

The GATE Embedded API

Module 5

Seventh GATE Training Course
June 2014

© 2014 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** GATE API Basics
- 2** The CREOLE Model
 - CREOLE Basics
 - Resources, Parameters, Features
 - Annotations, Documents, Corpora
- 3** Execution Control
 - Processing Resources and Language Analysers
 - Controllers

Before We Start

Prerequisites

- Java 7 or 8 JDK (OpenJDK or Oracle)
- latest GATE Developer/Embedded (version 8.0)
- Java Development Environment such as Eclipse (not compulsory but *highly* recommended!).

Hands-on resources

- Download hands-on resources from the participants' wiki
- Unzip it somewhere on your hard disk
- Have a look at the build.xml
- Create a Java project in your IDE (call it `module5`)
- Write a Hello World program to test your configurations

Your First GATE-Based Project

Libraries to include

- `<gate-install-dir>/bin/gate.jar`
- `<gate-install-dir>/lib/*.jar`

Documentation

GATE Documentation

Core documentation from Sheffield

- [The Developer and Embedded User Guide](#)
- [GATE Developer screencams](#)
- [The mailing list](#)
- The [core GATE plugins list](#)
- [Papers / theory](#) (and how to cite our work)
- Brochures: [4 pages](#) or [2-pages](#)
- Acres of [API and software docs](#), ([Javadocs](#)) and [example code](#)
- A [May 2010 summary presentation](#)
- For technical help please see [the support page](#) (or [below](#))
- [GATE's public wiki](#) (running on GATE wiki, of course)

For Eclipse users

- Open your Eclipse preferences (in the usual place on the application menu for Mac, under *Window* on other platforms)
- Java → Build Path → User Libraries
- Create a new library named *GATE*
- *Add JARs...*
 - add `bin/gate.jar` and all the JARs in `gate/lib`
- expand the `gate.jar` entry, edit the “Source attachment” and point it at the `src` directory in your GATE.
- OK
- Add this user library to the build path of your module5 project.
- Eclipse configurations for later modules assume this library.

Exercise 1: Loading a Document

Try this:

```
1 import gate.*;
2 public class Main {
3     public static void main(String[] args)
4         throws Exception{
5         Gate.init(); // prepare the library
6         // create a new document
7         Factory.newDocument("This is a document");
8     }
9 }
```

Interacting with GATE

Using GATE Developer



Interacting with GATE

Using GATE Developer



Using GATE API

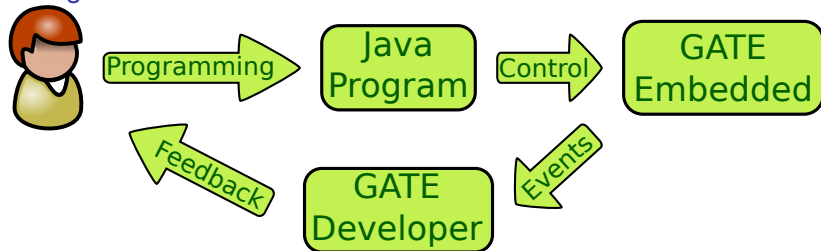


Interacting with GATE

Using GATE Developer



Using GATE API



Loading a Document (take 2)

```
1 package gatetutorial;
2
3 import gate.*;
4 import gate.gui.*;
5
6 public class Main {
7     public static void main(String[] args)
8         throws Exception{
9         //prepare the library
10        Gate.init();
11        //show the main window
12        MainFrame.getInstance().setVisible(true);
13        //create a document
14        Factory.newDocument("This is a document");
15    }
16 }
```

Note for (at least) Mac Users

If the previous slide doesn't work for you, you may need:

```
3 import javax.swing.SwingUtilities;

11 // show the main window
12 SwingUtilities.invokeLater(new Runnable() {
13     public void run() {
14         MainFrame.getInstance().setVisible(true);
15     }
16 });
```

Outline

- 1 GATE API Basics
- 2 The CREOLE Model
 - CREOLE Basics
 - Resources, Parameters, Features
 - Annotations, Documents, Corpora
- 3 Execution Control
 - Processing Resources and Language Analysers
 - Controllers

CREOLE

The GATE component model is called CREOLE (**C**ollection of **RE**usable **O**bjects for **L**anguage **E**ngineering).

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 Resources

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

A CREOLE resource:

- must implement the `gate.Resource` interface;
- must provide accessor methods for its parameters;
- must have associated CREOLE metadata.

