

Označevanje korpusov z orodjem WebAnno

JOTA, MPŠ IJS, Ljubljana, 13. 11. 2015

kaja.dobrovoljc@trojina.si



CLARIN.SI



Oris delavnice

▪ **predavanje** (15:00 – 15:45)

- glavne značilnosti
- primeri uporabe
- dostop do orodja

▪ **prikaz** (15:45 – 16:30)

- načini označevanja
- pregledovanje
- priprava projekta

▪ **vaje** (16:45 – 17:30)

- označevanje
- pregledovanje
- slovenski tviti

Označevanje korpusov

▪ analitični vidik

- analiza v kontekstu
- kvantitativna analiza
- empirična validacija teoretičnih izhodišč
- povratna informacija glede težavnih mest kategorizacije

▪ tehnološki vidik

- razvoj označevalnih aplikacij
- učna množica za strojne označevalnike
- samodejna izdelava leksikonov

▪ praktični vidik

- lažje iskanje in priklic
- hitrejša vizualizacija
- dolgoročna uporabnost

WebAnno

- <https://webanno.github.io/>
- večnamensko spletno orodje za (jezikoslovno) označevanje korpusov
- razvoj v okviru CLARIN-D (F-AG 7):
 - 2012-2013: Implementation of a web-based annotation platform
 - 2012-2013: Development of guidelines and best practices for annotation of non-standard varieties of German
 - 2014-2016: Semantic Annotation for Digital Humanities (**WebAnno 3**)



Glavne značilnosti

▪ Prilagodljivost

- večinivojsko označevanje
- poljuben nabor oznak
- različni načini označevanja
- označevanje (ne)označenih besedil

▪ Skupinsko delo

- en ali več označevalcev istega korpusa / besedila
- delitev vlog in pooblastil

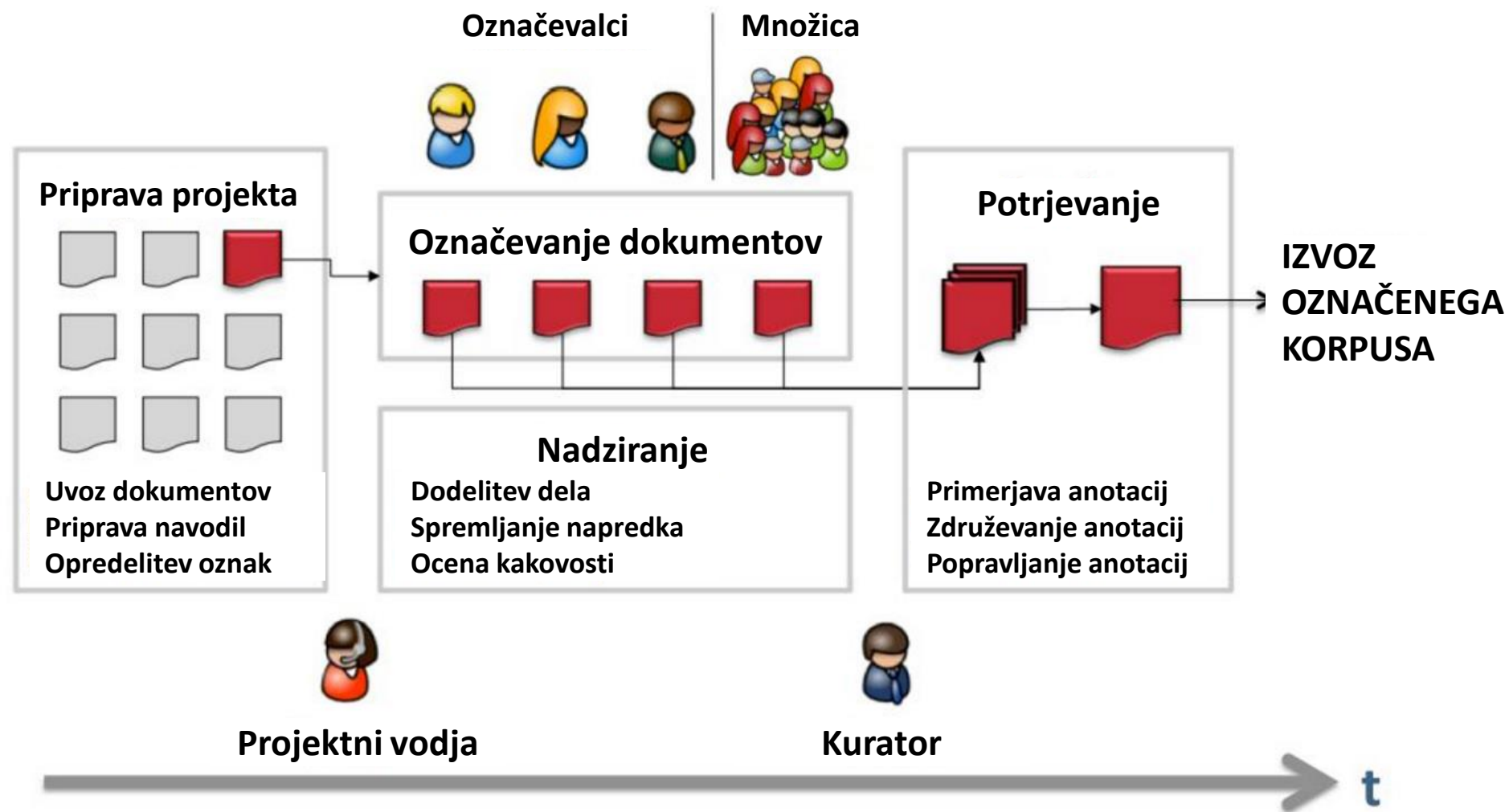
▪ Spletni dostop

- preprost dostop, brez namestitev programske opreme
- sprotno shranjevanje

▪ Odprta koda

- skupinsko delo
- dobra podpora in odzivnost

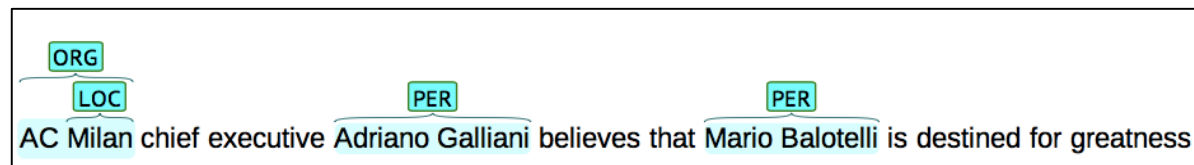
Delotok označevanja



Tipi označevanih pojavov (layers)

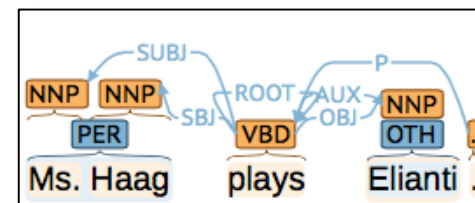
■ nizi (*span*)

- označevanje
- znakov, besed ali besednih zvez
- npr.: morfemska analiza, oblikoslovne lastnosti, frazeologija, imena, časovni izrazi, zaznamovanost besedišča, prevodne ustreznice itd.



