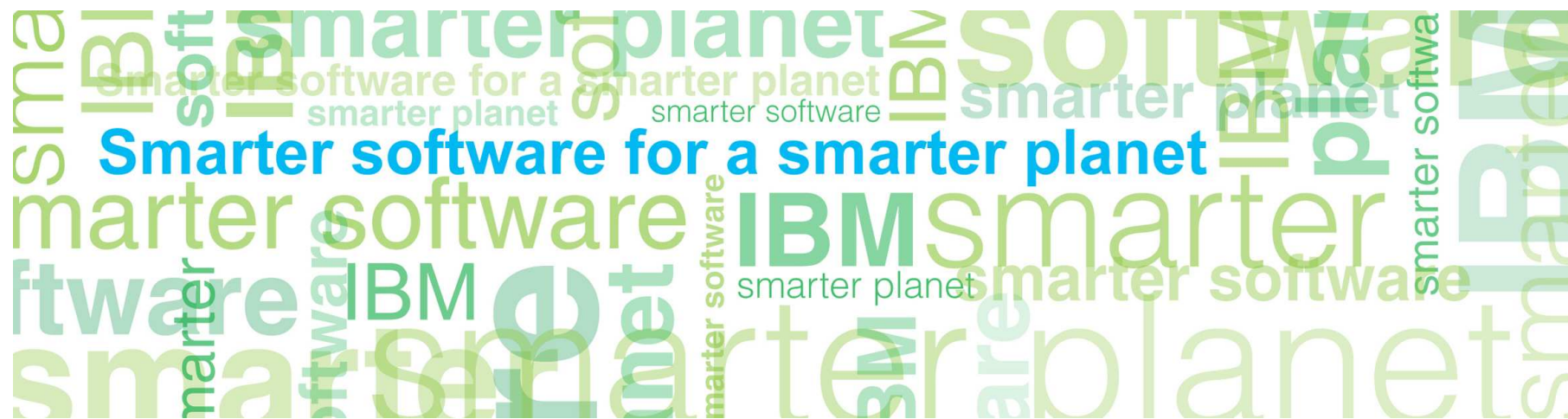


---

# LanguageWare Resource Workbench 7.2

## Annotate a document



© Copyright International Business Machines Corporation 2011. All Rights Reserved.  
US Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule  
Contract with IBM Corp.

## Introduction

### ■ **Module Overview**

- Add a document to the LRW workspace.
- Run an annotator on a document.
- See occurrences in the Outline.
- Use the clean up stage to hide annotations.
- Best practices

### ■ **Target Audience:**

- All audiences

### ■ **Prerequisites:**

- A project created within the LRW.
- Custom dictionaries and Parsing Rules databases
- An annotator (or UIMA pipeline configuration file, that uses the required dictionaries and parsing rules databases...as per the *Create a UIMA pipeline configuration* training module)

### ■ **Version Release Date:** LRW 7.2, ICA 2.2, released October, 2010

## Module objectives

After this module you will be able to:

- Add or create a document and annotate it within the LRW.
- Understand the results displayed in the Outline tab.
- Investigate and modify the results obtained.
- Use the Clean up stage to hide annotation types.

## Module roadmap

- **Annotate a document**
  - Add or create a document
  - Run an annotator
  - See the occurrences in the Outline tab
  - Clean up stage
- **Summary and best practices**
- **Sample exercises**

## Introduction

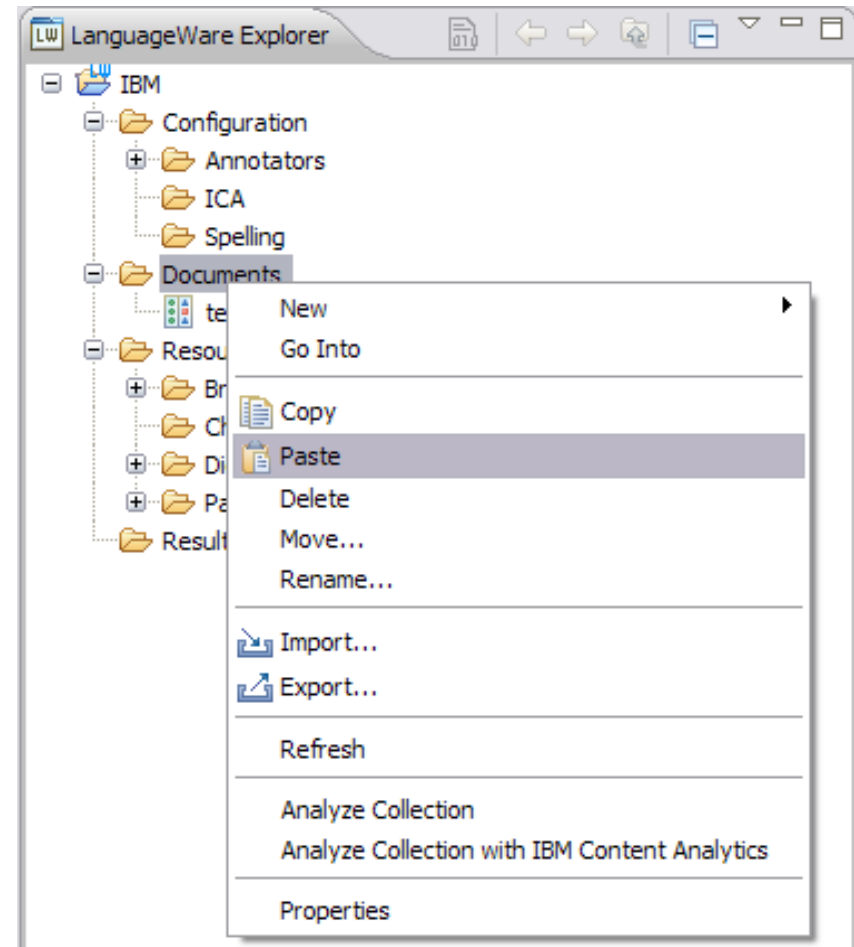
- Documents can be added or created within the LRW workspace.
- When annotating a document we are trying to identify the entities and relations between entities we have set your annotator to find in that document.
- In order to annotate a document we need an annotator with the relevant dictionaries and parsing rules defined to capture the parts of information we are interested in.
- When running that annotator on the document, the Outline tab will display the annotations found and we will be able to analyze the results and modify the dictionaries and parsing rules in order to improve those results.
- The output types can be customized in the Clean up stage of the annotator to only display the desired ones.

