

Korpusi in TEI XML

Od digitalizacije do konkordančnika

Katja Meden,
Inštitut Jožef Stefan, Inštitut za novejšo zgodovino
2. 12. 2024,
Predavanje v okviru prakse RI INZ PRINZ



Vsebina

- Uvod in (parlamentarni) korpusi
- XML, TEI in XSLT
 - EXtensible Markup Lanugage (XML)
 - Text Encoding Initiative (TEI)
 - TEI XML
- Od dokumenta do korpusa - Proces priprave korpusa:
 - Digitalizacija in OCR
 - Čiščenje podatkov
 - Označevanje podatkov
 - Strukturiranje podatkov
 - Objava korpusa - repozitorij, konkordančnik
- Raziskovanje korpusov

kórp^{us} -a m (ô)

SSKJ²



1. *vojaška enota iz divizij in posebnih enot*: formirati korpus; poveljnik korpusa / ekspedicijski korpus
2. jezikosl. *zbirka besedil v elektronski obliki s standardizirano obliko zapisa in izbranimi podatki*: analiza korpusa; delo s korpusom; besedila za korpus; korpus in slovar / referenčni korpus *ki je glede na vključena besedila zvrstno kar se da uravnotežen / besedilni korpus*
3. publ., navadno s prilastkom *skupina, zbor*: nastopili so najboljši korpusi; baletni, godalni korpus / diplomatski korpus *diplomatski zbor*; korpus miru zlasti v amerškem okolju *organizacija iz prostovoljcev, ki skrbi zlasti za gospodarski razvoj in pomoč nerazvitim deželam*
4. knjiž. *zbirka*: korpus srednjeveških fresk
 - ekspr. *ta ženska je pa korpus je zelo debela*

Tipi korpusov

- Referenčni korpusi (splošen korpus, ki ponuja pregled jezika, npr. [Gigafida](#), [metaFida](#))
- Vzporedni korpusi (korpusi v različnih jezikih):
 - [Japonsko-slovenski vzporedni korpus jaSlo](#)
- Specializirani korpusi:
 - Korpusi družbenih omrežij/forumov/nestandardnih jezikov: ([Janes Tweet](#), [Janes Forum](#), [Janes News](#), [Janes TwePo](#) – tviti politikov)
 - Korpusi spleta ([CLASSLA-web.sl](#),)
 - Korpusi Wikipedij ([CLASSLA-Wikipedia](#) ...)
 - Parlamentarni korpusi
- Korpusi so lahko hkrati vzporedni in specializirani (npr. korpusi ParlaMint)

Korpusi parlamentarnih razprav

- Parlamentarne razprave: transkripcije plenarnih zasedanj, zapisniki sej
- Prepisi govornega jezika
- Vsebujejo dragocene (sociodemografske) metapodatke:
 - Ime in priimek govorca
 - Datum rojstva
 - Politične / parlamentarne stranke
 - Spol govorca...
- Korpusi parlamentarnih razprav: digitalne zbirke prepisov parlamentarnih razprav
- Bogat vir za različne raziskovalne vede, predvsem za:
 - družboslovje in humanistiko
 - Raziskovanje parlamentarnega/političnega diskurza

Lastnosti

- Prepisi niso vedno dobesedni
- Izpuščene napake, neprimeren jezik
- Oblikovani v skladu z konvencijami parlamenta:
 - Grški parlament: prekinitve strogo prepovedane, sankcionirane
 - Britanski parlament: prekinitve so popolnoma običajne in niso sankcionirane
 - Poslanci imajo pogosto vnaprej pripravljene govore
- TEI kodiranje in jezikovna anotacija besedila

Official records

"Madam President, I think we all agreed that equal pay is, first of all, a matter of elementary fairness, and the fact that we still have this average pay gap of 16-17% also shows, among other things, the lack of respect for women, which is so deeply rooted in traditions and stereotypes. /.../."

(Věra Jourová, Member of the Commission)

Verbatim transcription

*"**Ladies and Gentlemen, yes gentlemen also are here still. Am...** I think we all agreed that equal pay is, first of all, a matter of elementary fairness, and **that** the fact that still **we** have this **pay gap am...** average 16-17% **this** also shows, among other things, the lack of respect to women, which is so deeply rooted in traditions and stereotypes. /.../."*

(Věra Jourová, Member of the Commission)

Darja Fišer in Kristina Pahor de Maiti. 2021. *Prvič, sem političarka in ne politik, drugič pa ---* : a corpus approach to parliamentary discourse research. Inštitut za novejšo zgodovino, Ljubljana. Dostopno na: <https://sidih.si/cdn/121/ch3.html>

Slovenski parlamentarni korpusi

- [Slovenski parlamentarni korpus \(1990 – 2022\) siParl 4.0](#)
- [Parlamentarni korpusi ParlaMint \(ParlaMint-SI, 2000 - 2022\)](#)
- [Parlamentarni korpus prve Jugoslavije \(1919-1939\) Yu1Parl](#)
- [Parlamentarni korpus Kranjskega deželnega zbora \(1861 – 1913\)](#)

	Kranjska 1.0	Yu1Parl 1.0	siParl 4.0	ParlaMint-SI 4.1
Obdobje	1861 – 1913	1919 – 1939	1990 – 2022	2000 - 2022
Zapisniki sej	Kranjski deželni zbor	Predstavništvo, senat in skupščina Kraljevine Jugoslavije	Državni zbor, Delovna telesa DZ, Kolegij predsednika DZ	Državni zbor
Jeziki	slovenščina (58%), nemščina (42%)	srbsščina (cirilica, 59%), hrvaščina (38%), slovenščina (3%)	slovenščina	slovenščina, angleščina (strojni prevodi)
Velikost korpusa (besede)	10.883.619	13.271.885	239.224.540	69.032.700

SiParl 4.0: Korpus parlamentarnih razprav Državnega zbora Republike Slovenije 1990-2022

- SiParl 4.0: <http://hdl.handle.net/11356/1936>
- Raziširitev do 13.5.2022 (8. mandat, seje delovnih teles in Kolegija predsednika DZ)
- Konkordančniki - SiParl 4.1 in ParlaMint 4.1
- 13233 posameznih transkriptov sej (1 dokument = 1 seja)
- 12.232.250 govorov
- 239.224.540 besed
- Podatki kodirani v skladu z Parla-CLARIN priporočili (TEI Speech)

ParlaMint

- ParlaMint I (Julij 2020 – Maj 2021):
 - 17 nacionalnih evropskih parlamentarnih korpusov
 - Vsak korpus vsebuje približno 20 milijonov besed
 - COVID/War/Reference podkorporusi
- ParlaMint II (December 2021 – Maj 2023)
 - Novi partnerji (skupaj 29 korpusov)
 - Novi metapodatki
 - Strojno prevajanje besedila
 - ParlaMint 4.1: <http://hdl.handle.net/11356/1912>
 - ParlaMint-SI (3. - 8. mandat)



siParl vs. ParlaMint-SI

siParl 4.0

- Nacionalni korpus
- Kodiranje odraža specifike pol. sistema (TEI Drama + TEI Speech)
- Obseg podatkov:
 - 1990 - 2022
 - 11. sklic Skupščine (SSK11)
 - 1. do 8. mandatno obdobje
 - Poleg plenarnih sej tudi seje delovnih teles in Kolegija
 - Prilagojeno kodiranje v TEI (Parla-CLARIN)

ParlaMint-SI 4.1

- primerljivi (comparable) korpusi
- Uniformno kodiranje (zanemarja določene specifike sistema za lažjo primerjavo)
- Obseg podatkov:
 - 2000 - 2022
 - 3. do 8. mandatno obdobje
 - Zgolj plenarne seje (SDZ)
 - Kodiranje po striktni ParlaMint shemi

Kako ustvarimo takšne (in podobne) korpuse?

