

End-to-End IE without Token-Level Supervision

Rasmus Berg Palm Technical University of Denmark

Dirk Hovy University of Copenhagen

Florian Laws Tradeshift

Ole Winther Technical University of Denmark

Let's create a travel concierge app.

60

Hi Freya, what can I do for you?

Freya

I need a flight from Oslo to JFK, landing tomorrow at 8pm.

Takes natural language input.

တေ

GO I found this flight for you:

Aug 9

GO

KLM

Oslo 12:55 PM

New York 7:20 PM

Proposes flights.

တေ

တေ

Behind the covers...

We use a flight search engine

The search engine accepts a fixed set of

fields, e.g. "from",

"to", "day", etc.

We return the top-1 hit

GO

I need a flight from Oslo to JFK, landing tomorrow at 8pm.

Hi Freya, what can I do for you?

GO

Aug 9

Oslo

12:55 PM

I found this flight for you:

Freya

New York

KLM

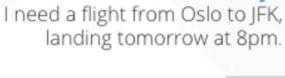
Behind the covers..

We hire human operators to extract the values for the search engine fields.

This labor is tedious for the operators and costly for us



GO Hi Freya, what can I do for you?





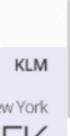
I found this flight for you:

Oslo

12:55 PM

Aug 9

JFK



Freya

So let's automate this extraction of information

Token level

TOROTT TO VOT		
Token	Label	
cheapest	B-PRICE	
oirforo		

airtare

from

tacoma B-FROM

to

st. B-TO

louis I-TO

Token level

Token	Label
cheapest	B-PRICE
airfare	0
from	0
tacoma	B-FROM
to	0
st.	В-ТО
louis	I-TO

Token	leve

Field

Field level

Value

tacoma

st. louis

cheapest

Token Label cheapest

airfare

tacoma

from

to

st.

louis

FROM

PRICE

MONTH

YEAR

AIRLINE

DAY

TO

Token level	
Token	Label

Field

Field level

cheapest B-PRICE

from

to

st.

