Rulelog: What the New Advances in Logic-Based AI Can Do for Data Science

Benjamin Grosof
Coherent Knowledge*

Slight revised version of Presentation (60-min.) at
Seattle Data Science Meetup**
February 24, 2016 (revised April 6, 2016)
Hosted at Galvanize, Seattle, Washington, USA

Slides also by Janine Bloomfield

* Web: http://coherentknowledge.com
Email: *firstname.lastname*@coherentknowledge.com

** Web: http://www.meetup.com/Seattle-Data-Science/events/227278878/?rv=me1

Smart Rules for Smart Data

Coherent Knowledge: Company Overview

- Logic-based artificial Intelligence (AI) software for advanced analytics, query answering, and decision automation
  - Knowledge representation and reasoning (KRR). Best-of-breed platform component (complements DB, ML, …).
- Leverages over a decade of major government/privately funded research advances in KRR. Founded 7/2013.
  - Most complete and optimized implementation of Rulelog
- Company offers: software product Ergo Suite™ + professional services (incl. dev)
  - Capabilities: smart rules with deep reasoning. Work thru partners to create applications.
  - Apps: policy/regulatory compliance for finance, defense intelligence analysis, e-commerce
- World-class founder team: created many industry-leading logic systems & standards
  - XSB Prolog, RuleML, W3C RIF, W3C OWL-RL, IBM Common Rules, SWRL, SweetRules, Rulelog
  - Extensive experience applying logic systems to numerous domains in government and business

Michael Kifer, PhD
Principal Engineer
Creator, Flora.
Co-Architect, W3C RIF.
Prof., Stonybrook Univ.

Benjamin Grosof, PhD
CTO & CEO
Creator, IBM Common Rules.
Co-Architect, RuleML.
Prof., MIT.
Advanced AI Prog. Mger. for Paul Allen.

Theresa Swift, PhD
Principal Engineer
Co-Architect, XSB Prolog.
Consultant, US CBP.

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

Janine Bloomfield, PhD
Dir., Marketing & Operations
Sr. Scientist, Climate Change, Environmental Defense Fund.
Mindexplorekids.org.

The Core of AI

Artificial Intelligence (AI)

Machine Learning (ML)

Knowledge Representation and Reasoning (KRR)

Combining ML and KRR

- Core AI = KRR + ML. KRR is required for Cognitive Computing too.
- The prediction step of ML requires reasoning
- The target of ML is a representation
- Getting value from ML requires reasoning for analysis and decisions
- KRR is required to combine results of ML, accumulate knowledge coherently, and explain knowledge
  - Weaknesses of ML today
- Reasoning to supply derived facts for ML to chew on
- Reasoning to focus ML's tasks and conjecture schemas
  - e.g., sets of relevant features, important questions, to drive ML

Outline

- Introduction: deep reasoning for data analysis
  - Technology overview: Rulelog, Ergo
- Case Study: financial regulatory/policy compliance
- Human-machine logic: combining English and logic
- More Case Studies: biomedical, e-commerce, defense
- Discussion and Conclusions

Frequent Pains in Analytics Today

- Shallow and disconnected
  - Failure to connect the dots well enough ("pieces on the floor")
  - Too separated from actual decision making
- Failure to leverage human expertise well enough
  - Humans know stuff that's not in the data; often pretty complex to say
    - Programming (vs. data entry/capture) is expensive
- Inflexibility and opacity of what questions can be answered
  - Humans need explanations, must check/debug and trust, ask drill-down Q's
- *Needed: critical technical capabilities to unlock business value*

Ergo Strengths (I)

- Represent more complex knowledge – encode & utilize it
  - Almost any sentence articulable in English natural language
  - Policies, regulations, science
  - Terminology mappings, and context, for data and system integration
  - The actual questions one wants to ask
  - *Capture & inject subject matter experts' (SMEs') insights, directly*

- Reason deeply – assemble & compose multiple analysis results
  - Many steps. Prioritize and weigh counter-arguments.
  - Orchestrate multiple knowledge sources & components
  - Supports high accuracy

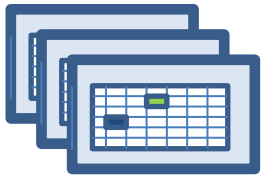
Ergo Strengths (II)

- Explain each answer – fully yet understandably
  - Every logical step is available, and described in English natural language
  - Interactively browsable – user chooses drill downs

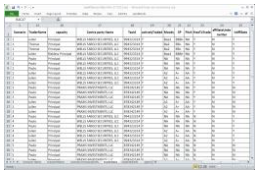
- Overall: *modeling* instead of *programming*
  - Faster, cheaper, more reusable

Evolution of Analytics

Relational DB



Spreadsheets



Domain-specific Apps
e.g., web logs,

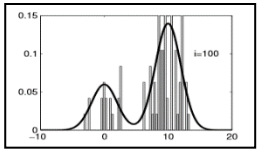
```
66.249.64.214 JRSMITH 873-541-4456 IBM
207.46.13.108 ECHASSE 321-898-9900 MICROSOFT
208.46.13.108 SDETTRA 407-345-8800 ORACLE
209.46.13.108 PROVENC 401-456-9900 HARRIS
199.58.86.206 JOHNSON 505-890-8899 INTERSIL
211.244.83.24 PAULSMI 310-888-9900 MICROSYSTEM
21.247.83.248 JOHNSON 510-887-7777 SUN
94.228.34.211 DJSMITH 178-454-4455 FIDELITY
180.76.150.57 SJOINER 321-787-4444 BANKOFAMERICA
```

Flexible/Graph DBMS tools

RDF, JSON, XML, …

Machine Learning & Statistics tools

Spark, R, Dato, …



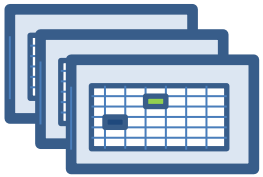
Human Expertise via Programming Alone

in Java, Python, Clojure, F#, …

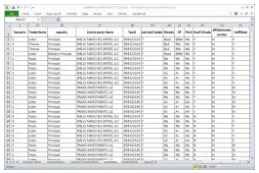
Analytics Results & Decisions

Evolution of Analytics

Relational DB



Spreadsheets

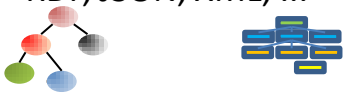


Domain-specific Apps
e.g., web logs,

```
66.249.64.214 JRSMITH 873-541-4456 IBM
207.46.13.108 ECHASSE 321-898-9900 MICROSOFT
208.46.13.108 SDETTRA 407-345-8800 ORACLE
209.46.13.108 PROVENC 401-456-9900 HARRIS
199.58.86.206 JOHNSON 505-890-8899 INTERSIL
211.244.83.24 PAULSMI 310-888-9900 MICROSYSTEM
21.247.83.248 JOHNSON 510-887-7777 SUN
94.228.34.211 DJSMITH 178-454-4455 FIDELITY
180.76.150.57 SJOINER 321-787-4444 BANKOFAMERICA
```

