Editorial

Open Access

Archives Of Breast Cancer

DOI: 10.19187/abc.20163370-73

The Challenge of Obtaining and Maintaining Expertise

David Krag*[a]

[a] Department of Surgical Oncology, University of Vermont, Vermont, USA

Expertise is a special skill or knowledge. An advanced expert also knows the boundaries of that special knowledge. Knowing the boundaries of what is known and not known about a topic puts the expert in the best position to teach, provide medical care or design new questions. Experts continually struggle to maintain their expertise. However, despite substantial efforts, the goal to be on top of the full set of information in a topic and know the boundaries of that information is elusive.

Expertise is highly valued. When faced with a new diagnosis of breast cancer, patients want the person with the most expertise to treat them. Confidence in the physician is high when that person is the recognized expert and knows the boundaries of treatment options. A grant review committee readily identifies an expertly written application. The evidence of expertise of the research topic includes well-chosen citations and a well-designed plan that that steps beyond the edge of what is known. The struggle to achieve expertise is not trivial. The methods used to achieve and maintain expertise are not discussed enough. We present tools to help ameliorate this struggle as it relates to published biomedical information.

Development of the National Library of Medicine (NLM) in Bethesda Maryland is one of the most profound events in the milestones of human knowledge. This is an incomparable data base of published biomedical information that is freely and instantly available around the world via the internet. This data base of published peer reviewed biomedical information provides the foundation of expertise in biomedical sciences. The rate of new published information continuously increases. By the year 2000, the number of new articles entered into the NLM began to exceed 500,000 per year. Within a decade, the number of new articles doubled to about 1,000,000 articles per year. As of August 2016, the total number of articles at NLM was 27,489,908. This is "big data" by any standard.

"Big data" typically refers to managing and standardizing large sets of biological information. Examples include sequences of nucleotides, amino acids, and multiple types of interactomes such as binding interactions between thousands of antibodies and their targets. Tools to manage this information are essential since a single lab can routinely produce sequence information in the billions. Tools to manage this information include an expanding repertoire of powerful algorithms to analyze and compare sequence information. The tools also address more mundane tasks for example submission of sequences, data storage, and interfaces that facilitate making sense of the data. Without these latter tools, the otherwise powerful analytic algorithms would not be functional.

Achieving and maintaining expertise in biomedicine means interacting with the 27,000,000 biomedical articles at NLM which increases by 1,000,000 each year. What tools are available for the user of this "big data" and how can they help establish and maintain expertise? There are excellent search tools such as PubMed to identify sets of articles. However, software tools to help a person interact with this massive amount of data have not kept pace. Simply finding a list of articles is no longer sufficient. Multiple complex tasks are involved to interact with this data set such as identification of articles, getting the full text version, reading the documents, summarizing the key observations, articulating the content in context with a person's intellectual data base, organizing

Address for correspondence:
David Krag, MD
Address: SD Ireland Professor of Surgical Oncology, 89 Beaumont Ave, University of Vermont, College of Medicine, Given Bldg Rm E309, Burlington, VT 05405, USA
Email: david.krag@uvm.edu
Tel: +1 802 6565830
Fax:+1 802 6565833

articles for projects, and sharing information. Ultimately, this information is used for clinical decisions, teaching, and research.

I was fortunate to participate in developing a new surgical method that is now widely used. In 1993 the first article was published describing the radiotracer method for identification and surgical removal of sentinel nodes.[1] The National Cancer Institute funded our research which led to the world's largest randomized surgical trial in breast cancer.[2] Clearly, there was motivation to maintain the highest level of expertise. The number of published articles on sentinel nodes and breast cancer exploded. When that number reached 500 it became obvious that the tools available to manage this data set were inadequate. When that number reached 1000 published articles a decision had to be made to either give up tracking it all or develop new tools. For the next 10 years I was fortunate to lead a group of software engineers and bioinformaticists to develop such tools. I also became the primary experimental subject. This research project in literature management included careful examination of the multitude of mundane steps involved with managing literature. Even on initial analysis, it was clear that the time involved with mundane tasks exceeded the time of the intellectual task to read and understand the key observations of an article. The finding that the average time involved in the tedious nonintellectual tasks was greater than the intellectual task of reading an article was the first of three important and unexpected findings of this research. This indicated that major gains could be made by software that focused on the nonintellectual tasks related to managing literature.

The end result of this research is an online software system, now commercially available, at www.refbin.com. This represents the current version of the software developed over 10 years of research related to tasks associated with managing published literature. The prototype clinical topic was all articles on sentinel nodes and breast cancer which now includes more than 5,600 articles. The prototype basic science topic was all articles involved with phage display which now includes more than 7,700 articles. Both of these data bases are freely available at www.treeofmedicine.com. Working with these two data bases helped established the core principles involved with managing published literature. The following includes some of the key features of the software designed to solve critical problems in managing published literature.

