Aid4Mail Filter Syntax Overview

1. Introduction

Aid4Mail's powerful filtering capabilities, specifically in the Investigator and Enterprise editions, are designed to meet the exacting needs of professionals in email forensics and eDiscovery. This document provides an overview of Aid4Mail's filter syntax, enabling you to efficiently search, analyze, and extract critical information from vast email datasets.

Please refer to Aid4Mail's User Guide for in-depth coverage of the filter syntax and additional examples.

Purpose of Filtering in Aid4Mail

In the fields of digital forensics and legal discovery, the ability to precisely locate and extract relevant emails is paramount. Aid4Mail supports a multitude of email formats and can even carve emails from unknown file formats, uncompressed disk images, and forensically extracted disk space. However, determining which emails are relevant to a case requires sophisticated filtering.

Aid4Mail's filtering feature allows you to:

- Rapidly identify emails pertinent to an investigation or legal case.
- Exclude irrelevant or privileged communications.
- Detect patterns of communication or specific content across large email sets.
- Isolate emails within specific date ranges or from particular senders.
- Uncover hidden or deleted emails that may be crucial to your investigation.

Familiar Syntax

Aid4Mail's filter syntax is similar to Gmail and Microsoft 365's, making it easy to learn and remember. However, Aid4Mail's syntax is richer, offering capabilities that are on par with, or even exceed, those of other eDiscovery and forensics tools.

Where Aid4Mail's syntax differs from Gmail and Microsoft 365's, it has been highlighted in the relevant parts of this document.

Basic Concepts

Before delving into the specifics of Aid4Mail's filter syntax, it's important to understand a few key concepts:

- **Basic concepts**
    - **Search Terms**: The words, phrases, or patterns you're looking for in emails.
    - **Wildcards**: Special characters that represent unknown or variable parts of a search term.
    - **Search Operators**: Refine search terms to target specific parts or attributes of an email.
    - **Boolean Operators**: Combine or exclude search terms using AND, OR, XOR, and NOT.

- **Advanced concepts**
    - **Proximity Searching**: Find words or phrases that appear near to each other.
    - **Deduplication**: Eliminate duplicate emails from search results.
    - **Unpurged Mail**: Emails that have been deleted but not permanently removed from the system.
    - **Tokenization:** Recognize and match similar characters and words within text.
    - **Stemming**: Find words that share the same root.
    - **Search Lists**: Groups of search terms, saved in external files. They improve organization and simplify search queries.
    - **Regular Expressions** (regex): Powerful search patterns (with a cryptic syntax) that enable otherwise impossible searches. Aid4Mail supports Perl Compatible Regular Expressions (PCRE2 version 10.44, at time of writing) in search terms.

For example, in an investigation into communication patterns, you might use a filter like:

Unset

```
Date>=2023 AND Type:Personal AND ("project alpha" OR
"confidential acquisition")
```

This filter would find all personal emails (excluding newsletters, marketing emails, automated notifications, etc.) from 2023 onwards that mention "project alpha" or "confidential acquisition". This could help isolate direct, person-to-person communications about sensitive topics, filtering out bulk emails or automated notifications that might use similar keywords.

As we progress through this manual, you'll learn how to construct increasingly sophisticated filters to support your forensic analysis or eDiscovery efforts. Whether you're searching for specific pieces of evidence, establishing communication patterns, or isolating relevant date ranges, Aid4Mail's filter syntax provides the precision and flexibility you need for your professional investigations.

2. Search Terms

Search terms are the foundation of Aid4Mail’s filtering capabilities. They allow forensics experts and eDiscovery professionals to pinpoint specific content within large email datasets.

Definition and Usage

A search term is a word, phrase, or pattern that Aid4Mail looks for in emails. These can be simple keywords or more complex expressions using wildcards and operators. Search terms are used to identify emails containing specific information relevant to an investigation or legal case.

Rules for Search Terms

1. **Case Sensitivity**: By default, search terms are case-insensitive. "EVIDENCE" will match "evidence", "Evidence", and "EVIDENCE".

2. **Whole Word Matching**: Aid4Mail searches for whole words by default. For example, searching for "car" will not match "cargo" or "scar".

3. **Phrases**: To search for an exact phrase, enclose it in double quotes. For example, "intellectual property".

4. **Special Characters**: Some characters have special meaning in Aid4Mail’s syntax. If you need to search for these characters literally, you may need to escape them or use specific syntax.

5. **Combining Terms**: Multiple search terms in a query can be combined using Boolean operators. If none are specified, AND is implicitly used between search terms.

Ordering of Search Terms

The order in which search terms appear in your query can significantly impact Aid4Mail’s processing speed. To enhance efficiency, place search terms with the smallest scope (e.g. date, sender, recipients, subject), and those that are less likely to be found, before those that are more common or covering a larger part of the email. For further details, refer to Best Practices and Tips at the end of this document.

