

Infrastructure for Human Language Technology and the Semantic Web

Http://gate.ac.uk

Open Source
Scalable
Robust

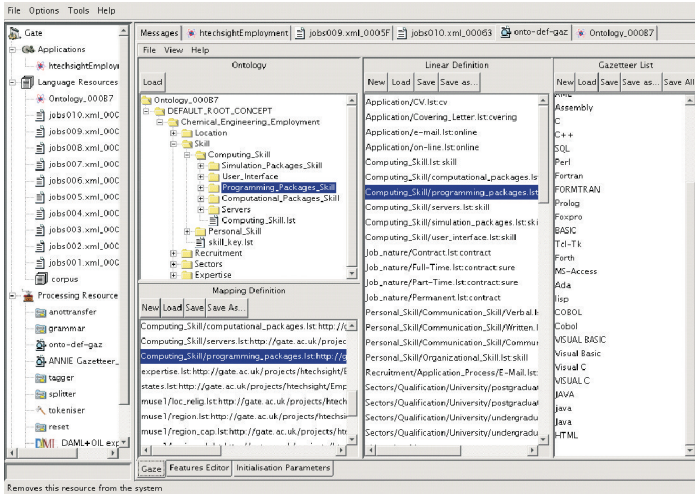
Stable
Customisable
Reusable

100% Java
XHL, HTML, RTF, ...
Oracle, PostgreSQL backends

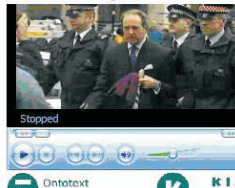
Contact: Hamish Cunningham (hamish@dcs.shef.ac.uk)

Semantic Annotation

- ◆ Given a domain ontology
- ◆ Identify instances in text
- ◆ Generate semantic metadata



Keys: **Person** **Location** **Organization**



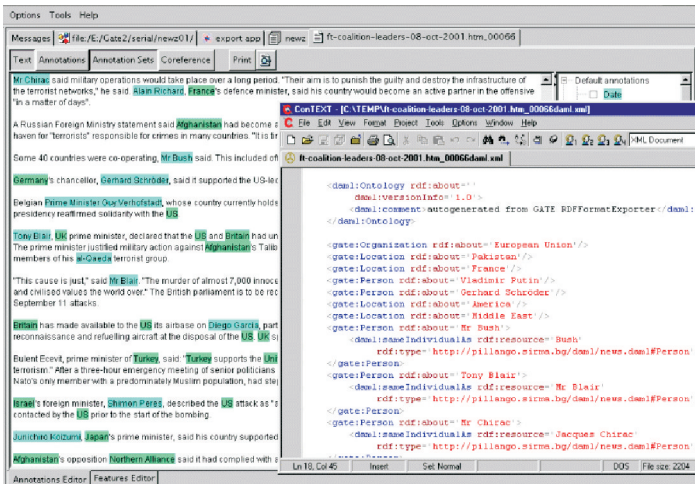
[...] entering them into a room in an area now which would be a wedding present the prince and princess [SIL] from the emir of **Sabah** [s] the two men stood in the dock of the number one court [SIL] and pleaded not guilty to all the charges against them [s] that was the cue [SIL] for the crown to begin its retreat [s] crown counsel william boyce qc explained how mr **Paul** had changed his story in relation to the dow several times [SIL] in the end claiming his butler colleague [SIL] **Bar al** had told him to sell it [SIL] and that was the crown's escape route [SIL] because mr powell [SIL] had had that conversation with **the Queen** [s] this is what william boyce told the court [SIL] around **the** real position was that his actions in relation to the dow

Headline: Another right royal embarrassment

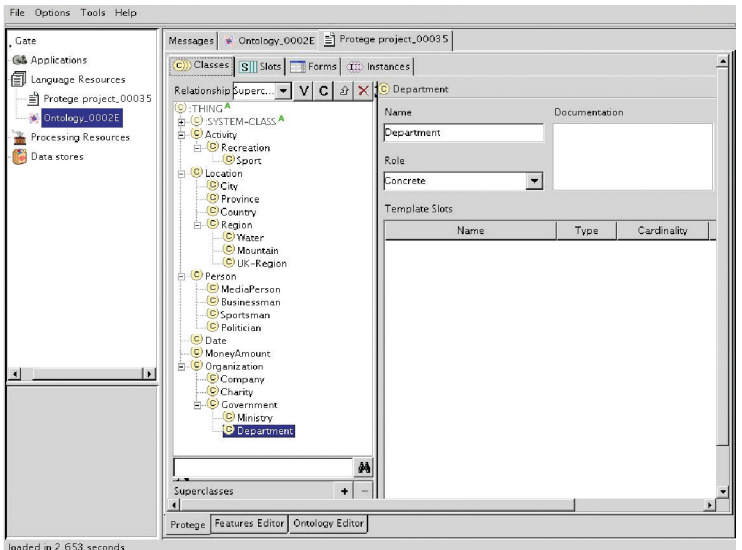
Description: The collapse of a second trial involving a royal butler raises more awkward questions.
BBC long class: (42)
Related stories:
<http://news.bbc.co.uk/1/hi/uk/2538419.stm>