Surovi podatki -----> Korpus v konkordančniku (iskanje)

SLAVKO GRČAR: Hvala. Najprej hvala za povabilo na to sejo. Rad bi zelo na kratko obvestil to komisijo o nekaterih takih stvareh, da v zadnjih skoraj dvajsetih letih, številnih naših dokumentih in govorih, smo nesmiselno obsodili poudarjam zaradi tega, ker se še vedno pojavljajo ugotovitve, da smo mi zagovorniki poveljnih pobojev. To ni res. O Druga stvar, ki bi jo rad izrečno povedal je, da podpiramo strokovno tehnično delovanje in druge strokovne aktivnosti vojnih grobiščih. Torej tudi strokovno tehnično delo te komisije o katerem je danes pogovor. Tudi nam se zdi, da se dolgo že vlečejo in da bi morda kakšne stvari bilo mogoče hitreje in temeljiteje opravljati v dosedanjih obdobjih pa aktivnost kaj bolj intenzivna. Seveda pa ne podpiramo strupenih agitk, ki se vežejo na odkritje teh tako imenovanih prikritih grobišč. Mislimo, da povezane z 2. svetovno vojno, začenjamo pogovarjati na kulturni način. Tudi v skladu s 63. členom slovenske Ustave, nestrpnosti in sovražstva. Mislimo, da je čas za to. Mislimo tudi, da je čas za kulturo dialoga, da se je treba začeti pogovarjati kot se je zdaj velikokrat dogajalo pa ne bom citiral posameznih izjav, ampak saj ste jim bili sami priča. To ni dobro za mednarodne odnose. Mnoge stvari, ki se pri nas dogajajo in potem prepotencirajo, potem izkoriščajo v tujini na našo škodo, recimo v Italiji, imeli smo jih na Hrvaškem pa morda še kje. Torej tu bi morali z eno odgovornostjo, bi rekel, se obrniti. Predvsem se nam tudi zdi sporno, da o tem razpravljamo marsikdaj izdvojevano, iz konteksta velikanskih morij in zločinov, ki so jih nad slovenskim in drugimi jugoslovanskimi narodi uresničevali okupatorji in njihovi sodelavci, iztrgano iz konteksta žrtev, ki so jih ti povzročili. V Sloveniji imamo zelo relevantno raziskavo, ki jo opravlja **Inštitut za novejšo zgodovino** s poimenskimi evidentiranjem vseh žrtev, slovenskih žrtev, mislim na slovenskih tleh tekom 2. svetovne vojne in veste, da je teh žrtev naštetih že 81 ali 82.000. Razen tega še 12.000 potem po vojni pobitih domobrancev. Od teh 82.000 je devet desetlin teh žrtev na račun okupatorjev in njihovih sodelavcev.

simple **Inštitut za novejšo zgodovino** • 49
0.10 per million tokens • 0.000018%