Examples

Here are some examples of search terms tailored for email forensics and eDiscovery scenarios:

1. Simple keyword:

Unset

```
confidential
```

This will find all emails containing the word "confidential".

2. Exact phrase:

Unset

```
"trade secret"
```

This will match emails containing the exact phrase "trade secret".

3. Multiple terms:

Unset

```
lawsuit settlement
```

This will find emails containing both "lawsuit" and "settlement" (in any order).

4. Using wildcards (which we'll cover in more depth later):

Unset

```
litigat*
```

This will match "litigation", "litigate", "litigator", etc.

5. Combining concepts:

Unset

```
"insider trading" OR "market manipulation"
```

This will find emails containing either phrase.

6. Complex example:

Unset

```
Date:2022 AND (embezzle* OR fraud*) AND (account* OR financ*)
```

This search term would be useful in a financial crimes investigation, looking for emails from 2022 that mention embezzlement or fraud in relation to accounts or finances.

Understanding how to construct effective search terms is crucial for efficiently sifting through large volumes of email data. As we progress through this manual, you'll learn how to combine these basic search terms with more advanced filtering techniques to create powerful, precise queries tailored to your specific investigative needs.

3. Wildcards

Wildcards are special characters that represent unknown or variable parts of a search term. They are particularly useful in forensics and eDiscovery when you need to account for variations in spelling, prefixes, suffixes, or unknown parts of an email address or domain.

Aid4Mail supports several types of wildcards, each with specific uses:

Character Wildcards

1. *** (Asterisk)**: Matches zero or more characters within a word.
   Example: `corrup*`
   Matches: corrupt, corruption, corrupted, corrupting

2. **? (Question Mark)**: Matches exactly one character.
   Example: `saniti?e`
   Matches: sanitise, sanitize (British and American spellings respectively)

3. **# (Hash)**: Matches zero or one non-alphanumeric characters.
   Example: `data#breach`
   Matches: data breach, data-breach, databreach

4. **~ (Tilde)**: When placed at the end of a word, it performs stemming (if a stemming dictionary has been set).
   Example: `steal~`
   Matches: steal, steals, stole, stolen

Known Variation Character Wildcards

These wildcards match known variations of characters, including accented letters and ligatures.

Example: `<e>vidence`
Matches: evidence, évidence

Word Proximity Wildcards

These are crucial for proximity searching, allowing you to find words near each other within emails. There are two variations of each of these wildcards: **<n>, <.>,** and **<*>** ignore the order of the words surrounding the wildcard whereas **<+n>, <+.>,** and **<+*>** respect word order.

1. **<n>** and **<+n>**: Match up to n words, where n is a number between 0 and 99.
   Example: `fraud<5>report`
   Matches: "fraud" and "report" within 5 words of each other.

2. **<.>** and **<+.>**: Match words in the same sentence.
   Example: `bribe<.>official`
   Matches: "bribe" and "official" in the same sentence.

3. **<*>** and **<+*>**: Match words in the same paragraph.
   Example: `insider<*>trading`
   Matches: "insider" and "trading" in the same paragraph.

Examples of Usage in Forensics and eDiscovery

1. Investigating various forms of financial misconduct:

Unset

```
embezzle* OR fraud* OR money<5>launder*
```

This search would find mentions of embezzlement, fraud, and money laundering, including variations of these terms.

2. Searching for potentially altered documents:

```
Unset

(modif* OR chang* OR alter*)<.>(document* OR file* OR record*)
```

This would find sentences mentioning modifications to documents, files, or records.

3. Identifying communication about insider information:

```
Unset

(insider<*>trading) OR (material<5>nonpublic<5>information)
```

This search would find paragraphs mentioning insider trading or discussions of material nonpublic information within close proximity.

4. Investigating international communications:

```
Unset

@*.?? OR @*.??? OR @*.????
```

This would match email addresses with two-letter (.us, .uk), three-letter (.com, .org), or four-letter (.info) top-level domains.

5. Searching for potential code words or deliberate misspellings:

```
Unset

th?ft OR fr??d
```

This could catch attempts to obfuscate discussions of theft or fraud.

Wildcards significantly enhance the power and flexibility of your searches, allowing you to cast a wider net in your investigations while still maintaining precision. They're particularly valuable when dealing with large datasets where you may not know the exact phrasing used in relevant communications.

4. Search Operators

Search operators in Aid4Mail allow you to refine your searches by specifying which parts of an email to search or by setting specific criteria. These operators are crucial for targeted investigations and efficient eDiscovery processes.

Folder Search Operators

These operators allow you to search within specific folders or types of folders.

