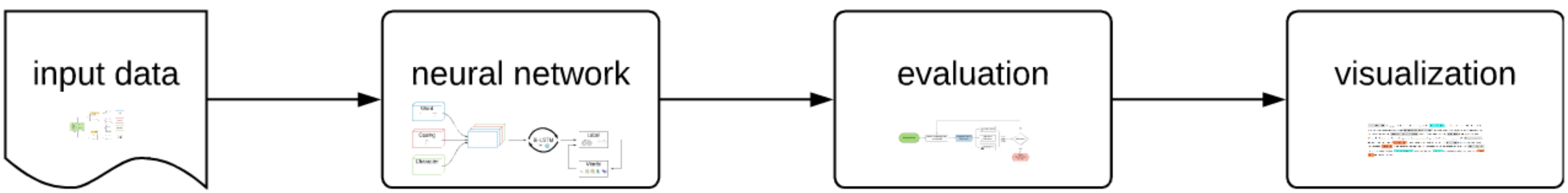
TUM Data Innovation Lab

Munich RE

Natural Language Processing with Deep Learning



Named Entity Recognition

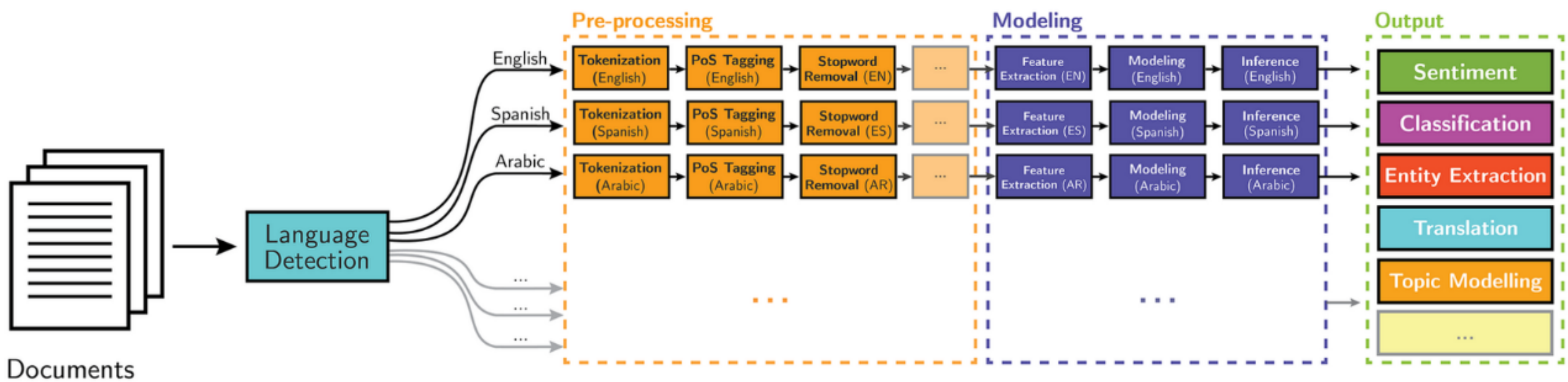
"named entities are real-world object, such as persons, locations, organizations, etc., that can be denoted with a proper name"

```
====================================================
=                    Input Text                    =
====================================================

Angela Merkel wohnt in Berlin

====================================================
=            Named Entity Visualization            =
====================================================
```

Angela Merkel **PER** wohnt in Berlin **LOC**

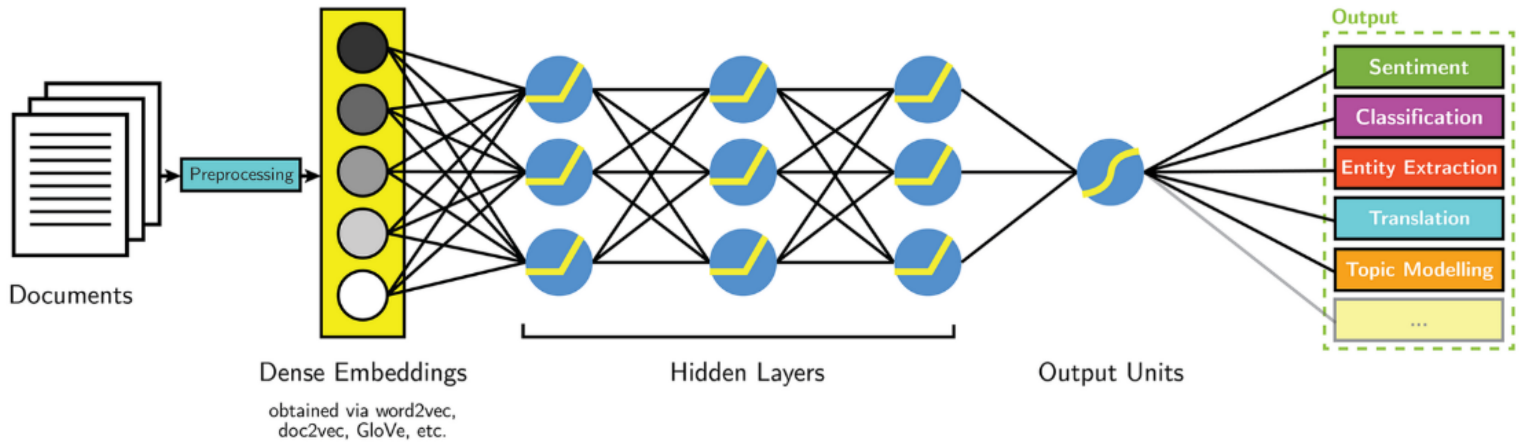
Classical Approaches

Example of Rule	Explanation
Rule: Person	
({Token.kind !=”number”})	Check whether it is number, to avoid address pattern is tagged.
(	
({Token.kind==word, Token.orth==upperInitial})	At least one Capital letter word
({Token.kind==word, Token.orth==upperInitial})?	? refer to Capital letter word exist anot
({Token.kind==word, Token.orth==upperInitial})?	? refer to Capital letter word exist anot
):label	
({Token.string==”,”})	
({Token.kind==”number”})	Number refer to the age
({Token.string==”;”) +	+ mean that repeat the pattern again
→	Match the LHS rules with RHS
:label.Person={rule=”Person”}	label as Person



Deep Learning

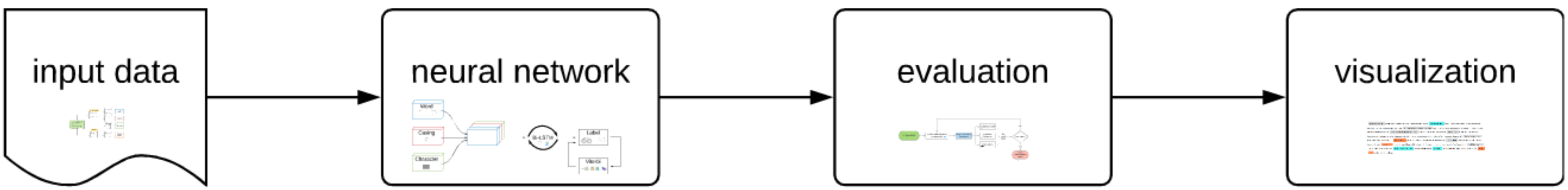
Infer features

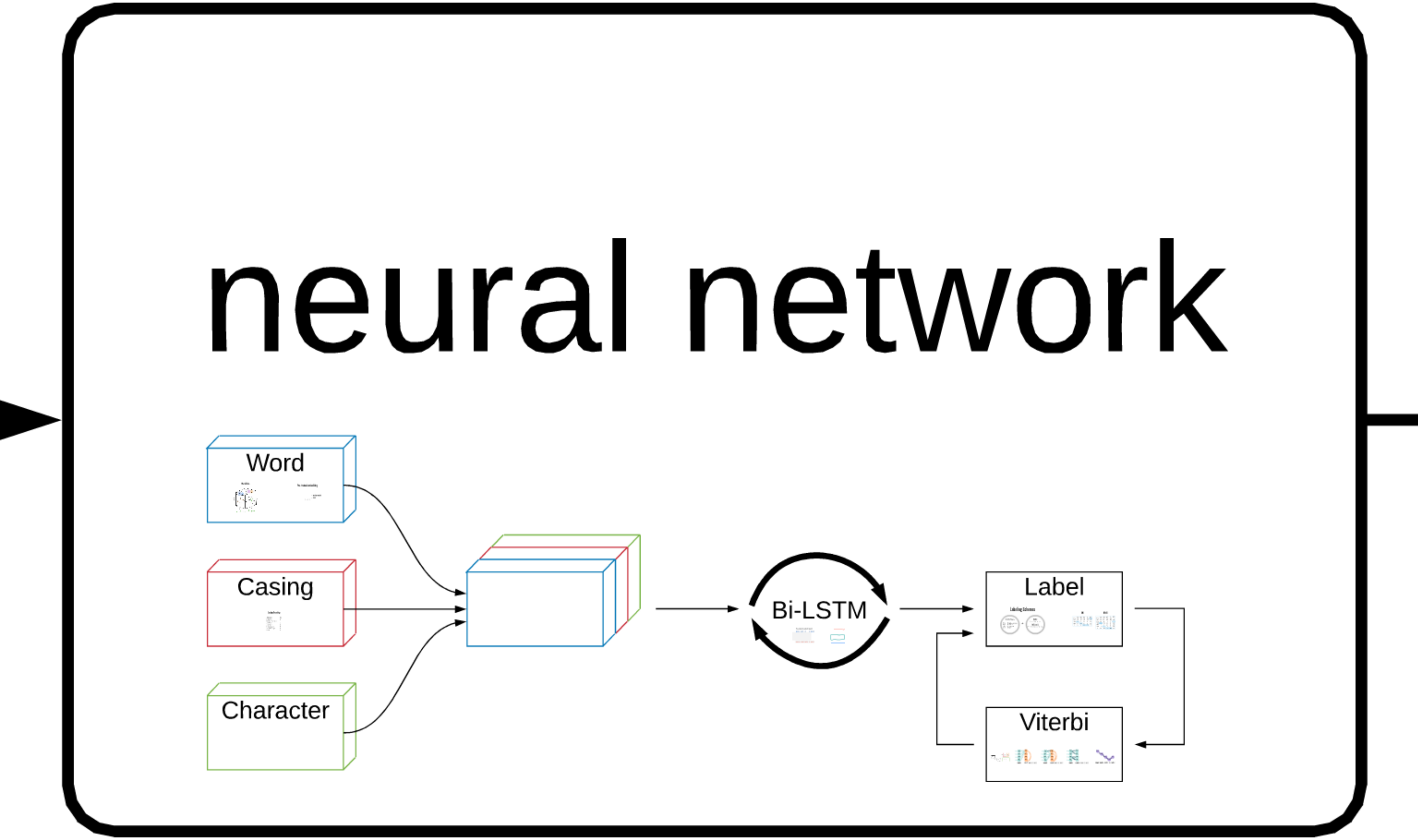


TUM Data Innovation Lab

Munich RE

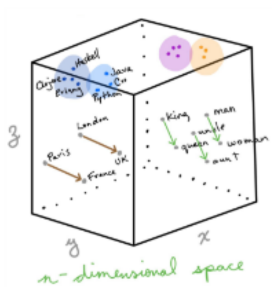
Natural Language Processing with Deep Learning




neural network
Word
Casing
Character
Bi-LSTM
Label
Viterbi

Word

Word2Vec

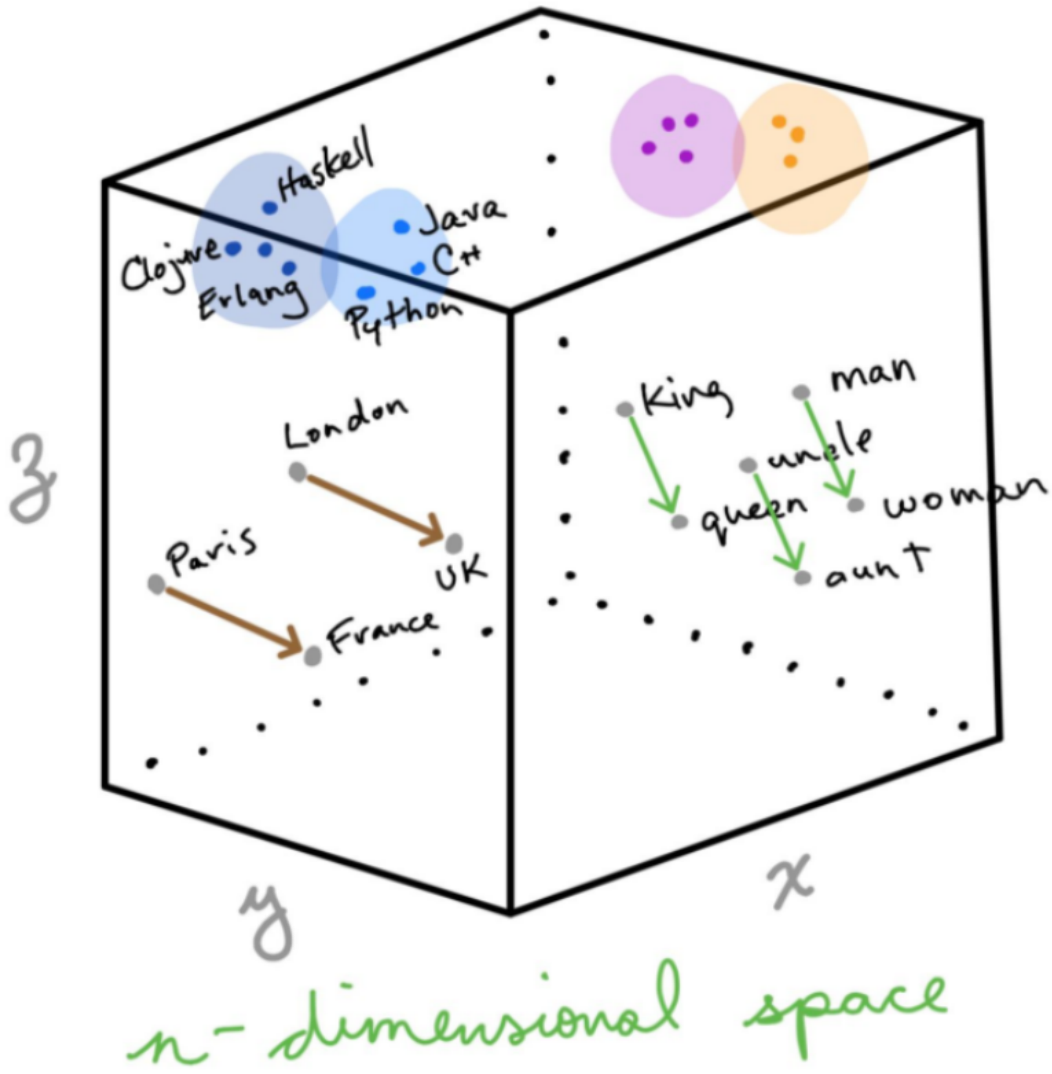


Pre-trained embedding

fastText by Facebook

- used as weights
- static

Word2Vec



Haskell
Clojure
Erlang
Java
C++
Python
London
UK
Paris
France
King
man
uncle
queen
woman
aunt
z
y
x
n-dimensional space

