Intella Connect 2.4 Release Notes

Released: **October 29, 2020.**

Highlights

- Added **Predictive Coding**.
- Added detection of **near-duplicates**.
- Added support for **single sign-on (SSO)**.
- Added **hash-based filtering** of items during indexing, e.g. for **DeNISTing**.
- Extended the **custom columns** functionality.

- Added support for indexing **AFF4** and **VHDX** files.
- Added support for **Oxygen 12 and 13** reports.
- Added **custodian-based deduplication** and **family-based deduplication**.
- Added an improved **tagging dialog**.
- **Performance** and **scalability** improvements across the board.

Predictive Coding

- Case managers can now create a review queue that is driven by our Predictive Coding engine. This engine uses artificial intelligence (AI) to learn from human coding decisions and prioritize a set of items to review. In typical eDiscovery cases with only a small fraction of responsive documents, this helps reviewers quickly identify the most pertinent data in the case, resulting in large time and cost savings.

- A case manager will typically first use the searching and filtering options available in Intella Connect to cull the data set (e.g. custodians, types, dates, keyword lists) and then launch a Predictive Coding-based review of the remaining items of interest.

- Reviewers working on such a review queue will be presented with documents selected by the engine and are asked to code them as Responsive or Non-Responsive. Initially, the engine will select diverse documents that make for good training samples while it tries to get an overview of the document space. When sufficient documents have been found for both categories, the engine starts looking for textual patterns that separate these categories. Once clear patterns start to emerge, the engine will use them to predict the responsiveness of the remaining documents, and the review queue will shift from showing training documents to presenting the most likely responsive documents to the reviewers. The engine will keep learning from additional coded documents and fine-tune its classification model along the way.

- Multiple reviews can be launched and used in parallel in a case, e.g. to represent different aspects of the case.

- As human coding progresses, several statistics are available that show the ratio of responsive documents, indicating the success rate of the predictions made by the engine.

- The Predictive Coding-based review queue provides an alternative to traditional batching and coding. Responsive documents will be identified earlier, especially in cases that have a small percentage of responsive documents.

- Once the overall responsiveness rate in the queue starts to drop, i.e. reviewers are being presented more and more non-responsive documents, case managers can opt to run an elusion test. This is a dedicated review process that statistically determines the relevance rate of the non-coded documents, and thereby establishes the recall rate. Elusion tests help defensibly justify a review team to stop reviewing the remainder of the queue.

- After an elusion test has been completed and the prediction model has been verified and accepted, the review team can choose whether to apply the engine's predictions to all unseen items.

- Preparing and running a Predictive Coding queue is a hardware-intensive process, particularly in terms of RAM usage, and secondary in terms of the CPU load. When running it on an item set in the tens of thousands of items or more, please consider using a server that exceeds the amount of RAM listed in the system requirements. Please consult the user manual for guidance.

- Predictive Coding is available to all Connect licensees – no additional license is needed but this may change in future versions.

General

- Various usability improvements.
- Various performance and memory usage optimizations aimed at large (multi-terabyte) cases.
- Resolved an issue with the server becoming unresponsive due to unexpected file locks on files that only need read access.
- Creation of minidump files (*.mdmp) is now disabled, as these files can be very large and are of little use.
- Any Java hs_err_PID.log files are now always written to the case folder.

Licensing

- Starting with Intella Connect 2.4, to ensure compliance with Intella Connect's licensing agreement, a Connect license will not allow a second instance of an older version to run on a different machine. One older version instance may still be run concurrently, but only on the same server running the 2.4 version, and only for the purpose of migrating to the newer version. Note that Upgrade Notes section on supported case versions: Intella Connect 2.4 can open cases made with versions 2.1.x, 2.2.x and 2.3.x, so there should be no need to run past versions in parallel.
- Updated the bundled haspupdate.exe, used for creating and applying C2V and V2C files. This resolves issues with the V2C files not applying successfully on some machines.
- Reliability improvements when checking for a valid license.
- When displaying network dongles, the Dongle Manager no longer recommends the user to generate a C2V file and send it to customer support, as that is no longer a necessity when activating or updating a network dongle.
- It is now possible to create a C2V file of any dongle (also non-network dongles) in the Dongle Manager, by right-clicking on the dongle.
- SL license keys were incorrectly displayed as "HASP Certificate dongle" keys in the License Chooser dialog. They are now displayed as "HASP software-based license" keys.

