

SCOPE and SCION

A Benchmark and an Auditable Reference Pipeline for Schema Induction and Fusion from Text

Miaobo Hu^{1, 2}, Xiaobo Guo¹, Shuhao Hu^{1, 2}, Bokun Wang^{1, 2}, Rui Chen^{1, 2}, Xin Wang¹, Jun Xiao³, Daren Zha¹

1 Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

2 School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

3 School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China

SCOPE = Schema Construction and Ontology-induction Pipeline Evaluation

SCION = Schema Construction and Induction with Ontology Normalization

TRAIN-TEXT-ONLY

AUDITABLE

SCHEMA FUSION

ICML 2026

Problem: the schema is usually assumed

MOTIVATION

What IE/KG systems need

- A KG stores instance facts, such as (Alice, works_for, OpenAI).
- A schema graph defines the allowed vocabulary, such as (Person, works_for, Organization).
- In practice, schema design and cross-source alignment are slow, expensive, and inconsistent.

fixed schema

standard IE

What we evaluate

- Standard IE benchmark: fixed schema -> score extracted instances.
- Our setting: raw train text -> induce and optionally fuse the schema itself.
- No validation or test evidence may enter the induction process.

raw train text

induced schema

SCOPE setting

SCOPE: a train-text-only benchmark

BENCHMARK

24 sources

15 relation extraction and 9 event extraction datasets.

Input

induction_texts.jsonl built from train documents only.

Target

Evaluation-only gold schema graphs in a unified typed-edge format.

RE edge

(head_type, relation, tail_type).

EE edge

(event_type, role, ARG) for within-event roles.

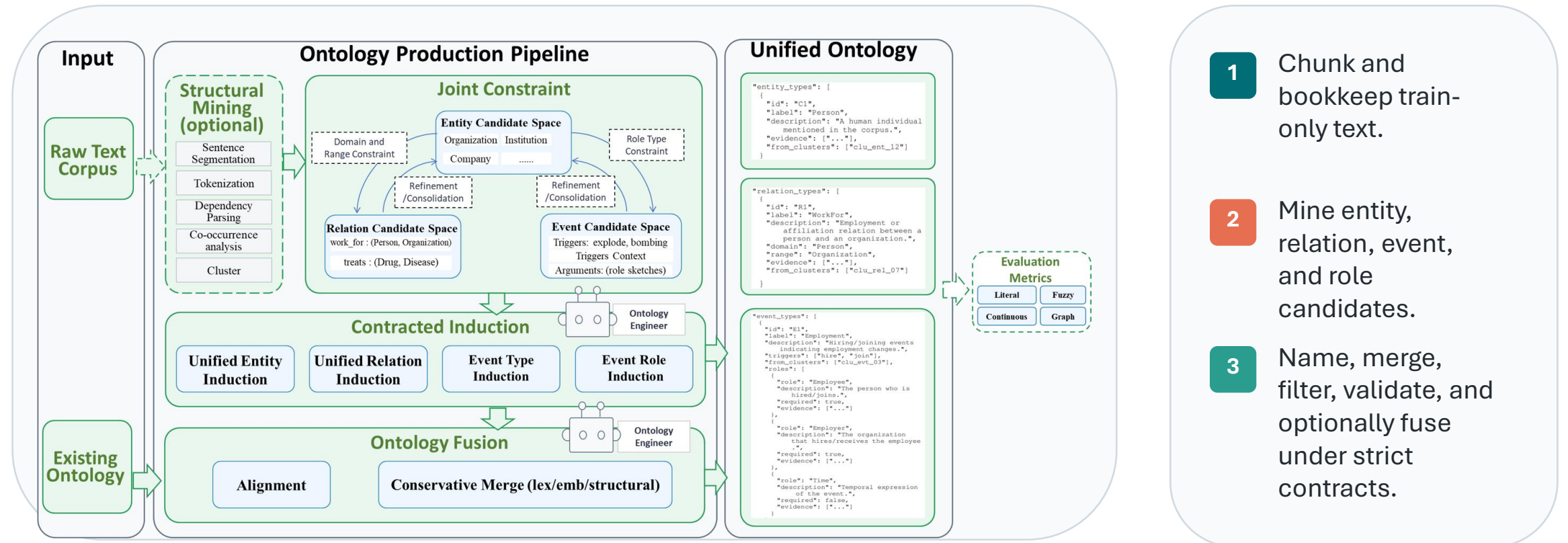
Metrics

Literal, Fuzzy, Continuous, and Graph schema-graph F1.

