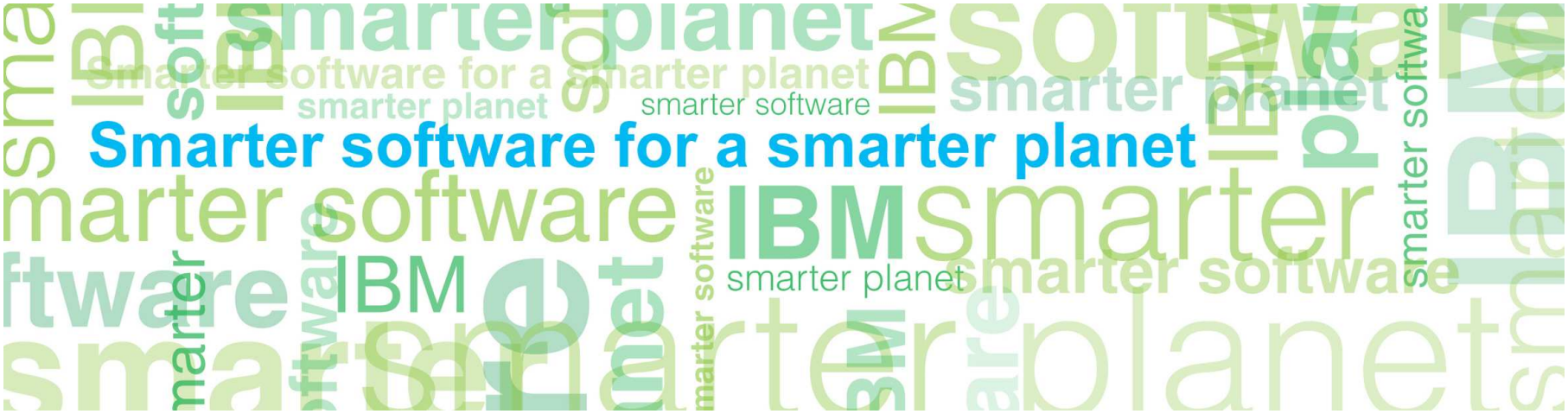
LanguageWare Resource Workbench 7.2

Export an annotator as a PEAR file for IBM Content Analytics V2.2



Introduction

- **Module overview**
  - How to export an annotator as an UIMA PEAR file for ICA
  - How to deploy the UIMA PEAR file in ICA
  - How to import search fields and create facets in ICA
  - Best practices
- **Target audience:**
  - All audiences
- **Prerequisites:**
  - Install LanguageWare® Resource Workbench (LRW)
  - Complete the examples at the end of the prior training modules to setup the AnalyseHelpline annotator.
  - Create the sample text analytics collection provided with ICA
- **Version Release Date** LRW 7.2, ICA 2.2, released October, 2010

Module objectives

After this course you will be able to:

- Export an annotator as a UIMA PEAR file for IBM Content Analytics (ICA) V2.2
- Deploy the UIMA PEAR file in ICA
- Import search fields and create facets mapped to the annotator in ICA
- View the analytics based on the annotator in ICA.

Module roadmap

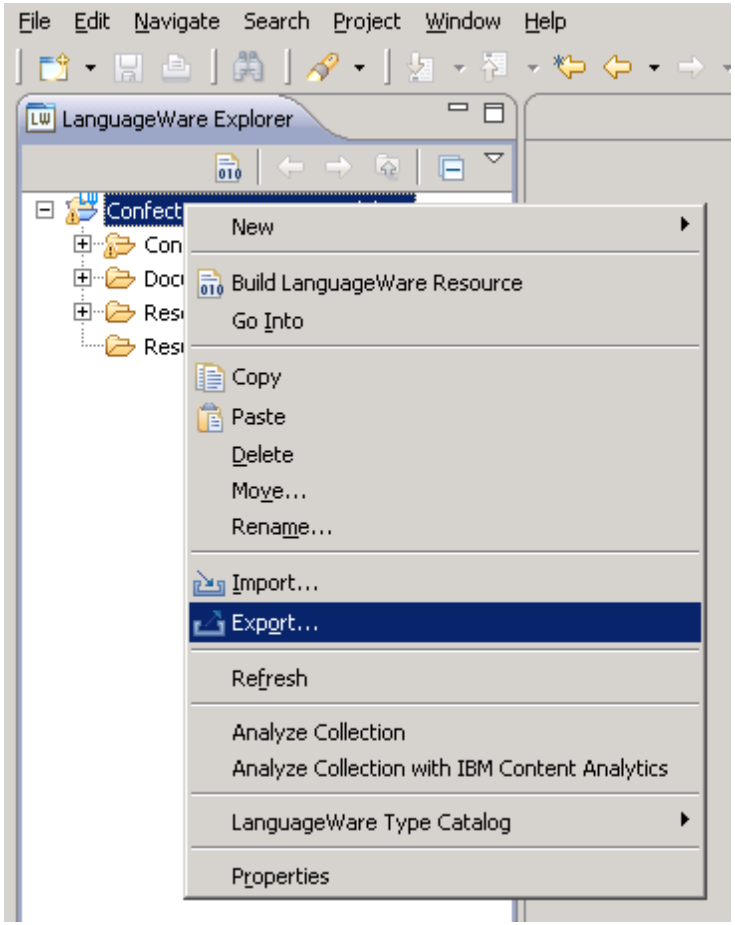
- **Export an annotator as a PEAR file for ICA**
  - What is it?
  - How to configure it?
- **How to deploy a PEAR file in ICA?**
- **How to associate a text engine with a collection?**
- **How to import search fields associated to the PEAR file?**
- **How to create facets and map them to search fields?**
- **How to redeploy resources and rebuild index?**
- **Module summary and best practices**
- **Sample exercises**

Export an annotator as a PEAR file for ICA
What is it?

- **General**
  - Export the annotator from LanguageWare Resource Workbench (LRW) as a PEAR (Processing Engine ARchive file) that contains information specific for ICA.
- **Specific**
  - An existing AnalyseHelpline annotator that was developed in LRW is exported to the file system as an Apache UIMA PEAR file named AnalyseHelplineICA.pear.
  - The exported UIMA PEAR file is deployed in ICA as a text analysis engine.
  - The text analysis engine is associated to a collection in ICA.
  - This functionality is useful if you want to work with your annotator in ICA to use text analytics and you currently do not have connectivity to ICA in order to deploy it. In particular, a business partner might want to use this option to deliver and deploy a previously created annotator in a customer's environment.

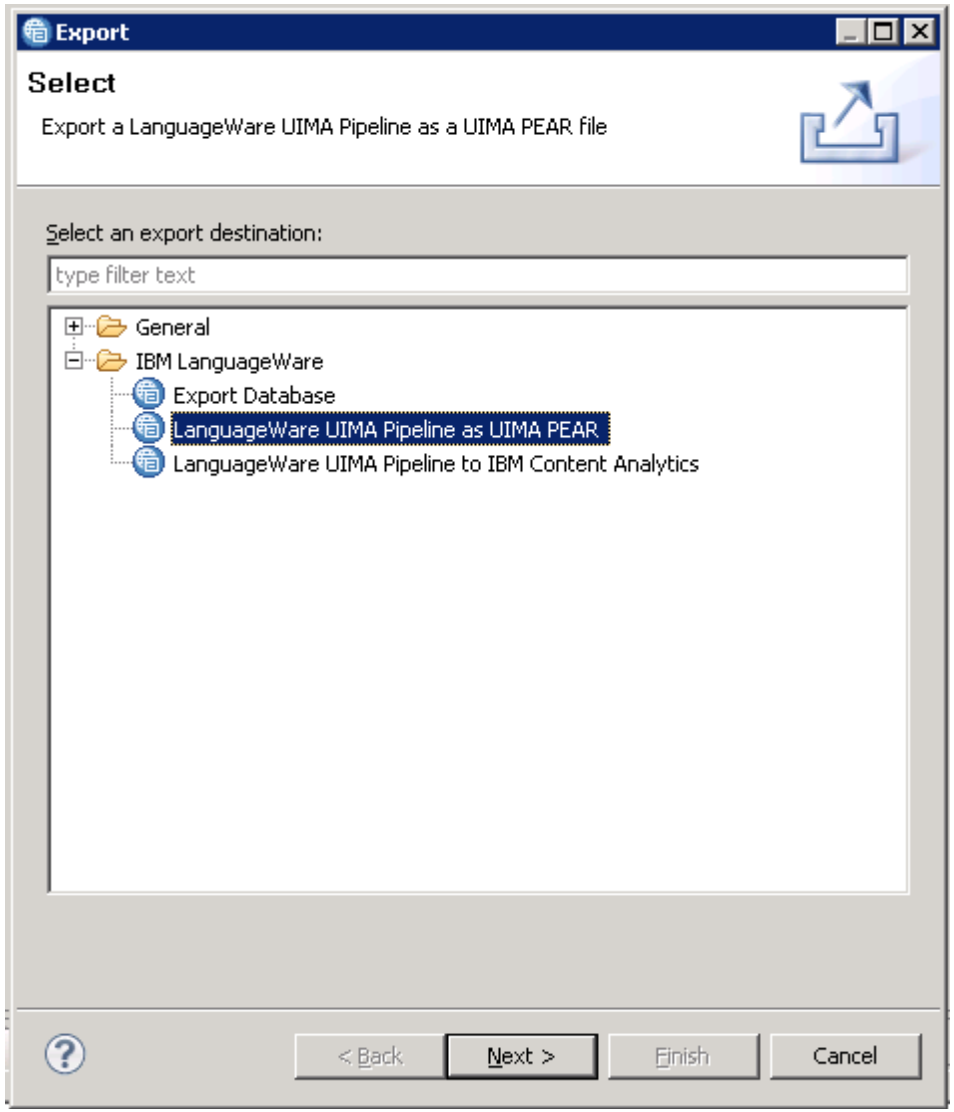
Export an annotator as a PEAR file for ICA how to configure it?

- Right click the project name for the annotator you want to export. For this example, right click **ConfectionaryCompanyHelpline**
- Select **Export**



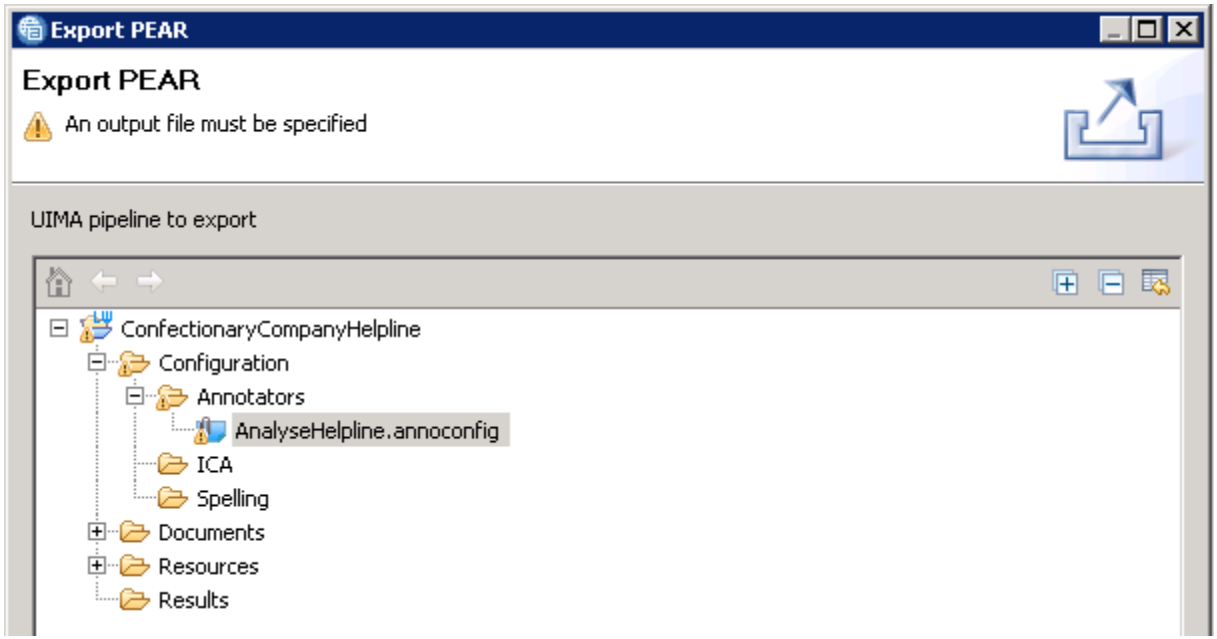
Export an annotator as a PEAR file for ICA how to configure it?

