Application of Search Analytics in the Healthcare Profession

Save to myBoK

By Daniel L. Regard and Ron Hedges, JD

As the healthcare industry moves into an electronic health record (EHR)-only environment, practitioners increasingly have to "search" electronically stored information (ESI). They may need to do this to find, for example, the date on which a patient was given a specific drug and the dosage given. It sounds simple to look through a given volume of ESI for a specific bit of information, but is it simple across a large data set?

As EHR use grows, the variety of "sources" of ESI multiply—consider wearable monitoring devices and the accelerating velocity of the flow of ESI into records-keeping systems. Analytics, defined as a variety of data mining and machine learning techniques, may offer answers to these perplexing questions.

Use the Right Tool for the Right Search

Over the past few years there has been a dramatic increase of interest in the area of Big Data. Big Data is a broad term for data sets so large or complex that traditional data processing applications are inadequate. This has stimulated the emergence of new tools for storing, managing, and searching very large data sets.

A natural question for health information management (HIM) professionals is whether or not these tools and techniques will benefit the healthcare arena, and if so, how. The emergence of Big Data—and the ever-increasing volume and variety of electronic information which that trend encompasses—imposes obligations on HIM personnel to embrace appropriate tools and techniques in order to make sense out of Big Data and to manage electronic information for the benefit of everyone.

The short answer is "Yes," emerging tools and techniques will help. But the devil is in the details. The truth is there are many different types of data involved in healthcare—individual healthcare records, large databases of pharmaceutical trials, and multi-decade collections of prescriptions and outcomes data.

Data sets are expected to grow with the popularity of fitness trackers such as the Fitbit, Nike Fuelband, and soon-to-be-released Apple Watch. In addition to variances in the data sets, different users may pose dramatically different questions and analyses to their data sets. Not surprisingly, not all tools or all techniques work for all data sets and all questions.

Before organizations look for the best tools, they should start by understanding their data. Some key considerations include:

- **Quantity.** Does data size exceed recommended specifications for standard storage, computer memory, or database record counts?
- **Structure.** Is the data highly structured, such as a database with well defined fields of information, or highly unstructured, such as physician notes and e-mail communications?
- **Integrity.** Is the data clean? Was it collected in a manner that had sufficient controls that maintained consistent and accurate values or use of fields in the data?

Once an organization has assessed its data, it should also consider what it's looking to get out of the data. Data searches could include the following, many of which come from the field of data mining:

- **Retrieval.** Are users trying to retrieve known records?
- **Classification.** Can certain patient behaviors be classified as more risky or less risky?
- **Prediction.** Are users trying to forecast outcomes from inputs?
- **Exploration.** Are users trying to understand trends and segments within a data set?
- **Affinity.** Are there two or more activities that are correlated? Is there a query as to causation?

As one might assume, there are different tools and techniques for these data characteristics and data mining objectives. Some real world choices might be:

- **Single search.** Users only want a single record, and it has to be the right one.
- **Single-best search.** Users want to retrieve many candidates for their search, but ranked in a list of most-likely to least-likely.
- **Groupings search.** Users need to retrieve all records that meet a certain criteria and all of these will be used for a secondary purpose.

'Availability' Under the IGPHC

In 2014, the American Health Information Management Association (AHIMA) published the Information Governance Principles for Healthcare (IGPHC), which outlines broad and comprehensive principles of information governance for healthcare. One of those principles is the "Principle of Availability," which states that "[a]n organization shall maintain information in a manner that ensures timely, accurate, and efficient retrieval."

The language and context of the IGPHC imply that the principles contemplate only a simple retrieval. But different organizations have different needs that may be more complex. Either way, a given organization's data characteristics will still impact their choices regarding tools and techniques.

The IGPHC expands on these concepts, but each one touches on a much larger discipline of knowledge that goes far beyond the 2014 publication. Whereas the principles appropriately identify the value of metadata and disaster recovery, as well as the challenges of conducting federated searches across multiple independently designed systems, its overall guidance speaks more to goals than specific techniques.

The actual techniques for measuring—and then improving—the timely, accurate, and efficient retrieval of information may be better addressed by referring to the science of information retrieval.

The Language of Information Retrieval

The science of information retrieval has developed a rich language of tools and techniques to facilitate the timely, accurate, and efficient retrieval of information. The terminology of information retrieval is a great place to find ways to implement "timely," "accurate," and "efficient" processes based on the IGPHC.

In information retrieval, accuracy is determined by measuring "recall" and "precision." These concepts can be confusing to the uninitiated, but, essentially, they represent the questions: Did you find ALL of what you were looking for (recall) and did you find ONLY what you were looking for (precision).

If you are searching for patients admitted after midnight and you identify 4 out of 5 of those patients, but also identify another 4 who were admitted before midnight, then your recall was 80 percent (you found most but not all) and your precision was 50 percent (you found most, but not only).

In practice, it is very possible to have 100 percent recall or 100 percent precision, but it is difficult to have both. This is where trade-offs occur. The need for timeliness and efficiency can affect recall and precision. Wanting results faster, or cheaper, can result in lowering the recall, lowering the precision, or lowering both.

Returning to our example of searching for patients admitted after midnight, you may be able to quickly identify many of the patients admitted after midnight, but to find both ALL and ONLY those patients may take more time or resources.

Using Search Options to Improve Performance

How should the right tool or mechanism for a record review be selected? The first key to selection is an understanding of the nature of the available ESI, how the ESI is maintained, and exactly what ESI is sought. The second key is leveraging information retrieval to measure and improve your performance. Together, these will allow organizations to make value-added choices of tools and techniques in this era of exponentially growing ESI.

ESI is now the “bread-and-butter” of health information management. ESI must be understood and retrieved to manage and improve healthcare. That understanding and retrieval are central to the role of the HIM professional today and into the future.

Dan Regard (dregard@idiscoverysolutions.com) is the CEO of iDiscovery Solutions, an electronic discovery consulting firm based in Washington, DC. He has national and international experience advising on such issues as electronic discovery, computer forensics, structured data, and information management. Ron Hedges (r_hedges@live.com) is a former US Magistrate Judge in the District of New Jersey and is currently a writer, lecturer, and consultant on topics related to electronic information.

Article citation:
Regard, Dan; Hedges, Ron. "Application of Search Analytics in the Healthcare Profession" *Journal of AHIMA* 86, no.5 (May 2015): 48-49.