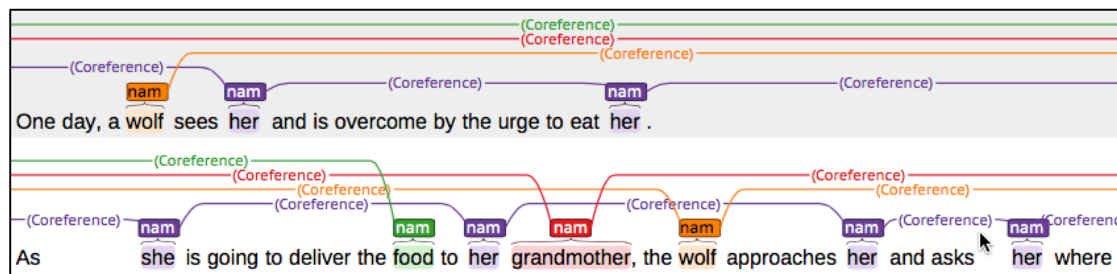
■ relacije (*relation*)

- povezovanje
- npr.: odvisnostna skladnja, glagolska vezljivost, semantična razmerja



■ verige (*chain*)

- nizanje
- npr.: koreferenčni nizi



Primer: imenske entitete

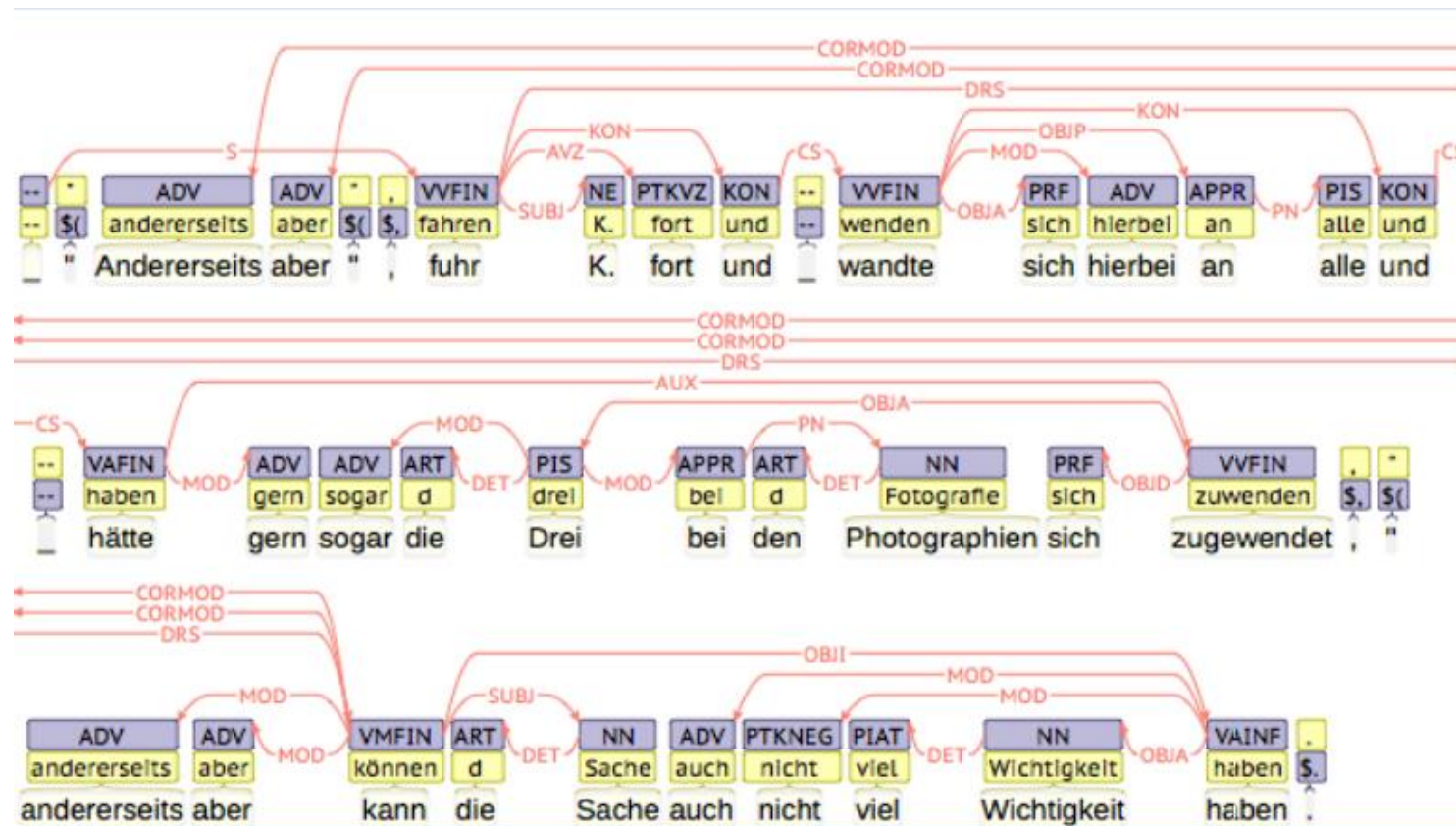
showing 6-15 of 500 sentences

Annotation

		LOC
		LOCderiv
6	Am Eingang ins Areal an der Freisinger Straße soll ein etwa 15 Meter hoher Neubau mit drei Vollgeschossen auf einem fünf Meter hohen Grundstück entstehen.	
7	Diese Struktur ermöglicht ein Verfeinern des reversen Adressraums in mehreren Schritten .	
8	Der Taifun schwächte sich bei der Annäherung auf das chinesische Festland ab ; die JMA stuft Nuri aber erst nach einem direkten tropischen Sturm zurück .	LOCderiv ORG OTH
9	Bundespräsident Horst Köhler würdigte den Verstorbenen als " vorbildlichen Diener seiner Kirche und Christen von einzigartiger Glaubwürdigkeit .	PER
10	Es ermöglicht das Sammeln und Einlösen von Bonuspunkten .	
11	Unterschiedliche Orientierungen der Polarisationsfilter in den Brillengläsern sorgen dafür , dass jedes Auge nur sein Bild sieht und das andere das Gegenbild .	
12	Da kamen beim Scheunenfest boshafte Charakterzüge ans Licht , wo man sie am wenigsten erwartet hätte .	
13	Georg Fischer wird in Kunshan , China , eine Eisengiesserei errichten .	PER LOC LOC
14	Als gesichert gilt wenigstens , dass sie im Consort gespielt Otterstedt , S. 121 bzw. um die Wende zum 17. Jahrhundert .	PER
15	Vor allem der bislang als beinahe sicherer Nachfolger geltende Schatzkanzler Gordon Brown gerät zunehmend in Verruf .	PER

Benikova et al. (2014): NoSta-D **Named Entity Annotation for German: Guidelines and Dataset**. Proceedings LREC 2014, Reykjavik, Iceland

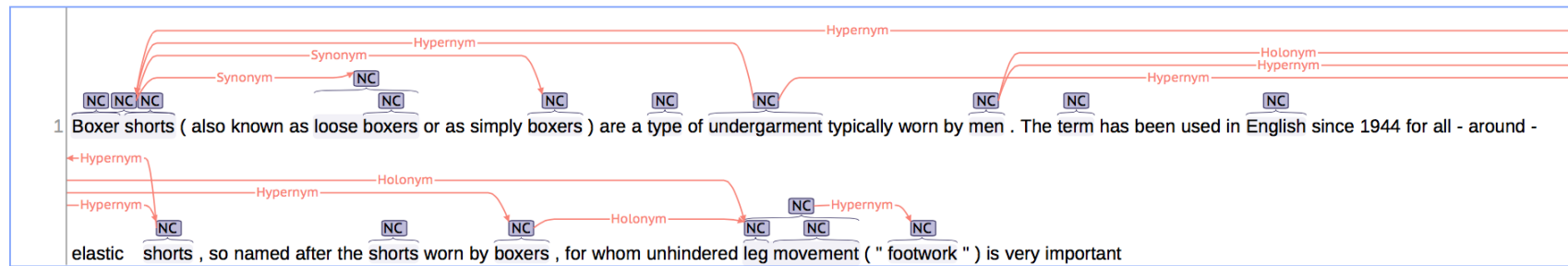
Primer: nestandardna nemščina



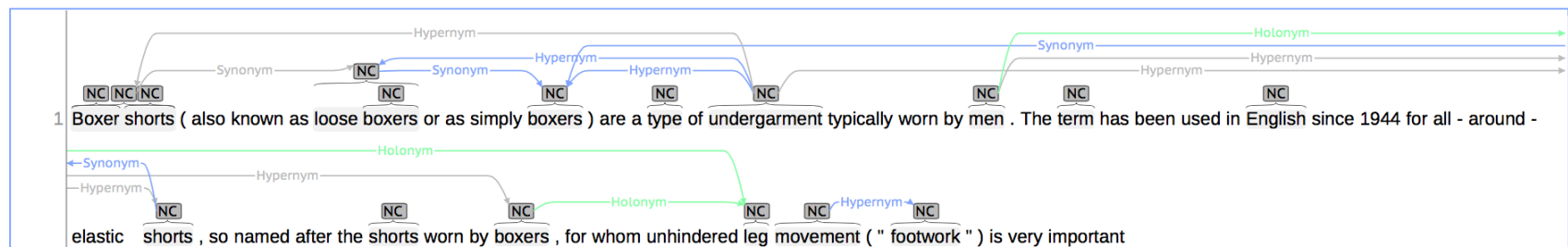
Dipper et al. NoSta-D: A corpus of German non-standard varieties. Non-Standard Data Sources in Corpus-Based Research (2013): 69-76.