Flexible/Graph DBMS tools

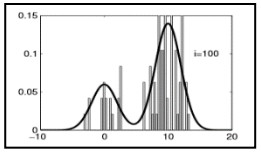
RDF, JSON, XML, …



"smart data"

Machine Learning & Statistics tools

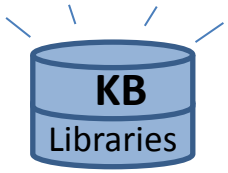
Spark, R, Dato, …



"smart rules"

Human Expertise via

ERGO



+ Programming

Analytics Results & Decisions

+ Explanations

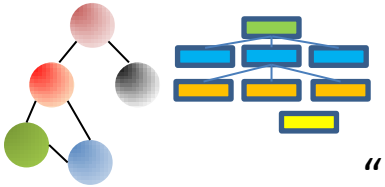
Deep Reasoning with Big Data

- Ergo addresses the 3 V's of big data, in combination with deep reasoning
  - Variety, volume, velocity

- Reasoning scales up
  - to millions of assertions/conclusions,
    for in-memory "hot zone" on one processor
  - to trillions of assertions/conclusions,
    by querying external DBs

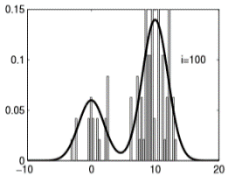
Actively Reason over Today's Gamut of Knowledge

Graph DB
& Semantic tech



Relational DB

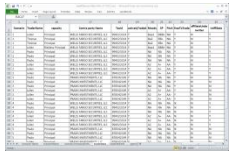
Machine
Learning



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

Domain Apps
& Legacy

Spreadsheets



"Business Rules"

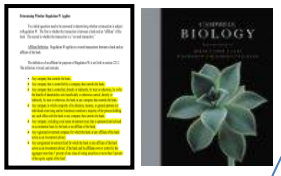
tier(X,1) ∧ supply(Y,X)
⇒ tier(Y,2)

Probabilistic engines

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

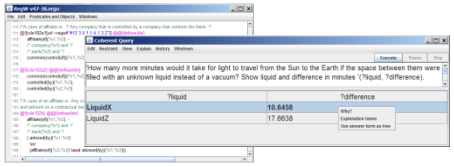
External Info
& Services

Text & Natural
Language processing



Application
Actions

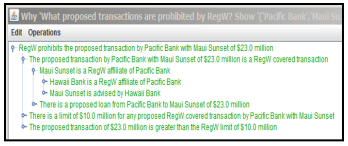
Queries
Assertions
Edits



ERGO

KB
Libraries

Answers, Views
Decisions, Alerts
Explanations



Ergo Architecture

Optional Custom Solutions

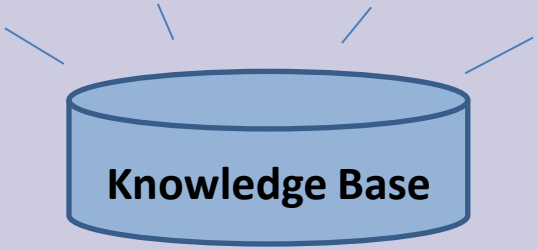
Ergo Suite

Ergo Studio

Rule Editor and Query UI

(Integrated Development Environment)

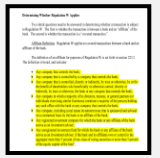
Ergo Reasoner



queries, assertions

answers, explanations

Complex Information
- English Doc.'s etc.
- Policies, Regulations
- Financial, Legal, Science



API's
Java
WS
C

External Services & Frameworks

- Relational DB
- RDF/Graph DB
- Other Sem. Tech
- Machine Learning
- Apps, Docker, …

External Info
(multi-source)

- Data
- Views, Rules
- Schemas & Ontologies
- Results of ML

events, decisions

App Actions

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

Ergo Reasoner & Ergo Studio (IDE/UI)

- Textual Rulelog: Implementation of major research advances in logic (Rulelog) and how to map between logic and English (Textual Logic)
  - **Ergo is the most complete & highly optimized implementation available**
  - Rulelog significantly extends Datalog. Adds Full Meta expressiveness: higher-order, general quantified formulas, exceptions, provenance, probabilistic.
    - Datalog is the logic of databases, business rule systems (production/ECA/Prolog), semantic web ontologies, and earlier-generation semantic web rules (SWRL/RIF/RuleML). Rulelog extends also declarative logic programs (LP).
- Ergo Reasoner component – with sophisticated algorithms & data structures
  - Smart cacheing with dependency-aware updating. Leverages LP & DBMS techniques.
  - Transformation, compilation, reordering, indexing, modularization, dependency analysis, performance monitoring, virtual machine, programming kernel, external import/querying
  - Java API. Other interfaces: command line, web, C, knowledge interchange (below).
  - Scales well: Millions of sentences on one processor; Trillions by querying external DBs
- Ergo Studio component – graphical Integrated Development Environment
  - Interactive editing, explanation, visualization of knowledge
  - Fast edit-test loop with award-winning toolset
- Knowledge interchange with leading and legacy systems: SQL, SPARQL, OWL, …
- Open, standards-based approach. Builds on open source components (incl. XSB Prolog).
  - Rulelog is draft industry standard from RuleML (submission to W3C & Oasis)

Series of Advances → Rulelog's Core Expressive Features

- Well-founded semantics; basic tabling algorithms
  - *Undefined* for paradox; smart cacheing; intuitionistic disjunction
- Higher-order syntax (Hilog); frame syntax
  - Associated optimizations of LP tabling etc. algorithms
- Statement id's for meta; argumentation meta-rules for defeasibility; provenance
- General formulas with all usual classical connectives and quantifiers (omniformity)
- Restraint bounded rationality
  - Use 3rd truth value *undefined* for "don't-care"
  - Radial, skipping; naf unsafety; external-query unsafety, unreturn

KRR Features Comparison: Rulelog Shines

Feature \ System	Rulelog Rules - e.g., Ergo	Datalog Rules - e.g., Jena, SWRL, Ontobroker, SPIN	Production Rules - e.g., IBM, Oracle, Red Hat	Prolog - e.g., SICStus, SWI, XSB	FOL & OWL-DL - e.g., Vampire, Pellet, Prover9	ASP Solvers - e.g., DLV, CLASP
Semantic & on standardization path	✓	✓	restricted case	restricted case	✓	✓
Basic expressiveness						
• Datalog LP	✓	✓	✓	✓	✓	✓
• Logical functions	✓	✗	✗	✓	✓	✓
• General formulas	✓	✗	✗	✗	✓	✗
Full Meta expressiveness						
• Higher-order syntax, provenance	✓	✗	✗	✗	✗	✗
• Defeasibility & well founded negation	✓	✗	✗	✗	✗	✗
• Restraint bounded rationality	✓	✗	✗	✗	✗	✗
• Probabilistic	✓	✗	✗	✗	✗	✗
Efficiency						
• Goal-directed	✓	✗ (except Jena)	✗	✓	✓	✓
• Full LP tabling with dependency-aware updating	✓	✗	✗	✗ (except XSB)	✗	✗
• Polynomial time complexity	✓	✓	✓	✗	✗	✗

