



Introduction to GATE

Andrés García-Silva, Elena Montiel-Ponsoda

Ontology Engineering Group
Facultad de Informática
Universidad Politécnica de Madrid
Campus de Montegancedo s/n
28660 Boadilla del Monte, Madrid, Spain

<http://www.oeg-upm.net>

emontiel@fi.upm.es
hagarcia@delicias.dia.fi.upm.es
Phone: 34.91.3363670
Fax: 34.91.3524819

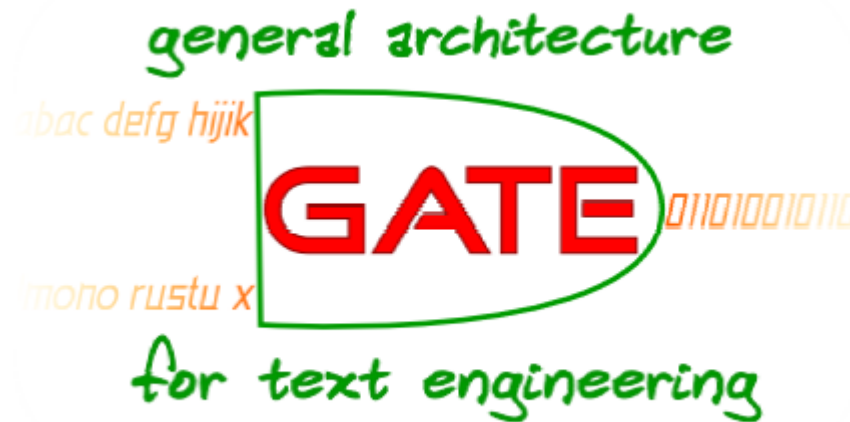
Date: December 10th, 2009

- Introduction to GATE
- ANNIE
 - ANNIE Demo
- Jape Rules
 - Jape Rules Demo
- Machine Learning
- Ontology support
 - Semantic Annotation
 - Ontology Learning
 - Semantic Annotation Demo
- LSPs Demo

- **Introduction to GATE**
- ANNIE
 - ANNIE Demo
- Jape Rules
 - Jape Rules Demo
- Machine Learning
- Ontology support
 - Semantic Annotation
 - Ontology Learning
 - Semantic Annotation Demo
- LSPs Demo

Introduction to GATE

- **GATE** stands for: **G**eneral **A**rchitecture for **T**ext **E**ngineering

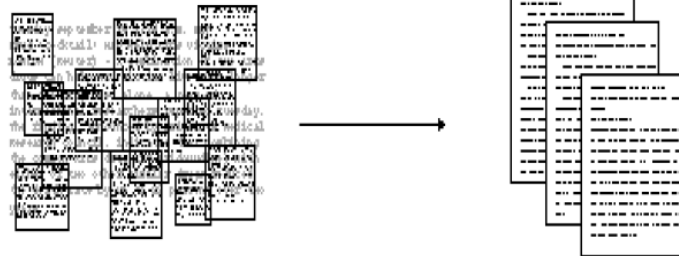


- It is an infrastructure for developing and deploying software components that process human language
- GATE has been in development at the University of Sheffield since 1995

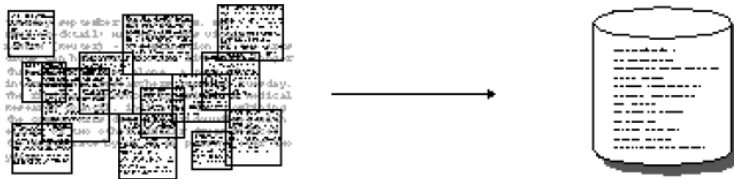
Introduction to GATE

- **GATE** was originally developed as an Information Extraction system (IE)
 - What does an IE system do?
 - Which is the difference between IE and IR (Information Retrieval)?

IR pulls **documents** from large text collections (usually the Web) in response to specific keywords or queries. You analyse the **documents**.

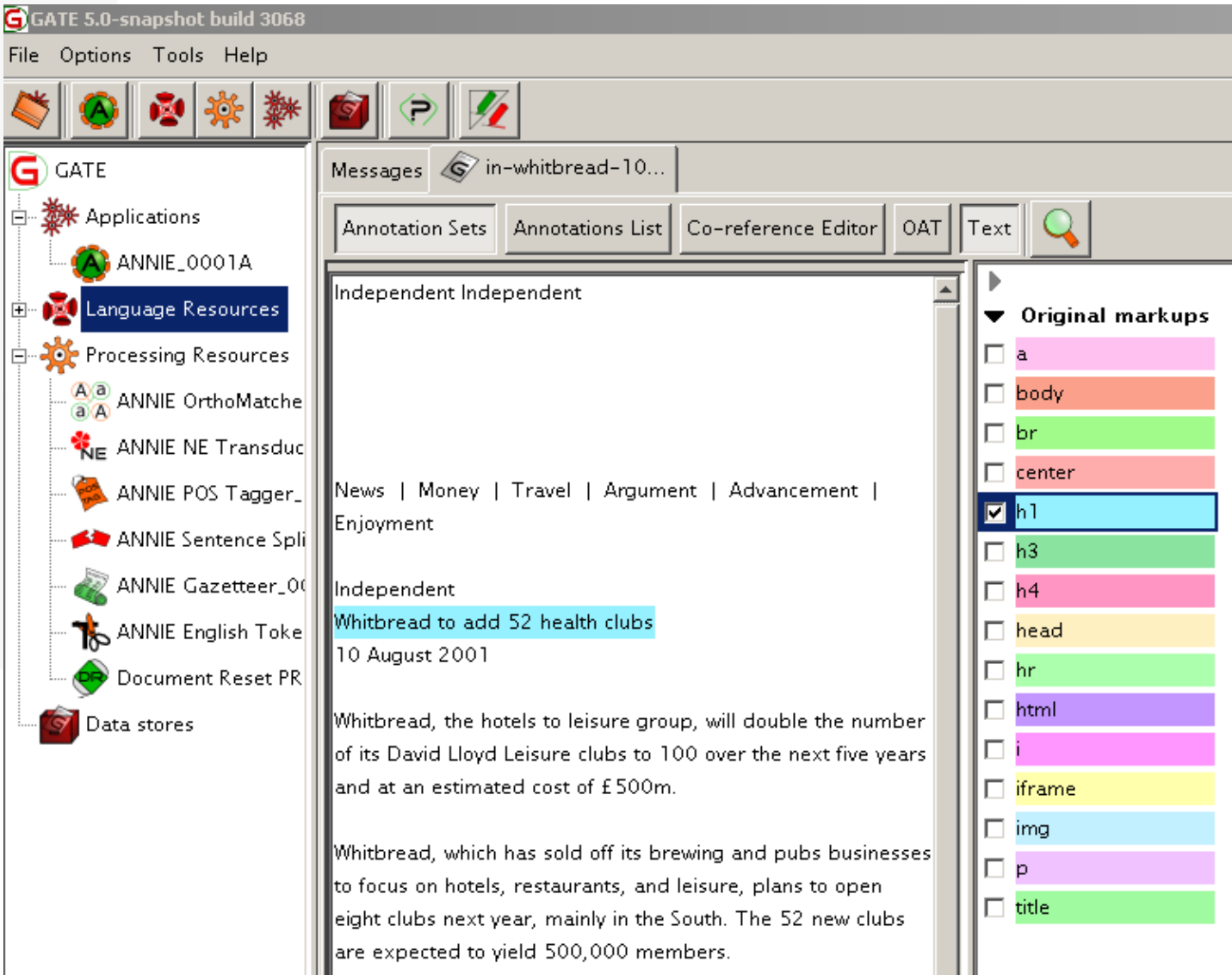


IE pulls **facts** and **structured information** from the content of large text collections. You analyse the **facts**.



