



# NER & Beyond

Branislava Šandrih, Ranka Stanković  
26. Novembar 2019



# NER & mnogo više...

<http://nerbeyond.jerteh.rs/>

- Tri modula:
  - Konverzije
  - Prepoznavanje i obeležavanje
  - Evaluacija



# Konverzije

- Prvo, pogledajmo standardne formate za predstavljanje imenovanih entiteta

# XML

```
<xml>На целом путу од <LOC>Београда</LOC> до <LOC>Крушевца  
</LOC> <PERS>Рајко</PERS> је био кипео од веселости. Срце му  
је било пуно необичне милине; очи су му светлеле неисказаном  
радошћу. Чинило му се као да је робовао у конаку <PERS>  
Мустаф-агином</PERS>, па се сад слободан и срећан враћа у  
завичај да види мајку, браћу и родбину.</xml>
```

# CoNLL02

|                 |               |             |   |                      |   |               |               |
|-----------------|---------------|-------------|---|----------------------|---|---------------|---------------|
| <xml>           | O             | Срце        | O | Чинило               | O | срећан        | O             |
| На              | O             | му          | O | му                   | O | враћа         | O             |
| целом           | O             | је          | O | се                   | O | у             | O             |
| путу            | O             | било        | O | као                  | O | завичај       | O             |
| од              | O             | пуно        | O | да                   | O | да            | O             |
| <b>Београда</b> | <b>B-Loc</b>  | необичне    | O | је                   | O | види          | O             |
| до              | O             | милине      | O | робовао              | O | мајку         | O             |
| <b>Крушевца</b> | <b>B-Loc</b>  | ;           | O | у                    | O | ,             | O             |
| <b>Рајко</b>    | <b>B-Pers</b> | очи         | O | конаку               | O | <b>Мустаф</b> | <b>B-Pers</b> |
| је              | O             | су          | O | <b>- I-Pers</b>      |   | браћу         | O             |
| био             | O             | му          | O | <b>агином I-PERS</b> |   | и             | O             |
| кипео           | O             | светлеле    | O | ,                    | O | родбину       | O             |
| од              | O             | неисказаном | O | па                   | O | .             | O             |
| веселости       | O             | радощу      | O | се                   | O | </xml>        | O             |
| .               | O             | .           | O | сад                  | O |               |               |
|                 |               |             |   | слободан             | O |               |               |

# BRAT

## TXT

На целом путу од Београда до Крушевца Рајко је био кипео од веселости. Срце му је било пуно необичне милине; очи су му светлеле неисказаном радошћу. Чинило му се као да је робовао у конаку Мустаф-агином, па се сад слободан и срећан враћа у завичај да види мајку, браћу и родбину.

## ANN

T1 LOC 22 30 Београда  
T2 LOC 34 42 Крушевца  
T3 PERS 43 48 Рајко  
T4 PERS 194 207 Мустаф-агином

# I BRAT ↔ CoNLL02

### BRAT → CoNLL02 (.py source)

Select an archive:  primer.zip  
primer.zip

Select tokenizer: Unspecified ▾

### CoNLL02 → BRAT (.py source)

Select an archive:  No file selected.

# I BRAT ↔ XML

### BRAT → XML

Select an archive:  No file selected.

---

### XML → BRAT

Tags to be ignored (separated by spaces):

Tags that contain NEs (separated by spaces):

Select an archive:  primer.zip



# I XML → CoNLL02

**XML → CoNLL02**

Select an archive:  No file selected.

# I Mapiranje etiketa

|              | DEMO                  | EVENT                 | FAC                              | ORG                              | OTHER                            | PERS                  | PLACE                            | ROLE                  | ROLEMISC              | TEMPORAL              | WORK                  | ignore                | remove                |
|--------------|-----------------------|-----------------------|----------------------------------|----------------------------------|----------------------------------|-----------------------|----------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| CARDINAL     | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| DATE         | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| EVENT        | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| FAC          | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input checked="" type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| GPE          | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| LANGUAGE     | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| LAW          | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| LOC          | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| MONEY        | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| NORP         | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| ORDINAL      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| ORG          | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| PERSON       | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| PERS_FIRST_M | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| PERS_FULL_M  | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| PERS_LAST_M  | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| PRODUCT      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| QUANTITY     | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| TIME         | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| WORK_OF_ART  | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| p            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| sample       | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| samples      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/>            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

[Download](#)

# II Prepoznavanje i obeležavanje entiteta

**Perform NER with spaCy (output in BRAT and XML formats)**

Select an archive:  No file selected.

