



LEXICAL DATA MASTERCLASS 2018 #LEXMC

BILJANA LAZIĆ,

UNIVERZITET U BEOGRADU, RUDARSKO-GEOLOŠKI FAKULTET

SEMINAR DRUŠTVA ZA JEZIČKE RESURSE I TEHNOLOGIJE – JERTEH,

BEOGRAD, 20.12.2018.

CILJ RADIONICE

- Deljenje iskustava, metoda i tehnika za stvaranje i upotrebu digitalnih leksičkih podataka
- Testiranje TEI (Text Encoding Initiative) Lex-0
- Proizvodnja dokumentacije i materijala za trening koji će postati dostupni široj publici, uz prikaz izazova i predloga rešenja prilikom modelovanja i obogaćivanja leksičkih resursa

2. PO REDU “LEXICAL DATA MASTERCLASS” 2018, OD. 3. DO 7. DECEMBRA 2018.

<https://lexmc18.sciencesconf.org/>

- 20 učesnika
- Organizatori: **DARIAH**, **Berlin Brandenburg Academy of Sciences (BBAW)**, **Inria** i **Belgrade Center for Digital Humanities (BCDH)**
- Uz podršku: **CLARIN** i **European Lexicographic Infrastructure (ELEXIS)**
- Deo zajedničkog francusko-nemačkog programa uz podršku **Ministarstva za visoko obrazovanje, istraživanje i inovacije Francuske (grant do 600 EUR)**

Predavači:

Laurent Romary (Inria), **Toma Tasovac** (BCDH, DARIAH-EU), **Marie Puren** (Inria), **Jack Bowers** (Austrian Academy of Sciences - ÖAW), **Axel Herold** (BBAW), **Mohamed Khemakhem** (Paris 7 University, Inria-ALMAAnCH lab)

	<u>Mon. 03</u>	<u>Tue. 04</u>	<u>Wed. 05</u>	<u>Thu. 06</u>	<u>Fri. 07</u>
9:00		Introduction to encoding dictionaries with the TEI guidelines Extracting content from legacy dictionaries with Groid-dictionary	Introduction to XSLT Advanced Oxygen techniques	Data management Advanced Path for searching dictionary content	Participants' Symposium
10:00		Coffee break	Coffee break	Coffee break	Coffee break
11:00		Workshop of participants' projects Workshop of participants' projects GROBID-Dictionary session	Workshop of participants' projects Workshop of participants' projects	Workshop of participants' projects Workshop of participants' projects	Participants' Symposium
12:00		Lunch	Lunch	Lunch	
13:00					
14:00	Coffee break				
15:00	Introduction to the Lexical Data Masterclass 2018 Project presentations	Workshop of participants' projects Workshop of participants' projects GROBID-dictionary session	Workshop of participants' projects Workshop of participants' projects	Workshop of participants' projects Workshop of participants' projects	
16:00		Coffee break	Coffee break	Coffee break	
17:00	Introduction to technical settings	Workshop of participants' projects Workshop of participants' projects GROBID-dictionary session	Workshop of participants' projects Workshop of participants' projects	Workshop of participants' projects Workshop of participants' projects	
18:00		The VICA Dictionaries: Working on a Virtual Research Environment			
19:00					
20:00		Dinner at restaurant Goodtime, Hausvogteiplatz 11			
21:00					

RASPORED

- 1. dan
 - Uvod u Lexical Data Masterclass
- Prezentacije učesnika
- Tehnička pomoć
- Tri dana predavanja – radionica
- 5. dan
 - Simpozijum učesnika

PREZENTACIJE PROJEKATA UČESNIKA

- Igniter prezentacija - 3 minuta – 12 slajdova po 15 sekundi
- Visoka zastupljenost “struje” slovenskih jezika - učesnici iz Poljske (1), Češke (1), Bugarske (2), Makedonije (1), Slovenije (1) i Srbije (4)
- Osim nas, tu si bili i učesnici iz Italije, Turske, Engleske, Škotske, Francuske, Nemačke i Portugala...

