EVERYONE KNOWS that lawyers like words. The legal vernacular sets the industry apart from other sectors and instills exactness in the language of law to consistently apply it to cases as they arise. It can confuse, however, when lawyers apply their love of language to technology, such as predictive coding, predictive ranking and computer or technology-assisted review.

Predictive coding started out simple enough. For document reviewers, it was easy to foresee predictive coding. We all found a document or a group of documents highly relevant to litigation or not, and looked for an easy method to find more documents like the one in hand and code them together. E-discovery vendors gave the legal industry an “easy button,” enabling the bulk coding of documents—and later automated the process with predictive coding.

Predictive coding is an automated review process using human expertise. An attorney or other subject-matter expert (SME) knowledgeable about a case codes a subset of documents, a seed set, as relevant or not relevant. The coded document is ingested into a computer with algorithms that apply the coding decisions across the corpus of documents. The computer suggests or predicts documents that may be relevant without a human reviewing every document in a linear fashion.

Predictive coding gave way to more descriptive words for the process, such as computer-assisted review and technology-assisted review (TAR). TAR appears to have outlasted other terms. Lawyers wanted to distance themselves from “predictive,” which gave the appearance of using a crystal ball to determine whether documents were relevant or not to litigation. That, and perhaps Recommind Inc., which coined the PC term and put it into a patent, adding apprehension to its use.

TAR IT’TIS

At base TAR can be viewed as a process that uses human knowledge and technology to find documents relevant to litigation or investigation. It works like many online app stores that accumulate your preferences over time and suggest items you may be interested in, such as books and music. The more you buy at such a store, the more the system gets to know you to suggest items relevant to your interest. This “training” time can be augmented and extended if you rate a purchase or give it thumbs up or down. The system takes your product or service review as relevance feedback and incorporates it into future suggestions.

Although the abstraction of TAR is simple, its application varies by e-discovery vendor and can be complex. Many approaches to TAR use different labels or names, such as Continuous Active Learning (CAL), Simple Passive Learning (SPL) and Simple Active Learning (SAL).
In an attempt to sort out various TAR methods, John Tredennick, founder and CEO at Catalyst Repository Systems Inc.; Mark Noel, managing director at Catalyst; Jeremy Pickens, Ph.D. and senior applied research scientist at Catalyst; and lawyer and legal journalist Robert Ambrogi wrote “TAR for Smart People: How Technology Assisted Review Works and Why It Matters for Legal Professionals.”

“TAR for Smart People” expounds CAL as instantiated in Catalyst Insight e-discovery, and compares it to SAL and SPL. SAL and SPL are first-generation TAR protocols. CAL is a second-generation protocol, popularly called TAR 2. This lays the first problem in sorting out TAR: labels. What’s better, TAR 2 or 1? Does continuous sound better than simple? Does active learning sound more, better than passive? Based on the language alone, a reader may decide that TAR 2 with continuous active learning is the best and let it rest.

TAR systems have common ground, said Tredennick. The processes are interactive. “A human reviews and tags a document as relevant or non-relevant. The computer takes the human input and uses it to draw inferences about other documents. Ultimately, the computer orders the documents by relevance to guide the review process. Humans then decide how many documents need to be reviewed.”

All TAR systems provide time and labor savings. “Review teams can work faster using prioritized (ordered) review because they are reviewing documents with similar content,” said Tredennick. It is easier to review a master service agreement and all its parts than it is to switch back and forth between calendar items, grocery lists and complex contracts.

All TAR systems save review costs because they provide “a reasonable basis to ‘cut off’ review once most of the relevant documents have been found,” said Tredennick. With TAR, you can remove large swaths of documents not relevant to the review and defend it by reasonable sampling techniques.

**TAR WORKINGS**

TAR works differently depending on the vendor and depending on what is being reviewed, e.g., relevance to litigation or internal investigation, or a search to remove privileged documents from a collection. From a high level, TAR systems have a similar workflow that includes: 1) Collect documents to review and ingest them into a computer system; 2) Review and code documents relevant or not and feed the coded documents back into the computer system for it to learn the terms in relevant and not relevant documents—the system develops a formula to predict relevance of other documents in the collection; 3) The system scores and orders documents by their relevance ranking—this step includes a quality assurance process to test the validity of the ranking so you can move forward to review, or drop back for more training; and 4) Analyze, batch and review documents in logical groups.

When you dig into the above workflow, differences among vendor offerings become apparent. In step 1, some systems prefer you to collect all the documents under review before it begins a predictive process. New documents added to an existing collection after a system begins ranking documents can jeopardize the stability of an overall ranking and mitigate the cost savings associated with TAR—additional sampling and training may be required with new documents.

In step 2, a religious debate ensues on how to best present training documents to the system. You can present documents that have a better chance of being relevant or not relevant using, e.g., keyword searches, which can inject bias in the selection; or you can let the computer select random samples.

Some systems, such as Insight, attempt to improve its understanding of the document population and recommend additional documents for training. Insight continuously learns from coding decisions and reranks the document population as document review ensues. This is “continuous, active” learning, said Tredennick to Legaltech News. “The more training and review you do, the better the system will be at ranking unseen documents.”

Other systems separate training from review, said Tredennick. SMEs handle the training and review until they are satisfied that the system is fully trained. Review teams then look at the higher ranking documents until “proportionality” sets in.

For any TAR system, completing a document review requires you to review documents until the concept of reasonableness and proportionality set in, which aim to save time and money. Federal Rule of Civil Procedure 26 states that discovery must “be proportional to what is at issue in the case” and “whether the burden or expense of the proposed discovery outweighs its likely benefit.”

If after reviewing a percentage of documents, the TAR system indicates you have reviewed the vast majority of predicted relevant documents, and to review more would be an undue burden and expense, review is complete. Whether reasonable, remains an open question for any TAR system, at any time.