The CREOLE metadata associated with a resource:

- can be provided inside the `creole.xml` file for the plugin;
- can be provided as special Java annotations inside the source code (recommended).

More details about this later today!

Outline

1 GATE API Basics

2 The CREOLE Model

- CREOLE Basics
- **Resources, Parameters, Features**
- Annotations, Documents, Corpora

3 Execution Control

- Processing Resources and Language Analysers
- Controllers

GATE Resource Types

There are three types of resources:

- **Language Resources (LRs)** used to encapsulate data (such as documents and corpora);
- **Processing Resources (PRs)** used to describe algorithms;
- **Visual Resources (VRs)** used to create user interfaces.

The different types of GATE resources relate to each other:

- PRs run over LRs,
- VRs display and edit LRs,
- VRs manage PRs, ...

These associations are made via CREOLE configuration.

GATE Feature Maps

Feature Maps...

- are simply Java Maps, with added support for firing events.
- are used to provide parameter values when creating and configuring CREOLE resources.
- are used to store metadata on many GATE objects.

All GATE resources are feature bearers

(they implement `gate.util.FeatureBearer`):

```
1 public interface FeatureBearer{
2     public FeatureMap getFeatures();
3
4     public void setFeatures(FeatureMap features);
5 }
```

FeatureMap Implementation

gate.FeatureMap

```
1 public interface FeatureMap extends Map<Object, Object>
2 {
3     public void removeFeatureMapListener(
4         FeatureMapListener l);
5     public void addFeatureMapListener(
6         FeatureMapListener l);
7 }
```

Events: gate.event.FeatureMapListener

```
1 public interface FeatureMapListener
2     extends EventListener
3 {
4     public void featureMapUpdated();
5 }
```

Resource Parameters

The behaviour of GATE resources can be affected by the use of parameters.

Parameter values:

- are provided as populated feature maps.
- can be any Java Object;
- This includes GATE resources!

Parameter Types

There are two types of parameters:

Init-time Parameters

- Are used during the instantiating resources.
- Are available for all resource types.
- Once set, they cannot be changed.

Run-time Parameters

- are only available for Processing Resources.
- are set before executing the resource, and are used to affect the behaviour of the PR.
- can be changed between consecutive runs.

Creating a GATE Resource

Always use the GATE Factory to create and delete GATE resources!

`gate.Factory`

```
1 public static Resource createResource(  
2     String resourceName,  
3     FeatureMap parameterValues,  
4     FeatureMap features,  
5     String resourceName){  
6     ...  
7 }
```

Only the first parameter is required; other variants of this method are available, which require fewer parameters.

Creating a GATE Resource

You will need the following values:

- **String resourceName**: the class name for the resource you are trying to create. This should be a string with the fully-qualified class name, e.g.
`"gate.corpora.DocumentImpl"`.
- **FeatureMap parameterValues**: the values for the init-time parameters. Parameters that are not specified will get their default values (as described in the CREOLE configuration). It is an error for a required parameter not to receive a value (either explicit or default)!
- **FeatureMap features**: the initial values for the new resource's features.
- **String resourceName**: the name for the new resource.

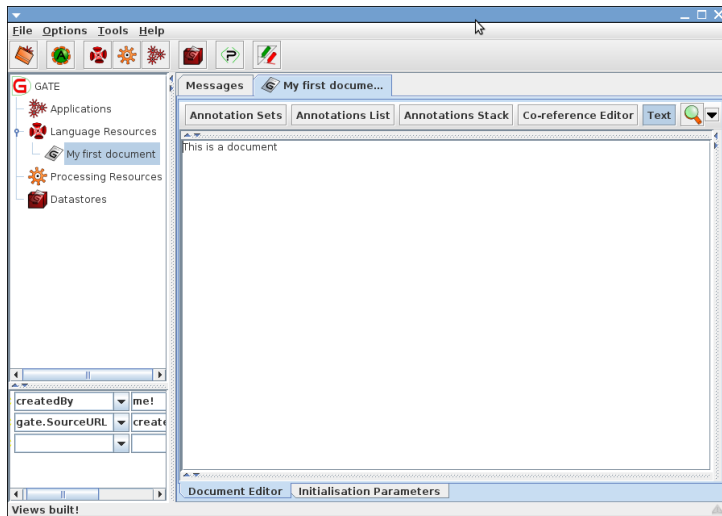
Example: Load a Document (take 3)

```
1 FeatureMap params = Factory.newFeatureMap();
2 params.put (
3     Document.DOCUMENT_STRING_CONTENT_PARAMETER_NAME,
4     "This is a document!");
5 FeatureMap feats = Factory.newFeatureMap();
6 feats.put ("createdBy", "me!");
7 Factory.createResource ("gate.corpora.DocumentImpl",
8     params, feats, "My first document");
```

TIP: Resource Parameters

The easiest way to find out what parameters resources take (and which ones are required, and what types of values they accept) is to use the GATE Developer UI and try to create the desired type of resource in the GUI!