- GATE integrates 3 types of resources:
 - **Language Resources (LRs) or DATA**, e.g. corpora, gazetteer, ontologies
 - **Processing Resources (PRs) or ALGORITHMS**, e.g. lemmatisers, parsers, taggers
 - **Visual Resources (VRs) or GUI**, i.e. visualization and editing components
- Algorithms + Data + GUI = **APPLICATIONS**

Introduction to GATE



general architecture
for text engineering






What are annotations?

- **Annotations** are linguistic information or metadata associated to the content in documents
- **Annotations** consist of **features** with a **name** and a **value**. E.g.: dog **TOKEN** **kind=word**
pos=NN (Part of speech)
length=3

Text: **Cyndi savored the soup.**

Nodes: | 0... | 5... | 10.. | 15.. | 20

Annotation spans: 

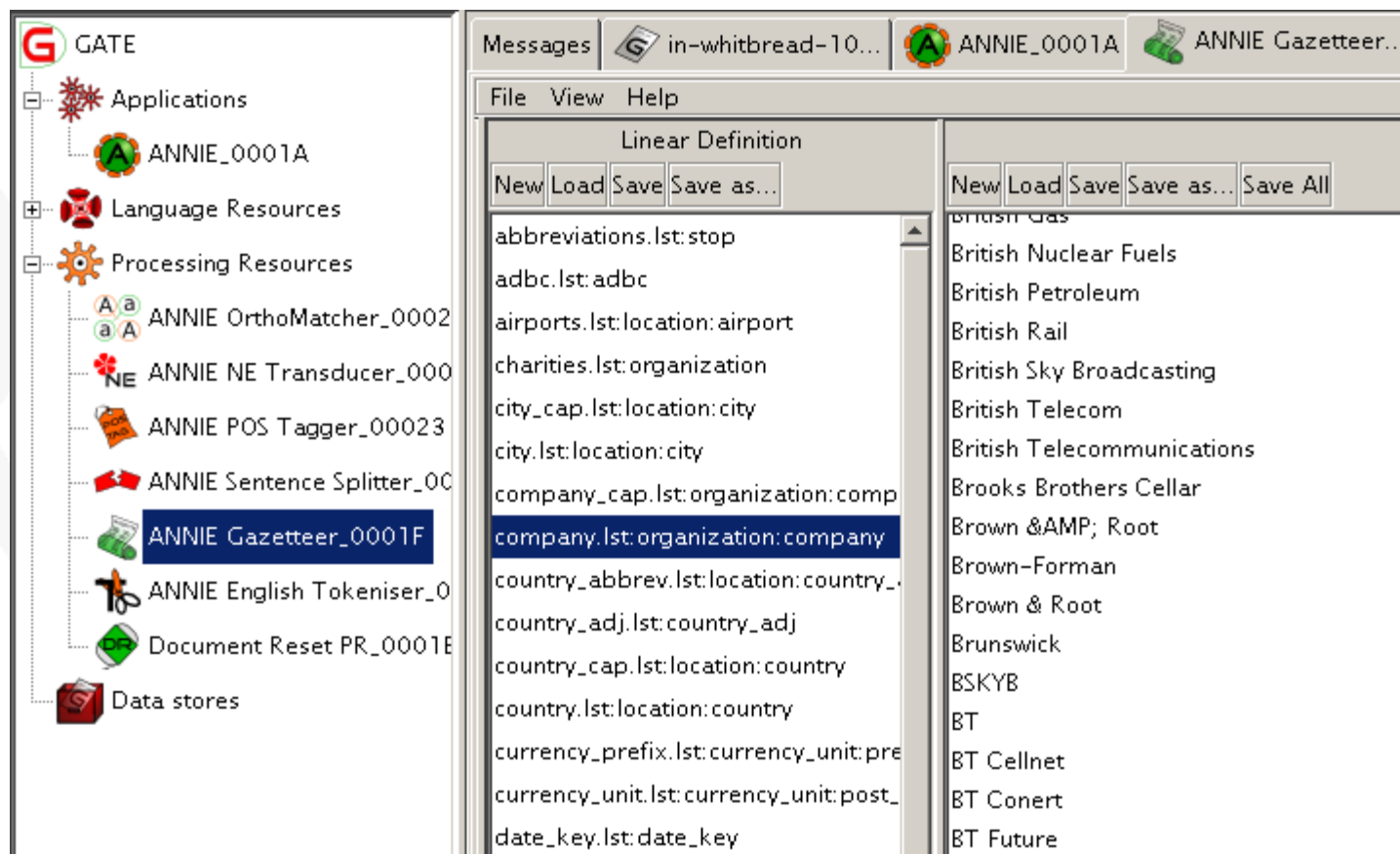
Annotation spans: 

Annotation descriptions:	Id	Type	Start	End	Features
	1	token	0	5	pos=NP
	2	token	6	13	pos=VBD
	3	token	14	17	pos=DT
	4	token	18	22	pos=NN
	5	token	22	23	
	6	name	0	5	type=person
	7	sentence	0	23	

- **NP: proper noun**
- **VBD: verb past tense**
- **DT: determiner**
- **NN: noun-singular**

What are Lexical Resources?

- **Lexical Resources** can be corpora, gazetteer, ontologies...
- Gazetteers are plain lists of entities grouped into categories.
 - E.g.: Location - European capital cities: Paris, Madrid, Berlin, etc.
 - (**Gazetteer PR** to conduct Named Entity or key phrase Lookup)



What are Processing Resources?

- PRs take as input LRs and produce as output **annotations** on those LRs
- E.g.: tokeniser, pos tagger, gazetteer, orthomatcher...

GATE

Applications

- ANNIE_0001A
- Language Resources
- Processing Resources
 - ANNIE OrthoMatcher_0002
 - ANNIE NE Transducer_0001
 - ANNIE POS Tagger_00023
 - ANNIE Sentence Splitter_0001
 - ANNIE Gazetteer_0001F
 - ANNIE English Tokeniser_0001
 - Document Reset PR_0001E
- Data stores

Messages in-whitbread-10... ANNIE_0001A

Annotation Sets Annotations List Co-reference Editor OAT Text

Independent Independent

News | Money | Travel | Argument | Advancement | Enjoyment

Independent

Whitbread to add 52 health clubs

Lookup

majorType	organization	minorType	company
organization	Whitbread	company	Whitbread

Open Search & Annotate tool

Number of its
s and at an

nesses to focus
clubs next year,
field 500,000

- ☐ Date
- ☐ FirstPerson
- ☐ JobTitle
- ☐ Location
- ☒ Lookup
- ☐ Money
- ☐ Organization
- ☐ Percent
- ☐ Person
- ☐ Sentence
- ☐ SpaceToken
- ☐ Split
- ☐ Token
- ☐ Unknown
- ☒ Original markups
- ☐ a
- ☐ body

- **Applications** execute PRs in a particular order
- Currently only sequential, or pipeline, execution is supported:

The screenshot displays the GATE application configuration window. It features two main panels for processing resources, a corpus selection dropdown, a status message, and a table for processing parameters.