Primer: semantična razmerja

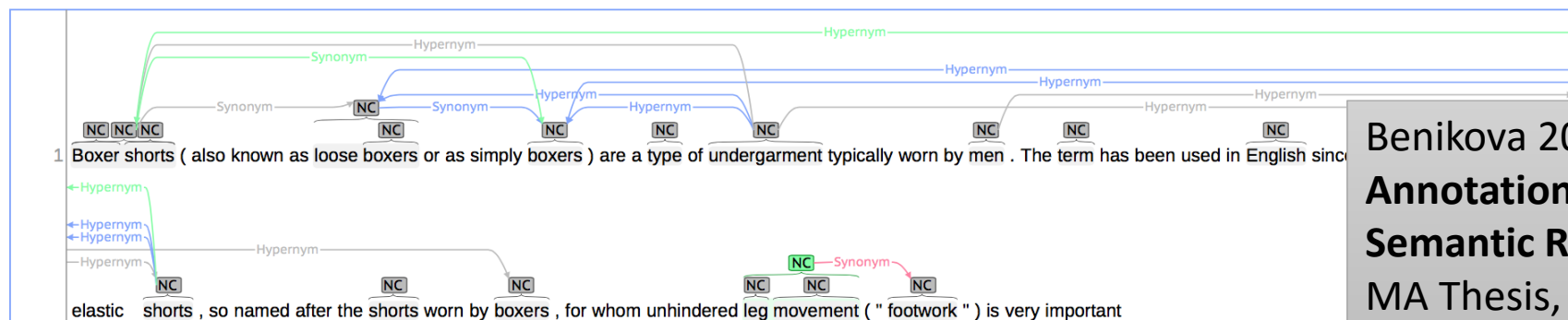
Annotation



User: semRel2



User: semRel3



Benikova 2015: Multilingual Contextual Annotation and Analysis of Classical Semantic Relations between Nominals. MA Thesis, Ling/LT, TU Darmstadt

Primer: razčlenjevanje oglasov za delo

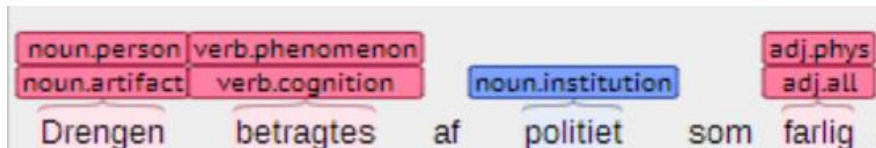
showing 2-2 of 220 sentences

company	ad_title
2 Domsel - Consulting - Stellenangebote	Software Entwickler VB (m / w) Mit neuen Ideen zum Ziel Als Personalberatung finden und vermitteln wir seit mehr als 15 Jahren
	Mitarbeiter in Festanstellung für unsere Kunden aus den Branchen Informationstechnologie und Vertrieb . Für unsere Kunden suchen wir talent und talent engagierte und qualifizierte Mitarbeiter in deren Fokus Sicherheit , Perspektiven und Innovation stehen . Unser Auftraggeber ist ein international tätiges Beratungsunternehmen , das vor allem in den Bereichen Unternehmensführung und Logistik Maßstäbe setzt . Als Kompetenzunternehmen stellt es eine Schnittstelle zwischen öffentlicher Hand und Markt dar und unterstützt seine Kunden im Vergabe - und Vertragsmanagement und beim Beschaffungsprozess von Leistungen und Waren . Durch optimale Ressourcenplanung und Geschäftsprozessoptimierung ist unser Auftraggeber ein Garant für den betrieblichen Erfolg seiner Kunden . Für die weitere Expansion des Unternehmens suchen wir
start_date	location_other_region ad_title
zum nächstmöglichen Zeitpunkt im	Rhein - Neckar - Raum eine / n Software Entwickler (m / w) . Ihr neues Aufgabengebiet sieht wie folgt aus :
task	task
Entwicklung einer Rich Internet Application mit Datenbankanbindung	Entwicklung von Desktop - und C / S - Anwendungen
task	skill_name skill_name skill_name skill_name
Entwicklung individueller Anwendungen (Point Solutions)	Programmiersprachen : Visual Basic , Flex3 / Action script 3 , T - SQL
task	necessity education_level
Eigenständiges Arbeiten in einem internationalen Team mit Anwendungsberatern und Methodenspezialisten	Idealerweise verfügen Sie über : eine qualifizierte Ausbildung im
education education_level level skill_name skill_name level skill_name skill_name skill_name	IT - Bereich oder eine technische Ausbildung sicherer Umgang mit Visual Basic , MS Office gute Kenntnisse in MS VisualStudio (VB) , ASP.NET und
skill_name level skill_name necessity	MS SQL Server 2005 Praxiserfahrungen mit Adobe Flex3 sind von Vorteil Unser Auftraggeber bietet Ihnen : Eine abwechslungsreiche und anspruchsvolle Tätigkeit in
position_type	talent
Festanstellung in einem interdisziplinären und internationalen Team mit Anwendungsberatern und Methodenspezialisten . Neben einer guten Dotierung bietet Ihnen unser	
position_type	talent talent
Auftraggeber außerdem ein flexibles Arbeitszeitmodell , in dem Sie sich mit neuen Ideen eigenständig und lösungsorientiert einbringen können . Ihr A	
location_city talent	zentraler Lage direkt am Hauptbahnhof Mannheim . Wenn Sie eine Leidenschaft für IT - Themen haben , dann freuen wir uns auf Ihre vollständigen Be
	Mail unter Angabe Ihrer Gehaltsvorstellung und Ihres möglichen Eintrittstermins an unsere E - Mail Adresse oder über unsere Homepage . Für Fragen
contact_person_name contact_tel	Aufgabe steht Ihnen Herr Domsel gerne unter der Telefonnummer 06241-985211 jederzeit zur Verfügung . Wir freuen uns auf Ihre Anfragen . Do
contact_tel contact_url	Flatenstraße 11 – 68623 Lampertheim – 06241-985-211 – www.domsel.de Bewerben Sie sich hier Vorname : Nachname : E - Mail : Laden Sie hier Ih
	Bewerbungsanschreiben

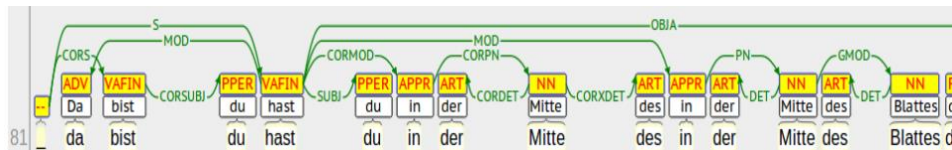
Tobias Krönke (2013): Job Ad Parsing for Applicant Qualification Prediction, MA Thesis, LT, TU Darmstadt, Sept. 2013

Še nekaj primerov

- **pomensko** označevanje glagolov in samostalnikov (danščina, Pedersen et al. 2014)



- različne ravni označevanja **govorjenih besedil** (nemščina, Lüdeling 2013)



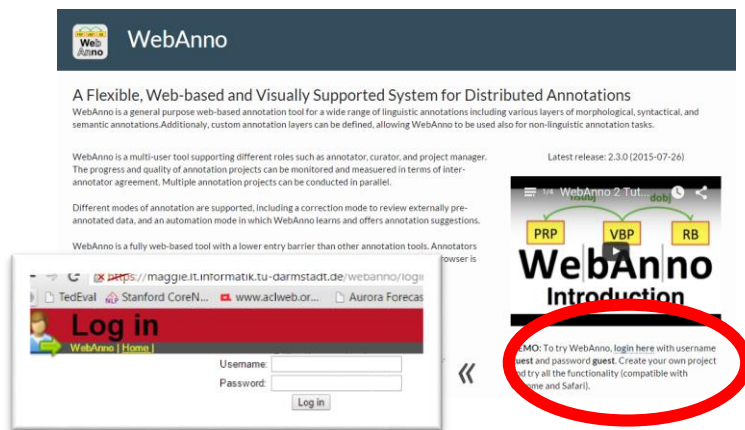
- **stalne besedne zveze** v spletni tehnični pomoči (portugalščina, Querido et al. 2015)
- **govorna dejanja** v pogovorih na Twitterju (nemščina, Zarisheva in Scheffer 2015)
- **biomedicinski izrazi** (angleščina, Yimam et al. 2015)

