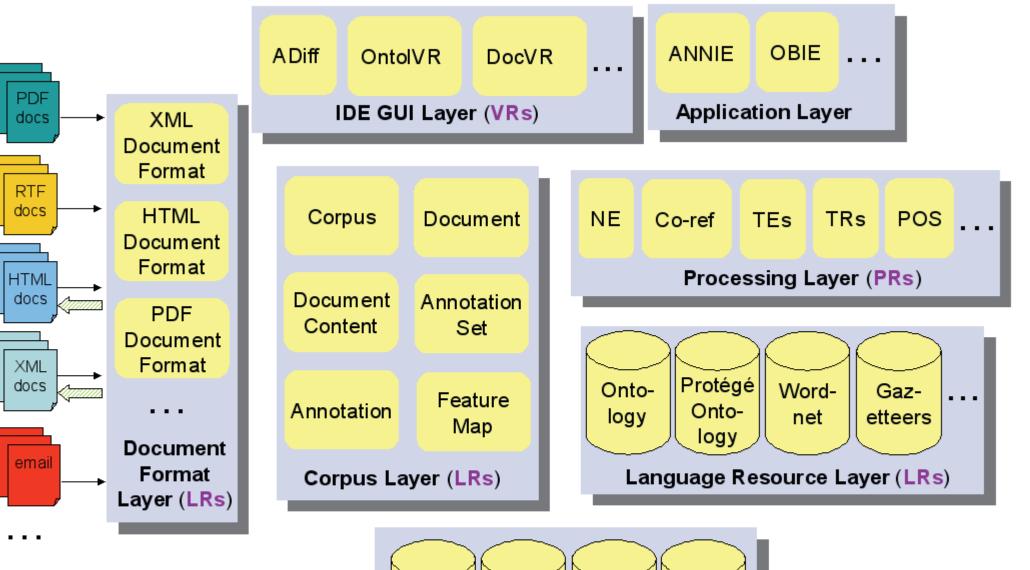






CONTACT: Prof: Hamish Cunningham hamish@dcs.shef.ac.uk



Infrastructure for Text Mining and Natural Language Processing

http://gate.ac.uk

DATA MANAGEMENT

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

NTEGRATION

- 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

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**

EFFICIENT

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

POPULAR

- "Outstanding" and "Internationally leading"
- Used at thousands of sites;

New Query : {Person}((Token))+3(Organization)