## Adding a document to the LRW workspace

- LRW supports text files with .txt extension.
- To add a document to the “**Documents**” directory in LanguageWare® project, you can:
  - **Copy and Paste** documents into the LRW workspace
  - **Drag and drop** documents into the LRW workspace
  - **Import** documents into the LRW workspace (**File/Import** from the menu)
  - **Create** new documents.
- You may need to **Refresh (F5)** the relevant folder (or right-click on it and select **Refresh** from the menu) for the updated contents to display.

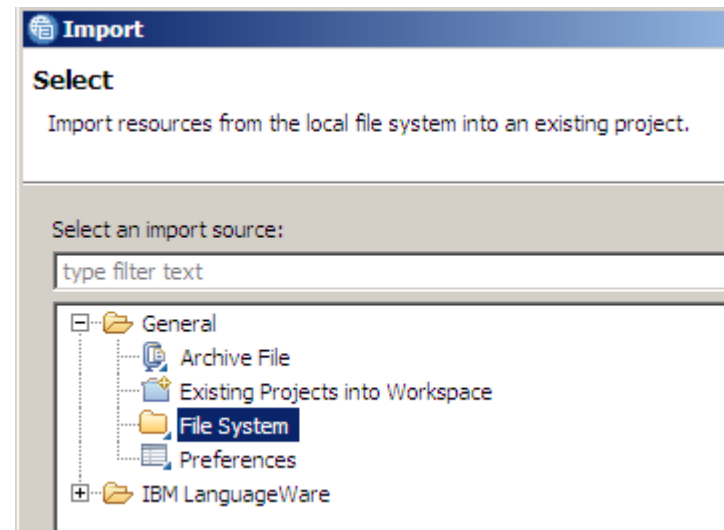
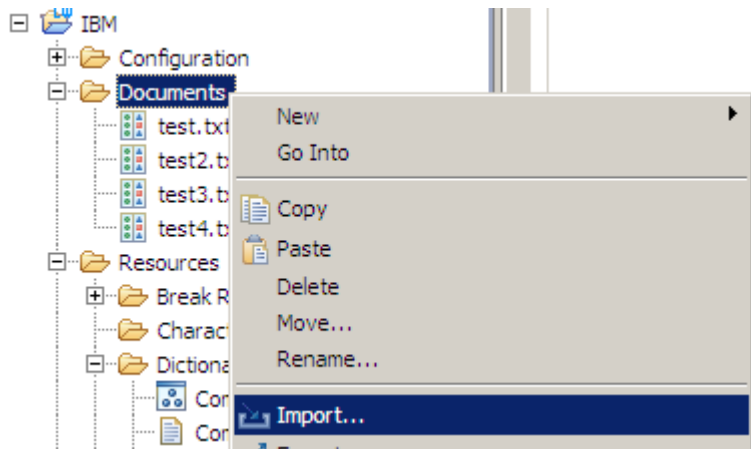
## Adding a document to the LRW workspace: Copy & Paste

- Every project created within the LRW includes a Documents folder for storing documents.
- Select the document from the file system and Copy it (Ctrl+C), select the Documents folder in the project on the LanguageWare Explorer, right-click and select Paste from the menu (or Ctrl+V).
- Text files can be dragged from a location in the file system and dropped into the desired folder in the project.



