

2012

Accessing Materials: Review of the MacCrate Report's Fundamental Lawyering Skill: Search

Michael McArthur

Follow this and additional works at: https://digitalcommons.law.uw.edu/law-lib_borgeson

Recommended Citation

Michael McArthur, *Accessing Materials: Review of the MacCrate Report's Fundamental Lawyering Skill: Search* (2012), https://digitalcommons.law.uw.edu/law-lib_borgeson/77

This Article is brought to you for free and open access by the Law Librarianship Program at UW Law Digital Commons. It has been accepted for inclusion in Borgeson Paper Archive by an authorized administrator of UW Law Digital Commons. For more information, please contact lawref@uw.edu.

Accessing Materials: Review of the MacCrate Report's
Fundamental Lawyering Skill: Search

Michael McArthur

Submitted to
Professor Penny A. Hazelton
to fulfill course requirements for Current Issues in Law Librarianship, LIS 595,
and to fulfill the graduation requirement of the
Culminating Experience Project for MLIS
University of Washington Information School
Seattle, Washington

May 21, 2012

a. The MacCrate Report

In 1992 the American Bar Association Section on Legal Education and Admissions to the Bar published a report evaluating the legal education system.¹ Twenty years later there remains glaring issues with legal education. One fundamental lawyering skill from the report that has a well-documented history of inadequacy is legal research.² Section 3.2 of the white paper calls for every lawyer to obtain a “knowledge of and the ability to use the most fundamental tools of legal research.”³ While this section covers a wide swath of insufficiencies, my focus is on subsection (d).⁴

Subsection (d) states that “a lawyer should be familiar with alternative forms of accessing the materials, including hard copy, microfiche and other miniaturization services, and computerized services (such as LEXIS and WESTLAW).”⁵ Access to information has changed drastically over the past twenty years. Print material has rapidly given way to online resources and microfiche has been relegated to an inferior archival medium. It seems timely to review some of the strengths and weaknesses of the various formats and re-evaluate the MacCrate standard found in section 3.2(d). This paper is not meant to elevate or diminish one search method or resource over another, but to broaden awareness of the current issues related to legal research.

Beginning with a brief review of the history of a user’s control over search activities and expanding into the LexisNexis and Westlaw interfaces, this paper will then discuss the characteristics of the current student stereotype. It will continue on to cover the fundamental evaluative tools used for legal research. There will be an attempt to raise awareness of a few current areas of frustration regarding the transition from print to digital search. Lastly is a focus on some important distinctions to make when learning the various search methods.

b. Power over Search

For centuries, the method of recording and organizing information was in printed material. Though in colonial times law books were relatively scarce, the law continued to develop and accessibility increased over time.⁶ Print materials of the

¹ ABA, Section of Legal Education and Admissions to the Bar, *Legal Education and Professional Development: An Educational Continuum: Report of the Task Force on Law Schools and the Profession: Narrowing the Gap, Statement of Fundamental Lawyering Skills and Professional Values* (1992) [hereinafter the MacCrate Report].

² Barbara Bintliff, *Legal Research: MacCrate’s “Fundamental Lawyering Skill” Missing in Action*, 28 Legal Reference Serv. Q. 1 (2009).

³ ABA, *supra* note 1, at 159.

⁴ ABA, *supra* note 1, at 159.

⁵ ABA, *supra* note 1, at 160.

⁶ J. MYRON JACOBSTEIN ET AL., *FUNDAMENTALS OF LEGAL RESEARCH* 11 (7th ed. 1998).

time already offered substantial organizational tools, such as the table of contents, allowing the reader to quickly locate desired information. As the amount of literature increased, special volumes devoted to locating information in corresponding books also developed. These resources, such as indexes, bibliographies and citators, worked efficiently as tools to assist in finding information.⁷

In the legal field, the West Publishing Company developed a series of regional case reporters and an intricate numbering system to classify and facilitate efficient retrieval of the intimidating amount of court opinions being issued across the country.⁸ Annotated copies of primary legal material such as the United State Code and those of the various States were also published.⁹

These print resources offered the user substantial control over searching activities. A user could choose to browse the table of contents or look for terms in an alphabetized index. Some volumes offered indexes organized in other ways, such as chronologically or by jurisdiction. Publishers would have been encouraged to produce material that was organized and easily navigable to increase popularity and ultimately sales.

West's National Reporter System worked to catch every written opinion and present it as evidence of the current status of the law. Classifying and organizing all of the expanding law was a truly labor intensive endeavor. Manual printing was a time consuming process and was complicated by the expansion of supplements and pocket parts. Things changed when advanced in technology offered a different medium.

The computer was developed, a new tool that organized data much more efficiently. Suddenly data that was recorded digitally could be quickly and efficiently searched. Rather than thumb through a dusty tome, a user was able to apply a few keystrokes and the desired information would suddenly appear on the monitor. Computers offered enormous savings in time and space, at what appeared to come at no real cost to many observers.

Of course the amount of work behind the scenes to manually input all the information was substantial. Once the data was populated in the database, however, multiple terminals could access it simultaneously. As technology progressed, the labor intensive process of hand-typing print documents into the system gave way to working with digital documents and more recently retrospective digitization scanning of print resources. This process facilitated access to materials while simultaneously acting in an archival capacity.

⁷ Jacobstein, *supra* note 7, at 378.

⁸ Jacobstein, *supra* note 7, at 79.

⁹ Jacobstein, *supra* note 7, at 135.

