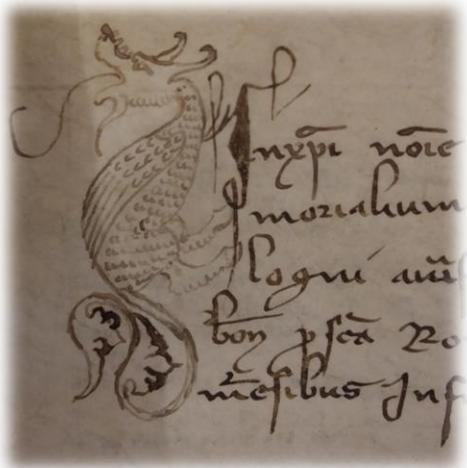


ENCODE Workshop
‘AI AND ANCIENT
WRITING CULTURES’
Bologna 23-27 January 2023

Challenges and issues of using *Transkribus* in large late medieval manuscript collections: The Memoriali Project (*MemoBo*)

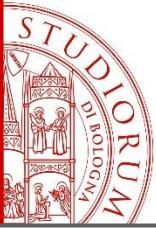


Edward Loss



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO



1) The *Memoriali*

Definition

The collection in the State Archive of Bologna

2) The *MemoBo* project

Aims and objectives

X-Dams, Regesta and the state of the art

Website layout

3) The *Memoriali* and *Transkribus*

Data input format: Diplomatic transcription + *tags vs* direct transcription

Overview of models : «pure» and «mixed» models

Memoriale 298 (1376)

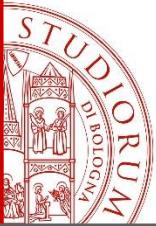
Memoriali 320 and 321 (1400-1436 – 40 different hands) – «mega model»

4) *Transkribus* and the *MemoBo* database

Customized Tags

Structural Tags/ Textual Tags

XML Transkribus/XML XDams



Definition:

- A series of notarial records – contracts, wills, dowry stipulations, emancipations etc – registered by a specific office, the «ufficio dei Memoriali» and created in order to avoid fraud and falsification.
- Product of a statutory (*statuta*) disposition: **April of 1265**



Rubrica XLIII- Qualiter contractus et ultime voluntates per notarios in memorialibus reducantur et qualiter ipsi notarii elegantur et qualiter ipsa memorialia fiant. Statuimus et ordinamus ut falsitatibus que circa instrumenta fiebant omnimode obvietur quod omnes contrahentes deinceps in civitate bononie et burgis, (...) Statuti di Bologna dall'anno 1245 all'anno 1267, III, a cura di Luigi Frati, Bologna, 1877, pp. 625-631.

- The series only contains documents in the value of 20 bolognese lire or more.

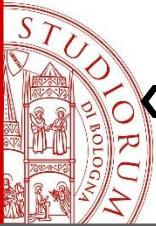


The *Memoriali* series in the State Archive of Bologna (ASBo)

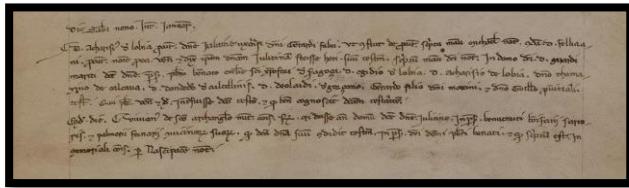
Composed of 322 volumes produced between 1265 and 1452, organized by semester.



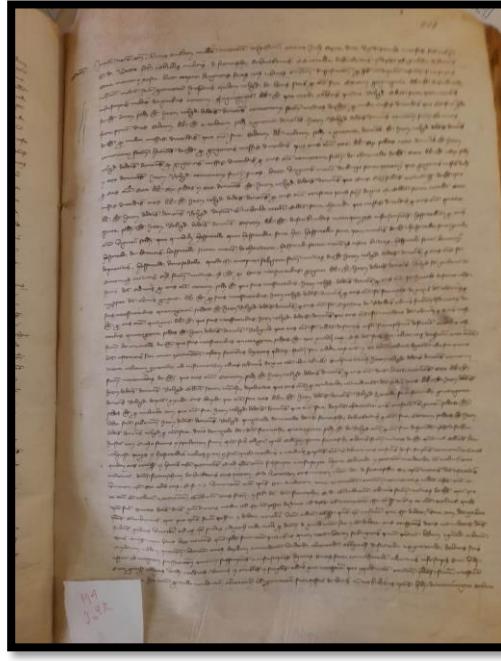
More than a million acts uninterruptedly covering 200 years of bolognese economic and social activities.



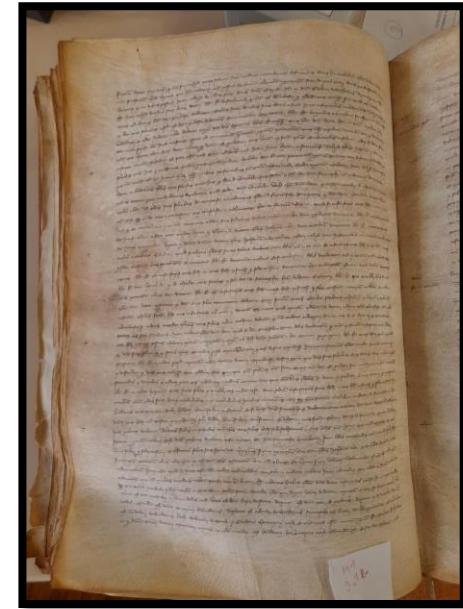
«...documents that were continuously evolving»

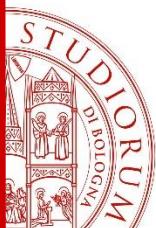


*Will of Jubetina, wife of Gerado,
blacksmith (09/01/1266)*



*Will of Dota, daughter of Francesco dei Clarissimi
– 20/02/1338*





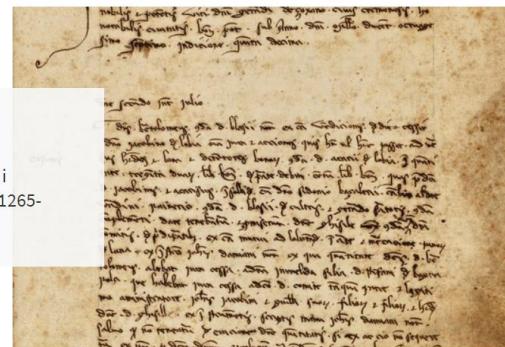
MemoBo

HOME PROGETTO PERSONE PARTNER DATABASE AGENDA PILOLE BIBLIOGRAFIA FOTOGALLERY LINK

Home / Progetto

Progetto

MemoBo. Un mare magnum di possibilità: i Memoriali bolognesi e la loro schedatura (1265-1452)



I Memoriali
bolognesi

✚ I libri memorialium del Comune di Bologna

L'Ufficio dei Memoriali fu istituito dal Comune di Bologna nel 1265 per registrare in

<https://site.unibo.it/memobo/it/progetto>



MemoBo

- 1) The creation of an **open-access search tool** with different access keys to the documentary series.
- 2) The organization of **serial data** that will provide scholars with a solid basis for more detailed research on economic, social, and legal history of a main medieval city.

- Together with *Regesta*, we created a data recording and indexing software in 2020 (on X-Dams)

The screenshot shows the X-Dams software interface for managing historical documents. The main window displays a list of entries under 'Volume 1 - Ognibene e Loss' with various document types like 'promessa', 'pace', and 'quieanza' from May 1265. The sidebar includes sections for 'Memorali', 'LIVELLO', 'IDENTIFICAZIONE', and 'NOTAIARI'. Navigation buttons at the bottom include 'SCHEDA', 'STRUTTURA', 'INIZIARSI', 'MODIFICA', 'SUCCESSIVO', 'SUPERIORE', 'INFERIORE', and 'XML'.



