Razvoj semantičnih tehnologij za slovenski jezik (RSDO DS3)

www.eu-skladi.si

rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

Agenda

- O sklopu 3
- Tehnologije in orodja (DEMO)
  - Imenske entitete in koreferenčnost
  - Ekstrakcija povezav
  - Baza znanja
  - Semantični premiki in diahrone analize
- „Infrastrukturna“ orodja
  - ANGLEr
  - SloBENCH

O Sklopu 3 - sodelujoči

UL FRI

Marko Robnik-Šikonja

Slavko Žitnik

UM FERI

Milan Ojsteršek

IJS

Senja Pollak

Erik Novak, Aljaž Košmerlj

UL FF, UNG, ZRC SAZU

Simon Krek

Prve verzije orodij na voljo do konca 2022.

rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

O Sklopu 3 – predvidena orodja

Nadgradnja “**digitalne slovarske baze**”

Prepoznavanje imenskih entitet

Ekstrakcija povezav

Odkrivanje koreferenčnosti

Izdelava baze znanja

Orodje za razdvoumljanje

Prepoznavanje semantičnih premikov in **izvajanje diahronih analiz**

Avtomatsko povzemanje krajših in daljših besedil

Avtomatsko odgovarjanje na vprašanja

rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

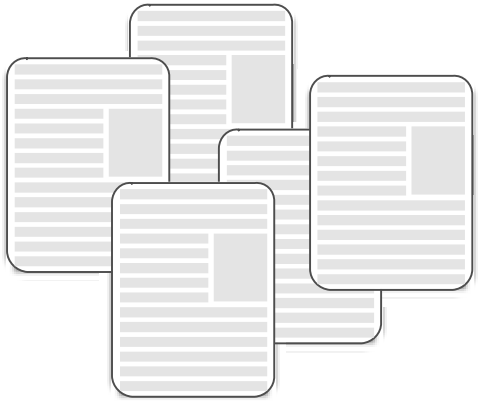
EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

Kaj so semantične tehnologije?

"**Semantične tehnologije** uporabljajo metode **umetne intelligence**, da simulirajo **razumevanje jezika** in **procesiranje informacij**."

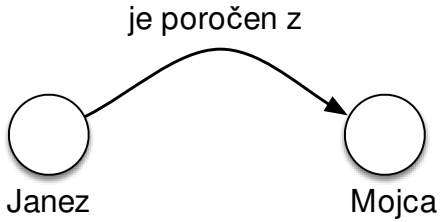
Podatek

Originalno iz 1600 s pomenom "nekaj danega"



Informacija

Originalno iz 1300 z nanašanjem na "akt informiranja"



rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

Prepoznavanje imenskih entitet in koreferenčnosti

Audi je izdelovalec luksuznih avtomobilov. Podjetje je bilo ustanovljeno v Nemčiji.

Ustanovil ga je August Horch v letu 1910. Horch je pred tem imel že drugo podjetje s svojimi popularnimi modeli. V Audiju so začeli s štiri-cilindrskimi modeli. Do leta 1914 so Horchovi avtomobili že dirkali in zmagovali.

August Horch je podjetje Audi zapustil leta 1920 in prevzel mesto predstavnika za združenje motornih vozil Nemčije.

Audi je trenutno hčerinsko podjetje skupine Volkswagen in proizvaja kvalitetne avtomobile.

rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

Prepoznavanje imenskih entitet in koreferenčnosti

Audi je izdelovalec luksuznih avtomobilov. **Podjetje** je bilo ustanovljeno v **Nemčiji**.

Ustanovil **ga** je **August Horch** v letu 1910. **Horch** je pred tem imel že drugo podjetje s **svojimi** popularnimi modeli. V **Audiju** so začeli s štiri-cilindrskimi modeli. Do leta 1914 so **Horchovi** avtomobili že dirkali in zmagovali.