Loaded Processing resources

Name	Type
ANNIE Gazetteer_00119	ANNIE Gazetteer

Selected Processing resources

!	Name	Type
	Document Reset PR_0003F	Document Reset PR
	ANNIE English Tokeniser_00040	ANNIE English Tokeniser
	ANNIE Sentence Splitter_00043	ANNIE Sentence Splitter
	ANNIE POS Tagger_00046	ANNIE POS Tagger
	GATE Morphological analyser_00047	GATE Morphological analyser
	Noun Phrase Chunker_0006E	Noun Phrase Chunker
	Flexible Gazetteer_0011A	Flexible Gazetteer
	Jape Transducer_0011C	Jape Transducer

Corpus: GATE corpus_00028

The **corpus** and **document** parameters are not available as they are automatically set by the controller!

No selected processing resource

Name	Type	Required	Value

Run this application

- Introduction to GATE
- **ANNIE**
 - ANNIE Demo
- Jape Rules
 - Jape Rules Demo
- Machine Learning
- Ontology support
 - Semantic Annotation
 - Ontology Learning
 - Ontology Learning Demo
- LSPs Demo

- ANNIE is GATE's basic building block
- ANNIE consists of several PRs that perform annotations (which range from basic annotations to more complex ones)
- The PRs in ANNIE are enough to create an NLP application

The screenshot displays the GATE software interface with three numbered annotations:

- 1**: Points to the 'File' menu in the top toolbar.
- 2**: Points to the 'Applications' folder in the left-hand tree view.
- 3**: Points to the 'Selected Processing resources' table on the right.

The 'Selected Processing resources' table lists the following resources:

!	Name	Type
	Document Reset PR_0005A	Document Reset PR
	ANNIE English Tokeniser_0005B	ANNIE English Tokeniser
	ANNIE Gazetteer_0005E	ANNIE Gazetteer
	ANNIE Sentence Splitter_0005F	ANNIE Sentence Splitter
	ANNIE POS Tagger_00062	ANNIE POS Tagger
	ANNIE NE Transducer_00063	ANNIE NE Transducer
	ANNIE OrthoMatcher_00064	ANNIE OrthoMatcher

- **Document Reset**: clears existing annotations in the document (no annotation is embedded in the document)
- **ANNIE English Tokeniser**: splits the text into tokens
- **ANNIE Gazetteer**: wordlists grouped in different categories to perform NE or Key Phrase Lookups
- **ANNIE Sentence Splitter**: assigns annotations to sentence boundaries
- **ANNIE PoS Tagger**: assigns PoS tags to tokens
- **ANNIE Orthomatcher**: adds identity relations between NE found by the NE Transducer to perform co-reference

Ryanair announced yesterday that it will make Shannon its next European base, expanding its route network to 14 in an investment worth around €180m. The airline says it will deliver 1.3 million passengers in the first year of the agreement, rising to two million by the fifth year.

- **ANNIE NE Transducers**: executes JAPE rules to create complex annotations based on the results of the previous PRs

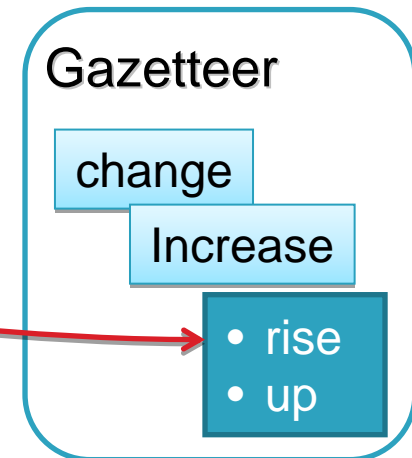
- Introduction to GATE
- ANNIE
 - **ANNIE Demo**
- Jape Rules
 - Jape Rules Demo
- Machine Learning
- Ontology support
 - Semantic Annotation
 - Ontology Learning
 - Semantic annotation Demo
- LSPs Demo

- Introduction to GATE
- ANNIE
 - ANNIE Demo
- **Jape Rules**
 - Jape Rules Demo
- Machine Learning
- Ontology support
 - Semantic Annotation
 - Ontology Learning
 - Semantic Annotation Demo
- LSPs Demo

- Gazetteer lists are designed for annotating simple, regular features
 - Some flexibility is provided by matching
 - Word roots
 - Whole/part words

HSBC shares **rose** after strong cashcall

New annotation
Lookup (
majorType=**change**,
minorType=**increase**
)



recognizing e-mail addresses using just a gazetteer
would be impossible

- JAPE provides pattern matching in GATE

LHS

Pattern to
match in the
text



RHS

New
annotations
(and opt.
features)

“contact information: JohnMalkovich@gmail.com”

```
{Token.kind == word}*  
{Token.string == "@"}  
{Token.kind == word}*  
{Token.string == "."}  
{Token.kind == word}
```

:email

New annotation

Email (
rule = “emailaddress”)

Email =
emailaddress”}

- JAPE rules combine to create a phase
- Phases combine to create a grammar

Phase: Email

Input: Token SpaceToken

Options: control = **appelt**

Macro: WORD_OR_NUMBER

```
(  
{Token.kind == word}|{Token.kind == number})  
)
```

Rule: emailaddress

Priority: 50

```
(  
(WORD_OR_NUMBER)+  
{Token.string == "."}(WORD_OR_NUMBER)+*  
{Token.string == "@"}  
(WORD_OR_NUMBER)+  
{Token.string == "."}(WORD_OR_NUMBER)+*  
)
```

:email -->

:email.**E****Mail** = {**r****ule** = "emailaddress"}

LHS

```
{Token.kind == word}*  
{Token.string == "@"}  
{Token.kind == word}*  
{Token.string == "."}  
{Token.kind == word}  
:email
```

- LHS:
 - Annotations
 - Optionally features and their values
- Any annotation to be used must be included in the input header
- Each annotation is enclosed in curly braces
- Annotations may be combined using traditional Kleene operators: | * + ?
- Each pattern to be matched is enclosed in round brackets and can have a label attached

LHS

```
{Token.kind == word}*  
{Token.string == "@"}  
{Token.kind == word}*  
{Token.string == "."}  
{Token.kind == word}  
:email
```

- JAPE rules can access annotation features
- Features can be compared with ==, !=, >, <, =~, !~, ==~ and !=~
- Ranges can be specified({Token})[1,3] or ({Token})[3]
- Contextual information can be specified in the same way, but has no label
- Contextual information will be consumed by the rule
({Annotation1})
({Annotation2}):match
({Annotation3})

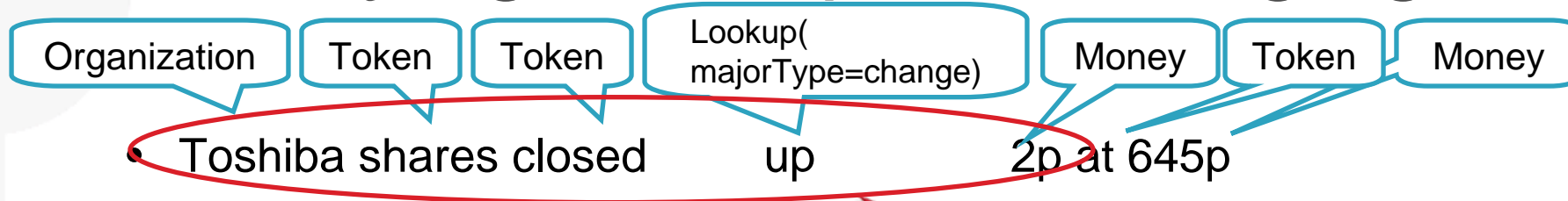
- LHS and RHS are separated by -->
- Label matches that on the LHS
- Annotation to be created follows the label

