



Introduction to IE and ANNIE

David Jones
Mehmet Bakir





About this tutorial

This tutorial comprises the following topics:

- 1) Introduction to Information Extraction (**IE**)
- 2) ANNIE
- 3) Evaluation and Corpus Quality Assurance

After lunch, you'll learn how to use JAPE, the pattern matching language that many Processing Resources use



Named Entity Recognition (NER): the cornerstone of IE

Traditionally, **NER is the identification of proper names in texts**, and their classification into a set of predefined categories of interest

- Person
- Organisation (companies, government organisations, committees, etc.)
- Location (cities, countries, rivers, etc.)
- Date and time expressions

Various other types are frequently added, as appropriate to the application, e.g. newspapers, ships, monetary amounts, percentages.



Why is NE important?

- **NE provides a foundation** from which to build more complex IE systems
- Relations between NEs can provide tracking, ontological information and scenario building
- Tracking (co-reference): “Dr Smith”, “John Smith”, “John”, “he”
- Ontologies: “Athens, Georgia” vs “Athens, Greece”



Typical NE pipeline

- **Pre-processing** (tokenisation, sentence splitting, morphological analysis, POS tagging)
- **Entity finding** (gazetteer lookup, NE grammars)
- **Coreference** (alias finding, orthographic coreference etc.)
- **Exporting** to database / XML / ontology



Example of IE

John lives in London . He works there for Polar Bear Design .



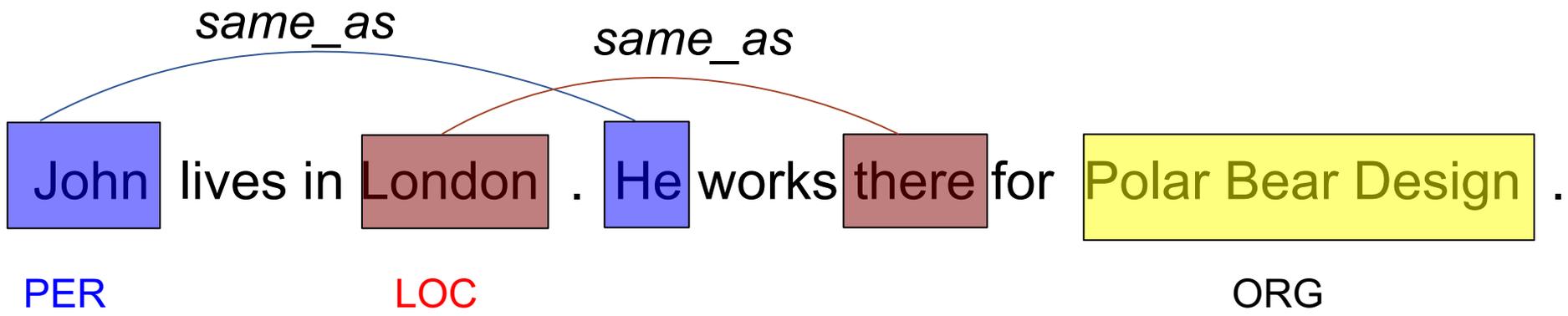
Basic NE Recognition

John lives in London . He works there for Polar Bear Design .

PER LOC ORG

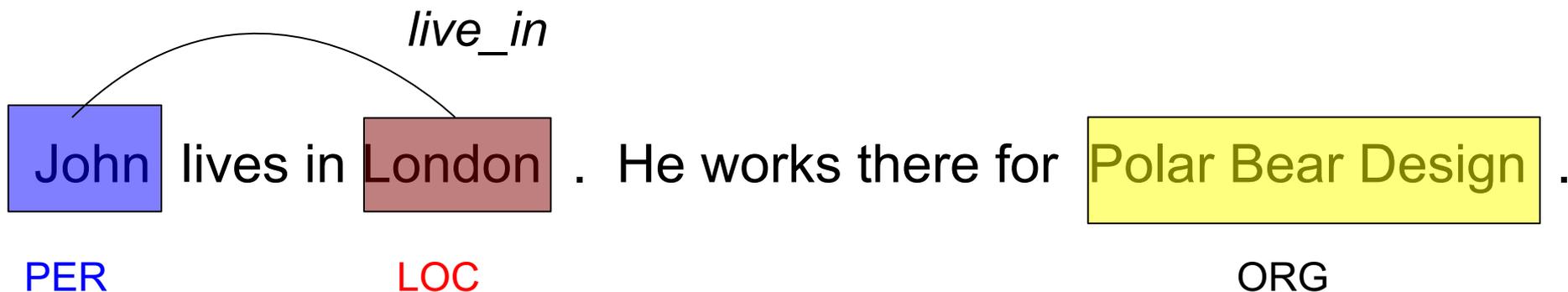


Co-reference



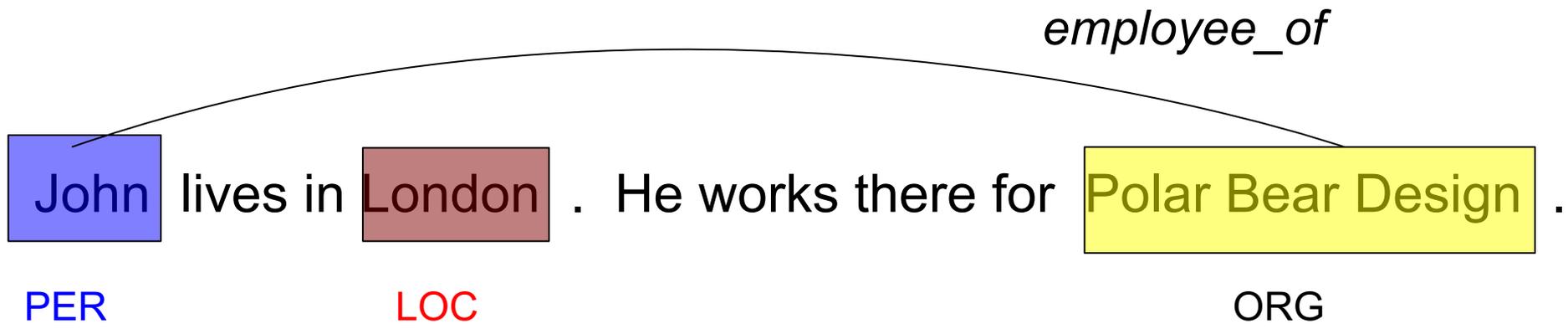


Relations



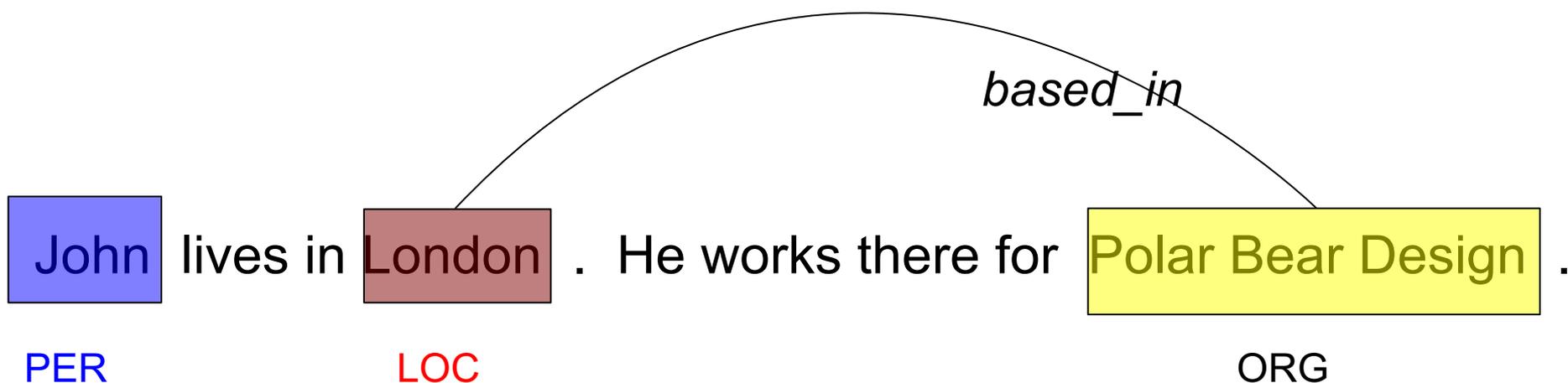


Relations (2)





Relations (3)



ANNIE: A Nearly New Information Extraction system



About this tutorial

- As before, this tutorial will be a hands on session with some explanation as you go.
- We will use a corpus of news texts in the file [annie-hands-on.zip](#)¹. Unzip this file if it isn't already.
- Things for you to try yourself are in **red**.
 - There will be instructions for you to follow for each step
 - Each step will be demonstrated
 - Correct answers will be shown before moving on
- **Start GATE on your computer now (if you haven't already)**

¹ <https://gate.ac.uk/wiki/TrainingCourseJune2018/>



A Nearly New Information Extraction system

ANNIE is a ready made collection of PRs that performs IE on unstructured text.