Gold schemas are evaluation-only; validation/test texts never provide candidates or evidence.

SCION: constrained schema engineering

METHOD

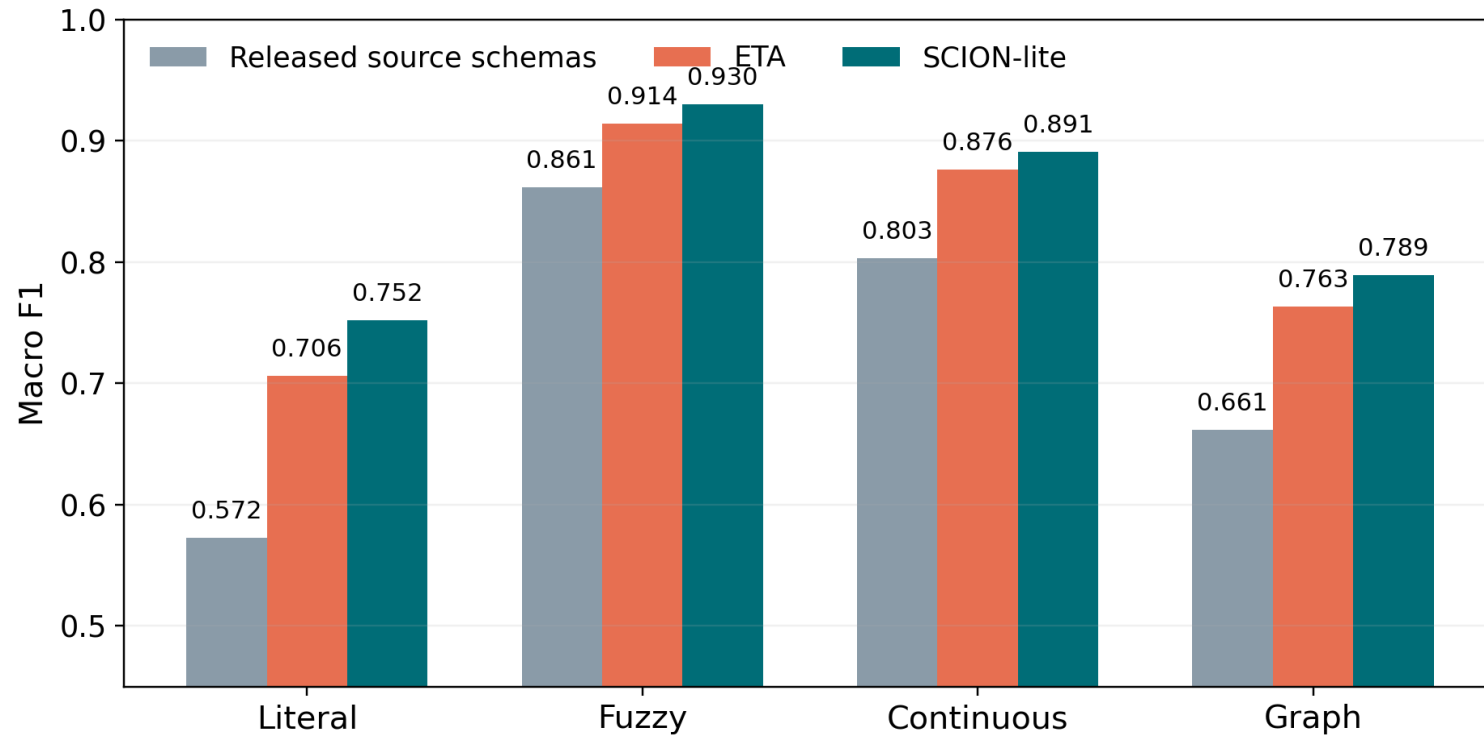


The schema engineer cannot invent arbitrary items; every output must link to candidate evidence.

Audit trail: parse success, fallback triggers, candidate retention, merge decisions, provenance.

