
Module 4: Teamware: A Collaborative, Web-based Annotation Environment

© The University of Sheffield, 1995-2011

This work is licenced under the

Creative Commons Attribution-NonCommercial-ShareAlike Licence



Outline

- Why Teamware?
- What's Teamware?
- Teamware for annotation
- Teamware for quality assurance and curation
- Teamware for defining workflows, running automatic services, managing annotation projects
- Outlook

From Annotation Tools to *Collaborative Annotation Workflows*

We have lots and lots of tools and algorithms for annotation; what we need is

1. methodological instead of purely technological
2. multi-role instead of single role
3. assistive instead of autonomous
4. service-orientated, not monolithic
5. usable by non-specialists

GATE Teamware

- Research users in several EU projects
 - External users at IRF and Matrixware
 - Interest from other commercial users as well
-

GATE Teamware: Annotation



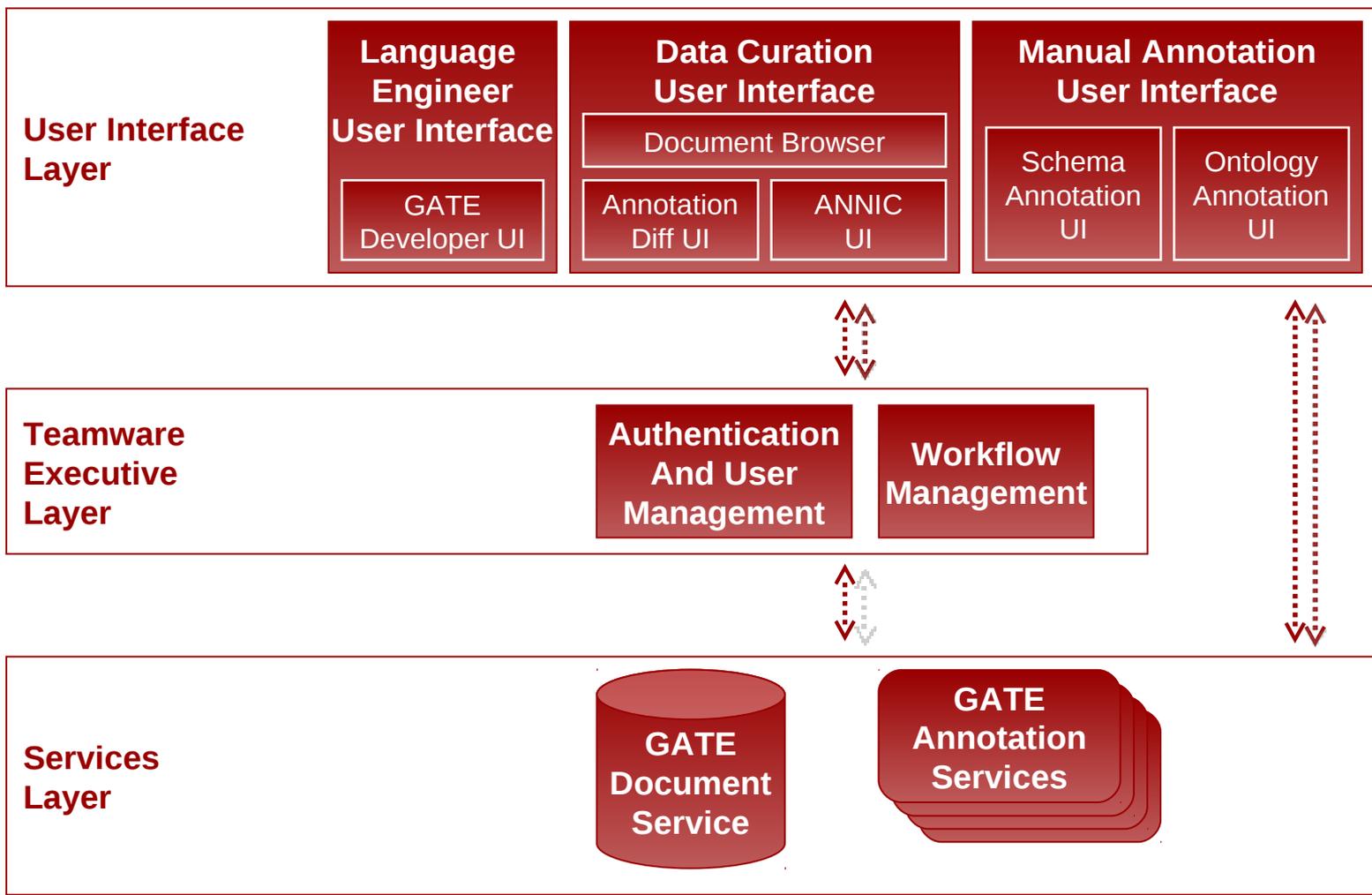
Workflows on the Web

GATE Teamware is:

- Collaborative, social, Web 2.0
- Parallel and distributed (using web services)
- Scalable (via service replication)
- Workflow based with business process integration



Teamware - Layer Cake

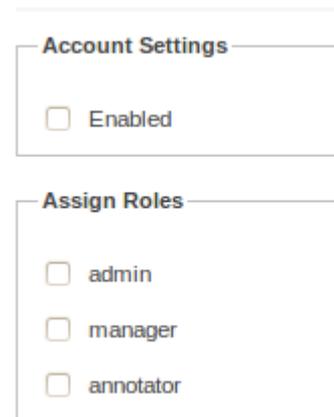


Division of Labour: A Multi-role Methodology

- **Annotators**
 - carries out manual annotation: from scratch or (faster) by correcting automatic annotation (generated by rule-based tools, machine learning, or a combination)
 - **Manager**
 - Defining **annotation guidelines and schemas**
 - **Choose relevant automatic services to pre-process**
 - Toolset including **performance benchmarking, progress monitoring tools, small linguistic customisations**
 - Define **workflow, manage annotators**, liaise with language engineers and sys admins
 - **Reconcile differences** between annotators, using IAA, AnnDiff
 - **Admin**
 - Setup the Teamware system, users, etc.
 - **Language engineer**
 - Uses **GATE Developer** to create bespoke services and deploy online
-

Hands-on: Administer Teamware

- You are given a Teamware installation running on GateCloud.net. It will only work during this course, but you can experiment further by registering on gatecloud.net
- Open the Teamware URL in your web browser.
- Log on with the provided user name and password.
- Select Admin/Manage Users, then press Add
- Create three users (remember to enable the accounts):
 - Two called test-annotator1 and test-annotator2, with only annotator rights (choose simple passwords)
 - One called test-manager, with annotator and manager rights
- Pass the details of your test-annotator1 account to the person on your left and test-annotator2 – to the person on your right

A screenshot of a web interface for account management. It features two sections: "Account Settings" and "Assign Roles". The "Account Settings" section has a single checkbox labeled "Enabled". The "Assign Roles" section has three checkboxes labeled "admin", "manager", and "annotator".

Account Settings	
<input type="checkbox"/>	Enabled

Assign Roles	
<input type="checkbox"/>	admin
<input type="checkbox"/>	manager
<input type="checkbox"/>	annotator

Teamware for Managers

- Defining workflows
- Running annotation projects
- Tracking progress
- Adjudication and measuring IAA

Teamware Workflows

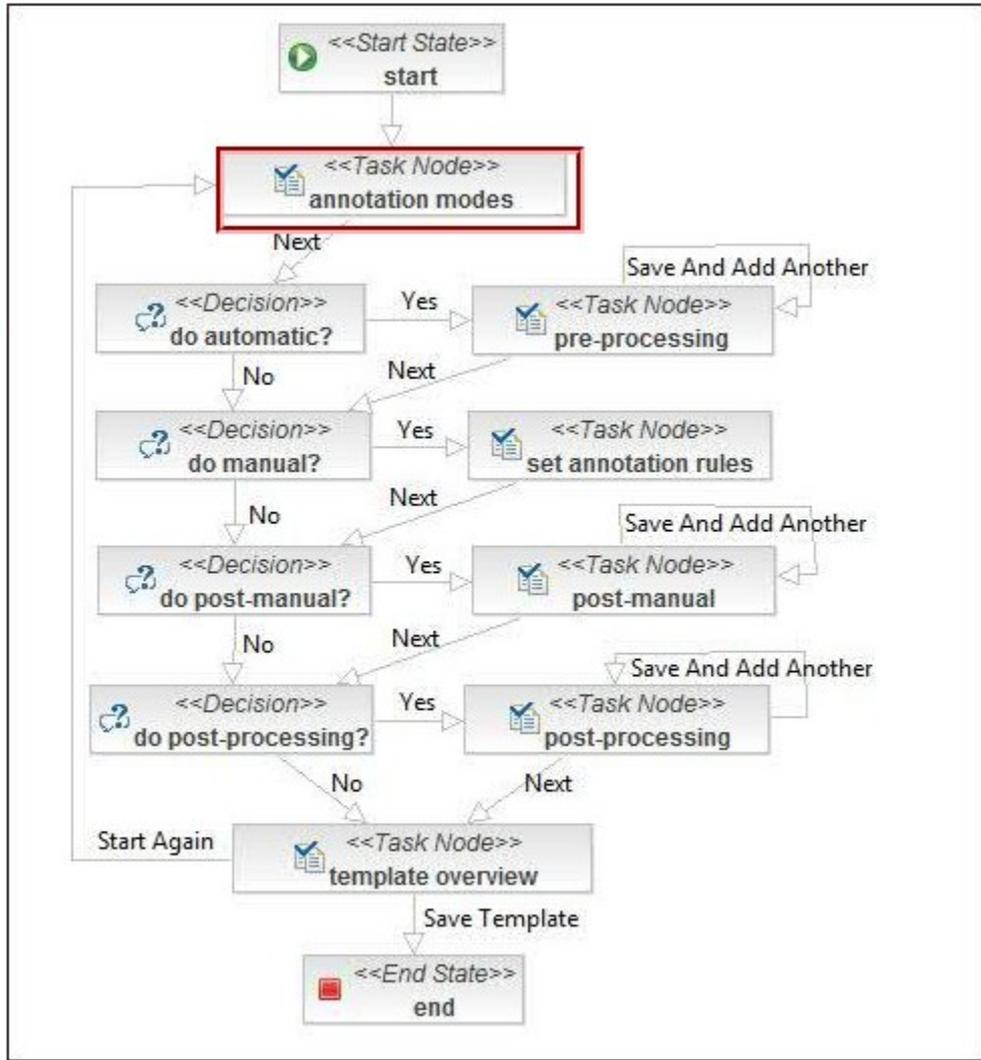
- Whole process is controlled by a **workflow manager**
- Workflow may be simple:
 - Give the document to n human annotators
- or more complex
 - Pre-process the documents (Gate Developer/Gate Cloud)
 - Pass each document to 3 annotators
 - Adjudicate the differences