- ANNIE is “nearly new” because:
 - It was based on an existing IE system, LaSIE.
 - However, being nearly 20 years old, it's not really new any more.



What's in ANNIE?

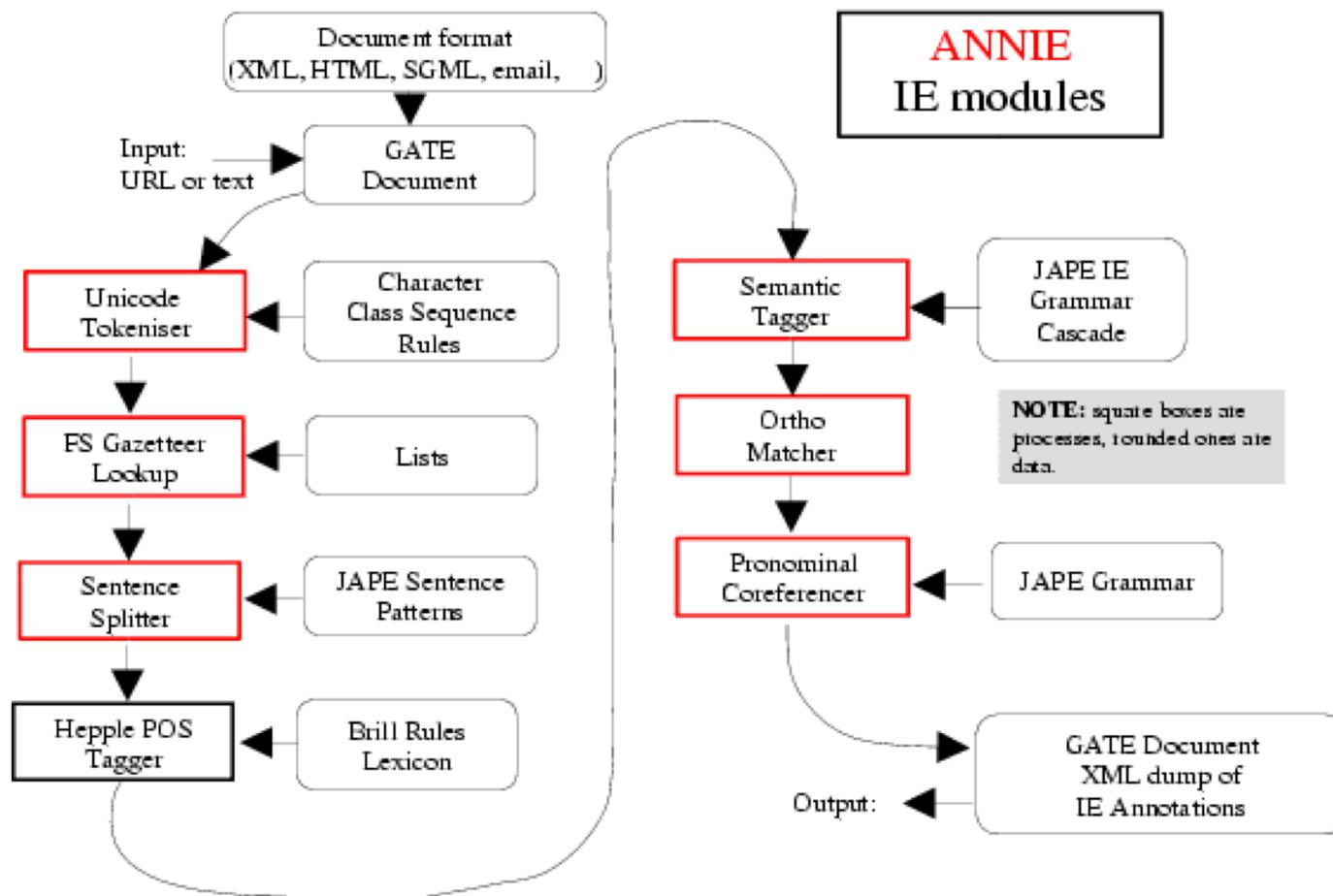
The ANNIE application contains a set of core PRs:

- Tokeniser
- Sentence Splitter
- POS tagger
- Gazetteers
- Named entity tagger (JAPE transducer)
- Orthomatcher (orthographic coreference)

There are also other useful PRs, which are not used in the default application, but can be added if necessary, e.g.

- NP chunker (in the Tagger:NP Chunking plugin)
- ANNIE VP Chunker (in the Tools plugin)

Core ANNIE components





Loading and running ANNIE

Because ANNIE is a ready-made application, we can just load it directly from the menu.

- Click the  icon from the top GATE menu.

OR

File → Ready Made Applications → ANNIE → ANNIE

OR

right-click ***Applications → Ready Made Applications → ANNIE → ANNIE***

- Create a new corpus and Populate it from the “**news-texts**” director.
- **Run ANNIE** and inspect the annotations.
- You should see a mixture of Named Entity annotations (Person, Location etc.) and some other linguistic annotations (Token, Sentence etc.).



Let's look at the PRs

Each PR in the ANNIE pipeline creates some new annotations, or modifies existing ones.

- **Document Reset** → removes annotations.
- **Tokeniser** → **Token** annotations.
- **Gazetteer** → **Lookup** annotations.
- **Sentence Splitter** → **Sentence**, **Split** annotations.
- **POS tagger** → adds **category** features to **Token** annotations.
- **NE transducer** → **Date**, **Person**, **Location**, **Organisation**, **Money**, **Percent** annotations.
- **Orthomatcher** → adds **match** features to NE annotations.



Document Reset

- This PR **should go at the beginning** of (almost) every application you create.
- It **removes annotations created previously**, to prevent duplication if you run an application more than once.
- It **does not remove the Original markup** set, by default.
- By default it also **keeps the “Key”** set (by convention the set used for evaluation).
- You can configure it to keep any other annotation sets you want, or to remove particular annotation types only.

Document Reset Parameters

Loaded Processing resources

Name	Type

Selected Processing resources

Name	Type
Document Reset PR_00016	Docur

Run "Document Reset PR_00016"?

Yes
 No
 If value of feature is

Corpus: <none>

Runtime Parameters for the "Document Reset PR_00016" Document Reset PR:

Name	Type	Required	Value
<?> annotationTypes	ArrayList		[]
<?> keepOriginalMarkupsAS	Boolean		true
<?> setsToKeep	ArrayList		[Key]

Specify any specific annotations to remove. By default, remove all.

Keep Original Markups set

Keep Key set

Tokenisation and sentence splitting



Tokeniser

- Tokenisation based on Unicode classes
- Declarative token specification language
- Produces **Token** and **SpaceToken** annotations with features **orthography** and **kind**
- **length** and **string** features are also produced
- Rule for a lowercase word with initial uppercase letter:

```
"UPPERCASE_LETTER" LOWERCASE_LETTER"* >  
  Token; orthography=upperInitial; kind=word
```

Document with Tokens

Annotation Sets Annotations List Annotations Stack Class Co-reference Editor Instance Text

Union Appeals For Talks To End BA Strike

Skip to navigation | Skip to content |
Home | Contact Us | News Search;
HubPage
Airwise News
Airport Guide
Airwise Travel
Search
Union Appeals For Talks To End BA Strike
March 22, 2010

Union leaders on Sunday called for talks with British Airways bosses to end strike action by cabin crew that has led to the cancellation of hundreds of flights and disrupted travel plans for thousands of passengers.

