

OpenIE6: Iterative Grid Labeling and Coordination Analysis for Open Information Extraction

**Keshav Kolluru^{1*}, Vaibhav Adlakha^{1*}, Samarth Aggarwal¹,
Mausam¹, and Soumen Chakrabarti²**

¹ Indian Institute of Technology Delhi

keshav.kolluru@gmail.com, vaibhavadlakha95@gmail.com
samarth.aggarwal.2510@gmail.com, mausam@cse.iitd.ac.in

² Indian Institute of Technology Bombay

soumen@cse.iitb.ac.in



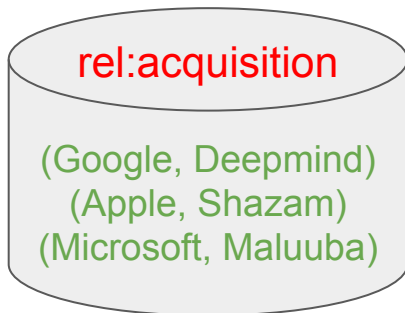
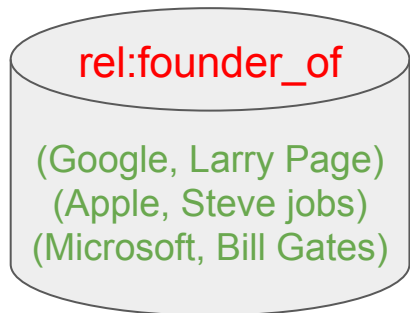
[github:dair-iitd/openie6](https://github.com/dair-iitd/openie6)

Ontological/Closed Information Extraction

“Apple’s founder Steve Jobs died of cancer following a...”



rel:founder_of(Apple, Steve Jobs)

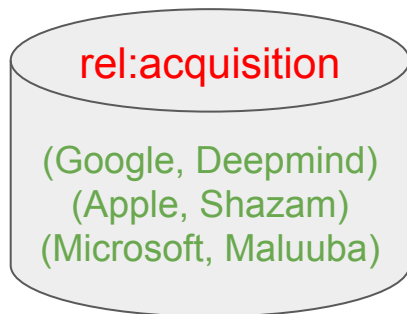
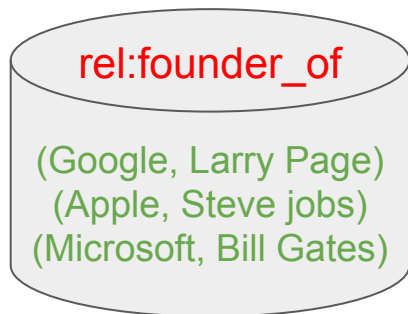


Ontological/Closed Information Extraction

“Apple’s founder Steve Jobs died of cancer following a...”



rel:founder_of(Apple, Steve Jobs)



.....

Only for tasks whose **knowledge needs** have been anticipated in advance

Open Information Extraction

“Apple’s founder Steve Jobs died of cancer following a...”

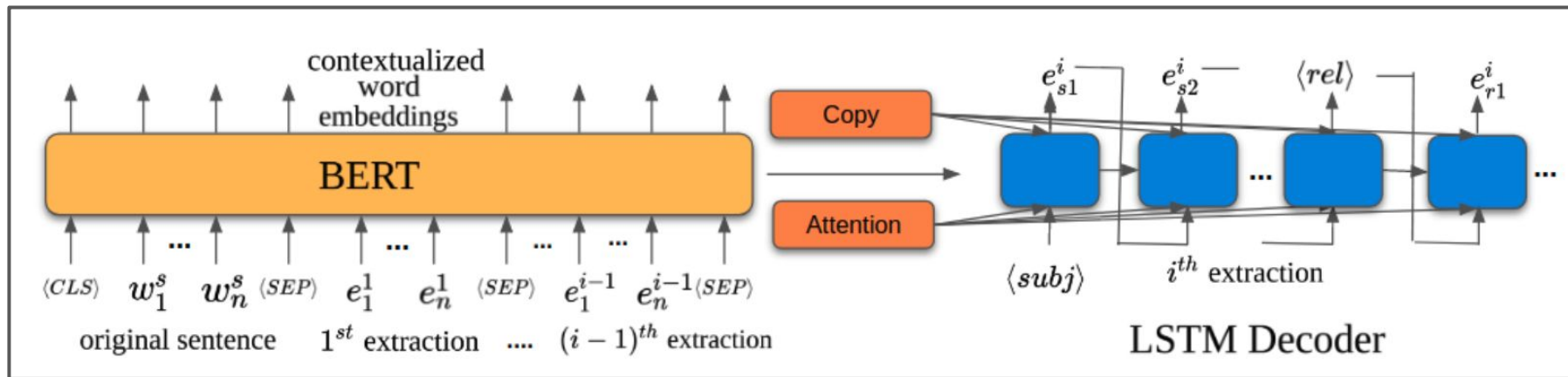
↓
Steve Jobs, [is] the founder [of], Apple
Steve Jobs, died of, cancer

↓
(Google, acquired, Deepmind)
(Oranges, contain, Vitamin C)
(Coronavirus, causes, loss of smell)

Approaches for OpenIE

- ***Generation***
 - Generate each OpenIE extraction one token after another
- ***Labeling***
 - Label words in the sentence with OpenIE tags

OpenIE Generation

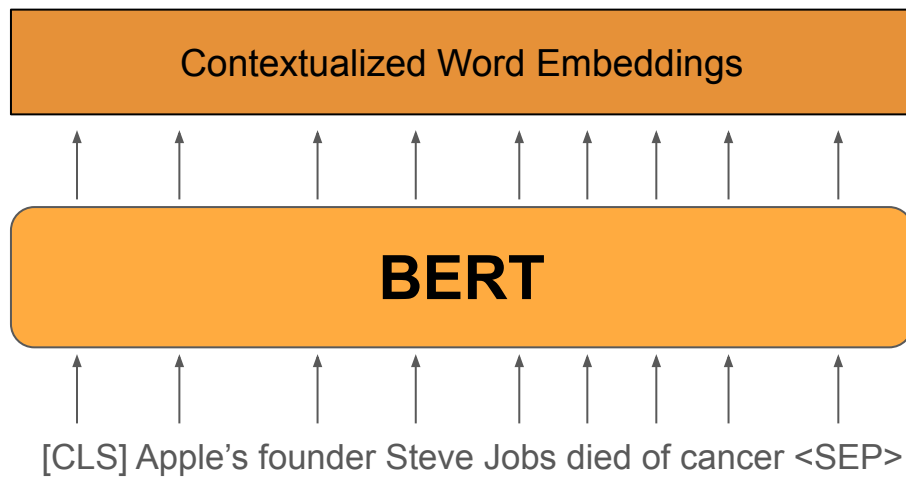


Terminology

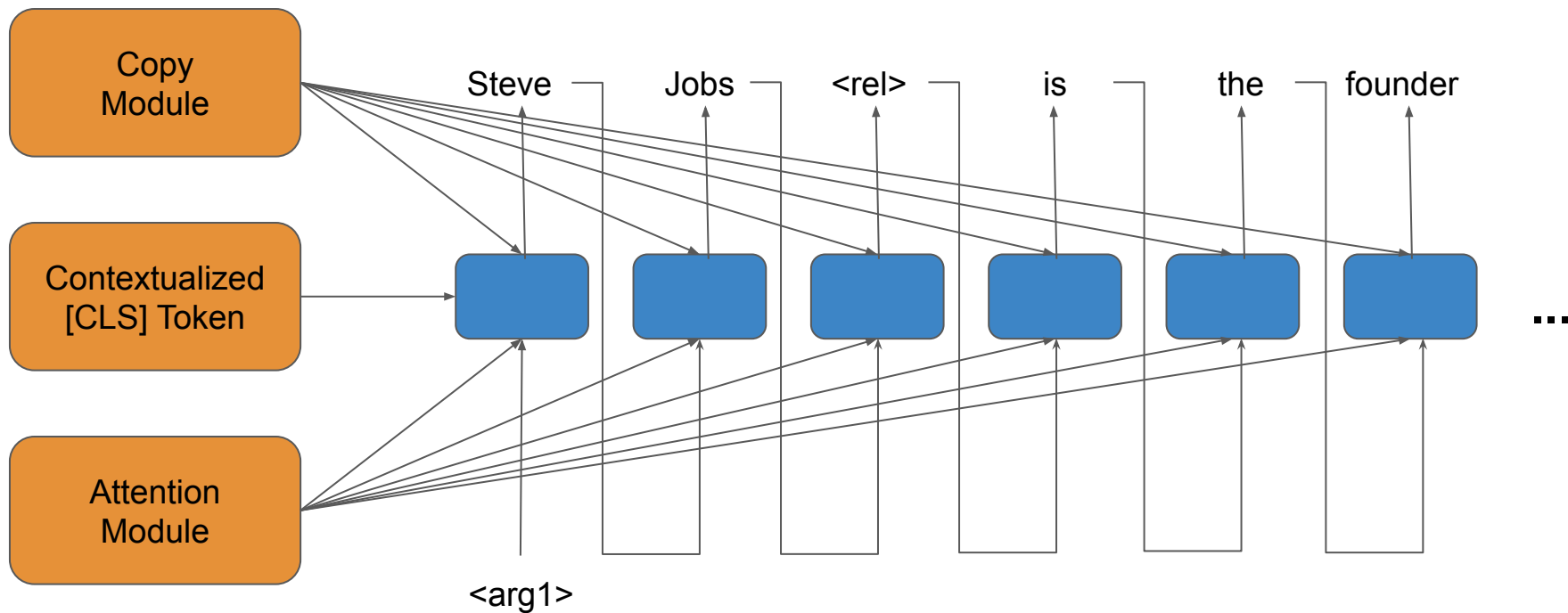
$\langle arg1 \rangle, \langle rel \rangle, \langle arg2 \rangle$

