Distant Reading

JePTex – (jerteh.rs)

Wikipedia: Wikiproject ELTeC

ELTeC@Wikidata

Belgrade Training School: 23.03.2023, 14.00-15.30

"Exploring ELTeC: Use-Cases for Information Extraction and Analysis"

Ranka Stanković & Milica Ikonić Nešić

University of Belgrade

Call for Belgrade Training School: "Exploring ELTeC: Use-Cases for Information Extraction and Analysis" – Distant Reading for European Literary History (distant-reading.net)

BelgradeTrening – Google диск

Key topics for ELTeC@Wikidata

- **Wiki-ELTeC data lecture session**:
  - Wikidata introduction on Serbian novels
  - Wiki-ELTeC schema
    - metadata from header
    - main characters
    - their relations
    - places
  - Pipeline for Wiki-ELTeC data population
  - Predefined SPARQL query exploration
- **Hands-on**:
  - OpenRefine tool for
  - QuickStatement for …
  - population of Wikidata with 10 novels for 6 languages (deu, eng, fra, hun, por, slv)

SrpELTeC data structure in Wikidata

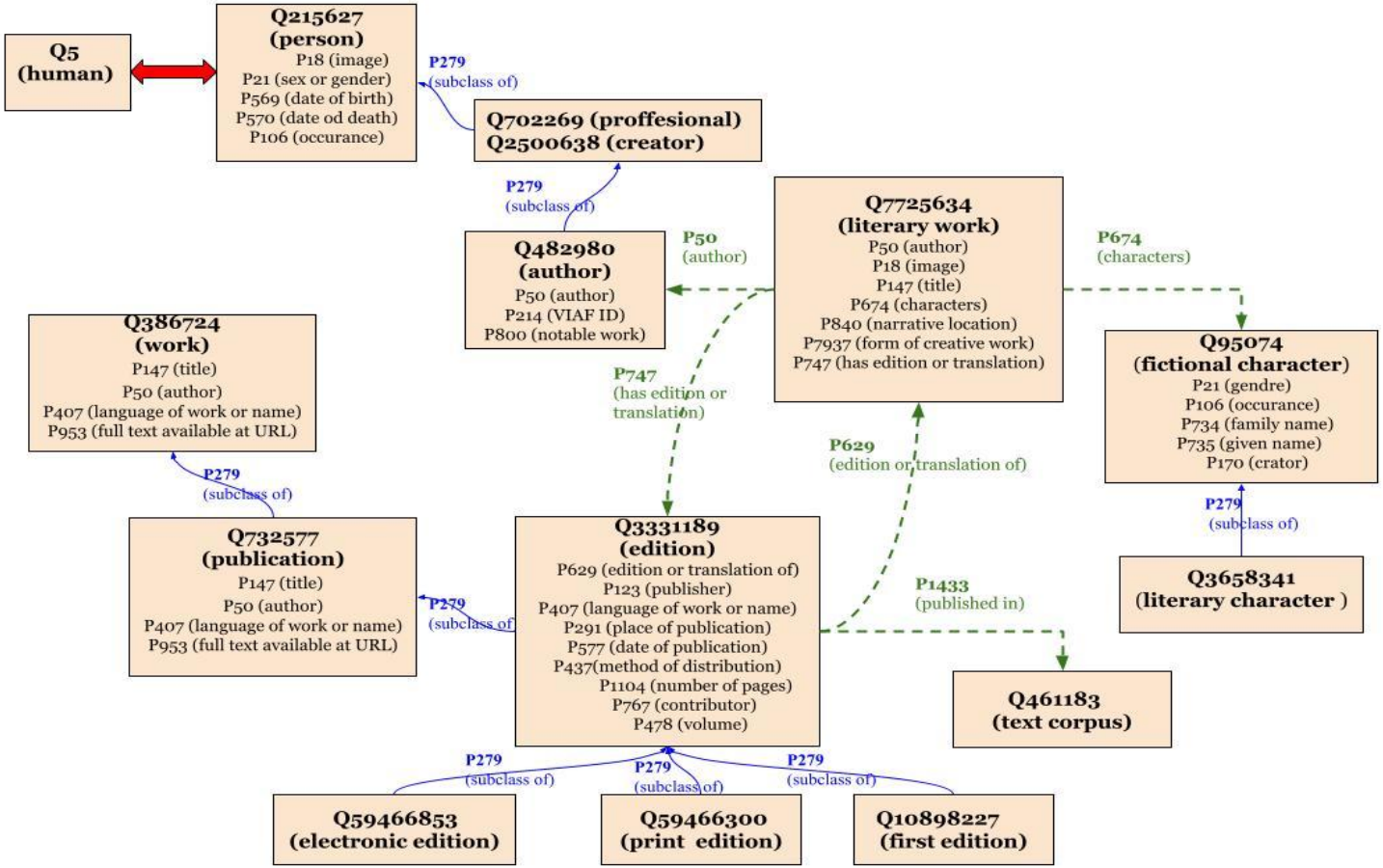
Data preparation and mapping with Wikidata

The work on the ELTeC corpus envisages a basic annotation for all subcollections, the so-called level 1 annotation. The annotation consists of marking the basic structural elements of the text (chapters and other units) and some basic textual elements. The annotation was made in accordance with the TEI recommendations (TEI Consortium 2021), where from a rich set of elements that define these recommendations, only a subset was selected as mandatory or allowed.

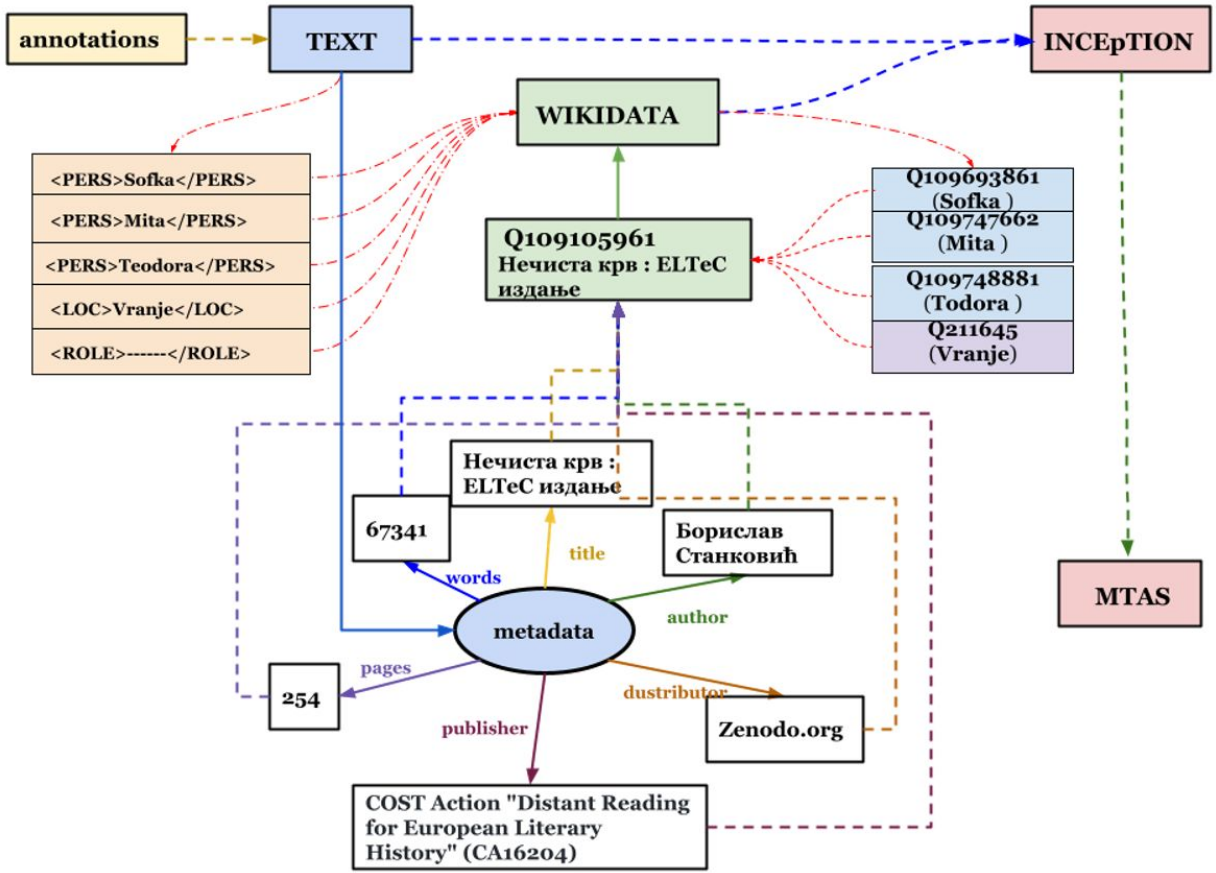
A metadata header <TeiHeader> is required for each text annotated in accordance with TEI recommendations, so it is the case with all ELTeC corpus novels. The mandatory header elements are uniform for all collections and they must contain:

1. Description of the electronic edition, which includes the title of the work and the name of the author as well as the statements of responsibility (scanning, correction, annotation), date of publication, size (measured by the number of words). The author and the work can be joined by identifiers, such as viaf and wikidata.
2. A brief catalog description of the first edition and the edition used as the source for ELTeC (if different from the first edition).
3. Description of the text in terms of meeting the balance criteria (e.g. author's gender, size, date of publication,...).
4. Review of all changes to the digital edition since its first publication.

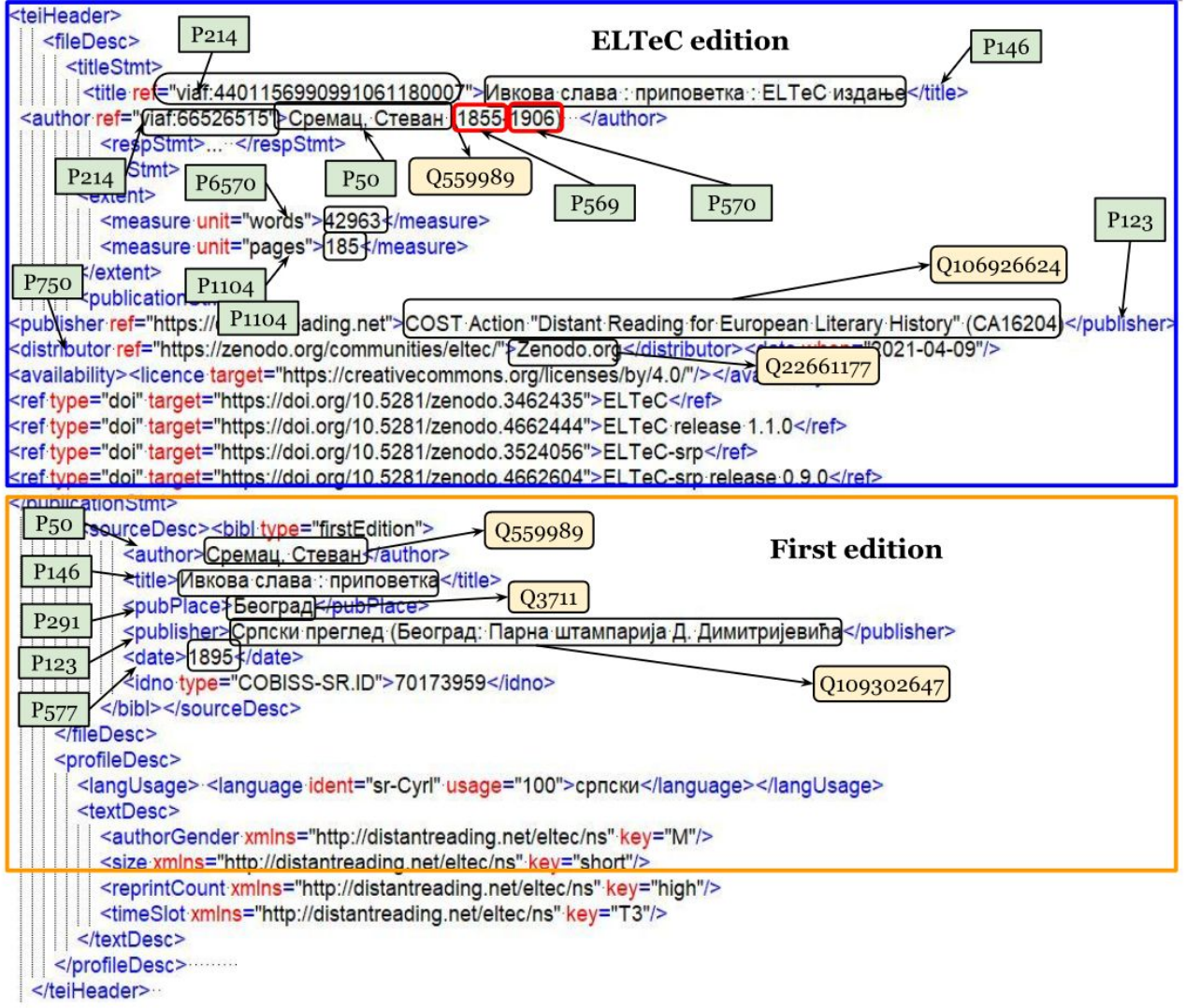
Classes and instances



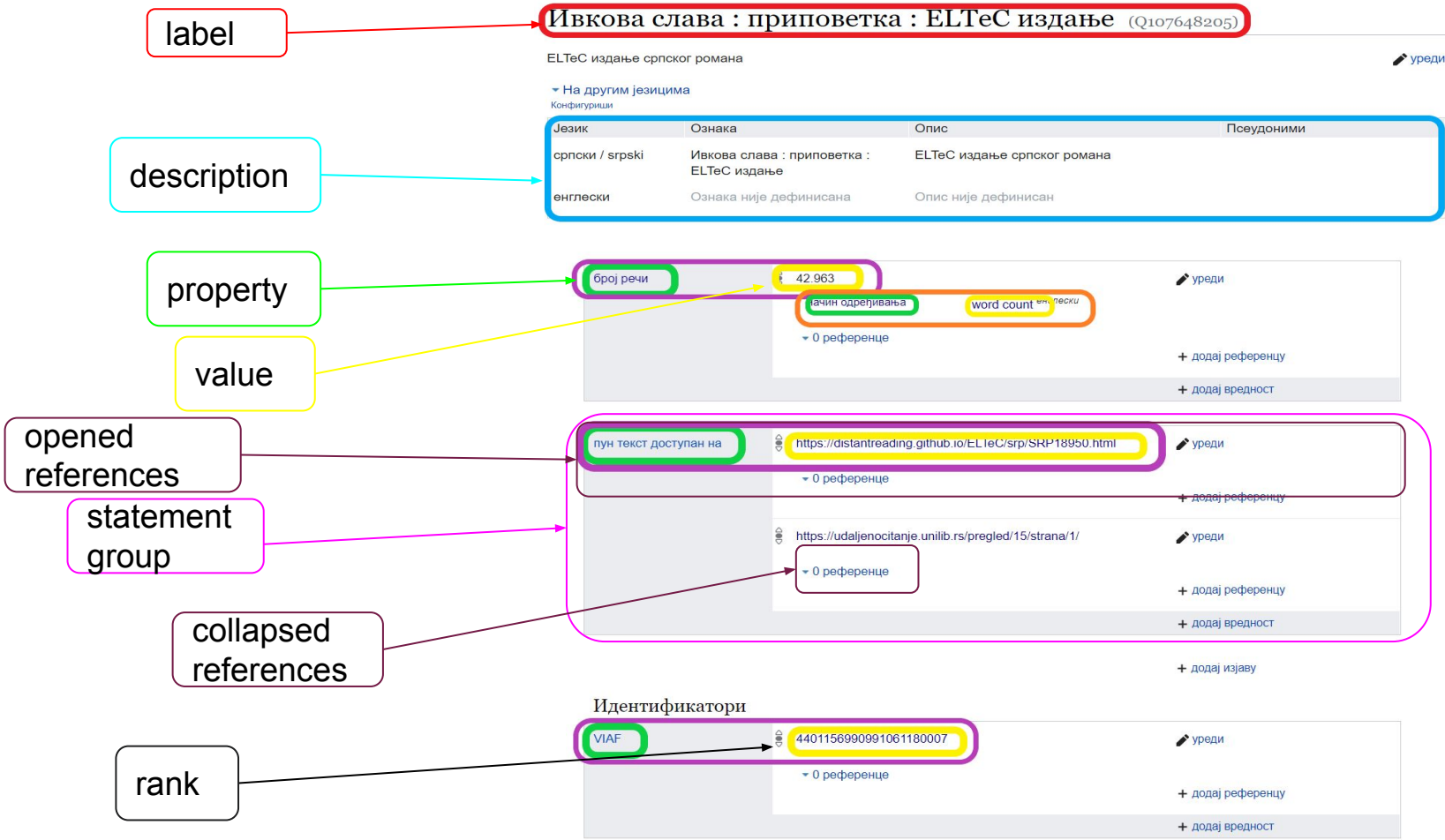
From metadata across Wikidata to INCEpTION