• Over 20,000 downloads in the last year

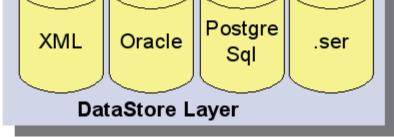
Performance Evaluation

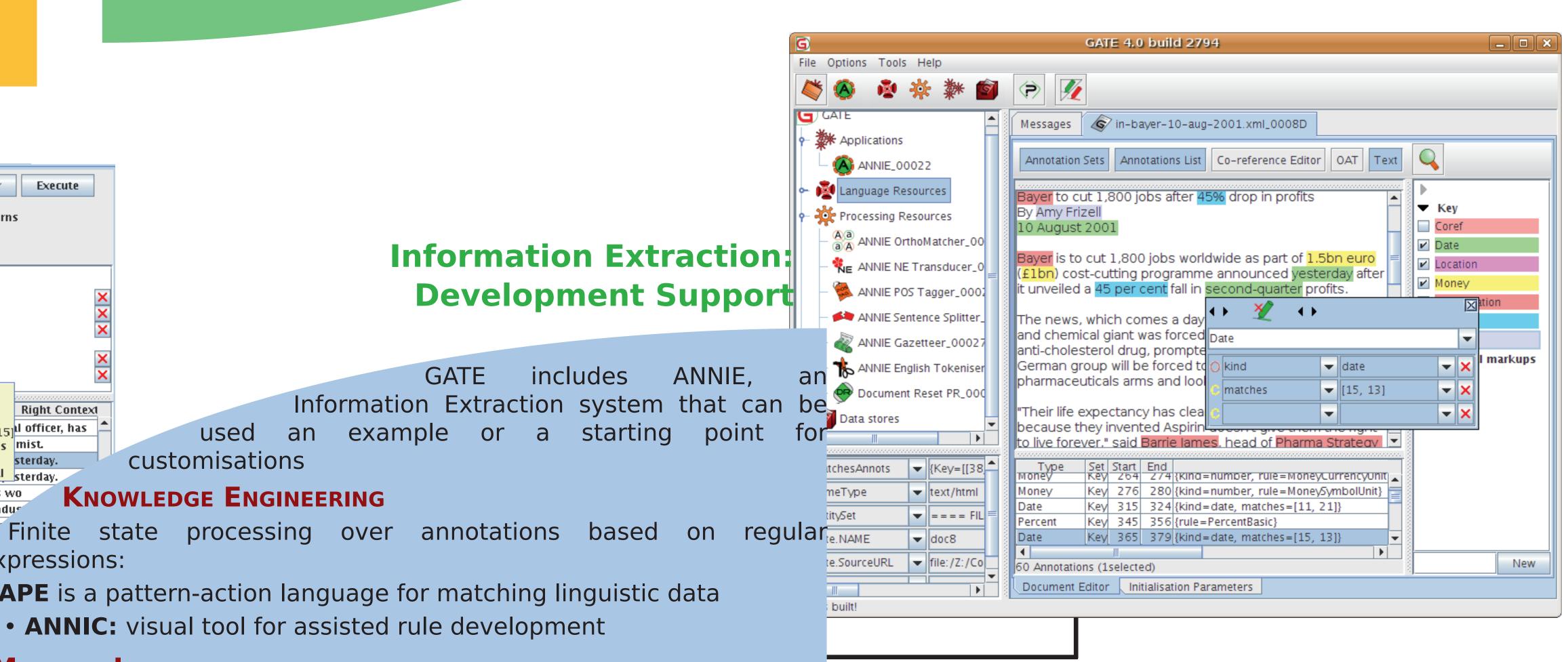
100% JAVA

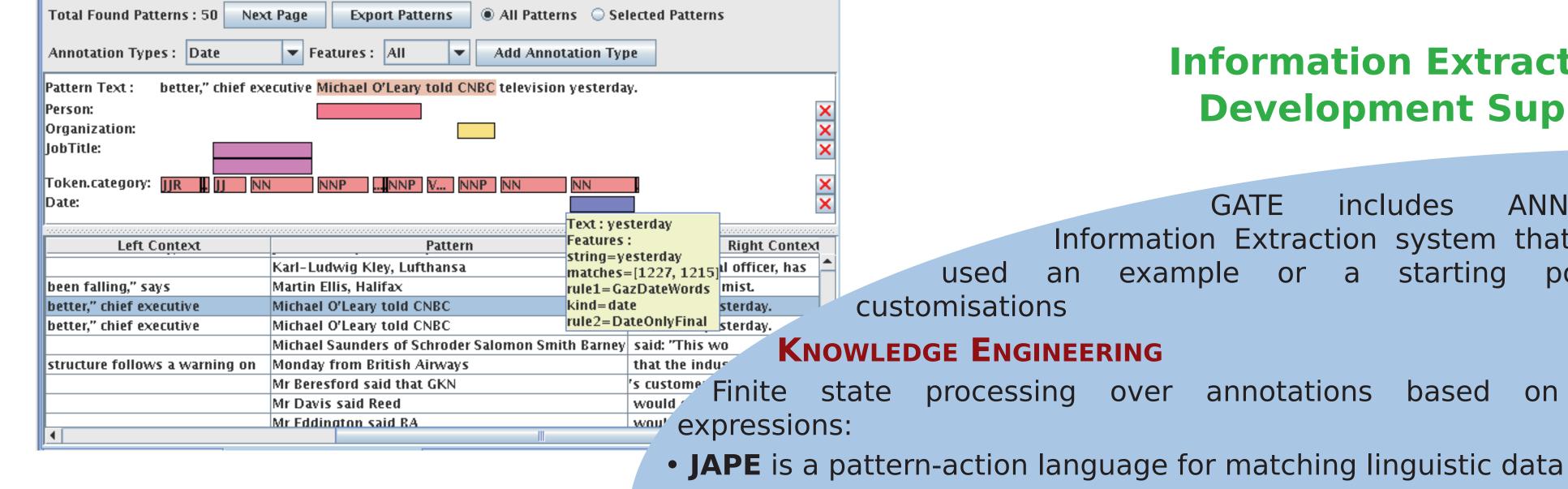
- **Platform-indepedent**
- Tested on Linux, Windows, Mac OS X
- Java 5 or newer

