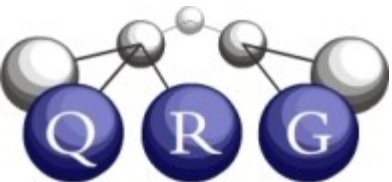# Combining Analogy with Language Models for Knowledge Extraction

Danilo N. Ribeiro

Kenneth D. Forbus

Northwestern University

AUTOMATED
KNOWLEDGE
BASE 2021 | irvine
CONSTRUCTION

Problem

- ## Knowledge base (KB) completion
  - Learn from text to populate exiting KB with new facts
  - Target common nouns (e.g., "dog" or "oxygen"), instead of named entities (e.g., "Joe Biden" or "Brazil")

- ## Challenges:
  - Schema and concepts defined by existing KB
    - not open domain
  - CycL representation
    - not simple triples
  - Few examples per relation
    - low resource learning
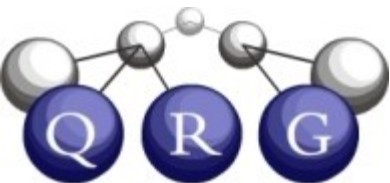
System Overview

- Combining Analogy and Language Models
  - Analogical Training
    - Able to learn from few training examples
  - Fact Scoring
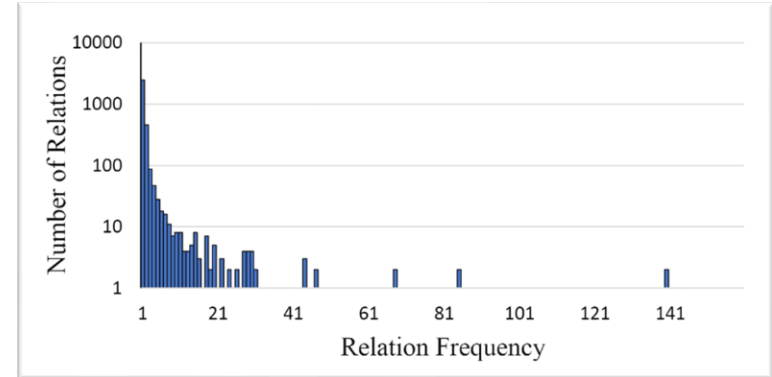   - Uses BERT to classify correctness of generated facts
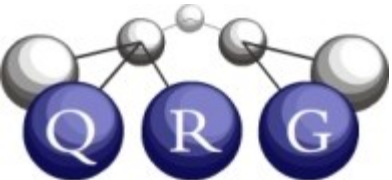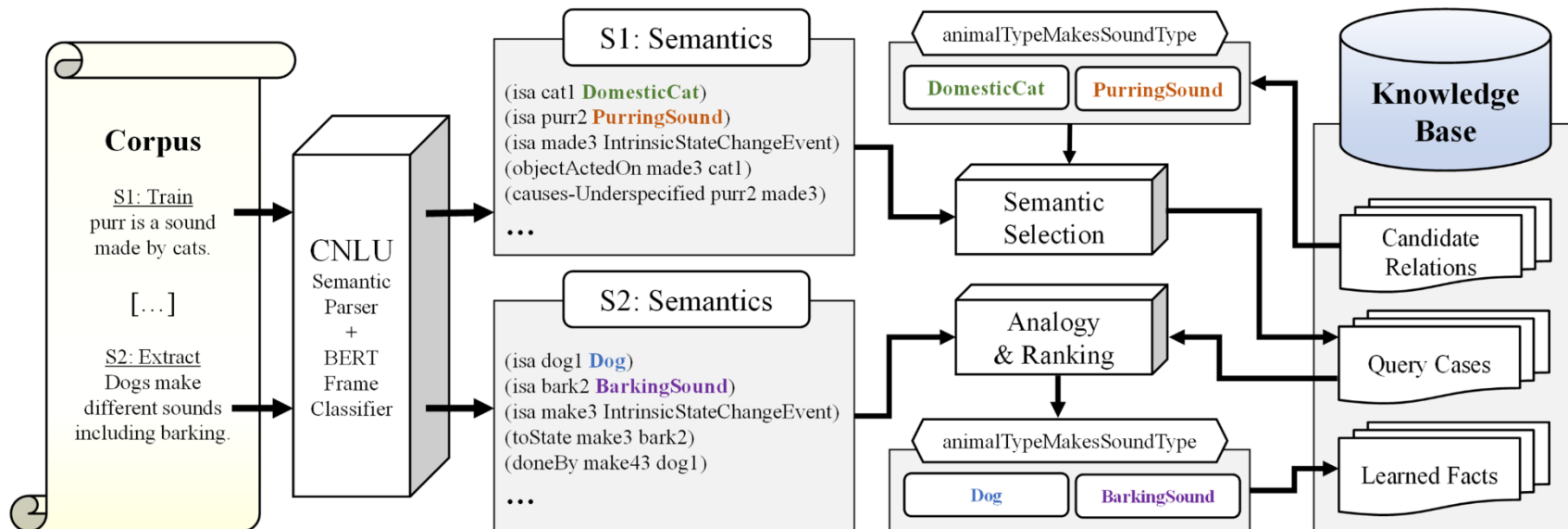
Data and Results

- ## Knowledge base (KB) completion

  - ### Knowledge Base: NextKB

    - 66,649 facts with
    - 3,745 distinct relations
    - Long tail distribution

  - ### Text Corpus: SE Wikipedia

    - 2,679 articles

- ## Baselines

  - ### Relation Extraction (CNN & BERT encoders)

  - ### Text-to-Text (T5-base)

Method	Estimated Precision
Relation Extraction (CNN) *	17,1%
Relation Extraction (BERT) *	20,8%
Text-to-Text (T5)	26.0%
Analogy (AKE)	45.7%
Analogy (AKE) + BERT fact classifier	**71.4%**