Type	Features
Token	{category=NNP, kind=word, length=5, orth=upperInitial, string=Union}
Token	{category=NNPS, kind=word, length=7, orth=upperInitial, string=Appeals}
Token	{category=IN, kind=word, length=3, orth=upperInitial, string=For}
Token	{category=NNS, kind=word, length=5, orth=upperInitial, string=Talks}
Token	{category=TO, kind=word, length=2, orth=upperInitial, string=To}

- Date
- FirstPerson
- JobTitle
- Location
- Lookup
- Money
- Organization
- Percent
- Person
- Sentence
- SpaceToken
- Split
- Title
- Token
- Unknown
- ▶ Original markups



ANNIE English Tokeniser

- The English Tokeniser is a slightly enhanced version of the Unicode tokeniser
- It comprises an additional JAPE transducer which adapts the generic tokeniser output for the POS tagger requirements
- It converts constructs involving apostrophes into more sensible combinations

don't → do + n't

you've → you + 've



Looking at Tokens

- **Tidy up GATE by removing all resources and applications (or just restart GATE)**
- **Load the news text hands-on corpus**
- **Create a new application (corpus pipeline)**
- **Load a Document Reset and an ANNIE English Tokeniser**
- **Add them (in that order) to the application and run on the corpus**
- **View the Token and SpaceToken annotations**
- **What different values of the “kind” feature do you see?**



Sentence Splitter

- The default splitter finds sentences based on Tokens.
- Creates **Sentence** annotations and **Split** annotations on the sentence delimiters.
- Uses a gazetteer of abbreviations etc. and a set of JAPE grammars which find sentence delimiters and then annotate sentences and splits.
- Load an **ANNIE Sentence Splitter PR** and add it to your application (**at the end**)
- **Run the application** and view the results

Document with Sentences

The screenshot displays the GATE software interface. The main window shows a document with several sentences highlighted in purple. The toolbar at the top includes tabs for 'Annotation Sets', 'Annotations List', 'Annotations Stack', 'Class', 'Co-reference Editor', 'Instance', and 'Text'. On the right side, there is a list of annotation types, with 'Sentence' checked. Below the text, there is a table with columns 'Type' and 'Features'.

The document text is as follows:

The opposition Conservatives, ahead in opinion polls, have been turning up the pressure on Labour over its links to Unite, saying the government had failed to take action quickly enough because it did not want to alienate its financial backers.

"We deplore the strike, and the prime minister and the transport secretary have said that absolutely clearly," Foreign Secretary David Miliband told Sky News.

"The way to resolve these disputes is through negotiation, it is damaging for the company, it is damaging for the crews and it is damaging for the country."

The dispute arose because BA, which has 12,000 cabin crew, wants to save an annual GBP£62.5 million pounds (USD\$95 million) to help cope with a fall in demand, volatile fuel prices and increased competition from low-cost carriers.

A spokesman said there was no estimate yet as to how much the industrial action would cost the company.

Type	Features
Sentence	{}



Sentence splitter variants

An alternate set of rules can be loaded with the regular sentence splitter.

- To do this, we can reload the sentence splitter using “**main.jape**” instead of “**main-single-nl.jape**” as the value of the grammar parameter
- The main **difference is the way it handles new lines**
- In some cases, you might want a new line to signal a new sentence, e.g. addresses
- In other cases, you might not, e.g. in emails that have been split by the email program
- A **regular expression Java-based splitter** is also available as a separate PR
- It is called **RegEx Sentence Splitter**, and is sometimes faster
- This handles new lines in the same way as the default sentence splitter

Shallow lexico-syntactic features



Part-of-Speech (POS) tagger

- ANNIE POS tagger is a Java implementation of Brill's transformation based tagger
- Previously known as **Hepple Tagger** (you may find references to this and to **heptag**)
- Trained on The Wall Street Journal, uses Penn Treebank tagset^{*}.
- Default ruleset and lexicon can be modified manually (with a little deciphering)
- Adds **category** feature to **Token** annotations
- Requires **Tokeniser** and **Sentence Splitter to be run first**

* https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html



Morphological analyser

- Not an integral part of ANNIE, but **can be found in the Tools plugin** as an “added extra”
- Flex based rules: can be modified by the user (instructions in the User Guide)
- Generates **root** feature on **Token** annotations
- **Requires Tokeniser** to be run first
- Requires **POS tagger to be run first if the considerPOSTag** parameter is set to true



Shallow lexico-syntactic features

- Add an **ANNIE POS Tagger** to your app.
- Add a **GATE Morphological Analyser** after the POS Tagger (If this PR is not available, **load the Tools plugin** first).
- **Re-run** your application.
- **Examine the features** of the **Token** annotations.
- New features of category and root have been added.



Gazetteers



Gazetteers

Gazetteers are plain text files containing lists of names (e.g. cities, rivers, people, ...).

- The ANNIE gazetteer has about 60,000 entries arranged in 80 lists.
- Each list reflects a certain category, e.g. airports, cities, first names etc.
- List entries might be entities or parts of entities, or they may contain contextual information (e.g. job titles often indicate people).
- Each gazetteer has an index file listing all the lists, plus features of each list (**majorType**, **minorType**, and **language**).
- Lists can be modified either internally using the Gazetteer Editor, or externally in your favourite editor.
- Gazetteers generate **Lookup** annotations with relevant features corresponding to the list matched.
- Lookup annotations are used primarily by the NE transducer.



Running the ANNIE Gazetteer

Various different kinds of gazetteer are available.

First we'll look at the **default ANNIE gazetteer**.

- **Add the ANNIE Gazetteer PR** to the end of your pipeline
- **Re-run the pipeline**
- Look for **“Lookup”** annotations and examine their features



ANNIE gazetteer - contents

- **Double click on the ANNIE Gazetteer PR** (under Processing Resources in the left hand pane) to open it.
- Make sure **“Gazetteer Editor”** is selected from the bottom tab.
- In the left hand pane (linear definition) you see the index file containing **all the lists**.
- In the right hand pane you see **the contents of the list** selected in the left hand pane.
- **The entries are read-only.**
- To edit ANNIE resources, we first need to make a copy of them that we can change.

Editing ANNIE gazetteer contents

1. Close *ANNIE Gazetteer* from PRs.
2. If you didn't download ANNIE before, go to the plugin manager and **select *Annie*, click the download button**  and **save it to a local folder.**
3. **Right-click PR and select *ANNIE Gazetteer*.**
4. Click the *listsURL* browse button 
5. Select the **file** tab, go to the location you downloaded *ANNIE*, go to the **gazetteer directory** and select **lists.def** and click OK.
6. Double click **ANNIE Gazetteer** in PR.

Now **each entry can be edited** by clicking in the box and typing.

New entries can be added by typing in the “**New list**” or “**New entry**” box respectively.



Gazetteer editor

GATE Developer 7.1-SNAPSHOT build 4319

File Options Tools Help

Messages Corpus Pipeline... in-whitbread-10... ANNIE Gazetteer...

airport.lst Add

List name	Major	Minor	Value
abbreviations.lst	stop		Aaccra
adbc.lst	adbc		Aalborg
airports.lst	location	airport	Aarhus
charities.lst	organization		Ababa
city.lst	location	city	Abadan
city_cap.lst	location	city	Abakan
company.lst	organization	company	Aberdeen
company_cap.lst	organization	company	Abha
country.lst	location	country	Abi Dhabi
country_abbrev.lst	location	country_abbrev	Abidjan
country_adj.lst	country_adj		Abilene
country_cap.lst	location	country	Abu
currency_prefix.lst	currency_unit	pre_amount	Abu Dhabi
currency_unit.lst	currency_unit	post_amount	Abuja
date_key.lst	date_key		Acapulco
date_unit.lst	date_unit		Acarigua
dav.lst	date	dav	Accra
			Adak Island

Filter Add +Cols 1989 entries Case Ins.

