Handwritten Text Recognition in the Archives

Maria Kallio Senior Research Officer maria.kallio@arkisto.fi









- READ (Recognition and Enrichment of Archival Documents) 2016-2019
- Horizon 2020 Research and Innovation Infrastructures
- 14 partners
- The core mission is to make archival material more accessible through the development and dissemination of Automated Text Recognition (ATR)





New research infrastructure – Transkribus



- Automated Text Recognition and other tools
- Enabling computers to automatically transcribe and search historical documents – of any date, language and layout!



Automated Text Recognition (ATR)

• Machine learning (Neural networks)

THE NATIONAL ARCHIVES

OF FINLAND

- Learns to read specific styles of writing
- Needs to be trained by being shown document images and transcripts
- More training data \rightarrow more accurate ATR
- Create a ATR model to transcribe and search a collection of documents







Court Records 19th century

of Borge aghen belages shall harmon them. Part Henrikelow the pande, deraf han bla foil, ginder dige fatter Handas themas spor Grinder day 29 decoute fordister in more gjorde Strendersyte, and setting At Carl Henrikston betaler for wal alla for halften alfahr kan Hustan Jow alf and Stelfen his stor Soften Come Coige Anderidalla Jentis Rabel Banga Objquational fig i all matter underfacter, of makes finder, and fin Brooks Henrik anderstow. Dotto Uppland afterpy king upit, have fund lever general Postorous in all muddeler

\$ 2.

Sidan Rullfallow miges Cast Prinsport in

"Topornalis Offittino for an Riagio dal of Status and Raffine " Binto bing Play Super, Jac Raffines Singer East Sister Cator and State Part Cator to Sand Rafie Barris Copp., Guitard Borger dan & John. 1927. Kum O'Kaplanda.

23. Market Upstand in affentingen och inste oppind Standard og den Engin oppind in hange det og Ste vender Standard for Standard in Stange det og Ste vender Standarden Standard Engen och der verder som stangen beren Standarden Standarden Standard at ble forder oppinder Standarden Engenen dens forsteret at ble som stand for ble for Standarden Engenen dens forsteret at Standard stal for bester Standarden Engenen dens forsteret at Standard stal for bester State of the State State State of State 125 State of the State State of State 125 State of the State State State State of State 125 gjørde of the State State State State State of State 125 State of the State State State State State of State 125 State of the State State State State State of State 125 State of the State State State State State of State 125 State of the State State State State State of State 125 State of the State State State State State of State 125 State of the State State State State State of State 125 State of the State State State State of State 125 State of the State State State State State of State 125 State of the State State State State of State 125 State of the State State State State State 125 State of the State State State State State 125 State of the State State State State State 125 State of the State State State State State 125 State of the State State State State 125 State State State State State State State State 125 State 125 State at go minder " at mind a gone of for the optime conserve Marine ad Gara & Walt and the algoes it does not go gold and go at the Marine Conserver and the conserver of Const Control Conserver, and the Confect and Charles Constructions, he actual into Glan for Ball gold atom a stage of the Glan for Ball gold atom a figure to are marine as a provide go for a figure to are of Charles of a figure to areas and figure and a for a figure to areas and the provide gold atom to for a signal Charles of the offer atom and for a figure to areas and the stage of the figure to a for a construct a stage of the figure to areas and the stage of the figure to areas and the stage of the figure to areas and the stage of the stage and the top had her a figure to areas a stage of the stage of the

Appendissen set info "grand fran grander This of an enter langer franker der ander an entered an terneter langer franker der ander an entered Bart in Annel & Bart Stage State beigen Bart franker, hange ten tafet to be and terneter an grander all first toper all an farm Stanle for franker beigen get alternet, and franker beigen get alternet, entergrader be for state of high to the house of the state of all house the state of the state of the state of state to be the house to first of the house of the house of the state of the house the state of the state of the state of the term of the house of the state and the term of the state of the state of the terms of the former of the state and the terms of the former of the state and the terms of the state of the state and the terms of the state and the state of the state terms of the state of the state of the state terms of the state of the terms of the state terms of the state of the

Trige Copen upplied no offentilizers of him ton the

- Collection of 660 000 images
- Current HTR model CER 9%
- Most of the GT produced by volunteers

med Gosefficiale afhandling forkearon jag mig two alla delar nojo, and forbinder 6 6 00 H alla decar; focus hanned forfaleres. Ohkoit for ofoan. Erik Johanfan Sucojala Vill withow. Trik Mattofon Ander ifrom Ohkois by hom.

- 23 Mmed förestående afhandling förklarr jag
- 24 nig·till·ålla·delar·nöjd, amt förbinder
- 25 nig åfven att uppfylla densanma uti∉
- 26 alla delar, som hämned försäkres. Ohkon som
- 27 gfon⇔