1. **FolderName:** Searches for messages in a particular folder.
   Example: `FolderName:Archive`

2. **In:** Searches for messages in a particular type of folder.
   Example: `In:Sent` (searches in the sent items folder, whatever its name)

Example in forensics:

Unset

```
FolderName:"Project X" AND Subject:confidential
```

This would search for confidential emails specifically within the "Project X" folder.

Sender/Recipient Search Operators

These operators allow you to filter emails based on the parties involved in the communication.

1. **From:** Searches for messages from a particular sender.
   Example: `From:john@company.com`

2. **To:** Searches for messages to a particular recipient.
   Example: `To:jane@company.com`

3. **CC:** Searches for messages where a particular recipient is cc'd.

4. **Bcc:** Searches for messages where a particular recipient is bcc'd.

5. **Sender:** Searches for messages where a particular sender appears in the From, Sender, or Reply-To fields.

6. **Recipients**: Searches for messages where a particular recipient appears in the To, CC, or BCC fields.

7. **Participants:** Searches for messages involving any of the specified email addresses.

Example in eDiscovery:

Unset

```
Participants:ceo@company.com AND Subject:(merger OR acquisition)
```

This would find all emails involving the CEO that mention mergers or acquisitions.

Date/Time Search Operators

These operators allow you to search within specific time frames.

1. **Date:** Searches for messages on a particular date or within a date range.
   Example: `Date:2023-01-01`

2. **Sent:** Searches for messages sent on or relative to a particular date.
   Example: `Sent>2023-06-01`

3. **Received:** Searches for messages received on or relative to a particular date.

Aid4Mail uses the International Date Format:

- YYYY
- YYYY-MM
- YYYY-MM-DD
- "YYYY-MM-DD hh"
- "YYYY-MM-DD hh:mm"
- "YYYY-MM-DD hh:mm:ss"

Note that double-quotes are required around dates that include the time, due to the space character. As in the example above, date/time search operators can use comparison operators:

- **>** (greater than)
- **<** (less than)
- **>=** (at least)
- **<=** (at most)

To specify a date range in Aid4Mail, use two separate conditions combined with the AND operator. For example:

Unset

```
Sent>=2023-04-14 AND Sent<=2024-03-21
```

This would search for emails sent between April 14, 2023, and March 21, 2024, inclusive.

⚠ Unlike Gmail and Microsoft 365, you cannot use the “..” notation to define a date range.

You can also search for partial dates:

Unset

```
Sent:2023-06
```

This would match all emails sent in June 2023.

Unset

```
Received:2023
```

This would match all emails received in the year 2023.

Example in forensics:

Unset

```
Sent>=2023-01-01 AND Sent<=2023-03-31 AND
From:suspect@company.com AND Subject:confidential
```

This would search for confidential emails sent by a suspect during the first quarter of 2023.

Additional date-related operators include:

- **SentDay:** Searches for emails sent on a particular day of the week.
  Example: `SentDay:(Saturday OR Sunday)`

- **SentTime:** Searches for emails sent at a particular time of day.
  Example: `SentTime>=18:00 AND SentTime<=23:59`

- **OlderThan:** and **NewerThan:** Search for emails older or newer than a specified time period.
  Example: `OlderThan:30d` (older than 30 days)

These date and time operators are crucial in forensics and eDiscovery for establishing timelines, identifying patterns of communication, and focusing investigations on specific periods of interest.

Email Section Search Operators

These operators allow you to search within specific parts of an email.

1. **Subject:** Searches within the subject line. Example: `Subject:"urgent meeting"`
2. **Header:** Searches the email header.
3. **Message**: Searches the message content.
4. **SenderMessage**: Searches the newest part of the message content (the part created by the sender), excluding older quoted emails in the thread.
5. **AttachmentNames:** Searches for specific attachment names.

Example in eDiscovery:

Unset

```
Subject:(contract OR agreement) AND AttachmentNames:*.pdf
```

This would find emails with "contract" or "agreement" in the subject that have PDF attachments.

⚠ Note that the parentheses in the example serve two purposes:

1. They allow two keywords to be used with the Subject operator, which is shorter than writing `Subject:contract OR Subject:agreement`.

2. They override the default precedence of the “AND” operator over “OR”. In other words, without the parentheses, “agreement” AND “AttachmentNames:*.pdf” would be evaluated before “Subject:contract”.

Email Attribute Search Operators

These operators search for emails with specific attributes.

1. **Type:** Searches for emails of a particular type. Examples:
   a. `Type:Personal` (finds emails that are not newsletters, automated messages, etc.).

    b. `Type:Unpurged` (finds emails that have been deleted but are still in the mail store).
    c. `NOT Type:Duplicate` (skips duplicates).