Search functions have evolved as well. Early computers used the same indexing methods as their print counterparts. Before long specific fields were populated with descriptive information that made selective field searching for an author or title available. Some services, such as the early Westlaw database, offered search of headnotes and synopsis. The first full-fledged legal database offered a significant leap in search; full text document searching.

While database search tools started out very basic, they soon evolved to what we know today as advanced search, allowing for the use of Boolean modifiers and filters. These advanced tools provide the user with significant control over the breadth and depth of the search. Computer databases, especially those with powerful search functions, often were without an easy to use table of contents or index however. Also lacking was the ability to quickly and easily browse through the material, a serious drawback for many users.

The personal computer and Internet explosion of the early 1990's brought about the idea of accessing information from one's own home. Search engines were developed to glean and organize all the Internet had to offer, both good and bad. Many of these search engines offered more and more advanced search tools, offering as much precision searching as their systems allowed.

Then one company came along that changed all that: Google. The enormously successful search engine that is heavily used all over the world today sought to simplify the search process, permitting natural language queries instead of Boolean functions. With it came the idea that putting all your search terms in a simple box is the ideal. The box would take the terms and magically produce a wonderful list of the most relevant results. The accuracy of the results was uncanny and the search engine became an international sensation.

Lost in the excitement was how exactly Google obtained relevant results from the words input in the box. The exact process is not disclosed to the user, nor are the factors used in weighing the relevance of each source. Based on its popularity and subsequent adoption on numerous sites however, most people have naturally come to rely on the apparent accuracy of the search engine. Virtually all control over the search process has been handed over to complex algorithms. Today this is increasingly accurate of legal research as well.

c. The Online Databases

Historically two online legal databases have demanded the attention of the legal community. LexisNexis is the product of the Reed Elsevier, a Anglo-Dutch publishing company. Westlaw is the online legal database for the Thompson Reuters publishing company based in Canada. Both databases have grown to be behemoths in the legal information marketplace and both compete for the business of law firms, courts, and law schools nationwide.

The online competition started in the late 1973 when the Lexis database first offered service.¹⁰ Westlaw followed with an online searchable database in 1975.¹¹ For both services the offerings were sparse, the search functions were clunky and only accessible by terminals connected by modem, often only one per school. By the time the Internet became widely available on law school campuses, both companies began making their databases accessible online.

The databases have grown and evolved over the years, often with better search in mind. While both databases utilize menu systems in a limited manner, the real focus has been on developing full-text search tools that dig down into individual databases, or across many at once. Buried deep or often absent are the indexes and topical guides of the publishing world. Proficiency in narrowing the scope of databases accessed, utilizing field and segment search fields and understanding advanced Boolean functions allows users to perform cost-effective legal research.

When the services were first made available, each database had a corresponding print equivalent. A user could use the database to locate a particular resource and then go to the print version for easier access and readability. Eventually legal resources were added to the databases that either did not have a print equivalent or the print resource was not widely recognized or available. The growing amount of accessible information and number of databases also complicated the search process, burying many resources and virtually mandating multiple database searches.

Somewhere along the way both LexisNexis and Westlaw recognized the need to move to a Google model. If they created a simple search box and let the algorithms decide what databases to search, their users would find relevant results each time.¹² The result was a reduction of “no results found” pages and (to the chagrin of many legal research instructors) the appearance of effective legal research. The transition is now in place with WestlawNext and Lexis Advance, but not without its criticism.¹³

It wasn't just the need for managing search that caused the movement to simple search. Both companies were targeting a new generation of users, the students now entering law school.

d. Current Students

¹⁰ William G. Harrington, *A Brief History of Computer-Assisted Legal Research*, 77 *Law Libr. J.* 543, 552 (1984-1985).

¹¹ *Id.* at 553.

¹² Thompson Reuters, *WestlawNext Quick Reference Guide*, 1, 4 (2011), <https://store.westlaw.com/pdf/wln2/L-356012.pdf>.

¹³ Ronald E. Wheeler, *Does WestlawNext Really Change Everything? The Implications of WestlawNext on Legal Research*, 103 *Law Libr. J.* 359 (2011).

Although not all incoming law students are coming in right after college, most will be members of a younger generation. “Digital native” is the term broadly accepted for those who have grown up exposed to the Internet and online society their whole life.¹⁴ It is a generation that has adopted online socialization as its own and has changed the teaching model in many ways. In response, online search has evolved in substantial ways. With the overwhelming adoption of social media and other technological trends, information is expanding at an unprecedented rate.

This mixed generation has reached law school ripe with expectations. They see themselves as naturally proficient in and comfortable with the vast expanse of the Internet. They have spent most of their lives connected to the web and feel eminently able to perform research of any kind through general search engines.¹⁵¹⁶ So comfortable in fact that they often show no interest in understanding how these search engines actually retrieve and organize results.¹⁷

Technology has made communication instantaneous. As things have sped up, so have the expectations of this younger generation. Slower methods of acquiring information come across as unreasonable or inefficient.¹⁸ Pre-digital techniques such as thumbing through an index are seen as inferior, unless it is on a touch screen within the “search in document” results. What cannot be done quickly and easily is classified as a dysfunctional process that someone needs to create an application for.

While computers and later gadgets have changed the flow of information generally, the transition has also influenced the way the younger generation gathers and processes information. The devices have been integrated into their lifestyle in a very real way.¹⁹ It is difficult for many of them to imagine life without constant access to immediate communication. The withdrawal from detachment can also be severe.²⁰

Students entering higher education often see themselves as efficient information sleuths. They are experienced with Google basic search and know how to draw out the results that they want. This confidence extends to legal research habits as well, as they often seek to jump right into the large legal database services right away. Suggesting that print resources still have value is often waived off as irrelevant.

