–	–	–	–	50	109	–	92	432	426	385	385	117	356
FOOD STRUCT	NA	141	NA	210	NA	NA	NA	–	NA	–	NA	NA	NA	–	NA	–	NA	NA	NA	–	NA	13	NA	NA
HORTS-CIENCE	4,723	5,061	13,635	2,251	61	10,871	–	–	240	43	–	99	–	232	282	43	4	280	259	–	226	63	–	231
J AGR FOOD CHEM	NA	30,775	34,778	28,297	28,638	37,498	NA	–	245	206	306	288	NA	695	492	556	–	734	NA	1,086	1,080	826	1,105	1,067
J DAIRY SCI	10,429	13,026	34,657	7,256	18,075	21,102	–	–	338	67	186	261	–	282	282	120	223	280	978	4	938	352	940	925
J FOOD BIOCHEM	NA	1,349	1,091	1,602	2	1,714	NA	–	–	–	–	–	NA	37	–	40	–	44	NA	51	79	75	1	77
J FOOD COMPOS ANAL	2,147	901	2,161	2,219	26	1,780	–	–	–	–	–	–	–	–	25	26	–	–	168	–	137	136	2	134
J FOOD ENG	6,728	7,527	6,063	7,723	3	7,371	–	–	–	–	–	–	–	143	1	118	–	117	405	376	378	373	1	363
J FOOD PROCESS PRES	NA	1,601	1,804	2,518	2	2,533	NA	–	–	–	–	–	NA	35	–	34	–	35	NA	72	155	155	–	154

(*Continued*)

New Horizons in Food Research

Table 5–Continued

Journal Titles[3]	Total Citations[4,5]						1976 Citations[4,6]						1996 Citations[4,6]						2016 Citations[4,6]					
	ASP	AGRICOLA	CAB Direct	FSTA	PubMed	WoS	ASP	AGRICOLA	CAB Direct	FSTA	PubMed	WoS	ASP	AGRICOLA	CAB Direct	FSTA	PubMed	WoS	ASP	AGRICOLA	CAB Direct	FSTA	PubMed	WoS
J FOOD PROTECT	4,842	8,528	8,397	11,833	6,527	9,676	–	–	–	–	–	–	–	207	172	235	32	236	278	271	274	266	276	264
J FOOD SAFETY	740	959	824	1,143	1	1,133	–	–	–	–	–	–	–	15	–	15	–	13	71	33	67	66	–	67
J FOOD SCI	17,765	7,974	10,725	18,458	4,281	17,403	367	2	115	362	–	301	288	225	118	286	–	287	399	175	357	357	357	353
J FOOD QUALITY	NA	1,423	809	1,313	1	1,561	NA	–	–	–	–	–	NA	39	–	–	–	39	NA	33	85	85	–	85
J AM SOC BREW CHEM	NA	435	604	1,520	NA	840	NA	–	1	42	NA	–	NA	–	4	56	NA	58	NA	–	32	31	NA	32
J SCI FOOD AGR	2,710	9,365	13,368	9,807	5,196	12,695	–	–	130	117	86	170	–	224	152	159	–	223	621	284	585	498	590	549
LWT-FOOD SCI TECHNOL	5,802	972	5,960	7,771	3	7,376	–	–	–	50	–	14	–	91	48	117	–	107	816	–	794	794	–	790
MEAT SCI	5,183	6,209	5,119	6,693	6,660	6,273	–	–	–	–	–	–	–	127	24	125	131	128	450	303	297	448	298	294
POSTHARVEST BIOL TEC	2,965	3,236	3,346	3,078	1	3,185	–	–	–	–	–	–	–	99	89	86	–	94	273	263	257	246	–	255
POTATO RES	137	646	1,758	390	NA	1,264	–	–	33	7	NA	11	–	–	52	7	NA	51	24	11	24	8	NA	24

