i2 TextChart

Welcome to the i2 TextChart documentation, where you can find information about how to use and administer i2 TextChart.

i2 TextChart is designed to let you seamlessly process and analyze unstructured documents in over 200 languages in Analyst's Notebook. Over three dozen important entity types, including people, organizations, and locations, as well as their relationships or links can be extracted.

After the automated process, you can modify extracted entities, tag a new entity or link, delete a system-tagged entity, or accept the entities that have been identified by TextChart. The combination of the automated text analysis and human input allows you to efficiently analyze the growing collections of unstructured documents and collaborate across your organization.

You can leverage all of the powerful visualization and analytical features of Analyst's Notebook to build charts from entities and expand the network of entities and links that were identified in documents that have been previously processed. In addition to helping you understand how entities are related, the application lets you view all documents that contain entities and links of interest.

Configuration

TextChart is shipped as an out-of-the-box offering. It includes standard term extraction rules and, if you use the embedded term storage option, requires very little configuration.

If needed, however, you can modify the term extraction rules and then merge your custom rules with later standard rule updates as they are released. You can also add your own database that allows you to store extracted terms for later use.

Upgrading to TextChart Premium increases the scale of your deployment significantly:

- Documents from different sources can be processed using different extraction rules.
- Terms can be loaded in bulk from multiple sources.
- Extracted terms are stored in a database that analysts can query as an external source in Analyst's Notebook Premium.

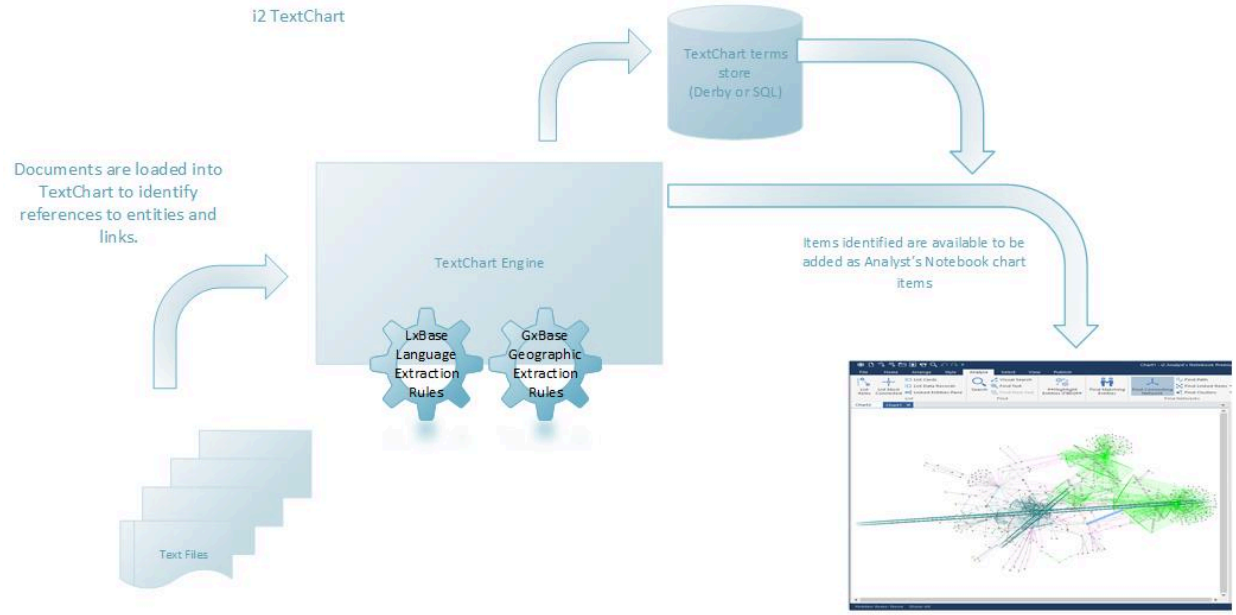
Key components

There are a number of key moving parts in every TextChart deployment:

- File locations: the file system paths that contain documents to be processed
- Extraction engine: where documents are processed to extract the terms that they contain
- LxBase: the lexicon definition, which is the list of rules that determine how terms are identified in the source material
- GxBase: the geographic definition, which is the list of rules that determine how geographic locations are identified in the source material

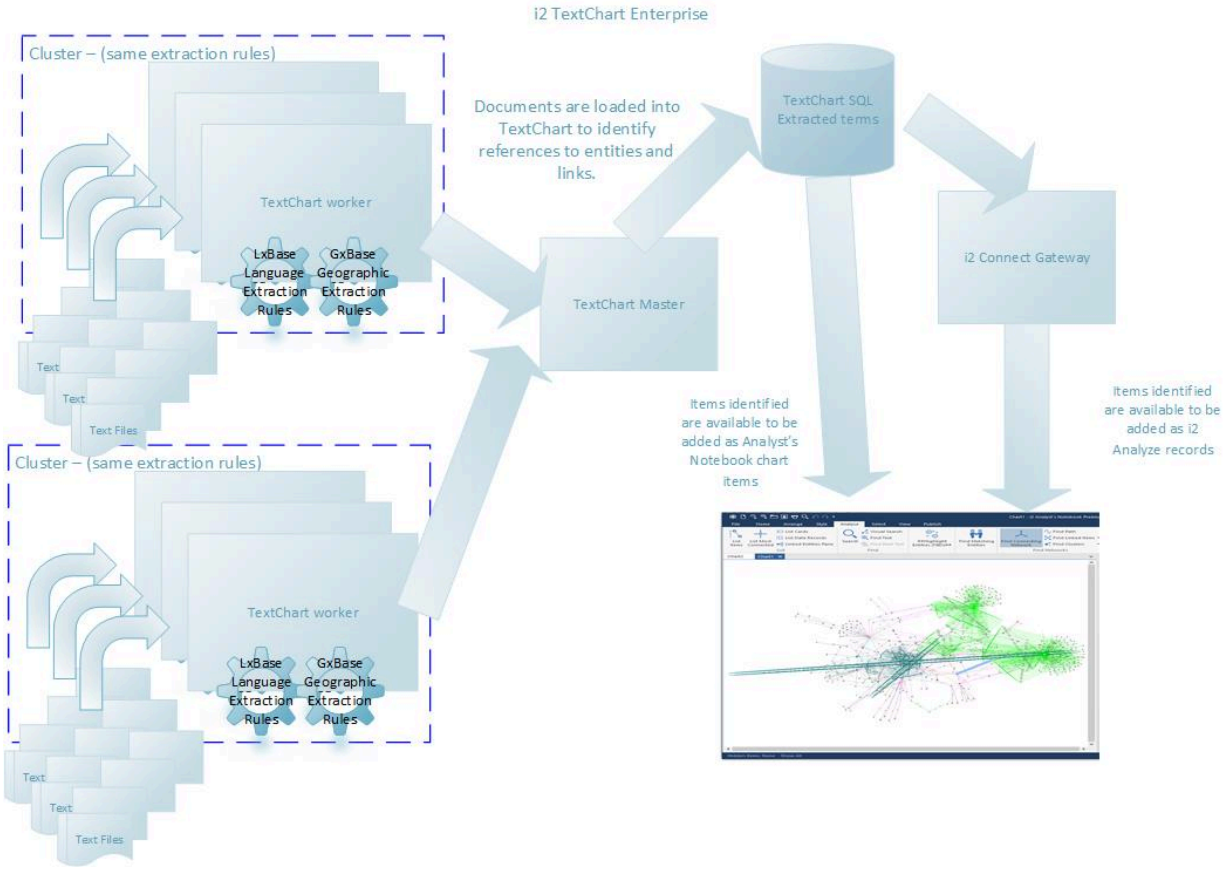
In addition, in a standard TextChart deployment you might also wish to set up a repository for the extracted terms, and external mapping applications to work with geographic data.

© N. Harris Computer Corporation



A TextChart Premium deployment can be scaled to match your needs, allowing different file types to be extracted using different rules, automated updates as files change or new documents become available, and integration with the i2 Connect gateway to create i2 Analyze records:

© N. Harris Computer Corporation



i2 TextChart Standard

This is the placeholder topic that introduces i2 TextChart Standard, and the components that are available with this product.

Installing i2 TextChart

This topic describes how to install and activate i2 TextChart.

Before you begin

Important: If you are upgrading from an earlier version of i2 TextChart, follow the instructions in Upgrading i2 TextChart before you complete the procedure below.

Prerequisite software

The release notes contain a full list of the required software for i2 TextChart. You *must* install i2 Analyst's Notebook before you install i2 TextChart.

Optional configuration

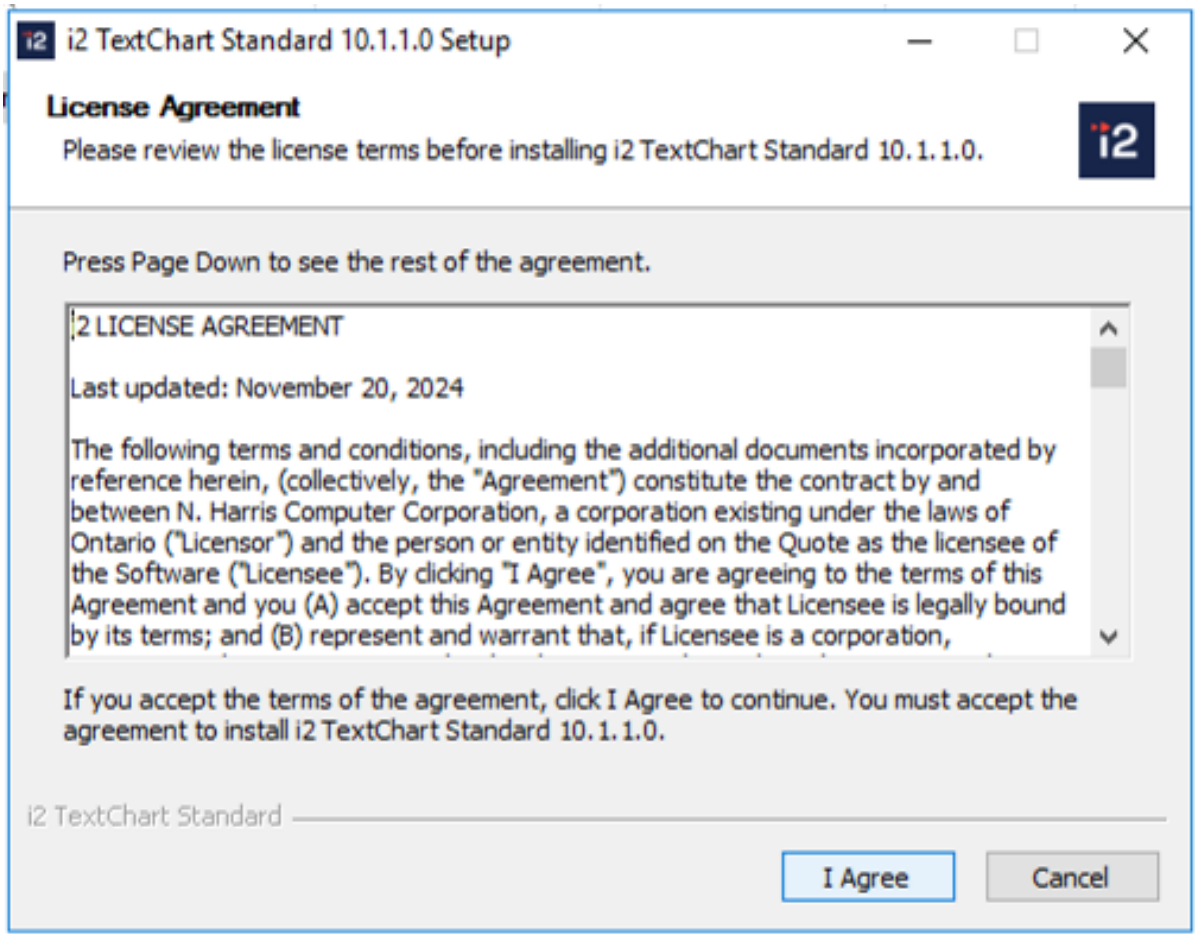
© N. Harris Computer Corporation

i2 TextChart is supplied with an embedded database for local workspace storage. Optionally, you can replace the embedded database with Microsoft SQL Server 2014 or later.

Procedure

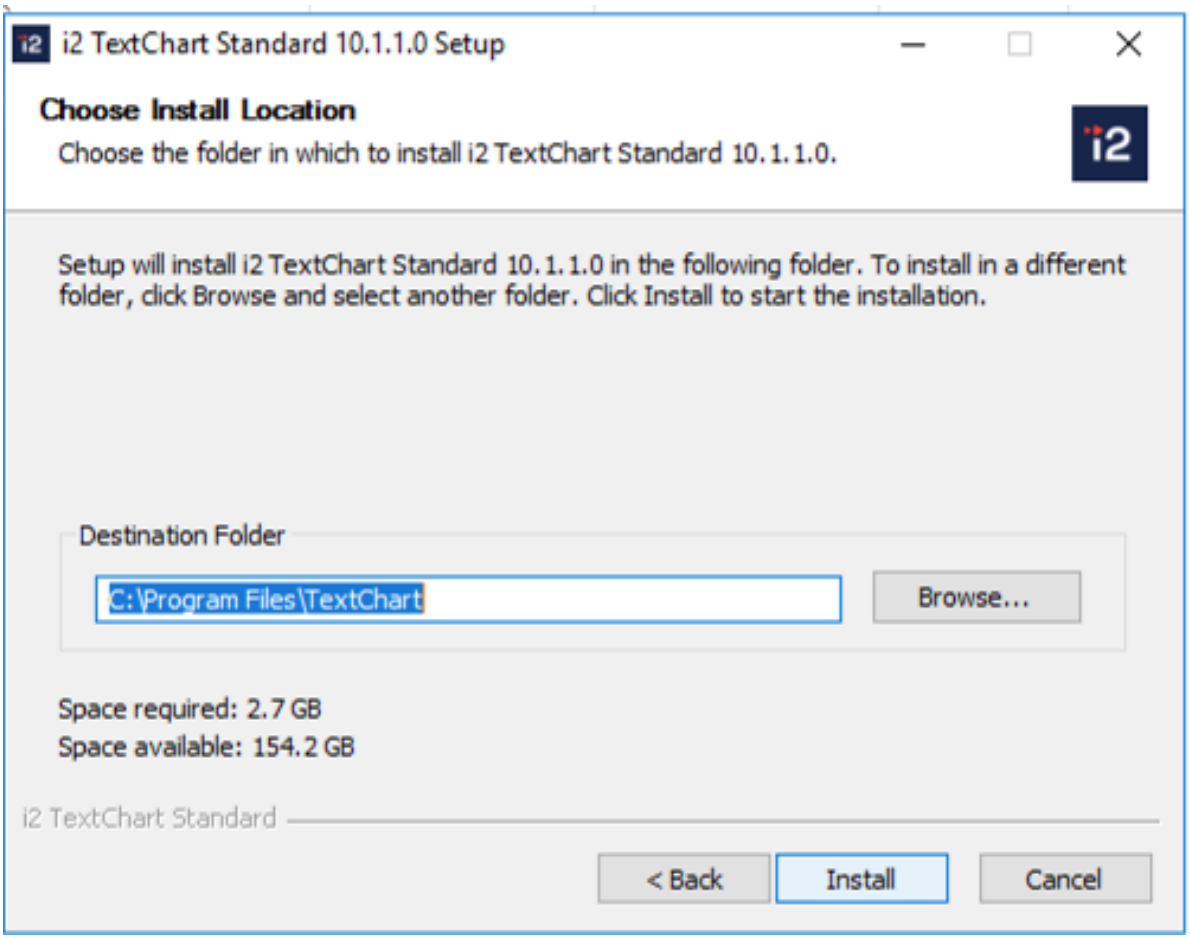
The following steps describe how to install i2 TextChart.

1. Extract the product files from your downloaded distribution.
2. Using Windows Explorer, browse to the root of the distribution and run the i2 TextChart Installer. Its name will contain the word "Standard" or "Premium", depending on which edition of i2 TextChart you're installing.
3. Follow the prompts. You must accept the license agreement to proceed. Click **I Agree**:



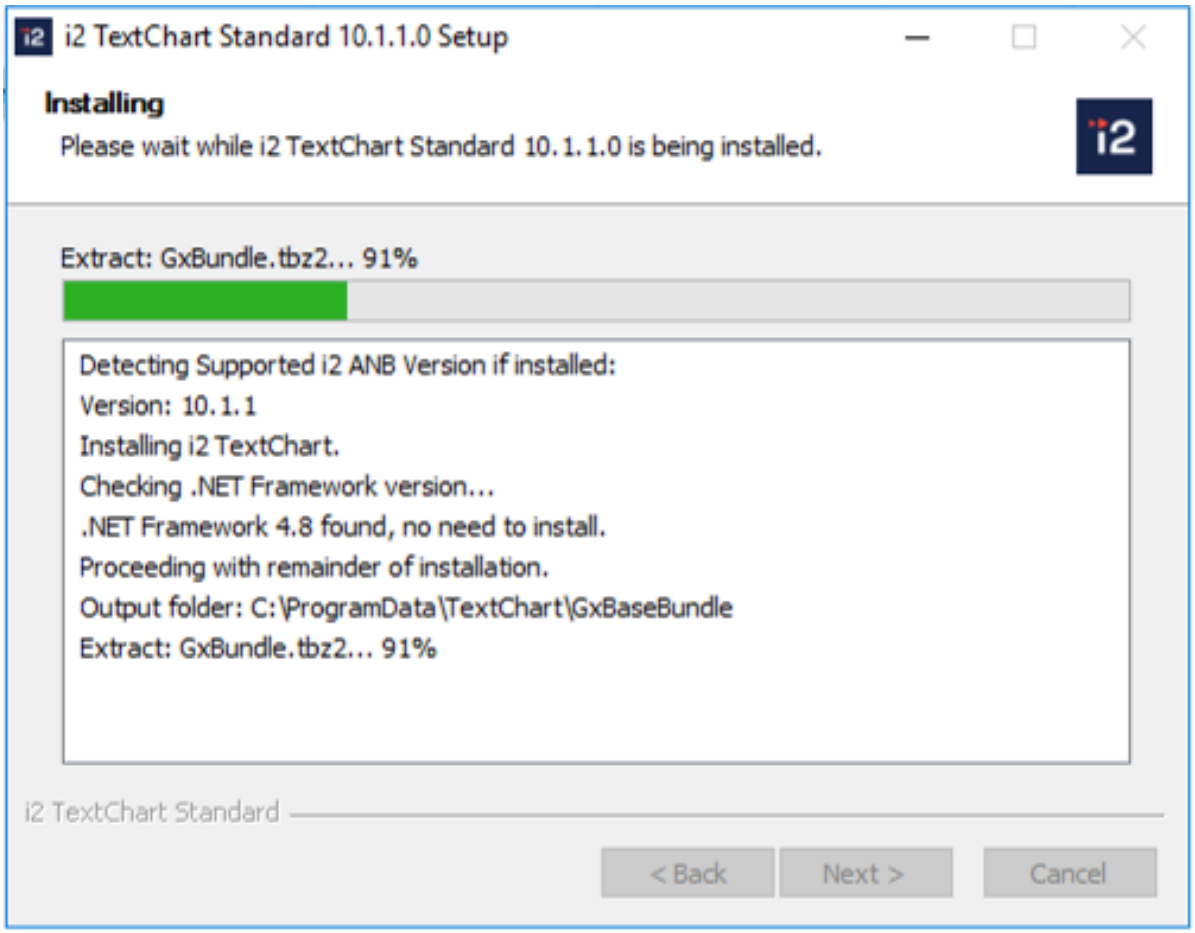
4. Choose the installation location and click **Install**:

© N. Harris Computer Corporation



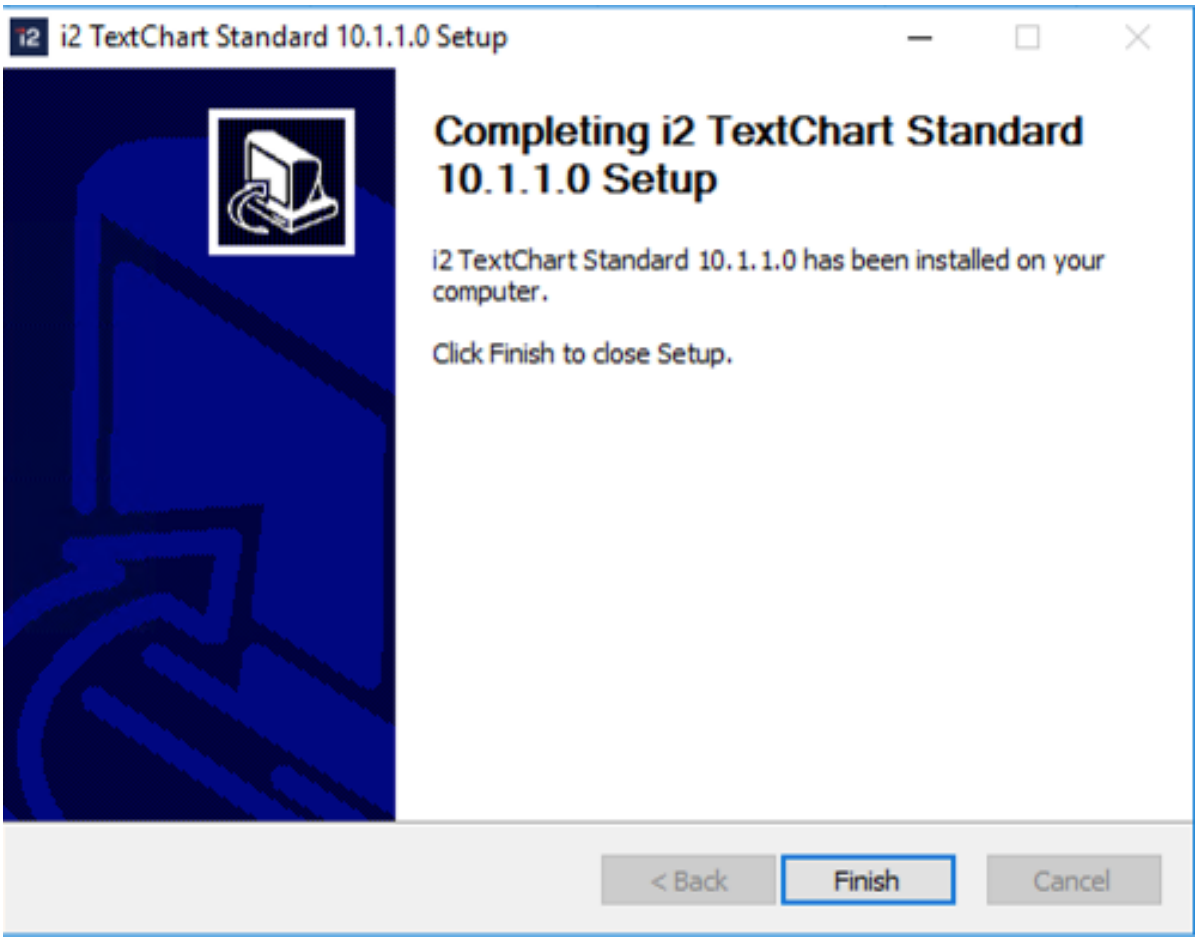
5. When the installer is complete, click **Next**.

© N. Harris Computer Corporation



6. When setup is complete, click **Finish**.

© N. Harris Computer Corporation



Installer Command-Line Parameters

The installer supports optional command-line flags that control how the application is installed. By default (no flags), the installer uses **command start/stop mode** with a **non-roaming (shared) profile**.

`/ROAMING`

Installs the application in roaming profile mode. Configuration and data are stored in the user's roaming `%APPDATA%` directory rather than the shared `C:\ProgramData\` location.

Use this when:

- Users have roaming profiles (e.g., VDI environments)
- The application data needs to follow the user across machines

`/SERVICE`

Installs the application as a Windows service (using Apache Commons Daemon). The service runs at the system level.

Use this when:

- The application should start automatically with Windows

© N. Harris Computer Corporation

- A system-level service is preferred over user-initiated start/stop

Incompatible combination

`/SERVICE` and `/ROAMING` **cannot be used together**. The Windows service runs at the system level and cannot access a user's roaming profile directory. If both flags are passed, the installer will:

- **Interactive mode**: Display an error dialog and abort.
- **Silent mode**: Write an error to `install_error.log` in the installer directory, set exit code 1, and abort.

Installing i2 TextChart Silently

The following steps describe how to install i2 TextChart silently:

1. Extract the product files from your downloaded distribution.
2. Using Windows Command Line Terminal, navigate to the root of the distribution and run the i2 TextChart Installer by issuing following command. Its name will contain the word "Standard" or "Premium", depending on which edition of i2 TextChart you're installing. Note that Analyst's Notebook is closed while installing the TextChart Text Analytics.

   ```
   "i2 TextChart Standard-X.X.X.X-Installer.exe" /S
   ```

   Note: If you need to uninstall i2 TextChart Text Analytics silently, you can do so by issuing following command. Note that Analyst's Notebook is closed while uninstalling the TextChart Text Analytics.

   ```
   Uninstall.exe /S
   ```

   Uninstall.exe should be available in installation directory, `C:\Program Files\TextChart` for example.

Installation Examples

```
:: Default install (command mode, non-roaming)
"i2 TextChart Standard-X.X.X.X-Installer.exe"

:: Roaming profile install (command mode)
"i2 TextChart Standard-X.X.X.X-Installer.exe" /ROAMING

:: Service install (non-roaming)
"i2 TextChart Standard-X.X.X.X-Installer.exe" /SERVICE

:: Silent default install
"i2 TextChart Standard-X.X.X.X-Installer.exe" /S

:: Silent roaming install
"i2 TextChart Standard-X.X.X.X-Installer.exe" /S /ROAMING

:: Silent service install
"i2 TextChart Standard-X.X.X.X-Installer.exe" /S /SERVICE
```

Defaults Summary

Setting	Default (no flag)	With flag
Profile mode	Non-roaming (`C:\ProgramData\...`)	`/ROAMING` = User roaming `%APPDATA%\...`

© N. Harris Computer Corporation

Setting	Default (no flag)	With flag
Run mode	Command start/stop	`/SERVICE` = Windows service

What to do next

When i2 TextChart is successfully installed, activated, and configured, the plug-in launches automatically when you open i2 Analyst's Notebook.

Note: If you need to stop i2 TextChart Text Analytics, you can do so through the Service Manager, which is available from the **i2 TextChart** menu in the Analyst's Notebook ribbon. To restart i2 TextChart Text Analytics, you must then restart Analyst's Notebook.

Configuring i2 TextChart

i2 TextChart is ready to use as soon as you install and activate it. However, to maintain and enhance its capabilities, TextChart also provides a range of options for configuring and updating it.

Changing the supported file extensions

The TextChart plug-in for Analyst's Notebook supports processing documents in a wide range of *formats* including plain text, PDF, and those that productivity applications generate. It uses file *extensions* to filter the documents that it presents for import, and you can change the configuration to display more or different files.

TextChart stores its list of file extensions in the file at `C:\ProgramData\TextChart\conf\filetypes.properties`. If an extension appears in this file, then documents with that extension appear in the Document View. By default, TextChart presents documents with the following extensions:

```
txt
rtf
pdf
doc
docx
pptx
```

The default `filetypes.properties` file also specifies that TextChart should present files that have no extension for import.

If your organization uses extensions that are not in the default list, you must add them to it. For example, you might have older Microsoft Powerpoint files with the `.ppt` extension, or plain text files that use a custom extension.

After you make changes to the list of supported file extensions, reinitialize the TextChart plug-in by closing and reopening the Analyst's Notebook application.

Changing the local database

By default, the standard edition of i2 TextChart uses its embedded database for storing the information that it generates during analysis. Optionally, you can configure TextChart standard edition to use a local Microsoft SQL Server database instead.

© N. Harris Computer Corporation

Before you begin

Choosing a SQL Server database in place of the embedded one enables you to inspect its contents directly. TextChart supports SQL Server 2014 and later, including the Express edition.

To configure TextChart to use a local SQL Server database, you use the Text Analytics Configuration application, which you also used to provide TextChart with a license key during installation.

To run the application after installation, you can start it from the Windows Start menu, or by launching the executable at `C:\Program Files\TextChart\RosokaConfiguration.exe`.

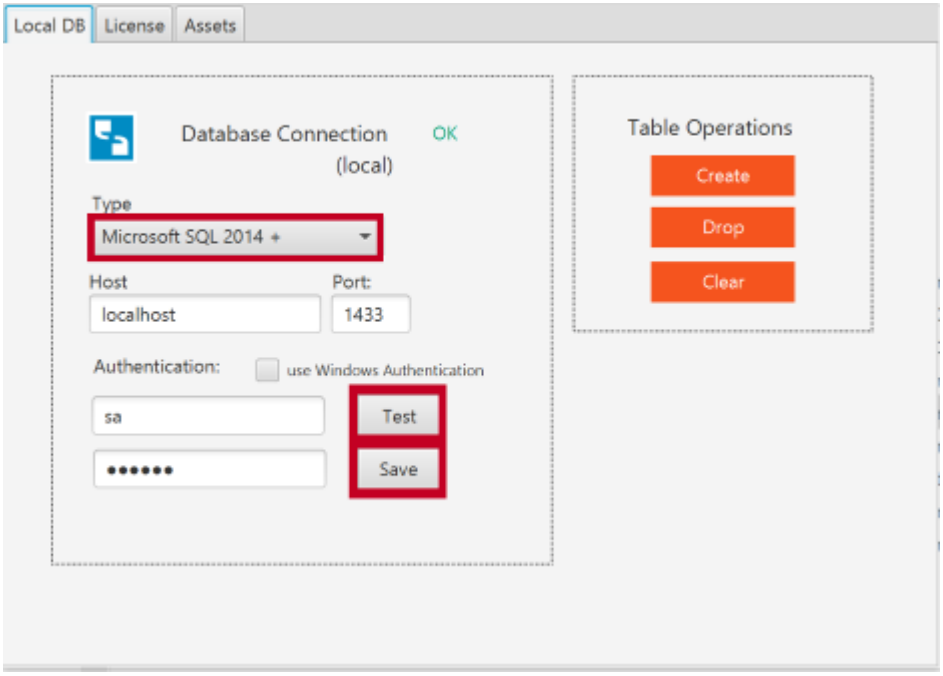
Procedure

Assuming that you have already set up a SQL Server database for TextChart to use, the procedure for connecting to it is as follows.

Note: If you want to create a database account for TextChart to use, or to change the port on which SQL Server accepts connections, see the instructions in Database configuration.

1. On the **Local DB** tab of the TextChart Text Analytics Configuration application, select your database from the **Type** drop-down, and ensure that the **Host** and **Port** details are correct.

   Note: For the standard edition of TextChart, the host must be the same workstation that TextChart and Analyst's Notebook are installed on.
2. In the **Authentication** fields, provide the user name and password of a database account with administrator privileges.



3. Click **Test** to verify that the information you provided is correct. If the test is successful, click **Save**.

© N. Harris Computer Corporation

Note: If you see the "Driver not Present" error message, you need to add the appropriate JAR file to the `C:\Program Files\TextChart\Drivers` directory, relaunch the configuration application, and try again.

Drivers for Microsoft databases are included, drivers for MySQL are available from the MySQL website, and drivers for Db2 are available as part of your Db2 installation.

What to do next

If you need to change the configuration of the local database after you've begun to use i2 TextChart, you must close Analyst's Notebook and ensure that the TextChart application is no longer running before repeating the above procedure.

Note: After you change the configuration to use a different database from the one you've been using, you might need to re-create your collections and reprocess the documents they contained.

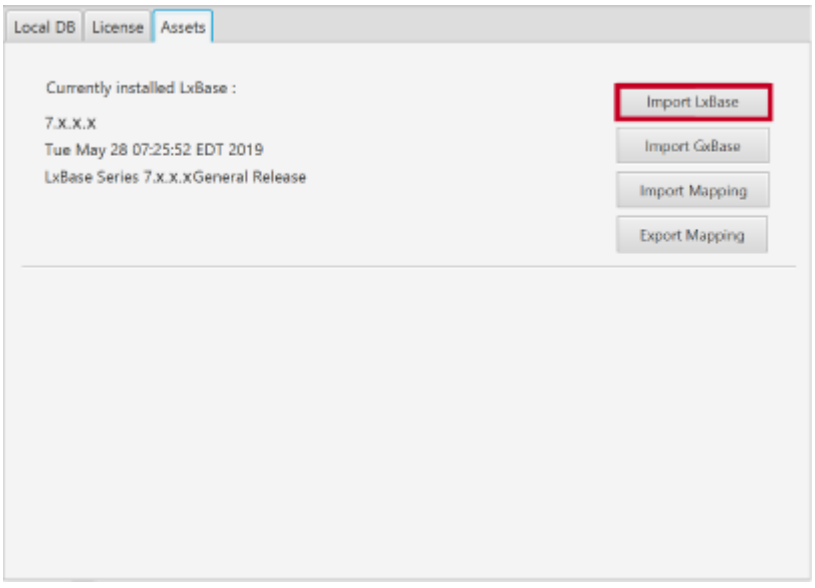
Updating an LxBase

In i2 TextChart, an LxBase is the collection of linguistic rules and dictionary entries that govern text extraction. From time to time, i2 publishes updates that you can apply to the LxBase in your installation.

Procedure

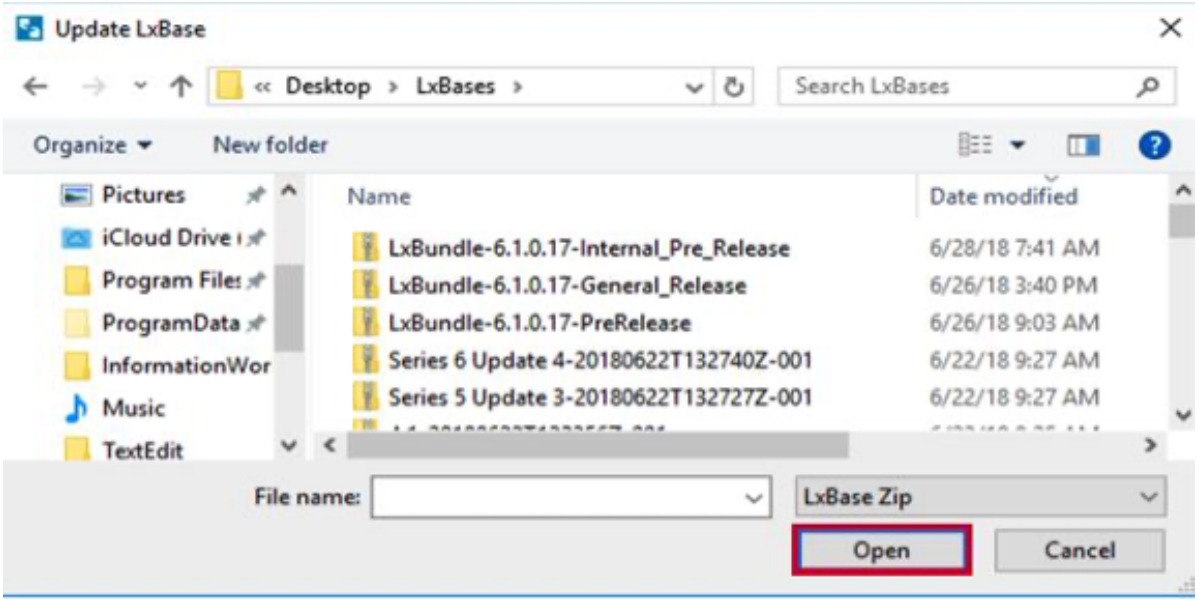
To update your LxBase, you use the TextChart Text Analytics Configuration application to provide i2 TextChart with the location of a ZIP file that contains the new information.

1. Open the configuration application and select the **Assets** tab:



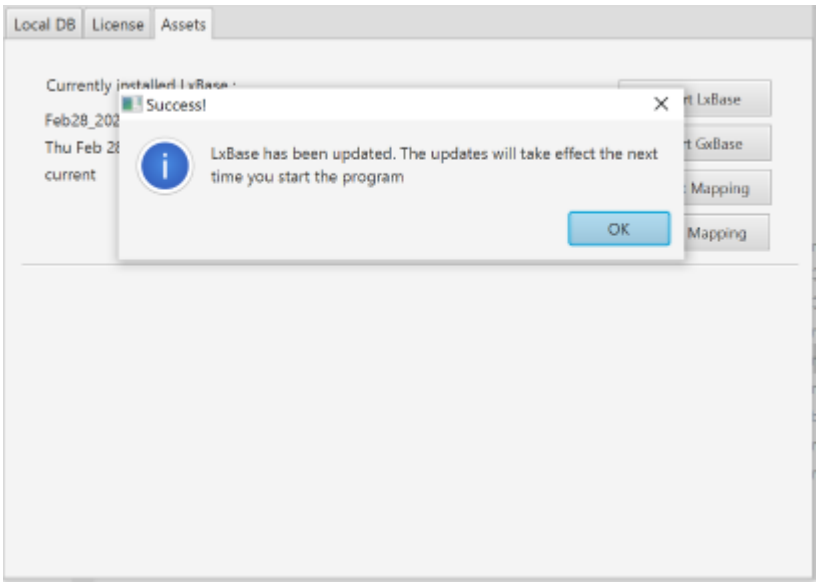
2. Click **Import LxBase**, navigate to the new ZIP file, and click **Open**.

© N. Harris Computer Corporation



i2 TextChart automatically updates your LxBase with information from the file.

3. When the following dialog appears, click **OK**:



The changes and updates will take effect when you restart Analyst's Notebook.

Note: i2 TextChart backs up the previous version of your LxBase in `C:\ProgramData\TextChart`, in a directory named `LxBase-backup-<TIMESTAMP>`. To restore a backup, delete the `LxBase` directory, and then rename the backup to `LxBase`.

© N. Harris Computer Corporation

Upgrading i2 TextChart Standard

Upgrading i2 TextChart Standard requires a complete uninstallation of the previous version before installing the new release. This procedure ensures a clean installation and prevents potential conflicts between versions.

About this task

The upgrade process consists of the following stages:

1. Back up your existing database configuration
2. Uninstall the current version of i2 TextChart Standard
3. Install the new version using the standard installation procedure

Prerequisites

Before beginning the upgrade:

- Ensure you have administrator privileges on the workstation
- Verify i2 TextChart Standard and i2 Analyst's Notebook are closed
- Obtain the installation package for the new version
- Close all related applications before starting the upgrade process

Procedure

Step 1: Prepare for the upgrade

1. Close i2 Analyst's Notebook.
2. If you are using a local TextChart database, create a backup:

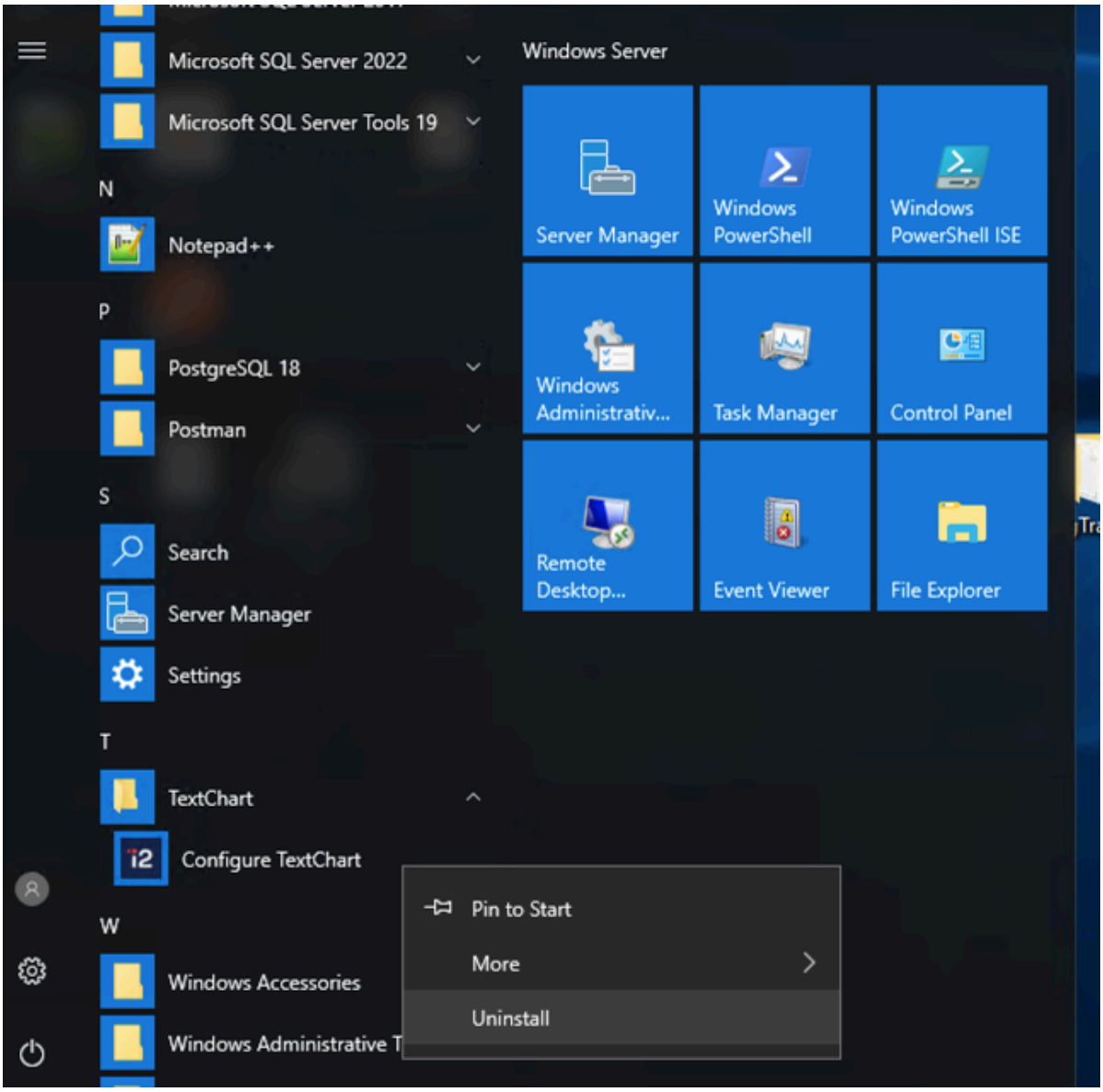
   a. **For SQL Server databases**: Use SQL Server Management Studio or other database administration tools to create a full backup of your TextChart database.

   b. **For embedded databases**: Navigate to the database directory at `C:\ProgramData\TextChart\databases\Rosoka` (or `C:\ProgramData\Rosoka\TextAnalytics\databases\Rosoka` for older versions) and create a copy of the entire folder structure.

Step 2: Uninstall the current version

1. Uninstall i2 TextChart Standard using one of the following methods:
   - From the Windows Start menu, locate and run the TextChart uninstaller
   - Navigate to `C:\Program Files\TextChart\` and execute `Uninstall.exe`
   - Use Windows **Add or Remove Programs** from the Control Panel

© N. Harris Computer Corporation



2. Follow the uninstaller prompts and wait for the process to complete before proceeding.

Step 3: Install the new version

After successful removal of the previous version, install the new version by following the procedures documented in Installing i2 TextChart Standard.

What's next

After completing the installation, restore your database content from the backup created in Step 1, if required. Test the installation by processing a sample document to verify that all components are functioning correctly.

© N. Harris Computer Corporation

Consider reviewing the release notes for the new version to understand any new features or configuration changes that may benefit your deployment.

Other configuration

i2 TextChart provides a few ways to modify its behavior so that it works better in your environment. For example, you can adjust the amount of memory available to the extraction process, and change the way TextChart authenticates with the database.

Memory configuration

If you experience poor performance, especially during processing of large files, you can change the size of the JVM heap by editing the configuration file at `C:\Program Files\TextChart \RosokaServiceCore.l4j.ini`.

To adjust memory allocation, which defaults to 4 GB, change:

```
-Xms4g
-Xmx4g
```

to:

```
-Xms<GB_OF_RAM_TO_ALLOCATE>g
-Xmx<GB_OF_RAM_TO_ALLOCATE>g
```

For example, to set it to 8 GB, the two lines would be:

```
-Xms8g
-Xmx8g
```

Database configuration (SQL Server)

TextChart authenticates with SQL Server through database user credentials rather than operating system credentials. You need to ensure that your installation of SQL Server is configured appropriately.

Setting mixed authentication

1. Open SQL Server Management Studio.
2. From the Authentication drop-down list, select **Windows Authentication** .
3. Click **Connect**.
4. Right-click the SQL instance and choose **Properties** to open the Server Properties window.
5. In the **Select a page** list, click **Security**.
6. In the **Server authentication** section, choose **SQL Server and Windows Authentication mode**.
7. Click **OK** to close the Server Properties window.
8. Right-click the SQL instance and choose **Restart**.
9. Click **Yes** to restart the SQL server.

Enabling the 'sa' user

1. Open SQL Server Management Studio.
2. From the Authentication drop-down list, select **Windows Authentication**.
3. Click **Connect**.
4. In the Object Explorer, expand **Security** and then expand **Logins**.
5. Right-click **sa**, and click **Properties**.

© N. Harris Computer Corporation

6. In the **Select a page** list, click **Status**, and then change the **Login** setting to **Enabled**.

   Note: You might need to set a password for this account as well.
7. Click **OK** to close the Login Properties window.
8. Press *F5* to refresh the information in the Object Explorer.

Changing TCP and port settings

1. From the **Start** menu, right-click **My Computer**, and then select **Manage**.
2. In the Computer Management window, expand **Services and Applications > SQL Server Configuration Manager > SQL Server Network Configuration**.
3. Select the instance for your installation. By default, this is **Protocols for SQLEXPRESS**.
4. Change the **Named Pipes** protocol option to **Enabled**. You do not need to change the default **Pipe Name** setting.
5. Change the TCP/IP protocol option to **Enabled**, and then open its **Properties** dialog.
6. In the **IP Addresses** tab, scroll down to **IPAll** at the bottom of the list.
7. Set **TCP Port** to 1433, and remove any value from **TCP Dynamic Ports**.
8. Click **Apply**. You will be prompted to restart SQL Server.
9. Restart the server for the changes to take effect.

Importing and processing documents

The i2 TextChart plug-in for Analyst's Notebook processes documents to extract information about the things they describe and the connections between them. You can add the extracted information to your charts in the form of entities and links.

Entity processing

Entities are the important, named items, such as the people, places, and events that a document describes. i2 TextChart uses the linguistic context of the document to determine what words or phrases are extracted as entities.

After processing, you can modify entity extraction results and apply your own knowledge before sending results to a chart.

Note: The LxBase documentation includes a list of all the entity types that TextChart understands, with their corresponding definitions.

Link processing

Links are relationships between two entities that i2 TextChart establishes through linguistic context. In extraction results, links appear in the form `<Entity>To<Entity>`. For example, `PersonToPerson` means that there is a link between one Person entity and another Person entity.

The predicate (that is, the reason why i2 TextChart identified the link) is stored with the link. If the predicate is "interviewed", then the link indicates that a Person interviewed another Person.

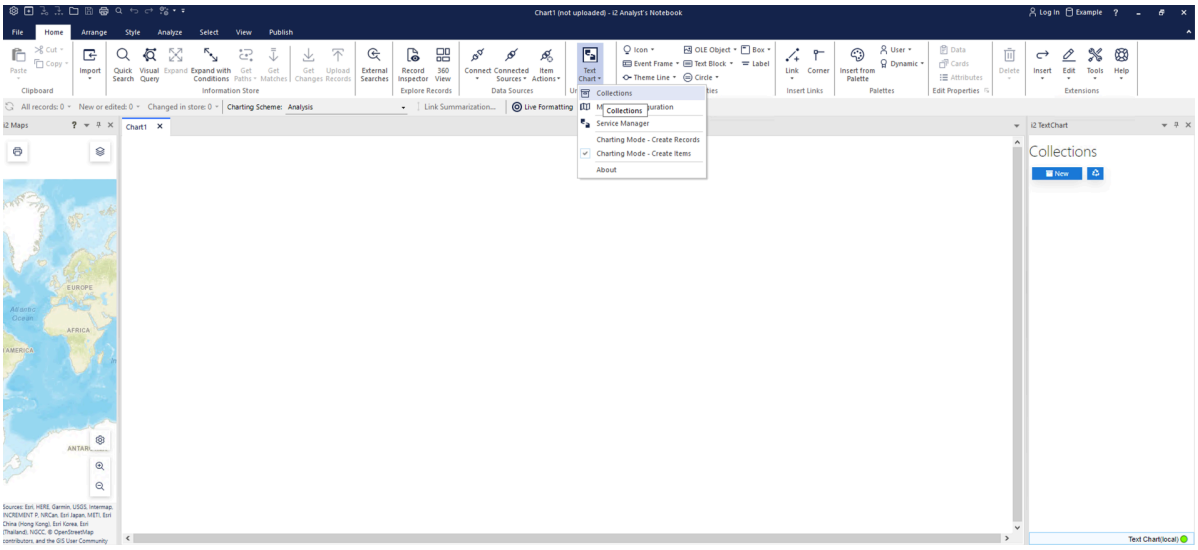
Note: The LxBase documentation includes a list of all the predicate types that TextChart understands.

Importing documents

To use i2 TextChart, the first step is to import the documents that you want to analyze.

© N. Harris Computer Corporation

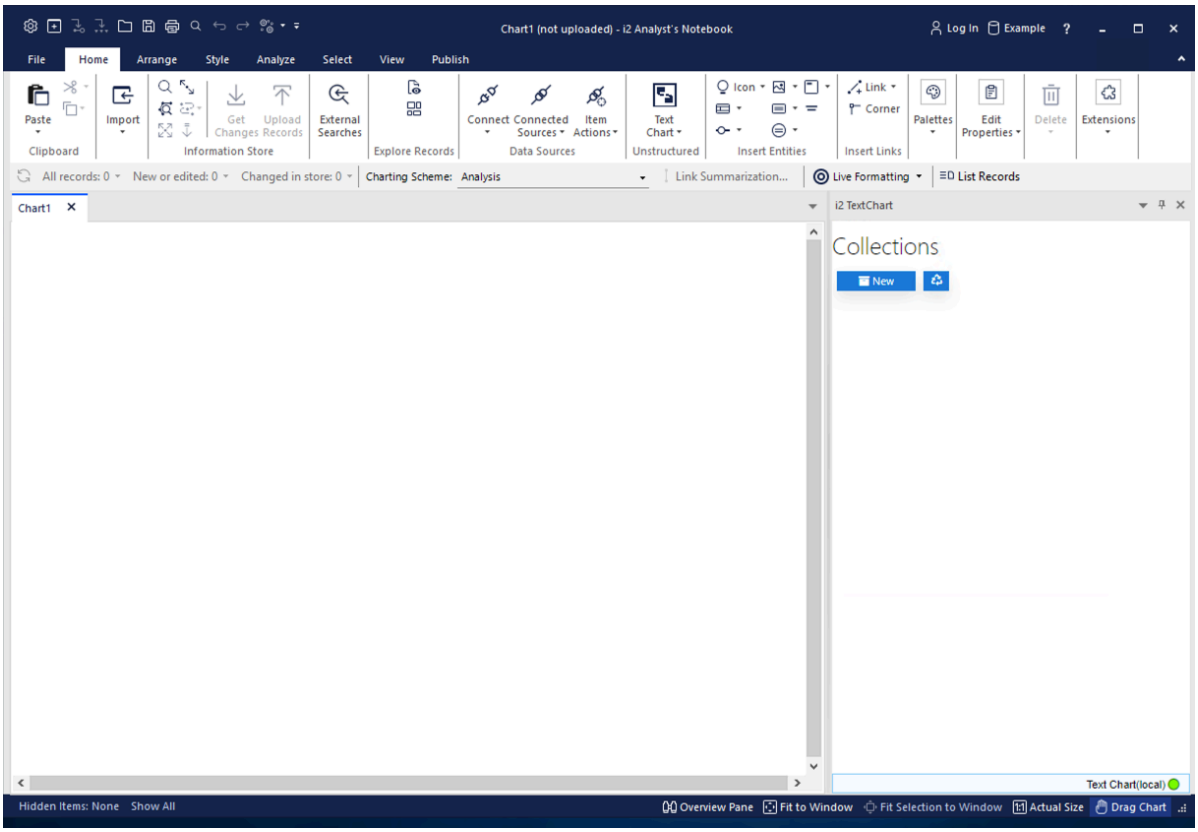
1. In Analyst's Notebook, select the **Text Analytics** menu in the **Home** tab, and then click **Collections**.



Note: The Mapping Configuration option is available only in the standard edition of TextChart. In TextChart Premium, mapping takes place on the server.

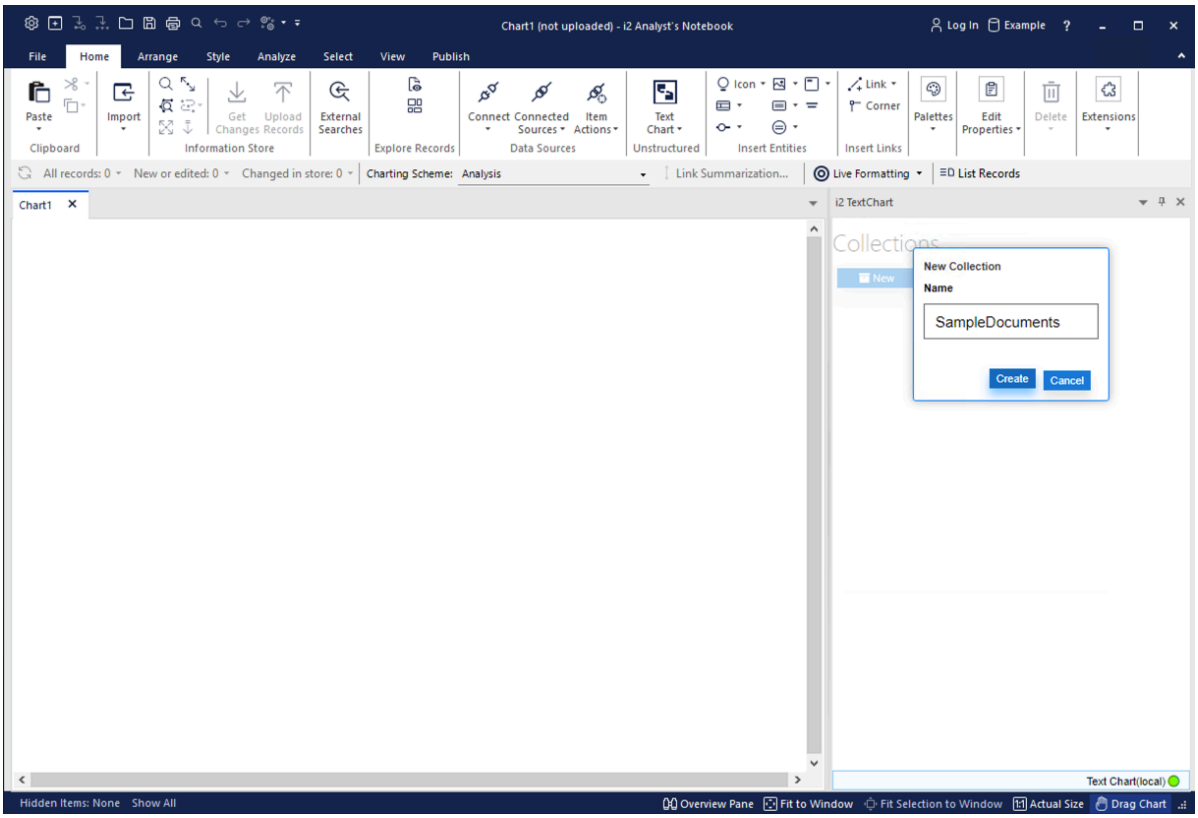
2. i2 TextChart stores documents in *collections*. A collection contains one or more documents that you choose to be in the same group for processing. You can start a new collection for the documents that you're importing, or add them to an existing one.

© N. Harris Computer Corporation



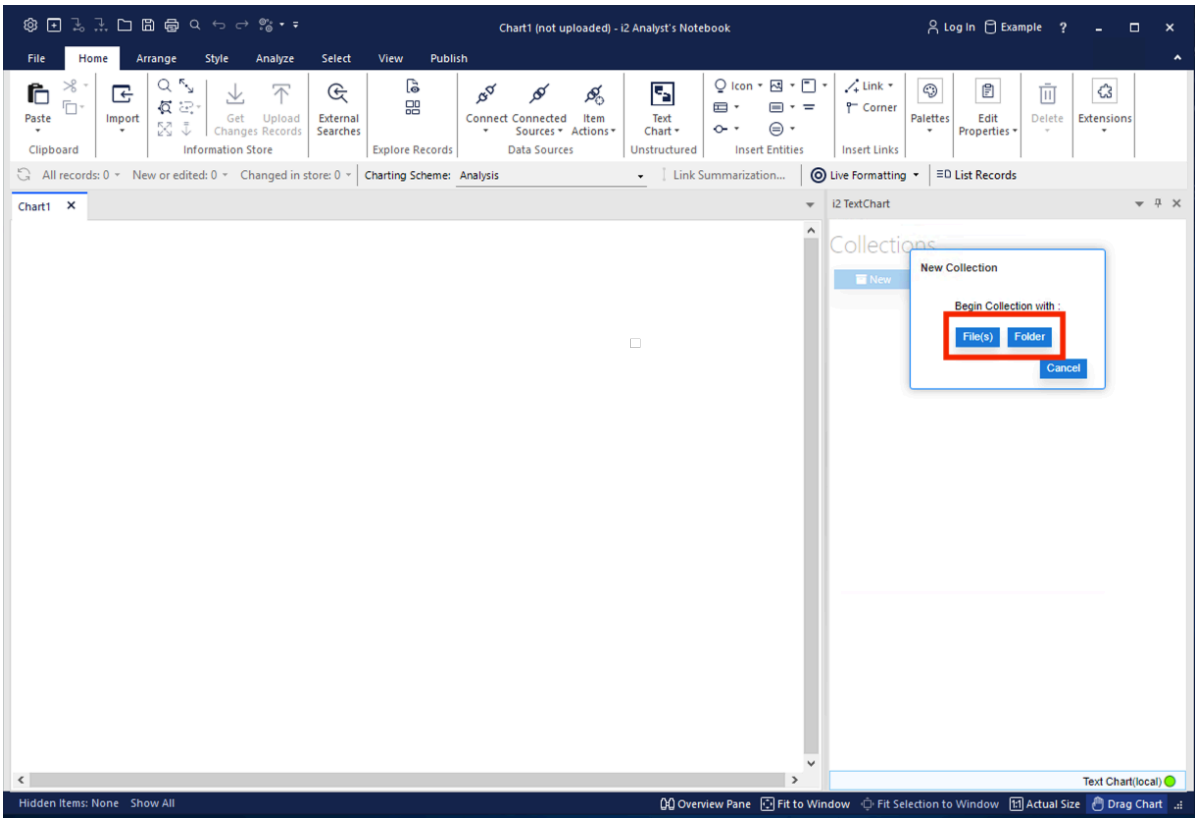
3. If you clicked **New**, provide a name for the new collection and click **Create**.

© N. Harris Computer Corporation



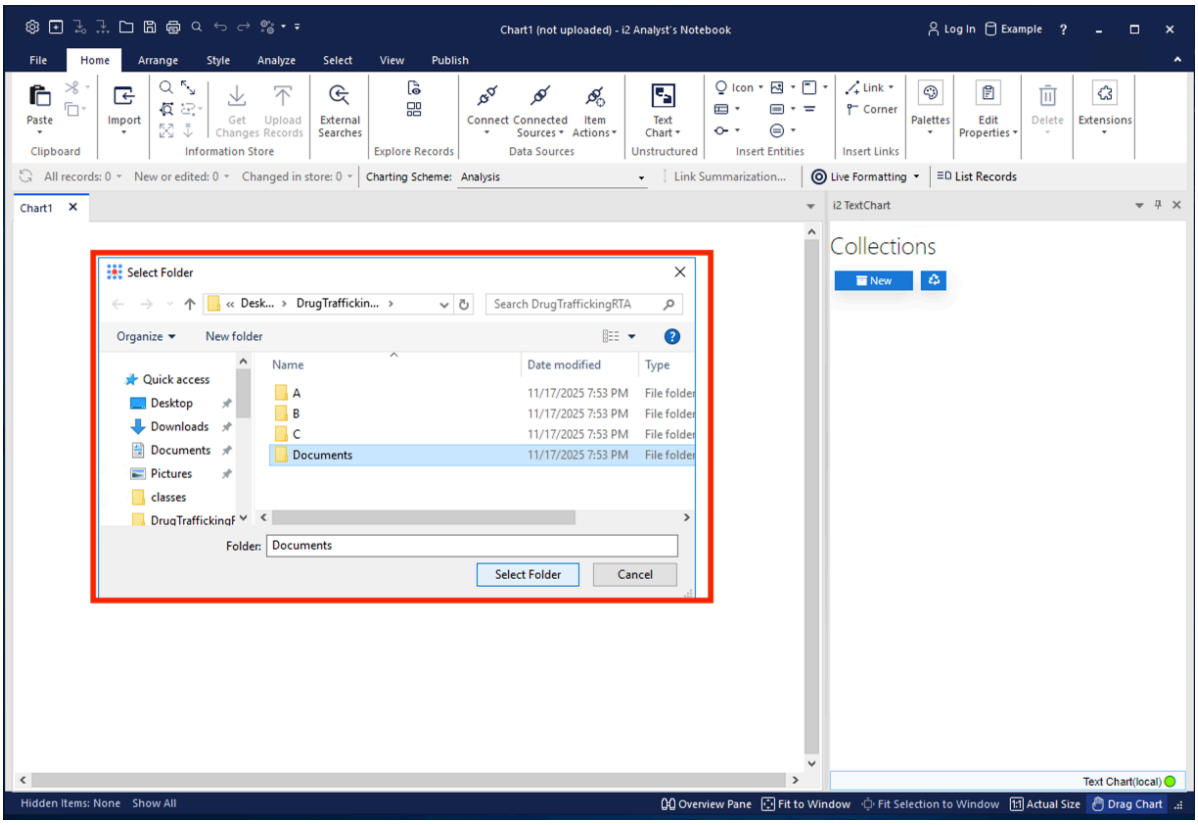
And then specify whether to populate the new collection with files or a folder.

© N. Harris Computer Corporation



4. Finally, select the files (or the folder that contains the files) that you want to import.

© N. Harris Computer Corporation



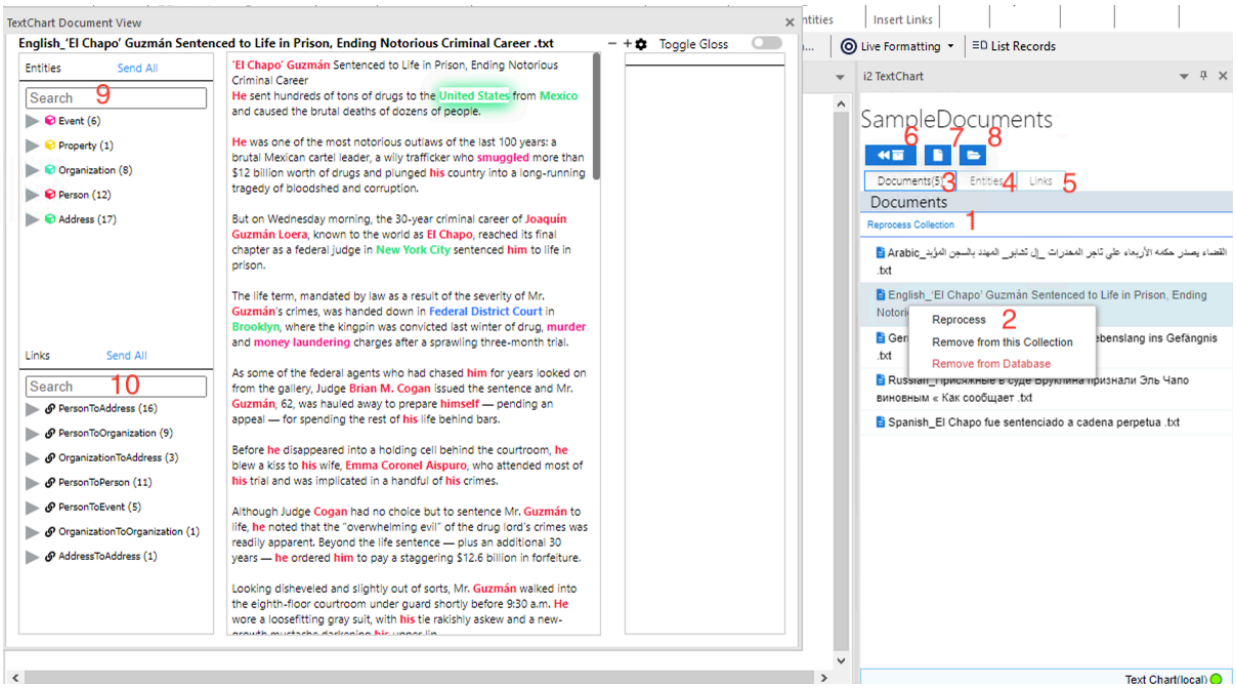
Processing documents

After you import documents into a collection, you can instruct TextChart to process them all at the same time, or to process each document separately.

Note: When you process documents for the first time, it can take a few seconds for the engine to load. Also, larger documents generally take longer to process than shorter ones. TextChart displays a spinning cursor while long-running operations take place.

The documents in a particular collection appear as a list in Analyst's Notebook, from which you can perform a number of tasks. Clicking a single document in the list processes that document and opens it in the Document View. Some of the other tasks are described below.

© N. Harris Computer Corporation



1. **Process Collection**

   When you click **Process Collection**, TextChart processes all the documents in the list that it has not already processed. The progress bar at the top of the pane shows the status of the operation. After each document is processed, its page icon turns blue.

   Canceling the operation stops processing for the current and subsequent documents in the collection. You can also ask TextChart to reprocess *all* documents in the collection, regardless of whether they've been processed before.

2. **Reprocess** / **Remove from this Collection** / **Remove from Database**

   If you right-click a single document in the list, the pop-up menu presents commands to reprocess the document, or to remove it from the collection, or to remove its processing results from the database.

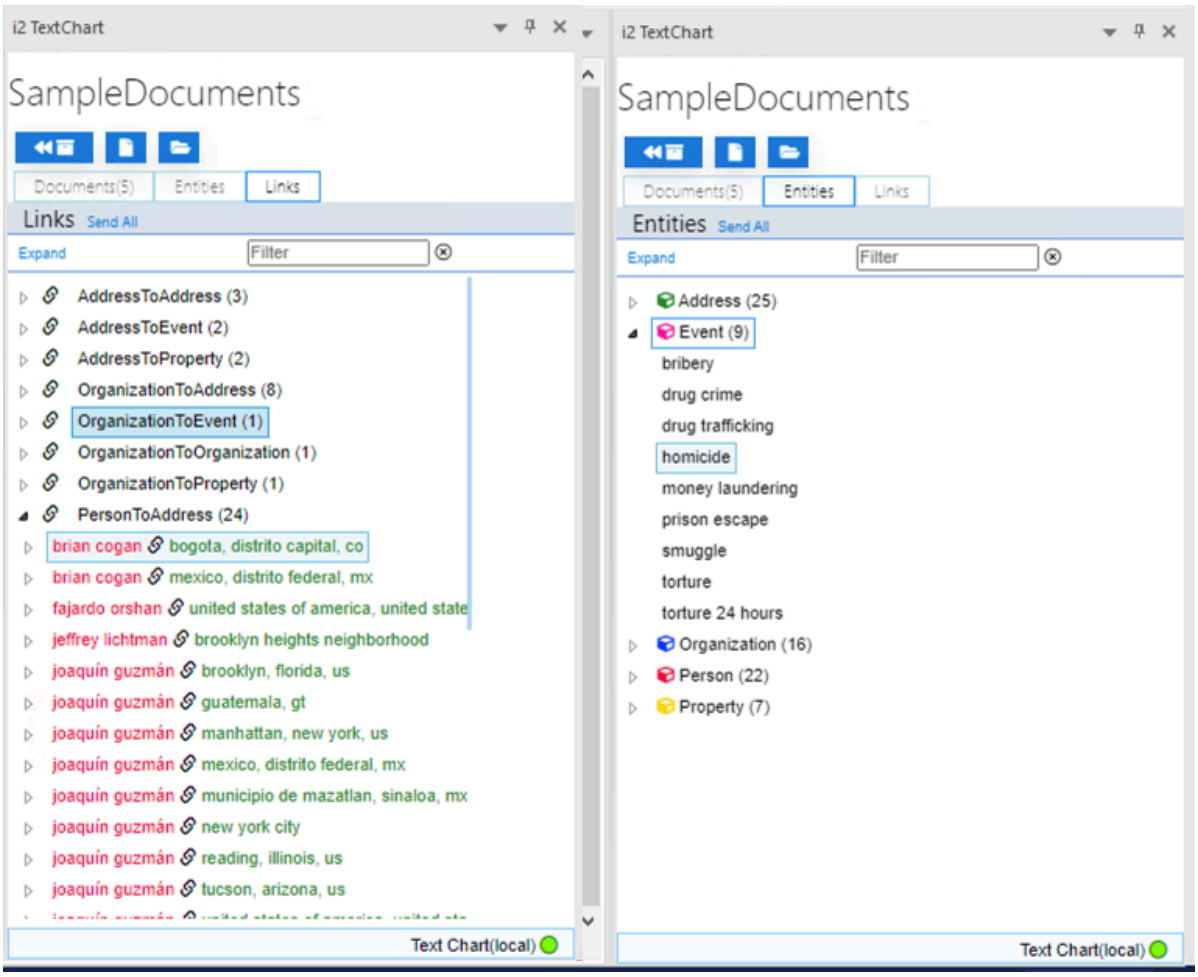
3. **Documents**

   The Documents tab displays a count of the documents in the collection.

4. **Entities**

   The Entities tab contains an entity tree, which lists the TextChart types of all the found entities in processed documents, as well as the count of each type of entity.

© N. Harris Computer Corporation



5. **Links**

   The Links tab contains a link tree that lists the TextChart types of all the found links in processed documents, as well as the count of each type of link.

   You can expand the link types in the tree view to see lists of the individual links and the predicate types that TextChart identified during processing.

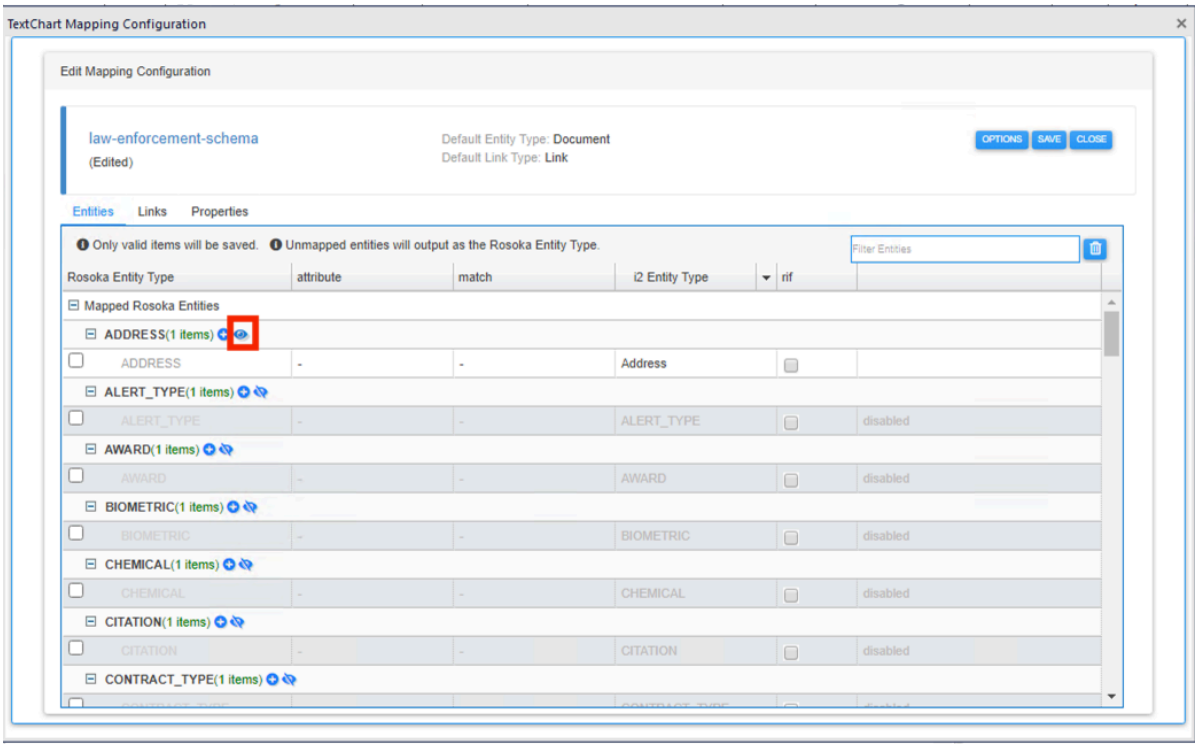
6. Click **Back to Collections** to return to the list of collections.
7. Click **Add More Files** to add more files to the current collection.
8. Click **Add Directory to Collection** to add all the documents in a particular directory to the current collection.

Ignoring TextChart entity, link, or attribute types

By default, TextChart attempts to find entities and links of all the types that it understands, and values for all the attributes they can contain. If there are some types that you don't want TextChart to extract, you can disable them through the Mapping Configuration.

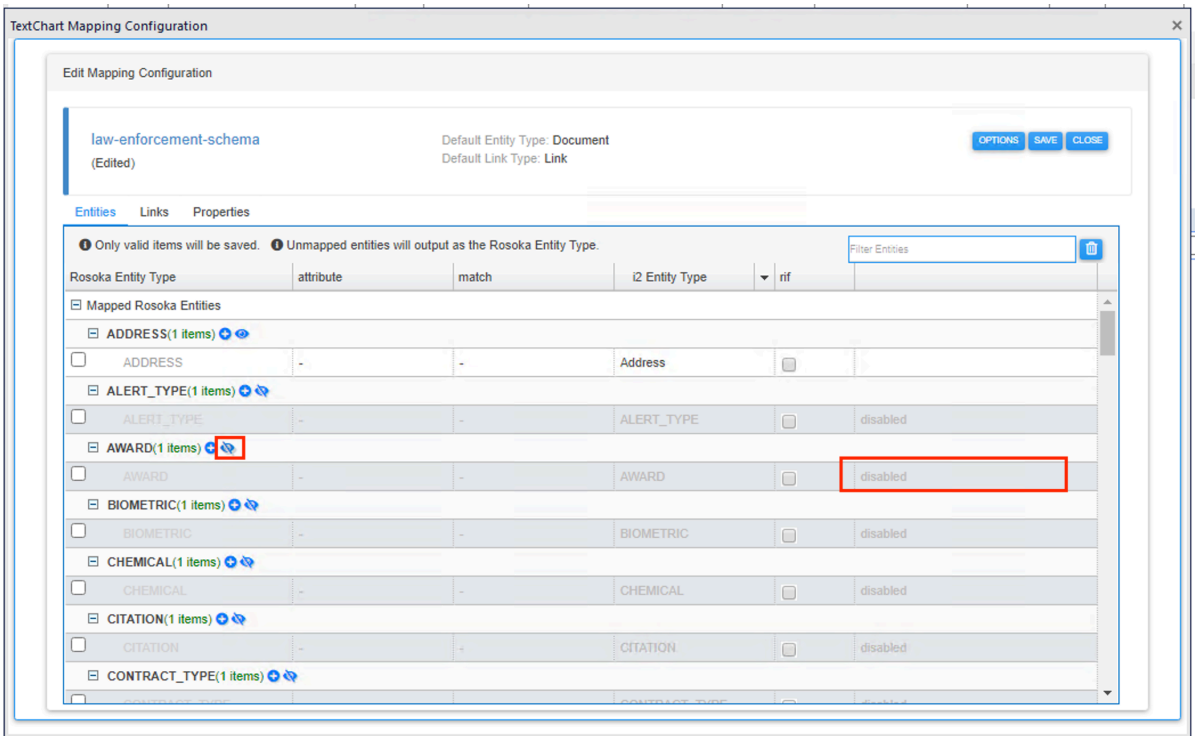
© N. Harris Computer Corporation

1. From the Analyst's Notebook ribbon, select **Text Analytics > Mapping Configuration** to open the Mapping Configuration window.
2. Click **Edit** to display the configuration settings for the active mapping. The left column of the **Entities**, **Links**, and **Properties** tabs contains a list of TextChart types.



3. Click the "eyeball" icon next to the name of any type that TextChart should not extract from the documents it processes.

© N. Harris Computer Corporation



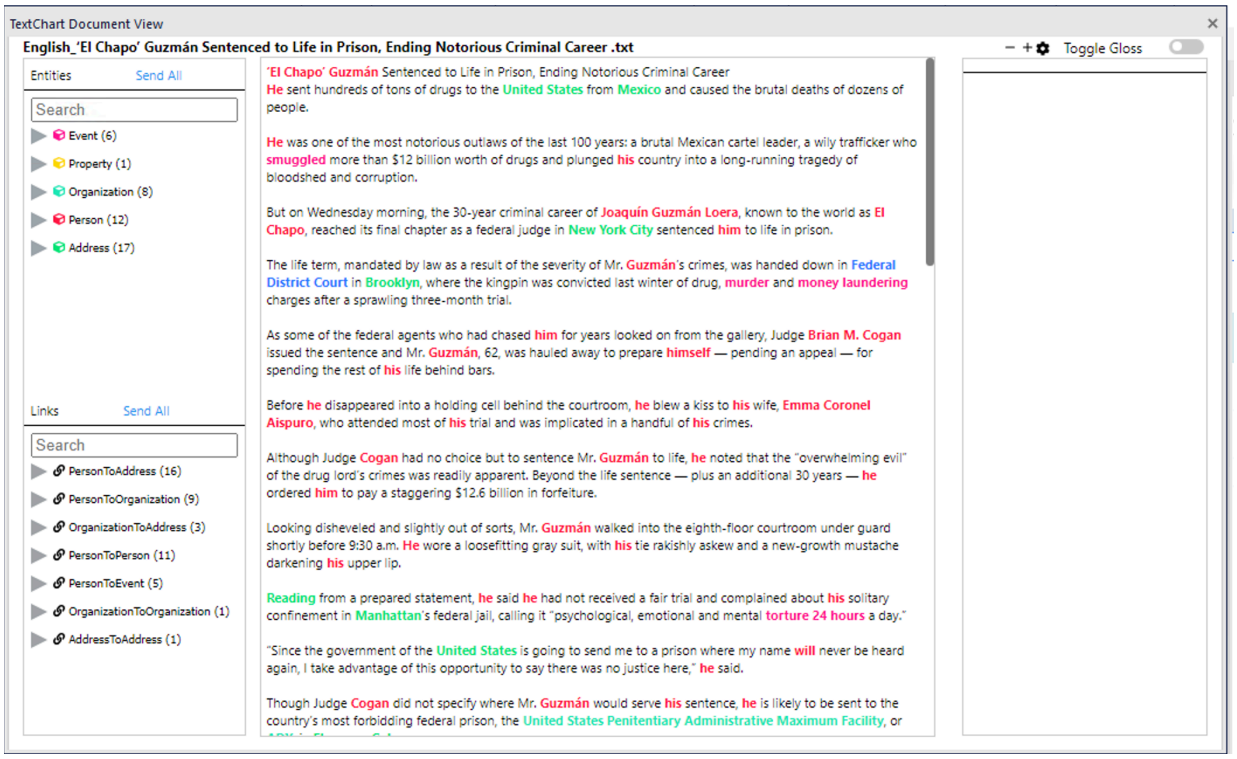
4. If you processed some documents before changing these settings, clear those results and reprocess the documents.

Viewing documents

After TextChart processes a document (or the first of several documents), it opens the Document View window, which presents an enriched view of the processed text.

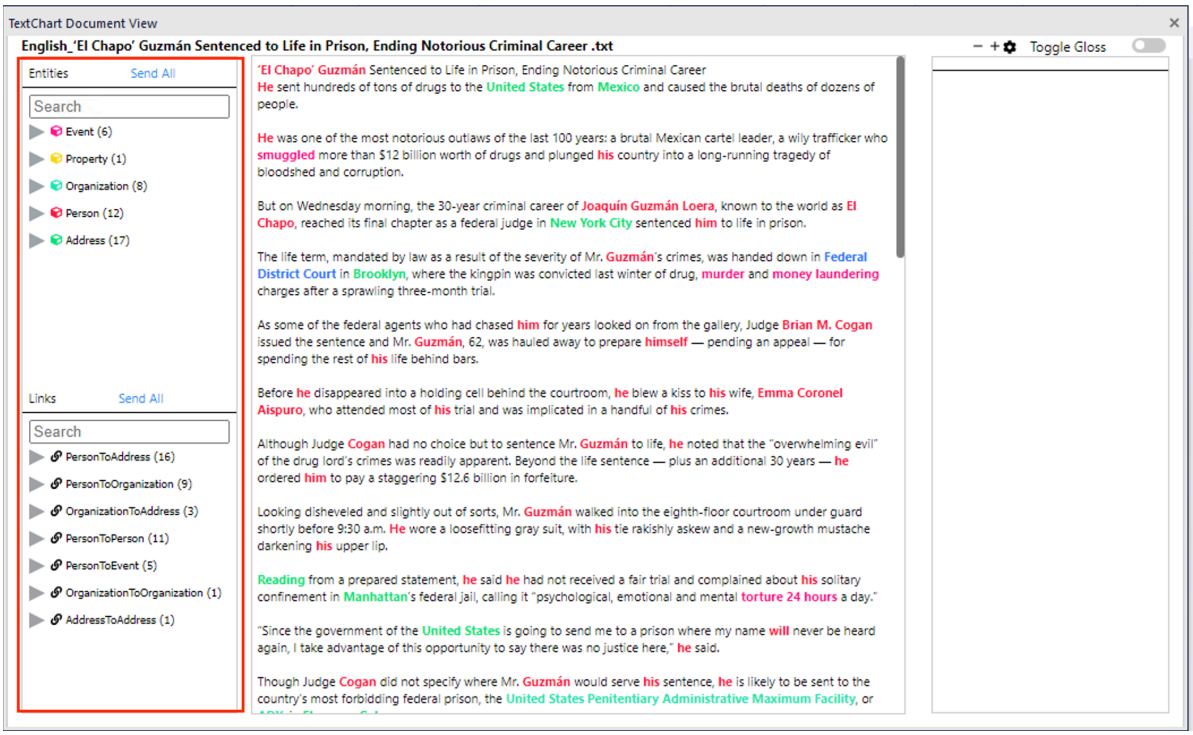
The Document View highlights all of the found entities in context, using different colors for different types. In this view, you can review and curate the entities that TextChart identified.

© N. Harris Computer Corporation



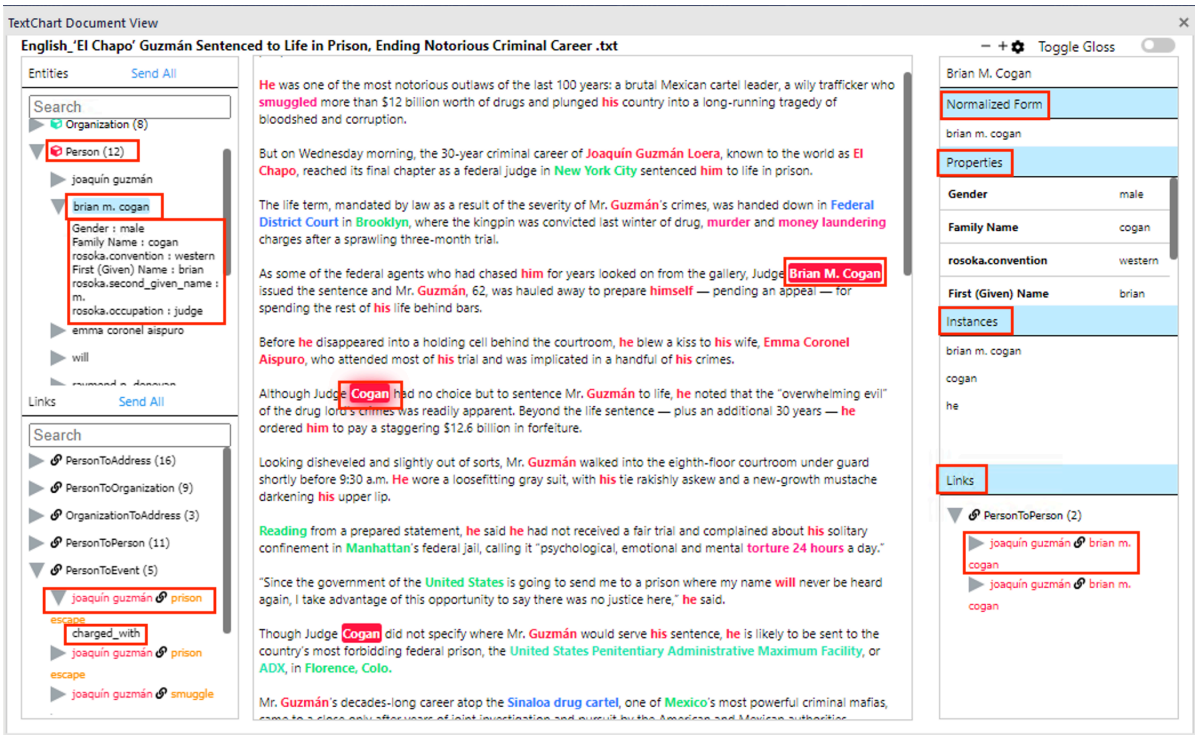
1. The button at the upper left of the window opens entity and link trees for the current document.

© N. Harris Computer Corporation



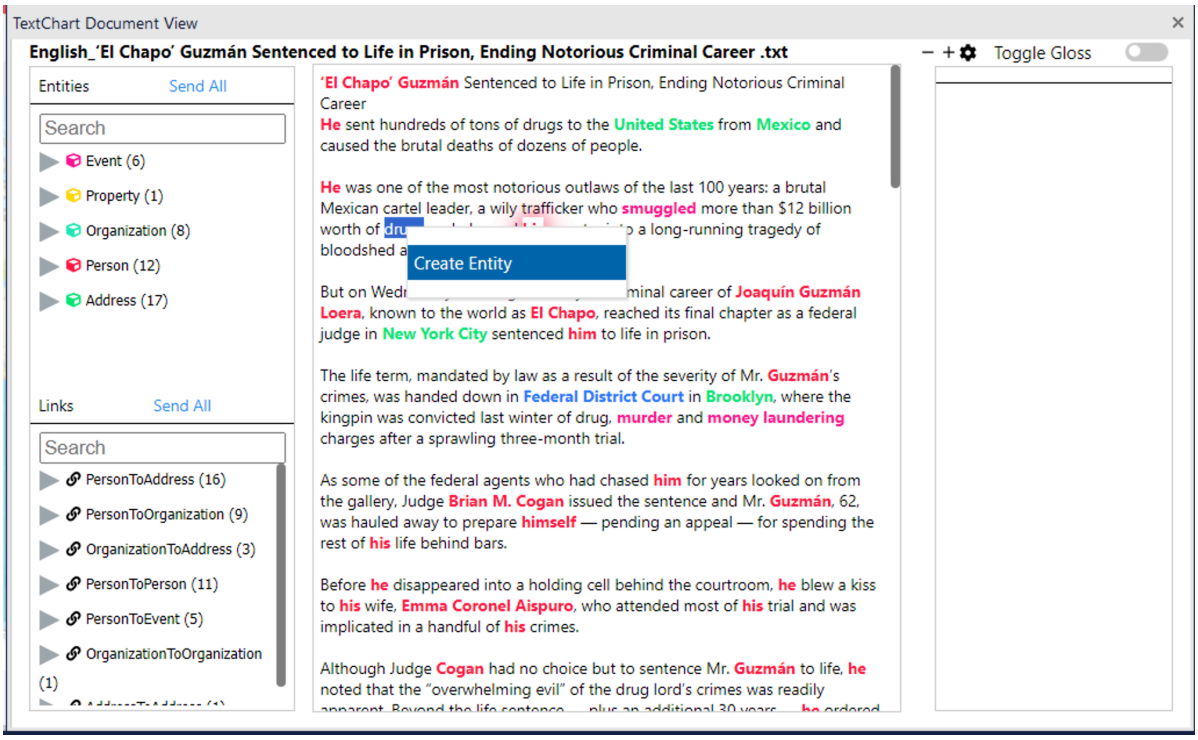
The trees behave initially like the ones in the list for all the documents in a collection. However, when you select entities and links (or groups of them), TextChart highlights them in the open document and displays a second panel that contains more information.

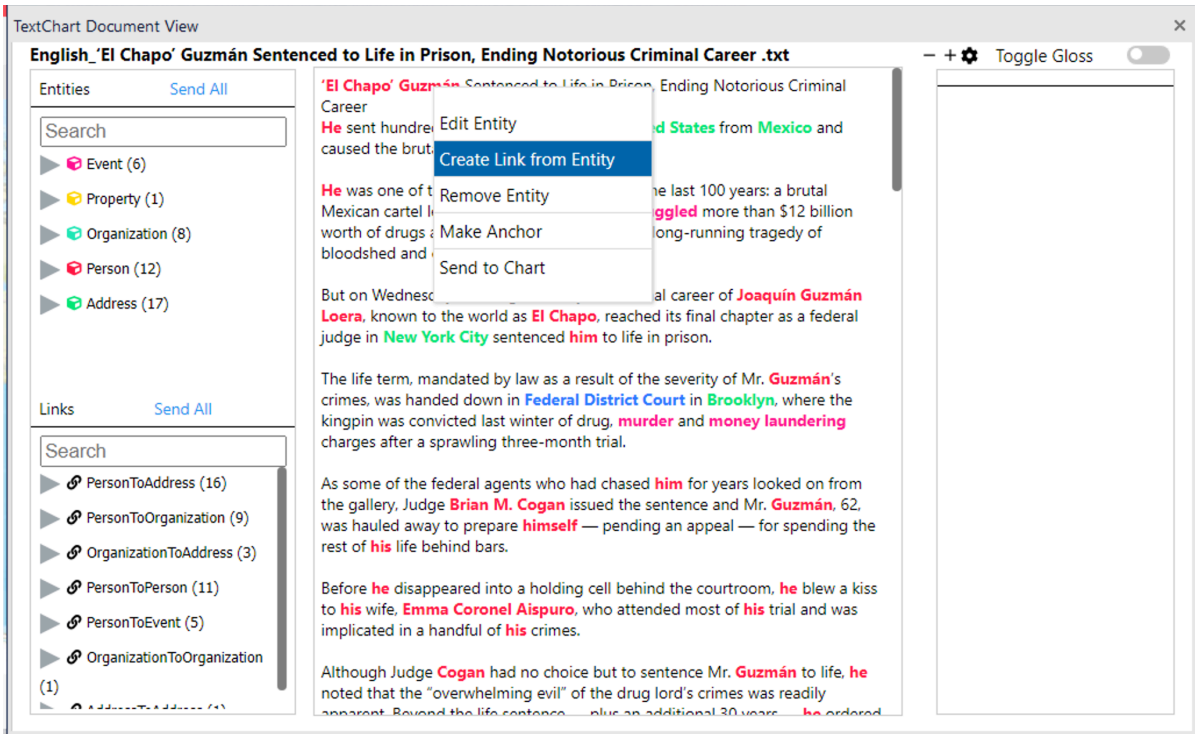
© N. Harris Computer Corporation



2. Use the **Tag** drop-down to tag a new entity or link.

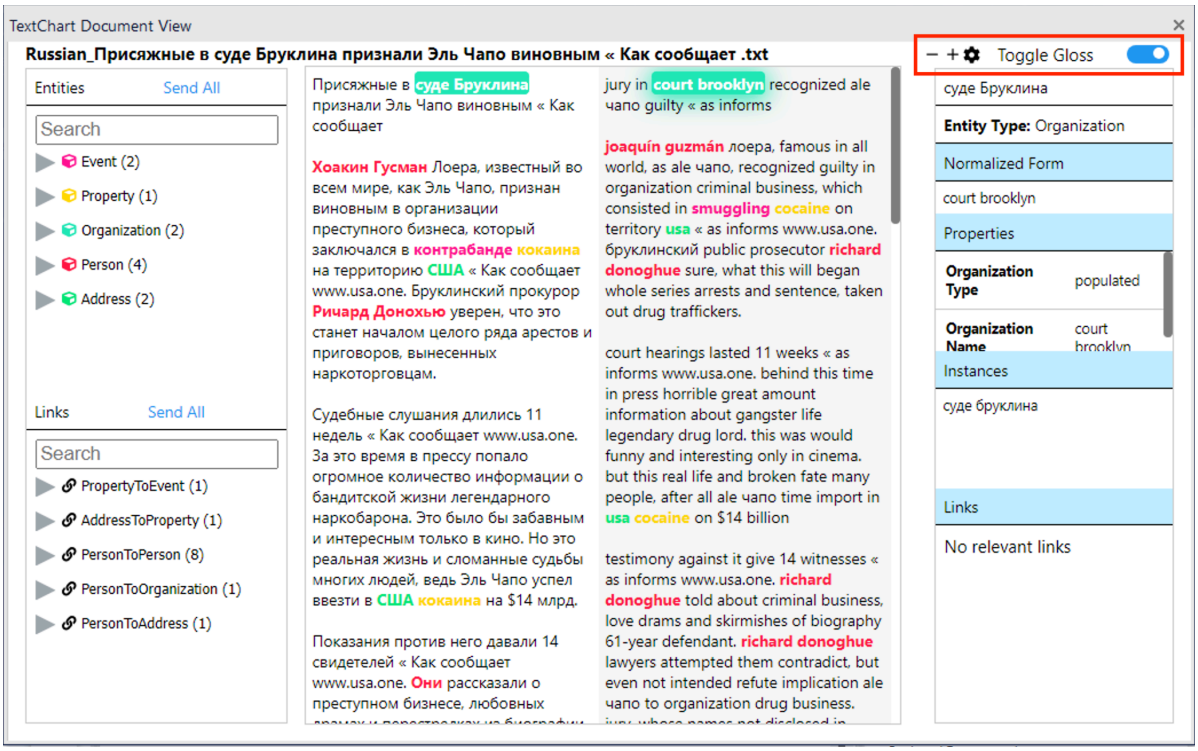
© N. Harris Computer Corporation





© N. Harris Computer Corporation

3. Through the **Options** button, you can:
   a. Toggle back and forth between an English version of the document, and the text in its original language.
   b. Remove whitespace from the document as displayed
   c. Change the font size of the text in the document.



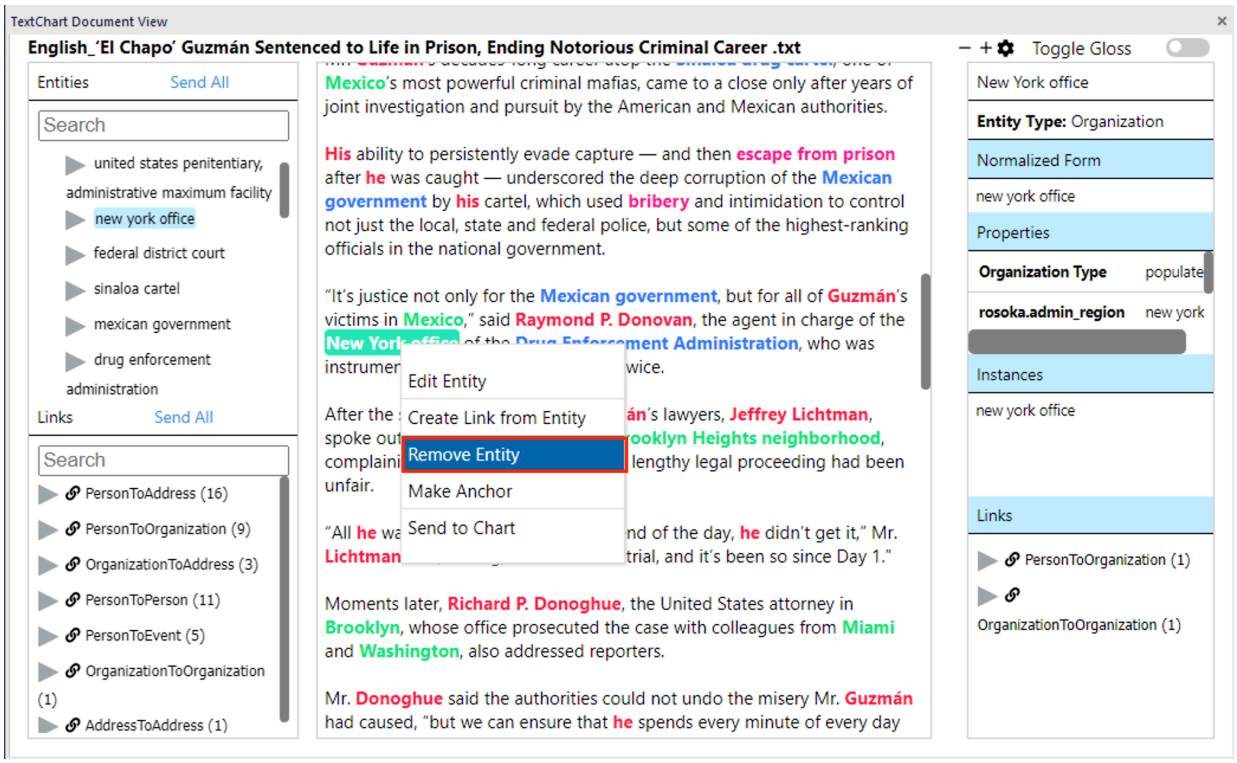
Curating extraction results

After i2 TextChart presents the results of processing a document in the Document View, you can review its output and make changes to the results. TextChart supports removing and modifying the records it finds, as well as identifying new records of your own.

Removing results

If TextChart identifies an entity or a link that you don't want to appear in the results, you can remove it from the Document View. For example, the following view contains a contrived entity extraction result for a "New York office", which has the Organization entity type.

© N. Harris Computer Corporation

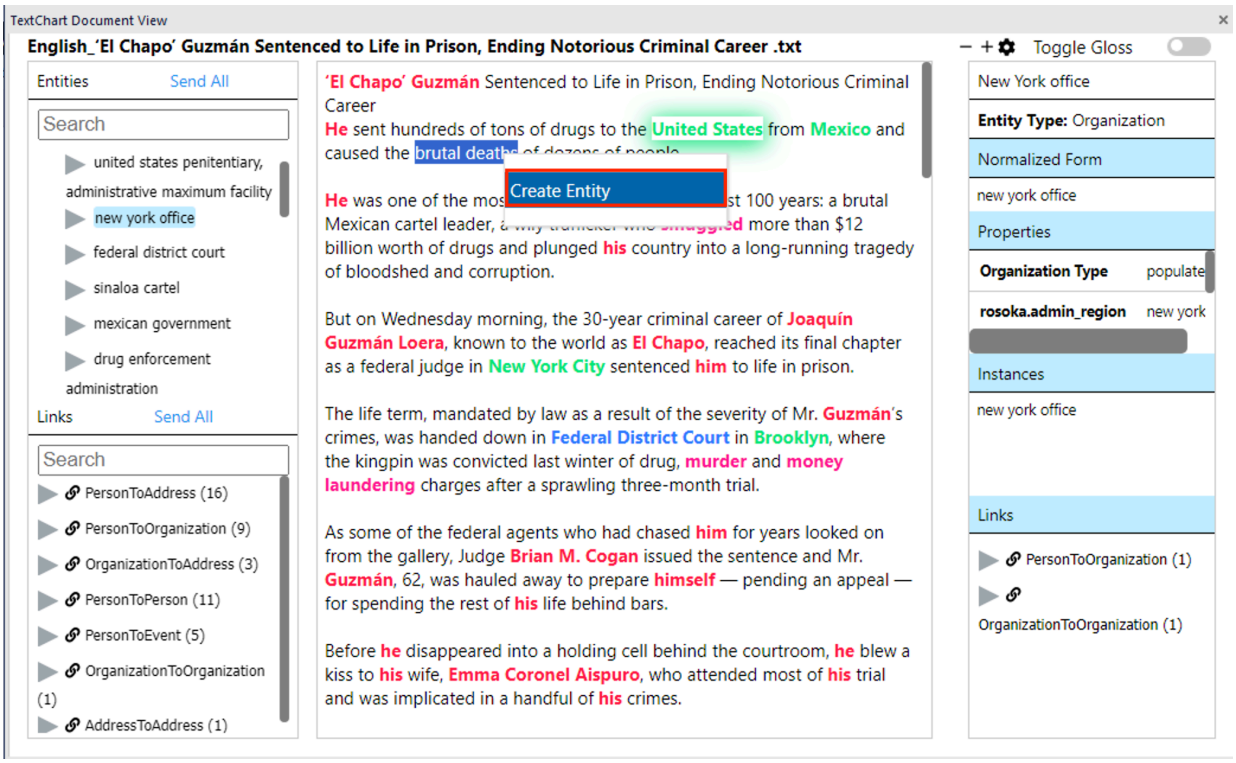


When you select the text in the Document View, TextChart displays information about the extracted entity. To remove this result, click **Remove Entity**.

Adding entities

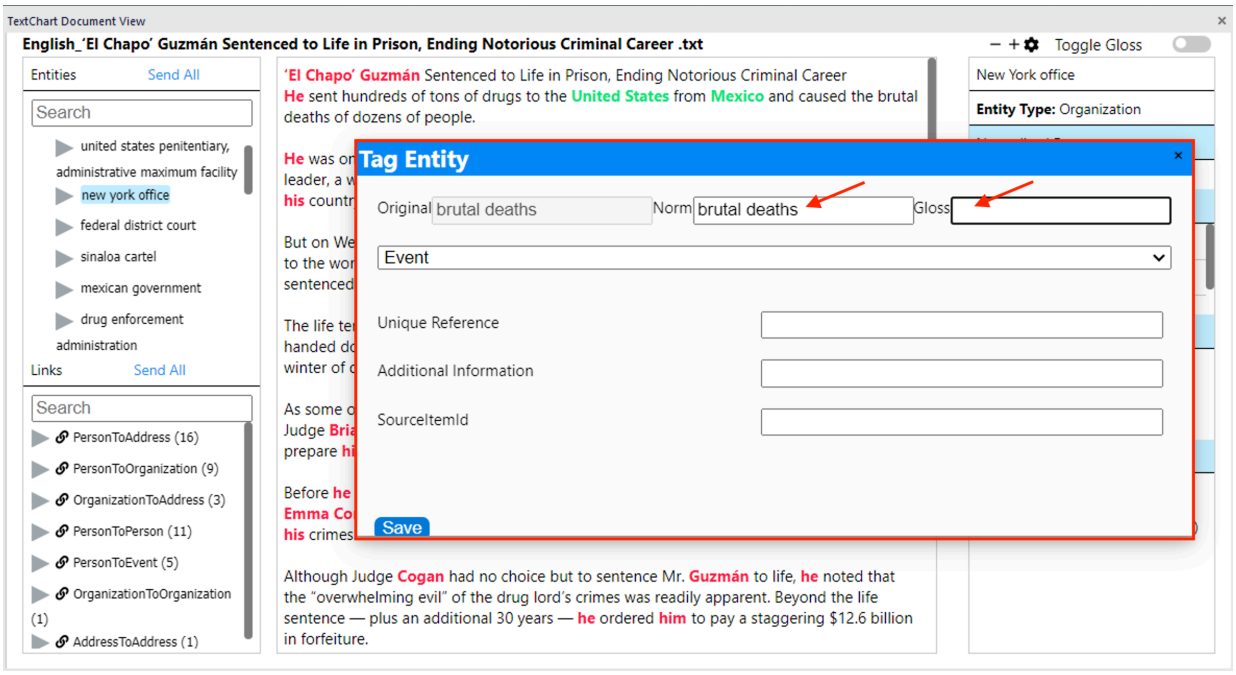
If TextChart failed to identify an entity that you *do* want to appear in the results, you can highlight the text to be extracted, and then select **Create Entity**.

© N. Harris Computer Corporation



TextChart displays the Tag Entity Tool for you to provide information about the entity, and populates the **Original** field with the text that you highlighted.

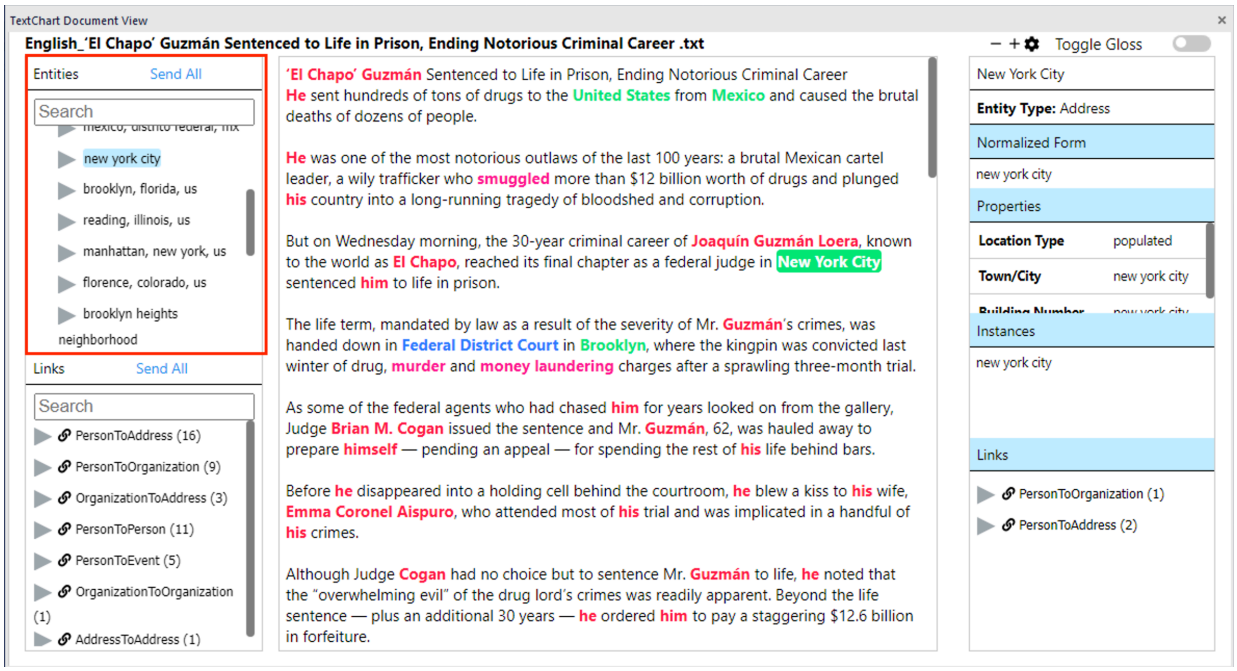
© N. Harris Computer Corporation



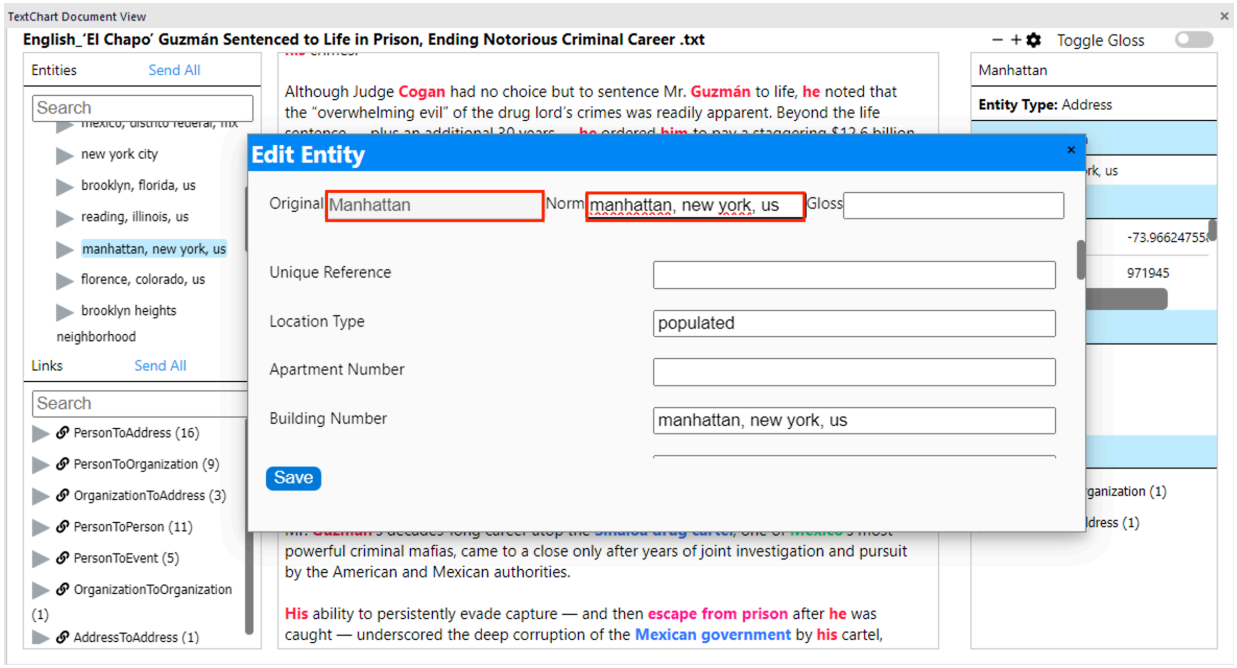
TextChart also populates the **Norm** field with the same text, but here you have the option of changing the text to normalize it when a document uses different terminology to refer the same piece of information.

For example, in the image below, "new york city" and "manhattan, new york, us" have been identified as separate places.

© N. Harris Computer Corporation



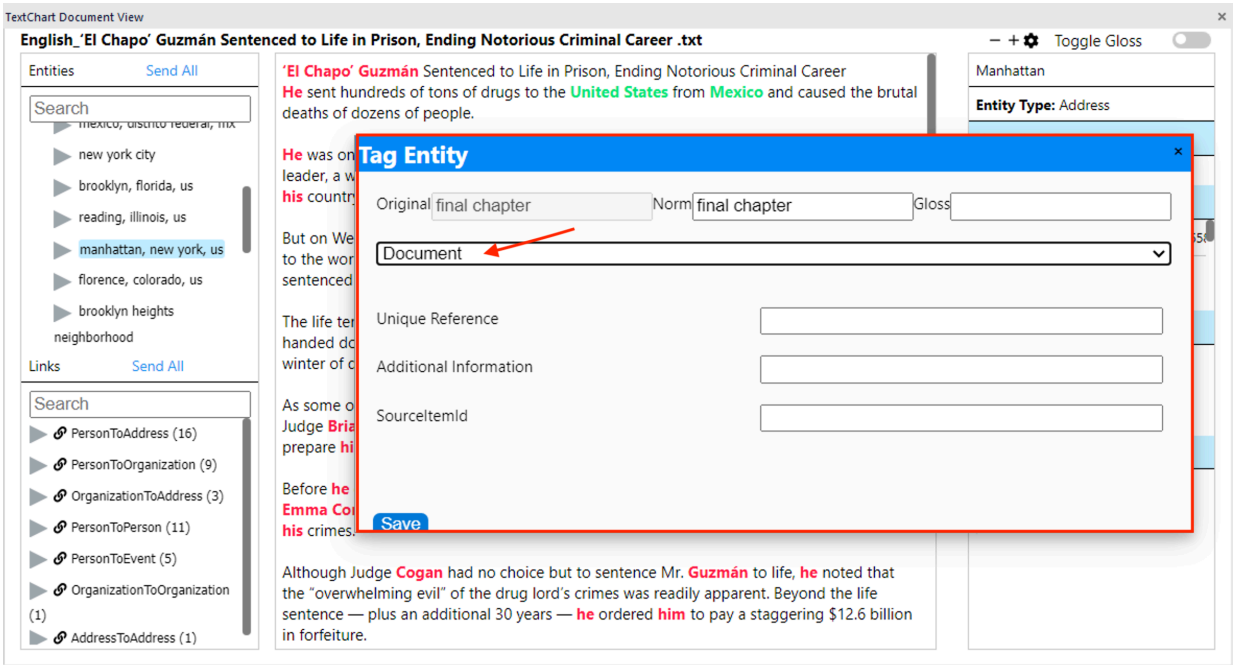
If you decide that TextChart should treat these instances as the same place, you can edit the **Norm** field of one so that it matches the other.



© N. Harris Computer Corporation

If you're processing documents in a language other than English, and you come across an important term that TextChart does not understand, you can add an English translation of that term to the **Gloss** field.

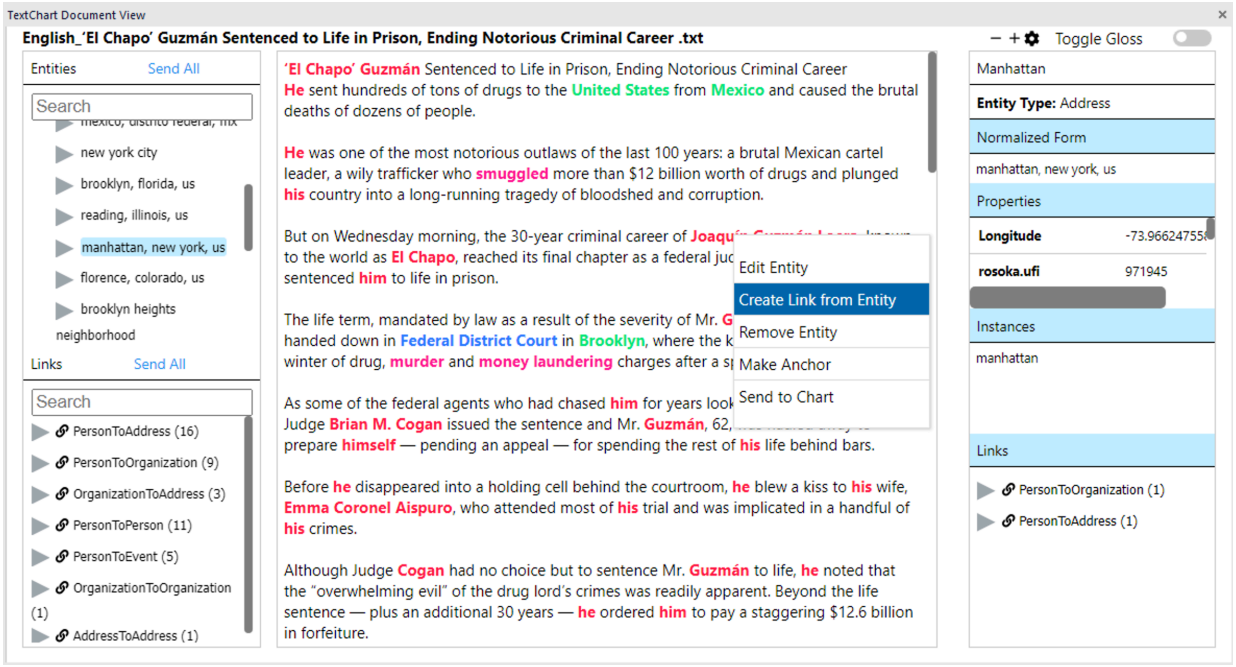
Finally, you must choose a TextChart type for your new entity. Different entity types have different attributes, and you can fill in additional information as you see fit.



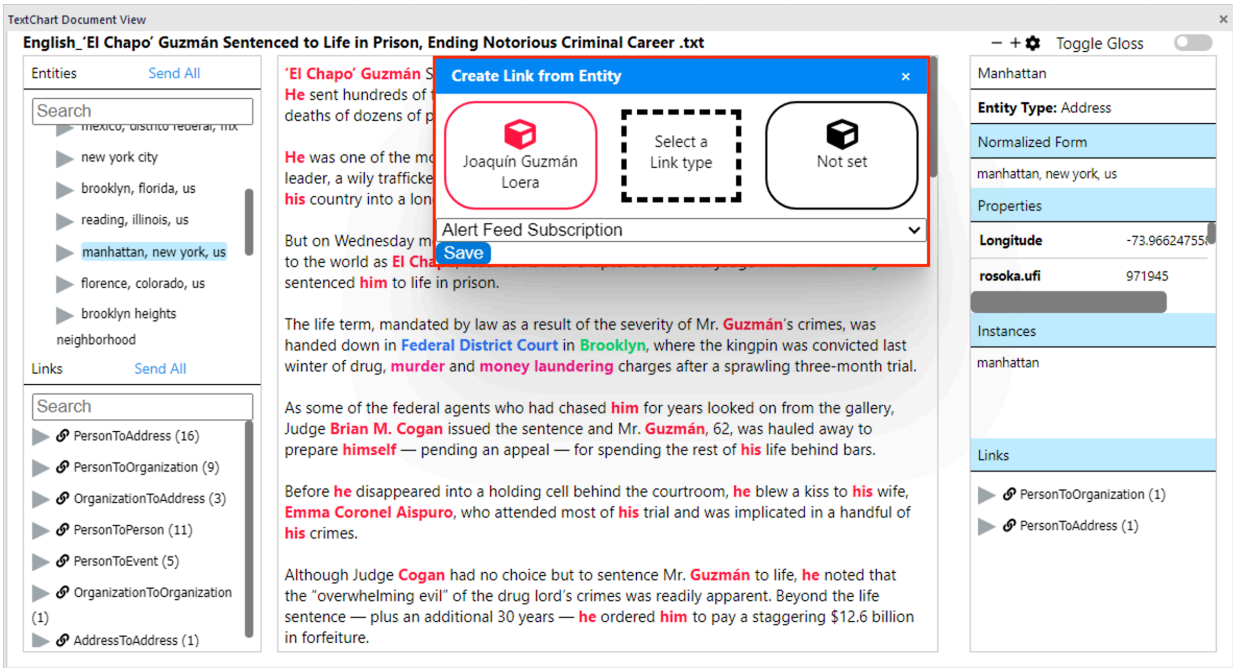
Adding links

Adding a link that TextChart failed to identify is similar to adding an entity in the same circumstances. You can highlight, right-click an extracted entity and select **Create Link from Entity** to use it as the "From" end of a new link.

© N. Harris Computer Corporation

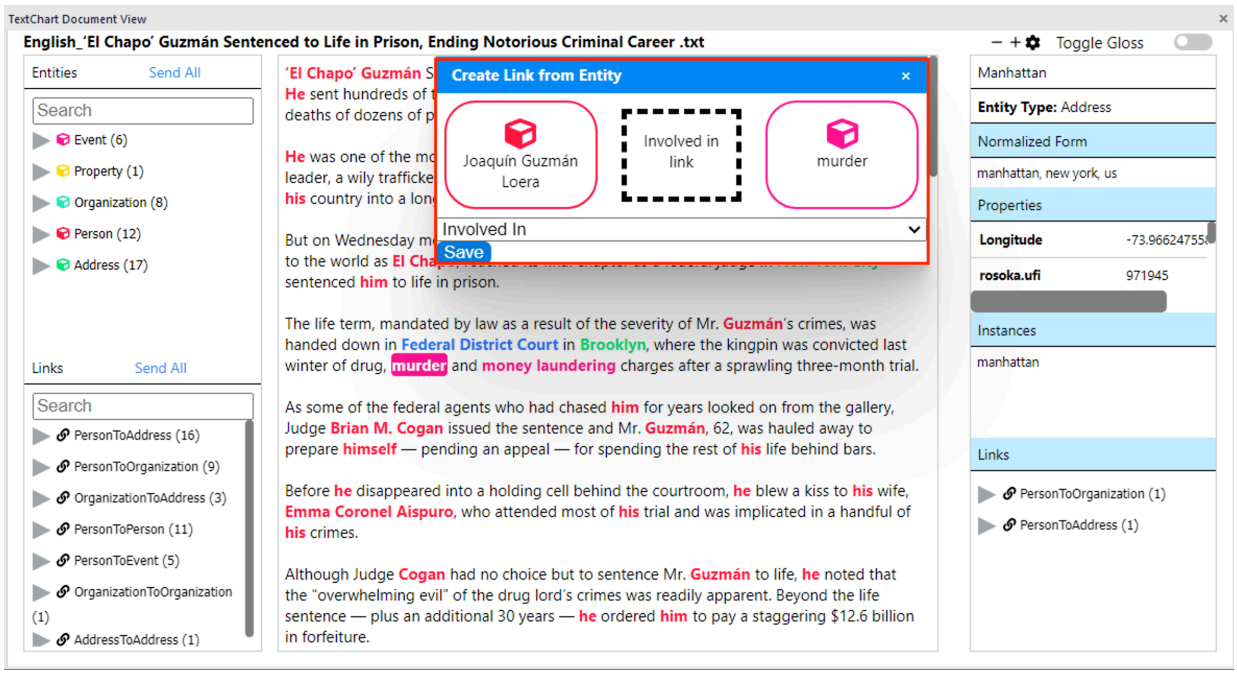


The Tag Link Tool starts by prompting you to select the entity at the "From" end.

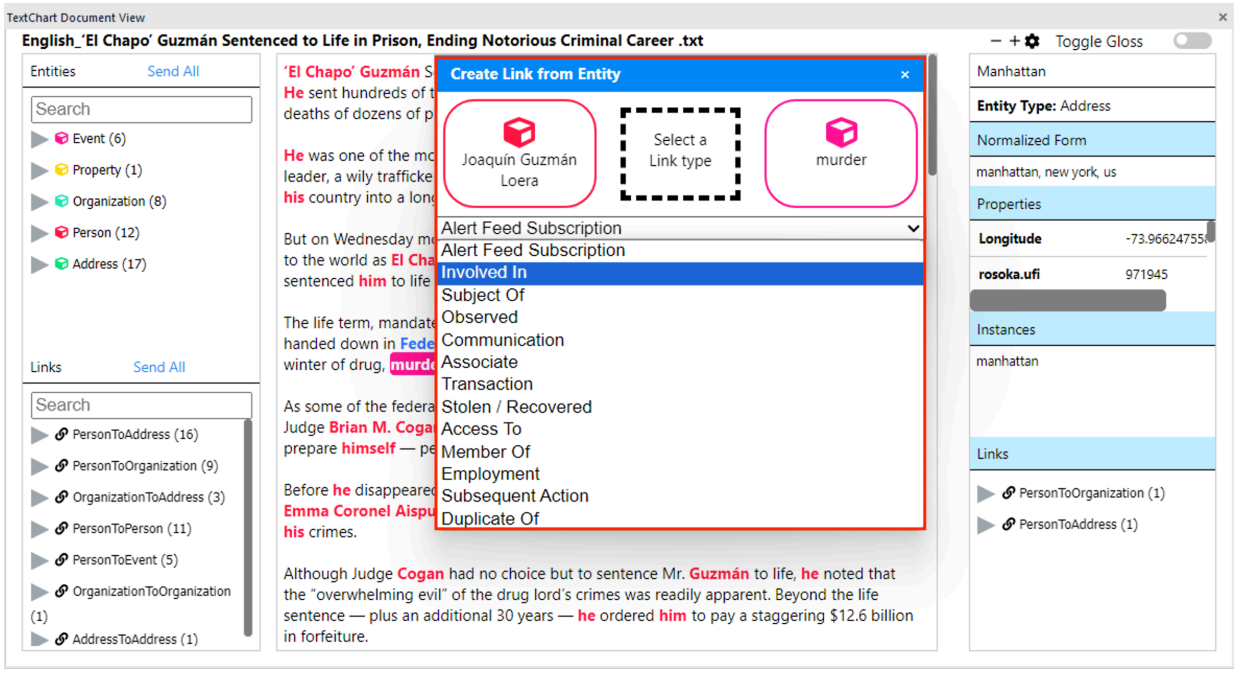


Then, you have to select the entity at the "To" end.

