

Original Text



mT5 NER-PG

Pseudonymized Text



GELECTRA NER (F1-score = 0.97)

On 27 Sep., 13:45, Wilfried Viebeg <hors  
Sorry - das war mir entgangen

In der Kürze deiner Mitteilungen ist es  
allerdings ein Kunststück,  
einen Tippfehler zu verstecken.

Gruß  
Thea

### Original Email Text

	Token_ID	Label	Start	End	Token
0	T1	DATE	3	10	27 Sep.
1	T2	MALE	19	27	Wilfried
2	T3	FAMILY	28	34	Viebeg
3	T4	FEMALE	177	181	Thea

### Original Annotations

On 25 Sep., 13:45, Frédéric Clahter <hors  
Sorry - das war mir entgangen

In der Kürze deiner Mitteilungen ist es  
allerdings ein Kunststück,  
einen Tippfehler zu verstecken.

Gruß  
Kaspar

### mT5 NER-PG Pseudonymized Email Text

	Token_ID	Label	Start	End	Token	Pseudonym
0	T1	DATE	3	10	27 Sep.	25 Sep.
1	T2	MALE	19	27	Wilfried	Frédéric
2	T3	FAMILY	28	34	Viebeg	Clahter
3	T4	FEMALE	177	181	Thea	Kaspar

### mT5 NER-PG Annotations with Pseudonyms

Operations  
DATE [TP] += 1  
MALE [TP] += 1  
FAMILY [TP] += 1  
FEMALE [TP] += 1  
**For NER,**

For every mis-match, for PG

Label [FP] = Label [FP] + 1

Label [FN] = Label [FN] + 1

instead of in NER case,  
Label [TP] = Label [TP] + 1

to penalize Precision, Recall and F1-score  
of type compliance of **P**seudonym **G**eneration  
component of mT5 NER-PG model

	Token_ID	Label	Start	End	Token
0	T1	DATE	3	10	25 Sep.
1	T2	MALE	19	27	Frédéric
2	T3	FAMILY	28	35	Clahter
3	T4	MALE	178	184	Kaspar

### GELECTRA NER Annotations

Operations  
DATE [TP] += 1  
MALE [TP] += 1  
FAMILY [TP] += 1  
FEMALE [TP] += 1  
**For PG,**

Penalty Operations  
FEMALE [FP] += 1  
FEMALE [FN] += 1