Veritas eDiscovery Platform™

Audio Search Guide

10.1.3

VERITAS™

Veritas eDiscovery Platform™: Audio Search Guide

The software described in this book is furnished under a license agreement and may be used only in accordance with the terms of the agreement.

Last updated: 2022-6-18.

Legal Notices

Copyright © 2022 Veritas Technologies LLC. All rights reserved.

Copyright © 2018 Nexidia Inc. All rights reserved, worldwide. Nexidia and the Nexidia logo are trademarks of Nexidia Inc. All other trademarks are the property of their respective owners. All Nexidia materials regardless of form, including without limitation, software applications, documentation and any other information relating to Nexidia Inc., and its products and services are the exclusive property of Nexidia Inc. or its licensors. The Nexidia products and services described in these materials may be covered by Nexidia's United States patents: 7,231,351; 7,263,484; 7,313,521; 7,324,939; 7,406,415, 7,475,065; 7,487,086, 7,640,161; 7,650,282; 7,769,587, 7,904,296; 7,949,527, 8,051,086; 8,311,828 and/or other patents pending and may be manufactured under license from the Georgia Tech Research Corporation USA

Veritas and the Veritas Logo are trademarks or registered trademarks of Veritas Technologies LLC or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners.

This product may contain third-party software for which Veritas is required to provide attribution to the third party ("Third-Party Programs"). Some of the Third-Party Programs are available under open source or free software licenses. The License Agreement accompanying the Software does not alter any rights or obligations you may have under those open source or free software licenses. Refer to the Third-Party legal notices for this product at:

https://www.veritas.com/about/legal/license-agreements

The product described in this document is distributed under licenses restricting its use, copying, distribution, and decompilation/reverse engineering. No part of this document may be reproduced in any form by any means without prior written authorization of Veritas Technologies LLC and its licensors, if any.

THE DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID. VERITAS TECHNOLOGIES LLC SHALL NOT BE LIABLE FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH THE FURNISHING, PERFORMANCE, OR USE OF THIS DOCUMENTATION. THE INFORMATION CONTAINED IN THIS DOCUMENTATION IS SUBJECT TO CHANGE WITHOUT NOTICE.

The Licensed Software and Documentation are deemed to be commercial computer software as defined in FAR 12.212 and subject to restricted rights as defined in FAR Section 52.227-19 "Commercial Computer Software - Restricted Rights" and DFARS 227.7202, et seq. "Commercial Computer Software and Commercial Computer Software Documentation," as applicable, and any successor regulations, whether delivered by Veritas as on premises or hosted services. Any use, modification, reproduction release, performance, display or disclosure of the Licensed Software and Documentation by the U.S. Government shall be solely in accordance with the terms of this Agreement.

Veritas Technologies LLC
2625 Augustine Drive
Santa Clara, CA 95054
http://www.veritas.com

Contents

About This Guide 7
Revision History 7
Technical Support 9
Documentation 9
Documentation Feedback 9

Overview 11

Audio Search Basics 11
What is Audio Search? 12
How Does Audio Search Work? 12
Multiple Language Support 13
Language Pack 13
Language-specific Search Documentation 13
Using Audio Search for Optimal Results 13

Getting Started 15

Setting Up Your System: Server-Side 16
Prerequisites 16
Installation 17
Organizing Your Audio Media 19
Folder Management 19
Folder Example 19
Folder Names 19
Optimize Audio Search for Better Processing 19

Audio Search Processing 21

Audio Search Processing Workflow 21
Setup 21
Step 1: Creating a Case 22
Step 2: Specifying Default Audio Processing Language Pack 22
Step 3: Adding Audio Media Sources 23
Step 4: Enabling Audio Processing 24
Step 5: Specifying Audio Processing Language Pack 24
Step 6: Processing Audio Media Source Data 25
Processing An Audio Media Source Data With Different Language Packs (Optional) 25
Generating Audio Processing Reports 26

Audio Search 29

Audio Search Workflow 29
Audio Search Steps 30
Using Audio Search 30
Steps 1,2, & 3: Constructing the Query 31
Steps 4, 5, & 6: Previewing & Validating Media Results 35
Considerations: 37
Step 7: Adjusting & Tuning Results 37
Step 8: Applying Work Product (Tags, Notes and Folders) 38
Creating the Audio Search Report 39
Tips For Creating Good Audio Search Queries 40
Query Construction 40
Query Accuracy 41
Exporting Audio Search Results 43
Audio Search Export Considerations 43

Appendix A: Phonemes 47

North American English 47

Appendix B: Media File Types (Formats) 55

Supported Processing & Search Media Formats 55
Unsupported Processing and Search Media Formats 56
Supported HTML5-Based Media Player Media Formats 56

Appendix C: Language Support 59

Appendix D: TCP Port Usage 61

Appendix E: Scaling Audio Search & Processing (Nexidia) 63

Scale Audio Processing 63
Install Compute Node on Utility Node (cw-util) 64
Modify Properties on Appliance (cw-appl) 65
Restart the Services 66
Removing Compute Node 66
Scale Audio Search 66
Assess Pros and Cons of Adding a Data Node for Audio Search IMPORTANT! 66
Install Data Node on Utility Node (cw-util) 66
Modify Properties on Appliance (cw-appl) 68
Redistribute Data to the New Data Node on the Utility Node (cw-util) 68
Redistribute Data From a Data Node Back to Appliance (cw-appl) 69

Appendix F: Product Documentation 71

Audio Search Guide

This guide explains how to set up audio search, process and search audio content, and use speech analytics to reduce and identify relevant audio media for review and export. Audio reporting capabilities are also discussed.

This section contains the following:

- *"About This Guide" in the next section*
- *"Revision History" on page 7*
- *"Technical Support" on page 9*
- *"Documentation" on page 9*
- *"Documentation Feedback" on page 9*

About This Guide

Refer to the following guides for useful information about audio search functions:

- Installation Guide
- System Administration Guide
- Case Administration Guide

Revision History

The following table lists the information that has been revised or added since the initial release of this document. The table also lists the revision date for these changes.

Revision Date	New Information
July 2022	• Updated the content and images related to Audio Search Results box that appears on Filters panel in search result page for refining search.
March 2022	• Updated the content as per UI change in Analysis & Review tab for 508 compliance.
December 2021	• Updated version for release 10.1
March 2021	• Minor edits
March 2020	• Minor edits
October 2018	• Minor edits
June 2017	• Minor edits
July 2016	• Branding and minor edits • Remove references to Apple Quicktime as Nexidia install provides all the codecs needed.

Revision Date	New Information
August 2015	• Added Flash/IE 11/Desktop Experience content • Remove Rights Management Guide
March 2015	• Image accessibility • Branding and minor edits
October 2014	• Updated graphics for Item Level View • Branding edits
May 2014	• Branding and minor edits.

Technical Support

Technical Support maintains support centers globally. All support services will be delivered in accordance with your support agreement and the then-current enterprise technical support policies.

For information about our support offerings and how to contact Technical Support, visit our website:

https://www.veritas.com/support

You can manage your Veritas account information at the following URL:

https://my.veritas.com

If you have questions regarding an existing support agreement, please email the support agreement administration team for your region as follows:

Worldwide (except Japan)	CustomerCare@veritas.com
Japan	CustomerCare_Japan@veritas.com

Documentation

Make sure that you have the current version of the documentation. The latest documentation is available from:

- **Documentation** link at the bottom of any page in the Veritas eDiscovery Platform landing page.
- **Veritas Technologies LLC. All rights reserved. Products Web site:** https://www.veritas.com/product/a-to-z

Documentation Feedback

Your feedback is important to us. Suggest improvements or report errors or omissions to the documentation. Include the document title, document version, chapter title, and section title of the text on which you are reporting. Send feedback to:

eDiscovery.InfoDev@veritas.com

You can also see documentation information or ask a question on the Veritas community site.

https://vox.veritas.com/

Overview

Release 7.1.4 of the Veritas eDiscovery Platform introduced audio search functionality to its processing, analysis and review (PAR) capabilities. Audio Search allows users to process the audio content of multimedia files (both audio and video) in order to search for spoken words occurring within them. This is easily accomplished without leaving the Veritas eDiscovery case management workspace and platform.

Powered by Nexidia's accurate and performance-driven phonetic speech technology, the system is able to efficiently process sound recordings. The phonetic search approach allows quick and easy access to segments of interest in sound recordings. Users can search, review, and analyze indexed audio data to produce relevant audio data subsets for legal, corporate compliance, government regulatory and forensic purposes. After review, the search results can be exported to downstream third-party tools for further analysis or for preparation for legal presentation.

This chapter contains the following topics:

- *"Audio Search Basics" on page 11*
  - *"What is Audio Search?" on page 12*
  - *"How Does Audio Search Work?" on page 12*
- *"Multiple Language Support" on page 13*
- *"Using Audio Search for Optimal Results" on page 13*

Audio Search Basics

More and more audio content is being created every day. This explosion of audio content comes from a variety of sources such as voice mail, call center conversations, and recordings of regulated financial, banking, trading floor and insurance transactions. A variety of industries and sectors ranging from commercial companies, contact centers, government, regulatory, medical, security and emergency services are interested in this data. Not surprisingly, audio data is playing an increasingly important role in the eDiscovery landscape. It is this growth in audio content that is driving the need for highly efficient audio search and analytic capabilities and tools.

What is Audio Search?

Audio processing and search is the ability to index an audio media file in order to perform useful queries such as:

- Locate the places where the words "*subprime*" and "*Bank of Antigua*" are spoken within 5 seconds of each other
- Find audio recording segments about trade option content of first-order derivatives that mention "*delta*" and "*Société Générale*" from French trader and custodian Jacques Enlaboîte
- Find the audio portions where "*customer rebate*" and "*problem*" and "*rebate link*" are spoken by a customer caller
- Show the positions (in seconds from the beginning of the audio file) where those phrases occur

Benefits

The results and analytics from such queries make audio search tremendously valuable to organizations that use and apply its findings and insights.

How Does Audio Search Work?

The more you know about how audio search works, its capabilities, and how it displays results, the better it can serve your needs.

Background

Currently, there are two dominant audio search approaches: Speech to Text Extraction (also referred to as Large Vocabulary Continuous Speech Recognition or LVCSR) and Phonetic Indexing and Search.

The Veritas eDiscovery Audio Search module is based on the phonetic indexing and search technology developed by Nexidia.

- **Speech-to-Text Extraction (LVCSR)**

  A speech-to-text based technology, LVCSR attempts to recognize words in an audio file and convert them into human readable text. The text is extracted and indexed using a text indexer. Users can then perform conventional keyword searches.

  This approach relies on the perfect translation of the spoken word into text which is not always possible and can have additional issues around creating searchable text that accurately renders spoken content from poor quality recordings and speakers that use jargon, slang or have accents.

- **Phonetic Indexing and Search**

  Phonetic indexing and search is a technology based on the phonetic representation of the pronunciation of a spoken word. This means that instead of looking for words, this approach looks for sounds, called *phonemes*, which are the fundamental building blocks that make up any spoken language.

  Phonetic search emphasizes how things sound and is not reliant on knowing what a particular sound means. For example, the English language has more than a million words while the North American English dialect is made up of only 40 phonemes. Searching for strings of this limited set of phonemes simplifies audio search and reduces dependence upon a particular dictionary or lexicon. A phonetic search performs a probabilistic search (how much does this sound actually *sound* like the search term?) and associates a probability with the match of a term to a spoken phrase.

  Once the audio data has been processed and indexed, searches can be done directly on phrases while applying operators like time-based proximity to the audio content. The audio search engine identifies and matches the phonetic equivalent of the search string and returns relevancy-ranked results.

Multiple Language Support

Language Pack

Do you need to search French, Mandarin, or Australian media content? Audio search supports these languages and many others. This functionality is accomplished through the concept of a "language pack". A language pack includes the phonetic content of the target language and also takes into account regional accents, dialects, gender and other differences in speech. Currently, the eDiscovery Platform supports and provides language packs for 14 languages. The primary and system default language pack is North American English. See *"Appendix C: Language Support" on page 59* for a complete list of supported languages and associated language packs.