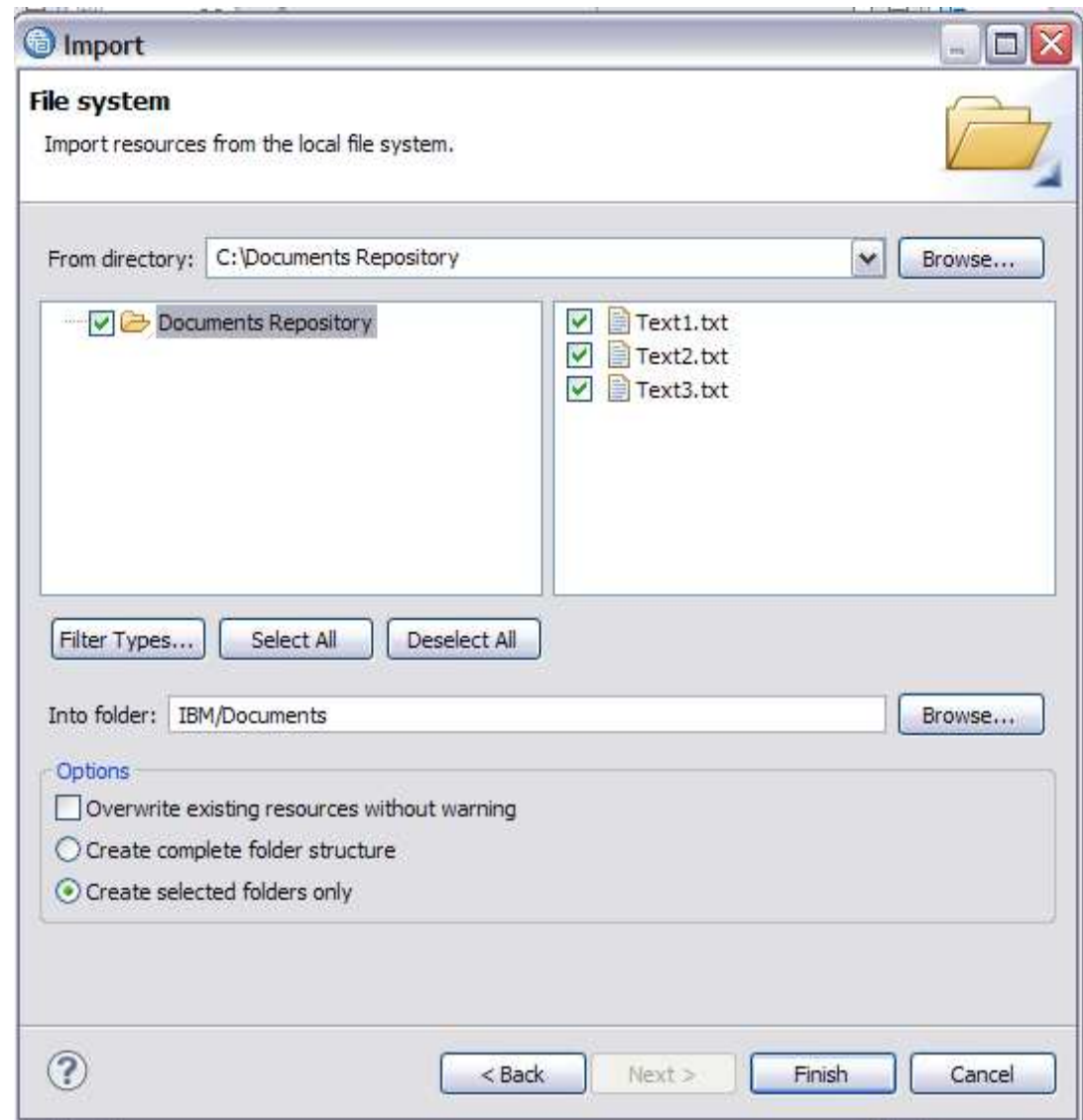
## Importing a document

- Select the destination folder (usually the Documents folder in the project), right-click and select Import.
- Select the source where documents are located.
- If the documents are stored somewhere in your machine, select File System.