[1]Searches were done during the first week of October, 2017; PubMed, Web of Science, and CAB Direct were accessed via their database platforms; AGRICOLA, FSTA and Academic Search Premier were accessed via EBSCOhost.
[2]Database acronyms as described in Table 2
[3]Journal abbreviations as indicated at "Web of Science – Journal Title Abbreviations" (https://images.webofknowledge.com/images/help/WOS/A_abrvjt.html)
[4]"NA" = not applicable due to journal not being indexed; "-" = journal currently indexed by database but no articles recovered for specified year
[5]Total number of articles in databases from specified journals (without filtering for year of publication)
[6]Number of articles in databases from specified journals for specified years (filtered with respect to specified year of publication)

Table 6–Search parameters for retrieval of articles dealing with "*in vitro* bile acid binding properties of dietary fibers."[1]

Search Parameters	Database: Academic Search Premier (via EBSCOhost)[2]	AGRICOLA (via EBSCOhost)[2]	CAB Direct	FSTA (via EBSCOhost)[2]	PubMed	Web of Science
Fields Searched[3] (includes every word in identified field)	Title (including source title), Abstract, Author, Author Keywords, Subject	Title, Abstract, Author, Author Keywords, Subject	All Fields[4]	Title, Abstract, Author, Author Keywords, Subject	Title, Abstract, Medical Subject Headings and Subheadings[5], Other Terms[6], Chemical Names, Secondary Source Identifiers[7]	Title, Abstract, Author Keywords, and KeyWords Plus[8]
Search Terms[9,10,11]	Group 1 (dietary fiber related terms): (Fiber OR Fibre OR Grain OR Wheat* OR Buckwheat OR Oat OR Bran OR Glucan OR Psyllium OR Cellulose OR Lignin OR Hemicellulose OR Pectin OR Hull OR Cereal) Group 2 (bile acid related terms): (bile OR cholic OR cholate OR glycocholic OR glycocholate OR taurocholic OR taurocholate OR ursodeoxycholic OR ursodeoxycholate OR glycochenodeoxycholic OR glycochenodeoxycholate OR taurochenodeoxycholic OR taurochenodeoxycholate OR lithocholic OR lithocholate OR deoxycholic OR deoxycholate) Group 3 (*in vitro* studies): Vitro Group 4 (binding/association related terms): (associa* OR dissocia* OR bind* OR affinit* OR sequest* OR absor* OR adsor* OR sorpti*)					
Searches Performed	Search 1 (S1) = Group 1 AND Group 2 Search 2 (S2) = Group 1 AND Group 2 AND Group 3 Search 3 (S3) = Group 1 AND Group 2 AND Group 4 Search 4 (S4) = Group 1 AND Group 2 AND Group 3 AND Group 4					

[1]Retrievals were limited to English language journal articles.
[2]Search parameters were dictated by Oregon State University's access via EBSCOhost.
[3]"Fields Searched" were based on those automatically accessed by Web of Science for "Topic" searches. "Fields Searched" for the other databases were chosen with the intent of keeping search fields as similar as possible across searches. Differences in "Fields Searched" were a result of each database having somewhat different field groupings that are automatically linked.
[4]"All Fields," in CAB Direct, includes article titles, abstracts, author names, author affiliations, and descriptors.
[5]"Medical Subject Headings" (MeSH) are based on the National Library of Medicine's controlled vocabulary thesaurus; they are used to categorize articles represented in MEDLINE. "Subheadings" are qualifiers often used with MESH for further clarification/division.
[6]"Other Terms" includes non-MeSH subject terms (assigned keywords) and author-supplied keywords.
[7]Secondary Source Identifiers" are those that supply further citation information, for example, other data sources, databanks, and accession numbers of molecular sequences (for example, GenBank, ClinicalTrials.gov)
[8]Keywords Plus are index terms generated from the titles of cited articles.
[9]"Search terms" (keywords) were chosen by the investigator based on the "research topic" specified in the text. Search terms were consistent for all database searches.
[10]Truncation: used to search the multiple forms of words. Enter the root of a search term with "*" at the beginning and/or the end of the term as appropriate (left- compared with right-hand truncation).
[11]The Boolean term "OR" dictates flexible inclusion (that is, at least one must be included); the Boolean term "AND" dictates mandatory inclusion (that is, all must be included).

important when deciding whether to search CAB Direct, FSTA, or both CAB Direct and FSTA—or any of the other databases listed in Table 3.