- Expand the **IBM LanguageWare** folder
- Select **LanguageWare UIMA Pipeline as UIMA PEAR**
- Click **Next**



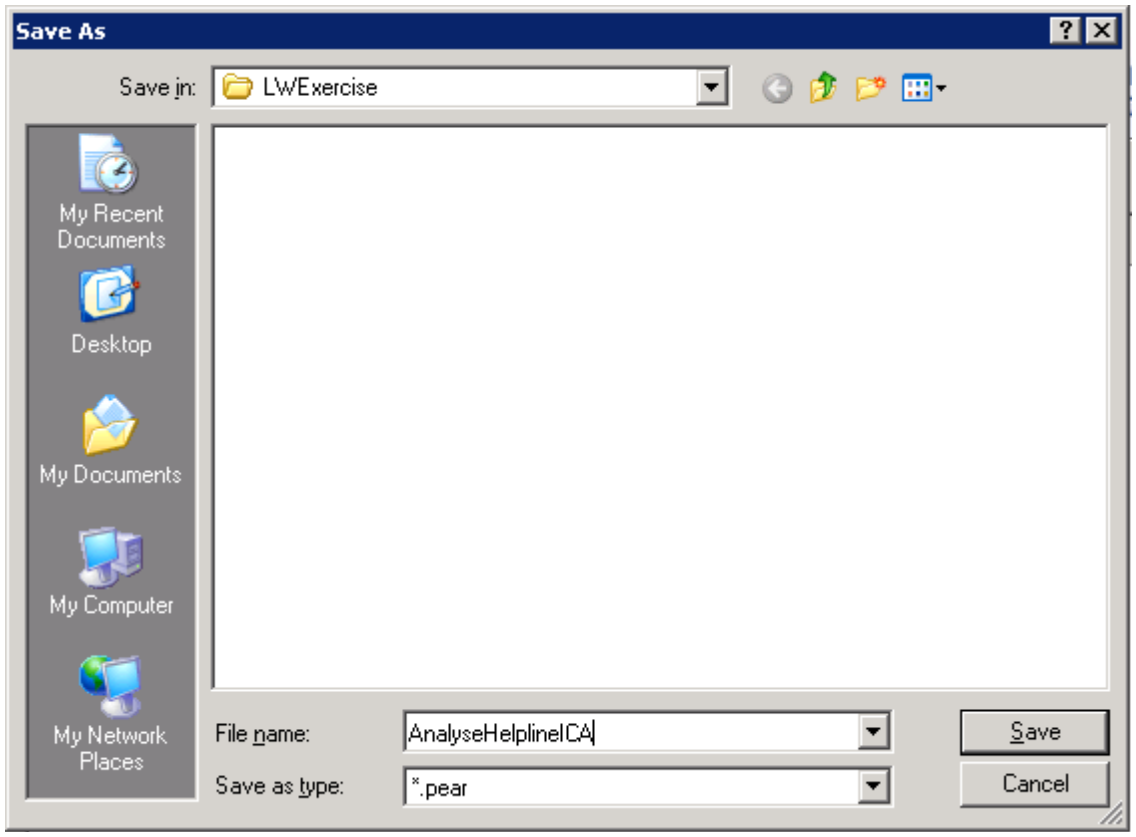
Export an annotator as a PEAR file for ICA how to configure it?

- Expand **ConfectionaryCompanyHelpline**
- Expand **Configuration**
- Expand **Annotators**
- Select **AnalyseHelpline.annoconfig**



Export an annotator as a PEAR file for ICA how to configure it?

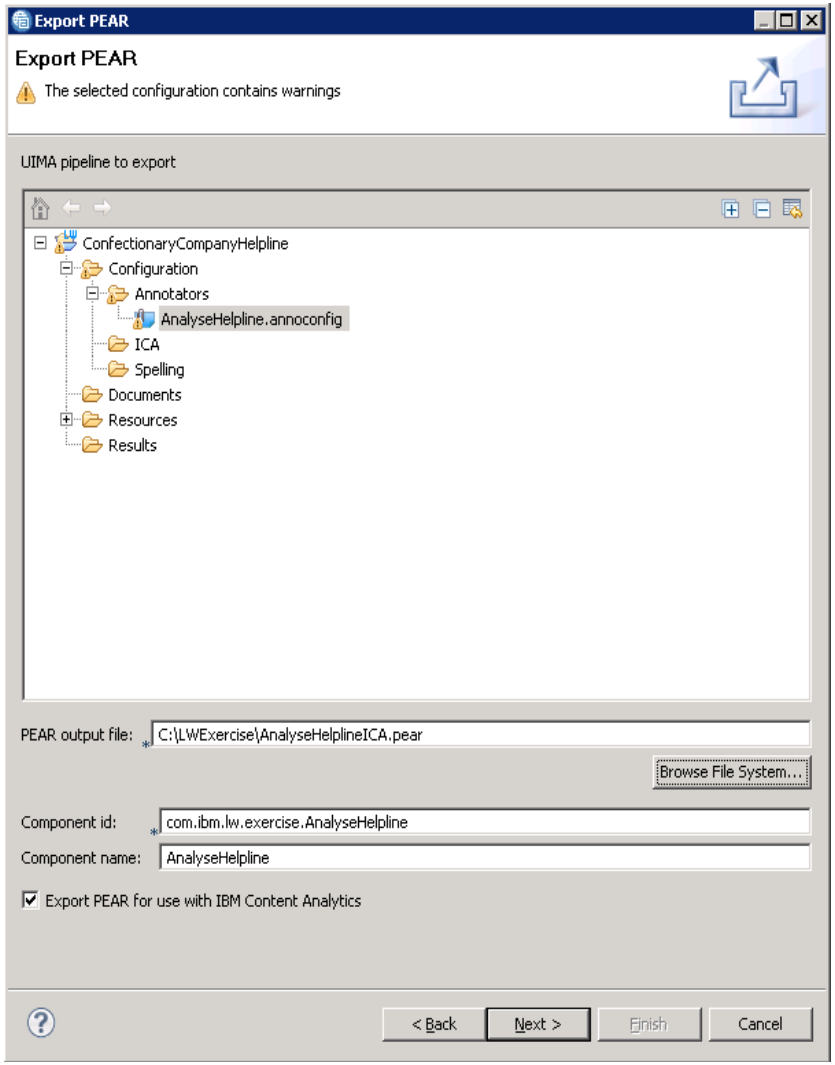
- Click **Browse File System...**
- Change the directory to your desired location of the PEAR file.
- For the File name field, type AnalyseHelplineICA
- Click **Save**



Export an annotator as a PEAR file for ICA how to configure it?

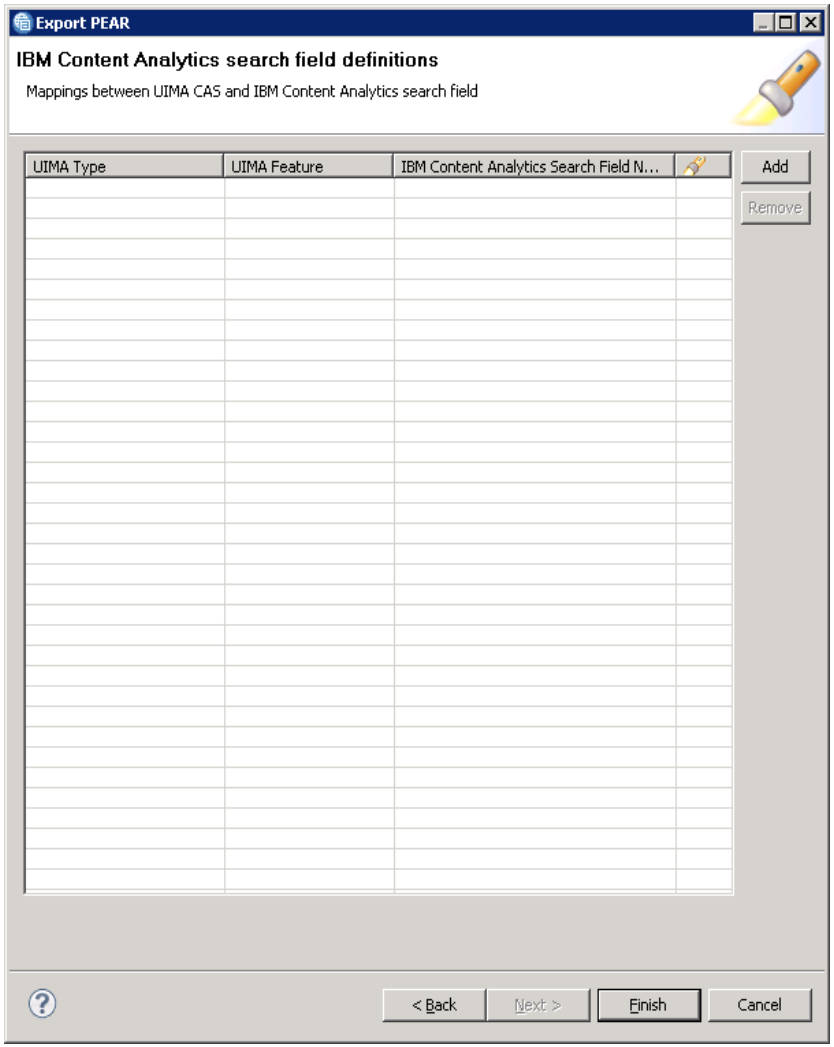
- The PEAR output file field is now populated
- Select the **Export PEAR for use with IBM Content Analytics** check box
- Click **Next**

Note: By selecting **Export PEAR for use with IBM Content Analytics** option, the PEAR includes files to integrate with ICA. In particular, the cas2index.xml and searchfield.xml files are included in the PEAR.