© N. Harris Computer Corporation

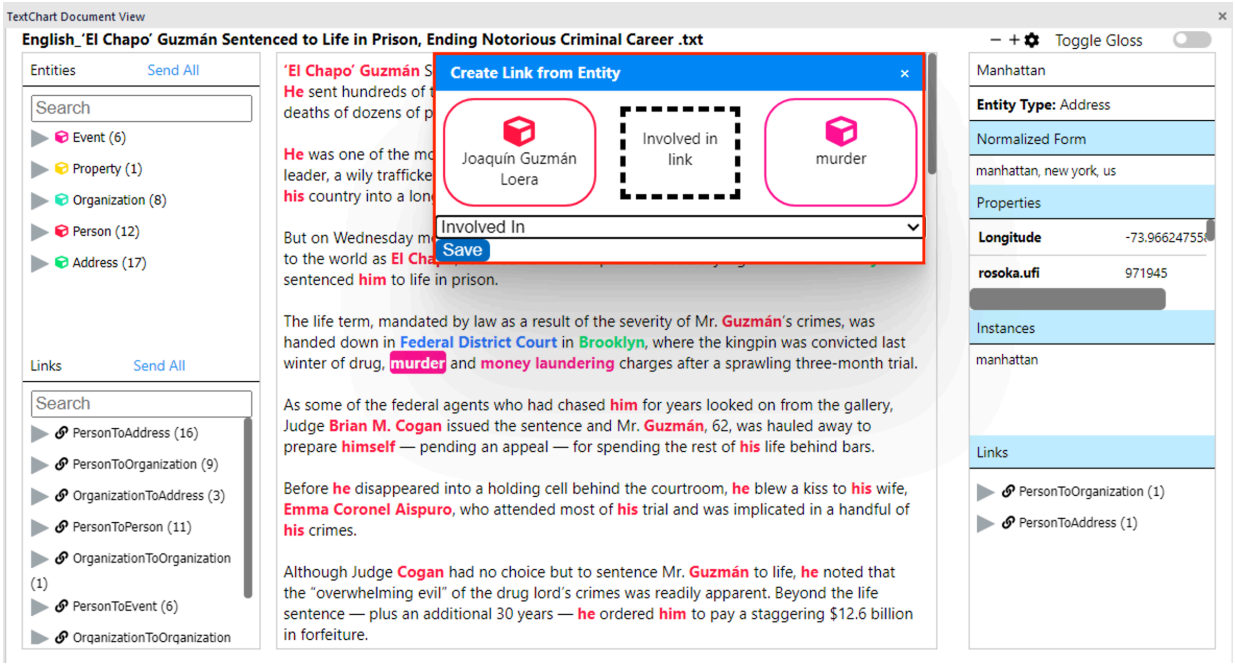


After you select both ends, the Tag Link Tool displays a drop-down that allows you to specify the TextChart type of the link to be created.



When all the information in the Tag Link Tool is complete, click **Save** to add the new link to the results.

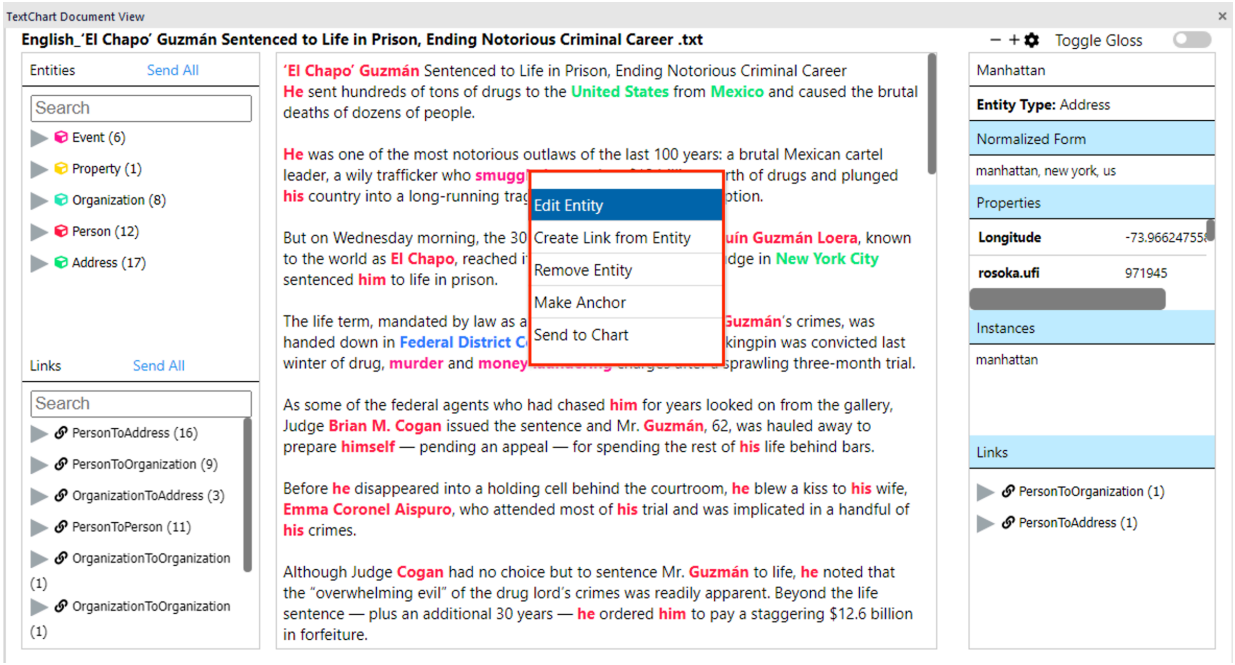
© N. Harris Computer Corporation



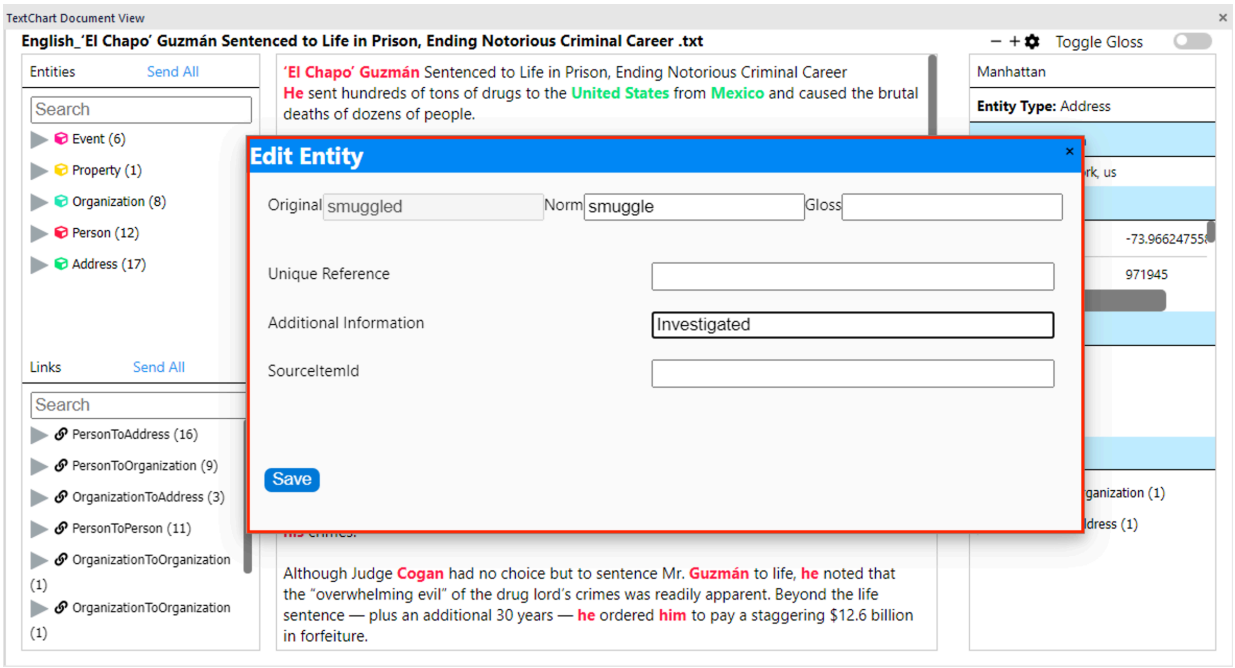
Editing results

If TextChart identified an entity or a link successfully, but the extracted result isn't exactly how you want it, you can right-click the highlighted text in the Document View and select **Edit Entity**.

© N. Harris Computer Corporation

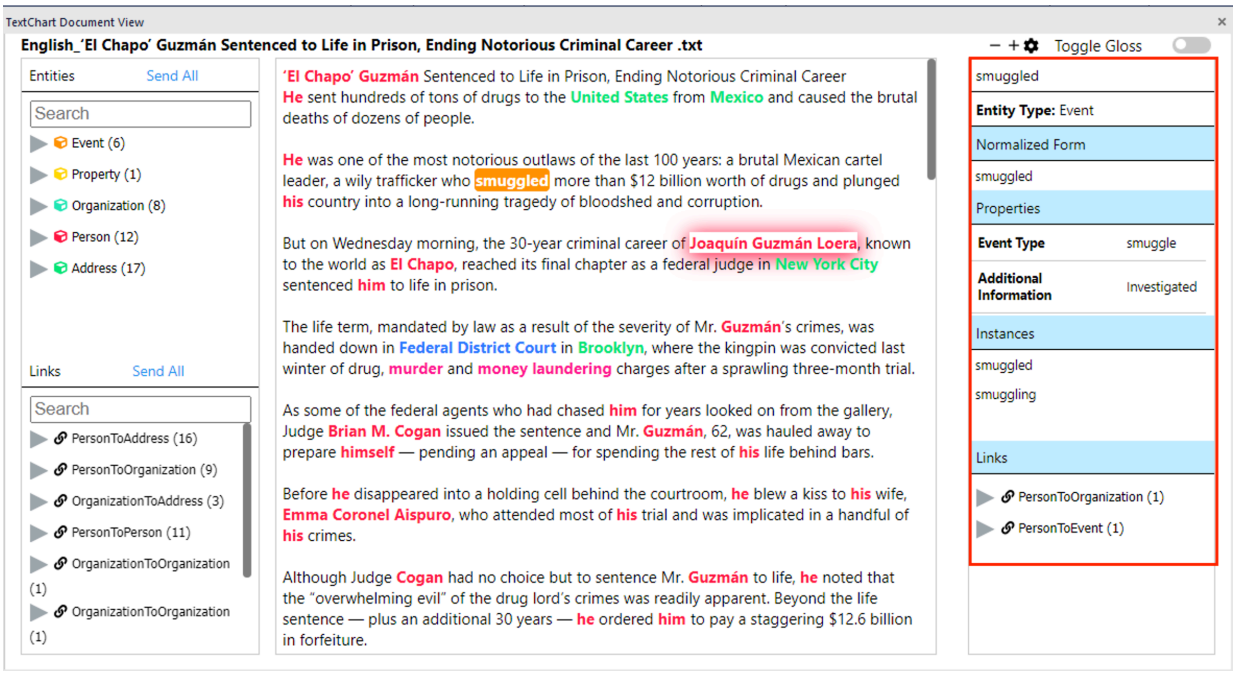


The behavior of the tool is the same when you're editing results as it is when you're adding them. You can edit the **Norm** field, add an English **Gloss**, change the **Type**, and modify attribute information.



© N. Harris Computer Corporation

When you finish editing, the changes you made are reflected in the feature bar on the right of the Document View.



Important: After you edit an extracted entity or link, you must reprocess the other documents in the collection so that they receive the same modifications.

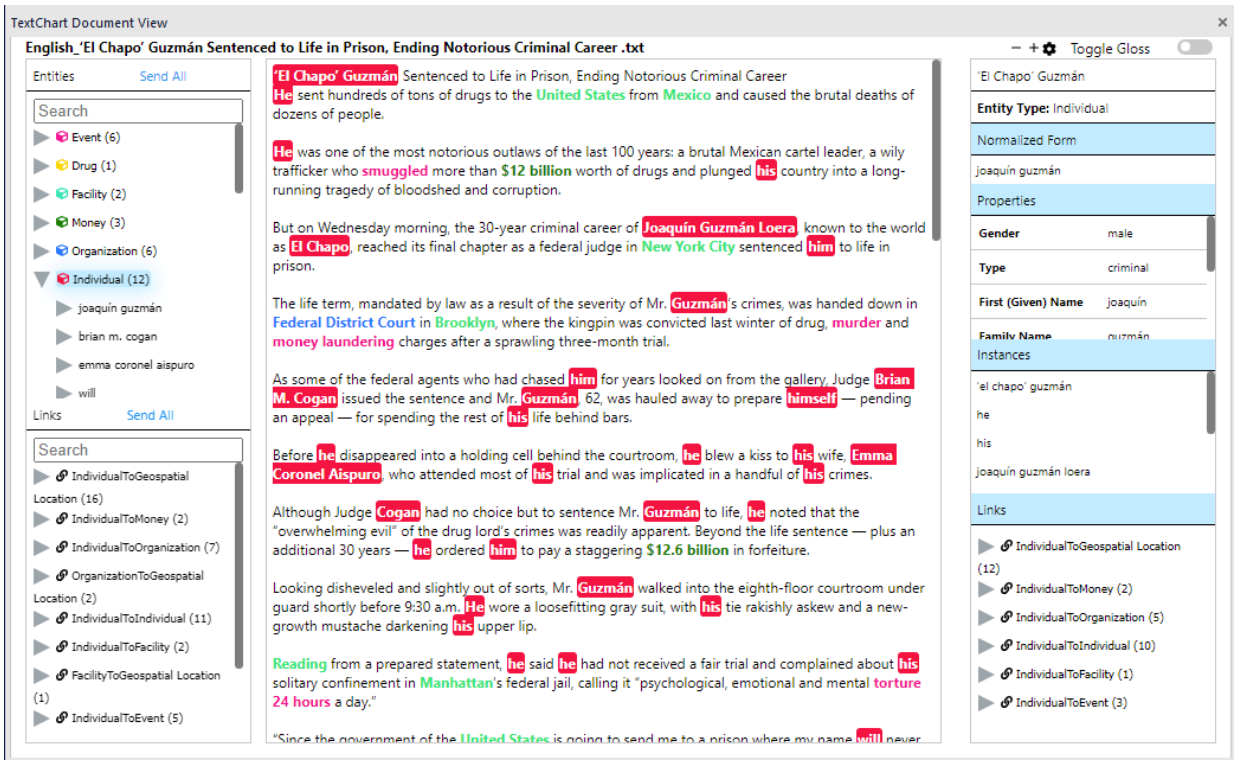
Analyzing extraction results

After i2 TextChart processes a document, it presents its results in the Document View for you to assess. The views that it presents for extracted entities and extracted links are slightly different from each other.

Viewing entities

When you select an entity type in the **Entities** list, TextChart highlights the entities of that type in the document.

© N. Harris Computer Corporation

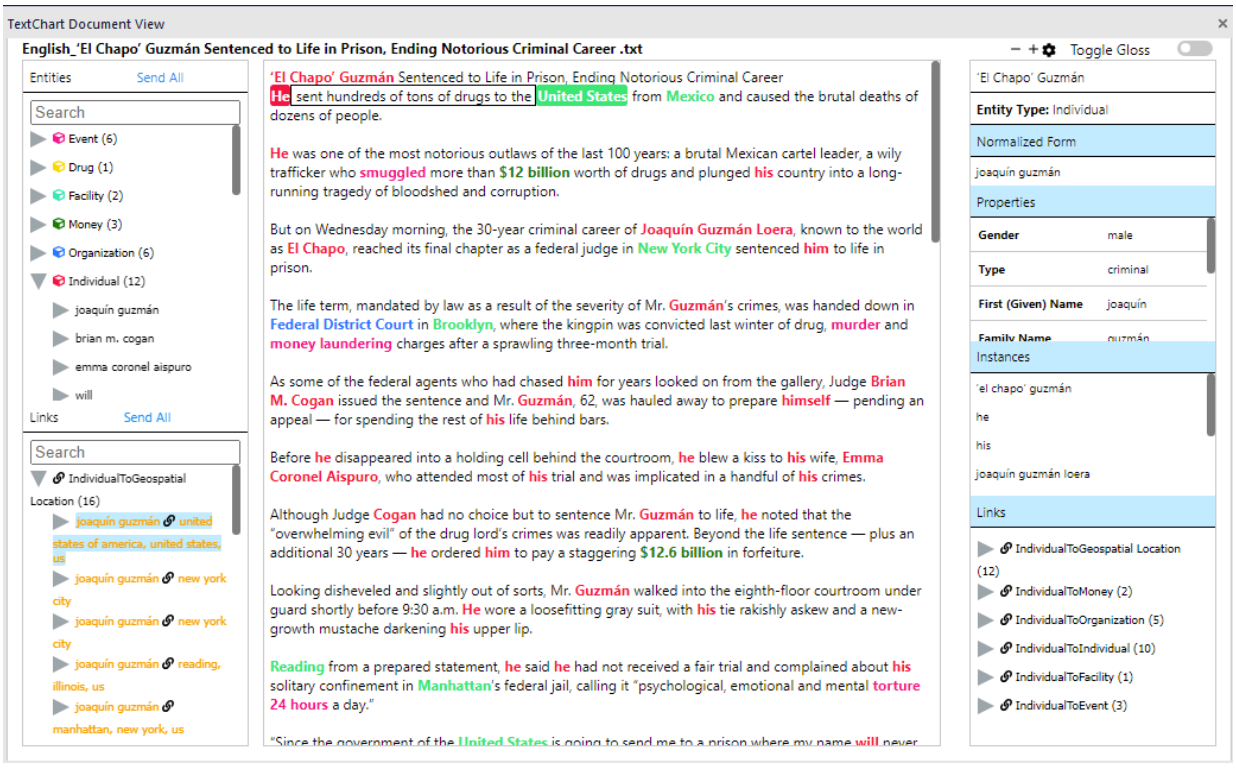


The list also displays the number of entities of each type in the document. If you expand an entity type and select a particular instance, TextChart opens the entity modification window on the right of the Document View.

Viewing links

When you select a link type in the **Links** list, TextChart highlights the entities and the predicates that contribute to links of that type.

© N. Harris Computer Corporation



As with the extracted entities, you can expand link types in the list to investigate and modify the information of any individual extracted link.

Charting extraction results

After i2 TextChart has imported and processed a collection of documents and you have curated the results, you can add the extracted entities and links to an Analyst's Notebook chart. TextChart supports sending results to the chart separately or in bulk.

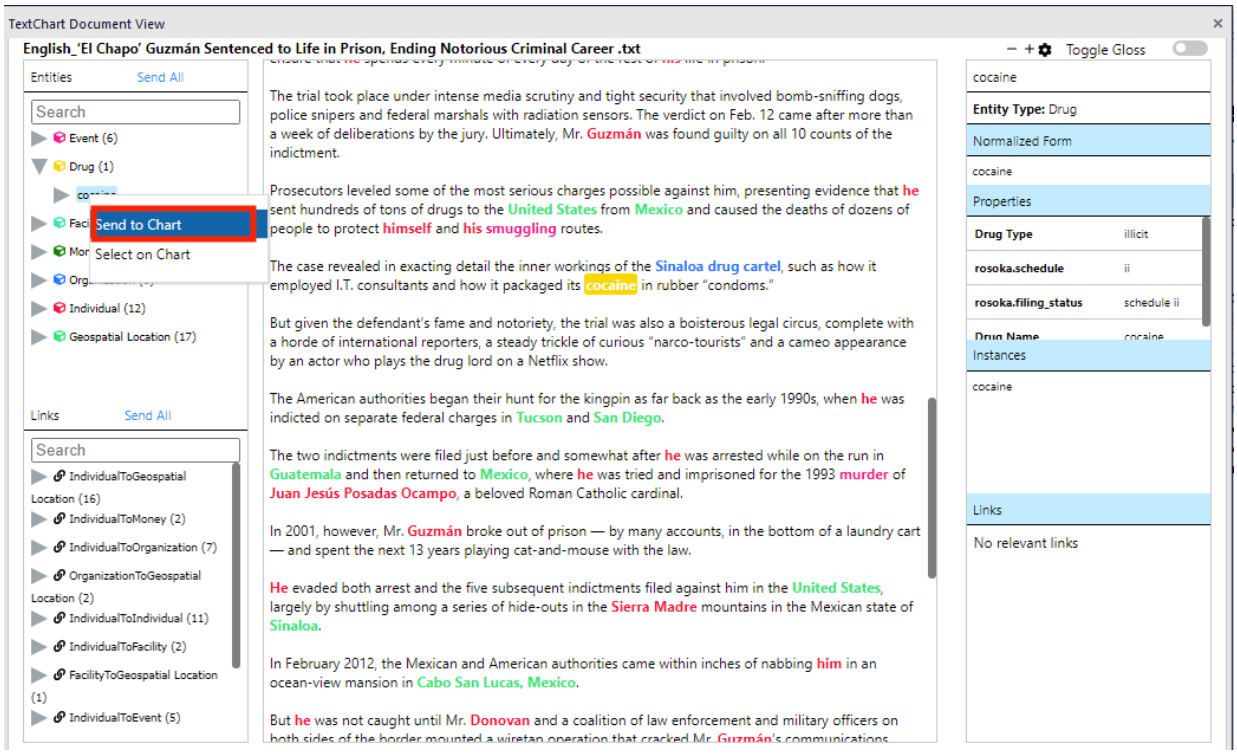
Sending entities to a chart

TextChart provides three options for sending entities to a chart. You can specify them individually, choose all entities of the same type, or simply send all the extracted entities at once.

Sending a single entity

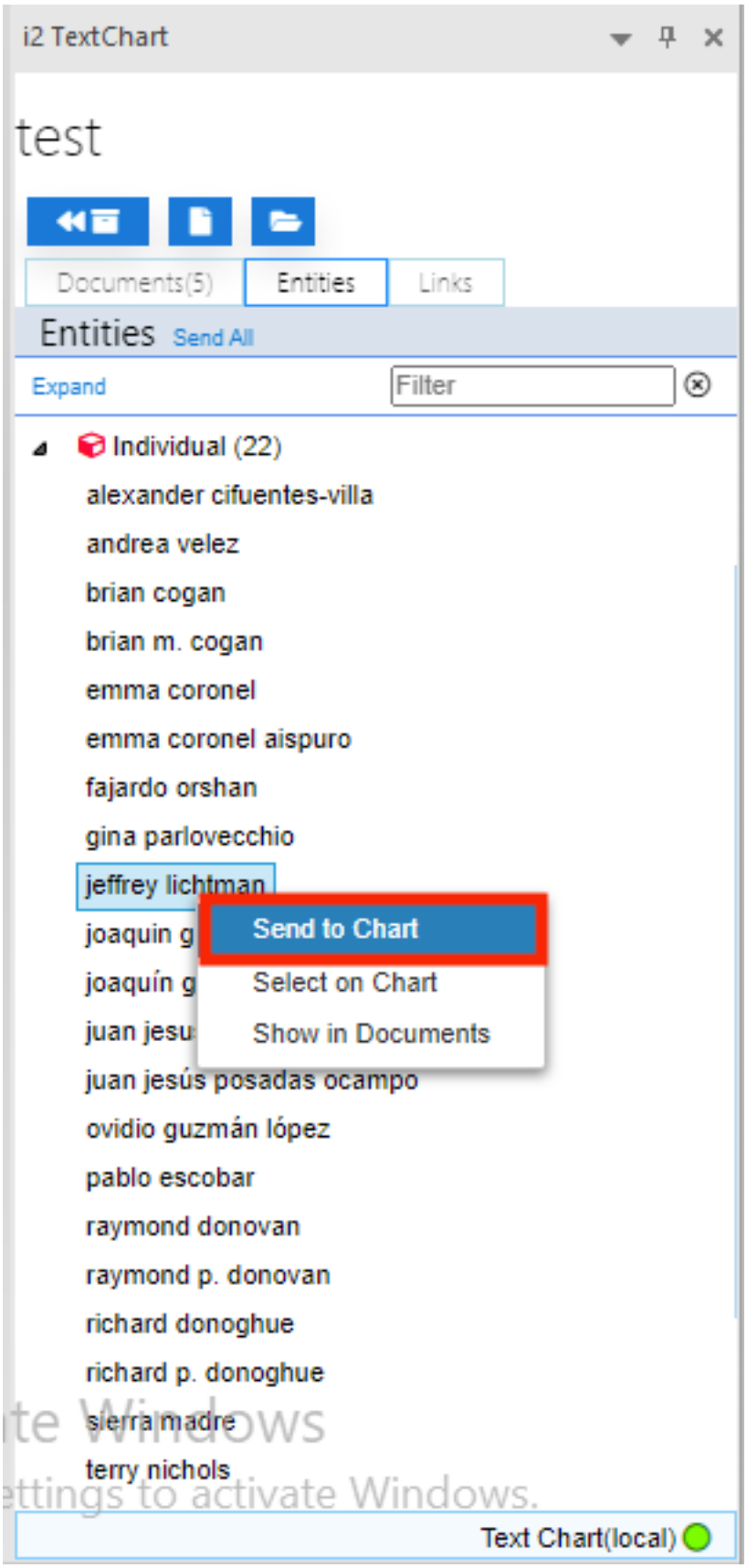
To send a single entity from the Document View to the current chart, you can use the pop-up menu for that entity in the **Entities** list on the left of the window, or the **Send To Chart** button from the entity details pane on the right of the window.

© N. Harris Computer Corporation



If an extracted entity is already on the chart, you can use the same pop-up menu to select it there.

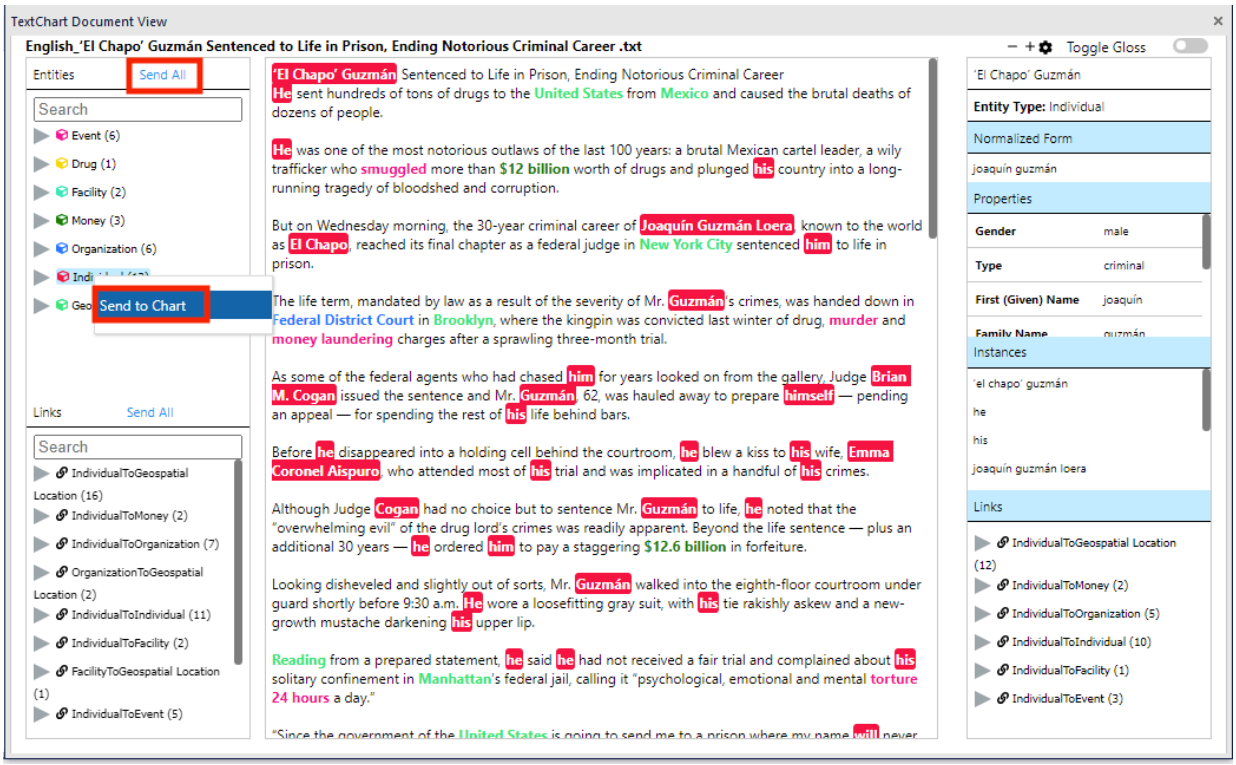
© N. Harris Computer Corporation



© N. Harris Computer Corporation

Sending several entities

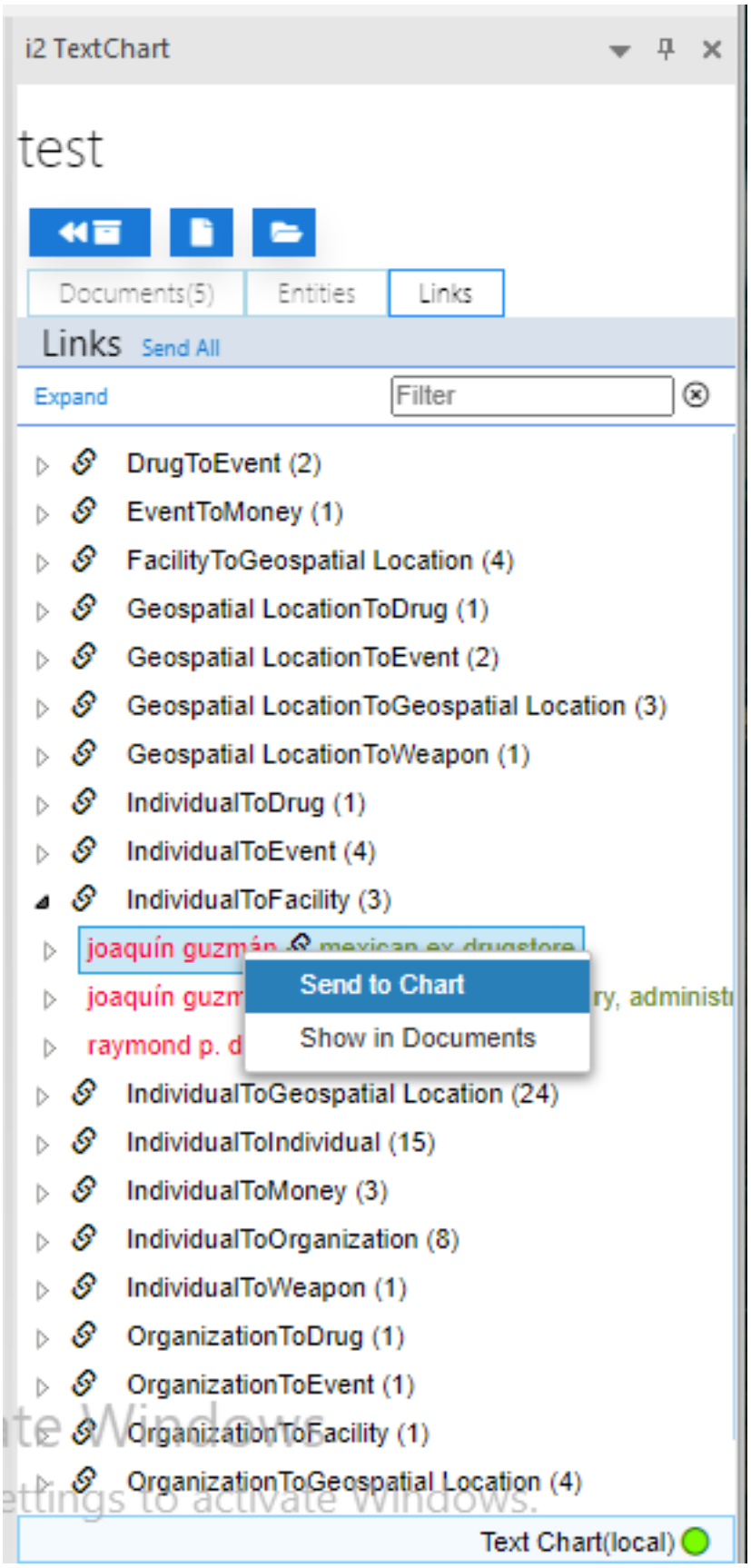
To send all the entities of the same type from the Document View to a chart, you can use the pop-up menu for that entity type in the **Entities** list. Alternatively, to send *all* the entities extracted from the document to the chart, you can click the **Send All** button at the top of the list.



Sending links to a chart

TextChart provides the same three options for sending extracted links to the chart as it does for sending entities, and you use them in the same way.

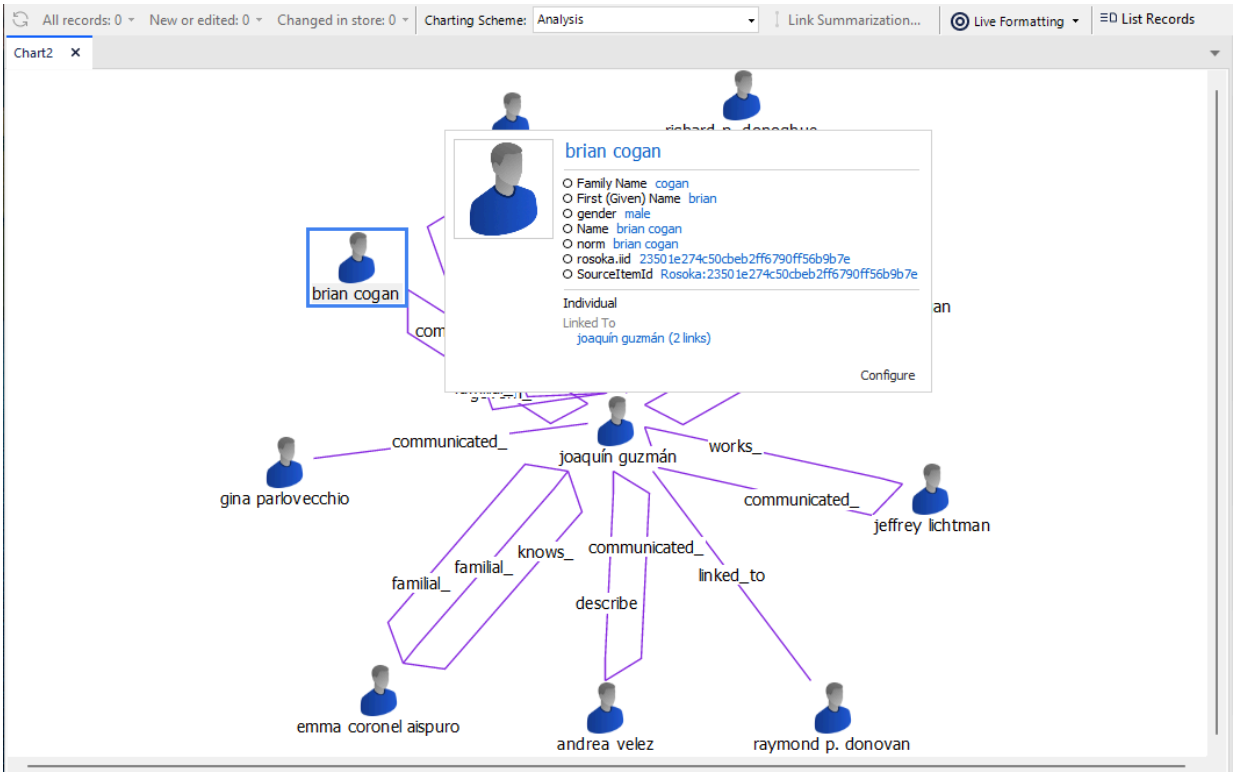
© N. Harris Computer Corporation



© N. Harris Computer Corporation

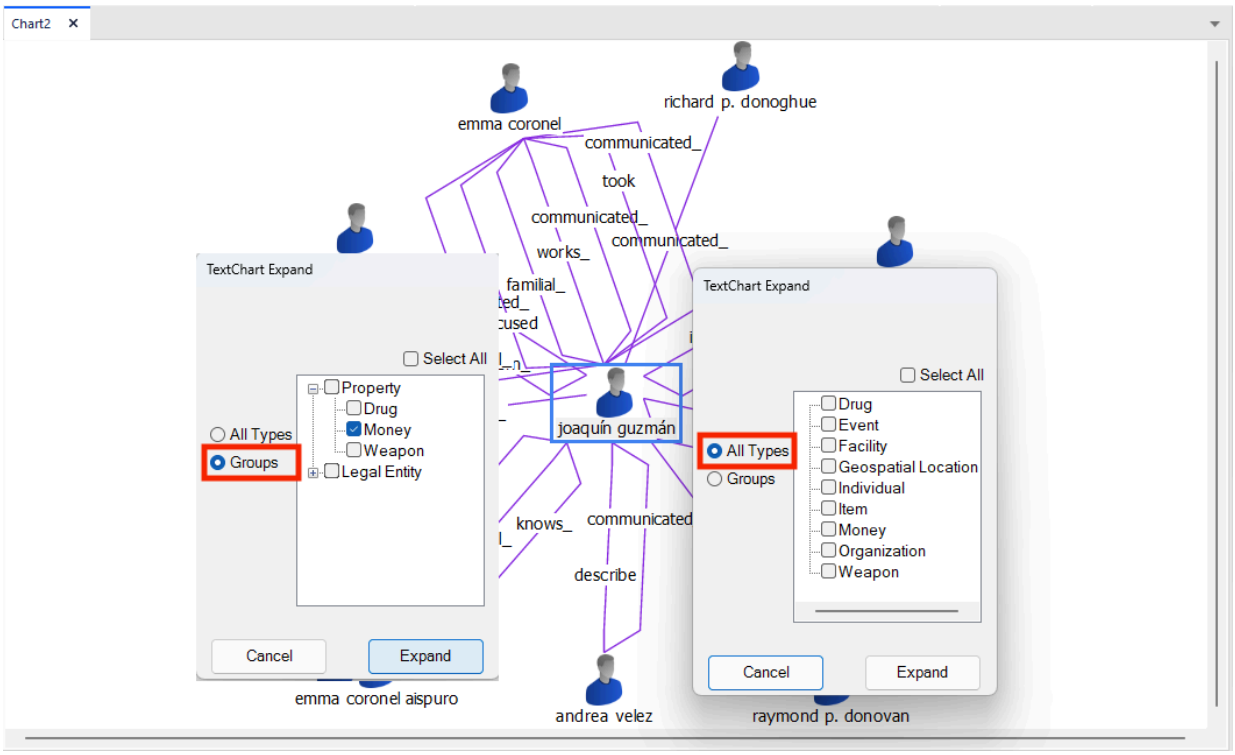
Viewing entities on the chart surface

You can hover over *any* entity on the Analyst's Notebook chart surface to display a window containing information about all the properties that are associated with that entity.



After you've added an entity *from TextChart* to the chart, you can use it to start adding related entities from the other documents that you've processed. Right-click the entity, and select **Rosoka Expand**.

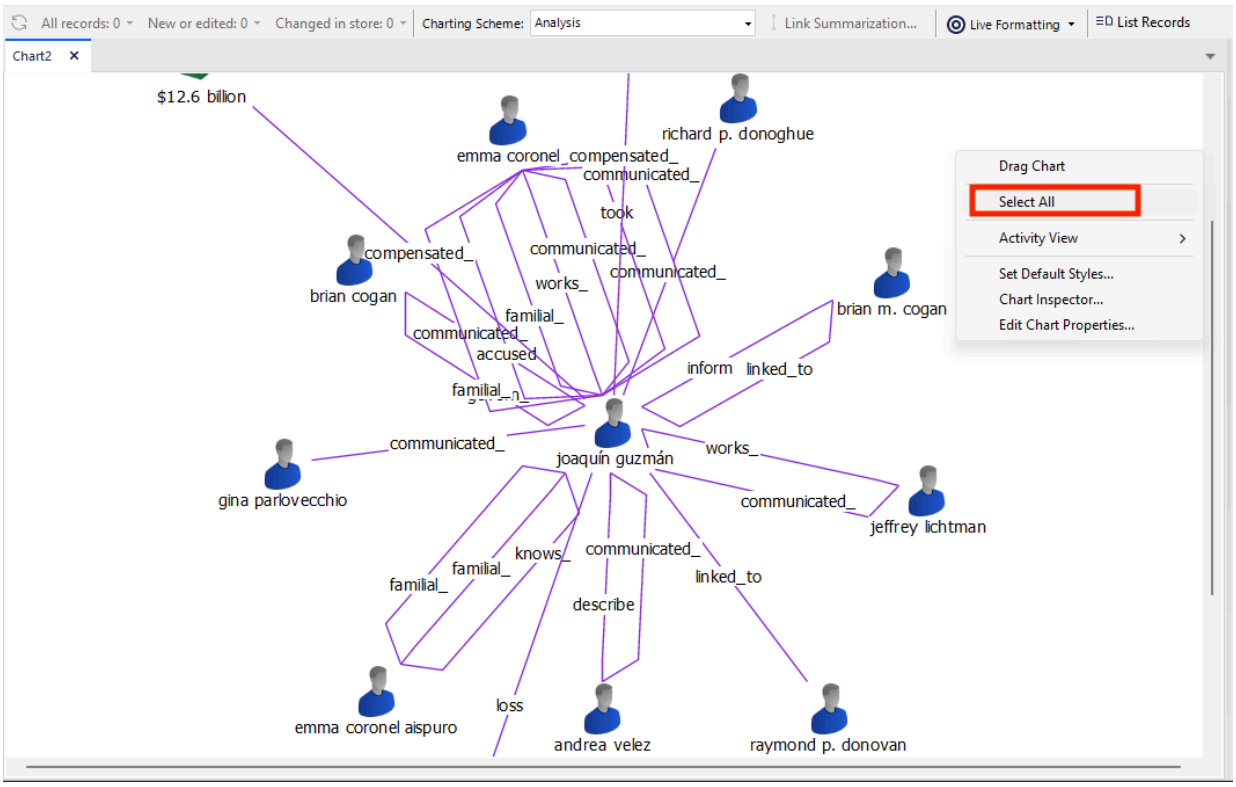
© N. Harris Computer Corporation



You can choose to add related entities of particular types, or groups of types; or you can use **Select All** to add all related entities in a single action.

When a chart contains multiple TextChart entities, you can perform the same **Rosoka Expand** operation on all of them by right-clicking the chart surface and choosing **Select All** first.

© N. Harris Computer Corporation

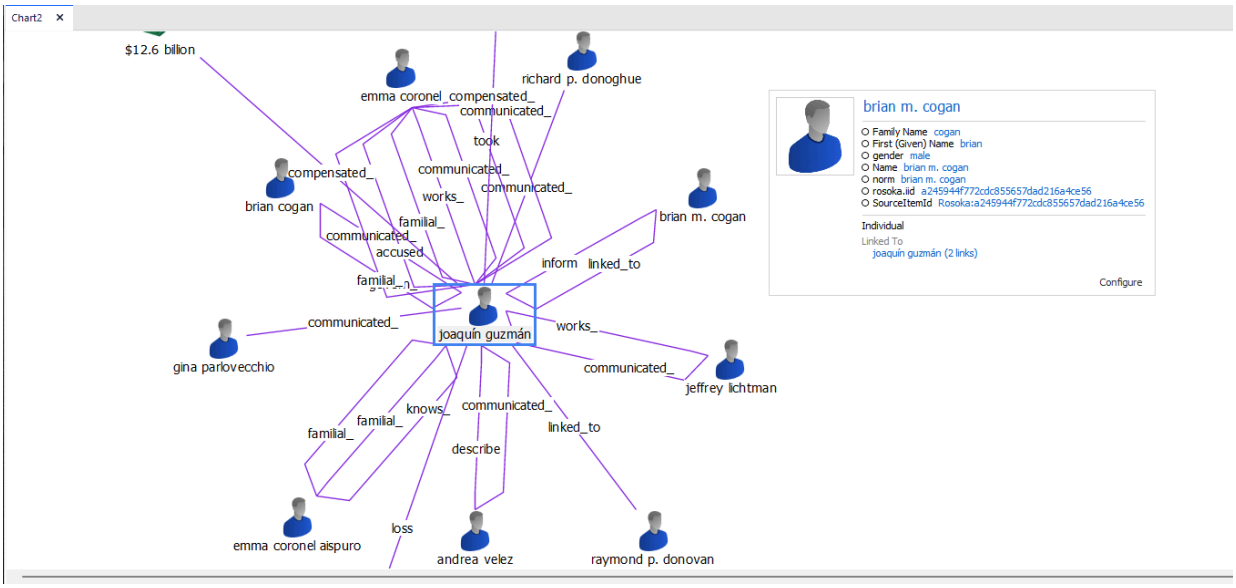


Viewing links on the chart surface

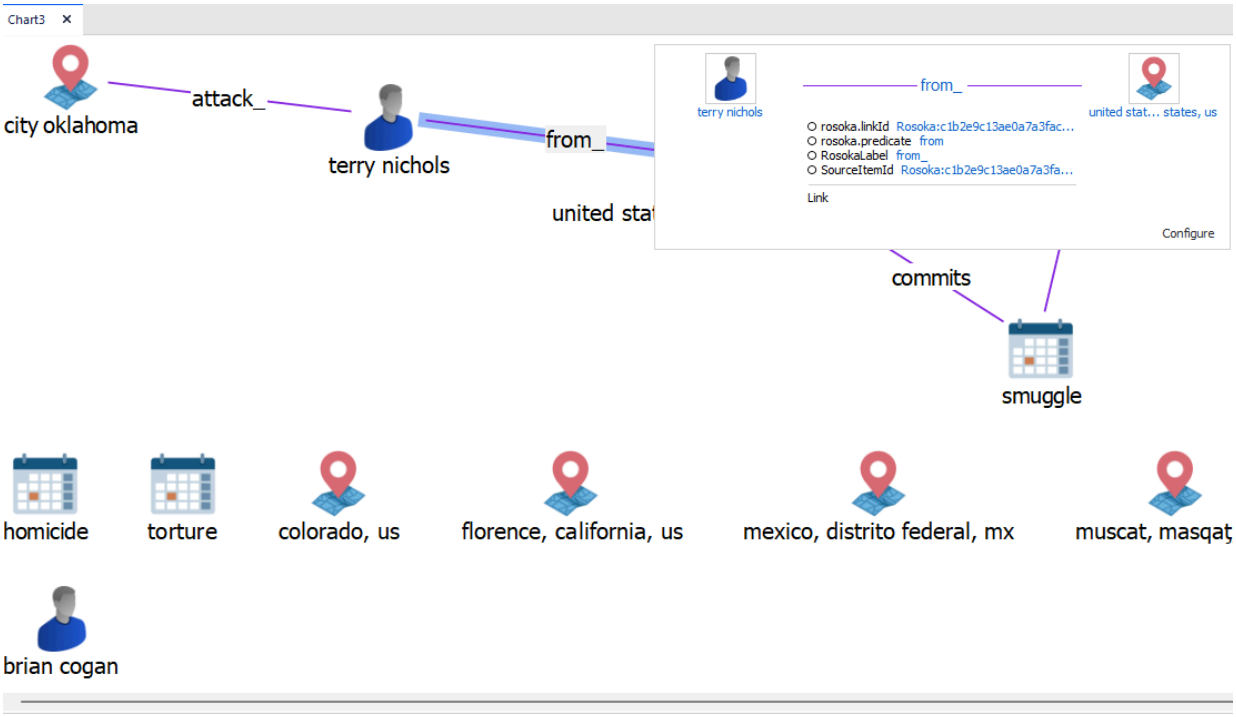
After you've added all the entities to the chart that you want to add, you can start to investigate the relationships between them, including the predicates that were responsible for the links being identified in the first place.

For example, hovering over a connected entity reveals a list of the entities that it is linked to, in addition to its properties.

© N. Harris Computer Corporation



Similarly, you can hover over a TextChart link to highlight the connection between two entities and display additional extraction information.

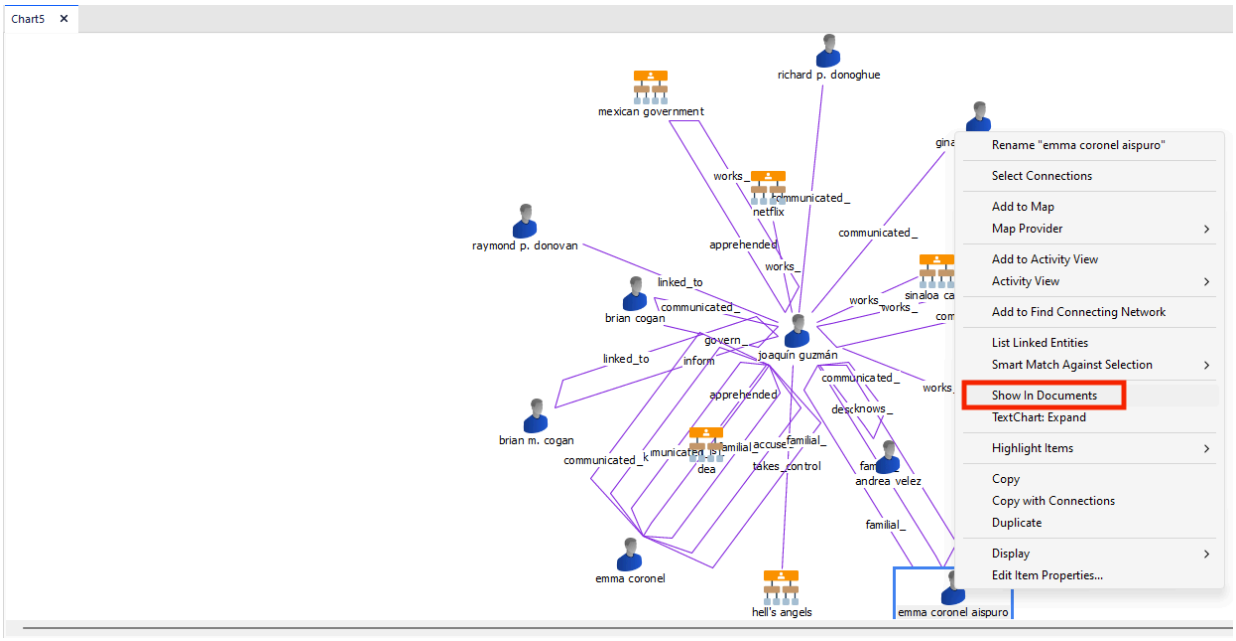


© N. Harris Computer Corporation

Listing source documents

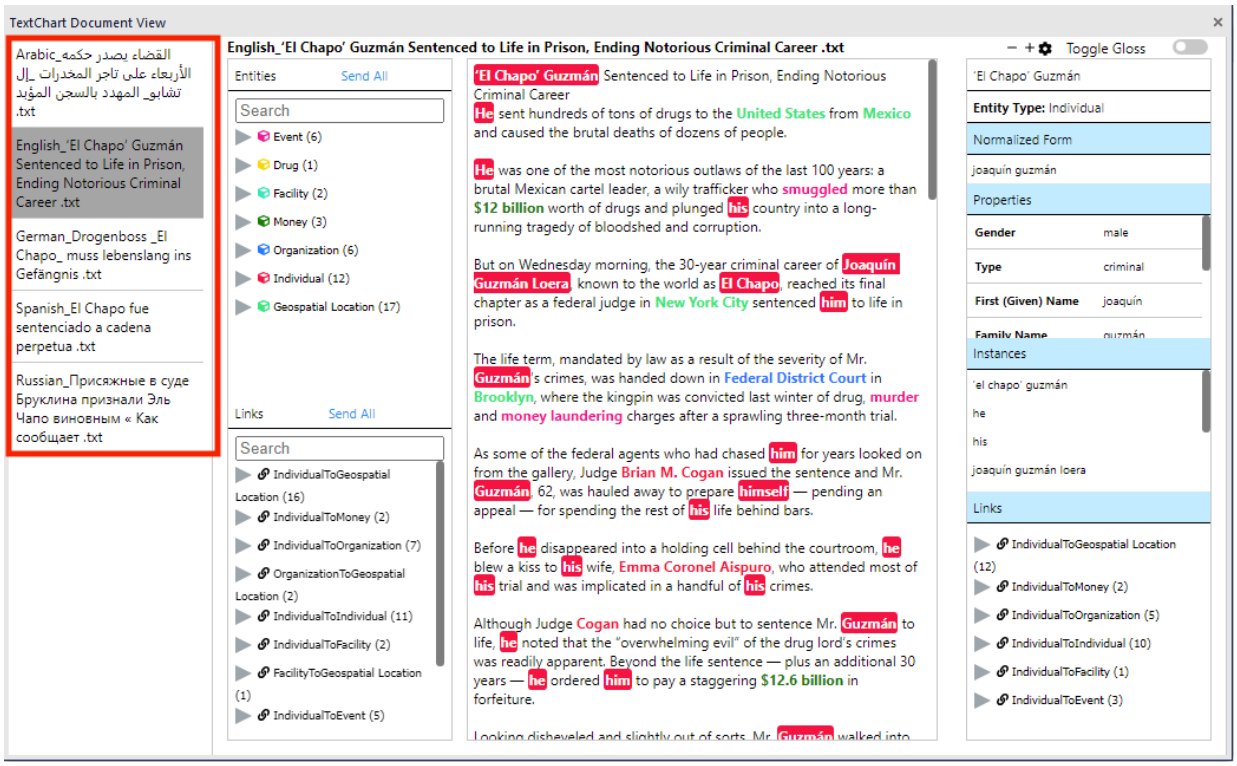
For any entity or link that was added to the Analyst's Notebook chart surface through TextChart, you can quickly refer back to the documents from which it was extracted.

To retrieve all the documents for an entity or a link, right-click and select **Show In Documents** to reopen the Document View.



On the left of the view, the list of documents allows you to select which document appears on the right. Each document appears with text highlighting that reveals the entity or link in question.

© N. Harris Computer Corporation



Mapping to i2 Analyze types

When Analyst's Notebook is connected to an i2 Analyze server, the types of the entities and links that users can add to a chart (and the types of the properties that they can contain) are governed by a *schema*. The schema has an associated *charting scheme* that controls the appearance of entities and links.

Individual deployments of i2 Analyze have custom schemas that define entity and link types that are appropriate to the analysis they perform. Different organizations will use different types: even if two organizations used entity types named Person, those types would have different definitions.

TextChart, on the other hand, has a fixed list of types that it can extract from documents, which is defined in the LxBase. In order for Analyst's Notebook to receive entities and links from TextChart, and to display and analyze them successfully, there must be a mapping from the TextChart types to an i2 Analyze schema.

Default type mapping

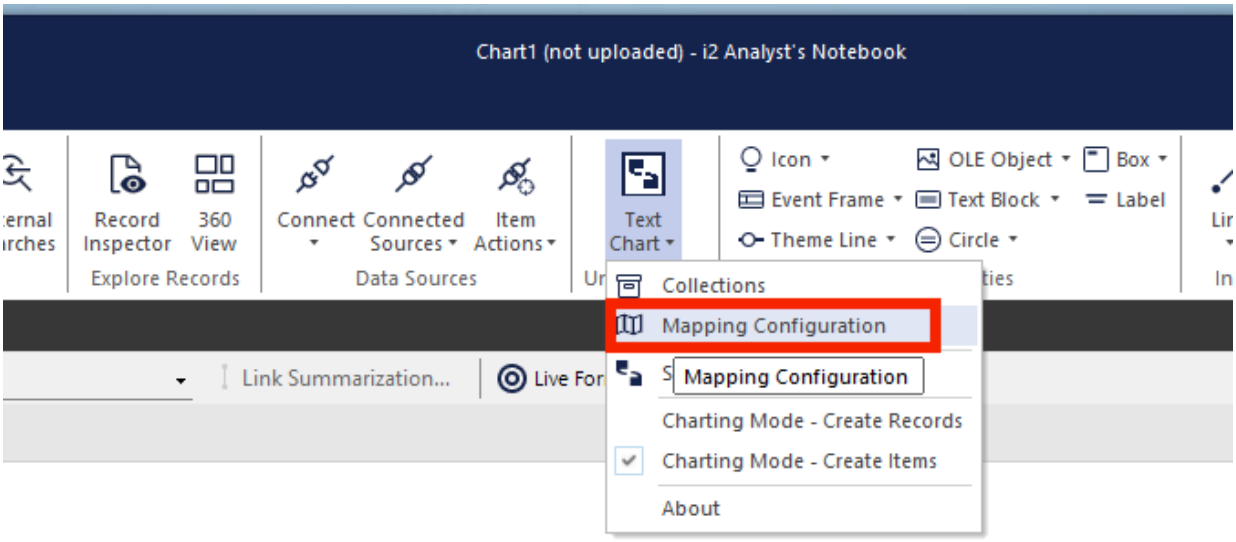
When you install i2 TextChart, it includes an i2 Analyze schema that defines all the same types that the LxBase defines, and a *mapping configuration* that provides a one-to-one mapping between the two sets of definitions. By default, TextChart uses that schema and mapping configuration when you send extracted entities and links to the chart.

Through Analyst's Notebook, you can make changes to the default mapping configuration so that some types are extracted differently - or so that they're not extracted at all. You can also create

© N. Harris Computer Corporation

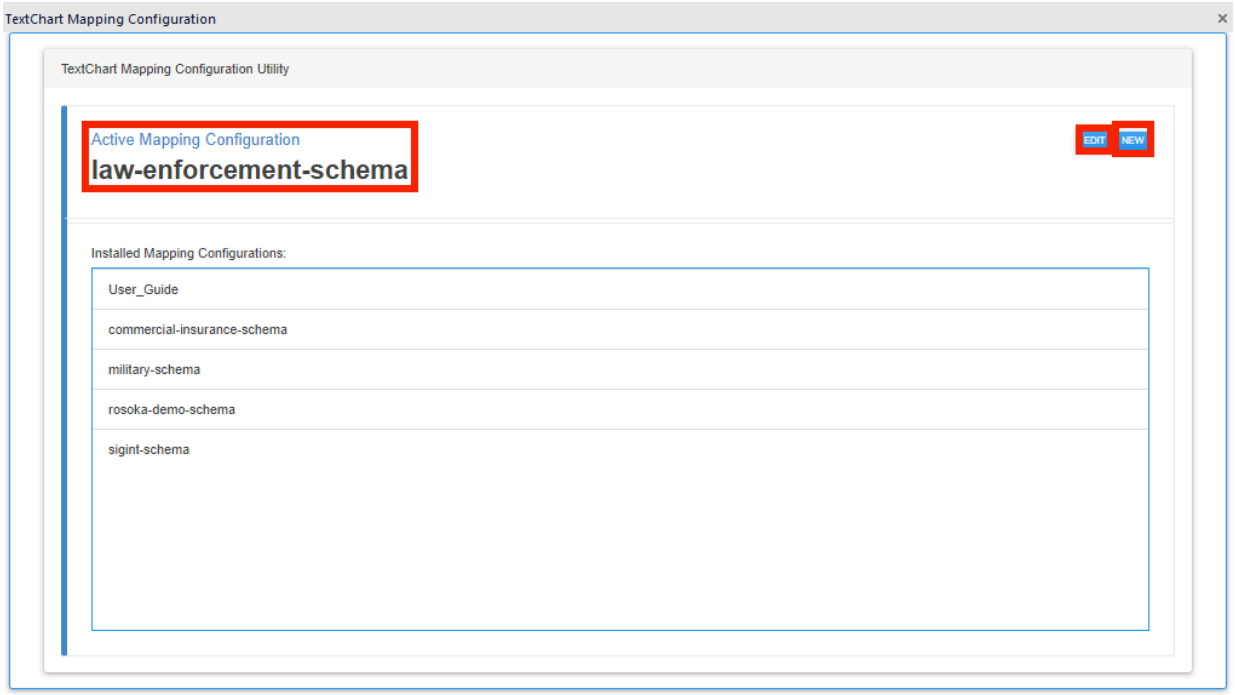
mapping configurations for different i2 Analyze schemas, so that the entities and links you extract from documents are aligned with those that you're already using.

To start the process, select **Text Analytics > Mapping Configuration**.



The command opens the Mapping Configuration window, where you can make changes to the active configuration, switch to a different configuration, or create a mapping configuration of your own.

© N. Harris Computer Corporation

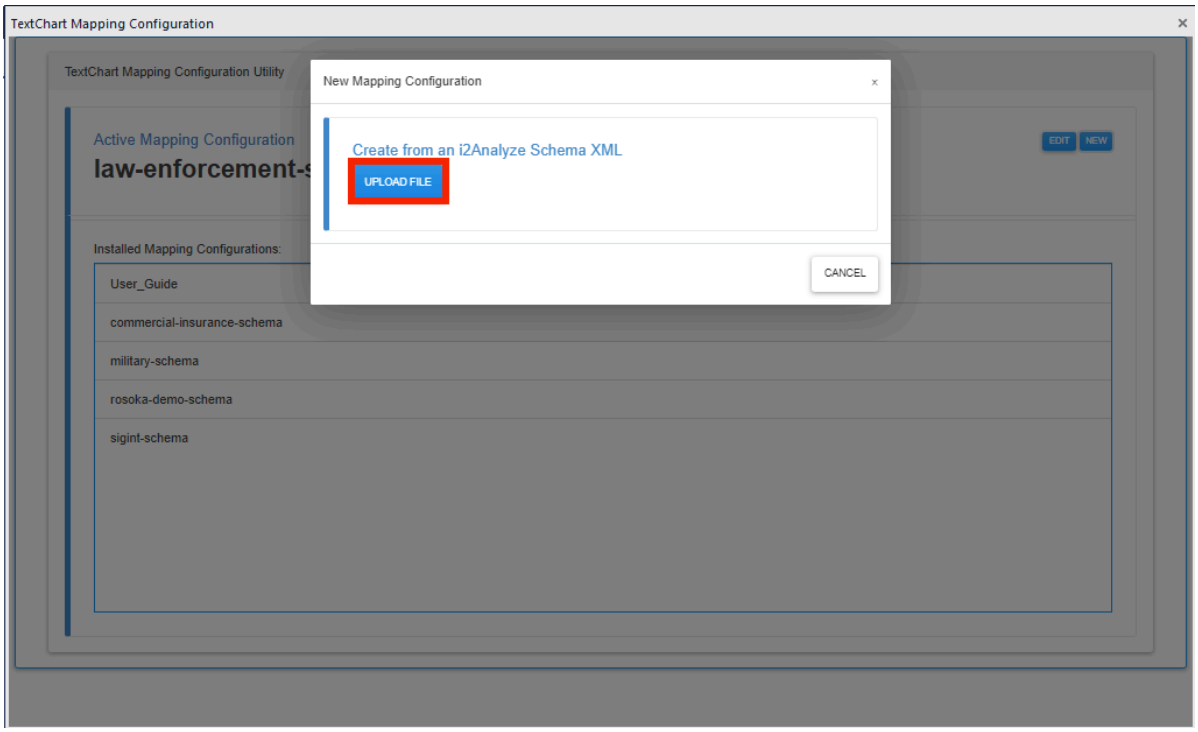


Mapping to a different schema

To start the process of mapping TextChart types to a different i2 Analyze schema, you must tell TextChart about the schema that you want to map to!

1. Click **New** to open the New Mapping Configuration window.

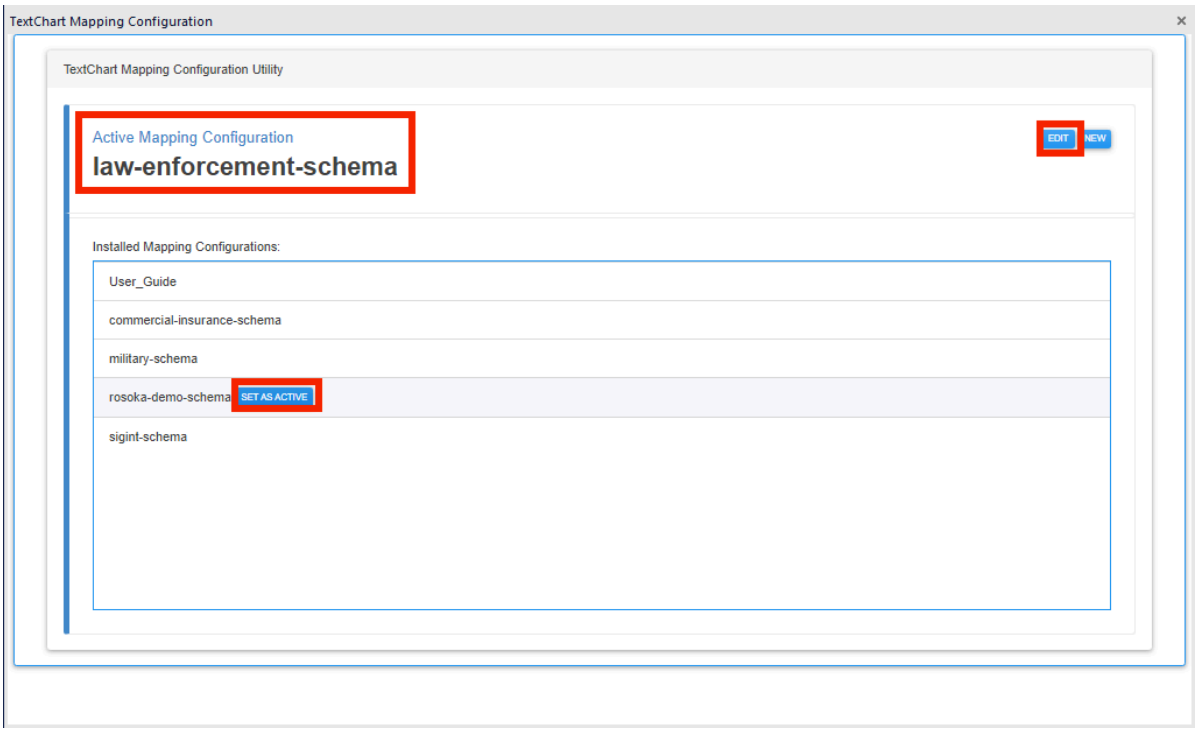
© N. Harris Computer Corporation



2. Click **UPLOAD FILE**, and locate the XML file that contains the i2 Analyze schema you want to use.

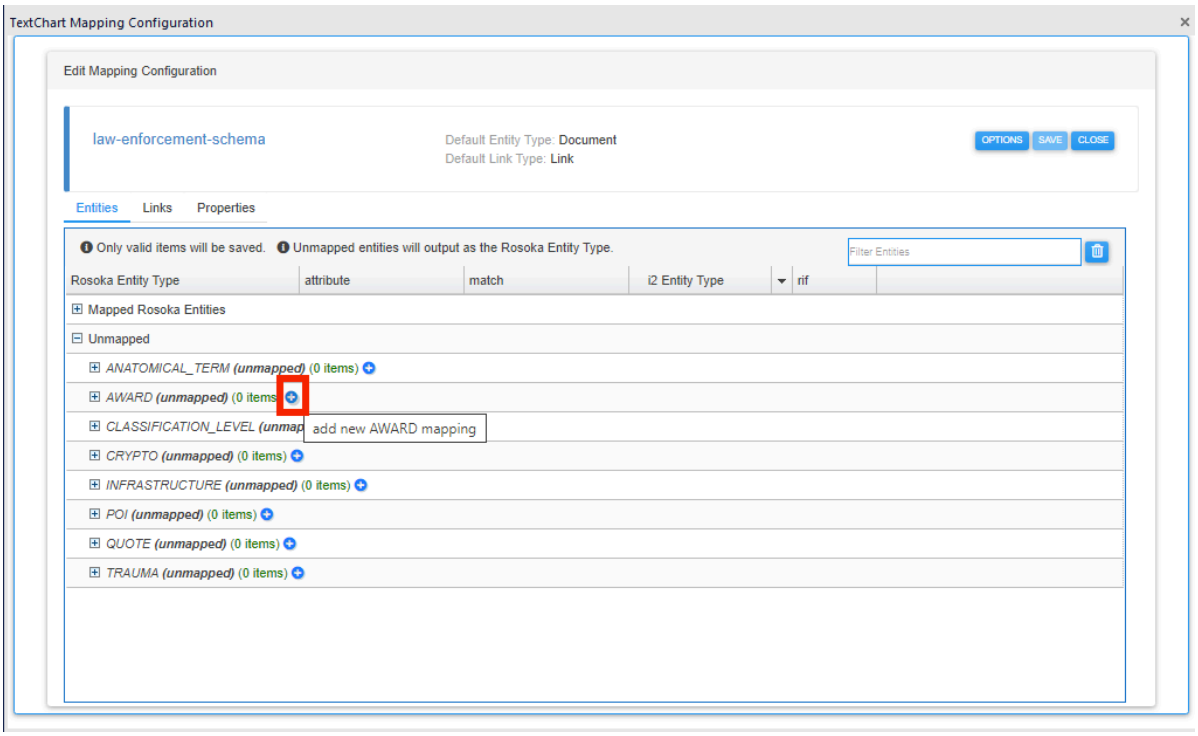
   When the upload succeeds, click **OK** to close the dialog. TextChart creates an empty mapping configuration for the schema and makes it active.

© N. Harris Computer Corporation



3. Click **Edit** to open the new mapping configuration. At this stage, all of the TextChart types are "unmapped".

© N. Harris Computer Corporation



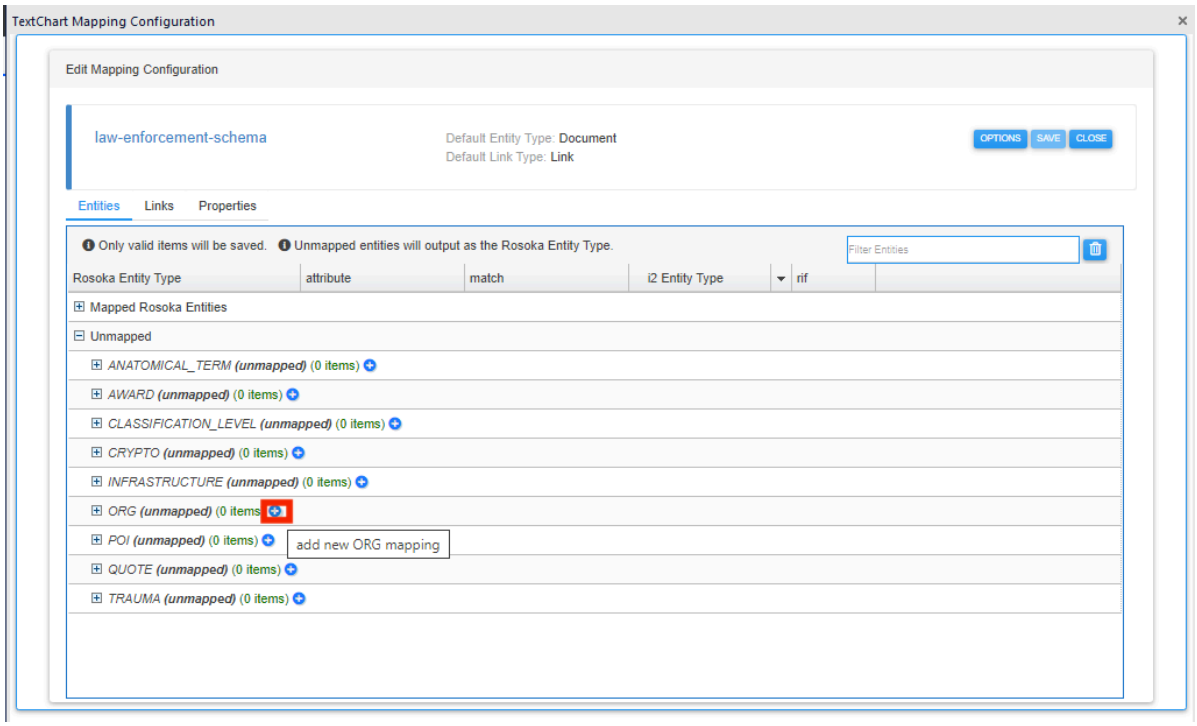
Note: When a configuration contains no mapping for a particular TextChart entity or link type, TextChart falls back to using the default mapping (and the default schema) for that type.

Mapping entity types

At this point, you can start to create the mappings from TextChart types to i2 Analyze schema types. It's usually easier to start with the entity types.

1. Click the "plus" icon next to the name of the TextChart type that you want to map. The type moves from the "unmapped" list to the "mapped" list:

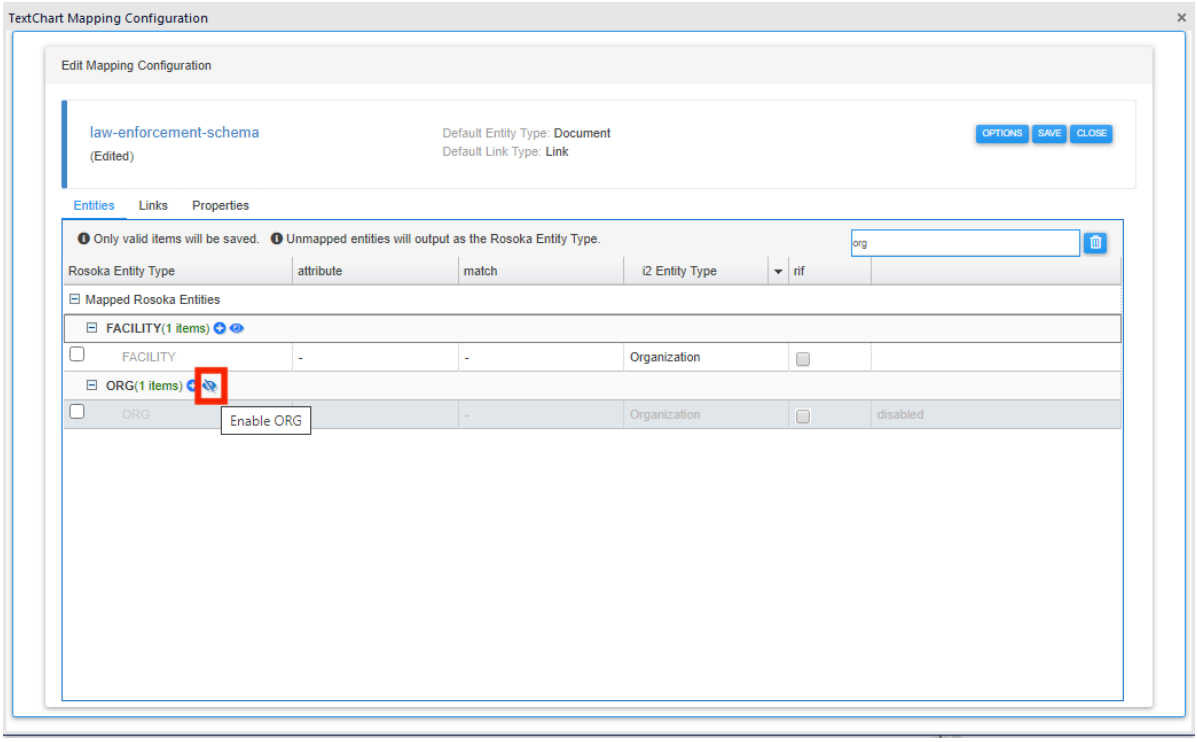
© N. Harris Computer Corporation



When you first move a type to the "mapped" list like this, the mapping - and, effectively, extraction - is disabled. Entities with this type do not appear in TextChart results.

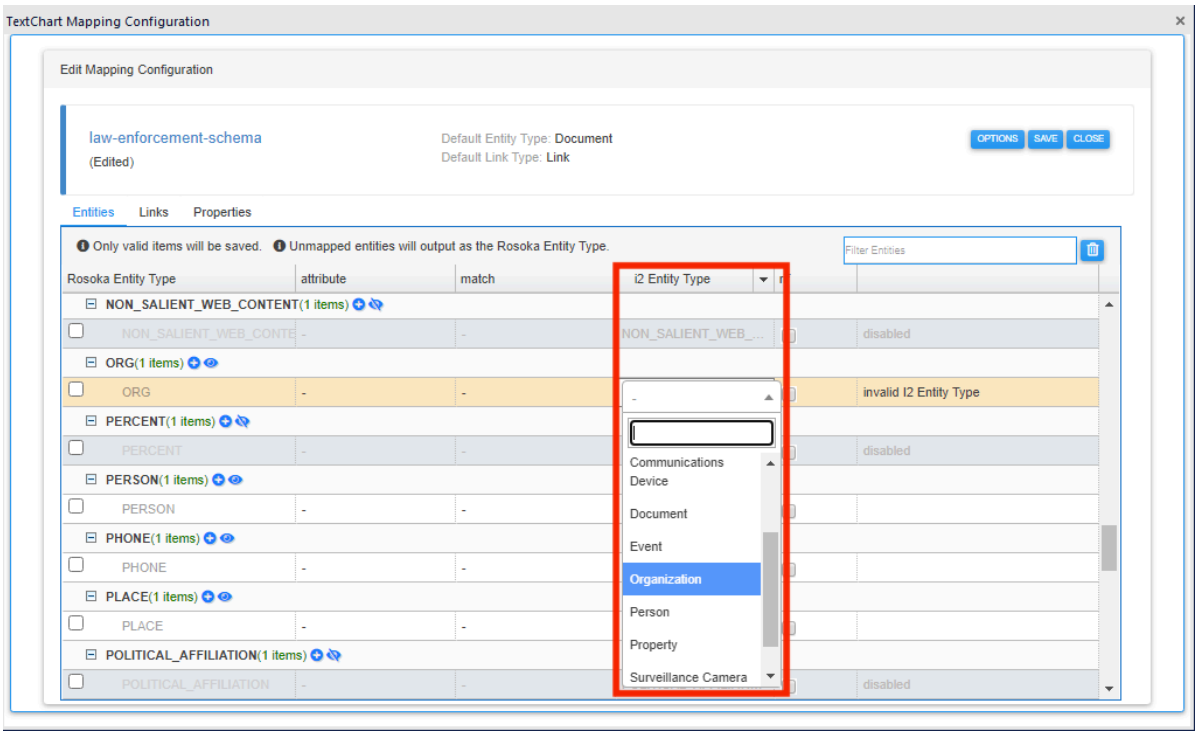
2. To enable the mapping, click the crossed-out "eyeball" icon next to the type name.

© N. Harris Computer Corporation



3. Now, you can choose the entity type in the i2 Analyze schema that you want the TextChart type to map to. Open the drop-down list in the **i2 Entity Type** column.

© N. Harris Computer Corporation



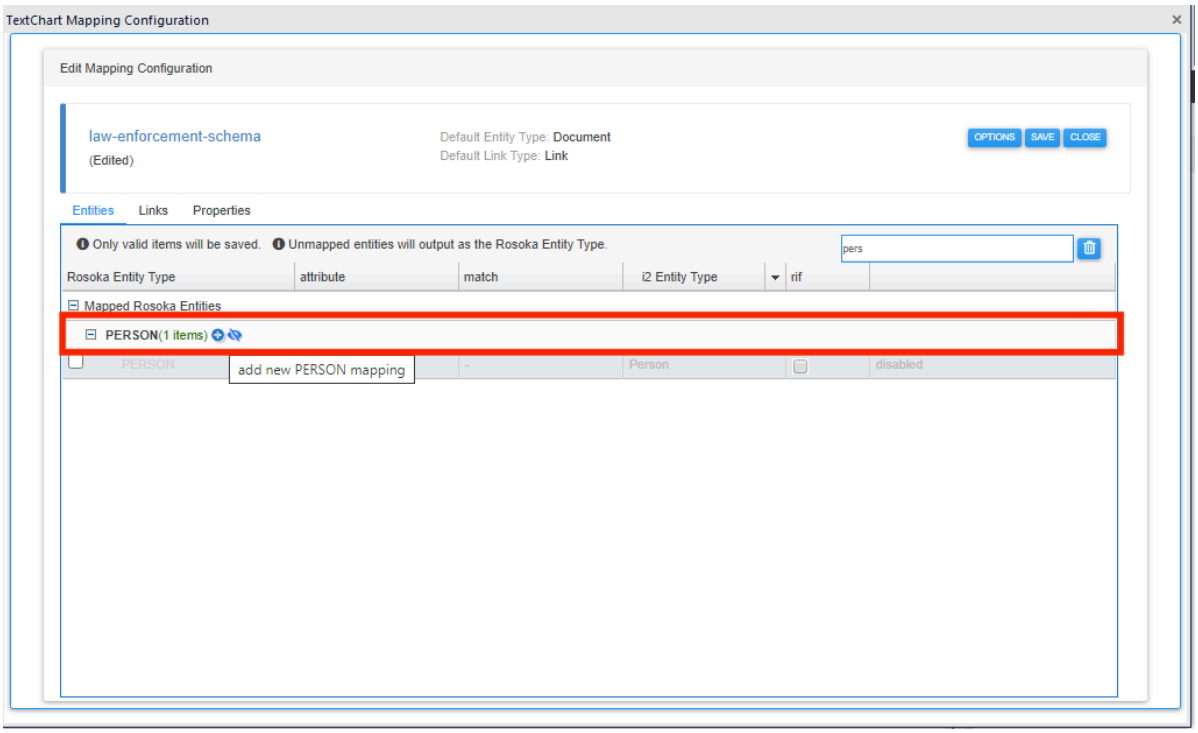
4. Select the i2 entity type that you want to use, and then click **Save** to update the mapping configuration.

Mapping entity types based on attribute values

Sometimes, you might want to map a TextChart type to different i2 Analyze types depending on the value of a particular attribute. For example, if the i2 Analyze schema has separate Male and Female entity types, you might map TextChart Person entities whose Gender attribute is "Male" to the former, and those whose Gender attribute is "Female" to the latter.

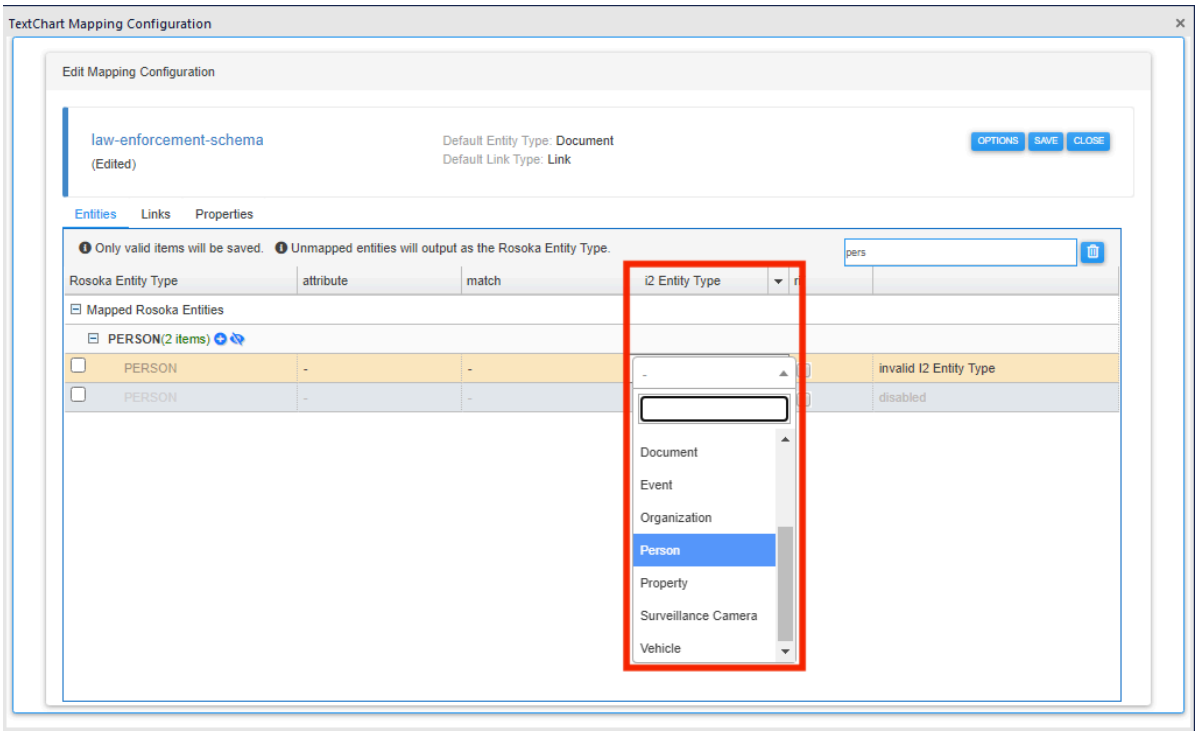
1. As before, click the "plus" icon next to the name of the TextChart entity type that you want to move to the "mapped" list.
2. Click the "plus" icon again to add a new row to the mapping configuration for the same TextChart entity type.

© N. Harris Computer Corporation



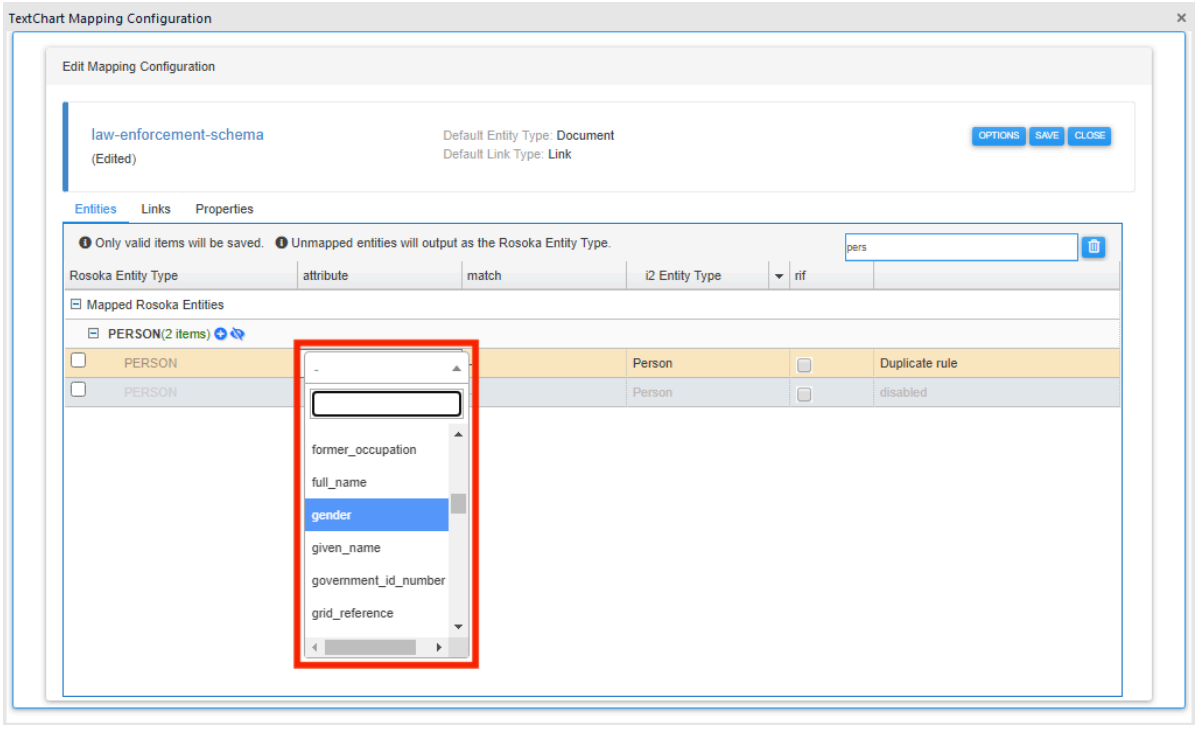
3. In the **i2 Entity Type** column, select one of the types that the TextChart entity type will be mapped to.

© N. Harris Computer Corporation



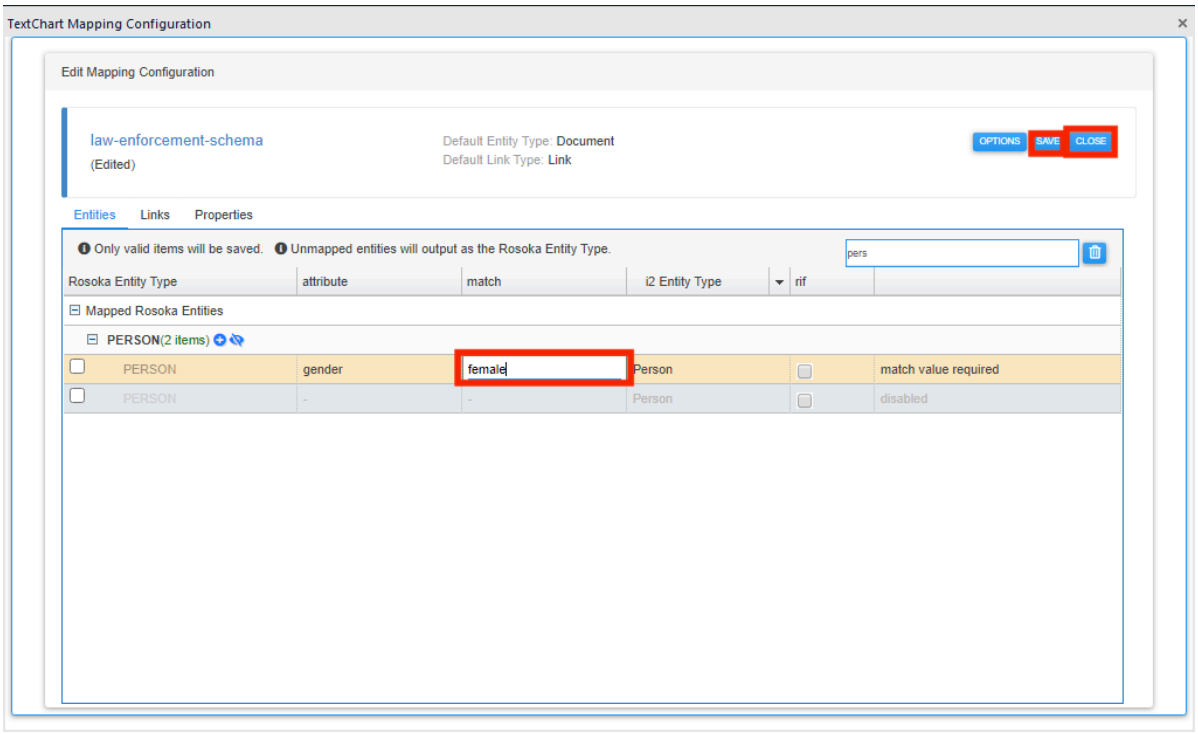
4. In the **attribute** column, select the TextChart attribute whose value will control whether this mapping is used.

© N. Harris Computer Corporation



5. In the **match** column, specify the attribute value that will cause this mapping to be used.

© N. Harris Computer Corporation



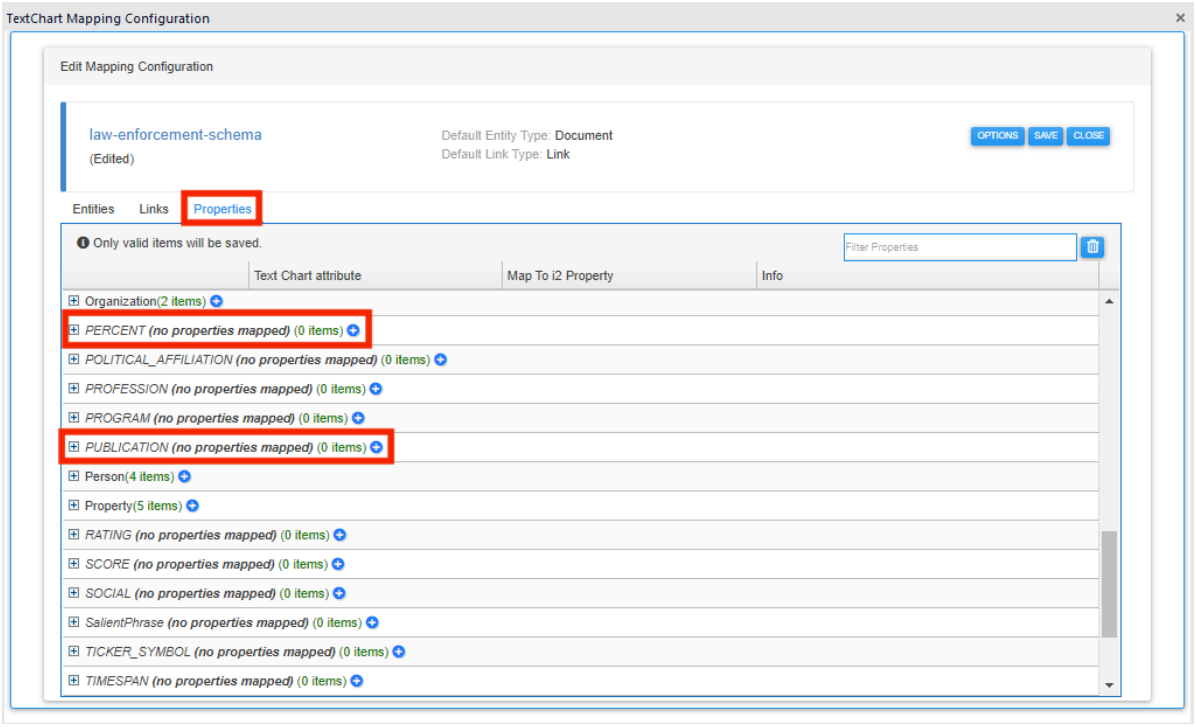
When you save this mapping configuration, the entities that you send to charts from TextChart extraction results will have different types, according to the criteria that you defined.

Mapping TextChart attributes to i2 Analyze property types

After you set up mappings between TextChart and i2 Analyze entity types, the next step is to configure the mappings between their attributes and property types.

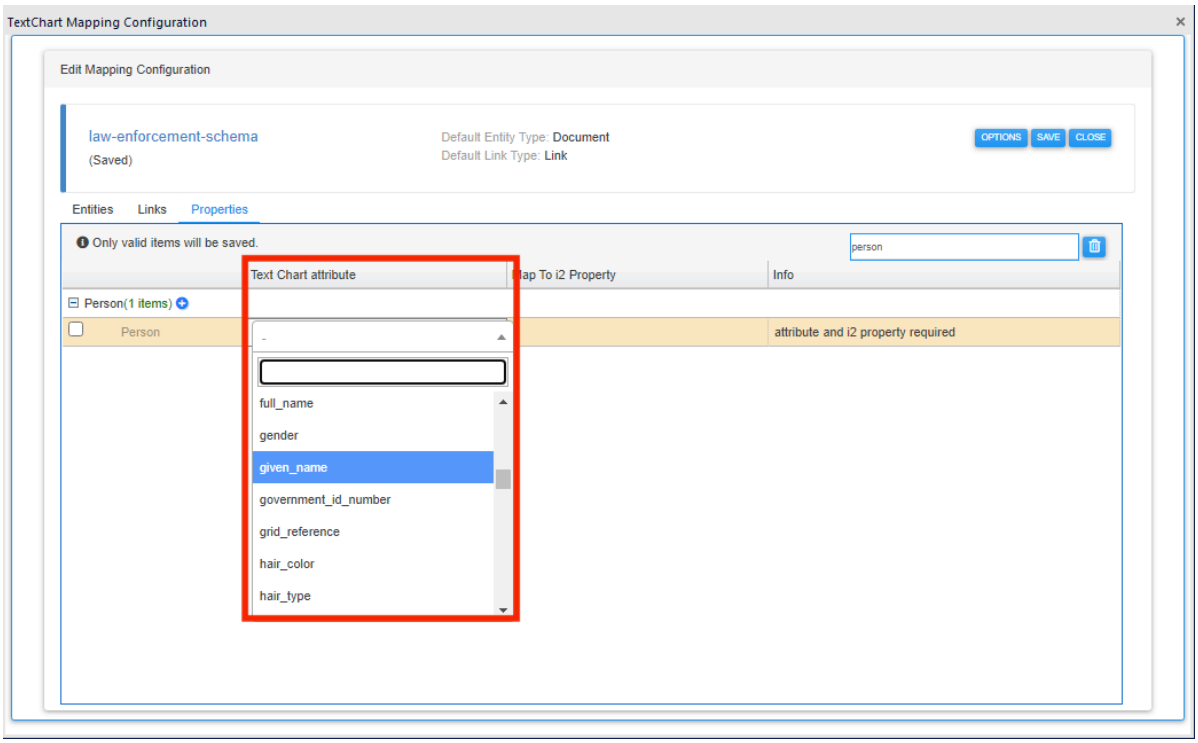
1. Select the **Properties** tab in the Mapping Configuration window to display a list of all the entity types that you've mapped.

© N. Harris Computer Corporation



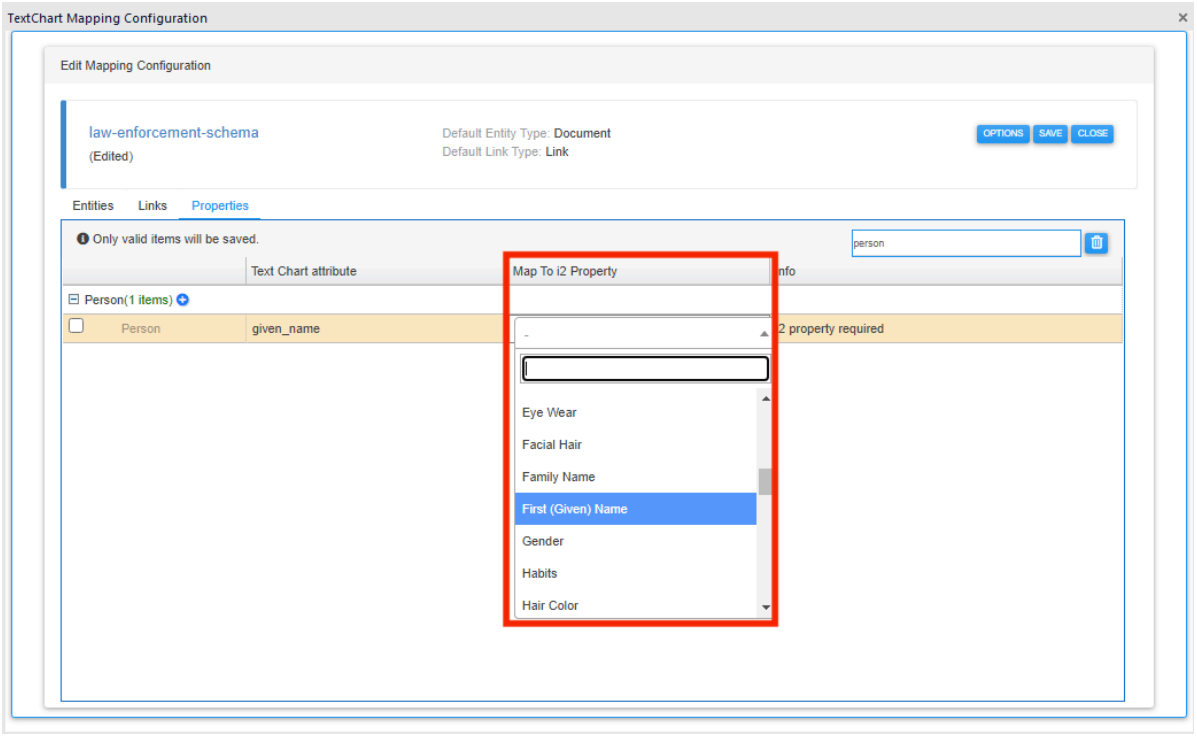
2. Click the "plus" icon next to the entity type to which you want to add a property type mapping. A new row appears below the entity type.
3. In the **TextChart attribute** column, select the attribute of the TextChart type for which you want to define a mapping.

© N. Harris Computer Corporation



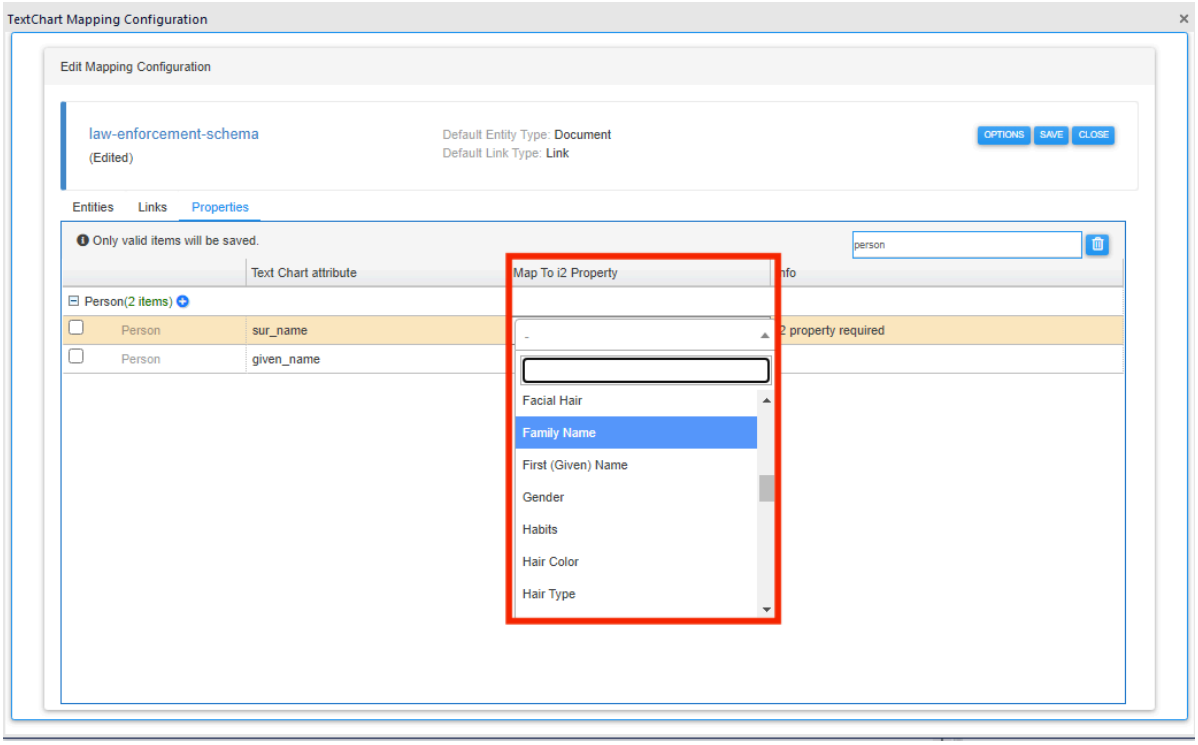
4. In the **Map To i2 Property** column, select the property type of the i2 entity type that you want to map the attribute to.

© N. Harris Computer Corporation



5. Continue this process until you have created mappings for all the property types that you want to see in Analyst's Notebook charts.

© N. Harris Computer Corporation



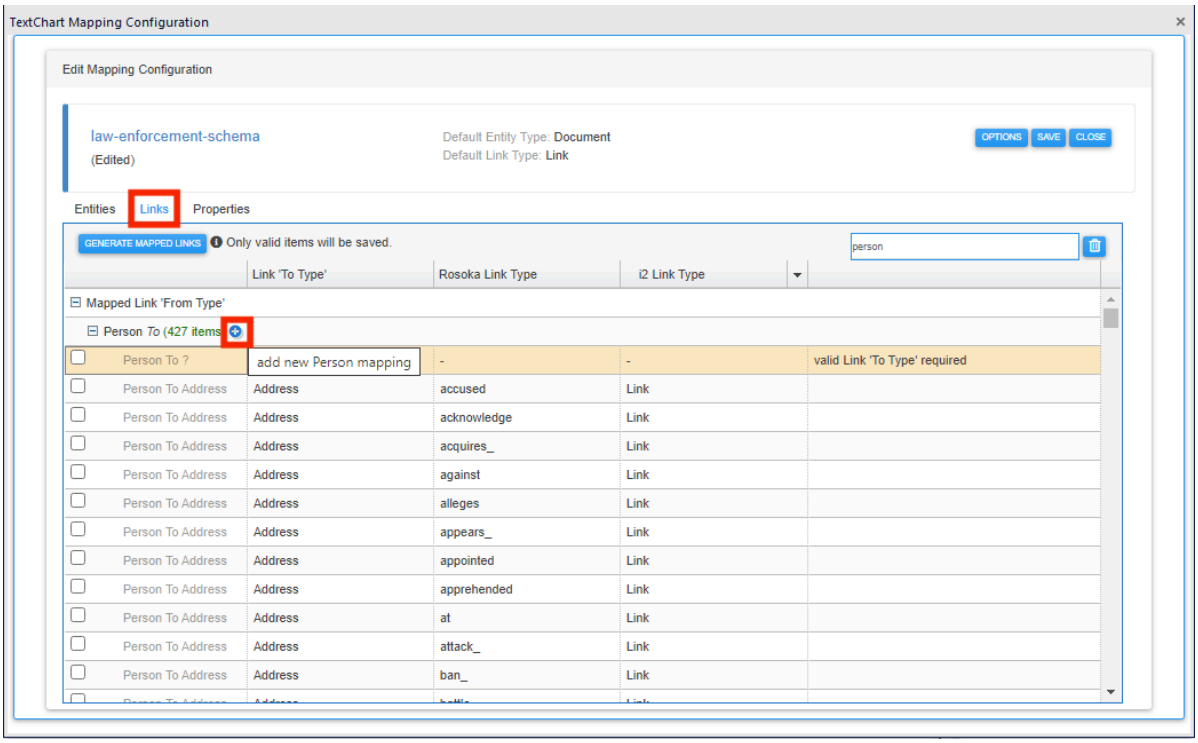
At any point in the mapping process (and always at the end), click **Save** to update the stored version of the mapping configuration.

Mapping link types

You can only create mappings from TextChart link types to i2 Analyze link types after you've created the necessary entity type mappings.

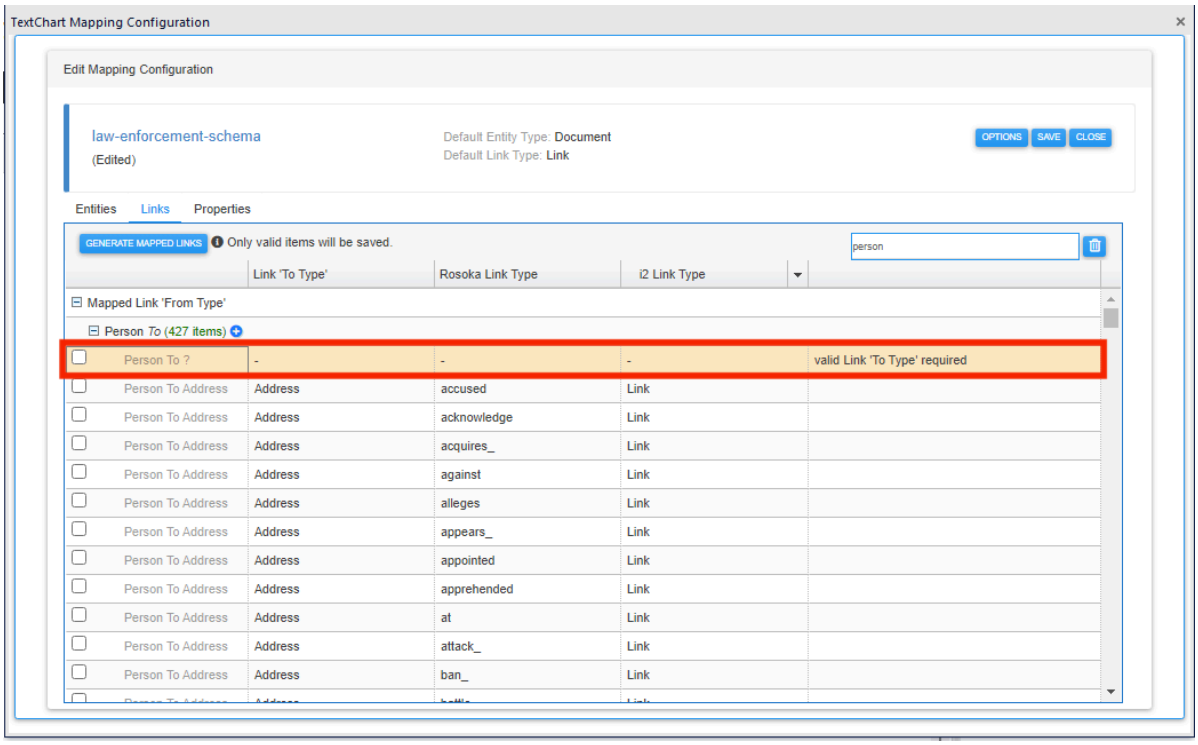
1. Select the **Links** tab in the Mapping Configuration window to display a list of all the entity types that you've mapped.

© N. Harris Computer Corporation



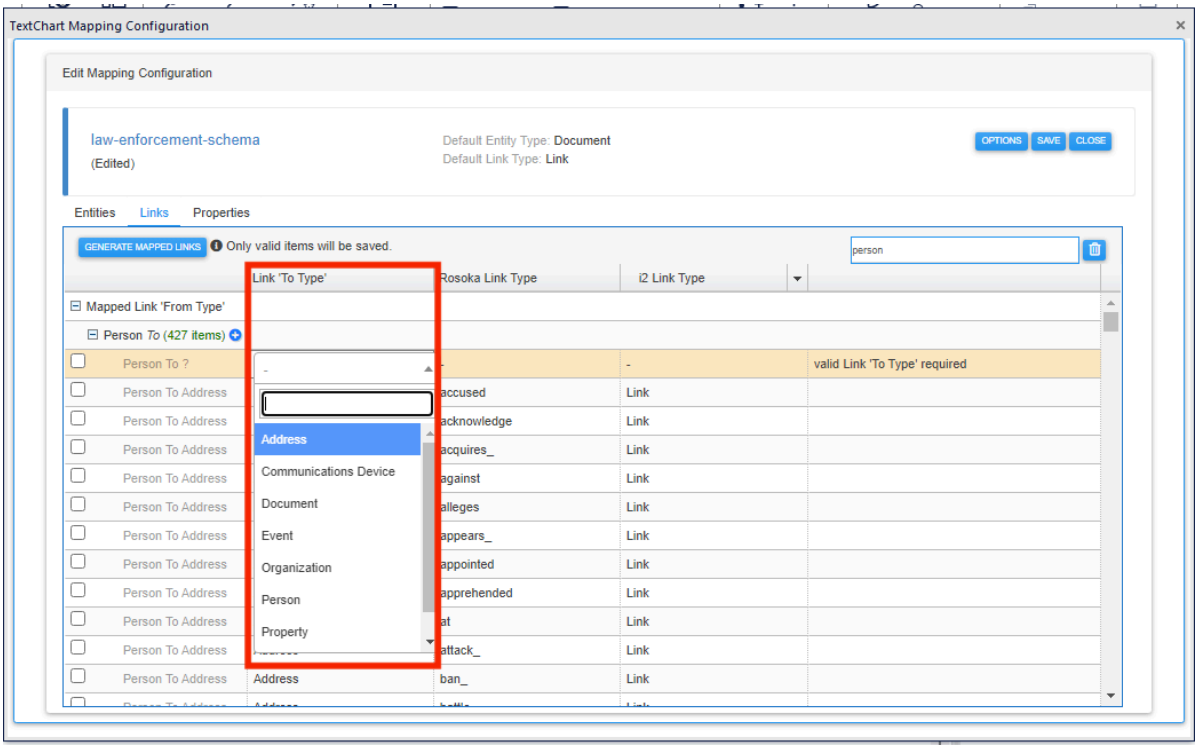
2. Click the "plus" icon next to the name of the entity type that forms the "From" end of the link type that you want to map. A new row appears below the entity type.

© N. Harris Computer Corporation



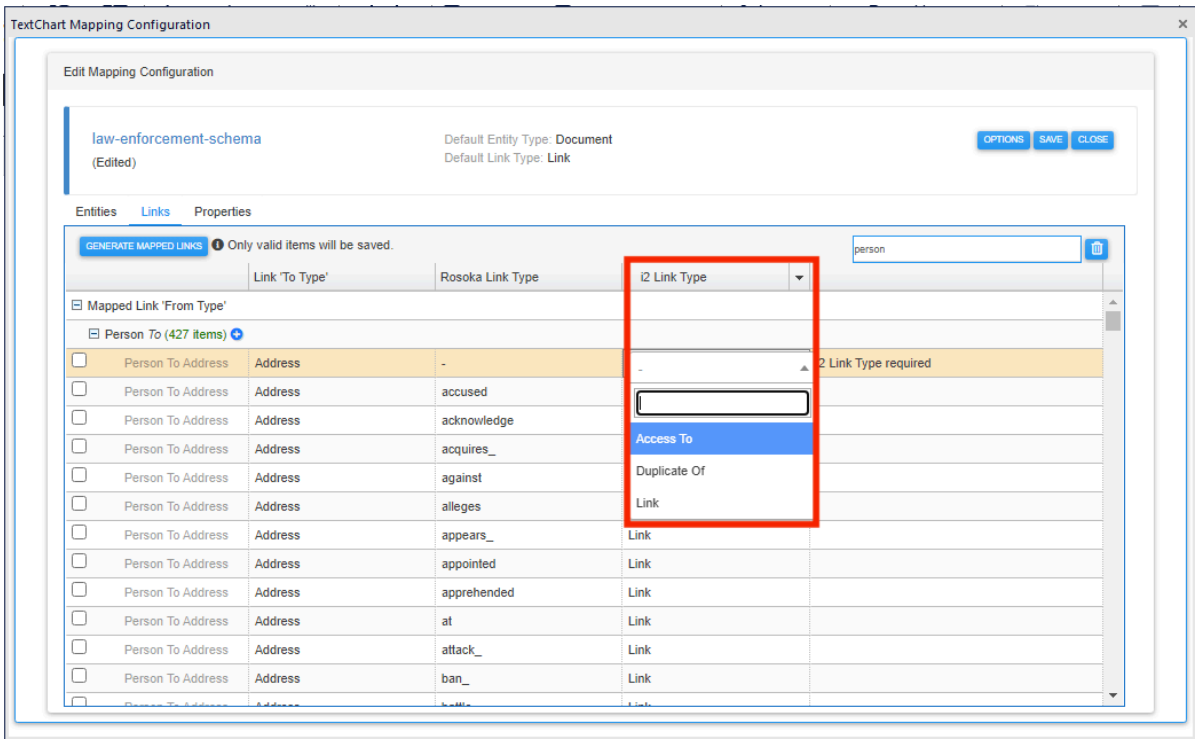
3. In the **Link 'To Type'** column, select the TextChart entity type that forms the "To" end of the link type that you want to map.

© N. Harris Computer Corporation



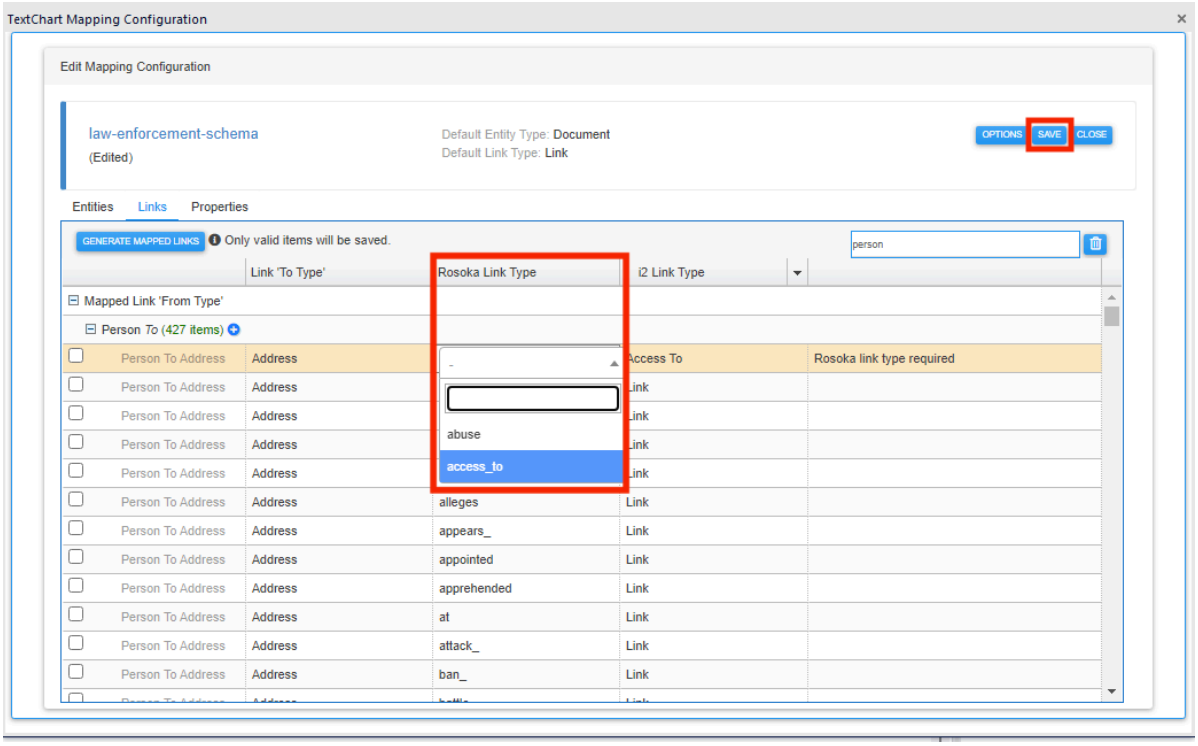
4. In the **i2 Link Type** column, select the i2 Analyze link type that you want to map to.

© N. Harris Computer Corporation



5. Finally, in the **Rosoka Link Type** column, select the TextChart link type that you want to map from, and then click **Save**.

© N. Harris Computer Corporation



Tesseract OCR parser

Tesseract Optical Character Recognition (OCR) parsing may be installed to provide seamless OCR support for many common graphical file formats, including JPEG, TIFF and GIF. Once Tesseract OCR is installed and running, TextChart will seamlessly process supported file formats in the ingestion process via the Tika interface. No further configuration is required.

Installing Tesseract on RHEL

1. Install the EPEL repository if it isn't already available:

   ```
   sudo dnf install epel-release
   ```

2. Install Tesseract:

   ```
   sudo dnf install tesseract
   ```

3. To add language packs, see what's available:

   ```
   dnf search tesseract
   ```

   And then install, for example:

   ```
   sudo dnf install tesseract-langpack-ara
   ```

Installing Tesseract on Ubuntu

1. Download package information from all configured sources:

   ```
   sudo apt update
   ```

© N. Harris Computer Corporation

2. Install Tesseract:

   ```
   sudo apt install tesseract-ocr
   ```

3. To add language packs, see what's available:

   ```
   apt search tesseract-ocr
   ```

   And then install, for example:

   ```
   sudo apt install tesseract-ocr-fra
   ```

Installing Tesseract on Windows

1. Download the latest Tesseract installer from the UB-Mannheim Tesseract repository. Older versions are available at `https://digi.bib.uni-mannheim.de/tesseract/`.
2. Run the installer and follow the prompts. Install Tesseract in the directory suggested by the installer or in a new directory.

   Important: Do not install into an existing directory that contains other files. The uninstaller removes the entire installation directory.
3. During installation, select any additional language packs that you require.
4. Add the Tesseract installation directory to the system `PATH` environment variable so that it is available from the command line.

Installing Tesseract on other platforms

For further details on Tesseract, including installation on Macintosh, see the information at https://cwiki.apache.org/confluence/display/tika/TikaOCR.

i2 TextChart Premium

i2 TextChart Premium Client

i2 TextChart Premium Client integrates with Analyst's Notebook to facilitate the processing, curation, and visualization of unstructured text data. The Premium client allows enhanced capabilities, particularly in scaled and enterprise-level deployments.

Installing i2 TextChart Premium

This topic describes how to install and activate i2 TextChart Premium.

Before you begin

Important: If you are upgrading from an earlier version of i2 TextChart, follow the instructions in Upgrading i2 TextChart before you complete the procedure below.

Prerequisite software

The release notes contain a full list of the required software for i2 TextChart. You *must* install i2 Analyst's Notebook before you install i2 TextChart.

Optional configuration

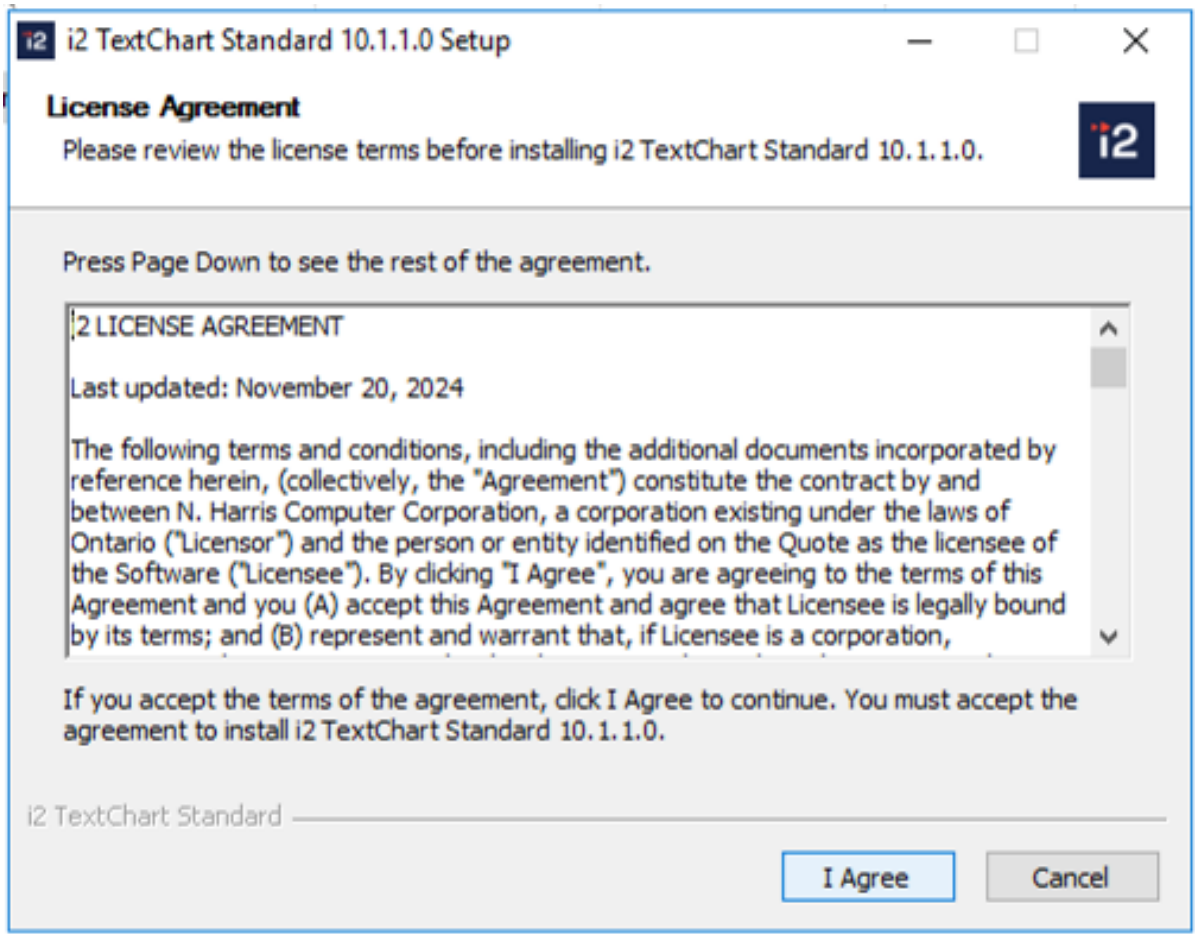
i2 TextChart is supplied with an embedded database for local workspace storage. Optionally, you can replace the embedded database with Microsoft SQL Server 2014 or later.

