

GATE Mimir and cloud services

Multi-paradigm indexing and search tool

Pay-as-you-go large-scale annotation



GATE Mimir

- GATE Mimir is an indexing system for GATE documents.
- Mimir can index:
 - Text: the original document content is indexed (based on Token annotations)
 - Annotations: annotations and features
 - Semantics: annotations can be linked to external ontologies which can be used at search time
- Mimir queries allow for any combination of these



Why Mimir?

- Standard search covers the text.
- GATE documents also have annotations, which give access to the document's:
 - Structure (sections, titles, etc.)
 - Linguistic features (nouns, verbs, etc.)
 - Semantics
 - Etc.
- Examples: <http://demos.gate.ac.uk/mimir>
 - BBC News demo



Mimir query language

- Simplest queries are free text
 - *text*
 - “quoted string”
- Searching against the document text



Plain text queries

Harriet Harman

Search

Documents 1 to 20 of 81:

UK childcare needs to be more affordable - CentreForum (cached)

quality' MP **Harriet Harman** was the architect

Birth weight among social mobility checks - Nick Clegg (cached)

's deputy leader **Harriet Harman** said Mr Clegg

Ed Miliband's shadow cabinet and ministerial teams (cached)

Miliband Opposition leader **Harriet Harman** Deputy Leader & in 2011. **HARRIET HARMAN** - DEPUTY LEADER

PM's response to Skinner Commons question 'shameful' (cached)

. Deputy leader **Harriet Harman** wrote on Twitter

Daily Politics and Sunday Politics highlights of 2012 (cached)

Sunday April 29 **Harriet Harman** on Hunt, Cameron and Clegg **Harriet Harman** struggles with bank on health by **Harriet Harman**
PMQs: Harriet ...

Leveson Inquiry: Jeremy Hunt 'sought News Corp guidance' (cached)

Shadow Culture Secretary **Harriet Harman** says Jeremy Hunt culture secretary, **Harriet Harman**, told the

No Rupert Murdoch deal says Alastair Campbell (cached)



Plain text queries



Harriet Harman says

Search

Documents 1 to 20 of 29:

Leveson Inquiry: Jeremy Hunt 'sought News Corp guidance' (cached)

Shadow Culture Secretary **Harriet Harman** says Jeremy Hunt was

Ed Miliband defends Iraq war condemnation (cached)

, says Harman **Harriet Harman** says Labour will be

Ed Miliband tells Labour: We're the optimists now (cached)

, says Harman **Harriet Harman** says Labour will be

Labour must have credible deficit plan, says Darling (cached)

, says Harman **Harriet Harman** says Labour will be

David Miliband says he won't join brother Ed's team (cached)

, says Harman **Harriet Harman** says Labour will be

Balls: Labour must fight cuts 'every inch of the way' (cached)

, says Harman **Harriet Harman** says Labour will be

Rob Ainsworth clashes with ex-aide over Trident (cached)



Token features

- Free text queries are actually searching the `string` features of the GATE Token annotations
- Can also search other token features, e.g. `root` (morphology) or `category` (POS)



Morphology

Harriet Harman root:say

Search

Documents 21 to 18 of 38:

[David Cameron criticised for 'calm down dear' jibe \(cached\)](#)

former equality minister **Harriet Harman** said Mr Cameron's

[Queen's Speech: Biggest change to voter registration \(cached\)](#)

, Labour's **Harriet Harman** said the government was

[Harriet Harman struggles with bank bonus and job figures \(cached\)](#)

in Coventry, **Harriet Harman** said: "I

[PMQs: Harriet Harman and Nick Clegg on unemployment \(cached\)](#)

Labour, but **Harriet Harman** said unemployment was falling

[Leveson Inquiry: Jeremy Hunt fair on BSkyB, says top civil servant \(cached\)](#)

Shadow culture secretary **Harriet Harman** said: "David

[Jeremy Hunt: I followed due process over BSkyB \(cached\)](#)

But Labour's **Harriet Harman** said Mr Hunt had

[Ed Miliband 'will marry' but politics 'got in the way' \(cached\)](#)

, says Harman **Harriet Harman** says Labour will be



Gaps

- Default combinator is *sequence* – terms must be adjacent
 - Different from typical search engines
- Can allow gaps with $[n..m]$
- Arbitrary gap with AND
 - $x \text{ AND } y$ finds the shortest span that covers both



Gaps



Harriet Harman [0..5] root:say

Search

Documents 1 to 20 of 43:

[Birth weight among social mobility checks - Nick Clegg](#) (cached)

's deputy leader **Harriet Harman said** Mr Clegg would

[Leveson Inquiry: Jeremy Hunt 'sought News Corp guidance'](#) (cached)

Shadow Culture Secretary **Harriet Harman says** Jeremy Hunt was

[Jeremy Hunt quit call after Leveson BSkyB evidence](#) (cached)

Labour deputy leader **Harriet Harman also said** Mr Hunt should

[000375_http://www.bbc.co.uk/news/uk-14481315](http://www.bbc.co.uk/news/uk-14481315) (cached)