Gazetteer Editor Initialisation Parameters

definition file entries

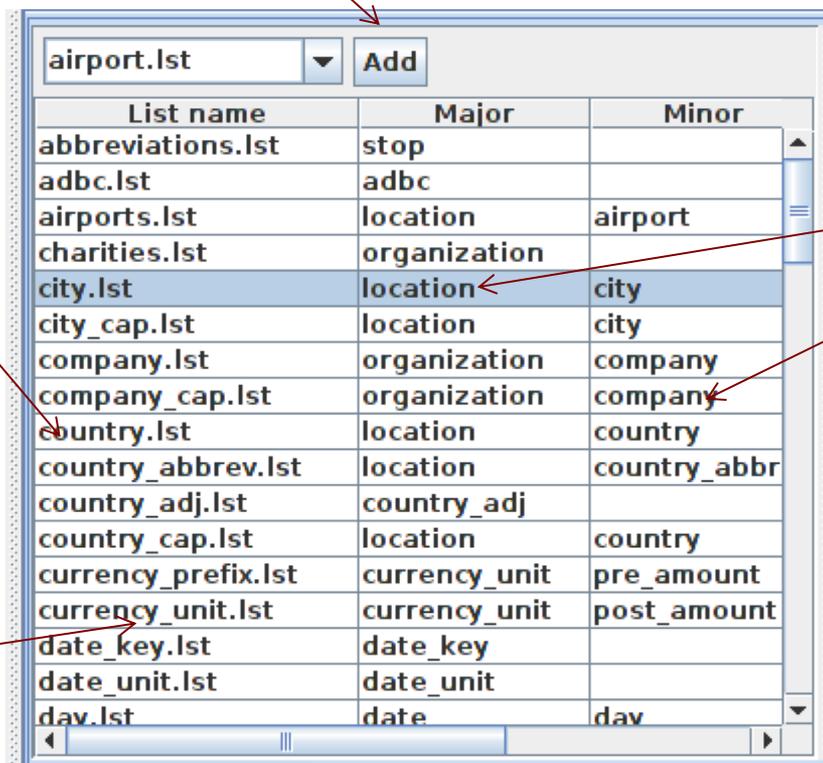
entries for selected list

Modifying the definition file

add a new list

edit an existing list name by typing here

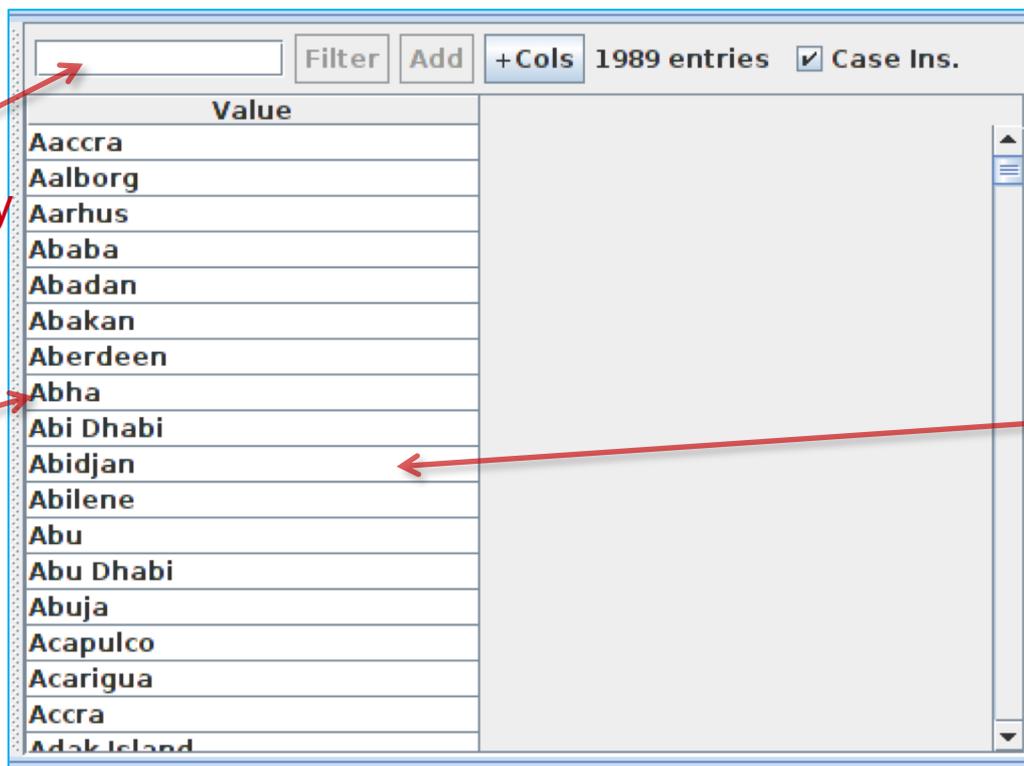
delete a list by right clicking on an entry and selecting 'Delete'



List name	Major	Minor
abbreviations.lst	stop	
adbc.lst	adbc	
airports.lst	location	airport
charities.lst	organization	
city.lst	location	city
city_cap.lst	location	city
company.lst	organization	company
company_cap.lst	organization	company
country.lst	location	country
country_abbrev.lst	location	country_abbrev
country_adj.lst	country_adj	
country_cap.lst	location	country
currency_prefix.lst	currency_unit	pre_amount
currency_unit.lst	currency_unit	post_amount
date_key.lst	date_key	
date_unit.lst	date_unit	
day.lst	date	day

edit the major and minor Types by typing here

Modifying a list



add a new entry
by typing here

edit an
existing entry
by typing here

Delete an entry by
right clicking and
selecting "Delete"



Editing gazetteer lists

- Click on any list to see the entries.
- Note that some lists are not very complete!
- **Try adding, deleting and editing** existing lists, or the list definition file.
- To save an edited gazetteer, **use *Ctrl-S* shortcut** or right click on the gazetteer name in the tabs at the top or in the resources pane on the right, and select “**Save and Reinitialise**” before running the gazetteer again.
- **Try adding a new word** from a document you have loaded (that is not currently recognised as a Lookup) into the gazetteer, re-run the gazetteer and check the results.



Editing gazetteers outside GATE

- You can also edit both the definition file and the lists outside GATE, in a text editor
- If you choose this option, **you will need to reinitialise the gazetteer** in GATE before running it again
- **To reinitialise** any PR, **right click on its name** in the Resources pane and **select “Reinitialise”**



List attributes

When something in the text matches a gazetteer entry, a **Lookup** annotation is created, with various features and values.

- The ANNIE gazetteer has the following default feature types: **majorType**, **minorType**, **language**.
- These features are used as a kind of classification of the lists: **in the definition file features are separated by “:”**
- For example, the “city” list has a majorType “location” and minorType “city”, while the “country” list has “location” and “country” as its types.
- Later, in the JAPE grammars, we can refer to all Lookups of type location, or we can be more specific and refer just to those of type “city” or type “country”.



NE transducers



NE transducer

- Gazetteers can be used to find terms that suggest entities.
- However, the entries can often be ambiguous.
 - “May Jones” vs “May 2010” vs “May I be excused?”
 - “Mr Parkinson” vs “Parkinson's Disease”.
 - “General Motors” vs. “General Smith”.
- Hand-crafted grammars can be used to **define patterns over the Lookups and other annotations.**
- These patterns can help disambiguate, and they can combine different annotations, e.g. Dates can be comprised of day + number + month.
- **NE transducer consists of a number of grammars written in the JAPE language.**
- This afternoon will be devoted to JAPE.



ANNIE NE Transducer

- **Load an ANNIE NE Transducer PR**
- **Add it to the end of the application**
- **Run the application**
- **Look at the annotations**
- You should see some new annotations such as **Person**, **Location**, **Date** etc.
- These will have features showing more specific information (e.g. what kind of location it is) and the rules that were fired (for ease of debugging)

Co-reference



Using co-reference

- Different expressions may refer to the same entity.
- Orthographic co-reference module (orthomatcher) matches proper names and their variants in a document.
- [*Mr Smith*] and [*John Smith*] will be matched as the same person.
- [*International Business Machines Ltd.*] will match [*IBM*].



Orthomatcher PR

- Performs co-reference resolution based on orthographical information of entities
- Produces a list of annotation IDs that form a **co-reference “chain”**
- List of such lists stored as a **document feature** named **“MatchesAnnots”**
- Improves results by assigning entity type to previously unclassified names, based on relations with classified entities
- May not reclassify already classified entities
- Classification of unknown entities very useful for surnames which match a full name, or abbreviations, e.g. “Bonfield” **<Unknown>** will match “Sir Peter Bonfield” **<Person>**
- A pronominal PR is also available



Looking at co-reference

- Add a new PR: **ANNIE OrthoMatcher**.
- Add it **to the end** of the application.
- **Run** the application.
- In a document view, **open the co-reference editor** by clicking the button above the text.

All the documents in the corpus should have some co-reference, but some may have more than others.

Co-reference editor

Annotation Sets Annotations List Annotations Stack Co-reference Editor Text

Seven **UK** airlines including British Airways, Virgin Atlantic, BMI British Midland and EasyJet, on Friday took over control of the air traffic control system, completing one of the government's most controversial public-private partnership deals.

Completion of the **National Air Traffic Services** deal comes at a critical time for the government as it tries to push through the PPP for the London Underground.

The sale to a strategic investor of a 46 per cent stake in **Nats** is the first time in Europe that management control of en route air traffic services has passed into private hands.

It has been carried out despite a pledge by Labour before the 1997 general election that **UK** air was "not for sale."

Under the terms of the deal, which was approved by the European competition authorities in May, the government has retained a 49 per cent stake and a golden share, while a 5 per cent stake is to be allocated to **Nats'** 5,700 staff.

The **Airline Group**, which also includes the charter carriers Airtours International Airways, Britannia Airways and Monarch Airlines, is paying GBP50m (\$71m) to acquire the 46 per cent stake.

Total government proceeds from the deal amount to about GBP800m, with the lion's share of the funds coming from new debt raised by **Nats**. The **Airline Group** has agreed financing facilities for **Nats** with a group of banks led by Barclays and Abbey National.

Completion of the deal has come about two months behind the original schedule announced at the end of **March**.