MemoBo - XDams



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

INSERISCI

CHIUDI X

è stata selezionata la scheda: Volume 1 - Ognibene e Loss

SALVA E CHIUDI

SALVA E

CONTINUA

CHIUDI

- informazioni di relazione
 - figlio
 - fratello precedente
 - fratello successivo

[+] inserimento multiplo di schede

- IDENTIFICAZIONE DELL'UNITÀ

visibilità della scheda

scegli... ▾

livello di descrizione

atto ▾

categoria

... ▾

- DATA DI REDAZIONE

da

/ /

a

/ /

forma visualizzata

APPLICA

forma normalizzata

SENZA DATA

indizione

nota alla data

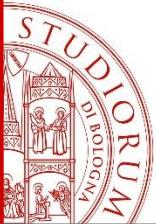
[+] DATA TOPICA: LUOGHI DI CITTÀ

[+] DATA TOPICA: GUARDIA CIVITATIS

[+] DATA TOPICA: LUOGHI DEL CONTADO

[+] DATA TOPICA: LUOGHI EXTRA GIURISDIZIONE

- NOTAIO REGISTRATORE



MemoBo - XDams



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

MODIFICA

CHIUDI X

[IDENTIFICAZIONE E DESCRIZIONE](#) | [INDICI NOMI](#) | [EDIZIONI, NOTE E COMPILAZIONE](#)

[PRECEDENTE](#)

[SUCCESSIVO](#)

[SUPERIORE](#)

[SALVA](#)

[SALVA E CHIUDI](#)

[CHIUDI](#)

Area indici nomi

PERSONE

genere

... ▾

[AGGIUNGI](#)

nome

...

cognome

...

/c/controlaccess[@altrender='persone']/list/item[1]/persname[@altrender='cognome']/text()

età

...

provenienza o residenza

...

qualifica

...

[INSERISCI](#)
[RIMUOVI](#)

ente di appartenenza/riferimento

[APRI](#)

...

condizione

...

mestiere

...

funzione

...

[APRI](#) note

...



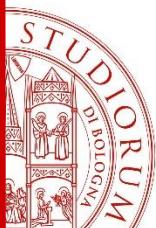


MemoBo - XDams

Some of the information catalogued and indexed:

- 1) Type of the act
- 2) People involved and their roles
 - 2.a) name, patronym and surname
 - 2.b) geographical provenance
 - 2.c) profession
- Etc
- 3) The object
 - 3.1) transferable goods, rights and etc
 - 3.2) their value and type of currency

- Main challenge:
the size of the collection!



MemoBo and A. I.

The search for A. I.

1) Supply the lack of manpower

2) Possible solution for the challenge of time

- Study of the options available in the market:



RESCRIBE



eScriptorium



MemoBo and Transkribus



Advantages:

Easier to install

Easier to use than most softwares

Sustainable structure (more possibilities of lasting)

Disadvantages:

Private for profit endeavour: no Italian university can establish a direct partnership according to current legislation, which complicates funding.

Costs considering the size of the collection



MemoBo and Transkribus

Models of data input for the creation of «ground truth»

1) Diplomatic transcription + «tags»

- Transcription of graphic elements as we see them on the parchment, copying medieval characters and specific abbreviation trace from MUFU (Medieval Unicode Font Initiative)
 - Expansion of abbreviations through the «tag» system.

MUFI	Medieval Unicode Font Initiative																			
	0020	0021	0022	0023	0024	0025	0026	0027	0028	0029	002A	002B	002C	002D	002E	002F	Q.			
Home	!	"	#	\$	%	&	'	()	*	+	,	-	.	/						
Browse by character	0030	0031	0032	0033	0034	0035	0036	0037	0038	0039	003A	003B	003C	003D	003E	003F				
Browse by code chart	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?				
Browse by range	0040	0041	0042	0043	0044	0045	0046	0047	0048	0049	004A	004B	004C	004D	004E	004F				
Browse by updates	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O				
Full code chart	P	Q	R	S	T	U	V	W	X	Y	Z	{ \ }	^	-						
	0060	0061	0062	0063	0064	0065	0066	0067	0068	0069	006A	006B	006C	006D	006E	006F				
	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o				
	0070	0071	0072	0073	0074	0075	0076	0077	0078	0079	007A	007B	007C	007D	007E	007F				
	p	q	r	s	t	u	v	w	x	y	z	{ }		{ }		~				
	0060	0061	0062	0063	0064	0065	0066	0067	0068	0069	006A	006B	006C	006D	006E	006F				
	i	ƒ	£	¤	¥	¦	§	„	„	©	ª	«	¬	®	–					
	0080	0081	0082	0083	0084	0085	0086	0087	0088	0089	008A	008B	008C	008D	008E	008F				
	o	±	²	³	‘	µ	¶	·	,	í	ó	»	¼	½	¾	¿				
	00C0	00C1	00C2	00C3	00C4	00C5	00C6	00C7	00C8	00C9	00CA	00CB	00CC	00CD	00CE	00CF				
	À	Á	Â	Ã	Ä	Å	È	É	Ê	Ê	Ê	Ê	Ê	Í	Í	Í				
	00D0	00D1	00D2	00D3	00D4	00D5	00D6	00D7	00D8	00D9	00DA	00DB	00DC	00DD	00DE	00DF				
	Ð	Ñ	Ó	Ó	Ó	Ó	Ó	Ó	Ó	Ó	Ó	Ó	Ó	Ó	Ó	Ó				
	00E0	00E1	00E2	00E3	00E4	00E5	00E6	00E7	00E8	00E9	00EA	00EB	00EC	00ED	00EE	00EF				
	À	Á	Â	Ã	Ä	Å	È	É	Ê	Ê	Ê	Ê	Ê	Í	Í	Í				

<https://mufi.info/m.php?p=muficodechart#&ui-state=dialog>



MemoBo and Transkribus: Models of data input

1) Diplomatic transcription + «tags»

Transkribus Expert Client v1.21.0 (22_09_2022_07:35), Loaded doc: ASBo, Memoriale 69, intero definizione minore, ID: 1258224, Page 345, file: Memoriali_69_0204_recto.jpg [Image Meta Info: (Resolution:150.0, w*h: 1872 * 2808)] [current line: w*h: 998 * 8]

Server Overview Layout **Metadata** Tools

Document Page Structural **Textual** Comments

Tags of current Transcript

Tag	Value	Text	Properties
1	abbrev	nōie	Illius nōie ihu xpī qui carnem suscepit de v
2	abbrev	ihu	egistrum sive memoriale contractuum et u
3	abbrev	xpī	1-3 per me henrigi⁹tum de querci⁹is notarium d
4	abbrev	ultima	1-4 bōnon deputatum tprē nobilis potentis viri c
5	abbrev	script⁹	1-5 civis cremonensis honorabilis civitatis Bonon⁹
6	abbrev	dct⁹	1-6 mill⁹ ducent⁹ octuāg⁹ septimo indictoē quin
7	abbrev	cōe	1-7 die scdō iñt jullio
8	abbrev	bōnor	1-8 dñs Gerarducius gdam magri Petri C S pculi pmisit scribere & complere dnō
9	abbrev	tprē	
10	abbrev	Bonon⁹	
11	abbrev	dni	
12	abbrev	mill⁹	
13	abbrev	ducen⁹	
14	abbrev	octuā	
15	abbrev	indict⁹	

Tags

Tag Specifications User tags Collection tags Tag specifications Customize...