Workflow Templates

Automatic Annotation:	<input type="checkbox"/>
Manual Annotation:	<input type="checkbox"/>
Post-Manual:	<input type="checkbox"/>
Review:	<input type="checkbox"/>
Post-Processing:	<input type="checkbox"/>

Next Help Quit



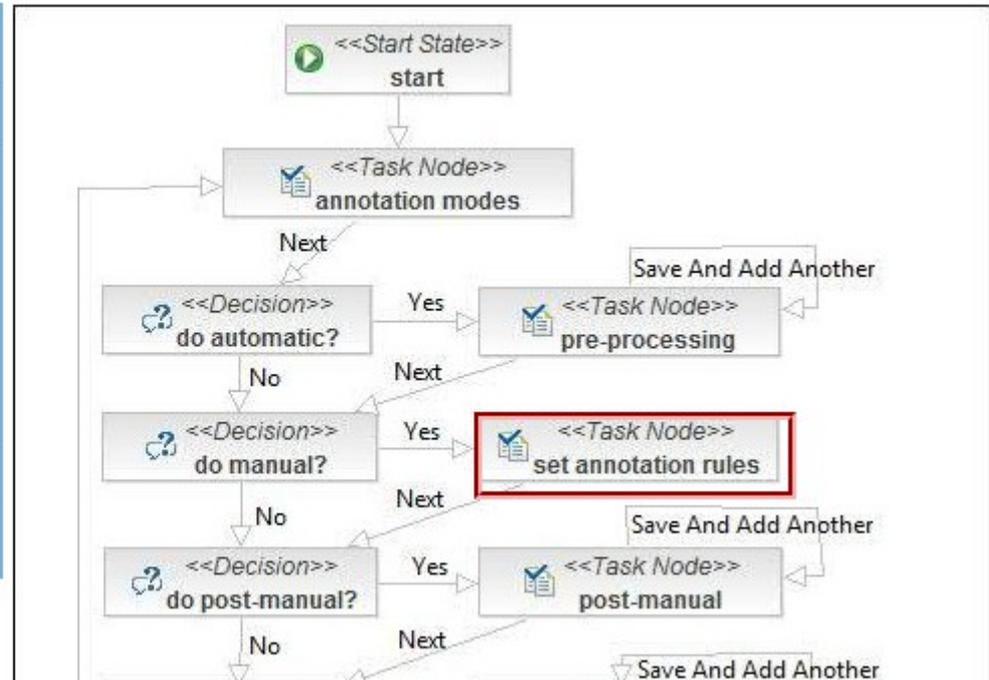
Defining new workflows

- Select Projects/WF Templates
 - Opens the WF wizard
 - Automatic annotation:
 - Runs one or more web services to pre-annotate
 - These need to be GATE Annotation Services (GAS), not just any service
 - The default Teamware installation includes ANNIE
 - Manual annotation
 - Post-manual: post-processing/merging service
 - Post-processing: Finalise and export corpus
-