Charly Moerth

-

GROBID-DICTIONARIES

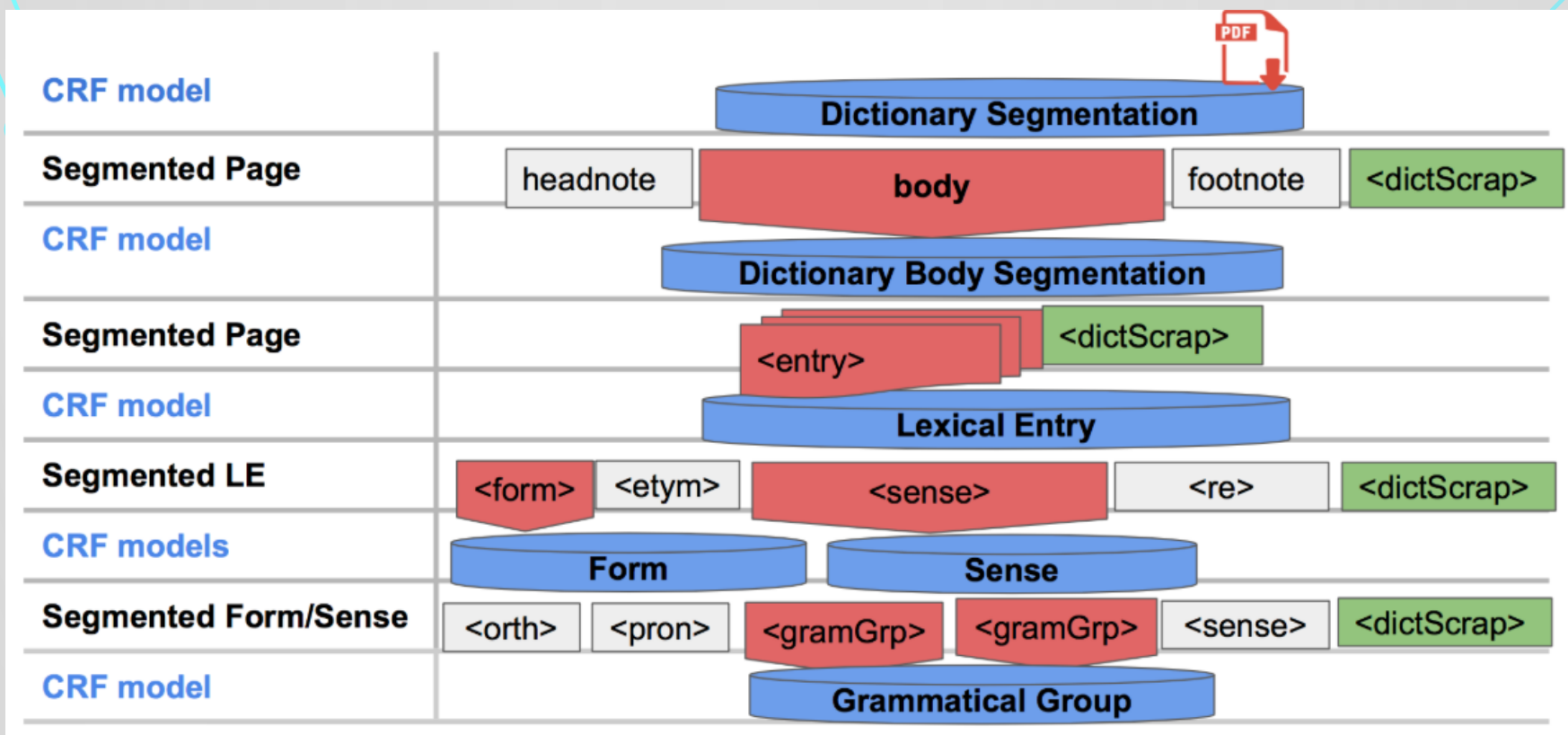
- GROBID – skr. od GeneRation Of Bibliographic Data
 - <https://grobid.readthedocs.io/en/latest/Introduction/>
 - Biblioteka za mašinsko učenje pisana u Javi (Java) namenjena ekstrakciji, parsiranju i restrukturiranju sirovih dokumenata, prvenstveno u PDF-u
 - Razvoj započet 2008. godine (od 2011. godine je program otvorenog koda)
 - Praktična primena: za potrebe citatnih menadžera (Mendeley), “društvenih mreža”

(ResearchGate), i drugih “digitalnih repozitorijuma” HAL Research Archive, the European Patent Office, INIST, CERN, Internet Archive

- GROBID-Dictionaries
 - [Mohamed Khemakhem, Luca Foppiano, Laurent Romary. Automatic Extraction of TEI Structures in Digitized Lexical Resources using Conditional Random Fields. Electronic lexicography, eLex 2017, Sep 2017, Leiden, Netherlands](#)
 - [Mohamed Khemakhem, Axel Herold, Laurent Romary. Enhancing Usability for Automatically Structuring Digitised Dictionaries. GLOBALEX workshop at LREC 2018, May 2018, Miyazaki, Japan](#)

TEHNIČKI ZAHTEVI KOJE JE TREBALO ISPUNITI PRE DOLASKA NA RADIONICU

- **“Docker technology**, which will ensure the setup of GROBID-Dictionaries on your machine, **is not supported by versions of Windows earlier than Windows 7. Windows 10 Home also does not support Docker.** Please mind also that "Surface Laptops" having **less than 8GB RAM will not allow you to run the tool.**
- So if you meet one of these conditions, please think of bringing another machine.
- Please make sure you execute the first 2 steps in the **Docker** instructions on github before the day of the workshop:
https://github.com/MedKhem/grobid-dictionaries/wiki/Docker_Instructions



Izvor: <https://github.com/MedKhem/grobid-dictionaries>

*CRF – Conditional Random Fields – Uslovna nasumična polja

KORACI ZA RAD NAKON POKRETANJA DOCKER-A

proizvodnja materijala za treniranje

```
java -jar /grobid/grobid-dictionaries/target/grobid-dictionaries-0.4.3-SNAPSHOT.one-jar.jar -dIn resources/DIRECTORY_OF_YOUR_PDF -dOut resources -exe createTrainingDictionarySegmentation
```