Language-specific Search Documentation

In addition to this guide, a helpful set of language-specific documents (PDFs) containing search tips and hints is automatically installed during the installation process. You can find these documents on: `C:\Program Files(86)\Nexidia\Language Packs\<language>.`

Using Audio Search for Optimal Results

- The Audio Search module requires the 7.1.4 release or later.
- While high quality recordings work best, audio search can work with recordings that have varying audio quality, accents and formats.

- A case folder containing audio content can only be processed with one language. To process audio content with more than one language, make a copy of the data for each required language and process them in separate case folders.
- Audio processing is a CPU intensive operation. During processing of large cases, ensure that other CPU-bound operations such as OCR processing are not running.

Getting Started

This chapter covers how to set up the basics in order to successfully configure the audio search software.

Note: Audio Search functionality is only available for new cases created on the 7.1.4 release. It is not available for cases created using older versions (even if the cases have been upgraded).

Refer to the following topics in this section:

- Setting Up Your System: Server-Side
  - *"Prerequisites" on page 16*
  - *"Installation" on page 17*
- Organizing Your Audio Media
  - *"Folder Management" on page 19*
  - *"Folder Example" on page 19*
  - *"Folder Names" on page 19*
- Optimize Audio Search for Better Processing

Setting Up Your System: Server-Side

Before you attempt to process your audio content, be sure that the system is licensed for audio processing and that the Audio Search module is installed with audio services running on your system. These prerequisites are necessary to later successfully pre-process, analyze, search and run analytics and reports on your audio content.

Prerequisites

Audio License

The eDiscovery Platform offers an Enterprise Audio Processing license which is a usage model based on the number of hours of audio content that has been processed. The system maintains an up-to-date inventory of the number of hours of audio content that has been consumed and the number of hours available.

To review general license information and see how much capacity you have used, see *"Managing Your License" in the -System Administration Guide*.

Note: The system does not charge for duplicate audio processed files that have the same language pack.

Antivirus Exclusions

The eDiscovery Platform installer preinstalls the Audio Search software components and a series of language packs and associated documentation (see *"Multiple Language Support" on page 13* and *"Appendix C: Language Support" on page 59*). By default, the Audio Search software is installed into the following directories and subdirectories. To avoid interference with critical media operations, be sure to disable virus and malware scanning software. In particular, Malwarebytes Anti-Malware, Kasperky Endpoint Security, and Microsoft Security Essentials are known to interfere with media operations. Make sure to exclude these directories from antivirus scans:

Audio Search Directories

Directory	Description
`C:\Program Files(86)\Nexidia`	Language Packs
`C:\Program Files(86)\Nexidia\Language Packs`	Language-specific Search documentation
`C:\Program Files\Nexidia\Search Grid 2.0`	Search Grid
`D:\Nexidia`	Search Grid data and logs
`C:\Users\<username>\AppData\Local\Temp`	Temporary folder for the account under which Search Grid services run

Firewall Configuration and TCP Port Usage

Make sure you configure any firewall software or other port filtering technology to allow incoming audio-related TCP connections on the ports listed in “Appendix D: TCP Port Usage” on this page.

Installation

Audio Search is deployed by the product installer. The *Installation Guide* (7.1.4 and later) covers these steps in detail. The following sections assume you have successfully run the installer to install the audio search component.

Note: Version 8.0 and later allows you to install audio services on a subnode. See Installation Guide.

Audio Search Services

After the installer installs the audio search components, the installation procedure creates (but does not start!) three Nexidia audio search grid services in the Services control panel. Before proceeding any further with the audio search setup, you must start these services.

Audio Search Services

Name	Service	Description
Nexidia Search Grid Agent Service	`EsaNxGridAgent`	Performs search and other CPU-intensive operations like phonetic index creation, classification, and language identification
Nexidia Search Grid Base Service	`EsaNxGridBase`	Manages data storage and communications for Nexidia Search Grid
Nexidia Search Grid Gateway Service	`EsaNxGridGateway`	Provides the public interface to Nexidia Search Grid

To start audio search services

When you are first starting audio services, use the start audio services command.

- **To start the audio services**

  Enter the following from a command prompt:

  - `b start-audio-services` (starts only the audio services)

- **To start all of the eDiscovery Platform services including audio**

  - `b start-services`

To stop and disable audio search services

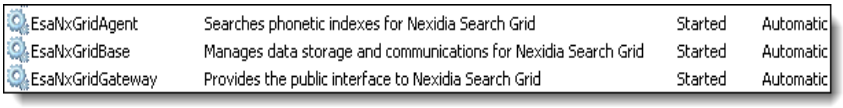
Use the stop command when audio processing and search is no longer needed.

- Stop audio search services from a command prompt:

  - `b stop-audio-services`

To check if audio search services are running

- If you see the three audio grid services running via the Windows Services control panel then you have successfully installed Audio Search.



Note: The procedure for accessing the Services control panel varies, depending upon the version of Microsoft Windows you are using. For example, on Windows Server 2008 R2, locate the Services control panel by selecting **Start > Control Panel > Administrative Tools > Services**.

For more audio search installation details, see *Installation Guide*.

Organizing Your Audio Media

Using folders for your audio sources, you can store your audio media in a helpful hierarchical manner for organizational and processing purposes.

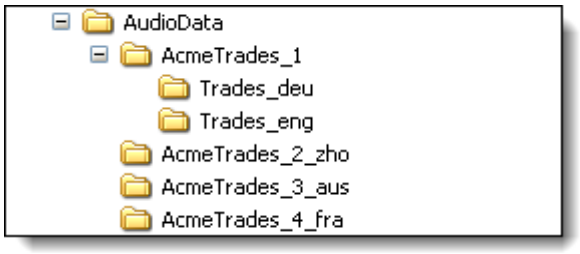
Folder Management

If your audio media sources are in different languages or if you have multiple language audio tracks in one source, you should create audio source folders based on the language of the recording. This folder setup allows you to efficiently process, iterate, and refine multi-language audio content.

Folder Example

Suppose you have a set of stock trading option audio media sources in Australian English, European French and Mandarin Chinese (Simplified Chinese) and there is one key audio source (Trades_1) that contains conversations in both North American English and German.

Your folder setup might look like this:



Example Processing & Search Notes

The audio source in the subfolders Trades_deu and Trades_eng is the same. This allows you to process once with the German language pack and then again with the North American English language pack (the system does not charge for duplicate audio processed files that have the same language pack). Once the audio content is processed, you can search in any of the languages (North American English and German).

Folder Names

When you have identified your media files, consider how and where your files may be used in a case. Naming your files simply and consistently makes it easier to identify media files, or duplicate them for multi-language processing.

Optimize Audio Search for Better Processing

Audio Search processing workloads can require significant system resources and time. We recommend that you do not run OCR document processing jobs or any other resource intensive operations during audio search processing in order to optimize resources for the audio search workload.

Note: The extent to which audio search affects system performance will depend on the size and composition of your audio content and the hardware resources available on the appliance.

Audio Search Processing

For information about how to process audio media files, refer to these steps:

- *"Step 1: Creating a Case" on page 22*
- *"Step 2: Specifying Default Audio Processing Language Pack" on page 22*
- *"Step 3: Adding Audio Media Sources" on page 23*
- *"Step 4: Enabling Audio Processing" on page 24*
- *"Step 5: Specifying Audio Processing Language Pack" on page 24*
- *"Step 6: Processing Audio Media Source Data" on page 25*
- *"Processing An Audio Media Source Data With Different Language Packs (Optional)" on page 25*
- *"Generating Audio Processing Reports" on page 26*

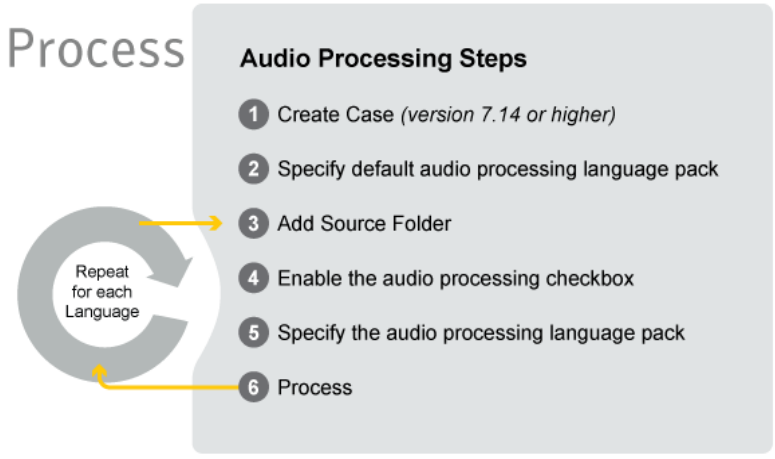
Audio Search Processing Workflow

Audio Search Processing assumes that you have completed the following prerequisites:

Setup

- Audio license installed
- Ensure audio processing services are running
- Organize audio files by language type
- Ensure no cpu-intensive (such as OCR) processing jobs are running

After setup, there are the 6 key steps in the Audio Search Processing workflow. The recommended audio search processing workflow is as follows:



Step 1: Creating a Case

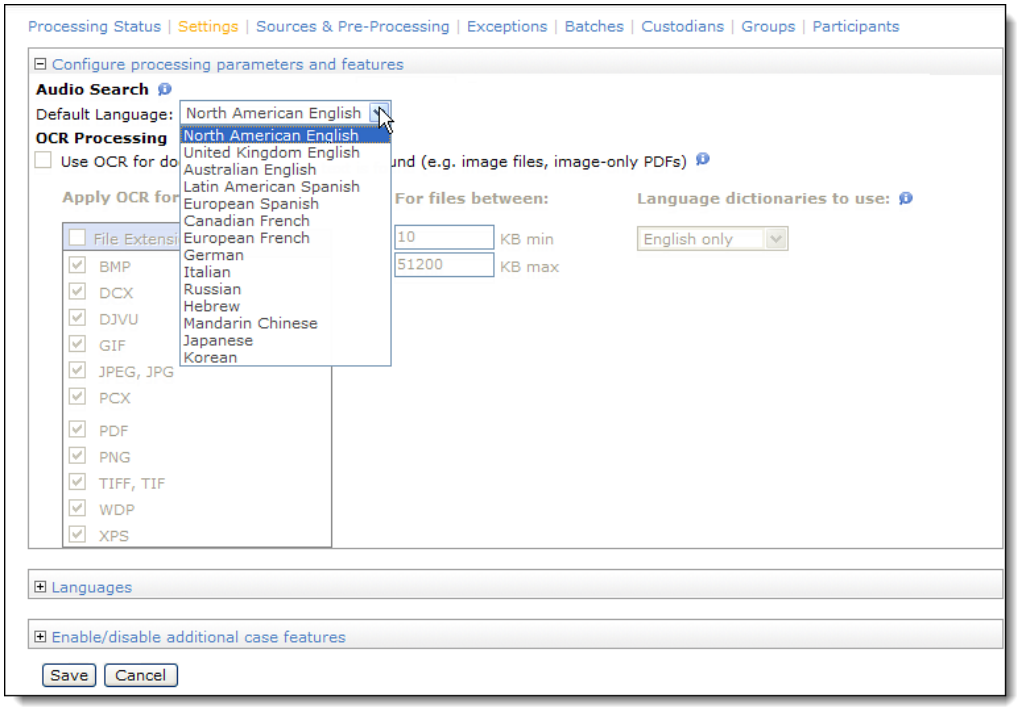
To create a new case

1. From the navigation bar, click **All Cases > New Case**. (Alternatively, from the drop-down menu, select **Create a new case**.)
2. Specify the new case information. For details, see *"Defining New Cases" in the Case Administration Guide*.
3. Click **Save & Set Up Processing**.

Step 2: Specifying Default Audio Processing Language Pack

To designate a default language pack

1. On the navigation bar, click **Processing > Settings** to display the new case.
2. Specify applicable non-audio case settings. For assistance, see *"Defining New Cases" in the Case Administration Guide*.
3. Under the **Configure processing parameters and features** menu, select the single **Default Language** to apply to your audio source from the drop-down list of available languages in the **Audio Search** section. If you need to change the language later on, you can override this case setting at the source level.



– The system default language is initially set to **North American English**

- If your system is not licensed for audio search, the language selections will not display.

4. Finalize your case setting selections and click **Save** to save the new source, or click **Cancel** to discard your changes.

Step 3: Adding Audio Media Sources

You can use the Add Case Folder Source screen to add audio media files to a case. Audio Search provides support for a wide range of audio formats. See *"Appendix B: Media File Types (Formats)" on page 55*.