The next phase of the study focused on article overlap between databases. Article overlap in database coverage is important because extents of "journal overlap," without consideration of "article overlap," can be misleading, because journal coverage by a database does not necessarily mean all of the articles from the "covered journal" are cited in the database. If a journal is considered a "core journal" for a particular database, then it is likely that all articles from that journal will be indexed in that particular database. Otherwise, database managers may be selective as to which articles from a "covered journal" are actually cited in their database. Thus, two databases may cover the same journal and yet differ with respect to the articles cited from that journal. This paradox was addressed herein by comparing the extent of article coverage from selected food science-discipline journals in the six major databases previously considered for journal overlap. The "test journals" chosen for this comparison ranged from those covering food science in general (for example, "Journal of Food Science"), to those covering an academic discipline within food science (for example, "Food Chemistry"), and to those covering a specific commodity (for example, "Potato Research"). All of the journals considered in this phase of the study are relatively well established. Presumably, these journals are more likely to be covered by major databases than are more obscure journals. Table 4 indicates which databases indexed these test journals. The majority of the test journals are covered by a majority of the databases. PubMed indexed the fewest of these journals, only nine of the 24; ASP indexed the second fewest (16 of 24). The other databases covered all but one journal. The most specialized database, FSTA, was the only database to formally cover the journal "Food Structure". Some of the database omissions are somewhat surprising; for example, the lack of coverage of "Journal of Agricultural and Food Chemistry" in ASP.

The data of Table 5 reflect the extent to which databases differ with respect to article coverage for specific journals. The data was collected by searching the databases for articles published in specific journals (those listed in Table 4) during selected years (1976, 1996, and 2016); thus, the range of coverage spans 40 years. Consideration of a few examples from Table 5 is sufficiently informative to make the point that journal coverage cannot be equated with article coverage. The first thing to note is the apparent lack of coverage of early papers as evidenced by the low numbers of database entries for 1976. Low numbers are indicative of relatively low numbers of papers being published at that time, relatively low numbers of published papers being indexed at that time, and/or, in some cases, journals not existing at that time (for example, "Journal of Food Processing" started in 1977, "Journal of Food Engineering" in 1982, "Postharvest Biology and Technology" in 1991, and "Food Structure" in 2014). Focusing on the more recent 2016 database entries, it can be seen that, in general, coverage of the test journals is similar, but not the same, for the different databases. The numbers of articles indexed in WoS, FSTA, and CAB Direct are within a few percent for the food-specific journals. In contrast, comparison of entries from "Journal of Dairy Science" in FSTA, WoS, and AGRICOLA shows considerable variability. This variability can be rationalized as WoS indexing more non-food articles than FSTA; the lower number for AGRICOLA appears to indicate a lack of consistent coverage of "Journal of Dairy Science". The numbers in Table 5 are to be taken as approximate. For example, the relatively high number of articles indexed from "Journal of Food Science" in ASP for 2016 reflects that this database indexes, in some cases, editorials, content descriptions, and so on. Recognize that even minor differences in database content can be important if the articles omitted from a given database are of particular importance to the research question being addressed. The following section presents a case study that illustrates this point.

The data presented in Table 3 to 5 illustrate the uniqueness of databases commonly recommended for information retrieval in the food sciences. A case study approach was used to ascertain the consequences of this uniqueness. Searches in individual and combinations of databases were used to determine the number of peer-reviewed articles published on the "in vitro bile acid binding properties of dietary fibers." This topic is of general interest with regard to the use of dietary fibers as functional food components; bile acid binding is a putative mechanism by which dietary fiber consumption effects cholesterol metabolism (Gunness & Gidley, 2010; Kahlon, 2011; Li, Mense, Brewer, Lau, & Shi, 2017; Liu et al., 2016). The topic is also appropriate for this study because it is sufficiently refined so as to yield manageable numbers of publications in database searches. Searches for English language journal articles dealing with this topic were done using the parameters described in Table 6. All databases were searched in an analogous manner. There were 4 independent searches for each database (S_1–S_4); searches within a given database differed with respect to the number of search terms required to satisfy the query (see S1, S2, S3, and S4 of Table 6).