© N. Harris Computer Corporation

Procedure

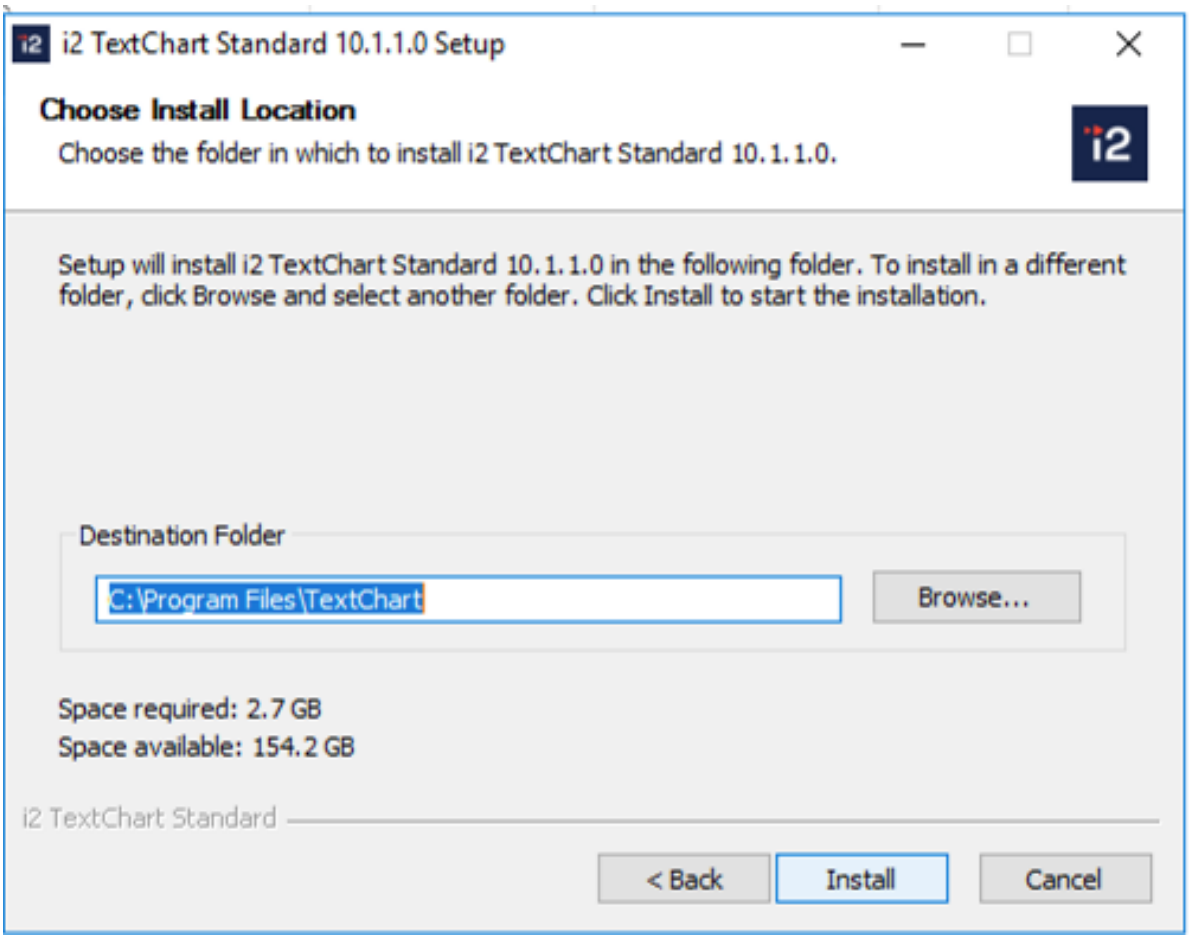
The following steps describe how to install i2 TextChart.

1. Extract the product files from your downloaded distribution.
2. Using Windows Explorer, browse to the root of the distribution and run the i2 TextChart Installer. Its name will contain the word "Premium", as this is the edition of i2 TextChart you're installing.
3. Follow the prompts. You must accept the license agreement to proceed. Click **I Agree**:

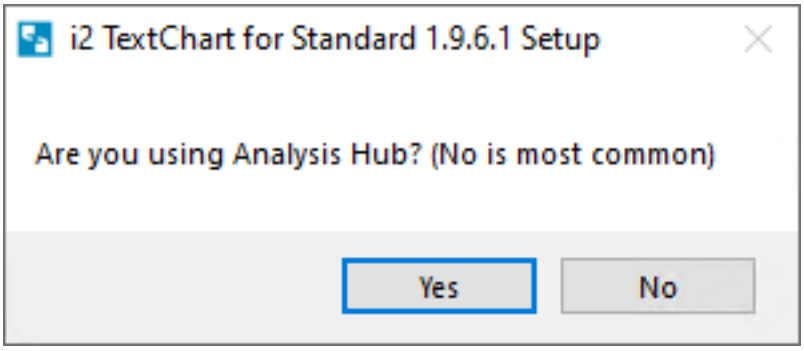


4. Choose the installation location and click **Install**:

© N. Harris Computer Corporation



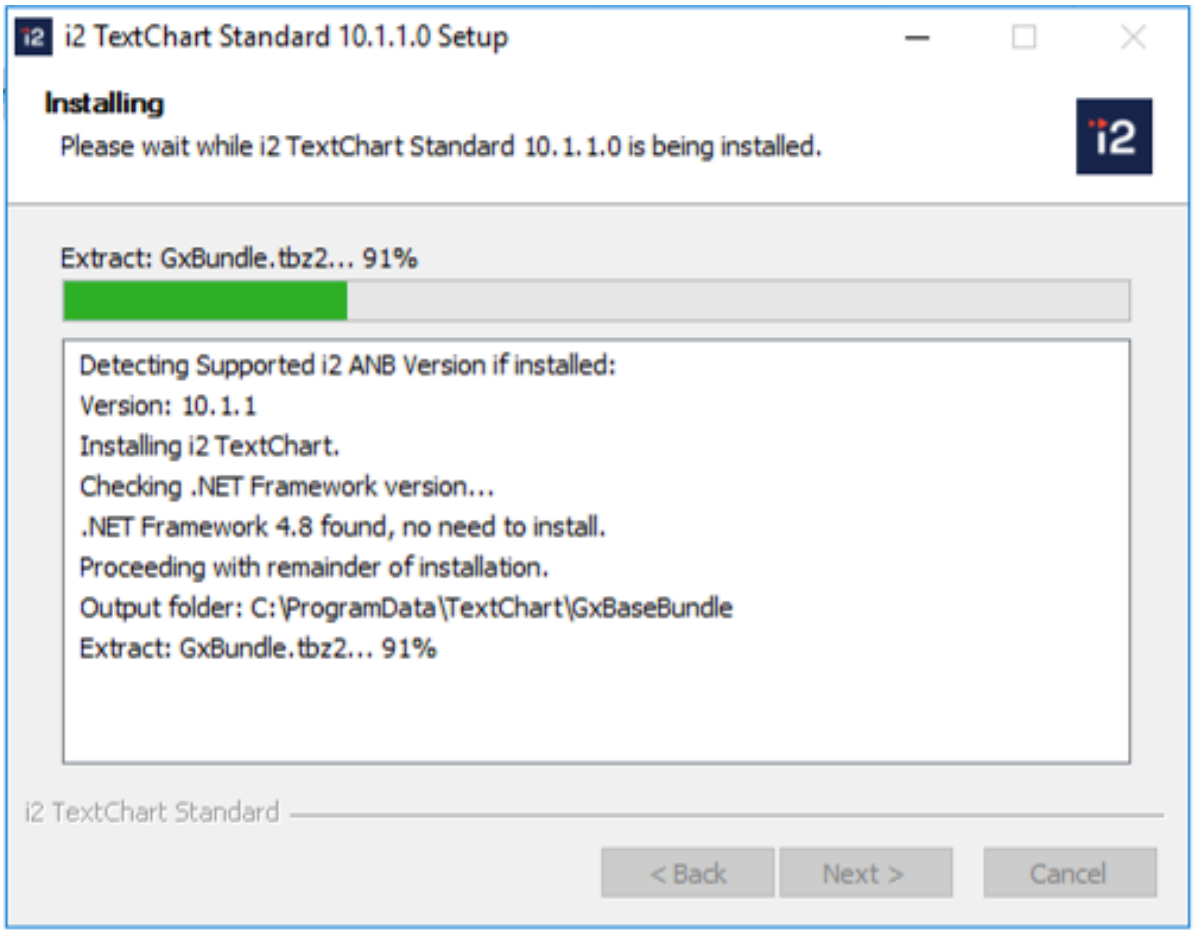
5. If you're installing the standard edition of i2 TextChart, the installer displays an extra question:



   If your workflow involves connecting to an i2 Analyze server from Analyst's Notebook, click **Yes**. Otherwise, click **No**.

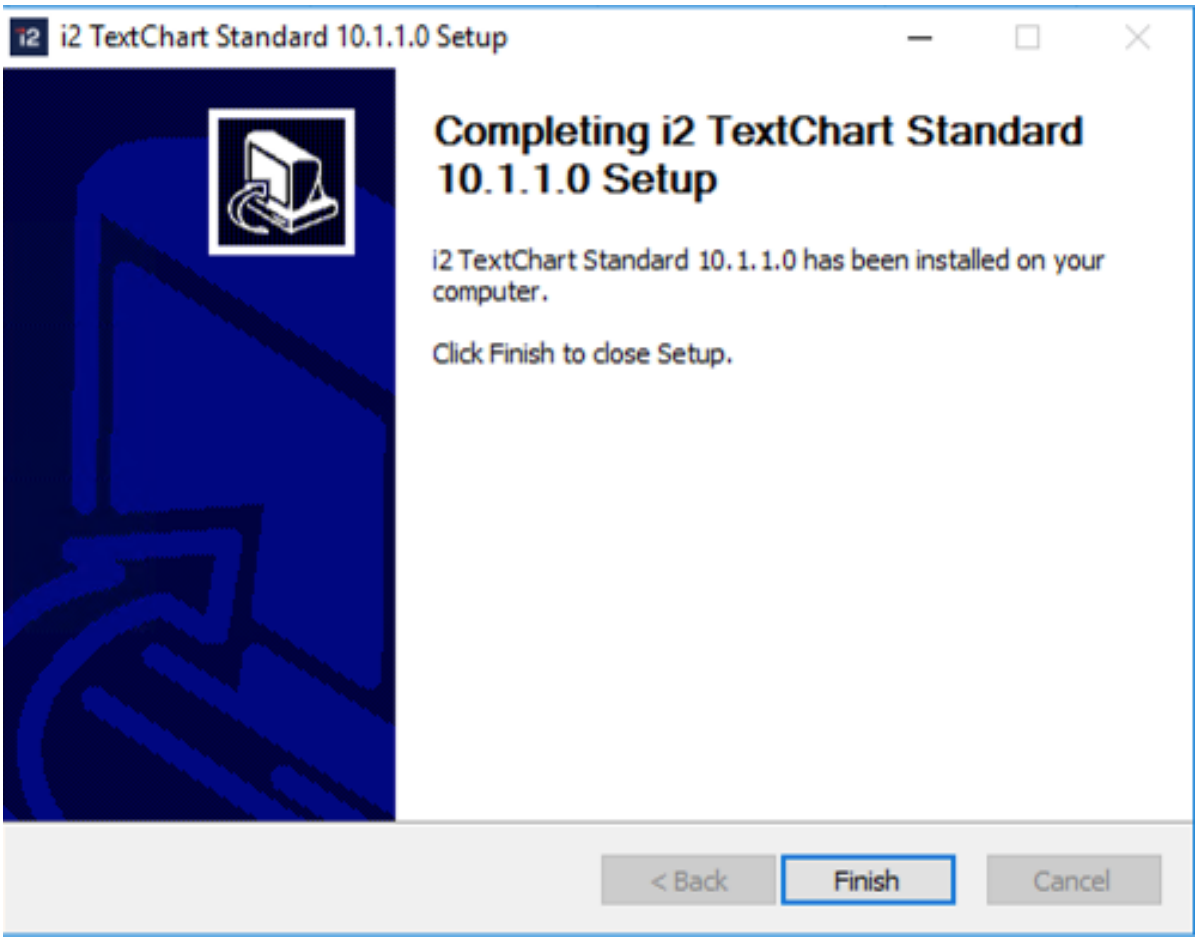
6. When the installer is complete, click **Next**.

© N. Harris Computer Corporation



7. When setup is complete, leave the **Configure TextChart** checkbox selected, and click **Finish**.

© N. Harris Computer Corporation



Installing i2 TextChart Silently

The following steps describe how to install i2 TextChart silently:

1. Extract the product files from your downloaded distribution.
2. Using Windows Command Line Terminal, navigate to the root of the distribution and run the i2 TextChart Installer by issuing following command. Note that Analyst's Notebook is closed while installing the TextChart Text Analytics.

   ```
   "i2 TextChart Premium-X.X.X.X-Installer.exe" /S
   ```

   Note: If you need to uninstall i2 TextChart Text Analytics silently, you can do so by issuing following command. Note that Analyst's Notebook is closed while uninstalling the TextChart Text Analytics.

   ```
   Uninstall.exe /S
   ```

   Uninstall.exe should be available in installation directory, `C:\Program Files\TextChart` for example.

© N. Harris Computer Corporation

What to do next

Before you launch i2 TextChart for the first time, refer to Configuring i2 TextChart for proper configuration.

When i2 TextChart is successfully installed, activated, and configured, the plug-in launches automatically when you open i2 Analyst's Notebook.

Note: If you need to stop i2 TextChart Text Analytics, you can do so through the Service Manager, which is available from the **i2 TextChart** menu in the Analyst's Notebook ribbon. To restart i2 TextChart Text Analytics, you must then restart Analyst's Notebook.

Configuring i2 TextChart Premium

i2 TextChart Premium is ready to use as soon as you install and activate it. However, to maintain and enhance its capabilities, TextChart also provides a range of options for configuring and updating it.

Connecting to TextChart Manager

This topic describes how to configure i2 TextChart Premium Clients to connect to a TextChart Manager infrastructure for enterprise-scale text analysis processing.

About this task

Configuring i2 TextChart Premium Client to connect to TextChart Server components enables distributed, scalable text processing across your enterprise environment. In this configuration, i2 TextChart Premium Client plugin functions as a client application that submits documents to the server's REST API and retrieves processed results.

This architecture provides several advantages:

- **Scalability**: Leverage the distributed processing power of TextChart Manager clusters
- **Performance**: Offload intensive processing tasks to dedicated server infrastructure
- **Consistency**: Maintain centralized text analysis configurations across multiple clients
- **Resource optimization**: Preserve local workstation resources for analysis and visualization tasks

Prerequisites

Before configuring the server connection, verify the following requirements:

- **Server infrastructure**: TextChart backend components (TextChart Manager, TextChart Worker nodes, and TextChart Data Access) are installed, configured, and operational
- **Network connectivity**: The client workstation has network access to the TextChart infrastructure
- **Connection details**: You have obtained the following information from your system administrator:
  - TextChart Manager hostname or IP address and port number
  - Cluster name for the TextChart Manager
  - TextChart Data Access hostname or IP address and port number
  - SSL configuration requirements (if applicable)
- **Certificates**: If using SSL/TLS encryption, the necessary certificates are installed and configured
- **Firewall configuration**: Network firewall rules permit communication between the client and server components

© N. Harris Computer Corporation

Procedure

To configure i2 TextChart Premium for server connectivity:

1. Launch the i2 TextChart Premium Configuration application.
2. Select the **Server Connection** tab.

   Note: If this tab is not available, verify that you have the Premium edition of i2 TextChart installed and licensed.
3. Configure the TextChart Manager connection:

   a. In the **Manager Connection** section, specify the following parameters:

   - **Manager Host**: Enter the fully qualified domain name or IP address of the TextChart Manager
   - **Port**: Enter the client service port number (default: 8081 for HTTP, 9443 for HTTPS)
   - **Cluster Name**: Enter the cluster identifier as configured on the TextChart Manager

   b. Click **Test** to verify connectivity to the TextChart Manager.

   c. Upon successful connection validation, click **Save** to store the configuration.
4. Configure the TextChart Data Access connection:

   a. In the **Data Access Connection** section, specify the following parameters:

   - **Data Access Host**: Enter the fully qualified domain name or IP address of the TextChart Data Access server
   - **Port**: Enter the client service port number (default: 9080 for HTTP, 9443 for HTTPS)

   b. Click **Test** to verify connectivity to the TextChart Data Access server.

   c. Upon successful connection validation, click **Save** to store the configuration.
5. **Troubleshooting connection failures**: If connection tests fail, verify the following:

   - Server components are running and accessible from the client workstation
   - Port numbers match the server configuration
   - Network routing and firewall rules permit the required connections
   - SSL certificates are properly installed (if using encrypted connections)
   - DNS resolution is functioning correctly for hostname-based connections
6. After successful configuration of both connections, restart i2 TextChart Premium to activate the server integration.

Verification

To confirm successful server integration:

1. Launch i2 TextChart Premium Client alongwith ANB and verify that the application indicates server connectivity status.
2. Process a test document to validate end-to-end functionality.
3. Monitor the processing workflow to confirm that tasks are being distributed to the server infrastructure.

© N. Harris Computer Corporation

What's next

With server connectivity established, i2 TextChart Premium Client will automatically route text analysis requests to the TextChart Server infrastructure. You can monitor processing status and analyze results through the standard i2 TextChart Premium interface while benefiting from the enhanced processing capabilities and scalability of the server environment.

For information about configuring SSL encryption, see Configuring SSL for i2 TextChart Server.

For advanced configuration options, see Other configuration.

Changing the supported file extensions

The TextChart plug-in for Analyst's Notebook supports processing documents in a wide range of *formats* including plain text, PDF, and those that productivity applications generate. It uses file *extensions* to filter the documents that it presents for import, and you can change the configuration to display more or different files.

TextChart stores its list of file extensions in the file at `C:\ProgramData\TextChart\conf\filetypes.properties`. If an extension appears in this file, then documents with that extension appear in the Document View. By default, TextChart presents documents with the following extensions:

```
txt
rtf
pdf
doc
docx
pptx
```

The default `filetypes.properties` file also specifies that TextChart should present files that have no extension for import.

If your organization uses extensions that are not in the default list, you must add them to it. For example, you might have older Microsoft Powerpoint files with the `.ppt` extension, or plain text files that use a custom extension.

After you make changes to the list of supported file extensions, reinitialize the TextChart plug-in by closing and reopening the Analyst's Notebook application.

Changing the local database

By default, the standard edition of i2 TextChart uses its embedded database for storing the information that it generates during analysis. Optionally, you can configure TextChart standard edition to use a local Microsoft SQL Server database instead.

Before you begin

Choosing a SQL Server database in place of the embedded one enables you to inspect its contents directly. TextChart supports SQL Server 2014 and later, including the Express edition.

To configure TextChart to use a local SQL Server database, you use the Text Analytics Configuration application, which you also used to provide TextChart with a license key during installation.

To run the application after installation, you can start it from the Windows Start menu, or by launching the executable at `C:\Program Files\TextChart\RosokaConfiguration.exe`.

© N. Harris Computer Corporation

Procedure

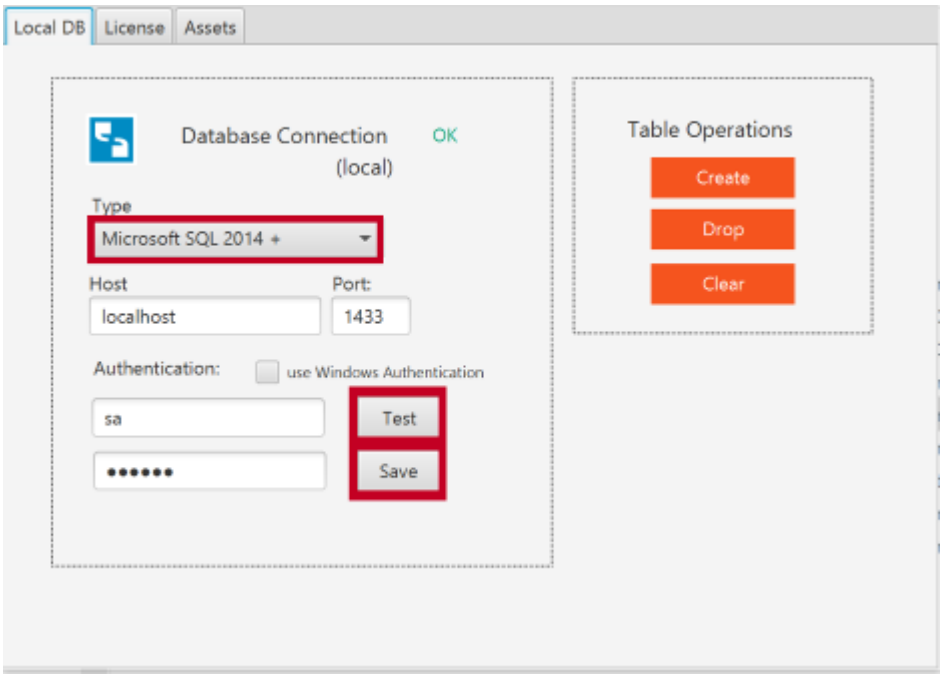
Assuming that you have already set up a SQL Server database for TextChart to use, the procedure for connecting to it is as follows.

Note: If you want to create a database account for TextChart to use, or to change the port on which SQL Server accepts connections, see the instructions in Database configuration.

1. On the **Local DB** tab of the TextChart Text Analytics Configuration application, select your database from the **Type** drop-down, and ensure that the **Host** and **Port** details are correct.

   Note: For the standard edition of TextChart, the host must be the same workstation that TextChart and Analyst's Notebook are installed on.

2. In the **Authentication** fields, provide the user name and password of a database account with administrator privileges.



3. Click **Test** to verify that the information you provided is correct. If the test is successful, click **Save**.

   Note: If you see the "Driver not Present" error message, you need to add the appropriate JAR file to the `C:\Program Files\TextChart\Drivers` directory, relaunch the configuration application, and try again.

   Drivers for Microsoft databases are included, drivers for MySQL are available from the MySQL website, and drivers for Db2 are available as part of your Db2 installation.

What to do next

If you need to change the configuration of the local database after you've begun to use i2 TextChart, you must close Analyst's Notebook and ensure that the TextChart application is no longer running before repeating the above procedure.

© N. Harris Computer Corporation

Note: After you change the configuration to use a different database from the one you've been using, you might need to re-create your collections and reprocess the documents they contained.

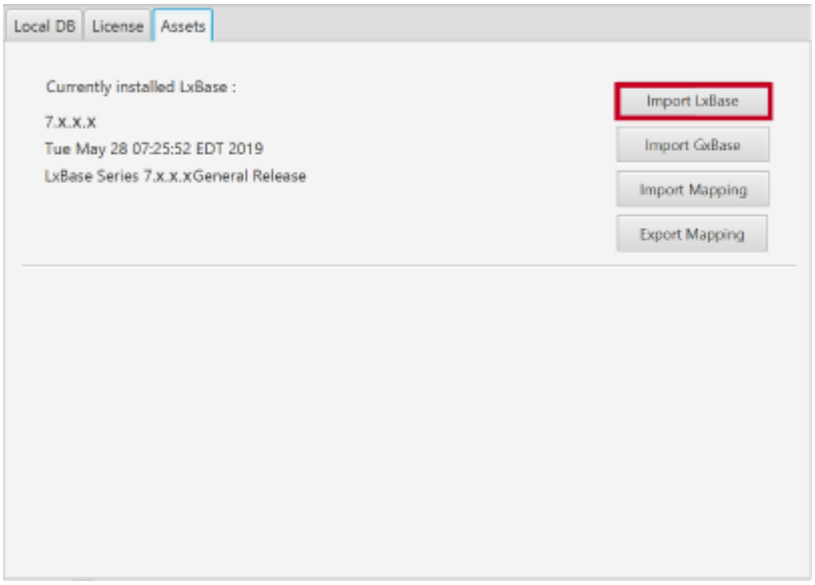
Updating an LxBase

In i2 TextChart, an LxBase is the collection of linguistic rules and dictionary entries that govern text extraction. From time to time, i2 publishes updates that you can apply to the LxBase in your installation.

Procedure

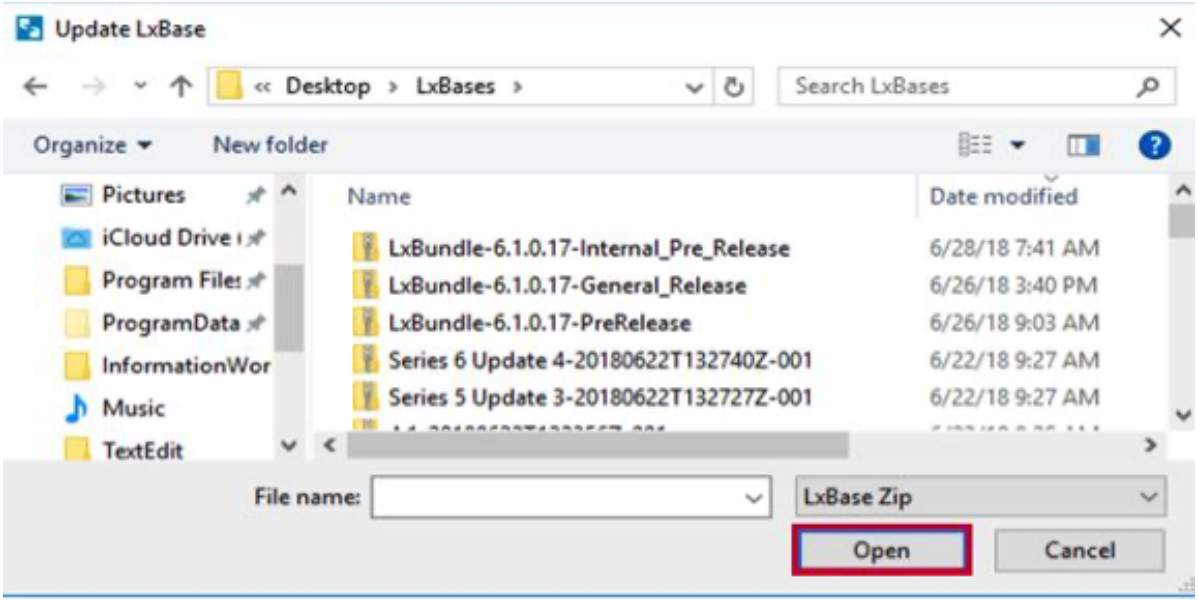
To update your LxBase, you use the TextChart Text Analytics Configuration application to provide i2 TextChart with the location of a ZIP file that contains the new information.

1. Open the configuration application and select the **Assets** tab:



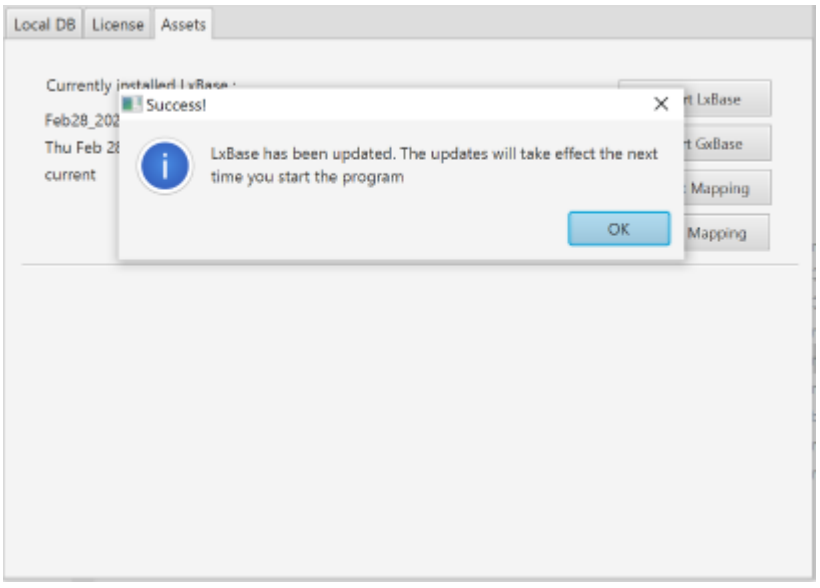
2. Click **Import LxBase**, navigate to the new ZIP file, and click **Open**.

© N. Harris Computer Corporation



i2 TextChart automatically updates your LxBase with information from the file.

3. When the following dialog appears, click **OK**:



The changes and updates will take effect when you restart Analyst's Notebook.

Note: i2 TextChart backs up the previous version of your LxBase in `C:\ProgramData\TextChart`, in a directory named `LxBase-backup-<TIMESTAMP>`. To restore a backup, delete the `LxBase` directory, and then rename the backup to `LxBase`.

© N. Harris Computer Corporation

Upgrading i2 TextChart Premium client

Upgrading i2 TextChart Premium requires a complete uninstallation of the previous version before installing the new release. This procedure ensures a clean installation and prevents potential conflicts between versions.

About this task

The upgrade process consists of the following stages:

1. Back up your existing database and configuration
2. Uninstall the current version of i2 TextChart Premium client
3. Install the new version using the standard installation procedure
4. Restore database content and reconfigure server connections

Prerequisites

Before beginning the upgrade:

- Ensure you have administrator privileges on the workstation
- Verify that no users are actively using i2 TextChart Premium or i2 Analyst's Notebook
- Obtain the installation package for the new version
- Note your current server connection settings for reconfiguration

Procedure

Step 1: Back up your existing database

1. Close all instances of i2 Analyst's Notebook and i2 TextChart Premium.
2. If you are using a local TextChart database, create a backup and clear the database:

   a. **For SQL Server databases**: Use SQL Server Management Studio or other database tools to create a full backup of your TextChart database.

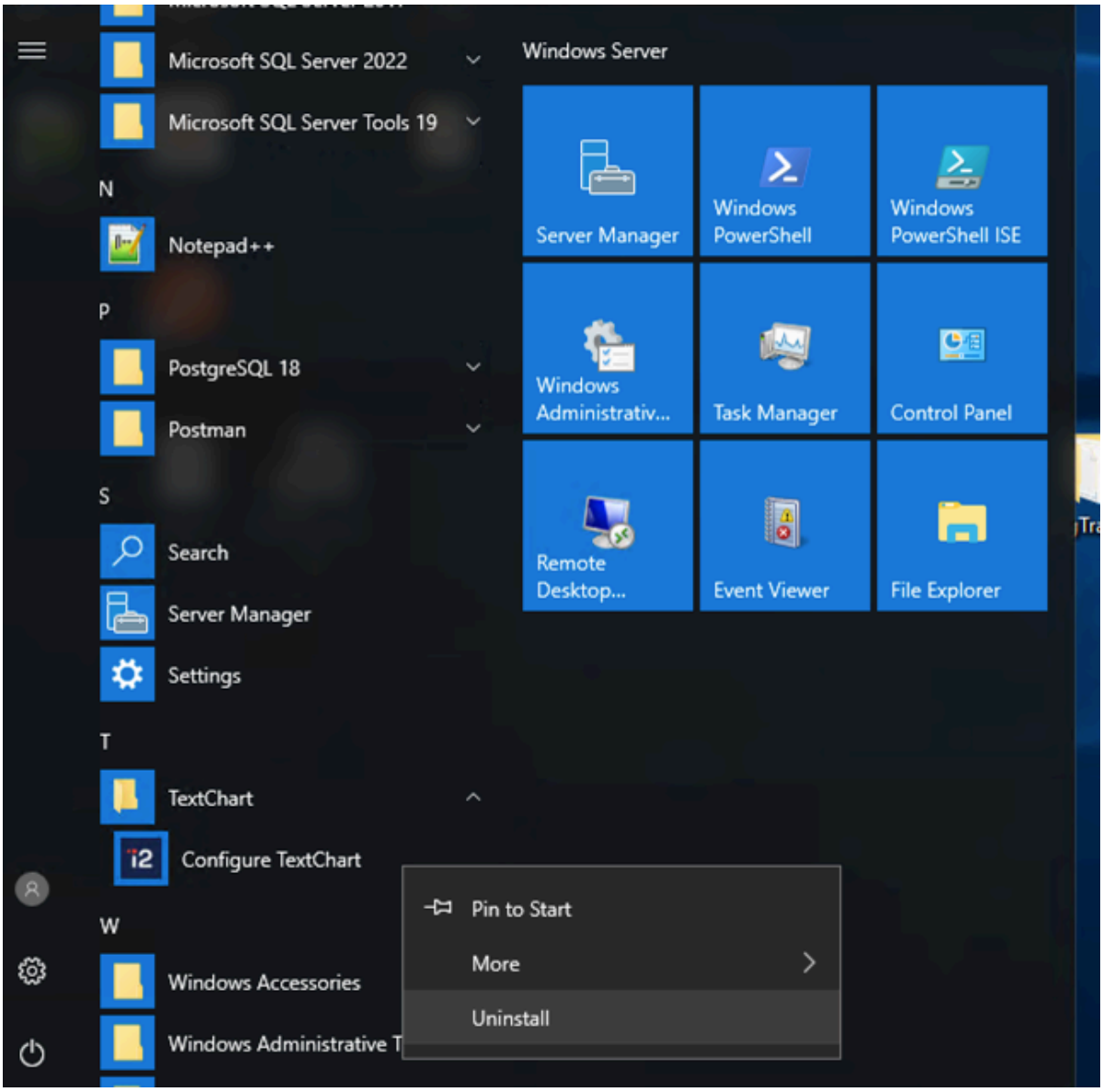
   b. **For embedded databases**: Navigate to the database directory at `C:\ProgramData\TextChart\databases\Rosoka` (or `C:\ProgramData\Rosoka\TextAnalytics\databases\Rosoka` for older versions) and create a copy of the entire folder.
3. Open the TextChart Text Analytics Configuration application.
4. Navigate to the **Local DB** tab and click **Clear** to clear the existing database content.

   Note: This step prepares the database for the new version while preserving your backup for restoration if needed.

Step 2: Uninstall the current version

1. Uninstall i2 TextChart Premium using one of the following methods:
   - From the Windows Start menu, locate and run the TextChart uninstaller
   - Navigate to `C:\Program Files\TextChart\` and run `Uninstall.exe`
   - Use Windows **Add or Remove Programs** from the Control Panel

© N. Harris Computer Corporation



2. Wait for the uninstallation process to complete before proceeding.

Step 3: Install the new version

Follow the standard installation procedure documented in Installing i2 TextChart Premium.

Step 4: Reconfigure server connections

After successful installation of the new version:

1. Follow the configuration steps in Connecting to Server to restore connectivity to the TextChart Manager and TextChart Data Access.
2. Restore your database content from the backup created in Step 1, if required.

© N. Harris Computer Corporation

3. Test the installation by processing a sample document to verify that all components are functioning correctly.

What's next

After completing the upgrade, consider reviewing the release notes for the new version to understand any new features or configuration changes that may benefit your deployment.

Other configuration

i2 TextChart provides a few ways to modify its behavior so that it works better in your environment. For example, you can adjust the amount of memory available to the extraction process, and change the way TextChart authenticates with the database.

Memory configuration

If you experience poor performance, especially during processing of large files, you can change the size of the JVM heap by editing the configuration file at `C:\Program Files\TextChart \RosokaServiceCore.l4j.ini`.

To adjust memory allocation, which defaults to 4 GB, change:

```
-Xms4g
-Xmx4g
```

to:

```
-Xms<GB_OF_RAM_TO_ALLOCATE>g
-Xmx<GB_OF_RAM_TO_ALLOCATE>g
```

For example, to set it to 8 GB, the two lines would be:

```
-Xms8g
-Xmx8g
```

Database configuration (SQL Server)

TextChart authenticates with SQL Server through database user credentials rather than operating system credentials. You need to ensure that your installation of SQL Server is configured appropriately.

Setting mixed authentication

1. Open SQL Server Management Studio.
2. From the Authentication drop-down list, select **Windows Authentication** .
3. Click **Connect**.
4. Right-click the SQL instance and choose **Properties** to open the Server Properties window.
5. In the **Select a page** list, click **Security**.
6. In the **Server authentication** section, choose **SQL Server and Windows Authentication mode**.
7. Click **OK** to close the Server Properties window.
8. Right-click the SQL instance and choose **Restart**.
9. Click **Yes** to restart the SQL server.

Enabling the 'sa' user

1. Open SQL Server Management Studio.
2. From the Authentication drop-down list, select **Windows Authentication**.

© N. Harris Computer Corporation

3. Click **Connect**.
4. In the Object Explorer, expand **Security** and then expand **Logins**.
5. Right-click **sa**, and click **Properties**.
6. In the **Select a page** list, click **Status**, and then change the **Login** setting to **Enabled**.

   Note: You might need to set a password for this account as well.
7. Click **OK** to close the Login Properties window.
8. Press *F5* to refresh the information in the Object Explorer.

Changing TCP and port settings

1. From the **Start** menu, right-click **My Computer**, and then select **Manage**.
2. In the Computer Management window, expand **Services and Applications > SQL Server Configuration Manager > SQL Server Network Configuration**.
3. Select the instance for your installation. By default, this is **Protocols for SQLEXPRESS**.
4. Change the **Named Pipes** protocol option to **Enabled**. You do not need to change the default **Pipe Name** setting.
5. Change the TCP/IP protocol option to **Enabled**, and then open its **Properties** dialog.
6. In the **IP Addresses** tab, scroll down to **IPAll** at the bottom of the list.
7. Set **TCP Port** to 1433, and remove any value from **TCP Dynamic Ports**.
8. Click **Apply**. You will be prompted to restart SQL Server.
9. Restart the server for the changes to take effect.

Tesseract OCR parser

Tesseract Optical Character Recognition (OCR) parsing may be installed to provide seamless OCR support for many common graphical file formats, including JPEG, TIFF and GIF. Once Tesseract OCR is installed and running, TextChart will seamlessly process supported file formats in the ingestion process via the Tika interface. No further configuration is required.

Installing Tesseract on RHEL

1. Install the EPEL repository if it isn't already available:

   ```
   sudo dnf install epel-release
   ```
2. Install Tesseract:

   ```
   sudo dnf install tesseract
   ```
3. To add language packs, see what's available:

   ```
   dnf search tesseract
   ```

   And then install, for example:

   ```
   sudo dnf install tesseract-langpack-ara
   ```

Installing Tesseract on Ubuntu

1. Download package information from all configured sources:

   ```
   sudo apt update
   ```
2. Install Tesseract:

   ```
   sudo apt install tesseract-ocr
   ```

© N. Harris Computer Corporation

3. To add language packs, see what's available:

   ```
   apt search tesseract-ocr
   ```

   And then install, for example:

   ```
   sudo apt install tesseract-ocr-fra
   ```

Installing Tesseract on Windows

1. Download the latest Tesseract installer from the UB-Mannheim Tesseract repository. Older versions are available at `https://digi.bib.uni-mannheim.de/tesseract/`.
2. Run the installer and follow the prompts. Install Tesseract in the directory suggested by the installer or in a new directory.

   Important: Do not install into an existing directory that contains other files. The uninstaller removes the entire installation directory.
3. During installation, select any additional language packs that you require.
4. Add the Tesseract installation directory to the system `PATH` environment variable so that it is available from the command line.

Installing Tesseract on other platforms

For further details on Tesseract, including installation on Macintosh, see the information at https://cwiki.apache.org/confluence/display/tika/TikaOCR.

PostgreSQL Integration Guide

This document provides instructions for setting up and using PostgreSQL integration across both **TextChart Extraction Manager**, **TextChart Extraction Worker**, and **TextChart Data Access** components. It consolidates all necessary setup, configuration, and troubleshooting steps into one cohesive guide.

Prerequisites:

- Microsoft Windows
- i2 Analyst's Notebook
- TextChart Data Access
- TextChart Extraction Manager
- TextChart Extraction Worker
- A PostgreSQL database that already has the base TextChart tables created.
- Valid Licensing for i2 products.

Overview

This guide helps customers connect their applications to PostgreSQL when using **TextChart Extraction Manager**, **TextChart Extraction Worker** and **TextChart Data Access** components. It focuses on user actions and practical setup steps rather than technical implementation details.

Installation and Requirements

Ensure the following requirements are met before beginning setup:

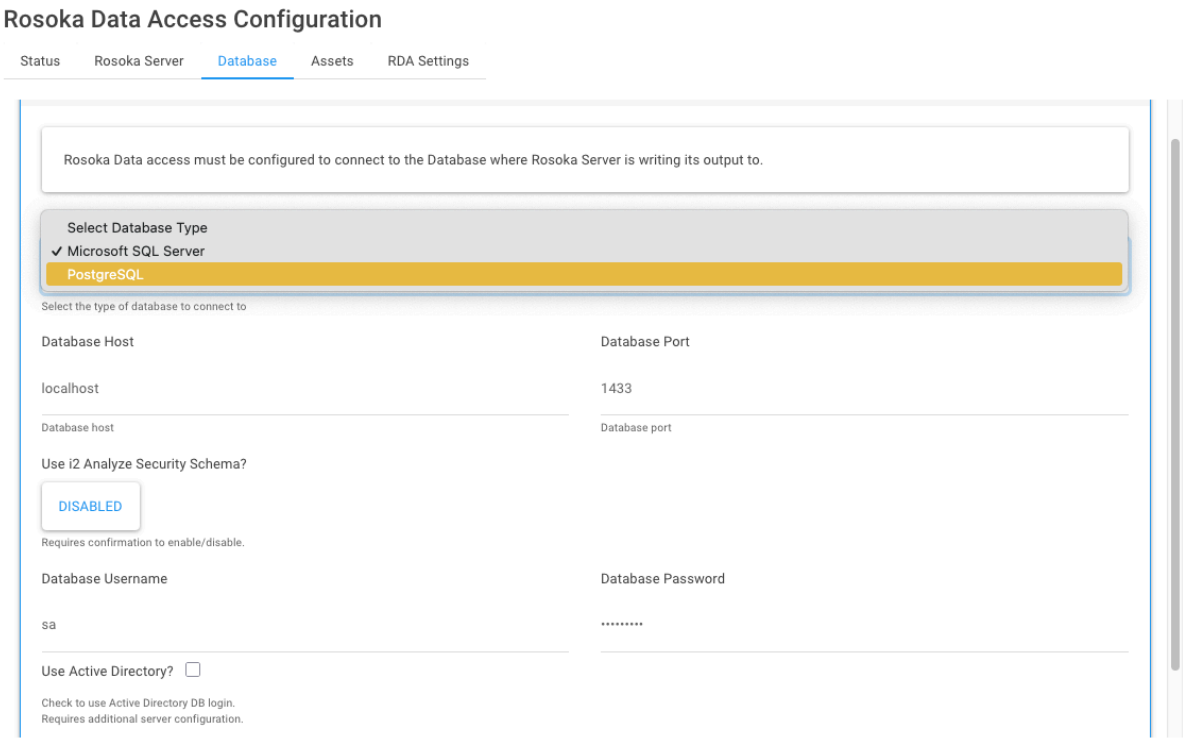
- PostgreSQL installed and running

© N. Harris Computer Corporation

- Network access to the PostgreSQL host and port (default: 5432)
- Properly configured database credentials (username, password)
- Appropriate permissions to create or modify configuration files
- Create TextChart Database Tables by running the provided SQL scripts against your PostgreSQL database.

Configuring PostgreSQL within the Admin UI

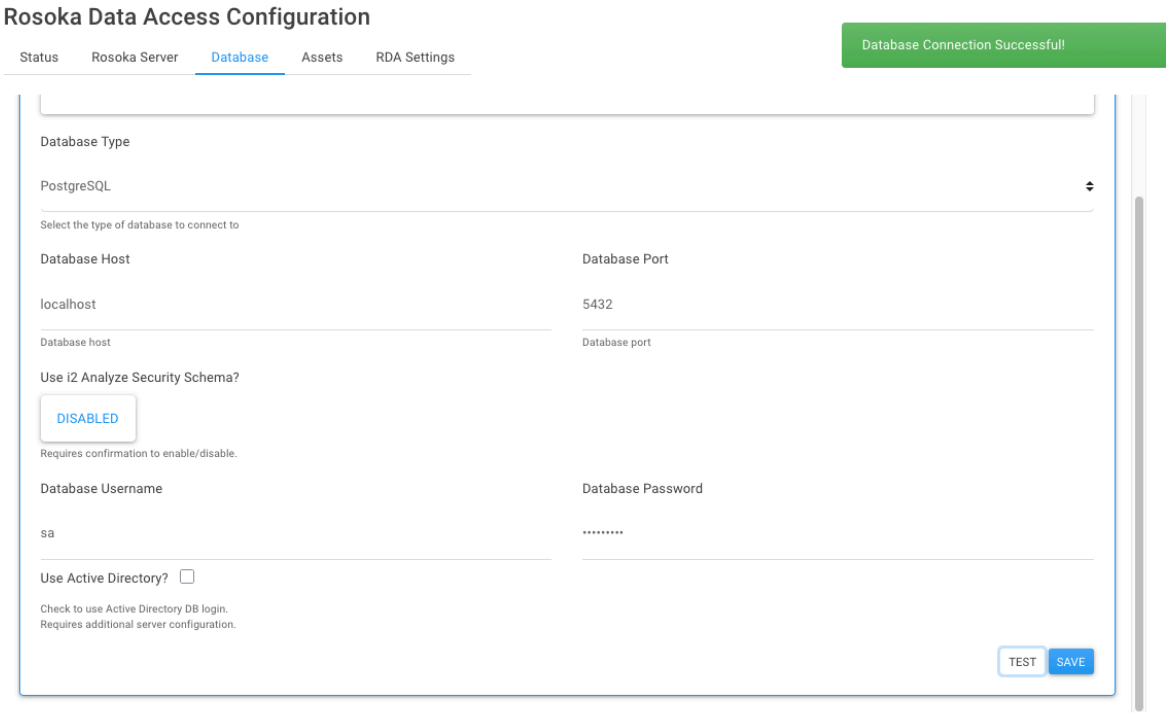
1. Navigate to the Database tab
2. Select PostgreSQL from the Database Type dropdown



1. Enter host, port (5432), username, and password
2. Click Test to verify connectivity

© N. Harris Computer Corporation

3. Click Save to persist configuration

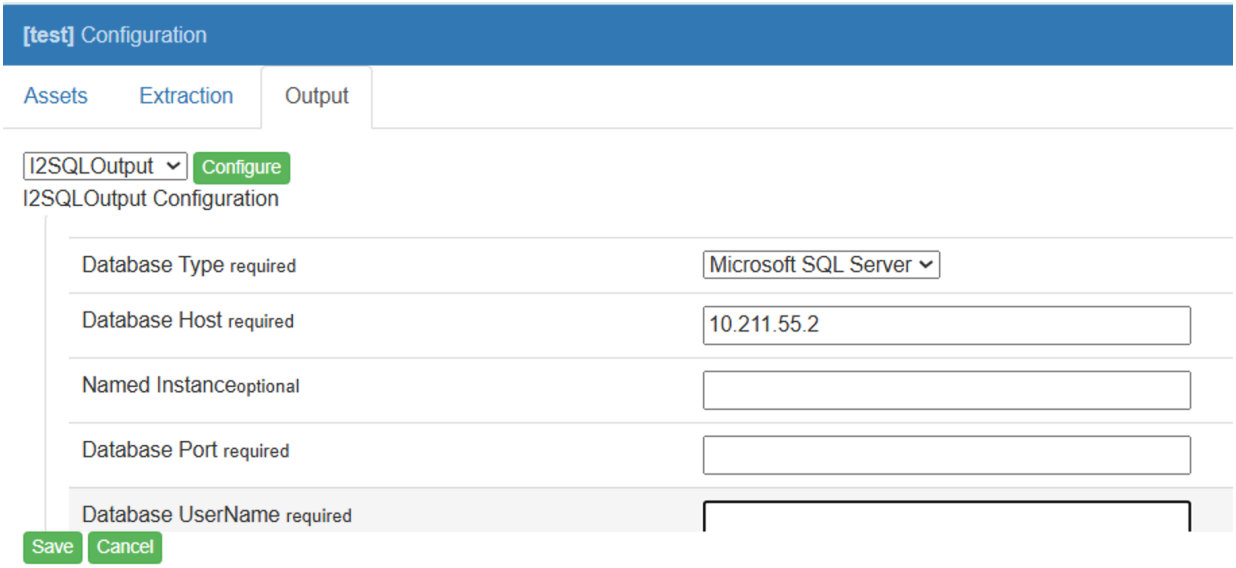


4. Restart the server when prompted

Configuring PostgreSQL within the Manager UI

1. Open Extraction Manager in a browser and navigate to the Clusters tab.
2. Then, click on the gear next to the cluster you want to enable security on and choose Configure.
3. Select the Output tab and then press the Configure button.
4. The first dropdown is Database Type; select PostgreSQL from that list.

© N. Harris Computer Corporation



1. From here, Fill in the Database Information:
   - **Database Host**: The hostname or IP address of the PostgreSQL server.
   - **Database Port**: The port on which the PostgreSQL server is listening (default is 5432).
   - **Database Named Instance**: The name of the PostgreSQL database to connect to.
   - **Database UserName**: The username for authenticating to the PostgreSQL database.
   - **Database Password**: The password for the specified database user.
2. Press the Save and then Save Changes button.

Additional Resources

- PostgreSQL JDBC Documentation: https://jdbc.postgresql.org/
- PostgreSQL System Catalogs: https://www.postgresql.org/docs/current/catalogs.html
- EclipseLink PostgreSQL Platform: https://www.eclipse.org/eclipselink/documentation/

Troubleshooting

Common issues and resolutions include:

- **Connection Refused:** Ensure the PostgreSQL server is running and the host/port are correct.
- **Authentication Failed:** Double-check credentials in the configuration files.
- **Table Not Found:** Confirm that all required tables exist in the database.
- **Driver Missing:** Ensure the PostgreSQL JDBC driver is installed in your environment.

© N. Harris Computer Corporation

Importing and processing documents

The i2 TextChart plug-in for Analyst's Notebook processes documents to extract information about the things they describe and the connections between them. You can add the extracted information to your charts in the form of entities and links.

Entity processing

Entities are the important, named items, such as the people, places, and events that a document describes. i2 TextChart uses the linguistic context of the document to determine what words or phrases are extracted as entities.

After processing, you can modify entity extraction results and apply your own knowledge before sending results to a chart.

Note: The LxBase documentation includes a list of all the entity types that TextChart understands, with their corresponding definitions.

Link processing

Links are relationships between two entities that i2 TextChart establishes through linguistic context. In extraction results, links appear in the form `<Entity>To<Entity>`. For example, `PersonToPerson` means that there is a link between one Person entity and another Person entity.

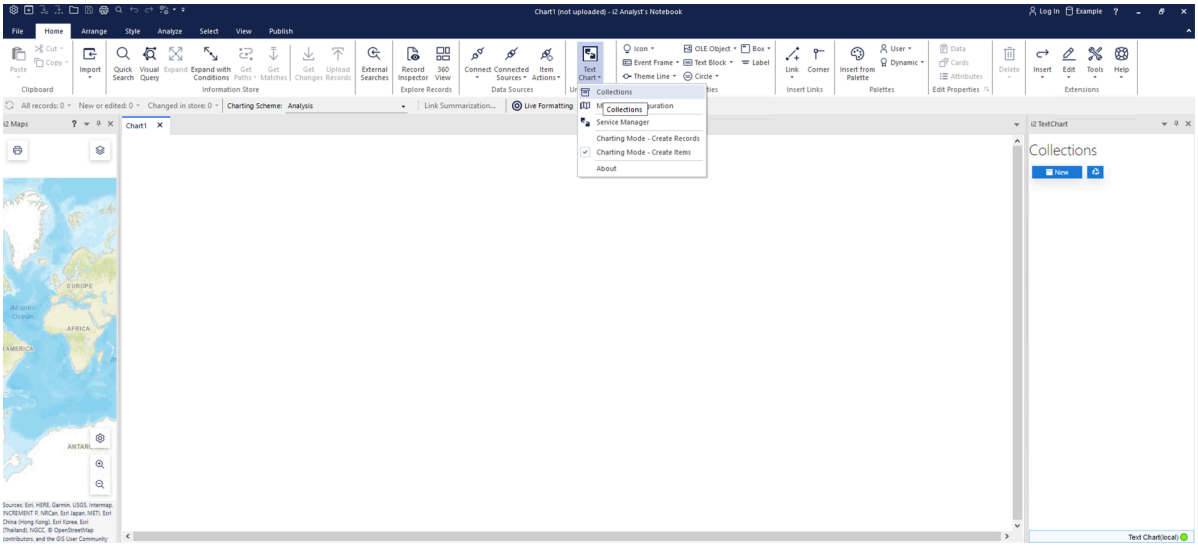
The predicate (that is, the reason why i2 TextChart identified the link) is stored with the link. If the predicate is "interviewed", then the link indicates that a Person interviewed another Person.

Note: The LxBase documentation includes a list of all the predicate types that TextChart understands.

Importing documents

To use i2 TextChart, the first step is to import the documents that you want to analyze.

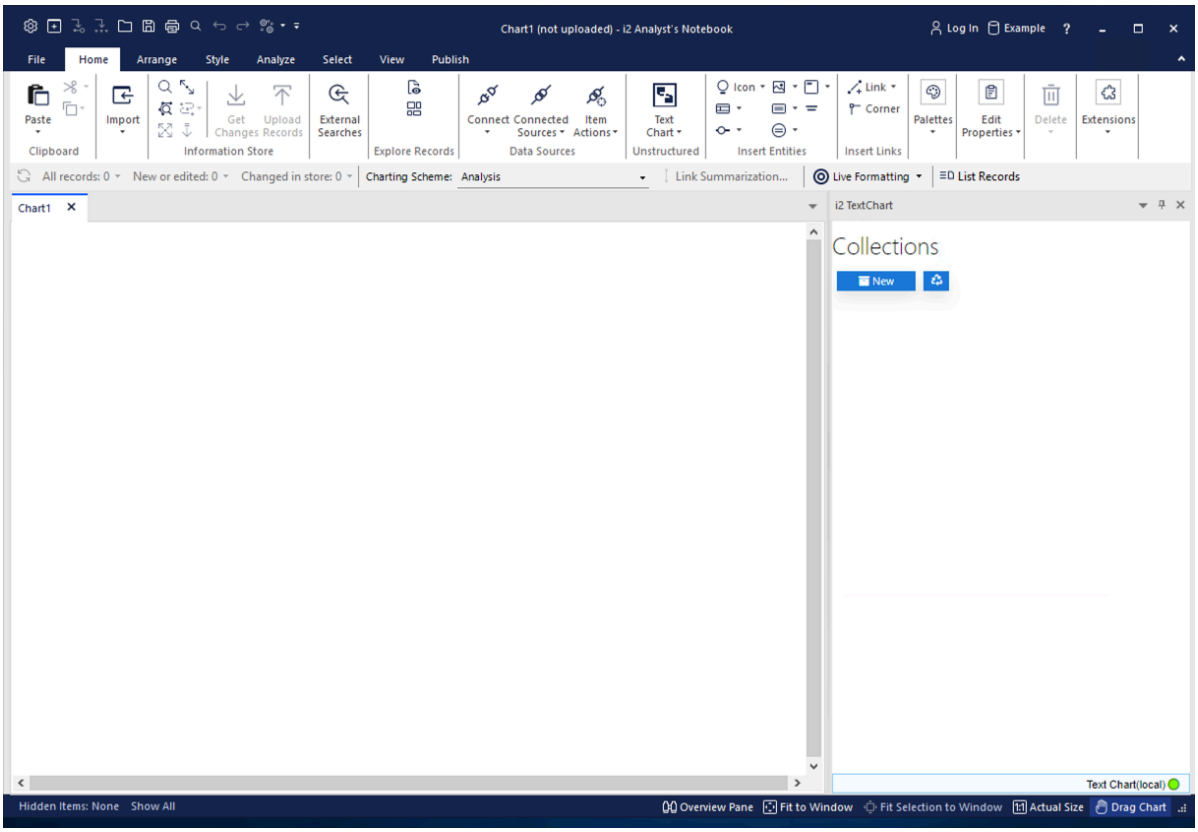
1. In Analyst's Notebook, select the **Text Analytics** menu in the **Home** tab, and then click **Collections**.



© N. Harris Computer Corporation

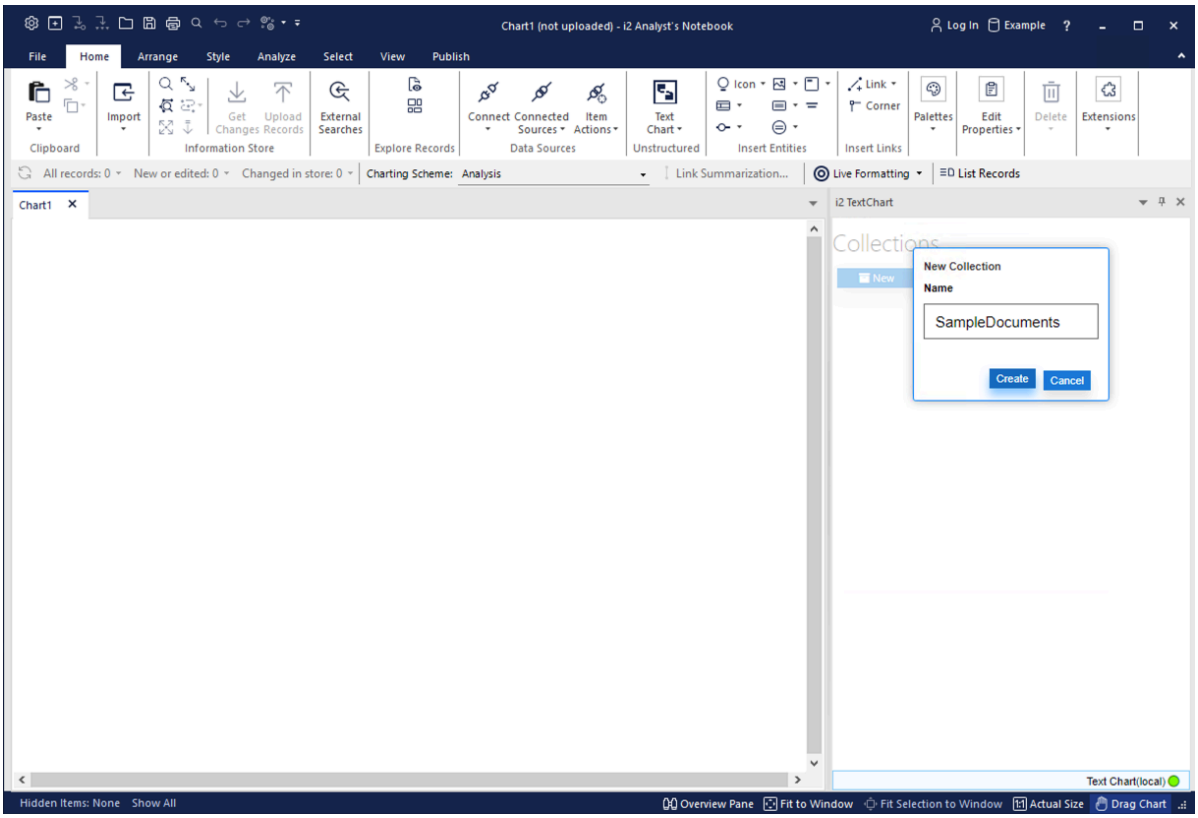
Note: The Mapping Configuration option is available only in the standard edition of TextChart. In TextChart Premium, mapping takes place on the server.

2. i2 TextChart stores documents in *collections*. A collection contains one or more documents that you choose to be in the same group for processing. You can start a new collection for the documents that you're importing, or add them to an existing one.



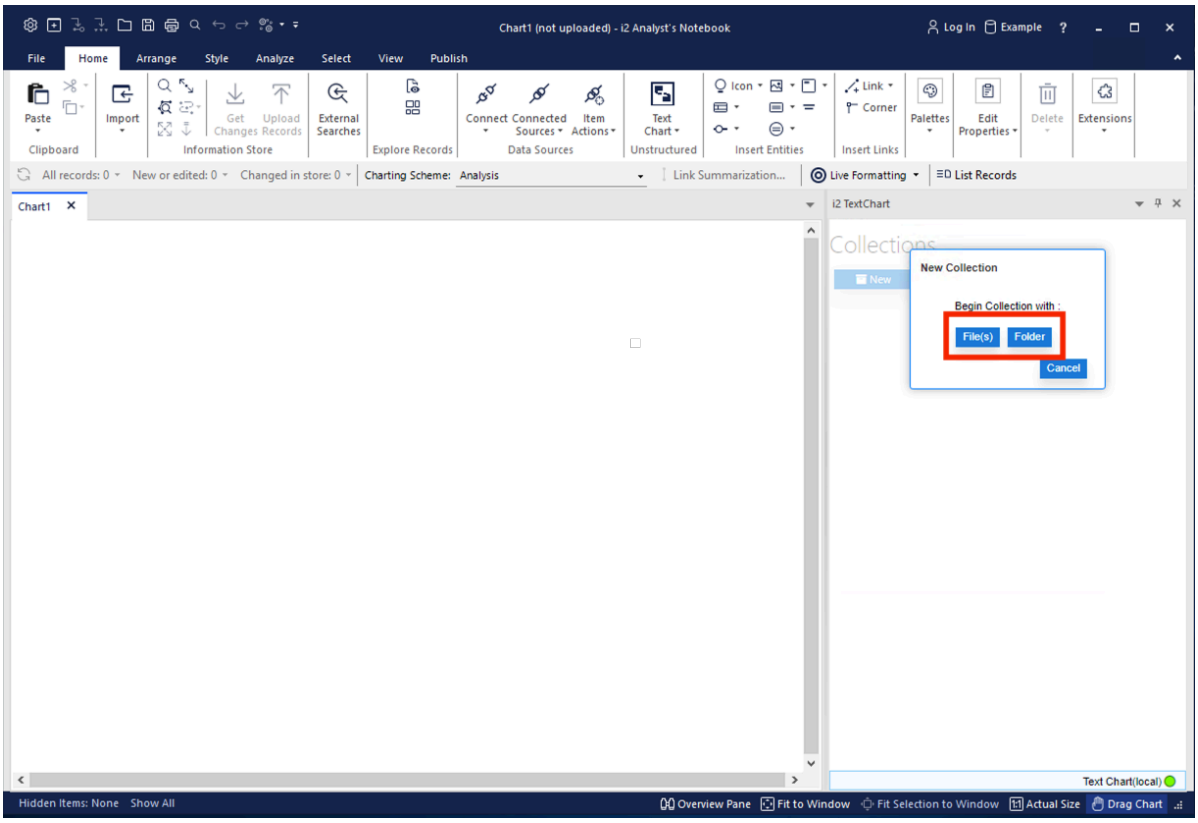
3. If you clicked **New**, provide a name for the new collection and click **Create**.

© N. Harris Computer Corporation



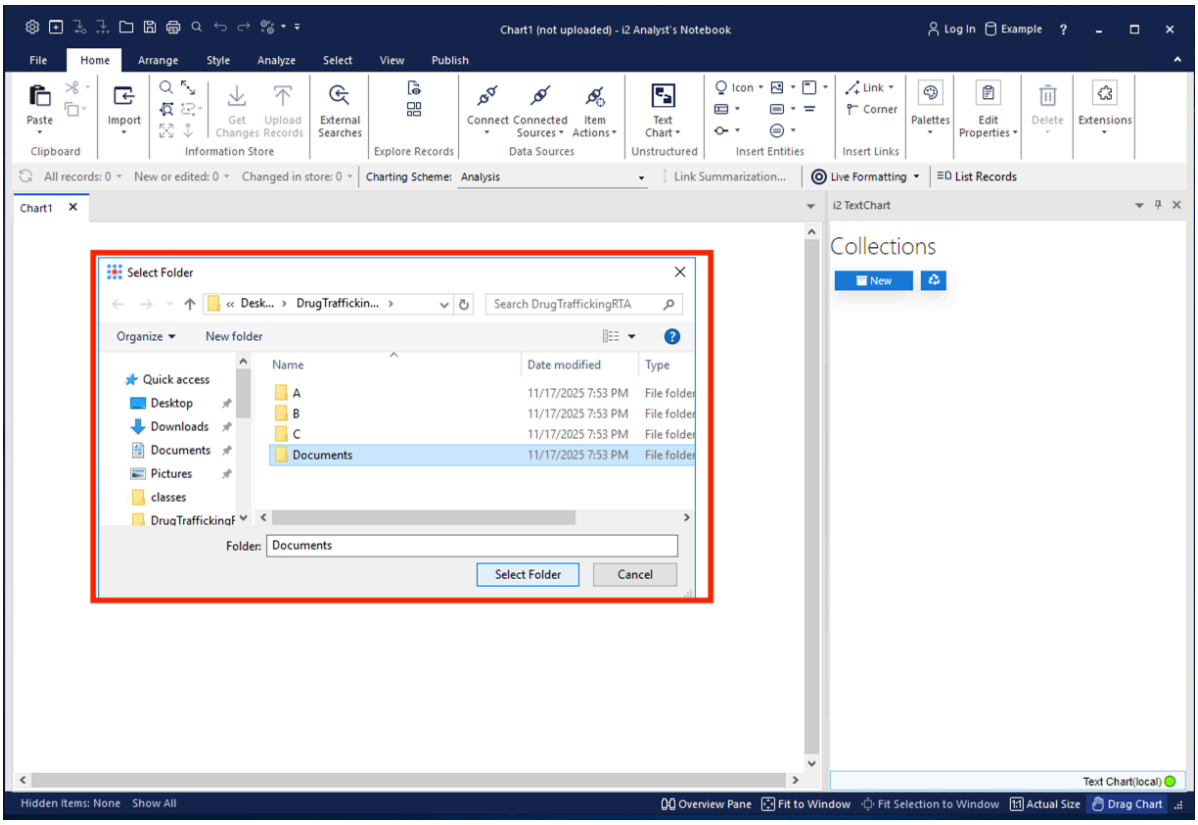
And then specify whether to populate the new collection with files or a folder.

© N. Harris Computer Corporation



4. Finally, select the files (or the folder that contains the files) that you want to import.

© N. Harris Computer Corporation



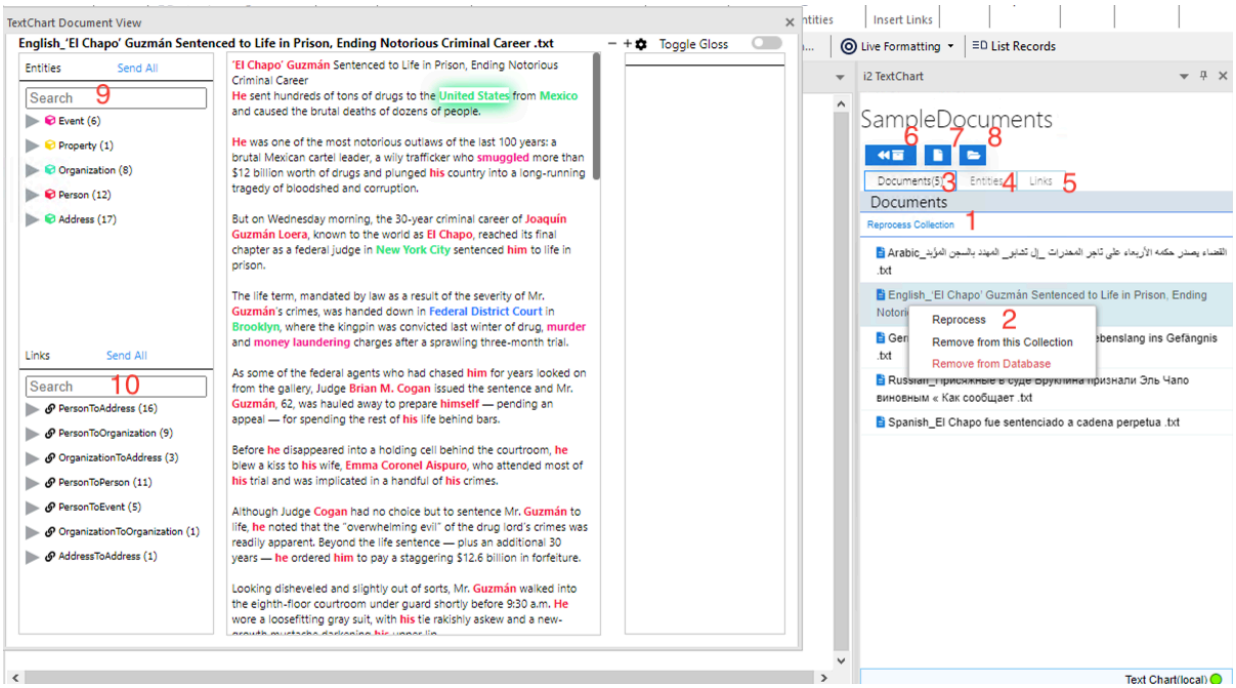
Processing documents

After you import documents into a collection, you can instruct TextChart to process them all at the same time, or to process each document separately.

Note: When you process documents for the first time, it can take a few seconds for the engine to load. Also, larger documents generally take longer to process than shorter ones. TextChart displays a spinning cursor while long-running operations take place.

The documents in a particular collection appear as a list in Analyst's Notebook, from which you can perform a number of tasks. Clicking a single document in the list processes that document and opens it in the Document View. Some of the other tasks are described below.

© N. Harris Computer Corporation



1. **Process Collection**

   When you click **Process Collection**, TextChart processes all the documents in the list that it has not already processed. The progress bar at the top of the pane shows the status of the operation. After each document is processed, its page icon turns blue.

   Canceling the operation stops processing for the current and subsequent documents in the collection. You can also ask TextChart to reprocess *all* documents in the collection, regardless of whether they've been processed before.

2. **Reprocess** / **Remove from this Collection** / **Remove from Database**

   If you right-click a single document in the list, the pop-up menu presents commands to reprocess the document, or to remove it from the collection, or to remove its processing results from the database.

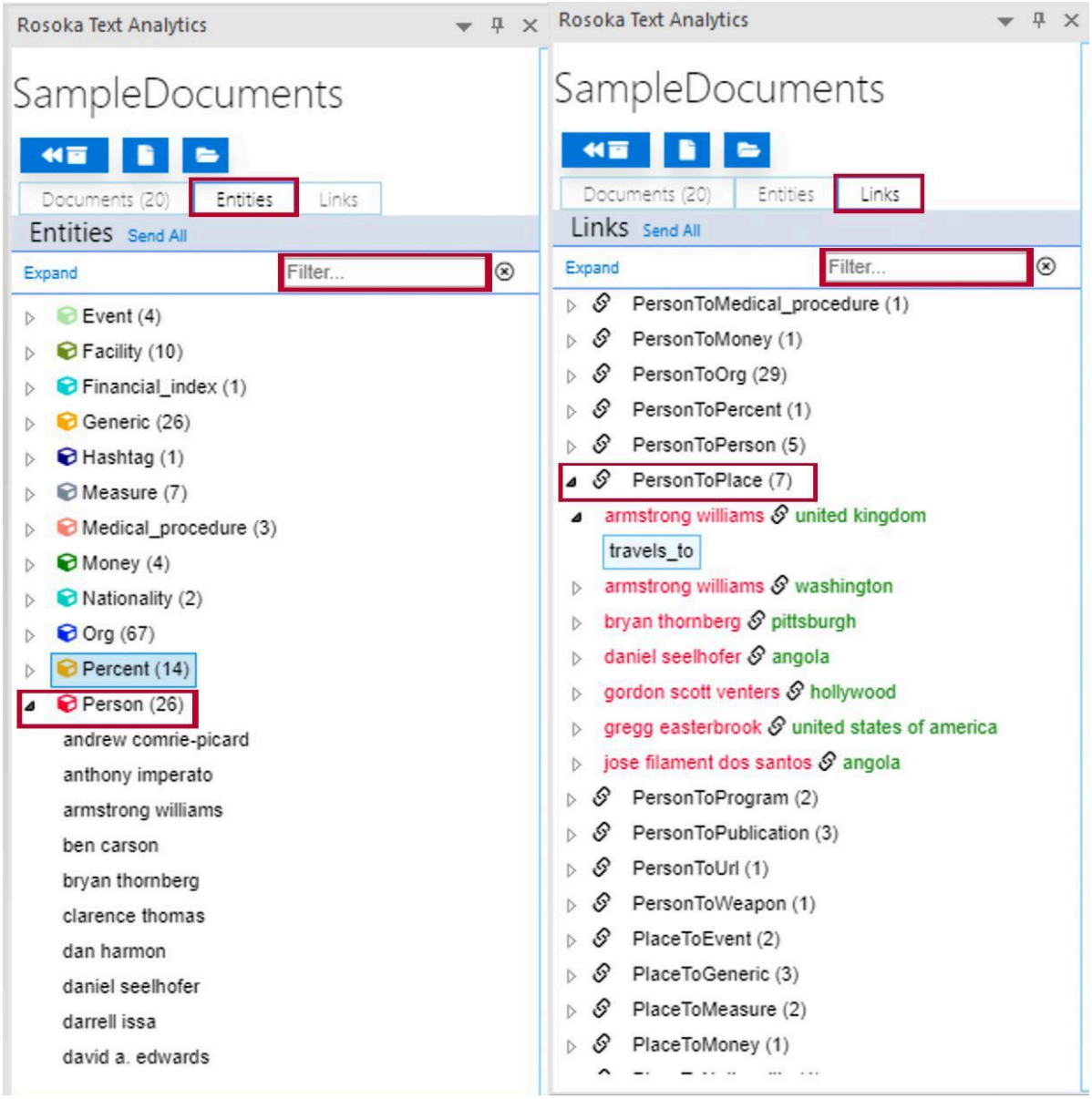
3. **Documents**

   The Documents tab displays a count of the documents in the collection.

4. **Entities**

   The Entities tab contains an entity tree, which lists the TextChart types of all the found entities in processed documents, as well as the count of each type of entity.

© N. Harris Computer Corporation



5. **Links**

   The Links tab contains a link tree that lists the TextChart types of all the found links in processed documents, as well as the count of each type of link.

   You can expand the link types in the tree view to see lists of the individual links and the predicate types that TextChart identified during processing.

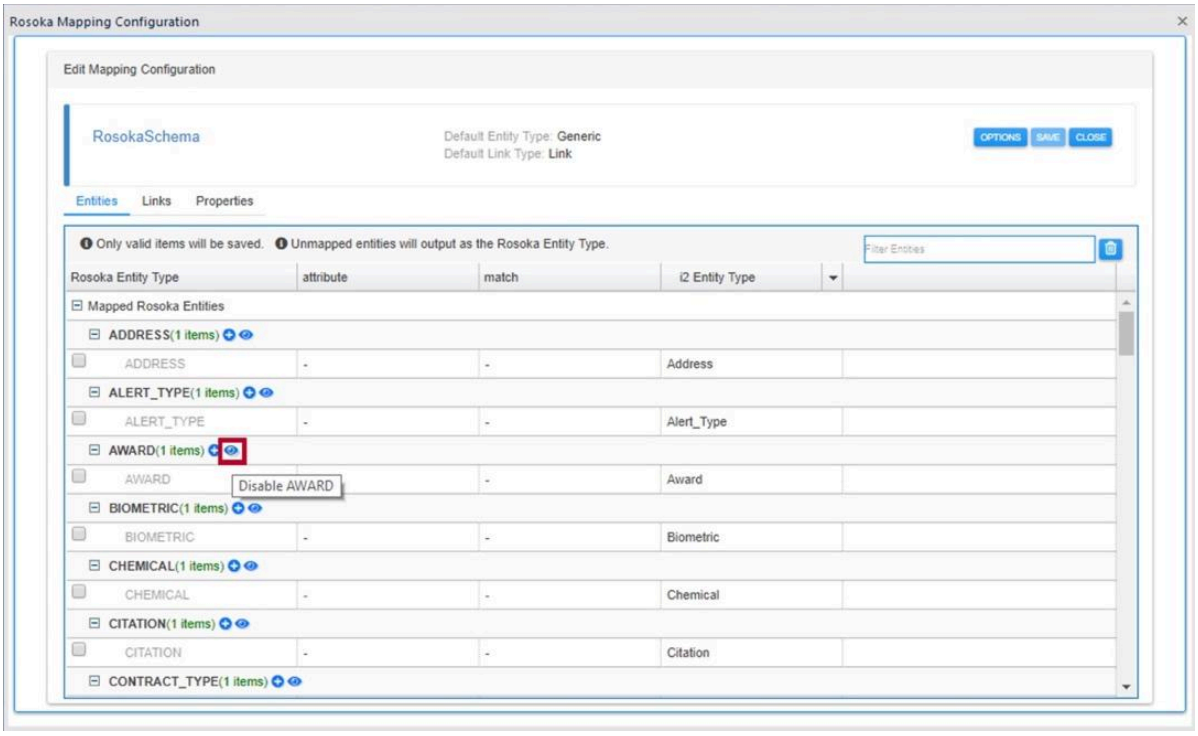
6. Click **Back to Collections** to return to the list of collections.
7. Click **Add More Files** to add more files to the current collection.
8. Click **Add Directory to Collection** to add all the documents in a particular directory to the current collection.

© N. Harris Computer Corporation

Ignoring TextChart entity, link, or attribute types

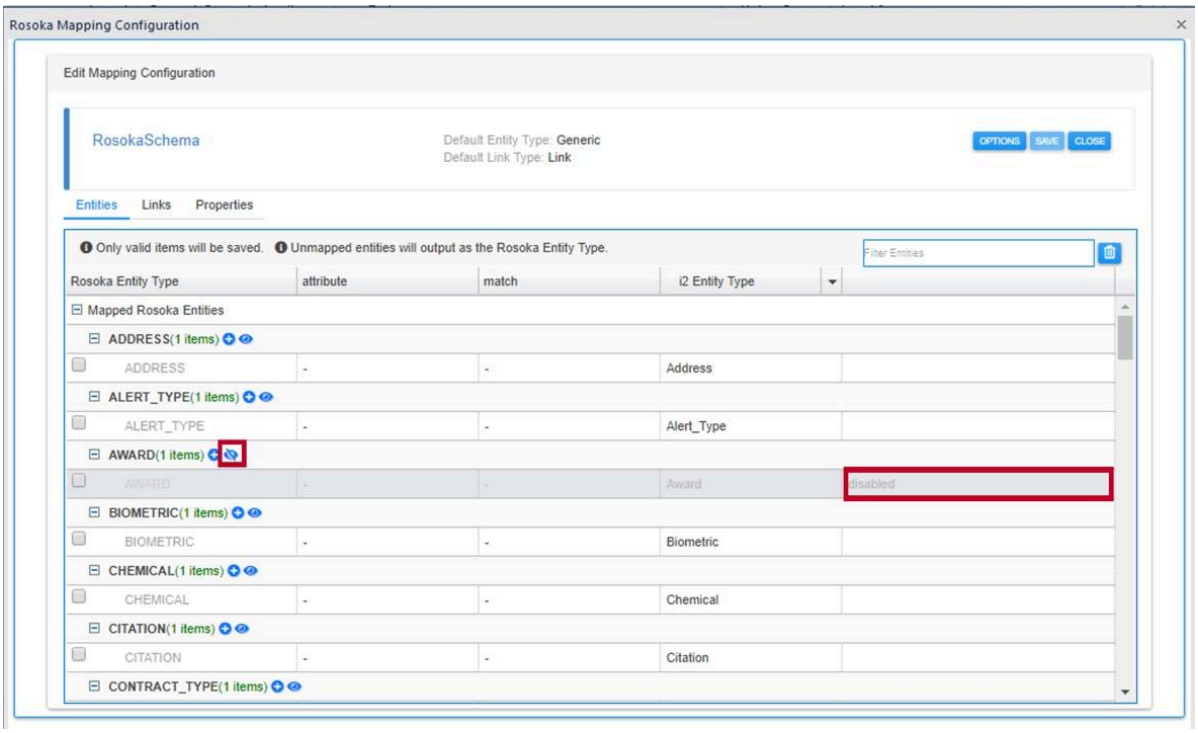
By default, TextChart attempts to find entities and links of all the types that it understands, and values for all the attributes they can contain. If there are some types that you don't want TextChart to extract, you can disable them through the Mapping Configuration.

1. From the Analyst's Notebook ribbon, select **Text Analytics > Mapping Configuration** to open the Mapping Configuration window.
2. Click **Edit** to display the configuration settings for the active mapping. The left column of the **Entities**, **Links**, and **Properties** tabs contains a list of TextChart types.



3. Click the "eyeball" icon next to the name of any type that TextChart should not extract from the documents it processes.

© N. Harris Computer Corporation



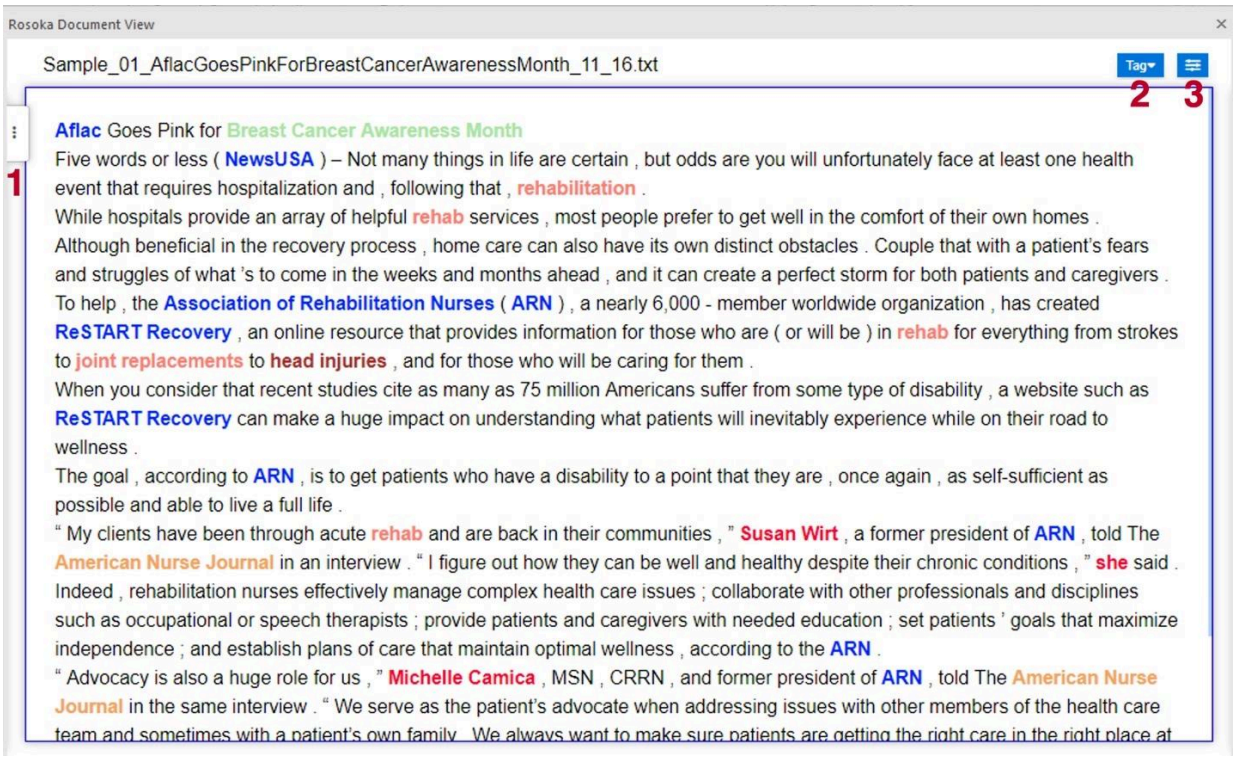
4. If you processed some documents before changing these settings, clear those results and reprocess the documents.

Viewing documents

After TextChart processes a document (or the first of several documents), it opens the Document View window, which presents an enriched view of the processed text.

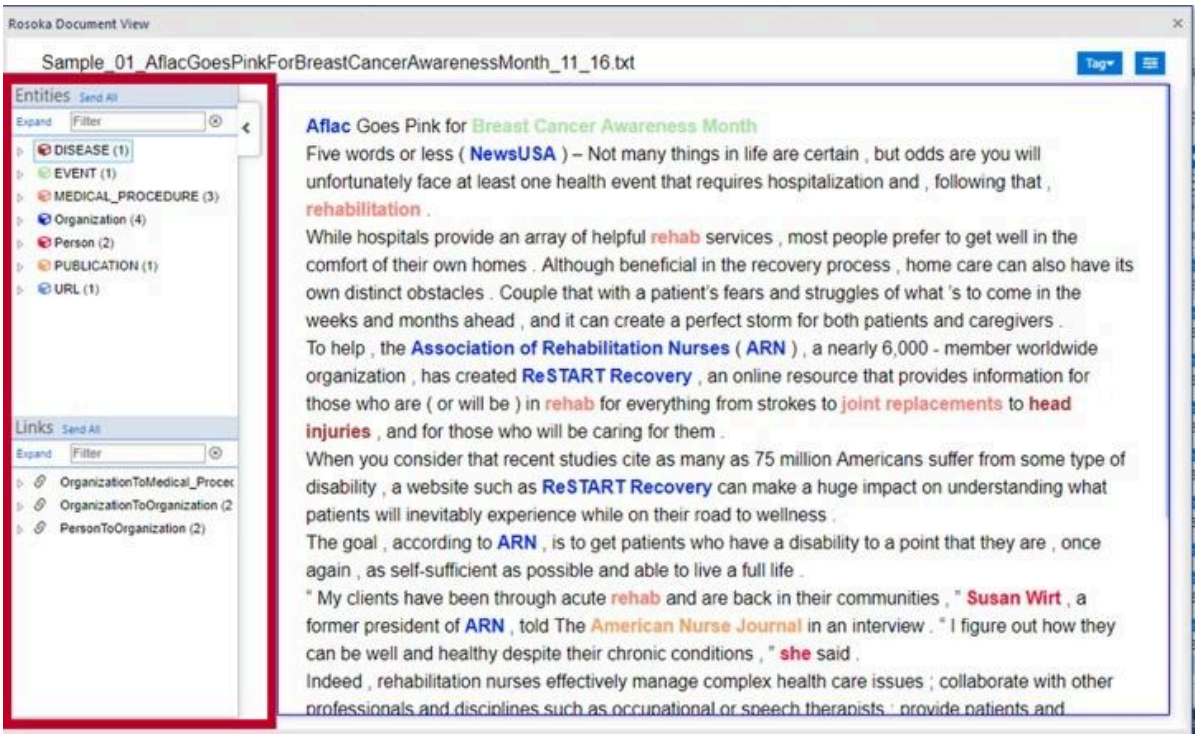
The Document View highlights all of the found entities in context, using different colors for different types. In this view, you can review and curate the entities that TextChart identified.

© N. Harris Computer Corporation



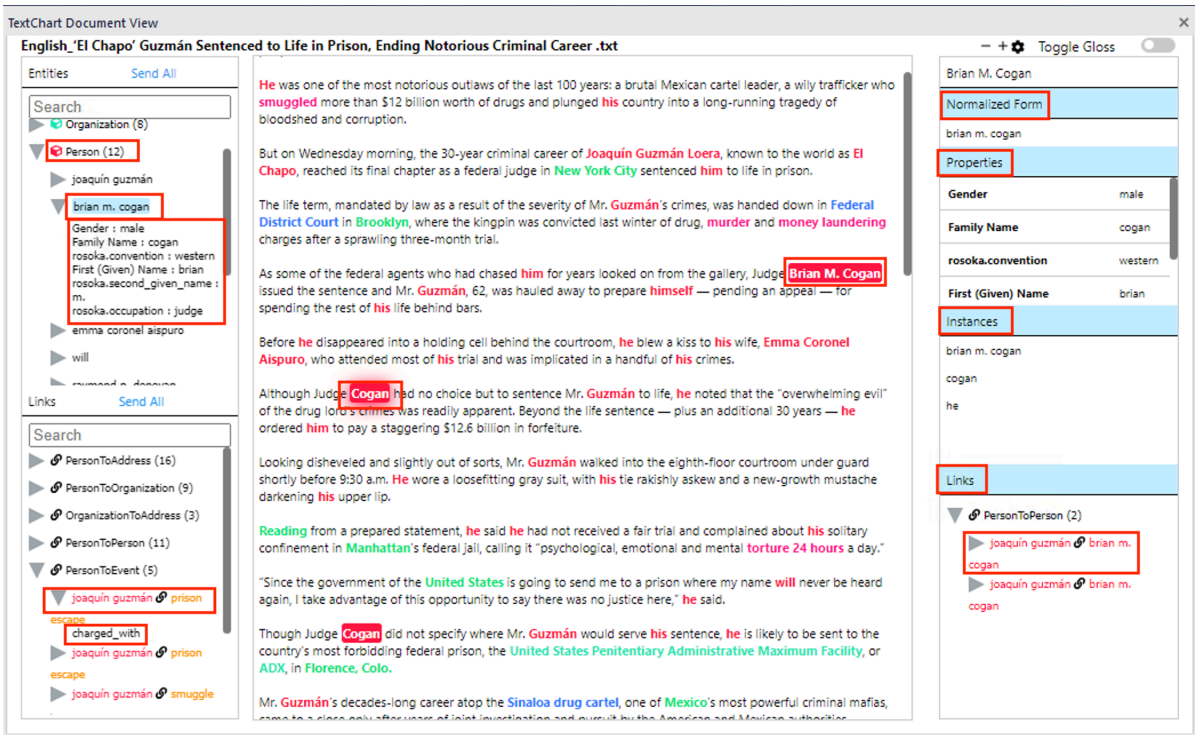
1. The button at the upper left of the window opens entity and link trees for the current document.

© N. Harris Computer Corporation



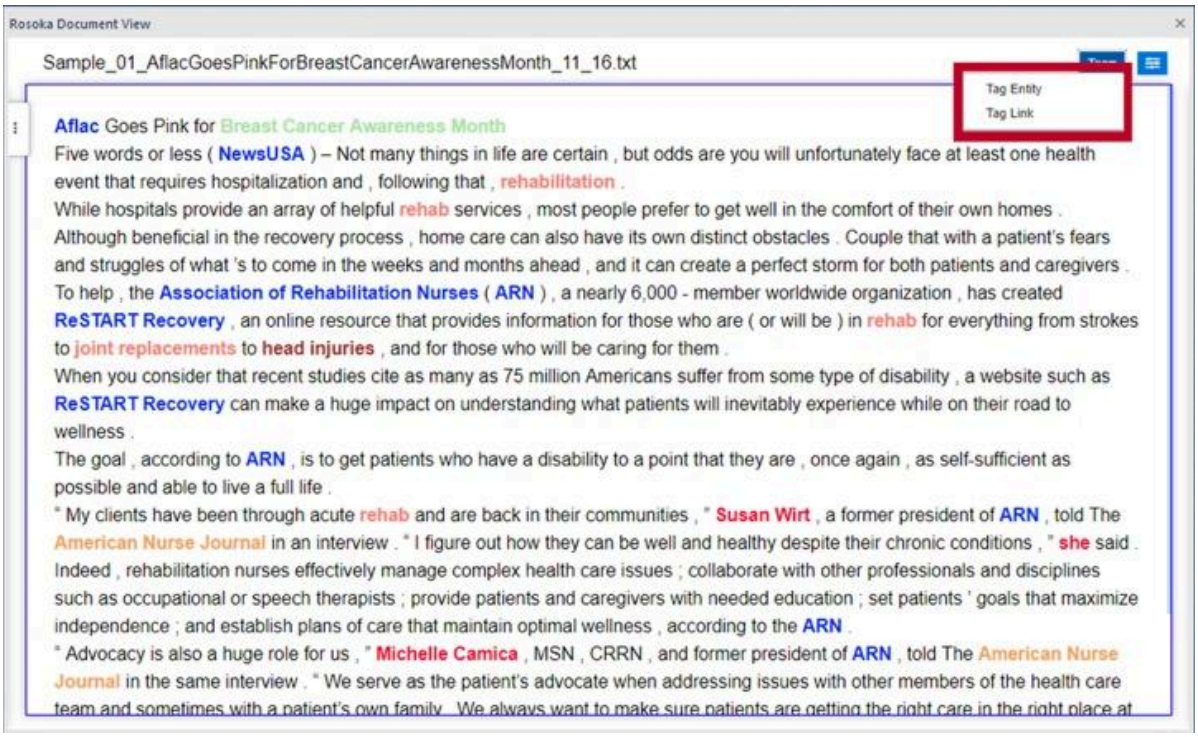
The trees behave initially like the ones in the list for all the documents in a collection. However, when you select entities and links (or groups of them), TextChart highlights them in the open document and displays a second panel that contains more information.

© N. Harris Computer Corporation



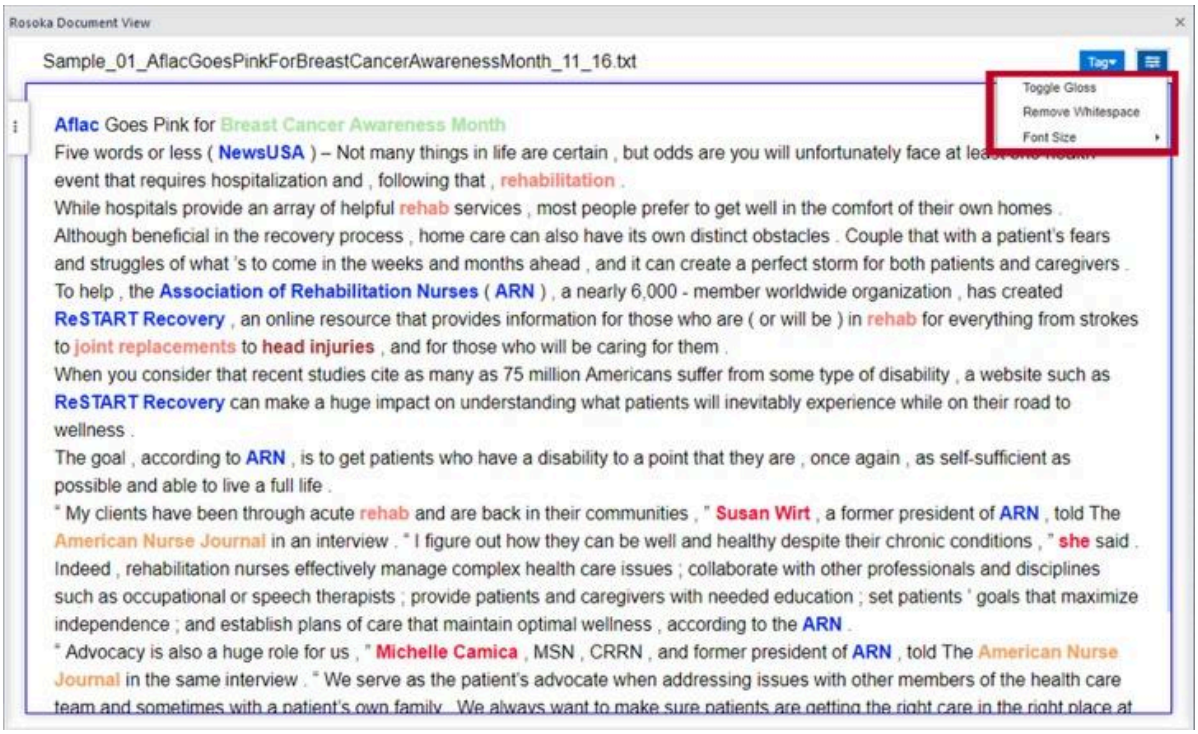
2. Use the **Tag** drop-down to tag a new entity or link.

© N. Harris Computer Corporation



3. Through the **Options** button, you can:
   a. Toggle back and forth between an English version of the document, and the text in its original language.
   b. Remove whitespace from the document as displayed
   c. Change the font size of the text in the document.

© N. Harris Computer Corporation



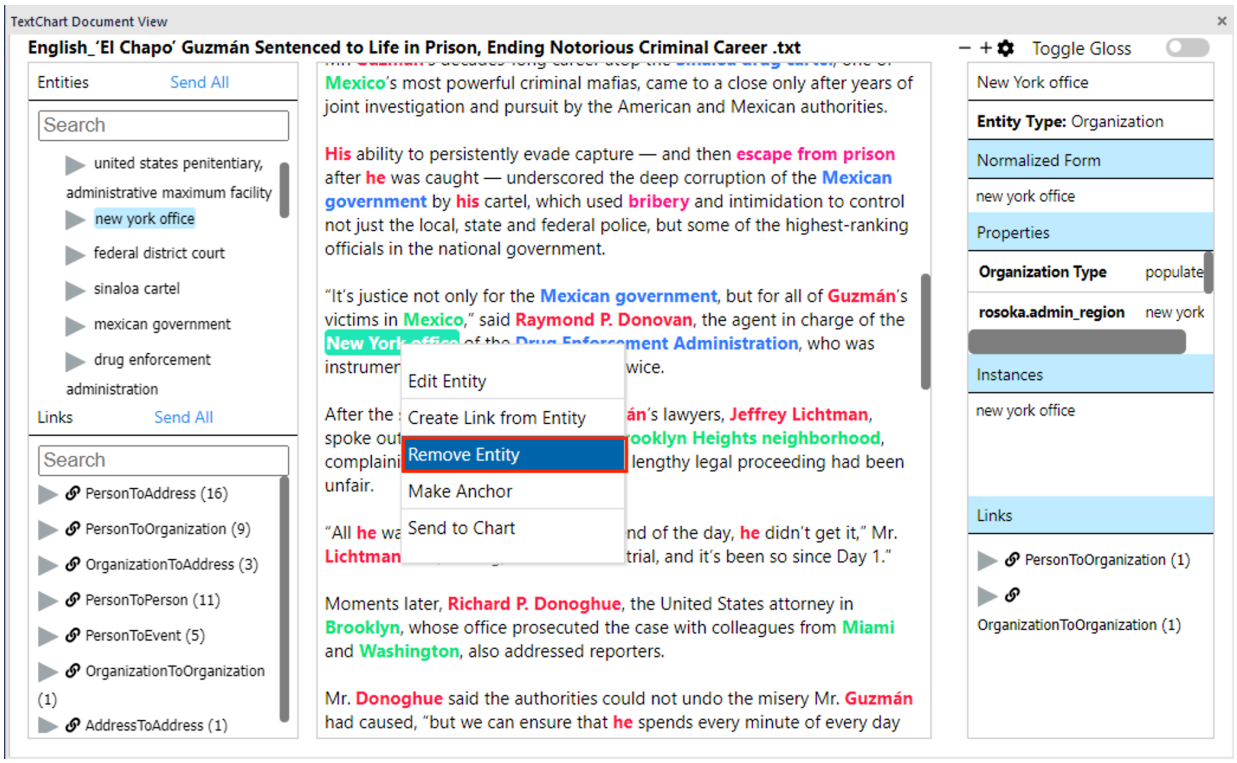
Curating extraction results

After i2 TextChart presents the results of processing a document in the Document View, you can review its output and make changes to the results. TextChart supports removing and modifying the records it finds, as well as identifying new records of your own.

Removing results

If TextChart identifies an entity or a link that you don't want to appear in the results, you can remove it from the Document View. For example, the following view contains a contrived entity extraction result for an "online resource", which has the Generic entity type.

© N. Harris Computer Corporation

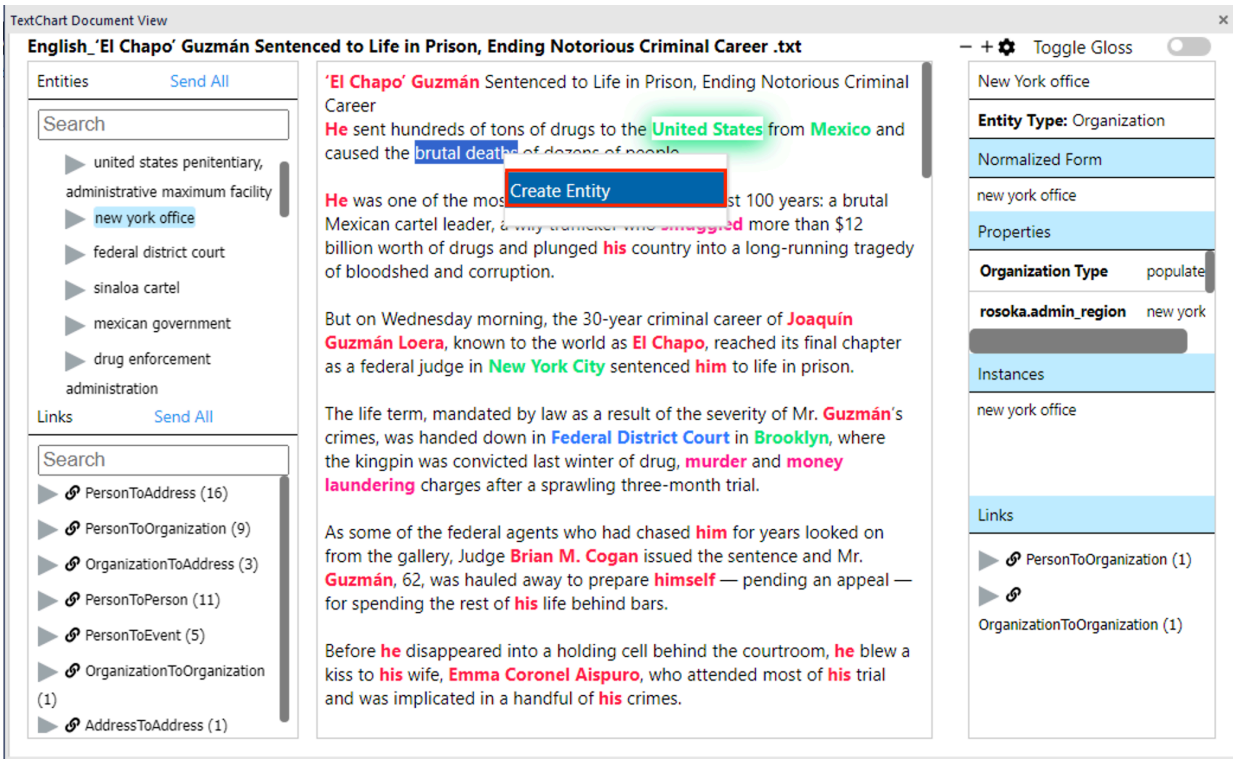


When you select the text in the Document View, TextChart displays information about the extracted entity. To remove this result, click **Remove Entity**.

Adding entities

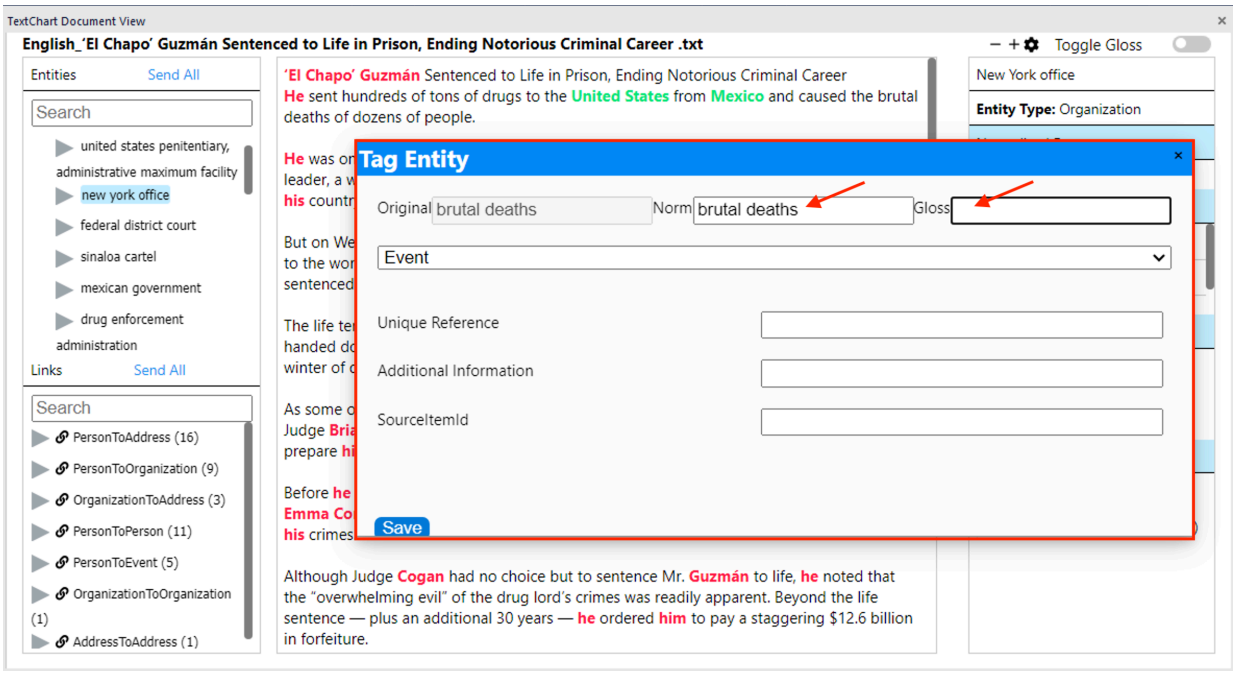
If TextChart failed to identify an entity that you *do* want to appear in the results, you can highlight the text to be extracted, and then select **Tag > Tag Entity**.

© N. Harris Computer Corporation



TextChart displays the Tag Entity Tool for you to provide information about the entity, and populates the **Original** field with the text that you highlighted.

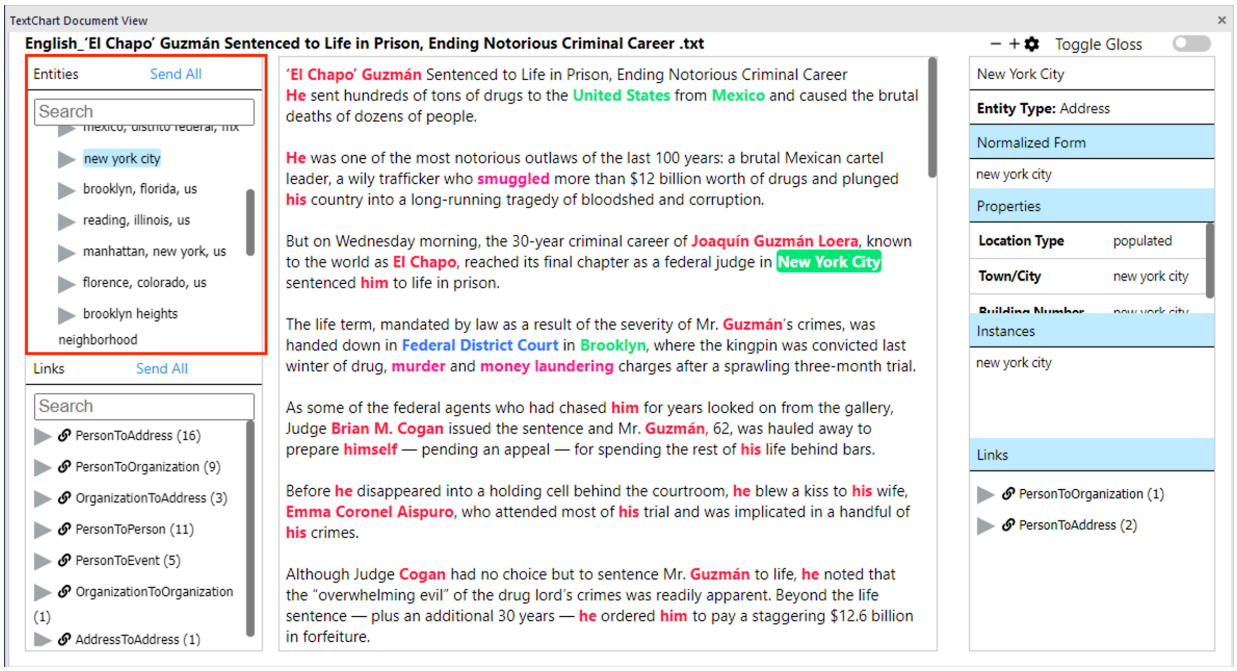
© N. Harris Computer Corporation



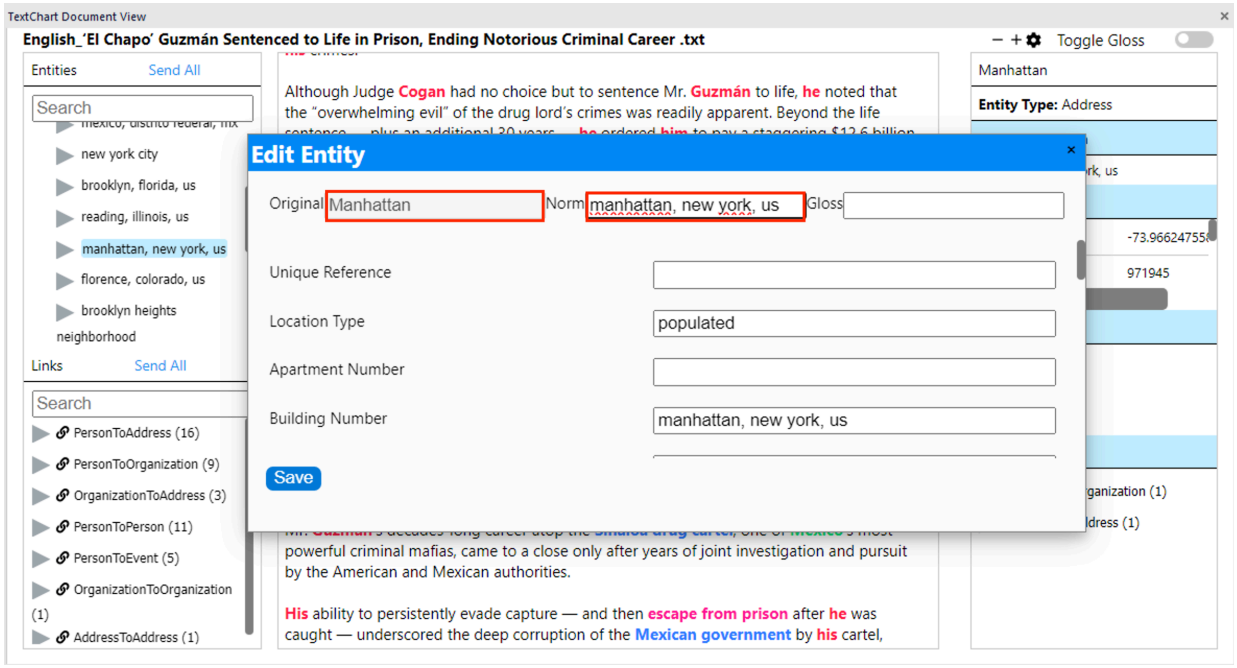
TextChart also populates the **Norm** field with the same text, but here you have the option of changing the text to normalize it when a document uses different terminology to refer the same piece of information.

For example, in the image below, "England, GB" and "United Kingdom, GB" have been identified as separate places.

© N. Harris Computer Corporation



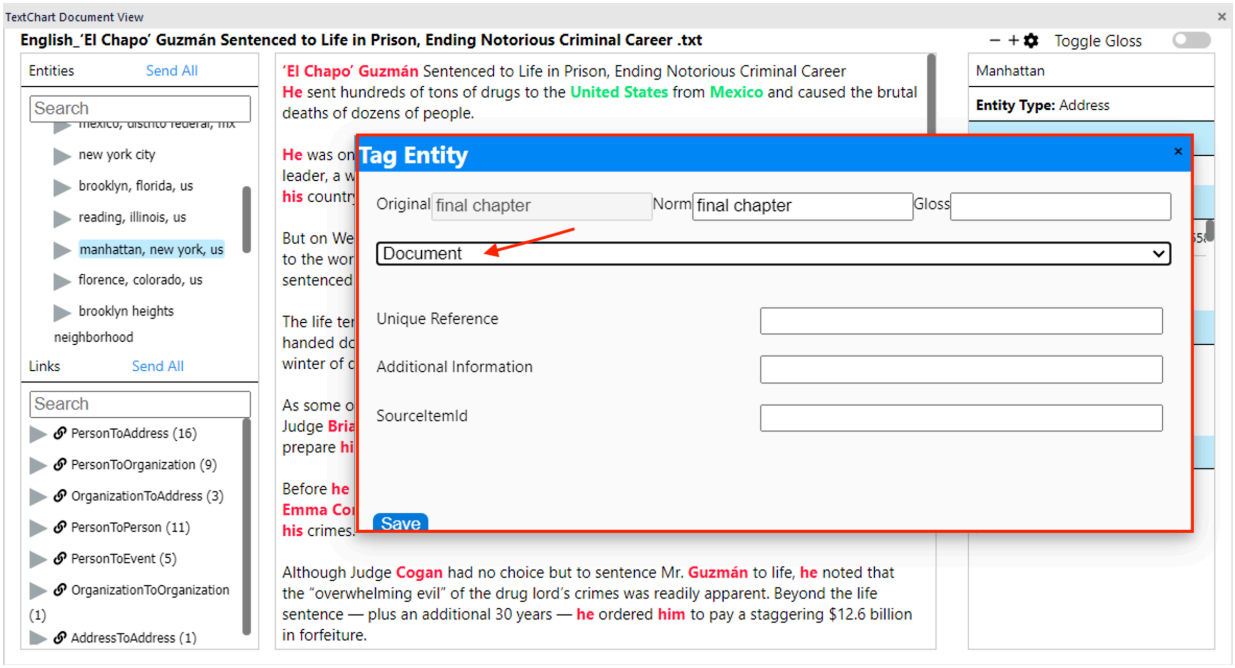
If you decide that TextChart should treat these instances as the same place, you can edit the **Norm** field of one so that it matches the other.



© N. Harris Computer Corporation

If you're processing documents in a language other than English, and you come across an important term that TextChart does not understand, you can add an English translation of that term to the **Gloss** field.

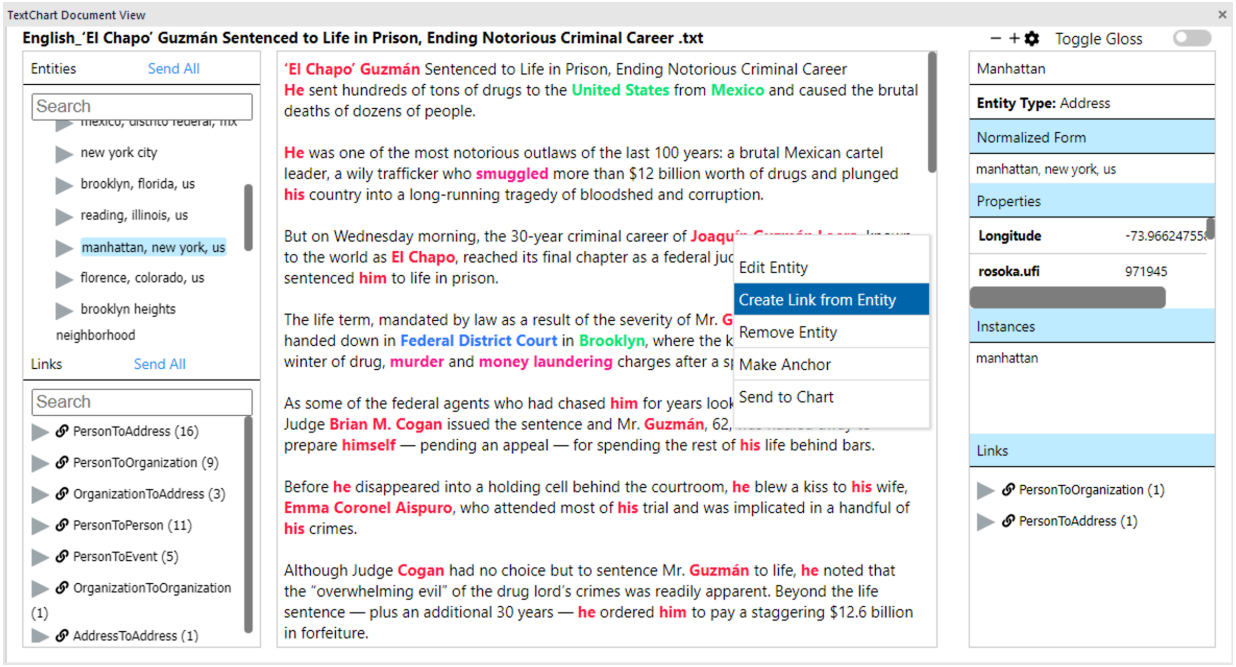
Finally, you must choose a TextChart type for your new entity. Different entity types have different attributes, and you can fill in additional information as you see fit.



Adding links

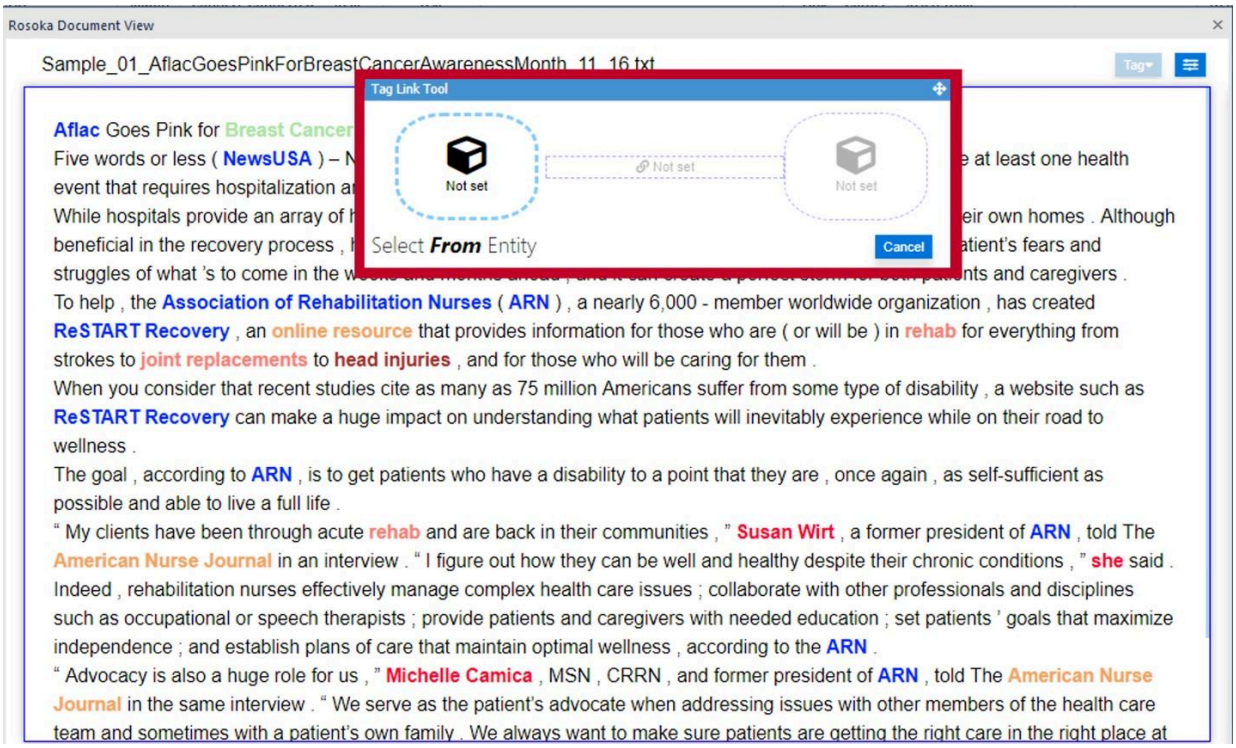
Adding a link that TextChart failed to identify is similar to adding an entity in the same circumstances. You can select **Tag > Tag Link** from the Document View header, or right-click an extracted entity to use it as the "From" end of a new link.

© N. Harris Computer Corporation



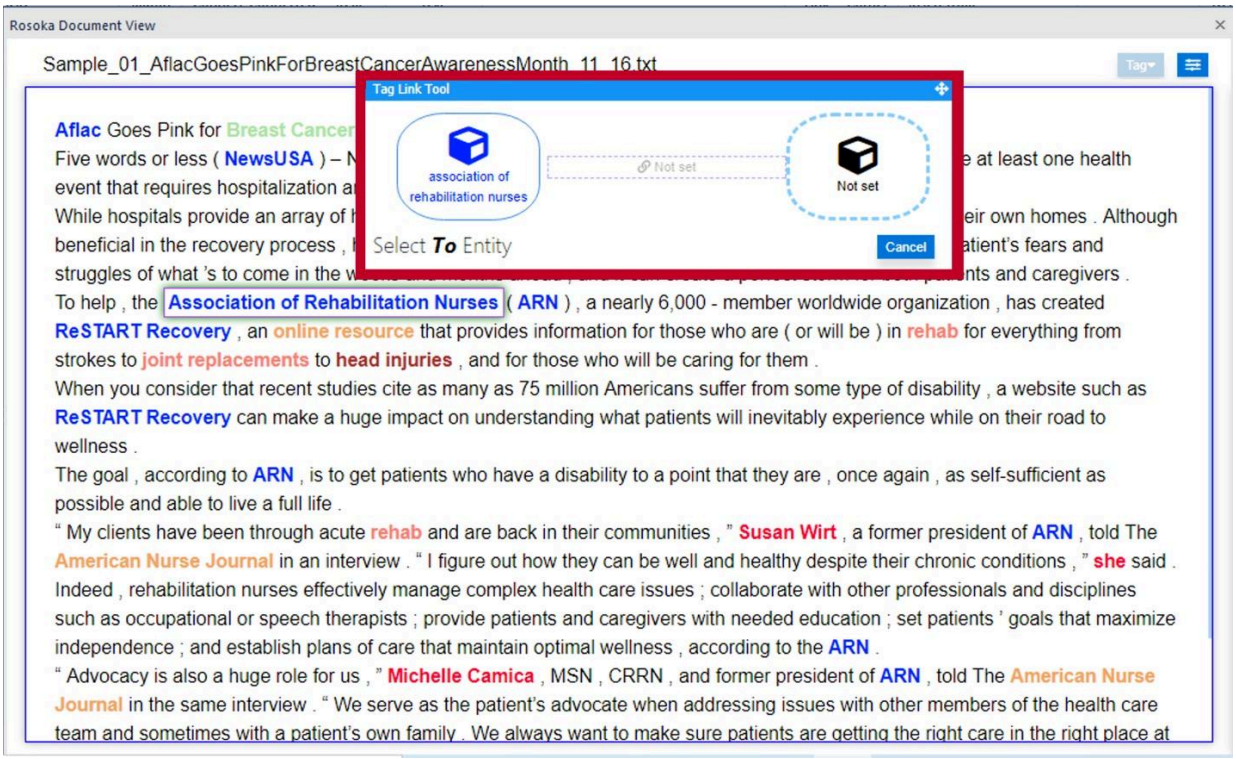
If you used the **Tag Link** command, the Tag Link Tool starts by prompting you to select the entity at the "From" end.

© N. Harris Computer Corporation



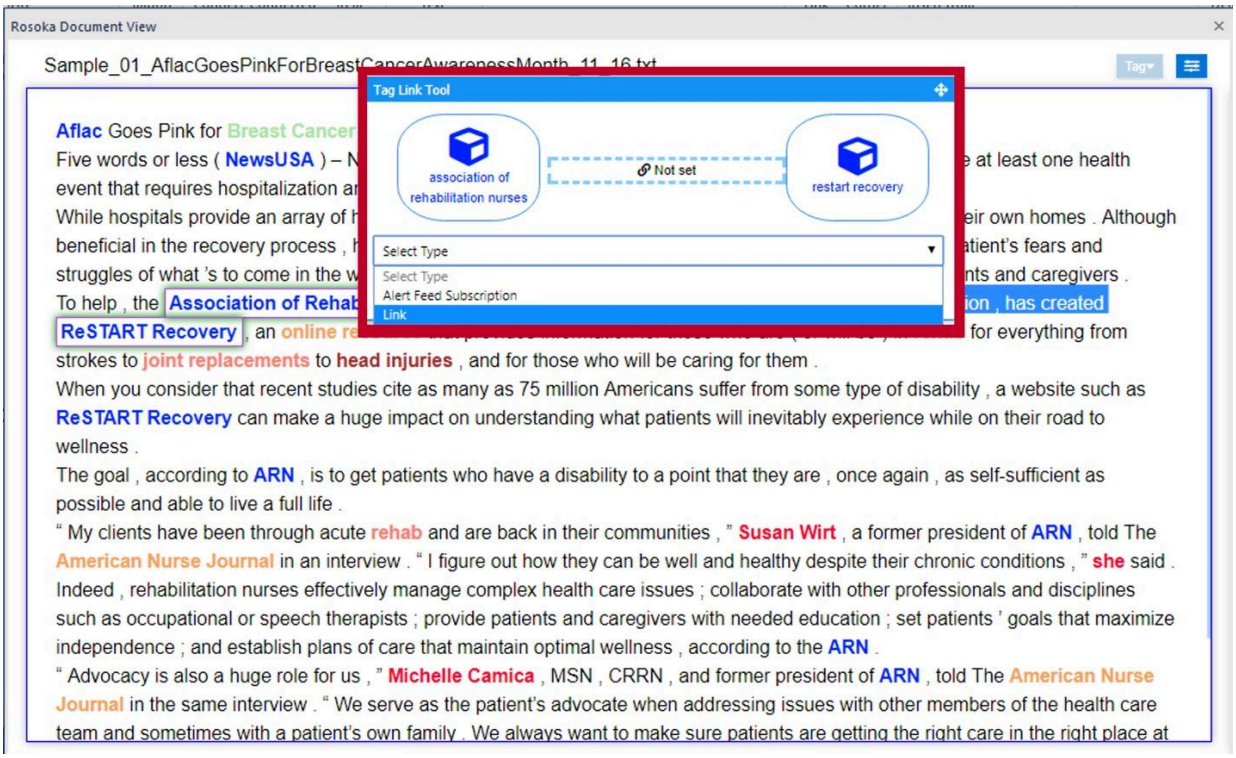
Then, you have to select the entity at the "To" end.

