BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

Analyzing Errors in OpenIE Systems

Rudolf Schneider
PhD. student @ DATEXIS
rudolf.schneider@beuth-hochschule.de

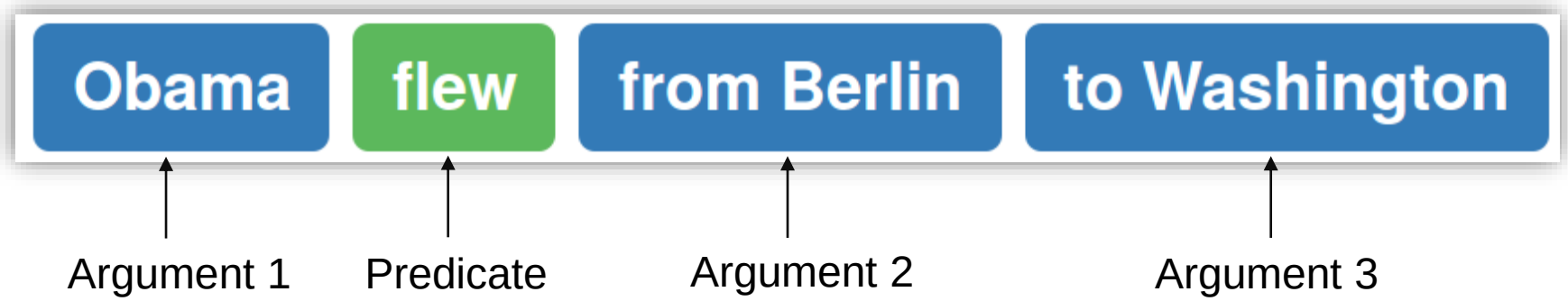
Open Information Extraction in a nutshell

"Open Information Extraction (OpenIE) is an extraction paradigm that facilitates **domain-independent** discovery of *relations extracted* from text and readily scales to the **diversity** and **size** of the Web corpus."

(Banko et. al., 2007)

Open Information Extraction

n-ary tupel consisting of *n Arguments* and one *Predicate.*



Methods do not agree on Common Benchmark

System	# Sentences	Domain	Metric	Type of Judge
TextRunner [2]	~ 400	Web	% Correct	**Human** (Authors)
ReVerb [5]	500	Web	Precision / Recall – AUC	**Human** (?)
KrakeN [3]	500	Web	% Correct	**Human** (?)
Ollie [1]	300	News, Wiki, Biology	Precision / Yield – AUC	**Human** (?)
ClausIE [4]	800	Web, Wiki, News	Precision / Yield	**Human** (?)
Stanford OpenIE [6]	2,01M Documents	News, Web, Forum	*TAC KBP Slot Filling 2013* - Precision	**Machine**
NestIE [8]	400	Wiki, News	Informativeness	**Human** (2 CS-Students)

Four well known OpenIE systems evaluated on four datasets

Evaluated OpenIE Systems

- N-ary
  - OpenIE 4.2 [7]
  - ClausIE [4]
  - PredPat [12]
- Binary
  - Stanford OpenIE [6]
  - ClausIE (binary mode)

Datasets

Name	Type	Domain	Sent.	# Tuple
NYT-222	n-ary	News	222	222
WEB-500	binary	Web/News	500	461
PENN-100	binary	Mixed	100	51
OIE2016	n-ary	Wiki	3200	10359

Systems are quantitative and qualitative evaluated in two experiments

Quantitative Benchmark

- 2590 sentences
- Precision
- Recall
- F2 Score

Qualitative Benchmark

- 68 sentences
- 749 manually evaluated extractions
- Qualitative error classes

What is correct?

Strict Containment Match

- Gold Predicate is fully contained.
- Number of Arguments is equal.
- All arguments are fully contained.

Relaxed Containment Match

- Gold Predicate is fully contained.
- All arguments are fully contained.

Strict Containment vs. Relaxed Containment

Strict Containment

Relaxed Containment

Gold Annotation

The airfield | hosted | an internationally recognised | Air Show | for several years .

Predicted Annotation

The airfield | hosted | an internationally recognised Air Show | for several years | .