(Annotation1):**label** -->
:label.NewAnnotation =
 {feature1 = value1,
 feature2 = value2}

RHS

:email.Email =
 {rule = "emailaddress"}

• Identifying share price changing:



New annotation

ShareChange(
rule = "ShareChange")

Phase: Shares

Input: Token Organization Lookup Money
Percent

Options: control = appelt

Rule:ShareChange

```
(
    {Organization}
    ({Token})[0,3]
    {Lookup.majorType=="change"}
    ({Token})[0,3]
    ({Money})|{Percent})
```

):**change** -->

```
:change.ShareChange = {rule =
"ShareChange"}
```

- Introduction to GATE
- ANNIE
 - ANNIE Demo
- Jape Rules
 - **Jape Rules Demo**
- Machine Learning
- Ontology support
 - Semantic Annotation
 - Ontology Learning
 - Semantic Annotation Demo
- LSPs Demo

- Corpus creation
- Default ANNIE application
- Gazetteer recognizing changes
- Run ANNIE app
- Jape transducer loading our jape rule.
- Run ANNIE App to review the results.

- The simple RHS of a JAPE rule can only add simple annotations and features
- Feature values are hard coded or can be copied from annotations matched by the LHS
- You may need more complex processing
 - Removing temporary annotations
 - Building complex features
- RHS of a rule can consist of arbitrary Java code

RHS

```
:email.Email =  
{rule = "emailaddress"}
```



- `public void doit(Document doc, Map bindings, AnnotationSet annotations, AnnotationSet inputAS, AnnotationSet outputAS, Ontology ontology) throws JapeException`
- Each labeled section of the LHS results in an Annotation Set that can be retrieved from the bindings map
 - `AnnotationSet set = (AnnotationSet)bindings.get("labelname");`
- All features of an annotation are stored in a map
 - `FeatureMap map = annotation.getFeatures()`
- Each feature is accessed by name
 - `Object obj = map.get("featurename")`
- New annotations should always be created in the outputAS
 - `outputAS.add(start,end,label,features)`

Rule:SC1_1

```
(
(LIST):subclass
{Lookup.minorType == be}
{{Token.category == DT}}?
{Lookup.majorType == CN}
({NounChunk}):superclass
)
-->
:superclass.Superclass = {rule="SC1_1"},
{
    // "subclass" matches LHS label
    List annList = new ArrayList((AnnotationSet)bindings.get("subclass"));
    //sort the list by offset
    Collections.sort(annList, new OffsetComparator());
    //iterate through the matched annotations
    for(int i = 0; i < annList.size(); i++)
    {
        Annotation anAnn = (Annotation)annList.get(i);
        // check that the new annotation is a NounChunk
        if ((anAnn.getType().equals("NounChunk")) )
        {
            FeatureMap features = Factory.newFeatureMap();
            // change this for a different rule name"
            features.put("rule", "SC1_1");
            // change "Subclass" for a different annotation name
            annotations.add(anAnn.getStartNode(), anAnn.getEndNode(), "Subclass", features);
        }
    }
}
```

LSP: (NP<subclass>)* and] NP<subclass> be [CN] NP<superclass>

Example: **Odometry, speedometry and GPS are types of sensors.**

NP: Noun Phrase
CN: Class Name

- Introduction to GATE
- ANNIE
 - ANNIE Demo
- Jape Rules
 - Jape Rules Demo
- **Machine Learning**
- Ontology support
 - Semantic Annotation
 - Ontology Learning
 - Semantic Annotation Demo
- LSPs Demo

Existing classified instances



- Typical supervised machine learning process (Classification):
 - We have items/instances and attributes/features
 - Car id=0001, brand=Ford, price=20000€, color=blue
 - Car id=0002, brand=Lexus, price=50000€, color=silver
 - We have classes: Car id=0001 -> Category = Sport
Car id = 0002 -> Category = Luxury
 - Machine learning algorithm learns a relationship between the attributes and classes