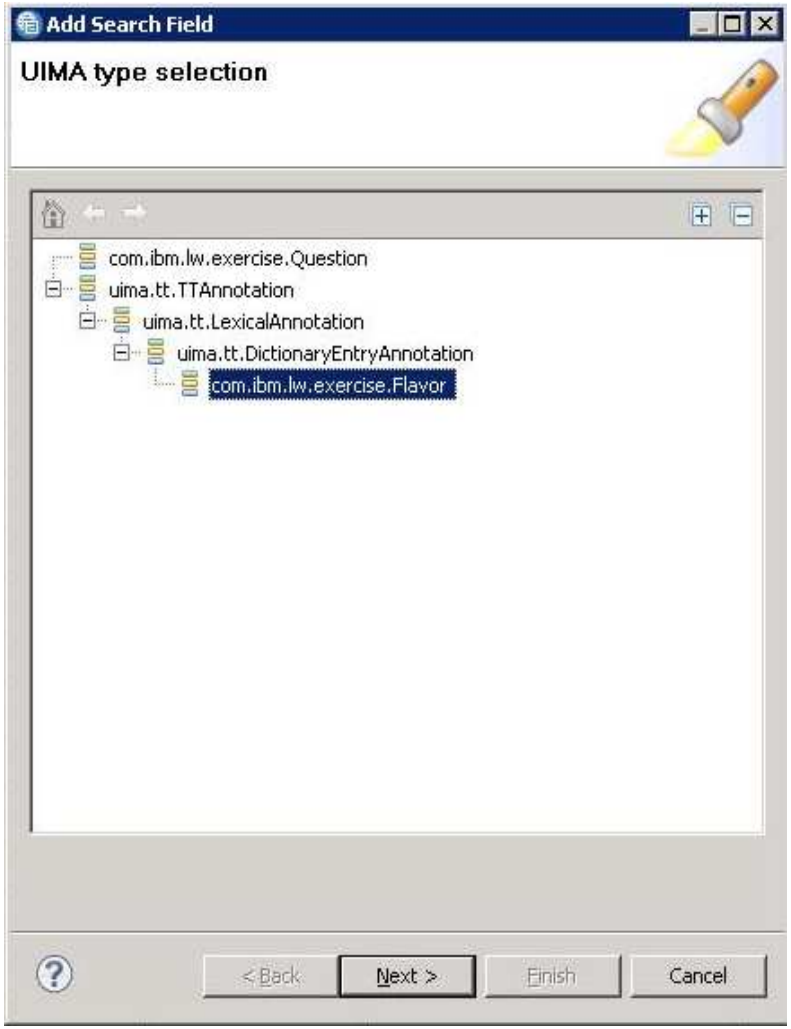
Export an annotator as a PEAR file for ICA how to configure it?

- Click **add** to add a mapping between UIMA CAS and a search field in IBM content analytics
- A dialog box is displayed (see next slide)



Export an annotator as a PEAR file for ICA
how to configure it?

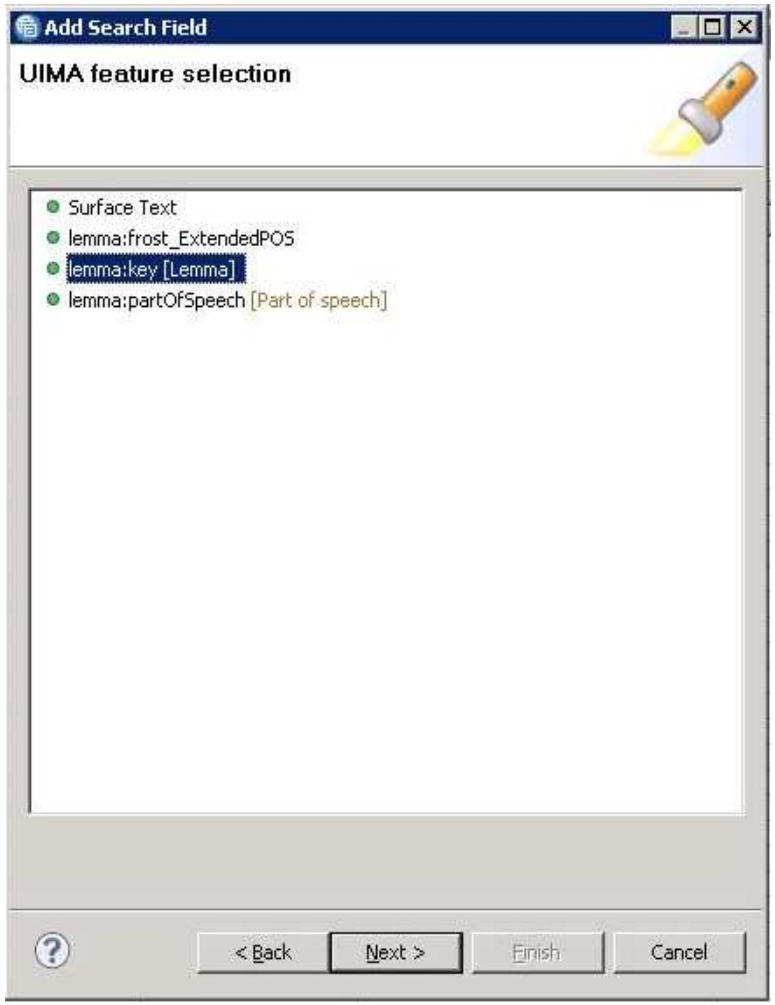
- Select the UIMA type that you want to use for the search field in ICA. For this example, select **com.ibm.lw.exercise.Flavors** as the UIMA type
- Click **Next**



Export an annotator as a PEAR file for ICA how to configure it?

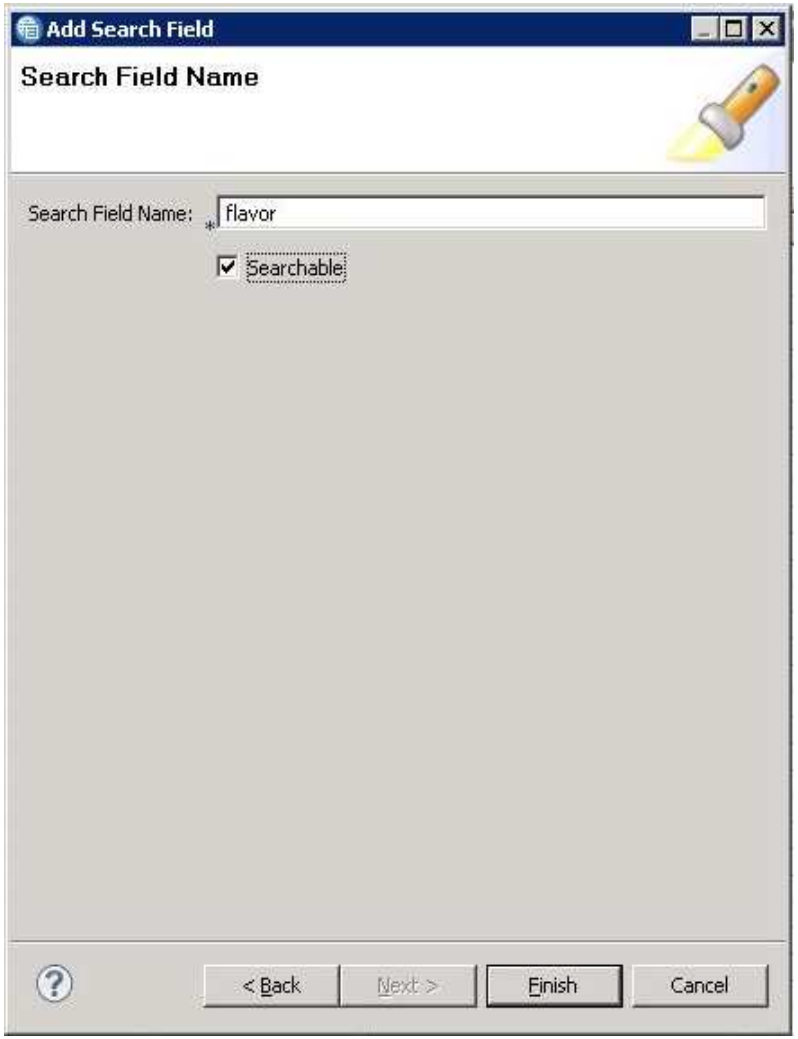
- Select **lemma:key** as the UIMA feature
- Click **Next**

Note: By selecting the lemma:key feature, you are allowing the lemma forms of the word to be included in the search results within ICA. For example, a search on the term “run” returns documents containing run, runs, running, and ran within the ICA text miner application



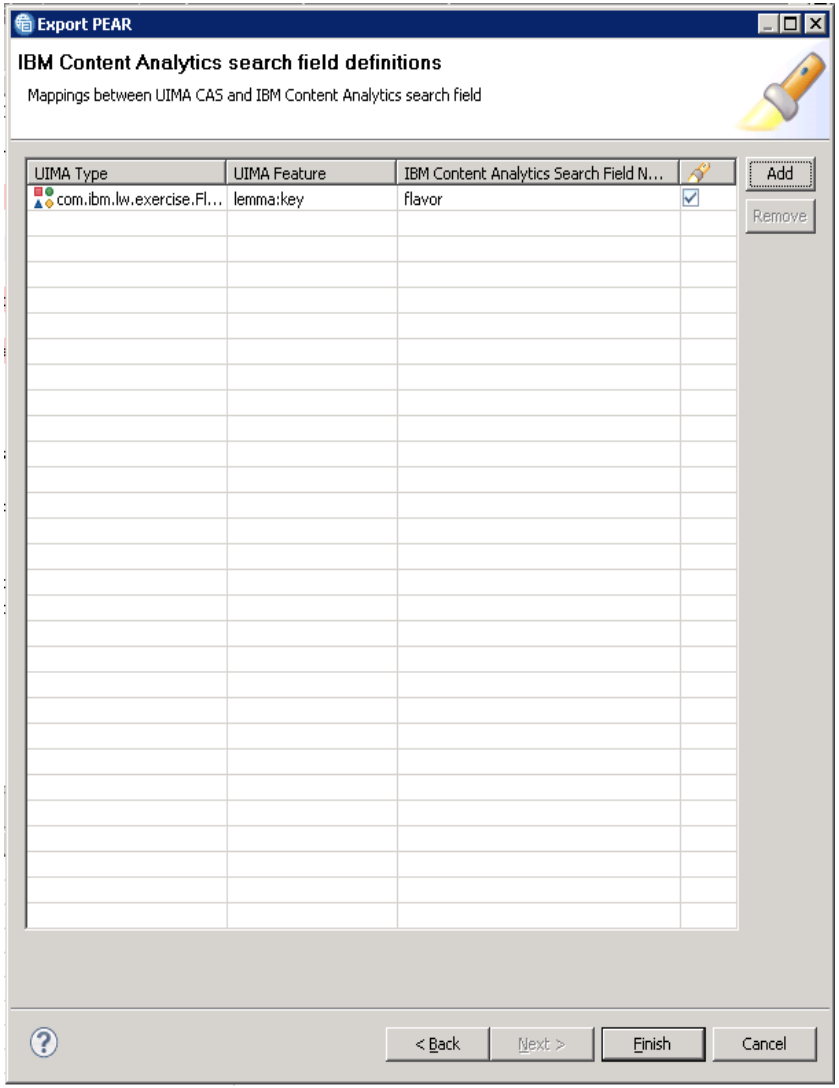
Export an annotator as a PEAR file for ICA how to configure it?

- Enter the search field name that you want mapped to the UIMA feature. For this example, type **flavor** for Search Field Name
- Select the **Searchable** check box so that the field name is able to be used as search criteria in ICA
- Click **Finish**



Export an annotator as a PEAR file for ICA how to configure it?

- The mapping between the UIMA CAS and IBM Content Analytics is displayed
- Click **Finish**



Export an annotator as a PEAR file for ICA how to configure it?

