

SEARCH AND REVIEW OF AUDIO

Sightline Help Documentation



Table of Contents

Table of Contents	2
Running and Refining Audio Searches in Sightline	2
Search Terms	3
Building Strong Search Terms	3
Run the First Iteration of Terms	4
Update the Terms	4
Sightline Audio Search Logic	5
Term Operator: ANY vs ALL	5
Wildcard Searching	5
Proximity Searching	5
Language Pack/Dialect	5
Numbers and Acronyms	6
Combining Audio Searches with Other Searches	6
Audio Review	7
Bifurcated Hit Review	7
Audio Review	7
Setting Up Assignments	7
Viewing and Listening to the Audio Files	8
DocView with Audio	8

Running and Refining Audio Searches in Sightline

Once the data has been loaded, it is now time to implement the searches and review the hits. Searching audio is more closely related to training a predictive coding model than it is to searching email. There is a period of judging recall and precision, along with adjusting the Confidence Levels and terms to find the best combination to provide the relevant hits.

Search Terms

Before any terms can be run in Sightline, it is important to review the list of terms. It may seem logical to use the same list of terms that were used for the text-based searches in the matter. This, however, is generally problematic, as people typically write differently than they speak. Additionally, text-based searching can be more complex, and the same types of searches used for text-based searches may not be feasible with Audio.

Building Strong Search Terms

When crafting your audio searches there are some general best practices that Consilio has developed. First and foremost, it is advised to not use monosyllabic search terms. Since the index is based on sounds, searching for a monosyllabic term will return any word or phrase that has that sound in it. Search phrases are far superior in finding relevant hits.

Determine the native language or location of the Custodians being searched to account for any accents or foreign language. Nexidia uses Language Packs to build the indices. If, for instance, the caller is based in the UK, we would load the International English Language Pack. If the user speaks Spanish, we would load the Spanish Language Pack. You can load multiple Language Packs to a database, but each will build its own separate index that will need to be searched and billed separately.

If there are any proper nouns in the search term list, find the common pronunciations of each term. Company names like, “Consilio” will need to be searched phonetically to ensure the correct pronunciation is captured, e.g. “con-sill-ee-oh”. This is the same for names like, “Smyth” which can be pronounced, “Smith” or “Sm-eye-th.”

When crafting phonetic searches, it is important to always search the exact term in addition to any phonetic variations, e.g. Libor should be searched as both “Libor” and “Lie-bore.” You should put a hyphen between the sounds of a single word. The hyphen joins the sounds together for the search. If you do not use the hyphen, the system will treat the space as a pause, which may affect the results.

It is possible that business specific jargon or acronyms may need to be searched. It is important to find out the correct pronunciation of the jargon term or terms being searched so you may craft the search accordingly. Additionally, if there are any acronyms in your searches you will need to find out if they are said as a word (like NASA or OSHA) or spoken as letters (like FBI or CIA). If they are said as a word, they should be phonetically spelled out (e.g. “nah-suh”). If they are spoken as letters they will need to be hyphenated (e.g. “f-b-l”) or spelled phonetically (e.g. “eff-bee-eye”).

Just like any text-based searching, you should first look for any obvious flood terms. For audio this could be any monosyllabic terms as mentioned above or words that are spoken regularly on the calls. If the name of the company was Dewey, Cheatem and Howe, you would not want to search for the term Cheatem, as any time someone answered the phone, they would most likely say the company name.

Lastly, it is useful to know the environment in which the calls were recorded. If the calls were recorded on a trading floor or in a board room, there could be multiple talkers conversing over one another. There could also be a lot of background noise to anticipate. In these instances, you may want to consider lowering the Confidence Level.

Run the First Iteration of Terms

By now you should have the terms formatted appropriately for audio searching. With the data loaded you can run the first iteration of the terms. Once the searches have completed, you can first look at any hits that have a very high number of responsive files. These are most likely flood terms, or the search is hitting on some commonly used word or phrase in the calls. Sample a few of these calls and see if you can determine where the hits are coming from.

If you can locate a flood term, you can go back to your term list and adjust. It is helpful if you can locate a true hit as well, so you can see if there are other words said around the term being searched. You can then add that word or words to your search to refine it. If you are unable to locate any common word or phrase causing the high number of false positives, or there is nothing obvious, you can also raise the Confidence Level and see if a higher threshold will return better hits.