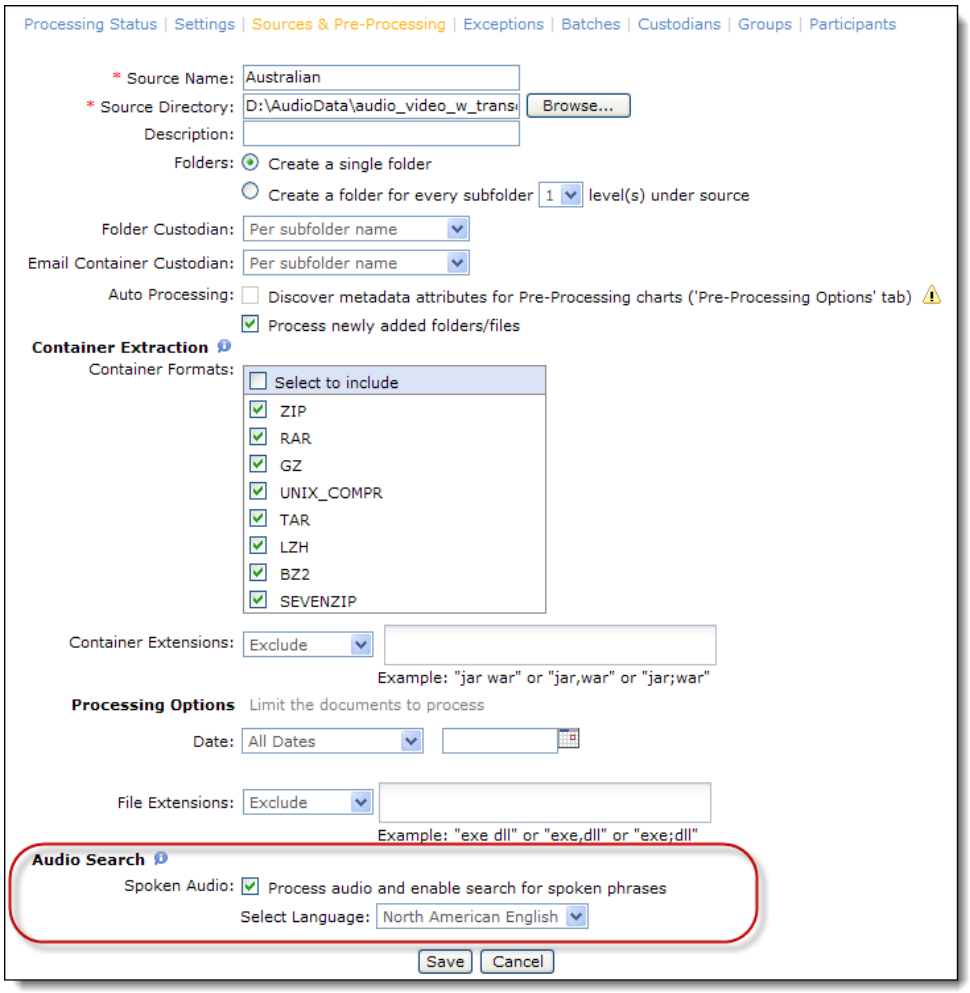
Note: If your audio media files contain multiple languages, you will need to create a separate physical copy of the files for each language and create a case folder for each copy. See *"Organizing Your Audio Media" on page 19*.

To add audio media sources to a case

1. On the top navigation bar, select your audio case, then click **Processing > Sources & Pre-Processing**.
2. On the Manage Sources screen, select **Add Case Folder Source** from the menu in the lower-left corner of the screen, and click **Go**.
3. Enter the settings information. Fields with an asterisk (*) are required.

Step 4: Enabling Audio Processing

4. Under the Audio Search section, select the check box to **Process audio and enable search for spoken phrases**.



Step 5: Specifying Audio Processing Language Pack

5. Select a language to apply to the audio source processing and indexing. The default is North American English but you can select any one of the fourteen language offerings.

6. Click **Save** to save the new source, or click **Cancel** to discard your changes.

Step 6: Processing Audio Media Source Data

After you have verified that your sources have been added correctly, you can process your audio source data.

To process your source data

1. On the top navigation bar, for a selected case, click **Processing > Sources & Pre-Processing**.
2. From the **For Selected Items** menu, select either **Start Processing Source without Discovery** or **Start Processing Source with Discovery**.
   - If no additional files have been added to the source since it was added, select Start Processing Source without Discovery.
   - If additional files have been added, Start Processing Source with Discovery must be used to discover the newly added files.
3. Click the **Go** button to start the selected task.

Processing An Audio Media Source Data With Different Language Packs (Optional)

If you have multiple languages spoken within the audio source, it is easy to add another language. You simply select a different language pack and process.

To iterate and process multi-language audio source

1. Copy the source folder and files.
2. Repeat Steps 3 through 6 for each audio language pack.

Generating Audio Processing Reports

The system provides a full audit of all the audio content and processing details of your multimedia files. The audio processing reports of interest are:

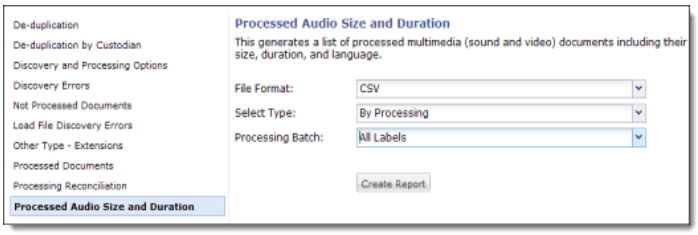
Audio Search Processing Reports

Report Name	Description
Discovery and Processing Options	Lists the Discovery and Processing options including the audio language pack selected for the processing batch and case folder.
Processed Audio Size and Duration	Lists processed multimedia (sound and video) files including their size, duration and language. Tells you how much audio data you have processed. • The report can be generated by source or by processing batch • The report is not visible if audio services are disabled or audio indexing license is not present • Batch audio reports can only be generated and accessed once processing for that batch completes There is a summary and detail report: **Summary Report** – The summary report aggregates the processing details by batch and case source folder **Detail Report** – Lists all audio data processed – All duplicate files will be displayed (irrespective of same language or not) – Loose files are listed one per line item – PST and NSF files appear as one line item, with a count of the total number of audio files contained inside – Audio files within a container file (like ZIP or RAR) are displayed as one line item for each audio file within the container file – Embedded files are rolled up into the loose file count

To generate audio processing reports

1. Under the **Processing** module for a selected case, click **Reports**.

   The Reports screen appears and lists available reports.



2. Choose one of the following report options:
   - **Discovery and Processing Options**
   - **Processed Audio Size and Duration**
3. Click **Create Report**.

   The report is generated and the job becomes available in the Jobs window for download.

Processed Audio Size and Duration Report Considerations

- The Processed Audio Size and Duration report generates a zip file containing a summary report and a detailed report.
- Report uses the date and time as part of the file name:
  - ZIP file:
    `<casename>_AudioSizeDuration_YYYYMMDDHHMMSS>`
  - Summary report:
    `<CaseName>_AudioSizeDuration_summary_YYYYMMDDHHMMSS_N`
  - Detail report:
    `<CaseName>_AudioSizeDuration_detail_YYYYMMDDHHMMSS_N`
- Report will roll over after one million records are written to the file.

Audio Search

This chapter describes the basic tasks you can perform to search audio media files to identify, prioritize, and review relevant data. Search tips are provided to help you locate the audio media information you need.

Refer to the following topics:

- *"Audio Search Workflow" on page 29*
  - *"Audio Search Steps" on page 30*
- *"Using Audio Search" on page 30*
  - *"Steps 1,2, & 3: Constructing the Query" on page 31*
  - *"Steps 4, 5, & 6: Previewing & Validating Media Results" on page 35*
  - *"Step 7: Adjusting & Tuning Results" on page 37*
  - *"Step 8: Applying Work Product (Tags, Notes and Folders)" on page 38*
- *"Creating the Audio Search Report" on page 39*
- *"Tips For Creating Good Audio Search Queries" on page 40*
  - *"Query Construction" on page 40*
  - *"Query Accuracy" on page 41*
- *"Exporting Audio Search Results" on page 43*

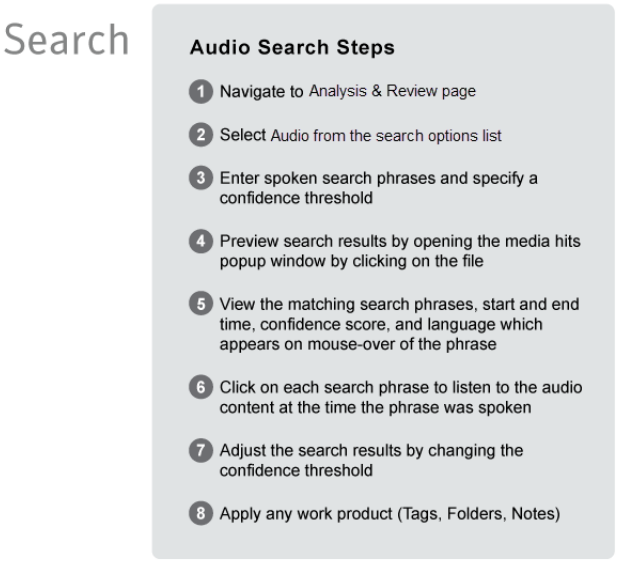
Audio Search Workflow

Audio search uses a probabilistic search model, which means that the returned results include ones that are *likely* to be relevant to the audio search phrase even though the search phrase may not be an exact match. This model is well-suited for audio content, as it allows a margin of error for noise in the recordings and variations in speech.

You should be aware that audio search requires a different, iterative work flow then the one you may be used to performing when locating *exact* text matches in documents. Typically, when performing an audio search you submit your audio search phrases, run the search, manually preview and validate the resulting sample of hits, and, if necessary, adjust the confidence threshold to ensure accurate results. Once you are satisfied with the results, you can apply tags, folder or make notes as you would in document review mode.

Audio Search Steps

The 8 key audio search steps are as follows:

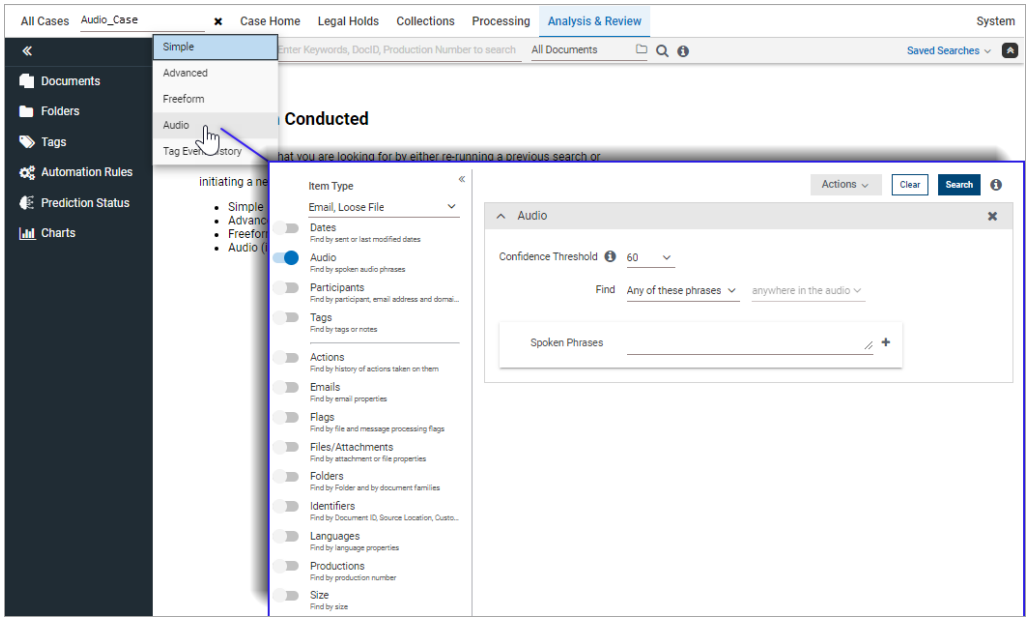


Using Audio Search

Using the Audio search options on the Analysis & Review page, audio search reviewers typing any combination of words or phrases can easily find and play any media clip where those words or phrases are spoken. The search results or hits from your media source file appear on the results page. Each result displays a confidence score, allowing you to focus on only the closest matches. It is possible to iterate through by adjusting the confidence threshold to further narrow or broaden the results. Once you arrive at the appropriate confidence setting and are satisfied with the results, you can apply tags (such as relevancy, privilege), add notes (annotations) and perform folder actions on the sound recordings.