used

an







Entire DataStore 🔽 50 5

Clear

Execute

GATE		
Q 💶 🔳 🖪 🔺 🕄 🔍 🔍		
Classes & Instances Properties	Resource Information	
Classes and Instances		
- GATEResource	URI http://gate.ac.uk/ns/gate-	
🔶 🧧 LanguageResourc	TYPE Ontology Instance	
- ProcessingResour	▼ Direct Types	
• Curre	GATEPlugin	
- 🖸 GUK - 🖸 JAPEGrammar 📃	▼ All Types	
- G Node	GATEPlugin	
- C ResourceParameter	Same Instances	
- GATEDeveloper	Property Types	=
	Property Values	
- A ANNIE - A Chemistry_Tagger	ANNIE	
- A Information_Retrieval	containsResource ANNIEAnnotationSchema	
- 🛦 Jape_Compiler	containsResource ANNIEGATEUnicodeTokeni:	
– 🛦 Kea	containsResource ANNIEANNIEEnglishTokenis	
- A Machine_Learning	containsResource ANNIEANNIEGazetteer	
 Minipar Minorthird 	containsResource ANNIEHashGazetteer	
- Montreal_Transducer	containsResource ANNIEJapeTransducer	
- A NLG_Tools	containsResource ANNIEANNIENETransducer-	
- 🛕 NP_Chunking	containsResource ANNIEANNIESentenceSplitte	
- A Ontology_Tools	containsResource ANNIEANNIEPOSTagger	
- A SUPPLE - A Stemmer	containsResource ANNIEANNIEOrthoMatcher	
	containsResource ANNIEANNIEPronominalCon	•

Nintendo's classic Famicom faces end of road - Yahoo! News UK - Mozilla Firefox 📃 🗖 🗙 <u>File Edit View History Bookmarks Tools Help</u> 🔻 🕨 🚺 Google 🖹 Nintendo's classic Famic... 🖸 General Architecture for Text Engineering ANNIE Output for http://uk.news.yahoo.com/afp/20071101/ttc-lifestyle-technology-japan-game-comp-5f79ecc_1.html Annotation Key: Person Location Organization Date Address Money Percent Nintendo's classic Famicom faces end of road

MACHINE LEARNING

 Machine Learning framework for text mining, entity recognition, and relation extraction