The data in Table 7 summarizes the number of articles retrieved from databases using different search term scenarios. The inverse correlation between the number of articles retrieved from any given database and the number of search terms required for recognition is obvious when reading across rows from S1 to S4. Under the most refined search scenario (that is, S4), Web of Science returned the most articles (261) and Academic Search Premier the least (80); the difference in number of articles retrieved being over three-fold. These values represent approximately 61% and 19% of the total number of articles retrieved by a combined search of all six databases (the total number of records retrieved from searching all six databases, after accounting for replicates, was 425). The number of articles "unique" to each database provides further insight into the amount of information that would be "missed" by not searching a given database, assuming the other five databases were searched (Table 7). These values can be unsettling because they point to the fact that a number of potentially relevant articles can be missed even if one searches multiple databases. An alternative way to think about the number of articles unique to different

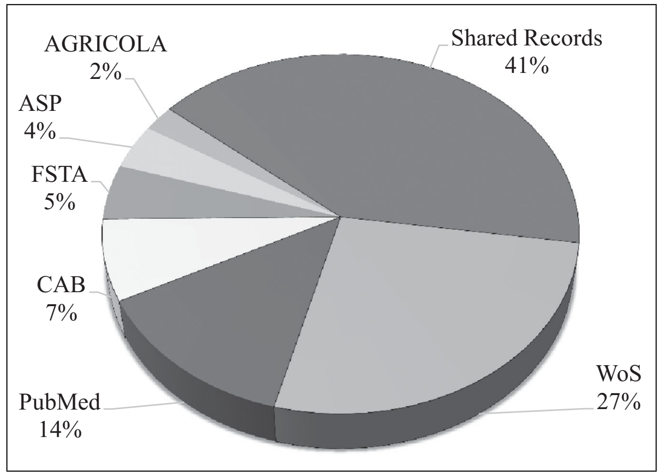

Figure 1–Percent of total articles retrieved from a combined search of AGRICOLA, Academic Search Premier (ASP), CAB Direct, Food Science and Technology Abstracts (FSTA), PubMed, and Web of Science (WoS) that are attributable to specific databases. Searches were done using S4 parameters (see Table 6 for details). Percent values for individual databases were calculated as 100 x (number of articles unique to database/sum of articles retrieved from all databases); values are rounded to the nearest whole number. "Shared records" are those articles indexed by at least two databases.

databases is to consider the percent they represent relative to the total number of records retrieved from a combined search of all 6 databases. These values are presented visually in Figure 1. The data shows that the percent of articles missed by not including FSTA, AGICOLA, ASP, or CAB Direct in such a search is less than 10%. One may be tempted to conclude that the numbers of articles unique to some of these databases are acceptably low such that those databases need not be included in searches. The difficulty in this rationale is that the importance of articles unique to different databases, no matter how few they may be, is generally not known until the articles have been examined. Clearly, one would not want to exclude all of these lower yielding databases (that is, FSTA, AGRICOLA, ASP, and CAB Direct) in the present case because together they account for nearly one-fifth of the total publications retrieved. When considering the absolute numbers of papers retrieved in this study (Table 7), it is important to keep in mind that these values are strictly dependent on the combinations of keywords used in the searches. The expectation is that adding additional permissible keywords (when using the "or" Boolean operator) or reducing the number of "required" keywords is likely to increase the number of papers recovered and, thus, broaden the overall search.

