Making Very Expressive Rules Practical in Logic and Text†

Benjamin Grosof* **

slides also by Paul Haley* ** and Janine Bloomfield*

Nov. 5, 2013

Decision CAMP 2013‡

San Jose, CA, USA

* Coherent Knowledge Systems, LLC, coherentknowledge.com

** Automata, Inc., paul@haleyai.com

† Work partly supported by Vulcan, Inc., http://www.vulcan.com

‡ http://decision-camp.com/

Outline

- Rulelog – more expressive logical KRR / advanced DB
- Textual Logic (TL): English ⟷ logic
- Case study with Demo: Federal Reserve Regulation W
  - Automate Decisions, with full Explanations
- Results from TL Authoring Experiment: Cell Membranes Bio
  - Rapid Collaborative Knowledge Acquisition (KA)
- Conclusions

Requirements on the logical KRR
for KA of Rich Logical K

- **The logic must be expressively rich – higher order logic formulas**
  - As target for the text interpretation
- **The logic must handle exceptions and change, gracefully**
  - Must be **defeasible**
    = K can have exceptions, i.e., be "defeated", e.g., by higher-priority K
  - For empirical character of K
  - For evolution and combination of KB's. I.e., for **social scalability**.
  - For causal processes, and "what-if's" (hypotheticals, e.g., counterfactual)
  - I.e., to represent change in K and change in the world
- **Inferencing in the logic must be computationally scalable**
  - Incl. **tractable** = **polynomial-time** in worst-case
  - (as are SPARQL and SQL databases, for example)

Past Difficulties with Rich Logical K

- **Hard to capture complex knowledge from English**
  - **… and manage change of knowledge**

- **KRR not defeasible & tractable**
- **E.g.**
  1. FOL-based – OWL, SBVR, CL: infer garbage
     - Perfectly brittle in face of conflict from errors, confusions, tacit context
  2. E.g., FOL and previous logic programs: run away
     - Recursion thru logical functions

- **KRR not higher-order and meta enough**

Knowledge often has Exceptions

- **A.k.a. knowledge is *defeasible* (i.e., can be "defeated")**
- **"A (eukaryotic) cell has a nucleus." … Except when it doesn't ☺**
  - A cell has no nucleus during anaphase. Red blood cells have no nuclei.
  - A cell has two nuclei between mitosis and cytokinesis. Some fungi are multinucleate.
- **Exceptions / special cases are inevitably realized over time**
  - E.g., knowledge is incomplete, multiple authors contribute, …
- **Requiring entered knowledge to be strictly / universally true (exception-free) is impractical**
  - Precludes stating generalities (the typical) and thus the population of authors
  - "The perfect is the enemy of the good"
- **Exceptions manifest as contradictions, i.e., conflict**
- **Leveraging multiple sources of knowledge (e.g., KB merging) requires conflict resolution**
  - Errors. Confusions. Omitted context.

Defeasibility is Indicated When…

- **Useful generalities – and potential exceptions – coexist**
  - Specify knowledge in detail/precision appropriate for various circumstances
- **Governing doctrine, definitions, or other knowledge, cannot be assured to be conflict-free, e.g.:**
  - Multiple sources of governing doctrine exist
    - Typically, no central authority resolves all conflict promptly
  - Truth depends on context
    - Yet context is rarely made fully explicit
- **Many broad realms are full of exceptions**
  - Policies, regulations, laws — and the workflows they drive
    - Multiple jurisdictions, organizations, contracts, origins
  - Learning and science. Updating. Debate.
    - May falsify previous hypotheses after observation or communication
  - Causal processes: changes to state, from interacting/multiple causes
  - Natural language (text interpretation): "there's a gazillion special cases"

Rulelog: Overview

- **First KRR to meet central challenge:**
  defeasible + tractable + rich
- **New rich logic: based on databases, not classical logic**
  - Expressively extends normal declarative logic programs (LP)
  - Transforms into LP
  - LP is the logic of databases (SQL, SPARQL) and pure Prolog
    - Business rules (BR) – production-rules -ish – has expressive power similar to databases
  - LP (not FOL) is "the 99%" of practical structured info management today
- **Advanced DB with new reasoning techniques to implement it**
  - LP "tabling". Optimizations for logical functions. Explanations, incl. of why-not. Transformations, incl. for expressive extensions. Bounded rationality.
- **Prototyped in Vulcan's SILK**
- **Commercially supported in Coherent 1.0**
  - Engine and HCI running on top of XSB Prolog
- **Rulelog in draft as industry standard (RuleML submission to W3C RIF)**

Textual Logic Approach: Overview

- **Logic-based text interpretation & generation, for KA & QA**
  - Map text to logic ("text interpretation"): for K and Q's
  - Map logic to text ("text generation"): for viewing K, esp. for justifications of answers (A's)
  - Map based on logic
- **Textual terminology – phrasal style of K**
  - Use words/word-senses directly as logical constants
  - Natural composition: textual phrase $\leftrightarrow$ logical term
- **Interactive logical disambiguation technique**
  - Treats: parse, quantifier type/scope, co-reference, word sense
  - Leverages lexical ontology – large-vocabulary, broad-coverage
  - Initial restriction to stand-alone sentences – "straightforward" text
    - Minimize ellipsis, rhetoric, metaphor, etc.
  - Implemented in Automata Linguist™
- **Leverage defeasibility of the logic**
  - For rich logical K: handle exceptions and change
    - Incl. for NLP itself: "The thing about NL is that there's a gazillion special cases" [Peter Clark]

KA = Knowledge Acquisition. QA = Question/Query Answering. NLP = Natural Language Processing.

Rulelog: more details

- Defeasibility based on ***argumentation theories (AT)*** [Wan, Grosof, Kifer 2009]
  - Meta-rules (~10's) specify principles of debate, thus when rules have exceptions
  - Prioritized conflict handling. Ensures consistent conclusions. Efficient, flexible, sophisticated defeasibility.
- ***Restraint***: semantically clean **bounded rationality** [Grosof & Swift, AAAI-13]
  - Leverages "undefined" truth value to represent "not bothering"
  - Extends well-foundedness in LP
- ***Omniformity***: higher-order logic formula syntax, incl. hilog, rule id's
  - Omni-directional disjunction. Skolemized existentials. [Grosof (invited), RuleML-2013]
  - Avoids general reasoning-by-cases (cf. unit resolution).
- Sound interchange of K with all major standards for sem web K
  - Both FOL & LP, e.g.: RDF(S), OWL-DL, SPARQL, CL
- Reasoning techniques based on extending tabling in LP inferencing
  - Truth maintenance, justifications incl. why-not, trace analysis for KA debug, term abstraction, delay subgoals [Andersen et al, RuleML-2013 (Challenge)]

For more info, see [Grosof, AAAI-13 Tutorial] – largely about Rulelog. Also see [Ontolog Forum 6/20/13, 10/31/13]

Example: Ontology Translation, leveraging hilog and exceptions

```
/* Company BB reports operating earnings using R&D operating cost which includes price of a
   small company acquired for its intellectual property. Organization GG wants to view
   operating cost more conventionally which excludes that acquisition amount. We use rules to
   specify the contextual ontological mapping. */
@{normallyBringOver} ?categ(GG)(?item) :- ?categ(BB)(?item).
@{acquisitionsAreNotOperating} neg ?categ(GG)(?item) :-
      acquisition(GG)(?item) and (?categ(GG) :: operating(GG)).
\overrides(acquisitionsAreNotOperating, normallyBringOver). /* exceptional */
acquisition(GG)(?item) :- price_of_acquired_R_and_D_companies(BB)(?item).
R_and_D_salaries(BB)(p1001). p1001[amount -> $25,000,000].
R_and_D_overhead(BB)(p1002). p1002[amount -> $15,000,000].
price_of_acquired_R_and_D_companies(BB)(p1003). p1003[amount -> $30,000,000].
R_and_D_operating_cost(BB)(p1003). /* BB counts the acquisition price item in this category */
R_and_D_operating_cost(GG) :: operating(GG).
Total(R_and_D_operating_cost)(BB)[amount -> $70,000,000]. /* rolled up by BB cf. BB's definitions */
Total(R_and_D_operating_cost)(GG)[amount -> ?x] :- … . /* roll up the items for GG cf. GG's definitions */
```

As desired: |= R_and_D_salaries(GG)(p1001)

|= neg R_and_D_operating_cost(GG)(p1003) /* GG doesn't count it */

|= Total(R_and_D_operating_cost)(GG)[amount -> $40,000,000]

Notation: @{…} declares a rule tag. ? prefixes a variable. :- means if. X :: Y means X is a subclass of Y. \overrides(X,Y) means X is higher priority than Y.

