

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.

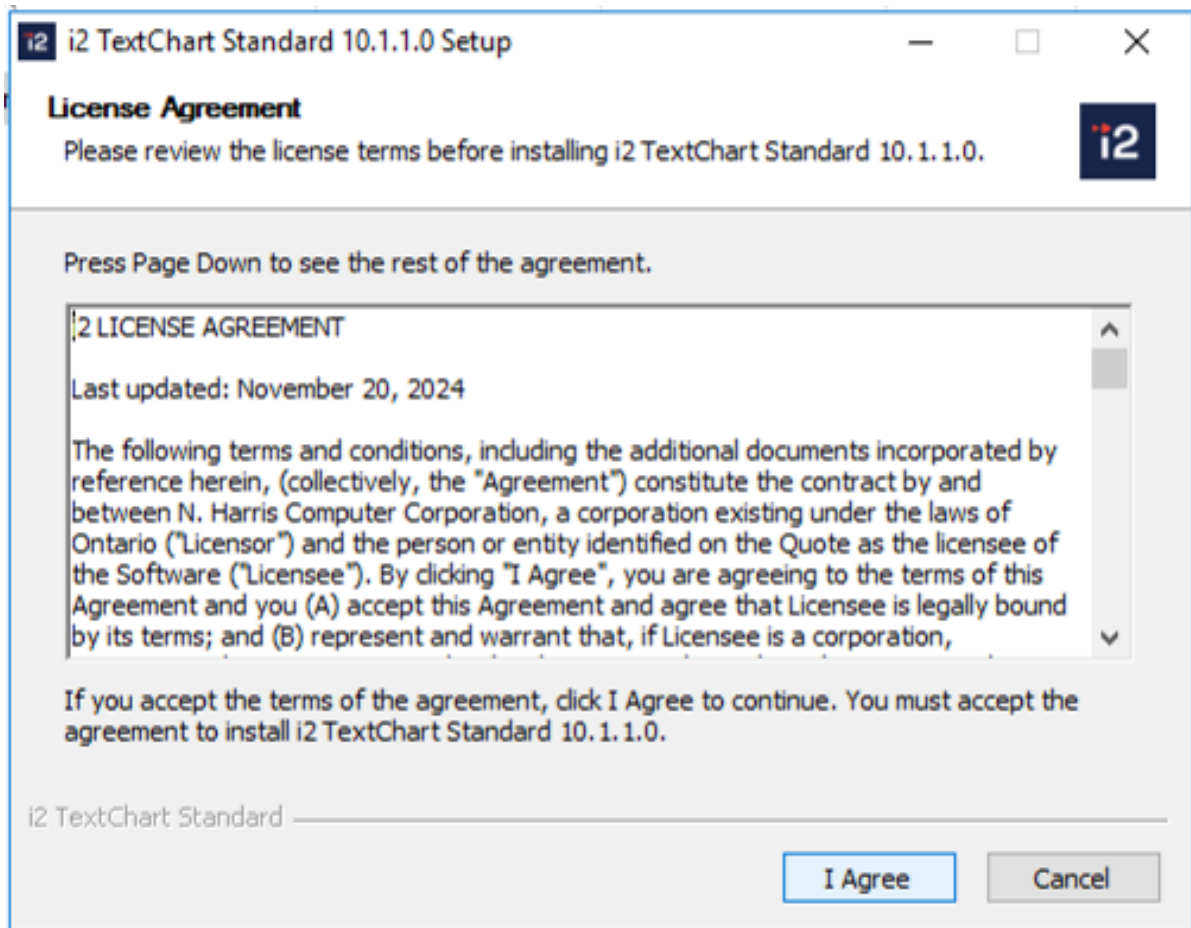
Optional configuration

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.

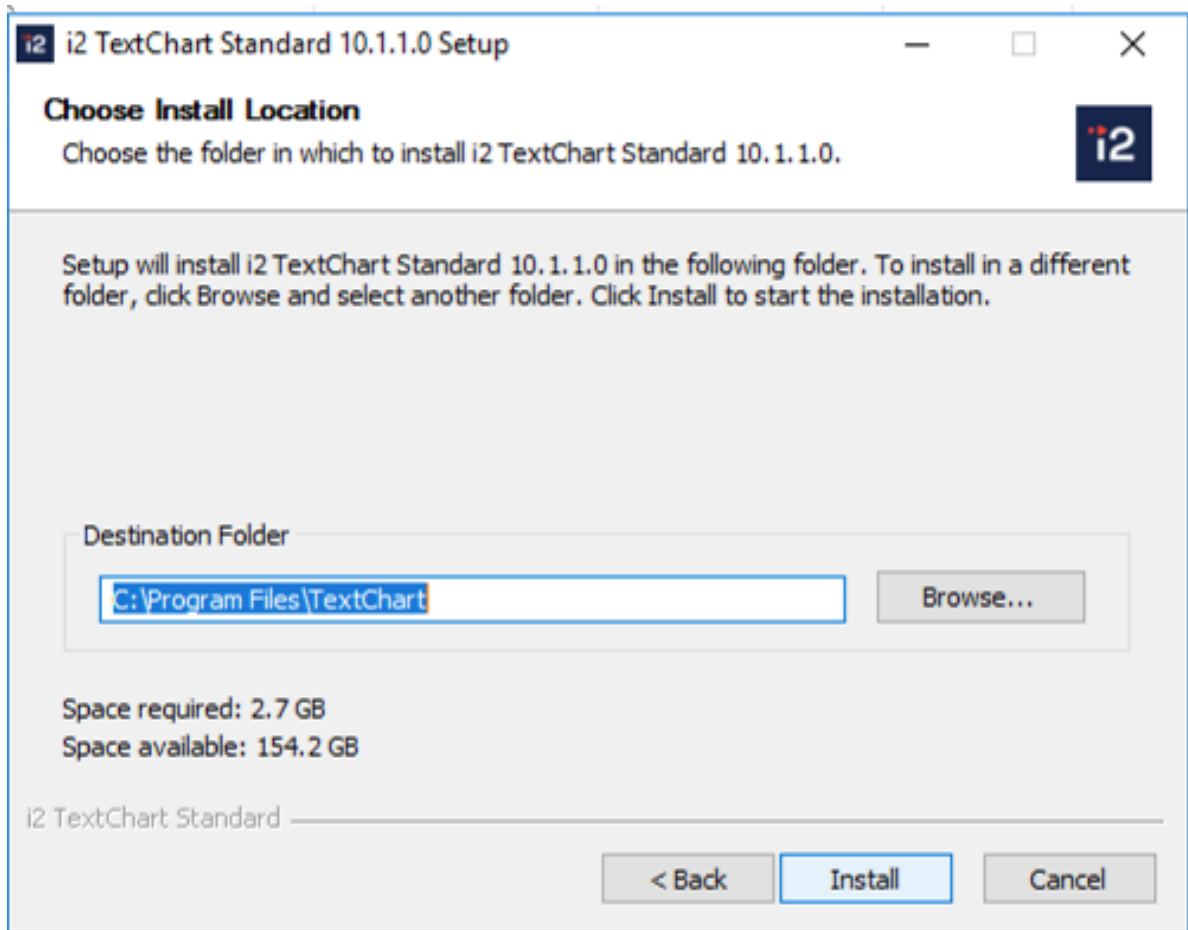
Procedure

The following steps describe how to install i2 TextChart.

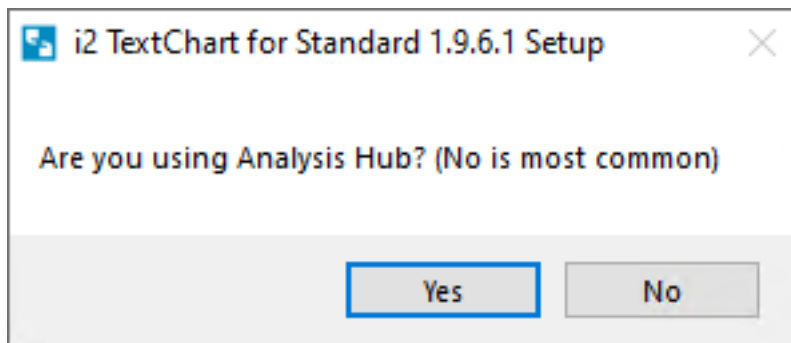
1. Extract the product files from your downloaded distribution.
2. Using Windows Explorer, browse to the root of the distribution and run the i2 TextChart Installer. Its name will contain the word "Premium", as this is the edition of i2 TextChart you're installing.
3. Follow the prompts. You must accept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



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

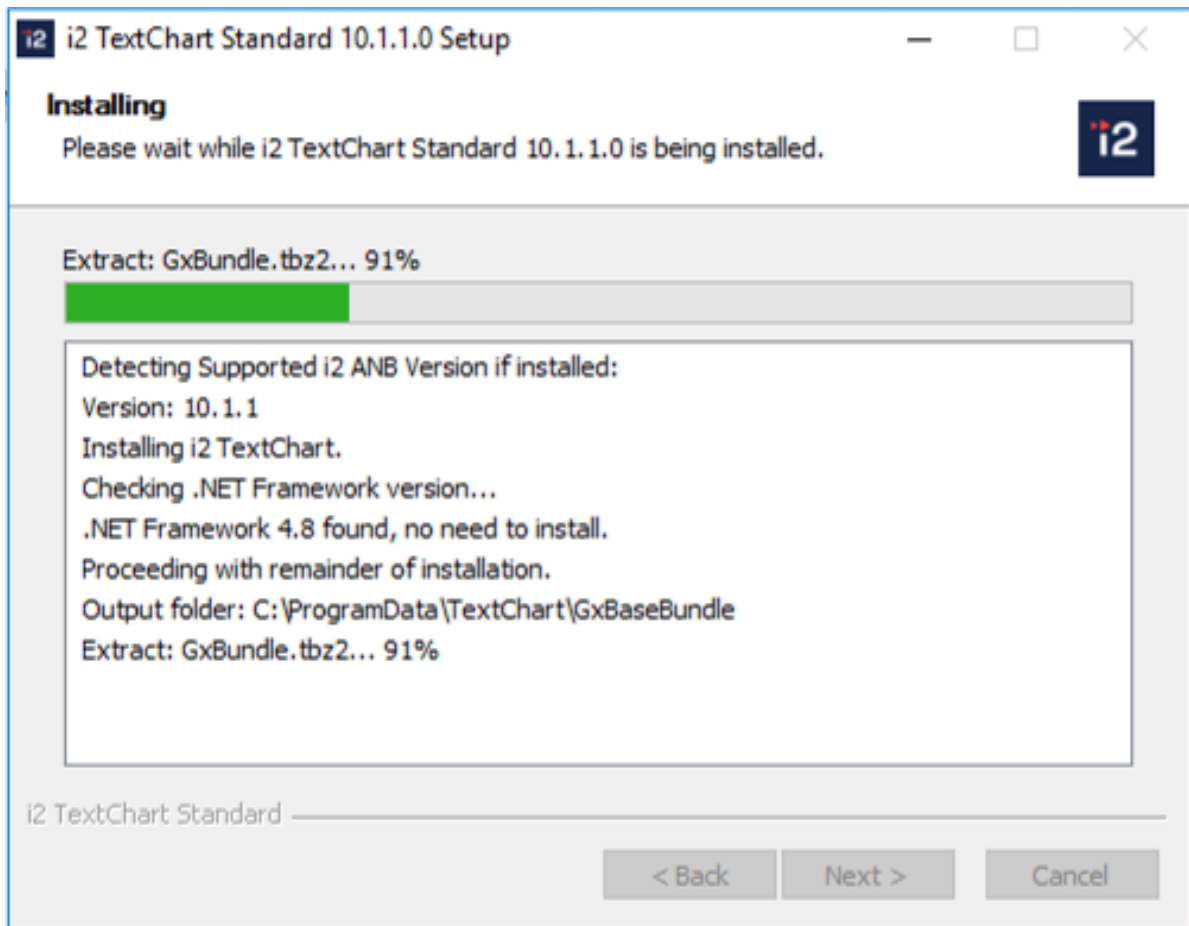


5. If you're installing the standard edition of i2 TextChart, the installer displays an extra question:

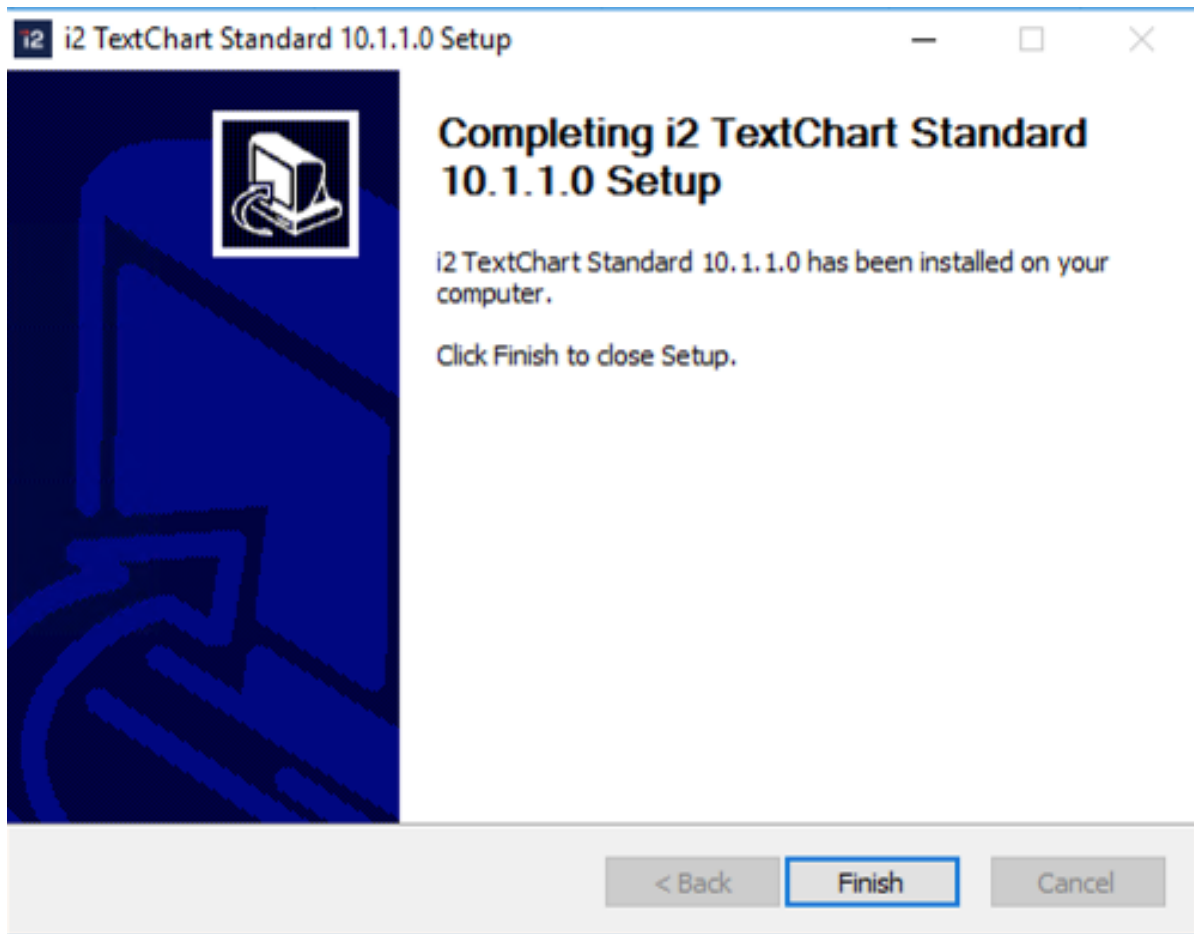


If your workflow involves connecting to an i2 Analyze server from Analyst's Notebook, click **Yes**. Otherwise, click **No**.

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



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



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.

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

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.

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.

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.

The screenshot shows the 'Local DB' configuration window. On the left, under 'Database Connection (local)', the 'Type' dropdown is set to 'Microsoft SQL 2014 +'. The 'Host' field contains 'localhost' and the 'Port' field contains '1433'. In the 'Authentication' section, the 'use Windows Authentication' checkbox is unchecked. The username field contains 'sa' and the password field is masked with dots. The 'Test' and 'Save' buttons are highlighted with red rectangles. On the right, the 'Table Operations' section contains three orange buttons: 'Create', 'Drop', and 'Clear'.

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.

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.

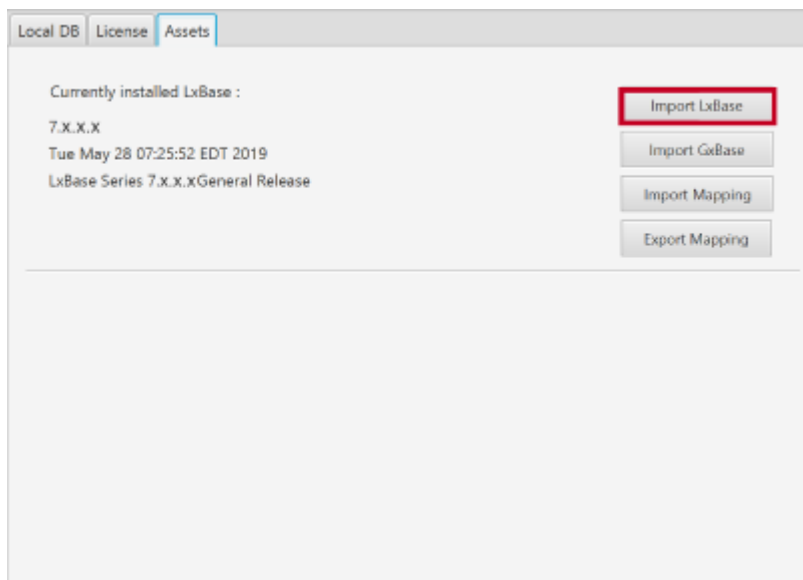
Updating an LxBase

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.

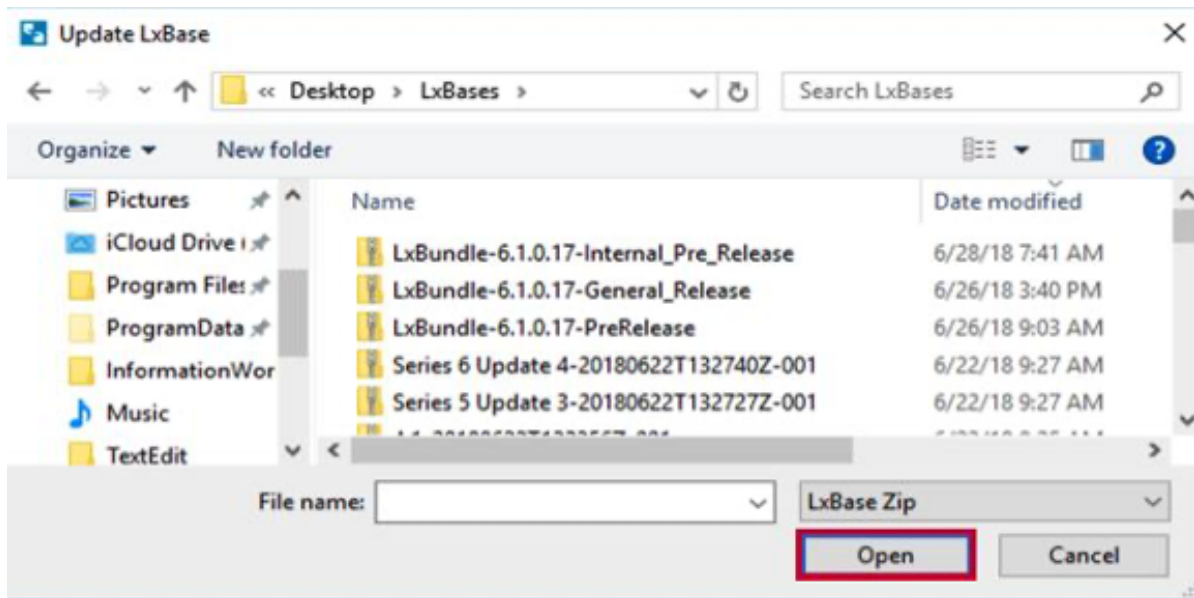
Procedure

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:

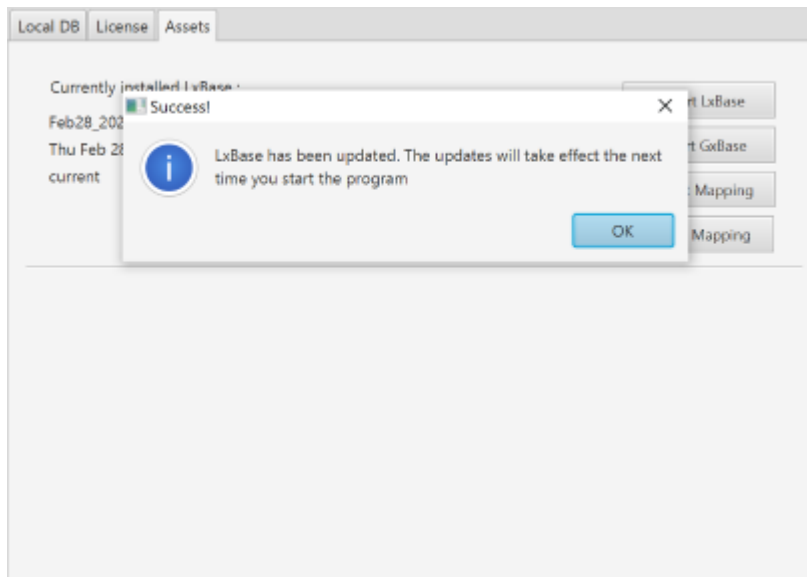


2. Click **Import LxBase**, navigate to the new ZIP file, and click **Open**.



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.

