

### DATA MANAGEMENT

- Pluggable input format filters
- Common data model of stand-off annotation
- Persistent storage: XML, database, etc.

### INTEGRATION

- **Information Retrieval:** Lucene, Google and Yahoo search APIs;
- **Machine Learning:** Weka and SVMLight;
- **Ontology Support:** Sesame and OWLIM;
- **Parsing:** RASP, Minipar, and SUPPLE;
- **Other:** UIMA, Wordnet, Snowball, etc.

### STANDARDS-BASED

- ISO TC37/SC4; founder member of OASIS/UIMA
- XCES, ACE, TREC formats

### EFFICIENT

- 20–40% speed and memory improvements
- Many components with linear performance

### POPULAR

- “Outstanding” and “Internationally leading”
- Used at thousands of sites;
- Over 20,000 downloads in the last year

### FREE

— Open source, LGPL, on SourceForge

### MATURE AND ACTIVELY SUPPORTED

— Over 12 yrs. old; version 4.0  
— Community support: over 20 active developers

### COMPREHENSIVE

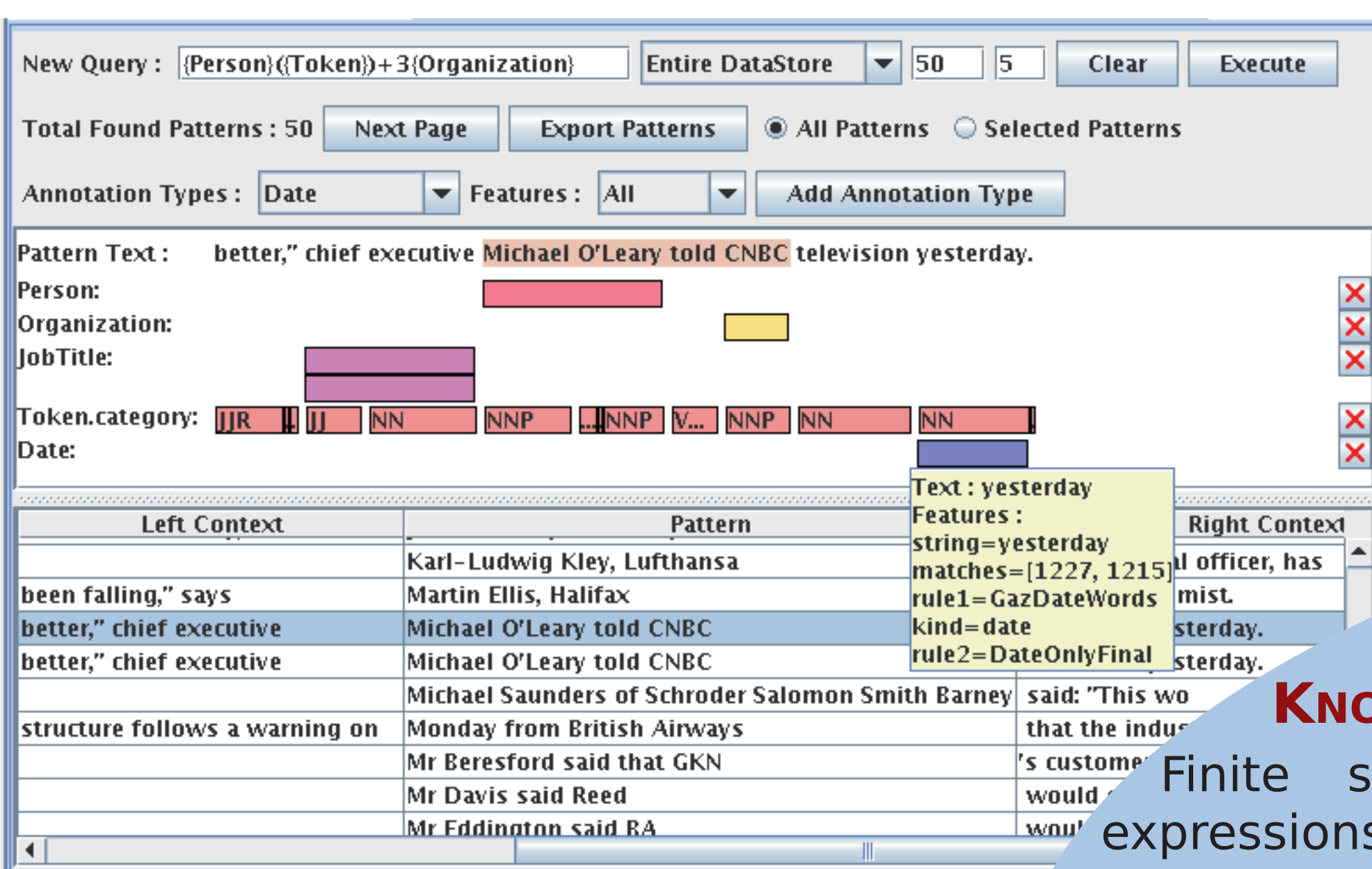
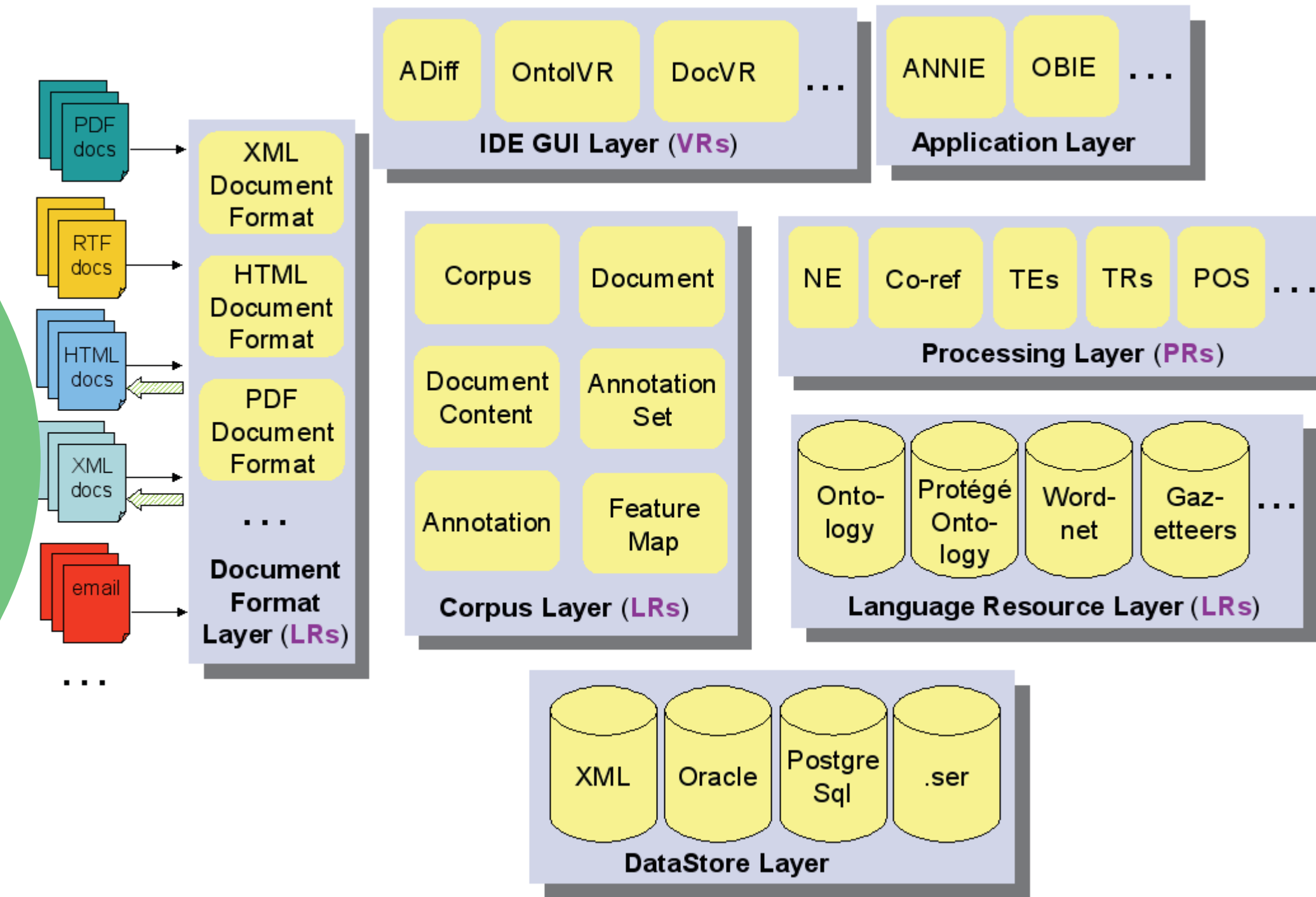
— Over **30 plugins** & 70 components

