netwrix

Netwrix Data Classification for Google Drive

Quick-Start Guide

Version: 5.5.2
3/10/2020

Table of Contents

Legal Notice

The information in this publication is furnished for information use only, and does not constitute a commitment from Netwrix Corporation of any features or functions, as this publication may describe features or functionality not applicable to the product release or version you are using. Netwrix makes no representations or warranties about the Software beyond what is provided in the License Agreement. Netwrix Corporation assumes no responsibility or liability for the accuracy of the information presented, which is subject to change without notice. If you believe there is an error in this publication, please report it to us in writing.

Netwrix is a registered trademark of Netwrix Corporation. The Netwrix logo and all other Netwrix product or service names and slogans are registered trademarks or trademarks of Netwrix Corporation. Microsoft, Active Directory, Exchange, Exchange Online, Office 365, SharePoint, SQL Server, Windows, and Windows Server are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. All other trademarks and registered trademarks are property of their respective owners.

Disclaimers

This document may contain information regarding the use and installation of non-Netwrix products. Please note that this information is provided as a courtesy to assist you. While Netwrix tries to ensure that this information accurately reflects the information provided by the supplier, please refer to the materials provided with any non-Netwrix product and contact the supplier for confirmation. Netwrix Corporation assumes no responsibility or liability for incorrect or incomplete information provided about non-Netwrix products.

1. Revision History

Revision #	Date	Summary of Changes
Revision 1	13/01/2020	Initial version of Google Drive QSG
Revision 2	14/01/2020	• Added Introduction topic • How It Works diagram hidden (will be updated) • Removed (up to 100,000 files) from deployment checklist • Deployment checklist updated (removed and updated several steps)
Revision 3	15/01/2020	• Updated Product Configuration Wizard description (4 separate steps)
Revision 4	16/01/2020	• Updated section Create App for GDrive Crawling
Revision 5	22/01/2020	• Updated section Create App for G Drive Crawling (added "select user type" step)
Revision 5	23/01/2020	QC comments: • Install - File system path (use default) • Initial configuration - modifed paragraph about web console url • Processing settings - use default values instead of describing all options in the table • Processing modes - keyword?

2. Introduction

This guide is intended for the first-time users of Netwrix Data Classification for Google Drive. It can be used for evaluation purposes, therefore, it is recommended to read it sequentially, and follow the instructions in the order they are provided. After reading this guide you will be able to:

- Install and configure Netwrix Data Classification
- Add a source to start crawling Google Drive
- Browse for classified documents
- Leverage reporting capabilities and export results for custom reports

NOTE: This guide only covers the basic configuration and usage options for crawling Google Drive with Netwrix Data Classification. For advanced installation scenarios and configuration options, as well as for information on various reporting possibilities and other product features, refer to Netwrix Data Classification Online Help Center.

2.1. Features and Benefits

Netwrix Data Classification is a platform that identifies data that's important for your organization and enables you to reduce risk and unleash the true value of this data.

Powered by unique compound term processing technology, it enriches your enterprise content with accurate and consistent metadata empowering you to work with data more confidently. By seeing which data is valuable, you can organize it in a way that promotes productivity and collaboration. By knowing where sensitive or regulated data is, you can reduce the risk of breaches and satisfy security and privacy requirements with less effort and expense. And by locating and getting rid of redundant and obsolete data, you can reduce storage and management costs.

Netwrix Data Classification in includes applications for Windows File Servers, Nutanix Files, Dell EMC, NetApp, SharePoint, Office 365, Exchange, SQL Server, Oracle Database, Box, Google Drive, MySQL and PostgreSQL. The platform provides a single panoramic view of your enterprise content, whether it's located in structured or unstructured data stores, on premises or in the cloud.

Major benefits:

- Identify sensitive information and reduce its exposure
- Improve employee productivity and decision making
- Reduce costs and risks by getting rid of unneeded data
- Meet privacy and compliance requirements for information governance
- Respond to legal requests without putting your business on hold

3. Configure G Suite for Crawling

Netwrix Data Classification for Google Drive uses the **OAuth 2.0** protocol to authenticate to your G Suite domain. You will need to create a service account and authorize it to access data in individual and shared Drives on behalf of users using the Google Drive API. Do the following:

In Google API Console:

1. Create a new project
2. Select Application type
3. Create a new service account
4. Create a service account key (JSON, save a copy for the data source configuration)
5. Enable G Suite domain-wide delegation for the service account (write down the Client ID)
6. Enable Google Drive API

In G Suite Admin Console:

1. Authorize service account to access the Google Drive API

To configure G Suite for crawling

IMPORTANT! Google administrative interfaces tend to change over time, so refer to the following guide for up-to-data instructions on creating OAuth 2.0 service accounts: Using OAuth 2.0 for Server to Server Applications.

