

Predictive Coding: E-Discovery Game Changer?

By *Melissa Whittingham, Edward H. Rippey and Skye L. Perryman*



Predictive coding promises more efficient e-discovery reviews, with significant cost savings, to combat the inflationary growth of electronically stored information (ESI). Should you consider using it? This article discusses whether predictive coding is a game changer or simply another tool in the

e-discovery arsenal. We also note factors to consider when assessing whether predictive coding could be helpful for your next case.

What is predictive coding?

“Predictive coding” refers generally to the use of automation to manipulate ESI during any stage of e-discovery. The technology is also known by such other names as automated document review, automated document classification, automatic categorization, predictive categorization, and predictive ranking. Predictive coding software, often used in conjunction with traditional early case assessment technologies, presents the possibility to increase rates of document review — in turn dramatically reducing review costs.

In short, predictive coding “(i) start[s] with a set of data, derived or grouped in any number of variety of ways (e.g., through keyword or concept searching); (ii) use[s] a human-in-the-loop iterative strategy of manually coding a seed or sample set of documents for responsiveness and/or privilege; (iii) employ[s] machine learning software to categorize similar documents in the larger set of data; (iv) analyze[s] user annotations for purposes of quality control feedback and coding consistency.”¹ This process can be used to aid reviewers during a traditional manual review or to select and produce documents that have not been viewed by attorneys.

Who offers predictive coding?

E-discovery vendors currently fall into three camps with respect to predictive coding. The first camp readily admits that they do not offer predictive coding. The second camp offers predictive coding services as the term generally has come to be understood. These include Kroll OnTrack, Servient, and Recommind. Epiq Systems offers Equivio>Relevance that “uses statistical and self-learning techniques to calculate graduated relevance scores for each document in the data collection.”² Several vendors offer software platform Relativity coupled with Content Analyst, marketed as Relativity Analytics, that provides concept search technology.

¹ Jason R. Baron, *Law in the Age of Exabytes: Some Further Thoughts on ‘Information Inflation’ and Current Issues in E-Discovery Search*, 17 RICH. J.L. & TECH. 9, 32 (Spring 2011).

² Equivio, Equivio>Relevance, <http://www.equivio.com/product.asp?ID=7> (last visited Aug. 26, 2011).

In June of this year, Recommind announced that it was issued a patent on predictive coding³ (although it has thus far been unsuccessful in its effort to trademark the term “predictive coding”⁴). According to the company, the patent covers “systems and methods for iterative computer-assisted document analysis and review.”⁵ A Recommind representative has announced that Recommind will seek to license the patent to other companies that already offer their own versions of predictive coding or that want to have the ability.⁶ Some view the patent as a marketing tactic before a possible IPO.⁷ Others note that the patent is broad and could be challenged on the basis of both novelty and non-obviousness.⁸ At this point, it is unclear how the patent will affect the predictive coding market.

The third set of vendors claim to offer predictive coding services — but, it is unclear whether they have true predictive coding as the term has come to be understood. These vendors may use traditional early case assessment tools in an iterative manner that mimics predictive coding outcomes. For example, the vendor may use filtering, keyword, and date searches to reduce a document set. Manual review can then yield additional information to further hone the search terms and group documents. Several repetitions of this process may further winnow the document set. Although this can be a helpful search technique, it is not predictive coding in the traditional sense of technology providing individual coding calls.

Is Predictive coding a game changer?

Although linear document review — in which individuals manually review and code documents ordered by date, keyword, and the like — has long been the accepted standard, this approach has become costly and, in many cases, inefficient given the exponential increase in ESI.⁹ Further, recent studies demonstrate that manual review may not be the “gold standard” it was previously thought to be.¹⁰ Unlike individual reviewers, computers do not get headaches or get tired.¹¹ Search technology

³ See Press Release, Recommind, Recommind Patents Predictive Coding (Jun. 8, 2011), http://www.recommind.com/releases/20110608/recommind_patents_predictive_coding.