Notes on KRR Features Comparison

- "System" means system type / approach of logical knowledge representation and reasoning (KRR).
- "Semantic" means in the sense of KRR, i.e., fully declarative and having a model theory in the logical sense.
- "FOL" means First Order Logic. "ASP" means Answer Set Programs.
  - ASP is recently emerging. The tasks for which it's suitable are more similar to FOL than to the other systems here.
- "Standardization" here means industry standardization. "On path to" means in process of being, or already, standardized.
- "Restricted case" means for a syntactic/expressive subset.
- Event-condition-action rules in this context are similar to, and lumped in with, production rules.
- "LP" means declarative logic programs.
- Datalog means LP without logical functions. Usually this is restricted to Horn. But here we permit negation(-as-failure).
- OWL-RL is pretty much a restricted case of Datalog LP.
- "Higher-order syntax" means Hilog, which enables probabilistic – and also 1) fuzzy and 2) frame syntax cf. F-Logic.
- "Provenance" means provenance info about assertions, via properties of rule id's that are within the logical language / KRR.
- "Full" applies to all four of the meta expressiveness features.
- Defeasibility includes flexible argumentation theories.
- "General formulas" means classical-logic-like formulas, including with head existentials and with head disjunction.
- "LP tabling" includes sophisticated: cacheing of intermediate reasoning results, inference control, and indexing.
- "Dependency-aware updating" means that when assertions are added or deleted, saved inferences are only recomputed if they depend on the changes to the assertions.
- Polynomial time "complexity" means worst-case computational complexity, with constant-bounded number of variables per rule. Polynomial-time is similar to database querying, and is a.k.a. "tractable".

- Datalog X defeasibility: Ontobroker has full well founded negation.
- Prolog X defeasibility: XSB has full well founded negation.
- ASP X defeasibility: ASP has restricted defeasibility & well founded negation.
- Datalog X goal-directed: Jena has a backward engine as well as a forward engine.
- ASP X general formulas: ASP has head disjunction.
- FOL X full LP tabling with dependency-aware updating: Some FOL theorem-provers cache intermediate results in a way that is analogous to LP tabling, and some do dependency tracking but we're not sure how analogous or sophisticated.

Outline

- Introduction: deep reasoning for data analysis
  - Technology overview: Rulelog, Ergo
- Case Study: financial regulatory/policy compliance
- Human-machine logic: combining English and logic
- More Case Studies: biomedical, e-commerce, defense
- Discussion and Conclusions

Case Study 1: Automated Decision Support for Financial Regulatory/Policy Compliance

Problem: Current methods are expensive and unwieldy, often inaccurate

Solution Approach – using Textual Rulelog software technology:

- Encode regulations and related info as semantic rules and ontologies
- Fully, robustly automate run-time decisions and related querying
- Provide understandable full explanations in English
  - *Proof*: Electronic audit trail, with provenance
- Handles increasing complexity of real-world challenges
  - Data integration, system integration
  - Conflicting policies, special cases, exceptions
  - What-if scenarios to analyze impact of new regulations and policies

Business Benefits – compared to currently deployed methods:

- More Accurate
- More Cost Effective – less labor; subject matter experts in closer loop
- More Agile – faster to update
- More Overall Effectiveness: less exposure to risk of non-compliance

Demo of Ergo Suite for Compliance Automation: US Federal Reserve Regulation W

- EDM Council Financial Industry Consortium Proof of Concept – ***successful and touted pilot***
  - Enterprise Data Management Council (Trade Assoc.)
  - Coherent Knowledge Systems (USA, Technology)
  - SRI International (USA, Technology)
  - Wells Fargo (Financial Services)
  - Governance, Risk and Compliance Technology Centre (Ireland, Technology)
- Reg W regulates and limits $ amount of transactions that can occur between banks and their affiliates. Designed to limit risks to each bank and to financial system.
- Must answer 3 key aspects:

1. *Is the transaction's counterparty an <u>affiliate</u> of the bank?*
2. *Is the transaction contemplated a <u>covered transaction</u>?*
3. *Is the amount of the transaction <u>permitted</u> ?*

Determining Whether Regulation W Applies

Two initial questions need to be answered in determining whether a transaction is subject to Regulation W. The first is whether the transaction is between a bank and an "affiliate" of the bank. The second is whether the transaction is a "covered transaction."

<u>Affiliate Definition</u>. Regulation W applies to covered transactions between a bank and an affiliate of the bank.

The definition of an affiliate for purposes of Regulation W is set forth in section 223.2. The definition is broad, and includes:

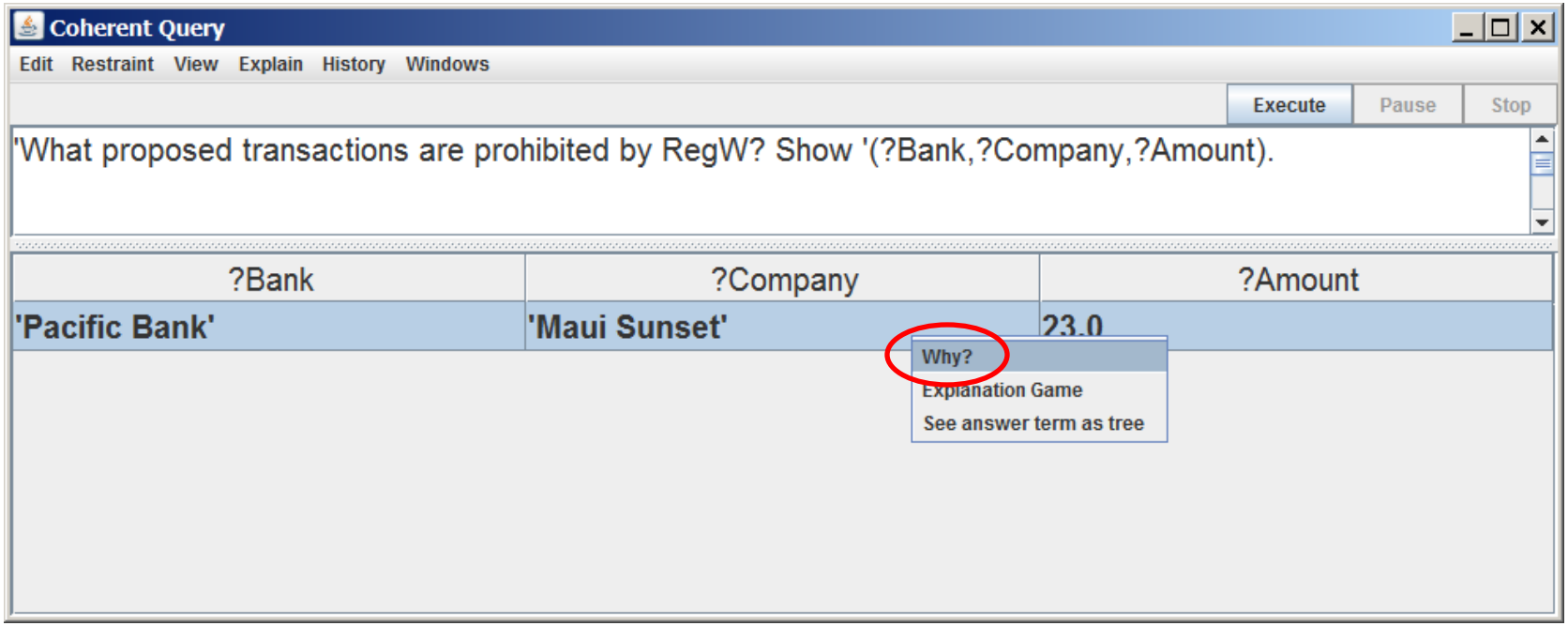
- Any company that controls the bank;
- Any company that is controlled by a company that controls the bank;
- Any company that is controlled, directly or indirectly, by trust or otherwise, by or for the benefit of shareholders who beneficially or otherwise control, directly or indirectly, by trust or otherwise, the bank or any company that controls the bank;
- Any company in which a majority of its directors, trustees, or general partners (or individuals exercising similar functions) constitute a majority of the persons holding any such office with the bank or any company that controls the bank;
- Any company, including a real estate investment trust, that is sponsored and advised on a contractual basis by the bank or an affiliate of the bank;
- Any registered investment company for which the bank or any affiliate of the bank serves as an investment adviser;
- Any unregistered investment fund for which the bank or any affiliate of the bank serves as an investment adviser, if the bank and its affiliates own or control in the aggregate more than 5 percent of any class of voting securities or more than 5 percent of the equity capital of the fund[1];

The Starting Point - Text of Regulation W

Demo goes here

- Note: The demo actually covers the next 10 slides that show explanation screenshots and executable rule assertions in Textual Rulelog. The deck includes those slides in order to be more self-contained. The demo also includes showing interactive syntax checking in the Ergo rule editor.

Query is asked in English



User Clicks the handles to expand the Explanations

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0) ?

Edit Operations Windows

- 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0)
  - RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0) ?

Edit Operations Windows

- 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0)
  - RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
    - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
    - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
    - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0) ?

Edit Operations Windows

- 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0)
  - RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
    - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
      - Maui Sunset is a RegW affiliate of Pacific Bank
      - There is a proposed loan from Pacific Bank to Maui Sunset of $23.0 million
    - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
    - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million

Why is the proposed transaction prohibited by Regulation W?

1. *Is the transaction's counterparty an "<u>affiliate</u>" of the bank?*

YES.

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0) ?

Edit Operations

- RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
  - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
    - Maui Sunset is a RegW affiliate of Pacific Bank
      - Hawaii Bank is a RegW affiliate of Pacific Bank
        - There is common control of Hawaii Bank and Pacific Bank
          - Hawaii Bank is controlled by Americas Bank
            - Hawaii Bank is a subsidiary of Americas Bank
          - Pacific Bank is controlled by Americas Bank
            - Pacific Bank is a subsidiary of Americas Bank
      - Maui Sunset is advised by Hawaii Bank
    - There is a proposed loan from Pacific Bank to Maui Sunset of $23.0 million
  - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
  - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million

And here's why …

Why is the proposed transaction prohibited by Regulation W?

2. *Is the transaction contemplated a "<u>covered transaction</u>"?*

YES.

And here's why …

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Su

Edit Operations

- RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
  - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
    - Maui Sunset is a RegW affiliate of Pacific Bank
      - Hawaii Bank is a RegW affiliate of Pacific Bank
      - Maui Sunset is advised by Hawaii Bank
    - There is a proposed loan from Pacific Bank to Maui Sunset of $23.0 million
  - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
  - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million

Why is the proposed transaction prohibited by Regulation W?

3. *Is the amount of the transaction <u>permitted</u>?*

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0) ?

Edit Operations

- RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
  - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
  - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
    - There is an aggregated-affiliates limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with any affiliate
    - There is an individual-affiliate limit of $250.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
    - The overall RegW limit of $10.0 million is the lesser of $10.0 million and $250.0 million
  - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million

And here's why …

NO.
It went over the limit.

Why is the proposed transaction prohibited by Regulation W?

3. *(continued) Why is the aggregate-affiliates limit $10 million?*

Why 'What proposed transactions are prohibited by RegW? Show '('Pacific Bank','Maui Sunset',23.0) ?

Edit Operations

- RegW prohibits the proposed transaction by Pacific Bank with Maui Sunset of $23.0 million
  - The proposed transaction by Pacific Bank with Maui Sunset of $23.0 million is a RegW covered transaction
  - There is a limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
    - There is an aggregated-affiliates limit of $10.0 million for any proposed RegW covered transaction by Pacific Bank with any affiliate
      - The aggregated total of previous RegW covered transactions by Pacific Bank with all affiliates is $490.0 million
      - The maximum threshold for aggregate RegW covered transactions by Pacific Bank with all affiliates is $500.0 million
        - The capital stock and surplus of Pacific Bank is $2500.0 million
        - The RegW threshold percentage for aggregate affiliates is 20.0 percent
        - $500.0 million is $2500.0 million multiplied by 20.0 percent
      - The limit of $10.0 million is the result of subtracting the previous RegW covered transactions total of $490.0 million from the RegW threshold $500.0 million
    - There is an individual-affiliate limit of $250.0 million for any proposed RegW covered transaction by Pacific Bank with Maui Sunset
    - The overall RegW limit of $10.0 million is the lesser of $10.0 million and $250.0 million
  - The proposed transaction of $23.0 million is greater than the RegW limit of $10.0 million

Examples of the Underlying Textual Rulelog Executable **Fact** Assertions

- subsidiary(of)('Pacific Bank','Americas Bank').
- advised(by)('Maui Sunset','Hawaii Bank').
- bank('Hawaii Bank').
- company('Maui Sunset').
- capital(stock(and(surplus)))('Pacific Bank',2500.0).
- proposed(loan) (from('Pacific Bank'))(to('Maui Sunset')) (of(amount(23.0))) (having(id(1101))).
- previous(loan)(from('Pacific Bank'))(to('Hawaii Bank')) (of(amount(145.0))) (having(id(1001))).
- proposed(asset(purchase))(by('Pacific Bank')) (of(asset(common(stock)(of('Flixado'))))) (from('Maui Sunset')) (of(amount(90.0)))(having(id(1202))).

