





## Near-Duplicate Detection Capability [continued]

Further analysis can be done using a text comparison tool to only review differences among near-duplicates as defined by the client's criteria. This level of quality control eliminates false positives which would hinder, rather than assist, the review process. Near-duplicates are identified as such in a separate database field in the document's record. "Like" documents can be subsequently assigned to the same reviewer for efficient and consistent treatment.

### **Electronically Stored Information (ESI)**

Identifying and eliminating true duplicates is common in processing ESI, whether it is within and/or across custodians. Using technology such as MD5 hash "digital fingerprinting," algorithms analyze both the content and format of digital files and identify identical duplicates through an automated process. Until recently, near-identical duplicates were left uncovered unless by chance a reviewer recognized a previous document or the use of sophisticated search strategies was employed. TechLaw Solutions offers a resolution to the near-duplicate detection dilemma that is based on our 25 years of providing litigation support services. The solution incorporates our extensive technical and project management skills coupled with technology that assists in automating the near-duplicate identification process.

TechLaw Solution's near-deduplication detection capability, which uses patent-pending technology powered by Equivio, works equally well with common proprietary formats like Microsoft Office and open-source or plain-text files. For ESI, detection is typically performed after the true duplicates have been identified and eliminated from further processing.

### **Hard Copy**

Near-duplicate detection is most commonly thought of as being applied against electronic files. It can also be used with hard copy materials that have been digitized and converted to full-text, usually through OCR processing. Such groups may include actual duplicates and near duplicates. Near-duplicates identified from hard copy or from electronically stored files can be reviewed simultaneously providing enhanced consistency in considering "like" duplicates over an entire collection of data.

## **What are the Benefits of Using TechLaw Solutions' Near-Duplicate Detection?**

Near-duplicate identification gives the reviewer the power to simultaneously review significantly similar documents. Along with the time and cost savings in organizing and reviewing the collection by "like" items, the ambiguity and risk of different reviewers assigning inconsistent review decisions to similar documents is considerably decreased. The return on investment can be substantial. Efficient discovery document management requires fast, accurate and consistent treatment of documents. Designing a document review protocol that will deliver the desired results is a major challenge in cases of all sizes and is particularly difficult in large-volume cases. TechLaw Solutions' near-duplicate detection capability augments our routine deduping during the preprocessing of electronic files, can be quickly implemented and extends your team's power to gain mastery of their discovery materials.

