

Structured Analytics – Email threading and TND Setup Basics

This guide contains workflows to provide you with basic information for setting up the following:

- Email threading only
- Textual near duplicate identification (TND) only
- Email threading and TND

For more detailed information, refer to the [documentation](#).

Email threading only setup

The setup for running email threading is comprised of five components:

1. Workspace fields
2. Saved search
3. Analytics profile
4. Structured analytics set
5. Views

1. Workspace fields setup

Your template should already have the fields listed below, so you may not need to create these fields. However, if you plan on using email threading with multiple structured analytics sets, you must create these relational fields for each structured analytics set (SAS).

Field Name	Field Type	Relational?
Email Duplicate ID	Fixed Length Text	Yes
Email Thread Group	Fixed Length Text	Yes

Click [here](#) to download the relational icons zip file.

2. Saved Search Setup

Use the following conditions and fields to create the saved search used for email threading. You don't need to set a sort order on this search.

Search Name

There is no recommendation for the saved search name. Follow your team’s normal protocol for naming searches.

Conditions

- From is set*
- Parent Document ID is not set**
- Include Families

*You can use any condition here which means “document is an email”. Other common choices are based on the Relativity Native File Type or file extension.

**Generally, the Parent ID field is empty for parent emails. However, there may be situations where it is populated with its own control number. In that case, you may omit this condition.

Fields

Any fields are acceptable. As a quality control step, we recommend including the fields that will be mapped on the analytics profile as there may be fields that are named similarly that could lead to errors. We also recommend reviewing the Parent Document ID field. The field should be populated for children and empty for parent documents.

- Email From
- Email To
- Email CC
- Email BCC
- Email Subject
- Email Date Sent
- Parent Document ID
- File Name field

3. Analytics Profile Setup

The analytics profile is where you map the fields containing input information for email threading. Map the fields as follows:

Profile Field Name	Mapped Document Object Field
Email from field:	Email From
Email to field:	Email To
Email cc field:	Email CC
Email bcc field:	Email BCC
Email subject field:	Email Subject

Email date sent field	Email Date/Time Sent
Parent document ID field:	Parent Document ID
Attachment name field:	File Name
Conversation ID field:	DO NOT MAP

Analytics Profile Information

Name: Standard Analytics Profile

Email Threading Settings ?

Email header fields:

Email from field: Email From

Email to field: Email To

Email cc field: Email CC

Email bcc field: Email BCC

Email subject field: Email Subject

Email date sent field: Sent Date/Time

Email metadata fields:

Parent document ID field: Parent Document ID

Attachment name field: File Name

Conversation ID field:

Notes about fields:

- The email date sent field should contain a date/time and not just a date. Email threading will produce degraded results if a date-only field is mapped.
- Set the Parent Document ID field to the control number of the parent for all child documents. The parent documents may have this field set to their own control number or set to blank.
- The Attachment name field should contain the file name for any attachment. It is ignored for non-attachments.

4. Creating a structured analytics set

Here are the steps and choices for creating a structured analytics set.

Structured Analytics Set Information

- **Structured analytics set name** - enter a name for the structured analytics set.
- **Set prefix** - keep the default prefix or add your own prefix. Shorter prefixes (even just two characters, e.g. "E1") take up less space in your views.
- **Select document set to analyze** - select the saved search you created above.
- **Select operations** - select **Email threading**.

Structured Analytics Set Information ?

Structured analytics set name:

Set prefix
The prefix you enter here is attached to SAS fields and objects to help you identify them:

Enter email addresses
Separate email addresses with a semi-colon:

Select document set to analyze:

Select operations:

- ☒ Email threading
- ☐ Textual near duplicate identification
- ☐ Language identification
- ☐ Repeated content identification

Email threading fields

- **Select profile for field mappings** - select the profile you created above.
- **Use email header fields** - select **Yes**.
- **Email header language** - select any language used in email headers in your extracted text. We recommend that you always include English, even for data sets in other languages.
- **Destination Email Thread Group** - select the default field or the custom field you created.
- **Destination Email Duplicate ID** - select the default field or the custom field you created.

Email Threading ?

Select profile for field mappings:

Use email header fields: ☒ Yes ☐ No

Email header languages: ☒ English ☐ Chinese ☐ French ☐ German ☐ Japanese ☐ Korean ☐ Portuguese ☐ Spanish

Destination Email Thread Group:

Destination Email Duplicate ID:

Advanced Settings

- **Analytics server** - select the appropriate Analytics server.

5. Email Threading views

Once you run the structured analytics set, create the following views to view the results of your email threading operations. There are two document list views and one relational items view.

