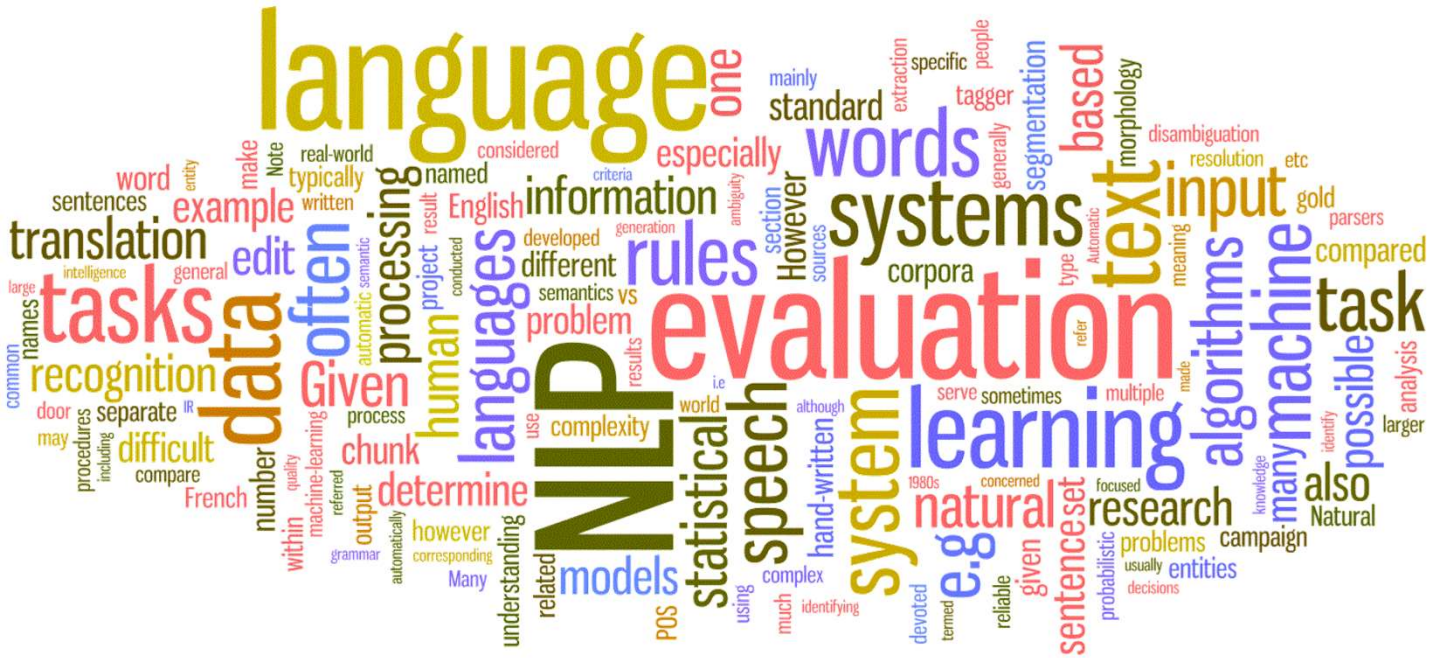


Lecture 14: Information Extraction

Xu Ruifeng

Harbin Institute of Technology, Shenzhen

Last Time

- Text Clustering: Background
- Text Clustering Algorithms
  - Partitional clustering
  - Hierarchical clustering
- Text Clustering Evaluation

Today's class

- What is Information Extraction
- Named Entity Recognition
- Relation Extraction
- Extracting Times
  - Temporal Expression Extraction
  - Temporal normalization
- Event Extraction

Information Extraction 信息抽取

- Is the process of identifying within text instances of specified classes of entities and of predications involving these entities. (Grishman, 1997)

- Information extraction (IE) systems
  - Find and understand limited relevant parts of texts
  - Gather information from many pieces of text
  - Produce a structured representation of relevant information:
    - ***relations*** (in the database sense), a.k.a.,
    - **a *knowledge base***
  - Goals
    - Organize information so that it is useful to people
    - Put information in a semantically precise form that allows further inferences to be made by computer algorithms

幻灯片 4

巫继鹏1 巫继鹏, 2020/12/8

Information Extraction 信息抽取

- IE systems extract clear, factual information
  - Roughly: who did what to whom when?

- Turns the **unstructured information** embedded in texts into **structured data**.

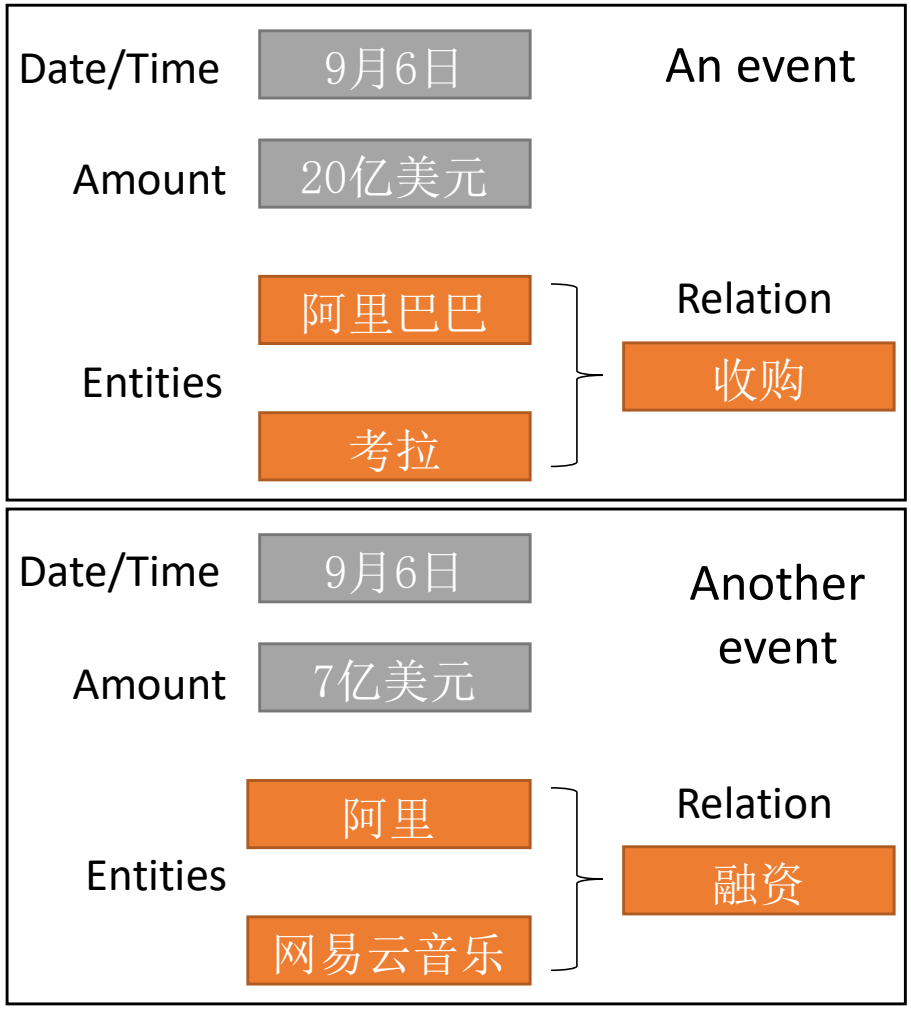
- E.g.,
  - Gathering earnings, profits, board members, head quarters, etc. from company reports / financial news.
  - Gathering patient's condition and treatment plan from the medical report.

Example 1

Text

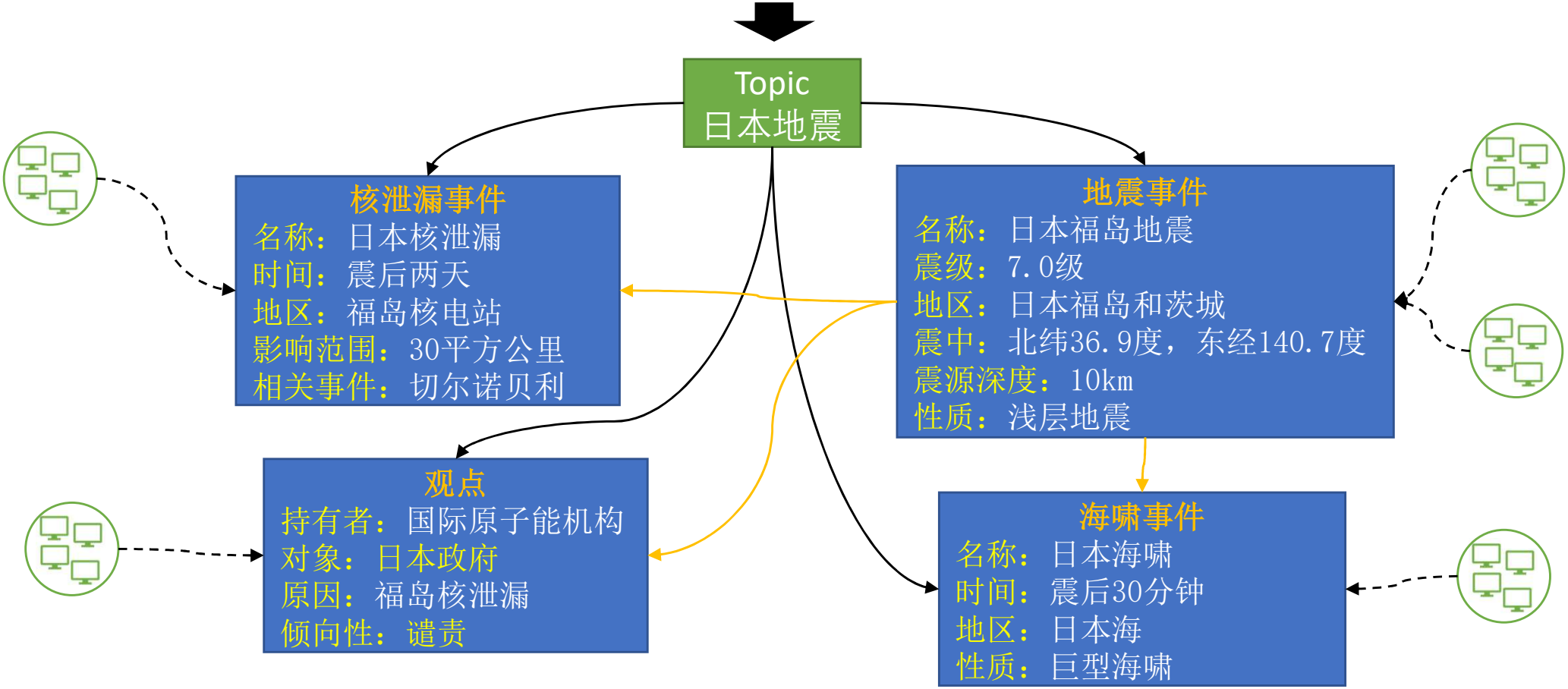
尘埃落定。9月6日，中国跨境电商行业最大的一起并购案浮出水面，阿里巴巴集团以20亿美元全资收购网易旗下跨境电商平台考拉。同时，阿里还作为领投方参与了网易云音乐B2轮7亿美元的融资。但网易仍会保持对网易云音乐的绝对控股。

Information/Knowledge



Example 2

2011年4月11日17点16分，日本东北部的福岛和茨城地区发生里氏7.0级强烈地震（震中北纬36.9度、东经140.7度，即福岛西南30公里左右的地方，震源深度10公里，属于浅层地震）当局已经发布海啸预警震后约30分钟后在日本海地区发生巨型海啸，同时造成福岛核电站出现核泄漏震后第十天，国际原子能机构对于日本政府反应迟钝进行了谴责。



Low-level Information Extraction

- Is now available in applications like Apple or Google mail, WeChat and web indexing.

A-Rain 邀请您参加腾讯会议
会议主题：A-Rain预定的会议
会议时间：2020/12/4 19:30-20:00

点击链接入会，或添加至会议列表：
https://meeting.tencent.com/s/0gNoZZvvJnsY

会议 ID：759 306 731

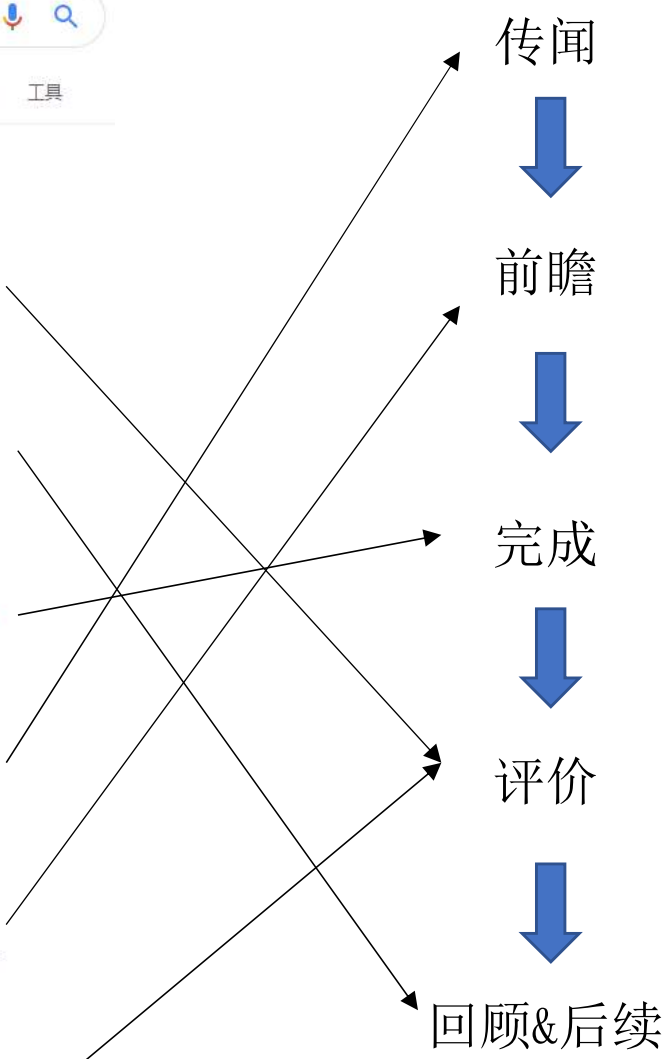
手机一键拨号入会
+8675536550000,,759306731# (中国大陆)
+85230018898,,,2,759306731# (中国香港)

根据您的位置拨号
+8675536550000 (中国大陆)
+85230018898 (中国香港)

- Often seems to be based on and name lists. **regular expressions**

Low-level Information Extraction

IE vs. IR(Information Retrieval)



Why is IE hard on the web?

When?

Who?

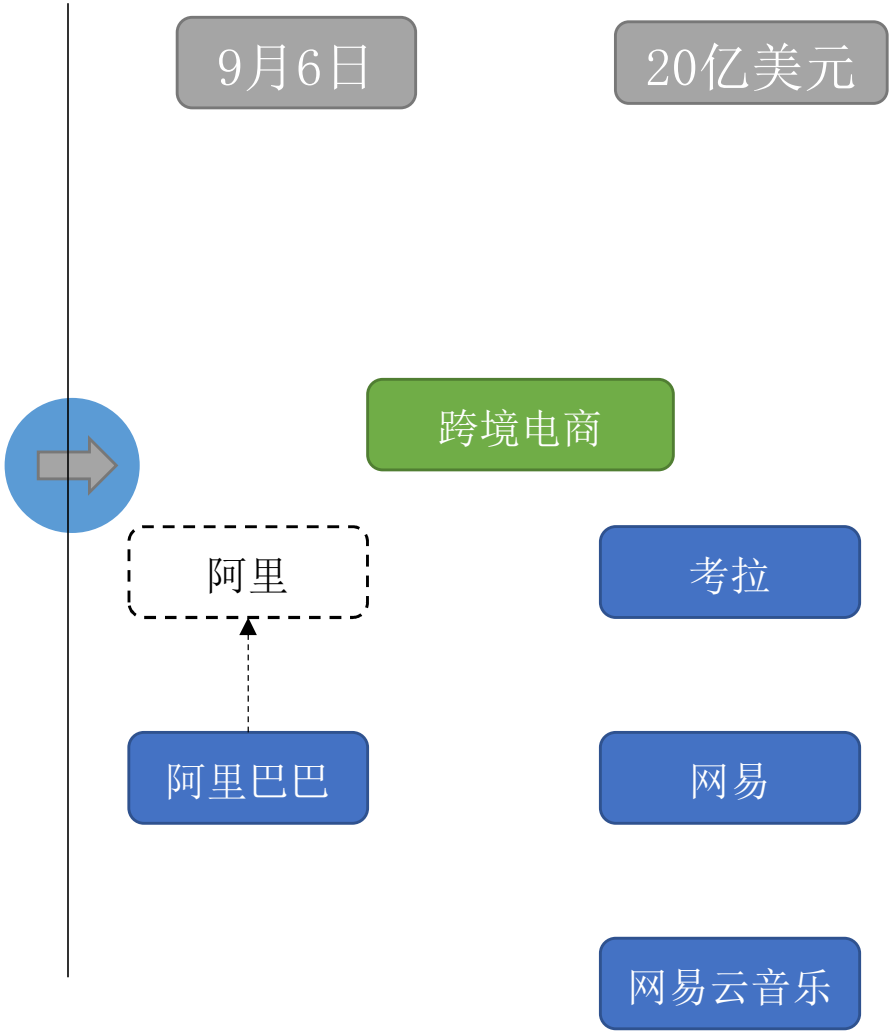
Where?

What?

尘埃落定。9月6日，中国跨境电商行业最大的一起并购案浮出水面，阿里巴巴集团以20亿美元全资收购网易旗下跨境电商平台考拉。同时，阿里还作为领投方参与了网易云音乐B2轮7亿美元的融资。但网易仍会保持对网易云音乐的绝对控股。

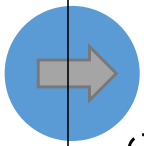
Why is IE hard on the web?

尘埃落定。9月6日，中国跨境电商行业最大的一起并购案浮出水面，阿里巴巴集团以20亿美元全资收购网易旗下跨境电商平台考拉。同时，阿里还作为领投方参与了网易云音乐B2轮7亿美元的融资。但网易仍会保持对网易云音乐的绝对控股。



Why is IE hard on the web?

尘埃落定。9月6日，中国跨境电商行业最大的一起并购案浮出水面，阿里巴巴集团以20亿美元全资收购网易旗下跨境电商平台考拉。同时，阿里还作为领投方参与了网易云音乐B2轮7亿美元的融资。但网易仍会保持对网易云音乐的绝对控股。



Why is IE hard on the web?

阿里巴巴20亿美元收购网易考拉 一场没有输家的比赛
2019-09-06 11:23:34 每日经济时报

Which year?

9月6日

2019年9月6日

尘埃落定。9月6日，中国跨境电商行业最大的一起并购案浮出水面，阿里巴巴集团以20亿美元全资收购网易旗下跨境电商平台考拉。同时，阿里还作为领投方参与了网易云音乐B2轮7亿美元的融资。但网易仍会保持对网易云音乐的绝对控股。

Why is IE hard on the web?

新华社北京9月6日电 中国外汇交易中心9月6日受权公布人民币对美元、欧元、日元、港元、英镑、澳元、新西兰元、新加坡元、瑞士法郎、加元、林吉特、卢布、兰特、韩元、迪拉姆、里亚尔、福林、兹罗提、丹麦克朗、瑞典克朗、挪威克朗、里拉、墨西哥比索及泰铢的市场汇价。

9月6日人民币汇率中间价如下：

100美元 708.55人民币

Which year? How much?

