Advanced Decision Analytics via Deep Reasoning on Diverse Data: For Health Care and More

Benjamin Grosof
Coherent Knowledge*

Presentation (35-min.) at
DecisionCAMP 2016**,
July 8, 2016,
Co-located with RuleML-2016 at Stony Brook University, USA

Slides also by Janine Bloomfield, Coherent Knowledge

* Web: http://coherentknowledge.com
Email: *firstname dot lastname* at coherentknowledge.com

** Web: https://sites.google.com/site/ruleml2016/decisioncamp

Deep Reasoning for Advanced Analytics

Bio

- AI researcher, turned entrepreneur
- CTO, CEO, Co-Founder, of Coherent Knowledge
  - AI software platform component startup
- Previously:
  - Directed advanced AI research program for Paul Allen
    - Developed Rulelog KRR theory, algorithms, UI approach
  - MIT Sloan professor and DARPA PI
    - Co-Founder of RuleML, key contributor to W3C OWL-RL and RIF standards
  - IBM Research, creator IBM Common Rules
    - 1st successful semantic rules product in industry
  - Stanford AI PhD, combining ML with logical and probabilistic reasoning
- http://www.linkedin.com/in/benjamingrosof
- http://benjamingrosof.com

Coherent Knowledge

Outline

- Intro: Textual Rulelog is a next step on semantic approach to analytics for “complex” knowledge
- Case Study with Demo
  - Health care: treatment guidance
  - Decision automation + query answering with full explanation
- Technology for Textual Rulelog: Coherent’s Ergo
  - Human-machine logic: combining English and logic syntax
  - Virtualization of diverse knowledge sources, via “connectors”
- Discussion & Conclusions

Problem: Analytics for *Complex* Knowledge

Examples: policies, regulations, contracts; terminology mappings; science, causality

Existing *Non-Semantic* Technologies tend to be:

- Shallow
- Siloed
- Costly, and Slow
- Patchily automated
- Opaque
- Inaccurate
- End users not empowered to modify

Based on:

- *Conventional programming languages*
- *Production/ECA rules*
- *Prolog*

Coherent Knowledge™

The Semantic Approach

- Modeling, declaratively, rather than programming
- E.g., via DMN-based Decision Tables
  - That's "a first step"
- Benefits:
  - Greater integration and reusability
  - More transparent, i.e., explainable
  - Easier to modify, end users* more empowered
  - More cost-effective and agile

* esp. subject matter experts (SMEs)

Ergo is a Next Step on Semantic

- Compared to decision tables:
  - Deeper in reasoning & knowledge
    - Support many-step inferencing
    - Model complex sentences with high fidelity, via high expressiveness, e.g., higher-order, existentials
    - Map to/from natural language
    - Map between ontologies, schemas, terminologies
    - Principled defeasibility (exceptions)
  - Fuller, more understandable explanations
  - Greater scope of automation

➔ Extends the benefits of the semantic approach

Based on Rulelog

Rulelog summary:

- *Semantic rules and ontologies*
- *Very flexibly expressive, yet querying is poly-time*

Coherent Knowledge

Outline

- Intro: Textual Rulelog is a next step on semantic approach to analytics for “complex” knowledge
- Case Study with Demo
  - Health care: treatment guidance
  - Decision automation + query answering with full explanation
- Technology for Textual Rulelog: Coherent’s Ergo
  - Human-machine logic: combining English and logic syntax
  - Virtualization of diverse knowledge sources, via “connectors”
- Discussion & Conclusions

Health Care Case Study: Task Domain

- Task: Treatment Guidance for
  - Delivery of care, e.g., by medical staff or self-service
  - Insurance
  - Oversight of quality (e.g., “care measurement”)
- Guidance takes form of policies
  - Portions are often based closely on clinical studies
  - Top ~100 diseases have “protocols” written up in considerable detail