It is understood that negotiations were held up by concerns expressed by the banks financing the deal about revised traffic forecasts presented by **Nats** after the selection of the **Airline Group** as the government's partner was announced at the end of **March**.

The **Airline Group** is taking over **Nats** at a difficult time with air traffic control capacity under increasing pressure from rising air traffic volumes.

Sets : Default

Types : Organization Show

Co-reference Data

Default

- National Air Traffic Services
- Airline Group
- UK
- Swanwick
- March



Using the co-reference editor

- **Select the annotation set** you wish to view (Default)
- A list of all the co-reference chains that are based on annotations in the currently selected set is displayed
- **Select an item in the list** to highlight all the member annotations of that chain in the text (you can select more than one at once)
- **Hovering over** a highlighted annotation in the text **enables you to delete** an item from the co-reference chain
- **Try it!**

Evaluation



“We didn’t underperform. You over expected.”



Evaluation exercises: preparation

- Restart GATE, or **close all documents** and PRs to tidy up
- **Load the annie-hands-on/news-texts** into a corpus
- Take a **look at the annotations**.
- There is a set called “**Key**”. This is a set of annotations against which we want to evaluate ANNIE. In practice, **they could be manual annotations, or annotations from another application**.
- **Load ANNIE and run it**
- You should have annotations in the **Default** set from ANNIE, and in the **Key** set, against which we can compare them.



Annotation Diff

- Graphical comparison of 2 sets of annotations
- Visual diff representation, like tkdiff
- Compares one document at a time, one annotation type at a time

Annotations are like squirrels...



Annotation Diff helps with “spot the difference”

Annotation Diff Exercise

- Open the document “**ft-airlines-27-jul-2001.xml**”
- Open the ***Annotation Diff*** (Tools → Annotation Diff or **click the icon** )
- **For the Key set** (may contain the manual annotations) select **Key annotation set**
- **For the Response set** (containing annotations from ANNIE) select the **Default annotation set**
- **For the Type option** select the **Organization annotation.**
- Click on “Compare”
- Scroll down the list, to see correct, partially correct, missing and false positive annotations



Annotation Diff

Annotation Diff Tool

Key doc: Key set: Type: Weight:

Resp. doc: Resp. set: Features: all some none

Start	End	Key	Features	=?	Start	End	
1932	1936	Nats	{}	=	1932	1936	Nats
2456	2460	Nats	{}	=	2456	2460	Nats
2070	2075	LATCC	{}	=	2070	2075	LATCC
1354	1362	Barclays	{}	=	1354	1362	Barclays
1784	1788	Nats	{}	=	1784	1788	Nats
1751	1768	The·Airline·Group	{}	~	1755	1768	Airline·Gro
938	955	The·Airline·Group	{}	~	942	955	Airline·Gro
1669	1686	the·Airline·Group	{}	~	1673	1686	Airline·Gro
2412	2429	The·Airline·Group	{}	~	2416	2429	Airline·Gro
1266	1283	The·Airline·Group	{}	~	1270	1283	Airline·Gro
1052	1068	Monarch·Airlines	{}	~	1030	1068	Britannia·A
2029	2068	London·Area·and·Terminal·Control·Centre	{}	~	2045	2068	Terminal·C
634	640	Labour	{}	-?			
1030	1047	Britannia·Airways	{}	-?			
				?-	2029	2040	London·Are
				?-	2386	2395	Hampshire

10 documents loaded

Correct:	19	Recall	Precision	F-measure	
Partially correct:	7	Strict:	0.68	0.68	0.68
Missing:	2	Lenient:	0.93	0.93	0.93
False positives:	2	Average:	0.80	0.80	0.80



Comparing the individual annotations

- In the Annotation Diff, colour codes indicate whether the annotation pair shown are correct, partially correct, missing (false negative) or false positive.
- You can sort the columns however you like.



Comparing the annotations

Annotation Diff Tool

Key doc: ft-airlines-27-jul-2001.xml_00068 Key set: Key Type: Organization Weight: 1.0

Resp. doc: ft-airlines-27-jul-2001.xml_00068 Resp. set: [Default set] Features: all some none Compare

Start	End	Key	Features	=?	Start	End	Response	Features
45	60	Virgin:Atlantic	{rule2=OrgCountryFinal, rule1=null}	=	45	60	Virgin:Atlantic	{orgType=company, ru...nal=OrgCountryFinal}
265	294	National:Air:Traffic:Services	{rule1=OrgXKey, rule...al, orgType=unknown}	=	265	294	National:Air:Traffic:Services	{orgType=unknown, ru..., 1652, 1659, 1664}
1367	1381	Abbey:National	{rule1=OrgXBase, rul...al, orgType=unknown}	=	1367	1381	Abbey:National	{orgType=unknown, ru... ruleFinal=OrgFinal}
28	43	British:Airways	{rule1=GazOrganizati...al, orgType=company}	=	28	43	British:Airways	{orgType=company, ru... ruleFinal=OrgFinal}
1641	1645	Nats	{}	=	1641	1645	Nats	{orgType=other, rule..., 1652, 1659, 1664}
1784	1788	Nats	{}	=	1784	1788	Nats	{orgType=other, rule..., 1652, 1659, 1664}
1932	1936	Nats	{}	=	1932	1936	Nats	{orgType=other, rule..., 1652, 1659, 1664}
2070	2075	LATCC	{}	=	2070	2075	LATCC	{orgType=other, rule... ruleFinal=OrgFinal}
2277	2281	Nats	{}	=	2277	2281	Nats	{orgType=other, rule..., 1652, 1659, 1664}
2456	2460	Nats	{}	=	2456	2460	Nats	{orgType=other, rule..., 1652, 1659, 1664}
388	406	London:Underground	{}	=	388	406	London:Underground	{orgType=company, ru... ruleFinal=OrgFinal}
998	1028	Airtours:International:Airways	{}	=	998	1028	Airtours:International:Airways	{orgType=unknown, ru... ruleFinal=OrgFinal}
1260	1264	Nats	{}	=	1260	1264	Nats	{orgType=other, rule..., 1652, 1659, 1664}
1320	1324	Nats	{}	=	1320	1324	Nats	{orgType=other, rule..., 1652, 1659, 1664}
1354	1362	Barclays	{}	=	1354	1362	Barclays	{orgType=company, ru... ruleFinal=OrgFinal}
62	81	BMI:British:Midland	{}	=	62	81	BMI:British:Midland	{orgType=company, ru... ruleFinal=OrgFinal}
86	93	EasyJet	{}	=	86	93	EasyJet	{orgType=company, ru... ruleFinal=OrgFinal}
468	472	Nats	{}	=	468	472	Nats	{orgType=other, rule..., 1652, 1659, 1664}
918	922	Nats	{}	=	918	922	Nats	{orgType=other, rule..., 1652, 1659, 1664}
1751	1768	The:Airline:Group	{}	~	1755	1768	Airline:Group	{orgType=unknown, ru..., 1647, 1649, 1663}
2412	2429	The:Airline:Group	{}	~	2416	2429	Airline:Group	{orgType=unknown, ru..., 1647, 1649, 1663}
938	955	The:Airline:Group	{}	~	942	955	Airline:Group	{orgType=unknown, ru..., 1647, 1649, 1663}
1266	1283	The:Airline:Group	{}	~	1270	1283	Airline:Group	{orgType=unknown, ru..., 1647, 1649, 1663}
1669	1686	the:Airline:Group	{}	~	1673	1686	Airline:Group	{orgType=unknown, ru..., 1647, 1649, 1663}
1030	1047	Britannia:Airways	{}	~	1030	1068	Britannia:Airways and Monarch:Airlines	{orgType=unknown, ru... ruleFinal=OrgFinal}
2029	2068	London:Area and Terminal:Control:Centre	{}	~	2045	2068	Terminal:Control:Centre	{orgType=unknown, ru... ruleFinal=OrgFinal}
634	640	Labour	{}	~?				
1052	1068	Monarch:Airlines	{}	~?				
				?	376	379	PPP	{rule =AcronymOrg, orgType =unknown}
				?	2036	2040	Area	{rule =OrgConjOrg1, orgType =unknown}
				?	1129	1134	Total	{orgType=company, ru... ruleFinal=OrgFinal}

Correct: 19 Recall Precision F-measure 9 documents loaded

Partially correct: 7 Strict: 0.68 0.66 0.67

Missing: 2 Lenient: 0.93 0.90 0.92

False positives: 3 Average: 0.80 0.78 0.79

Statistics Adjudication

Key annotations (indicated by a red arrow pointing to the left column)

Response annotations (indicated by a red arrow pointing to the right column)



Measuring success

- In IE, we classify the annotations produced in one of 4 ways:

Correct = things annotated correctly

e.g. annotating “Sheffield” as a Location

Missing = things not annotated that should have been

e.g. not annotating “Sheffield” as a Location

False positive (spurious) = things annotated wrongly

e.g. annotating “Sheffield” as a Location in “Sheffield United F.C.”