☐ Details

Left context

KWIC

Right context

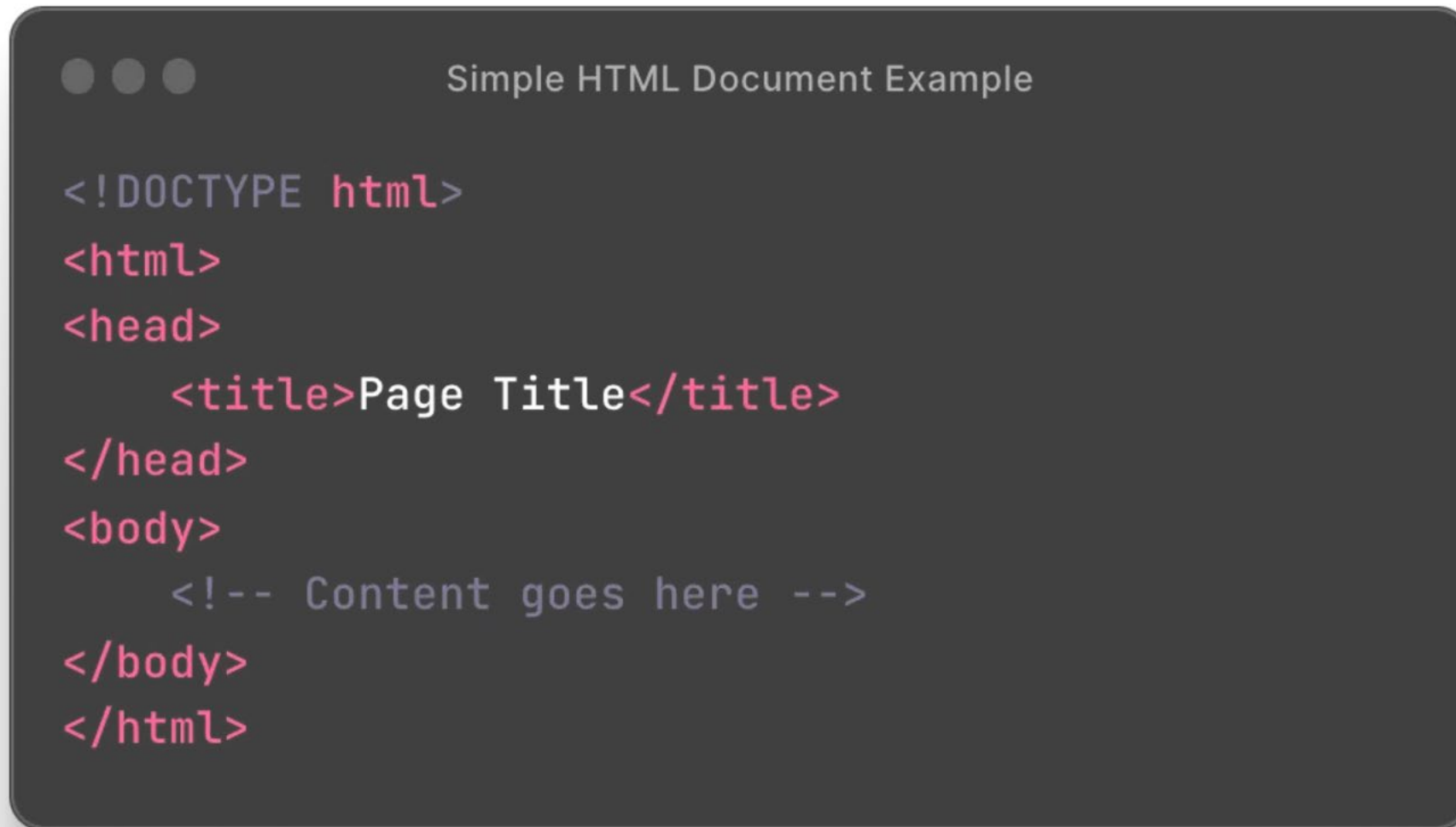
1	☐ ☐ mečMatjaž • 2...	vini, takrat se je ta odbor dejansko izrazil bistveno bolj demokratično kot tokrat, takrat se je	Inštitut za novejšo zgodovino	zaprlo za posvet o tem kateri datum bi bil najbolj primeren in seveda inštitut za novejšo z
2	☐ ☐ NemečMatjaž • 2...	novejšo zgodovino zaprosilo za posvet o tem kateri datum bi bil najbolj primeren in seveda	Inštitut za novejšo zgodovino	se je opredelil do tega vprašanja in tega z vsem dolžnim spoštovanjem so naredili tako spo
3	☐ ☐ NemečMatjaž • 2...	razpravo in naj predlagajo tako kot je bilo pri primorskem zakonu in drugih zakonih s strani	Inštitut za novejšo zgodovino	predlaganih štiri datume, ki so zgodovinsko utemeljeni, kajti vemo, če ste prebrali dopis in t
4	☐ ☐ SiterPrimož • 2...	omne probleme, prostorske, infrastrukturne, zgodovinske, Arhiv Ljubljana recimo, knjižnica	Inštituta za novejšo zgodovino	sta lepa primera teh težav, pa so tukaj še druge ustanove. Zato v Levici torej predlagamo, c
5	☐ ☐ PavlinBlaž • 20...	kvidirala politične nasprotne, tako vojne ujetnike kot tudi civilno prebivalstvo. Po podatkih	Inštituta za novejšo zgodovino	je število teh žrtev samo med Slovenci 14 tisoč 885. Žrtve poveljnih pobojev so bile usmrče
6	☐ ☐ PriklUroš • 201...	omenikov in spominskih znamenj in se sprejmejo ustrezni ukrepi. Dalje. Zagotovi se, da bo	Inštitut za novejšo zgodovino	lahko dokončal register in analizo slovenskih smrtnih žrtev druge svetovne vojne. Dalje. Za
7	☐ ☐ TurnšekTit • 20...	o umrli, zakaj so umrli. 17 podatkov je o vsakem človeku zbranih in ta register je 5 let delal	Inštitut za novejšo zgodovino	v Ljubljani, in se ta register še vedno vzdržuje. Po tem registru ali v skladu s tem registrom,
8	☐ ☐ PriklUroš • 201...	omenikov in spominskih obeležij ter se sprejmejo ustrezni ukrepi. Zagotoviti je treba, da bo	Inštitut za novejšo zgodovino	lahko dokončal register in analizo slovenskih smrtnih žrtev v 2. svetovni vojni. Zavezati je tr
9	☐ ☐ KozlovičLilijan...	emo na način, da razumemo, da so v času vojne ta dejanja počeli vsi. In to je povedal tudi	Inštitut za novejšo zgodovino	. Povedal je, da so to počeli vsi, ki si bili oboroženi takrat na našem ozemlju: okupatorji, dor
10	☐ ☐ BonKlanjščekMir...	ič, direktorica Inštituta za narodnostna vprašanja, gospod doc. dr. Damijan Guštin, direktor	Inštituta za novejšo zgodovino	, gospa Tamara Vonta, bivša članica Upravnega odbora EMUNI, gospod dr. Laris Gaiser, n
11	☐ ☐ BonKlanjščekMir...	anosti in umetnosti, predstavniki Inštituta Jožef Štefan, Inštituta za narodnostna vprašanja,	Inštituta za novejšo zgodovino	ter predstavniki Slovenske Matic. Besedo najprej dajem predlagateljem zahteve za sklic s
12	☐ ☐ BonKlanjščekMir...	uta Jožef Štefan, ni? Predstavniki Inštituta za narodnostna vprašanja? Tudi ne. Predstavniki	Inštituta za novejšo zgodovino	? Izvolite. Lepo pozdravljeni. Damjan Guštin, direktor inštituta. Predstavniki Slovenske Mat
13	☐ ☐ RajšpVincenc • ...	di odnosi z Dunajem skozi čas, ki je izšla meseca aprila in sicer v so-izdaji našega inštituta,	Inštituta za novejšo zgodovino	in znanstveno raziskovalnega centra. Knjiga pa nekako sovpada tudi z jubileji patra Ivana T
14	☐ ☐ GuštinDamijan • ...	za kakšne spremembe, ki bi bile na tem področju. Hvala. Najlepša hvala. Izvolite. Direktor	Inštituta za novejšo zgodovino	. Rad bi se vrnil nazaj k samemu gradivu, ki ga imamo danes pred seboj. Naj povem, da tis
15	☐ ☐ KomarDagmar • 2...	rada opozorila. Prvo je re		
16	☐ ☐ KomarDagmar • 2...	in sankcijami. Vse je pov		
17	☐ ☐ FrangežMatevž • ...	danese soočeni s tem, da		
18	☐ ☐ ZalarAleš • 200...	čjučne značilnosti slovens		
19	☐ ☐ ValičAndreja • ...	raziskovalni projekt o nasi		
20	☐ ☐ GrčarSlavko • 2...	eksta žrtev, ki so jih ti pov		

iz konteksta velikanskih morij in zločinov druge svetovne vojne, iz konteksta zločinov, ki so jih nad slovenskim in drugimi jugoslovanskimi narodi uresničevali okupatorji in njihovi sodelavci, iztrgano iz konteksta žrtev, ki so jih ti povzročili. V Sloveniji imamo zelo relevantno raziskavo, ki jo opravlja **Inštitut za novejšo zgodovino** s poimenskimi evidentiranjem vseh žrtev, slovenskih žrtev, mislim na slovenskih tleh tekom 2. svetovne vojne in veste, da je teh žrtev naštetih že 81 ali 82.000. Razen tega še 12.000 potem po vojni pobitih domobrancev. Od teh 82.000 je devet desetlin teh žrtev na račun

XML, TEI in XSLT

HyperText Markup Language (HTML)

- HTML (HyperText Markup Language) : standardni označevalni jezik za oblikovanje in prikazovanje vsebin na spletu.
- Glavni namen: prikazovanje besedil, slik, tabel, povezav in drugih elementov na spletnih straneh.
 - Osnovni gradnik Interneta
- Fiksen nabor označevalcev (označevanje -> markup -> Markup language)
 - Uporablja vnaprej določene oznake (npr. `<p>`, `<h1>`, `<img>`) za strukturo in predstavitev vsebine.



```
<!DOCTYPE html>
<html>
<head>
  <title>Page Title</title>
</head>
<body>
  <!-- Content goes here -->
</body>
</html>
```

Example of a simple HTML document

eXtensible Markup Language (XML)

- XML (eXtensible Markup Language) je prilagodljiv, na besedilu temelječ jezik, zasnovan za shranjevanje in prenos podatkov v strukturirani obliki.
- Podatke organizira v elemente, zaprte v oznakah (npr. <name>Jane Doe</name>), kar omogoča berljivost tako za ljudi kot za računalnike.
 - Ustvarjen za preprosto in prilagodljivo predstavitev hierarhičnih podatkov.
 - Leta 1998 priznan kot W3C standard
- Za razliko od HTML-ja: XML namenjen izključno opisovanju in strukturiranju podatkov (in ne prikazu podatkov na spletni strani)

For example, a simple two-paragraph document could be encoded as follows in XML:

```
<?xml version="1.0" encoding="UTF-8"?>
<document>
  <!-- paragraphs go here -->
  <paragraph number="1">Paragraph one of <title>an XML example</title>.</paragraph>
  <paragraph number="2">Paragraph two of this example.</paragraph>
</document>
```

Example 1. A sample XML document.

