Rulelog: Highly Expressive yet Scalable Logic enabling Cost-Efficient Ontologies†

Benjamin Grosof*

Authored also by Paul Haley**

June 20, 2013
Ontolog Forum‡
Globally accessible webconference session


** Automata, Inc., paul@haleyai.com and Coherent Knowledge Systems http://www.coherentknowledge.com

† Work partly supported by Vulcan, Inc., http://www.vulcan.com
‡ http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2013_06_20

© Copyright 2013 by Benjamin Grosof & Associates, LLC, and Automata, Inc. All Rights Reserved.
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

• **KRR** not defeasible & tractable

• ... even when not target of text-based KA

• **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
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

• **RIF-Rulelog in draft as industry standard** *(RuleML submission to W3C)*

• **Associated new reasoning techniques to implement it**

• **Prototyped in Vulcan’s SILK**
  • Mostly open source: Flora-2 and XSB Prolog
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 et al, AAAI-13 Tutorial]* – largely about Rulelog * preprint/prelim-v. already avail.
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.
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]
# Digital Aristotle and Project Halo

Gunning et al, AAAI & IAAI (August 2011)

## Pilot
- Partial AP Syllabus
  - Chemistry
  - AP Biology
  - Chemistry
  - Physics

## Phase II
- 3 Partial AP Syllabi

## HaloBook
- Single Textbook

## DA
- Complete Domain

### Authors
<table>
<thead>
<tr>
<th>Pilot</th>
<th>Phase II</th>
<th>HaloBook</th>
<th>DA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logic Queries</td>
<td>AP Question Answering</td>
<td>AP QA General QA Education</td>
<td>AP QA General QA Education Research</td>
</tr>
</tbody>
</table>

### Uses
<table>
<thead>
<tr>
<th>Pilot</th>
<th>Phase II</th>
<th>HaloBook</th>
<th>DA</th>
</tr>
</thead>
</table>
IBM Watson FAQ on QA using logic or NLP

• Classic knowledge-based AI approaches to QA try to logically prove an answer is correct from a logical encoding of the question and all the domain knowledge required to answer it. Such approaches are stymied by two problems:
  • the prohibitive time and manual effort required to acquire massive volumes of knowledge and formally encode it as logical formulas accessible to computer algorithms, and
  • the difficulty of understanding natural language questions well enough to exploit such formal encodings if available.

• Techniques for dealing with huge amounts of natural language text, such as Information Retrieval, suffer from nearly the opposite problem in that they can always find documents or passages containing some keywords in common with the query but lack the precision, depth, and understanding necessary to deliver correct answers with accurate confidences.
Why not QA using logic and NLP?

• What if it was “cheap” to acquire massive volumes of knowledge formally encoded as logical formulas?

