



Key-Value Information Extraction from Full Handwritten Pages

Solène Tarride, Mélodie Boillet and Christopher
Kermorvant

Teklia, Paris, France

ICDAR 2023

August 21, 2023

T E K L I A

Key-value information extraction

B.4.
Marie
Alice
Cora
Charlotte
Dufault.

Le vingt-sept février mil neuf cent un, sous seing privé, prêtre curé, de cette paroisse, avons baptisé Marie Alice Cora Charlotte, née la veille fille légitime de Napoléon Dufault, et de Alice Mercier de cette paroisse. Le parrain a été Pierre Dufault, grand père de l'enfant et la marraine son épouse Marie Larolle qui ainsi que le père ont signé avec nous; le parrain a déclaré ne savoir signer. Lecture faite.

Marie Larolle
Napoléon Dufault
J. D. Bernier (pasteur)

Key	Value
Child name	Marie Alice Cora Charlotte
Birth date	26/02/1901
Father name	Napoléon
Father last name	Dufault
Mother name	Alice
Mother last name	Mercier

Use-cases

- Extract information from different types of documents
 - Forms
 - Records
 - Tables
 - Finding aids
 - ...

NOMS DES ACCUSÉS	ÂGE	QUALITÉS	DEMEURES	JURISDICTIONS	ACCUSATIONS	PENES PROPOSÉES PAR LES JURY	DECRET DU ROI	ARRÊTES DE
A.
B.
C.
D.
E.
F.
G.
H.
I.
J.
K.
L.
M.
N.
O.
P.
Q.
R.
S.
T.
U.
V.
W.
X.
Y.
Z.

Troisième feuillet

S. 2.
Lea
Breault,

Le douze février mil neuf cent un nous soussigné, prêtre, curé, de cette paroisse, avons inhumé dans le cimetière de cette paroisse, le corps de Lea fille de feu Jean Baptiste Breault et de Azilda l'épouse de cette paroisse, décédée le avant veille, âgé de dix ans. Présents à l'inhumation Jean Breault et Eliophas Blanchette qui ont signé avec nous - Lecteur pasteur -

Jean Breault
E. D. Bernier (curé)

B. 4.
Marie
Alice
Cora
Charlotte
Dufault.

Le vingt-sept février mil neuf cent un nous soussigné, prêtre curé, de cette paroisse, avons baptisé Marie Alice Cora Charlotte, née la veille fille légitime de Napoléon Dufault, et de Alice Mercier de cette paroisse. Le parrain a été Pierre Dufault, grand père de l'enfant et la marraine son épouse Marie Larallée qui avec nous, le parrain a déclaré avec nous, le parrain a déclaré

Amadouche 1810

1916

ÉTAT CIVIL

N. le ... à ...

PROFESSEUR

Marié à ...

DÉCISION DU CONSEIL DE REVISION ET MOTIFS.

Inscrit sous le n° ... de la liste de cession d ...

DÉTAIL DES SERVICES ET MUTATIONS DIVERSES.

1° ...

2° ...

3° ...

4° ...

5° ...

6° ...

7° ...

8° ...

9° ...

10° ...

11° ...

12° ...

13° ...

14° ...

15° ...

16° ...

17° ...

18° ...

19° ...

20° ...

21° ...

22° ...

23° ...

24° ...

25° ...

26° ...

27° ...

28° ...

29° ...

30° ...

31° ...

32° ...

33° ...

34° ...

35° ...

36° ...

37° ...

38° ...

39° ...

40° ...

41° ...

42° ...

43° ...

44° ...

45° ...

46° ...

47° ...

48° ...

49° ...

50° ...

51° ...

52° ...

53° ...

54° ...

55° ...

56° ...

57° ...

58° ...

59° ...

60° ...

61° ...

62° ...

63° ...

64° ...

65° ...

66° ...

67° ...

68° ...

69° ...

70° ...

71° ...