Miliband and deputy **Harriet Harman are in effect saying** Prime Minister David



Annotations

- So far nothing a standard search engine couldn't do...
- But Mimir also indexes annotations
- Syntax:
 - `{AnnotationType feat1=val1 feat2=val2}`
 - Feature comparisons can be =, <, <=, >=, >
 - UI offers pop-up with available types/features
- Let's generalise – *any* person, not just Harriet



Annotations



```
{Person} [0..5] root:say
```

Documents 1 to 20 of 5495:

[Apple's Sir Jonathan Ive reaffirms desire to stay at company \(cached\)](#)

Today programme, **Sir Jonathan said** he would stay partner". **Sir Jonathan said** that Apple's

[Diamond Jubilee Tube train was faulty \(cached\)](#)

be happening' **Ms Siggs said**: "It

[Warning over deep-ocean stowaways \(cached\)](#)

using the famous **Alvin sub say** the vehicle picked embarrassment. But **Dr Voight says** the experience is it," **Dr Voight said**. "We

[School building system not fit for purpose, review says \(cached\)](#)

government-commissioned review by **Sebastian James of Dixons Group said** value for money by Education Secretary **Michael Gove, Mr James said**. Schools with shadow education secretary **Andy Burnham said** Mr Gove had ...



Other operators

- Containment (use parentheses to group)
 - Query1 IN Query2
 - Query1 OVER Query2

```
( {Person} [0..5] root:say ) IN {Content}
```

Search

Documents 1 to 20 of 4924:

[Apple's Sir Jonathan Ive reaffirms desire to stay at company \(cached\)](#)
Today programme, **Sir Jonathan said** he would stay partner". **Sir Jonathan said** that Apple's

[Diamond Jubilee Tube train was faulty \(cached\)](#)
be happening' **Ms Siggs said**: "It



Other operators

- Alternative
 - Query1 OR Query2
 - Query1 | Query2

```
{(Person} | {Organization}) [0..2] root:say
```

Search

Documents 1 to 20 of 8611:

[Apple's Sir Jonathan Ive reaffirms desire to stay at company \(cached\)](#)

Today programme, **Sir Jonathan said** he would stay partner". **Sir Jonathan said** that Apple's

[Diamond Jubilee Tube train was faulty \(cached\)](#)

for London (TfL) **has said**. Passengers were be happening' **Ms Siggs said**: "It broke down. **London Ambulance Service said** two ambulance crews ...



Other operators

- Set difference
 - `Query1 MINUS Query2`
 - Returns all spans that match Query1 but are not also matches of Query2
 - E.g. sentences that *don't* mention a location

```
{Sentence} MINUS  
( {Sentence} OVER {Location} )
```



Try it!

-
- <http://demos.gate.ac.uk/mimir>
 - BBC News demo
 - Find:
 - Document titles
 - Date expressions
 - Amounts of money **being paid**
 - Amounts of money **being received**
 - Hint: if you get too much noise, try
 - restricting to matches within a sentence
 - using `IN {Content}` to ignore boilerplate



Semantics

- Annotations may be linked to a knowledge base, e.g. DBpedia (<http://dbpedia.org>)
- Annotation refers to an instance
 - http://dbpedia.org/resource/Harriet_Harman
- KB knows that this instance belongs to the class of politicians (and people, ...)
 - <http://dbpedia.org/ontology/Politician>
- SPARQL query language can retrieve instances that match constraints
- More tomorrow...



Semantics

- In news demo, Person, Location and Organization have `class` and `inst` features

```
{Person inst="http://dbpedia.org/resource/Harriet_Harman"}
```

Search

Documents 1 to 20 of 81:

[UK childcare needs to be more affordable - CentreForum](#) (cached)

quality' MP **Harriet Harman** was the architect

[Birth weight among social mobility checks - Nick Clegg](#) (cached)

's deputy leader **Harriet Harman** said Mr Clegg



Semantic search

- Can use SPARQL to query the KB at search time
- E.g. to find *all* politicians

```
{Person sparql="
SELECT DISTINCT ?inst WHERE {
?inst a :Politician }"}
```



Semantic search



```
{Person sparql="SELECT DISTINCT ?inst WHERE { ?inst a :Politician }"}  
/
```

Search

Documents 1 to 20 of 4373:

[000001_http://www.bbc.co.uk/news/](http://www.bbc.co.uk/news/) (cached)

Climate Change minister **Gregory Barker** is making a

[In pictures: Royal arts gathering](#) (cached)

Shirley Bassey and **Joan Collins** were thrilled to

[EU wants Greece to stay in eurozone, says Van Rompuy](#) (cached)

European Council President **Herman Van Rompuy**: "The , European Council **President Herman** Van Rompuy has European Council President **Herman Van Rompuy** has said.

[000022_http://www.bbc.co.uk/news/world/europe/](http://www.bbc.co.uk/news/world/europe/) (cached)

order EX-IMF head **Dominique Strauss-Kahn** asks a French



Custom UI

- Not the sort of query you want to construct by hand...
- Mimir provides an XML-over-HTTP query API to allow programmatic querying
- Can build custom UIs that hide the query language from users
 - <http://demos.gate.ac.uk/pin>
 - Try out the query builder, look at the underlying query



Building an index - you need:

- Some annotated GATE documents
- A description of which annotations and features you want to index
 - “Index template”
 - Only the features you specify in the template will be available for searching
- A running instance of the Mímir webapp
 - You can download and build your own (requires Grails)
- A way to push the documents to the server



Pushing documents

- GATE PR
 - “Mímir indexing PR” available as part of the Mímir source distribution
- GCP – the “GATE Cloud Parallelizer”
 - Tool to deploy a saved GATE application multi-threaded on your own machine
 - Includes various “output handlers” to save annotations to disk, or push them into Mímir
 - <http://gate.ac.uk/gcp>
- or let us do it for you...