Review the following for additional information:

To...	Do...
Create a new project	1. Navigate to https://console.developers.google.com (Google API Console) while logged in as a G-Suite administrator within the domain to be crawled (if the user is not added within the correct domain then the correct data will not be identified). 2. Create a new project.
Select Application type	1. Once a new project has been created, navigate to **APIs&Services** → **OAuth consent screen**. 2. Set **User type** to "*Internal*". 3. Provide the name for new application. 4. Click **Save**.

To...	Do...
Create a new service account	1. In **Google API** console, navigate to **IAM & Admin→Service Accounts**. 2. Create service account as described in Google official article. 3. On the **Grant this service account access to project (optional)** step, do not select any roles. 4. On the **Grant users access to this service account (optional)** step, do not grant any user access. Click **Done**.
Create a service account key	1. On the **Service accounts** page, select the account you want to create a key for. 2. Click ⋮ icon under **Actions** and select **Create key**. 3. In the **Create private key for <Service account name>** dialog, select **JSON** format, and download the file to a known location as it will be required later. **NOTE:** Your new public / private keypair is generated and downloaded to your machine; it serves as the only copy of this key. You are responsible for storing it securely. If you lose this keypair, you will need to generate a new one.
Delegate domain-wide authority to the service account	1. On the **Service accounts** page, select your service account and click **Edit**. 2. Click the **Show Domain-Wide Delegation** link and tick the **Enable G Suite Domain-wide Delegation** checkbox. 3. Click **Save**. 4. Once completed, review the "*Domain wide delegation*" column for this account and make sure that it enabled. 5. Click the **View Client ID** link. 6. Copy your Client ID, you will need it later.
Enable Google Drive API	1. In **Google API** console, navigate to the **API Dashboard** and select **Enable APIs and Services** (if APIs have not previously been enabled). 2. Search for Google Drive API and click **Enable** (or **Manage**). 3. Switch to **G Suite Admin Console**.

To...	Do...
	4. Navigate to **Security** → **Advanced Settings** → **Manage API Client Access** within the Google admin portal. 5. Set the client name to the **Client ID** you copied on the previous step. 6. Set the API scope to *"https://www.googleapis.com/auth/drive"* and select **Authorize**.

4. Install Netwrix Data Classification

1. Run **Netwrix_Data_Classification.exe**.
2. Review minimum system requirements and then read the License Agreement. Click **Next**.
3. Follow the instructions of the setup wizard. When prompted, accept the license agreement.
4. On the **Product Settings** step, specify path to install Netwrix Data Classification. For example, *C:\Program Files\NDC*.
5. On the **Configuration** step, specify the directory where **Index files** reside. For example, *C:\Program Files\NDC\Index*.
6. On the **SQL Database** step, provide SQL Server database connection details.

   Complete the following fields:

Option	Description
Server Name	Provide the name of the SQL Server instance that hosts your NDC SQL database. For example, *"WORKSTATIONSQL\SQLSERVER"*.
Authentication Method	Select Windows or SQL Server authentication method.
Username	Specify the account name.
Password	Provide your password.
Database Name	Enter the name of the SQL Server database. Netwrix recommends using **NDC_database** name.

7. On the **Licensing** step, add license. You can add license as follows:
   - Click the **Import** button and browse for your license file

     OR
   - Open your license file with any text editor, e.g., **Notepad** and paste the license text to the **License** field.
8. On the **Administration Web Application** step, review default IIS configuration.
9. On the **Services** step, configure Netwrix Data Classification services:

- Select all services to be installed.
- **File System Path**—Use default path or provide a custom one to store Netwrix Data Classification's Services files. For example, *C:\Program Files\NDC Services.*
- Provide user name and password for the product services service account.

  NOTE: This account is granted the **Logon as a service** privilege automatically on the computer where NDC is going to be installed.
- Select additional service options, if necessary.

