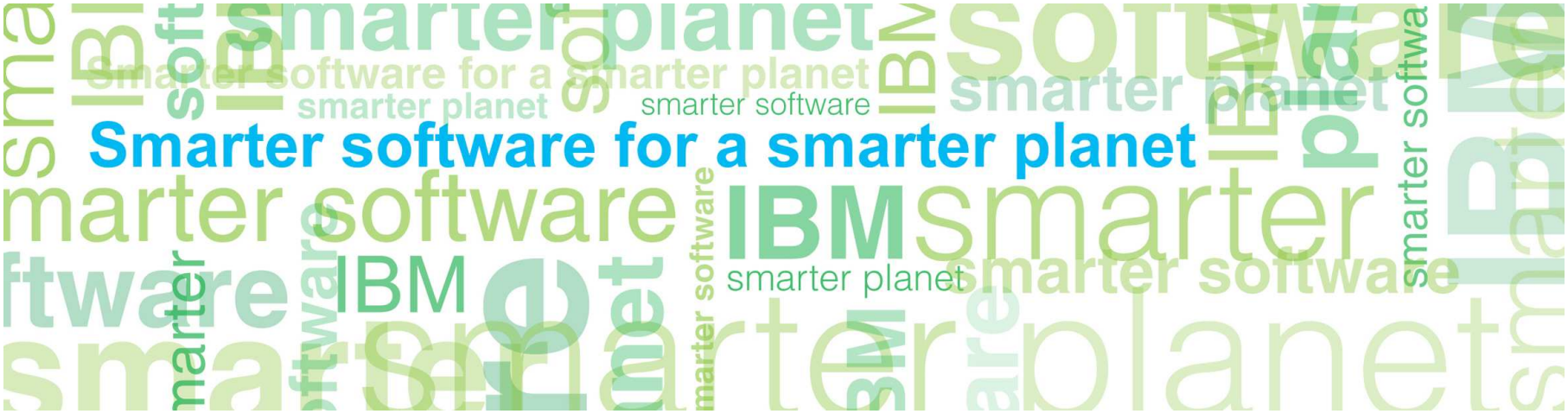
LanguageWare Resource Workbench 7.2

Export PEAR to ICA



Introduction

- **Module overview**
  - How to use LanguageWare® Resource Workbench (LRW) 7.2 to export a PEAR file directly to IBM Content Analytics 2.2
  - Best practices
- **Target audience**:
  - Technical Sales professionals with a need to customize the text analytics in ICA for a customer POC, demo, or engagement
- **Prerequisites**:
  - Students should have completed the prior modules in this training to set up a connection to ICA and create an annotation pipeline (.annoconfig) in LRW
  - The prior models had an ongoing training exercise that should have been completed in order to fully exercise the knowledge in this module
  - Student must have created the sample text analytics collection in ICA
- **Version release date:** LRW 7.2, ICA 2.2, released October, 2010

Module objectives

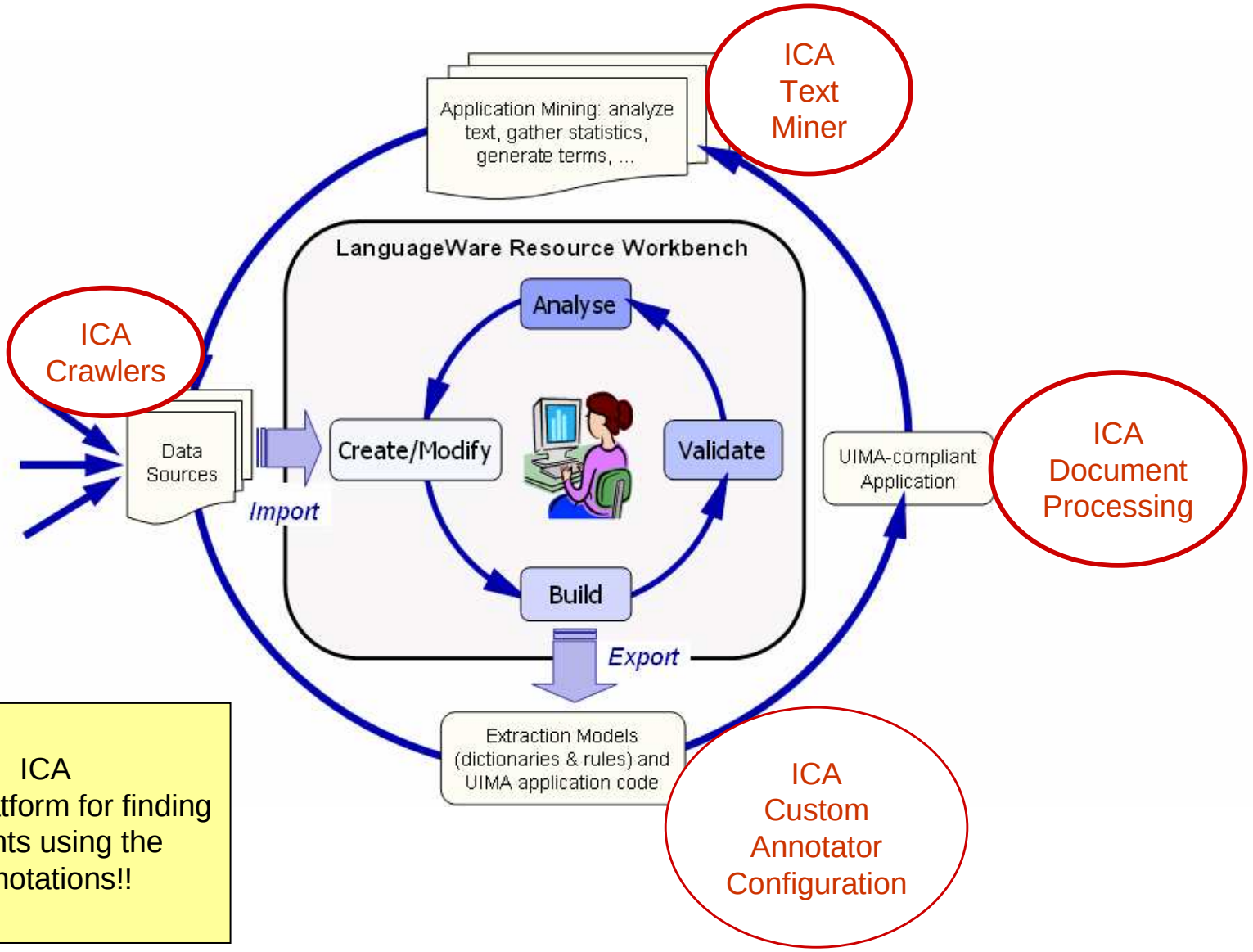
After this module you will be able to:

- Export a custom annotation pipeline as a PEAR file directly to ICA
- See your annotation results in the ICA text miner interface

Module roadmap

- **Exporting a custom annotation pipeline to ICA from LRW**
  - Background for the custom annotator slot in ICA
  - How to use the ICA connection previously created
  - How to map the annotations (types) created by the custom annotation pipeline to displayable facets in ICA
- **Summary and best practices**
- **Sample exercises**

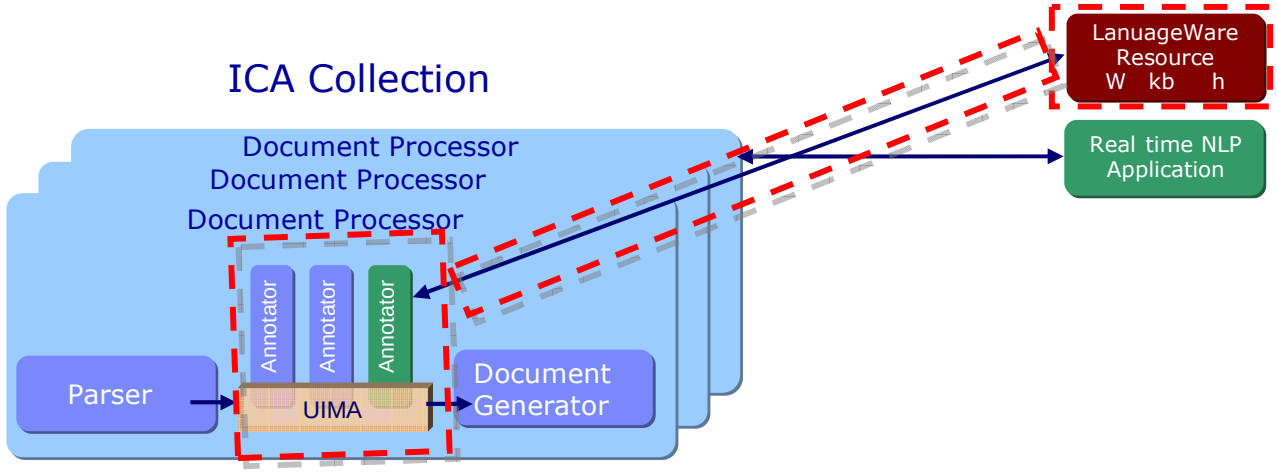
LRW develops UIMA annotators, isn't that enough?



ICA
is the platform for finding insights using the annotations!!

Exporting a custom text annotator to ICA from LRW
The What and the Why

- Once you have created a pipeline (.annoconfig) in LRW, you now need to set it up for document ingestion and application access.
- ICA has a custom slot in each collection which was designed for adding custom annotators
- As of ICA 2.2, you can deploy what you build in LRW directly to ICA

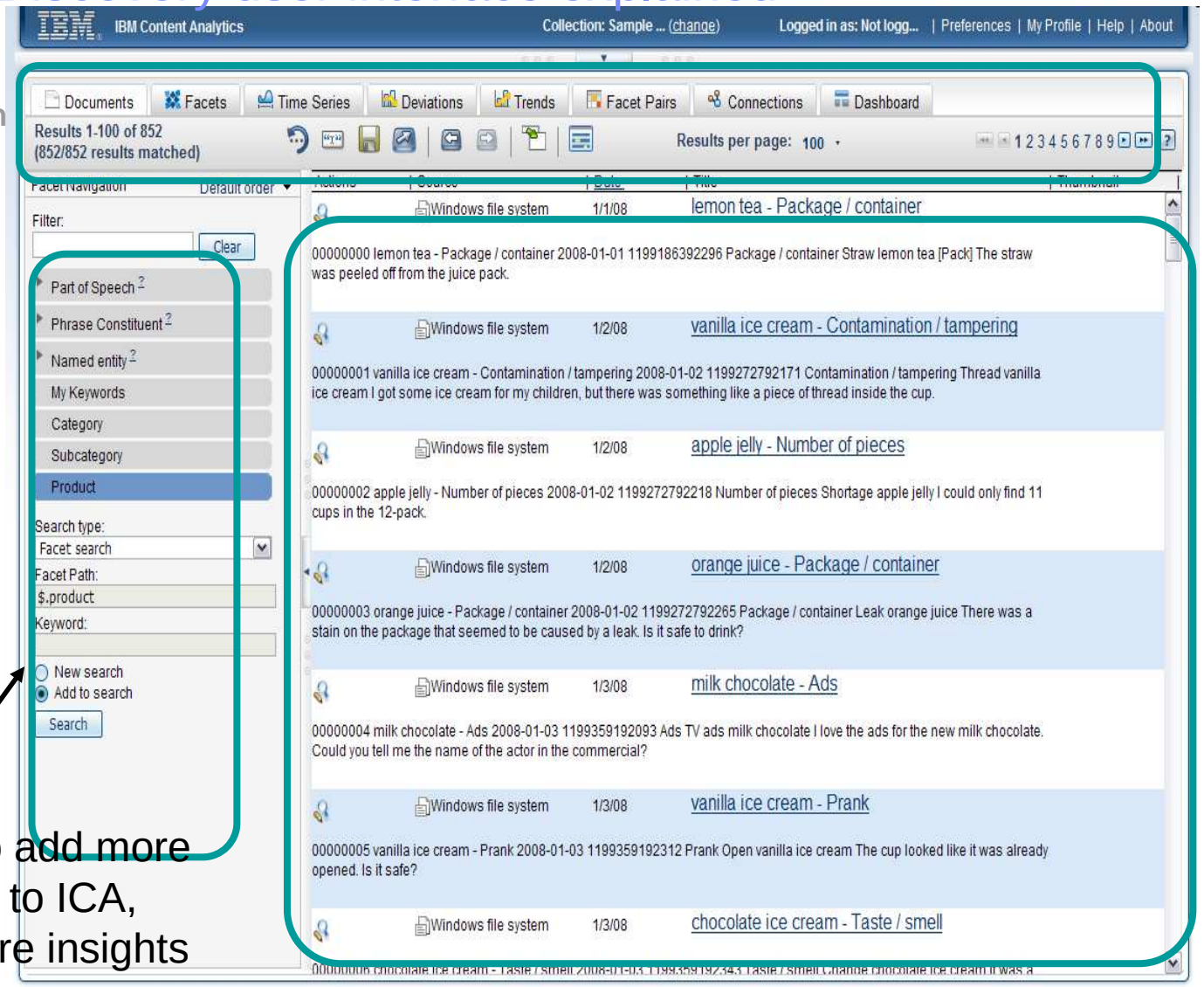


The Interactive Discovery user interface explained

Query-based Exploration

Views, Filters and Thresholds

Automatically Extracted and Analyzed Concepts, Entities, Relationships, Meta Data and Classifications