72° ...

73° ...

74° ...

75° ...

76° ...

77° ...

78° ...

79° ...

80° ...

81° ...

82° ...

83° ...

84° ...

85° ...

86° ...

87° ...

88° ...

89° ...

90° ...

91° ...

92° ...

93° ...

94° ...

95° ...

96° ...

97° ...

98° ...

99° ...

100° ...

ABRAM (Etienne Jean), prêtre, curé de la paroisse de St. Jean de Pont. (Vendée)

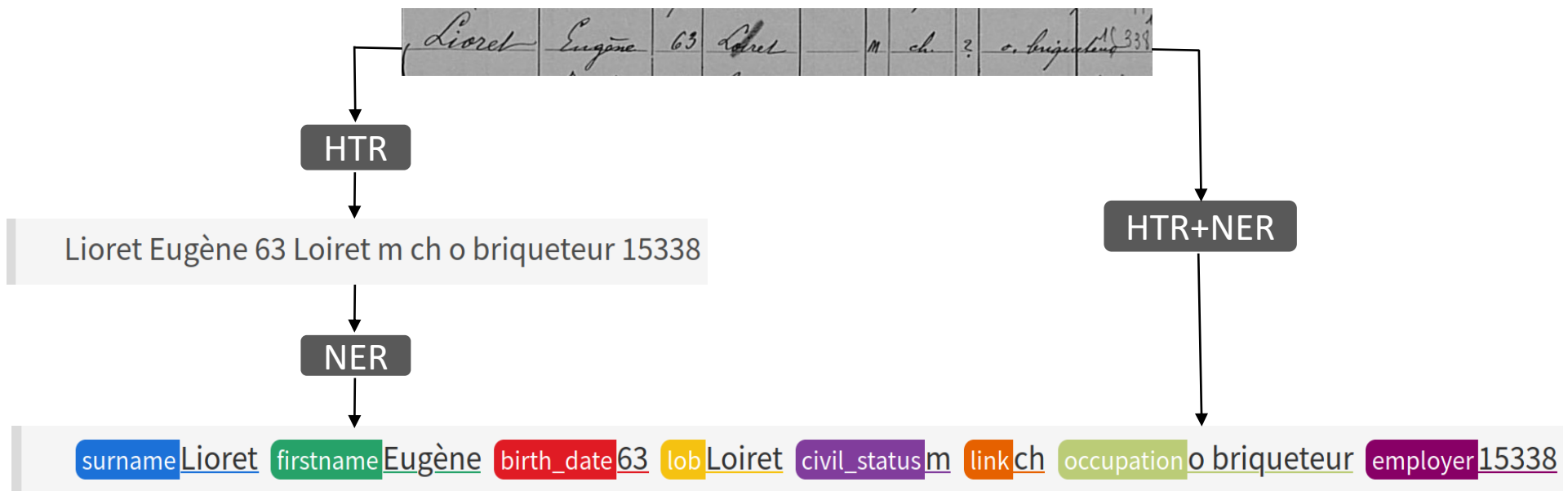
Jacques PELLOQUIN

23 mars 1770

X^{1A} 4678, 1020

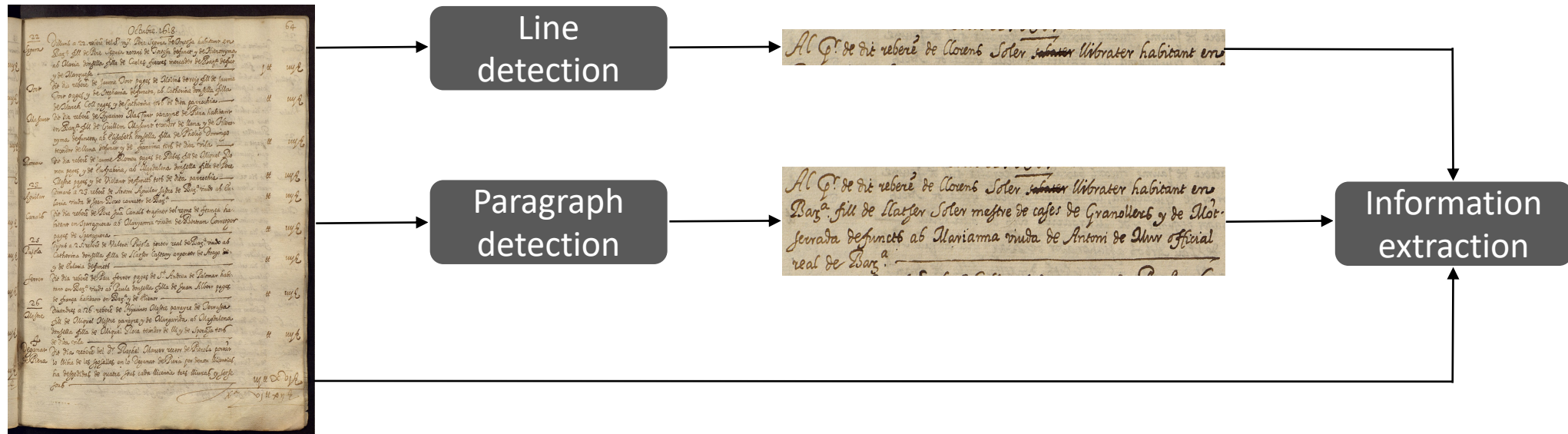
Research questions

- Should HTR and NER be performed jointly or in separate steps?



Research questions

- Should HTR and NER be performed jointly or in separate steps?
- Can we directly process full pages? Are line/paragraph detection and recognition models still needed?



Research questions

- Should HTR and NER be performed jointly or in separate steps?
- Can we directly process full pages? Are line/paragraph detection and recognition models still needed?
- Are end-to-end models able to learn from partial key-value annotations?