Dokumenti tipa **tei.xml** služe za ručnu anotaciju
Nakon anotacije se proizvedeni dokumenti pomeraju u odgovarajući katalog (za trening i evaluaciju):
tei.xml u toyData/dataset/MODEL_NAME/corpus/tei katalog
Sve ostalo osim **.rng** i **.css** u toyData/dataset/MODEL_NAME/corpus/raw katalog

treniranje modela

```
mvn generate-resources -P train_naziv_modela -e
```

pokretanje servisa za veb kako bi se video izlaz modela (TEI)

```
mvn -Dmaven.test.skip=true jetty:run-war
```

<http://localhost:8080> ili <http://192.168.99.100:8080>

ŠTA JE BILJANA RADILA...

Bergeisen (Eisen); рударски сечивасти чекић, чекић, цепач, сечивац.

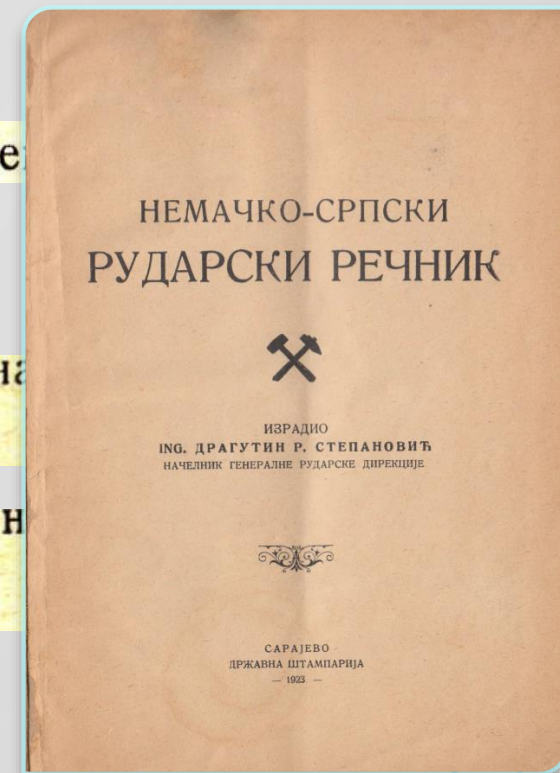
Kohlenbergmann; (рудар) угљар, угљарац; в. Kohlenhauer

Klaubarbeit; одбирање, пребирање, требљење, требез. Klaubarbeiter, пребирач.

Erzgang; рудна жица, рудни пруд, рудна пруда, рудна матица (у Лесковачком).

Feder; пругло (у Пожарев. и Крушевачком), пружина, опруга.

Weißbleierz; церусит, бистра (у Подрињу) в. Cerussit.



NETIPIČNI LEKSIČKI ZAPISI

Göpel (Göppel, Gapel); витао, чекрк, мотовило (у ширем смислу); то је увек построј за извлачење а имено: **Handgöpel** (Haspel, Handhaspel); ручни витао, в. Haspel.
Pferdegöpel (Göpel insbesondere); коњски витао.
Wasserradgöpel (Förderungs-Wasserrad) витао са воденим точком.
Turbinengöpel (Förderungs-Turbine) турбински витао.
Wassersäulengöpel (Förderungs-Wassersäulenmaschine); витао са воденом стублином.
Dampföpel (Förderungs-Dampfmaschine); парни витао.
Göpel; по новој немачкој терминологији разуме се: коњски витао.

Kohlenaufbereitung; пребиралиште угља, пребор угља:

a) пребирање при руднику (Aufbereitung bei der Grube) има задатак одељивања угља по величини

зрна (Klassifikation nach der Korngröße) и избацивања камена нарочито из грубљег комађа; затим долази према потреби;

b) преправа механичка или препирање угља (Mechanische Aufbereitung oder Kohlenwäsche), која се врши на мокром. путу, помоћу воде, те да се на тај начин одстрани страни делови у угљу; в. Separation; Kohlenseparation.

Примедба. Пребором на руднику(a) или класирањем (Klassierung) при коме се већ и у руднику угљено комађе ручним пребором одваја, издваја се из извађеног угља (Förderkohle) прорештавањем: осредњи угаљ (Mittelkohle); груменасти угаљ (Würfelkohle); орашастасти угаљ (Nußkohle); крупастасти или зрнасти угаљ (Kohlengries, Grieskohle); угљени прах (Kohlenstaub, Lösche). Понекад се одељује само на: груби угаљ или угљено комађе, осредњи угаљ и ситан угаљ (Kleinkohle, Grubeklein), — сем угљене прашине. Пребор на руднику дакле није ништа друго него одељивање врсте (класе), — нем. Klassierung; али се често, погрешно употребљују речи „Sortierung“ или „Separation“ који изрази одговарају једино механичкој преправи угља, то јест преправи на мокром путу. Види под b.

