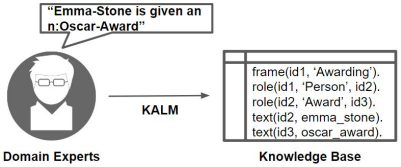
Authoring Knowledge via Natural Language

Yuheng Wang, Nathanael Payen, Paul Fodor, and Michael Kifer
Department of Computer Science, Stony Brook University

Stony Brook University
Computer Science

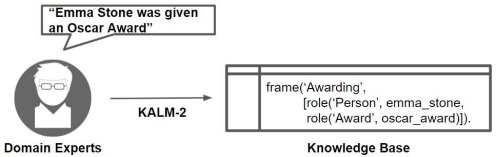
INTRODUCTION

The Knowledge Authoring Logic Machine[1] (KALM) enables domain experts who do not have knowledge representation skills to **write their domain knowledge into knowledge base** via Controlled Natural Language (CNL). In this way, domain experts are able to store their knowledge and reason with it.



However, by relying on CNL, KALM only takes **restricted syntactic forms** and **present tense** as input, which are too burdensome for users.

KALM-2 is proposed to solve these problems using a natural language parsing toolkit, **Stanza**[2].

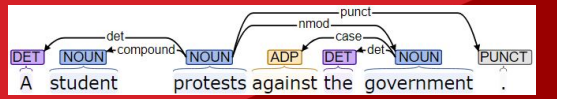


It is demonstrated that KALM-2 performs as well as the original KALM (**95.6%** accuracy).

PROBLEMS AND SOLUTIONS

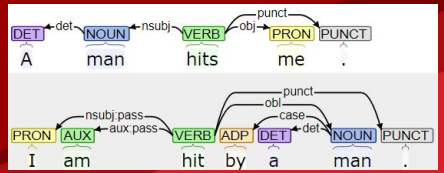
There are a bunch a problems emerging when migrating from CNL to NL using Stanza. The problems are-two fold: 1) **Stanza Model Issues** and, 2) **Stanza Semantic Mismatch**.

> **Stanza Model Issues** refers to the accuracy problems existing in the pipeline.



"Protests" is wrongly identified as a noun, which causes wrong dependencies. KALM-2 detects such errors and modifies them to be correct.

> **Stanza Semantic Mismatch** refers to the problems that equivalent sentences are represented differently.



Two sentences represent the same meaning but with different dependencies. Using **Para-parsing**, KALM-2 can merge them into the same representation.

EVALUATION

Metrics	Description	Results
FrSynC	All frames and roles are correct, and all role-fillers are disambiguated	239 (95.6%)
FrC	All frames and roles are correct	248 (99.2%)
PFrC	Some frames/ roles are correct	0
Wrong	Some frames/roles are wrong	2 (0.008%)
Total	All frames	250

REFERENCES

[1] Gao, Tiantian, Paul Fodor, and Michael Kifer. "Knowledge authoring for rule-based reasoning." OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Springer, Cham, 2018.
[2] Qi, Peng, et al. "Stanza: A Python natural language processing toolkit for many human languages." arXiv preprint arXiv:2003.07082 (2020).