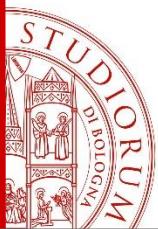
Tag specification Color Shortcut

Props for tag: 'abbrev' - value: 'Bonon⁹'
expansion Bonon⁹

die scdō iñt jullio

Copy text
Paste text
All tags
Add a comment
Delete all tags for current selection
Delete tag: abbrev {offset:40; length:6; expansion:Bonon⁹;}

The screenshot shows a medieval manuscript page from the 'Memoriale 69' with a large initial 'R'. The text is written in a Gothic script with some red rubrication. A context menu is open over the text 'die scdō iñt jullio', showing options like 'Copy text', 'Paste text', and a list of tags. The 'Tags' panel on the left lists abbreviations and their expansions, such as 'nōie' for 'nōie ihu xpī qui carnem suscepit de v'. The 'Tag Specifications' panel at the bottom shows a single entry for 'abbrev' with the value 'Bonon⁹'. The status bar at the bottom right shows the date and time: '23/01/2023 12:49'.

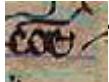


MemoBo models of data input: Diplomatic transcription + «tags»

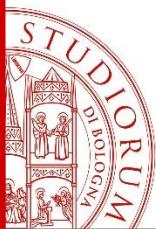
Advantages:

- Smaller number of pages necessary to create a model
- More precise reading models (in theory)
- Especially useful for the production of critical editions

Disadvantages:

- **Time-consuming process:** 2-3 hours of preparation for each page.
- **Very arbitrary process:** the selection of characters input depends on the observer's individual understanding of the script. Es. coē , coe or coē for ?
- Not all characters available on MUFI are supported by transkribus: «□»

Put aside in the initial test



MemoBo and Transkribus

Models of data input for the creation of «ground truth»

2) Direct transcription

- Expansion of abbreviations directly in the main text

1-9 aixerunt et denunciaverunt predicta in memorialibus predictis die quinto decimo aucti mensis aprilis et ea in
ipsis memorialibus ponit
1-10 scribi et registrari petierunt secundum formam statutorum dicti communis Bononie
2-1 Millesimo trecentesimo septuagesimo sexto indictione quartadecima die sextodecimo mensis aprilis
tempore pontificatus sanctissimi in Christo patris et domini
2-2 domini Gregorii divina providentia pape undecimi | domina Ymiglia filia quondam Johannis de Terra Sassuni
et uxor Bevolli quondam ser Abelli de



MemoBo models of data imput: Direct transcription

Advantages:

- **Quicker to produce ground truth:**
estimated 10 minutes a page
- No need to use external characters and fonts (avoid subsequent errors)

Disadvantages:

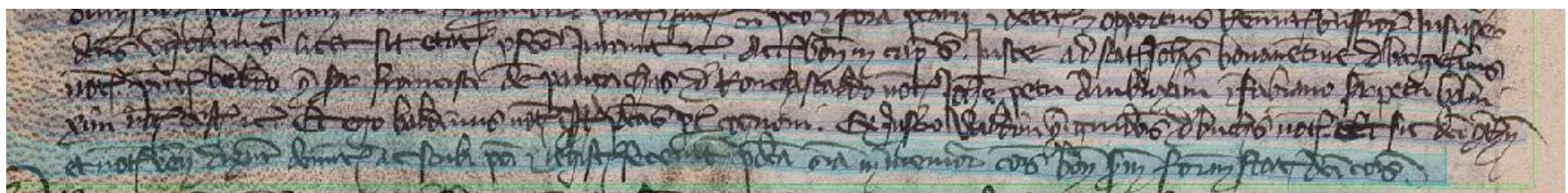
- **Larger amount of material necessary to produce first model**
- **Necessity of more experienced and prepared trainers (paleography and latin)**
- **Long preparatory phases (decisions on how to transcribe vowels and consants, on how to expand abbreviations and etc.)**

Chosen for the initial test

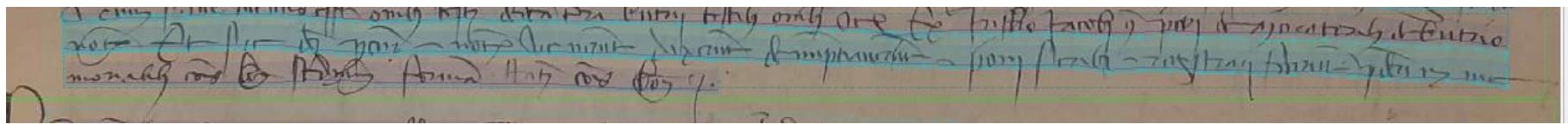


MemoBo models of data imput: Direct transcription

«Et sic dicti procurator et notarius venerunt dixerunt denumptiaverunt et poni scribi et registrari fecerunt predicta in memorialibus comunis Bononie secundum formam statutorum communis Bononie»



Memoriale 298 (1376)



Memoriale 320 (atto del 1414)



MemoBo and Transkribus

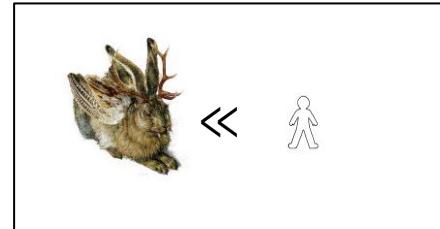


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

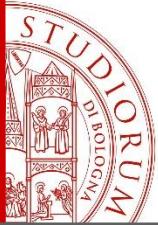
First phase of the experiment:

Specific goals:



- a) Establish the minimum material quantity (*Ground Truth*) necessary to create functioning models specifically for the Memoriali.

- b) Verify the software's sensibility to the quality of image reproduction
 - Professional reproductions
 - Amatorial reproduction
 - Microfilm (future step)



MemoBo and Transkribus

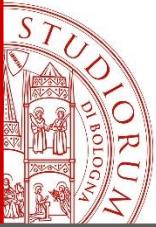


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

First step of the experiment

- The creation of 26 models based on Memoriali from the Thirteenth-Century
- Insistence of the creation of «pure» models. (Generation of «base models» for future training)
- Verify any difficulties the software might have with the bolognese abbreviation system.



MemoBo and Transkribus



PROGETTO DI RICERCA
MEMOBO

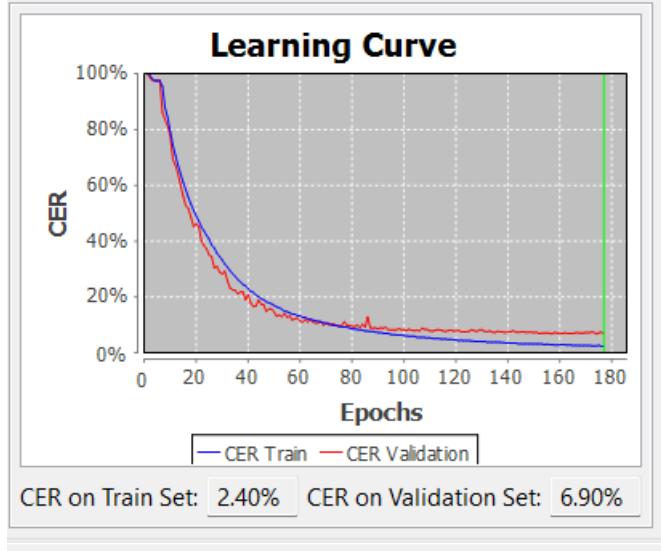
1) *Memoriale Nascimpece Petrizani* (Memorial 1)

- 5 models
- The most efficient: *MemoBo*, Me. 1,
Model. 3

Number of pages: 50 pag. (25 c.)

Time spent: 30 hours (in all steps –
caricamento, *layout generation*, etc.)

