**Pretvorba RTF v TEI P5  
Navodila in testno besedilo**

Tomaž Erjavec  
Odsek za tehnologije znanja  
Institut »Jožef Stefan«  
[tomaz.erjavec@ijs.si](mailto:tomaz.erjavec@ijs.si)

2013-01-14

Na <http://nl.ijs.si/e-zrc/rtf2tei/> najdete servis za pretvorbo Word datotek v [TEI P5](http://www.tei-c.org/Guidelines/P5/). Spletna stran vsebuje tudi navodila za uporabo same spletnega servisa, pričujoče besedilo pa vsebuje navodila za oblikovanje Word datotek za servis in obenem služi kot testno besedilo, ki ga lahko pošljete servisu, da vam vrne ustrezno stavljen rezultat v XML oz. HTML.

# Splošna navodila

1. Besedilo naj uporablja naslove (Heading 1, Heading 2, itn.), saj bo le tako pravilno razčlenjeno na razdelke (<div>) v TEI. Besedilo pred prvim naslovom gre v element <front>.   
   Namig: Kaj točno so naslovi lahko preverite preko »prikaži ¶«. Pravilni naslovi imajo na začetku vrstice napisan kvadratek, ▪   
   Pozor: »Naslovi« lahko pomotoma postanejo tudi prazne vrstice ali pa prelomi strani, kar preverite skozi kvadratek. Ali se naslovi res pravilno pretvorijo, preverite v kazalu generiranega HTML, pa tudi v sami datoteki Word lahko naredite kazalo.
2. Večina formatiranja (font, velikost, barva, vrsta podčrtaja, razmaki, zamik odstavkov, itd.) se pri pretvorbi izgubi.
3. Za bolj specifične elemente TEI je potrebno uporabiti vnaprej določene sloge, predstavljeni v naslednjih razdelkih. Slogi uporabljajo različne barve in podčrtanja, da se vizualno ločijo od drugega besedila. Ime sloga je tipično enako imenu elementa TEI v katerega se pretvori.   
   Namig: Pri delu s slogi pomaga, če vklopite podokno s slogi (Oblika – Slogi in oblikovanje).
4. Pretvorba ne ohrani paginacije, razen v primeru, ko vstavimo ročni prelom strani, ki se prevede v <pb/>; ta je primeren kot nosilec kazalke na faksimile.
5. Odstavek (Enter) in ročni prelom vrstice (Ctrl-Enter) se prevedeta v različna elementa (<p> in <lb/>).   
   Namig: Znake za odstavke, prelom strani in vrstic lahko vidite, če vklopite »prikaži ¶« v Wordu, oz. skozi generirani HTML.
6. Tabele in opombe naj bi se v pretvorbi ohranile, čeprav bodo verjetno drugače stavljene v HTML, kot pa so bile v Word.
7. Vrnjeno besedilo v TEI vsebuje tudi kolofon TEI - tega je treba v TEI ročno popraviti in izpolniti.

Tule je <pb/>, prelom strani!

# Uporaba slogov

V tem poglavju so podani primerov uporabe slogov, ki so definirani posebej za pretvorbo v TEI; te sloge spoznamo po tem, da se vsi začnejo z nizom »tei:«. Najprej opišemo sloge, ki so definirani na ravni odstavkov, nato pa take, ki jih uporabljamo znotraj besedila za označevanje posameznih besed ali fraz.

## Slogi, definirani na ravni odstavkov

Spodnji slogi se vedno uporabijo za označevanje celotnih odstavkov.

### Avtor: *tei:docAuthor*

Matija Ogrin

Avtor se preslika v <docAuthor>.

Pozor: avtor se sme pojaviti samo neposredno po naslovu poglavja!

### Bibliografski seznam: *tei:listBibl*

Bibliografski seznam se preslika v <listBibl>, s postavkami <bibl>:

Matija Ogrin. Uvod v Brižinske spomenike. Acta Analytica 12/4, 1994, str. 3-120.

Matija Ogrin. Uvod v zbrana dela. Acta Analytica 12/5, 1995, str. 13-15.

### Citirano besedilo: *tei:cit*

Citirano besedilo, <cit>, lahko vsebuje več odstavkov. V spodnjem primeru je na koncu podan tudi vir s slogom bibl, glej naslednje poglavje:

Kar so govorili, to so mislili.

Česar niso, niso. Wittgenstein, 1931.

### Dramski govor: *tei:sp*

Govori v gledališkem delu; <sp> (posamezni govori) so ločeni s praznim odstavkom, prvi odstavek govora se preslika v <speaker> (govorec), vsi naslednji pa v <l> (vrstica):

Romeo

Biti ali ne biti,

to je sedaj vprašanje.

Julija

Kakšno vprašanje le?

Bodi, pa je!

### Pesem: *tei:lg*

Pesmi se preslikajo v serijo <lg> (kitica); če ima pesem več kitic, naj bodo med sabo ločene s praznim odstavkom. Vsak odstavek znotraj kitice se preslika v <l> (vrstica):

Prva vrstica prve kitice pesmi,

druga vrstica prve kitice pesmi.

