Far From the Black Box: Explaining Equivio Relevance to Lawyers

White Paper

by
Chris Dale of the e-Disclosure Information Project

e-Disclosure Information Project

EDIP

equivio
zoom in. find out

This paper is written by Chris Dale of the UK-based eDisclosure Information Project[1] in conjunction with Equivio[2]. The eDisclosure Information Project brings objective and informed comment to lawyers, judges, suppliers and clients aimed at encouraging the better use of technology in electronic disclosure for litigation. Equivio is a software company whose focus is on the defensible reduction of data volumes for use in litigation, regulatory investigations and analogous circumstances. Its applications cover near-duplicate detection, e-mail threading and a predictive coding application called Relevance[3]. These applications are available as separate components for inclusion in applications and processes built by others and, since January 2012, in a single end-user application called Zoom[4].

PURPOSE OF THIS PAPER

Equivio Relevance is a market-leading example of software whose generic function is known variously as predictive coding, technology-assisted review, computer-assisted review and other names indicative of the joint application of human skill and computing power to document review. Put as broadly as possible, software of this kind takes input from senior lawyers or subject-matter experts on a subset of documents and uses their coding decisions (most obviously as to whether documents are relevant or not relevant) to make provisional decisions about the rest of the document population. The technology generalises from that human input across the wider population, and the output will be either categorisation as responsive or not, or a score which reflects the likelihood that documents will be responsive.

That broad description leaves some lawyers nervous about reliance on what is easily thought of as "black box" technology relative to the tested (as they see it) approach which relies on human eyes both for the primary selection of potentially-relevant documents and for final review. This nervousness has multiple root causes, including a misplaced faith in the alleged "gold standard" of keyword-driven human review, misunderstandings about the process and its purpose, lack of awareness of the quality control functionality which is included in applications like Equivio Relevance, and an assumption that all these applications are the same.

Above all, particularly in the US, lawyers are nervous of the use of technology which has not been endorsed by a court, notwithstanding that no US court has ever specifically endorsed the use of any particular technology. That hurdle has now been cleared by a written Opinion of US Magistrate Judge Andrew Peck in in the matter of Da Silva Moore v. Publicis Groupe & MSL Group, No. 11 Civ. 1279 (ALC) (AJP) (S.D.N.Y) which expressly approved the use of computer-assisted review.

---

[1] http://edisclosureinformation.co.uk/edisclosureproject.htm
[2] http://www.equivio.com/
[3] http://www.equivio.com/product.php?ID=38
[4] http://www.equivio.com/product.php?ID=35

That increases the focus on the other factors. This paper is aimed at those within law firms and their clients who may be more willing than previously to consider the use of predictive coding software in the light of Judge Peck's Opinion, but who need reassurance on the other points mentioned above. Its purpose is to encourage lawyers and their clients to understand that the use of Equivio's predictive coding technology as part of a properly managed process (something emphasised by Judge Peck) is a proper and defensible way to conduct cost-effective discovery in any jurisdiction.

WHAT THIS PAPER IS NOT ABOUT

This paper takes it for granted that the reader is familiar with the by now well-established fact that manual review by expensive lawyers is both inefficient and expensive. It does not purport to consider the circumstances in which sanctions may be suffered by those who give defective disclosure, nor does it analyse the specific requirements of any particular jurisdiction. Apart from one quotation, it does not consider the detail of Judge Peck's opinion in Da Silva Moore.

It avoids also any detailed statistical analysis of what underlies the predictive coding process, and includes no formulae or other mathematical expressions. These things are important, and reference is made to other sources[5] for those who want to understand them. This paper concentrates more on helping a non-technical lawyer to understand, at a high level, how a document review is undertaken using Equivio Relevance and what safeguards and checks are built into the process to ensure that the job is done properly.