⁴ USPTO, Office Action About Applicant’s Trademark Application (May 18, 2011), <http://tdr.uspto.gov/jsp/DocumentViewPage.jsp?77899686/OOA20110518144912/Offc%20Action%20Outgoing/31/18-May-2011/sn/false#p=1> (“Registration is refused because the applied-for mark merely describes a characteristic, feature or function of applicant’s goods.”); Evan Koblentz, “Predictive Coding” is Not a Registered Trademark, EDD UPDATE, Jun. 10, 2011, <http://www.eddupdate.com/2011/06/predictive-coding-is-not-a-recommind-trademark.html>.

⁵ See Press Release, Recommind, Recommind Patents Predictive Coding (Jun. 8, 2011), http://www.recommind.com/releases/20110608/recommind_patents_predictive_coding; see also U.S. Patent No. 7,933,859 (filed May 25, 2010).

⁶ Evan Koblentz, *Recommind Intends to Flex Predictive Coding Muscles*, LAW.COM, Jun. 8, 2011, <http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202496430795&slreturn=1&hbxlogin=1>.

⁷ *Id.*

⁸ *Id.*; Christopher Danzig, *Predictive Coding Patented, E-Discovery World Gets Jealous*, ABOVE THE LAW BLOG, Jun. 9, 2011, <http://abovethelaw.com/2011/06/predictive-coding-patented-e-discovery-world-gets-jealous/>.

⁹ See Craig Carpenter, *Predictive Coding Explained*, RECOMMIND BLOG, Mar. 10, 2010, http://www.recommind.com/blogs/20100310/predictive_coding_explained.

¹⁰ See Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, 17 RICH. J.L. & TECH. 11 (2011), <http://jolt.richmond.edu/v17i3/article11.pdf>;

has become closer to approximating human reasoning.¹² Given that predictive coding technology has the potential to be the most proportionate way of managing a case, this is an important new technology that must be considered.

Should you consider using predictive coding?

We think that predictive coding may be a terrific tool for some cases — but, for others, may be less suitable than a traditional review. It depends on the circumstances. The list below offers factors to consider when assessing whether to use predictive coding.

1. Volume of Documents

Whether you are producing documents or processing a document dump, predictive coding generally works best with a large quantity of documents. The algorithms behind predictive coding generally need a significant amount of data to work properly. There is also a greater benefit to avoid manual review of all documents with a larger amount of data.

The same predictive coding techniques used to identify responsive documents may also be used to assess whether the documents are privileged. It remains to be seen whether predictive coding will work well with technical documents — but it is certainly possible that predictive coding could be used here as well.

2. Initial Work

Unlike traditional reviews where the bulk of work can occur during the middle or towards the end of the review, predictive coding requires the most work at the start of the review. Instead of a team of contract attorneys diving into the documents and passing their selections to more senior attorneys, predictive coding works best when experienced reviewers review documents until the system is sufficiently trained. Using the skills and knowledge of more senior lawyers to train properly the system will provide long-term benefits. Although this may take a few days of effort at the start, it should be followed by a long tail of cost savings.

3. Get What You Expect

As discussed, vendors offer a variety of services and may have different flavors of predictive coding. To ensure you are getting the services you expected, you should ask your vendor to take you through the proposed workflow and explain how it would work. This will guarantee that everyone — including the client— is on the same page for the review.

Herbert L. Roitblat, Anne Kershaw, & Patrick Oot, *Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review*, 61 J. AM. SOC'Y. FOR INFO. SCI. AND TECH. 1 (2010).

¹¹ See John Markoff, *Armies of Expensive Lawyers, Replaced by Cheaper Software*, N.Y. TIMES, Mar. 4, 2011, <http://www.nytimes.com/2011/03/05/science/05legal.html>.

¹² See IBM, Watson Research Center, <http://www.watson.ibm.com/index.shtml> (last visited Aug. 21, 2011).

