

Near-Duplicates: The Elephant In The Document Review Room

The Editor interviews Warwick Sharp, Vice-President – Marketing and Business Development, Equivio Ltd.

Editor: De-duping has been around forever. So what's new about Equivio?

Sharp: You're right. There's nothing new about de-duping. De-duping identifies exact duplicates. We're focused on near-duplicates. Near-duplicates are documents which differ by a few words or paragraphs. Near-duplicates are well camouflaged in litigation processes. They're probably the elephant in the document review room.

Editor: Why should corporate counsel be concerned about near-duplicates?

Sharp: Litigation costs are spiraling, and near-duplicates are a key part of that. It turns out that there are a lot of near-duplicates, much more than most people would expect. In the 150 cases that we've handled over the last year, near-duplicates have accounted for 20 to 30 percent of the documents to be reviewed. We recently had a case with 45 percent near-duplicates. This is on top of the exact duplicates. So, near-duplicates represent a huge hidden cost in document review.

Editor: Why do you say that near-duplicates are a hidden cost?

Sharp: This is a classic chicken-and-the-egg scenario. Without the software, you have no idea how much pain near-duplicates are causing you. This is the reason a lot of firms like to run a sample of their data through Equivio before deciding whether to purchase the service. We're fine with this. It's a new technology, so people need to feel comfortable. From our experience, once the customer sees how many near-duplicates they have, they will always decide to use the service.

Editor: So, how do you actually reduce the cost of review?

Sharp: In a nutshell, Equivio enables set-centric review, as opposed to traditional document-centric review. First of all, we run all the documents through the Equivio software. This groups the corpus into sets of very similar documents. Let's say we found a group of 10 near-duplicates, all versions of a 50-page contract. This set can then be assigned to one lawyer to review. So, he or she can review all these very similar documents together in a systematic fashion.

Editor: OK. You're reviewing sets of near-duplicates. How does this actually cut costs?

Sharp: Firstly, your review process is much more coherent and organized. Secondly, within each set of near-duplicates, our software suggests a "pivot" document. This is the document the lawyer should read first. It's the most representative document of the set. So you can prioritize the initial review process, reading just the pivots to cover the entire collection. If the pivot is clearly irrelevant to the matter, you can skip the other documents in the near-duplicate set. After all,



Warwick Sharp

they differ by just a few words. Obviously it's a question of common sense. For example, we recently had a case which had one set of over 30,000 near-duplicates, all of which were training forms. They obviously had no bearing on the case. So, the lawyer read the pivot, and then could skip the other 29,999 members in the set.

Editor: With bulk handling of groups of documents, isn't there a risk of skipping critical information? One of those near-duplicates might have one word which is different, which could be crucial to the entire case.

Sharp: Precisely the opposite. Equivio actually reduces the risk of inadvertently missing important data. Going back to our set of ten 50-page contracts – we read the pivot. We decide that the pivot is relevant, or that similar documents may be pertinent to the case. Therefore, we need to review all the other documents in the near-duplicate set. However, we wouldn't need to read the 50-page contract ten times. Once we've read the pivot, we can simply use a redline tool which will highlight the differences. We'll compare each document to the pivot, zooming in on the words which have been added or deleted.

We might have a version of the contract to which just two words were added. We can immediately focus on these two words. Obviously, it's a lot quicker reading two words than 50 pages. More importantly, you're avoiding the dreaded "glazed eyes effect." By zooming in on the differences, you pinpoint each item of unique information in each document. Bottom line: you've saved a lot of time in the review process. But, perhaps even more importantly, you've reduced the risk of missing data which might be crucial to your case.

Editor: Who would you say benefits from Equivio's near-duplicate capability in the litigation value chain?

Sharp: From our discussions with both corporate counsel and law firms, it seems that this is really win-win capability. The corporation benefits because it slashes litigation review costs. In the cases we've worked on, Equivio has reduced litigation review costs by 20 to 40 percent. These are very significant savings on direct out-of-pocket expenses.