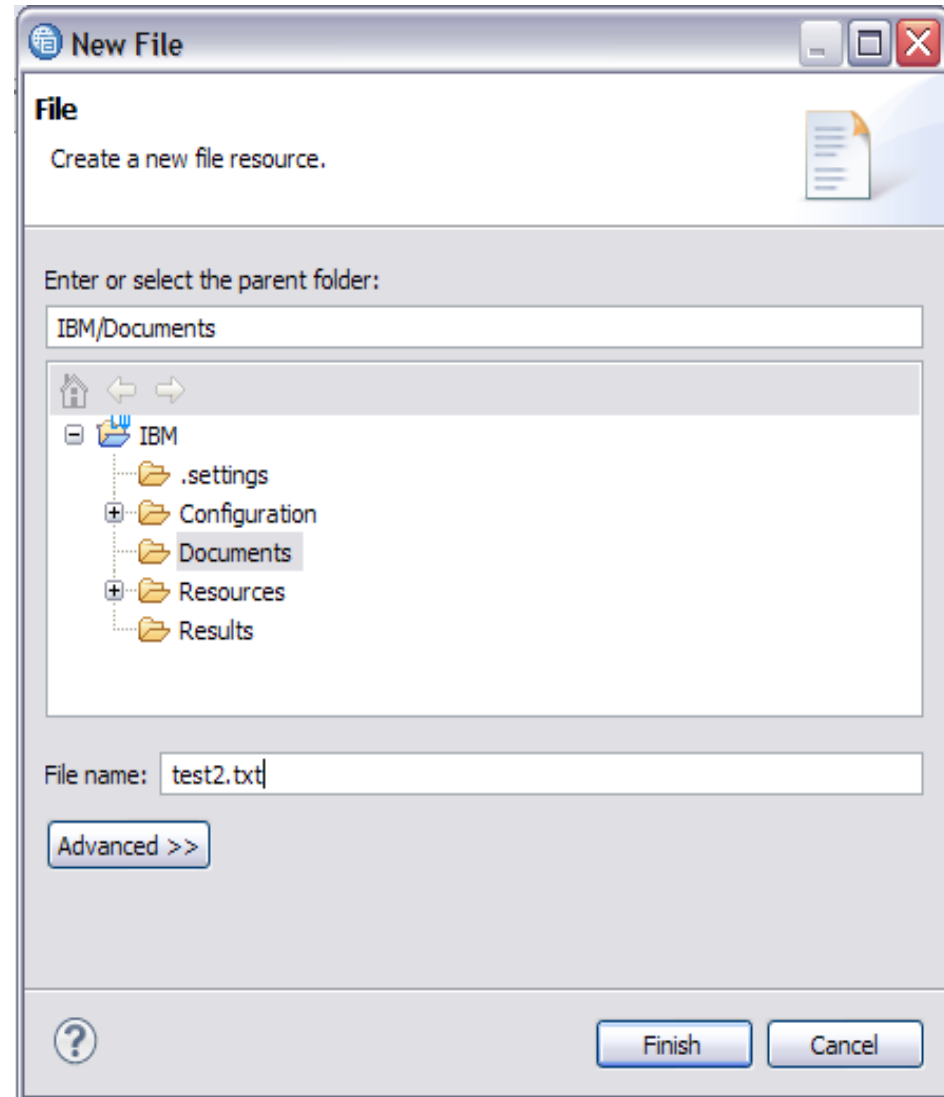


- Browse and select the specific documents or folders to be imported. The destination folder you had selected is automatically displayed, but it can be changed.
- Click on the Finish button and the selected files will be imported.

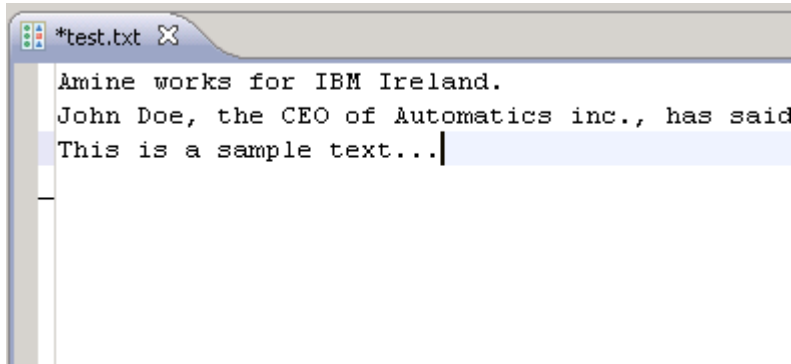


## Creating a new document

- From the LRW menu, select **File/New/File**. Alternatively, you can select the folder where you want to store the documents, right click, and select **New/File**
- Name the file and make sure you provide the extension .txt.
- Click on the Finish button.



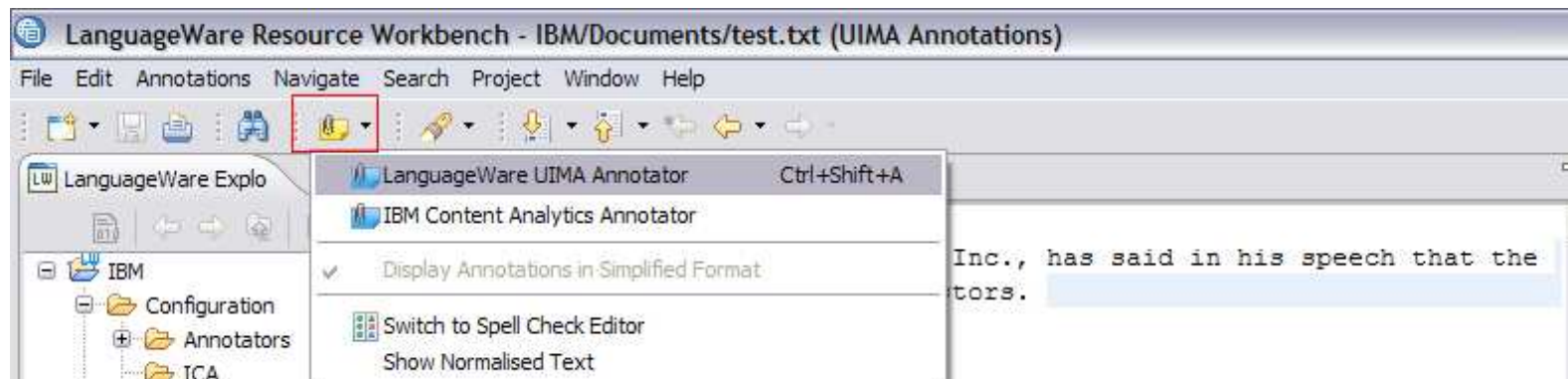
- The new document can be edited as a regular .txt file (type text in or paste copied text into it.)
- Enter some text sample representing the patterns you want to capture with your annotator (like person names associated to places, car models associated to registration plates, locations and company names...)
- This document can help you to develop and test the annotator you want to create.



