GATE

Infrastructure for Human Language Technology and the Semantic Web

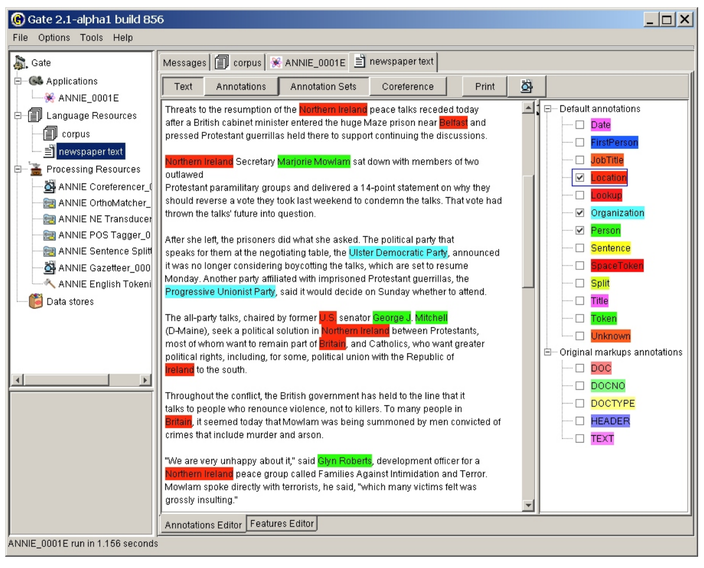
http://gate.ac.uk

Open Source
Scalable
Robust

Stable
Customisable
Reusable Components

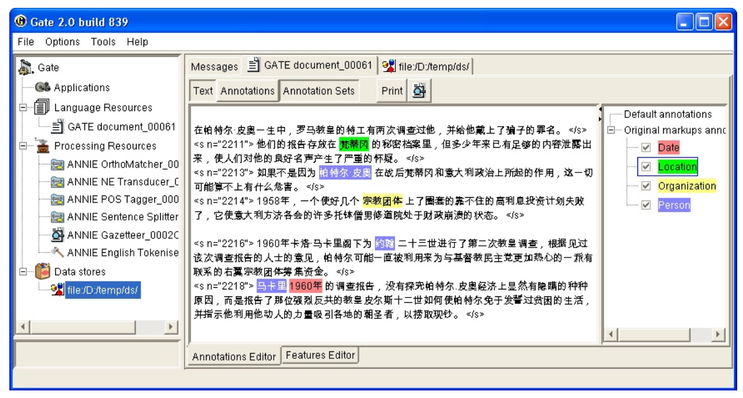
100% Java
XML, HTML, RTF, ...
Oracle, PostgreSQL backends

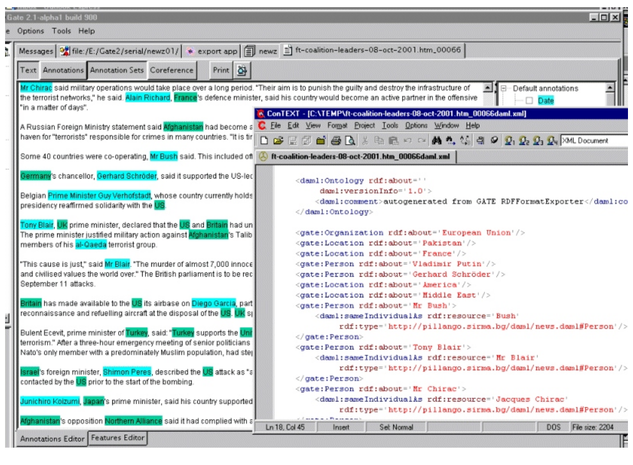
Contact: Hamish Cunningham (hamish@dcs.shef.ac.uk)



Information Extraction

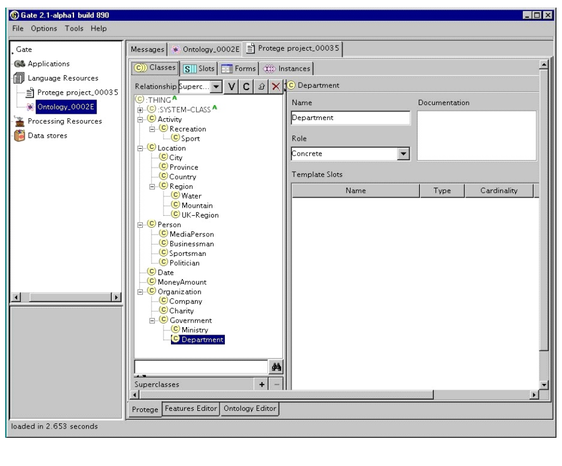
- High Performance
- Robust across genres and formats
- Multilingual
- Portable