© N. Harris Computer Corporation



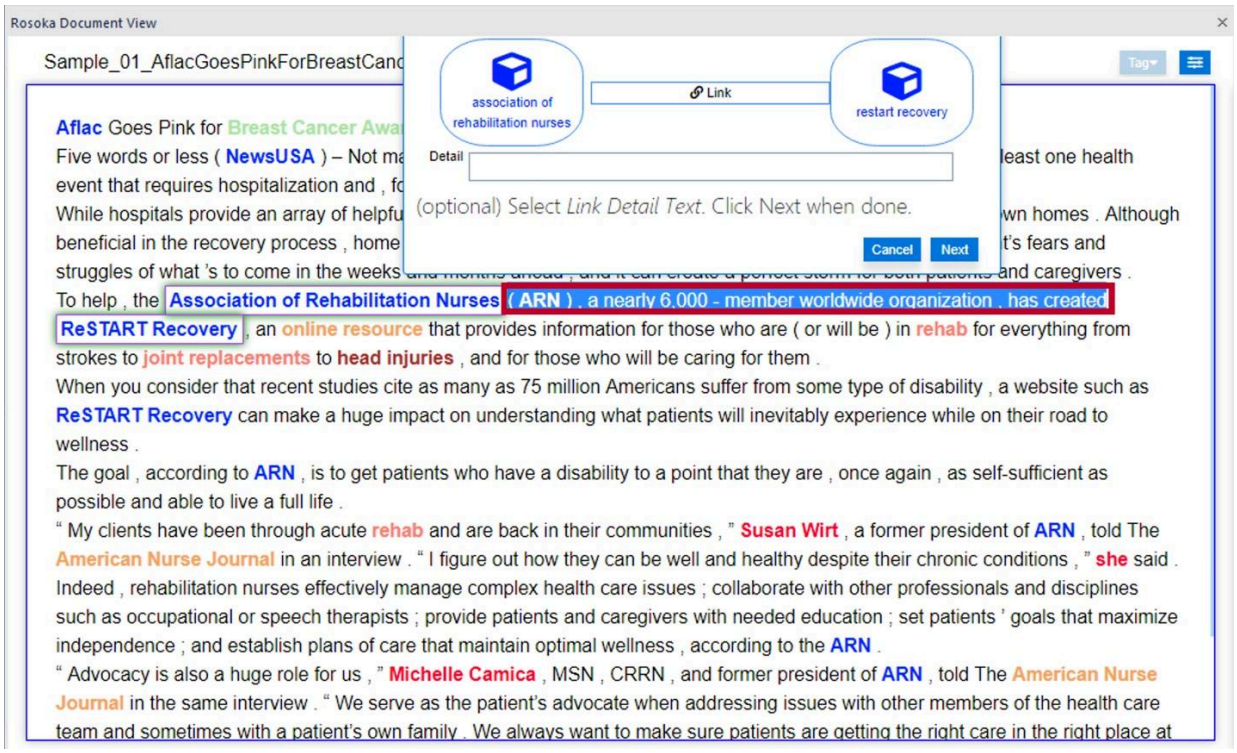
After you select both ends, the Tag Link Tool displays a drop-down that allows you to specify the TextChart type of the link to be created.

© N. Harris Computer Corporation



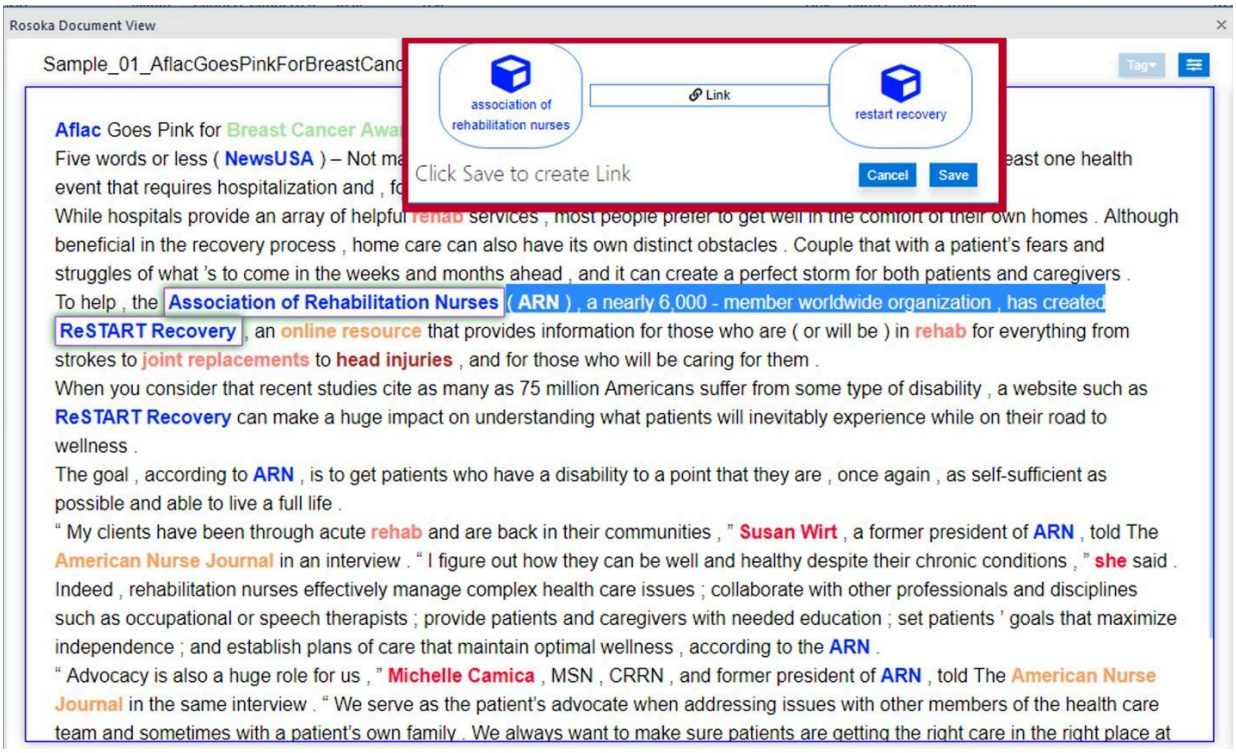
You can also provide the *predicate* (that is, the supporting text) that prompted you to create the relationship between the two entities.

© N. Harris Computer Corporation



When all the information in the Tag Link Tool is complete, click **Save** to add the new link to the results.

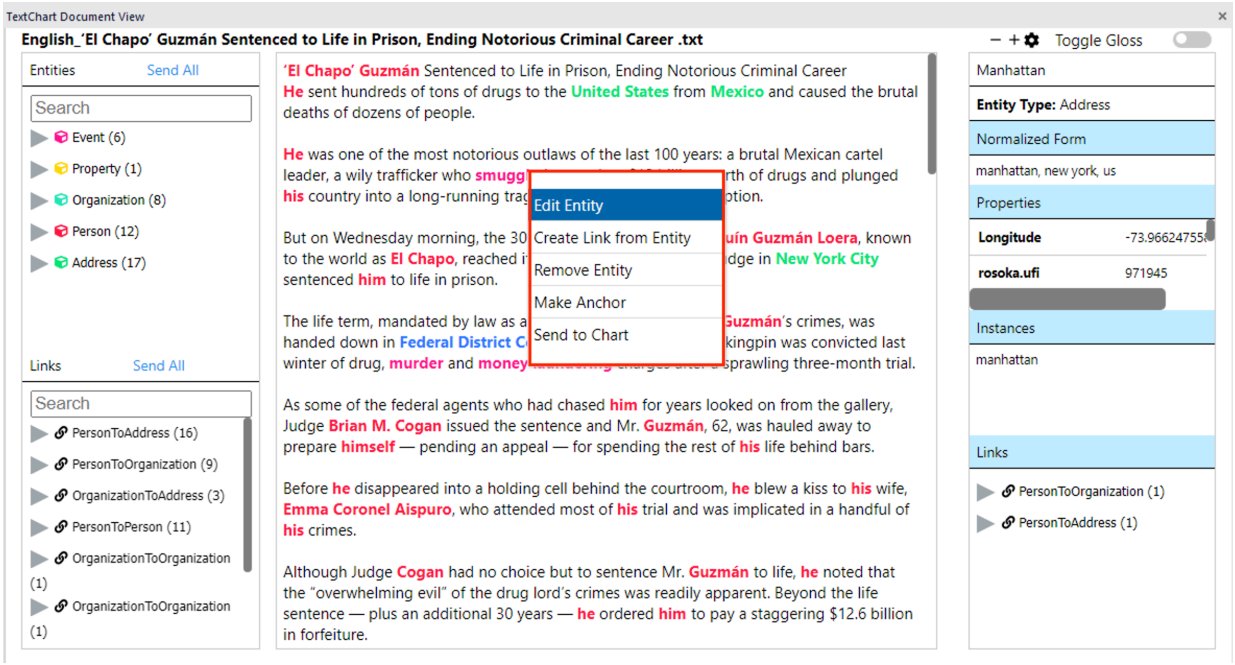
© N. Harris Computer Corporation



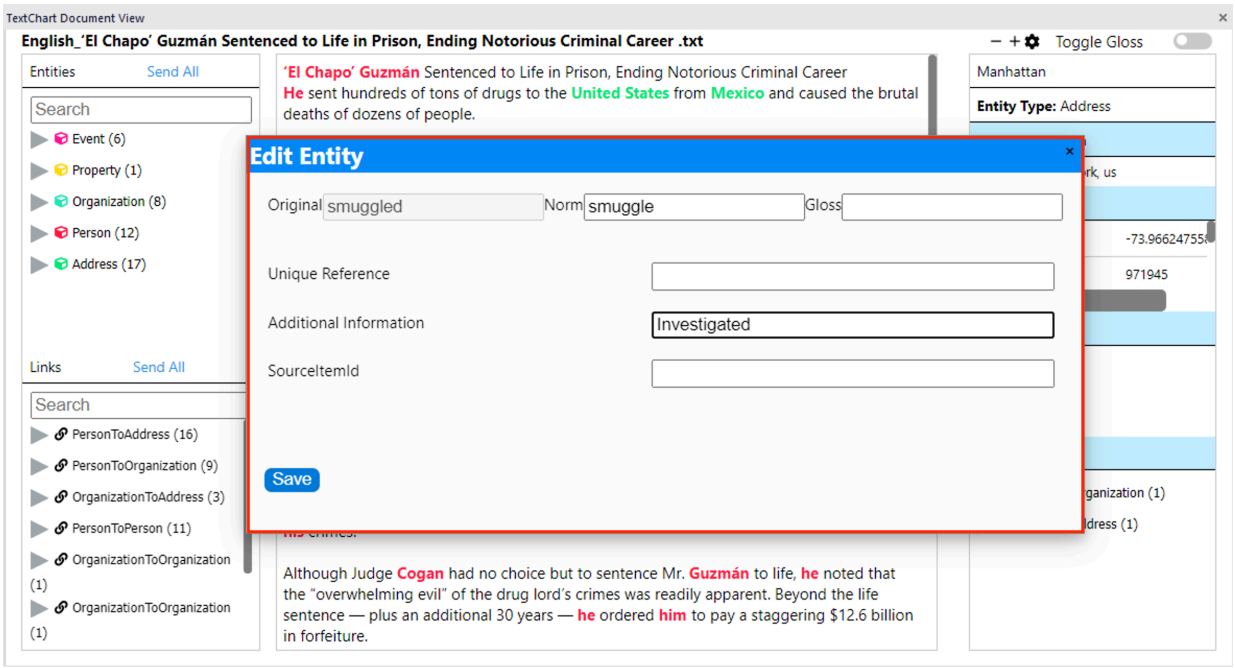
Editing results

If TextChart identified an entity or a link successfully, but the extracted result isn't exactly how you want it, you can right-click the highlighted text in the Document View and select **Edit**.

© N. Harris Computer Corporation

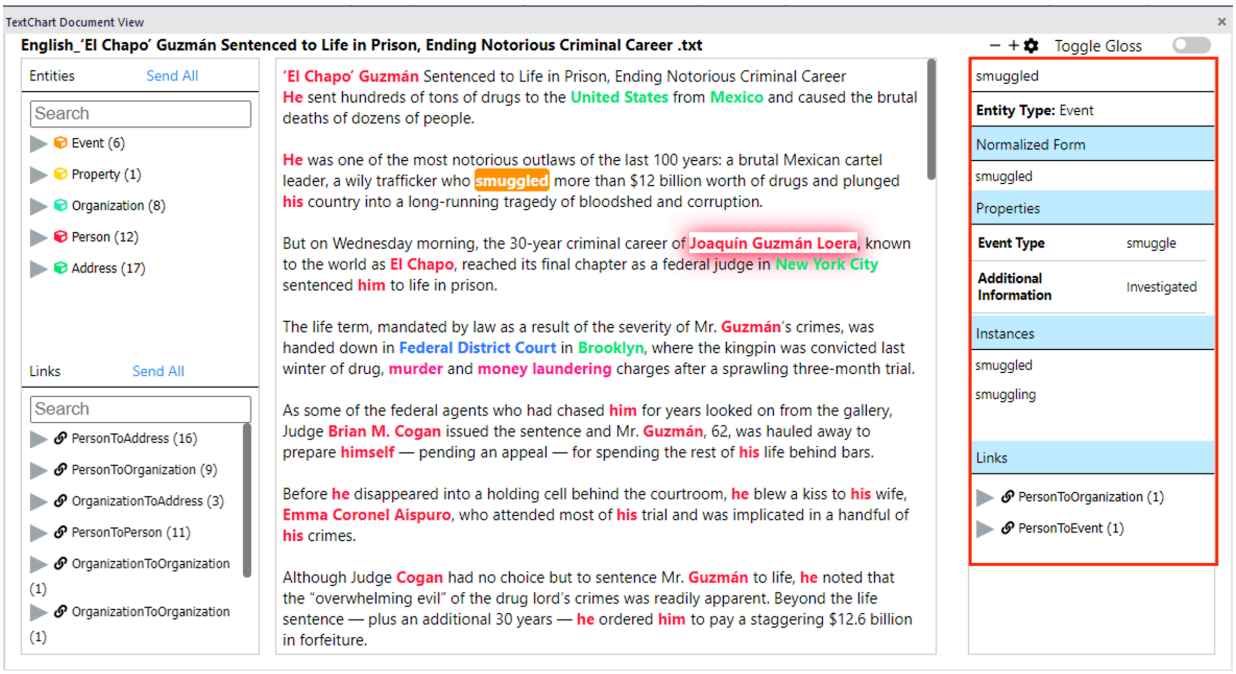


The behavior of the tool is the same when you're editing results as it is when you're adding them. You can edit the **Norm** field, add an English **Gloss**, change the **Type**, and modify attribute information.



© N. Harris Computer Corporation

When you finish editing, the changes you made are reflected in the feature bar on the right of the Document View.



Important: After you edit an extracted entity or link, you must reprocess the other documents in the collection so that they receive the same modifications.

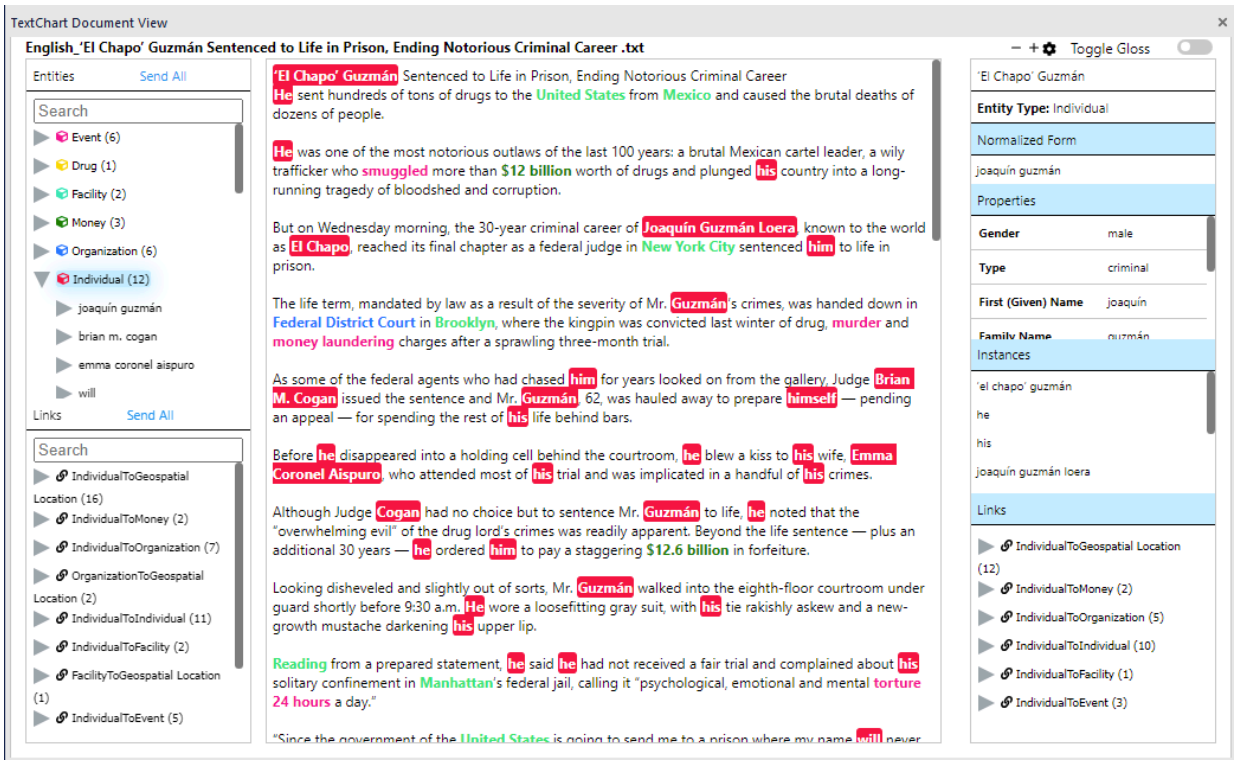
Analyzing extraction results

After i2 TextChart processes a document, it presents its results in the Document View for you to assess. The views that it presents for extracted entities and extracted links are slightly different from each other.

Viewing entities

When you select an entity type in the **Entities** list, TextChart highlights the entities of that type in the document.

© N. Harris Computer Corporation

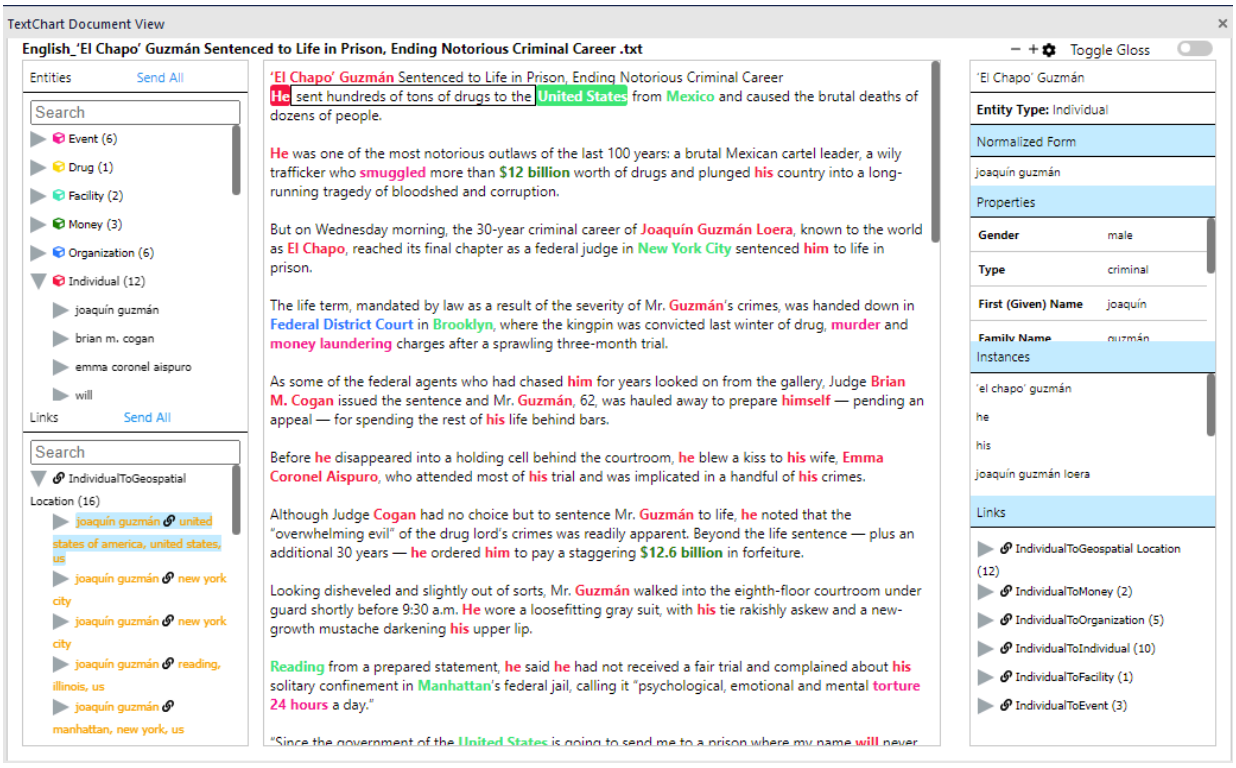


The list also displays the number of entities of each type in the document. If you expand an entity type and select a particular instance, TextChart opens the entity modification window on the right of the Document View.

Viewing links

When you select a link type in the **Links** list, TextChart highlights the entities and the predicates that contribute to links of that type.

© N. Harris Computer Corporation



As with the extracted entities, you can expand link types in the list to investigate and modify the information of any individual extracted link.

Charting Extraction Results

After i2 TextChart has imported and processed a collection of documents and you have curated the results, you can add the extracted entities and links to an Analyst's Notebook chart. TextChart supports sending results to the chart separately or in bulk.

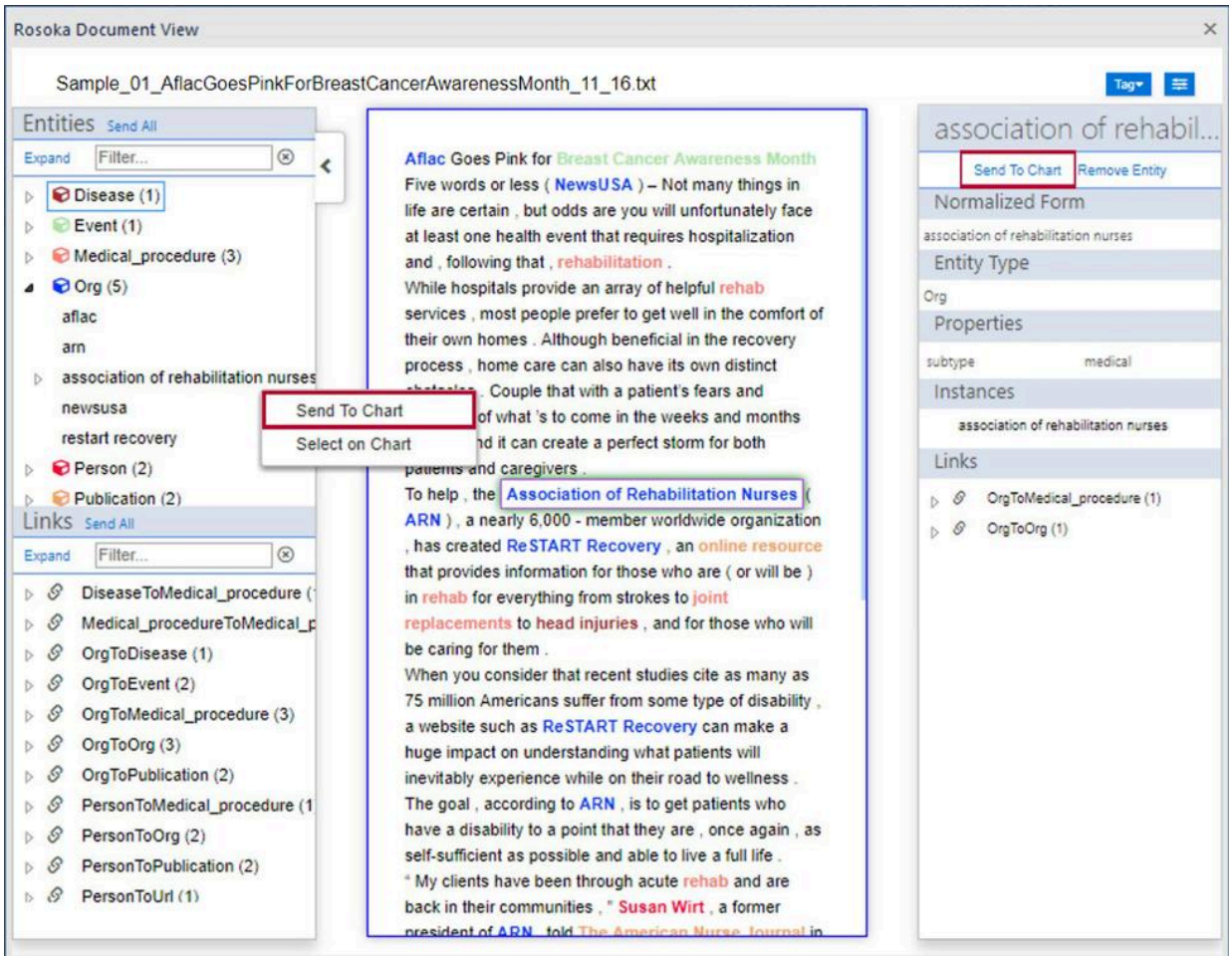
Sending Entities To A Chart

TextChart provides three options for sending entities to a chart. You can specify them individually, choose all entities of the same type, or simply send all the extracted entities at once.

Sending a single entity

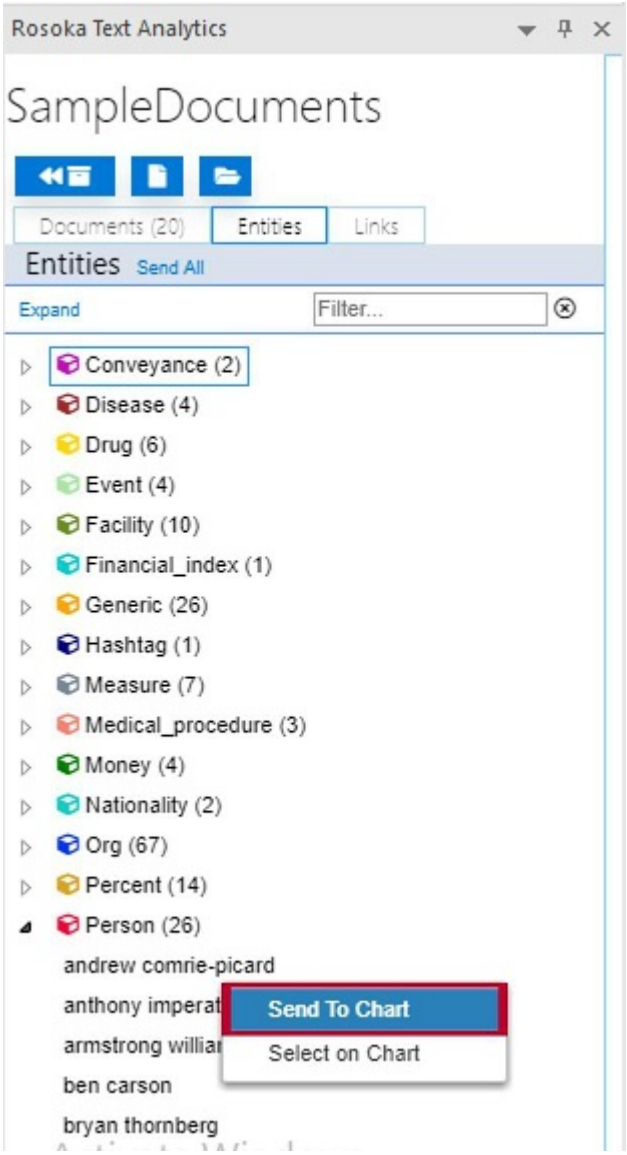
To send a single entity from the Document View to the current chart, you can use the pop-up menu for that entity in the **Entities** list on the left of the window, or the **Send To Chart** button from the entity details pane on the right of the window.

© N. Harris Computer Corporation



If an extracted entity is already on the chart, you can use the same pop-up menu to select it there.

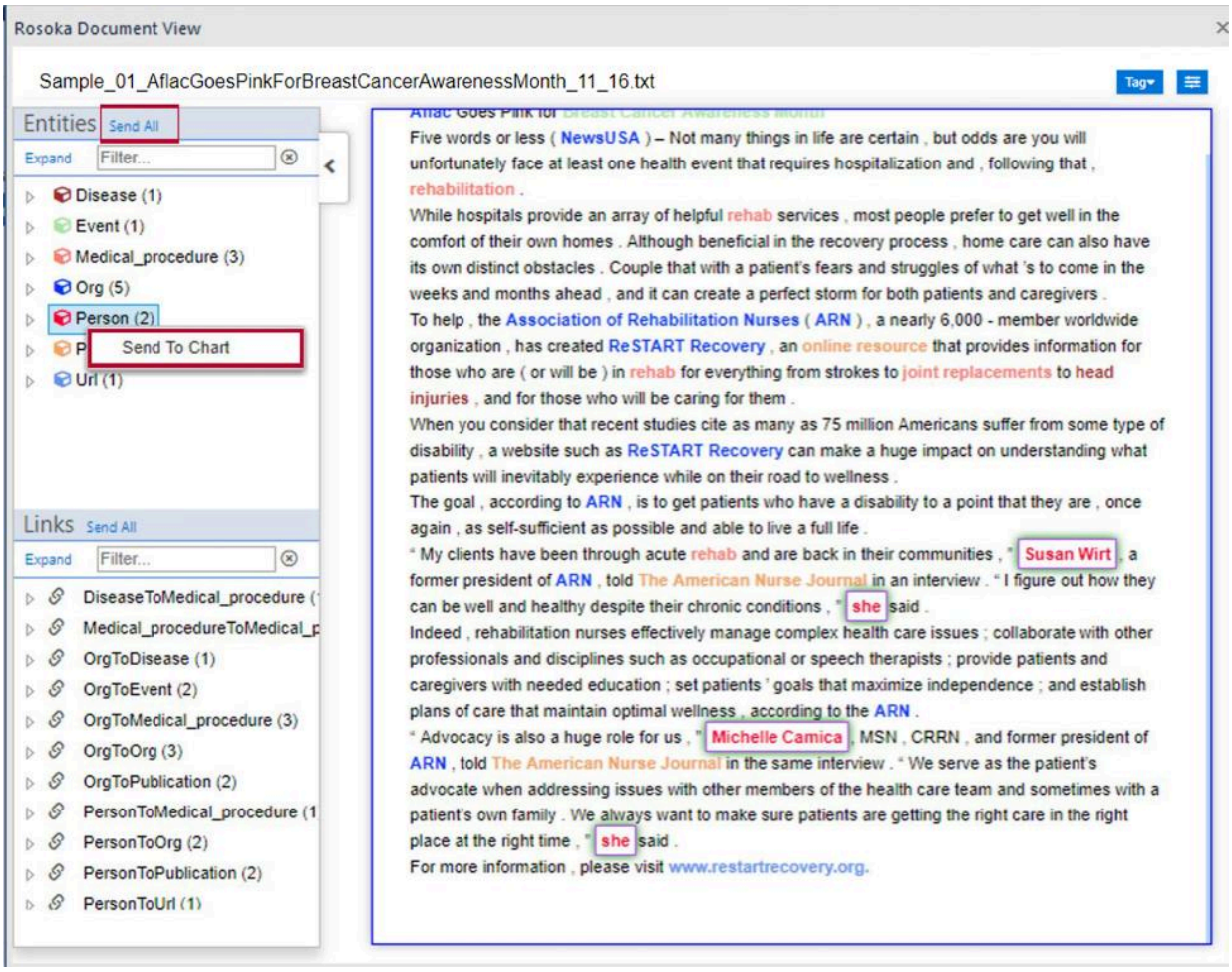
© N. Harris Computer Corporation



Sending several entities

To send all the entities of the same type from the Document View to a chart, you can use the pop-up menu for that entity type in the **Entities** list. Alternatively, to send *all* the entities extracted from the document to the chart, you can click the **Send All** button at the top of the list.

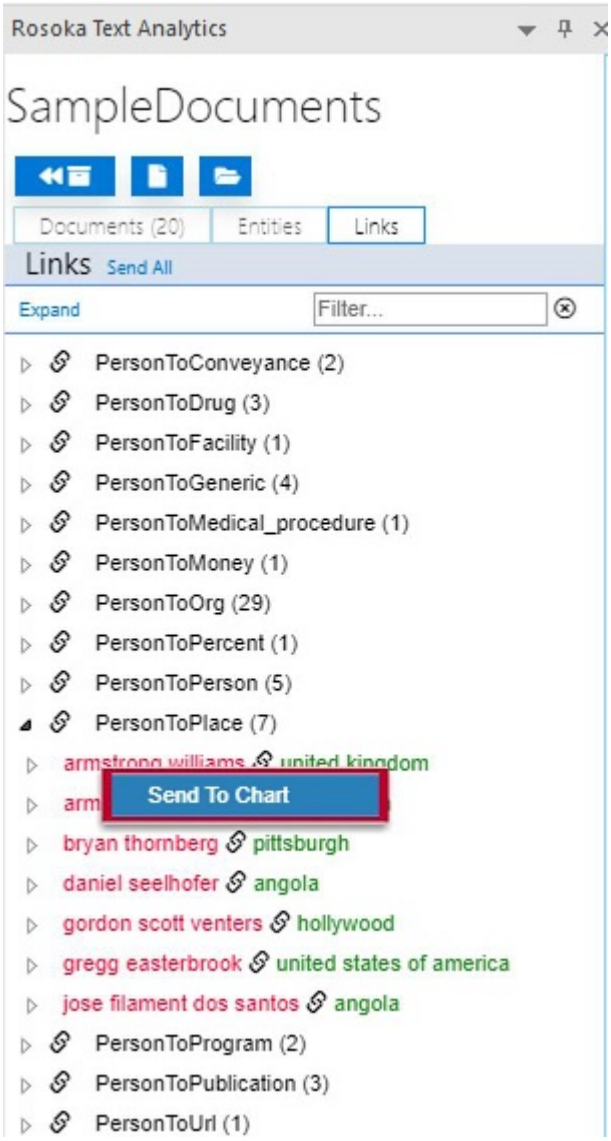
© N. Harris Computer Corporation



Sending Links To A Chart

TextChart provides the same three options for sending extracted links to the chart as it does for sending entities, and you use them in the same way.

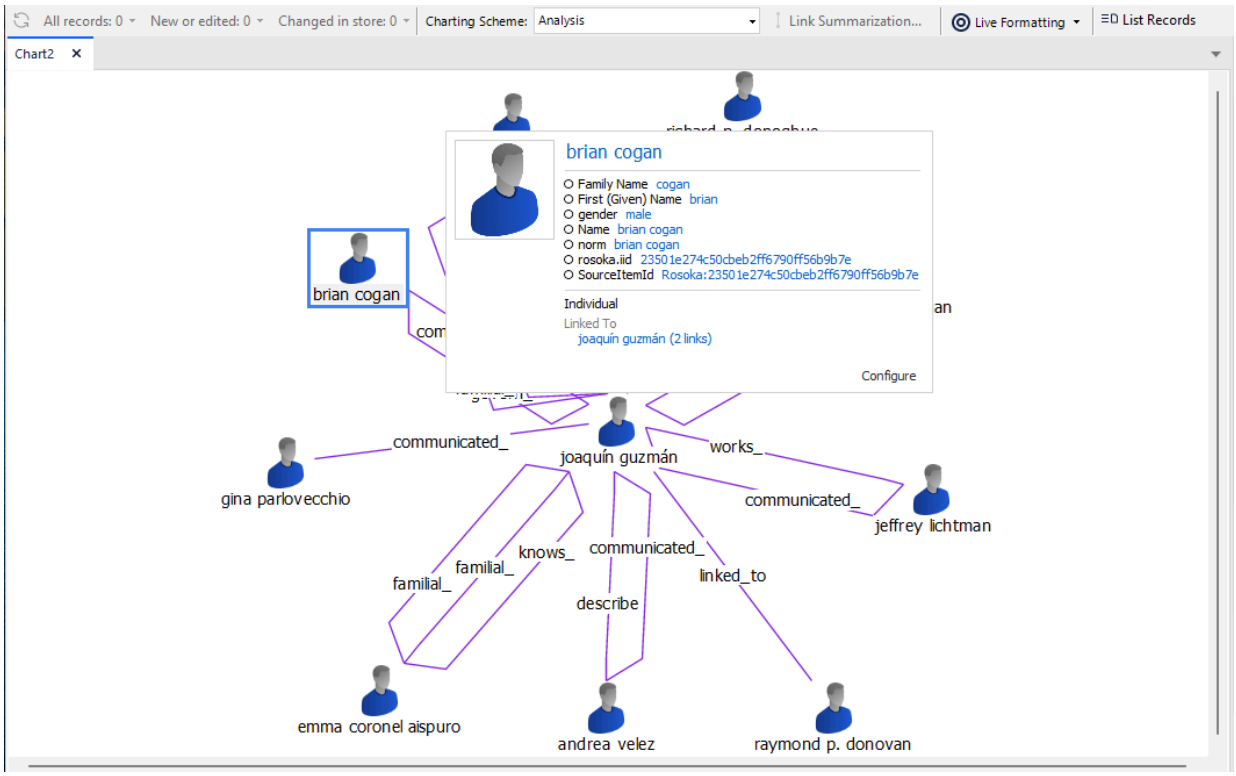
© N. Harris Computer Corporation



Viewing Entities On The Chart Surface

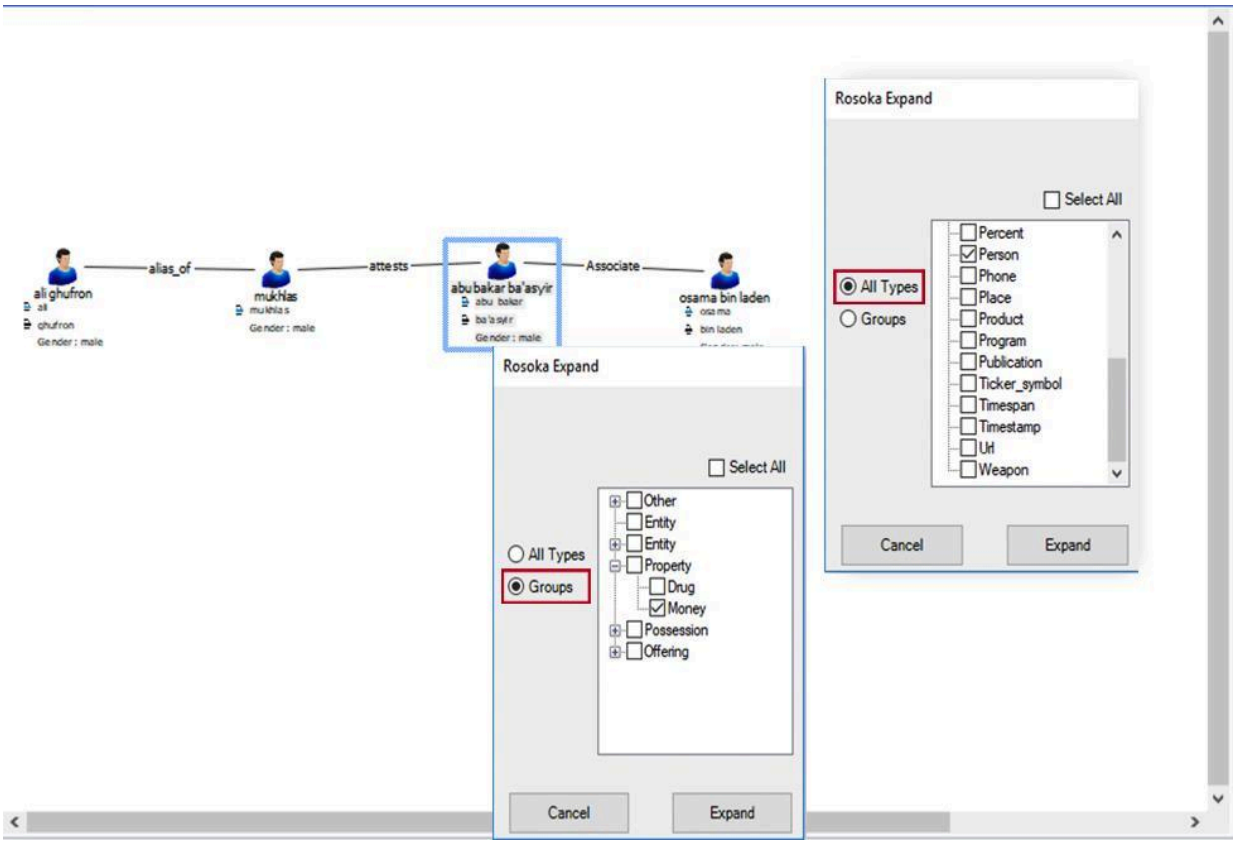
You can hover over *any* entity on the Analyst's Notebook chart surface to display a window containing information about all the properties that are associated with that entity.

© N. Harris Computer Corporation



After you've added an entity *from TextChart* to the chart, you can use it to start adding related entities from the other documents that you've processed. Right-click the entity, and select **Rosoka Expand**.

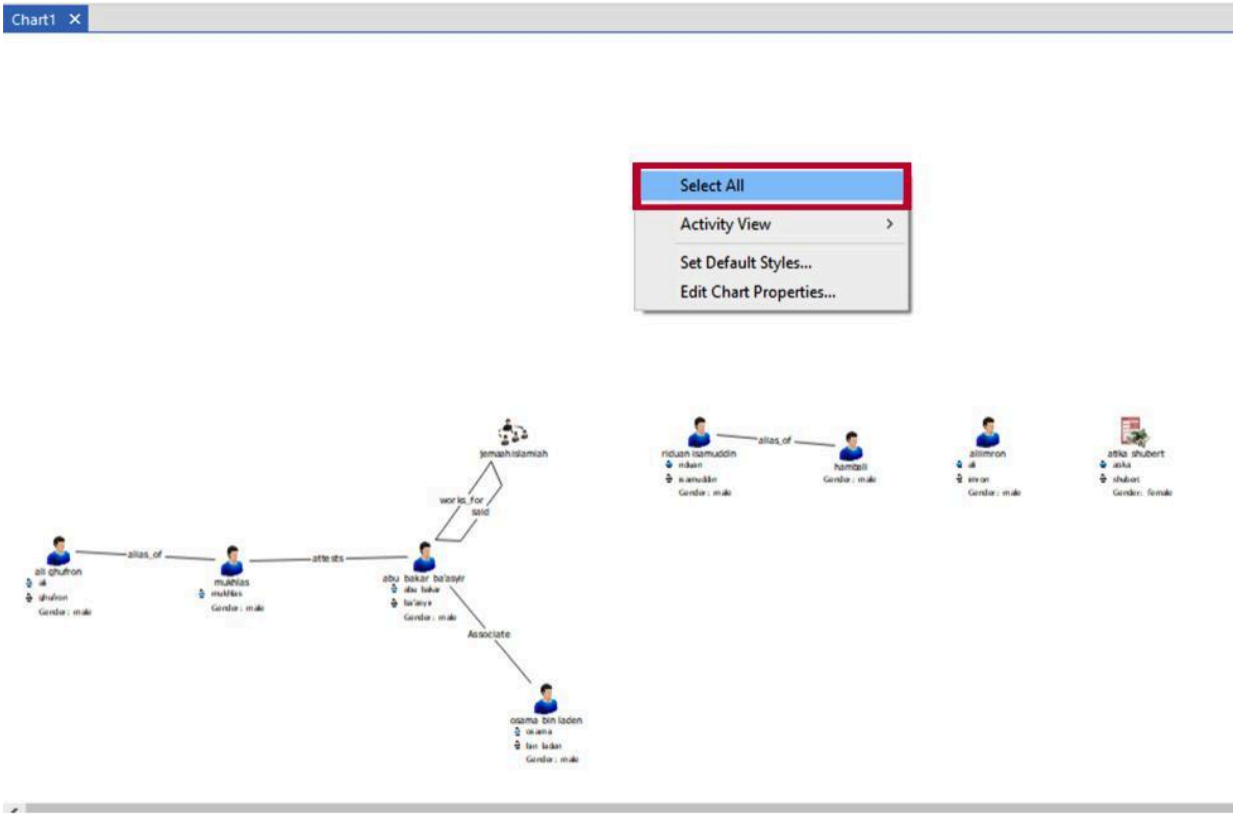
© N. Harris Computer Corporation



You can choose to add related entities of particular types, or groups of types; or you can use **Select All** to add all related entities in a single action.

When a chart contains multiple TextChart entities, you can perform the same **Rosoka Expand** operation on all of them by right-clicking the chart surface and choosing **Select All** first.

© N. Harris Computer Corporation

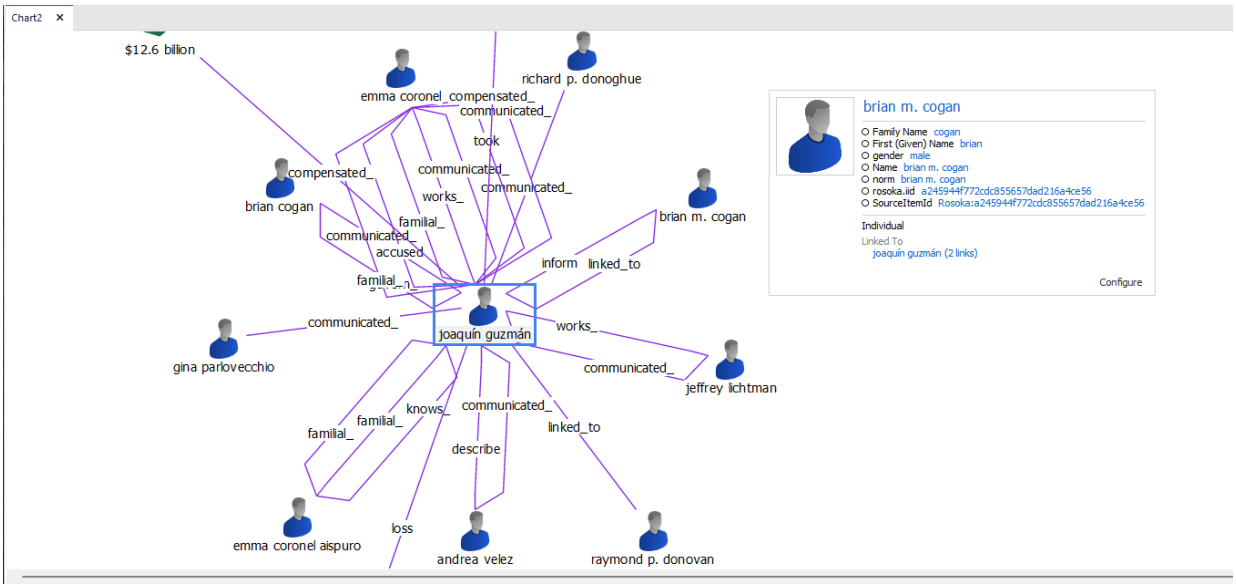


Viewing Links On The Chart Surface

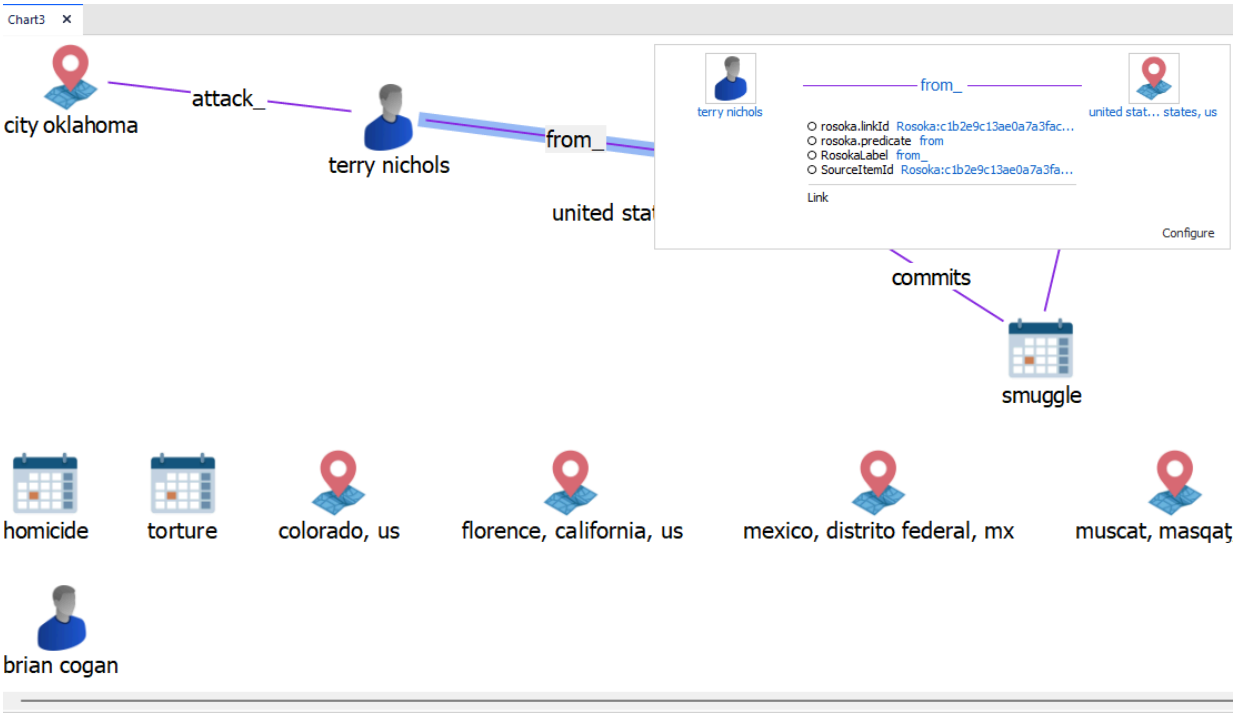
After you've added all the entities to the chart that you want to add, you can start to investigate the relationships between them, including the predicates that were responsible for the links being identified in the first place.

For example, hovering over a connected entity reveals a list of the entities that it is linked to, in addition to its properties.

© N. Harris Computer Corporation



Similarly, you can hover over a TextChart link to highlight the connection between two entities and display additional extraction information.

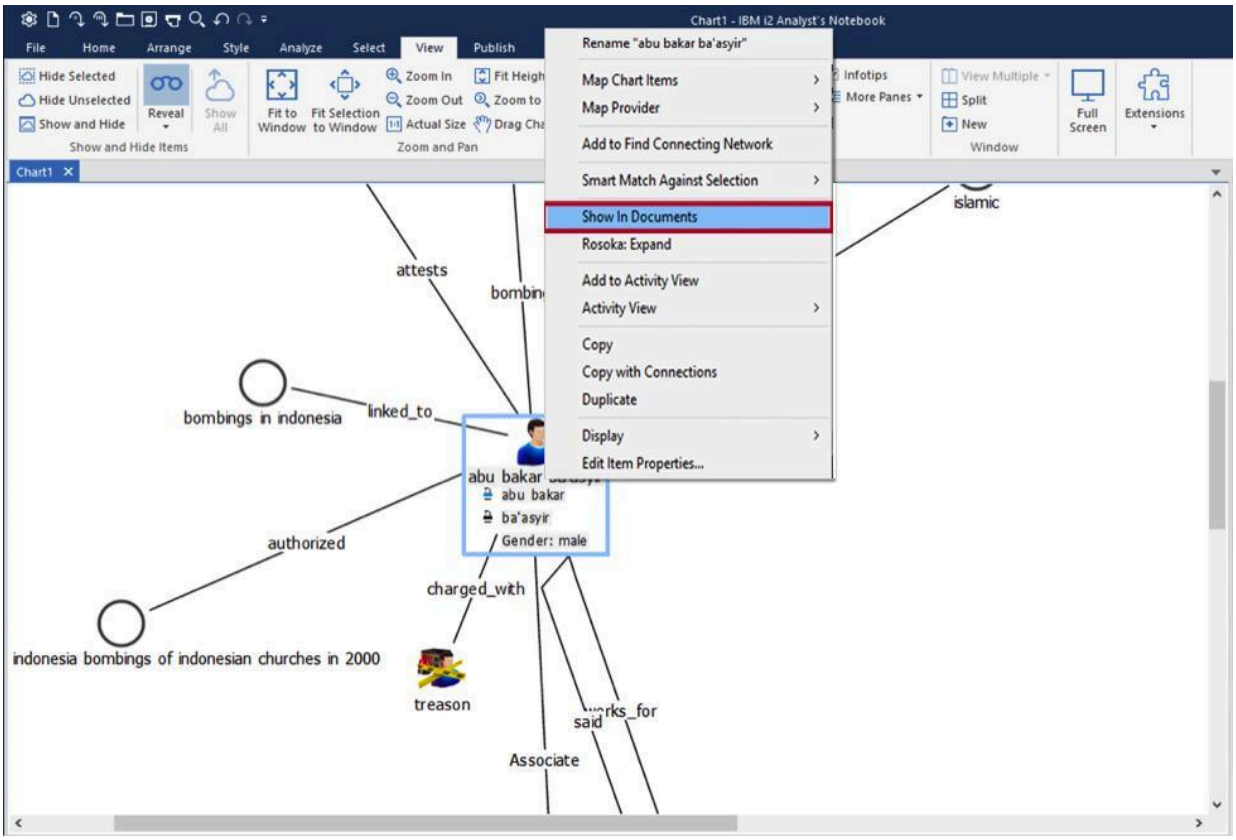


© N. Harris Computer Corporation

Listing Source Documents

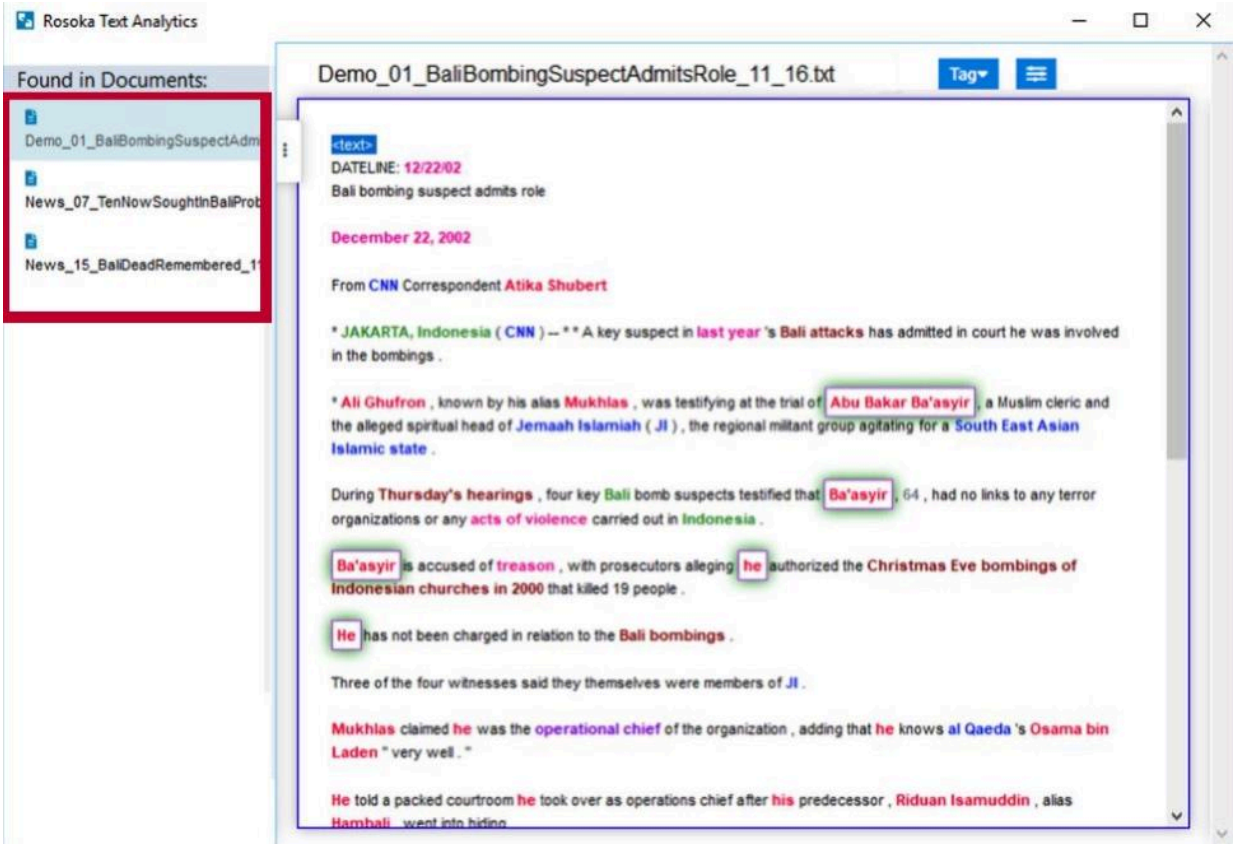
For any entity or link that was added to the Analyst's Notebook chart surface through TextChart, you can quickly refer back to the documents from which it was extracted.

To retrieve all the documents for an entity or a link, right-click and select **Show In Documents** to reopen the Document View.



On the left of the view, the list of documents allows you to select which document appears on the right. Each document appears with text highlighting that reveals the entity or link in question.

© N. Harris Computer Corporation



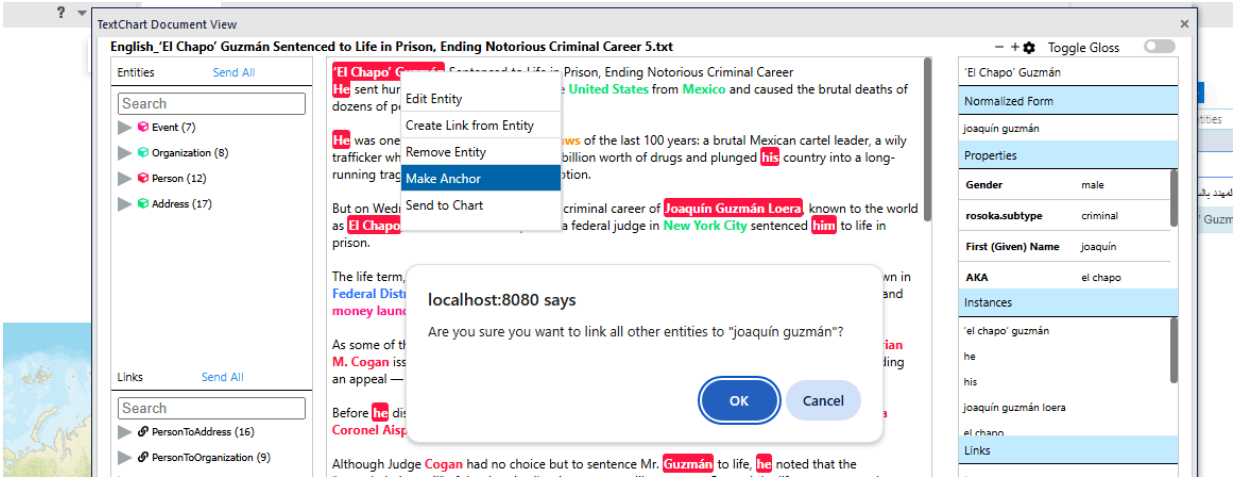
Creating an Anchor Entity

The **Anchor Entity** feature allows users to quickly create links between a selected entity and all other entities in a document. This feature is designed to establish a "hub" entity that connects to multiple other entities, creating a star-pattern relationship graph.

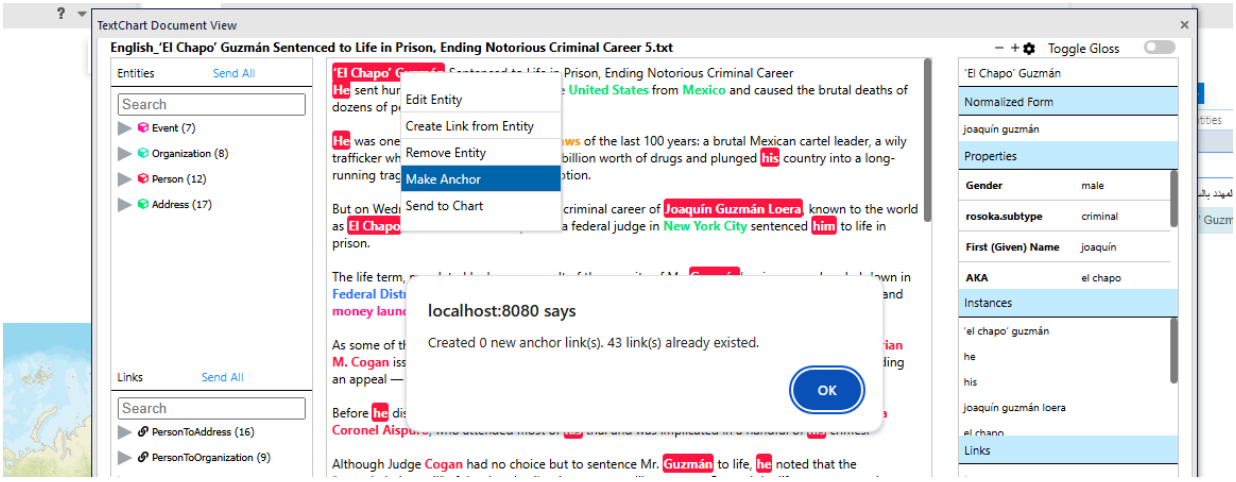
How to Use the Anchor Entity Feature

1. **Right-click on an Entity**: In the document viewer, right-click on any highlighted entity text
2. **Select "Make Anchor"**: From the context menu that appears, click on "Make Anchor"
3. **Confirm Action**: A confirmation dialog will appear asking:

© N. Harris Computer Corporation



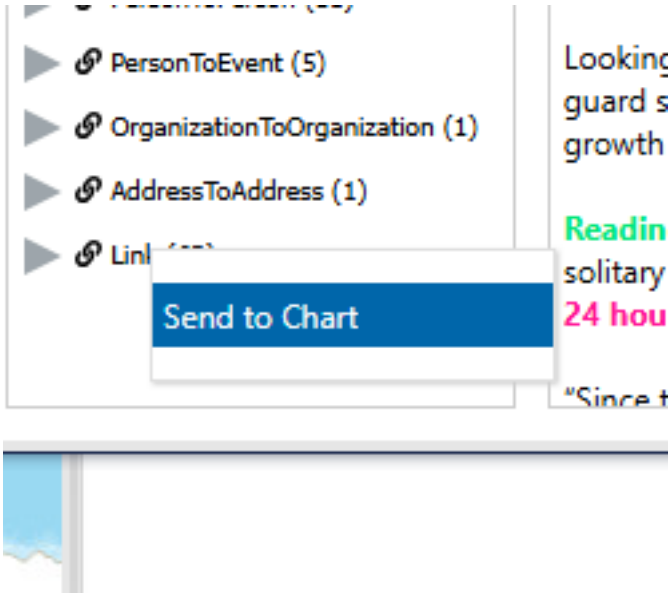
1. **Review Results**: After confirming, an alert will display how many new links were created



This will also prevent duplicate links from being created.

1. **Sending an Anchor Entity to Chart**: Once an entity has been designated as an Anchor, you can send it to the chart using the standard methods (right-click menu or Send to Chart button).

© N. Harris Computer Corporation



TextChart Security

TextChart Premium comes with a built-in security option that utilizes **i2 Analyze** to protect information provided to it. This security option is a deny-by-default token-based authentication system.

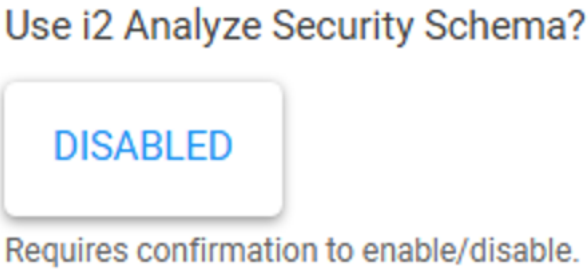
Prerequisites:

- Microsoft Windows
- i2 Analyst's Notebook
- TextChart Premium
- TextChart Data Access
- TextChart Extraction Manager
- TextChart Extraction Worker
- Either a PostgreSQL or Microsoft SQL Server database that already has the base TextChart tables created.
- SQL scripts containing the security schema.
- Valid Licensing for i2 products.
- i2 Analyze set up with security enabled.

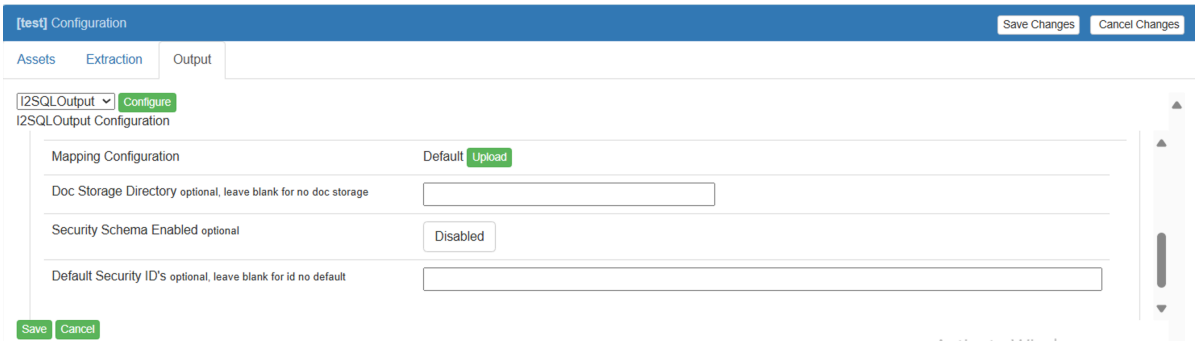
Step-By-Step Setup Instructions

1. Run the SQL security scripts in your PostgreSQL or Microsoft SQL Server database.
2. Open **Rosoka Data Access** in a browser and navigate to the Database tab. Press on the *Use **i2 Analyze** Security Schema?* button and type in SECURITY and press Confirm.

© N. Harris Computer Corporation



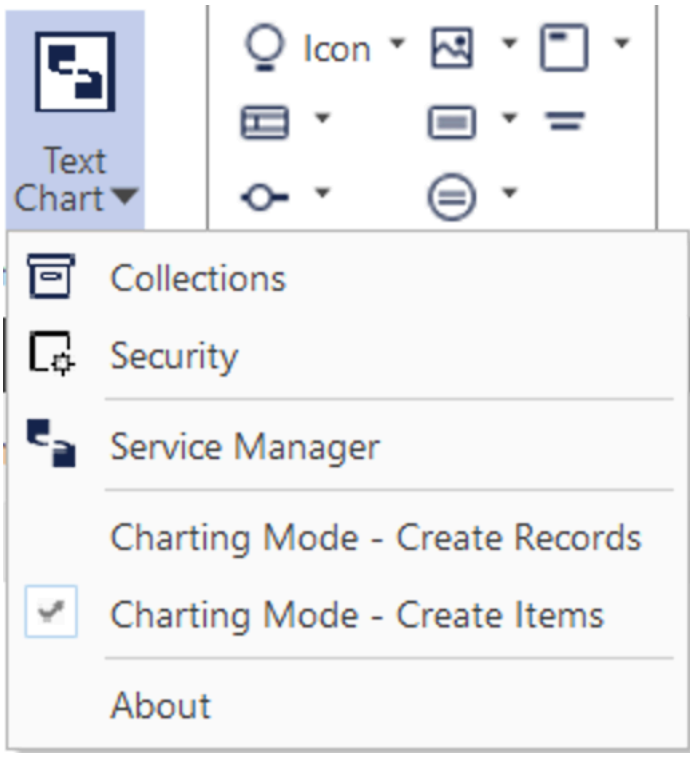
1. Open **Extraction Manager** in a browser and navigate to the Clusters tab. Then, click on the gear next to the cluster you want to enable security on and choose Configure. Select the Output tab and then press the Configure button. Scroll down to Security Schema Enabled and press on the button next to it. Type SECURITY and press confirm. Press the Save and then Save Changes button.



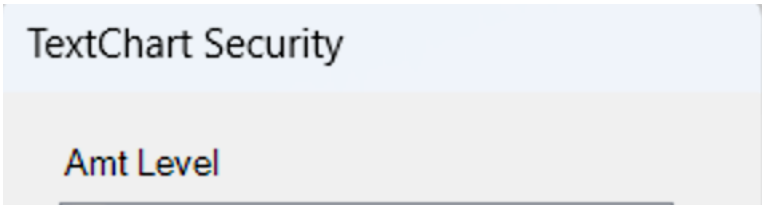
2. Restart TextChart Premium, **Rosoka Data Access**, **Extraction Manager**, and **Extraction Worker**.

To verify TextChart Security is functioning, go to the TextChart ribbon, press on it, and look for the Security option.

© N. Harris Computer Corporation



If that option is visible, then security has been successfully enabled. Additionally, if a user tries to process a document with security enabled, a security pop-up will appear.

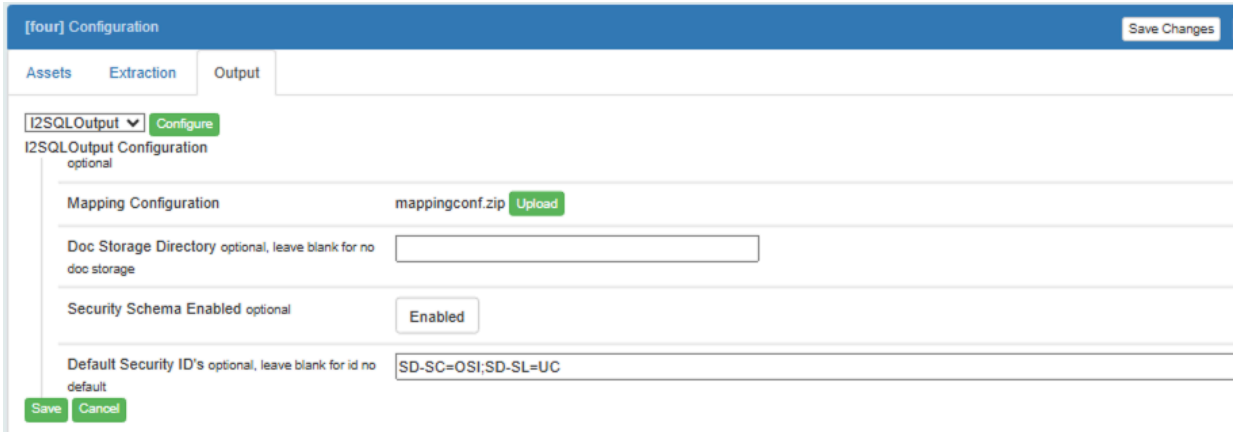


Usage:

- Security is added to individual documents. There is no security on a document until it is processed. The security of a collection depends on the security of the documents within it. If a user does not have access to any of the documents within the collection, the collection will not be visible. If a user does not have access to the document, the user will not be able to open the document viewer.
- Once logged in, users can modify security settings by selecting the Security option in the TextChart ribbon. The contents of the pop-up that appears will depend on settings in **i2 Analyze**.
- When processing documents, a pop-up will appear allowing a user to choose what security the document will have. This pop-up may appear behind i2 Analyst's Notebook.

© N. Harris Computer Corporation

- If a user does not process the documents within a collection before navigating away from that collection, then that collection will not be visible afterward. This will prevent a user from creating a collection of the same name. In order to create a collection with the same name, the user will have to clear the local database, which can be done using the TextChart Configuration UI that comes with TextChart and is available in the TextChart installation folder. Be aware that clearing the local database will remove ALL collections, but will not delete any processed documents. Any processed documents can be readded to new collections.
- The user can provide Default Security Ids in **Extraction Manager** by navigating to the Clusters tab, clicking on the gear next to the desired cluster, choosing Configure, selecting the Output tab, and pressing the Configure button. Scroll down to Default Security Ids. Type in the desired Default Security Ids and press confirm. Press Save and then Save Changes button. Here is an example Default Security Ids value: SD-SL=CON;SD-SC=HI. SD-SL is the Security Level and SD-SC is the Security Category. Multiple values are separated by semicolons.



i2 TextChart Server

i2 TextChart Server provides a scalable environment for performing data extraction from free-form text. Important features include:

- A REST interface for processing text with Java and C/C++ clients
- Scalable up or down depending on throughput requirements
- Browser-based administration interface for setup, management, and monitoring
- Input and output connectors for interfacing with different document sources and databases.

The TextChart Server system is broken up into two major pieces: the *manager node* and the *worker nodes*.

Manager node

An installation of i2 TextChart Server contains a single manager node. This node acts as the central point for submitting data for processing and for the administration and monitoring of the overall system.

The manager provides these services:

© N. Harris Computer Corporation

- A REST endpoint for submitting documents for extraction processing, either in "immediate" mode (where a result is returned to the caller) or "ingest" mode (where results are pushed to a data repository, such as a database)
- A browser-based user interface that allows an administrator to set up, manage, and monitor the entire system from a single point
- Automatic document queuing and load balancing
- Data collection from sources such as the file system

Worker nodes

An installation of i2 TextChart Server contains one or more worker nodes. These nodes accept documents from the manager, process them using i2 TextChart Server technology, and either return the results to the user or send them to a predefined data repository.

A worker provides these services:

- A zero-configuration instance of the i2 TextChart Server engine. All configuration is managed by the worker node in conjunction with the manager node.
- Adjustable multi-threaded processing.
- Storage of results in a defined data repository, such as a database.

Clusters

The main unit of configuration in the i2 TextChart Server system is the *cluster*, which defines how content will be processed and where the results of that processing will be sent. Worker nodes can be added to or removed from clusters according to throughput needs.

Each worker that belongs to a particular cluster shares the same configuration. Data extraction settings and the destination for the processing results are same across all workers in the cluster.

After the details of a processing configuration have been established for a cluster, any worker belonging to that cluster is automatically configured to reflect those details. If a worker is moved from one cluster to another, its configuration is automatically updated to that of the new cluster.

An installation of i2 TextChart Server might define only one cluster, especially if there is only one type of processing that needs to be performed on the incoming documents.

A cluster contains the following configurable elements:

- **LxBase**: The LxBase is the set of lexical data (dictionaries and patterns) that controls what elements are extracted from documents by the extraction engine. A default LxBase is distributed with the i2 TextChart Server system, or one can be produced with TextChart Studio.
- **GxBase**: The GxBase is a set of geographic data that enables coordinate and other metadata lookup for locations around the globe. A default GxBase is distributed with the i2 TextChart Server system.
- **Properties**: Properties act to make minor adjustments to how the extraction engine and geographic lookup operate. These settings are independent of the LxBase and GxBase.
- **Output connector** (optional): An output connector is a piece of code that takes the results of data extraction and moves it to a defined repository. i2 TextChart Server includes two such connectors: one that writes results to the file system, and one that pushes results to an Elasticsearch database.

All of these elements can be modified through the manager's administrative interface. In the case of the LxBase and GxBase, the manager provides a facility to upload a new version to a particular cluster.

© N. Harris Computer Corporation

Processing documents

There are two ways of processing documents through the i2 TextChart Server system: the REST API, and input connectors.

REST API

In i2 TextChart Server, the manager provides a REST API for submitting documents for processing. There are two ways to submit document text:

- **Raw**

  Raw documents are provided as a simple array of bytes. These documents may be of any supported encoding and document type, including complex formats like Microsoft Word and PDF. The i2 TextChart Server system first converts these documents to simple text before they are analyzed by the extraction engine.

- **String**

  Documents that are already in simple text format can be submitted as a UTF-8 encoded string. Doing so avoids the processing overhead of the conversion.

And there are two types of processing that the server can perform:

- **Immediate**

  In immediate processing, documents are pushed to the head of the queue on the manager so that they are processed as soon as a worker node has free cycles. The results are then returned as the result of the REST call.

  In general, use immediate mode for cases where the extraction results are needed immediately, such as an application with a user interface.

- **Ingest**

  In ingest processing, documents are pushed to the end of a queue (one queue per cluster), from which they are processed in order on the next available worker node that has free cycles. The results are then pushed to the cluster's output connector, which is typically a database.

  The result from the REST call is a simple status value indicating whether the document was successfully queued for processing.

  Note: If no output connector is defined for a cluster, and documents are submitted in ingest mode, then the extraction results are discarded.

Input connectors

Another way of getting documents into the system is to use an input connector. Input connectors are small pieces of code that are loaded dynamically by the i2 TextChart Server manager node at startup.

You can start and stop input connectors through the manager administration interface, or you can set them to start automatically when the manager starts.

This version of i2 TextChart Server includes the FileMonitor input connector, which can monitor one or more directories on the manager file system. For each directory, it can process new and changed documents immediately, or do so on a schedule as the contents change. Optionally, the connector can delete or move processed files to another location in the file system.

Installing i2 TextChart Server

Installing the i2 TextChart Server system involves two steps:

© N. Harris Computer Corporation

1. Choose a target system and install and start the manager software.
2. Choose one or more systems on which to install the worker software.

Manager installation

Before you install the manager software, make sure you have the following information, which you will need during the installation process:

- **Installation directory**: A directory where the software will be installed. The directory's file system should have at least 2 GB of free space.
- **Client port**: A network port where the REST API will communicate. Choose an unused port on your system.
- **Admin port**: A network port where the administration user interface will be hosted. Choose an unused port on your system.
- **Worker service port**: A network port that the worker nodes will contact to interact with the manager. Choose an unused port on your system.
- **Queue directory**: A directory where any backlog of documents submitted for processing will be written. If left unspecified during installation, a directory will be created under the installation directory.

  The file system where this directory is located should have as much free space as you anticipate needing to store submitted documents until they are processed. If the space is full, the manager rejects documents until more becomes available.
- **License key file**: An XML file that you received from i2 containing your license for the i2 TextChart Server software.

  Note: You *can* install and administer the software without a license key, but it will not process documents without a valid license.

On Linux, you must run the manager installation with root privileges. (Use the `sudo` command to start the process.)

When the manager software is installed, you must start the service, which is named `rosokamgr` on Linux or `RosokaManager` on Windows. The service is set to start automatically at operating system startup.

Start the service by issuing the command appropriate for your system. For example, the following command starts the service on most variants of Linux:

```
sudo service rosokamgr start
```

While on Windows, you can use this command:

```
net start RosokaManager
```

On Windows, you can also use the Services control panel to start, stop, and manage the service.

Verify that the manager is running before you proceed to installing workers.

Worker installation

Before you install the worker software, make sure you have the following information:

- **Manager host**: The hostname or IP address of the system on which you installed the manager software.
- **Worker service port**: The network port that you specified for the same setting on the manager node.

© N. Harris Computer Corporation

- **Worker name**: The name of this worker node. Typically, this value is the network hostname of the node, but you can choose something else.
- **Worker port**: A network port that the manager can use to communicate with the worker. Choose an unused port on your system.
- **Threads**: The number of threads to use when processing documents on this node. Each thread processes one document, so the thread count determines how many simultaneous documents can be processed.

  The default value of `0` automatically uses all available cores for processing.

On Linux, you must run the worker installation with root privileges. (Use the `sudo` command to start the process.)

During installation, the procedure attempts to verify that the provided manager host and port information is correct. The manager service does not have to be running when you install the worker software, but the verification will fail.

When the worker software is installed, you must start the service, which is named `rosokawrk` on Linux or `RosokaWorker` on Windows. The service is set to start automatically at operating system startup.

Start the service by issuing the command appropriate for your system. For example, the following command starts the service on most variants of Linux:

```
sudo service rosokawrk start
```

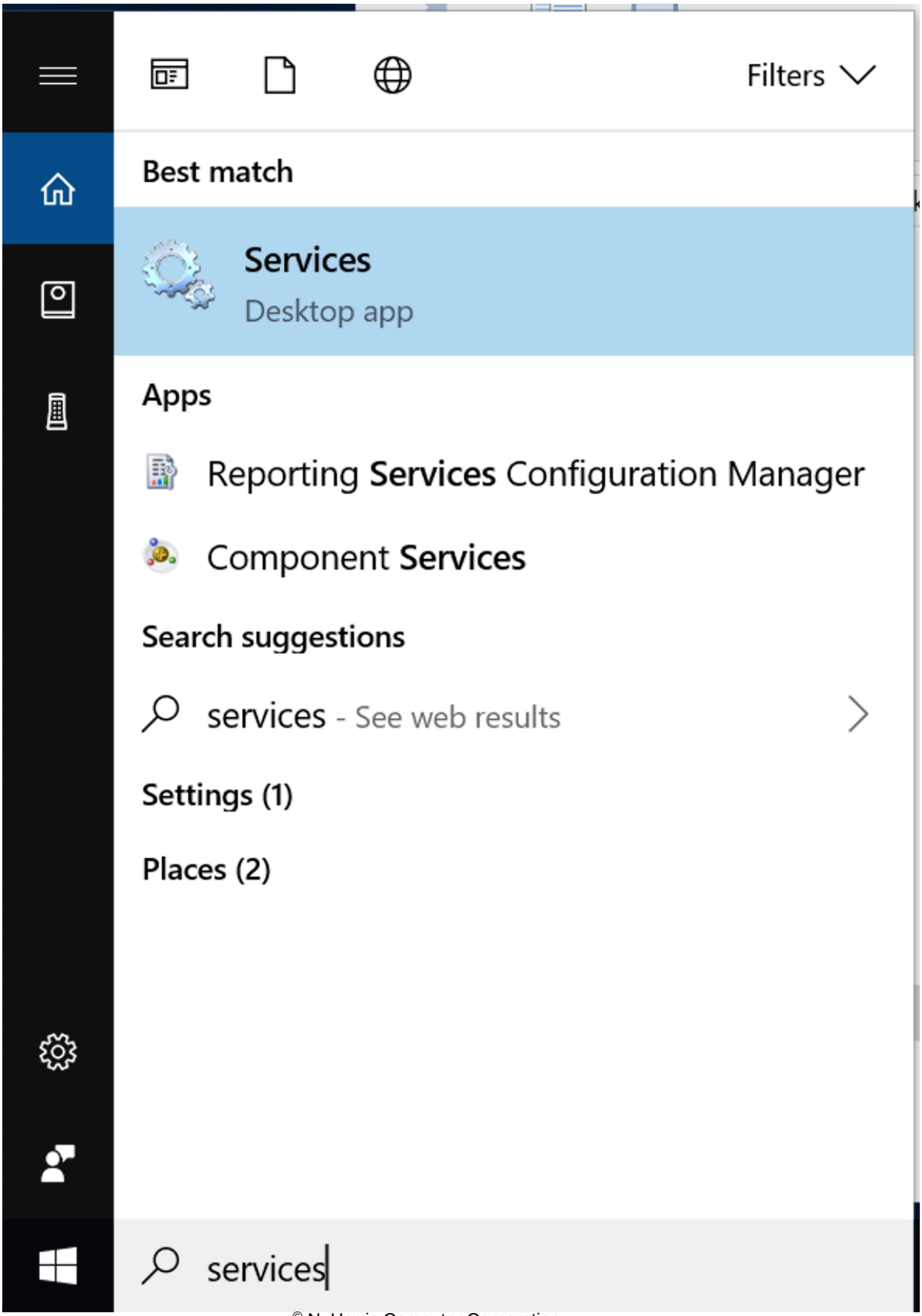
While on Windows, you can use this command:

```
net start RosokaWorker
```

On Windows, you can also use the Task Manager or the Services control panel to start, stop, and manage the service.

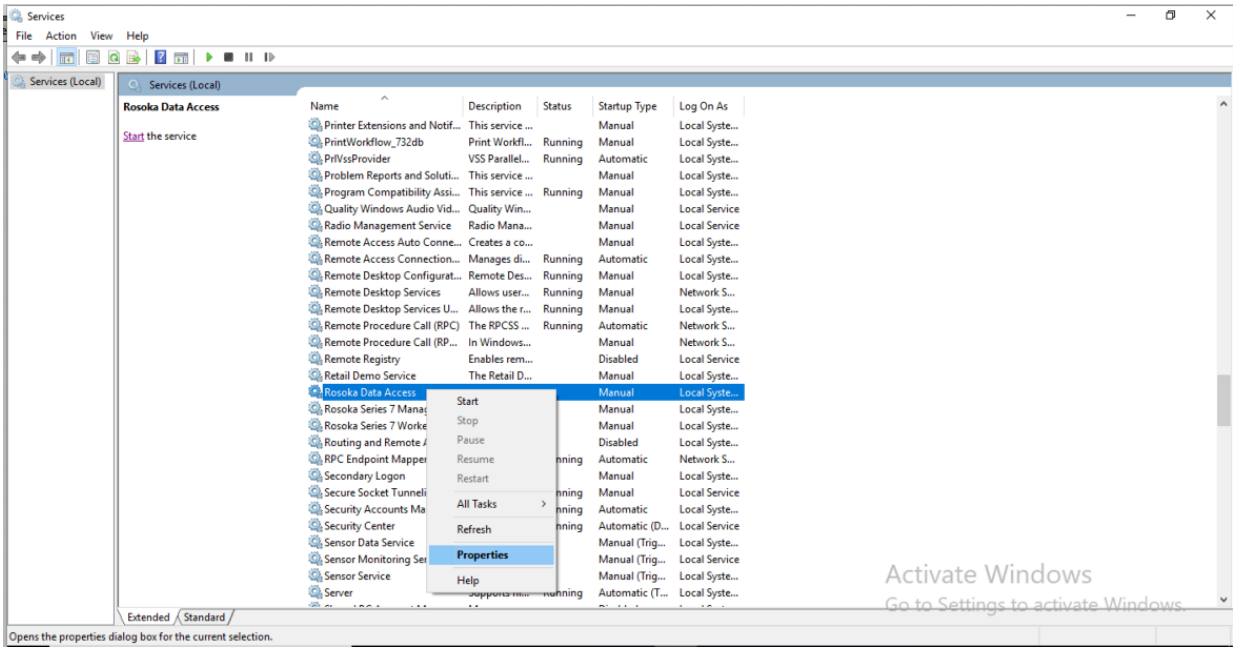
Also on Windows, you must configure the services to restart automatically in the event of failure. To do so, access the services by launching the Services control panel:

© N. Harris Computer Corporation



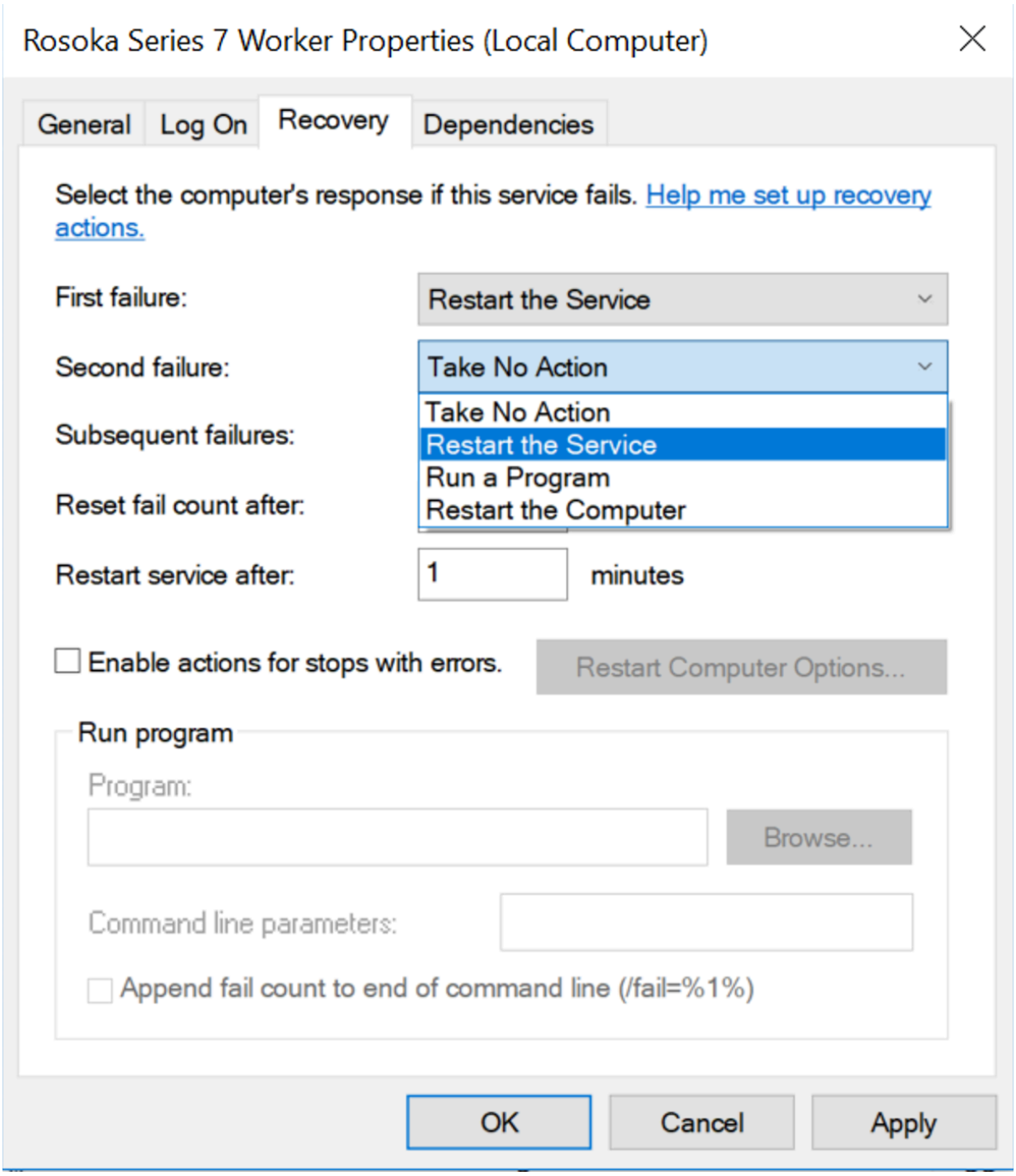
© N. Harris Computer Corporation

Locate the "i2 TextChart Server Worker" service, and view its properties:



In the Properties dialog, open the **Recovery** tab, and set the **First failure**, **Second failure**, and **Subsequent failures** settings to **Restart the Service**:

© N. Harris Computer Corporation



Starting the services

On most variants of Linux and Microsoft Windows, the installed services are configured to start automatically. For other operating systems, you must issue commands in a terminal window.

© N. Harris Computer Corporation

To run the *manager* software manually, follow these steps after installation:

1. Start a terminal or command prompt window on the installation system.
2. Navigate to the installation directory. Then, issue the following command:

   ```
   java -Djava.util.logging.config.file=logger.properties --jar
    RosokaServerManager.jar
   ```

   The manager software then stays running in the terminal window.

To run the *worker* software manually, follow these steps after installation:

1. Start a terminal or command prompt window on the installation system.
2. Navigate to the installation directory. Then, issue this command:

   ```
   java --Xmx4g --Djava.util.logging.config.file=logger.properties --jar
    RosokaServerWorker.jar
   ```

   The worker software then stays running in the terminal window.

Getting started

At this point, you have installed the manager software on one machine and at least one instance of the worker software on another. This topic describes how to set up your first cluster and begin processing documents.

Creating a new cluster

A cluster is a set of nodes that share the same type of extraction processing and data destination. To process any documents with i2 TextChart Server, you must define at least one cluster.

1. Open the manager user interface by navigating to `http://<manager_host>:<admin_port>/manager` in a web browser, replacing `<manager_host>` and `<admin_port>` with the values for your installation.

   If you're asked to log in to the administrator UI, the default user is `admin` and the default password is `rosoka`. You can change these credentials later.
2. The left side of the user interface contains contains links to different features. Click **Clusters** to navigate to the Clusters page.
3. Initially, no clusters are defined. Create one by clicking **New Cluster**.
4. In the dialog that appears, enter a name and an optional short description for the new cluster. Click **Add**.

At this point, you have a new cluster with a default set of LxBase, GxBase, and properties. To configure the cluster further, click the drop-down menu next to its name and select **Configure**.

Adding workers to a cluster

For now, we will leave the cluster settings as they are, and add one or more nodes to it.

1. Click **Workers** to navigate to the Workers page.

   The page is divided into two sections: known or *registered* workers appear at the top, while workers that are unknown but requesting access to the manager appear at the bottom. You should see a list of the workers that you started in the bottom section.
2. For each of the unknown workers, select the drop-down menu and click **Register Worker**.

© N. Harris Computer Corporation

As you register the worker nodes, they disappear from the bottom and move to the top of the page with a status of **registered**.

3. In the **Cluster** column for each registered worker, click the drop-down and select the cluster that you just created. The worker then moves to the **configured** state.

   Depending on the speed of your network, this operation might take a few minutes, as the lexical and geographic assets are quite large.
4. When all the nodes are in the **configured** state, click the drop-down menu next to the worker name and select **Enable/Disable**. Each worker moves to the **starting** state, and eventually to the **ready** state. At this point, the workers are ready to process documents.

Processing documents

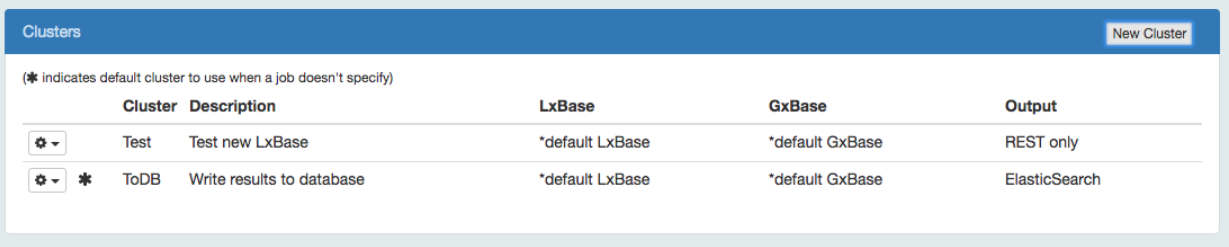
To process some documents, you can use the manager's built-in web client to select a few files on your local file system.

1. Navigate to `http://<manager_host>:<client_port>`, where `<manager_host>` is the hostname or IP address of the manager, and `<client_port>` is the client port number that you specified during installation.
2. In the page that appears, select a cluster to use for processing from the **Select a cluster** drop-down menu.
3. Click **Drop files here to upload** to open a file selection dialog, or simply drag and drop a file from your local file system to the drop location.

   The file is uploaded and processed, and the results appear on the right side of the display.

Managing clusters

To manage clusters in i2 TextChart Server, you use the Clusters page of the manager's user interface:



Adding a new cluster

To add a new cluster, click **New Cluster** in the upper-right corner of the page to display a dialog. Enter a unique name for the cluster and an optional short description. Then, click **Add** to create it.

A new cluster is automatically configured with the default LxBase, the default GxBase, and a default set of properties. At this stage, you can get output from the cluster only by providing documents directly through REST API calls.

© N. Harris Computer Corporation

Setting the default cluster

One cluster in the system is always designated as the *default* cluster, which means that it's used if no specific cluster is requested when documents are sent for processing through the REST API.

Unless you change it, the first cluster that you created is the default To change the default cluster, click the drop-down menu in the row that contains the cluster you want to use, and select **Set as Default**. The asterisk moves to the selected cluster, indicating that it is now the default.

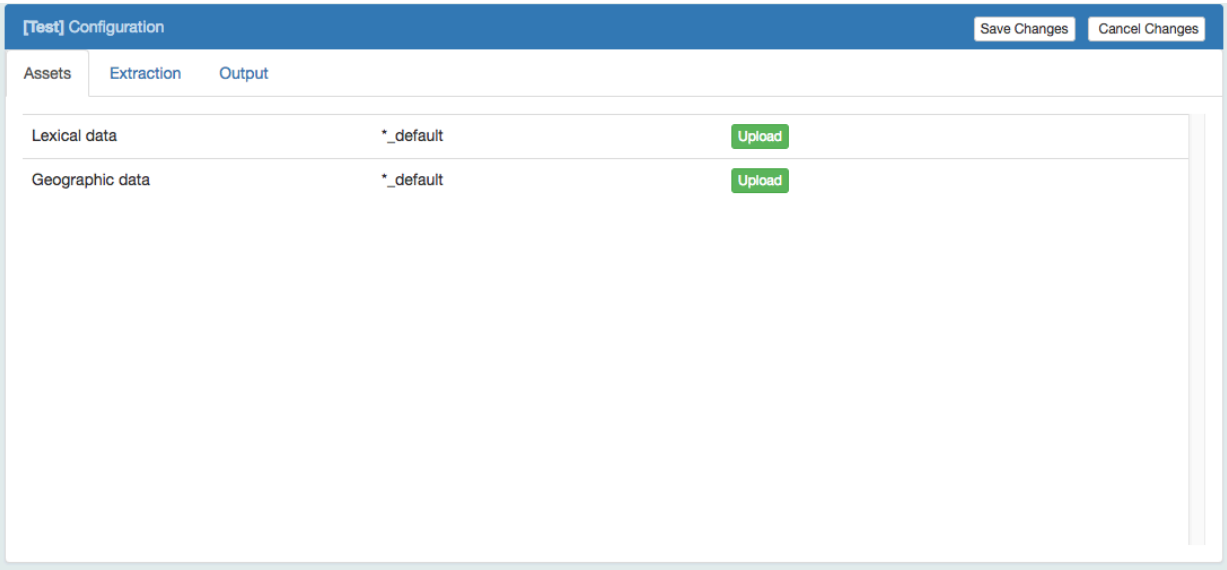
Removing a cluster

To remove a cluster that you no longer need, click the drop-down menu in the row that contains the cluster and select **Remove Cluster**. Then, click **OK** in the confirmation dialog to remove the cluster from the system.

You can remove a cluster only if it has no assigned workers. Use the Workers page to remove all worker nodes from the cluster in question before you attempt to remove the cluster itself.

Configuring a cluster

To configure a cluster, click the drop-down menu in the row that contains the cluster and select **Configure**. A panel opens below the cluster table that allows you to see and change all available cluster options:



The configuration settings are split across three tabs: **Assets**, **Extraction**, and **Output**. After you make changes, click **Save Changes** button in the upper-right corner of the panel to propagate them to all the worker nodes in the cluster. Alternatively, click **Cancel Changes** to discard your changes.

Assets

A cluster's *assets* are the two large pieces of data that i2 TextChart Server uses for extraction and tagging:

© N. Harris Computer Corporation

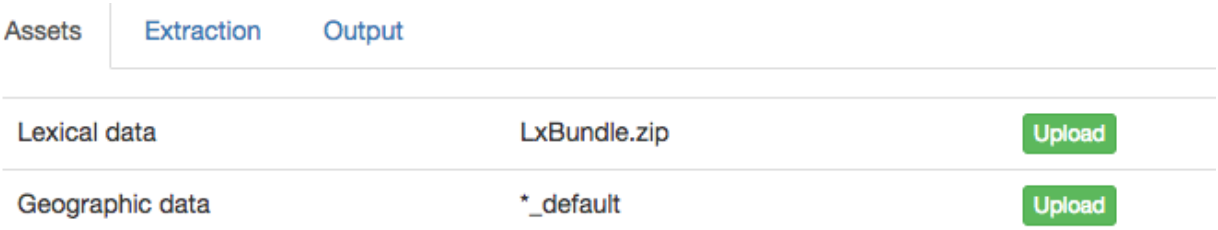
- **LxBase**: The LxBase consists of multi-lingual dictionaries and rules that control how the i2 TextChart Server engine finds named entities in unstructured text. The LxBase is a single ZIP file that is provided by i2 or created through TextChart Studio.
- **GxBase**: The GxBase consists of a geographic database that is used to augment information on extracted place entities. The GxBase is a single ZIP file that i2 provides.

By default, i2 TextChart Server comes with the latest version of each of these assets. When you create a cluster, these default versions are used.

To use a new LxBase or a new GxBase, you can upload it by using the **Assets** tab in the cluster Configuration pane:



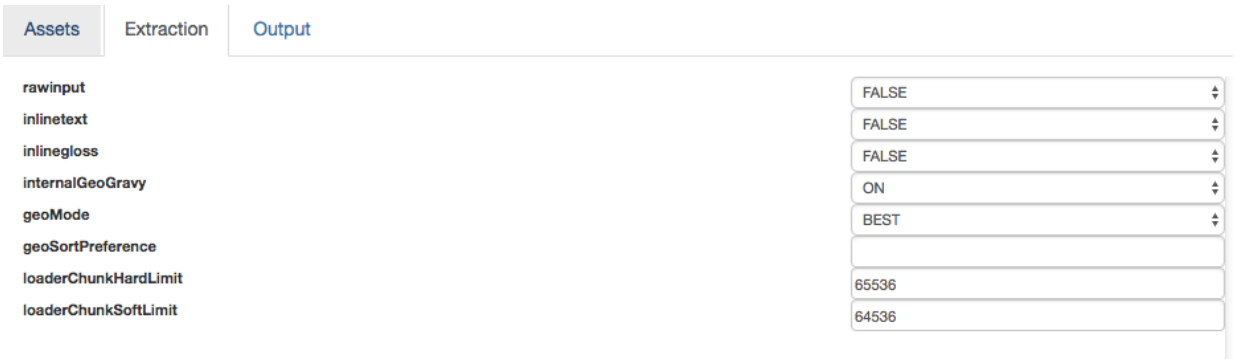
Click **Upload** next to the appropriate asset to select a file to upload. After upload, the file is checked to validate the data it contains. If validation is successful, the name of the asset changes:



Extraction

Use the **Extraction** tab to set run-time operational properties for the i2 TextChart Server engine:

© N. Harris Computer Corporation



You can modify each property by selecting a value from a drop-down or typing into a field:

- **rawinput**

  Set to TRUE to force the engine never to convert documents from non-text formats such as Word, Excel, or PDF.
- **inlinetext**

  Set to TRUE to populate the `inlinetext` field of the `RosokaFullObject` output, so that extraction tags are embedded in the source text.
- **inlinegloss**

  Set to TRUE to populate the `inlinegloss` field of the `RosokaFullObject` output, so that the gloss is embedded in the source text.
- **internalGeoGravy**

  Set to ON to enable geographic name lookup.
- **geoMode**

  Set to BEST to disambiguate place names with multiple matches solely by contained priority value and type.

  Set to COLOCATED to take other locations in the document into account, to find locations in a close geographic area.
- **geoSortPreference**

  Indicates how multiple geographic name matches are disambiguated. Can be blank (no preference) or one of:

  - **CONUS** (prefer locations in the continental US)
  - **OCONUS** (prefer locations outside of the continental US)
  - ***region*** (prefer locations in the given region, including AFRCA, AMER, ANTAR, ASIA, BALK, CAFR, CAMER, CARIB, CAUC, CEURA, CEURO, EAFR, EASIA, EURA, EURO, MDEDT, MEAST, NAFR, NAMER, NEURO, OCEAN, OCENA, RUSSA, SAFR, SAMER, SASIA, SEASA, SEURO, and WAFR)
  - **country: *country*** (prefer locations in the given country)
  - ***N, W, S, E*** (prefer locations in the region bounded by the given lat/long measures in decimal degrees)

© N. Harris Computer Corporation

- **loaderChunkHardLimit**

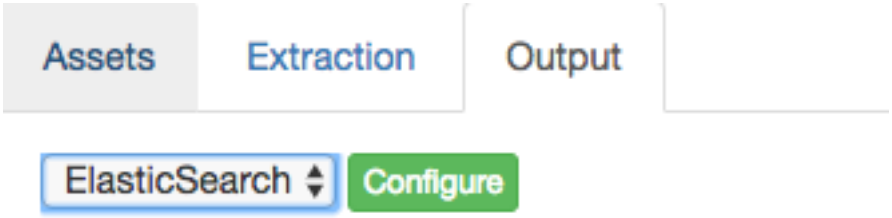
  Maximum length, in characters, of text that the engine can process in one go. A value of -1 indicates no limit (not recommended).

- **loaderChunkSoftLimit**

  Minimum length, in characters, to be processed before the loader attempts to break the document at a reasonable location (such as a paragraph break or a sentence break). A value of -1 indicates no limit. Must be set to a value lower than the hard limit.

Output

Use the **Output** tab to set the output connector for extraction output when documents are processed in "ingest" mode. A cluster that has no output connector can be used only for "immediate" processing, by returning the extraction results through the REST API.



To change the output connector for a cluster, select a new one from the drop-down list. If you don't want to use an output connector, select **None**.

After you select an output connector, click **Configure**, enter the information that the connector needs, and the click **Save** below the configuration panel.

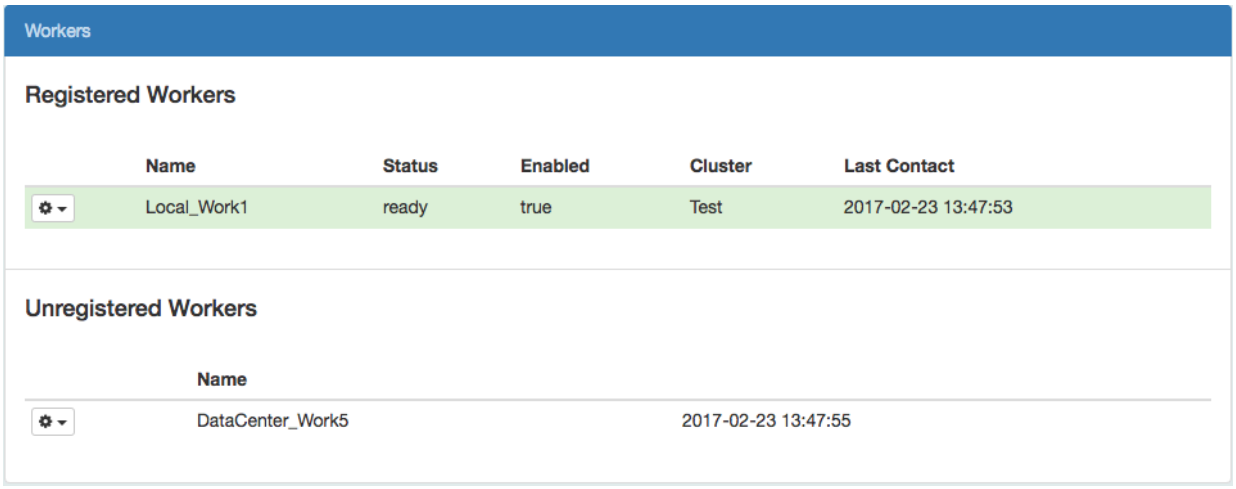
Note: You must click **Save** in the configuration before you click **Save Changes** at the top of the panel in order for your changes to be applied.

See Storing Results using Output Connectors for information about the configuration options for the built-in output connectors.

Managing workers

To manage worker nodes in i2 TextChart Server, you use the Workers page of the manager's user interface:

© N. Harris Computer Corporation



Registering a worker

All worker nodes have a configured manager, and they attempt to register with that manager when they start. Worker nodes that are not formally registered with the manager appear in the Unregistered Workers section of the page.

To register a worker node, select the drop-down next to it in the Unregistered Workers section, and select **Register** option. The worker is moved to the Registered Workers list and will soon enter the **registered** state. It can take up to 30 seconds for registration to complete.

Removing a worker

To remove a registered worker node from the system, select **Remove Worker** from the drop-down in the row that contains the worker in the Registered Workers table. The worker is moved to the Unregistered Workers list, and it will no longer process documents.

Assign a worker to a cluster

To assign a worker node to a cluster, click the drop-down menu in the **Cluster** column in the Registered Workers table. Select the desired cluster.

If the worker is already assigned to a cluster, then click the left-most drop-down in the corresponding row of the Registered Workers table, and select **Unassign Worker**. The worker is removed from the cluster, and you can assign it to a different cluster.

Enable or disable a worker

By default, when you register a worker node with the manager, it connects to the manager and completes all necessary configuration, but stays in the **configured** state.

Note: A *disabled* worker like this uses a license instance. A worker that you don't plan to use for processing should be removed from the system entirely.

To move a worker from (or back to) the **configured** state, click the drop-down in the row that corresponds to the worker in the Registered Workers table. Select the **Enable/Disable** option to change its state.

© N. Harris Computer Corporation

When a worker node is enabled, it will start the Rosoka Series 7 engine and be ready to process documents.

Worker states

A worker can be in one of the following states:

State	Description
off	The worker is registered with the manager, but it is not currently connected. This is typically the case for a worker node that has been shut down, or where the worker process is not running.
awaiting registration	The worker is registered, and the manager is waiting for the worker node to contact it to complete the connection process.
registered	The worker is connected to the manager but has no assigned cluster.
configuring	The worker is downloading configuration assets corresponding to the selected cluster.
configured	The worker is fully configured with the current cluster configuration.
starting	The worker is starting the i2 TextChart Server engine.
ready	The worker is ready to process documents.
stopping	The worker is shutting down the i2 TextChart Server engine.
error	A configuration error has occurred. Typically, errors are caused by incorrect settings in an output connector. Correct the error by configuring the cluster and changing the problematic settings. Upon saving the cluster changes, the assigned worker nodes will automatically reconfigure.

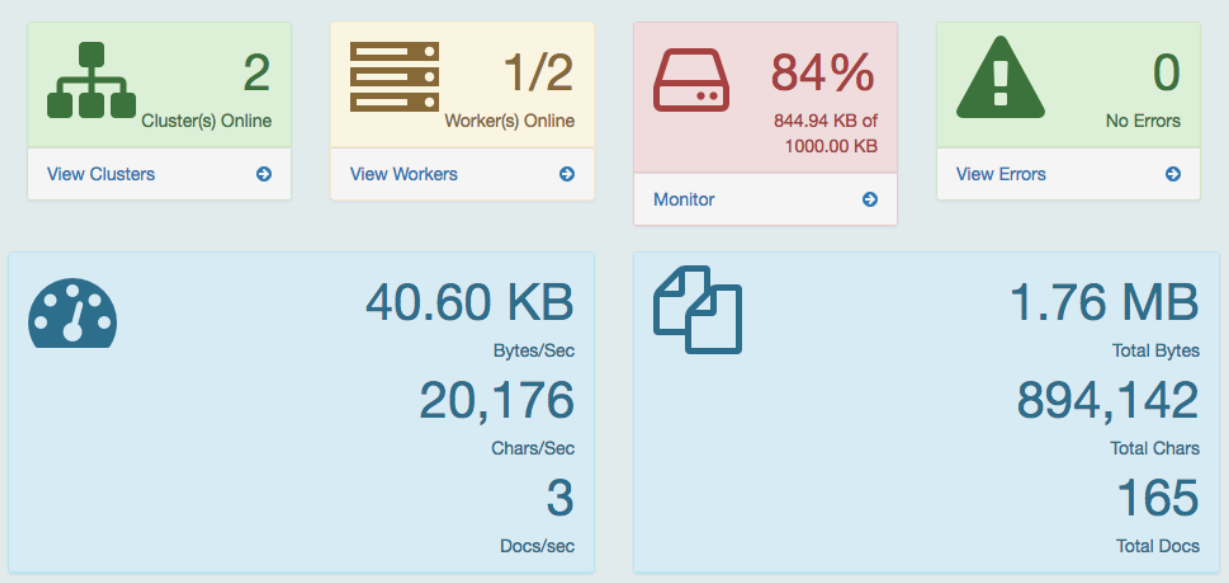
Monitoring activity

The manager user interface allows you to monitor all i2 TextChart Server activity for performance analysis, or to know when an error has occurred.

System overview (Dashboard)

To see an overview of the operation of the i2 TextChart Server system, navigate to the Dashboard page:

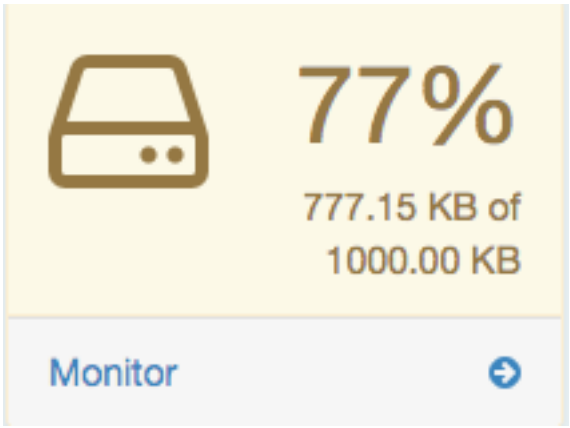
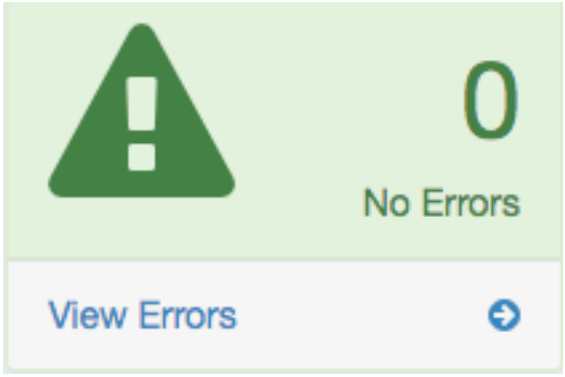
© N. Harris Computer Corporation



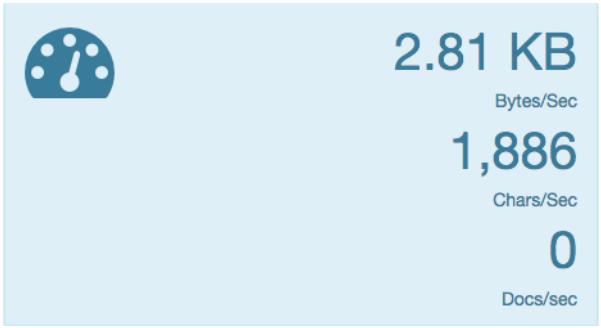
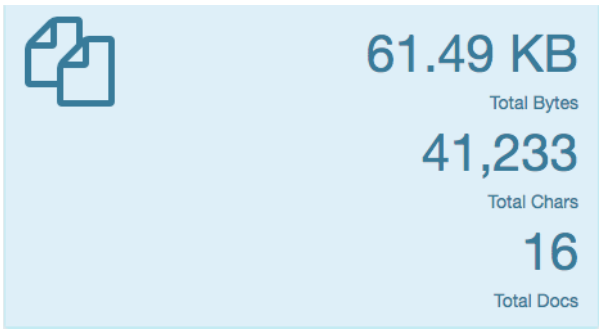
The panels in this display are updated every few seconds. They provide the following information:

Panel	Description
	Shows the number of clusters that are active and ready to accept documents.

© N. Harris Computer Corporation

Panel	Description
	Shows the number of worker nodes that are active, relative to the number of registered nodes.
	Shows the amount of data queued for processing. If a maximum queue size has been specified, shows the percentage of that size that is queued. If no maximum size is set, shows the percentage of disk space used on the file system where the documents are queued.
	Shows the number of processing errors that have occurred.

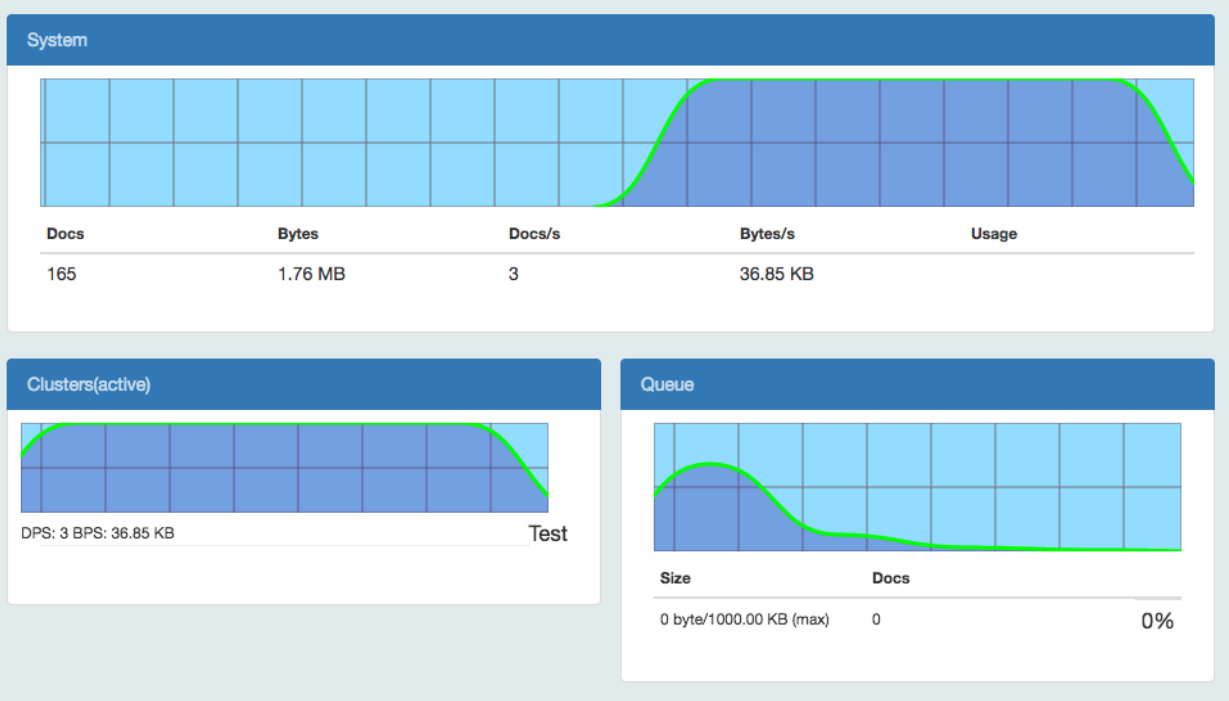
© N. Harris Computer Corporation

Panel	Description
	Shows the overall throughput performance of the i2 TextChart Server system since manager startup.
	Shows the amount of data and the number of documents that have been processed through the i2 TextChart Server system since manager startup.

Real-time monitoring (Monitor)

To see a real-time graphical view of system performance, navigate to the Monitor page:

© N. Harris Computer Corporation

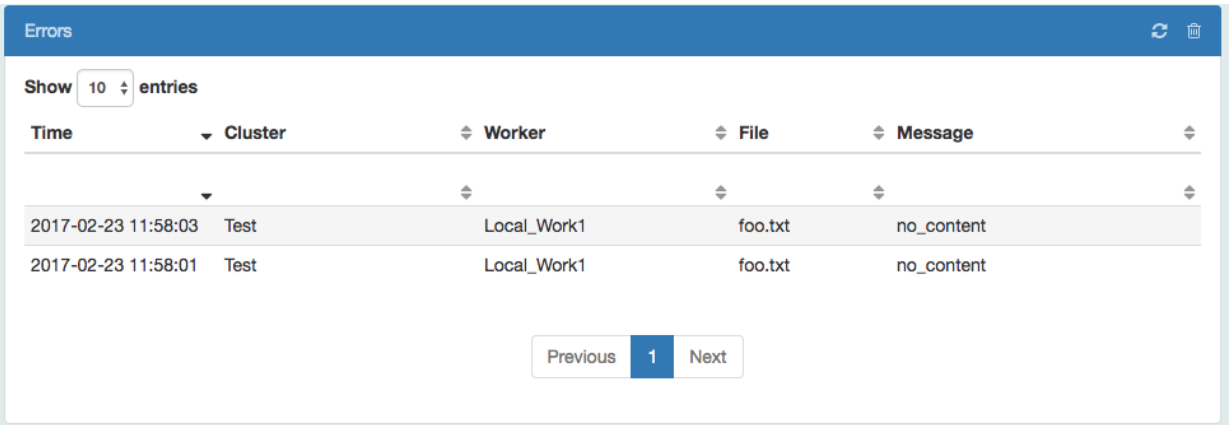


The panels show different aspects of system performance:

- **System**: Shows thread utilization across the entire system.
- **Clusters**: Shows thread utilization on a per-cluster basis.
- **Queue**: Shows the overall amount of data that has been queued for processing.

Errors

If any errors occur during document processing, a message appears in the Errors page:



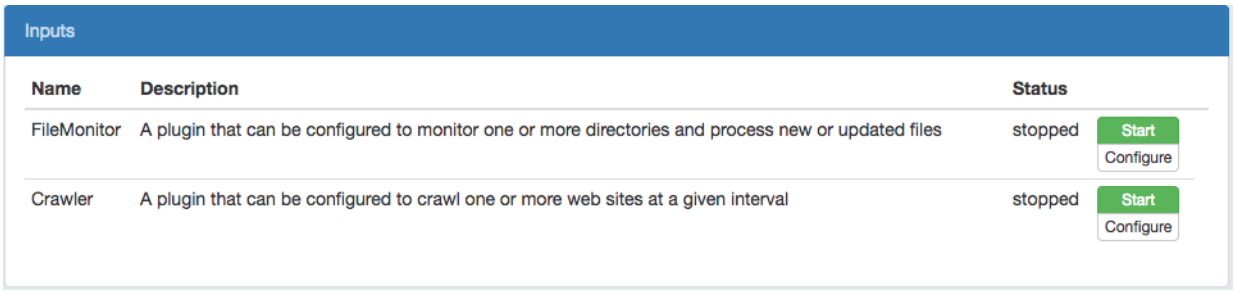
© N. Harris Computer Corporation

The panel shows when the error occurred, which cluster and worker node it happened on, the name of the file being processed, and a message indicating the type of error.

The table shows only the simple filename, but you can hover over the name to see the full path. Click the trashcan icon to clear the error log.

Processing documents using input connectors

The i2 TextChart Server manager can discover and queue documents for processing through small pieces of dynamically loaded software called *input connectors*. These connectors run on the manager, and you can administer them through the manager.



i2 TextChart Server is supplied with the file monitor input connector.

File monitor

The file monitor input connector lets you set up one or more directories on the manager to be monitored for new and changed files. These files are then queued up for processing by a specified cluster of worker nodes.

To set up the file monitor, open the Inputs page of the manager's user interface. You'll see a list of the available input connectors. Find the **FileMonitor** connector and click **Configure**.

The page lists all the directories that are currently being monitored, with the option to edit or delete directories from the list. To add a directory to monitor, click **Add**.

Enter the following information about the directory to be monitored:

- **Name**: A descriptive name for the directory. Use a name that helps you to identify the type of content that is being processed.
- **Path**: The absolute path to the directory, including the directory itself.
- **Recurse**: Select this option to explore not only the directory at the specified path, but also any subdirectories.
- **Doc ID**: For each file in the directory, a document identifier is passed to the output connector configured for the cluster. This setting determines what the identifier looks like:
  - **relative**: The document identifier is the path to the file from the monitored directory. For example, if the directory `/a/b/c` is monitored, and a file is found at `/a/b/c/d/file.txt`, then the document identifier is `d/file.txt`.
  - **absolute**: The document identifier is the absolute path of the file.
  - **name**: The document identifier is the name of the file, without any preceding directory path.

© N. Harris Computer Corporation

- **Format Conversion**: Select this option to submit files as raw data, so that i2 TextChart Server determines their encoding and format automatically. When the option is not selected, the server assumes the files to contain only UTF-8 encoded text.
- **Interval**: The monitor interval. Use **monitor** to process any new or changed files in the directory path immediately, or specify a time interval in the form `DDdHHhMMmSSs`, where `DD` is the number of days, `HH` is the number of hours, `MM` is the number of minutes, and `SS` is the number of seconds.