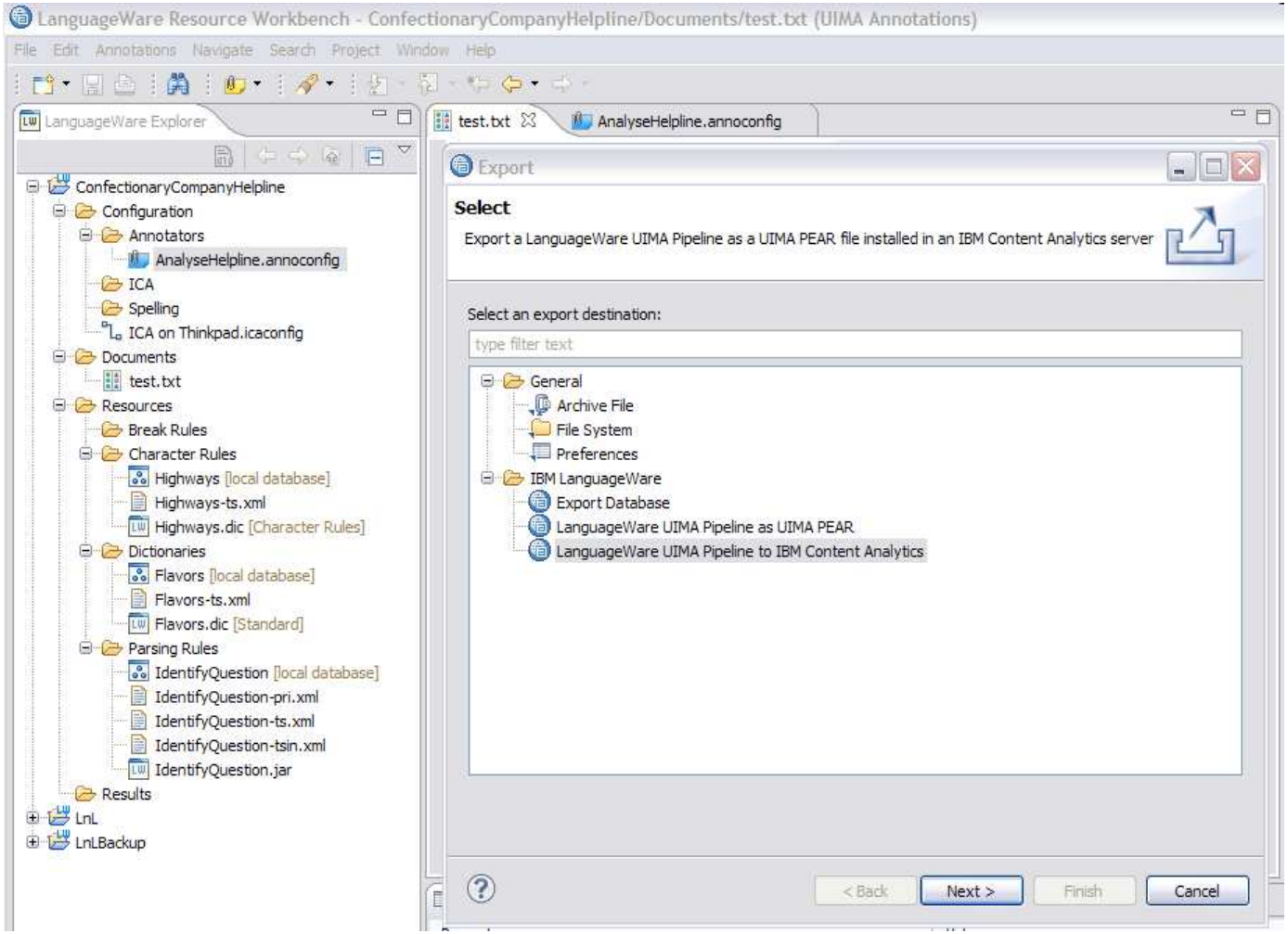


LRW can be used to add more facets for navigation to ICA, leading to finding more insights

