Information Literacy

©THOMAS J FROEHLICH, PH.D.
PROFESSOR EMERITUS
SCHOOL OF INFORMATION
KENT STATE UNIVERSITY
KENT OH 44240
TFROEHLI@KENT.EDU
HTTP://PERSONAL.KENT.EDU/~TFROEHLI/

Outline

- Regular approaches to information literacy:
  - determining the credibility of web sites and other online sources;
  - learning how to locate, evaluate, and use effectively the information sought for information needs;
  - CRAAP test
  - Wikipedia and the Wisdom of the crowd
- Learning the merits, defects, and effective use of search engines;
- The merits and problems of intellectual technologies
  - Database searching
  - Citation Searching
  - Bias in Classification Schemes

Information Literacy

The American Library Association (ALA) characterizes information literacy as the: "set of skills needed to find, retrieve, analyze, and use information," including "competencies in formulating research questions and in their [students'] ability to use information as well as an understanding of ethical and legal issues surrounding information" and skills "in critical thinking" (Information literacy glossary, 2006).

With information literacy training, information seekers would:

1. know when they have a need for information
2. identify information needed to address a given problem or issue
3. find needed information and evaluating the information
4. organize the information
5. use the information effectively to address the problem or issue at hand. (adapted from Presidential committee on information literacy: Final report, 2006)

IFLA: How to Spot Fake News

Information Literacy

- There are many aspects of information literacy that can be detailed. You are undoubtedly familiar with the kind of chart the IFLA has provided.
- We must determine the credibility of web sites, especially those espousing fake news,
- by analyzing their currency (appropriate up-to-dateness),
- the authorship (if available) or sponsoring agency,
- the quality of their links and supporting resources such as bibliographic references
- by checking with experts or with fact-checking sites, such as PolitiFact (http://www.politifact.com/), FactCheck (https://www.factcheck.org/) or Snopes (https://www.snopes.com/fact-check/). 2020 Ten Best Fact-Checking Sites: https://mediabiasfactcheck.com/2020/04/12/the-10-best-fact-checking-websites-for-2020/
- These sites, too, can reflect bias (though not necessarily an invalidating bias, one that it ignores or distorts the interpretation of the facts or evidence): https://www.makeuseof.com/tag/true-5-factchecking-websites/ (Eillis, 2019).
- For a good approach to web site evaluation, see https://www.citationmachine.net/apa/cite-a-website/ (Citation Machine, n.d.).
- See also http://knight.org/vision https://www.cip.uw.edu/ Center for Informed Policy

Information Literacy

- Many libraries post information about the CRAAP test, a simple guide for evaluating sources found on the internet.
- CRAAP is an acronym for evaluating such properties as
  - Currency
  - Relevance
  - Authority
  - Accuracy
  - Purpose
  - An example can be found at https://guides.library.illinoisstate.edu/evaluating/craap.

CRAAP Illustration

Use the questions to evaluate the source.

Currency: To determine if the date of publication of the information is suitable for your project.

What is the copyright, publication, or posting date?
Why is or isn't the date important for the message or content of the source?
Is the information outdated in relation to the topic?

Relevance: To determine how applicable the information is for the purpose of your project.

For what audience or level is the information written (general public, experts/scholars, etc.)?
Explain why you would or would not quote/reference the information from this source in your project.

Authority: To determine if the source author, creator, or publisher of the information is the most knowledgeable.

Who is the author, creator, or publisher of the source or what organization is responsible for the source? How do you know if the author is an expert on the topic (e.g. examine the author's credentials, experience, and/or organizational affiliation)? Is there contact information available?
From where does the money for the research or programming come, if relevant?

Accuracy: To determine the reliability, truthfulness and correctness of the content.

What indications do you see that the information is or is not well researched or provides sufficient evidence? Are facts and claims documented within the text, as notes, or in a bibliography?
What kind of language, imagery and/or tone is used (e.g. emotional, objective, professional, etc.)?

Purpose: To determine the reason why the information exists.

Why was this source written (e.g.to inform, teach, entertain, persuade)?
How might the author's affiliation affect the point of view, slant, or potential bias of the source?
How might the intended audience affect the point of view, slant, or potential bias of the information?
What conclusions are presented, and is the information complete? Is anything major excluded?
How does this resource compare to others on the same topic?