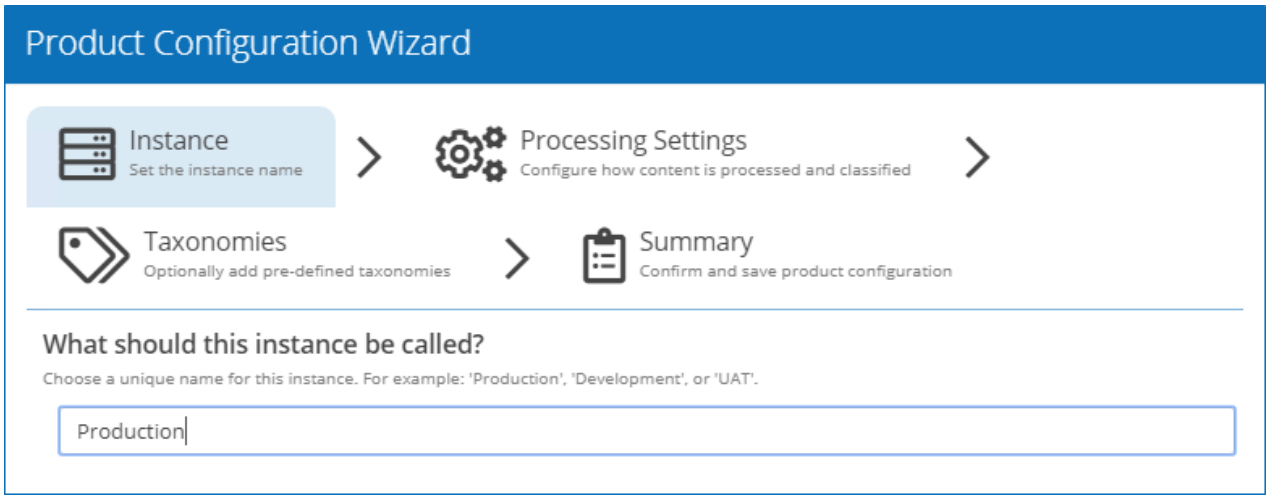
10. On the **Pre-Installation Tasks and Checks** step, review your configuration and select **Install**.
11. When the installation completes, open a web browser and navigate to the following URL: *http://localhost/conceptQS* where **localhost** is the name or IP address of the computer where Netwrix Data Classification is installed. For example, *http://workstationndc/conceptQS*.

5. Initial Product Configuration

The **Product Configuration Wizard** allows you quickly configure basic Netwrix Data Classification settings such as processing mode, taxonomies, etc.

In your web browser, navigate to the following URL: http://hostname/conceptQS where **hostname** is the name or IP address of the computer where Netwrix Data Classification is installed and perform initial configuration steps.

On the **Instance** step, provide the unique name for your Netwrix Data Classification instance. For example, *"Production"*.

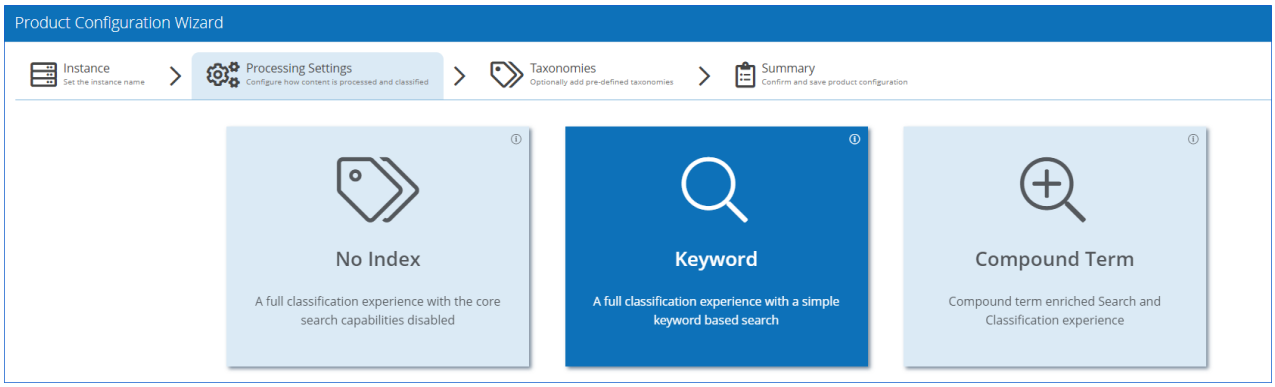


Click **Next** to proceed. See also:

- Select Processing Mode
- Processing Settings
- Add Taxonomy
- Review Your Configuration