CER on Train: 2,40% CER on Validation:
6,90%



In practical terms:
1 mistake every 18 words

The screenshot shows a medieval manuscript page with handwritten Latin text in two columns. Red boxes highlight specific words or lines for transcription. A sidebar on the left contains icons and labels for different character sets: TR, L, BL, W, U, H, V, L, and other symbols. Below the manuscript, several lines of transcribed text are displayed in blue. The first line reads: "ipsius indecum corbum friae. ex instantem nga". The second line is numbered 1-23 and reads: "ini fi quondam ednis dipii notari facto hodie. subtotim domius dicti notari presentibus bevenuto martiniti et johanne filio johannini. de lalana". The third line is numbered 1-24 and reads: "et domino c̄boninio de stella. testibus". The fourth line is numbered 1-25 and reads: "Eodem die". The fifth line is numbered 2-1 and reads: "Matheus martini renovati de flesso dixerunt conduxisse ad laborandum a domino montanario quondam domini berardi debargaça duas vachis et duas". The sixth line is numbered 2-2 and reads: "vitellas et unum vitellum pretii et extimaonis viginti sex librarum ad unum annum ceipolabore sex corbeies frumenti".

Page transcribe by the software in 33.02 seconds



2) *Transkribus e i Memoriali bolognesi*



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

Some abbreviations

Documents Fulltext (Solr) Tags KWS (CITlab)

Search transcribed text for words or phrases

Search for Anyword All words Exact phrase

Input: dominus

Show word preview Current document Word-based text Line-based text

Case sensitive Fuzzy search

Search results

Showing lines 1 to 100 (total hits: 263)

All collections All documents All authors All uploads

Context

Document

dominus Iohannes et Bittinus testes predictus procurator dicte domine Zanvere ex instrume Memoriale di Nic

et uxor quondam domini Tantidinari **1 dominus** frater Matheus quondam domini Peregrini de Memoriale di Nico

. **1 Dominus** Matheus quondam domini Iacobini de Castello de capella Sancti Luche fuerunt Memoriale di Nic

venerunt denunciaverunt et scribi fecerunt Lune quartodecimo intrante Jullio **Dominus** Alar Memoriale di Nic

Dominus Rolandus quondam Guillielmi butrigarii infirmus suum condidit testamentum in pte ASBo, Memoriale di Nic

Bertholacio Morandi testibus et sic dictus **dominus** presibiter Ugolinus qui dixit cognoscere ASBo, Memoriale di Nic

facti dicti die loco et testibus et sic dicti contrahentes dixerunt et scribi fecerunt. **1 Dominus** TRAINING_VALIDA

sic dicti contrahentes dixerunt et scribi fecerunt. **Dominus** Neri quondam Bonicnti caplio Sar TRAINING_VALIDA

contrahentes dixerunt et scribi fecerunt **1 Dominus** Jacobinus cui dicitur Minus filia quondam TRAINING VALIDA

Dominus, Domini

Search for Anyword All words Exact phrase

Input: bononie

Show word preview Current document Word-based text Line-based text

Case sensitive Fuzzy search

Search results

Showing lines 1 to 100 (total hits: 307)

All collections All documents All authors All uploads

Context

Document

. die nono. exeunte decembre.. actum **bononie**. sub porticus domus pallatii veteris comuniſ b parte 2

. domino Symone de guaçola. et Naximbene iohannis. testibus Eodem die Blasius bannitor cc parte 2

februarii. in pallatio veteri comuniſ **bononie**. coram domino Juliano de castaldo. iudice et assessori d parte 2

Guillelmi de sesso. potestatis. **bononie** presentibus. domino Gandulfo petri de ymola iudice et assessori d parte 2

bononie. cridasſe. in pallatio et in scalis pallatii. comuniſ et ante domum dicti domini Guillelmo d parte 2

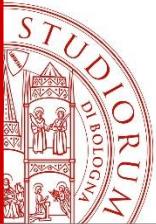
sextagesimo tercio indicione sexta. die tercio exeunte. madio. super solario pallatii. comuniſ et ante domum dicti domini Guillelmo d parte 2

. domino Gallerano iudice et assessore dicti Jacobi tavernerii. potestatis. **bononie**. presentibus. d parte 2

. sexagesimo primo. indicione. quarta die quarto. intrante novembris. in pallatio veteri comuniſ et ante domum dicti domini Guillelmo d parte 2

. coram. domini. imperiali iudicie et assessore domini mathei de coricio. potestatis **bononie**. d. f. p. b. parte 2

Bononia, Bononie



Transkribus and MemoBo



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

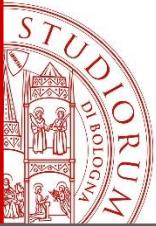
PROGETTO DI RICERCA
MEMOBO

Attempt to apply the *MemoBo, Me. 1, Model.* su altri Memoriali

Memoriale of Nicola da Lastignano (1285)

3-5 quixa tivanus suum secett teessomnum scriptum marinu Sancti cu hioli notini creserto Bancti
3-6 domi Jacobini prebiteri Barbaibri Ieri juliali prsentibus Pesimnrbino de baubis et senqis heendas
3-7 Detam
4-1 sa Faciolus flius comine Alborii e casstri feretri prts emancipatis et fax suxo postii
4-2 u dominus tetimni e itrno miunu Ysauli Betantimur notarius promisi darti et restemitu de-
4-3 Biandino e Sao irmnestei uintia triptem libras bononinorum hinc ad iraspisle esia lerbis
4-4 en instrmrno Inuti et Barmibiaoli tuni hodii fratri in prptuns ipsius notri presente dehanibus
4-5 cu amdi Aaralba surict scr quend atribitini mantrisur t Buarçencii tesas rius
4-6 lati eit
4-7 Constatu deo dicitur dicitur

Colossal failure:
mistakes in almost
every word



Transkribus and MemoBo



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

Second phase of the experiment: Mixed Models

Memoriali of Nascimpace Petrizani and of Nicola da Lastignano (Memoriali 1 and 58)

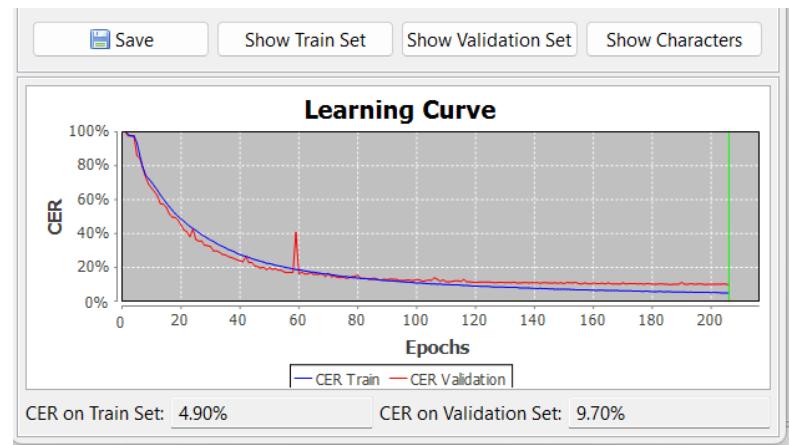
- 3 models
- The most efficient (to this date): *MemoBo, Me. 1 and 58, Model. 3*

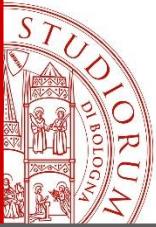
Number of pages: 50 pag (25c.) from each:
100pag (50c.)