  You must specify at least one unit. For example, the following are all valid interval specifications: `4d3h`, `30m`, `1d12h30m20s`.
- **Cluster**: The cluster to use for processing. If left blank, the default cluster is used.
- **Post-process**: Determines what happens to a file after it is processed. The options are:
  - **none**: Do nothing to the file after it is processed.
  - **delete**: Delete the file after it is processed.
  - **move**: Move the file to another directory after it is processed. The **Move Path** setting determines the destination.
- **Move Path**: If **Post-process** is set to **move**, then this path determines where the processed files move to. If files were found in subdirectories of **Path**, that structure is maintained so that the resulting directory mirrors the original source directory.

To start the file monitor input connector, click **Start**. To learn how to start the file monitor when the manager starts up, see Manager Settings.

Storing results using output connectors

i2 TextChart Server worker nodes can store the results of extraction processing using small, dynamically loaded modules called *output connectors*. Output connectors are attached to a particular cluster, and you can administer them per-cluster through the manager.

i2 TextChart Server is supplied with two output connectors: the file writer, and the Elasticsearch connector.

File writer

The file writer output connector instructs the workers in a cluster to write the results of extraction processing to a particular file system directory in XML or JSON format. Where a cluster has several workers, it's common to use a network directory that's mounted in the same location on all worker nodes.

To use the file writer, configure the target cluster and select the Output tab. From the drop-down menu, select **FileWriter**. Then, click **Configure** to customize its settings:

- **Directory**: The absolute path to the directory where the result files will be written.
- **File Type**: Select **xml** or **json** output format.

The file writer generates filenames from the name portions of the incoming document identifiers. All output is placed in the specified directory; no subdirectories are created.

Elasticsearch

The Elasticsearch output connector instructs the workers in a cluster to send extraction results to an existing Elasticsearch database instance.

© N. Harris Computer Corporation

To use the Elasticsearch connector, configure the target cluster and select the Output tab. From the drop-down menu, select **Elasticsearch**. Then, click **Configure** to customize its settings:

- **Database host**: The hostname or IP address of the server running the Elasticsearch database instance.
- **Database port**: The network port on which the Elasticsearch server is listening.
- **Database cluster**: The name of the Elasticsearch cluster to use.
- **RFO index**: The name of the index in which to store the full data associated with a processed document. ("RFO" stands for `RosokaFullObject`.) The connector creates one entry per document in this index, using the document identifier as the key.
- **Entity index**: The name of the index in which to store the extracted entity data associated with a processed document. The connector creates one entry per document in this index.
- **PSO index**: The name of the index in which to store the extracted PSO (relationship) data associated with a processed document. The connector creates one entry per PSO in this index.
- **Store tokens (RFO)**: Select this option to store all of the individual tokens that were identified during extraction from the entry in the RFO index, alongside the full data. Due to the potentially high volume of data, the default is to not store tokens.
- **Store text (RFO)**: Select this option to store the original text of the entry in the RFO index. This is the text that was used as the input to the extraction engine, after encoding and formatting took place. By default, this text is not stored.
- **Store gloss (RFO)**: Select this option to store the "gloss" - that is, the rough translation - of the entry in the RFO index. By default, the gloss is not stored.
- **Store PSO (RFO)**: Select this option to store the PSO (relationship) data from the entry in the RFO index, alongside the full data. By default, this information *is* stored.

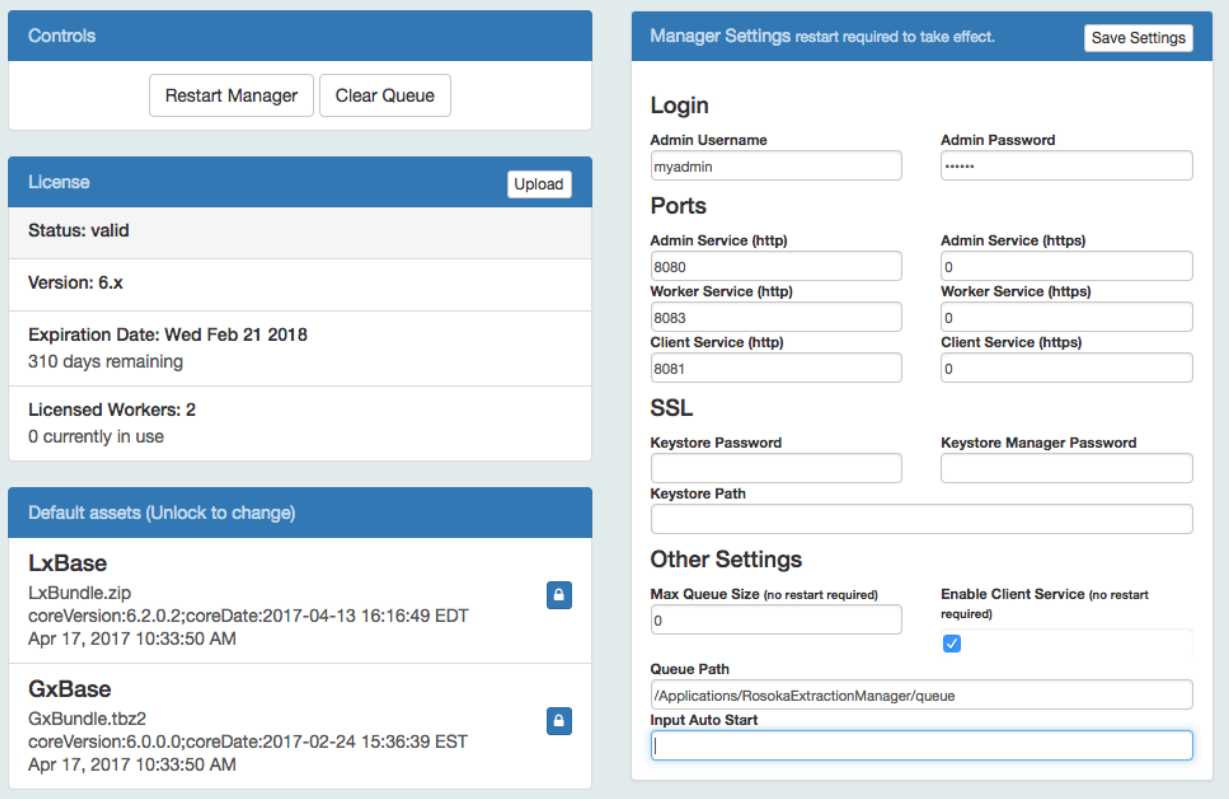
You must specify at least one index name. Specifying more than one index causes the connector to populate each index with the appropriate data.

Manager settings

The TextChart Server manager software supports a number of global settings that you can modify through the Settings tab in the administration user interface.

You can use the manager to update your license, to upload new versions of the LxBase and GxBase, to provide access to SSL certificates, to change the ports on which clients connect to TextChart Server, and more.

© N. Harris Computer Corporation



Controls

Use the Controls panel to restart the manager or to clear the processing queues:

- **Restart Manager**: Forces the manager process to restart and adopt your changes to settings. Restarting the manager pauses all processing in all clusters, and resets the status of any running input connectors.
- **Clear Queue**: Clears the queues associated with each cluster. All pending documents are removed. If a worker has become unresponsive, clearing the queue can help to restore the system.

License

The License panel shows the current license status of the i2 TextChart Server system, which *requires* a valid license file to operate. If the license is missing, expired, or not valid, then no worker nodes can connect to the manager and no documents will be accepted for processing.

To install a new license file that you received from i2 Group, click **Upload** and select it. The License panel is updated to reflect the new license status.

Default assets

The Default Assets panel shows the status of the default LxBase, GxBase, and SSP templates assets.

© N. Harris Computer Corporation

When you create a cluster in the manager, it's configured to use the default LxBase and the default GxBase. To use a different asset from the default, you can change the cluster configuration settings on the Clusters page.

Use the Default Assets panel to upload a new LxBase, GxBase, or SSP templates bundle for use by all clusters for which a custom asset is not configured. Be aware that uploading new assets here causes all clusters to be updated immediately. All worker nodes that belong to these clusters will pause to download and reinitialize with the new assets.

To upload a new asset, click the appropriate lock icon, and then click **Upload**. The file that you select in the resulting dialog is uploaded and verified, and the asset information is updated to match it.

SSP templates

The manager supports uploading a ZIP file containing SSP (Semi-Structured Processing) template XML files as the `ssp_templates` asset. When the asset is deployed to a worker, the ZIP is extracted and the template files are made available to the extraction engine for structured document parsing.

The ZIP file is identified as an SSP templates bundle if its filename starts with `ssp` or `rssp` (case-insensitive). You can produce a suitable bundle from the `RsspBundle` Maven module in rsk-api-core, or assemble your own ZIP containing template XML files at the root level.

Settings

The Manager Settings panel allows you to edit the global parameters for the i2 TextChart Server manager. In the following table, the settings marked with an asterisk (*) require that the manager is restarted for changes to take effect.

Property	**Description**
Admin Username*	The user name for logging into the manager UI. If the authorization is set to use LDAP, this field does not appear.
Admin Password*	The password for the admin user. If the authorization is set to use LDAP, this field does not appear.
Admin Service (http)*	The unsecured network port on which to access the admin service, or 0 to disable HTTP access.
Worker Service (http)*	The unsecured network port on which to access the worker service, or 0 to disable HTTP access.
Client Service (http)*	The unsecured network port on which to access the client REST services and user interface, or 0 to disable HTTP access.
Admin Service (https)*	The secure network port on which to access the admin service, or 0 to disable HTTPS access.
Worker Service (https)*	The secure network port on which to access the worker service, or 0 to disable HTTPS access.

© N. Harris Computer Corporation

Property	Description
Client Service (https)*	The secure network port on which to access the client REST services and user interface, or 0 to disable HTTPS access.
Keystore Password*	The password for the SSL keystore.
Keystore Manager Password*	The password for the SSL keystore manager. If not specified, the manager uses the keystore password.
Keystore Path*	The absolute path to the SSL keystore file.
Max Queue Size	The maximum size of the files that can be queued up for processing. If set to zero, then the free space on the file system that contains the queue directory is used as the maximum. By default the size is in bytes, but you can also use typical suffixes (G = gigabyte, M = megabyte, k = kilobyte).
Enable Client Service	Select this option to enable the client REST services.
Queue Path*	To absolute path to the directory that contains queued documents.
Input Auto Start	A comma-separated list of input connectors that should be started automatically at manager startup. If the list is empty, all the installed input connectors start automatically.

See Configuring the manager for LDAP authentication for more information about LDAP, and Configuring SSL for i2 TextChart Server for more information about setting up secure HTTPS access.

Manager REST services

This topic describes the REST services that are available on the manager server at `http://<manager_host>:<client_port>/manager/services/client`.

The following guidance applies to all the endpoints apart from `getClusters`:

- Endpoints support only POST methods.
- Endpoints accept multi-part MIME content.
- Endpoints produce XML (`text/xml`) or JSON (`application/json`) results, depending on the MIME type passed in the `Accept` header.
- The `flags` parameter to processing is a string of space-separated strings that control which parts of the extraction analysis are passed back to the caller. Use this parameter to remove information from the results to improve performance. To accept the default behavior, pass an empty string.

© N. Harris Computer Corporation

  - `-text`: do not include the original text (included by default).
  - `+gloss`: include the glossed text (not included by default).
  - `+tokens`: include the raw token list (not included by default).
  - `-pso`: do not include the relationships (included by default).
- The results are returned in complex objects that contain the following top-level fields:
  - `status`: The queuing or processing status.
  - `errorFound`: `true` if the processing cluster reports an error; `false` otherwise.
  - `results`: The extraction analysis results, as a `RosokaFullObject`.

getClusters

A GET method that returns a JSON-formatted list of the available clusters by name.

processString

Accepts simple text content for processing and returns results immediately to the caller.

- `cluster`: The name of the cluster to use for processing. If unspecified or blank, use the default cluster.
- `flags`: A space-separated list of flags that control what content is returned from extraction.
- `content`: The text to be processed.

processRaw

Accepts raw, binary content for processing and returns results immediately to the caller. Content is analyzed for encoding and format before being converted to simple text for extraction analysis.

- `cluster`: The name of the cluster to use for processing. If unspecified or blank, use the default cluster.
- `flags`: A space-separated list of flags that control what content is returned from extraction.
- `content`: The raw byte stream to be processed.

ingestString

Accepts simple text content and queues it for processing. Results from extraction analysis are sent to the output connector for the specified cluster. The caller receives only a status in return.

- `cluster`: The name of the cluster to use for processing. If unspecified or blank, use the default cluster.
- `meta`: A JSON-formatted set of key-value pairs to pass to the output connector. Some output connectors look for the `_docID` key to determine the name or URI of the incoming text.
- `content`: The text to be processed.

For example:

```
curl -X POST http://127.0.0.1:8011/manager/services/client/ingestString -F
 'content="Terry Miller flew to Rome, Italy on March 12, 1991. Mr. Miller is
 an employee of Tisher & Associates and travels frequently for business."' -
F 'meta={'_docID':'record001'}'
```

© N. Harris Computer Corporation

ingestFile

Accepts raw, binary content and queues it for processing. Results from extraction analysis are sent to the output connector for the specified cluster. The caller receives only a status in return.

- `cluster`: The name of the cluster to use for processing. If unspecified or blank, use the default cluster.
- `meta`: A JSON-formatted set of key-value pairs to pass to the output connector. Some output connectors look for the `_docID` key to determine the name or URI of the incoming stream.
- `content`: The raw byte stream to be processed.

Under the hood

Not all i2 TextChart Server settings are available via the management UI. Some manager settings (and all settings on worker nodes) can only be changed by editing a local properties file. Changing any of these files requires a restart of the appropriate service.

Manager settings

The `manager.properties` file is located in the manager's installation directory. It's a simple Java settings file (one setting per line, `<name>=<value>`) that you can open in any text editor. It contains these properties that you can change in the file but not in the user interface:

Property	Description
authType	Determines the type of authorization for accessing the administration UI. Options are `form` (the default), which uses a single username-password combination (from the `adminUser` and `adminPassword` properties), or `ldap`, which uses the LDAP server defined in the `conf/ldap.conf` file for authentication and authorization.
authGroup	(LDAP only) Specifies the LDAP group of which the authenticated user must be a member in order to gain access to the UI. Setting this to `**` (the default) authenticates the username and password but does *not* require any specific group membership.

Worker settings

The `worker.properties` file is located in the worker's installation directory. It's a simple Java settings file (one setting per line, `<name>=<value>`) that you can open in any text editor.

Typically, it is not necessary to edit the properties for a worker node. However, if you need to change a node port, or if the host or port of the manager changes, then you can make the edits to this file without having to reinstall the worker software.

© N. Harris Computer Corporation

Property	Description
name	A descriptive name for the worker.
ip	The hostname or IP address of the worker.
port	The port on the worker for communicating with the manager.
secure	`true` to use HTTPS communication between the worker and the manager. Secure communication requires a keystore file on both the manager and the worker.
keyStorePath	The absolute path to the keystore file for SSL communication. If secure communication is enabled but this parameter is not set, then the default is to use the `keystore.jks` file in the worker installation directory.
keyStorePassword	The password for the SSL keystore.
keyManagerPassword	The manager password for SSL keystore. If not set, this value defaults to the keyStorePassword value.
slots	The number of threads to use for extraction processing.
managerHost	The hostname or IP address of the manager.
workerServicePort	The port on the manager for communicating with the worker.
heapSize	The maximum Java heap usage for the worker process when started as a Linux or Windows service. Use the same syntax as would be passed to the `-Xmx` option. For example, to specify a heap size of 6 GB, use `6144m`. If not present, defaults to `4096m`.

General settings

Workers and the manager all contain a `logger.properties` file that you can use to configure service logging. By default, both services write a rotating set of log files to the `logs` directory in their respective installation directories.

To change how logging operates on either system, edit the `logger.properties` file in a text editor.

© N. Harris Computer Corporation

Configuring the manager for LDAP authentication

If a site-wide authentication and authorization system is available via LDAP (Lightweight Directory Access Protocol), then you can set up the manager to use it for authorizing access to the management UI.

To set up LDAP, first set the `authType` property in the `manager.properties` file to the value `ldap`:

```
authType=ldap
```

To restrict access to the management UI to members of a particular LDAP group, set the `authGroup` property to the name of the group that's allowed to access it. For example, if the name of the allowed group is `RosokaUsers`, then set the property like this:

```
authGroup=RosokaUsers
```

To allow *all* authenticated users to access the management UI, either don't set the `authgroup` property, or set it to the default value, `**`.

Save these changes to the `manager.properties` file.

Next, you need to configure the manager to access the LDAP server by editing the sample `conf/ldap.conf` file to adapt to the site environment.

Note: Interfacing with LDAP can be complex, and i2 recommends contacting your system administrator for help with the settings in the LDAP configuration file.

To use i2 TextChart Server with an LDAPS (LDAP with SSL) server, see Configuring SSL for i2 TextChart Server.

After you finalize changes to `manager.properties` and `conf/ldap.conf`, restart the manager service to make them take effect.

Configuring SSL for i2 TextChart Server

i2 recommends using SSL to secure communications between TextChart clients and the manager server, and between the manager and its worker nodes. The instructions in this topic describe how to create a self-signed certificate for use with the system.

About this task

Enabling SSL for i2 TextChart Server involves the following sequence of actions:

1. Create a certificate `keystore` file.
2. Set keystore and keystore manager passwords.
3. Enable SSL through the i2 TextChart Server management UI, and on the worker nodes.

The procedure below describes this sequence in more detail.

Procedure

To enable SSL functionality with a self-signed certificate, perform the following steps:

1. Use the `keytool` command provided with the Java JDK to create a new keystore:

   ```
   keytool -keystore keystore -alias jetty -genkey -keyalg RSA
   ```

   Answer the prompts as required for your site, and make a note of the password that you provided.

© N. Harris Computer Corporation

When the command finishes, it generates a `keystore` file in the directory that you ran it from.

Note: On Microsoft Windows, you might need to add the JDK binary directory to your path in order to run the `keytool` command:

```
set PATH=%PATH%;"c:\Program Files\Java\jdk1.8.0_65\bin"
```

Adjust the path and version number to match your JDK installation.

2. Copy the `keystore` file to the manager server, and place it in the `conf` subdirectory of the i2 TextChart Server installation directory.
3. In the i2 TextChart Server management UI, set the values of the following fields:
   - **Key Store Path**: The absolute path to the `keystore` file.
   - **Key Store Password**: The password that you entered when you created the keystore.
   - **Key Manager Password**: The manager password that you entered when you created the keystore. If this password is the same as the **Key Store Password**, leave the field blank.
4. You can enable SSL separately for three different areas of communication:
   - To use SSL for client-to-manager communication, enter a valid port number in the **Client Service (https)** field.
   - To use SSL for admin-to-manager communication, enter a valid port number in the **Admin Service (https)** field.
   - To use SSL for manager-to-worker communication, enter a valid port number in the **Worker Service (https)** field.

   In all three cases, you can disable unsecure communication by setting the corresponding **http** field to `0`.
5. For SSL manager-to-worker communication, you also need to copy the `keystore` file to the installation directory on *every worker node*, and edit the `worker.properties` file to add the following properties:

   ```
   secure=true
   keyStorePassword=<password>
   keyManagerPassword=<manager_password>
   keyStorePath=/opt/RosokaExtractionWorker/conf/keystore
   ```

   Note: Set `keyManagerPassword` only if the manager password is different from the keystore password.

Any change to client- or admin-to-manager communication requires a restart of the manager server. Any change to manager-to-worker communication requires a restart of the manager *and* all workers.

If you access the manager UI or the client UI via SSL with a self-signed certificate, it is likely that the browser display a warning, and you will have to allow the browser to navigate to the page. To remove the warning, use a certificate that's signed by a valid certification authority.

Similarly, a self-signed certificate will generate warnings if you use the REST API over SSL through most common communications libraries (including the i2 TextChart Server client tools). To use a self-signed certificate without error in this context, install it as a trusted certificate in the client Java system.

i2 TextChart Data Access

i2 TextChart Data Access provides a scalable environment for performing data extraction from free-form text. Important features include:

© N. Harris Computer Corporation

- Turnkey operation.
- Straightforward REST interface for processing text with Java and C/C++ clients.
- Easily scale up or down depending on throughput requirements.
- Browser-based administration interface for setup, management, and monitoring.
- Dynamic input and output connectors for interfacing with document sources and databases.

TextChart Data Access Server (TDAS)

An installation of TextChart Series 7 may include installation of the TextChart Data Access Server (TDAS). This service acts as the central data access point of for systems using TextChart Analysis in an enterprise configuration.

The TDAS provides these important services:

- A collection of REST endpoint for submitting and accessing data.
- A browser-based user interface that allows an administrator to set up, manage, and monitor the entire system from a single point.

REST API

The TextChart Series 7 TextChart Data Access Server (TDAS) provides a collection of REST services for submitting documents for processing. For administrators and developers that want to integrate third-party applications with the TDAS, the REST endpoint documentation is provided as a collection of javadocs that provide descriptions of each endpoint.

Installation

Before beginning the TDA installation, please make sure you have the following information, which you will need to enter during the installation process:

- **Installation Directory**: A directory where the software will be copied. The directory's file system should have at least 2GB of free space.
- **Client Port**: A network port where the REST API will communicate. Choose an unused network port on your system.
- **Admin Port**: A network port where the administration UI will be hosted. Choose an unused port on your system.
- **Manager Host**: The hostname or IP address of the system on which you installed the Manager software.
- **Worker Service Port**: The network port on the Manager Node you specified for the *Worker Service Port* field.
- **Worker Name**: The name of this Worker Node. Typically, this value would be the network hostname of the node. However, the name can be arbitrary.
- **Worker Port**: A network port on the Worker node that the Manager can use to communicate with the Worker. Select an unused port on the Worker Node.
- **Threads**: The number of threads to use when processing documents on this node. Each thread will process one document, so the thread count determines how many simultaneous documents can be processed. Typically, one would set a thread count to the same number of CPU cores that are present on the system. The default value of "0" will automatically use all available cores for processing.

© N. Harris Computer Corporation

On Linux, you must run the TDAS installation with root privileges (use the **sudo** command to start the Java installation process).

After installation, you will need to start the service. The service is called **rosokarda** on Linux or **RosokaDataAccess** on Windows. The service will be set to start automatically at OS startup.

Start the service by issuing the command appropriate for your system type. For example,

```
sudo service rosokarda start
```

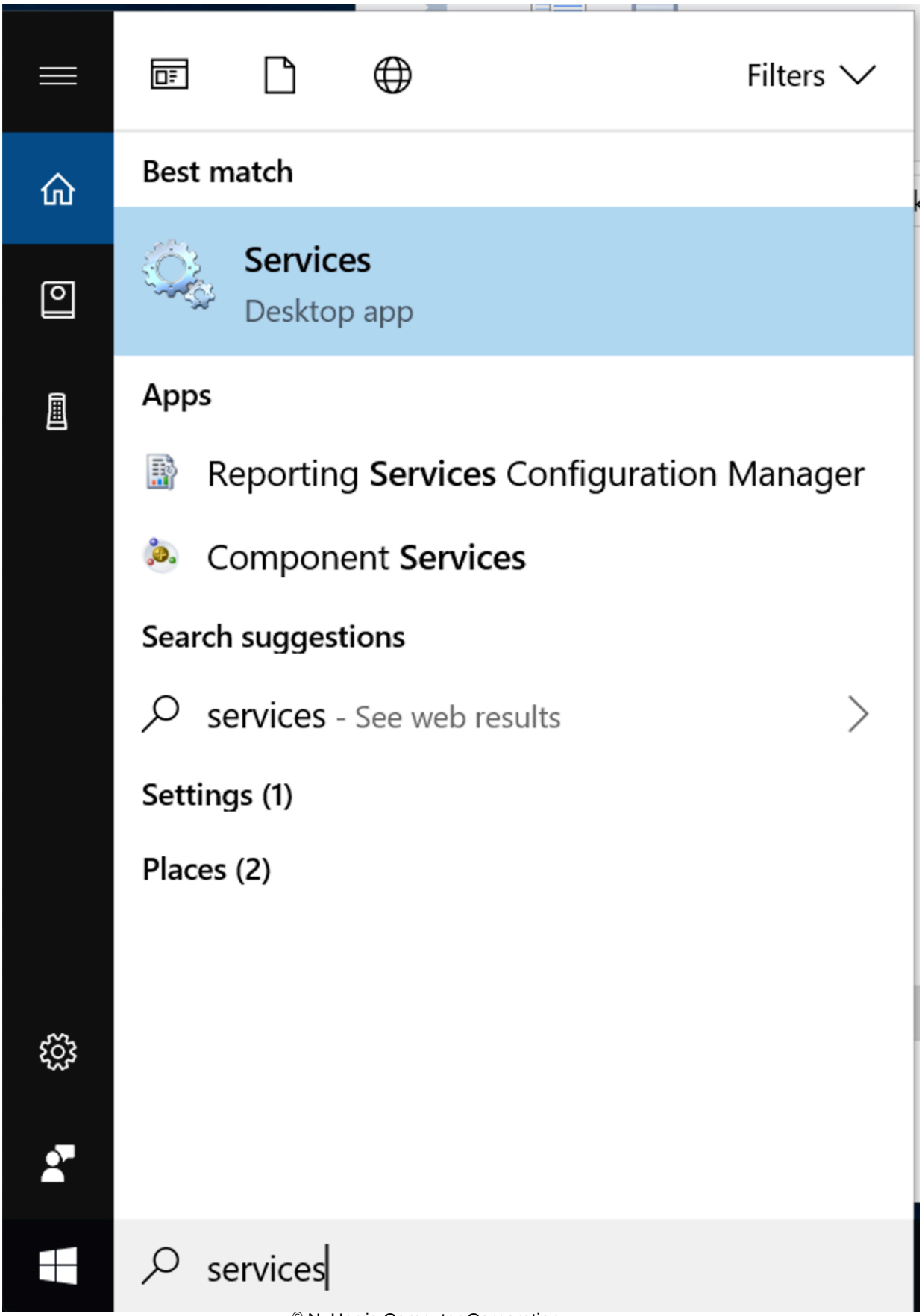
will start the service on most variants of Linux, while on Windows, use a command like

```
net start RosokaDataAccess
```

On Windows, one may also use the Task Manager or the Services control panel to start, stop, and manage the service.

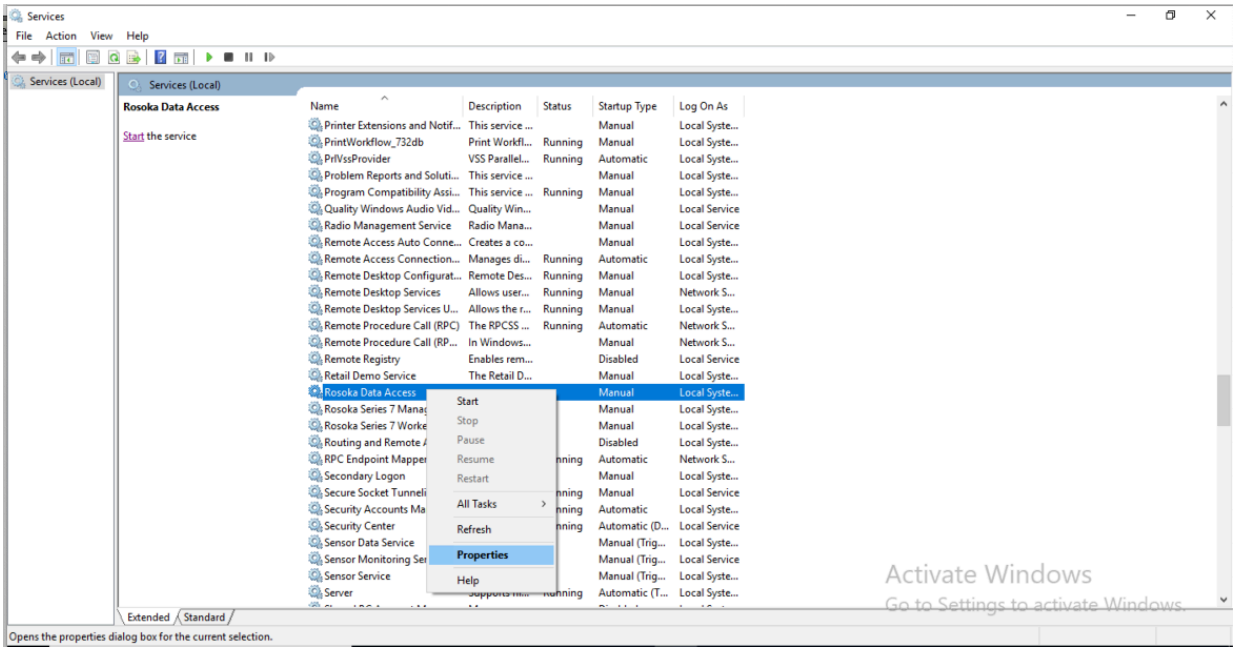
On Windows the services need to be configured to restart the services. To do so, access the services by launching the services manager.

© N. Harris Computer Corporation



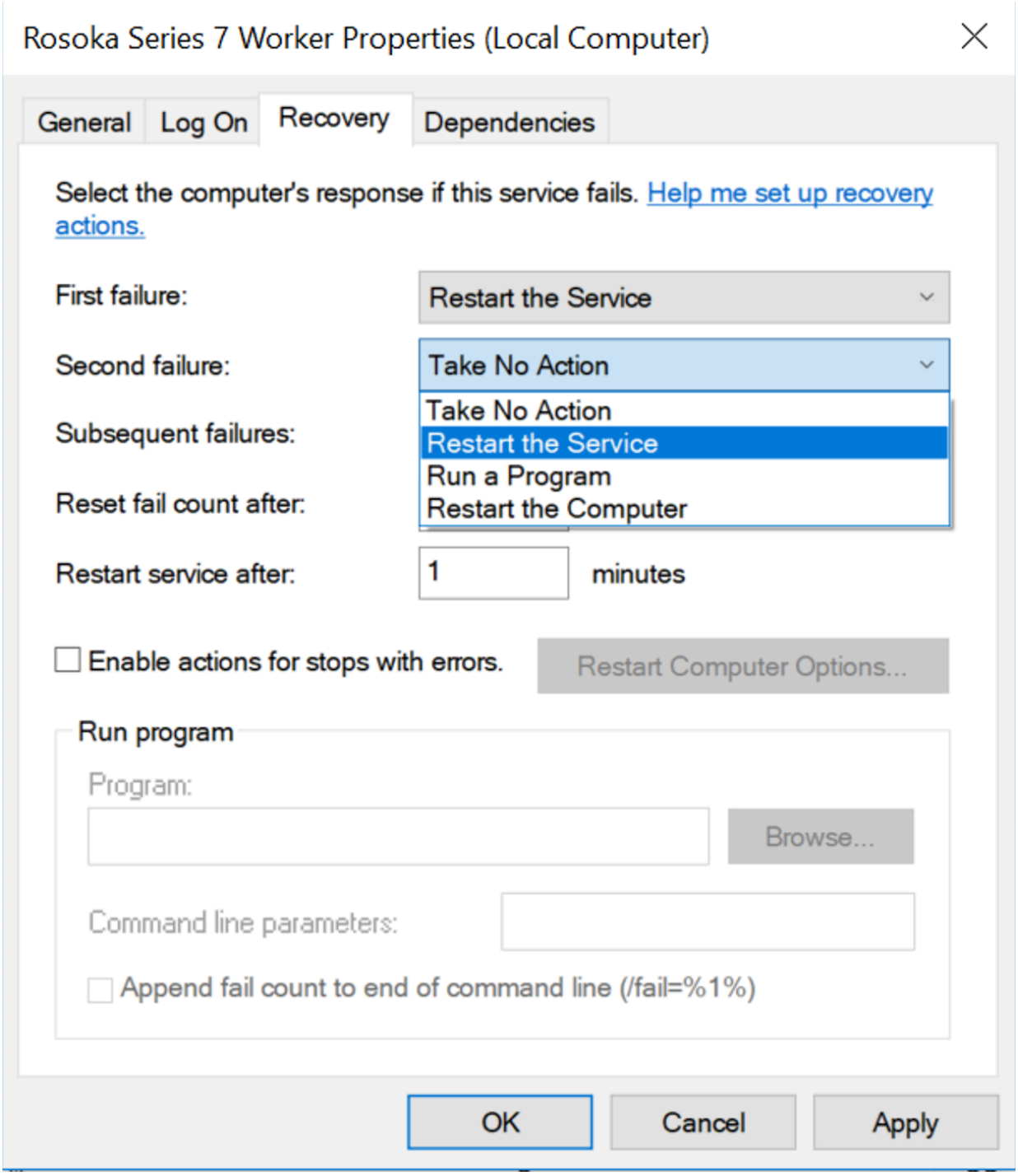
© N. Harris Computer Corporation

Locate the "Rosoka Data Access" service and select the properties for the service.



In the properties dialog go to the "recovery" tab set the "First Failure", "Second Failure", and "Subsequent Failures" to "Restart the Service" using the selector.

© N. Harris Computer Corporation



Non-standard Installations

TextChart Series 7 was designed to be installed and run on Linux-based server systems. Due to the fact that both the installation procedure and the software itself are Java-based, it is possible to install and run the system on other operating systems, such as Microsoft Windows or Mac OSX.

© N. Harris Computer Corporation

Automatically configured and started services are created on most variants of Linux and Microsoft Windows. For other operating systems, one must issue commands in a terminal window.

To run the TDAS software manually, follow these steps after installation:

1. Start a terminal or command-prompt window on the installation system.
2. Navigate to the installation directory such that the default directory in the terminal window is the installation directory. Then issue the following command:

   ```
   java -Djava.util.logging.config.file=logger.properties –jar
    RosokaDataAccessServer.jar
   ```

   The TDAS software will stay running in the terminal window.

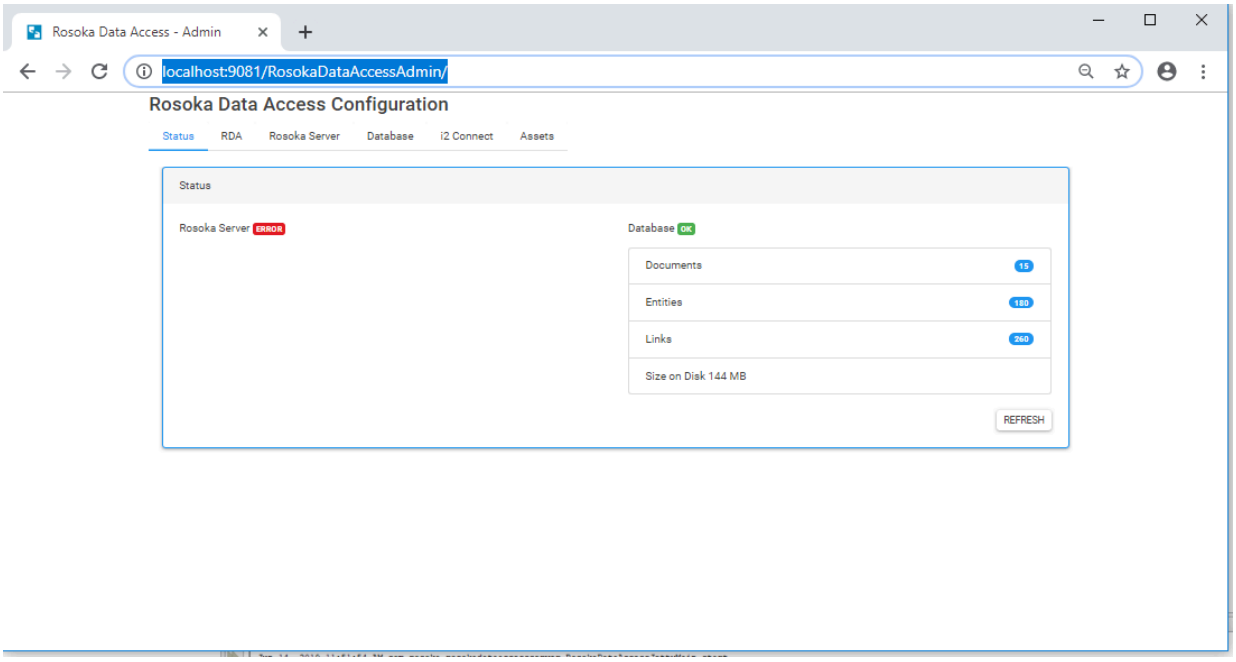
Getting started

At this point, you should have installed the TextChart Data Access software. This section will guide you through the steps to set up your data connections.

Bring up the admin user interface by opening a browser and navigating to `http://<tdas_host>:<admin_port>/RosokaDataAccessAdmin` (for example, `http://localhost:9081/RosokaDataAccessAdmin`).

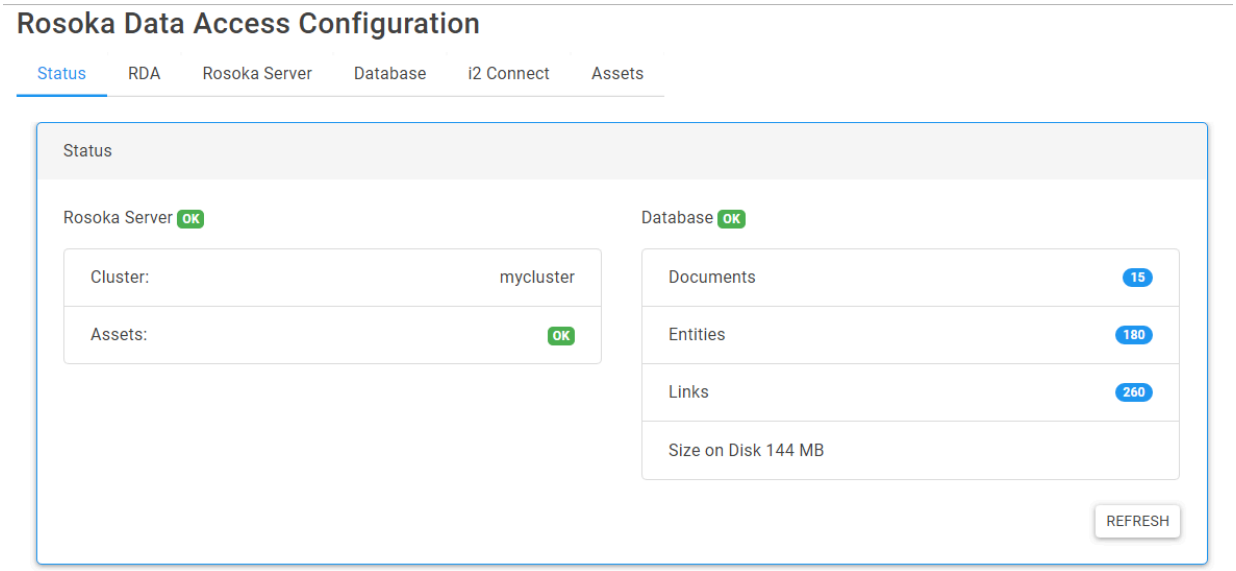
You may be asked to log into the administrator UI. The default user is "admin" and the default password is "rosoka". You may change these credentials at a later time.

When TDAS is first accessed you will get indicators letting you know that the connections to TextChart Manager and/or the database or not made. Once they are connected these are connected these will be turn to green status. The figure below is of the main page showing there is no connection to TextChart Manager but the database connection is green.

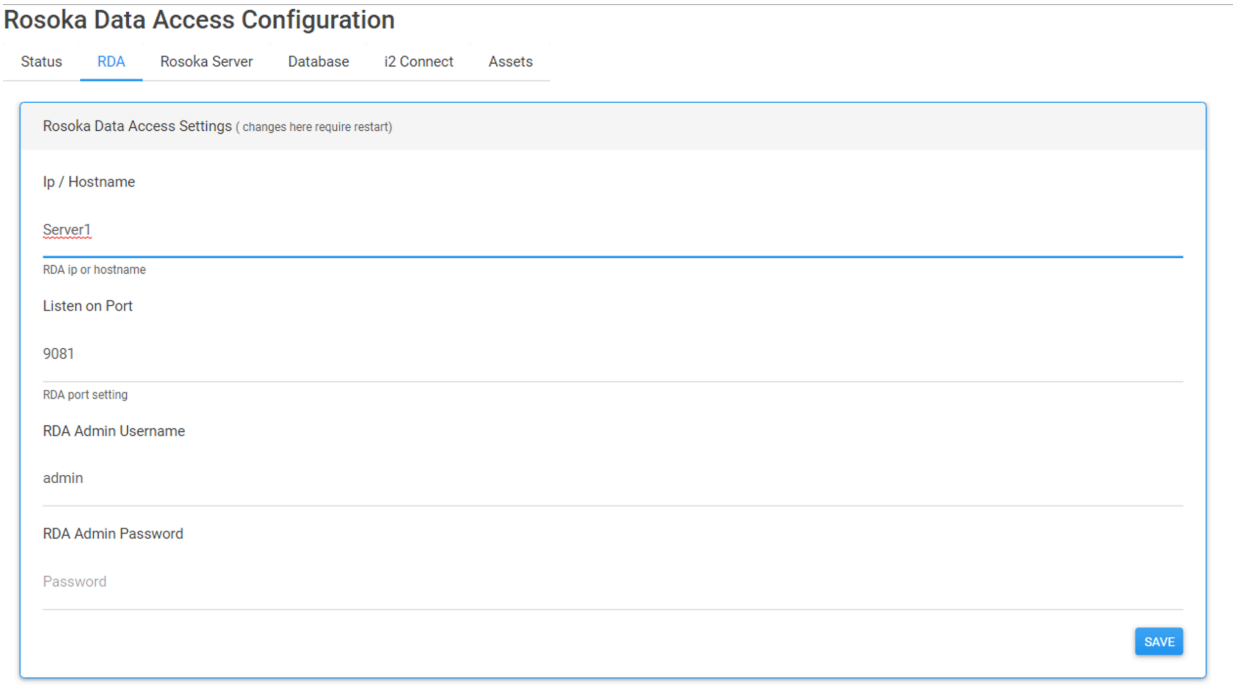


© N. Harris Computer Corporation

Once all of the connection are configured you will see the green status as shown.

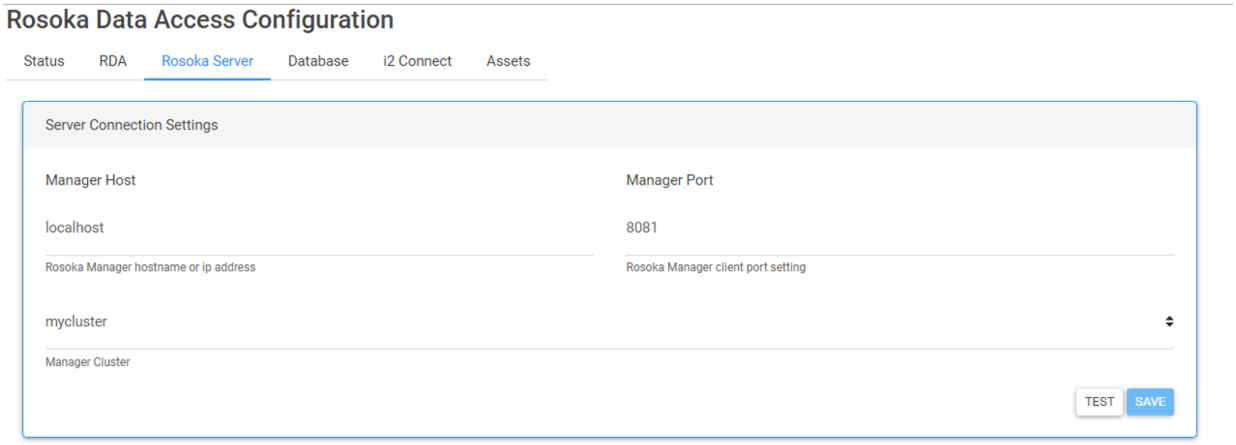


The **RDA** tab allows the RDAS server name (or IP) to be set the port to use, and the admin password.

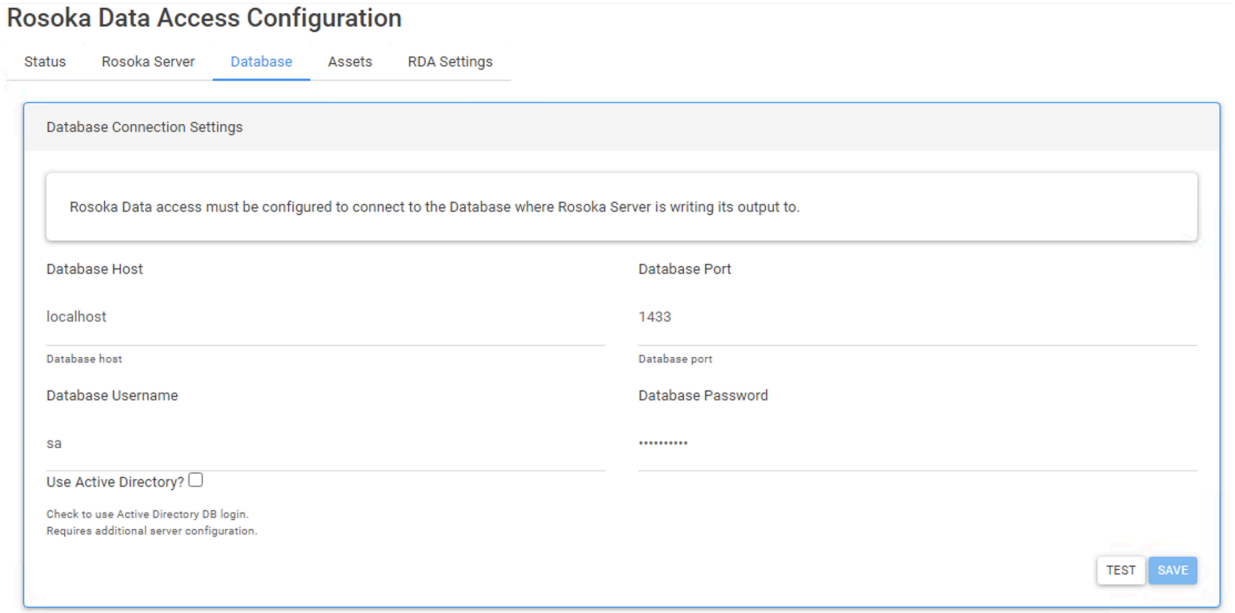


The **Rosoka Server** tab is used to set the properties to a Rosoka Server (Manager/Worker cluster) and test the connections.

© N. Harris Computer Corporation

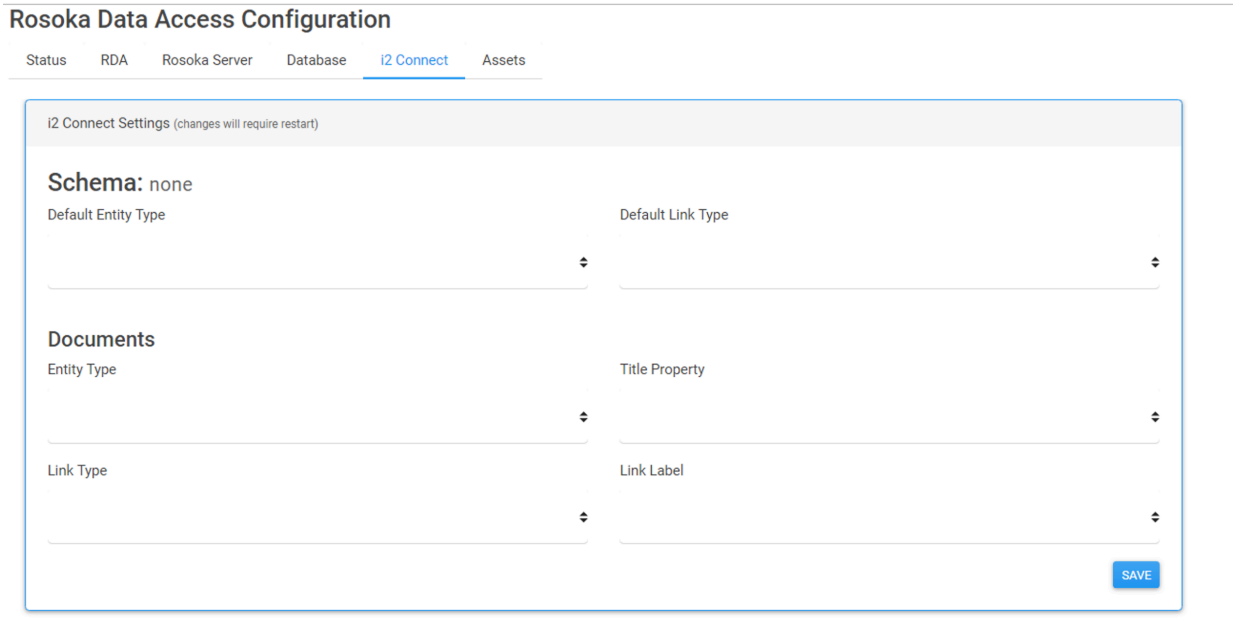


The **Database** tab is used to set and test the connection the database where processing results are stored.

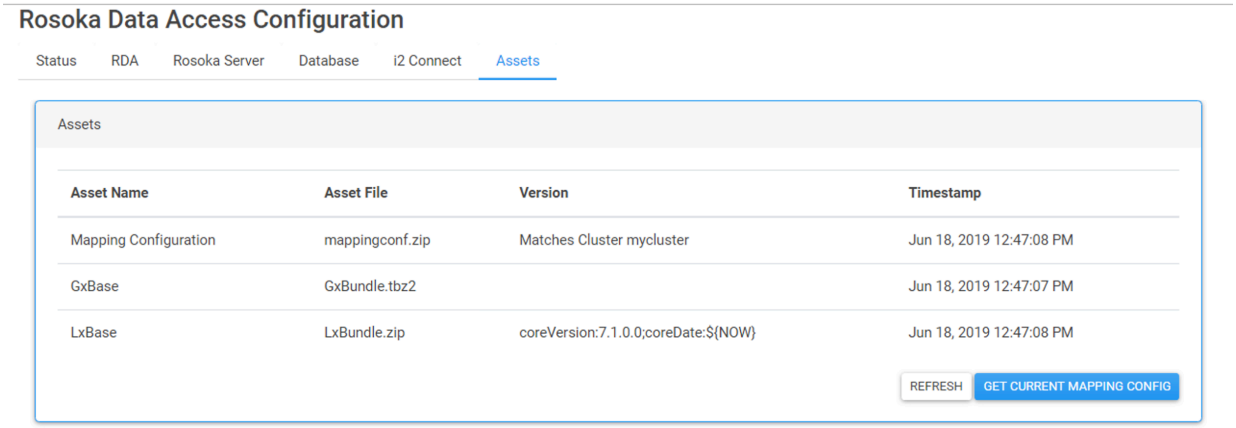


The **i2 Connect** tab allows you to configure an i2 connector.

© N. Harris Computer Corporation



The **Assets** tab provides information about the assets (GxBase, LxBase, and Mapping Configuration). The asset tab allows you to sync the assets to have the same assets used by Rosoka Server.



Logging

TDAS contain a `logger.properties` file that can be used to configure how service logging is done. By default, both services write a rotating set of log files to the `logs` directory in their respective installation directories.

© N. Harris Computer Corporation

To change how logging operates on either system, edit `logger.properties` with your favorite text editor.

Configuring TextChart Data Access Server for LDAP authentication

If a site-wide authentication and authorization system is available via LDAP (Lightweight Directory Access Protocol), then the TextChart Manager may be set up to access this directory to authorize access to the management UI.

To set up LDAP, first set the `authType` property in the file `manager.properties` to the value `ldap`, such as

```
authType=ldap
```

If you wish to restrict access to the TextChart Manager UI to members of a particular LDAP group, then also set the `authGroup` property to the name of the group that is allowed access to the TextChart Manager. For example, if the name of the allowed group is "TextChartUsers", then set the property such as

```
authGroup=TextChartUsers
```

To allow all authenticated users to access the Server UI, either do not set the property or set it to the default value, "`**`".

Save these changes to the `manager.properties` file.

Next, one must configure settings so that the TextChart Manager can access the LDAP server. To do this, edit the file `conf/ldap.conf` in the Manager installation. A sample file is provided, and one should only need to change the property values in order to adapt to the site environment. Because of the complexity of interfacing with LDAP, TextChart recommends contacting your site IT department for assistance in setting the values in the LDAP configuration file.

For using Studio with an LDAPS (LDAP with SSL) server, *Configuring SSL for TextChart Series 7* provides information for creating and managing SSL certificates and key stores.

After changes to both the `manager.properties` and `conf/ldap.conf` file have been finalized, restart the Manager service (either via the UI or via Linux service management methods) to cause the changes to take effect.

Configuring SSL

This step is not needed to run TextChart Series 7 if you will only be accessing the system using the non-secure HTTP protocol. However, if you wish to enable more secure communications between any TextChart Series 7 clients and the TextChart server, or between the TextChart server and its Worker nodes, then follow the instructions in this section, which describe how to create a self-signed certificate for use with the system.

Enabling SSL for TextChart Series 7 involves the following basic steps:

1. Create a certificate `keystore` file.
2. Set keystore and keystore manager passwords
3. Enable SSL functionality using the TextChart Series 7 Manager admin UI, and optionally enabling SSL functionality on some or all of the Worker nodes.

To enable SSL functionality with a self-signed certificate, perform the following steps:

1. Use the `keytool` program provided with the Java JDK to create a new keystore:

© N. Harris Computer Corporation

```
keytool -keystore keystore -alias jetty -genkey -keyalg RSA
```

Answer the prompts as required for your site. Make note of the password you enter for the key/cert generation process.

When this program completes, it will have generated a `keystore` file in the directory.

Note: On Microsoft Windows, you may need to add the JDK binary directory to your path in order to run the `keytool` command. If you cannot run the `keytool` program from the command line, then add the `bin` directory of your JDK installation to the path using a command like:

```
set PATH=%PATH%;"c:\Program Files\Java\jdk1.8.0_65\bin"
```

Adjust the path and version number to match your JDK installation.

2. Copy the `keystore` file to the Manager system, preferably in the `conf` sub-directory of the Manager installation directory.
3. Use the TextChart Series 7 Manager admin UI to set the values for the fields:
   - **Key Store Path**: the full absolute path to the keystore file.
   - **Key Store Password**: the password entered when creating the keystore.
   - **Key Manager Password**: the manager password entered when creating the keystore. If this password is the same as the Key Store Password, leave the field blank.
4. Determine whether you wish to use SSL for client-to-Manager communication, admin-to-Manager communication, or Manager-to-Worker communication. One can enable SSL on any combination of these three:
   - To use SSL for client-to-Manager communication, enter "0" for the **Client Service (http)** field, and enter a valid port number for the **Client Service (https)** field.
   - To use SSL for admin-to-Manager communication, enter "0" for the **Admin Service (http)** field, and enter a valid port number for the **Admin Service (https)** field.
   - To use SSL for Manager-to-Worker communication, enter "0" for the **Worker Service (http)** field, and enter a valid port number for the **Worker Service (https)** field.

   Save these values in the Manager admin UI. You will need to restart the Manager server, either via the UI or by restarting the service on the Manager node.
5. For SSL Manager-to-Worker communication, copy the `keystore` file to the Worker installation directory on *each Worker node*.
6. For SSL Manager-to-Worker communication, edit the `worker.properties` file in the Worker installation directory on *each Worker node* and add the following properties:

   ```
   secure=true
   keyStorePassword=<password>
   keyManagerPassword=<manager_password>
   ```

   The `keyManagerPassword` property only needs to be set if the manager password differs from the keystore password.

Any change to admin- or client-to-Manager communication requires a restart of the Manager server. Any change to Worker-to-Manager communication requires a restart of the Manager and all affected Workers.

If you access the Manager UI or the client UI via SSL with a self-signed certificate, it is likely that the browser will complain, and you will have to explicitly allow the browser to navigate to the page. This problem can be alleviated by having your certificate signed by a valid CA.

© N. Harris Computer Corporation

Similarly, if you attempt to use the REST API over SSL via most common communications libraries (the TextChart Series 7 Client Tools included), you will likely also have problems with a self-signed certificate. If you wish to use a self-signed certificate, then you should install the certificate in the client Java system as a trusted cert so that the communication can succeed without error.

i2 TextChart Studio

Welcome to i2 TextChart Studio. This software enables you to tune TextChart processing for your organization by customizing or creating the LxBases and GxBases that TextChart uses. Users can apply the new or updated LxBases and GxBases to both TextChart and TextChart Premium

Note: i2 TextChart Studio is only supported with the Google Chrome browser. Other browsers may cause unusual errors when using TextChart Studio.

Administering TextChart Studio

This part of the documentation describes how to install, configure, and secure access to i2 TextChart Studio.

Installing TextChart Studio

i2 TextChart Studio is a Java application that you can install on Microsoft Windows, Apple macOS, and Linux operating systems. The procedure in this topic describes the installation process.

1. Double-click the `Studio-Installer.jar` file to begin, and then click **Next**.

   Note: You can also type `java -jar Studio-Installer.jar` at a command prompt to start the installer.
2. Read the license agreement, select **I accept the terms of this license agreement**, and click **Next** to continue.
3. Provide the installer with an installation path, and click **Next** to continue.

   By default, TextChart Studio is installed in a directory under `C:\Program Files` on Windows, under `/Applications` on macOS, and under `/opt` on Linux.

   Important: Always install TextChart Studio into a new directory. Do not install the software over previous installations.
4. On the **Select Path** page, choose a directory in which to store user data, and click **Next** to continue.
5. On the **User Data** page, select the amount of memory to devote to the Java virtual machine running the TextChart Studio process and, if desired, modify the network port on which to connect. Click **Next** to continue.

   Note: To change these settings after installation, you can configure the JVM memory in the `runtime.properties` file, and the HTTP port in the `studio.properties` file.
6. The installer displays a summary of your chosen settings. Click **Next** to continue.
7. The installation process begins. As it proceeds through each phase, click **Next** to continue, until you see the **Installation Finished** page.
8. Click **Done** to finish and exit the installer.

© N. Harris Computer Corporation

User data

TextChart Studio stores user data (including configuration settings, the LxBase, and regression points) in an operating-system-specific user area. Uninstalling TextChart Studio doesn't delete this data, and reinstalling TextChart Studio doesn't overwrite it.

To examine or delete TextChart Studio user data, see the following locations:

- Windows: The `%APPDATA%/TextChartStudio` directory
- Linux: The `$HOME/.textchartstudio` directory
- macOS: The `$HOME/Library/Application Support/TextChart Studio` directory

Starting and stopping TextChart Studio

i2 TextChart Studio is a web application that runs on the (included) Jetty web server, on the user's workstation. The application must be running before the user can access it through their web browser.

The commands for starting and stopping the TextChart Studio application are slightly different, depending on whether the user has a Windows, Linux, or macOS workstation.

Note: On all platforms, to restart the TextChart Studio application, use the "stop" and "start" commands in sequence.

Windows

To start TextChart Studio on Microsoft Windows:

- From the Start menu, select **TextChart Studio Start**. Alternatively, open a command window, navigate to the installation directory, and run the following command:

  ```
  rosokastudio.bat start
  ```

To stop TextChart Studio:

- From the Start menu, select **TextChart Studio Stop**. Alternatively, run the following command:

  ```
  rosokastudio.bat stop
  ```

Linux

To start TextChart Studio on Linux:

- From the installation directory, run the following command:

  ```
  ./rosokastudio.sh start
  ```

To stop TextChart Studio:

- From the installation directory, run the following command:

  ```
  ./rosokastudio.sh stop
  ```

macOS

On Apple macOS, you can use the same commands for starting and stopping TextChart Studio as on Linux. i2 also provides commands that you can run from the Finder.

To start TextChart Studio on macOS:

- Double-click the `StartStudio.command` file in the installation directory.

To stop TextChart Studio:

© N. Harris Computer Corporation

- Double-click the `StopStudio.command` file in the installation directory.

Managing TextChart Studio users

A fresh installation of i2 TextChart Studio comes with a pre-populated user account named **rosoka** that has administrator privileges. The initial password is the same as the username.

To add users to the system, or to change the password for a user, you use a command-line utility from the TextChart Studio installation directory. On Microsoft Windows, the utility is `rosokastudio.bat`; on Apple macOS and Linux, the utility is `rosokastudio.sh`.

To manage TextChart users, navigate to the TextChart Studio installation directory and run the command for your operating system, providing three parameters:

- The command name (**passwd**)
- The name of the user to add or modify
- The new password for the named user

For example, to create a user named **joe** and set their password to `xx7799`, run this command on Windows:

```
rosokastudio.bat passwd joe xx7799
```

Or this command on macOS and Linux:

```
./rosokastudio.sh passwd joe xx7799
```

On completion, the command provides the following output:

```
The new encrypted password for the user 'joe' is

   CRYPT:joKVahDE8hMXY

Copy this into the file 'conf/auth.properties' on the appropriate
line for the user 'joe'. Here is the full line for a user:

joe: CRYPT:joKVahDE8hMXY,rosokauser
```

Follow the instructions in the command output:

1. Using a text editor, open the `auth.properties` file, which is located in the `conf` directory under the TextChart Studio installation directory.
2. Paste the text from the output into the file, either replacing a line (to change the password for an existing user) or adding a line to the end (to add a new user).

   For example, after adding **joe** to the default user provided with TextChart Studio, the edited file would look like this:

   ```
   rosoka: CRYPT:roVsWZ.bWXgsk,rosoka
   joe: CRYPT:joKVahDE8hMXY,rosoka
   ```
3. Save the file, and then start or restart the TextChart Studio server.

On completion of the above steps, the new or modified user becomes available for logging in to i2 TextChart Studio.

Configuring user access through LDAP

If your organization uses LDAP (Lightweight Directory Access Protocol) as a site-wide authentication and authorization system, then you can configure TextChart Studio to use LDAP in place of its own user management system.

© N. Harris Computer Corporation

1. To configure TextChart Studio to access the LDAP server, you need to edit the supplied `conf/ldap.conf` file.

   Because interfacing with LDAP can be complex, i2 recommends contacting your IT department for assistance in setting the values in the LDAP configuration file.

   Note: To use TextChart Studio with an *LDAPS* (LDAP with SSL) server, see Configuring SSL for Studio for information about creating and managing SSL certificates and key stores.
2. To set up LDAP authentication, open the `conf/studio.properties` file in a text editor, and set the `authType` property to `ldap`:

   ```
   authType=ldap
   ```
3. To restrict access to the TextChart Studio user interface to members of a particular LDAP group, also set the `authGroup` property to the name of that group. For example:

   ```
   authGroup=TextChartUsers
   ```

   Alternatively, to allow all LDAP users to access TextChart Studio, either don't set the `authGroup` property or set it to the default value, `**`.
4. When your changes to `conf/ldap.conf` and `conf/studio.properties` are complete, restart TextChart Studio to make them take effect.

Changing the default HTTP port

By default, users connect to i2 TextChart Studio through HTTP on port 8080. If this port is already in use in your organization - or for any other reason - you can configure TextChart Studio to use a different port.

1. Open the `conf/studio.properties` file in a text editor.
2. Find the line that contains `httpPort=8080`, and change the port number to the new value. For example, to change the port to 8081, the line would look like this:

   ```
   httpPort=8081
   ```
3. Save your changes and close the file.

To verify your changes, restart TextChart Studio, add the new port number to the URL in your browser, and reopen the application.

Configuring SSL access to TextChart Studio

If you will access TextChart Studio only from the the workstation it's installed on, then setting up access through SSL is probably unnecessary. However, if others will use TextChart Studio from other workstations, i2 recommends configuring SSL connections.

Note: These instructions assume that you are familiar with the general use of SSL certificates and key stores. Following the instructions sets up the Jetty server that runs TextChart Studio to use SSL with a self-signed certificate. For more advanced use, see Managing SSL keys and certificates.

Enabling SSL for TextChart Studio involves creating a certificate *keystore* file, setting passwords for the keystore and the keystore manager, and enabling SSL functionality in the `conf/studio.properties` file:

1. Use the `keytool` program from the Java JDK to create a new keystore. Navigate to the `etc` directory and run the following command:

   ```
   keytool -keystore keystore -alias studio -genkey -keyalg RSA
   ```

© N. Harris Computer Corporation

Answer the questions appropriately for your site, and make a note of the password that you provide for the certificate generation process. When the command finishes, it generates a file named `keystore` file in the same directory.

Note: On Microsoft Windows, you might need to add the JDK binary directory to your path in order to run the `keytool` command. You can do so with a command like this:

```
set PATH=%PATH%;"c:\Program Files\Java\jdk1.8.0_65\bin"
```

Adjust the path and the version number to match your JDK installation.

2. Next, you must add the password that you entered during the certificate creation process to the TextChart Studio properties file. So that the password is not visible in plain text, you obfuscate it.

   Run the following command at the command line, substituting the password you entered during the certificate creation process in place of `<password>`:

   ```
   java -cp RosokaStudio.jar org.eclipse.jetty.util.security.Password
    <password>
   ```

   Note: Run this command from the TextChart Studio installation directory.

   The command outputs the obfuscated password to the console. Copy it (including the initial `OBF:`) for use in the next step.

3. Open the `conf/studio.properties` file in a text editor, remove the # signs from the lines below, and enter the port you want to use, the path to the keystore, and the obfuscated password as values for the following settings:

   ```
   httpsPort=8443
   keyStorePath=
   keyStorePassword=
   keyManagerPassword=
   ```

   To prevent unsecured HTTP access to TextChart Studio, comment out or remove the `httpPort` setting, and then save the modified file.

4. Restart TextChart Studio.

After making these changes, you can access TextChart Studio on the local workstation at `https://localhost:8443/RosokaStudio`. (If you changed the port number, you'll need to make the same change to the URL.)

Using a self-signed certificate like this generates a warning when you access TextChart Studio through a web browser. You can instruct your browser to ignore the error and proceed to the page, but to remove it completely you must get your certificate signed by a valid certification authority. See Managing SSL keys and certificates for more information.

Managing SSL keys and certificates

This topic contains additional information about generating and managing SSL keys and certificates for use with deployments of i2 TextChart Studio.

Generating keys and certificates with keytool

The simplest way to generate keys and certificates is to use the `keytool` application that comes with the JDK, as it generates keys and certificates directly into the keystore.

If you already have keys and certificates, jump to Loading keys and certificates to load them into a Java Secure Socket Extension (JSSE) keystore. That section also applies if you have a renewal certificate to replace one that is expiring. The examples below generate only basic keys and certificates.

© N. Harris Computer Corporation

The following command generates a key pair and a certificate directly into a file named `keystore`:

```
keytool -keystore keystore -alias jetty -genkey -keyalg RSA
```

The command prompts for information about the certificate, and for passwords to protect both the keystore and the keys within it. The only mandatory response is to provide the fully qualified host name of the server at the "first and last name" prompt.

You now have the minimum requirements for creating an SSL connection, and you could proceed directly to configuring one. However, the browser will not trust the certificate you generated, and will prompt the user to this effect. What you have at this point is sufficient for testing, but not for production.

If you want to use only a self-signed certificate, you can add `-validity <days>` to the `keytool` call above to extend the validity of the certificate beyond a month, which is the default period.

You can also use the SAN extension to set one or more names that the certificate applies to:

```
keytool -keystore keystore -alias jetty -genkey -keyalg RSA -sigalg
 SHA256withRSA --ext 'SAN=dns:jetty.eclipse.org,dns:*.jetty.org'
```

Requesting a trusted certificate

To obtain a certificate that most common browsers will trust, you need to ask a well-known certificate authority (CA) to sign it. Trusted CAs include AddTrust, Entrust, GeoTrust, RSA Data Security, Thawte, VISA, ValiCert, Verisign, and beTRUSTed.

Each CA has its own instructions (look for information about "JSSE" or "OpenSSL"), but all involve a step that generates a certificate-signing request (CSR).

Generating a CSR with keytool

The following command uses `keytool` to generate the file `jetty.csr` for a key or certificate that's already in the keystore:

```
keytool -certreq -alias jetty -keystore keystore -file jetty.csr
```

Loading keys and certificates

When a CA has sent you a certificate, or if you generated your own certificate without `keytool`, you need to load it into a JSSE keystore.

Note: You need both the private key and the certificate in the JSSE keystore. You should load the certificate into the keystore that was used to generate the CSR with `keytool`. If your key pair is not already in a keystore (for example, because it was generated with OpenSSL), you need to use the PKCS12 format to load both the key and certificate.

Loading certificates with keytool

You can use `keytool` to load a certificate in PEM format directly into a keystore. The PEM format is a text encoding of certificates; it is produced by OpenSSL, and is returned by some CAs.

The following command loads a PEM-encoded certificate in the `jetty.crt` file into a JSSE keystore:

```
keytool -keystore keystore -import -alias jetty -file jetty.crt -
trustcacerts
```

If the certificate that you receive from the CA is not in a format that `keytool` understands, you can use the `openssl` command to convert formats:

```
openssl x509 -in jetty.der -inform DER -outform PEM -out jetty.crt
```

Loading keys and certificates via PKCS12

© N. Harris Computer Corporation

If you have a key and certificate in separate files, you need to combine them into a PKCS12 format file to load into a new keystore. The certificate can be one you generated yourself or one returned from a CA in response to your CSR.

The following OpenSSL command combines the keys in `jetty.key` and the certificate in the `jetty.crt` file into a file named `jetty.pkcs12`:

```
openssl pkcs12 -inkey jetty.key -in jetty.crt -export -out jetty.pkcs12
```

If you have a chain of certificates because your CA is an intermediary, build the PKCS12 file as follows:

```
cat example.crt intermediate.crt [intermediate2.crt] ... rootCA.crt > cert-
chain.txt
openssl pkcs12 -export -inkey example.key -in cert-chain.txt -out
 example.pkcs12
```

The order of certificates must be from server to rootCA, as described in RFC2246 section 7.4.2.

OpenSSL asks for an *export password*. A non-empty password is required to make the next step work. Then load the resulting PKCS12 file into a JSSE keystore with `keytool`:

```
keytool -importkeystore -srckeystore jetty.pkcs12 -srcstoretype PKCS12 -
destkeystore keystore
```

Renewing certificates

If you are updating your configuration to use a newer certificate because the old one is expiring, just load it as described in Loading keys and certificates.

If you originally imported the key and certificate using the PKCS12 method, use an `alias` of `1` rather than `jetty`, because that is the alias the PKCS12 process enters into the keystore.

Tesseract OCR parser

Tesseract Optical Character Recognition (OCR) parsing may be installed to provide seamless OCR support for many common graphical file formats, including JPEG, TIFF and GIF. Once Tesseract OCR is installed and running, TextChart will seamlessly process supported file formats in the ingestion process via the Tika interface. No further configuration is required.

Installing Tesseract on RHEL

1. Install the EPEL repository if it isn't already available:

   ```
   sudo dnf install epel-release
   ```
2. Install Tesseract:

   ```
   sudo dnf install tesseract
   ```
3. To add language packs, see what's available:

   ```
   dnf search tesseract
   ```

   And then install, for example:

   ```
   sudo dnf install tesseract-langpack-ara
   ```

Installing Tesseract on Ubuntu

1. Download package information from all configured sources:

   ```
   sudo apt update
   ```
2. Install Tesseract:

© N. Harris Computer Corporation

```
sudo apt install tesseract-ocr
```

3. To add language packs, see what's available:

   ```
   apt search tesseract-ocr
   ```

   And then install, for example:

   ```
   sudo apt install tesseract-ocr-fra
   ```

Installing Tesseract on Windows

1. Download the latest Tesseract installer from the UB-Mannheim Tesseract repository. Older versions are available at `https://digi.bib.uni-mannheim.de/tesseract/`.
2. Run the installer and follow the prompts. Install Tesseract in the directory suggested by the installer or in a new directory.

   Important: Do not install into an existing directory that contains other files. The uninstaller removes the entire installation directory.
3. During installation, select any additional language packs that you require.
4. Add the Tesseract installation directory to the system `PATH` environment variable so that it is available from the command line.

Installing Tesseract on other platforms

For further details on Tesseract, including installation on Macintosh, see the information at https://cwiki.apache.org/confluence/display/tika/TikaOCR.

Using TextChart Studio

This part of the documentation describes what i2 TextChart Studio does, how it works, and how to use it.

Getting started

When you start to use i2 TextChart Studio for the first time, you need to log in, provide your license information, and perform some initial customization.

Logging in

To log in to TextChart Studio, enter your username and password on the front page. A development system includes a default account with the following credentials:

- **Username:** rosoka
- **Password:** rosoka



© N. Harris Computer Corporation

Accessing settings

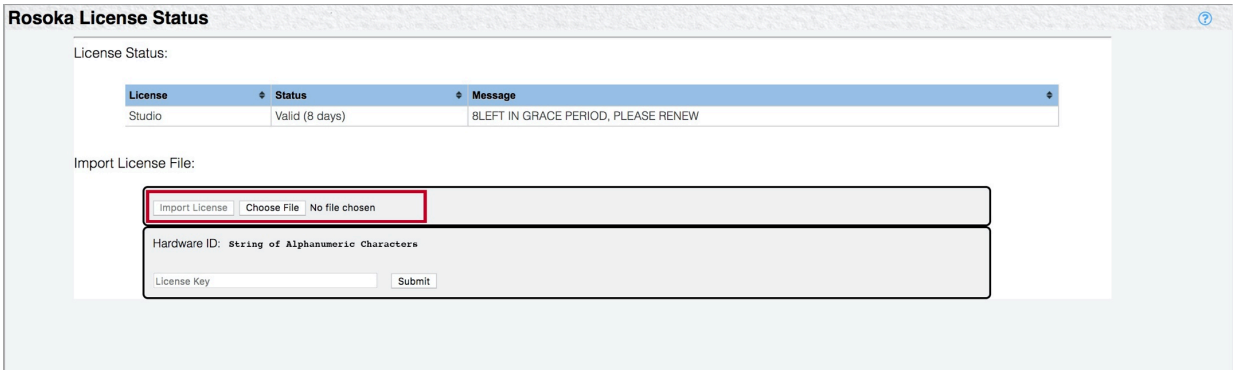
To access TextChart Studio settings (including license provision), use the toolbar on the left of the application window. The button at the top of the column opens the **Licenses & LxBase Settings** section of the toolbar:



The first button in the open section allows you to **Manage Licenses**; the second lets you **Edit Property Settings**.

Providing a license key

To provide TextChart Studio with a license key, click **Manage Licenses** to open the **Rosoka License Status** page.

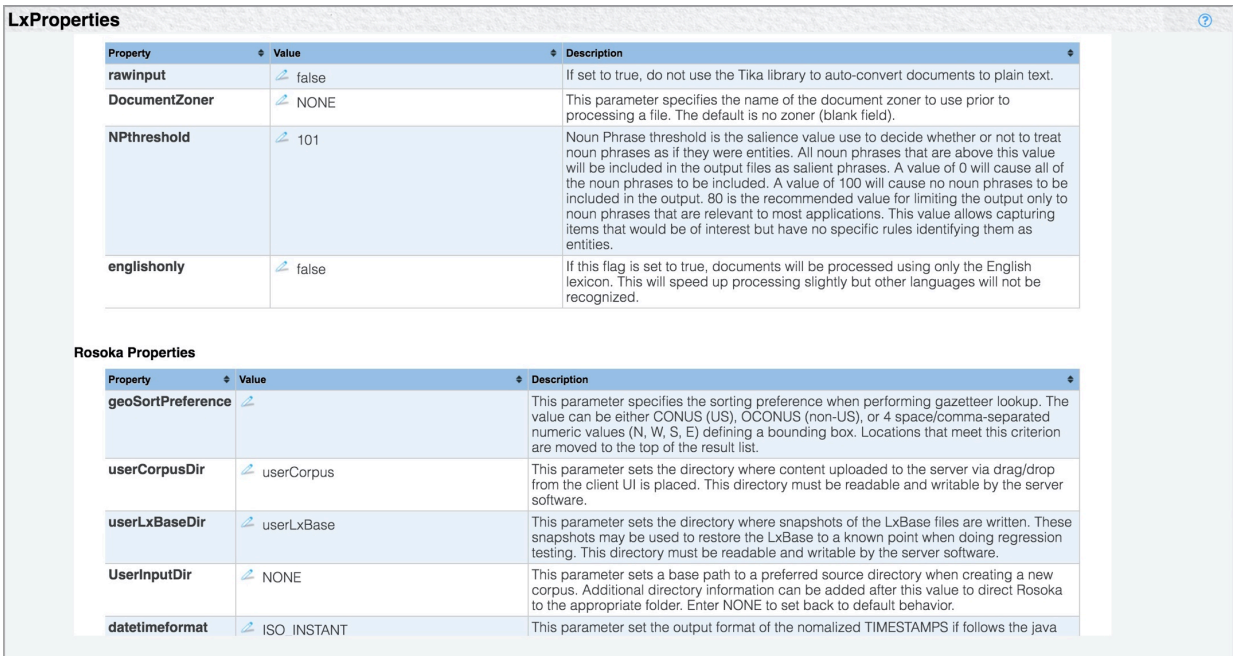


The top part of the page indicates how long a particular license is valid for. The bottom part allows you to provide a new license key. Click **Choose File** to locate your `licenseKeys.xml` file, and then click **Import License**.

Changing settings

To change TextChart Studio settings, click **Edit Property Settings** to open the **LxProperties** page.

© N. Harris Computer Corporation



Some of the most frequently modified properties are the following:

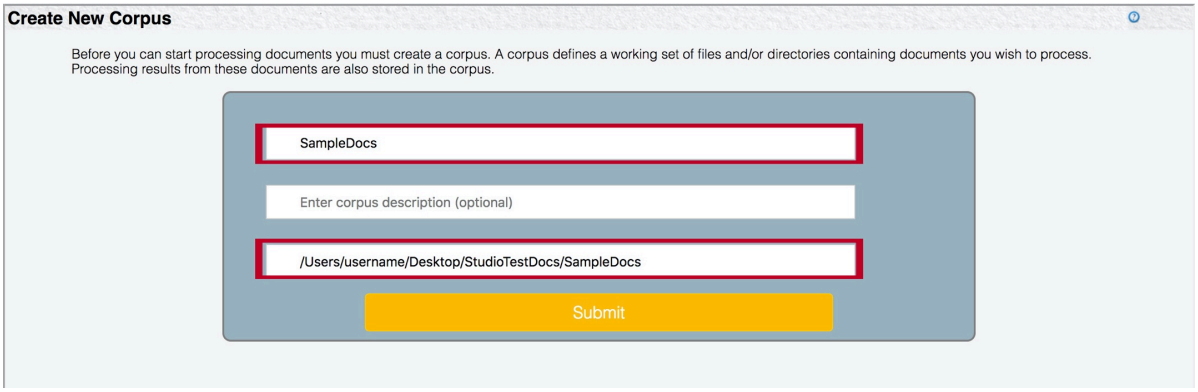
- **proximityrelationship**

  When `proximityrelationship` is `true`, relationship extraction includes relationships based on the entities' proximity to one another.

- **UserInputDir**

  The `UserInputDir` setting specifies the location of the directory where TextChart Studio looks for sets of documents (*corpora*) by default.

  The directory that you provide here is used to pre-populate the path on the **Create New Corpus** page.

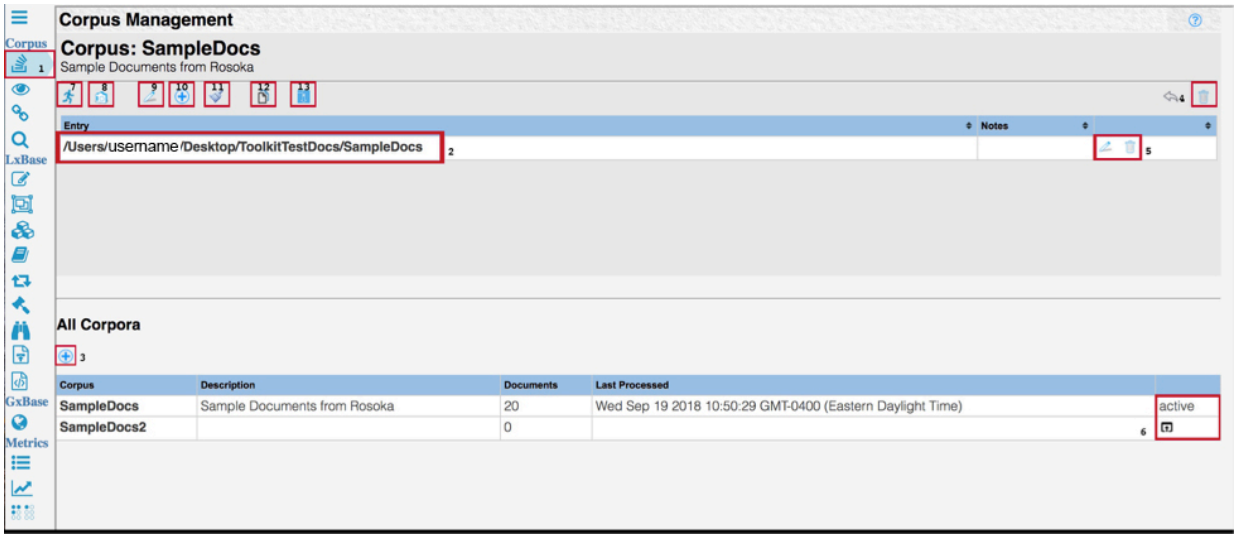


© N. Harris Computer Corporation

Corpus management

When you first log in to i2 TextChart Studio, you're prompted to create an initial corpus of documents that you'll use to evaluate the changes you make to the LxBase. optionally, you can create multiple, named corpora with different sets of documents.

The **Corpus Management** page in TextChart Studio allows you to add, remove, and process corpora; to toggle between different saved corpora; and to perform a variety of other document management functions.



1. To open the the **Corpus Management** page, click the **View Corpus List** icon located towards the top of toolbar.
2. The list at the top of the page indicates the *active* corpus. Corpus management commands act only on the active corpus.
3. To return to the **Create New Corpus** page, click the **Add a New Corpus** button above the **All Corpora** list at the bottom of the page.
4. To remove the active corpus from the **Corpus Management** page, click **Delete Corpus**.
5. To edit the path of the active corpus, use the **Edit Entry** and **Delete Entry** buttons.
6. To make a different corpus active, click the **Open Corpus** button on the right of the **All Corpora** list.
7. To make TextChart Studio process all the documents in the active corpus, click **Process all Documents** in the horizontal toolbar.
8. To take a snapshot of the LxBase in its current form for later comparisons, click **Create New Regression Point**.

   You can use TextChart Studio's regression testing feature to measure the impact that your changes to the LxBase have on the output. For more information, see Regression Testing.
9. To change the name or the description of the active corpus, click **Edit Corpus Name/Description**.
10. To add more documents to the active corpus, click **Add More Files**.
11. If you modify any dictionaries or linguistic rules after you process the active corpus, you must clear the previous results before you reprocess the corpus. To do so, click **Clear Processing Results**.

© N. Harris Computer Corporation

12. TextChart Studio automatically removes duplicate documents upon processing. Click **Show Duplicate Documents** to see a list of these duplicates. Raw documents are documents that are the same before Tika processing; content duplicates are documents that are the same after Tika processing.
13. Sometimes, TextChart finds words that it does not understand. Typically, these words belong to subject-specific vocabularies, or they pertain to proprietary industry information, or they are non-English words for which TextChart has no translation.

    To download a list of these words, potentially to add them to a custom LxBase, click **Download Unknown Words**

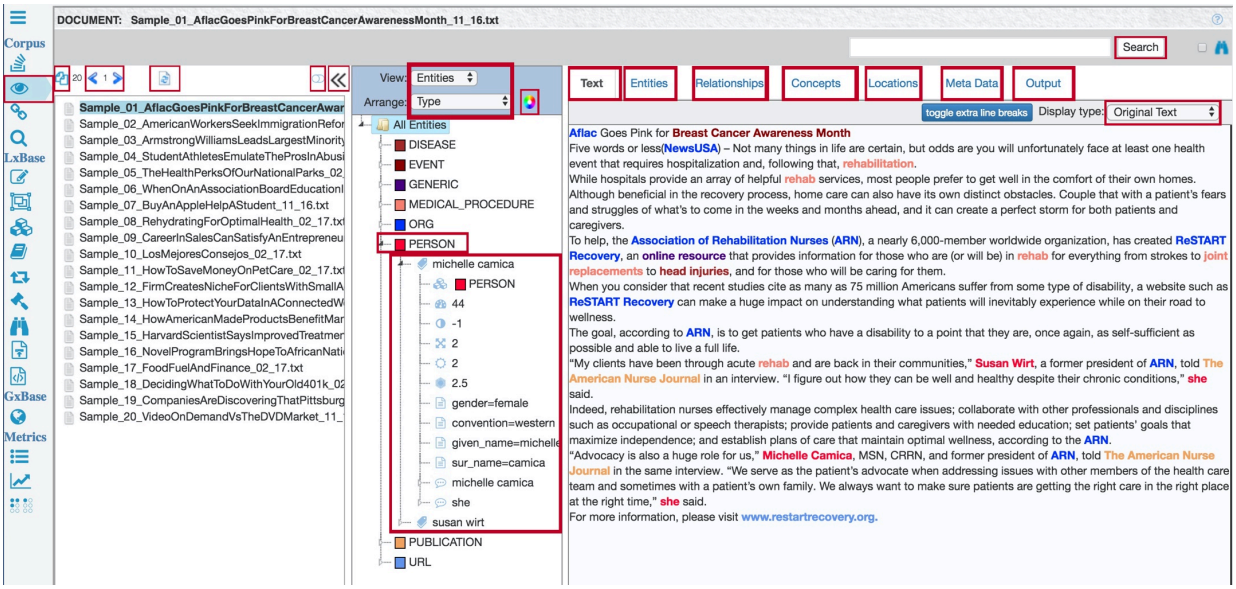
    

    above the entry information line.

    Select which language to download an unknown word list for, and then click **Submit** to compile the list, which can take several minutes. When you see a notification, click the icon to download the list as a ZIP file.

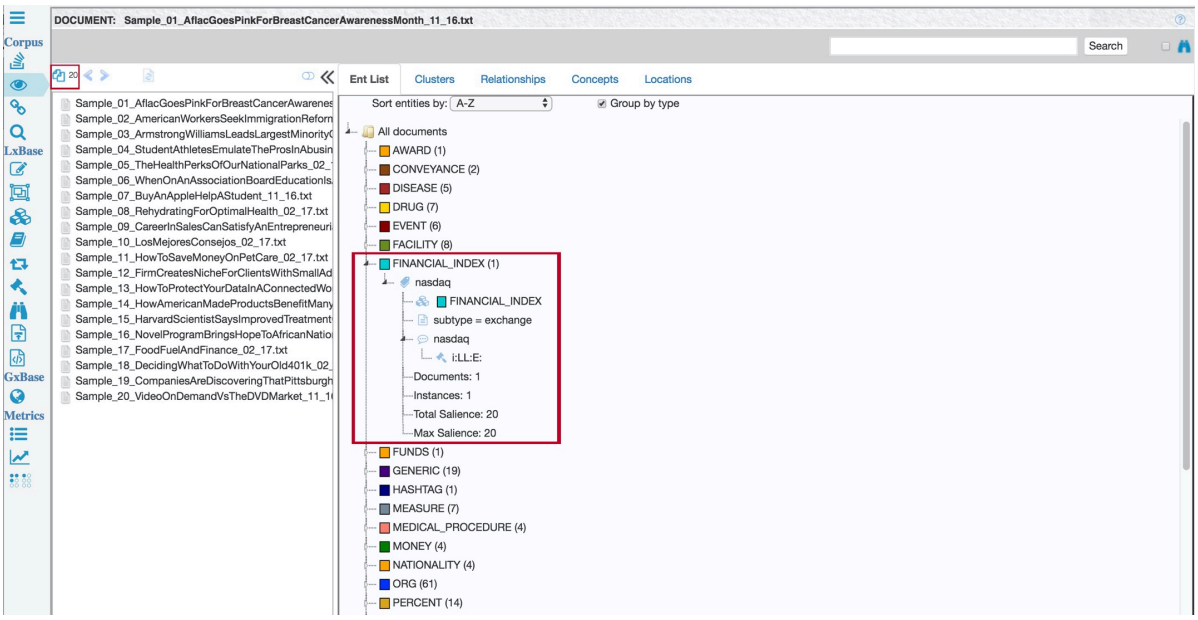
Viewing corpus results

By default, the **Active Corpus** page contains information about a single document in that corpus. Here, you can view a single document with hit highlights, select certain entities to view, explore the lexical items, and view rule traces.



1. To open the **Active Corpus** page, click the **View Active Corpus** icon near the top of the vertical toolbar.
2. To see a summary of the extraction results for the whole corpus, click **Aggregate View**.

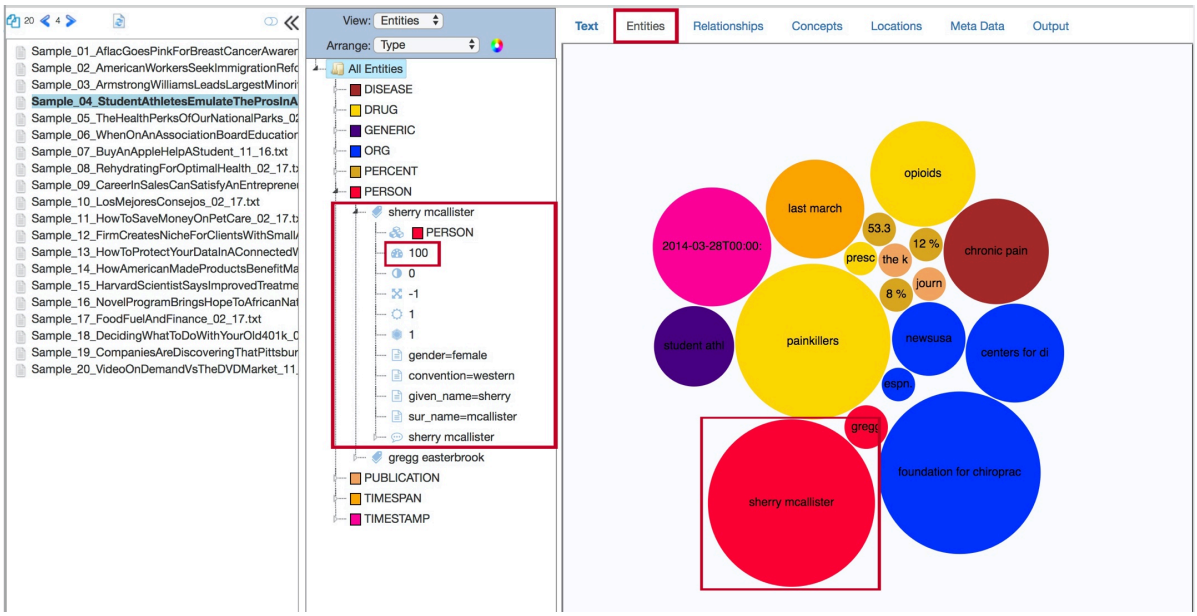
© N. Harris Computer Corporation



Each entity type that TextChart Studio found in the corpus is listed, along with its total number of extraction instances. You can open entries in the list to see individual extraction results and their corresponding metadata.

3. To view the results for each document in the corpus, click the **Next Document** and **Previous Document** buttons. TextChart Studio displays the text for the document on the right of the page, with highlights corresponding to the extracted entities.
4. When you make changes to the LxBase, and you want to see the effect of those changes on a single document without reprocessing the whole corpus, click **Reprocess the Current Document**.
5. Instead of navigating through the documents in the corpus, click **Switch to Search Mode** to view the results for a specific document by name.
6. When you select a single document in the corpus, the tree view in the center of the **Active Corpus** page displays the extraction results for that document. Use the **View** and **Arrange** lists to change how the results are represented.
7. On the right of the **Active Corpus** page, the **Text** tab contains the text of the selected document with hit highlights corresponding to the extracted entities.
8. The **Entities** tab contains a visualization of extracted entities and their overall importance within the selected document.

© N. Harris Computer Corporation



The larger the circle that represents the entity, the more important (or *salient*) that entity is to the document. Salience is scored from 0 to 100, with 100 representing the most entity.

The **Entities** tab is also available in the aggregate view, where its contents represent the salience of extracted entities to to the corpus as a whole.

9. The **Relationships** tab contains a visualization of extracted entities and their relationships to each other in the selected document.

© N. Harris Computer Corporation

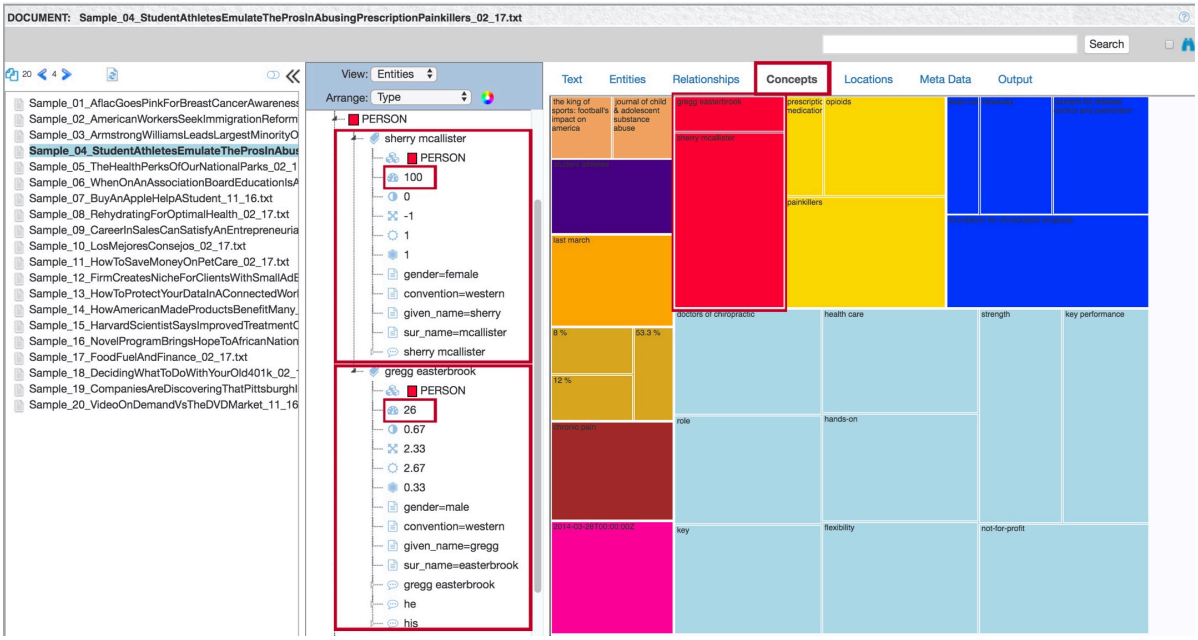


You can interact with the contents of the **Relationships** tab to gain information about the subjects, objects, and predicates that prompted each extracted relationship.

The tab is also available in the aggregate view, where its contents represent relationships between entities across the whole corpus.

10. The **Concepts** tab contains a visualization of extracted entities, as well as other words or phrases that TextChart has identified as salient by their overall importance to the selected document.

© N. Harris Computer Corporation



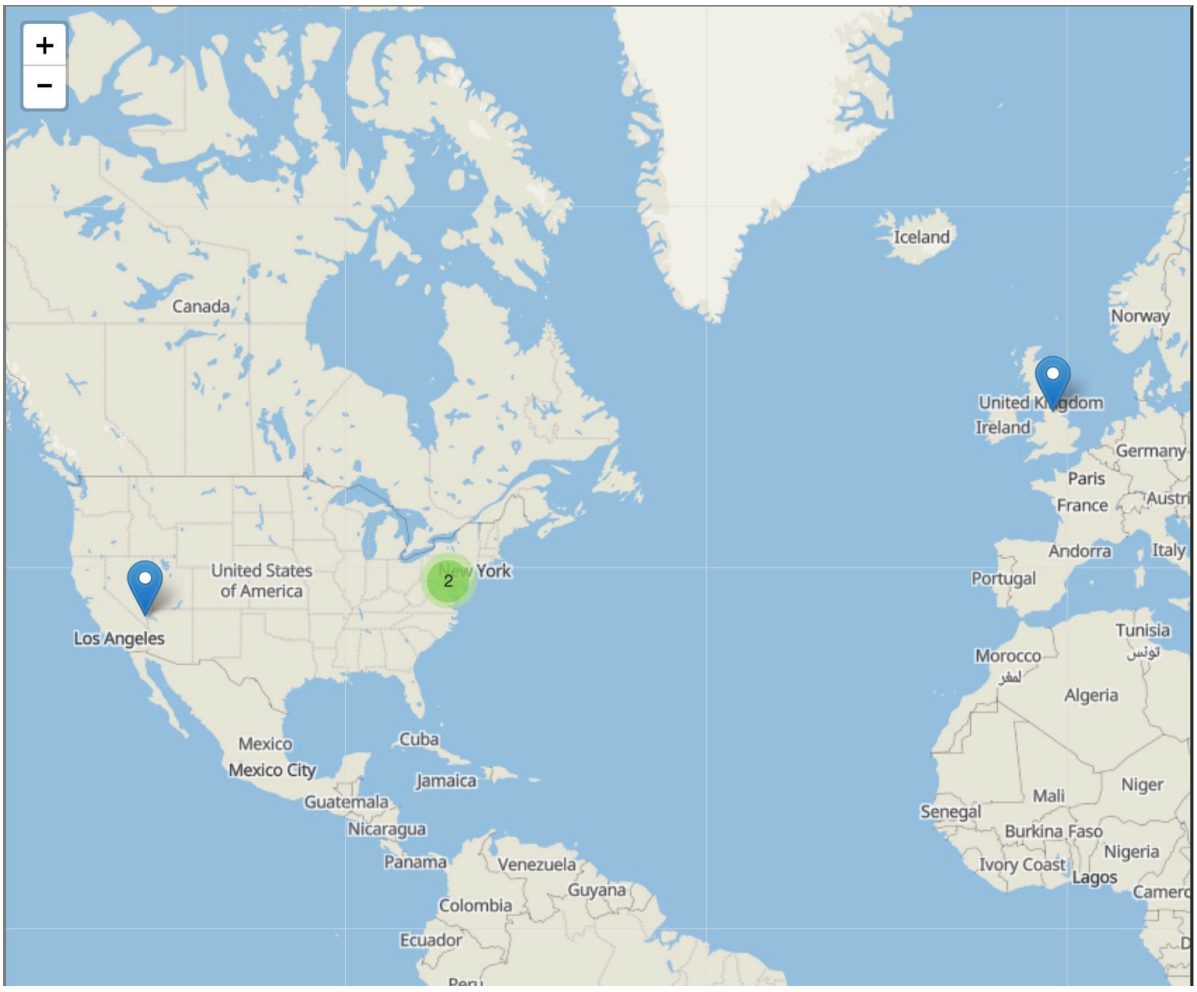
The larger the box that represents the entity, the more important (or *salient*) that entity is to the document.

The brightly colored boxes represent individually extracted entities, while the light blue boxes represent salient words or phrases. You can use this information to see if there are any additional items that need to be modified to reach your extraction goals.

The **Concepts** tab is also available in the aggregate view, where its contents represent the salience of words and phrases to the corpus as a whole.

11. The **Locations** tab contains a visualization of extracted entities and their geographic locations from the selected document.

© N. Harris Computer Corporation



This tab is also available in the aggregate view, where its contents represent the locations of entities from across the corpus as a whole.

12. The **Metadata** tab presents the user with document-level metadata, including overall sentiment scores and the languages identified in the selected document.

© N. Harris Computer Corporation

Document	
Anchor Date	2018-03-28T16:38:05Z
Anchor Method	OSFILEDATE
Polarity	-0.78
Mood	1.43
Intensity	1.47
Aspect	1.79

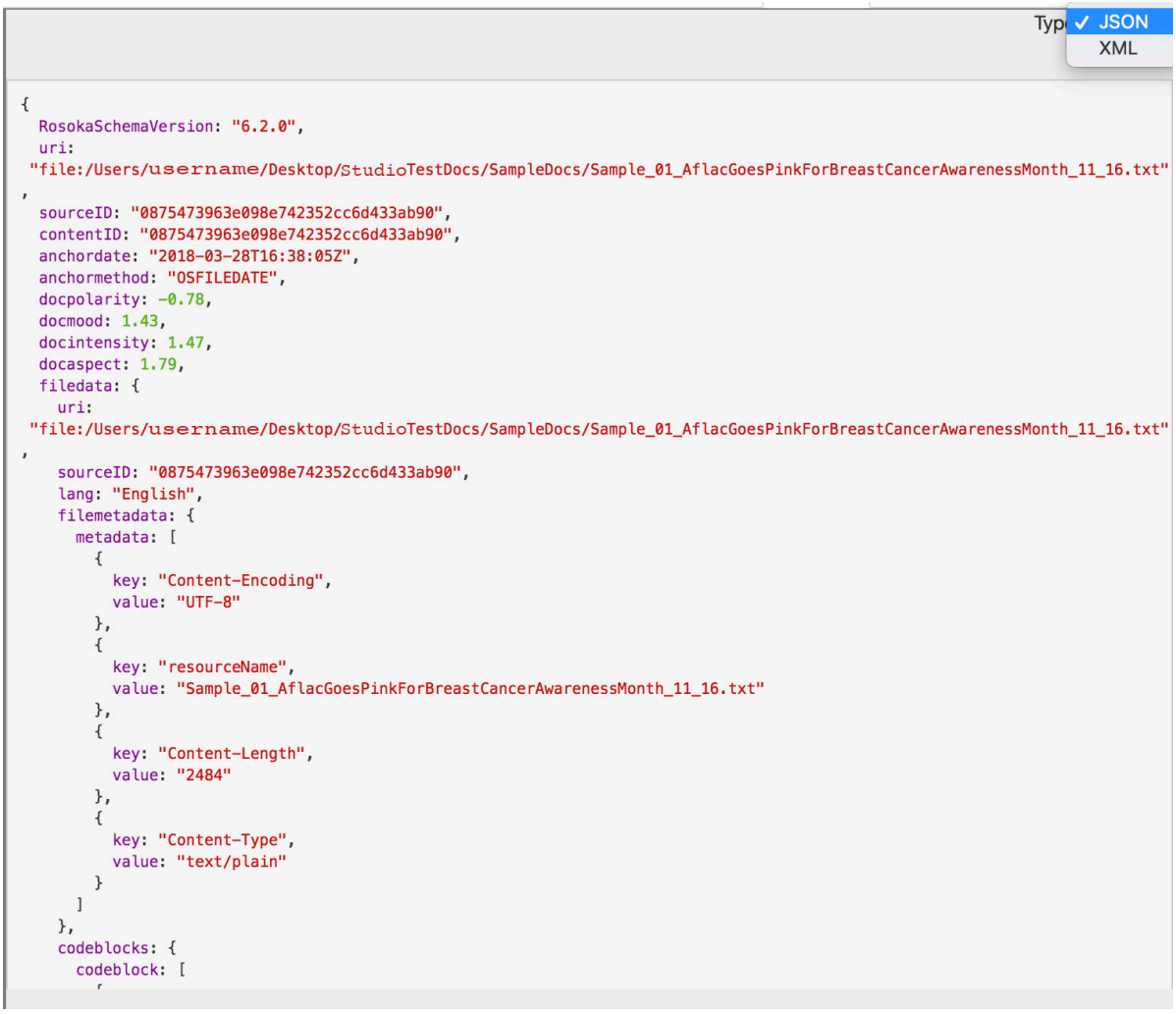
Languages	
English	100
French	7

Codeblock	**Count**
Basic_Latin	2469
General_Punctuation	14

Field	**Value**
Content-Encoding	UTF-8
resourceName	Sample_01_AflacGoesPinkForBreastCancerAwarenessMonth_11_16.txt
Content-Length	2484
Content-Type	text/plain

13. The **Output** tab contains the raw output generated from the extraction process for the selected document, in JSON or XML format.

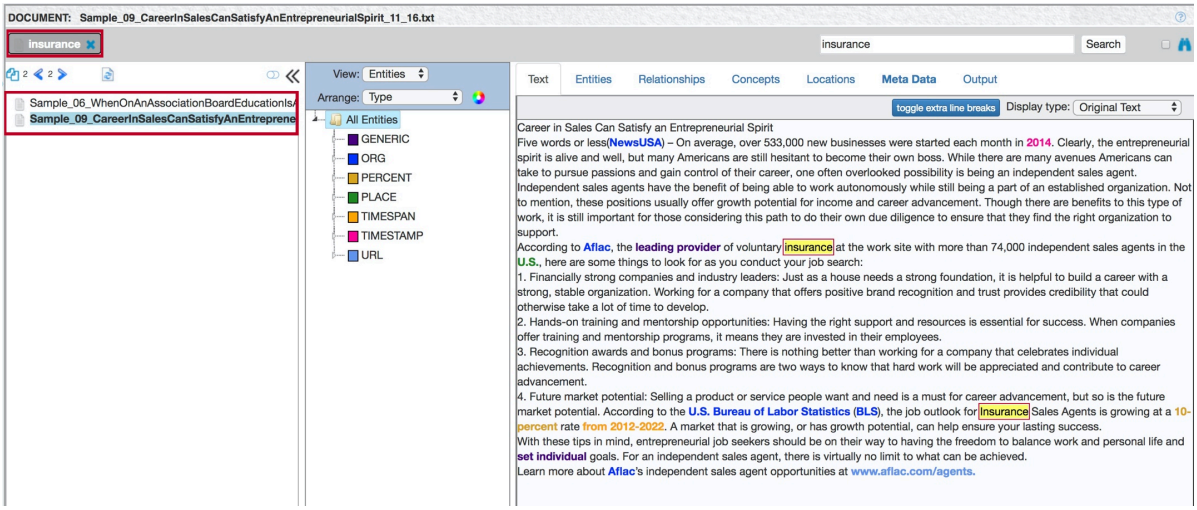
© N. Harris Computer Corporation



This information is what would be sent to the database if the current LxBase had processed the document in a production environment. You can view the output in its source language or with an English gloss, and get a breakdown of the individual tokens, entities, sentiment scores, and relationship extraction results.

14. Use the **Search** function to find all the documents in the corpus that contain a particular word or phrase, effectively filtering the document list.

© N. Harris Computer Corporation



When you select a document from search results, the contents of the **Text** tab highlight not only extracted entities but also the terms that you searched for.

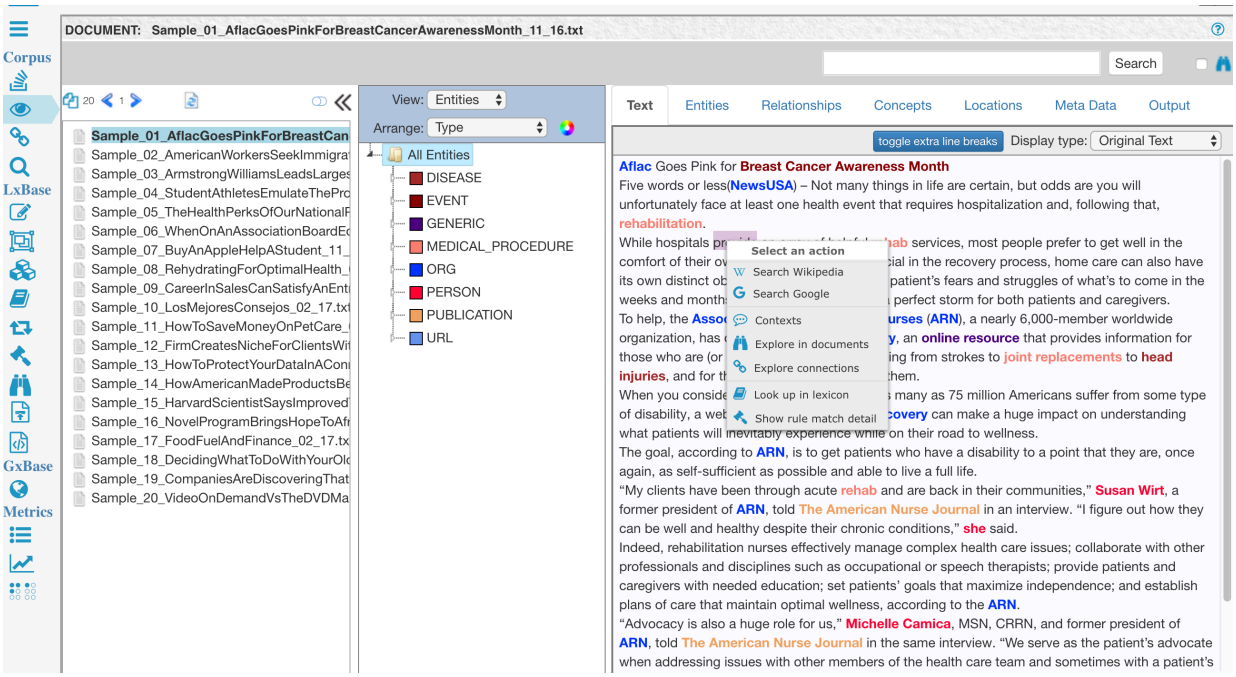
15. When you're viewing a single document, you can use the **Display type** options to view the original text in its source language, with an English gloss, or both.

    In the side-by-side view, hovering over an element on either side highlights the corresponding information in both views.

Exploring processed text

TextChart Studio's **Text** tab allows you to modify the current LxBase directly, by interacting with the highlighted extraction results. Right-click any result to see a pop-up menu with a list of the available options.

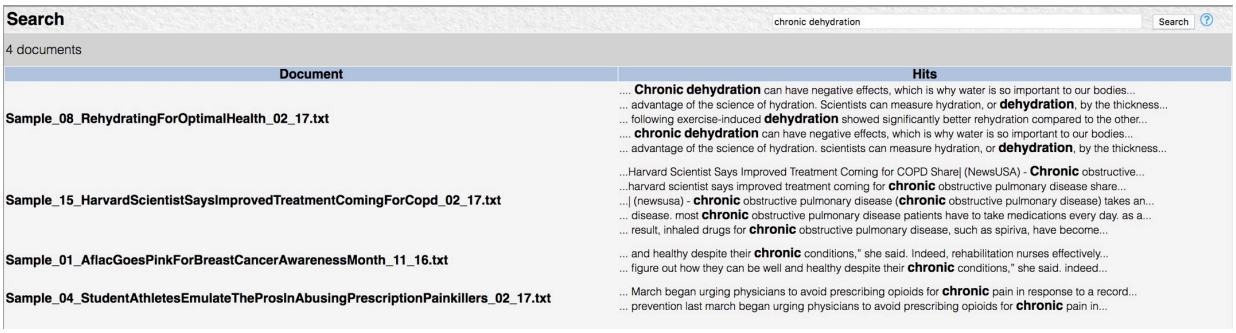
© N. Harris Computer Corporation



Among other things, the menu enables you to search for the result in both Wikipedia and Google. You can also open different views in TextChart Studio, as well as modify the associated semantic vectors.

Document search

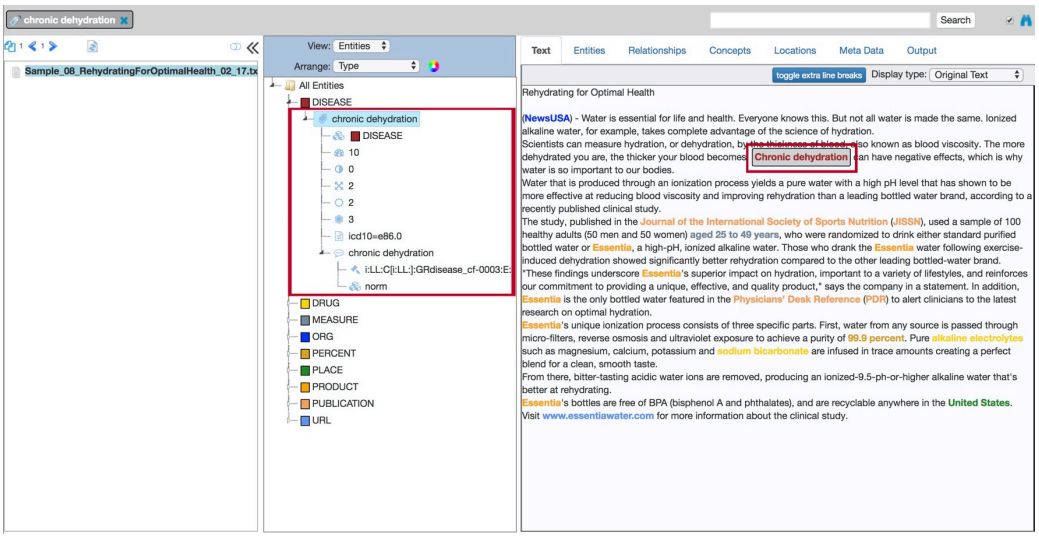
Select **Document search** from the pop-up menu to perform a fuzzy search for the extraction result across all documents in the corpus. The results are a list of the documents that matched the search, along with a brief sample of the context in which the result was found.



Explore in documents

Select **Explore in documents** from the pop-up menu to generate a list of all the documents that contain the extraction result. The contents of the **Text** tab change to display the first document in the list, with the first matching result selected.

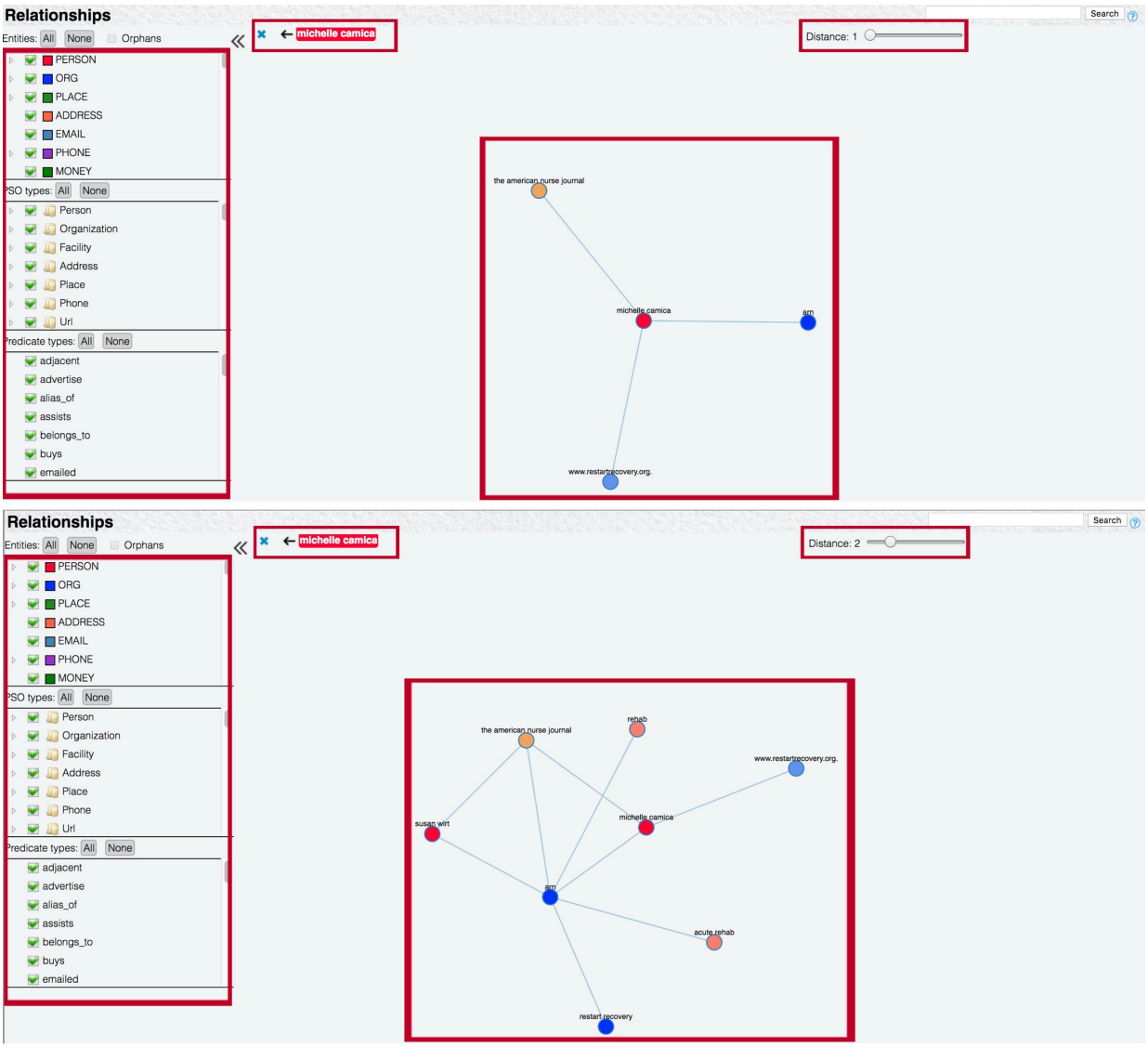
© N. Harris Computer Corporation



Explore connections

Select **Explore connections** from the pop-up menu to switch to the **Relationships** page, which contains information about the relationships in which the selected entity is involved.

© N. Harris Computer Corporation

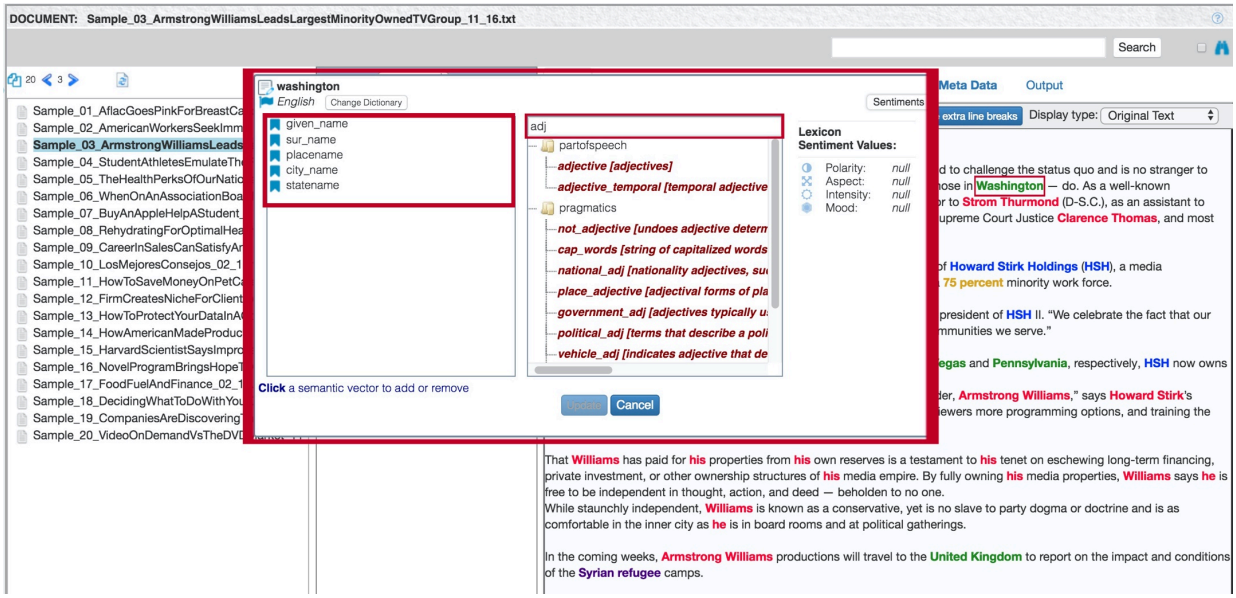


You can adjust the **Distance** setting to display relationships at up to six degrees of separation from the selected entity.

Look up in lexicon

Select **Look up in lexicon** from the pop-up menu to display a dialog where you can modify the highlighted term.

© N. Harris Computer Corporation



The dialog shows the term, the language of the dictionary that TextChart found it in, and a list of any semantic vectors (SVs) that are associated with it.

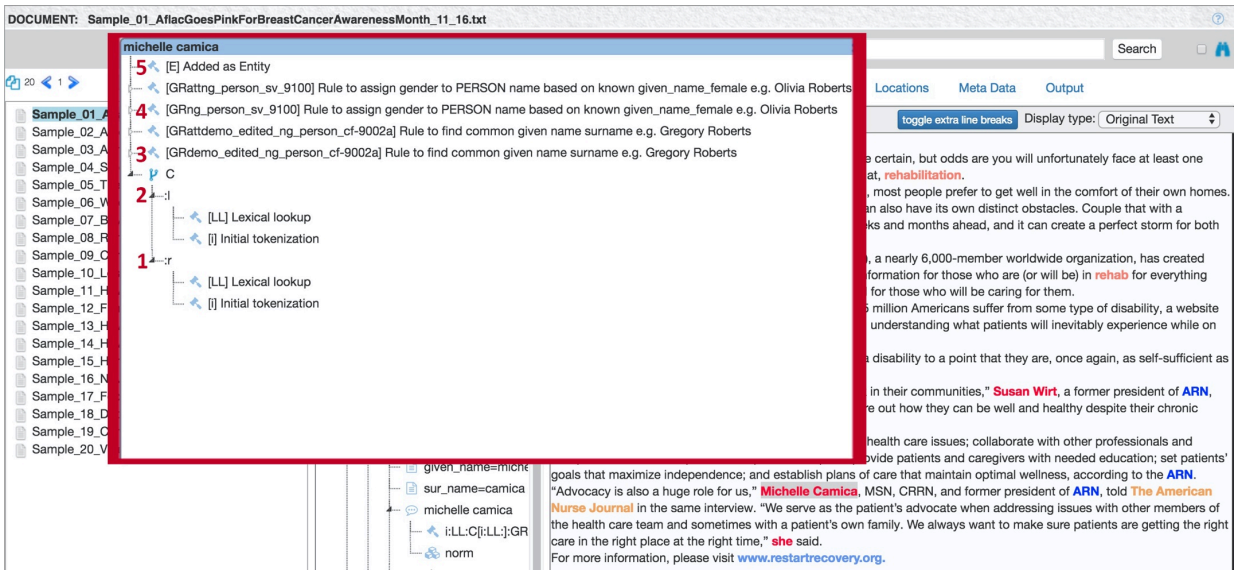
To add a new semantic vector to the term, click **Add SV** and then use the box on the right to find the semantic vector that you want to add. Alternatively, click an existing semantic vector to remove it from the term.

When your changes are complete, click **Update** to make the TextChart engine recognize them, and then reprocess either the individual document or the entire corpus.

Show rule match detail

Select **Show rule match detail** from the pop-up menu to display a dialog that contains the steps the TextChart engine took in order to extract (or not extract) a particular result. For example, to extract the PERSON entity "Michelle Camica", the engine took five steps:

© N. Harris Computer Corporation



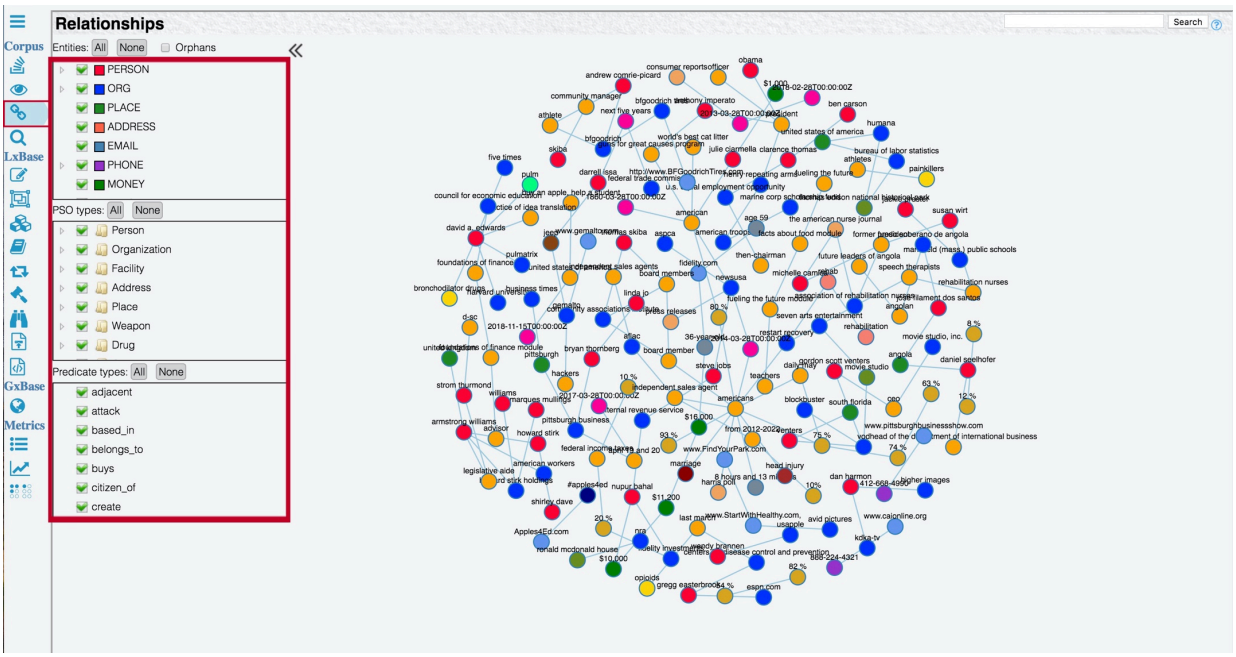
1. The TextChart engine tokenized the term "Michelle" and looked it up in the dictionaries.
2. The TextChart engine tokenized the term "Camica" and looked it up in the dictionaries.
3. The engine executed a linguistic rule whose description contains "...to find common given name and surname..."
4. The engine executed a second linguistic rule whose description contains "...to assign gender to PERSON name based on known given name female..."
5. The engine assigned the two tokens "Michelle" and "Camica" to one PERSON entity, based on the linguistic rules that it executed.