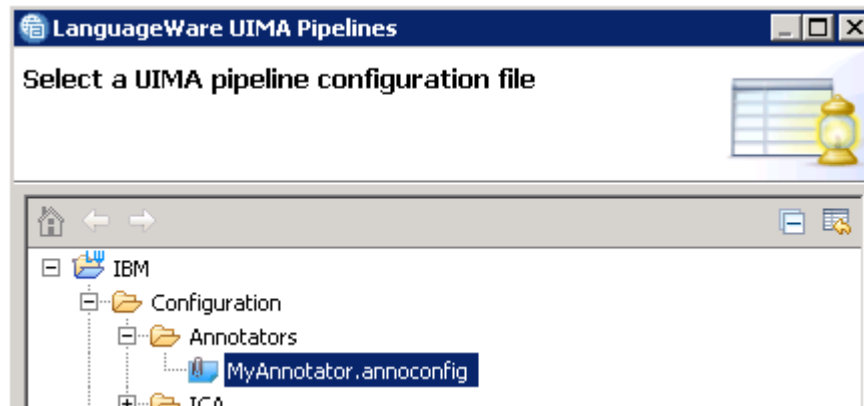
## Annotating a document

### How to run the annotator:

- Open the document in the editor, and make sure it is in focus (place the cursor inside the text field and click inside the text)..
- Select Annotations from the LRW menu and LanguageWare UIMA Annotator.
- Alternatively, you can right-click on the document and select “Analyze Document” , or you can simply click “Ctrl+ Shift+ A”



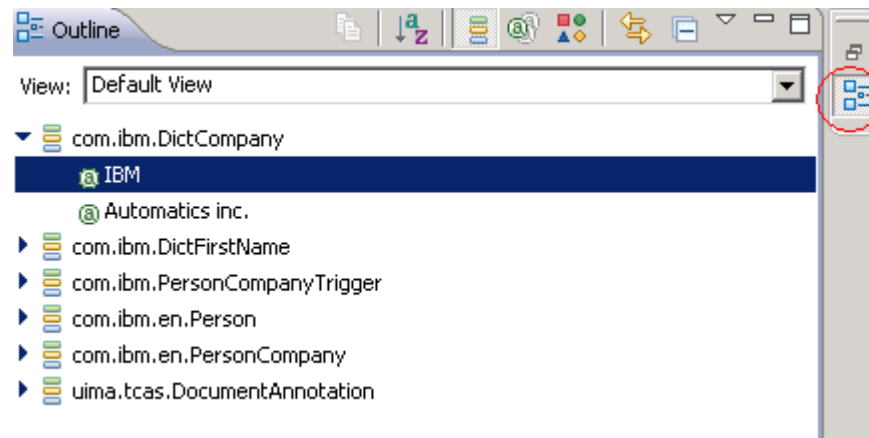
- Select the LanguageWare UIMA Annotator ( .annoconfig file) you created in the previous training module. Usually these files are stored in the Configuration/Annotators folder of the project. Click on OK.



## Annotating a document

### See occurrences in the Outline tab

- The **Outline** pane on the right hand side of the LRW should now display annotations as below:



- You may need to click on the **Outline icon** highlighted on the right hand side bar for the Outline pane to display.
- The collapsed **annotation types** found in the text are listed. When expanded, they display the particular **annotations** found in the text.
- The DictCompany and DictFirstName annotations displayed in the example come from the custom dictionaries in your Annotator Pipeline.

## Dictionary-based annotations

- These annotations are called **dictionary-based** (using the Dict prefix in their naming helps to identify them) annotations, as they come from matches found in the custom dictionaries.
- All annotations of the same type are highlighted in the document when clicking on one of them on the right hand side.
- All information associated with the dictionary entry is displayed in the **Properties** tab.

The screenshot displays an IDE interface with a text editor, an Outline panel, and a Properties panel. The text editor shows the following text:

```
Amine works for IBM Ireland.
John Doe, the CEO of Automatics inc., has said in his speech that the
company is looking for new investors.
```

The Outline panel on the right shows a tree view of annotations. The following entries are circled in red:

- com.ibm.DictCompany
  - @ IBM
  - @ Automatics inc.
- com.ibm.DictFirstName
  - @ Amine
  - John** (selected)

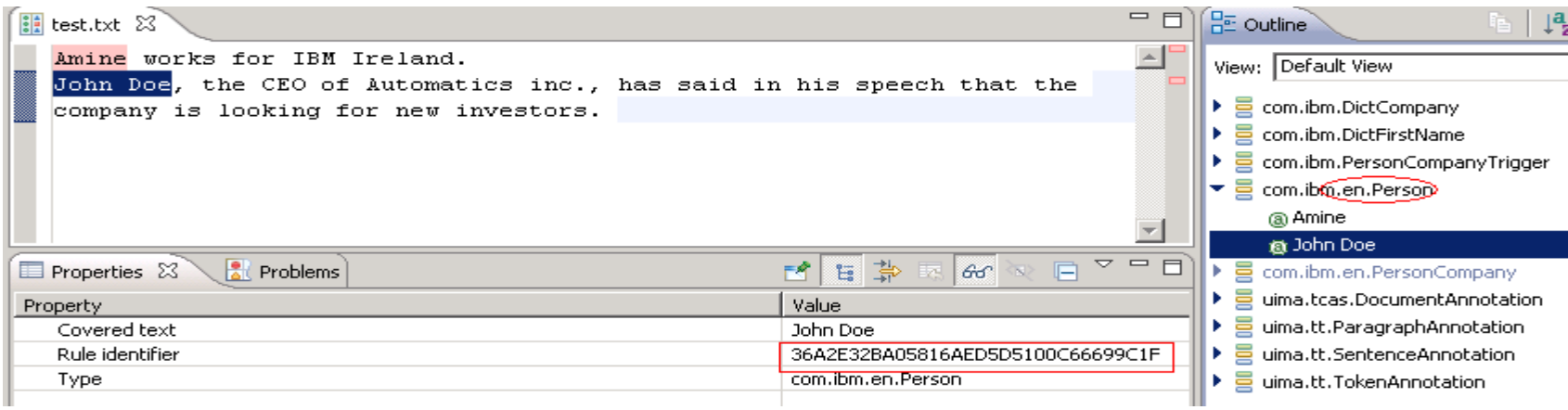
The Properties panel at the bottom shows the following table:

Property	Value
Covered text	John
gender	Male
Lemma	John
Part of speech	Noun
Type	com.ibm.DictFirstName

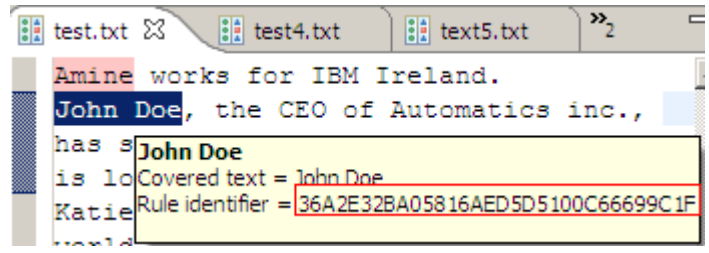
## Rule-based annotations\*

\* Covered in the next tutorial

- The highlighted annotations are rule-based. Note the type **en** that indicates English as the language used in the Annotator configuration.
- The Properties tab displays the ID of the rule that matched the **Person** type in the text..



- If you hover over the annotation in the text, you see the same information.





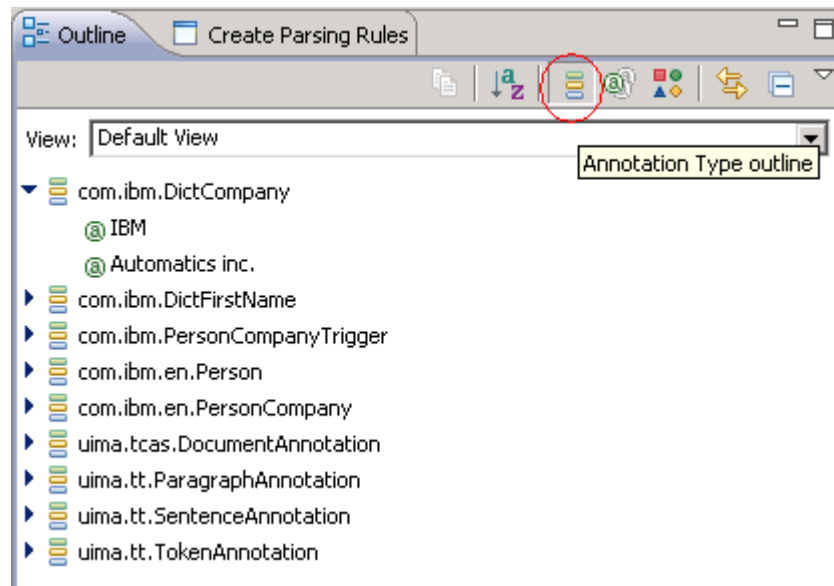
## Rule identifier

- The **Rule identifier** or Rule ID identifies the rule in the Parsing Rules database used in the annotator.
- Open the relevant Parsing Rules database and look for the same rule ID reference that was displayed in the annotation you are investigating.
- Double click on it to open the rule. It can be reviewed, edited and changed in order to improve or modify current annotations.
- After any changes, the rule has to be saved, the Parsing Rules database rebuilt and the Outline should display changes in the annotations affected by that rule.

X	Set	Type	Label	Original Text	Created	Modified	Id
<input type="checkbox"/>	Default	Aggregate		[Person Pers...	2011-01-17 ...	2011-01-17 ...	1674C82FBA803E1AC690297ABEECD177
<input type="checkbox"/>	Default	Phrases		John Doe	2011-01-17 ...	2011-01-17 ...	36A2E32BA05816AED5D5100C66699C1F

