Paper SAS2015-2018

Diamonds in the Rough: New Discoveries with the SAS® Visual Investigator Text Analytics Workspace

Danielle Davis, SAS Institute Inc., Cary, NC

ABSTRACT

Take your investigative process to a new dimension and incorporate the analysis of your unstructured data. Adding text analytics to your process gives you the potential to discover new relationships between existing or newly discovered entities. Discover new locations, organizations, or people that can further your investigative process. Uncover new patterns of common phrases that were previously undetected. Let the analytics reveal common themes that might cluster the unstructured data into meaningful results. The new “Text Analytics Workspace” in SAS® Visual Investigator provides these features to enhance your investigation.

INTRODUCTION

Unstructured information (emails, claim documents, social media posts, text messages, videos, and so on) plays an important role in the investigation process. Unstructured data can easily represent 80% or more of potential data to investigate. Unstructured data is where undetected fraudulent activities can often be found. Text analytics plays an important role in uncovering hidden patterns from unstructured data and preventing wrongdoing. Text analytics is a tremendously effective technology in any domain where the majority of information collected is unstructured. SAS® Visual Investigator now includes a new investigative workspace for text analytics. This workspace enables the investigator to analyze any type of text field defined in an entity object. The text analytics workspace parses this text to reveal meaningful data and analyzes the newly created data to provide a deeper understanding of the content. This new workspace incorporates natural language processing (NLP) to reveal people, places, and organizations mentioned in the text. This workspace facilitates the integration of unstructured text data with structured data.

This paper explores the use and benefits of the new text analytics workspace of SAS Visual Investigator. This workspace can extract entities of interest, detecting people, places, organizations, companies, and entity characteristics that enable analysts to uncover relationships that would previously go unidentified. In addition to all of this, it enables investigators to organize and group data into a topic, theme, or category. Each topic is further represented by words, phrases, entities, and so on. The presence of these words or phrases in a document implies the occurrence of a concept or theme in a cluster of documents. The end result of using the text analytics workspace is a potential fraud dictionary that is a repository of concepts and suspicious key words.

To show the workspace features, we use the sample data of some known fraudulent emails as our entity objects to analyze.

UNSTRUCTURED DATA

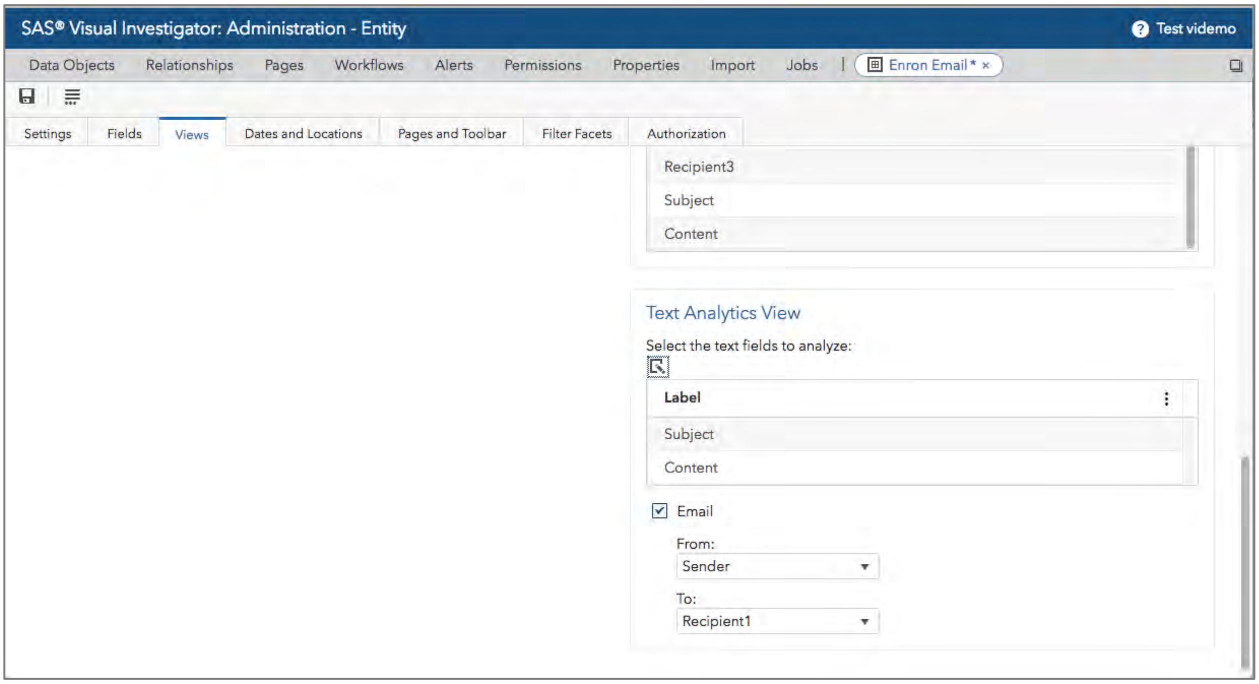
Intelligence analysts and investigators have access to large volumes of text-based information that can come in many forms, such as the following:

- Criminal intelligence reports
- Arrest records
- Adjuster notes
- Emails
- Field reports
- Case management files
- Customer service calls
- Claimant interviews
- Social media

The goal of analyzing this data is to turn these large data stores into actionable investigation intelligence. Analyzing this type of data is more than just keyword searching. The analysis can identify new patterns, trends, and topics under which fraudulent activity takes place. Traditional keyword and search-based approaches simply fail to uncover these patterns.

REGISTERING UNSTRUCTURED DATA

In order to take advantage of the text analytics workspace, your unstructured data has to be defined in the administrator component of SAS Visual Investigator. Two key areas need to be configured. The first is under the **Permissions** tab. The user's group must have permission to the following capability: "Access text analytics view in workspaces". Second, the SAS Visual Investigator administrator needs to define certain fields of an entity object to be used for textual analysis. This is done under the **Views** tab of an entity registration. If the entity is an email, the sender and recipient (To/From) fields can be defined as well, as shown in Figure 1:



Text Analytics View

Select the text fields to analyze:

Label
Subject
Content

☑ Email

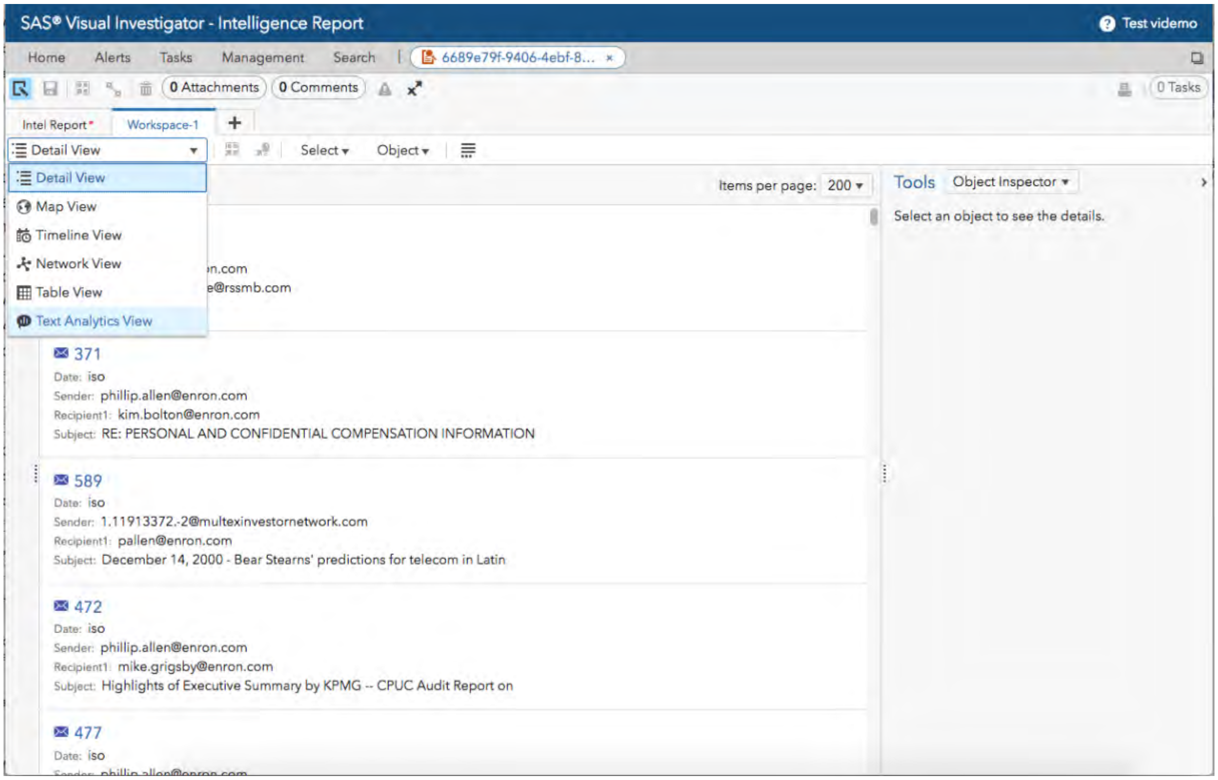
From: Sender

To: Recipient1

Figure 1: Registering Unstructured Data in the Administrator

ANALYZING THE UNSTRUCTURED DATA

Now that the entity has been configured for the **Text Analytics View**, the investigator can add those entity types to a workspace, and the workspace option **Text Analytics View** appears, as shown in Figure 2:



371
Date: iso
Sender: phillip.allen@enron.com
Recipient1: kim.bolton@enron.com
Subject: RE: PERSONAL AND CONFIDENTIAL COMPENSATION INFORMATION

589
Date: iso
Sender: 1.11913372.-2@multexinvestornetwork.com
Recipient1: pallen@enron.com
Subject: December 14, 2000 - Bear Stearns' predictions for telecom in Latin

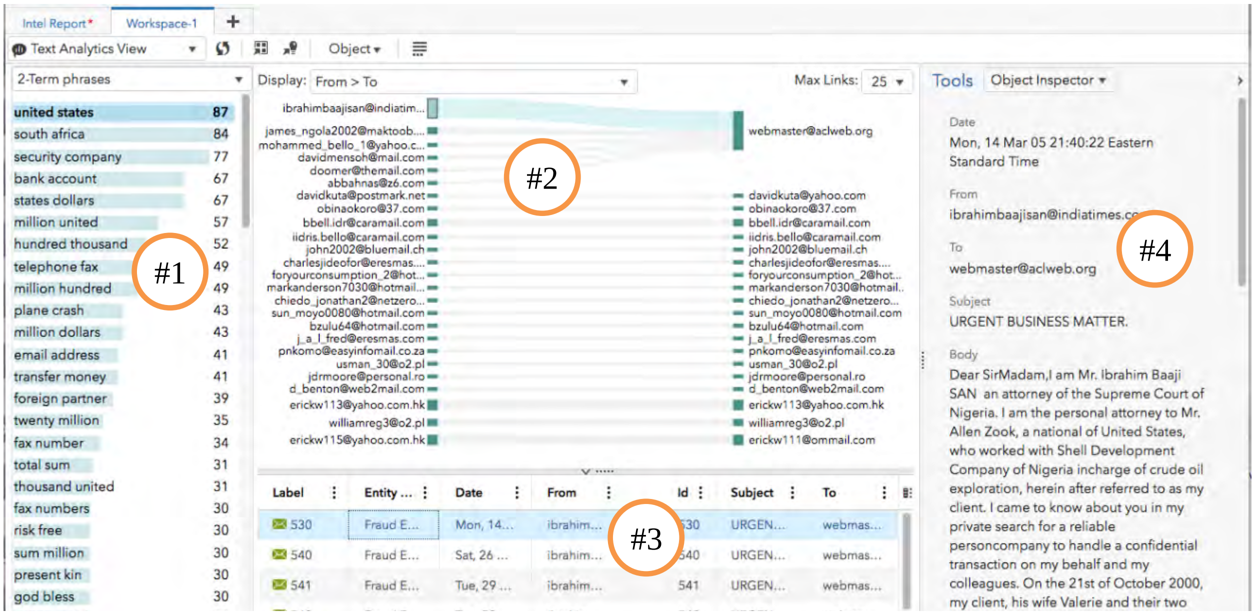
472
Date: iso
Sender: phillip.allen@enron.com
Recipient1: mike.grigsby@enron.com
Subject: Highlights of Executive Summary by KPMG -- CPUC Audit Report on

477
Date: iso
Sender: phillip.allen@enron.com

Figure 2: Selecting the Text Analytics View

TEXT ANALYTICS VIEW

Text analysis can be computationally intensive, depending on the volume of text and the number of entities and topics generated. When analysis is complete, the investigator is shown an initial view of the results. If some of the entity types have been defined as email, then the default first view of the results is similar to that shown in Figure 3:



united states	87
south africa	84
security company	77
bank account	67
states dollars	67
million united	57
hundred thousand	52
telephone fax	49
million hundred	49
plane crash	43
million dollars	43
email address	41
transfer money	41
foreign partner	39
twenty million	35
fax number	34
total sum	31
thousand united	31
fax numbers	30
risk free	30
sum million	30
present kin	30
god bless	30

Label	Entity ...	Date	From	Id	Subject	To
530	Fraud E...	Mon, 14...	ibrahim...	530	URGEN...	webmas...
540	Fraud E...	Sat, 26 ...	ibrahim...	540	URGEN...	webmas...
541	Fraud E...	Tue, 29 ...	ibrahim...	541	URGEN...	webmas...

Date
Mon, 14 Mar 05 21:40:22 Eastern Standard Time

From
ibrahimbaajisan@indiatimes.co

To
webmaster@aclweb.org

Subject
URGENT BUSINESS MATTER.

Body
Dear SirMadam,I am Mr. Ibrahim Baaji SAN an attorney of the Supreme Court of Nigeria. I am the personal attorney to Mr. Allen Zook, a national of United States, who worked with Shell Development Company of Nigeria incharge of crude oil exploration, herein after referred to as my client. I came to know about you in my private search for a reliable personcompany to handle a confidential transaction on my behalf and my colleagues. On the 21st of October 2000, my client, his wife Valerie and their two

Figure 3: Default Text Analytics View That Contains Emails

The following sections explain each of the four different areas depicted in Figure 3.

Bar Chart Area (#1)

This area of the results reveals term frequencies. The drop-down menu in Figure 4 enables the investigator to navigate between 1-,2-, and 3-term frequencies as well as the entities that have been extracted and organized by type.

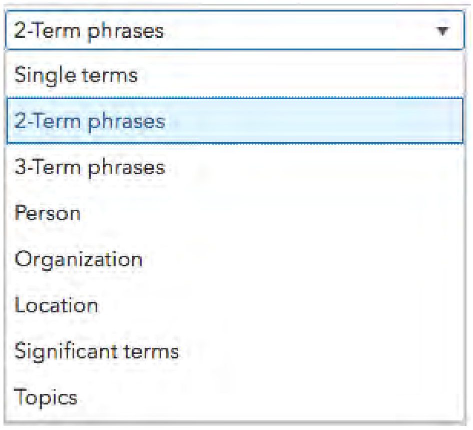


Figure 3: Drop-Down Menu for Bar Chart Results

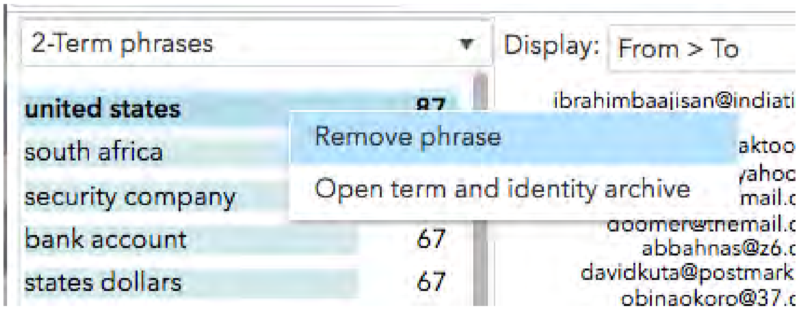
The options are the following:

- the most common 1-, 2-, and 3-term phrases
- potential discovered named entities categorized by person, location, or organization
- text mining results of statistically relevant terms and topics or themes discovered in the text

For common phrases and discovered entities, a frequency count indicates the number of occurrences for the term or phrase. **Significant terms** and **Topics** show a weighting to reveal the importance or strength of the term or topic. Common terms or phrases might be important in some cases, but if the investigator views the **Significant terms**, this list might be very different.