### MANY NLP ALGORITHMS

- Named Entity Recognition
- Information Extraction
- Semantic Annotation
- Text and Web Mining
- Document Summarisation
- Performance Evaluation

### 100% JAVA

- Platform-independent
- Tested on Linux, Windows, Mac OS X
- Java 5 or newer



### Information Extraction: Development Support

GATE includes ANIE, an Information Extraction system that can be used as an example or a starting point for customisations

### KNOWLEDGE ENGINEERING

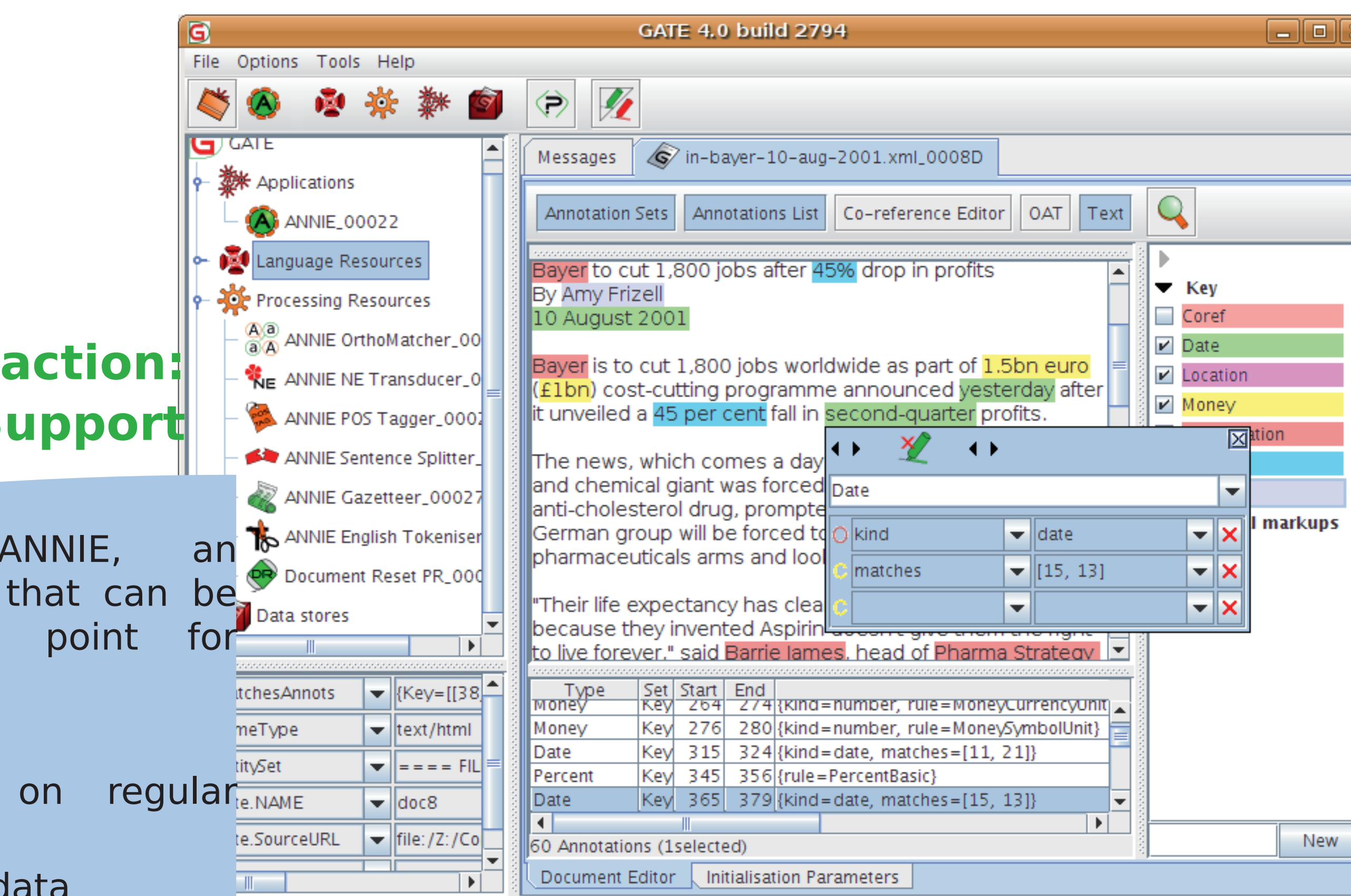
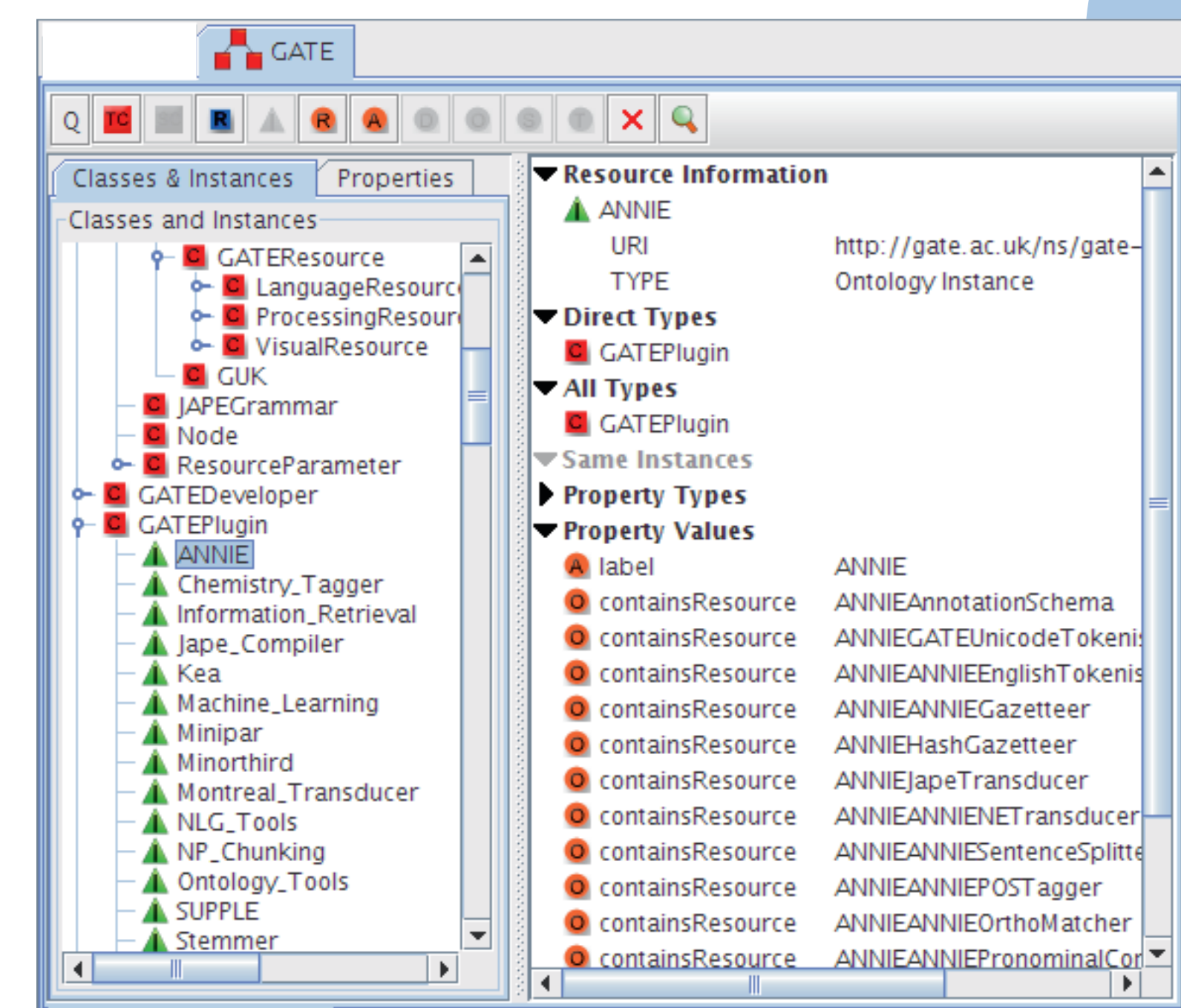
Finite state processing over annotations based on regular expressions:

- **JAPE** is a pattern-action language for matching linguistic data
- **ANNIC**: visual tool for assisted rule development

### MACHINE LEARNING

- Machine Learning framework for text mining, entity recognition, and relation extraction
- An abstraction layer for unified access to different ML algorithms

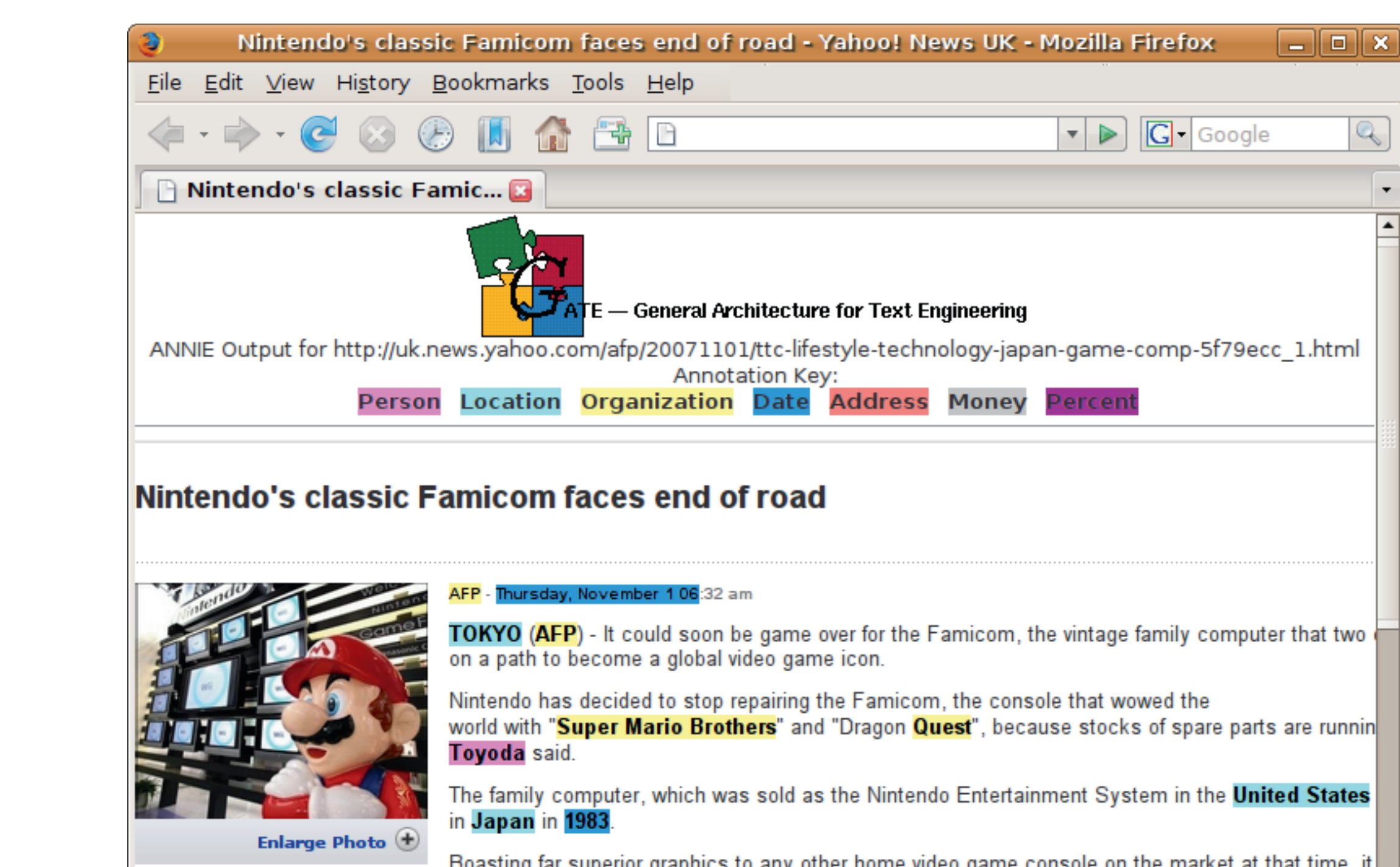
### PERFORMANCE EVALUATION TOOLS



Start	End	Key	Features	Start	End	Response	Features
1192	1204	last quarter	(kind=date)	=	1192	1204	last quarter
2431	2447	previous quarter	(kind=date)	=	2431	2447	previous quarter
2022	2028	Friday	(rule2=DateOnlyFinal, kind=date, rule1=GazDate)	=	2022	2028	Friday
1877	1888	end of June	(kind=date)	=	1884	1888	June
643	654	this autumn	(kind=date)	>	50	54	fall
1906	1922	three months ago	(kind=date)	>	-7		
225	231	Monday	(rule2=DateOnlyFinal, kind=date, rule1=GazDate)	<>	225	231	Monday

Summary statistics:

- Correct: 3
- Partially Correct: 1
- Missing: 3
- False Positives: 2
- Average: 0.50, 0.5833, 0.5385



### Applications

### SEMANTIC ANNOTATION

- (Semi-)automatic production of semantic meta-data for text and multimedia
- Identify mentions of known concepts and instances from an ontology
- Enables a search by meaning paradigm to enhance traditional retrieval methods

### RICHNEWS

- Applying semantic annotation to multimedia broadcasts
- Combines timing information from video and semantics from web pages



### SELECTED COMMERCIAL USERS

- Glaxo Smith Kline PLC
- AT&T
- Master Foods NV
- British Gas PLC
- Syntalex Ltd.
- Lernout Hauspie GmbH.
- Thompson Corp
- Innovantage
- Garlik
- Fizzback
- Spock

### RESEARCH PROJECTS

- **E-Science:** AKT, MultiFlora, MiAKT, CLEF
- **Digital Libraries:** Prestospace, GATE/ETCSL, EMILLE, OldBaileyE

- **Semantic Web and Knowledge Management:** SEKT, TAO, NeON, Musing, MediaCampaign
- **Human Language Technology:** SAFE, LIRICS, MUSE