Outline

- Rulelog – more expressive logical KRR / advanced DB
- Textual Logic (TL): English ⟷ logic
- Case study with Demo: Federal Reserve Regulation W
  - Automate Decisions, with full Explanations
- Results from TL Authoring Experiment: Cell Membranes Bio
  - Rapid Collaborative Knowledge Acquisition (KA)
- Conclusions

Financial Regulatory Compliance:
Using Coherent software for Regulation W

Case Study from

Coherent Knowledge Systems, LLC

http://coherentknowledge.com

Banking Regulation

Problem:

- Complex set of regulations govern wide range of operations and activities that financial institutions engage in every day
- Compliance and *Proof* of compliance are essential
  - External: Outside regulators
  - Internal: Company's management chain
- Automated support needed –
  - Current methods are expensive and unwieldy

Solution:

- Coherent Knowledge Systems – software and services
  - Automates decisions for compliance with banking regulation
  - Advanced database with automated logical reasoning
  - Explains fully how compliance decisions were reached

USA Federal Reserve Act: Regulation W

- Concerns activities/transactions between a bank and "affiliates"
- Designed to limit risks
- Defines who is an "affiliate", what is a "covered transaction", whether a particular transaction is permitted

San Francisco's Federal Reserve building

Interpreting Regulation W: 3 key aspects

1. *Is the transaction's counterparty an affiliate of the bank?*

2. *Is the transaction contemplated a covered transaction?*

3. *Is the amount of the transaction permitted ?*

Document from Federal Reserve

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."

Affiliate Definition. 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];

Example Scenario:
A Loan to the Maui Sunset Hotel Group

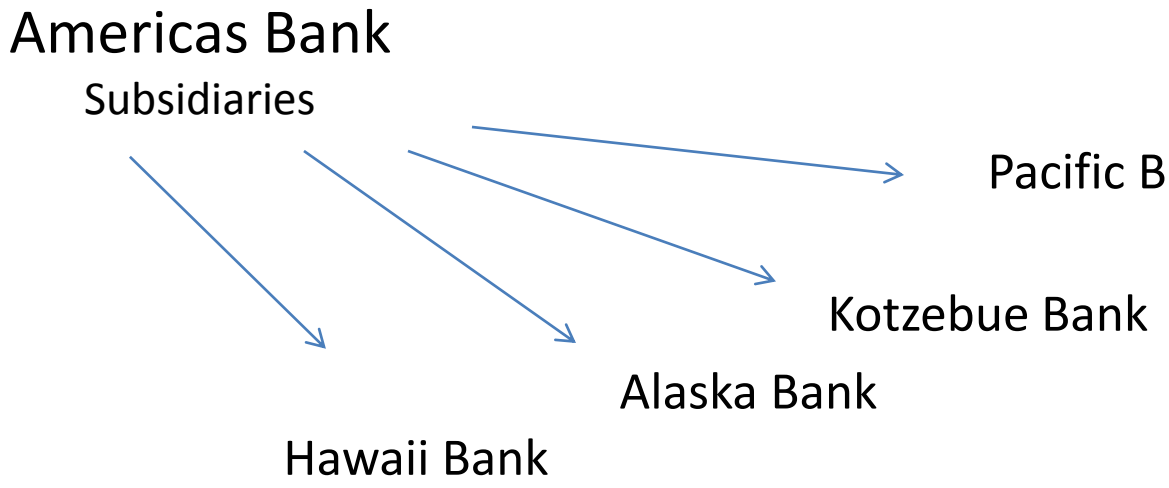
Pacific Bank is considering a loan of $23 million dollars to the Maui Sunset hotel group to open a new location on the island.

Is this transaction allowed under Regulation W?

As part of that, one must ascertain if Maui Sunset could be considered an affiliate under Regulation W.

To watch the demo video, click here <TBD, on Coherent website>

Relationships and Other Data in the Demo Scenario



Capital Stock and Surplus

Pacific Bank	$2500 million

Previous Loans

Pacific Bank	Hawaii Bank	$145 million
Pacific Bank	Alaska Bank	$245 million
Pacific Bank	Kotzebue Bank	$100 million

Pertinent Regulation W Rules and Definitions

For this scenario, the following specific rules and definitions from the Federal Reserve Act were utilized:

Section 223.2 Definition of Affiliate :

- 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.
- A financial subsidiary of the bank

Section 223.3(h) Definition of Covered Transaction:

- An extension of credit to an affiliate

Quantitative Limitations.

- A bank may not engage in a new covered transaction with an affiliate if the aggregate amount of covered transactions between the bank and the affiliate would be in excess of 10 percent of the bank's capital stock and surplus after consummation of the new transaction.
- Aggregate covered transactions between the bank and all affiliates are limited to 20 percent of the bank's capital stock and surplus.

Regulation W Becomes Coherent Logic

Using Coherent tools:

- Regulation W is translated from English into logic, rapidly.
- A knowledge base is created, ready to make decisions and provide detailed explanations.

Sample English Text:

> Any company that is advised on a contractual basis by the bank or an affiliate of the bank is considered an affiliate of the bank.

Logical representation:

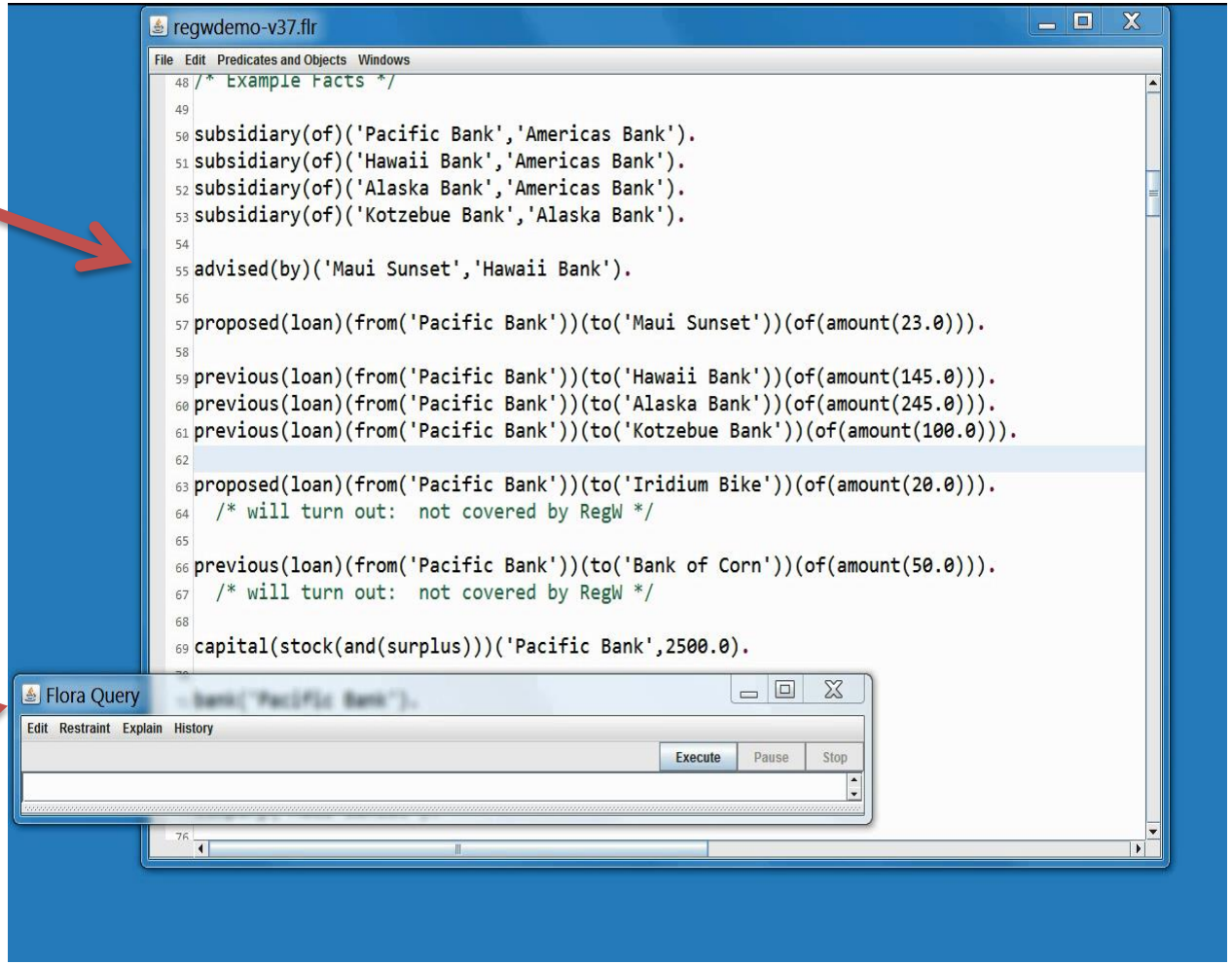
```
114    affiliate(of)(?x1,?x2) :-
115    ( advised(by)(?x1,?x2)
116      or
117      (affiliate(of)(?x3,?x2) and advised(by)(?x1,?x3))).
```