Executable Assertions: non-fact **Rules**

```
/* A company is controlled by another company when the first company
    is a subsidiary of a subsidiary of the second company. */
@!{rule103b} /* declares rule id */
@@{defeasible} /* indicates the rule can have exceptions */
 controlled(by)(?x1,?x2)
 :- /* if */
   subsidiary(of)(?x1,?x3) \and
   subsidiary(of)(?x3,?x2).

/*A case of an affiliate is: Any company that is advised on a contractual basis by
   the bank or an affiliate of the bank. */
@!{rule102b} @@{defeasible}
 affiliate(of)(?x1,?x2) :-
  ( advised(by)(?x1,?x2)
     \or
    (affiliate(of)(?x3,?x2) \and advised(by)(?x1,?x3))).
```

Executable Assertions: **Exception** Rule

```
@!{rule104e}
@{‘ready market exemption case for covered transaction'} /* tag  for prioritizing */
\neg covered(transaction)(by(?x1))(with(?x2))
          (of(amount(?x3)))(having(id(?Id))) :-
  affiliate(of)(?x2,?x1) \and
  asset(purchase)(by(?x1))(of(asset(?x6)))(from(?x2))(of(amount(?x3)))
     (having(id(?Id))) \and
  asset(?x6)(has(ready(market))).

/* prioritization info, specified as one tag being higher than another */
\overrides(‘ready market exemption case for covered transaction',
                 'general case of covered transaction').

/* If a company is listed on the New York Stock Exchange (NYSE), then the
    common stock of that company has a ready market. */
@!{rule201} @@{defeasible}
asset(common(stock)(of(?Company)))(has(ready(market))) :-
     exchange(listed(company))(?Company)(on('NYSE')).
```

Executable Assertions: **Import of OWL**

```
:- iriprefix fibof =  /* declares an abbreviation */
    "http://www.omg.org/spec/FIBO/FIBO-Foundation/20120501/ontology/".

/* Imported OWL knowledge:  from Financial Business Industry Ontology (FIBO) */
rdfs#subClassOf(fibob#BankingAffiliate, fibob#BodyCorporate).
rdfs#range(fibob#wholly OwnedAndControlledBy, fibob#FormalOrganization).
owl#disjointWith(edmc#Broad_Based_Index_Credit_Default_Swap_Contract,
                 edmc#Narrow_Based_Index_Credit_Default_Swap_Contract).

/* Ontology Mappings between textual terminology and FIBO OWL vocabulary */
company(?co) :- fibob#BodyCorporate(?co).
fibob#wholly OwnedAndControlledBy(?sub,?parent) :- subsidiary(of)(?sub,?parent).

/* Semantics of OWL - specified as general Rulelog axioms */
?r(?y) :- rdfs#range(?p,?r), ?p(?x,?y).
?p(?x,?y) :- owl#subPropertyOf(?q,?p), ?q(?x,?y).
```

Lessons Learned from Case Studies

Financial, E-commerce, and Defense customers 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 Smart Data tech: RDF, SPARQL, OWL

Outline

- Introduction: deep reasoning for data analysis
  - Technology overview: Rulelog, Ergo
- Case Study: financial regulatory/policy compliance
- Human-machine logic: combining English and logic
- More Case Studies: biomedical, e-commerce, defense
- Discussion and Conclusions

Textual terminology

- Leverage Rulelog to much more simply and closely map between natural language and logic

- English phrase ↔ logical term in Rulelog
- English word ↔ logical functor in Rulelog

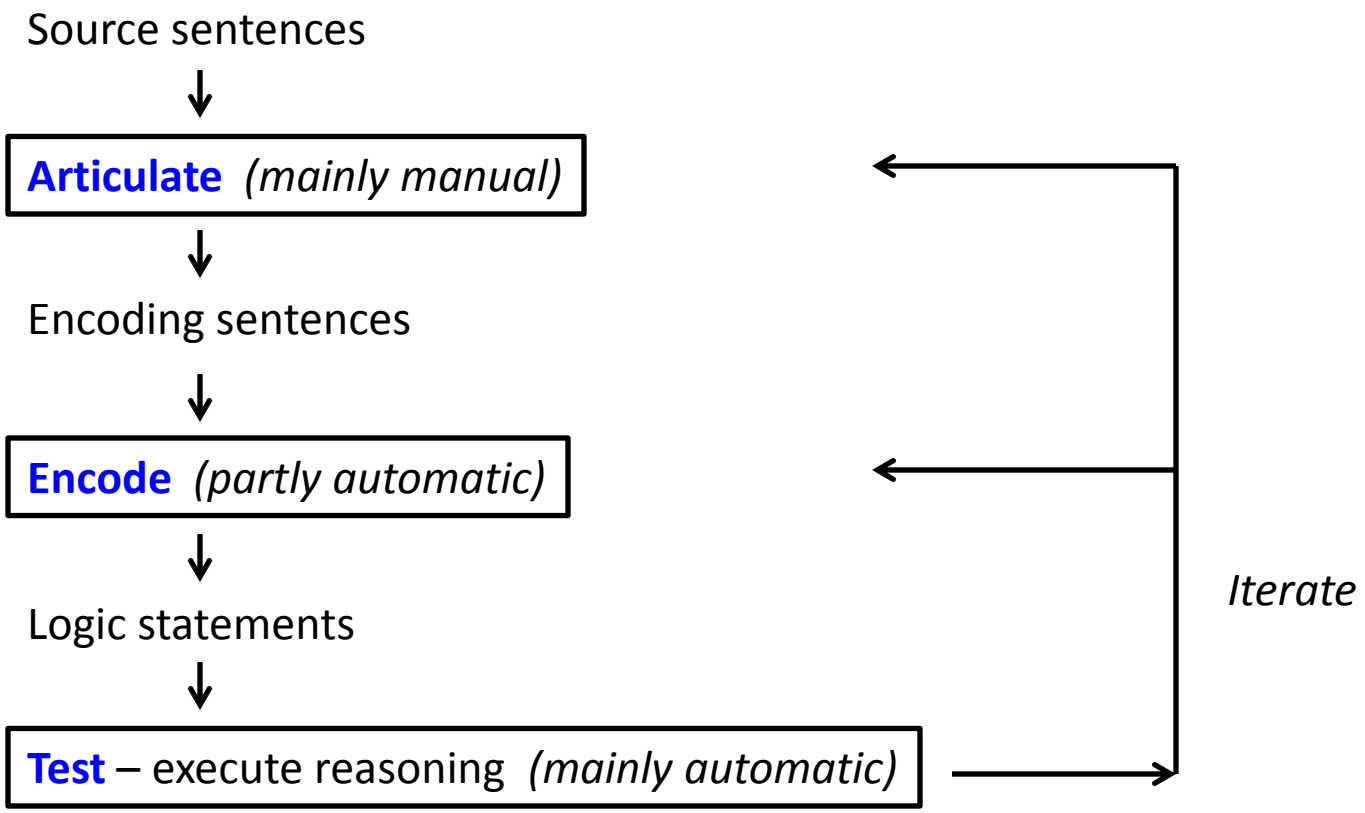
- Basis for textual templates

Knowledge Authoring Process using Ergo Suite

- Start with source text in English – e.g., textbook or policy guide
  - A sentence/statement can be an assertion or a query
- Articulate: create <u>encoding</u> sentences (text) in English.