Pre-trained embedding

fastText
by Facebook

- used as weights
- static

Casing

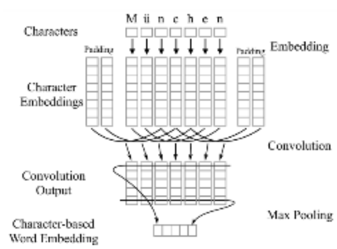
Casing Encoding

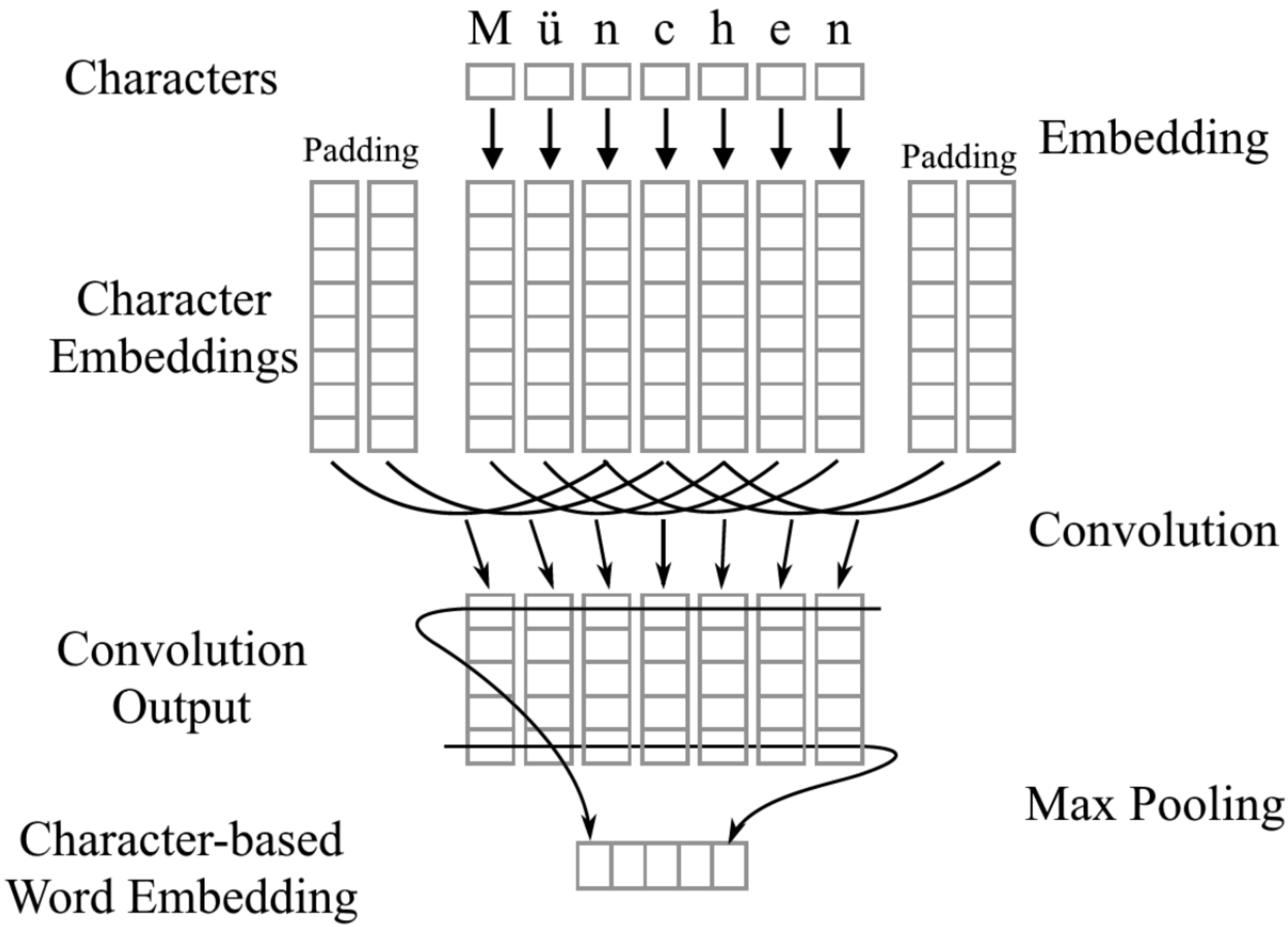
Options	ID
numeric	0
mainly_numeric	1
allLower	2
allUpper	3
initialUpper	4
contains_digit	5
other	6

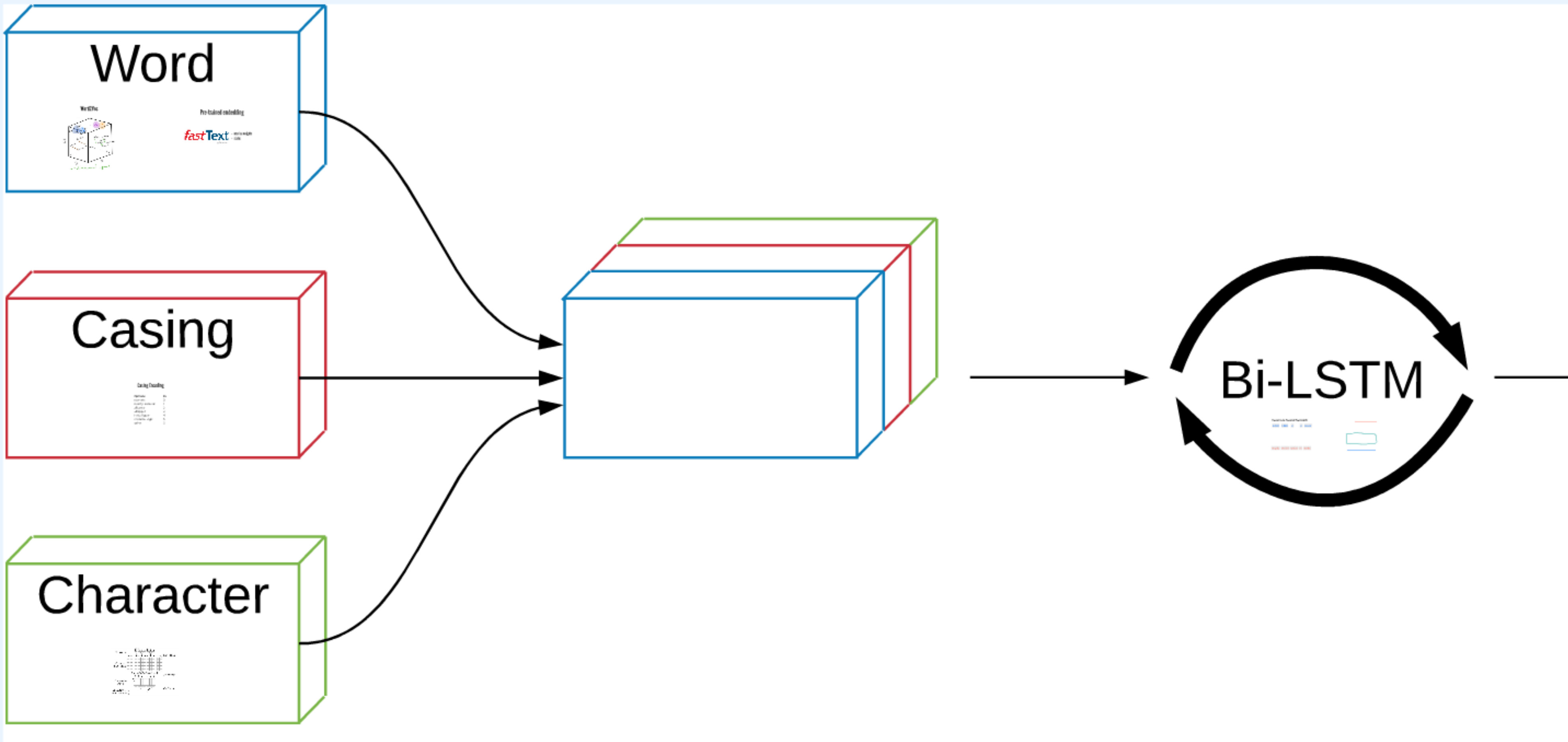
Casing Encoding

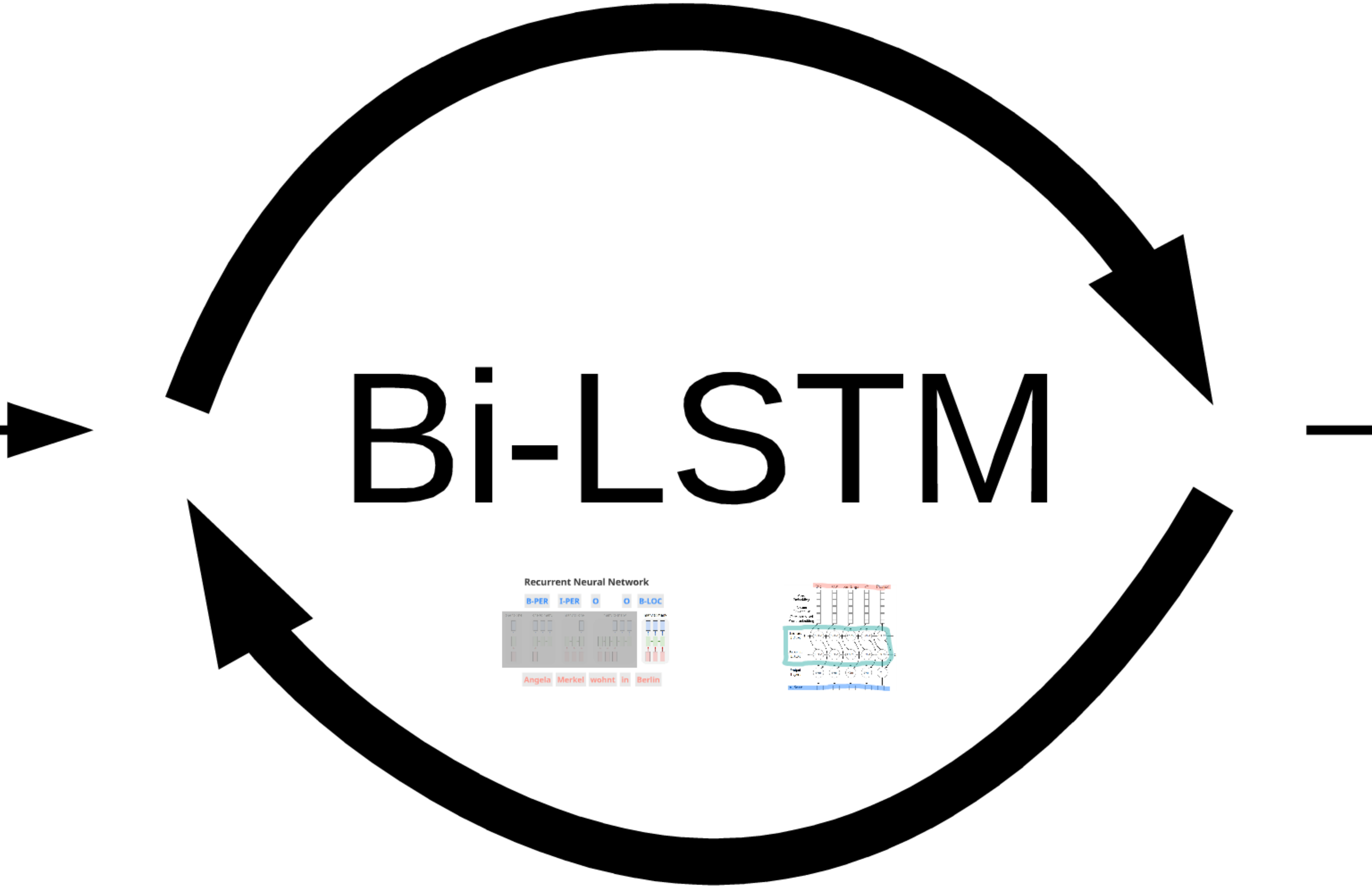
Options	ID
numeric	0
mainly_numeric	1
allLower	2
allUpper	3
initialUpper	4
contains_digit	5
other	6

Character


M ü n c h e n
Characters
Embedding
Character
Embeddings
Convolution
Convolution
Output
Max Pooling
Character-based
Word Embedding

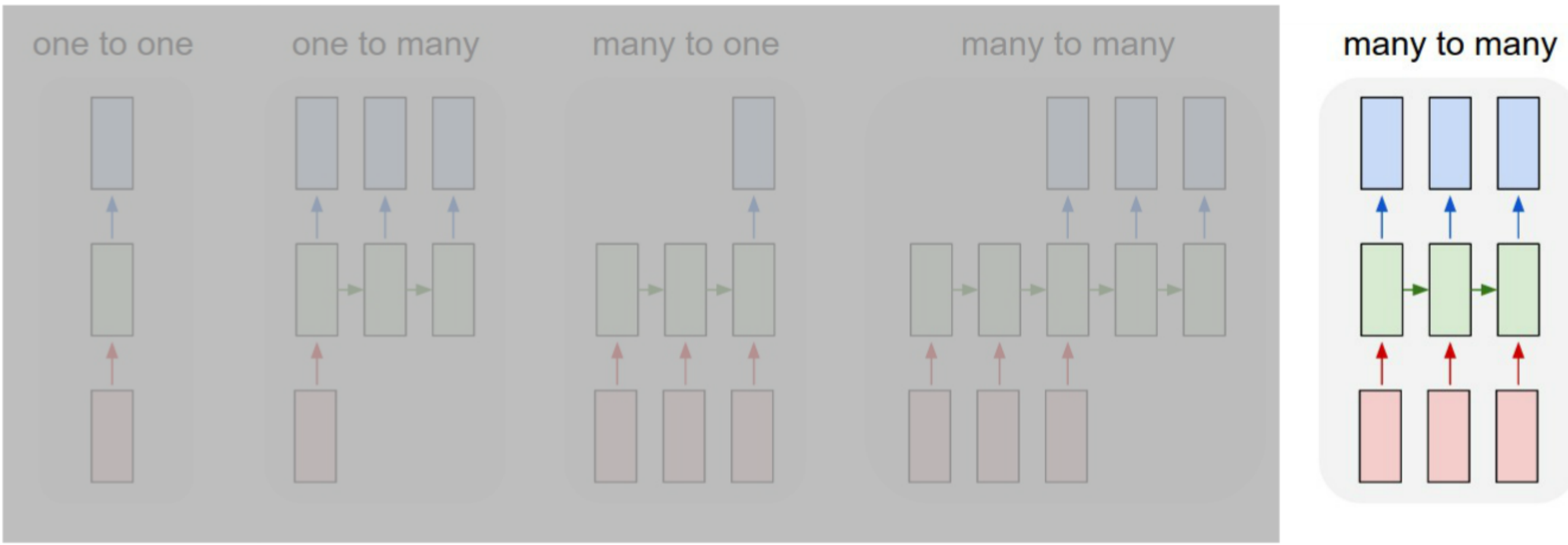

M ü n c h e n
Characters
Embedding
Padding
Padding
Character Embeddings
Convolution
Convolution Output
Max Pooling
Character-based Word Embedding


Word
Casing
Character
Bi-LSTM

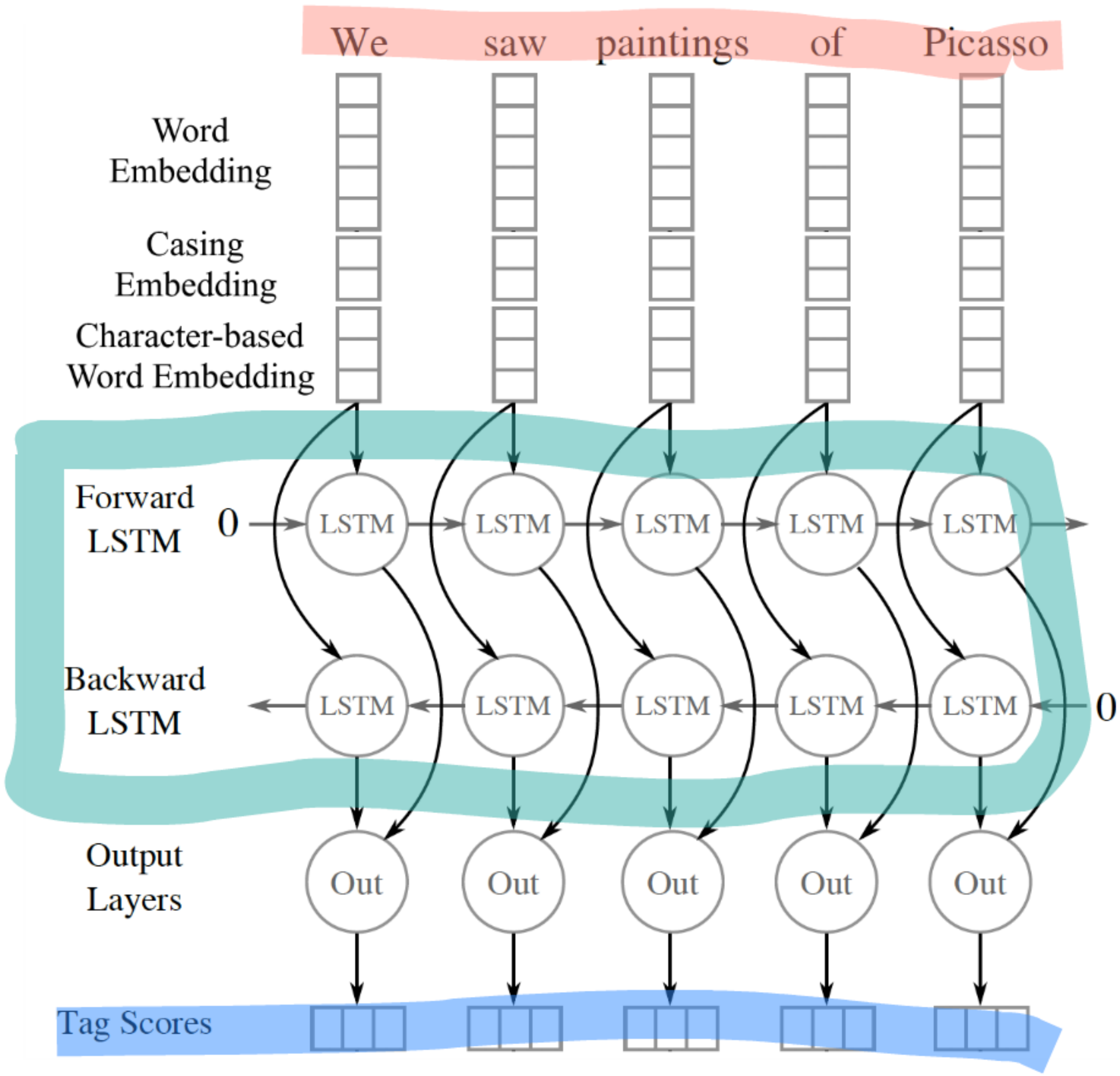

Bi-LSTM
Recurrent Neural Network
B-PER
I-PER
O
O
B-LOC
Angela
Merkel
wohnt
in
Berlin

