Searching for Veterinary Information: Balancing Precision and Recall

I am often asked what the best way is to search for veterinary information. The short, catchy answer is that the best way is the easiest search for you to create that also gets you to the results you need. As with most short, catchy answers, there is a problem: What does that mean?

A caveat when discussing the best search strategy is that there is, of course, no one single best search strategy for all situations. The best will vary depending on your question or objective, where you are searching and the type of information you need.

Discussions about improving searching generally stem from discussions about frustration with results. The most common frustrations I hear from practitioners, researchers and students are "finding too few results" and "finding too many results." In information retrieval literature, this is the tension between precision and recall. It is a balancing act.

A simple explanation is that high-precision searches have a smaller number of results overall and some of them are relevant. A high-recall search has a larger overall number of results, including more that are relevant and more that are irrelevant.

Precision is the amount of (potentially) relevant citations in your search results. The tradeoff is that there is a high risk that relevant citations are missing from the results. The search has identified some of the relevant literature and has a lower burden on the

By Heather K. Moberly
Coordinator of Veterinary Services, Medical Sciences Library, Texas A&M University

part of the searcher to spend time looking through the results to identify the relevant. This is a good search strategy when the searcher needs to identify some relevant citations but does not need comprehensive results. Examples of situations that work well for precise searches are those that have clearly defined concepts and questions that can be answered with a high degree of certainty. Ultimately, the searcher only needs a few good citations.

Recall is the amount of (potentially) relevant citations missed in your search results. High-recall search results are larger result sets than precision searches. The high-recall results include more relevant citations than the precise search and also include more irrelevant citations. In trade for retrieving a higher number of relevant citations, the searcher needs to spend more time and energy looking through irrelevant citations to locate the relevant ones. Examples of situations that work well for recall searches are those with difficult to define concepts in which the question requires a higher number of sources.

Strategies
Start with the basics. Thinking about your question or objective and the kind of results you need before you search

can help get the best results quickly. Beginning with a few search terms that represent the basic concepts in your topic can be the best idea. Do a search of the basic concepts, which would be high recall, and then narrow your results by publication date, language or publication type.

After the initial search and refining, decide if you want to increase precision or recall.
• To increase precision, add an additional concept to your search.
• To increase recall, remove a concept or increase the synonyms used to describe the concepts.

It is uncommon to find what you need in a single search. In today's universe of online searching, it is quick to do a search, peek at the results, tweak the search and run the new search. Don't think of running more than one search as a failure. Think in terms of iterative searching, and use precision and recall techniques to get the results you need.

Comparing precision and recall in information retrieval is a topic broad enough and deep enough to be covered in articles, chapters and books. For those who want a description of the mathematical relationship between precision and recall, try Buckland and Gey's 1994 article, "The relationship between recall and precision," in the *Journal of the American Society for Information Retrieval* (vol 45, pgs 12-19). **TV**