Results can be quickly previewed in a media player without having to scroll through numerous clips in order to find a specific sound segment.

Steps 1,2, & 3: Constructing the Query

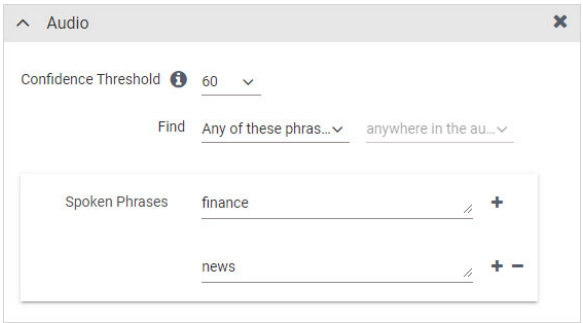


To submit an audio search query

The following audio search assumes that your case administrator has enabled audio search and has granted user access to play media files. The latter is accomplished by setting the permission option: "**Allow media streaming**" in Document access rights. See *"Defining User Roles" in the System Administration Guide*.

1. Select **Audio** in the search options list at the top of the **Analysis & Review** page.
2. By-default the search Style is set to **Audio**.

   The audio search options display with the 60 percent preset as the default Confidence Threshold setting.



3. Enter the following information:

Audio Search Criteria

Field	Description
Confidence Threshold	A number between 0 and 100 that indicates the confidence level that the hit is accurate. The confidence threshold will limit results to those with a confidence score above the specified threshold value. Reducing the threshold will help find relevant audio files that have low scores because of noisy backgrounds, dialects, or accents. Enter a value that best fits the balance between precision and recall. You may have to iterate through and preview the results in order to adjust this threshold to meet your review requirements and find the suitable trade-off between false positive and false negative results. For example, a threshold of zero may return everything while one hundred may return nothing. **Considerations:** Run audio search in background option **Scenario 1**: Setting a low confidence threshold, (such as 40), can impact search retrieval time. In this situation, the system gives the option of running the search in the background. This allows you to continue with other tasks while monitoring its progress. Try to avoid using a low threshold value since it will return a lot of false positive results. **Scenario 2**: If the case has more than 10,000 hours of media indexed then every search submitted will take a few minutes (if the threshold is set to a low value, it may take even more time). For cases with more than 10,000 hours of content, the system gives the option of running the search in the background.

Audio Search Criteria

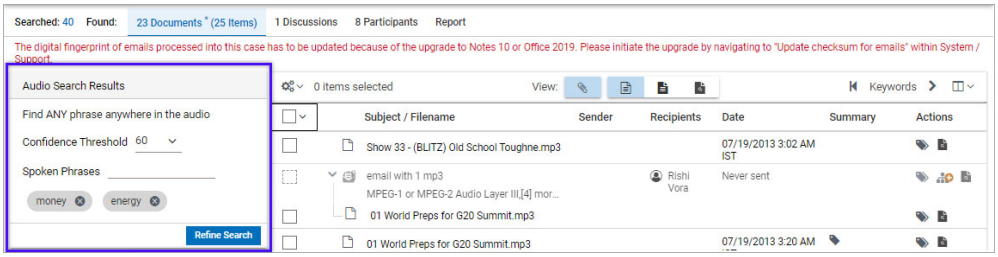
Spoken phrases	Enter one or more spoken phrases or the exact sequence of phonemes that you want to search. in the text box. You can add or remove query phrases by clicking or . **Considerations:** The phonetic engine uses your spelling to construct a phonemic representation. You can improve the accuracy of this representation in several ways: • **Important:** Spell things exactly as they sound. • For more accurate results, try to enter a phrase rather than a single word • If a search phrase contains punctuation characters or numeric characters, it is considered as invalid. Use your own pronunciation of the phrase. • Numbers must be spelled out – Incorrect: "`$200 rebate`" – Correct: "`two hundred dollar rebate`" • Searches are performed against all language packs that have been processed with the audio source. For example, if you have processed audio content with English and Japanese, then you can search in both languages. See the ANY and ALL multiple language considerations outlined below. • A phrase is considered as invalid if there are no audio documents indexed for the specific language. For example searching for a word in Japanese script in English audio indexed case is not allowed. • If a search phrase is not valid for a particular language then it is considered as invalid for that language pack • For searches that involve multiple phrases, enter each phrase on a separate line. Multi-phrase searches entered in a single line (even with quotes) are invalid. For more key search tips, see *"Tips For Creating Good Audio Search Queries" on page 40*.
Find ANY of these phrases	Finds **ANY** phrases from the drop-down list to match. Considerations: If there are multiple languages used for indexing then user can enter any valid phrase with "ANY" constraint. Example, If both Chinese and English indexed in one case, both "Thank you" and "ありがとう" can be used in search with "ANY" constraint.

Audio Search Criteria

Find All of these phrases within [numeric] sec/ min/hrs	Finds **ALL** phrases from the drop-down list that are found within a designated time range from each other. Considerations: If multiple languages are used, then to effectively use the "All" constraint, both words must belong to the same language pack. This means both "Thank you" and "ありがとう" cannot be used in "All" search.
Find ALL of these phrases anywhere in the audio	Finds **ALL** phrases from the drop-down list that match

4. Click **Search** to view your audio search results.

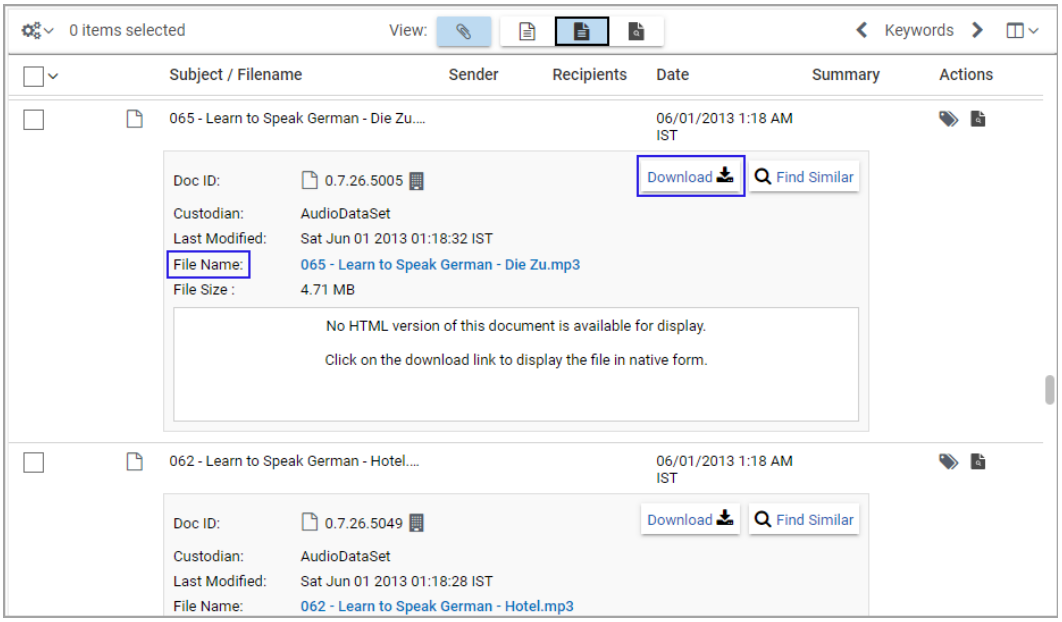
   The Audio Search Results box shows the Confidence threshold and lists the spoken search phrases.



Steps 4, 5, & 6: Previewing & Validating Media Results

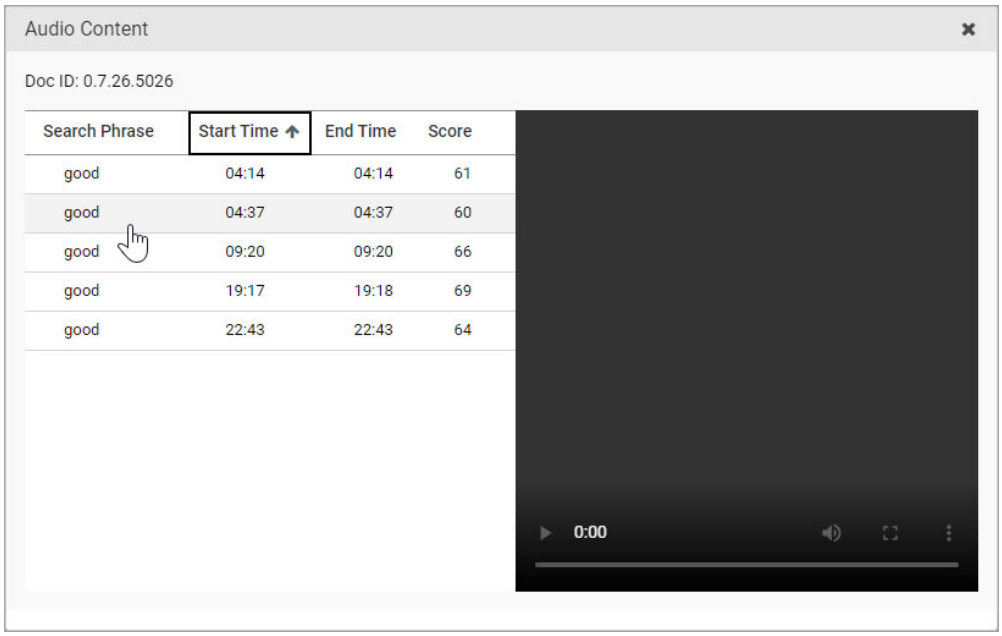
To preview search results for accuracy

1. Select a specific item or click **Details View** to drill down on the media results.



- **Filename** lists the media file.
- Download file can be useful when you need to play an unsupported file in an external media player. See *"Trouble Playing Media Files" on page 37*.
- The icon (which indicates that no indexed text or content was found for multimedia files), is only displayed for pre-7.1.4 cases that have been upgraded. 7.1.4 cases will not see this flag displayed.

2. From the Details View, click **Filename** to open Audio Content dialog box that shows list of media hits and a media player.



In this example, the user selects the search phrase "good" segment with a confidence score of 60 for the media player to play.

- On the left-hand side are the **Search Phrase** results, **Start Time** (beginning of the search phrase within the media segment), **End Time** (time offset for the end of the search phrase) and a **Score** that the search phrase occurs as indicated, between the start/end times.
- Select which audio/video track to play by clicking on the search phrase. All of the matches in each media file are ranked by a score that indicates how well the search term matches the specific audio segment.
- To find out the language of the matching phrase, hover the mouse pointer over the search phrase.
- On the right-hand side is the HTM5-based media player with intuitive buttons and controls making it easy to play and analyze tracks from your audio search results.

3. Click on the search phrase of interest to playback the audio/video segment.

Considerations:

Trouble Playing Media Files

If you have a media file that you cannot play with the system launched HTML5-based media player (assuming the file is not damaged), it may be because the media file format is unsupported or that media streaming is not enabled in the user role settings.

- Be aware that certain supported media files may not be playable by the media player. In this situation, a list of hits will display but clicking on the hit does not result in the content being played by the media player. For a list of media files that the HTML5-based media player supports, *"Supported HTML5-Based Media Player Media Formats" on page 56*.
- If you cannot play a media file through the system launched media player, you may want to try downloading it for play by another, external media player. You can do this by clicking the ⋮ icon on the media player and selecting download. The file will download and can be played by another media player. One example of this approach would be to download an unsupported RealAudio media file (.RM) for play in RealPlayer.
- Verify that the permission "Allow media streaming" is enabled. This permission must be enabled in order to play audio files in the media player. For details, see *"Defining User Roles" System Administration Guide*.