Update the Terms

Once you have addressed any possible flood terms you should also look at any terms with very low hits or no hits. Here you may need to broaden the search term by removing words or sounds to see if you can get any or better hits. You can also move the Confidence Level down to a lower threshold to see if you can locate additional responsive hits.

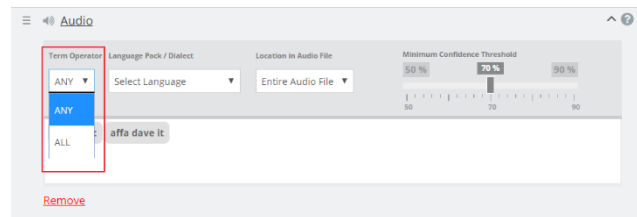
Using the predetermined knowledge gleaned by using the tactics provided in this document and the knowledge gained from listening to the search terms you can now make your adjustments and rerun the updated terms. You may need to run this exercise multiple times until you are able to isolate the expected hits. You do not need to get all the hits correct, but you should aim to get close.

Sightline Audio Search Logic

The audio search in Sightline is user friendly and straight forward so long as you understand how the index is built. You can search for a word by simply typing in that word, or the phonetic spelling of the word, into the search window, e.g., you can type “affidavit” or “affa-dave-it” into the search. Always use a hyphen for phonetic spelling of a single word. If you have more than one term, or a phrase, they can be entered as separate words with a space between them.

Term Operator: ANY vs ALL

In addition to typing out your terms into the search window, you are also given the option for run the search as an ANY search or an ALL search. ANY will return any audio file in with one or more of the listed terms is found. ALL will return any audio file in which every term listed is present in the file.

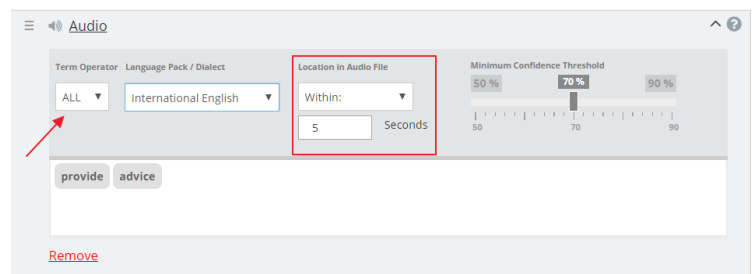


Wildcard Searching

Since Nexidia builds its index on phonetic building blocks there is no wildcard logic that needs to be used. Running a search for, “tech” will return any audio file that contains that sound, including technology, technologist, and nanotech. Think of searching in audio as running all your text-based search terms as bracketed wildcard searches; “tech” in audio searching is similar to searching “*tech*” in text-based searching.

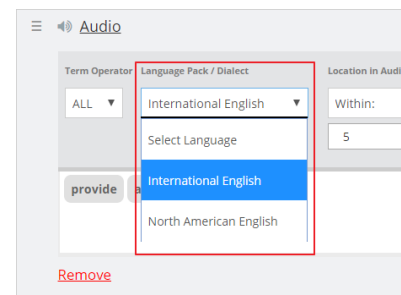
Proximity Searching

Sightline allows for basic proximity searches to be run. When you add more than one term to the search box as an “ALL” search, you can use the “Location in Audio” dropdown to choose the number of seconds that can exist between the terms.



Language Pack/Dialect

As discussed previously, Nexidia builds its phonetic index using Language Packs. Multiple Language Packs can be used in the same database; however, they must be run separately. If multiple Language Packs are used, you will need to select the appropriate Language Pack from the dropdown.



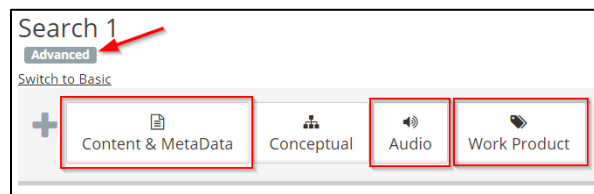
Numbers and Acronyms

When searching for numbers, they will need to be spelled out phonetically, e.g., “100” should be searched as “one hundred.” Please also take note of how people generally speak numbers. Most times people do not say, “zero” and instead say “oh”; e.g., “90210” might be spoken, “nine oh two one oh” rather than “nine zero two one zero.”