Prva vrstica druge kitice pesmi,

druga vrstica druge kitice pesmi.

### Kustoda: *tei:catch-word*

Čeprav se slog imenjuje »tei:catch-word«, tega elementa v TEI ne bomo našli. TEI namreč vsebuje element [fw](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-fw.html) (forme work), vendar je pomen tega elementa širši kot samo kustoda, zato je mora biti za oznaka dodatno kvalificirana z atributom za tip, ki ima za kustodo vrednost »catch«, torej »<*fw type="catch">.* Primer uporabe:

Kust-

### Slike: *tei:figure*

Uporaba sloga povzroči, da se v dokument TEI vpiše element <figure>, ki vsebuje podnapis k sliki (<head>). V datoteki TEI bo nato na roke potrebno vstaviti še element <graphic>, ki bo vseboval kazalko na dejansko slikovno datoteko, npr. *<graphic url="slika.jpg"/>*. Primer uporabe:

Naslov slike

## Slogi, definirani znotraj odstavka

* Sloga za <date> (datum) in <name> (ime): 21. rujna 1845 se je rodila Jožica Širca.
* Slog za <foreign> (tujejezični citat): Quis custodiet ipsos custodes?
* Slog za <q> (besedilo v narekovajih, manj določni navedki): In rekla si sonce!
* Slog za <quote> (določen navedek iz drugega teksta): Prešeren: O Vrba!
* Slog za <bibl> (bibliografska enota): kot piše v Ogrin, 2009
* Slog za <gap> (vrzel): velika packa
* Slog za <damage> (poškodovano): Sloni in druge živali.
* Slog za <surplus> (odvečno): Sloni in drudruge živali.
* Slog za <supplied> (vstavil urednik): Sloni in druge živali.
* Slog za <unclear> (nejasno): Sloni in druge živali.

### Slogi za parne elemente

Če sta prisotna oba elementa sledečih parov in med njima ni besedila, ju združi <subst> (nadomestno) oz. <choice> (izbira):

* Sloga za <del> (prečrtano) in <add> (dodano): prečrtano dodano.
* Sloga za <sic> (napaka) in <corr> (uredniški popravek): napaka popravek.
* Sloga za <abbr> (okrajšava) in <expan> (razširjeno): okrajšava razširjeno.
* Sloga za <orig> (izvorna oblika) in <reg> (regularizirano): izvorno regularizirano.

# Uporaba standardnega oblikovanja

V tem poglavju pokažemo še uporabo nekaterih standardnih načinov za oblikovanje besedila v urejevalniku Word, ki se tudi upoštevajo pri pretvorbi:

* Opomba pod črto.[[1]](#footnote-1)
* [Spletna povezava](http://nl.ijs.si/e-zrc/rtf2tei/).

Spodnji primeri se preslikajo v element <hi>, pri čemer je način oblikovanja izražen kot vrednost atributa @rend.

*Pomembno*: spremenjeni tisk ne sme zajemati celotnega besedila odstavka oz. alineje; vsaj končno ločijo ali presledek mora biti v privzetem zapisu.

* Primer <hi rend="italic"> *Ležeči tisk.*
* Primer <hi rend="ul"> Podčrtano besedilo.
* Primer <hi rend="bold"> **Krepko besedilo**.

Primer tabele:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Prva vrstica | 1. stolpec | 2. stolpec | 3. stolpec | 4. stolpec |
| Druga vrstica | 1. stolpec | 2. stolpec | 3. stolpec | 4. stolpec |

## Uporaba kazala

Za daljša besedila je lahko koristno uporabiti Wordovo kazalo (Vstavi / Reference / Navzkrižne reference / Kazalo), ki nam pokaže, ali je besedilo pravilno strukturirano po poglavjih oz. razdelkih. Kazalo se v TEI ne ohrani, vendar se ponovno avtomatsko generira v obliki HTML.

Kazalo za pričujoče besedilo:

[1. Splošna navodila 1](#_Toc317600600)

[2. Uporaba slogov 2](#_Toc317600601)

[Slogi, definirani na ravni odstavkov 2](#_Toc317600602)

[Avtor: *tei:docAuthor* 2](#_Toc317600603)

[Bibliografski seznam: *tei:listBibl* 2](#_Toc317600604)

[Citirano besedilo: *tei:quote* 2](#_Toc317600605)

[Dramski govor: *tei:sp* 2](#_Toc317600606)

[Pesem: *tei:lg* 3](#_Toc317600607)

[Kustoda: *tei:catch-word* 3](#_Toc317600608)

[Slike: *tei:figure* 3](#_Toc317600609)

[Slogi, definirani znotraj odstavka 3](#_Toc317600610)

[Slogi za parne elemente 3](#_Toc317600611)

[3. Uporaba standardnega oblikovanja 4](#_Toc317600612)

[Uporaba kazala 4](#_Toc317600613)

1. Primer opombe pod črto. [↑](#footnote-ref-1)