¹⁴ Marc Prensky, Digital Natives, Digital Immigrants, 9 *On The Horizon* No. 5, 1 (2001).

¹⁵ Richard A. Danner, Contemporary and Future Directions in American Legal Research: Responding to the Threat of the Available, 31 *Int'l J. Legal Info.* 179, 191 (2003).

¹⁶ Sarah Valentine, Legal Research as a Fundamental Skill: A Lifeboat for Students and Law Schools, 39 *U. Balt. L. Rev.* 173, 189 (2009-2010).

¹⁷ Danner, *supra* note 15, at 192.

¹⁸ Michael Robinson, Digital Nature and Digital Nurture: Libraries, Learning and the Digital Native, 29 *Libr. Mgmt.* 67, 68 (2008).

¹⁹ *Id.* at 68.

²⁰ Jaclyn Cabral, Is Generation Y Addicted to Social Media, 2 *The Elon J. of Undergraduate Res. in Comm.* 5 (2011).

Another perception shift has been that a valid education can be obtained through almost any format or method. Real-life concepts that appear in media or news are embraced as truth, regardless of their actual accuracy or bias. This new generation is quick to search for connections to something they have seen or heard online. Their reception and connection to social media is also formidable. They push for recognition that any one opinion carries the same weight as another, no matter the perceived background, and a collaborative understanding trumps any single voice.

New students have been increasingly dismissive of the traditional information structure, where the assumed value of having physical copies of information stored in libraries (as opposed to openly available on the Internet) has diminished significantly. Often students believe and trust their own search results more than information available within the library, even when confronted with proof otherwise.²¹²² If a resource doesn't come up under their search it is easily dismissed as unimportant or non-existent.²³ They may also miss valuable information because of their reliance on online search and preference for ease above other factors.²⁴

This new generation believes information that is not freely available on the Internet must lack intrinsic value or is not worth the effort to track down.²⁵ The prevalence of Wikipedia along with the ease of website design, video and photo editing give some youth the impression that when information or resources are valuable or of interest, someone will scan or digitize it. They display a growing acceptance of the power of collaborative information and the reliability expressed through tools like Wikipedia.

The concepts of copyright and intellectual property rights are often presumed malleable, on display through the proliferation of peer-to-peer music and video downloads. The ability to use, modify and pass on information without restriction, regardless of source or ownership is presumed.²⁶ A commercial database may be a somewhat foreign concept to some, while others see access to them as a step towards professional acceptance.

Upon entering law school, these students quickly seek out the more advanced Westlaw and LexisNexis legal information systems, assuming their digital skills will easily get them by.²⁷ Their expectation is that the same algorithms that generate results for general search engines will power these databases, that the same rules

²¹ Robinson, *supra* note 18, at 69.

²² Valentine, *supra* note 16, at 190.

²³ Lutishoor Salisbury et al., Information Seeking Pattern and "Googlization" of Information: the Issues Facing Libraries, 1st Int'l Conf. on Info. Mgmt. 412 (2007).

²⁴ *Id.*

²⁵ *Id.*

²⁶ Robinson, *supra* note 18, at 69.

²⁷ Valentine, *supra* note 16, at 190.

will apply.²⁸ They often seek to utilize their existing information world to access legal information rather than learn and rely on a new set of tools and methods developed for and inherent to the legal community.²⁹

What tools and methods does the legal community espouse? The following is a discussion of some of the fundamental elements of search.

e. Comprehensiveness

The number of sources of legal information is overwhelming. Primary law, created by legislatures and agencies or refined through the judicial process, is an expansive concept. Laws are legislated and passed by any number of legislative bodies. Administrative agencies hold hearings and make rules that are then published. Courts hear cases and provide rulings and opinions that need to be accessed. A comprehensive resource, even in a single jurisdiction, is difficult to imagine let alone find, and all of this happens at the federal, state and local levels.

Each of these bodies create information that must be studied and distilled to understand and apply appropriately. Secondary sources explain and clarify the law and are even more diverse, crossing hundreds of publishers and multiple formats. Students must understand the full spectrum of legal sources and recognize how to access each type of information. The format and tools to access each of these materials can differ drastically, some more efficiently than others.

Legal research must be comprehensive, embracing the full substance of an issue without leaving room for error. Reliance on a single source can often be fatal. Competent comprehensive research skills can prevent relying on overruled cases, discounted legal theories or from incorrectly distinguishing the holding of an opinion.³⁰

Access to legal materials, once concentrated in locations such as law libraries, can now be easily accessed online from anywhere. Convenience can be a powerful obstacle to proper research. Knowing when the research does not reach the depth it should or is not directly on point can be difficult when a simple search falls short of results.³¹ The ease of access and simplified search tools have proliferated lazy research skills.³²

The next generation legal search model has only further blurred the lines between the various legal sources. WestlawNext offers results from all the different databases in one fell swoop, forcing the user to then filter and identify the types of

²⁸ Danner, *supra* note 15, at 191.

²⁹ Danner, *supra* note 15, at 191.

³⁰ Valentine, *supra* note 16, at 196.

³¹ Robinson, *supra* note 18, at 70.

³² Valentine, *supra* note 16, at 189.

information that are applicable in their situation. A user may be tempted to rely on the top 10 hits, which are often cases, when a statute or agency ruling directly addresses the issue.