Mapping between **SrpELTeC metadata** and **Wikidata**



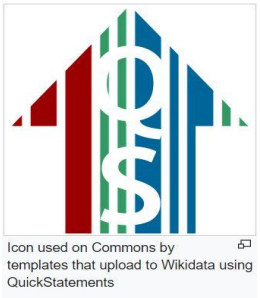
Structure of SrpELTeC Wikidata items (example of ELTeC edition)



Automatization with web resources

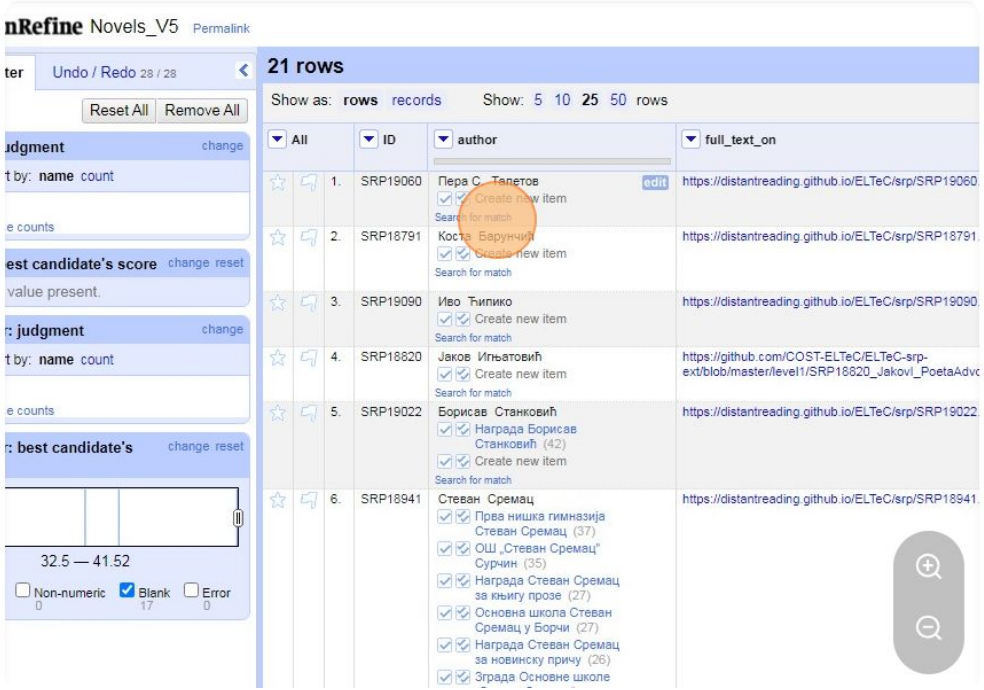
Since the automation of the process of data preparation and import was envisaged, the different solutions were analysed and finally synergy of OpenRefine and QuickStatements tools was chosen as the best choice.

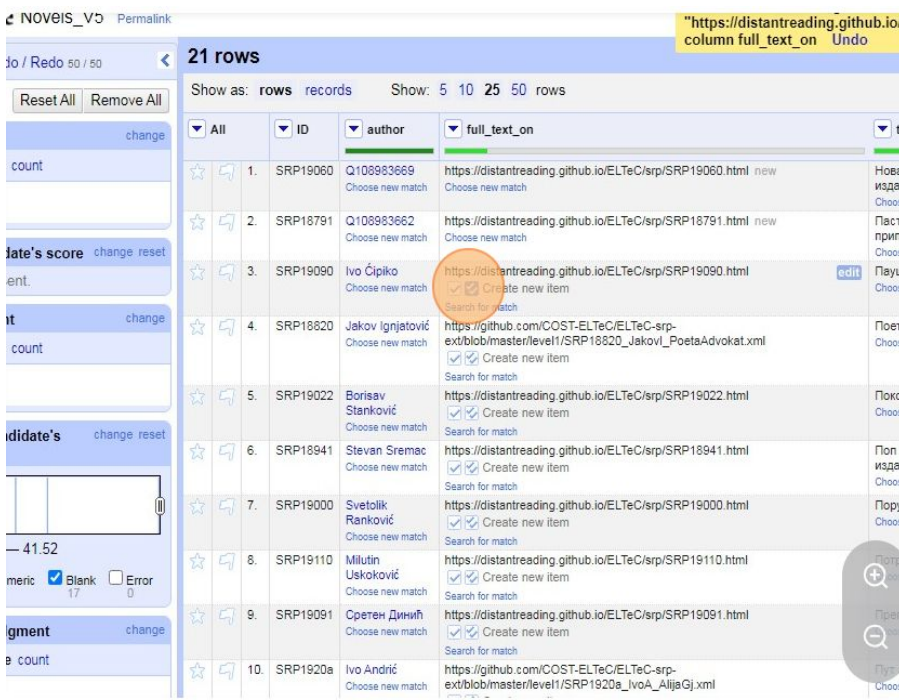
[Tutorial for automatization using OpenRefine and QuickStatement]