Text Encoding Initiative

- TEI (Text Encoding Initiative): standard za označevanje struktur besedilnih podatkov
 - Razvit in standardiziran leta 1994
 - De facto standard na področju digitalne humanistike
 - Večje število modulov (naborov označevalcev) za različne tipe besedil (dramsko besedilo, prozno, korpusna besedila...)
 - TEI Guidelines = 2552 strani
- TEIByExample: <https://www.teibyexample.org/exist/examples/>

```

<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>A sample TEI document</title>
      </titleStmt>
      <publicationStmt>
        <publisher> KANTL</publisher>
        <pubPlace>Ghent</pubPlace>
        <date when="2009" />
      </publicationStmt>
      <sourceDesc>
        <p>No source, born digital</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>This is a sample paragraph, illustrating a <name
        type="organisation">TEI</name> document.</p>
    </body>
  </text>
</TEI>

```

Example 12. A sample (valid) TEI document.

```

<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <text>
    <body>
      <p>This is a sample paragraph, illustrating a
        <orgName>TEI</orgName> document.</p>
    </body>
  </text>
</TEI>

```

Example 13. A sample (invalid) TEI document.

TEI Drama in TEI Speech

- TEI vsebuje 21 glavnih modulov (npr. tei modul: glavni (obvezni) modul, ki diktira osnovno strukturo)
- Za parlamentarne korpuse sta pomembna dva specifična modula:
 - performance (TEI Drama): označevanje izvedbenih del
 - spoken (TEI Speech): reprezentacija govornega jezika
- 2 modula = 2 filozofiji:
 - Parlament kot **performans** (poslanci kot akterji v parlamentu, pripravljeni govori kot skript...)
 - Parlamentarni **govor** (jezik v fokusu, transkripti govornega jezika)

```
<head>Vsebina zapisa seje</head>
<stage type="title">REPUBLIKA SLOVENIJA,
DRŽAVNI ZBOR, KOLEGIJ PREDSEDNIKA
DRŽAVNEGA ZBORA</stage>
<stage type="session">1. seja</stage>
<stage type="date">(29. junij 2018)</stage>
<stage type="chairman">Sejo je vodil mag. Matej Tonin,
predsednik Državnega zbora.</stage>
<stage type="time">Seja se je pričela ob 10.01.</stage>
<sp who="ToninMatej">
  <speaker>PRESEDNIK MAG. MATEJ TONIN:</speaker>
  <p>Dober dan, spoštovani vodje poslanskih skupin.</p>
  <p>Začnjam 1. sejo Kolegija predsednika
Državnega zbora.</p>
  <p>S tem lahko preidemo kar na
  <title>1. TOČKO DNEVNEGA REDA</title>.</p>
</sp>
```

(a) TEI Drama

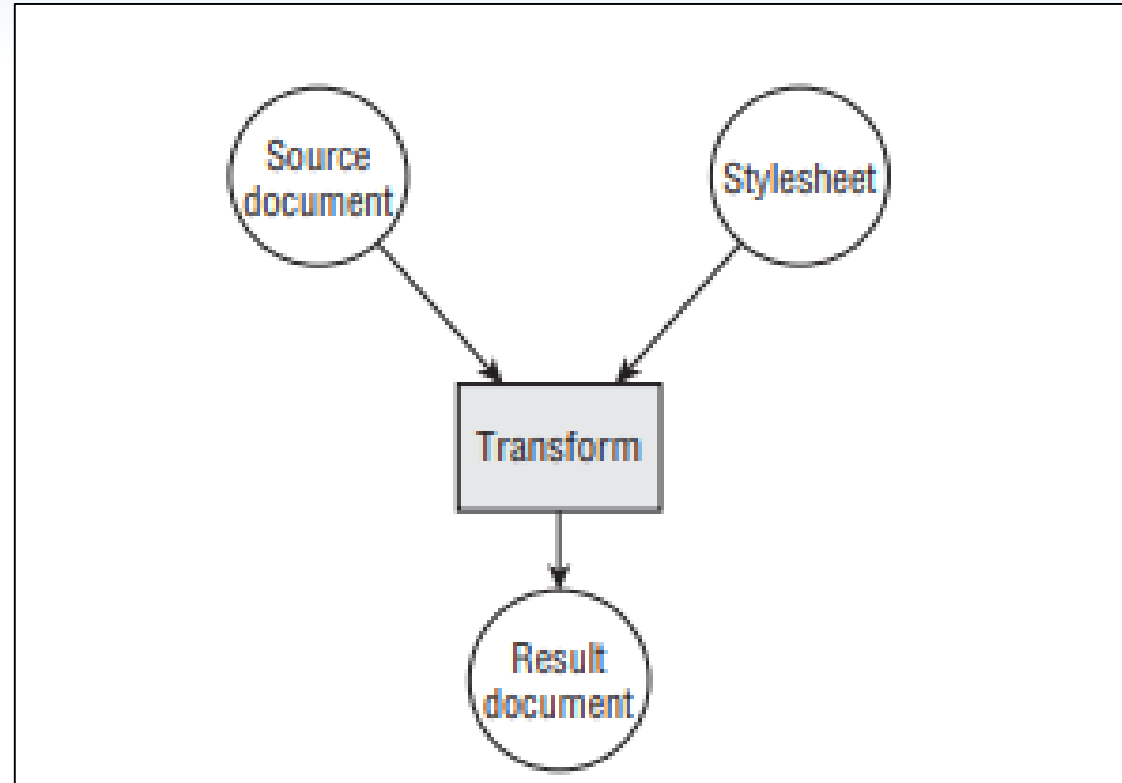
```
<list type="agenda">
  <item xml:id="KPDZ-001-Redna-2018-06-29.toc-item0">
    <title>Pred dnevnim redom</title>
    <ptr target="#KPDZ-001-Redna-2018-06-29.seg1"/>
    <ptr target="#KPDZ-001-Redna-2018-06-29.seg2"/>
    <ptr target="#KPDZ-001-Redna-2018-06-29.seg3"/>
  </item>
  ...
<note type="time">Seja se je pričela ob 10.01.</note>
<note type="speaker">PRESEDNIK MAG. MATEJ TONIN:</note>
<u who="#ToninMatej">
  xml:id="KPDZ-001-Redna-2018-06-29.u1"
  ana="#chair">
    <seg xml:id="KPDZ-001-Redna-2018-06-29.seg1">Dober dan,
    spoštovani vodje poslanskih skupin.</seg>
    <seg xml:id="KPDZ-001-Redna-2018-06-29.seg2">
    Začnjam 1. sejo Kolegija predsednika
    Državnega zbora.</seg>
    <seg xml:id="KPDZ-001-Redna-2018-06-29.seg3">S tem lahko
    preidemo kar na 1. TOČKO DNEVNEGA REDA</seg>
  </u>
```

(b) TEI Speech

Slika 1: Primer razlik med TEI Drama in TEI Speech kodiranimi parlamentarnimi razpravami.