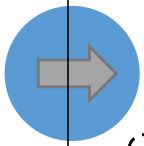
9月6日 20亿美元

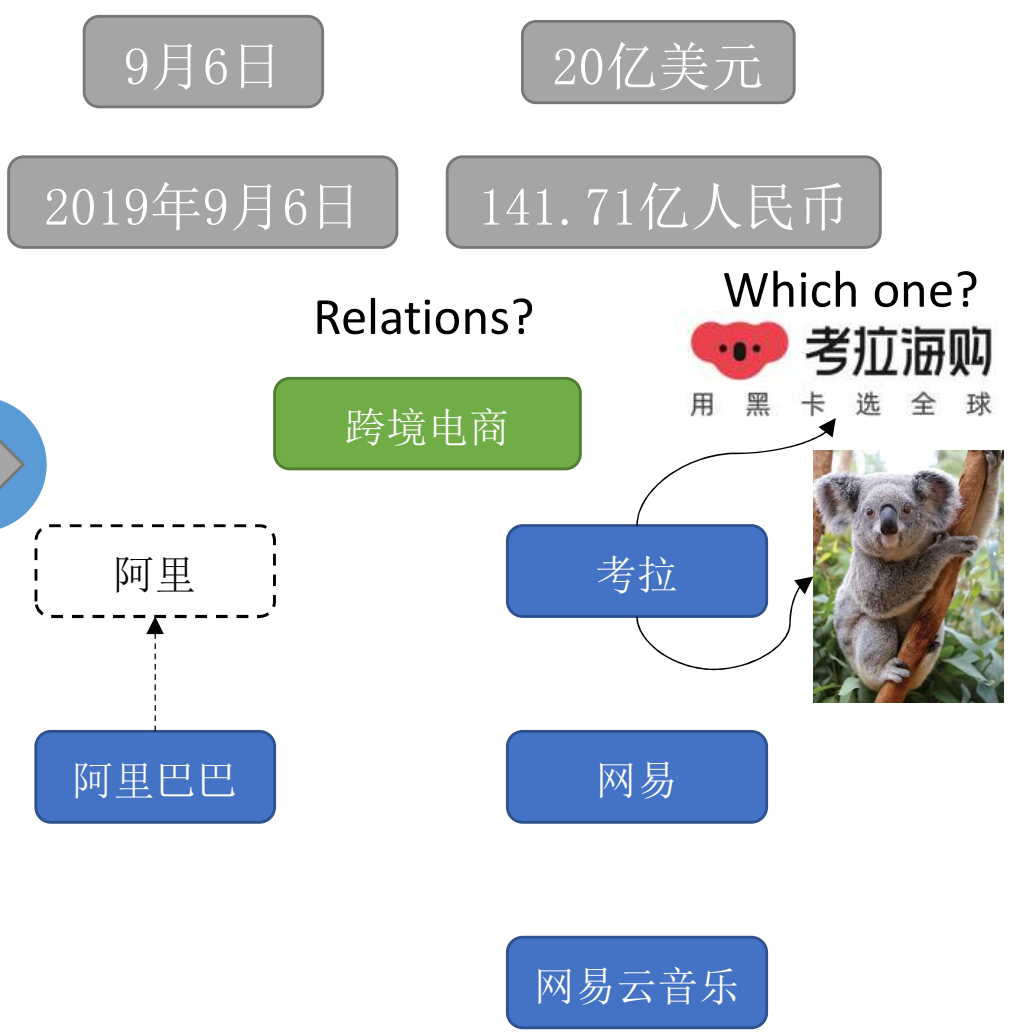
2019年9月6日 141.71亿人民币

尘埃落定。9月6日，中国跨境电商行业最大的一起并购案浮出水面，阿里巴巴集团以20亿美元全资收购网易旗下跨境电商平台考拉。同时，阿里还作为领投方参与了网易云音乐B2轮7亿美元的融资。但网易仍会保持对网易云音乐的绝对控股。

Why is IE hard on the web?

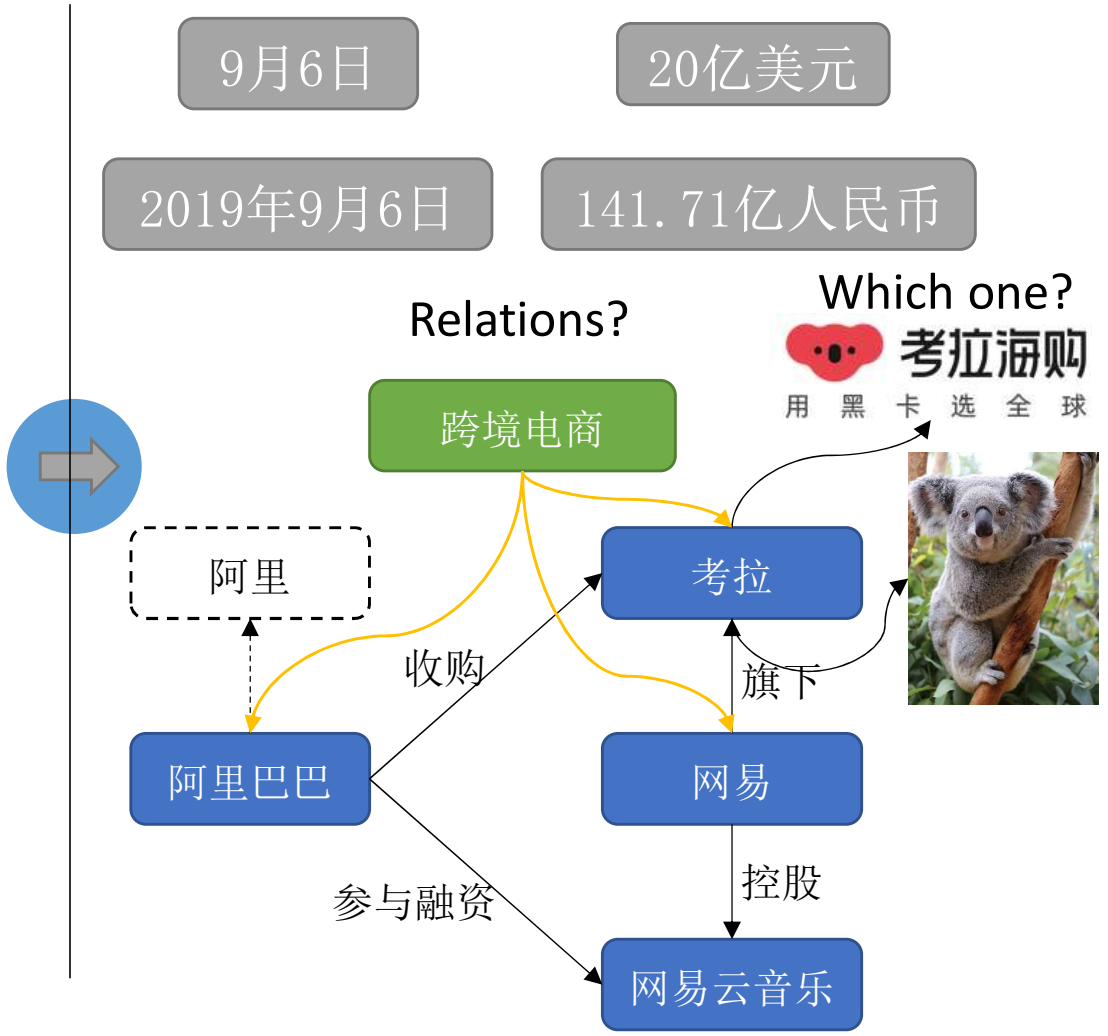
尘埃落定。9月6日，中国跨境电商行业最大的一起并购案浮出水面，阿里巴巴集团以20亿美元全资收购网易旗下跨境电商平台考拉。同时，阿里还作为领投方参与了网易云音乐B2轮7亿美元的融资。但网易仍会保持对网易云音乐的绝对控股。





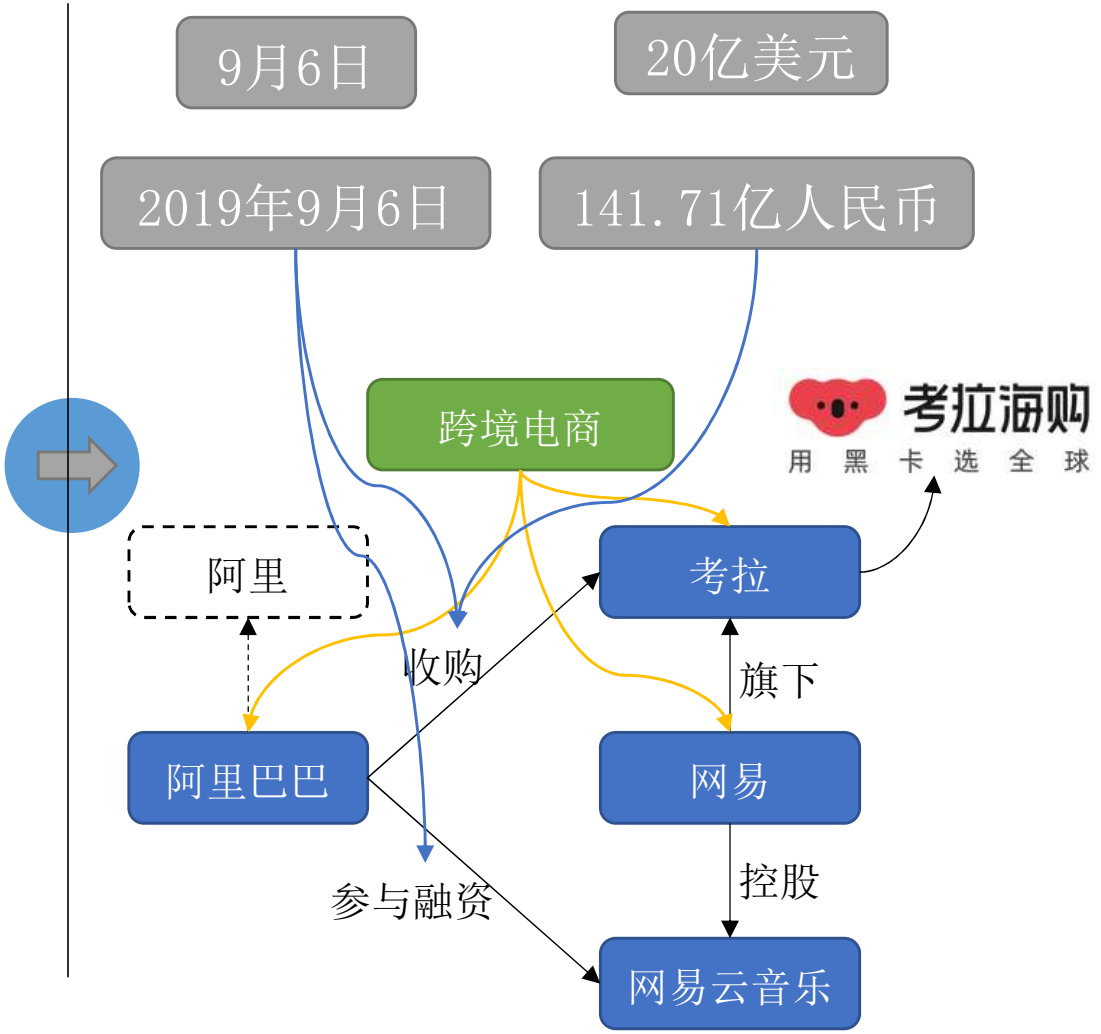
Why is IE hard on the web?

尘埃落定。9月6日，中国跨境电商行业最大的一起并购案浮出水面，阿里巴巴集团以20亿美元全资收购网易旗下跨境电商平台考拉。同时，阿里还作为领投方参与了网易云音乐B2轮7亿美元的融资。但网易仍会保持对网易云音乐的绝对控股。



Why is IE hard on the web?

尘埃落定。9月6日，中国跨境电商行业最大的一起并购案浮出水面，阿里巴巴集团以20亿美元全资收购网易旗下跨境电商平台考拉。同时，阿里还作为领投方参与了网易云音乐B2轮7亿美元的融资。但网易仍会保持对网易云音乐的绝对控股。



Named Entity Recognition (NER)

- A very important sub-task: find and classify names in text, for example:

蚂蚁集团实际控制人马云被证监会约谈

2020年11月02日 22:06 新浪网 作者 大众网

大众网·海报新闻记者 刘璐 北京报道

11月2日，证监会官方微信发布消息称，中国人民银行、中国银保监会、中国证监会、国家外汇管理局对蚂蚁集团实际控制人马云、董事长井贤栋、总裁胡晓明进行了监管约谈。

Named Entity Recognition (NER)

- A very important sub-task: find and classify names in text, for example:

蚂蚁集团实际控制人马云被证监会约谈

2020年11月02日 22:06 新浪网 作者 大众网

大众网·海报新闻记者 刘璐 北京报道

11月2日，证监会官方微信发布消息称，中国人民银行、中国银保监会、中国证监会、国家外汇管理局对蚂蚁集团实际控制人马云、董事长井贤栋、总裁胡晓明进行了监管约谈。

Named Entity Recognition (NER)

- A very important sub-task: find and classify names in text, for example:

蚂蚁集团实际控制人马云被证监会约谈

2020年11月02日 22:06 新浪网 作者 大众网

大众网·海报新闻记者 刘璐 北京报道

11月2日，证监会官方微信发布消息称，中国人民银行、中国银保监会、中国证监会、国家外汇管理局对蚂蚁集团实际控制人马云、董事长井贤栋、总裁胡晓明进行了监管约谈。

Person **Date** **Location** **Organization**

Named Entity Recognition (NER)

- A very important sub-task: find and classify names in text, for example:

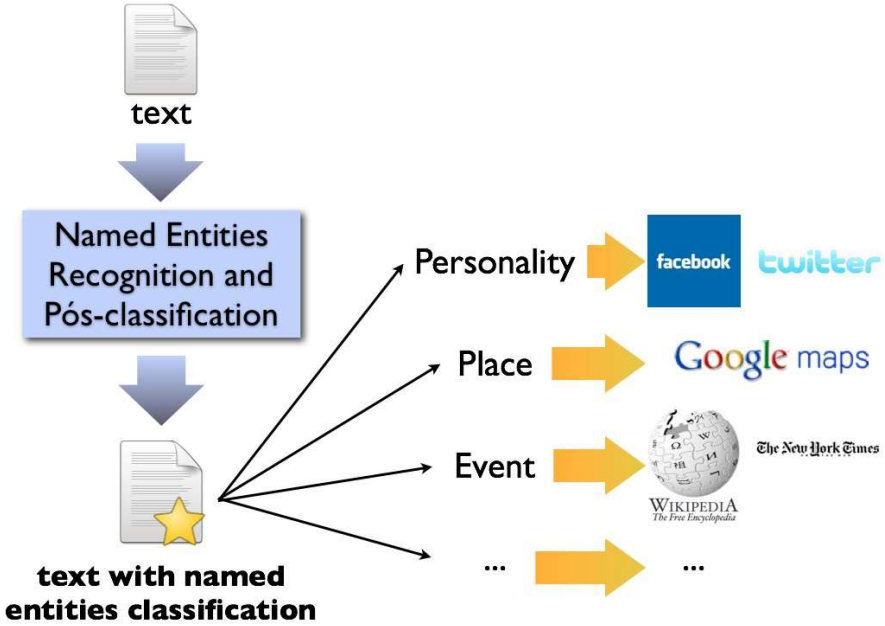
  Citing high fuel prices, **[ORG United Airlines]** said **[TIME Friday]** it has increased fares by **[MONEY $6]** per round trip on flights to some cities also served by lower-cost carriers**. [ORG American Airlines],** a unit of **[ORG AMR Corp.],** immediately matched the move, spokesman **[PER Tim Wagner]** said. [ORG United], a unit of **[ORG UAL Corp.]**, said the increase took effect **[TIME Thursday]** and applies to most routes where it competes against discount carriers, such as **[LOC Chicago]** to **[LOC Dallas]** and **[LOC Denver]** to **[LOC San Francisco].**

Typical Generic Named Entity Types

Type	Tag	Sample Categories	Example sentences
People	PER	people, characters	**Turing** is a giant of computer science.
Organization	ORG	companies, sports teams, government apartment	The **IPCC** warned about the cyclone.
Location	LOC	regions, mountains, seas	The **Mt. Sanitas** loop is in **Sunshine Canyon.**
Geo-Political Entity	GPE	countries, states, provinces	**Palo alto** is raising the fees for parking.
Facility	FAC	bridges, buildings, airports	Consider the **Golden Gate Bridge.**
Vehicles	VEH	planes, trains, automobiles	It was a classic **Ford Falcon.**

Named Entity Recognition (NER)

- The uses:
  - Named entities can be indexed, linked off, etc.
  - Sentiment can be attributed to companies or products
  - A lot of IE relations are associations between named entities
  - For question answering, answers are often named entities.



NER as Sequence Labeling

- The standard algorithm for NER is as a word-by-word sequence labeling task, in which the assigned tags capture both the boundary and the type.

- Consider the following simplified excerpt from our running example.

[ORG American Airlines], a unit of **[ORG AMR Corp.]**, immediately matched the move, spokesman **[PER Tim Wagner]** said.

NER as Sequence Labeling

- Typical labeling encodings:
  - BIO: Beginning(B), Inside(I), Outside(O)
  - IO: Inside(I) and Outside(O)
  - BMES/BIOES: Beginning(B), Middle(M), Ending(E), Single(S)
- Without B tag, IO tagging is unable to distinguish between entities of the same type that are right next to each other.
- With E tag, BMES tagging is able to identify the end of the entity more accurately.
- But the BIO tagging may be sufficient in many cases.

NER as Sequence Labeling

[ORG American Airlines], a unit of **[ORG AMR Corp.]**, immediately matched the move, spokesman **[PER Tim Wagner]** said.

Words	BIO Label	IO Label	BMES Label
American	B-ORG	I-ORG	B-ORG
Airlines	I-ORG	I-ORG	E-ORG
,	O	O	O
a	O	O	O
unit	O	O	O
of	O	O	O
AMR	B-ORG	I-ORG	B-ORG
Corp.	I-ORG	I-ORG	E-ORG
,	O	O	O

Words	BIO Label	IO Label	BMES Label
immediately	O	O	O
matched	O	O	O
the	O	O	O
move	O	O	O
,	O	O	O
spokesman	O	O	O
Tim	B-PER	I-PER	B-PER
Wagner	I-PER	I-PER	E-PER
said	O	O	O

NER as Sequence Labeling

- Metrics (Exact-match Evaluation)
  - A named entity is considered correctly recognized only if its both boundaries and type match ground truth.
  - Precision, Recall, and F-score are computed on the number of true positives(TP), false positives(FP), and false negatives(FN).
  - True Positive (TP): entities that are recognized by NER and match ground truth.
  - False Positive (FP): entities that are recognized by NER but do not match ground truth.
  - False Negative (FN): entities annotated in the ground truth that are not recognized by NER.

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \quad F - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Rule-based Approaches

- Rely on **hand-crafted rules** which can be designed based on **domain-specific gazetteers**(地名词典), and **syntactic-lexical patterns**.