August Horch je **podjetje Audi** zapustil leta 1920 in prevzel mesto predstavnika za združenje motornih vozil **Nemčije**.

Audi je trenutno hčerinsko podjetje **skupine Volkswagen** in proizvaja kvalitetne avtomobile.

rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

Prepoznavanje imenskih entitet in koreferenčnosti

Audi je izdelovalec luksuznih avtomobilov. **Podjetje** je bilo ustanovljeno v **Nemčiji**.

Ustanovil **ga** je **August Horch** v letu 1910. **Horch** je pred tem imel že drugo podjetje s **svojimi** popularnimi modeli. V **Audiju** so začeli s štiri-cilindrskimi modeli. Do leta 1914 so **Horchovi** avtomobili že dirkali in zmagovali.

August Horch je **podjetje Audi** zapustil leta 1920 in prevzel mesto predstavnika za združenje motornih vozil **Nemčije**.

Audi je trenutno hčerinsko podjetje **skupine Volkswagen** in proizvaja kvalitetne avtomobile.

rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

Ekstrakcija povezav

Janez Drnovšek je bil predsednik Liberalne demokracije Slovenije.

rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

Ekstrakcija povezav

Janez Drnovšek je bil predsednik **Liberalne demokracije Slovenije**.

rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

Ekstrakcija povezav

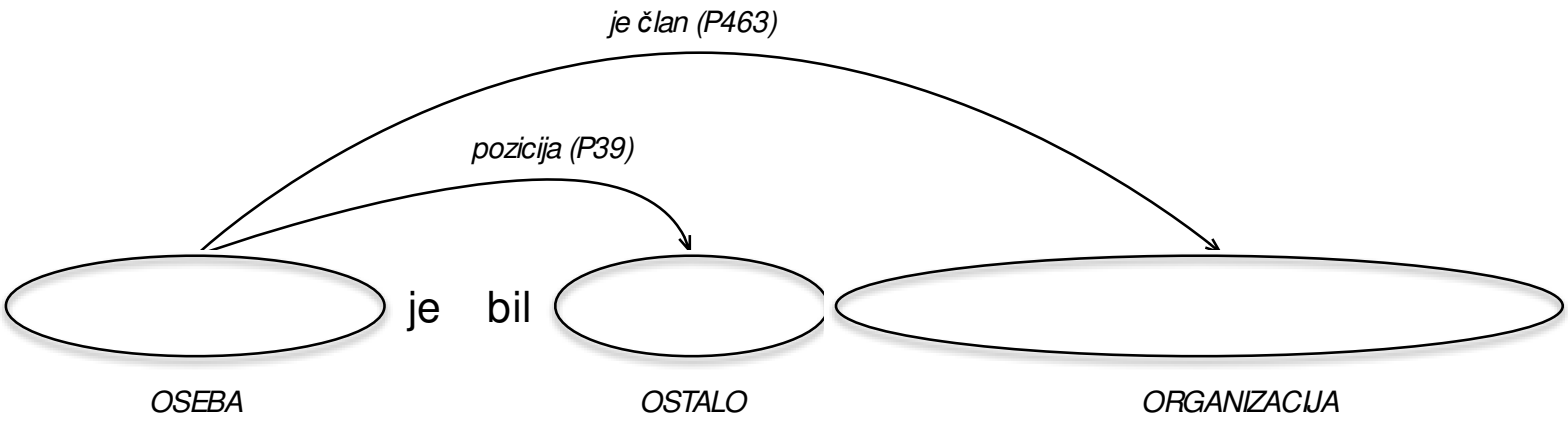
Janez Drnovšek je bil ***predsednik*** **Liberalne demokracije Slovenije**.

rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

Ekstrakcija povezav



Ekstrakcija povezav

Avtomatsko izdelan korpus za ekstrakcijo povezav

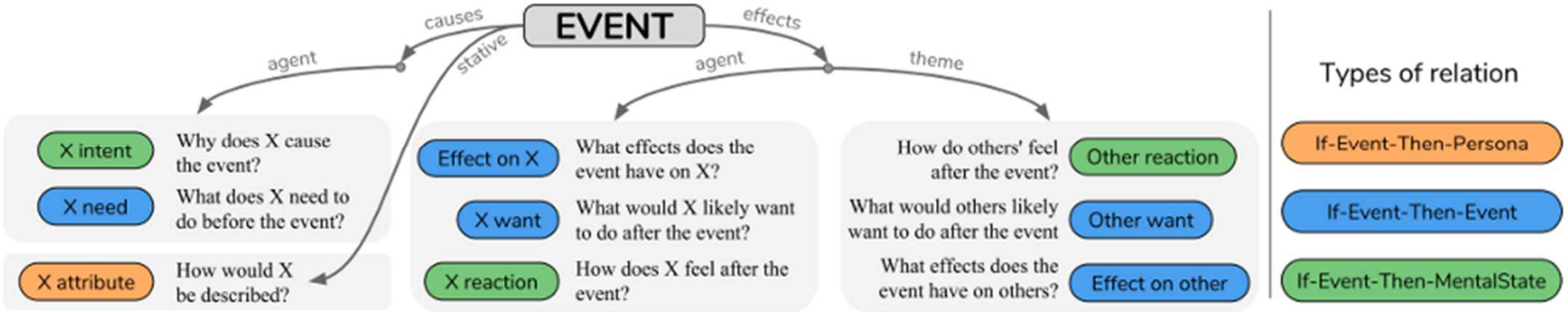
29 tipov povezav

Wikidata značka	ime relacije	celoten korpus	učna množica	validacijska množica	testna množica
P0	prazna	0	30721	4455	8901
P131	se nahaja v	31745	28644	4110	8217
P150	administrativne podenote	22061	15400	2165	4496
P31	primerek od	18463	13064	1819	3580
P106	poklic	16605	11543	1670	3392
P527	vsebuje	13521	9460	1410	2651
P17	država, kateri objekt pripada	72806	9299	1333	2701
P156	naslednik	10841	7733	1041	2067
P155	predhodnik	10709	7647	1030	2032
P361	je del	9930	6988	1005	1937
P19	kraj rojstva	9812	6094	880	1799
P3450	sezona športne prireditve	5820	4105	603	1112
P20	kraj smrti	7041	3650	491	989
P138	poimenovano po	4459	3083	440	936
P641	šport povezan s tem	4147	2910	410	827
P172	etnična skupina	4015	2837	366	812
P27	država državljanstva	8861	2830	389	762
P276	kraj, lokacija	3641	2575	361	705
P3373	sorojenec	3139	2206	322	611
P607	bitka povezana s tem	3013	2093	301	619
P1001	spada pod upravo	2667	1884	258	525
P279	podpomenka od	12621	1818	271	500
P50	avtor	1344	1444	193	431
P140	vera	2052	1421	219	414
P136	žanr	1673	1191	160	322
P40	otrok	2455	1047	133	238
P39	uradni položaj	1294	910	129	255
P463	član organizacije	1231	902	113	216
P22	oče	1989	837	102	203
P25	mati	451	220	28	38

rsdo
REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO
EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

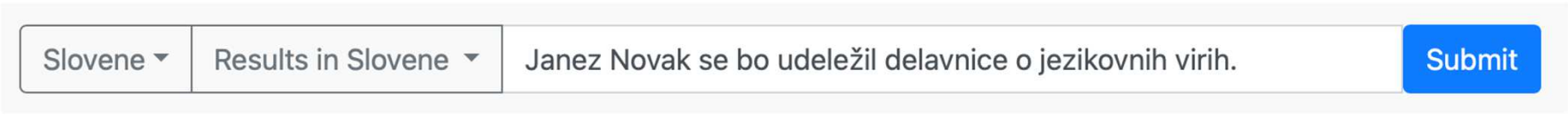
Baza znanja

Allen Institute for AI (2020): *Commonsense Inferences about People and Events*



rsdo
REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO
EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