Coherent User Interface

Coherent software includes various tools. Two are:

- **Knowledge-base editor**
  Regulations, data, and linguistic information are entered here as logical formulations. Additional editing tools (not shown here) start directly from English.

- **Query window**
  Ask the knowledge base and get answers for decision making.

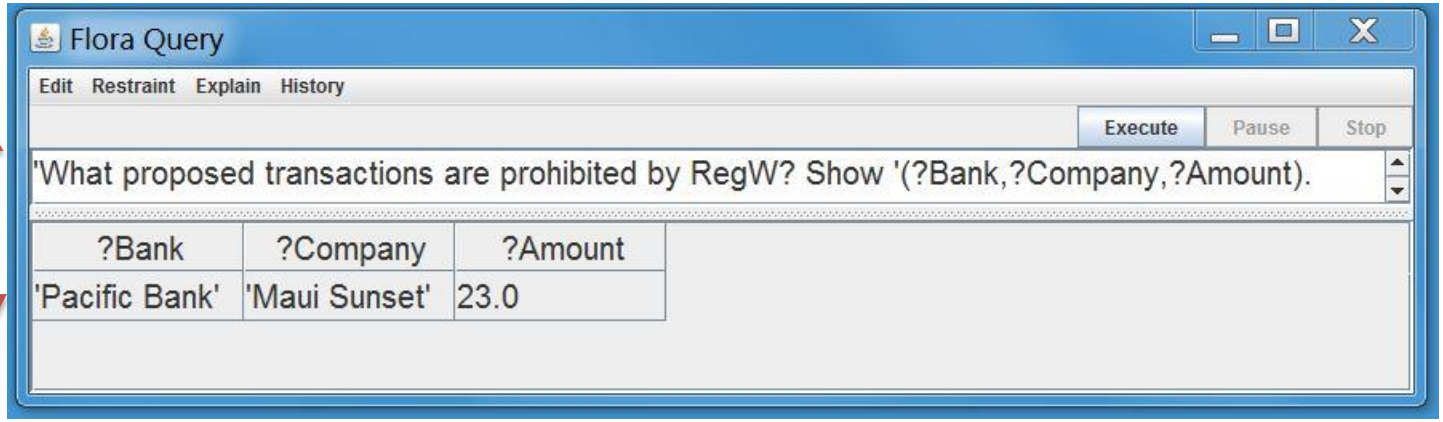


Sample Question:

"What proposed transactions are prohibited by Regulation W?"

Query

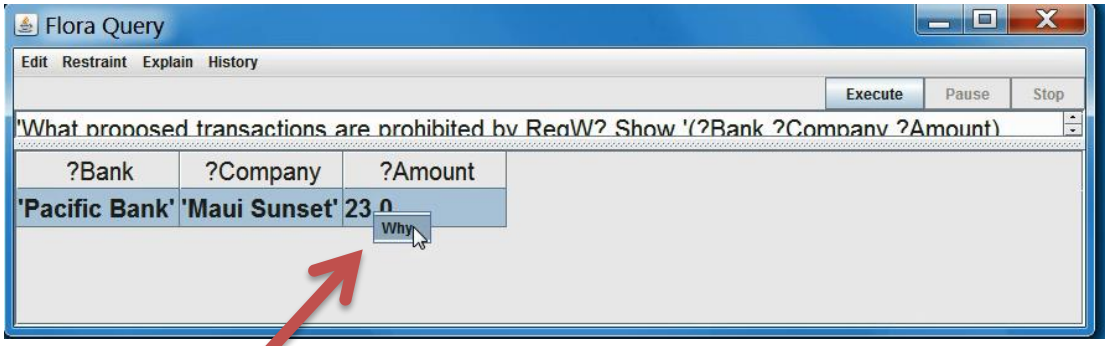
A decision answer is generated



Coherent software automatically makes a decision using the Regulation W knowledge base, including data facts from the example scenario:

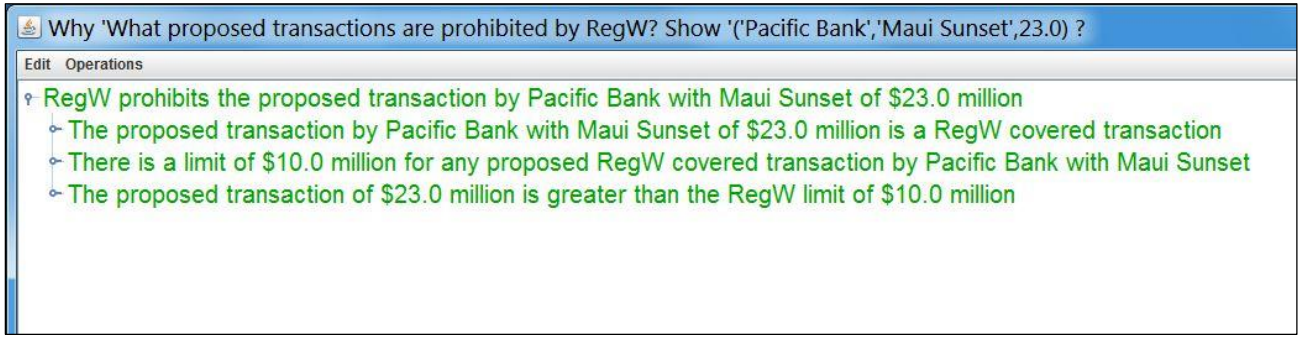
The proposed transaction between Pacific Bank and Maui Sunset in the amount of $23.0 million is prohibited.

Explanation
of How Decisions Are Reached



Explanation Window

Why?