- Works well when lexicon is exhaustive, **high precision and low recall**. (due to domain-specific rules and incomplete dictionaries)

- Well-known rule-based NER systems: **LaSIE-II, NetOwl, Facile, SAR, FASTUS, and LTG systems.**

Feature-Based Approaches

- Training
  - Collect a set of representative training documents
  - Label each token for its entity class or other (O)
  - Design feature extractors appropriate to the text and classes
  - Train a sequence classifier(MEMM/CRF) to predict the labels from the data

- Testing
  - Receive a set of testing documents
  - Run sequence model inference to label each token
  - Appropriately output the recognized entities

Linguistic Features

Words	POS	Chunk	Short shape	Label
American	NNP	B-NP	Xx	B-ORG
Airlines	NNPS	I-NP	Xx	I-ORG
,	,	O	,	O
a	DT	B-NP	x	O
unit	NN	I-NP	x	O
of	IN	B-PP	x	O
AMR	NNP	B-NP	X	B-ORG
Corp.	NNP	I-NP	Xx.	I-ORG
,	,	O	,	O

Words	POS	Chunk	Short shape	Label
immediately	RB	B-ADVP	x	O
matched	VBD	B-VP	x	O
the	DT	B-NP	x	O
move	NN	I-NP	X	O
,	,	,	,	O
spokesman	NN	B-NP	X	O
Tim	NNP	I-NP	Xx	B-PER
Wagner	NNP	I-NP	Xx	I-PER
said	VBD	B-VP	x	O

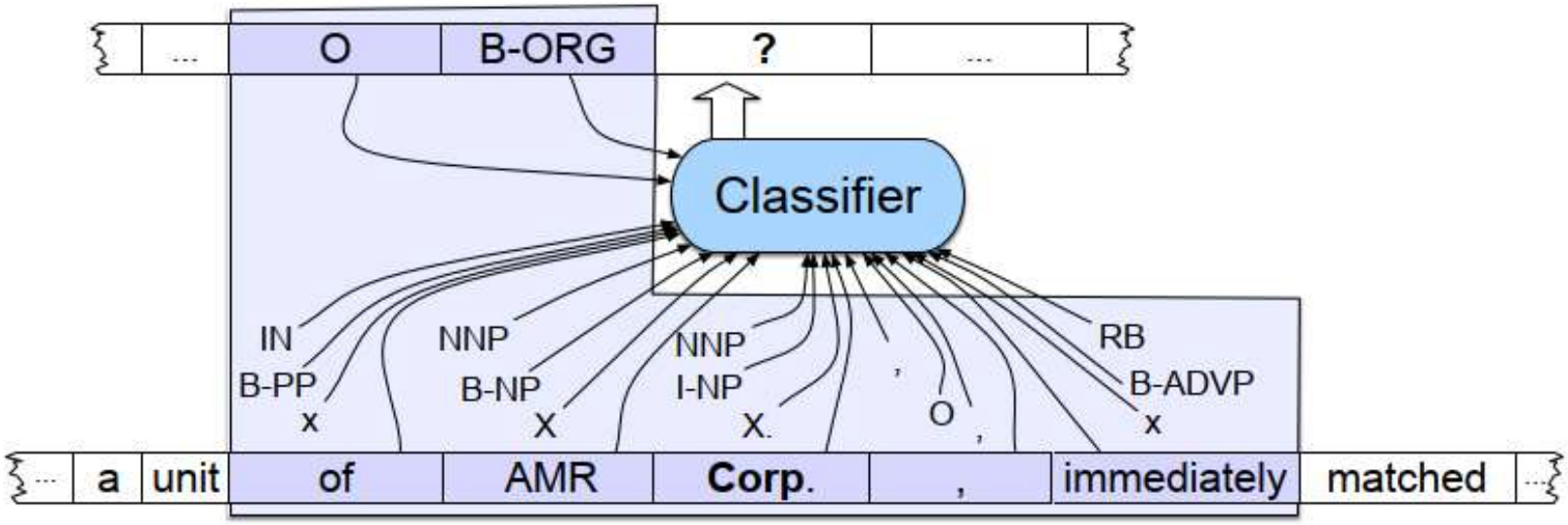
Features for Sequence Labeling

- Words
  - Current word (essentially like a learned dictionary)
  - Previous/next word (context)
- Other kinds of inferred linguistic features
  - POS tags
  - Chunks
  - Short shapes
- Label context
  - Previous (and perhaps next) label

Architecture

- A sequence classifier like an MEMM can be trained to label new sentences.

- If we assume a context window that includes the two preceding and following words, then the features available to the classifier are those shown in the boxed area.



Feature-based NER Approach

- For a Conditional Markov Model (CMM) a.k.a. a Maximum Entropy Markov Model (MEMM), the classifier makes a single decision at a time, conditioned on evidence from observations and previous decisions.

- A larger space of sequences is usually explored via search.

Local Context

Decision Point

pos	-2	-1	0	1	2
Word	**of**	**AMR**	**Corp.**	**,**	**immediately**
Tag	O	B-ORG	???	???	???
POS	IN	NNP	NNP	,	RB
Chunk	B-PP	B-NP	I-NP	O	B-ADVP
Short shape	x	x	X.	,	x

Features

w_0	**Corps.**
w_{-1}	AMR
w_{+1}	,
w_{-2}	of
w_{+2}	immediately
P_0	NNP
Short	X.
…	

Feature-based NER Approach

- Scoring individual labeling decisions is no more complex than standard classification decisions
  - We have some assumed labels to use for prior positions
  - We use features of those and the observed data (which can include current, previous, and next words) to predict the current label

Local Context

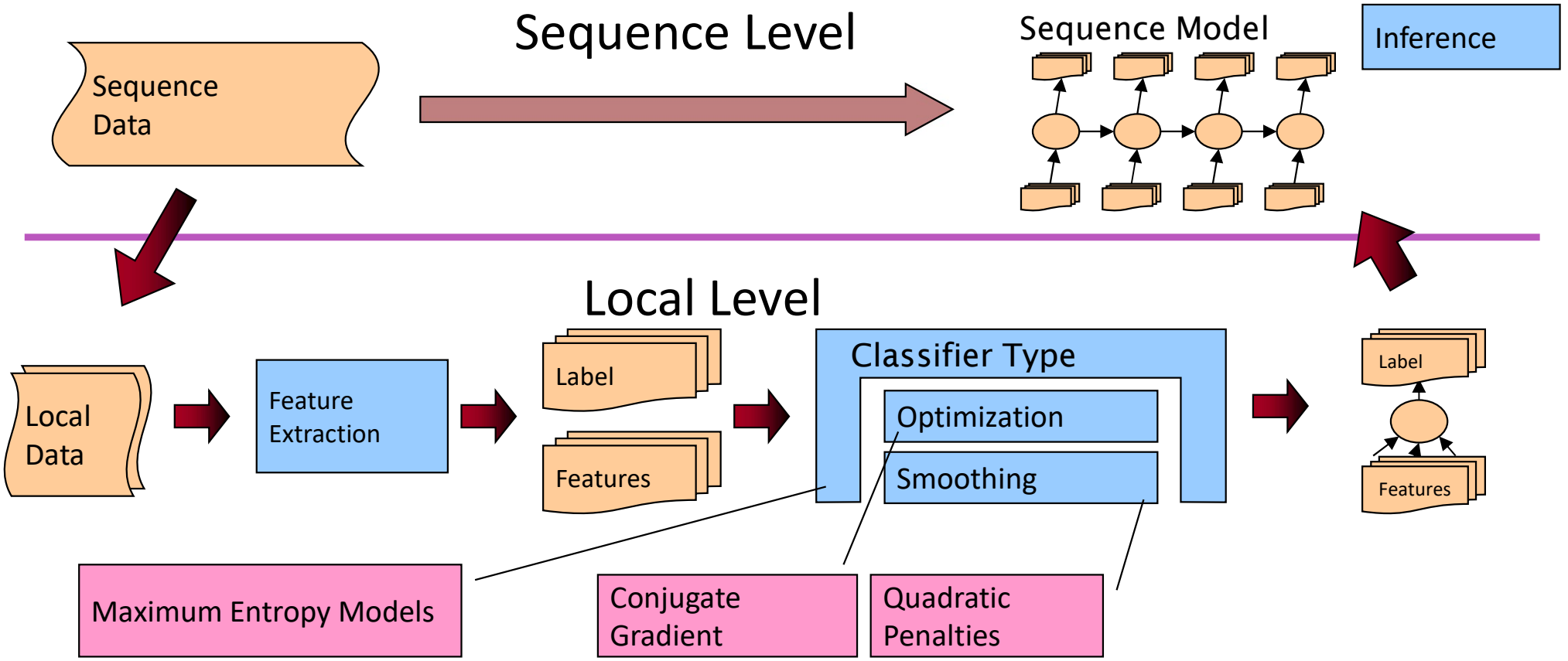
Decision Point

pos	-2	-1	0	1	2
Word	**of**	**AMR**	**Corp.**	**,**	**immediately**
Tag	O	B-ORG	???	???	???
POS	IN	NNP	NNP	,	RB
Chunk	B-PP	B-NP	I-NP	O	B-ADVP
Short shape	x	x	X.	,	x

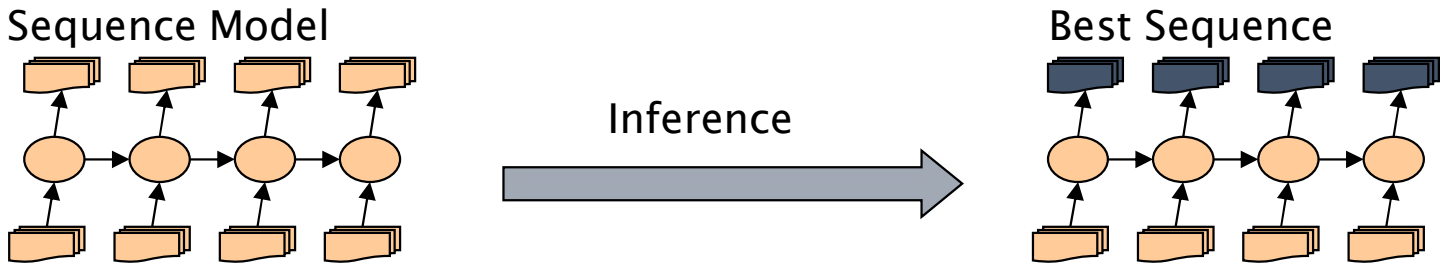
Features

w_0	**Corps.**
w_{-1}	AMR
w_{+1}	,
w_{-2}	of
w_{+2}	immediately
P_0	NNP
Short	X.
…	

Inference in Systems

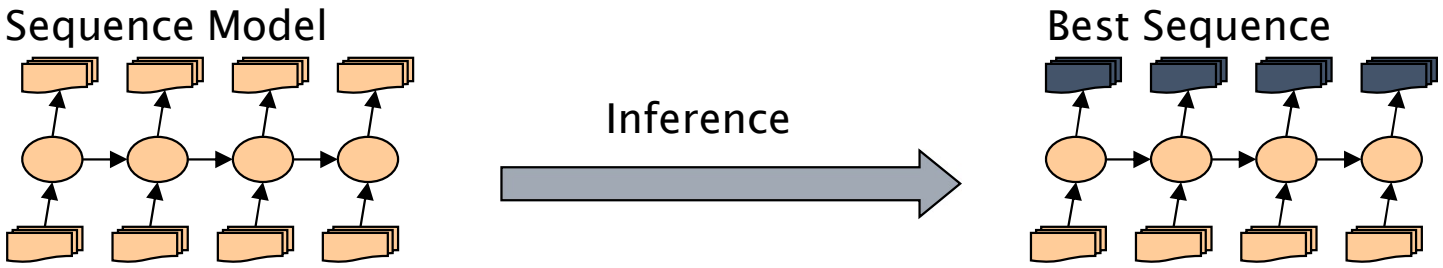


Greedy Inference



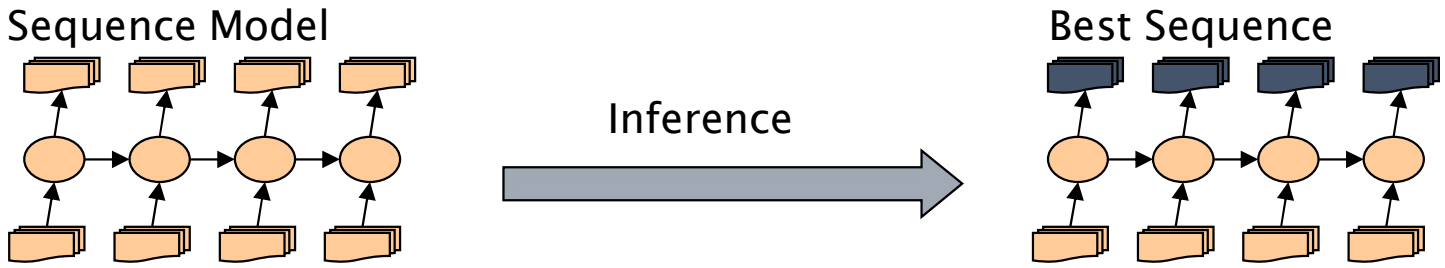
- Greedy inference:
  - We just start at the left, and use our classifier at each position to assign a label
  - The classifier can depend on previous labeling decisions as well as observed data
- Advantages:
  - Fast, no extra memory requirements
  - Very easy to implement
  - With rich features including observations to the right, it may perform quite well
- Disadvantage:
  - Greedy. We make commit errors we cannot recover from

Beam Inference



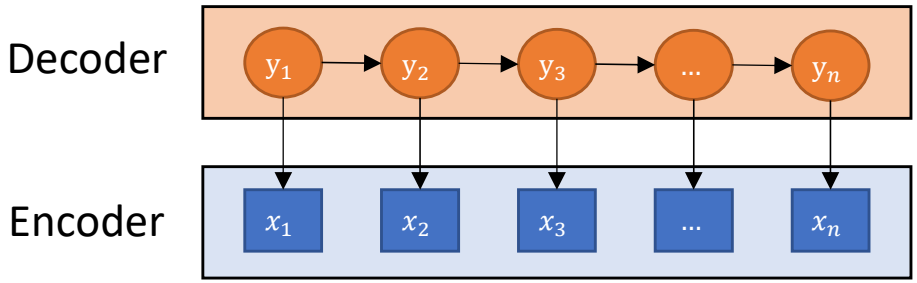
- Beam inference:
  - At each position keep the top k complete sequences.
  - Extend each sequence in each local way.
  - The extensions compete for the k slots at the next position.
- Advantages:
  - Fast; beam sizes of 3–5 are almost as good as exact inference in many cases.
  - Easy to implement (no dynamic programming required).
- Disadvantage:
  - Inexact: the globally best sequence can fall off the beam.

Viterbi Inference



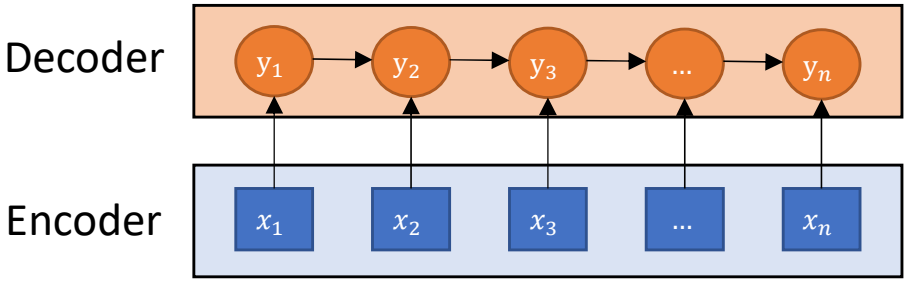
- Viterbi inference:
  - Dynamic programming or memoization.
  - Requires small window of state influence (e.g., past two states are relevant).
- Advantage:
  - Exact: the global best sequence is returned.
- Disadvantage:
  - Harder to implement long-distance state-state interactions (but beam inference tends not to allow long-distance resurrection of sequences anyway).

HMM vs. MEMM vs. CRFs



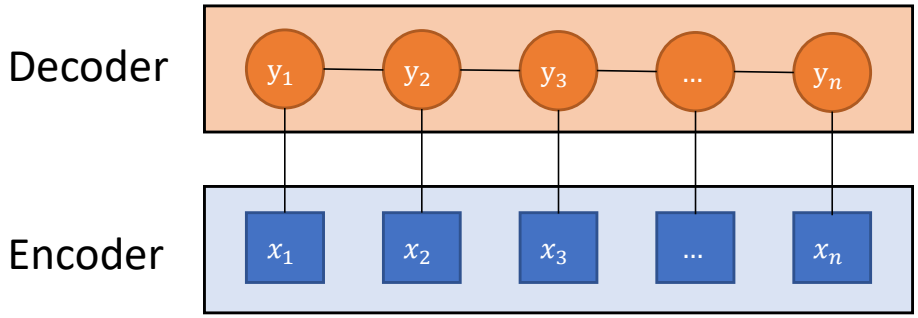
Hidden Markov Model(HMM)

- Generative directed probabilistic graphical model.
- Only dependent on every state and its corresponding observed object



Maximum Entropy Markov Model(MEMM)

- Discriminative directed probabilistic graphical model.
- Takes into account the dependencies between neighboring states and the entire observed sequence.

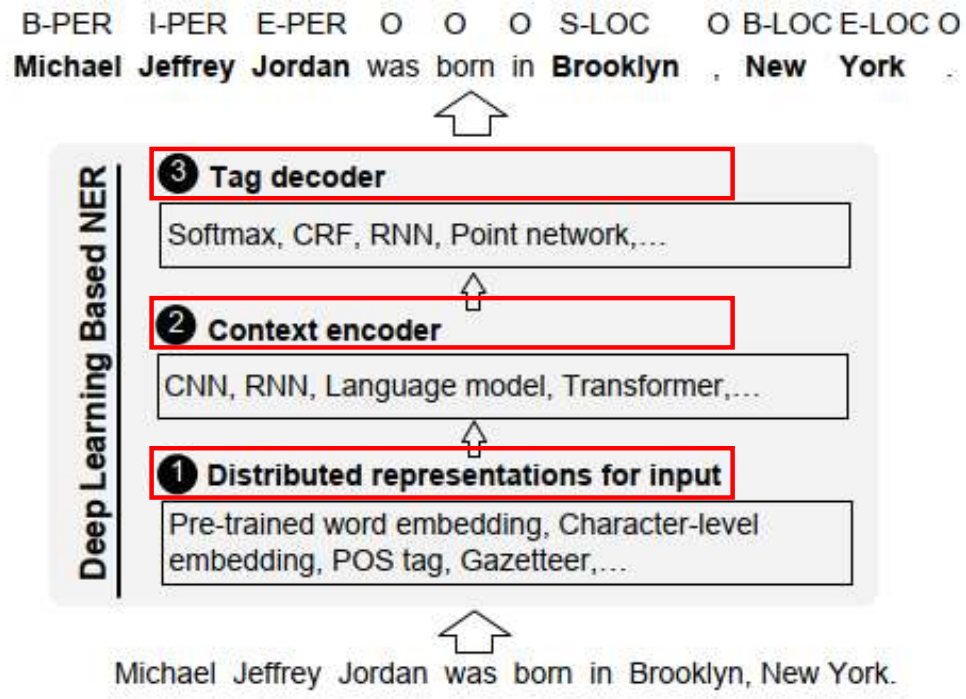


Conditional Random Field(CRF)

- Discriminative undirected probabilistic graphical model
- Takes into account the dependencies between the predictions

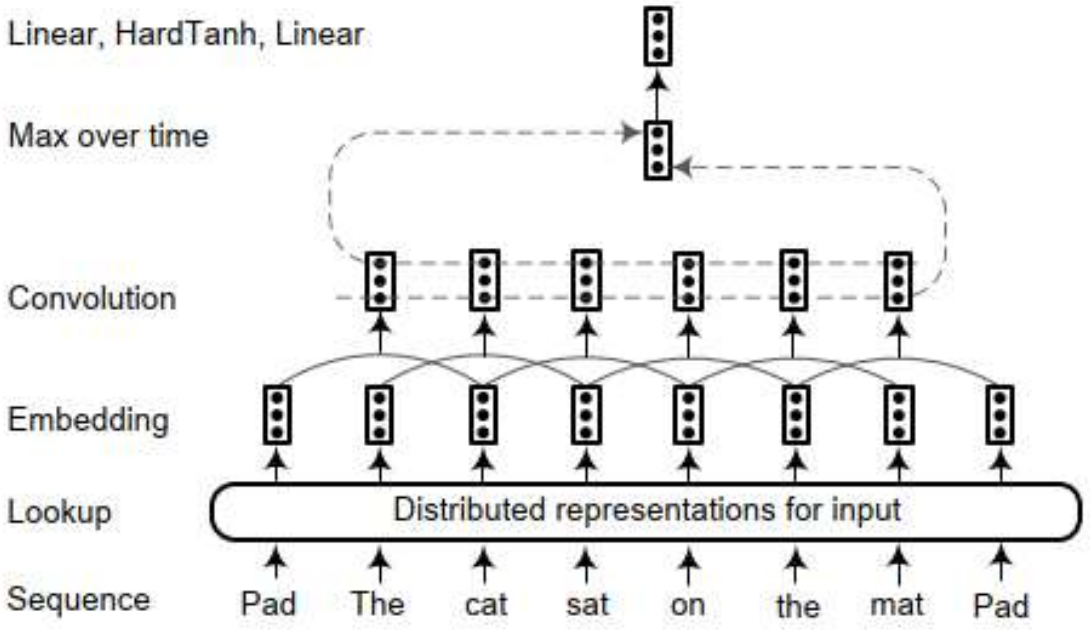
Deep Learning Techniques for NER

- DL-based models are
  - Able to learn complex and intricate features from data via non-linear activation functions
  - Effective in automatically learning useful representations and underlying factors from raw data.
  - Able to be trained in an end-to-end paradigm by gradient descent.