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

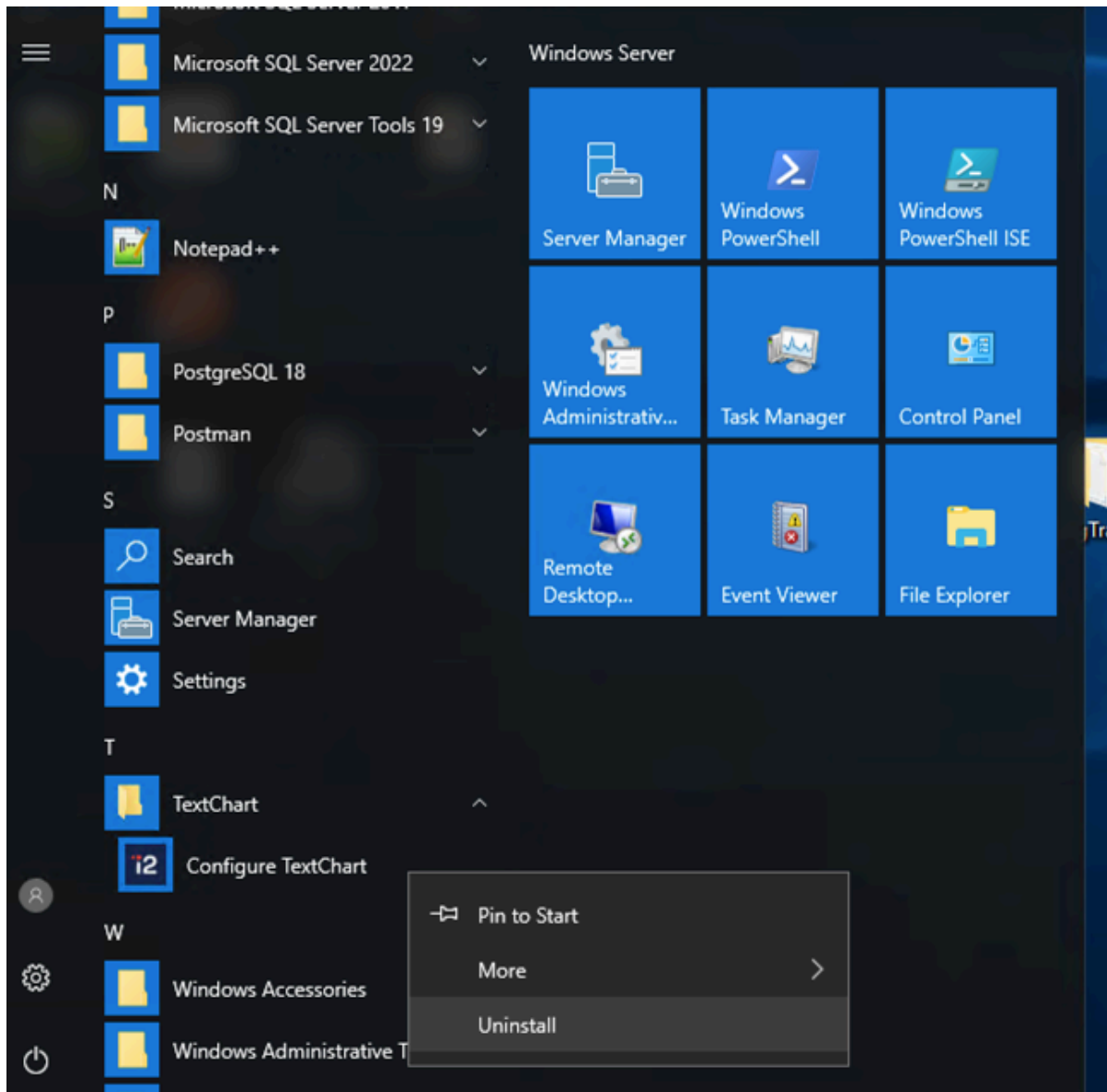
Step 1: Back up your existing database

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



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.

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**.