2. **Status information:** Searches for emails based on user actions. Examples:
    a. `IsRead:True` or `Is:Read` (finds emails that have been read by the recipient).
    b. `Is:Starred` (finds emails that the custodian has flagged or starred).
    c. `Is:Replied` (finds emails that the custodian has replied to).

3. **Importance:** Searches for emails marked with a specific importance level.

4. **Size:** Searches for emails of a particular size.

Example in forensics:

Unset

```
Type:Personal AND Size>5M AND AttachmentNames:*.zip
```

This would find personal emails (not bulk messages) larger than 5MB with zip attachments, which could be relevant in data exfiltration investigations.

Email ID Search Operators

These operators allow you to search for specific emails by their unique identifiers.

1. **MessageId:** Searches for a specific Message-ID.
2. **Uid:** Searches for a specific IMAP UID.

Example in eDiscovery:

Unset

```
MessageId:<abc123@company.mail.com> OR
MessageId:<def456@company.mail.com>
```

This would retrieve specific emails identified by their Message-IDs, which could be useful when following up on particular pieces of evidence.

You can also write the above example like this:

Unset

```
MessageId:{<abc123@company.mail.com>|<def456@company.mail.com>}
```

The curly braces “{}” and vertical tab “|” notation can be used to shorten scripts when combining multiple “OR” criteria. See the Aid4Mail User Guide for a more in-depth coverage of this syntax.

These search operators provide powerful tools for forensics experts and eDiscovery professionals to narrow down their searches and quickly locate relevant emails within large datasets. By combining these operators with search terms and wildcards, you can create highly specific and effective search queries.

5. Boolean Operators

Boolean operators are essential tools in Aid4Mail for combining or excluding search terms. They allow forensics experts and eDiscovery professionals to create complex, precise queries to pinpoint relevant emails within large datasets.

Main Boolean Operators

1. **AND**: Finds emails that contain all specified terms.
   Syntax: `term1 AND term2`
   Example: `embezzlement AND "financial records"`

2. **OR**: Finds emails that contain at least one of the specified terms.
   Syntax: `term1 OR term2`
   Example: `bribery OR kickback`

3. **XOR**: Finds emails that contain either one term or the other, but not both.
   Syntax: `term1 XOR term2`
   Example: `"intellectual property" XOR patent`

4. **NOT**: Excludes emails containing the specified term.
   Syntax: `NOT term`
   Example: `confidential AND NOT draft`

Symbols for Boolean Operators

Aid4Mail also supports symbolic representations of Boolean operators:

- `+` : Equivalent to AND

- `|` : Equivalent to OR
- `-` : Equivalent to NOT
- `^` : Equivalent to XOR

These symbols can be used without spaces, making queries more concise:

Example: `confidential+urgent-draft`
This is equivalent to: `confidential AND urgent AND NOT draft`

Order of Precedence

The order in which Aid4Mail processes boolean operators is important:

1. Parentheses ()
2. NOT
3. AND
4. XOR
5. OR

⚠ Note that this order is different from Google/Gmail's, which gives OR precedence over AND.

Example #1:

Unset

```
Subject:contract OR agreement AND Sender:john.doe@aid4mail.com
```

The "AND" operator has higher precedence than "OR". As a result, this search query will match emails that have the word "contract" in the subject line, as well as emails from john.doe@aid4mail.com that have the word "agreement" anywhere in the email. In other words, this search query is equivalent to:

Unset

```
Subject:contract OR (agreement AND Sender:john.doe@aid4mail.com)
```

Example #2:

Unset

```
Subject:(contract OR agreement) AND Sender:john.doe@aid4mail.com
```

This query will match emails sent by john.doe@aid4mail.com that have either “contract” or “agreement” in the subject line.

⚠ Understanding this order is crucial for creating accurate, complex search queries.

Examples of Complex Queries

1. Investigating potential insider trading:

Unset

```
(("insider information" OR "material nonpublic") AND (trade OR
stock OR share)) AND NOT (newsletter OR "press release")
```

This query looks for discussions of insider information or material nonpublic information in relation to trades, stocks, or shares, while excluding newsletters and press releases.

2. Examining communication patterns in a fraud case:

Unset

```
Date>=2023-01-01 AND Date<=2023-06-30 AND
(From:suspect@company.com OR To:suspect@company.com) AND (money
OR payment OR transfer)
```

This query focuses on emails to or from a suspect, involving financial terms, within a specific six-month period.

3. Identifying potential data breaches:

Unset

```
("data leak" OR "information breach" OR "unauthorized access")
AND (customer OR client OR patient) AND NOT (drill OR test OR
exercise)
```

This search looks for mentions of data breaches involving customer, client, or patient information, while excluding emails about security drills or tests.