Kako dostopamo

- brskalnik Chrome ali Safari
- 3 prostodostopne "različice"

DEMO

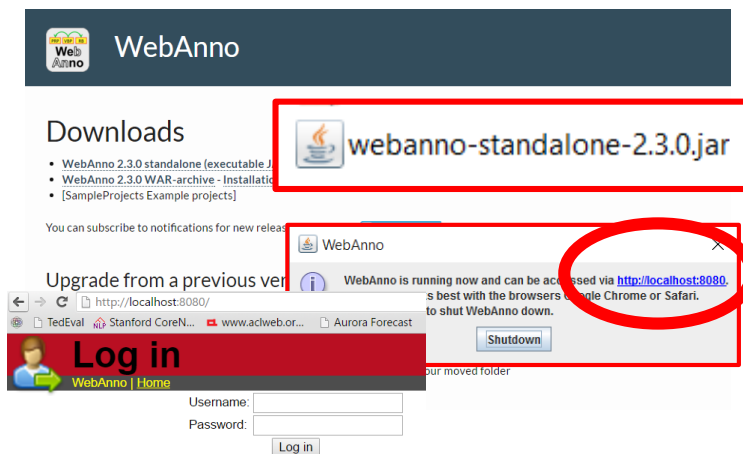
(nemška inštalacija)
guest / guest



preizkušanje, novosti

STANDALONE

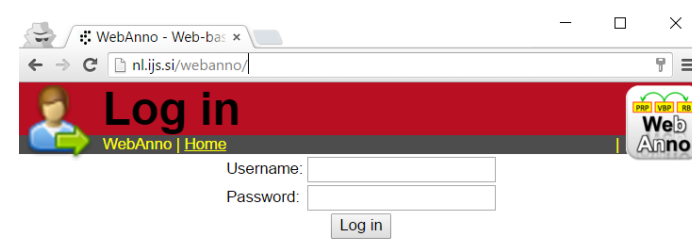
(samostojna aplikacija)
admin / admin



samostojno označevanje

CLARIN.SI

<http://nl.ijs.si/webanno/>
info@clarin.si



samostojno označevanje in
skupinski projekti z več
označevalci,
tehnična podpora

CLARIN.SI

DEMO

Glavni meni



[Annotation](#)

- [način 1] Označevanje neoznačenega besedila ali pregledovanje predhodno označenega besedila



[Curation](#)

- Primerjava in usklajevanje anotacij več označevalcev



[Correction](#)

- [način 2] Potrjevanje in popravljanje predhodno označenega besedila



[Automation](#)

- [način 3] Označevanje s sprotnim učenjem



[Projects](#)

- Priprava projekta



[Monitoring](#)

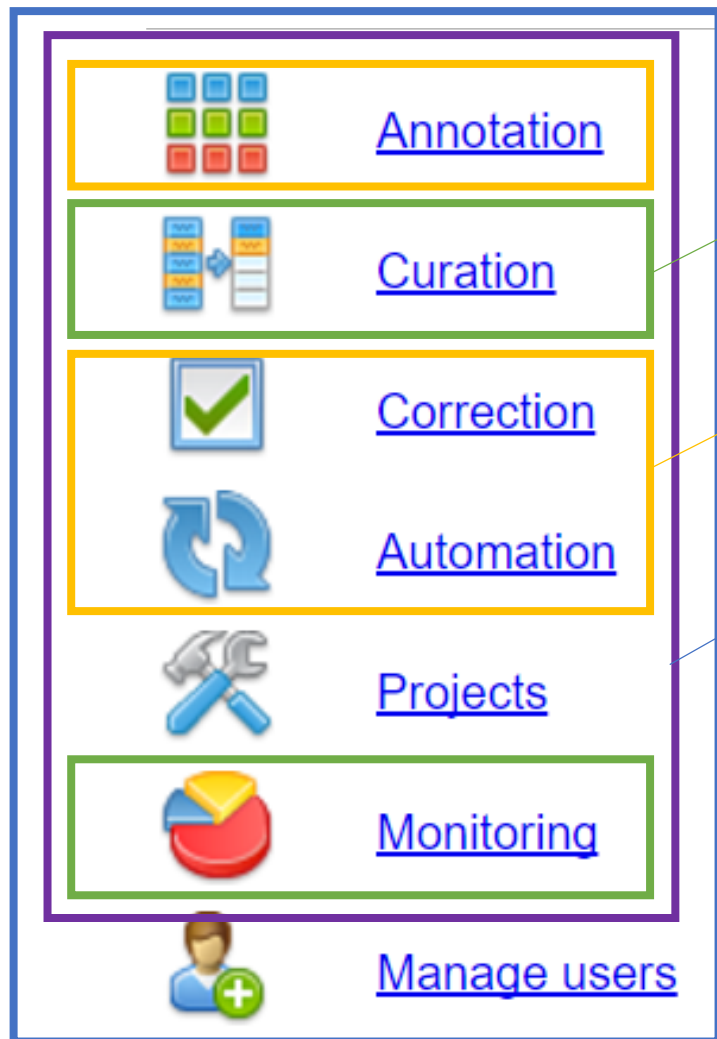
- Delitev dela in spremljanje napredka



[Manage users](#)