RelVis – Benchmarking Tool for OpenIE Systems

RelVis | Documents | Gold Standards | Results | About

Random Selection (1)

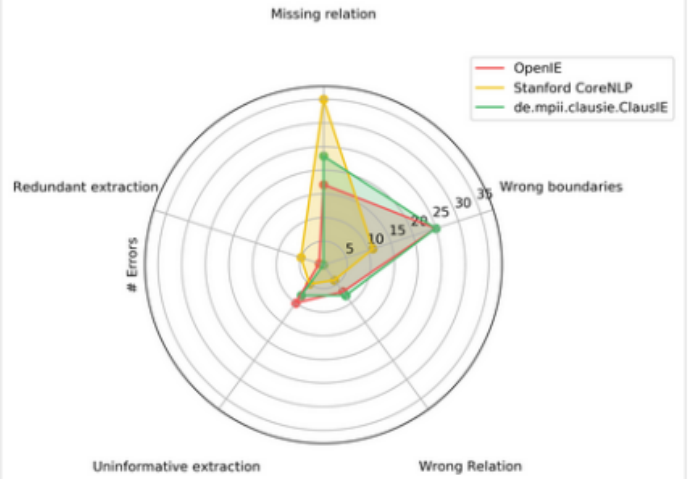
ID	Sentence	# Annotations	Action
2088-2211	Introduction Chagas disease (CD) is endemic to the American continent, and 25 million people are at risk in Latin America.	0 2 2	View
2212-2531	QVC Network Inc. said it completed its acquisition of CVN Cos. for about $ 423 million .	0 3 4	View
2532-2714	Mr. Moon 's support for a Watergate - beleaguered Richard Nixon , the Koreagate scandal , and his prison sentence for income - tax evasion did not help the church 's recruitment efforts .	0 1 3	View
2715-2838	On Cambodia : `` Let 's assume that { former Cambodian leader Prince Norodom } Sihanouk does what the press wants him to do and joins up with { Vietnamese - backed Cambodian leader } Hun Sen .	0 2 2	View
2839-3008	Texas Eastern Transmission will build and operate the system , which will connect the Arkoma Basin with several interstate pipelines .	0 1 2	View
3009-3269	It has been shown that dogs develop diffuse chronic myocarditis with histological and electrocardiographic changes that are also found in humans ; therefore, this animal represents a useful experimental model that is gaining attention in the CD research field.	0 3 5	View
3270-3402	Greek and Roman pagans , who saw their relations with the gods in political and social terms , scorned	0 1 1	View

Legend (2)

