

A PROBABILISTIC KNOWLEDGE BASE SYSTEM

Daisy Zhe Wang

CISE, University of Florida

Data Science Lab, Database Research Group

01/07/2013

Knowledge Bases





















- A *knowledge base* is a collection of entity, facts, relationships that conforms with a certain data model.
- A knowledge base helps machine understand humans, languages, and the world.
- Examples 1: Google Knowledge Graph



Knowledge Bases

- A *knowledge base* is a collection of entity, facts, relationships that conforms with a certain data model.
- A knowledge base helps machine understand humans, languages, and the world.
- Examples 2: NELL (Never Ending Language Learner)

Recently-Learned Facts twitter Refresh

instance	iteration	date learned	confidence	
albert_iii is a monarch	623	13-aug-2012	99.9	 
kaseberg_creek is a river	620	01-aug-2012	99.1	 
senior_e_commerce_intelligence_consultant is a profession	622	09-aug-2012	98.4	 
deonna_anderson is a journalist	620	01-aug-2012	97.3	 
rose_city_comic_con is a convention	620	01-aug-2012	98.6	 
candidas is a drug worked_on by merck	622	09-aug-2012	93.8	 
softball is a sport with fans in the country new_zealand_wellington	623	13-aug-2012	100.0	 
mammals is an animal that eats berries	622	09-aug-2012	100.0	 
wccu is a TV affiliate of the network fox	625	21-aug-2012	100.0	 
cats is an animal that can develop respiratory_infection	622	09-aug-2012	100.0	 

Knowledge Bases

- A *knowledge base* is a collection of entity, facts, relationships that conforms with a certain data model.
- A knowledge base helps machine understand humans, languages, and the world.
- Examples: Google Knowledge Graph, NELL, DBpedia...
- Applications:
 - Search Engines
 - Conversation Systems (Chatterbots & Turing Test)
 - Electronic Medical Notes Analysis
 - Human Belief Systems

Life of A Knowledge Base (KB)

- A *knowledge base system* is a special kind of database management system for knowledge base management.
- *KB extraction*: knowledge extraction using statistical models in NLP/ML literature
- *KB expansion*: knowledge inference using deductive rules and knowledge with uncertainty
- *KB evolution*: knowledge update given new evidence and new data sources
- *KB integration*: knowledge integration from multiple sources (e.g., human, data sets, models)

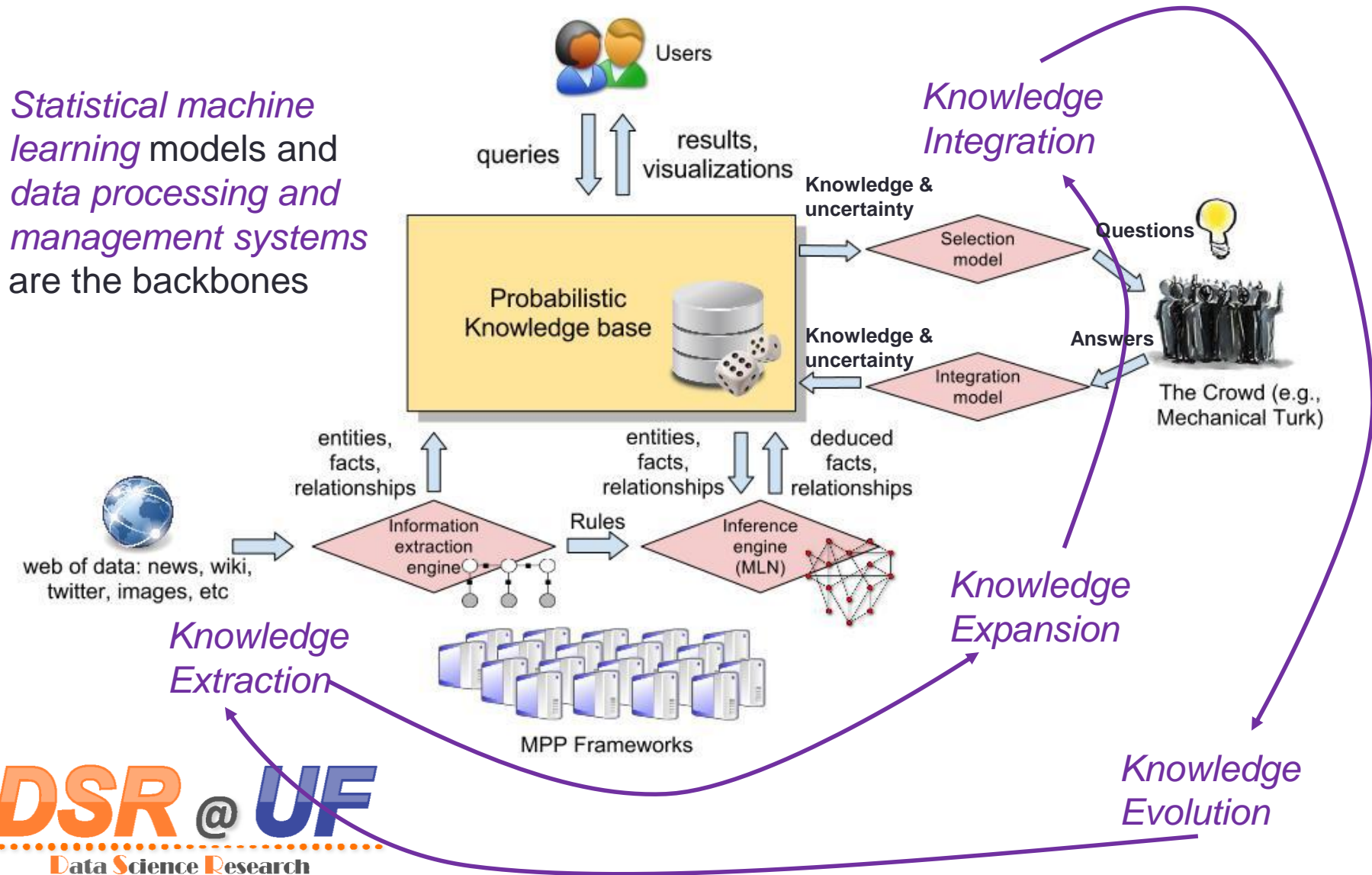
Uncertainty Management

- Where does Uncertainty come from?
 - Inherent in the state-of-the-art NLP results
 - Incorrect, conflicting data sources
 - Derived facts and query results
- Uncertainty vs. MAP/majority voting
- Uncertainty vs. NULL

- *Probability Theory* should lay the foundation for uncertainty management → *Probabilistic Knowledge Base System*

Probabilistic Knowledge Base System

Statistical machine learning models and data processing and management systems are the backbones



Challenges & Projects

- Probabilistic Database for Knowledge Extraction → *BayesStore* (2006-2011) [VLDB08, VLDB10, ICDE10, SIGMOD11]
- Deduction over Probabilistic KB *ProbKB*
 - Scalability, Efficient Updates
- Uncertain Knowledge Integration *CAMeL-DB*
 - Data sources (e.g., datasets, turkers, models)
 - Beliefs and Distributions
- Query-Driven/Smart Filtering *Archer*
 - Ad-hoc queries over static dataset
 - Continuous queries over streaming dataset
- Scalable NLP over MPP frameworks *MADLib-Text*