Sometimes a term appears so often in the documents that it becomes irrelevant—mere “noise.” Therefore, looking a term that the basic text mining results have shown as statistically relevant might prove to be more beneficial.

Since we are looking at term or phrase frequency, many times terms appear in the results that are not relevant. The **Text Analytics View** enables you to remove those unwanted terms and save that removal as part of the workspace. Therefore, you don’t have to repeat your cleanup each time you come back to the view. The bars also re-scale themselves in proportion to the results with the terms removed. You can bring those removed terms back into the analysis by selecting the **Open term and identity archive** option, as shown in Figure 5.



2-Term phrases

Display: From > To

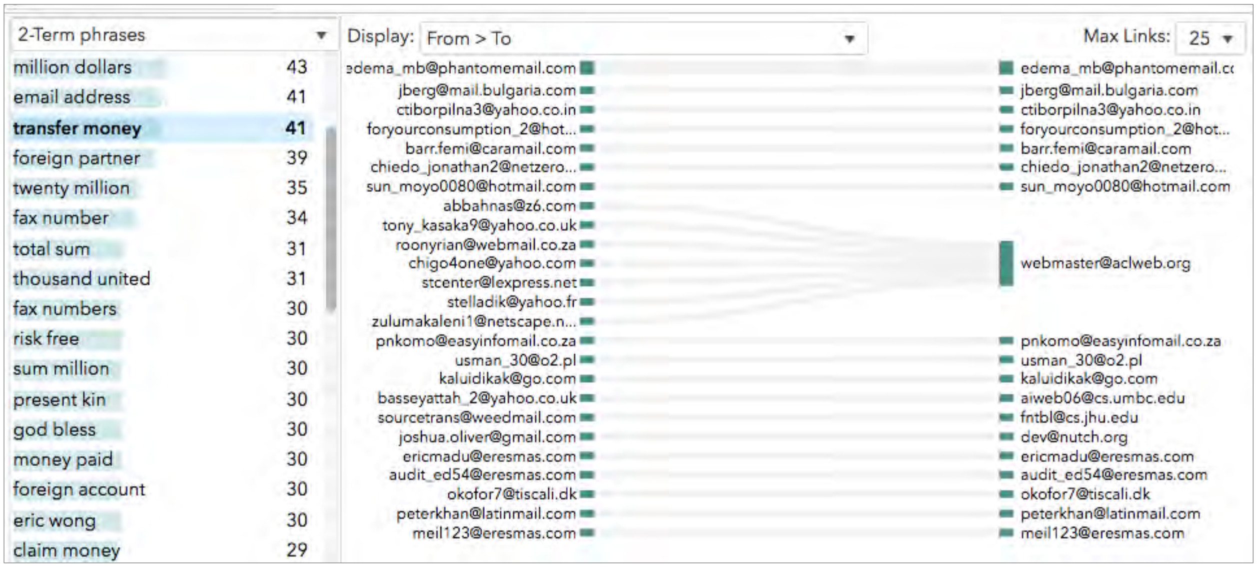
united states	87
south africa	
security company	
bank account	67
states dollars	67

Figure 4: Ability to Remove a Term or Phrase from Analysis

When an investigator selects a bar of interest, the area to the right (#2) updates its results based on that selection.

Display Area (#2)

When the user selects the bar of interest, the display area reveals a Sankey diagram if some of the entities have been defined as email.



2-Term phrases

Display: From > To

Max Links: 25

million dollars	43
email address	41
transfer money	41
foreign partner	39
twenty million	35
fax number	34
total sum	31
thousand united	31
fax numbers	30
risk free	30
sum million	30
present kin	30
god bless	30
money paid	30
foreign account	30
eric wong	30
claim money	29

Figure 5: All Email Correspondence Containing the Phrase "Transfer Money"

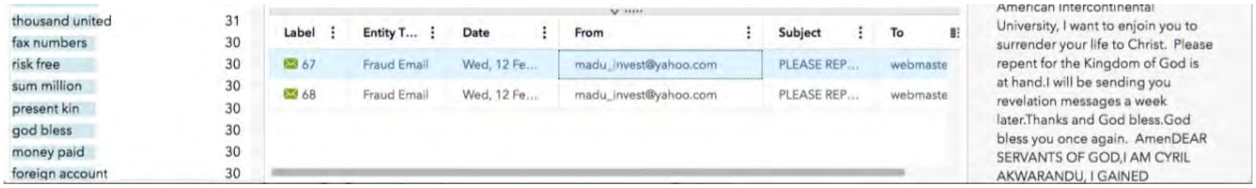
The results in Figure 6 show all the emails that contain the phrase “transfer money”. The results also show the volume of emails that were sent and received. The width of the flow from the sender to the receiver is proportional to the number of documents that were sent. Therefore, you can quickly determine which emails contain a key term or phrase that might be relevant to the investigation. This Sankey diagram always shows the top *N* links (the document count that is represented by the flow width between the emails). The **Max Links** option at the top can control the volume that is shown.

Viewing common phrases among the documents might assist in detecting a person or group of people telling the same story for different claims or statements. This can be a flag for fraudulent behavior.

Table Area (#3)

This list always shows the entities that contain the terms or phrases of interest. When the investigator selects a bar, the list reveals all the documents that contain that term or phrase. If the user selects a flow in the Sankey diagram, the list reveals all documents that contain the term or phrase *and* are from the sender and are sent to the recipient. Figure 7 shows emails that meet the following criteria:

- They contain the phrase “bank account”.
- The sender was “madu_invest@yahoo.com”.
- The recipient was “webmaster@aclweb.org”.



thousand united	31
fax numbers	30
risk free	30
sum million	30
present kin	30
god bless	30
money paid	30
foreign account	30

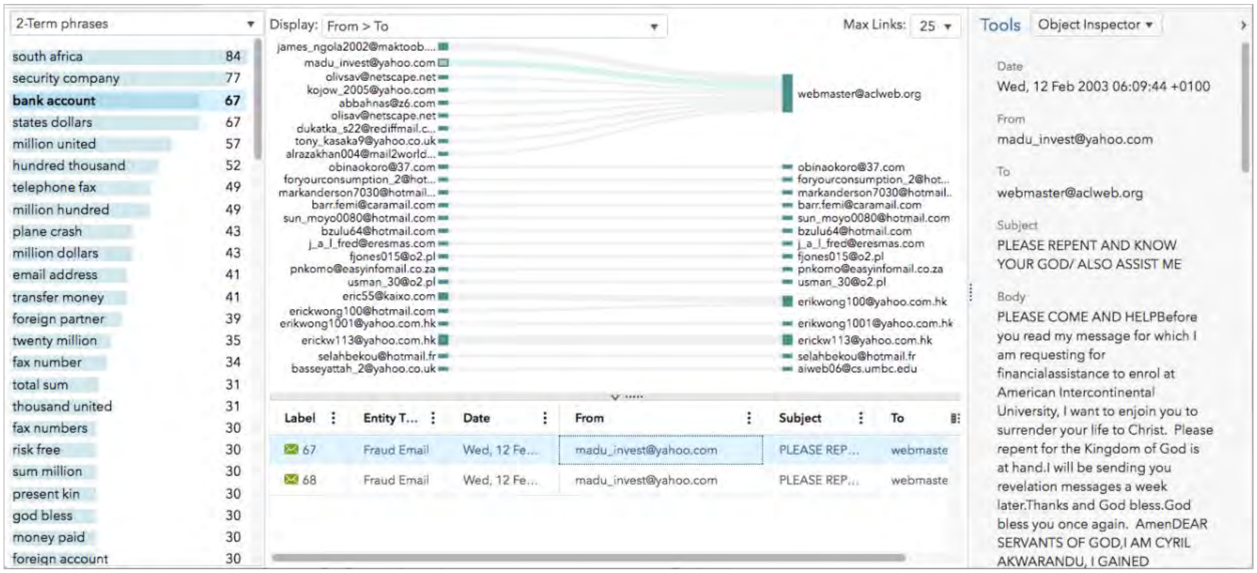
Label	Entity T...	Date	From	Subject	To
67	Fraud Email	Wed, 12 Fe...	madu_invest@yahoo.com	PLEASE REP...	webmaste
68	Fraud Email	Wed, 12 Fe...	madu_invest@yahoo.com	PLEASE REP...	webmaste