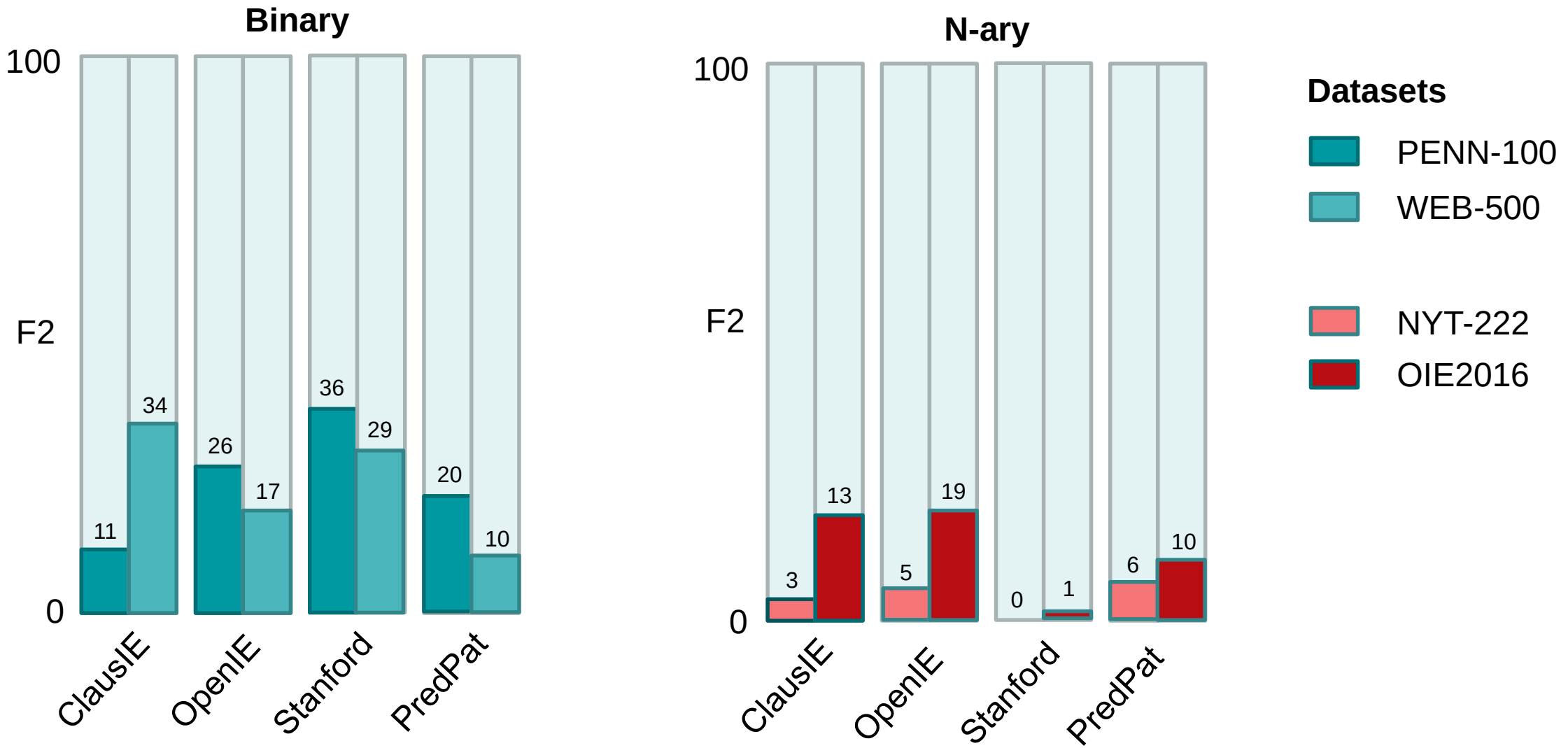
User Annotation | Stanford OpenIE | OpenIE 4.2.1 | PredPatt | ClausIE 1.0

Error Statistics (3)