XSLT

- eXtensible Stylesheet Language Transformations
- Transformacija XML dokumentov v druge XML dokumente
 - Tudi v druge formate (HTML, TXT, ...)
- Za razliko od drugih programskih jezikov je XSLT zelo specializiran jezik



Delotok

eXtensible Markup Language (XML) in Text Encoding Initiative (TEI)

Delotok - Faze

1. Digitalizacija in optično prepoznavanje znakov (OCR)/Ekstrakcija podatkov (HTML)
2. Čiščenje in označevanje podatkov
3. Razločevanje govorcev
4. Popis in dodajanje metapodatkov
5. Jezikoslovno označevanje korpusa
6. Objava korpusa (repozitorij + konkordančnik)

Pridobivanje podatkov

- Vir podatkov:
 - Digitalizirani sejni zapisi
 - Ekstrakcija objavljenih zapisnikov sej
 - spletne strani Državnega zbora
 - Format podatkov v HTML
- Način pridobivanja podatkov:
 - Digitalizirani podatki: optično prepoznavanje znakov (OCR)
 - Ekstrakcija: avtomatizacija, skripta za zajem podatkov

Pridobivanje podatkov – Digitalizirano gradivo

- Korpus Kranjskega deželnega zbora – digitalizirano gradivo
- OCR (Optical Character Recognition) – računalniški model, ki omogoča prepoznavanje znakov v digitaliziranem (slikovnem) gradivu
 - Pretvori besedilo na slikah ali skeniranih dokumentih v digitalno besedilo.
 - Podpora več jezikom: različne jezike in pisave, vključno z ročno pisanimi črkami
 - Prepoznavanje struktur (tabele, naslovi...)
- Natančnost modela je odvisna od marsičesa:
 - Jezik gradiva
 - Pisava gradiva
 - Debelina strani, kvaliteta skeniranja
- Strojno učenje modela

Pridobivanje podatkov – HTML gradivo

- Slovenski parlamentarni korpus siParl
- Skripta, ki omogoča pobiranje (scraping) podatkov s spletne strani
- Rezultat: Podatki posameznih sej v formatu XML

```
<text>
  <front>
    <div type="preface">
      <head>DRUŽBENOPOLITIČNI ZBOR SKUPŠČINE REPUBLIKE SLOVENIJE</head>
      <head type="session">1. seja</head>
      <docDate>7. maj 1990</docDate>
      <note type="chairman">Sejo je vodil dr. France Bučar</note>
    </div>
  </front>
  <body>
    <div>
      <note type="time">Seja se je pričela ob 14.20 uri</note>
      <note type="speaker">PRESEDUJOČI DR. FRANCE BUČAR:</note>
      <u who="#BučarFrance" xml:id="DruzPolZb.1990-05-07.s001-01.u1" ana="#chair">
        <seg xml:id="DruzPolZb.1990-05-07.s001-01.seg1">Pričenjam 1. sejo Družbenopolitičnega zbora Skupščine Republike Slovenije, ki je po določbi prvega odstavka 2. člena Poslovnika Družbenopolitičnega zbora Skupščine Republike Slovenije sklicala predsednica zbora, ki je doslej opravljala to funkcijo. Dovolite, da se vam predstavim. Sem France Bučar, kot najstarejši član tega zbora po drugem odstavku 2. člena Družbenopolitičnega zbora Republike Slovenije, zato vodim sejo do izvolitve predsednika zbora. Pri vodenju seje zbora mi bo pomagala dosedanja sekretarka zbora Tina Bitenc-Pengov.</seg>
      </u>
    </div>
  </body>
</text>
```

Vsebina zapisa seje

DRUŽBENOPOLITIČNI ZBOR

SKUPŠČINE REPUBLIKE SLOVENIJE

1. seja

(7. maj 1990)

Sejo je vodil dr. France Bučar

Seja se je pričela ob 14.20 uri

PRESEDUJOČI DR. FRANCE BUČAR: Pričenjam 1. sejo Družbenopolitičnega zbora Skupščine Republike Slovenije, ki je po določbi prvega odstavka 2. člena Poslovnika Družbenopolitičnega zbora Skupščine Republike Slovenije sklicala predsednica zbora, ki je doslej opravljala to funkcijo. Dovolite, da se vam predstavim. Sem France Bučar, kot najstarejši član tega zbora po drugem odstavku 2. člena Družbenopolitičnega zbora Republike Slovenije, zato vodim sejo do izvolitve predsednika zbora. Pri vodenju seje zbora mi bo pomagala dosedanja sekretarka zbora Tina Bitenc-Pengov. Pred nadaljevanjem našega dela sprašujem delegate v zboru, ali so svojo prisotnost javili službi zbora. Menim, da smo sedaj to ugotovili in bi za vsak slučaj še enkrat pregledali in oddali potrdila o izvolitvi. Menim, da je to sedaj opravljeno. Če kdo tega še ni storil, naj to potem stori. bora. Zato prehajamo na predlog dnevnega reda, ki ste ga

Čiščenje in označevanje podatkov

- „Surovi“ podatki (HTML ali OCR podatki) so v veliki večini zelo neurejeni
- Pogosto problem že pri sami sestavi prepisa
- Pogoste napake:
 - Zatipki govorcev (Alenka Bratušek, Alena Bratusek, Alenkra Bratusek)
 - Rekorder v siParl: Zmago Jelinčič Plemeniti (10+ variacij zapisa imena)
 - Neustrezna struktura v originalnih transkriptih
 - Ostanki HTML/drugih problematičnih znakov
- Odstranimo s skripto (in izjemoma minimalnimi ročnimi popravki struktur)

Prepoznavanje govorcev

- Del čiščenja in pravilnega strukturiranja podatkov
- Problemi že v samih podatkih
- Zatipki govorcev (Alenka Bratušek, Alena Bratusek, Alenkra Bratusek)
- Rekorder v siParl: Zmago Jelinčič Plemeniti (10+ variacij zapisa imena)
- Zatipkanih imen v transkripcijah ne popravljamo (integriteta originalnih podatkov)
- Dodamo jim identifikatorje, tako da vse variacije imen vežemo na enega (pravilnega) govorca
 - Alenka Bratušek, Alena Bratusek, Alenkra Bratusek --> Alenka Bratušek
 - <u who=„#Alenka Bratušek“>

Dodajanje metapodatkov

- Dodatne informacije, ki bolje opišejo gradivo/podatke
- Metapodatki govorcev:
 - Ime govorcev
 - Afiliacija govorcev (vlada, politična stranka, menjave afilijacij)
 - Tip seje (redna, izredna)
 - Vrsta govornika (redni/poslanec, gost, predsednik)
- Večjo količino metapodatkov je mogoče dobiti že iz samih transkriptov
- Ostale je potrebno pridobiti iz zunanjih virov:
 - Poročila mandatov DZ
 - Pretekle Vlade RS
 - Drugi viri

Metapodatki o govornih

- ime in priimek govornika
- spol govornika (generirano avtomatično skupaj z ročnim pregledom)
- datum rojstva (če je le ta na voljo)
- afilijacija govornika (s pol. organizacijami)

Metapodatki o organizacijah

- podatki o Državnem zboru xml:id="SK", xml:id="DZ")
- seznam delovnih teles (xml:id="workingBodies")
- seznam političnih strank ter drugih org. (npr. koalicija DEMOS)
- seznam Vlad RS (xml:id="GOV")
- Vsak seznam vsebuje nadaljnje metapodatke:
 - i. ime organizacije (<orgName>)
 - ii. podatek o časovnem obdobju, v katerem organizacija obstaja (<event>)
 - iii. povezava na vir podatkov,
 - iv. seznam "dogodkov"(<listEvent>/<event>)