Broadening access to resources has heightened the need to plan out a search strategy. For each problem or issue, comprehensive searches require planning and execution. One needs to be able to recognize the different types of legal authority that impact the issue, connecting all the dots. In court, an argument not raised cannot be brought up retrospectively. Proactively identifying and researching issues is a critical skill.

f. Accuracy

Errors in legal research can lead to malpractice or worse. With the incredible influx of accessible information, how can one be sure that one's research accurately reflects the state of the law?

When legal resources were limited only to print, the accuracy of a work was of equal concern. Publishers of legal resources would examine a work to make sure that the author provided an accurate analysis or that a case was, to the best of their ability, free of transcription error. Any mistake would reflect poorly on a publisher and could reduce their overall perception. This vetting process was an important step to evaluating the accuracy of a legal work or report.

Online access to legal resources has frustrated this issue. The process used by publishers appears much less pronounced. Computer programs, while decent with spelling and grammar checks, allow for substantive mistakes. Format errors are not particularly uncommon on the digital platform, and one is often not informed if a human body has actually reviewed the content at all for conceptual errors.

Historically it took a significant amount of time and money to publish a print volume. The investment required was well out of the means of most commenters and virtually required a standing patron base. This investment was a key step in maintaining a quality driven collection of law and legal analysis.

Publishing online requires far less investment to accomplish. The threshold for disseminating works online is almost trivial now. As a result much more information is being published or re-published as updated but these works are often not checked for accuracy.

An additional concern for the user is that content is often created as a service by an interest group or an interested individual. Many of these legal resources can provide valuable commentary or critique. Without an editorial or evaluative process however, it can be difficult to distinguish between value added content and haphazard analysis. Determining the original author is also becoming more difficult.

Perhaps the most concerning of all is the movement to create collaborative legal information hubs. There are already a multitude of free sites, each with its respective disclaimer. These sites are often updated by non-professionals who copy-paste text from another source. Wikipedia is rife with examples of definitions of legal principles that may apply in some jurisdictions but not apply in others. Though not usually taken as the conclusive source, the trajectory of one's research can be impacted significantly by an inaccurate resource.

It is more important than ever to learn to distinguish vetted and accurate information sources.³³

g. Relevance

We may think we know what we are looking at when we see it. Many legal terms seem to speak for themselves. Despite our best efforts the actual relevance of the result may actually come into question.

One of the most difficult concepts in law is the relationship between a manifestation of the law and the issue one is presented with, making the leap to attach the two.³⁴ It is in making these connections that identifies a competent professional with a specialized degree in law.

In search terms, the relevance of the results is dependent on precision, finding more relevant results than irrelevant, and recall, finding all of the relevant results that are within the source database. Both principles are important in measuring the quality of a search engines results.³⁵ While it is easy to see how many relevant document were retrieved in a search, it is much more difficult to determine how many relevant document were not retrieved in a search.³⁶

To maximize recall, a user must anticipate the descriptive language used in the database for their issue. Using anything but the actual term used in the database will cause relevant cases to be overlooked.³⁷ Also, words that have many synonyms are problematic since multiple meanings will naturally force irrelevant results to appear. Information professionals have historically used controlled subject headings to alleviate this issue, but it is not an option when performing full-text

³³ Nancy P. Johnson, Best Practices: What First-Year Law Students Should Learn in a Legal Research Class, 28 Legal Reference Serv. Q. 77, 80 (2009).

³⁴ Danial P. Dabney, The Curse of Thamus: An Analysis of Full-Text Document Retrieval, 78 Law. Libr. J. 5, 15 (1986).

³⁵ Jeffrey Beall, How Google Uses Metadata to Improve Search Results, 59 The Ser. Libr. 40, 43 (2010).

³⁶ Dabney, *supra* note 34, at 15.

³⁷ Dabney, *supra* note 34, at 19.

searches.³⁸ A term used in one resource is not always consistent with use in other resources, and can vary in meaning.³⁹

Where the search terms are distinct enough the algorithm may be able to distinguish relevant matches.⁴⁰ It is when terms have multiple meanings or are vague that problems arise. Innovations such as tagging and filtering can help clarify ambiguous terms but take additional steps and often reduce results in unanticipated ways.

Relevance also work on a scale from most relevant to least relevant. Even though a blog post contains the proper terms and seems to be on point, it will not replace the need for an official version of a state regulation. A search engine may tweak its algorithm to prioritize certain types of sites, such as .gov or .edu. Unfortunately this process is not transparent.

Other influences can change the apparent relevance of a result as well. Current trends in social media have prompted most of the large search engines to take a user's social media use into account when formatting results. These connections may substantially change the information accessed through search engines. If someone was friends with a number of patent attorneys, a search for "procedure" could bring up multiple hits for the patent filing procedure.

Relevance, often determined by complicated algorithms, is one of the main factors in ranking. Most search engines today rank results by relevance. The relevance score or any individual result may vary from day to day as the large search engines are tweaking their algorithms constantly. As most users do not usually get past the first page of results this may not appear to be as substantial as it really is. Ultimately it is the search engine's relevance rater that will ultimately determine the full extent of a user's research.

h. Currency

Information is valuable when it is current. Many of the laws passed by legislatures are put into effect for only a certain time period with sunset provisions or other legislative tools. Many laws are revised each legislative term. Laws that are in force need to be distinguished with laws that have expired.

Cases are similar. When a court makes a decision it could be reinforcing current trends or it could be overruling a prior decision. When a court makes a ruling that is binding, prior opinions can be overruled and precedent can change. When this

³⁸ Alan Dawson & Val Hamilton, Optimising Metadata to Make High-Value Content More Accessible to Google Users, 62 J. of Documentation 307 (2006).

