



IBM Watson Group

Using Machine-Learning Annotators from Watson Knowledge Studio in Watson Explorer

IBM Watson Explorer Version 11.0.1

Contents

Overview	1
Language support	1
Limitations	2
System requirements	2
Information resources	2
Install SIRE rpm	3
Export the machine-learning model from Watson Knowledge Studio	3
Upload the machine-learning model to Watson Explorer	4
Associate the machine-learning model with a collection	5
Automatically generated facet definitions	7
Explore analytical results in Watson Explorer Content Analytics	9
Explore analytical results in Watson Explorer applications	10
Notices	11
Trademarks	13

Overview

Beginning with IBM® Watson Explorer Version 11.0.1, you can configure a machine-learning annotator to annotate documents that you add to Watson Explorer collections. After you train a machine-learning annotator component (also known as a model) in IBM Watson Knowledge Studio, you can export it as a ZIP file. You can then import the model into Watson Explorer Content Analytics or Watson Explorer Annotation Administration Console, and enable it to be used as a machine-learning annotator in your collections.

Watson Explorer supports three types of entities:

- **Mentions.** A mention is a span of text that is relevant in your collection data. For example, in a collection that contains documents about automobiles, terms like airbag, Ford Explorer, and child restraint system might be labeled by a machine-learning annotator as relevant mentions.
- **Relations.** A relation identifies a binary, ordered relationship between two entities. For example, in documents about automobiles, the machine-learning annotator might use the relation “occupantOf” to identify people who are occupants of a vehicle. For another example, the relation “employedBy” might identify people and the company they work for.
- **Coreferences.** Coreferences are mentions that mean the same thing, thus helping to ensure consistency when words are not identical. Examples of co-referenced mentions include the name of a U.S. state and its abbreviation, the name of a company and its acronym, or a person's name and a pronoun that refers back to that person.

Based on the entity information in the model, Watson Explorer automatically creates facet definitions for exploring content in a content analytics collection.

Enabling Watson Explorer to use a machine-learning annotator involves the following steps:

1. Installing the Statistical Information and Relation Extraction (SIRE) runtime
2. Exporting a trained machine-learning model from Watson Knowledge Studio
3. Uploading the exported model into Watson Explorer
4. Associating the machine-learning model with a content analytics collection

Language support

The Machine-Learning Annotator supports annotating text in the following languages:

- English (the default language for creating models in Watson Knowledge Studio)
- Arabic
- German
- Japanese
- Spanish

Limitations

The Machine-Learning Annotator cannot be used by the following Watson Explorer Advanced Edition collection features:

- Solution Gallery

System requirements

Operating system:

- Red Hat Enterprise Linux Server 7

The SIRE runtime requires the following libraries:

- apr
- apr-util
- boost-filesystem
- boost-iostreams
- boost-program-options
- boost-regex
- boost-serialization

Memory:

The SIRE runtime consumes memory outside Watson Explorer processes. Around 4GB of memory are required per SIRE runtime process on the Watson Explorer server.

In Watson Explorer Content Analytics, this amount of memory is needed on each server that has the document processing role, depending on the model size, regardless of whether the same model is associated with multiple collections. For example, if the same machine-learning model is associated with two collections, the SIRE runtime consumes memory for each collection. If you let both collections run, you need 8GB memory on each document processing server beyond the Watson Explorer usage requirements.

Information resources

- IBM Watson Explorer Content Analytics documentation in [IBM Knowledge Center](#).
- IBM Watson Explorer Annotation Administration Console documentation in [IBM Knowledge Center](#).
- IBM Watson Knowledge Studio documentation in [IBM Watson Developer Cloud](#).
- Video that demonstrates the integration between Watson Explorer and Watson Knowledge Studio: <https://www.youtube.com/watch?v=1VoS-xczBow&feature=youtu.be>

Install SIRE rpm

After you install Watson Explorer Content Analytics or Watson Explorer Annotation Administration Console, SIRE rpm is provided in the ES_INSTALL_ROOT/bin/sire/sire-20160429-2.x86_64.rpm directory. Before you can configure a machine-learning annotator, you must install SIRE rpm and prerequisite libraries on the server where you installed Watson Explorer Content Analytics or Watson Explorer Annotation Administration Console. In a distributed server installation of Watson Explorer Content Analytics, install SIRE rpm on all servers that are configured to support the document processing role.