Jezikoslovno označevanje

- Transkripti/korpus v TEI XML je sicer že sam po sebi precej uporaben, vendar zahteva določeno predznanje za uporabo
- Za lažje pregledovanje korpus naložimo v konkordančnike
- Konkordančnih = orodje za pregledovanje in osnovno obdelavo korpusnih podatkov
- Da korpus dodamo v konkordančnik, je s TEI XML označenim podatkom potrebno dodati tudi jezikoslovne označbe
- Konkordančniki: [NoSketch Engine](#), KonText

```
<s xml:id="KPDZ-redna-1-2014-08-22.seg10.1">
<w xml:id="KPDZ-redna-1-2014-08-22.seg10.1.1"
  msd="UPosTag=NOUN|Case=Nom|Gender=Fem|
  Number=Sing" ana="mte:Ncfsn"
  lemma="hvala">Hvala</w>
<w xml:id="KPDZ-redna-1-2014-08-22.seg10.1.2"
  msd="UPosTag=ADP|Case=Acc"
  ana="mte:Sa" lemma="za">za</w>
<w xml:id="KPDZ-redna-1-2014-08-22.seg10.1.3"
  msd="UPosTag=NOUN|Case=Acc
  |Gender=Fem|Number=Sing" ana="mte:Ncfsa"
  lemma="beseda" join="right">besedo</w>
<pc xml:id="KPDZ-redna-1-2014-08-22.seg10.1.7"
  msd="UPosTag=PUNCT" ana="mte:Z">.</pc>
<linkGrp type="UD-SYN" targFunc="head argument">
  <link ana="ud-syn:amod"
    target="#KPDZ-redna-1-2014-08-22.seg10.1.2
    #KPDZ-redna-1-2014-08-22.seg10.1.1"/>
  <link ana="ud-syn:root"
    target="#KPDZ-redna-1-2014-08-22.seg10.1
    #KPDZ-redna-1-2014-08-22.seg10.1.2"/>
  <link ana="ud-syn:cc"
    target="#KPDZ-redna-1-2014-08-22.seg10.1.4
    #KPDZ-redna-1-2014-08-22.seg10.1.3"/>
</linkGrp>
</s>
```

Slika 3: Primer jezikoslovno označenega stavka "Hvala za besedo."

Attributes **ana** and **msd** both encode the morphosyntactic characteristics of the word in focus (in blue). The **ana** attribute contains Slovene-specific tags according to MULTEXT-East Specifications, while the **msd** attribute contains the Universal Dependency tagset which greatly simplifies cross-lingual comparison.

The attribute values Q and PART both stand for *particle*. The difference between the two tags is in the tagset used.

The **word form** also known as a token from the running text in the corpus.

The attribute **lemma** indicates the basic word form of the token, that is of the word in focus (in blue).

The corpus also encodes syntactic parses and named entities but since they are not used in this tutorial they were omitted from this illustrative example.

```
<s>
<w ana="mte:Q" msd="UpoSTag=PART" lemma="zlasti">Zlasti</w>
<w ana="mte:Sg" msd="UpoSTag=ADP|Case=Gen" lemma="glede">glede</w>
<w ana="mte:NcmSg" msd="UpoSTag=NOUN|Case=Gen|Gender=Masc|Number=Sing"
lemma="nadzor">nadzora</w>
<w ana="mte:Va-r3s-n"
msd="UpoSTag=AUX|Mood=Ind|Number=Sing|Person=3|Polarity=Pos|Tense=Pres|VerbForm=Fin"
lemma="biti">je</w>
<w ana="mte:Pd-fsn" msd="UpoSTag=DET|Case=Nom|Gender=Fem|Number=Sing|PronType=Dem"
lemma="ta">ta</w>
<w ana="mte:NcfSg" msd="UpoSTag=NOUN|Case=Nom|Gender=Fem|Number=Sing"
lemma="stvar">stvar</w>
<w ana="mte:Rgp" msd="UpoSTag=ADV|Degree=Pos" lemma="zelo">zelo</w>
<w ana="mte:AgpfSg" msd="UpoSTag=ADJ|Case=Nom|Degree=Pos|Gender=Fem|Number=Sing"
lemma="kočljiv">kočljiva</w>
<pc ana="mte:Z" msd="UpoSTag=PUNCT">.</pc>
</s>
```

One pair of the opening and closing structural tags which, in our case, indicate punctuation (**pc**), word (**w**) or sentence (**s**).

Objava korpusa

- Korpus (oziroma označene podatke) je potrebno validirati
- Validacija poteka na nivoju označevanja
- Seveda pa to ne pomeni, da so vsi podatki 100% urejeni in ni možnosti za napake
- Korpus nato še pretvorimo v druge formate (plain text, CoNNL-U format, vert):
 - CoNNL-U:
 - Vert format: format podatkov za konkordančnike
- Objava korpusa na CLARIN.SI (ali drugih) repozitoriju
- Objava korpusa na konkordančnih (3 produkcijski strežniki, 3 beta strežniki)
- Korpus je na voljo uporabnikom

phrase **Parlament je** • 1,184
4.34 per million tokens • 0.00043%

☐ Details

	sentence
1 ☐ ⓘ RajićBranislav ...	In drugo, parlament je predstavnik ljudi in mi smo tukaj tudi, da smo neka korelacija z nekim narodom s katerim smo v bližini ali bliskosti že stoletja.
2 ☐ ⓘ TrčekFranc • 20...	Parlament je tudi neka procedura, vlagajo se takšni in drugačni zakon, dobri, slabi.
3 ☐ ⓘ GregorčičMonika...	Vemo, nekaj parlamentom v Evropi je že sprejelo podobne dokumente in pa konec koncev tudi Evropski parlament je že obsodil agresorsko dejanje.
4 ☐ ⓘ ŽnidaričMojca •...	Tudi vi imate možnost, vsi lahko kandidirate, tako da tudi če pridete mladi v parlament je vaš glas seveda toliko bolj sliši.
5 ☐ ⓘ PogačnikMarko •...	In parlament je , bi rekel, zakonodajna veja oblasti.
6 ☐ ⓘ ŠkrinjarMojca •...	Ce vse te stvari vpletamo v svoje razprave, kajti parlament je predmet razprave ne pa operative, potem si lahko od tega da res zagotovo bo veliko mest v domovih za ostarele, ker bomo vsi prej umrli, preden nas bi kakšen ukrep dosegel.
7 ☐ ⓘ SukičNataša • 2...	Da bi prezrli dejstvo, da če se zbere cel parlament je 47 poslank in poslanecev proti frekingu.
8 ☐ ⓘ KociperMaša • 2...	<u>Parlament je</u> po teoriji polje kontroliranega konflikta.
9 ☐ ⓘ PrebilNik • 202...	Ne glede na to, če imamo na določeno politično situacijo povsem drugačno stališče, je to prostor – to tukaj, parlament je soočamo.
10 ☐ ⓘ GolobRobert • 2...	Baje, da ista oseba, ki sprejema direktorje distribucij v parlamentu tako, da tokrat ni hodil v parlament je samo po telefonu.
11 ☐ ⓘ StarovićVojko •...	Parlament je najnižje na lestvic.
12 ☐ ⓘ SukičNataša • 2...	Tukaj je resna debata, so resni pomisleki in parlament je prostor, kjer se takšne pomisleke izrazi in to ni nobena žalitev, že gre malo počasi na živce, ker vsak to potem začne opletati, ko ni argumenta postanete žrtve.
13 ☐ ⓘ SukičNataša • 2...	Parlament je avtonomen.