5.1. Select Processing Mode

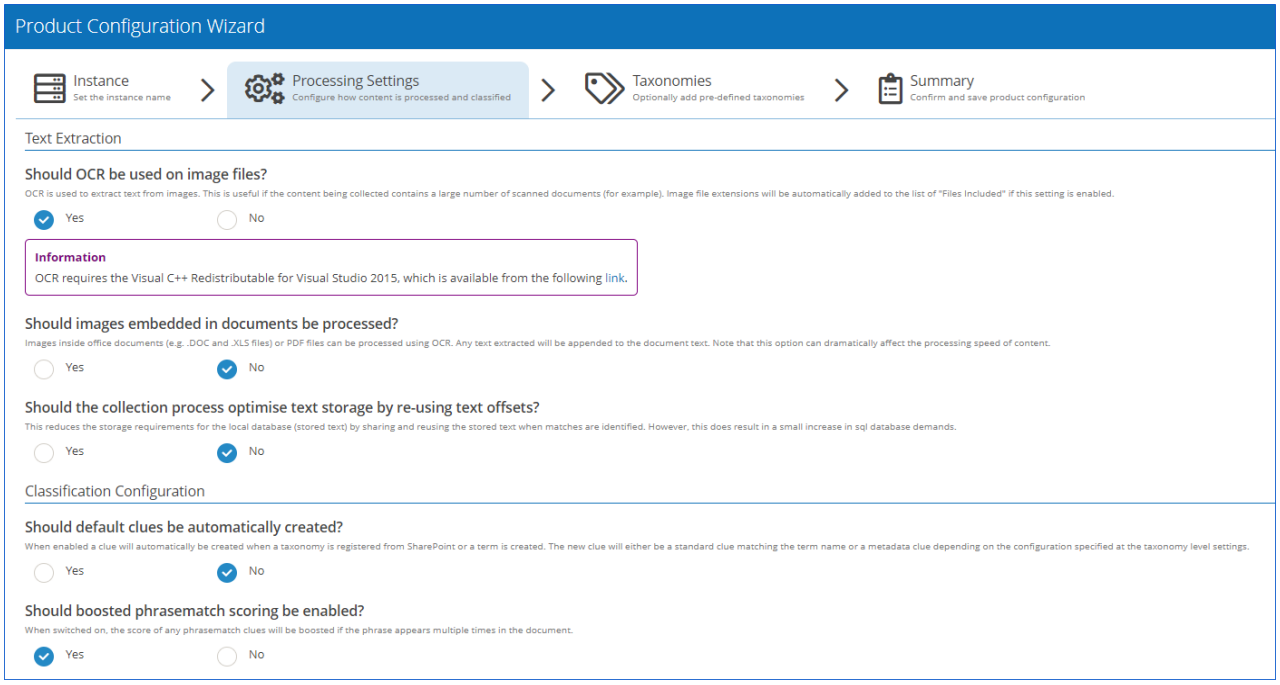
At this step of the wizard, select processing (indexing) mode for your environment.



For starter and evaluation purpose, select **Keyword** mode.

5.2. Processing Settings

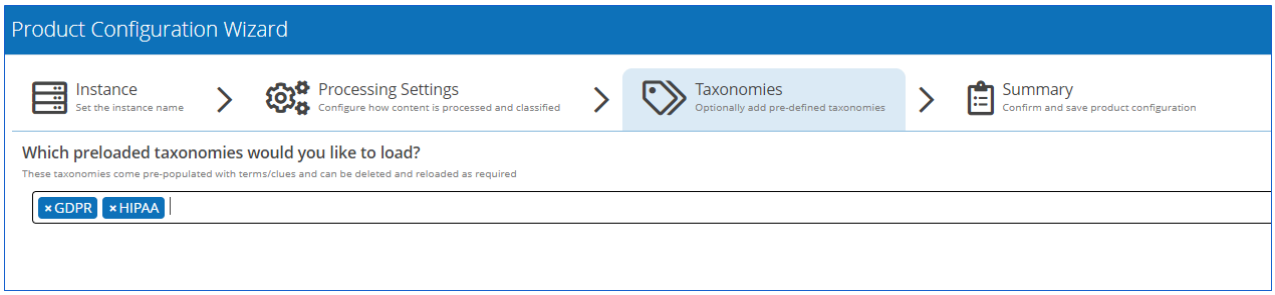
On the **Processing Settings** step, review options for data processing and classification. For test and evaluation purposes, Netwrix recommends use default values.



Proceed with adding taxonomies.

5.3. Add Taxonomy

On this step, you are prompted to load predefined taxonomies.