To modify a rule, you can double-click it in the list to open the rule editor.

Relationships

The **Relationships** page allows you to interact with relationships between the entities in your extraction results. To open the page, click the **View Relationships** button in the vertical toolbar.

© N. Harris Computer Corporation



Use the check boxes on the left of the page to select and deselect entity and PSO types in order to customize the visualization.

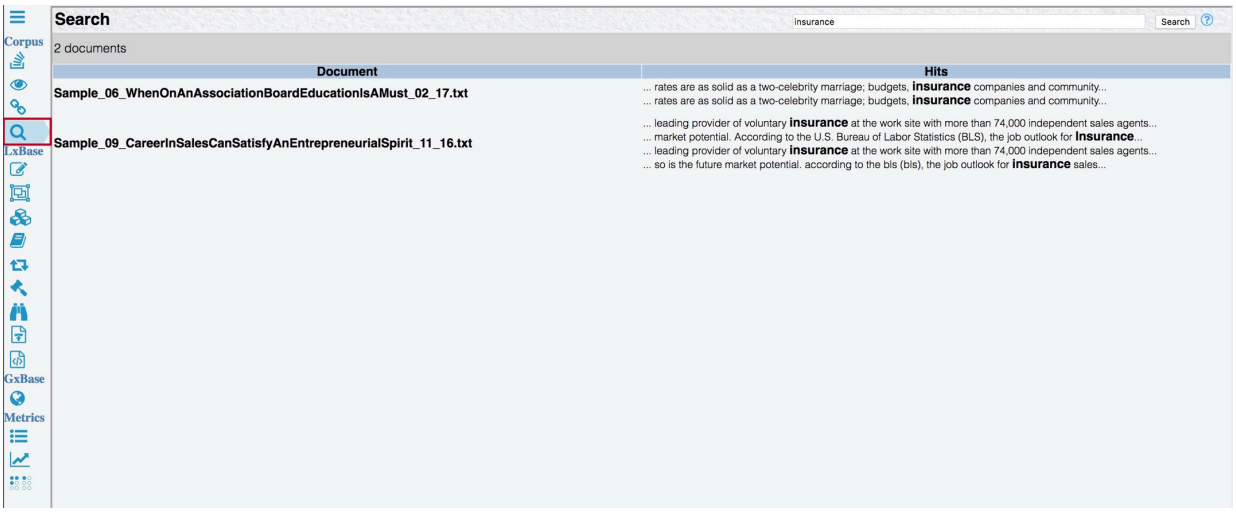
From a single entity in the visualization, you can step out from first-level connections to see a more complete picture by using the **Distance** setting.

You can also see connections in other views throughout TextChart Studio by right-clicking entities and selecting **Explore connections**.

Search page

You can use the **Search** page in TextChart Studio to perform a fuzzy search across all the documents in a corpus. To open it, click the **Search** button in the vertical toolbar.

© N. Harris Computer Corporation



This list of results is presented on the left, with a brief summary of the context surrounding the search result on the right. Click a document in the results to open it in the **Text** tab.

Token definition editor

TextChart Studio's **Token Definition Editor** page enables you to add or modify semantic vectors in the LxBase. To open the page, click the **Edit Token Definitions** button in the vertical toolbar.

The page presents the current set of semantic vectors in separate tabs according their type.

Semantic vector types

After processing with TextChart, the terms in a document are associated with one or more semantic vectors. An LxBase defines each semantic vector as one of the following: an *entity*, a *no-output entity*, a *part of speech*, *pragmatic*, or *relational*.

- **Entities**

  A term that's associated with an entity semantic vector such as PERSON, ORG, or PLACE is extracted as an entity and included in the output.

- **No-output entities**

  A term that's associated with a no-output entity semantic vector *is* extracted as an entity but *is not* included in the output.

  By changing a semantic vector from "entity" to "no-output entity", you can prevent entities of the affected type from appearing in the output.

- **Parts of speech**

  A term that's associated with a part-of-speech semantic vector has been identified as a part of speech, such as a verb or an adjective.

- **Pragmatic**

  A term that's associated with a pragmatic semantic vector has been identified as belonging to a pragmatic category such as "given name" or "possible identification number".

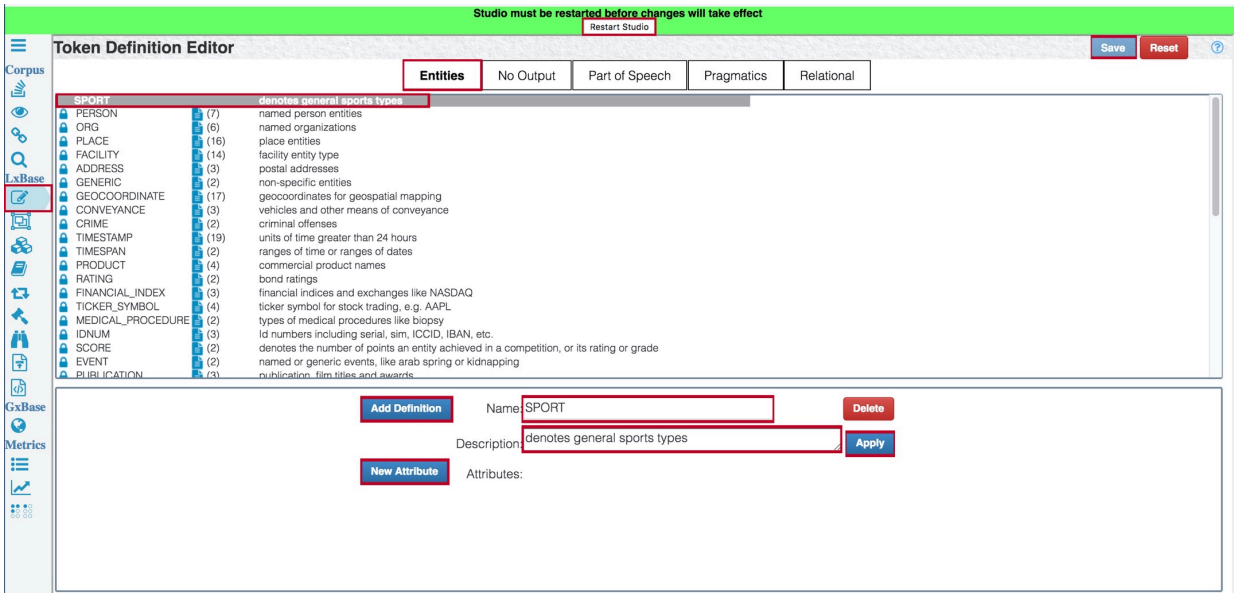
© N. Harris Computer Corporation

- **Relational**

  A term that's associated with a relational semantic vector such as "interviewed" or "identified as" is used as the predicate in a predicate-subject-object (PSO) relationship.

Editing entity types

To add to the default set of entity types in an LxBase, or to modify one of the existing entity types, you use the **Entities** tab in the token definition editor.



To add a new entity type, click **Add Definition**. To add an attribute to a new or existing entity type, click **New Attribute**.

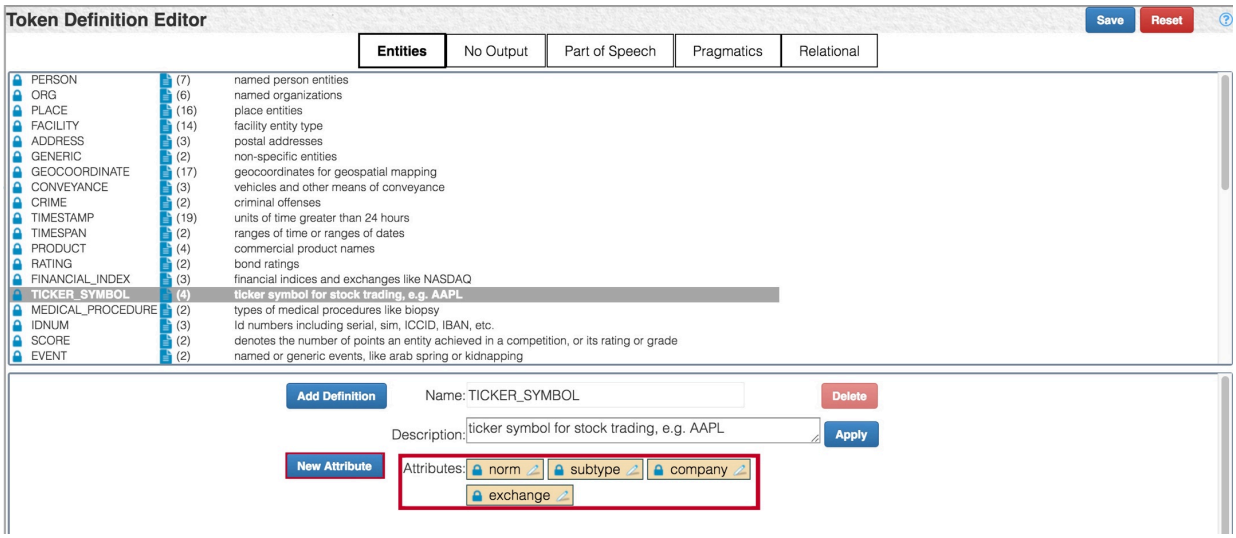
Adding and modifying attributes

Clicking **New Attribute** displays a dialog where you can select attributes to add to the entity type. To modify the behavior of an attribute that you've already added - to change its default value, for example - click the pencil icon



next to its name.

© N. Harris Computer Corporation



Note: Semantic vectors and attributes with a lock



next to their name are a core part of TextChart and cannot be deleted. You can, however, add attributes to a locked semantic vector.

Updating LxBase

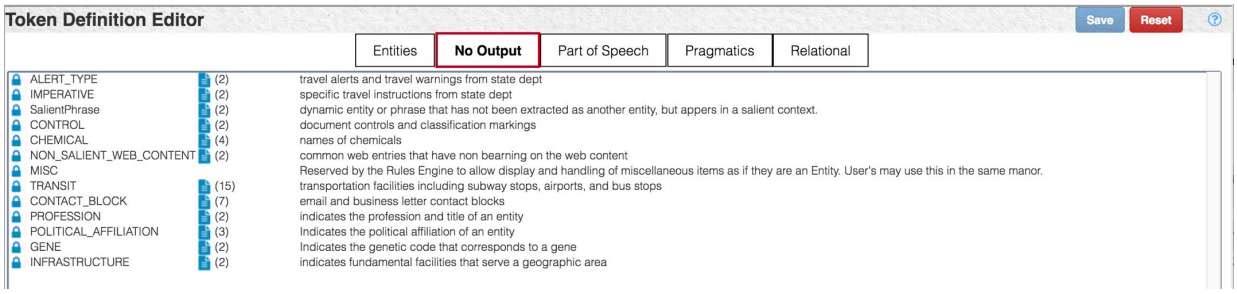
To commit your changes to the LxBase, click **Apply** and then **Save**. Then, since modifying entity types is a schema change, you must restart TextChart Studio, clear any previous processing results from the **Corpus Management** page, and reprocess the corpus.

Important: When you add a new entity type, you must associate lexical entries or linguistic rules with it in order for TextChart to extract entities of that type from documents. You should create any additional semantic vectors that will inform that extraction at the same time.

No-output entity types

By default, the LxBase is configured to suppress entities of certain entity types. Those types still have linguistic rules and lexical entries associated with them, and TextChart still identifies entities with those types, but they do not appear in extraction results.

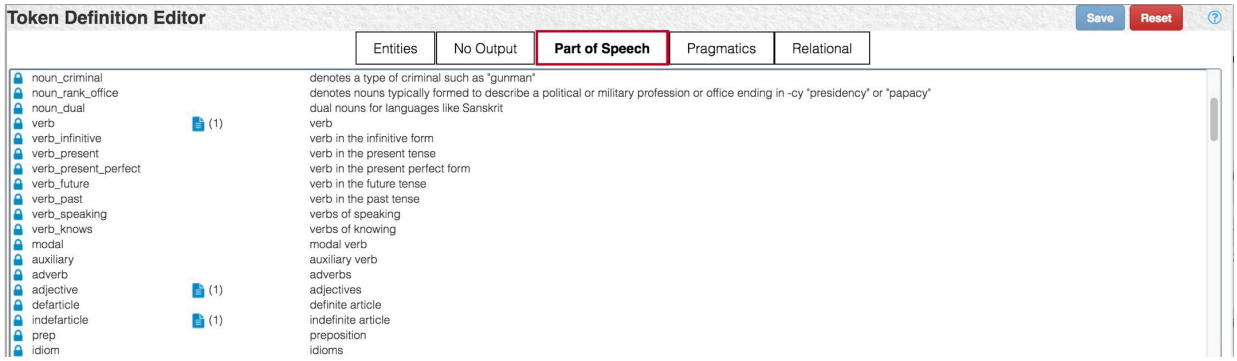
© N. Harris Computer Corporation



If your extraction results contain entities of types that you're not interested in, or if you want entities of suppressed types to appear in your results, you can move entity types between these categories. To do so, drag an entity type from the **Entities** tab to the **No Output** tab, or vice versa.

Parts of speech

Part-of-speech semantic vectors assign syntactic meaning such as "noun", "verb", and "adjective" to lexical entries. Semantic vectors of this type appear on lexical entries and in linguistic rules to help establish linguistic context.

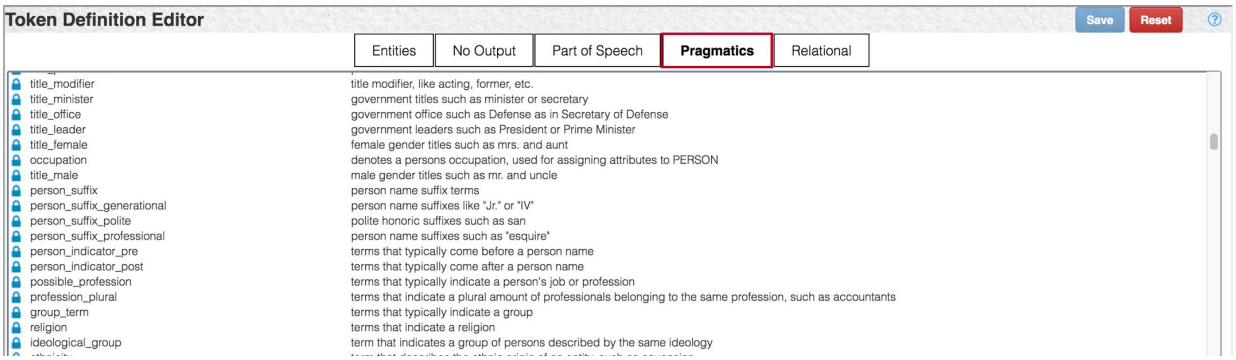


Pragmatics

Pragmatic semantic vectors assign meanings like "given name", "city name", or "office title" to lexical entries. As such, they are more specific than their part-of-speech counterparts.

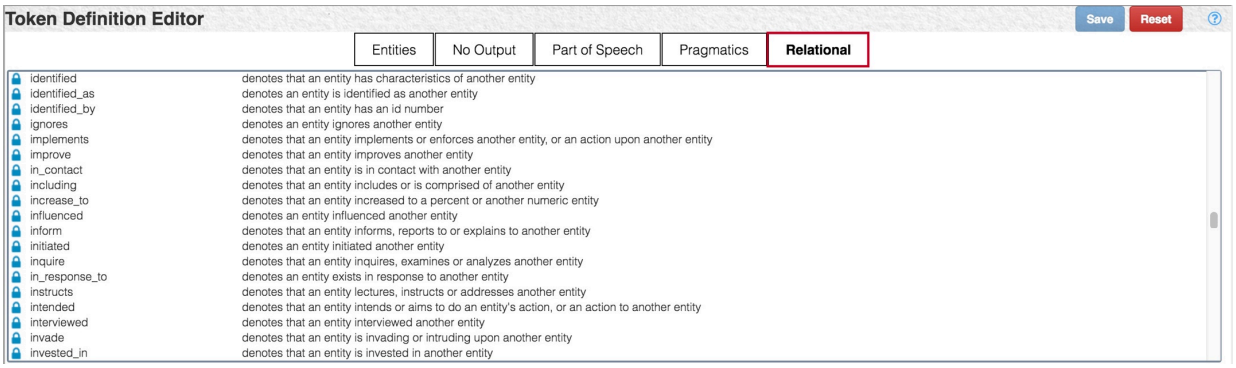
Semantic vectors of this type appear on lexical entries and in linguistic rules to make use of linguistic context and find entity-specific context.

© N. Harris Computer Corporation



Relationals

Relational semantic vectors assign predicates like "identified as", "alias of", or "influenced" to a word or a small phrase. Semantic vectors of this type appear on lexical entries, where they're used to prompt extraction of a relationship between two entities.



Modifying relationships

TextChart extracts relationships between entities through the use of predicate-subject-object (PSO) types that identify those relationships. It uses relational semantic vectors to designate words or small phrases as predicates.

For example, the TextChart dictionaries might associate a relational semantic vector like "interviewed" with a number of lexical entries:

```
<lex><word>interview</word><sv><interviewed/></sv></lex>
<lex><word>interviews</word><sv><interviewed/></sv></lex>
<lex><word>interviewing</word><sv><interviewed/></sv></lex>
<lex><word>interviewed</word><sv><interviewed/></sv></lex>
<lex><word>interrogate</word><sv><interviewed/></sv></lex>
<lex><word>interrogates</word><sv><interviewed/></sv></lex>
<lex><word>interrogating</word><sv><interviewed/></sv></lex>
<lex><word>interrogated</word><sv><interviewed/></sv></lex>
```

© N. Harris Computer Corporation

If TextChart tags all of these words in a document with the "interviewed" semantic vector, then a relationship will be extracted any time one of them appears as a predicate in context with two entities.

*During **David Frost**'s **interview** of former President **Richard Nixon**, Frost very bluntly asked why he didn't burn the tapes.*

In the example phrase above, both "David Frost" and "Richard Nixon" are extracted as PERSON entities. A PERSON to PERSON relationship is also extracted, because the predicate "interview" is tagged with the "interviewed" relational semantic vector.

In PSO terms, the *subject* of the relationship is "David Frost", the *object* is "Richard Nixon", and the *predicate* is "interview".

Modifying PSO relationship types

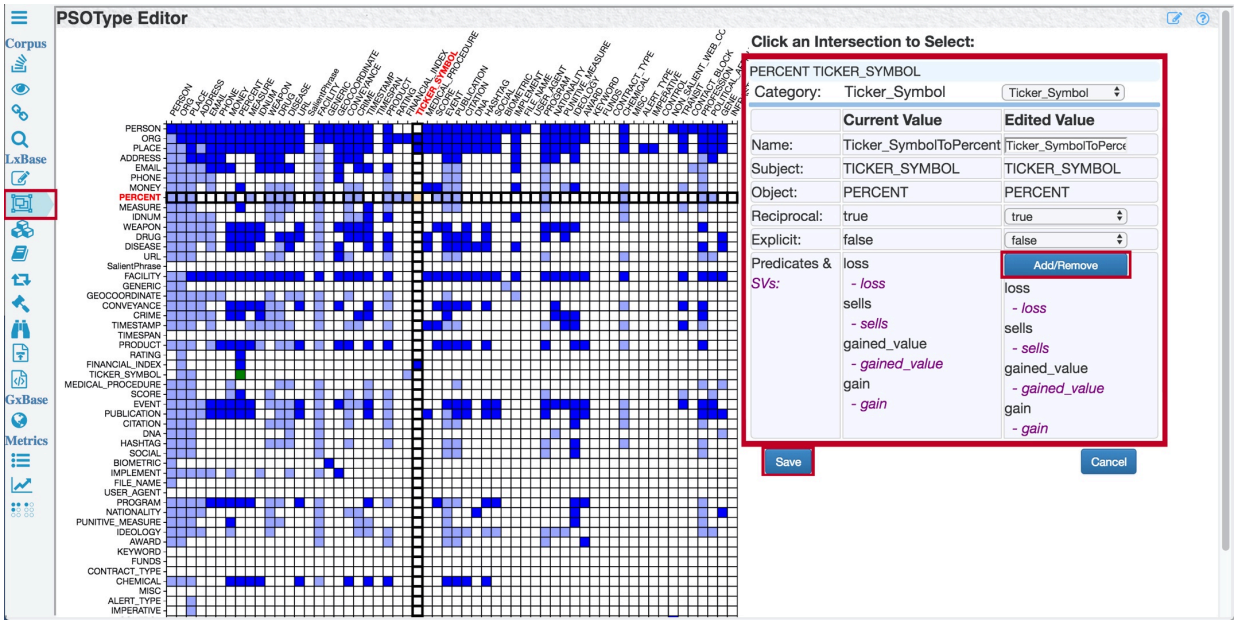
To create or edit PSO relationship types in TextChart Studio, click **Edit PSO types** in the **LxBase** section of the vertical toolbar. You can make modifications through a graphical interface, or by modifying an XML file directly.

The sections below demonstrate using both approaches to add a relational semantic vector named `closed_at` to the list of semantic vectors that are associated with relationships between TICKER_SYMBOL and PERCENT entities.

Before you begin, make sure that the `closed_at` semantic vector is present in the list in the **Relational** tab in TextChart Studio. If it's not there, create it, and then follow either of the following procedures.

GUI

The **PSOType Editor** page displays the graphical user interfaces for modifying relationship types by default.



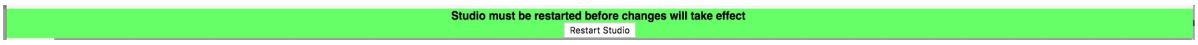
1. In the grid, click on the intersecting point between TICKER_SYMBOL and PERCENT.

© N. Harris Computer Corporation

2. In the section on the right, click **Add/Remove** to open the **Predicate Editor** dialog.



3. In the bar at the top, type a predicate that you'll associate with the new semantic vector, and then click **Add**. In this example, the predicate is also `closed_at`.
4. To associate a semantic vector with the new predicate, click the new **closed_at** button. In the new window on the right, type `closed_at` in the field at the top of the **Semantic Vectors** list. Select `closed_at` from the list.
5. Close all the new windows, and then click **Save**. TextChart Studio displays a message at the top of the application window, prompting a restart. Click **Restart Studio**.

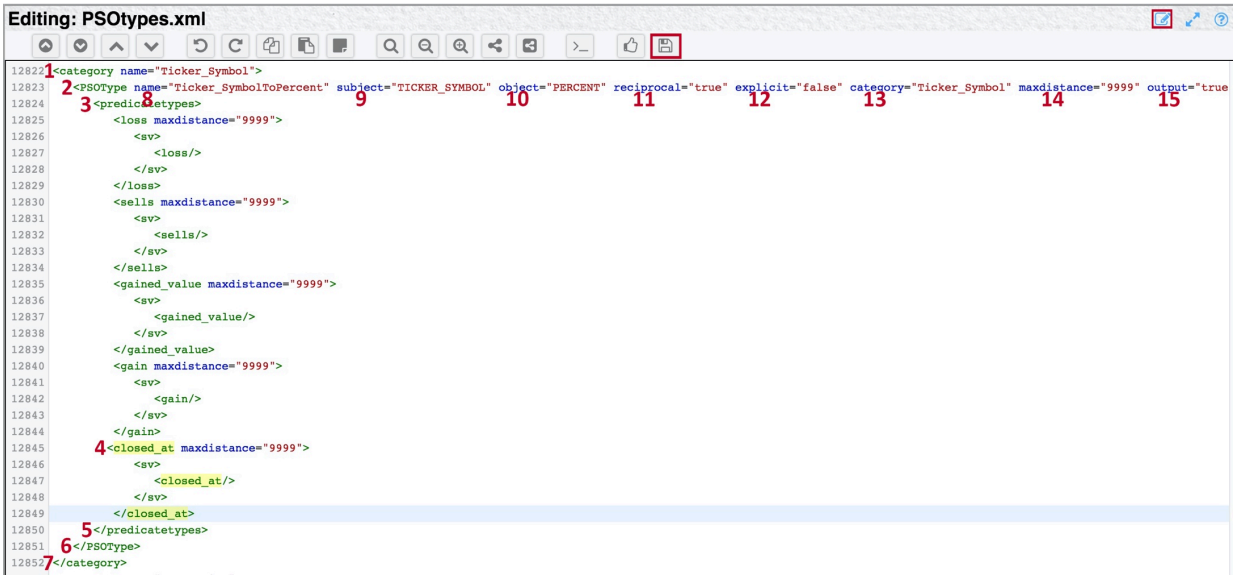


6. Add the `closed_at` semantic vector to relevant entries in the lexicon.

XML

To modify PSO relationship types directly, click **Edit PSO types in XML Mode** in the top right corner of the **PSOType Editor** window.

© N. Harris Computer Corporation



Then, you need to find the definition of the relationship type that you want to modify, and add a block of XML like this to it:

```
<users_predicate maxdistance="9999">
  <sv>
    <users_semantic_vector/>
  </sv>
</users_predicate>
```

For example, to add a predicate named `closed_at` (and its associated semantic vector) to the Ticker_SymbolToPercent relationship type, find the <PSOType> element for that type in the file, and then add the following before the `</predicatetypes>` tag:

```
<closed_at maxdistance="9999">
  <sv>
    <closed_at/>
  </sv>
</closed_at>
```

After you edit the file, click the **Save** button, and restart TextChart Studio.

Creating a PSO relationship type

To create a relationship type when a category for one of the entity types involved is not present in the XML file, you need to add a complete `<category>` element.

For example, to create a type for relationships between TICKER_SYMBOL and PERCENT entities in the case where neither has its own category, you'd need to add (and modify) the following code:

```
<category name="Entity_Type1">
  <PSOType name="Entity_Type1ToEntity_Type2" subject="Entity_Type1"
           object="Entity_Type2" reciprocal="true" output="true"
           explicit="false" category="Entity_Type1">
    <predicatetypes>
      <users_predicate maxdistance="9999">
        <sv>
```

© N. Harris Computer Corporation

```
          <users_semantic_vector/>
        </sv>
      </users_predicate>
    </predicatetypes>
  </PSOType>
</category>
```

On the <PSOType> element, the `reciprocal` attribute is `"true"` by default, which means that in our example, TextChart extracts both TICKER_SYMBOL to PERCENT relationships and PERCENT to TICKER_SYMBOL relationships. Setting it to `"false"` means extracting only TICKER_SYMBOL to PERCENT relationships.

Conversely, the `explicit` attribute is `"false"` by default, which makes TextChart extract relationships that are explicitly defined as TICKER_SYMBOL to PERCENT, *and* other relationships that involve entities of those types, as in both examples below:

1. AAPL **closed at** a 3% increase yesterday.
2. AAPL, +3%, 11-01-2018.

Setting the attribute to `"false"` means extracting only explicitly TICKER_SYMBOL to PERCENT relationships, as in the first example above.

Finally, setting the `output` attribute to `"true"` means that the result of extracting this relationship will be output.

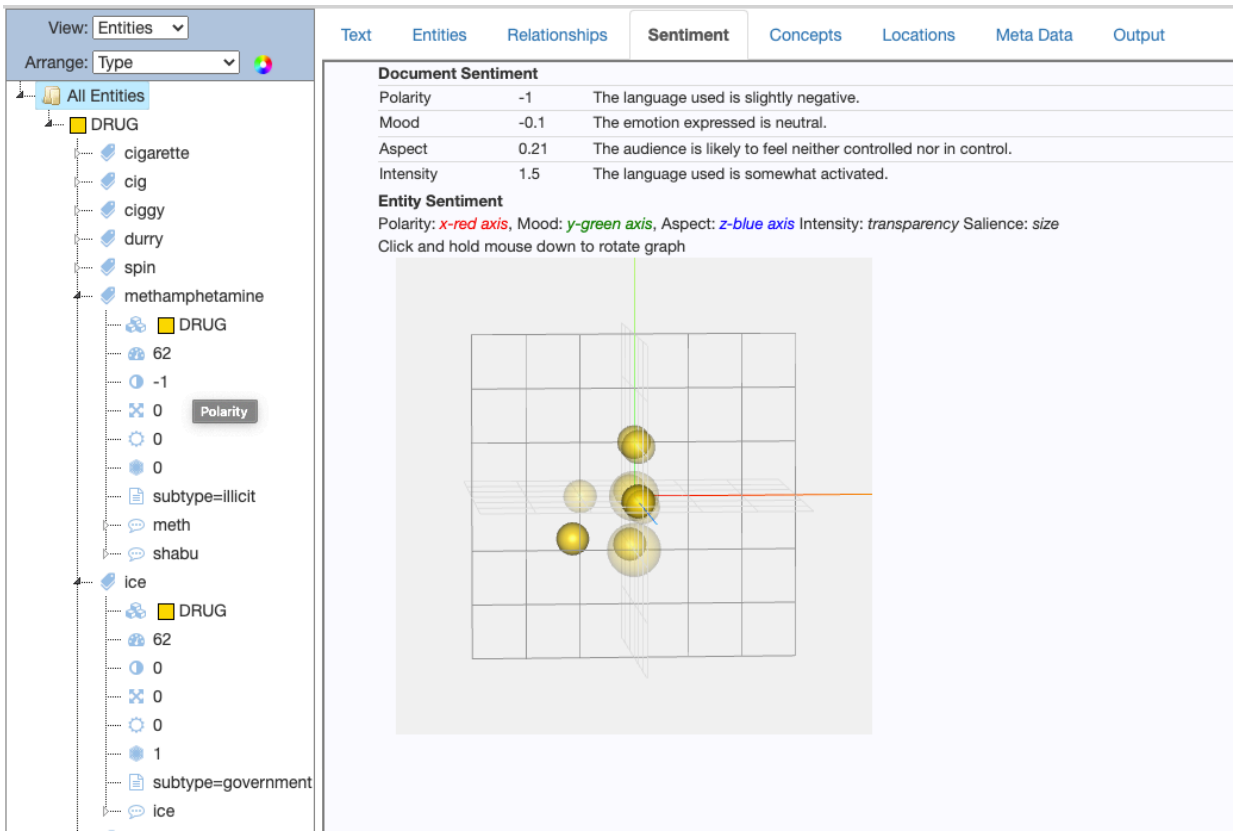
Modifying sentiment scores

i2 TextChart performs sentiment analysis by assigning scores to the polarity, mood, aspect, and intensity of each dictionary entry. You can use TextChart Studio to modify the sentiment scores of individual dictionary entries manually.

In the following example, a list of slang terms for DRUG entities has been added to a new dictionary file using the word import tool.

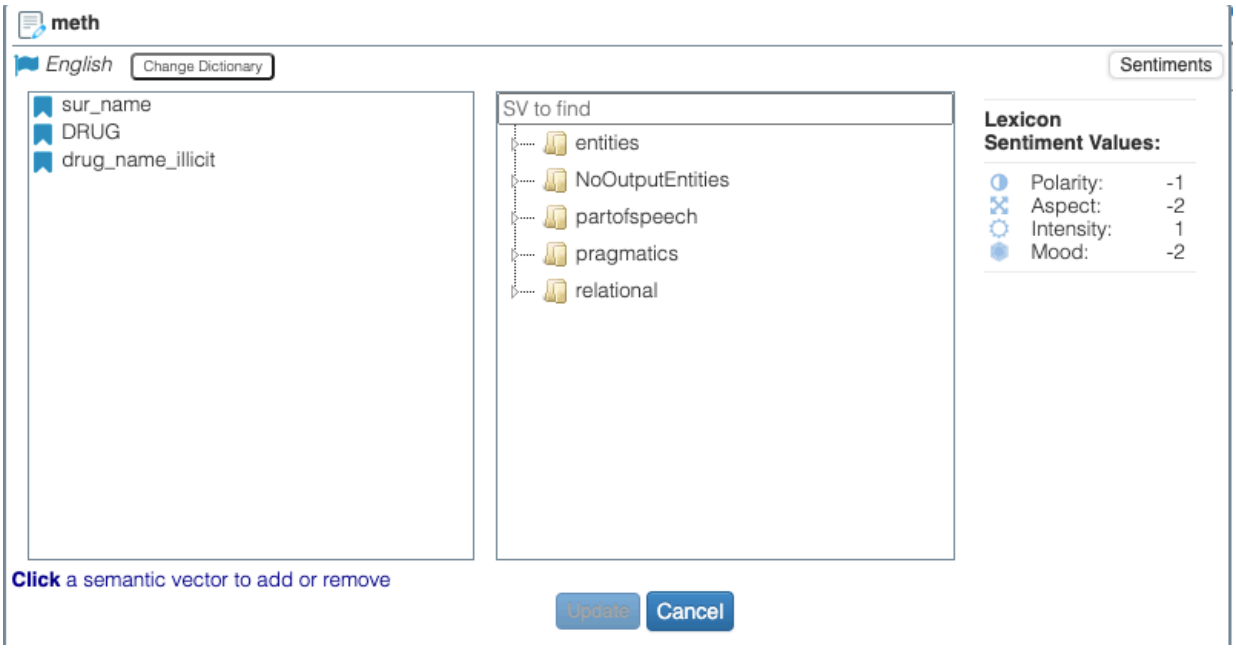
The next image shows the **Sentiment** tab displaying results for these DRUG terms, as well as the individual sentiment results for "methamphetamine" and "ice." In this case, "methamphetamine" has a negative polarity score out-of-the-box, while "ice" does not have a sentiment score.

© N. Harris Computer Corporation

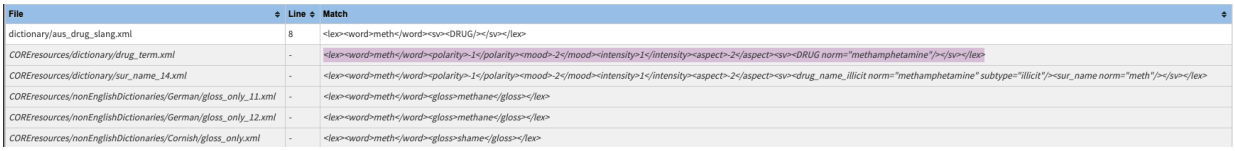


To see what out-of-the-box information is associated with a particular term (and to view the XML of the dictionary entry), right-click its name in the document view and select **Look up in lexicon** to display the following dialog.

© N. Harris Computer Corporation



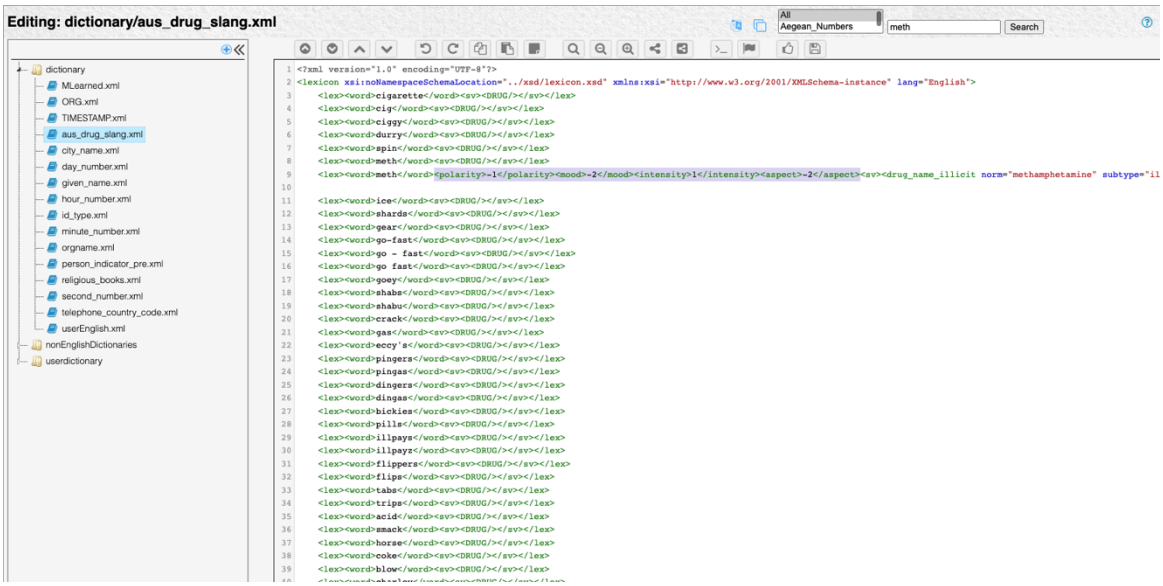
Click the "Notepad" icon in the upper left corner of the the dialog to display the individual dictionary entries associated with the selected result. Then, you can simply copy the relevant information. In this example, the XML for "meth" has been highlighted and copied.



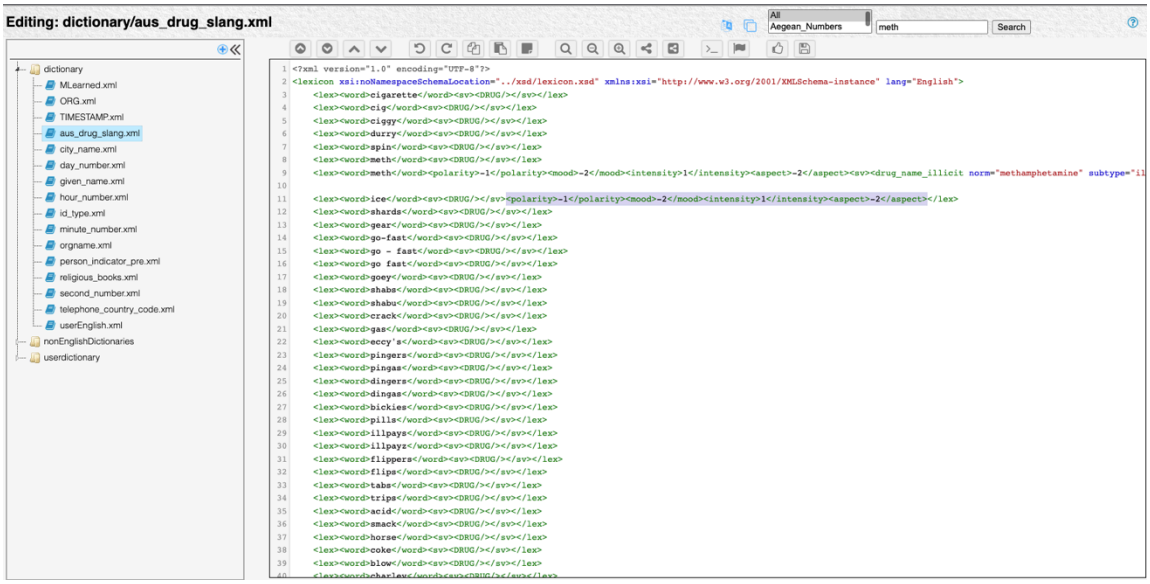
File	Line	Match
dictionary/aus_drug_slang.xml	8	<lex><word>meth</word><sv><DRUG/></sv></lex>
CORE resources/dictionary/drug_term.xml	-	<lex><word>meth</word><polarity>-1</polarity><mood>-2</mood><intensity>1</intensity><aspect>-2</aspect><sv><DRUG norm="methamphetamine"/></sv></lex>
CORE resources/dictionary/sur_name_14.xml	-	<lex><word>meth</word><polarity>-1</polarity><mood>-2</mood><intensity>1</intensity><aspect>-2</aspect><sv><drug_name_illicit norm="methamphetamine" subtype="illicit"/><sur_name norm="meth"/></sv></lex>
CORE resources/nonEnglishDictionaries/German/gloss_only_11.xml	-	<lex><word>meth</word><gloss>methane</gloss></lex>
CORE resources/nonEnglishDictionaries/German/gloss_only_12.xml	-	<lex><word>meth</word><gloss>methane</gloss></lex>
CORE resources/nonEnglishDictionaries/Cornish/gloss_only.xml	-	<lex><word>meth</word><gloss>shame</gloss></lex>

Now you can navigate to another dictionary file and paste the XML where you choose. In this example, the XML for "meth" has been pasted into the drug slang dictionary.

© N. Harris Computer Corporation

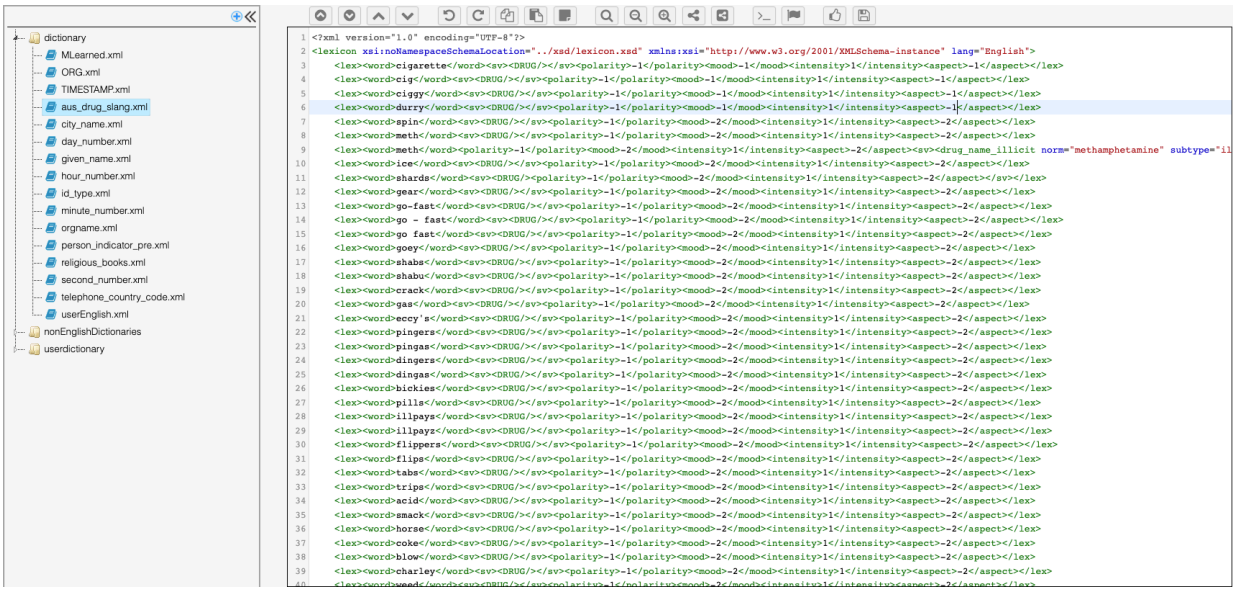


You can add the same sentiment scores to other terms in the same file just by copying the XML from one term to another.

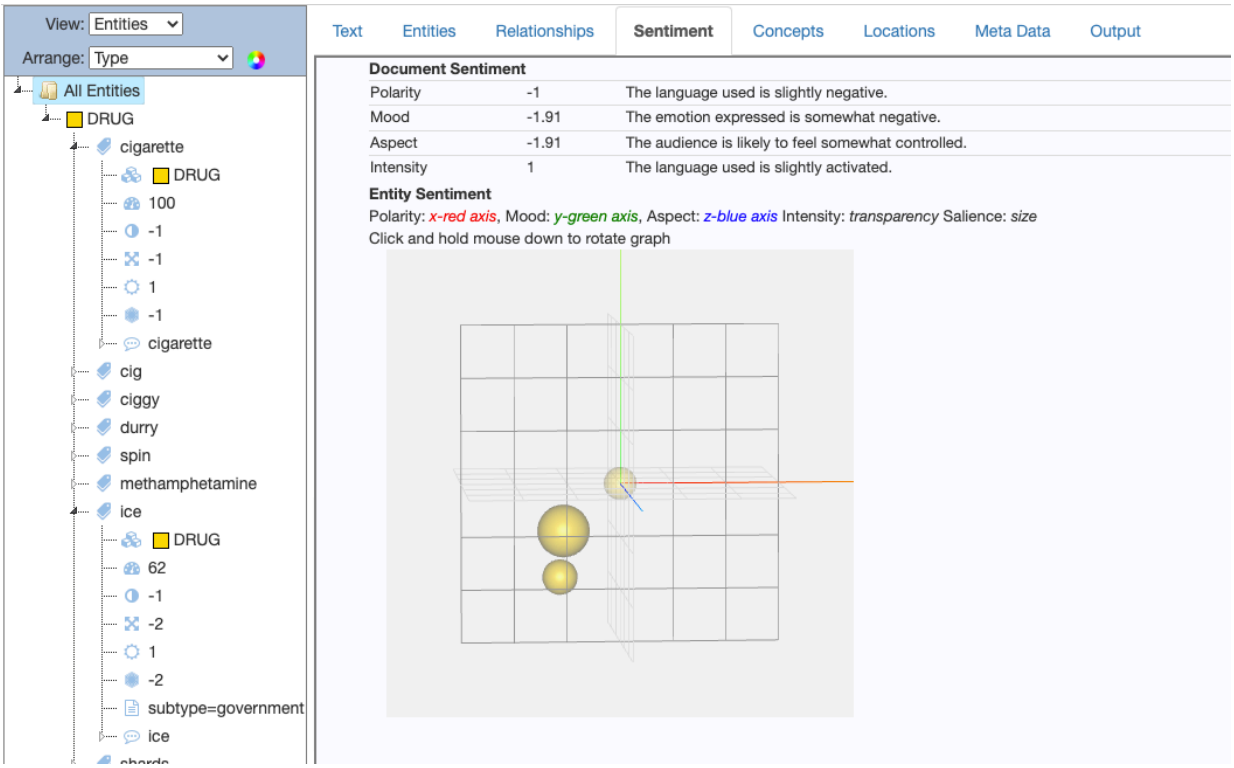


In the next example, all of the drug entries are associated with the same scores, except for the tobacco-type terms, which have less intense scores.

© N. Harris Computer Corporation



After you reprocess the document, the scores in the **Sentiment** tab are different, as are the scores for individual DRUG entity results.



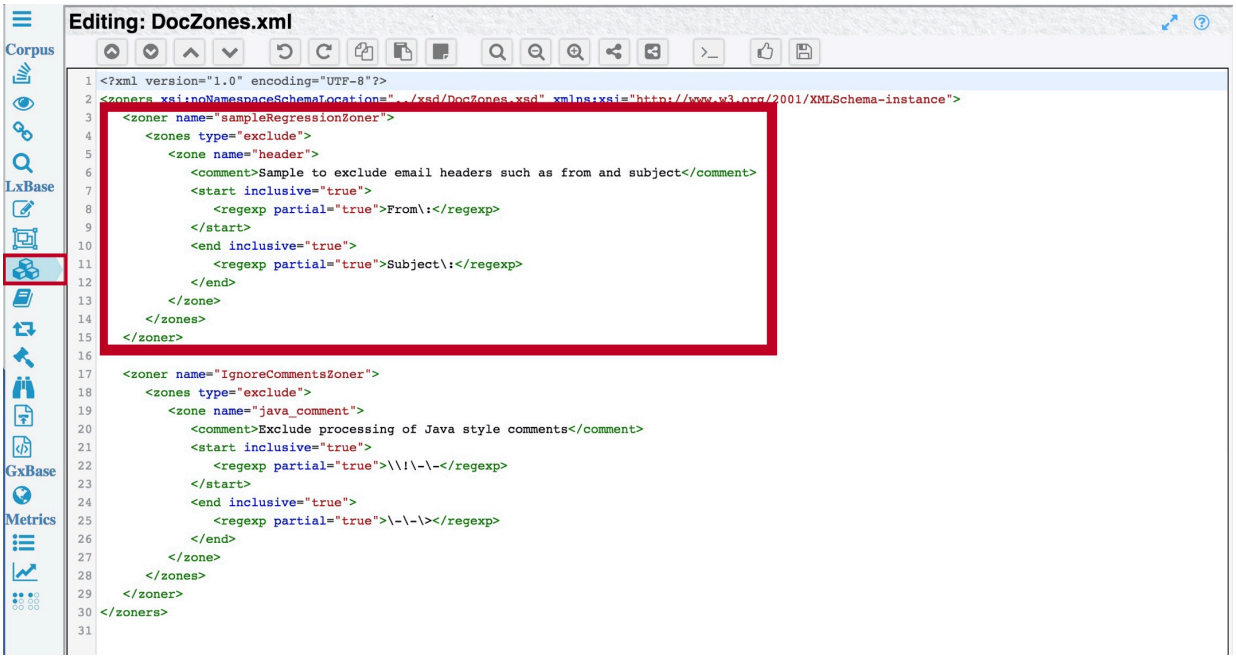
© N. Harris Computer Corporation

Document zoning

If the corpora that you process with TextChart regularly contain documents with sections that you want to skip, you can configure LxBase to ignore them. TextChart Studio enables this behavior through *document zoning*.

In zoning, you provide the definitions of some text that marks the beginning and the end of zones, and say whether you want to include those zones in processing. These definitions appear in an XML file that you can edit through TextChart Studio.

To view or modify the document zoning XML file, click **Configure document zoning** in the **LxBase** section of the vertical toolbar.



The default zoning file contains the definitions of two *zoners*: collections of zone definitions that have similar aims. In this file, each zoner contains a single zone.

The zone definitions themselves contain regular expressions describing the text that starts and ends the document zones that you want to control processing for.

The first zone definition in the file demonstrates a way of ignoring some email headers during processing. As it appears above, the zone is excluded from processing by a setting in its enclosing `<zones>` element:

```
<zones type="exclude">
```

To include the headers in processing, you can permanently delete the zone definition, or temporarily edit the element:

```
<zones type="include">
```

For the `<start>` and `<end>` elements that define the start and end of document zones, you can use the `inclusive` attribute to say whether the text that matched the regular expression is a part of the zone. `inclusive="true"` means that it is; `inclusive="false"` means the opposite.

© N. Harris Computer Corporation

For the regular expressions, you can specify whether a match effectively selects only the matching text (`<regexp partial="true">`), or the whole line that contains the matching text (`<regexp partial="false">`). In other words, you control whether the document zone can start or end in the middle of a line, or if it always includes whole lines.

XML zoners

If your corpora include documents in XML format, then you can use zoners to target specific elements for inclusion and exclusion from processing, instead of using regular expressions.

You define XML zoners in the same file as the zoners that use regular expressions, although only one zoner can be active at a time.

To create an XML zoner, use `<includedXml>` and `<excludedXml>` elements in place of the `<zones>` element in the `<zoner>` definition:

```
<zoner name="datastream">
  <includedXml>
    <tag include="false">legis-body</tag>
  </includedXml>
  <excludedXml>
    <tag>section</tag>
  </excludedXml>
</zoner>
```

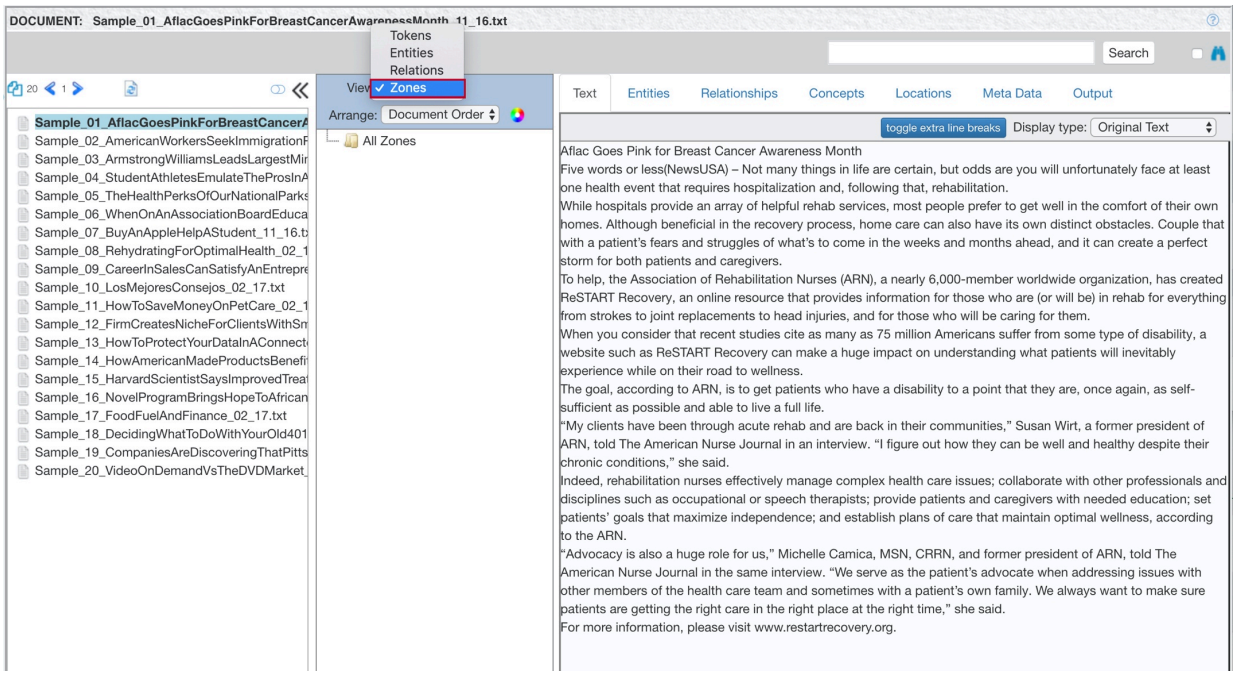
With this zoner definition, TextChart processing creates document zones for the `<legis-body>` elements in an XML document. However, if a zone contains a `<section>` element, then its contents are excluded, effectively splitting the zone into two or more pieces.

The `include` attribute of the `<tag>` element in the definition controls whether the opening and closing tags of the specified element are included in the zone. In general, you'll use `include="false"` for inclusion, but `include="true"` for exclusion.

For the XML zoner to work, it needs XML. For the moment, set the LxProperties item "rawinput" to "true" to avoid Tika processing.

To view the parts of a document that match a particular zone definition, use the **Zones** setting in the single document view.

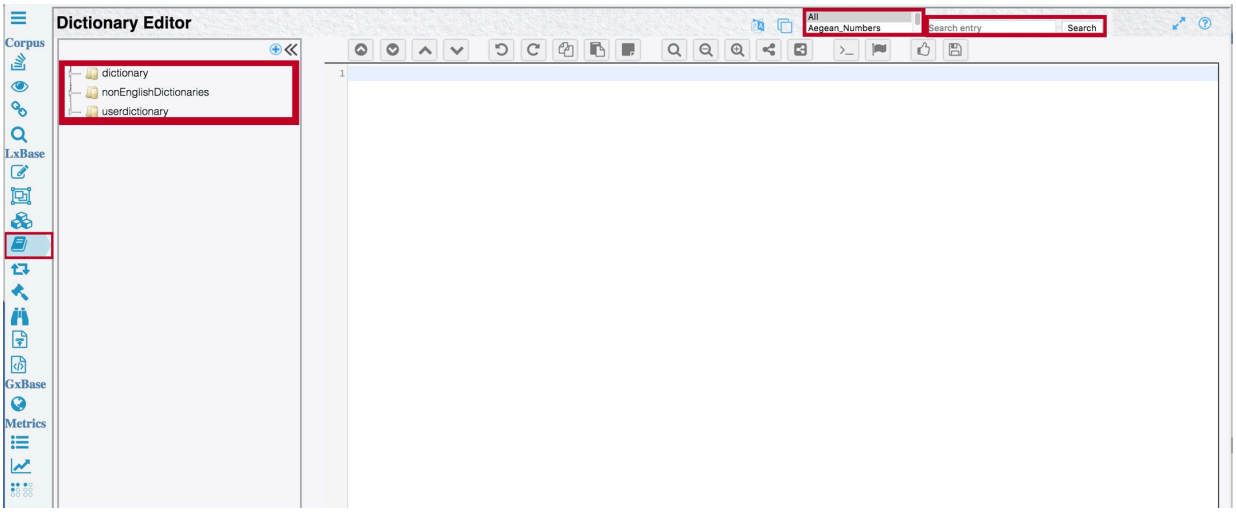
© N. Harris Computer Corporation



Modifying lexicons

In TextChart studio, you can modify lexical items directly in the lexicon, or from the document view. To open the dictionary editor, click **Edit lexicon** in the **LxBase** section of the vertical toolbar.

The **Dictionary Editor** page includes a search box at the top for searching in specific dictionaries or the lexicon as a whole. Lexical items that are not in the core can be modified from this view.



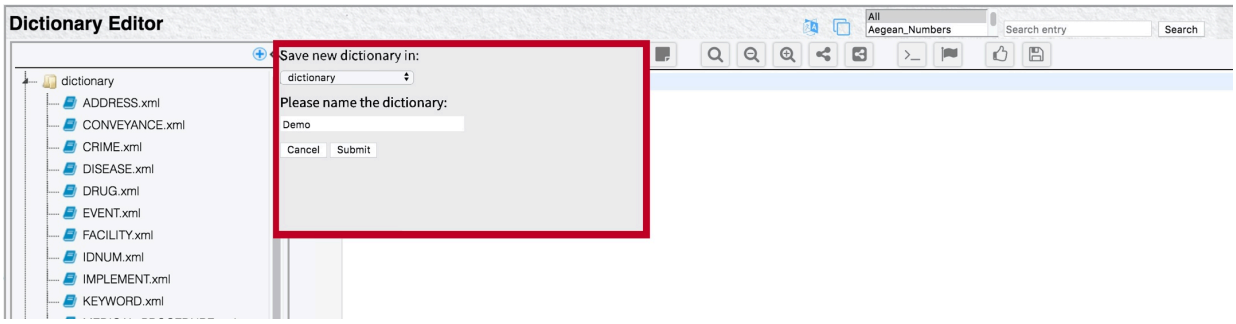
© N. Harris Computer Corporation

The toolbar below the search box provides standard features to assist with navigating and customizing dictionaries. The buttons next to the search box enable you to enable or disable lexical entries, and to check for duplicate entries across multiple dictionaries.

Creating dictionaries

i2 TextChart can only extract information from documents when they contain words and terms that appear in its dictionaries. If your organization regularly uses industry-specific terms that are not present in the standard TextChart dictionaries, you can create your own dictionary and add those terms to it.

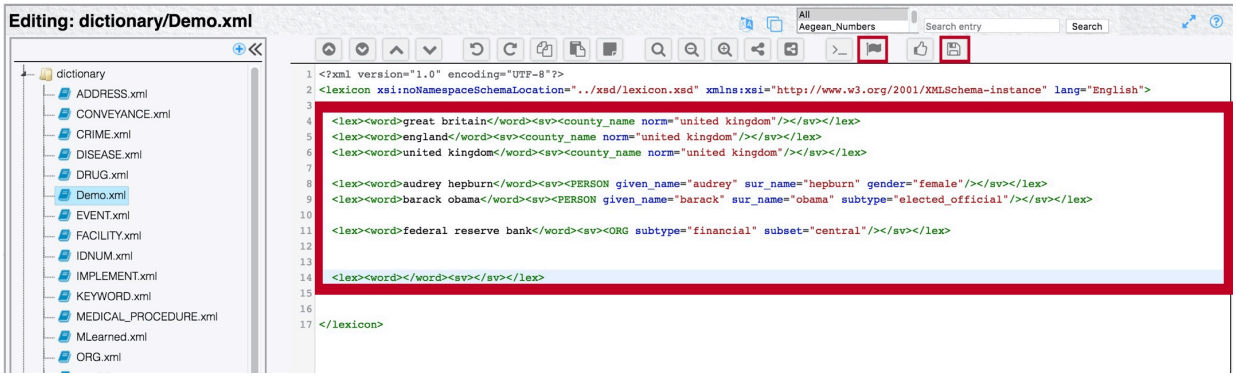
To create a dictionary, clicking the "plus" icon above the tree view in the dictionary editor. TextChart Studio displays a dialog where you can provide the name and location of the new dictionary.



To add a lexical entry to the new dictionary, click the "flag" button in the toolbar to generate an XML template.

```
<lex><word></word><sv></sv></lex>
```

As well as the term itself, which you type in the `<word>` element, there are additional attributes that you can use to enrich your extraction results.



For example, if you wanted to normalize processing so that all references in documents to "England", "Great Britain", or "United Kingdom" resolve to just "United Kingdom", you can use the `norm` attribute on the `country_name` semantic vector:

```
<lex><word>great britain</word><sv><country_name norm="united kingdom"/></
sv></lex>
```

© N. Harris Computer Corporation

```
<lex><word>england</word><sv><country_name norm="united kingdom"/></sv></
lex>

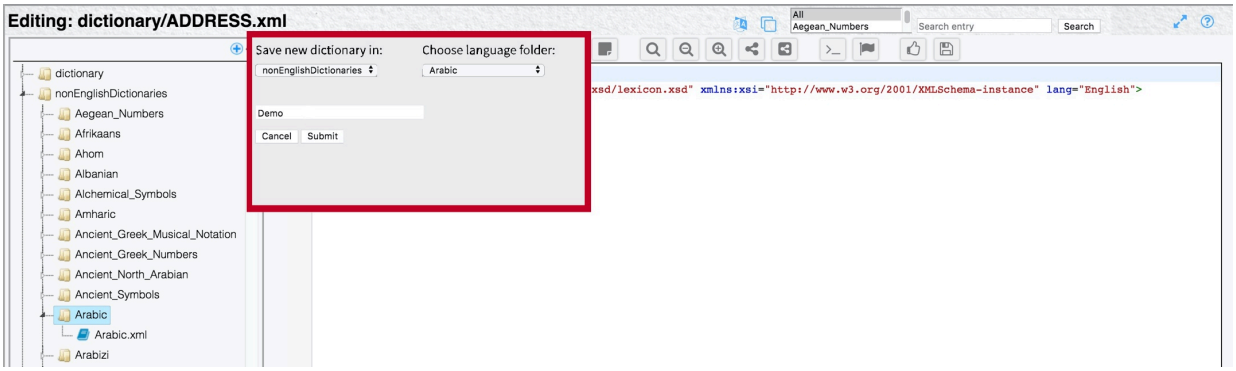
<lex><word>united kingdom</word><sv><country_name norm="united kingdom"/></
sv></lex>
```

Alternatively, you can configure lexical entries so that terms to be extracted as PERSON entities always produce results with the same form. For example, to extract "Audrey Hepburn" as "Audrey Hepburn" and not "V. Audrey Hepburn", or to arrange that "Barack Obama" always has the subtype "elected official":

```
<lex><word>audrey hepburn</word><sv><PERSON given_name="audrey"
 sur_name="hepburn" gender="female"/></sv></lex>

<lex><word>barack obama</word><sv><PERSON given_name="barack"
 sur_name="obama" subtype="elected_official"/></sv></lex>
```

You can also create non-English, industry-specific dictionaries. When you click the "plus" icon above the tree view in the dictionary editor, TextChart Studio displays a dialog where you can select the appropriate language for your dictionary.

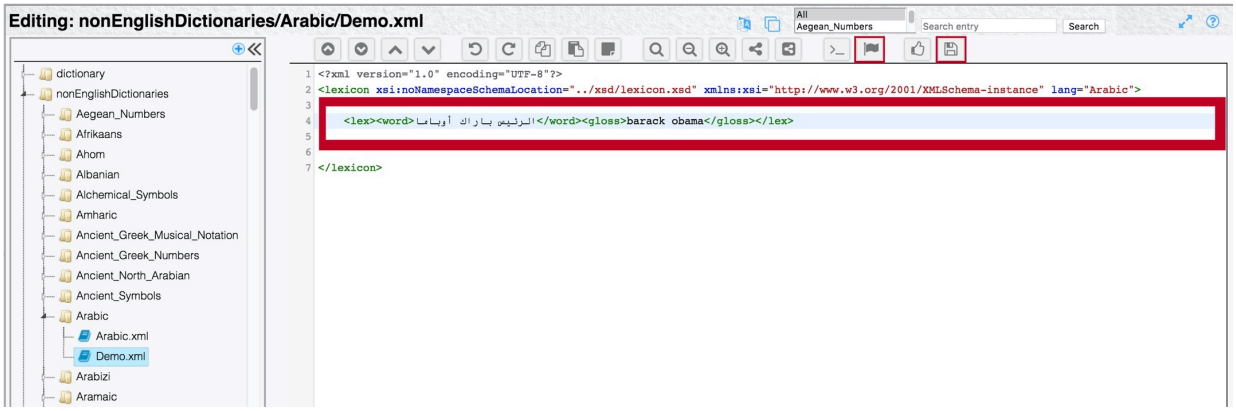


You can industry-specific knowledge to any available non-English dictionary. To do so, click the "flag" button in the toolbar to generate an XML template:

```
<lex><word></word><gloss></gloss></lex>
```

To add an entry to a non-English dictionary, add the term (in the dictionary's language) to the `<word>` element, and the English gloss to the `<gloss>` element.

© N. Harris Computer Corporation



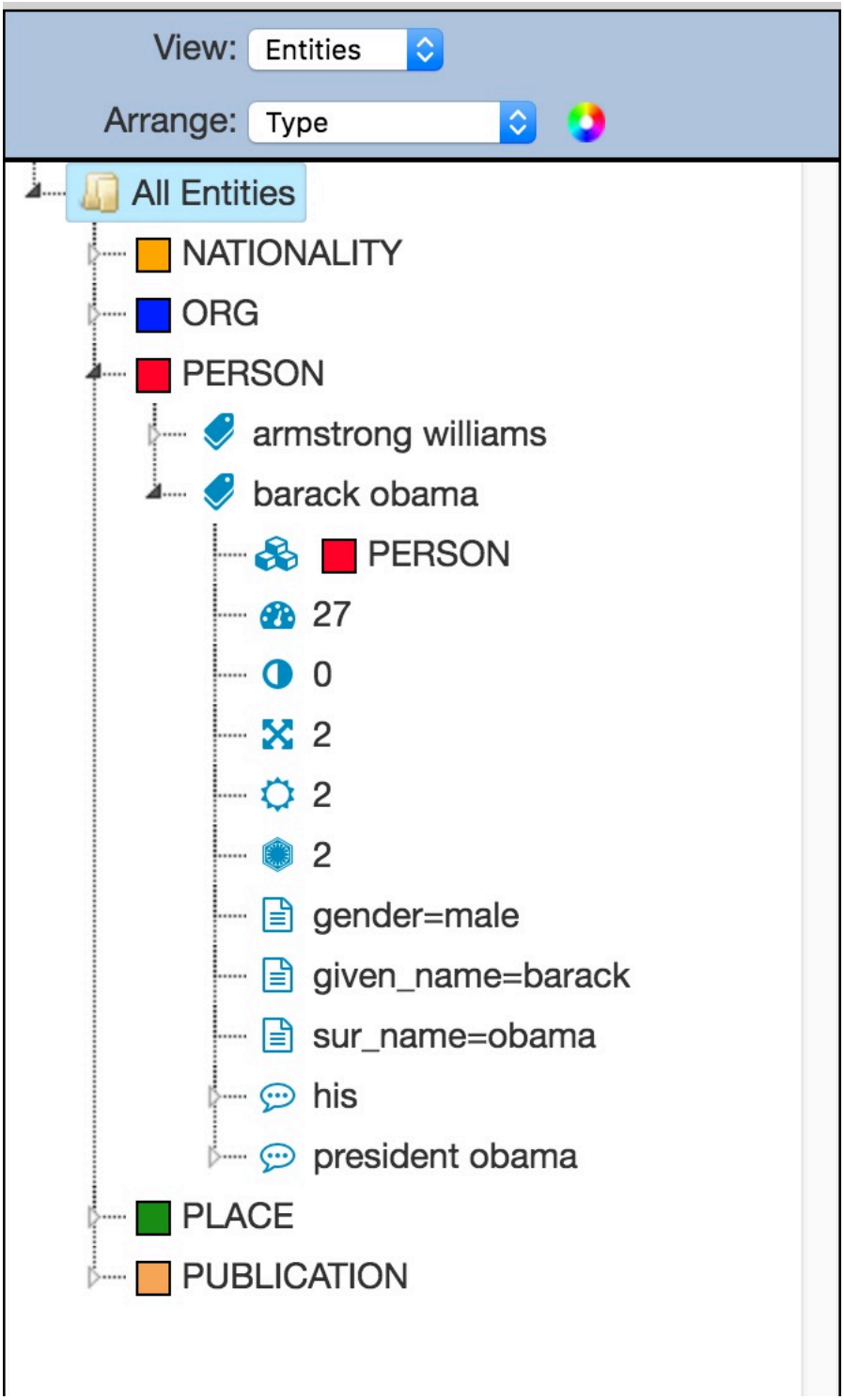
If the English gloss is already an entry in an English dictionary, TextChart uses the semantic vectors that are attached to the English entry. Optionally, you can also add semantic vectors to the non-English entry.

Anaphora resolution

Anaphora refers to words that have already been used. If a document mentions a person by name in one place, it's likely that the same person is referred to again, later in the document, through the use of pronouns or in a different variant.

TextChart uses algorithms to determine which anaphora refer to which entities.

© N. Harris Computer Corporation

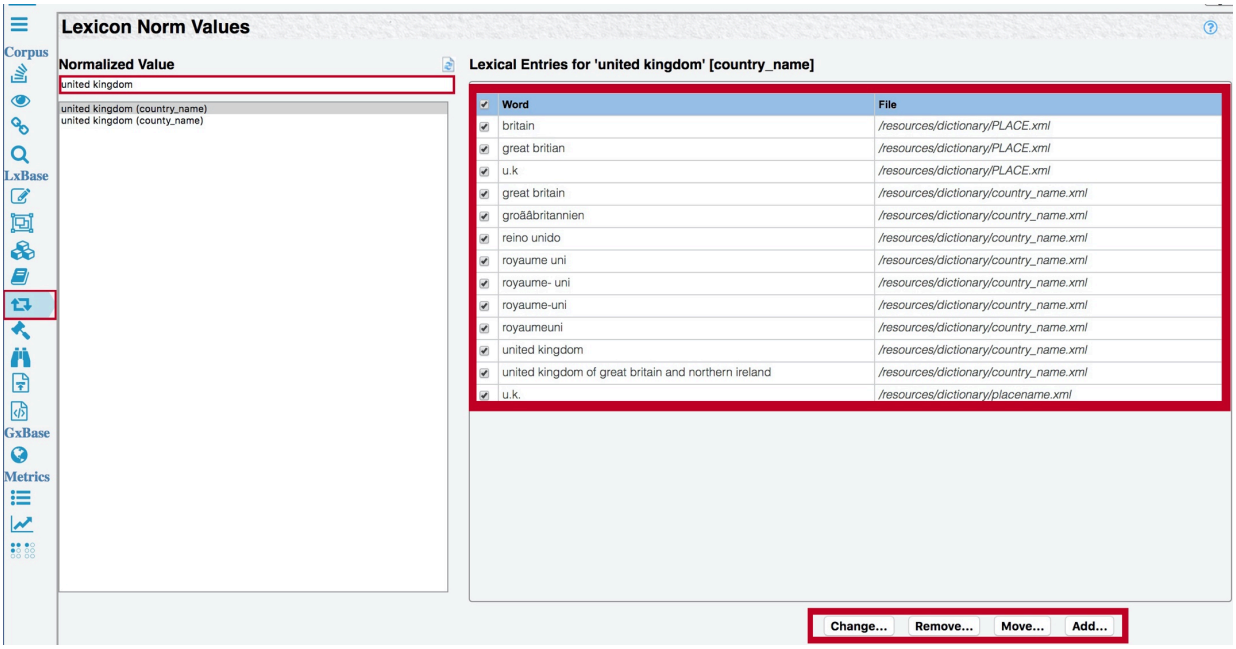


© N. Harris Computer Corporation

Managing normalization

When you create dictionaries in TextChart Studio, you can arrange for lexical entries to be *normalized* to a standard entry. TextChart Studio also provides the ability to manage all such normalization in one place.

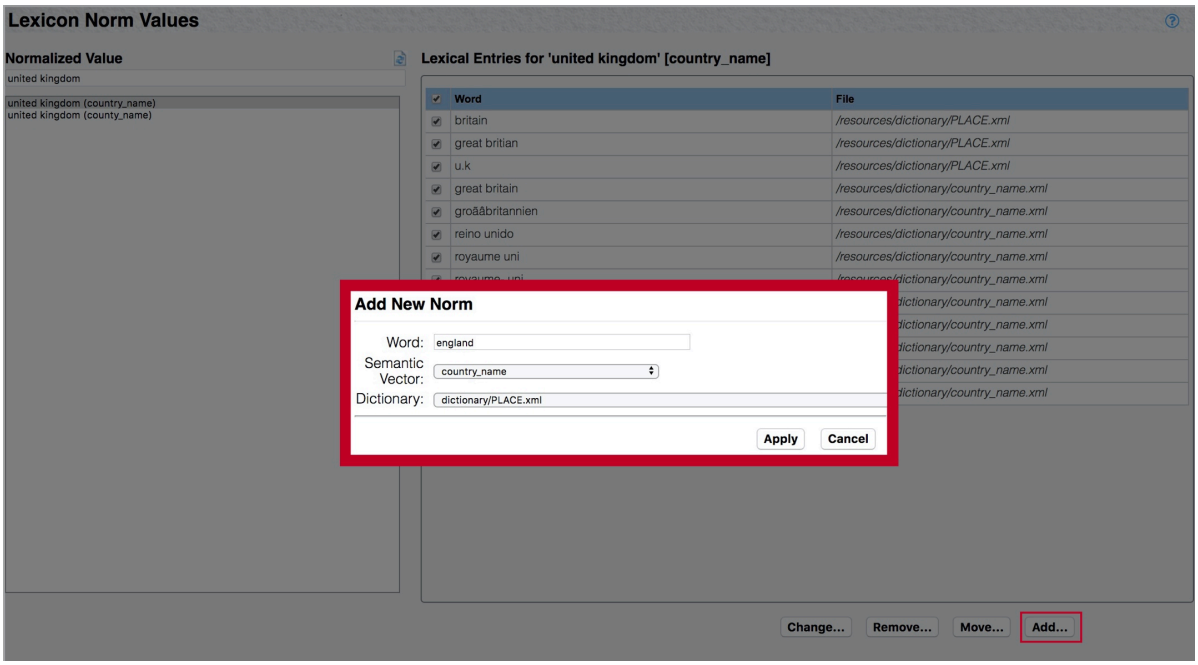
To see a list of the lexical entries that have been normalized, click **Edit lexical norm values** in the **LxBase** section of the vertical toolbar to open the **Lexical Norm Values** page.



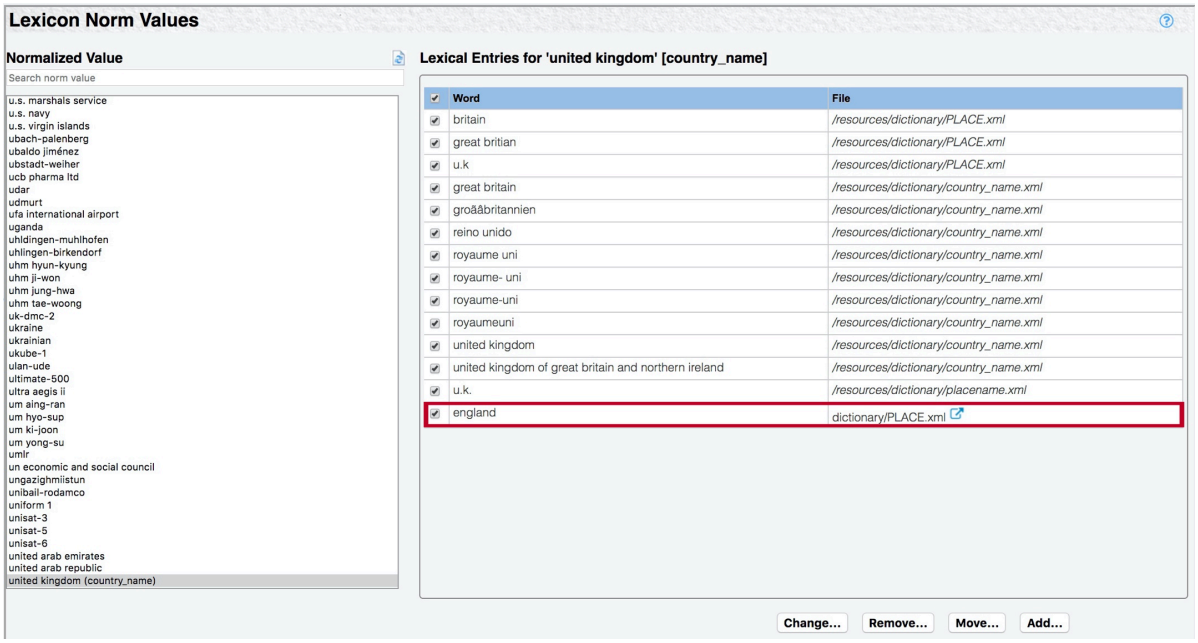
Selecting a normalized value on the left of the page populates the list on the right with a list of the lexical entries that normalize to it. The buttons below the list provide the following features:

- **Change** changes the normalization value for all the selected entries.
- **Remove** removes the selected entries.
- **Move** moves the value from one semantic vector to another in the same entry.
- **Add** creates a lexical entry that uses this normalized value. You can specify the term, and assign it a semantic vector and a dictionary. For example:

© N. Harris Computer Corporation



When you click **Apply**, the new entry appears in the list:

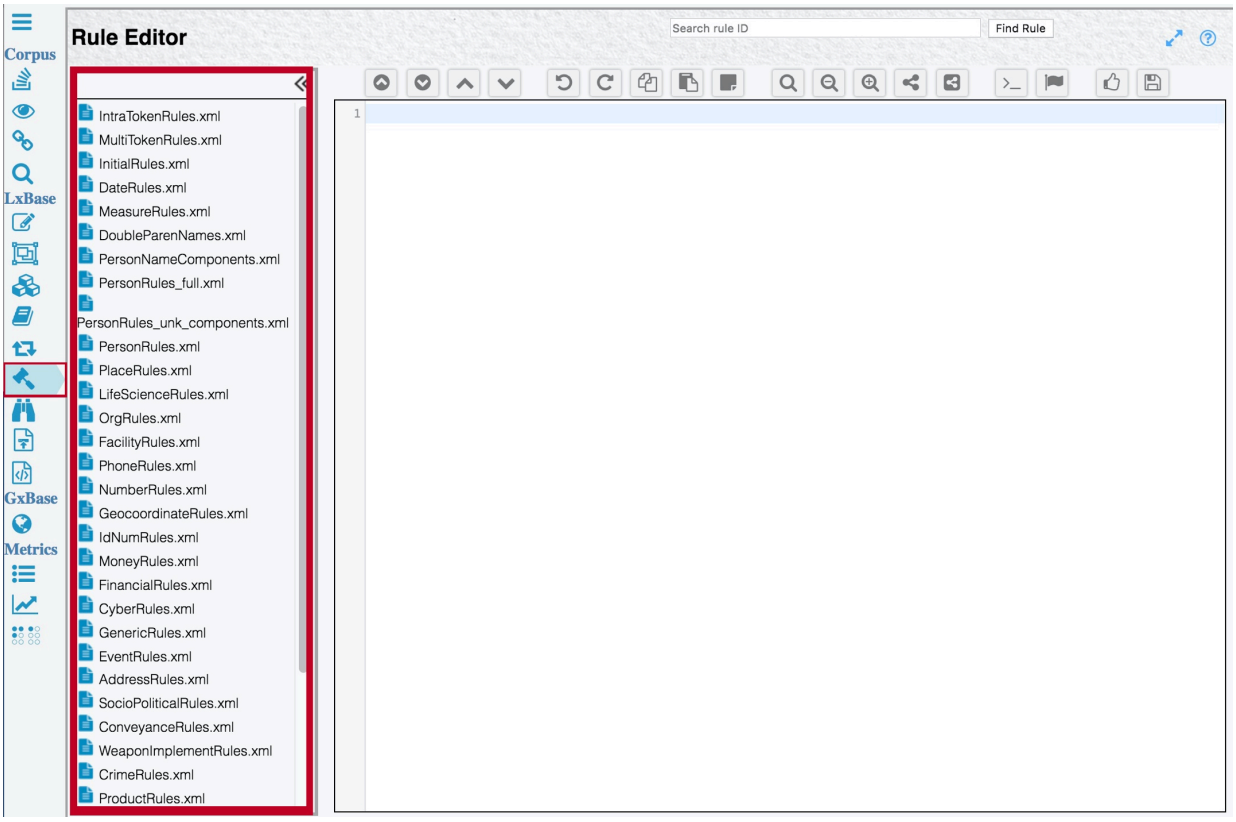


© N. Harris Computer Corporation

Editing TextChart rules

i2 makes all of the rules that control extraction available in a set of XML files. You can use TextChart Studio to modify the rules in the supplied set, or to write new rules of your own.

TextChart rules specify the linguistic patterns that entities must match in order to be extracted. TextChart reads and applies rule files in the order they appear in `RuleFileList.xml`. To access this file, click **Manage LxBase** in the **LxBase** section of the vertical toolbar.



Within each rule file, TextChart applies the rules in the order in which they occur. To view and edit a particular rule file, click the **Edit rules** button in the menu.

To add a new rule to the open rule file, click the "Flag" button in the horizontal toolbar. TextChart Studio adds an XML template for the new rule to the file:

```
<Rule ID="ADD ID">
  <description>ADD DESCRIPTION</description>
  <comment>ADD ANY NECESSARY COMMENTS</comment>
  <example>ADD AN EXAMPLE OF THE DESIRED RESULT</example>
  <result>
    <combine></combine>
    <sv></sv>
    <attributes></attributes>
    <nolonger></nolonger>
  </result>
  <when>
    <T offset="0">
```

© N. Harris Computer Corporation

```
      <IS><sv></sv></IS>
      <ISNOT><sv></sv></ISNOT>
    </T>
  </when>
</Rule>
```

Every TextChart rule has a unique identifier and a description. You can also provide additional comments, and an example of the desired extraction result.

The rule's logic is composed of a `<result>` clause and a `<when>` clause. The `<result>` clause includes the semantic vector (`<sv>`) that the rule creates. For example, the matching pattern might be a PERSON entity or a `sur_name` semantic vector.

The `<result>` element also optionally includes `<combine>`, `<nolonger>`, and `<attributes>` elements:

- Use `<combine>` to merge multiple tokens into the same resulting semantic vector. The value, which can be positive or negative, is the number of tokens to combine together as a match, starting with a count of 0. If the value is negative, combination happens backwards from the 0 token. This is used to look backwards for recursion.
- Use `<nolonger>` to remove semantic vectors from tokens that match the pattern.
- Use `<attributes>` to assign attributes to a particular token in a match.

The `<when>` element contains the pattern itself. For each token in the match, semantic vectors that must match are specified in the `<IS>` element, while semantic vectors that must not match are specified in the `<ISNOT>` element.

If an `<IS>` element contains two semantic vectors, they're ORed together during processing. The following example illustrates a token that must be a given name or a surname, but not a verb:

```
<T offset="0">
  <IS><sv><given_name/><sur_name/></sv></IS>
  <ISNOT><sv><verb/></sv></ISNOT>
</T>
```

To AND semantic vectors, the `<when>` element must contain instances of tokens with the same offset. In this example, the token must be both a surname and a word that starts with a capital letter:

```
<T offset="0">
  <IS><sv><sur_name/></sv></IS>
</T>

<T offset="0">
  <IS><sv><cap_word/></sv></IS>
</T>
```

You can use a negative offset to find a match in the rule, without including that particular token in the extraction result.

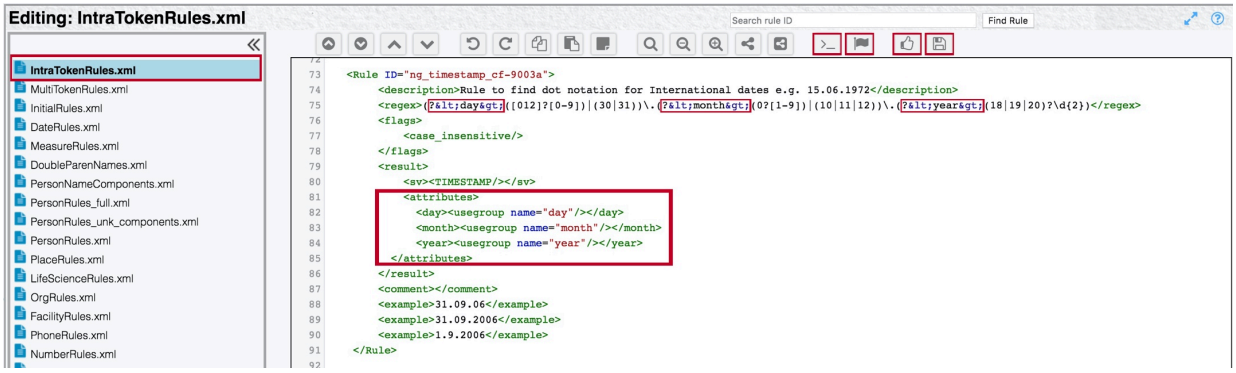
Regular expression rules

Intra-token rules and multi-token rules are used to define character-based regular expressions. The `IntraTokenRules.xml` file should only include regular expressions without character type change. Regular expressions *with* character type change should be multi-token rules. For example, if a pattern includes both letters and numbers, it should be a multi-token rule.

© N. Harris Computer Corporation

Intra-token rules

The image below contains an intra-token rule that adds attribute information to the extraction result. The regular expression and the `<attribute>` element work together to enable the rule to enrich the result with metadata.



The regular expression in the rule matches attributes for "day", "month", and "year", and automatically adds the attributes to the extraction result:

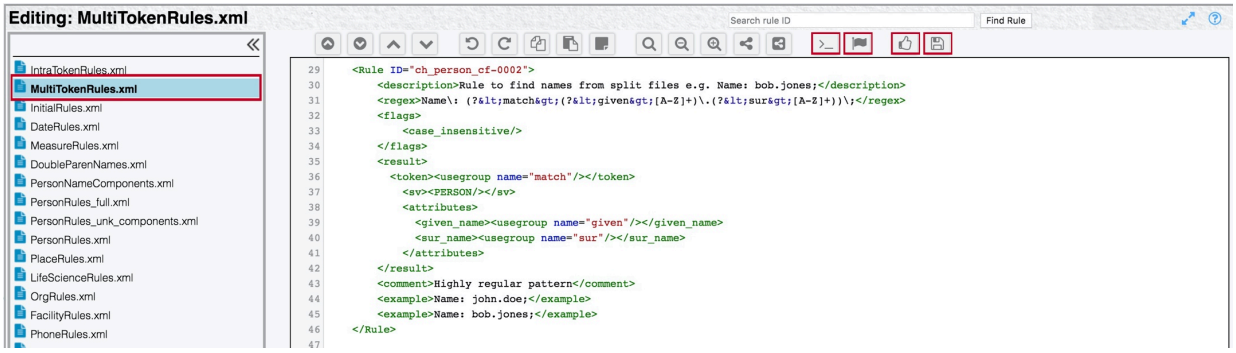
```
(?&lt;day&gt;([012]?[0-9])|(30|31))\.(?&lt;month&gt;(0?[1-9])|(10|11|12))\.
(?&lt;year&gt;(18|19|20)?\d{2})
```

The following attribute information is then completed within the rule:

```
<attributes>
  <day><usegroup name="day"/></day>
  <month><usegroup name="month"/></month>
  <year><usegroup name="year"/></year>
</attributes>
```

Multi-token rules

The image below presents an example of a multi-token rule, where there is a character type change, and given name and surname attributes are automatically added to the extraction result. The rule is listed after the image.



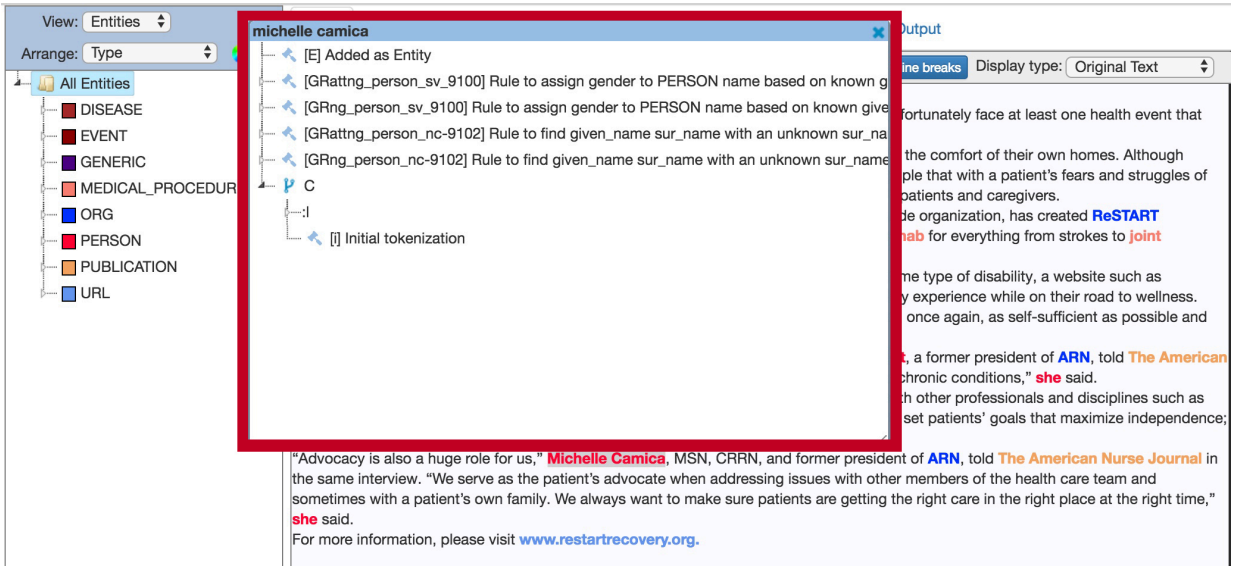
© N. Harris Computer Corporation

```
<Rule ID="ch_person_cf-0002">
  <description>Rule to find names from split files, such as Name:
 bob.jones;</description>
  <regex>Name\: (?&lt;match&gt;(?&lt;given&gt;[A-Z]+)\.(?&lt;sur&gt;[A-
Z]+))\;</regex>
  <flags>
    <case_insensitive/>
  </flags>
  <result>
    <token><usegroup name="match"/></token>
    <sv><PERSON/></sv>
    <attributes>
      <given_name><usegroup name="given"/></given_name>
      <sur_name><usegroup name="sur"/></sur_name>
    </attributes>
  </result>
  <comment>Highly regular pattern</comment>
  <example>Name: john.doe;</example>
  <example>Name: bob.jones;</example>
</Rule>
```

Tracing TextChart rules

In TextChart Studio, you can use the document view to see which rules were responsible for an entity being extracted. Highlight the entity, right-click, and select **Show rule match detail**.

TextChart Studio displays a window that lists the order in which the TextChart engine executed lexical lookups and rules. From here, double-clicking a rule opens the corresponding rule file in the rule editor.



The rule trace contains the sequence of processing steps that took place when a string or document was processed. The following list provides a summary of how to interpret the steps:

- `:` = Delimiter between rule actions
- `i` = Initial tokenization
- `LL` = Lexical Lookup

© N. Harris Computer Corporation

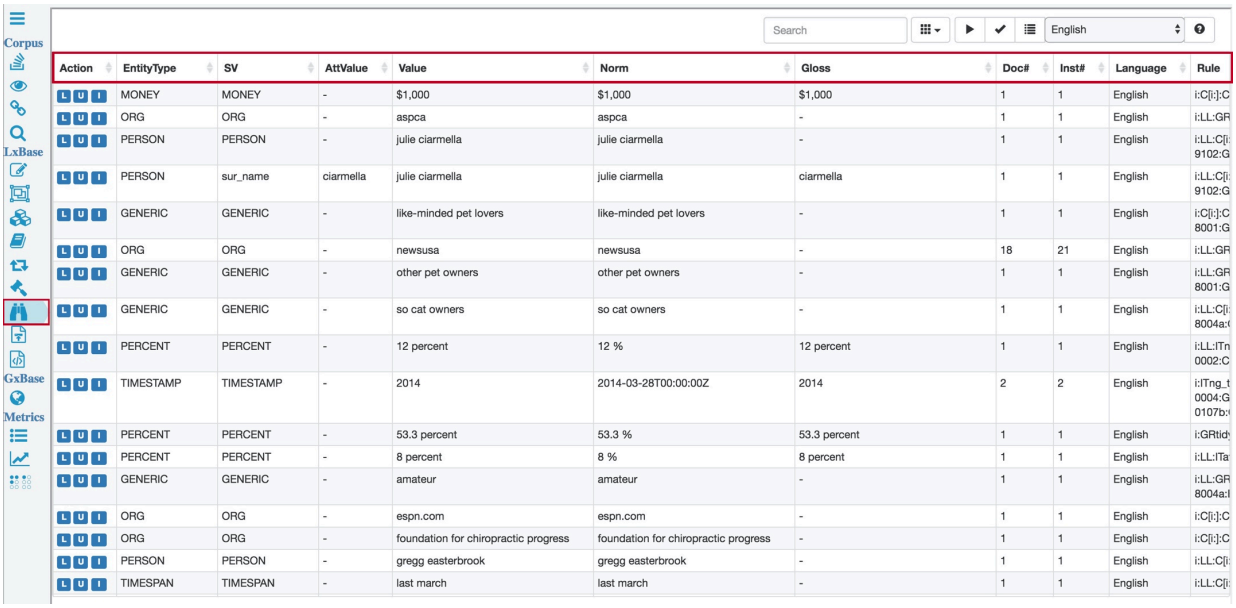
- `C` = The token was combined
- `[...]` = The contents of the brackets after the `C` includes the rule trace of the combined tokens
- `MT` = Multi-token rule
- `IT` = Intra-token rule
- `GR` = General rules
- `HC` = Hard-coded rule
- `BCE` = Back chaining entity detection
- `att` = Attribute portion of a rule assigning attribution to a token

Adding terms in bulk

TextChart Studio provides two ways of adding terms to the lexicon in bulk. Through machine discovery, you can let the TextChart engine alert you to likely new terms in your documents so that you can add them. Alternatively, if you have a list of terms that you know TextChart does not understand, you can import and classify those terms by hand.

Machine discovery

In addition to matching lexical items and patterns, TextChart uses *machine discovery* to extract other entities and lexical entries based on their linguistic context. Using their known semantic vectors, you can quickly add them to a lexicon as another meaningful semantic vector or entity, as well as unlearn or ignore them.



Action	EntityType	SV	AttValue	Value	Norm	Gloss	Doc#	Inst#	Language	Rule
L U I	MONEY	MONEY	-	$1,000	$1,000	$1,000	1	1	English	i:C[i:]:C
L U I	ORG	ORG	-	aspca	aspca	-	1	1	English	i:LL:GR
L U I	PERSON	PERSON	-	julie ciarmella	julie ciarmella	-	1	1	English	i:LL:C[i: 9102:G
L U I	PERSON	sur_name	ciarmella	julie ciarmella	julie ciarmella	ciarmella	1	1	English	i:LL:C[i: 9102:G
L U I	GENERIC	GENERIC	-	like-minded pet lovers	like-minded pet lovers	-	1	1	English	i:C[i:]:C 8001:G
L U I	ORG	ORG	-	newsusa	newsusa	-	18	21	English	i:LL:GR
L U I	GENERIC	GENERIC	-	other pet owners	other pet owners	-	1	1	English	i:LL:GR 8001:G
L U I	GENERIC	GENERIC	-	so cat owners	so cat owners	-	1	1	English	i:LL:C[i: 8004a:
L U I	PERCENT	PERCENT	-	12 percent	12 %	12 percent	1	1	English	i:LL:ITn 0002:C
L U I	TIMESTAMP	TIMESTAMP	-	2014	2014-03-28T00:00:00Z	2014	2	2	English	i:ITng_t 0004:G 0107b:
L U I	PERCENT	PERCENT	-	53.3 percent	53.3 %	53.3 percent	1	1	English	i:GRtidy
L U I	PERCENT	PERCENT	-	8 percent	8 %	8 percent	1	1	English	i:LL:ITa
L U I	GENERIC	GENERIC	-	amateur	amateur	-	1	1	English	i:LL:GR 8004a:I
L U I	ORG	ORG	-	espn.com	espn.com	-	1	1	English	i:C[i:]:C
L U I	ORG	ORG	-	foundation for chiropractic progress	foundation for chiropractic progress	-	1	1	English	i:C[i:]:C
L U I	PERSON	PERSON	-	gregg easterbrook	gregg easterbrook	-	1	1	English	i:LL:C[i:
L U I	TIMESPAN	TIMESPAN	-	last march	last march	-	1	1	English	i:LL:C[i:

After TextChart has processed a corpus, you can view the entities that it found through machine discovery by clicking **Perform discovery** in the LxBase section of the vertical toolbar.

© N. Harris Computer Corporation

Click **Discovery settings**



to select which entities or semantic vectors to learn.

Perform discovery by clicking **Start discovery**



.

You can choose to learn, unlearn, or ignore items from machine discovery by selecting the corresponding checkbox for each item, or by pressing the *L*, *U*, or *I* keys. You can also press *Space* to move to the next item without making a selection.

To add your selections to the lexicon, click **Commit actions**



.

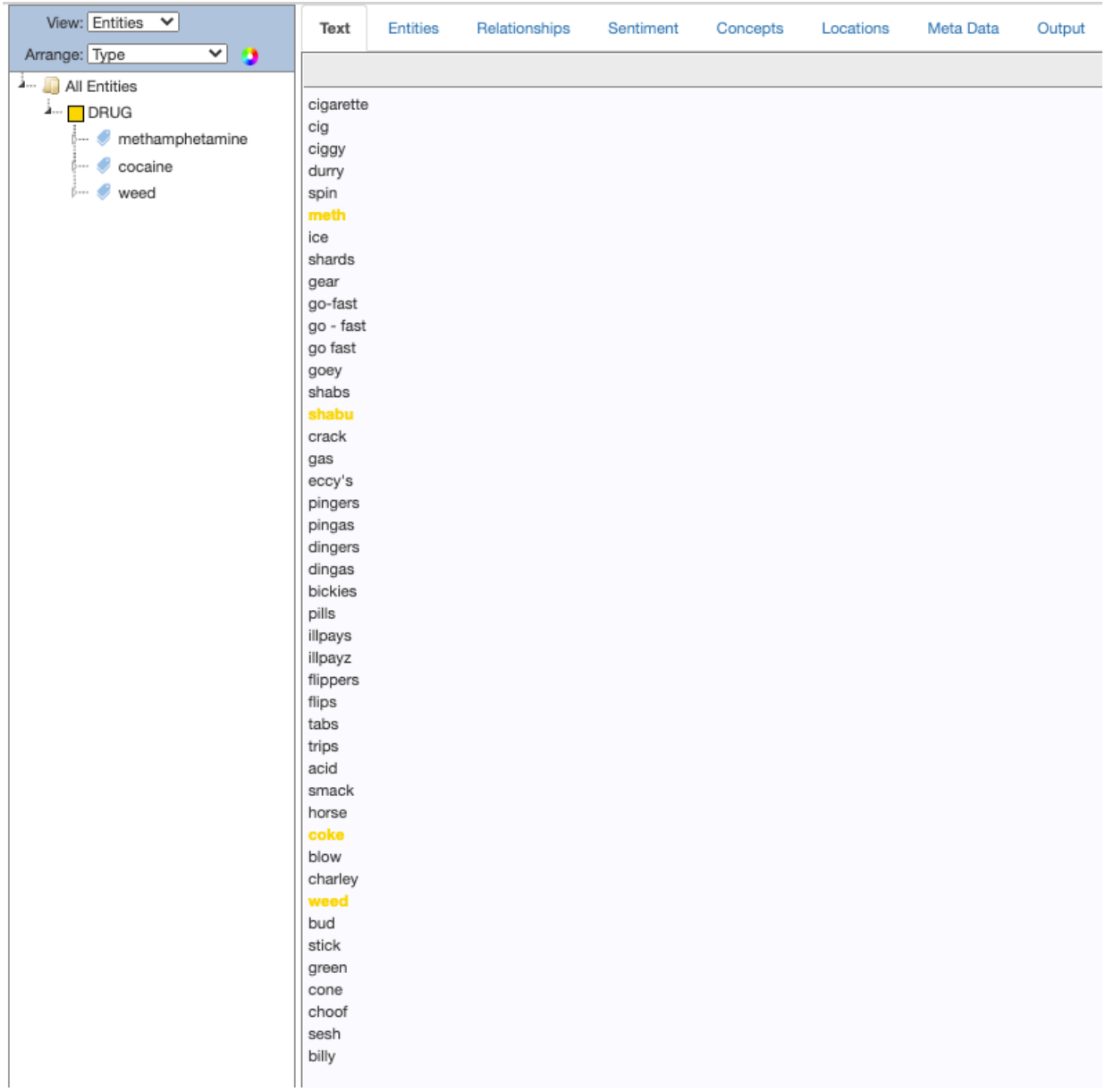
Note: Double-clicking an item in the **Value** or **Norm** column opens a window containing the surrounding context for each item. As well as the text, this context also includes the rule trace and the associated semantic vectors for each token.

Importing words

Another way of adding lexical items to LxBase is through TextChart Studio's word import tool, which you open by clicking **Import word list** in the vertical toolbar. The tool allows you to add a large list of lexical entries by appending them to an existing file, overwriting an existing file, or creating an additional file.

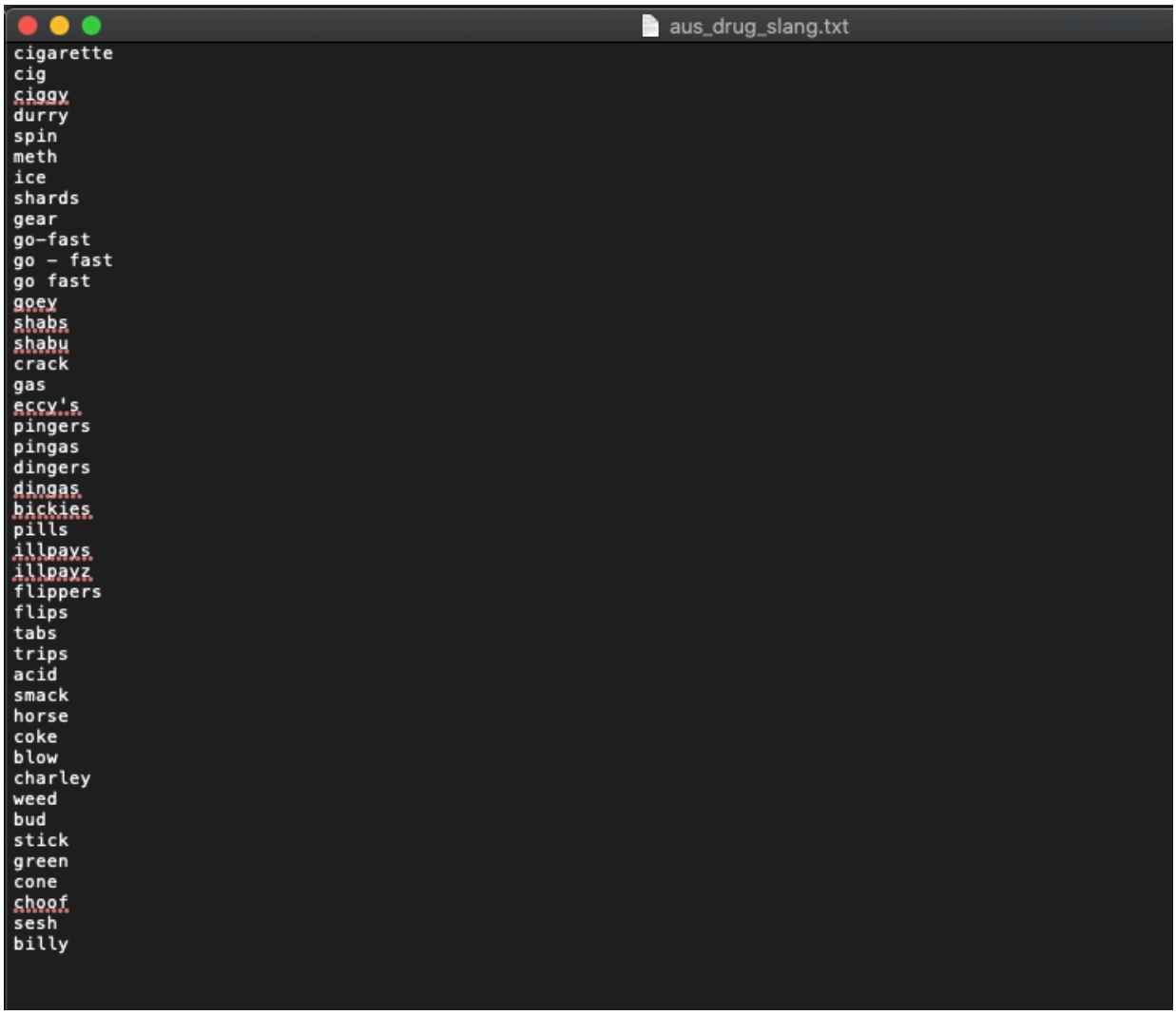
For example, the next image shows a list of drug slang terms that the TextChart engine has processed. Some of the results are highlighted in yellow, which represents an extracted DRUG entity result. The others do not have this "hit highlighting" and were therefore not extracted as an entity.

© N. Harris Computer Corporation



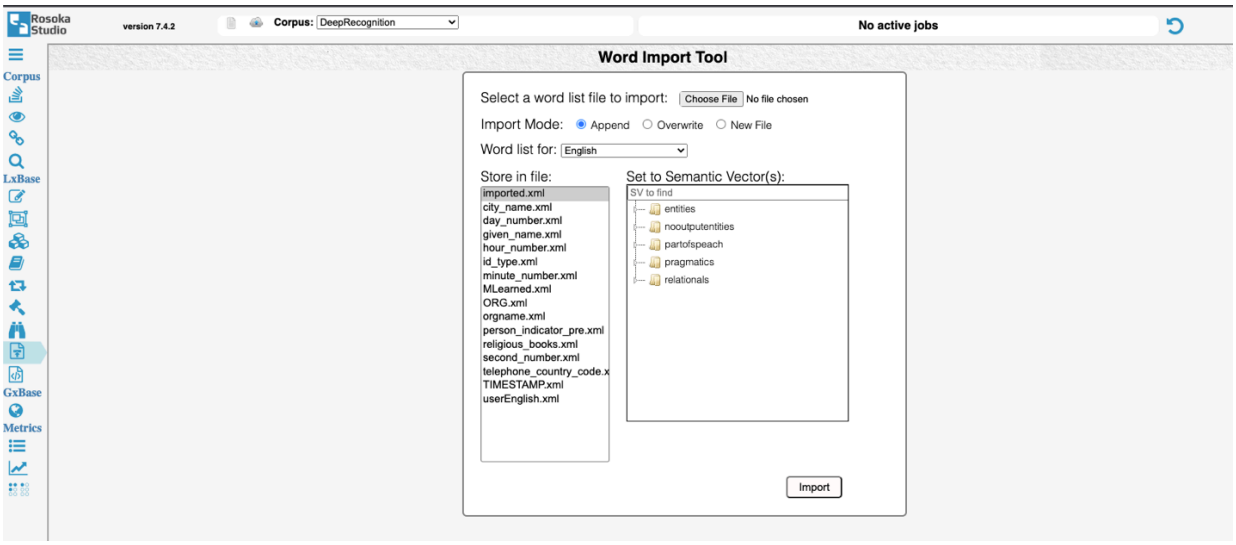
To add some or all of the words that were not extracted to the LxBase, you can create a text file that contains them. Entries in the text file must be formatted with a line break between each entry, as in the next image.

© N. Harris Computer Corporation



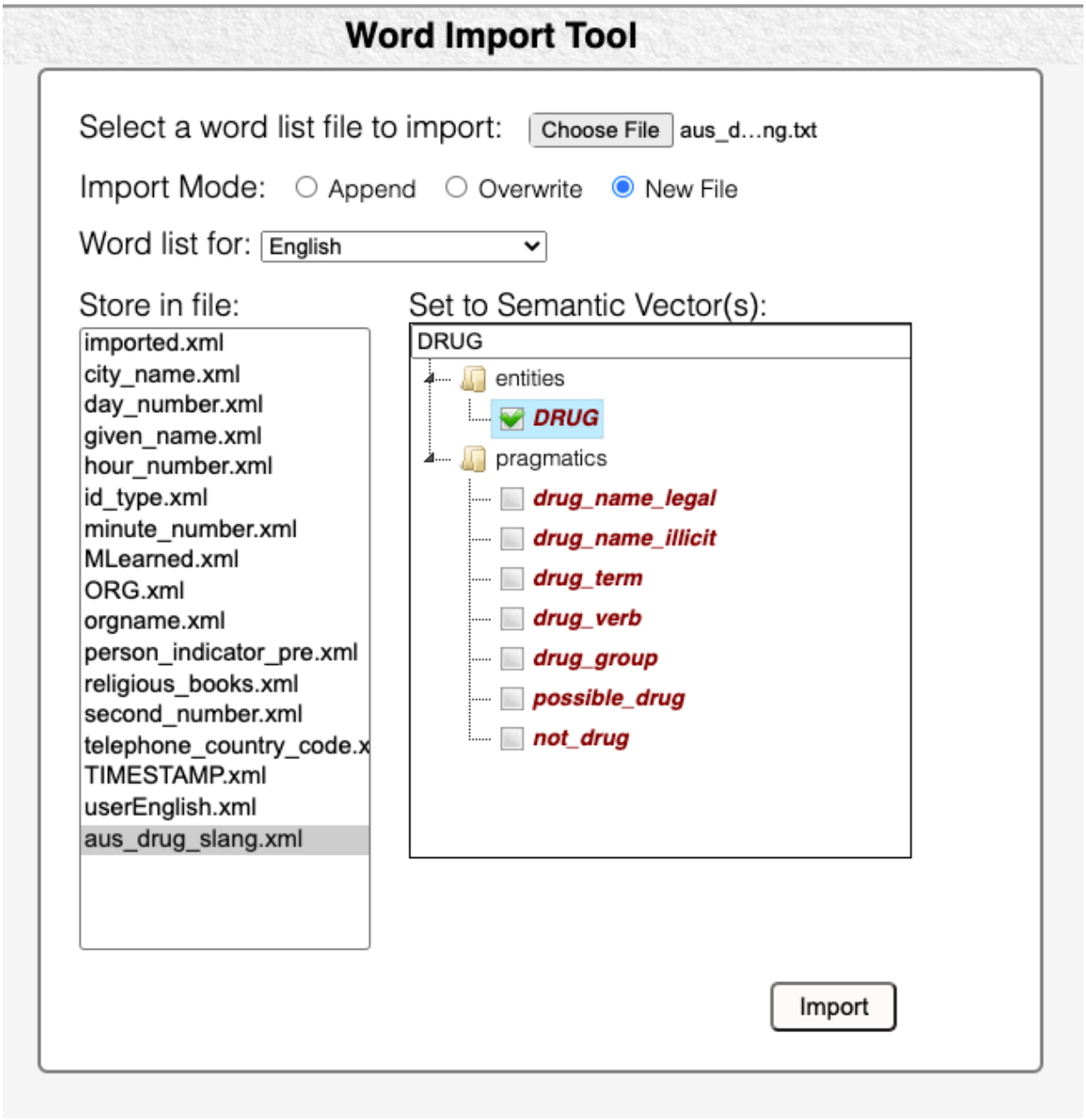
When you open the word import tool, you can click **Choose File** to locate the text file that you created, and then choose the import mode.

© N. Harris Computer Corporation



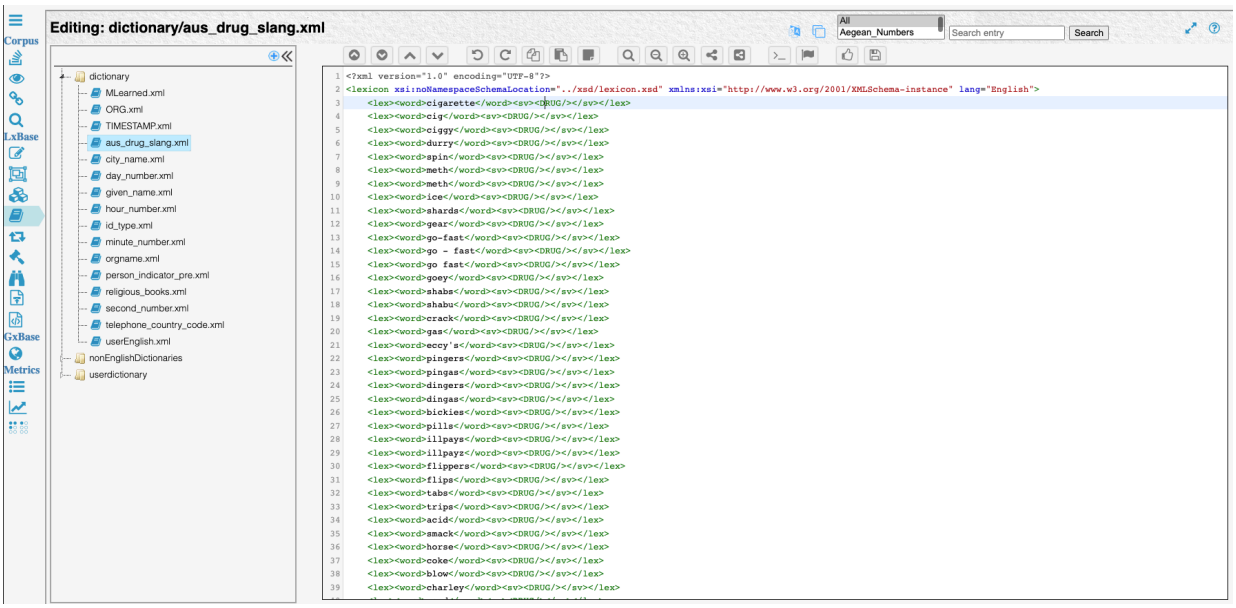
In the next image, the word import tool is configured to create a file for storing the new terms, and to associate each term with the DRUG entity type.

© N. Harris Computer Corporation



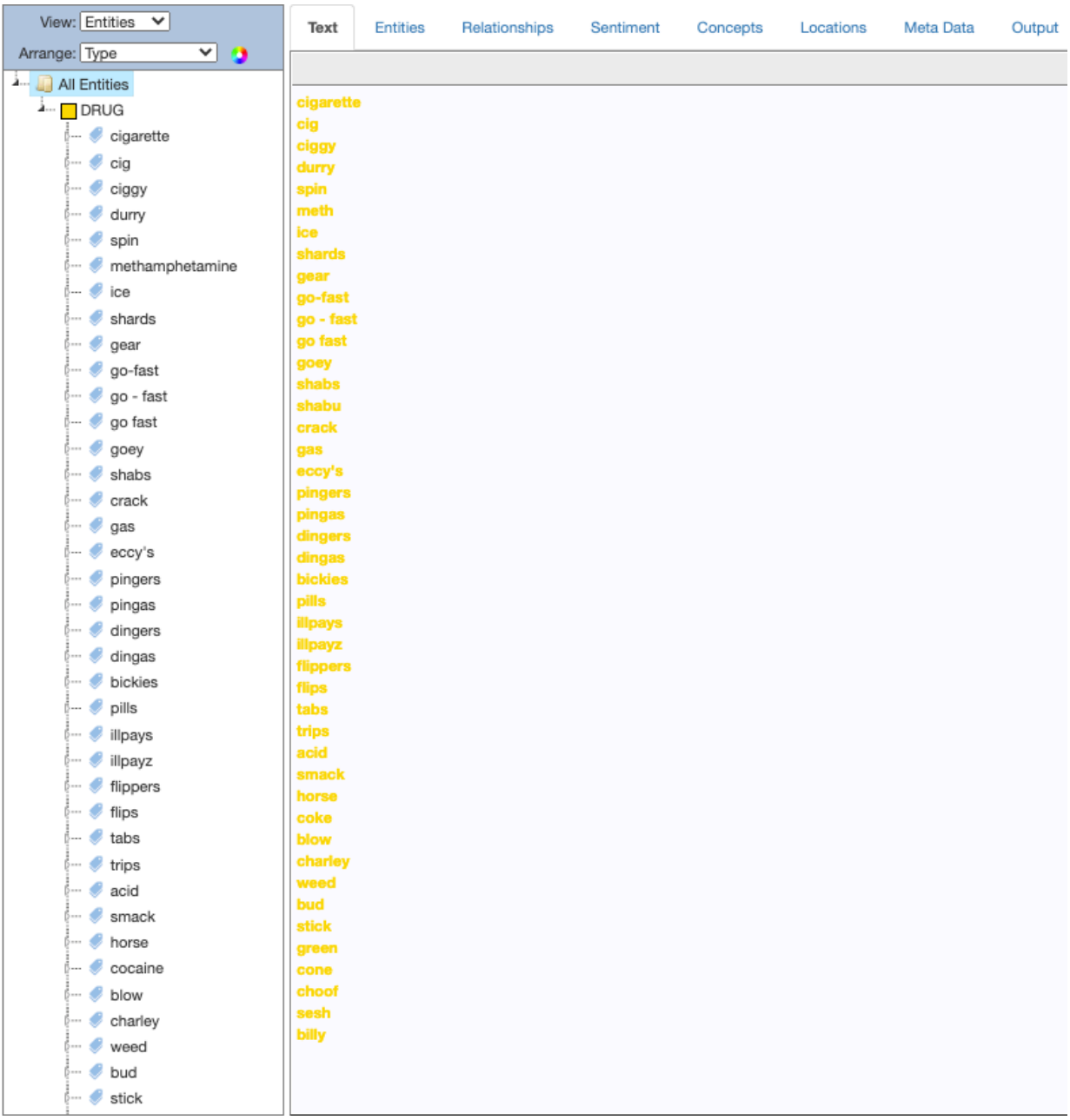
A few seconds after you click **Import**, the TextChart engine creates a dictionary file, complete with the appropriate XML. If you want to see the file, click the "Book" icon in the LxBase section of the vertical toolbar.

© N. Harris Computer Corporation



If you now reprocess the original source document inside TextChart Studio, all the terms are displayed in yellow, indicating that they are being extracted as entities of type DRUG.

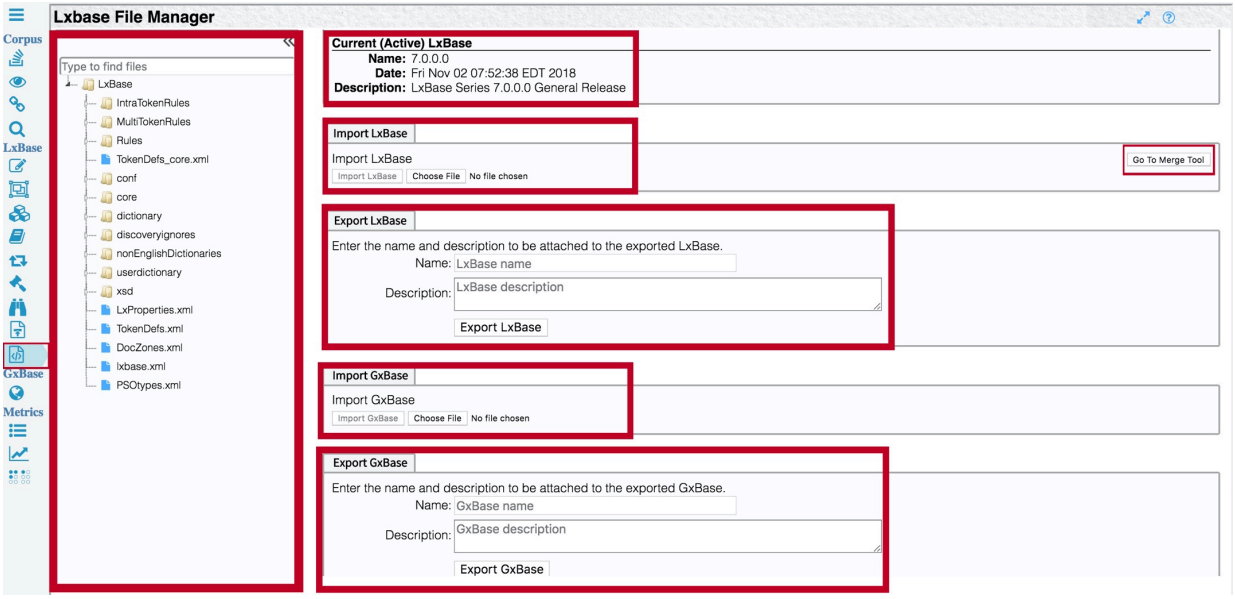
© N. Harris Computer Corporation



Importing and exporting LxBases

TextChart Studio provides a tool for managing LxBase files, and for importing, exporting, and merging LxBases. To open the **LxBase File Manager** page, click **Manage LxBase Files** in the LxBase section of the vertical toolbar.

© N. Harris Computer Corporation



The tree view on the left of the page contains a searchable list of LxBase files, while the right side provides information about the current LxBase and a series of actions that you can perform on it.

Import LxBase

To import a new LxBase, click **Choose File** in the **Import LxBase** panel, locate the ZIP file containing the LxBase you want to import, and then click **Import LxBase**.

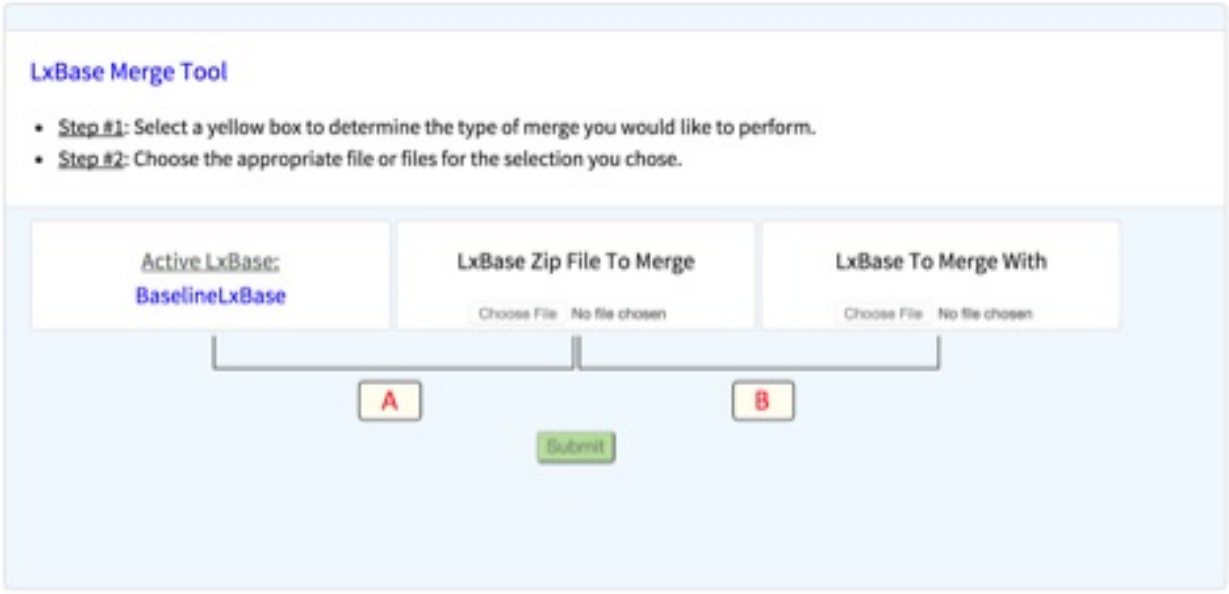
Export LxBase

To export an LxBase, provide a name and a description in the **Export LxBase** panel, and then click **Export LxBase** to download a ZIP file containing the LxBase along with necessary metadata.