GATECloud.net

- A cloud based service from the GATE team
- Usual cloud benefits:
 - Pay-as-you-go, no upfront hardware costs
 - No sysadmin work
 - Web-based management tools
 - Always latest version, maintained by us
- Not-so-usual cloud benefits
 - Based on open-source software
 - Bring your own pipeline

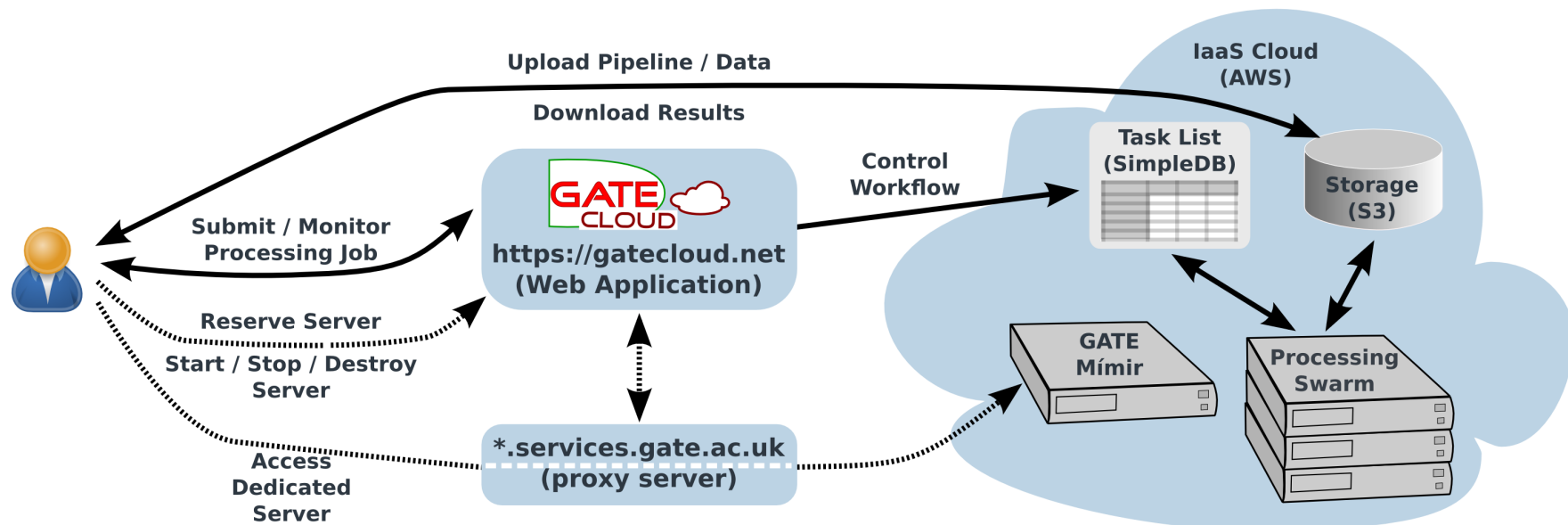


Features

- On demand document processing (a.k.a. *Annotation Jobs*)
 - Parallel processing on Amazon EC2
 - On-line job definition tool
 - Many output formats, including Mimir
- On demand servers, including Mimir
- Top up your account with vouchers from the University online shop



Architecture





Dedicated servers

- Rent a dedicated Mimir server for your private use
- Start and stop it as required
- Pay only for the hours it is running
- Data (i.e. indexes) persistent across reboots
- Backup and restore facility available



The shop

The screenshot shows a web browser window with the URL `gatecloud.net/shopfront`. The page features the GATE Cloud logo, navigation links for Home, Shop, and Dashboard, and links for Log in, Register, and Help. A shopping basket icon shows it is empty. The main content area is titled "GATECloud Products" and lists four items with their prices and descriptions. A "Discounts" section is also present, listing conditions for a discount and a "send us" link. An illustration of a mug with the GATE logo is shown next to the discounts.

Product Name	Description	Price
Annotation Job - Custom	Execute your own pipeline on the GATE cloud.	GBP0.00 (plus GBP0.99 per hour)
ANNIE (Named entity annotation service)	Upload your own documents and run our pre-packaged named entity annotator (ANNIE) on GATE Cloud	GBP0.00 (plus GBP1.49 per hour)
ANNIE plus Measurements and Numbers (annotation service)	Upload your documents and run our pre-packaged named entity annotator with the Numbers and Measurements add-ons	GBP0.00 (plus GBP1.75 per hour)
GATE Mimir 5.0 Server	Multiparadigm indexing server	GBP99.00 GBP0.00 (plus GBP0.99 per hour)

Discounts:

- high volume
- research or non-profit use
- to apply create an account and [send us](#) your details



Reserving a server

- The usual e-commerce experience
 - Sign up for an account
 - Buy a top-up voucher
 - Add item(s) to your basket
 - Checkout to complete the order
- Server appears in your *dashboard*
- Behind the scenes, creates a persistent data *volume* for your data



Dashboard

Gatecloud.net dashboard screenshot showing user Ian Roberts, navigation menu (Home, Shop, Dashboard), and sections for Cloud Machines and Annotation Jobs.