Raziskovanje korpusov

Slovenski parlamentarni korpus (1990 – 2022) siParl 4.0

Parlamentarni korpus Kranjskega deželnega zbora

Korpus slovenske periodike (1771-1914) sPeriodika 1.0

Seje delovnih teles - Izbrani zapis seje (v pregledu)

Polni naziv telesa - št. in vrsta seje

26 - Komisija za odnose s Slovenci v zamejstvu in po svetu - 017. Redna

Datum zasedanja

7. 4. 2022

Vsebina zapisa seje (v pregledu)

REPUBLIKA SLOVENIJA
DRŽAVNI ZBOR
KOMISIJA ZA ODNOSE S SLOVENCMI V ZAMEJSTVU IN PO SVETU
17. seja

(7. april 2022)

Sejo je vodil Dušan Šiško, predsednik komisije.

Seja se je začela ob 11.04.

1. TRAK: (DAG)

PREDSEDNIK DUŠAN ŠIŠKO: Spoštovani! Začenjam 17. sejo Komisije za odnose s Slovenci v zamejstvu in po svetu. Lepo pozdravljeni vsi prisotni! Dovolite mi, da vas na začetku seje vljudno prosim, da glede na zdravstvena napotila tekom celotne seje vzdržujete in spoštujete predpisano medsebojno razdaljo in nosite zaščitne maske.

Članice in člane komisije prosim, da preverijo, ali imajo s sabo identifikacijsko kartico.

Obveščam vas, da so zadržani in se seje ne morejo udeležiti naslednje članice in člani komisije: Jurij Lep, Marko Bandelli, Tadeja Šuštar Zalar, Lidija Divjak Mirnik, Igor Peček, Felice Žiža in Suzana Lep Šimenko.

2. TRAK: (DAG) – 11.05

(nadaljevanje) Na seji ni nadomestnih članov.

PREDSEDNIK DUŠAN ŠIŠKO: Hvala.

Želi kdo od ostalih vabljenih predstavnikov ministrstva besedo? (Da.) Ja, gospa, izvolite.

7. TRAK: (DAG) – 11.30

GOSPA_____: Hvala lepa za besedo.

Spoštovane poslanke in poslanci, spoštovani gostje!

Jaz bi seveda rekla, da z Uradom Vlade Republike Slovenije za Slovence v zamejstvu in po svetu tesno sodelujemo in tako tudi pomagamo pri nekaterih zadevah, kar se konkretno tiče Ministrstva za kulturo. Mogoče bi samo še izpostavila, da preko Javne agencije za knjigo, Javnega sklada za kulturne dejavnosti se nameni okrog 130 tisoč evrov, preko Slovenskega kulturno-informacijskega centra, ki odlično sodeluje s Pavlovo hišo, člen 7, prav tako podpira in sodeluje pri določenih kulturnih projektih. Kot je bilo že omenjeno, tudi pri obnovi Pangerčeve domačije ter na vseh področjih, ki jih pokriva Ministrstvo za kulturo. Tako je to sodelovanje z uradom odlično in kot vidimo po rezultatih, tudi pripomore k večji učinkovitosti na teh področjih.

Hvala lepa.

PREDSEDNIK DUŠAN ŠIŠKO: Hvala, gospa.

Prosil bi vas samo – in tudi za bodoče prosim vse –, da se predstavite zaradi zapisnika.

GOSPA_____: / mikrofoni ni vključen/

PREDSEDNIK DUŠAN ŠIŠKO: Naprej, gospod...

VLADIMIR ČELIGOJ: Hvala lepa.

(Vladimir Čeligoj. Ministrstvo za kmetištvo, gozdarstvo in prehrano.)

root h5 span

```
1 <root>
2 <h5 class="label"><span>Vsebina zapisa seje (v pregledu)</span></h5><div align="center">
3 <p><b><font size="4">REPUBLIKA SLOVENIJA</font></b><br/>
4 <b><font size="4">DRŽAVNI ZBOR</font></b><br/>
5 <b><font size="4">KOMISIJA ZA ODNOS S SLOVENC V ZAMEJSTVU IN PO SVETU</font></b><br/>
6 <b><font size="4">17. seja</font></b><br/>
7 <br/>
8 <b><font size="4">(7. april 2022)</font></b></p></div><br/>
9 <b><font size="4">Sejo je vodil Dušan Šiško, predsednik komisije.</font></b><br/>
10 <b><font size="4">Seja se je začela ob 11.04. </font></b><br/>
11 <br/>
12 <b><font size="4">1. TRAK: (DAG) </font></b><br/>
13 <br/>
14 <b><font size="4">    PREDSEDNIK DUŠAN ŠIŠKO: </font></b><font size="4">Spoštovani! Začenjam 17. sejo Komisije za odnose s Slovenci v zamejstvu in po svetu. Lepo pozdravljeni vsi prisotn
15 <font size="4">    Dovolite mi, da vas na začetku seje vljudno prosim, da glede na zdravstvena napotila tekom celotne seje vzdržujete in spoštujete predpisano medsebojno razdaljo in nosi
16 <font size="4">    Članice in člani komisije prosim, da preverijo, ali imajo s sabo identifikacijsko kartico. </font><br/>
17 <font size="4">    Obveščam vas, da so zadržani in se seje ne morejo udeležiti naslednje članice in člani komisije: Jurij Lep, Marko Bandelli, Tadeja Šuštar Zalar, Lidija Divjak Mirnik,
18 <br/>
19 <b><font size="4">2. TRAK: (DAG) – 11.05</font></b><br/>
20 <br/>
21 <b><font size="4">    </font></b><font size="4">(nadaljevanje) Na seji ni nadomestnih članov. </font><br/>
22 <font size="4">    Glede na to, da je seznam vabljenih obširen, ga ne bom bral, saj je razviden iz sklica seje. Svojo udeležbo je opravičil predsednik Skupnosti koroških Slovencev in Slo
23 <br/>
24 <font size="4">    Prehajamo na določitev dnevnega reda seje. </font><br/>
25 <font size="4">    S sklicem seje ste prejeli predlog dnevnega reda. Ker v poslovniškem roku nisem prejel predlogov za njegovo spremembo, je določen, kot je bil predlagan s sklicem seje.
26 <br/>
27 <font size="4">    Prehajamo na </font><b><font size="4">1. TOČKO DNEVNEGA REDA – SEZNANITEV S POROČILOM VLADE REPUBLIKE SLOVENIJE O REALIZACIJI NAČRTOVANIH DEJAVNOSTI IN SREDSTVIH ZA SL
28 <b><font size="4">    </font></b><font size="4">Gradivo k tej točki, ki smo ga prejeli od Vlade Republike Slovenije dne 24. 3. 2022, je objavljeno na spletnih straneh Državnega zbora s s
29 <font size="4">    Za uvodno predstavitev obravnavanega gradiva dajem besedo državnemu sekretarju dr. Valentinčiču. Izvolite. </font><br/>
30 <br/>
31 <font size="4">    </font><b><font size="4">DR. DEJAN VALENTINČIČ: </font></b><font size="4">Hvala lepa, gospod predsednik. </font><br/>
32 <font size="4">    Spoštovani poslanke in poslanci, člani komisije ter vsi ostali gostje, predvsem naši rojaki!</font><br/>
33 <font size="4">    V skladu s sklepom te komisije iz leta 2019 Vlada enkrat letno, običajno do konca meseca maja, pripravi poročilo o dejavnostih in sredstvih, namenjenih za Slovence izv
34 <font size="4">    Kot vsi vemo, je odnos do rojakov izven Slovenije ustavna obveza Republike Slovenije. Urad za Slovence v zamejstvu in po svetu je tu centralni in koordinativni organ,
35 <font size="4">    Lahko rečemo, da vse aktivnosti države in tudi našega urada so nekako skoncentrirane v ohranjanje in krepitev slovenske identitete v najširšem pomenu izven Slovenije,
36 <br/>
37 <b><font size="4">3. TRAK: (DAG) – 11.10</font></b><br/>
38 <br/>
39 <b><font size="4">    </font></b><font size="4">(nadaljevanje) Tu je sedaj prišlo do tretje gradbene faze, to financira urad skupaj z Ministrstvo za kmetijstvo. Zgledujemo se po dobrem z
```