WHAT JUDGE PECK ACTUALLY SAID

The opinion is worth reading in full[6], both for its explanation of what is expected from parties and for its summary of the iterative processes which may be required to arrive at a seed set to train it and to test the results. The central explanation, and the primary significance of the Opinion, lies in this passage:

> This Opinion appears to be the first in which a Court has approved of the use of computer-assisted review. That does not mean computer-assisted review must be used in all cases, or that the exact ESI protocol approved here will be appropriate in all future cases that utilize computer-assisted review. Nor does this Opinion endorse any vendor (the Court was very careful not to mention the names of the parties' vendors in the body of this Opinion, although it is revealed in the attached ESI Protocol), nor any particular computer-assisted review tool. What

---

[5] See Wikipedia article *Precision and recall* for an explanation of these terms
http://en.wikipedia.org/wiki/Precision_and_recall.
[6] Rob Robinson has helpfully collected all the formal court documents in the Da Silva Moore case into a single indexed .PDF on the Complex Discovery web site. The link and index are at
http://www.complexdiscovery.com/info/2012/03/02/peck-parties-and-predictive-coding/

the Bar should take away from this Opinion is that computer-assisted review is an available tool and should be seriously considered for use in large-data-volume cases where it may save the producing party (or both parties) significant amounts of legal fees in document review. Counsel no longer have to worry about being the "first" or "guinea pig" for judicial acceptance of computer-assisted review. As with keywords or any other technological solution to ediscovery, counsel must design an appropriate process, including use of available technology, with appropriate quality control testing, to review and produce relevant ESI while adhering to Rule 1 and Rule 26(b)(2) proportionality. Computer-assisted review now can be considered judicially-approved for use in appropriate cases.

OBJECTIONS TO THE PROTOCOL

A protocol was annexed to Judge Peck's Opinion describing the process which was to be followed. It is important to emphasise that the parties had agreed on the use of predictive coding but took different views as to the protocol to be adopted. The plaintiffs objected to parts of the protocol and, specifically, to what they considered to be inadequate safeguards in circumstances where the defendants held all the documents. The detail of those objections can be found in the Plaintiff's Rule 72(a) Objection to the Magistrate's February 8, 2012 Discovery Rulings and in the supporting statement of their expert.

It is irrelevant for the purposes of this paper what the final outcome is of this disagreement. The protocol and the objections to it are both detailed and represent extreme positions; they are referred to here to illustrate the issues which can arise between parties as the context for the explanation which follows in this paper.

The question to be addressed in this paper is not whether the Da Silva Moore Objections are well-founded on the facts of this case, but what steps and processes are available to lawyers using Equivio Relevance who want to be satisfied that they are doing a good job as a preliminary to persuading others – their partners, clients, opponents and courts - of that.

WHAT IS A "GOOD JOB" IN EDISCOVERY TERMS?

Before describing how one approaches an eDiscovery task, it is helpful to define what that task is. In any relevant jurisdiction, and in regulatory investigations as well as litigation, it is the identification of "relevant" documents (however relevance may be defined), and the removal of privileged ones. In the US in particular there is a gulf between the perceived aim and the methods used to achieve it: the perceived aim is that counsel can certify that discovery is "complete" and "correct" with the fear of sanctions as the punishment for shortcomings; the traditional methods depend on keywords, Boolean searches and manual review. The reality is first that this perceived target is not

the right one – as Judge Peck pointed out, the requirement is one of "reasonableness" - and second that the conventional approach using keywords and manual review is not merely time-consuming and expensive but extremely unreliable[7]. Sanctions are indeed a threat, but not for the failure of a good-faith and cooperative attempt to do the job properly.

The fact that many lawyers have been aiming at an unreasonably high target and with inadequate tools is not an excuse for lowering standards, nor does it make it safe or defensible to delegate decision-making to some kind of "black box" technology. The use of predictive coding implies neither of these things, subject to two provisos – choosing the right technology and adopting a suitable process. The key lies in Judge Peck's words:

> *As with keywords or any other technological solution to ediscovery, counsel must design an appropriate process, including use of available technology, with appropriate quality control testing, to review and produce relevant ESI while adhering to Rule 1 and Rule 26(b)(2) proportionality.*

TECHNOLOGY, PROCESS AND PEOPLE

The application of technology to business needs involves more than the bundle of interfaces and algorithms which comprise the lay perception of the "black box". Products like Equivio Relevance certainly use the best available technology, but the raw components are valueless without a set of workflows and processes built around them. Some of these are inherent in the technology package; others can be customised using tools provided with the technology; yet others, and perhaps the most important components in defensibility terms, are those devised by the users – what Judge Peck refers to when he says that "Counsel must design an appropriate process". This in turn emphasises the importance of the people involved. The New York Times article of March 2011 Armies of Expensive Lawyers, Replaced by Cheaper Software[8] promoted the erroneous idea that technology would do it all at the expense of lawyer jobs; the reality is that legal skills are vital.