- Dodajanje uporabnikov

Glavni meni



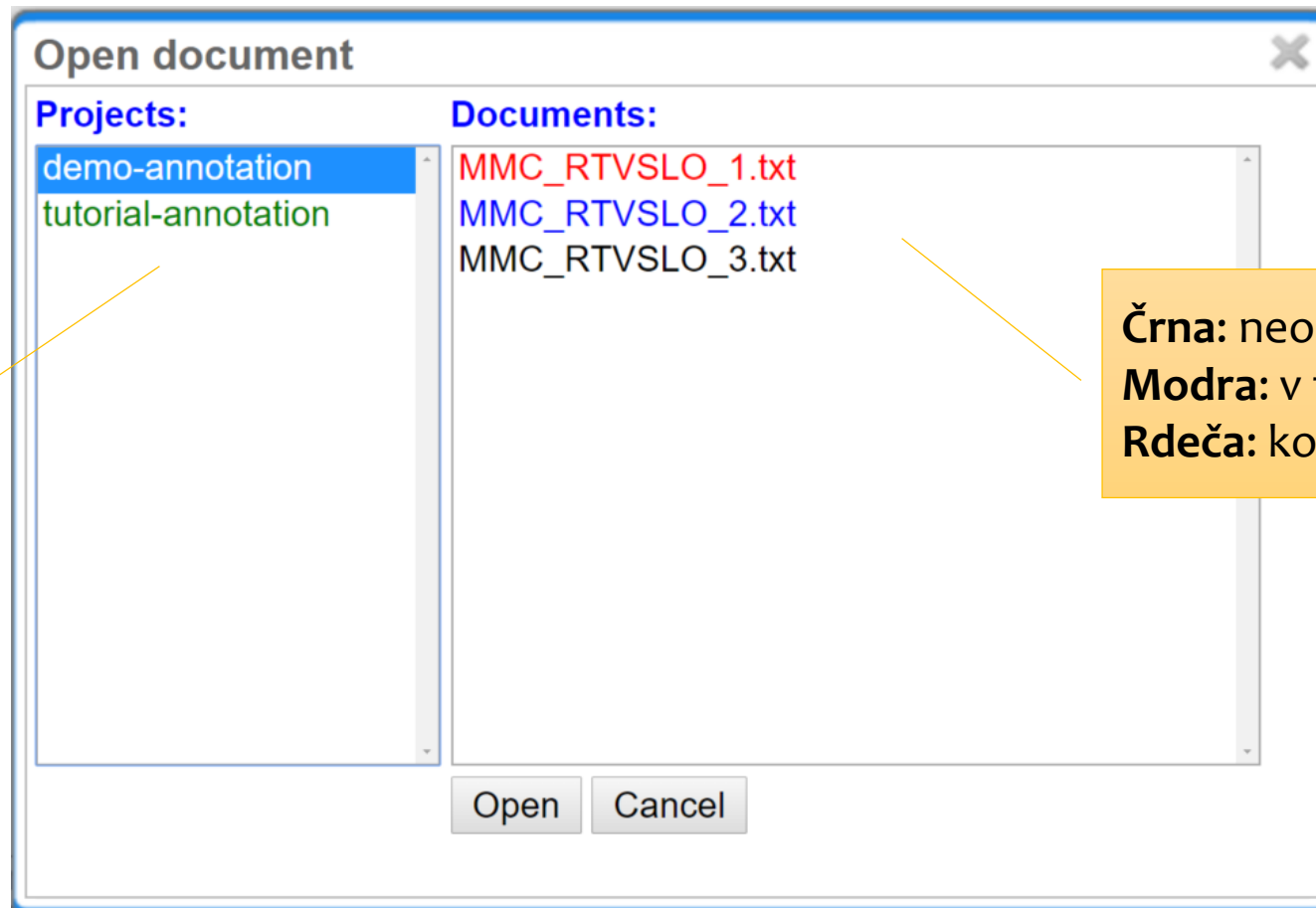
KURATOR

OZNAČEVALEC

VODJA PROJEKTA

ADMINISTRATOR

Odpiranje dokumenta



Izbira projekta in dokumenta za označevanje.

Črna: neodprto
Modra: v teku
Rdeča: končano

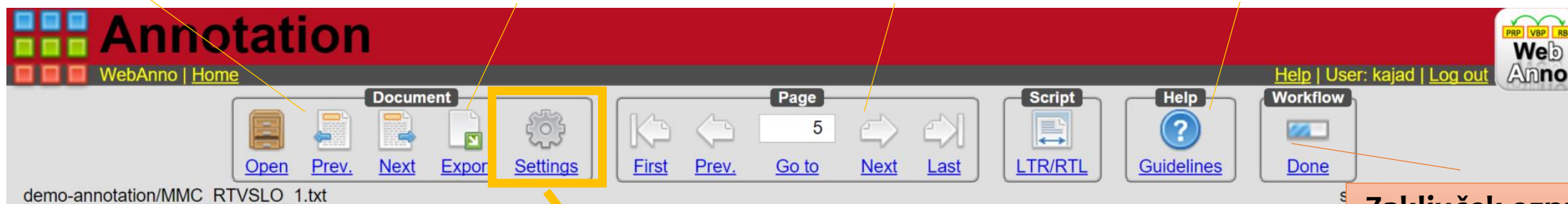
Navigacija

Premikanje med dokumenti.

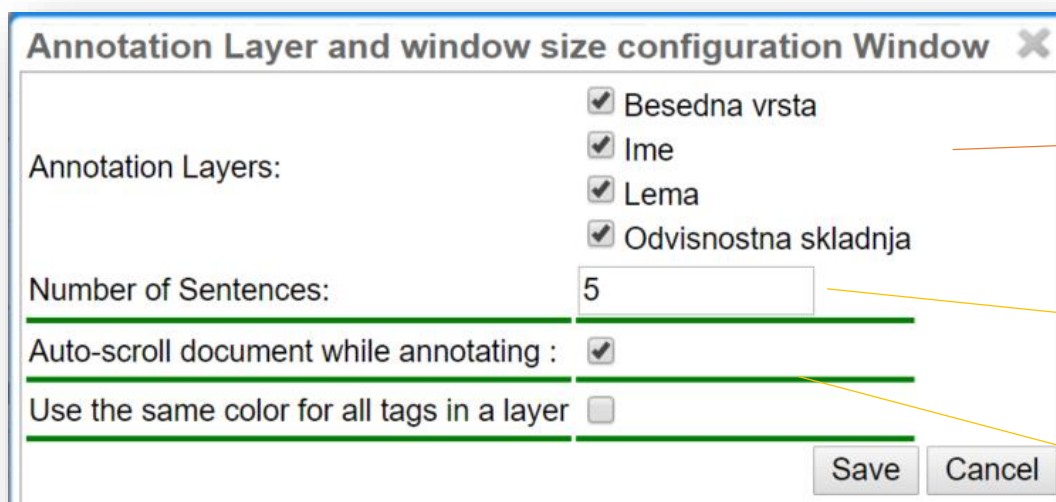
Izvoz trenutno označenega dokumenta

Premikanje po dokumentu

Ogled navodil za označevalce



Zaključek označevanja dokumenta



Vklapljanje in izklapljanje nivojev označevanja

Poljubna nastavitev števila naenkrat prikazanih povedi

Nastavitev možnosti samodejnega premikanja in barvanja oznak



Klasično označevanje (Annotation)

V besedilu kliknemo ali označimo niz (morfem, besedo ali besedno zvezo), ki mu želimo pripisati oznako.



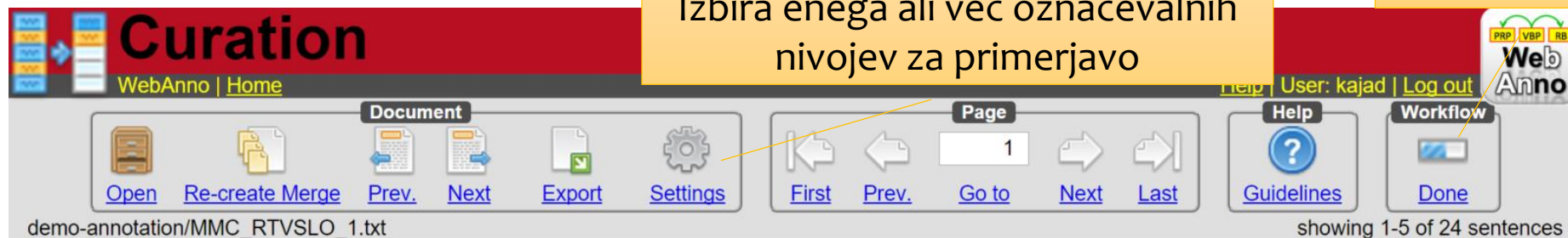
Izbira nivoja označevanja

Izbira ali vpis oznake

Potrditev oznake (tudi ENTER).

The 'New Span Annotation' dialog box is shown. It has a close button (X) in the top right corner. The 'Selected text' field contains 'Glucksmann'. The 'Layer' dropdown menu is set to 'Ime'. The 'Features' section shows a list of possible annotations: 'osebno', 'organizacija', 'osebno', 'stvarno', and 'zemljepisno'. The 'Annotate' button is at the bottom left of the dialog box.

Pregledovanje (Curation)



Izbira enega ali več označevalnih nivojev za primerjavo

Zaključek pregledovanja dokumenta

Sentences

- 1 V 79. letu starosti je umrl francoski filozof Andre Glucksmann.
Eden najvidnejših sodobnih francoskih filozofov je bil v mladosti goreč marksist in maoistični levičar, pozneje pa se je v svojih polemičnih zapisih začel bojevati proti vsakršni obliki diktature in totalitarizma.
- 2 Francoski predsednik Francois Hollande se je Glucksmannu poklonil kot zagovorniku zatiranih.
- 3 "Na sramotilnem stebru je izpostavljal tiste, ki so povzročali trpljenje," je tvitnil Hollande.
- 4 Glucksmann je že v svoji prvi knjigi Discours de la guerre (Razprava o vojni) leta 1968 na muho vzel vojno in politiko jedrskega zastraševanja.

Annotation

osebno
5 Glucksmann je že v svoji prvi knjigi Discours de la guerre (Razprava o vojni) leta 1968 na muho vzel vojno in politiko jedrskega zastraševanja.

User: anno1

osebno osebno
5 Glucksmann je že v svoji prvi knjigi Discours de la guerre (Razprava o vojni) leta 1968 na muho vzel vojno in politiko jedrskega zastraševanja.