Error Class Distribution by Tool

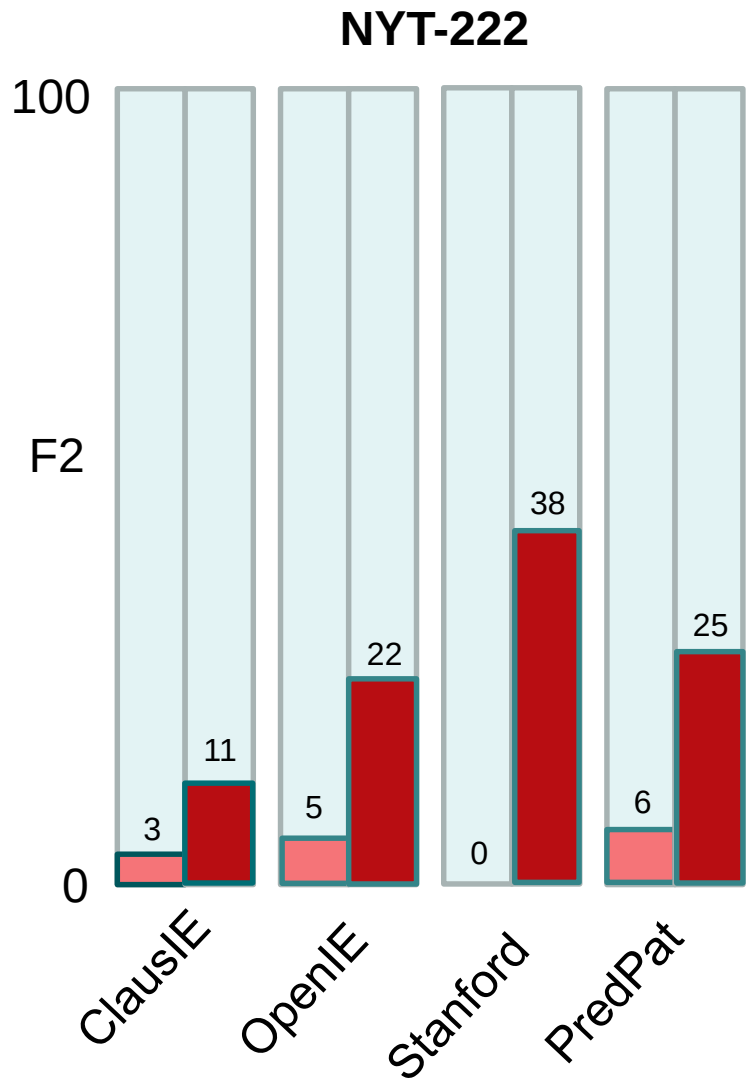


OpenIE systems seem to be biased towards binary tuples and not robust against noise



Evaluating boundaries is important

- Stanford outperforms all systems if relaxed match is applied
  - Over-specific arguments
  - Additional effort for downstream applications



Impact of Error Classes from Literature

Error Class	True Positives	False Positives	False Negatives
Wrong Boundaries [2]	Less	More	More
Redundant Extraction [4]	No Impact	More	No Impact
Missing Relation [5]	Less	No Impact	More
Uninformative Extraction [5]	No Impact	More	No Impact
Wrong Relation [6]	No Impact	More	No Impact
Out of Scope	No Impact	No Impact	No Impact

Wrong Boundaries and Missing Extractions are the main causes for errors

- **Wrong Boundaries**
  - Annotation style
  - Overestimating argument spans
- **Missing extraction**
  - Noise
  - Wrong intermediate structure
- **Wrong and uninformative**
  - Missed negations
  - Co-reference
  - Processing of adjectival triggers

Conclusion

We need adaptable and generalizing OpenIE Systems with a well defined Annotation policy

Task in general

- We need a stringent formalized annotation policy

Datasets

- We observe many Out of Scope “errors”
  - Datasets do not cover the capabilities of current OpenIE systems.
- Large datasets with consistent annotation policy
- Datasets for idiosyncratic domains

Next generation OIE systems

- Apply normal forms form data base theory
- Be (fast) adaptable to downstream tasks
- Effective resolution of co-references
- (Do not rely on intermediate structure)
- Generalize better to unseen text
  - Robust against noisy data
  - Perform well on idiosyncratic texts

Literature

- [1] M. Schmitz, R. Bart, S. Soderland, O. Etzioni, and others, “Open language learning for information extraction,” in Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012, pp. 523–534.
- [2] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni, “Open Information Extraction from the Web.,” in IJCAI, 2007, vol. 7, pp. 2670–2676.
- [3] A. Akbik and A. Löser, “Kraken: N-ary facts in open information extraction,” in Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction, 2012, pp. 52–56.
- [4] L. Del Corro and R. Gemulla, “Clausie: clause-based open information extraction,” in Proceedings of the 22nd international conference on World Wide Web, 2013, pp. 355–366.
- [5] A. Fader, S. Soderland, and O. Etzioni, “Identifying relations for open information extraction,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2011, pp. 1535–1545.
- [6] G. Angeli, M. J. Premkumar, and C. D. Manning, “Leveraging linguistic structure for open domain information extraction,” Linguistics, no. 1/24, 2015.
- [7] “allenai/openie-standalone,” GitHub. [Online]. Available: https://github.com/allenai/openie-standalone.
- [8] N. Bhutani, H. V. Jagadish, and D. R. Radev, “Nested Propositions in Open Information Extraction,” in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016, 2016, pp. 55–64.
- [9] G. Stanovsky and I. Dagan, “Creating a Large Benchmark for Open Information Extraction,” in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), Austin, Texas, 2016, p. (to appear).
- [10] F. Mesquita, J. Schmidek, and D. Barbosa, “Effectiveness and Efficiency of Open Relation Extraction,” in Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2013, pp. 447–457.
- [11] M. Mausam, “Open Information Extraction Systems and Downstream Applications,” presented at the International Joint Conference on Artificial Intelligence (IJCAI), New York, 2016.
- [12] Keisuke, Aaron Steven White, Drew Reisinger, et al. "Universal Decompositional Semantics on Universal Dependencies."