Click the search bar and select one or several taxonomies you want to add. See Built-in Taxonomies Overview for the full list of built-in taxonomies supported by Netwrix Data Classification.

5.4. Review Your Configuration

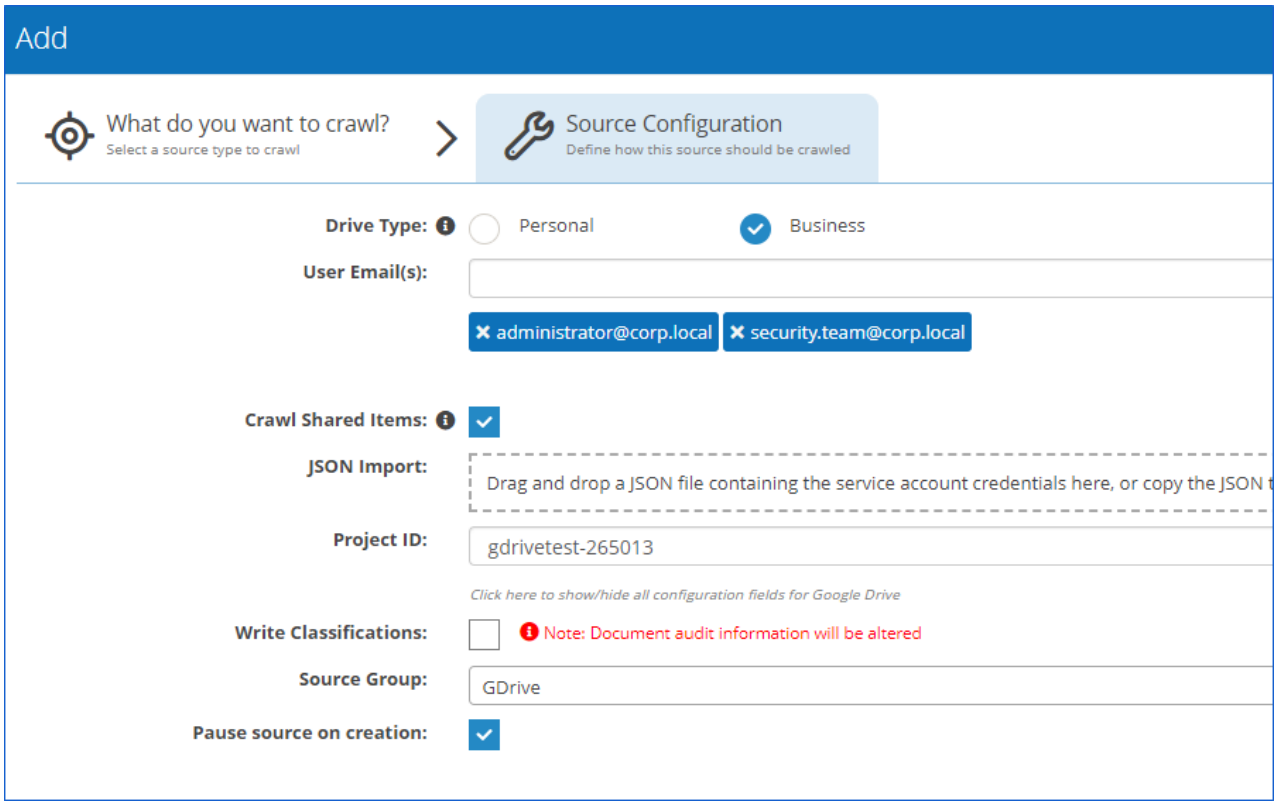
On this step, review your configuration. Once you complete the wizard, you can:

- Add a Source
- Add a Taxonomy
- Take the Product Tour
- Get Help

6. Add Google Drive Source

The **Google Drive** source configuration screen allows you to enable the crawling and classification of content stored in both G-Suite repositories and Google Drive personal accounts.

IMPORTANT! Make sure you created App for GDrive crawling prior to start adding the source. See Configure G Suite for Crawling for more information.



Complete the following fields:

Option	Description
	Basic settings
Drive Type	Select *Business*.
User Email(s)	When adding a G-Suite source, enter the email address of the user's drive that you wish to crawl (via impersonation).
Crawl Shared Items	Select to crawl all files shared with the specified user in addition to any team drives shared with the user.