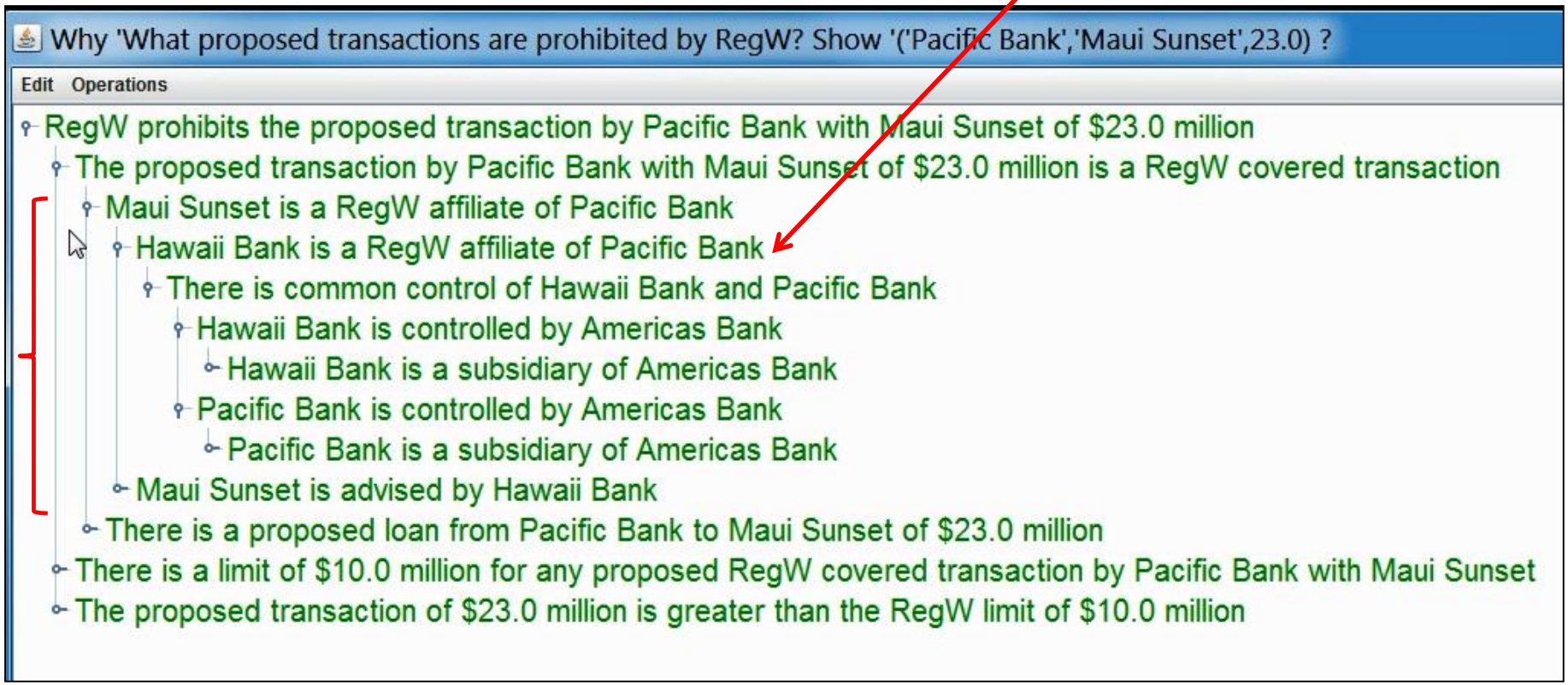


- Clicking on the 'why' button for a decision answer opens an explanation window.
- Clicking on any line in the explanation drills down to its supporting steps.

Why is the proposed transaction prohibited by Regulation W?

1. *Is the transaction's counterparty an "affiliate" of the bank?*

YES.



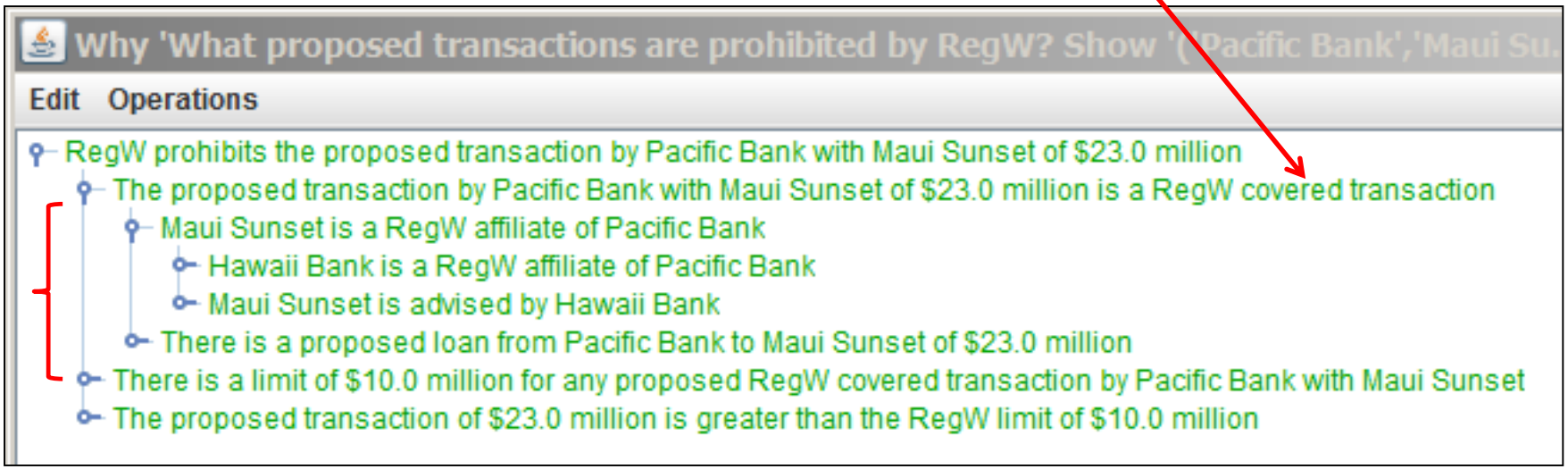
And here's why …

Why is the proposed transaction prohibited by Regulation W?

2. *Is the transaction contemplated a "covered transaction"?*

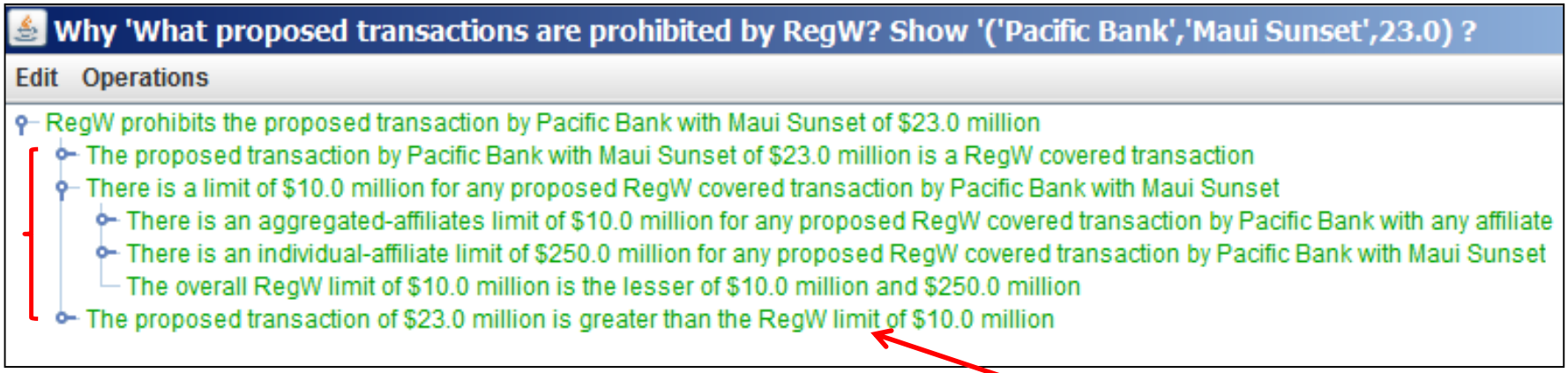
YES.

And here's why …



Why is the proposed transaction prohibited by Regulation W?

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

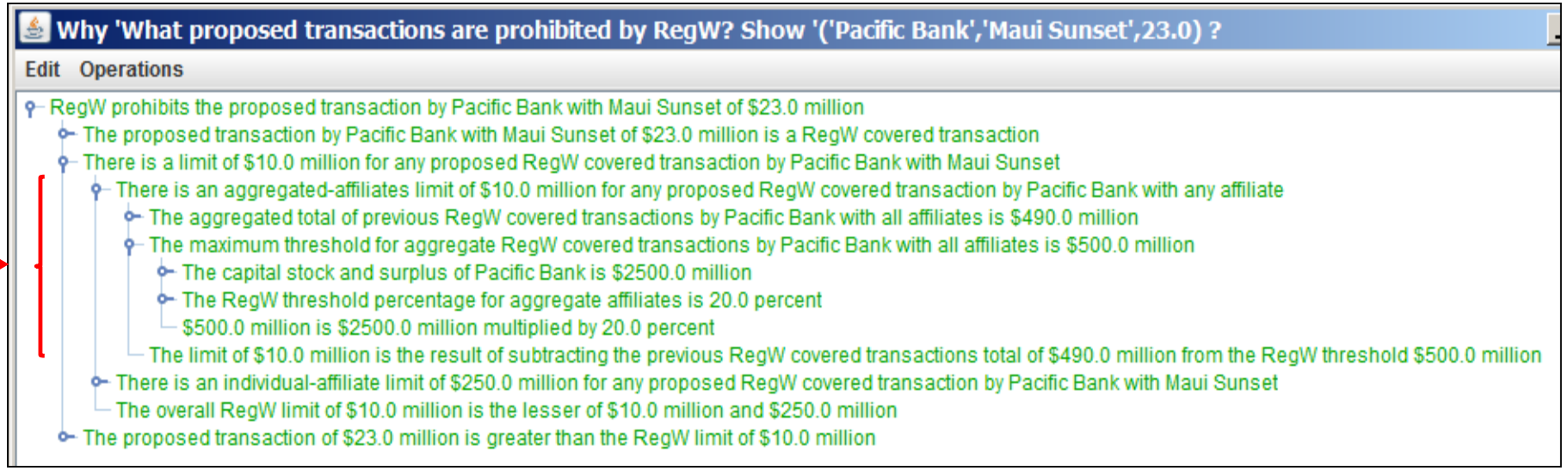


And here's why …

NO.
It went over the limit.

