

One Decision[®] Document Review Accelerator
Integrated Near Duplicate Identification Technology
Orange Legal Technologies



By the Numbers: The Need for Technology in Attorney Review

Seventy.

Integrated near-duplicate detection has the potential of reducing review costs by up to 70%.¹

Sixty.

Industry studies estimate up to 60% of documents in legal reviews may be near-duplicates.²

One.

One Decision® *Document Review Accelerator* from OrangeLT™ simplifies reviews with advanced near-duplicate identification to deliver more consistent, less costly and higher quality results.



One Decision®: Document Review Accelerator from OrangeLT™

Orange Legal Technologies' proprietary *One Decision Document Review Accelerator* leverages advanced near-duplicate identification technology to enhance reviewer efficiency by grouping similar documents and allowing them to be considered together during legal document reviews as part of the eDiscovery process. This grouping and ability to propagate coding decisions throughout all near-duplicate documents results in more efficient and economical document reviews. *Advantages to One Decision from OrangeLT include:*

Increased Consistency.

Reduced review redundancy by focusing reviewers on appropriate content.

Decreased Time.

Reduced number of documents to be reviewed without sacrificing review quality.

Decreased Risk.

Reduced risk of multiple reviewers reviewing similar documents.

Decreased Cost.

Reduced dollar and hour resources required for a complete review.

Increased Quality.

Reduced inconsistent tagging decisions on similar documents after first-pass review.

Fully integrated into the OneO® Discovery Platform, *One Decision* enhances the review capabilities of OneO to deliver a more consistent, less costly and higher quality approach to technology enhanced review. *One Decision* does this by empowering reviewers to select **ONE DOCUMENT**, make **ONE DECISION** and have that decision propagated throughout all near-duplicates for **ONE RESULT**.

¹ *Orange Legal Technologies' Service and Technology Pricing*. Orange Legal Technologies. January 2013.

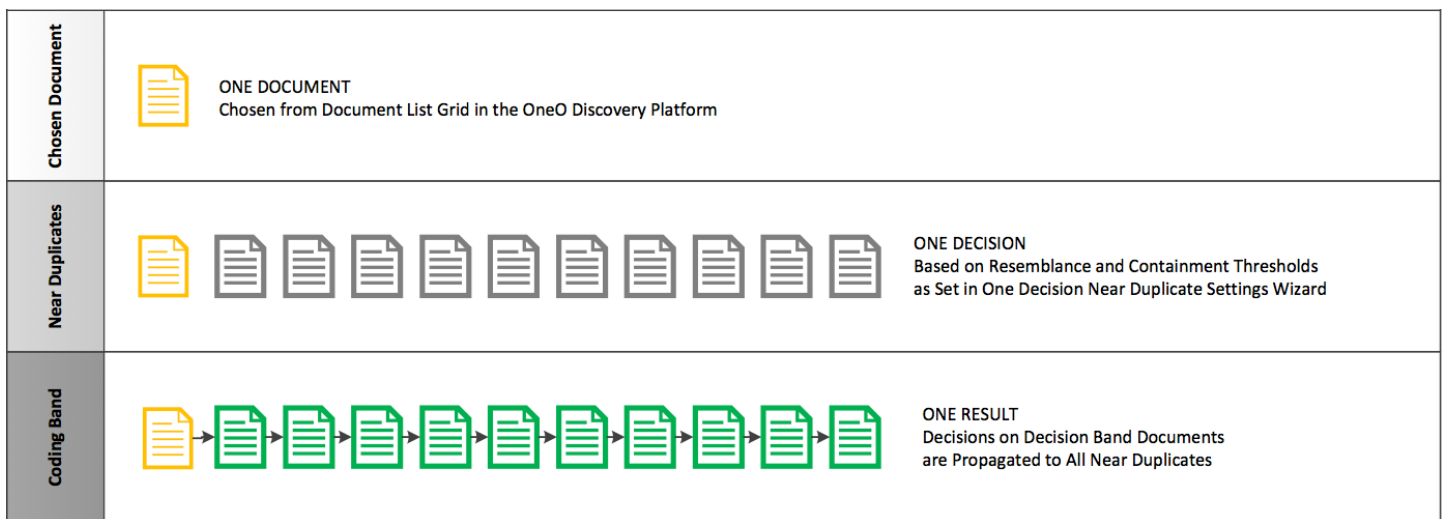
² *Where The Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery*. Nicholas M. Pace and Laura Zakaras. RAND. April 2012. 53.