Cloud Machines

Recently purchased machines may take a few minutes to appear in this list.

[Filter view...](#)

Reservation ID	Machine Type	State	Order Number	Order Date
R-000178	Mimir Server 5.0	inactive	O-000350	June 11, 2014

Annotation Jobs

[Filter view...](#)

Name	Created At	State
------	------------	-------



Reservation control panel

The screenshot shows a web browser window with the URL `gatecloud.net/yourAccount/machineReservationDetails/178`. The page header includes the GATE Cloud logo, the user's name "Ian Roberts's account", and links for "Log out" and "Help". A navigation bar contains "Home", "Shop", and "Dashboard" icons, along with a "Your account" link.

Machine Reservation R-000178

ID	R-000178	Destroy Reservation
Machine type	GATE Mimir 5.0 Server	
Hourly price	GBP0.99	
State	inactive	Start Instance
Instance ready	no	

Backups

Slot 1	<empty>	Create new backup
Slot 2	<empty>	Create new backup

Reservation Details:

URL1:	<code>https://mimir-2z17eea4hs.services.gate.ac.uk/</code>
Admin account name:	<code>admin</code>
Admin account password:	<code>12345678</code>



Controlling the server

- Start and stop instance
 - Startup/shutdown takes a few minutes – system will email you when server is ready
 - You pay the hourly price whenever the instance is running
- Backup and restore
 - Save the state of your data volume so you can roll back later
- Destroy reservation
 - If you no longer need the server, destroy it to discard the data volume and all backups
 - *This cannot be undone*