1. Enter the following command to install the required libraries, which are listed in [System requirements](#):
yum -y install apr apr-util boost-filesystem boost-iostreams boost-program-options boost-regex boost-serialization

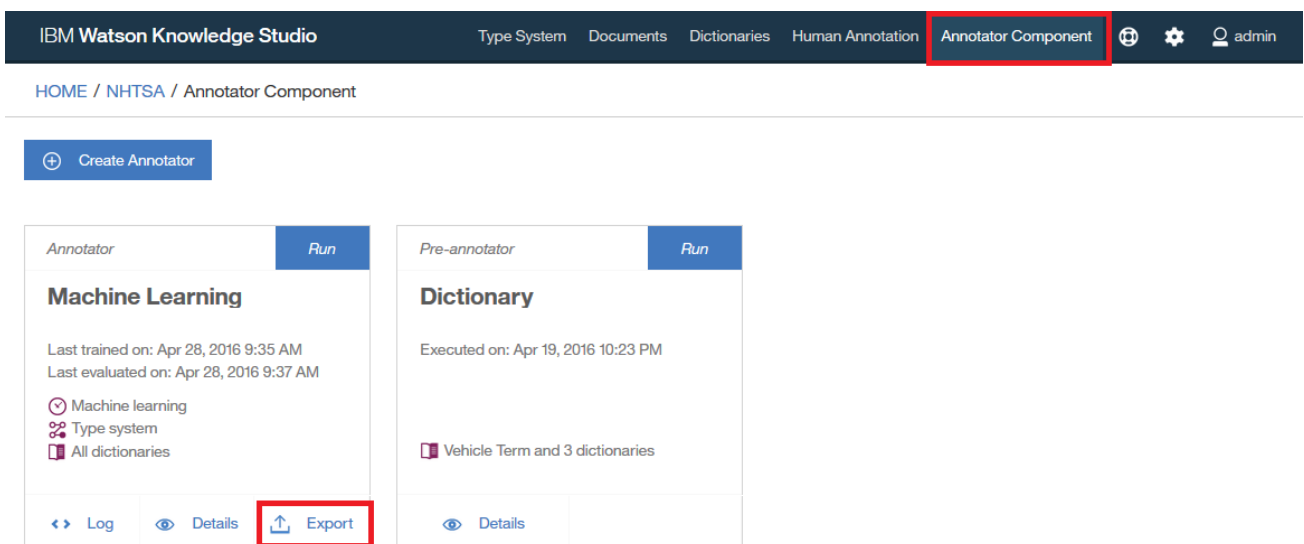
2. Enter the following command to install SIRE rpm:
rpm -ivh sire-20160429-2.x86_64.rpm

3. After SIRE rpm is installed, log in again as the default content analytics administrator (e.g., as user esadmin, if you accepted the default installation settings). To set the SIRE environment variables, enter the following commands to restart the system:

```
esadmin system stopall  
esadmin system startall
```