Merge LxBases

To consolidate the work of multiple users, or to incorporate an update to the default LxBase into a customized one, TextChart Studio can merge two LxBases together. To begin, click **Go To Merge Tool** in the **Import LxBase** panel.

© N. Harris Computer Corporation



Here, you can either merge a single LxBase ZIP file into the active LxBase (**A**), or merge two LxBase ZIP files together (**B**).

Important: To restore an existing LxBase after completing a merge, you must save a regression checkpoint beforehand.

After you select one or two LxBase ZIP files, TextChart Studio displays a table of their contents, organized by category. For each category, the table indicates how many files need merging, how many are unique to each LxBase, and how many are identical. To start merging, click in the leftmost column in the first row.

© N. Harris Computer Corporation



	Total	Pending	UniqueA	UniqueB	Identical	Resolved
Conf Files	5	3	0	1	1	0
MultiTokenRules Files	1	1	0	0	0	0
IntraTokenRules Files	1	1	0	0	0	0
Rules Files	21	17	0	3	1	0
Dictionary Files (English)	18	1	5	12	0	0
Non-English Dictionary Files	217	3	0	0	214	0
User Dictionary Files	1	1	0	0	0	0
Discovery Ignores Files	1	0	0	0	1	0

- **Pending**

  Pending files are not identical between the two LxBases, and therefore you need to resolve their differences.

  Click **Edit** to open a dialog that allows you to view the differences between the files. You can decide whether the merged LxBase should use the contents of one of the LxBases in the merge, or you can edit to create a hybrid. Click **Back** to return to the table without saving changes, or **Submit** to validate your changes and save them if validation was successful.

- **Unique**

  Unique files appear in one of the LxBases being merged, but not the other. To include a unique file in the result, click **Include**; to exclude it, click **Exclude**. Either decision resolves the merge process for that file.

- **Identical**

  Identical files appear in both of the LxBases being merged, and have the same name and the same contents. You don't need to do anything to identical files, apart from clicking **Resolve** to acknowledge them.

  To resolve all identical files at once, click **Resolve All Identicals**.

To change which LxBase appears in which pane on the screen, click **Reset Pane**. To abandon all current merges and delete the selected ZIP files, click **Reset Merge**.

© N. Harris Computer Corporation

When you've resolved all the files in one category, you can move on to the next. When you've resolved all the files in all the categories, click **Final Validation** to make sure that your completed file is valid. Provided that it is, you can then complete the merge operation.

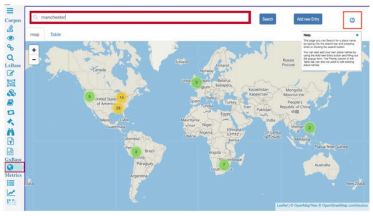
GxBase

TextChart uses GxBase to retrieve names and geocoordinates for the places that it finds in documents in your corpora. The standard GxBase includes the National Geospatial Agency's (NGA) GNS, the United States Geological Survey's (USGS) GNIS, and i2's own internally developed gazetteer to provide worldwide coverage.

In TextChart Studio, you can search a GxBase, customize it, and add features that are specific to your domain. To begin, click **GxBase** in the **GxBase** section of the vertical toolbar..

Searching for places

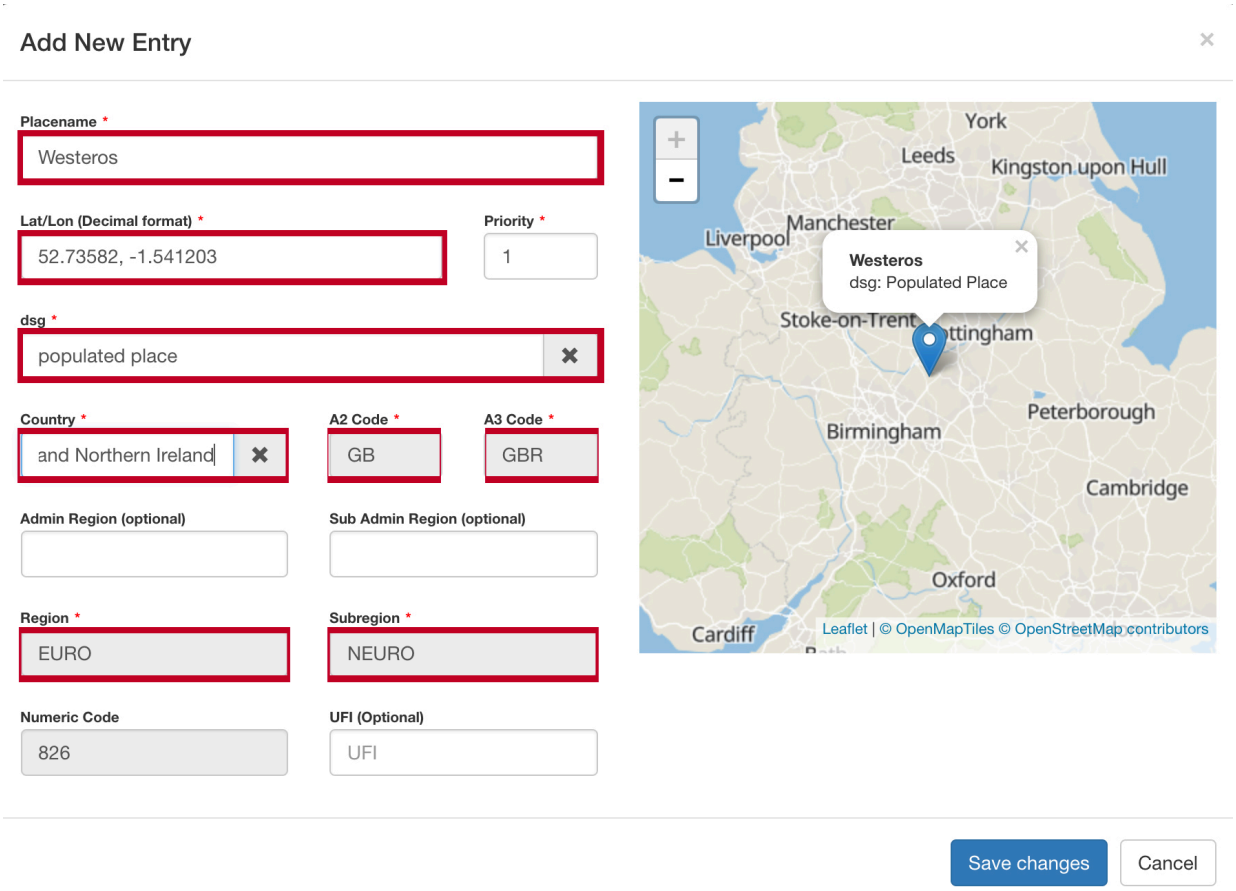
To search for a place in GxBase, type its name into the search bar and press *Enter*. All instances of the name that exist in the gazetteer are shown on the map, and you can zoom in and out as you need.



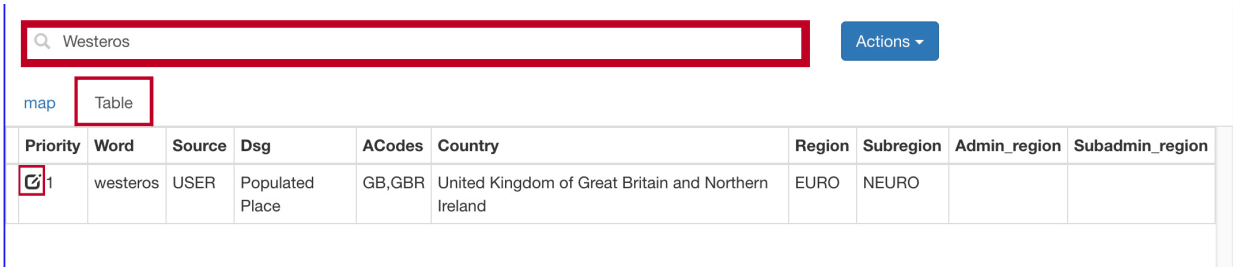
Adding a place name

To add a place to GxBase, type its name ito the search bar and click **Add new Entry** to display the **And New Entry** dialog. Complete the required information, and then click **Save changes**.

© N. Harris Computer Corporation



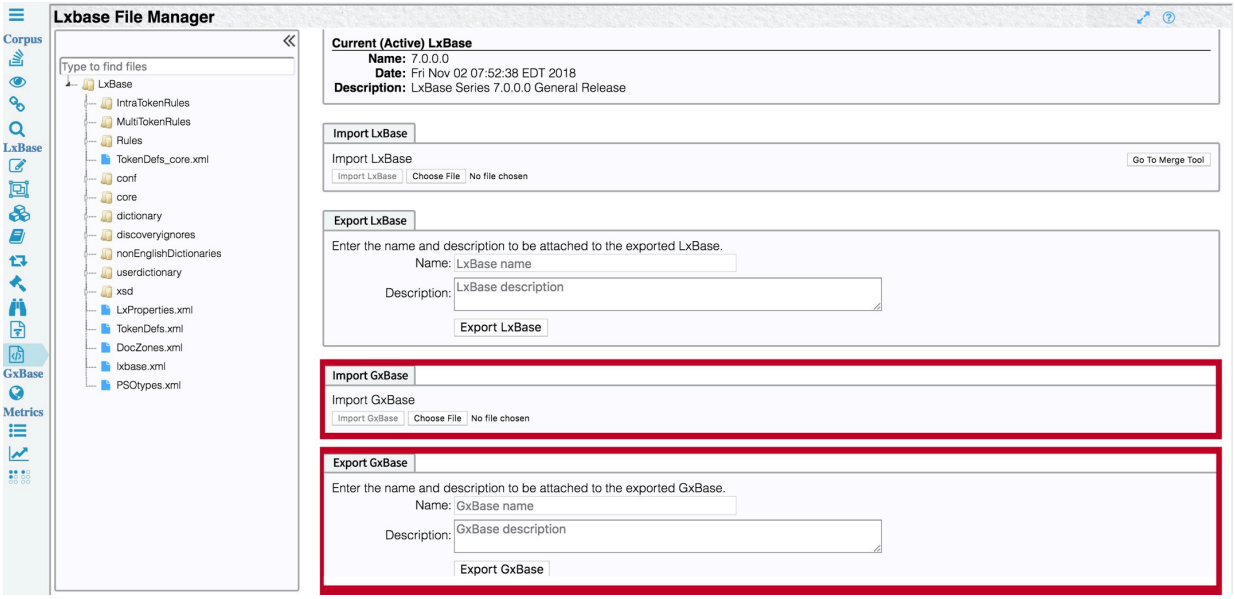
After you add the new place, you can open the **Table** tab to search for and view the entry for it in GxBase. To make adjustments, click the button in the **Priority** column.



Importing and exporting GxBases

You can import and export customized GxBases through the same interface that you use to import and export LxBases. Open the **LxBase File Manager** page by clicking **Manage LxBase Files** in the LxBase section of the vertical toolbar, and use the **Import GxBase** and **Export GxBase** panels according to your needs.

© N. Harris Computer Corporation



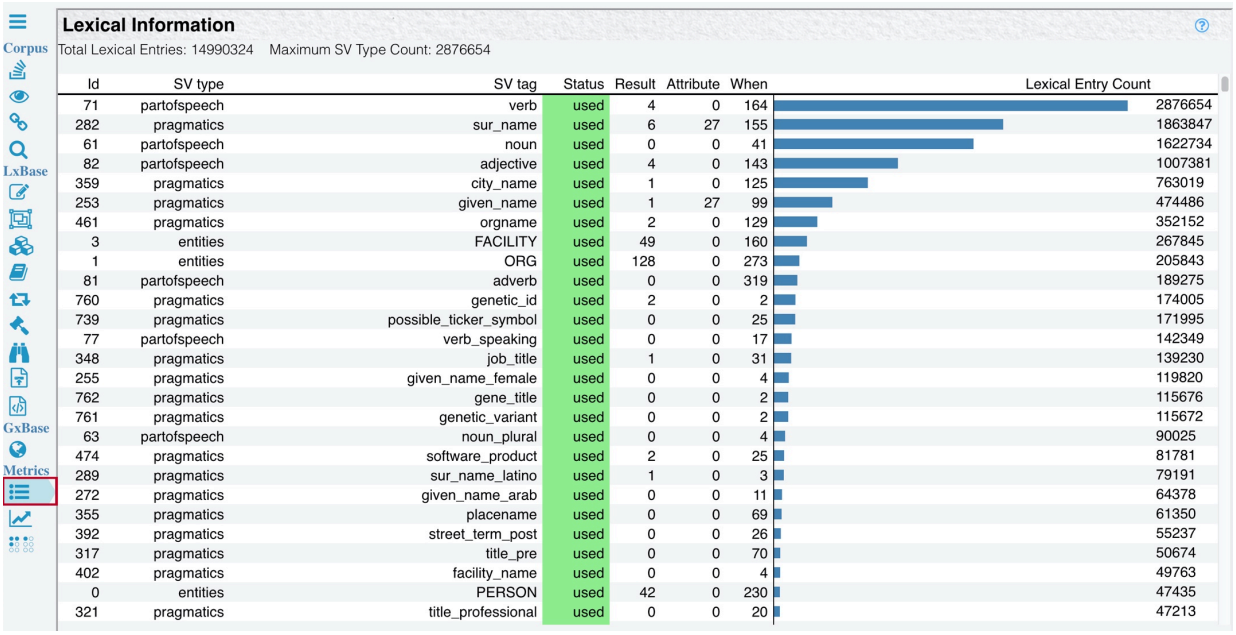
Metrics

TextChart Studio provides metrics about the active LxBase that you can review and act upon to improve the efficiency of the system. The items in the **Metrics** menu reveal information about lexical entries (including semantic vectors) and LxBase rules.

Lexical information

Click **Analyze lexicon** in the **Metrics** section of the vertical toolbar to open the **Lexical information** page. TextChart Studio displays information including the total number of lexical entries in all dictionaries, and the counts of the semantic vector types that occur most frequently.

© N. Harris Computer Corporation



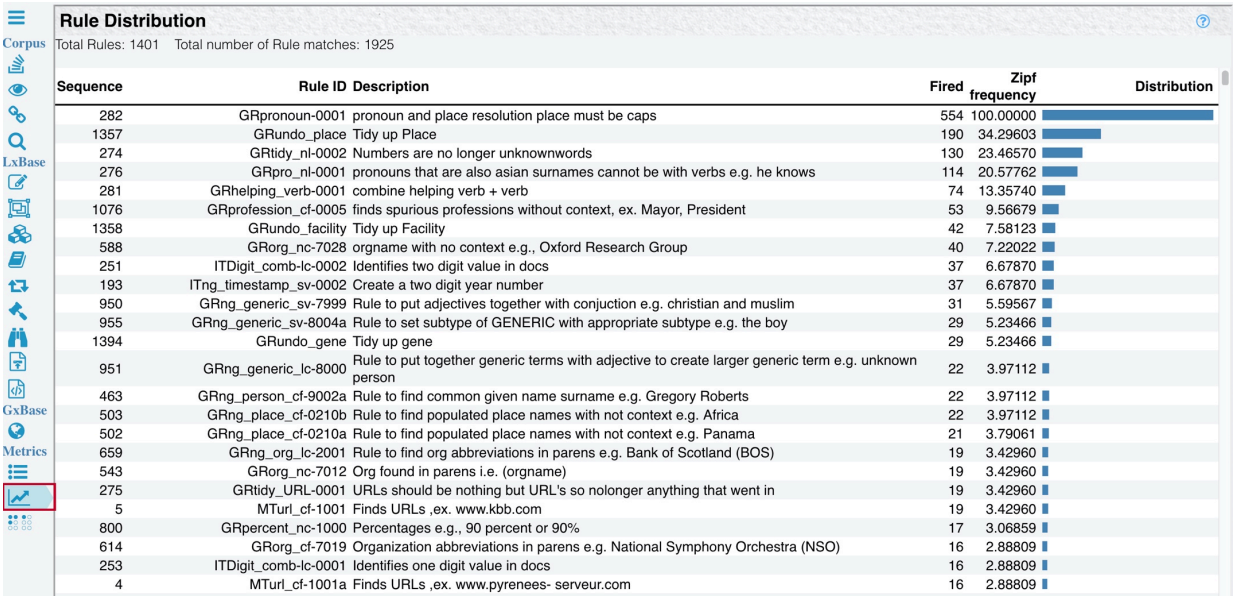
Id	SV type	SV tag	Status	Result	Attribute	When	Lexical Entry Count
71	partofspeech	verb	used	4	0	164	2876654
282	pragmatics	sur_name	used	6	27	155	1863847
61	partofspeech	noun	used	0	0	41	1622734
82	partofspeech	adjective	used	4	0	143	1007381
359	pragmatics	city_name	used	1	0	125	763019
253	pragmatics	given_name	used	1	27	99	474486
461	pragmatics	orgname	used	2	0	129	352152
3	entities	FACILITY	used	49	0	160	267845
1	entities	ORG	used	128	0	273	205843
81	partofspeech	adverb	used	0	0	319	189275
760	pragmatics	genetic_id	used	2	0	2	174005
739	pragmatics	possible_ticker_symbol	used	0	0	25	171995
77	partofspeech	verb_speaking	used	0	0	17	142349
348	pragmatics	job_title	used	1	0	31	139230
255	pragmatics	given_name_female	used	0	0	4	119820
762	pragmatics	gene_title	used	0	0	2	115676
761	pragmatics	genetic_variant	used	0	0	2	115672
63	partofspeech	noun_plural	used	0	0	4	90025
474	pragmatics	software_product	used	2	0	25	81781
289	pragmatics	sur_name_latino	used	1	0	3	79191
272	pragmatics	given_name_arab	used	0	0	11	64378
355	pragmatics	placename	used	0	0	69	61350
392	pragmatics	street_term_post	used	0	0	26	55237
317	pragmatics	title_pre	used	0	0	70	50674
402	pragmatics	facility_name	used	0	0	4	49763
0	entities	PERSON	used	42	0	230	47435
321	pragmatics	title_professional	used	0	0	20	47213

The page also displays a chart that lists all entities and semantic vectors, as well as associated usage information. You can find the the number of times a semantic vector occurs in different parts of a rule (results clause, attribute clause, and when clause), and the number of lexical entries that contain that semantic vector.

Rule distribution

Clicking **View rule distribution** in the **Metrics** section of the vertical toolbar to open the **Rule Distribution** page. TextChart Studio displays information about the frequency and distribution of rule matches.

© N. Harris Computer Corporation



Sequence	Rule ID	Description	Fired	Zipf frequency	Distribution
282	GRpronoun-0001	pronoun and place resolution place must be caps	554	100.00000	
1357	GRundo_place	Tidy up Place	190	34.29603	
274	GRtidy_nl-0002	Numbers are no longer unknownwords	130	23.46570	
276	GRpro_nl-0001	pronouns that are also asian surnames cannot be with verbs e.g. he knows	114	20.57762	
281	GRhelping_verb-0001	combine helping verb + verb	74	13.35740	
1076	GRprofession_cf-0005	finds spurious professions without context, ex. Mayor, President	53	9.56679	
1358	GRundo_facility	Tidy up Facility	42	7.58123	
588	GRorg_nc-7028	orgname with no context e.g., Oxford Research Group	40	7.22022	
251	ITDigit_comb-lc-0002	Identifies two digit value in docs	37	6.67870	
193	ITng_timestamp_sv-0002	Create a two digit year number	37	6.67870	
950	GRng_generic_sv-7999	Rule to put adjectives together with conjuction e.g. christian and muslim	31	5.59567	
955	GRng_generic_sv-8004a	Rule to set subtype of GENERIC with appropriate subtype e.g. the boy	29	5.23466	
1394	GRundo_gene	Tidy up gene	29	5.23466	
951	GRng_generic_lc-8000	Rule to put together generic terms with adjective to create larger generic term e.g. unknown person	22	3.97112	
463	GRng_person_cf-9002a	Rule to find common given name surname e.g. Gregory Roberts	22	3.97112	
503	GRng_place_cf-0210b	Rule to find populated place names with not context e.g. Africa	22	3.97112	
502	GRng_place_cf-0210a	Rule to find populated place names with not context e.g. Panama	21	3.79061	
659	GRng_org_lc-2001	Rule to find org abbreviations in parens e.g. Bank of Scotland (BOS)	19	3.42960	
543	GRorg_nc-7012	Org found in parens i.e. (orgname)	19	3.42960	
275	GRtidy_URL-0001	URLs should be nothing but URL's so nolonger anything that went in	19	3.42960	
5	MTurl_cf-1001	Finds URLs ,ex. www.kbb.com	19	3.42960	
800	GRpercent_nc-1000	Percentages e.g., 90 percent or 90%	17	3.06859	
614	GRorg_cf-7019	Organization abbreviations in parens e.g. National Symphony Orchestra (NSO)	16	2.88809	
253	ITDigit_comb-lc-0001	Identifies one digit value in docs	16	2.88809	
4	MTurl_cf-1001a	Finds URLs ,ex. www.pyrenees- serveur.com	16	2.88809	

The chart on this page lists all rules in the LxBase, the number of times the rule matched, and the Zipf frequency of the rule.

Regression testing

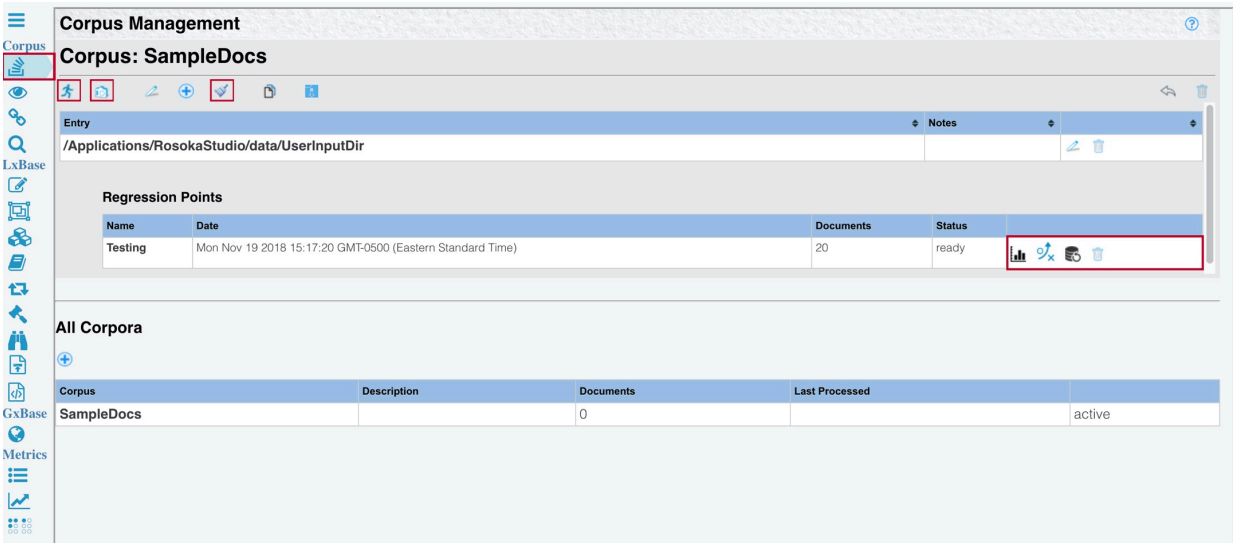
TextChart Studio has a built-in regression testing feature for measuring the impact that changes to the LxBase have on the output. When a corpus is active, all saved regression points are listed in a chart below the corpus entry in the **Corpus Management** page.

To create a new regression point, click **Create New Regression Point**



in the horizontal toolbar above the corpus entry.

© N. Harris Computer Corporation



After you make lexical or rule modifications to the LxBase, you can score a current run against a regression point. First, click **Clear Processing Results** to remove any previous results from the comparison. Then you need to reprocess the corpus.

To compare the results of the current run with a regression point, click **Score against current results**



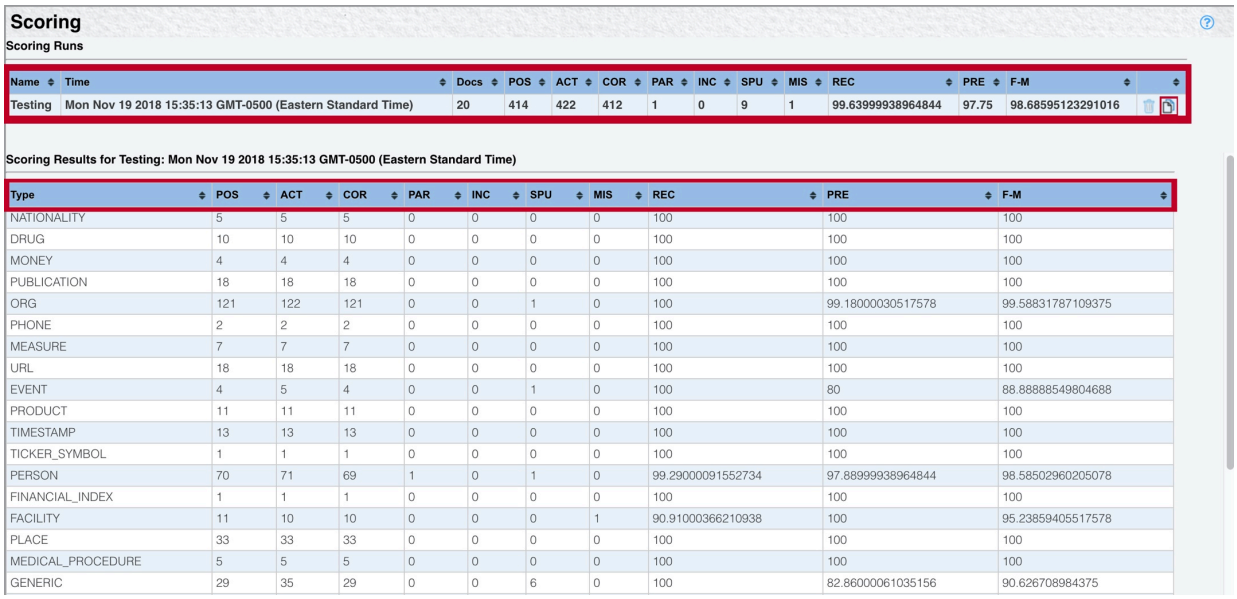
in the appropriate row of the corpus list, or **Compare and score results** in the **Metrics** section of the vertical toolbar. You can save multiple regression points and test against any of them.

TextChart Studio retains regression points and scoring results until you delete them. To view scoring results without prompting a new run, click **View Score Results** in the appropriate row. You can also revert the active LxBase to an earlier version by clicking **Restore saved LxBase**.

Interpreting results

As you view the scoring results of a regression run, you can click the table at the top of the **Scoring** page to see a breakdown for each entity type.

© N. Harris Computer Corporation



Scoring

Scoring Runs

Name	Time	Docs	POS	ACT	COR	PAR	INC	SPU	MIS	REC	PRE	F-M	
Testing	Mon Nov 19 2018 15:35:13 GMT-0500 (Eastern Standard Time)	20	414	422	412	1	0	9	1	99.63999938964844	97.75	98.68595123291016	

Scoring Results for Testing: Mon Nov 19 2018 15:35:13 GMT-0500 (Eastern Standard Time)

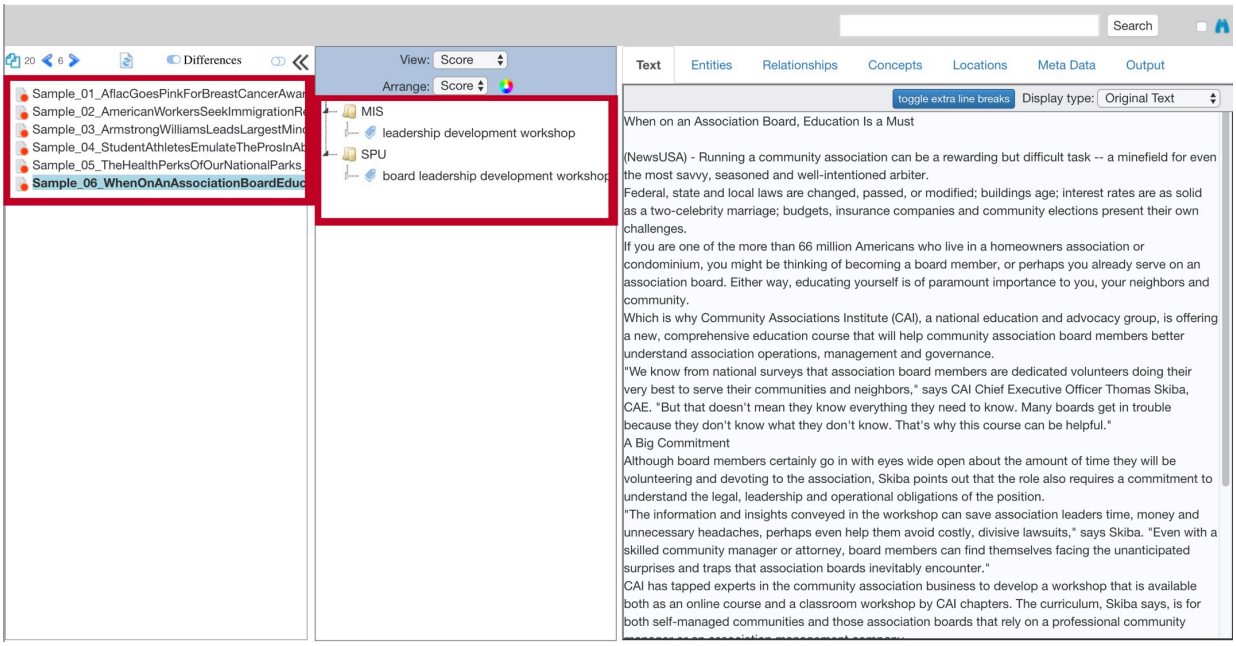
Type	POS	ACT	COR	PAR	INC	SPU	MIS	REC	PRE	F-M
NATIONALITY	5	5	5	0	0	0	0	100	100	100
DRUG	10	10	10	0	0	0	0	100	100	100
MONEY	4	4	4	0	0	0	0	100	100	100
PUBLICATION	18	18	18	0	0	0	0	100	100	100
ORG	121	122	121	0	0	1	0	100	99.18000030517578	99.58831787109375
PHONE	2	2	2	0	0	0	0	100	100	100
MEASURE	7	7	7	0	0	0	0	100	100	100
URL	18	18	18	0	0	0	0	100	100	100
EVENT	4	5	4	0	0	1	0	100	80	88.88888549804688
PRODUCT	11	11	11	0	0	0	0	100	100	100
TIMESTAMP	13	13	13	0	0	0	0	100	100	100
TICKER_SYMBOL	1	1	1	0	0	0	0	100	100	100
PERSON	70	71	69	1	0	1	0	99.29000091552734	97.88999938964844	98.58502960205078
FINANCIAL_INDEX	1	1	1	0	0	0	0	100	100	100
FACILITY	11	10	10	0	0	0	1	90.91000366210938	100	95.23859405517578
PLACE	33	33	33	0	0	0	0	100	100	100
MEDICAL_PROCEDURE	5	5	5	0	0	0	0	100	100	100
GENERIC	29	35	29	0	0	6	0	100	82.86000061035156	90.626708984375

The scoring run produces the following results:

- POS (possible) - The number of possible entities (based on key)
- ACT (actual) - The number of entities extracted during the current run.
- COR (correct) - The number of correct entities (matching the key)
- PAR (partial) - The number of partial entities, where a portion of the string overlaps with an entity from the key.
- INC (incorrect) - The number of incorrect entities, where the string overlaps exactly with an entity in the key, but the entity type is different in the current run.
- SPU (spurious) - The number of new entities not in the key.
- MIS (missing) - The number of missing entities, where the entity *is* in the key, but not in the current run.
- REC (recall) - The proportion of actual entities (from the key) that are extracted as entities (in the current run).
- PRE (precision) - The proportion of postulated entities (from the current run) that are actual entities (from the key).
- F-M (F-measure) - A weighted average of precision and recall, between 0 and 1, with 1 being a perfect score. The formula uses # = 1, meaning that precision and recall are weighted equally.

As an alternative to the breakdown table, you can click **View Detailed Score Results in Documents** in the top table to view a list of the documents that have extraction differences.

© N. Harris Computer Corporation



If there are no differences, the view is empty. However, a document with a red dot next to its name means that TextChart Studio has detected an extraction change.

Click each document to review the extraction changes. Click the **Differences** button above the list of documents to move between the original extraction results and the new ones. This view displays scoring changes in the center column.

Glossary

The definitions in this list of terms that appear in i2 TextChart Studio apply only to it (and to other TextChart software). Some of the terms have different meanings in other i2 software.

Corpus

A *corpus* is a set of documents. TextChart Studio allows a user to save the file paths to multiple corpora on the **Corpus Management** page.

Entity

Entities are the important items such as persons, places, and events that TextChart finds within a document. The linguistic context of the document determines what words or phrases are extracted as entities.

Users have the ability to modify entity extraction results and to apply their own real-world knowledge. The LxBase documentation includes a list of all the types of entities that TextChart can extract.

GxBase

TextChart uses *GxBase* to retrieve the place names and geocoordinates that it finds in a set of documents. It uses the National Geospatial Agency's GNS and the United States Geological Survey's

© N. Harris Computer Corporation

GNIS, as well as i2's own, internally developed gazetteer, to provide world-wide coverage and use linguistic context to decipher between ambiguous location names.

Through TextChart Studio, users can customize and import their own client-specific features.

LxBase

The *LxBase* is the set of underlying linguistic rules and dictionaries that TextChart uses as the foundation for entity and relationship extraction.

TextChart Studio allows users to modify and add to the LxBase in order to fine tune for industry-specific extraction goals.

Normalized form

Whether one term is equivalent to another is sometimes a choice for an individual user. For example, it might or might not be appropriate for the terms "Britain" and "England" to be considered equivalent to "United Kingdom".

When they do judge terms to be equivalent, users can arrange for TextChart to extract entities in the same, *normalized form*.

The standard TextChart dictionaries already contain many normalized lexical entries. Through TextChart Studio, users can modify and add normalized forms of their own. See Normalization management for more information.

Relationship (or PSO)

A predicate-subject-object statement, or *PSO*, is a *relationship* between two entities established by the linguistic context.

Relationships have names of the form EntityToEntity. For example, PersonToPerson is the name of a relationship between one person entity and another.

In TextChart, every relationship has a *predicate* that describes the nature of that relationship. For example, a PersonToPerson relationship might have the predicate "interviewed". For a full list of predicate types, see the LxBase documentation.

Semantic vector (or SV)

Semantic vectors represent a vector space of possible meanings for individual terms or phrases, allowing the same term to have various meanings depending on the linguistic context.

For example, a term such as "Washington" has many different semantic vectors associated with it: it might be a city name, a given name, or a surname. Some linguistic rules might even determine that it is a place.

In TextChart Studio, you can find (and modify) a list of semantic vectors and their corresponding definitions through the Token Definition Editor.

Token

A *token* is the smallest unit of meaning that TextChart can extract from a document.

For example, a TextChart dictionary might include the term "bank" as a noun, while the term "Bank of America" is an ORG. TextChart then considers both terms as one token, because they are both listed in the dictionary.

© N. Harris Computer Corporation

If "Bank of America" was not listed in the dictionary, then each unit of meaning would be parsed individually, resulting in three unique tokens: "Bank", "of", and "America".

You can modify this in TextChart Studio.

LxBase

This documentation describes the entities and relationships that i2 TextChart software can extract from text documents.

Entities

Entities are the named items that i2 TextChart can find in a document. The context of the document is the only vehicle for determining any associated knowledge. Real-world knowledge (or externally encoded knowledge) does not determine associated knowledge with these entities.

The term *identity* is the collection of entities in a document that describes or references a particular entity in the real world.

ADDRESS

The ADDRESS entity type represents postal addresses such as "*123 Main St., Seattle, WA 98101*".

Attributes

- norm
- internal_location
- norm
- building_name
- street_name
- house_number
- city
- state_county_level
- province
- postal_code
- country
- suite_number
- assignee_name
- cntry_code
- grid_reference
- mgrs_coordinate
- utm_coordinate
- datum_code
- ups_coordinate
- bng_coordinate
- dms_coordinate

© N. Harris Computer Corporation

- decimal_degree
- easting
- northing
- hemisphere
- utm_zone
- altitude
- owner
- phone_number
- fax_number
- po_box
- source_description
- source_grade
- street_direction
- street_type
- monetary_amount
- abstract_flag
- abstract_binary
- purchase_date
- built
- surface_composition
- target_details
- subtype
  - physical_address
  - po_box
  - internal_location
- [[customer_specific]]

CONVEYANCE

The CONVEYANCE entity type represents vehicles or other means of conveyance. This entity type also refers to generic as well as specific instances, including both "*several aircraft*" and "*Volkswagen Jetta*".

Attributes

- norm
- year
- idnum
- type
- make
- model
- manufacturer
- designator

© N. Harris Computer Corporation

- vin
- color
- owner
- number_of_openings
- number_of_seats
- registered_state
- registered_country
- state_county_level
- suspicious
- assignee_name
- collateral
- filing_office
- filing_type
- filing_status
- address
- grid_reference
- mgrs_coordinate
- utm_coordinate
- datum_code
- ups_coordinate
- bng_coordinate
- dms_coordinate
- decimal_degree
- easting
- northing
- hemisphere
- utm_zone
- information_grade
- keyword
- keyword_type
- altitude
- prev_address
- serial_number
- source_description
- source_grade
- trademark
- body_style
- hull_number
- tail_number

© N. Harris Computer Corporation

- weight
- monetary_amount
- quantity
- timestamp
- abstract_flag
- abstract_binary
- purchase_date
- size
- callsign
- capabilities
- condition
- freq
- freq_range
- fuel_type
- isr_platform
- license_plate_num
- manufacture_date
- operational_status
- order_of_battle
- specialty_equipment
- target_details
- subtype
  - air
  - land
  - pipeline
  - space
  - vessel
- [[customer_specific]]

CRIME

The CRIME entity type represents criminal offenses, such as "*armed robbery*" or "*DUI*".

Attributes

- norm
- subtype
- date_of_crime
- start_date
- end_date
- lat (latitude in decimal degrees)
- lon (longitude in decimal degrees)

© N. Harris Computer Corporation

- coordinate
- location
- weapon_used
- crime_code
- verdict
- charging_organization
- defendant_name
- plaintiff_name
- crime_description
- classification_description
- classification_title
- information_grade
- keyword
- keyword_type
- source_description
- source_grade
- duration
- timestamp
- event_date
- abstract_flag
- abstract_binary
- idnum
- investigator
- motive
- items_taken
- casualties
- demographics
- [[customer_specific]]

DRUG

The DRUG entity type includes both legal and illicit drug names and general drug types, such as "*beta blocker*", "*marijuana*", and "*omeprazole*".

Attributes

- norm
- subtype
  - illicit
  - legal
- active_ingredients
- product_owner

© N. Harris Computer Corporation

- amount
- drug_alt_name
- weight
- price
- side_effects
- interactions
- nick_name
- drug_family
- classification_description
- classification_title
- filing_office
- filing_type
- filing_status
- color
- information_grade
- keyword
- keyword_type
- serial_number
- source_description
- source_grade
- trademark
- version
- monetary_amount
- quantity
- timestamp
- abstract_flag
- abstract_binary
- distinguishing_feature
- manufacture_date
- idnum
- dea_number
- schedule
- [[customer_specific]]

EMAIL

The EMAIL entity type represents email addresses such as "`john.doe@yahoo.com`". It does not include generic references to email or email addresses, such as "*her email*" or "*Gmail*".

Attributes

- norm

© N. Harris Computer Corporation

- subtype
- owner
- platform
- country
- organization
- ip_address
- mac_address
- source_description
- source_grade
- abstract_flag
- abstract_binary
- password
- service_provider
- [[customer_specific]]

EVENT

The EVENT entity type represents specifically named social, professional, or religious gatherings; and major world events of a specific date. EVENT also includes generic references to non-specific events such as "*the bowling tournament*", "*the last war*", or "*the attack*".

An event such as "*the 23rd annual Wing Bowl*" is extracted as "*Wing Bowl*" (without the numeral), while an event such as "*the 39th Annual Penn Linguistics Conference*" is extracted as "*39th Annual Penn Linguistics Conference*" (including the numeral), since the temporal term "annual" is capitalized and part of the title.

Attributes

- norm
- subtype
  - generic
  - named
  - weather
- start_date
- end_date
- lat (latitude in decimal degrees)
- lon (longitude in decimal degrees)
- coordinate
- location
- weapon_used
- defendant_name
- participating_individual
- holder_name
- plaintiff_name

© N. Harris Computer Corporation

- subject_name
- arrest_summons_reference
- suit_cause
- event_description
- event_title
- grid_reference
- mgrs_coordinate
- utm_coordinate
- datum_code
- ups_coordinate
- bng_coordinate
- dms_coordinate
- decimal_degree
- easting
- northing
- hemisphere
- utm_zone
- information_grade
- keyword
- keyword_type
- source_description
- source_grade
- duration
- timestamp
- event_date
- abstract_flag
- abstract_binary
- idnum
- casualties
- demographics
- desired_effect
- type
- target_details
- [[customer_specific]]

FACILITY

The FACILITY entity type represents named buildings, monumental structures, and collections of buildings. This entity type does not include generic, nominal references to facilities, such as "*the building*", "*the police station*", or "*the airport*".

© N. Harris Computer Corporation

Metonymic references for organizations that are facility names are extracted as FACILITY, not ORG. For example, in "*The Pentagon announced spending cuts*", where "*Pentagon*" is used to represent the US Department of Defense, "*Pentagon*" is extracted as a FACILITY.

Attributes

- norm
- subtype
  - academic
  - infrastructure
  - law_enforcement
  - medical
  - military
  - religious
  - stadium
  - transit
  - government
  - internal_location
- gaz_id
- dsg
- dsgrank
- dsgfeature
- pc
- country
- province
- admin_region
- lat (latitude in decimal degrees)
- lon (longitude in decimal degrees)
- elevation_m (elevation above sea level in meters)
- source
- region
- subregion
- ufi
- coordinate
- parent_organization
- date_established
- date_closure
- alternative_name
- idnum
- altitude
- address

© N. Harris Computer Corporation

- assignee_name
- attorney_name
- building_name
- chapter
- chapter_number
- classification_description
- classification_title
- collateral
- cntry_code
- email_address
- phone_number
- social_id
- grid_reference
- mgrs_coordinate
- utm_coordinate
- datum_code
- ups_coordinate
- bng_coordinate
- dms_coordinate
- decimal_degree
- easting
- northing
- hemisphere
- utm_zone
- information_grade
- keyword
- keyword_type
- line_of_business
- owner
- fax_number
- po_box
- prev_address
- secured_party_name
- source_description
- source_grade
- web_address
- credit_limit
- monetary_amount
- assets

© N. Harris Computer Corporation

- liabilities
- number_employees
- timestamp
- abstract_flag
- abstract_binary
- purchase_date
- built
- freq
- freq_range
- operational_status
- specialty_equipment
- surface_composition
- target_details
- [[customer_specific]]

GEOCOORDINATE

The GEOCOORDINATE entity type represents geocoordinates for geospatial mapping, such as "*30.99727°N 120.00213°E*" and "*1.877301, 116.171488*".

Attributes

- norm
- subtype
  - be_number (basic encyclopedic number)
  - dds (decimal degree system)
  - dms (degrees, minutes, seconds)
  - georef
  - maidenhead
  - mgrs
  - ups
  - utm
  - usng
- country
- lat (latitude in decimal degrees)
- lon (longitude in decimal degrees)
- placename
- gaz_id
- dsg
- dsgrank
- dsgfeature
- pc

© N. Harris Computer Corporation

- province
- admin_region
- elevation_m (elevation above sea level in meters)
- source
- region
- subregion
- ufi
- coordinate
- city
- cntry_code
- grid_reference
- mgrs_coordinate
- utm_coordinate
- datum_code
- ups_coordinate
- bng_coordinate
- dms_coordinate
- decimal_degree
- easting
- northing
- hemisphere
- utm_zone
- information_grade
- altitude
- source_description
- source_grade
- abstract_flag
- abstract_binary
- target_details
- [[customer_specific]]

IDNUM

The IDNUM entity type represents identification numbers such as social security numbers, account numbers, serial numbers, SIM numbers, ICCIDs, IBANs, and ISBNs.

Attributes

- norm
- country
- state
- company

© N. Harris Computer Corporation

- legal_code
- suspicious
- type
- id_user
- id_issuer
- currency_type
- source_description
- source_grade
- timestamp
- abstract_flag
- abstract_binary
- org_name
- person_name
- security_designation
- class_level
- subtype
  - accession
  - account
  - cve
  - document
  - hash
  - iban
  - isbn
  - issn
  - mac
  - serial
  - sim
  - social
  - ssn
  - swift
  - telecode
  - vin
- subset (law_enforcement_ids)
  - felony
  - misdemeanor
  - violent
  - nonviolent
- [[customer_specific]]

© N. Harris Computer Corporation

IMPLEMENT

The IMPLEMENT entity type represents tools and instruments such as farm, medical, or law enforcement equipment, as well as musical or scientific instruments.

Attributes

- norm
- company
- product_name
- type
- make
- model
- idnum
- height
- weight
- country
- suspicious
- collateral
- ip_address
- mac_address
- imsi_number
- imei_number
- iccid_number
- filing_office
- filing_type
- filing_status
- color
- owner
- serial_number
- source_description
- source_grade
- trademark
- version
- monetary_amount
- quantity
- abstract_flag
- abstract_binary
- condition
- status
- distinguishing_feature
- manufacture_date

© N. Harris Computer Corporation

- subtype
  - law_enforcement
  - medical
  - scientific
  - farm
  - military
  - miscellaneous
- [[customer_specific]]

KEYWORD

The KEYWORD entity type denotes a particular keyword or coded speech used to mask the intended content.

This entity type is intentionally left "blank" out-of-the-box so that users can customize it to their industry-specific needs.

Attributes

- norm
- subtype
- [[customer_specific]]

MONEY

The MONEY entity type represents numeric amounts of money, labeled by the currency type.

Attributes

- converted_amount
- currency
- sender
- receiver
- creditor
- debtor
- monetary_payee
- monetary_payer
- payment_method
- check_number
- currency_type
- source_description
- source_grade
- date_passed
- date_cleared
- transaction_date
- idnum

© N. Harris Computer Corporation

- abstract_flag
- abstract_binary
- routing_number
- numeric_format
- subtype
- norm
- [[customer_specific]]

ORG

The ORG entity type represents organization names. This entity type does not include nominal references to organizations, such as "*terrorists*", "*the police*", or "*the government*".

Organization names that are immediately adjacent to a PLACE include the PLACE name as part of the ORG, unless separated by a preposition. For example, "*University of Illinois, Urbana-Champaign*" includes "*Urbana-Champaign*" as part of the ORG, but "*Bradley University in Peoria, Illinois*" does not include "*Peoria, Illinois*" as part of the ORG.

Some exceptions to the above rule are "*Al Qaida in the Arabian Peninsula*", "*Al Qaida in Yemen*", and so on.

Attributes

- norm
- country
- lat (latitude in decimal degrees)
- lon (longitude in decimal degrees)
- subtype & subset
  - academic
  - commercial
  - criminal
  - education
  - financial
    - central_bank
    - commercial_bank
    - development_bank
    - investment_bank
  - government
    - executive
    - intelligence
    - judicial
    - legislative
  - inter_department
  - law_enforcement
  - medical

© N. Harris Computer Corporation

    - military
    - political_party
    - ngo
    - political
    - publisher
    - social
    - tribal
    - weapon_manufacturer
- additional_information
- affiliation
- date_established
- date_closure
- alternative_name
- idnum
- registered_name
- related_company_registration_number
- related_organization_name
- attorney_name
- chapter
- chapter_number
- classification_description
- classification_title
- company_size
- cntry_code
- email_address
- phone_number
- social_id
- address
- grid_reference
- mgrs_coordinate
- utm_coordinate
- datum_code
- ups_coordinate
- bng_coordinate
- dms_coordinate
- decimal_degree
- easting
- northing
- hemisphere

© N. Harris Computer Corporation

- utm_zone
- information_grade
- keyword
- keyword_type
- line_of_business
- owner
- fax_number
- po_box
- prev_address
- currency_type
- secured_party_name
- source_description
- source_grade
- state_county_level
- web_address
- credit_limit
- monetary_amount
- assets
- liabilities
- number_employees
- timestamp
- abstract_flag
- abstract_binary
- critical_industry
- demographics
- influenced_by
- operational_status
- order_of_battle
- specialty_equipment
- target_details
- [[customer_specific]]

PERSON

The PERSON entity type represents named people. This entity type does not include nominal references to people, such as "*terrorist*", "*the judge*", or "*the president*".

Attributes

- norm
- gender
- given_name

© N. Harris Computer Corporation

- second_given_name
- sur_name
- second_name
- nick_name
- full_name
- cause_of_death
- date_of_birth
- date_of_death
- hair_color
- hair_type
- eye_color
- height
- weight
- occupation
- second_occupation
- other_alias
- government_id_number
- nationality
- age
- physical_feature
- criminal_status
- race
- alias
- face_feature
- known_injury
- ideology
- marital_status
- blood_type
- idnum
- mother_name
- father_name
- children
- number_of_children
- allegiance
- language_spoken
- military_experience
- tribe_subtribe
- known_associate
- attorney_name

© N. Harris Computer Corporation

- relationship
- place_of_birth
- email_address
- phone_number
- social_id
- address
- grid_reference
- mgrs_coordinate
- utm_coordinate
- datum_code
- ups_coordinate
- bng_coordinate
- dms_coordinate
- decimal_degree
- easting
- northing
- hemisphere
- utm_zone
- body_build
- information_grade
- keyword
- keyword_type
- mo_identifier
- pncid_number
- cro_id_number
- national_insurance_number
- passport_number
- ssn_number
- suffix
- title
- work_number
- home_number
- cell_number
- fax_number
- pager_number
- prev_address
- secured_party_name
- source_description
- source_grade

© N. Harris Computer Corporation

- monetary_amount
- salary
- assets
- liabilities
- timestamp
- time_of_death
- weapon_used
- abstract_flag
- abstract_binary
- works_for
- account
- kia
- wia
- abd
- det
- education_level
- drivers_license
- former_occupation
- influenced_by
- language_proficiency
- license_plate_num
- operational_status
- password
- political_affiliation
- target_details
- affiliation
- convention
  - western
  - asian
  - latin
  - iberian_es
  - iberian_pt
  - hungarian
  - slavic
  - arabic
  - brazilian
  - tribal
  - mixed
- subtype

© N. Harris Computer Corporation

    - author
    - elected_official
- [[customer_specific]]

PHONE

The PHONE entity type represents any kind of device with a telephone number. The PHONE entity type includes only actual devices, not generic references to phones or phone numbers, such as "*the phone*", "*Jane's phone*", or "*dial the fax*".

Attributes

- norm
- subtype
    - domestic
    - international
- country
- device
- international_code
- area_code
- extension
- main_number
- make
- model
- provider
- city_registered
- country_registered
- belongs_to
- idnum
- phone_exchange_number
- call_type
- imsi_number
- imei_number
- iccid_number
- international_phone_region
- source_description
- source_grade
- abstract_flag
- abstract_binary
- timestamp
- sms_text_message
- [[customer_specific]]

© N. Harris Computer Corporation

PLACE

The PLACE entity type represents places with names. This entity type does not include generic, nominal references to places, such as "*the river*" or "*the big city*", but it does include generic references to specific places, as in "*the Iraqi capital*".

The PLACE entity type also does not include FACILITY, which is a separate entity. Place names adjacent to each other are considered two separate places, except in cases where the two place names adjacent to each other represent a single place, as in "*City, Province*" or "*City, Country*".

Attributes

- norm
- gaz_id
- dsg
- dsgrank
- dsgfeature
- pc
- country (ISO 3166-1 country digraph)
- state_county_level
- province (ISO 3166-2 province digraph)
- admin_region (string that corresponds to US counties or foreign equivalent)
- lat (latitude in decimal degrees)
- lon (longitude in decimal degrees)
- elevation_m (elevation above sea level in meters)
- source (string of the source)
- region
- ufi
- coordinate
- subtype
  - terrain
  - hydro
  - manmade
  - island
  - populated
  - military
  - road
  - unknown
- subregion
  - sasia (south asia)
  - seasa (south east asia)
  - easia (east asia)
  - neuro (north europe)

© N. Harris Computer Corporation

    - balk (balkans)
    - seuro (south europe)
    - ceuro (central europe)
    - nafr (north africa)
    - safr (south africa)
    - wafr (west africa)
    - cafr (central africa)
    - eafr (east africa)
    - ocean (oceania)
    - samer (south americas)
    - carib (caribbean)
    - camer (central america)
    - namer (north americas)
    - antsb (antarctica)
    - ceura (entral eurasia)
    - russa (russia)
    - cauc (caucasus)
    - meast (middle east)
- assignee_name
- cntry_code
- grid_reference
- mgrs_coordinate
- utm_coordinate
- datum_code
- ups_coordinate
- bng_coordinate
- dms_coordinate
- decimal_degree
- easting
- northing
- hemisphere
- utm_zone
- information_grade
- keyword
- keyword_type
- altitude
- secured_party_name
- source_description
- source_grade

© N. Harris Computer Corporation

- monetary_amount
- abstract_flag
- abstract_binary
- demographics
- affiliation
- surface_composition
- target_details
- [[customer_specific]]

POI

The POI entity type represents persons of interest used in social media, texts, chats, and other shorter discussions.

Attributes

- norm
- subtype
- given_name
- second_given_name
- second_name
- sur_name
- nick_name
- gender
- convention
- norm
- full_name
- cause_of_death
- date_of_birth
- date_of_death
- hair_color
- hair_type
- eye_color
- height
- weight
- occupation
- second_occupation
- other_alias
- government_id_number
- nationality
- age
- physical_feature

© N. Harris Computer Corporation

- criminal_status
- race
- alias
- face_feature
- known_injury
- ideology
- marital_status
- blood_type
- idnum
- mother_name
- father_name
- children
- number_of_children
- allegiance
- language_spoken
- military_experience
- tribe_subtribe
- known_associate
- attorney_name
- relationship
- place_of_birth
- email_address
- phone_number
- social_id
- address
- grid_reference
- mgrs_coordinate
- utm_coordinate
- datum_code
- ups_coordinate
- bng_coordinate
- dms_coordinate
- decimal_degree
- easting
- northing
- hemisphere
- utm_zone
- body_build
- information_grade

© N. Harris Computer Corporation

- keyword
- keyword_type
- mo_identifier
- pncid_number
- cro_id_number
- national_insurance_number
- passport_number
- ssn_number
- suffix
- title
- work_number
- home_number
- cell_number
- fax_number
- pager_number
- prev_address
- source_description
- source_grade
- monetary_amount
- salary
- assets
- timestamp
- weapon_used
- abstract_flag
- abstract_binary
- works_for
- account
- kia
- wia
- abd
- det
- education_level
- drivers_license
- former_occupation
- influenced_by
- language_proficiency
- license_plate_num
- operational_status
- password

© N. Harris Computer Corporation

- political_affiliation
- target_details
- affiliation
- [[customer_specific]]

PRODUCT

The PRODUCT entity type represents commercial product names, such as "*google+*" and "*chromebook*". It does not include generic references to products, such as "*wine*" or "*pretzels*".

Attributes

- norm
- subtype
- company
- product_name
- make
- model
- idnum
- height
- weight
- country
- suspicious
- collateral
- ip_address
- mac_address
- imsi_number
- imei_number
- iccid_number
- filing_office
- filing_type
- filing_status
- color
- owner
- serial_number
- source_description
- source_grade
- trademark
- version
- monetary_amount
- quantity
- timestamp

© N. Harris Computer Corporation

- abstract_flag
- abstract_binary
- purchase_date
- size
- condition
- status
- distinguishing_feature
- manufacture_date
- [[customer_specific]]

PUBLICATION

The PUBLICATION entity type represents communication works, such as written works, online sermons, news publications, and movies.

Attributes

- norm
- publication_date
- story_start_date
- title
- type
- idnum
- url
- story_end_date
- reported_date
- entered_date
- author
- review_status
- district
- division
- case_status
- classification_code
- publisher
- arrest_summons_reference
- suit_cause
- classification_description
- classification_title
- language
- filing_office
- filing_type
- filing_status

© N. Harris Computer Corporation

- currency_type
- source_description
- source_grade
- awards
- filing_amount
- transaction_date
- abstract_flag
- abstract_binary
- encryption_type
- summary
- subtype
  - form
  - report
  - publisher
  - government
  - travel
  - financial
  - law
  - medical
  - academic
  - publication_date
  - propaganda
  - issued_identification
- [[customer_specific]]

PUNITIVE_MEASURE

The PUNITIVE_MEASURE entity type represents a punishment made against an entity, such as imprisonment, torture, or fines.

Attributes

- norm
- start_date
- end_date
- lat (latitude in decimal degrees)
- lon (longitude in decimal degrees)
- coordinate
- location
- weapon_used
- holder_name
- participating_individual

© N. Harris Computer Corporation

- source_description
- source_grade
- event_date
- monetary_amount
- abstract_flag
- abstract_binary
- subtype
  - death
  - prison
  - torture
  - on_record
  - financial
- [[customer_specific]]

SOCIAL

The SOCIAL entity type represents identifiers relevant to social media platforms, such as usernames and handles. It does not include email addresses, hashtags, or website URLs, which are separate entity types.

Attributes

- norm
- subtype
- owner
- platform
- country
- organization
- source_description
- source_grade
- abstract_flag
- abstract_binary
- password
- service_provider
- [[customer_specific]]

TIMESPAN

The TIMESPAN entity type indicates a range of time or ranges of dates, such as "*Mon-Thur*" or "*9:00-11:00pm*".

Attributes

- norm
- subtype
- date_range_and_location

© N. Harris Computer Corporation

- [[customer_specific]]

TIMESTAMP

The TIMESTAMP entity type represents date & time expressions.

Date & time expressions that are adjacent to each other and represent a narrowing or broadening of the same time are extracted as a single entity. Date & time expressions that are adjacent to each other but represent distinct temporal expressions are extracted as separate entities.

Attributes

- norm
- holiday
- weekday
- day
- month
- year
- hour
- minute
- second
- ampm
- timezone
- epoch
- era
- century
- scope
- direction
- multiplier
- to_timestamp
- from_timestamp
- subtype & subset
  - absolute
  - holiday
  - range
  - relative
    - issue_date
    - expiration_date
- [[customer_specific]]

TRAUMA

The TRAUMA entity type describes possible lethal injury, such as dismemberment.

Attributes

© N. Harris Computer Corporation

- norm
- subtype
- location
- size
- trauma_level
- vitals_affected
- treatment_required
- burn_type
- break_type
- recovery_time
- medication_required
- subject_name
- [[customer_specific]]

URL

The URL entity type represents web addresses such as websites and IP addresses. It does not include other locators in cyberspace such as email addresses, social media usernames, and so on.

URL also excludes generic references to web addresses, such as "*the company's website*" or "*Google*".

Attributes

- norm
- intelligence_date
- reference
- source
- provider
- related_web_address
- belongs_to
- source_description
- source_grade
- abstract_flag
- abstract_binary
- service_provider
- account_type
- subtype
  - URL
  - ipv4
  - ipv6
  - ip_range
- [[customer_specific]]

© N. Harris Computer Corporation

WEAPON

The WEAPON entity type represents weapons of a conventional, CBRNE, or cyber nature. Unlike other entity types, WEAPON includes generic weapon terms such as "*3 rifles*" as well as specific weapon terms such as "*3 AK-47s*".

Attributes

- norm
- manufacturer
- place_of_manufacture
- size
- weapon_type
- ammunition
- country
- idnum
- user
- coordinate
- lat
- lon
- name
- range
- speed
- model
- filing_office
- filing_type
- filing_status
- grid_reference
- mgrs_coordinate
- utm_coordinate
- datum_code
- ups_coordinate
- bng_coordinate
- dms_coordinate
- decimal_degree
- easting
- northing
- hemisphere
- utm_zone
- color
- information_grade
- keyword

© N. Harris Computer Corporation

- keyword_type
- altitude
- owner
- serial_number
- source_description
- source_grade
- trademark
- version
- weight
- monetary_amount
- quantity
- timestamp
- abstract_flag
- abstract_binary
- purchase_date
- condition
- status
- distinguishing_feature
- manufacture_date
- operational_status
- target_details
- subset
- subtype
  - cbrne
  - conventional
  - cyber
  - ammunition
- [[customer_specific]]

No-output entities

No-output entities are entities that i2 TextChart extracts, but whose output it suppresses.

ALERT_TYPE

The ALERT_TYPE entity type represents travel alerts and travel warnings from the State Department.

Attributes

- norm
- subtype
- description

© N. Harris Computer Corporation

- idnum
- organization
- level
- area_of_alert
- alert_delivery_method
- alert_feed_subscription
- alert_retrieval
- [[customer_specific]]

ANATOMICAL_TERM

The ANATOMICAL_TERM entity type describes parts of the body.

Attributes

- norm
- subtype
- location
- size
- injury_type
- condition
- weapon
- [[customer_specific]]

AWARD

The AWARD entity type represents a mark of recognition in honor of an achievement, such as "*Grammy*", "*World Series Champion*", or "*Magna Cum Laude*".

Attributes

- norm
- issuing_authority
- monetary_amount
- recipient
- nominee
- date_of_award
- award_item
- subtype
  - sport
  - music
  - academic
  - beauty
  - film
  - animal

© N. Harris Computer Corporation

    - military
    - design
    - technology
- [[customer_specific]]

BIOMETRIC

The BIOMETRIC entity type extracts mentions of biometric data present in text, such as "*fingerprints*" and "*iris scan*".

Attributes

- norm
- subtype (by default = unknown)
    - match
    - no_match
    - unknown
- [[customer_specific]]

CHEMICAL

The CHEMICAL entity type represents named chemicals.

Attributes

- norm
- common_name
- cas
- symbol
- atomic_mass
- atomic_number
- melting_point
- density
- appearance
- formula
- ph_level
- iupac_id
- molar_mass
- more_general_ingredient
- symptoms
- health_effects
- toxicity
- subtype
    - drug_active_ingredient
- [[customer_specific]]

© N. Harris Computer Corporation

CITATION

The CITATION entity type denotes legal citation numbers, such as "*Smith v. Doe, 2000 VA. 3, Ariz. 53, 120 Miss 12*".

Attributes

- norm
- publication_date
- title
- type
- idnum
- url
- reported_date
- entered_date
- author
- review_status
- district
- division
- case_status
- classification_code
- publisher
- defendant_name
- holder_name
- plaintiff_name
- subject_name
- suit_cause
- currency_type
- awards
- filing_amount
- abstract_flag
- abstract_binary
- summary
- subtype
- [[customer_specific]]

CLASSIFICATION_LEVEL

The CLASSIFICATION_LEVEL entity type represents a used classification level.

Attributes

- norm
- subtype
- class_level

© N. Harris Computer Corporation

- clear_access
- owner
- assignee
- issuer
- confidence_evaluation
- [[customer_specific]]

CONTRACT_TYPE

The CONTRACT_TYPE entity type represents a type of contracting vehicle, such as "*idiq*", "*firm-fixed price*", and so on.

Attributes

- norm
- subtype
- [[customer_specific]]

CONTROL

The CONTROL entity type represents document controls and classification markings.

Attributes

- norm
- subtype
  - TS
  - S
  - C
  - U
- [[customer_specific]]

DISEASE

The DISEASE entity type represents names of diseases and medical conditions or episodes, such as "*hemophilia*" and "*heart attack*".

Attributes

- norm
- subtype
  - disorder
  - hiv_aids
  - cancer
  - viral
  - bacterial
  - encephalitis
  - injury

© N. Harris Computer Corporation

- icd10
- icd9
- treatment
- mortality rate
- symptoms
- associated_prescriptions
- treatment_success
- subject_name
- classification_description
- classification_title
- information_grade
- keyword
- keyword_type
- source_description
- source_grade
- timestamp
- abstract_flag
- abstract_binary
- [[customer_specific]]

DNA

The DNA entity type represents DNA sequences, such as "*forward (position 552-570) 5#-CCAAACCCCAAAGACACCC-3#*".

Attributes

- norm
- subtype
- gender
- identified_illness
- sequence
- modification_site
- [[customer_specific]]

FILE_N#ME

The FILE_N#ME entity type represents file names found between quotation marks, such as `"updated_img0912.jpg"` or `"documentation.html"`.

Attributes

- norm
- subtype
- publication_date
- story_start_date

© N. Harris Computer Corporation

- title
- type
- idnum
- url
- story_end_date
- reported_date
- entered_date
- author
- review_status
- district
- division
- classification_code
- case_status
- source_description
- source_grade
- abstract_flag
- abstract_binary
- encryption_type
- size
- [[customer_specific]]

FINANCIAL_INDEX

The FINANCIAL_INDEX entity type represents financial market indices such as "*Nasdaq*" or "*S&P 500*".

Attributes

- norm
- subtype
  - exchange
  - index
- country
- price
- index_name
- index_high
- index_low
- prev_close
- [[customer_specific]]

FUNDS

The FUNDS entity type denotes a named or general reference to a monetary fund, grant, or tax.

Attributes

© N. Harris Computer Corporation

- norm
- subtype
- issuing_authority
- monetary_amount
- recipient
- idnum
- fund_awarded_on
- fund_type
- [[customer_specific]]

GENE

The GENE entity type indicates the genetic code that corresponds to a gene, such as "*LYRM4*" and "*FARS2*".

Attributes

- norm
- identified_illness
- sequence
- modification_site
- therapy_applied
- subtype
  - nucleotide
  - locus
  - genetic_id
  - gene_variant
  - gene_product
- [[customer_specific]]

GENERIC

The GENERIC entity type serves to extract non-specific entities such as "*suspect*".

Attributes

- norm
- subtype
  - person
  - org
  - place
  - facility
- [[customer_specific]]

© N. Harris Computer Corporation

HASHTAG

The HASHTAG entity type represents social media hashtags for crowd source topics, such as "*ShareaCoke*".

Attributes

- norm
- subtype
- [[customer_specific]]

IDEOLOGY

The IDEOLOGY entity type represents religious entities, such as "*muslim*" or "*neo-evangelicalism*".

Attributes

- norm
- source_description
- source_grade
- subtype
  - western
  - eastern
  - historic
  - native
  - pagan
  - fictional
  - islamic
  - judaism
  - christian
  - political
- [[customer_specific]]

INFRASTRUCTURE

The INFRASTRUCTURE entity type represents fundamental facilities that serve a geographic area.

Attributes

- norm
- subtype
- city
- country
- state
- type
- style
- functional_usage
- owner

© N. Harris Computer Corporation

- company
- builder
- price
- idnum
- made_from
- adjacent_structures
- [[customer_specific]]

MEASURE

The MEASURE entity type represents measured quantities with units, such as "*12 years*" or "*14.5 km*".

Attributes

- norm
- hours
- minutes
- seconds
- amp
- degree
- joule
- pound
- kg
- psi
- volt
- watt
- liter
- gram
- gallon
- ounce
- meter
- kilometer
- mile
- acre
- celsius
- fahrenheit
- kelvin
- altitude
- depth
- unit
- numeric_format
- subtype

© N. Harris Computer Corporation

    - temperature
    - orbit_location
    - distance
    - speed
    - volume
    - altitude
    - weight
    - time
    - energy
    - length
    - age
    - dimension
    - frequency
    - physical
    - height
    - radiation
- [[customer_specific]]

MEDICAL_PROCEDURE

The MEDICAL_PROCEDURE entity type represents medical procedures such as "*biopsy*" or "*angiography*".

Attributes

- norm
- subtype
- icd10
- icd9
- surgical
- related_procedures
- rate_of_success
- physician_to_perform
- region_of_body
- item_removed
- item_implanted
- equipment_needed
- medication_prescribed
- timestamp
- event_date
- monetary_amount

© N. Harris Computer Corporation

- abstract_flag
- abstract_binary
- [[customer_specific]]

MISC

The MISC entity type allows for the display and handling of miscellaneous items as if they are entities. Users can use this entity type in the same manner.

Attributes

- [[customer_specific]]

NATIONALITY

The NATIONALITY entity type represents the nation or region of origin for a person or an object, such as "*Japanese*" or "*Asian*".

Attributes

- norm
- gender
- country
- region
- source_description
- source_grade
- subtype
  - national_of
  - ethnicity
- [[customer_specific]]

NON_SALIENT_WEB_CONTENT

The NON_SALIENT_WEB_CONTENT entity type represents terms that are frequently found on a website but have no bearing on its content, such as "*downloads*", "*connect on...*", and "*cookie policy*".

Attributes

- norm
- subtype
- [[customer_specific]]

PERCENT

The PERCENT entity type represents percentages. This entity type also includes "per mille", which the extraction process converts to percent automatically.

The word "percent" is normalized to the percent symbol (%) during extraction.

Attributes

- norm
- subtype

© N. Harris Computer Corporation

- [[customer_specific]]

POLITICAL_AFFILIATION

The POLITICAL_AFFILIATION entity type represents the political affiliation of an entity, such as "*R-AK*" or "*republican*".

Attributes

- norm
- admin_region
- party
- person_affiliated
- [[customer_specific]]

PROFESSION

The PROFESSION entity type indicates the title and profession of an entity, such as "*cashier*" or "*House Majority Whip*".

Attributes

- norm
- subtype
- location
- profession_org
- salary
- reporting_profession
- [[customer_specific]]

PROGRAM

The PROGRAM entity type indicates organized programs, plans, and projects, such as "*Abandoned Land Mine Reclamation Program*", "*Adult Education State Grant Program*", and "*Beach Erosion Control Project*".

Attributes

- norm
- type
- idnum
- url
- program_start_date
- program_end_date
- issuing_organization
- program_status
- program_value
- issuing_person
- program_location

© N. Harris Computer Corporation

- subtype
  - curriculum
  - program
  - project
  - grant
- [[customer_specific]]

QUOTE

The QUOTE type entity indicates and highlights possible quotes of interest by an entity.

Attributes

- norm
- subtype
- [[customer_specific]]

RATING

The RATING entity type represents bond rating scores of the kind found in financial documents, such as "*BBB-*" and "*Lowered to A+ from AA-*".

Attributes

- norm
- subtype
- credit_quality
- rating_organization
- grade
- description
- [[customer_specific]]

SALIENT_PHRASE

The SALIENT_PHRASE entity type represents a dynamic entity or phrase that has not been extracted as another entity, but appears in a salient context.

Attributes

- norm
- subtype
- [[customer_specific]]

SCORE

The SCORE entity type represents the number of points an entity received in a competition, its rating, or its grade.

Attributes

- norm
- subtype

© N. Harris Computer Corporation

  - rating
  - grade
  - ranking
  - record
  - nomination
  - value
  - competition_result
- [[customer_specific]]

TICKER_SYMBOL

The TICKER_SYMBOL entity type represents ticker symbols used on financial exchanges, such as "*AAPL*" or "*AZO*".

Attributes

- norm
- company (the company represented by the ticker symbol)
- exchange (the exchange on which the company/symbol is traded)
- subtype
- country
- price
- index_name
- index_high
- index_low
- prev_close
- [[customer_specific]]

TRANSIT

The TRANSIT entity type represents transportation facilities such as subway stops, airports, and bus stops.

Attributes

- norm
- subtype
- gaz_id
- dsg
- dsgrank
- dsgfeature
- pc
- country
- province
- admin_region

© N. Harris Computer Corporation

- lat (latitude in decimal degrees)
- lon (longitude in decimal degrees)
- elevation_m (elevation above sea level in meters)
- source
- region
- subregion
- ufi
- coordinate
- [[customer_specific]]

USER_AGENT

The USER_AGENT entity type represents software user agents such as bots or email readers. The entity is extracted when preceded by an indicator such as "*user agent:*" and contained within quotes.

Attributes

- norm
- subtype
- [[customer_specific]]

Relationships

In i2 TextChart, *relationships* are binary connections between entities found in a single document. TextChart defines relationships in terms of *predicate-subject-object* (PSO) statements.

PSO relationships have an order of precedence: PERSON is the most important entity type, and all other types are subordinate to it. All relationships are reciprocal by default.

Each relationship that TextChart extracts has a predicate from a set of predicate types that are valid for that relationship. The valid set for each relationship is drawn from the full list of predicate types.

PersonToPerson

PersonToPerson relationships are those relationships that connect two persons together. These relationships can be between different people or with the same person who has multiple names.

Predicate types

- acquires_
- disagreed
- govern_
- accused
- complaints
- inform
- select
- alias_of
- dissent
- appointed

© N. Harris Computer Corporation

- convert
- acknowledge
- alleges
- against
- follows_
- cared_for
- manipulates
- works_
- identified_
- confronted
- burial
- negotiated_with
- employs
- contact
- invested_in
- knows_
- works_with
- donate
- took
- inquire_
- like
- supports_
- competes_with
- communicated_
- discredit
- mandate
- penalize_
- researched
- seen
- kidnapped_
- awarded_
- assists
- advise
- instructs
- crime
- win
- died_from
- enclose
- witnessed

© N. Harris Computer Corporation

- caused
- emailed
- apprehended
- sponsored
- attack_
- interviewed
- battle
- meet
- trusts
- compensated_
- visited_
- died_from
- united_
- describe
- separated_
- terrified
- criticize
- received_information
- received_legal_assistance
- reported_to
- sexual_relations
- threaten
- surrendered
- abuse
- deceive
- linked_to
- familial_
  - familial_
  - cousin_of
  - child_parent_of
  - grandchild_grandparent_of
  - greatgrandchild_greatgrandparent_of
  - pibling_of
  - nibling_of
  - sibling_of
  - spouse_of
  - related_by_marriage
  - related_to
- [[customer_specific]]

© N. Harris Computer Corporation

PersonToOrg

PersonToOrg relationships are those relationships that connect a person with an organization. These relationships are *specific relationships*, meaning that this relationship will only be extracted when a predicate type is specified.

Predicate types

- belongs_to
- acquires_
- communicated_
- discredit
- inform
- select
- works_with
- compensated_
- attests
- funded
- negotiated_with
- dispatched
- disagreed
- against
- alleges
- dissent
- threaten
- complaints
- appointed
- surrendered
- confronted
- apprehended
- witnessed
- employs
- contact
- kidnapped_
- issued_
- made_
- follows_
- host
- acknowledge
- accused
- owns
- create

© N. Harris Computer Corporation

- takes_control
- donate
- consider
- works_
- depart
- knows_
- cared_for
- make_improvements
- inquire_
- penalize_
- invested_in
- failed_
- announce
- works_with
- improve
- formed
- launches
- mandate
- contribute
- appears_
- amplify
- researched
- transform
- assists
- advise
- terrified
- instructs
- access_to
- awarded_
- studies
- win
- removed_from
- monetary_earning
- enclose
- visited_
- works_with
- sells_
- supplied
- emailed

© N. Harris Computer Corporation

- manages
- united_
- escaped_from
- endorse
- meet
- trusts
- abuse
- deceive
- died_from
- describe
- separated_
- criticize
- received_information
- received_legal_assistance
- reported_to
- identified_
- linked_to
- [[customer_specific]]

PersonToFacility

PersonToFacility relationships are those relationships that connect a person with a facility. These relationships are specific relationships.

Predicate types

- acquires_
- inform
- select
- communicated_
- discredit
- funded
- issued_
- at
- closed
- apprehended
- works_with
- complaints
- confronted
- dissent
- alleges
- threaten

© N. Harris Computer Corporation

- accused
- compensated_
- employs
- disagreed
- acknowledge
- contact
- cared_for
- made_
- terrified
- owns
- kidnapped_
- donate
- works_
- depart
- make_improvements
- location_of
- invested_in
- announce
- visited_
- move_to
- attack_
- travels_
- appears_
- targets
- researched
- access_to
- operates_in
- enclose
- removed_from
- invade
- sells_
- manages
- escaped_from
- services
- move_from
- meet_at
- damaged
- demolished
- describe

© N. Harris Computer Corporation

- separated_
- criticize
- received_information
- reported_to
- abuse
- death
- deceive
- linked_to
- [[customer_specific]]

PersonToAddress

PersonToAddress relationships are those relationships that connect a person with a physical address.

Predicate types

- maintains
- belongs_to
- acquires_
- kidnapped_
- manages
- location_of
- services
- works_
- terrified
- move_from
- operates_in
- acknowledge
- meet
- move_to
- improve
- damaged
- knows_
- visited_
- made_
- demolished
- access_to
- owns
- depart
- make_improvements
- removed_from
- criticize

© N. Harris Computer Corporation

- linked_to
- [[customer_specific]]

PersonToPlace

PersonToPlace relationships are those relationships that connect a person with a place. These relationships are specific relationships.

Predicate types

- belongs_to
- death
- travels_
- at
- accused
- govern_
- dissent
- surrendered
- dispatched
- threaten
- funded
- apprehended
- from_
- alleges
- escaped_from
- location_of
- against
- appears_
- disagreed
- appointed
- complaints
- takes_control
- targets
- damaged
- acknowledge
- researched
- burial
- studies
- terrified
- knows_
- operates_in
- depart

© N. Harris Computer Corporation

- kidnapped_
- monetary_earning
- seen
- discover
- invade
- birth
- citizen_of
- visited_
- move_from
- meet
- move_to
- demolished
- attack_
- battle
- describe
- communicated_
- discredit
- separated_
- criticize
- works_
- embargo
- abuse
- deceive
- identified_
- linked_to
- [[customer_specific]]

PersonToWeapon

The PersonToWeapon relationship represents relationships that connect a person with a physical or virtual weapon.

Predicate types

- discover
- acquires_
- expert_on
- location_of
- sends_
- surrendered
- made_
- attack_

© N. Harris Computer Corporation

- authorized
- researched
- lost
- gave_
- compensated_
- access_to
- owns
- has
- died_from
- acknowledge
- witnessed
- took
- inquire_
- sells_
- seen
- invested_in
- died_from
- uses
- dispose
- describe
- criticize
- took
- linked_to
- [[customer_specific]]

PersonToDrug

The PersonToDrug relationship represents relationships between people and legal or illicit drugs.

Predicate types

- acquires_
- expert_on
- authorized
- gave_
- sends_
- alleges
- complaints
- appears_
- made_
- researched
- branded

© N. Harris Computer Corporation

- access_to
- acknowledge
- has
- seen
- consider
- sells_
- died_from
- affected_by
- discover
- took
- inquire_
- invested_in
- provider_of
- uses
- dispose
- describe
- advertise
- criticize
- treating
- apprehended
- supplied
- survived_
- [[customer_specific]]

PersonToChemical

PersonToChemical relationships link persons to chemicals.

Predicate types

- researched
- acquires_
- sells_
- expert_on
- access_to
- gave_
- uses
- dispose
- has
- made_
- died_from
- criticize

© N. Harris Computer Corporation

- treating
- [[customer_specific]]

PersonToConveyance

The PersonToConveyance relationship represents relationships between persons and methods of conveyance, including land, air, sea, and space vehicles.

Predicate types

- took
- inquire_
- authorized
- gave_
- complaints
- canceled
- funded
- is_evaluating
- death
- acquires_
- sells_
- made_
- sends_
- attack_
- researched
- branded
- acknowledge
- demolished
- dispatched
- access_to
- uses
- seen
- dispose
- has
- describe
- consider
- depart
- launches
- make_improvements
- separated_
- criticize
- travels_

© N. Harris Computer Corporation

- [[customer_specific]]

PersonToCrime

The PersonToCrime relationship denotes relationships between persons and crimes.

Predicate types

- witnessed
- authorized
- inquire_
- mandate
- belongs_to
- funded
- penalize_
- accused
- conviction_
- threaten
- attack_
- researched
- assists
- compensated_
- charged_with
- describe
- affected_by
- commits
- seen
- criticize
- denied
- involvement
- suspected
- alleges
- prevented_
- participated_in
- survived_
- plotted_
- absolved_
- linked_to
- [[customer_specific]]

PersonToDisease

The PersonToDisease relationship denotes relationships between persons and diseases.

© N. Harris Computer Corporation

Predicate types

- researched
- expert_on
- separated_
- appears_
- gave_
- afflicted_with
- has
- describe
- acknowledge
- affected_by
- communicated_
- died_from
- criticize
- suspected
- studies
- isolate
- originated_
- variant_of
- treating
- survived_
- immunized_
- [[customer_specific]]

PersonToProduct

The PersonToProduct relationship denotes relationships between persons and commercial products.

Predicate types

- acquires_
- select
- authorized
- funded
- unveil
- canceled
- expert_on
- complaints
- contribute
- sends_
- appears_
- made_

© N. Harris Computer Corporation

- researched
- alleges
- acknowledge
- branded
- compensated_
- access_to
- has
- consider
- sells_
- took
- inquire_
- invested_in
- guaranteed
- announce
- damaged
- uses
- dispose
- describe
- advertise
- criticize
- donate
- [[customer_specific]]

PersonToMedical_Procedure

The PersonToMedical_Procedure relationship represents relationships between a person and a medical procedure.

Predicate types

- discover
- inquire_
- authorized
- unveil
- expert_on
- perform
- contribute
- received
- researched
- assists
- compensated_
- studies

© N. Harris Computer Corporation

- create
- has
- acknowledge
- consider
- affected_by
- undergo
- died_from
- criticize
- treating
- linked_to
- [[customer_specific]]

PersonToEvent

The PersonToEvent relationship represents relationships between persons and events.

Predicate types

- implements
- communicated_
- inform
- travels_
- funded
- discredit
- dissent
- unveil
- contribute
- attended
- appears_
- complaints
- works_with
- researched
- visited_
- assists
- advise
- employs
- contact
- instructs
- host
- depart
- win
- died_from

© N. Harris Computer Corporation

- witnessed
- inquire_
- manages
- sponsored
- participated_in
- competes_with
- attack_
- describe
- affected_by
- launches
- removed_from
- criticize
- initiated
- involvement
- compensated_
- linked_to
- [[customer_specific]]

PersonToPublication

The PersonToPublication relationships are those relationships that connect a person with a publication.

Predicate types

- communicated_
- inform
- subject_of
- contribute
- accused
- complaints
- funded
- discredit
- disagreed
- against
- alleges
- signed
- sponsored
- mentioned_in
- sends_
- appears_
- acknowledge
- published_in

© N. Harris Computer Corporation

- branded
- referenced
- contact
- instructs
- author_of
- has
- sells_
- affected_by
- works_
- advertise
- criticize
- denied
- submitted
- linked_to
- [[customer_specific]]

PersonToEmail

The PersonToEmail relationship represents the relationship between a person and an email address.

Predicate types

- discover
- belongs_to
- inform
- originated_
- subject_of
- sells_
- emailed
- sends_
- opened
- received
- mentioned_in
- location_of
- appears_
- communicated_
- identified_by
- contact
- author_of
- access_to
- uses
- has

© N. Harris Computer Corporation

- linked_to
- [[customer_specific]]

PersonToSocial

PersonToSocial relationships are those relationships between a person and a social media identifier.

Predicate types

- maintains
- belongs_to
- communicated_
- inform
- disagreed
- confronted
- complaints
- acknowledge
- subject_of
- accused
- alleges
- caused
- manages
- mentioned_in
- appears_
- identified_by
- referenced
- contact
- follows_
- uses
- affected_by
- criticize
- linked_to
- [[customer_specific]]

PersonToPhone

PersonToPhone relationships are those relationships that connect a person to a phone number.

Predicate types

- belongs_to
- inform
- communicated_
- identified_by
- originated_

© N. Harris Computer Corporation

- contact
- uses
- location_of
- has
- linked_to
- [[customer_specific]]

PersonToURL

PersonToURL relationships are those relationships that connect a person to a website or IP address.

Predicate types

- maintains
- belongs_to
- acquires_
- inform
- identified_by
- create
- owns
- appears_
- launches
- linked_to
- [[customer_specific]]

PersonToContact_Block

The PersonToContact_Block relationship represents relationships between a person and a contact block.

Predicate types

- mentioned_in
- contact
- [[customer_specific]]

PersonToGeocoordinate

The PersonToGeocoordinate relationship represents relationships between a person and a geocoordinate.

Predicate types

- location_of
- appears_
- [[customer_specific]]

© N. Harris Computer Corporation

PersonToMoney

PersonToMoney relationships are those relationships that connect a person to a numeric amount of money.

Predicate types

- took
- inquire_
- belongs_to
- penalize_
- made_
- sends_
- received
- gave_
- guaranteed
- monetary_earning
- laundered
- donate
- gain_
- loss
- acquires_
- lost
- compensated_
- access_to
- funded
- invested_in
- has
- sells_
- deceive
- win
- [[customer_specific]]

PersonToPercent

The PersonToPercent relationship represents a relationship between a person and a percentage.

Predicate types

- [[customer_specific]]

PersonToMeasure

The PersonToMeasure relationship connects a person to a numeric measure and unit.

Predicate types

- identified_

© N. Harris Computer Corporation

- death
- age_of
- weight_of
- height_of
- [[customer_specific]]

PersonToIDNUM

PersonToIDNUM relationships are those relationships that connect a person to an identification number.

Predicate types

- belongs_to
- awarded_
- identified_by
- referenced
- compensated_
- has
- [[customer_specific]]

PersonToCitation

The PersonToCitation relationship denotes a relationship between a person and a legal citation number.

Predicate types

- referenced
- author_of
- mentioned_in
- [[customer_specific]]

PersonToDNA

The PersonToDNA relationship represents those relationships that connect a person to a DNA sequence.

Predicate types

- discover
- belongs_to
- expert_on
- appears_
- convert
- researched
- isolate
- transform
- identified_
- identified_by

© N. Harris Computer Corporation

- compensated_
- studies
- ethnic_group
- afflicted_with
- has
- affected_by
- communicated_
- died_from
- is_evaluating
- linked_to
- [[customer_specific]]

PersonToHashtag

The PersonToHashtag relationship represents those relationships that connect a person to a social media hashtag.

Predicate types

- belongs_to
- referenced
- mentioned_in
- sends_
- communicated_
- criticize
- linked_to
- [[customer_specific]]

PersonToTimestamp

The PersonToTimestamp relationship represents a connection between a person and a date & time expression.

Predicate types

- death
- birth
- conviction_
- burial
- meet
- attack_
- apprehended
- linked_to
- [[customer_specific]]

© N. Harris Computer Corporation

PersonToProgram

The PersonToProgram relationship represents a connection between a person and a program, plan, or project.

Predicate types

- belongs_to
- inform
- contribute
- appears_
- funded
- amplify
- works_with
- discredit
- against
- assists
- compensated_
- advise
- contact
- instructs
- access_to
- donate
- works_
- depart
- removed_from
- inquire_
- emailed
- manages
- invested_in
- describe
- affected_by
- formed
- launches
- complaints
- criticize
- denied
- involvement
- linked_to
- [[customer_specific]]

© N. Harris Computer Corporation

PersonToPunitive_Measure

PersonToPunitive_Measure relationships are those relationships that connect a person with a punitive measure.

Predicate types

- witnessed
- implements
- amount
- mandate
- date_of
- penalize_
- against
- death
- perform
- expert_on
- conviction_
- received
- died_from
- enforce
- has
- afflicted_with
- affected_by
- criticize
- [[customer_specific]]

PersonToNationality

The PersonToNationality relationship represents a connection between a person and a nationality.

Predicate types

- belongs_to
- works_with
- united_
- communicated_
- citizen_of
- apprehended
- manipulates
- disagreed
- identified_
- compensated_
- employs
- seen

© N. Harris Computer Corporation

- against
- ethnic_group
- attack_
- access_to
- knows_
- works_
- creed
- monetary_earning
- criticize
- sexual_relations
- abuse
- deceive
- linked_to
- [[customer_specific]]

PersonToProfession

The PersonToProfession relationship represents a connection between a person and a profession.

Predicate types

- inquire_
- inform
- select
- manages
- contribute
- works_with
- disagreed
- announce
- trusts
- advise
- employs
- contact
- against
- instructs
- describe
- consider
- works_
- depart
- communicated_
- removed_from
- criticize

© N. Harris Computer Corporation

- [[customer_specific]]

PersonToPolitical_Affiliation

The PersonToPolitical_Affiliation relationship represents a connection between a person and a political affiliation.

Predicate types

- belongs_to
- inform
- select
- complaints
- subject_of
- govern_
- disagreed
- convert
- amplify
- works_with
- assists
- compensated_
- negotiated_with
- advise
- contact
- instructs
- has
- works_
- depart
- penalize_
- sponsored
- participated_in
- announce
- supports_
- formed
- separated_
- criticize
- discredit
- linked_to
- [[customer_specific]]

PersonToTransit

The PersonToTransit relationship represents a connection between a person and a transit entity.

© N. Harris Computer Corporation

Predicate types

- took
- travels_
- improve
- manages
- access_to
- location_of
- attack_
- appears_
- made_
- works_
- make_improvements
- [[customer_specific]]

PersonToGene

The PersonToGene relationship represents a connection between a person and a gene entity.

Predicate types

- discover
- like
- expert_on
- appears_
- convert
- targets
- isolate
- researched
- transform
- identified_
- identified_by
- received
- gave_
- damaged
- studies
- ethnic_group
- afflicted_with
- has
- affected_by
- communicated_
- died_from
- is_evaluating

© N. Harris Computer Corporation

- linked_to
- [[customer_specific]]

PersonToImplement

The PersonToImplement relationship represents a connection between a person and an implement.

Predicate types

- took
- funded
- discover
- belongs_to
- acquires_
- sells_
- made_
- apprehended
- sends_
- access_to
- owns
- died_from
- uses
- invested_in
- has
- separated_
- linked_to
- [[customer_specific]]

PersonToNon_Salient_Web_Content

The PersonToNon_Salient_Web_Content relationship represents a connection between a person and non-salient web content.

Predicate types

- participated_in
- takes_action
- [[customer_specific]]

PersonToIdeology

The PersonToIdeology relationship represents a connection between a person and an ideology.

Predicate types

- follows_
- belongs_to
- expert_on

© N. Harris Computer Corporation

- participated_in
- convert
- amplify
- works_with
- against
- researched
- identified_
- takes_action
- instructs
- commits
- depart
- creed
- separated_
- abuse
- criticize
- [[customer_specific]]

PersonToFile_Name

The PersonToFile_Name relationship represents a connection between a person and a file name.

Predicate types

- emailed
- access_to
- create
- mentioned_in
- sends_
- [[customer_specific]]

PersonToAward

The PersonToAward relationship represents a connection between a person and an award.

Predicate types

- acquires_
- belongs_to
- received
- has
- win
- lost
- criticize
- [[customer_specific]]

© N. Harris Computer Corporation

PersonToScore

The PersonToScore relationship represents a connection between a person and a score.

Predicate types

- maintains
- acquires_
- received
- has
- [[customer_specific]]

PersonToBiometric

The PersonToBiometric relationship represents a connection between a person and a biometric entity.

Predicate types

- linked_to
- [[customer_specific]]

PersonToGeneric

The PersonToGeneric relationship represents a connection between a person and a generic entity.

Predicate types

- linked_to
- emailed
- alleges
- [[customer_specific]]

OrgToOrg

OrgToOrg relationships are those relationships that connect an organization with another organization.

Predicate types

- acquires_
- belongs_to
- inform
- select
- against
- funded
- originated_
- complaints
- dissent
- accused
- alleges
- disagreed
- alias_of

© N. Harris Computer Corporation

- compensated_
- discredit
- acknowledge
- threaten
- negotiated_with
- dispatched
- takes_action
- employs
- contact
- owns
- create
- works_with
- donate
- consider
- depart
- inquire_
- penalize_
- made_
- invested_in
- competes_with
- formed
- awarded_
- mandate
- contribute
- amplify
- surrendered
- assists
- advise
- takes_control
- win
- monetary_earning
- sells_
- supplied
- manages
- united_
- sponsored
- endorse
- battle
- contracted_with

© N. Harris Computer Corporation

- meet
- trusts
- describe
- merger
- separated_
- criticize
- received_information
- received_legal_assistance
- reported_to
- abuse
- deceive
- linked_to
- [[customer_specific]]

OrgToFacility

The OrgToFacility relationship connects an organization with a facility.

Predicate types

- acquires_
- belongs_to
- funded
- competes_with
- govern_
- inform
- authorized
- select
- complaints
- threaten
- dissent
- compensated_
- disagreed
- negotiated_with
- takes_action
- acknowledge
- accused
- contact
- alleges
- made_
- owns
- has

© N. Harris Computer Corporation

- donate
- depart
- contracted_with
- make_improvements
- invested_in
- opened
- announce
- works_with
- move_to
- improve
- location_of
- attack_
- unveil
- damaged
- transform
- operates_in
- takes_control
- enclose
- maintains
- invade
- sells_
- manages
- move_from
- services
- move_from
- meet
- demolished
- closed
- uses
- describe
- separated_
- criticize
- merger
- received_information
- received_legal_assistance
- reported_to
- abuse
- deceive
- linked_to

© N. Harris Computer Corporation

- [[customer_specific]]

OrgToAddress

The OrgToAddress relationship represents relationships between organizations and physical addresses.

Predicate types

- belongs_to
- acquires_
- location_of
- services
- damaged
- move_from
- operates_in
- meet
- acknowledge
- move_to
- made_
- has
- received_information
- linked_to
- [[customer_specific]]

OrgToPlace

The OrgToPlace relationship represents a connection between an organization and a place.

Predicate types

- invade
- belongs_to
- originated_
- funded
- travels_
- dispatched
- confronted
- govern_
- dissent
- alleges
- discredit
- complaints
- against
- move_from
- accused

© N. Harris Computer Corporation

- threaten
- location_of
- contracted_with
- mentioned_in
- disagreed
- acknowledge
- attack_
- invested_in
- services
- founded_
- terrified
- move_from
- operates_in
- battle
- region_served
- meet
- awarded_
- move_to
- researched
- monetary_earning
- enclose
- separated_
- criticize
- demolished
- embargo
- surrendered
- abuse
- deceive
- identified_
- linked_to
- [[customer_specific]]

OrgToWeapon

OrgToWeapon relationships are those relationships that connect organizations to weapons.

Predicate types

- took
- acquires_
- sells_
- made_

© N. Harris Computer Corporation

- funded
- gave_
- authorized
- invested_in
- sends_
- appears_
- compensated_
- lost
- takes_action
- acknowledge
- owns
- uses
- dispose
- has
- describe
- criticize
- linked_to
- [[customer_specific]]

OrgToDrug

The OrgToDrug relationship represents a connection between an organization and a legal or illicit drug.

Predicate types

- took
- acquires_
- sells_
- researched
- funded
- made_
- authorized
- complaints
- ceased
- invested_in
- alleges
- gave_
- appears_
- branded
- compensated_
- uses
- dispose

© N. Harris Computer Corporation

- has
- describe
- consider
- acknowledge
- advertise
- criticize
- provider_of
- apprehended
- supplied
- [[customer_specific]]

OrgToChemical

The OrgToChemical relationship represents a connection between an organization and a chemical.

Predicate types

- acquires_
- sells_
- uses
- sends_
- appears_
- gave_
- dispose
- has
- describe
- made_
- criticize
- [[customer_specific]]

OrgToConveyance

OrgToConveyance relationships are those relationships that connect an organization to a method of conveyance.

Predicate types

- witnessed
- took
- funded
- discover
- is_evaluating
- authorized
- acquires_
- complaints

© N. Harris Computer Corporation

- dispatched
- ceased
- sells_
- canceled
- made_
- gave_
- manages
- sends_
- acknowledge
- appears_
- branded
- attack_
- owns
- died_from
- uses
- dispose
- has
- describe
- consider
- launches
- criticize
- [[customer_specific]]

OrgToCrime

The OrgToCrime relationship represents relationships between an organization and a criminal offense.

Predicate types

- assists
- compensated_
- charged_with
- funded
- authorized
- commits
- acknowledge
- denied
- involvement
- suspected
- alleges
- prevented_
- threaten

© N. Harris Computer Corporation

- compensated_
- attack_
- absolved_
- plotted_
- participated_in
- penalize_
- involvement
- linked_to
- [[customer_specific]]

OrgToDisease

The OrgToDisease relationship represents a connection between an organization and a disease.

Predicate types

- researched
- expert_on
- gave_
- describe
- accused
- communicated_
- criticize
- battle
- targets
- treating
- studies
- [[customer_specific]]

OrgToProduct

The OrgToProduct relationship represents a connection between an organization and a commercial product.

Predicate types

- took
- acquires_
- canceled
- authorized
- select
- funded
- alleges
- complaints
- sells_

© N. Harris Computer Corporation

- made_
- ceased
- sends_
- appears_
- guaranteed
- amplify
- announce
- researched
- branded
- acknowledge
- uses
- dispose
- has
- describe
- consider
- launches
- advertise
- criticize
- donate
- supplied
- [[customer_specific]]

OrgToImplement

The OrgToImplement relationship represents relationships between an organization and an implement.

Predicate types

- discover
- belongs_to
- acquires_
- supplied
- made_
- unveil
- expert_on
- apprehended
- sends_
- appears_
- researched
- provider_of
- access_to
- owns

© N. Harris Computer Corporation

- studies
- invested_in
- has
- make_improvements
- funded
- linked_to
- [[customer_specific]]

OrgToFinancial_Index

The OrgToFinancial_Index relationship represents relationships between an organization and a financial index.

Predicate types

- ranks
- tracks
- criticize
- [[customer_specific]]

OrgToTicker_Symbol

The OrgToTicker_Symbol relationship represents relationships between an organization and a ticker symbol.

Predicate types

- uses
- tracks
- has
- [[customer_specific]]

OrgToMedical_Procedure

The OrgToMedical_Procedure relationship represents relationships between an organization and a medical procedure.

Predicate types

- compensated_
- uses
- describe
- consider
- is_evaluating
- criticize
- [[customer_specific]]

OrgToEvent

OrgToEvent relationships represent those relationships that connect an organization to an event.

© N. Harris Computer Corporation

Predicate types

- implements
- authorized
- inform
- assists
- contact
- host
- funded
- attack_
- describe
- depart
- launches
- works_with
- announce
- criticize
- initiated
- involvement
- participated_in
- compensated_
- conclude
- studies
- affected_by
- [[customer_specific]]

OrgToPublication

OrgToPublication relationships represent a connection between an organization and a publication.

Predicate types

- inform
- branded
- contact
- authorized
- owns
- complaints
- alleges
- funded
- mentioned_in
- disagreed
- sends_
- sells_

© N. Harris Computer Corporation

- advertise
- published_in
- criticize
- denied
- submitted
- [[customer_specific]]

OrgToEmail

The OrgToEmail relationship represents a relationship between an organization and an email address.

Predicate types

- belongs_to
- inform
- originated_
- emailed
- sends_
- received
- mentioned_in
- opened
- location_of
- communicated_
- contact
- access_to
- linked_to
- [[customer_specific]]

OrgToSocial

The OrgToSocial relationship represents a connection between an organization and a social media identifier.

Predicate types

- belongs_to
- inform
- contact
- confronted
- complaints
- disagreed
- acknowledge
- criticize
- follows_
- linked_to

© N. Harris Computer Corporation

- [[customer_specific]]

OrgToPhone

The OrgToPhone relationship represents a connection between an organization and a phone number.

Predicate types

- belongs_to
- inform
- communicated_
- identified_by
- originated_
- contact
- uses
- has
- linked_to
- [[customer_specific]]

OrgToURL

OrgToURL relationships are those relationships that represent a connection between an organization and a website or IP address.

Predicate types

- maintains
- belongs_to
- acquires_
- inform
- identified_by
- create
- owns
- launches
- linked_to
- [[customer_specific]]

OrgToGeocoordinate

The OrgToGeocoordinate relationship represents a connection between an organization and a geocoordinate.

Predicate types

- location_of
- [[customer_specific]]

© N. Harris Computer Corporation

OrgToMoney

OrgToMoney relationships are those relationships that represent a connection between an organization and a numeric amount of money.

Predicate types

- took
- belongs_to
- govern_
- is_evaluating
- ceased
- penalize_
- gave_
- manages
- sends_
- donate
- received
- guaranteed
- awarded_
- laundered
- gain_
- price_
- loss
- monetary_earning
- acquires_
- value_of
- funded
- invested_in
- compensated_
- has
- sells_
- deceive
- [[customer_specific]]

OrgToPercent

The OrgToPercent relationship represents a connection between an organization and a percentage.

Predicate types

- gain_
- decrease_to
- owns
- guaranteed

© N. Harris Computer Corporation

- announce
- [[customer_specific]]

OrgToNationality

The OrgToNationality relationship represents a connection between an organization and a nation.

Predicate types

- belongs_to
- works_with
- united_
- communicated_
- manipulates
- region_served
- identified_
- compensated_
- employs
- commits
- kidnapped_
- access_to
- works_
- monetary_earning
- criticize
- battle
- abuse
- deceive
- linked_to
- [[customer_specific]]

OrgToProfession

The OrgToProfession relationship represents a connection between an organization and a profession.

Predicate types

- select
- united_
- contribute
- appears_
- works_with
- announce
- trusts
- kidnapped_
- negotiated_with

© N. Harris Computer Corporation

- employs
- contact
- consider
- formed
- works_
- depart
- criticize
- [[customer_specific]]

OrgToMeasure

OrgToMeasure relationships are those relationships that denote a connection between an organization and a numeric measure.

Predicate types

- age_of
- announce
- [[customer_specific]]

OrgToIDNUM

The OrgToIDNUM relationship represents a connection between an organization and an identification number.

Predicate types

- belongs_to
- awarded_
- identified_by
- compensated_
- ein
- has
- sic
- [[customer_specific]]

OrgToCitation

The OrgToCitation relationship represents relationships between organizations and legal citation numbers.

Predicate types

- mentioned_in
- referenced
- [[customer_specific]]

© N. Harris Computer Corporation

OrgToHashtag

The OrgToHashtag relationship represents those relationships that connect an organization to a social media hashtag.

Predicate types

- belongs_to
- referenced
- mentioned_in
- linked_to
- [[customer_specific]]

OrgToTimestamp

The OrgToTimestamp relationship represents a connection between an organization name and a date & time expression.

Predicate types

- founded_
- failed_
- formed
- linked_to
- funded
- [[customer_specific]]

OrgToProgram

The OrgToProgram relationship represents a connection between an organization and a program, project, or plan.

Predicate types

- manages
- invested_in
- acquires_
- works_with
- assists
- compensated_
- negotiated_with
- contact
- describe
- donate
- affected_by
- formed
- depart
- launches

© N. Harris Computer Corporation

- criticize
- denied
- involvement
- funded
- [[customer_specific]]

OrgToPunitive_Measure

OrgToPunitive_Measure relationships represent a connection between an organization and a punitive measure.

Predicate types

- implements
- amount
- mandate
- penalize_
- takes_action
- compensated_
- enforce
- has
- afflicted_with
- affected_by
- criticize
- [[customer_specific]]

OrgToPolitical_Affiliation

OrgToPolitical_Affiliation relationships represent a connection between an organization and a political party.

Predicate types

- govern_
- inform
- select
- penalize_
- sponsored
- amplify
- works_with
- announce
- assists
- compensated_
- negotiated_with
- contact
- depart

© N. Harris Computer Corporation

- criticize
- linked_to
- [[customer_specific]]

OrgToTransit

OrgToTransit relationships represent a connection between an organization and a transit entity.

Predicate types

- closed
- location_of
- [[customer_specific]]

OrgToDNA

OrgToDNA relationships represent a connection between an organization and a DNA entity.

Predicate types

- discover
- expert_on
- invested_in
- targets
- manipulates
- researched
- isolate
- transform
- compensated_
- create
- studies
- uses
- communicated_
- monetary_earning
- [[customer_specific]]

OrgToGene

OrgToGene relationships represent a connection between an organization and a gene entity.

Predicate types

- discover
- made_
- expert_on
- invested_in
- convert
- targets

© N. Harris Computer Corporation

- manipulates
- isolate
- researched
- transform
- compensated_
- create
- studies
- uses
- communicated_
- monetary_earning
- linked_to
- [[customer_specific]]

OrgToIdeology

OrgToIdeology relationships represent a connection between an organization and an ideology.

Predicate types

- compensated_
- owns
- create
- afflicted_with
- conviction_
- formed
- creed
- abuse
- criticize
- [[customer_specific]]

OrgToAward

OrgToAward relationships represent a connection between an organization and an award.

Predicate types

- acquires_
- belongs_to
- awarded_
- received
- sends_
- has
- win
- criticize
- [[customer_specific]]

© N. Harris Computer Corporation

OrgToScore

OrgToScore relationships represent a connection between an organization and a score.

Predicate types

- acquires_
- received
- has
- [[customer_specific]]

OrgToRating

OrgToRating relationships represent a connection between an organization and a rating.

Predicate types

- acquires_
- received
- has
- [[customer_specific]]

OrgToGeneric

OrgToGeneric relationships represent a connection between an organization and a generic entity.

Predicate types

- [[customer_specific]]

FacilityToFacility

FacilityToFacility relationships are those relationships that connect a facility with another facility.

Predicate types

- move_from
- move_to
- location_of
- competes_with
- merger
- received_information
- [[customer_specific]]

FacilityToAddress

FacilityToAddress relationships are those relationships that connect a facility with a physical address.

Predicate types

- move_from
- move_to
- location_of
- operates_in

© N. Harris Computer Corporation

- [[customer_specific]]

FacilityToPlace

The FacilityToPlace relationship represents a connection between a facility and a place.

Predicate types

- move_from
- originated_
- move_to
- location_of
- operates_in
- monetary_earning
- contracted_with
- identified_
- linked_to
- [[customer_specific]]

FacilityToWeapon

FacilityToWeapon relationships are those relationships that connect facilities to weapons.

Predicate types

- acquires_
- sells_
- location_of
- uses
- appears_
- made_
- [[customer_specific]]

FacilityToDrug

The FacilityToDrug relationship represents a connection between a facility and a legal or illicit drug.

Predicate types

- acquires_
- sells_
- supplies
- location_of
- uses
- sends_
- appears_
- made_
- [[customer_specific]]

© N. Harris Computer Corporation

FacilityToChemical

The FacilityToChemical relationship represents a connection between a facility and a chemical.

Predicate types

- acquires_
- sells_
- supplies
- location_of
- uses
- sends_
- appears_
- made_
- [[customer_specific]]

FacilityToConveyance

FacilityToConveyance relationships are those relationships that connect a facility to a method of conveyance.

Predicate types

- supplies
- location_of
- sends_
- canceled
- appears_
- made_
- launches
- donate
- [[customer_specific]]

FacilityToCrime

The FacilityToCrime relationship represents relationships between a named facility and a criminal offense.

Predicate types

- manages
- uses
- location_of
- denied
- enforce
- penalize_
- threaten
- witnessed

© N. Harris Computer Corporation

- accused
- linked_to
- [[customer_specific]]

FacilityToDisease

The FacilityToDisease relationship represents a connection between a facility and a disease.

Predicate types

- linked_to
- researched
- caused
- expert_on
- gave_
- describe
- battle
- accused
- originated_
- communicated_
- criticize
- targets
- treating
- [[customer_specific]]

FacilityToImplement

The FacilityToImplement relationship represents a connection between a facility and an implement.

Predicate types

- sells_
- unveil
- create
- uses
- sponsored
- appears_
- has
- made_
- linked_to
- [[customer_specific]]

FacilityToProduct

The FacilityToProduct relationship represents a connection between a facility and a commercial product.

Predicate types

© N. Harris Computer Corporation

- acquires_
- sells_
- supplies
- canceled
- location_of
- uses
- sends_
- appears_
- has
- made_
- donate
- [[customer_specific]]

FacilityToMedical_Procedure

The FacilityToMedical_Procedure relationship represents relationships between a facility and a medical procedure.

Predicate types

- location_of
- linked_to
- perform
- [[customer_specific]]

FacilityToEvent

FacilityToEvent relationships represent those relationships that connect a facility to a named event.

Predicate types

- move_from
- move_to
- host
- funded
- location_of
- affected_by
- depart
- is_evaluating
- launches
- works_with
- announce
- initiated
- involvement
- [[customer_specific]]

© N. Harris Computer Corporation

FacilityToPublication

FacilityToPublication relationships are those relationships that connect a facility with a publication.

Predicate types

- identified_by
- mentioned_in
- sends_
- published_in
- submitted
- [[customer_specific]]

FacilityToEmail

The FacilityToEmail relationship represents a relationship between a facility and an email address.

Predicate types

- originated_
- emailed
- mentioned_in
- location_of
- [[customer_specific]]

FacilityToSocial

The FacilityToSocial relationship represents a connection between a facility and a social media identifier.

Predicate types

- mentioned_in
- [[customer_specific]]

FacilityToPhone

The FacilityToPhone relationship represents a connection between a facility and a phone number.

Predicate types

- belongs_to
- communicated_
- identified_by
- originated_
- location_of
- has
- linked_to
- [[customer_specific]]

© N. Harris Computer Corporation

FacilityToURL

FacilityToURL relationships are those relationships that represent a connection between a facility and a website or IP address.

Predicate types

- belongs_to
- linked_to
- [[customer_specific]]

FacilityToGeocoordinate

The FacilityToGeocoordinate relationship represents relationships between a facility and a geocoordinate.

Predicate types

- location_of
- [[customer_specific]]

FacilityToMoney

FacilityToMoney relationships are those relationships that connect a facility to a numeric amount of money.

Predicate types

- loss
- acquires_
- penalize_
- lost
- compensated_
- received
- appears_
- has
- gain_
- donate
- [[customer_specific]]

FacilityToPercent

The FacilityToPercent relationship represents a relationship between a facility and a percentage.

Predicate types

- [[customer_specific]]

FacilityToNationality

The FacilityToNationality relationship represents a relationship between a facility and a nationality.

Predicate types

© N. Harris Computer Corporation

- discover
- belongs_to
- acquires_
- damaged
- works_
- operates_in
- made_
- attack_
- has
- takes_control
- linked_to
- [[customer_specific]]

FacilityToProfession

The FacilityToProfession relationship represents a relationship between a facility and a profession.

Predicate types

- works_
- [[customer_specific]]

FacilityToIdeology

The FacilityToIdeology relationship represents a relationship between a facility and an ideology.

Predicate types

- move_from
- acquires_
- move_to
- demolished
- made_
- location_of
- attack_
- opened
- [[customer_specific]]

FacilityToMeasure

The FacilityToMeasure relationship represents a relationship between a facility and a numeric measure.

Predicate types

- age_of
- height_of
- [[customer_specific]]

© N. Harris Computer Corporation

FacilityToIDNUM

The FacilityToIDNUM relationship represents a connection between a facility and an identification number.

Predicate types

- awarded_
- identified_by
- compensated_
- ein
- [[customer_specific]]

FacilityToCitation

The FacilityToCitation relationship represents a connection between a facility and a legal citation number.

Predicate types

- mentioned_in
- referenced
- [[customer_specific]]

FacilityToHashtag

The FacilityToHashtag relationship represents those relationships that connect a facility to a social media hashtag.

Predicate types

- belongs_to
- linked_to
- referenced
- [[customer_specific]]

FacilityToTimestamp

The FacilityToTimestamp relationship represents a connection between a facility and a date & time expression.

Predicate types

- linked_to
- [[customer_specific]]

FacilityToProgram

The FacilityToProgram relationship represents a connection between a facility and a program, project, or plan.

Predicate types

- belongs_to
- acquires_

© N. Harris Computer Corporation

- manages
- location_of
- opened
- move_from
- compensated_
- move_to
- made_
- demolished
- access_to
- owns
- operates_in
- depart
- make_improvements
- linked_to
- [[customer_specific]]

FacilityToPunitive_Measure

FacilityToPunitive_Measure relationships are those relationships that connect a facility with a punitive measure.

Predicate types

- mandate
- affected_by
- [[customer_specific]]

FacilityToPolitical_Affiliation

FacilityToPolitical_Affiliation relationships are those relationships that connect a facility with a political party.

Predicate types

- works_
- belongs_to
- has
- [[customer_specific]]

FacilityToTransit

FacilityToTransit relationships are those relationships that connect a facility with a transit entity.

Predicate types

- location_of
- [[customer_specific]]

© N. Harris Computer Corporation

FacilityToDNA

FacilityToDNA relationships are those relationships that connect a facility with a DNA entity.

Predicate types

- discover
- location_of
- appears_
- [[customer_specific]]

FacilityToGene

FacilityToGene relationships are those relationships that connect a facility with a gene entity.

Predicate types

- discover
- isolate
- transform
- made_
- conclude
- create
- convert
- linked_to
- [[customer_specific]]

FacilityToAward

FacilityToAward relationships are those relationships that connect a facility with an award.

Predicate types

- acquires_
- received
- has
- win
- lost
- awarded_
- [[customer_specific]]

FacilityToScore

FacilityToScore relationships are those relationships that connect a facility with a score.

Predicate types

- acquires_
- received
- has
- [[customer_specific]]

© N. Harris Computer Corporation

AddressToAddress

AddressToAddress relationships are those relationships that connect a physical address with another physical address.

Predicate types

- [[customer_specific]]

AddressToPlace

The AddressToPlace relationship represents a connection between a physical address and a place.

Predicate types

- location_of
- operates_in
- [[customer_specific]]

AddressToGeocoordinate

The AddressToGeocoordinate relationship represents a connection between a physical address and a geocoordinate.

Predicate types

- location_of
- [[customer_specific]]

AddressToPhone

The AddressToPhone relationship represents a connection between a physical address and a phone number.

Predicate types

- belongs_to
- identified_by
- originated_
- location_of
- has
- linked_to
- [[customer_specific]]

AddressToConveyance

The AddressToConveyance relationship represents a connection between a physical address and a conveyance.

Predicate types

- location_of
- sends_
- appears_
- [[customer_specific]]

© N. Harris Computer Corporation

AddressToMoney

The AddressToMoney relationship represents a connection between a physical address and money.

Predicate types

- invested_in
- penalize
- value_of
- sells_
- [[customer_specific]]

AddressToCrime

The AddressToCrime relationship represents a connection between a physical address and a crime.

Predicate types

- location_of
- affected_by
- linked_to
- [[customer_specific]]

AddressToEvent

The AddressToEvent relationship represents a connection between a physical address and an event.

Predicate types

- location_of
- initiated
- [[customer_specific]]

AddressToPublication

The AddressToPublication relationship represents a connection between a physical address and a publication.

Predicate types

- location_of
- sends_
- made_
- [[customer_specific]]

AddressToEmail

The AddressToEmail relationship represents a connection between a physical address and an email address.

Predicate types

- originated_
- location_of

© N. Harris Computer Corporation

- mentioned_in
- [[customer_specific]]

AddressToProgram

The AddressToProgram relationship represents a connection between a physical address and a program, project, or plan.

Predicate types

- move_from
- move_to
- location_of
- linked_to
- [[customer_specific]]

AddressToImplement

The AddressToImplement relationship represents a connection between a physical address and an implement.

Predicate types

- discover
- sends_
- location_of
- linked_to
- made_
- sells_
- [[customer_specific]]

AddressToNationality

The AddressToNationality relationship represents a connection between a physical address and a nationality.

Predicate types

- move_from
- operates_in
- move_to
- location_of
- [[customer_specific]]

AddressToPolitical_Affiliation

The AddressToPolitical_Affiliation relationship represents a connection between a physical address and a political party.

Predicate types

- works_

© N. Harris Computer Corporation

- belongs_to
- location_of
- linked_to
- [[customer_specific]]

AddressToWeapon

The AddressToWeapon relationship represents a connection between a physical address and a weapon.

Predicate types

- discover
- location_of
- sends_
- appears_
- made_
- sells_
- [[customer_specific]]

AddressToDrug

AddressToDrug relationships are those relationships that connect an address with a drug.

Predicate types

- discover
- location_of
- sends_
- appears_
- made_
- sells_
- [[customer_specific]]

AddressToIDNUM

The AddressToIDNUM relationship represents a connection between a physical address and an identification number.

Predicate types

- linked_to
- [[customer_specific]]

AddressToIdeology

The AddressToIdeology relationship represents a connection between a physical address and an ideology.

Predicate types

- move_from

© N. Harris Computer Corporation

- move_to
- location_of
- [[customer_specific]]

PlaceToPlace

The PlaceToPlace relationship represents a connection between a place and another place.

Predicate types

- battle
- belongs_to
- confronted
- govern_
- location_of
- has
- complaints
- competes_with
- surrendered
- alleges
- disagreed
- criticize
- demolished
- embargo
- linked_to
- [[customer_specific]]

PlaceToWeapon

PlaceToWeapon relationships are those relationships that connect places to weapons.

Predicate types

- move_from
- move_to
- location_of
- sends_
- funded
- appears_
- criticize
- [[customer_specific]]

PlaceToDrug

The PlaceToDrug relationship represents a connection between a place and a legal or illicit drug.

Predicate types

© N. Harris Computer Corporation

- move_from
- acquires_
- sells_
- ceased
- funded
- move_to
- compensated_
- location_of
- uses
- sends_
- appears_
- made_
- apprehended
- [[customer_specific]]

PlaceToChemical

The PlaceToChemical relationship represents a connection between a place and a chemical.

Predicate types

- move_from
- move_to
- location_of
- appears_
- [[customer_specific]]

PlaceToConveyance

PlaceToConveyance relationships are those relationships that connect a place to a method of conveyance.

Predicate types

- move_from
- discover
- ceased
- move_to
- funded
- attack_
- made_
- complaints
- location_of
- sends_
- appears_

© N. Harris Computer Corporation

- [[customer_specific]]

PlaceToCrime

The PlaceToCrime relationship represents relationships between a specific place and a criminal offense.

Predicate types

- charged_with
- commits
- accused
- denied
- alleges
- funded
- prevented_
- authorized
- occurs
- participated_in
- involvement
- involvement
- location_of
- penalize_
- absolved_
- plotted_
- linked_to
- [[customer_specific]]

PlaceToImplement

The PlaceToImplement relationship represents a connection between a place and an implement.

Predicate types

- move_from
- move_to
- made_
- location_of
- sends_
- appears_
- [[customer_specific]]

PlaceToDisease

The PlaceToDisease relationship represents a connection between a place and a disease.

Predicate types

© N. Harris Computer Corporation

- travels_
- originated_
- location_of
- made_
- death
- affected_by
- isolate
- variant_of
- researched
- damaged
- expert_on
- describe
- accused
- battle
- targets
- treating
- [[customer_specific]]

PlaceToDNA

The PlaceToDNA relationship represents a connection between a place and DNA.

Predicate types

- belongs_to
- identified_
- originated_
- [[customer_specific]]

PlaceToGene

The PlaceToGene relationship represents a connection between a place and a gene.

Predicate types

- belongs_to
- identified_
- originated_
- [[customer_specific]]

PlaceToProduct

The PlaceToProduct relationship represents a connection between a place and a commercial product.

Predicate types

- move_from
- move_to

© N. Harris Computer Corporation

- lost
- damaged
- made_
- location_of
- received
- sends_
- appears_
- launches
- ceased
- [[customer_specific]]

PlaceToEvent

PlaceToEvent relationships represent those relationships that connect a place to an event.

Predicate types

- location_of
- initiated
- describe
- attended
- funded
- affected_by
- targets
- participated_in
- compensated_
- linked_to
- [[customer_specific]]

PlaceToHashtag

PlaceToHashtag relationships represent those relationships that connect a place to a hashtag.

Predicate types

- mentioned_in
- [[customer_specific]]

PlaceToPublication

PlaceToPublication relationships are those relationships that connect a place with a publication.

Predicate types

- location_of
- sends_
- published_in
- denied

© N. Harris Computer Corporation

- funded
- [[customer_specific]]

PlaceToEmail

The PlaceToEmail relationship represents a relationship between a place and an email address.

Predicate types

- location_of
- [[customer_specific]]

PlaceToScore

The PlaceToScore relationship represents a relationship between a place and a score.

Predicate types

- acquires_
- received
- has
- [[customer_specific]]

PlaceToSocial

The PlaceToSocial relationship represents a connection between a place and a social media identifier.

Predicate types

- [[customer_specific]]

PlaceToPhone

The PlaceToPhone relationship represents a connection between a place and a phone number.

Predicate types

- originated_
- location_of
- [[customer_specific]]

PlaceToURL

PlaceToURL relationships are those relationships that represent a connection between a place and a website or IP address.

Predicate types

- location_of
- [[customer_specific]]

PlaceToGeocoordinate

The PlaceToGeocoordinate relationship represents relationships between a place and a geocoordinate.

Predicate types

© N. Harris Computer Corporation

- location_of
- [[customer_specific]]

PlaceToMoney

PlaceToMoney relationships are those relationships that connect a place to a numeric amount of money.

Predicate types

- location_of
- ceased
- sends_
- funded
- [[customer_specific]]

PlaceToPercent

The PlaceToPercent relationship represents a relationship between a place and a percentage.

Predicate types

- location_of
- [[customer_specific]]

PlaceToMeasure

PlaceToMeasure relationships are those relationships that denote a connection between a place and a numeric measure.

Predicate types

- age_of
- [[customer_specific]]

PlaceToIDNUM

The PlaceToIDNUM relationship represents a connection between a place and an identification number.

Predicate types

- [[customer_specific]]

PlaceToCitation

The PlaceToCitation relationship represents a connection between a place and a legal citation number.

Predicate types

- location_of
- [[customer_specific]]

PlaceToProfession

The PlaceToProfession relationship represents a connection between a place and a profession.

Predicate types

© N. Harris Computer Corporation

- travels_
- location_of
- works_
- kidnapped_
- employs
- [[customer_specific]]

PlaceToNationality

The PlaceToNationality relationship represents a connection between a place and a nationality.

Predicate types

- travels_
- escaped_from
- citizen_of
- battle
- move_from
- region_served
- move_to
- works_with
- affected_by
- kidnapped_
- works_
- communicated_
- employs
- linked_to
- [[customer_specific]]

PlaceToMedical_Procedure

The PlaceToMedical_procedure relationship represents a connection between a place and a medical procedure.

Predicate types

- location_of
- [[customer_specific]]

PlaceToIdeology

The PlaceToIdeology relationship represents a connection between a place and an ideology.

Predicate types

- region_served
- govern_
- operates_in

© N. Harris Computer Corporation

- location_of
- [[customer_specific]]

PlaceToTimestamp

The PlaceToTimestamp relationship represents a connection between a place and a timestamp.

Predicate types

- founded_
- linked_to
- [[customer_specific]]

PlaceToTimespan

The PlaceToTimespan relationship represents a connection between a place and a timespan.

Predicate types

- from_
- [[customer_specific]]

PlaceToTicker_Symbol

The PlaceToTicker_Symbol relationship represents a connection between a place and a ticker symbol.

Predicate types

- operates_in
- location_of
- linked_to
- [[customer_specific]]

PlaceToAlert_Type

The PlaceToAlert_Type relationship represents a connection between a place and a travel alert or warning issued by the State Department.

Predicate types

- [[customer_specific]]

PlaceToImperative

The PlaceToImperative relationship represents a connection between a place and travel instructions from the State Department.

Predicate types

- [[customer_specific]]

PlaceToProgram

The PlaceToProgram relationship represents a connection between a place and a program, project, or plan.

Predicate types

© N. Harris Computer Corporation

- location_of
- operates_in
- denied
- [[customer_specific]]

PlaceToPunitive_Measure

The PlaceToPunitive_Measure relationship represents a connection between a place and a punitive measure.

Predicate types

- affected_by
- [[customer_specific]]

PlaceToPolitical_Affiliation

The PlaceToPolitical_Affiliation relationship represents a connection between a place and a political party.