Authentication

- Restyled the login page.
- Connect admins can now configure custom messages and links on the login page. This can be used for purposes such as linking to customer-specific help pages, listing support contact details, stating messages on planned server downtime, etc.
- Added support for Single Sign-on (SSO). This lets users log in to Intella Connect and several other services by logging in only once with an account shared across all these services. The authentication provider needs to be OpenID Connect (OIDC) complaint. Example setups are included for setting up Google G Suite and Okta as a provider.

- Added an alternative LDAP integration module (still turned off by default) that can help avoid time-outs and other connectivity issues with LDAP connections.
- Resolved an issue with local (non-LDAP) users not being able to change their password when the username they logged in with differed in casing from the exact registered name (e.g. "John" vs "john").
- Resolved an issue with the login form not showing properly in Internet Explorer 11.

Security

- Resolved an issue where user-defined hosts were not applied when using SSL certificates with wildcard domains.

Case Management

- A warning is now shown in the Admin UI when the amount of cases grows to an unsupported amount.
- Improved the error logging when the user tries to open a case located in a shared folder that has already been opened on a different machine.
- Resolved an issue with folder creation not working properly in the Create Case dialog.
- Resolved some corner cases where users could not add new sources to a case.
- Resolved an issue with failing case merges due to custom column types being configured incorrectly in the merge configuration.

Indexing - General

- Added support for filtering files during indexing based on a repository of known hashes. This can be used for DeNISTing items, among other tasks.
- Added a Custom Columns Manager UI, for defining custom columns that are populated with data from selected email headers, raw data fields, and load file columns.
- Several improvements to the processing of MS Exchange databases:
  - Resolved decoding issues for encoded email body texts.
  - Decoding of properties stored in new Exchange property blob formats.
  - Resolved a performance issue that made processing of some EDB files very time consuming.
  - Improved extraction of email-attachment relations.
- Several improvements to the processing of PDF documents, EML files, plain text files, and Slack reports. The improvements address data completeness, the ability to process damaged files, and performance.
- Added support for Oxygen 12 and 13 reports.
- Indexing of Windows Event Log files has been made optional. By default, it is turned off.
- Added support for note attachments in Cellebrite reports.
- Added detection MSI installer packages, IE crash recovery stores and Acrobat Forms Data.

- The user interface for downloading the GeoLite2 database has been adjusted to incorporate the use of a license key. Due to changes in MaxMind's policies, one must now register for such a license key to be able to download this database.

- Resolved an issue with missing WeChat message participants in Cellebrite reports.

- Improved file type and character encoding detection for item data, resulting in more complete processing as well as a reduction of false-positive identifications of file types like NSF, MBOX, EXE, XML and more.

- Improved the speed of indexing Windows Event Log files.

- Resolved a reliability issue when validating Notes ID files.

- Resolved an issue where exceptionally large Windows Event Log files would cause the indexing process to run out of memory.

- Resolved a synchronization issue with the Key Store that caused reliability issues on encrypted items with duplicates.

- Resolved various email header and body decoding issues and font issues, particularly with emails in Asian languages.

- Resolved an issue with incorrect timestamp parsing of iTunes backups due to changes in this storage format.

- Resolved an issue with incorrect GeoLite2 databases (damaged files or incorrect database type) interfering with the correct indexing of certain items.

- Resolved a case database issue when the maximum item ID value in a case exceeded 165,191,047.

- Reduced disk space usage of temporary files.

Indexing – Disk images

- Added support for indexing AFF4 files.

- Added support for indexing VHDX files.

- When adding a Disk Image source, the Find Parts button no longer adds the .log file to the list of image parts.

- Resolved an issue with structurally sound L01 images being incorrectly marked as broken files.

- Resolved an issue where an E01 and DMG image of the same physical disk resulted in different item counts when indexed.

- Resolved a memory leak when indexing BitLocker-encrypted disk images that caused the disk image verification process to fail.

- Resolved an issue with items representing folders in the Recycle Bin getting Size and MD5 Hash properties.

Indexing – Load files

- Resolved an issue with load file overlays not importing custodian information properly.

- Resolved an issue with load file verification validating the text and native path fields even when the corresponding checkboxes were not checked.

- Resolved an issue with load files failing to import when they contain relative paths that start with a backslash character.