Email Threading View

This view returns emails that have been threaded along with any duplicate spares and family members.

Fields

- Edit
- Control Number
- [SAS Prefix]::Email Threading Display
- [SAS Prefix]::Email Thread Group
- [SAS Prefix]::Email Threading ID
- [SAS Prefix]::Inclusive Email
- [SAS Prefix]::Inclusive Reason
- [SAS Prefix]::Email Duplicate Spare*
- [SAS Prefix]::Email Duplicate ID*
- [SAS Prefix]::Indentation*

*Fields marked with an asterisk are not required for an email threading view, but they will be helpful for viewing duplicate spares.

Conditions

- [SAS Prefix]::Email Thread Group is Set

Sort

- [SAS Prefix]::Email Thread Group – Ascending Order
- [SAS Prefix]::Indentation – Ascending Order (some prefer Descending)
- [SAS Prefix]::Email Threading ID – Ascending Order

Group Definition

- Email Thread Group

#		Control Number	SAS01::Email Threading Display	SAS01::Email Thread Group	SAS01::Inclusive Email	SAS01::Inclusive Reason	SAS01::Email Duplicate Spare	SAS01::Indentation
		Filter	Filter	Filter	(All)	(All)	(All)	= Filter
1	Edit	Kaminski_0000003046	Shirley Crenshaw Greg Whalley (Bi-Weekly Meeti	A00000001	Yes	MESSAGE	No	1
2	Edit	Kaminski_0000020349	Shirley Crenshaw Greg Whalley (Bi-Weekly Meeti	A00000002	Yes	MESSAGE	No	1

Inclusive Non-Duplicate Spare Emails with Family Members

Fields

- Edit
- Control Number
- [SAS Prefix]::Email Threading Display
- [SAS Prefix]::Email Thread Group
- [SAS Prefix]::Inclusive Reason

- [SAS Prefix]::Email Duplicate ID
- [SAS Prefix]::Indentation

Conditions

Create the following saved search, and then associate the search with the view using the Saved Search condition.

- Name
 - [SAS Prefix] Inclusive Non-Dup
- Search conditions
 - [SAS Prefix]::Email Thread Group is Set AND
 - [SAS Prefix]::Inclusive Email is Yes AND
 - [SAS Prefix]::Email Duplicate Spare is No AND
 - Related Items = + Family
- Fields
 - Control Number

Sort

- [SAS Prefix]::Email Thread Group – Ascending Order
- [SAS Prefix]::Indentation – Ascending Order (some prefer Descending)
- [SAS Prefix]::Email Threading ID – Ascending Order

Group Definition

- Email Thread Group

#		Control Number	SAS01::Email Threading Display	SAS01::Email Thread Group	SAS01::Inclusive Email	SAS01::Inclusive Reason	SAS01::Email Duplicate Spare	SAS01::Email Duplicate ID	SAS01::Indentation
		Filter	Filter	Filter	(All)	(All)	(All)	Filter	= Filter
1	Edit	Kaminski_0000031885	1 Vince J Kaminski	A00000003	Yes	MESSAGE	Yes	email-1	1
2	Edit	Kaminski_0000010836	1 Vince J Kaminski	A00000004	Yes	MESSAGE	Yes	email-2	1

Relational Items View

The Relational Items View is located within the viewer and is used to navigate to documents within a specific thread. The view is associated with the Email Thread Group relational field.

Fields

- Control Number
- [SAS Prefix]::Email Threading Display

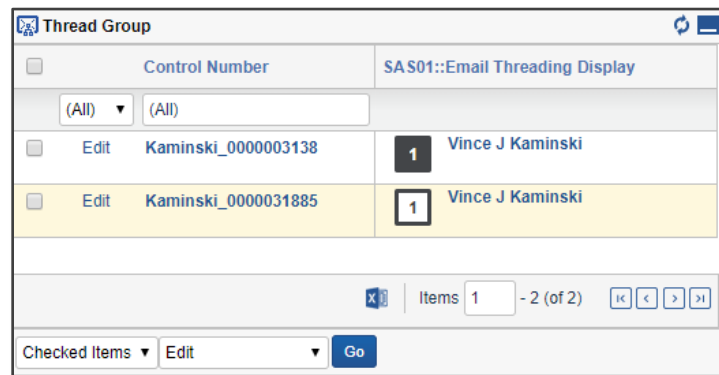
Conditions

- No conditions can be set on a relational view

Sort

- [SAS Prefix]::Email Thread Group – Ascending Order

- [SAS Prefix]::Indentation – Ascending Order (some prefer Descending)
- [SAS Prefix]::Email Threading ID – Ascending Order



Textual near duplicate identification (TND) only setup

The setup for running textual near duplicate identification is comprised of four components:

1. Workspace fields
2. Saved search
3. Structured analytics set
4. Views

1. Workspace Fields Setup

Your template should already have the fields listed below, so you may not need to create these fields. However, if you plan on using TND within multiple structured analytics sets, you must create these relational fields for each structured analytics set.