Why is the proposed transaction prohibited by Regulation W?

3. *(continued) How was the limit calculated, using the bank's capital, to determine whether the covered transaction was permitted?*



Here's how the aggregate-affiliates limit was determined

Demo Summary: Coherent software …

- Translates Regulation W into a Coherent Logic knowledge base, integrated with financial data
- Automates decisions for regulatory compliance
- Explains how conclusions were drawn
  - In readable English
  - The user can select the level of detail wanted
  - Supports human decision making, review, assurance, and *proof* of compliance

Outline

- Rulelog – more expressive logical KRR / advanced DB
- Textual Logic (TL): English ⟷ logic
- Case study with Demo: Federal Reserve Regulation W
  - Automate Decisions, with full Explanations
- Results from TL Authoring Experiment: Cell Membranes Bio
  - Rapid Collaborative Knowledge Acquisition (KA)
- Conclusions

Knowledge Acquisition for Deep QA: Expt.

- Goal 1: represent the knowledge in one chapter of a popular college-level science textbook, at 1st-year college level
  - Chapter 7 on cell membranes, in Biology 9th ed., by Campbell et al
- Goal 2: measure what KA productivity is achieved by KE's
  - Assess level of effort, quality of resulting logic, and coverage of textbook
- Software used in this case study:
  - for translating English to logic
    - Automata Linguist™ and KnowBuddy™ (patents pending)
    - English Resource Grammar (http://www.delph-in.net/erg/)
  - for knowledge representation & reasoning
    - Vulcan, Inc.'s SILK (http://www.projecthalo.com/): prototype implementation of Rulelog

Summary of Effort & Results

- Captured 3,000+ sentences concerning cellular biology
  - hundreds of questions (2 examples herein)
  - 600 or so sentences directly from Campbell's Biology textbook
  - 2,000 or so sentences of supporting or background knowledge
- Sentence length averaged 10 words up to 25 words
  - background knowledge tends to be shorter
  - disambiguation of parse typically requires a fraction of a minute
  - hundreds of parses common, > 30 per sentence on average
  - the correct parse is typically not the parse ranked best by statistical NLP
- Sentences disambiguated and formalized into logic in very few minutes on average
  - resulting logic is typically more sophisticated than skilled logicians typically produce
- Collaborative review and revision of English sentences, disambiguation, and formalization approximately doubled time per sentence over the knowledge base