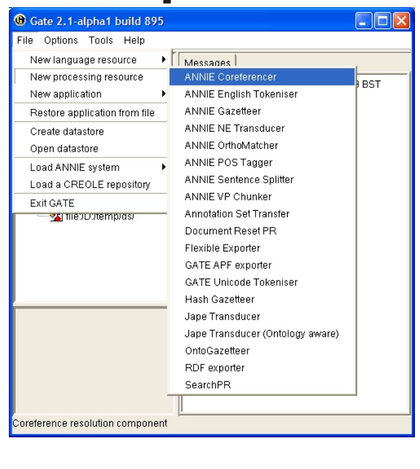


Ontologies

- Ontology management through Protégé & Sesame integration.
- RDF, OWL, DAML+OIL support.



Components



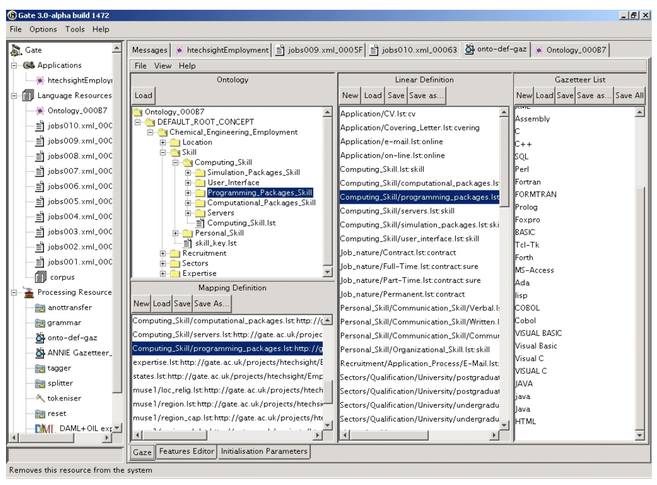
Machine Learning

Algorithms:

- WEKA - decision trees, Bayesian methods
- Maximum Entropy
- SVM Light

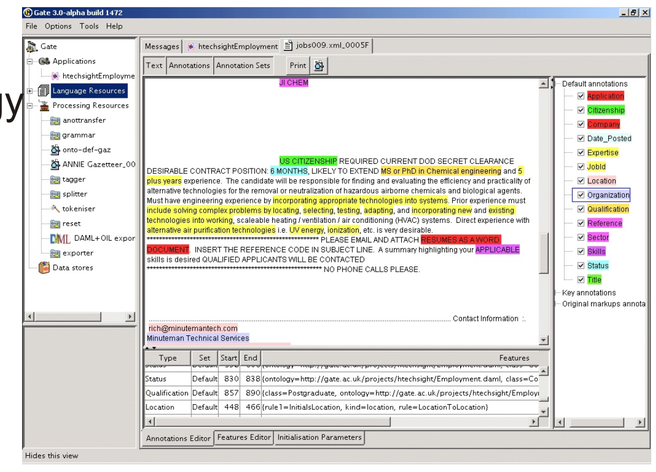
Used for:

- Information Extraction
- IE-based ontology population

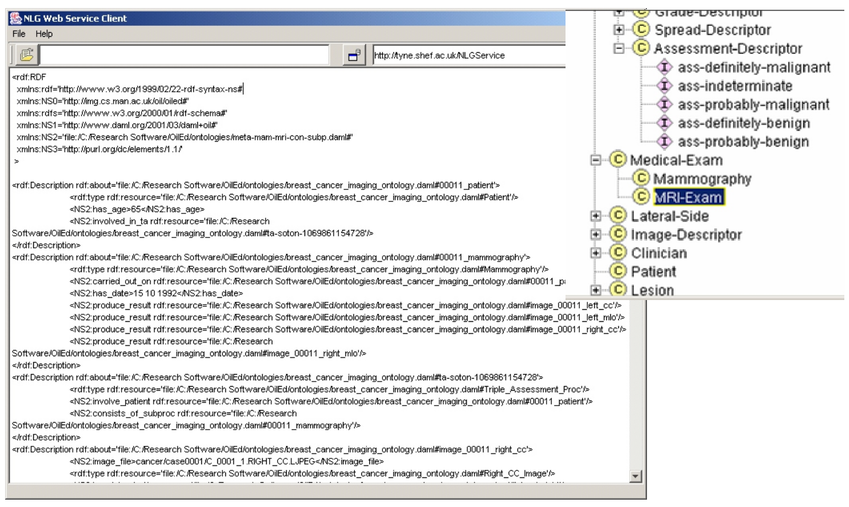


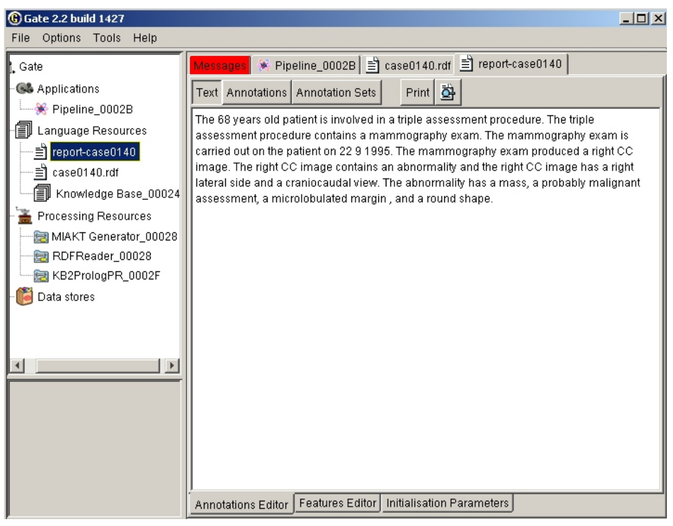
Ontology Population using IE

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

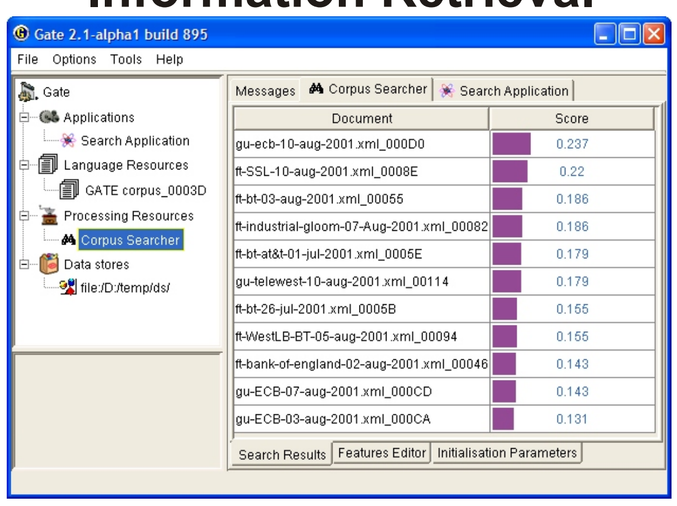


Automatic Text Generation from Ontologies





Information Retrieval



Full featured implementation based on

Research

- Repeatability
- Collaboration
- Reuse vs. Reinvention
- Quantitative evaluation

Semantic Web projects:

- SEKT - http://sekt.semanticweb.org
- **KnowledgeWeb** -knowledgeweb.semanticweb.org
- hTechSight - http://www.h-techsight.org
- AKT- http://www.aktors.org

Digital library projects:

- PrestoSpace - http://www.prestospace.org
- ETCSL - http://www-etcsl.orient.ox.ac.uk

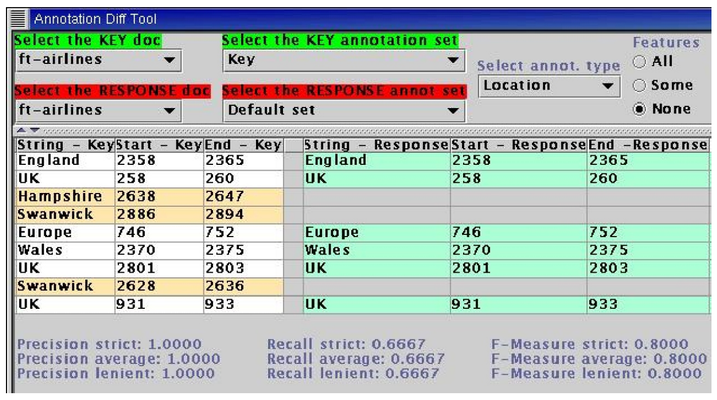
Commercial

Used at:

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

GATE research is funded by EU & EPSRC grants

Performance Evaluation



Integrated tools for performance evaluation.