Three groups of concepts are described here – some definitions, the stages in the process, and the context in which the tools are used. Taking each of these in turn:

- TERMINOLOGY

What do we mean by Recall and Precision and how do these computer science terms relate to the legal task and the client's objective?

---

[7] See Conor Crowley *Judge Peck Endorses Predictive Coding* http://www.equivio.com/files/files/Crowley%20Memo%20re%20Peck%20Decision.pdf for a succinct summary of Judge Peck's analysis of the true aim of discovery, including the footnote references to the sources of as to the defects of keywords and manual review.
[8] http://www.nytimes.com/2011/03/05/science/05legal.html?pagewanted=all

- THE STAGES IN THE PROCESS

  There are four stages in a discovery exercise involving Equivio Relevance - Assessment, Training, Decision and Verification. They are described here in terms sufficient only to contradict the "black box" idea.

- CONTEXT

  Within the broad context of the determination of relevance leading to a discovery production which complies with Rule 1 lie crucial procedural components such as the meet and confer, the obligation of cooperation of which the meet and confer is part, and the protocol arrived at as a result.

This paper takes each of these in turn.

TERMINOLOGY

However non-technical the aim of this paper, one cannot avoid using certain terms of art when trying to relate the everyday experience of lawyers to what is available from the technology. The terms which matter most are Recall and Precision. In lay terms, stripped of the formulae properly required by statisticians, these have straightforward meanings which one can relate to any kind of search exercise.

- **Recall** means: of the relevant documents in the collection, how many did we find?
- **Precision** means: of the documents thought to be relevant, how many are in fact relevant?

 To understand how these terms matter, one needs two more elements. These are:

- How much **time and budget** have you got? and
- How willing are you to **risk** overlooking some relevant documents?

Although the terms Recall and Precision derive from computer science, they can equally be applied to a pre-computer discovery exercise involving only paper. A team of lawyers might go to a warehouse stacked with boxes of documents looking for potentially relevant ones. By opening every box and reading through every file they might achieve a high level of Recall. The highest possible level of Recall, however, could only be achieved by a combination of a) a wide definition of "relevance", b) a very big team and/or all the time in the world and c) consistency of quality, both as between the searchers and across the whole period of the search, Monday to Friday, morning till night.

The high Recall, however, will have two downsides – first, the wide definition of relevance, erring on the side of inclusion, will necessarily sweep up many documents which are at best of marginal relevance and, second, the time and cost would almost certainly be disproportionate to the objective.

The trade-off for high Recall is likely to be low Precision – that generous definition of "relevance" will bring back many documents which, when subjected to the closer examination of a review exercise, are not relevant to at all.   To introduce another term of art, the collection will include a high proportion of "false positives", and the review stage, expensive enough as it is, will be made much more so by the need to wade through a high proportion of useless documents.

Have you done a good job thus far with this safe-looking but expensive exercise? Two points arise here: the requisite input depends, in part, on how important the objective is, and few discovery exercises warrant infinite budgets, whatever view one takes on either the formal duties of disclosure or the chances of finding the "killer document"; equally important is the fact that you really have no idea how good a job you have done.

In this respect, a "paper in the warehouse" exercise and a "key words plus manual review" equivalent are much the same. The only safe way to check the validity of what you have done is to redo it, perhaps with a different team, and compare the results.

One could refine this explanation of the shortcomings of traditional approaches by reference to the detailed comparisons in, and the conclusions of, the studies referred to by Conor Crowley in the paper mentioned above. Our purpose here is at a higher, less granular level: all discovery exercises involve a trade-off between cost and risk – the expense of over-inclusion against the risk of omission – and both cost and risk increase with volume while the average value of claims remains constant; how do you use tools like Equivio Relevance to strike a balance between these competing pressures?

