Human Language Technology for Automatic Annotation and Indexing of Digital Library Content

Kalina Bontcheva, Hamish Cunningham
Dept. of Computer Science
University of Sheffield
Regent Court, 211 Portobello St
Sheffield, S1 4DP, UK
kalina,hamish@dcs.shef.ac.uk

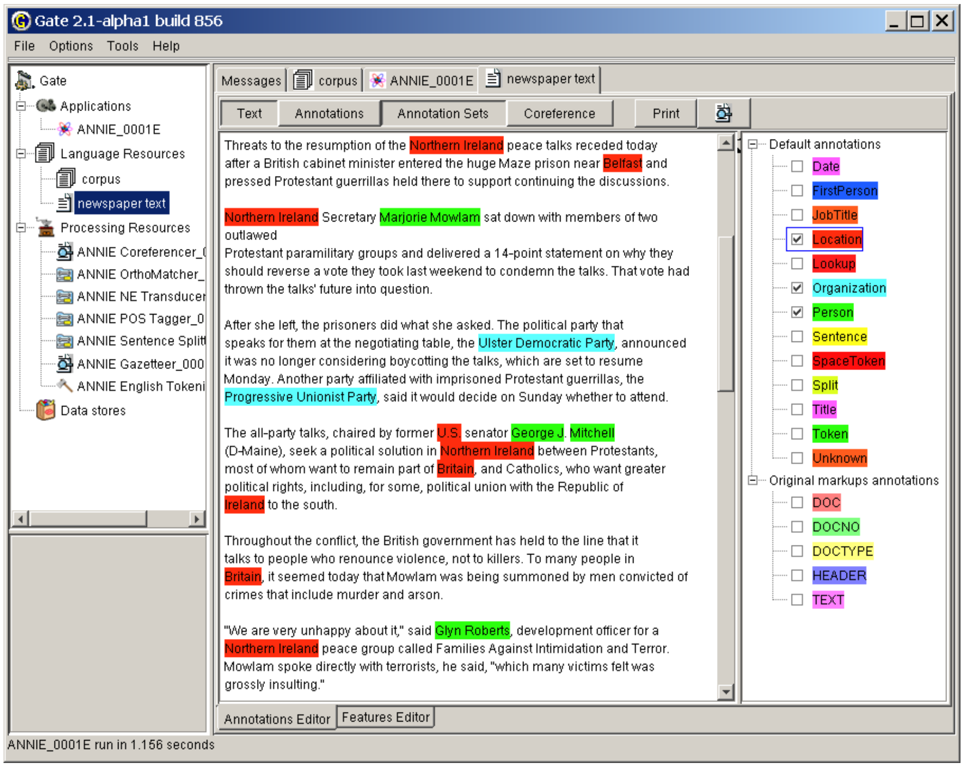
May 3, 2002

This demo is intended to present the domain-independent and customisable human language technology and the way it was applied for annotating 18*th* century OldBailey proceedings and indexing multimedia content. This demo is intended to accompany the paper with the same title, submitted to ECDL.

Each of these applications posed a unique challenge: the court trials required adapting the language processing components to the non-standard written conventions of old English, while the football collection presented the challenge of indexing material in multiple modalities - video, audio, semi-structured documents, and free text (newspaper reports). In both cases, we used as a basis our robust and customisable named entity recogniser, which comes as part of ANNIE- A Nearly New Information Extraction system. In OldBaileyIE we also used the graphical environment of GATE, which allows manual annotation verification and correction to be carried out on the processed texts.

First we will demonstrate the set of language technology tools which constitute ANNIE and the visual environment which is used for creating and customising new applications (see Figure 1).

Next we will demonstrate the automatic annotation of the OldBailey texts (Figure 2) and present the visual environment used for annotation error correction (Figure 3). We will also demonstrate how this environment can be used for manual annotation, independent of the language technology tools.



Threats to the resumption of the Northern Ireland peace talks receded today after a British cabinet minister entered the huge Maze prison near Belfast and pressed Protestant guerrillas held there to support continuing the discussions.