Partially correct = the annotation type is correct, but the span is wrong

e.g. annotating just “Trump” as a Person (too short) or annotating “Unfortunately Donald Trump” as a Person (too long)

Finding Precision, Recall and F-measure

Annotation Diff Tool

Key doc: ft-airlines-27-jul-200... Key set: Key Type: Organization Weight: 1.0

Resp. doc: ft-airlines-27-jul-200... Resp. set: [Default set] Features: all some none

Start	End	Key	Features	=?	Start	End	
1932	1936	Nats	{}	=	1932	1936	Nats
2456	2460	Nats	{}	=	2456	2460	Nats
2070	2075	LATCC	{}	=	2070	2075	LATCC
1354	1362	Barclays	{}	=	1354	1362	Barclays
1784	1788	Nats	{}	=	1784	1788	Nats
1751	1768	The·Airline·Group	{}	~	1755	1768	Airline·Gro
938	955	The·Airline·Group	{}	~	942	955	Airline·Gro
1669	1686	the·Airline·Group	{}	~	1673	1686	Airline·Gro
2412	2429	The·Airline·Group	{}	~	2416	2429	Airline·Gro
1266	1283	The·Airline·Group	{}	~	1270	1283	Airline·Gro
1052	1068	Monarch·Airlines	{}	~	1030	1068	Britannia·A
2029	2068	London·Area·and·Terminal·Control·Centre	{}	~	2045	2068	Terminal·C
634	640	Labour	{}	-?			
1030	1047	Britannia·Airways	{}	-?			
				?-	2029	2040	London·Are
				?-	2386	2395	Hampshire

10 documents loaded

	Correct:	Recall	Precision	F-measure
Partially correct:	7	Strict: 0.68	0.68	0.68
Missing:	2	Lenient: 0.93	0.93	0.93
False positives:	2	Average: 0.80	0.80	0.80

Statistics Adjudication

← scores displayed



Precision

- How many of the entities your application found were correct?

$$\textit{Precision} = \frac{\textit{Correct}}{\textit{Correct} + \textit{False positive}}$$



Recall

- How many of the entities that exist did your application find?
- Sometimes recall is called **coverage**

$$\textit{Recall} = \frac{\textit{Correct}}{\textit{Correct} + \textit{Missing}}$$



F-Measure

- Precision and recall tend to trade off against one another.
- If you specify your rules precisely to improve precision, you may get a lower recall.
- If you make your rules very general, you get good recall, but low precision.
- This makes it difficult to compare applications, or to check whether a change has improved or worsened the results overall.
- **F-measure combines precision and recall into one measure.**



F-Measure

- Also known as the “**harmonic mean**”.
- Usually, precision and recall are equally weighted.
- This is known as F1.
- To use F1, set the value of the F-measure weight to 1, this is the default setting in Annotation Diff tool.

$$F = 2 \cdot \left(\frac{\textit{precision} \cdot \textit{recall}}{\textit{precision} + \textit{recall}} \right)$$



Annotation Diff defaults to F1

Annotation Diff Tool

Key doc: ft-airlines-27-jul-200... Key set: Key Type: Organization Weight: 1.0

Resp. doc: ft-airlines-27-jul-200... Resp. set: [Default set] Features: all some none

Start	End	Key	Features	Start	End	Key	
1932	1936	Nats	{}	1932	1936	Nats	
2456	2460	Nats	{}	2456	2460	Nats	
2070	2075	LATCC	{}	2070	2075	LATCC	
1354	1362	Barclays	{}	1354	1362	Barclays	
1784	1788	Nats	{}	1784	1788	Nats	
1751	1768	The·Airline·Group	{}	1755	1768	Airline·Gro	
938	955	The·Airline·Group	{}	942	955	Airline·Gro	
1669	1686	the·Airline·Group	{}	1673	1686	Airline·Gro	
2412	2429	The·Airline·Group	{}	2416	2429	Airline·Gro	
1266	1283	The·Airline·Group	{}	1270	1283	Airline·Gro	
1052	1068	Monarch·Airlines	{}	1030	1068	Britannia·A	
2029	2068	London·Area·and·Terminal·Control·Centre	{}	2045	2068	Terminal·C	
634	640	Labour	{}	-?			
1030	1047	Britannia·Airways	{}	-?			
				?-	2029	2040	London·Are
				?-	2386	2395	Hampshire

10 documents loaded

Correct:	19	Recall	Precision	F-measure
Partially correct:	7	Strict:	0.68	0.68
Missing:	2	Lenient:	0.93	0.93
False positives:	2	Average:	0.80	0.80

Statistics Adjudication

F-measure weight set to 1



Statistics can mean what you want them to....

How we want **to measure partially correct annotations** may differ, depending on our goal.

In GATE, there are **3 different ways** to measure them

- The most usual way is to consider them to be “**half right**”.
- **Strict:** Only perfectly matching annotations are counted as correct.
- **Lenient:** Partially matching annotations are counted as correct. This makes your scores look better :)
- **Average:** Strict and lenient scores are averaged (this is the same as counting a half weight for every partially correct annotation).

Strict, Lenient and Average

Annotation Diff Tool

Key doc: ft-airlines-27-jul-200... Key set: Key Type: Organization Weight: 1.0

Resp. doc: ft-airlines-27-jul-200... Resp. set: [Default set] Features: all some none

Compare

Start	End	Key	Features	=?	Start	End	
1932	1936	Nats	{}	=	1932	1936	Nats
2456	2460	Nats	{}	=	2456	2460	Nats
2070	2075	LATCC	{}	=	2070	2075	LATCC
1354	1362	Barclays	{}	=	1354	1362	Barclays
1784	1788	Nats	{}	=	1784	1788	Nats
1751	1768	The·Airline·Group	{}	~	1755	1768	Airline·Gro
938	955	The·Airline·Group	{}	~	942	955	Airline·Gro
1669	1686	the·Airline·Group	{}	~	1673	1686	Airline·Gro
2412	2429	The·Airline·Group	{}	~	2416	2429	Airline·Gro
1266	1283	The·Airline·Group	{}	~	1270	1283	Airline·Gro
1052	1068	Monarch·Airlines	{}	~	1030	1068	Britannia·A
2029	2068	London·Area·and·Terminal·Control·Centre	{}	~	2045	2068	Terminal·C
634	640	Labour	{}	-?			
1030	1047	Britannia·Airways	{}	-?			
				?-	2029	2040	London·Are
				?-	2386	2395	Hampshire