³⁹ Beall, *supra* note 35, at 49.

⁴⁰ Dabney, *supra* note 34, at 20.

happens any discussion on the prior case law should also change to reflect the current state of the law. This updating takes time and effort.

The computer introduced the ability to transfer data and communicate immediately via the Internet. While incredibly convenient, one misconception with the Internet deals with the currency of the information found online. Even though information appears as if it should be or easily could be current, many times it is not. Many websites and online access points do not record updates with a timestamp.

The print publishing world is not required to produce a copyright statement in order to obtain copyright protection anymore. This requirement went away in 1978, but the practice still exists. One reason it may continue in printed legal works is to validate the time period of the work. By including a notation of the edition and when it was printed, the user can get an idea of whether the material is too dated and can decide how to weigh the validity of what is contained.

When the publishing date is found on a website, it may have been automatically rendered at certain points without regard to updating. It could be changed or deleted at any time through only a few keystrokes. Sometimes databases are updated but the timestamp is not kept up to date. It is difficult to know how to rely on a timestamp located in a database or on a website. While many of the large commercial databases do update currency timestamps, they often include lengthy disclaimers.

Many government websites have been moving toward providing online access to their legal documents. A significant number of jurisdictions also try to provide the most current information in a timely manner. Often this comes with disclaimers that it is not the official version.

These fundamental elements of search have brought awareness to some growing concerns in the transition from print to digital resources regarding search.

- i. Information Overload

The amount of legal information is expanding at an exponential rate. There is no sign that this surge in information is going to level off or decrease in the coming years. The legal information market historically required a large investment to enter. Now distribution and publishing tools are freely available through the Internet. The cost of disseminating information is also at historic lows. Outside of the commercial market, free legal information sites directed toward the public are rapidly increasing. Blogs and social media tools give virtually anyone a voice, many who choose to write on legal topics. The source of new information appears endless.

With so many people able to create or reproduce legal information, the marketplace and information receptacles have been overloaded. The expansion of legal materials

has reached a critical level. One can only wonder how far off the serious consequences of the “threat of the available” may be.⁴¹

The way people access information has been shifting as well. Where once there was a few access points to news, the local newspaper or television station along with any information by word of mouth, this notion has been turned on its head. Information and opinions on virtually any topic imaginable can be found on the open Internet through common searches. The breadth of information is not the only facet, but also the depth of discussion for many topics. The law is more accessible than it even has been.

As adherents to the common law system, legislative actions as well as decisions of the court impact the state of the law. It is therefore important to understand how these bodies work. While legislative actions have been recorded and published by the government, for case law it was important to find a reporter that had compiled the cases.

For over 200 years court reporters have been busy recording the expansion of the law. Now there are more than 7,000,000 published judicial opinions with that number growing by more than 200,000 per year.⁴² This is just the case law. State and federal legislation adds another 50,000 pages every year.⁴³

Most of this information is now available over the Internet free of charge. The publically accessible material is just the tip of the iceberg however. There is an incredible amount of information being produced that is not open to the public, but locked behind commercial pay-walls or stored in commercial archives.

While search tools available through computer interfaces seem to have kept the growing amount of information in check, basic principles such as Zipf’s Law of Distribution are displaying the strain of too much information. While much is being made available, much is becoming unrecoverable through all but advanced searching techniques.

Our reliance on full-text search is obvious. Each search takes an amount of processing power to complete. As the collection size increases, the more processing power that is required and the less efficient the search will be.⁴⁴ Westlaw dealt with the issue by breaking up the information into databases. The demand for simple search and accessing all information at once has led to a reversal of course.⁴⁵

⁴¹ Danner, *supra* note 15, at 182 (quoting Karl Llewellyn).

⁴² Jacobstein, *supra* note 7, at 32.

⁴³ Danner, *supra* note 15, at 181.

⁴⁴ Dabney, *supra* note 34, at 21.

⁴⁵ Lada A. Adamic and Bernardo A. Huberman, Zipf’s Law and the Internet, 3 *Glottometrics* 143 (2002).

With all this information, legal databases and search engines are struggling to provide precise results. Fallout is a term used for the number of irrelevant documents produced in a single search.⁴⁶ As the sheer volume of results in any basic search is extreme, these search engines are fine tuned to limit almost all fallout.⁴⁷ It is unclear exactly how the companies are accomplishing this however.

One way of dealing with this concept is by forcing many of the most used resources to appear at the top of the search results. An example would be a search for “President of the United States” in Google. No matter how the algorithm weighs the factors, the official site at www.whitehouse.gov will display on the first page of results, as it should. For every source that is hardcoded however, a potentially relevant source that may have ranked higher will slip to the second page into obscurity. In the situation above that may not matter so much but it is a form of censorship, a type that is concealed from plain view.

The large legal databases and search engines also track search patterns. The resources that get more hits crawl up to the top of relevance whether accurately reflecting the quality of the resources or not. A catchy phrase or subtle joke in a title can impact the veracity of the algorithm.

The popularity of Wikipedia has provided it with a unique position. In almost any Google search, the Wikipedia page will rise in the results. The same is true for a definition when a single term is entered. The partnership between a search engine and other organizations can raise more concerns that important information may be superseded or left off the list.

With the exponential growth of information, databases and search engines are constantly working to improve their results, but it is unclear how they are doing it. Tools to organize and sort the information are available but bring other issues.

j. Reliance on Metadata

Computers changed the face of information storage and access. Computers can store vast amounts of information while quickly indexing this data in very complicated ways that benefit the user. Each digital resource has descriptive information attached to it called metadata. Metadata, in this context, is “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.”⁴⁸ This data is often not located within the body of the document and may not show up in full-text searches.