GATE

example or a

• An abstraction layer for unified access to different ML algorithms

PERFORMANCE EVALUATION TOOLS

Annotation Diff Tool Document Annotation Set Key: Document_00016 Key Document_00016 Key Peatures: Annotation Type: Date F-Measure Weight Document_00016 Key Features: Annotation Type: Date F-Measure Weight Document_00016 Key Features: Annotation Type: Start End Key Features: </th <th>ginci</th> <th></th> <th>17</th> <th></th>	ginci		17													
Key: Document_00016 Key Annotation Type: Date F-Measure Weight Response: Document_00016 [Default set] Features: All Some None 1.00 Start End Key Features: Start End Response Features 1192 1204 last quarter (kind=date) = 1192 1204 last quarter (rule2=DateOnlyFinal, kind=date, rule2 2022 028 Friday (rule2=DateOnlyFinal, kind=date, rule1=GazDate) = 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rule1=GazDate) = 225 231 Monday <	2								Annota	tion Dif	if Tool					_ _ X
Document_00016 [Default set] Features: All © Some © None 1.00 Start End Key Features Start End Response Features 1192 1204 last quarter {kind=date} = 1192 1204 last quarter (rule2=DateOnlyFinal, kind=date, rul 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rule1=GazDate) = 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rul 1877 1888 end of June (kind=date) - 1884 1888 June (rule2=DateOnlyFinal, kind=date, rul 643 654 this autumn (kind=date) -? - - - 1906 1922 three months ago (kind=date) -? - - - 225 231 Monday (rule2=DateOnlyFinal, kind=date, rule1=GazDate) <> 225 231 Monday (kind=time) - Correct: 3 Recall Precision F-Measure Export to HTML Partially Correct: 1 Strict: 0.4286 0.50 0.4615 Missing: 3 Lenient: 0.5714 0.6667 0.6154			Docum	nent		Annota	tion Set									
Response: Document_00016 IDefault set] Features: All Some None 1.00 Start End Key Features Start End Response Features 1192 1204 last quarter (kind=date) = 1192 1204 last quarter (rule2=DateOnlyFinal, kind=date, rul 2431 2447 previous quarter (kind=date) = 2431 2447 previous quarter (rule2=DateOnlyFinal, kind=date, rul 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rul = 2431 2447 previous quarter (rule2=DateOnlyFinal, kind=date, rul 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rul = 2431 2447 previous quarter (rule2=DateOnlyFinal, kind=date, rul 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rul = 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rul 2025 2031 Monday (rule2=DateOnlyFinal, kind=date, rul=GazDate) -? -? -? -? -? -? -? -? -? -?	Key:		Docun	nent_00(016 🔻	Кеу		▼ A	nnotatio	on Type:	Date			💌 F-Mea	_	Diff
1192 1204 last quarter (kind=date) = 1192 1204 last quarter (rule2=DateOnlyFinal, kind=date, rule] 2431 2447 previous quarter (kind=date) = 2431 2447 previous quarter (rule2=DateOnlyFinal, kind=date, rule] 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rule]=GazDate) = 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rule] 1877 1888 end of June (kind=date) ~ 1884 1888 June (rule2=DateOnlyFinal, kind=date, rule] 643 654 this autumn (kind=date) ~ 1884 1888 June (kind=date) 643 654 this autumn (kind=date) -? 2 2 2 231 Monday (kind=date) -? 2 2 2 231 Monday (kind=tate) -? 2 2 2 231 Monday (kind=tate) -? 2 2 2 231 Monday (kind=tate) -? 2 2 2 2 2 2 2 <td colspan="3">Response: Document_00016</td> <td>016 💌</td> <td colspan="5">▼ [Default set] ▼ Features: ○</td> <td>⊖ All</td> <td colspan="5"></td>	Response: Document_00016			016 💌	▼ [Default set] ▼ Features: ○					⊖ All						
2431 2447 previous quarter (kind=date) = 2431 2447 previous quarter (rule2=DateOnlyFinal, kind=date, rule 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rule1=GazDate) = 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rule1=GazDate) 1877 1888 end of June (kind=date) ~ 1884 1888 June (rule2=DateOnlyFinal, kind=date, rule1=GazDate) 1877 1888 end of June (kind=date) ~ 1884 1888 June (rule2=DateOnlyFinal, kind=date, rule1=GazDate) 643 654 this autumn (kind=date) -? 50 54 fall (kind=date) 643 654 this autumn (kind=date) -?	Start I	End		Key			Fe	atures				Start	End	Response		Features
2022 2028 Friday {rule2=DateOnlyFinal, kind=date, rule1=GazDate} = 2022 2028 Friday (rule2=DateOnlyFinal, kind=date, rule1=GazDate) = 1877 1888 end of June {kind=date} ~ 1884 1888 June (rule2=DateOnlyFinal, kind=date, rule1=GazDate) = 1877 1888 end of June {kind=date} ~ 1884 1888 June (rule2=DateOnlyFinal, kind=date, rule1=GazDate) = 643 654 this autumn { -? -? -	11921	204	last qu	Iarter	{kin	d=date}					=	119 2	1204	last quarter	{rule2=DateOnlyFir	ial, kind=date, rul
1877 1888 end of June (kind=date) ~ 1884 1888 June (rule2=DateOnlyFinal, kind=date, rule) 643 654 this autumn (kind=date) -? 643 644 (kind=date) 1906 1922 three months ago (kind=date) -? 1 1 (kind=date) 225 231 Monday (rule2=DateOnlyFinal, kind=date, rule1=GazDate) -? 1 1 1 1906 1922 three months ago (kind=date) -? 1 <t< td=""><td>2431 2</td><td>447</td><td>previo</td><td>us quart</td><td>er {kin</td><td>d=date}</td><td></td><td></td><td></td><td></td><td>=</td><td>2431</td><td>2447</td><td>previous quarter</td><td>{rule2=DateOnlyFir</td><td>nal, kind=date, rul</td></t<>	2431 2	447	previo	us quart	er {kin	d=date}					=	2431	2447	previous quarter	{rule2=DateOnlyFir	nal, kind=date, rul
Image: Contract of the second seco	2022 2	028	Friday	y {rule2=DateOnlyFinal, kind=date, rule1=Gaz!						e1=GazD) ate} =	2022	2028	Friday	{rule2=DateOnlyFir	nal, kind=date, rul
643 654 this autumn 0 -? 0	1877 1	888	end of June {kind=date}						~	1884	1888	June	{rule2=DateOnlyFir	nal, kind=date, rul		
1906 1922 three months ago {kind=date} -? 23 23 Monday {rule2=DateOnlyFinal, kind=date, rule1=GazDate} <> 225 231 Monday {kind=time} 1											?-	50	54	fall	{kind=date}	
225 231 Monday {rule2=DateOnlyFinal, kind=date, rule1=GazDate}< <> 225 231 Monday {kind=time} 4 Image: Correct: 3 Recall Precision F-Measure Export to HTML Partially Correct: 1 Strict: 0.4286 0.50 0.4615 Missing: 3 Lenient: 0.5714 0.6667 0.6154	643 6	54	this au	itumn	8						-?					
Image: Correct: 3 Recall Precision F-Measure Export to HTML Partially Correct: 1 Strict: 0.4286 0.50 0.4615 Missing: 3 Lenient: 0.5714 0.6667 0.6154	1906 1922 three months ago (kind=date)								-?							
Correct:3RecallPrecisionF-MeasureExport to HTMLPartially Correct:1Strict:0.42860.500.4615Missing:3Lenient:0.57140.66670.6154	225 2	31	Monda	ıy	{rul	e2=Date	OnlyFinal, I	kind=a	date, rule	e1=GazD) ate} <>	225	231	Monday	{kind=time}	
Correct:3RecallPrecisionF-MeasureExport to HTMLPartially Correct:1Strict:0.42860.500.4615Missing:3Lenient:0.57140.66670.6154																
Partially Correct: 1 Strict: 0.4286 0.50 0.4615 Missing: 3 Lenient: 0.5714 0.6667 0.6154																
Missing: 3 Lenient: 0.5714 0.6667 0.6154	Correct: 3 Recall Precision F-Measure Export to HTML															
	Partial	ly Co	orrect:	1 S	trict:	0.4286	0.50	0.461	5							
False Positives: 2 Average: 0.5833 0.5385	Missin	g:		3 L	enient:	0.5714	0.6667	0.61 54	4							
	False F	Posit	ives:	2 A	verage:	0.50	0.5833	0.538	5]