Tutorial for automation using OpenRefine and QuickStatement

… Step by step, ilustrated tutorial with SrpELTeC text collection

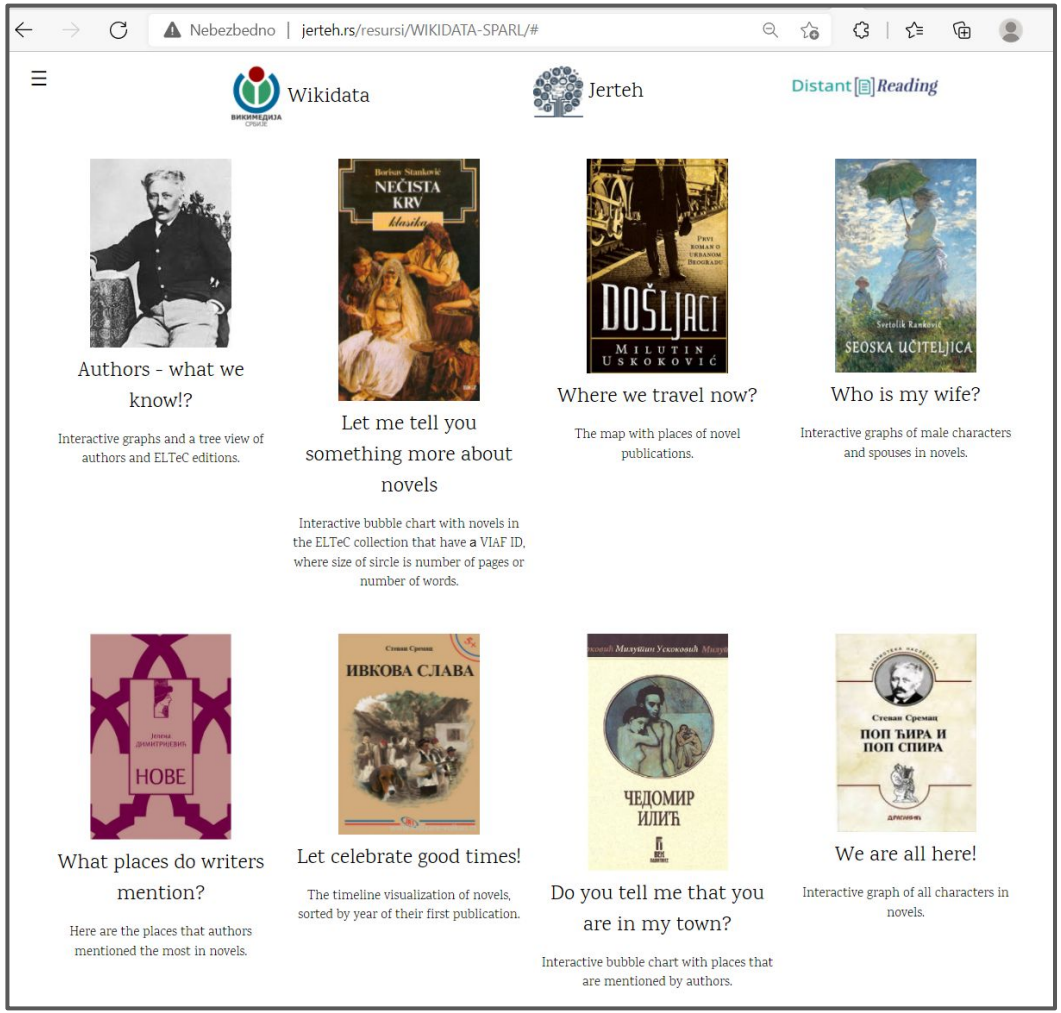




When creating new item "Create new item" manual option was chosen for each cell, to check the results.
There is an option to automatically create a new item, for each cell in the column.

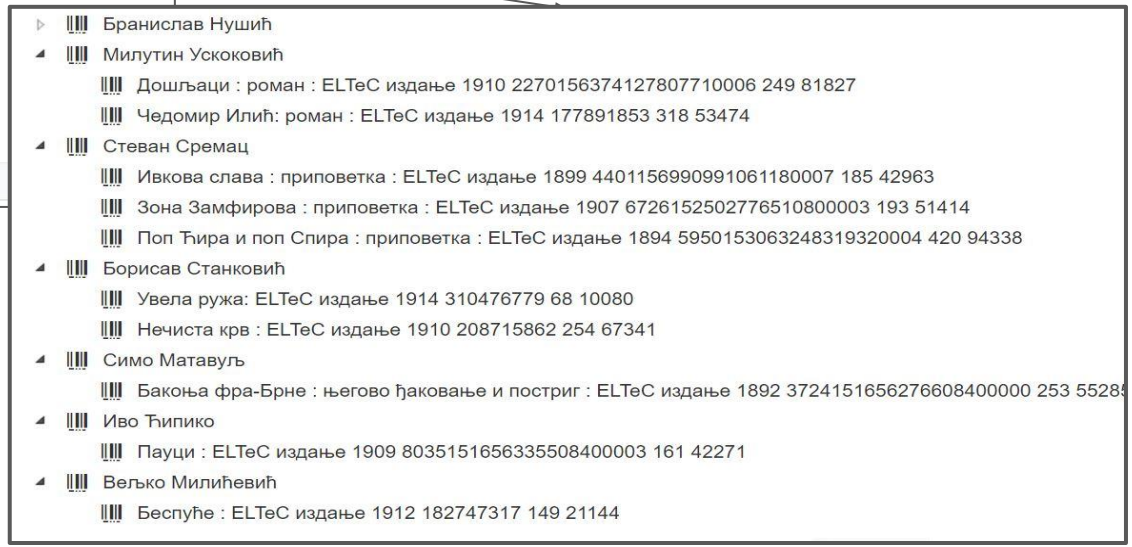
Visualization

http://jerteh.rs/resursi/WIKIDATA-SPARL/



Overview of SrpELTeC@Wikidata by SPARQL queries

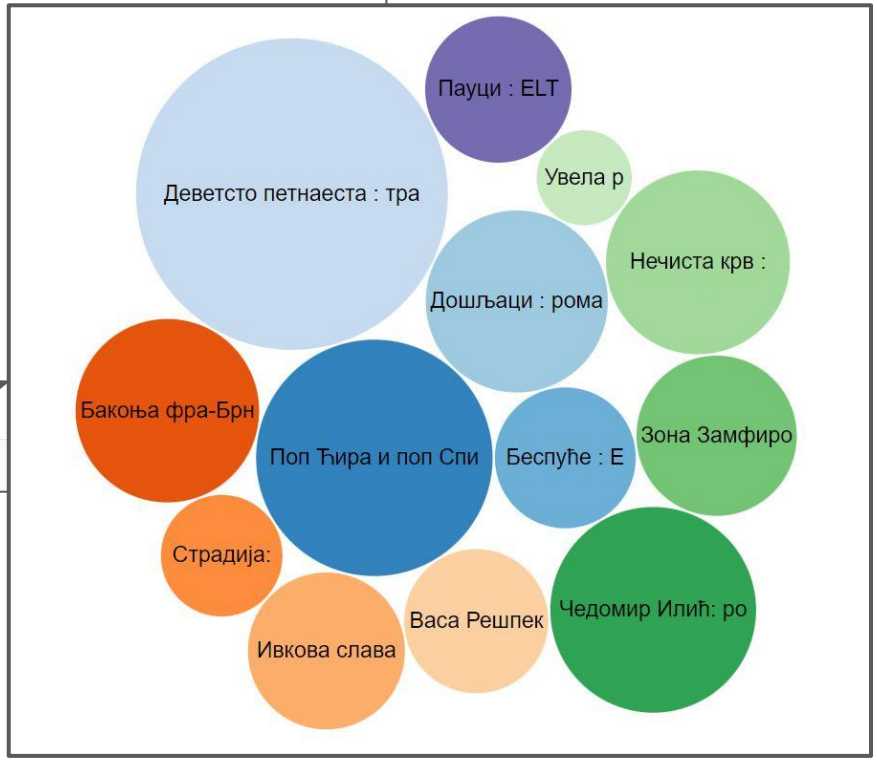
Visualization of authors of novels.
https://w.wiki/4Lja



Wikidata Query Service | Примери | Query Builder | Помоћ | Још алата

```
#defaultView:BubbleChart
SELECT DISTINCT ?novel ?novelLabel ?num_pages ?num_words ?viaf
WHERE {
  # novel published in (P1433) SrpELTeC collection (Q106936149)
  ?novel wdt:P1433 wd:Q106936149;
         # number of pages (P1104)
         wdt:P1104 ?num_pages;
         # number of words (P6570)
         wdt:P6570 ?num_words;
         # viaf id (P214)
         wdt:P214 ?viaf.
SERVICE wikibase:label
{ bd:serviceParam wikibase:language "sr,[AUTO_LANGUAGE],en". }}
```