From the law firm's point of view, they are able to provide better value for money to their corporate clients. This is obviously important in competitive situations, or where you are trying to build client loyalty. Another important consideration for the law firm is the time factor. Time saved in review is crucial when you are facing tight discovery deadlines. And, that's most of the time.

We have also found review quality to be an important consideration, both for law firms and corporate counsel. The near-duplicate groupings enable the consistent treatment of documents. The lawyer will be tagging documents as responsive, privileged, hot and so on. If, for example, the pivot document is deemed to be privileged, the Equivio groupings allow the reviewer to tag all the documents in that near-duplicate set as privileged. This is an important quality assurance function.

Editor: There are many "compare" or "redline" tools on the market. How does Equivio's solution differ from these tools?

Sharp: You're right. There are many redline tools around. In fact, we also provide a tool like this as an add-on module. Redline tools show the differences between document A and document B. This is straightforward. Once you know that document A is fairly similar to document B, you can use these tools to display the differences. But, first of all, you have to know that they are in fact similar. There's no point comparing totally different documents. This is the problem we solve. Basically, our software is telling you that documents A and B are very similar, and it will make sense to compare them with a redline tool.

Editor: How receptive is the market to Equivio's new product?

Sharp: The market's response speaks for itself. The product was introduced at LegalTech New York in January 2005. We are already working with over 20 service providers, including most of the top ten providers in the discovery arena. People seem to recognize that we have hit on a solution for a very painful problem.

Editor: Could you give an example of a case where Equivio's solution was used.

Sharp: We recently worked on a regulatory matter that entailed more than 10 million documents. The case included paper documents that had been OCR'd, as well as electronic documents and emails. We found almost 30 percent near-duplicates. Using our tool, the law firm was able to reduce its review costs by over \$5 million. This is a very significant saving.

Editor: People are used to working with the popular review repositories like CT Summation, Concordance, iCONNECT, Ringtail, Introspect and so on. Do they have to use something else if they work with Equivio's solution?

Sharp: Definitely not. Equivio's solution

is fully integrated with the standard review repositories. In addition to the standard grid, each document gets an Equivio number. This Equivio number – we call it the EquiSet – is the near-duplicate group to which the document belongs. This way, users can assign, sort, search and review documents by the near-duplicate groupings, all in the familiar surroundings of their standard review platform.

Editor: How much training is required to use Equivio's solution?

Sharp: Training is minimal. Since they're using their standard review tool, users can be up and running in 15 minutes.

Editor: What kind of feedback have you been getting from lawyers?

Sharp: The feedback has been excellent. Near-duplicates are actually a very intuitive concept. It's very concrete and objective. People like the fact that they can visually see the differences between documents. This is very objective and gives them confidence to use the tool.

They also like the fact that you don't have any surprises in a near-duplicate set. There's nothing counter-intuitive about a near-duplicate. In fact, we can guarantee no false positives. A false positive would mean that document A is totally different from document B, but it was identified, falsely, as a near-duplicate. There is obviously zero tolerance for false positives in a litigation environment. It would be a recipe for disaster. Our ability to ensure zero false positives is very important for litigators.

Editor: Is this a service that you offer?

Sharp: We are a software vendor. We sell our software to discovery vendors, and they offer the near-duplicate grouping capability as a service within their overall service package. Essentially, the service providers are delivering a near-duping service, based on the Equivio software, to the corporate legal departments and law firms. We also work directly with law firms and corporations, especially where they are doing a lot of their discovery processing in-house.

Editor: Who has been pushing for the introduction of near-duping?

Sharp: Corporate counsels have been leading the way here – particularly in addressing the near-duplicate issue in pretrial conferences and also by demanding the near-duplicate service from their law firm and discovery vendors.

Editor: One last question – Equivio has very quickly established an exceptionally strong presence in the litigation industry. How do you explain the speed that this has happened?

Sharp: It's about bottom line. Corporations are pushing for near-duping because it can take a huge chunk off their litigation costs.

Please email the interviewee at warwick.sharp@equivio.com with questions about this interview.