Correct: 19 Recall Precision F-measure

Partially correct: 7

Missing: 2

False positives: 2

Strict: 0.68 0.68 0.68

Lenient: 0.93 0.93 0.93

Average: 0.80 0.80 0.80

10 documents loaded

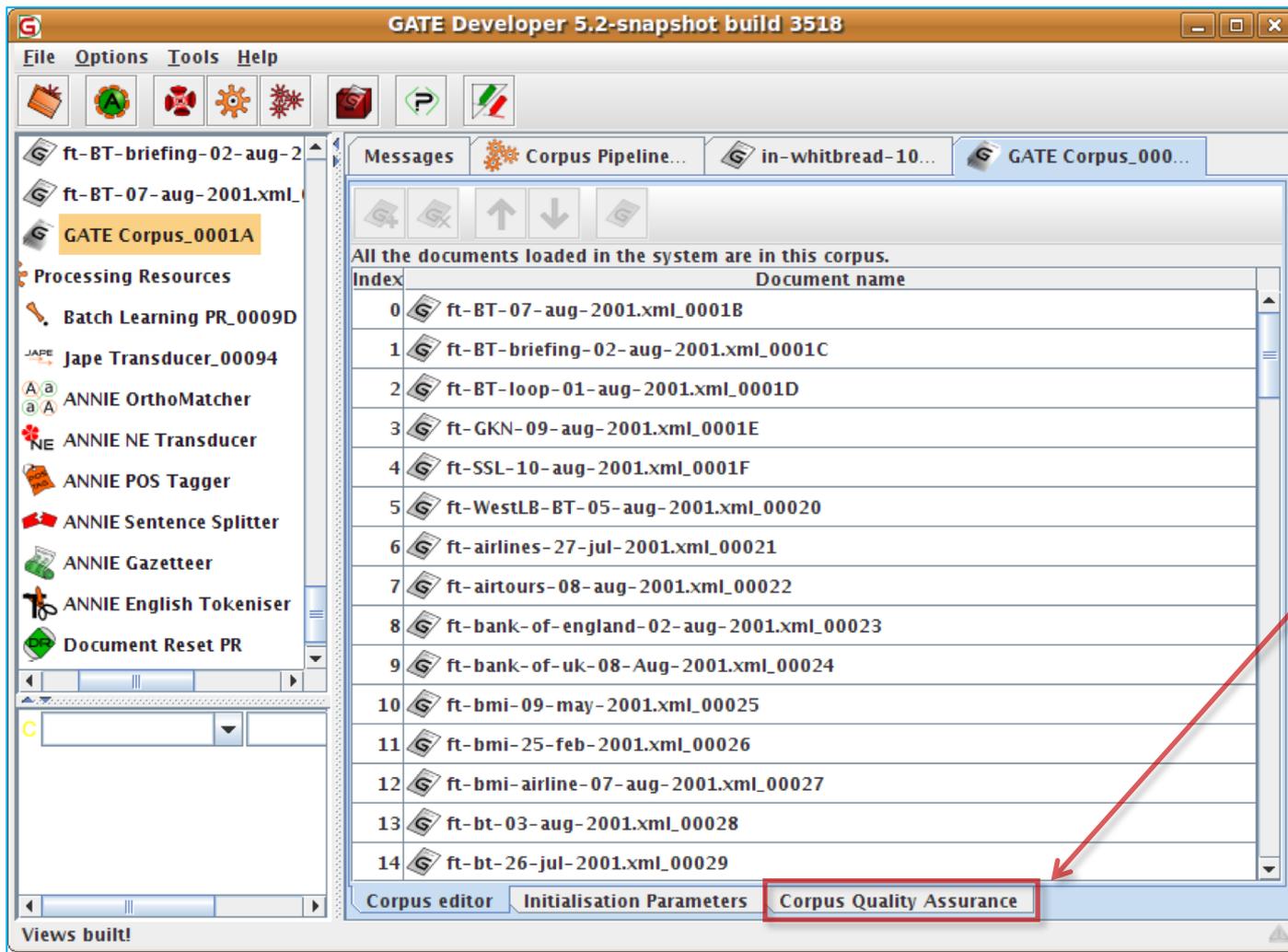
Statistics Adjudication



Corpus Quality Assurance

- **Corpus Quality Assurance tool extends the Annotation Diff functionality to the entire corpus**, rather than on a single document at a time.
- It **produces statistics both for the corpus as a whole** (Corpus statistics tab) **and for each document separately** (Document statistics tab).
- It **compares two annotation sets, but makes no assumptions about which (if either) set is the gold standard**. It just labels them A and B.
- This is because it can be used to measure **Inter Annotator Agreement (IAA)** where there is no concept of “correct” set.

Try out Corpus Quality Assurance



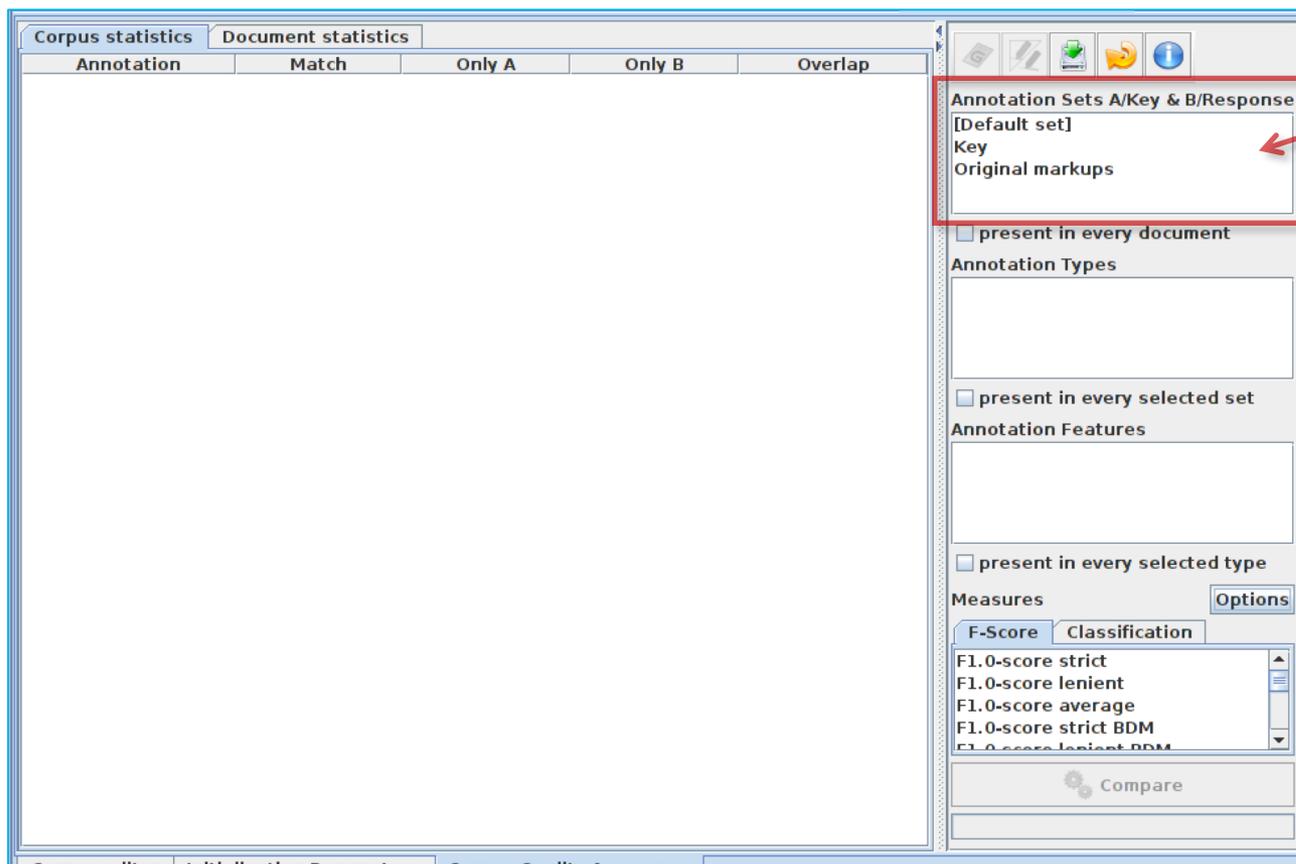
The screenshot shows the GATE Developer 5.2 interface. The main display pane is titled "Messages" and "Corpus Pipeline...". Below the toolbar, there is a list of documents loaded in the system, with the following table:

Index	Document name
0	ft-BT-07-aug-2001.xml_0001B
1	ft-BT-briefing-02-aug-2001.xml_0001C
2	ft-BT-loop-01-aug-2001.xml_0001D
3	ft-GKN-09-aug-2001.xml_0001E
4	ft-SSL-10-aug-2001.xml_0001F
5	ft-WestLB-BT-05-aug-2001.xml_00020
6	ft-airlines-27-jul-2001.xml_00021
7	ft-airtours-08-aug-2001.xml_00022
8	ft-bank-of-england-02-aug-2001.xml_00023
9	ft-bank-of-uk-08-Aug-2001.xml_00024
10	ft-bmi-09-may-2001.xml_00025
11	ft-bmi-25-feb-2001.xml_00026
12	ft-bmi-airline-07-aug-2001.xml_00027
13	ft-bt-03-aug-2001.xml_00028
14	ft-bt-26-jul-2001.xml_00029

At the bottom of the display pane, there are three tabs: "Corpus editor", "Initialisation Parameters", and "Corpus Quality Assurance". The "Corpus Quality Assurance" tab is highlighted with a red box, and a red arrow points to it from the text on the right.