Northern Ireland Secretary Marjorie Mowlam sat down with members of two outlawed
Protestant paramilitary groups and delivered a 14-point statement on why they should reverse a vote they took last weekend to condemn the talks. That vote had thrown the talks' future into question.

After she left, the prisoners did what she asked. The political party that speaks for them at the negotiating table, the Ulster Democratic Party, announced it was no longer considering boycotting the talks, which are set to resume Monday. Another party affiliated with imprisoned Protestant guerrillas, the Progressive Unionist Party, said it would decide on Sunday whether to attend.

The all-party talks, chaired by former U.S. senator George J. Mitchell (D-Maine), seek a political solution in Northern Ireland between Protestants, most of whom want to remain part of Britain, and Catholics, who want greater political rights, including, for some, political union with the Republic of Ireland to the south.

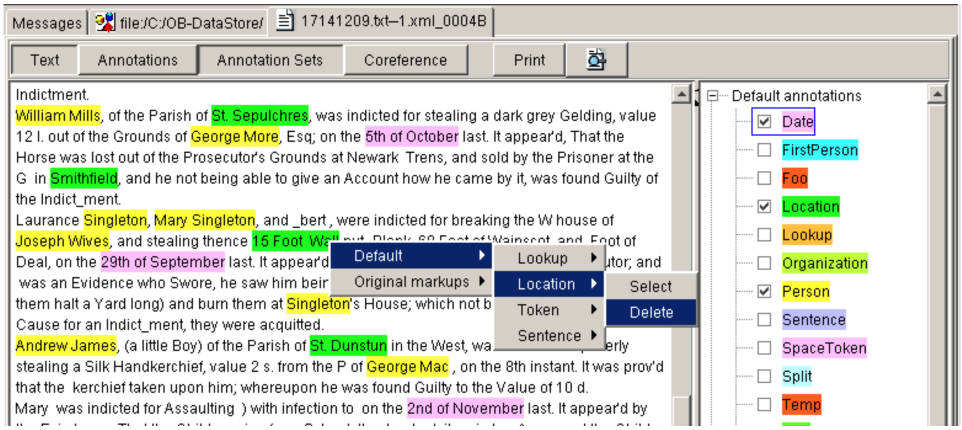
Throughout the conflict, the British government has held to the line that it talks to people who renounce violence, not to killers. To many people in Britain, it seemed today that Mowlam was being summoned by men convicted of crimes that include murder and arson.

"We are very unhappy about it," said Glyn Roberts, development officer for a Northern Ireland peace group called Families Against Intimidation and Terror. Mowlam spoke directly with terrorists, he said, "which many victims felt was grossly insulting."