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
```

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](https://digi.bib.uni-mannheim.de/tesseract/). 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

- 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

Rosoka Data Access Configuration

Status Rosoka Server **Database** Assets RDA Settings

Rosoka Data access must be configured to connect to the Database where Rosoka Server is writing its output to.

Select Database Type

✓ Microsoft SQL Server

PostgreSQL

Select the type of database to connect to

Database Host	Database Port
localhost	1433
Database host	Database port

Use i2 Analyze Security Schema?

DISABLED

Requires confirmation to enable/disable.

Database Username	Database Password
sa	*****
Database Username	Database Password

Use Active Directory? ☐

Check to use Active Directory DB login.
Requires additional server configuration.

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

3. Click Save to persist configuration

Rosoka Data Access Configuration

Status Rosoka Server **Database** Assets RDA Settings

Database Connection Successful!

Database Type

PostgreSQL

Select the type of database to connect to

Database Host Database Port

localhost 5432

Database host Database port

Use i2 Analyze Security Schema?

DISABLED

Requires confirmation to enable/disable.

Database Username Database Password

sa

Use Active Directory? ☐

Check to use Active Directory DB login.
Requires additional server configuration.

TEST SAVE

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.

[test] Configuration

Assets

Extraction

Output

I2SQLOutput ▾

Configure

I2SQLOutput Configuration

Database Type required	Microsoft SQL Server ▾
Database Host required	10.211.55.2
Named Instance optional	
Database Port required	
Database UserName required	

Save

Cancel

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.

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.

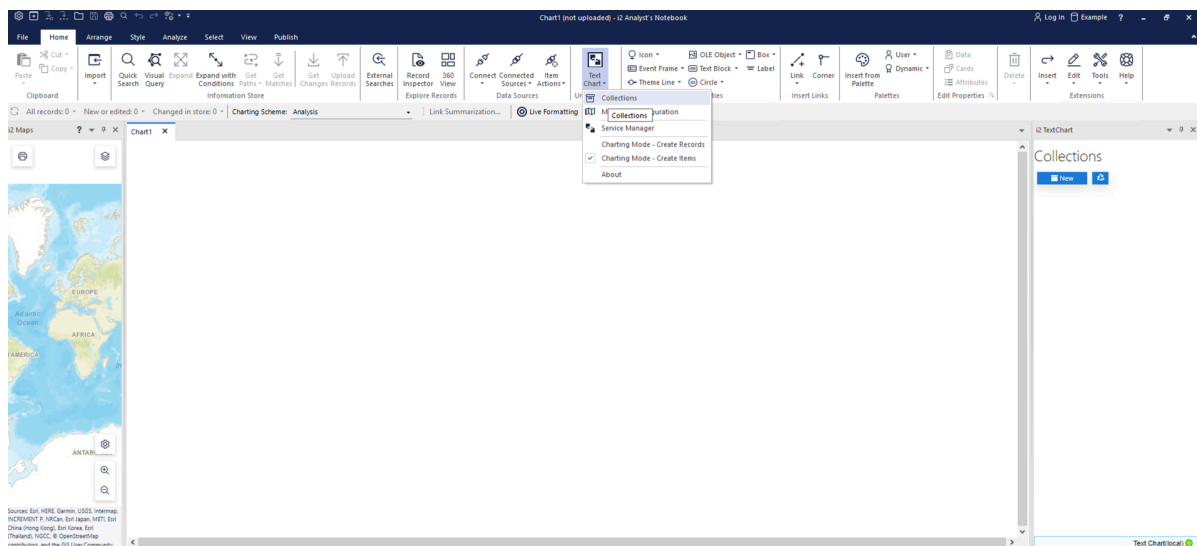
The predicate (that is, the reason why i2 TextChart identified the link) is stored with the link. If the predicate is "interviewed", then the link indicates that a Person interviewed another Person.

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

Importing documents

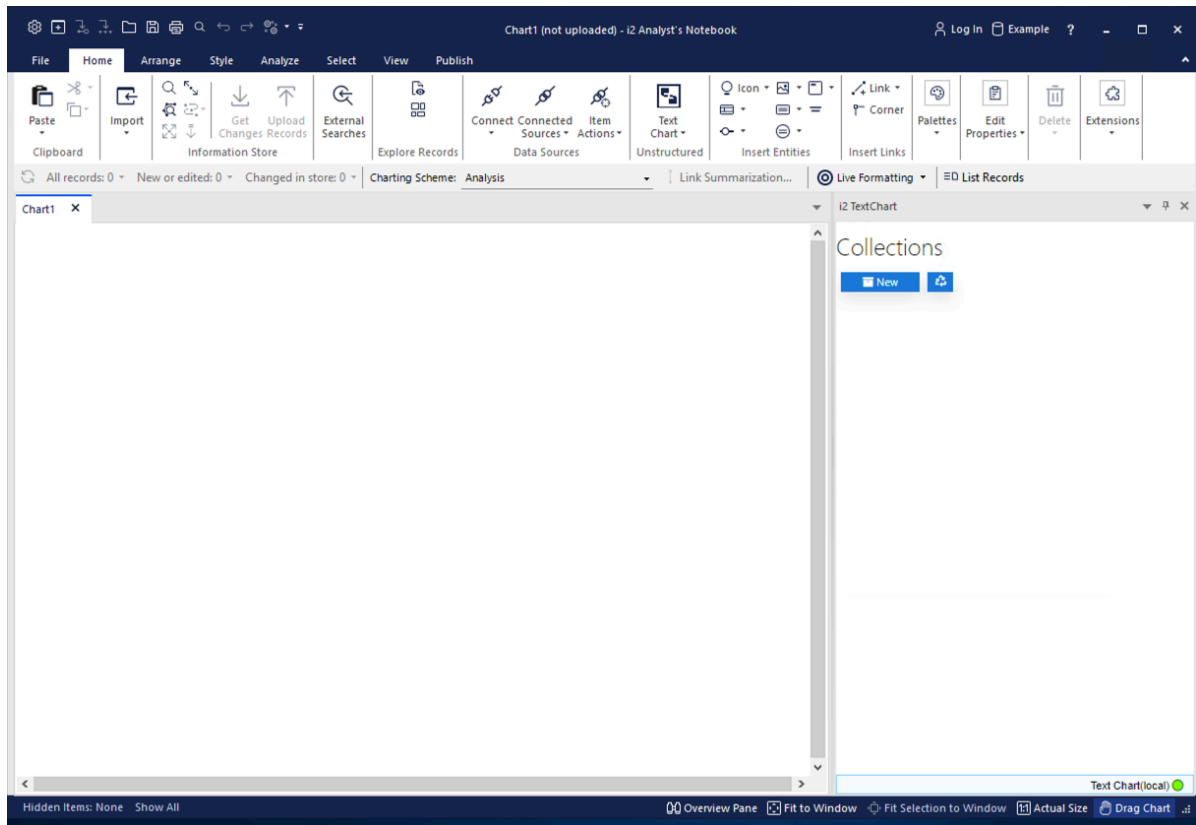
To use i2 TextChart, the first step is to import the documents that you want to analyze.

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

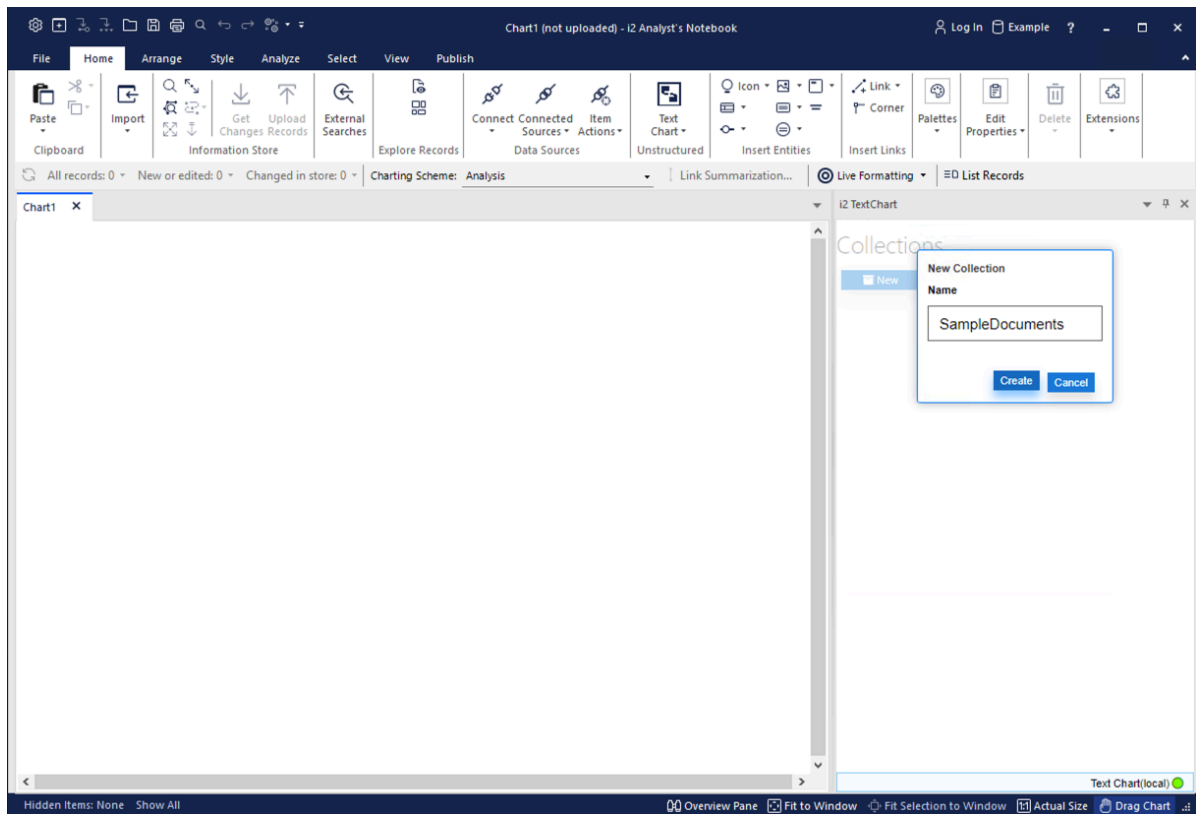


Note: The Mapping Configuration option is available only in the standard edition of TextChart. In TextChart Premium, mapping takes place on the server.

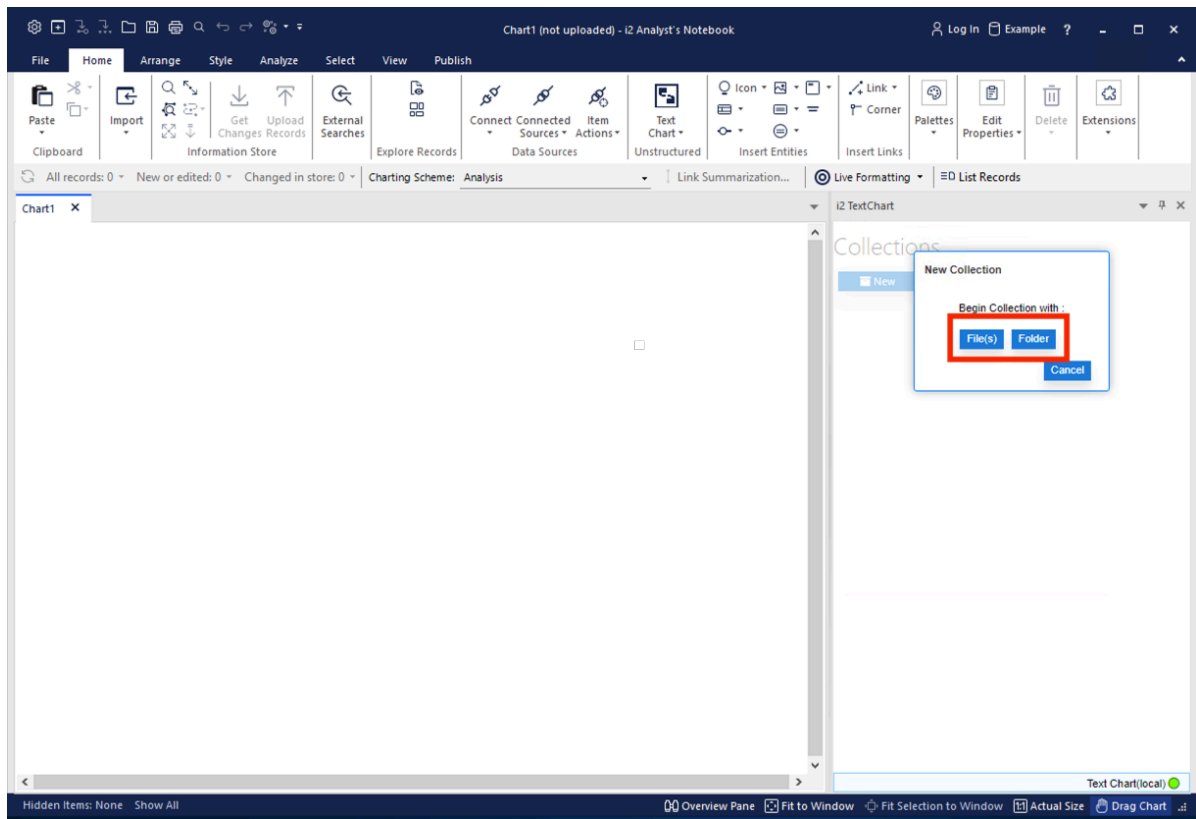
2. i2 TextChart stores documents in *collections*. A collection contains one or more documents that you choose to be in the same group for processing. You can start a new collection for the documents that you're importing, or add them to an existing one.



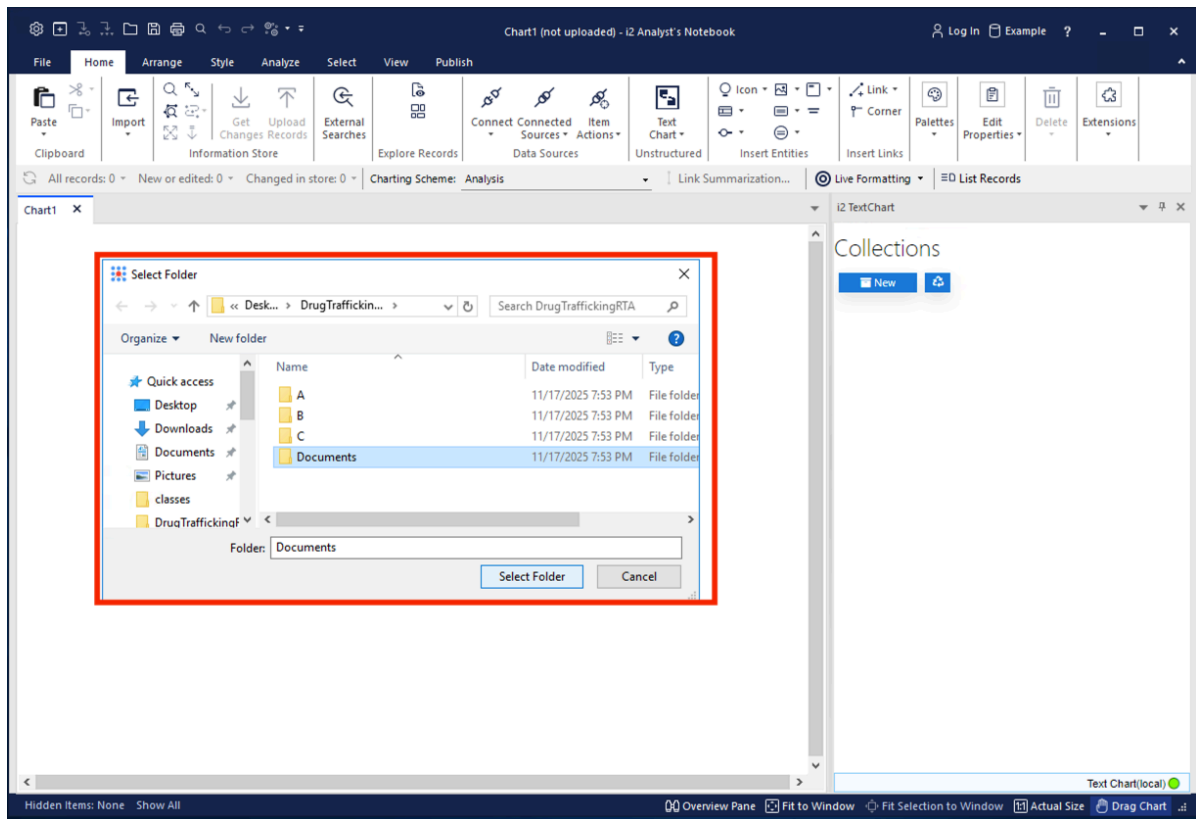
3. If you clicked **New**, provide a name for the new collection and click **Create**.



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



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

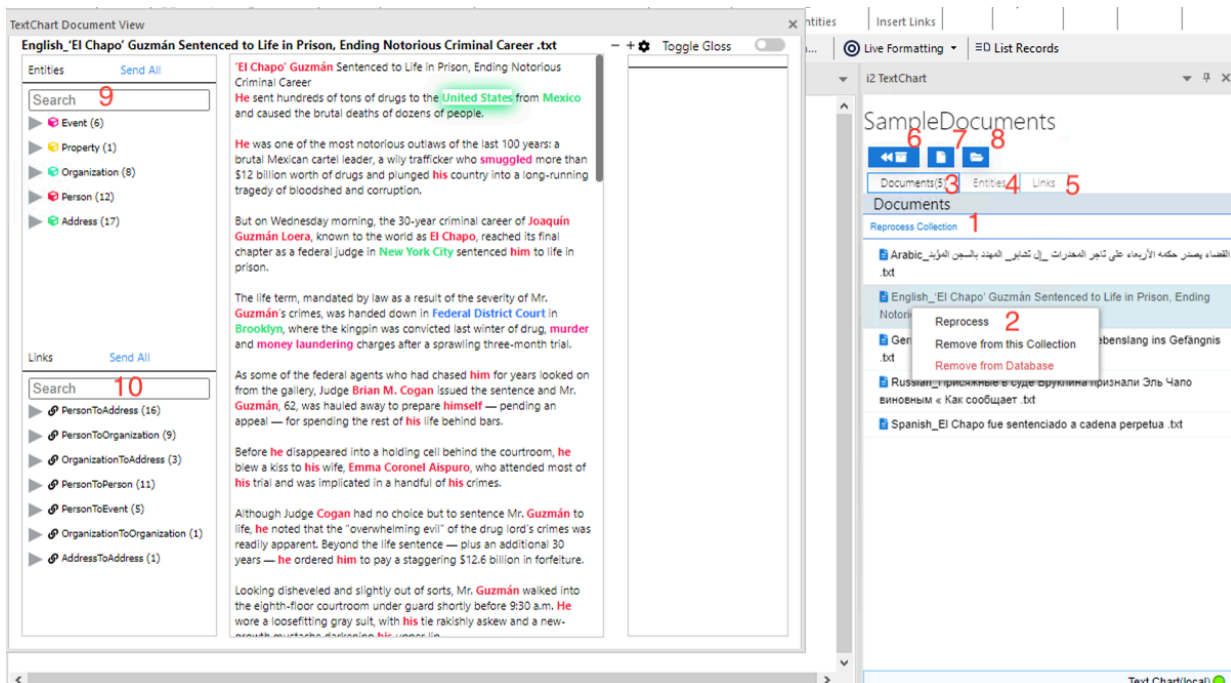


Processing documents

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.



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

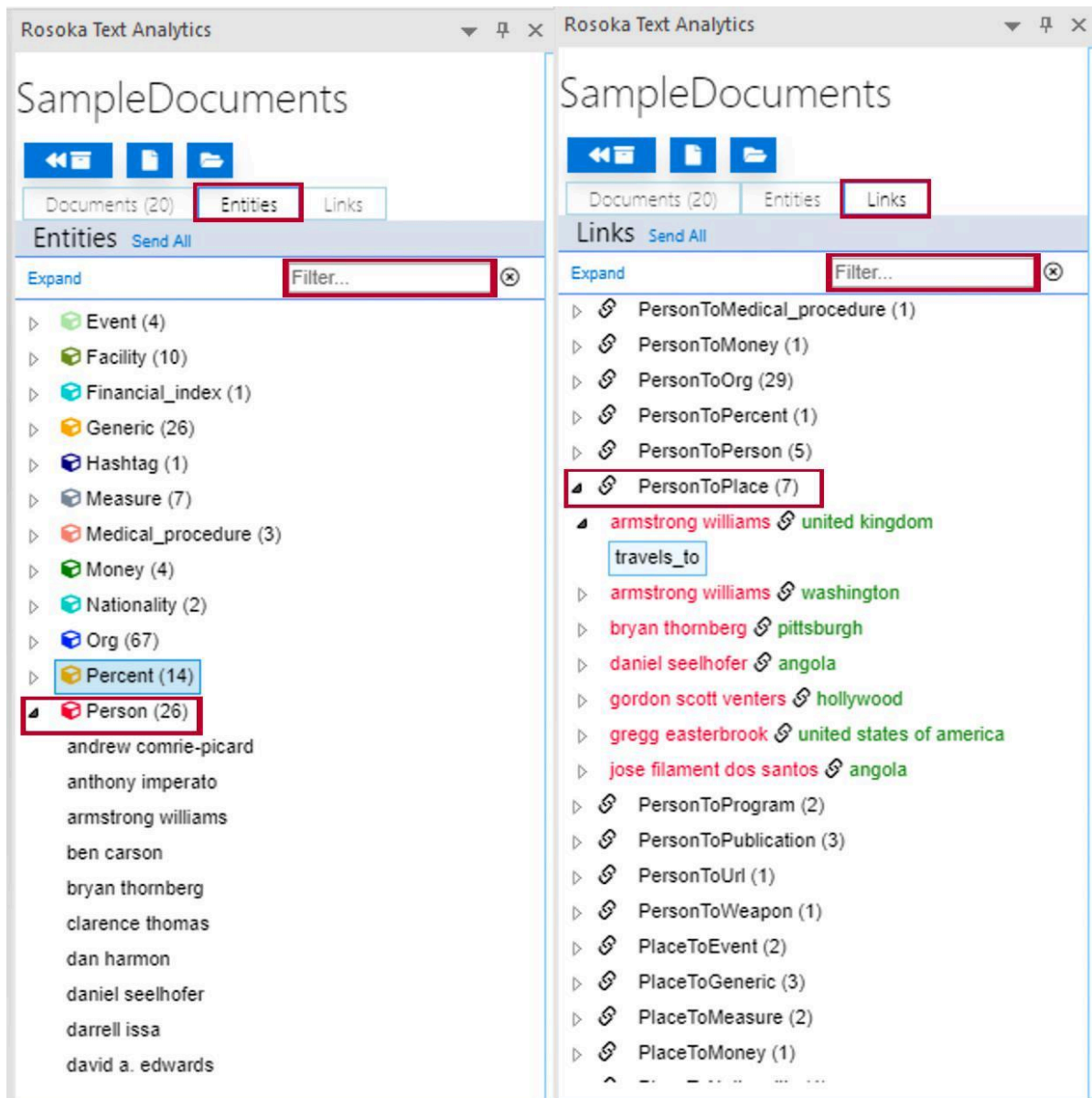
If you right-click a single document in the list, the pop-up menu presents commands to reprocess the document, or to remove it from the collection, or to remove its processing results from the database.

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.



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.

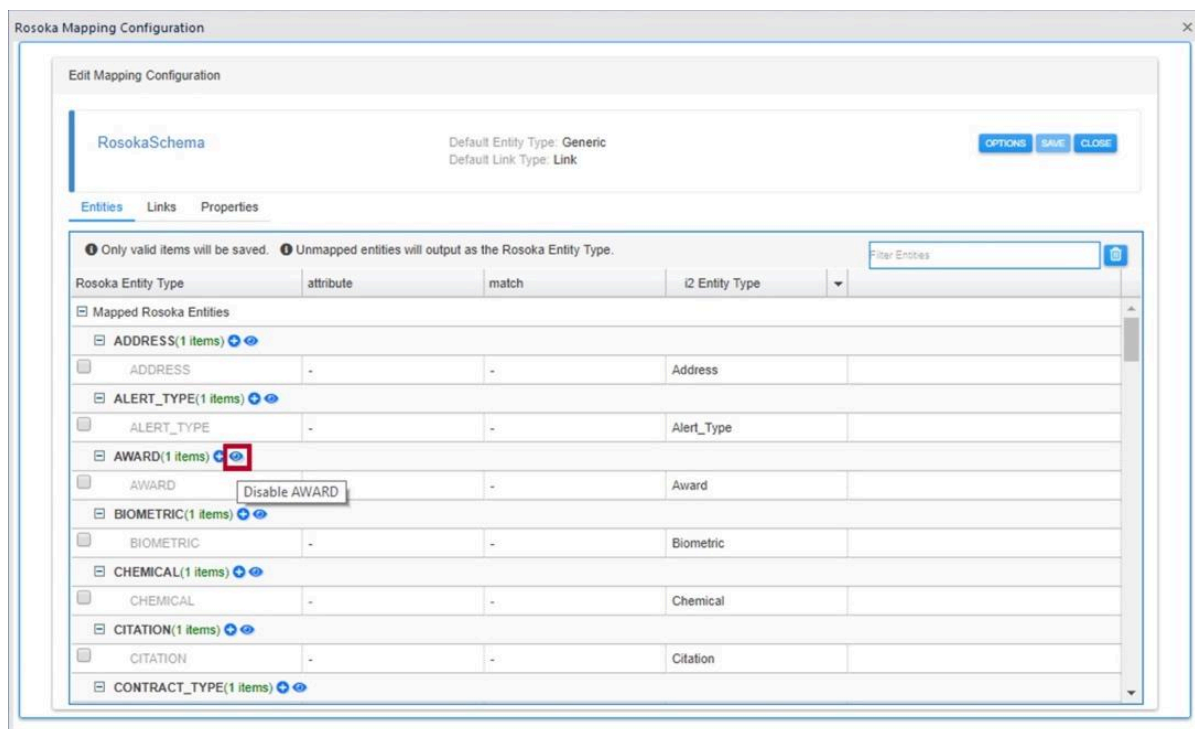
You can expand the link types in the tree view to see lists of the individual links and the predicate types that TextChart identified during processing.

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.

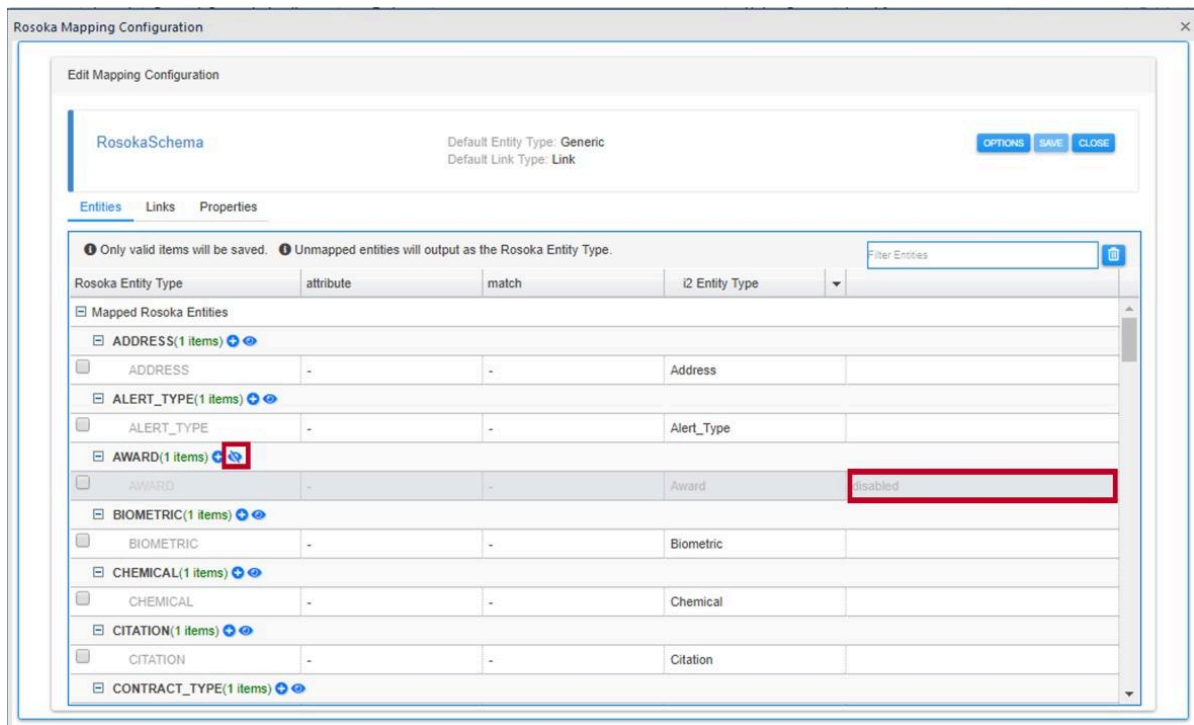
Ignoring TextChart entity, link, or attribute types

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.

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.



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.

Rosoka Document View

Sample_01_AflacGoesPinkForBreastCancerAwarenessMonth_11_16.txt

1

2

3

Aflac Goes Pink for Breast Cancer Awareness Month

Five words or less (NewsUSA) – Not many things in life are certain , but odds are you will unfortunately face at least one health event that requires hospitalization and , following that , rehabilitation .

While hospitals provide an array of helpful rehab services , most people prefer to get well in the comfort of their own homes . Although beneficial in the recovery process , home care can also have its own distinct obstacles . Couple that with a patient's fears and struggles of what 's to come in the weeks and months ahead , and it can create a perfect storm for both patients and caregivers . To help , the Association of Rehabilitation Nurses (ARN) , a nearly 6,000 - member worldwide organization , has created ReSTART Recovery , an online resource that provides information for those who are (or will be) in rehab for everything from strokes to joint replacements to head injuries , and for those who will be caring for them .

When you consider that recent studies cite as many as 75 million Americans suffer from some type of disability , a website such as ReSTART Recovery can make a huge impact on understanding what patients will inevitably experience while on their road to wellness .

The goal , according to ARN , is to get patients who have a disability to a point that they are , once again , as self-sufficient as possible and able to live a full life .

" My clients have been through acute rehab and are back in their communities , " Susan Wirt , a former president of ARN , told The American Nurse Journal in an interview . " I figure out how they can be well and healthy despite their chronic conditions , " she said . Indeed , rehabilitation nurses effectively manage complex health care issues ; collaborate with other professionals and disciplines such as occupational or speech therapists ; provide patients and caregivers with needed education ; set patients ' goals that maximize independence ; and establish plans of care that maintain optimal wellness , according to the ARN .

" Advocacy is also a huge role for us , " Michelle Camica , MSN , CRRN , and former president of ARN , told The American Nurse Journal in the same interview . " We serve as the patient's advocate when addressing issues with other members of the health care team and sometimes with a patient's own family . We always want to make sure patients are getting the right care in the right place at

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

Entities Send All

Expand Filter

- DISEASE (1)
- EVENT (1)
- MEDICAL_PROCEDURE (3)
- Organization (4)
- Person (2)
- PUBLICATION (1)
- URL (1)

Links Send All

Expand Filter

- OrganizationToMedical_Procec
- OrganizationToOrganization (2)
- PersonToOrganization (2)

Aflac Goes Pink for Breast Cancer Awareness Month

Five words or less (**NewsUSA**) – Not many things in life are certain , but odds are you will unfortunately face at least one health event that requires hospitalization and , following that , **rehabilitation** .

While hospitals provide an array of helpful **rehab** services , most people prefer to get well in the comfort of their own homes . Although beneficial in the recovery process , home care can also have its own distinct obstacles . Couple that with a patient's fears and struggles of what 's to come in the weeks and months ahead , and it can create a perfect storm for both patients and caregivers .

To help , the **Association of Rehabilitation Nurses (ARN)** , a nearly 6,000 - member worldwide organization , has created **ReSTART Recovery** , an online resource that provides information for those who are (or will be) in **rehab** for everything from strokes to **joint replacements** to **head injuries** , and for those who will be caring for them .

When you consider that recent studies cite as many as 75 million Americans suffer from some type of disability , a website such as **ReSTART Recovery** can make a huge impact on understanding what patients will inevitably experience while on their road to wellness .

The goal , according to **ARN** , is to get patients who have a disability to a point that they are , once again , as self-sufficient as possible and able to live a full life .

" My clients have been through acute **rehab** and are back in their communities , " **Susan Wirt** , a former president of **ARN** , told **The American Nurse Journal** in an interview . " I figure out how they can be well and healthy despite their chronic conditions , " **she** said .

Indeed , rehabilitation nurses effectively manage complex health care issues ; collaborate with other professionals and disciplines such as occupational or speech therapists ; provide patients and

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.