User: kajad

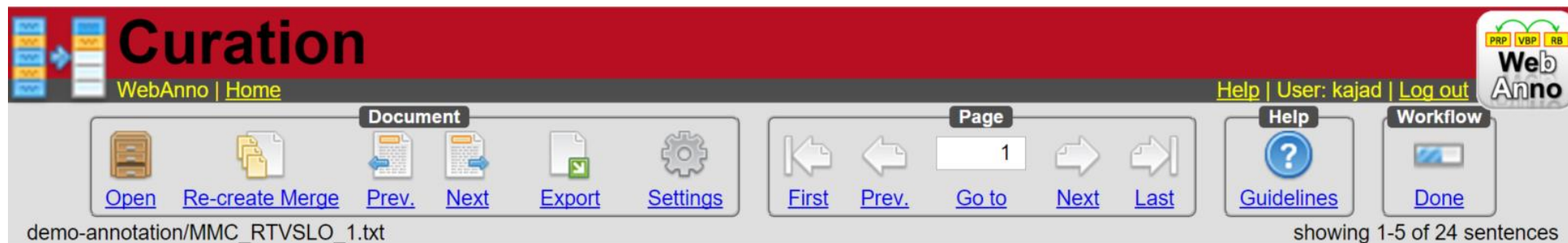
osebno stvarno
5 Glucksmann je že v svoji prvi knjigi Discours de la guerre (Razprava o vojni) leta 1968 na muho vzel vojno in politiko jedrskega zastraševanja.

Okno s končnimi oznakami

Okno z oznakami označevalcev
Sivo: ujemanje
(samodejni prenos v zgornje okno)
Modro: neujemanje
(s klikom izberemo pravo oznako)

Rdeče obarvane povedi kažejo na neujemanje označevalcev.

Pregledovanje (Curation)



Sentences

- 1 V 79. letu starosti je umrl francoski filozof Andre Glucksmann.
- Eden najvidnejših sodobnih francoskih filozofov je bil v mladosti goreč marksist in maoistični
- 2 levičar, pozneje pa se je v svojih polemičnih zapisih začel bojevati proti vsakršni obliki diktature in totalitarizma.
- Francoski predsednik Francois Hollande se je
- 3 Glucksmannu poklonil kot zagovorniku zatiranih.
- 4 "Na sramotilnem stebru je izpostavljaj tiste, ki so povzročali trpljenje," je tvtinil Hollande.
- Glucksmann je že v svoji prvi knjigi Discours de
- 5 la guerre (Razprava o vojni) leta 1968 na muho vzel vojno in politiko jedrskega zastraševanja.

Annotation

5 Glucksmann je že v svoji prvi knjigi Discours de la guerre (Razprava o vojni) vojno in politiko jedrskega zastraševanja.

Annotations: osebno, stvarno, stvarno

Možnost dodajanja novih oznak.

User: anno1

5 Glucksmann je že v svoji prvi knjigi Discours de la guerre (Razprava o vojni) leta 1968 na muho vzel vojno in politiko jedrskega zastraševanja.

Annotations: osebno, osebno

User: kajad

5 Glucksmann je že v svoji prvi knjigi Discours de la guerre (Razprava o vojni) vojno in politiko jedrskega zastraševanja.

Annotations: osebno, stvarno

Po izbiri ene izmed neujemajočih se oznak se pravilna oznaka obarva **zeleno**, napačna pa **rdeče**.

Označevanje s popraviljanjem (Correction)

 **Correction**
WebAnno | Home

 Open

 Prev

 Next

 Export

 Settings

 First

 Prev

Page

6

 Next

 Last

 Help

Guidelines

 Workflow

Done

Okno s potrjenimi oznakami

Annotation

1 Ob blesteči predstavi centrske linije so košarkarji Miamijs s 101:88 premagali Los Angeles Lakerse, ki so prvič v sezoni igrali brez Kobeja Bryanta.

2 Goran Dragič tudi tokrat ni blestel.

3 Slovenski organizator igre je v 30 minutah ob metu 2 / 10 dosegel 4 točke ter dodal še 5 skokov in 3 podaje.

4 Na zadnjih dveh točkah je vknjižil le 6 točk.

5 Rekord sezone je s 30 točkami in 11 skoki postavil Chris Bosh, četrti zaporedni dvojni dvojček (19 točk, 15 skokov) na domačem parketu pa je dosegel Hassan Whiteside.

User: Suggestion

1 Ob blesteči predstavi centrske linije so košarkarji Miamijs s 101:88 premagali Los Angeles Lakerse, ki so prvič v sezoni igrali brez Kobeja Bryanta.

2 Goran Dragič tudi tokrat ni blestel.

3 Slovenski organizator igre je v 30 minutah ob metu 2 / 10 dosegel 4 točke ter dodal še 5 skokov in 3 podaje.

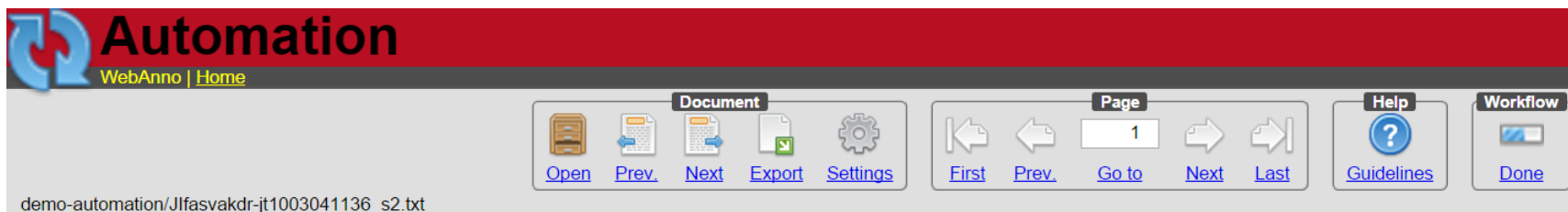
4 Na zadnjih dveh točkah je vknjižil le 6 točk.

5 Rekord sezone je s 30 točkami in 11 skoki postavil Chris Bosh, četrti zaporedni dvojni dvojček (19 točk, 15 skokov) na domačem parketu pa je dosegel Hassan Whiteside.

Okno s predlaganimi oznakami
S klikom se predlagana (**modra**) oznaka prenese v zgornje okence in obarva **sivo**. Če je predlagana oznaka napačna in jo spremenimo, se ta obarva **rdeče**.



Označevanje s sprotnim učenjem (Automation)



The screenshot shows the top interface of the Automation tool. It features a red header with the 'Automation' logo and 'WebAnno | Home' link. Below the header is a toolbar with two main sections: 'Document' and 'Page'. The 'Document' section includes icons for 'Open', 'Prev', 'Next', 'Export', and 'Settings'. The 'Page' section includes icons for 'First', 'Prev', 'Go to' (with a text input showing '1'), 'Next', and 'Last'. To the right of these are 'Help' (Guidelines) and 'Workflow' (Done) buttons. The address bar shows the file path: 'demo-automation/Jlfasvakdr-jt1003041136_s2.txt'.

