# Probabilistic Reasoning in the Semantic Web using Markov Logic

**Pedro Carvalho de Oliveira**
*pcoliv@student.dei.uc.pt*

MSc Thesis

*University of Coimbra*
*Faculty of Sciences and Technology*
*Department of Informatics Engineering*

*Knowledge and Intelligent Systems Laboratory*
*Cognitive and Media Systems Group*
*Centre of Informatics and Systems of the University of Coimbra*

# Overview

- **Semantic Web**
- **Markov Logic**
- **Markov Logic + Semantic Web**
  - Formulas
  - Weights
    - Annotated
    - Learned Using Individuals
    - Learning Individuals/Probabilities
- **System**
- **Conclusions**

# Semantic Web

- **Semantic Web**
  - Bring structure to the meaningful content of Web pages
  - Knowledge represented by ontologies
    - DAML+OIL, OWL, OWL2
  - None of them provide means of dealing with <u>uncertainty</u>
    - Probabilistic reasoning tries to solve this problem through the use of probability theory

# Objectives

- **Study mechanisms to perform probabilistic reasoning in the Semantic Web**
  - Markov Logic = First-order Logic + Markov Networks
- **Develop a system that provides a Semantic Web interface to Markov Logic reasoning and learning capabilities**

# Markov Logic

- **Markov Logic**
  - First-order logic and Markov networks in the same representation
  - A world that violates a formula is not invalid, but less probable
  - Attaching weights to first-order logic formulas
    - The higher the weight, the bigger is the difference between a world that satisfies the formula and one that does not

$$P(X = x) = \frac{1}{Z} exp\left(\sum_{i=1}^{F} w_i n_i(x)\right)$$

# Markov Logic + Semantic Web

- **Markov Logic = Formulas+Weights**
- **Formulas**
  - Semantic Web ontology languages like OWL2 follow a model-theoretic semantics

| OWL2 Axiom | First-order logic formula |
|---|---|
| $SubClassOf(CE_1, CE_2)$ | $\forall x : CE_1(x) \Rightarrow CE_2(x)$ |
| $TransitiveProperty(OPE)$ | $\forall x, y, z : OPE(x, y) \land OPE(y, z) \Rightarrow OPE(x, z)$ |
| $ClassAssertion(CE, a)$ | $CE(a)$ |

# Markov Logic + Semantic Web

- **Weights**
  - Provided
  - Learned Using Individuals
  - Learning Individuals/Probabilities

# Markov Logic + Semantic Web

- **Provided Weights**
  - Ontology axioms can be annotated with a value that can be used as weight
  - Problem: Probabilities ≠ Weights
    - Discriminative weight learning

$$\frac{\partial}{\partial w_i} log\, P_w(y|x) = n_i(x, y) - E_w[n_i(x, y)]$$

$$n_i(x, y) = count(i) * p_i$$

# Markov Logic + Semantic Web

- **Provided Weights**
  - Applications
    - User-created probabilistic ontologies
      - E.g., Body Gestures ontology
        - HeadScratch(A) → P(Recalling(A)) = 0.96
    - Ontology learning
      - E.g., Taxonomies automatically learned from web search engines
        - Acetone(A) → P(Ethanol(A)) = 0.91

# Markov Logic + Semantic Web

- **Learned weights**
  - Weights can be learned generatively or discriminatively through example data
  - Applications
    - Using a financial ontology, predict the probability of a loan being problematic
    - Semantic Web Social Network analysis
      - Link prediction, link-based classification (64%) and cluster analysis
    - Pure machine learning tasks
      - Classification of mushrooms (94%) and Titanic passengers (75%)

# Markov Logic + Semantic Web

- **Learn Individuals**
  - Ontology population
    - Given a source ontology and a corpus, extract individuals of that ontology from the corpus, with their class and property assertions

**Animals** such as **dogs** and **cats**
**Lions** are **animals**
**Lions** are the **predators** of **zebras**

→

**Animal(Dog)**
**Animal(Cat)**
**Animal(Lion)**
**Predator(Lion,Zebra)**

# Markov Logic + Semantic Web

- **Learn Individuals**
  - Applications
    - Automatically populate ontology about diseases and their symptoms and cluster diseases by their symptoms
      - 140 diseases (66-70%), 459 symptoms (63%)
      - Cluster examples
        - Depression (migraines, anxiety, alzheimers, bipolar disorder)
        - Respiratory system (lung cancer, tuberculosis, asthma, pneumonia, diphteria)

# Markov Logic + Semantic Web

- **Learn Probabilities**
  - Use semantic similarity techniques

$$Pointwise\ Mutual\ Information\ (P,Q) = \log_2\left(\frac{H(P,Q)}{H(P)H(Q)} * N\right)$$

  - Given *(P,R,Q)*
    - H(P) = “PR*”
    - H(Q) = “*RQ”
    - H(P,Q) = “PRQ”