louis

tacoma

FROM

tacoma

Value

airfare

Chunking

TO

st. louis cheapest

B-TO

I-TO

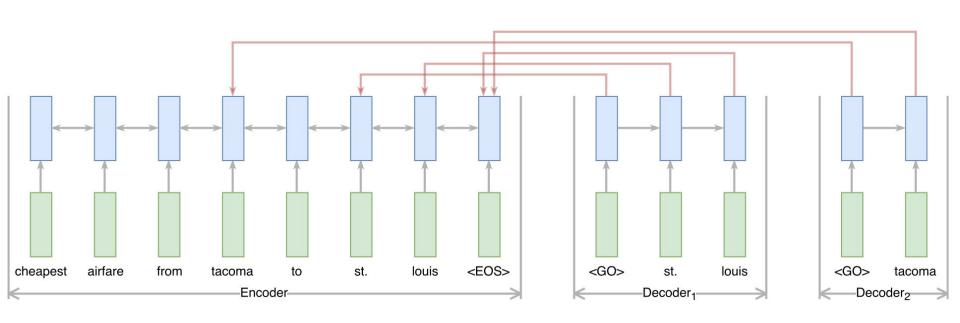
B-FROM

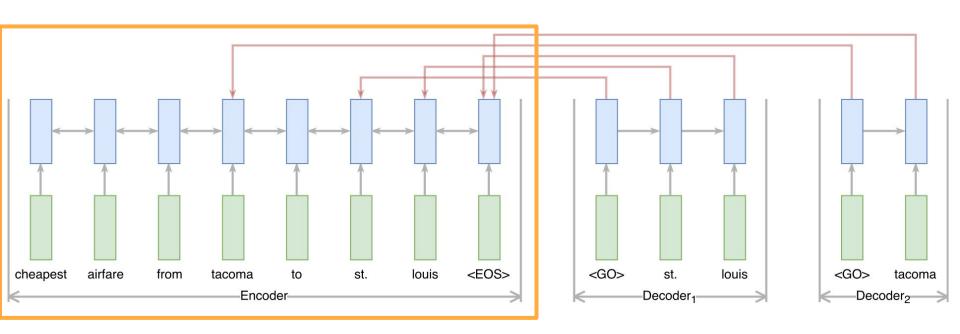
PRICE DAY

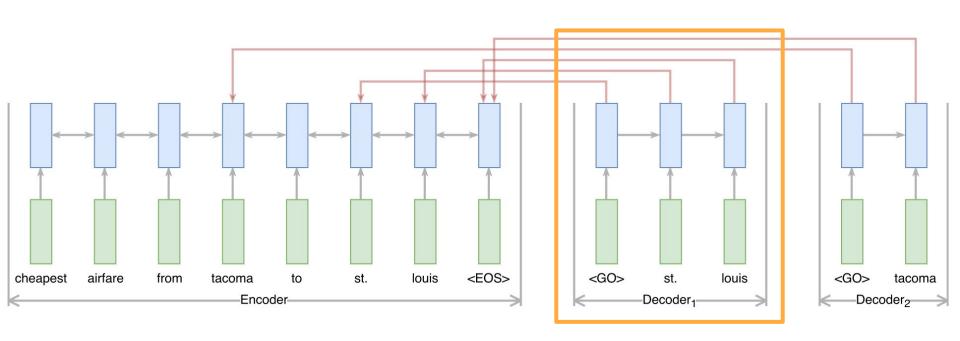
MONTH

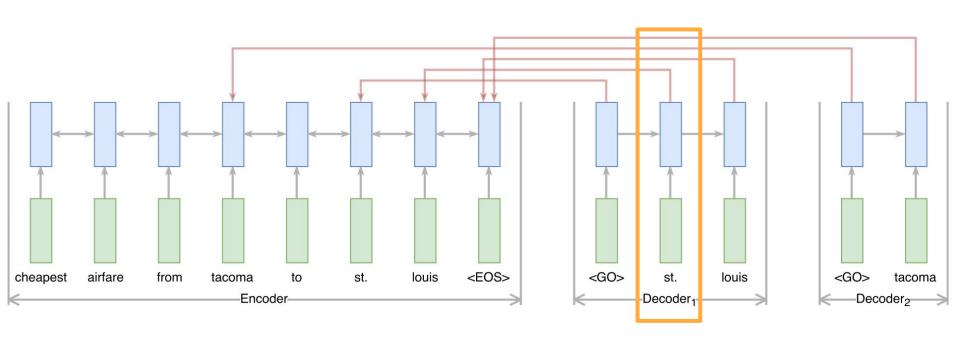
YEAR