- Removed Custom ID and Custom Family ID columns from both the regular and overlay load file import, as it cannot be guaranteed that such imported data will produce a correct load file on a subsequent export.

Indexing – Cloud sources

- The bundled instructions for adding cloud sources have been moved to the Knowledge Base section on the Vound Support portal. This has been done so that these instructions can be updated as soon as Microsoft, Google or Dropbox change their administrative interface for obtaining authentication tokens.

- When indexing Gmail sources, potentially security-sensitive information such as authentication tokens are no longer being logged.

Indexing – Error reporting

- Resolved an issue where crawler crashes were not reported if a subsequent attempt to index the item took place. The second attempt may fail due to different reasons, thereby obfuscating the cause of the first crash. The second attempt may in fact succeed, in which case the fact that a crawler did crash could remain hidden.

Indexing – Intella Node

- The Nodes overview screen in the Connect Admin UI has been redesigned to accommodate the addition of more configuration options.

- Added options for configuring the SSL settings of a Node server.

- Added options for letting a Node server validate the presence of a geolocation database and downloading the GeoLite2 database.

- Resolved an issue with Notes validation failing on the Node server due to incorrect handling of preference files.

- Resolved an issue with Notes validation failing on the Node server resulting in a non-informative error message.

- Resolved an issue where, when browsing the local disks of a Node server through the Connect Admin UI, the local Connect server's disks were also shown in the folder tree.

OCR

- Updated the bundled ABBYY FineReader OCR library, addressing issues with several documents reported to us.

- Resolved an issue with paragraphs being reported in an incorrect order on some documents.

Analysis

- Added support for detecting near-duplicates. This functionality groups documents with similar textual content, e.g. emails with similar bodies, or multiple revisions of the same conceptual document. Near-duplicates can be used to find related information that will not be found by looking for exact duplicates (binary copies) of items of interest, matching message hashes, or email threading. It can also be used to sort items in a large review job, by grouping similar items and focusing on their differences.

- Users can invoke near-duplicate processing by selecting "Near-Duplicate Detection" in the Table's right-click menu. Intella Connect will compare the selected items for similarity using a user-configurable similarity threshold.

- The item in a near-duplicate set that has the longest text is assigned the status of "master item". All items in a near-duplicate set will have a similarity with this item that is larger than or equal to the specified similarity threshold.

- Additional table columns and Features facet categories have been added for listing the items that received near-duplicate processing, what near-duplicate group an item belongs to, and what its master item and similarity score are.

- Near-duplicate sets can be queried via the new Near-Duplicates facet, or by clicking the Show Near-Duplicates link in the Previewer.

- Improved the accuracy of highlighting credit card numbers, social security numbers (SSNs) and phone numbers found by Content Analysis.

- Resolved an issue with paragraph hashing not storing correct results when over 2 million (2²¹) items contained the same paragraph text.

- Resolved an issue with values not being removed from the Content Analysis facet when the corresponding items were removed from the case.

Searching

- Added a Unique Families search option. This search option performs family-based deduplication: it searches for all top-level parents of the selected items, deduplicates the parents, and then extends that deduplicated parent set with all their children. This can be used in eDiscovery workflows where unique families need to be produced. The deduplication step can optionally perform custodian-based deduplication (see the Results section).

- Added a Near-Duplicates facet, listing the group names of the detected near-duplicate sets, and allowing for them to be queried. See the Analysis section for a full description of near-duplicate processing.

- Added Has Near-Duplicates and Analyzed for Near-Duplicates categories to the Features facet. See the Analysis section for a full description of near-duplicate processing.

- Speed improvements in the initialization time of several facets.

- Resolved an issue with exceptionally large keyword lists that would fail to evaluate, by processing the keyword list in a streaming fashion.

- Resolved an issue with entire keyword lists failing to evaluate due to one or more invalid queries in the list.

- Phone numbers of SMS and MMS messages in the Chat Account facet are now normalized, like they already were in the Phone Number facet.

- Resolved an issue with wildcards being ignored when used in phrase queries holding only a single term.

- Resolved an issue with tag descriptions no longer being displayed in the Tags facet.

- The Geolocation map now shows a continuous map; one can keep panning in all directions. This resolves an issue with certain parts of the world map (especially Australia and New Zealand) being obscured by the Searches panel.