Example: Load a Document (take 3)



Shortcuts for Loading GATE Resources

Loading a GATE document

```
1 import gate.*;  
2 //create a document from a String content  
3 Document doc = Factory.newDocument("Document text");  
4 //...or a URL  
5 doc = Factory.newDocument(new URL("http://gate.ac.uk"));  
6 //...or a URL and a specified encoding  
7 doc = Factory.newDocument(new URL("http://gate.ac.uk"),  
8     "UTF-8");
```

Loading a GATE corpus

```
1 Corpus corpus = Factory.newCorpus("Corpus Name");
```

Exercise 2: Loading a Document (again!)

Load a document:

- using the GATE home page as a source;
- using the UTF-8 encoding;
- having the name "This is home";
- having a feature named "date", with the value the current date.

TIP: Make sure the GATE Developer main window is shown to test the results!

Outline

1 GATE API Basics

2 The CREOLE Model

- CREOLE Basics
- Resources, Parameters, Features
- Annotations, Documents, Corpora

3 Execution Control

- Processing Resources and Language Analysers
- Controllers

GATE Documents

A GATE Document comprises:

- a DocumentContent object;
- a Default annotation set (which has no name);
- zero or more named annotation sets;

A Document is also a type of Resource, so it also has:

- a name;
- features.

Main Document API Calls

```
1 // Obtain the document content
2 public DocumentContent getContent();
3 // Get the default annotation set.
4 public AnnotationSet getAnnotations();
5 // Get a named annotation set.
6 public AnnotationSet getAnnotations(String name);
7 // Get the names for the annotation sets.
8 public Set<String> getAnnotationSetNames();
9 // Get all named annotation sets.
10 public Map<String, AnnotationSet>
11     getNamedAnnotationSets();
12 // Convert to GATE stand-off XML
13 public String toXml();
14 // Convert some annotations to inline XML.
15 public String toXml(Set aSourceAnnotationSet,
16     boolean includeFeatures);
```

Annotation Sets

GATE Annotation Sets...

- maintain a set of **Node** objects (which are associated with offsets in the document content);
- and a set of annotations (which have a start and an end node).
- implement the **gate.AnnotationSet** interface;
- ... which extends **Set<Annotation>**.
- implement several **get ()** methods for obtaining the included annotations according to various constraints.
- are created, deleted, and managed by the Document they belong to.

TIP: always use a Document object to create a new annotation set! Do not use the constructor!

Main AnnotationSet API Calls

Nodes

```
1 // Get the node with the smallest offset.  
2 public Node firstNode();  
3 // Get the node with the largest offset.  
4 public Node lastNode();
```

Creating new Annotations

```
1 // Create (and add) a new annotation  
2 public Integer add(Long start, Long end,  
3     String type, FeatureMap features);  
4 // Create (and add) a new annotation  
5 public Integer add(Node start, Node end,  
6     String type, FeatureMap features)
```


AnnotationSet API (continued)

Getting Annotations by ID, or type

```
1 // Get annotation by ID
2 public Annotation get(Integer id);
3 // Get all annotations of one type
4 public AnnotationSet get(String type)
5 // Get all annotation types present
6 public Set<String> getAllTypes()
7 // Get all annotations of specified types
8 public AnnotationSet get(Set<String> types)
```

AnnotationSet API (continued)

Getting Annotations by position

```
1 // Get all annotations starting at a given
2 // location, or right after.
3 public AnnotationSet get(Long offset)
4 // Get all annotations that overlap an interval
5 public AnnotationSet get(Long startOffset,
6     Long endOffset)
7 // Get all annotations within an interval.
8 public AnnotationSet getContained(Long startOffset,
9     Long endOffset)
10 // Get all annotations covering an interval.
11 public AnnotationSet getCovering(String neededType,
12     Long startOffset, Long endOffset)
```

AnnotationSet API (continued)

Combined get methods

```
1 // Get by type and feature constraints.
2 public AnnotationSet get(String type,
3     FeatureMap constraints)
4 // Get by type, constraints and start position.
5 public AnnotationSet get(String type,
6     FeatureMap constraints, Long offset)
7 // Get by type, and interval overlap.
8 public AnnotationSet get(String type,
9     Long startOffset, Long endOffset)
10 // Get by type and feature presence
11 public AnnotationSet get(String type,
12     Set featureNames)
```

Exercise 3: The AnnotationSet API

For the document loaded in exercise 2:

- find out how many named annotation sets it has;
- find out how many annotations each set contains;
- for each annotation set, for each annotation type, find out how many annotations are present.