Visualization with Drill Down for Exploration and Assessment

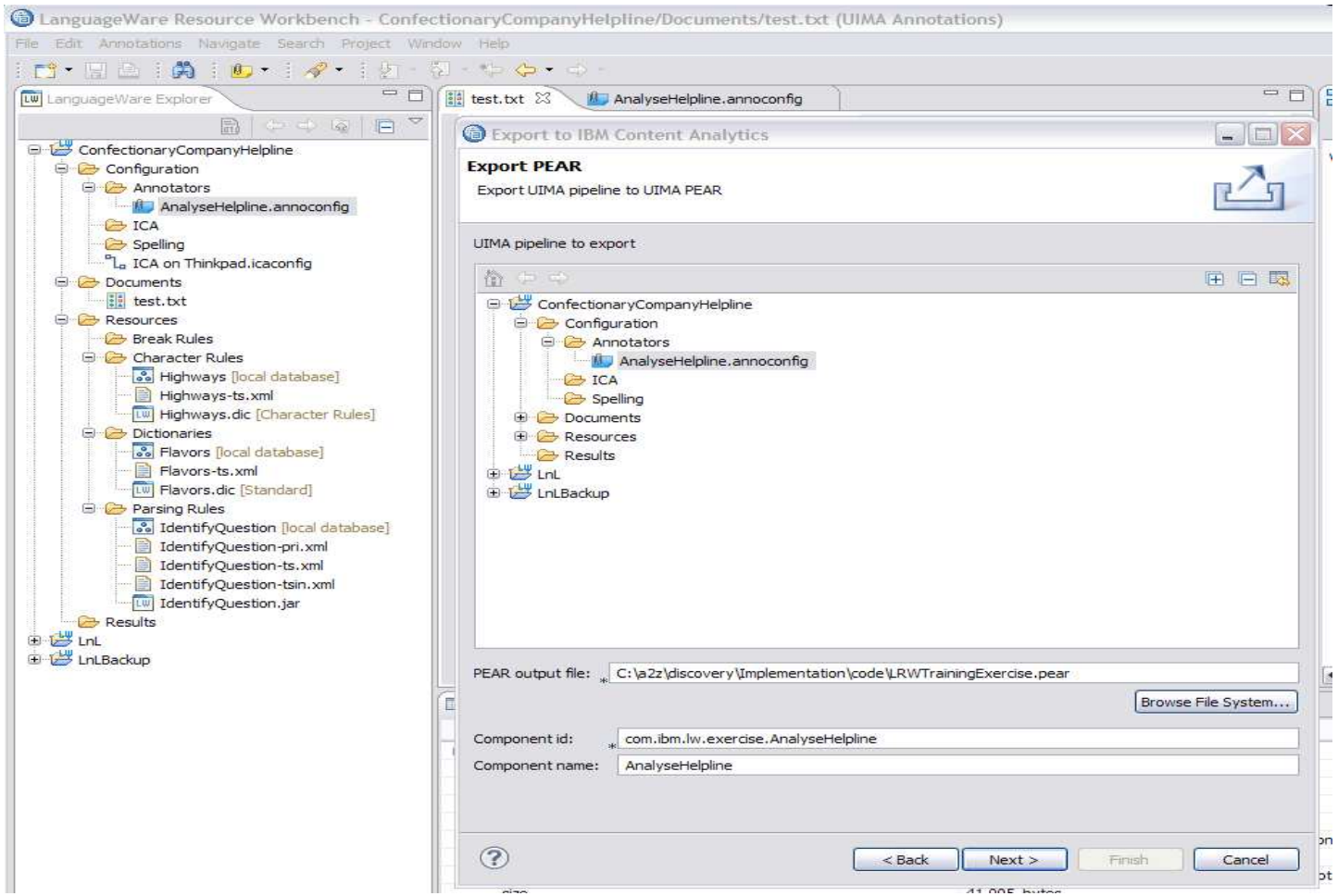
Exporting a custom text annotator to ICA from LRW

- Right-click your annotation pipeline and choose Export, then choose “LanguageWare UIMA Pipeline to IBM Content Analytics” and click “Next”



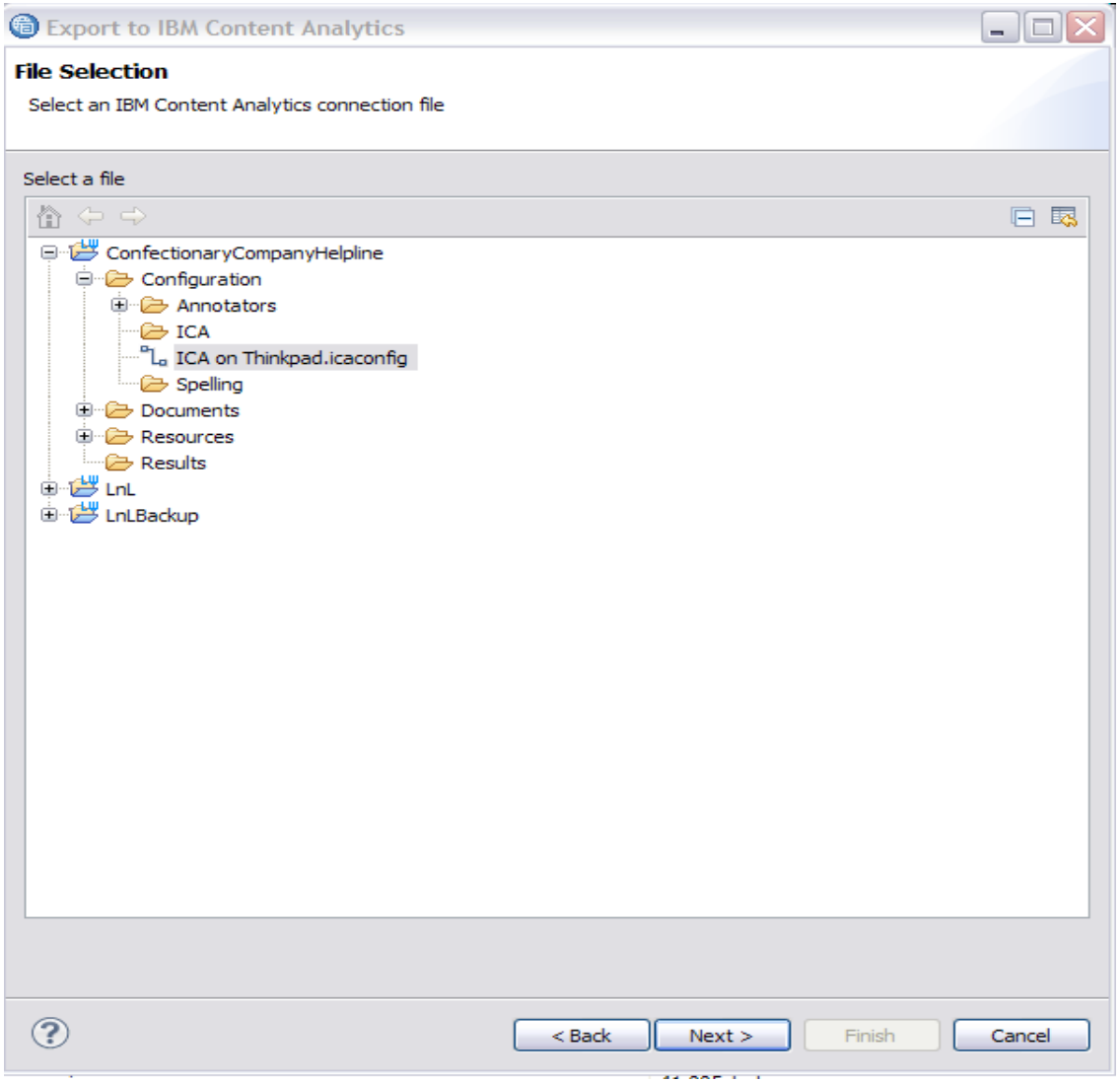
Exporting a custom text annotator to ICA from LRW

- Fill in values for PEAR output file (use Browse File System to choose a directory, or else it will go into the directory where LRW was installed)
- Fill in a value for Component name (same as the annoconfig file name is fine)