Time spent: ±50 hours

CER in training: 4,90 % CER in validation:
9,70%

In practical terms: 5 mistakes every 15 words





Transkribus and MemoBo

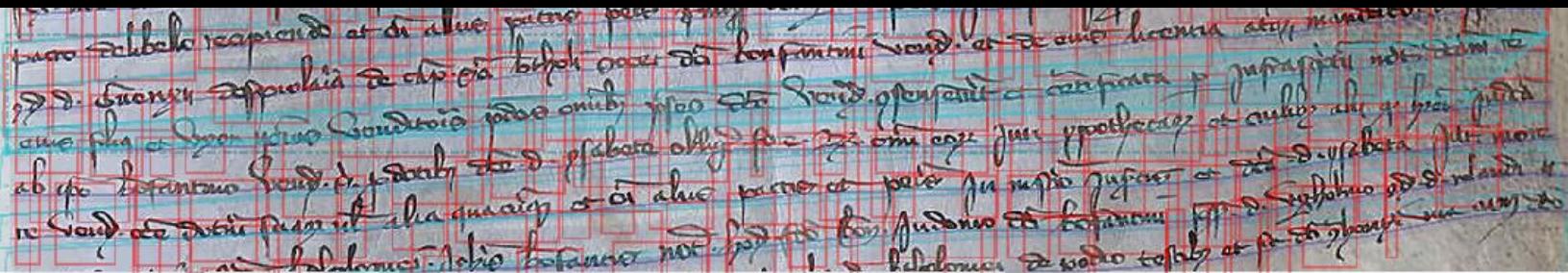


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

Application capability of this model to other Memoriali?

- Es: Memoriale 100 (1300) – never seen before by the software

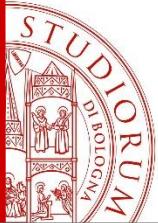


17. quondam domini Ginixii depola de capella Sancti Baholi soror dicti Bonfantini venditer et de eius licentia Amadicuppletta
18. eius filia et uxor ipsius venditionis predictis omnibus pres dicte venditionis consenserit et conficata per frassestum notti
te-

2-38 quondam domini Ginixii depola de capella Sancti Baholi soror dicti Bonfantini venditer et de eius licentia Amadicuppletta

2-39 eius filia et uxor ipsius venditionis predictis omnibus pres dicte venditionis consenserit et conficata per frassestum notti
te-

*picture taken by cellphone on a crooked support



Transkribus and MemoBo

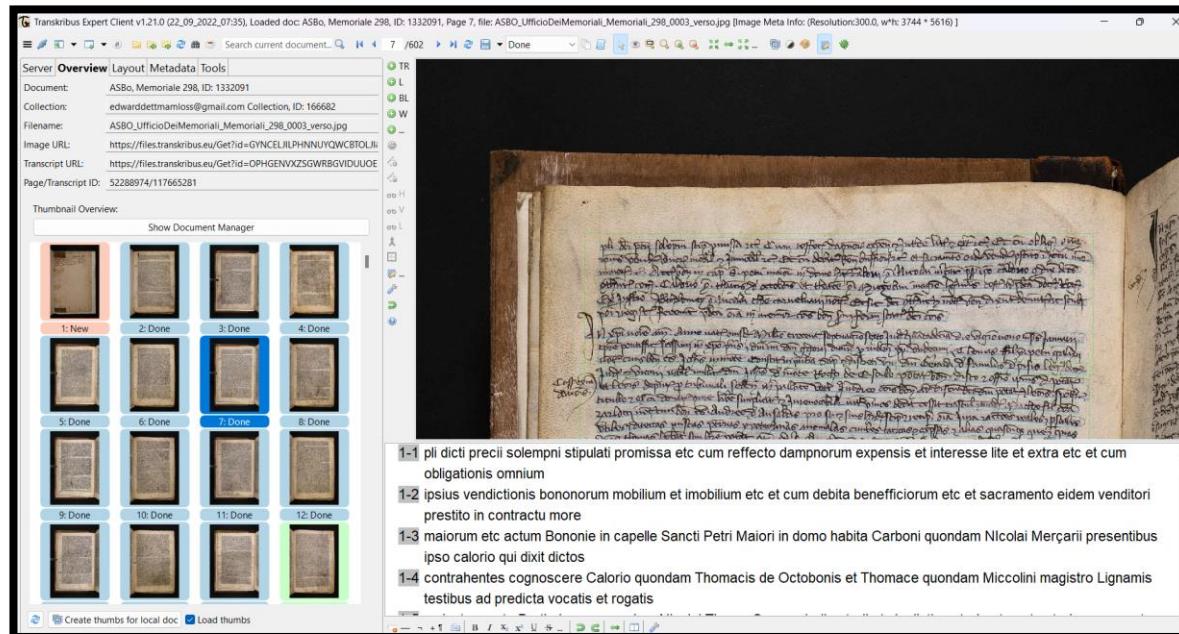


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

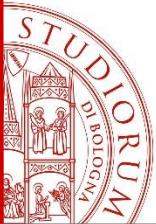
PROGETTO DI RICERCA
MEMOBO

Third phase of the experiment

1) Creation of fourteenth-century models



Memoriale 298 (1376)
4 different hands



Transkribus and MemoBo



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

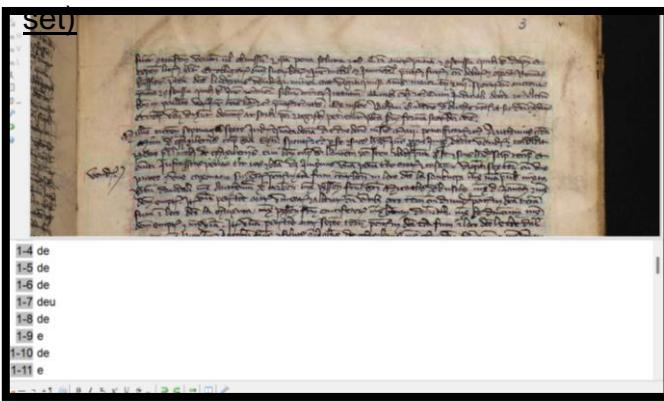
MemoBo, Me. 298, Model. 1

CER in training: 97% CER in validation: 97,1% (disaster)

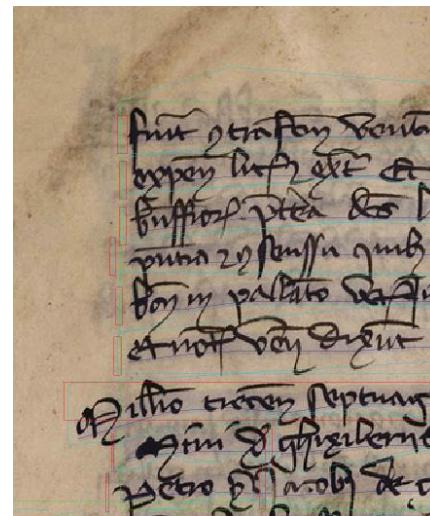
Number of pages: 50p (25c)