TIP: Make sure the GATE Developer main window is shown to test the results!

Annotations

GATE Annotations...

- are metadata associated with a document segment;
- have a type (**String**);
- have a start and an end Node (**gate.Node**);
- have features;
- are created, deleted and managed by annotation sets.

TIP: always use an annotation set to create a new annotation! Do not use the constructor.

Annotation API

Main Annotation methods:

```
1 public String getType();  
2 public Node getStartNode();  
3 public Node getEndNode();  
4 public FeatureMap getFeatures();
```

gate.Node

```
1 public Long getOffset();
```

Exercise 4: Annotation API

Implement the following:

- Use the document created in exercise 3;
- Use the annotation set **Original markups** and obtain annotations of type **a** (anchor).
- Iterate over each annotation, obtain its features and print the value of **href** feature.

TIP: Before printing the value of **href** feature, use the **new URL(URL context, String spec)** constructor such that the value of the **href** feature is parsed within the context of the document's source url.

Outline

- 1 GATE API Basics
- 2 The CREOLE Model
 - CREOLE Basics
 - Resources, Parameters, Features
 - Annotations, Documents, Corpora
- 3 Execution Control**
 - Processing Resources and Language Analysers
 - Controllers

GATE Processing Resources

Processing Resources (**PRs**) are java classes that can be executed.

gate.Executable

```
1 public interface Executable {  
2     public void execute() throws ExecutionException;  
3     public void interrupt();  
4     public boolean isInterrupted();  
5 }
```

gate.ProcessingResource

```
1 public interface ProcessingResource  
2     extends Resource, Executable  
3 {  
4     public void reInit()  
5         throws ResourceInstantiationException;  
6 }
```

Language Analysers

Analysers are PRs that are designed to run over the documents in a corpus.

```
1 public interface LanguageAnalyser
2     extends ProcessingResource {
3
4     // Set the document property for this analyser.
5     public void setDocument(Document document);
6
7     // Get the document property for this analyser.
8     public Document getDocument();
9
10    // Set the corpus property for this analyser.
11    public void setCorpus(Corpus corpus);
12
13    // Get the corpus property for this analyser.
14    public Corpus getCorpus();
```

Loading a CREOLE Plugin

- Documents and corpora are built in resource types.
- All other CREOLE resources are defined as plugins.
- Before instantiating a resource, you need to load its CREOLE plugin first!

Loading a CREOLE plugin

```
1 //get the root plugins dir
2 File pluginsDir = Gate.getPluginsHome();
3 //Let's load the Tools plugin
4 File aPluginDir = new File(pluginsDir, "Tools");
5 //load the plugin.
6 Gate.getCreoleRegister().registerDirectories(
7     aPluginDir.toURI().toURL());
```

Exercise 5: Run a Tokeniser

Implement the following:

- Load the plugin named "ANNIE";
- Instantiate a Language Analyser of type `gate.creole.tokeniser.DefaultTokeniser` (using the default values for all parameters);
- set the **document** of the tokeniser to the document created in exercise 2;
- set the **corpus** of the tokeniser to `null`;
- call the **execute ()** method of the tokeniser;
- inspect the document and see what the results were.

Outline

- 1 GATE API Basics
- 2 The CREOLE Model
 - CREOLE Basics
 - Resources, Parameters, Features
 - Annotations, Documents, Corpora
- 3 Execution Control**
 - Processing Resources and Language Analysers
 - Controllers**

GATE Controllers

- Controllers provide the implementation for execution control in GATE.
- They are called *applications* in GATE Developer.
- The implementations provided by default implement a *pipeline* architecture (they run a set of PRs one after another).
- Other kind of implementations are also possible.
 - e.g. the Groovy plugin provides a *scriptable* controller implementation (more details in module 8).
- A controller is a class that implements **gate.Controller**.