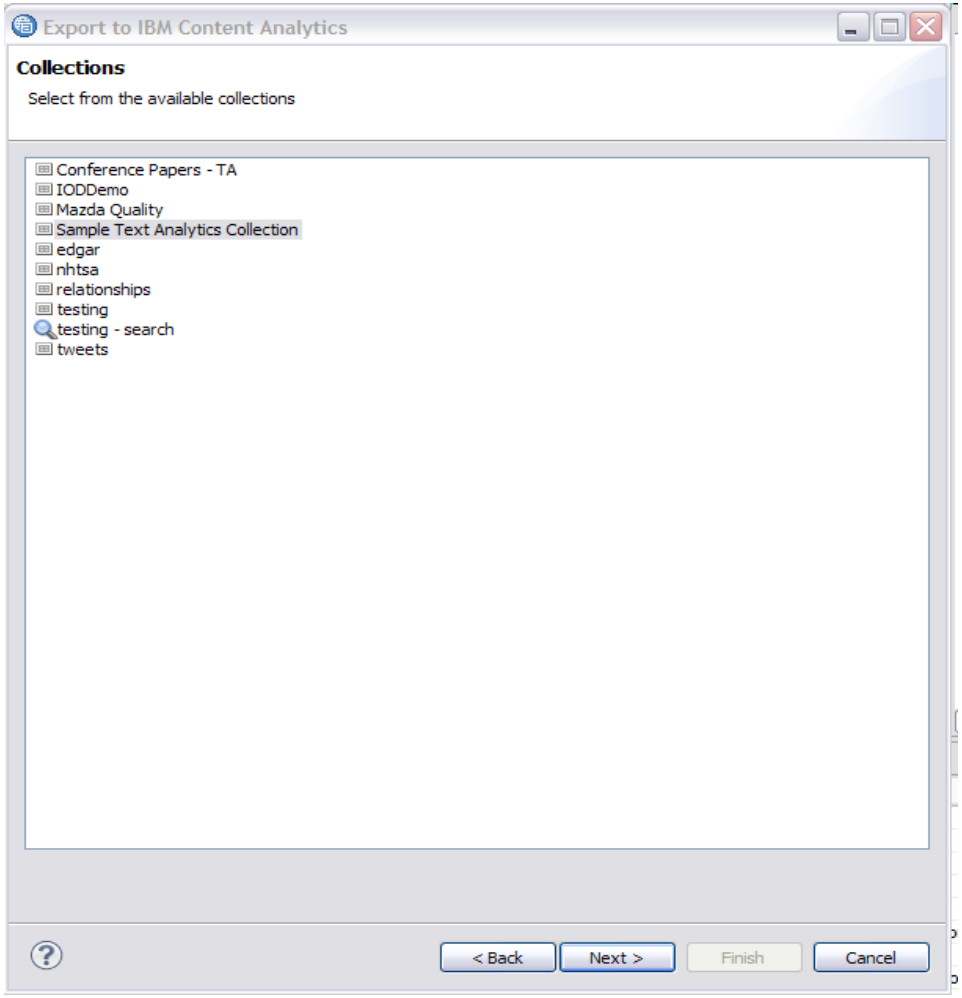
Exporting a custom text annotator to ICA from LRW

- Select the ICA connection to your ICA system (in the Configuration folder)



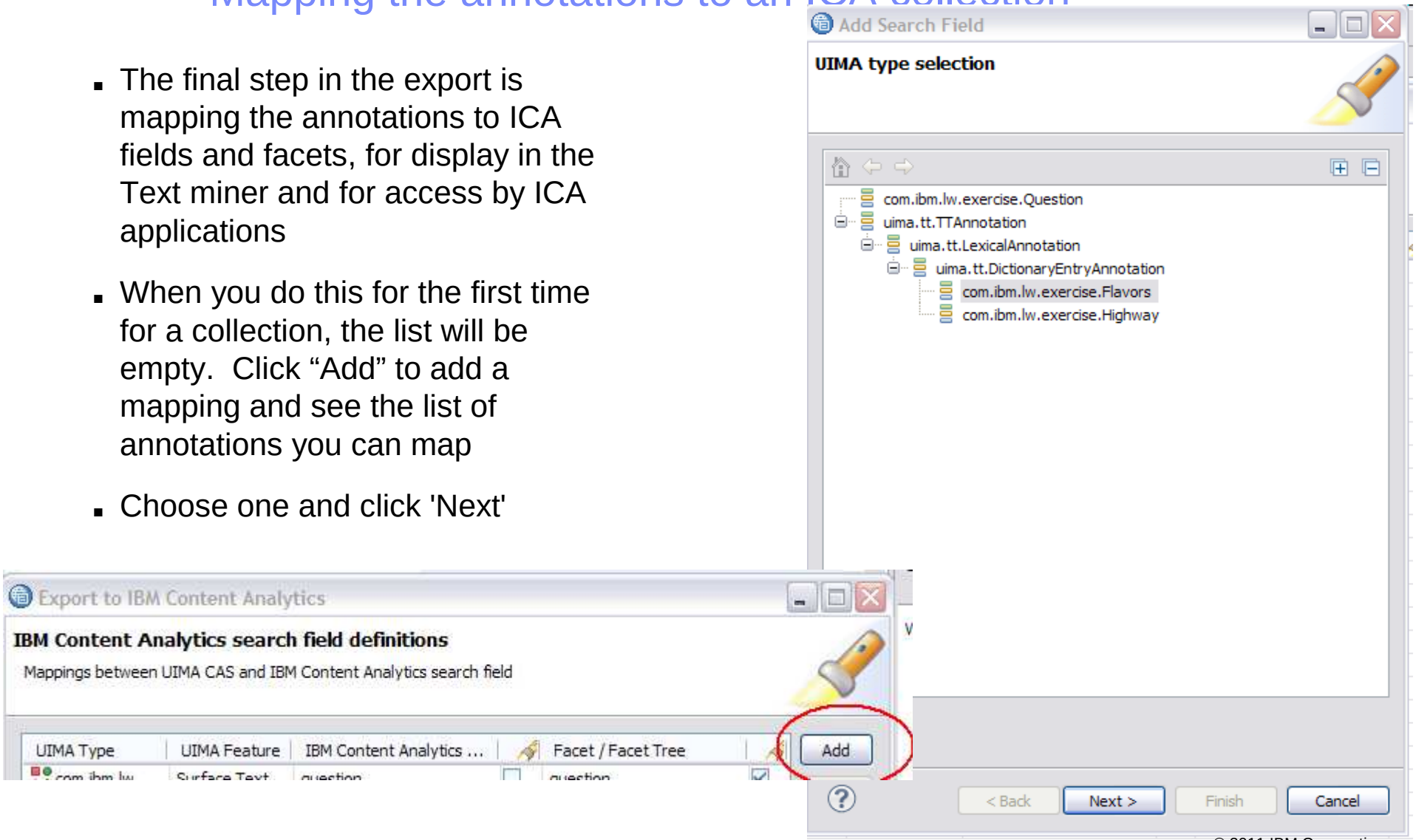
Exporting a custom text annotator to ICA from LRW
Using the ICA Connection

