

## Stanford CoreNLP OpenIE

What is OpenIE? .....	1
System requirements .....	2
Java.....	2
Input .....	2
Passive sentences .....	2
Output.....	2
References .....	2

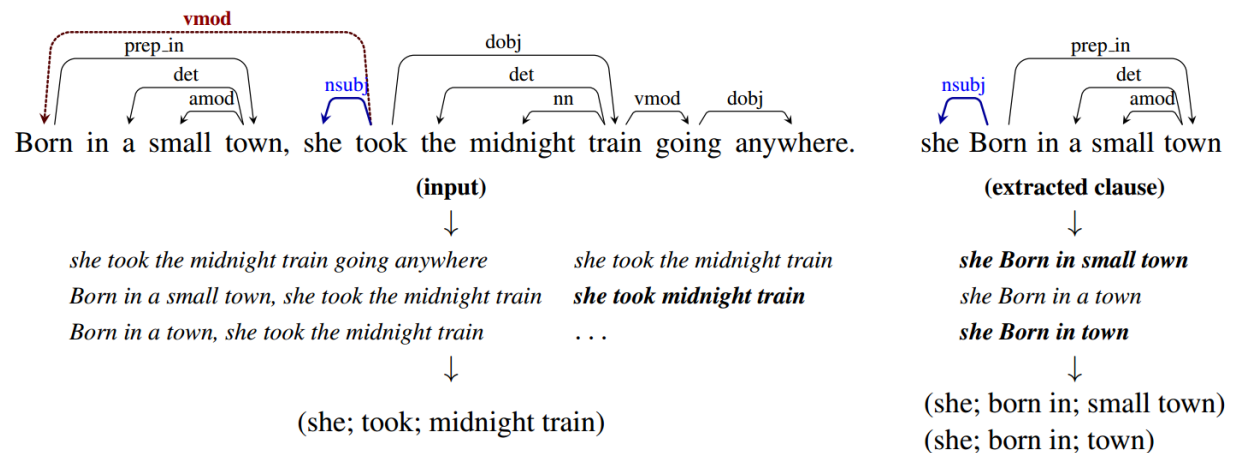
### What is OpenIE?

From the Stanford CoreNLP website (<https://nlp.stanford.edu/software/openie.html>) we read:

Open information extraction (open IE) refers to the **extraction of relation tuples**, typically binary relations, from plain text, such as (Mark Zuckerberg; founded; Facebook). The central difference from other information extraction is that the schema for these relations does not need to be specified in advance; typically the relation name is just the text linking two arguments. For example, Barack Obama was born in Hawaii would create a triple (Barack Obama; was born in; Hawaii), corresponding to the open domain relation was-born-in (Barack-Obama, Hawaii). This software is a Java implementation of an open IE system described in the paper:

Gabor Angeli, Melvin Johnson Premkumar, and Christopher D. Manning.  
Leveraging Linguistic Structure For Open Domain Information Extraction. In  
Proceedings of the Association of Computational Linguistics (ACL), 2015.

The system first splits each sentence into a set of entailed clauses. Each clause is then maximally shortened, producing a set of entailed shorter sentence fragments. These fragments are then segmented into OpenIE triples, and output by the system. An illustration of the process is given for an example sentence below:



## System requirements

OpenIE requires Java 8+ to be installed, and generally requires around 50MB of memory in addition to the memory used by the part of speech tagger and dependency parser (and optional named entity recognizer). Stanford CoreNLP recommends running java with around 1gb of memory (2gb if using NER) to be safe (i.e., `java -mx1g`).

## Java

**You will need to download and install Java. See the TIPS file TIPS\_NLP\_Java download install run.pdf**

## Input

The Stanford CoreNLP OpenIE takes in input a single text file or a set of text files in a directory.

## Passive sentences

The SVO script converts passive sentences to active ones. When the passive subject (the agent) of the sentence is not available, the script will insert the default unknown subject as **Someone?**

## Output

The NLP Suite uses Stanford CoreNLP OpenIE in the SVO script. Contrary to the NLP Suite Stanford CoreNLP parsers, the OpenIE script does not produce in output a CoNLL table. Just a csv file with the SVO triplets.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Sentence Index	S	V	O/A	LOCATION	PERSON	TIME	TIME_STAMP	Sentence							
2	2	It	was	day			the day after Christmas		It was the day after Christmas.							
3	3	rows	looming	hospital corridor			this day 45 years since the day		As I was wheeled backward through the glaring hospital corridor, rows of overhead lig							
4	3	rows	looming	me			this day 45 years since the day		As I was wheeled backward through the glaring hospital corridor, rows of overhead lig							
5	3	rows	looming	rows			this day 45 years since the day		As I was wheeled backward through the glaring hospital corridor, rows of overhead lig							
6	5	anesthetist	had warned	me					The anesthetist had warned me that the operating room would feel cold.							
7	6	He	introduce	himself					He came to introduce himself to me in the little cubicle where I lay on the gurney.							
8	6	I	lay	gurney					He came to introduce himself to me in the little cubicle where I lay on the gurney.							
9	7	wife	held	hand		Kathleen			My wife, Kathleen, held my hand and tried to look calm.							
10	8	corridor	felt	hospital gown					But even the corridor already felt cold in my hospital gown.							
11	9	Everything	was	motion					Everything was in slow motion.							
12	13	I	looked	family book					I looked it up in our family medical book, reading about paralysis, blindness, seizures,							
13	13	I	looked	it					I looked it up in our family medical book, reading about paralysis, blindness, seizures,							
14	14	I	remembered	recital		John Wright			I remembered the clear, calm voice of Dr John Wright explaining how he was going to							
15	14	I	remembered	recital after-effects		John Wright			I remembered the clear, calm voice of Dr John Wright explaining how he was going to							
16	14	I	remembered	voice		John Wright			I remembered the clear, calm voice of Dr John Wright explaining how he was going to							
17	14	I	remembered	voice Dr John Wright		John Wright			I remembered the clear, calm voice of Dr John Wright explaining how he was going to							
18	14	he	cut	hole		John Wright			I remembered the clear, calm voice of Dr John Wright explaining how he was going to							
19	14	he	cut	skull		John Wright			I remembered the clear, calm voice of Dr John Wright explaining how he was going to							
20	15	I	graded	papers		Kathleen	Christmas		I recalled Kathleen's marvelous calm as Christmas approached, as I frantically graded							
21	15	I	graded	students		Kathleen	Christmas		I recalled Kathleen's marvelous calm as Christmas approached, as I frantically graded							
22	15	I	graded	students university		Kathleen	Christmas		I recalled Kathleen's marvelous calm as Christmas approached, as I frantically graded							
23	15	I	recalled	Kathleen calm		Kathleen	Christmas		I recalled Kathleen's marvelous calm as Christmas approached, as I frantically graded							

## References

Klein, Dan and Christopher D. Manning. 2003. "Accurate Unlexicalized Parsing." *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pp. 423-430.

TIPS\_NLP\_Java download install run.pdf

TIPS\_NLP\_Stanford CoreNLP parser.pdf

TIPS\_NLP\_Stanford CoNLL table.pdf