Kinds of Domain Knowledge & Reasoning

- Knowledge & reasoning about:
  - Patient characteristics and history
  - Symptoms
  - Diseases and diagnoses
  - Drug treatments
  - Non-drug treatments
  - Medical tests
  - Intended effects
  - Side effects
  - Interactions between treatments, e.g., drug-drug
  - Risks
  - Interactions between risks; aggravation and severity of risks

Challenges & Requirements

- Challenge: personalization
- Patients undergo multiple diseases and treatments, but protocols are developed, based on clinical studies, for
  - One disease (e.g., diagnosis) at a time
  - One treatment (e.g., drug) at a time
- Requirements, both beforehand and post-play, for
  - Correctness / competence
    - Maximize benefit to patient
    - Minimize harm to patient, incl. avoid potential treatment errors
    - Minimize costs
  - Verifiability, therefore
  - Explainability to:
    - Medical staff performing care delivery – e.g., combat “alert fatigue”
    - Patients – e.g., improve compliance by knowing why to avoid an easy-to-obtain drug
    - Insurers
    - Oversight staff, incl. for audits

Treatment Scenario

- A busy intern encounters an elderly woman in a rehabilitation facility complaining of knee pain.
- What treatment should be given -- or not given – and why?
- EHR records show:
  - The elderly woman is currently taking Coumadin to treat the pre-existing condition of atrial fibrillation which increases the risk of blood clot and stroke.

- *Automatically gives both alerts and educates.*

DEMO GOES HERE

Outline

- Intro: Textual Rulelog is a next step on semantic approach to analytics for "complex" knowledge
- Case Study with Demo
  - Health care: treatment guidance
  - Decision automation + query answering with full explanation
- Technology for Textual Rulelog: Coherent's Ergo
  - Human-machine logic: combining English and logic syntax
  - Virtualization of diverse knowledge sources, via "connectors"
- Discussion & Conclusions

Textual Rulelog Approach in *ERGO*

- Rulelog knowledge representation and reasoning (KRR).
- + Natural language processing
  - Maps text syntax <- -> logic syntax, using logic
- + Explanations that are fully detailed, SME-understandable
- + Connectors to most kinds of enterprise knowledge:
  - Relational DB
  - Graph DB
  - Spreadsheets
  - …