- LRW will present you with a list of collections – chose the one you want
  - Tip: If you do not see a collection you created after you created the connection, open the connection configuration, click 'Refresh', and then CNTL-S to save it



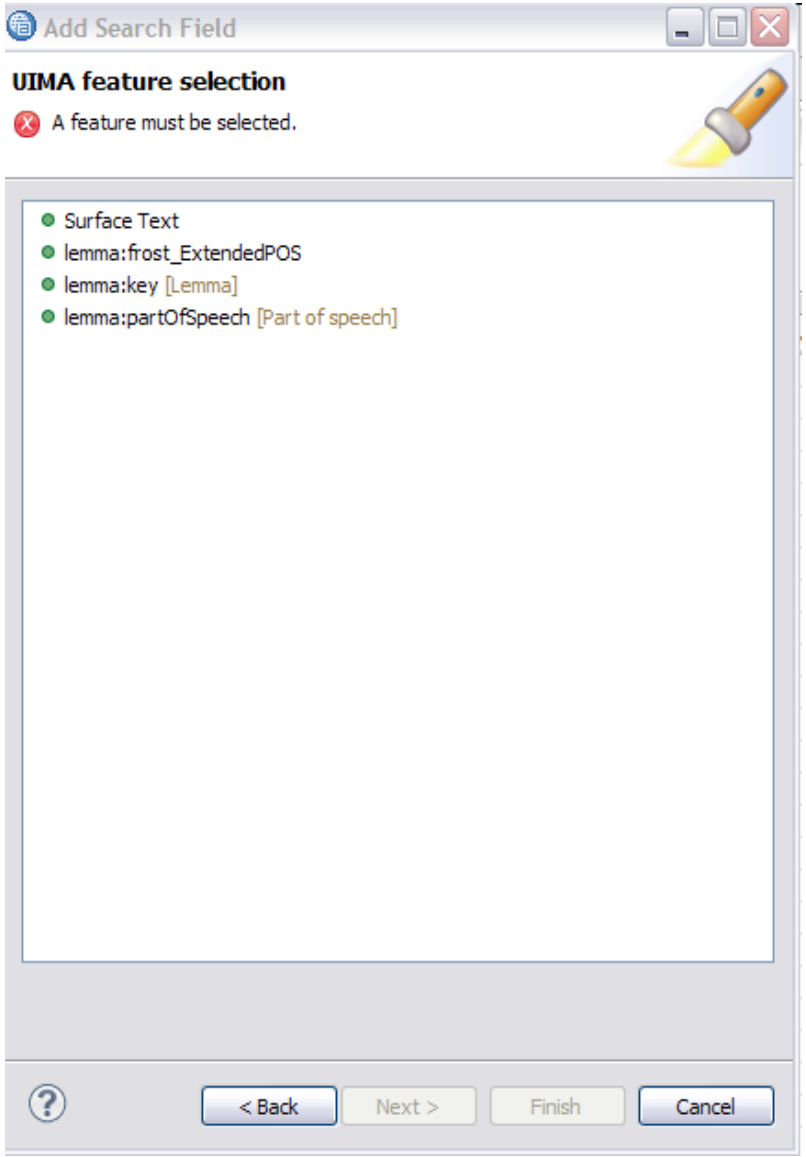
Exporting a custom text annotator to ICA from LRW
Mapping the annotations to an ICA collection

- The final step in the export is mapping the annotations to ICA fields and facets, for display in the Text miner and for access by ICA applications
- When you do this for the first time for a collection, the list will be empty. Click “Add” to add a mapping and see the list of annotations you can map
- Choose one and click 'Next'



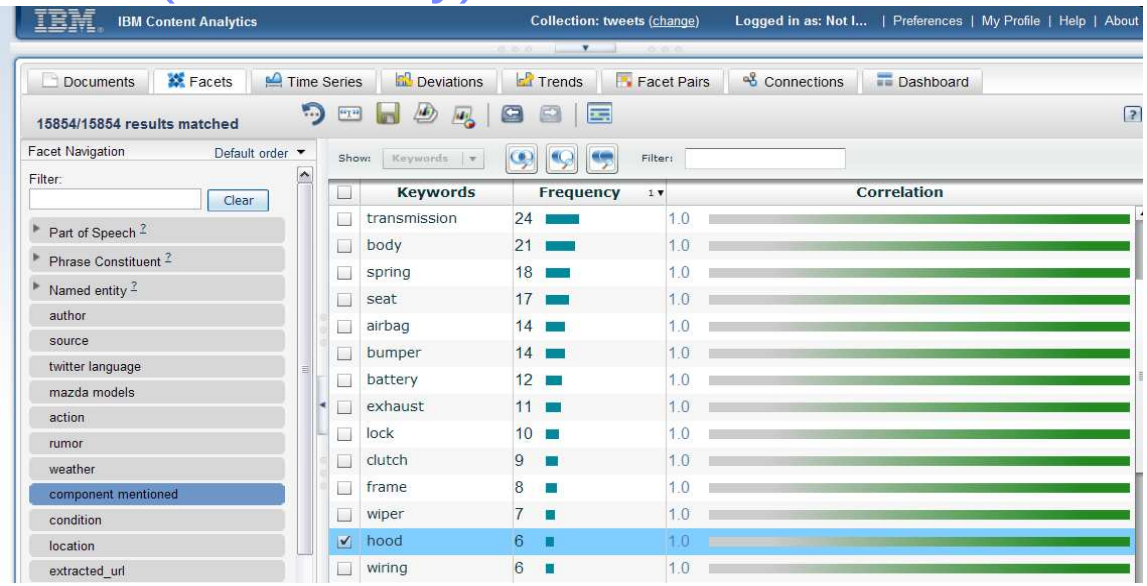
Exporting a custom text annotator to ICA from LRW
Mapping the annotations to an ICA collection

- You will be presented a choice of what value you want for the annotation
- Usually you want to choose lemma:key, which is the normal form
- In some cases, Surface Text may be what you want
- Some annotations will have more choices than others
- This choice affects what shows up in the facets view of ICA, the original surface text will always remain in the documents view of ICA, and will be highlighted when a facet is selected.