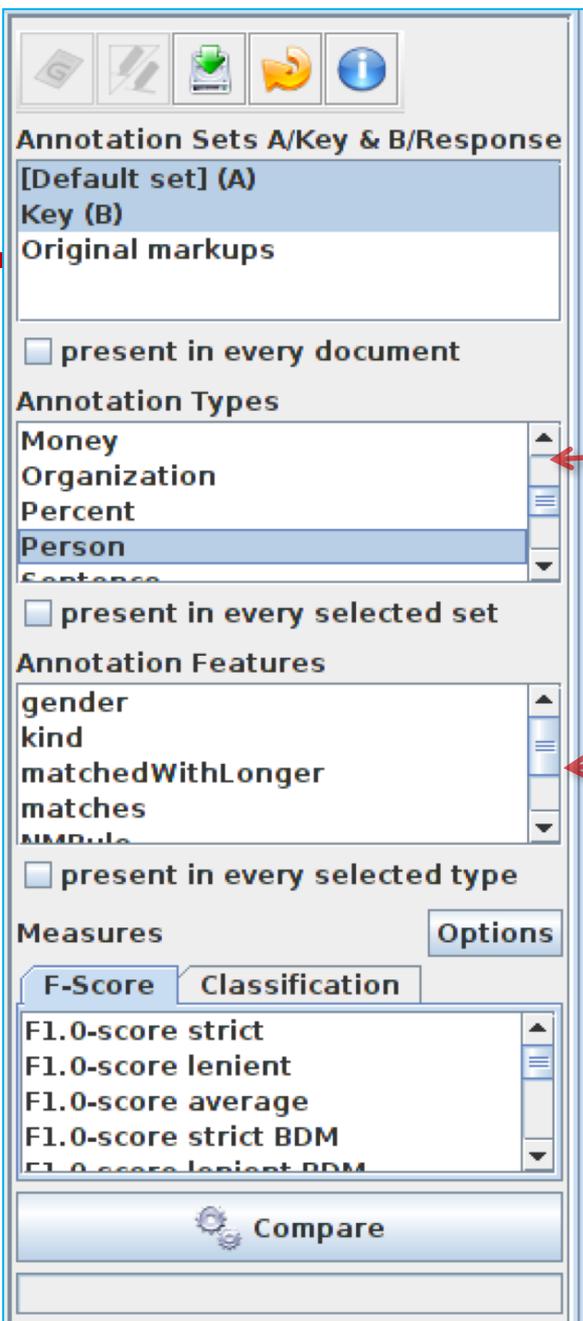
Open your hands-on corpus and click the Corpus Quality Assurance tab at the bottom of the Display pane.

Select Annotation Sets



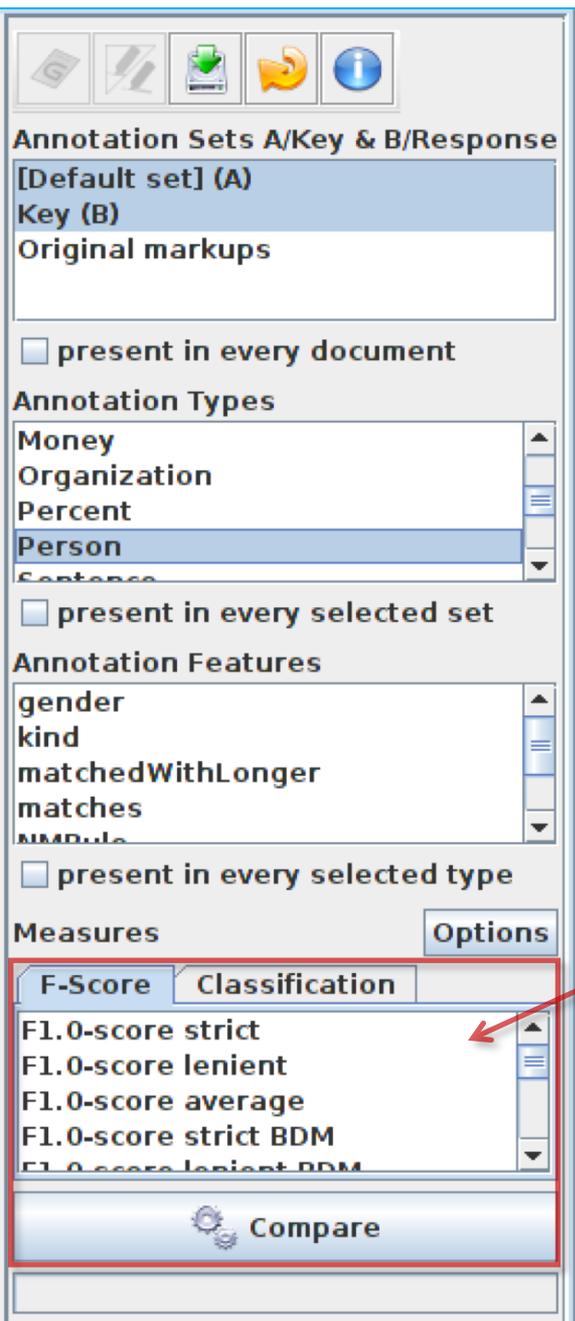
- **Select the annotation sets you wish to compare.**
- **Click on the Key annotation set – this will label it set A.**
- **Now click on the default annotation set - this will label it set B.**

Select Type



- **Select the annotation type to compare** (suggestion: select Organisation, Person and Location for now)
- **Select the features to include** (if any – leave unselected for now)
- **You can select as many types and features as you want.**

Select measure



- In the “**Measures**” box, **select the kind of F score** you want “Strict, Lenient, Average” or any combination of them.
- Suggestion: **try just “lenient” at first.**
- **Select Compare**

Corpus Statistics Tab

Corpus statistics		Document statistics					
Annotation	Match	Only A	Only B	Overlap	Prec.B/A	Rec.B/A	F1.0-l.
Location	55	6	0	5	1.0000	0.9091	0.9524
Organization	76	5	15	11	0.8529	0.9457	0.8969
Person	28	2	1	0	0.9655	0.9333	0.9492
Macro summary					0.9395	0.9294	0.9328
Micro summary	159	13	16	16	0.9162	0.9309	0.9235

- Each annotation type is listed separately
- Precision, recall and F measure are given for each
- Two summary rows provide micro and macro averages



Micro and Macro Averaging

- Micro averaging treats the entire corpus as one big document, for the purposes of calculating precision, recall and F.
- Macro averaging takes the average of the rows.
Here: the average over different annotation types

Document Statistics Tab

Corpus statistics		Document statistics					
Document	Match	Only A	Only B	Overlap	Prec.B/A	Rec.B/A	F1.0-l.
ft-airlines-27-jul-2001.xml_00030	28	4	2	7	0.9459	0.8974	0.9211
ft-airtours-08-aug-2001.xml_00031	19	0	0	0	1.0000	1.0000	1.0000
ft-bank-of-england-02-aug-2001.xml_00032	22	2	2	1	0.9200	0.9200	0.9200
ft-bmi-09-may-2001.xml_00033	24	1	2	2	0.9286	0.9630	0.9455
ft-claims-direct-10-aug-2001.xml_00034	21	2	1	0	0.9545	0.9130	0.9333
ft-commerzbank-10-aug-2001.xml_00035	10	2	4	2	0.7500	0.8571	0.8000
ft-equitable-07-auf-2001.xml_00036	9	1	4	0	0.6923	0.9000	0.7826
ft-house-price-08-aug-2001.xml_00037	9	1	0	1	1.0000	0.9091	0.9524
ft-industrial-gloom-07-Aug-2001.xml_00038	17	0	1	3	0.9524	1.0000	0.9756
Macro summary					0.9049	0.9288	0.9145
Micro summary	159	13	16	16	0.9162	0.9309	0.9235

- Each document is listed separately.
- Precision, recall and F measure are given for each.
- Two summary rows provide micro and macro (here: over documents) averages.



Summary

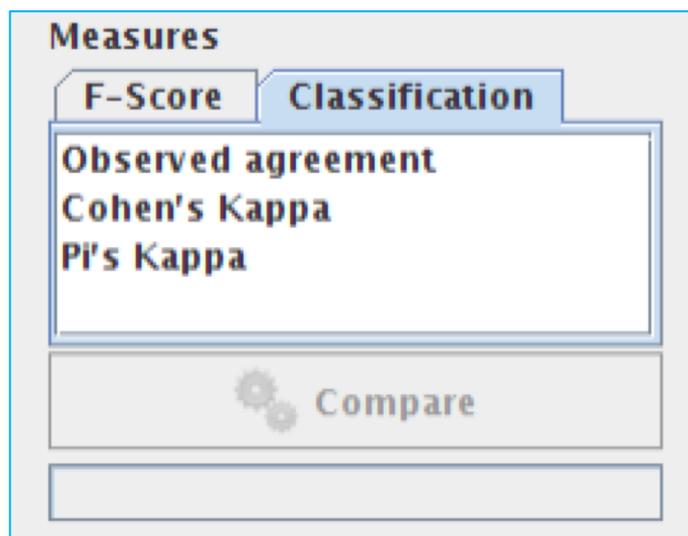
This session has been devoted to IE and ANNIE

- You should now have a basic understanding of:
what IE is
- how to load and run ANNIE,
- what each of the ANNIE components do,
- how to modify ANNIE components,
- evaluation using Annotation Diff and Corpus QA.



Optional

Classification Measures



- By default, Corpus Quality Assurance presents the F-measures.
- However, classification measures are also available.
- These are not suitable for entity extraction tasks.