Analyze

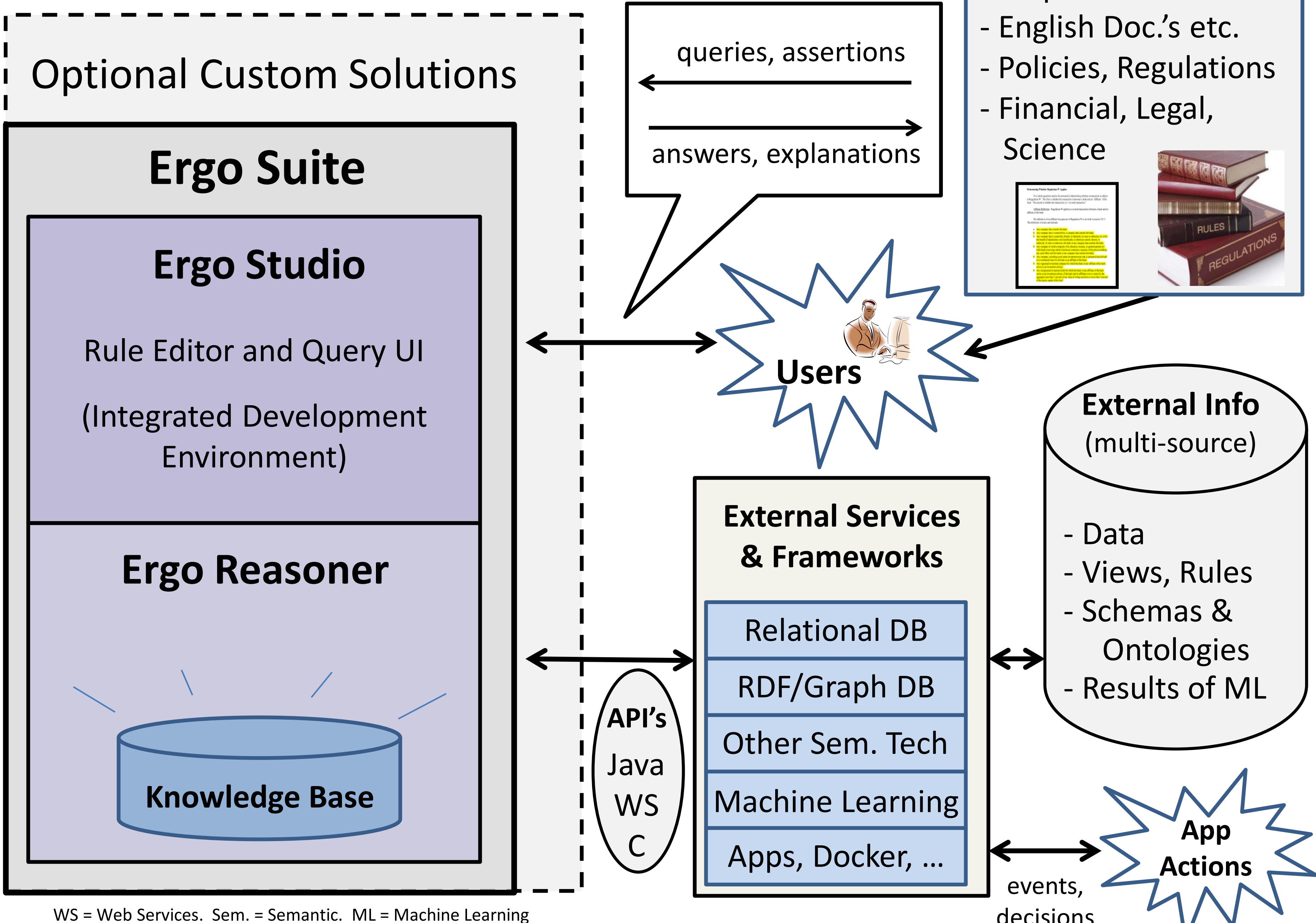
Explain

Make Decisions

Monitor & Alert

Answer Questions

Ergo Architecture



WS = Web Services. Sem. = Semantic. ML = Machine Learning

Textual Rulelog is Ergo's KRR basis

- *Rulelog*: a major research advance in KRR theory & algorithms, which culminated in 2012
- Ergo is the most complete & optimized implementation available of Rulelog
- Rulelog features very high/flexible expressiveness: logical chaining, higher-order, general quantified formulas, defeasibility/exceptions, provenance, probabilistic, restraint bounded rationality, and more
- Yet Rulelog scales well: reasoning is polynomial time (as in databases)

- *Textual* Rulelog extends Rulelog with natural language processing (NLP)
  - Logic itself is utilized to map between logic syntax and English syntax
- ErgoText templates aid knowledge authoring and explanation generation

- *More background:* Rulelog adds "full meta" expressiveness to *Datalog*
  - Datalog is the logic of databases, business rule systems (production/ECA/Prolog), semantic web ontologies, and earlier-generation semantic web rules (e.g., SWRL and RIF-BLD)
  - Rulelog extends also declarative logic programs (LP)

Coherent Knowledge

Ergo Suite: Reasoner, Studio, Connectors

- Ergo Reasoner has sophisticated algorithms & data structures
  - Smart cacheing with dependency-aware updating. Leverages LP & DBMS techniques.
  - Transformation, compilation, reordering, indexing, modularization, dependency/loop analysis, performance monitoring/analysis, pausing, virtual machine, programming kernel, external import/querying
  - Java API. Other interfaces: command line, web, C.
  - <u>Scales</u> well: Millions of sentences on 1 processor; Trillions on distributed nodes
- Ergo Studio is a graphical Integrated Development Environment
  - Interactive editing, querying, explanation, visualization of knowledge
  - Fast edit-test loop with award-winning advanced knowledge debugging/monitoring
- Ergo Connectors federate knowledge & reasoning
  - Import/query dynamically via: SPARQL, OWL, RDF; SQL; DSV; XML; and more
  - Federation distributes reasoning (i.e., its processing) across multiple nodes
- Open, standards-based approach; a portion is open source
  - Rulelog is draft industry standard from RuleML (submission to W3C & Oasis)