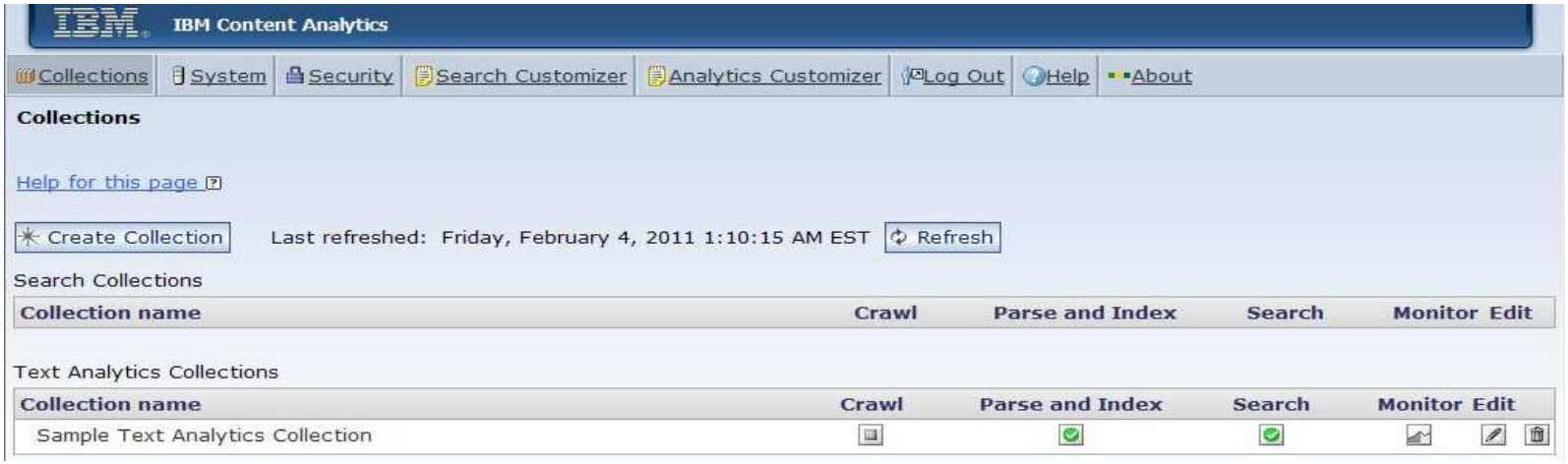
- A confirmation message indicating that the export is complete will appear
- Click **OK**

Module roadmap

- Export an annotator as a PEAR file for ICA
  - What is it?
  - How to configure it?
- **How to deploy a PEAR file in ICA?**
- How to associate a text engine with a collection?
- How to import search fields associated to the PEAR file?
- How to create facets and map them to search fields?
- How to redeploy resources and rebuild index?
- Module summary and best practices
- Sample exercises

Deploy PEAR file in ICA

- Open IBM Content Analytics administration console. In a web browser, type: http://<IP Address>:<port>/ESAdmin
- Log in using the ICA system administrator user ID

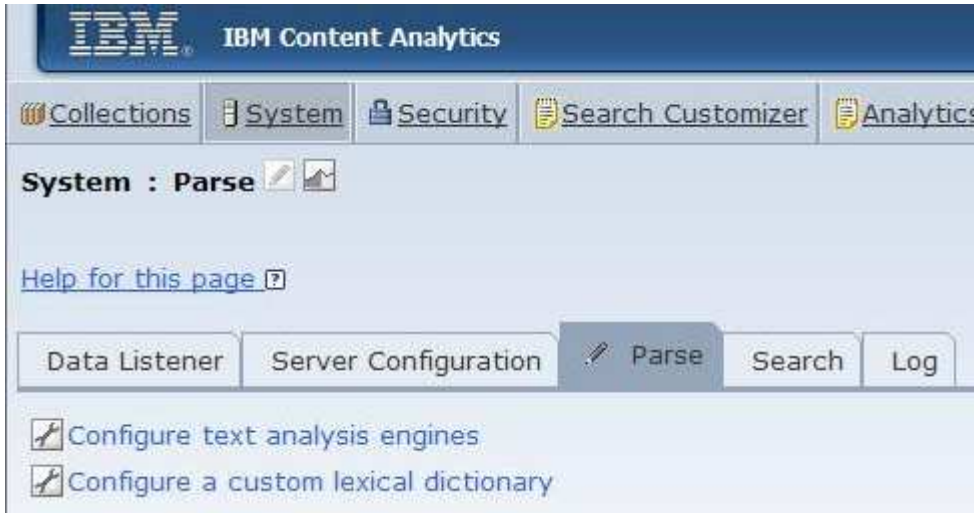


Deploy PEAR file in ICA
how to configure it?

- Click the **System** tab.
- Click the **Edit** icon.

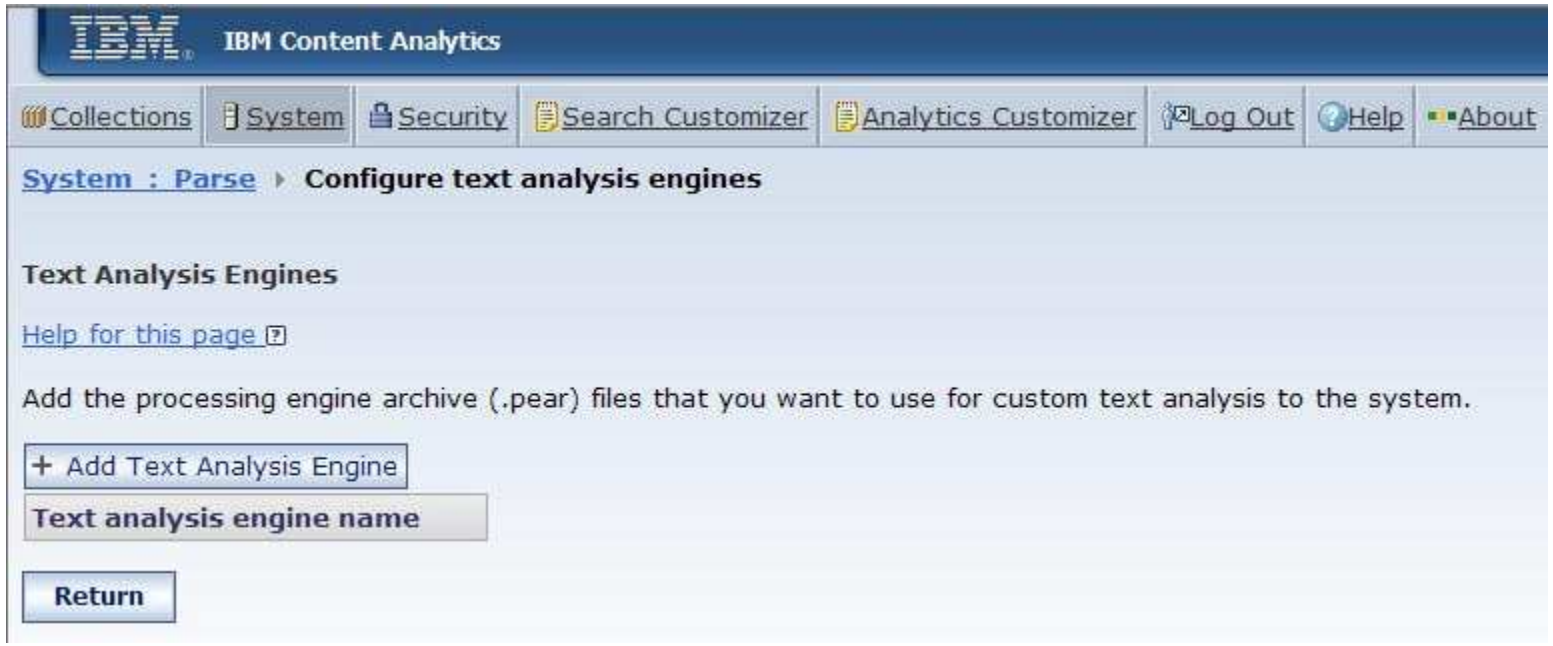
Deploy PEAR file in ICA
how to configure it?

- Click the **Parse** tab
- Click **Configure text analysis engines**



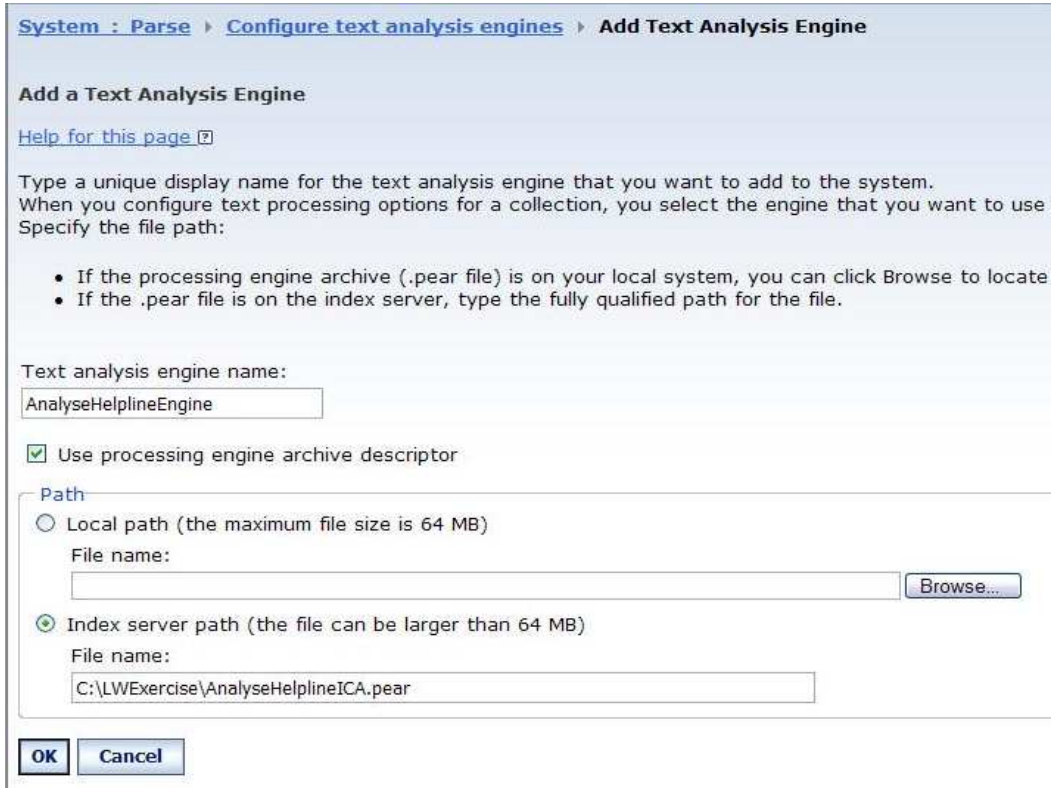
Deploy PEAR file in ICA
how to configure it?

- Click **Add Text Analysis Engine**



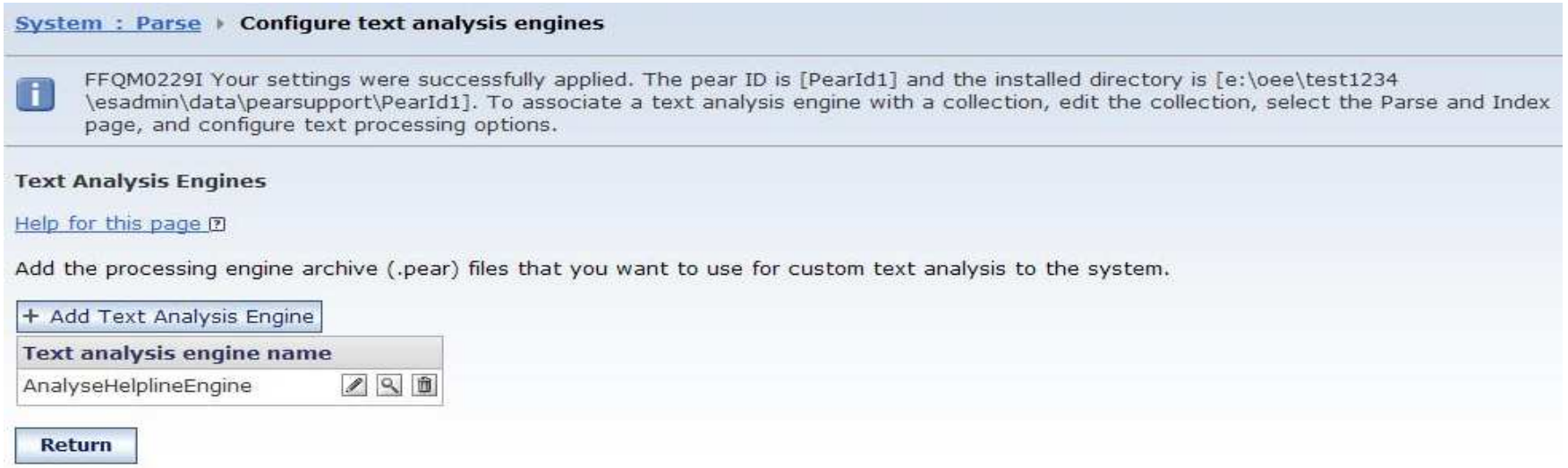
Deploy PEAR file in ICA
how to configure it?