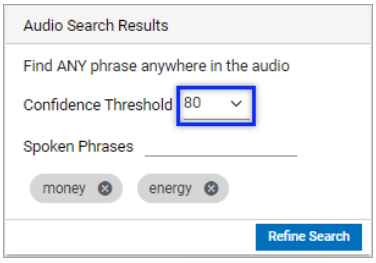
Step 7: Adjusting & Tuning Results

Based on your findings in Step 6, you may need to refine the confidence threshold to a higher level (include less results) or to a lower level (include more results).

Note: Audio discovery requirements, which tend to be more inclusive and err on over-inclusion (higher recall), may have this level set lower than other types of end-user searches.

To refine audio search results

1. To refine results, in the Audio Search Results box, adjust the **Confidence Threshold**.



2. Click **Refine Search** to display results based on the new threshold.
3. Repeat Steps 4 through 6 as necessary.

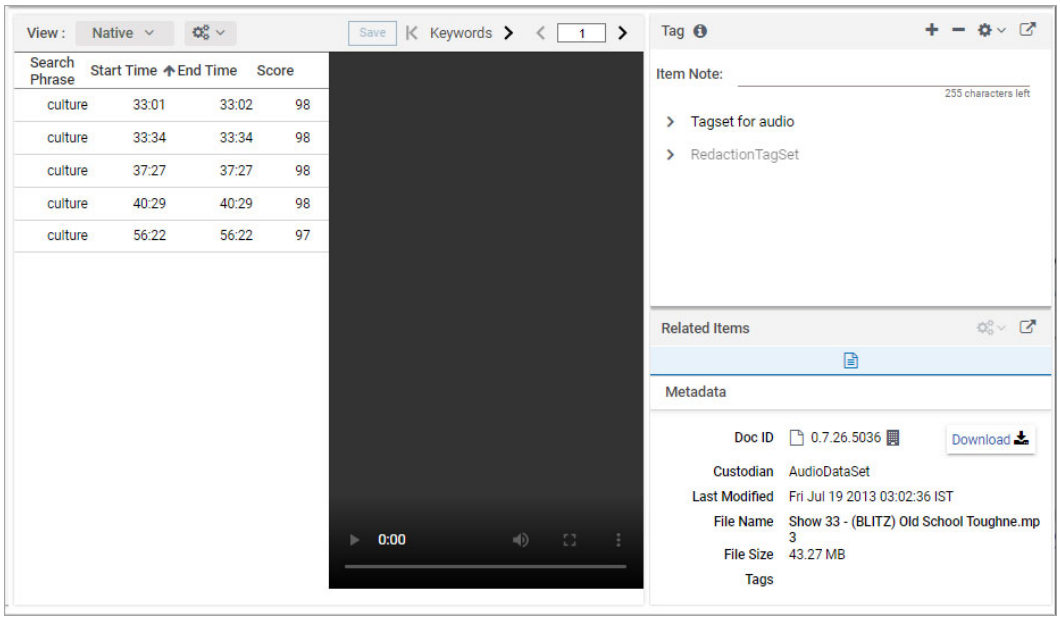
Step 8: Applying Work Product (Tags, Notes and Folders)

In review mode, you can apply tags to audio content, add an item note to explain tagging decisions and assign audio search results to specific folders. These are familiar tasks for reviewers who are used to analyzing, culling, and preparing data for relevance, responsiveness, privilege.

When preparing audio content for review, you may want to consider audio quality when determining what to put in review folders. For example, you may want to separate poor audio quality recordings from higher quality recordings. See *"Query Accuracy" on page 41*

For more details on how to perform these tasks, refer to the *User Guide*.

Review Mode



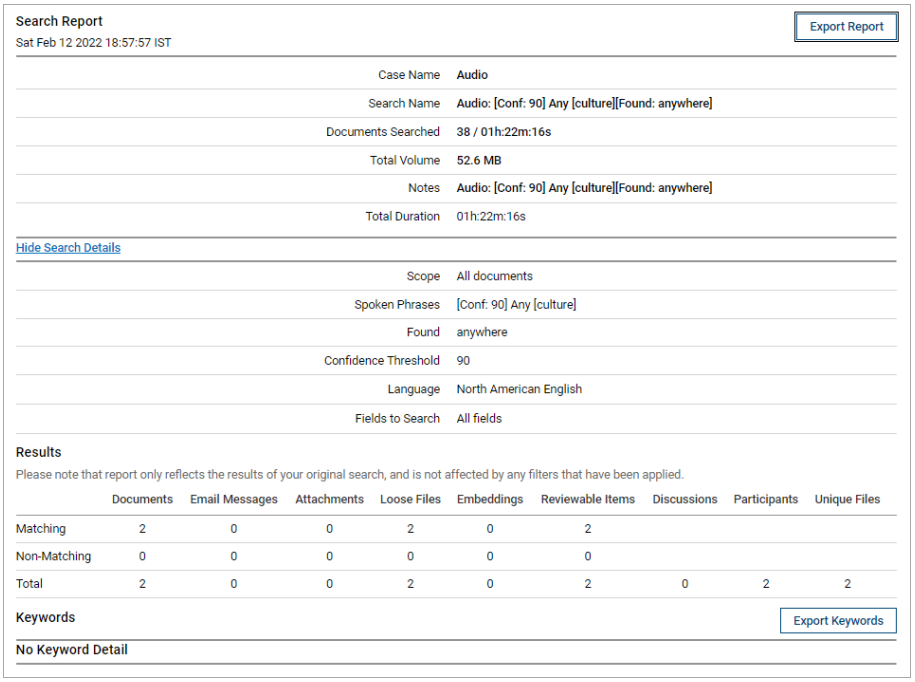
Creating the Audio Search Report

The audio search report provides information on the specific audio search criteria, counts and results of a search. For more details, see *"Using the Search Reports Screen" Veritas eDiscovery Platform User's Guide*.

To view the Audio Search Results Report

1. Click **Report** under the **Analysis & Review** module.

   The Search Report displays.



Search Report
Sat Feb 12 2022 18:57:57 IST

Export Report

Case Name	Audio
Search Name	Audio: [Conf: 90] Any [culture][Found: anywhere]
Documents Searched	38 / 01h:22m:16s
Total Volume	52.6 MB
Notes	Audio: [Conf: 90] Any [culture][Found: anywhere]
Total Duration	01h:22m:16s

Hide Search Details

Scope	All documents
Spoken Phrases	[Conf: 90] Any [culture]
Found	anywhere
Confidence Threshold	90
Language	North American English
Fields to Search	All fields

Results

Please note that report only reflects the results of your original search, and is not affected by any filters that have been applied.

	Documents	Email Messages	Attachments	Loose Files	Embeddings	Reviewable Items	Discussions	Participants	Unique Files
Matching	2	0	0	2	0	2			
Non-Matching	0	0	0	0	0	0			
Total	2	0	0	2	0	2	0	2	2

Keywords

Export Keywords

No Keyword Detail

   - **Documents Searched** lists the length of the media searched
   - **Total Duration** lists the total length of the media matches found

2. View the keyword search results information.
3. To export the report, click **Export Report**.

Tips For Creating Good Audio Search Queries

Query Construction

The most useful strategy for creating a good audio search query is to enter phrases that reflect how people pronounce or say a term instead of how that term is written.

The following tips may prove useful for handling numbers, punctuation, acronyms, and other special cases:

Spell it the way it sounds

If the spelling is tricky or unusual, ignore the proper spelling of the word and spell it the way it sounds. You can do this in two ways:

- Enter the letters that represent the sounds in the most straightforward way possible (the way it sounds). For example, “four hours” can locate an audio segment as well as “fore ours”. This is because both resolve to the same string of phonemes. Similarly, “my sequel” is better than “MySQL”.
- Build a long, complex word from a series of short, simple ones that have unambiguous pronunciations, separated by hyphens. For example, spell disbandment as diss-band-mint.

Try variations

If you cannot find *Chicago airport*, can you find *O’Hare*? Try to imagine other ways someone might have phrased the same idea.

Omit punctuation

The phonetic engine ignores commas, periods, colons —everything (except an ampersand) that is not a letter or a number. (The other exception—a special use of the underscore—is described below.)

Omit capitalization

The phonetic engine ignores case. Search accuracy is the same whether you enter “South Pacific” or “south pacific”.

Spell out numbers

There are many ways to pronounce numbers, so the system does not make any assumptions for the user. Instead, all numeric queries must be spelled out.

Every number with more than two digits can be pronounced in at least two ways: 23, for example, can be pronounced “twenty-three” or “two three”; 1000 can be pronounced “a thousand”, “one thousand”, “one zero zero zero” or even “one oh oh oh”. The pronunciation file specifies most common pronunciations for the numbers it includes, but someone may have pronounced that number in an uncommon way. If you know what someone said, spell it out that way.

Put spaces between the letters of acronyms spoken as acronyms

Many acronyms can be pronounced: WHO, for example, could stand for the World Health Organization or the word "who". If you think the speaker said "double-you aitch oh", enter: W H O. Unless you put a space between each pair of letters, the phonetic engine assumes you are writing a word, not an acronym, and applies its usual pronunciation rules.

In the case of one-letter words, even that might not be enough. If an acronym includes the letter "A" for example, the phonetic engine will look for the usual pronunciation of the indefinite article—more like "uh" as in "Have a drink". In such a case, enter the phonemes explicitly – type "_ey" to tell the phonetic engine that you're looking for the sound in "hey" or "day."

Spell out abbreviations

People do not say etc., whether as "ets" or "etk". They say "et cetera".

Spell it in phonemes (Advanced)

When in doubt, spell it out phonetically. Try specifying the exact sequence of phonemes in the phrase you want to locate. The search engine will bypass its own pronunciation rules to translate your query and will use the representation you supply.

To tell the phonetic engine to interpret your input as phonemes, enter an underscore before each phoneme. For example, you can specify the first phoneme in the word "pop" like this: _p.

The underscore is a signal to the phonetic engine to accept the input immediately following as literal phonemes—until it encounters a space. Then, if it doesn't find another underscore, it returns to its usual operation. For example, to specify the four- phoneme phrase "payday", enter:
_p _ey _d _ey

To construct queries directly into phonemes, refer to the list of phonemes that the phonetic engine recognizes. See *"Appendix A: Phonemes" on page 47*.

Note: The same phrase may not be applicable to all language packs. For example, using the same search term may be valid for English, but is invalid for Japanese.

Query Accuracy

These three factors affect the accuracy and thoroughness of your query results:

- Media quality
- Query length
- Exotic or unusual spelling

Audio quality affects every query conducted on that audio. If the sound is poor (scratchy, distorted, or full of background noise), the phonetic engine will be less successful in finding accurate results.

If you are searching through many files at once, it is best if they are all of more or less the same quality. If one is notably poorer than the others, results found in that file will have a lower confidence score than the others, though they will be perfectly valid, while some higher-ranking results from better quality files may be false alarms.

Query length affects accuracy as well. Longer queries are more accurate, up to a point. Just as with text-based search engines, a longer, more specific query such as *Presidential Election* yields better results than a short, vague one such as Election because it gives the phonetic engine more to go on.

Even if what someone actually said was *Presidential, uh, Election*, there is a good chance of finding it, because the phonetic engine presents close matches as well as perfect ones. Queries of ten syllables or more are generally less affected by these small variations. If you're sure that someone said a particular phrase, it's usually a good idea to search for the whole phrase, unless it's extremely long.

To help you gauge query length, the table below provides phoneme counts for sample queries:

Query Length in Number of Phonemes

Query	Number of Phonemes
mob, rear, bought, loose, cake, jet	3
crowded, withdraw, precious, ownership	6
save you money, someone says, dromedary	9
the new standards, took away our rights, Washington today	12
stock market quotes, maximum strength, astounding profits	15

Exporting Audio Search Results

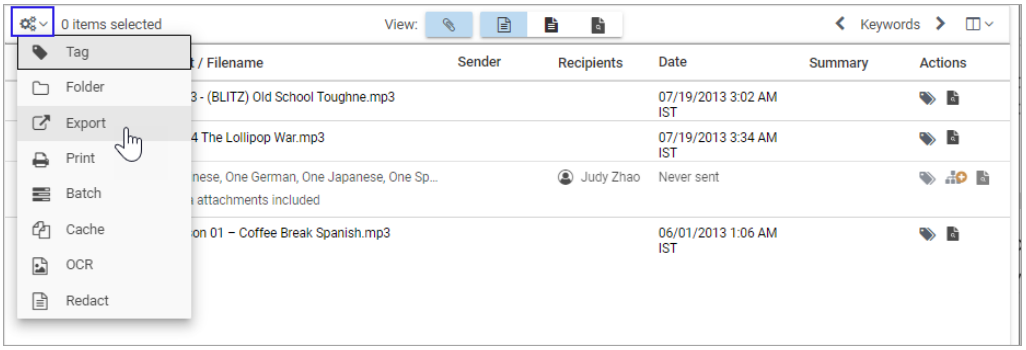
Once you have identified your audio search results, you easily export the contents of the search results pane to a CSV file for further analysis or for third-party tool processing.