Recurrent Neural Network

B-PER I-PER O O B-LOC



Angela Merkel wohnt in Berlin


We
saw
paintings
of
Picasso
Word
Embedding
Casing
Embedding
Character-based
Word Embedding
Forward
LSTM
0
LSTM
Backward
LSTM
Output
Layers
Out
Tag Scores

Label

Labeling Schemes

Entity Types

ORG - Organization
PER - Person
LOC - Location
OTH - Other

+

BIO

Begin - Inside - Outside

BILOU

Begin - Inside - Last - Outside - Unitary

BIO

	Angela	Merkel	wohnt	in	Berlin
B-LOC	0.08	0.08	0.03	0.08	0.78
I-LOC	0.02	0.07	0.02	0.05	0.03
B-PER	0.6	0.2	0.04	0.01	0.05
I-PER	0.2	0.55	0.01	0.01	0.05
O	0.1	0.1	0.9	0.85	0.09

BILOU

	Angela	Merkel	wohnt	in	Berlin
B-LOC	0.008	0.002	0.03	0.08	0.3
I-LOC	0.001	0.008	0.02	0.05	0.03
U-LOC	0.001	0.01	0.0	0.0	0.4
B-PER	0.4	0.25	0.04	0.01	0.05
I-PER	0.2	0.3	0.01	0.01	0.05
L-PER	0.3	0.4	0.0	0.0	0.08
O	0.09	0.03	0.9	0.85	0.09

Labeling Schemes

Entity Types

```
ORG - Organization
PER - Person
LOC - Location
OTH - Other
```

+

BIO

Begin - Inside - Outside

BILOU

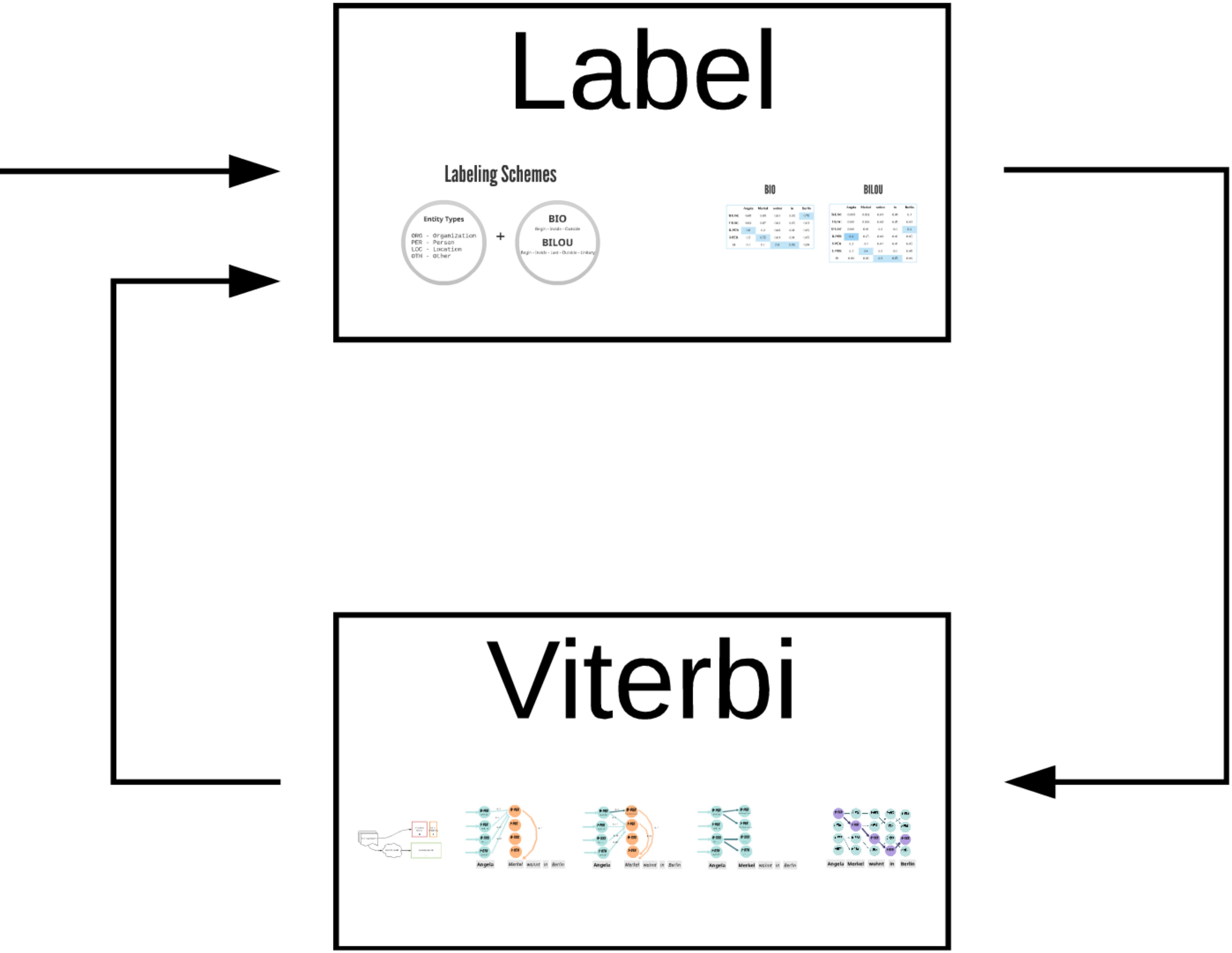
Begin - Inside - Last - Outside - Unitary

BIO

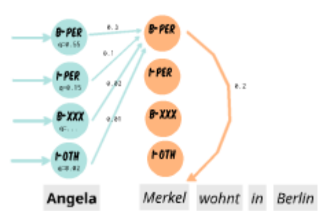
	Angela	Merkel	wohnt	in	Berlin
B-LOC	0.08	0.08	0.03	0.08	0.78
I-LOC	0.02	0.07	0.02	0.05	0.03
B-PER	0.6	0.2	0.04	0.01	0.05
I-PER	0.2	0.55	0.01	0.01	0.05
O	0.1	0.1	0.9	0.85	0.09

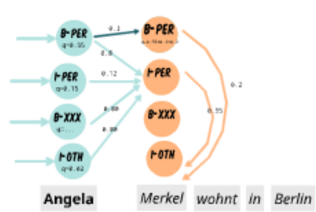
BILOU

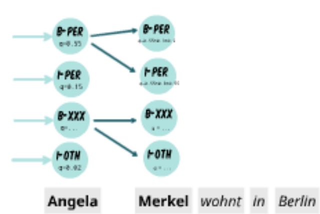
	Angela	Merkel	wohnt	in	Berlin
B-LOC	0.008	0.002	0.03	0.08	0.3
I-LOC	0.001	0.008	0.02	0.05	0.03
U-LOC	0.001	0.01	0.0	0.0	0.4
B-PER	0.4	0.25	0.04	0.01	0.05
I-PER	0.2	0.3	0.01	0.01	0.05
L-PER	0.3	0.4	0.0	0.0	0.08
O	0.09	0.03	0.9	0.85	0.09

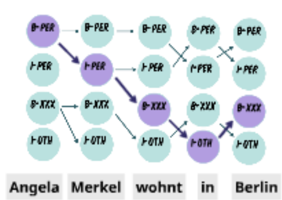

Label
Labeling Schemes
Entity Types
ORG - Organization
PER - Person
LOC - Location
OTH - Other
+
BIO
BILOU
BIO
BILOU
Viterbi