A pairwise comparison of the numbers of articles retrieved from the different databases is presented in Table 8. Pairwise comparisons are useful when considering database similarity and prioritizing databases for searches. In the present case, it is apparent that the database pair generating the most articles in response to the specified query is WoS and PubMed, and that both databases contributed substantially to that total (approximately 67% of the 261 articles retrieved from WoS were not indexed in PubMed; approximately 48% of the 163 articles retrieved from PubMed were not indexed in WoS). "Relative overlap" is defined for a pair of databases as the number of articles, excluding replicates, indexed by both databases divided by the total number of articles indexed in the one database of the pair for which the overlap is relative to (see "Methods" for the equation). Relative overlap is thus a measure of the percentage of the retrieved articles from one database

Table 7–Number of English language journal articles retrieved from selected bibliographic databases using combined search terms

Database[1]	Selected search terms[2] S1	S2	S3	S4	Number of articles unique to database[3] (search scenario S4)
Web of Science	2,136	941	468	261	114
PubMed	2,092	908	288	163	57
CAB Direct	1,414	570	255	159	31
Food Science and Technology Abstracts	533	287	153	106	21
Academic Search Premier	948	393	156	80	17
AGRICOLA	610	291	146	103	10
Combined Databases[4]	4,160	1,793	739	425	–

[1]Databases are listed in descending order of predominance by number of unique articles retrieved using search scenario S4.
[2]Search terms are defined in detail in the lower portion of Table 6. "S1" refers to "Search 1" as described in of Table 6, which was based on combined search terms from the fiber grouping and the bile acid grouping; "S2" was based on combined search terms from the fiber grouping, the bile acid grouping, and the term *vitro*; "S3" was based on combined search terms from the fiber grouping, the bile acid grouping, and the association/binding grouping; "S4" was based on combined search terms from all 4 groupings: fiber, bile acid, *vitro*, and association/binding.
[3]Number of individual articles retrieved from specified database using search parameters corresponding to S4 (see Table 6) which were not available in other databases when using the same search parameters.
[4]Sum of articles retrieved from PubMed, Web of Science, AGRICOLA, Food Science and Technology Abstracts, Academic Search Premier, and CAB Direct; replicates were deleted prior to summation.

Table 8–Pairwise comparison of numbers of retrieved articles and article overlap for selected databases (subject area: "*in vitro* bile acid binding properties of dietary fibers").[1]

Overlap analysis	Database pairings[2] WoS-PbM	WoS-Agrc	WoS-FSTA	WoS-ASP	WoS-CAB	PbM-Agrc	PbM-FSTA	PbM-ASP	PbM-CAB	Agrc-FSTA	Agrc-ASP	Agrc–CAB	FSTA-ASP	FSTA-CAB	ASP-CAB
Total Articles[3]	339	283	291	285	314	221	227	218	251	150	145	187	148	185	194
TO (%)[4]	25.1	28.6	26.1	19.6	33.8	20.4	18.5	11.5	28.3	39.3	26.2	40.1	25.7	43.2	23.2
ROA (%)[5]	32.6	31.0	29.1	21.5	40.6	27.6	25.8	15.3	43.6	57.3	36.9	72.8	35.8	75.5	56.3
ROB (%)[6]	52.1	78.6	71.7	70.0	66.7	43.7	39.6	31.3	44.7	55.7	47.5	47.2	47.5	50.3	28.3

[1]Search parameters are defined in Table 6, scenario S4.
[2]Database pairs are presented in columns, the upper database designated "A," the lower database is designated "B;" PbM = PubMed, Agrc = AGRICOLA, CAB = CAB Direct, other acronyms as used in Table 2.
[3]The number of retrieved articles for a given combination of databases after removing duplicates (that is, the number of articles retrieved from A and B after removing duplicates).
[4]TO (%) = "traditional overlap" calculated as the percent of the total articles retrieved (excluding duplicates) that were indexed in both databases (see "Methods" for equation)
[5]ROA (%) = overlap relative to database A, "A" being the upper database of the pair listed at the top of the column; calculated as the percent of records in database A that are also indexed in database B.
[6]ROB (%) = overlap relative to database B, "B" being the lower database of the pair listed at the top of the column; calculated as the perent of records in database B that are also indexed in database A.

that are also indexed in the other database of the pair. For example, the data of Table 8 shows that nearly 80% of the articles retrieved from AGRICOLA were also indexed in WoS. This database pair provides an instructive example of why overlap is best considered relative to both databases; note that only approximately 31% of the articles retrieved from WoS are indexed in AGRICOLA. In some cases the relative overlap is strikingly low, for example only approximately 15% of the articles retrieved from PubMed were indexed in ASP. These relative values are particularly useful when considering the likelihood that a database will contribute articles that are distinctive to the combined retrieval list. The traditional overlap values of Table 8 provide a general picture of the overall similarity of databases (Gluck, 1990). It is not particularly surprising that the two most specialized databases, FSTA and AGRICOLA, were among those databases with the highest traditional overlap.