$\langle subj \rangle, \langle rel \rangle, \langle obj \rangle$

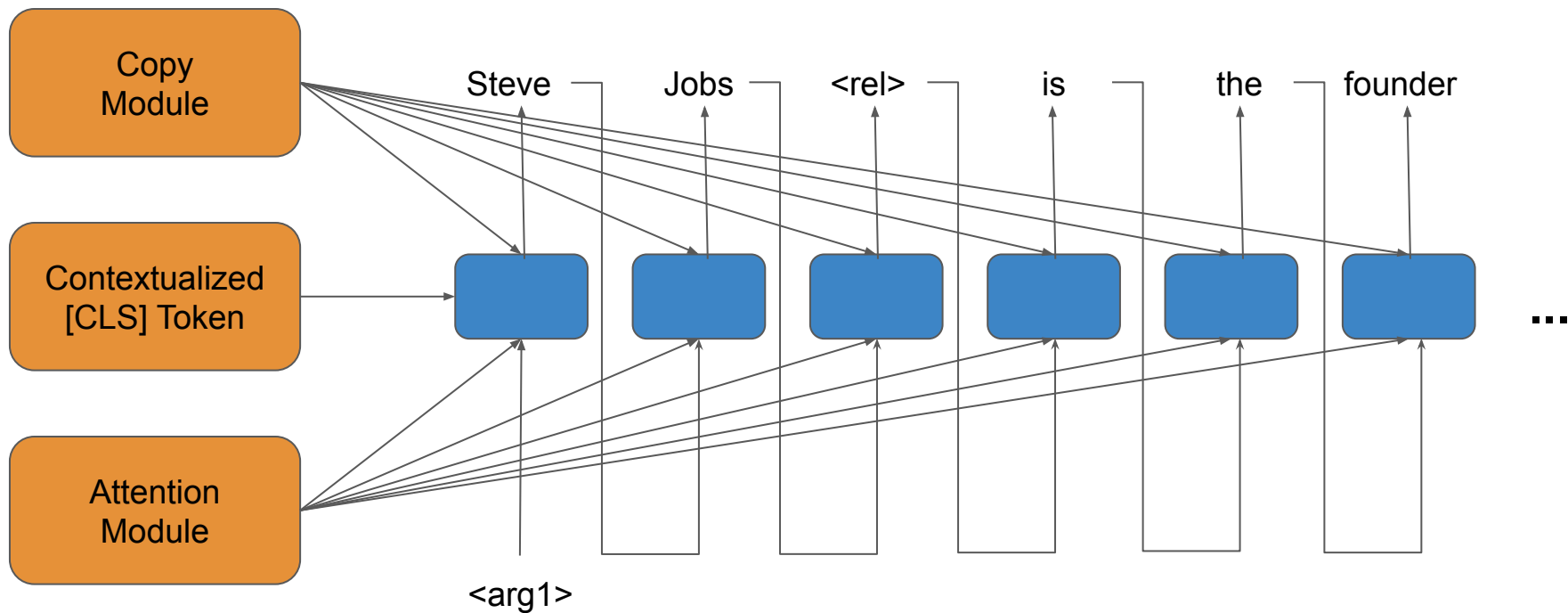
IMoJIE - Step 1 (Encoder)



IMoJIE - Step 1 (Decoder)

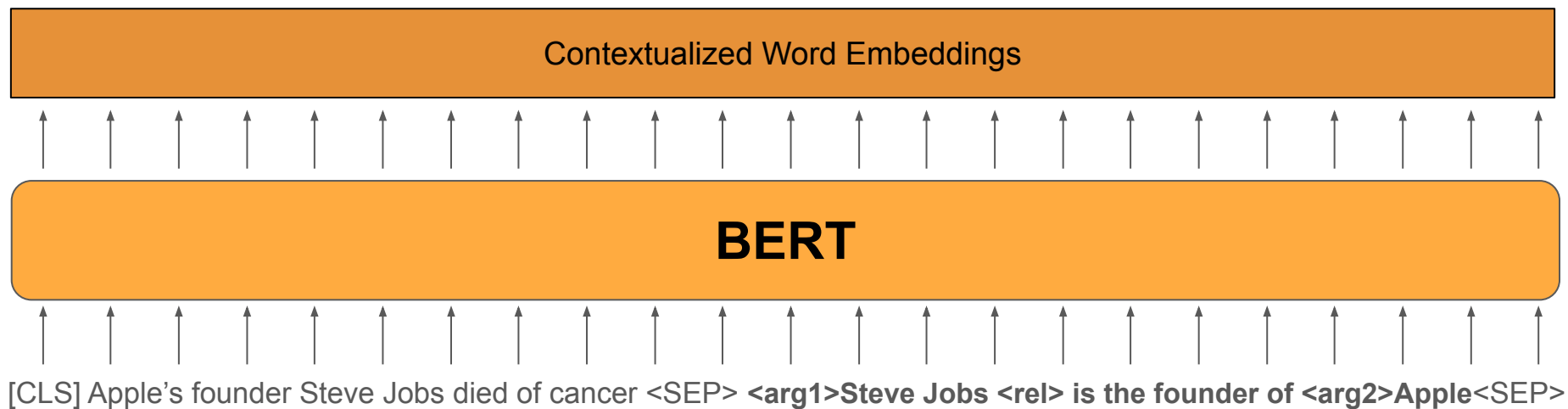


IMoJIE - Step 1 (Decoder)



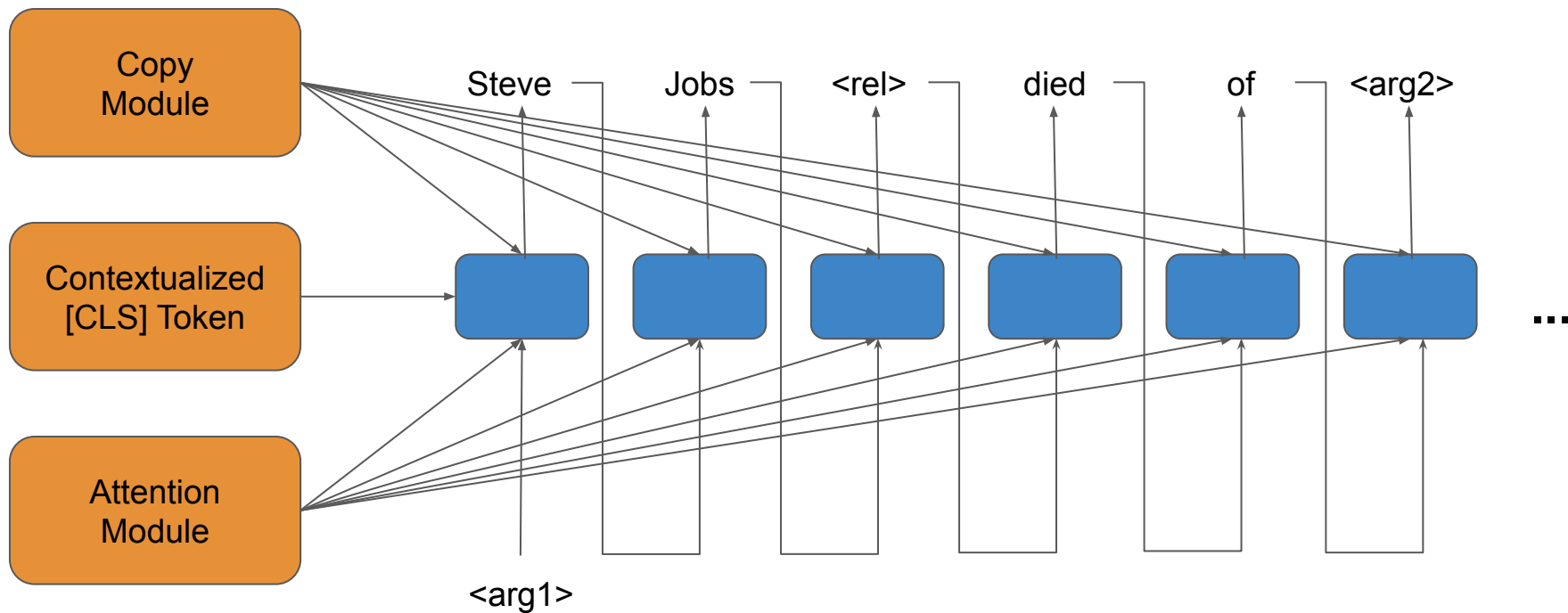
Extraction 1 : <arg1> Steve Jobs <rel> is the founder of <arg2> Apple