One Document. One Decision. One Result.

The One Decision Review Accelerator allows users to experience and benefit from the advanced near-duplication identification technology developed by Orange Legal Technologies by providing them a simple, three step near-duplicate approach that is accessible directly from the OneO Discovery Platform Review Module. This One Decision Near-Duplicate Review Accelerator approach consists of:

- 1) Reviewers selecting ONE DOCUMENT from the review module Document List Grid for near-duplicate evaluation.
- 2) Reviewers applying ONE DECISION on resemblance³ and containment⁴ thresholds as set in the Near Duplication Setting Wizard to establish a targeted set of near-duplicate documents.
- 3) Reviewers making coding decisions on ONE DOCUMENT and having those decisions propagated throughout associated near-duplicates for ONE RESULT for all documents.



Three Step *One Decision* Review Discovery Accelerator Approach

This highly intuitive, fully integrated approach simplifies and expedites document reviews while delivering more consistent, less costly and higher quality document reviews.



About the One Decision[®] Document Review Accelerator

What is Near-Duplicate Detection of electronically stored information (ESI)?

Near-Duplicate Detection is a technology enabled feature of advanced eDiscovery platforms that allows for the identification and grouping together of documents that have strong similarities in regard to content and context, yet are not completely identical.

³ *Resemblance* (or Jaccard coefficient) of two documents is defined as the size of the intersection of their shingle sets divided by the size of the union of shingle sets. (A *Shingle* is defined as a contiguous sequence of words.)

⁴ *Containment* of two documents is defined as the size of the intersection of shingle sets divided by the size of the shingle of the initial shingle set.

Common academic definitions of near-duplicate detection include:

- **Near-Duplicate Detection** is a method of grouping together nearly identical documents.⁵
- **Near-Duplicate Detection** groups together documents that contain mostly identical blocks of text or other information while differing in some way (if two or more documents were, in fact, completely identical, all but one should have been dropped during the prior processing phase as part of deduplication efforts). Such differences can include minor amounts of additional or deleted text, altered formats, or variation in file types. One example of a near-duplicate would be a word processing document file and a scan of a printed version of that same document after being subjected to OCR. Another example of near-duplicates would be multiple drafts of the same document, with only slight differences between revisions. When grouped in this way, the reviewer might decide that, for example, because the “master” or “pivot” document in the set (the one that is judged to be most representative of the entire group) appears to be relevant and responsive, it is therefore not necessary to examine the other documents in the same set. Some applications highlight the differences between the master document and the related ones, thus allowing the reviewer to more quickly determine whether others in the group should be coded differently from the first viewed. Whether two documents are near-duplicates is essentially a subjective judgment, and applications allow users to adjust the similarity threshold (sometimes referred to as a likeness threshold or resemblance threshold), the statistical value that determines how close to an exact match the documents must be to be classified as a near-duplicate.⁶

Why is Near-Duplicate Detection important in eDiscovery?

When one considers that vendor estimates show that the number of near-duplicate documents in complex reviews may range from 20% to 60%⁷ of reviewable documents, Near-Duplicate Detection technology is important in the field of eDiscovery as it can significantly increase the speed of legal document reviews by decreasing the number of documents a review team must consider to gain a complete understanding of available information.

How is Near-Duplicate Detection executed within the OneO Discovery Platform?

Fully integrated into the architecture of the OneO Discovery Platform, *One Decision* Near-Duplicate Identification (NDI) capability is enabled by two wizard driven interfaces allowing user input and adjustment into NDI process. These wizards consist of the Near-Duplicate Settings Wizard and the NDI Wizard.

- **Near-Duplicate Settings Wizard:** The *One Decision* Near-Duplicate Settings Wizard is a companion application to the NDI Wizard and is used to specify the algorithm parameters for near-duplicate identification for a given OneO case and to prepare the case to utilize the NDA wizard. (*Figures 1 and 2*)
- **NDI Wizard:** After documents have been imported by the OneO Conversion Wizard into a case database (standard procedure in all OneO matters), the NDI Wizard can be used to update the case database with a list of near-duplicate document pairs. OneO Review can take advantage of this near-duplicate information to display the near-duplication for any selected document(s), as well as to include near-duplicates of search hits into search results. The NDI Wizard is designed for use in an incremental batch-based processing environment. When new document batches are imported, the NDI Wizard uses stored results from previous processing so that it does not need to re-process the entire case. Near-duplicate pairs are identified only for newly imported documents, and the case database is incrementally updated. The NDI Wizard also provides the ability to “roll-back” batch processing. It will automatically detect when previously processed batches have been deleted or are being re-processed, and it will automatically repair data files and database records.