4. Price

How predictive coding services will be priced in the future remains uncertain. As demonstrated by fees for de-duplication services, vendors tend to charge by the cost savings to the firm — not the effort it takes the vendor to provide the service. This means that even if use of predictive coding would cost significantly less than manual review, vendor fees eventually could reduce the amount of savings. In the short term, firms could reap the benefits. Established predictive coding vendors may be firm on pricing, but many vendors may be eager to break into the field and establish a reputation. If you shop around, you may be able to find a vendor willing to offer predictive coding at reduced rates or even for free on a trial basis.

5. Other Uses for Predictive Coding

Use of predictive coding software does not have to be an all or nothing commitment. Using predictive coding techniques could enhance a traditional review. Predictive coding software can rank responsive documents — meaning that a document could be, for example, 84% responsive. Documents can be ranked by responsiveness so that reviewers see more likely responsive documents first. This information could greatly assist manual reviewers when reviewing the documents and speed up the review. After a certain time, if reviewers find that they are agreeing with the computer, the computer assessments could be used for the remainder of the documents without having actual attorney review. Even if this step is not ultimately taken, predictive coding could aid reviewers in a traditional review.

6. Consider Safeguards.

Many lawyers (and their clients) are hesitant to be the “test case” for the use of predictive coding when there is no reported decision on this or a similar technology, especially when e-discovery sanctions are at an all-time high.¹³ As U.S. Magistrate Judge Andrew Peck recently noted at the Carmel Valley eDiscovery Retreat, none of the important cases on searches has actually endorsed the use of keywords.¹⁴ All of these cases involved criticism of the way that keywords had been used.¹⁵

To protect yourself in court when using predictive coding, cooperation could be key. Similar to discussing keyword search lists, it may be best to discuss use of predictive coding with the opposing side. Judge Peck advocates being up-front about use of the technology and to get the opponents’ agreement and buy-in.¹⁶ If agreement is not forthcoming, the court could be consulted to rule quickly

¹³ Dan H. Willoughby, Jr., Rose Hunter Jones, & Gregory R. Antine, *Sanctions for E-Discovery Violations: By the Numbers*, 60 DUKE L.J. 789 (2010), <http://www.abajournal.com/files/DukeLaw.pdf>; Ashby Jones, *Study: Lawyer Sanctions Over Electronic Discovery on the Rise*, WSJ LAW BLOG, Jan. 13, 2011, <http://blogs.wsj.com/law/2011/01/13/study-lawyer-sanctions-over-electronic-discovery-on-the-rise/>.

¹⁴ Chris Dale, *Judge Peck and Predictive Coding at the Carmel eDiscovery Retreat*, THE E-DISCLOSURE INFORMATION PROJECT, Aug. 2, 2011, <http://chrisdale.wordpress.com/2011/08/02/judge-peck-and-predictive-coding-at-the-carmel-ediscovery-retreat/>.

¹⁵ *Id.*

¹⁶ *Id.*

on the issue. Determining how much of the process to reveal will be established in each case, but some discussion of the technology and how it will be used should provide a measure of protection in court.

Conclusion

The growth of ESI (and the importance of these documents to litigations) make finding a manageable solution to reviewing such material an urgent need. Predictive coding is at the forefront of the response to this issue. Technologies differ from vendor to vendor — and the use of any product necessitates diligence from its users to ensure accurate and defensible document review. Use of predictive coding will likely grow more common in the near future.

Edward Rippey (erippey@cov.com) is a partner at Washington-based Covington & Burling. He handles complex commercial litigation, including patent cases, and is Chair of the E-Discovery Practice Group. Melissa Whittingham (mwhittingham@cov.com) is a litigation and food & drug associate at the firm, and has expertise in predictive coding and related e-discovery issues. Skye Perryman (sperryman@cov.com) is a litigation associate at the firm, and is a member of the e-discovery practice group.