IMoJIE - Step 2 (Encoder)

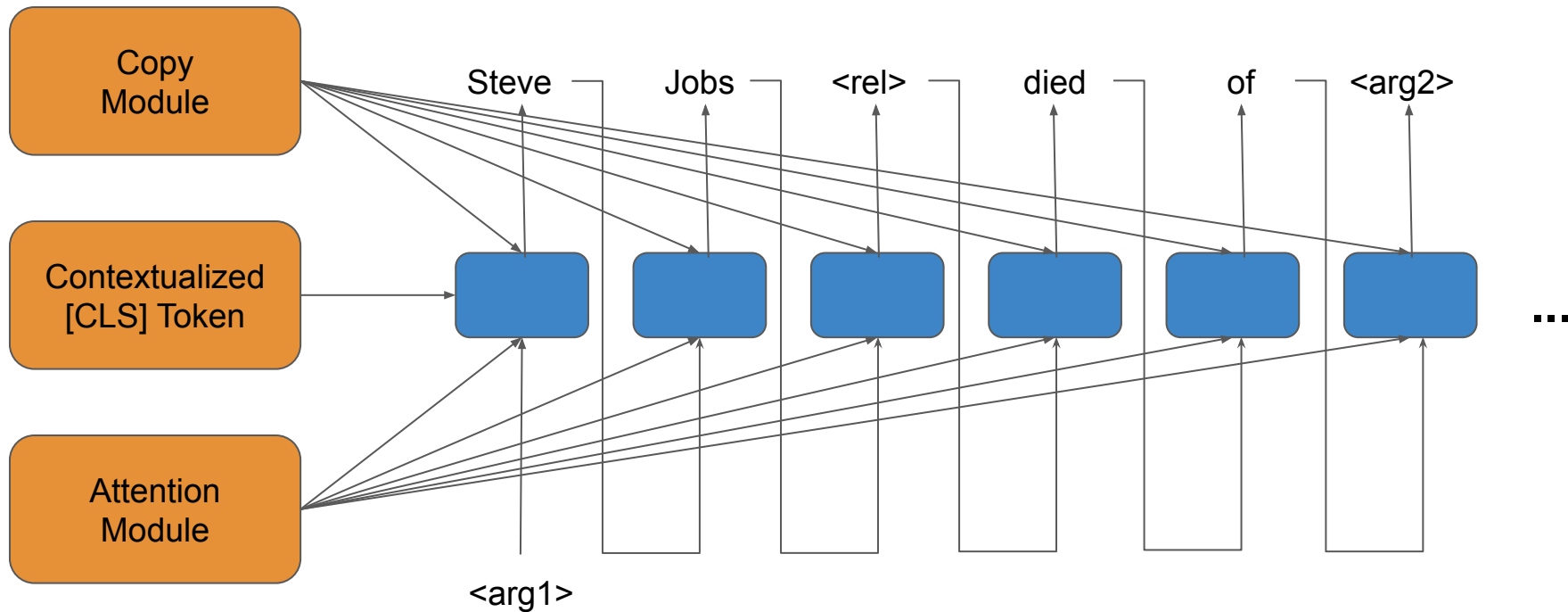


Extraction 1

IMoJIE - Step 2 (Decoder)



IMoJIE - Step 2 (Decoder)

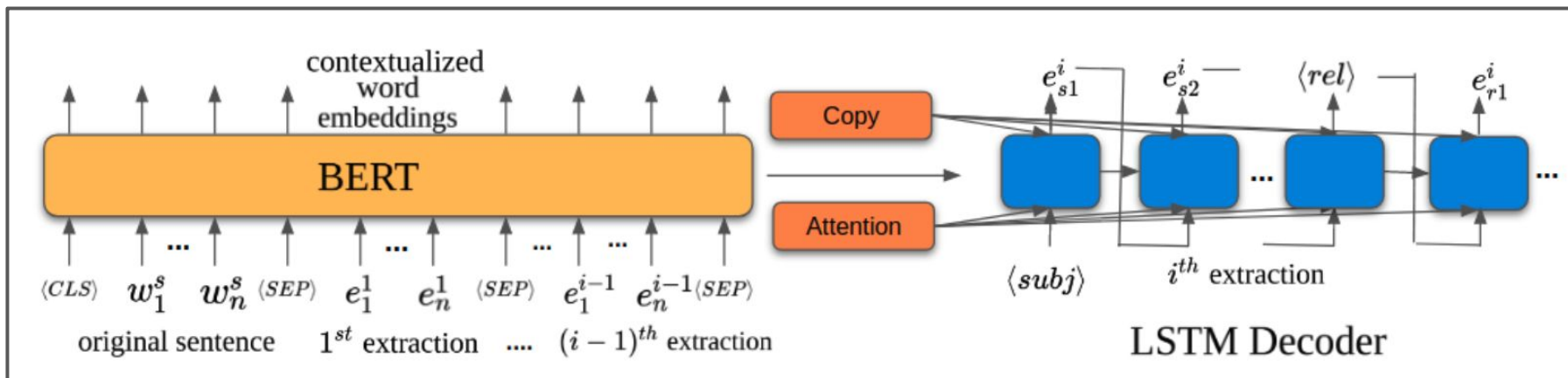


Extraction 2 : <arg1> Steve Jobs <rel> died of <arg2> cancer

IMoJIE

Extraction 1 : <arg1> Steve Jobs <rel> is the founder of <arg2> Apple

Extraction 2 : <arg1> Steve Jobs <rel> died of <arg2> cancer



Terminology

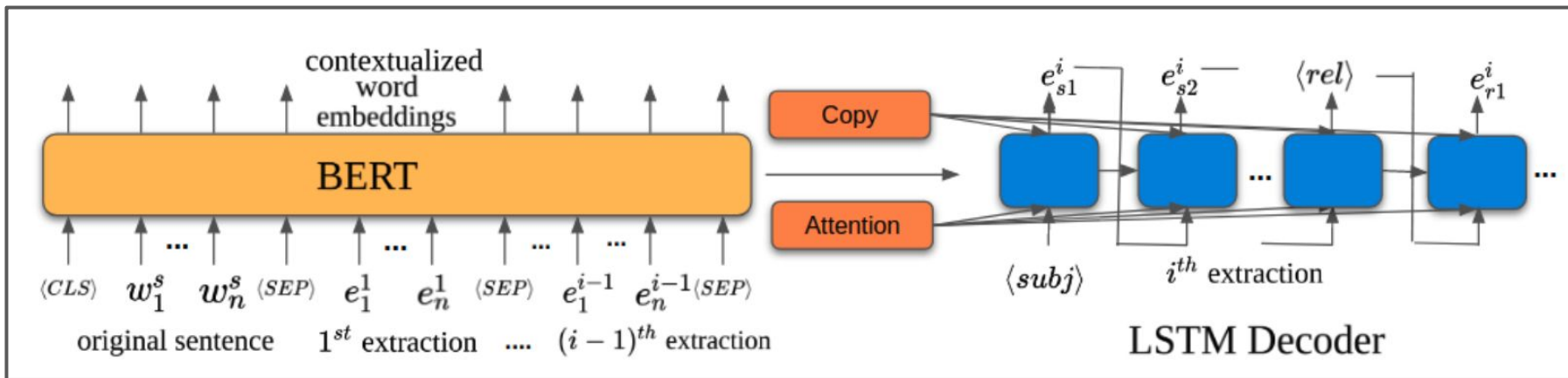
<arg1>, <rel>, <arg2>
<subj>, <rel>, <obj>

IMoJIE

Slow!