• What if it was “easy” to understand natural language questions well enough to exploit such formal encodings?
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
<table>
<thead>
<tr>
<th>Text</th>
<th>Atom</th>
</tr>
</thead>
<tbody>
<tr>
<td>The environment of a cell is the solution surrounding it.</td>
<td>Ψ(h8)(cell(h6)⇒Ψ(h6)(environment(h8)(h6)⇒solution(h6)∧ surround(h6,h8)))</td>
</tr>
<tr>
<td>Enzymes are produced by cells.</td>
<td>Ψ(h6)(enzyme(h2)⇒Ψ(h6)(cell(h6)⇒produce(h6,h2,h5)))</td>
</tr>
<tr>
<td>An enzyme is a complex protein.</td>
<td>Ψ(h6)(enzyme(h2)⇒complex(h2)∧protein(h2))</td>
</tr>
<tr>
<td>The endoplasmic reticulum is an organelle of cells in eukaryotic organisms.</td>
<td>Ψ(h6)(endoplasmic(reticulum(h8))⇒Ψ(h6)(eukaryotic(oOrganism(h9))⇒Ψ(h6)(cell(h6)⇒organ(h6)⇒organelle(h8))))</td>
</tr>
<tr>
<td>A eukaryotic cell is not a prokaryotic cell.</td>
<td>Ψ(h6)(eukaryotic(cell(h8))⇒Ψ(h6)(prokaryotic(cell(h8))))</td>
</tr>
<tr>
<td>Diffusion is a result of the constant motion of molecules.</td>
<td>Ψ(h8)(molecule(h8)⇒Ψ(h8)(diffusion(h8)⇒result(h8)(of(h8)⇒constant(vibration(h8)(of(h8))))))</td>
</tr>
<tr>
<td>Cholesterol is a sterol.</td>
<td>Ψ(h6)(cholesterol(h6)⇒sterol(h6))</td>
</tr>
<tr>
<td>An oxygen molecule is dioxygen.</td>
<td>Ψ(h6)(oxygen(molecule(h6))⇒dioxygen(h6))</td>
</tr>
<tr>
<td>A membrane's permeability to a species is the ratio of its diffusion rate through the membrane to its concentration difference across the membrane.</td>
<td>Ψ(h6)(membrane(h6)⇒Ψ(h6)(species(h15)⇒membrane(h6)⇒Ψ(h6)(permeability(h6)(of(h6)⇒Ψ(h6)(to(h15))⇒Ψ(h6)(between(h6))))))</td>
</tr>
<tr>
<td>Endocytosis is cellular ingestion.</td>
<td>Ψ(h6)(endocytosis(h3)⇒cellular(ingestion(h3)))[[1]]</td>
</tr>
<tr>
<td>A thing regulates something that it adjusts to some requirement.</td>
<td>Ψ(h8)(requirement(h8)⇒Ψ(h8)(adjust(to(h8))⇒Ψ(h8)(h6,Ψ(h6)(h8)))⇒regulate(h8,h6,Ψ(h6)(h8))))</td>
</tr>
<tr>
<td>The ability of phospholipids to form membranes is inherent in their molecular structure.</td>
<td>Ψ(h6)(ability(h6)(of(h6)⇒Ψ(h6)(molecular(structure(h6))⇒Ψ(h6)(h22)⇒inherent(h6,Ψ(h8)(h22))))</td>
</tr>
<tr>
<td>Are the tails of phospholipids in a membrane oriented towards the interior of it?</td>
<td>Ψ(h6)(membrane(h6)⇒Ψ(h6)(toward(h6)(Ψ(h6)(tail(h6))))⇒Ψ(h6)(membrane(h6)⇒Ψ(h6)(orientation(h6)(Ψ(h6)(tail(h6))))))</td>
</tr>
<tr>
<td>There are two major populations of membrane proteins integral proteins and peripheral proteins.</td>
<td>Ψ(h6)(membrane(h6)⇒Ψ(h6)(population(h6)⇒Ψ(h6)(major(h6)⇒Ψ(h6)(integral(h6)⇒Ψ(h6)(protein(h6))))⇒Ψ(h6)(integral(h6)⇒Ψ(h6)(protein(h6)))) B⇒Ψ(h6)(peripheral(h6)⇒Ψ(h6)(protein(h6))))</td>
</tr>
<tr>
<td>An envelope encloses something.</td>
<td>Ψ(h6)(envelope(h6)⇒Ψ(h6)(enclose(h6)(of(h6)⇒Ψ(h6)(thing(h8))))))</td>
</tr>
<tr>
<td>A protein is an organic macromolecule that is composed of polymers of amino acids that are connected by peptide bonds.</td>
<td>Ψ(h6)(protein(h6)⇒Ψ(h6)(macromolecule(h6)⇒Ψ(h6)(organic(h6)⇒Ψ(h6)(macromolecule(h6)⇒Ψ(h6)(composition(h6)(Ψ(h6)(polymers(h6)⇒Ψ(h6)(amino(h6)⇒Ψ(h6)(acid(h6))))⇒Ψ(h6)(connected(h6)(of(h6)⇒Ψ(h6)(by(h6))))))))))</td>
</tr>
<tr>
<td>A structure has one organizational pattern.</td>
<td>Ψ(h6)(structure(h6)⇒Ψ(h6)(one(Ψ(h6)(organizational(pattern(h6)⇒Ψ(h6)(h6)))))</td>
</tr>
<tr>