⁴⁶ Dabney, *supra* note 34, at 16.

⁴⁷ Dabney, *supra* note 34, at 17.

⁴⁸ National Information Standards Organization (NISO), *Understanding Metadata*, 1 (2004), available at <http://www.niso.org/>.

Libraries were some of the first innovators of metadata and indexing. As they developed standards, the ideas were often freely available and exchanged. A hold-over from the traditional card catalog, library records are still often indexed by hand.⁴⁹

The Internet has brought new challenges to the open standards that librarians developed over time. In many areas metadata and their governing standards have morphed into encoded scripts and are protected as trade secrets.⁵⁰ The processes have also been streamlined and new metadata is often lifted directly from the content without human input.⁵¹

The National Information Standards Organization (NISO) has defined good metadata as “(1) conforms to community standards; (2) supports interoperability; (3) uses authority control and content standards; (4) includes a clear statement of the conditions and terms of use; (5) supports long-term curatorship and preservation; and (6) should have the qualities of good objects, including authority, authenticity, archivability, persistence, and unique identification.”⁵²

There is still no single standard that has been generally adopted for metadata. Numerous professions and organizations have developed their own standards to organize the different types of information to their needs. Information such as title, author and keywords associated with the information resource are high priority elements and obviously included in most all standards. How detailed and necessarily complex the data needs to be from there is largely disputed.⁵³ One well recognized standard, the Dublin Core Metadata Element Set, lays out 15 common elements that should be attached to each digital resource.⁵⁴

The amount of metadata attached to any particular resource can vary drastically, but as long as it is accurate, more elements will provide higher precision in search. Each metadata field gives key points of information for the type of resource being described. These elements may contain information that is not in the actual text of the digital resource, thus the only place to find these terms would be in the metadata. Even when following a defined standard however, there is no guarantee that each and every record field contains information.

What happens when metadata is incomplete? Any missing elements may influence the results of a search, making it more difficult to locate or miss a key term that normally would have resulted in a search hit. Ideally important digital resources

⁴⁹ Beall, *supra* note 35, at 45.

⁵⁰ Beall, *supra* note 35, at 43.

⁵¹ Beall, *supra* note 35, at 49.

⁵² Jung-Ran Park, Metadata Quality in Digital Repositories: A Survey of the Current State of Art, 47 *Cataloging & Classification Q.* 213 (2009).

⁵³ NISO, *supra* note 47.

⁵⁴ NISO, *supra* note 47.

will have complete metadata records, but this isn't always the case. Sometimes the information isn't known at the time the metadata is created. Sometimes it can be corrupted or lost.

Metadata attached to a document can often contain mistakes. User input information is subject to typos and spelling errors. Computer generated information can be pulled from the wrong source or can be corrupted. Once an error is created in metadata it can be very difficult to correct.

Enormous amounts of digital material are being made available online every day. Despite recent improvements, the digitization of print material by scanning still is not perfectly accurate. This makes the metadata attached to scanned digitized works even more crucial.⁵⁵ One study using the popular search tool Google Books highlights the difficulties of inserting correct metadata and the obstacles involved with correcting faulty metadata.⁵⁶ Similar issues have been found in Google Scholar as well.⁵⁷

Search engines rely on metadata to evaluate and sort their results for relevance. A term that shows up in a full-text search may appear in the results, but when a search term shows up in the topic section of metadata it is much more likely to raise the ranking of that item within the search results. Metadata is used to distinguish items that have the same titles, separate various editions of the same material and locate similar or related works.⁵⁸ These issues are especially poignant in legal research due to the need for accuracy and currency discussed above.

In addition to documents, websites also use metadata in certain fields such as meta tags. A search engine could originally use these tags to measure the relevance of a site based on the terms used. Other tags in website encoding can also be flagged and used for indexing and searching as well.

Data pulled from a website for use by a search engine is limited however. User abuse of the meta tag has caused the functionality of the tag to be marginalized. Webmasters have the ability to alter and supplement these tags with irrelevant or commercially advantageous terms, masking the actual nature of the website. Due to the potential for abuse, search engines are hesitant to include specific meta tags in their algorithms.⁵⁹ Notably, Google only recognizes and weighs the HTML title tag in its search results.⁶⁰

⁵⁵ Julia T. Pope & Robert P. Holley, Google Book Search and Metadata, 49 *Cataloging & Classification Q.* 1 (2011).

⁵⁶ *Id.*

⁵⁷ *Id.*

⁵⁸ NISO, *supra* note 47.

⁵⁹ Dawson, *supra* note 38, at 310.

⁶⁰ Dawson, *supra* note 38, at 315.

While numerous standards for metadata exist, such as the Dublin Core Set, each must adjust and change with to keep up with new technologies and meet user needs. Metadata that was created previously may not provide sufficient information to meet the future needs. Search engines are constantly fine-tuning their products, changing the weight of fields and data, complicating the efforts to create a uniform standard. The definition of an official source or document online for metadata standardization has also been unreliable.

k. Authentication

The Internet provides instant access to many valuable stores of legal information. Many of these sites provide access to primary legal documents one may assume is reliable. Government legislative, executive and even judicial websites have increasingly provided access to primary law as an informative service to their citizen base. The information provided is most often free and open to simple search through search engines.