Defining Manual Annotation WF

- Select only the Manual Annotation box in the WF template
- Configure further the number of annotators per document, whether they can reject documents, whether annotation sets are anonymous (annotator1, annotator2, ...)
- Select or upload annotation schemas

Annotators per Document:	2
Cancel Task Allowed:	<input checked="" type="checkbox"/>
Anonymous Annotation:	<input checked="" type="checkbox"/>
Annotation Schemas:*	<ul style="list-style-type: none"> OrganizationSchema.xml PersonSchema.xml AddressSchema.xml Non-FunctionalRequiren ConcordiaDataGroupsSc ConcordiaRequirementsS
Add Schema:	Add Schema
Pre-Manual Service:*	None



Annotation Schemas

- GATE's annotation schemas can be used to constrain manual annotation for consistency
 - You can use the standard ANNIE schemas or develop your own
 - Schemas should be consistent with your annotation guidelines
-

Annotation Schemas

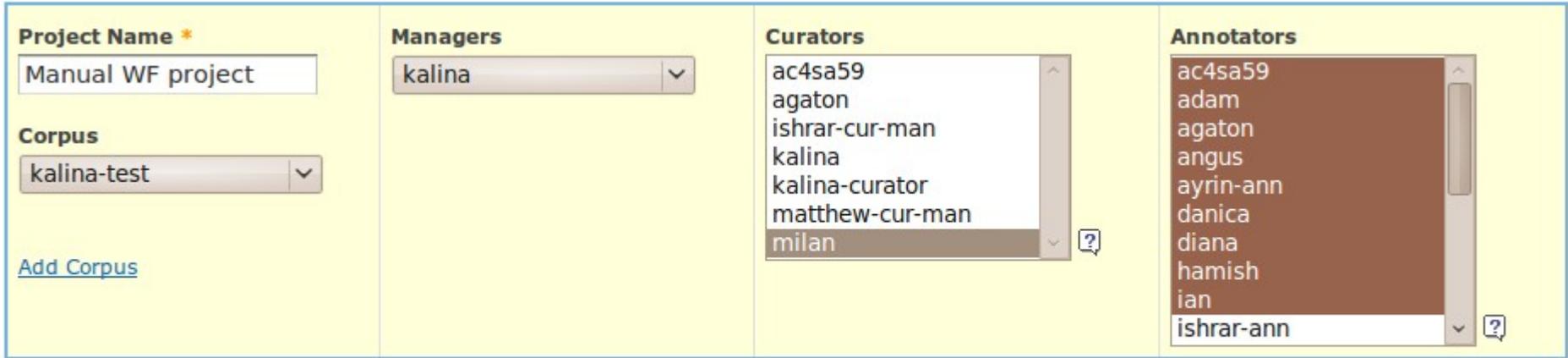
- Define types of annotations and restrict annotators to use specific feature-values
 - Person
 - Person.gender = male
 - Person.gender = female
- Uses the XML Schema language supported by W3C for these definitions

```
<?xml version="1.0"?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
<element name="Person">
  <complexType>
    <attribute name="gender" use="optional">
      <simpleType>
        <restriction base="string">
          <enumeration value="male"/>
          <enumeration value="female"/>
        </restriction>
      </simpleType> ...
```

Running a Manual Annotation WF

- The WF template defined in the previous steps can be saved and the project started or revised, as necessary
 - To actually run a manual annotation WF, you need to create a project, which defines:
 - Which corpus is going to be annotated
 - Who are the managers and annotators
 - The corpus can be uploaded via the link provided and is a ZIP file of documents
 - can be in any format supported by GATE
-

Running a Manual Annotation WF



The screenshot shows the configuration interface for a manual annotation workflow. It is divided into four main sections: Project Name, Managers, Curators, and Annotators. Below these sections are two buttons: 'Save & Start' and 'My Workflow Templates'.

Project Name *	Managers	Curators	Annotators
Manual WF project	kalina	ac4sa59 agaton ishrar-cur-man kalina kalina-curator matthew-cur-man milan	ac4sa59 adam agaton angus ayrin-ann danica diana hamish ian ishrar-ann

[Add Corpus](#)

Save & Start My Workflow Templates

- Once project is started, annotators can login and ask for tasks

Hands-On

- Login as a manager (using your test-manager account)
 - Schemas:
 - In the resources directory you will find 2 annotation schemas: Person and Organization
 - Edit the Person one to include optional gender
 - Define your own manual WF with your schemas, specifying 2 annotators per document
 - Start a project with your WF template
 - Choose yourself as a manager and add your 2 annotators to the project
 - For corpus: copy and save some text from a news site like the bbc (no more than 3-4 paragraphs), zip the saved document and upload it as your corpus
-