Concept: Humagic Knowledge

- Humagic = human-machine logic
- A humagic KB consists of a set of linked sentences
  - Assertions, queries, conclusions (answers & explanations)
- NL-syntax sentence may have 1 or more logic-syntax sentences associated with it
  - E.g., that encode it, or give its provenance
- Logic-syntax sentence may have 1 or more NL-syntax sentences associated with it
  - E.g., that result from text generation on it
- Other sentences can be in a mix of NL-syntax and logic-syntax
  - ErgoText: templates used for text interpretation and text generation

Textual extension of Rulelog (I)

- Leverage Rulelog to much more simply and closely map between natural language (NL) and logic
- Textual terminology
  - English phrase ↔ logical term in Rulelog
  - English word ↔ logical functor in Rulelog
  - Leverages the <u>higher-order</u> feature of Rulelog
- Textual templates

ErgoText

- ErgoText:

\(The proposed transaction ?Id by ?Bank with ?Affiliate of $?Amount is a RegW covered transaction\)

- ErgoText Template:

```
template(headbody,
            \(The proposed transaction ?Id by ?Bank with ?Affiliate of $?Amount
              is a RegW covered transaction\),

           covered(proposed(transaction))(by(?Bank))(with(?Affiliate))
             (of(amount(?Amount)))(having(id(?Id)))
           ).
```

- The templates are self-documenting

Coherent Knowledge

Textual Rulelog (II)

- Almost any NL sentence can be represented as a logical sentence
  - Leverages the logical quantifiers feature of Rulelog
  - Ex.: “each large company has some talented CEO”
    - forall(?x)^((?x \isa \(large company\)) ==>
      exists(?y)^((\?x has ?y\) \and
      (?y \isa \(talented CEO\)))).

DEMO Cont.'d GOES HERE

Concept: Virtual Rulelog

- Ergo orchestrates overall federated reasoning by sub-goaling dynamically

- A variety of other structured information systems are virtualized as Rulelog via Ergo federation connectors, which import/query and translate