Tracked effort & collaboration per sentence

Sentences (2322) | Relations

all types | axiomatic | all editors | 7/ 1/2012

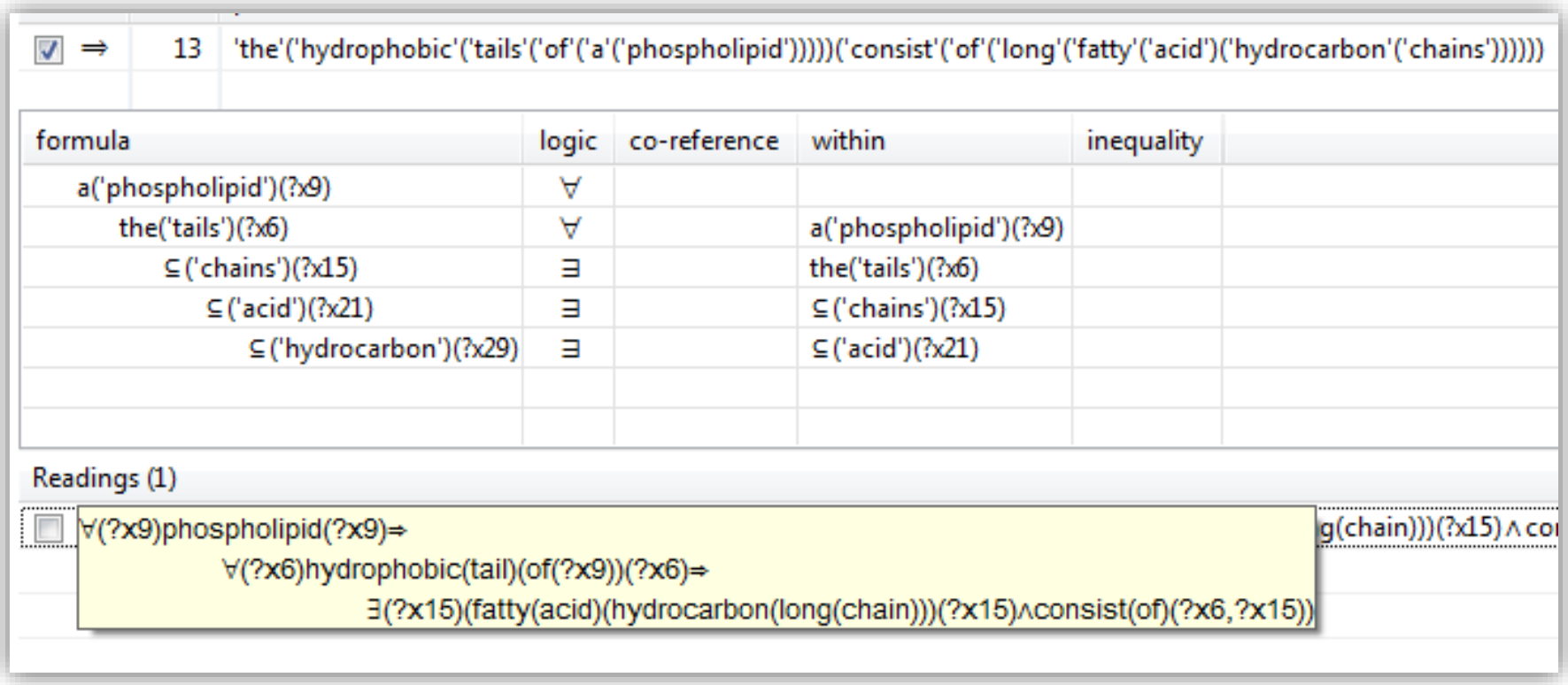
Creator	Created	Editor	Edited	Noted by	Noted	Status	Type	Words	Parses	Warnings	Relations	Supports	Supporters	Based On	Basis For	Related To	Edits	hits	Sessions	Editors	Total Time	Com
tathan	Feb 13	tathan	Feb 13			axiomatic?	background	10	4	1	44	44					7	5	2	2	00:00:56	
tathan	Feb 11	tathan	Feb 11	cogbuji	Mar 7	axiomatic	encoding	5	2		40	39		1			5	5	1	2	00:00:16	
tathan	Feb 11	tathan	Feb 11	tathan	Mar 7	axiomatic	encoding	6	2		39	38		1			4	4	1	2	00:00:21	
pvhaley	Jan 2	cogbuji	Jan 8	tathan	Mar 11	axiomatic	source	11	3		36	36					17	8	2	2	00:03:03	
tathan	Jan 4	tathan	Jan 18			axiomatic	encoding	8	2		35	32		3			7	5	2	2	00:01:18	
dwitting	Jan 4	tathan	Mar 11	pfodor	Mar 10	axiomatic	deprecated	10	4		26	2	21		3		18	8	2	4	00:02:09	
tathan	Feb 11	tathan	Feb 11			axiomatic	encoding	4	1		22	20	1	1			3	4	1	2	00:00:27	
tathan	Jan 28	tathan	Feb 1	tathan	Feb 1	axiomatic	background	5	1		20	16	4				17	9	2	2	00:04:59	
tathan	Feb 13	tathan	Feb 13			axiomatic???	background	23	200	3	19	18				1	30	7	1	2	00:10:38	
dwitting	Sep 12	dwitting	Jan 31			axiomatic	background	4	1		17	17					20	10	4	3	00:06:12	
tathan	Jan 29	tathan	Jan 29			axiomatic	background	10	14		17	17					8	5	1	2	00:01:25	
tathan	Jan 16	tathan	Mar 8	tathan	Mar 8	axiomatic	source	13	6		16	4	6		5	1	22	16	4	5	00:04:19	
tathan	Feb 16	tathan	Feb 28	tathan	Mar 7	axiomatic	question	14	106		16		15	1			15	6	3	2	00:04:55	
dwitting	Jan 3	dwitting	Feb 26	tathan	Mar 4	axiomatic	source	13	93		16		15			1	11	7	2	3	00:03:00	
bulicny	Feb 21	bulicny	Feb 26	cogbuji	Mar 7	axiomatic	background	4	1		16	16					3	8	2	3	00:00:06	
bulicny	Jan 9	bulicny	Jan 9			axiomatic	encoding	20	100		15	14		1			52	10	1	3	00:15:24	
tathan	Jan 8	tathan	Jan 8			axiomatic	background	6	1		14	12	2				12	7	1	2	00:08:54	
tathan	Feb 7	tathan	Feb 7			axiomatic?	background	16	140	1	13	12	1				10	5	1	2	00:04:24	
bulicny	Jan 20	dwitting	Feb 19			axiomatic	encoding	12	18		12			1	11		24	10	2	4	00:03:29	
dwitting	Jan 4	tathan	Feb 6	bulicny	Mar 5	axiomatic	source	8	18		12	1	5	1	4	1	10	9	1	3	00:01:22	
dwitting	Jan 4	dwitting	Feb 27	pfodor	Mar 13	axiomatic	source	10	4		11	4	5		2		6	8	2	3	00:05:49	
tathan	Feb 21	tathan	Feb 21			axiomatic	question	8	102		11		3		8		7	6	1	2	00:00:48	
tathan	Jan 8	tathan	Jan 8	tathan	Mar 7	axiomatic	background	13	1		11	6	5				21	6	1	2	00:08:12	
tathan	Feb 1	tathan	Feb 1			axiomatic?	encoding	13	100	1	10		6	4			13	6	1	2	00:07:24	
tathan	Jan 30	tathan	Jan 30			axiomatic?	encoding	12	26	1	10		6	4			18	7	2	2	00:03:38	
cogbuji	Jan 23	cogbuji	Jan 23			axiomatic	question	6	1		10		10				3	4	1	2	00:00:18	
tathan	Jan 8	tathan	Jan 8	tathan	Mar 7	axiomatic	encoding	12	1		10	1	8	1			36	6	1	2	00:10:14	
tathan	Jan 8	tathan	Jan 8	bulicny	Mar 1	axiomatic	encoding	3	1		10	4	4	1		1	2	4	1	3	00:00:07	
tathan	Jan 16	tathan	Mar 8	tathan	Mar 7	axiomatic	source	15	16		9		5		3	1	34	10	3	2	00:08:21	
tathan	Jan 23	tathan	Jan 23			axiomatic?	encoding	12	100	1	9	1	6	2			11	5	1	2	00:02:25	
tathan	Jan 16	tathan	Jan 16			axiomatic	encoding	11	20		9	5	1	1		2	24	6	1	2	00:06:59	
tathan	Jan 8	tathan	Jan 8			axiomatic	background	6	1		9	8	1				3	4	1	2	00:00:16	
dwitting	Sep 3	bulicny	Nov 5	cogbuji	Mar 7	axiomatic	encoding	4	2		9	3	4	2			6	5	1	3	00:00:43	
dwitting	Jan 2	pvhaley	Mar 10	pvhaley	Feb 26	axiomatic	source	9	6		8	1	2	1	3	1	8	22	6	3	00:01:00	
dwitting	Jan 3	tathan	Mar 9			axiomatic	encoding	9	1		8	2	3	3			20	12	4	4	00:02:15	
tathan	Feb 20	tathan	Mar 8	tathan	Mar 8	axiomatic	encoding	16	200		8		7	1			19	8	2	2	00:05:40	
dwitting	Jan 3	tathan	Mar 7	tathan	Mar 7	axiomatic	source	7	6		8	1	3	1	3		4	9	2	4	00:00:21	
dwitting	Jan 3	bulicny	Mar 1	tathan	Mar 8	axiomatic	source	7	25		8		5	2	1		14	9	2	3	00:06:12	
tathan	Feb 21	tathan	Feb 21	cogbuji	Mar 7	axiomatic?	background	5	3	1	8	7	1				5	3	1	2	00:00:14	
tathan	Feb 20	tathan	Feb 20			axiomatic	encoding	15	24		8		7	1			11	8	1	2	00:02:16	
dwitting	Feb 14	dwitting	Feb 14	cogbuji	Mar 7	axiomatic	background	7	6		8	8					12	2	1	2	00:00:59	
tathan	Jan 29	tathan	Jan 29			axiomatic	encoding	20	100		8		7	1			24	4	1	2	00:09:21	
dwitting	Dec 20	cogbuji	Jan 15	cogbuji	Mar 8	axiomatic	encoding	9	1		8		7	1			5	6	1	3	00:00:07	
cogbuji	Aug 24	cogbuji	Jan 15			axiomatic	encoding	5	1		8	6		2			9	7	2	2	00:00:30	
bulicny	Jan 10	bulicny	Jan 10	bulicny	Mar 4	axiomatic	encoding	11	12		8	5		2		1	11	9	1	2	00:01:45	
bulicny	Jan 9	bulicny	Jan 9	pvhaley	Jan 20	axiomatic	encoding	13	100		8	6		1		1	10	20	4	3	00:01:11	