Audio Search Export Considerations

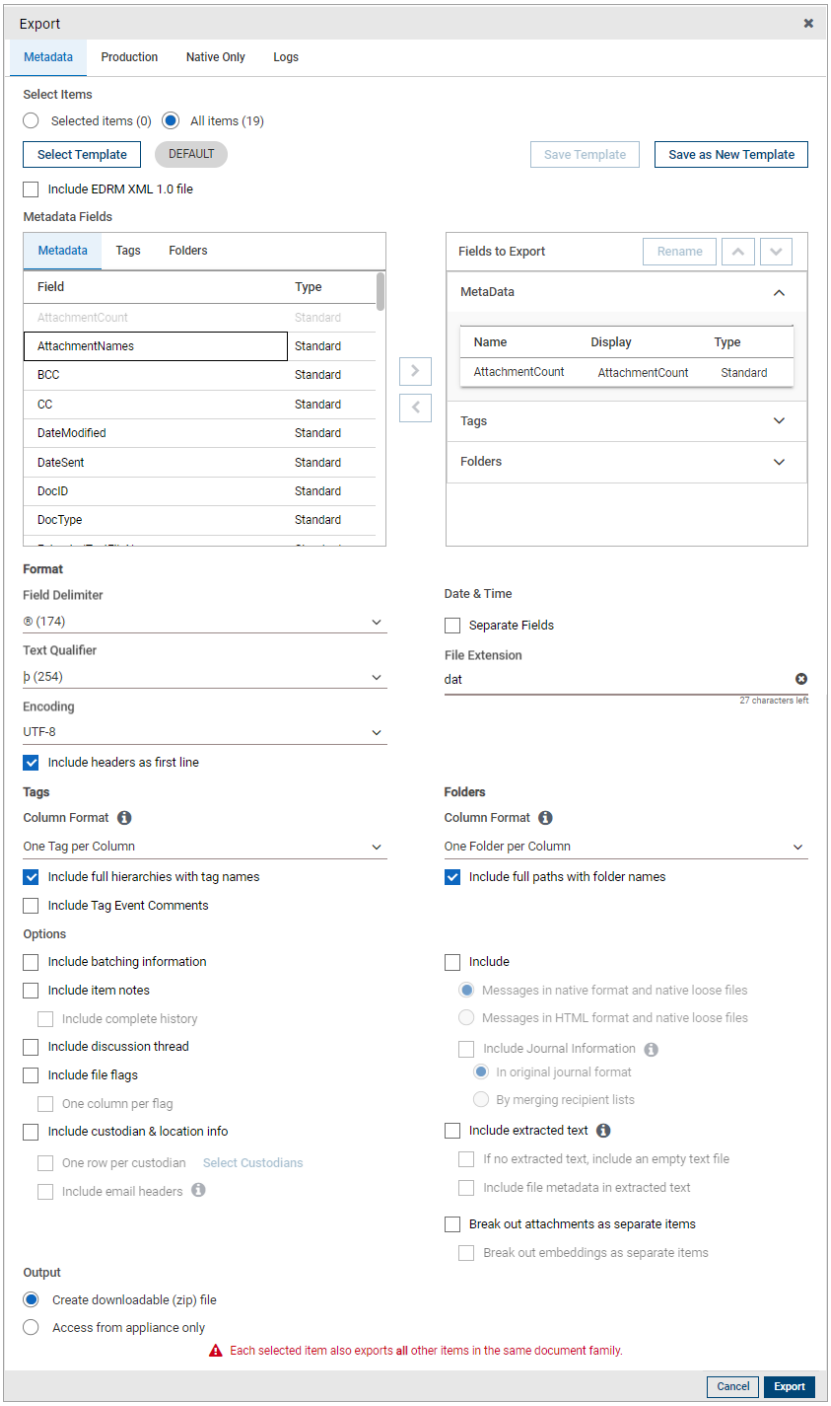
- Exports of audio search results are performed from the search results when you click **Export** from the **Bulk actions menu** icon.
- The export CSV file matches the delimiters and extensions that you chose in your search.
- Export works on a document family level and not on an item level. When an error is encountered while trying to write audio hits for a specific item of a document family, the entire document family is marked as failed. You can retry the export for the failed document.
- The export list is sorted by Doc ID and the hits are sorted by their associated score.
- Audio search export files (audio.csv) along with other export files are only created if the current search is an audio search. For non-audio searches, even though there may be audio processed items present, the system does not export any audio search (audio.csv) files.

To export audio search data

1. Submit an audio search to find the audio segments you want to export. See *"Using Audio Search" on page 30*.
2. From the search results, you can export all the audio content in your search results set, or select one or more audio hits on the current screen by clicking the check box next to the appropriate audio content to be exported. To select all the audio content on the current screen, click the check box in the column heading.
3. Above the search results screen, click **Bulk actions menu** icon **> Export** to open the Export options window.



The export options on the **Metadata** tab:



4. Click **Selected Items** or **All Items** (for the number of documents selected on the current screen, or the total number of documents).

5. Click **Export** to open Export Documents dialog box.

6. Choose an option:

   A. If you want to create a batch, to include a description, enter a label for the batch (optional), click **Create an export batch**.

   B. If you do not want a batch created, click **Do not create an export batch**.

   – To create the export batch, click **Continue to Export**.

7. An export job is launched and you can monitor its progress from the Jobs window.

8. Click on the audio.csv file to view the Audio Search Export.

   The following is a summary of the export fields

Audio Search Export

Field	Description
Searchable DOC ID	Unique number identifying the audio search source
Phrase	Search phrase
Language	Language pack
Start Time	Start of the audio content
End Time	End of the audio content
Score	Confidence threshold score

Appendix A: Phonemes

North American English

This table lists the phonemes of North American English. To assist in translating queries into phonemes, there is a sample word for each phoneme and its corresponding phonetic transcription.

Some of the phonemes have a parenthesis in the phoneme name. This notation specifies whether the phoneme only occurs in the beginning of the syllable (before the vowel) or at the end of the syllable. For example, the phoneme _(t can only occur in the beginning part of the syllable, like in the words **t**alk or s**t**op. The counterpart _t) only occurs at syllable endings, like in the words lis**t** and lo**t**s.

Note: The phonetic engine is case-sensitive for any queries where you enter phonemes (instead of the native language spelling) for the phrase or word. When entering *phoneme* search queries, be sure to use correct upper and lower case.

.

North American English Phoneme Chart

Phoneme	Letter	Example (Phonemes)	Example
STOPS			
p	p	_p _ah _f	puff
	pp	_hh _ae _p _iy	happy
(t	t	_(t _ah _f	tough
	tt	ah _(t _ey _n)	attain
t)	d	_ae _ng _g _w _ih _sh _t)	anguished
	t	_(l _ih _s _t)	list
	tt	_b _eh _t)	Bette
k	c	_k _ah _f	cuff
	cc	ah _k _aw _n) _t)	account
	ck	_(l _aa _k	lock
	k	_k _iy	key
	lk	_f _ow _k	folk
	q	_(l _ih _k _w _ih _d	liquid
	x	_s _ih _k _s	six
b	b	_b _ah _f	buff
	bb	_r _ae _b _ih _t)	rabbit

North American English Phoneme Chart

Phoneme	Letter	Example (Phonemes)	Example
d	d	_d _ah _f	duff
	dd	_aa _d	odd
g	g	g _ah _f	guff
	gg	_s _(t _ae _g _er	stagger
	gh	_g _ow _s _t)	ghost
FLAP/TAP			
d\	t	_r _ay _d\ _er	writer
	tt	_b _ih _d\ _er	bitter
	d	_r _ay _d\ _er	rider
	dd	_b _ih _d\ _er	bidder
FRICATIVE			
s	c	_(l _eh _g _ah _s _iy	legacy
	cc	_eh _k _s _eh _p _t)	accept
	ps	_s _ay _k _aa _(l _ah _jh _iy	psychology
	s	_s _ih _n)	sin
	sc	_s _eh _n) _t)	scent
	ss	_p _r _aa _g _r _eh _s	progress
	x	_s _ih _k _s	six
sh	c	_s _p _eh _sh _ah _l)	special
	ch	_m _ah _sh _iy _n)	machine
	s	_sh _uh _g _er	sugar
	sh	_sh _ih _n)	shin
	ss	_p _r _eh _sh _er	pressure
z	s	_y _uw _z _er	user
	ss	_s _ih _z _er _z	scissors
	x	_z _iy _r _aa _k _s	Xerox
	z	_z _ih _ng	zing
	zz	_f _ah _z	fuzz
zh	s	_ey _zh _ah	Asia
	z	_ah _zh _uh _r	azure

North American English Phoneme Chart

Phoneme	Letter	Example (Phonemes)	Example
f	f	_f _ih _n)	fin
	ff	_ao _f _er	offer
	gh	_ih _(n _ah _f	enough
	lf	_k _ae _f	calf
	ph	_f _ow _n)	phone
v	v	_v _ey _k _ey _sh _n)	vacation
	f	_ah _v	of
th	th	_th _ih _n)	thin
dh	th	_dh _ih _s	this
AFFRICATIVES			
ch	cc	_k _ae _p _ah _ch _iy _(n _ow	cappuccino
	ch	_ch _ey _n)	chain
	t	_(n _ey _ch _er	nature
	tch	_b _ae _ch	batch
jh	dg	_d _aa _jh	dodge
	g	_jh _eh _l)	gel
	j	_jh _ey _n)	Jane
NASALS			
m	m	_b _ae _m	bam
	mb	_(n _ah _m	numb
	mm	_hh _ae _m _er	hammer
	mn	_ao _d\ _ah _m	autumn
(n	gn	_(n _ao	gnaw
	kn	_(n _ow	know
	n	_(n _ow _z	nose
	nn	_m _ae _(n _er	manner
	mn	_(n _eh _m _aa _(n _ih _k	mnemonic
n)	gn	_d _ih _z _ay _n)	design
	n	_b _ae _n)	ban
	nn	_ae _n)	Anne
	on	_p _ah _z _ih _sh _n)	position
	en	_h _ih _d _n)	hidden

North American English Phoneme Chart

Phoneme	Letter	Example (Phonemes)	Example
ng	n	_b _ae _ng _k	bank
	ng	_b _ae _ng	bang
SEMIVOWELS			
(l	l	_(l _aa _t)	lot
	ll	_ae _k _sh _ah _(l _iy	actually
l)	l	_m _ey _l)	mail
	ll	_ao _l)	all
r	r	_r _aa _t)	rot
	rh	_r _ow _d _pau _ay _(l _ah _n) _d	Rhode Island
	rr	_ae _r _ow	arrow
w	o	_k _w _ay _r	choir
	u	_iy _k _w _ah _l)	equal
	w	_w _aa _t)	watt
	wh	_w _ah _t)	what
y	i	_ah _(n _y _ah _n)	onion
	ia	_f _ah _m _ih _l) _y _er	familiar
	u	_y _uw _z	use
	y	_y _aa _t)	yacht
hh	h	_hh _aa _t)	hot
	wh	_hh _uw	who
VOWELS			
ae	a	_s _ae _t)	sat
	ai	_p _(l _ae _d	plaid
	au	_(l _ae _f	laugh
aa	a	_r _ih _g _aa _r _d	regard
	e	_aa _n) _(t _r _ey	entree
	o	_s _aa _t)	sot
	u	_s _er _k _aa _m _f _r _ih _n) _s	circumference