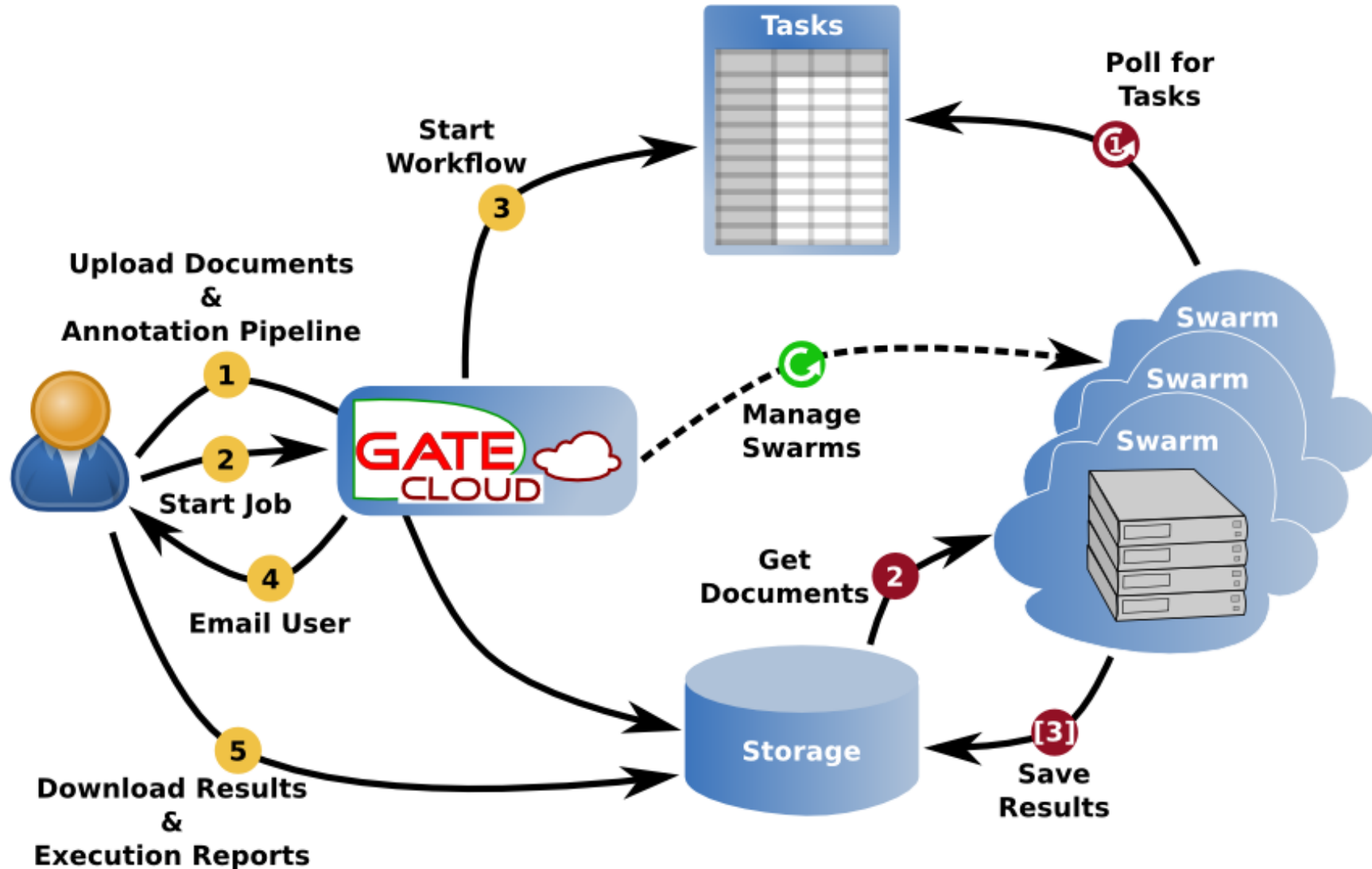


Annotation jobs

- Parallel and distributed annotation of documents with a GATE application
- Upload your own documents
 - zip, tar, arc/warc archives
- Upload your own pipeline
 - “Export for GATECloud.net”
- ... or use a standard one
- Output annotations in various formats, or send documents directly to Mimir



Job lifecycle





Execution environment

- Amazon EC2
- Ubuntu LTS, 64-bit
- Oracle Java 7
- ~2GB/thread RAM on average
- GCP 2.4, based on GATE Embedded 8.0



Reserving a job

- Same process as servers
- Choose the job you want from the shop
- Add to basket
- Checkout
- Job appears in dashboard



Managing a job - application

Annotation Job "Annotation Job - Custom"

Details	Inputs	Outputs	Execution Progress	Download Results	Status
<p>Use this tab to specify details about your job. Click the "Change" button below to change the GATE application to be used for this job.</p>					<ul style="list-style-type: none">✓ An application file has been provided.✓ Some input documents have been provided.✓ Some outputs have been defined.
Job Name:	Annotation Job - Custom	<input type="button" value="Rename"/>			
Application File:	annie-with-morph.zip	<input type="button" value="Change"/>			
Job State:	Job fully defined and ready to be started.				
ID:	J-000303				
UUID:	a5a6b952-bd62-4059-9631-0af130ec9956				
Date Created:	11 June 2014 22:44:47 BST				
<input type="button" value="Advanced options"/>					



Managing a job - input

Annotation Job "Annotation Job - Custom"

Details

Inputs

Outputs

Execution Progress

Download Results

Status

Use this tab to specify which documents should be processed.
Select one or more inputs in the list on the left to modify their parameters.

news-corpus-large.zip (Type: ZIP)
 Input Filters: none (process all)
 Input mime type: text/html
 Input encoding: UTF-8

Add New

Delete Selected

Options for the selected inputs:

Type:

File Suffixes:

Mime Type Override:

Encoding:

Cancel Changes

Apply Changes

- ✔ An application file has been provided.
- ✔ Some input documents have been provided.
- ✔ Some outputs have been defined.

Start Job

Stop Job

Delete Job



Outputting to Mimir

- Start up the Mimir server we reserved earlier
- Create an appropriate index template
- Create an index



Create a template

The screenshot shows a web browser window with the URL `mimir-2z17eea4hs.services.gate.ac.uk/admin/indexTemplate/create`. The page title is "Create IndexTemplate".

Form fields:

- Name:
- Comment:
- Configuration:

```
semanticAnnotations = {  
  index {  
    annotation helper:new DefaultHelper(annType:'Sentence')  
    annotation helper:new DefaultHelper(annType:'Document',  
mode:Mode.DOCUMENT)  
  }  
  index {  
    annotation helper:new DefaultHelper(annType:'Person',  
nominalFeatures:['gender'])  
    annotation helper:new DefaultHelper(annType:'Location',  
nominalFeatures:['locType'])  
    annotation helper:new DefaultHelper(annType:'Organization')  
  }  
  index {  
    annotation helper:new DefaultHelper(annType:'Percent')  
    annotation helper:new DefaultHelper(annType:'Money')  
    annotation helper:new DefaultHelper(annType:'Date',  
nominalFeatures:['kind'])  
    annotation helper:new DefaultHelper(annType:'Address',  
nominalFeatures:['kind'])  
  }  
}  
documentRenderer = new OriginalMarkupMetadataHelper()  
documentMetadataHelpers = [documentRenderer]
```

At the bottom of the form is a "Create" button. The footer of the page reads "Mimir 5.0, © GATE 2014."



Create an index

A screenshot of a web browser window showing the 'Create Local Index' page in the Mimir admin interface. The browser's address bar shows the URL 'mimir-2z17eea4hs.services.gate.ac.uk/admin/localIndex/create'. The page features a 'Search' header with a magnifying glass icon and the text 'powered by Mimir'. A user is logged in as 'admin' with a '(Log out)' link. Below the header is an 'Admin Home' link. The main content area is titled 'Create Local Index' and contains a form with the following fields: 'Name' (text input with 'fig7'), 'Index template' (dropdown menu with 'annie'), 'Document URIs are external links' (checkbox), 'Sub-bindings enabled' (checkbox), and 'Index ID (optional)' (text input). A 'Create' button is located at the bottom of the form. The footer of the page reads 'Mimir 5.0, © GATE 2014.'

Machine Reservation R-000 Create Local Index

mimir-2z17eea4hs.services.gate.ac.uk/admin/localIndex/create

Search powered by Mimir

You are logged in as **admin**.
(Log out)

Admin Home

Create Local Index

Name:

Index template:

Document URIs are external links:

Sub-bindings enabled:

Index ID (optional):

Create

Mimir 5.0, © GATE 2014.



