Creating new Resource Types

Module 5

GATE Training Course
June 2016

© 2016 The University of Sheffield
This material is licenced under the Creative Commons Attribution-NonCommercial-ShareAlike Licence
(http://creativecommons.org/licenses/by-nc-sa/3.0/)
Outline

1. CREOLE Metadata
   - CREOLE Recap

2. Creating CREOLE Resources
   - Your First Language Analyser
   - Best Practice
   - Your First Visual Resource
   - Ready Made Applications

3. Advanced CREOLE
   - CREOLE Management
   - Corpus-level processing
   - Adding actions to the GUI
   - Distributing Your Plugins
Outline

1. CREOLE Metadata
   - CREOLE Recap

2. Creating CREOLE Resources
   - Your First Language Analyser
   - Best Practice
   - Your First Visual Resource
   - Ready Made Applications

3. Advanced CREOLE
   - CREOLE Management
   - Corpus-level processing
   - Adding actions to the GUI
   - Distributing Your Plugins
The GATE component model is called CREOLE (Collection of REusable Objects for Language Engineering).

CREOLE uses the following terminology:

- **CREOLE Plugins**: contain definitions for a set of resources.
- **CREOLE Resources**: Java objects with associated configuration.
- **CREOLE Configuration**: the metadata associated with Java classes that implement CREOLE resources.
CREOLE Plugins

CREOLE is organised as a set of plugins.

Each CREOLE plugin:

- is a directory on disk (or on a web server);
- is specified as a URL pointing to the directory;
- contains a special file called creole.xml;
- may contain one or more .jar files with compiled Java classes.
  - alternatively, the required Java classes may simply be placed on the application classpath.
- contains the definitions for a set of CREOLE resources.
CREOLE Plugins

A typical creole.xml:

```
<CREOLE-DIRECTORY>
  <JAR SCAN="true">this-plugin.jar</JAR>
  <JAR>lib/some-dependency.jar</JAR>
</CREOLE-DIRECTORY>
```

- Each `<JAR>` element specifies a JAR file to add to the plugin’s classloader
- `SCAN="true"` identifies JAR(s) that will be scanned to find CREOLE resource classes.
CREOLE Resources

A CREOLE resource is a Java Bean with some additional metadata.

A CREOLE resource class:

- must implement the `gate.Resource` interface (or one of its sub-interfaces `LanguageResource`, `ProcessingResource`, `VisualResource` or `Controller`);

- must be annotated with `@CreoleResource` to identify it as a resource class (this is what `SCAN="true"` looks for);

- must provide accessor methods for its parameters.
Outline

1. CREOLE Metadata
   - CREOLE Recap

2. Creating CREOLE Resources
   - Your First Language Analyser
   - Best Practice
   - Your First Visual Resource
   - Ready Made Applications

3. Advanced CREOLE
   - CREOLE Management
   - Corpus-level processing
   - Adding actions to the GUI
   - Distributing Your Plugins
Exercise 1: Create an Empty Processing Resource

Create a Java class:

```java
package module7;
import gate.creole.AbstractLanguageAnalyser;
import gate.creole.metadata.*;

@CreoleResource
public class DocStats extends AbstractLanguageAnalyser { }
```

- make sure it compiles;
- create a .jar file with the compiled class;
- **TIP:** see the build.xml file in your hands-on!
Exercise 1: Create an Empty Processing Resource

In order to make your resource available to GATE we need a plugin

Create a corresponding creole.xml file:

```
1 <CREOLE-DIRECTORY>
2   <JAR SCAN="true">module5.jar</JAR>
3 </CREOLE-DIRECTORY>
```
Exercise 1 (part 2): Implementation

Implement:

1. `public Resource init()` 
2. `throws ResourceInstantiationException { }`

...to print out a message;

Implement:

1. `public void execute() throws ExecutionException { }`

...to count the number of Token annotations in the input document, and set the value as a feature on the document.
Exercise 1: Solution