Extraction 1 : <arg1> Steve Jobs <rel> is the founder of <arg2> Apple

Extraction 2 : <arg1> Steve Jobs <rel> died of <arg2> cancer



Terminology

<arg1>, <rel>, <arg2>
<subj>, <rel>, <obj>

Labeling for OpenIE

<i>Apple's</i>	<i>founder</i>	<i>Steve</i>	<i>Jobs</i>	<i>died</i>	<i>of</i>	<i>cancer</i>	<i>[is]</i>	<i>[of]</i>	<i>[from]</i>
ARG2	REL	ARG1	ARG1	NONE	NONE	NONE	REL	REL	NONE
NONE	NONE	ARG1	ARG1	REL	REL	ARG2	NONE	NONE	NONE

Labeling for OpenIE

<i>Apple's</i>	<i>founder</i>	<i>Steve</i>	<i>Jobs</i>	<i>died</i>	<i>of</i>	<i>cancer</i>	<i>[is]</i>	<i>[of]</i>	<i>[from]</i>
ARG2	REL	ARG1	ARG1	NONE	NONE	NONE	REL	REL	NONE
NONE	NONE	ARG1	ARG1	REL	REL	ARG2	NONE	NONE	NONE



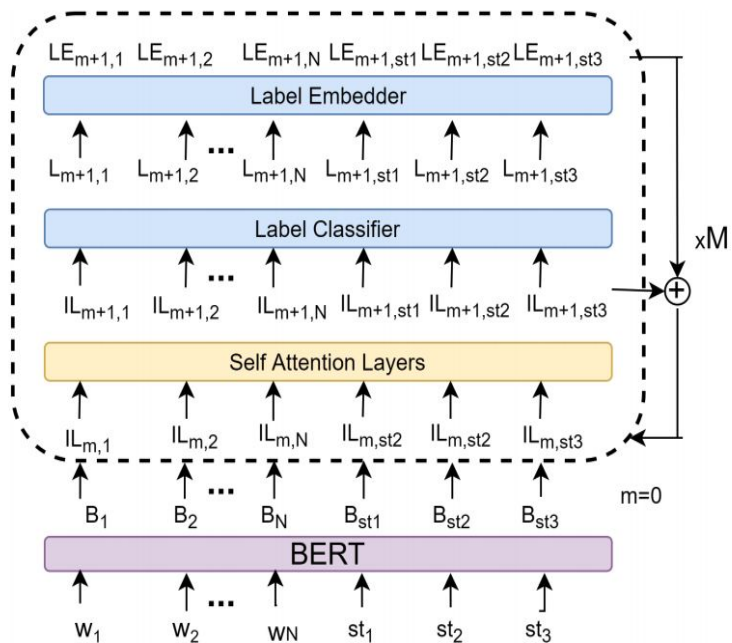
(Steve Jobs, [is] the founder [of], Apple)
(Steve Jobs, died of, cancer)

Results

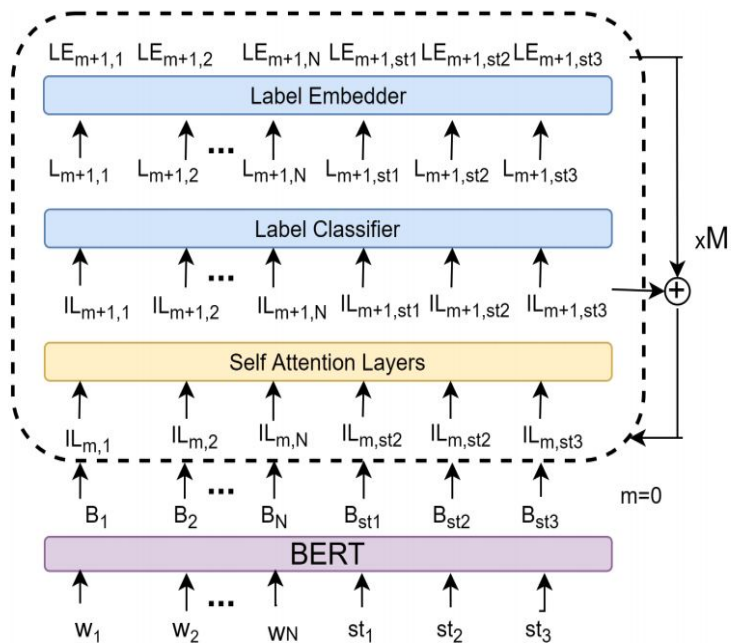
System	CaRB		Speed
	F1	AUC	Sentences/sec.
RnnOIE	49.0	26.0	149.2
IMoJIE	53.5	33.3	2.6

- Trade-off between speed and accuracy
- IMoJIE is 4.5 F1 better than RnnOIE
- RnnOIE is 60x faster than IMoJIE