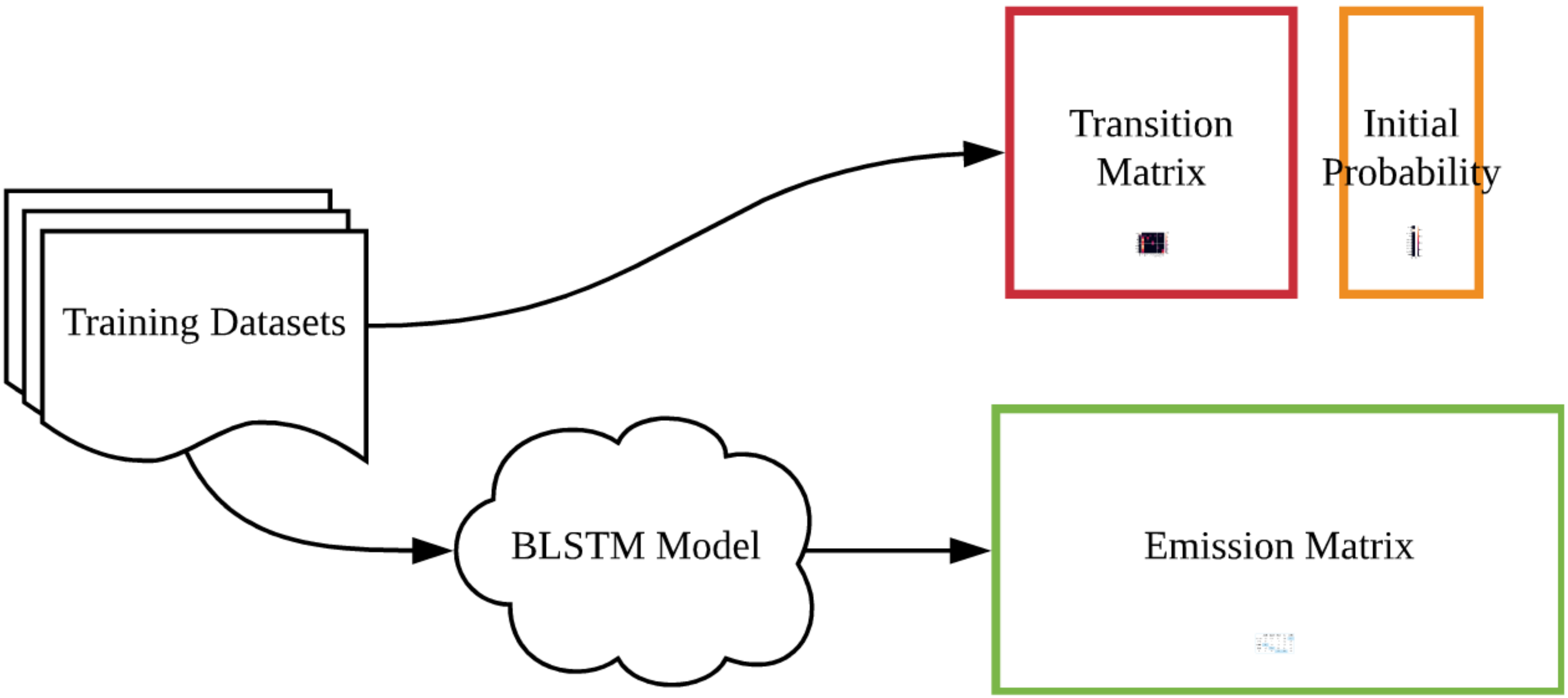
Viterbi

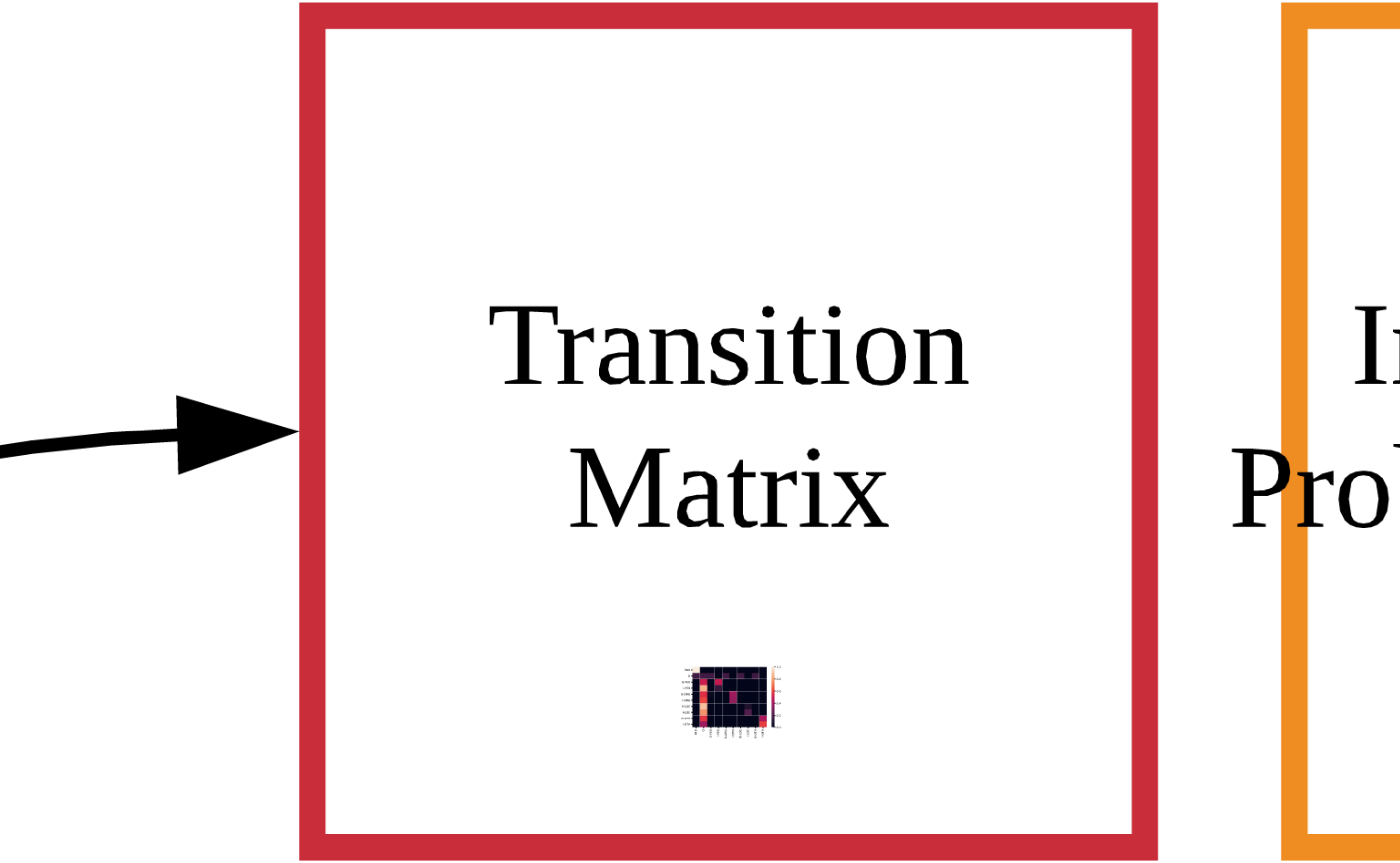

B-PER
I-PER
B-XXX
I-OTH
B-PER
I-PER
B-XXX
I-OTH
Angela
Merkel
wohnt
in
Berlin

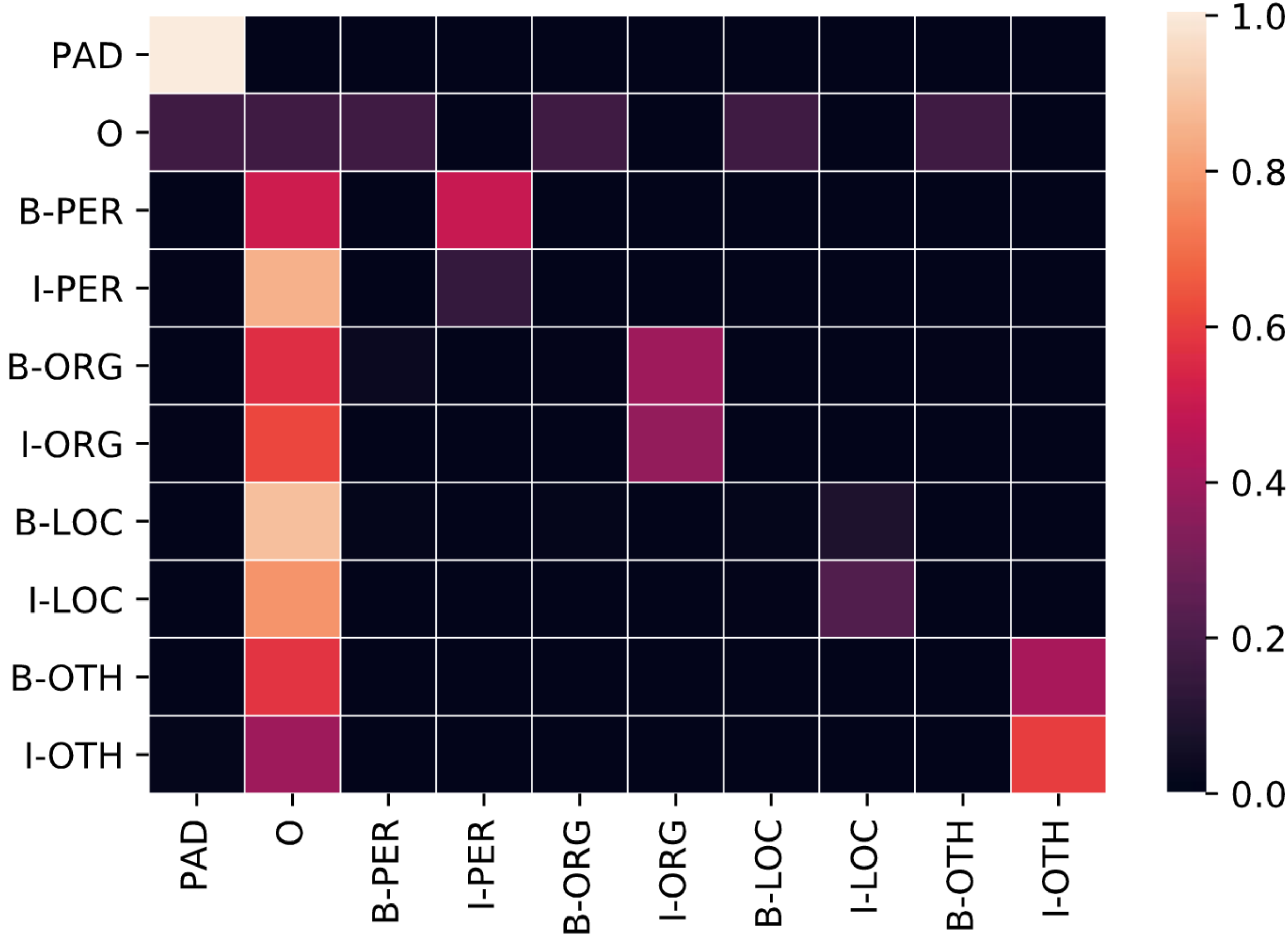

B-PER
I-PER
B-XXX
I-OTH
B-PER
I-PER
B-XXX
I-OTH
Angela
Merkel
wohnt
in
Berlin


B-PER
I-PER
B-XXX
I-OTH
B-PER
I-PER
B-XXX
I-OTH
Angela
Merkel
wohnt
in
Berlin


Angela
Merkel
wohnt
in
Berlin

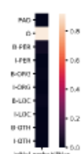

Training Datasets
BLSTM Model
Transition Matrix
Initial Probability
Emission Matrix

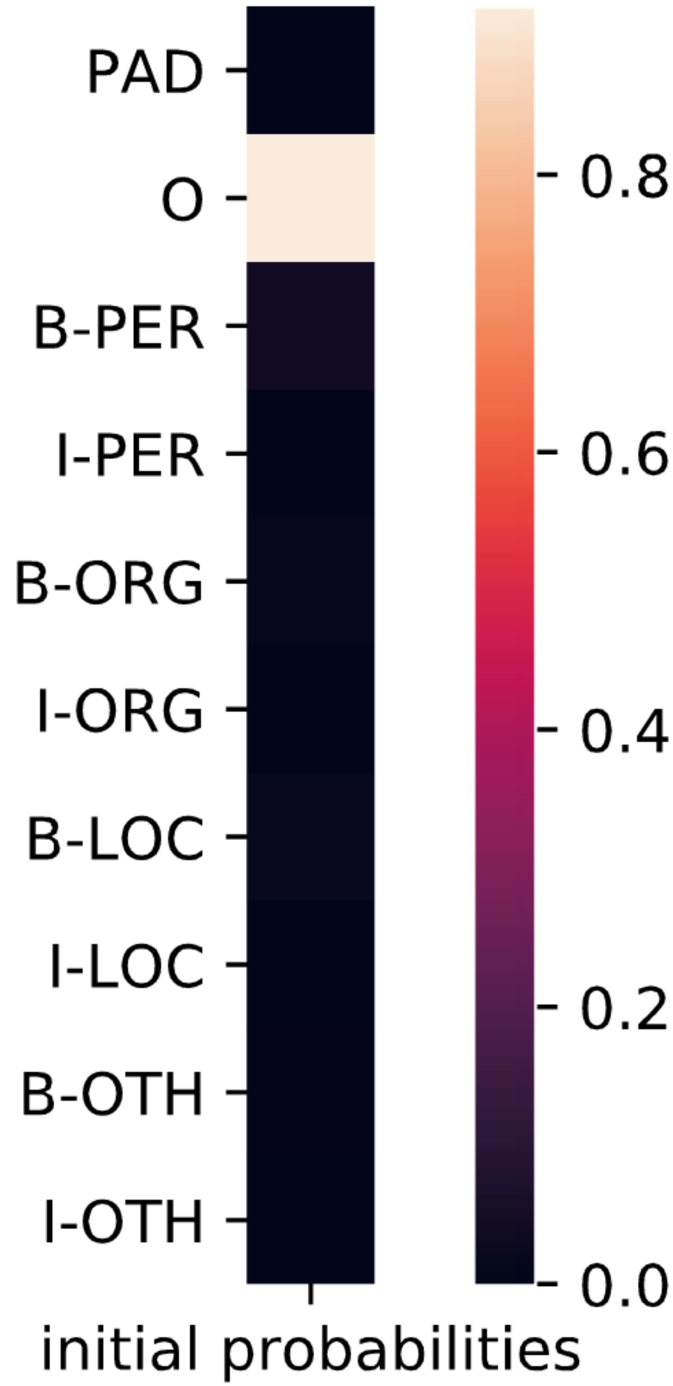

Transition
Matrix

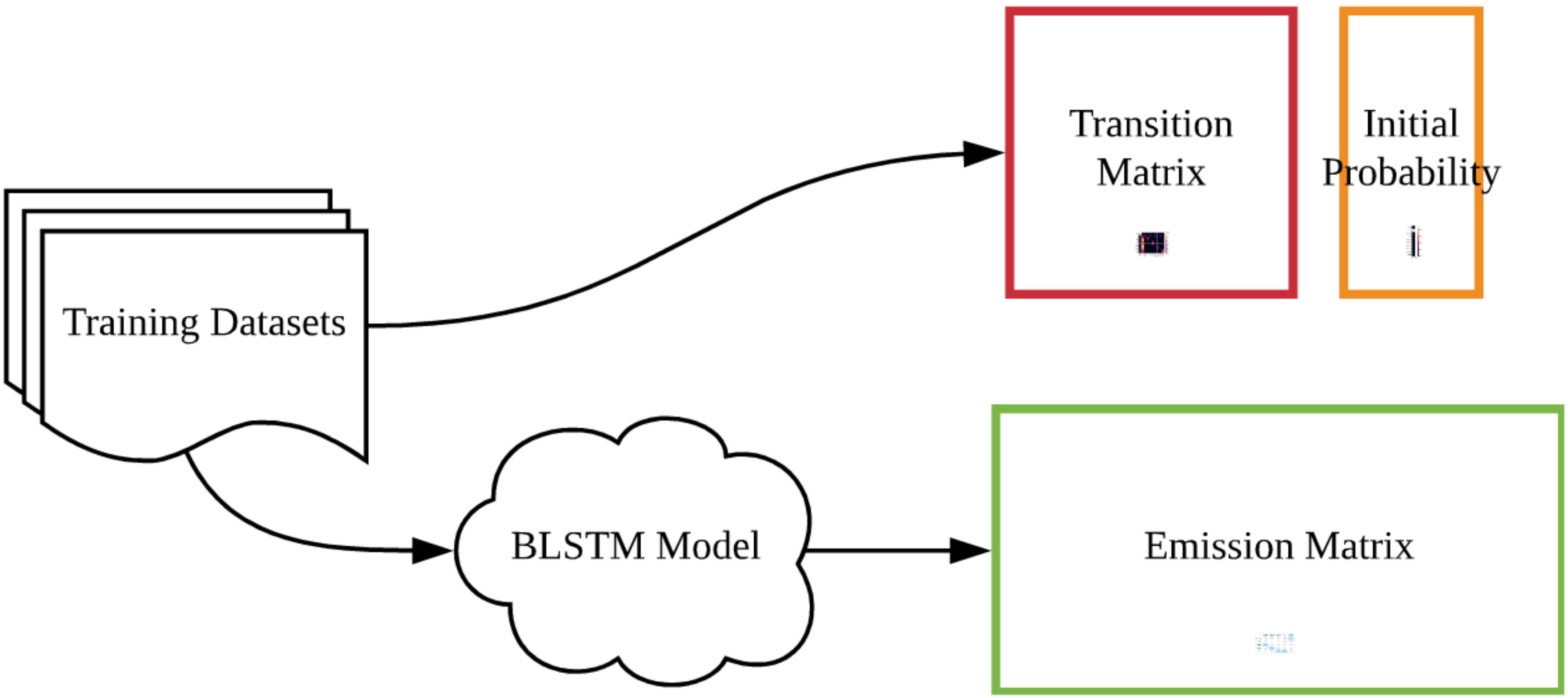

PAD
O
B-PER
I-PER
B-ORG
I-ORG
B-LOC
I-LOC
B-OTH
I-OTH
PAD
O
B-PER
I-PER
B-ORG
I-ORG
B-LOC
I-LOC
B-OTH
I-OTH
1.0
0.8
0.6
0.4
0.2
0.0