Export the machine-learning model from Watson Knowledge Studio

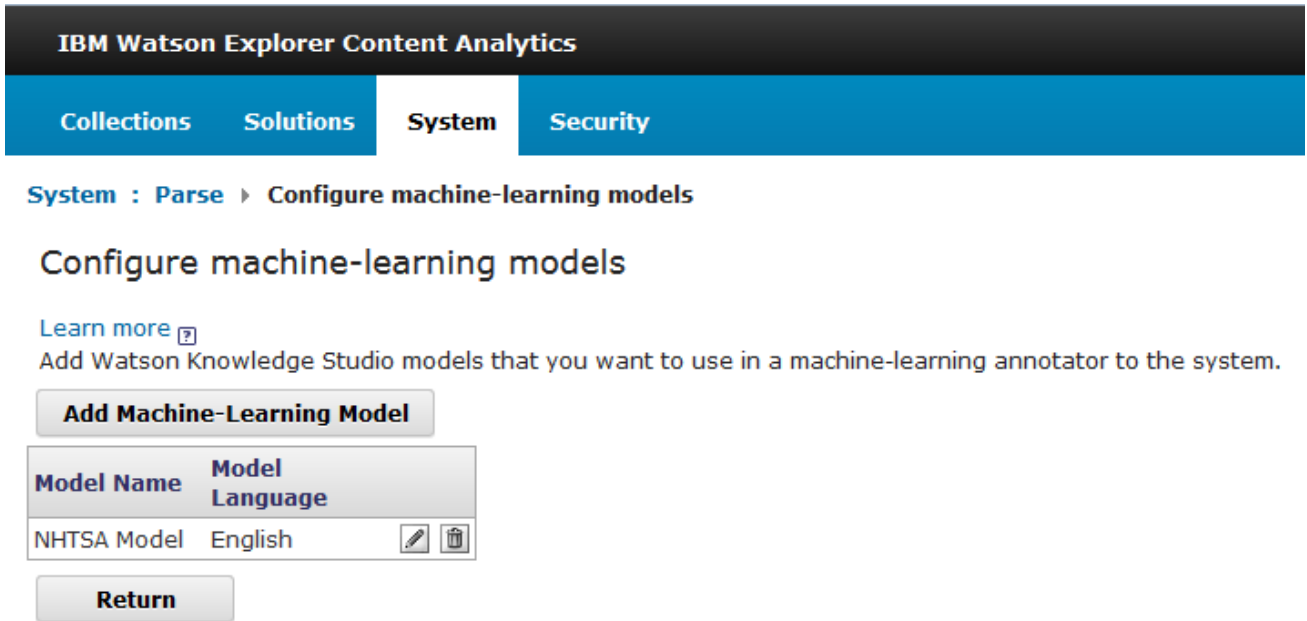
In Watson Knowledge Studio, open the project that contains the trained machine-learning annotator model that you want to export. Open the **Annotator Component** page and click **Export** in the Machine Learning annotator tile. Save the ZIP file that gets created and copy the file to the server where you installed Watson Explorer Content Analytics or Watson Explorer Annotation Administration Console.



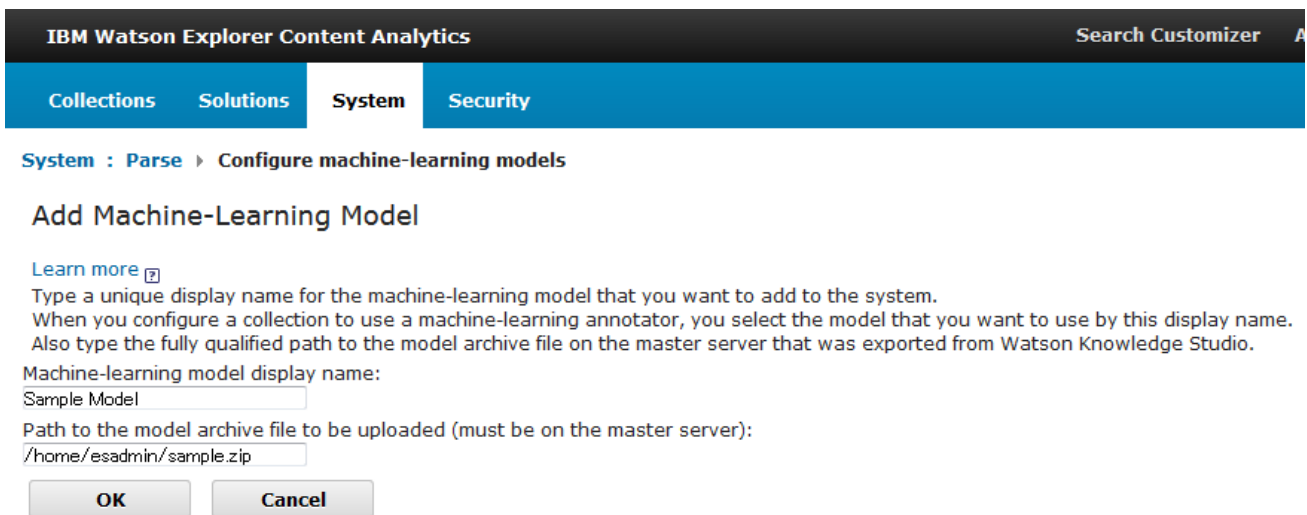
Upload the machine-learning model to Watson Explorer

In the Content Analytics administration console or Annotation Administration Console, open the **System -> Parse** page and click **Configure machine-learning models**. A list of machine-learning models that were previously added to the system is displayed, if any.

To upload a new machine-learning model, click **Add Machine-Learning Model**. The following sample screens show the Content Analytics administration console, but the steps are the same in Annotation Administration Console.