Try not to use this!
Exercise 1: Solution

Try not to use this!

```java
package module5;

import gate.Resource;
import gate.creole.*;
import gate.creole.metadata.*;

@CreoleResource
public class DocStats extends AbstractLanguageAnalyser {

    @Override
    public void execute() throws ExecutionException {
        int tokens = document.getAnnotations().get("Token").size();
        document.getFeatures().put("token_count", tokens);
    }

    @Override
    public Resource init() throws ResourceInstantiationException {
        System.out.println(getClass().getName() + " is initialising.");
        return this;
    }
}
```
CREOLE Annotations: @CreoleResource

The @CreoleResource annotation takes attributes giving metadata:

- **name** (String) the name of the resource.
- **comment** (String) a descriptive comment about the resource.
- **helpURL** (String) a URL to a help document on the web for this resource.
- **icon** (String) the icon to use to represent the resource in GATE Developer.

**Example**

```java
@CreoleResource (name = "Document Stats",
                comment = "Calculates document statistics.")
public class DocStats extends AbstractLanguageAnalyser {
    ...
}
```
CREOLE Annotations: @CreoleResource

Attributes for Visual Resources

If the resource being configured is a Visual Resource, you can also use the following attributes:

- `guiType` (GuiType enum) the type of GUI this resource defines.
- `resourceDisplayed` (String) the class name of the resource type that this VR displays, e.g. "gate.Corpus".
- `mainViewer` (boolean) is this VR the most important viewer for its displayed resource type?
Outline

1. CREOLE Metadata
   - CREOLE Recap

2. Creating CREOLE Resources
   - Your First Language Analyser
   - Best Practice
   - Your First Visual Resource
   - Ready Made Applications

3. Advanced CREOLE
   - CREOLE Management
   - Corpus-level processing
   - Adding actions to the GUI
   - Distributing Your Plugins

Creating new Resource Types
Best Practice: Use Parameters!

- Do not hardcode values, specify them as parameters.
- Values that change internal data structures, built when the PR is created, should be **init-time** parameters. These cannot be changed once the PR was created.
- Values that can be changed between executions should be **run-time** parameters.
- Try to make as many parameters as possible into **run-time** parameters!
- Provide *sensible defaults* for most parameters.
- If you have too many init-time parameters, use a config file instead!
- If you have too many run-time parameters, provide a Visual Resource!
- Make sure the parameters are well documented!
Best Practice: Input/Output

Specify Input/Output!

- If your PR uses annotations, always specify input and output annotation sets:
  - use a parameter `inputASName` for the input annotation set name;
  - use a parameter `outputASName` for the output annotation set name;

OR

- use a parameter named `annotationSetName` (if the PR only modifies existing annotations).
Defining Parameters

Creole parameters are Java Bean properties (a pair of get/set methods), with @CreoleParameter annotations on the setter method. Main attributes include:

- **comment** (String) an optional descriptive comment about the parameter.
- **defaultValue** (String) the optional default value for this parameter.
- **suffixes** (String) for URL-valued parameters, a semicolon-separated list of default file suffixes that this parameter accepts.

**Example**

```java
@CreoleParameter(defaultValue="", comment="The name for the input annotation set.")
public void setInputASName(String inputASName) {
    this.inputASName = inputASName;
}
```
CREOLE Annotations: Parameter Types

You can also use the following annotations to mark the type of a CREOLE parameter:

- `@Optional` for parameters that are not required.
- `@RunTime` for run-time parameters.

Corrected Example

```java
@Optional
@RunTime
@CreoleParameter(defaultValue="", comment="The name for the input annotation set.")

public void setInputASName(String inputAsName) {
    this.inputAsName = inputAsName;
}
```

**TIP:** More info at http://gate.ac.uk/userguide/sec:creole-model:config
Exercise 2: Develop/Test Cycle

Change the implementation from Exercise 1 to:

- add proper metadata on the resource class;
- use a parameter for the input annotation set;
- use a parameter for the Token annotation type;
- make sure these parameters have good defaults, and documentation;

Test it!