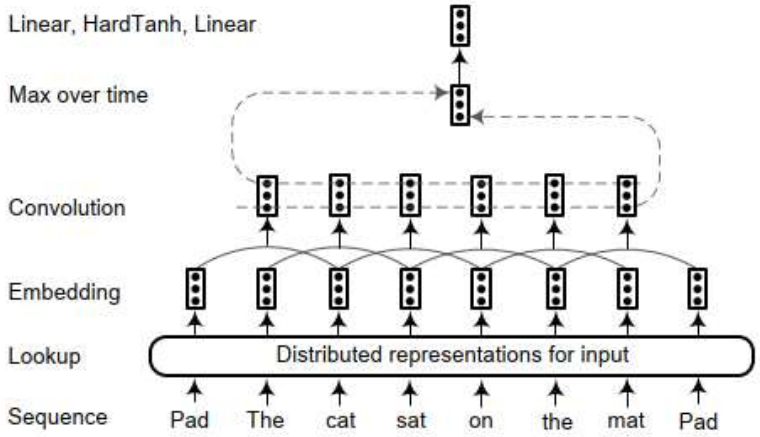
Distributed Representations for Input

- Word-level Representation
- Character-level Representation
- Hybrid Representation

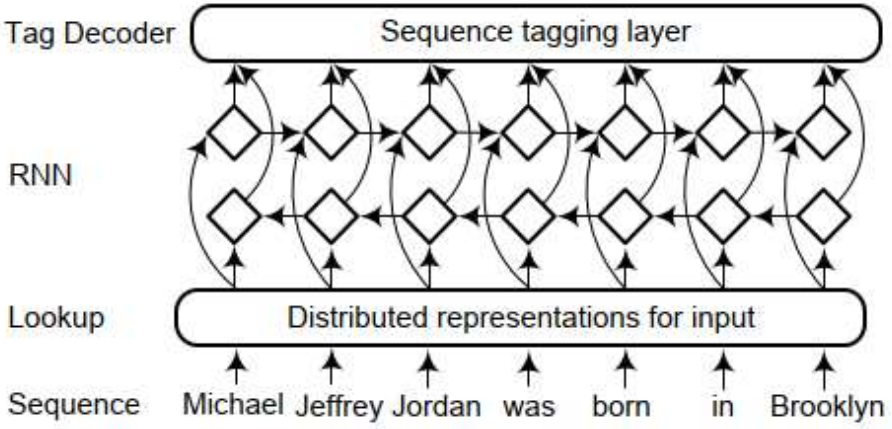


R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, “Natural language processing (almost) from scratch,” J. Mach. Learn. Res., vol. 12, no. Aug, pp. 2493–2537, 2011.

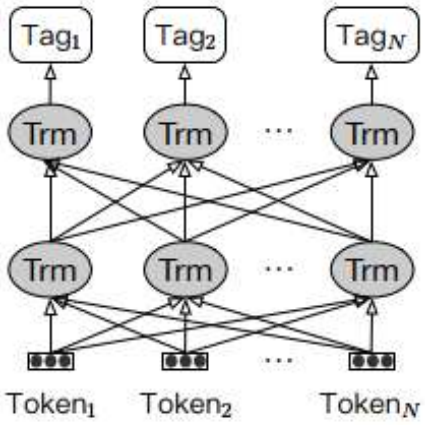
Context Encoder Architectures



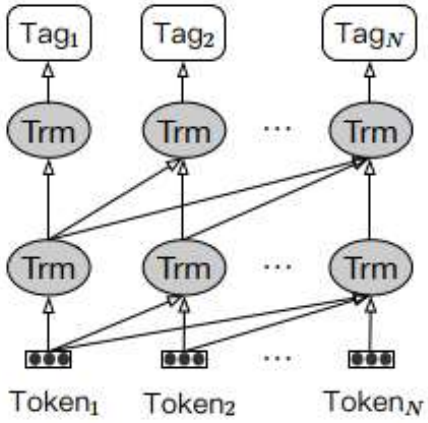
CNNs



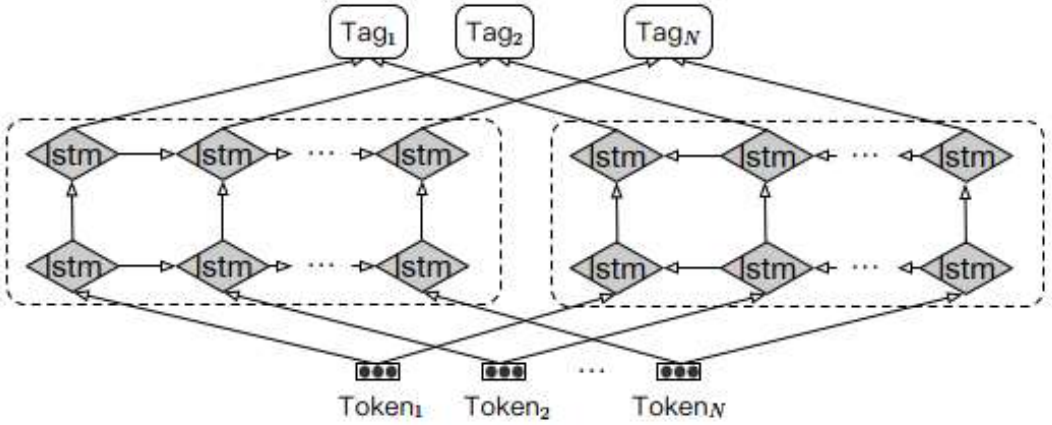
RNNs



(a) Google BERT



(b) OpenAI GPT

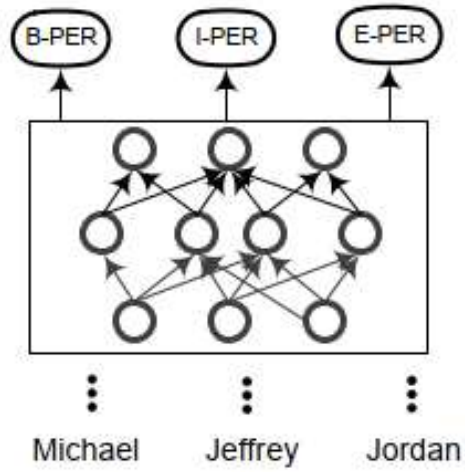


(c) AllenNLP ELMo

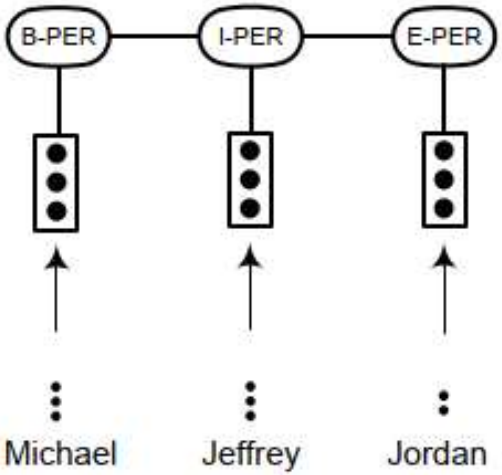
LMs

Tag Decoder Architectures

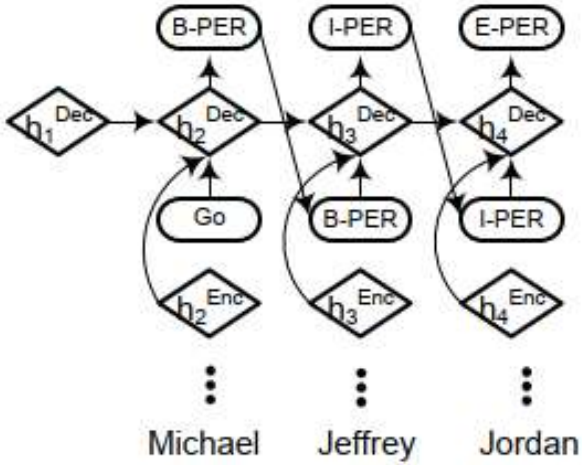
Seq2Seq



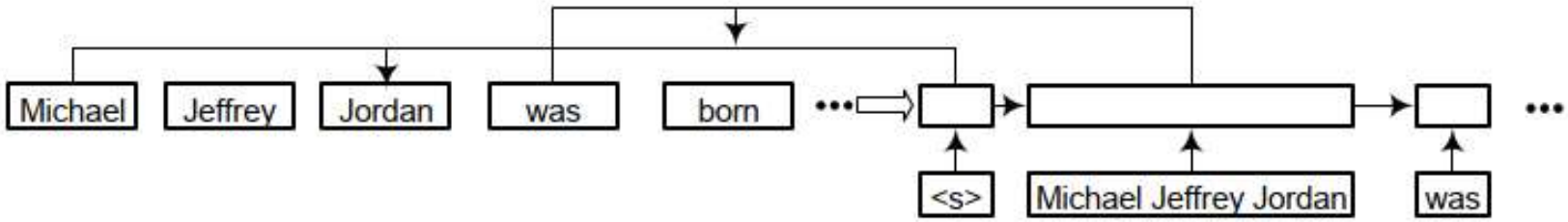
(a) MLP+Softmax



(b) CRF

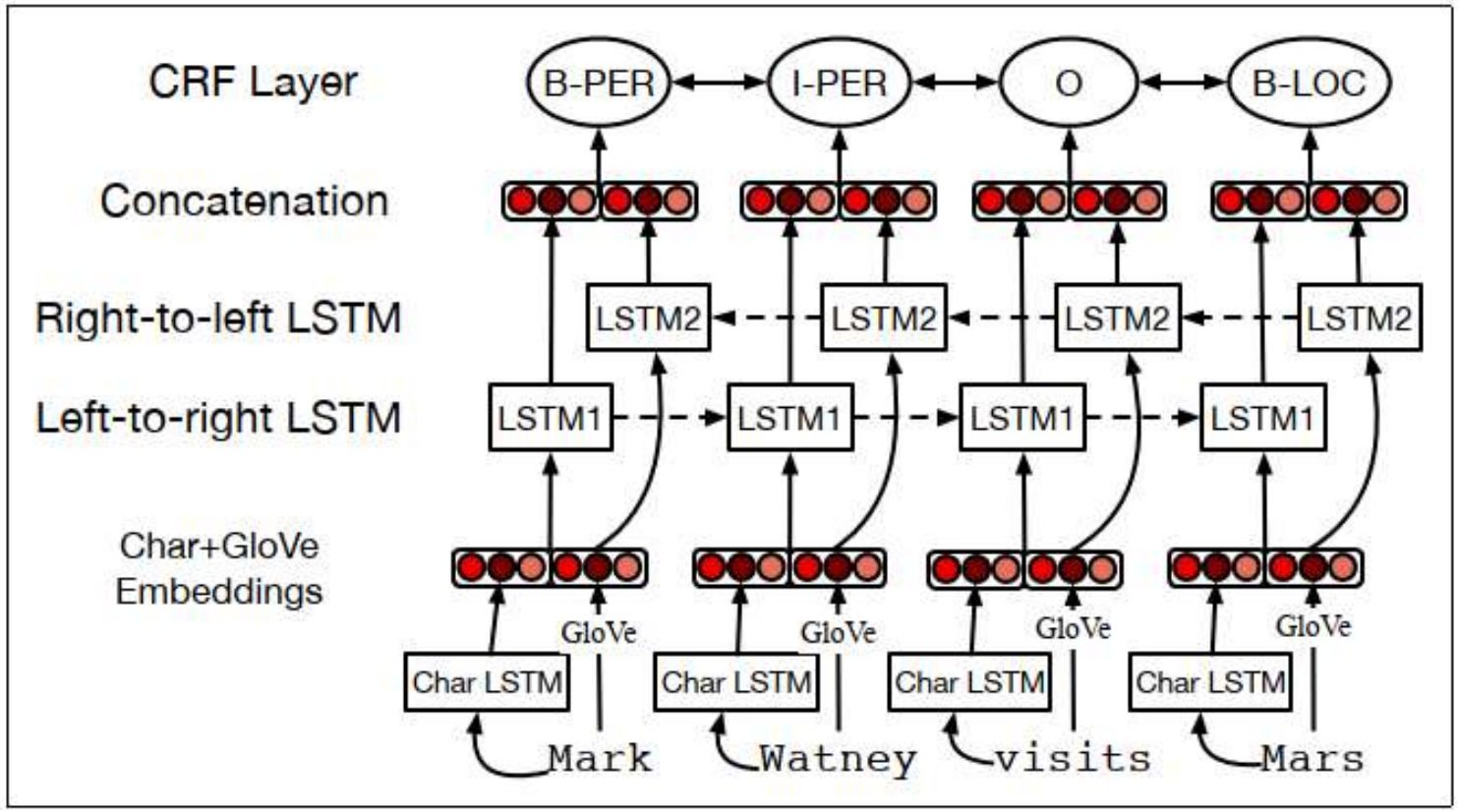


(c) RNN



(d) Pointer Network

The Most Common Architecture for NER



A Bi-LSTM + CRF sequence model. After Lample et al. (2016)

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. In NAACL HLT 2016.

The Most Common Architecture for NER

- The standard DL model for NER is based on the Bi-LSTM and CRF. (which is the most common architecture for NER)
  - Word and character embeddings are computed for input word w_i.
  - These are passed through a left-to-right LSTM and a right-to-left LSTM.
  - Two group of outputs are concatenated (or otherwise combined) to produce a single output layer at position i.
  - In the simplest method, this layer can then be directly passed onto a softmax that creates a probability distribution over all NER tag, and the most likely tag is chosen as t_i.
  - For NER task, a CRF layer is normally used on top of the Bi-LSTM output, and the Viterbi decoding algorithm is used to decode.

Benchmarks

TREND	DATASET	BEST METHOD	PAPER TITLE	PAPER	CODE	COMPARE
	CoNLL 2003 (English)	LUKE	LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention			See all
	Ontonotes v5 (English)	BERT-MRC+DSC	Dice Loss for Data-imbalanced NLP Tasks			See all
	ACE 2005	BERT-MRC	A Unified MRC Framework for Named Entity Recognition			See all
	GENIA	Biaffine-NER	Named Entity Recognition as Dependency Parsing			See all
	ACE 2004	Biaffine-NER	Named Entity Recognition as Dependency Parsing			See all
	CoNLL++	CrossWeigh + Pooled Flair	CrossWeigh: Training Named Entity Tagger from Imperfect Annotations			See all
	Long-tail emerging entities	Flair embeddings	Contextual String Embeddings for Sequence Labeling			See all
	BC5CDR	NER+PA+RL (PubMed)	Reinforcement-based denoising of distantly supervised NER with partial annotation			See all
	JNLPBA	BLSTM-CNN-Char (SparkNLP)	Biomedical Named Entity Recognition at Scale			See all
	NCBI-disease	BioBERT	BioBERT: a pre-trained biomedical language representation model for biomedical text mining			See all
	SciERC	SpERT	Span-based Joint Entity and Relation Extraction with Transformer Pre-training			See all
	CoNLL 2003 (German)	ACE + document-context	Automated Concatenation of Embeddings for Structured Prediction			See all

Benchmarks

	CoNLL 2002 (Spanish)	🏆 ACE + document-context	Automated Concatenation of Embeddings for Structured Prediction			See all
	CoNLL 2002 (Dutch)	🏆 ACE + document-context	Automated Concatenation of Embeddings for Structured Prediction			See all
	CoNLL 2003 (German) Revised	🏆 ACE	Automated Concatenation of Embeddings for Structured Prediction			See all
	WLPC	🏆 DyGIE	A General Framework for Information Extraction using Dynamic Span Graphs			See all
	WetLab	🏆 BiLSTM-CRF with ELMo	Using Similarity Measures to Select Pretraining Data for NER			See all
	Species-800	🏆 BioFLAIR	BioFLAIR: Pretrained Pooled Contextualized Embeddings for Biomedical Sequence Labeling Tasks			See all
	LINNAEUS	🏆 BLSTM-CNN-Char (SparkNLP)	Biomedical Named Entity Recognition at Scale			See all
	Code-Switching English-Spanish NER	🏆 HME (word + BPE + char)	Hierarchical Meta-Embeddings for Code-Switching Named Entity Recognition			See all
	ontontoes chinese v5	🏆 DGLSTM-CRF	Dependency-Guided LSTM-CRF for Named Entity Recognition			See all
	CoNLL 2000	🏆 SWEM-CRF	Baseline Needs More Love: On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms			See all
	French Treebank	🏆 CamemBERT (subword masking)	CamemBERT: a Tasty French Language Model			See all
	SoSciSoCi	🏆 Bi-LSTM-CRF (SSC->GSC)	Investigating Software Usage in the Social Sciences: A Knowledge Graph Approach			See all
	LeNER-Br	🏆 LSTM-CRF	LeNER-Br: a Dataset for Named Entity Recognition in Brazilian Legal Text			See all

Summary of NER Approaches

- In recent years, DL-based NER models become dominant and achieve state-of-the-art results.
- Transformer encoder is more effective than LSTM when Transformer is pre-trained on huge corpora. Transformers fail on NER task if they are not pre-trained and when the training data is limited.
- A major disadvantage of RNN and Pointer Network decoders lies in greedily decoding, which means that the input of current step needs the output of previous step. (CRF could be computationally expensive when the number of entity types is large)

Applied DL for NER

- Deep Multi-task Learning for NER
  - By considering the relation between different tasks, multi-task learning algorithms are expected to achieve better results than the ones that learn each task individually. (Y. Lin et al., ACL 2018)
- Deep Transfer Learning for NER
  - Recently, a few approaches have been proposed for low-resource and across-domain NER using deep neural networks. (G. Beryozkin et al., ACL 2019)
- Deep Active Learning for NER
  - Deep learning typically requires a large amount of training data which is costly to obtain. Thus, combining deep learning with active learning is expected to reduce data annotation effort. (Y. Shen et al., ICLR 2017)

Applied DL for NER

- Deep Reinforcement Learning for NER
  - RL is concerned with how software agents take actions in an environment so as to maximize some cumulative rewards. (Y. Yang et al., COLING 2018)
- Deep Adversarial Learning for NER
  - Some studies considered the instances in a source domain as adversarial examples for a target domain, and vice versa.(J. Li et al., IJCAI 2019; P. Cao et al., EMNLP 2018)
  - Another option is to prepare an adversarial sample by adding an original sample with a perturbation. (J.T. Zhou et al., ACL 2019)
- Neural Attention for NER
  - Neural attention mechanism allows neural networks have the ability to focus on a subset of its inputs. (Q. Zhang et al., AAAI 2018)

Datasets

Corpus	Year	Text Source	#Tags	URL
MUC-6	1995	Wall Street Journal	7	https://catalog.ldc.upenn.edu/LDC2003T13
MUC-6 Plus	1995	Additional news to MUC-6	7	https://catalog.ldc.upenn.edu/LDC96T10
MUC-7	1997	New York Times news	7	https://catalog.ldc.upenn.edu/LDC2001T02
CoNLL03	2003	Reuters news	4	https://www.clips.uantwerpen.be/conll2003/ner/
ACE	2000 - 2008	Transcripts, news	7	https://www.ldc.upenn.edu/collaborations/past-projects/ace
OntoNotes	2007 - 2012	Magazine, news, web, etc.	18	https://catalog.ldc.upenn.edu/LDC2013T19
W-NUT	2015 - 2018	User-generated text	6/10	http://noisy-text.github.io
BBN	2005	Wall Street Journal	64	https://catalog.ldc.upenn.edu/LDC2005T33
WikiGold	2009	Wikipedia	4	https://figshare.com/articles/Learning_multilingual_named_entity_recognition_from_Wikipedia/5462500
WiNER	2012	Wikipedia	4	http://rali.iro.umontreal.ca/rali/en/winer-wikipedia-for-ner
WikiFiger	2012	Wikipedia	112	https://github.com/xiaoling/figer
HYENA	2012	Wikipedia	505	https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/hyena/
N^3	2014	News	3	http://aksw.org/Projects/N3NERNEDNIF.html
Gillick	2016	Magazine, news, web, etc.	89	https://arxiv.org/e-print/1412.1820v2
FG-NER	2018	Various	200	https://fgner.alt.ai/
NNE	2019	Newswire	114	https://github.com/nickyringland/nested_named_entities
GENIA	2004	Biology and clinical text	36	http://www.geniaproject.org/home
GENETAG	2005	MEDLINE	2	https://sourceforge.net/projects/bioc/files/
FSU-PRGE	2010	PubMed and MEDLINE	5	https://julielab.de/Resources/FSU_PRGE.html
NCBI-Disease	2014	PubMed	1	https://www.ncbi.nlm.nih.gov/CBBresearch/Dogan/DISEASE/
BC5CDR	2015	PubMed	3	http://bioc.sourceforge.net/
DFKI	2018	Business news and social media	7	https://dfki-lt-re-group.bitbucket.io/product-corpus/

Tools

NER System	URL
StanfordCoreNLP	https://stanfordnlp.github.io/CoreNLP/
OSU Twitter NLP	https://github.com/aritter/twitter_nlp
Illinois NLP	http://cogcomp.org/page/software/
NeuroNER	http://neuroner.com/
NERsuite	http://nersuite.nlplab.org/
Polyglot	https://polyglot.readthedocs.io
Gimli	http://bioinformatics.ua.pt/gimli
spaCy	https://spacy.io/api/entityrecognizer
NLTK	https://www.nltk.org
OpenNLP	https://opennlp.apache.org/
LingPipe	http://alias-i.com/lingpipe-3.9.3/
AllenNLP	https://demo.allennlp.org/
IBM Watson	https://natural-language-understanding-demo.ng.bluemix.net/
FG-NER	https://fgner.alt.ai/extractor/
Intellexer	http://demo.intellexer.com/
Repustate	https://repustate.com/named-entity-recognition-api-demo/
AYLIEN	https://developer.aylien.com/text-api-demo
Dandelion API	https://dandelion.eu/semantic-text/entity-extraction-demo/
displaCy	https://explosion.ai/demos/displacy-ent
ParallelDots	https://www.paralleldots.com/named-entity-recognition
TextRazor	https://www.textrazor.com/named_entity_recognition

Entity Linking

- Is the task of linking mention to specific entity in **Knowledge Base**
- Also referred to as Named-Entity Linking (NEL), Named-Entity Disambiguation(NED), Named-Entity Normalization (NEN).