Automated searches

To establish and maintain expertise the expert must read existing relevant literature and continue to read newly published literature. Automating the acquisition of these articles is not trivial. A simple list is not good enough. The software must interact with the expert to make sure that over time, only new articles are presented. The software must have a memory of all prior articles reviewed by the expert. The output of multiple automated searches will overlap and the software needs to sort this out so that the expert is not confronted with redundant articles. The articles must be presented to the user so that decisions on importing or excluding are rapid. The average time in refbin for the expert to make a decision on an article is about 5 seconds. This means that very large sets of articles can be screened. At this rate 2,880 articles can be screened in 4 hours. Once the initial screen of a newly created automated search is complete, the expert can be confident that they have pulled virtually every key article related to their topic out of the 27,000,000 at NLM. The boundaries of published information on that topic will then have been have been defined. Every 24 hours refbin repeats the search but only displays new articles to the expert. The work load then drops dramatically and the expert remains up to date on that topic.

Getting full text PDFs

This is one of the tedious and frustrating tasks that is largely addressed by the refbin software. In about 10 seconds, assuming the article is available electronically to the expert; a copy of the full text PDF is uploaded to their account and correctly affiliated with the citation.

Retrieval of single articles of interest

When reading an article, a citation in the reference list of that article is commonly of interest. In refbin, the expert copies a portion of the citation and a query to PubMed is generated. The desired article is instantly uploaded to the expert's account.

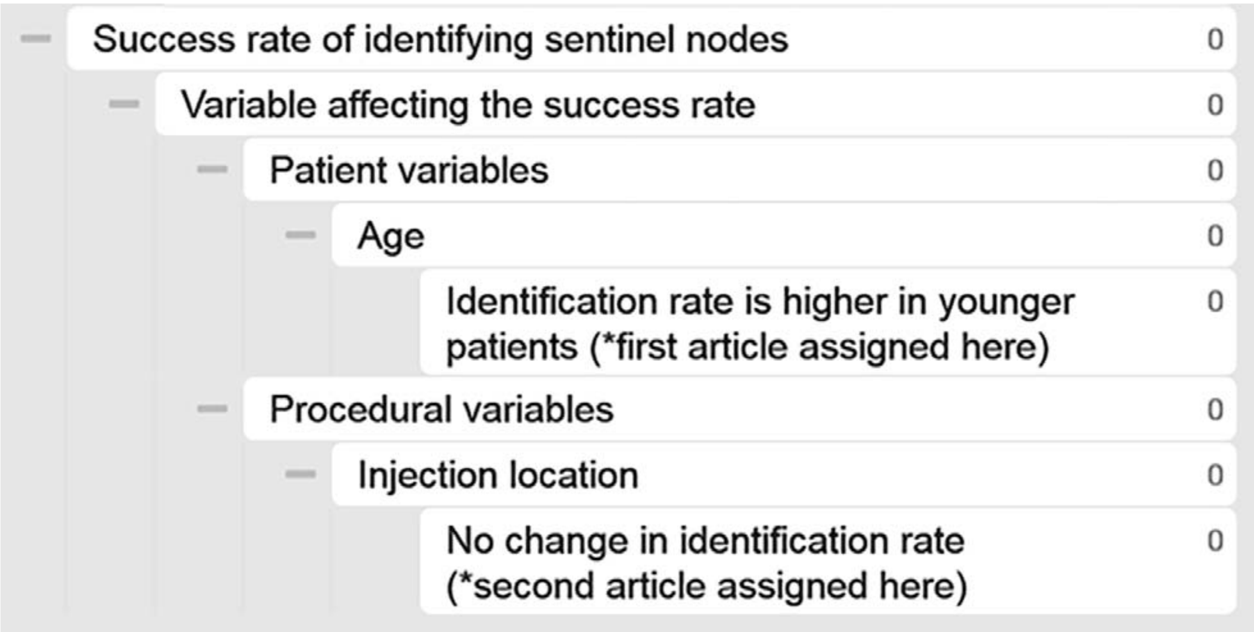
Describing key observations from an article

The second important finding of this research is related to the intellectual events that take place when an expert reads an article. The expert reaches a moment when an observation presented in the article becomes clear. At this moment the expert grasps the idea as an element of content. This element of content is important and the expert readily identifies the various key observations of the article. The key observation is a circumscribed unit of information. Putting a boundary around this unit of information is a uniquely human intellectual task. Computers are not yet in a position to accomplish this task. Identification and expression of these key observations are also not readily expressible through key words or by an ontological approach. The expert can extract key observations from the article because the expert already has background knowledge that allows understanding of the observation presented in the article. This means the expert has an intellectual framework to put the new idea into context. The new observation gets incorporated into this framework making the framework just a bit bigger. Imagine visualizing the structure that represents the intellectual framework of an expert. This would allow externalization of the knowledge of the expert and provide a remarkable tool to share that information.