The combined results from this study are helpful when considering the design of literature searches. The productivity of initial searches is typically evaluated based on the number of relevant articles recovered. Assuming this criterion, the results of the case study suggests the Web of Science database is a logical place to start. Over half of the total citations retrieved in the case study were accessible through Web of Science. In many situations, an initial search in a single database will provide access to sufficient information to answer a researcher's question(s) and thus no further data mining is necessary. However, in those cases where reproducible, in-depth surveys of the literature are required, such as for systematic reviews and meta-analyses, one would certainly want to continue mining other databases. The need to search multiple databases to obtain a thorough representation of the literature on a given topic is obvious from the number of unique citations found in each database for the case study. There will eventually be a point of diminishing returns where further searches will yield few additional citations. The study reported herein focused on the use of abstracting and indexing databases for information retrieval. These are "curated" (or "purposefully managed") databases in the sense that they are selective as to what is indexed; the goal is not to include all scholarly activity in these databases. Their selectivity is a result of each database having unique quality guidelines and topical constraints. Database-specific guidelines and constraints account for the differences in the numbers and types of articles included in the databases. An alternative approach to using controlled abstracting and indexing databases is to search for information using an academic web search engine, such as Google Scholar or Microsoft Academic (Halevi, Moed, & Bar-Ilan, 2017; Thelwall, 2017). These search engines contain web crawler-derived data. The advantage of using an academic web search engine is that they typically access far more data than do abstracting and indexing services. A major reason for the higher numbers of citations retrieved by academic web search engines is that they do not filter source materials to the same extent as the abstracting and indexing services. Some of the data included comes from traditional journal sources, but other sources of quasi-scholarly materials may also be retrieved. Thus, the sources retrieved by these services will not have undergone the level of quality control typically associated with curated abstracting and indexing databases. Drawbacks to using these search engines, beyond the perceived quality control issues, include difficulties in determining the sources of materials crawled and the need for the researcher to spend additional time filtering the relevant from the nonrelevant retrievals. The inability to identify specifically which journal sources were included and for what date ranges made the inclusion of academic web search engines unfeasible for this case study. Even with these limitations, academic web search engines can be reasonable choices for initial information retrieval activities, particularly when searching broad topic areas.

Conclusions

The data presented illustrates the importance of database selection and the need to work with multiple databases when doing knowledge assessment in the food sciences. All of the databases evaluated in this study indexed articles unique to them; this proved to be true in the general sense and in the specific "case study" section of the study. A logical extension of this finding is that databases not included in the present study may also include articles that are both unique to those databases and relevant to the stated search query. Thus, it should be recognized that even after searching multiple databases there exists the possibility that pertinent information/articles may not have been recovered. There are several approaches for addressing the comprehensiveness of an information retrieval exercise (Papaioannou, Sutton, Carroll, Booth, & Wong, 2010). "Comprehensiveness" is often assessed by following citation trails, either forward or backward, of key papers in the target field of research; essentially asking the extent to which one's database mining has retrieved articles cited in accessible papers (Wright, Golder, & Rodriguez-Lopez, 2014).

Author Contributions

Tuba Karaarslan Urhan contributed to the design of study, collected data, contributed to data analysis, prepared the first draft of manuscript and participated in the manuscript revisions. Hannah Gascho Rempel participated in the design of the study, provided guidance pertaining to practical aspects of information retrieval, and participated in the manuscript revisions. Lisbeth Meunier-Goddik contributed to the design of the study and participated in the manuscript revisions. Michael H. Penner conceived the study, directed parts of the experimental work, contributed to data analysis and participated in the manuscript revisions.