Example of Cross-Type Confusion

Name	Possible Categories
Washington	Person, Location, Political Entity, Organization, Vehicle
Downing St.	Location, Organization
IRA	Person, Organization, Monetary Instrument
Louis Vuitton	Person, Organization, Commercial Product

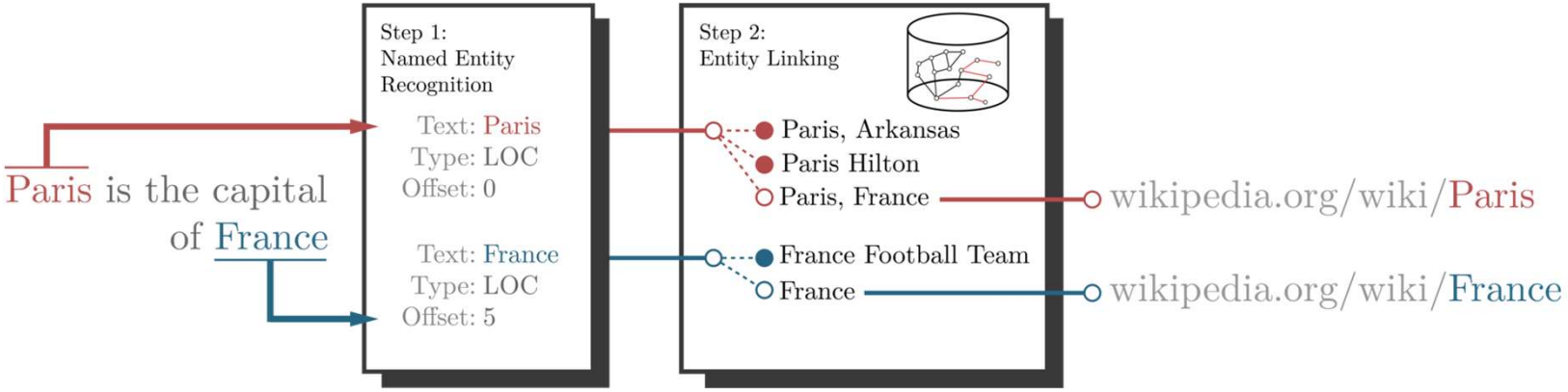
- **[PER Washington]** was born into slavery on the farm of James Burroughs.
- **[ORG Washington]** went up 2 games to 1 in the four-game series.
- Blair arrived in **[LOC Washington]** for what may well be his last state visit.
- In June, **[GPE Washington]** passed a primary seatbelt law.
- The **[VEH Washington]** had proved to be a leaky ship, every passage I made...

Challenges

- **Name variations:** the same entity might appear in textual representations. i.e. 腾讯–鹅厂、小米–猴厂、网易–猪厂
- **Ambiguity:** the same mention can often refer to many different entities. i.e.苹果售价是多少?
- **Absence:** some entities might not have a correct entity link in the target knowledge base. i.e. 信条 (a new movie)
- **Scalability and Speed:** It is desirable for an industrial entity linking system to provide results in a reasonable time, and often in real-time.
- **Evolving Information**
- **Multiple Languages**

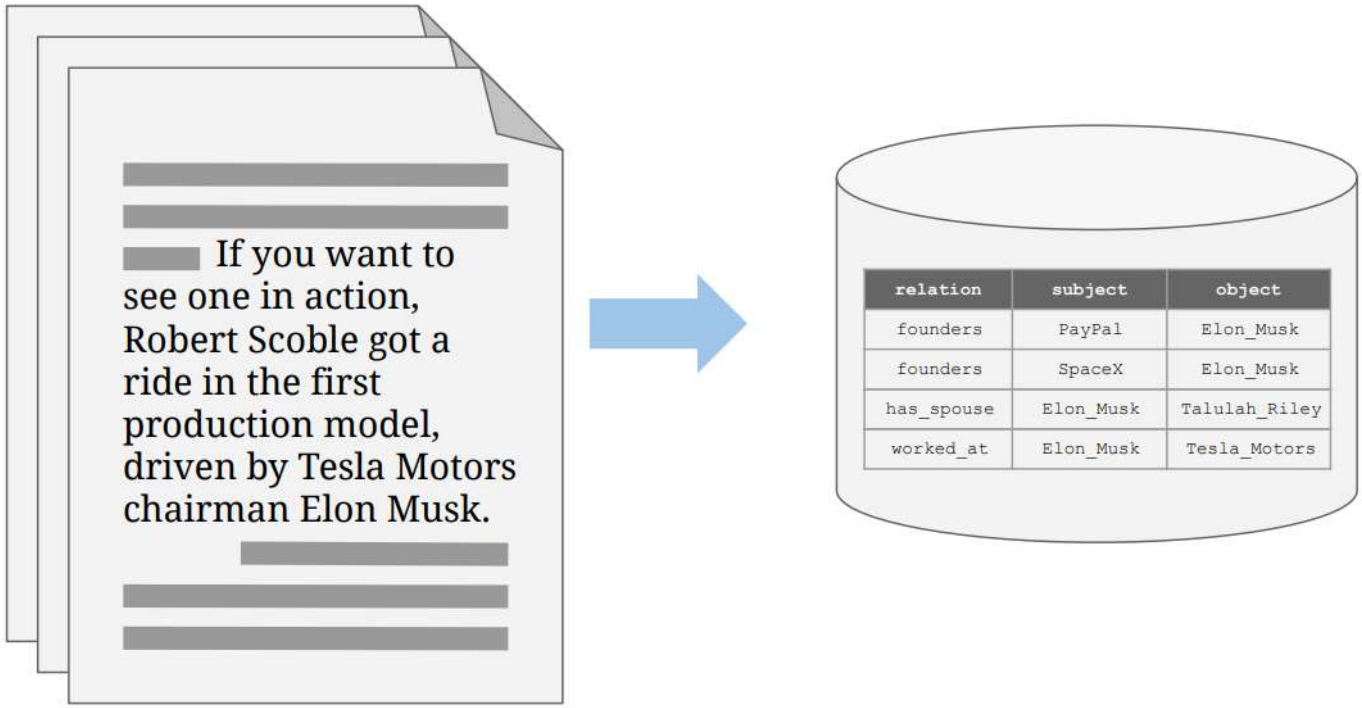
Approaches

- Text-based approaches
  - Make use of textual features extracted from large text corpora (e.g. TF-IDF, n-grams…)
- Graph-based approaches
  - Exploit the structure of knowledge graphs to represent the context and the relation of entities.



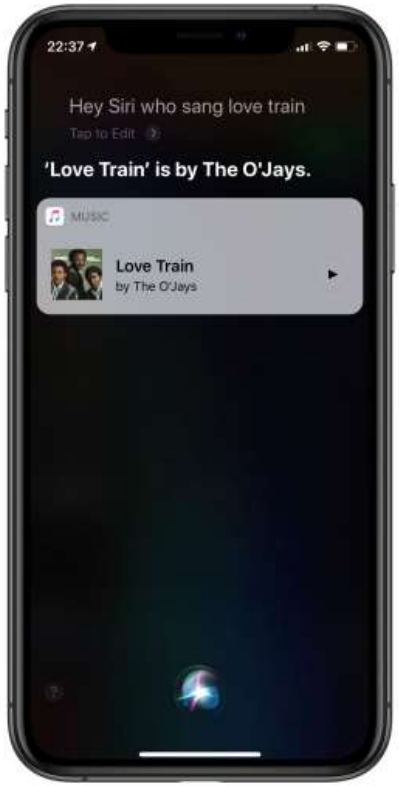
Relation Extraction

- Detection and classification of semantic relationship mentions within a set of entities.



Applications

- Intelligent assistants



/music/artist/track

The O'Jays	Love Train
Cardi B	Bodak Yellow
Selena Gomez	Bad Liar

/film/film/starring

Wonder Woman	Gal Gadot
Dunkirk	Tom Hardy
Tomb Raider	Alicia Vikander

/organization/organization/parent

tbh	Facebook
Kaggle	Google
LinkedIn	Microsoft

/people/person/date_of_death

Barbara Bush	2018-04-17
Milos Forman	2018-04-14
Winnie Mandela	2018-04-11

"Love Train" is a hit single by The O'Jays, written by Kenny Gamble and Leon Huff. Released in 1972, it reached number one on both the R&B Singles and the Billboard Hot 100, in February and March 1973 respectively, number 9 on the UK Singles Chart and was certified gold by the RIAA. It was The O'Jays' first and only number-one record on the US pop chart.

Applications

- Building ontologies

```
video game
  action game
    ball and paddle game
      Breakout
    platform game
      Donkey Kong
    shooter
      arcade shooter
        Space Invaders
      first-person shooter
        Call of Duty
      third-person shooter
        Tomb Raider
  adventure game
    text adventure
    graphic adventure
  strategy game
    4X game
      Civilization
    tower defense
      Plants vs. Zombies
```

Mirror ran a headline questioning whether the killer's actions were a result of playing Call of Duty, a first-person shooter game ...

Melee, in video game terms, is a style of elbow-drop hand-to-hand combat popular in first-person shooters and other shooters.

Tower defense is a kind of real-time strategy game in which the goal is to protect an area or place and prevent enemies from reaching ...

Applications

- Gene regulation

textual abstract:
summary for human

relation	subject	object
is_a	p53	protein
is_a	Bax	protein
has_function	p53	apoptosis
has_function	Bax	induction
involved_in	apoptosis	cell_death
is_in	Bax	cytoplasm
related_to	apoptosis	caspase_activation
...	...	...

structured knowledge extraction:
summary for machine

Relation Types

- Relation types from ACE 2003
  - ROLE: relates a person to an organization or a geopolitical entity
    - Subtypes: member, owner, affiliate, client, citizen
  - PART: generalized containment
    - subtypes: subsidiary, physical part-of, set membership
  - AT: permanent and transient locations
    - subtypes: located, based-in, residence
  - SOCIAL: social relations among persons
    - subtypes: parent, sibling, spouse, grandparent, associate

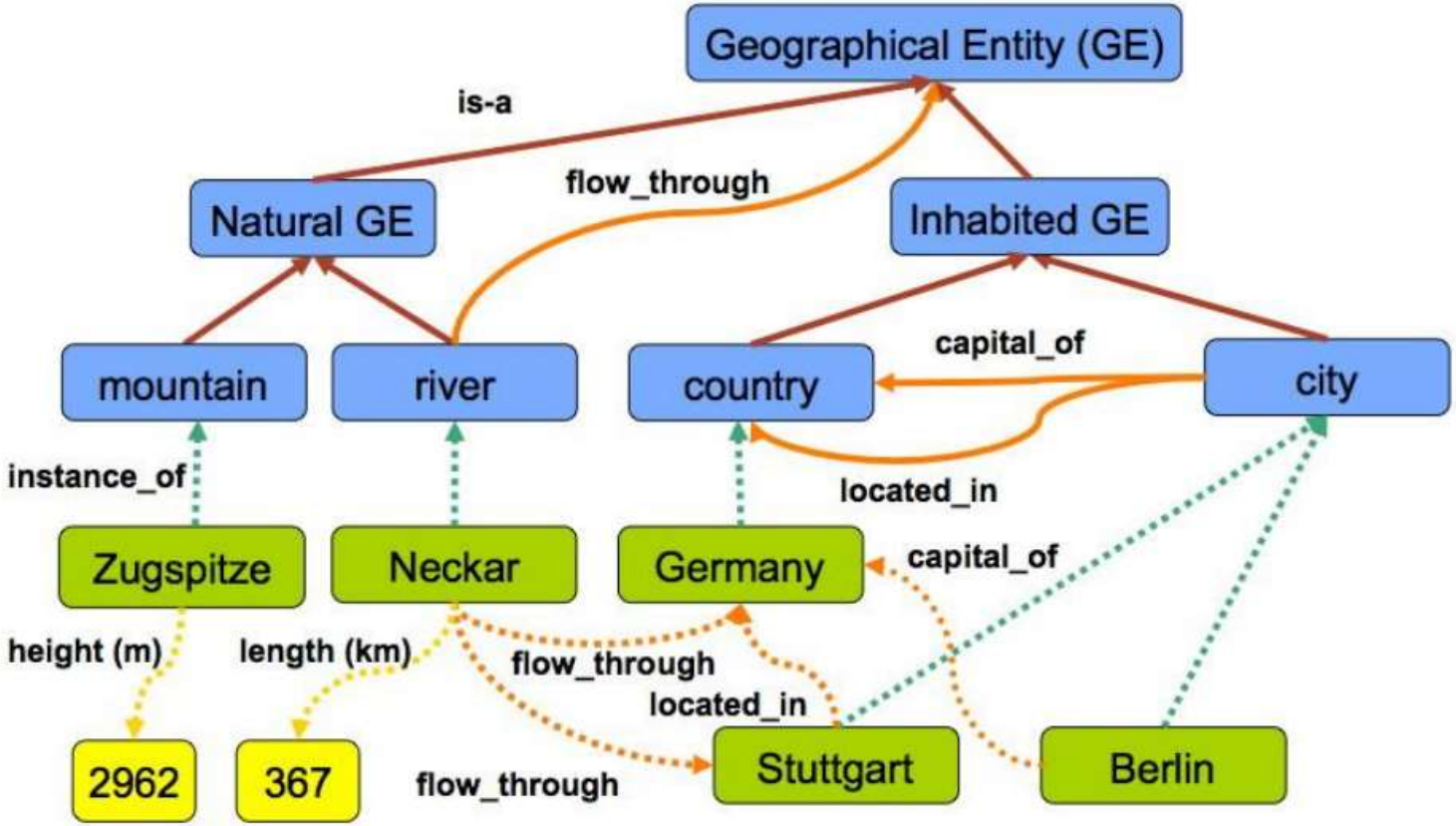
Relation Types

- Freebase – 23 million entities, thousands of relations

Relation name	Size	Example
/people/person/nationality	281,107	John Dugard, South Africa
/location/location/contains	253,223	Belgium, Nijlen
/people/person/profession	208,888	Dusa McDuff, Mathematician
/people/person/place_of_birth	105,799	Edwin Hubble, Marshfield
/dining/restaurant/cuisine	86,213	MacAyo's Mexican Kitchen, Mexican
/business/business_chain/location	66,529	Apple Inc., Apple Inc., South Park, NC
/biology/organism_classification_rank	42,806	Scorpaeniformes, Order
/film/film/genre	40,658	Where the Sidewalk Ends, Film noir
/film/film/language	31,103	Enter the Phoenix, Cantonese
/biology/organism_higher_classification	30,052	Calopteryx, Calopterygidae
/film/film/country	27,217	Turtle Diary, United States
/film/writer/film	23,856	Irving Shulman, Rebel Without a Cause
/film/director/film	23,539	Michael Mann, Collateral
/film/producer/film	22,079	Diane Eskenazi, Aladdin
/people/deceased_person/place_of_death	18,814	John W. Kern, Asheville
/music/artist/origin	18,619	The Octopus Project, Austin
/people/person/religion	17,582	Joseph Chartrand, Catholicism
/book/author/works_written	17,278	Paul Auster, Travels in the Scriptorium
/soccer/football_position/players	17,244	Midfielder, Chen Tao
/people/deceased_person/cause_of_death	16,709	Richard Daintree, Tuberculosis
/book/book/genre	16,431	Pony Soldiers, Science fiction
/film/film/music	14,070	Stavisky, Stephen Sondheim
/business/company/industry	13,805	ATS Medical, Health care

Relation Types

- Geographical



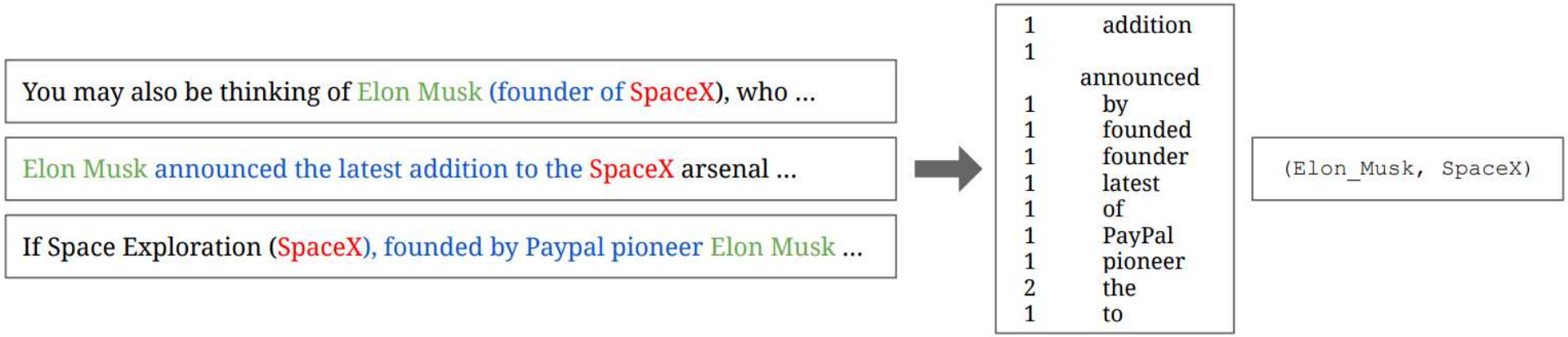
Problem Formulation

- Inputs and outputs
  - What is the input to the prediction?
    - A pair of entity mentions in the context of a sentence?
    - A pair of entities, independent of any specific context?
  - What is the output to the prediction?
    - A single relation (multi-class classification)?
    - Or multiple relations (multi-label classification)?

Problem Formulation

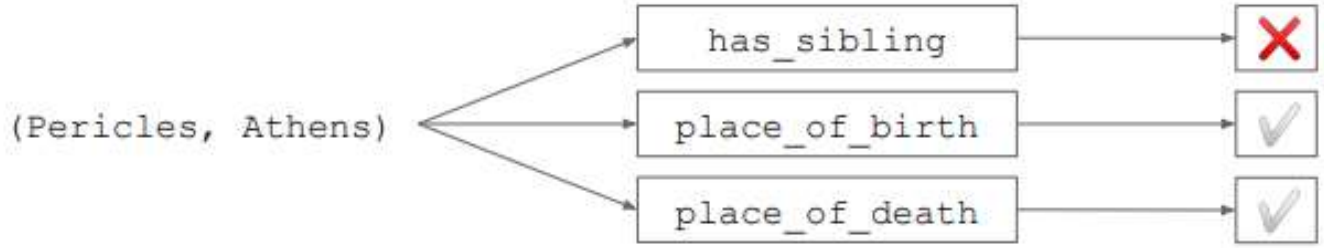
- Joining the corpus and the KB
  - Classifying a pair of entity mentions in Corpus? Get labels from KB.

  - Classifying a pair of entities for the KB? Get features from corpus.



Problem Formulation

- Multi-label classification
  - Many possible approaches to multi-label classification.
  - The most obvious is the binary relevance method: just train a separate binary classifier for each label.



