More efficient document production
LexisNexis® Near Dupe

Increasing control over e-discovery is critical to your success. LexisNexis® supports you in this effort by providing powerful, flexible solutions for early data assessment, processing, review, and seamless chain of custody.

For example:

- **LexisNexis® Early Data Analyzer** helps you reduce your e-discovery files, assess risk, and explore the merits of each case at the earliest possible point.
- **LAW PreDiscovery™** combines imaging and electronic discovery processing into one easy-to-use application to help you produce and organize both paper and electronic files more efficiently.
- **Concordance® Evolution** helps you reduce your overall e-discovery cost and boost your litigation workflow productivity while managing your largest and most complex cases.

And now we're taking another step forward by offering near-duplicate, email-threading, and automated coding tools. One of these tools is LexisNexis® Near Dupe.

Announcing LexisNexis® Near Dupe

LexisNexis is now the exclusive provider of LexisNexis® Near Dupe (formerly known as Polaris® ND). This tool filters out irrelevant documents from the e-discovery process by flagging documents that are nearly identical. It recognizes documents within an e-discovery collection that are similar—based on parameters you determine—by evaluating the actual content of the documents.

Key features include:

- Identification and grouping of exact and near duplicates
- Identification of the one document that is the most representative within each near-dupe family
- Easy integration from OCR-only databases, extracted electronic text databases, or common load files

You can reveal document relationships beyond those available from traditional near-dupe analysis. This means faster and more efficient lawyer review, more consistent review decisions, fewer documents hosted in a review environment, and increased quality control of outgoing productions.

LexisNexis®

LexisNexis® Near Dupe field names and definitions

Once your documents are loaded and processed, LexisNexis Near Dupe provides culled information separated into the following fields:

- **DocID:** This is the unique ID of the document being described by the row of data. The DocID is assigned when data is initially loaded.
- **ND_Cluster:** This is a unique ID for each Cluster. A Cluster is a broader grouping than a Family. If two documents belong to the same Cluster, they will have some relationship, but it may be loose or indirect.
- **ND_ExactDupeSet:** This identifies exact duplicates. If two documents share an ExactDupeSet ID, they share exactly the same text. This grouping is based solely on textual content, so the documents may have different metadata or be in different file formats. Because of this, programs that generate MD5 or SHA hashes may not assign the same hashes to documents that belong to the same ExactDupeSet.
- **ND_Family:** This a unique ID for each near-duplicate Family. Families are sets of documents that are all highly similar, if not identical, to each other. The Family ID is composed of two numbers separated by a hyphen. The first part identifies the Cluster to which the Family belongs. The second part identifies the Family within that Cluster.
- **ND_IsMaster:** This indicates whether the document is a "Master document." A Master document is the one identified as being most representative of the near-duplicate Family to which it belongs. A Master document is indicated by a "Y" in this field. For all others, it will contain a blank space.
- **ND_ReviewSet:** This is an ID for each Review Set. LexisNexis Near Dupe tries to place more closely related documents and Families together in the same Review Set.
- **ND_ResultSet:** This is the name of the Result Set being exported. This is a name you can assign, which can be used for tracking purposes. If you don't assign a name, LexisNexis Near Dupe will create a name based on the date the review set was generated.
- **ND_Similarity:** This indicates the document's level of similarity to the Master document of its Family. Master documents are 100 percent similar to themselves.
- **ND_Sort:** This provides a sort order based on document similarity. Families and Clusters are kept together, and more closely related documents and Families are placed closer together.

Want more information?

Learn more about LexisNexis near-duplicate, email-threading, and automated coding tools by checking out www.lexisnexis.ca/litigation-reporting

LexisNexis®