Index details – note URL

The screenshot shows a web browser window with the following content:

- Browser tabs: "Machine Reservation R-000" and "Mimir index 'fig7'"
- Address bar: `mimir-2z17eea4hs.services.gate.ac.uk/admin/actions/97ebd97c-f44e-4453-9c8b-fe877bd096e0/ac`
- Page header: "Search" logo on the left, "Mimir powered by" logo on the right, and "You are logged in as admin. (Log out)" text.
- Navigation: "Admin Home" link.
- Section: "Mimir index 'fig7'" (highlighted in a blue bar).
- Index Name: `fig7`
- Index URL: `http://ip-10-34-141-217.eu-west-1.compute.internal:8080/97ebd97c-f44e-4453-9c8b-fe877bd096e0` (circled in red)
- Text: "This URL can be used for API access to the index, for example when creating a federated index or when configuring the output of a GATECloud.net Annotation Job."
- Text: "When running inside the GATECloud.net cloud, the URL of your index changes every time you restart your Mimir server! If you use it to create a federated index, then you will need to re-create your Remote index instance every time you restart."
- Text: "For security reasons, the firewall only allows this URL to be accessed by other Mimir servers that you own (so that you can create federated indexes) and by your Annotation Jobs. If you have other requirements please contact the GATECloud.net team."
- State: `ready`
- Annotations indexed: [Detail...](#)
- Scorer: `No Scoring`
- Links: [Search this index using the web interface.](#), [Search this index using the XML service interface.](#), [Manage deleted documents.](#)



Managing a job - output

Annotation Job "Annotation Job - Custom"

Details Inputs **Outputs** Execution Progress Download Results Status

On this tab you can specify what should happen to the output of your annotation job.
 Select one of the outputs on the left to change its parameters.

MIMIR: http://ip-10-34-141-217.eu-west-1. No filters: all annotations are saved!	Details for the selected output: Type: MIMIR Index URL: <u>http://ip-10-34-141.</u> Username: <input type="text"/> Password: <input type="password"/> Output filters (which annotations should be saved): No current filters are defined: all annotations will be saved! Annotation Set: Annotation Type: <input type="text"/> <input type="text"/> Add New Filter
--	--

Cancel Changes Apply Changes

Add New Delete Selected

Status

- ✓ An application file has been provided.
- ✓ Some input documents have been provided.
- ✓ Some outputs have been defined.

Start Job Stop Job Delete Job



Start the job

Annotation Job "Annotation Job - Custom"

Details Inputs Outputs **Execution Progress** Download Results

Status

On this tab you can review the execution progress of your job.
No need to refresh: the execution log is **refreshed automatically**.

Job progress N/A

Execution Log

Time	Message
2014 June 11 22:44:47 UTC+1	Job purchased.

- ✓ An application file has been provided.
- ✓ Some input documents have been provided.
- ✓ Some outputs have been defined.

Start Job

Stop Job

Delete Job



When job completes...

Annotation Job "Annotation Job - Custom"

Details
Inputs
Outputs
Execution Progress
Download Results
Status

On this tab you can review the execution progress of your job. No need to refresh: the execution log is **refreshed automatically**.