Name	Surname	Birth date
Gammelin	Robert	2-12-02
Rouillat	Paul	16-10-02
Laforêt	Alfred	10-5-10
Proffit	Marcel	10-4-13
...

Experimental setup

Datasets

	IAM [1]	ESPOSALLES [2]	POPP [3]
Type	English, free-text, modern	Catalan, records, historical	French, tables, modern
Pages	747 pages	125 pages	160 pages
Writers	500 writers	1 writer	Unknown
Levels	Line and page	Line, record and page	Line and page
Entities	18 entities (event, date, ...)	6 categories (name, surname, ...) + 8 persons (wife, husband, ...)	10 entities (surname, name, ...)
Entities/words	<10%, unstructured (free-text)	>50%, semi-structured	100%, structured

[1] Marti, UV., Bunke, H. (2002). The IAM-database: an English sentence database for offline handwriting recognition. *IJDAR* 5, 39–46 (2002). <https://doi.org/10.1007/s10032020007>

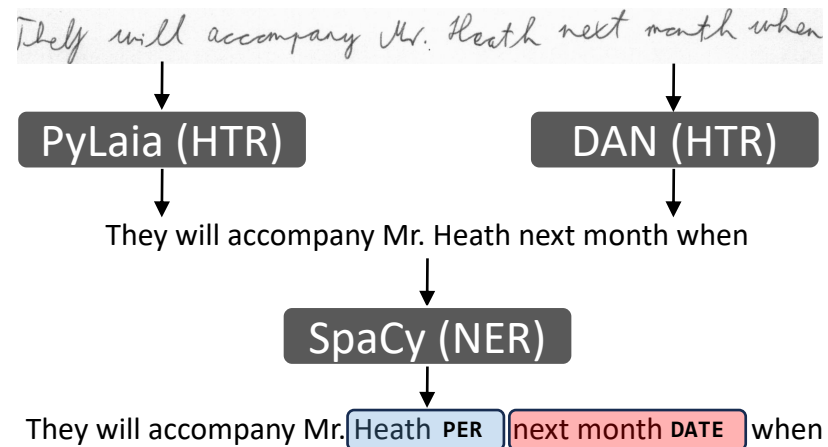
[2] Verónica Romero *et al.* (2013). The ESPOSALLES database: An ancient marriage license corpus for off-line handwriting recognition. *Pattern Recognition*, 46(6), 1658-1669.