American Intercontinental University, I want to enjoin you to surrender your life to Christ. Please repent for the Kingdom of God is at hand.I will be sending you revelation messages a week later.Thanks and God bless.God bless you once again. AmenDEAR SERVANTS OF GOD,I AM CYRIL AKWARANDU, I GAINED

Figure 6**: List of Emails That Contain the Selected Term and Email Correspondents**

Object Inspector (#4)

Notice that if the investigator selects one of the entities in the list, then the **Object Inspector** reveals the details of the email. Figure 8 shows the actual email details of the selected email in the table list.



south africa	84
security company	77
bank account	**67**
states dollars	67
million united	57
hundred thousand	52
telephone fax	49
million hundred	49
plane crash	43
million dollars	43
email address	41
transfer money	41
foreign partner	39
twenty million	35
fax number	34
total sum	31
thousand united	31
fax numbers	30
risk free	30
sum million	30
present kin	30
god bless	30
money paid	30
foreign account	30

Label	Entity T...	Date	From	Subject	To
67	Fraud Email	Wed, 12 Fe...	madu_invest@yahoo.com	PLEASE REP...	webmaste
68	Fraud Email	Wed, 12 Fe...	madu_invest@yahoo.com	PLEASE REP...	webmaste

Date
Wed, 12 Feb 2003 06:09:44 +0100

From
madu_invest@yahoo.com

To
webmaster@aclweb.org

Subject
PLEASE REPENT AND KNOW YOUR GOD/ ALSO ASSIST ME

Body
PLEASE COME AND HELPBefore you read my message for which I am requesting for financialassistance to enrol at American Intercontinental University, I want to enjoin you to surrender your life to Christ. Please repent for the Kingdom of God is at hand.I will be sending you revelation messages a week later.Thanks and God bless.God bless you once again. AmenDEAR SERVANTS OF GOD,I AM CYRIL AKWARANDU, I GAINED

Figure 7: List of Emails That Contain the Selected Phrase and Email From/To

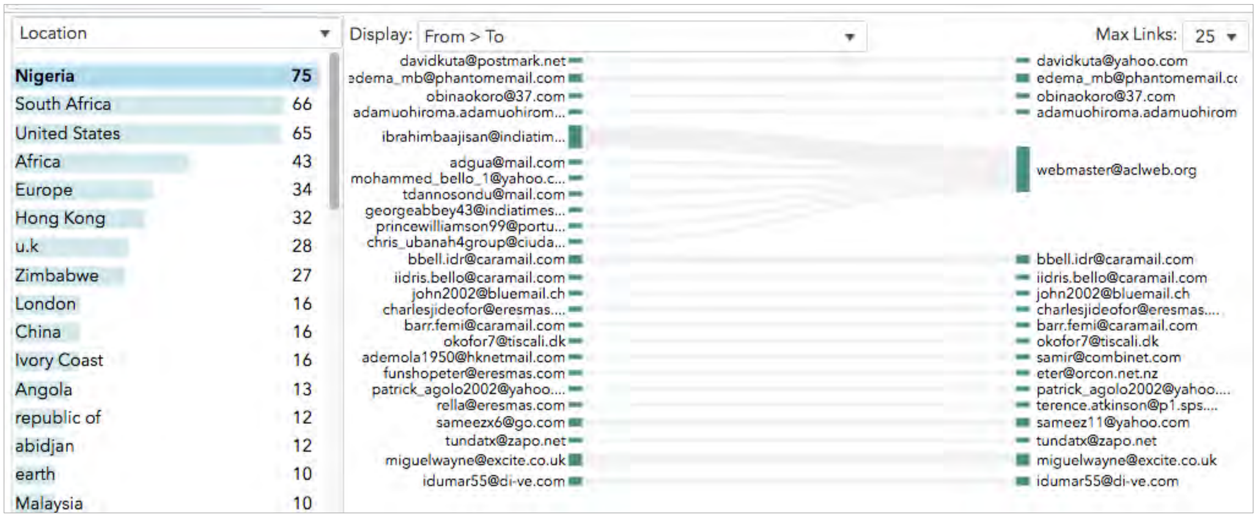
Notice that the workspace can easily group emails that contain the same phrases or terms. The investigator can also discover who is communicating this information. We can take this a step further.

TEXT ANALYTICS VISUALIZATIONS

An investigator can look at all the potential discovered named entities as well. (We refer to these entities as “identities” in this view so that they are not confused with actual entity objects defined in the SAS Visual Investigator system.) The named identities that are revealed fall into the following categories:

- people
- organizations
- locations

This information can help you discover new relationships to other persons of interest or organizations that might or might not already be in your system. In Figure 9, we see the frequency bars showing the different locations discovered in the text. We also see the emails that contained the location “Nigeria” and who sent and received them.



Location

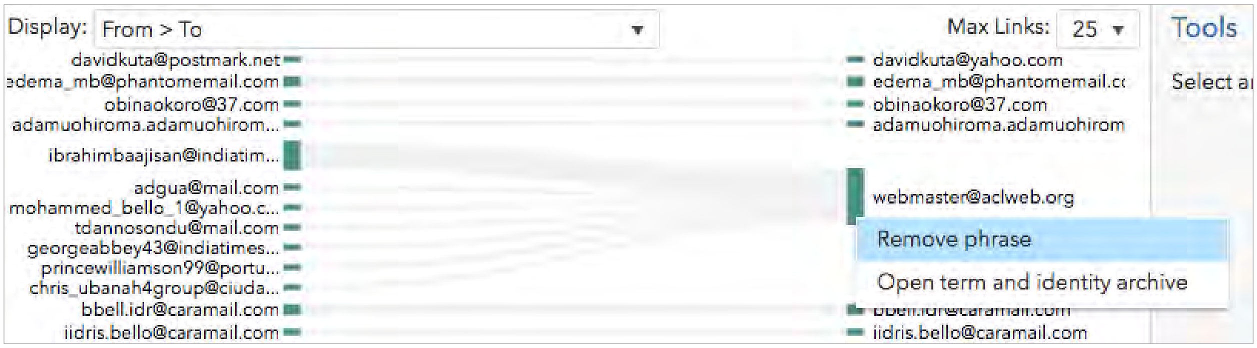
Nigeria	75
South Africa	66
United States	65
Africa	43
Europe	34
Hong Kong	32
u.k	28
Zimbabwe	27
London	16
China	16
Ivory Coast	16
Angola	13
republic of	12
abidjan	12
earth	10
Malaysia	10

Display: From > To

Max Links: 25

Figure 8: Documents Containing the Location "Nigeria"

Our entity discovery is not always accurate, and some of the discoveries that are revealed might not have any relevance to the investigation. These discoveries can easily be removed from the analysis, as mentioned earlier. Removal of terms or phrases can be performed on the Sankey diagram as well. Figure 10 shows the results before the removal, and Figure 11 shows the results after the removal.