Job progress 100%

Execution Log

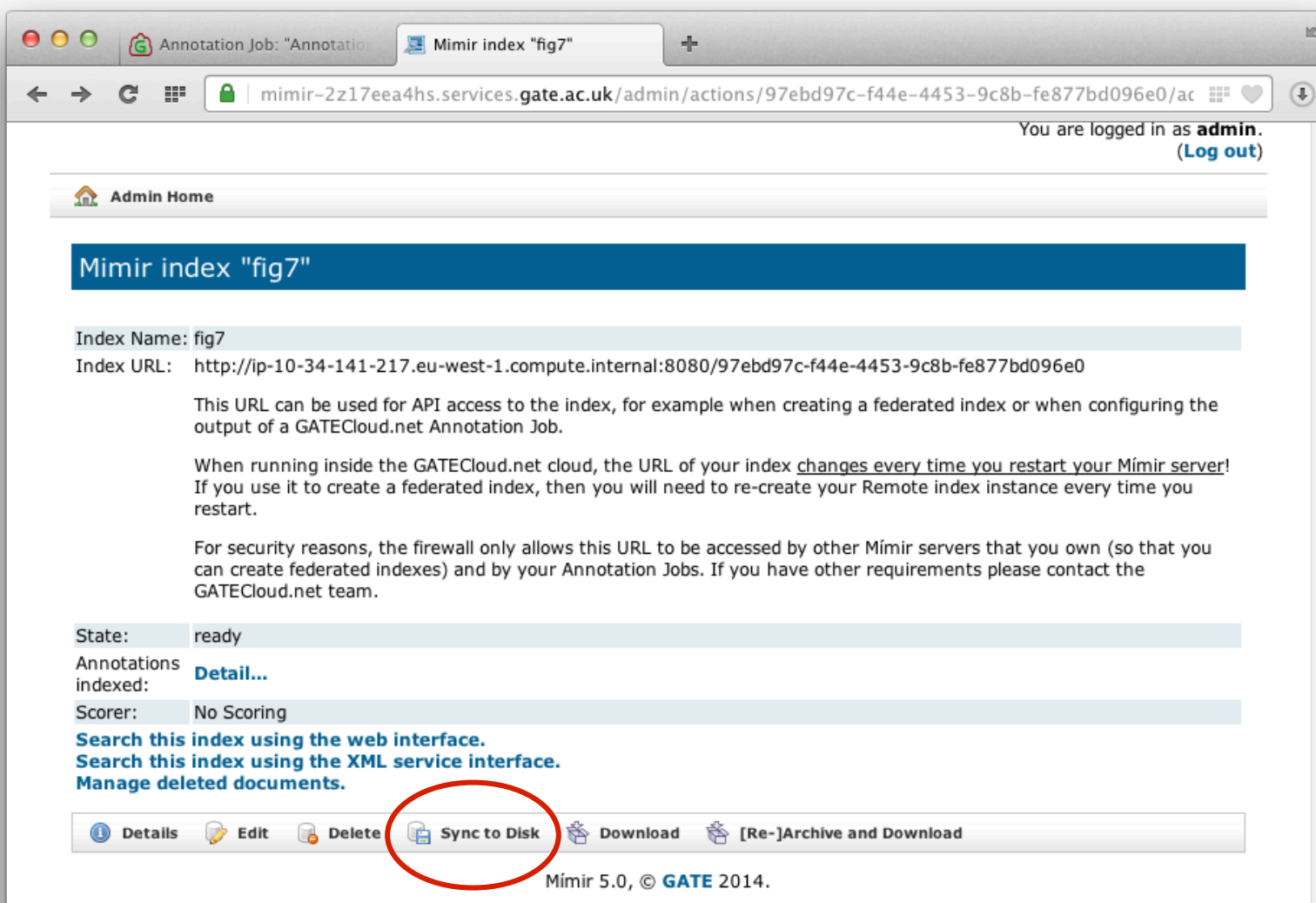
Time	Message
2014 June 11 23:15:20 UTC+1	Job completed.
2014 June 11 23:15:20 UTC+1	Task 'join' completed with an exit status of 'success', execution time: 0 seconds.
2014 June 11 23:14:20 UTC+1	Task 'gcp' completed with an exit status of 'success', successful documents: 761, failed documents: 0, execution time: 2 minutes 30 seconds.
2014 June 11 23:11:20 UTC+1	Task 'split' completed with an exit status of 'success', execution time: 0 seconds.
2014 June 11	Job started

✓ The job has finished running.

- Start Job
Stop Job
Reset Job
Delete Job



“Sync” the index



The screenshot shows a web browser window with the following elements:

- Browser Tabs:** "Annotation Job: 'Annotation'" and "Mimir index 'fig7'".
- Address Bar:** `mimir-2z17eea4hs.services.gate.ac.uk/admin/actions/97ebd97c-f44e-4453-9c8b-fe877bd096e0/ac`
- User Status:** "You are logged in as **admin**. (Log out)"
- Page Header:** "Admin Home" with a home icon.
- Section Header:** "Mimir index 'fig7'" in a blue bar.
- Index Information:**
 - Index Name: fig7
 - Index URL: `http://ip-10-34-141-217.eu-west-1.compute.internal:8080/97ebd97c-f44e-4453-9c8b-fe877bd096e0`
- Usage Instructions:**
 - "This URL can be used for API access to the index, for example when creating a federated index or when configuring the output of a GATECloud.net Annotation Job."
 - "When running inside the GATECloud.net cloud, the URL of your index changes every time you restart your Mimir server! If you use it to create a federated index, then you will need to re-create your Remote index instance every time you restart."
 - "For security reasons, the firewall only allows this URL to be accessed by other Mimir servers that you own (so that you can create federated indexes) and by your Annotation Jobs. If you have other requirements please contact the GATECloud.net team."
- Index Status:**
 - State: ready
 - Annotations indexed: [Detail...](#)
 - Scorer: No Scoring
- Actions:**
 - [Search this index using the web interface.](#)
 - [Search this index using the XML service interface.](#)
 - [Manage deleted documents.](#)
- Toolbar:** Details, Edit, Delete, **Sync to Disk** (circled in red), Download, [Re-]Archive and Download.
- Footer:** Mimir 5.0, © GATE 2014.



“Sync” the index

- Mimir accumulates documents in RAM
- Documents saved to disk after (by default) one hour, or when memory threshold reached
- Documents become searchable once saved to disk
- “Sync” button forces an immediate save, if you know no more documents due
 - Can continue to send more documents, but only sync-ed ones available for search



Search your new index

The screenshot shows a web browser window with the following elements:

- Browser Tabs:** "Annotation Job: 'Annotation'" and "Mimir Index 'fig7'".
- Address Bar:** `mimir-2z17eea4hs.services.gate.ac.uk/97ebd97c-f44e-4453-9c8b-fe877bd096e0/search/index#qui`
- Page Header:** "Search" with a magnifying glass icon and "powered by Mimir".
- User Status:** "You are logged in as admin. (Log out)".
- Search Bar:** A text input field containing the query `{Person gender = "female"}` and a "Search" button.
- Results Header:** "Documents 1 to 20 of 760:"
- Search Results:**
 - Result 1: [www.guardian.co.uk/business/2009/dec/20/shoppers-defy-big-freeze?](http://www.guardian.co.uk/business/2009/dec/20/shoppers-defy-big-freeze?INTCMP=ILCNETTXT3487)
[INTCMP=ILCNETTXT3487](#)
39 5. **Sarah Raven's** Wild Flowers Wild Flowers by **Sarah Raven** £29.
 - Result 2: [www.guardian.co.uk/business/2008/dec/27/high-street-retailers-retail?](http://www.guardian.co.uk/business/2008/dec/27/high-street-retailers-retail?INTCMP=ILCNETTXT3487)
[INTCMP=ILCNETTXT3487](#)



Try it!