Annotation

Okno za označevanje

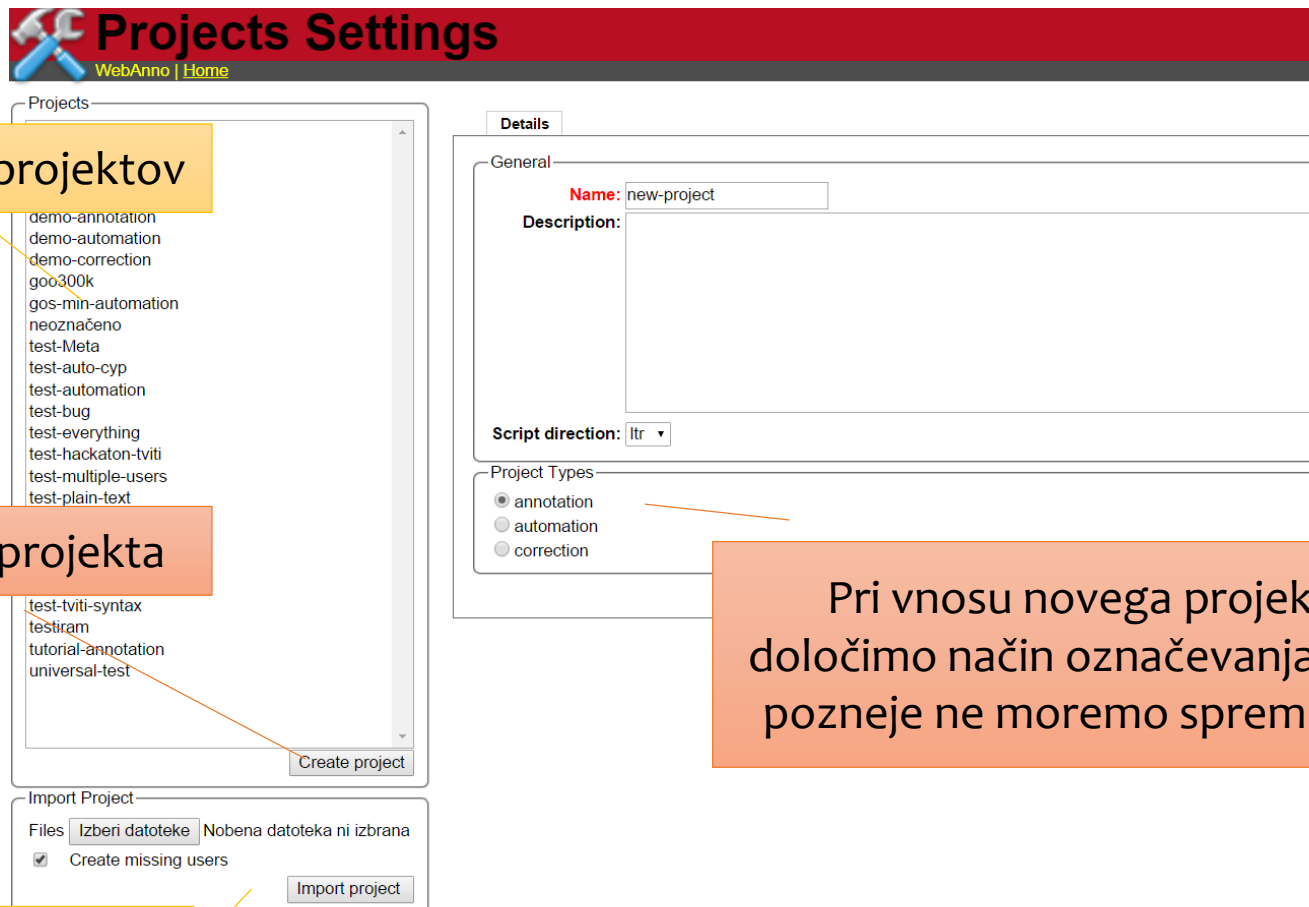
- | | |
|---|--|
| 1 | [Pf-pred-07541:] pozdravljeni danes nas je kr mal več zbranih a ne tk da vas je prav fajn pogledat \$. DM |
| 2 | [Pf-pred-07541:] en čist nekaj bi rekli en mogoče res zaključek en rezime vsega kar smo na teh preteklih treh oziroma zdaj na četrtem srečanju govorili a ne \$. |
| 3 | [Pf-pred-07541:] dejansko zakaj je tolk pomembno da se mi sploh ukvarjamo s tem kako otroka vzgojit \$. |
| 4 | [Pf-pred-07541:] zakaj je tok pomembno da mi tam se dost z njim pogovarjamo pa nekaj vzpostavmo do njega odnos da spoznamo kdo ta otrok v bistvu sploh je \$. |
| 5 | [Pf-pred-07541:] to je tist o čemer bomo dans govorili a ne \$. |
| 6 | [Pf-pred-07541:] ker vsi vemo a ne v kakšnem svetu živimo v bistvu v svetu ki je kr zahtevn za preživetje a ne \$. |

User: Suggestion

- | | |
|---|--|
| 1 | [Pf-pred-07541:] pozdravljeni danes nas je kr mal več zbranih a ne tk da vas je prav fajn pogledat \$. DM |
| 2 | [Pf-pred-07541:] en čist nekaj bi rekli en mogoče res zaključek en rezime vsega kar smo na teh preteklih treh oziroma zdaj na četrtem srečanju gov |
| 3 | [Pf-pred-07541:] dejansko zakaj je tolk pomembno da se mi sploh ukvarjamo s tem kako otroka vzgojit \$. |
| 4 | [Pf-pred-07541:] zakaj je tok pomembno da mi tam se dost z njim pogovarjamo pa nekaj vzpostavmo do njega odnos da spoznamo kdo ta otrok v bi |
| 5 | [Pf-pred-07541:] to je tist o čemer bomo dans govorili a ne \$. DM |
| 6 | [Pf-pred-07541:] ker vsi vemo a ne v kakšnem svetu živimo v bistvu v svetu ki je kr zahtevn za preživetje a ne \$. DM |

Okno s sprotno posodobljenimi predlogi
Z vsako vneseno oznako v okno za označevanje
sistem posodobi svoje znanje in **predlaga oznake
naslednjih podobnih pojavitev**. Predlagane
(modre) oznake v spodnjem okencu v zgornje
okence prenesemo s klikom.

Nastavitve projekta (Projects)



Projects Settings
WebAnno | Home

Projects

demo-annotation
demo-automation
demo-correction
goo300k
gos-min-automation
neoznačeno
test-Meta
test-auto-cyp
test-automation
test-bug
test-everything
test-hackaton-tviti
test-multiple-users
test-plain-text

test-tviti-syntax
testiram
tutorial-annotation
universal-test

Create project

Import Project

Files Nobena datoteka ni izbrana

☒ Create missing users

Details

General

Name: new-project

Description:

Script direction: ltr

Project Types

☒ annotation
☐ automation
☐ correction

Urejanje že obstoječih projektov

Začetek novega projekta

Uvoz zunanjih projektov

Pri vnosu novega projekta določimo način označevanja, ki ga pozneje ne moremo spreminjati.

Dodajanje uporabnikov

The screenshot shows the 'Users' tab of the CLARIN.SI interface. The 'Users' list contains four entries: 'jaka [user]', 'kaja [admin]', 'katja [user]', and 'spela [curator, user]'. The 'spela' entry is highlighted. A yellow callout box points to the list with the text 'Dodajanje sodelavcev s seznama vseh registriranih uporabnikov'. An orange callout box points to the 'Change Permissions' button with the text 'Določitev vlog in pooblastil' followed by a legend: 'user: označevalec', 'curator: kurator (potrjevalec)', and 'admin: vodja projekta'. At the bottom right are buttons for 'Add Users', 'Remove User', and 'Change Permissions'.

Details Users Documents Layers Tagsets Guidelines Export

Users

- jaka [user]
- kaja [admin]
- katja [user]
- spela [curator, user]

Dodajanje sodelavcev s seznama vseh registriranih uporabnikov

Določitev vlog in pooblastil

- user:** označevalec
- curator:** kurator (potrjevalec)
- admin:** vodja projekta

Add Users Remove User Change Permissions

Uvoz dokumentov

Details Users Documents Layers Tagsets Guidelines Export

Import new documents

Format: Binary format
Files: Binary format ka ni izbrana
CoNLL 2000
CoNLL 2002
CoNLL 2006
CoNLL 2009
Plain text
WebAnno CPH TEI reader
WebAnno TSV Format
WebAnno TSV Format - Old
Weblicht TCF Format
XMI format

Documents

Import document

Izbira formata enega ali več uvoženih dokumentov projekta.

Specifikacija nivojev označevanja

The screenshot shows a web-based configuration tool for linguistic layers. The interface includes a top navigation bar with tabs: Details, Users, Documents, Layers (active), Tagsets, Guidelines, and Export. The main content area is divided into several sections:

- Layers (Help):** A list of existing layers: Chunk, Coreference, Dependency, Lemma, Multi-word units, Named Entity, and POS. A 'Create layer' button is at the bottom.
- Properties (Help):** Fields for 'Layer name' (Multi-word units), 'Description', and 'Enabled' (checked).
- Technical Properties (Help):** Fields for 'Type' (span) and 'Attach to layer' (-NONE-).
- Behaviors (Help):** Checkboxes for 'Lock to token offsets', 'Allow stacking', 'Allow crossing sentence boundary', and 'Allow multiple tokens' (checked).
- Import layer:** A section with a 'Files' input and an 'Import layer' button.
- Feature overview (Help):** A list of features, currently showing 'mwe type : [String]'. A 'New feature' button is at the bottom.
- Feature details (Help):** Fields for 'Type' (uima.cas.String), 'Feature name' (mwe type), 'Description', 'Enabled' (checked), 'Show' (checked), 'Remember' (unchecked), and 'Tagset' (MWE-LexSem). 'Save layer' and 'Export layer' buttons are at the bottom.