Teamware: Manual Annotation Tool

Document Editor

Type	Set	Start	End	Id	Features
Organization		4	19	159	{}
Date		23	31	170	{kind= date, rule1= GazDate, rule2= DateOnlyFinal}
Percent		104	119	157	{}
Percent		159	169	169	{rule= PercentBasic}
Date		212	222	138	{kind= date}
Organization		276	297	145	{orgType= unknown, rule1= OrgXBase, rule2= OrgFinal}
Organization		344	359	139	{}

34 Annotations (1 selected) Select:

The **Bank of England** on **Thursday** surprised the financial markets and delighted British businesses with a **0.25 percentage** point cut in its main interest rate to **5 per cent**. Interest rates have not been lower since **early 1964**.

But the mood inspired by the cut was marred by the **European Central Bank's** refusal to cut rates, and by concern at the **Bank of England's** growing pessimism about the global outlook.

Forecasters had expected the Bank's monetary policy committee to leave rates unchanged, but fears more cuts would fuel consumer demand and threaten higher inflation.

But the **MPC** said the unexpected severity of the global slowdown means that inflation on track to hit its **2.5 per cent** target in two years.

"Monetary policy needs to balance the weaker external environment but we believe that strong growth," it said. The **MPC** believes a two-speed economy of weak international growth and strong consumer demand is better than an economy in recession.

Since **last month's** **MPC** meeting, which voted eight to one to leave rates unchanged, signs domestic demand is cooling. Retail sales, household borrowing and new house sales are all down.

Task: 162 Document: ft-bank-of-england-02-aug-2001.xml

Annotation Editor Dialog

Type: Address Date Location Money Organization Percent Person Phone

Features: Case Regexp Whole Highlights

Bank of England

First Prev. Next Annotate Ann. all next

Manual Annotation

- Annotator: log into Teamware
 - Click on “Open Annotation Editor”
 - Request an annotation task (first button)
 - Annotate the assigned document
 - Save work in progress occasionally
 - Click on the “Finish task” button when done
 - If you want to save work and return to this task later – “Save” button, then close the UI. Next time you request a task, you get the same document back to finish it.
 - Depending on the project setup, you may be able to reject a document and then ask for another one to annotate (Reject button)
-

Hands-On

- Log out as manager and login as each of the two annotators which you were given, making sure you open the correct Teamware URL in your browser
 - Each manager has their own Teamware, at a different URL, so you need to swap this info as well
 - Open the annotation UI by clicking on the link provided
 - You should now receive a task from the project defined by the respective manager. You need to annotate 2 documents in total (1 from each project/Teamware)
 - If you have put more than 1 document in your corpus, please log back in as a manager and delete them from the corpus!
 - Annotate these documents for Person and Organization names
 - Make a note of any issues you had when annotating and let's discuss them
 - Note that annotators cannot access the manager functionality
-

Going Back to Managing Projects in Teamware

- Log back in with your management account, into your Teamware to continue
-

Monitoring Project Progress

Process Monitoring: Annotation Status

[Detailed View](#)[Back to Project](#)

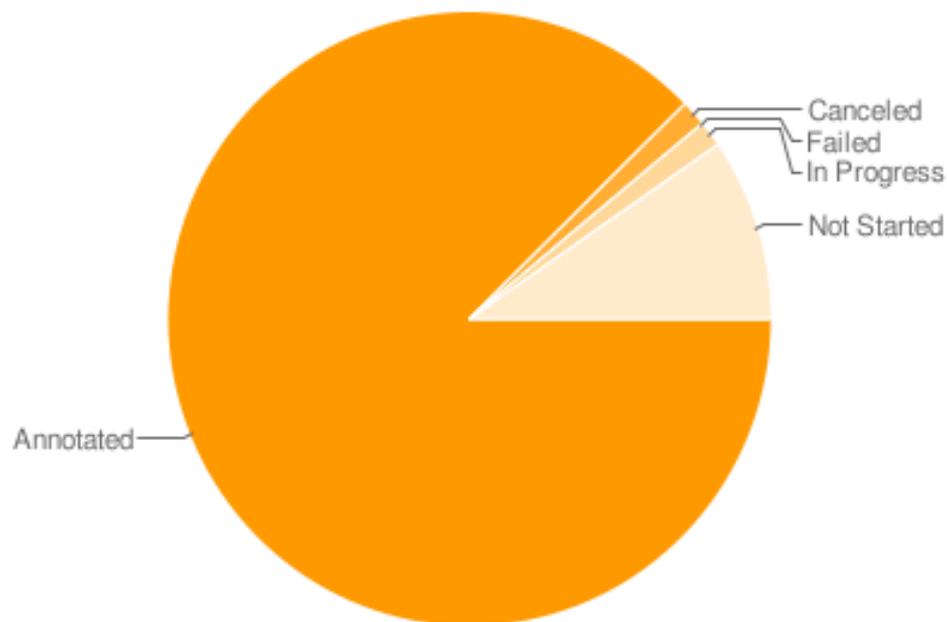
Status	#
Annotated	64
Canceled	1
Failed	0
In Progress	1
Not Started	7