- In the Text analysis engine name field, type an unique name for the engine. For this example, type AnalyseHelplineEngine
- Select the **Use processing engine archive descriptor** check box
- Enter the path to the PEAR file that you exported from LRW. If the PEAR file is located on your local machine, select the **Local Path** radio button. If the PEAR file is located on the ICA machine, select the **Index server path** radio button.
- Click **OK**



Deploy PEAR file in ICA
how to configure it?

- The text analysis engine name now displays in the list of engines
- Click **Return**
- Record the path to the pear ID that is displayed in the success message at the top of the screen. For this example, you would write down: e:\oee\test1234\esadmin\data\pearsupport\PearId1 as shown in the image below

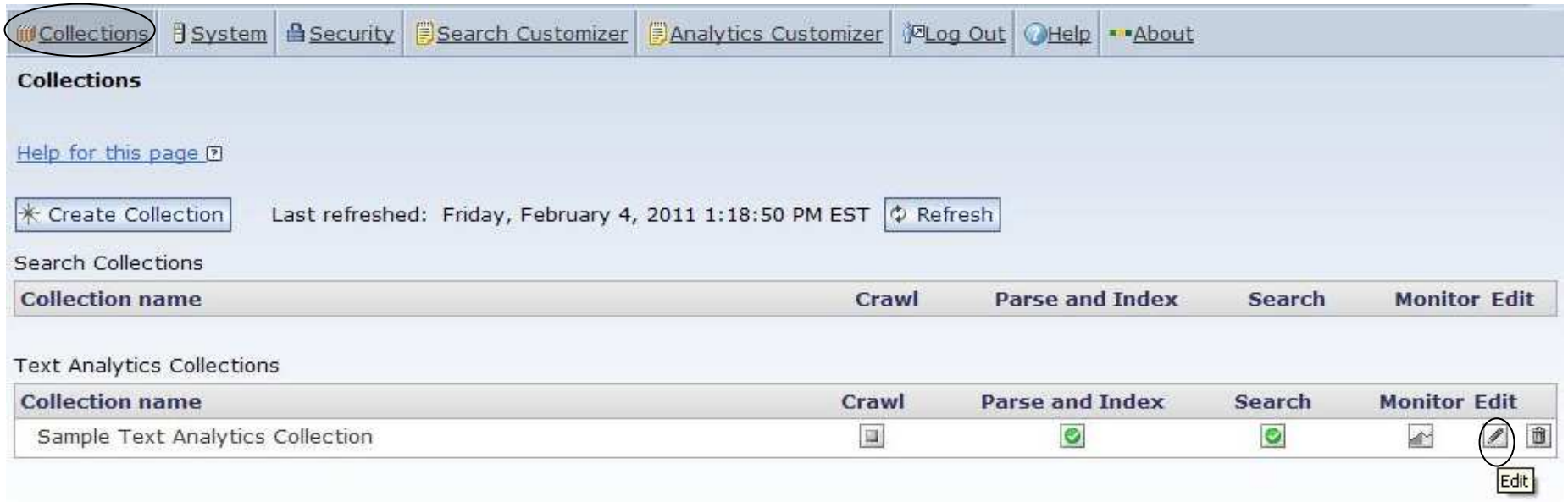


Module roadmap

- Export an annotator as a PEAR file for ICA
  - What is it?
  - How to configure it?
- How to deploy a PEAR file in ICA?
- **How to associate a text engine with a collection?**
- How to import search fields associated to the PEAR file?
- How to create facets and map them to search fields?
- How to redeploy resources and rebuild index?
- Module summary and best practices
- Sample exercises

Associate a text engine with an ICA collection how to configure it?

- Now that the PEAR file is deployed in ICA, it needs to be associated to a collection so that it can be used for text analytics. Click the **Collections** tab
- Click the **Edit** icon for the collection that you want the PEAR file to be associated with. For this example, use the Sample Text Analytics Collection

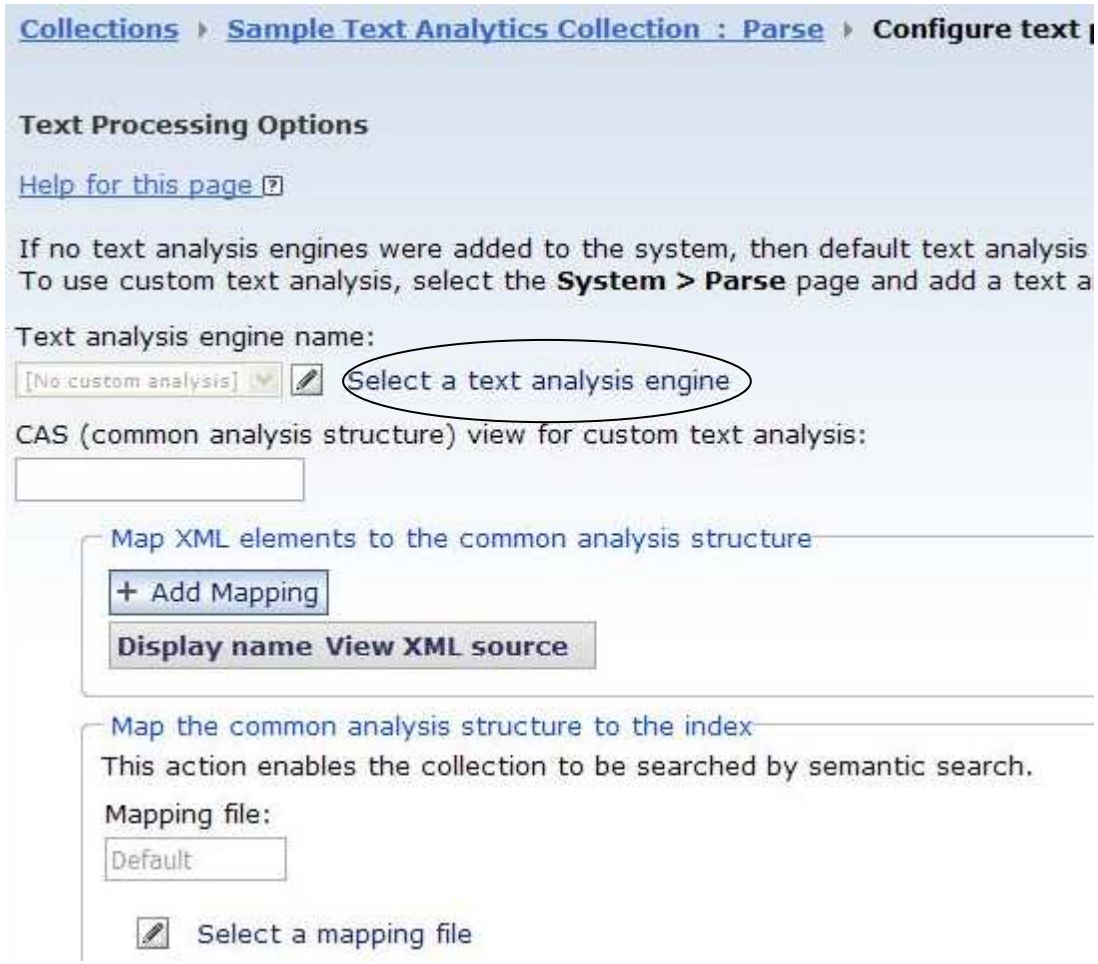


Associate a text engine with an ICA collection how to create and configure it?

- Click the **Parse and Index** tab
- Click **Configure text processing options**

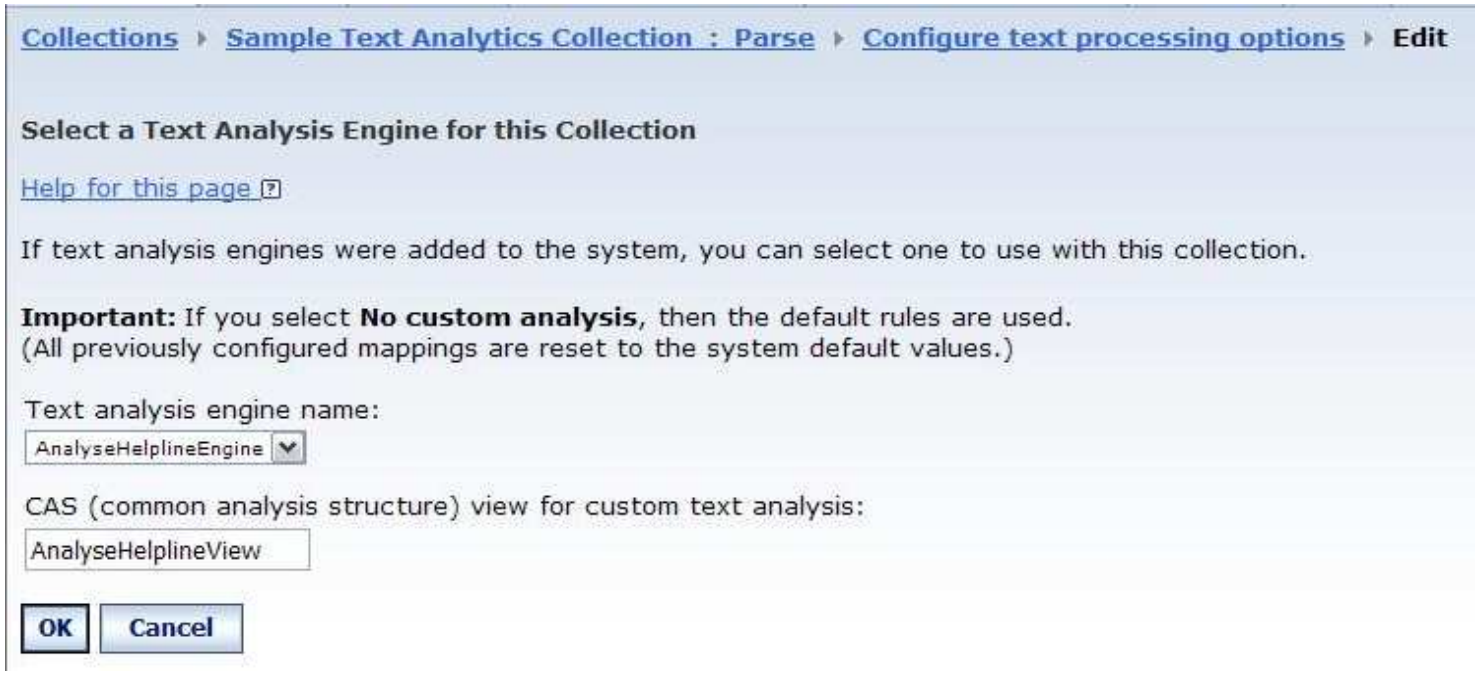
Associate text engine with an ICA collection how to configure it?

