The demand for fast and efficient methods to extract pivotal information encourages researchers towards IE tasks. Open Information Extraction (OIE) is the process of extracting relation tuples from text, targets to ease the process of identifying domain-independent relations in texts.

Example

"Barack Obama born August 4, 1961 in Hawaii served as the 44th President of the United States."

OIE relation triplets format: (argument 1, relation, argument 2):
(Barack Obama, BornIn-Year, August 4, 1961)
(Barack Obama, BornIn-Loc, Hawaii)
(Barack Obama, Served-as, President of the United States)

Challenges

1. **Uninformative Extractions:** Incorrect handling of relational phrases that results in leaving out crucial information.
2. **Incoherent Extractions:** Purposeless extractions that are derived from opaque relation phrases that the extractor fails to correctly identify.

Machine-Learning Classifiers

**TextRunner [1]:**
- 1st OIE system.
- Naïve Bayes model.
- Two-stage Technique that learns relation mapping rules and domain relations.

**WEOPOS [2]:**
- Employs a CRF extractor
- Utilizes Wikipedia to train data for their extractors.
- Automatic assembly of training examples from Wikipedia info box.

**WEOParse [2]:**
- Dependency path patterns acquired from from Wikipedia extraction.
- Parses handles complicated distance relationships.

**OLLIE [3]:**
- Bootstraps training set from seed tuples to learn pattern templates.
- Pattern templates determine the argument and the relation phrase.
- Learning component ensures that all the important information had been captured.
- Confidence function is trained to rule out non-factual extractions.

**ReVerb [4]:**
- Utilizes syntactic and lexical constraints.
- Confidence score is then allocated to extractions.

**RZA2 [5]:**
- Merges ReVerB with an argument enhance argument extraction.
- Identifies arguments by utilizing patterns.

**ExtrHec [6]:**
- Applies syntactic constraints as regular expressions.
- Multi-lingual (Spanish and English)

**LSOE [7]:**
- Implements lexical-syntactic patterns to POS-tagged texts to extract relation triples.
- Uses Generic patterns and Rule-based patterns.

**KraKen [8]:**
- Extracts N-ary facts.
- Examines fact completeness and correctness.

**DepOIE [9]:**
- Multi-lingual system (Portuguese, Spanish, English)
- Extract verb-based triples from Wikipedia.

**ClausIE [10]:**
- Identify clauses in an input sentence.
- Determines category of each clause to be consistent with the grammatical function of surrounding text.

**CSD-IE [11]:**
- Tree expressing the semantics is derived.
- Tree constituents are combined to form the contexts creating the phrase.

Future Trends and Conclusion

**Neural Networks**
- OIE paradigm that implements an encoder-decoder framework [12].
  - Employs recurrent neural network
  - Utilizes three-layer LSTM.

**Future Research:**
- Multilingual & N-ary extractions.
- Analysis supports Hand-Crafted approaches and the novel Neural Network approaches.

References