Annotations in Slovenian explain the interface elements:

- Urejanje obstoječih nivojev**
zeleno: privzeti nivoji
modro: nivo po meri
rdeče: izklopljen nivo
- Uvoz zunanjih nivojev**
- Priprava novega nivoja po meri**
- 1. določitev tipa nivoja (niz, povezava, veriga)**
- 2. določitev omejitev nivoja**
- 3. določitev enega ali več tipov oznak**
- 4. določitev vnaprej definiranega nabora oznak (sicer NONE)**

Zamejitev nabora oznak

Urejanje obstoječih naborov

Uvoz zunanjega nabora

1. Priprava novega nabora

2. vnos posameznih
oznak v danem naboru

Users Documents Layers Tagsets Guidelines Export

Tagsets

BART
MWE-LexSem
NER_WebAnno
STTS
Tiers

Create tagset

Tagset details

Name: MWE-LexSem
Language: sl
Description:

Extendible tagset?: ☒

Export as: JSON Export tagset Save tagset Delete tagset

Import tagsets

Format: JSON
Files: Izberi datoteke Nobena datoteka ni izbrana
Import tagset

Tags

PP modifier
compound
fixed phrase
named entity
proverb
semi-fixed VP
verb-particle
verb-preposition

Create tag

Tag details

Name: conjunction
Description:

Save tag Delete tag

Izvoz projekta

DetailsUsersDocumentsLayersTagsetsGuidelinesExport

Export

Warning:

Do not leave this page or close the browser while export is running!
Leaving this page or closing the browser may lock the project settings. Unlocking them again requires a restart of the server!

Progress:

0%

Cancel

Export format:

WebAnno TSV Format

AUTO

Binary format

CoNLL 2000

CoNLL 2002

CoNLL 2006

CoNLL 2009

Plain text

WebAnno TSV Format

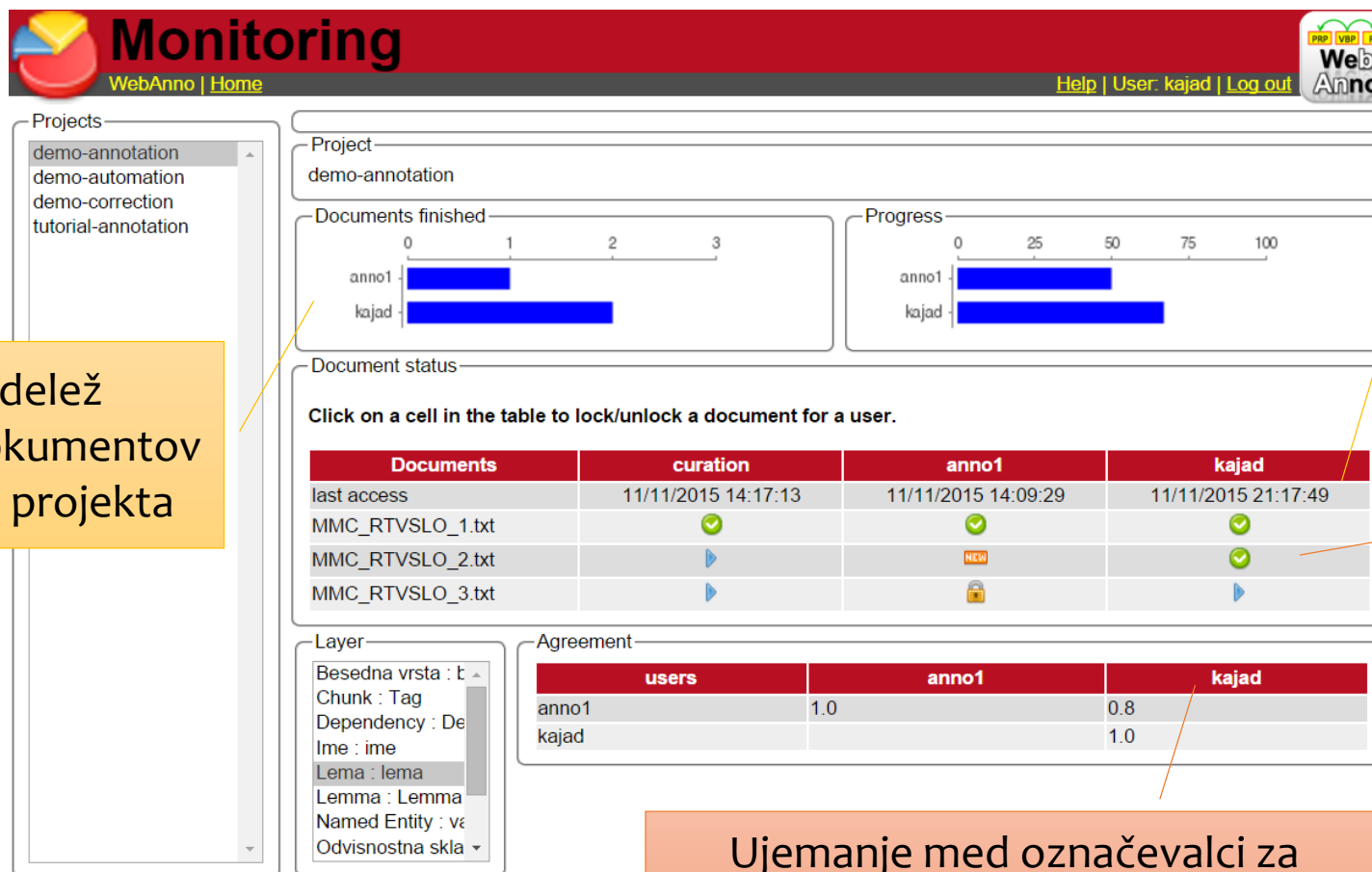
Weblicht TCF Format

XML format

Export the whole project to the file system

Izvoz projekta v izbranem formatu
Izvozijo se vse oblike dokumentov projekta,
vključno z začetnimi besedili, anotacijami
posameznih označevalcev in končnimi združenimi
anotacijami. Izvozijo se tudi vse nastavitve projekta,
kot so nastavitve nivojev, nabori oznak ipd.

Spremljanje (Monitoring)



Čas zadnje prijave

Dodeljevanje in spremljanje statusa dokumentov

new: neodprto
puščica: v teku
kljukica: zaključeno
ključavnica: onemogočeno

Ujemanje med označevalci za opazovani nivo (kappa)

Povzetek

- spletni dostop
 - en ali več nivojev označevanja
 - poljubne ali vnaprej opredeljene oznake
 - različni načini označevanja
 - preprosta priprava projekta
-
- jezikoslovna in nejezikoslovna raba
 - samostojno označevanje ali označevalni projekt

Podpora

- spletna stran
<https://webanno.github.io/>
- priročnik za uporabnike
<https://webanno.github.io/webanno/releases/2.3.0/docs/user-guide.html>
- video predstavitve
<https://www.youtube.com/user/webanno>
- poštni seznam
<https://groups.google.com/forum/#!forum/webanno-user>
<https://github.com/webanno/webanno/issues>
- CLARIN.SI
info@clarin.si
- publikacije

VAJE

Normalizacija nestandardne slovenščine

▪ **vaja 1: označevanje** (Annotation)

- vpis v sistem
- projekt *tutorial-annotation*
- dva dokumenta po 7 tvitov
- nivo *Normalisation*
- prilagoditev vmesnika
- označevanje
- zaključek

Npr.

Annotation

1 @janez ^{hehe} Hehehe, to res ^{ne | vem} nvem , ampak ^{ona} una štacuna že ^{zdavnaj} zdaavni ni ^{več} vec odprta lol :) ^{pojdi} Pejt ^{jutri} ju3
^{Ljubljano} v ljubano ^{ali} čekirat al pa ^{vprašaj} vpraši ^{boss-a} big bossa #goodluck

Normalizacija nestandardne slovenščine

▪ **vaja 2: pregled tretjega označevalca (Curation)**

- zavihek *Curation*
- nivo *Normalisation*
- projekt *tutorial-annotation*
- izbira dodeljenega besedila
- pregled in potrditev končnih oznak
- izvoz in zaključek