[3] Thomas Constum *et al.* (2022). Recognition and Information Extraction in Historical Handwritten Tables: Toward Understanding Early 20th Century Paris Census. In: Uchida, S., Barney, E., Eglin, V. (eds) Document Analysis Systems. DAS 2022. Lecture Notes in Computer Science, vol 13237. Springer, Cham. https://doi.org/10.1007/978-3-031-06555-2_10

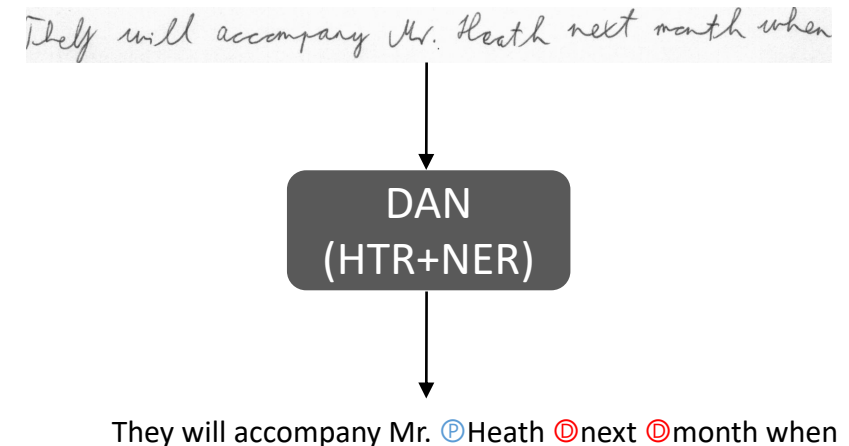
Should HTR and NER be performed jointly or in separate steps?

Method

- Sequential approach
 - DAN [4] or PyLaia [5] for HTR
 - SpaCy [6] for NER



- End-to-end approach
 - DAN [4] for HTR+NER
 - Special tokens are used to tag entities



[4] Denis Coquenat, Clement Chatelain, & Thierry Paquet (2023). DAN: a Segmentation-free Document Attention Network for Handwritten Document Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–17.

[5] Joan Puigcerver. Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition ? In : *14th ICDAR*, p. 67-72.

[6] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, Adriane Boyd. SpaCy: Industrial-strength Natural Language Processing in Python. 2020. 10.5281/zenodo.1212303

Results

- Sequential and end-to-end models yield similar results

Method	ESPOSALLES				IAM			POPP		
	CER (%) ↓	F1 (%) ↑		Level	CER (%) ↓	F1 (%) ↑	Level	CER (%) ↓	F1 (%) ↑	Level
		<i>Person</i>	<i>Category</i>							
GT + SpaCy	0.00	98.5	98.5	Record	0.0	75.5	Page	0.0	96.4	Line
PyLaia + SpaCy	0.76	94.9	95.0	Line	7.79	49.0	Line	17.19	76.3	Line
DAN + SpaCy	0.46	97.9	97.8	Record	4.30	57.9	Page	8.18	84.0	Page
DAN end-to-end	0.39	98.1	97.6	Record	5.12	33.7	Page	7.83	85.9	Page

Results

- Sequential and end-to-end models yield similar results

Sequential methods

- Pros
 - Benefit from pre-trained NER embeddings
- Cons
 - Two models to train
 - Errors accumulate
 - No information about word localization

→ Any type of documents

End-to-end methods

- Pros
 - Train only one model
 - Larger context (localization ↔ text ↔ entities)
- Cons
 - Train almost from scratch
 - Not efficient with a low entities/words ratio

→ Tables, forms, records

Can we directly process full pages?
Are line/paragraph detection and recognition
models still needed?