Implementation

gate.Controller

```
1 public interface Controller extends Resource,  
2     Executable, NameBearer, FeatureBearer {  
3     public Collection getPRs();  
4     public void setPRs(Collection PRs);  
5     public void execute() throws ExecutionException;  
6 }
```

- all default controller implementations also implement **gate.ProcessingResource** (so you can include controllers inside other controllers!);
- like all GATE resources, controllers are created using the **Factory** class;
- controllers have names, and features.

Default Controller Types

The following default controller implementations are provided (all in the `gate.creole` package):

- **SerialController**: a pipeline of PRs.
- **ConditionalSerialController**: a pipeline of PRs. Each PR has an associated **RunningStrategy** value which can be used to decide **at runtime** whether or not to run the PR.
- **SerialAnalyserController**: a pipeline of **LanguageAnalysers**, which runs all the PRs over all the documents in a **Corpus**. The **corpus** and **document** parameters for each PR are set by the controller.
- **RealtimeCorpusController**: a version of **SerialAnalyserController** that interrupts the execution over a document when a specified timeout has lapsed.

SerialAnalyserController API

SerialAnalyserController is the most used type of Controller. Its most important methods are:

```
1 // Adds a new PR at a given position
2 public void add(int index, ProcessingResource pr);
3 // Adds a new PR at the end
4 public void add(ProcessingResource pr);
5 // Replaces the PR at a given position
6 public ProcessingResource set(int index,
7     ProcessingResource pr);
8 // Remove a PRs by position
9 public ProcessingResource remove(int index);
10 // Remove a specified PR
11 public boolean remove(ProcessingResource pr);
12 // Sets the corpus to be processed
13 public void setCorpus(gate.Corpora corpus);
14 // Runs the controller
15 public void execute() throws ExecutionException;
```

Exercise 6: Run a Tokeniser (again!)

Implement the following:

- Create a `SerialAnalyserController`, and add the tokeniser from exercise 5 to it;
- Create a corpus, and add the document from exercise 2 to it;
- Set the **corpus** value of the controller to the newly created corpus;
- Execute the controller;
- Inspect the results.

Controller Persistency (or *Saving Applications*)

- The configuration of a controller (i.e. the list of PRs included, as well as the features and parameter values for the controller and its PRs) can be saved using a special type of XML serialisation.
- This is done using the **`gate.util.persistence.PersistenceManager`** class.
- This is what *GATE Developer* does when saving and loading applications.

Implementation

gate.util.persistence.PersistenceManager

```
1 // Serialises the configuration of a GATE object
2 // to a special XML format.
3 public static void saveObjectToFile(Object obj,
4     File file) throws PersistenceException,
5     IOException ;
6
7 // Re-creates the serialised GATE object from the saved
8 // configuration data.
9 public static Object loadObjectFromFile(File file)
10     throws PersistenceException, IOException,
11     ResourceInstantiationException;
12 // Loads a GATE object from a [remote] location.
13 public static Object loadObjectFromUrl(URL url)
14     throws PersistenceException, IOException,
15     ResourceInstantiationException;
```

Thank you!

Questions?

More answers at:

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

Exercise 2: Solution

A possible solution

```
1 import gate.*;
2 import java.net.URL;
3 import java.util.Date;
4
5 public class Main {
6     public static void main(String[] args)
7         throws Exception{
8
9         // prepare the library
10        Gate.init();
11        // show the main window
12        MainFrame.getInstance().setVisible(true);
13
14        // init-time parameter for document
15        FeatureMap params = Factory.newFeatureMap();
16        params.put(Document.DOCUMENT_URL_PARAMETER_NAME,
17            new URL("http://www.gate.ac.uk"));
```

Exercise 2: Solution (continued)

```
19  params.put(Document.DOCUMENT_ENCODING_PARAMETER_NAME,  
20      "UTF-8");  
21  
22      // document features  
23  FeatureMap feats = Factory.newFeatureMap();  
24  feats.put("date", new Date());  
25  Factory.createResource("gate.corpora.DocumentImpl",  
26      params, feats, "This is home");  
27  }  
28 }
```

Exercise 2: Solution (continued)

The screenshot shows the GATE application window. The menu bar includes File, Options, Tools, and Help. The toolbar contains icons for file operations and navigation. The left sidebar shows a tree view with categories: Applications, Language Resources, Processing Resources, and Databases. The main area is divided into a Messages pane at the top and a Document Editor at the bottom. The Messages pane shows the URL 'GATE.ac.uk - index.html' and a list of search results. The Document Editor shows the text content of the page.