TextChart Document View

English, 'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career .txt

Entities Send All

Search

- Organization (8)
- Person (12)
 - joaquin guzmán
 - brian m. cogan
 - Gender : male
 - Family Name : cogan
 - rosoka.convention : western
 - First (Given) Name : brian
 - rosoka.second_given_name : m.
 - rosoka.occupation : judge
 - emma coronel aispuro
 - will
 - emma coronel aispuro

Links Send All

Search

- PersonToAddress (16)
- PersonToOrganization (9)
- OrganizationToAddress (3)
- PersonToPerson (11)
- PersonToEvent (5)
 - joaquin guzmán prison
 - charged_with
 - joaquin guzmán prison
 - escape
 - joaquin guzmán smuggle

He was one of the most notorious outlaws of the last 100 years: a brutal Mexican cartel leader, a wily trafficker who smuggled more than \$12 billion worth of drugs and plunged his country into a long-running tragedy of bloodshed and corruption.

But on Wednesday morning, the 30-year criminal career of Joaquín Guzmán Loera, known to the world as El Chapo, reached its final chapter as a federal judge in New York City sentenced him to life in prison.

The life term, mandated by law as a result of the severity of Mr. Guzmán's crimes, was handed down in Federal District Court in Brooklyn, where the kingpin was convicted last winter of drug, murder and money laundering charges after a sprawling three-month trial.

As some of the federal agents who had chased him for years looked on from the gallery, Judge Brian M. Cogan issued the sentence and Mr. Guzmán, 62, was hauled away to prepare himself — pending an appeal — for spending the rest of his life behind bars.

Before he disappeared into a holding cell behind the courtroom, he blew a kiss to his wife, Emma Coronel Aispuro, who attended most of his trial and was implicated in a handful of his crimes.

Although Judge Cogan had no choice but to sentence Mr. Guzmán to life, he noted that the "overwhelming evil" of the drug lord's crimes was readily apparent. Beyond the life sentence — plus an additional 30 years — he ordered him to pay a staggering \$12.6 billion in forfeiture.

Looking disheveled and slightly out of sorts, Mr. Guzmán walked into the eighth-floor courtroom under guard shortly before 9:30 a.m. He wore a loosefitting gray suit, with his tie rakishly askew and a new-growth mustache darkening his upper lip.

Reading from a prepared statement, he said he had not received a fair trial and complained about his solitary confinement in Manhattan's federal jail, calling it "psychological, emotional and mental torture 24 hours a day."

"Since the government of the United States is going to send me to a prison where my name will never be heard again, I take advantage of this opportunity to say there was no justice here," he said.

Though Judge Cogan did not specify where Mr. Guzmán would serve his sentence, he is likely to be sent to the country's most forbidding federal prison, the United States Penitentiary Administrative Maximum Facility, or ADX, in Florence, Colo.

Mr. Guzmán's decades-long career atop the Sinaloa drug cartel, one of Mexico's most powerful criminal mafias, came to a close only after years of international and domestic law enforcement and Mexican authorities.

Toggle Gloss

Brian M. Cogan

Normalized Form

brian m. cogan

Properties

Gender	male
Family Name	cogan
rosoka.convention	western
First (Given) Name	brian

Instances

brian m. cogan

cogan

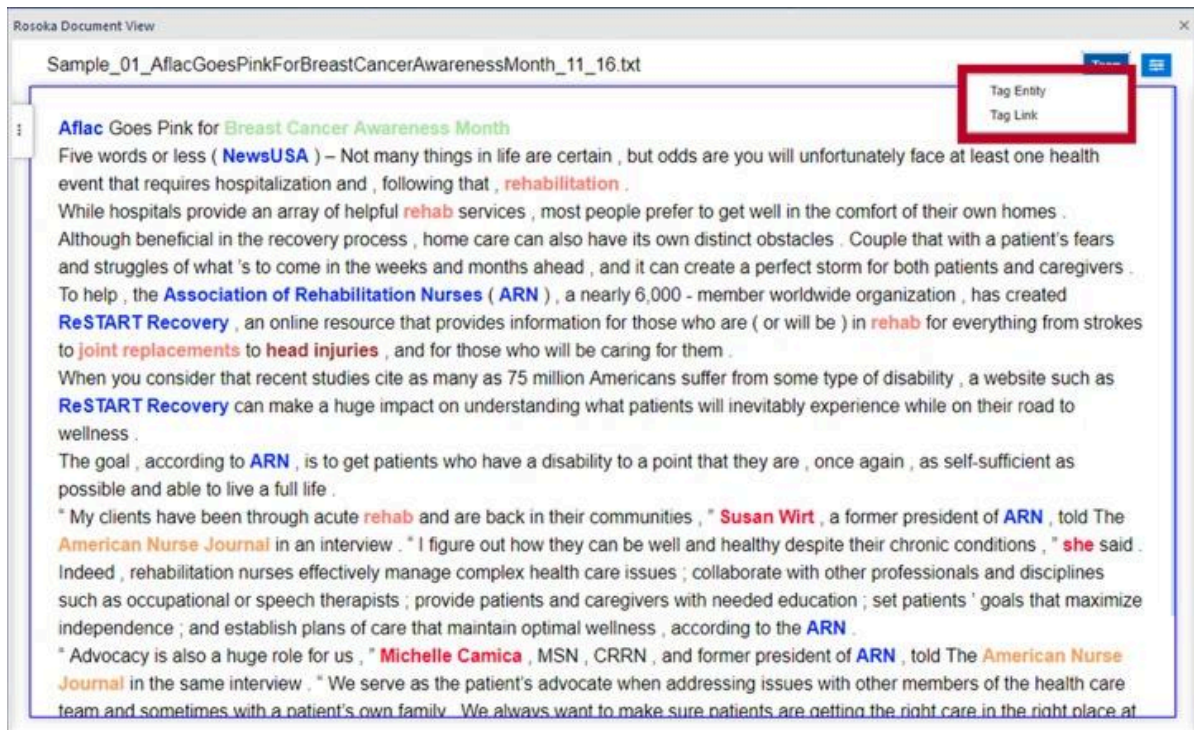
he

Links

PersonToPerson (2)

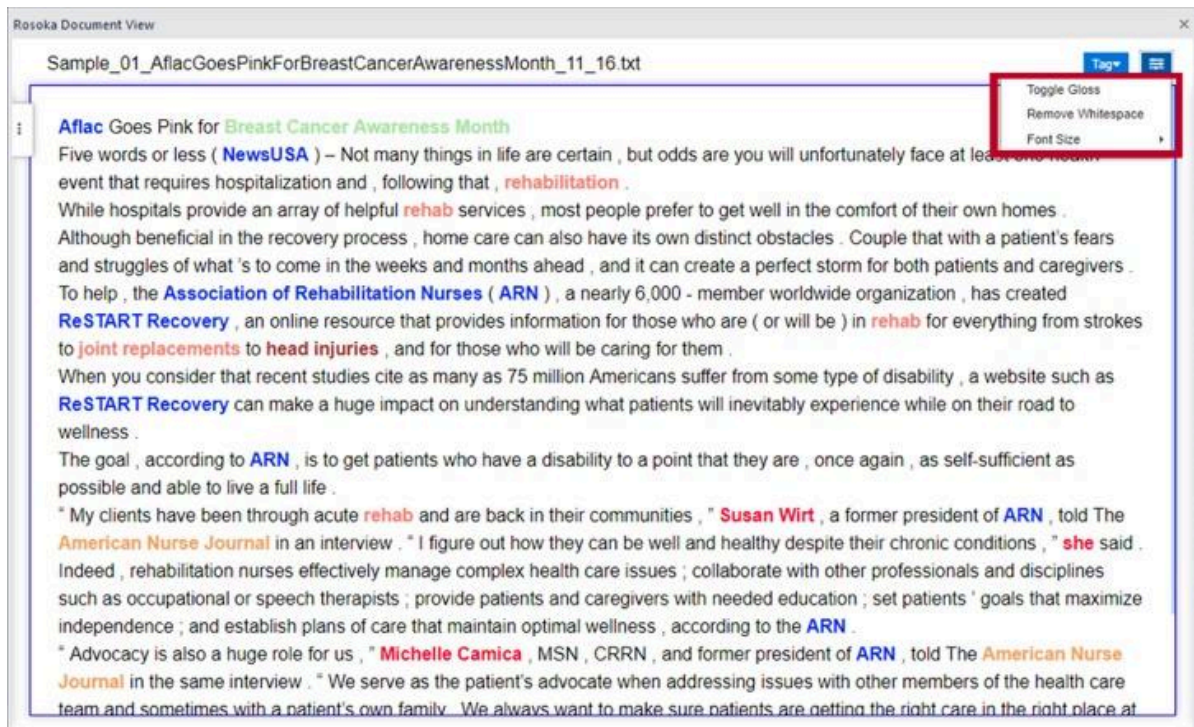
- joaquin guzmán brian m. cogan
- joaquin guzmán brian m. cogan

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



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 an "online resource", which has the Generic entity type.

The screenshot displays the TextChart Document View interface. The central pane shows a document titled "English, 'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career .txt". A context menu is open over the text "New York office", with the "Remove Entity" option highlighted. The left sidebar contains two panels: "Entities" and "Links". The "Entities" panel lists various entities like "united states penitentiary, administrative maximum facility", "new york office", "federal district court", "sinaloa cartel", "mexican government", "drug enforcement", and "administration". The "Links" panel lists various link types like "PersonToAddress (16)", "PersonToOrganization (9)", "OrganizationToAddress (3)", "PersonToPerson (11)", "PersonToEvent (5)", "OrganizationToOrganization (1)", and "AddressToAddress (1)". The right sidebar shows the details for the selected entity "New York office", including its "Entity Type" (Organization), "Normalized Form", "Properties", "Organization Type" (populated), "rosoka.admin_region" (new york), "Instances" (new york office), and "Links" (PersonToOrganization (1), OrganizationToOrganization (1)).

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**.

TextChart Document View

English_ 'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career .txt

Entities [Send All](#)

Search

- ▶ united states penitentiary, administrative maximum facility
- ▶ new york office
- ▶ federal district court
- ▶ sinaloa cartel
- ▶ mexican government
- ▶ drug enforcement administration

Links [Send All](#)

Search

- ▶ PersonToAddress (16)
- ▶ PersonToOrganization (9)
- ▶ OrganizationToAddress (3)
- ▶ PersonToPerson (11)
- ▶ PersonToEvent (5)
- ▶ OrganizationToOrganization (1)
- ▶ AddressToAddress (1)

'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career

He sent hundreds of tons of drugs to the **United States** from **Mexico** and caused the **brutal deaths** of dozens of people.

He was one of the most **dangerous** men in **the world** for over 100 years: a brutal Mexican cartel leader, **he** smuggled more than \$12 billion worth of drugs and plunged **his** country into a long-running tragedy of bloodshed and corruption.

But on Wednesday morning, the 30-year criminal career of **Joaquín Guzmán Loera**, known to the world as **El Chapo**, reached its final chapter as a federal judge in **New York City** sentenced **him** to life in prison.

The life term, mandated by law as a result of the severity of Mr. **Guzmán's** crimes, was handed down in **Federal District Court** in **Brooklyn**, where the kingpin was convicted last winter of drug, **murder** and **money laundering** charges after a sprawling three-month trial.

As some of the federal agents who had chased **him** for years looked on from the gallery, Judge **Brian M. Cogan** issued the sentence and Mr. **Guzmán**, 62, was hauled away to prepare **himself** — pending an appeal — for spending the rest of **his** life behind bars.

Before **he** disappeared into a holding cell behind the courtroom, **he** blew a kiss to **his** wife, **Emma Coronel Aispuro**, who attended most of **his** trial and was implicated in a handful of **his** crimes.

Toggle Gloss

New York office

Entity Type: Organization

Normalized Form

new york office

Properties

Organization Type populate

rosoka.admin_region new york

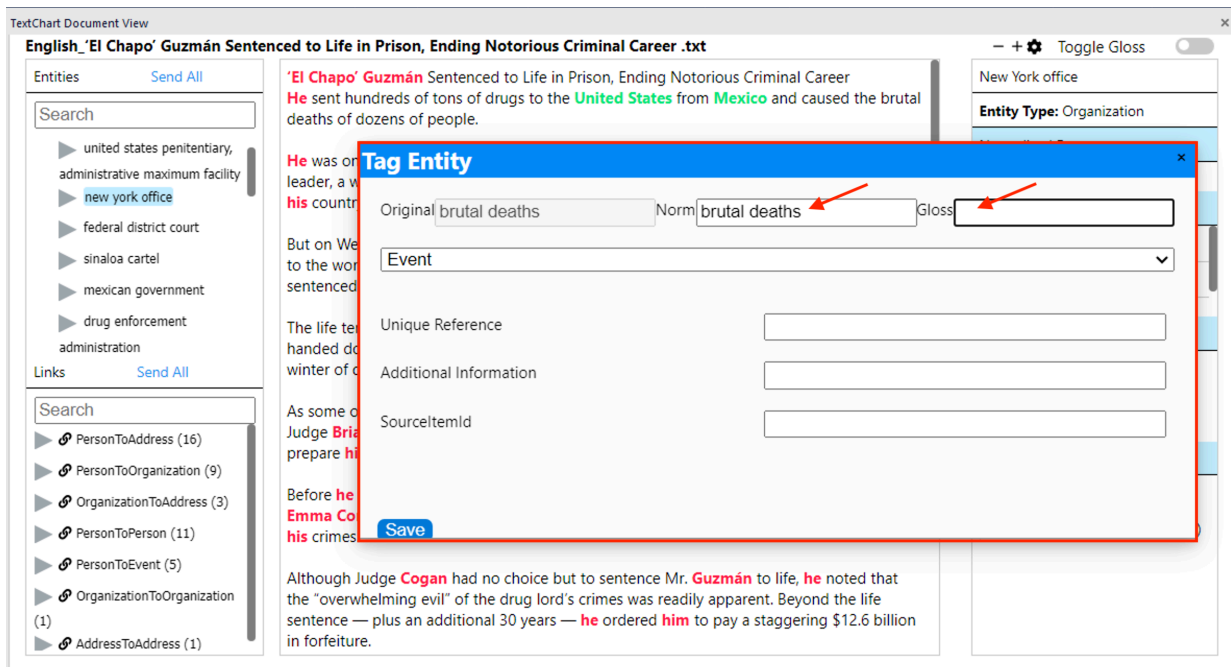
Instances

new york office

Links

- ▶ PersonToOrganization (1)
- ▶ OrganizationToOrganization (1)

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.



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.

English: 'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career .txt

Entities [Send All](#)

Search

- mexico, distrito federal, mx
- new york city
- brooklyn, florida, us
- reading, illinois, us
- manhattan, new york, us
- florence, colorado, us
- brooklyn heights neighborhood

Links [Send All](#)

Search

- PersonToAddress (16)
- PersonToOrganization (9)
- OrganizationToAddress (3)
- PersonToPerson (11)
- PersonToEvent (5)
- OrganizationToOrganization (1)
- AddressToAddress (1)

'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career

He sent hundreds of tons of drugs to the **United States** from **Mexico** and caused the brutal deaths of dozens of people.

He was one of the most notorious outlaws of the last 100 years: a brutal Mexican cartel leader, a wily trafficker who **smuggled** more than \$12 billion worth of drugs and plunged **his** country into a long-running tragedy of bloodshed and corruption.

But on Wednesday morning, the 30-year criminal career of **Joaquín Guzmán Loera**, known to the world as **El Chapo**, reached its final chapter as a federal judge in **New York City** sentenced **him** to life in prison.

The life term, mandated by law as a result of the severity of Mr. **Guzmán**'s crimes, was handed down in **Federal District Court** in **Brooklyn**, where the kingpin was convicted last winter of drug, **murder** and **money laundering** charges after a sprawling three-month trial.

As some of the federal agents who had chased **him** for years looked on from the gallery, Judge **Brian M. Cogan** issued the sentence and Mr. **Guzmán**, 62, was hauled away to prepare **himself** — pending an appeal — for spending the rest of **his** life behind bars.