North American English Phoneme Chart

Phoneme	Letter	Example (Phonemes)	Example
ao	a	_f _ao _l)	fall
	al	_w _ao _k	walk
	au	_ao _th _er	author
	aw	_s _ao	saw
	o	_s _ao _l) _v	solve
	oa	_b _r _ao _d	broad
	ou	_k _ao _f	cough
aw	ou	_r _iy _b _aw _n) _d	rebound
	ow	_aw _l)	owl
ay	ey	_g _ay _z _er	geyser
	eye	_ay	eye
	i	_s _ay	sigh
	ie	_(l _ay	lie
	ui	_g _ay _d	guide
	uy	_b _ay	buy
	y	_k _(l _ae _s _ih _f _ay	classify
	ye	_d _ay	dye
eh	a	_m _eh _(n _iy	many
	ai	_s _eh _d	said
	e	_s _eh _t)	set
	ea	_b _r _eh _d	bread
	ie	_f _r _eh _n) _d	friend
	ue	_g _eh _s	guess
er	ar	_g _r _ae _m _er	grammar
	ear	_er _th	earth
	er	_b _ow _l) _d _er	boulder
	eur	_sh _ow _f _er	chauffeur
	ir	_s _er	sir
	or	_w _er _s _t)	worst
	ur	_b _er _n)	burn
	yr	_m _er _d\ _ah _l)	myrtle

North American English Phoneme Chart

Phoneme	Letter	Example (Phonemes)	Example
ey	a	_p _ey _jh	page
	ai	_r _ey _d	raid
	ay	_s _ey	say
	ea	_g _r _ey _t)	great
ih	e	_p _r _ih _d\ _iy	pretty
	ee	_b _ih _n)	been
	i	_s _ih _t)	sit
	u	_b _ih _z _iy	busy
	ui	_b _ih _l) _d	build
	y	_m _ih _th	myth
iy	e	_s _iy _d	cede
	ea	_m _iy _(n _ih _ng	meaning
	ee	_s _iy	see
	eo	_p _iy _p _ah _l)	people
	ey	_k _iy	key
	i	_m _ah _sh _iy _n)	machine
	ie	_f _iy _l) _d	field
	y	_p _ae _n) _(t _r _iy	pantry
ow	ew	_s _ow	sew
	o	_m _ow _d	mode
	oa	_r _ow _d	road
	oe	_f _ow	foe
	oo	_d _ow _r	door
	ou	_s _ow _l)	soul
	ow	_(n _ae _r _ow	narrow
oy	oi	_p _oy _n) _t)	point
	oy	_s _oy	soy
uh	o	_w _uh _l) _f	wolf
	ou	_sh _uh _d	should
	oo	_s _uh _t)	soot
	u	_p _uh _t	put

North American English Phoneme Chart

Phoneme	Letter	Example (Phonemes)	Example
uw	ew	_k _r _uw	crew
	oe	_k _ah _(n _uw	canoe
	oo	_(t _uw	too
	ou	_s _uw _p	soup
	u	_k _r _uw _d	crude
	ui	_s _uw _t)	suit
	wo	_(t _uw	two
ah	a	_ah _m _eh _r _ih _k _ah	America
	e	_ae _n) _th _ah _m	anthem
	o	_s _ah _m _th _ih _ng	something
	u	_s _ah _n)	sun

Appendix B: Media File Types (Formats)

A compendium of supported and unsupported media formats.

- *Supported Processing & Search Media Formats*
- *Unsupported Processing and Search Media Formats*
- *Supported HTML5-Based Media Player Media Formats*

Supported Processing & Search Media Formats

Audio Search Processing and Search supports the following file formats.

Supported Media Formats

File Format	File Extension
AAC	aac
ADTS	
AIFF	.aiff, .aif, .aifc
AMR NB	.amr
ASF	.asf,.wma, .wmv
AU	.au,.snd
AVCHD	.mts, .m2ts
AVI	.avi
BWF	.wav
F4V	.f4v, .f4p, .f4a, .f4b
GXF	.gxf
M4A	.mp4, .m4a, .m4p, .m4p, .m4b, .m4r
M4V	.m4v
MOV	.mov, .qt
MP2	.mp2
MP3	.mp3
MPEG	.mpg, .mpeg, .m2v
MXF	.mxf
NMF	.nmf
OGG	.ogg, .ogv, .oga, .ogx, .spx, .opus
OMF	.omf
PCM	.pcm

Supported Media Formats

File Format	File Extension
RM	.rm
SWF	.swf
WAV	.wave, .wav
WMV	.wmv

Unsupported Processing and Search Media Formats

These four file formats are supported by the Nexidia but are not supported by the platform's processing and search modes.

Unsupported Media Formats for Processing and Search

File Format
3GPP
3GPP2
AES3-331
FLV

Supported HTML5-Based Media Player Media Formats

The HTML5-based media player does not support the playback of every type of media file.

This is a list of the supported video media files.

Supported Video Media Formats for HTML5-based Media Player

Browser	MP4	WebM	OGG
Microsoft Edge	Yes	No	No
Google Chrome	Yes	Yes	Yes
Apple Safari	Yes	No	Yes

This is a list of the supported audio media files..

Supported Audio Media Formats for HTML5-based Media Player

Browser	MP3	WAV	OGG
Microsoft Edge	Yes	No	No

Supported Audio Media Formats for HTML5-based Media Player

Browser	MP3	WAV	OGG
Google Chrome	Yes	Yes	Yes
Apple Safari	Yes	No	No

Note: If you cannot play a media file through the launched HTML5-based media player, you may want to try downloading it for play by an external media player. See *"Trouble Playing Media Files" on page 37*.

Appendix C: Language Support

This section lists the current languages and language packs that the eDiscovery Platform supports. A helpful set of documents (PDFs) for each of the supported languages is automatically installed during the installation process. You can find these language-specific guides in the following directory:

```
C:\Program Files(86)\Nexidia\Language Packs\<language>
```

Supported Language Packs

Language
Australian English
North American English (default language)
United Kingdom English
Canadian French
European French
Castilian Spanish
Latin American Spanish
German
Hebrew
Italian
Japanese
Korean
Mandarin Chinese (Simplified)
Russian

Appendix D: TCP Port Usage

This section lists default port assignments that support firewall configurations. Grid servers must allow incoming connections on the ports listed below:

Default TCP Ports

Component	Port #
esa.firewall.port.nexidiapublic.desc=**Nexidia Search Grid Gateway Public Port** esa.firewall.port.nexidiapublic.port=25002	25002
esa.firewall.port.nexidiamsgbrkr.desc=**Nexidia Search Grid Message Broker Port** esa.firewall.port.nexidiamsgbrkr.port=25100	25100
esa.firewall.port.nexidiadatabase.desc=**Nexidia Search Grid Gateway Database Port** esa.firewall.port.nexidiadatabase.port=25101	25101
esa.firewall.port.nexidiagtwyhttp.desc=**Nexidia Search Grid Gateway HTTP Port** esa.firewall.port.nexidiagtwyhttp.port=25102	25102
esa.firewall.port.nexidiabasehttp.desc=**Nexidia Search Grid Base HTTP Port** esa.firewall.port.nexidiabasehttp.port=25122	25122

Appendix E: Scaling Audio Search & Processing (Nexidia)

This section describes the steps for distributing the audio processing and search workload across multiple systems such as utility nodes. These instructions follow the scalable audio search infrastructure and clustering mechanism of the Nexidia Search Grid application.

The Nexidia Search Grid application has its own mechanism for clustering which differs and should not be confused with the Distributed Architecture techniques and guidelines for the eDiscovery Platform.

The high-level steps are:

- *Scale Audio Processing*
  - *Install Compute Node on Utility Node (cw-util)*
  - *Modify Properties on Appliance (cw-appl)*
  - *Restart the Services*
  - *Removing Compute Node*
- *Scale Audio Search*
  - *Assess Pros and Cons of Adding a Data Node for Audio Search IMPORTANT!*
  - *Install Data Node on Utility Node (cw-util)*
  - *Modify Properties on Appliance (cw-appl)*
  - *Redistribute Data to the New Data Node on the Utility Node (cw-util)*
  - *Redistribute Data From a Data Node Back to Appliance (cw-appl)*

Scale Audio Processing

When the system is processing a batch of audio files, the CPU load can be quite heavy since processing is a CPU bound operation. The Search Grid installation on each node is responsible for processing the audio of all cases that are resident on that node. This may lead to CPU contention issues especially when multiple cases are processing audio at the same time, or when a single case is processing a large amount of audio data.

Search Grid provides a mechanism to scale audio processing by farming out the processing work to one or more additional systems such as utility nodes. This document describes the steps to set up and configure such additional systems.

Note: These directions are specific to Search Grid and the audio processing and search workload. Moreover, the eDiscovery Platform Distributed Architecture techniques are not relevant or applicable to audio processing and audio search node configurations.

At a high level, Search Grid consists of the following logical components:

- Gateway Node - used for all client interactions
- Data Node - stores, organizes, and searches phonetic indexes. Data Nodes also perform CPU-intensive operations like phonetic index creation

- Compute Node - perform CPU-intensive operations like phonetic index creation, classification, and language identification. Unlike Data Nodes, it does not store data locally.

Each eDiscovery installation has one Gateway Node and a Data Node. To facilitate audio processing scaling, you must install a standalone compute node on a separate system and configure it to work with a single eDiscovery installation.

Note: A Compute Node can be configured to work with only one Gateway Node at a time and can be configured to point to a single Search Grid installation. The Compute Node essentially pulls the work from the Search Grid installation that it is points to.

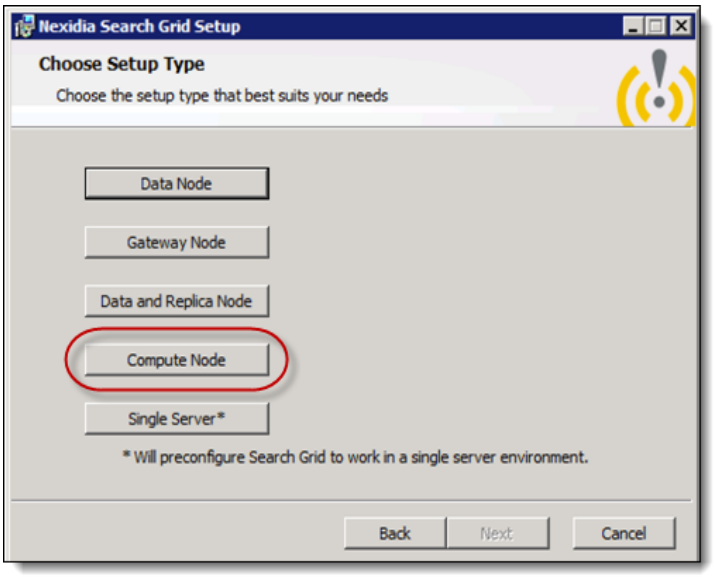
The following steps are required for scaling audio indexing before starting the processing operation. For the purposes of these instructions and explanation, assume that the eDiscovery appliance is installed on `cw-appl` and there is a utility node called `cw-util` which will be used to install a Compute Node for audio processing scale out.

Install Compute Node on Utility Node (cw-util)

On the utility node `cw-util`, install the Compute Node

Compute Node Installation Considerations

1. Run the Search Grid installer and choose the Compute Node setup type.



   Note: The Search Grid installer is located at:
   `<CW_Installer>\packages\Nexidia\NexidiaSearchGrid-2.0.2.2`

2. Choose the same service credential (user name and password) that were used for the *ESAApplicationService* on the appliance (`cw-appl`). These service credentials must have access rights to read the audio files that need to be processed - similar to the credential of the Agent Service on the appliance (`cw-appl`).
   - Note that you can provide a different data directory during the installation.

– Data directory location - this will contain the log files that will be needed for any troubleshooting.

3. Copy the license file `SearchGrid.lic` from the appliance to the utility node (`cw-util`) `<installdir>/Search Grid 2.0/etc`.

   There are two possible locations for the license file:

   1. `<CW_Installer>\packages\Nexidia\NexidiaSearchGrid-2.0.2.2`

   2. `C:\Program Files\Nexidia\Search Grid 2.0\etc\SearchGrid.lic` (on `cw-appl`)

4. Change `<installdir>/Search Grid 2.0/etc/local.properties.xml` to the following on the cw-util node.

   Note: Only the machine name (for example, `cw-util`) is required. A fully qualified domain name is not necessary or certified.

```
<properties>
  <entry key="nexidia.searchgrid.this.public.bindAddress">cw-util</entry>
  <entry key="nexidia.searchgrid.this.gridprivate.bindAddress">cw-util</entry>
  <entry key="nexidia.searchgrid.gateway.public.address">cw-appl</entry>
  <entry key="nexidia.searchgrid.gateway.gridprivate.address">cw-appl</entry>
</properties>
```

Modify Properties on Appliance (cw-appl)

Modify the properties from two files on the main node appliance (`cw-appl`).