Wikipedia?

- With caution I often recommend Wikipedia for a start of many topics.
  - The good news: when it is good, it reflects “the wisdom of the crowd, ” contributions edited by many familiar with a topic
  - The bad news: it is difficult to know the author, authors or sponsoring agency, the background of those contributing to its creation.
  - If one knows something of the content of the topic and the topic fits with what they know, it might be useful.
  - One can checks am entry’s sources or the trustworthiness of the links embedded in it.

Wisdom of Crowd: Wikipedia?

If Wikipedia reflects the “wisdom of the crowd” what does that mean?

- In Aristotle, wisdom was a property of an individual.
- With the advent of Web 2.0 tools, such as Wikis and blogs and twitter, we have the social creation of information or knowledge
- James Surowiecki in 2004 wrote ***The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations,*** in which he argues that aggregation of information in groups, resulting in decisions that, he argues, are often better than could have been made by any single member of the group.
- The notion of the crowd has evolved and there have been many psychological experiments to show when it is successful and where there are problems. See: https://en.wikipedia.org/wiki/Wisdom_of_the_crowd
- Google’s page ranking algorithm is an example of the frequent effectiveness of the wisdom of the crowds, because it relies on link popularity.
- Trial by jury is a legal instance of the wisdom of the crowd (problems: CBS show *Bull* on jury manipulation)

Search Engines

Even with the ease of access to search engines, such engines are poorly used, and the nature of the results is poorly understood. The next important information literacy tool is learning the merits, defects, and effective use of search engines. The following is an outline of key points:

(1) The choice of vocabulary in a search engine is important. A search on *kidney neoplasms* will generally produce qualitatively better results than *kidney cancer* because the former is the accepted medical terminology, used in scientific studies, and is likely to occur in research-based web sites or resources. Having said that, *kidney cancer* sites may be more accessible to the layperson. The point is that the choice of search terms can greatly affect the nature and quality of the results.

(2) The use of search engine qualifiers will improve the quality of one's search, such as these Google techniques, Refine Web Searches (https://support.google.com/websearch/answer/2466433?hl=en&ref_topic=3081620) or Advanced Search Techniques (https://support.google.com/websearch/answer/35890?hl=en&ref_topic=3081620). One can restrict searches to specific domains (e.g., .gov), to specific time frames, to particular words or phrases, to alternative words or phrases, to language, to file type, to image type, or to image color, or to exclude any of these, to mention a few options).

Search Engines

- All search engines exhibit bias. There are 200 factors that affect how Google ranks its search results (https://backlinko.com/google-ranking-factors), but most factors do so only slightly (https://optinmonster.com/seo-ranking-factors/).
- For example, new sites often rank low, the most popular sites (built on the notion of link popularity – the more sites that link to a particular site are call link popular) are high on the output list.
- However, what is popular may not be the best. Sites that load slowly on mobile phones are ranked low but may have good information. Since 95% of searchers never go beyond the first page of search results, this is a serious problem because there may be more valuable resources below the splash page or pages or the fold of the splash page (Santora, 2019).
- See also Fowler, G. (2020, October 19). Perspective | How does Google's monopoly hurt you? Try these searches. Retrieved October 19, 2020, from https://www.washingtonpost.com/technology/2020/10/19/google-search-results-monopoly/?utm_campaign=wp_post_most

Information Systems

- Despite the biases of search engines like Google, it often supplies resources that satisfices the information (resources good enough). We need to train users to use resources beyond Google that often will supply them with more accurate and detailed information by using library catalogs and library databases.
- These intellectual tools have come a long way – with the introduction menu-driven and ranking systems that often produce reasonably good output, often leading directly to primary resources. The output is often better than Google alone.

Intellectual Technologies

- Patrick Wilson notes that library and information work is comprised of a variety of “intellectual technologies.”
- Abstracting, indexing, cataloging, reference work are all intellectual technologies, which, if done well, help the information seeker or mystery lover to the best sources.
- Abstracting reduces the intellectual content of an article to a concise summary of the content of an article so that the potential user can determine whether a article may be relevant to their research.
- Indexing tries to assign key signifiers (terms, phrases) about what an article is most likely to be about.
  - It can be an intellectual process by an person training the art of abstracting
  - It can be a computer ranking process that is not attached to the meaning of terms but their frequency and location (e.g., in the text or in the headings of subheadings).