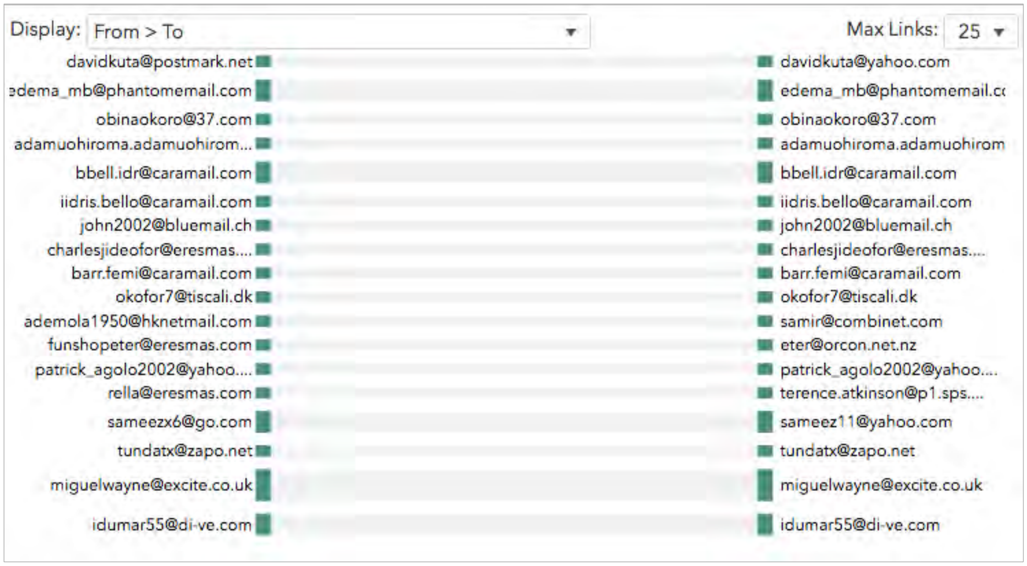


Display: From > To

Max Links: 25

Tools
Select a
davidkuta@postmark.net
edema_mb@phantomemail.com
obinaokoro@37.com
adamuohiroma.adamuohirom...
ibrahimbaajisan@indiatim...
adgua@mail.com
mohammed_bello_1@yahoo.c...
tdannosondu@mail.com
georgeabbey43@indiatimes...
princewilliamson99@portu...
chris_ubanah4group@ciuda...
bbell.idr@caramail.com
iidris.bello@caramail.com
davidkuta@yahoo.com
edema_mb@phantomemail.c
obinaokoro@37.com
adamuohiroma.adamuohirom
webmaster@aclweb.org
Remove phrase
Open term and identity archive
bbell.idr@caramail.com
iidris.bello@caramail.com

Figure 9: Before Email Is Removed from the Diagram



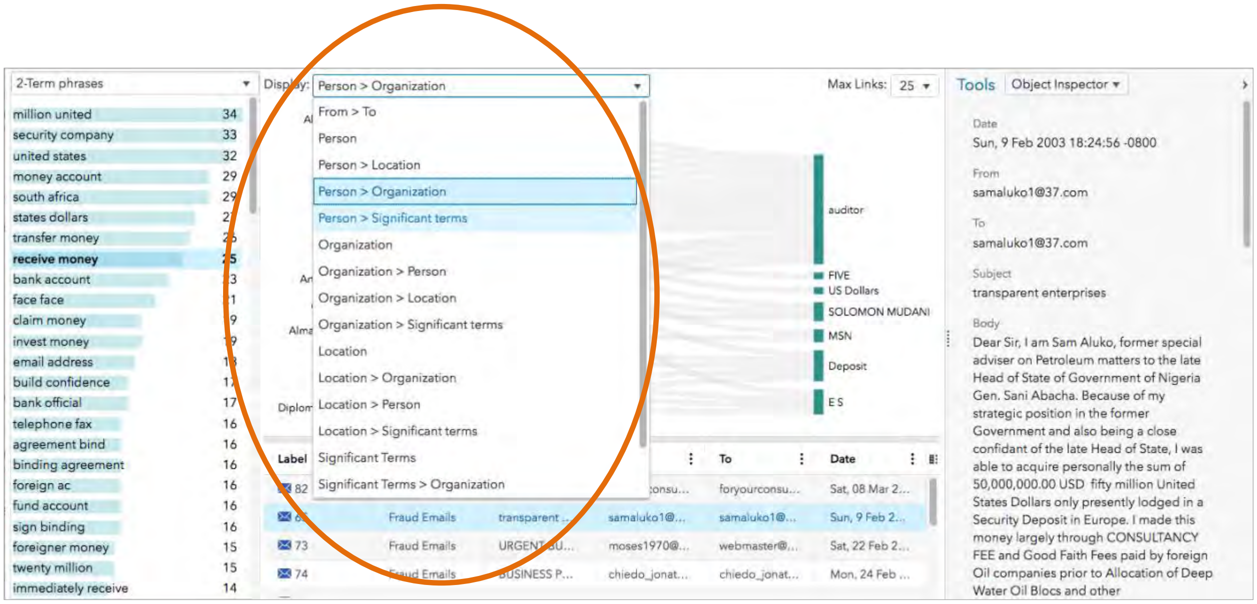
Display: From > To

Max Links: 25

Figure 10: After Email Is Removed

OTHER DISPLAY OPTIONS

On the main view, an investigator can change the diagram display also. This enables the user to view the resulting parsing and analysis from many different viewpoints. Figure 12 shows all possible options. Any option containing a “>” is displayed as a Sankey diagram to show the relationships between the two category types. All other options are displayed as a word cloud.



2-Term phrases	
million united	34
security company	33
united states	32
money account	29
south africa	29
states dollars	27
transfer money	25
receive money	**25**
bank account	23
face face	21
claim money	19
invest money	19
email address	18
build confidence	17
bank official	17
telephone fax	16
agreement bind	16
binding agreement	16
foreign ac	16
fund account	16
sign binding	16
foreigner money	15
twenty million	15
immediately receive	14

Display: Person > Organization

Max Links: 25

Tools Object Inspector

Date
Sun, 9 Feb 2003 18:24:56 -0800

From
samaluko1@37.com

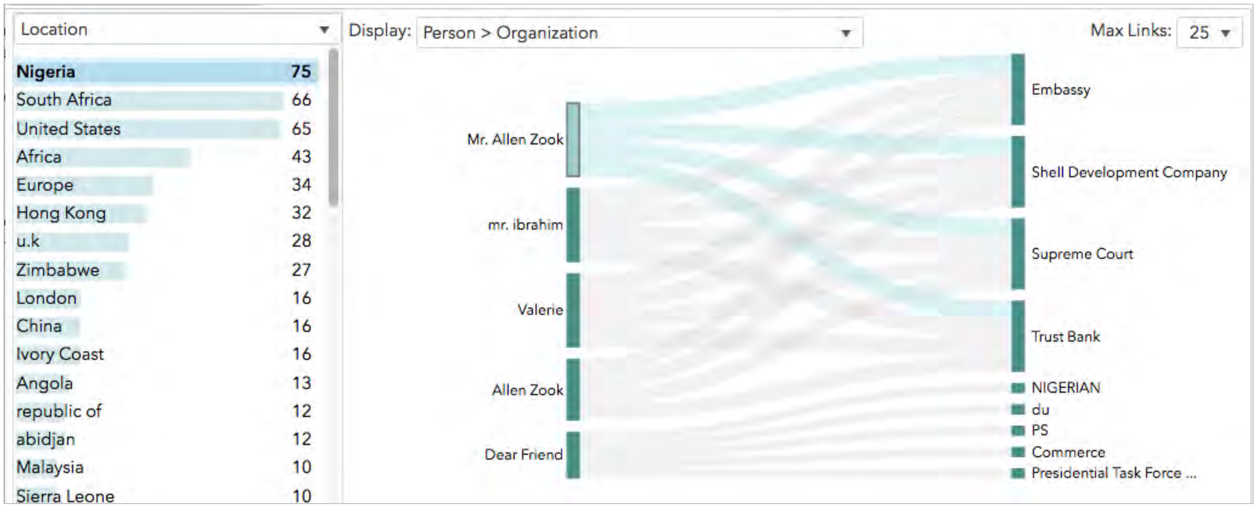
To
samaluko1@37.com

Subject
transparent enterprises

Body
Dear Sir, I am Sam Aluko, former special adviser on Petroleum matters to the late Head of State of Government of Nigeria Gen. Sani Abacha. Because of my strategic position in the former Government and also being a close confidant of the late Head of State, I was able to acquire personally the sum of 50,000,000.00 USD fifty million United States Dollars only presently lodged in a Security Deposit in Europe. I made this money largely through CONSULTANCY FEE and Good Faith Fees paid by foreign Oil companies prior to Allocation of Deep Water Oil Blocs and other

Figure 11: Display Options for Main View of the Text Analytics Workspace

In Figure 13, we can quickly see that emails that contain the location "Nigeria" also contain a number of top persons and organizations. In other words, we can see that anytime "Nigeria" is mentioned, a "Mr. Allen Zook" is also mentioned. If you select "Mr. Allen Zook", you can see that four different organizations are mentioned as well. This functionality enables the investigator to quickly discover new relationships that might have gone undetected otherwise.