Base model: *MemoBo, Me. 1 and 58, Model. 3*

«Page» transcribed by the software («training validation

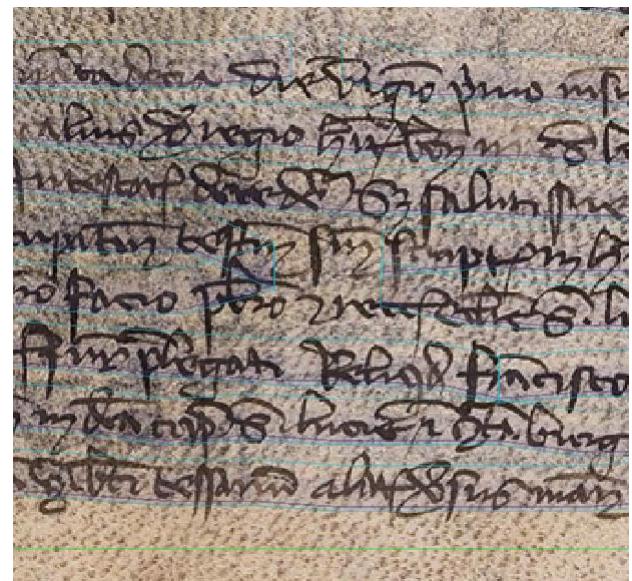


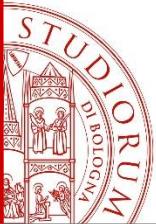
*not capable of identifying words



Probable explanations

- Serious issues in the automatic layout generation





Transkribus and MemoBo



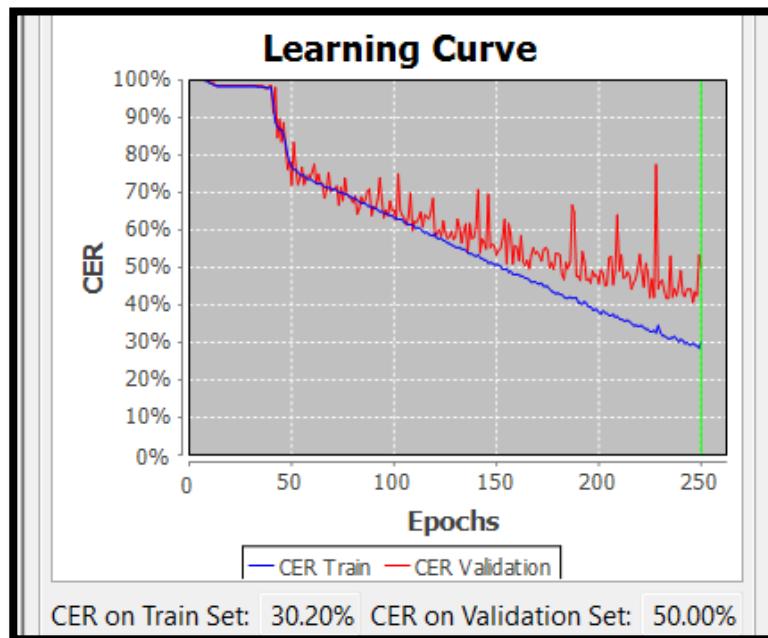
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

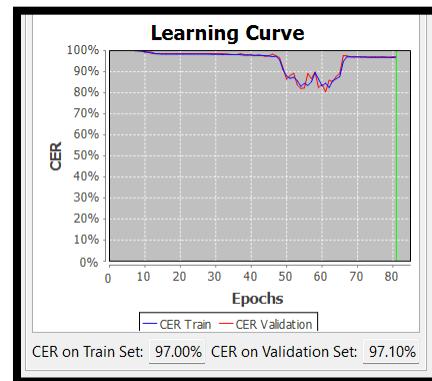
- After correcting approximately 20% of all lines (71 lines per page)

MemoBo, Me. 298, Model. 2

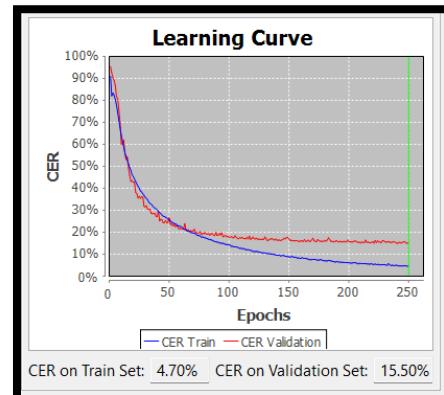
CER in training: 30,20% CER in validation: 50%

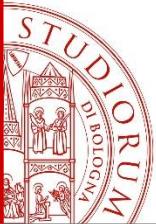


Before the corrections



After 60% of layout and line corrections





Transkribus and MemoBo



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

Fourth Phase: a fifteenth-century mega model (ongoing)

Memoriali 320 and 321 (the last ones of the series)

- Particularly interesting for the experimentation because they are amongst the most heterogeneous of all the Memoriali: composed of nearly 40 different hands.

Iacobus Gregori de Sachis (1400)

Franciscus quondam Nicholai de Plantavignis (1400)

Baldesar quondam Thome de Tentaquattro (1400)

Iohannes Magistri Lencii (1400)

Iohannes quondam Dominici de Brazarola (1401)

Laurentius Francischini de Chanonicis (1402)

Rustiganus quondam Soldani de Rustiganis (1403)

Monsinus quondam Bechadelli de Aliotis (1403)

Guilielus quondam ser Petri olim magistri Guilielmi de Varignana (1403)

Nicholaus Jacobi de Maronibus (1404)

Iohannes Jacobi de Maronibus (1404)

Iohannes Dominici de Braçarola (1404)

Nicolaus quondam Simonis Lotti (1405)

Guido quondam Francisci de Paganelis (1407 e 1409)

Bartholomeus Maghinardi de Hosbertis (1408)

Iohanes magistri Lentii (1409 e 1416)

Dinus Andree de Mezacvachis (1410)

Iacobus quondam Dominici olim ser Pauli de Castagnolo (1410)

Bernardinus Francisci de Mulettis (1410)

Iohannes quondam ser Federici olim Francisci de Sanctis (1411)

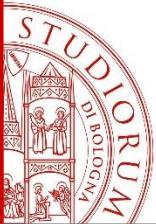
Iacobus Nanini de Maymelinis (1412, 1415, 1416)

Franciscus quondam Bondi de Dachis (1416)

Petrus Dini de Rapegiis de Vargnana (1416)

Iohannes ser Iacobi de Castelanis (1416)

ecc.



Transkribus and MemoBo

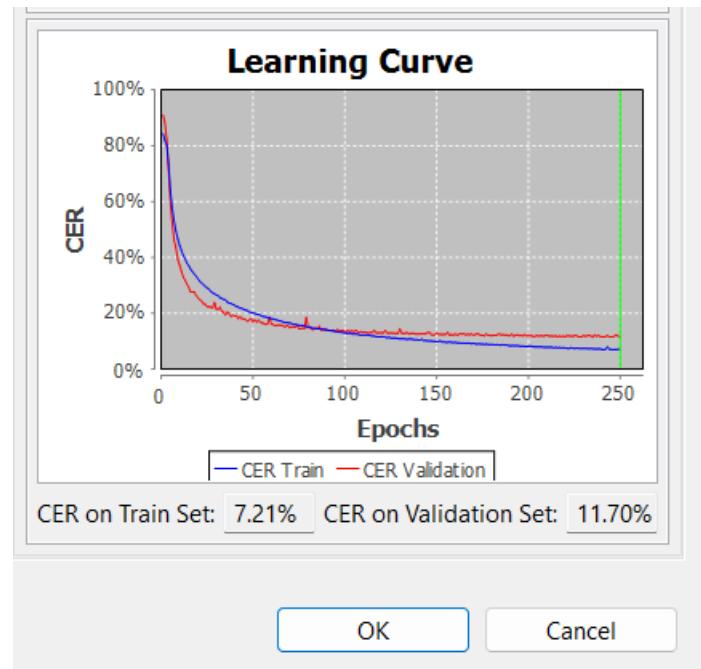


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

Fourth phase: a fifteenth-century mega model (ongoing)

- The most efficient: *MemoBo, Me. 320 and 321, Model. 6*
- *Ground Truth*: 10% of pages available from each notary (sometimes only 4-5 pages) – 70 000 words
CER on Train: 7,21% CER on Validation: 11,70%

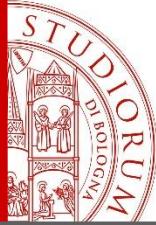


Base models:

MemoBo, Me. 1 and 58, Model. 3

MemoBo, Me. 298, Model. 5

Still non optimal, but....