- Click **Select a text analysis engine**



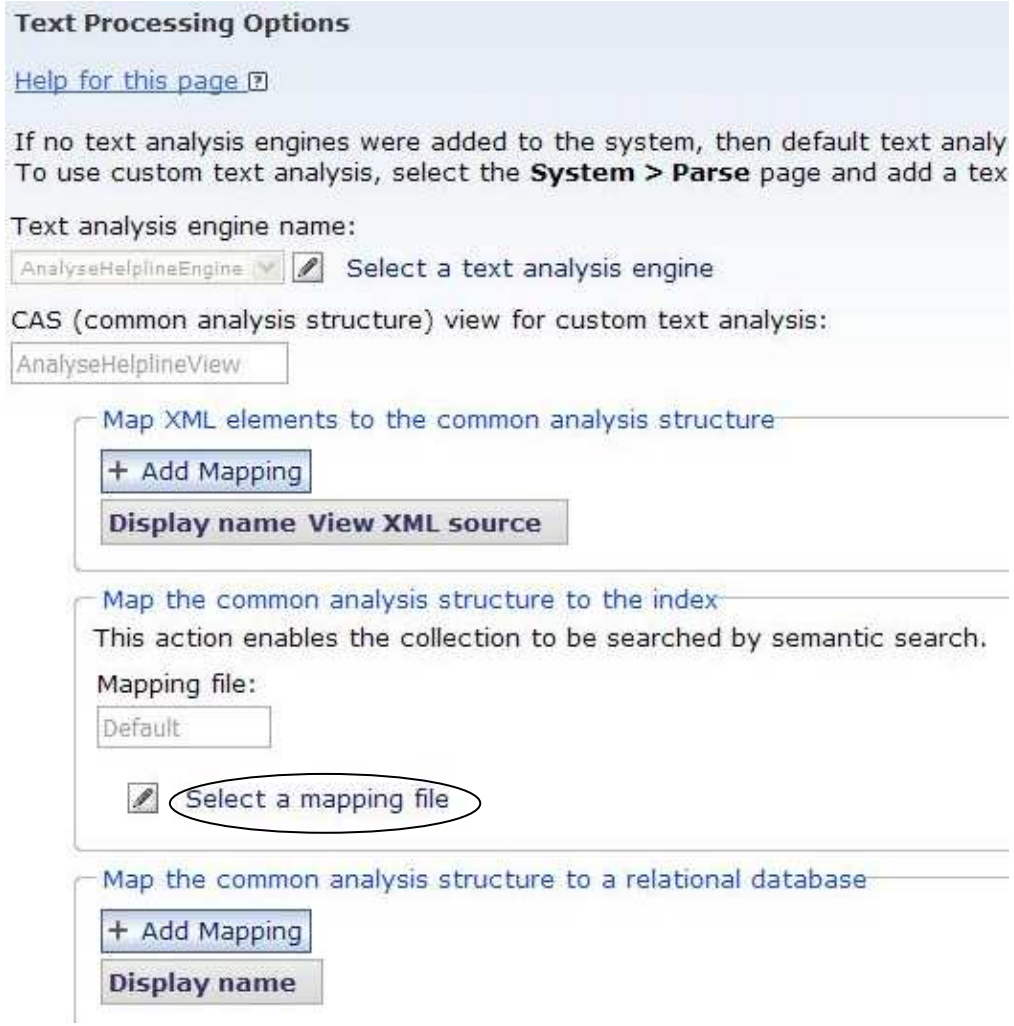
Associate text engine with an ICA collection how to configure it?

- Select the engine that you created using your PEAR file in the Text analysis engine name field. For this example, select **AnalyseHelplineEngine**
- Provide a unique name in the CAS view for custom text analysis field. For this example, type AnalyseHelplineView
- Click **OK**



Associate text engine with an ICA collection how to configure it?

- Click **Select a mapping file**



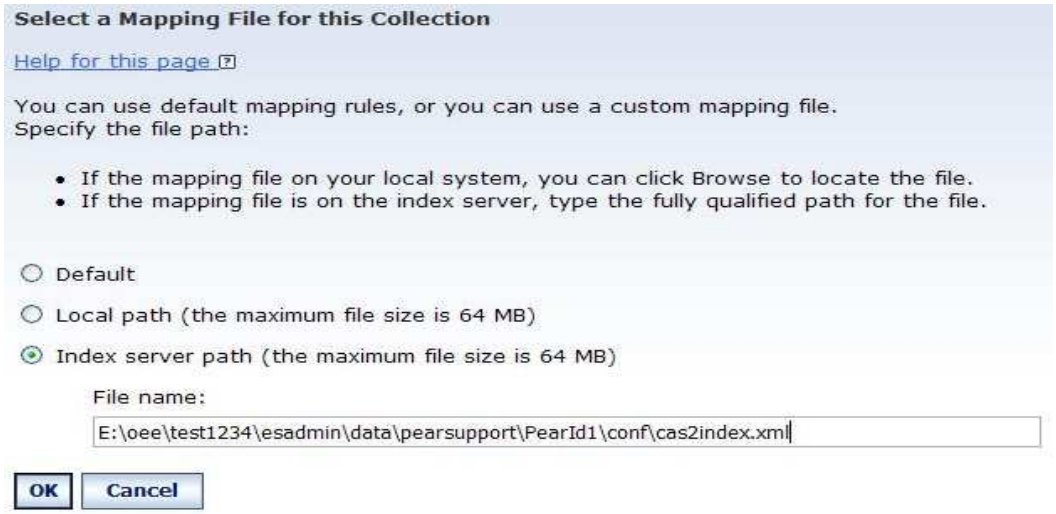
Associate text engine with an ICA collection how to configure it?

- Select the **Index server path** radio button
- Type the path to the cas2index.xml file that is located on the ICA server machine. The cas2index.xml file is located in the following location:

  <ES_Node_Root>\data\pearsupport\<PearId>\conf\cas2index.xml

  ES_Node_Root represents the data directory defined during install. PearId represents the ID for the PEAR file that you deployed in ICA. You recorded this path after the successful text engine deployment in slide 22

- Click **OK**
- Click **Return**

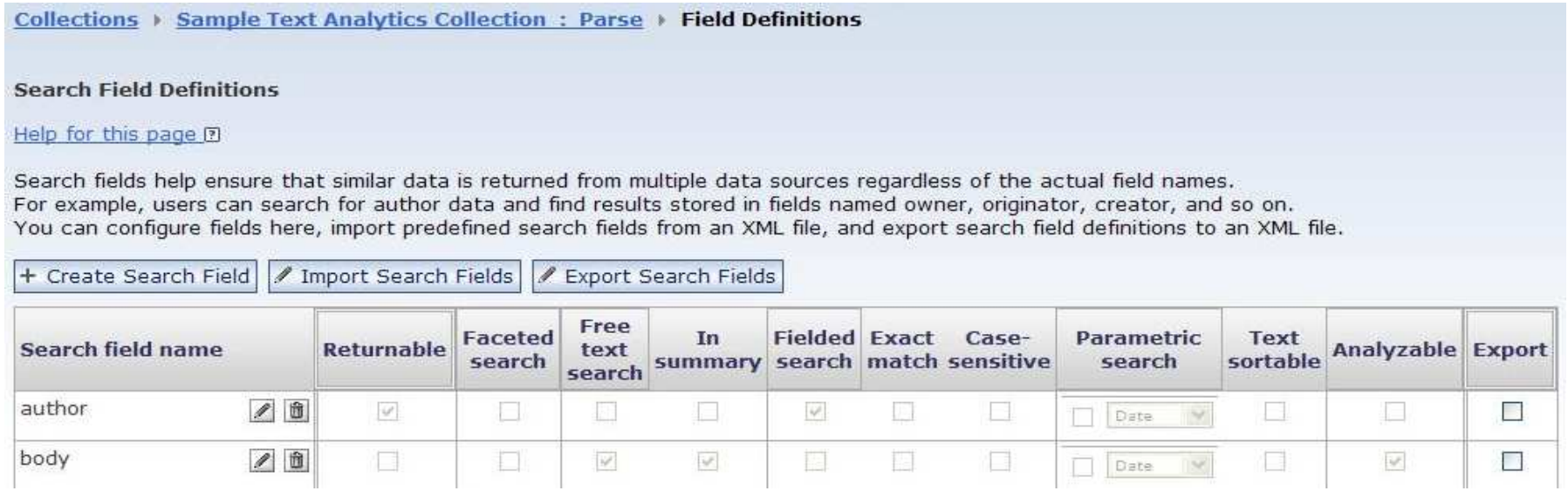


Module roadmap

- Export an annotator as a PEAR file for ICA
  - What is it?
  - How to configure it?
- How to deploy a PEAR file in ICA?
- How to associate a text engine with a collection?
- **How to import search fields associated to the PEAR file?**
- How to create facets and map them to search fields?
- How to redeploy resources and rebuild index?
- Summary and best practices
- Sample exercises

Import search fields associated to the annotator how to configure it?

- On the Parse and Index tab, click **Configure Search Field**
- Click **Import Search Fields**



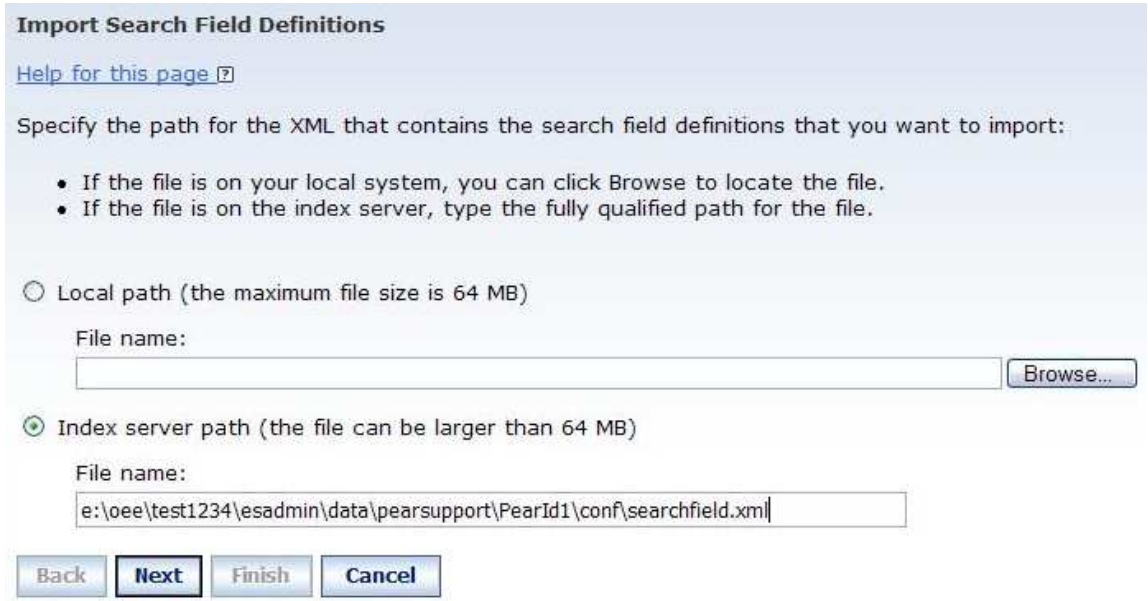
Import search fields associated to the annotator how to configure it?