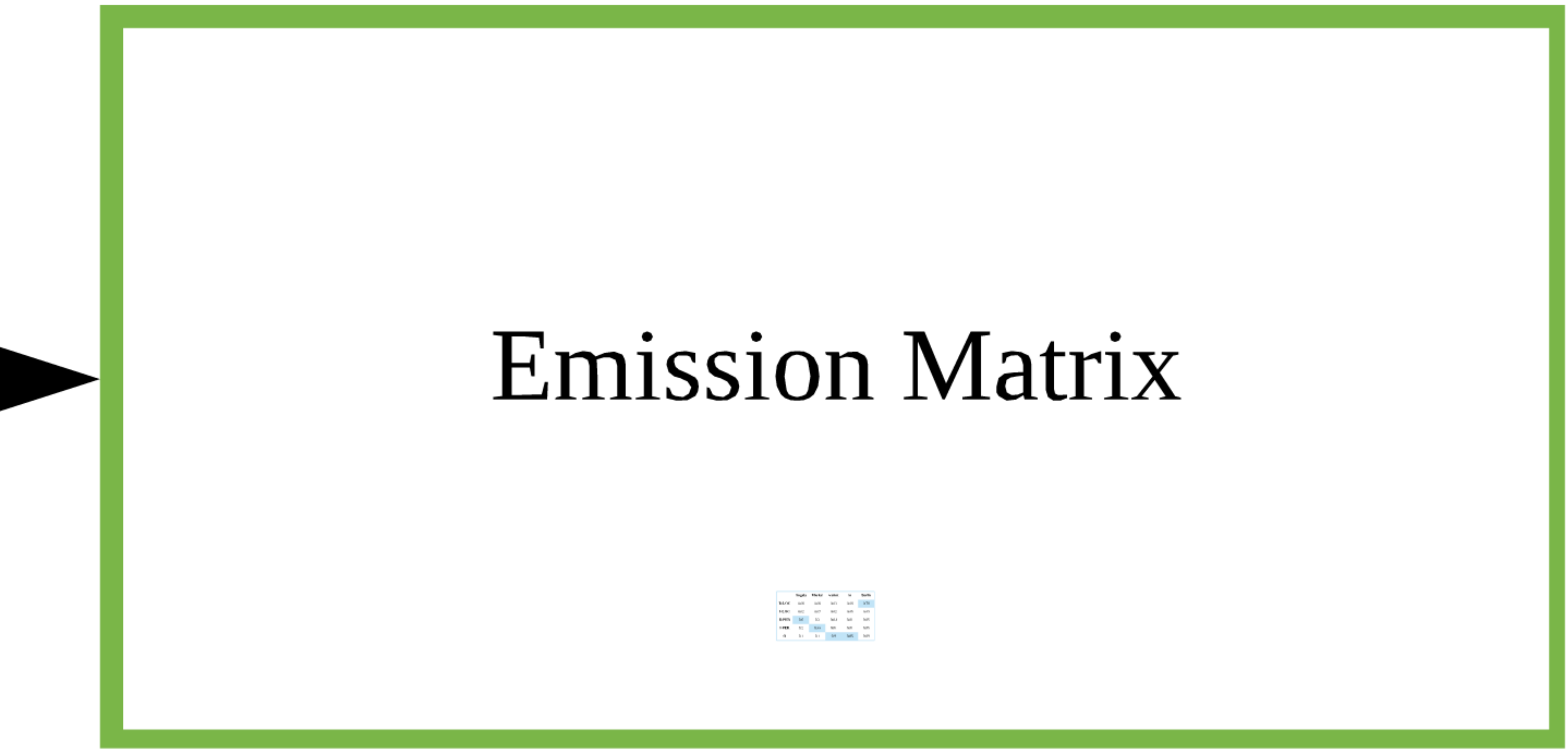
tion
ix

Initial
Probability

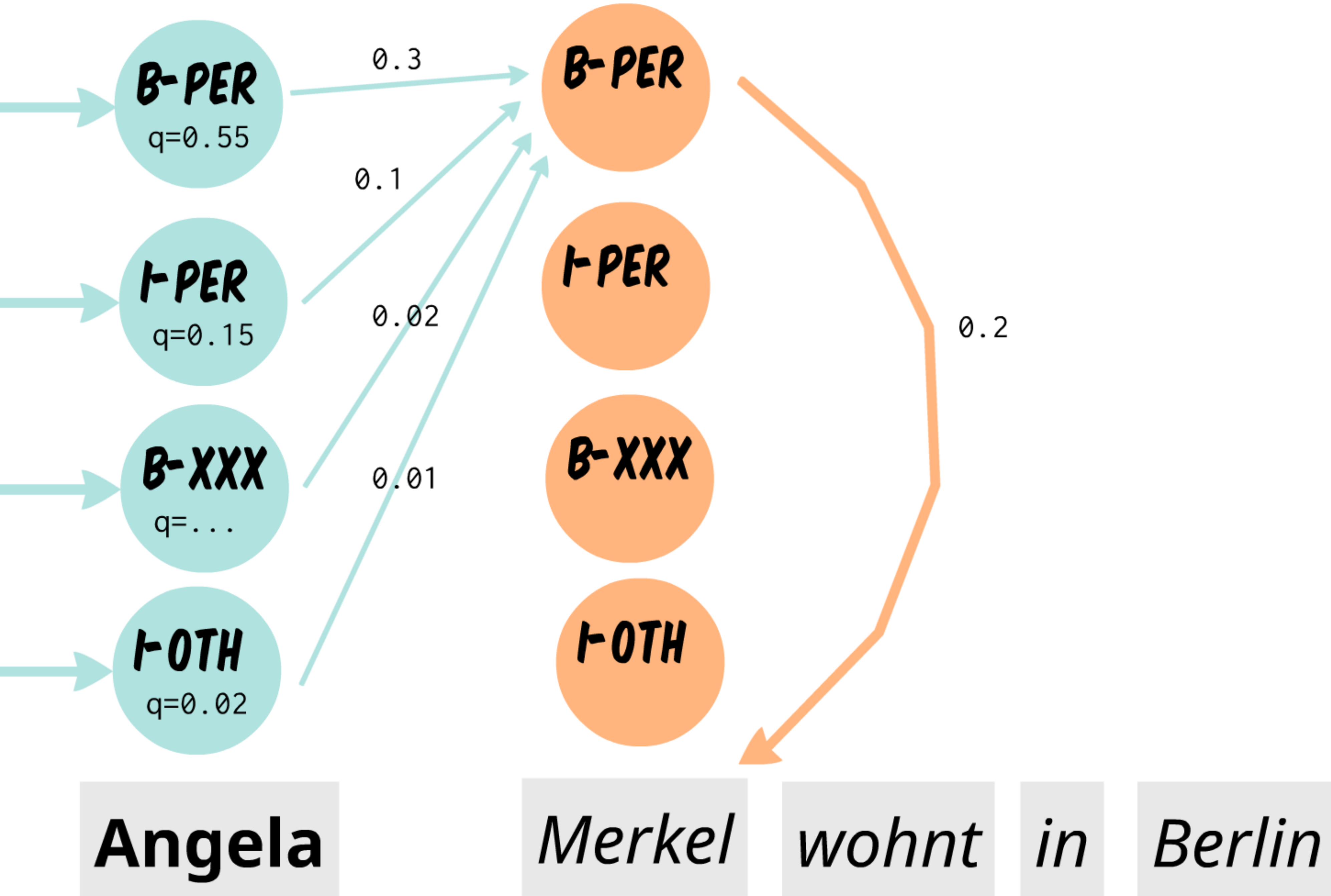


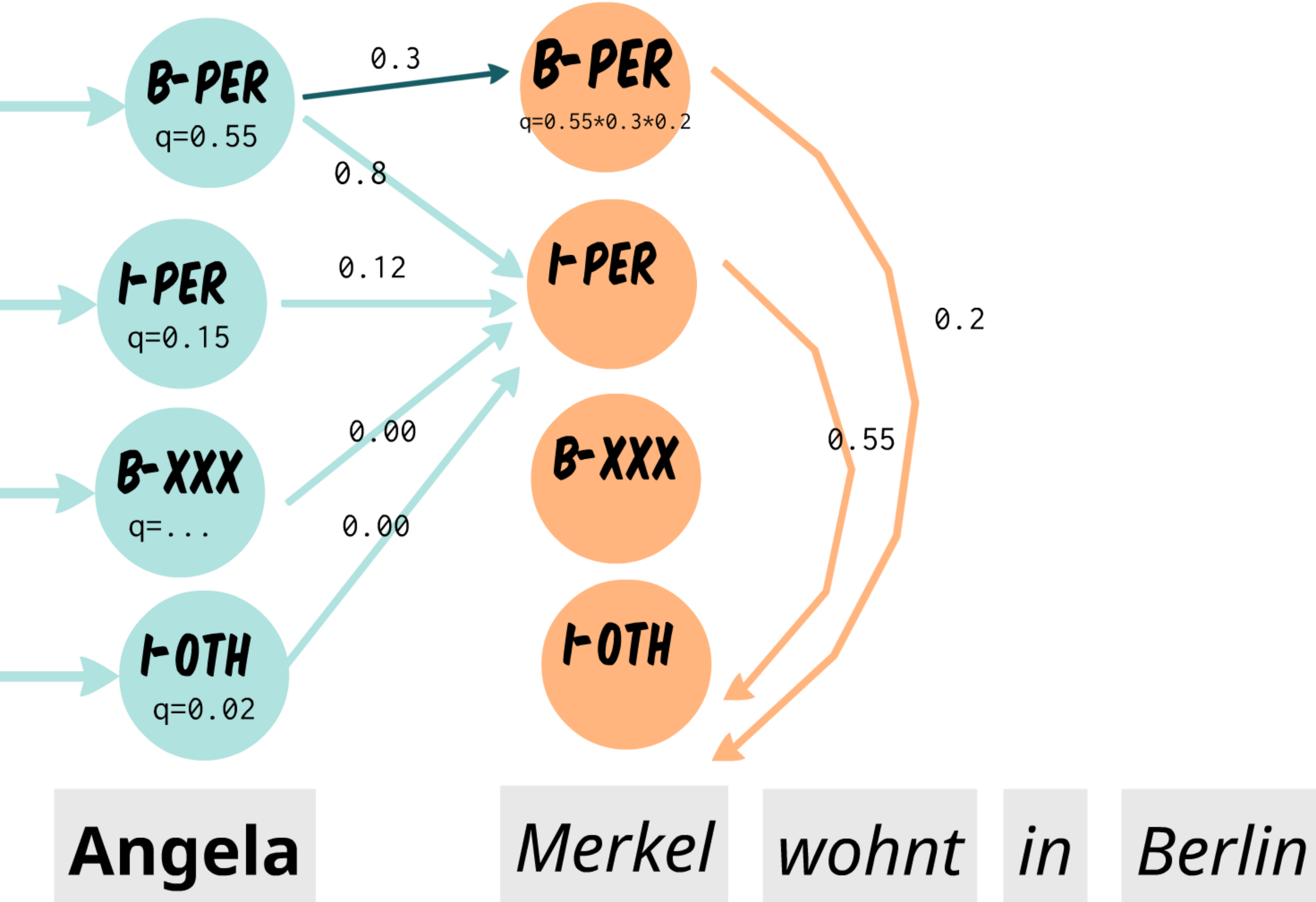

PAD
O
B-PER
I-PER
B-ORG
I-ORG
B-LOC
I-LOC
B-OTH
I-OTH
initial probabilities
0.8
0.6
0.4
0.2
0.0

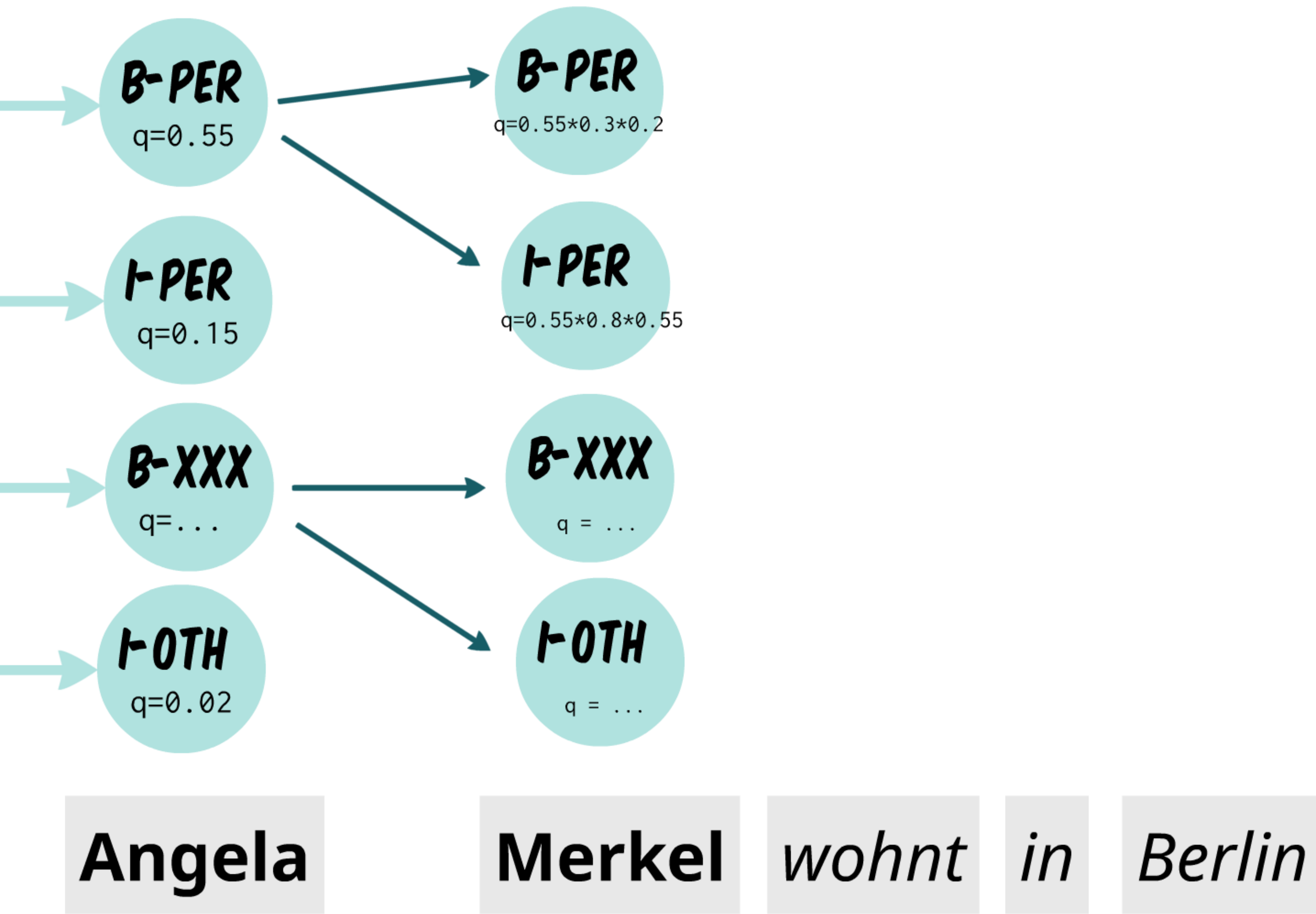

Training Datasets
Transition
Matrix
Initial
Probability
BLSTM Model
Emission Matrix

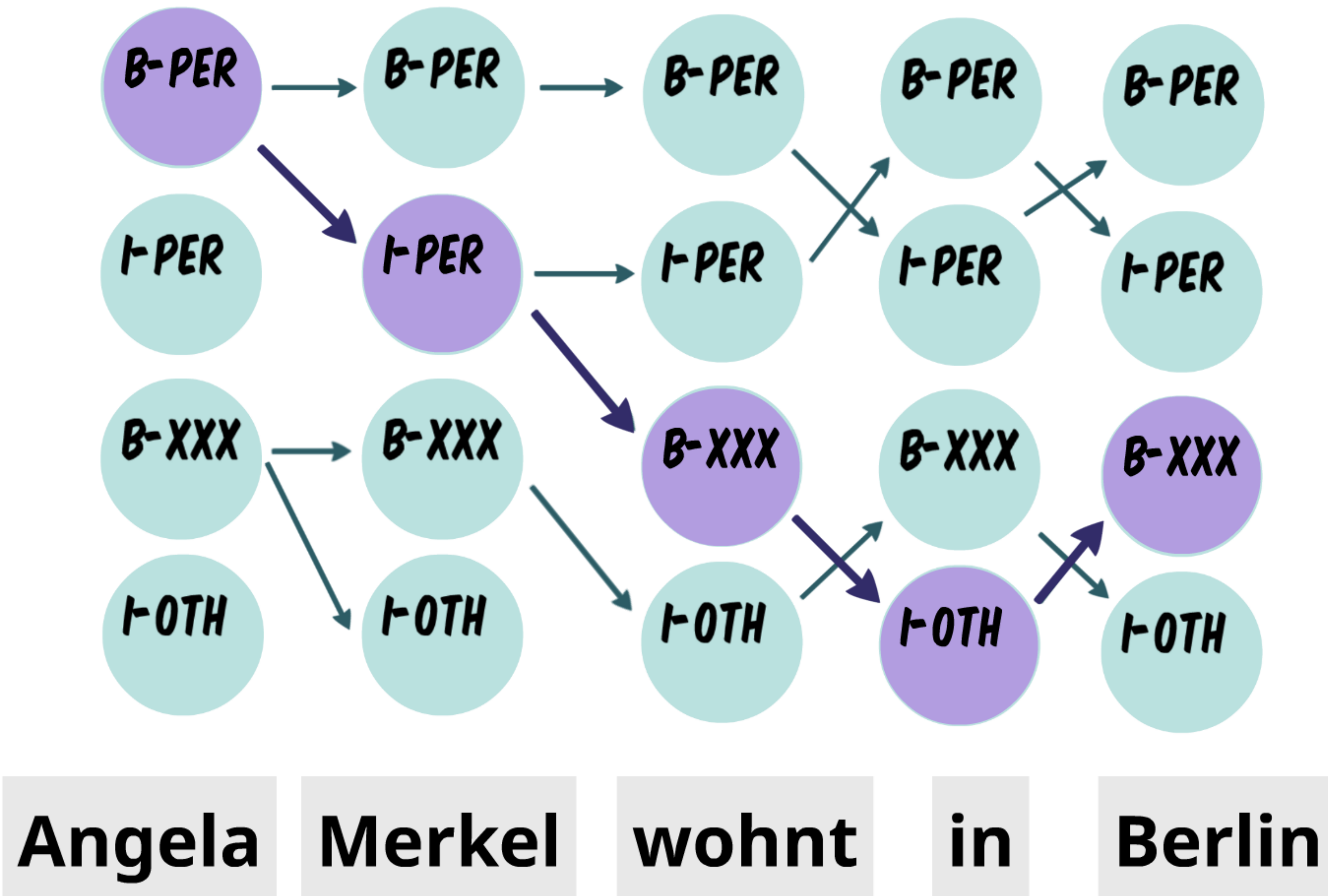

Emission Matrix

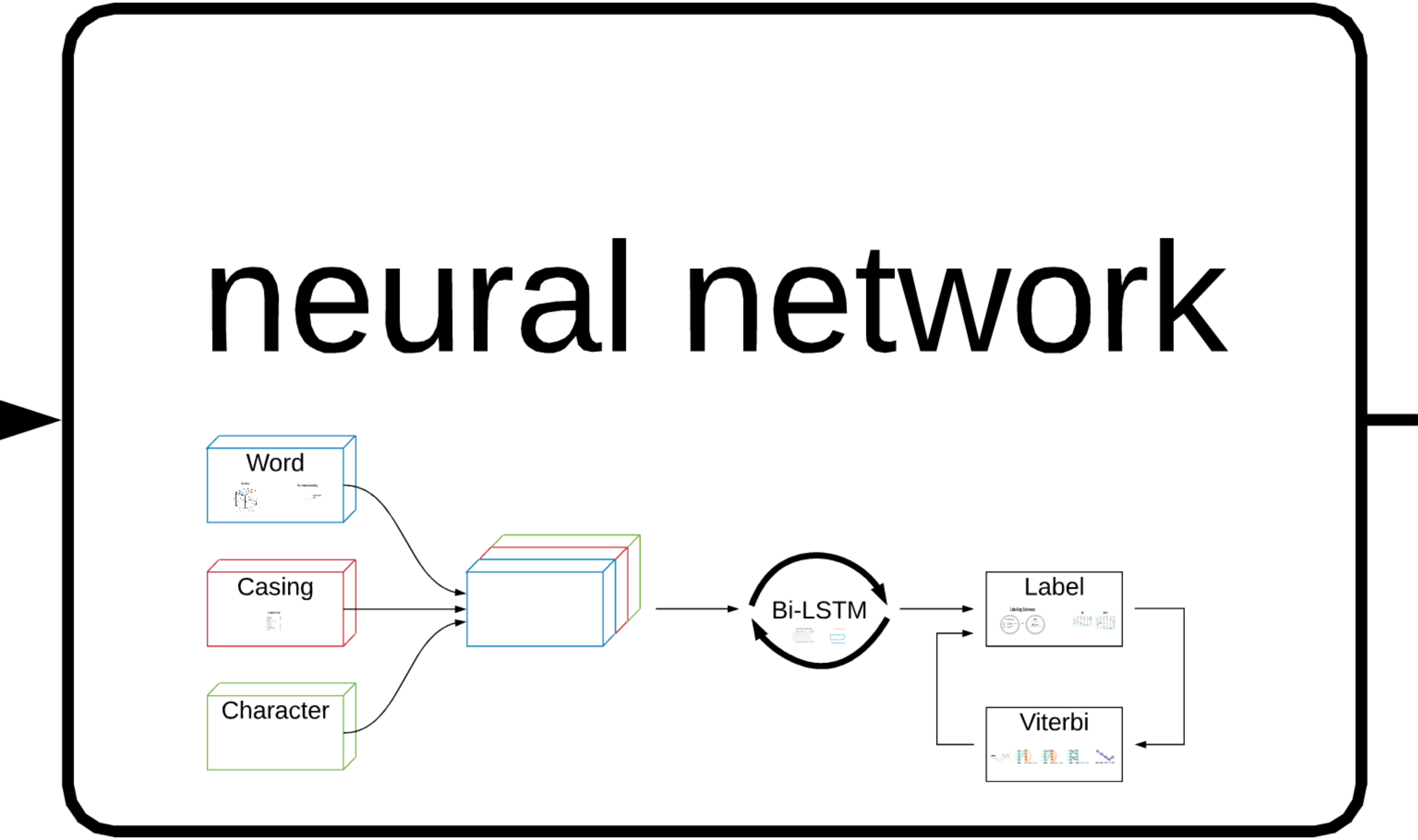
	Angela	Merkel	wohnt	in	Berlin
B-LOC	0.08	0.08	0.03	0.08	0.78
I-LOC	0.02	0.07	0.02	0.05	0.03
B-PER	0.6	0.2	0.04	0.01	0.05
I-PER	0.2	0.55	0.01	0.01	0.05
O	0.1	0.1	0.9	0.85	0.09