- Start GATE Developer, load a document, create an instance of the Unicode Tokeniser;
- load the module5 CREOLE plugin, create an instance of your PR; create a Corpus Pipeline and add the two PRs to it;
- run the pipeline over the document and check it works.
Exercise 3: Better Statistics

Change the implementation from *Exercise 2* to also calculate counts for all *words*, all *nouns*, all *verbs*.

**TIPs:**
You will need to run a Sentence Splitter, and POS Tagger after the Tokeniser, in order to get the part-of-speech information.

**Definitions:**

```python
word  {Token.kind=="word"}
noun  {Token.category.startsWith("NN")}
verb  {Token.category.startsWith("VB"))}
```

Test it!
Outline

1. CREOLE Metadata
   - CREOLE Recap

2. Creating CREOLE Resources
   - Your First Language Analyser
   - Best Practice
   - Your First Visual Resource
   - Ready Made Applications

3. Advanced CREOLE
   - CREOLE Management
   - Corpus-level processing
   - Adding actions to the GUI
   - Distributing Your Plugins

Creating new Resource Types
Visual Resources

- Visual Resources provide UI elements (Swing components) for building user interfaces.
- They are classes that implement the `gate.VisualResource` interface.
- They are associated with a type of resource via CREOLE metadata (which is used as a model for the view represented by the VR).
- The abstract class `gate.creole.AbstractVisualResource` can be used as a starting point.
Visual Resource API

Visual resources extend the `gate.Resource` interface, with:

```java
/**
 * set the object to be displayed */
public void setTarget(Object target);
```

AbstractVisualResource

- `extends` `javax.swing.JPanel`;
- implements all the methods required by `gate.Resource`;
- extending classes only need to implement:
  - `public Resource init()`: initialise the resource (i.e. build the required UI elements);
  - `public void setTarget(Object target)`: sets the model for this view.
Visual Resource CREOLE Metadata

- A Visual Resource is associated with a given type of object that it can display (or edit, configure). This association is done via CREOLE metadata on the VR implementation.

- From the API, the VR is populated by calling `setTarget(Object target)`.

- In GATE Developer, the appropriate VR types are instantiated on demand when a resource is double-clicked in the tree. E.g., when a Document is double-clicked, all VR registered as capable of displaying `gate.Document` targets are instantiated.
VR Metadata Example

CREOLE Annotations:

```java
@CreoleResource(name="Statistics Viewer",
    comment="Shows document statistics",
    resourceDisplayed="gate.Document",
    guiType=GuiType.LARGE,
    mainViewer=true)

public class StatsViewer extends AbstractVisualResource
```
Exercise 4: Show the Statistics

- Create a VR that, given a document, can show the statistics produced by the DocStats language analyser.
- add CREOLE metadata to associate the new VR with the `interface gate.Document`;

You can use a simple `JTextPane` to show a `toString()` value for the document’s features.
Exercise 4: Solution (metadata omitted)

Try not to use this!
Exercise 4: Solution (metadata omitted)

Try not to use this!

```java
package module7;
import javax.swing.*;
import gate.*;
import gate.creole.*;
import gate.event.FeatureMapListener;
public class StatsViewer extends AbstractVisualResource
    implements FeatureMapListener{
    private JTextPane textPane;
    private FeatureMap targetFeatures;
    public Resource init() throws ResourceInstantiationException {
        textPane = new JTextPane();
        add(new JScrollPane(textPane));
        return this;
    }
    public void setTarget(Object target) {
        if(targetFeatures != null) targetFeatures.removeFeatureMapListener(this);
        targetFeatures = ((Document)target).getFeatures();
        targetFeatures.addFeatureMapListener(this);
        featureMapUpdated();
    }
    public void featureMapUpdated() {
        textPane.setText(targetFeatures.toString());
    }
}
```
Outline

1. CREOLE Metadata
   - CREOLE Recap

2. Creating CREOLE Resources
   - Your First Language Analyser
   - Best Practice
   - Your First Visual Resource
   - Ready Made Applications

3. Advanced CREOLE
   - CREOLE Management
   - Corpus-level processing
   - Adding actions to the GUI
   - Distributing Your Plugins
Many CREOLE plugins contain one or more example applications

- they may be used to show how the processing resources can be used
- some plugins might only contain applications, i.e. the language plugins

Making these applications easily available through the GUI will make your processing resources easier for others to use.