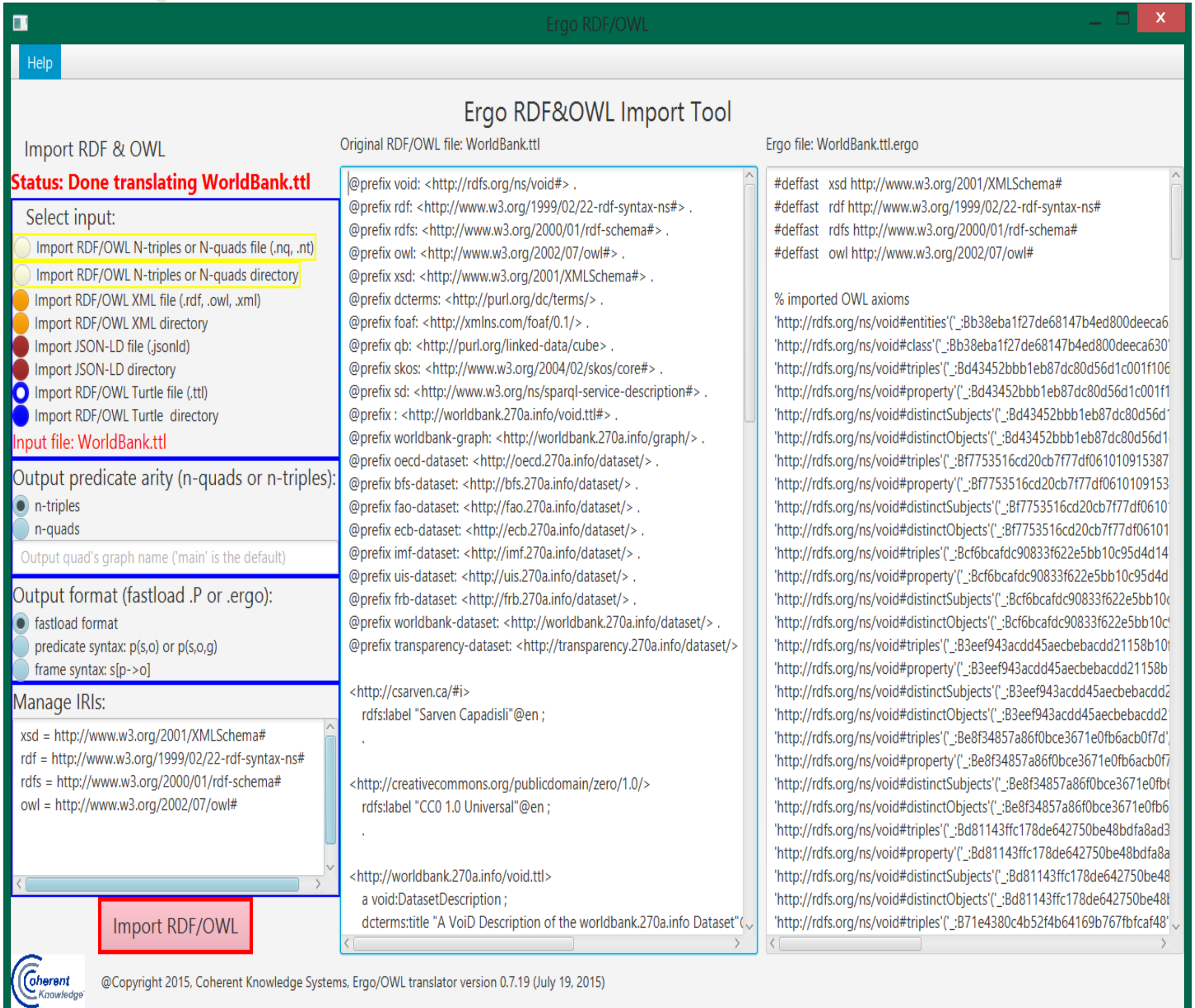
Coherent Knowledge

Kinds of Virtual Ergo

- Graph databases: via SPARQL/RDF connector
  - Description logic ontologies: via OWL connector
- Relational databases: via SQL connector
- Spreadsheets and web logs: via DSV connector
- Web services: via XML connector (JSON is under dev)
- Extensible to almost any kind of (semi-)structured info
  - E.g., Machine Learning (ML) and NLP systems
    - Represent prob(content_sentence, lower_bound, upper_bound, confidence_level, statistical_procedure) as an Ergo sentence
  - E.g., legacy applications in Java
    - Get method is treated like a query

Importing RDF & OWL knowledge into Ergo

- Screenshot of Ergo OWL connector part of Ergo Studio



Translates RDF & OWL to Ergo

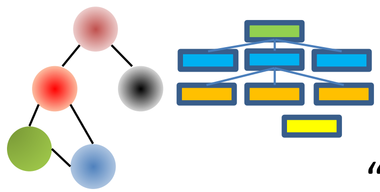
Define IRIs in Ergo Studio

N-triples and N-quads

RDF/OWL XML, JSON-LD, or Turtle as input. Predicate or Frame syntax output.

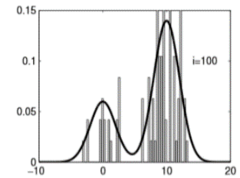
Actively Reason over Today's Gamut of Knowledge

Graph DB
& Semantic tech



Relational DB

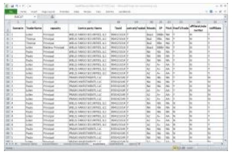
Machine
Learning



$\text{Prob}_N(x) = \Sigma_{i=1..N}\ \delta_{si}(x)\ /\ N$

Domain Apps
& Legacy

Spreadsheets



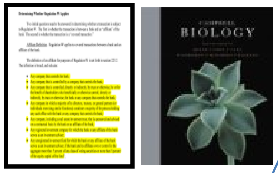
"Business Rules"

