

## TNA ACTIVITY REPORT

## Building an R Package and Web Application to Interact Within a Digital Ecosystem for Literary Studies

Author: Pozdniakov, Ivan Current position: Independent researcher Affiliation: none Host institution: University of Potsdam, Germany Mentor(s): Prof. Peer Trilcke Period of stay: 2022-04-15 to 2022-06-14

## Overview of my TNA stay

My stay at the University of Potsdam, Germany from April 15th, 2022 to June 14th, 2022 was dedicated to the development of "rdracor", a package in the programming language R, that could be used to analyze the literary data provided by the Drama Corpora Project (DraCor). DraCor is a prototype of the Programmable Corpora concept that is explored within the CLS Infra Project in Work Package 7 led by the University of Potsdam and provides data on theater plays in several European languages. Every play is encoded using TEI that helps to analyze different plays and the whole corpora by access to the project's API. I am excited by insights given by the distant reading approach and I believe that providing a complex but easy-to-use infrastructure for computational research in Literary Studies will not only make it less challenging and faster but also will lead to new hypotheses and ways to test them. Thus my particular interest was to develop a package that could be used by the R community within CLS in native way to the programming language and build on commonly used data structures, like datatables. For the package I could build upon an early prototype that I created in the course of my involvement in a student research group at my University some years ago.



CLS INFRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004984



During my stay in Potsdam, in several meetings with my TNA mentor, Prof. Peer Trilcke and team members of CLS INFRA project and the Potsdam Network for Digital Humanities we went through the functions provided by the DraCor API and discussed which of them needed to be represented in the package. I then updated some of the legacy code to the current version of API, but also focused on making the code more independent of possible future changes of the API by implementing a generic function that could send requests not only to the current production version of the platform, but also to local instances running in a Docker container.

## First impressions of rdracor (still in development)

The main feature of rdracor is that it translates the DraCor data to R data objects which are familiar to R users and convenient to process in the R environment. For example, the metadata for all plays in each corpus can be retrieved as a dataframe with a single 'get\_dracor' command:

d2* ×	veche	r_df ×	Untitled3* × O example.R × revizor_df ×	examp	le2.R × 0	main.R ×	Cls_infra_showcase.R ×	🗌 rus_corpus × 🛛 🔉 👝 🗖		
4	1218	Filter C	lo s: ≪ < 0 - 50 > ≫					Q,		
*	corpus 🔅	id ÷	playName	yearNormalized *	title			subtitle		
1	rus	rus000001	turgenev-zavtrak-u-predvoditelja	1849	Завтрак у і	предводителя		NA		
2	rus	rus000002	turgenev-vecher-v-sorrente	1852	Вечер в Со	орренте		Сцена		
3	rus	rus000003	turgenev-razgovor-na-bolshoj-doroge	1851	Разговор н	на большой до	роге	Сцена		
4	rus	rus000004	turgenev-provintsialka	1851	Провинци	алка		Комедия в одном действии		
5	rus	rus000005	turgenev-neostorozhnost	1843	Неосторож	кность		NA		
6	rus	rus000006	turgenev-nahlebnik	1857	Нахлебник	к		Комедия в двух действиях		
7	rus	rus000007	turgenev–holostjak	1849	Холостяк			Комедия в трех действиях		
8	rus	rus000008	turgenev-gde-tonko-tam-i-rvetsja	1848	Где тонко,	, там и рвётся		Комедия в одном действии		
9	rus	rus000009	tolstoy-vlast-tmy	1887	Власть тьм	ны, или 'когото	к увяз, всей птичке пропасть'	Драма в пяти действиях		
10	rus	rus000010	sumarokov-semira	1751	Семира			Трагедия		
11	rus	rus000011	sumarokov-horev	1747	Хорев			Трагедия		
12	rus	rus000012	sumarokov-dimitrij-samozvanets	1771	Димитрий Самозванец		Трагедия			
13	rus	rus000013	sukhovo-kobylin-svadba-krechinskogo	1856	Свадьба Кречинского		Комедия в трех действиях			
14	rus	rus000014	sukhovo-kobylin-smert-tarelkina	1869	Смерть Та	релкина		Комедия-шутка в трех дейст		
15	rus	rus000015	sukhovo-kobylin-delo	1861	Дело			Драма в пяти действиях		
Show	ing 1 to 15 c	f 212 entries	50 total columns	1017		****		1		
-										
Con	sole Term	inal × Jot	bs ×							
~/R	/parse_tei/	n(v df)								
+	recur	n(x_u)								
+ 1		Cuduce								
> 1	ibrary	Charace								
> 1	ibrary	(tiayve	erse)							
> 1	'us_cor	pus <-	get_dracor(corpus = "rus")							
Err	or in	curl::c	curl_fetch_memory(url, handl	e = handle)	:					
(	ould n	ot resc	olve host: dracor.org							
> 1	ibrary	(rdrace	or)							
> 1	ibrary	(rdrace	or)							
> 1	ibrary	(tidyve	erse)							
> 1	us_cor	pus <-	<pre>get_dracor(corpus = "rus")</pre>							
> View(rus_corpus)										
>										

Moreover, the TEI source for each play can also be retrieved as a line-by-line dataframe using Rdracor package. I believe this to be the most considerable advantage of the package,



since it converts an XML (which is usually not easy to process with data analysis tools) to a dataframe without any loss of information. In this dataframe, we still have all the data about who said what and in which order.