SEMANTIC ANNOTATION										
• (Semi-)automa	produ	0								
semantic meta-	data	for	text	and						
multimedia										

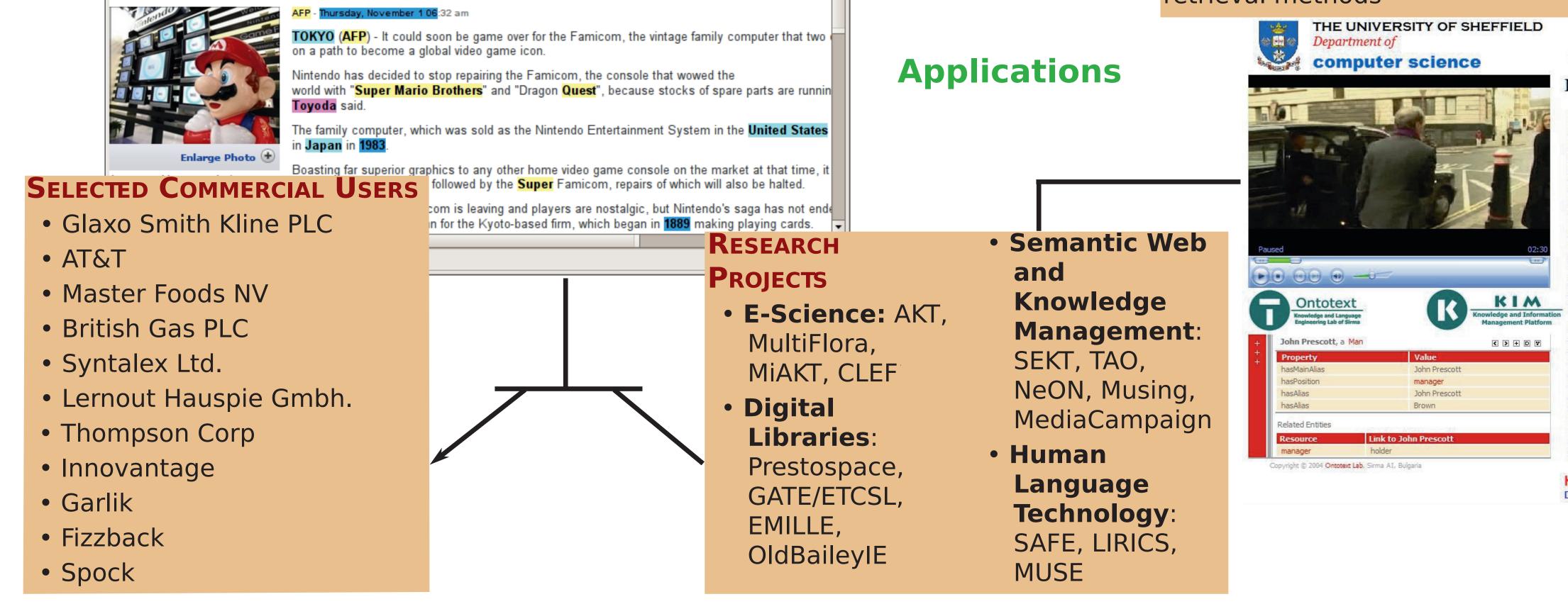
• Identify mentions of known concepts video and semantics from web pages and instances from an ontology