Translation and merging of ideas and observations

The third unanticipated outcome of this research is related to how the expert can merge multiple new ideas learned from biomedical articles. Put aside keywords for a moment. For most readers and authors key words do not have much utility. What does have utility is the language we use every day to communicate with students and our colleagues and in our own internal thinking. We use narrative statements that are not bounded by strict rules affiliated with keywords or ontology. We simply make a statement. The result is an unambiguous description of a key observation. For example, a key observation in an article is "the success rate of identifying sentinel nodes is higher in younger patients than older patients with breast cancer". There are many different ways to say the same thing with slightly different words but it will remain a clear and unambiguous statement. A next article describes "the success rate of identifying sentinel nodes in breast cancer patients is not affected by the site of injection of tracer". Refbin begins with blank data fields where each idea is typed as a complete narrative statement. The two above ideas work well in that they are both unambiguous. However, after entry of multiple narrative statements this approach becomes burdensome and less useful. The solution is to fragment and merge the narrative statements.

Narrative statement one and two are merged as follows:



From the words "Identification rate is higher in younger patients", the parent phrases, "age, patient variables, variables affecting the success rate, and success rate of identifying the SNs" are the fragments of the original narrative statement. The next narrative statement related to injection location shares the first two parent headings "Success rate of identifying sentinel nodes" and "Variables affecting the success rate".

This strategy merges and condenses the two observations. Adding observations from additional articles on the success rate of identifying SNs becomes much simpler with this framework in place.

The result of this third observation relating to articulating and merging narrative statements results in a written display of an intellectual framework of knowledge. This structure shows the thinking process of the expert. Development and merging of observations based on intact narrative statements begins free form with an unambiguous statement. It becomes obvious to the expert how to merge observations. This allows intact narrative ideas to be manipulated as units of information in ways that are very helpful to the expert. The first benefit is the generation of a visual framework of information. The second is that this allows unexpected and massive condensation of information. Basically all observations in articles that present data on the variables affecting success rate of SN surgery can be merged together in a short section. The expert is then free to see all at once the variables that collectively affect success rate. This allows easier thinking by the expert about everything that might affect the success rate of SN surgery.

How does the process of setting up this framework begin? It does not require premeditation.

Simply making the narrative statements and merging them as they present themselves during literature reading begins to create a structure. There are no limits to the number of observations that can be made for an article. The refbin software allows moving and merging statements at will. The expert lets the observations "self-declare" through narrative statements. This begins to replicate and externalize the structure of the ideas that are in the mind of the expert. We have found this framework to be dynamic and help the expert rethink ideas. Ideas merge and when visualized can be pondered further to help create new ideas.

Workflow

Multiple tasks are associated with published articles. Examples include organizing articles for preparing a grant or a manuscript. Typically this can be one of the most tedious challenges associated with grant or manuscript writing. A parent and child field-based structure in refbin, identical to the structure for entering and merging narrative statements about articles, is used for project management. For example a first level parent heading describes the project, "R01 grant application on immunology of sentinel nodes due March". Multiple child headings may include "first pass articles to read", "articles likely to use", "articles not to use", "articles published by grant review committee members", or "articles not available in the library". Articles are assigned to and moved about to additional headings as appropriate. Such entries are at the complete discretion of the expert. This important application saves considerable time and keeps articles organized during the entire process of grant or manuscript preparation.

Sharing

Sharing ideas, citations and PDFs are integral to clinical care, research and teaching. In refbin, the primary account holder allows others to share some or all of the information in their data base. Sharing is done online from anywhere which considerably facilitates joint projects. This allows clinical, teaching or research groups to combine efforts and jointly build larger data bases.

In summary, the software tools in refbin considerably reduce the time involved with frustrating mundane tasks associated with managing published literature. This liberates experts and students to focus on the intellectual aspects of reading and thinking. Refbin software supports the expert to extract key observations from articles using only common narrative descriptions without needing to learn or confine thinking to strict ontology or a simple file system. Essentially, each expert defines their own ontology. Importantly, using a field-based parent-child system allows free-form merging of ideas and observations. Remarkably this results in massive condensation of information that is built on a structure defined by the expert. Through simplification of large amounts of data, this structure opens up new opportunities to think about biomedical problems. It also allows the expert to share this structure with others in a way not otherwise possible.

Conflict of interest

The author declares ownership in Plomics which produces Refbin.

References

1. Krag DN, Weaver DL, Alex JC, Fairbank JT. Surgical resection and radiolocalization of the sentinel lymph node in breast cancer using a gamma probe. Surg Oncol 1993; 2(6): 335-9; discussion 40.
2. Krag DN, Anderson SJ, Julian TB, Brown AM, Harlow SP, Costantino JP, *et al.* Sentinel-lymph-node resection compared with conventional axillary-lymph-node dissection in clinically node-negative patients with breast cancer: overall survival findings from the NSABP B-32 randomised phase 3 trial. Lancet Oncol 2010; 11(10): 927-33.