Main result: higher schema-graph F1

RESULTS



Best under all four metrics

Literal, Fuzzy, Continuous, Graph

+0.0455 Literal F1 over ETA

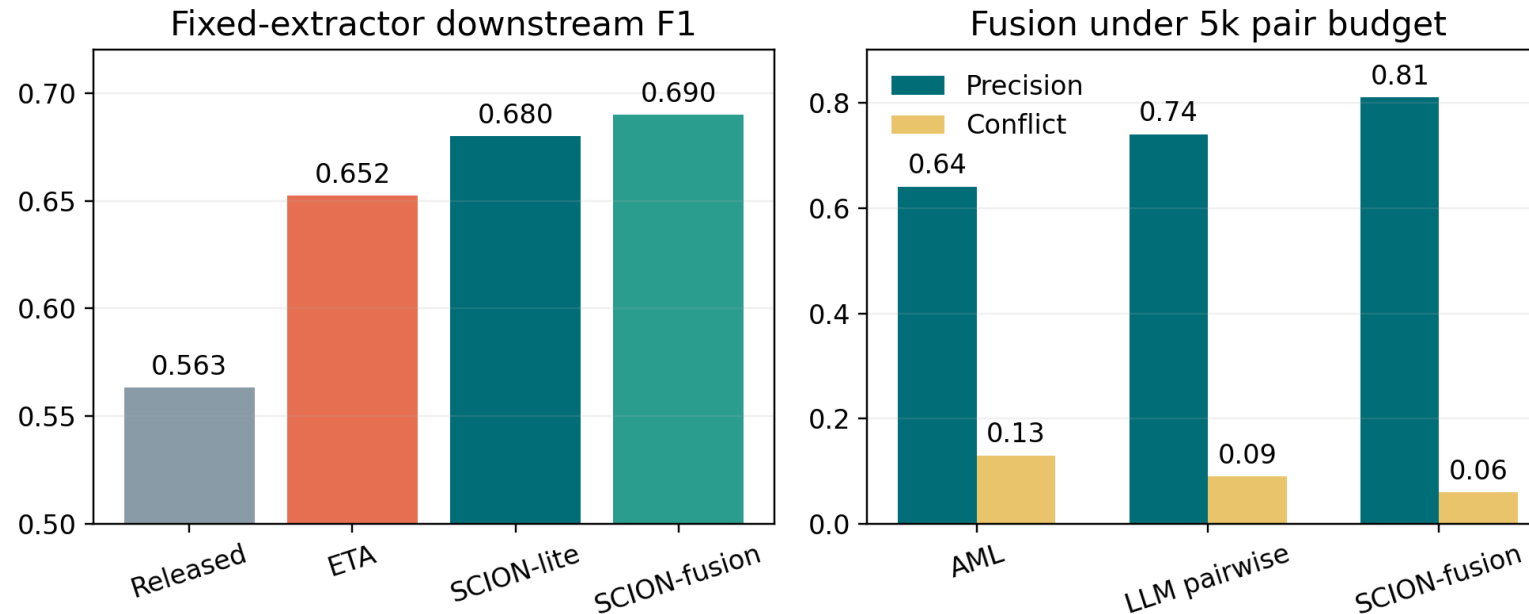
+0.0150 Continuous F1 over ETA

+0.0259 Graph F1 over ETA

SCION-full reaches 0.8257 Graph F1; SCION-RL (Qwen3-8B) reaches 0.8228.

Extraction flow and practical effect

WORKFLOW



Reachable target

SCION-lite stays above ETA on Continuous and Graph F1.

Fusion reliability

0.81 precision, 0.06 conflict under the shared 5k pair budget.

Efficiency

4 calls, 35.5s, \$0.12 per source for induction.

The induced schema also helps actual extraction when the extractor is fixed.

Takeaway

CLOSING

- 1 SCOPE makes corpus-to-schema induction and fusion measurable without gold-schema leakage.
- 2 SCION shows that candidate constraints plus JSON contracts produce stronger, auditable schemas.
- 3 The claims are about normalized typed-edge targets, not surpassing human ontology design.
- 4 Next: richer event links, larger fusion audits, and long-term schema maintenance under drift.

Thank you