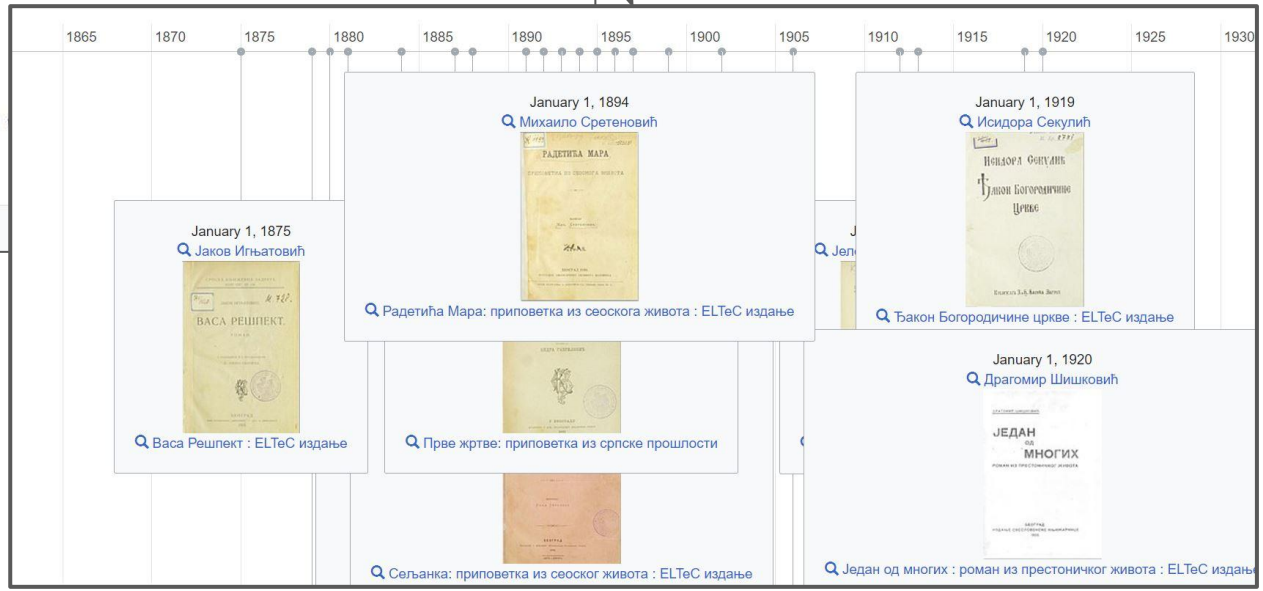
Bubble chart of novels with number of pages, number of words and VIAF ID. https://w.wiki/4LjC



Wikidata Query Service | Примери | Query Builder | Помоћ | Још алата

```
#defaultView:Timeline
SELECT DISTINCT ?novel ?novelLabel ?image ?date ?author ?authorLabel
WHERE {
  # novel published in (P1433) ELTeC collection (Q106927517)
  ?novel wdt:P1433 wd:Q106927517;
         # has edition or translation (P747)
         wdt:P747 ?edition.
  # edition instance of (P31) first edition (Q10898227)
  ?edition wdt:P31 wd:Q10898227;
           # image (P18)
           wdt:P18 ?image.
  # optional date of publication (P577)
  OPTIONAL { ?edition wdt:P577 ?date. }
  # optional author (P50)
  OPTIONAL { ?novel wdt:P50 ?author. }
  SERVICE wikibase:label { bd:serviceParam
```

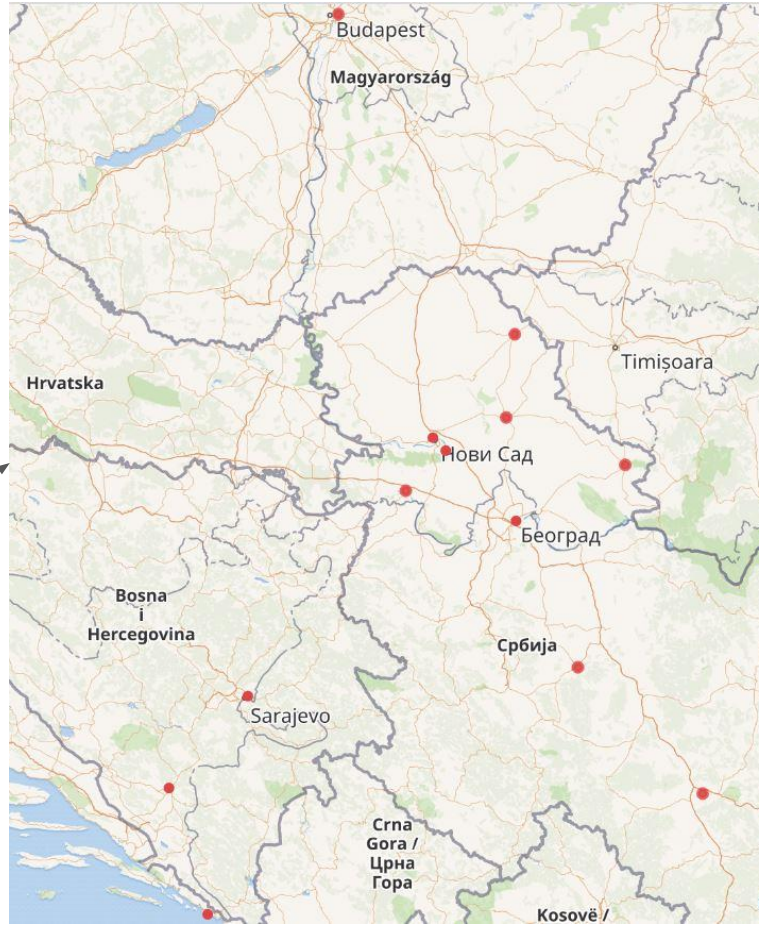
Timeline visualization of novels sorted by year of first publication
https://w.wiki/4LjZ