Acronyms should also be spelled out phonetically. If you are trying to find someone talking about the FBI, you would need to search for, “f-b-i” or “eff-bee-eye.” As previously discussed, if the acronym is also a word, like NASA or OSHA, you should spell out the word, e.g., “nah-suh” and “oh-sha.”

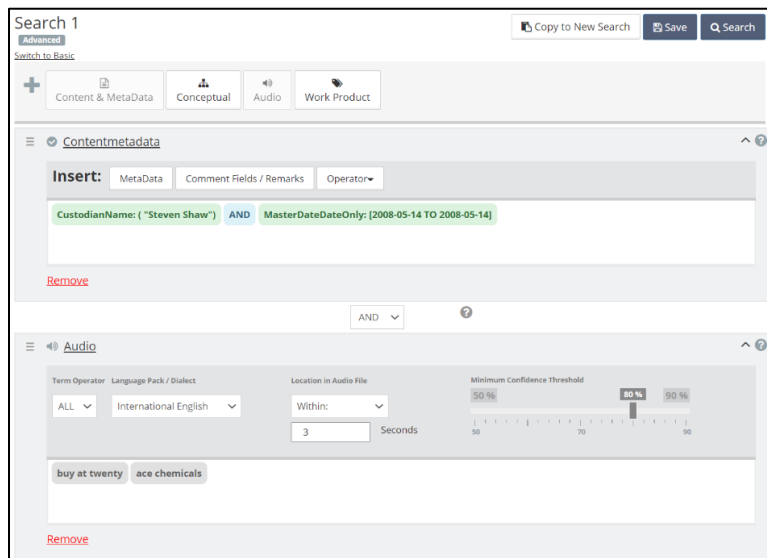
Combining Audio Searches with Other Searches

It is possible to combine Audio Searches with both Content & Metadata Searches and Work Product Searches. Conceptual Searching is not allowed with Audio Searching. To combine searches, you simply need to click the search you wish to add to your Audio search from the Advanced Search window.



Please note, that the Audio Search must be last when combining search logic.

In the below example, the User is searching for any audio files from Custodian Steven Shaw that were made on May 14, 2008 and contain the phrases “Buy at Twenty” and “Ace Chemicals” spoken within 3 seconds of each other at no less than 80% Confidence Level. The Audio search portion must be last when combining searches for the search hits to be properly highlighted in the audio files.



Audio Review

Once the terms have been refined, the data can now be released and placed into assignments for review. When the data is assigned out from a search result, the search terms will be “highlighted” in the waveform, along with a listing of the Confidence Level of the hit and it’s time stamp.

Bifurcated Hit Review

Consilio can offer to bifurcate the review to a low-cost review center to perform a Yes/No review of the hits. By using off-shore, or other low-cost review alternatives, Consilio can quickly determine if the hit in the Audio is, in fact, the term that was being searched. If the hit matches the term, the review would check yes and only those hits would be promoted for full review. This solution would limit the more expensive resources having to review audio in which there was no verified hit.

Audio Review

The power of reviewing Audio in Sightline is found in the ability to eliminate non-hits through both phonetic searching and by being able to quickly jump to the term in the Audio file and make a determination without having to listen to the entire file. In this way, the reviewer only needs to listen to the files with responsive hits on the terms; non-hits and non-responsive hits can easily and quickly be eliminated, focusing the review on only the files that have responsive hits.

Setting Up Assignments

In order for the Audio hits to be visible in the viewer, the documents must be either viewed from a search – either a Session or a Saved Search – or placed into an assignment from a Session or Saved Search. As Sightline is centered around review from within an Assignment, we will focus our attention on these methods. There are 3 basic options for placing Audio with hits into an Assignment.