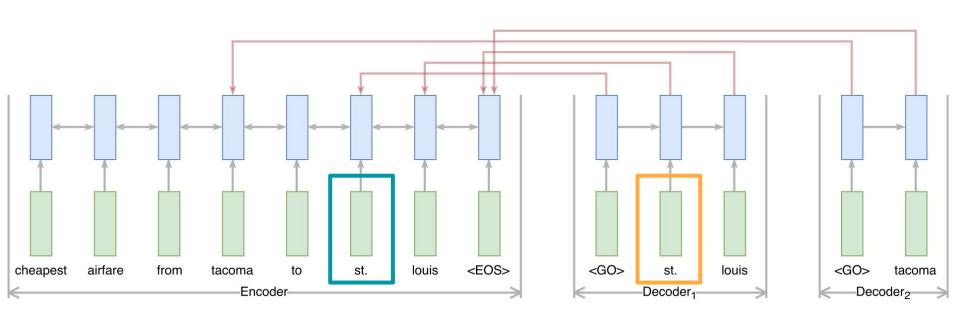
AIRLINE

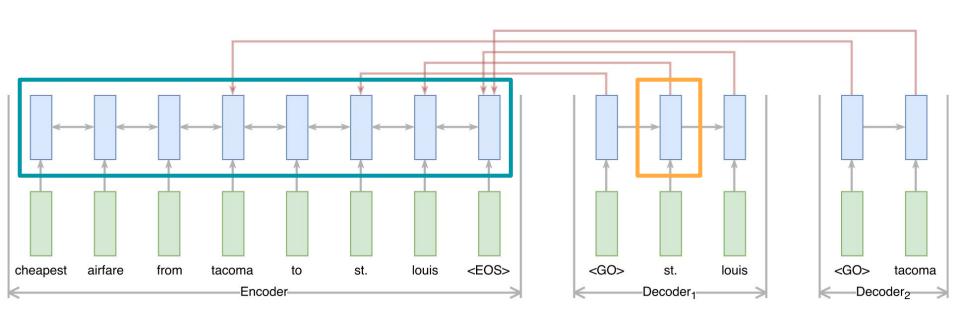


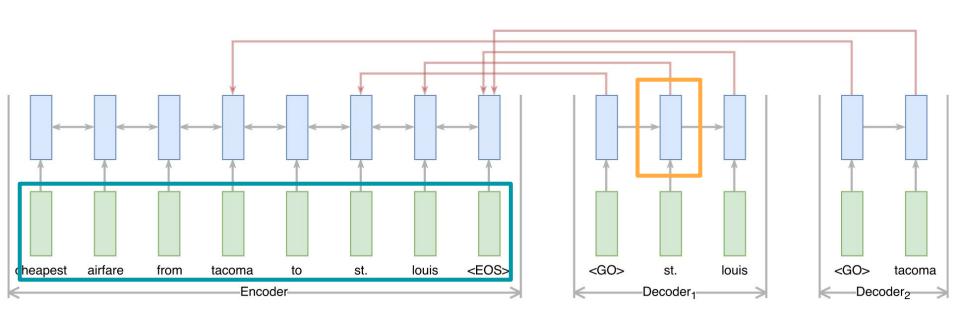


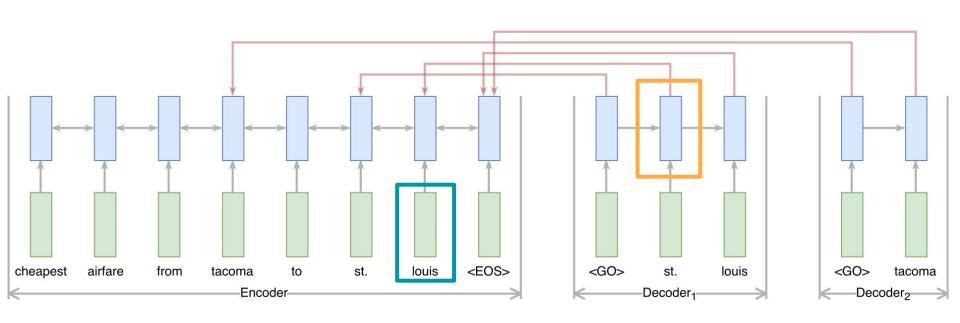


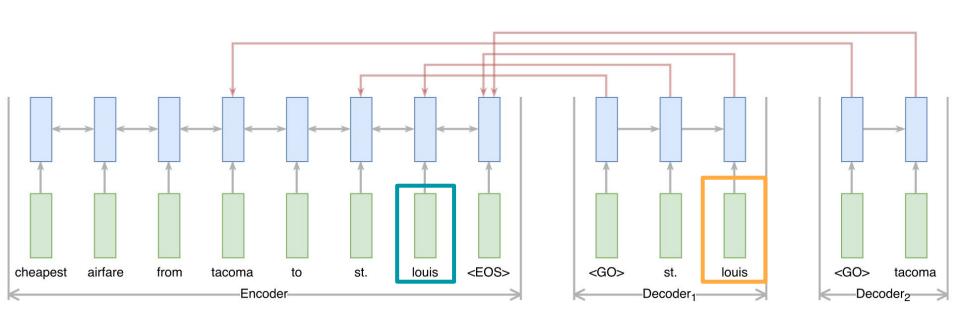


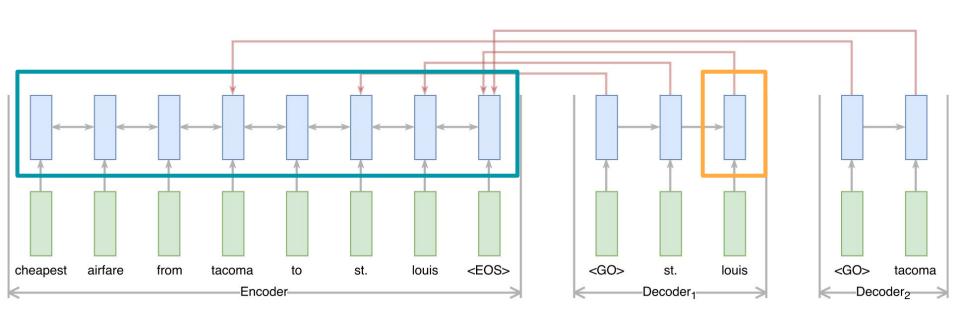


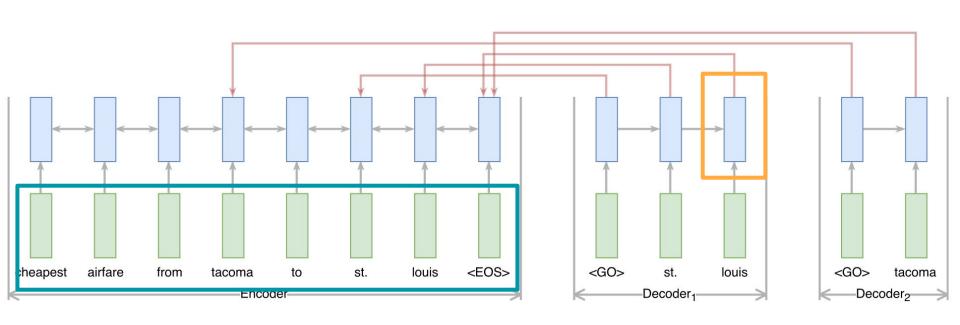