Predictive coding brings the labels Recall and Precision to something which has always existed. What it brings in addition is the ability to validate the outcome and, if necessary, to adjust both these elements on the fly. The equivalent of the trips back to the warehouse, the cross-checks and the re-scoping of the review exercise can be done instantly. The lowest target of a predictive coding exercise is to do no worse a job at no higher cost than the equivalent traditional exercise. The ideal (and it is an achievable ideal with the proper use of this kind of technology) is to do a better job and at a lower cost and, crucially, to be able to demonstrate what you have done. It is NOT (as Judge Peck emphasised) to achieve some notional level of perfection.

THE STAGES IN THE PROCESS

It was suggested above that there are four stages in a discovery exercise involving Equivio Relevance - Assessment, Training, Decision and Verification.

Together they are a set of processes far removed from the common perception of a "black box" with its implication that documents are tipped in at one end and disgorged ready-coded at the other. Such an approach to Discovery would be rightly condemned. That is not what Equivio is used for.

The aim of the assessment stage is to arrive at a pool of documents which serve as the "gold standard" against which the subsequent exercise is tested and measured. Control documents are documents whose relevance or non-relevance is known – known because an expert (a senior lawyer or a subject-matter expert) has reviewed them personally and tagged them as relevant or not relevant. A standard control set consists of 500 documents, though more may be required from collections of low richness – that is, where the relevant documents are a small proportion of the whole.

This stage is crucial and depends as much on the quality of the human input as on the ability of the system to determine when it has enough information to act as a statistically valid control sample.

The system then builds a custom classifier based on the relevance decisions which have been made. Although the ultimate objective may appear to be a binary relevant / not relevant decision, the system is capable of measuring degrees of relevance on a scale from zero to 100, leaving it to the lawyers to decide in due course where to set the threshold. This is called "prioritisation" and has multiple benefits: one is that it allows the lawyers to see first those documents which are (provisionally) deemed most relevant; another is that documents can be allocated to lawyers of appropriate seniority (and therefore cost) based on their provisional ranking; a third is that this ability to choose a threshold at a later stage in the process allows proportionality to be injected into the decision-making because it is easy to see how many documents lie above and below the cut-off point and therefore what the cost or saving is of making a different decision.

TRAINING FOR AS LONG AS IT TAKES – AND NO MORE

The previous sections showed the creation of a control set of documents whose status is known. They also identified two concepts - Recall (the probability that a randomly selected relevant document is retrieved in a search) and Precision (the probability that a randomly selected retrieved document is relevant). Cost (how much time and budget have you got?) and risk (how willing are you to risk overlooking the some relevant documents?) are factors which should apply to any eDiscovery exercise. All these factors are inter-related – high Recall reduces your risk, but it also increases cost unless Precision is also high. These two factors are often combined into a single measure known as the F-measure. The Equivio training process is designed to optimize the F-measure, and in so doing obtain the best possible results for Recall and Precision.

The object of the tools in Equivio is to help the lawyers strike a balance between cost and risk, allowing them to make an informed set of provisional decisions. Without hammering the point too much, it is this which distinguishes an Equivio predictive coding exercise from the so-called "black box" approach. It is the lawyers who decide that the control set is representative; the system does not make coding decisions but offers a transparent view of provisional decisions for the lawyers to accept or adjust; as we shall

now see, it is the lawyers who decide, with the help of Equivio's tools, how many training iterations are required.

"Training" involves several rounds in which lawyers make a series of document decisions in phases which "teach" the system what the characteristics are of a relevant and of a not relevant document. This is best understood by reference to other environments in which decision-making skills are passed on by an expert. In a manual review, the team is instructed in its decision-making by being given examples (perhaps in a coding manual) of what is and what is not relevant. Spam filters "learn" from user input that e-mails with certain characteristics are to be barred whilst others are let through. The key here is that accumulation of experience can be used to inform generalised decision-making. The use of words like "train" and "learn" does not impute anthropomorphism to the software; its "learning" consists of analysing each selected document, and the relevance decision made about it, to identify the multiple characteristics which define a document of that kind. It is the multiplicity of characteristics which enable ranking to take place.

How does one make the decision as to when to stop training? This is a cost versus risk point - how much expensive lawyer time do you need to put in to be reasonably sure that the system is adequately informed as to the characteristics of a relevant and a non-relevant document? In any such exercise, there comes a point when further training adds nothing to the system's understanding. This is known in Equivio Relevance as the "stabilisation point", illustrated in the picture below. It is important to appreciate first that this is a human decision - the stabilisation point is a suggestion, not an enforced stopping-point - and second that the decision is not a final one because further training can be given if subsequent quality assessment tests, or valid objections from opponents, suggest that this is necessary.