Average Execution Time

627.609375

[Detailed View](#)[Back to Project](#)

Annotation Status Chart



Hands-On

- Log back in as a manager
 - Go to Projects/My Projects
 - Select the Process Button The icon for the 'Process' button is a small graphic showing a green square above a red square, connected by a thin black line that curves from the bottom of the green square to the top of the red square.
 - You will see the manual annotation task
 - Click on Monitoring to see the statistics inside
 - If both your annotators have completed their tasks, you should also receive a notification email
-

Adjudication

- Only available to the manager role
 - Identify differences between annotators using IAA
 - Inspect differences in detail using AnnDiff
 - Edit and reconcile differences if required
 - More sophisticated adjudication UI (the Annotation Stack View) in GATE Developer
-

IAA: Recap

- The IAA on IE tasks, such as named entity recognition, should be measured using f-measure across all annotators
- For classification tasks, use Kappa to measure IAA
- For details, see Module 2 slides and the GATE user guide

IAA: Do my annotators agree?


[Home](#)
[Resources](#) →

[Projects](#) →

[Support](#) →

IAA Caculation on the document ft-airtours-08-aug-2001.xml

Please note you can NOT do caculation if there is no common annotation type on the two selected annotation sets or more.

Available Annotation Sets

- Key
- Original mark
- safe.preproce
- <Default>



Annotation Sets to Compare

- annotator1
- annotator2
- annotator3



Annotation Type

Date

Algorithm

all-ways-f-measure

Feature

Calculate

Cancel

IAA: Results

IAA Calculation on the document ft-airtours-08-aug-2001.xml

Please note you can NOT do calculation if there is no common annotation type on the two selected annotation sets or more.

Available Annotation Sets

Key
annotator2
Original mark
annotator1
annotator3
safe.preproc
<Default>



Annotation Sets to Compare



Annotation Type

Date

Algorithm

all-ways-f-measure

Feature

Calculate

Cancel

Overall F-Measure

Precision	Recall	F1	Precision Lenient	Recall Lenient	F1 Lenient	Correct	Partially Correct	Missing	Spurious
1.0	0.9411765	0.96875	1.0	0.9411765	0.96875	16.0	0.0	1.0	0.0

F-Measure values against the Key annotator1

Response	SP / SR / SF	LP / LR / LF	C / P / M / S
annotator2	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	17.0 / 0.0 / 0.0 / 0.0
annotator3	1.0 / 0.88235295 / 0.9375	1.0 / 0.88235295 / 0.9375	15.0 / 0.0 / 2.0 / 0.0



AnnDiff: Finding the differences

Annotation Diff GUI [Connected in DIRECT mode: ft- airtours-08- aug-2001.xml__1248214749368__610]

Annotation Diff GUI

Document: ft- airtours-08- aug-2001.xml | Annotation Set: annotator1 | Annotation Type: Date | F-Measure Weight: 1.00

Response: ft- airtours-08- aug-2001.xml | Annotation Set: annotator3 | Features: All Some None

Start	End	Key	Features	=?	Start	End	Response
123	136	third quarter	{kind= date}	-?			
180	219	first nine months of its financial year	{kind= date}	-?			
2473	2484	summer 2001	{kind= date, rule1= DateName, rule2= DateOnlyFinal}	=	2473	2484	summer 2001 {kind= date, rule1=
2121	2130	this year	{kind= date}	=	2121	2130	this year {kind= date}
237	255	quarter to June 30	{kind= date}	=	237	255	quarter to June 30 {kind= date}
1204	1220	end of this year	{kind= date}	=	1204	1220	end of this year {kind= date}
2797	2824	second quarter of next year	{}	=	2797	2824	second quarter of next year {}
2409	2418	last year	{kind= date}	=	2409	2418	last year {kind= date}
2270	2281	summer 2001	{kind= date, rule1= DateName, rule2= DateOnlyFinal}	=	2270	2281	summer 2001 {kind= date, rule1=
337	346	last year	{kind= date}	=	337	346	last year {kind= date}
2167	2176	next year	{kind= date}	=	2167	2176	next year {kind= date}
1560	1584	nine months of this year	{kind= date}	=	1560	1584	nine months of this year {kind= date}
2326	2332	winter	{kind= date, rule1= GazDate, rule2= DateOnlyFinal}	=	2326	2332	winter {kind= date, rule1=

Correct: 15 Recall Precision F-Measure

Partially Correct: 0 Strict: 0.8824 1.00 0.9375

Missing: 2 Lenient: 0.8824 1.00 0.9375

False Positives: 0 Average: 0.8824 1.00 0.9375

Document: ft- airtours-08- aug-2001.xml__1248214749368__610

Where are these in Teamware?