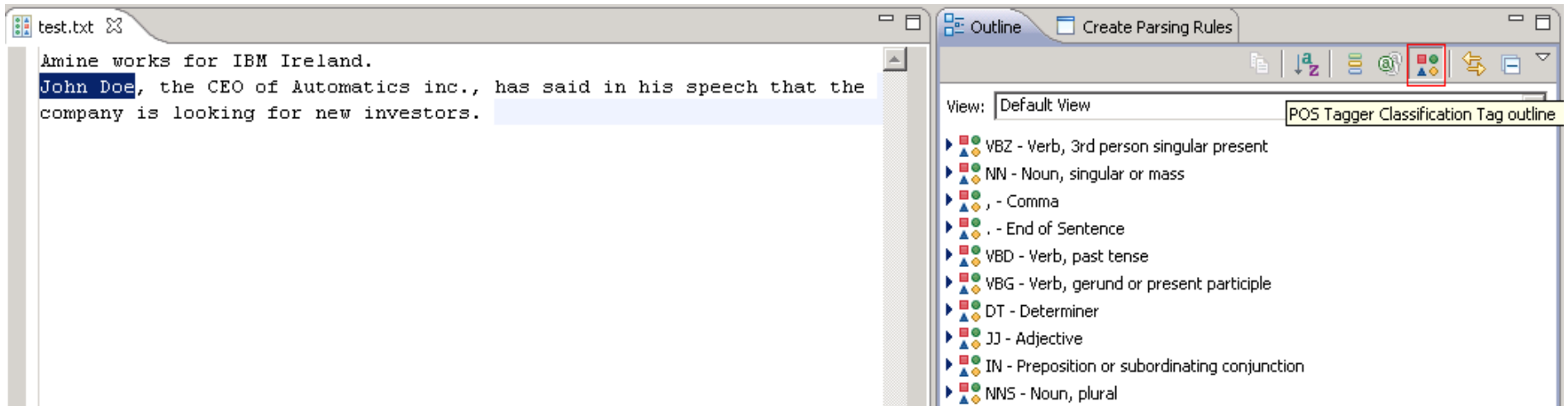
## Outline updates

- Once the changes to a rule are saved and the Parsing Rules database is rebuilt, the annotations from your annotator will be updated accordingly to your changes...
- Results in the Outline tab display as Annotation Types by default. This is the icon highlighted below:



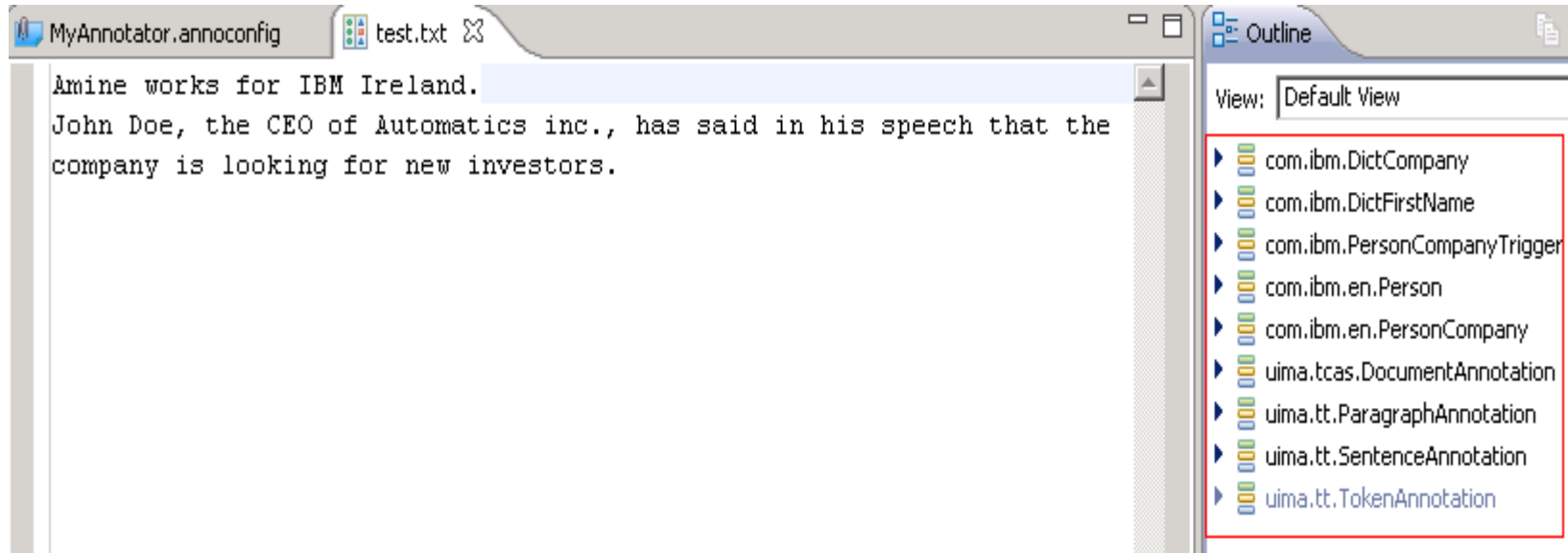
## The Outline view: Different options

- Other views are available (like Part Of Speech, POS Tagger Classification Tag) when clicking on the different icons on the tab bar.