Enables a search by meaning paradigm to enhance traditional retrieval methods

RICHNEWS

 Applying semantic annotation to multimedia broadcasts

Combines timing information from



dinbailante RichNews Keys: Person Location Organization

date and that his seriously ill in hospital on the people of the buses travelling on this why stop [SIL] now is to be met [s] the lib dems' continues to eat his hat [SIL] into the break [SIL] and [SIL] be mature [SIL] enough to use [SIL] them [s] no welcome to the bbc's news at one o'clock [s] the trial of another war but has collapsed at the opening [SIL] hour brown face charges of stealing valuables including an ornate what we're able in sailing ship [SIL] worth more than half a million pounds from diana princess of wales [s] the case against him the society jeweller edge and having collapsed before a jury was sworn in [s] the prosecution said there was no realistic prospect of conviction [s] so if the collapse of the trial of paul but last month [s] now this to be an internal review of the metropolitan police's investigations [SIL] are royal correspondent nicholas witchell [SIL] is at the old bailey [SIL] make [SIL] and this was a decision taken [SIL] personally by v. director of public prosecutions david calvert smith [s] it will i think kirby the cause of a huge relief within [SIL] the royal palaces have been [SIL] those of the royal palaces to be in desperate really is a kick [SIL] this whole further into touch and the crown prosecution service [SIL] has now obliged [s] the metropolitan police though are not happy they believe that in this [SIL] instance then it carried out [SIL] a four and thorough investigation [SIL] and most unusually in explaining the decision to drop the case [SIL] the prosecution went through

many of the facets of what would of been the prosecution case [s] now they say haroldbrown [SIL] kept changing his story how initially said [SIL] that these items had been disposed off [SIL] before diana princess of wales's death of the money had been given to her [s] in fact the evidence was clearly they were disposed off [s] after [SIL] her death [SIL] and then [SIL] haroldbroken apparently saying that he be passed to dispose of these items by [SIL] another [SIL] man [SIL] paul butler and it was that [SIL] of course that gave [SIL] the crown their way out of this [SIL] the second instalment of the saga [SIL] of the crown [SIL] and the butler's [SIL] haroldbrown [SIL] and other buffalo who work for the prince and princess of wales [SIL] a mother of a clue who's experienced the curious way in which some royal households [SIL] have dealt with gifts and their disposal [s] his co accused and have any [SIL] society you learned was charged with receiving items from haroldbrown [SIL] among them this model of an arab dow which would be a wedding present the prince and princess [SIL] from the emir of bahrain [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 brown had changed his story in relation to the dow several times [SIL] in the end claiming his butler colleague [SIL] paul bar all 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] narrowed brown's final position was that his actions in relation to the dow were effectively [SIL] authorized by poor bowel [s] if however paul barrow was willing to inform the queen that he was holding on to property belonged to diana princess of wales [SIL] without discernible disapproval from her [s] it enhances the credibility of hell brown's account [s] but he reasonably accepted that poor barrel [SIL] was acting in good faith [s] that was the prosecution's way [SIL] of throwing in the town [s] both men were acquitted [s] mr brown's lawyers said his client was delighted at [SIL] our ground [SIL]many of the

Headline: Another right royal embarrassment Description: The collapse of a second trial involving a royal butler raises more awkward questions.

> GATE has been funded by EU, EPSRC, AHRC, and BBSRC grants