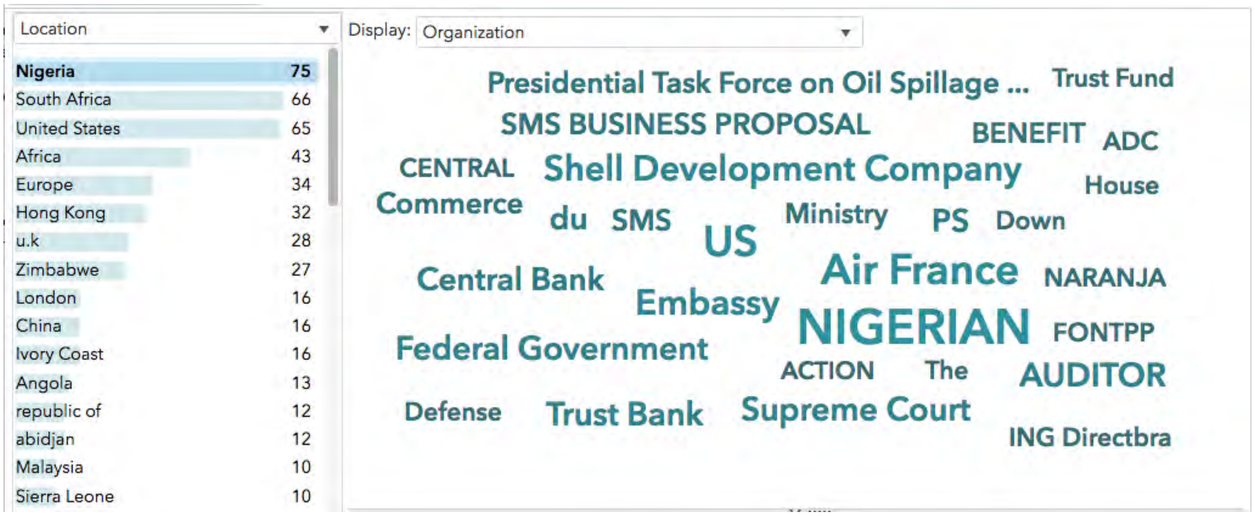
Location	
Nigeria	**75**
South Africa	66
United States	65
Africa	43
Europe	34
Hong Kong	32
u.k	28
Zimbabwe	27
London	16
China	16
Ivory Coast	16
Angola	13
republic of	12
abidjan	12
Malaysia	10
Sierra Leone	10

Display: Person > Organization

Max Links: 25

Figure 12: Details of Entity Discovery Relationships

If the investigator is only interested in *all* of the organizations that are discovered in the emails that contain the location "Nigeria", then a word cloud is the more appropriate display. Figure 14 shows the word cloud display. The size of the word represents the number of times it has occurred in the documents.



Location

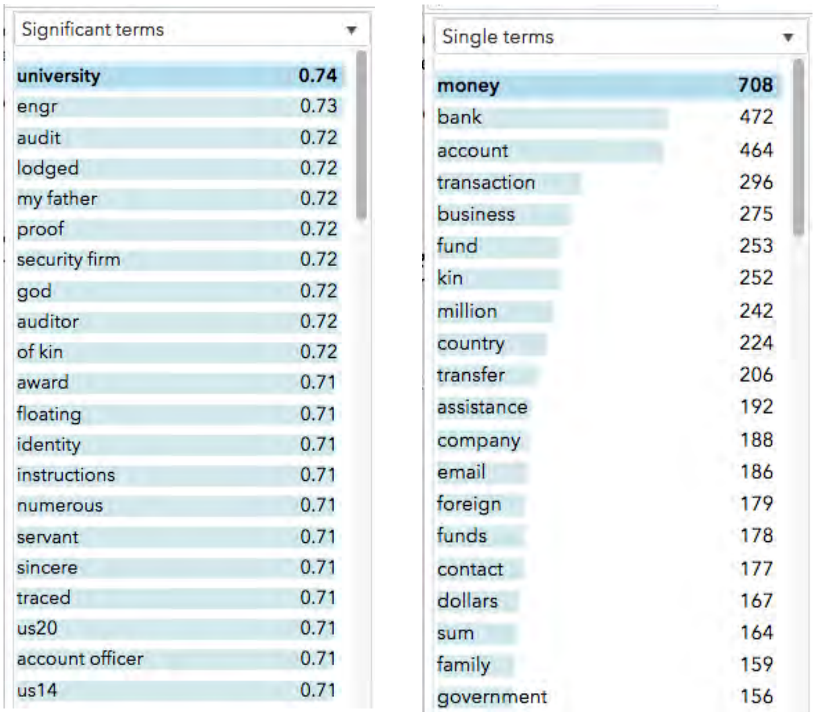
Location	
Nigeria	75
South Africa	66
United States	65
Africa	43
Europe	34
Hong Kong	32
u.k	28
Zimbabwe	27
London	16
China	16
Ivory Coast	16
Angola	13
republic of	12
abidjan	12
Malaysia	10
Sierra Leone	10

Display: Organization

Figure 13: A Word Cloud of Discovered Organizations Found in Emails Containing "Nigeria"

TEXT MINING

There are two display options for the bar charts that reveal the results of text mining analysis performed on the documents. The first is the **Significant terms** option. These terms or phrases are results of the text mining analysis and have shown statistical relevance to the documents as a whole. This significance doesn't always correlate to the term frequency, so a weighting or importance is represented as a bar value instead of a frequency count. Figure 15 shows a comparison of significant terms versus single term frequency.



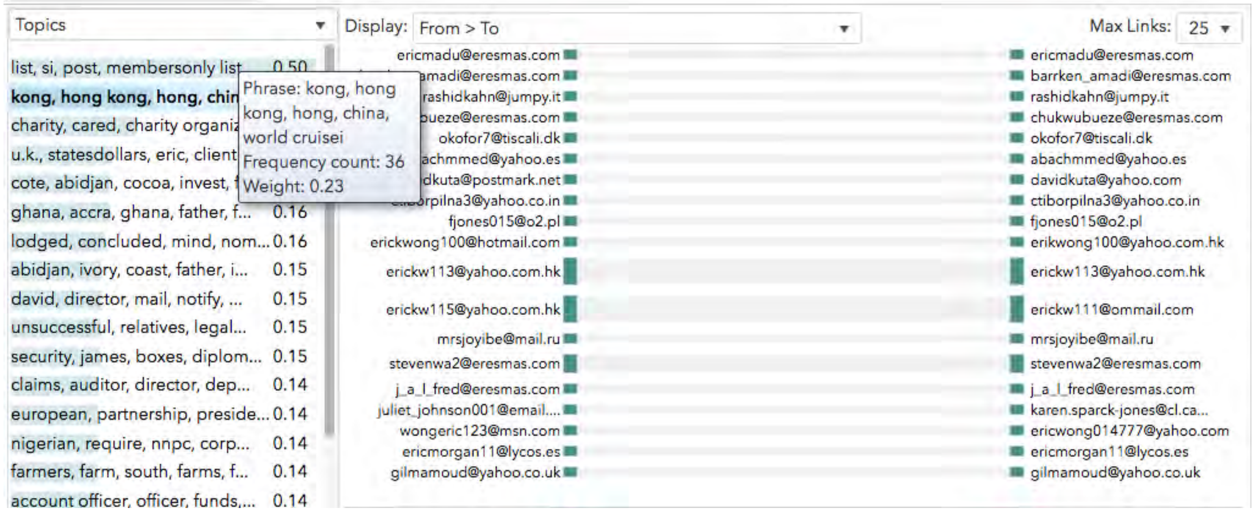
Significant terms	
university	0.74
engr	0.73
audit	0.72
lodged	0.72
my father	0.72
proof	0.72
security firm	0.72
god	0.72
auditor	0.72
of kin	0.72
award	0.71
floating	0.71
identity	0.71
instructions	0.71
numerous	0.71
servant	0.71
sincere	0.71
traced	0.71
us20	0.71
account officer	0.71
us14	0.71

Single terms	
money	708
bank	472
account	464
transaction	296
business	275
fund	253
kin	252
million	242
country	224
transfer	206
assistance	192
company	188
email	186
foreign	179
funds	178
contact	177
dollars	167
sum	164
family	159
government	156

Figure 14: Statistically Significant Terms Versus High-Frequency Terms

In Figure 15, notice that the word “money” doesn’t appear in the list of significant terms. This is probably because it occurs so frequently that it does not make a statistical impact. Frequent terms such as “money” become “noise” in the documents, and thus aren’t significant.