- o-ce-erovan pdf
- Abby Fine Reader

Streckenbau, Örterbau; добивање поткопима, ме-стимично преривање. Abbau mit breitem Blick, преривање широким прочељем, -ширимице. 6 Stoß bau, дуварско вађење, н. пр. прочељем, -зачељем по протез / у слеме, по паду -падомично. Firstenbau; преривање узвишно, - узвишице на ступњеве. Sohlenbau (Strossenbau); преривање падимице, подношке на ступњеве. Pfeilerbau; преривање на стубове, стубимично преривање. Strebbau (Diagonal-Stoßbau); дуварско ускосно преривање, косимично преривање. Schachbrettartiger Abbau; шахасто прери-вање, вреривање на прескоке. Etagenbau; преривање на подстате, катове, лесе. Q u erbau; преривање попречице, попречно. Ulmstrassenbau; преривање побочним прерубом. Kammerbau; преривање на избе. Stockwerksbau; преривање на катове. Abbauen; добивати, вадити, преривати, прерубљивати, прекопавати. Abbauen feldwärts; преривати у поље, прер, од себе; heimwärts abbauen преривати к себи. Abbaufeld (Baufeld); преровно поље, преровиште, прери-валиште. Abbaugrenze (Baugrenze) преровна међа; -граница. Abbauhöhe (Bauhöhe, Höhe des Abbaufeldes, der Bau abteilung); висина прерова. Abbauhorizont; (Bauhorizont, Bausohle) хоризонт прерова, кат прерова. Abbaukosten; трошкови преривања. Abbaulänge; дужина прерова. Abbaumethode (Baumethode); в. Abbauart.

DICTIONARY SEGMENTATION MODEL

...aining.dictionarySegmentation.tei.xml [C:\Users\Biljana\Desktop\toyData\chunk004.training.dictiona...

Project Options Tools Document Window Help

ecute XPath on 'Current File'

...aining.dictionarySegmentation.tei.xml x ...aining.dictionarySegmentation.tei.xml x

tei text body

lb lb Auszimmerung (Rüstung); подграда дрветом.

lb lb Auszug-Schacht-Stollen; в. Wetterschacht, Wetterstollen.

lb lb Autochthon; на месту поставше в. Bodeneigen.

lb lb body headnote 21 headnote body Axt (Zimmeraxt);

секира, брадва.

lb lb Azimut; азимут, зреник.

lb lb Azimutalkreis; азимутски круг.

lb lb Azimutalwinkel; азимутски угао.

lb lb body dietScrap , в.

lb lb dietScrap body Backe; чељуст, шкрипац; в.

Bremsbacke.

lb lb Backen (Packen), Oberbacken, Unterbacken, в. Bank.

lb lb Backenquetsche; чељусни, шкрипасти разбијач.

кљечкалица.

lb lb Backkohle, backende Kohle; мрсни угаљ, угаљ пржуљ.

lb lb препек.

lb lb Backkoks; мрсни кокс.

lb lb Baggern; изривати, багеровати.

lb lb Baggermaschine; изривача, ривача, багер.

lb lb Bahn, Eisenbahn, Förderbahn; пут, цеста, железница, до-

правница (н. пр. жична).

lb lb Bahn v. Schram; bahnen v. Schrämen.

lb lb Bahn (Hammerbahn, Fäustelbahn. .); затиљак на чекињу.

lb lb тупар, челце на бургији у које се удара чекињем.

lb lb Bahns (Schlechte, Schnitt); разметак претрг, пресек

Text Grid Author

U+002E ① Days left for evaluation: 115 31 new me.

chunk004.pdf

file:///C:/Users/Biljana/Desktop/lexmc2018/Chunks/r

2 of 5

21

Axt (Zimmeraxt); секира, брадва.
Azimut; азимут, зреник.
Azimutalkreis; азимутски круг.
Azimutalwinkel; азимутски угао.

B.

Backe; чељуст, шкрипац; в. Bremsbacke.
Backen (Packen), Oberbacken, Unterbacken, в. Bank.
Backenquetsche; чељусни, шкрипасти разбијач, кљечкалица.
Backkohle, backende Kohle; мрсни угаљ, угаљ пржуљ, препек.
Backkoks; мрсни кокс.
Baggern; изривати, багеровати.
Baggermaschine; изривача, ривача, багер.
Bahn, Eisenbahn, Förderbahn; пут, цеста, железница, до-правница (н. пр. жична).
Bahn v. Schram; bahnen v. Schrämen.
Bahn (Hammerbahn, Fäustelbahn. .); затиљак на чекињу, тупар, челце на бургији у које се удара чекињем.
Bahne (Schlechte, Schnitt); разметак, претрг, пресек.
Bahnschiene v. Schiene.
Bahnseil (bei der Seilbahn) в. Führungsseil.
Bahnstrecke; колосечна галерија, ходник.
Bahnwechsel v. Wechsel.
Balancer; ојница, вагало, балансје.
Band; трака, пџс, појас.
Bandhacke (Gezimmerbeil); тесачка секира, брадва.
Bandseil (Flachseil); плонте уже, појасасто уже.
Bandseilstrang (Flachseilstrang); струка појасастог ужета.
Bank (Backen); спруд слоја, преслога, слог.
Oberbank (Oberbacken); спруд, горњи преслог.
Unterbank (Unterbacken) доњи преслог, спруд.
Bär (Bohrklotz b. Erdbohren); маљ.
Barfuß (Zimmerung); проста подграда без подножница (потпетица) и слеменаџа.