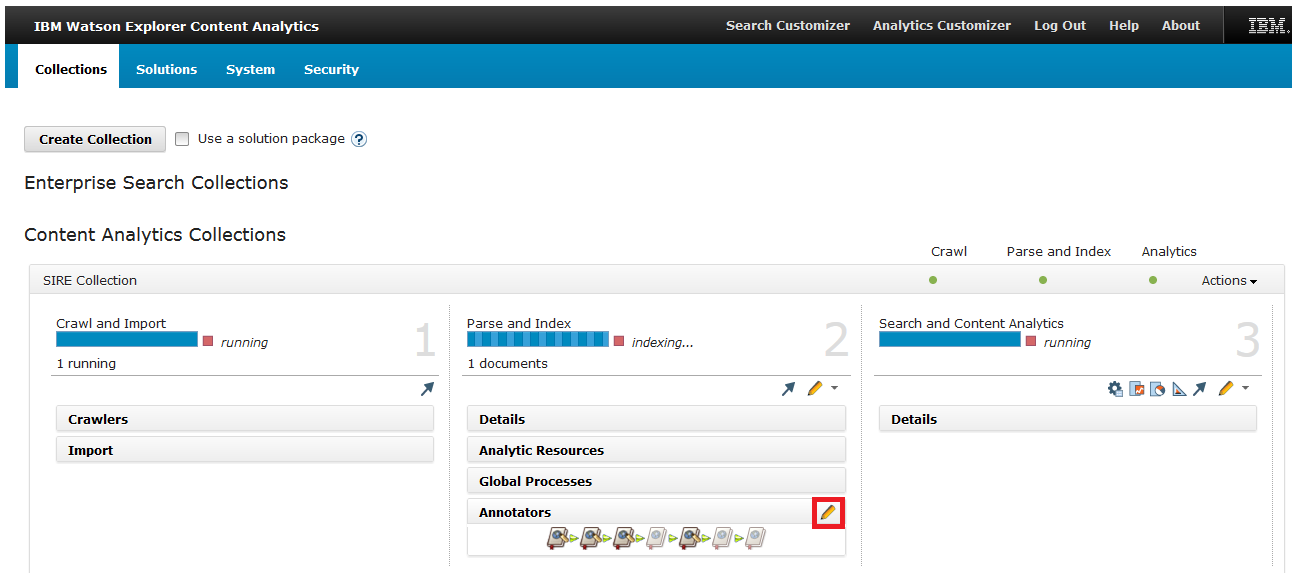
Enter a display name for the model, specify the path where you copied the ZIP file for the exported model, and click **OK**.



Associate the machine-learning model with a collection

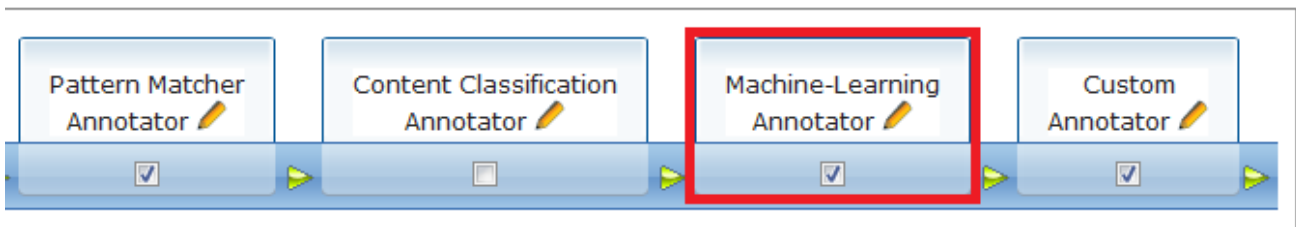
You can associate the same machine-learning model with multiple collections.

To associate a model with a Watson Explorer Content Analytics collection, expand the collection where you want to use a machine-learning annotator that you previously uploaded to the system. In the **Parse and Index** pane, click the **Edit** icon to configure **Annotators**.