Method

- Comparing HTR+NER models trained at different levels

	IAM	ESPOSALLES	POPP
Line*	✓	✓	✓
Record*		✓	
Page	✓	✓	✓

* Note:

- Models have been trained on ground truth polygons → not fair
- Fully automatic workflow would introduce errors (line/record detection + reading-order retrieval)

Results

- Page-level recognition is possible and efficient
- Pros: remove the need for detection models (simplify workflow and avoid error accumulation)

Method	ESPOSALLES				IAM			POPP		
	CER (%) ↓	F1 (%) ↑		Level	CER (%) ↓	F1 (%) ↑	Level	CER (%) ↓	F1 (%) ↑	Level
		<i>Person</i>	<i>Category</i>							
DAN end-to-end	0.48	96.1	96.9	Line*	5.12	33.7	Line*	7.83	85.9	Line*
	0.39	98.1	97.6	Record*	-	-	-	-	-	-
	3.61	96.6	95.9	Page	4.82	31.3	Page	11.74	85.3	Page

* Note:

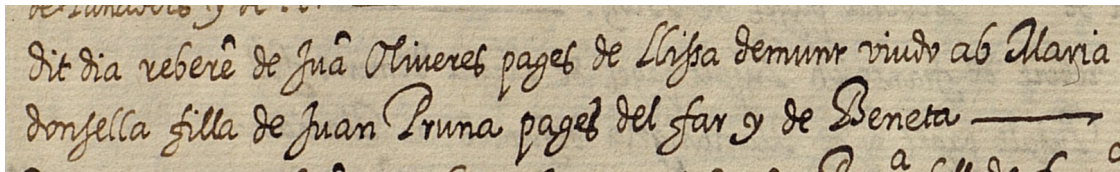
- Models have been trained on ground truth polygons → not fair
- Fully automatic workflow would introduce errors (line/record detection + reading-order retrieval)

Are end-to-end models able to learn from partial key-value annotations?

Method

- Are end-to-end models able to learn from key-value annotations?

Esposalles → 50% of words are entities → learning from key-value annotations could work

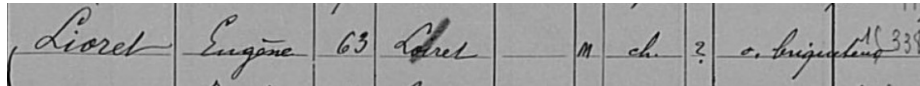


HTR	Dit dia rebere de Jua Oliveres pages de Lissa demunt viudo ab Maria donfella filla de Juan Pruna pages del far y de Beneta
HTR+NER	Dit dia rebere de <N-H>Jua <SN-H>Oliveres <O-H>pages de <L-H>Lissa demunt <S-H>viudo ab <N-W>Maria <S-W>donsella filla de <N-WF>Juan <SN-WF>Pruna <O-WF>pages del <L-WF>far y de <N-WM>Beneta
Key-value	<N-H>Jua <SN-H>Oliveres <O-H>pages <L-H>Lissa <S-H>viudo <N-W>Maria <S-W>donsella <N-WF>Juan <SN-WF>Pruna <O-WF>pages <L-WF>far <N-WM>Beneta

Method

- Are end-to-end models able to learn from key-value annotations?