Field Name	Field Type	Relational?
Textual Near Duplicate Group	Fixed Length Text	Yes

Click [here](#) to download the relational icons zip file.

2. Saved Search Setup

Use the following conditions and fields to create the saved search used for TND. You don't need to set a sort order on this search.

Search Name

There is no recommendation for the saved search name. Your team's normal protocol for naming searches will work here.

Conditions*

- Record type is not email**

*We recommend not setting an extracted text size on the search. As a result, all documents without text are grouped together, which can help you identify documents that don't have text but should.

**You can use whatever condition is necessary to identify documents that are only attachments or loose files.

Fields

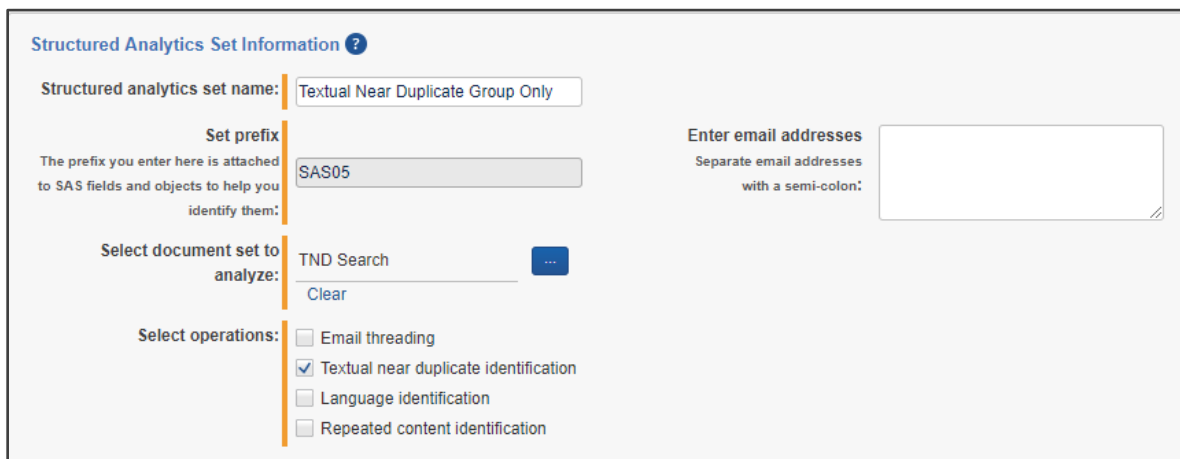
Any fields are acceptable.

3. Creating a structured analytics set

Here are the steps and choices for creating a structured analytics set.

Structured Analytics Set Information

- **Structured analytics set name** - enter a name for the structured analytics set.
- **Set prefix** - keep the default prefix or add your own prefix. Shorter prefixes (even just two characters, such as "E1") take up less space in your views.
- **Select document set to analyze** - select the saved search you created above.
- **Select operations** - select **Textual near duplicate identification**.



Structured Analytics Set Information ?

Structured analytics set name:

Set prefix
The prefix you enter here is attached to SAS fields and objects to help you identify them:

Enter email addresses
Separate email addresses with a semi-colon:

Select document set to analyze:

Select operations:

- ☐ Email threading
- ☒ Textual near duplicate identification
- ☐ Language identification
- ☐ Repeated content identification

Textual Near Duplicate Identification fields

- **Minimum similarity percentage** - you can keep the default value or change it to a value between 80 and 100.
- **Ignore numbers** - select **No** unless you have a very large document set that will take a long time to run.
- **Destination Textual Near Duplicate Group** - Textual Near Duplicate Group

Textual Near Duplicate Identification ?

Minimum similarity percentage
Enter a value between 80 and 100:

Ignore numbers: ☒ Yes ☐ No

Destination Textual Near Duplicate Group:

Advanced Settings

- Choose the appropriate Analytics server

4. TND views

A document list view and a relational view can be set up after the structured analytics set has been run.