- Select the **Index server path** radio button
- Type the path to the searchfield.xml file that is located on the ICA server machine. The searchfield.xml file is located in the following location:

  <ES_Node_Root>\data\pearsupport\<PearId>\conf\searchfield.xml

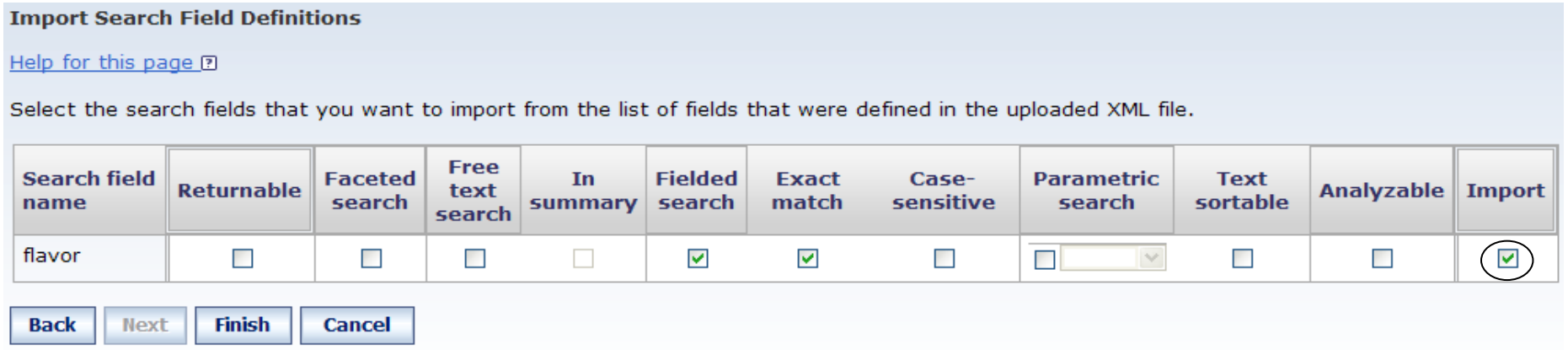
  ES_Node_Root represents the data directory defined during install. PearId represents the ID for the PEAR file that you deployed in ICA. You recorded this path after the successful text engine deployment in slide 22

- Click **Next**



Import search fields associated to the annotator how to configure it?

- The search fields to import are shown. Select the search fields that you want to import by selecting the Import check box. In this example, ensure that the **Import** check box for the flavor search field is selected
- Click **Finish**
- Click **Return**

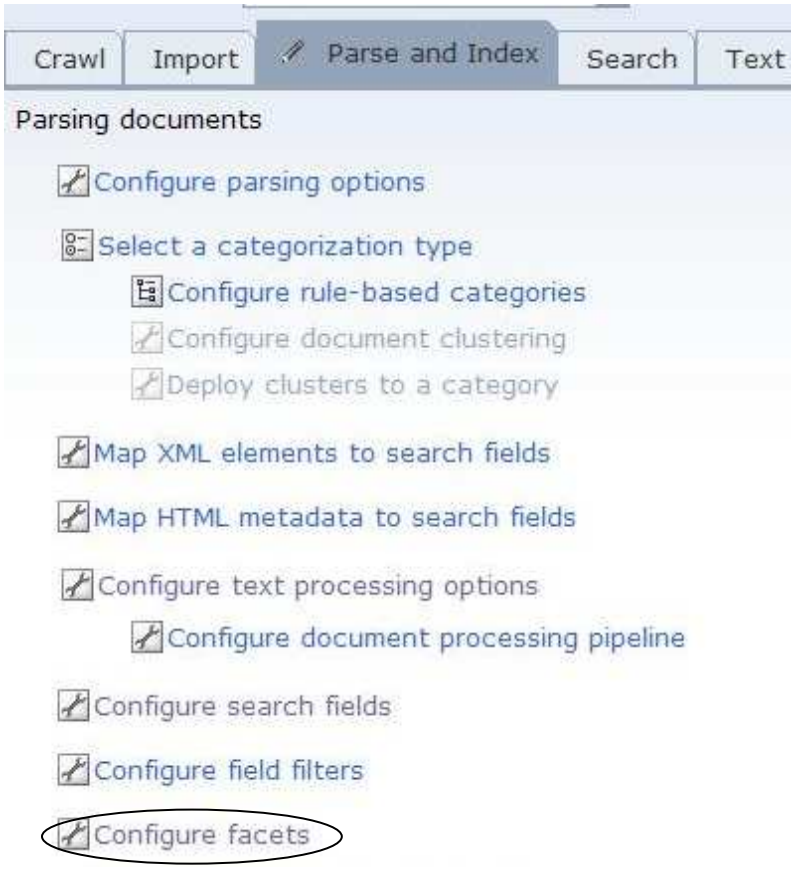


Module roadmap

- Export an annotator as a PEAR file for ICA
  - What is it?
  - How to configure it?
- How to deploy a PEAR file in ICA?
- How to associate a text engine with a collection?
- How to import search fields associated to the PEAR file?
- **How to create facets and map them to search fields?**
- How to redeploy resources and rebuild index?
- Module summary and best practices
- Sample exercises

Create facets mapped to the search fields how to configure it?

- On the Parse and Index screen, click **Configure facets**

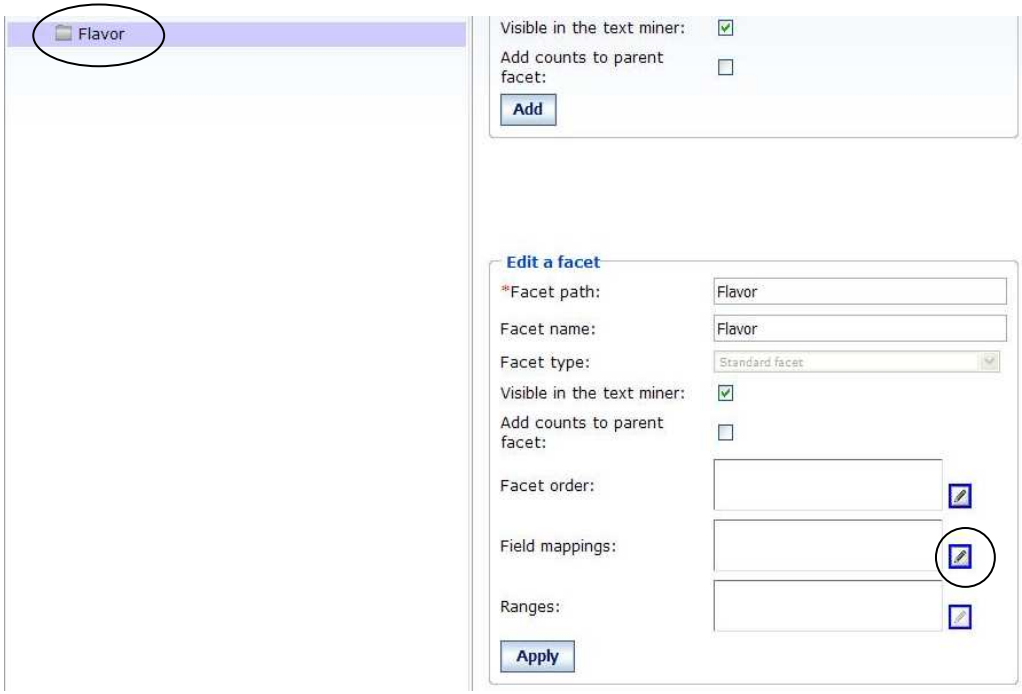


Create facets mapped to the search fields how to configure it?

- For the Facet path, type a path for the Facet. In this example, type Flavor
- For the Facet name, type the name that you want displayed in the text miner for the facet. In this example, type Flavor
- Click **Add**

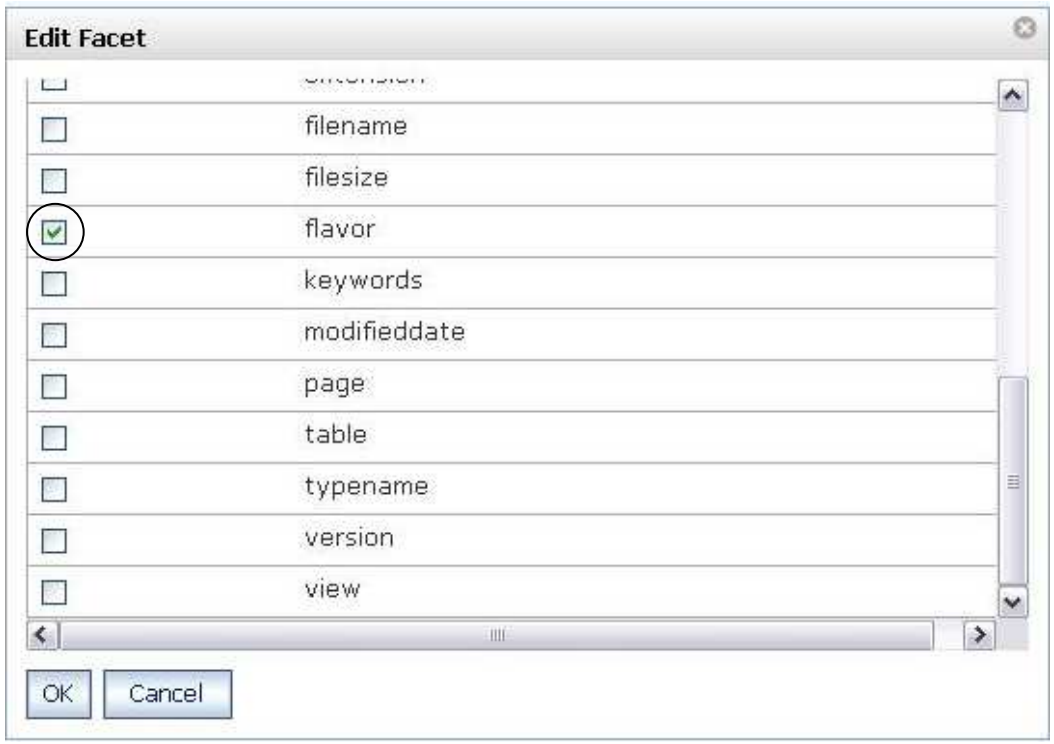
Create facets mapped to the search fields how to configure it?

- Select the new facet in the Facet tree. For this example, select Flavor
- Click the **Edit** icon next to the Field mappings text box under the Edit a facet section. This allows you to map the flavor search field to the Flavor facet