Messages This is home

Annotation Sets Annotations List Annotations Stack Co-reference Editor Text

GATE.ac.uk - index.html

Search: Log in Help
Print
Products GATE Developer
GATE Embedded
GATE Teamware
The GATE Process
GATE Cloud
GATE Mimir
GATE Wiki
Training/Certification

Services Customer stories
Why GATE?
Training/Certification

Support Community-based
For business

Download
Documentation User guide
Mailing list
Wiki JAPE tips
Example code
Beginners' quick-start guide
Groovy recipes

Screencams
Books
Core plugins
Roachura

Document Editor Initialisation Parameters

MimeType text/html
date Wed May 11 15:58:42 BST 2011
gate.SourceURL http://www.gate.ac.uk

Views built!

Exercise 3: Solution

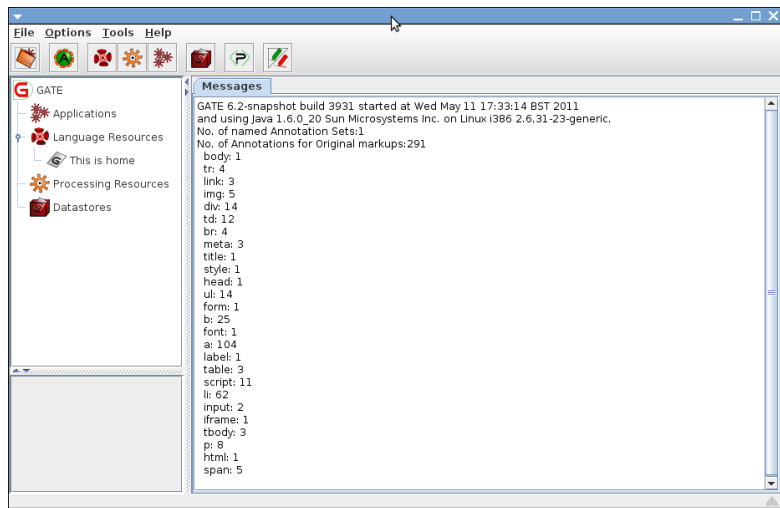
Additions to the solution of Exercise 2

```
1      ...
2
3      // obtain a map of all named annotation sets
4      Map<String, AnnotationSet> namedASes =
5          doc.getNamedAnnotationSets();
6      System.out.println("No. of named Annotation Sets:"
7          + namedASes.size());
8
9      // no of annotations each set contains
10     for (String setName : namedASes.keySet()) {
11         // annotation set
12         AnnotationSet aSet = namedASes.get(setName);
13         // no of annotations
14         System.out.println("No. of Annotations for " +
15             setName + ":" + aSet.size());
```

Exercise 3: Solution (Continued..)

```
17 // all annotation types
18 Set<String> annotTypes = aSet.getAllTypes();
19 for(String aType : annotTypes) {
20     System.out.println(" " + aType + ": "
21         + aSet.get(aType).size());
22 }
23 }
24
25 ...
```

Exercise 3: Solution (continued)



The screenshot shows the GATE application window. The menu bar includes File, Options, Tools, and Help. The left sidebar shows a tree view with categories: Applications, Language Resources, This is home, Processing Resources, and Datastores. The main Messages window displays the following text:

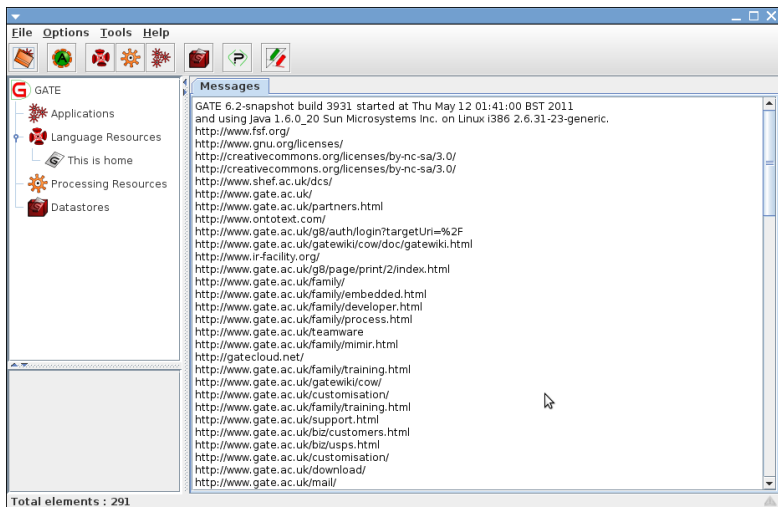
```
GATE 6.2-snapshot build 3931 started at Wed May 11 17:33:14 BST 2011
and using java 1.6.0_20 Sun Microsystems Inc. on Linux i386 2.6.31-23-generic.
No. of named Annotation Sets:1
No. of Annotations for Original markups:291
body: 1
tr: 4
link: 3
img: 5
div: 14
td: 12
br: 4
meta: 3
title: 1
style: 1
head: 1
ul: 14
form: 1
b: 25
font: 1
a: 104
label: 1
table: 3
script: 11
li: 62
input: 2
iframe: 1
tbody: 3
p: 8
html: 1
span: 5
```