1 💭   🚈   🌱 Filter					Q,
text 🗘		type_attributes	id 🗘	who	segment
1 Комната в доме городничего.	stage	NA	1	NA	Действие первое
2 Городничий, попечитель богоугодных заведений, с	stage	NA Y	2	NA	Действие первое   Явление
<b>3</b> Я пригласил вас, господа, с тем чтобы сообщить ва	р	NA	3	#gorodnichij	Действие первое   Явление
4 Как ревизор?	р	NA	4	#ammos_fedorovich_ljapkin_tjapkin	Действие первое   Явление
5 Как ревизор?	р	NA	5	#artemij_filippovich_zemljanika	Действие первое   Явление
Ревизор из Петербурга, инкогнито. И еще с секретн	р	NA	6	#gorodnichij	Действие первое   Явление
<sup>7</sup> Вот те на!	р	NA	7	#ammos_fedorovich_ljapkin_tjapkin	Действие первое   Явление
Вот не было заботы, так подай!	р	NA	8	#artemij_filippovich_zemljanika	Действие первое   Явление
Господи Боже! еще и с секретным предписаньем!	р	NA	9	#luka_lukich	Действие первое   Явление
Я как будто предчувствовал: сегодня мне всю ночь	р	NA	10	#gorodnichij	Действие первое   Явление
(бормочет вполголоса, пробегая скоро глазами)	stage	NA	10	#gorodnichij	Действие первое   Явление
2 и уведомить тебя». А! вот: «Спешу, между прочи	р	NA	10	#gorodnichij	Действие первое   Явление
(значительно поднимает палец вверх)	stage	NA	10	#gorodnichij	Действие первое   Явление
<ol> <li>Я узнал это от самых достоверных людей, хотя он</li> </ol>	р	NA	10	#gorodnichij	Действие первое   Явление
(остановясь)	stage	NA	10	#gorodnichij	Действие первое   Явление
uing 1 to 15 of 1 522 optring. 7 total columns		***	10	Anna an Antalatt	R.X
wing 1 to 15 of 1,525 entries, 7 total columns					
nsole Terminal × Jobs ×					_
R/parse_tei/ 🗇					
[] "The Government Inspector"	n_text	c(corpus =	"מווכ"	ulau "aaal aadaan")	

And as one of DraCor's main purposes is network analysis, one can of course use rdracor to retrieve networks of characters for the DraCor plays and analyze them making use of the rich network analysis infrastructure available in R. Here is an example of a network visualization with rdracor:



🔍 -   🔩   🛫 -   🕁 🙀   🍏   🎓 Coto file/function 🔢   🔯 - Addins -	parse_tei +		
d2*x vecher_df x O`Untitled3*x O example.R x revizor_df x .x O`example2.R x O`main.R x O`ds_infra_showcase.R x rus_corpus x >>	Environment History Connections Tutorial		
(+) 1	🛎 🚰 🔒 📅 Import Dataset 👻 🔬 📃 List 👻 🎯 🗸		
Q vecher (Next) (Prev) All revizor Replace All	R 👻 🍓 Global Environment 👻 🔍		
In selection Match case Whole word Regex 🗹 Wrap	Data		
1 library(rdracor)	• revizor_ig List of 10 a		
2 library(tidyverse)	Ins_corpus 212 obs. of 50 variables □		
3	Functions		
4 rus_corpus <- get_dracor(corpus = "rus")	parse_line function (line)		
<pre>5 revizor_igraph &lt;- play_igraph(corpus = "rus", play = "gogol-revizor")</pre>	segment no function (tei line sen = "   🖻		
6 plot(revizor_igraph)	Files Plots Packages Help Viewer		
7	<ul> <li>Image: A state of the state of</li></ul>		
<pre>8 get_play_metadata(corpus = "rus", play = "gogol-revizor")</pre>			
9			
<pre>10 revizor_text &lt;- get_play_spoken_text(corpus = "rus", play = "gogol-revizor")</pre>			
11 revizor_text[1]			
12	•		
13 revizor_tei <- get_play_tei(corpus = "rus", play = "gogol-revizor")			
14			
8:1 (Top Level) C R Scrip	it :		
Console Terminal × Jobs ×			
- (R/parse_tel/ ⇒			
	Аныа Андреевна		
> tubrary(tubyerse)			
Franciscurpus <- get_uracor(corpus = rus)	Городничий		
Could not receive heart, denore one			
Library(rdgarcan)			
	Хлестаков		
rue construction of the co			
Viaucijis comis			
<pre>&gt; rection is a proph &lt;- play is a proph (corpus = "rus" play = "agaal-revizor")</pre>			

> |