4. Investigating intellectual property theft:

Unset

```
("trade secret" OR patent OR copyright) AND (steal OR theft OR
misappropriate OR "unauthorized use") AND NOT legal
```

This query searches for discussions about stealing or misusing intellectual property, excluding emails that might be from the legal department discussing these terms in a different context.

5. Examining conflicts of interest:

Unset

```
("conflict of interest" OR "competing interest" OR "personal
benefit") AND (disclosure OR report OR notify) XOR conceal
```

This complex query looks for emails discussing conflicts of interest in relation to disclosure or reporting, or alternatively, concealment of such conflicts, but not both simultaneously.

By leveraging these Boolean operators and understanding their order of precedence, forensics and eDiscovery professionals can craft highly specific search queries. This allows for efficient filtering of large email datasets, helping to quickly identify the most relevant communications for an investigation or legal proceeding.

6. Proximity Searching

Proximity searching allows you to find words or phrases that appear near each other in an email. This is crucial for identifying relevant conversations and context in investigations. Wildcards that respect word order (**<+n>**, **<+.>**, and **<+*>**) are often more useful for proximity searching than those that ignore it (**<n>**, **<.>**, and **<*>**). This is because there is a higher risk of false positives when word order is ignored.

1. **Using <n> or <+n> Wildcard**
   Syntax: `word1<n>word2` or `word1<+n>word2`
   Example: `bribe<5>official`
   Finds: "bribe" and "official" appearing in any order, within 5 words of each other.

2. **Using <.> or <+.> Wildcard**
   Syntax: `word1<.>word2` or `word1<+.>word2`
   Example: `insider<+.>trading`
   Finds: "insider" and "trading" appearing in the order specified, in the same sentence.

3. **Using <*> or <+*> Wildcard**
   Syntax: `word1<*>word2` or `word1<+*>word2`
   Example: `confidential<+*>agreement`
   Finds: "confidential" and "agreement" appearing in the order specified, in the same paragraph.

Example in forensics:

Unset

```
(trade<+2>secret) AND (steal<.>proprietary)
```

This query would find emails discussing trade secrets within five words of each other, in the same email as discussions of stealing proprietary information within the same sentence.

7. Searching by Email Type

Searching by an email type can be significantly more powerful than it may initially seem. Three common examples of this are:

1. Deduplication
2. Searching unpurged mail
3. Searching personal mail

Deduplication

Deduplication is the process of eliminating duplicate emails from your search results. This is crucial for efficiency in large-scale investigations.

To skip duplicates in Aid4Mail, use the following search term:

Unset

```
NOT Type:Duplicate
```

or its shorter form:

Unset

```
-Type:Duplicate
```

Example in eDiscovery:

Unset

```
(contract OR agreement) AND NOT Type:Duplicate
```

This would find unique emails about contracts or agreements, eliminating duplicates to streamline review.

Searching Unpurged Mail

Unpurged mail refers to emails that have been deleted but not permanently removed from the system. These can be crucial in forensic investigations.

To include only unpurged mail:

Unset

```
Type:Unpurged
```

To exclude unpurged mail:

Unset

```
NOT Type:Unpurged
```

or

Unset

```
-Type:Unpurged
```

Example in forensics:

```
Unset
Type:Unpurged AND Date>=2023-01-01 AND Date<=2023-06-30 AND
From:suspect@company.com
```

This would search for unpurged emails from a suspect within a specific date range, potentially uncovering deleted evidence.

Searching Personal Mail

Aid4Mail classifies emails as personal based on specific criteria, separating them from newsletters, marketing emails, and automated notifications.

To include only personal emails:

```
Unset
Type:Personal
```

To exclude personal emails:

```
Unset
NOT Type:Personal
```

or

```
Unset
-Type:Personal
```

Example in eDiscovery:

```
Unset
Date>=2023 AND Type:Personal AND (confidential OR proprietary)
```

This query focuses on personal communications (excluding bulk emails) that mention confidential or proprietary information from 2023 onwards.

Combining Email Types

Email-type search terms can be combined for highly targeted results:

Unset

```
Date>=2023-01-01 AND Date<=2023-06-30 AND Type:Unpurged AND NOT
Type:Duplicate AND Type:Personal AND (insider<5>trading OR
material<.>nonpublic) AND (From:executive1@company.com OR
From:executive2@company.com)
```

This complex query:

1. Restricts the date range for efficiency.
2. Includes mail that has been deleted.
3. Eliminates duplicates.
4. Focuses on personal emails.
5. Searches for discussions of insider trading or material nonpublic information using proximity.
6. Narrows down to emails from specific executives.

These advanced filtering techniques allow forensics and eDiscovery professionals to create highly targeted searches, improving the efficiency and effectiveness of their investigations. By combining filtering techniques, investigators can quickly isolate the most relevant communications from large email datasets, even when dealing with deleted items or attempting to separate personal communications from automated messages.