Replace < and > with < and > for XML files

Select language model:

**Perform NER with StanfordNER (output in XML format)**

Select an archive:  No file selected.

Replace < and > with < and > for XML files

Select language model:

# III Statistika

## NER stats on .ann files

Select an archive:

Browse...

No file selected.

Upload

# III Evaluacija

**Gemini tool for NER evaluation, [GitHub](#)**

Matching type score:  
 Strict  Weighted

Alignment:  
 Greedy  Max

Visualise by annotation type (at least one file has to be XML, separated with spaces):

Other options:  
 Generate a CSV  Compare annotations by each type

Select an archive:  No file selected.

# Bonus: vizuelizacija

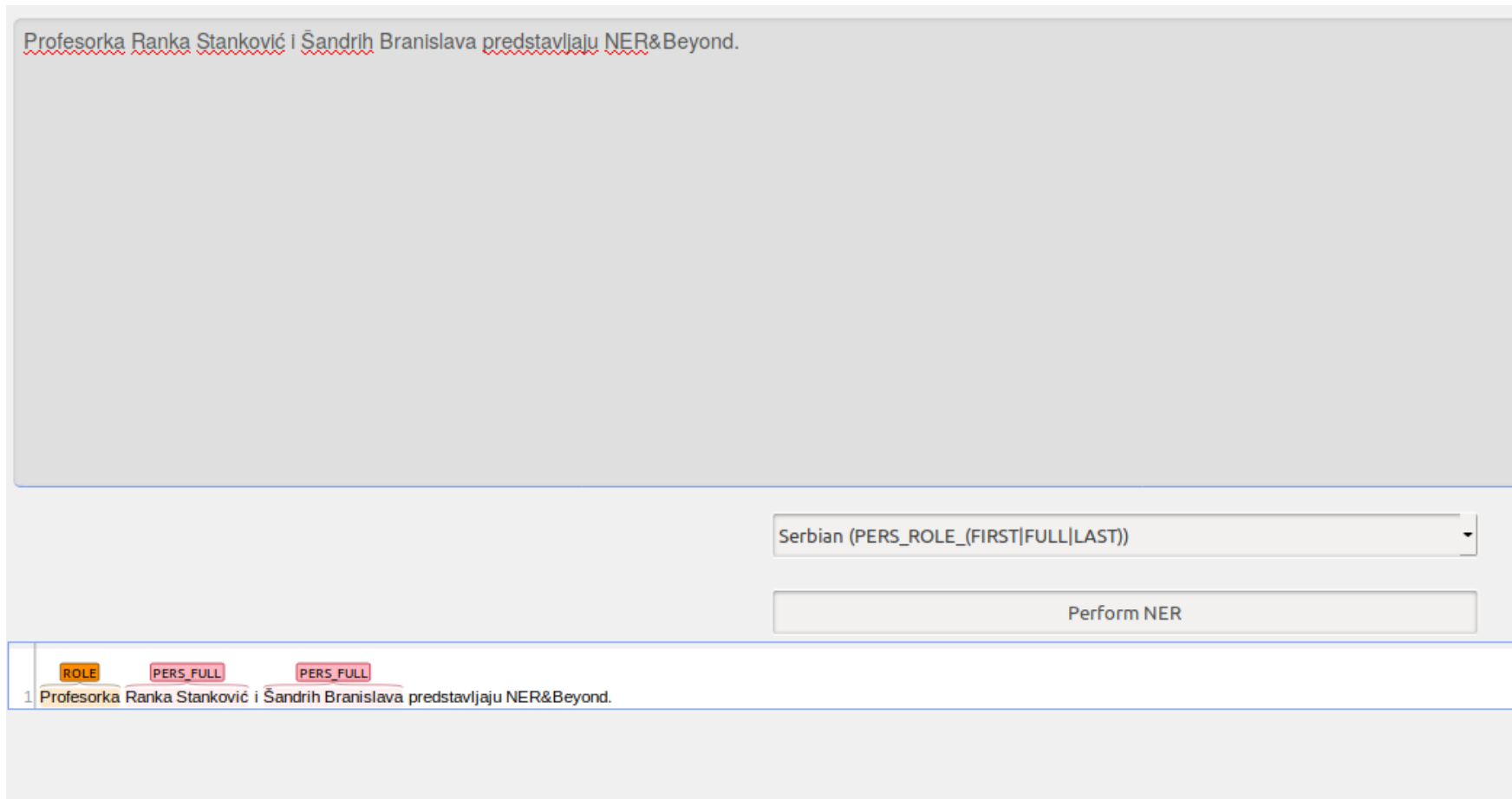
<http://ner.jerteh.rs/>

Profesorka Ranka Stanković i Šandrih Branislava predstavljaju NER&Beyond.

Serbian (PERS\_ROLE\_(FIRST|FULL|LAST))

Perform NER

1 **ROLE** **PERS\_FULL** **PERS\_FULL**  
Profesorka Ranka Stanković i Šandrih Branislava predstavljaju NER&Beyond.



# BRAT

<http://brat.jerteh.rs/>

6 <p n="SRP1892051">На целом путу од Београда до Крушевца Рајко је био кипео од веселости. Срце му је било пуно необичне милине; очи су му светлеле неисказаном радошћу. Чинило му се као да је робовао у конаку Мустаф-агином, па се сад слободан и срећан враћа у завичај да види мајку, браћу и родбину. Као да га негде на путу или на крају тога пута чека неко неизмерно благо или нека изванредна срећа, тако га је радосна жудња вукла некуда — и сам није знао куда, и к нечему — и сам није знао к чему. Сва му се Србија чинила као да је баштина његове задруге; свако му је дрво изгледало као да је кита босиљка, свака стена румена; и саме вране и свраке нису му се више чиниле да су тице злослутнице, него само као старије у црнину завијене сестре плавосивих дивљих голубица, што падаху преда њ на пут само да опет прну у вис, те да се зрак потресе и фијуком зајечи кад се сребрни таламбас витих крила њихових узлепрша. Ништа се око његовом измакло није: ни орао који је високо под небом господствено кружио, ни зец који погурен и сав услахирен скакаше од џбуна до џбуна. Сва су му поља мирисала као