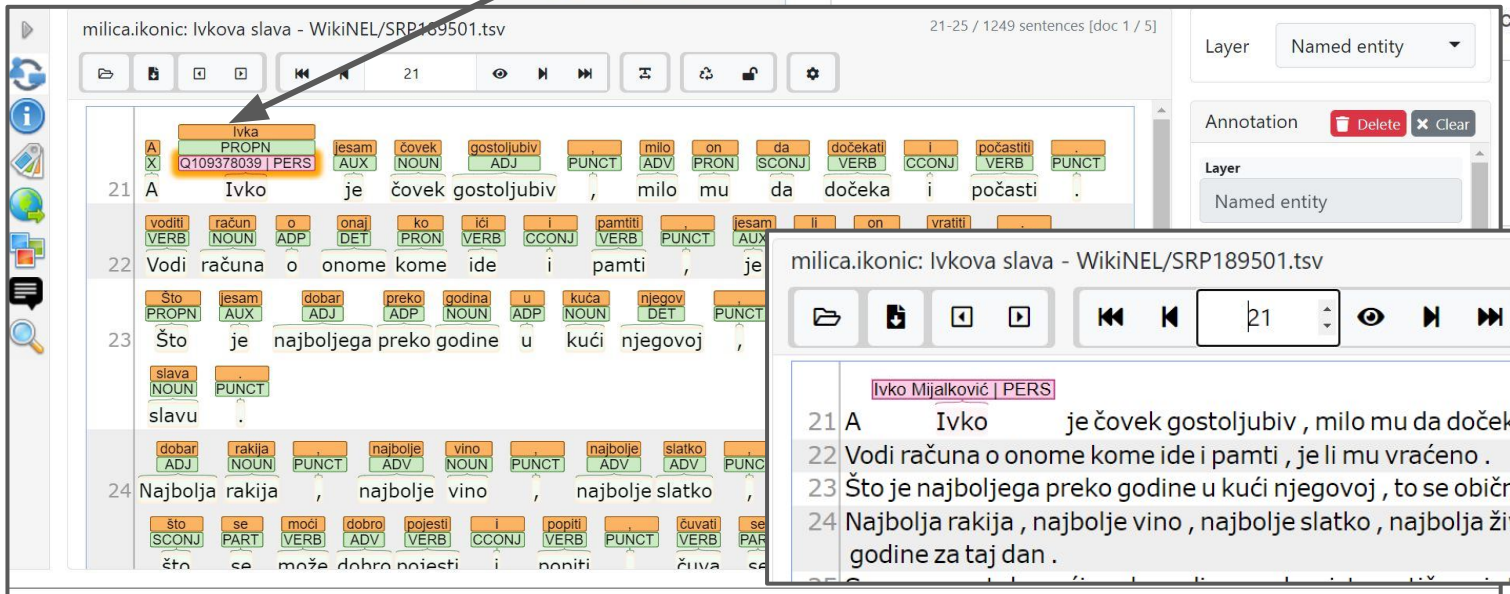
Wikidata Query Service | Примери | Query Builder | Помоћ | Још алата

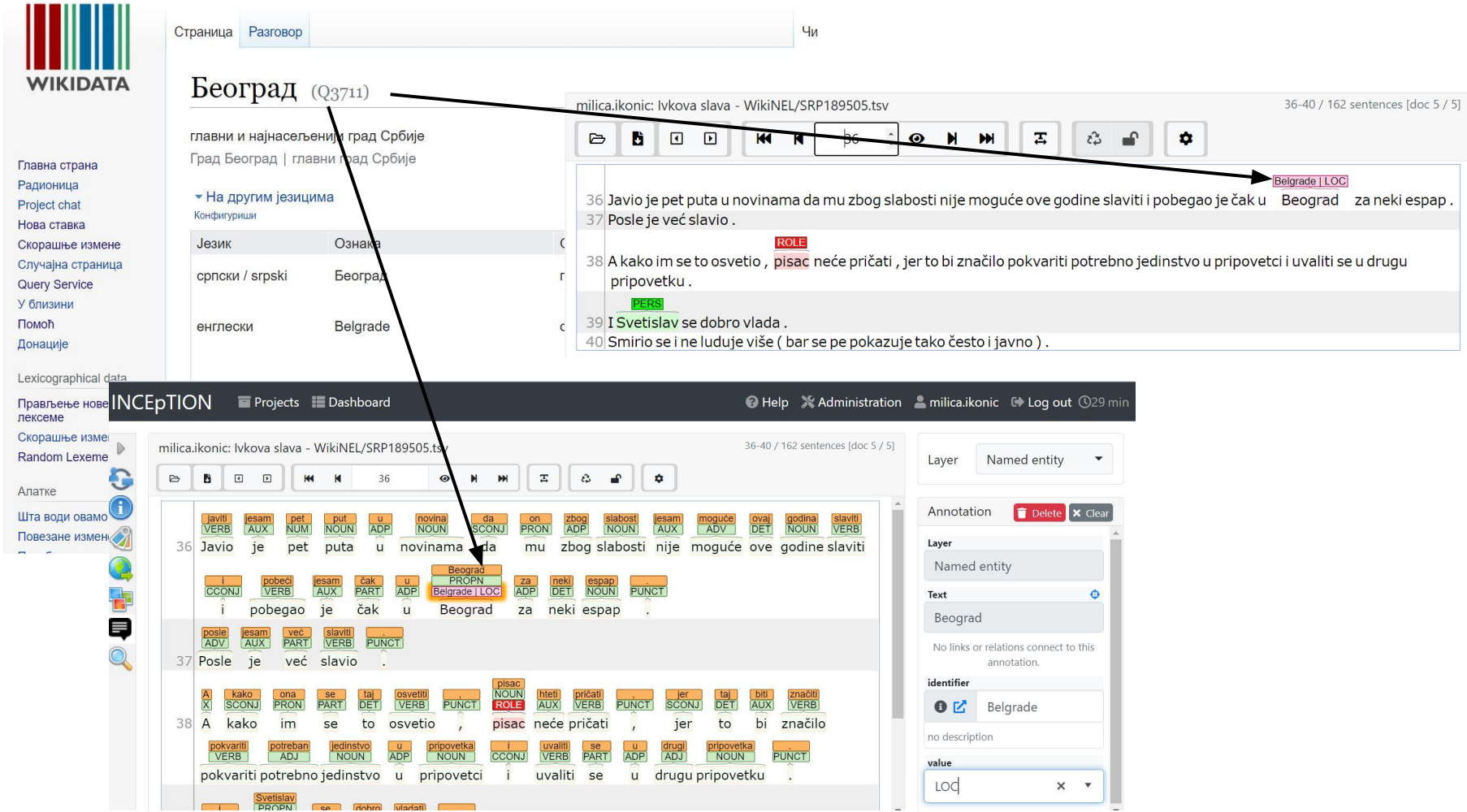
```
#defaultView:Map
# presenting place of first edition
SELECT DISTINCT ?place ?coor
WHERE {
  # edition instance of (P31) edition (Q3331189)
  ?edition wdt:P31 wd:Q3331189;
        # published in (P143) ELTeC collection (Q106927517)
         wdt:P1433 wd:Q106927517;
         # publication place (P291)
         wdt:P291 ?place.
         # place coordinate location (P625)
         ?place wdt:P625 ?coor.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "sr,[AUTO_LANGUAGE],en". }
}
```

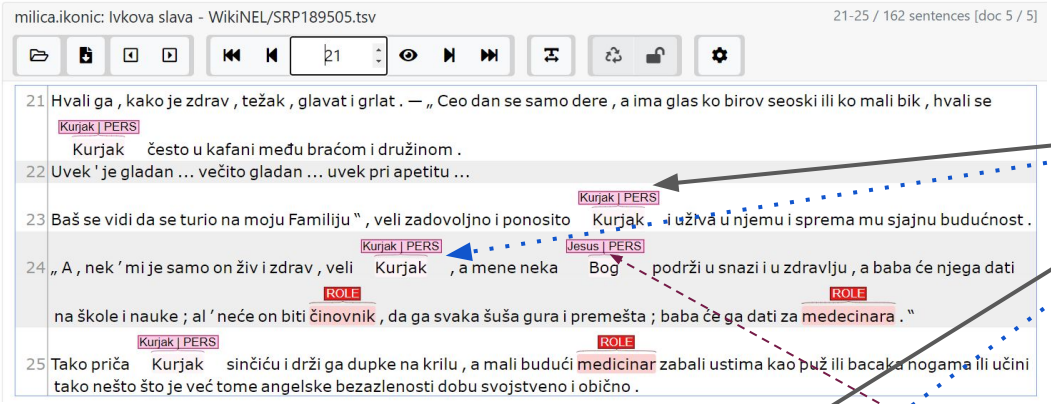
Map with places of novel publication https://w.wiki/4Lje