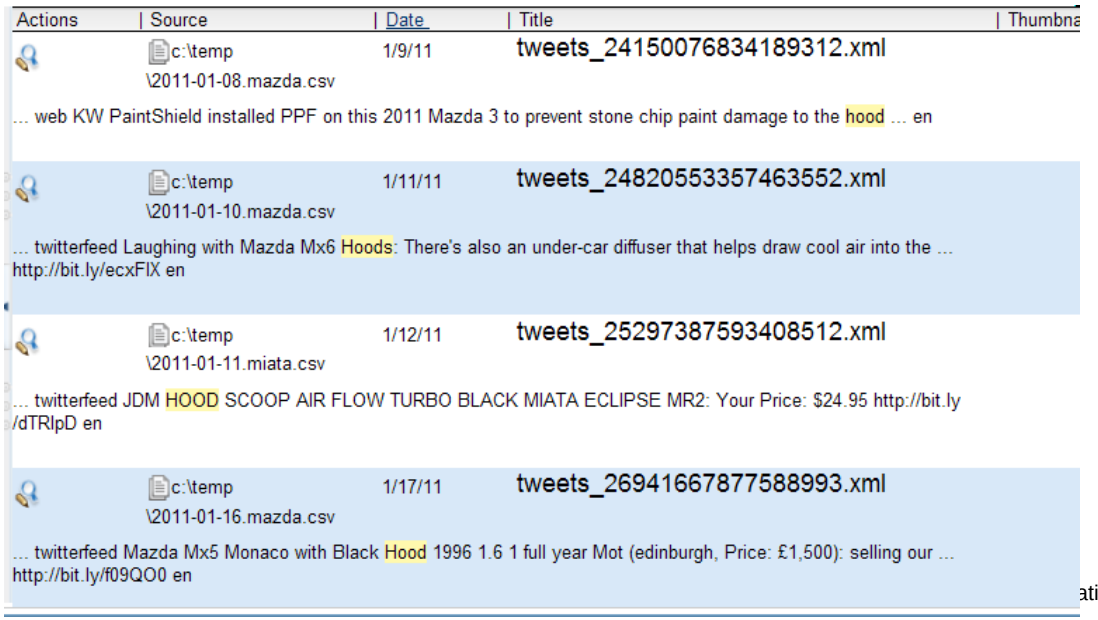


Example of using normal form (lemma:key) for facet value

- The value 'hood' is chosen in the facet view

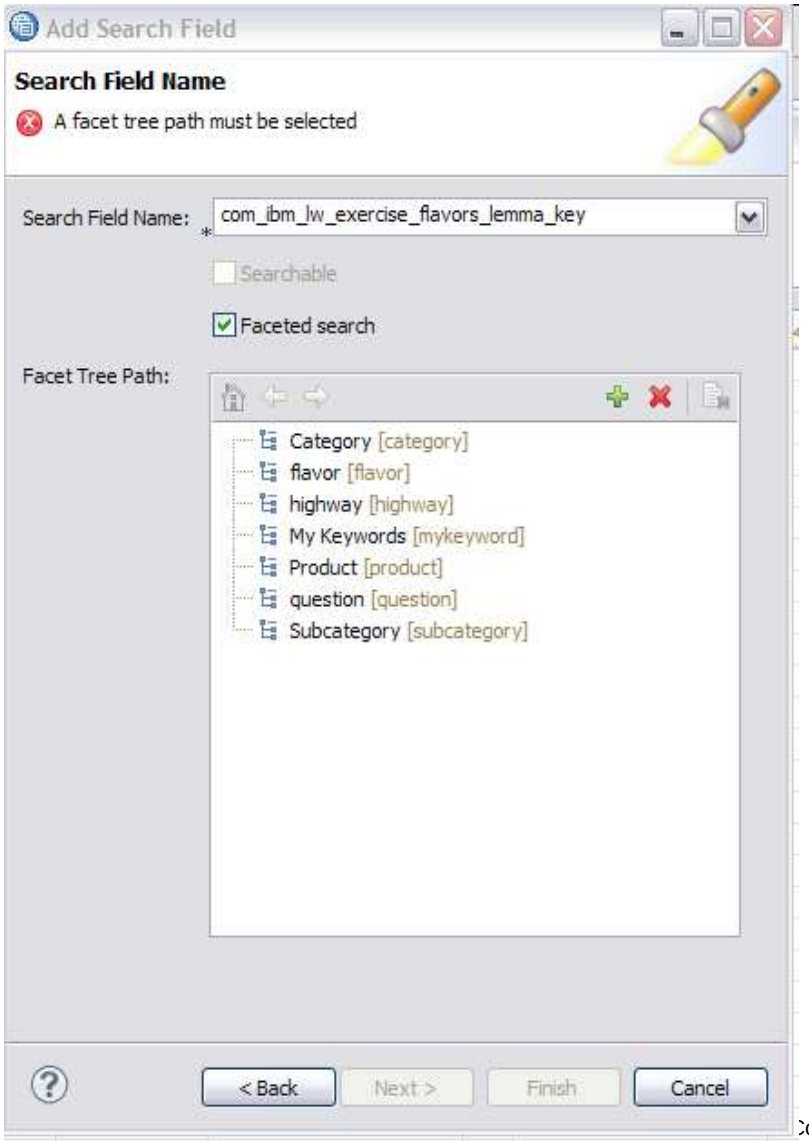


- Variations hood, Hoods, HOOD, Hood all show up highlighted in the documents view