  As necessary:
  - Clarify & simplify – be prosaic and grammatical, explicit and self-contained
  - State relevant background knowledge – that's not stated directly in the source text
- Encode: create executable logic statements
  - Each encoding text sentence results in one executable logic statement ("rules")
  - Ergo Suite has tools and methodology
- Test and debug, iteratively
  - Execute reasoning to answer queries, get explanations, perform other actions
  - Find and enter missing knowledge
  - Find and fix incorrect knowledge
  - Optionally: further optimize reasoning performance, where critical

Knowledge Authoring Steps using Ergo Suite



In-development: methods to greatly increase the degree of automation in encoding

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

Outline

- Introduction: deep reasoning for data analysis
  - Technology overview: Rulelog, Ergo
- Case Study: financial regulatory/policy compliance
- Human-machine logic: combining English and logic
- More Case Studies: biomedical, e-commerce, defense
- Discussion and Conclusions

ErgoText: Self-documenting, templates

ErgoText:

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

ErgoText Template:

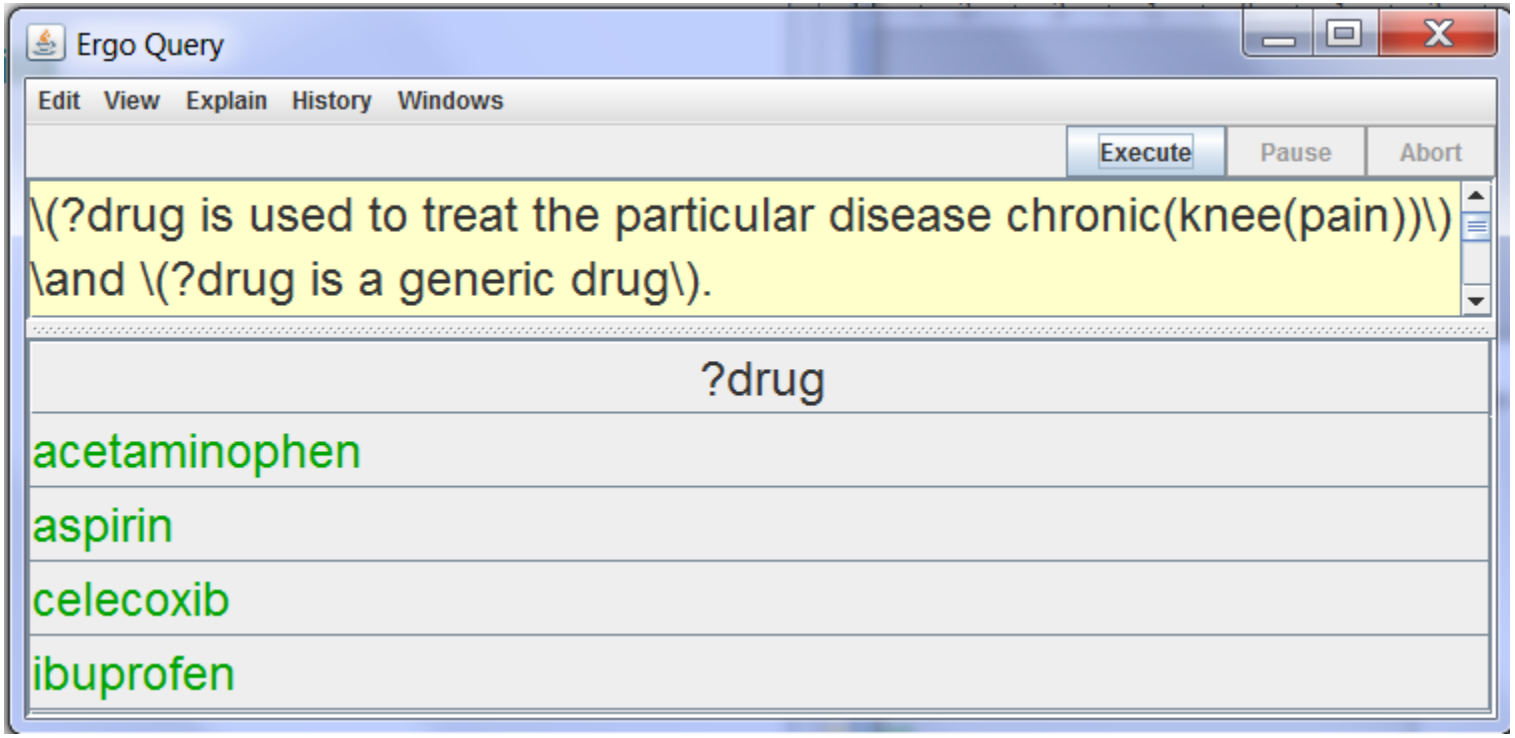
```
template(headbody,
        \(The proposed transaction ?Id by ?Bank with ?Affiliate of $?Amount
          is a RegW covered transaction\),

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

Treatment Scenario 1

- A busy intern encounters an elderly woman in a rehabilitation facility complaining of knee pain.
- What treatment should be given?
- 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.

Which drugs are used in general to treat the chronic knee pain?

Step 1: What are possible drug classes to treat the condition?

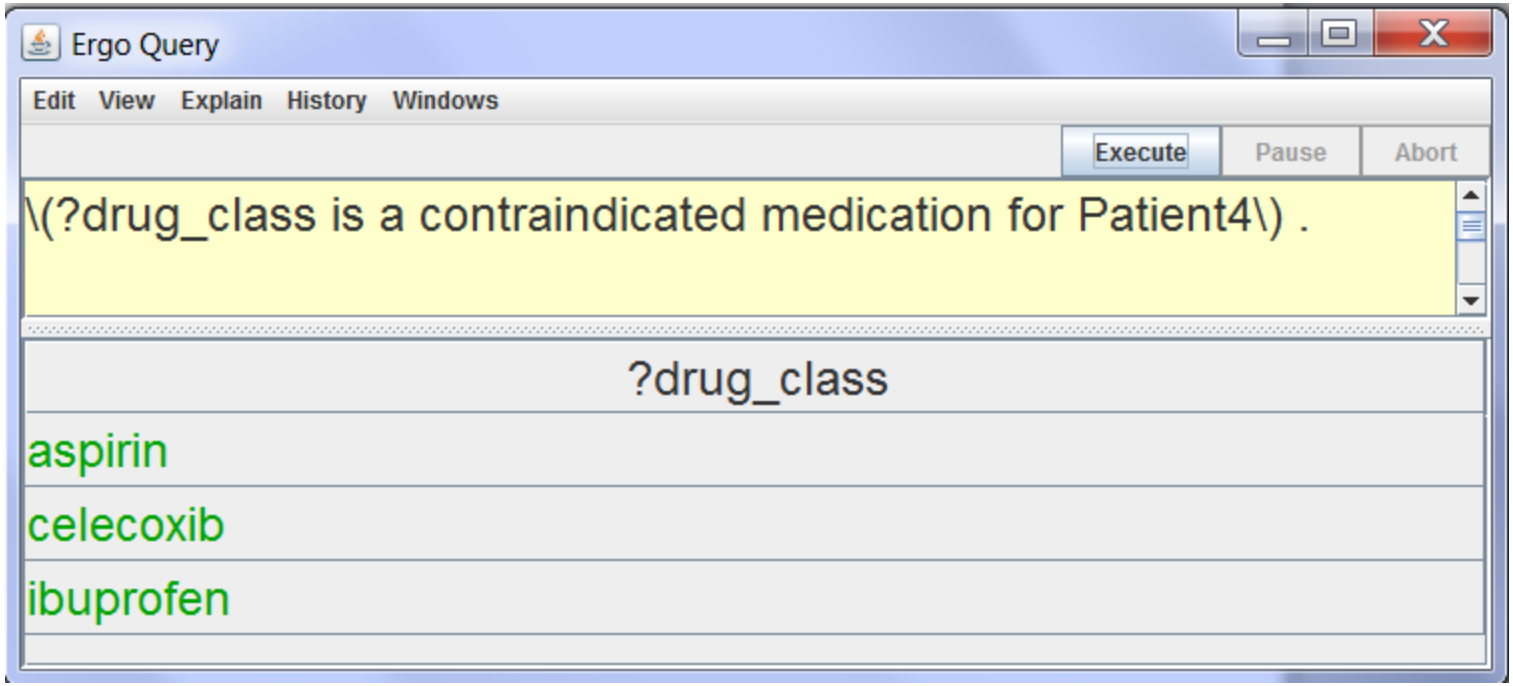


Which drugs are used in general to treat the chronic knee pain?

Step 2: Why is the generic drug ibuprofen recommended for treating chronic knee pain?

ibuprofen is used to treat the particular disease chronic(knee(pain))
- ibuprofen is in the anti-inflammatory class of drugs
- the anti-inflammatory class of drugs is used to treat inflammatory(disease)
- chronic(knee(pain)) is a subclass of inflammatory(disease)

Which drug classes are contraindicated for this patient?



Why is ibuprofen contraindicated for this patient?

Justifications

Windows

Why contraindicated_medication(ibuprofen,Patient4) ?

- ibuprofen is a contraindicated medication for Patient4