- Sign up for an account on <https://gatecloud.net>
- Use your voucher code
- Reserve a Mimir 5.0 server
- Start it up, log in
- Create a new index template using the contents of `index-template.groovy`
- Create a new local index using this template
- Visit index admin page and note the URL



Try it!

- Reserve a “custom annotation job”
- Application zip file is annie-with-morph.zip
- Input is news-corpus-large.zip
 - Mime type: text/html, Encoding: UTF-8
- Set one output to MIMIR, using the Index URL you noted above
 - Make sure to not include any spaces in the index URL
- Run the job, and when finished sync the index



Try it!

- Try some searches on your new index
 - E.g. stock price movements
{Organization} (up | down) ({Money} | {Percent})
- When finished, make sure you stop the Mimir server and destroy the reservation



AnnoMarket.com

- New development based on GATECloud.net, adding
 - A much wider range of pre-packaged pipelines
 - Simplified UI, plus REST APIs for job management and processing single documents
 - “Test this pipeline” function to quickly try different pipelines on your text
 - Access to crawled web data



AnnoMarket.com

The screenshot shows the AnnoMarket.com website interface. At the top, there is a navigation bar with 'Home', 'Shop', 'Data sources', and 'Dashboard'. Below this is a search bar and a list of filters for various languages and tools. The main content area displays six service cards, each with an icon, a description, and a price.


Log In (or Register) | Help

Marketplace

Home Shop Data sources Dashboard


Show only items tagged: Arabic Biomed Bulgarian Chunker Danish DBpedia Dutch English Finnish Freebase French Genes GeoNames German Hungarian Hunpos Italian LOD Measurements Morphology Named Entity News Norwegian OpenNLP Opinion Mining Parser Part-of-Speech Politics Portuguese Romanian Russian Spanish Stanford Parser Stemmer Swedish Tagger Turkish Twitter

ANNIE Named Entity Recognizer

 The ANNIE named entity recognition service. Identifies names of *persons*, *locations*, *organizations*, as well as *money amounts*, *time and date expressions*.


EUR0.80 / CPU hour

AnnoMarket News Pipeline

 The AnnoMarket News Pipeline is a named entity recognition pipeline dedicated to the news domain. It combines text mining tools from the University of Sheffield with the Press Association's extensive knowledge base of newsworthy entities.


EUR0.80 / CPU hour

TwitIE Named Entity Recognizer for Tweets


 Named entity recognition service for Twitter data. Identifies *person*, *location*, *organization* etc. and also performs normalization of abbreviations and common shorthands ("brb", "gr8", "2day", etc.).

EUR0.80 / CPU hour


AnnoMarket Political Speech Categoriser

 The AnnoMarket Political Speech Categoriser is a text classification pipeline dedicated to analysing political speech. It will attempt to associate each input sentence with one of the categories it was trained to recognise.

Generic Opinion Mining (English)

 Recognises "opinionated" sentences in English text or Tweets and classifies them as positive or negative, with a score indicating the strength of the sentiment. This pipeline was developed by the **Arcomem** project.

UK political entities and keywords

 Annotates various terms relevant to UK political news, including names of all current (and some former) politicians, political offices, government departments, constituencies and significant political buildings.



Test this pipeline

AnnoMarket News Pipeline

annomarket.com/shopfront/displayItem/3

Test this pipeline

Type the content to annotate:

The chancellor is also expected to establish a joint review by the Treasury, Bank of England and the Financial Conduct Authority (FCA) to look into how these markets operate.

Mr Osborne will say: "The integrity of the City matters to the economy of Britain.

"Markets here set the interest rates for people's mortgages, the exchange rates for our exports and holidays, and the commodity prices for the goods we buy. I am going to deal with abuses."

A crackdown on malpractice in the financial industry has been gathering pace.

Or select a text file: No file chosen

Output type:

Document format:

Content Date DateRange Location Measurement Money Number Organization Percent Person Ratio Sentence Token

[download](#)

Annotation colors **Osborne** pledges criminal action against banks and traders More action to tackle wrongdoing in the financial sector will be unveiled by Chancellor **George Osborne** on **Thursday**. The measures will include making the manipulation of the foreign currency markets by banks a criminal offence. In a keynote speech he will promise to target "the unacceptable behaviour of the



For developers

- REST APIs for integration with other systems
- Sell your own pipelines on the platform and take a cut of the proceeds
 - Talk to us for more details
- Vouchers available if you want to try it!



Questions?



- More info
 - <https://gatecloud.net>
 - <https://annomarket.com>