

# 森羅パネルディスカッション 「構造化知識を使った 言語処理応用」

①What?

③Why? ④How?

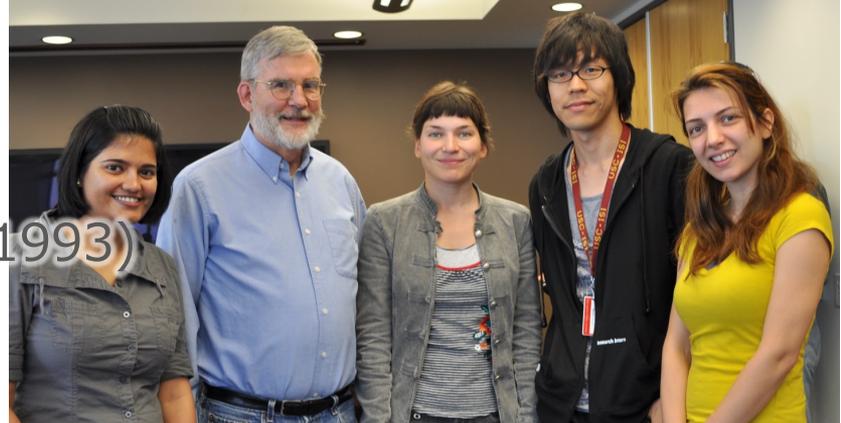
井之上 直也 (JAIST/RIKEN)

②Stance?

# What: 人の知識に基づいて組織化された世界知識

- **オントロジ**: 人が設計した知識 (概念や概念間の関係) の記述言語
  - 概念: モノ, コト
  - 関係: instance of, born in, agent, theme, purpose of, cause, ...
- **構造化知識**: オントロジに基づいて世界の知識を記述した情報
  - 森羅, WordNet, FrameNet, 格フレーム, CoceptNet, Wikidata, ATOMIC, ...
- **構築方法**: 計算機による自動構築 and/or 人手による整備
- **Wikidata の例**
  - Q2: Earth/地球, Q321: Milky Way/天の川銀河/, Q405: Moon/月
  - (Q2, *location*, Q321), (Q2, *child astronomical body*, Q405), ...

# Stance: 構造化知識推し



- Interpretation as Abduction (Hobbs 1993)  
との出会い (2010/D1の秋)
  - 言語理解  $\Leftrightarrow$  文に対する最小の説明の生成
- 構造化知識に基づく言語言語処理のための仮説推論エンジン (2011)
- 仮説推論に基づく含意関係認識、共参照解析 (2012, 2014)
- スクリプト知識の分散表現化 (2016)
- 論理構造付き論述データセット (2017, 2022)
- 半構造化知識に基づく説明可能な機械読解データセット (2020, 2021, 石井+2023)

# Why: 人とAIの共通基盤として

## 1. テキストに書かれない常識的知識の源をAIに教える手段として

- c.f. Reporting Bias 問題
- ダウンストリームの高精度化

## 2. AIと人のコミュニケーションのための言語として

- AI: 人の知っている知識に基づいて予測を説明する
  - 人: 知識の編集によりAIの非合理的な推論を修正する
- 
- 構造化知識ですべてが解決するわけではない
    - 「使いこなすための工夫」が必要

# How: 言語モデルとうまく付き合わせたい

	言語モデル	構造化知識 + 記号推論
構築の難易度	低	難
クエリの柔軟性	高 (自然言語)	難
ダウンストリーム タスクとの接続	易	難
精度/再現率	中/高	高/中
知識のソース	文章	文章+人
解釈性/説明性	低	高
知識の検証	難	易
更新 (削除, 修正)	難	易
他モーダルとの紐付け	難	易

## その他融合パラダイム:

- Language Model as Knowledge Base (Petroni 2019)
- Retrieval-augmented Language Model (Gua 2020)