ANNIE_0001E run in 1.156 seconds

Figure 1: Named Entities recognised by ANNIE

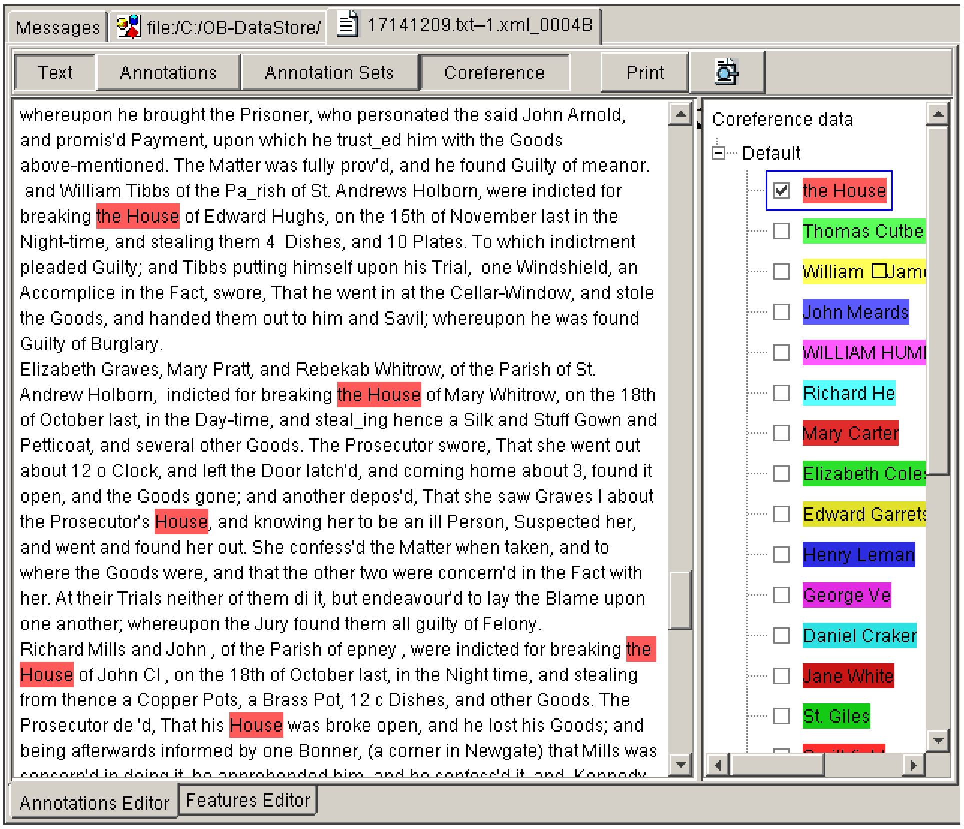
Finally we will demonstrate how multimedia content can be indexed and searched using human language technology. The demonstrator is in the football domain, where video and textual sources about matches are indexed and can be accessed via a user-friendly interface (see Figure 4).



Indictment.
William Mills, of the Parish of St. Sepulchres, was indicted for stealing a dark grey Gelding, value
12 l. out of the Grounds of George More, Esq; on the 5th of October last. It appear'd, That the
Horse was lost out of the Prosecutor's Grounds at Newark Trens, and sold by the Prisoner at the
G in Smithfield, and he not being able to give an Account how he came by it, was found Guilty of
the Indict_ment.
Laurance Singleton, Mary Singleton, and _bert , were indicted for breaking the W house of
Joseph Wives, and stealing thence 15 Foot Wa
Deal, on the 29th of September last. It appear'd
was an Evidence who Swore, he saw him beir
them halt a Yard long) and burn them at Singleton's House; which not b
Cause for an Indict_ment, they were acquitted.
Andrew James, (a little Boy) of the Parish of St. Dunstun in the West, wa
stealing a Silk Handkerchief, value 2 s. from the P of George Mac , on the 8th instant. It was prov'd
that the kerchief taken upon him; whereupon he was found Guilty to the Value of 10 d.
Mary was indicted for Assaulting) with infection to on the 2nd of November last. It appear'd by

Default annotations
☑ Date
☐ FirstPerson
☐ Foo
☑ Location
☐ Lookup
☐ Organization
☑ Person
☐ Sentence
☐ SpaceToken
☐ Split
☐ Temp

Figure 2: An Old Bailey example text from 1714



whereupon he brought the Prisoner, who personated the said John Arnold, and promis'd Payment, upon which he trust_ed him with the Goods above-mentioned. The Matter was fully prov'd, and he found Guilty of meanor. and William Tibbs of the Pa_rish of St. Andrews Holborn, were indicted for breaking the House of Edward Hughs, on the 15th of November last in the Night-time, and stealing them 4 Dishes, and 10 Plates. To which indictment pleaded Guilty; and Tibbs putting himself upon his Trial, one Windshield, an Accomplice in the Fact, swore, That he went in at the Cellar-Window, and stole the Goods, and handed them out to him and Savil; whereupon he was found Guilty of Burglary.

Elizabeth Graves, Mary Pratt, and Rebekab Whitrow, of the Parish of St. Andrew Holborn, indicted for breaking the House of Mary Whitrow, on the 18th of October last, in the Day-time, and steal_ing hence a Silk and Stuff Gown and Petticoat, and several other Goods. The Prosecutor swore, That she went out about 12 o Clock, and left the Door latch'd, and coming home about 3, found it open, and the Goods gone; and another depos'd, That she saw Graves I about the Prosecutor's House, and knowing her to be an ill Person, Suspected her, and went and found her out. She confess'd the Matter when taken, and to where the Goods were, and that the other two were concern'd in the Fact with her. At their Trials neither of them di it, but endeavour'd to lay the Blame upon one another; whereupon the Jury found them all guilty of Felony.