Confidence	Page Nr.	Line transcription	Previ	
0.2496	1041	Ftörmyndares	Not my and a	
0.2494	156	manna lagens städganeden bom förupudarenädd	Firmfle	
0.2486	863	feor gndare for sin onm yndig	foringeda	
0.2482	163	de mot Bonden Nikadenmn Neloson Is Bruuni, Loeyrdare för	Su former again	
0.2452	447	di förnni ndae b,fåttni p skilas	first inte	
0.2446	483	ålle Carn till färmyndaren :	Bengaran.	
0.2439	802	söärmyndne ör sknna dredrika Alianders datten	trangente	
0.2437	363	ganden bim foyrdalevåt, till efterrattelse.	To margan the la	1
0.2429	816	Lovisa Borgström sig vara villig att attråda förmynder.	gite my me	
0.2380	479	Förmyndares	For my all	
0.2371	1026	gens talganden om örmyndarevart, gy=	participa a second	
0.2351	867	salorm vodon soig dare sor b alde	forengellar	
0.2287	274	gjens tlgmden oom förvmjndarvar till o=	Philippers.	
0.2285	942	å igende sörnydarno tillförortrande	for apartico	
0.2244	449	fönngndare or bernälde ngndirg utse och	Jun gular	
0.2215	449	sördngndare sön	friend a	
0.2198	354	tillförannåd sönmjrdere för sormessevage Aen=	for Back	
0.2190	812	Hemmala sig villig att ifrågakamne förmyndlr	Horne you	



ne 01 a

Close

7





Poll Tax Records 19th century

Kylia nini.	T x i w m Maatta 2 Thanka Thanka Xugen	Sava Sava Galemeta Malameta Malameta	Anishin and Anishin and Anishin and Anishin and Anishin anishi	se, pilitin istra- salletto percetter. Xangan S.	Tilm +	ckā mniek stait je	kirja ous a syntymi	6 nom po alenosi yi	dhinisead) anii 1971-194	in sainte Britaine	aint keis o heidirahasta	yris kenkli 1 y. m.	Bu m	akiraha desertis.	Joilla R Add a Co To that a Co	barphin ole and motor action a	When the dramy.	A B Mubbs	netia talent, Natio	mennet yfficiaeth verzin, pennet yfficiaeth verzin, meanefficiaeth yn enneth in fræerikentik yn enklori min efficiernels orski hrj min efficiernels orski hrj dda valmena yr inpernet	dan pateritanan hill anzi petatuna, oda bri dar naturna ja tapona	Mar valtalory 1986s d Missony vises	20 22 EreikValaisin	10 krostijos. 10 Nasa kriečejes	1000000
Ilautalahs	1		1	114	Lois	SU	at M	Va <i>st</i> t.	tunen	19:00	Varia	6.8.4.1	-1. 1	11	1			2	1			1			
	I. K	/	Rinn	* 45%	hoir	rula	mha	rsjie.	str	am l	Josen	peril	4. 1	3 2	/	/		4	2		9 4				
					-li.	shi_	tesh	e chi	usiai	nen 4	yh 2	Abre	tere.												
					ha	mil	l Da	vid	74.10:-	ella g	ash	lain	an -												
					-11	. tere	cal A	lann	renen	nale ?	Mais	lospe	nen												
1			1		1.	This	nas	rans 1° 11	- 39. 1	1.5	1.2	i.	/		1	1		1	1			2			
			1		sures a	ich h		Vinil	Ender	1841	our in	nerie		1	1			1	1			2			
			/		Alto	Mil	10.9	tubb	e pp n	Ame	1. 11	120	1	1/	2			3	1			4			
			/		Jois .	Ale	1 ha	risa	15 als	Innas	7		1	11				1	1			2			
			/		Mtp.	Ade	am e	Ma	tilais	ren i	6. The	hika	56		4	3		4	3			1			
					19	3.2	egel						Ξ.,		0				1	- ;	,	/			
		/			In f	cho	har	rhus	nemi	UL1	to the	Anni	w 4	D	2			1	2	//					
					162	1.11	1. 1.	Mail	8a.6%.	ann	Chul	73.alc	Ing -					ľ							
	514	,		11- 99	ma	151	Il.	1 0	A		1	~11	0 1	1 1	3	1		7	5	- /	1 9				
	010	1	*	14	zasi	ha 1	fus	hords A-Da	umar	Mart	Jul	2.112 C-10.	i.		0	/		1		11	× ×				
					11	in fil	a juce	Kon	hanes	. 57/0	dward	2 44C	6.												
					di	comie	n d	ville	Kash	inen la	bhi i	and h	ink.												
					1/h	men	2 bul	6 1. c	Iman	de 2	Koljon.	in 11.													
					Ol	gais	Hau	sho	nen j	1/-	(//													
			/		Lols	Will	le ch	Valie	ainen	The La	la 1/2		/	11				1	1			Z.			
		/			Je j	Tel 1	Wall	15.4.1	edma	linst	Mag	tilda.	\$1	2	1	1		2	3	0	?				
2			-/	-	ja	Pl.					111	10-0			1	,		0	0			11			
		1	/		Lois	Juk	o ja	ash	elain	= 1/2 C	aillia 1	2.9.1		1	1	/		Z	2		-	_ 7.			
		-/			1. S.	Emo	far	nthi	1220	ALKS .	a spl	1 h. L.	2	1	1			1	/	l					
			1		0.	mo	21 ja	X.L.		14. 6	Vall.	11	1	1		1		1	9			3			
			1		Lois	TAR.	a p	Maggin	enen 4	1 de la	- and -	aa ji	2 1	/		/		1	L			4			
		91	q								8	liste aun		1 10	01	0		11	100	1 12	11 7	40			÷

- Collection of 1,5 million images
- In house GT production
- Layout analysis and document understanding → digital register and research data





Our goals

- To provide 19th century Court Records in searchable and computer readable form by the end of the project
- To enable new versatile possibilities for research by giving our customers the chance to download all the produced data for their own use
- To integrate text recognition in to our digital services on large scale





For more information:

Transkribus website: https://transkribus.eu/

READ website: <u>https://read.transkribus.eu/</u>

Contact: maria.kallio@arkisto.fi or email@transkribus.eu

Twitter: @Transkribus