1. You can assign out from each search into a search specific Assignment. Please note that this will only highlight the term, or terms, from that search, and will not highlight any other terms in subsequent searches. Also, if you assign out from each individual search there is the possibility for overlap within the Assignments. You can do this be either using placing the hits into the Shopping Cart and then using the Action dropdown from a Session Search, or by selecting the Saved Search you want to assign and click the Assign button from the Search Ribbon. Please ensure that the “Persist Content Hits Within Assignment:” is checked when assigning the files from the Assignment window.
2. You can assign out from a Search Group. If you place all your searches into a Search Group, you can click on the Saved Search Group and then click Assign from the Search Ribbon. Please ensure that the “Persist Content Hits Within Assignment:” is checked when assigning the files

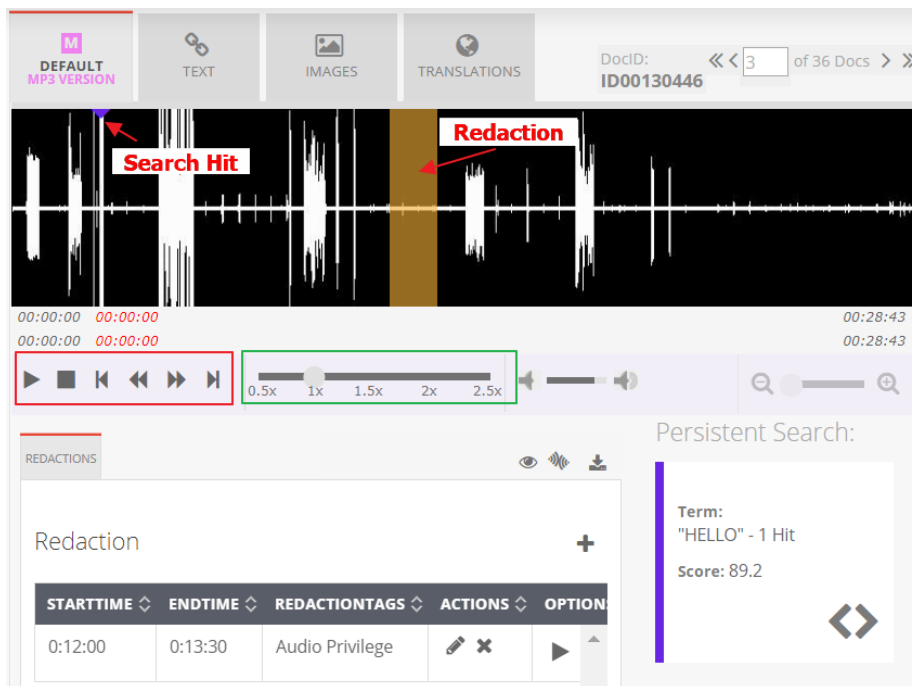
from the Assignment window. By assigning the records out in this way, all the hits, from all the searches in the Group will be visible in the audio files in the Assignment. This acts in the same way as adding Keyword Highlighting to a non-audio assignment.

3. You can assign out from the Search Term Report. Navigate to the Reports Page and select the Search Term Report. From here you can select your audio searches from the list on the left-hand side. Once you have your searches selected, click Apply Changes. This will generate the Search Term Report. Grab the Doc Hits from the Totals row at the top of the report and using the Action dropdown, choose Bulk Assign. Please ensure that the “Persist Content Hits Within Assignment:” is checked when assigning the files from the Assignment window. By assigning the records out in this way, all the hits, from all the searches in the Group will be visible in the audio files in the Assignment. This acts in the same way as adding Keyword Highlighting to a non-audio assignment.

Viewing and Listening to the Audio Files

Reviewing Audio files in Sightline uses the same DocView document viewer as email and efiles. The difference is that the Default View tab will display the Audio Viewer.

DocView with Audio



The Audio Player will be automatically displayed for any Audio file. This will appear in the Default Tab of the DocView. Within the Audio player are the following:

- A visual representation of the Audio file in a waveform
- Any search term hit will be indicated with a colored arrow on the top of the waveform that matches the color of the hit in the Persistent Search section
- Redactions will appear highlighted in Orange
- Below the waveform is the player controls (Red Box)
- Next to the player controls are the speed controls (Green Box)
- Below the player controls are the Redaction controls and then Persistent Search

The reviewer can navigate to the search hit by either user the player control buttons or the arrows in the Persistent Search. If you use the player control buttons the file will cycle between all the hits whereas if you use the arrows in the term tiles in the Persistent Search area you will only cycle between the individual term hits.

Redactions can be made directly in Sightline on the Audio files. The user will add a redaction using the plus and then mark the start and stop time of the redaction. The user will also be asked to apply a Redaction Tag. Once this is complete, the redaction can be saved and can then be applied to the file when it is produced.