  - Advantage: simple
  - Disadvantage: fails to exploit correlations between labels

Problem Formulation

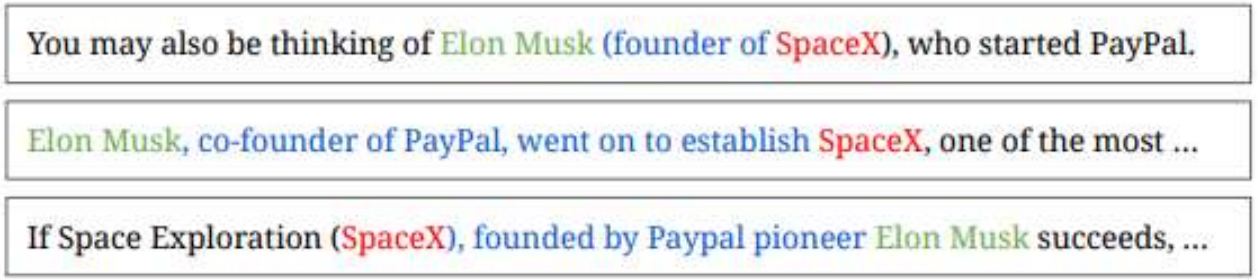
- Binary classification of KB triples
  - Input: an entity pair and a candidate relation
  - Output: does the entity pair belong to the relation?
  - That is, given a candidate KB triple

(worked_at, Elon_Musk, SpaceX) ?

Approaches

- Hand-built patterns
  - Idea: define some extraction patterns

  - Problem: most occurrences do not fit simple patterns



Approaches

- Supervised Learning
  - Idea: label examples, train a classifier

  - Problem: labeling examples is expensive

Approaches

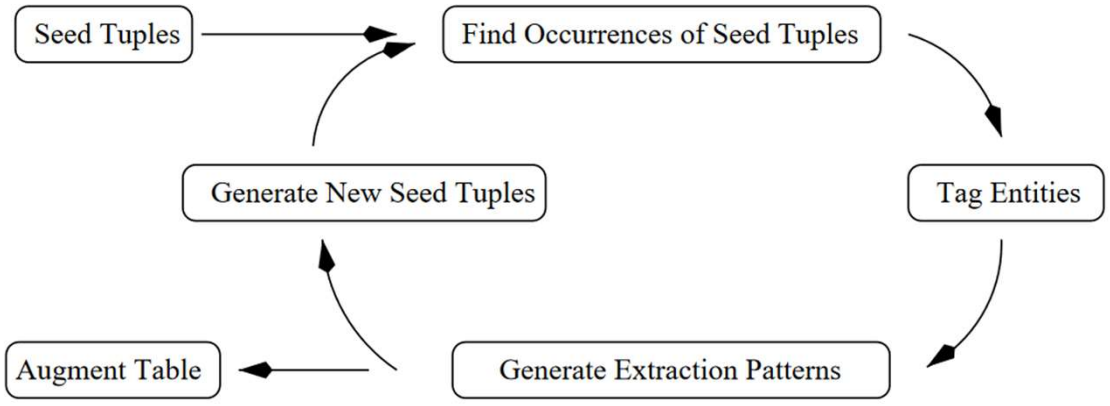
- Distant supervision
  - Idea: derive labels from an existing KB
    - Assume sentences with related entities are positive examples
    - Assume sentences with unrelated entities are negative examples

  - Problem: are those assumptions reliable?
    - Not all sentences with related entities are truly positive examples

    (but the benefit of more data outweighs the harm of noisier data)
    - Need an existing KB to start from – can't start from scratch

Milestones

- 1998 *Beginning*
  - At the 7th Message Understanding Conference (MUC). Since this is considered as part of **template filling**, they call it **template relations**. Relations are limited to organizations: employee_of, product_of, and location_of.
- 2000 *Snowball*
  - Agichtein and Gravano propose **Snowball**, a semi-supervised approach to generating patterns and extracting relations from a small set of seed relations.



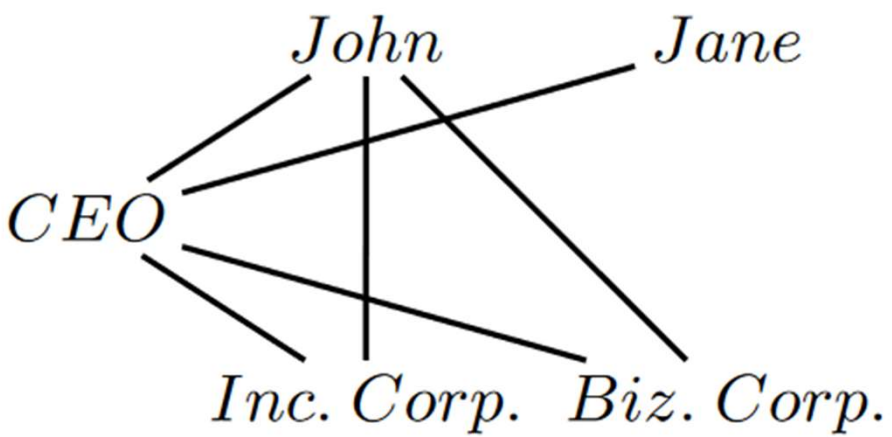
Milestones

- 2003 *Kernel-based approaches*
  - **Shallow parse trees**. Zelenko et al., 2003.
  - **Dependency parse trees.** Culotta and Sorensen, 2004;
  - **Denpendency path kernels**. Bunescu and Mooney, 2005
  - **Convolutional parse kernels.** Zhang et al., 2006
  - **Composite kernels.** Choi et al., 2009.
- 2004 *Feature-based approaches*
  - **A MaxEnt model** is used along with lexical, syntactic and semantic features. Kambhatla et al., 2004.
  - **A SVM model** with more syntactic features using kernels. Zhao and Grishman. 2005.

Milestones

- 2005 *Higher-order relations (McDonald et al.)*
  - Binary relations are represented as a graph, from which cliques are extracted.

a. Relation graph G



b. Tuples from G

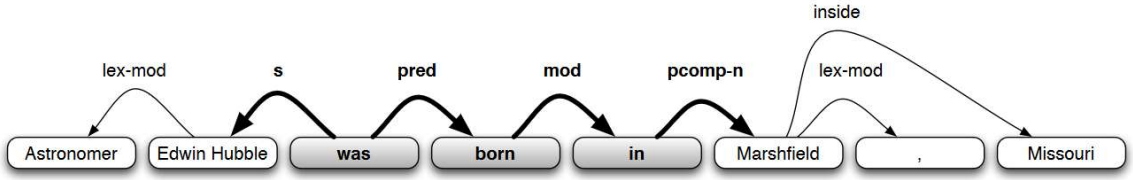
$(John, CEO, \perp)$
$(John, \perp, Inc.\ Corp.)$
$(John, \perp, Biz.\ Corp.)$
$(Jane, CEO, \perp)$
$(\perp, CEO, Inc.\ Corp.)$
$(\perp, CEO, Biz.\ Corp.)$
$(John, CEO, Inc.\ Corp.)$
$(John, CEO, Biz.\ Corp.)$

"John and Jane are CEOs at Inc. Corp. and Biz. Corp. respectively."

Milestones

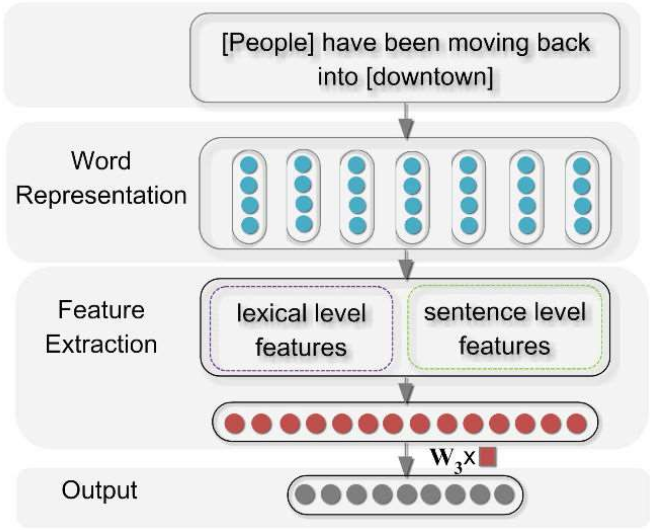
- 2009 *Distant supervision (Mintz et al.)*
  - Proposed to avoid the cost of producing hand-annotated corpus.
  - Using entity pairs that appear in **Freebase**
  - In the early 2010s, distant supervision becomes an active area of research.

Feature type	Left window	NE1	Middle	NE2	Right window
Lexical	[]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[]
Lexical	[Astronomer]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[,]
Lexical	[#PAD#, Astronomer]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[, Missouri]
Syntactic	[]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[]
Syntactic	[Edwin Hubble $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[]
Syntactic	[Astronomer $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[]
Syntactic	[]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{lex-mod}$,]
Syntactic	[Edwin Hubble $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{lex-mod}$,]
Syntactic	[Astronomer $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{lex-mod}$,]
Syntactic	[]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{inside}$ Missouri]
Syntactic	[Edwin Hubble $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{inside}$ Missouri]
Syntactic	[Astronomer $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{inside}$ Missouri]

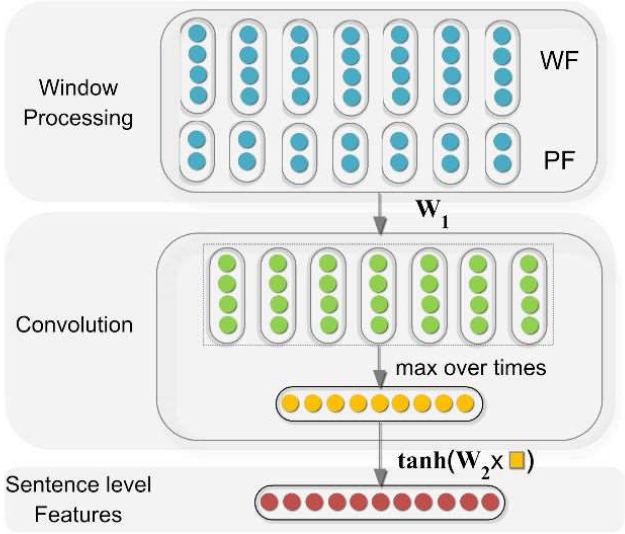


Milestones

- 2014 *Neural networks (Zeng et al.)*
  - Apply **word embeddings** and **CNN** to relation classification (multi-class classification).



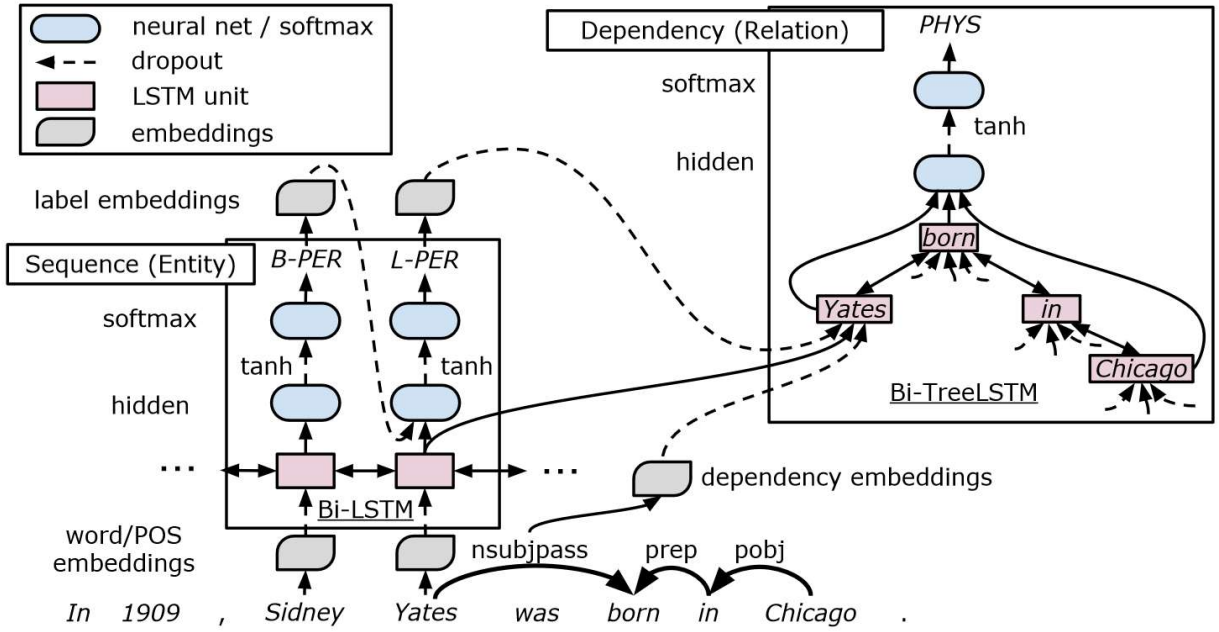
(a) Neural Network Architecture for Relation Classification.



(b) Extracting Sentence-level Features using CNN.

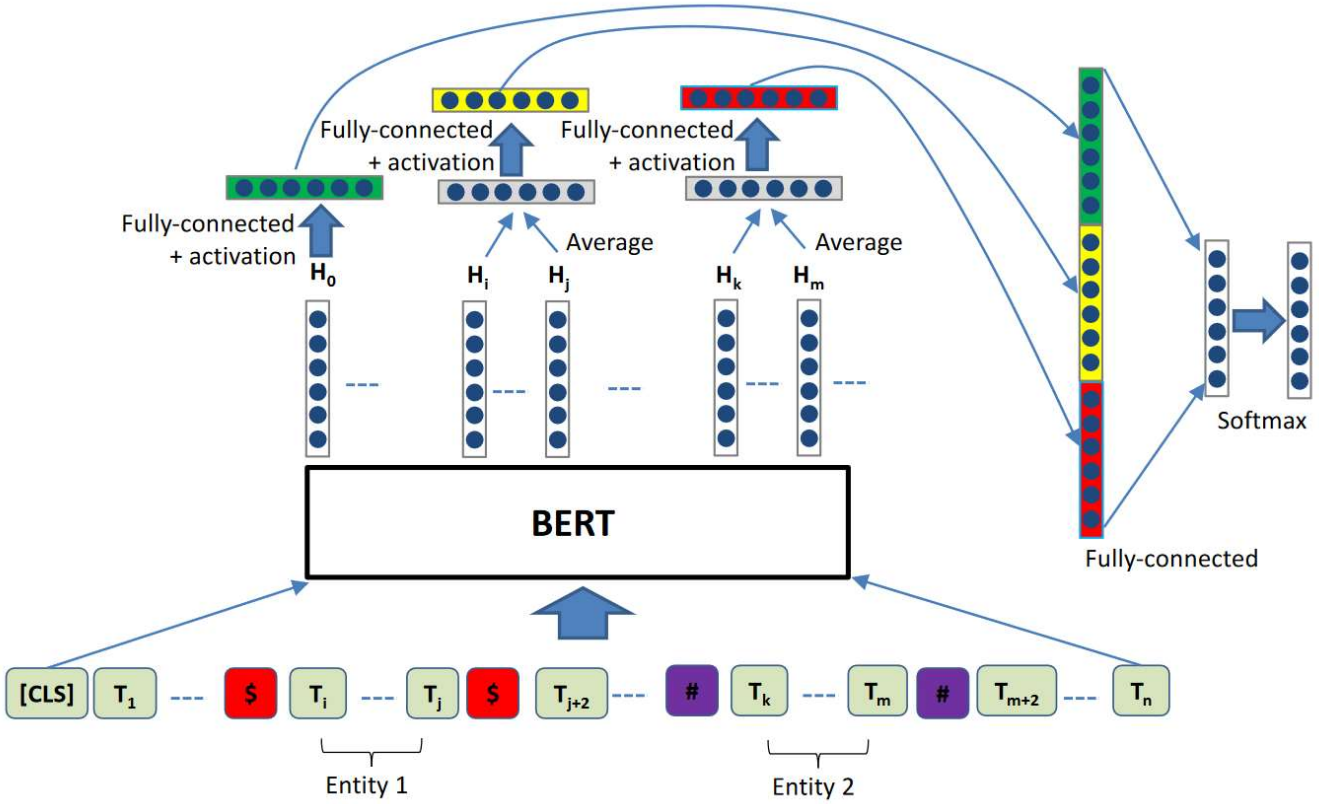
Milestones

- 2016 *Jointly model the tasks of NER and RE (Miwa and Bansal)*
  - A Bi-LSTM is used on word sequences to obtain the named entities.
  - Another Bi-LSTM is used on dependency tree structures to obtain the relations.



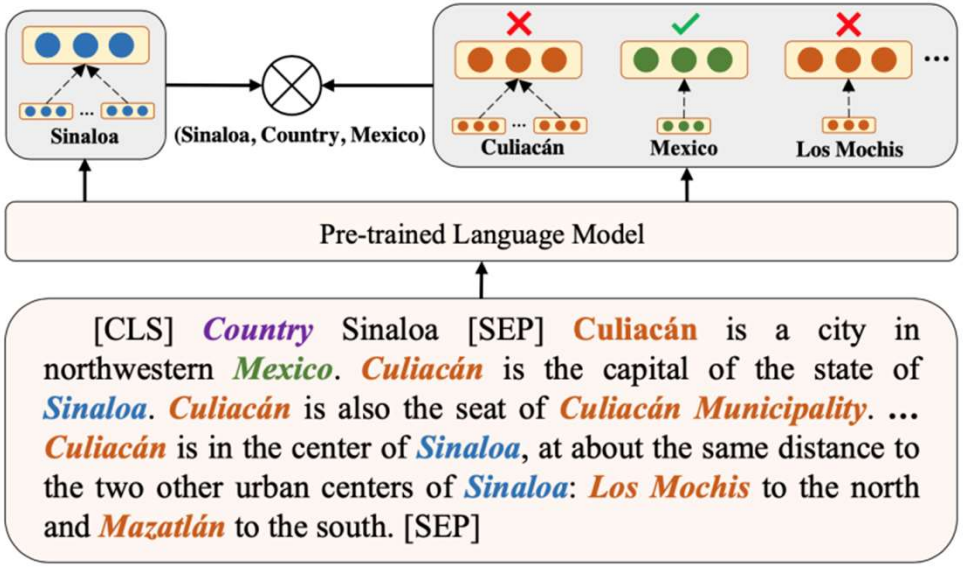
Milestones

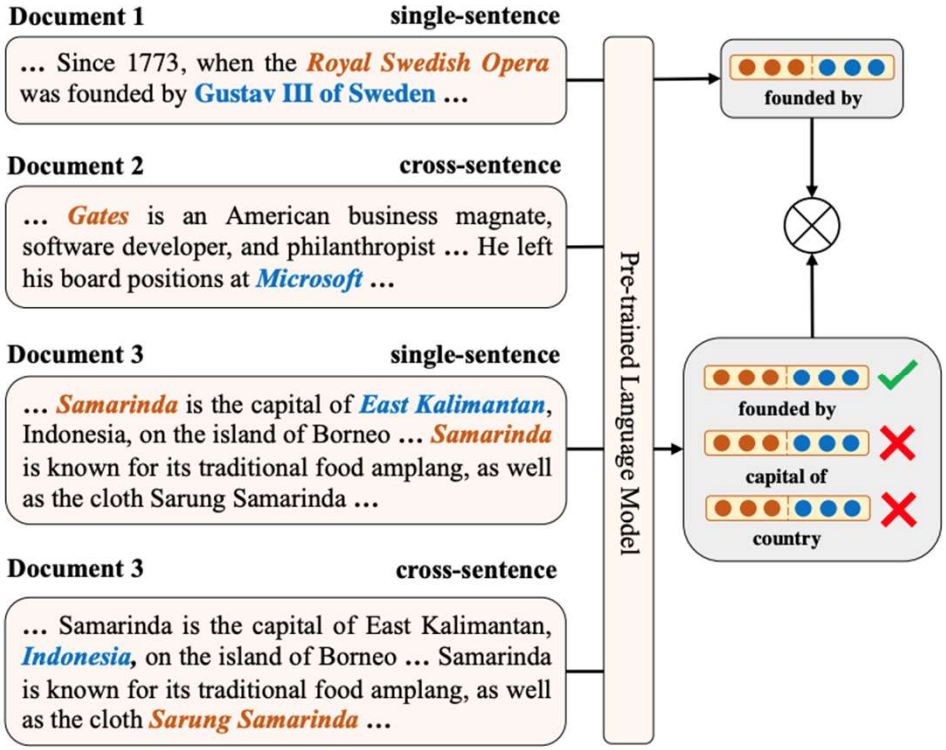
- 2019 R-*BERT (Wu and He)*
  - Achieves SOTA on SemEval-2010 Task 9



Milestones

- 2021 ERICA *(Qin et al.)*
  - Propose a contrastive learning framework to obtain a deep understanding of the entities and their relations in text
  - Two discrimination task: Entity and Relation





Extracting Times

- Times and dates are a particularly important kind of named entity.
- Two-steps
  - Temporal Expression Extraction
  - Temporal Normalization

Extracting Times

- Temporal Expression Extraction
  - **Absolute** temporal expressions. Can be mapped directly to calendar dates, times of day, or both.
  - **Relative** temporal expressions. Map to particular times through some other reference point (as in a week from last Tuesday)
  - **Durations**. Denote spans of time at varying levels of granularity (seconds, minutes, days, weeks, centuries, etc.)