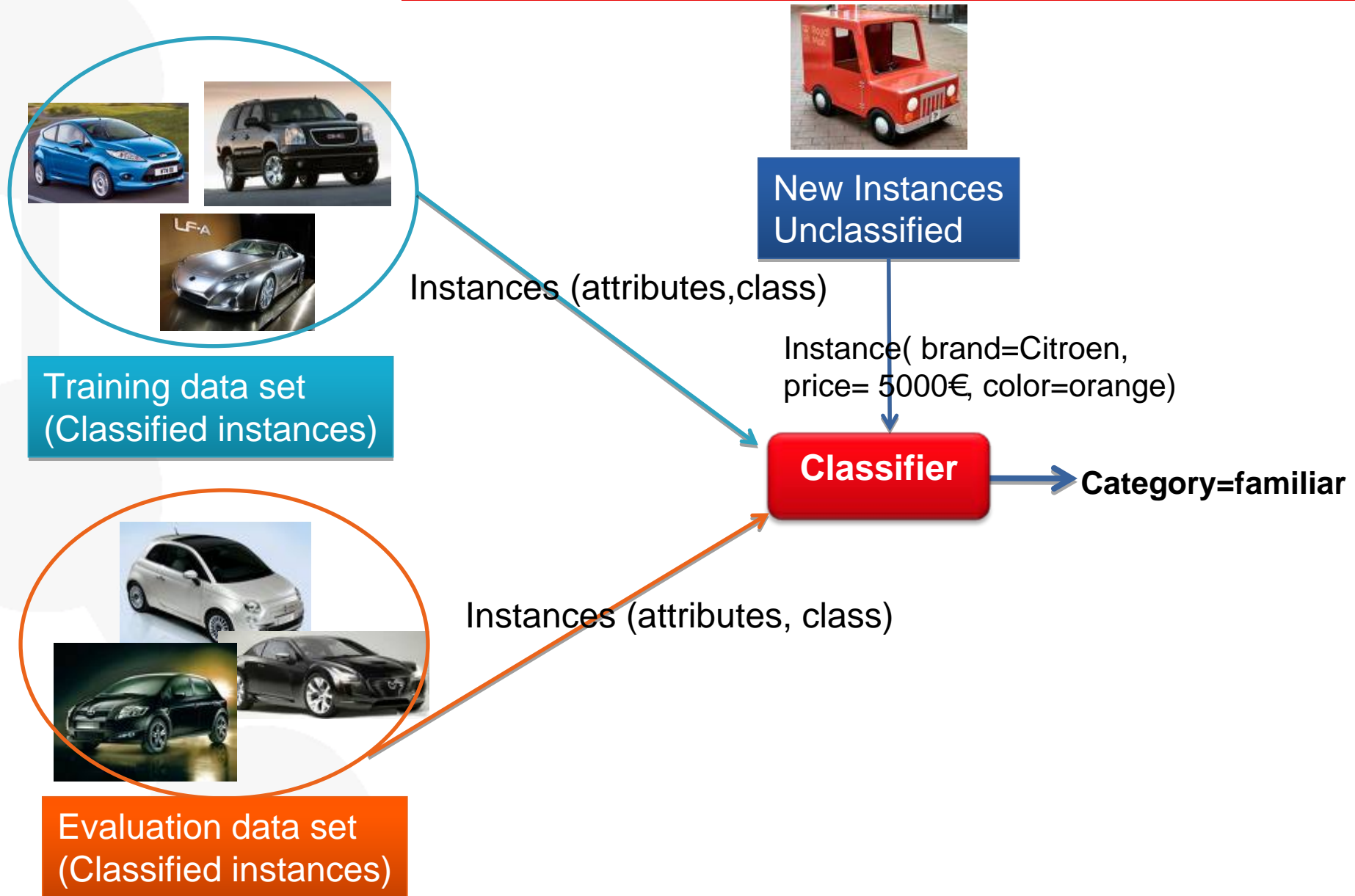
New instance



- Car id=003, brand=Citroen, price= 5000€, color=orange, category=?

Machine Learning

Supervised classification





Token.String="California"
Lookup.majorType="Location"

Lookup.majorType
="Job"

Entity.type=?

Token.category="Verb"

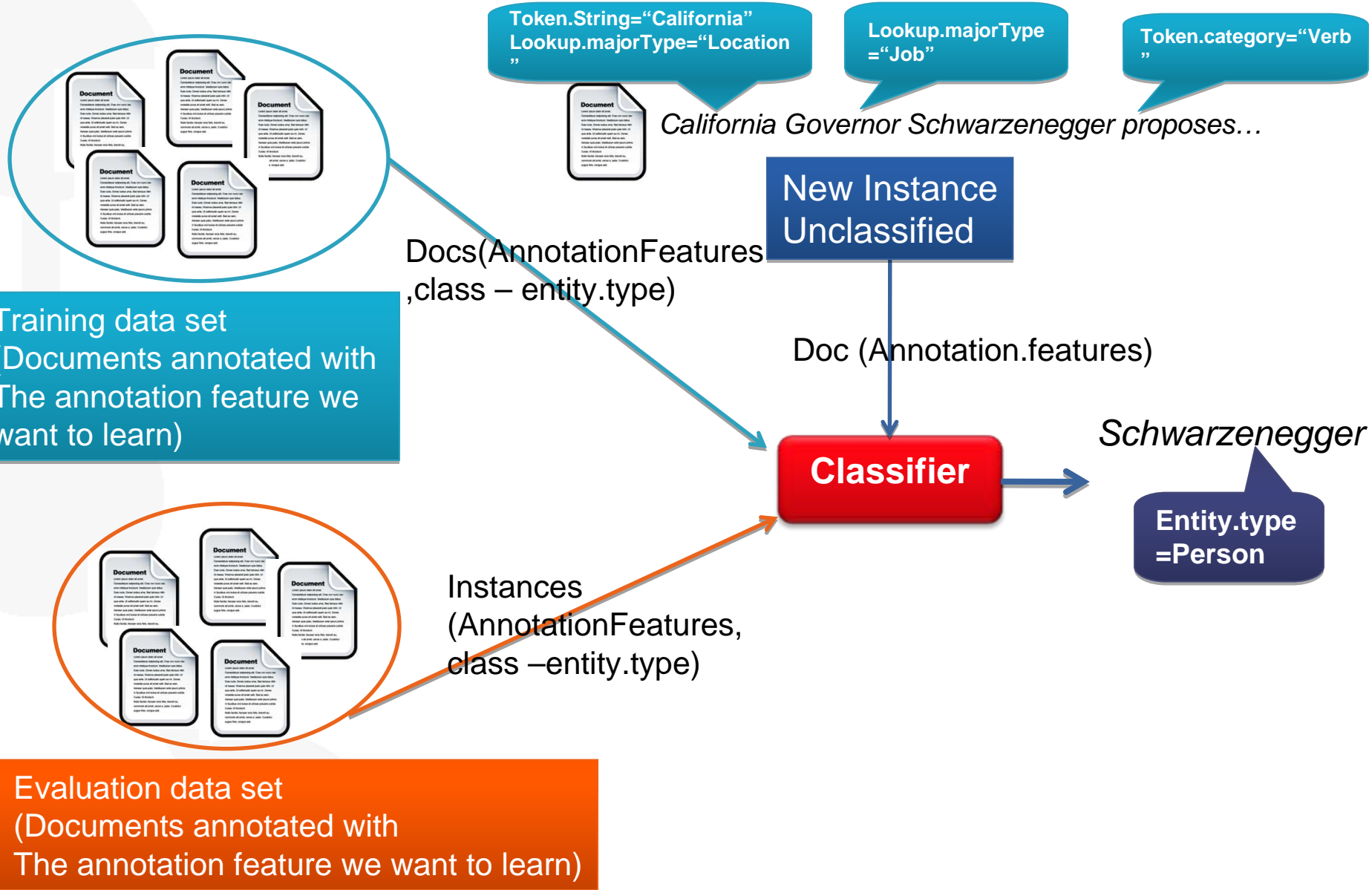
"California Governor Schwarzenegger proposes deep cuts"

Car example

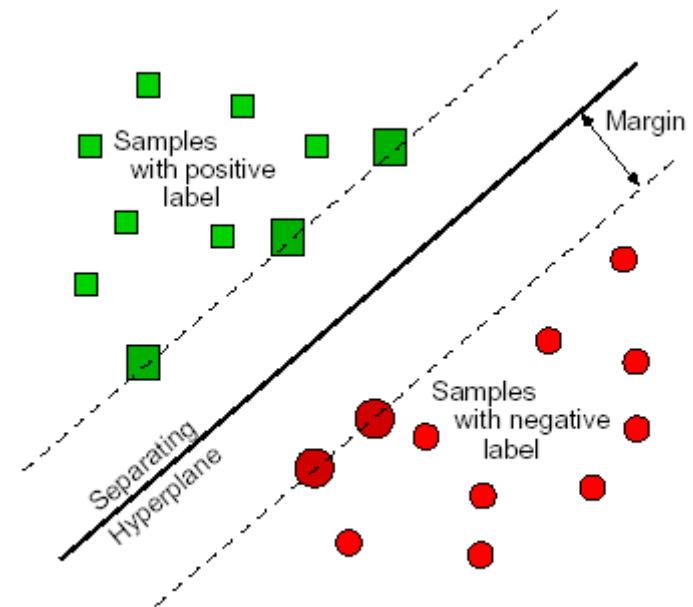
- Instances
 - Cars
- Attributes
 - Color, Brand, Prize
- Class
 - Category

Annotations in Gate

- Instances: Annotations
 - Token, Lookup, Organization
- Attributes: Annotation Features
 - Token.string, Token. Root, Token.category(POS), Lookup.majorType, etc.
- Class: Annotation Feature
 - Entity.type = Person?