  - ibuprofen aggravates the risk of bleeding for Patient4
    - ibuprofen interacts with warfarin for Patient4 to increase the risk of bleeding
      - ibuprofen and warfarin are distinct drugs
      - ibuprofen is a drug
      - warfarin is a drug
      - The drug ibuprofen increases the risk of bleeding
      - The drug warfarin increases the risk of bleeding
      - Patient4 takes warfarin as an existing medication
      - It is proposed that Patient4 takes ibuprofen as a new medication

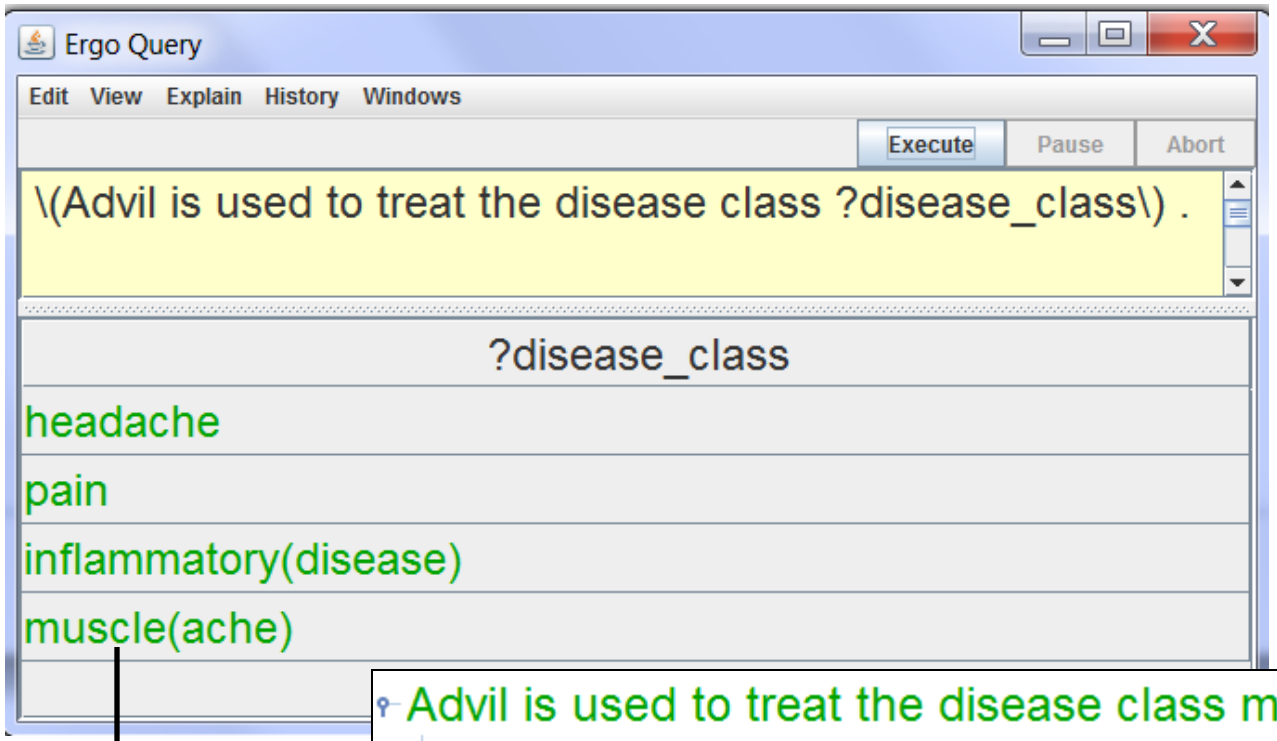
More details: Why is ibuprofen contraindicated?

Anti-inflammatories are contraindicated for the patient because there is a drug-drug interaction between anticoagulants and anti-inflammatories. Anti-inflammatories increase bleeding risk to patients taking anticoagulants.

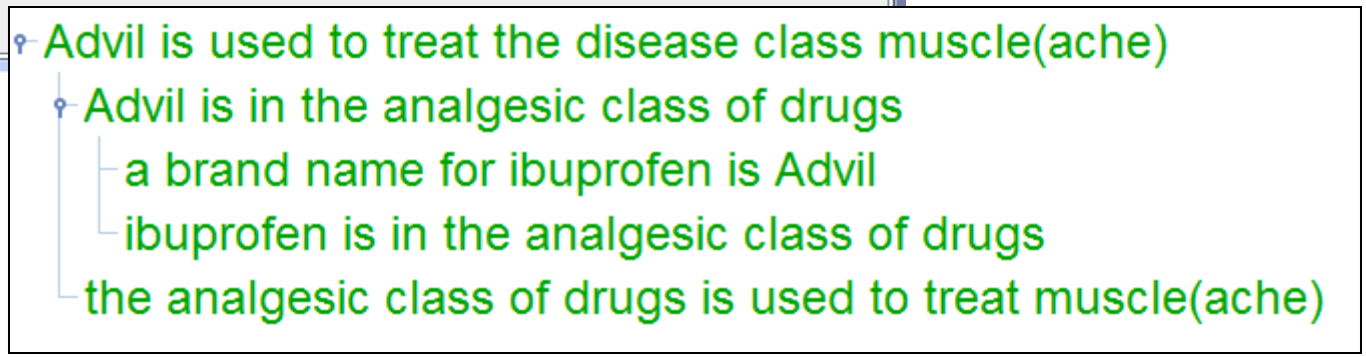
- ibuprofen interacts with warfarin for Patient4 to increase the risk of bleeding
  - ibuprofen and warfarin are distinct drugs
  - ibuprofen is a drug
  - warfarin is a drug
  - The drug ibuprofen increases the risk of bleeding
    - ibuprofen is a drug
    - anti-inflammatory is a class of drugs
    - The drug class anti-inflammatory increases the risk of bleeding
    - ibuprofen is in the anti-inflammatory class of drugs
  - The drug warfarin increases the risk of bleeding
    - warfarin is a drug
    - anticoagulant is a class of drugs
    - The drug class anticoagulant increases the risk of bleeding
    - warfarin is in the anticoagulant class of drugs

Easily query the database for both clinical decision support and education.

What conditions can a particular drug treat?



Why?



Case Study: E-commerce marketing: Configuring Complex Products

- Very large manufacturing company
- Product catalog: highly technical, complex
- Over 1 Million products, in over 50 countries
- Problem: Creating configurators and updating catalog too slow & costly. Thus siloed by sales regions.
- Solution (with smart data partner) on $multi-Billion segment
  - Developed product ontology: compatible with legacy and standards
  - Converted all product data into live triple store
  - Added configuration rules, in Rulelog, on top of that catalog
  - Leveraged Ergo expressiveness and RDF import capability
- Benefits: faster, cheaper, agile updating, reusable globally
  - Enable greatly enhanced up-sell and cross-sell, across product lines

E-commerce marketing: Pricing & Promotions

- Example 1: Loyal customer gets 5% discount