8. Tokenization and Stemming

Tokenization and stemming are advanced linguistic processing techniques that Aid4Mail employs to enhance search capabilities. These features are particularly valuable for forensics and eDiscovery professionals dealing with large volumes of email data where variations in language use can be critical to investigations.

Tokenization

Tokenization in Aid4Mail is the process of recognizing and matching similar lexical units (both characters and whole words) within text. It can be turned on or off using the “Tokenize” option, located directly under the “Search Query” field in the “Item filtering” settings.

Key Features:

1. **Automatic Character Matching**:
    - Matches certain characters to other similar characters. For example, punctuated terms, diacritical marks, ligatures, typographic apostrophes, quotes, and

double-quotes.

2. **Word List Tokenization**:
   - Allows specification of whole-word alternatives in a separate word list. This list can be set in the “Filters” section of the “Project Settings”.

Examples Relevant to Forensics:

1. Matching variations in company names:

```
Unset

Acme Corp
```

This could match "Acme Corp", "Acme Corp.", "ACME CORP", etc.

2. Handling different spellings:

```
Unset

naïve
```

This would match both "naive" and "naïve".

3. Dealing with punctuation variations in sensitive terms:

```
Unset

trade-secret
```

This could match "trade secret", "trade-secret", "tradesecret".

Stemming

Stemming finds words with the same root as the specified word. A dictionary for stemming can be set in the “Filters” section of the “Project Settings”.

Key Features:

1. **Dictionary-Based**:

    - Aid4Mail uses dictionaries listing words with common roots.

2. **Language Support**:
    - Multiple language dictionaries are available.

3. **Customization**:
    - Users can modify existing dictionaries or create new ones.

Usage:

- Use the ~ wildcard at the end of a word to apply stemming.

Examples in an eDiscovery Context:

1. Investigating financial misconduct:

Unset

```
embezzle~
```

This would match "embezzle", "embezzled", "embezzling", etc.

2. Searching for communication about financial irregularities:

Unset

```
steal~
```

This would find "steal", "stole", "stolen", etc.

3. Identifying discussions about legal matters:

Unset

```
sue~
```

This could match "sue", "sued", and "suing".

Combining Tokenization and Stemming

When both tokenization and stemming are enabled, Aid4Mail provides powerful linguistic processing capabilities. It handles variations in spelling, punctuation, and word forms, increasing the chances of finding relevant information even when the exact phrasing is unknown.

Best Practices for Forensics and eDiscovery

1. **Use with Caution**: While these features can broaden searches, they may also introduce false positives. Always review results carefully.

2. **Document Your Settings**: Clearly document tokenization and stemming settings used in your searches for defensibility.

3. **Language Consideration**: Ensure you're using appropriate stemming dictionaries for the language(s) of your investigation.

4. **Iterative Approach**: Use these features to discover variations in how key concepts are discussed, then refine your searches accordingly.

5. **Combine with Other Techniques**: Use tokenization and stemming in conjunction with proximity searches and Boolean operators for highly targeted queries.

By effectively leveraging tokenization and stemming, forensics and eDiscovery professionals can create more comprehensive and nuanced search strategies, potentially uncovering relevant communications that might be missed by exact-match searches alone. However, it's crucial to balance the broader reach these techniques provide with the precision required in legal and investigative contexts.

9. Search Lists

Search lists in Aid4Mail are a powerful feature that allow forensics and eDiscovery professionals to manage and reuse large sets of search terms efficiently. They are particularly useful when dealing with complex investigations involving numerous keywords, email addresses, EDRM MIH values, or other identifiers.

Creating and Using Search Lists

1. **File Format**:
    - Search lists are simple text files, with one search term per line.
    - Save them with a .txt extension.
    - Optimize your list by ordering search terms from most to least common.
    - Do not use double-quotes or parentheses around search terms unless you're looking for literal double-quote or parenthesis characters.
    - Do not use Boolean operators in search lists. Instead, prefix lines with '+' and '-' characters to group them (use with MIXED option; see the user guide for details).

    - Lines starting with a semicolon (;) or a hash (#) are treated as comments and are ignored during processing. Blank lines are also ignored.

2. **Location**:
    - Place search list files in the `SearchLists` subfolder of one of the following Aid4Mail locations:
        i. Project folder (…\Documents\Aid4Mail\Projects\[ProjectName]\SearchLists) for search lists that are project-specific.
        ii. Application data folder (…\AppData\Roaming\Aid4Mail6\SearchLists) for search lists to be shared by all projects.
        iii. Program folder (C:\Program Files (x86)\Aid4Mail6\SearchLists)
    - Alternatively, place them in any accessible location on your system. In this case you will need to specify the fully qualified path to the file.