Absolute	Relative	Durations
April 24, 1916	Yesterday	Four hours
The summer of ’77	Next semester	Three weeks
10:15 AM	Two weeks from yesterday	Six days
The 3rd quarter of 2006	Last quarter	The last three quarters

Extracting Times

- Temporal Expression Extraction
  - **Lexical triggers**
    - Nouns, proper nouns, adjectives, and adverbs.
    - Noun phrases, adjective phrases, and adverbial phrases

Category	Examples
Noun	Morning, noon, night, winter, dusk, dawn
Proper noun	January, Monday, Ides, Easter, Rosh Hashana, Ramadan, Tet
Adjective	Recent, past, annual, former
Adverb	Hourly, daily, monthly, yearly

Extracting Times

- Temporal Expression Extraction
  - Approaches
    - Rule-based approaches. Use cascades of automata to recognize patterns at increasing levels of complexity.

Feature	Explanation
Token	The target token to be labeled
Tokens in window	Bag of tokens in the window around a target
Shape	Character shape features
POS	Parts of speech of target and window words
Chunk tags	Base-phrase chunk tag for target and words in a window
Lexical triggers	Presence in a list of temporal terms

    - Sequence-labeling approaches

A	fare	increase	initiated	last	week	by	UAL	Corp's	…
O	O	O	O	B	I	O	O	O	

Extracting Times

- Temporal Normalization
  - **TimeML annotation scheme**
    - In which temporal expressions are annotated with an XML tag.

<TIMEX3 id = “t 1” type= “DATE” value=“2007−07−02” functionInDocument=“CREATION TIME”> July 2, 2007 </TIMEX3> A fare increase initiated <TIMEX3 id=“t2” type=“DATE” value=“2007−W26” anchorTimeID=“t1”>last week</TIMEX3> by United Airlines was matched by competitors over <TIMEX3 id=“t3” type=“DURATION” value=“P1WE” anchorTimeID=“t1”> the weekend </TIMEX3>, marking the second successful fare increase in <TIMEX3 id=“t4” type=“DURATION” value=“P2W” anchorTimeID=“t1”> two weeks </TIMEX3>.

Extracting Times

- Temporal Normalization
  - **ISO 8601 standard** for encoding temporal values

Unit	Pattern	Sample value
Fully specified dates	YYYY-MM-DD	1993-11-24
Weeks	YYYY-Wnn	2007-W27
Weekends	PnWe	P1WE
24-hour clock times	HH:MM:SS	11:13:45
Dates and times	YYYY-MM-DDTHH:MM:SS	1993-11-24T11:00:00
Financial quarters	Qn	1999-Q3

Extracting Times

- Temporal Normalization
  - Rule-based approaches
    - Normalization information for patterns
    - Reference time detection (DCT, previous expression)
    - Relation to reference time -> domain-dependent
    - News domain: tense information can be helpful
    - Narrative domain: chronology assumption (for short passages between underspecified expressions and reference times)
  - Temporal taggers
    - SUTime. [Chang and Manning 2012, 2013.]
    - HeidelTime. [Strötgen & Gertz 2010, 2013; Strötgen et al. 2013]
    - ClearTK-TimeML with Timenorm. [Bethard 2013]

Event Extraction

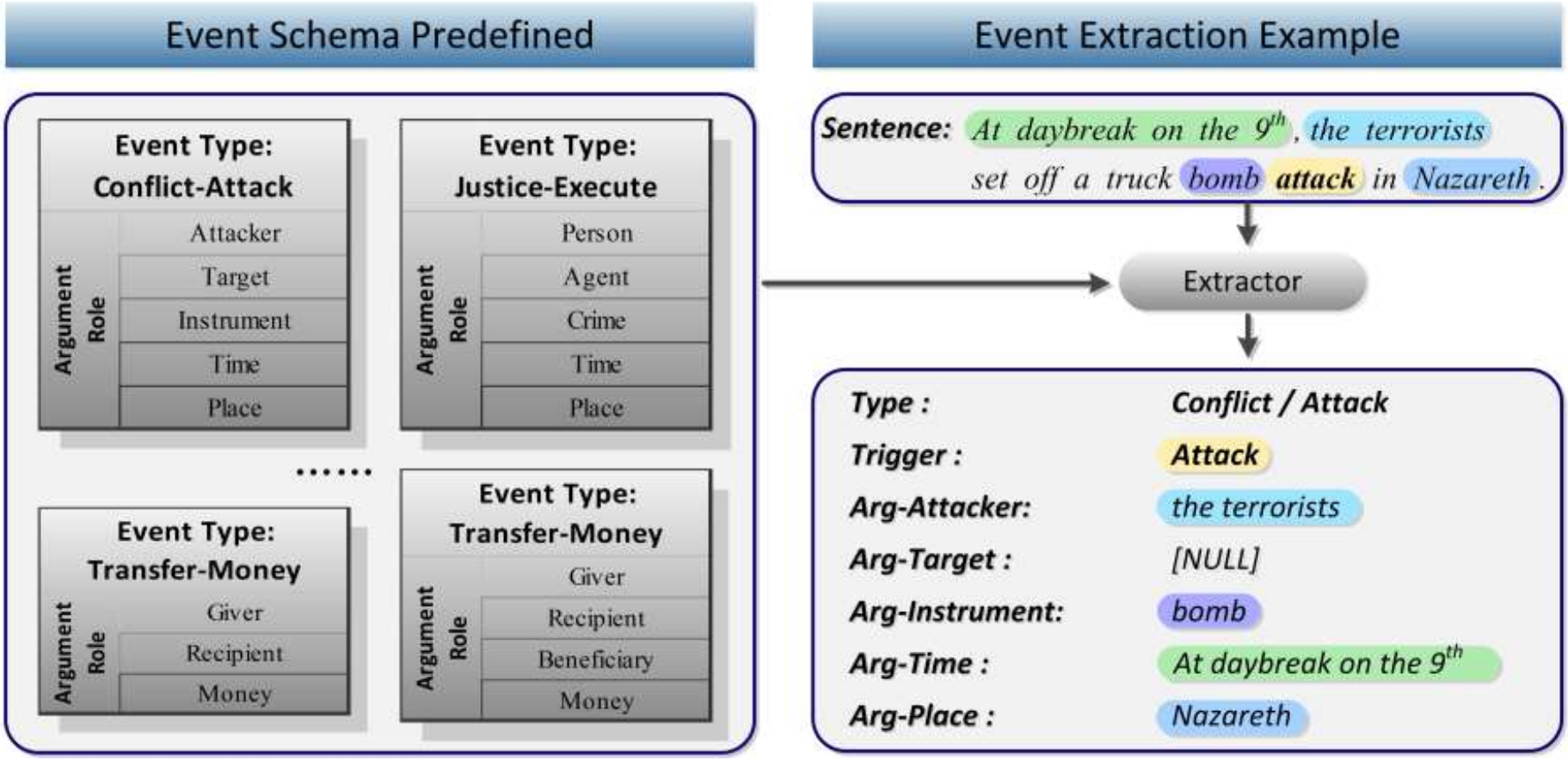
- To identify **mentions of events** in texts.
- And if existing, identify the **event type** as well as all of its **participants** and **attributes**.
- **5W1H:** Who? When? Where? What? Why? How?
- Most event mentions correspond to **verbs**, and most verbs introduce events.
- Closed-domain event extraction & Open-domain event extraction

Closed-domain Event Extraction

- Uses predefined event schema to discover and extract desired events of particular type from text.
- Event structure
  - ACE(Automatic Content Extraction) terminologies
    - **Event mention:** a phrase or sentence describing an event, including a trigger and several arguments.
    - **Event trigger:** the main word that most clearly expresses an event occurrence, typically a verb or a noun.
    - **Event argument:** an entity mention, temporal expression or value that serves as a participant or attribute with a specific role in an event.
    - **Argument role:** the relationship between an argument to the event in which it participants.

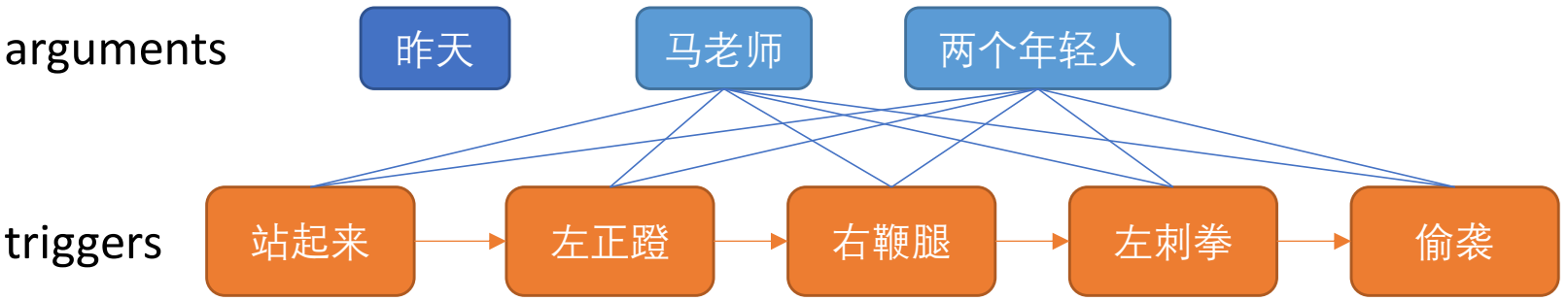
Closed-domain Event Extraction

- An illustration



Open-domain Event Extraction

- Without predefined event schemas
- Detecting events from texts and clustering similar events via extracted **event keywords**.
  - **Event keywords** refers to words/phrases mostly describing an event, and sometimes keywords are further divided into **triggers** and **arguments**.



Open-domain Event Extraction

- Topic Detection and Tracking (TDT)
  - **Story.** A segment of news article describing a specific event
  - **Topic.** A set of events in articles yet strongly related to some real-world topic.
  - **Tasks**
    - **Story segmentation.** Detecting the boundaries of a story from news articles.
    - **First story detection.** Detecting the story that discuss a new topic in the stream of news.
    - **Topic detection.** Grouping the stories that discuss a new topic in the stream of news.
    - **Topic tracking.** Detecting stories that discuss a previously known topic.
    - **Story link detection.** Deciding whether a pair of stories discuss the same topic

Event Extraction Corpus

- ACE 2005 Corpus
  - 599 annotated documents
  - About 6000 labeled events
  - Including English, Arabic and Chinese
  - From newswire articles, broadcast news, weblog…

English annotation example:

Sentence:	At daybreak on the 9th, the terrorists set off a truck bomb attack in Nazareth .
Annotation:	At daybreak on the 9th, the terrorists set off a truck bomb attack in Nazareth .
Label:	Arg-Time Arg-Attacker Arg-Instrument Tri-Attack Arg-Place

Chinese annotation example:

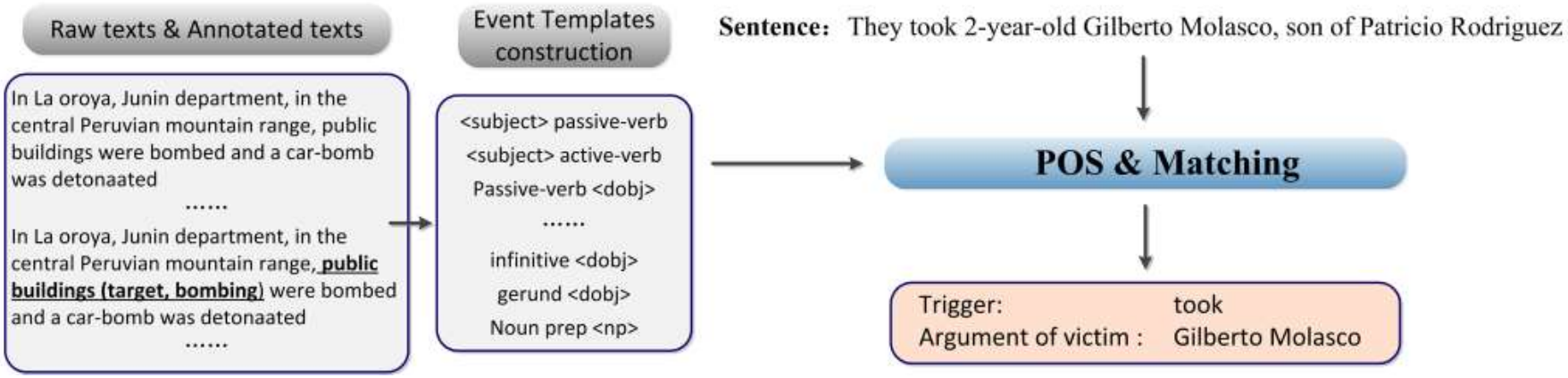
Sentence:	10 日 凌 晨， 恐 怖 分 子 在 拿 撒 勒 镇 制 造 了 一 起 汽 车 炸 弹 爆 炸 事件。
Annotation:	10 日 凌 晨， 恐 怖 分 子 在 拿 撒 勒 镇 制 造 了 一 起 汽 车 炸 弹 爆 炸 事件。
Label:	[B] [I] [I] [I] O [B] [I] [I] [I] O [B] [I] [I] [I] O O O O O O O [B] [I] [B] [I] O O O

Event Extraction Corpus

- The TDT Corpus
  - From TDT-1 to TDT5
  - Including English and Chinese
  - Each corpus contains millions of news stories and annotated with hundreds of topics
  - From newswire, broadcast articles…
- Other domain-specific corpus
  - **Biological Domain.** BioNLP-ST, GENIA, BioInfer, Gene regulation event corpus, GeneReg corpus and PPI corpora.
  - **Breaking news.** TimeBank, CEC(Chinese Event Corpus)
  - **Military intelligence, Terrorist attacks, Chip technology and financial.** MUC series copus

Approaches

- Pattern matching
  - First constructs some specific event templates
  - Then performs template matching to extract an event with a single argument from text

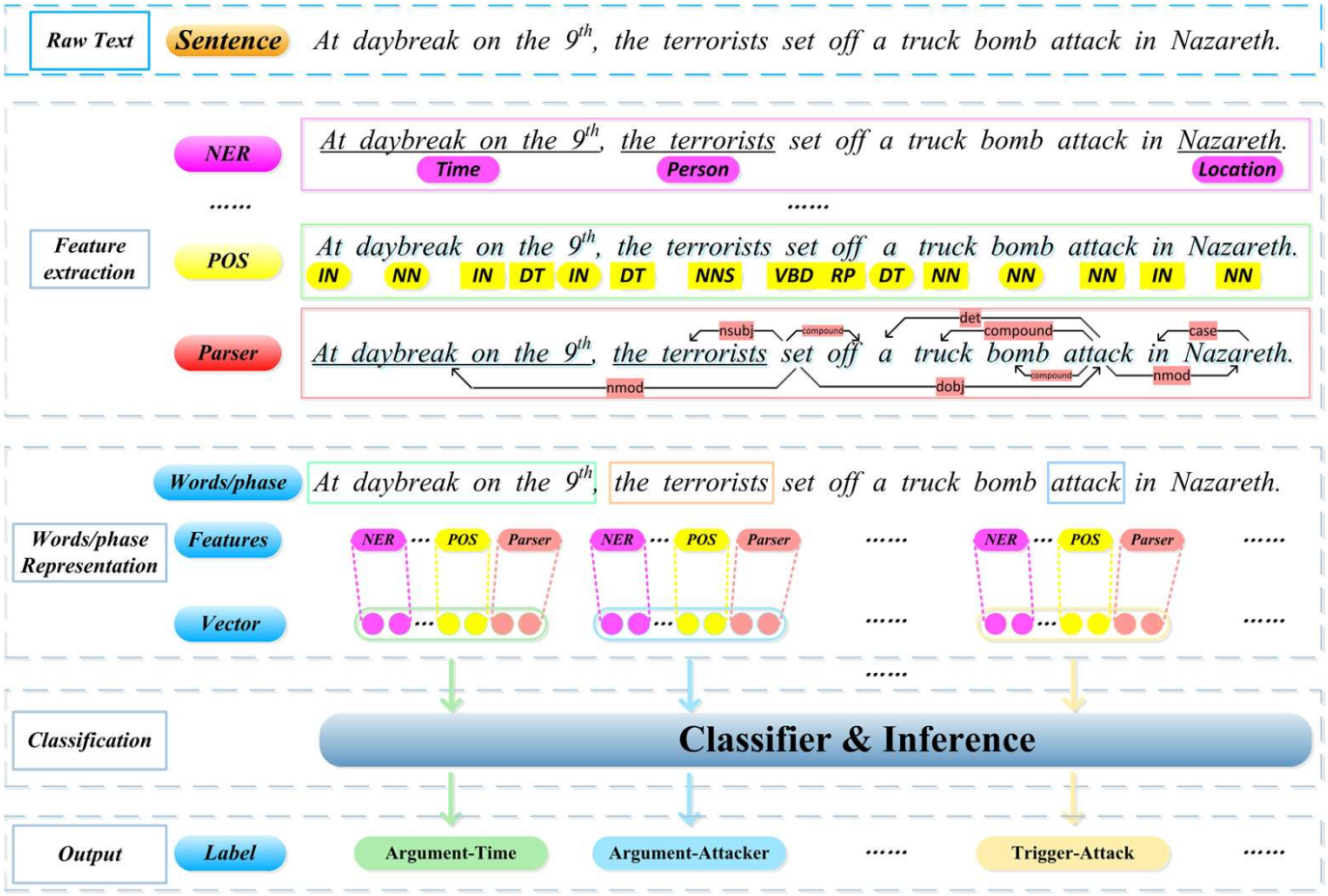


Approaches

- Machine learning
  - Learning classifiers from training data
  - Applying classifiers for event extraction from new text.
  - Two stages and Four subtasks
    - Stage 1:
      - Trigger detection
      - Trigger/Event type identification
    - Stage 2:
      - Argument detection
      - Argument role identification