Before **he** disappeared into a holding cell behind the courtroom, **he** blew a kiss to **his** wife, **Emma Coronel Aispuro**, who attended most of **his** trial and was implicated in a handful of **his** crimes.

Although Judge **Cogan** had no choice but to sentence Mr. **Guzmán** to life, **he** noted that the "overwhelming evil" of the drug lord's crimes was readily apparent. Beyond the life sentence — plus an additional 30 years — **he** ordered **him** to pay a staggering \$12.6 billion in forfeiture.

Toggle Gloss

New York City

Entity Type: Address

Normalized Form

new york city

Properties

Location Type populated

Town/City new york city

Building Number new york city

Instances

new york city

Links

- PersonToOrganization (1)
- PersonToAddress (2)

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.

English: 'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career .txt

Entities [Send All](#)

Search

- mexico, distrito federal, mx
- new york city
- brooklyn, florida, us
- reading, illinois, us
- manhattan, new york, us
- florence, colorado, us
- brooklyn heights neighborhood

Links [Send All](#)

Search

- PersonToAddress (16)
- PersonToOrganization (9)
- OrganizationToAddress (3)
- PersonToPerson (11)
- PersonToEvent (5)
- OrganizationToOrganization (1)
- AddressToAddress (1)

Although Judge **Cogan** had no choice but to sentence Mr. **Guzmán** to life, **he** noted that the "overwhelming evil" of the drug lord's crimes was readily apparent. Beyond the life sentence — plus an additional 30 years — **he** ordered **him** to pay a staggering \$12.6 billion in forfeiture.

His ability to persistently evade capture — and then **escape from prison** after **he** was caught — underscored the deep corruption of the **Mexican government** by **his** cartel,

Edit Entity

Original: **Manhattan** Norm: **manhattan, new york, us** Gloss:

Unique Reference:

Location Type: populated

Apartment Number:

Building Number: manhattan, new york, us

[Save](#)

Toggle Gloss

Manhattan

Entity Type: Address

Normalized Form

manhattan, new york, us

Properties

Location Type populated

Town/City manhattan, new york, us

Building Number manhattan, new york, us

Instances

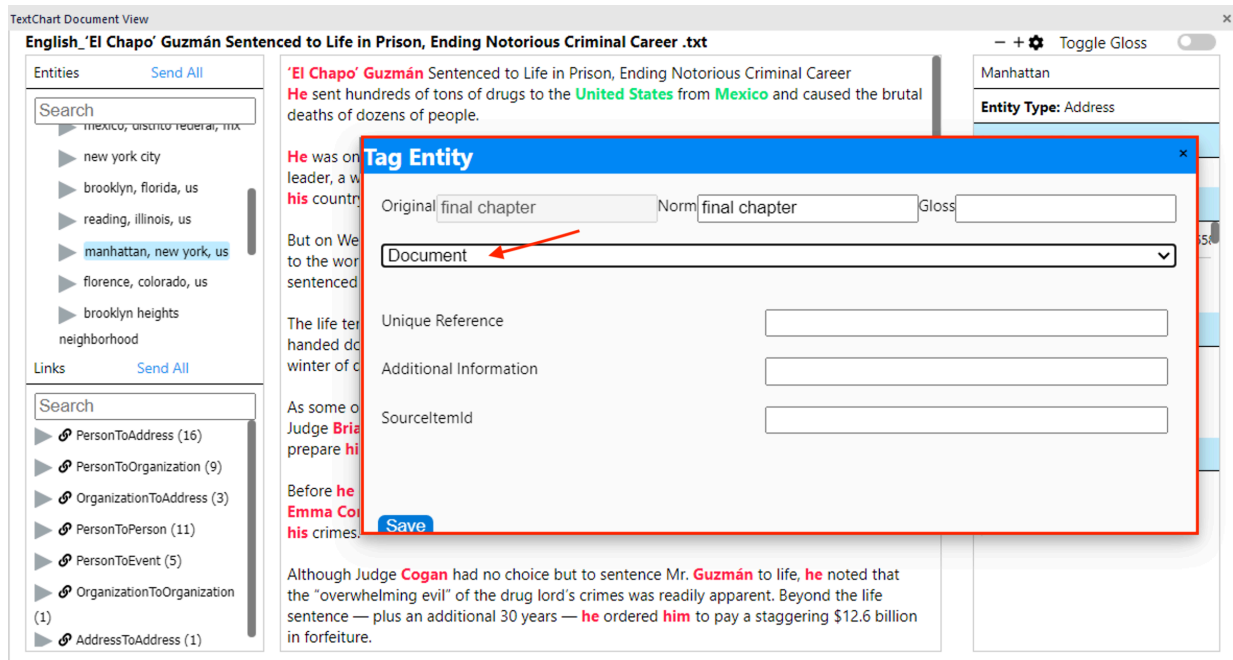
manhattan, new york, us

Links

- PersonToOrganization (1)
- PersonToAddress (2)

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.

TextChart Document View

English_‘El Chapo’ Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career .txt

Entities Send All

- mexico, distrito federal, mx
- new york city
- brooklyn, florida, us
- reading, illinois, us
- manhattan, new york, us
- florence, colorado, us
- brooklyn heights neighborhood

Links Send All

- PersonToAddress (16)
- PersonToOrganization (9)
- OrganizationToAddress (3)
- PersonToPerson (11)
- PersonToEvent (5)
- OrganizationToOrganization (1)
- AddressToAddress (1)

‘El Chapo’ Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career
He sent hundreds of tons of drugs to the United States from Mexico and caused the brutal deaths of dozens of people.

He was one of the most notorious outlaws of the last 100 years: a brutal Mexican cartel leader, a wily trafficker who smuggled more than \$12 billion worth of drugs and plunged his country into a long-running tragedy of bloodshed and corruption.

But on Wednesday morning, the 30-year criminal career of Joaquín Guzmán Loera came to the world as El Chapo, reached its final chapter as a federal judge sentenced him to life in prison.

The life term, mandated by law as a result of the severity of Mr. Guzmán's crimes, was handed down in Federal District Court in Brooklyn, where the kingpin spent the winter of drug, murder and money laundering charges after a stay in New York City.

As some of the federal agents who had chased him for years looked on, Judge Brian M. Cogan issued the sentence and Mr. Guzmán, 62, prepared himself — pending an appeal — for spending the rest of his life behind bars.

Before he disappeared into a holding cell behind the courtroom, he blew a kiss to his wife, Emma Coronel Aispuro, who attended most of his trial and was implicated in a handful of his crimes.

Although Judge Cogan had no choice but to sentence Mr. Guzmán to life, he noted that the “overwhelming evil” of the drug lord’s crimes was readily apparent. Beyond the life sentence — plus an additional 30 years — he ordered him to pay a staggering \$12.6 billion in forfeiture.

Toggle Gloss

Manhattan	
Entity Type:	Address
Normalized Form	
manhattan, new york, us	
Properties	
Longitude	-73.966247558
rosoka.ufi	971945
Instances	
manhattan	
Links	
PersonToOrganization (1)	
PersonToAddress (1)	

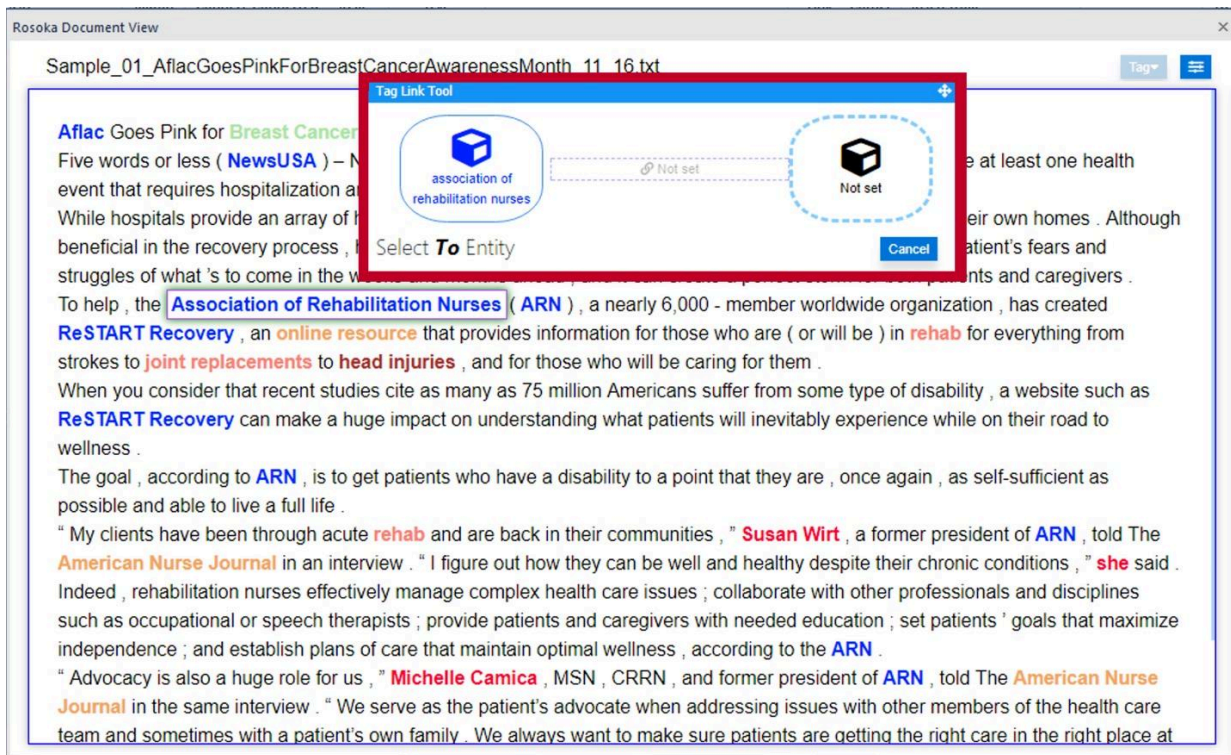
Context menu options over the word "him":

- Edit Entity
- Create Link from Entity
- Remove Entity
- Make Anchor
- Sent to Chart

If you used the **Tag Link** command, 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.



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.

Rosoka Document View

Sample_01_AflacGoesPinkForBreastCancerAwarenessMonth_11_16.txt

Tag Link Tool

association of rehabilitation nurses

restart recovery

Select Type

Select Type

Alert Feed Subscription

Link

Not set

Aflac Goes Pink for Breast Cancer Awareness Month 11-16

Five words or less (NewsUSA) – N

event that requires hospitalization an

While hospitals provide an array of h

beneficial in the recovery process , f

struggles of what 's to come in the w

To help , the Association of Rehab

ReSTART Recovery , an online re

strokes to joint replacements to head injuries , and for those who will be caring for them .

When you consider that recent studies cite as many as 75 million Americans suffer from some type of disability , a website such as

ReSTART Recovery can make a huge impact on understanding what patients will inevitably experience while on their road to

wellness .

The goal , according to ARN , is to get patients who have a disability to a point that they are , once again , as self-sufficient as

possible and able to live a full life .

" My clients have been through acute rehab and are back in their communities , " Susan Wirt , a former president of ARN , told The

American Nurse Journal in an interview . " I figure out how they can be well and healthy despite their chronic conditions , " she said .

Indeed , rehabilitation nurses effectively manage complex health care issues ; collaborate with other professionals and disciplines

such as occupational or speech therapists ; provide patients and caregivers with needed education ; set patients ' goals that maximize

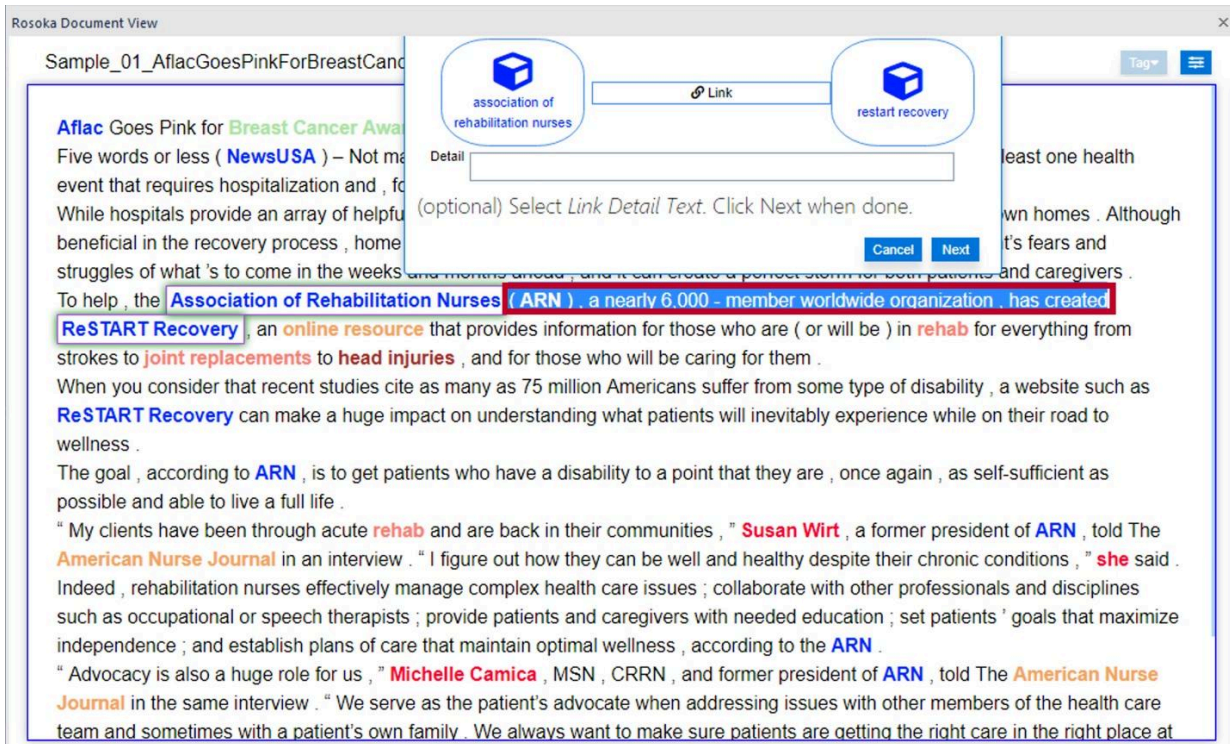
independence ; and establish plans of care that maintain optimal wellness , according to the ARN .

" Advocacy is also a huge role for us , " Michelle Camica , MSN , CRRN , and former president of ARN , told The American Nurse

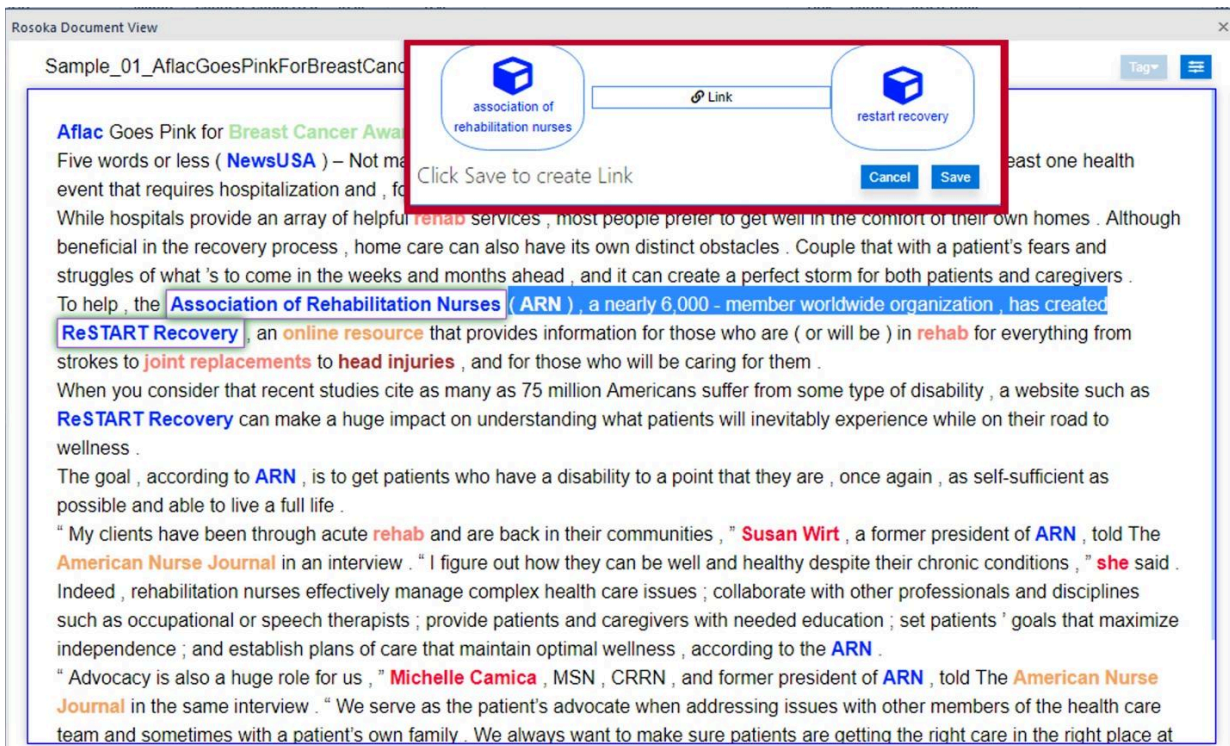
Journal in the same interview . " We serve as the patient's advocate when addressing issues with other members of the health care

team and sometimes with a patient's own family . We always want to make sure patients are getting the right care in the right place at

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



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



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**.

The screenshot displays the i2 TextChart Premium interface. The main window shows a document titled "English_ 'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career .txt". The left sidebar contains a list of entities: Event (6), Property (1), Organization (8), Person (12), and Address (17). The right sidebar shows a detailed view of an entity, "Manhattan", with fields for Entity Type (Address), Normalized Form, Properties, Longitude, and Instances. A context menu is open over the word "smuggled" in the main text, with options: Edit Entity, Create Link from Entity, Remove Entity, Make Anchor, and Send to Chart.

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.

The screenshot displays the i2 TextChart Premium interface with the "Edit Entity" dialog box open. The dialog box has a title bar "Edit Entity" and a close button. It contains the following fields: Original (smuggled), Norm (smuggle), Gloss (empty), Unique Reference (empty), Additional Information (Investigated), and SourceId (empty). A "Save" button is at the bottom left of the dialog. The background shows the same document and sidebar as the previous screenshot.

TextChart Document View

English, 'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career .txt

Entities [Send All](#)

Search

- Event (6)
- Drug (1)
- Facility (2)
- Money (3)
- Organization (6)
- Individual (12)
 - joaquin guzmán
 - brian m. cogan
 - emma coronel aispuro
 - will

Links [Send All](#)

Search

- IndividualToGeospatial Location (16)
- IndividualToMoney (2)
- IndividualToOrganization (7)
- OrganizationToGeospatial Location (2)
- IndividualToIndividual (11)
- IndividualToFacility (2)
- FacilityToGeospatial Location (1)
- IndividualToEvent (5)

Entity Details (Right Sidebar):

Entity Type: Individual

Normalized Form

joaquin guzmán

Properties

Gender	male