Sentences translated from English to logic

Sentences (2322) | Relations | all types | axiomatic | all editors | 7/ 1/2012

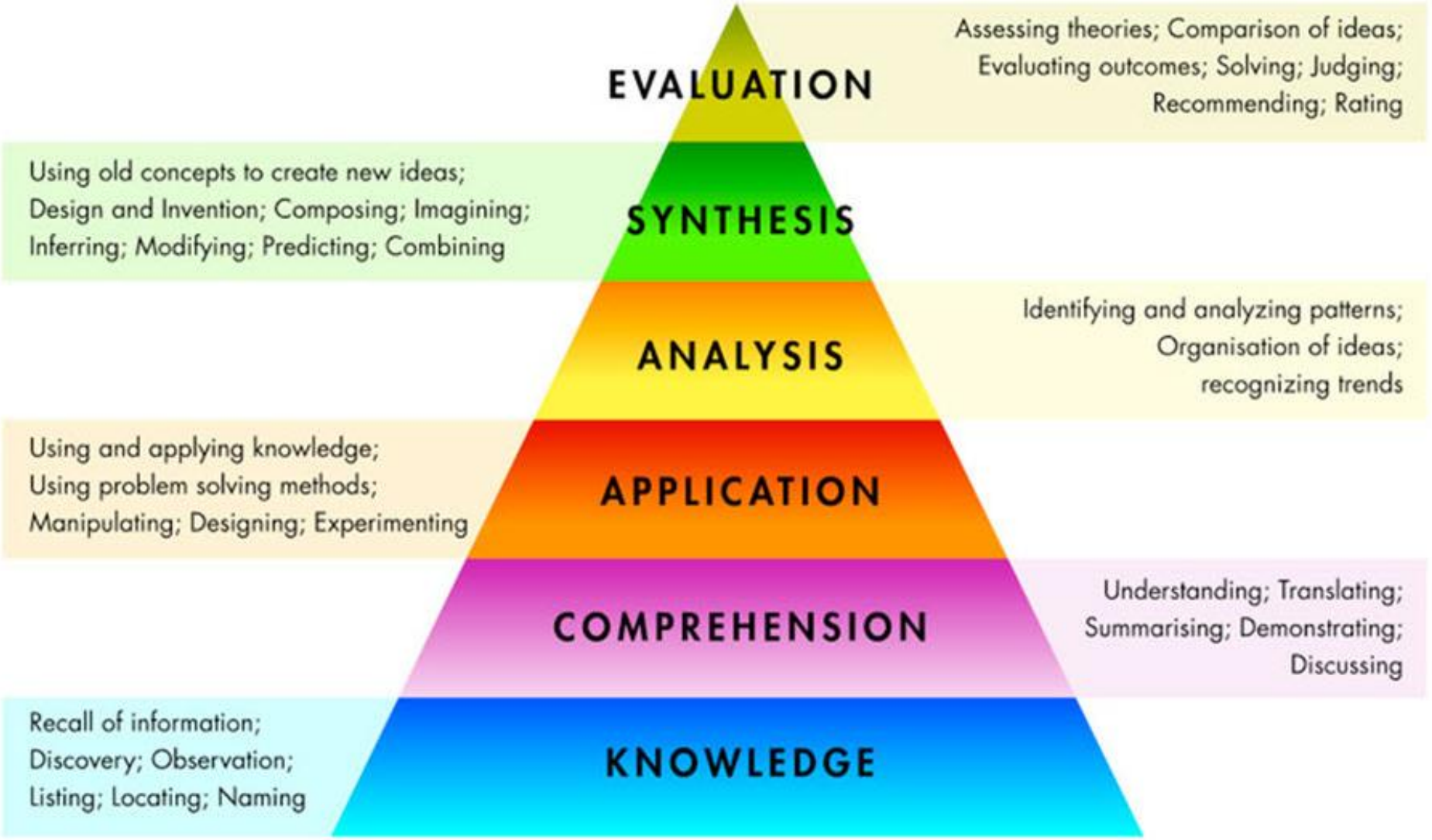
Text	Axiom
The environment of a cell is the solution surrounding it.	∀(?x8)cell(?x8)⇒∀(?x6)environment(of(?x8))(?x6)⇒solution(?x6)∧surround(?x6,?x8)
Enzymes are produced by cells.	∀(?x5)enzyme(?x5)⇒∃(?x8)(cell(?x8)∧produce(?x8,?x5))
An enzyme is a complex protein.	∀(?x6)enzyme(?x6)⇒complex(protein)(?x6)
The endoplasmic reticulum is an organelle of cells in eukaryotic organisms.	∀(?x6)endoplasmic(reticulum)(?x6)⇒∃(?x19)(eukaryotic(organism)(?x19)∧∃(?x14)(cell(in(?x19))(?x14)∧organ
A eukaryotic cell is not a prokaryotic cell.	¬(∃(?x6)(eukaryotic(cell)(?x6)∧prokaryotic(cell)(?x6)))
Diffusion is a result of the constant motion of molecules.	∀(?x8)molecule(?x8)⇒∀(?x5)diffusion(of)(?x5,?x8)⇒∀(?x18)constant(vibration)(of(?x8))(?x18)⇒result(of)(?x5,?
Cholesterol is a steroid.	?x5≡cholesterol→?x5≡steroid
An oxygen molecule is dioxygen.	∃(?x6)(oxygen(molecule)(?x6)∧dioxygen(?x6))
A membrane's permeability to a species is the ratio of its diffusion rate through the membrane to its concentration difference across the membrane.	∀(?x6)membrane(?x6)⇒∀(?x15)species(?x15)⇒membrane(?x6)∧∃₁(?x11)(permeability(of(?x6))(to(?x15))(?x11)
Endocytosis is cellular ingestion.	∃(?x5)(endocytosis(?x5)∧cellular(ingestion)(?x5))
A thing regulates something that it adjusts to some requirement.	∀(?x6)∀(?x8)∃(?x18)(requirement(?x18)∧adjust(to)(?x6,?x8,?x18))⇒regulate(?x6,?x8)
The ability of phospholipids to form membranes is inherent in their molecular structure.	∀(?x8)∃(?x6)(ability(of(?x8))(?x6)∧∃(?e2)(∀(?x22)molecular(structure)(of(?x8))(?x22)⇒in(?e2,?x22)∧inherent(?
Are the tails of phospholipids in a membrane oriented towards the interior of it?	∀(?x14)membrane(?x14)⇒∀(?x22)interior(of(?x14))(?x22)⇒∀(?x9)phospholipid(in(?x14))(?x9)⇒∀(?x4)tail(of(?x
There are two major populations of membrane proteins: integral proteins and peripheral proteins.	∃(?x3)(#(?x3,2)∧major(population)(?x3)∧∃(?x30)(integral(protein)(?x30)∧∃(?x36)(peripheral(protein)(?x36)∧a
An envelope encloses something.	∀(?x6)envelope(?x6)⇒∃(?x8)enclose(?x6,?x8)
A protein is an organic macromolecule that is composed of polymers of amino acids that are connected by peptide bonds.	∀(?x6)protein(?x6)⇒∃(?x15)(∃(?x21)(∃(?x32)(peptide(bond)(?x32)∧amino(acid)(?x21)∧be(connect(to)(with))
A structure has one organizational pattern.	∀(?x6)structure(?x6)⇒∃(?x8)(#(?x8,1)∧organizational(pattern)(?x8)∧have(?x6,?x8))
A direction that is down a gradient is the opposite of the direction of the gradient.	∀(?x9)gradient(?x9)⇒gradient(?x9)∧∃(?x20)(direction(of)(?x20,?x9)∧∀(?x6)direction(down(?x9))(?x6)⇒opposi
A hydrocarbon is an organic chemical compound that comprises carbon and hydrogen.	∀(?x6)hydrocarbon(?x6)⇒∃(?x8)(∃(?x21)(∃(?x27)(carbon(?x27)∧∃(?x31)(hydrogen(?x31)∧and(?x21,?x27,?x31
Passive transport aided by proteins is facilitated diffusion.	∀(?x10)protein(?x10)⇒∀(?x5)aid(?x10,?x5)∧passive(transport)(?x5)⇒facilitated(diffusion)(?x5)
Diffusion is a spontaneous process, needing no input of energy.	∀(?x5)diffusion(?x5)⇒¬∃(?x16)(∃(?x21)(energy(?x21)∧input(of(?x21))(?x16))∧need(?x5,?x16)∧spontaneous(pr
Do white blood cells engulf bacteria through exocytosis?	∃(?x5)(blood(white(cell))(?x5)∧∃(?x15)(bacterium(?x15)∧∃(?x20)(exocytosis(?x20)∧engulf(through(?x20))(?x5
An organizational level of a structure is a level of its organizational pattern.	∀(?x9)structure(?x9)⇒∀(?x6)organizational(level)(of(?x9))(?x6)⇒∀(?x21)organizational(pattern)(of())(?x21)⇒le
Carrier proteins use diffusion of protons into the cell to drive sucrose uptake.	∃₁(?x27)(cell(?x27)∧∃₋(?x15)(∃₋(?x21)(proton(?x21)∧diffusion(of)(into(?x27))(?x15,?x21))∧∃₁(?x33)(sucrose(up
Carrier proteins use proton diffusion into the cell to drive sucrose uptake.	∃(?x5)(carrier(protein)(?x5)∧∃(?x35)(sucrose(uptake)(?x35)∧∃(?x15)(∃(?x29)(cell(?x29)∧proton(diffusion)(int
Do some biological membranes contain cellulose?	∃(?x6)(biological(membrane)(?x6)∧∃(?x9)(cellulose(?x9)∧contain(?x6,?x9)))
An organizational level of supramolecular structures is higher than the molecular level.	∀(?x9)supramolecular(structure)(?x9)⇒∃(?x6)(organizational(level)(of(?x9))(?x6)∧∀(?x17)molecular(level)(?x17
Phospholipids are amphipathic.	∀(?x5)phospholipid(?x5)⇒amphipathic(?x5)
A supramolecular structure is composed of many molecules ordered into a higher level of organization.	∀(?x6)supramolecular(structure)(?x6)⇒∃(?x27)(organization(?x27)∧∃(?x18)(≳(high)(level)(of(?x27))(?x18)∧∃
Lipid bilayers are somewhat permeable to nonpolar particles that are not small.	∀(?x5)lipid(bilayer)(?x5)⇒∀(?x9)¬(small(?x9))∧nonpolar(particle)(?x9)⇒∃(?e2)to(?e2,?x9)∧somewhat(permea
Membrane carbohydrates are attached to proteins or lipids of the membrane.	∀(?x8)membrane(?x8)∧∀(?x5)membrane(carbohydrate)(?x5)⇒∃(?x23)(protein(?x23)∧∃(?x27)(lipid(?x27)∧∃(
An organizational pattern is an arrangement.	∀(?x6)organizational(pattern)(?x6)⇒arrangement(?x6)
Eukaryotic cells contain mitochondria.	∀(?x5)eukaryotic(cell)(?x5)⇒∃(?x9)(mitochondrion(?x9)∧contain(?x5,?x9))
Lipids and proteins are the staple ingredients of membranes.	∀(?x29)membrane(?x29)⇒∃(?x18)(staple(ingredient(of))(?x18,?x29)∧∃(?x5)(∃(?x10)(lipid(?x10)∧∃(?x15)(prot
A supramolecular structure is an assemblage of several molecules.	∀(?x6)supramolecular(structure)(?x6)⇒∃(?x15)(several(molecule)(?x15)∧assemblage(of(?x15))(?x6))
Cellulose is made by enzymes that are located within the plasma membrane of a plant cell.	∀(?x23)plant(cell)(?x23)⇒∃(?x17)(plasma(membrane)(of(?x23))(?x17)∧∃(?x8)(enzyme(?x8)∧be(locate)(within
Proteins are embedded in the phospholipid bilayer.	∀(?x8)phospholipid(bilayer)(?x8)⇒∃(?x5)(protein(?x5)∧be(embed(in))(?x5,?x8))
Membranes must be fluid to function properly.	∀(?x5)membrane(?x5)⇒in(order(to))(must(fluid(?x5)),proper(function)(?x5))
Phagocytosis is engulfment and digestion.	∀(?x5)phagocytosis(?x5)⇒∃(?x8)(∃(?x14)(nominal(engulfment)(?x14)∧∃(?x18)(digestion(?x18)∧and(?x8,?x14,
Enzymes in plasma membranes that make cellulose deposit it on the outer surface of them.	∀(?x9)plasma(membrane)(?x9)⇒∀(?x26)outer(surface)(of(?x9))(?x26)⇒?x15≡cellulose→∀(?x5)enzyme(in(?x9))
Ribosomes carry out the synthesis of protein.	∀(?x5)ribosome(?x5)⇒∃(?x8)(∃(?x13)(protein(?x13)∧synthesis(of)(?x8,?x13))∧carry(out)(?x5,?x8))
Transportation across a cell's membrane of some compound is a key component of the regulation of transport in a cell.	∀(?x13)cell(?x13)⇒cell(?x13)∧∀(?x44)transport(in(?x13))(?x44)⇒∃(?x39)(regulation(of(?x44))(?x39)∧∀(?x8)me
Internal membranes compartmentalize the functions of a eukaryotic cell.	∀(?x5)internal(membrane)(?x5)⇒∃(?x14)(eukaryotic(cell)(?x14)∧∀(?x9)function(of(?x14))(?x9)⇒compartment
Carrier proteins are transport proteins.	∀(?x5)carrier(protein)(?x5)⇒transport(protein)(?x5)
A cell membrane consists of a lipid bilayer with embedded proteins.	∀(?x6)cell(membrane)(?x6)⇒∃(?x8)(∃(?x15)(be(embed)(?x15)∧protein(?x15)∧lipid(bilayer)(with(?x15))(?x8))∧
A bilayer is a double layer of molecules that are closely packed together.	∀(?x6)bilayer(?x6)⇒∃(?x15)(molecule(?x15)∧close(be(pack))(together)(?x15)∧double(layer(of))(?x6,?x15))