Transkribus and MemoBo



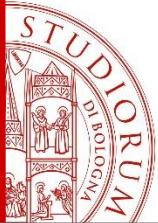
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

2-4 ac saluti aive sue providere dchiis et super hiis et universo suo patrimonio per presens nuncupatum sive scriptis testamentum inhuum notarium facieore poravit et fecit paimo quode Bebidi

2-5 pro anima sua pro male ablatis incertis cui de iure debentur solidisquinque bononinorum Item reliquit clesie Sancti Seani de terra vigni pro reparacione ipsius tunc pro anima sua et

In practical terms: 4 mistakes every 30 words



Transkribus and the MemoBo database



PROGETTO DI RICERCA
MEMOBO

Università di Bologna - Progetto Memoriali

NUOVA RICERCA ▾

Benvenuto Edward Loss (livello: 2) set in: Archivi / Memoriali

visualizza 50 ▾ elementi per pagina

- Volume 1 - Ognibene e Loss
 - promessa 05 maggio 1265
 - pace 05 maggio 1265
 - promessa 05 maggio 1265
 - promessa 05 maggio 1265
 - compravendita 05 maggio 1265
 - quietanza 05 maggio 1265
 - quietanza 05 maggio 1265
 - pace 05 maggio 1265
 - promessa 05 maggio 1265
 - rinuncia 05 maggio 1265
 - promessa 05 maggio 1265
 - promessa 05 maggio 1265
 - compravendita 05 maggio 1265
 - promessa 06 maggio 1265
 - quietanza 06 maggio 1265
 - promessa 06 maggio 1265
 - promessa 06 maggio 1265
 - compravendita 06 maggio 1265
 - promessa 07 maggio 1265
 - promessa 07 maggio 1265
 - promessa 07 maggio 1265
 - quietanza 07 maggio 1265
 - quietanza 07 maggio 1265
 - quietanza 07 maggio 1265

unibo.xdams.org/xdams-unibo/viewTab/uniboXDamsHistReg001/shortTab.html?physDoc=3171&pageToShow=8&perpage=50

xdams

icerca libera

SELEZIONA | SPOSTA | RIORDINA | ELIMINA | TAGLIA | COPIA | STAMPA | FIND|REPLACE

> Memoriali

LIVELLO

volume

IDENTIFICAZIONE

Volume 1 - Ognibene e Loss

NOTAIOI

Nascimpare qd. Petriziani

SCHEDA | STRUTTURA |

| SUCCESSIVO | SUPERIORE | INFERIORE |

INSERISCI | MODIFICA | XML

INSERISCI

è stata selezionata la scheda: Volume 1 - Ognibene e Loss

SALVA E CHIUDI

SALVA E CONTINUA

CHIUDI

I informazioni di relazione

tipo

fratello precedente

fratello successivo

inserimento multiplo di schede

IDENTIFICAZIONE DELL'UNITÀ

visibilità della scheda

livello di descrizione

categoria

DATA DI REDAZIONE

da

a

forma visualizzata

forma nominativa

indirizzo

nota alla data

DATA TOPICA: LUOGHI DI CITTÀ

DATA TOPICA: GUARDIA CIVITATIS

DATA TOPICA: LUOGHI DEL CONTADO

DATA TOPICA: LUOGHI EXTRA GIURISDIZIONE

NOTAIO REGISTRATORE

MODIFICA

IDENTIFICAZIONE E DESCRIZIONE | INDICI NOME | EDIZIONI, NOTE E COMPIAZIONE

Area indici nomi

PRECEDENTE

SUCCESSIVO

SUPERIORE

SALVA

SALVA E CHIUDI

CHIUDI

PERSONE

genero

nome

cognome

età

provenienza o residenza

qualifica

ente di appartenenza/riferimento
[APRI](#)

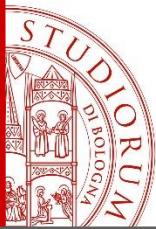
condizione

mestiere

funzione

[APRI](#) note

XDams – Regesta.exe



Transkribus and the MemoBo database



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

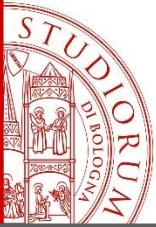
The «tags» system

The screenshot shows a digital image of a medieval manuscript page from the Transkribus application. The page is filled with dense handwritten text in a Gothic script. A floating context menu is open over a portion of the text, listing various categories for tagging: Address, abbrev, add, blackening, comment, date, div, gap, organization, person, place, sic, speech, supplied, unclear, and work. Below the menu, a list of numbered tags is displayed, corresponding to specific lines of text:

- 2-3 anus et dominus Bonjohanninus promisseru
- 2-4 dominus Damianus pro duabus partibus et d
- 2-5 dictum terminum ex instrumento domini Ben
- 2-6 Bononie sub portichu pallatii veteris comunis
- 2-7 de Argelata domino Iohanne de Forantibus
- 2-8 Querçis no
- 2-9 averunt et r
- 2-10 Domina Bo

At the bottom of the menu, there are three options: Copy text, Paste text, and Delete all tags for current selection.

Possibility of customizing all categories:
«author»,
«residence», etc



Transkribus and the *MemoBo* database



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

«Customized Tags»

Tags

User tags Collection tags Tag specifications

Tag specification	Color	Shortcut
abbrev	Red	
add	Cyan	
Altro	Brown	
Arbitro	Light Yellow	
Autore	Teal	
Beni-Immobili	Grey	
Beni-Mobili	Maroon	
blackening	Black	
Cappella	Light Green	
Carta	Brown	
Categoria	Cyan	
Condizione	Dark Teal	
Conduttore	Dark Green	
Confinante	Green	
Creditore	Blue	
Curatore	Dark Grey	
Data-Topica	Brown	
date	Blue	

Props for tag: no tag selected

Tag specification	Color	Shortcut
Debitore	Pink	
Destinatario	Green	
Diritti	Purple	
div	Blue	
E-Appartenenza	Yellow	
Erede	Grey	
Esecutore	Red	
Fideiussore	Olive Green	
gap	Cyan	
Legatario	Teal	
Macro-toponimo	Cyan	
Mestiere	Dark Blue	
Misure	Dark Grey	
N-rogatario	Pink	
organization	Pink	
person	Green	
place	Purple	
Presente	Orange	

Tag specification	Color	Shortcut
Presente	Orange	
Procuratore	Pink	
Provenienza	Green	
Qualifica	Light Blue	
Residenza	Brown	
sic	Yellow	
Sostituto-Testamen...	Purple	
speech	Dark Red	
supplied	Red	
Testimone	Orange	
Tipologia	Red	
Titolare-precedente	Light Blue	
Tutore	Dark Blue	
unclear	Yellow	
Valore	Dark Purple	
Valuta	Dark Blue	
work	Dark Green	



Transkribus and the MemoBo database



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

Transkribus®

edwardde...collection > 321 - (3 carte) > #2

Feedback

Done

4.9.2023, 08:27

Share

Region 2

Millesimo quadringentesimo decimo septimo indictione decima die ultimo mensis Januarii.

Johannes quondam Dominici Boellahabitator Terre Vida-

galle comitatus Bononiesuo proprio et principali nomine et per se et suos heredes ac etiam vice et

nomine Antonii eius fratratis et fratri

dicti quondam Dominici pro quo de ratho et rathi habitatione promisit in omnem casum et

eventum omnium singulorum infrascriptum et sub infrascripta

pena et obligatione omnium suorum bonorum. Sponte et ex certa scientia nulloque iuris vel facti

errore ductus nominibus et modis

pluris iure proprio et imperpetuum dedit et tradidit insolutum et titullo pro soluto

Federico quondam Putii de Florentia mercatorii habitatore bononiensi

in capella Sancti Johannis in Monteibidem presenti et pro se et suis heredibus stipulanti et

recipienti unam petiam terre aratorie arborate et vidate octo tornaturarum

vel circa positam in guardia terre viaragole comitatus Bononie in loco dicto lebelelle sive armole

iuxta viam publicam a duobus

donationis

lateribus iuxta Stephanum Maxii Guirixinii iuxta ser Ugolinum Bidinellis de Viadagolla et iuxta