Type	criminal
First (Given) Name	joaquin
Family Name	guzmán

Instances

- 'el chapo' guzmán
- he
- his
- joaquin guzmán loera

Links (Right Sidebar):

- IndividualToGeospatial Location (12)
- IndividualToMoney (2)
- IndividualToOrganization (5)
- IndividualToIndividual (10)
- IndividualToFacility (1)
- IndividualToEvent (3)

Document Text:

'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career

He sent hundreds of tons of drugs to the United States from Mexico and caused the brutal deaths of dozens of people.

He was one of the most notorious outlaws of the last 100 years: a brutal Mexican cartel leader, a wily trafficker who smuggled more than \$12 billion worth of drugs and plunged his country into a long-running tragedy of bloodshed and corruption.

But on Wednesday morning, the 30-year criminal career of Joaquín Guzmán Loera, known to the world as El Chapo, reached its final chapter as a federal judge in New York City sentenced him to life in prison.

The life term, mandated by law as a result of the severity of Mr. Guzmán's crimes, was handed down in Federal District Court in Brooklyn, where the kingpin was convicted last winter of drug, murder and money laundering charges after a sprawling three-month trial.

As some of the federal agents who had chased him for years looked on from the gallery, Judge Brian M. Cogan issued the sentence and Mr. Guzmán, 62, was hauled away to prepare himself — pending an appeal — for spending the rest of his life behind bars.

Before he disappeared into a holding cell behind the courtroom, he blew a kiss to his wife, Emma Coronel Aispuro, who attended most of his trial and was implicated in a handful of his crimes.

Although Judge Cogan had no choice but to sentence Mr. Guzmán to life, he noted that the "overwhelming evil" of the drug lord's crimes was readily apparent. Beyond the life sentence — plus an additional 30 years — he ordered him to pay a staggering \$12.6 billion in forfeiture.

Looking disheveled and slightly out of sorts, Mr. Guzmán walked into the eighth-floor courtroom under guard shortly before 9:30 a.m. He wore a loosefitting gray suit, with his tie rakishly askew and a new-growth mustache darkening his upper lip.

Reading from a prepared statement, he said he had not received a fair trial and complained about his solitary confinement in Manhattan's federal jail, calling it "psychological, emotional and mental torture 24 hours a day."

"Since the government of the United States is going to send me to a prison where my name will never

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.

TextChart Document View

English: 'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career .txt

Entities [Send All](#)

Search

- Event (5)
- Drug (1)
- Facility (2)
- Money (3)
- Organization (6)
- Individual (12)
 - joaquin guzmán
 - brian m. cogan
 - emma coronel aispuro
 - will

Links [Send All](#)

Search

- IndividualToGeospatial Location (16)
 - joaquin guzmán united states of america, united states, us
 - joaquin guzmán new york city
 - joaquin guzmán new york city
 - joaquin guzmán reading, illinois, us
 - joaquin guzmán manhattan, new york, us

Entity Details: 'El Chapo' Guzmán

Entity Type: Individual

Normalized Form: joaquin guzmán

Properties:

Gender	male
Type	criminal
First (Given) Name	joaquin
Family Name	guzmán

Instances:

- 'el chapo' guzmán
- he
- his
- joaquin guzmán loera

Links:

- IndividualToGeospatial Location (12)
- IndividualToMoney (2)
- IndividualToOrganization (5)
- IndividualToIndividual (10)
- IndividualToFacility (1)
- IndividualToEvent (3)

Document Text:

'El Chapo' Guzmán Sentenced to Life in Prison, Ending Notorious Criminal Career

He sent hundreds of tons of drugs to the United States from Mexico and caused the brutal deaths of dozens of people.

He was one of the most notorious outlaws of the last 100 years: a brutal Mexican cartel leader, a wily trafficker who smuggled more than \$12 billion worth of drugs and plunged his country into a long-running tragedy of bloodshed and corruption.

But on Wednesday morning, the 30-year criminal career of Joaquín Guzmán Loera, known to the world as El Chapo, reached its final chapter as a federal judge in New York City sentenced him to life in prison.

The life term, mandated by law as a result of the severity of Mr. Guzmán's crimes, was handed down in Federal District Court in Brooklyn, where the kingpin was convicted last winter of drug, murder and money laundering charges after a sprawling three-month trial.

As some of the federal agents who had chased him for years looked on from the gallery, Judge Brian M. Cogan issued the sentence and Mr. Guzmán, 62, was hauled away to prepare himself — pending an appeal — for spending the rest of his life behind bars.

Before he disappeared into a holding cell behind the courtroom, he blew a kiss to his wife, Emma Coronel Aispuro, who attended most of his trial and was implicated in a handful of his crimes.

Although Judge Cogan had no choice but to sentence Mr. Guzmán to life, he noted that the "overwhelming evil" of the drug lord's crimes was readily apparent. Beyond the life sentence — plus an additional 30 years — he ordered him to pay a staggering \$12.6 billion in forfeiture.

Looking disheveled and slightly out of sorts, Mr. Guzmán walked into the eighth-floor courtroom under guard shortly before 9:30 a.m. He wore a loosefitting gray suit, with his tie rakishly askew and a new-growth mustache darkening his upper lip.

Reading from a prepared statement, he said he had not received a fair trial and complained about his solitary confinement in Manhattan's federal jail, calling it "psychological, emotional and mental torture 24 hours a day."

"Since the government of the United States is going to send me to a prison where my name will never

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.

The screenshot displays the Rosoka Document View interface. The main window shows a text document titled "Sample_01_AflacGoesPinkForBreastCancerAwarenessMonth_11_16.txt". A section of the text is highlighted in blue, and a context menu is open over it, showing options like "Send To Chart" and "Select on Chart". The left sidebar contains a tree view of entities and links. The right sidebar shows a detailed view of the selected entity, "association of rehabilitation nurses", including its normalized form, entity type, properties, instances, and links.

Entities Send All

Expand Filter...

- Disease (1)
- Event (1)
- Medical_procedure (3)
- Org (5)
 - aflac
 - arn
 - association of rehabilitation nurses
 - newsusa
 - restart recovery
- Person (2)
- Publication (2)

Links Send All

Expand Filter...

- DiseaseToMedical_procedure (1)
- Medical_procedureToMedical_procedure (1)
- OrgToDisease (1)
- OrgToEvent (2)
- OrgToMedical_procedure (3)
- OrgToOrg (3)
- OrgToPublication (2)
- PersonToMedical_procedure (1)
- PersonToOrg (2)
- PersonToPublication (2)
- PersonToUrl (1)

association of rehabil...

Send To Chart Remove Entity

Normalized Form

association of rehabilitation nurses

Entity Type

Org

Properties

subtype medical

Instances

association of rehabilitation nurses

Links

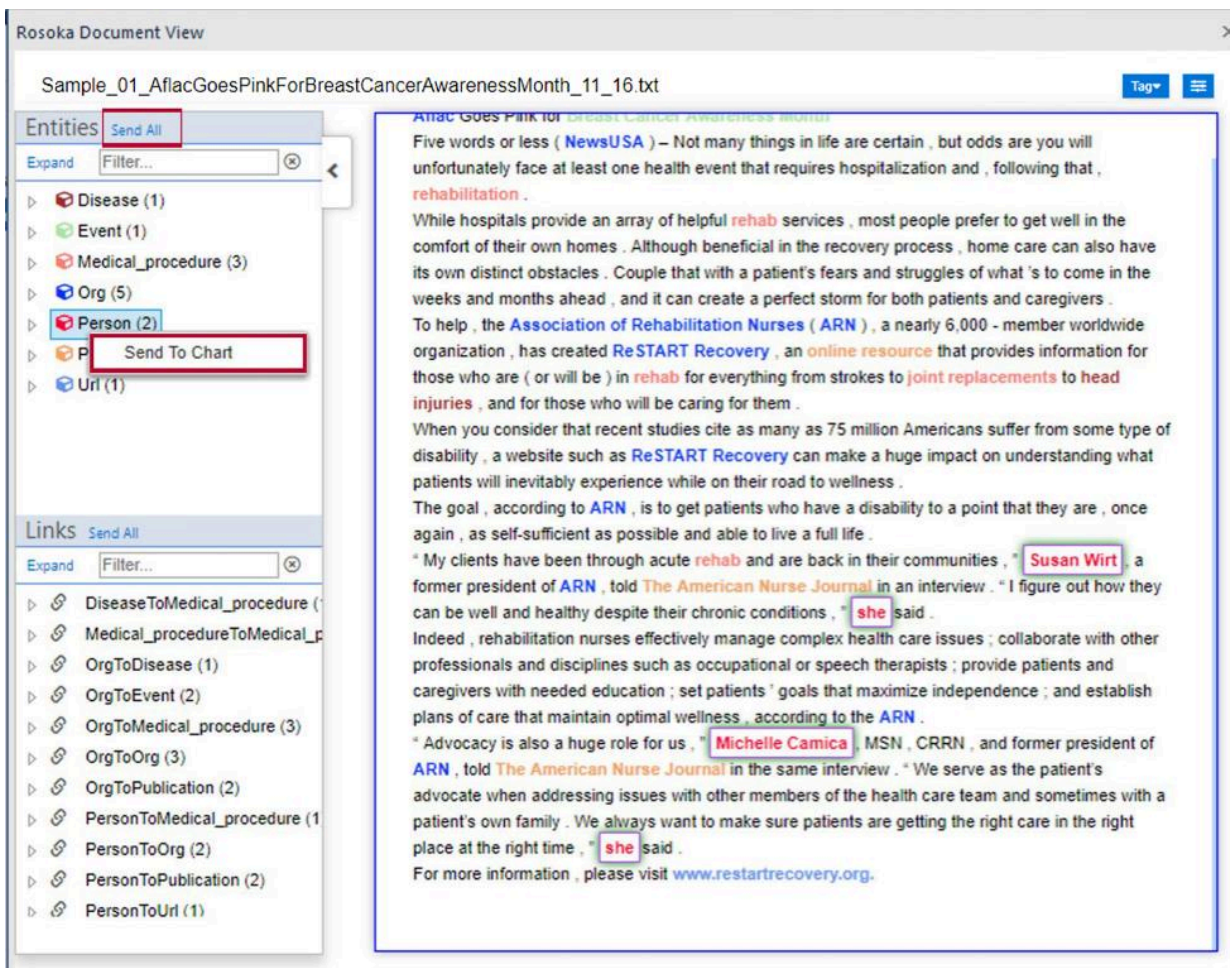
- OrgToMedical_procedure (1)
- OrgToOrg (1)

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



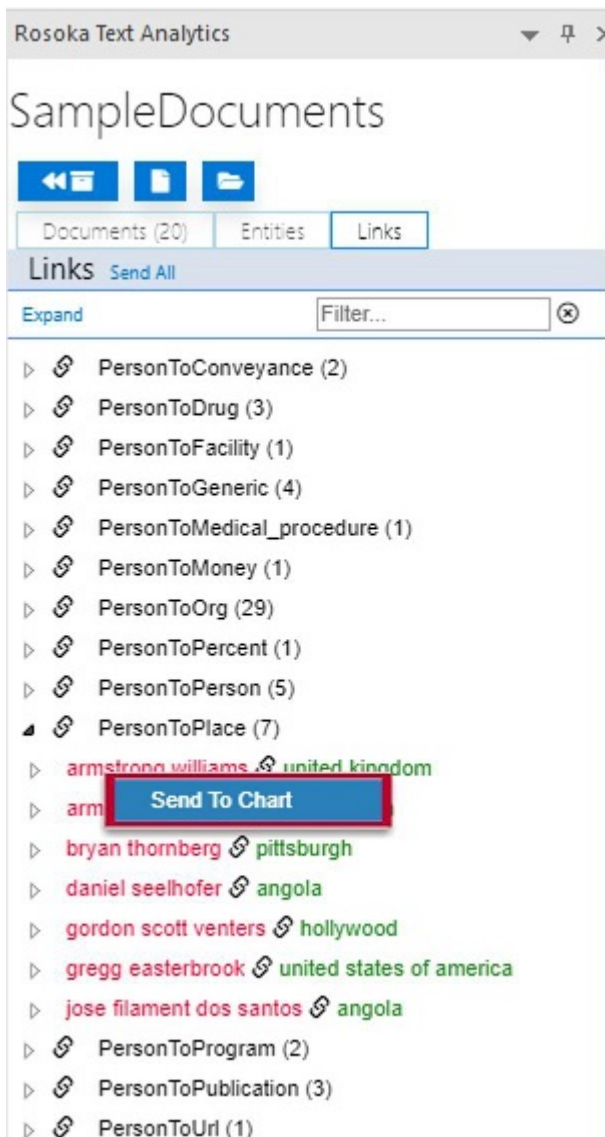
Sending several entities

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.



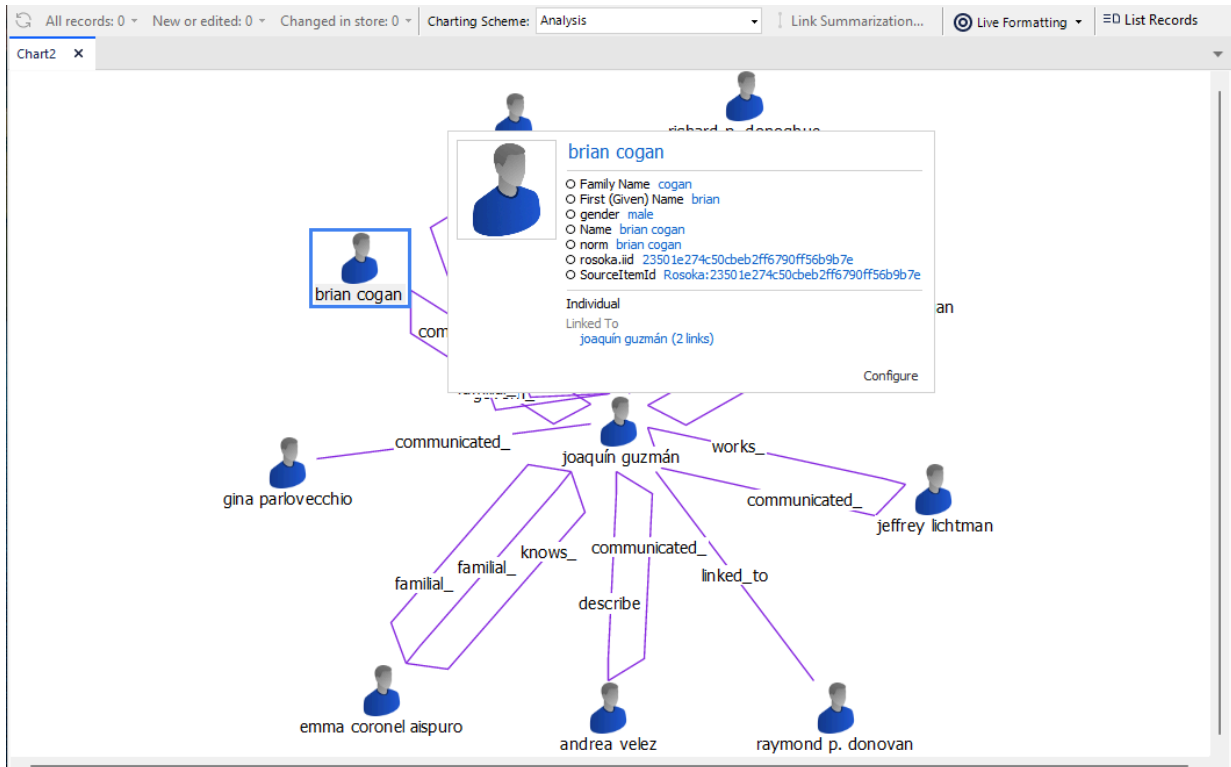
Sending Links To A Chart

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.

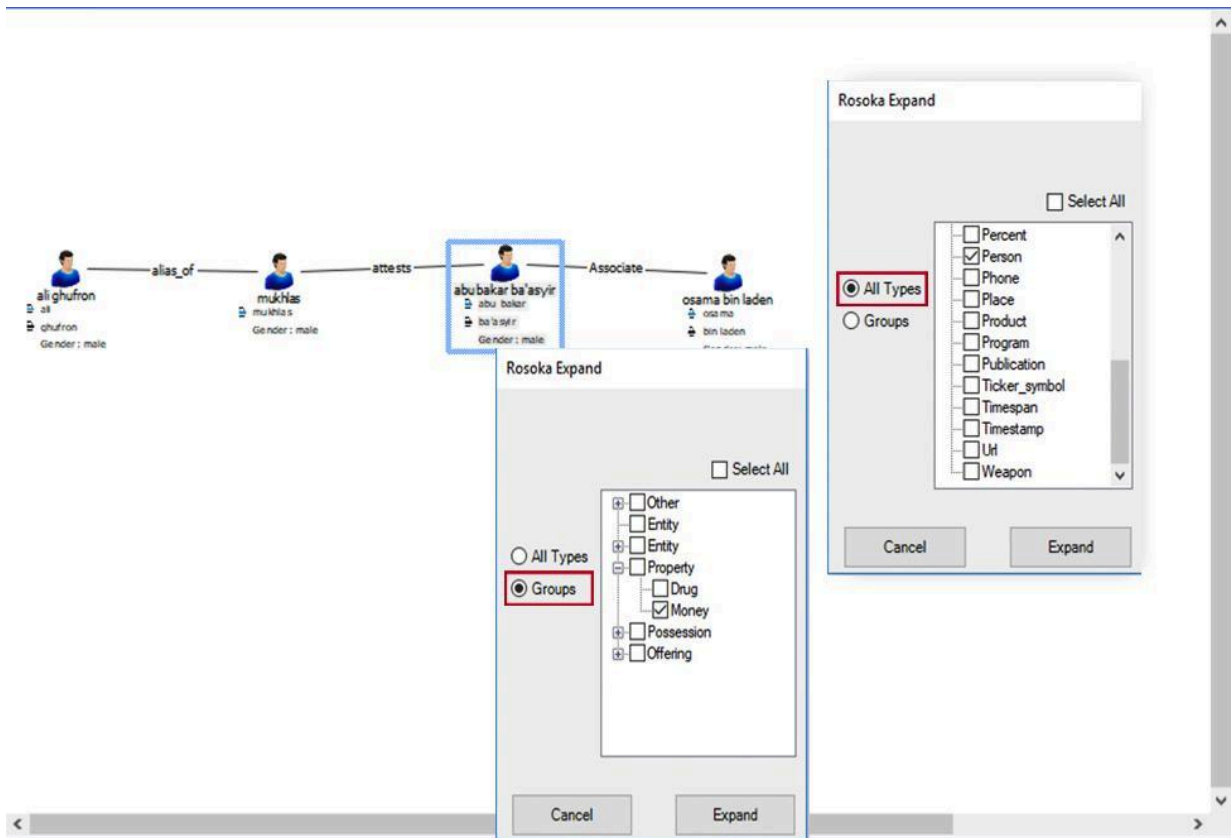


Viewing Entities On The Chart Surface

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.

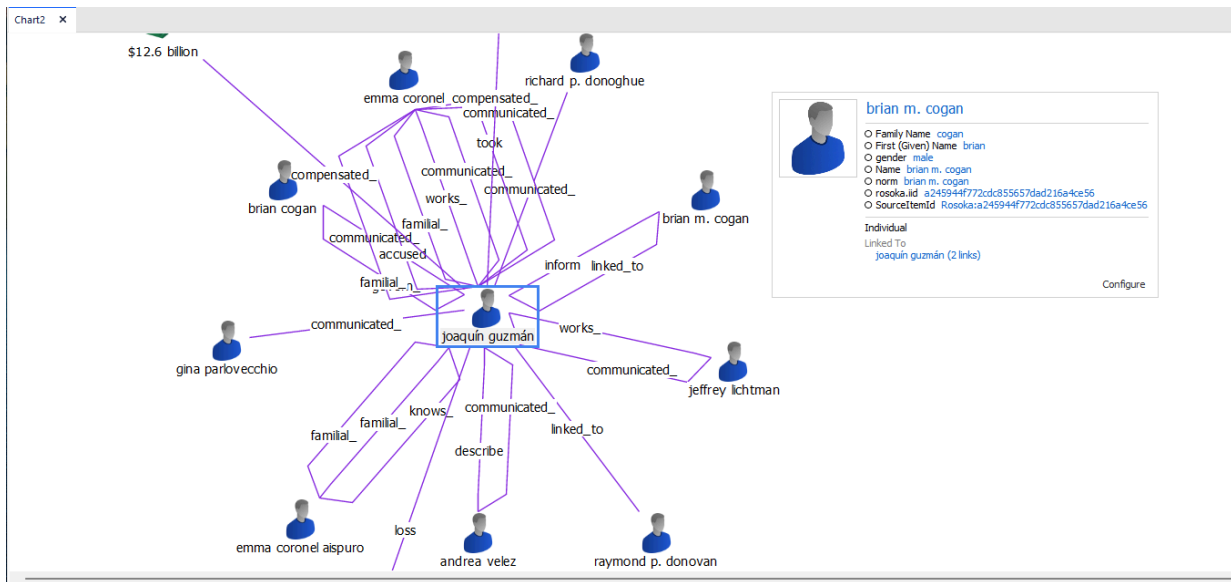


After you've added an entity *from TextChart* to the chart, you can use it to start adding related entities from the other documents that you've processed. Right-click the entity, and select **Rosoka Expand**.

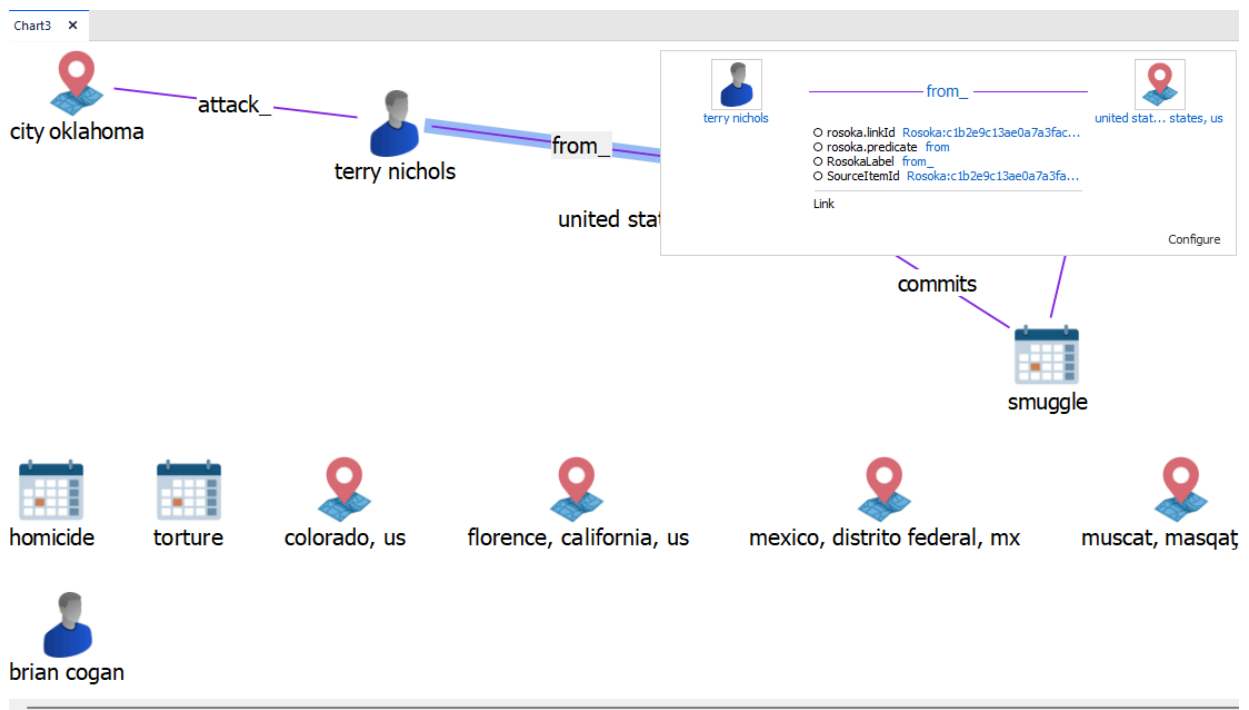


You can choose to add related entities of particular types, or groups of types; or you can use **Select All** to add all related entities in a single action.

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.



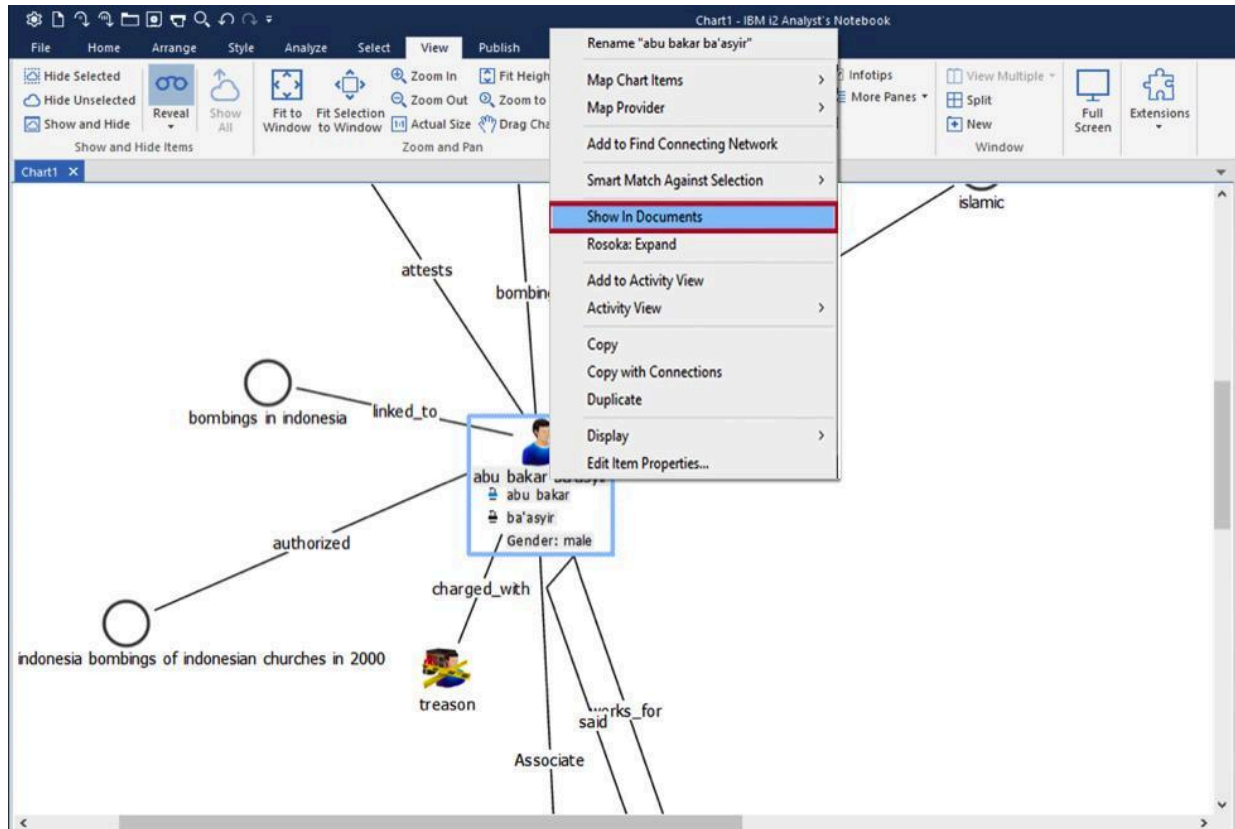
Similarly, you can hover over a TextChart link to highlight the connection between two entities and display additional extraction information.



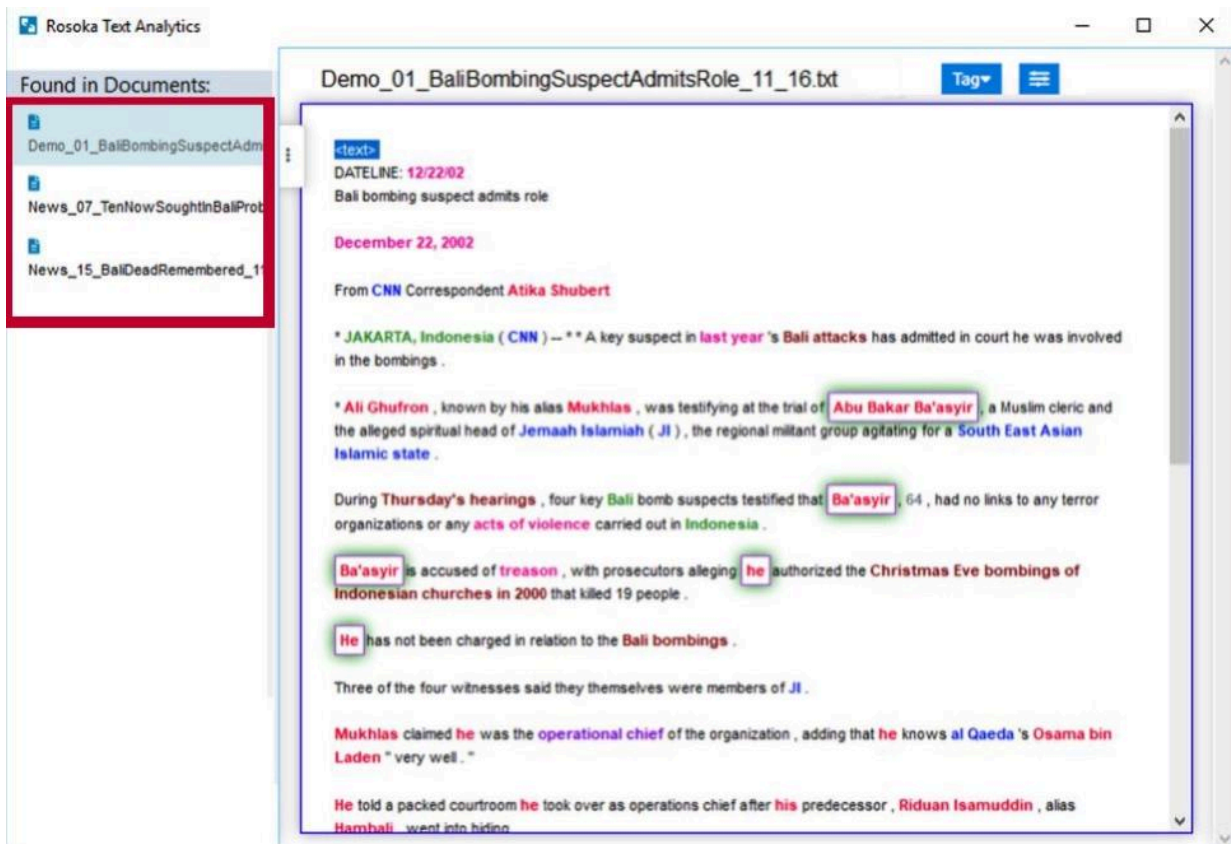
Listing Source Documents

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.

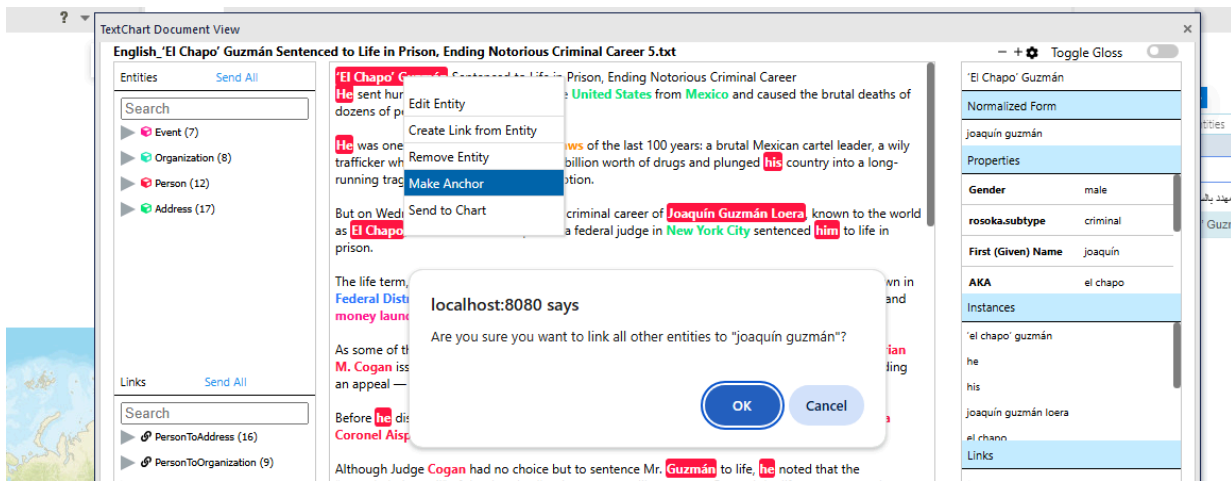


Creating an Anchor Entity

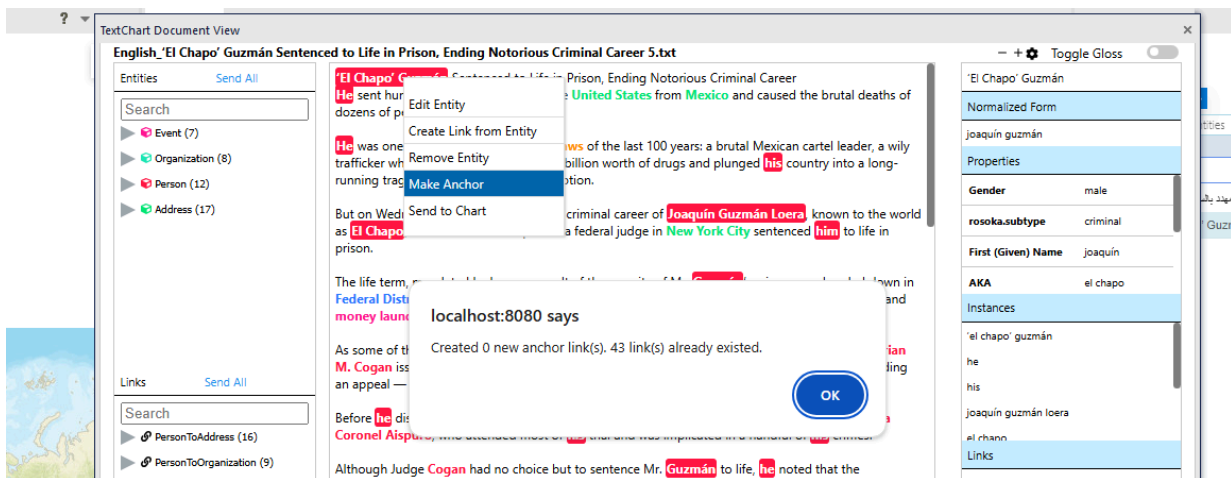
The **Anchor Entity** feature allows users to quickly create links between a selected entity and all other entities in a document. This feature is designed to establish a "hub" entity that connects to multiple other entities, creating a star-pattern relationship graph.

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:

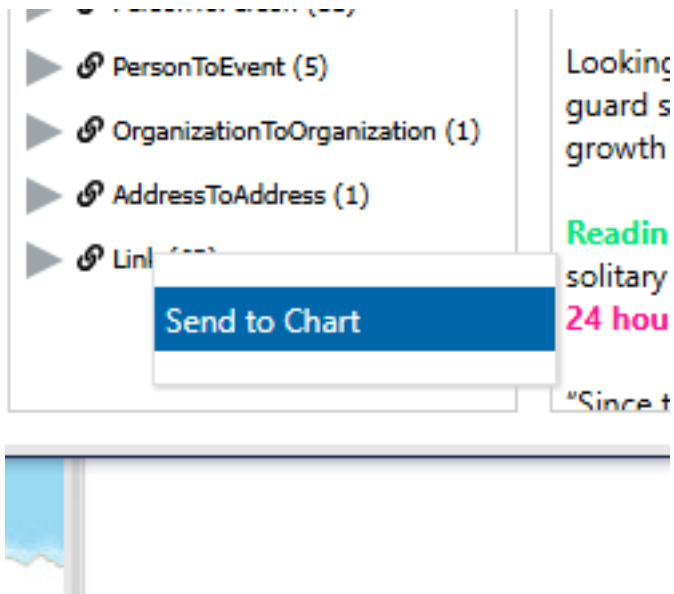


1. **Review Results:** After confirming, an alert will display how many new links were created



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).