Predicate types

- linked_to
- location_of
- [[customer_specific]]

PlaceToTransit

The PlaceToTransit relationship represents a connection between a place and a transit entity.

Predicate types

- location_of
- services
- [[customer_specific]]

PhoneToGeocoordinate

The PhoneToGeocoordinate relationship represents a connection between a phone number and a geocoordinate.

Predicate types

- [[customer_specific]]

URLToGeocoordinate

The URLToGeocoordinate relationship represents a connection between a URL and a geocoordinate.

Predicate types

- linked_to
- [[customer_specific]]

© N. Harris Computer Corporation

WeaponToMoney

WeaponToMoney relationships are those relationships that connect a weapon to a numeric amount of money.

Predicate types

- gain_
- loss
- acquires_
- penalize_
- price_
- value_of
- sells_
- [[customer_specific]]

WeaponToPercent

The WeaponToPercent relationship represents a relationship between a weapon and a percentage.

Predicate types

- loss
- sells_
- gain_
- [[customer_specific]]

WeaponToMeasure

WeaponToMeasure relationships are those relationships that denote a connection between a weapon and a numeric measure.

Predicate types

- weight_of
- speed
- [[customer_specific]]

WeaponToConveyance

WeaponToConveyance relationships are those relationships that denote a connection between a weapon and a conveyance.

Predicate types

- appears_
- sells_
- has
- linked_to
- [[customer_specific]]

© N. Harris Computer Corporation

WeaponToEvent

WeaponToEvent relationships are those relationships that denote a connection between a weapon and an event.

Predicate types

- caused
- sends_
- appears_
- affected_by
- [[customer_specific]]

WeaponToIdeology

WeaponToIdeology relationships are those relationships that denote a connection between a weapon and an ideology.

Predicate types

- died_from
- uses
- supplied
- dispose
- [[customer_specific]]

WeaponToNationality

WeaponToNationality relationships are those relationships that denote a connection between a weapon and a nationality.

Predicate types

- supplied
- dispose
- died_from
- uses
- [[customer_specific]]

WeaponToPunitive_Measure

WeaponToPunitive_Measure relationships are those relationships that denote a connection between a weapon and a punitive measure.

Predicate types

- uses
- [[customer_specific]]

WeaponToGeocoordinate

The WeaponToGeocoordinate relationship represents a relationship between a weapon and a geospatial coordinate.

© N. Harris Computer Corporation

Predicate types

- discover
- took
- sells_
- made_
- sends_
- received
- appears_
- lost
- damaged
- uses
- removed_from
- [[customer_specific]]

WeaponToCitation

The WeaponToCitation relationship represents a connection between a weapon and a legal citation number.

Predicate types

- [[customer_specific]]

WeaponToHashtag

The WeaponToHashtag relationship represents those relationships that connect a weapon to a social media hashtag.

Predicate types

- [[customer_specific]]

WeaponToTimestamp

The WeaponToTimestamp relationship represents a connection between a weapon and a date & time expression.

Predicate types

- [[customer_specific]]

WeaponToCrime

The WeaponToCrime relationship represents a connection between a weapon and a crime.

Predicate types

- linked_to
- uses
- [[customer_specific]]

© N. Harris Computer Corporation

WeaponToProfession

The WeaponToProfession relationship represents a connection between a weapon and a profession.

Predicate types

- discover
- belongs_to
- acquires_
- sends_
- invested_in
- made_
- researched
- compensated_
- access_to
- studies
- uses
- has
- linked_to
- [[customer_specific]]

WeaponToProgram

The WeaponToProgram relationship represents those relationships that connect a weapon to a program, project, or plan.

Predicate types

- acquires_
- compensated_
- invested_in
- affected_by
- launches
- [[customer_specific]]

WeaponToIDNUM

The WeaponToIDNUM relationship represents a connection between a weapon and an identification number.

Predicate types

- belongs_to
- identified_by
- has
- linked_to
- [[customer_specific]]

© N. Harris Computer Corporation

WeaponToScore

The WeaponToScore relationship represents those relationships that connect a weapon to a score.

Predicate types

- acquires_
- received
- has
- [[customer_specific]]

WeaponToURL

The WeaponToURL relationship represents those relationships that connect a weapon to a URL.

Predicate types

- sells_
- acquires
- [[customer_specific]]

DrugToDrug

The DrugToDrug relationship represents a connection between a legal or illicit drug and another drug.

Predicate types

- variant_of
- linked_to
- [[customer_specific]]

DrugToProduct

The DrugToProduct relationship represents a connection between a legal or illicit drug and a commercial product.

Predicate types

- linked_to
- [[customer_specific]]

DrugToMedical_Procedure

The DrugToMedical_Procedure relationship represents a connection between a legal or illicit drug and a medical procedure.

Predicate types

- contribute
- treating
- [[customer_specific]]

DrugToDisease

The DrugToDisease relationship represents a connection between a legal or illicit drug and a disease.

© N. Harris Computer Corporation

Predicate types

- battle
- assists
- caused
- improve
- alleges
- affected_by
- damaged
- treating
- attack_
- targets
- uses
- [[customer_specific]]

DrugToCrime

The DrugToCrime relationship represents a connection between a legal or illicit drug and a crime.

Predicate types

- uses
- alleges
- assists
- linked_to
- [[customer_specific]]

DrugToPunitive_Measure

The DrugToPunitive_Measure relationship represents a connection between a legal or illicit drug and a punitive measure.

Predicate types

- uses
- [[customer_specific]]

DrugToGene

The DrugToGene relationship represents a connection between a legal or illicit drug and a gene.

Predicate types

- improve
- [[customer_specific]]

DrugToConveyance

DrugToConveyance relationships are those relationships that connect a legal or illicit drug to a method of conveyance.

Predicate types

© N. Harris Computer Corporation

- dispose
- discover
- linked_to
- [[customer_specific]]

DrugToEvent

DrugToEvent relationships represent those relationships that connect a legal or illicit drug to an event.

Predicate types

- sends_
- treating
- acquires_
- [[customer_specific]]

DrugToPublication

DrugToPublication relationships are those relationships that connect a legal or illicit drug with a publication.

Predicate types

- [[customer_specific]]

DrugToMoney

DrugToMoney relationships are those relationships that connect a legal or illicit drug to a numeric amount of money.

Predicate types

- gain_
- loss
- acquires_
- penalize_
- price_
- value_of
- sells_
- [[customer_specific]]

DrugToPercent

The DrugToPercent relationship represents a relationship between a legal or illicit drug and a percentage.

Predicate types

- gain_
- loss
- sells_
- decrease_to

© N. Harris Computer Corporation

- weight_of
- [[customer_specific]]

DrugToMeasure

DrugToMeasure relationships are those relationships that denote a connection between a legal or illicit drug and a numeric measure.

Predicate types

- amount
- weight_of
- [[customer_specific]]

DrugToCitation

The DrugToCitation relationship represents a connection between a legal or illicit drug and a legal citation number.

Predicate types

- [[customer_specific]]

DrugToHashtag

The DrugToHashtag relationship represents those relationships that connect a legal or illicit drug to a social media hashtag.

Predicate types

- [[customer_specific]]

DrugToTimestamp

The DrugToTimestamp relationship represents a connection between a legal or illicit drug and a date & time expression.

Predicate types

- [[customer_specific]]

DrugToProgram

The DrugToProgram relationship represents a connection between a legal or illicit drug and a program, project, or plan.

Predicate types

- compensated_
- [[customer_specific]]

DrugToIDNUM

The DrugToIDNUM relationship represents a connection between a legal or illicit drug and an identification number.

Predicate types

- linked_to

© N. Harris Computer Corporation

- [[customer_specific]]

DrugToURL

The DrugToURL relationship represents a connection between a legal or illicit drug and a URL.

Predicate types

- sells
- [[customer_specific]]

ChemicalToConveyance

ChemicalToConveyance relationships are those relationships that connect a chemical to a method of conveyance.

Predicate types

- [[customer_specific]]

ChemicalToProduct

The ChemicalToProduct relationship represents a connection between chemical and a commercial product.

Predicate types

- linked_to
- [[customer_specific]]

ChemicalToEvent

ChemicalToEvent relationships represent those relationships that connect a chemical to an event.

Predicate types

- [[customer_specific]]

ChemicalToPublication

ChemicalToPublication relationships are those relationships that connect a chemical with a publication.

Predicate types

- [[customer_specific]]

ChemicalToMoney

ChemicalToMoney relationships are those relationships that connect a chemical to a numeric amount of money.

Predicate types

- price_
- value_of
- sells_
- [[customer_specific]]

© N. Harris Computer Corporation

ChemicalToPercent

The ChemicalToPercent relationship represents a relationship between a chemical and a percentage.

Predicate types

- [[customer_specific]]

ChemicalToMeasure

ChemicalToMeasure relationships are those relationships that denote a connection between a chemical and a numeric measure.

Predicate types

- amount
- weight_of
- [[customer_specific]]

ChemicalToURL

ChemicalToURL relationships are those relationships that denote a connection between a chemical and a URL.

Predicate types

- sells_
- [[customer_specific]]

ChemicalToIDNUM

The ChemicalToIDNUM relationship represents a connection between a chemical and an identification number.

Predicate types

- [[customer_specific]]

ChemicalToCitation

The ChemicalToCitation relationship represents a connection between a chemical and a legal citation number.

Predicate types

- [[customer_specific]]

ChemicalToHashtag

The ChemicalToHashtag relationship represents those relationships that connect a chemical to a social media hashtag.

Predicate types

- [[customer_specific]]

© N. Harris Computer Corporation

ChemicalToTimestamp

The ChemicalToTimestamp relationship represents a connection between a chemical and a date & time expression.

Predicate types

- [[customer_specific]]

ConveyanceToEvent

ConveyanceToEvent relationships represent those relationships that connect a method of conveyance to an event.

Predicate types

- [[customer_specific]]

ConveyanceToPublication

ConveyanceToPublication relationships are those relationships that connect a method of conveyance with a publication.

Predicate types

- consider
- [[customer_specific]]

ConveyanceToMoney

ConveyanceToMoney relationships are those relationships that connect a method of conveyance to a numeric amount of money.

Predicate types

- gain_
- loss
- acquires_
- penalize_
- price_
- value_of
- sells_
- [[customer_specific]]

ConveyanceToPercent

The ConveyanceToPercent relationship represents a relationship between a method of conveyance and a percentage.

Predicate types

- loss
- sells_
- gain_
- [[customer_specific]]

© N. Harris Computer Corporation

ConveyanceToMeasure

ConveyanceToMeasure relationships are those relationships that denote a connection between a method of conveyance and a numeric measure.

Predicate types

- [[customer_specific]]

ConveyanceToIDNUM

The ConveyanceToIDNUM relationship represents a connection between a method of conveyance and an identification number.

Predicate types

- belongs_to
- identified_by
- has
- linked_to
- [[customer_specific]]

ConveyanceToConveyance

The ConveyanceToConveyance relationship represents a connection between a method of conveyance and another conveyance entity.

Predicate types

- dispatched
- [[customer_specific]]

ConveyanceToCrime

The ConveyanceToCrime relationship represents a connection between a method of conveyance and a crime entity.

Predicate types

- suspected
- died_from
- uses
- [[customer_specific]]

ConveyanceToCitation

The ConveyanceToCitation relationship represents a connection between a method of conveyance and a legal citation number.

Predicate types

- [[customer_specific]]

© N. Harris Computer Corporation

ConveyanceToHashtag

The ConveyanceToHashtag relationship represents those relationships that connect a method of conveyance to a social media hashtag.

Predicate types

- [[customer_specific]]

ConveyanceToTimestamp

The ConveyanceToTimestamp relationship represents a connection between a method of conveyance and a date & time expression.

Predicate types

- [[customer_specific]]

ConveyanceToTransit

The ConveyanceToTransit relationship represents a connection between a method of conveyance and a transit entity.

Predicate types

- owns
- location_of
- [[customer_specific]]

ConveyanceToProfession

The ConveyanceToProfession relationship represents a connection between a method of conveyance and a profession.

Predicate types

- linked_to
- [[customer_specific]]

ConveyanceToGeocoordinate

The ConveyanceToGeocoordinate relationship represents a connection between a method of conveyance and a geocoordinate.

Predicate types

- discover
- move_from
- attack_
- move_to
- lost
- move_from
- damaged
- made_
- location_of

© N. Harris Computer Corporation

- sends_
- attack_
- [[customer_specific]]

ConveyanceToNationality

The ConveyanceToNationality relationship represents a connection between a method of conveyance and a nationality.

Predicate types

- acquires_
- made_
- [[customer_specific]]

ConveyanceToScore

The ConveyanceToScore relationship represents a connection between a method of conveyance and a score.

Predicate types

- acquires_
- received
- has
- [[customer_specific]]

ConveyanceToURL

The ConveyanceToURL relationship represents a connection between a method of conveyance and a URL.

Predicate types

- sells_
- acquires_
- [[customer_specific]]

CrimeToTimestamp

The CrimeToTimestamp relationship represents a connection between a crime and a date & time expression.

Predicate types

- date_of
- occurs
- linked_to
- [[customer_specific]]

CrimeToEvent

The CrimeToEvent relationship represents a connection between a crime and an event.

© N. Harris Computer Corporation

Predicate types

- affected_by
- date_of
- occurs
- location_of
- linked_to
- [[customer_specific]]

CrimeToProfession

The CrimeToProfession relationship represents a connection between a crime and a profession.

Predicate types

- supports_
- compensated_
- uses
- participated_in
- affected_by
- apprehended
- witnessed
- authorized
- conviction_
- inquire_
- mandate
- funded
- accused
- threaten
- attack_
- assists
- charged_with
- commits
- denied
- involvement
- suspected
- alleges
- absolved_
- linked_to
- [[customer_specific]]

CrimeToCrime

The CrimeToCrime relationship represents a connection between one crime and another.

© N. Harris Computer Corporation

Predicate types

- affected_by
- caused
- linked_to
- [[customer_specific]]

CrimeToPunitive_Measure

The CrimeToPunitive_Measure relationship represents a connection between a crime and a punitive measure.

Predicate types

- enforce
- conviction_
- linked_to
- [[customer_specific]]

CrimeToPercent

The CrimeToPercent relationship represents a connection between a crime and a percentage.

Predicate types

- gain_
- maintains
- decrease_to
- affected_by
- [[customer_specific]]

CrimeToMoney

The CrimeToMoney relationship represents a connection between a crime and money.

Predicate types

- acquires_
- commits
- funded
- penalize_
- compensated_
- received
- made_
- laundered
- lost
- monetary_earning
- deceive
- linked_to

© N. Harris Computer Corporation

- [[customer_specific]]

CrimeToIdeology

The CrimeToIdeology relationship represents a connection between a crime and an ideology.

Predicate types

- compensated_
- commits
- died_from
- affected_by
- laundered
- involvement
- suspected
- participated_in
- charged_with
- funded
- authorized
- threaten
- plotted_
- linked_to
- [[customer_specific]]

CrimeToNationality

The CrimeToNationality relationship represents a connection between a crime and a nationality.

Predicate types

- govern_
- compensated_
- died_from
- funded
- commits
- has
- charged_with
- affected_by
- laundered
- denied
- involvement
- suspected
- alleges
- conviction_
- participated_in

© N. Harris Computer Corporation

- plotted_
- authorized
- absolved_
- threaten
- [[customer_specific]]

CrimeToIDNUM

The CrimeToIDNUM relationship represents a connection between a crime and an identification number.

Predicate types

- belongs_to
- identified_by
- has
- linked_to
- [[customer_specific]]

CrimeToSocial

The CrimeToSocial relationship represents a connection between a crime and a social media identifier.

Predicate types

- communicated_
- acknowledge
- alleges
- accused
- mentioned_in
- caused
- referenced
- linked_to
- [[customer_specific]]

CrimeToImplement

The CrimeToImplement relationship represents a connection between a crime and an implement.

Predicate types

- uses
- linked_to
- [[customer_specific]]

IDNUMToTimestamp

The IDNUMToTimestamp relationship represents relationships between an identification number and a timestamp.

Predicate types

© N. Harris Computer Corporation

- identified_by
- issued_
- expiration_date
- linked_to
- [[customer_specific]]

IDNUMToImplement

The IDNUMToImplement relationship represents relationships between an identification number and an implement.

Predicate types

- identified_by
- has
- [[customer_specific]]

IDNUMToProduct

The IDNUMToProduct relationship represents relationships between an identification number and a product.

Predicate types

- has
- identified_by
- belongs_to
- [[customer_specific]]

MeasureToPercent

The MeasureToPercent relationship represents relationships between a measure and a percentage.

Predicate types

- has
- [[customer_specific]]

MeasureToTimestamp

The MeasureToTimestamp relationship represents relationships between a measure and a timestamp.

Predicate types

- date_of
- has
- [[customer_specific]]

DiseaseToGene

The DiseaseToGene relationship denotes relationships between diseases and gene entities.

Predicate types

- caused

© N. Harris Computer Corporation

- damaged
- targets
- attack_
- [[customer_specific]]

DiseaseToDNA

The DiseaseToDNA relationship denotes relationships between diseases and DNA.

Predicate types

- caused
- like
- damaged
- targets
- attack_
- [[customer_specific]]

DiseaseToDisease

The DiseaseToDisease relationship denotes relationships between one disease and another.

Predicate types

- caused
- variant_of
- like
- linked_to
- originated_
- [[customer_specific]]

DiseaseToProduct

The DiseaseToProduct relationship denotes relationships between diseases and commercial products.

Predicate types

- linked_to
- [[customer_specific]]

DiseaseToImplement

The DiseaseToImplement relationship denotes relationships between diseases and an implement.

Predicate types

- perform
- uses
- has
- affected_by
- linked_to

© N. Harris Computer Corporation

- isolate
- [[customer_specific]]

DiseaseToProfession

The DiseaseToProfession relationship denotes relationships between diseases and a profession.

Predicate types

- researched
- works_with
- studies
- has
- linked_to
- expert_on
- gave_
- afflicted_with
- describe
- acknowledge
- affected_by
- communicated_
- died_from
- criticize
- suspected
- isolate
- originated_
- variant_of
- treating
- [[customer_specific]]

DiseaseToEvent

DiseaseToEvent relationships represent those relationships that connect a disease to an event.

Predicate types

- like
- formed
- caused
- gave_
- originated_
- affected_by
- formed
- linked_to
- [[customer_specific]]

© N. Harris Computer Corporation

DiseaseToPublication

DiseaseToPublication relationships are those relationships that connect a disease with a publication.

Predicate types

- communicated_
- researched
- studies
- subject_of
- alleges
- criticize
- denied
- referenced
- appears_
- acknowledge
- [[customer_specific]]

DiseaseToGeocoordinate

The DiseaseToGeocoordinate relationship represents relationships between a disease and a geocoordinate.

Predicate types

- travels_
- formed
- originated_
- location_of
- [[customer_specific]]

DiseaseToMoney

DiseaseToMoney relationships are those relationships that connect a disease to a numeric amount of money.

Predicate types

- price_
- [[customer_specific]]

DiseaseToMedical_Procedure

DiseaseToMedical_Procedure relationships are those relationships that connect a disease to a medical procedure.

Predicate types

- create
- uses
- has

© N. Harris Computer Corporation

- death
- affected_by
- treating
- linked_to
- [[customer_specific]]

DiseaseToPercent

The DiseaseToPercent relationship represents a relationship between a disease and a percentage.

Predicate types

- has
- affected_by
- linked_to
- [[customer_specific]]

DiseaseToMeasure

The DiseaseToMeasure relationship represents a relationship between a disease and a measure.

Predicate types

- weight_of
- [[customer_specific]]

DiseaseToCitation

The DiseaseToCitation relationship represents a connection between a disease and a legal citation number.

Predicate types

- [[customer_specific]]

DiseaseToHashtag

The DiseaseToHashtag relationship represents those relationships that connect a disease to a social media hashtag.

Predicate types

- [[customer_specific]]

DiseaseToTimestamp

The DiseaseToTimestamp relationship represents a connection between a disease and a date & time expression.

Predicate types

- discover
- occurs
- [[customer_specific]]

© N. Harris Computer Corporation

ProductToEvent

ProductToEvent relationships represent those relationships that connect a commercial product to an event.

Predicate types

- unveil
- launches
- ceased
- [[customer_specific]]

ProductToGeocoordinate

ProductToGeocoordinate relationships are those relationships that connect a commercial product with a geocoordinate.

Predicate types

- discover
- move_from
- move_to
- lost
- damaged
- made_
- location_of
- sends_
- received
- [[customer_specific]]

ProductToPublication

ProductToPublication relationships are those relationships that connect a commercial product with a publication.

Predicate types

- select
- mentioned_in
- describe
- consider
- published_in
- [[customer_specific]]

ProductToPhone

ProductToPhone relationships are those relationships that represent a connection between a product and a phone number.

Predicate types

- [[customer_specific]]

© N. Harris Computer Corporation

ProductToConveyance

ProductToConveyance relationships are those relationships that represent a connection between a product and a conveyance.

Predicate types

- [[customer_specific]]

ProductToURL

ProductToURL relationships are those relationships that represent a connection between a product and a website or IP address.

Predicate types

- sells_
- linked_to
- [[customer_specific]]

ProductToMoney

ProductToMoney relationships are those relationships that connect a commercial product to a numeric amount of money.

Predicate types

- gain_
- loss
- acquires_
- monetary_earning
- penalize_
- price_
- made_
- value_of
- sells_
- [[customer_specific]]

ProductToPercent

The ProductToPercent relationship represents a relationship between a commercial product and a percentage.

Predicate types

- gain_
- loss
- monetary_earning
- decrease_to
- [[customer_specific]]

© N. Harris Computer Corporation

ProductToMeasure

The ProductToMeasure relationship represents a relationship between a commercial product and a numeric measure.

Predicate types

- age_of
- weight_of
- sells_
- height_of
- price_
- [[customer_specific]]

ProductToCitation

The ProductToCitation relationship represents a connection between a commercial product and a legal citation number.

Predicate types

- [[customer_specific]]

ProductToHashtag

The ProductToHashtag relationship represents those relationships that connect a commercial product to a social media hashtag.

Predicate types

- mentioned_in
- referenced
- [[customer_specific]]

ProductToTimestamp

The ProductToTimestamp relationship represents a connection between a commercial product and a date & time expression.

Predicate types

- ceased
- launches
- [[customer_specific]]

ProductToProduct

The ProductToProduct relationship represents a connection between a commercial product and another product.

Predicate types

- linked_to
- [[customer_specific]]

© N. Harris Computer Corporation

ProductToProgram

The ProductToProgram relationship represents a connection between a commercial product and a program, project, or plan.

Predicate types

- acquires_
- made_
- [[customer_specific]]

ProductToProfession

The ProductToProfession relationship represents a connection between a commercial product and a profession.

Predicate types

- acquires_
- made_
- contribute
- describe
- [[customer_specific]]

ProductToAward

The ProductToAward relationship represents a connection between a commercial product and an award.

Predicate types

- acquires_
- received
- has
- win
- [[customer_specific]]

ProductToScore

The ProductToScore relationship represents a connection between a commercial product and a score.

Predicate types

- acquires_
- received
- has
- [[customer_specific]]

ProductToIdeology

The ProductToIdeology relationship represents a connection between a commercial product and an ideology.

Predicate types

© N. Harris Computer Corporation

- belongs_to
- acquires_
- identified_by
- compensated_
- access_to
- [[customer_specific]]

EventToPublication

EventToPublication relationships are those relationships that connect an event with a publication.

Predicate types

- visited_
- contact
- mentioned_in
- published_in
- [[customer_specific]]

EventToEvent

EventToEvent relationships are those relationships that connect one event with another.

Predicate types

- linked_to
- contact
- [[customer_specific]]

EventToNationality

EventToNationality relationships are those relationships that connect an event with a nationality.

Predicate types

- works_with
- subject_of
- compensated_
- supports_
- employs
- access_to
- participated_in
- attended
- works_
- monetary_earning
- initiated
- involvement
- travels_

© N. Harris Computer Corporation

- funded
- discredit
- dissent
- contribute
- visited_
- died_from
- witnessed
- sponsored
- attack_
- criticize
- initiated
- linked_to
- [[customer_specific]]

EventToEmail

The EventToEmail relationship represents the relationship between an event and an email address.

Predicate types

- belongs_to
- emailed
- contact
- mentioned_in
- linked_to
- [[customer_specific]]

EventToSocial

EventToSocial relationships are those relationships that connect an event to a social media identifier.

Predicate types

- belongs_to
- emailed
- contact
- mentioned_in
- linked_to
- [[customer_specific]]

EventToPhone

EventToPhone relationships are those relationships that connect an event to a phone number.

Predicate types

- [[customer_specific]]

© N. Harris Computer Corporation

EventToURL

EventToURL relationships are those relationships that represent a connection between an event and a website or IP address.

Predicate types

- [[customer_specific]]

EventToGeocoordinate

The EventToGeocoordinate relationship represents relationships between an event and a geocoordinate.

Predicate types

- [[customer_specific]]

EventToMoney

EventToMoney relationships are those relationships that connect an event to a numeric amount of money.

Predicate types

- loss
- amount
- acquires_
- penalize_
- made_
- price_
- sends_
- received
- [[customer_specific]]

EventToPercent

The EventToPercent relationship represents a relationship between an event and a percentage.

Predicate types

- [[customer_specific]]

EventToMeasure

The EventToMeasure relationship represents a relationship between an event and a numeric measure.

Predicate types

- [[customer_specific]]

EventToCitation

The EventToCitation relationship represents a connection between an event and a legal citation number.

Predicate types

© N. Harris Computer Corporation

- [[customer_specific]]

EventToHashtag

The EventToHashtag relationship represents those relationships that connect an event to a social media hashtag.

Predicate types

- referenced
- mentioned_in
- [[customer_specific]]

EventToTimestamp

The EventToTimestamp relationship represents a connection between an event and a date & time expression.

Predicate types

- date_of
- occurs
- targets
- linked_to
- [[customer_specific]]

EventToTimespan

The EventToTimespan relationship represents a connection between an event and a duration of time.

Predicate types

- date_of
- occurs
- targets
- linked_to
- [[customer_specific]]

EventToIdeology

The EventToIdeology relationship represents a connection between an event and an ideology.

Predicate types

- compensated_
- host
- invested_in
- sponsored
- participated_in
- formed
- launches
- works_

© N. Harris Computer Corporation

- works_with
- supports_
- initiated
- involvement
- funded
- discredit
- dissent
- contribute
- attack_
- criticize
- initiated
- linked_to
- [[customer_specific]]

EventToProgram

The EventToProgram relationship represents a connection between an event and a program, project, or plan.

Predicate types

- compensated_
- contact
- host
- depart
- launches
- [[customer_specific]]

EventToPunitive_Measure

The EventToPunitive_Measure relationship represents a connection between an event and a punitive measure.

Predicate types

- affected_by
- [[customer_specific]]

EventToPolitical_Affiliation

The EventToPolitical_Affiliation relationship represents a connection between an event and a political party.

Predicate types

- compensated_
- contact
- host
- sponsored

© N. Harris Computer Corporation

- participated_in
- depart
- [[customer_specific]]

EventToTransit

The EventToTransit relationship represents a connection between an event and a transit entity.

Predicate types

- location_of
- sends_
- [[customer_specific]]

EventToAward

The EventToAward relationship represents a connection between an event and an award.

Predicate types

- acquires_
- received
- has
- win
- [[customer_specific]]

EventToProfession

The EventToProfession relationship represents a connection between an event and a profession.

Predicate types

- visited_
- implements
- works_
- speaks_at
- contact
- attended
- contribute
- charged_with
- depart
- communicated_
- linked_to
- [[customer_specific]]

EmailToEmail

The EmailToEmail relationship represents the relationship between one email address and another.

Predicate types

© N. Harris Computer Corporation

- originated_
- sends_
- opened
- received
- location_of
- communicated_
- [[customer_specific]]

EmailToIDNUM

The EmailToIDNUM relationship represents the relationship between an email address and an identification number.

Predicate types

- belongs_to
- identified_by
- has
- [[customer_specific]]

EmailToPhone

The EmailToPhone relationship represents the relationship between an email address and a phone number.

Predicate types

- mentioned_in
- [[customer_specific]]

EmailToTimestamp

The EmailToTimestamp relationship represents the relationship between an email address and a timestamp.

Predicate types

- belongs_to
- originated_
- emailed
- opened
- received
- location_of
- has
- date_of
- [[customer_specific]]

© N. Harris Computer Corporation

EmailToGeocoordinate

The EmailToGeocoordinate relationship represents the relationship between an email address and a geocoordinate.

Predicate types

- originated_
- mentioned_in
- location_of
- [[customer_specific]]

EmailToMoney

The EmailToMoney relationship represents the relationship between an email address and money.

Predicate types

- mentioned_in
- [[customer_specific]]

EmailToDrug

The EmailToDrug relationship represents the relationship between an email address and a drug.

Predicate types

- mentioned_in
- emailed
- [[customer_specific]]

EmailToDisease

The EmailToDisease relationship represents the relationship between an email address and a disease.

Predicate types

- mentioned_in
- [[customer_specific]]

EmailToCrime

The EmailToCrime relationship represents the relationship between an email address and a crime.

Predicate types

- emailed
- mentioned_in
- linked_to
- [[customer_specific]]

EmailToWeapon

The EmailToWeapon relationship represents the relationship between an email address and a weapon.

Predicate types

© N. Harris Computer Corporation

- mentioned_in
- emailed
- [[customer_specific]]

EmailToProduct

The EmailToProduct relationship represents the relationship between an email address and a product.

Predicate types

- emailed
- linked_to
- mentioned_in
- [[customer_specific]]

EmailToImplement

The EmailToImplement relationship represents the relationship between an email address and an implement.

Predicate types

- mentioned_in
- emailed
- linked_to
- [[customer_specific]]

PublicationToEmail

The PublicationToEmail relationship represents the relationship between a publication and an email address.

Predicate types

- [[customer_specific]]

PublicationToSocial

The PublicationToSocial relationship represents a connection between a publication and a social media identifier.

Predicate types

- mentioned_in
- communicated_
- [[customer_specific]]

PublicationToPublication

The PublicationToPublication relationship represents a connection between one publication and another.

Predicate types

- mentioned_in

© N. Harris Computer Corporation

- published_in
- submitted
- linked_to
- [[customer_specific]]

PublicationToProfession

The PublicationToProfession relationship represents a connection between a publication and a profession.

Predicate types

- inform
- contact
- mentioned_in
- describe
- consider
- communicated_
- submitted
- [[customer_specific]]

PublicationToPhone

PublicationToPhone relationships are those relationships that connect a publication to a phone number.

Predicate types

- [[customer_specific]]

PublicationToURL

PublicationToURL relationships are those relationships that represent a connection between a publication and a website or IP address.

Predicate types

- maintains
- belongs_to
- inform
- owns
- [[customer_specific]]

PublicationToMoney

PublicationToMoney relationships are those relationships that connect a publication to a numeric amount of money.

Predicate types

- loss
- acquires_
- price_

© N. Harris Computer Corporation

- made_
- value_of
- sells_
- received
- [[customer_specific]]

PublicationToPercent

The PublicationToPercent relationship represents a relationship between a publication and a percentage.

Predicate types

- announce
- [[customer_specific]]

PublicationToMeasure

The PublicationToMeasure relationship represents a relationship between a publication and a numeric measure.

Predicate types

- age_of
- price_
- [[customer_specific]]

PublicationToCitation

The PublicationToCitation relationship represents a connection between a publication and a legal citation number.

Predicate types

- referenced
- has
- [[customer_specific]]

PublicationToMedical_Procedure

The PublicationToMedical_Procedure relationship represents a connection between a publication and a medical procedure.

Predicate types

- mentioned_in
- consider
- [[customer_specific]]

PublicationToHashtag

The PublicationToHashtag relationship represents those relationships that connect a publication to a social media hashtag.

Predicate types

© N. Harris Computer Corporation

- [[customer_specific]]

PublicationToTimestamp

The PublicationToTimestamp relationship represents a connection between a publication and a date & time expression.

Predicate types

- issued_
- published_in
- date_of
- [[customer_specific]]

PublicationToProgram

The PublicationToProgram relationship represents a connection between a publication and a program, project, or plan.

Predicate types

- inform
- referenced
- mentioned_in
- describe
- published_in
- [[customer_specific]]

PublicationToPolitical_Affiliation

The PublicationToPolitical_Affiliation relationship represents a connection between a publication and a political party.

Predicate types

- inform
- subject_of
- mentioned_in
- published_in
- [[customer_specific]]

PublicationToGene

The PublicationToGene relationship represents a connection between a publication and a gene entity.

Predicate types

- discover
- researched
- identified_by
- conclude
- studies

© N. Harris Computer Corporation

- linked_to
- [[customer_specific]]

PublicationToAward

The PublicationToAward relationship represents a connection between a publication and an award.

Predicate types

- acquires_
- received
- has
- win
- [[customer_specific]]

PublicationToIdeology

The PublicationToIdeology relationship represents a connection between a publication and an ideology.

Predicate types

- discover
- author_of
- studies
- describe
- communicated_
- linked_to
- [[customer_specific]]

RatingToPercent

The RatingToPercent relationship represents a relationship between a financial rating and a percentage.

Predicate types

- loss
- sells_
- gain_
- [[customer_specific]]

Financial_IndexToPercent

The Financial_IndexToPercent relationship represents a relationship between a financial index or market and a percentage.

Predicate types

- gain_
- loss
- sells_
- decrease_to

© N. Harris Computer Corporation

- compensated_
- announce
- [[customer_specific]]

Financial_IndexToTicker_Symbol

The Financial_IndexToTicker_Symbol relationship represents a relationship between a financial index or market and a ticker symbol.

Predicate types

- gain_
- loss
- decrease_to
- has
- [[customer_specific]]

Ticker_SymbolToPercent

The Ticker_SymbolToPercent relationship represents a relationship between a ticker symbol and a percentage.

Predicate types

- loss
- sells_
- gain_
- [[customer_specific]]

ProfessionToPunitive_Measure

The ProfessionToPunitive_Measure relationship represents a relationship between a profession and a punitive measure.

Predicate types

- takes_action
- implements
- [[customer_specific]]

ProfessionToProfession

The ProfessionToProfession relationship represents a relationship between one profession and another.

Predicate types

- works_with
- inform
- manages
- convert
- announce
- acquires_

© N. Harris Computer Corporation

- identified_by
- trusts
- negotiated_with
- advise
- conviction_
- contact
- instructs
- has
- consider
- works_
- removed_from
- linked_to
- [[customer_specific]]

ProfessionToMoney

The ProfessionToMoney relationship represents a relationship between a profession and money.

Predicate types

- penalize_
- made_
- sends_
- received
- monetary_earning
- laundered
- gain_
- loss
- acquires_
- lost
- value_of
- compensated_
- access_to
- invested_in
- has
- [[customer_specific]]

ProfessionToEmail

The ProfessionToEmail relationship represents a relationship between a profession and an email address.

Predicate types

- contact

© N. Harris Computer Corporation

- inform
- [[customer_specific]]

ProfessionToSocial

The ProfessionToSocial relationship represents a relationship between a profession and a social media identifier.

Predicate types

- contact
- inform
- [[customer_specific]]

ProfessionToPhone

The ProfessionToPhone relationship represents a relationship between a profession and a phone number.

Predicate types

- [[customer_specific]]

ProfessionToURL

The ProfessionToURL relationship represents a relationship between a profession and a URL.

Predicate types

- contact
- inform
- [[customer_specific]]

ProfessionToPercent

The ProfessionToPercent relationship represents a relationship between a profession and a percentage.

Predicate types

- [[customer_specific]]

ProfessionToMeasure

The ProfessionToMeasure relationship represents a relationship between a profession and a measure.

Predicate types

- [[customer_specific]]

ProfessionToCitation

The ProfessionToCitation relationship represents a relationship between a profession and a citation.

Predicate types

- [[customer_specific]]

© N. Harris Computer Corporation

ProfessionToHashtag

The ProfessionToHashtag relationship represents a relationship between a profession and a hashtag.

Predicate types

- [[customer_specific]]

ProfessionToTimestamp

The ProfessionToTimestamp relationship represents a relationship between a profession and a timestamp.

Predicate types

- [[customer_specific]]

ProfessionToAddress

The ProfessionToAddress relationship represents a relationship between a profession and an address.

Predicate types

- [[customer_specific]]

ProfessionToPolitical_Affiliation

The ProfessionToPolitical_Affiliation relationship represents a relationship between a profession and a political party.

Predicate types

- belongs_to
- inform
- advise
- contact
- instructs
- works_
- depart
- linked_to
- [[customer_specific]]

ProfessionToMedical_Procedure

The ProfessionToMedical_Procedure relationship represents a relationship between a profession and a medical procedure.

Predicate types

- discover
- works_with
- unveil
- perform
- contribute

© N. Harris Computer Corporation

- received
- researched
- instructs
- create
- studies
- uses
- has
- consider
- undergo
- died_from
- linked_to
- [[customer_specific]]

ProfessionToIdeology

The ProfessionToIdeology relationship represents a relationship between a profession and an ideology.

Predicate types

- researched
- belongs_to
- instructs
- works_
- depart
- creed
- linked_to
- [[customer_specific]]

ProfessionToProgram

The ProfessionToProgram relationship represents a relationship between a profession and a program.

Predicate types

- belongs_to
- communicated_
- advise
- manages
- instructs
- has
- compensated_
- works_
- depart
- endorse
- linked_to

© N. Harris Computer Corporation

- [[customer_specific]]

ProfessionToNationality

The ProfessionToNationality relationship represents a relationship between a profession and a nationality.

Predicate types

- works_with
- belongs_to
- united_
- researched
- identified_by
- compensated_
- advise
- studies
- has
- works_
- formed
- communicated_
- monetary_earning
- [[customer_specific]]

NationalityToIdeology

The NationalityToIdeology relationship represents a relationship between a nationality and an ideology.

Predicate types

- belongs_to
- follows_
- compensated_
- advise
- ethnic_group
- formed
- creed
- citizen_of
- abuse
- linked_to
- [[customer_specific]]

NationalityToGene

The NationalityToGene relationship represents a relationship between a nationality and a gene.

Predicate types

© N. Harris Computer Corporation

- belongs_to
- identified_
- identified_by
- ethnic_group
- has
- afflicted_with
- affected_by
- died_from
- linked_to
- [[customer_specific]]

NationalityToDNA

The NationalityToDNA relationship represents a relationship between a nationality and DNA.

Predicate types

- belongs_to
- identified_
- identified_by
- ethnic_group
- has
- afflicted_with
- affected_by
- died_from
- linked_to
- [[customer_specific]]

ProgramToEmail

The ProgramToEmail relationship represents a relationship between a program and an email address.

Predicate types

- emailed
- author_of
- mentioned_in
- linked_to
- [[customer_specific]]

ProgramToSocial

The ProgramToSocial relationship represents a relationship between a program and a social media identifier.

Predicate types

- mentioned_in

© N. Harris Computer Corporation

- [[customer_specific]]

ProgramToPhone

The ProgramToPhone relationship represents a relationship between a program and a phone number.

Predicate types

- belongs_to
- identified_by
- uses
- has
- linked_to
- [[customer_specific]]

ProgramToURL

The ProgramToURL relationship represents a relationship between a program and a URL.

Predicate types

- has
- [[customer_specific]]

ProgramToPercent

The ProgramToPercent relationship represents a relationship between a program and a percentage.

Predicate types

- [[customer_specific]]

ProgramToMeasure

The ProgramToMeasure relationship represents a relationship between a program and a measure.

Predicate types

- age_of
- [[customer_specific]]

ProgramToHashtag

The ProgramToHashtag relationship represents a relationship between a program and a hashtag.

Predicate types

- [[customer_specific]]

ProgramToTimestamp

The ProgramToTimestamp relationship represents a relationship between a program and a timestamp.

Predicate types

- founded_
- date_of

© N. Harris Computer Corporation

- occurs
- [[customer_specific]]

ProgramToMoney

The ProgramToMoney relationship represents a relationship between a program and money.

Predicate types

- acquires_
- price_
- sends_
- [[customer_specific]]

ProgramToAward

The ProgramToAward relationship represents a relationship between a program and an award.

Predicate types

- acquires_
- received
- has
- win
- [[customer_specific]]

ProgramToScore

The ProgramToScore relationship represents a relationship between a program and a score.

Predicate types

- acquires_
- received
- has
- [[customer_specific]]

ProgramToIdeology

The ProgramToIdeology relationship represents a relationship between a program and an ideology.

Predicate types

- maintains
- belongs_to
- compensated_
- advise
- create
- invested_in
- sponsored
- launches

© N. Harris Computer Corporation

- [[customer_specific]]

ScoreToPercent

The ScoreToPercent relationship represents a relationship between a score and a percentage.

Predicate types

- gain_
- loss
- decrease_to
- [[customer_specific]]

TimestampToTransit

The TimestampToTransit relationship represents a relationship between a timestamp and a transit entity.

Predicate types

- founded_
- opened
- [[customer_specific]]

TimestampToIdeology

The TimestampToIdeology relationship represents a relationship between a timestamp and an ideology.

Predicate types

- founded_
- [[customer_specific]]

TimestampToMedical_Procedure

The TimestampToMedical_Procedure relationship represents a relationship between a timestamp and a medical procedure.

Predicate types

- date_of
- occurs
- [[customer_specific]]

TimestampToScore

The TimestampToScore relationship represents a relationship between a timestamp and a score.

Predicate types

- date_of
- [[customer_specific]]

© N. Harris Computer Corporation

TimestampToPunitive_Measure

The TimestampToPunitive_Measure relationship represents a relationship between a timestamp and a punitive measure.

Predicate types

- date_of
- conviction_
- [[customer_specific]]

HashtagToIdeology

The HashtagToIdeology relationship represents a relationship between a hashtag entity and an ideology.

Predicate types

- belongs_to
- referenced
- mentioned_in
- communicated_
- linked_to
- [[customer_specific]]

MoneyToIdeology

The MoneyToIdeology relationship represents a relationship between money and an ideology.

Predicate types

- made_
- lost
- access_to
- funded
- has
- monetary_earning
- laundered
- donate
- [[customer_specific]]

MoneyToImplement

The MoneyToImplement relationship represents a relationship between money and an implement.

Predicate types

- gain_
- funded
- loss
- acquires_

© N. Harris Computer Corporation

- penalize_
- price_
- value_of
- sells_
- [[customer_specific]]

MoneyToMedical_Procedure

The MoneyToMedical_Procedure relationship represents a relationship between money and a medical procedure.

Predicate types

- price_
- funded
- [[customer_specific]]

MoneyToFinancial_Index

The MoneyToFinancial_Index relationship represents a relationship between money and a financial index.

Predicate types

- price_
- is_evaluating
- [[customer_specific]]

MoneyToPercent

The MoneyToPercent relationship represents a relationship between money and a percentage.

Predicate types

- gain_
- loss
- sells_
- decrease_to
- compensated_
- [[customer_specific]]

Punitive_MeasureToIdeology

The Punitive_MeasureToIdeology relationship represents a relationship between a punitive measure and an ideology.

Predicate types

- mandate
- conviction_
- endorse
- [[customer_specific]]

© N. Harris Computer Corporation

Punitive_MeasureToMoney

The Punitive_MeasureToMoney relationship represents a relationship between a punitive measure and money.

Predicate types

- amount
- penalize_
- [[customer_specific]]

SocialToIdeology

The SocialToIdeology relationship represents a relationship between a social media identifier and an ideology.

Predicate types

- belongs_to
- mentioned_in
- uses
- communicated_
- [[customer_specific]]

SocialToTimestamp

The SocialToTimestamp relationship represents a relationship between a social media identifier and a timestamp.

Predicate types

- linked_to
- [[customer_specific]]

SocialToWeapon

The SocialToWeapon relationship represents a relationship between a social media identifier and a weapon.

Predicate types

- sells_
- acquires_
- [[customer_specific]]

Contact_BlockToEmail

The Contact_BlockToEmail relationship represents a relationship between a contact block and an email address.

Predicate types

- has
- [[customer_specific]]

© N. Harris Computer Corporation

Political_AffiliationToIdeology

The Political_AffiliationToIdeology relationship represents a relationship between a political affiliation and an ideology.

Predicate types

- belongs_to
- works_with
- compensated_
- negotiated_with
- participated_in
- [[customer_specific]]

Political_AffiliationToScore

The Political_AffiliationToScore relationship represents a relationship between a political affiliation and a score.

Predicate types

- acquires_
- received
- has
- announce
- [[customer_specific]]

Political_AffiliationToURL

The Political_AffiliationToURL relationship represents a relationship between a political affiliation and a URL.

Predicate types

- maintains
- belongs_to
- owns
- linked_to
- [[customer_specific]]

Political_AffiliationToPercent

The Political_AffiliationToPercent relationship represents a relationship between a political affiliation and a percentage.

Predicate types

- gain_
- acquires_
- decrease_to
- received
- has

© N. Harris Computer Corporation

- announce
- [[customer_specific]]

Political_AffiliationToMoney

The Political_AffiliationToMoney relationship represents a relationship between a political affiliation and money.

Predicate types

- acquires_
- compensated_
- access_to
- received
- uses
- has
- affected_by
- amplify
- linked_to
- [[customer_specific]]

GeneToGene

The GeneToGene relationship represents a relationship between one gene entity and another.

Predicate types

- transform
- like
- made_
- create
- uses
- convert
- targets
- linked_to
- [[customer_specific]]

GeneToMedical_Procedure

The GeneToMedical_Procedure relationship represents a relationship between a gene and a medical procedure.

Predicate types

- transform
- made_
- create
- uses
- convert

© N. Harris Computer Corporation

- targets
- [[customer_specific]]

GeneToDNA

The GeneToDNA relationship represents a relationship between a gene and DNA.

Predicate types

- transform
- made_
- create
- has
- convert
- targets
- [[customer_specific]]

CitationToIdeology

The CitationToIdeology relationship represents a relationship between a citation and an ideology.

Predicate types

- mentioned_in
- referenced
- [[customer_specific]]

TransitToScore

The TransitToScore relationship represents a relationship between a transit entity and a score.

Predicate types

- acquires_
- received
- has
- [[customer_specific]]

TransitToMoney

The TransitToMoney relationship represents a relationship between a transit entity and money.

Predicate types

- amount
- acquires_
- gain_
- lost
- price_
- [[customer_specific]]

© N. Harris Computer Corporation

TransitToMeasure

The TransitToMeasure relationship represents a relationship between a transit entity and a measure.

Predicate types

- [[customer_specific]]

AwardToAward

The AwardToAward relationship represents a relationship between one award and another.

Predicate types

- acquires_
- awarded_
- received
- has
- [[customer_specific]]

ImplementToGeocoordinate

The ImplementToGeocoordinate relationship represents a relationship between an implement and a geocoordinate.

Predicate types

- discover
- move_from
- move_to
- lost
- damaged
- made_
- location_of
- sends_
- received
- appears_
- [[customer_specific]]

BiometricToGeneric

The BiometricToGeneric relationship represents a relationship between a biometric entity and a generic entity.

Predicate types

- linked_to
- [[customer_specific]]

ControlToNon_Salient_Web_Content

The ControlToNon_Salient_Web_Content relationship represents a relationship between a control entity and a non-salient web content entity.

© N. Harris Computer Corporation

Predicate types

- takes_action
- [[customer_specific]]

Sentiment analysis

During extraction, i2 TextChart provides sentiment analysis at the entity and the document level to provide a detailed representation of the sentiment of the text.

This analysis leverages psycholinguistic research across multiple sentiment vectors to provide precise information about an author's language, and pinpoint documents that do not follow predicted patterns. Multi-vector sentiment analysis uses four metrics to help analysts identify outliers in their data, recognize heightened emotional language, sort data by media type, and subset large data sets into only the documents that require further assessment.

```
<lex>
  <word>arson</word>
  <aspect>-2</aspect>
  <intensity>3</intensity>
  <mood>-3</mood>
  <polarity>-3</polarity>
  <sv><CRIME/></sv>
</lex>
```

Aspect

What level of control is a reader likely to feel?

Aspect measures how controlled or in control the language in a text makes a reader feel.

Aspect is measured on a floating point scale ranging from #3 (controlled) to +3 (in control). Very positive or negative aspect values generally indicate that the author is part of the narrative, while neutral aspect values generally indicate that the author is not part of the narrative. Very negative values for aspect tend to indicate that the author is attempting to bully or dominate the reader, while very positive values tend to indicate that the author is attempting to persuade the reader.

For example, words like *abandonment* and *abduction* have a very negative aspect because these words tend to make readers feel a loss of control. By contrast, words like *accomplish* and *motivate* have a very positive aspect because readers tend to feel in control when reading these words.

Intensity

How motivated is a reader likely to be to react?

Intensity is a measurement of the level of activation used to describe a particular entity; in other words, whether an entity is activated/aroused or deactivated/calm. It is measured on a floating point scale ranging from 0 (no activation) to +3 (high activation).

Unlike polarity, mood, and aspect, intensity does not have a negative range.

Words like *catastrophic* or *prestigious* evoke a high level of activation with a value of +3, while words like *thermometer* and *syllabus* evoke a low level of activation with a value of 0.

Mood

How happy or sad is your text?

© N. Harris Computer Corporation

Mood is a measurement of the level of emotion of the language used to describe a particular entity in a document.

This measure roughly indicates the degree of happiness or sadness associated with the language.

While polarity measures the valence of the events themselves, mood measures an emotional response typically elicited in readers or listeners upon hearing the words. This metric is also computed at both the entity and document level, and represented as a floating point value on a scale from #3 (negative mood - "sad") to +3 (positive mood - "happy").

For example, words like *funny* and *affection* have a positive mood, while words like *cancer* and *racist* have a negative mood.

Polarity

How positive or negative is your text?

Polarity is a measurement of how positive, negative, or neutral the language is about a particular entity or the document as a whole, taking into consideration the salience of various words and phrases.

This metric is calculated at both the entity and document level, and is measured on a floating point scale from #3 (negative) to +3 (positive), where 0 represents neutral polarity.

For example, words like *elegant* and *strongest* have positive polarity, while words like *arson* and *lousy* have negative polarity.

Salience

How important is your text?

Salience refers to how important a word or phrase (or entity) is within a document or set of documents.

Entity extraction examples

This topic provides examples of how the i2 TextChart extraction process generates entities from plain text.

1. "i2 Group is located on **Milton Road** in Cambridge."

```
<entity type="ADDRESS">
  <value>Milton Road</value>
  <att name="subtype">physical_address</att>
</entity>
```

2. "**123456 N QD 123456 123456**"

```
<entity type="GEOCOORDINATE">
  <value>123456 N QD 123456 123456</value>
</entity>
```

3. "...which has received more than **$5 billion** in American aid..."

```
<entity type="MONEY">
  <value>$5 billion</value>
  <norm>$5,000,000,000</norm>
</entity>
```

© N. Harris Computer Corporation

Relationship extraction examples

This topic provides examples of how the i2 TextChart extraction process identifies relationships from plain text.

1. "...Imam Samudra claimed **he was beaten by Indonesian police** after his capture."

```
<PSO type="PersonToOrg">
  <predicate>was beaten by</predicate>
  <subject>he</subject>
  <object>indonesian police</object>
</PSO>
```

2. "...**CBS News, reporting on the SARS** outbreak and other stories."

```
<PSO type="OrgToDisease">
  <predicate>, reporting on the</predicate>
  <subject>cbs news</subject>
  <object>sars</object>
</PSO>
```

3. "Admission: $5 and **$8. Ogden Memorial Church**, Main Street."

```
<PSO type="FacilityToMoney">
  <predicate>.</predicate>
  <subject>ogden memorial church</subject>
  <object>$8</object>
</PSO>
```

Predicate types

PSO Type	Definition
absolved_	An entity was set free from/by another entity
abuse	An entity was abused by another entity or committed abuse by or at another entity
access_to	An entity has access to another entity
accused	An entity is accused of or by another entity
acknowledge	An entity acknowledges another entity
acquires_	An entity acquired another entity
advertise	An entity promotes or sponsors another entity
advise	An entity advises, recommends, or warns another entity
affected_by	An entity is affected by another entity
afflicted_with	An entity is afflicted with another entity

© N. Harris Computer Corporation

PSO Type	Definition
against	An entity used another entity against something, or an entity is against another
age_of	The age of an entity
alleges	An entity alleges something about another entity
alias_of	The alias of a person entity
amount	The amount of an entity
amplify	An entity amplified or made another entity stronger
announce	An entity announces or declares intentions for another entity
appears_	An entity appears or emerges from, in, or at another entity
appointed	An entity appointed another entity
apprehended	An entity apprehended another
assists	An entity assists or aids another
at	An entity is located at another
attack_	An entity attacked another
attended	A person entity attended a named event
attests	An entity testifies, or attests to, another entity
author_of	An entity is the author of another entity
authorized	An entity authorized another entity to do something, or to occur
awarded_	An entity has been granted another
ban_	An entity placed an embargo on another entity
battle	An entity fights with another entity
belongs_to	An entity belongs to another entity

© N. Harris Computer Corporation

PSO Type	Definition
birth	The time or place of birth of an entity
branded	An entity has trademarked, labeled, or committed other proprietary action to another entity
burial	The date or place of burial of an entity
canceled	An entity canceled another entity
cared_for	An entity cared for another entity
caused	An entity caused, derived, relates to, or led to another entity
ceased	The time or place that an entity expired or ended
charged_with	An entity was a crime committed by another entity
child_parent_of	A son-of, daughter-of, or parent-of relationship
citizen_of	An entity belongs to a particular nationality
classifies	An organization, document, or other entity has been classified by another
closed	An entity closed an account of another entity
commits	An entity committed an action on another
complaints	An entity has issued complaints about another entity, or received them
communicated_	An entity communicated to, about, or with another entity
compensated_	An entity compensated monetarily or otherwise to another entity
competes_with	An entity competes against another
conclude	The conclusions, results, or findings of one entity on another
confronted	An entity confronted another entity
consider	An entity considers or scrutinizes another entity

© N. Harris Computer Corporation

PSO Type	Definition
contact	An entity contacts another entity
contracted_with	An entity is contracting with another entity
contribute	An entity contributes to another entity
cousin_of	A cousin-of relationship
convert	An entity converts another entity
conviction_	A relationship between two entities involving conviction
creed	An entity has a particular religious belief
create	An entity is created, started, originated, or began with another entity
criticize	An entity criticized another entity
damaged	An entity is damaged or altered in a manner below an ideal state
date_of	The date when an event/entity occurred/ completed an action
death	The place, age, or date of an entity's death
decrease_to	An entity decreased to a percent or another numeric entity
deceive	An entity has deceived or lied to another entity
demolished	An entity demolished another entity
denied	An entity denied another entity, or that another entity took place
depart	An entity leaves, resigns from, or vacates from another entity
describe	An entity describes another entity
died_from	An entity dies from the result of another entity
disagreed	An entity disagreed with or about another entity

© N. Harris Computer Corporation

PSO Type	Definition
discredit	An entity discredited another entity
discover	An entity discovered another entity
dispatched	An entity was dispatched to another entity
dispose	An entity disposes of an entity, or disposes at another entity
dissent	An entity dissented at the location of or about another entity
donate	An entity makes a donation of an entity, or donates to another entity
ein	An employer identification number for businesses
emailed	An entity emailed another entity
embargo	An entity placed an embargo on another entity
enclose	An entity enclosed upon or surrounded another entity
endorse	An entity promotes, endorses, or recommends another entity
enforce	An entity enforces a punishment or enforced its power upon another entity
employs	An entity employs another entity
entered	An entity entered another entity
escaped_from	An entity escaped from another entity
ethnic_group	An entity belongs to an ethnic group
evaluate_	An entity is evaluating a medical procedure or another entity
expert_on	An entity is an expert on another entity
expiration_date	An entity is an expiration date
failed_	An entity ended

© N. Harris Computer Corporation

PSO Type	Definition
familial_	An entity has a familial relationship with another
follows_	An entity is affiliated with, practices with, follows, or associates with another entity
formed	An entity formed another entity
founded_	An entity was founded in or on another entity
from_	An entity is from a place entity
funded	An entity funded another entity
gain_	A financial gain
gave_	An entity gave something to another entity, or gave an entity itself
govern_	An entity governs another entity
grandchild_grandparent_of	A grandparent-of or grandchild-of relationship
greatgrandchild_greatgrandparent_of	A great-grandparent-of or great-grandchild-of relationship
guaranteed	An entity guarantees or secures another entity
has	An entity has another entity
height_of	The height of an entity
hijacked	An entity has been hijacked by or at another entity
host	An entity is hosting another entity
identified_	An entity is identifiable to another entity
identified_by	An entity has an identification number
immunized_	An entity has been vaccinated, immunized, or inoculated against another entity
implements	An entity implements or enforces another entity, or an action upon another entity
improve	An entity improves another entity

© N. Harris Computer Corporation

PSO Type	Definition
incite_	An entity incites another entity
inform	An entity informs, reports to, or explains to another entity
initiated	An entity initiated another entity
inquire_	An entity inquires, examines, or analyzes another entity
influenced	An entity influenced another entity
instructs	An entity lectures, instructs, or addresses another entity
interviewed	An entity interviewed another entity
invade	An entity is invading or intruding upon another entity
invested_in	An entity is invested in another entity
involvement	An entity has involvement with another entity
is_applying	An entity is applying a medical procedure to another entity
is_banker_of	A person or organization entity is the banker of an organization, person, or POI
is_equivalent_to	An entity is a representation of the same real-world entity as another
is_subject_of	A person or organization entity is the subject of a legal filing
is_subscriber_of	A person or organization entity is the subscriber to a service
isolate	An entity isolates another entity
issued_	An entity is an issued another entity
kidnapped_	An entity kidnapped or was kidnapped by another
knows_	An entity knows another entity
launches	An entity has/is launching an event or product

© N. Harris Computer Corporation

PSO Type	Definition
like	An entity is like, similar to, associated with, connected to, or common to another entity
linked_to	An entity is linked to another entity
laundered	An entity laundered another entity
location_of	The place where an organization or facility entity is located
loss	A financial loss or loss in value
lost	An entity lost an entity
made_	An entity made another entity
mandate	An entity mandated another entity
maintains	An entity maintains another entity
make_improvements	An entity makes improvements to another entity
manages	An entity manages another entity
manipulates	An entity is negatively effecting another entity by means of control and deception
meet	An entity meets another entity
mentioned_in	An entity is mentioned in a publication
merger	An entity merges with or acquires another
monetary_earning	An entity has benefited monetarily from another entity
move_from	An entity moves from a place or facility
move_to	An entity moves to a place or facility
negotiated_with	An entity negotiated with another entity
nibling_of	A niece-of or nephew-of relationship
occurs	An entity occurs on or at another entity
opened	An entity opened an account or another entity

© N. Harris Computer Corporation

PSO Type	Definition
operates_in	An entity is active in a particular place
originated_	An entity originated from another
owes_	An entity owes another entity
owns	An entity owns another entity
participated_in	An entity participated in an event
penalize_	An entity penalized another entity
perform	An entity performs another entity
pibling_of	An aunt-of or uncle-of relationship
plotted_	An entity plotted or schemed against another entity
prevented_	An entity stopped another entity
price_	The price of an entity
prohibited_	An entity is prohibited by another entity
provider_of	An entity is a provider of another entity
published_in	The publication date of an entity
pursue	An entity is pursuing another entity
ranks	An entity ranks on a financial index
received	An entity received another entity
received_information	An entity received information from or about another entity
received_legal_assistance	An entity received legal assistance from another entity
recovered_from_	An entity has recovered, is in remission, is cured, or has been healed from another entity
referenced	An entity referred to or used a hashtag

© N. Harris Computer Corporation

PSO Type	Definition
region_served	The location and effect of an entity on a particular place
related_by_marriage	A relationship-in-law, or a step-relationship
related_to	A familial relationship between people
removed_from	A person entity is no longer with an organization
reported_to	An entity reported to another entity
researched	An entity researched another entity
seen	An entity was seen by or at another entity
select	An entity has selected another entity
sells_	An entity sells or sold another entity
sends_	An entity sent another entity, or sent something to another entity
separated_	An entity was separated from an entity, or separated entities from each other
services	The relationship or actions of an entity on a place
sexual_relations	An entity was involved in sexual relations with, by, or at another entity
sibling_of	A brother-of or sister-of relationship
sic	A standard industrial classification code for industry areas
signed	An entity signed another entity
speed	The speed of an entity
sponsored	An entity sponsored another entity
spouse_of	A husband-of or wife-of relationship
subject_of	An entity is the subject of a publication
served_by	An entity has been served by another entity

© N. Harris Computer Corporation

PSO Type	Definition
submitted	An entity submitted an entity, or submitted something to an entity
supplied	An entity supplied another entity
supports_	An entity has supported another entity in some capacity
studies	An entity studied, analyzed, diagnosed, or was processed by another entity
surrendered	An entity gave up or surrendered to another entity
surveil_	An entity is surveilling another entity
survived_	An entity survived or made it through an action caused by another entity
suspected	An entity is suspected by or suspects another entity
takes_action	An entity takes a particular action on another
targets	An entity targets, affects, connects, fuses, or bonds to another entity
terrified	An entity is terrified by another entity
threaten	An entity uses threatening actions against another entity
took	An entity took another entity
tracks	An entity tracks in the financial market
transform	An entity transforms another entity
travels_	Ad entity will travel or move
takes_control	An entity took control of another entity
treating	An entity treated another entity
trusts	An entity trusts or believes in another entity
undergo	An entity undergoes another entity, especially a medical procedure

© N. Harris Computer Corporation

PSO Type	Definition
united_	An entity has united with another entity
unveil	An entity unveils another entity
uses	An entity uses another entity, such as a drug or weapon
value_of	The value of an entity
variant_of	A known spelling variation of a person name
visited_	An entity visited a place entity
weight_of	The weight of an entity
win	An entity wins or triumphs over another entity
witnessed	An entity witnessed another entity
works_	A working relationship between entities
works_with	An entity works with another entity

TextChart SDK

Overview

The TextChart SDK (Software Development Kit) is a comprehensive Java-based toolkit for integrating advanced text analytics and entity extraction capabilities into your applications. The SDK allows you to programmatically process documents, extract entities with their types and relationships, and integrate TextChart's powerful analysis engine into custom applications.

The TextChart SDK builds upon the Rosoka Text Analytics Platform and provides a well-documented API for:

- **Document Processing**: Process files and strings containing unstructured or semi-structured text
- **Entity Extraction**: Automatically identify and classify entities (people, organizations, locations, etc.)
- **Relationship Discovery**: Find connections and relationships between extracted entities
- **Multi-Language Support**: Process text in over 100 languages
- **Customizable Rules**: Extend and customize entity extraction using rule-based configurations
- **Multiple Output Formats**: Get results in XML, JSON, or direct Java object access

Key Components

The TextChart SDK architecture consists of several key components:

© N. Harris Computer Corporation

Core Engine

The Core Engine is the central processing component that coordinates all text analysis operations. It manages the initialization of the system, license validation, and orchestrates the various analysis engines.

Extraction Engine

The extraction engine processes documents to identify and extract entities from text. It applies linguistic rules defined in the LxBase (lexicon) to identify entities and their types.

LxBase (Lexicon Database)

The LxBase is a collection of linguistic rules and patterns that define how entities are identified. It includes:

- Entity type definitions
- Extraction rules (exact match, regex patterns, etc.)
- Rule precedence and conflict resolution
- Support for custom rules and extensions

Output Objects

Results are returned as JAXB-mapped Java objects or XML structures containing:

- Entity lists with types and attributes
- Relationship information
- Confidence scores and metadata
- Document metadata

Supported File Formats

The TextChart SDK can process documents in over 30 file formats including:

- **Office Documents**: .docx, .xlsx, .pptx, .doc, .xls, .ppt
- **PDF**: .pdf
- **Web**: .html, .htm, .xml
- **Text**: .txt, .csv, .tsv
- **Archives**: .zip, .tar, .gz
- **Other**: .rtf, .odt, and more

File parsing is handled through Apache Tika integration for maximum compatibility.

System Requirements

- **Java**: Java 21 (OpenJDK or Oracle JDK recommended)
- **Memory**: 4GB RAM minimum (8GB recommended for production)
- **Disk Space**: 500MB for SDK installation plus additional space for LxBase and configuration files
- **Operating Systems**: Linux, macOS, Windows
- **Build Tool** (for development): Maven 3.6 or higher

© N. Harris Computer Corporation

Quick Start Summary

1. Install the TextChart SDK
2. Set up ROSOKA_HOME environment variable
3. Configure license keys
4. Initialize the Rosoka instance in your code
5. Process documents or strings
6. Retrieve and analyze results

See SDK Installation and Configuration for detailed setup instructions, SDK Usage Guide for development examples, and RosokaProperties Configuration Reference for all available configuration options.

SDK Installation and Configuration

Installation

Prerequisites

Before installing the TextChart SDK, ensure you have:

- Java 21 JDK installed and configured
- Maven 3.6 or higher (for building from source)
- At least 500MB of free disk space
- Administrative access to install software and create directories
- A valid TextChart license

Download and Installation

1. **Obtain the SDK Package**
   - Download the TextChart SDK distribution package
   - Extract the archive to your desired installation directory
   - Example: `/opt/textchart-sdk` or `C:\Program Files\TextChart SDK`
2. **Verify Installation**
   - Navigate to the SDK installation directory
   - Verify that the following directories exist:
     - `lib/` - JAR files and dependencies
     - `conf/` - Configuration files
     - `examples/` - Sample code and documentation

Configuration

Setting ROSOKA_HOME Environment Variable

The TextChart SDK requires the `ROSOKA_HOME` environment variable to locate configuration files and license keys.

© N. Harris Computer Corporation

Linux/macOS

Add the following to your `.bashrc`, `.bash_profile`, or `.zshrc`:

```
export ROSOKA_HOME=/path/to/textchart-sdk
```

Then reload your shell configuration:

```
source ~/.bashrc
# or for macOS:
source ~/.zshrc
```

Windows

Set the environment variable through the System Properties dialog:

1. Open System Properties (Win+Pause or Settings # System # Advanced system settings)
2. Click "Environment Variables"
3. Click "New" under User variables
4. Variable name: `ROSOKA_HOME`
5. Variable value: `C:\Program Files\TextChart SDK` (or your installation path)
6. Click OK and restart any open terminals

Alternatively, set it via Command Prompt:

```
setx ROSOKA_HOME "C:\Program Files\TextChart SDK"
```

License Configuration

1. **Obtain License Key**
   - Contact i2 for your license key file
   - You will receive a file named `licensekeys.xml`
2. **Install License Key**
   - Copy `licensekeys.xml` to `${ROSOKA_HOME}/conf/`
   - Ensure the file has read permissions
3. **Verify License**
   - Run the license verification tool:

     ```
     java -jar ${ROSOKA_HOME}/lib/LicenseCheck.jar
     ```
   - You should see output confirming your license is valid

Configuration Files

The TextChart SDK reads configuration from several XML files in `${ROSOKA_HOME}/conf/`.

RosokaProperties.xml

This is the main configuration file controlling SDK behavior. The SDK uses flat, top-level elements:

```
<?xml version="1.0" encoding="UTF-8"?>
<RosokaProperties>
    <!-- Output configuration -->
    <fileOutputMode>FULL</fileOutputMode>
    <fileOutputType>XML</fileOutputType>

    <!-- Input/output directories -->
```

© N. Harris Computer Corporation

```
    <inputDir>./inputdir/</inputDir>
    <outputDir>./outdir/</outputDir>

    <!-- LxBase (lexicon) directory -->
    <lxbasedir>./LxBase/</lxbasedir>

    <!-- Optional user data directories -->
    <userDataDir>./userData</userDataDir>
</RosokaProperties>
```

Embedding in Applications

When embedding the TextChart SDK in another application, you can set the ROSOKA_HOME path programmatically:

```
import com.rosoka.RosokaAPI.Rosoka;

// Set the Rosoka home directory
Rosoka.setRosokaHome("/path/to/rosoka/home");

// Then get an instance
Rosoka rosoka = Rosoka.getRosokaInstance();
```

The SDK will also check for ROSOKA_HOME in the following order:

1. Programmatically set via `setRosokaHome()`
2. Environment variable `ROSOKA_HOME`
3. Environment variable `ROSOKA_BASE`
4. Current working directory
5. Directory from which the SDK is running

Memory Configuration

For applications that process large documents or high volumes, configure JVM memory settings.

Linux/macOS

```
export JAVA_OPTS="-Xms2G -Xmx4G"
java $JAVA_OPTS -jar your-application.jar
```

Windows

```
set JAVA_OPTS=-Xms2G -Xmx4G
java %JAVA_OPTS% -jar your-application.jar
```

Or in your Java code before initializing Rosoka:

```
// Note: This is illustrative; JVM memory must be set before JVM starts
System.setProperty("java.awt.headless", "true");
```

Performance Tuning

RosokaProperties does not expose thread pool or document cache settings in the current SDK codebase. For performance tuning, adjust JVM options and deployment settings (heap size, CPU allocation, and I/O throughput) and test with representative workloads.

Troubleshooting Installation

License Errors

© N. Harris Computer Corporation

Problem: "License not found" or license validation fails

Solution:

1. Verify `licensekeys.xml` is in `${ROSOKA_HOME}/conf/`
2. Check file permissions (should be readable)
3. Verify the license hasn't expired
4. Run: `java -jar lib/LicenseCheck.jar` to see detailed error

Configuration Not Found

Problem: "RosokaProperties.xml not found"

Solution:

1. Ensure `ROSOKA_HOME` environment variable is set correctly
2. Verify `${ROSOKA_HOME}/conf/` directory exists
3. Check that configuration files are readable
4. On Linux/macOS: `echo $ROSOKA_HOME` to verify the path
5. On Windows: `echo %ROSOKA_HOME%` to verify the path

Memory Issues

Problem: "OutOfMemoryError" when processing large documents

Solution:

1. Increase JVM heap size (see Memory Configuration section above)
2. Process documents in smaller batches
3. Enable document streaming for very large files
4. Check available system memory

Java Version Issues

Problem: "Unsupported class version" or version mismatch errors

Solution:

1. Verify Java version: `java -version`
2. Must be Java 21
3. Update Java if necessary
4. Verify `JAVA_HOME` environment variable points to correct installation

Next Steps

After successful installation and configuration:

1. Review SDK Usage Guide for development examples
2. Examine sample code in `${ROSOKA_HOME}/examples/`
3. Read Entity Types Reference
4. Explore Properties Reference

© N. Harris Computer Corporation

SDK Usage Guide

This guide covers the most common use cases for the TextChart SDK with practical code examples.

Overview

The TextChart SDK provides a straightforward API for text processing. The typical workflow involves:

1. Initialize the Rosoka instance
2. Process a document or string
3. Extract entities from results
4. Access relationships and metadata

Basic Setup

Adding Dependencies

If you're using Maven, add the following to your `pom.xml`:

```
<dependency>
    <groupId>com.rosoka</groupId>
    <artifactId>RosokaAPI</artifactId>
    <version>10.1.2</version>
</dependency>

<dependency>
    <groupId>com.rosoka</groupId>
    <artifactId>RosokaJAXB</artifactId>
    <version>10.1.2</version>
</dependency>
```

Initializing Rosoka

The simplest way to initialize the TextChart SDK is to use the default ROSOKA_HOME location:

```
import com.rosoka.RosokaAPI.Rosoka;
import com.rosoka.JAXB.RosokaFullObject;

try {
    // Initialize with default ROSOKA_HOME
    Rosoka rosoka = Rosoka.getRosokaInstance();

    // Or specify a custom path
    Rosoka rosoka = Rosoka.getRosokaInstance("/path/to/rosoka/home");

} catch (FileNotFoundException | RosokaLicenseException |
 RosokaConfigurationException e) {
    System.err.println("Failed to initialize Rosoka: " + e.getMessage());
    e.printStackTrace();
}
```

Processing Text

Process a String

Extract entities from a text string:

© N. Harris Computer Corporation

```
import com.rosoka.RosokaAPI.Rosoka;
import com.rosoka.JAXB.RosokaFullObject;
import com.rosoka.JAXB.Entitylist;
import com.rosoka.JAXB.Entity;

Rosoka rosoka = Rosoka.getRosokaInstance();

String inputText = "John Smith works at Acme Corporation in New York. " +
                   "He can be reached at john.smith@acme.com or (555)
 123-4567.";

try {
    // Process the string and get full object
    RosokaFullObject result =
 rosoka.processStringRosokaFullObject(inputText);

    // Get entities
    Entitylist entities = result.getEntities();

    // Iterate through extracted entities
    for (Entity entity : entities.getEntity()) {
        System.out.println("Entity: " + entity.getValue());
        System.out.println("Type: " + entity.getEntitytype());
        System.out.println("Confidence: " + entity.getConfidence());
        System.out.println("---");
    }

} catch (Exception e) {
    e.printStackTrace();
}
```

Process a File

Extract entities from a file:

```
import com.rosoka.RosokaAPI.Rosoka;
import com.rosoka.JAXB.RosokaFullObject;
import java.io.File;

Rosoka rosoka = Rosoka.getRosokaInstance();

try {
    // Process file (supports 30+ formats)
    File inputFile = new File("document.pdf");
    RosokaFullObject result = rosoka.processFileRosokaFullObject(inputFile);

    // Process results...

} catch (Exception e) {
    e.printStackTrace();
}
```

Process Multiple Files

Batch processing for efficiency:

```
import com.rosoka.RosokaAPI.Rosoka;
import com.rosoka.JAXB.RosokaFullObject;
import java.io.File;
import java.nio.file.Files;
import java.nio.file.Paths;
```

© N. Harris Computer Corporation

```
Rosoka rosoka = Rosoka.getRosokaInstance();

File directory = new File("documents/");

try {
    Files.walk(Paths.get(directory.getPath()))
        .filter(Files::isRegularFile)
        .forEach(filePath -> {
            try {
                File file = filePath.toFile();
                RosokaFullObject result =
 rosoka.processFileRosokaFullObject(file);

                System.out.println("Processed: " + file.getName());
                System.out.println("Entities found: " +
                    result.getEntities().getEntity().size());

            } catch (Exception e) {
                System.err.println("Error processing " + filePath + ": " +
 e.getMessage());
            }
        });

} catch (Exception e) {
    e.printStackTrace();
}
```

Working with Results

Accessing Entity Information

Entities contain detailed information about extracted values:

```
import com.rosoka.JAXB.Entity;
import com.rosoka.JAXB.atomic.KeyValue;

Entity entity = /* ... from results ... */;

// Basic properties
String value = entity.getValue();              // The extracted text
String type = entity.getEntitytype();          // Entity type (PERSON, ORG,
 etc.)
String confidence = entity.getConfidence();    // Confidence score
String offset = entity.getOffset();            // Position in original text

// Attributes (type-specific properties)
List<KeyValue> attributes = entity.getAttribute();
for (KeyValue attr : attributes) {
    System.out.println(attr.getKey() + ": " + attr.getValue());
}
```

Getting XML Output

Get the raw XML representation of results:

```
import com.rosoka.RosokaAPI.Rosoka;

Rosoka rosoka = Rosoka.getRosokaInstance();
```

© N. Harris Computer Corporation

```
try {
    RosokaObject result = rosoka.processString(inputText);

    // Get as XML string
    String xmlString = rosoka.getPSOAsXMLstring(result);
    System.out.println(xmlString);

    // Get entities as XML
    String entityXml = rosoka.getEntitiesAsXMLstring(result);
    System.out.println(entityXml);

} catch (Exception e) {
    e.printStackTrace();
}
```

Getting JSON Output

Convert results to JSON format:

```
import com.fasterxml.jackson.databind.ObjectMapper;
import com.rosoka.JAXB.RosokaFullObject;

RosokaFullObject result = /* ... from processing ... */;

try {
    ObjectMapper mapper = new ObjectMapper();
    String jsonString = mapper.writeValueAsString(result);
    System.out.println(jsonString);

} catch (Exception e) {
    e.printStackTrace();
}
```

Entity Types

The TextChart SDK can extract many entity types. Common types include:

- **PERSON**: Names of people (John Smith, Mary Johnson)
- **ORGANIZATION**: Company names (Acme Corp, Google Inc.)
- **LOCATION**: Geographic locations (New York, United States)
- **EMAIL**: Email addresses (user@example.com)
- **PHONE**: Phone numbers ((555) 123-4567)
- **URL**: Web addresses (https://example.com)
- **DATE**: Dates (January 15, 2024)
- **CURRENCY**: Money amounts ($1,000, €500)
- **PERCENT**: Percentages (25%, 3.14%)
- **IDENTIFIER**: ID numbers, SSN, etc.

For a complete list, see Entity Types Reference.

Working with Relationships

Extract and analyze relationships between entities:

```
import com.rosoka.JAXB.RosokaFullObject;
import com.rosoka.JAXB.Relationshiplist;
```

© N. Harris Computer Corporation

```
import com.rosoka.JAXB.Relationship;

RosokaFullObject result = /* ... from processing ... */;

Relationshiplist relationships = result.getRelationshiplist();

if (relationships != null) {
    for (Relationship rel : relationships.getRelationship()) {
        System.out.println("From: " + rel.getFromvalue());
        System.out.println("Type: " + rel.getRelationshiptype());
        System.out.println("To: " + rel.getTovalue());
        System.out.println("Confidence: " + rel.getConfidence());
        System.out.println("---");
    }
}
```

Error Handling

The SDK can throw several exception types:

```
import com.rosoka.RosokaAPI.Rosoka;
import com.rosoka.RosokaAPI.RosokaException;
import com.rosoka.EntityExtraction.genericrules.RosokaRuleParserException;

try {
    Rosoka rosoka = Rosoka.getRosokaInstance();
    RosokaFullObject result = rosoka.processStringRosokaFullObject(text);

} catch (FileNotFoundException e) {
    // Configuration or license file not found
    System.err.println("Setup error: " + e.getMessage());

} catch (RosokaLicenseException e) {
    // License validation failed
    System.err.println("License error: " + e.getMessage());

} catch (RosokaConfigurationException e) {
    // Configuration error
    System.err.println("Configuration error: " + e.getMessage());

} catch (RosokaRuleParserException e) {
    // Rule parsing error
    System.err.println("Rule error: " + e.getMessage());

} catch (Exception e) {
    // General exception
    System.err.println("Processing error: " + e.getMessage());
    e.printStackTrace();
}
```

Advanced Topics

Custom Rules

Customize entity extraction by modifying rule files in the LxBase. See LxBase Configuration for details.

Performance Optimization

For high-volume processing:

© N. Harris Computer Corporation

- Use thread pools for parallel processing
- Configure document caching
- Adjust JVM heap size appropriately
- Process documents in batches

See Properties Reference for configuration details.

Examples

See the `examples/` directory in your TextChart SDK installation for complete working examples including:

- Basic text processing
- File processing
- Batch processing
- Custom rule configuration
- Integration examples
- Web service examples

Getting Help

- Review the included documentation
- Check the examples directory
- Contact i2 support for assistance

TextChart Entity Types Reference

TextChart entity types are defined in the default `TokenDefs.xml` files under `rsk-api-core/LxBase/conf/`. This reference reflects those default settings.

Output Entity Types

These entities are defined in the `entities` section of `TokenDefs.xml` and are emitted by default:

- `PERSON` - Named person entities.
- `ORG` - Named organizations.
- `PLACE` - Geographic places.
- `FACILITY` - Facilities (buildings, installations, infrastructure).
- `ADDRESS` - Postal addresses.
- `GEOCOORDINATE` - Coordinates for geospatial mapping.
- `CONVEYANCE` - Vehicles and other means of conveyance.
- `CRIME` - Criminal offenses.
- `TIMESTAMP` - Units of time greater than 24 hours (dates, date-like references).
- `TIMESPAN` - Ranges of time or date ranges.
- `PRODUCT` - Commercial product names.
- `IDNUM` - Identification numbers (serial, SIM, ICCID, IBAN, etc.).
- `EVENT` - Named or generic events.

© N. Harris Computer Corporation

- `PUBLICATION` - Publications, film titles, awards.
- `WEAPON` - Weapons.
- `DRUG` - Drugs (legal and illicit).
- `EMAIL` - Email addresses and social media account names.
- `SOCIAL` - Social media usernames (e.g., handles).
- `PHONE` - Phone numbers.
- `URL` - Web addresses.
- `MONEY` - Numeric amounts of money.
- `IMPLEMENT` - Implements and instruments.
- `PUNITIVE_MEASURE` - Punitive measures taken against an entity.
- `KEYWORD` - Specific keywords or coded speech.
- `TRAUMA` - Injuries or trauma descriptions.
- `POI` - Person of interest (used in short or informal text sources).

Non-Output Entity Types

These entities are defined in the `NoOutputEntities` section of `TokenDefs.xml` and are **not output by default**. They may still be used for internal matching or can be enabled via rule customization:

- `AWARD`
- `BIOMETRIC`
- `CONTRACT_TYPE`
- `CITATION`
- `DNA`
- `GENERIC`
- `ALERT_TYPE`
- `SalientPhrase`
- `CONTROL`
- `DISEASE`
- `FILE_NAME`
- `HASHTAG`
- `IDEOLOGY`
- `CHEMICAL`
- `NON_SALIENT_WEB_CONTENT`
- `MEASURE`
- `MEDICAL_PROCEDURE`
- `MISC`
- `SCORE`
- `PERCENT`
- `NATIONALITY`
- `TRANSIT`

© N. Harris Computer Corporation

- PROGRAM
- PROFESSION
- POLITICAL_AFFILIATION
- GENE
- RATING
- QUOTE
- USER_AGENT
- FINANCIAL_INDEX
- TICKER_SYMBOL
- INFRASTRUCTURE
- ANATOMICAL_TERM
- FUNDS
- CRYPTO
- CLASSIFICATION_LEVEL

Attributes and Subtypes

Attributes and subtypes are defined per-entity in `TokenDefs.xml`. The list is extensive and varies by entity type. For the authoritative set of attributes and descriptions, consult:

- `rsk-api-core/LxBase/conf/TokenDefs.xml`
- `rsk-api-core/LxBase/conf/TokenDefs_core.xml`

Customizing Entity Extraction

You can customize which entity types are extracted and how they are identified by:

1. **Modifying the LxBase**: Edit linguistic rules to add or remove entity type extractions
2. **Creating Custom Rules**: Define custom patterns specific to your domain

See `sdk-lxbase.md` for instructions on customizing entity extraction.

Entity Type Relationships

Relationship extraction depends on rule configuration and is not hard-coded to specific entity pairs. If you need specific relationships (e.g., PERSON # ORG), add or adjust rules in the LxBase.

RosokaProperties Configuration Reference

This is the complete reference for all configuration properties available in TextChart SDK's RosokaProperties.xml file.

Configuration File Format

RosokaProperties configuration is loaded from the XML file `RosokaProperties.xml` in `${ROSOKA_HOME}/conf/`.

Note: JSON parsing utilities exist for programmatic use (PropertyIO.getPropertiesFromJSON), but the standard configuration file format is XML only.

© N. Harris Computer Corporation

Output Configuration

Basic Output Options

```
<RosokaProperties>
    <!-- File output mode: ENTITY, PSO, ENTITY_PSO, COMPOSITE, INLINE,
 INLINEXML, FULL -->
    <fileOutputMode>FULL</fileOutputMode>

    <!-- File output type: XML or JSON -->
    <fileOutputType>XML</fileOutputType>

    <!-- Include inline text in output -->
    <inlinetext>false</inlinetext>

    <!-- Include inline gloss in output -->
    <inlinegloss>false</inlinegloss>

    <!-- Output raw input before processing -->
    <rawinput>false</rawinput>
</RosokaProperties>
```

fileOutputMode Options:

- `ENTITY` - Output only entities
- `PSO` - Output only PSO (Processed Source Objects)
- `ENTITY_PSO` - Output both entities and PSO
- `COMPOSITE` - Output composite format
- `INLINE` - Output inline format
- `INLINEXML` - Output inline XML format
- `FULL` - Output full information (default)

fileOutputType Options:

- `XML` - XML format (default)
- `JSON` - JSON format

Tagging and Processing

```
<!-- Generate PSO (Processed Source Objects) -->
<doPSO>true</doPSO>

<!-- Tag categories in output -->
<dotagcategories>false</dotagcategories>

<!-- Include identity IDs for entity management across documents -->
<includeidentityID>true</includeidentityID>

<!-- Output processing timestamps -->
<useprocesstimestamps>false</useprocesstimestamps>

<!-- Log unknown time formats encountered -->
<logunknowntimeformats>false</logunknowntimeformats>

<!-- Log processed filenames -->
<logprocessedfilenames>false</logprocessedfilenames>
```

© N. Harris Computer Corporation

```
<!-- Enable rule tracing for debugging -->
<ruletrace>false</ruletrace>

<!-- Output token phrase list -->
<tokenphraselist>true</tokenphraselist>

<!-- Include semantic vectors (SV) in output for internal entity
 representations -->
<svoutput>false</svoutput>

<!-- Log unparsable timestamps -->
<logunparsabletimestamps>false</logunparsabletimestamps>
```

Processing Options:

- `doPSO` - Controls whether Processed Source Objects (PSO) are generated. PSOs represent processed and analyzed document structure.
- `dotagcategories` - When enabled, adds category tags to entity output for better classification.
- `includeidentityID` - Enables inclusion of identity IDs to track and correlate entities across multiple documents.
- `useprocesstimestamps` - When enabled, includes timestamps indicating when processing occurred.
- `logunknowntimeformats` - Logs any date/time formats that could not be parsed.
- `logprocessedfilenames` - Logs the names of all files successfully processed.
- `ruletrace` - Enables detailed rule execution tracing for debugging rule behavior.
- `tokenphraselist` - Outputs the list of tokens and phrases detected during processing.
- `svoutput` - Includes the internal semantic vectors used for entity classification and analysis.
- `logunparsabletimestamps` - Logs timestamps that could not be parsed during processing.

Directory Configuration

```
<RosokaProperties>
    <!-- Input directory for documents -->
    <inputDir>./inputdir/</inputDir>

    <!-- Output directory for results -->
    <outputDir>./outdir/</outputDir>

    <!-- User input directory -->
    <userInputDir>NONE</userInputDir>

    <!-- LxBase (lexicon) directory -->
    <lxbasedir>./LxBase/</lxbasedir>

    <!-- LxBase core bundle location -->
    <lxCoreBundle></lxCoreBundle>

    <!-- User corpus directory -->
    <userCorpusDir>userCorpus</userCorpusDir>

    <!-- User LxBase directory -->
    <userLxBaseDir>userLxBase</userLxBaseDir>

    <!-- User data directory -->
    <userDataDir>./userData</userDataDir>
```

© N. Harris Computer Corporation

```
    <!-- Statistics output directory -->
    <statsDir>./statsdir/</statsDir>

    <!-- SSP directory -->
    <sspDir>./sspDir/</sspDir>

    <!-- Temporary files directory -->
    <tempDir>./tempdir/</tempDir>
</RosokaProperties>
```

Directory Options:

- `inputDir` - Source directory containing documents to process. Must end with "/" (forward slash).
- `outputDir` - Destination directory where processed results are written. Must end with "/".
- `userInputDir` - Directory for user-supplied input data.
- `lxbasedir` - Location of the linguistic knowledge base (lexicon) used for entity extraction and linguistic analysis. Must end with "/".
- `lxCoreBundle` - Optional path to the core linguistic bundle used by LxBase import/export utilities.
- `userCorpusDir` - Directory containing user-created corpus files.
- `userLxBaseDir` - Directory for user-customized LxBase data (e.g., unknown words, edits).
- `userDataDir` - General purpose directory for user data and configurations.
- `statsDir` - Directory where processing statistics and metrics are written. Must end with "/".
- `sspDir` - Directory used by semi-structured processing when enabled.
- `tempDir` - Temporary directory for intermediate processing files. Must end with "/".

Geographic Configuration

GeoGravy Settings

```
<RosokaProperties>
    <!-- Enable/disable GeoGravy integration -->
    <internalGeoGravy>OFF</internalGeoGravy>

    <!-- Connection mode: EMBEDDED, CLIENT, WEBSERVICE, JNDI, ES_EMBEDDED,
 ES_CLIENT -->
    <geoGravyConnectionMode>EMBEDDED</geoGravyConnectionMode>

    <!-- Geographic mode: BEST, ALL, COLOCATED -->
    <geoMode>BEST</geoMode>

    <!-- Coordinate format: See Coordinate Formats below -->
    <coordinateFormat>MGRS</coordinateFormat>

    <!-- Geographic sort preference -->
    <geoSortPreference></geoSortPreference>
</RosokaProperties>
```

GxBase (Geographic Reference) Configuration

```
<RosokaProperties>
    <!-- Geographic reference data directory -->
    <gxBaseDir>./GxBase/</gxBaseDir>

    <!-- GxBase host (for remote database) -->
```

© N. Harris Computer Corporation

```
    <gxBaseHost>localhost</gxBaseHost>

    <!-- GxBase port (for remote database) -->
    <gxBasePort>1527</gxBasePort>
</RosokaProperties>
```

internalGeoGravy Options:

- `ON` - GeoGravy integration enabled
- `OFF` - GeoGravy integration disabled (default)

geoGravyConnectionMode Options:

- `EMBEDDED` - Embedded connection with internal DB control (default)
- `CLIENT` - GeoGravy separately controlled
- `WEBSERVICE` - Not implemented in GeoGravyRosokaBridge (logs an error and returns null)
- `JNDI` - App server pooled connection
- `ES_EMBEDDED` - Elasticsearch embedded connection
- `ES_CLIENT` - Elasticsearch client connection

geoMode Options:

- `BEST` - Single best location (default)
- `ALL` - All possible locations
- `COLOCATED` - Colocated locations

GxBase Configuration Options:

- `gxBaseDir` - Path to the geographic reference database directory containing location and coordinate data. Must end with "/".
- `gxBaseHost` - Hostname or IP address of the remote GxBase database server. Used when connecting to a remote geographic reference database instead of an embedded one.
- `gxBasePort` - Port number for the remote GxBase database connection. Default port is 1527 (Derby database default).
- `geoSortPreference` - Controls how GeoGravy sorts competing location matches. Supported values:
  - `CONUS` or `OCONUS` (continental U.S. vs outside U.S. preference).
  - Region or subregion code (case-insensitive): `ASIA`, `EURO`, `AFRCA`, `OCENA`, `AMER`, `ANTAR`, `EURA`, `MDEST`, `SASIA`, `SEASA`, `EASIA`, `NEURO`, `SEURO`, `CEURO`, `BALK`, `NAFR`, `SAFR`, `WAFR`, `CAFR`, `EAFR`, `OCEAN`, `SAMER`, `NAMER`, `CAMER`, `CARIB`, `ANTSB`, `CEURA`, `RUSSA`, `CAUC`, `MEAST`.
  - Country list: `COUNTRY:US,CA,MX` (comma-separated list of ISO country codes).
  - Bounding box: four numbers in the order `N W S E`, separated by spaces or commas.

  Default behavior: if `geoSortPreference` is empty or unset, GeoGravy uses `OCONUS`. No combined preferences: only one preference type is supported at a time (the only multi-value form is the `COUNTRY:` list).

  Examples:

  - `CONUS`
  - `OCONUS`
  - `MEAST`

© N. Harris Computer Corporation

- `COUNTRY:US,CA`
- `45.0 -120.0 30.0 -100.0`

Coordinate Formats

```
<!-- Available coordinate formats -->
<coordinateFormat>DD</coordinateFormat>  <!-- Decimal Degrees -->
<coordinateFormat>DM</coordinateFormat>  <!-- Degrees Minutes -->
<coordinateFormat>DMS</coordinateFormat> <!-- Degrees Minutes Seconds -->
<coordinateFormat>NDMS</coordinateFormat> <!-- Named DMS (e.g., 35°18#29#S
 149°07#28#E) -->
<coordinateFormat>UTM</coordinateFormat>  <!-- Universal Transverse Mercator
 -->
<coordinateFormat>UPS</coordinateFormat>  <!-- Universal Polar Stereographic
 -->
<coordinateFormat>USNG</coordinateFormat> <!-- US National Grid (same as
 MGRS) -->
<coordinateFormat>GEOREF</coordinateFormat> <!-- World Geographic Reference
 -->
<coordinateFormat>MAIDENHEAD</coordinateFormat> <!-- Maidenhead Locator -->
<coordinateFormat>MGRS</coordinateFormat> <!-- Military Grid Reference
 System (default) -->
<coordinateFormat>BNG</coordinateFormat>  <!-- British National Grid -->
<coordinateFormat>ING</coordinateFormat>  <!-- Irish National Grid -->
```

Semi-Structured Processing Configuration

These properties control the semi-structured processing (SSP) engine, which extracts entities and relationships from documents with predictable layouts using XML templates. For full details on writing SSP templates, see Semi-structured processing.

```
<RosokaProperties>
    <!-- Enable semi-structured processing -->
    <sspEnabled>false</sspEnabled>

    <!-- Directory containing SSP template files -->
    <sspDir>./sspDir/</sspDir>

    <!-- Parser implementation to use -->
    <sspParser>utah</sspParser>

    <!-- Anchor entity type for document-level anchor linking -->
    <anchorEntityType></anchorEntityType>
</RosokaProperties>
```

SSP Options:

- `sspEnabled` - When `true`, the SSP engine runs on each document after standard NLP extraction. The engine evaluates every template in `sspDir` and applies the first one whose <`match`> criteria pass. Default: `false`.
- `sspDir` - Directory containing SSP template XML files. All `.xml` files in this directory are loaded at startup. Accepts relative or absolute paths:
  - **Relative** (e.g. `./sspDir/`, `../sspDir`): resolved against the JVM working directory, which is typically the directory from which the application was launched.
  - **Absolute** (e.g. `/opt/textchart/sspDir` or `C:\textchart\sspDir`): used as-is.

© N. Harris Computer Corporation

The directory must exist and be readable at startup. Absolute paths are recommended for production deployments to avoid ambiguity when the application is launched from different directories. Default: `./sspDir/`.

- `sspParser` - Parser implementation used by the SSP engine. The only supported value is `utah`. Default: `utah`.
- `anchorEntityType` - When set to an entity type (e.g. `IDNUM`, `ORG`), the engine locates the first entity of that type in the NLP output and creates relationships linking every other entity back to it. This is a document-level anchor independent of the per-template `isAnchor` attribute in SSP templates. Default: empty (disabled).

Processing Configuration

Entity and Token Processing

```
<RosokaProperties>
    <!-- NP (noun phrase) threshold -->
    <nPthreshold>101</nPthreshold>

    <!-- Maximum depth for chained entities -->
    <maxChainedEntityDepth>4</maxChainedEntityDepth>
</RosokaProperties>
```

Entity Processing Options:

- `nPthreshold` - Controls salient phrase extraction. Values 0-100 enable extraction of noun phrases with salience scores above the threshold (higher = more selective). Default value of 101 disables salient phrase extraction.
- `maxChainedEntityDepth` - Specifies the maximum number of related entities to chain together in output. A value of 1 returns only the primary entity, 2 includes the primary and one related entity, etc. Minimum value is 1 (enforced in code). Used when tracking relationships between consecutive entities in documents.

Regression Testing

```
<RosokaProperties>
    <!-- Enable regression testing -->
    <doRegressionTest>false</doRegressionTest>
</RosokaProperties>
```

Regression Testing Options:

- `doRegressionTest` - When enabled, runs the processor in regression testing mode to validate output against baseline results. Useful for quality assurance and ensuring updates don't break existing functionality.

File Management

```
<RosokaProperties>
    <!-- Delete input files after processing -->
    <deleteOnProcess>false</deleteOnProcess>

    <!-- Delete web uploads after processing -->
    <deleteWebuploadsOnProcess>true</deleteWebuploadsOnProcess>

    <!-- Apply ignore list when true -->
    <ignorefileextensions>false</ignorefileextensions>
```

© N. Harris Computer Corporation

```
    <!-- List of file extensions to ignore -->
    <ignorefileextensionlist>
        <extension>.tmp</extension>
        <extension>.bak</extension>
    </ignorefileextensionlist>
</RosokaProperties>
```

File Management Options:

- `deleteOnProcess` - When true, deletes input files after successful processing.
- `deleteWebuploadsOnProcess` - When true, deletes uploaded files after processing in web-based flows.
- `ignorefileextensions` - Boolean flag (XML element: `ignorefileextensions`). When true, skip processing files whose extensions appear in `ignorefileextensionlist`.
- `ignorefileextensionlist` - Extension list (XML elements: `ignorefileextensionlist` with `extension` children). Matched against the suffix after the last dot, e.g., `.tmp`.

Date and Time Configuration

```
<RosokaProperties>
    <!-- DateTime format: ISO_INSTANT or other Java DateTimeFormatter
 formats -->
    <datetimeformat>ISO_INSTANT</datetimeformat>

    <!-- DateTime locale (if not specified, uses system default) -->
    <!-- Example: en_US, fr_FR, de_DE -->
    <datetimeLocale>en_US</datetimeLocale>
</RosokaProperties>
```

Date/Time Options:

- `datetimeformat` - Specifies the format for date/time values in output. Uses Java DateTimeFormatter patterns. Default is ISO_INSTANT which produces timestamps in ISO 8601 format (e.g., "2023-02-26T10:30:45Z").
- `datetimeLocale` - Specifies the locale for date/time parsing and formatting. When not specified, uses the system default locale. Format is language_COUNTRY (e.g., "en_US", "fr_FR", "de_DE").

Logging and Monitoring

```
<RosokaProperties>
    <!-- Logger properties file location -->
    <loggerpropertiesfile>conf/logger.properties</loggerpropertiesfile>
</RosokaProperties>
```

Logging Options:

- `loggerpropertiesfile` - Path to the Java logging configuration file. This file controls log levels, output destinations, and formatting for all SDK logging.

Text Processing

Transliteration

```
<RosokaProperties>
    <!-- Transliteration values (comma-separated entity types) -->
    <!-- If empty, no transliteration is performed -->
```

© N. Harris Computer Corporation

```
    <transliterationValues></transliterationValues>
</RosokaProperties>
```

Transliteration Options:

- `transliterationValues` - Comma-separated list of entity types that should output transliterated text instead of glossed text. Transliteration converts non-Latin scripts to Latin characters (e.g., Cyrillic "######" # "Moskva", Chinese "##" # "Beijing"). For example, setting this to "PERSON" will output person entity names in transliterated form. Leave empty to disable transliteration and use glossed (dictionary) forms for all entity types. Common use cases: "PERSON,FACILITY" to transliterate names and places.

Complete Example Configuration

```
<?xml version="1.0" encoding="UTF-8"?>
<RosokaProperties>
    <!-- Output Configuration -->
    <fileOutputMode>FULL</fileOutputMode>
    <fileOutputType>XML</fileOutputType>
    <inlinetext>false</inlinetext>
    <inlinegloss>false</inlinegloss>
    <rawinput>false</rawinput>

    <!-- Tagging and Processing -->
    <doPSO>true</doPSO>
    <dotagcategories>false</dotagcategories>
    <includeidentityID>true</includeidentityID>
    <useprocesstimestamps>false</useprocesstimestamps>
    <logunknowntimeformats>false</logunknowntimeformats>
    <logprocessedfilenames>false</logprocessedfilenames>
    <ruletrace>false</ruletrace>
    <tokenphraselist>true</tokenphraselist>
    <svoutput>false</svoutput>

    <!-- Directories -->
    <inputDir>./inputdir/</inputDir>
    <outputDir>./outdir/</outputDir>
    <userInputDir>NONE</userInputDir>
    <lxbasedir>./LxBase/</lxbasedir>
    <lxCoreBundle></lxCoreBundle>
    <userCorpusDir>userCorpus</userCorpusDir>
    <userLxBaseDir>userLxBase</userLxBaseDir>
    <userDataDir>./userData</userDataDir>
    <statsDir>./statsdir/</statsDir>
    <sspDir>./sspDir/</sspDir>
    <tempDir>./tempdir/</tempDir>

    <!-- Geographic Configuration -->
    <internalGeoGravy>OFF</internalGeoGravy>
    <geoGravyConnectionMode>EMBEDDED</geoGravyConnectionMode>
    <geoMode>BEST</geoMode>
    <coordinateFormat>MGRS</coordinateFormat>
    <gxBaseDir>./GxBase/</gxBaseDir>
    <gxBaseHost>localhost</gxBaseHost>
    <gxBasePort>1527</gxBasePort>

    <!-- Semi-Structured Processing -->
    <sspEnabled>false</sspEnabled>
    <sspDir>./sspDir/</sspDir>
    <sspParser>utah</sspParser>
```

© N. Harris Computer Corporation

```
    <anchorEntityType></anchorEntityType>

    <!-- Processing Configuration -->
    <nPthreshold>101</nPthreshold>
    <maxChainedEntityDepth>4</maxChainedEntityDepth>
    <englishonly>false</englishonly>

    <!-- File Management -->
    <deleteOnProcess>false</deleteOnProcess>
    <deleteWebuploadsOnProcess>true</deleteWebuploadsOnProcess>
    <ignorefileextensions>false</ignorefileextensions>
    <ignorefileextensionlist>
        <extension>.tmp</extension>
        <extension>.bak</extension>
    </ignorefileextensionlist>

    <!-- Date/Time -->
    <datetimeformat>ISO_INSTANT</datetimeformat>
    <datetimeLocale>en_US</datetimeLocale>

    <!-- Logging -->
    <loggerpropertiesfile>conf/logger.properties</loggerpropertiesfile>

    <!-- Text Processing -->
    <transliterationValues></transliterationValues>
</RosokaProperties>
```

Property Defaults Summary

Property	Default Value
fileOutputMode	FULL
fileOutputType	XML
inlinetext	false
inlinegloss	false
rawinput	false
doPSO	true
dotagcategories	false
includeidentityID	true
useprocesstimestamps	false
logunknowntimeformats	false
logprocessedfilenames	false
ruletrace	false

© N. Harris Computer Corporation

Property	**Default Value**
tokenphraselist	true
svoutput	false
logunparsabletimestamps	false
inputDir	./inputdir/
outputDir	./outdir/
userInputDir	NONE
lxbasedir	./LxBase/
lxCoreBundle	
userCorpusDir	userCorpus
userLxBaseDir	userLxBase
userDataDir	./userData
statsDir	./statsdir/
sspDir	./sspDir/
tempDir	./tempdir/
internalGeoGravy	OFF
geoGravyConnectionMode	EMBEDDED
geoMode	BEST
coordinateFormat	MGRS
gxBaseDir	./GxBase/
gxBaseHost	localhost
gxBasePort	1527
nPthreshold	101
maxChainedEntityDepth	4
deleteOnProcess	false

© N. Harris Computer Corporation

Property	Default Value
deleteWebuploadsOnProcess	true
ignorefileextensions	false
ignorefileextensionlist	(empty)
datetimeformat	ISO_INSTANT
loggerpropertiesfile	conf/logger.properties
doRegressionTest	false
sspEnabled	false
sspParser	utah
anchorEntityType	(empty)
englishonly	false

Unimplemented Properties

The following properties are defined in the RosokaProperties configuration but have no verified usage in the current codebase. They may be legacy properties from earlier versions, or properties reserved for future features that have not yet been implemented.

Unimplemented Data Store Properties

- `dossierIndexname` - Dossier index name. Default: "dossierindex"
- `embeddeddatastorehttpaccess` - Enable HTTP access to embedded data store. Default: false

Notes

- All directory paths should end with "/" (forward slash) for consistency
- File extensions to ignore should be specified as `.extension` format
- Geographic coordinate format default is MGRS (Military Grid Reference System)
- Entity identity ID tracking is enabled by default for document correlation
- Most file processing defaults preserve original files (deleteOnProcess=false)

Related Documentation

- SDK Installation and Configuration
- LxBase Configuration

© N. Harris Computer Corporation

LxBase Configuration and Customization

What is LxBase?

The LxBase (Lexicon Base) is the linguistic knowledge base that defines how TextChart identifies and extracts entities from text. It contains:

- **Entity definitions**: What constitutes a PERSON, ORGANIZATION, LOCATION, etc.
- **Extraction rules**: Patterns and conditions for identifying entities
- **Linguistic patterns**: Grammar rules, word lists, and terminology
- **Custom extensions**: Domain-specific terms and patterns
- **Rule precedence**: How to handle conflicts between rules

The LxBase is organized as a collection of XML configuration files that work together to guide the extraction process.

LxBase Structure

The LxBase is typically located in `./LxBase/` (as specified in `RosokaProperties.xml` by the `lxbasedir` property) and organized as follows:

```
LxBase/
### Rules/                          # All extraction rules
#   ### RuleFileList.xml           # Main rule file list (specifies loading
 order)
#   ### InitialRules.xml           # Initial rules
#   ### PersonRules.xml            # Person entity extraction rules
#   ### PlaceRules.xml             # Place/location entity rules
#   ### OrgRules.xml               # Organization entity rules
#   ### FacilityRules.xml          # Facility entity rules
#   ### AddressRules.xml           # Address extraction rules
#   ### DateRules.xml              # Date/time entity rules
#   ### PhoneRules.xml             # Phone number rules
#   ### NumberRules.xml            # Number extraction rules
#   ### PersonNameComponents.xml   # Person name parsing
#   ### ChemicalRules.xml          # Chemical entity rules
#   ### ConveyanceRules.xml        # Conveyance/vehicle rules
#   ### EventRules.xml             # Event entity rules
#   ### PublicationRules.xml       # Publication entity rules
#   ### CitationRules.xml          # Citation rules
#   ### ProgramRules.xml           # Program/software rules
#   ### CategoryRules.xml          # Category classification rules
#   ### RecursiveRules.xml         # Recursive/complex rules
#   ### DoubleParenNames.xml       # Names in parentheses
#   ### OtherRules.xml             # Miscellaneous rules
#   ### UndoCleanup.xml            # Final cleanup (must be last)
### dictionary/                     # English dictionary files
#   ### *.xml                      # English lexicon entries
#   ### ...
### nonEnglishDictionaries/         # Non-English language support
#   ### Spanish/
#   ### French/
#   ### German/
#   ### Arabic/
#   ### Chinese/
#   ### ... (other languages)
### dictionary_CORE/                # Core English dictionary (optimized)
```

© N. Harris Computer Corporation

```
#   ### ... (compiled/optimized version)
### nonEnglishDictionaries_CORE/   # Core non-English dictionaries
#   ### Spanish/
#   ### French/
#   ### ... (optimized versions)
### userdictionary/               # User-defined custom dictionaries
#   ### ... (user additions)
### lxbase.xml                    # LxBase metadata
```

Key Points:

- Rules are loaded in the order specified in `Rules/RuleFileList.xml`
- Dictionary files are organized by language
- `*_CORE/` directories contain compiled/optimized versions for better performance
- `userdictionary/` is where custom user additions are stored
- Rule files are XML-based and contain entity extraction patterns

How Rules Work

TextChart rules are defined in XML files and use a combination of pattern matching and semantic vectors to identify and extract entities. The basic rule structure is:

Basic Rule Structure

```
<RuleSet
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    language="english"
    domain="common"
    author="your_name"
    copyright="Your Organization">

    <Rule ID="unique_rule_id">
        <description>What this rule does</description>
        <order>0</order>
        <langconstraint><japanese/></langconstraint>  <!-- Optional:
 language constraints -->

        <!-- Result: what happens when pattern matches -->
        <result>
            <combine>2</combine>  <!-- Combine N tokens into 1 entity (0 =
 no combining) -->
            <sv>
                <PERSON/>  <!-- Assign PERSON semantic vector -->
            </sv>
            <nolonger>
                <!-- Remove these semantic vectors if matched -->
                <ORGANIZATION/>
                <LOCATION/>
            </nolonger>
            <attributes>
                <!-- Extract sub-components -->
                <given_name><T offset="0"/></given_name>
                <sur_name><T offset="1"/></sur_name>
            </attributes>
        </result>

        <!-- When: pattern matching conditions -->
        <when>
```

© N. Harris Computer Corporation

```
            <T offset="0">
                <IS><sv><given_name/></sv></IS>
                <ISNOT><sv><verb/></sv></ISNOT>
            </T>
            <T offset="1">
                <IS><sv><sur_name/></sv></IS>
            </T>
        </when>
    </Rule>
</RuleSet>
```

Rule Components

- **Rule ID**: Unique identifier for the rule
- **Disabled**: Optional attribute to disable rule without removing it (disabled="true")
- **Combine**: Number of consecutive tokens to merge into a single entity (0 = update semantic vectors without combining tokens)
- **Semantic Vectors (sv)**: Entity type tags (PERSON, ORGANIZATION, LOCATION, etc.)
- **nolonger**: Remove these semantic vectors to prevent conflicting classifications
- **langconstraint**: Optional language constraints (e.g., `<langconstraint><spanish/></langconstraint>`). When specified, the rule only applies to documents in those languages.
- **Attributes**: Extract and label components of the matched entity
- **T offset**: Token position relative to current token (0 = current, 1 = next, -1 = previous)

Common Pattern Types

1. Semantic Vector Matching (IS) Match specific semantic vectors:

```
<when>
    <T offset="0">
        <IS><sv><given_name/></sv></IS>
    </T>
    <T offset="1">
        <IS><sv><sur_name/></sv></IS>
    </T>
</when>
```

2. Negative Matching (ISNOT) Exclude specific semantic vectors:

```
<when>
    <T offset="0">
        <IS><sv><given_name/></sv></IS>
        <ISNOT><sv><verb/></sv></ISNOT>
    </T>
    <T offset="1">
        <IS><sv><sur_name/></sv></IS>
    </T>
</when>
```

3. Literal Text Matching Match specific literal text values:

```
<when>
    <T offset="-1">
        <IS><sv><locative_prep/></sv></IS>
        <ISNOT><literal>by</literal></ISNOT>
    </T>
    <T offset="0">
        <IS><sv><cap_word/></sv></IS>
```

© N. Harris Computer Corporation

```
    </T>
</when>
```

4. Multiple Semantic Vector Options (OR Logic) Match tokens with any of multiple semantic vectors:

```
<when>
    <T offset="0">
        <IS><sv><province_name/><statename/><city_name/><placename/></sv></
IS>
    </T>
</when>
```

Rule Processing Order

Rules are processed in the order specified by `Rules/RuleFileList.xml`:

1. `InitialRules.xml` - Basic patterns
2. Language and name component rules
3. Specific entity type rules (Person, Place, Organization, etc.)
4. Cross-domain and complex rules
5. `UndoCleanup.xml` - Final corrections (always last)

Creating Custom Rules

Step 1: Create a Custom Dictionary File

Add custom terms to the LxBase by creating new dictionary files in the appropriate location:

For English terms, create or edit files in `LxBase/dictionary/`:

```
<?xml version="1.0" encoding="UTF-8"?>
<Lexicon xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <LexiconEntry>
        <Headword>John Smith</Headword>
        <semanticVector>
            <given_name/>
            <sur_name/>
            <PERSON/>
        </semanticVector>
    </LexiconEntry>

    <LexiconEntry>
        <Headword>ACME Corporation</Headword>
        <semanticVector>
            <ORGANIZATION/>
            <organization_name/>
        </semanticVector>
    </LexiconEntry>
</Lexicon>
```

For other languages, create files in `LxBase/nonEnglishDictionaries/{Language}/`

For user additions, add files to `LxBase/userdictionary/`

Step 2: Create Custom Rules

Create or modify rule files in `LxBase/Rules/`. Most customization involves adding rules to existing rule files:

© N. Harris Computer Corporation

```
<?xml version="1.0" encoding="UTF-8"?>
<RuleSet
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="../xsd/Rules.xsd"
    language="english"
    domain="common"
    author="your_name">

    <Rule ID="custom_domain_pattern">
        <description>Extract domain-specific identifiers like REQ-12345</
description>
        <order>10</order>
        <result>
            <combine>1</combine>
            <sv>
                <IDENTIFIER/>
            </sv>
            <attributes>
                <identifier><T offset="0"/></identifier>
            </attributes>
        </result>
        <when>
            <T offset="0">
                <!-- Match literal text "REQ-" -->
                <IS><sv><cap_word/></sv></IS>
                <ISNOT><literal>REQ</literal></ISNOT>
            </T>
        </when>
    </Rule>
</RuleSet>
```

Step 3: Register Custom Rules in RuleFileList

If you create a new rule file, add it to `LxBase/Rules/RuleFileList.xml`:

```
<?xml version="1.0" encoding="UTF-8"?>
<rulefilelist>
    <file>InitialRules.xml</file>
    <file>PersonRules.xml</file>
    <!-- ... other rules ... -->
    <file>YourCustomRules.xml</file>  <!-- Add your custom rule file -->
    <file>UndoCleanup.xml</file>      <!-- Must be last -->
</rulefilelist>
```

Step 4: Test Your Customizations

Restart the TextChart SDK and test extraction with sample documents:

```
import com.rosoka.RosokaAPI.Rosoka;
import com.rosoka.JAXB.RosokaFullObject;

// After modifying LxBase
Rosoka rosoka = Rosoka.getRosokaInstance();

String testText = "Contact John Smith at ACME Corporation";
RosokaFullObject result = rosoka.processStringRosokaFullObject(testText);

// Verify new entities are extracted
result.getEntities().getEntity().forEach(entity ->
    System.out.println(entity.getValue() + " -> " + entity.getEntitytype())
);
```

© N. Harris Computer Corporation

Advanced Rule Features

Using Nolonger Vectors

The `nolonger` element prevents conflicting semantic vector assignments:

```
<Rule ID="person_vs_organization">
    <description>Ensure person names don't conflict with organizations</
description>
    <result>
        <combine>2</combine>
        <sv>
            <PERSON/>
            <person_name/>
        </sv>
        <nolonger>
            <!-- Remove these if this person pattern matches -->
            <ORGANIZATION/>
            <organization_name/>
            <PLACE/>
            <placename/>
        </nolonger>
        <attributes>
            <given_name><T offset="0"/></given_name>
            <sur_name><T offset="1"/></sur_name>
        </attributes>
    </result>
    <when>
        <!-- Pattern matching conditions -->
    </when>
</Rule>
```

Multi-Token Matching and Combining

Combine multiple tokens into a single entity:

```
<Rule ID="three_part_name">
    <description>Match first name, middle initial, last name</description>
    <result>
        <combine>3</combine>  <!-- Merge 3 tokens into 1 -->
        <sv>
            <PERSON/>
        </sv>
        <attributes>
            <given_name><T offset="0"/></given_name>
            <middle_initial><T offset="1"/></middle_initial>
            <sur_name><T offset="2"/></sur_name>
        </attributes>
    </result>
    <when>
        <T offset="0">
            <IS><sv><given_name/></sv></IS>
        </T>
        <T offset="1">
            <IS><sv><initial/></sv></IS>
        </T>
        <T offset="2">
            <IS><sv><sur_name/></sv></IS>
        </T>
    </when>
```

© N. Harris Computer Corporation

```
</Rule>
```

Domain-Specific Customization

Law Enforcement and Criminal Justice

Add specialized rules for legal entities:

```
<RuleSet language="english" domain="law_enforcement">
    <Rule ID="criminal_case_number">
        <description>Extract case numbers preceded by CASE keyword</
description>
        <result>
            <combine>2</combine>
            <sv>
                <CASE_NUMBER/>
            </sv>
            <attributes>
                <case_keyword><T offset="0"/></case_keyword>
                <case_number><T offset="1"/></case_number>
            </attributes>
        </result>
        <when>
            <T offset="0">
                <IS><sv><cap_word/></sv></IS>
                <literal>CASE</literal>
            </T>
            <T offset="1">
                <IS><sv><NUMBER/></sv></IS>
            </T>
        </when>
    </Rule>

    <Rule ID="state_vs_pattern">
        <description>Extract "State v. Defendant" patterns</description>
        <result>
            <combine>3</combine>
            <sv>
                <LEGAL_CASE/>
            </sv>
            <attributes>
                <defendant><T offset="2"/></defendant>
            </attributes>
        </result>
        <when>
            <T offset="0">
                <IS><sv><cap_word/></sv></IS>
                <literal>State</literal>
            </T>
            <T offset="1">
                <IS><sv><locative_prep/></sv></IS>
            </T>
            <T offset="2">
                <IS><sv><PERSON/></sv></IS>
            </T>
        </when>
    </Rule>
</RuleSet>
```

© N. Harris Computer Corporation

Medical and Healthcare Domain

```
<RuleSet language="english" domain="medical">
    <Rule ID="medication_reference">
        <description>Extract medication names - match capitalized words with
 medical semantic vectors</description>
        <result>
            <combine>1</combine>
            <sv>
                <MEDICATION/>
            </sv>
            <nolonger>
                <PERSON/>
                <ORGANIZATION/>
            </nolonger>
        </result>
        <when>
            <T offset="0">
                <IS><sv><cap_word/><medication_name/></sv></IS>
                <ISNOT><sv><PERSON/><ORGANIZATION/></sv></ISNOT>
            </T>
        </when>
    </Rule>
</RuleSet>
```

Financial Services

```
<RuleSet language="english" domain="financial">
    <Rule ID="account_reference">
        <description>Extract account references - pattern "Account" followed
 by number</description>
        <result>
            <combine>2</combine>
            <sv>
                <ACCOUNT_NUMBER/>
            </sv>
            <attributes>
                <account_number><T offset="1"/></account_number>
            </attributes>
        </result>
        <when>
            <T offset="0">
                <IS><sv><cap_word/></sv></IS>
                <literal>Account</literal>
            </T>
            <T offset="1">
                <IS><sv><NUMBER/></sv></IS>
            </T>
        </when>
    </Rule>
</RuleSet>
```

Managing Rule Conflicts

When multiple rules could match the same text, TextChart uses these strategies:

File Order Based Resolution

Rules are evaluated in the order specified in `RuleFileList.xml`. Earlier files have priority:

```
<!-- In RuleFileList.xml -->
```

© N. Harris Computer Corporation

```
<rulefilelist>
    <file>InitialRules.xml</file>      <!-- Evaluated first -->
    <file>PersonRules.xml</file>       <!-- Evaluated second -->
    <!-- ... -->
    <file>UndoCleanup.xml</file>       <!-- Evaluated last -->
</rulefilelist>
```

Within each file, rules execute sequentially in the order they appear in the XML.

Using Nolonger for Conflict Resolution

The `nolonger` element removes conflicting semantic vectors:

```
<Rule ID="person_vs_organization">
    <description>Ensure person classification wins over organization</
description>
    <result>
        <sv>
            <PERSON/>
        </sv>
        <nolonger>
            <!-- Remove these conflicting classifications -->
            <ORGANIZATION/>
            <PLACE/>
        </nolonger>
    </result>
    <!-- ... -->
</Rule>
```

Disabling and Enabling Rules

You can disable individual rules using the `disabled` attribute without removing them from the file:

```
<RuleSet>
    <!-- This rule is active -->
    <Rule ID="active_rule">
        <!-- ... -->
    </Rule>

    <!-- This rule is disabled but preserved -->
    <Rule ID="disabled_rule" disabled="true">
        <description>This rule will not be used</description>
        <!-- ... -->
    </Rule>
</RuleSet>
```

Alternatively, you can comment out or remove rules from rule files without affecting the overall structure, or create separate rule files for different domains or use cases and load them conditionally via `RuleFileList.xml`.

Performance Considerations

When creating custom rules:

- **Use specific patterns**: Specific patterns are faster than overly general ones
- **Dictionary size**: Larger dictionaries take longer to match against
- **Rule complexity**: Simpler rules execute faster
- **Order matters**: Place frequently-matching rules early in `RuleFileList.xml`
- **Token range**: Rules that check many token offsets are slower than those checking nearby tokens

© N. Harris Computer Corporation

Monitor performance using:

```
long startTime = System.currentTimeMillis();
RosokaFullObject result = rosoka.processStringRosokaFullObject(text);
long endTime = System.currentTimeMillis();
System.out.println("Processing time: " + (endTime - startTime) + "ms");
```

Testing Custom Rules

Create a test harness for validating rule behavior:

```
import com.rosoka.RosokaAPI.Rosoka;
import com.rosoka.JAXB.RosokaFullObject;

public class RuleTest {
    public static void main(String[] args) throws Exception {
        Rosoka rosoka = Rosoka.getRosokaInstance();

        String[] testCases = {
            "Contact John Smith at john@example.com",
            "The case is State v. Brown",
            "Call the office at (555) 123-4567"
        };

        for (String test : testCases) {
            RosokaFullObject result =
 rosoka.processStringRosokaFullObject(test);
            System.out.println("Input: " + test);
            result.getEntities().getEntity().forEach(entity ->
                System.out.println("  " + entity.getValue() + " (" +
 entity.getEntitytype() + ")")
            );
        }
    }
}
```

Versioning and Maintenance

When updating the LxBase:

1. **Backup existing rules**: Keep a copy of current rule files before making changes
2. **Version your changes**: Add version comments to rule files
3. **Document changes**: Keep a changelog of rule modifications
4. **Test thoroughly**: Validate with representative samples
5. **Deploy carefully**: Roll out to production in phases

Best Practices

1. **Be specific**: Write specific rules rather than overly broad patterns
2. **Document rules**: Include descriptions and examples
3. **Use dictionaries**: For word lists, use dictionary files instead of embedding in rules
4. **Avoid duplication**: Reuse common patterns
5. **Test early**: Test new rules before deploying to production
6. **Monitor performance**: Watch for rules that degrade extraction speed
7. **Keep organized**: Organize rules by domain or entity type

© N. Harris Computer Corporation

8. **Review regularly**: Periodically review rules for effectiveness and relevance

Unimplemented Rule Attributes

The following attributes are defined in the rule structure but are **not used** in actual rule processing:

The order Attribute

The `order` attribute is parsed and stored in Rule objects but is **NOT used to control rule execution sequence** or priority. Rules execute in the order they appear in rule files as specified by `RuleFileList.xml`, regardless of `order` values.

Status: Parsed but unused (likely metadata/documentation placeholder for future use)

```
<!-- These order values have NO EFFECT on execution sequence -->
<Rule ID="example_rule" order="0">
    <!-- ... -->
</Rule>
```

Getting Help

For assistance with rule customization:

- Check the examples in the custom rules directory
- Consult the online documentation
- Contact i2 support

Tesseract OCR parser

Tesseract Optical Character Recognition (OCR) parsing may be installed to provide seamless OCR support for many common graphical file formats, including JPEG, TIFF and GIF. Once Tesseract OCR is installed and running, TextChart will seamlessly process supported file formats in the ingestion process via the Tika interface. No further configuration is required.

Installing Tesseract on RHEL

1. Install the EPEL repository if it isn't already available:

   ```
   sudo dnf install epel-release
   ```
2. Install Tesseract:

   ```
   sudo dnf install tesseract
   ```
3. To add language packs, see what's available:

   ```
   dnf search tesseract
   ```

   And then install, for example:

   ```
   sudo dnf install tesseract-langpack-ara
   ```

Installing Tesseract on Ubuntu

1. Download package information from all configured sources:

   ```
   sudo apt update
   ```
2. Install Tesseract:

   ```
   sudo apt install tesseract-ocr
   ```

© N. Harris Computer Corporation

3. To add language packs, see what's available:

   ```
   apt search tesseract-ocr
   ```

   And then install, for example:

   ```
   sudo apt install tesseract-ocr-fra
   ```

Installing Tesseract on Windows

1. Download the latest Tesseract installer from the UB-Mannheim Tesseract repository. Older versions are available at `https://digi.bib.uni-mannheim.de/tesseract/`.
2. Run the installer and follow the prompts. Install Tesseract in the directory suggested by the installer or in a new directory.

   Important: Do not install into an existing directory that contains other files. The uninstaller removes the entire installation directory.
3. During installation, select any additional language packs that you require.
4. Add the Tesseract installation directory to the system `PATH` environment variable so that it is available from the command line.

Installing Tesseract on other platforms

For further details on Tesseract, including installation on Macintosh, see the information at https://cwiki.apache.org/confluence/display/tika/TikaOCR.

Download PDFs

i2 TextChart documentation is also available in PDF documents.

- i2 TextChart Standard
- i2 TextChart Premium
- i2 TextChart Studio
- i2 TextChart LxBase
- i2 TextChart Full Collection

© N. Harris Computer Corporation