да су босиљком засађена. Чинило му се као да кроз растове шуме шумадијске бруји поветарац, који се понео чак отуда са сињег мора, па преко снежног Дурмитора пао у четинарске горе ужичке, да се у њима напоји мирисом од јела и борова, па да онда проструји кроз Шумадију, те да је као неку цркву тамњаном окади. На све то и на сваку ситницу обраћао је он пажњу мађарској господи, јер му се чинило да свега тога нигде ван Србије нема..</p>  
7 </sample>

# Rezultati evaluacije za spaCy

## [http://portal.jerteh.rs/dr\\_gemini](http://portal.jerteh.rs/dr_gemini)

<p n="SRP1892051">На целом путу од Београда до Крушевца Рајко је био кипео од веселости. Срце му је било пуно необичне милине; очи су му светлеле неисказаном радошћу. Чинило му се као да је робовао у конаку Мустаф-агином, па се сад слободан и срећан враћа у завичај да види мајку, браћу и родбину. Као да га негде на путу или на крају тога пута

<sample><p n="ENG184501331">"Indeed!"</p>

<p n="ENG184501332">"Yes: I could not help thinking of that to-day," said Gerard, "when he questioned me with his mincing voice and pulled the wool with his cursed white hands and showed it to his dame, who touched it with her little finger; and his daughters who tossed their heads like pea-hens—Lady Joan and Lady Maud. Lady Joan and Lady Maud!" repeated Gerard in a voice of bitter sarcasm. "I did not care for the rest; but I could not stand that Lady Joan and that Lady Maud. I wonder if my Sybil saw them."</p>

<p n="ENG184501333">In the meantime, Sybil had been sent for by Mrs Trafford. She had inferred from the message that the guests had departed, and her animated cheek showed the eagerness with which she had responded to the call. Bounding along with a gladness of the heart which lent additional lustre to her transcendent brightness, she suddenly found herself surrounded in the garden by Lady Maud and her friends. The daughter of Lord de Mowbray, who could conceive nothing but humility as the cause of her alarmed look, attempted to re-assure her by condescending volubility, turning often to her friends and praising in admiring interrogatories Sybil's beauty.</p>