3. **Using Search Lists in Queries**:
    - Use the SearchList modifier inside braces:

Unset

```
{SearchList=KeywordList.txt}
```

    - Specify the full path unless the file is in the `SearchLists` subfolder of your Aid4Mail project folder, application data folder, or program folder:

Unset

```
{SearchList=C:\Investigations\Case123\KeywordList.txt}
```

Advanced Search Techniques in Lists

Search lists support various advanced search techniques:

1. **Wildcards**: Example: `confiden*` (matches confident, confidential, etc.)
2. **Stemming**: Example: `policy~` (matches policy and policies)
3. **Regular Expressions**: Example: `{[R]=\b[A-Z0-9._%+-]+@[A-Z0-9.-]+\.[A-Z]{2,}\b}` (matches email addresses)

These techniques allow for powerful and flexible searches directly from your search lists.

Search List Behavior

Modify search list behavior by adding properties to the first line, prefixed with an exclamation mark (!). If a property is absent, its default value will be used. Available properties include:

1. **CASE**: Controls case sensitivity.
   - CASE=DEFAULT (follows Aid4Mail's general settings)
   - CASE=SENSITIVE
   - CASE=INSENSITIVE

2. **OPERATOR**: Defines how search terms are combined.
   - OPERATOR=MIXED (default option; similar to OR option, but allows term grouping)
   - OPERATOR=OR (matches any search term in the list)
   - OPERATOR=AND (requires all search terms in the list to match)

3. **STEMMING**: Controls word stemming.
   - STEMMING=DEFAULT (follows Aid4Mail's general settings)
   - STEMMING=[File path] (turns on stemming for this list, using the specified dictionary file)
   - STEMMING=NO (turns off stemming for this list)

4. **TOKENIZATION**: Controls tokenization.
   - TOKENIZATION=DEFAULT (follows Aid4Mail's general settings)
   - TOKENIZATION=YES (turns on tokenization for this list)
   - TOKENIZATION=NO (turns off tokenization for this list)

Example of a search list file with properties:

Unset

```
!CASE=INSENSITIVE OPERATOR=OR STEMMING=DEFAULT TOKENIZATION=YES
dirty money
placement layering integration
smurf*
structured transaction*
launder~
conceal~
offshore<+2>account*
real<+2>estate<+3>transaction*
```

Examples of Search Lists in Forensics and eDiscovery

1. **List of Suspects**:

Unset

```
Subject:(confidential OR classified) AND
{SearchList=SuspectEmails.txt}
```

Searches for emails from suspects with specific subject terms.

2. **Industry-Specific Terminology**:

Unset

```
Type:Personal AND {SearchList=FinancialTerms.txt}
```

Finds personal emails containing financial terms from your list.

3. **Code Names in Corporate Investigations**:

Unset

```
From:executive@company.com AND {SearchList=ProjectCodeNames.txt}
```

Searches for executive emails mentioning project code names.

4. **Combining Multiple Lists**:

Unset

```
Date>=2023 AND {SearchList=SuspectEmails.txt} AND
{SearchList=SensitiveKeywords.txt}
```

Combines lists of suspects and sensitive keywords, focusing on recent communications.

5. **Exclusion List**:

Unset

```
("trade secret" OR confidential) AND NOT
{SearchList=ExclusionTerms.txt}
```

Finds emails about trade secrets or confidential information, excluding certain terms.

Sample Search Lists

Aid4Mail 6 includes sample Search Lists to streamline your investigations. You can find them in the `SearchLists` subfolder of your Aid4Mail program folder (C:\Program Files (x86)\Aid4Mail6\SearchLists). Samples include:

- Compliance Monitoring.txt
- Corruption and Bribery.txt
- Email-Based Attack Vectors.txt
- HR Investigation.txt
- IP Theft.txt
- Missing Person.txt
- Money Laundering.txt
- Sexual Harassment.txt

These sample search lists serve as examples and are not intended for direct use. They are starting points, and you may want to adjust, remove, or add search terms to suit your specific needs.

You can customize a sample search list by copying it to the `SearchLists` subfolder of either your Aid4Mail application data folder or, if it's project-specific, your project folder. Modifying it in either of these locations will leave the original unaltered.

To use a sample search list in your query, apply the following syntax:

Unset

```
{SearchList=Money Laundering.txt}
```

This allows you to integrate pre-defined keyword lists directly into your searches, enhancing the efficiency and accuracy of your email investigations.

Benefits of Search Lists for Forensics and eDiscovery

- Easy updating of search terms without modifying the main query.

- Reusability across multiple investigations or team members.
- Efficient management of complex searches involving numerous terms.
- Version control and documentation as part of the investigative process.
- Flexibility to incorporate advanced search techniques within lists.