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.

Use i2 Analyze Security Schema?

DISABLED

Requires confirmation to enable/disable.

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.

[test] Configuration Save Changes Cancel Changes

Assets Extraction Output

i2SQL Output Configure

i2SQL Output Configuration

Mapping Configuration Default Upload

Doc Storage Directory optional, leave blank for no doc storage

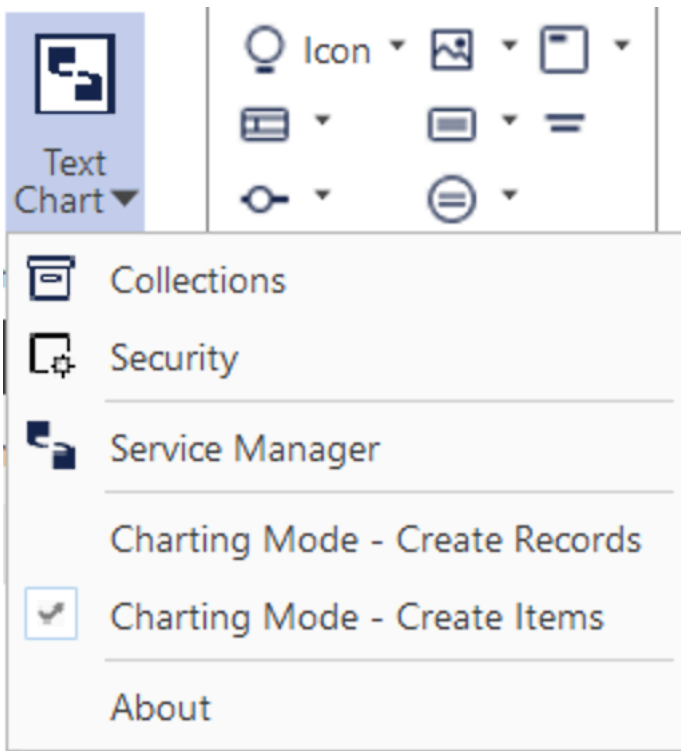
Security Schema Enabled optional Disabled

Default Security ID's optional, leave blank for id no default

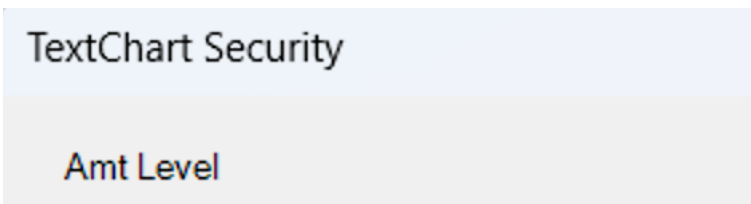
Save Cancel

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.



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.

- 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.

The screenshot shows the 'Output' configuration tab for 'I2SQLOutput'. It includes a 'Configure' button and a 'Save Changes' button. The configuration fields are:

- Mapping Configuration:** mappingconf.zip (with an 'Upload' button)
- Doc Storage Directory:** optional, leave blank for no doc storage (text input field)
- Security Schema Enabled:** optional (set to 'Enabled' via a dropdown)
- Default Security ID's:** optional, leave blank for id no default (text input field containing 'SD-SC=OSI;SD-SL=UC')

At the bottom left, there are 'Save' and 'Cancel' buttons.

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:

- 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.

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:

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.

- **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:

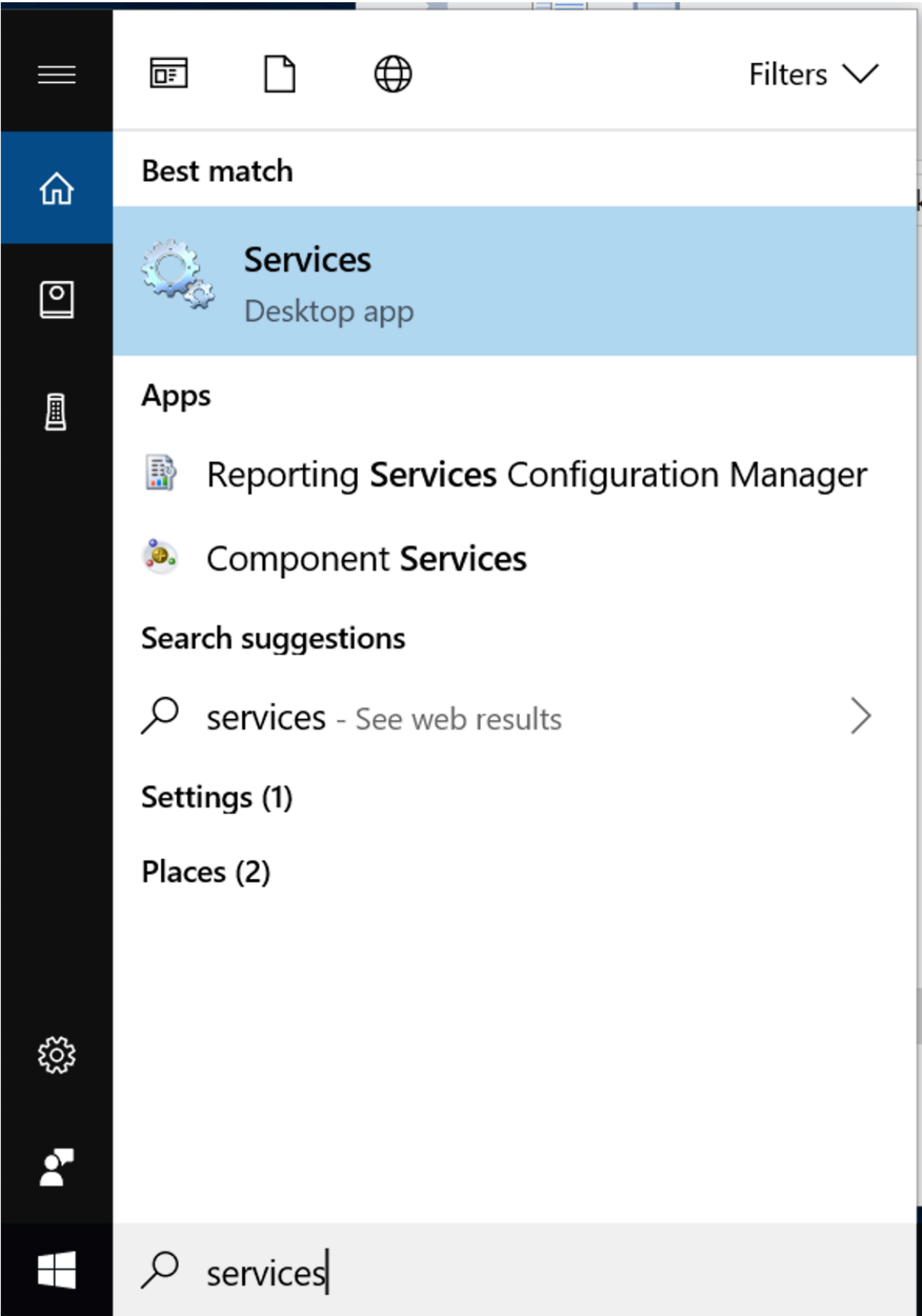
```
sudo service rosokawrk start
```

While on Windows, you can use this command:

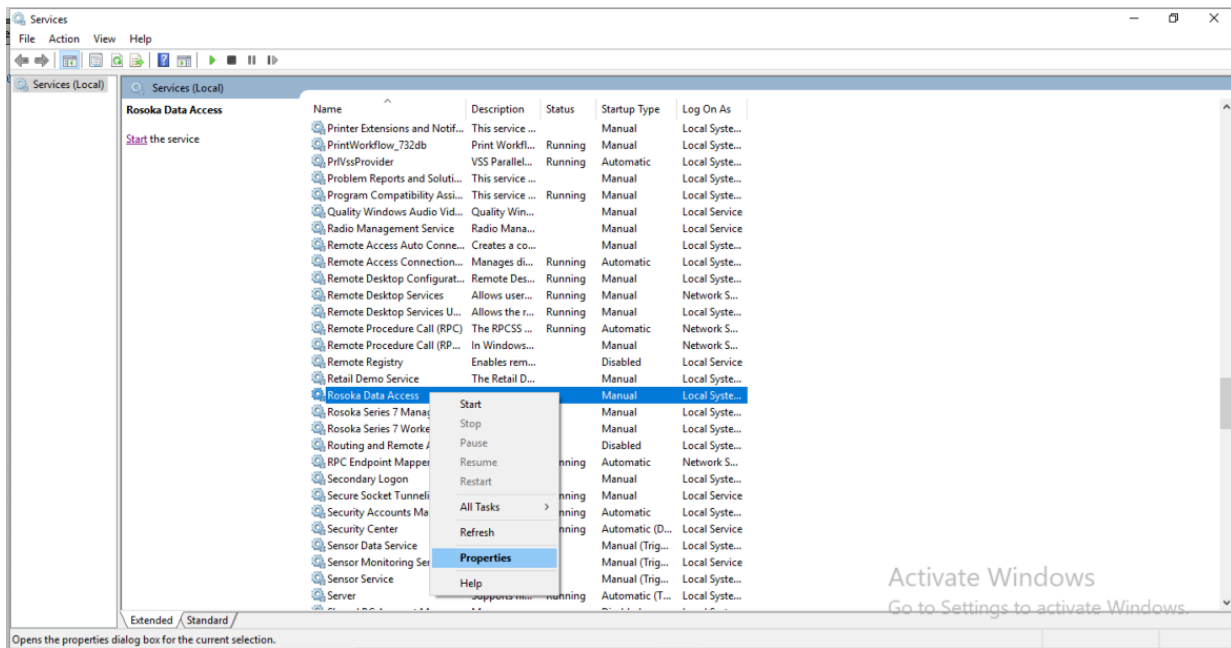
```
net start RosokaWorker
```

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:



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**:

Rosoka Series 7 Worker Properties (Local Computer) ✕

General Log On Recovery Dependencies

Select the computer's response if this service fails. [Help me set up recovery actions.](#)

First failure: Restart the Service

Second failure: Take No Action

Subsequent failures: Take No Action
Restart the Service

Reset fail count after: Run a Program
Restart the Computer

Restart service after: 1 minutes

☐ Enable actions for stops with errors. Restart Computer Options...

Run program

Program: Browse...

Command line parameters:

☐ Append fail count to end of command line (/fail=%1%)

OK Cancel Apply

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.

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 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**.

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:

Clusters

New Cluster

(*) indicates default cluster to use when a job doesn't specify)

	Cluster	Description	LxBase	GxBase	Output
	Test	Test new LxBase	*default LxBase	*default GxBase	REST only
*	ToDB	Write results to database	*default LxBase	*default GxBase	ElasticSearch

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.

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 screenshot shows a configuration panel titled "[Test] Configuration" with a blue header bar. In the top right corner of the header are two buttons: "Save Changes" and "Cancel Changes". Below the header is a tabbed interface with three tabs: "Assets", "Extraction", and "Output". The "Extraction" tab is currently selected. The main content area of the "Extraction" tab contains a table with two rows. The first row is for "Lexical data" and the second row is for "Geographic data". Both rows have a value of "*_default" in the second column and a green "Upload" button in the third column.

[Test] Configuration		
Assets Extraction Output		
Lexical data	*_default	Upload
Geographic data	*_default	Upload

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:

- **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:

Assets	Extraction	Output
Lexical data	*_default	Upload
Geographic data	*_default	Upload

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:

Assets	Extraction	Output
Lexical data	LxBundle.zip	Upload
Geographic data	*_default	Upload

Extraction

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

Assets	Extraction	Output
rawinput		FALSE
inlinetext		FALSE
inlinegloss		FALSE
internalGeoGravy		ON
geoMode		BEST
geoSortPreference		
loaderChunkHardLimit		65536
loaderChunkSoftLimit		64536

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)

- **loaderChunkHardLimit**

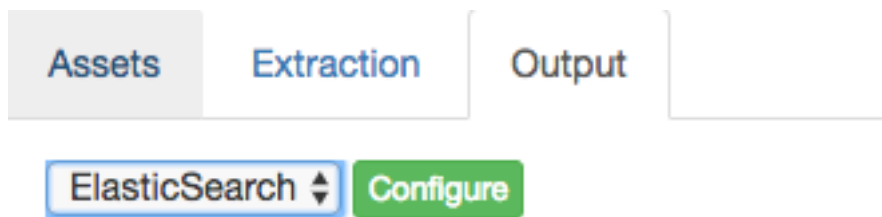
Maximum length, in characters, of text that the engine can process in one go. A value of -1 indicates no limit (not recommended).

- **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 then 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:

Workers					
Registered Workers					
	Name	Status	Enabled	Cluster	Last Contact
	Local_Work1	ready	true	Test	2017-02-23 13:47:53
Unregistered Workers					
	Name				
	DataCenter_Work5				2017-02-23 13:47:55

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.

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:

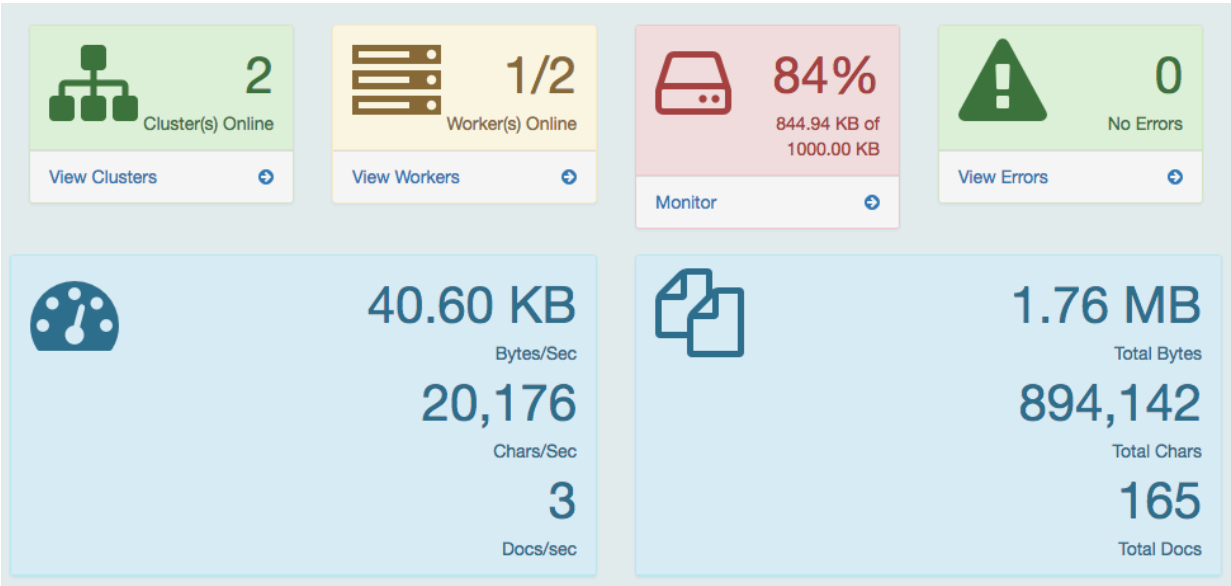
State	Description
off	The worker is registered with the manager, but it is not currently connected. This is typically the case for a worker node that has been shut down, or where the worker process is not running.
awaiting registration	The worker is registered, and the manager is waiting for the worker 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.

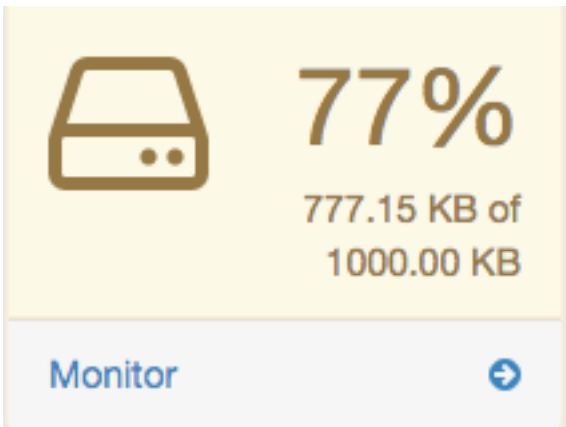
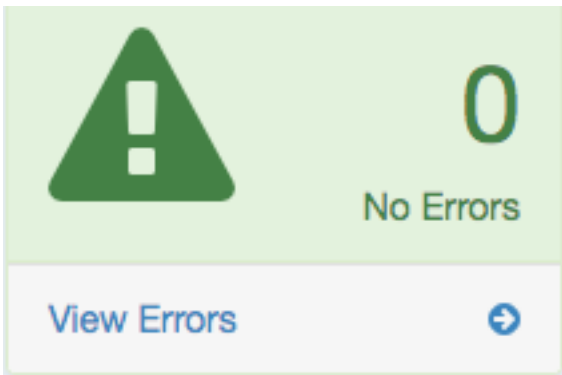
System overview (Dashboard)

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



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.

Panel	Description
 The panel displays a server rack icon, the fraction 1/2, and the text 'Worker(s) Online'. Below this is a button labeled 'View Workers' with a right-pointing arrow icon.	Shows the number of worker nodes that are active, relative to the number of registered nodes.
 The panel displays a disk icon, the percentage 77%, and the text '777.15 KB of 1000.00 KB'. Below this is a button labeled 'Monitor' with a right-pointing arrow icon.	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.
 The panel displays a green warning triangle icon, the number 0, and the text 'No Errors'. Below this is a button labeled 'View Errors' with a right-pointing arrow icon.	Shows the number of processing errors that have occurred.

Panel	Description
 2.81 KB Bytes/Sec 1,886 Chars/Sec 0 Docs/sec	Shows the overall throughput performance of the i2 TextChart Server system since manager startup.
 61.49 KB Total Bytes 41,233 Total Chars 16 Total Docs	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:



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:

Errors				
Show 10 entries				
Time	Cluster	Worker	File	Message
2017-02-23 11:58:03	Test	Local_Work1	foo.txt	no_content
2017-02-23 11:58:01	Test	Local_Work1	foo.txt	no_content

Previous 1 Next

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.

Inputs		
Name	Description	Status
FileMonitor	A plugin that can be configured to monitor one or more directories and process new or updated files	stopped Start Configure
Crawler	A plugin that can be configured to crawl one or more web sites at a given interval	stopped Start Configure

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.

- **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.

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.

The screenshot displays the i2 TextChart Server web interface, divided into three main sections:

- Controls Panel:** Contains two buttons: "Restart Manager" and "Clear Queue".
- License Panel:** Shows the current license status. It includes an "Upload" button and displays the following information:
 - Status: valid
 - Version: 6.x
 - Expiration Date: Wed Feb 21 2018, 310 days remaining
 - Licensed Workers: 2, 0 currently in use
- Default assets (Unlock to change) Panel:** Lists two default assets:
 - LxBase:** LxBundle.zip, coreVersion:6.2.0.2;coreDate:2017-04-13 16:16:49 EDT, Apr 17, 2017 10:33:50 AM.
 - GxBase:** GxBundle.tbz2, coreVersion:6.0.0.0;coreDate:2017-02-24 15:36:39 EST, Apr 17, 2017 10:33:50 AM.
- Manager Settings Panel:** Contains various configuration options:
 - Login:** Admin Username (myadmin), Admin Password (masked).
 - Ports:** Admin Service (http: 8080, https: 0), Worker Service (http: 8083, https: 0), Client Service (http: 8081, https: 0).
 - SSL:** Keystore Password, Keystore Manager Password, Keystore Path.
 - Other Settings:** Max Queue Size (no restart required: 0), Enable Client Service (no restart required: checked), Queue Path (/Applications/RosokaExtractionManager/queue), Input Auto Start.

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.

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.

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.

- `-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).
- `-psa`: 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"}'
```