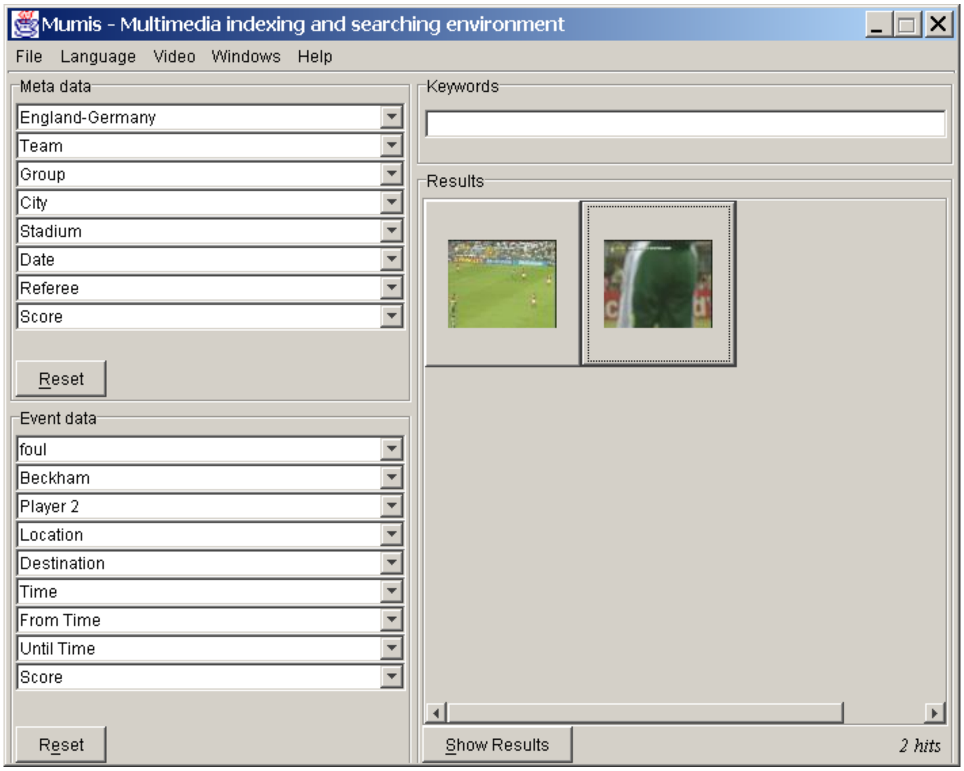
Richard Mills and John , of the Parish of epney , were indicted for breaking the House of John Cl , on the 18th of October last, in the Night time, and stealing from thence a Copper Pots, a Brass Pot, 12 c Dishes, and other Goods. The Prosecutor de 'd, That his House was broke open, and he lost his Goods; and being afterwards informed by one Bonner, (a corner in Newgate) that Mills was concern'd in doing it, he apprehended him, and he confess'd it, and Kennedy

Coreference data

⊟ Default

- ☑ the House
- ☐ Thomas Cutbe
- ☐ William ☐Jam
- ☐ John Meards
- ☐ WILLIAM HUMI
- ☐ Richard He
- ☐ Mary Carter
- ☐ Elizabeth Coles
- ☐ Edward Garrets
- ☐ Henry Leman
- ☐ George Ve
- ☐ Daniel Craker
- ☐ Jane White
- ☐ St. Giles

Figure 3: Viewing and deleting all annotations for an entity



Meta data

England-Germany
Team
Group
City
Stadium
Date
Referee
Score

Event data

foul
Beckham
Player 2
Location
Destination
Time
From Time
Until Time
Score

Figure 4: The MUMIS interface