Finding the Stabilisation Point

One obvious situation in which testing may have to be resumed is when rolling loads arrive, perhaps because a new source is found, or because a staged approach was always planned. It is not good enough, in Equivio's terms, simply to assume that the

new documents have the same range of characteristics as those already in the system. New arrivals are tested against the classifier and the system recommends if further training is required.

Support for lawyer decision-making in Equivio Relevance

This theme – that the system provides transparent views to enable the lawyers to make decisions – continues to the end of the training process and through into the subsequent decision-making and quality assessment. Examples include warnings if different tagging is given to two very similar documents, or to different documents within the same e-mail thread, and in other examples of inconsistency.

In addition, the system deploys an "active learning" approach designed to minimise the risks implicit in a human sampling exercise. This is best illustrated by going back to the human-in-a-warehouse example. A truly random, bran-tub, approach to sampling the files and boxes stands as good a chance of missing as it does of hitting a truly representative selection; equally, an approach which is led by preconceived ideas of significance runs the risk of the self-fulfilling prophecy – the human may indeed find what he or she was expecting to find, but may miss other areas of significance.

There comes a point in any eDiscovery exercise where a decision must be made about which documents to send for review. This is where the balance between cost and risk appears most sharply; every document sent for review increases time and cost, whilst the exclusion of documents brings the risk of omitting something important.

The lawyer wants to know the ratio between reviewed documents and relevant documents, where the ideal is the smallest reviewed percentage which will catch the highest percentage of relevant documents. Equivio Relevance visualizes this decision-making process, allowing the user to choose between alternatives: for example, review 16% of the document population to yield 73% of the relevant documents, or review 20% to yield 82% of the relevant documents. The decision-making lawyer can analyze the effect of reviewing more or fewer documents. Reviewing more documents will reduce risk (by increasing Recall) but increase cost. That cost is not simply a straight-line per-document cost because the distribution of relevant documents will diminish as relevance rankings reduce; in human review terms, the gaps between relevant documents will increase and the per-document relevance will diminish, and the cost per relevant document goes up as their distribution thins out.

The Equivio system also calculates costs, allowing users to make the serious decisions that need to be made. Importantly, these decisions are made by a lawyer; the lawyer armed with the cost and risk information provided by the Relevance application is making decisions on an informed basis which can be shared with opposing counsel and, if necessary, the court.

VERIFYING THE RESULTS

Equivio Relevance provides two key quality assurance tools, *Test the Rest* and *Test the Method*. The "Rest" is everything below the cut-off point, that is, those which, subject to this verification round, are to be excluded from review. Shorn of its technical detail, *Test the Rest* generates a random sample of documents from the documents provisionally marked to be excluded with the sample's size dictated by the level of statistical confidence set by the user. A lawyer reviews the sample documents and marks them as relevant or not. This information is used to make a calculation of Precision in the Rest zone. The outcome allows the user either to confirm or to modify the selected cut-off point.

Test the Method allows comparison between the provisional conclusions made by Equivio Relevance and some alternative culling or review method such as manual review or keyword searching. It cross-matches between the two results, drawing attention to discrepancies between the outcomes of the two approaches, and allowing a decision to be made between them.

This summary of the QA tools in Equivio Relevance is deliberately brief. The facilities exist to cross-check and validate both parts of the collection – the Review set and the Rest - to whatever extent and with whatever amount of statistical validation is required by that circumstance and justified by the same balance of cost and risk as applied earlier in the process.

One point which is easily overlooked is that the time saved by the use of predictive coding technology will, relative to other methods, leave time in hand for QA and validation.