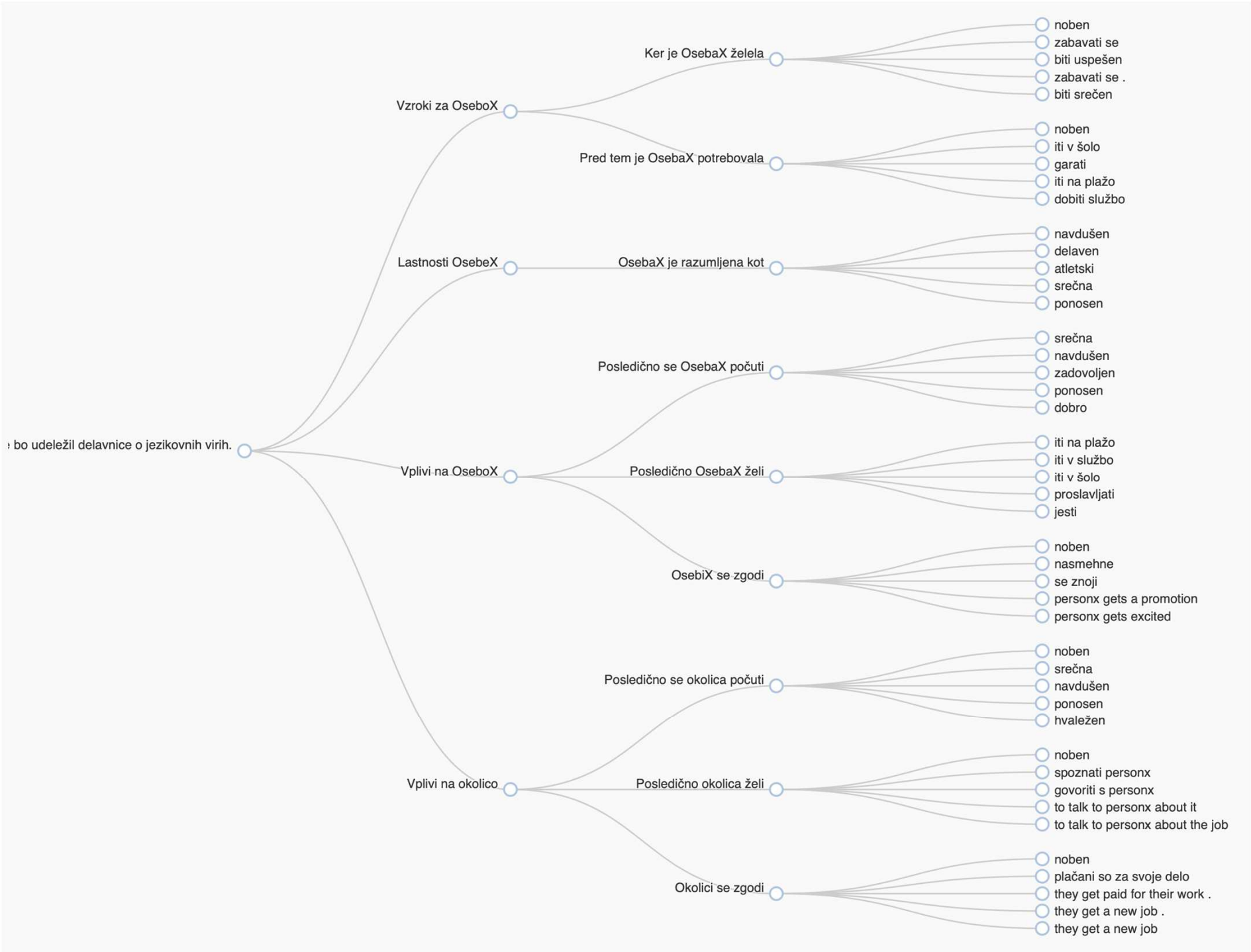
Baza znanja

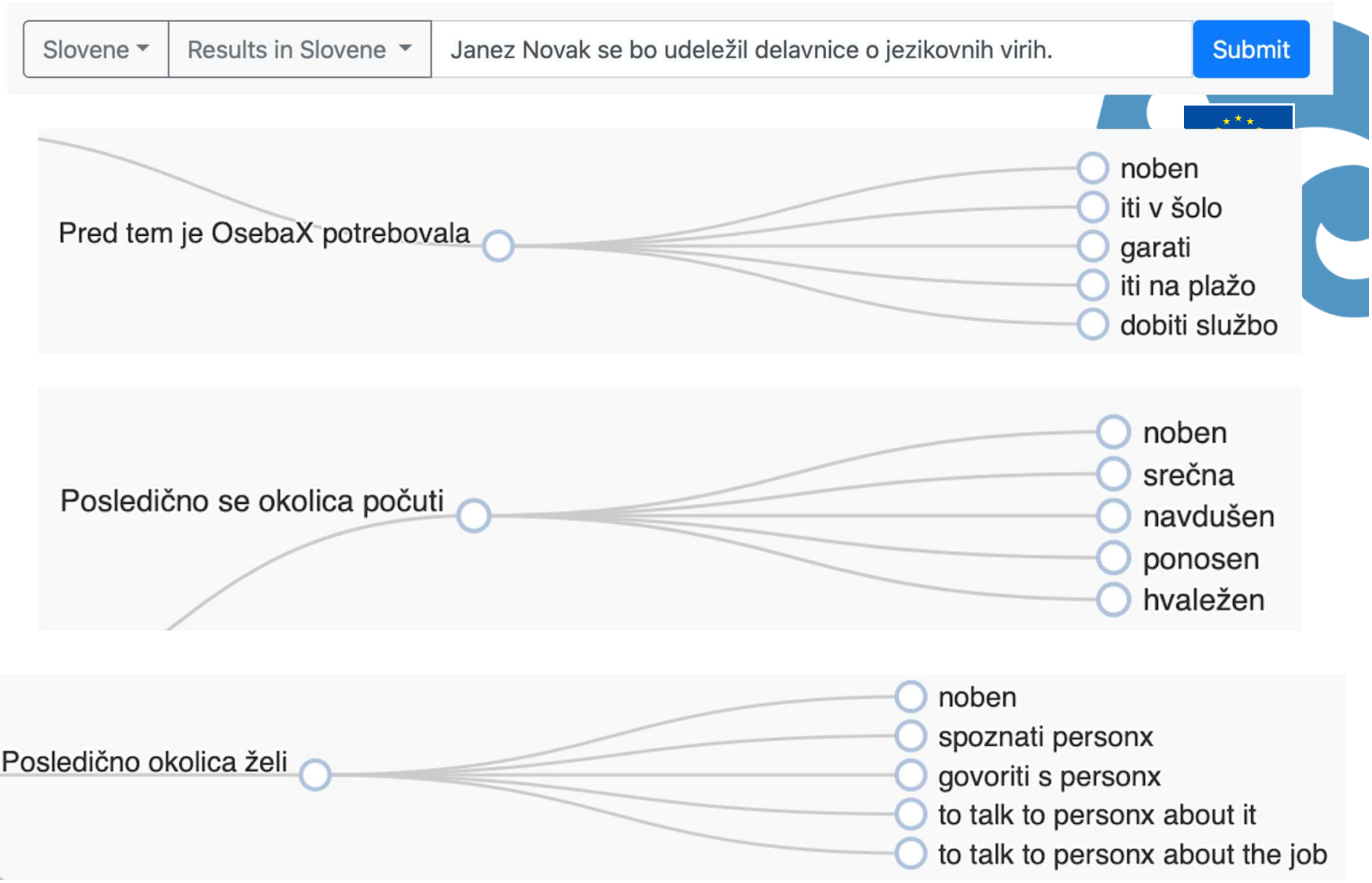


rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

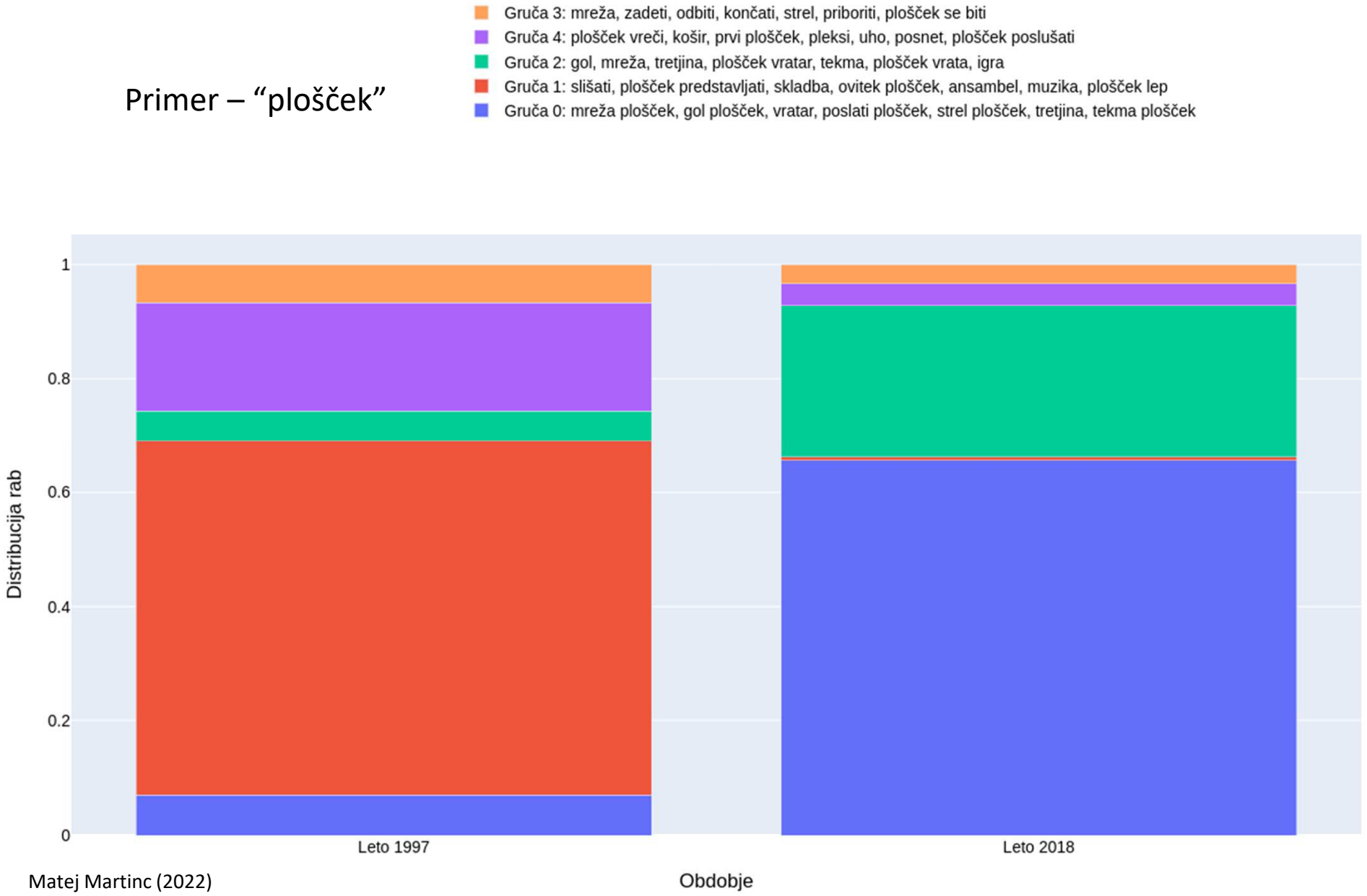

bo udeležil delavnice o jezikovnih virih.
Vzroki za OseboX
Ker je OsebaX želela
noben
zabavati se
biti uspešen
zabavati se .
biti srečen
Pred tem je OsebaX potrebovala
noben
iti v šolo
garati
iti na plažo
dobiti službo
Lastnosti OsebeX
OsebaX je razumljena kot
navdušen
delaven
atletski
srečna
ponosen
Vplivi na OseboX
Posledično se OsebaX počuti
srečna
navdušen
zadovoljen
ponosen
dobro
Posledično OsebaX želi
iti na plažo
iti v službo
iti v šolo
proslavljati
jesti
OsebiX se zgodi
noben
nasmehne
se znoji
personx gets a promotion
personx gets excited
Vplivi na okolico
Posledično se okolica počuti
noben
srečna
navdušen
ponosen
hvaležen
Posledično okolica želi
noben
spoznati personx
govoriti s personx
to talk to personx about it
to talk to personx about the job
Okolici se zgodi
noben
plačani so za svoje delo
they get paid for their work .
they get a new job .
they get a new job



rsdo
REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO
EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST

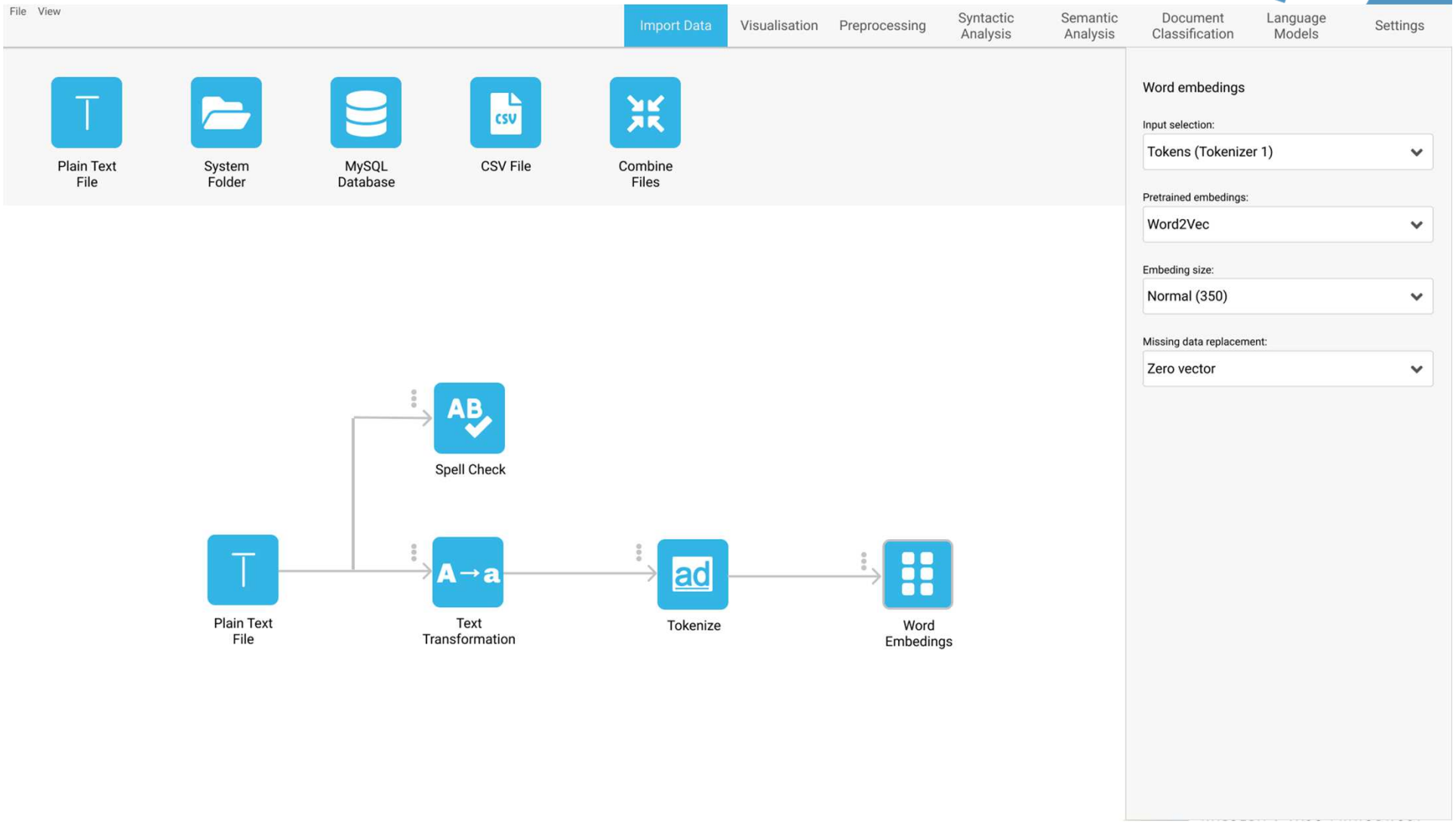
Semantični premiki in diahrone analize

Primer – “plošček”



Matej Martinc (2022)

Splošno orodje za izvajanje analiz nad besedili



Knez, Bajec, Žitnik (2022): *ANGLEr: A Next-Generation Natural Language Exploratory Framework*, RCIS 2022

Vrednotnik SloBENCH

cjvt slobench 1.0 Login Register

Leaderboards

Question answering (SuperGLUE)
This question answering leaderboard is a Slovene (translated) version of an existing English version of SuperGLUE benchmark.
15 February 2022 | 1 | 1.0

Machine Translation (ENG -> SLO)
This machine translation leaderboard is measuring a success of automatic machine translation from English to Slovene language.
18 March 2022 | 4 | 1.0

Machine Translation (SLO -> ENG)
This machine translation leaderboard is measuring a success of automatic machine translation from Slovene to English language.
18 March 2022 | 4 | 1.0

Frequently Asked Questions

- How can I make a submission?
- I have questions or want to report bugs
- How can my submission get special tags - "Verified", ...?

About SloBench - Slovenian NLP Benchmark

SloBENCH is an evaluation platform for benchmarking the development of Slovene natural language processing technologies. The main goal of SloBENCH evaluation is that evaluation data labels are not publicly published. Additionally, submission limitation are defined to avoid overfitting of models. Still we ask all the submitters to respect ethics and to not manually annotate the test data.

The evaluation scripts are publicly available. To support the community we propose to open-source as much of the submitted systems as possible. Initially, up to ten different leaderboards will be set up and gradually updated later.

The system was developed as one of Clarin.si's 2021 projects and was proposed by Slavko Žitnik, Marko Robnik-Šikonja and Simon Krek. The implementation was done by Frenk Dragar.

Contact

Centre for Language Resources and Technologies
Večna pot 113
SI-1000 Ljubljana

Phone
+386 1 479 82 99

Email
info@cjvt.si

Uporabne povezave

https://slovenscina.eu

https://github.com/orgs/RSDO-DS3/repositories

https://multicomet.ijs.si

https://slobench.cjvt.si

rsdo

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO

EVROPSKA UNIJA
EVROPSKI STRUKTURNI IN INVESTICIJSKI SKLADI
NALOŽBA V VAŠO PRIHODNOST