Nadaljevanje ednajste seje
deželnega zbora Ljubljanskega

dné 15. oktobra 1874.

Nazoči: Prvosednik: Deželni glavar dr. Friderek vitez Kaltenegger. — Vladini zastopnik: Deželne vlade načelnik, dvorni svetovalec Bohuslav vitez Widmann. — Vsi članovi razun: knezoškof dr. Widmer in: dr. Bleiweis, grof Thurn, Lavrenčič Andrej, Kramarič.

Dnevni red:

Nadaljevanje dopoldanjega.

Obseg: Odgovor na interpelacijo gosp. poslanca viteza Langerja gledé po toči poškodovanih Dolenjcev, — poročilo o peticijah, — dalje glej dnevni red.

Seja sa začne o 15. minuti čez 6. uro.

Fortsetzung der eilften Sitzung
des Landtages zu Laibach

am 15. Oktober 1874.

Anwesende: Vorsitzender: Landeshauptmann Dr. Friedrich Ritter v. Kaltenegger. — Vertreter der k. k. Regierung: Regierungsleiter Hofrath Bohuslav Ritter v. Widmann. — Sämmtliche Mitglieder mit Ausnahme von: Fürstbischof Dr. Widmer und: Dr. Bleiweis, Graf Thurn, Lavrenčič Andreas, Kramarič.

Tagesordnung:

Fortsetzung der vormittägigen.

Inhalt: Beantwortung der Interpellation des Herrn Abgeordneten Ritter v. Langer den Nothstand in Unterfrain betreffend. — Berichte über Petitionen. — Weiter siehe Tagesordnung.

Beginn der Sitzung 6 Uhr 15 Minuten.

```

<body>
  <div type="debateSection">
    <meeting xml:lang="sl">Mahanje eñajste seje deželnega zbora Ljubljanskega dne 15. oktobra 1874.</meeting>
    <meeting xml:lang="de">Fortsetzung der eiftten Sitzung des Landtages zu Laiöae m 15. Oktober 1874.</meeting>
    <list type="attendance" xml:lang="sl">
      <head>Navzoči</head>
      <label>Prvosednik</label>
      <item>Deželni glavar dr. Friderik vitez Kaltenegger.</item>
      <label>Vladini zastopnik</label>
      <item>Deželne vlade načelnik, dvorni svetovalec Bohuslav vitez Widmann.</item>
      <label>Vsi članovirazun</label>
      <item>knezoškof dr. Widm er in: dr. Blei weis, grof Thurn, Lavrenčič Andrej, Kramarič.</item>
    </list>
    <list type="attendance" xml:lang="de">
      <head>Anwesende</head>
      <label>Vorsitzender</label>
      <item>Landeshauptmann Dr. Friedrich Ritter v. Kaltenegger.</item>
      <label>Vertreter der t,!. Regierung</label>
      <item>Regiernngsleiter Hofrath Bohuslav Ritter v. Widmann.</item>
      <label>Sämmtliche Mitglieder mit Ausnahme von</label>
      <item>Fürstbischof Dr. Widmer und: Dr. Bleiweis, Graf Thurn, Lavrenčič Andreas, Kramarič.</item>
    </list>
    <list type="agenda" xml:lang="sl">
      <head>Dnevni red</head>
      <item n="1">Nadaljevanje dopoldanjega.</item>
    </list>
    <list type="content" xml:lang="sl">
      <head>Obseg</head>
      <item n="1">Odgovor na interpelacijo gosp. poslanca viteza Langer-ja glede po toči poškodovanih Dolenjcev,</item>
      <item n="2">poročilo o peticijah,</item>
      <item n="3">dalje glej dnevni red.</item>
    </list>
    <list type="agenda" xml:lang="de">
      <head>Tagesordnung</head>
      <item n="1">Fortsetzung der vormittägigen.</item>
    </list>
    <list type="content" xml:lang="de">
      <head>Inhalt</head>
      <item n="1">Beantwortung der Interpellation des Herrn Abgeordneten Ritter v. Langer den Nothstand in Nnterkrain betreffend.</item>
      <item n="2">Berichte über Petitionen.</item>
      <item n="3">Weiter siehe Tagesordnung.</item>
    </list>
    <note type="time" xml:lang="sl">Seja sa začne o 15. minuti čez 6. uro.</note>
    <note type="time" xml:lang="de">Beginn der Sitzung 6 Uhr 15 Minuten.</note>
    <note type="speaker">Landeshauptmann:</note>
    <u>
      <seg xml:id="DZK_1874-10-15_15_11-2_seg1" n="2">
        <s xml:lang="de" xml:id="DZK_1874-10-15_15_11-2_seg1.s1">Da wir in beschlußfähiger Anzahl vorhanden sind, so erkläre ich die früher unterbrochene Sitzung für
      </s>
      </seg>
    </u>
    <note type="speaker">Neqierunagskeiter Kofrath stiff er v. ZSidurann:</note>
  </div>

```

Kakvo vsek dan, izvršeni poučevalni dnevi po pravnih, ter velja po postopku prejemana za avstro-ogorske dežele za eno leto 16 gld., za pol leta 8 gld., za četrto leto 4 gld. — Za Ljubljano brez posiljanja na dom za celo leto 13 gld., za četrto leto 3 gld. 30 kr., za en mesec 1 gld. 10 kr. Za posiljanje na dom se računa 10 kr. za mesec, 30 kr. za četrto leto. — Za tuje dežele tolko več, kolikor poštena izkaza. — Za gospode udeležne na ljudskih šolah in za dijake velja znižana cena in sicer: Za Ljubljano za četrto leto 2 gld. 50 kr., po postopku prejemana za četrto leto 8 gld. — Za ostanila se plačuje od četnostopne vrednosti — vsakega, ki se je umaknil kmalu po 10 kr., če se dvakrat in 4 kr. če se tri-ali večkrat toka.

Dopis naj se vrne vskraj! Rokovani se ne smejo. Uredništvo. — Posredovanje v "Narodni shod" za 10 gld. — Posredovanje v "Narodni shod" za 10 gld.

Opravišev, na katero naj se blagovoljijo poslati naročnice, reklamacije, ostanila, i. j. administrativne reči, je v "Narodni shod" v Kolovosovi 18.