IGL - Iterative Grid Labeling

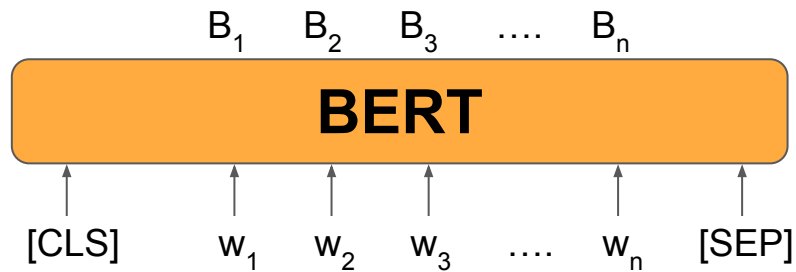


IGL - Iterative Grid Labeling

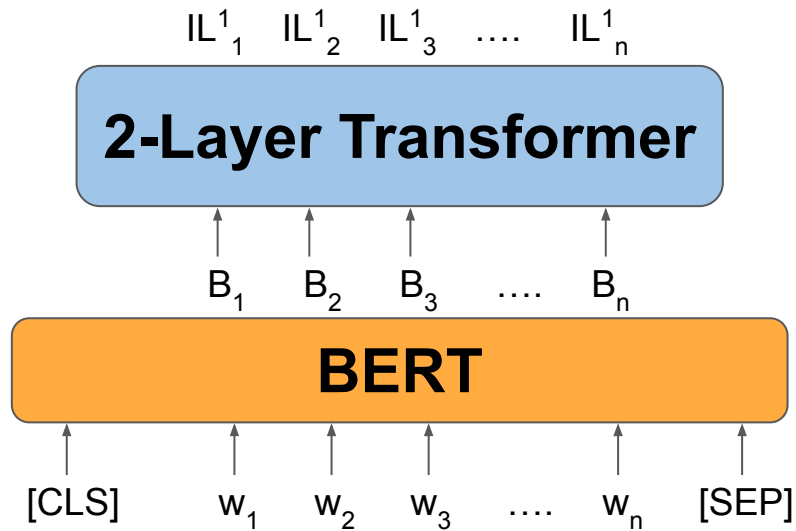


Have to capture dependencies between extractions

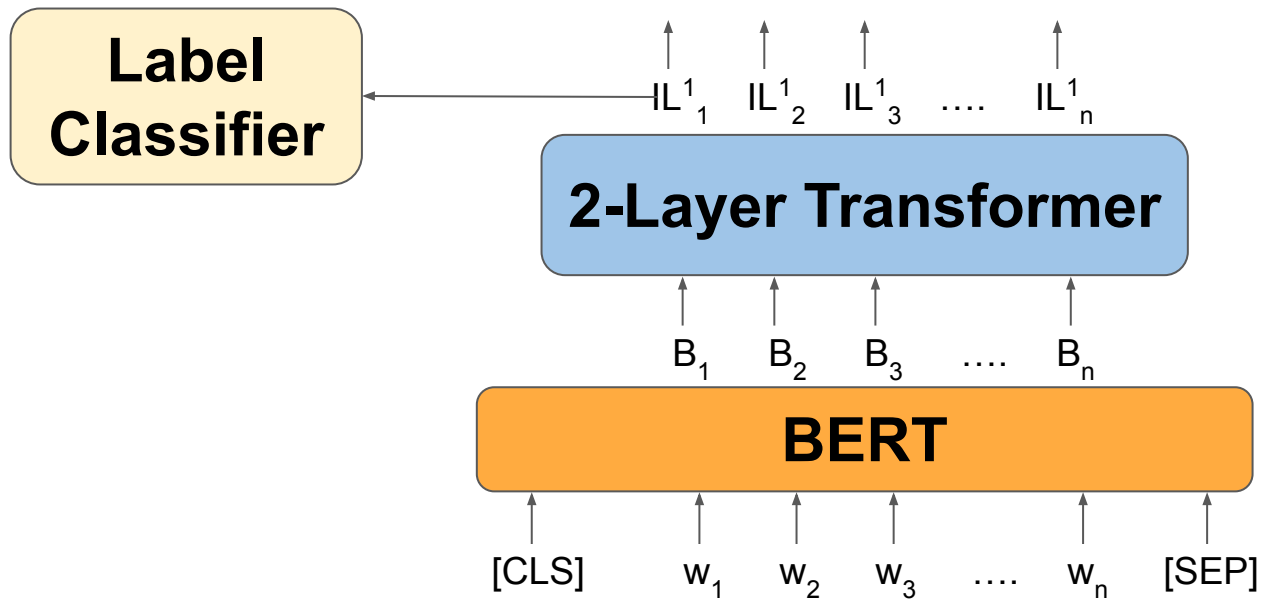
IGL - *Iterative* Grid Labeling



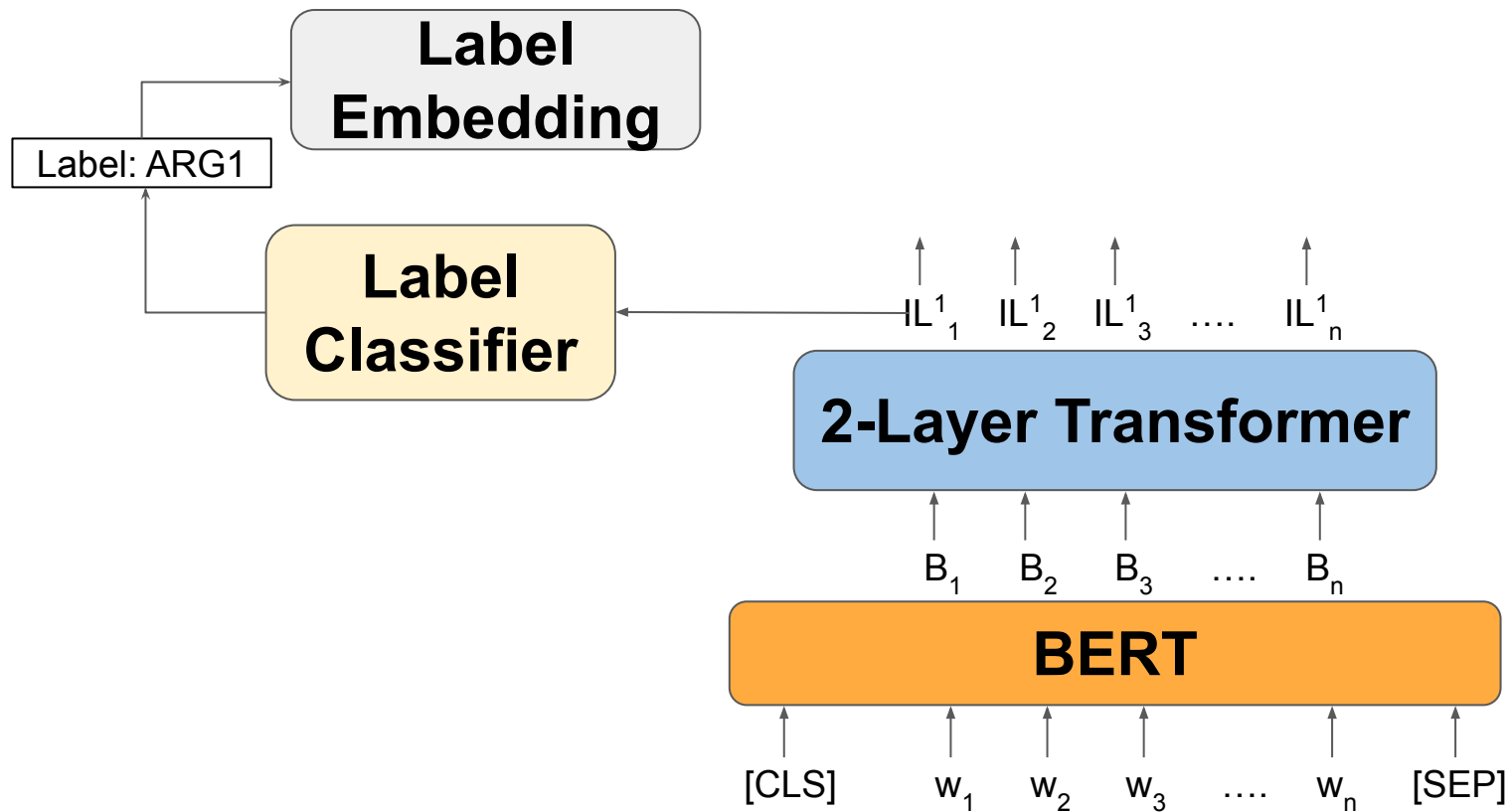
IGL - *Iterative* Grid Labeling



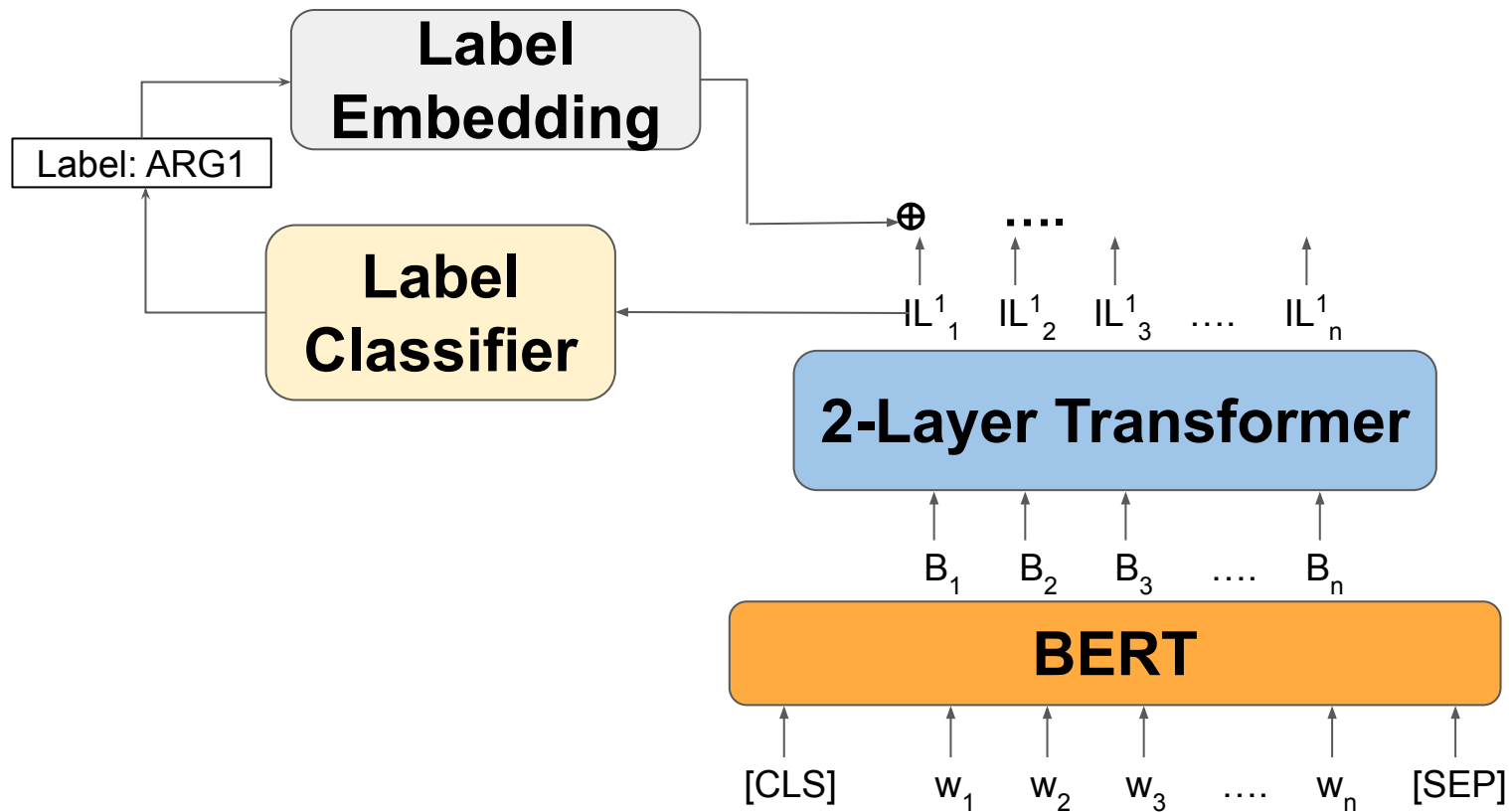
IGL - *Iterative* Grid Labeling



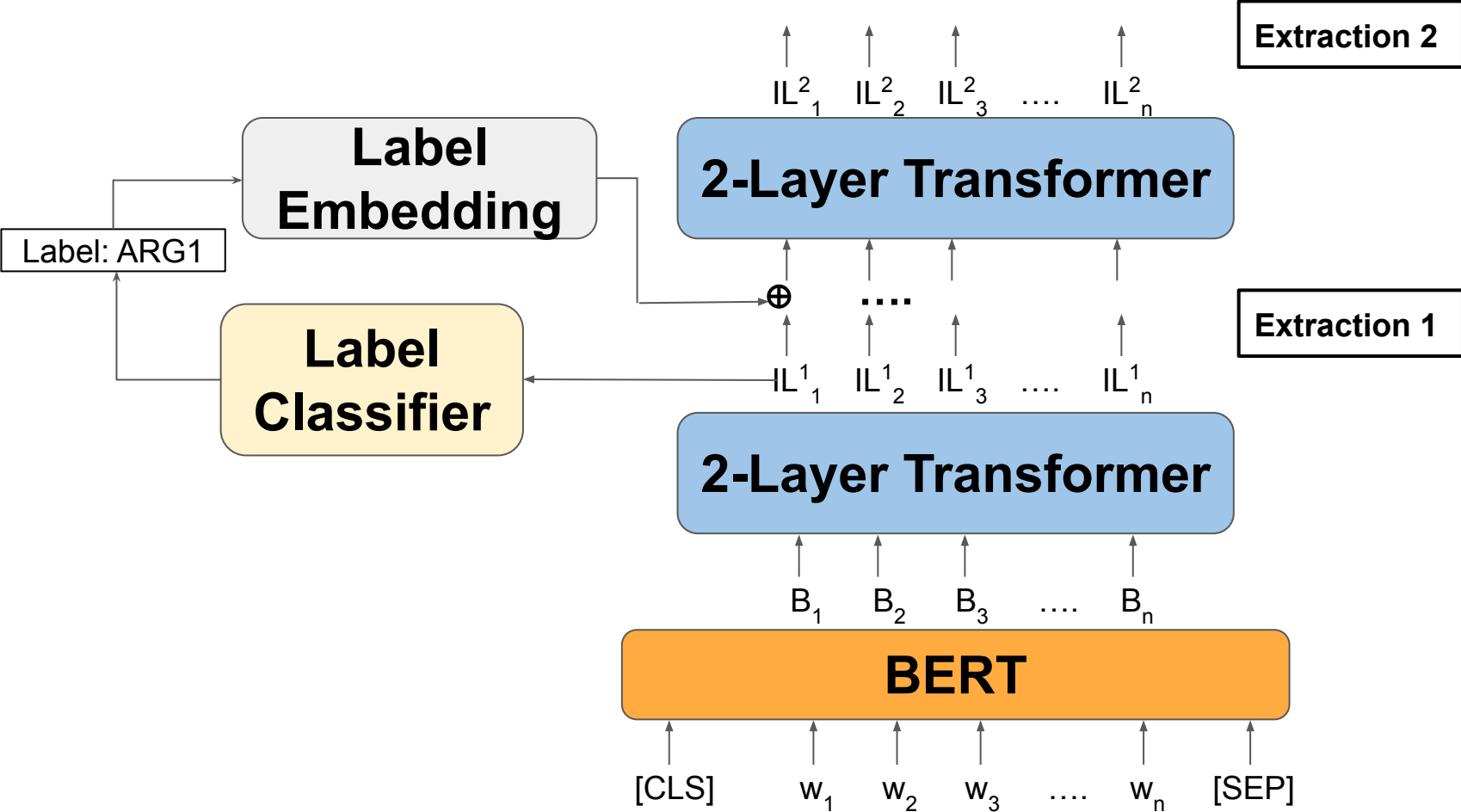
IGL - *Iterative* Grid Labeling



IGL - *Iterative* Grid Labeling



IGL - Iterative Grid Labeling



IGL - Iterative *Grid* Labeling

E5	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
E4	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
E3	ARG1	NONE	REL	REL	REL	ARG2	ARG2	NONE
E2	ARG1	NONE	REL	REL	NONE	ARG2	ARG2	NONE