Exercise 4: Solution

Additions to the solution of Exercise 2

```
1      ...
2
3      // obtain the Original markups annotation set
4      AnnotationSet origMarkupsSet =
5          doc.getAnnotations("Original markups");
6
7      // obtain annotations of type 'a'
8      AnnotationSet anchorSet = origMarkupsSet.get("a");
9
10     // iterate over each annotation
11     // obtain its features and print the value of href feature
12     for (Annotation anchor : anchorSet) {
13         String href = (String) anchor.getFeatures().get("href");
14         if(href != null) {
15             // resolving href value against the document's url
16             System.out.println(new URL(doc.getSourceUrl(), href));
17         }
18     }
19
20     ...
```

Exercise 4: Solution (continued)



The screenshot shows the GATE application window. The menu bar includes File, Options, Tools, and Help. The left sidebar displays a tree view of resources under the GATE icon, including Applications, Language Resources, This is home, Processing Resources, and Databases. The main area is titled 'Messages' and contains a log of startup information and a list of URLs. At the bottom left, it indicates 'Total elements : 291'. At the bottom right, there are navigation icons.

File Options Tools Help

GATE

- Applications
- Language Resources
 - This is home
- Processing Resources
- Databases

Messages

GATE 6.2-snapshot build 3931 started at Thu May 12 01:41:00 BST 2011 and using java 1.6.0_20 Sun Microsystems Inc. on Linux i386 2.6.31-23-generic.

<http://www.fsf.org/>
<http://www.gnu.org/licenses/>
<http://creativecommons.org/licenses/by-nc-sa/3.0/>
<http://creativecommons.org/licenses/by-nc-sa/3.0/>
<http://www.shef.ac.uk/dcs/>
<http://www.gate.ac.uk/>
<http://www.gate.ac.uk/partners.html>
<http://www.ontotext.com/>
<http://www.gate.ac.uk/g8/auth/login?targetUri=%2F>
<http://www.gate.ac.uk/gatewiki/cow/doc/gatewiki.html>
<http://www.ir-facility.org/>
<http://www.gate.ac.uk/g8/page/print/2/index.html>
<http://www.gate.ac.uk/family/>
<http://www.gate.ac.uk/family/embedded.html>
<http://www.gate.ac.uk/family/developer.html>
<http://www.gate.ac.uk/family/process.html>
<http://www.gate.ac.uk/teamware>
<http://www.gate.ac.uk/family/mimir.html>
<http://gatecloud.net/>
<http://www.gate.ac.uk/family/training.html>
<http://www.gate.ac.uk/gatewiki/cow/>
<http://www.gate.ac.uk/customisation/>
<http://www.gate.ac.uk/family/training.html>
<http://www.gate.ac.uk/support.html>
<http://www.gate.ac.uk/biz/customers.html>
<http://www.gate.ac.uk/biz/usps.html>
<http://www.gate.ac.uk/customisation/>
<http://www.gate.ac.uk/download/>
<http://www.gate.ac.uk/mail/>

Total elements : 291

Exercise 5: Solution

Additions to the solution of Exercise 2

```
1      ...
2
3      //get the root plugins dir
4      File pluginsDir = Gate.getPluginsHome();
5      //Let's load the ANNIE plugin
6      File aPluginDir = new File(pluginsDir, "ANNIE");
7      //load the plugin.
8      Gate.getCreoleRegister().registerDirectories(
9          aPluginDir.toURI().toURL());
10
11     // create tokenizer
12     LanguageAnalyser pr = (LanguageAnalyser)
13         Factory.createResource(
14             "gate.creole.tokeniser.DefaultTokeniser");
15
16     pr.setDocument(doc); // set the document
17     pr.setCorpus(null); // set the corpus to null
18     pr.execute(); // execute the PR
19
20     ...
```