Users may assume that accessing the website of the originator of the law or rule would provide the most accurate version. The expectation is that the information provided digitally, whether on the webpage or downloaded in a document, is exactly as it would be if it the user located it in print. Unfortunately this isn't always the case. Once a book is printed it is difficult to change or fix without reprinting. Digital content is much easier to change quickly and without raising alarm.

The validity of an online document is established through a process called authentication.⁶¹ Authentication is the method a publisher uses to convey to the user that the document received by the user is unaltered from the original version.⁶² The goal is to make the digital version of a document as trustworthy and resistant to lapses in management and control, corruption or tampering as the print equivalent.⁶³ Authenticating is not just a digital stamp of approval however, but includes a series of steps to secure the access through backend management, chain of custody, integrity of transmission and long-term repository management.⁶⁴

While true that most of the time authentication may seem like a trivial matter as the information is accurate as presented in the digital format, without a monitoring system in place, unethical or malicious individuals could try to disrupt the flow of accurate information from the source. Authentication also helps mitigate human error.

⁶¹ ALAN S. KOWLOWITZ, *OPENING GOVERNMENT'S OFFICIAL LEGAL MATERIALS: AUTHENTICITY AND INTEGRITY IN THE DIGITAL WORLD* (2012).

⁶² UNIF. ELEC. LEGAL MATERIAL ACT (2011), http://www.uniformlaws.org/Shared/Docs/AM2011_Prestyle%20Finals/UELMA_PreStyleFinal_Jul11.pdf.

⁶³ Richard J. Matthews & Mary Alice Baish, American Ass'n of Law Libraries, Elec. Legal Info. Access & Citation Comm., *State-by-State Report on Authentication of Online Legal Resources* 7 (2007), http://www.aallnet.org/Documents/Government-Relations/authen_rppt/authenfinalreport.pdf.

⁶⁴ Kowlowitz, *supra* note 60.

Today authentication of primary legal material available online is of enormous concern. An American Association of Law Libraries report concluded that in 2007, “no state’s online primary legal resources [were] *authentic* – meaning that none of those resources [were] *authenticated* or afford[ed] ready authentication by standard methods.”⁶⁵ A follow up report published in 2010 noted that while awareness had been raised, many of the same authentication concerns had not yet addressed.⁶⁶

The immediate response in many states had been the introduction of various disclaimers on government websites as to the accuracy of the legal information contained in them.⁶⁷ This facilitates basic research through simplified search of the legal materials but then requires the user to take additional steps to verify the accuracy of the legal material through the equivalent print resource.⁶⁸

Recently a model law called UELMA was drafted to address the concerns of authentication.⁶⁹ While it does not designate a standard that all government bodies must adhere to, it would require a state to authenticate all of its official documents.⁷⁰ Until these protections are put into place, government websites or other online resources should not be completely relied on as the official versions and knowledge of the corresponding print versions is critical.

Reliance on online search features may provide access to primary legal materials, but few are authenticated and confirmation by going to the official print resource will be critical.

I. Current Status

After two decades, the same fundamental lawyering skill, legal research, is required for the practice of law. But how would MacCrate characterize our progress in the past twenty years? Do lawyers graduating today have better familiarity with the basic research tools? The answer would likely be very similar to the concerns expressed in the original report. Despite the transition from basic computerized services to a widespread online interface, the key principles remain unchanged.

⁶⁵ Matthews, *supra* note 62, at 13.

⁶⁶ American Ass’n of Law Libraries, Elec. Legal Info. Access & Citation Comm., *2009-10 Updates to the State-by-State Report on Authentication of Online Legal Resources*, 1 (Tina S. Ching ed., 2010), <http://www.aallnet.org/Documents/Government-Relations/authen-rprt-updates/2009aallauthenticationreportupdates.pdf>.

⁶⁷ *Id.* at 1.

⁶⁸ Matthews, *supra* note 62, at 7.

⁶⁹ UNIF. ELEC. LEGAL MATERIAL ACT (2011) http://www.uniformlaws.org/Shared/Docs/AM2011_Prestyle%20Finals/UELMA_PreStyleFinal_Jul11.pdf.

⁷⁰ *Id.*

The challenge today is developing easier access to the variety of sources without compromising quality or comprehensiveness. The students arriving at law school today may not be prepared to do anything beyond simple search in Google. The goal of MacCrate was to improve legal education and one aspect of that will be teaching this generation to use proper research methods even if it conflicts with their previous experience online or their natural tendencies.⁷¹

Print is by no means obsolete, but there are an ever-increasing number of advantages to do research online. Ultimately it is up to the student to decide which format provides the most effective for the situation. No matter the tool one chooses to use, good study habit and preparation before initiating a search will save time and effort in the long run. Poorly crafted search terms, whether online or in print, can result in significant lost time and energy.⁷²

Some materials are easier to understand or find in a well-organized print volume. Pinpoint searching in a legal database may be ideal for locating a single statute, but seeing where that statute fits in its section or within the whole may be harder to comprehend than if visualized in the volume or set of a print resource.

Often the legal research professor fights an uphill battle, where students feel that electronic research is the always the most efficient method. By demonstrating the pros and cons of both print and online materials however, students are provided with the knowledge and tools to implement efficient research as the need arises.⁷³

The number of innovative tools used for finding and managing data is growing rapidly, so staying aware of the changes is a constant challenge. Each new technological step changes how law and information are accessed.⁷⁴ It is for the legal research instructor to be familiar with and clearly teach the strengths and limitations of each format.⁷⁵

Although the majority of students may feel proficient in natural language searching, some understanding of the mechanics of a search engine and how to evaluate the weight of results may be particularly useful to explain. Search engines can examine amazing amounts of information and provide results in the blink of an eye, but they do little to help the user figure out the relevancy of the information displayed.⁷⁶ A student who understands some of the shortcomings of simple search is better prepared to look deeper when the answer doesn't seem to match the situation.