$\text{tier}(X,1) \wedge \text{supply}(Y,X) \Rightarrow \text{tier}(Y,2)$

Probabilistic engines

$p(H|C) = \Pi\ p(x_i|Y_j)\ /\ \Sigma_C\ \Pi\ p(x_i|Y_j)$

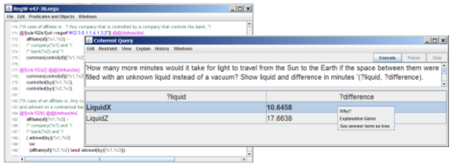
Text & Natural
Language processing



**External Info
& Services**

**Application
Actions**

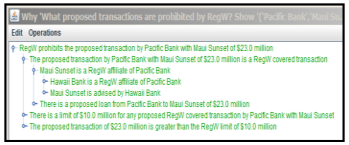
**Queries
Assertions
Edits**



ERGO

KB
Libraries

**Answers, Views
Decisions, Alerts
Explanations**



Other Ergo Features

- Advanced knowledge base debugging: pause/resume, performance monitor, analysis of attempted infinite loops ("terminyzer")
- Probabilistic reasoning, e.g., evidential
- Longer-term directions, already under dev:
  – Semantically augmented NL parsing, for authoring
  – Optimize restricted cases of probabilistic
  – Complement ML via:
    - feed derived data to ML, query ML, supply features to ML

Outline

- Intro: Textual Rulelog is a next step on semantic approach to analytics for "complex" knowledge
- Case Study with Demo
  - Health care: treatment guidance
  - Decision automation + query answering with full explanation
- Technology for Textual Rulelog: Coherent's Ergo
  - Human-machine logic: combining English and logic syntax
  - Virtualization of diverse knowledge sources, via "connectors"
- Discussion & Conclusions

Other Case Studies

- Financial regulatory compliance decisions: with databases/ontologies
- Defense intelligence analysis: with text extraction, databases/ontologies
- Personalized tutoring in continuing/higher ed: answering science questions
- E-commerce marketing: with product databases/ontologies, promotion/pricing policies

Lessons Learned from Case Studies

Customers in these multiple domains benefited from:

- Agility: Flexibility and ease of authoring, fast updating
- High accuracy and transparency
  - Explanations and provenance
  - Lower risk of non-compliance or confusion
- More Cost Effectiveness – less labor, SMEs in closer loop
- Leveraging investment in semantic tech: RDF, SPARQL, OWL

Thank you.

Deep Reasoning for Advanced Analytics

http://coherentknowledge.com

Coherent Knowledge: Company Overview

- Fundamentally new kind of logic/rule based AI software platform component for advanced analytics: flexible deep reasoning + natural language processing
- Radical business benefits: accuracy/competence, cost, agility, transparency
- Company offers: software product Ergo + professional services for custom solution dev
  - Capabilities: engine + development environment, for executable knowledge bases (logic/rules) embedded in apps
- World-class founder team: created many industry-leading logic systems & standards
  - Extensive experience applying logic systems to financial, regulation/policy, and other domains
  - Former/current professors at Stony Brook University and MIT

Benjamin Grosof, PhD
CTO & CEO
Prof., MIT Sloan. DARPA PI.
Advanced AI Prog. Mger.
for Paul Allen.
Creator, IBM Common Rules.

Michael Kifer, PhD
Principal Engineer
Prof., Stonybrook Univ.
Winner, 3 ACM & ALP
test-of-time research awards.

Theresa Swift, PhD
Principal Engineer
Co-lead dev, mission-
critical rules system,
US Customs.
Co-Architect, XSB Prolog.

Paul Fodor, PhD
Senior Engineer
Prof., Stonybrook Univ.
IBM Watson team.

Janine Bloomfield, PhD
Director of Operations
Sr. Scientist, Climate Change,
Environmental Defense Fund.
Data Science at Yale, US Forest Service.

Coherent Knowledge