E1	ARG1	ARG1	NONE	NONE	REL	NONE	ARG2	NONE

w1	w2	w3	w4	w5	w6	w7	w8
----	----	----	----	----	----	----	----

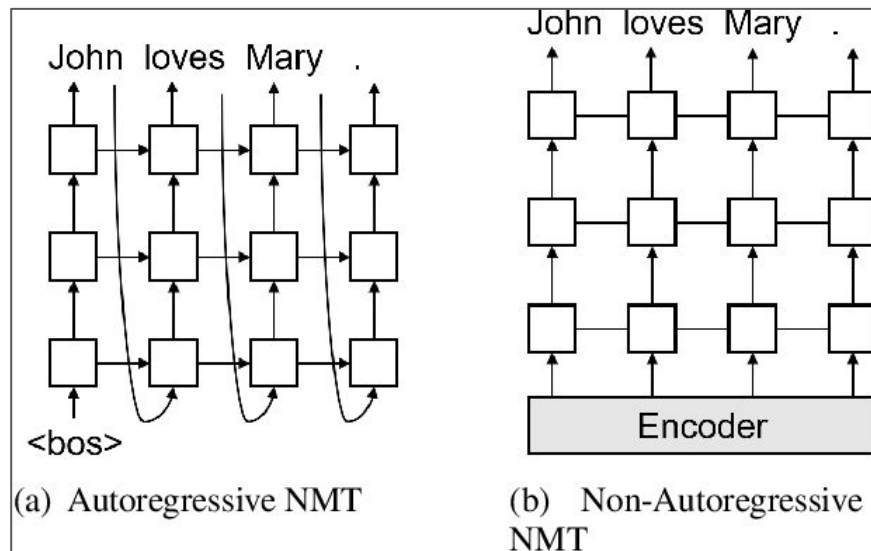
Results

System	CaRB		Speed
	F1	AUC	Sentences/sec.
RnnOIE	49.0	26.0	149.2
IMoJIE	53.5	33.3	2.6
IGL-OIE	52.4	33.7	142.0

- **60x faster** than generation based systems
- 1.1 F1 lower than IMoJIE

Autoregressive vs Non-Autoregressive

- No *sequential* nature in non-autoregressive
- Trade-off of *speed vs accuracy*

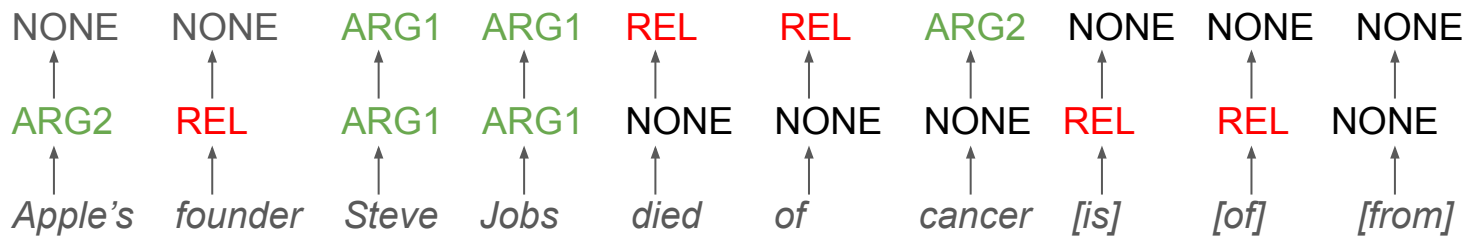


[arxiv:1906.02041](https://arxiv.org/abs/1906.02041)

(Machine Translation terminology)

Autoregressive vs Non-Autoregressive

- Iterative Grid Labeling
 - Autoregressive at *extraction* level
 - Non-autoregressive at *word* level



IGL for OpenIE

- Not accurate enough. *Why?*
 - Non-Autoregressive?