DICTIONARY SEGMENTATION MODEL

- Prepoznate i obeležene strane u rečniku
- Obeleženi brojevi strana
- Statistika

```
<fw type="header">6 </fw>
<ab type="page">Stoß bau, дуварско вађење, н. пр. прочељем, -зачељем по протез / у с
Fistenbau; преривање узвишно, -узвишце на ступњеве. Sohlenbau (Strossenbau); преривање падимиче
Pfeilerbau; преривање на стубове, стубимично преривање. Strebbau (Diagonal-Stoßbau); дуварско ус
преривање. Schachbrettartiger Abbau; шахасто прери-вање, вреривање на прескоке. Etagenbau; прери
лесе. Q u erbau; преривање попречице, попречно. Ulmstrassenbau; преривање побочним прерубом. Kam
Stockwerksbau; преривање на катове. Abbauen; добивати, вадити, преривати, прерубљивати, прекопава
преривати у поље, прер, од себе; heimwärts abbauen преривати к себи. Abbaufeld (Baufeld); преров
валиште. Abbaugrenze (Baugrenze) преровна међа; -граница. Abbauhöhe (Bauhöhe, Höhe des Abbaufelde
висина прерова. Abbauhorizont; (Bauhorizont, Bausohle) хоризонт прерова, кат прерова. Abbaukoster
Abbaulänge; дужина прерова. Abbaumethode (Baumethode); в. Abbauart. Abbaumittel (Baumittel); међ
Abbauort (Abbaustrasse); место прерова; в. Abbaustoß, Abbaustrecke. Abbaupfeiler; преровни стуб.
Bauhorizont); преровни спрат, кат, бој, подиште, запад Abbaustoß; преровно чело, прочеље прерова.
<fw type="header">7 </fw>
<ab type="page">Abbaustrasse; преровни подстат; преруба. Abbaustrecke (Vorrichtungsst
делна ходба, разгодница. Abbausystem (Bausystem); в. Abbauart. Abbauverlust; губитак при прерива
приносан преров, издашан преров. Abbauwürdigkeit (Bauwürdigkeit); приносност, издашмог. Abbauzir
подграда прерова. Abbiättern, в. Ablösen, Blättern; откорити. Abbinden ein Gezimmer (Schachtkran
припојити, везати. Abbohren Bohrlöcher; извртити, избушити рупу, -буше-тину. Äbbohren eine Gegenc
крај. Abbohren einen Schacht; бушити, дубити окно, јаму, спустити окно, -вртло. Abbremsen; кочити
наку терет. #bbohren (einen Sprengschuß, ein Bohrloch); опалчти, опуцати метак, -мину. Abbrüster
с-нути, запретати, загатити. Abdekarbeit в. Abraum; прер
en. Abdrücke в. Fossilien. Abend (Westen); запад. Abend
nd); преградити, за-градити; в. Abdämmen. Abfahren; отић
S; отпадак, ивер. Abfallen; падати, отицати (кад је реч
а, падни поткоп. </ab>
(Fallstück bei d. Erdbohrer) однаца. Abfangen; хватати,
nge b. Erdbohrer); прихватити, обе-сити бушаћи алат. Abfa
в. Abkehren. Abfluß; отицај (воде), одилазак, оток. Abfl
d; преривено поље, прекопано, прерађено поље. Abgelöste
```

==== Token-level results ====

label	accuracy	precision	recall	f1	
<body>	100	100	100	100	
<headnote>	100	100	100	100	
all fields	100	100	100	100	(micro average)
	100	100	100	100	(macro average)

DICTIONARY BODY SEGMENTATION MODEL

- Rezultat zadovoljavajući
- Prepoznati obeleženi leksički zapisi
- Statistika

```
<pb/>
<fw type="header">7</fw>
<entry>Abbaustrasse; преровни подстат; преруба</entry>
<pc>.</pc>
<entry>Abbaustrecke (Vorrichtungsstrecke); преривна ходба; по- делна ходба, разгодница</ent
<pc>.</pc>
<entry>Abbausystem (Bausystem); в. Abbauart</entry>
<pc>.</pc>
<entry>Abbauverlust; губитак при преривању</entry>
<pc>.</pc>
<entry>Abbauwürdig (bauwürdig); приносан преров, издашан преров</entry>
<pc>.</pc>
<entry>Abbauwürdigkeit (Bauwürdigkeit); приносност, издашмос</entry>
<pc>.</pc>
<entry>Abbauzimmerung, Verhauzimmerung; подграда прерова</entry>
<pc>.</pc>
<entry>Abbiättern, в. Ablösen, Blättern; открити</entry>
<pc>.</pc>
<entry>Abbinden ein Gezimmer (Schachtkranz, Türstock); спојити, припојити, везати</entry>
<pc>.</pc>
<entry>Abbohren Bohrlöcher; извртити, избушити рупу, -буше- тину</entry>
```

```
итати неки крај</entry>
убити окно, јаму, спустити окно, -вртло</entry>
у кочила некч терет</entry>
Bohrloch); опалчити, опуцати метак, -мину</entry>
и, заптити, загис- нути, запретати, загатити</entry>
```