POPP → 100% of words are entities → key-value = HTR+NER

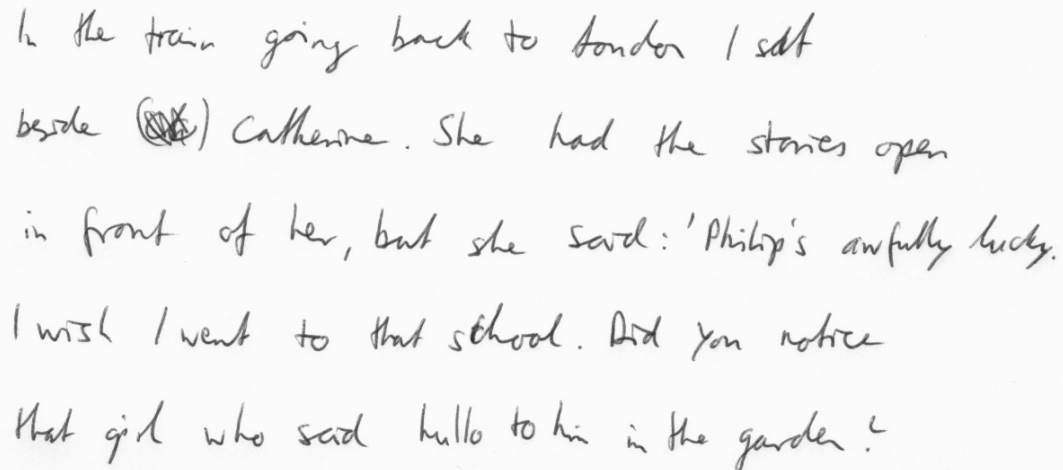


HTR	Lioret Eugène 63 Loiret m ch o briqueteur 15338
HTR+NER	<sn>Lioret <fn>Eugène <a>63 <l>Loiret <s>m <l>ch <o>o <o>briqueteur <n>15338
Key-value	<sn>Lioret <fn>Eugène <a>63 <l>Loiret <s>m <l>ch <o>o <o>briqueteur <n>15338

Method

- Are end-to-end models able to learn from key-value annotations?

IAM → <10% of words are entities → learning from key-value annotations will be difficult



In the train going back to London I sat beside ~~(the)~~ Catherine. She had the stories open in front of her, but she said: 'Philip's awfully lucky. I wish I went to that school. Did you notice that girl who said hullo to him in the garden?'

HTR

In the train going back to London I sat beside # Catherine. She had the stories open in front of her, but she said: 'Philip's awfully lucky. I wish I went to that school. Did you notice that girl who said hullo to him in the garden?'

HTR+NER

In the train going back to <gpe>London I sat beside # <per>Catherine . She had the stories open in front of her, but she said: ' <per>Philip 's awfully lucky. I wish I went to that school. Did you notice that girl who said hullo to him in the garden?'

Key-value

<gpe>London <per>Catherine <per>Philip

Results

- Key-value learning is efficient, although not as efficient as full text learning
- **Pros:** (a lot) easier to collect ground truth (form filling)
- **Cons:** only works on structured documents, with a sufficient ratio of entities/words

Method	ESPOSALLES			IAM	
	F1 (%) ↑		Level	F1 (%) ↑	Level
	<i>Person</i>	<i>Category</i>			
PyLaia + SpaCy	94.9	95.0	Record	52.5	Page
PyLaia + DAN	97.9	97.8	Record	61.8	Page
DAN end-to-end	98.1	97.6	Record	31.3	Page
DAN key-value	97.2	96.9	Record	0.0	Page

Conclusion

Conclusion

- Transformers are able to perform many tasks in a single step
 - Recognize text from full-pages
 - Understand the layout [4]
 - Tag named entities
 - **[under conditions]** Learn to extract specific key-value information, ignoring irrelevant text
- Information extraction can be done in one or two steps, depending on the document structure
- Training from key-value information simplifies the annotation process (form-filling)

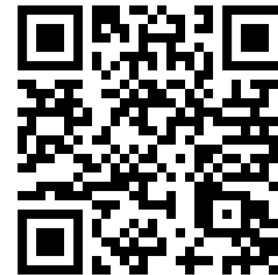
Thank you for your attention!



Our paper



New database
for key-value IE
(Simara)



Annotation tool
(Callico)