⁷¹ Valentine, *supra* note 16, at 189.

⁷² Johnson, *supra* note 33, at 86.

⁷³ Johnson, *supra* note 33, at 97.

⁷⁴ Valentine, *supra* note 16, at 175.

⁷⁵ Johnson, *supra* note 33, at 80.

⁷⁶ Robinson, *supra* note 18, at 69.

Choosing to begin with Google may provide a bird's eye view of an issue or topic. The simple search is likely to pull a wide variety of results, often from outside the desired topic, due to the breadth of the sources Google pulls from. Narrowing results or focusing can be very difficult. More filtering tools are available after the initial search than ever before. Tagging and collaborate evaluative tools are also growing in popularity but may not be as effective for legal research.

Advanced search features using Boolean operators and other functions can be very powerful in full-text search.⁷⁷ When targeting a specific document or using distinct phrases this type of search can be very effective.⁷⁸ Westlaw and LexisNexis still provide access to these tools. Students may not be as familiar with pinpoint searching so additional instruction may be necessary. Law firms do not provide the luxury of multiple searches with vague terms.⁷⁹

Regardless of the search functions used, a student must be aware of the types of legal sources that are available. For example, WestlawNext will provide eleven different types of documents in a basic search; cases, statutes, regulations, administrative decisions, trial court orders, secondary sources, forms, briefs, trial court documents, expert materials and jury verdicts and settlements. A student who cannot distinguish which resource is best suited for their need will not be able to efficiently research the law using such a complicated result list.

It is important to understand the amount of control one is relinquishing in the various search activities and weigh the reliability in turn. Turning to Google, the user maintains very little control over their search, relying solely on Google's algorithm to get them the most relevant results. With an increasing amount of fixed results (Wikipedia, online dictionaries, etc.) and other balancing schemes, skewed or mandated results near the top are not uncommon.

Westlaw or LexisNexis offer both simple and advanced search options, but act very differently due to the specialized scope of the databases. Students need to be shown the differences between these databases and Google, as they do not function the same way.⁸⁰

Advanced search tools offer more control over the existence and placement of exact terms. "Law school programs should emphasize both the natural language and basic Boolean searching, such as how to expand a word, how to look for alternate spellings, and how to use proximity operators."⁸¹

⁷⁷ Jill L. K. Brooks, Great Expectations: New Associates' Research Skills from Law School to Law Firm, 28 Legal Reference Serv. Q. 291, 293 (2009).

⁷⁸ Johnson, *supra* note 33, at 87.

⁷⁹ Brooks, *supra* note 76, at 299.

⁸⁰ Brooks, *supra* note 76, at 294.

⁸¹ Brooks, *supra* note 76, at 293.

Students should be aware of the trend for government resources to be made available online. Official of authenticated government resources are not available in all jurisdictions. Even when there are, search engines may not always prioritize the official or authenticated resources in their results. Free resources provided by governmental websites often provide a different organizational structure than the big legal databases and can often be very effective research tools.⁸²

Search within databases and online resources has added an amazingly productive facet to legal research but it still is important to understand how to find things outside of the “app for that” mentality. Despite what may sound like their death knell with the persistent introduction and adoption of all things digital, print resources still persist and can be effectively utilized.

m. Future Application

One thing that the MacCrate Report does emphasize that still applies today is the value found in maintaining a variety of sources. The transition to online legal databases has been swift but recognizing the shortcomings outlined above, we shouldn't be so quick to disregard the tools that have been used for so long.

The reality is that now most legal research occurs over the Internet and utilizes a combination of search methods. A user may not always be aware or informed of the processes involved with their search. A user may believe they are using a full text search feature on a website when that may not be the case at all. The search process for most databases is far from transparent.

Complicated mathematical algorithms dictate the results in the largest search engines. Commercial and social media inputs are increasingly influencing the order of these results. Many of the factors are undisclosed, meaning that a user cannot be completely sure how the results are being populated. But as long as the results appear to be accurate, many of us trust with blind reliance in the results.

The overwhelming popularity of the Google search box has become a phenomenon. Many sites have adopted the algorithm as their own, placing a Google powered search box on the front page and integrating their pages to cater to the search giant. Many users have expressed frustration with the current model however. The overwhelming influence of simple search may be on the decline as many users are beginning to recognize the need for recall and precision in their search activities.⁸³

Google rose to dominate the industry by tracking better indicators of content quality than anyone else.⁸⁴ It will continue to evolve in a way that benefits users or it will

⁸² Brooks, *supra* note 76, at 294.

⁸³ Beall, *supra* note 35, at 52.

⁸⁴ Shane Snow, *How Social Media Affects Content Relevance in Search*, MASHABLE (May 10, 2012, 10:45 PM), <http://mashable.com/2011/09/09/seo-social-media/>.

disappear. It has already taken on a “library like” appearance through tools such as Google Books and Google Scholar.⁸⁵ Yet many of the large initiatives started by Google could fall by the wayside as well.

Unless we provide our students with knowledge of the various means and formats of legal research, we may see that within a generation, “electronically manufactured” law is the accepted norm.⁸⁶

⁸⁵ Robinson, *supra* note 18, at 70.

⁸⁶ Valentine, *supra* note 16, at 176.