1. Shut down the audio services using the command:

```
b stop-audio-services-dont-disable
```

2. Modify the `gridprivate` properties of the `<installdir>/Search Grid 2.0/etc/local.properties.xml` to use the system name.

   Note: Only the `gridprivate` properties that are set as 127.0.0.1 need to change.

```
<properties>
<entry key="nexidia.searchgrid.this.public.bindAddress">cw-appl</entry>
<entry key="nexidia.searchgrid.this.gridprivate.bindAddress">cw-appl</entry>
<entry key="nexidia.searchgrid.gateway.public.address">cw-appl</entry>
<entry key="nexidia.searchgrid.gateway.gridprivate.address">cw-appl</entry>
</properties>
```

3. Modify the `<installdir>/Search Grid 2.0/etc/data-node/data-node-properties.xml` as follows

```
<properties>
....
<entry key="nexidia.searchgrid.journal.webservice.gridprivate.address">cw-appl</entry>
...
</properties>
```

Restart the Services

1. Start all audio services on the appliance (`cw-appl`) from the command prompt:

   ```
   b start-audio-services
   ```

2. Start the Nexidia Search Grid Compute Service on the utility node (`cw-util`). The successful startup of compute service is written to Compute Service log at the location: `<SearchGrid-Data-Dir>/compute-node/logs directory`

3. Once the Search Grid Compute Service starts, the utility node (`cw-util`) will start pulling audio processing work from the appliance node (`cw-appl`).

Removing Compute Node

Compute Nodes can be removed after the initial deployment without reconfiguring or downtime. If the Search Grid software is no longer required, you can uninstall the utility node (`cw-util`) or stop the Compute Service. No change needs to be made on to the appliance configurations (`cw-appl`).

Scale Audio Search

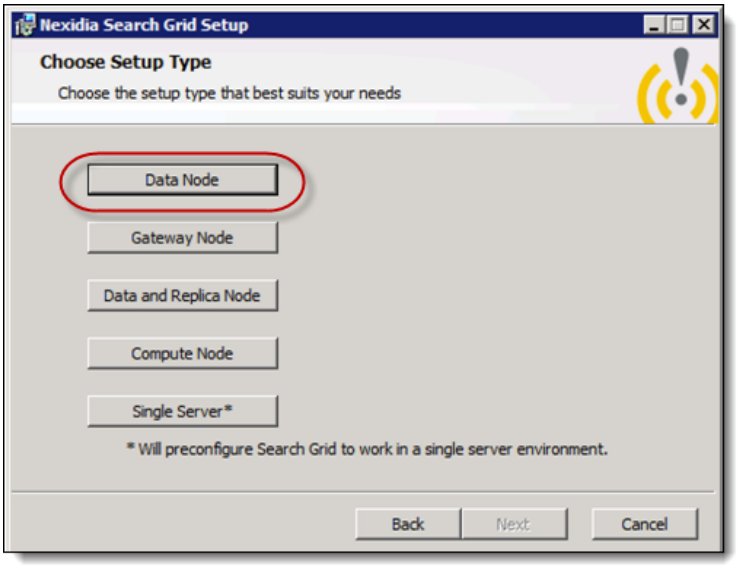
The ability to scale audio search is achieved by adding a new data node. Since a data node also contains compute capabilities, adding a data node means scaling both the processing and search capabilities. However, there are a number of considerations to weigh before adding a new data node and these are discussed in the next section.

Assess Pros and Cons of Adding a Data Node for Audio Search **IMPORTANT!**

Important criteria to consider when adding a data node for audio search is that, while it is relatively easy to add a data node to scale audio search, it is considerably more complex to remove it. This is because once a data node is added, a part of the audio index is distributed on the new data node and it cannot be removed without following manual and time-consuming processes to redistribute the data back to the original node. For these reasons, adding a node for audio search should only be attempted when there is a clear and identifiable scaling audio search performance need.

Install Data Node on Utility Node (cw-util)

4. Run the Search Grid installer and choose the Data Node setup type.



Note: The Search Grid installer is located at:
`<CW_Installer>\packages\Nexidia\NexidiaSearchGrid-2.0.2.2`

5. Choose the same service credential (user name and password) that were used for the *ESAApplicationService* on the appliance (`cw-appl`). These service credentials must have access rights to read the audio files that need to be processed - similar to the credential of the Agent Service on the appliance (`cw-appl`).
   - Note that you can provide a different data directory during the installation.
   - Data directory location - this will contain the log files that will be needed for any troubleshooting.
6. Copy the license file `SearchGrid.lic` from the appliance to the utility node (`cw-util`) `<installdir>/Search Grid 2.0/etc`.

   There are two possible locations for the license file:

   1. `<CW_Installer>\packages\Nexidia\NexidiaSearchGrid-2.0.2.2`
   2. `C:\Program Files\Nexidia\Search Grid 2.0\etc\SearchGrid.lic` (on `cw-appl`)

7. Change `<installdir>/Search Grid 2.0/etc/local.properties.xml` to the following on the `cw-util` node.

   Note: Only the machine name (for example, `cw-util`) is required. A fully qualified domain name is not necessary or certified.

```
<properties>
  <entry key="nexidia.searchgrid.this.public.bindAddress">cw-util</entry>
  <entry key="nexidia.searchgrid.this.gridprivate.bindAddress">cw-util</entry>
  <entry key="nexidia.searchgrid.gateway.public.address">cw-appl</entry>
  <entry key="nexidia.searchgrid.gateway.gridprivate.address">cw-appl</entry>
</properties>
```

8. Change the following properties for the `<installdir>/Search Grid 2.0/etc/data-node/data-node-properties.xml` to the following:

   IMPORTANT! Do not use non-alphanumeric characters. For example, `my-PC` is an invalid machine name.

```
<properties>
....
 <entry key="nexidia.searchgrid.node.name">cwutil</entry>
...
</properties>
```

Modify Properties on Appliance (cw-appl)

1. Stop all audio services

```
b stop-audio-services-dont-disable
```

2. Modify the following property in `<installdir>/Search Grid 2.0/etc/gateway-node/global-properties.xml` to add the new node name. This should be set to the same name that was added in the data node property as described above and set the `maintenanceMode` property to true.

```
<properties>
....
<entry key="nexidia.searchgrid.nodeNameList">DefaultNode,cwutil</entry>
...
<entry key="nexidia.searchgrid.maintenanceMode">true</entry>
</properties>
```

Redistribute Data to the New Data Node on the Utility Node (cw-util)

To redistribute data to the new Data Node, follow these steps:

1. Start Services (Gateway Node and all Data Nodes)
   A. Wait for Gateway Node to completely load before starting Data Nodes
   B. Look for "completed startup activities" in the Gateway Service log

2. Start Management Console and log in as administrator. Click **Start > All Programs > Nexidia > Search Grid 2.0**.
3. From the Management Console, enter "request-redistribution"
   A. This will move media from existing nodes to the new nodes, and balance data volume across all nodes.
   B. A list of which nodes are sending media to which other nodes will be displayed.
   C. If no nodes are displayed, this probably means there is no need to redistribute. Check the Gateway Service log as described in Step 4.
4. Check the Gateway Service log to determine when redistribution has completed. One of the following status messages will display:
   A. "Data redistribution complete" - indicates that data distribution ran and completed.
   B. "No need to redistribute data" - indicates that data was already distributed appropriately
   C. Check individual Data Node logs to monitor the progress of the redistribution
5. When redistribution is complete, stop all services including Gateway.
6. Modify `<install location>\etc\gateway-node\global-properties.xml` and set `nexidia.searchgrid.maintenanceMode` to false.
7. Start all Search Grid Services on all servers.
   A. Wait for Gateway Node on `cw-appl` to completely load before starting newly configured `cw-util` notes (Data Nodes).
   B. Look for "completed startup activities" in the Gateway Service log
8. Once Search Grid Services have started, resume normal operation

Redistribute Data From a Data Node Back to Appliance (cw-appl)

To decommission a data node, you must move all the data back to the appliance (`cw-appl`). To redistribute the data, follow the steps described above with the following changes:

1. Set the following property in the `data-node.properties` file on the utility node (`cw-util`)

   ```
   <entry key="nexidia.searchgrid.node.capacity">0.0</entry>
   ```

2. Modify `<install location>\etc\gateway-node\global-properties.xml` and set `nexidia.searchgrid.maintenanceMode` to true.
3. Restart all Search Grid Services on all servers.
4. Run the redistribution Steps 1 through 5 as outlined above
5. For Step 6, remove the Data Nodes' name from `Nexidia.searchgrid.nodeNameList`.
6. Uninstall the Data Node from the utility node (`cw-util`)

7. Restart all audio services on cw-app and resume normal use

```
b start-audio-services
```

Appendix F: Product Documentation

The table below lists the administrator and end-user documentation that is available for the Veritas eDiscovery Platform product.

Veritas eDiscovery Platform Documentation

Document	Comments
Installation and Configuration	
Installation Guide	Describes prerequisites, and how to perform a full install of the Veritas eDiscovery Platform application
Upgrade Overview Guide	Provides critical upgrade information, by version, useful prior to upgrading an appliance to the current product release
Upgrade Guide	Describes prerequisites and upgrade information for the current customers with a previous version of the software application
Utility Node Guide	For customers using utility nodes, describes how to install and configure appliances as utility nodes for use with an existing software setup
Distributed Architecture Deployment Guide	Provides installation and configuration information for the Review and Processing Scalability feature in a distributed architecture deployment
Getting Started	
Navigation Reference Card	Provides a mapping of review changes from 10.x compared to 9.x, 8.x compared to 7.x and 7.x compared to 6.x
Administrator's QuickStart Guide	Describes basic appliance and case configuration
Reviewer's QuickStart Guide	A reviewer's reference to using the Analysis & Review module
Tagging Reference Card	Describes how tag sets and filter type impact filter counts
User and Administration	
Legal Hold User Guide	Describes how to set up and configure appliance for Legal Holds, and use the Legal Hold module as an administrator
Identification and Collection Guide	Describes how to prepare and collect data for processing, using the Identification and Collection module
Case Administration Guide	Describes case setup, processing, and management, plus pre-processing navigation, tips, and recommendations. Includes processing exceptions reference and associated reports, plus file handling information for multiple languages, and supported file types and file type mapping
System Administration Guide	Includes system backup, restore, and support features, configuration, and anti-virus scanning guidelines for use with Veritas eDiscovery Platform
Load File Import Guide	Describes how to use and produce exports, productions, and logs (privilege and redaction logs)
User Guide	Describes how to perform searches, analysis, and review, including detailed information and syntax examples for performing advanced searches

Veritas eDiscovery Platform Documentation

Document	Comments
Imaging Tool Upgrade Guide	Provides details about the Imaging Tool Upgrade feature and how to perform Imaging Tool Upgrade after the eDiscovery Platform appliance is upgraded to version 10.1, workflows affected when the cases are upgraded or not upgraded, and frequently asked questions (FAQs).
Export and Production Guide	Describes how to use and produce exports, productions, and logs (privilege and redaction logs)
Transparent Predictive Coding User Guide	Describes how to use the Transparent Predictive Coding feature to train the system to predict results from control data and tag settings
Audio Search Guide	Describes how to use the Audio Search feature to process, analyze, search and export search media content
Reference and Support	
Audio Processing	A quick reference card for processing multimedia sources
Audio Search	A quick reference card for performing multimedia search tasks
Legal Hold	A quick reference card of how to create and manage holds and notifications
Collection	A quick reference card of how to collect data
OnSite Collection	A quick reference for performing OnSite collection tasks
Review and Redaction	Reviewer's reference card of all redaction functions
Keyboard Shortcuts	A quick reference card listing all supported shortcuts
Production	Administrator's reference card for production exports
User Rights Management	A quick reference card for managing user accounts
Online Help	
Includes all the above documentation (excluding Installation and Configuration) to enable search across all topics. To access this information from within the user interface, click **Help**.	
Release	
Release Notes	Provides latest updated information specific to the current product release