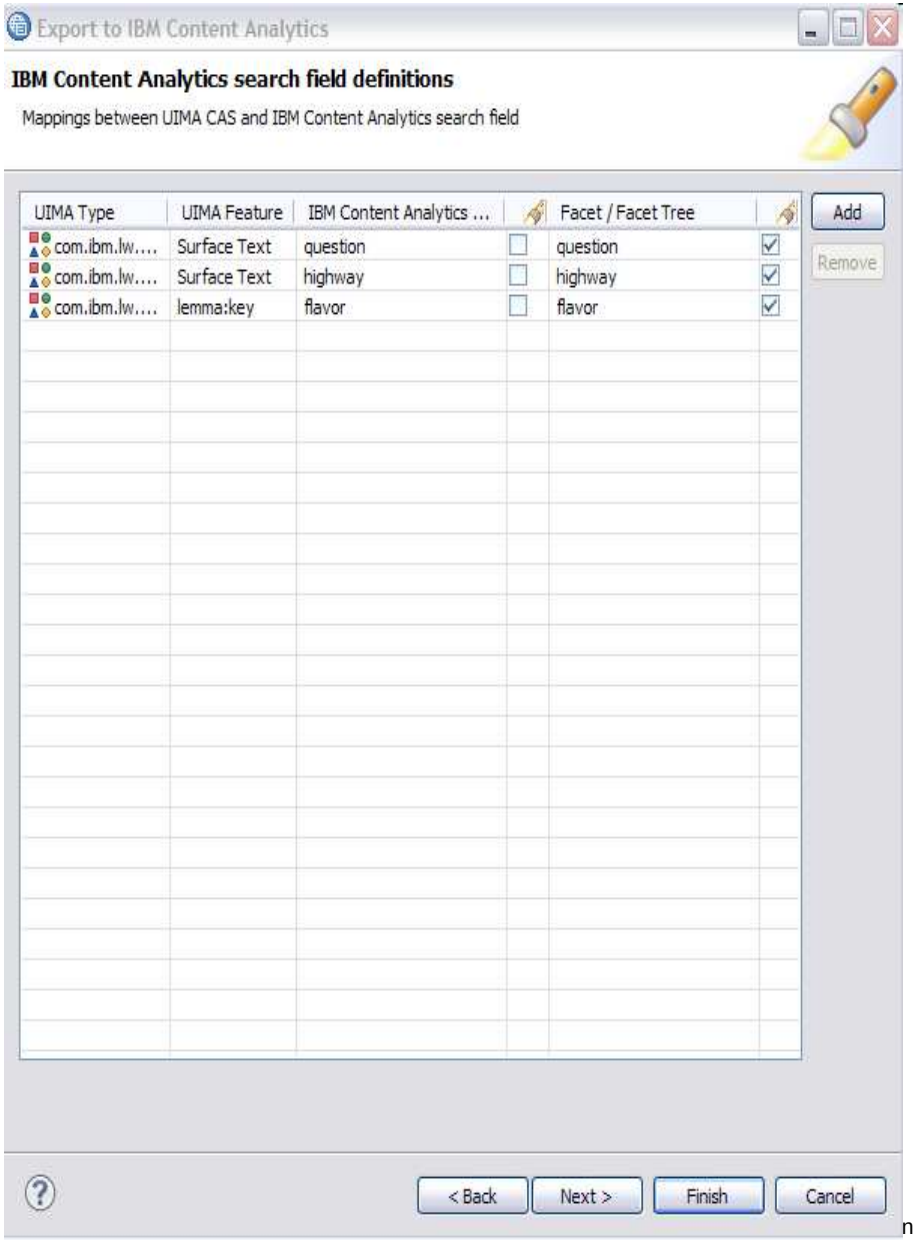
Specifying fields and facets

- The default field name will be the whole long UIMA type name
  - You probably want to change this to the short name, 'flavors' in this case
- Check the “Faceted search” box.
- Click the green + to add a facet
- Specify facet path and facet name
- Click “finish”
- Repeat these steps for all of the new annotations that you want to map to ICA fields/facets



Completing the deployment

- When the field definitions are complete, you will have a table of UIMA type to field and facet
- Click “Finish”
- Optionally choose whether to start a re-index of the collection
  - No reason not to unless you have something else you want to do first on the server?
- PEAR will install on the ICA server!



Module roadmap

- **Exporting a custom annotation pipeline to ICA from LRW**
  - Background for the custom annotator slot in ICA
  - How to use the ICA connection previously created
  - How to map the annotations (types) created by the custom annotation pipeline to displayable facets in ICA
- **Summary and best practices**
- **Sample exercises**

Module summary

You have completed this module and can:

- Export a custom annotation pipeline as a PEAR file directly to ICA
- See your annotation results in the ICA text miner interface

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

Best practices

- You may need to refresh your text miner a couple of times to see your results after re-indexing finishes.
- If you need to make a change to your annotator after you see the results, you must manually remove it from ICA first
  1) Disassociate the PEAR from the collection, using the edit view of the Parse and Index tab and choosing “No custom analysis” in “Configure text processing options”
  2)Remove the PEAR from the system, using the edit view of the Parse tab
  3)Follow the steps here again. LRW will remember your choices so it's easier the next time.

Module roadmap

- **Exporting a custom annotation pipeline to ICA from LRW**
  - Background for the custom annotator slot in ICA
  - How to use the ICA connection previously created
  - How to map the annotations (types) created by the custom annotation pipeline to displayable facets in ICA
- **Summary and best practices**
- **Sample exercises**

Practice exercises

- This module used the sample exercise from all of the prior modules. For your convenience, those steps are repeated here.
- **Create project**
  - Create a new Project in the Workspace called "ConfectionaryCompanyHelpline". The default UIMA Type prefix should be "com.ibm.lw.exercise"
- **Create a text file**
  - Create a file called "test.txt", containing text from the ICA sample text analytics collection which includes questions and a Highway name
- **Create dictionary**
  - Create a dictionary of Flavors.
  - Add a few words to the dictionary (chocolate, vanilla, mint, orange).
  - Import a CSV file containing more flavors
- **Create a rules DB**
  - Create a Rules DB called "IdentifyQuestion"
- **Create a UIMA Annotator**
  - Create an annotator called "AnalyseHelpline"
  - Add "Flavor" Dictionary and "IdentifyQuestion" Parsing Rules to the pipeline
- **Annotate a document**
  - Annotate the "test.txt" text file with the "AnalyseHelpline" annotator
- **Create rules**
  - Create some rules in the "IdentifyQuestion" database to identify the questions in the "SampleQuestion" document
- **Character rules**
  - Create a Character Rule database called Highways
  - Add the Character Rules file to the "AnalyseHelpline" annotator
  - Write a character rule to identify Interstate highways - for example "I-123"
- **Create an ICA connection file**
  - Create an ICA Connection file to an ICA server which has the "Sample Text Analytics Collection"

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.