Document List View

Fields

- Edit
- Control Number
- [SAS Prefix]::Textual Near Duplicate Group
- [SAS Prefix]::Textual Near Duplicate Similarity
- [SAS Prefix]::Textual Near Duplicate Principal
- Relativity Compare

Conditions

- [SAS Prefix]::Textual Near Duplicate Group is Set

Sort

- [SAS Prefix]::Textual Near Duplicate Group – Ascending Order
- [SAS Prefix]::Textual Near Duplicate Similarity – Descending Order
- [SAS Prefix]::Textual Near Duplicate Principal – Descending Order

Group Definition

- Textual Near Duplicate Group

#		Control Number	SAS05::Textual Near Duplicate Group	SAS05::Textual Near Duplicate Principal	SAS05::Textual Near Duplicate Simil...
		<input type="text" value="Filter"/>	<input type="text" value="Filter"/>	(All) <input type="button" value="Filter"/>	<input type="button" value="Filter"/>
1		Edit B_RAPP0000000021.0001	B_RAPP0000000021.0001	Yes	100
2		Edit B_RAPP00000000374.0001	B_RAPP0000000021.0001	No	100
3		Edit B_RAPP0000000031.0001	B_RAPP0000000031.0001	Yes	100
4		Edit B_RAPP00000000119.0001	B_RAPP0000000031.0001	No	100

Relational Items View

The Relational Items View is located within the viewer and is used to navigate to documents within a specific thread. The view is associated to the [SAS Prefix]::Textual Near Duplicate Group field.

Fields

- Control Number
- [SAS Prefix]::Textual Near Duplicate Group
- [SAS Prefix]::Textual Near Duplicate Similarity
- [SAS Prefix]::Textual Near Duplicate Principal
- Relativity Compare

Conditions

- Conditions cannot be set on a relational items view.

Sort

- [SAS Prefix]::Textual Near Duplicate Group – Ascending Order
- [SAS Prefix]::Textual Near Duplicate Similarity – Descending Order
- [SAS Prefix]::Textual Near Duplicate Principal – Descending Order

<

Email Threading & TND Setup

The setup for running email threading and TND together is comprised of five components:

1. Workspace fields
2. Saved search
3. Analytics profile
4. Structured analytics set
5. Views

1. Workspace Fields Setup

The template should already have the fields listed below so there may not be a need to create these fields. However, if you plan on using multiple structured analytics sets, you will need to create these fields for each structured analytics set.

Field Name	Field Type	Relational?
Email Duplicate ID	Fixed Length Text	Yes
Email Thread Group	Fixed Length Text	Yes
Textual Near Duplicate Group	Fixed Length Text	Yes

Click [here](#) to download the relational icons zip file.

2. Saved Search Setup

Here are the conditions and fields for the saved search that will be used for email threading and TND. It is not necessary to set a sort order on this search.

Search Name

There is no recommendation for the saved search name. Your team's normal protocol for naming searches will work here.

Conditions

No conditions are set for this search as you want to include all attachments and all loose files.

Fields

Any fields are acceptable. As a QC step, we recommend including the fields that will be mapped on the analytics profile as there may be fields that are named similarly that could lead to errors.

- Email From
- Email To
- Email CC
- Email BCC
- Email Subject
- Email Date Sent
- Parent Document ID
- File Name field

3. Analytics Profile Setup

The analytics profile is where you map the fields containing input information for email threading. Map the fields as follows:

Profile Field Name	Mapped Document Object Field
Email from field:	Email From
Email to field:	Email To
Email cc field:	Email CC
Email bcc field:	Email BCC
Email subject field:	Email Subject
Email date sent field	Email Date/Time Sent
Parent document ID field:	Parent Document ID
Attachment name field:	File Name
Conversation ID field:	DO NOT MAP

4. Structured Analytics Set

Here are the steps and choices for creating your structured analytics set.

Structured Analytics Set Information

- **Structured analytics set name** - enter a name for the structured analytics set.
- **Set prefix** - keep the default prefix or add your own prefix. Shorter prefixes (even just two characters, such as “E1”) take up less space in your views.
- **Select document set to analyze** - select the saved search you created above.
- **Select operations** - select **Email threading** and **Textual near duplicate identification**.

Email Threading fields

- **Select profile for field mappings** - Select the profile you created above.
- **Use email header fields** - select **Yes**.
- **Email header language** - select any language used in email headers in your extracted text. We recommend that you always include English, even for data sets in other languages.
- **Destination Email Thread Group** - select the default field or the custom field you created.
- **Destination Email Duplicate ID** - select the default field or the custom field you created.

Textual Near Duplication fields

- **Minimum similarity percentage** - you can keep the default value or change it to a value between 80 and 100.
- **Ignore numbers** - select **No** unless you have a very large document set that will take a long time to run.

- **Destination Textual Near Duplicate Group** - Textual Near Duplicate Group

Advanced Settings

- Choose the appropriate Analytics server

5. Views

Refer to the Email threading only and Textual near duplicate identification only sections for creating views.