A Unifying Approach

- **Bootstrap** from easiest extractable knowledge
- Conquer the long tail via **self-supervised learning**
- **Markov logic**: Knowledge representation / joint inference
- Govern self-supervised learning by a probabilistic model: domain knowledge + large-scale joint inference
- Scale up joint inference with **coarse-to-fine inference**
- Induce a **probabilistic ontology** from text for better generalization and tractable coarse-to-fine inference
- **Continuous learning**: Combine bootstrapping and crowdsourcing to continuously improve reading performance

A Success Story: USP

- USP [Poon & Domingos, EMNLP-09]
- Unsupervised, end-to-end machine reading system
- First approach for unsupervised semantic parsing
- Encoded in a few formulas in higher-order Markov logic
- Recursive extension of relational clustering
- OntoUSP = USP + Ontology Induction [Poon & Domingos, ACL-10]

Solution: Statistical Relational AI

- **End-to-End**: Raw text → knowledge → support end tasks (e.g., QA)
- **High quality**: Extract knowledge with high accuracy
- **Large-scale**: Scale up to the Web and be open to domains, genres, languages
- **Maximally autonomous**: Require minimal human effort
- **Continuous learning from experience**: Constantly integrate new information sources, and learn from user questions and feedback

Key Challenges

- **Uncertainty** + Long-tailed distribution of textual knowledge:
  - Variations for the same meaning
  - Kale contains calcium
  - Calcium is found in kale
  - Kale is rich in calcium
  - “Dark matter” in the natural language universe
  - Kale contains calcium + calcium prevent osteoporosis → kale prevents osteoporosis
- **Lack of labeled data** → Need unsupervised learning

Desiderata for Machine Reading systems

- **End-to-End**: Raw text → knowledge → support end tasks (e.g., QA)
- **High quality**: Extract knowledge with high accuracy
- **Large-scale**: Scale up to the Web and be open to domains, genres, languages
- **Maximally autonomous**: Require minimal human effort
- **Continuous learning from experience**: Constantly integrate new information sources, and learn from user questions and feedback

A Unifying Approach

- **Bootstrap** from easiest extractable knowledge
- Conquer the long tail via **self-supervised learning**
- **Markov logic**: Knowledge representation / joint inference
- Govern self-supervised learning by a probabilistic model: domain knowledge + large-scale joint inference
- Scale up joint inference with **coarse-to-fine inference**
- Induce a **probabilistic ontology** from text for better generalization and tractable coarse-to-fine inference
- **Continuous learning**: Combine bootstrapping and crowdsourcing to continuously improve reading performance

A Success Story: USP

- USP [Poon & Domingos, EMNLP-09]
- Unsupervised, end-to-end machine reading system
- First approach for unsupervised semantic parsing
- Encoded in a few formulas in higher-order Markov logic
- Recursive extension of relational clustering
- OntoUSP = USP + Ontology Induction [Poon & Domingos, ACL-10]