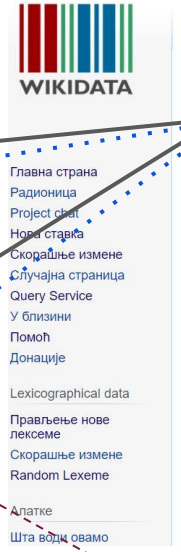


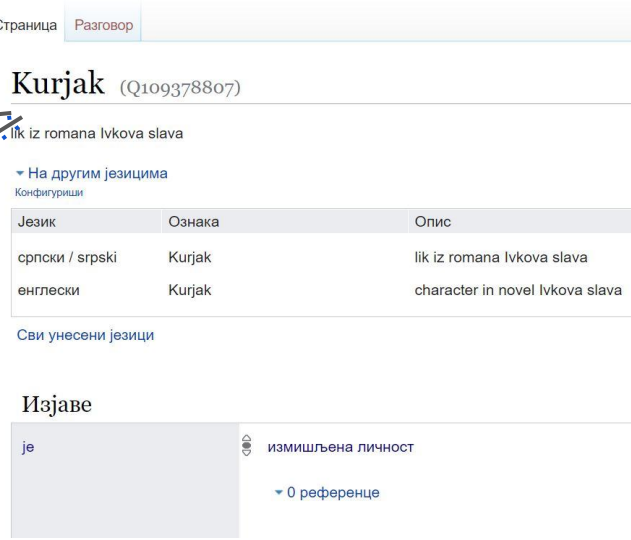
Integration Wikidata with [INCEpTION]

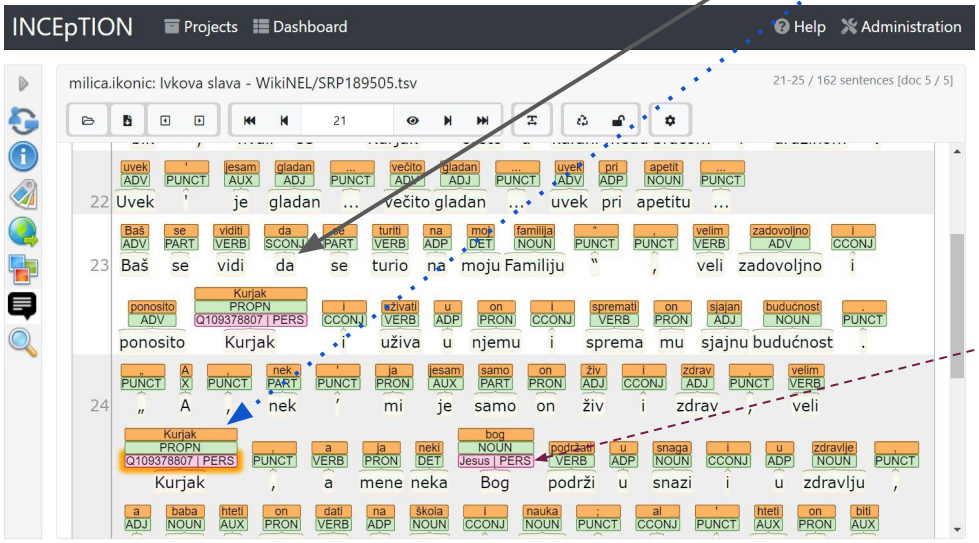


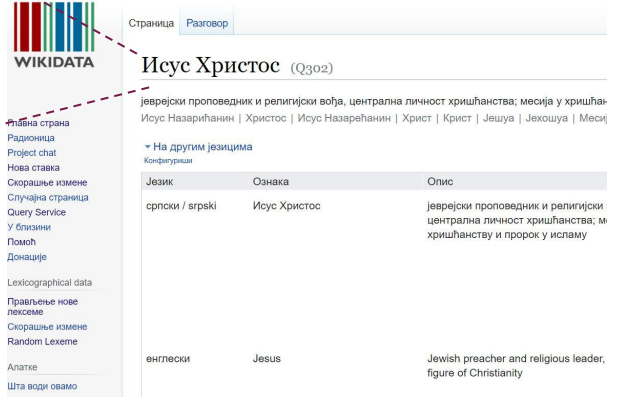

WIKIDATA
Страница
Разговор
Београд (Q3711)
главни и најнасељенији град Србије
Град Београд | главни град Србије
На другим језицима
Конфигуриши
Језик
Ознака
српски / srpski
Београд
енглески
Belgrade
Главна страна
Радионица
Project chat
Нова ставка
Скорашње измене
Случајна страница
Query Service
У близини
Помоћ
Донације
Lexicographical data
Прављење нове лексеме
Скорашње измене
Random Lexeme
Алатке
Шта води овамо
Повезане измене
milica.ikonic: Ivkova slava - WikiNEL/SRP189505.tsv
36-40 / 162 sentences [doc 5 / 5]
36 Javio je pet puta u novinama da mu zbog slabosti nije moguće ove godine slaviti i pobegao je čak u Beograd za neki espap .
37 Posle je već slavio .
38 A kako im se to osvetio , pisac neće pričati , jer to bi značilo pokvariti potrebno jedinstvo u pripovetci i uvaliti se u drugu pripovetku .
39 I Svetislav se dobro vlada .
40 Smirio se i ne luduje više (bar se pe pokazuje tako često i javno) .
Belgrade | LOC
ROLE
PERS
INCEpTION
Projects
Dashboard
Help
Administration
milica.ikonic
Log out
29 min
Layer
Named entity
Annotation
Delete
Clear
Text
Beograd
No links or relations connect to this annotation.
identifier
Belgrade
no description
value
LOC


milica.ikonic: lvkova slava - WikiNEL/SRP189505.tsv
21-25 / 162 sentences [doc 5 / 5]
21
21 Hvali ga , kako je zdrav , težak , glavat i grlat . — „ Ceo dan se samo dere , a ima glas ko birov seoski ili ko mali bik , hvali se
Kurjak | PERS
Kurjak često u kafani među braćom i družinom .
22 Uvek ' je gladan ... večito gladan ... uvek pri apetitu ...
Kurjak | PERS
23 Baš se vidi da se turio na moju Familiju " , veli zadovoljno i ponosito Kurjak , i uživa u njemu i sprema mu sjajnu budućnost .
Kurjak | PERS
Jesus | PERS
24 „ A , nek ' mi je samo on živ i zdrav , veli Kurjak , a mene neka Bog podrži u snazi i u zdravlju , a baba će njega dati
ROLE
na škole i nauke ; al ' neće on biti činovnik , da ga svaka šuša gura i premešta ; baba će ga dati za
ROLE
medecinara . "
Kurjak | PERS
25 Tako priča Kurjak sinčiću i drži ga dupke na krilu , a mali budući
ROLE
medicinar zabali ustima kao puž ili bacaka nogama ili učini
tako nešto što je već tome angelske bezazlenosti dobu svojstveno i obično .


WIKIDATA
Главна страна
Радионица
Project chat
Нова ставка
Скорашње измене
Случајна страница
Query Service
У близини
Помоћ
Донације
Lexicographical data
Прављење нове лексеме
Скорашње измене
Random Lexeme
Алатке
Шта води овамо


Страница
Разговор
Kurjak (Q109378807)
lik iz romana lvkova slava
На другим језицима
Конфигуриши
Језик
Ознака
Опис
српски / srpski
Kurjak
lik iz romana lvkova slava
енглески
Kurjak
character in novel lvkova slava
Сви унесени језици
Изјаве
је
измишљена личност
0 референце


INCEpTION
Projects
Dashboard
Help
Administration
milica.ikonic: lvkova slava - WikiNEL/SRP189505.tsv
21-25 / 162 sentences [doc 5 / 5]
22 Uvek ' je gladan ... večito gladan ... uvek pri apetitu ...
23 Baš se vidi da se turio na moju Familiju " , veli zadovoljno i
ponosito Kurjak i uživa u njemu i sprema mu sjajnu budućnost .
24 „ A , nek ' mi je samo on živ i zdrav , veli
Kurjak , a mene neka Bog podrži u snazi i u zdravlju ,


WIKIDATA
Страница
Разговор
Исус Христос (Q302)
јеврејски проповедник и религијски вођа, централна личност хришћанства; месија у хришћ
Исус Назарећанин | Христос | Исус Назарећанин | Христ | Крист | Јешуа | Јехошуа | Месиј
На другим језицима
Конфигуриши
Језик
Ознака
Опис
српски / srpski
Исус Христос
јеврејски проповедник и религијски
централна личност хришћанства; м
хришћанству и пророк у исламу
енглески
Jesus
Jewish preacher and religious leader,
figure of Christianity

SrpELTeC @Wikidata status

Wikidata has 100 novels from SrpELTeC plus 19 novels from SrpELTeC-ext collection with associated items for the first edition and electronic SrpELTeC editions.