Database Information Systems

- Unfortunately there is also a down-side to menu-driven systems and to the ranking algorithms, both of which are flawed (but better than nothing).
- While menu-driven systems are useful in searching online databases, there are unknown hazards if one is trying to do a comprehensive search or precise search.
- For example, if one is looking in the research database, ARTbibliographies Modern, for a list of publications, by Yves-Alain Bois from 1980 to the present, one would typically enter the author name as given or interpreted in the search query: e.g., Bois, Yves-Alain. However, it turns out that the database has six variations of the author's name: ("Bois, Yve Alain" OR "Bois, Yve-Alain" OR "Bois, Y -A" OR "Blois, Yve-Alain" OR "Bois, Y A" OR "Bois, Yves-Alain"). If one used only the name given to them (Bois, Yves-Alain), one would get a partial result because they would get results only for the one variation of author name that they used, not any from any of the other variations of author name.
- Many, if not all, users think that computers automatically map all variations of an author's name to a single entry, but it does not. A few systems which have what is called strong authority control, such as the Library of Congress, do link, for example, Jacqueline Kennedy Onassis (their preferred entry) with Jacqueline Bouvier Kennedy, Jackie Kennedy, Jackie Onassis, Jacqueline Bouvier, Jiagulin (Chinese variant) or Jackie, all of the forms of her name that an information seeker may use to find information by or about her.
- But these systems are few, and many information databases do not have this feature.

Information Systems

- This problem of a lack of authority control is not only true for ARTbibliographies Modern but many other databases and systems. It is easy to fail to make a comprehensive or precise search in databases such as these or miss a relevant entry because of not using the correct form of author name or not using all forms of the author name in the database. The same can be said of subject terms (Medline uses *kidney cancer* whereas Embase uses *kidney neoplasms* as descriptors and not using the proper form for each database will produce widely different and imprecise results)
- What magnifies this problem is that different database producers do different forms of indexing of author name and may have different entries for a particular author name (e.g., last name with first name versus last name with initials for first and/or middle name) so that when does multiple database searching (for which most libraries provide), especially database searching from different vendors, the results are severely flawed (because the computer *does not map* to different forms of author name).
- [Computers are stupid, stupid, stupid! Or rather the software developers are often not very smart!].

Information Systems

- Furthermore, one can increase the precision of one's results from organized information collections, such as information databases, by learning about the indexing or subject terms used to construct the database.
- For controlled databases with a controlled indexing vocabulary, the indexers try to be consistent in assigning subject term vocabulary to the intellectual content of articles in the database.
- If one uses the assigned term for a particular concept for a particular database, one can achieve a precise result, i.e., all articles that have been assigned a particular subject term will be clustered in the result. The result will be the consequence of an intellectual process undertaken by indexers and not a computer algorithm that does not understand the meaning of terms.
- Unfortunately, the assignment of subject terms varies among different databases and database producers, so that terms used in one database may not be used in another. Many databases have no controlled subject terms. Multidatabase searching using a single search term or phrase will produce flawed results, unless one takes the trouble to use the correct term, if it is available, for each of the databases being searched.
- There are many other issues to learn about databases and their construction that could enhance one's ability to search more effectively. However, it is important to note that if the information seeker just wants anything related to the search topic (i.e., anything about a particular concept or anything by a particular author), something that "satisfices" their information need (i.e., seek the minimum acceptable outcome or choose the first satisfactory option that one comes across), then rigor in using search systems, seeking what is called high precision (i.e., looking for many articles directly on target) or high recall (i.e., looking for many articles closely related to their information need) is not required. Google satisfices many information needs, which is why it is so popular.

Issues in Citation Indexing

- With regard to citations, the internet also offers librarians (and information seekers) many tools, as do libraries or information centers.
- Again, however, librarians should exercise caution when recommending a particular tool.
- For example, high citation counts of a particular journal article or author may in fact successfully lead to related work.
- But there are flaws in citation work, for example, that citation of a particular work indicates that the author used that work.
  - There are many reasons why an author may include a reference in their paper that in fact he or she does not use.
  - For example, he or she may be hoping to benefit from the "halo effect." By citing more prominent authors in the field, they may be hoping that their work will gain prestige from the citation.