- Only visible to managers
 - Resources/Documents menu
 - Select the corpus worked on
 - Iterate through each document
 - Run IAA and AnnDiff, as required
 - These are clumsy, as are on a document by document basis
-

GATE Developer adjudication facilities



- Download the corpus as a zip file
 - Corpus QA Tool
 - A corpus-level view of IAA (F-measure or Kappa)
 - Extended AnnDiff to allow easy reconciliation of the differences between 2 annotators
 - Annotation Stack View to reconcile difference visually, of 2 or more annotators
-



Adjudication in AnnDiff

Annotation Diff Tool

Key doc: ft-pirelli-10-aug-2... Key set: annotator1 Type: Date Weight: 1.0

Resp. doc: ft-pirelli-10-aug-2... Resp. set: annotator2 Features: all some none 1.0 Compare

Start	End	Key	Features	K=?	R	Start	End	Response
590	597	July · 28	{kind= date, rule1= Da...rule2= DateOnlyFinal}	<input type="checkbox"/> =	<input checked="" type="checkbox"/>	590	597	July · 28 {kind= date, r
1313	1321	Thursday	{kind= date, rule1= Ga...rule2= DateOnlyFinal}	<input type="checkbox"/> =	<input checked="" type="checkbox"/>	1313	1321	Thursday {kind= date, r
1262	1266	2002	{kind= date, rule1= Ye...rule2= DateOnlyFinal}	<input type="checkbox"/> ~	<input checked="" type="checkbox"/>	1254	1275	between · 2002 · and · 2004 {kind= date, r
1271	1275	2004	{}	<input type="checkbox"/> -?	<input type="checkbox"/>			

Target set: consensus Copy selection to target set

2 documents loaded Show document Export to HTML

Statistics Adjudication

Adjudication in AnnDiff

- Select the annotations which are correct by ticking the boxes (see previous screen shot)
 - Provide the name of the target consensus set
 - Click on the button to copy them into that set
 - Once copied, they are removed from the list of annotations to adjudicate, so the curator can focus on the remaining ones
 - Adjudication works one annotation type at a time and only for 2 annotators, whose results are stored into 2 different annotation sets
-

Annotation Stack View

Annotation Sets Annotations List Annotations Stack Co-reference Editor OAT Text

But when the Telecom Italia payment of E15.5m was proposed last week, its board rejected it. Pirelli and its ally, the Benetton family, gained control on **July 28** by buying 23 per cent of Olivetti from Bell, an investment group headed by Mr Colaninno and others. The E31m payouts would be in lieu of options and any other compensation due to him. Mr Colaninno was also chief executive of Olivetti.

Several board members objected that Telecom Italia shares had not risen since Mr Colaninno took control and that the figure was too high.

Mr Colaninno had a salary last year of E3.5m from Telecom Italia and E819,000 from Olivetti. He also had cash bonuses of E3.6m from Telecom Italia and E700,000 from Olivetti and other companies controlled by Olivetti.

In addition, he received options to buy 12.6m Olivetti shares **between 2002 and 2004** at E2.81 a share. Olivetti shares on **Thursday** closed at E1.94 compared with the E4.17 – an 80 per cent premium to the market – paid by Pirelli and the Benetton family when they bought their controlling stake. The shares fell when investors realised they would not benefit from a takeover.

Pirelli is expected to bring in a new board this autumn but several members are likely to object again to the payment. The Olivetti board might approve its E15.5m payout but minority shareholders in both companies may

Previous boundary Next boundary

Context 6m Olivetti shares between 2002 and 2004 at E2.81 a share. Olivetti shares on Thu

annotator1#Date   

annotator2#Date  

Original markups

- annotator1
 - Date
 - Money
 - Organization
 - Percent
 - Person
- annotator2
 - Date
 - Money
 - Organization
 - Percent
 - Person

Annotation Stack View

- Open the Document Viewer
 - Select the “Annotations Stack” button
 - Select the annotation types you’d like to reconcile, from as many annotation sets as needed
 - The Stack view window is at the bottom
 - Context shows the part of the document which we are working on now
-

Annotation Stack View

- The Previous/Next boundary buttons position the text we're working on (the Context) on the previous/next annotation of the chosen type
 - Note that such annotation may only exist in one of the sets, i.e., the other annotator might have missed it!
 - Hover the mouse over the coloured annotation rectangles to see further details
 - Right click would open the annotation editor
 - Double click copies the annotation to the target consensus set (you choose which one it is)
-



Annotation Stack View

Mr Colaninno had a salary last year of E3.5m from Telecom Italia and E819,000 from Olivetti. He also had cash bonuses of E3.6m from Telecom Italia and E700,000 from Olivetti and other companies controlled by Olivetti.