The second text mining analysis we provide is topic or theme clustering. Topics represent a cluster of documents that contain a similar theme. For example, Figure 16 shows a topic that is characterized by the following terms or phrases: “kong, hong kong, hong, china, world cruisei”. Therefore, an investigator can determine that the 36 documents that have been classified under this topic have a similar theme that discusses China or Hong Kong. This provides the investigator a quick way to group the documents together and understand in general what the topic of discussion might be.



Topics ▾

list, si, post, membersonly list	0.50
kong, hong kong, hong, chin	
charity, cared, charity organiz	
u.k., statesdollars, eric, client	
cote, abidjan, cocoa, invest, f	
ghana, accra, ghana, father, f...	0.16
lodged, concluded, mind, nom...	0.16
abidjan, ivory, coast, father, i...	0.15
david, director, mail, notify, ...	0.15
unsuccessful, relatives, legal...	0.15
security, james, boxes, diplom...	0.15
claims, auditor, director, dep...	0.14
european, partnership, preside...	0.14
nigerian, require, nnpc, corp...	0.14
farmers, farm, south, farms, f...	0.14
account officer, officer, funds,...	0.14

Phrase: kong, hong kong, hong, china, world cruisei
Frequency count: 36
Weight: 0.23

Display: From > To ▾

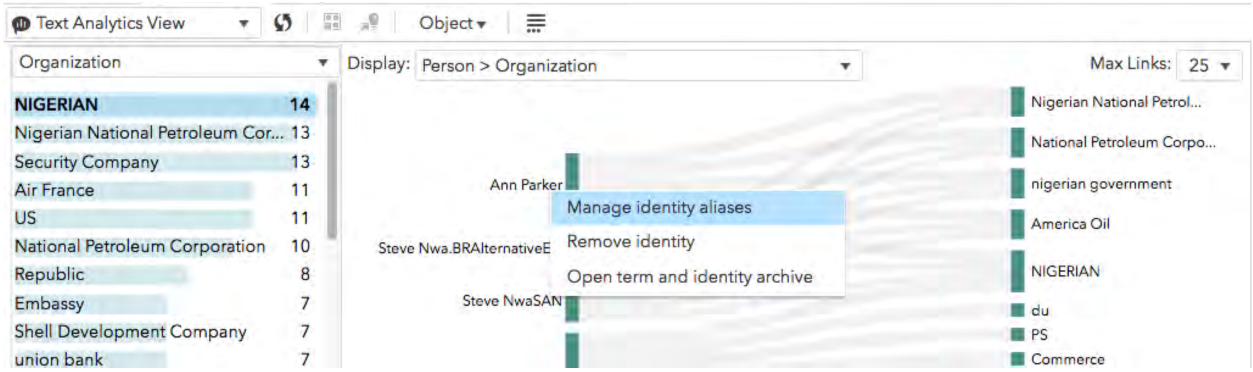
Max Links: 25 ▾

Figure 15: Topics View

The text mining results of the **Test Analytics View** are only a starting point for text mining analytics. The SAS® Visual Text Analytics solution can provide a much richer set of results for text mining.

IDENTITY MANAGEMENT

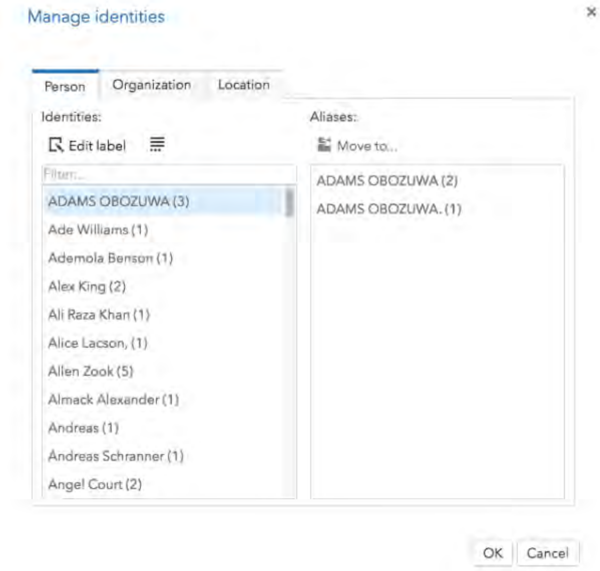
When discovering potential named entities or identities, you might encounter several issues that require some adjustments. When the text analysis is initially performed, we attempt to group like entities together and create aliases. However, the basic fuzzy matching is not always 100% correct, and some are missed. The text analytics workspace enables investigators to manage the identities found during the analysis. An investigator can access all discovered identities via a pop-up menu available on any discovered identity. Figure 17 shows the pop-up menu.



NIGERIAN	14
Nigerian National Petroleum Cor...	13
Security Company	13
Air France	11
US	11
National Petroleum Corporation	10
Republic	8
Embassy	7
Shell Development Company	7
union bank	7

Figure 16: Pop-Up Menu for Identities

Figure 18 shows the window that enables an investigator to create aliases and to rename the identity. Notice that the highlighted identity contains two aliases. They look very similar, but the second one contains a period (.) at the end. The number in parentheses represents the number of occurrences of the alias. By default, all identities have a label that corresponds to the name of the first alias in the list. This can be changed by the investigator.



Identities:

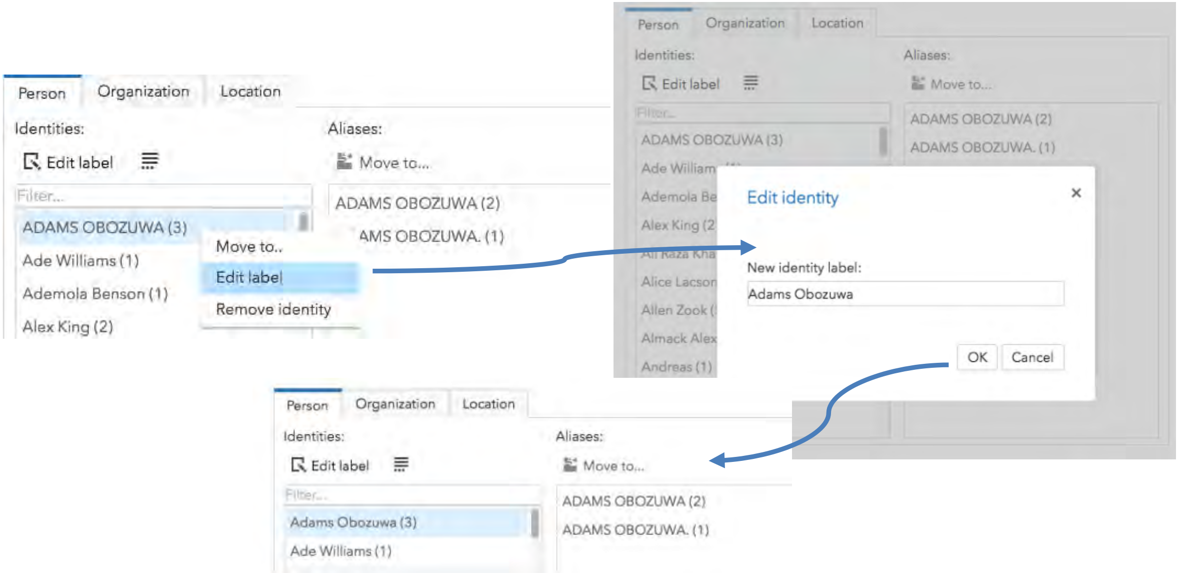
ADAMS OBOZUWA (3)
Ade Williams (1)
Ademola Benson (1)
Alex King (2)
Ali Raza Khan (1)
Alice Lacson, (1)
Allen Zook (5)
Almack Alexander (1)
Andreas (1)
Andreas Schranner (1)
Angel Court (2)

Aliases:

ADAMS OBOZUWA (2)
ADAMS OBOZUWA. (1)

Figure 17: Identity Management Window

Identities and aliases can be moved and deleted from this list as well. Figure 19 illustrates the steps of changing an identity label. The changed label becomes the label that is used in all of the views.