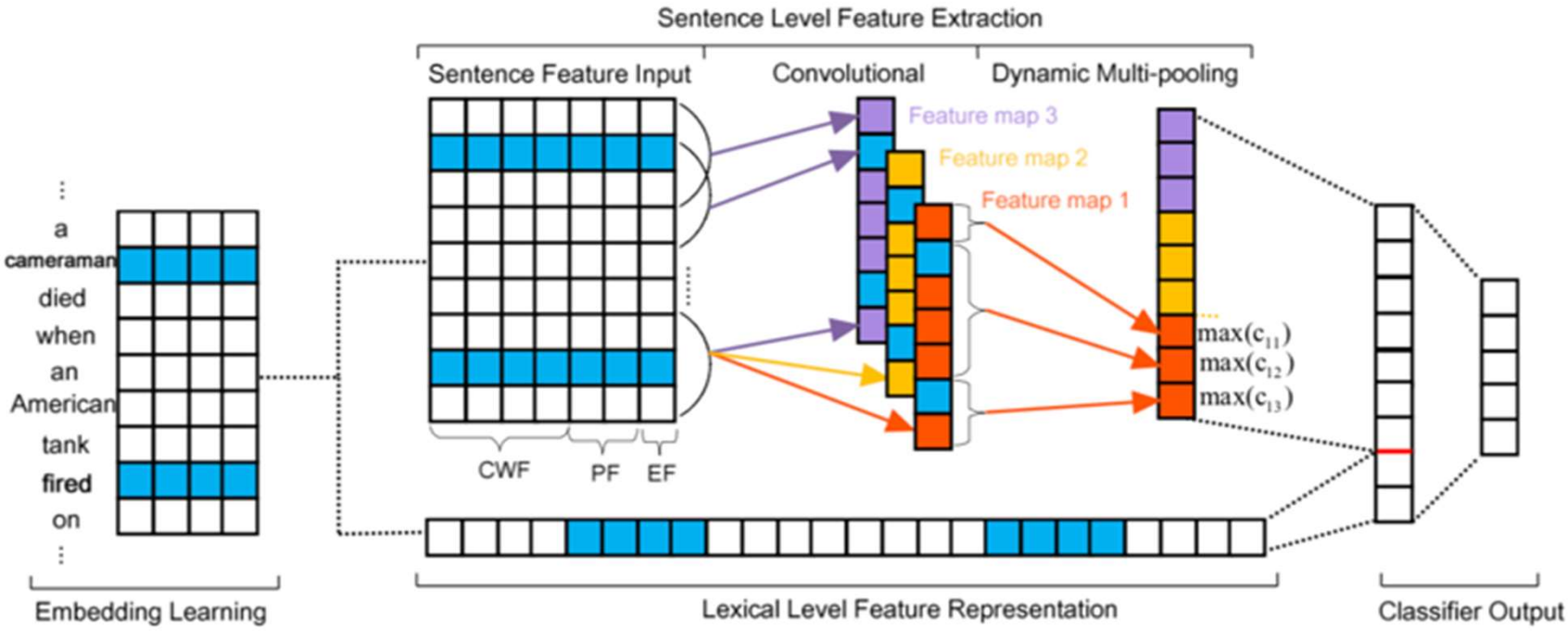
Approaches

- Illustration of machine learning-based approach



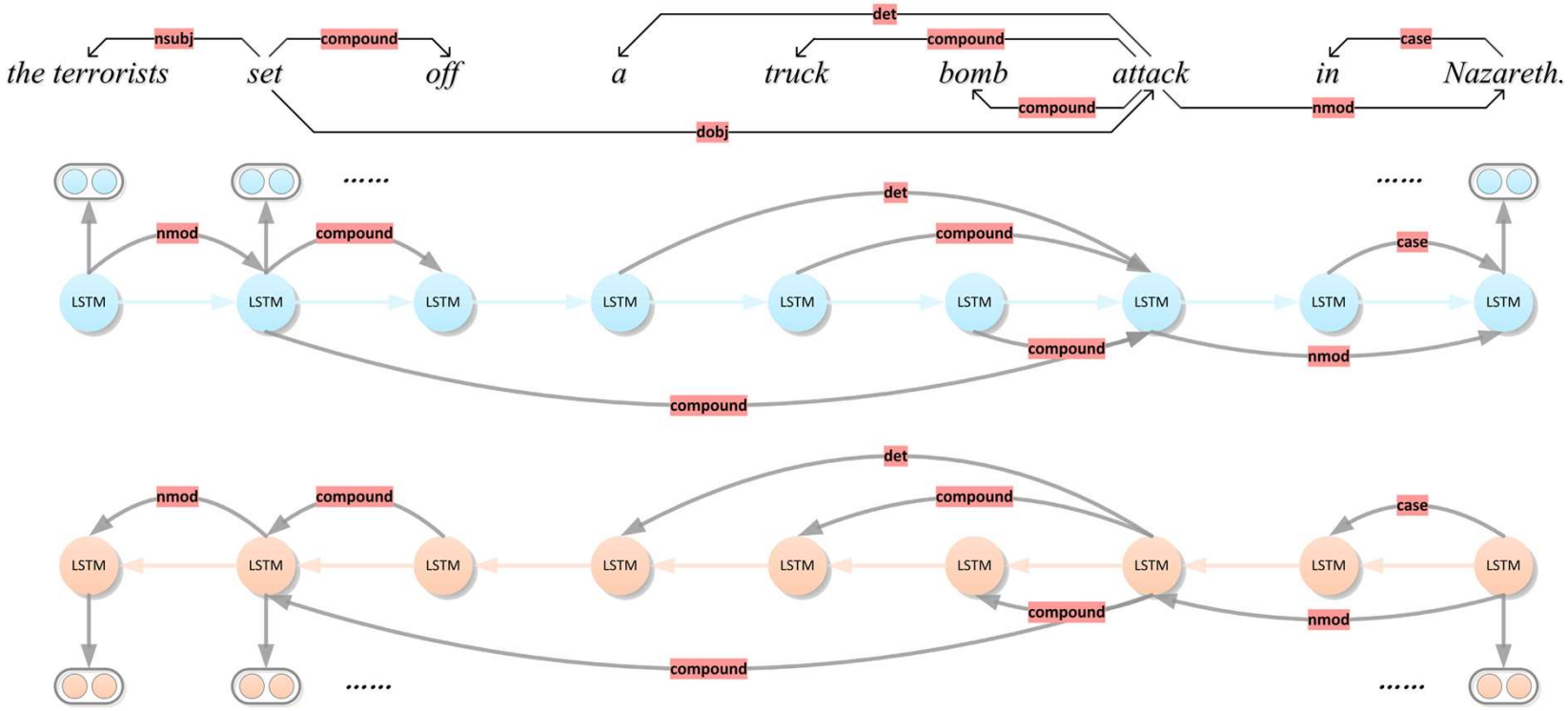
Approaches

- Deep learning
  - DMCNN(Dynamic Multi-Pooling CNN Model. Zhao and Liu, ACL 2015)
    - Evaluate each part of a sentence via a dynamic multi-pooling layer extracting both lexical-level and sentence-level features.



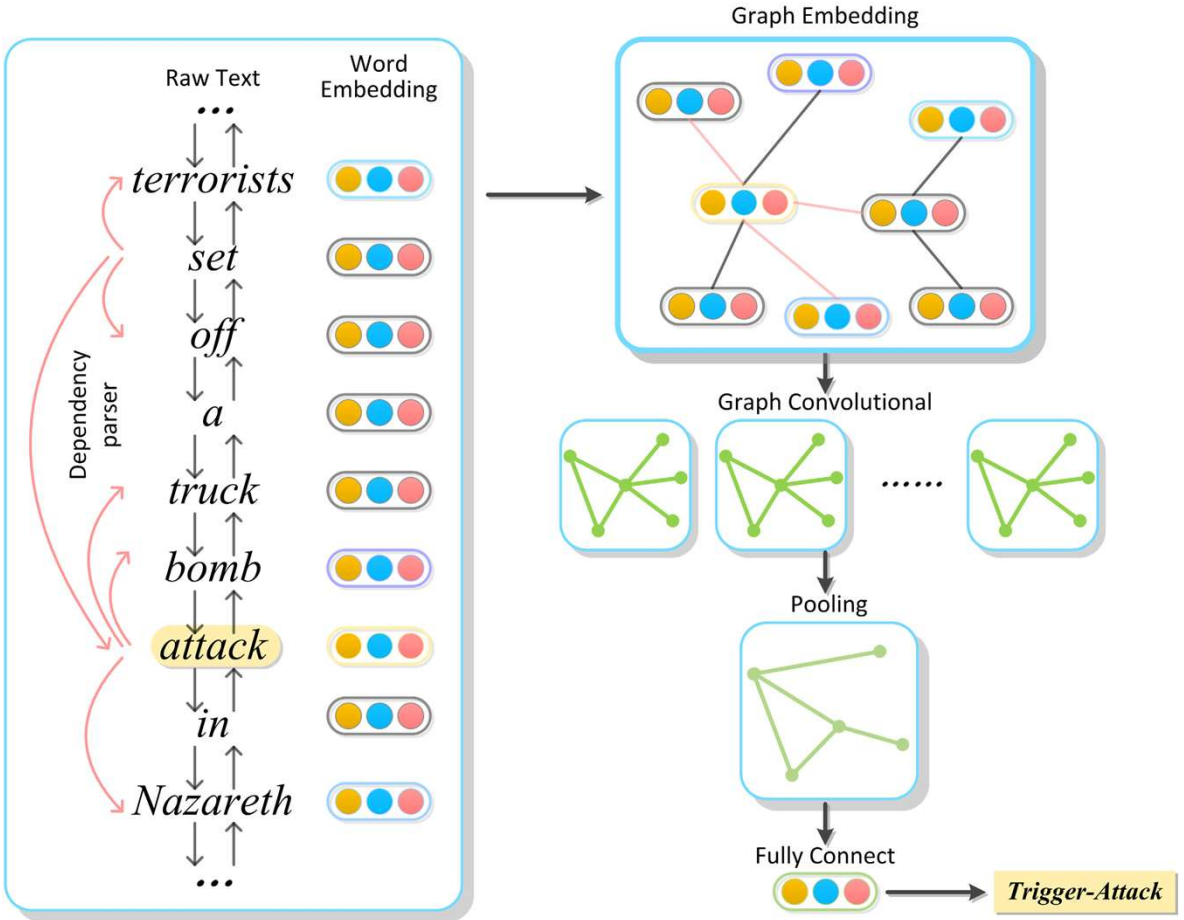
Approaches

- Deep learning
  - dbRNN(dependency bridge RNN Model. Sha et al., AAAI 2018)
    - Using dependency bridges to build a tree-structured RNN.



Approaches

- Graph Neural Networks



Approaches

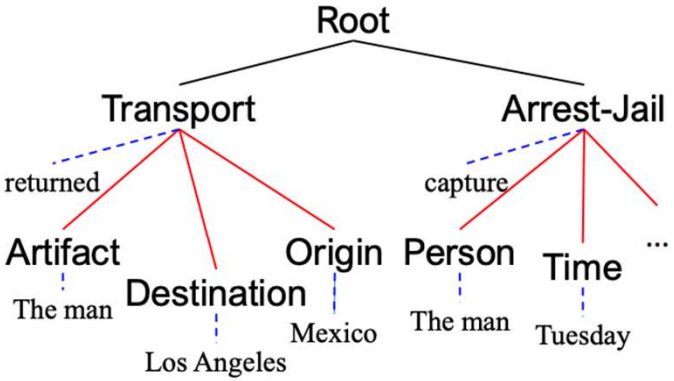
- Deep learning
  - Text2Event(Sequence2Structure generation Model. Lu et al., ACL 2021)
    - Directly extract events from the text in an end-to-end manner
    - Achieve competitive performance using only record-level annotations

The man returned to Los Angeles from Mexico following his capture Tuesday by bounty hunters.

Event Type	Transport
Trigger	returned
Artifact	The man
Destination	Los Angeles
Origin	Mexico

Event Type	Arrest-Jail
Trigger	capture
Person	The man
Time	Tuesday
Agent	bounty hunters

(a) Record format.



(b) Tree format.

```
((Transport returned
  (Artifact The man)
  (Destination Los Angeles)
  (Origin Mexico))
 (Arrest-Jail capture
  (Person The man)
  (Time Tuesday)
  (Agent bounty hunters))
```

(c) Linearized format.

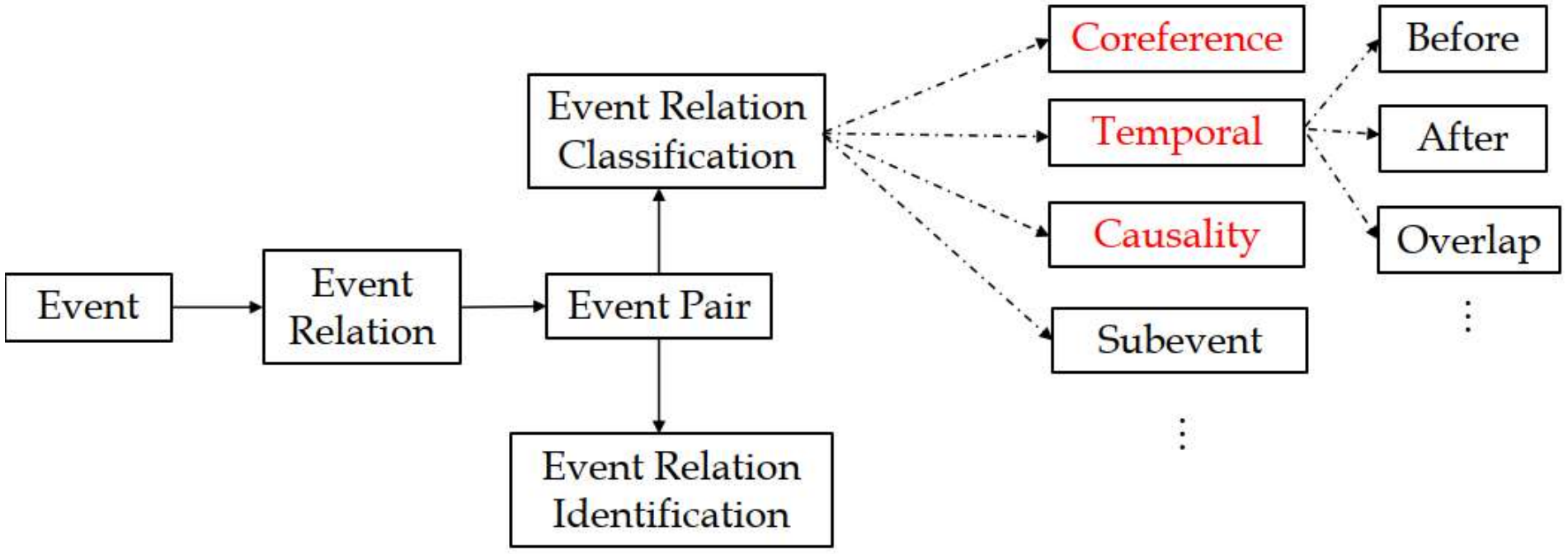
Approaches

- Semi-supervised Learning
  - Expand a small set of labeled data to a larger corpus
  - Train extraction models from mixture data.
  - Joint data expansion and model training
    - L. Huang et al., ACL 2018; Wang et al., ACL 2019
  - Data expansion from knowledge bases
    - Y. Zeng et al., AAAI 2018; J. Araki and T. Mitamura, COLING 2018
  - Data expansion from multi-language data
    - M. Li et al., NAACL 2019

Approaches

- Unsupervised Learning
  - Focus on open-domain event extraction tasks. i.e. detecting trigger and arguments
  - Event mention detection and tracking
    - T. Ge et al., COLING 2016
  - Event extraction and clustering
    - L. Huang et al., ACL 2016; Q. Yuan et al., CIKM 2018.
  - Event extraction from social media
    - Twitter, Facebook and etc.

Event Relation Classification



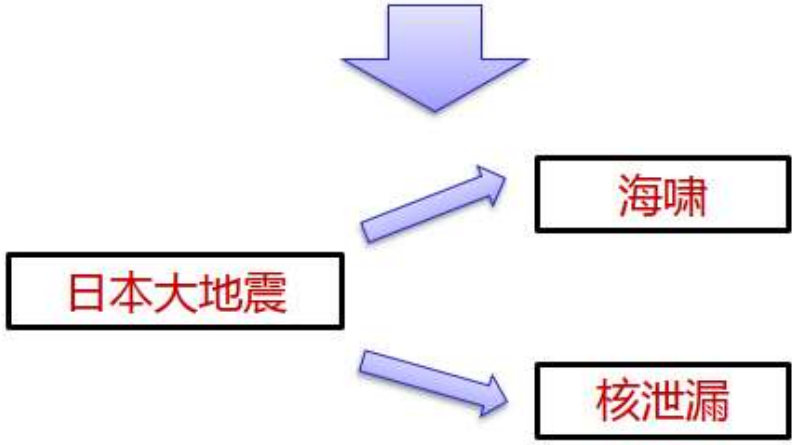
Event Relation Classification

- Coreference relation

Event Relation Classification

- Causality relation

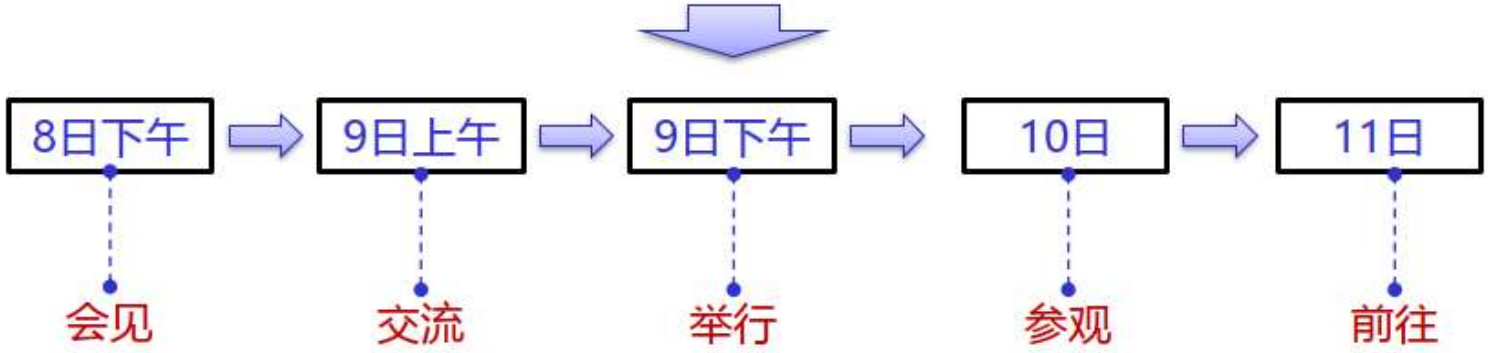
2011年3月11日13时46分，日本发生里氏9.0级大地震，地震随即引发了海啸，后来更使福岛核电站发生核泄漏危机。地震震中位于宫城县以东太平洋海域，震源深度20公里，东京有强烈震感。



Event Relation Classification

- Temporal Relation

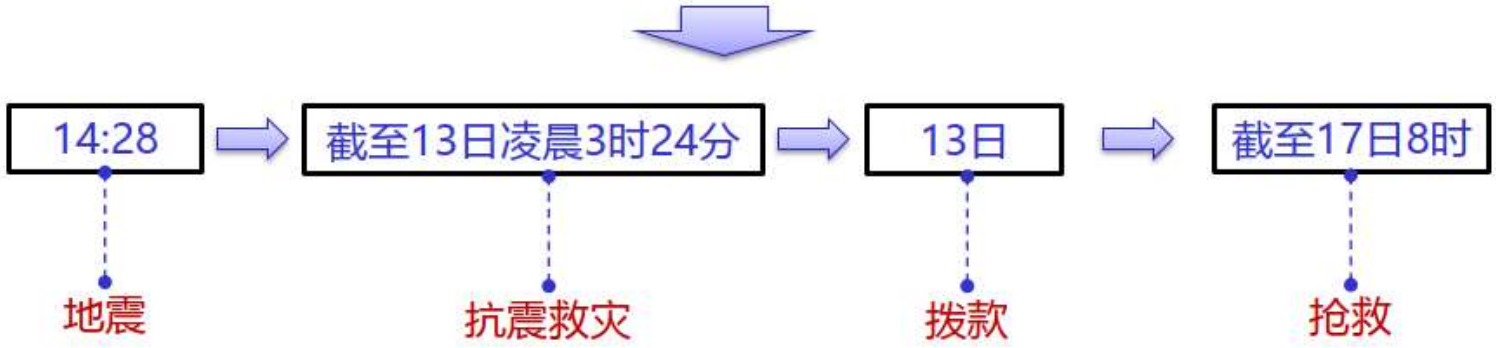
据了解，代表团将在乌鲁木齐停留３天。活动暂定为８日下午，新疆维吾尔自治区领导会见代表团成员；９日上午在新疆师范大学与新疆６所高校师生交流，下午在新疆人民会堂举行报告会和体育表演；１０日参观乌鲁木齐市容。“奥运健儿祖国西部行”代表团在结束乌鲁木齐的活动后，将于１１日前往南疆喀什。



Event Relation Classification

- Temporal Relation

5月12日14:28，四川汶川发生7.8级地震。截至13日凌晨3时24分，武警部队已出动13000余名官兵急赴灾区抗震救灾。13日，为帮助地震灾区开展紧急救灾工作，中央财政紧急下拨款地震救灾资金8.6亿元。截至17日8时，参加救援的民兵预备役部队人员已从灾区废墟中一共抢救出1352名幸存者，发现和掩埋遇难者遗体2756人，救治伤员7600多人。



Event Relation Classification

- Subevents relation

偷袭珍珠港使第二次世界大战的规模迅速扩大,发展成世界大战。
斯大林格勒大会战是苏联军队在苏联卫国战争中对德国军队的一次决定性战役，也是欧洲东线战场的转折。苏军在这场保卫战中所取得的胜利具有巨大的战略意义，不仅扭转了苏德战场的整个形势，而且成为第二次世界大战的根本转折点。
而诺曼底登陆是规模最大,最成功的战役.是苏联免于单独作战,同时让德国成为两线作战.开辟第 二战场，为二战结束奠定基础.