- Resolved an issue with the Sorting drop-down list in the Location facet not being fully visible.

Results

- Added support for custodian-based deduplication. When the case has one or more custodians defined, the Deduplicate button in the results list will be extended with a drop-down button, enabling the user to toggle between case-wide deduplication and custodian-based deduplication. When the latter mode is used, each custodian set is deduplicated separately; duplicates that exist across those sets will be retained.

- Added columns for Near-Duplicate Group, Near-Duplicate Master Item and Near-Duplicate Score. See the Analysis section for a full description of near-duplicate processing.

- Added columns for Duplicate Custodians, All Custodians, Duplicate Locations, All Locations. By default, these columns are empty. A "Generate Duplicate Custodians and Locations" task for populating these fields has been added to the Desktop version. This task will also be added in a future Intella Connect version.

- Added a Shadow Copies column. This shows the number of shadow copies of the item. Indexing of Volume Shadow Copy (VSS) information is currently only available in the Desktop version. This will also be added in a future Intella Connect version.

- Resolved an issue with the date filter in the results list not resetting properly when it was cleared and the page was subsequently refreshed.

- Resolved an issue with the tag tooltip in the List view not displaying properly.

- Resolved an issue with the case preferences not persisting any custom columns that were selected for display in the Table.

Previewer

- Added a Show Near-Duplicates link, for locating the near-duplicates of the displayed item.

- Improvements in the native rendering of MS Office documents.

- Improved the rendering of individual chat messages that are not part of a larger conversation.

- Improved the rendering of the tag tooltip that is shown when hovering over the Tags line in the Previewer.

- Resolved an issue with large lists of recipients not truncating properly.

- Resolved an issue with the native preview of HTML emails not displaying fully.

- Resolved an issue with HTML markup in IBM Sametime chat dumps interfering with their rendering, such as missing scrollbars and blocked mouse wheel scrolling.

- Resolved an issue with "Mark paragraph as unseen" not unsetting this state for duplicates of that paragraph.
- Resolved an issue with the "More" submenu on paragraphs collapsing before the user could choose an option in the menu.

Tagging

- Added an improved tagging interface. The new interface is optimized for searching for existing tag names, defining new tags, and removing tags. It improves how tags present in the current item set are displayed and what their respective volumes are. It is designed to better handle large tag hierarchies. The overall responsiveness of the dialog has been improved.
- The old tagging user interface is still available; see the toggle in the lower left corner to switch between the Simple View and Classic View.
- Added indentation to the tag tree in the Classic View, reflecting the tag hierarchy.
- Several improvements to the handling of tag colors throughout the user interface.
- Permissions have been added for adding and editing tags. Previously, only tag deletions were subject to permissions.

Batching

- Resolved an issue with wrong items being added to a batch when the Timeline in the Search tab was used to select the appropriate items.

Coding

- The items in the Review tab's item list now use color-coded swatches that indicate how that item has been coded. This makes it possible to, for example, easily spot the relevant items in the list of coded items.
- Improved the display of item families, by rendering dividers in the item list that separates the families, as well as indenting nested items. Custom IDs need to be generated on these items to make these features appear.
- The Review tab now allows for applying a coding decision to an entire family at once.
- Improved the handling of coding layouts that refer to deleted or moved tags.
- Resolved an issue with the Preview tab in the Review tab not loading properly on emails under certain circumstances.
- Resolved an issue with the coding form not clearing a "cannot advance" error message after the reviewer had fixed the reason for the error.
- Resolved an issue where the attachments of the current item were shown both at the top and bottom of the item view.

Redacting

- The redaction editor and underlying services now prevent the creation of multiple redactions on the exact same spot (same coordinates and dimensions). This addresses issues with automatically made redactions, using mass-redaction or the Redact All button.

- Improved performance of applying full page redactions to multiple pages.
- Added a "Draw border around white redaction" option to redaction templates.
- The event log now makes a distinction between redaction additions, removals, and changes. In older versions these were all listed as generic redaction events.
- Resolved an issue with the "Process queued items" task not being able to launch until a tag color was selected.

Dashboard – Insight tab

- Resolved an issue with case re-indexing not resetting the state of the Email Threading and Content Analysis checkboxes in the Workflow panel.
- Performance improvements in populating the "Top 10 Email Addresses" list. This used to be a potentially long and expensive operation.