Knowledge Acquisition



- Note: the “parse” ranked first by machine learning techniques is usually not the correct interpretation

BLOOMS TAXONOMY



A Bloom level 4 question

- If a Paramecium swims from a hypotonic environment to an isotonic environment, will its contractile vacuole become more active?

∀(?x9)paramecium(?x9)
⇒∃(?x13)(hypotonic(environment)(?x13)
∧∃(?x21)(isotonic(environment)(?x21)
$\wedge\forall_1$(?x31)contractile(vacuole)(of(?x9))(?x31)
⇒if(then)(become(?x31,more(active)(?x31)),swim(from(?x13))(to(?x21))(?x9))))

  - The above formula is translated into a hypothetical query, which answers “No”.

TL KA – Study Results

- **Axiomatized ~2.5k English sentences during 2013:**
  - One defeasible axiom in Rulelog (SILK syntax) per sentence
  - On average, each of these axioms correspond to > 5 "rules"
    - e.g., "rule" as in logic programs (e.g., Prolog) or business rules (e.g., PRR, RIF-PRD)
- **<< 10 minutes on average to author, disambiguate, formalize, review & revise a sentence**
- **The coverage of the textbook material was rated "A" or better for >95% of its sentences**
- **Collaboration resulted in an average of over 2 authors/editors/reviewers per sentence**
- **Non-authors rated the logic for >90% of sentences as "A" or better; >95% as "B+" or better**
- **TBD: How much will TL effort ↑ during QA testing?**
- **TBD: How much will TL effort ↓ as TL tooling & process mature?**

TL KA – Study Results (II)

- **Expressive coverage: very good, due to Rulelog**
  - All sentences were representable but some (e.g., modals) are TBD wrt reasoning
  - This and productivity were why background K was mostly specified via TL
  - Small shortfalls (< few %) from implementation issues (e.g., numerics)

- **Terminological coverage: very good, due to TL approach**
  - Little hand-crafted logical ontology
  - Small shortfalls (< few %) from implementation issues
  - Added several hundred mostly domain-specific lexical entries to the ERG

TL KA: KE labor, roughly, per Page

- **(In the study:)**
- **~~$3-4/word** **(actual word, not simply 5 characters)**
- **~~$500-1500/page** **(~175-350 words/page)**
- **Same ballpark as: labor to author the text itself**
- **… for many formal text documents**
  - E.g., college science textbooks
  - E.g., some kinds of business documents
  - "Same ballpark" here means same order of magnitude
- **TBD: How much will TL effort ↑ when K is debugged during QA testing?**
- **TBD: How much will TL effort ↓ as its tooling & process mature?**

Outline

- Rulelog – more expressive logical KRR / advanced DB
- Textual Logic (TL): English ⟷ logic
- Case study with Demo: Federal Reserve Regulation W
  - Automate Decisions, with full Explanations
- Results from TL Authoring Experiment: Cell Membranes Bio
  - Rapid Collaborative Knowledge Acquisition (KA)
- Conclusions

KA Advantages of Approach

- **Approach = Rulelog + Textual Logic**
- **Rulelog as rich target logic**
  - Can handle exceptions and change, and is tractable
- **Textual terminology: logical ontology emerges naturally**
  - From the text's phrasings, rather than needing effort to specify it explicitly and become familiar with it
  - Perspective: Textual terminology is also a bridge to work in text mining and "textual entailment"
- **Interactive disambiguation: relatively rapidly produces rich K**
  - With logical and semantic precision
  - Starting from effectively unconstrained text
- **Rulelog supports K interchange (translation and integration)**
  - Both LP and FOL; all the major semantic tech/web standards (RDF(S), SPARQL, OWL, RIF, CL, SBVR); Prolog, SQL, and production rules. (Tho' for many of these, with restrictions.)

Conclusions

- **Leveraging recent research breakthroughs on:**

  reasoning with rich knowledge in logic and text

- **Making it practical**
  - **Implement highly expressive reasoning behavior**
  - **Meta knowledge**
  - **Explanations**
  - **Computational efficiency and scalability**
  - **Support English, incl. in authoring too**

- **Appears to be significant progress on the famous "KA bottleneck" of AI**
  - "Better, faster, cheaper" logic. Usable on a variety of KRR platforms.
- **It's early days still, so lots remains to do**
  - Tooling, e.g.: leverage inductive learning
  - More experiments, e.g., scale up

Application Areas

- Financial, Regulatory Compliance
- Intelligence Analysis, Defense, Security
- Health Care, Clinical Guidance
- Education, Science
- E-Commerce
- Policies, Contracts, Legal
- Info Integration, Data Analytics
- Natural Language Processing

Acknowledgements

- Thanks to:
  - The whole Coherent Knowledge Systems team
  - RuleML and W3C, for their cooperation on Rulelog
  - The Project Sherlock team at Automata

- This work was supported in part by Vulcan, Inc., as part of the Halo Advanced Research (HalAR) program, within overall Project Halo. HalAR included SILK. In particular, the KA experiment case study in Biology was supported by Vulcan, Inc. Thanks to:
  - The Project Halo team at Vulcan, incl. the HalAR/SILK team

Thank You

Disclaimer: The preceding slides represent the views of the author(s) only.
All brands, logos and products are trademarks or registered trademarks of their respective companies.