Exercise 5: Solution (continued)

The screenshot displays the GATE software interface. On the left is a project tree with folders for Applications, Language Resources, Processing Resources, and Datasets. The main window shows a document editor with a text document open. A 'Token' window is overlaid on the text, displaying a table of tokenization parameters. The table has columns for 'kind', 'value', and a red 'X' icon. The parameters are: kind: word, length: 9, orth: upperInitial, string: Developer. Below the table is a button labeled 'Open Search & Annotate tool'. On the right side of the interface, there are tabs for 'Annotation Sets', 'Annotations List', 'Annotations Stack', 'Co-reference Editor', and 'Text'. A 'Messages' pane at the top shows 'This is home'. A 'Properties' pane at the bottom left shows fields for 'MimeType', 'date', and 'gate.SourceURL'. A 'Document Editor' pane at the bottom right shows 'Initialisation Parameters'.

| kind | value | X |
|--------|--------------|---|
| word | word | X |
| length | 9 | X |
| orth | upperInitial | X |
| string | Developer | X |

Exercise 6: Solution

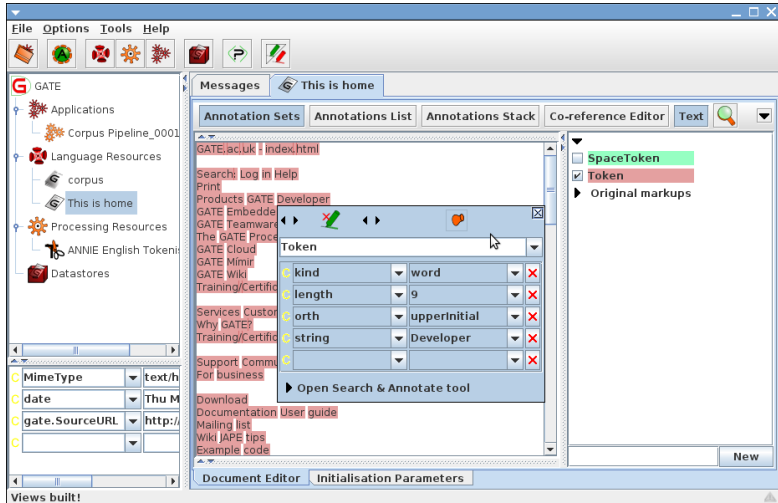
Additions to the solution of Exercise 2

```
1      ...
2
3      // get the root plugins dir
4      File pluginsDir = Gate.getPluginsHome();
5      // Let's load the ANNIE plugin
6      File aPluginDir = new File(pluginsDir, "ANNIE");
7      // load the plugin.
8      Gate.getCreoleRegister().registerDirectories(
9          aPluginDir.toURI().toURL());
10
11     // create tokenizer
12     LanguageAnalyser pr = (LanguageAnalyser)
13         Factory.createResource(
14         "gate.creole.tokeniser.DefaultTokeniser");
```


Exercise 6: Solution (Continued..)

```
16
17 // create serialAnalyzerController
18 SerialAnalyserController controller =
19     (SerialAnalyserController) Factory.createResource(
20         "gate.creole.SerialAnalyserController");
21 // add pr to the corpus
22 controller.add(pr);
23
24 // create a corpus
25 Corpus corpus = Factory.newCorpus("corpus");
26 corpus.add(doc); // add document to the corpus
27 controller.setCorpus(corpus); // set corpus
28 controller.execute(); // execute the corpus
29
30 ...
```

Exercise 6: Solution (continued)



The screenshot shows the GATE software interface. On the left is a project tree with folders like Applications, Corpus Pipeline_0001, Language Resources, Processing Resources, and Datasets. The main window displays a document editor with a web page from GATE.ac.uk. A 'Token' annotation tool is overlaid on the text 'Developer', showing the following details:

| Property | Value | Action |
|----------|--------------|--------|
| kind | word | X |
| length | 9 | X |
| orth | upperinitial | X |
| string | Developer | X |

Below the table is a button labeled 'Open Search & Annotate tool'. The right sidebar shows a list of annotation sets, with 'Token' checked and highlighted in red. Other sets include 'SpaceToken' and 'Original markups'. The bottom status bar indicates 'Views built!'.