==== Token-level results ====

label	accuracy	precision	recall	f1	
<dictScrap>	99.88	0	0	0	
<entry>	99.78	99.83	99.93	99.88	
<pc>	99.84	98.82	99.21	99.02	
all fields	99.83	99.75	99.75	99.75	(micro average)
	99.83	66.22	66.38	66.3	(macro average)

LEXICAL ENTRY MODEL

Options Tools Document Window Help

Current File

chunk001.training.LexicalEntry.tei.xml x ...g.dictionaryBodySegmentation.tei.xml* x chun

tei text body entry

lb lb зачељем по протез / у слеме, по паду -падомично re.

lb lb re Firstenbau; преривање узвишно, -узвишице на

lb lb ступњеве re.

lb lb re Sohlenbau (Strossenbau); преривање падимце,

lb lb подношке на ступњеве re.

lb lb re Pfeilerbau; преривање на стубове, стубимично

lb lb преривање re.

lb lb re Strebau (Diagonal-Stoßbau); дуварско ускосно

lb lb преривање, косимично преривање re.

lb lb re Schachbrettartiger Abbau; шахасто прери-

lb lb вање, вреривање на прескоке re.

lb lb re Etagenbau; преривање на подстате, катове, лесе re.

Qu erbau; преривање попречице, попречно.

lb lb Ulmstrassenbau; преривање побочним прерубом.

lb lb Kammerbau; преривање на избе.

lb lb Stockwerksbau; преривање на катове

file:///C:/Users/Biljana/Desktop/toyData/pdfDictionary

Abbau mit breitem Blick, преривање широким прочељем, — ширинице.

6

Stoßbau, дуварско вађење, н. пр. прочељем, — зачељем по протезу у слеме, по паду — падомично.

Firstenbau; преривање узвишно, — узвишице на ступњеве.

Sohlenbau (Strossenbau); преривање падимце, подношке на ступњеве.

Pfeilerbau; преривање на стубове, стубимично преривање.

Strebau (Diagonal-Stoßbau); дуварско ускосно преривање, косимично преривање.

Schachbrettartiger Abbau; шахасто преривање, преривање на прескоке.

Etagenbau; преривање на подстате, катове, лесе.

Querbau; преривање попречице, попречно.

Ulmstrassenbau; преривање побочним прерубом.

Kammerbau; преривање на избе.

Stockwerksbau; преривање на катове.

Abbauen; добивати, вадити, преривати, прерубљивати, прекопавати.

Abbauen feldwärts; добивати у поље, преривати од себе:

DICTIONARY BODY SEGMENTATION MODEL

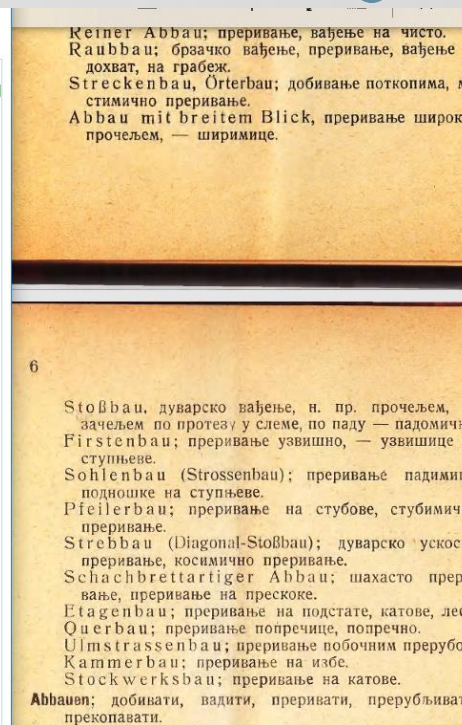
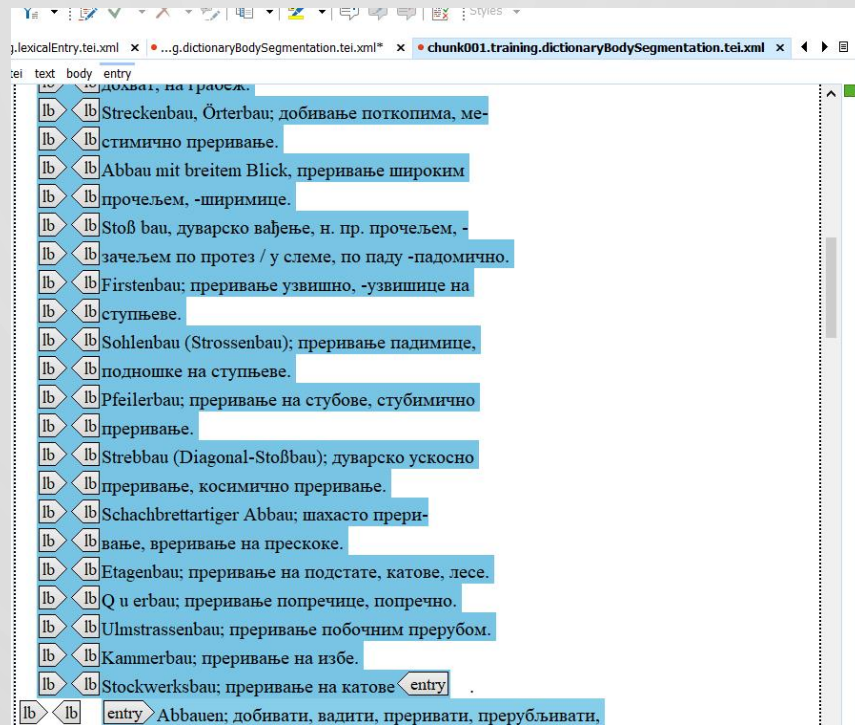
```
<entry>Abbau (Bau); преров, прерив, обров, преруба (рудне жице или слоја), такође вађење, оброњавање, оба- рање, добивање</entry>
<pc>.</pc>
<entry>Abbauabteilung в. Bauabteilung</entry>
<pc>.</pc>
<entry>Abbauart, Abbaumethode (Bauart, Baumethode); начин пре- ривања (оброва и т. д), прерова</entry>
<pc>.</pc>
<entry>Abbauarten (Bauarten); начин преривања: Tagabraum (Abraum, Tagbau, Abdecarbeit, Auf deckarbeit); вађење, добивање обданично, -на обданици, обданични рад. Schwebender Bau (schwebender Abbau); сле- мено вађење, преров у слеме. Diagonaler Bau (diagon. Abbau) косимично Ba uerte, преров искосирке. Streichender Bau (streich. Abbau) преривање по протезању (полошке). Tonnlägiger Bau (tonnlägiger Abbau, verfla chender Abbau); вађење падомице, по паду, падо- мични преров. Versatzbau (Abbau mit Bergversatz); запунско, засипно вађење, -преривање. Bruchbau; преривање на поломе, на претрп. Reiner Abbau; преривање, вађење на чисто. Raubbau; брзачко вађење, преривање, вађење на дохват, на грабеж. Streckenbau, Örterbau; добивање поткопима, ме- стимично преривање. Abbau mit breitem Blick, преривање широким прочељем, -ширимице.
<dictScrap>2*</dictScrap>
<pb/>
<fw type="header">6</fw>Stoß bau, дуварско вађење, н. пр. прочељем, - зачељем по протез / у слеме, по паду - падомично. Firstenbau; преривање узвишно, -узвишице на ступење. Sohlenbau (Strossenbau); преривање падимице, подношке на ступење. Pfeilerbau; преривање на стубове, стубимично преривање. Strebbau (Diagonal-Stoßbau); дуварско ускосно преривање, косимично преривање. Schachbrettartiger Abbau; шахасто прери- вање, вреривање на прескоке. Etagenbau; преривање на подстате, катове, лесе
</entry>
<pc>.</pc>
<entry>Q u erbau; преривање попречице, попречно. Ulmstrassenbau; преривање побочним прерубом. Kammerbau; преривање на избе. Stockwerksbau; преривање на катове</entry>
<pc>.</pc>
<entry>Abbauen; добивати, вадити, преривати, прерубљивати, прекопавати</entry>
<pc>.</pc>
<entry>Abbauen feldwärts; преривати у поље, прер, од себе; heimwärts abbauen преривати к себи</entry>
<pc>.</pc>
<entry>Abbaufeld (Baufeld); преровно поље, преровиште, прери- валиште</entry>
<pc>.</pc>
<entry>Abbaugrenze (Baugrenze) преровна међа; -граница</entry>
<pc>.</pc>
```

IAKO...

je ovaj primer
ispravno anotiran
za potrebe
treninga 2. modela

- Objašnjenje: u
rečniku postoji
jako malo
primera koji
odstupaju od
obrasca

- Boja papira?