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
<code>authType</code>	Determines the type of authorization for accessing the administration UI. Options are <code>form</code> (the default), which uses a single username-password combination (from the <code>adminUser</code> and <code>adminPassword</code> properties), or <code>ldap</code> , which uses the LDAP server defined in the <code>conf/ldap.conf</code> file for authentication and authorization.
<code>authGroup</code>	(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 <code>**</code> (the default) authenticates the username and password but does <i>not</i> 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.

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	<code>true</code> 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 <code>keystore.jks</code> 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 <code>keyStorePassword</code> 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 <code>-Xmx</code> option. For example, to specify a heap size of 6 GB, use <code>6144m</code> . If not present, defaults to <code>4096m</code> .

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.

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.

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:

- 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.

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,

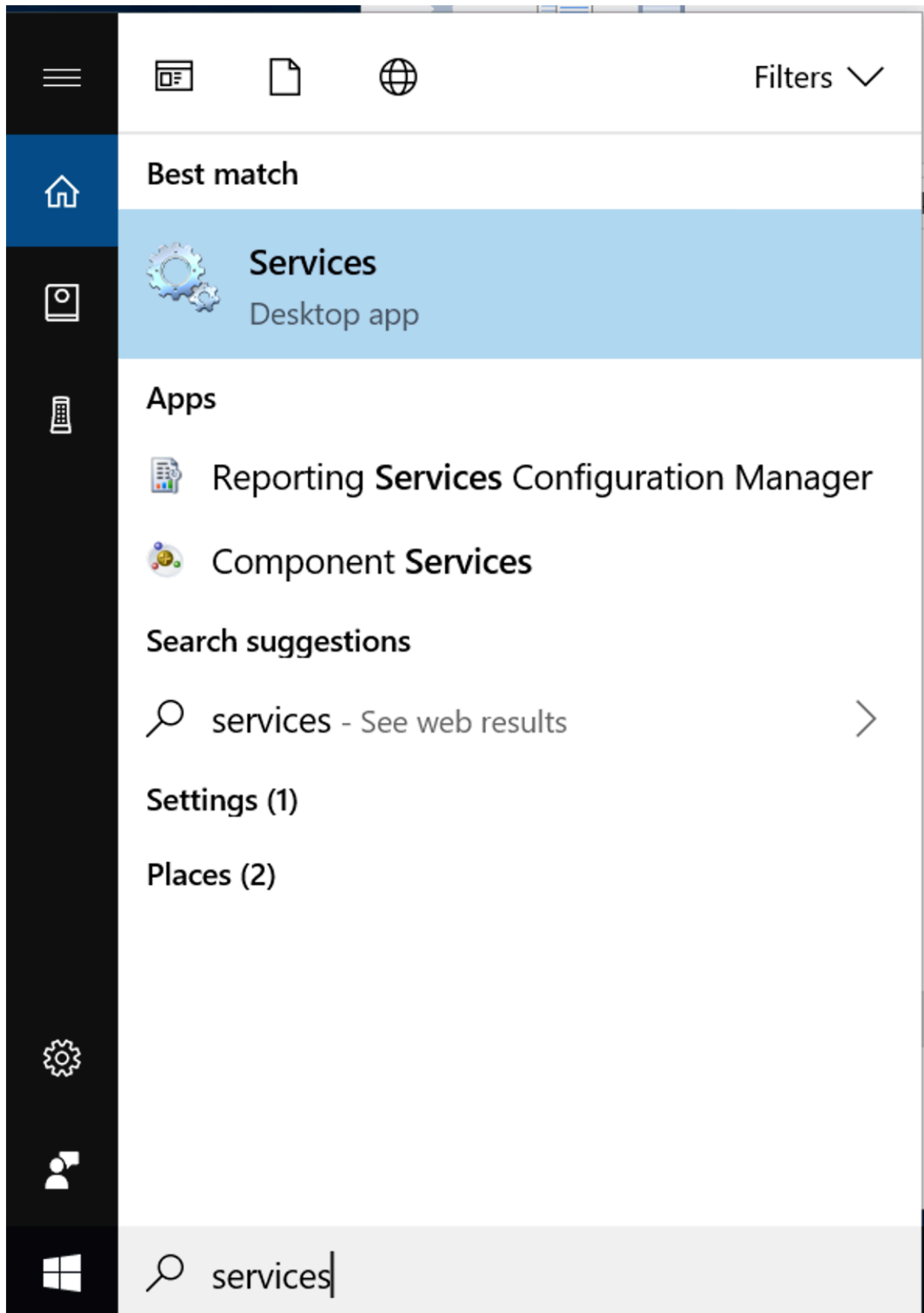
```
sudo service rosokarda start
```

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

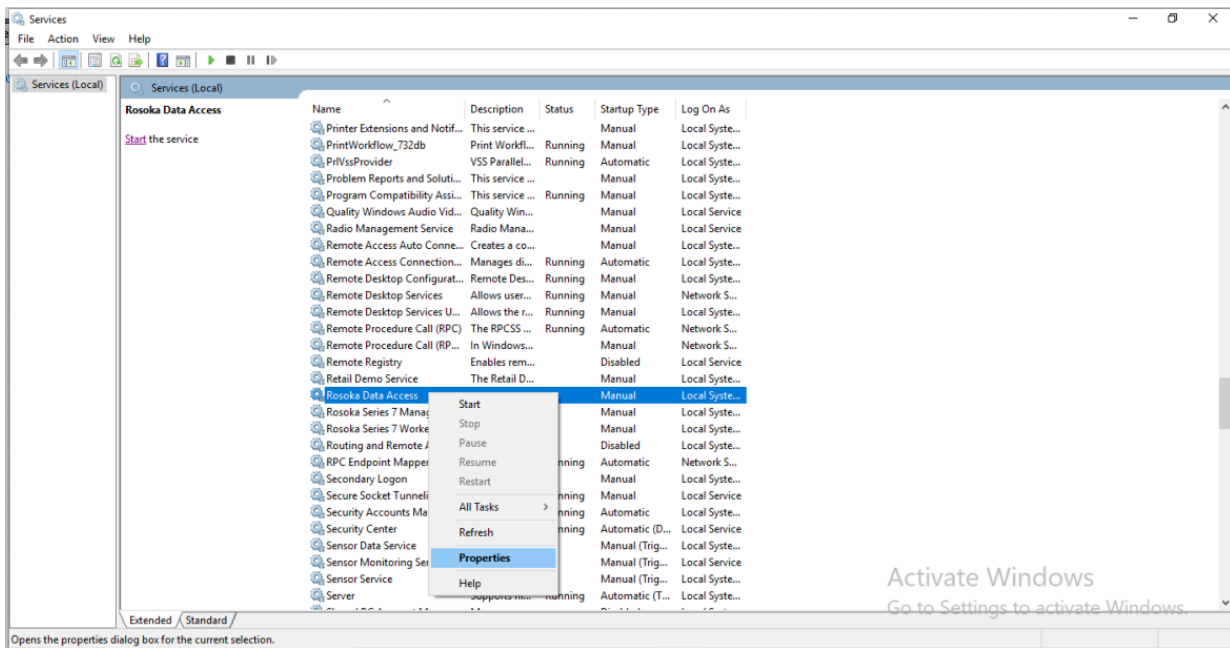
```
net start RosokaDataAccess
```

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.



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.

Rosoka Series 7 Worker Properties (Local Computer) ✕

General Log On Recovery Dependencies

Select the computer's response if this service fails. [Help me set up recovery actions.](#)

First failure: Restart the Service

Second failure: Take No Action

Subsequent failures: Take No Action
Restart the Service
Run a Program
Restart the Computer

Reset fail count after:

Restart service after: 1 minutes

☐ Enable actions for stops with errors. Restart Computer Options...

Run program

Program: Browse...

Command line parameters:

☐ Append fail count to end of command line (/fail=%1%)

OK Cancel Apply

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.

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.

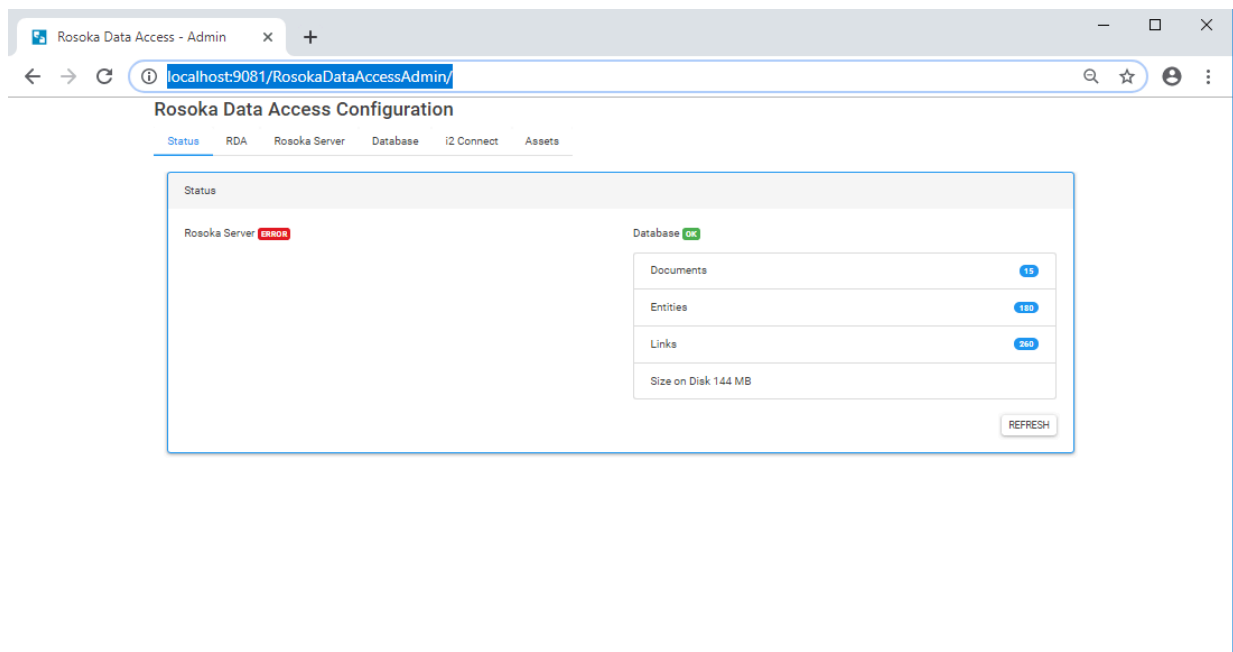
Getting started

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.



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

Rosoka Data Access Configuration

[Status](#)
[RDA](#)
[Rosoka Server](#)
[Database](#)
[i2 Connect](#)
[Assets](#)

Status

Rosoka Server OK

Cluster: mycluster

Assets: OK

Database OK

Documents 15

Entities 180

Links 260

Size on Disk 144 MB

REFRESH

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

Rosoka Data Access Configuration

[Status](#)
[RDA](#)
[Rosoka Server](#)
[Database](#)
[i2 Connect](#)
[Assets](#)

Rosoka Data Access Settings (changes here require restart)

Ip / Hostname

Server1

RDA ip or hostname

Listen on Port

9081

RDA port setting

RDA Admin Username

admin

RDA Admin Password

Password

SAVE

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

Rosoka Data Access Configuration

Status RDA **Rosoka Server** Database i2 Connect Assets

Server Connection Settings

Manager Host	Manager Port
localhost	8081
Rosoka Manager hostname or ip address	Rosoka Manager client port setting
mycluster	
Manager Cluster	

TEST SAVE

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

Rosoka Data Access Configuration

Status Rosoka Server **Database** Assets RDA Settings

Database Connection Settings

Rosoka Data access must be configured to connect to the Database where Rosoka Server is writing its output to.

Database Host	Database Port
localhost	1433
Database host	Database port
Database Username	Database Password
sa	*****
Use Active Directory? ☐	
Check to use Active Directory DB login. Requires additional server configuration.	

TEST SAVE

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

Rosoka Data Access Configuration

Status RDA Rosoka Server Database **i2 Connect** Assets

i2 Connect Settings (changes will require restart)

Schema: none

Default Entity Type ⌵ Default Link Type ⌵

Documents

Entity Type ⌵ Title Property ⌵

Link Type ⌵ Link Label ⌵

SAVE

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.

Rosoka Data Access Configuration

Status RDA Rosoka Server Database i2 Connect **Assets**

Assets			
Asset Name	Asset File	Version	Timestamp
Mapping Configuration	mappingconf.zip	Matches Cluster mycluster	Jun 18, 2019 12:47:08 PM
GxBase	GxBundle.tbz2		Jun 18, 2019 12:47:07 PM
LxBase	LxBundle.zip	coreVersion:7.1.0.0;coreDate:\${NOW}	Jun 18, 2019 12:47:08 PM

REFRESH **GET CURRENT MAPPING CONFIG**

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.

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:

```
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.

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.