That means that we automatically added approx **4300** statements.

Towards hands-on

- The presented example can be seen as succesful show cases
- Similar approach can be used for other language subcollection

Vol 21 No 2 (2021): Infotheca - Journal for Digital Humanities | Infotheca - Journal for Digital Humanities (bg.ac.rs)

Serbian ELTeC Sub-Collection in Wikidata | Infotheca - Journal for Digital Humanities (bg.ac.rs)

Step by step authors:
https://scribehow.com/shared/OpenRefine__0K5TU3J9SmalqRBYePhdhQ

Step by step novels:
https://scribehow.com/shared/OpenRefine__AOPgpu7dRqmxUkYBtVbGgQ

Hands-on: Data

- Wikidata account creation
- Dataset for each participant separate
- Wikidata exploration
- Insert of one novel manualy
- OpenRefine automatic preparation
- QuickStatement insert
- SPARQL exploration
- INCEPTION:tsbgd:ELTec@Wikidata, http://inception.jerteh.rs/

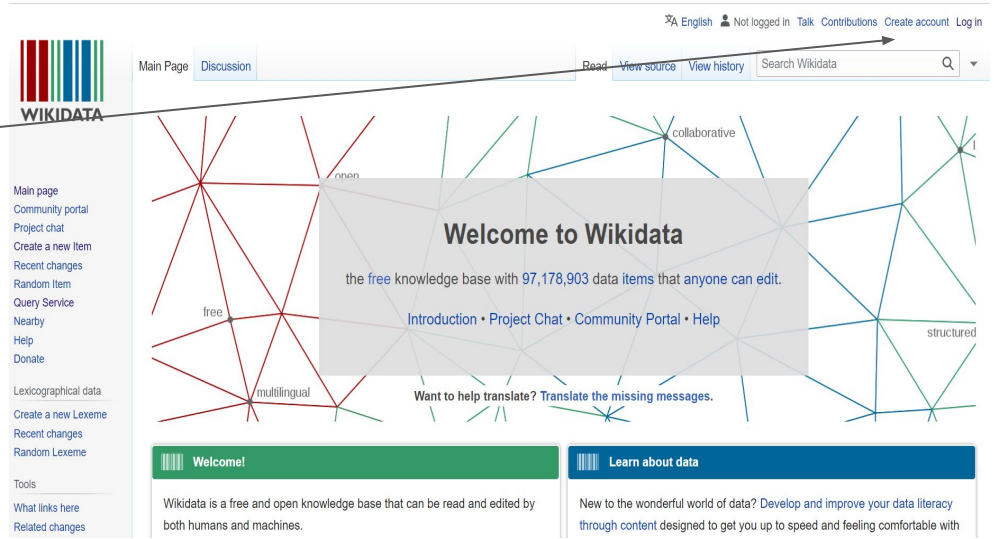
Working google sheets for hands-on
https://docs.google.com/spreadsheets/d/10Nl4w6VOF7NN3MBuugDy0rCd84uPLjHV/edit?usp=drive_web&ouid=101444873968805165718&rtpof=true

Wikidata account creation

Follow steps:

Go to https://www.wikidata.org/wiki/Wikidata:Main_Page

Create account



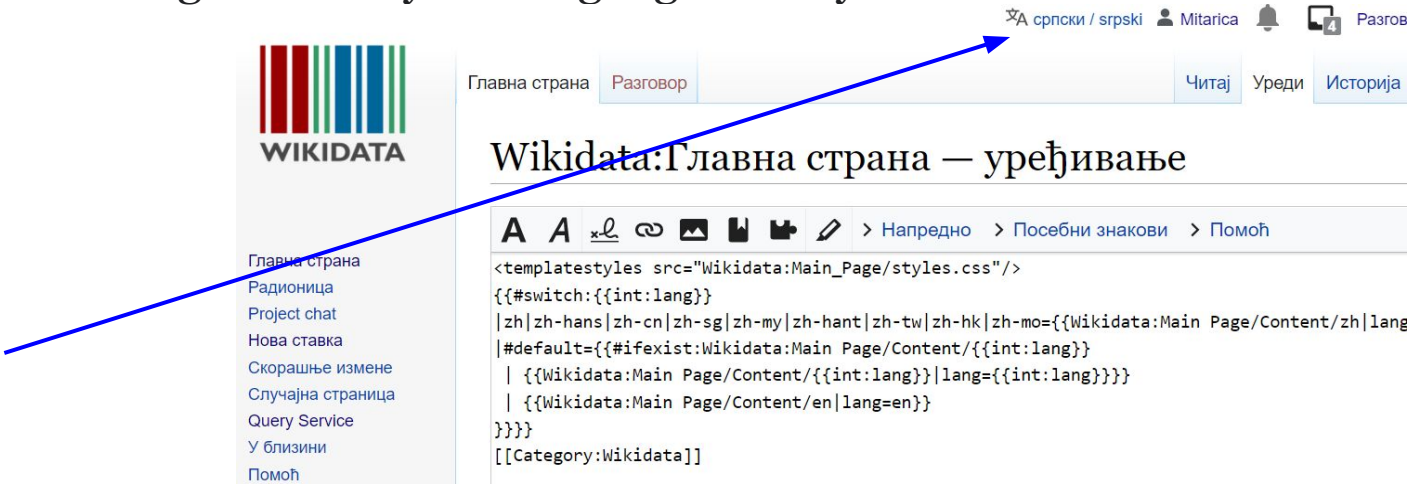
Insert of one novel manualy from ELTeC-fra

Label (name of the novel)
Description
instance of P31 literary work Q772563
title P1476 (name of novel)
author P50 (see in label)
language of work or name P407 fra
form of creative work P7937 novel

For example see:
https://www.wikidata.org/wiki/Q2699195

How to add a new language in the decription

To use the labelLister gadget on an item, **click on the "Label List" tab at the top right of the item page**. After selecting it, you will see a pop-up with a list of all labels, descriptions, and aliases available by language code. Choose language that you need, and than go back to your language. Now you can see different labels.



Automate the editing of ELTeC editions of entered novels

Using **OpenRefine** and **QuickStatements** a faster way to add ELteC editions of novels that participants added manually will be shown.

Participants must fill column **QID_novel** in shared Excel file, where will add QID of novel which they manualy created.

Than starts the magic!

It’s time to do by yourself!

Wikidata is a multilingual knowledge base of structured data that anyone can edit. The project relies on the efforts of people — just like you! — from all around the world who work together to collect and maintain data in **more than 200 languages**.

Participient will translate titles of ELTeC editions (Serbian, Franch, English…). Then using **OpenRefine** and **QuickStatements** they will add **all descriptions in Wikidata and connect with ELTeC editions.**

At the end let’s SPARQL!
https://query.wikidata.org/

ELTeC@Wikidata

Rankas (Wikimedia Serbia) ranka@rgf.rs
Mitarica (Wikimedia Serbia) milica.ikonic.nesic.fil@gmail.com