DICTIONARY BODY SEGMENTATION MODEL

- Rezultat zadovoljavajući

- Prepoznati obzapi

- Statistika

==== Token-level results ====

label	accuracy	precision	recall
<dictScrap>	99.88	0	0
<entry>	99.78	99.83	99.83
<pc>	99.84	98.82	98.82
all fields	99.83	99.75	99.75
	99.83	66.22	66.38
			99.75
			66.3

```
<entry>Abbau (Bau); преров, прерив, обров, преруба (рудне жице или слоја), такође вађење, оброчавање, оба- рање, добивање</entry>
<pc>.</pc>
<entry>Abbaubteilung в. Bauabteilung</entry>
<pc>.</pc>
<entry>Abbaumethode (Bauart, Baumethode); начин прери- вања (оброва и т. д), прерова</entry>
<pc>.</pc>
<entry>Abbauarten (Bauarten); начин прери- вања: Tagabraun (Abraum, Tagbau, Abdekarbeit, Auf deckarbeit); вађење, добивање обданично, -на обданици, обданични рад. Schwebender Bau (schwebender Abbau); сле- мено вађење, преров у слеме. Diagonaler Bau (diagon. Abbau) косичично Ba uerte, преров ископирке. Streichender Bau (streich. Abbau) прери- вање по протезану (полошке). Tonnlägeriger Bau (tonnlägeriger Abbau, verfla chender Abbau); вађење падомике, по паду, падо- мични преров. Versatzbau (Abbau mit Bergversatz); запунско, засипно вађење, -прери- вање. Bruchbau; прери- вање на поломе, на претрп. Reiner Abbau; прери- вање, вађење на чисто. Raubbau; брзачко вађење, прери- вање, вађење на дохват, на грабеж. Streckenbau, Ortenbau; добивање поткопина, ме- стично прери- вање. Abbau mit breitem Blick, прери- вање широким прочељем, -ширици.
<dictScrap>2*</dictScrap>
<pb/>
<fw type="header">6</fw>Stoß bau, дуварско вађење, н. пр. прочељем, - зачељем по протез / у слеме, по паду - падомично. Firstenbau; прери- вање узвишно, -узвишице на ступље. Sohlenbau (Strossenbau); прери- вање падмиче, подношке на ступље. Pfeilerbau; прери- вање на стубове, стубично прери- вање. Strebbau (Diagonal-Stoßbau); дуварско ускосно прери- вање, косичично прери- вање. Schachbrettartiger Abbau; шахасто прери- вање, врери- вање на прескоке. Etagenbau; прери- вање на подстате, катове, лесе
</entry>
<pc>.</pc>
<entry>Q u erbau; прери- вање попречице, попречно. Ulstrassenbau; прери- вање побочним прерубом. Kammerbau; прери- вање на избе. Stockwerksbau; прери- вање на катове</entry>
<pc>.</pc>
<entry>Abbauen; добивати, вадити, прери- вати, прери- вати, прери- вати, прери- вати</entry>
<pc>.</pc>
<entry>Abbauen feldwärts; прери- вати у поље, прер, од себе; heimwärts abbauen прери- вати к себи</entry>
<pc>.</pc>
<entry>Abbaufeld (Baufeld); преровно поље, прери- вати, прери- вати</entry>
<pc>.</pc>
<entry>Abbaugrenze (Baugrenze) прерова међа; -граница</entry>
<pc>.</pc>
```

```
<pb/>
<fw type="header">7</fw>
<entry>Abbaustrasse; преровни подстат; прери- вање</entry>
```

strecke); прери- вати ходба; по- делна ходба, разгодница</ent

. Abbauart</entry>

прери- вање</entry>

риносан преров, издашан преров</entry>

keit); приносност, издашмост</entry>

erung; подграда прерова</entry>

ättern; открити</entry>

achtkranz, Türstock); спојити, припојити, везати</entry>

ити, избушити рупу, -буше- тину</entry>

њем испитати неки крај</entry>

шити, дубити окно, јаму, спустити окно, -вртло</entry>

и помоћу кочица некч терет</entry>

uß, ein Bohrloch); опалчти, опуцати метак, -мину</entry>

```
<pc>.</pc>
<entry>Abbrüsten в. Anbrüsten</entry>
<pc>.</pc>
<entry>Abdämmen (Wasser, Feuer); запунити, заптити, загис- нути, запрети, загатити</entry>
<pc>.</pc>
```

SSK - STANDARDIZATION SURVIVAL KIT

- <http://ssk.huma-num.fr/#/>



[HOME](#) [EXPLORE](#) [CONTRIBUTE](#) [LEARN MORE](#) [PARTHENOS](#)

[Log in](#) 

the editing environment

Setting up the publishing environment

Provide for long-term availability



#LEXMC

GROBID-dictionaries Gang... #lexmc 🏴‍☠️🏴‍☠️🏴‍☠️



Fraser Dallachy @FraserDallachy · Dec 7

Final day of [#lexmc](#) features participants presenting on their projects, including (people whose Twitter handles I have) [@mrhzc](#) on using GROBID and XSLT on Turkish dictionaries, [@knjigohranitelj](#) on a German-Serbian mining dictionary, and [@MartinJWynne](#) on French variants



1



5



Emrah Özcan @mrhzc · Dec 7

[@I_m_Grobid](#) [@Med_Khemakhem](#) [@mrhzc](#) [@nmickoski1](#) [@knjigohranitelj](#)
#MarijaGmitrović Proudly presenting the GROBID-dictionaries gang... #lexmc

LEXICAL DATA MASTERCLASS

- Blog radionice: <https://digilex.hypotheses.org/date/2018/12>



The image features a light gray background with a large, faint, light blue circle in the center. In the four corners, there are decorative elements consisting of thin blue lines and small circles, resembling circuit traces or stylized trees. The text "Hvala na pažnji!" is centered in a bold, dark blue font.

Hvala na pažnji!