B-PER
q=0.55
I-PER
q=0.15
B-XXX
q=...
I-OTH
q=0.02
0.3
0.1
0.02
0.01
B-PER
I-PER
B-XXX
I-OTH
0.2
Angela
Merkel
wohnt
in
Berlin


B-PER
q=0.55
I-PER
q=0.15
B-XXX
q=...
I-OTH
q=0.02
0.3
0.8
0.12
0.00
0.00
B-PER
q=0.55*0.3*0.2
I-PER
B-XXX
I-OTH
0.2
0.55
Angela
Merkel
wohnt
in
Berlin

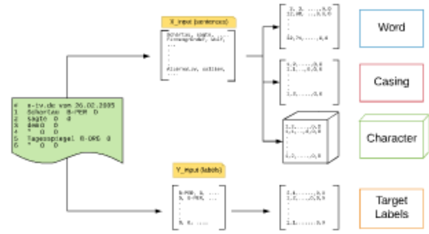

B-PER
q=0.55
I-PER
q=0.15
B-XXX
q=...
I-OTH
q=0.02
B-PER
q=0.55*0.3*0.2
I-PER
q=0.55*0.8*0.55
B-XXX
q = ...
I-OTH
q = ...
Angela
Merkel
wohnt
in
Berlin

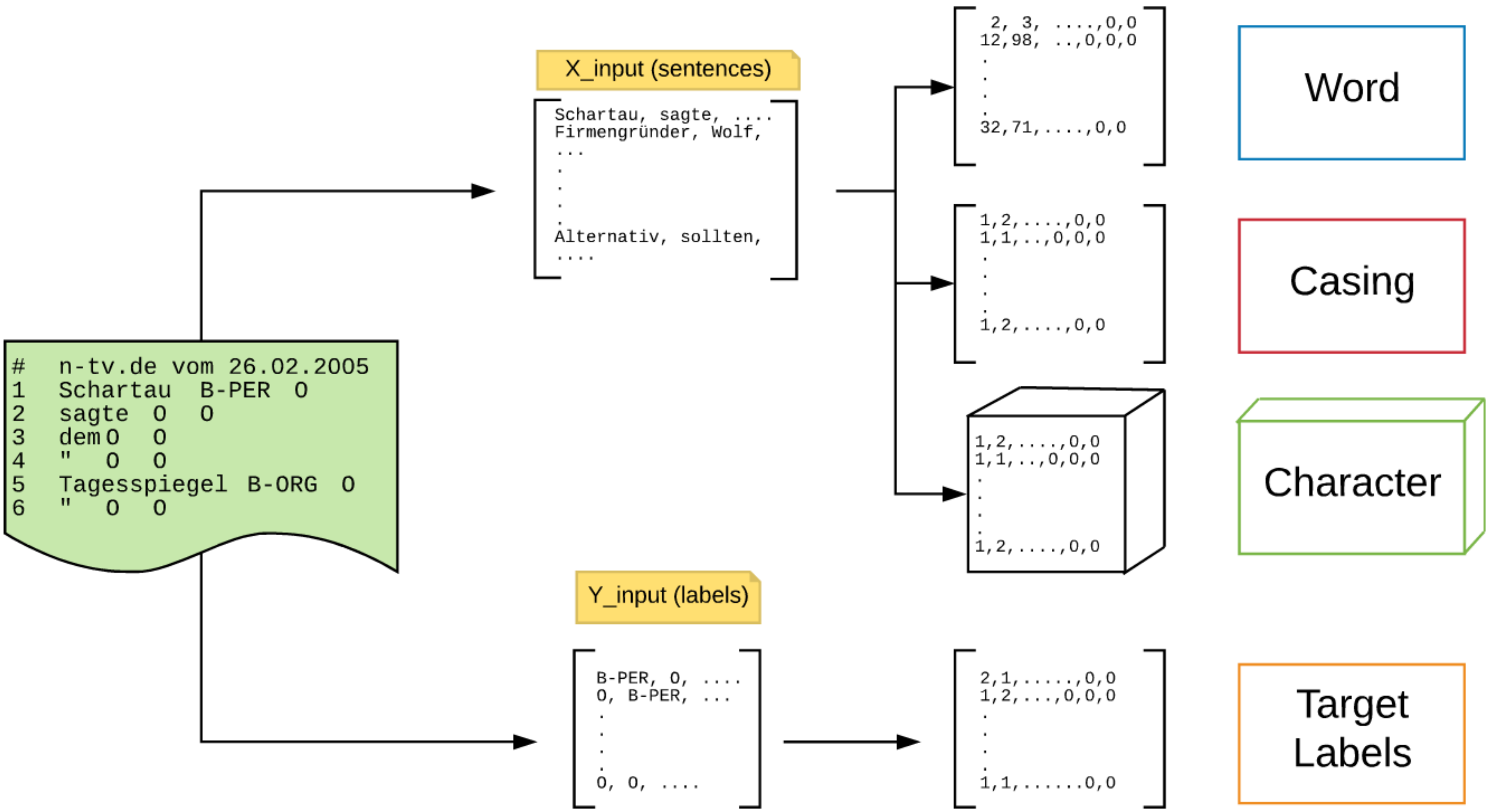

B-PER
I-PER
B-XXX
I-OTH
B-PER
I-PER
B-XXX
I-OTH
B-PER
I-PER
B-XXX
I-OTH
B-PER
I-PER
B-XXX
I-OTH
B-PER
I-PER
B-XXX
I-OTH
Angela
Merkel
wohnt
in
Berlin



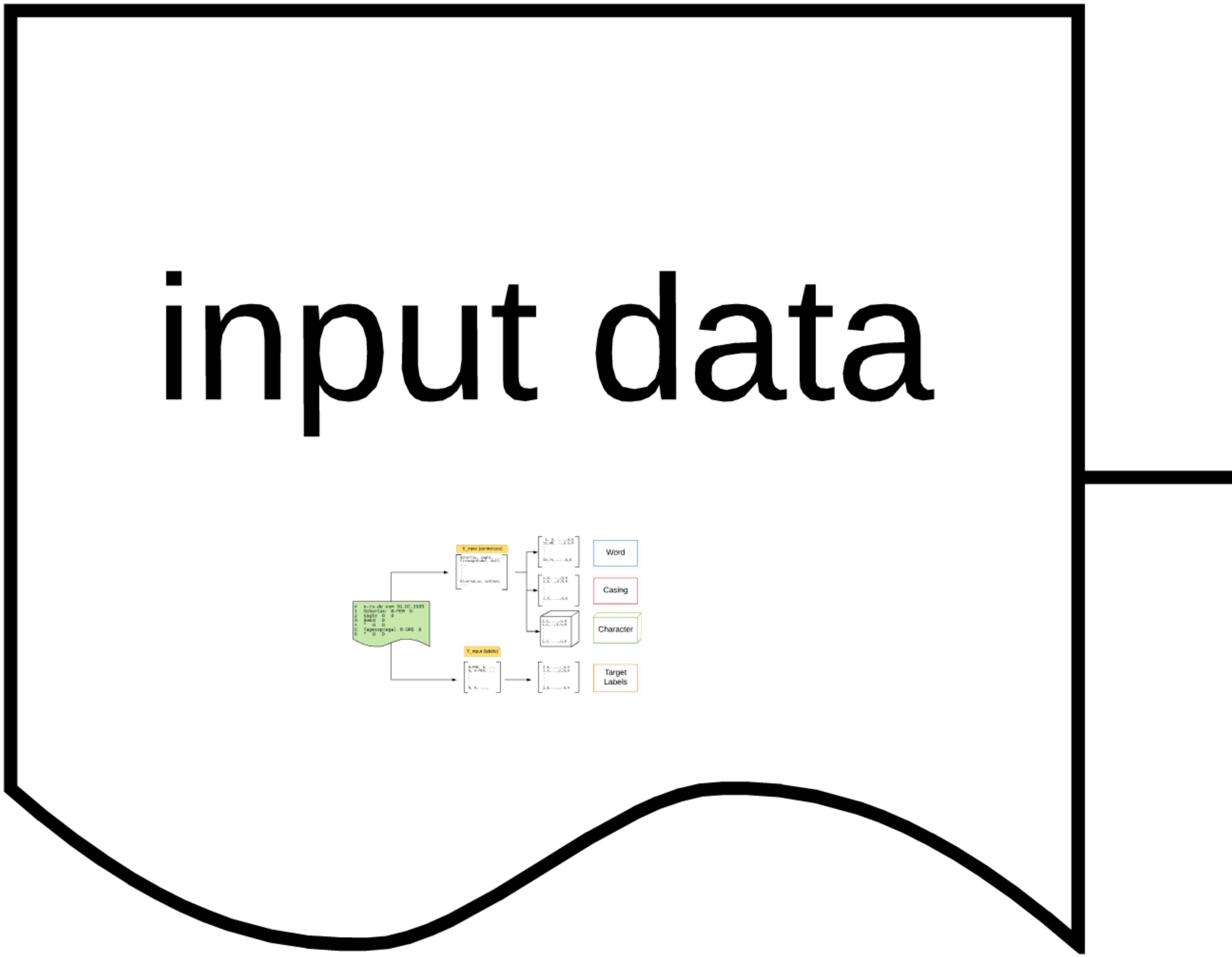
neural network
Word
Casing
Character
Bi-LSTM
Label
Viterbi

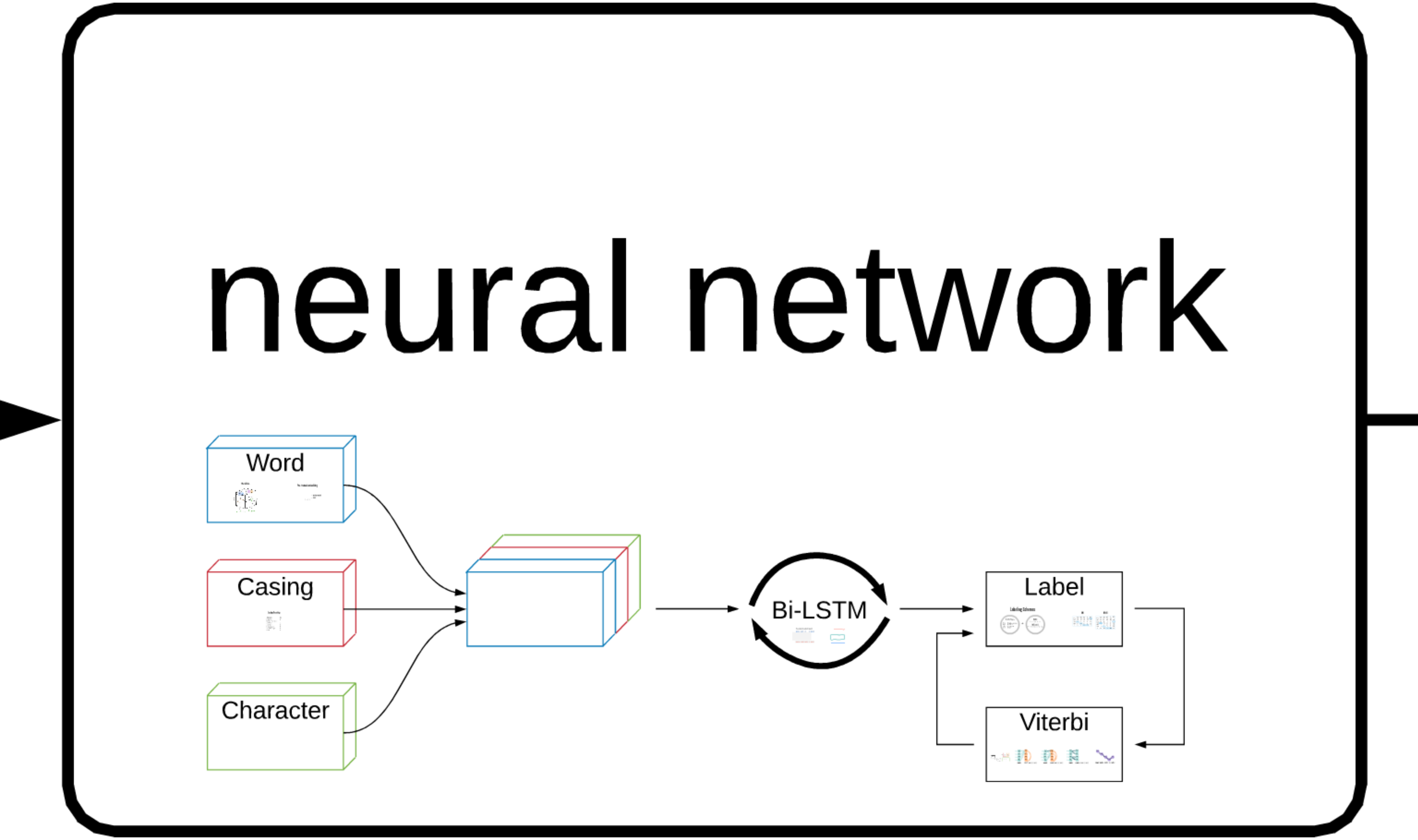
input data



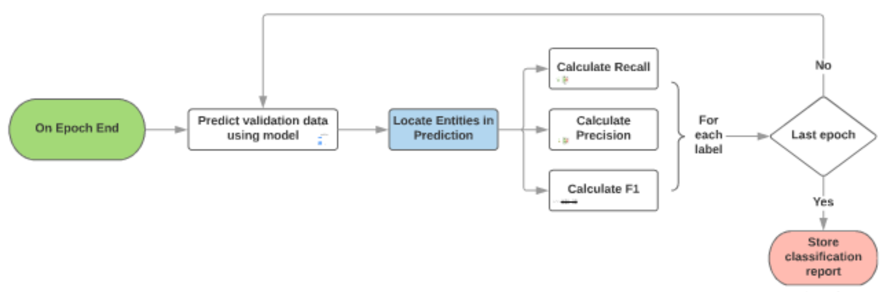


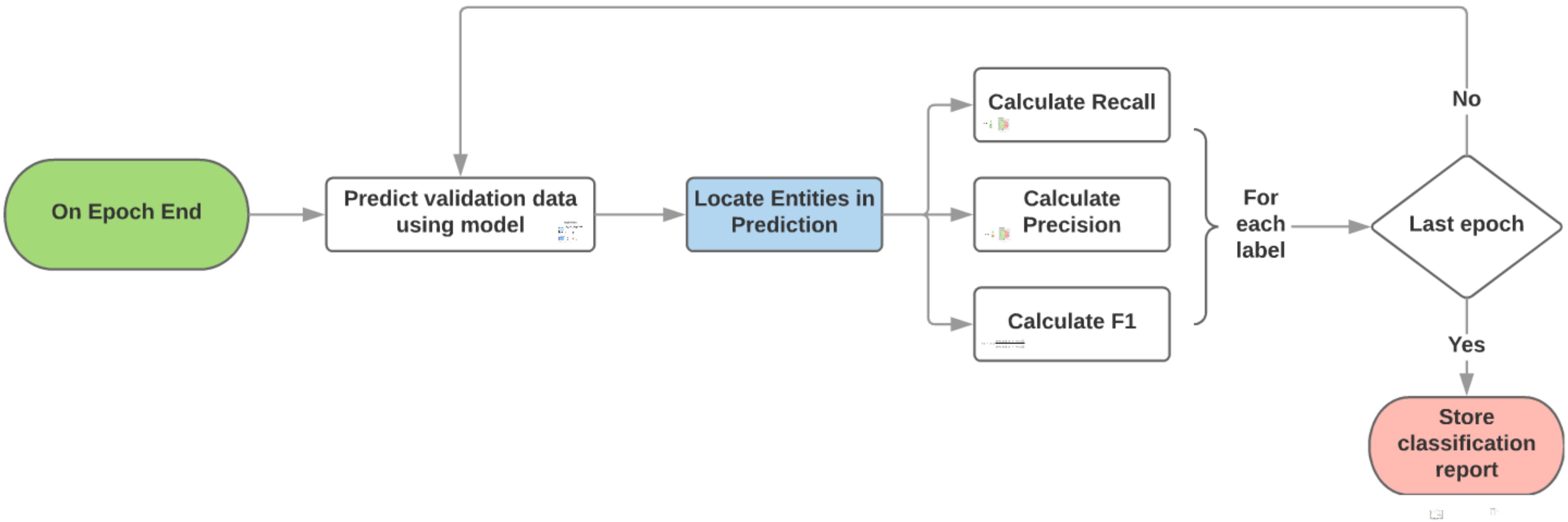
n-tv.de vom 26.02.2005
1 Schartau B-PER O
2 sagte O O
3 dem O O
4 " O O
5 Tagesspiegel B-ORG O
6 " O O
X_input (sentences)
Schartau, sagte, ...
Firmengründer, Wolf,
...
Alternativ, sollten,
....
2, 3, ...,0,0
12,98, ..,0,0,0
32,71,....,0,0
1,2,....,0,0
1,1,..,0,0,0
1,2,....,0,0
1,2,....,0,0
1,1,..,0,0,0
1,2,....,0,0
Word
Casing
Character
Y_input (labels)
B-PER, O, ...
O, B-PER, ...
O, O, ...
2,1,.....,0,0
1,2,...,0,0,0
1,1,......0,0
Target Labels