Issues in Citation Indexing

Linda Smith in her article on "Citation Analysis," (**Library Trends**, 1981, pp. 83-106) delineates five major points about the problems with assumptions about citations and references.

- “Citation of a document implies use of that document by the citing author.”
- “Citation of a document (author, journal, etc.) reflects the merit (quality, significance, impact) of that document (author, journal, etc.).” Obviously there are references that occur because an article is particularly bad.
- “Citations are made to the best possible works. This assumes that authors know and sift the literature of their field and choose the best. But the principle of least effort seems to obtain” Writers use most often works they know or available – of course, electronic access has broadened choices.
- “A cited document is related in content to the citing document; if the documents are bibliographically coupled, they are related in content; and if two documents are cocited, they are related in content.” The problem is the parts of a particular document used may not refer to the same parts of the document. The coupling is therefore at the document level but not necessarily at the subdocument level – hence the notion of coupling is problematic
- “All citations are equal.” Clearly this is problematic, because the variety of citations reflect a variety of values – from highly significant, somewhat significant to not significant at all (e.g., significant in its badness – the worst study on this subject).

Intellectual Technologies

- We must assert that citation indexing can, in fact, be useful, especially in trying to find related work. But there are also problems -- we cannot understand the connection between the cited and citing works until we do an intellectual analysis – looking at each text and seeing what parts of the citing work are connected to what parts of the cited work. In some sense, citation indexing is simply mechanical – we connect the citing and cited work simply because the cited work is cited, not because of any intellectual analysis per se. This is how the citation databases are in fact constructed.
- What is the point of this? These tools, such as citation indexing or controlled vocabulary searching are useful, but an information professional should know the merits and defects of a particular tool, and make sure that, if consulted, they explain the merits and defects to end-users; otherwise, they mislead the information seekers and exhibit a level of incompetence.

Bias in Classification Schemes

- Philosophically, one could show that any classification scheme or cataloging process is full of prejudgments or biases, which is a problem when they are not acknowledged.
- For example, Siamese-twins (or rather "Twins conjoined") still falls under the heading "Monsters" in **Medical Subject Headings (MESH)**. The problem emerges when
  - such prejudgments are not recognized
  - erroneous and unacceptable conceptions persist
  - pejorative and outdated terms continue in use, partly because of the inertia built into systems, but more importantly because those who maintain and apply such systems do not modify them and users are unable to find references in these systems

Classification Schemes and Bias

- ***Illustration***
  - Sanford Berman, the former head cataloger of the Hennepin County library in Minnesota, attacked the 19th edition of Dewey Decimal Classification for inadequate coverage of popular music and gay and lesbian issues and optional use of numbers for North American Indians [*Berman 1982, pp. 178-180*].
  - He also points out that many people-oriented subject headings either obscure or demean the group to which they refer. For example, of the use of the subdivision "Management" under the heading "Children" he says: Children "... require guidance and special care. But 'guidance' and 'care' do not equal 'management,' a term dear to manipulators and repugnant to anyone who refuses to class *people*, including the young with *things*." [*Berman 1993, pp. 171-172*].

- We see that widely used classification schemes contain biases and what we now see as structural errors: e.g., philosophy as a central phenomenon.
- The problem is that schemes cannot be overthrown and started from scratch. Even so any current classification scheme would reflect our understanding of the world (e.g., the centrality of computers and information technology) and future classifications involving new phenomena would be difficult to classify.

Classification Schemes and Prejudgments

- In 1983 Celeste West found the heading, "Literature, Immoral" in use in the San Francisco Public Library [*West, p. 1652-1653*].
- One wonders which librarian undertook the task to place materials into that category.
- When such dubious categories are used, it makes it difficult or impossible for patrons to find information.
- Erroneous perceptions and misbegotten values distort the domain of available information.

Intellectual Technologies

About these intellectual technologies several points can be made:

(1) They can be done well or they can be done badly

(2) They can be understood well or badly and communicated well or badly

(3) They are not morally neutral, often reflecting the bias, orthodox framework at the time they were created.

(4) When information professionals use these tools or train end-users in these tools, they must be aware of both the benefits and limitations of such tools.