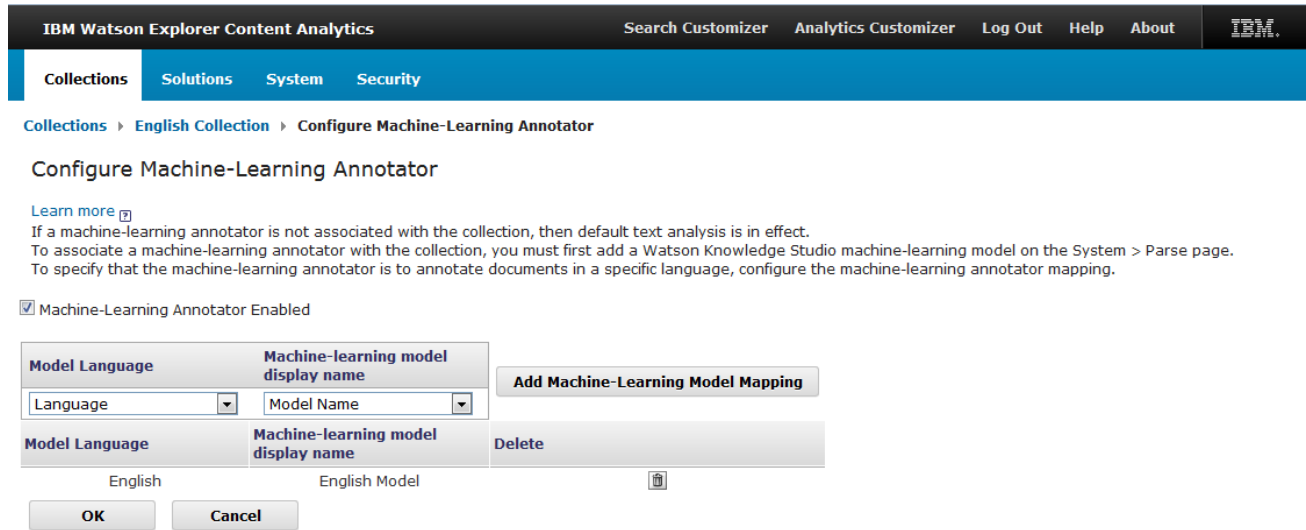


To associate a model with a collection in Annotation Administration Console, expand the collection that you want to configure and click **Actions > Annotators** or click the **Edit** icon in the **Annotators** area.

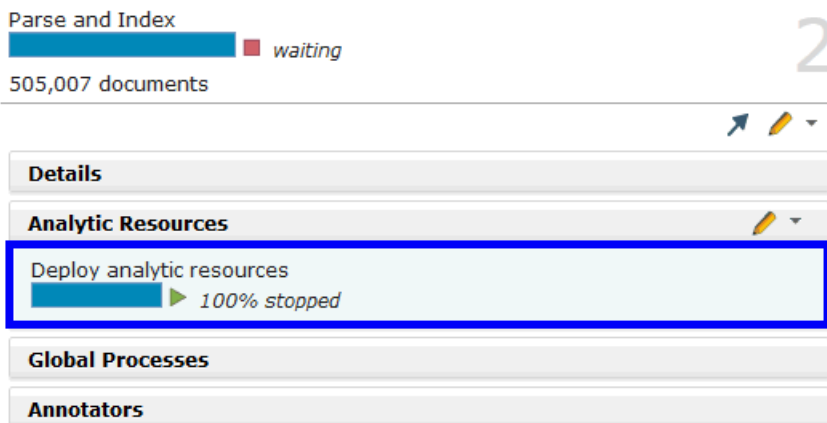
In either interface, select the check box in the **Machine-Learning Annotator** tile to enable the annotator, and then click the **Edit** icon.