Example applications can easily be added to the *Ready Made Applications* menu by creating an instance of `gate.creole.PackagedController`
Packaged Controller API

- Packaged Controllers extend the `gate.creole.PackagedController` class

```java
/** the URL of the pipeline XGAPP file */
public URL getPipelineURL();

/** the menu under which the application appears */
public List<String> getMenu();
```

- `gate.creole.PackagedController` is also a GATE resource so we can provide these values using CREOLE annotations
Example: Chinese IE

```java
package chinese;

import gate.creole.PackagedController;
import gate.creole.metadata.AutoInstance;
import gate.creole.metadata.AutoInstanceParam;
import gate.creole.metadata.CreoleParameter;
import gate.creole.metadata.CreoleResource;
import java.net.URL;
import java.util.List;

@CreoleResource(name = "Chinese IE System", icon = "ChineseLanguage",
autoinstances = @AutoInstance(parameters = {
    @AutoInstanceParam(name="pipelineURL", value="resources/chinese.gapp"),
    @AutoInstanceParam(name="menu", value="Chinese")}))
public class ChineseIE extends PackagedController {
    // all without writing any code!
}
```
Exercise 5: Show Off Your New Plugin

- create, and save, an application that shows how to use your statistics PR
- create a `gate.creole.PackagedController` instance to make the application available through the GUI.

You can use the Chinese IE example as a starting point.
Outline

1. CREOLE Metadata
   - CREOLE Recap

2. Creating CREOLE Resources
   - Your First Language Analyser
   - Best Practice
   - Your First Visual Resource
   - Ready Made Applications

3. Advanced CREOLE
   - CREOLE Management
   - Corpus-level processing
   - Adding actions to the GUI
   - Distributing Your Plugins
The CREOLE and DataStore Registers

The CREOLE Register

- Stores all CREOLE data, including:
  - which plugins are loaded;
  - which types of CREOLE Resources have been defined;
  - loaded instances of each resource type;
  - which Visual Resources can display any resource type;

- fires events when resources are loaded and deleted;
- forwards all events from the DataStore Register (see below).

The DataStore Register

- is a java.util.Set of DataStore objects.
- fires events when datastores are created, opened and closed.
CREOLE Register and its Events

```java
// Obtain a pointer to the CREOLE Register
CreoleRegister cReg = Gate.getCreoleRegister();

// listen to CREOLE events
// listen to CREOLE events
cReg.addCreoleListener(new CreoleListener() {
    public void resourceUnloaded(CreoleEvent e) { ... }  
    public void resourceRenamed(Resource resource, String oldName, String newName) { ... }  
    public void resourceLoaded(CreoleEvent e) { ... }  
    public void datastoreOpened(CreoleEvent e) { ... }  
    public void datastoreCreated(CreoleEvent e) { ... }  
    public void datastoreClosed(CreoleEvent e) { ... }  
});

// remove a registered listener
// remove a registered listener
cReg.removeCreoleListener(aListener);
```
Other CREOLE APIs

Plugins Management

```
// load a new CREOLE plugin
try {
    cReg.registerDirectories(new URL("..."));
    // register a single resource class without using creole.xml
    cReg.registerComponent(MyResource.class);
} catch (GateException e1) { ... }

// get all loaded plugins
cReg.getDirectories();

// remove a loaded plugin
cReg.removeDirectory(...);
```
Other CREOLE APIs (continued)