References

Booth, A. (2016). Searching for qualitative research for inclusion in systematic reviews: A structured methodological review. *Systematic Reviews*, *5*(1), 74. https://doi.org/10.1186/s13643-016-0249-x

Duran, N., & MacDonald, K. (2006). Information sources for food studies research. *Food, Culture & Society*, *9*(2), 233–243.

Gasparyan, A. Y., Yessirkepov, M., Voronov, A. A., Trukhachev, V. I., Kostyukova, E. I., Gerasimov, A. N., & Kitas, G. D. (2016). Specialist bibliographic databases. *Journal of Korean Medical Science*, *31*(5), 660–673. https://doi.org/10.3346/jkms.2016.31.5.660

Gasparyan, A. Y., Ayvazyan, L., & Kitas, G. D. (2013). Multidisciplinary bibliographic databases. *Journal of Korean Medical Science*, *28*(9), 1270–1275. https://doi.org/10.3346/jkms.2013.28.9.1270

Gluck, M. (1990). A review of journal coverage overlap with an extension to the definition of overlap. *Journal of the Association for Information Science and Technology*, *41*(1), 43–60. https://doi.org/10.1002/(SICI)1097-4571(199001)41:1<43::AID-ASI4>3.0.CO;2-P

Gunness, P., & Gidley, M. J. (2010). Mechanisms underlying the cholesterol-lowering properties of soluble dietary fibre polysaccharides. *Food & Function*, *1*(2), 149–155. https://doi.org/10.1039/c0fo00080a

Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature. *Journal of Informetrics*, *11*(3), 823–834.

Hart, C. (2001). *Doing a Literature Search: A Comprehensive Guide for the Social Sciences*. London: Sage.

Hood, W. W., & Wilson, C. S. (2003). Overlap in bibliographic databases. *Journal of the Association for Information Science and Technology*, *54*(12), 1091–1103.

Jensen, L. J., Saric, J., & Bork, P. (2006). Literature mining for the biologist: From information retrieval to biological discovery. *Nature Reviews Genetics*, 7(2), 119–129. https://doi.org/10.1038/nrg1768

Kahlon, T. S. (2011). Health-promoting potential of cereals, grain fractions, and beans as determined by their in vitro bile acid binding. *Cereal Foods World*, *56*(4), 151–155.

Li, C. Y., Mense, A. L., Brewer, L. R., Lau, C., & Shi, Y. C. (2017). In vitro bile acid binding capacity of wheat bran with different particle sizes. *Cereal Chemistry*, *94*(4), 654–658. https://doi.org/10.1094/CCHEM-08-16-0211-R

Liu, C., Lin, X. L., Wan, Z., Zou, Y., Cheng, F. F., & Yang, X. Q. (2016). The physicochemical properties, in vitro binding capacities and in vivo hypocholesterolemic activity of soluble dietary fiber extracted from soy hulls. *Food & Function*, 7(12), 4830–4840. https://doi.org/10.1039/C6FO01340F

Papaioannou, D., Sutton, A., Carroll, C., Booth, A., & Wong, R. (2010). Literature searching for social science systematic reviews: Consideration of a range of search techniques. *Health Information & Libraries Journal*, *27*(2), 114–122. https://doi.org/10.1111/j.1471-1842.2009.00863.x

Stanbury, H., & Selman, J. (2008). Database publishing and increasing access to food science information. *Journal of Agriculture & Food Information*, *9*(1), 21–40. https://doi.org/10.1080/1049650080212286

Stevinson, C., & Lawlor, D. A. (2004). Searching multiple databases for systematic reviews: Added value or diminishing returns? *Complementary Therapies in Medicine*, *12*(4), 228–232. https://doi.org/10.1016/j.ctim.2004.09.003

Thelwall, M. (2017). Microsoft Academic: A multidisciplinary comparison of citation counts with scopus and mendeley for 29 journals. *Journal of Informetrics*, *11*(4), 1201–1212.

Wright, K., Golder, S., & Rodriguez-Lopez, R. (2014). Citation searching: A systematic review case study of multiple risk behaviour interventions. *BMC Medical Research Methodology*, *14*(1), 73. https://doi.org/10.1186/1471-2288-14-73.