Reporting – Keywords tab

- One can now query for the exclusive items of a specific query row. The resulting items will then be shown in the Search tab.
- Performance optimizations when evaluating a keyword list.
- Resolved an issue with incorrect hit counts for certain types of queries.

Exporting – General

- Improved the rendering of exported chat messages that are not part of a larger conversation.
- Performance improvements when exporting items using an export set.
- Resolved an issue with incorrect file names for items downloaded in original format, due to broken character encodings in certain HTTP headers.
- Resolved an issue with custom fields failing to export to a CSV.
- Resolved an issue with cases not going into IDLE mode after an export task completed and there was no other activity that should keep the case alive.
- Resolved an issue with exporting tasks not taking the selected export template into account.

Exporting – PDF

- Improvements in the native rendering of MS Office documents.

Exporting – PST

- Resolved an issue with calendar items failing to export to a PST file.
- Resolved an issue with calendar items lacking Sent and Received date attributes.

Exporting – Load files

- The resolution unit of TIFF images has been changed from 3 (centimeters) to 2 (inches). This resolves an issue with loading these images into the Relativity platform.

- Resolved an issue with tag columns being exported using an incorrect value separator.
- Resolved an issue with file names sometimes getting double file extensions (e.g. "document.pdf.pdf") in certain export configurations.

Exporting – Report

- Resolved an issue with the "Include original format files" and "Include extracted text" options in the Sections sheet not being applied when generating the report.

Printing

- Resolved an issue with printouts not looking well when using Microsoft Edge. This only affected the older, non-Chromium based Edge versions.

Upgrade Notes

Case versions 2.1.x to 2.3.x – Intella Connect 2.4 can directly open cases made the 2.3.x, 2.2.x and 2.1.x versions of Intella and Intella Connect.

Upon first startup, Intella Connect 2.4 will migrate all existing coding layouts to a new format. This process reads each coding layout XML file stored in C:\Users\[USERNAME]\AppData\Roaming\Intella\coding-layouts and creates a new XML file in its place, making sure that there will not be any naming conflicts. A copy of the original coding layout is preserved in the "migrated" subfolder.

When a case made with Intella or Intella Connect 2.3.0 or older is opened in Intella Connect 2.4, adding Custom ID tasks to the case in version 2.4 may render the case unopenable in that older version. Changes have been made in the 2.3.1 version to prevent such forward compatibility issues in future releases.

When a case made with Intella or Intella Connect 2.1 or older is opened, all Content Analysis results are automatically migrated to a new data storage format. This migration happens only once. The old store is retained and will still be used when using version 2.1 or older. New results will not be added to the old store though, and new results added to the old store will not be migrated. Cases made with 2.1.1 or later already use the new data store and are therefore not affected.

When items in cases made with a 2.1.x version are exported to a separate case, the registered case size of the target case will be incremented with the size of the original case. This may be problematic for users with licenses that have a case size limitation. To resolve the inflated case size, the source case needs to be re-indexed before exporting items from it.

In Intella Connect 2.2.2, a new method for calculating message hashes was introduced. While this change is transparent, please be aware that message hashes will change when re-indexing a case that has been made with an older version.

Case versions 1.9.x to 2.0.x – Cases made with the 1.9.x and 2.0.x versions of Intella and Intella Connect can be opened, but these cases first require conversion. Case conversion will create a copy of the case in which all item data is converted, and all tags, comments and flags are imported. The existing case will not be altered in any way and can afterwards still be opened in the older Intella version. Access to the original evidence files is not required for case conversion.

Caveats concerning the case contents:

- Case conversion will not transfer the geolocation metadata extracted from emails when the "Determine geographic location of emails" option was used. Re-indexing of the converted case is required to restore such metadata.
- Multi-page TIFFs will be displayed in the converted case as if they were single-page TIFFs. Exporting and printing of the item report does reveal the other pages. Re-indexing of the converted case will make all pages displayable again.

Case conversion will require sufficient disk space. As a rule of thumb, please reserve twice the amount of the evidence size for your case folder.

Other case versions – Cases made with Intella 1.8.x or older are not supported. Cases made with beta versions are not supported and should be recreated.

Software versions – Vound will provide technical support for one major past version. For this release that will mean the 2.3.x range of products. Vound will always recommend that users upgrade to the latest version.