Find Loaded Resources

```java
// find all resources of a given type
try {
    cReg.getAllInstances("gate.LanguageAnalyser");
} catch (GateException e1) { ... }
```

Resource Types

```java
// get PR types (class names)
cReg.getPrTypes();
// get LR types (class names)
cReg.getLrTypes();
// get VR types (class names)
cReg.getVrTypes();
```
Other CREOLE APIs (continued)

CREOLE Metadata

```java
// Obtain the Resource Data about a resource
ResourceData rData = cReg.get("resource.class.name");

// get the list of instances
List<Resource> instances = rData.getInstantiations();

// get the list of parameters
ParameterList pList = rData.getParameterList();

// get the Init-time / Run-time parameters
List<List<Parameter>> someParams;
someParams = pList.getRuntimeParameters();
someParams = pList.getInitimeParameters();
```
Exercise 6: CREOLE Metadata

- load the ANNIE application;
- find out which plugins are loaded;
- find out which PR instances exist;
- find out which PR types are known to the system;
- find out what parameters they have.

You may find this useful:

```java
public void main(String[] args) throws Exception{
    Gate.init();
    // load the ANNIE application
    File annieGappFile = new File(
        new File(Gate.getPluginsHome(), "ANNIE"),
        "ANNIE_with_defaults.gapp");
    PersistenceManager.loadObjectFromFile(annieGappFile);
    // ...
}
```
Outline

1. CREOLE Metadata
   - CREOLE Recap

2. Creating CREOLE Resources
   - Your First Language Analyser
   - Best Practice
   - Your First Visual Resource
   - Ready Made Applications

3. Advanced CREOLE
   - CREOLE Management
   - Corpus-level processing
   - Adding actions to the GUI
   - Distributing Your Plugins
Corpus-level processing

- When running a PR over a corpus of more than one document, you may want to do some additional pre- and post-processing before the first and after the last document.

- To do this, implement `gate.creole.ControllerAwarePR`

- Three callback methods called at key points in the execution of the controller that contains the PR:
  - `controllerExecutionStarted`
  - `controllerExecutionFinished`
  - `controllerExecutionAborted`

- Parameter is the `Controller`.

- “aborted” callback also receives the `Throwable` that caused the abort.
Corpus-level processing

- So if the controller is a `CorpusController`, these correspond to:
  - before the first document
  - after the last document
  - when something goes wrong
ControllerAwarePR example

```java
public class ExampleAnalyser
    extends AbstractLanguageAnalyser
    implements ControllerAwarePR {

    public void controllerExecutionStarted(Controller c) {
        if (c instanceof CorpusController) {
            System.out.println("Processing corpus " +
                               ((CorpusController)c).getCorpus().getName());
        } else {
            System.out.println("Running in a simple pipeline");
        }
    }

    // controllerExecutionFinished is similar
}
```
Exercise 7: Corpus statistics

Add corpus statistics to your DocStats PR:

- Add private fields to keep a running total count of words (and nouns/verbs).
- **Implement** `ControllerAwarePR`.
- In the “started” callback, initialize these totals to 0.
- In the “finished” callback
  - check whether you are running in `CorpusController`
  - if so, put the total counts into features on the controller’s `Corpus`.
- You can leave the “aborted” callback empty (or just print a message).
Exercise 7: Solution

```java
package module7;

// imports omitted for space reasons

@CreoleResource(name = "Corpus statistics")
public class CorpusStats extends AbstractLanguageAnalyser
    implements ControllerAwarePR {
    private int totalTokens;

    public void execute() throws ExecutionException {
        int tokens = document.getAnnotations().get("Token").size();
        document.getFeatures().put("token_count", tokens);
        totalTokens += tokens; // keep a running total
    }

    public void controllerExecutionStarted(Controller c) {
        totalTokens = 0;
    }

    public void controllerExecutionFinished(Controller c) {
        if (c instanceof CorpusController) {
            ((CorpusController)c).getCorpus().getFeatures()
                .put("token_count", totalTokens);
        }
    }

    // controllerExecutionAborted omitted
}
```
Outline