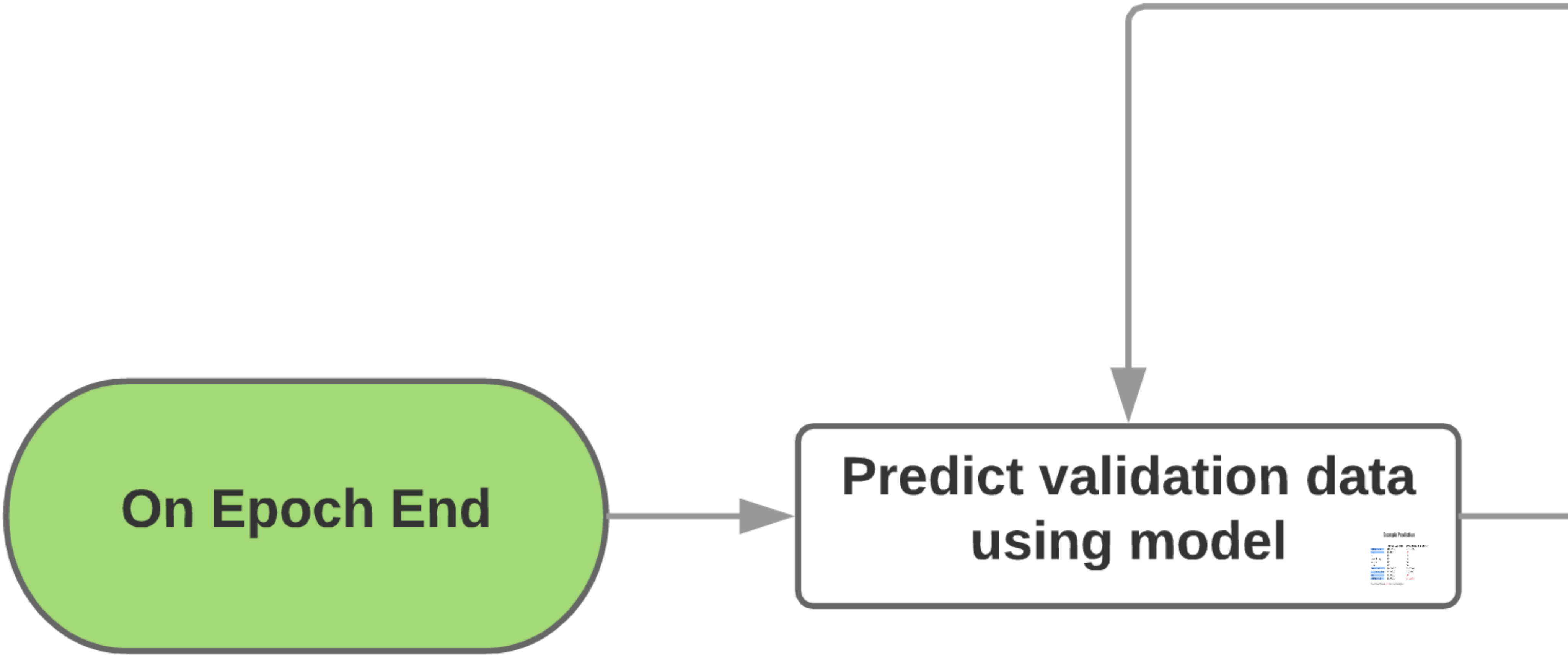

input data
X_input (sentences)
Y_input (labels)
Word
Casing
Character
Target
Labels


neural network
Word
Casing
Character
Bi-LSTM
Label
Viterbi

evaluation


On Epoch End
Predict validation data using model
Locate Entities in Prediction
Calculate Recall
Calculate Precision
Calculate F1
For each label
Last epoch
No
Yes
Store classification report

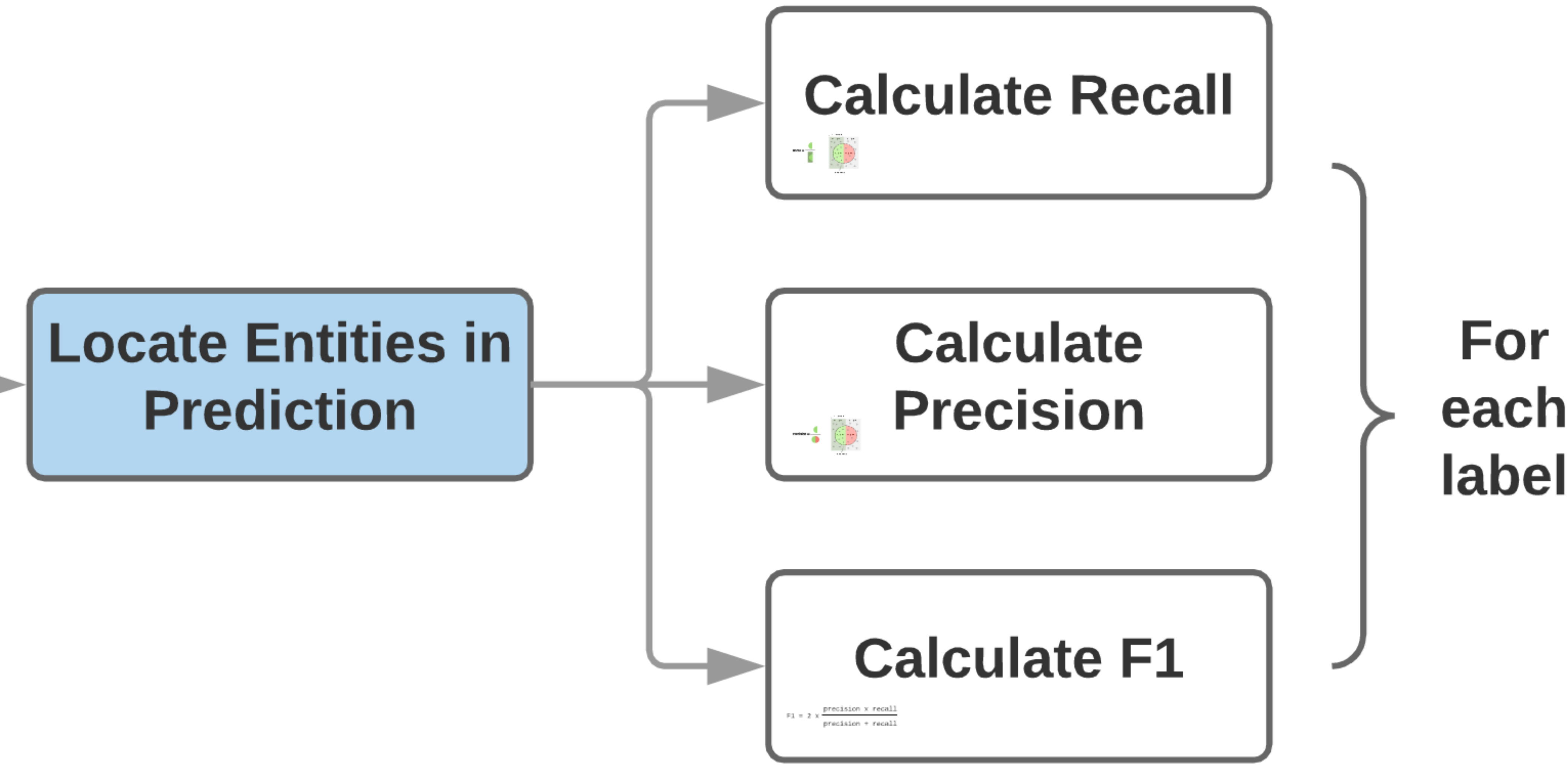

On Epoch End
Predict validation data using model
Locate Entities in Prediction
Calculate Recall
Calculate Precision
Calculate F1
For each label
Last epoch
No
Yes
Store classification report

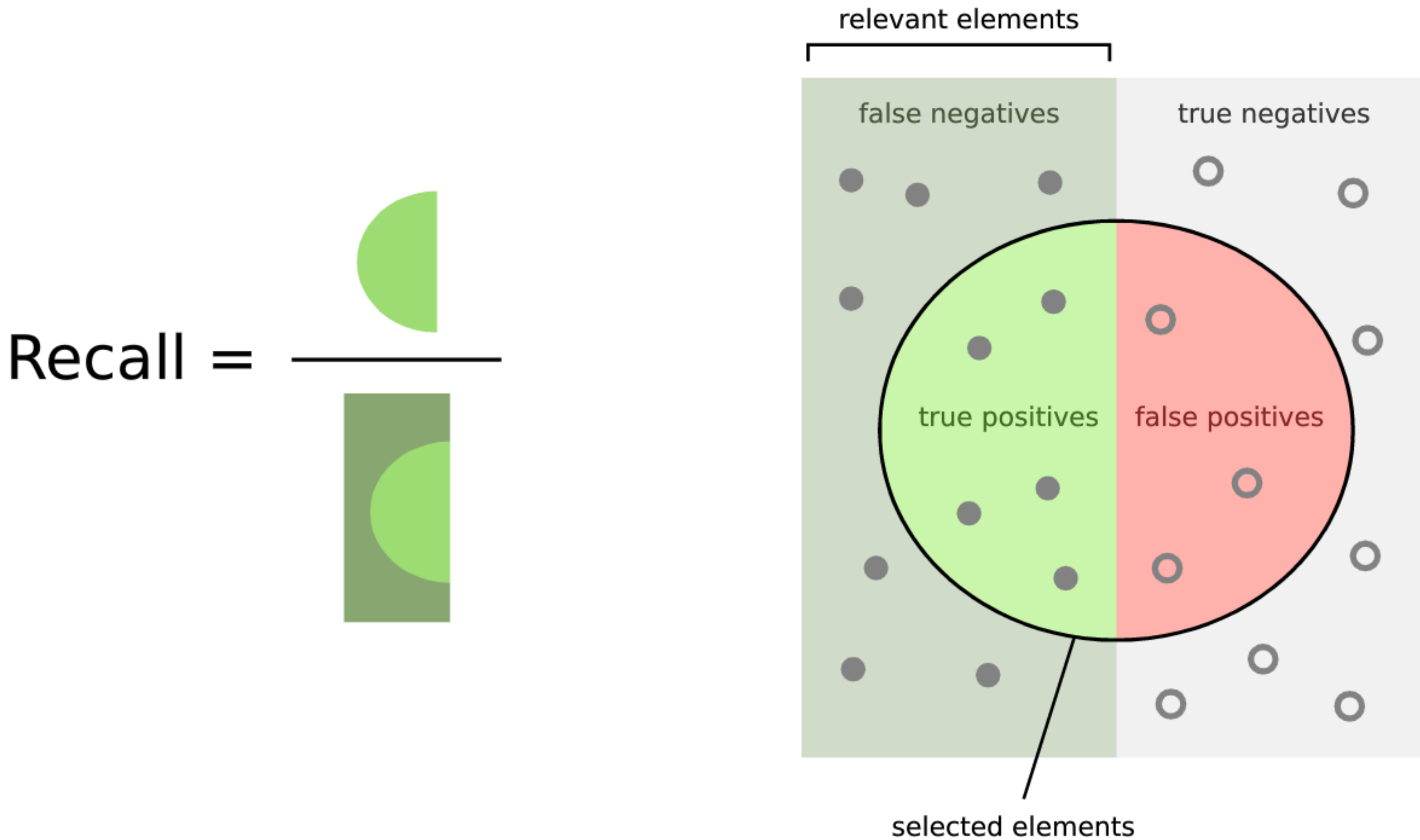

On Epoch End
Predict validation data using model

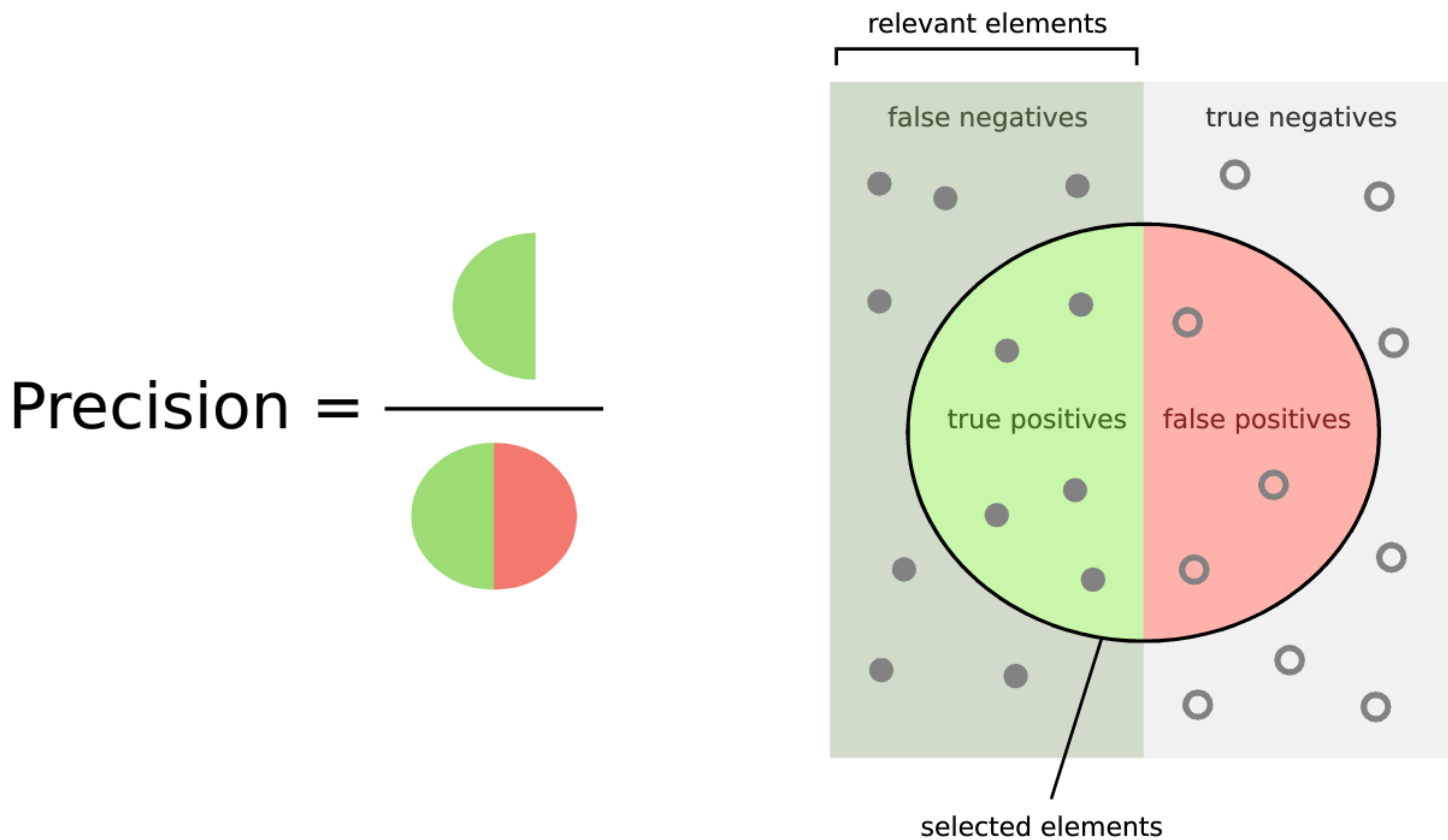
Example Prediction

	True Labels	*Predicted Labels*
Munich	B-ORG	B-ORG
Re	I-ORG	O
is	O	O
working	O	O
with	O	O
the	O	O
Technical	B-ORG	B-ORG
University	I-ORG	I-ORG
of	I-ORG	O
Munich	I-ORG	B-LOC