<p n="ENG184501334">"And we took advantage of your absence," said Lady Maud in a tone of amiable artlessness, "to find out all about you. And what a pity we did not know you when you were at the convent, because then you might have been constantly at the castle; indeed I should have insisted on it. But still I hear we are neighbours; you must promise to pay me a visit, you must indeed. Is not she beautiful?" she added in a lower but still distinct voice to her friend. "Do you know I think there is so much beauty among the lower order."</p>

<p n="ENG184501335">Mr Mountchesney and Lord Milford poured forth several insipid compliments, accompanied with some speaking looks which they flattered themselves could not be misconstrued. Sybil said not a word, but answered each flood of phrases with a cold reverence.</p>

<p n="ENG184501336">Undeterred by her somewhat haughty demeanour, which Lady Maud only attributed to the novelty of her situation, her ignorance of the world, and her embarrassment under this overpowering condescension, the good-tempered and fussy daughter of Lord de Mowbray proceeded to re-assure Sybil, and to enforce on her that this perhaps unprecedented descent from superiority was not a mere transient courtliness of the moment, and that she really might rely on her patronage and favourable feeling.</p>

### Annotation <PERS>

Text File1 : ENG18450\_Disraeli\_sample.xml

Text File2 : GOLD STANDARD



# F1

## Annotation <PERS>

Text File1 : ENG18450\_Disraeli\_sample.xml

Text File2 : GOLD STANDARD

| A   | B    | C    | D        | E  | F           | G                      | H     | I     | J    | K    |
|-----|------|------|----------|--|-------------|------------------------|-------|-------|------|------|
| T1  | 135  | 141  | ORG      | Marney                                   | 0.047272727 | Marney                 | PLACE | T904  | 135  | 141  |
| T2  | 254  | 262  | ORG      | Egremont                                 | 0.14529915  | Egremont               | PERS  | T907  | 254  | 262  |
| T3  | 266  | 275  | PERS     | Lady Maud                                | 0.46420047  | Lady Maud              | PERS  | T908  | 266  | 275  |
| T4  | 362  | 374  | ORG      | Marney Abbey                             | 0.061728396 | Marney Abbey           | FAC   | T909  | 362  | 374  |
| T5  | 861  | 870  | DEMO     | Christian                                | 0           | Christian architecture | WORK  | T9014 | 861  | 883  |
| T6  | 891  | 900  | PERS     | Lady Maud                                | 0.46420047  | Lady Maud              | PERS  | T9015 | 891  | 900  |
| T7  | 1255 | 1263 | ORG      | Egremont                                 | 0.14529915  | Egremont               | PERS  | T9018 | 1255 | 1263 |
| T8  | 1532 | 1539 | FAC      | Nobles—                                  | 0           | Nobles                 | ROLE  | T23   | 1532 | 1538 |
| T9  | 1566 | 1575 | PERS     | Lady Maud                                | 0.46420047  | Lady Maud              | PERS  | T9022 | 1566 | 1575 |
| T11 | 1920 | 1928 | ORG      | Egremont                                 | 0.14529915  | Egremont               | PERS  | T9026 | 1920 | 1928 |
| T12 | 2254 | 2263 | PERS     | Lady Maud                                | 0.46420047  | Lady Maud              | PERS  | T9028 | 2254 | 2263 |
| T13 | 2468 | 2480 | PERS     | Lady Mowbray                             | 0.46420047  | Lady Mowbray           | PERS  | T9033 | 2468 | 2480 |
| T14 | 2482 | 2490 | ORG      | Egremont                                 | 0.14529915  | Egremont               | PERS  | T9034 | 2482 | 2490 |
| T15 | 2576 | 2587 | DEMO     | Hippocrates                              | 0.012731481 | Hippocrates            | PERS  | T9036 | 2576 | 2587 |
| T16 | 2659 | 2666 | ROLEMISC | English                                  | 0.21875     | English                | DEMO  | T9038 | 2659 | 2666 |
| T17 | 2719 | 2725 | PERS     | Norman                                   | 0.013157895 | Norman                 | DEMO  | T9040 | 2719 | 2725 |