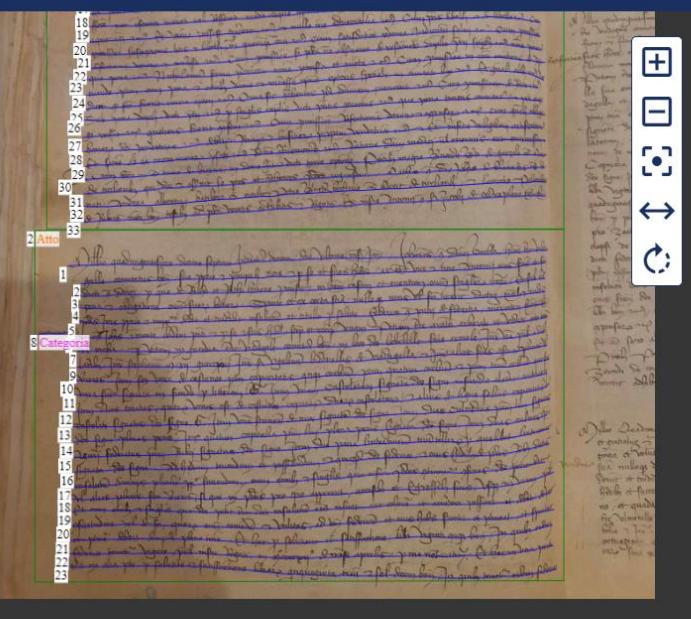
alios suos confines plures vel

veriores. Item

sex vaxa de castaneo partim capacitatibus quinque corbium partim quatuor corbium et partim

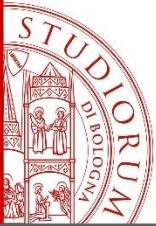
trium corbium quescumque

vaxa sunt sianata in fundo per litteram .G. Item unam castelatam sianatam dicto siano infundo



Structural Tag: Act

Textual tags: author, residence and etc.



Transkribus and the MemoBo database



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PROGETTO DI RICERCA
MEMOBO

Exporting the tagged material in XML

Screenshot of the Transkribus software interface showing the export dialog for a manuscript page.

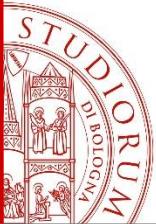
Export document dialog:

- Server export | Client export** (selected)
- TRAINING_VALIDATION_SET_MemoBo, Me. 69, Model 4 (1329791)**
- Current document Current collection
- Choose documents to export...**
- Choose export formats:**
 - Transkribus Document
 - PDF
 - TEI
 - DOCX
 - Simple TXT
 - Tag Export (Excel)
 - Tag Export (IOB)
 - Table Export into Excel
 - Page metadata into Excel
 - Export ALL formats
 - Export Selected as ZIP
- Export options:**
 - Export Page
 - Export ALTO
 - Export ALTO (Split Lines Into Words)
 - Export Image
 - Export text files
 - Export structural elements to Mets
 - Image type:** Original
 - Version status:** Latest version
 - Use word layer
 - Do blackening
 - Create Title Page
- Filename pattern:**
 - pageNr + filename
 - filename (warning: filenames must be unique for each page)
 - docId + pageNr + pageld**Pattern:** \${filename}
Placeholder: \${docId}, \${filename}, \${pageld}, \${pageNr}
- Pages:** (6) 1-6
- OK** | **Cancel**

The main window shows a photograph of a medieval manuscript page with handwritten text in two columns. A large blue arrow points from the export dialog towards the right side of the screen, where a smaller window titled "Tagname Choice" is displayed.

Tagname Choice window:

- Export all tags in doc
- Choose tags for export
- Select Tags**



Transkribus and the MemoBo database



PROGETTO DI RICERCA
MEMOBO

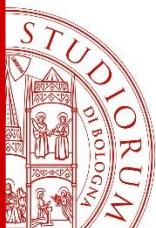
```
inferiis descriptis et specificatis
</ab>
<ab facs="#facs_2_TextRegion_1693516038427_1541" type="Atto">
...
<ab>
<ab facs="#facs_2_TextRegion_1693516053682_1549" type="Atto">
...
<ab facs="#facs_3_tr_2" type="Atto">
Millesimo trecentesimo septuagesimo sexto indictione quarta decima die vigesimo nono mensis
<Autore>Nicolaus quondam</Autore>
<lb facs="#facs_3_tr_2_t1_2" n="N002"/>
<Autore>magistri Johannis de Barberis notarius communis Bononie</Autore>
<Residenza>capella sancti Andree de Ansaldi</Residenza>
sponte etc per se et suos heredes titulio et
<lb facs="#facs_3_tr_2_t1_3" n="N003"/>
ex causa donationis mere libere simpliciter et inrevocabiliter inter vivos
<Tipologia>cessit transtulit et mandavit</Tipologia>
<lb facs="#facs_3_tr_2_t1_4" n="N004"/>
<Destinatario>domine Zanne quondam Calderini de Caldaria civi bononiensi</Destinatario>
capelle
<lb facs="#facs_3_tr_2_t1_5" n="N005"/>
ibidem presenti per se et suis heredibus stipulanti
<lb facs="#facs_3_tr_2_t1_6" n="N006"/>
et recipienti
<Diritti> omnia iura et actiones reales et personales</Diritti>
utiles et directas aypotecarias atque mistas et alterius cuiuscumque
<lb facs="#facs_3_tr_2_t1_7" n="N007"/>
Cessionis ex causa
<lb facs="#facs_3_line_1693553061288_358" n="N008"/>
donationis
<lb facs="#facs_3_line_1693553061288_357" n="N009"/>
generis que et quas ipse Nicolaus habet et habere potest et posset et eidem spectant et compre-
<lb facs="#facs_3_tr_2_t1_9" n="N010"/>
rentur et posset ac eidem quesita forent in futurum contra et adversus
<Altro>
...
<Altro>
```



```
<c id="IT-UNIBO-REG001-005758" level="fonds">
  <did>
    <unitid countrycode="IT"/>
    <unittitle>Scollegati</unittitle>
  </did>
  <processinfo>
    <list>
      <item>
        <date>15/05/2023</date>
        <persname>Admin xDams - open source</persname>
        inserimento
      </item>
    </list>
  </processinfo>
  <use>
    <c id="IT-UNIBO-REG001-002160" level="item">
      <controlaccess altrender="persone">
        <list>
          <item>
            <geoname altrender="provenienza">terra Pollicini, de</geoname>
            <persname altrender="genere">M</persname>
            <persname altrender="nome">Morinus f. qd. Nigri</persname>
            <persname role="funzione">autore</persname>
          </item>
          <item>
            <geoname altrender="provenienza">terra Pollicini, de</geoname>
            <persname altrender="genere">F</persname>
            <persname altrender="nome">Blonda f. Nigri</persname>
            <persname role="funzione">autore</persname>
          </item>
          <item>
            <persname altrender="cognome">Manticis, de</persname>
            <persname altrender="genere">M</persname>
            <persname altrender="nome">Gerardus qd. d. Tolomey</persname>
            <persname role="funzione">destinatario</persname>
          </item>
        </list>
      </controlaccess>
      <note>
        <p>i medesimi dell'atto immediatamente precedente (Compravendita c. 345r. n. 2)</p>
      </note>
      <persname altrender="genere">M</persname>
      <persname altrender="nome">testes</persname>
      <persname role="funzione">testimone</persname>
    </item>
  </list>
</controlaccess>
```

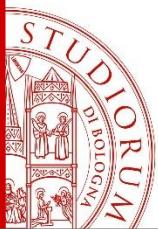
iTranskribus with tags structure in XML

Xdams structure in XML



Some (partial) conclusions of the work in progress:

- *Transkribus* is very sensitive to the change of hands even inside the same volume of the *memoriali*
- The software finds it very difficult to transcribe words it has never seen (necessity to augment the sample)
- The homogeneity of the script seems to be more important than the abbreviation system (the software learnt quickly how to associate abbreviations with the expanded words in the direct transcription)



Questions?

edward.loss2@unibo.it