Copyright © 2004 Ontotext Lab, Simsa AI, Bulgaria

Ontologies



- ◆ Ontology management through OWLIM, Jena, KAON and Protégé integration.
- ◆ RDF, OWL support.



Research

Semantic Web projects:

- ◆ SEKT - <http://www.sekt-project.com>
- ◆ TAO - <http://www.tao-project.eu>
- ◆ MUSING - <http://www.musing-project.eu>
- ◆ MediaCampaign - <http://www.media-campaign.eu>
- ◆ NeOn - <http://www.neon-project.org>
- ◆ KnowledgeWeb - <http://knowledgeweb.semanticweb.org>

Digital library projects:

- ◆ PrestoSpace - <http://www.prestospace.org>
- ◆ ETCSL - <http://www-etcs.orient.ox.ac.uk>

Commercial

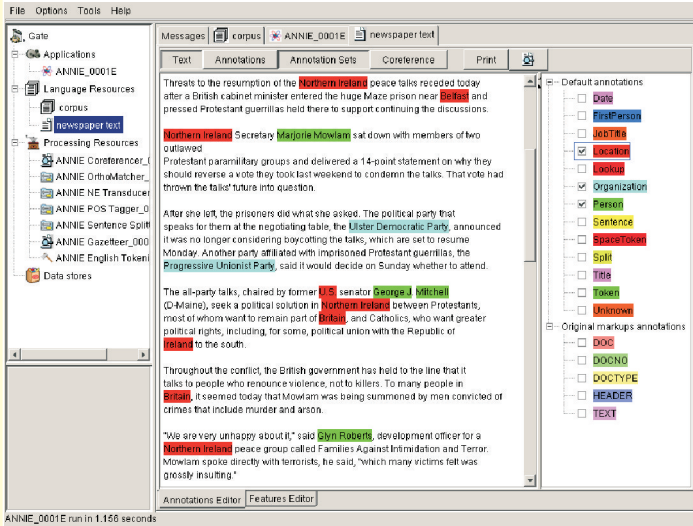
Used at:

- ◆ BT Exact
- ◆ Reuters plc.
- ◆ GlaxoSmithKline plc.
- ◆ Merck KgAa
- ◆ British Gas Plc.

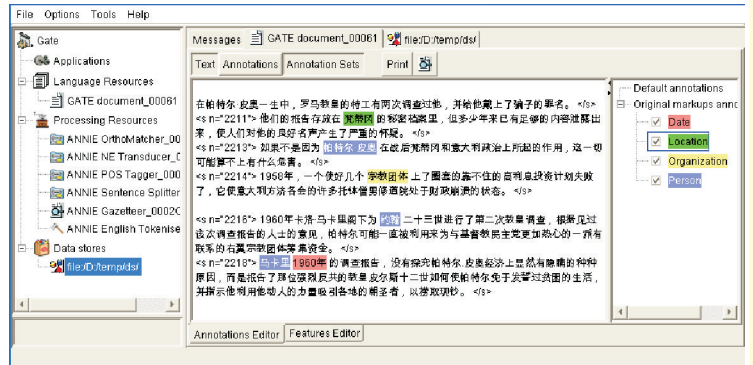
Web Mining

- ◆ Additional knowledge mined from the web and merged with semantic annotation of original content
- ◆ Uses Google and Yahoo APIs; new ones easy to add