THE TECHNOLOGY PROCESSES IN THE PROCEDURAL CONTEXT

Any methods which a party intends to use to arrive at the production set should be the subject of discussion between the parties both before the processes are initiated and as they take place. This is not just because the rules surrounding the FRCP meet and confer process (and its equivalent in other jurisdictions) require such cooperation, but because cooperation tends to mutual benefit. This is forcefully argued in a paper by US Magistrate Judge David Waxse called *Cooperation – What Is It and Why Do It?* published by the Richmond Journal of Law and Technology[9]. Judge Waxse also considers the

---

[9] David J. Waxse, *Cooperation—What Is It and Why Do It?,* XVIII RICH. J. L. & TECH. 8 (2012), http://jolt.richmond.edu/v18i3/article8.pdf

question "how does the idea of cooperation exist in an adversarial system where each lawyer has the duty of zealous advocacy?" and concludes (citing the Sedona conference Cooporation Guidance for Litigators and In-House Counsel) that the principles of co-operative discovery outweigh any ethical duty of zealous advocacy which conflicts with them. Judge Waxse goes on to cite the relevant rules imposing the obligation to cooperate.

Even if we are content to leave jurists to argue about the scope of the duty to cooperate, the most compelling motive for cooperation is to save costs, principally by ensuring that the process does not have to be redone because of post-production arguments on matters which should have been discussed earlier. Enough has been said in this paper to make it clear that Equivio Relevance offers frequent and continuing opportunities for decision-making at which the choices and their effects are entirely transparent.

That in turn allows discussion on a mutually-informed basis whose end-purpose is a protocol setting out what is expected of the party giving discovery. That might include, for example, mutual input into the make-up of the control set, as well as provision for future input into stages not yet reached – the number of training iterations and the placing of the cut-off point. That does not, of course, guarantee agreement on all or any of the decisions, but it does mean that the discussions take place on the basis of transparent shared information at least as to some of the factors which may subsequently have to be argued in front of the judge.

SUMMARY

This paper has deliberately avoided both the detailed arguments in the Da Silva Moore case and the technical and scientific elements which underlie a fully-informed use of Equivio Relevance. Its primary purpose is to address the fear felt by many lawyers and their clients to the effect that the use of predictive coding technology deprives them of control or, indeed, a role in the decision-making.

The reality is quite the opposite. The tools described above bring transparency and the basis for informed decision-making throughout the process, increasing rather than diminishing the role of the skilled lawyer in making decisions about the scope of discovery. Furthermore, those same tools allow cooperative and collaborative discussion both about what is being done and about the costs and risks of the alternative decisions which might be made.

Transparency does not end with the election to set the cut-off point in one place rather than another. Equivio Relevance is capable not only of absorbing and taking account of fresh information – rolling loads or revised views about issues etc – but also of cross-checking results of both those provisionally identified as to be produced and "the Rest".

If anything resembles a "black box" approach to eDiscovery, it is the use of methods which allow no sight of the potential end result as the process takes place, and which allow no quantifiable means of checking what has been done. Equivio Relevance, when coupled with the sort of process envisaged by Judge Peck and managed by the right people, provides transparency, predictability and the verification tools to enable the lawyer to answer the question posed at the beginning of this paper – "Am I doing a good job?"

The word "provisionally" has recurred throughout this paper. It emphasizes a point which is critical to allaying the fears of some potential users of tools like Equivio Relevance. The application's function is not to make decisions nor to send out documents which have not been reviewed by lawyers. Its purpose is to support lawyer decisions, at the beginning by learning from their input and magnifying that across volumes which cannot be managed by traditional methods, and at the end by offering transparent validation. It allows lawyers to get back to being lawyers, on top of the issues, the facts and the documents and freed from manual drudgery to focus on the tactics and the strategy which will meet the objective of the clients.

ABOUT THE eDISCLOSURE INFORMATION PROJECT

The e-Disclosure Information Project is run by Chris Dale, a former commercial litigation partner turned e-Disclosure consultant. The Project aims to bring together lawyers, suppliers, courts and corporations with an interest in electronic disclosure, and to disseminate information about the court rules, the problems and the software and services available to handle them.

http://chrisdale.wordpress.com
chrisdaleoxford@gmail.com
www.edisclosureinformation.co.uk

ABOUT EQUIVIO

Equivio develops text analysis software. It's widely used in e-discovery. Users include the DoJ, the FTC, KPMG, Deloitte, plus hundreds of law firms and corporations. Equivio offers Zoom, an integrated web platform for analytics and predictive coding. Zoom organizes collections of documents in meaningful ways. So you can zoom right in and find out what's interesting, notable and unique. Request a demo at info@equivio.com or visit us at www.equivio.com.

Zoom in. Find out.