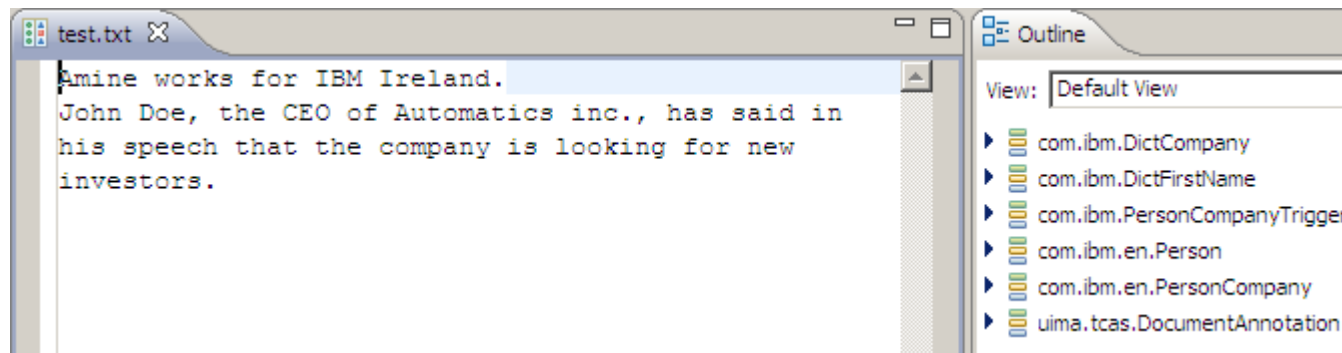


## Clean up stage

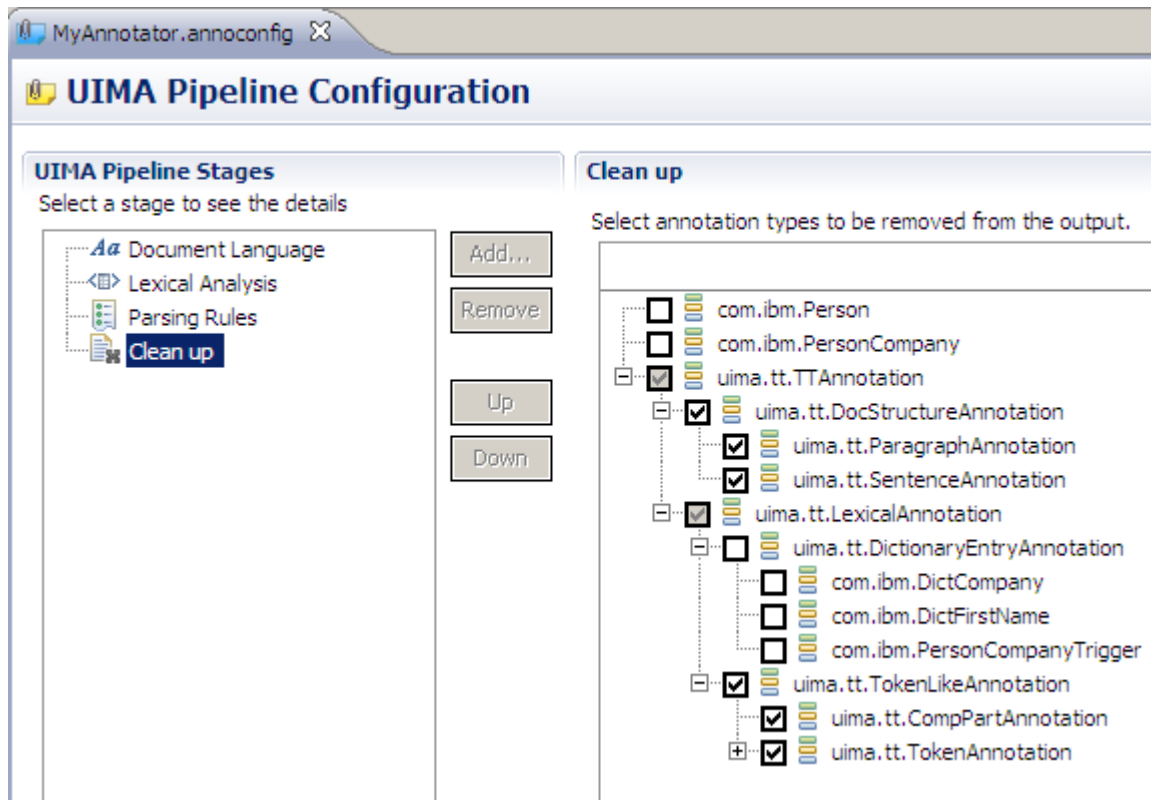
- By default, all annotation types found in a document are displayed in the Outline.



- However this can be edited to display only the desired annotation types.



- To hide/show the displayed annotations, edit the Clean-up stage in the pipeline configuration.
- Check the annotations you want to hide, then save the file.
- The outline will refresh automatically and you can see those changes.



## Module roadmap

- **Annotate a document**
  - Add or create a document
  - Run an annotator
  - See the occurrences in the Outline tab
  - Clean up stage
- **Summary and best practices**
- **Sample exercises**

## Module summary

You have completed this module and can:

- Add or create a document into the LRW workspace
- Annotate a document
- Review the annotations in the Outline tab
- Understand and trace the origins of those annotations

See the LanguageWare help for more tips and advanced use cases

## Best practices

- Creating a new document with examples of annotations you want to capture is a good way to start developing an annotator.
- Annotating documents in an iterative way will help you to perfect the parsing rules and improve your models.
- Reviewing the annotations will show you the shortcomings of your rules and will help you to find ways to improve them.
- Once you are happy with the annotations you are capturing in your document, try annotating a new document that contains the same type or similar annotations...The results may show areas to improve or problems to be fixed...



## Module roadmap

- **Annotate a document**
  - Add or create a document
  - Run an annotator
  - See the occurrences in the Outline tab
  - Clean up stage
- **Summary and best practices**
- **Sample exercises**

## Practice exercises

- Copy the SampleQuestion.txt file (supplied in Annex.zip) into the Documents folder of your project.
- Open the text file into the editor and annotate it with the "AnalyseHelpline" annotator and see the results in the Outline.

## Contacts

- If you have any questions, comments or suggestions, contact us using the LanguageWare email address [EMEALAN@ie.ibm.com](mailto:EMEALAN@ie.ibm.com) or on the developerWorks® forum.

## 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](http://www.ibm.com/legal/copytrade.shtml)" 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.

© Copyright International Business Machines Corporation 2011. All rights reserved.