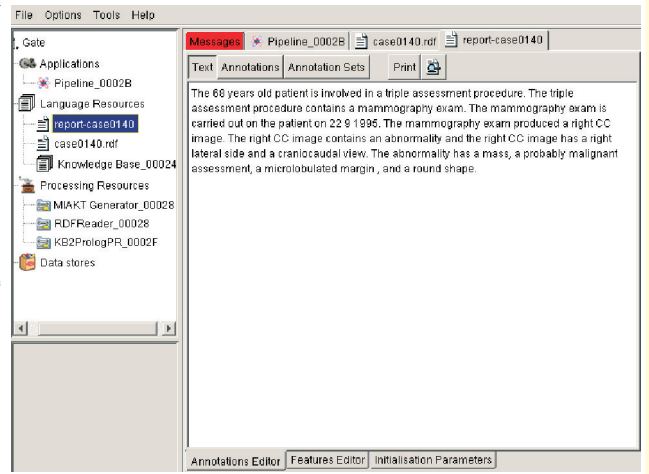
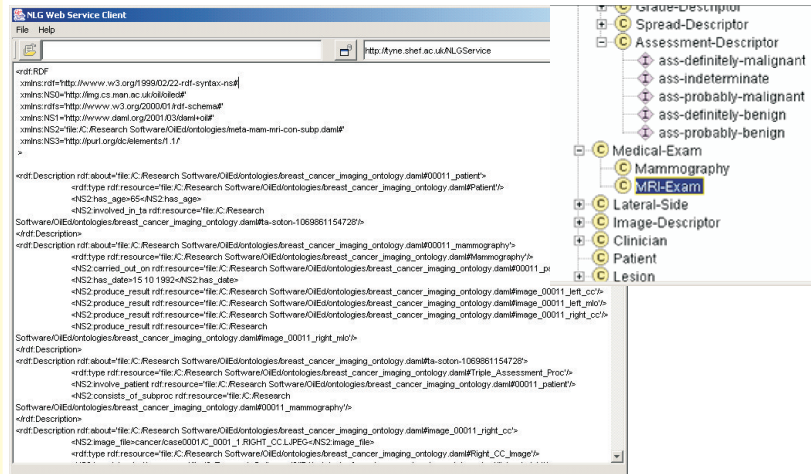
Information Extraction



- ◆ High Performance
- ◆ Robust across genres and formats
- ◆ Multilingual
- ◆ Portable



Automatic Text Generation from Ontologies

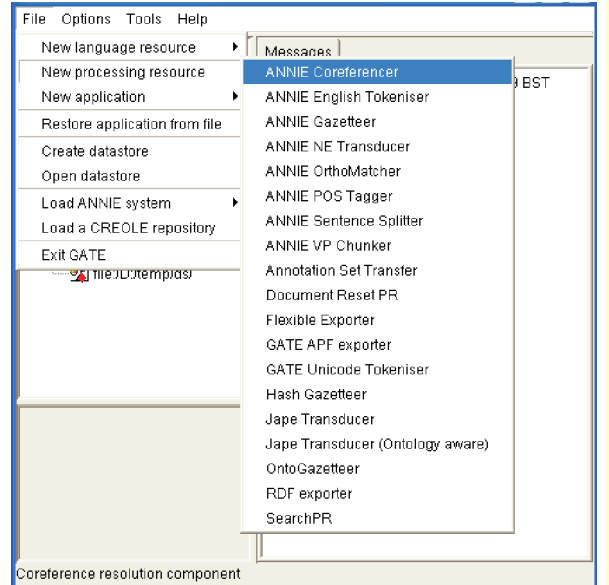


Performance Evaluation

String - Key	Start - Key	End - Key	String - Response	Start - Response	End - Response
England	2358	2365	England	2358	2365
UK	258	260	UK	258	260
Hampshire	2638	2647			
Swanwick	2886	2894			
Europe	746	752	Europe	746	752
Wales	2370	2375	Wales	2370	2375
UK	2801	2803	UK	2801	2803
Swanwick	2628	2636			
UK	931	933	UK	931	933

Precision strict: 1.0000 Recall strict: 0.6667 F-Measure strict: 0.8000
 Precision average: 1.0000 Recall average: 0.6667 F-Measure average: 0.8000
 Precision lenient: 1.0000 Recall lenient: 0.6667 F-Measure lenient: 0.8000

Components



GATE supports: Machine Learning

- ◆ Repeatability
 - ◆ Collaboration
 - ◆ Reuse vs. Reinvention
 - ◆ Quantitative evaluation
- Algorithms:**
- ◆ WEKA - decision trees, Bayesian methods
 - ◆ Maximum Entropy
 - ◆ SVM Light
- Used for:**
- ◆ Information Extraction
 - ◆ Semantic Annotation

GATE research is funded by EU & EPSRC grants