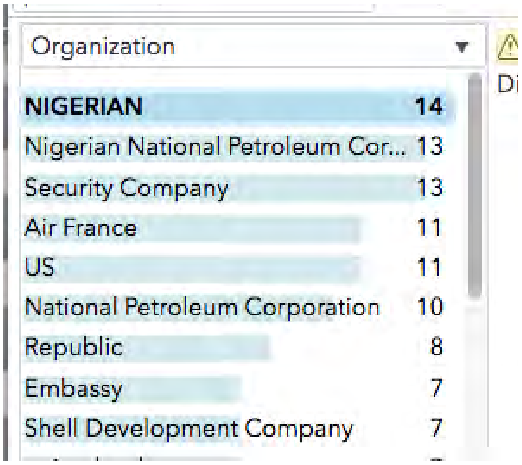
New identity label: Adams Obozuwa

Figure 18: Changing the Identity Label

Some examples of potential issues with named identities are the following:

- Identity overlap:
  An identity is defined in two different categories. In our sample data, the phrase “Angel Court” appears as a person and an organization.
- Invalid identity:
  The identity is not a valid person, organization, or location. These items need to be removed from the analysis.
- Missed aliases:
  In our example, we have three unique identities: “Nigerian National Petroleum CorporationNNPC.”, “National Petroleum Corporation”, and “Nigerian National Petroleum Corporation”. These three identities might need to be aliases of an identity with a label of “NNPC”.

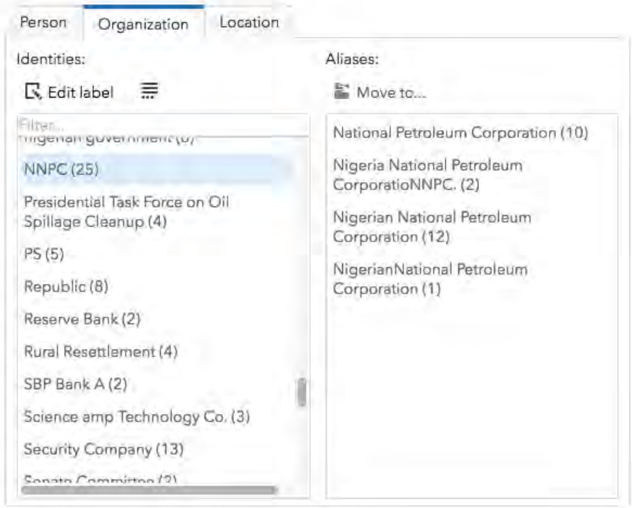
Figure 20 shows the organizations before they are managed. Notice that both identities show up separately: “Nigerian National Petroleum CorporationNNPC.” and “National Petroleum Corporation”.



NIGERIAN	14
Nigerian National Petroleum Cor...	13
Security Company	13
Air France	11
US	11
National Petroleum Corporation	10
Republic	8
Embassy	7
Shell Development Company	7

Figure 19: Identities without User-Defined Aliasing

In Figure 21, we have a new identity label “NNPC” for all the aliases listed.



Identities:

- nigerian government (6)
- NNPC (25)
- Presidential Task Force on Oil Spillage Cleanup (4)
- PS (5)
- Republic (8)
- Reserve Bank (2)
- Rural Resettlement (4)
- SBP Bank A (2)
- Science amp Technology Co. (3)
- Security Company (13)
- Senate Committee (2)

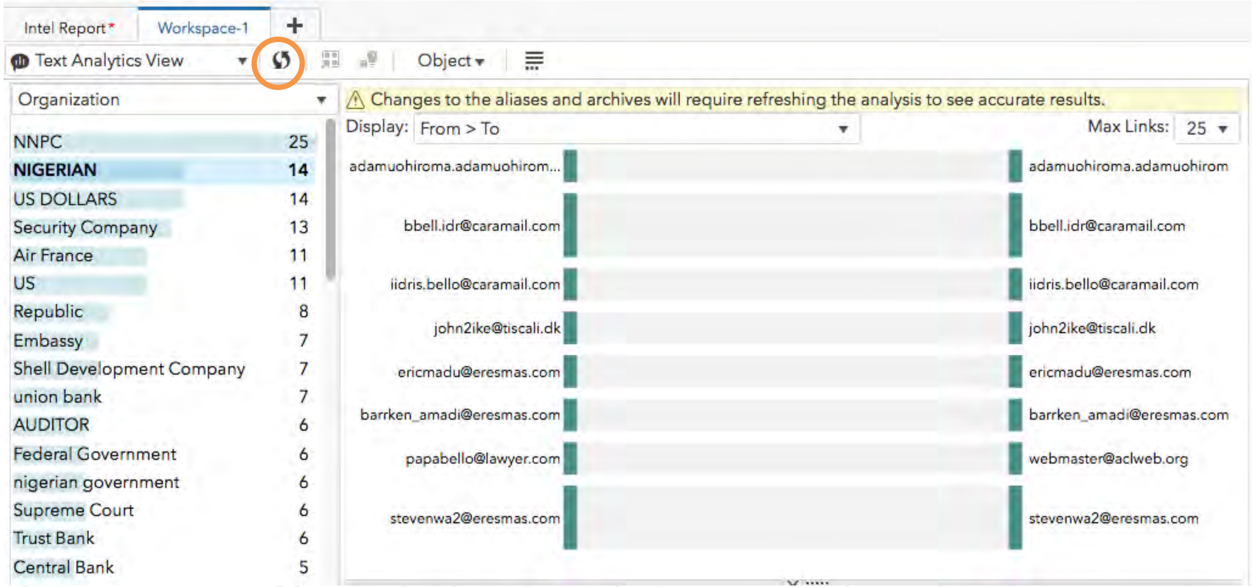
Aliases:

- National Petroleum Corporation (10)
- Nigeria National Petroleum CorporatioNNPC. (2)
- Nigerian National Petroleum Corporation (12)
- NigerianNational Petroleum Corporation (1)

Figure 20: Identity Aliasing

Figure 22 shows the new identity label. All aliases defined under this identity are grouped together when they appear in the analysis.

Note: A highlighted note also appears at the top of the display. This note informs you that there are changes to the identities and you need to refresh the analysis to fully incorporate your changes into the analysis. In Figure 22, the Refresh button at the top is circled.



⚠ Changes to the aliases and archives will require refreshing the analysis to see accurate results.

NNPC	25
NIGERIAN	14
US DOLLARS	14
Security Company	13
Air France	11
US	11
Republic	8
Embassy	7
Shell Development Company	7
union bank	7
AUDITOR	6
Federal Government	6
nigerian government	6
Supreme Court	6
Trust Bank	6
Central Bank	5

adamuohiroma.adamuohirom...	adamuohiroma.adamuohirom
bbell.idr@caramail.com	bbell.idr@caramail.com
iidris.bello@caramail.com	iidris.bello@caramail.com
john2ike@tiscali.dk	john2ike@tiscali.dk
ericmadu@eresmas.com	ericmadu@eresmas.com
barrken_amadi@eresmas.com	barrken_amadi@eresmas.com
papabello@lawyer.com	webmaster@aclweb.org
stevenwa2@eresmas.com	stevenwa2@eresmas.com

Figure 21: Results with New Identity Label and Aliases Grouped Together

One of the big advantages of the identity management feature is that all the changes that are made are saved with the workspace. Therefore, the next time the investigator opens this workspace, all the term or phrase removal and identity management will be done as part of the initial analysis. This enables the investigator to start with the investigative process right away instead of having to repeat the cleanup process.

CONCLUSION

The text analytics workspace can provide the first step in understanding unstructured data that might be related to your investigation. In this workspace, relationships are revealed, clusters of documents are defined by a topic or theme, and common phrases can be exposed to assist in identifying fraudulent behavior. The workspace also provides an identity management tool to manage aliases and enable you to consolidate multiple terms or names into one key category. Your investigative process can be enhanced by examining the hidden gems that are sometimes buried deep in the voluminous text of unstructured data.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Danielle Davis
SAS Institute
919-531-9702
danielle.davis@sas.com
http://www.sas.com