Option	Description
Crawl Shared Items	Select to enable crawling of any types of documents shared with the specified user.
JSON Import	Drag the JSON connection file you downloaded while creating Google service account in the form.
Project ID	Open the JSON connection file and copy file contents to **Project ID** field.
Write Classifications	Leave this checkbox empty.
Source Group	Netwrix recommends creating a dedicated source group for Google Drive.
Pause source on creation	Select if you want to make other configuration changes before collection of the source occurs.

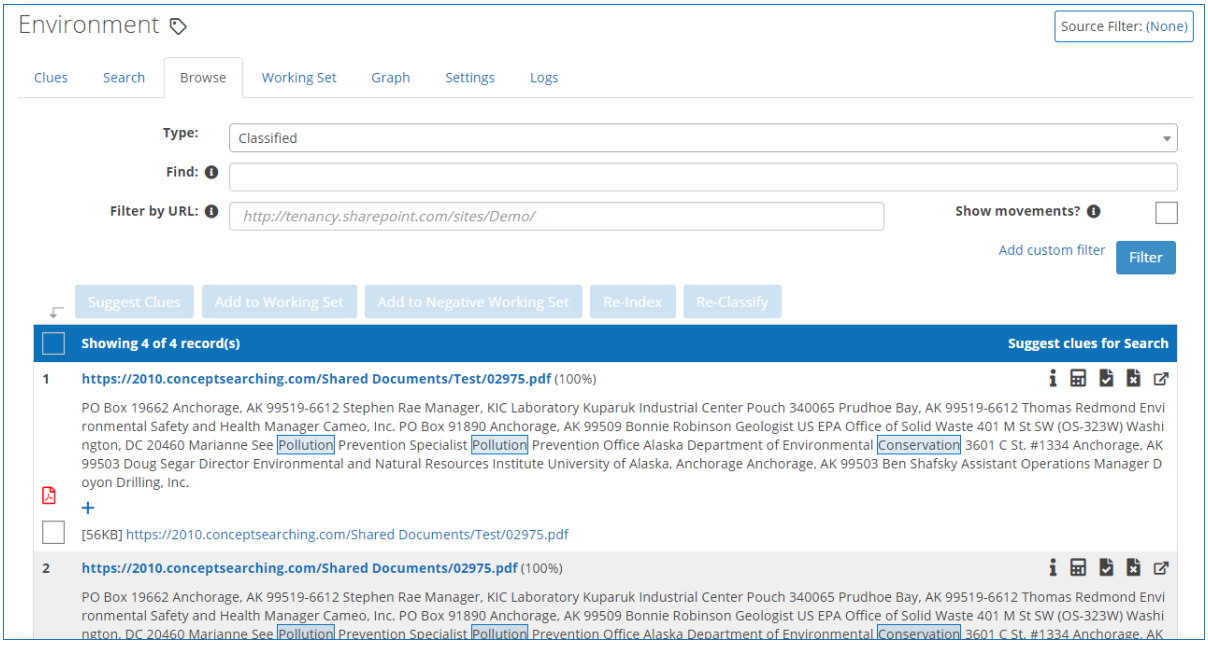
7. Review Reports and Browse Classified Documents

Once your documents are classified, you can identify sensitive information and reduce its exposure. Netwrix recommends starting with the **Document Tagging** report to see automatic and manual classifications of the documents within the reporting set. Further, you can browse your documents to see a list of documents achieving the minimum score set for classification in the term. Review the following for additional information:

- To browse classification results
- To review the Document Tagging report

To browse classification results

1. In administrative web console, navigate to **Taxonomies** → **Term Management**.
2. Select **Taxonomy** in the dropdown on the left and then expand specific term you are interested in.
3. Switch to **Browse** tab:



4. Click **Filter** to start browsing your documents.

To review the Document Tagging report

1. In administrative web console, navigate to **Reports** and expand the **Document Reports** set.
2. Select the **Document Tagging** report and click **Show filters** to narrow report scope.

Filter	Description
Taxonomy	By default, the report shows results for all taxonomies. Select the taxonomy you are interested in to restrict report scope.
Score Range	Select the score. Review Scoring for more information.
Classification	By default, the report shows results for all terms within a taxonomy. Limit your results by selected term.
Page URL	Filter your results by selected page URL.
Source	Select source group you created for Google Drive.

3. Click **Generate** and review results.
4. You can also export displayed page to .csv and .xlsx table or download the whole results.

TIP: Upon export, you will be prompted to include any associated document metadata to the report. It can be useful if you want to generate custom security reports. Specify metadata fields and click **Export** to download report.