⁵ The Grossman-Cormack Glossary of Technology Assisted Review. Version 1.03. By Maura Grossman and Gordon Cormack. December 2012.

⁶ See Nicholas M. Pace and Laura Zakaras. RAND. April 2012. 51.

⁷ See Nicholas M. Pace and Laura Zakaras. RAND. April 2012. 53.

Figure 1 - Example of Near-Duplicate Settings Wizard (Pre-processing Setting).

Near-Duplicate Settings Wizard - Version 1.0.3

Wizard Page 2
Please specify the parameters to be used for this case. Parameter values will be saved to the database on the next page. Note that values cannot be changed after they have been saved to the database!

Shingle Parameters

Shingle Size (Words): 5
Minimum Shingles For Containment: 1

Near-Duplicate Thresholds

Minimum Resemblance (Percent): 70
Minimum Containment (Percent): 80

Shingle Sampling

None
 Modulus
 Adaptive Reduction Factor: 1

Text Analysis

Include Numbers In Shingles
 Use Stop Words
 Use Language Stemming
Language: English

Maximum Length Of Metadata File (Bytes): 0 (Zero implies no limit to maximum length.)

Help < Back Next > Cancel

Server: CALIFORNIA Case: 539

Figure 2 - Example of Near-Duplicate Settings Wizard - Viewing Filter (Post-processing Setting).

Near-Duplicate Viewing Filter

When this filter is enabled, it limits the display of Near-Duplicate documents so that only documents with Resemblance and Containment values that are inside the filter range will be shown. The filter also sets a limit to the number of near-duplicates that can be displayed.

Disable Viewing Filter Reset Viewing Filter

Filter By Resemblance

Resemblance

100%
0%

Set Maximum Maximum: 100 %
 Set Minimum Minimum: 50 %

AND
 OR

Filter By Containment

Containment

100%
0%

Set Maximum Maximum: 100 %
 Set Minimum Minimum: 50 %

Maximum Near-Duplicates To Display: 10000

OK Cancel

How is Orange Legal Technologies' delivery of Near-Duplicate Detection technology different than other approaches?

Differentiation factors for One Decision from OrangeLT include:

- **Transparency:** User visibility and access to near-duplicate identification algorithm parameters.
- **Adaptability:** User adjustable thresholds for near-duplicate resemblance and containment
- **Integration:** Fully integrated into OneO and presented as an optional part of a comprehensive eDiscovery workflow.
- **Affordability:** Developed by OrangeLT and delivered without the third party licensing cost constraints typically associated with near duplication technology.

To learn more about how OrangeLT can augment your eDiscovery capabilities with the *One Decision Review Accelerator*, contact us today for an introductory briefing and demonstration.



About Orange Legal Technologies

Formed in 2008 as an eDiscovery-centric outgrowth of the 1995 founded Litigation Document Group, Orange Legal Technologies is a leading provider of electronic discovery litigation, audit, and investigation services for law firms and corporations. Having served over 1,000 clients since inception and with over 250 clients leveraging the OneO® Discovery Platform since its introduction, OrangeLT™ has worked with some of the world's most well known corporations and law firms and has been recognized in leading analyst and media publications from such organizations as Gartner, IDC, 451 Research and Forbes.

OrangeLT™ is a leading provider of electronic discovery litigation, audit, and investigation services for law firms and corporations. OrangeLT offers a complete suite of electronic discovery technology and services, including collection, analysis, processing, review and production of digital and paper-based information. OrangeLT is enabled by the OneO® Discovery Platform – an integrated, web-accessible electronic discovery platform that provides online analysis, processing, and review of unstructured data from the security of a hosted centralized repository.

To learn more about Orange Legal Technologies, contact us today at:

General Contact Information:

Website: OrangeLT.com

General Information – info@orangelt.com / 801-328-4566

Customer Support – customersupport@orangelt.com / 801-328-4566

Sales Support – sales@orangelt.com / 512-934-7531

Headquarters Location

30 East Broadway, Suite300 | Salt Lake City, UT 84111

801-328-4566 telephone | 801-355-3770 fax