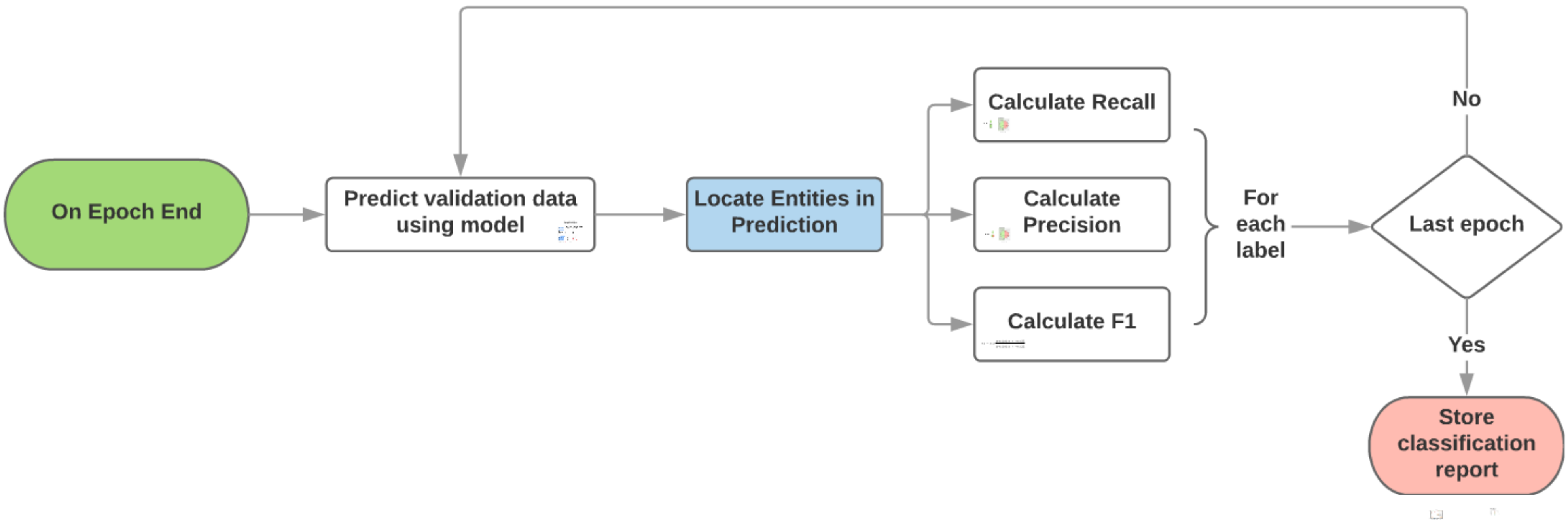
**labels written in red are misclassified*


Locate Entities in Prediction
Calculate Recall
Calculate Precision
Calculate F1
For each label

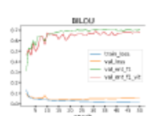

Recall =
relevant elements
false negatives
true negatives
true positives
false positives
selected elements

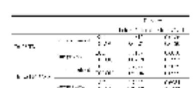

Precision =
relevant elements
false negatives
true negatives
true positives
false positives
selected elements

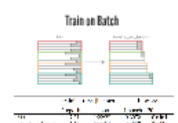
$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

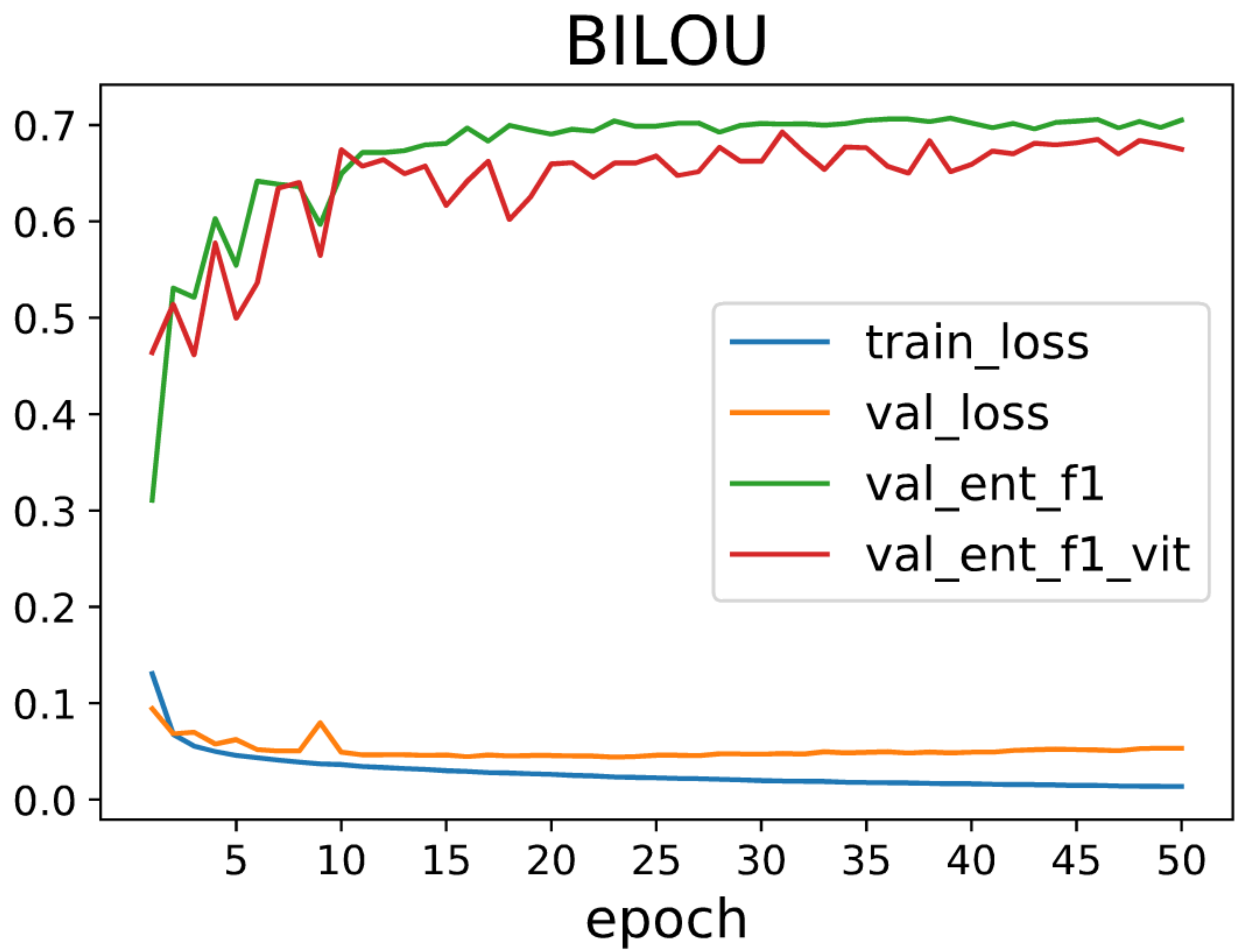

On Epoch End
Predict validation data using model
Locate Entities in Prediction
Calculate Recall
Calculate Precision
Calculate F1
For each label
Last epoch
No
Yes
Store classification report

Store classification report



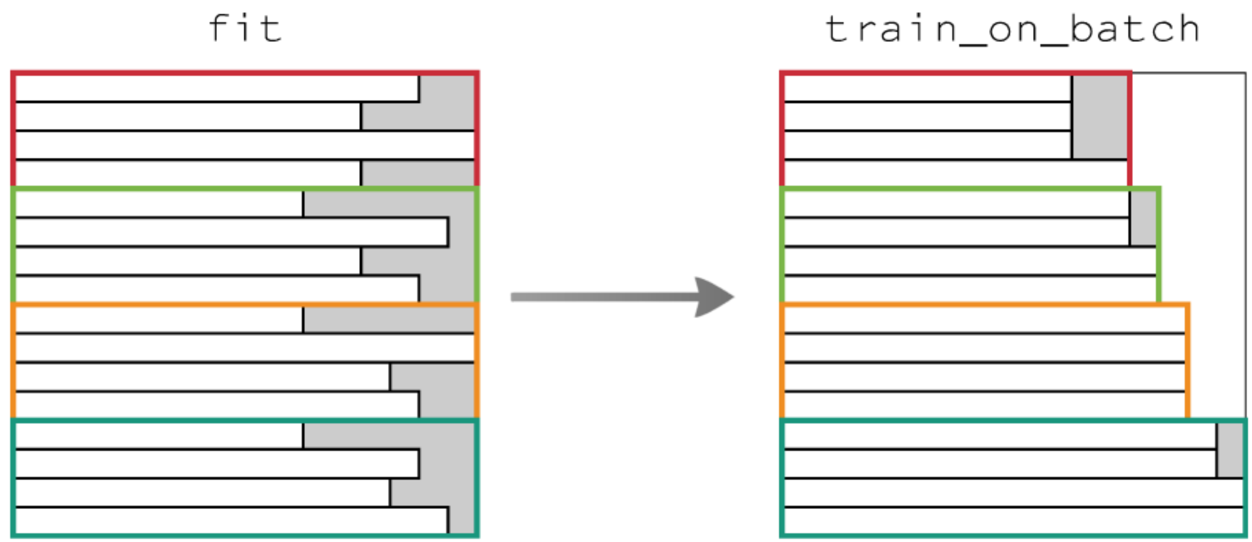




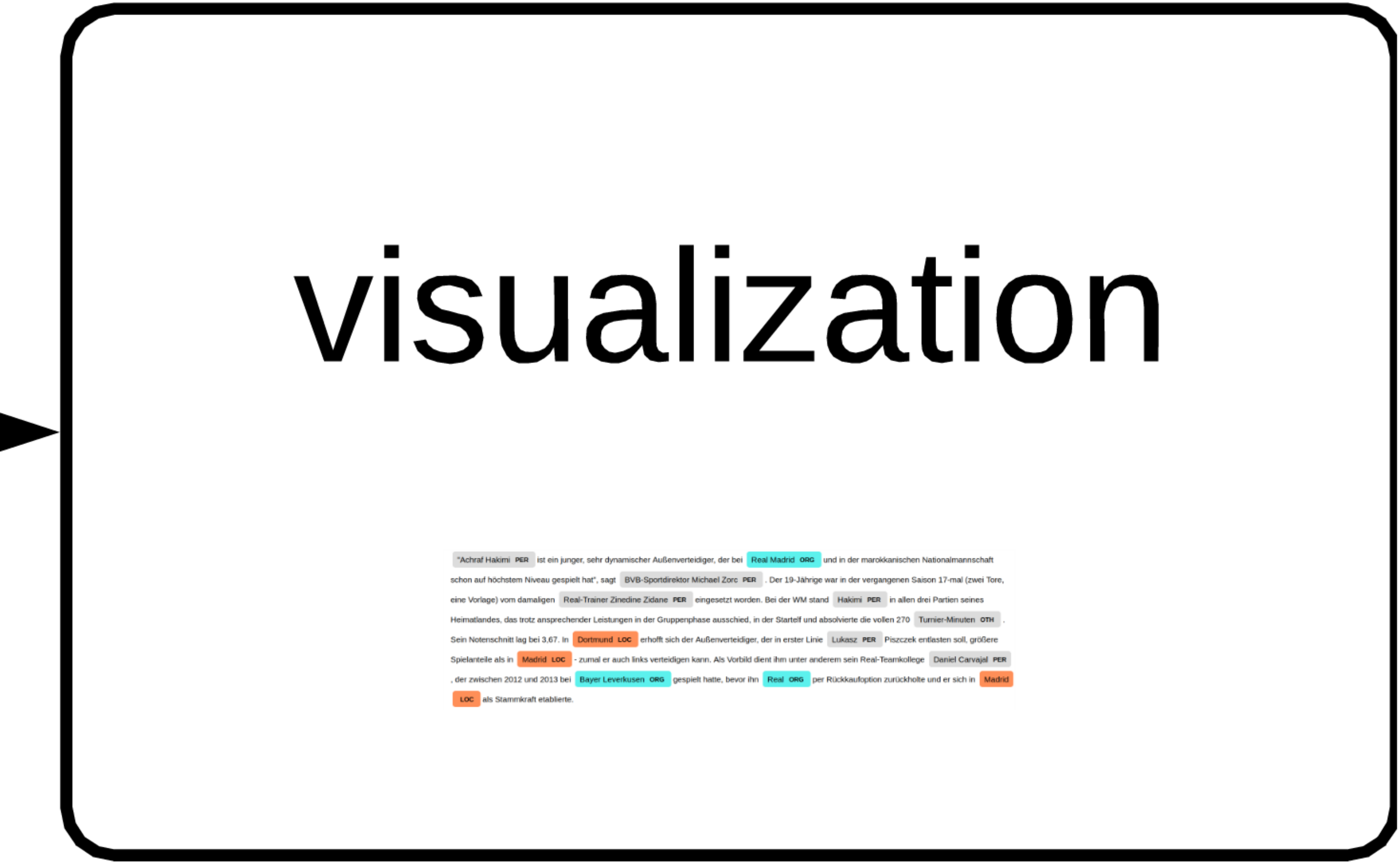

BILOU
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
5
10
15
20
25
30
35
40
45
50
epoch
train_loss
val_loss
val_ent_f1
val_ent_f1_vit

			F_1 score	
			before Viterbi	after Viterbi
Bi-LSTM	coarse-grained	BIO	0.6315	0.6186
		BILOU	**0.6421**	0.6120
	fine-grained	BIO	0.6167	0.6050
		BILOU	**0.6243**	0.5997
Bi-LSTM CNN	coarse-grained	BIO	0.7077	0.7079
		BILOU	**0.7095**	0.6929
	fine-grained	BIO	0.7002	0.6934
		BILOU	**0.7008**	0.6552

Train on Batch



	training time [h:mm:ss]		F_1 score	
	1 epoch	20 epochs	1 epoch	20 epochs
`fit`	0:07:51	1:22:28	0.3972	0.6464
`train_on_batch`	0:05:16	0:44:14	0.4087	0.6328


visualization
"Achraf Hakimi PER ist ein junger, sehr dynamischer Außenverteidiger, der bei Real Madrid ORG und in der marokkanischen Nationalmannschaft
schon auf höchstem Niveau gespielt hat", sagt BVB-Sportdirektor Michael Zorc PER . Der 19-Jährige war in der vergangenen Saison 17-mal (zwei Tore,
eine Vorlage) vom damaligen Real-Trainer Zinedine Zidane PER eingesetzt worden. Bei der WM stand Hakimi PER in allen drei Partien seines
Heimatlandes, das trotz ansprechender Leistungen in der Gruppenphase ausschied, in der Startelf und absolvierte die vollen 270 Turnier-Minuten OTH .
Sein Notenschnitt lag bei 3,67. In Dortmund LOC erhofft sich der Außenverteidiger, der in erster Linie Lukasz PER Piszczek entlasten soll, größere
Spielanteile als in Madrid LOC - zumal er auch links verteidigen kann. Als Vorbild dient ihm unter anderem sein Real-Teamkollege Daniel Carvajal PER
, der zwischen 2012 und 2013 bei Bayer Leverkusen ORG gespielt hatte, bevor ihn Real ORG per Rückkaufoption zurückholte und er sich in Madrid
LOC als Stammkraft etablierte.

"Achraf Hakimi **PER** ist ein junger, sehr dynamischer Außenverteidiger, der bei Real Madrid **ORG** und in der marokkanischen Nationalmannschaft schon auf höchstem Niveau gespielt hat", sagt BVB-Sportdirektor Michael Zorc **PER** . Der 19-Jährige war in der vergangenen Saison 17-mal (zwei Tore, eine Vorlage) vom damaligen Real-Trainer Zinedine Zidane **PER** eingesetzt worden. Bei der WM stand Hakimi **PER** in allen drei Partien seines Heimatlandes, das trotz ansprechender Leistungen in der Gruppenphase ausschied, in der Startelf und absolvierte die vollen 270 Turnier-Minuten **OTH** . Sein Notenschnitt lag bei 3,67. In Dortmund **LOC** erhofft sich der Außenverteidiger, der in erster Linie Lukasz **PER** Piszczek entlasten soll, größere Spielanteile als in Madrid **LOC** - zumal er auch links verteidigen kann. Als Vorbild dient ihm unter anderem sein Real-Teamkollege Daniel Carvajal **PER** , der zwischen 2012 und 2013 bei Bayer Leverkusen **ORG** gespielt hatte, bevor ihn Real **ORG** per Rückkaufoption zurückholte und er sich in Madrid **LOC** als Stammkraft etablierte.

Demo Time

Summary

- Implemented NER with recurrent neural networks
- Using character and casing information improves results
- Applied Viterbi post-processing
- Got competitive results on GermEval data
- BILOU labeling scheme works best
- Sorting sentences by length yields a significant speedup

Future Work

- Use trainable word embeddings
- Fine-tune Viterbi algorithm
- Optimize neural network parameters

Thank you!