IGL for OpenIE

- Not accurate enough. *Why?*
 - Non-Autoregressive?
 - Likely! But we need that for system to be *fast!*

IGL for OpenIE

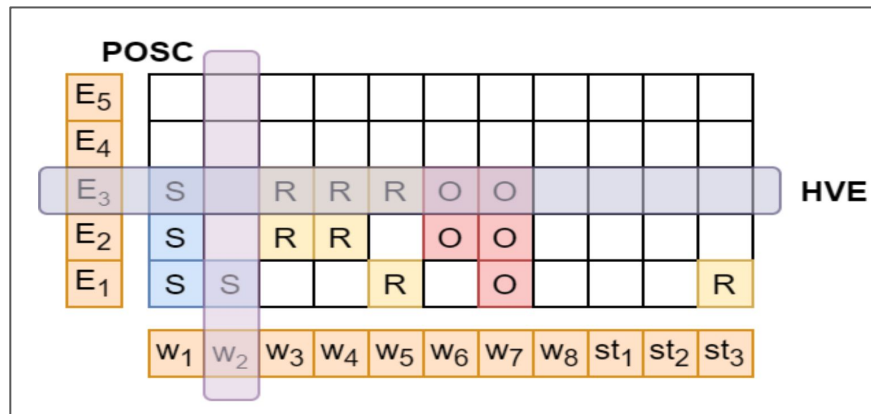
- Not accurate enough. *Why?*
 - Non-Autoregressive?
 - Likely! But we need the system to be *fast!*

Solution - **Constraints!**



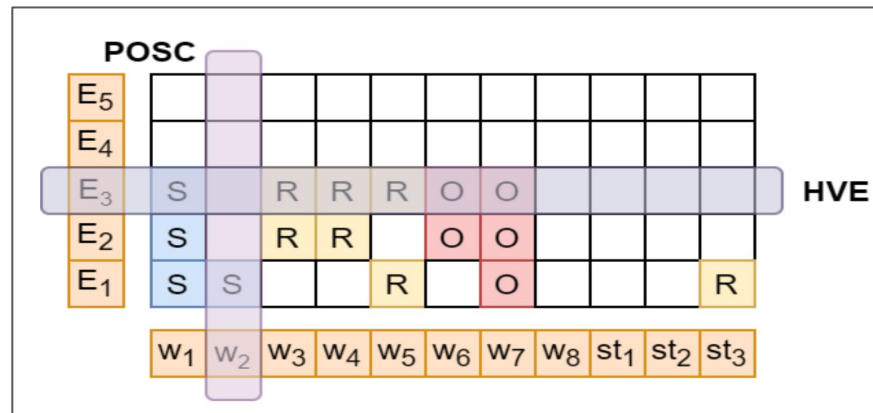
Constraints for OpenIE

- Part-of-Speech Constraint (**POSC**)
- Head Verb Coverage (**HVC**)
- Head Verb Exclusivity (**HVE**)
- Extraction Count (**EC**)



Constraints for OpenIE

- **Part-of-Speech Constraint (POSC)**
- **Head Verb Coverage (HVC)**
- **Head Verb Exclusivity (HVE)**
- **Extraction Count (EC)**



What makes a good set of extractions (for English)?

“Obama gained popularity after Oprah endorsed him for the presidency”

(Obama, gained, popularity)



What makes a good set of extractions (for English)?

“Obama gained popularity after Oprah endorsed him for the presidency”

(Obama, gained, popularity)

(Oprah, endorsed, him)

What makes a good set of extractions (for English)?

“Obama gained popularity after Oprah endorsed him for the presidency”

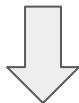
(Obama, gained, popularity)

(Oprah, endorsed him for, the presidency)

What makes a good set of extractions (for English)?

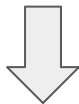
“Obama gained popularity after Oprah endorsed him for the presidency”

(Obama, gained, popularity)



(Obama, gained, popularity)

(Oprah, endorsed, him)



(Obama, gained, popularity)

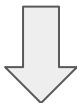
(Oprah, endorsed him for, the presidency)

What changed?

What makes a good set of extractions (for English)?

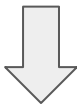
“Obama gained popularity after Oprah endorsed him for the presidency”

(Obama, gained, popularity)



(Obama, gained, popularity)

(Oprah, endorsed, him)



(Obama, gained, popularity)

(Oprah, endorsed him for, the presidency)

“Oprah”, “endorsed”, “presidency”

should have been in the set of extractions

Because they convey *information!*

POSC: All words with POS tags as *nouns (N)*, *verbs (V)*, *adjectives (JJ)*, and *adverbs (RB)* should be part of at least one extraction.

Constrained Iterative Grid Labeling (CIGL)

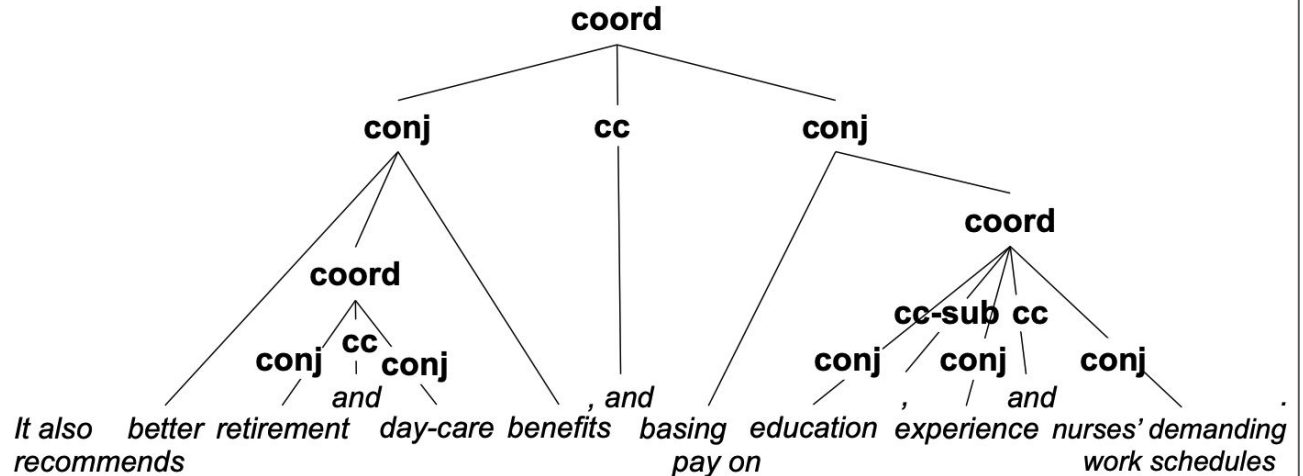
System	CaRB		Speed
	F1	AUC	Sentences/sec.
RnnOIE	49.0	26.0	149.2
IMoJIE	53.5	33.3	2.6
IGL-OIE	52.4	33.7	142.0
CIGL-OIE	54.0	35.7	142.0

- **1.6 F1** improvement over IGL-OIE
- **60x faster** than generation based systems

IGL - Scope beyond OpenIE

Coordination Analysis

e.g.) It also recommends better retirement and day-care benefits, and basing pay on education, experience and nurses' demanding work schedules.



IGL for Coordination Analysis

Jeff Bezos founded Amazon and Blue Origin and invested in Google, Grail and Zocdoc

- The sentence contains 3 coordination structures:

IGL for Coordination Analysis

Jeff Bezos *founded Amazon and Blue Origin* *and* *invested in Google, Grail and Zocdoc*

- The sentence contains 3 coordination structures:
 - CC: *and*,
 - CONJ: (*founded Amazon and Blue Origin; invested in Google, Grail and Zocdoc*)

IGL for Coordination Analysis

Jeff Bezos founded Amazon and Blue Origin and invested in Google, Grail and Zocdoc

- The sentence contains 3 coordination structures:
 - CC: *and*,
CONJ: (*founded Amazon and Blue Origin; invested in Google, Grail and Zocdoc*)
 - Coordinating Conjunction (CC): *and*,
Conjuncts (CONJ): (*Amazon; Blue Origin*)

IGL for Coordination Analysis

Jeff Bezos founded Amazon and Blue Origin and invested in Google, Grail and Zocdoc

- The sentence contains 3 coordination structures:
 - CC: *and*,
CONJ: (*founded Amazon and Blue Origin; invested in Google, Grail and Zocdoc*)
 - Coordinating Conjunction (CC): *and*,
Conjuncts (CONJ): (*Amazon; Blue Origin*)
 - CC: *and*,
CONJ: (*Google; Grail; Zocdoc*)

IGL for Coordination Analysis

Jeff Bezos founded Amazon and Blue Origin and invested in Google, Grail and Zocdoc

Jeff	founded	Amazon	and	Blue	Origin	and	invested	in	Google	Grail	and	Zocdoc
	CONJ	CONJ	CONJ	CONJ	CONJ	CC	CONJ	CONJ	CONJ	CONJ	CONJ	CONJ
		CONJ	CC	CONJ	CONJ				CONJ	CONJ	CC	CONJ

Iterative Grid Labeling - Coordination Analyzer (IGL-CA)