In addition, he received options to buy 12.6m Olivetti shares between 2002 and 2004 at E2.81 a share. Olivetti shares on Thursday closed at E1.94 compared with the E4.17 - an 80 per cent premium to the market - paid by Pirelli and the Benetton family when they bought their controlling stake. The shares fell when investors realised they would not benefit from a takeover.

- Date
- Money
- Organization
- Percent
- Person

Copy annotation to another set

- Percent
- Person
- consensus
- Date

In addition, he received options to buy 12.6m Olivetti shares between 2002 and 2004 at E2.81 a share. Olivetti shares on Thursday closed at E1.94 compared with the E4.17 - an 80 per cent premium to the market - paid by Pirelli and the Benetton family when they bought their controlling stake. The shares fell when investors realised they would not benefit from a takeover.

Previous boundary

Context

annotator1: Previous boundary Next boundary Target set: consensus

annotator2: Context 6m Olivetti shares between 2002 and 2004 at E2.81 a share. Olivetti shares on Thu

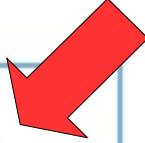
annotator1#Date			
annotator2#Date			
consensus#Date			

Semi-automatic Projects

- You can pre-process the corpus with pipelines in GATE Developer and then have people improve the annotations manually
 - Make sure in GATE Developer that you store all automatically produced annotations in an annotation set called **safe.preprocessing**
 - Then in Teamware, when you specify the manual workflow, make sure that you select to run the pre-manual service, which is called **gas:pre-manual gas**
 - This service will copy all annotations from the safe.preprocessing set to the annotation set storing the results from the manual annotation, for each of the annotators. This is done at the moment when annotatorX requests this document for annotation, not earlier.
-

Configuration: "manual annotation"

Next

Pre-Manual Service:*	gas: pre-manual gas ▼ 
Annotators per Document:	2 ▼
Cancel Task Allowed:	<input checked="" type="checkbox"/>
Anonymous Annotation:	<input checked="" type="checkbox"/>
Annotation Schemas:*	Address.xml Money.xml Percent.xml Organization.xml Location.xml Date.xml
Add Schema:	Add Schema

Hands-On

- As a manager, create a new corpus and upload the documents contained in the resources folder, annie-corpus.zip
 - This contains documents pre-processed with ANNIE, with the output saved in the safe.preprocessing set. You can inspect this from the corpus viewer
 - Create a new workflow template, still selecting only “Manual annotation” in the first step
 - Configure the pre-manual gas service
 - Run the annotation project as normal
 - When you login as test-annotator1, for example, you will now see all automatically created annotations which you will then have to edit to create the gold standard
-

● Annotator GUI [Connected to POOL mode: admin]

Document Editor

Type	Set	Start	End	Id	Features
Organization		23	40	3589	{}
Organization		259	262	3592	{}
Organization		271	278	3593	{}
Organization		283	290	3594	{}
Organization		482	488	3595	{}
Organization		514	524	3596	{}

30 Annotations (0 selected) Select:

Semiconductor designer **ARC International** said yesterday that sales levels remained stagnant and it would no longer give revenue forecasts. Shares in the company closed up 2p at 52.2p despite analyst concerns that the company might not break even until 2005.

ARC follows **Parthus** and **Bookham** in giving a cautious outlook on the microchip market. About 60 of the 290 staff face the axe after results showing a pre-exceptionals net loss of £6.1m for the quarter on revenues of £3.5m.

Amazon group to sue

A group of **Amazon.com** shareholders has filed a class action suit in the US district court for the southern district of New York against **Morgan Stanley** and its internet analyst **Mary Meeker**, charging them with issuing "false and misleading" statements on **Amazon** stock.

The suit, filed on behalf of **Amazon** shareholders who purchased shares between August 1 1998 and January 22 2001, claims that **Morgan Stanley** failed to keep a "Chinese wall" between its investment banking and equity research activities.

Task: 4 Document: ft-BT-briefing-02-aug-2001.xml__1304973189093__2157



How can I use Teamware?

- Available from <http://gatecloud.net/> for experimentation and use