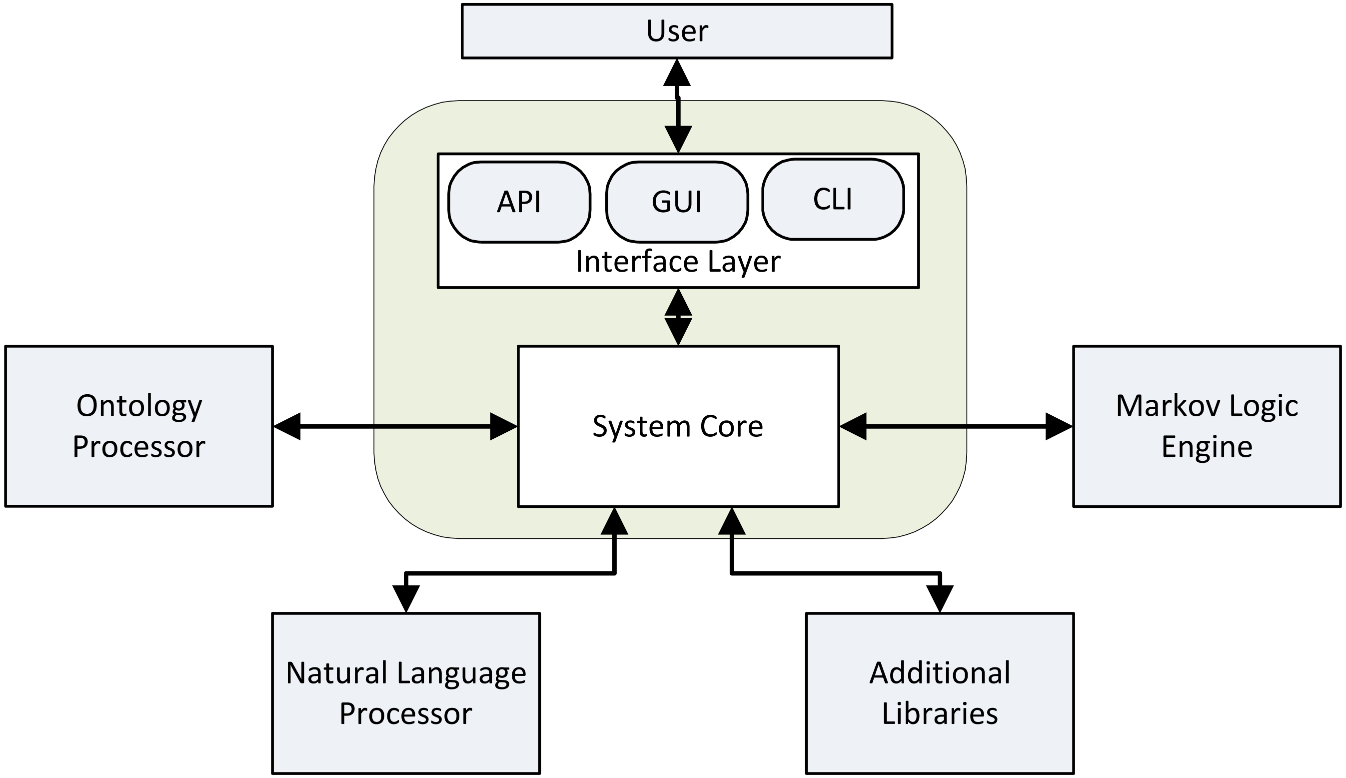
| | Query | Result Count |
|---|---|---|
| **H(P)** | “dog is a *” | 287,000 |
| **H(Q)** | “* is a pet” | 182,000 |
| **H(P,Q)** | “dog is a pet” | 785 |

# Markov Logic + Semantic Web

- **Learn Probabilities**
  - Applications
    - Infer the uncertainty of automatically learned taxonomies
      - E.g.
        - P(Animal(Dog)) = 82%
        - P(Dog(Cat)) = 9%

# System

- **Incerto**
  - http://code.google.com/p/incerto (LGPL)



# Conclusions

- **Contributions**
  - Applied Statistical Relational Learning techniques to Semantic Web reasoning
  - Studied not only reasoning under uncertainty in the Semantic Web, but also how to learn this uncertainty
  - Unlike other approaches, use undirected probabilistic models
  - New method to transform probabilities in Markov logic weights

# Conclusions

- **Contributions**
  - New method to learn the probabilities of OWL2 axioms using a web search engine
  - Useful for many interesting tasks
    - Ontology learning, reasoning, mapping, refining

# Conclusions

- **Publications**
  - Pedro Oliveira, Paulo Gomes: "Instance-based Probabilistic Reasoning in the Semantic Web", Poster at the ***18th International World Wide Web Conference***, Madrid, Spain, April 2009
    - AR: 31%
  - Pedro Oliveira, Paulo Gomes: "Learning and Reasoning about Uncertainty in the Semantic Web", ***14th Portuguese Conference on Artificial Intelligence***, Aveiro, Portugal, October 2009
    - AR: 66%

# Conclusions

- **Future Work**
  - More experimentation
  - Other ways to automatically learn the uncertainty of ontology axioms
  - Study reasoning/learning with transitivity and cardinality restrictions

# Probabilistic Reasoning in the Semantic Web using Markov Logic

**Pedro Carvalho de Oliveira**
***pcoliv@student.dei.uc.pt***

MSc Thesis

# Thank You!