System	Precision	Recall	F1
(Teranishi et al., 2017)	71.5	70.7	71.0
(Teranishi et al., 2019)	75.3	75.6	75.5
BERT-Base:			
(Teranishi et al., 2019)	83.1	83.2	83.1
IGL-CA	86.3	83.6	84.9
BERT-Large:			
(Teranishi et al., 2019)	86.4	86.6	86.5
IGL-CA	88.1	87.4	87.8

- IGL-CA shows **12.3 F1** improvement compared to previous SoTA
- Improvements reduce when SoTA augmented with BERT

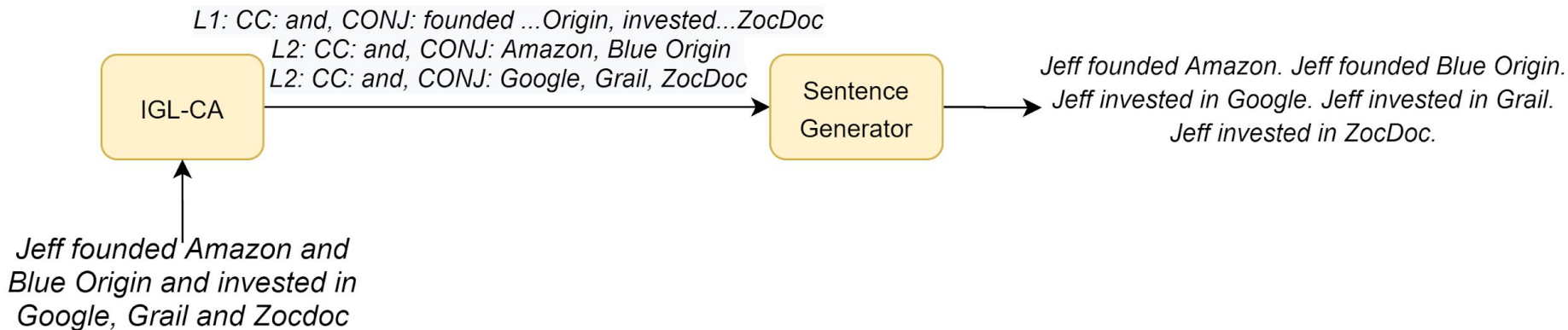
Augmenting OpenIE with Coordination Analysis

*Jeff Bezos founded
Amazon and Blue Origin
and invested in Google,
Grail and Zocdoc*

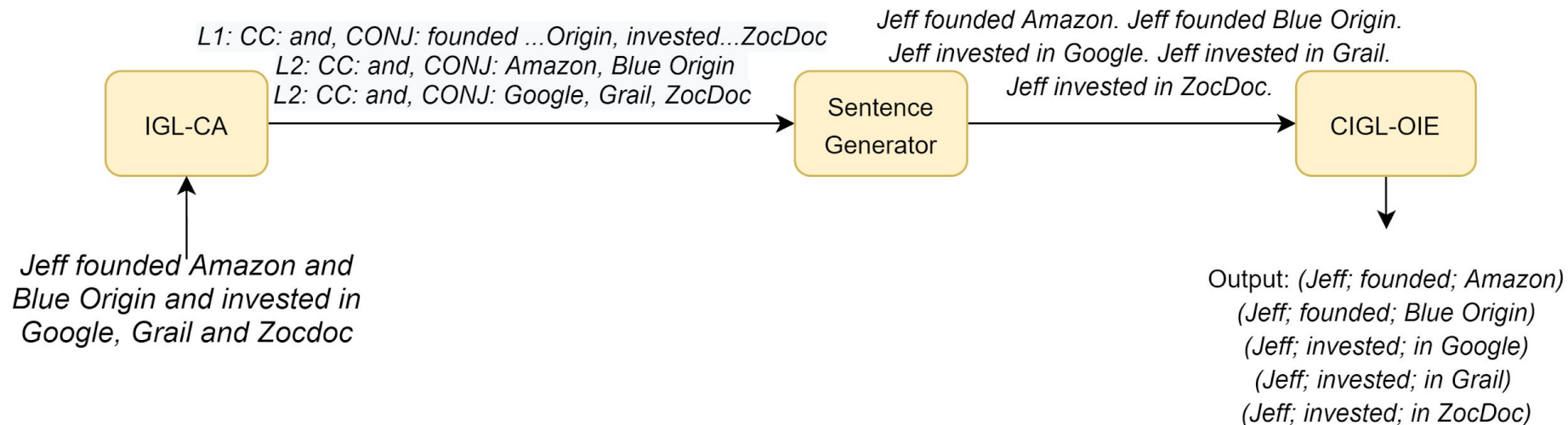
IGL-CA

*L1: CC: and, CONJ: founded ...Origin, invested...ZocDoc
L2: CC: and, CONJ: Amazon, Blue Origin
L2: CC: and, CONJ: Google, Grail, ZocDoc*

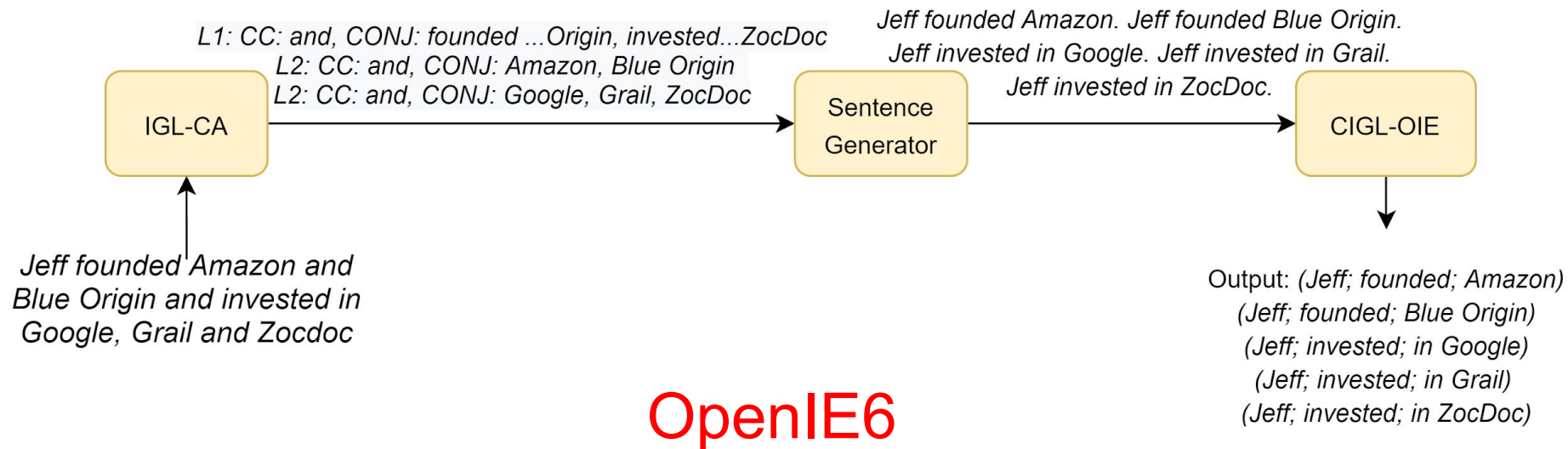
Augmenting OpenIE with Coordination Analysis



Augmenting OpenIE with Coordination Analysis



Augmenting OpenIE with Coordination Analysis



OpenIE6 = CIGL-OIE + IGL-CA

System	CaRB		CaRB(1-1)		OIE16-C		Wire57-C	Speed
	F1	AUC	F1	AUC	F1	AUC	F1	Sentences/sec.
RnnOIE	49.0	26.0	39.5	18.3	56.0	32.0	26.4	149.2
IMoJIE	53.5	33.3	41.4	22.2	56.8	39.6	36.0	2.6
IGL-OIE	52.4	33.7	41.1	22.9	55.0	36.0	34.9	142.0
CIGL-OIE	54.0	35.7	42.8	24.6	59.2	40.0	36.8	142.0
CIGL-OIE + IGL-CA (OpenIE6)	52.7	33.7	46.4	26.8	65.6	48.4	40.0	31.7

- OpenIE-6 shows improvement in 3 out of 4 metrics
- Observe drop in speed due to additional module involved

Conclusions

- *Iterative Grid Labeling:*
 - 60x faster than generation-based systems for Open Information Extraction
 - Establishes a new state-of-art for task of Coordination Analysis
- *Constrained Iterative Grid Labeling:*
 - Establishes a new state-of-art for task of Open Information Extraction
- *OpenIE6:*
 - Fast and accurate OpenIE system that handles conjunctive sentences
- Code, training data and pretrained models are available at <https://github.com/dair-iitd/openie6>