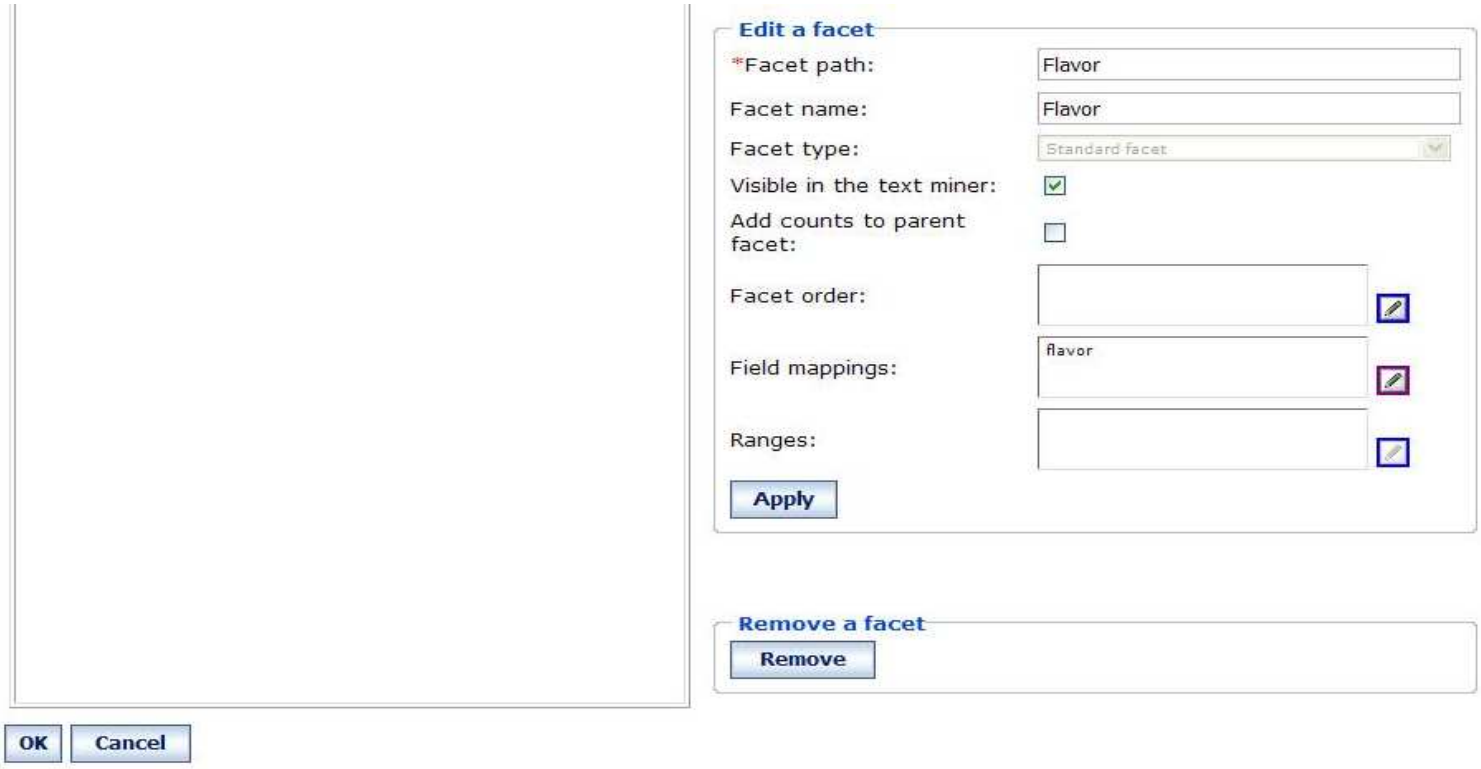
Create facets mapped to the search fields how to configure it?

- Select the search field check box that you want associated with the facet. For this example, select the check box associated with the flavor search field.
- Click **OK**



Create facets mapped to the search fields how to configure it?

- Click **Apply**
- Click **OK**

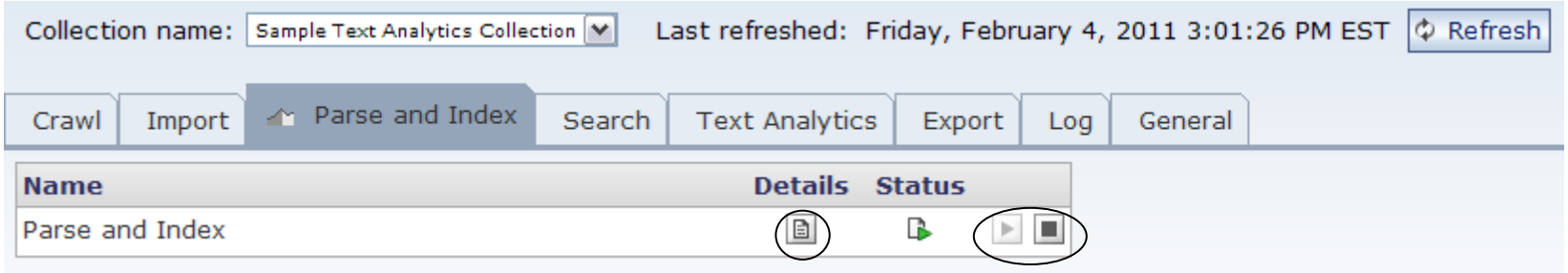


Module roadmap

- Export an annotator as a PEAR file for ICA
  - What is it?
  - How to configure it?
- How to deploy a PEAR file in ICA?
- How to associate a text engine with a collection?
- How to import search fields associated to the PEAR file?
- How to create facets and map them to search fields?
- **How to redeploy resources and rebuild index?**
- Summary and best practices
- Sample exercises

Redeploy resources and rebuild the index in ICA

- On the Parse and Index screen, click the **Monitor** icon
- Click the **Stop** icon
- Once the server is stopped, click the **Start** icon
- Once the server is started, click the **Details** icon

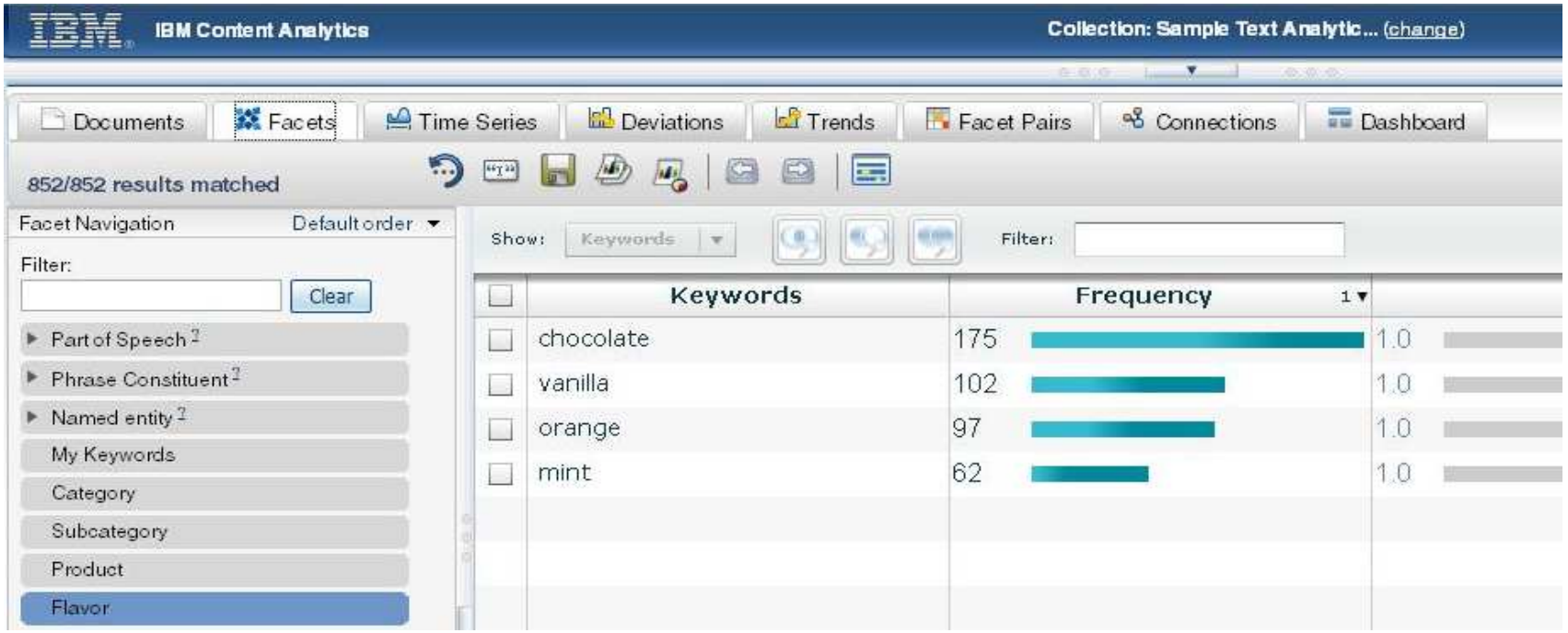


Redeploy resources and rebuild the index in ICA

- Redeploy the resources by clicking the **Start** icon under the Resource deployment status section
- Rebuild a full index by clicking the **Restart a full index build** icon under the Rebuild index status section
- Click **OK** to the confirmation message

View annotation results in the text miner application

- Open the ICA text miner
- Click **Facets**
- Select your facet associated with the annotator. For this example, select **Flavor** in the Facet Navigation



Module roadmap

- Export an annotator as a PEAR file for ICA
  - What is it?
  - How to configure it?
- How to deploy a PEAR file in ICA?
- How to associate a text engine with a collection?
- How to import search fields associated to the PEAR file?
- How to create facets and map them to search fields?
- How to redeploy resources and rebuild index?
- **Module summary and best practices**
- Sample exercises

Module summary

You have completed this module and can:

- Export an annotator as a UIMA PEAR file.
- Deploy the UIMA PEAR file in ICA
- Import search fields and create facets mapped to the annotator
- Work with analytics based on the annotator.

See the LanguageWare help for more tips and advanced use cases.

Best practices

- If you need to modify your annotator in LRW, you need to deploy the new version of the annotator in ICA as a new text engine and associate it to the collection. The prior deployed text engine will still remain in ICA. It is advisable to delete the prior deployed text engine if you will no longer use it in ICA. To delete a text engine:
  - Click System -> Edit icon -> Parse -> Configure text analysis engines
  - Select the delete icon associated with the text analysis engine that is no longer being used.
- When importing the search fields, you can select the facet check box associated with the search field to automatically create the facet. This will result in a facet added under the root facet with the same name as the search field. To create specific facet names or hierarchies, do not select the facet check box when importing the search field. Instead, use the Configure facet option on the Parse and Index screen.

Module roadmap

- Export an annotator as a PEAR file for ICA
  - What is it?
  - How to configure it?
- How to deploy a PEAR file in ICA?
- How to associate a text engine with a collection?
- How to import search fields associated to the PEAR file?
- How to create facets and map them to search fields?
- How to redeploy resources and rebuild index?
- Module summary and best practices
- **Sample exercises**

Practice exercises

- Create an annotator based on the examples in the prior modules.
  - Create project named: ConfectionaryCompanyHelpline. Set the UIMA type prefix to com.ibm.lw.exercise.
  - Create a dictionary named: Flavors. It should contain the following noun terms:
    - Chocolate, vanilla, mint, orange
  - Create a sample document and annotate the document.
  - Create a parsing rule named: IdentifyQuestion.
  - Create a UIMA annotator named AnalyseHelpline. Add the Flavor dictionary and IdentifyQuestion parsing rule to the pipeline.
- Export the AnalyseHelpline annotator as a UIMA PEAR file for ICA.
- Configure ICA with the sample collection.
- Deploy the PEAR file in ICA as a text engine.
- Associate the text engine with the Sample Text Analytics Collection in ICA.
- Import the search fields in ICA.
- Create a facet for each search field and map it to the search field in ICA.
- Redeploy resources and rebuild the full index in ICA.
- View and work with the analytics in the text miner application.

Contacts

- If you have any questions, comments or suggestions, contact us using the LanguageWare email address *EMEALAN@ie.ibm.com* or on the developerWorks® forum

ICA and LRW resources

- IBM Content Analytics Site:
  - http://www-01.ibm.com/software/data/content-management/analytics/
- LRW, alphaWorks
  - http://www.alphaworks.ibm.com/tech/lrw
- Content Analytics MicroSite:
  - www.ibm/ecm/content-analytics
- Information Center:
  - http://publib.boulder.ibm.com/infocenter/analytic/v2r2m0/index.jsp
- Redbook:
  - http://www.redbooks.ibm.com/abstracts/sg247877.html?Open
- Medical Records Text Analytics
  - http://www.youtube.com/watch?v=_Ku1rWU_Jxs
- Jstart team
  - http://www-01.ibm.com/software/ebusiness/jstart/textanalytics/index.html
- Text Analytics Group (Lab Services) Team
  - James Mobley (jmobley@us.ibm.com)

Trademarks, copyrights, and disclaimers

IBM, the IBM logo, ibm.com, developerWorks, and LanguageWare 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 other IBM trademarks is available on the web at "Copyright and trademark information" at http://www.ibm.com/legal/copytrade.shtml

Other company, product, or service names may be trademarks or service marks of others.

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS OR SOFTWARE.