1. CREOLE Metadata
   - CREOLE Recap

2. Creating CREOLE Resources
   - Your First Language Analyser
   - Best Practice
   - Your First Visual Resource
   - Ready Made Applications

3. Advanced CREOLE
   - CREOLE Management
   - Corpus-level processing
   - Adding actions to the GUI
   - Distributing Your Plugins
Adding actions to the GUI

- Any (language, processing or visual) resource can contribute *actions* to the GATE developer GUI.

- These appear as items on the resource’s right-click menu. For example:
  - The “Run” option for controllers comes from the controller editor VR
  - The “Save as...” and “Delete ontology data” options for an ontology LR come from the LR itself.

- This is done by implementing the interface `gate.gui.ActionsPublisher`

- One method, returning a `List` of `javax.swing.Action` objects.
Exercise 8: ActionsPublisher

Implement cumulative statistics for your DocStats PR:

- keep a running total as before, but rather than resetting it in `controllerExecutionStarted`, provide an action to reset it explicitly.
- provide another action to display the current total.
Exercise 8: Solution

```java
package module7;
// imports omitted for space reasons

@CreoleResource(name = "Cumulative statistics")
public class CumulativeStats extends AbstractLanguageAnalyser implements ActionsPublisher {
    // totalTokens and execute() method exactly as in exercise 7

    public List<Action> getActions() {
        if (actions == null) {
            actions.add(new AbstractAction("Reset counter") {
                @Override
                public void actionPerformed(ActionEvent e) {
                    totalTokens = 0;
                }
            });

            actions.add(new AbstractAction("Show current total") {
                @Override
                public void actionPerformed(ActionEvent e) {
                    JOptionPane.showMessageDialog(
                       MainFrame.getInstance(), totalTokens + " token(s) since last reset."));
                }
            });
        }
        return actions;
    }

    private List<Action> actions;
```
Outline

1. CREOLE Metadata
   - CREOLE Recap

2. Creating CREOLE Resources
   - Your First Language Analyser
   - Best Practice
   - Your First Visual Resource
   - Ready Made Applications

3. Advanced CREOLE
   - CREOLE Management
   - Corpus-level processing
   - Adding actions to the GUI
   - Distributing Your Plugins
Distributing Your Plugins

- GATE comes with a lot of plugins (61 at last count)
- There are a lot of other 3rd party plugins available but...
  - they can be difficult to find
  - they need to be manually installed
- GATE plugins can also be automatically installed from a CREOLE plugin repository
  - four plugin repositories, providing 16 plugins, currently listed
  - it is easy to create and host your own repository
  - we’ll happily include your repository in GATE to advertise your plugins!
Creating a CREOLE Plugin Repository

- A CREOLE plugin repository is just a single XML file which
  - points to a CREOLE directory
  - optionally provides a download link (assumes creole.zip if not specified)

- A few extra details also need to be added to the CREOLE-DIRECTORY element of your creole.xml file. You need to provide at least
  - ID a unique ID for this plugin which should look like a Java class/package name. The class portion will be used as the plugin name in the GUI.
  - version the version number for this release of the plugin
Example CREOLE Plugin Repository

```xml
<?xml version="1.0"?>
<UpdateSite>
  <CreolePlugin
    url="http://example.com/plugins/sample1/" />

  <CreolePlugin
    url="sample2/"
    downloadURL="http://example.com/sample2.zip" />
</UpdateSite>
```

- for all the gory details see the userguide, specifically
  - http://gate.ac.uk/userguide/sec:development:pluginrepository
  - http://gate.ac.uk/userguide/sec:creole-model:config
Thank you!

Questions?

More answers at:

- http://gate.ac.uk (Our website)
- http://gate.ac.uk/mail/ (Our mailing list)