LOME: Large Ontology Multilingual Extraction

Patrick Xia*, Guanghui Qin*, Siddarth Vashishtha, Yunmo Chen, Tongfei Chen, Chandler May, Craig Harman, Kyle Rawlins, Aaron Steven White, Benjamin Van Durme

2021 EACL System Demonstrations
Background

• Recent advances in IE →
  • More detailed and fine-grained predictions
  • Document-level > Sentence-level
  • Multilingual

• Previous multilingual IE limitations:
  • separate subsystems per language (→)
  • limited scope of the task (Pan et al., 2017)
LOME: Large Ontology Multilingual Extraction

• A single, modularized multilingual system

• **Input:** document in any language (supported by XLM-R)

• **Output:**
  1. FrameNet parse (events and arguments)
  2. Coreference linking
  3. Entity typing (fine-grained, cluster-level)
  4. Temporal relation between events
The rabbit ate a carrot.
兔子喝了水
Translation: The rabbit drank water

Entities
the rabbit (animal)
Mentions: “The rabbit ate a carrot.”, “兔子喝了水”
carrot.
Mentions: “兔子喝了水”

Events
Before
Ingestion: 吃
Ingestor: The rabbit
Ingestible: a carrot

Ingestion: 吃
Ingestor: 兔子
Ingestible: a carrot

LOME Architecture

Multilingual coreference resolution

Fine-grained hierarchical entity typing

Temporal relation prediction

Third-party systems (e.g. relation extraction)
The rabbit ate a carrot.

**Mentions:** "The rabbit ate a carrot."

**Entities:**
- **the rabbit** (animal)
- **carrot**

**Ingestible:** a carrot

**Ingestion:** ate

**Ingestor:** The rabbit

**Before:**

**Translation:** The rabbit drank water

**Multilingual coreference resolution**

**Fine-grained hierarchical entity typing**

**Temporal relation prediction**

**Third-party systems (e.g. relation extraction)**
The rabbit ate a carrot.

Entities

**the rabbit** (animal)
- Mentions: “The rabbit ate a carrot.”, “兔子喝了水”
- Mentions: “兔子喝了水”

**carrot**
- Mentions: “The rabbit ate a carrot.”

Events

**Before**

- Ingestion: **ate**
- Ingestor: **The rabbit**
- Ingestible: **a carrot**

**Ingestion:** **喝了**
- Ingestor: **兔子**

Translation: *The rabbit drank water*
The rabbit ate a carrot.

"The rabbit drank water"

**Entities**

the rabbit (animal)

Mentions: “The rabbit ate a carrot.”, “兔子喝了水”

carrot

Mentions: “兔子喝了水”

**Events**

Before

Ingestion: 耶了
Ingestor: 兔子

Ingestible: a carrot

Ingestion: ate
Ingestor: The rabbit

**LOME Architecture**

1. Raw Text
2. Multilingual Tokenization
3. FrameNet Parser
   - The rabbit ate a carrot.
   - 兔子喝了水

**Output**

- Multilingual coreference resolution
- Fine-grained hierarchical entity typing
- Temporal relation prediction
- Third-party systems (e.g. relation extraction)
FrameNet Parser

- Goal: Find trigger spans and arguments, then label with FrameNet roles
- Model:
  - XLM-R encodings
  - BIO tagger to find trigger spans
  - Typing module for labeling spans
- Data: FrameNet v1.7
- SOTA on Frame ID
- First to report on full FrameNet parsing
The rabbit ate a carrot.

Mentions:
- "The rabbit ate a carrot."
- "兔子喝了水"

Translation: The rabbit drank water

Entities
- the rabbit (animal)
  - Mentions: "The rabbit ate a carrot.
  - Mentions: "兔子喝了水"

Events
- Before
  - Ingestion: "喝"
  - Ingestor: "兔子"
  - Ingestible: "a carrot"

FrameNet Parser

FrameNet Parser

Ingestion:
- "兔子喝了水"

Ingestor:
- "兔子"

Ingestible:
- "a carrot"

Multilingual coreference resolution

Fine-grained hierarchical entity typing

Temporal relation prediction

Third-party systems (e.g. relation extraction)

LOME Architecture

Raw Text

Multilingual Tokenization

FrameNet Parser

Multilingual coreference resolution

Fine-grained hierarchical entity typing

Temporal relation prediction

Third-party systems (e.g. relation extraction)
Coreference Resolution

- Goal: Determine which **given** mentions refer to the same entity
- Model: Modified Incremental Coreference Model *(Xia et al., 2020)* + XLM-R
- Data: OntoNotes 5.0, SemEval 2011 Task 1, Russian RuCor/AnCor
- Results: matches SpanBERT *(Joshi et al., 2020)* in English, comparable to RuBERT for Russian, sets neural baselines
Hierarchical Entity Typing

• Goal: associate entities (*mention clusters*) with fine-grained hierarchical entity types

• Model: Coarse-to-fine decoder (Chen et al., 2020) + XLM-R + *Borda Voting*

• Data: various (AIDA Ontologies, BBN, FIGER, UltraFine)
The rabbit ate a carrot.

<table>
<thead>
<tr>
<th>Entities</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>the rabbit</strong> (animal)</td>
<td>Before</td>
</tr>
</tbody>
</table>
| Mentions: “The rabbit ate a carrot.” “兔子吃了水” | Ingestion: **ate**
|                   | Ingestor: The rabbit |
|                   | Ingestible: a carrot |
Temporal Relation Prediction

• Goal: given two events, predict their temporal relation (e.g. before, overlap, etc)

• Model: Real-valued event pairs (Vashishtha et al., 2019) + XLM-R

• Data: TimeBank Dense, TempEval3, Chinese corpus

• Cross-lingual performance
The rabbit ate a carrot.

Ingestion: ate
Ingestor: The rabbit
Ingestible: a carrot

Third-party systems (e.g., relation extraction)
Applications

• TAC SM-KBP 2020 Task 1
  • Pipeline: GAIA (Li et al., 2020) → Coref → Entity Type → argument linking

• Schema inference tasks also needs relations
  • Attach OneIE (Lin et al., 2020) to output of LOME

LOME is modular: designed to be mix/matched with other systems
Website: [https://nlp.jhu.edu/demos/](https://nlp.jhu.edu/demos/)

- Available on [Docker Hub](https://nlp.jhu.edu/demos/)
- Concrete (Ferraro et al., 2014) format outputs
- [Web Demo](https://nlp.jhu.edu/demos/)
- Interactive output visualization
- Structured Frame/entity output
The rabbit ate a carrot. 兔子吃了胡萝卜。
Summary

• Single system for multilingual information extraction
  • Highly reliant on strong multilingual encoders and their cross-lingual transfer ability

• Available as a web demo and on Docker

Website: https://nlp.jhu.edu/demos/