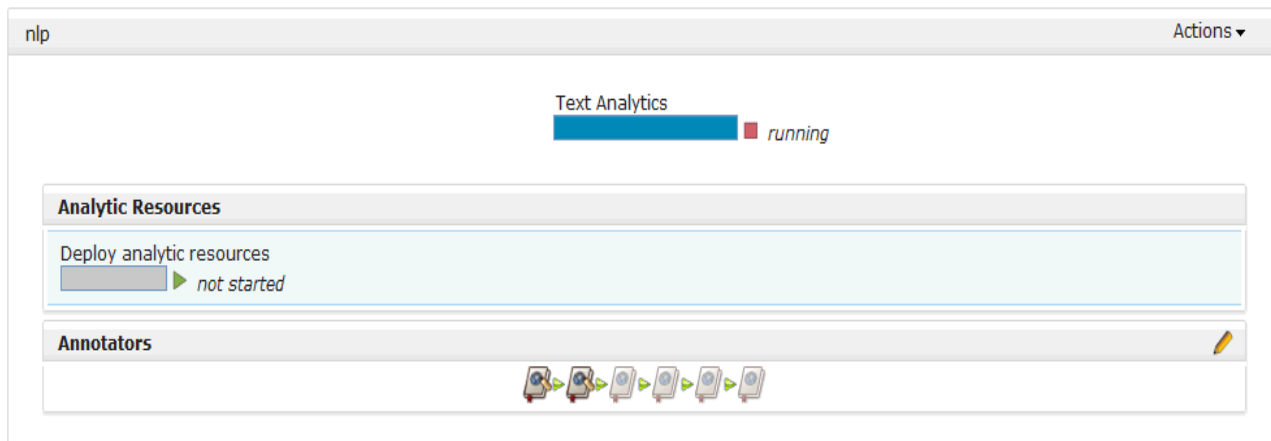
You can associate multiple machine-learning models with a collection. However, you can associate only one model per language. Select the language of a model that you want to use with this collection, and then select the model by the display name that you assigned when you added it to the system. To add another model to the collection, click **Add Machine-Learning Model Mapping**, and the repeat the steps to select the language and model name.



After you set or change the association between a collection and a machine-learning model in Watson Explorer Content Analytics, you must restart the collection’s Parse and Index session and redeploy the analytic resources. For example:

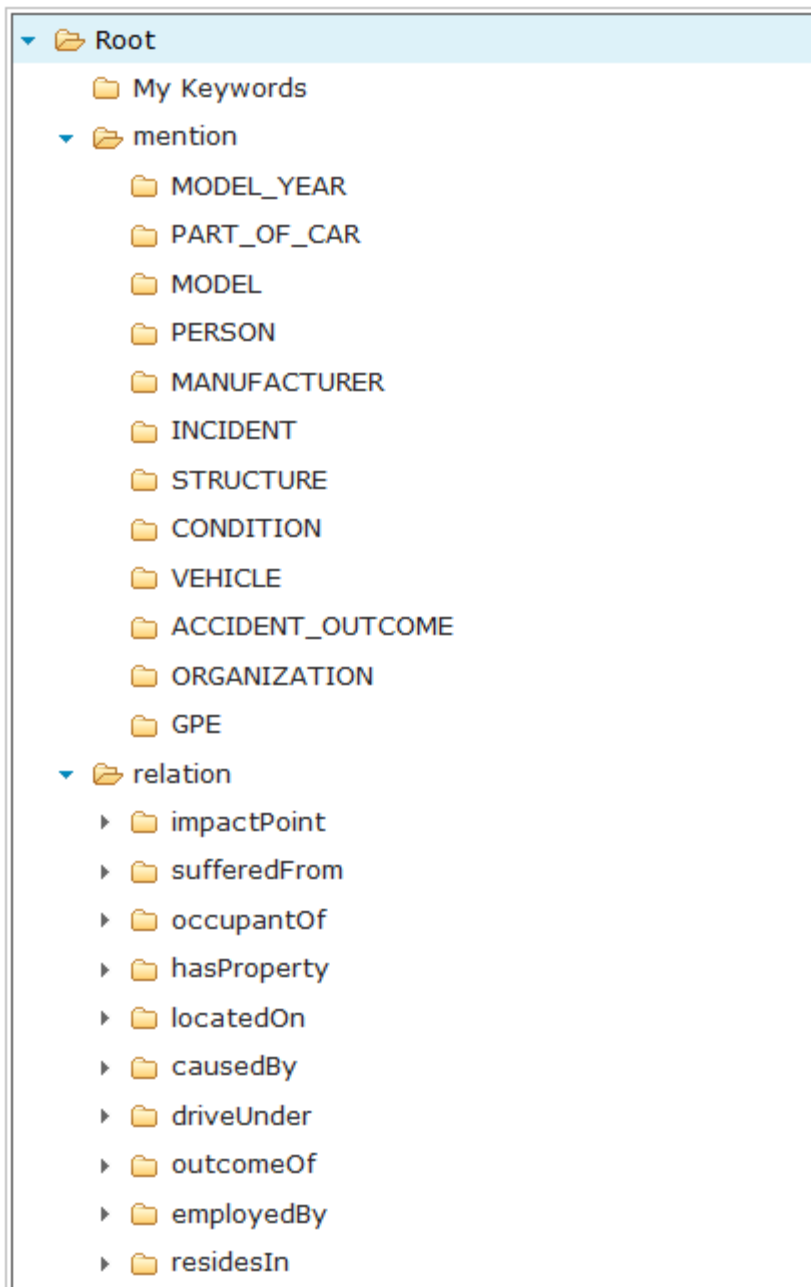


After you set or change the association between a collection and a machine-learning model in Annotation Administration Console, you must stop and restart the Text Analytics session and redeploy the analytic resources. For example:



Automatically generated facet definitions

When you associate machine-learning models with a collection, facet definitions for the collection are automatically created. In Watson Explorer Content Analytics, you can view the facet definitions by viewing the collection's facet tree. To check the facet definitions, click the **Edit** icon to configure **Analytic Resources** -> **Facet Tree** in the **Parse and Index** pane of the collection. The following facet tree example shows mention facets and relation facets that were created for a machine-learning model that was trained to annotate documents about traffic incidents.



You can change the facet names and delete facet definitions that you do not need. However, you cannot merge multiple facet definitions to create one facet. To preserve indexed data, these generated facet definitions are not deleted even if you disassociate the machine-learning model from the collection.

Anytime that you change the facet definitions, you must restart the collection's Parse and Index session and redeploy the analytic resources, as shown on page 6.

Explore analytical results in Watson Explorer Content Analytics

After crawling and indexing documents, you can explore the collection by selecting mention facets in the content analytics miner. If relations were annotated in the model before the model was exported, you can also view and select relation facets in the content analytics miner. The following examples show the imported mention and relation facets in the Facets view:

The screenshot shows the 'Facets' view in Watson Explorer. The left sidebar is titled 'Facet Navigation' and has a search filter. Under 'mention', several facets are listed, with 'PART_OF_CAR' highlighted. The main area shows a table of facets with columns for 'Values', 'Frequency', and 'Correlation'. The table contains five rows of facets, each with a frequency of 1 and a correlation of 1.0.

Values	Frequency	Correlation
antilock brakes	1	1.0
front door-mounted side impact air bags	1	1.0
side impact Inflatable Curtain (IC) air bags	1	1.0
side impact inflatable occupant protection system	1	1.0
Certified Advanced 208-Compliant (CAC) frontal air bag system	1	1.0

The screenshot shows the 'Facets' view in Watson Explorer. The left sidebar is titled 'Facet Navigation' and has a search filter. Under 'relation', several facets are listed, with 'relation' highlighted. The main area shows a table of facets with columns for 'Values', 'Frequency', and 'Correlation'. The table contains eight rows of facets, each with a frequency of 1 and a correlation of 1.0.

Values	Frequency	Correlation
impactPoint : crash - that	1	1.0
locatedOn : Certified Advanced 208-Compliant (CAC) frontal air bag system - vehicle	1	1.0
locatedOn : antilock brakes - vehicle	1	1.0
locatedOn : front door-mounted side impact air bags - vehicle	1	1.0
locatedOn : side impact Inflatable Curtain (IC) air bags - vehicle	1	1.0
locatedOn : side impact inflatable occupant protection system - BMW	1	1.0
quantityOf : 多< - 人	1	1.0

Usage Guidelines

- Facet values represent the actual results from the machine-learning annotator without text normalization, such as lemmatization. For example, facet values might include 'occupant' and 'occupants' or 'she' and 'She'.
- Coreferenced mentions appear only in relation facets, not as stand-alone mention facets. In the relation facet, a representative mention is displayed instead of the original mention if the mention is part of a coreference chain. A representative mention is the referent mention in the coreference chain. For example, if a relation was defined as '**She**' - '**sufferedFrom**' - '**injuries**' when the machine-learning model was trained in Watson Knowledge Studio, Watson Explorer might display the value 'driver' instead of 'She' if 'She' and 'driver' belong to the same coreference chain.

Explore analytical results in Watson Explorer applications

After configuring a collection use a machine-learning annotator, your applications can use it the same way that they use a custom annotator. The text analytics API, which is provided with Watson Explorer Engine, pulls analysis results from Annotation Administration Console and makes the results available to your Watson Explorer applications.

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make

improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows: © (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. 2004, 2016. All rights reserved.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.