Analyzing Errors in OpenIE Systems

Rudolf Schneider
PhD. student @ DATEXIS
rudolf.schneider@beuth-hochschule.de
“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)
n-ary tupel consisting of \( n \) Arguments and one Predicate.
Methods do not agree on Common Benchmark

<table>
<thead>
<tr>
<th>System</th>
<th># Sentences</th>
<th>Domain</th>
<th>Metric</th>
<th>Type of Judge</th>
</tr>
</thead>
<tbody>
<tr>
<td>TextRunner [2]</td>
<td>~ 400</td>
<td>Web</td>
<td>% Correct</td>
<td>Human (Authors)</td>
</tr>
<tr>
<td>KrakeN [3]</td>
<td>500</td>
<td>Web</td>
<td>% Correct</td>
<td>Human (?)</td>
</tr>
<tr>
<td>Ollie [1]</td>
<td>300</td>
<td>News, Wiki,</td>
<td>Precision / Yield – AUC</td>
<td>Human (?)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ClausIE [4]</td>
<td>800</td>
<td>Web, Wiki, News</td>
<td>Precision / Yield</td>
<td>Human (?)</td>
</tr>
<tr>
<td>NestIE [8]</td>
<td>400</td>
<td>Wiki, News</td>
<td>Informativeness</td>
<td>Human (2 CS-Students)</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Domain</th>
<th>Sent.</th>
<th># Tuple</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYT-222</td>
<td>n-ary</td>
<td>News</td>
<td>222</td>
<td>222</td>
</tr>
<tr>
<td>WEB-500</td>
<td>binary</td>
<td>Web/News</td>
<td>500</td>
<td>461</td>
</tr>
<tr>
<td>PENN-100</td>
<td>binary</td>
<td>Mixed</td>
<td>100</td>
<td>51</td>
</tr>
<tr>
<td>OIE2016</td>
<td>n-ary</td>
<td>Wiki</td>
<td>3200</td>
<td>10359</td>
</tr>
</tbody>
</table>
Experiments

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**

**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.

**Relaxed Containment**
RelVis – Benchmarking Tool for OpenIE Systems

### Random Selection

<table>
<thead>
<tr>
<th>ID</th>
<th>Sentence</th>
<th>Annotations</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>2086-2211</td>
<td>Introduction Chagas disease (CD) is endemic to the American continent, and 25 million people are at risk in Latin America.</td>
<td>2</td>
<td>View</td>
</tr>
<tr>
<td>2212-2931</td>
<td>OVC Network Inc. said it completed its acquisition of CVN Cos. for about $423 million.</td>
<td>4</td>
<td>View</td>
</tr>
<tr>
<td>2532-2714</td>
<td>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.</td>
<td>3</td>
<td>View</td>
</tr>
<tr>
<td>2839-3008</td>
<td>Texas Eastern Transmission will build and operate the system, which will connect the Arkoma Basin with several interstate pipelines.</td>
<td>2</td>
<td>View</td>
</tr>
<tr>
<td>3009-3269</td>
<td>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.</td>
<td>3</td>
<td>View</td>
</tr>
<tr>
<td>3270-3402</td>
<td>Greek and Roman pagans, who saw their relations with the gods in political and social terms, scorned</td>
<td>1</td>
<td>View</td>
</tr>
</tbody>
</table>

### Error Statistics

- **Missing relation**: OpenIE 4.2.1, Stanford CoreNLG, de.mpi.lit.ClausiE
- **Redundant extraction**: Stanford CoreNLG, OpenIE 4.2.1, de.mpi.lit.ClausiE
- **Wrong boundaries**: Stanford CoreNLG, de.mpi.lit.ClausiE, OpenIE 4.2.1

---

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

Matching Strategy
- Strict Containment
- Relaxed Containment

NYT-222

<table>
<thead>
<tr>
<th></th>
<th>ClausIE</th>
<th>OpenIE</th>
<th>Stanford</th>
<th>PredPat</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2</td>
<td>3</td>
<td>11</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>NYT-222</td>
<td>22</td>
<td>38</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Error Class</td>
<td>True Positives</td>
<td>False Positives</td>
<td>False Negatives</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>Wrong Relation [6]</td>
<td>No Impact</td>
<td>More</td>
<td>No Impact</td>
<td></td>
</tr>
<tr>
<td>Out of Scope</td>
<td>No Impact</td>
<td>No Impact</td>
<td>No Impact</td>
<td></td>
</tr>
</tbody>
</table>
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
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 from database 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

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