- Attempt to find a hyperplane that separates data
- Goal: maximize margin separating two classes
- Wider margin = greater generalisation

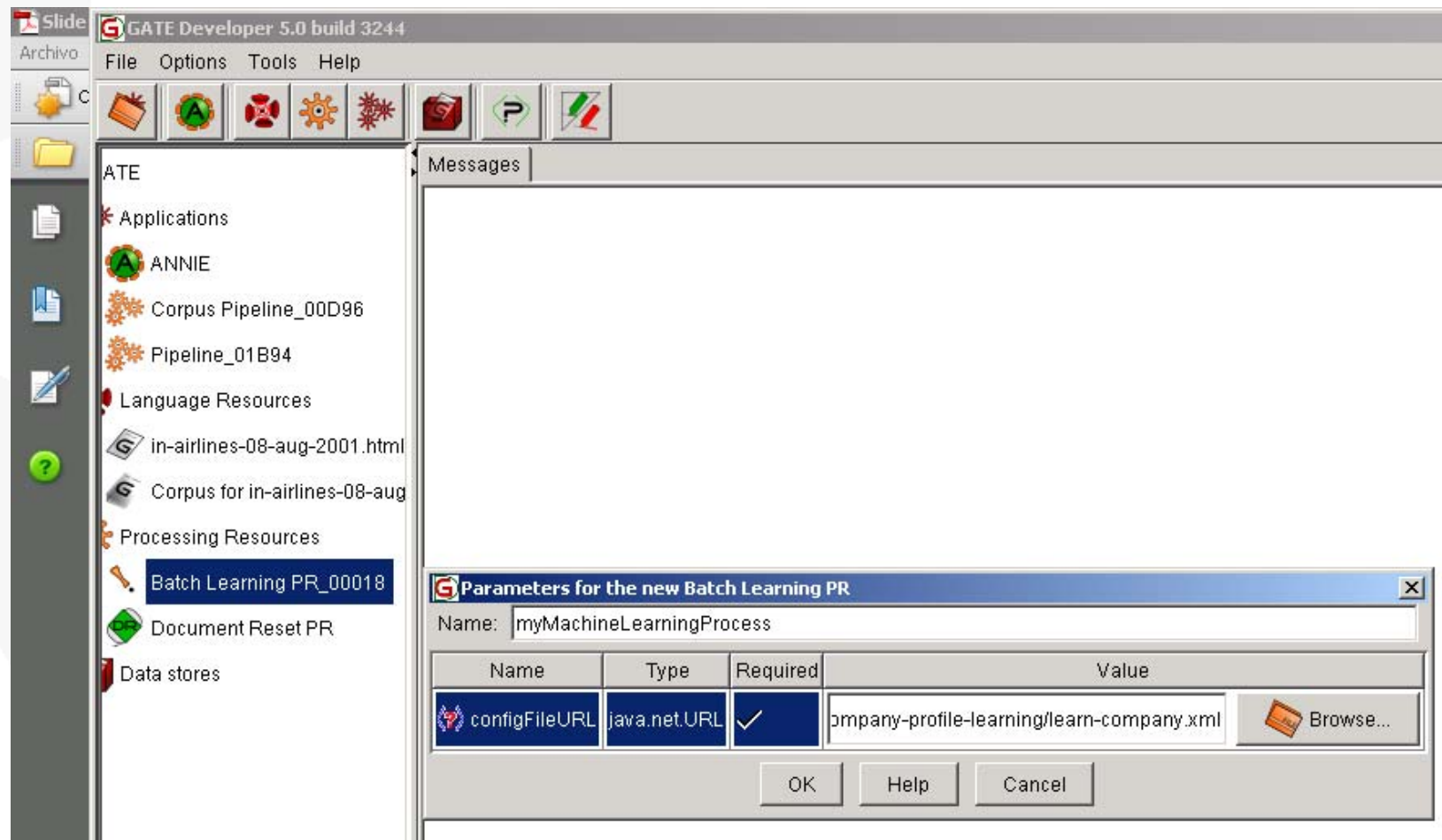


```
<?xml version="1.0" encoding="UTF-8"?>
<ML-CONFIG>
  <EVALUATION method="split" runs="2"
ratio="0.66"/>
  <ENGINE nickname="SVM"
implementationName="SVMLibSvmJava" options="
-c 0.7 -t 0 -m 100 -tau 1  "/>

<DATASET>
  <INSTANCE-TYPE>Token</INSTANCE-TYPE>
  <ATTRIBUTELIST>
    <NAME>Form</NAME>
    <SEMTYPE>NOMINAL</SEMTYPE>
    <TYPE>Token</TYPE>
    <FEATURE>string</FEATURE>
    <RANGE from="-5" to="5"/>
  </ATTRIBUTELIST>
  <ATTRIBUTELIST>
    <NAME>category</NAME>
    <SEMTYPE>NOMINAL</SEMTYPE>
    <TYPE>Token</TYPE>
    <FEATURE>category</FEATURE>
    <RANGE from="-5" to="5"/>
  </ATTRIBUTELIST>
  <ATTRIBUTELIST>
    <NAME>Tokenkind</NAME>
    <SEMTYPE>NOMINAL</SEMTYPE>
    <TYPE>Token</TYPE>
    <FEATURE>kind</FEATURE>
    <RANGE from="-5" to="5"/>
  </ATTRIBUTELIST>

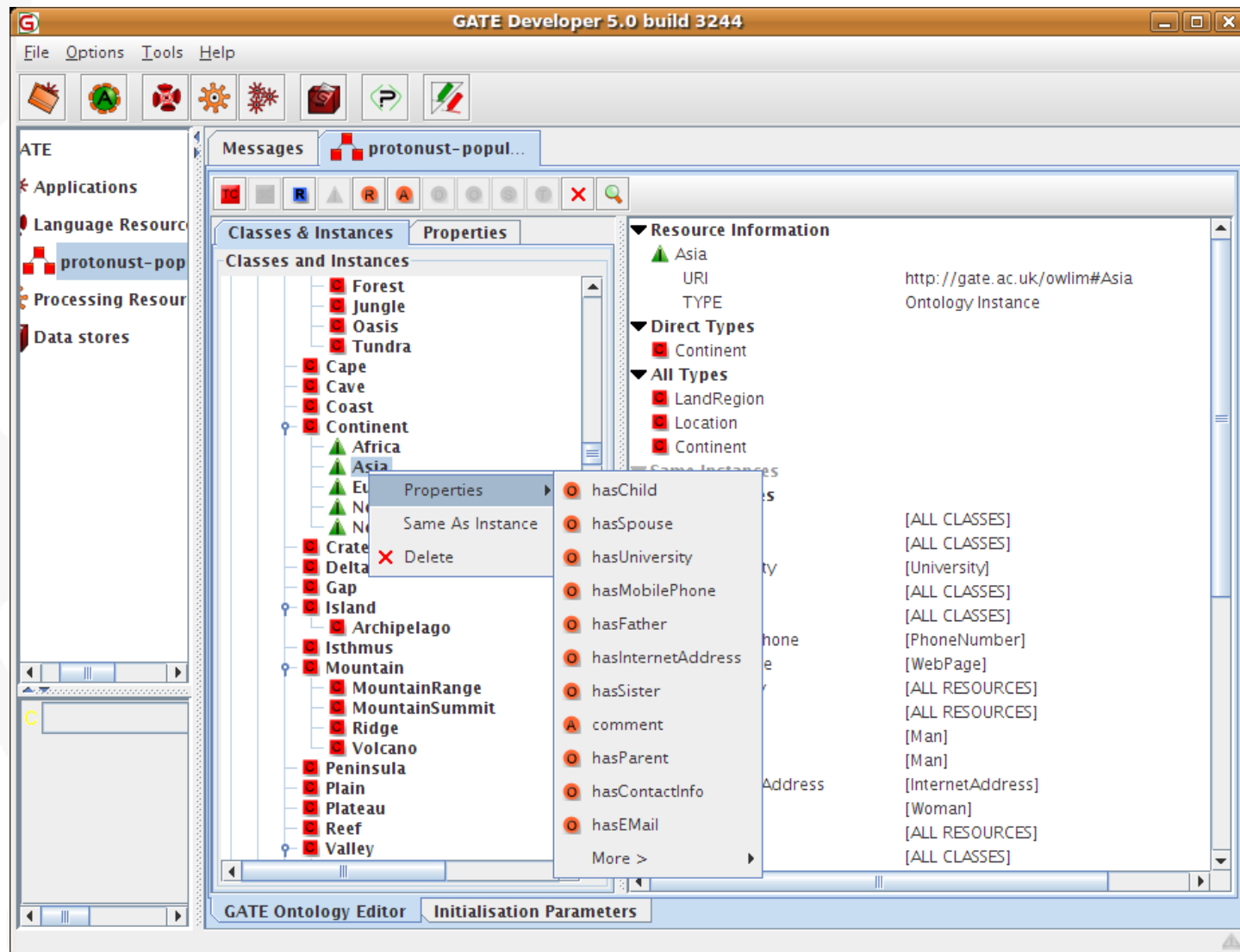
```

```
<ATTRIBUTELIST>
  <NAME>Lemma</NAME>
  <SEMTYPE>NOMINAL</SEMTYPE>
  <TYPE>Token</TYPE>
  <FEATURE>root</FEATURE>
  <RANGE from="-5" to="5"/>
</ATTRIBUTELIST>
<ATTRIBUTELIST>
  <NAME>Gaz</NAME>
  <SEMTYPE>NOMINAL</SEMTYPE>
  <TYPE>Lookup</TYPE>
  <FEATURE>majorType</FEATURE>
  <RANGE from="-5" to="5"/>
</ATTRIBUTELIST>
<ATTRIBUTE>
  <NAME>ENTITY</NAME>
  <SEMTYPE>NOMINAL</SEMTYPE>
  <TYPE>Entity</TYPE>
  <FEATURE>type</FEATURE>
  <POSITION>0</POSITION>
</ATTRIBUTE>
<ATTRIBUTE>
  <NAME>Class</NAME>
  <SEMTYPE>NOMINAL</SEMTYPE>
  <TYPE>Mention</TYPE>
  <FEATURE>class</FEATURE>
  <POSITION>0</POSITION>
  <CLASS/>
</ATTRIBUTE>
</DATASET>
</ML_CONFIG>
```