| T18 | 2742 | 2747 | ORG      | Saxon                                    | 0.044217687 | Saxon                  | DEMO  | T9041 | 2742 | 2747 |
| T19 | 3027 | 3034 | ORG      | England                                  | 0.047272727 | England                | PLACE | T9042 | 3027 | 3034 |
| T21 | 3349 | 3354 | ORG      | Dandy                                    | 0.14529915  | Dandy                  | PERS  | T9048 | 3349 | 3354 |
| T22 | 3495 | 3533 | ORG      | n=""ENG184503028"">"The National Holiday | 0           | The National Holiday   | EVENT | T902  | 3513 | 3533 |
| T23 | 3556 | 3560 | PERS     | Mick                                     | 0.46420047  | Mick                   | PERS  | T9054 | 3556 | 3560 |
| T24 | 3644 | 3654 | PERS     | de Mowbray                               | 0           | Lord de Mowbray's pa   | FAC   | T903  | 3639 | 3661 |
| T25 | 3911 | 3921 | ORG      | Dandy Mick                               | 0.14529915  | Dandy Mick             | PERS  | T9058 | 3911 | 3921 |
| T26 | 4998 | 5004 | ORG      | Bishop                                   | 0.033426184 | Bishop                 | ROLE  | T9064 | 4998 | 5004 |
| T27 | 5069 | 5086 | PERS     | George the Fourth                        | 0.46420047  | George the Fourth      | PERS  | T9065 | 5069 | 5086 |
| T28 | 5114 | 5133 | DEMO     | the Roman Catholics                      | 0           | Roman Catholics        | ROLE  | T9066 | 5118 | 5133 |
| T29 | 5269 | 5279 | PERS     | de Mowbray                               | 0           | Lord de Mowbray        | PERS  | T8    | 5264 | 5279 |
| T30 | 5367 | 5374 | PERS     | Mowbray                                  | 0.03526971  | Mowbray                | FAC   | T49   | 5367 | 5374 |
| T31 | 5403 | 5409 | ORG      | Bishop                                   | 0.033426184 | Bishop                 | ROLE  | T9072 | 5403 | 5409 |
| T32 | 5457 | 5465 | ORG      | Mowedale                                 | 0           | Mowedale works         | FAC   | T50   | 5457 | 5471 |
| T33 | 5473 | 5479 | PERS     | Gerard                                   | 0.46420047  | Gerard                 | PERS  | T9074 | 5473 | 5479 |
| T34 | 5746 | 5761 | PERS     | Ursula Trafford                          | 0.46420047  | Ursula Trafford        | PERS  | T9077 | 5746 | 5761 |
| T35 | 6533 | 6539 | PERS     | Gerard                                   | 0.46420047  | Gerard                 | PERS  | T9086 | 6533 | 6539 |
| T36 | 6756 | 6765 | PERS     | Lady Joan                                | 0.46420047  | Lady Joan              | PERS  | T9087 | 6756 | 6765 |
| T37 | 6770 | 6779 | PERS     | Lady Maud                                | 0.46420047  | Lady Maud              | PERS  | T9088 | 6770 | 6779 |
| T38 | 6781 | 6790 | PERS     | Lady Joan                                | 0.46420047  | Lady Joan              | PERS  | T9089 | 6781 | 6790 |
| T39 | 6795 | 6804 | PERS     | Lady Maud                                | 0.46420047  | Lady Maud              | PERS  | T9090 | 6795 | 6804 |
| T40 | 6816 | 6822 | PERS     | Gerard                                   | 0.46420047  | Gerard                 | PERS  | T9091 | 6816 | 6822 |
| T41 | 6910 | 6919 | PERS     | Lady Joan                                | 0.46420047  | Lady Joan              | PERS  | T9092 | 6910 | 6919 |
| T42 | 6929 | 6938 | PERS     | Lady Maud                                | 0.46420047  | Lady Maud              | PERS  | T9093 | 6929 | 6938 |
| T43 | 7040 | 7052 | PERS     | Mrs Trafford                             | 0.46420047  | Mrs Trafford           | PERS  | T9095 | 7040 | 7052 |
| T44 | 7364 | 7373 | PERS     | Lady Maud                                | 0.46420047  | Lady Maud              | PERS  | T9096 | 7364 | 7373 |
| T45 | 7410 | 7410 | PERS     | de Mowbray                               | 0.46420047  | de Mowbray             | PERS  | T910  | 7410 | 7410 |

# Radovi u toku