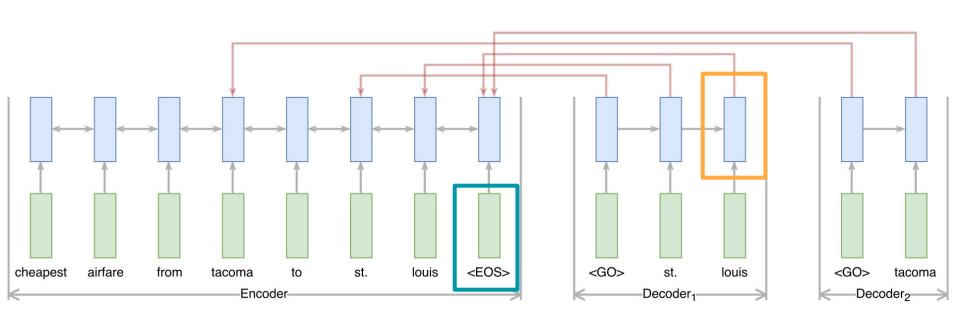










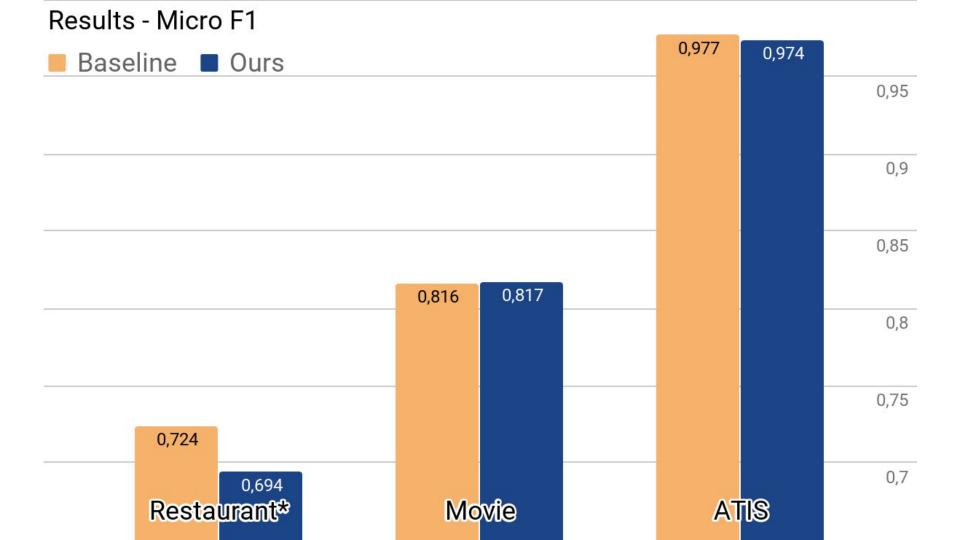


See the paper for details

arxiv.org/abs/1707.04913

Code is available at

github.com/rasmusbergpalm/e2e-ie-release



There's one major limitation Normalization

'17 Jan 2012' → '2012-01-17'

TRADESHIFF

Thank you!

Questions?