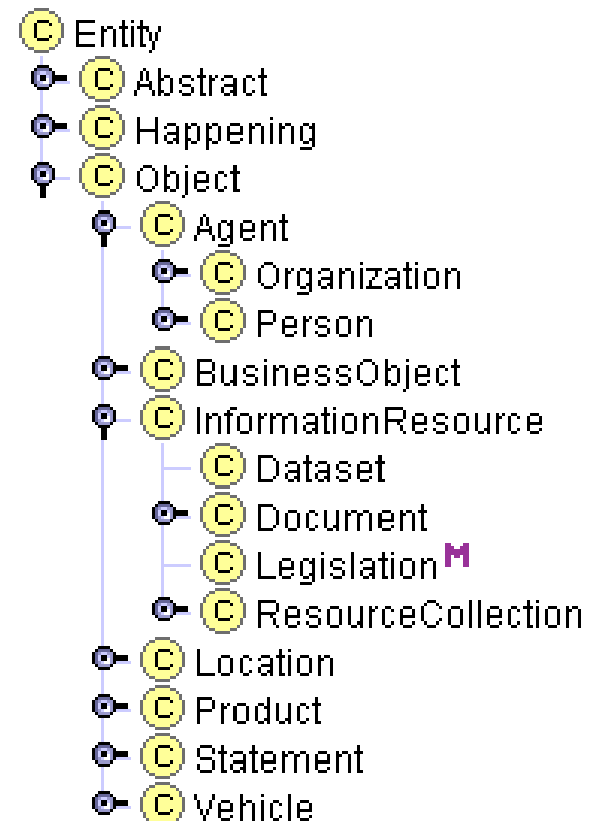


- Introduction to GATE
- ANNIE
 - ANNIE Demo
- Jape Rules
 - Jape Rules Demo
- Machine Learning
- **Ontology support**
 - Semantic Annotation
 - Ontology Learning
 - Semantic Annotation Demo
- LSPs Demo

- Comes with one concrete implementation
 - preinstalled: Sesame
- Comes with several tools:
 - Ontology Visualizer/Editor
 - OntoRootGazetteer
 - Ontology support in JAPE
- Fast in memory repository, scales to millions of statements (depending on RAM)
- Ontology API: JAPE RHS can access Ontology object
 - Useful to ontology population and ontology Learning.



- A light-weight upper-level ontology:
 - 250 NE classes;
 - 100 relations and attributes;
 - 200.000 entity descriptions;
 - covers mostly **NE classes**, and ignores general concepts;
 - includes classes representing **lexical resources**



```
:London a City ;  
:Company a :Organization .  
XYZ-02FA a :Company ;  
    rdfs:label "XYZ"@en ;  
    :basedIn :London-UK  
XYZ-98 a :Company ;  
    rdfs:label "XYZ"@en ;  
    :basedIn :Boston-US  
...
```

XYZ was
established on
03 November 1978
in London. The
company opened a
plant in
Bulgaria in ..

GATE Developer 5.0 build 3244

File Options Tools Help

ATE

- Applications
- Language Resources
 - gu-ECB-03-aug
 - corpus
 - proton
- Processing Resources
- Data stores
 - file:/home/johan

Messages proton gu-ECB-03-aug-2...

Annotation Sets Annotations List Co-reference Editor OAT Text

Ontology Tree(s) Options

proton

- ☒ JobPosition
- ☒ ContactInformation
- ☒ InformationResource
- ☒ Number
- ☒ Organization
- ☒ PoliticalEntity
- ☒ Team
- ☒ Charity
- ☒ EducationalOrganization
- ☒ ReligiousOrganization
- ☒ SportOrganization
- ☒ GovernmentOrganization
- ☒ StockExchange
- ☒ Division
- ☒ ResearchOrganization
- ☒ CommercialOrganization
- ☒ InternationalOrganization
- ☒ Document
- ☒ Location
- ☒ TimeInterval
- ☒ Event
- ☒ Agent

The European Central Bank yesterday shrugged off evidence of a worse than expected slowdown in the global economy and kept interest rates in the 12-nation zone unchanged at 4.5%.

Although Bank of England fears about the darkening outlook for the world economy prompted a surprise cut in British interest rates yesterday, the ECB declined the opportunity to join global efforts to boost flagging growth.

Its decision came despite data which showed economic confidence in Europe continuing to collapse and a further fall in US manufacturing orders as American industry struggles to climb out of recession.

The ECB has cut interest rates once this year, compared with six cuts by the US Federal Reserve and four by the Bank of England's monetary policy committee.

"Compared with more of a prior

Survey evidence confidence in and Alcatel a

In Germany, commerce ex

showed consumer confidence at its lowest level for two years.