```
price(of(item(?i))((for(customer(?c))(is(?p)) :-
    normal(price)(of(item(?i))(is(?pn)) \and
    ?c(is(a(loyal(customer)))) \and
    ?p \is ?pn * (1 - 0.05).
```

- Example 2: New customer gets $20 coupon on first $100 of purchases
- Via API, query Ergo as business logic within application managing overall customer experience

Data cleaning – examples

- An Ergo rule can detect a schema violation
  - implausibly(high(salary))(?person,XYZCo,2015) :-
    employee(?person,XYZCo,?position),
    annual(salary)(?person,XYZCo,2015,?amount(USD)),
    ?amount > 500000,
    \neg CXO(?position).

- An Ergo rule can treat a missing value, e.g., draw an inference about it
  - @{'presume device owner is application owner'}
    application(owner)(?person,?app) :-
    smartphone(owner)(?person,?device),
    installed(application)(?app,?device).

Case Study: Defense Intelligence Analysis

- Challenge: managing, accessing, integrating knowledge
  - Huge amounts of information as text, RDF, triple stores
- Current text extraction methods:
  - Noisy/inaccurate, shallow, patchy
  - Lacks contextualization, e.g., is a date past, present, or future?
  - Events are more complex than Entities, and are only shallowly treated

Rulelog enriches Text Extraction

- Leverage Rulelog's high expressiveness and flexibility
- Mappings between multiple terminologies or ontologies

"moving a bomb" —*implies*→ "transporting weaponized material"

Lessons Learned from Case Studies

Financial, E-commerce, and Defense customers 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 Smart Data tech: RDF, SPARQL, OWL

Outline

- Introduction: deep reasoning for data analysis
  - Technology overview: Rulelog, Ergo
- Case Study: financial regulatory/policy compliance
- Human-machine logic: combining English and logic
- More Case Studies: biomedical, e-commerce, defense
- Discussion and Conclusions

Ergo Strengths (I) [repeat slide]

- Represent more complex knowledge – encode & utilize it
  - Almost any sentence articulable in English natural language
  - Policies, regulations, science
  - Context and mappings for data and system integration
  - The actual questions one wants to ask
  - *Capture & inject subject matter experts' (SMEs') insights, directly*

- Reason deeply – assemble & compose multiple analysis results
  - Many steps. Prioritize and weigh counter-arguments.
  - Orchestrate multiple knowledge sources & components
  - Supports high accuracy

Ergo Strengths (II) [repeat slide]

- Explain each answer – fully yet understandably
  - Every logical step is available, and described in English natural language
  - Interactively browsable – user chooses drill downs

- Overall: *modeling* instead of *programming*
  - Faster, cheaper, more reusable

Topics for Discussion

- How might Rulelog/Ergo apply in your work?
- Stuff we didn't have time to get into
  - More domains: continuing education, financial reporting, natural language human-computer interaction (NL HCI), insurance, games, Internet of Things (IoT)
  - Probabilistic reasoning

Thank you.

Deep Reasoning for Advanced Analytics

http://coherentknowledge.com

OPTIONAL/BACKUP SLIDES FOLLOW

Probabilistic Knowledge & Reasoning, in Rulelog

- Leverage Hilog and restraint

- Probabilistic knowledge has tuple of parameters
  - Prob(<formula-term>, <parameters>)
  - Flexible in regard to what are the <parameters>:
    - Point value
    - Interval
    - Mean, standard-deviation
    - Interval, confidence-level, sample-size, statistical-technique

- Evidential reasoning: weighted or prioritized combination

- Distribution semantics: semantics/foundation of Probabilistic LP

END of OPTIONAL/BACKUP SLIDES