By leveraging search lists effectively, forensics and eDiscovery professionals can create more flexible, maintainable, and powerful search strategies in Aid4Mail. This enhances the ability to efficiently process large volumes of email data and adapt to the specific needs of each investigation.

This is only a brief overview of what you can do with search lists. Please refer to the Aid4Mail User Guide to learn more.

10. Best Practices and Tips

Optimizing Filter Performance

1. **Query Structure**: For optimal results, structure your search queries in the following order.
    a. Start with date restrictions (Date:, Received:, Newer_than:, etc.) and folder restrictions (In:, Label:, FolderName:).
    b. Follow with metadata filters (Type:Replied, Type:Personal, -Type:Duplicate).
    c. Then header field searches (From:, To:, Subject:).
    d. Next, whole-header and message-body searches (Header:, SenderMessage:, Message:).
    e. Broad searches that include the contents of local email attachments.
    f. The slowest option involves searching the contents of cloud attachments, which must be downloaded before their contents can be analyzed.

Note that you should always place broad content searches, especially those using proximity operators, last.

Example:

Unset

```
Date>=2023-01-01 AND Date<=2023-12-31 AND Type:Personal AND NOT
Type:Duplicate AND From:executive@company.com AND
(confidential<10>project)
```

2. **Use Folder Filters**: If you're only interested in specific folders, use folder filters early in your query to reduce the dataset.

3. **Leverage Search Lists**: For complex investigations with many search terms, use search lists to manage your queries more efficiently.

4. **Incremental Searching**: Start with broader searches and progressively narrow down results, especially in large datasets.

Enhancing Search Accuracy

1. **Combine Techniques**: Use a mix of exact matches, wildcards, proximity searches, and stemming to catch various forms of relevant content.

2. **Regular Expressions**: For complex patterns (like specific formatting of sensitive information), use regular expressions.

3. **Tokenization Awareness**: Remember that tokenization can affect search results. Document your tokenization settings for defensibility.

4. **Date Formats**: Use the most efficient date format for your needs. Remember Aid4Mail supports [multiple variations of the International Date Format](#).

Handling Large Datasets

1. **Segmented Approach**: For extremely large datasets, consider breaking your search into segments (e.g., by date ranges or sender domains).

2. **Use of Exclusions**: Sometimes it's more efficient to exclude irrelevant data first before searching for relevant content. Example:

Unset

```
NOT Type:Newsletter AND NOT From:*@marketing.com AND
(confidential OR proprietary)
```

3. **Deduplication**: Always consider using `NOT Type:Duplicate` to eliminate redundant data, unless you specifically need to examine duplicates.

Forensic Considerations

1. **Chain of Custody**: Document all search parameters, including any modifications to search lists or stemming dictionaries.

2. **Reproducibility**: Ensure your search process is documented in a way that allows for exact reproduction of results.

3. **False Positives/Negatives**: Be aware of the potential for both. Regularly sample your results to verify accuracy.

4. **Unpurged Mail**: In investigations where deleted content is crucial, remember to use `Type:Unpurged` to include this data.

eDiscovery Specific Tips

1. **Custodian Focus**: Use the From: and To: operators effectively to focus on key custodians.
   Example:

Unset

```
(From:custodian1@company.com OR To:custodian1@company.com) AND
Subject:(contract OR agreement)
```

2. **Privileged Content**: Develop robust search lists for identifying potentially privileged content early in the process.

3. **Date Ranges**: Be precise with date ranges to align with the scope of discovery requests.

4. **Iterative Process**: Work closely with legal teams to refine searches based on initial results and emerging case strategy.

Common Pitfalls to Avoid

1. **Overly Broad Searches**: Avoid using very common words or short strings that may appear in many irrelevant emails.

2. **Ignoring Context**: Remember that keywords alone may miss contextual relevance. Use proximity searches to maintain context.

3. **Overlooking Variations**: Don’t rely solely on exact matches. Consider regional spelling differences, common typos, and abbreviations. Take advantage of Aid4Mail's advanced

linguistic processing.

4. **Neglecting Non-Text Content**: Remember to search for relevant non-textual content like specific attachment types, names, or sizes.
   Example:

Unset

```
AttachmentNames:*.pdf AND Size>5M
```

5. **Assuming Completeness**: Always be aware that no search is perfect. Regularly review and refine your approach based on results.

By following these best practices and tips, forensics and eDiscovery professionals can maximize the effectiveness of their Aid4Mail searches, ensuring more thorough, efficient, and defensible investigations. Remember that the key to successful email analysis often lies in the iterative refinement of search strategies based on initial findings and evolving case requirements.

This document has provided an overview of the most important elements of Aid4Mail's filter syntax. But there's much more, including powerful, time-saving features like native pre-acquisition filtering. For full details, please refer to the Aid4Mail User Guide.