Number

Apply To All Create Instance Dehighlight

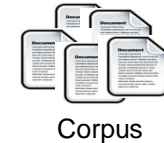
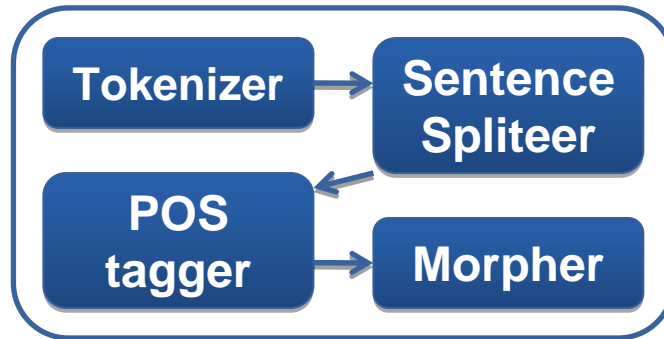
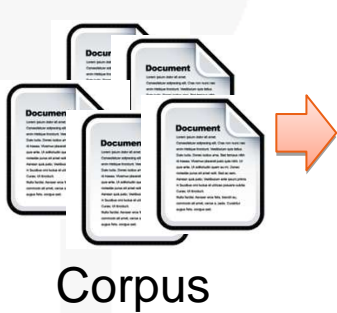
ontology	http://proton.semanticweb.org/2005/04/protont	X
class	Number	X
label	{}	X
seeAlso	{}	X

Document Editor Initialisation Parameters

Views built!

- **Onto Root Gazetteer PR**
 - Automatic semantic annotation
 - Creates a gazetteer PR that can be used with the FlexibleGazetteerPR

"Samsung corporation shares closed up ..."



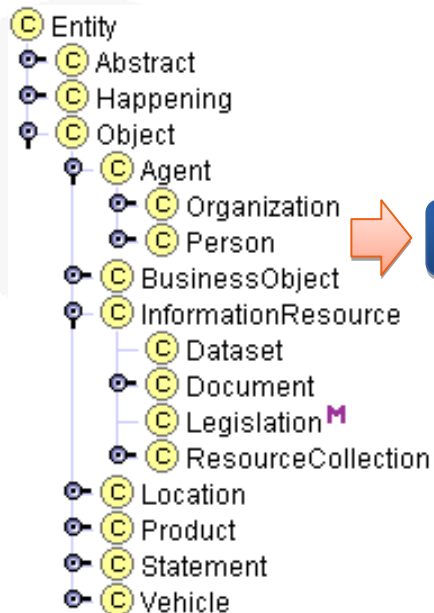
"Samsung corp...."

Token.root
="samsung"

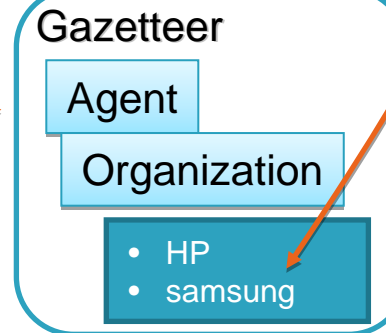
Ontology=
"myLocalOnt"
(class="Organ
ization")

FlexibleGazetteer

"Samsung corp..."



OntoRootGazetteer



Parameters for the new Onto Root Gazetteer

Name:

Name	Type	Required	Value
caseSensitive	java.lang.Boolean	✓	false
considerHeuristicRules	java.lang.Boolean	✓	false
considerProperties	java.lang.Boolean	✓	true
morpher	gate.creole.morph.Morph	✓	MorphAnal
ontology	gate.creole.ontology.Ontology	✓	<none>
posTagger	gate.creole.POSTagger	✓	<none>
propertiesToExclude	java.lang.String		
propertiesToInclude	java.lang.String		
separateCamelCasedWords	java.lang.Boolean	✓	true
tokeniser	gate.creole.tokeniser.DefaultTokeniser		<none>
useResourceUri	java.lang.Boolean	✓	true

OK Help Cancel

Ontology LR

POS Tagger
PR

Tokenizer
PR

Jape Rules Method signature:

```
public void doit(Document doc, Map bindings, AnnotationSet annotations,  
AnnotationSet inputAS, AnnotationSet outputAS, Ontology ontology)  
    throws JapeException
```

Ontology API

```
URI uri = new URI("http://my.uri/#Class1",false);  
Oclass c = ontology.addClass(uri);  
Oinstance i = ontology.addOInstance(uri2,c);  
Datatype dt = new Datatype(XMLStringURI);  
DatatypeProperty dtp =  
    ontology.addDatatypeProperty(uri3,domain,dt);  
i.addDatatypePropertyValue(dtp,new Literal("thevalue"));
```

- Introduction to GATE
- ANNIE
 - ANNIE Demo
- Jape Rules
 - Jape Rules Demo
- Machine Learning
- Ontology support
 - Semantic Annotation
 - Ontology Learning
 - **Semantic Annotation Demo**
- LSPs Demo

Manual semantic annotation

- Load ontology_tools plugging
- Load ontology_OWLIM2
- Language resource OWLIMOntologyLR
- Use OAT in one of the docs in Corpus
- Create annotation

- Introduction to GATE
- ANNIE
 - ANNIE Demo
- Jape Rules
 - Jape Rules Demo
- Machine Learning
- Ontology support
 - Semantic Annotation
 - Ontology Learning
 - Semantic Annotation Demo
- **LSPs Demo**



Messages

LSPs

Loaded Processing resources

Name	Type
ANNIE Gazetteer_00119	ANNIE Gazetteer

Selected Processing resources

!	Name	Type
	Document Reset PR_0003F	Document Reset PR
	ANNIE English Tokeniser_00040	ANNIE English Tokeniser
	ANNIE Sentence Splitter_00043	ANNIE Sentence Splitter
	ANNIE POS Tagger_00046	ANNIE POS Tagger
	GATE Morphological analyser_00047	GATE Morphological analyser
	Noun Phrase Chunker_0006E	Noun Phrase Chunker
	Flexible Gazetteer_0011A	Flexible Gazetteer
	Jape Transducer_0011C	Jape Transducer

>>

<<

ANNIE

LSPs

↑

↓

Corpus: GATE corpus_00028

The **corpus** and **document** parameters are not available as they are automatically set by the controller!

No selected processing resource

Name	Type	Required	Value
------	------	----------	-------

Run this application

- **GATE Morphological Analyser**: finds the root and affixes of a token
 - E.g.: infinitive form of the verb – to be (is, are, was, etc.)
 - singular form of a noun – company (companies)
- **Noun Phrase Chunker**: identifies nominal groups that have the same function
 - E.g.: **Thyroid medicines** belong to the general group of hormone medicines
- **Flexible Gazetteer**: wordlists grouped in different categories to perform NE or Key Phrase Lookups



Introduction to GATE

Andrés García-Silva, Elena Montiel-Ponsoda

Ontology Engineering Group
Facultad de Informática
Universidad Politécnica de Madrid
Campus de Montegancedo s/n
28660 Boadilla del Monte, Madrid, Spain

<http://www.oeg-upm.net>

emontiel@fi.upm.es
hagarcia@delicias.dia.fi.upm.es
Phone: 34.91.3363670
Fax: 34.91.3524819

Date: October 22nd, 2009