<td>A direction that is down a gradient is the opposite of the direction of the gradient.</td>
<td>Ψ(h6)(gradient(h6)⇒Ψ(h6)(direction(h6)(Ψ(h6)(up(h6))))⇒Ψ(h6)(direction(h6)(Ψ(h6)(down(h6))))))</td>
</tr>
<tr>
<td>A hydrocarbon is an organic compound that comprises carbon and hydrogen.</td>
<td>Ψ(h6)(hydrocarbon(h6)⇒Ψ(h6)(organic(h6)⇒Ψ(h6)(composition(h6)(Ψ(h6)(carbon(h6))))⇒Ψ(h6)(composition(h6)(Ψ(h6)(hydrogen(h6)))))))</td>
</tr>
<tr>
<td>Passive transport aided by proteins is facilitated diffusion.</td>
<td>Ψ(h6)(protein(h6)⇒Ψ(h6)(aided(h6)(by(h6))))⇒Ψ(h6)(facilitated(diffusion(h6))))</td>
</tr>
<tr>
<td>Diffusion is a spontaneous process, needing no input of energy.</td>
<td>Ψ(h6)(diffusion(h6)⇒Ψ(h6)(energy(h21)⇒Ψ(h6)(input(h6)(Ψ(h6)(energy(h21))))⇒Ψ(h6)(spontaneous(h6))))</td>
</tr>
<tr>
<td>Do white blood cells engulf bacteria through endocytosis?</td>
<td>Ψ(h6)(white(h6)⇒Ψ(h6)(blood(h6)⇒Ψ(h6)(cell(h6)⇒Ψ(h6)(esophagus(h6)⇒Ψ(h6)(endoctysis(h3))))))</td>
</tr>
<tr>
<td>An organizational level of a structure is a level of its organizational pattern.</td>
<td>Ψ(h8)(structure(h8)⇒Ψ(h8)(organizational(pattern(h8))⇒Ψ(h8)(organizational(level(h8))))))</td>
</tr>
<tr>
<td>Carrier proteins use diffusion of proteins into the cell to drive sucrose uptake.</td>
<td>Ψ(h6)(carrier(h6)⇒Ψ(h6)(protein(h6))⇒Ψ(h6)(use(h6)(Ψ(h6)(diffusion(h6)(Ψ(h6)(protein(h6))))⇒Ψ(h6)(use(h6)(Ψ(h6)(cell(h6))))⇒Ψ(h6)(drive(h6)(Ψ(h6)(sucrose(h6)⇒Ψ(h6)(uptake(h6))))))))</td>
</tr>
<tr>
<td>Carrier proteins use proton diffusion into the cell to drive sucrose uptake.</td>
<td>Ψ(h6)(carrier(h6)⇒Ψ(h6)(protein(h6))⇒Ψ(h6)(use(h6)(Ψ(h6)(diffusion(h6)(Ψ(h6)(protein(h6))))⇒Ψ(h6)(use(h6)(Ψ(h6)(cell(h6))))⇒Ψ(h6)(drive(h6)(Ψ(h6)(sucrose(h6)⇒Ψ(h6)(uptake(h6))))))))</td>
</tr>
<tr>
<td>Do some biological membranes contain cellulose?</td>
<td>Ψ(h6)(biological(membrane(h6))⇒Ψ(h6)(contains(h6)(Ψ(h6)(cellulose(h6))))))</td>
</tr>
<tr>
<td>An organizational level of supramolecular structures is higher than the molecular level.</td>
<td>Ψ(h6)(supramolecular(structure(h6))⇒Ψ(h6)(organizational(level(h6))⇒Ψ(h6)(organizational(level(h6))))))</td>
</tr>
<tr>
<td>Phospholipids are amphipathic.</td>
<td>Ψ(h8)(phospholipid(h8)⇒Ψ(h8)(amphipathic(h8))))</td>
</tr>
<tr>
<td>A supramolecular structure is composed of many molecules ordered into a higher level of organization.</td>
<td>Ψ(h6)(supramolecular(structure(h6))⇒Ψ(h6)(organizational(level(h6))⇒Ψ(h6)(organizational(level(h6))))))</td>
</tr>
<tr>
<td>Lipid bilayers are somewhat permeable to nonpolar particles that are not small.</td>
<td>Ψ(h6)(lipid(h6)⇒Ψ(h6)(bilayer(h6))⇒Ψ(h6)(somewhat(Ψ(h6)(permeable(h6))))⇒Ψ(h6)(not(Ψ(h6)(small(h6))))))</td>
</tr>
<tr>
<td>Membrane carbohydrates are attached to proteins or lipids of the membrane.</td>
<td>Ψ(h6)(membrane(h6)⇒Ψ(h6)(carbohydrate(h6)⇒Ψ(h6)(linked(h6)(Ψ(h6)(protein(h6))))⇒Ψ(h6)(linked(h6)(Ψ(h6)(lipid(h6))))))))</td>
</tr>
<tr>
<td>An organizational pattern is an arrangement of components.</td>
<td>Ψ(h6)(organizational(pattern(h6))⇒Ψ(h6)(arrangement(h6)(Ψ(h6)(component(h6))))))</td>
</tr>
<tr>
<td>Eukaryotic cells contain mitochondria.</td>
<td>Ψ(h6)(eukaryotic(cell(h6))⇒Ψ(h6)(contain(h6)(Ψ(h6)(mitochondria(h6))))))</td>
</tr>
<tr>
<td>Lipids and proteins are the staple ingredients of membranes.</td>
<td>Ψ(h6)(lipid(h6)⇒Ψ(h6)(protein(h6))⇒Ψ(h6)(ingredients(h6)(Ψ(h6)(membrane(h6))))))</td>
</tr>
<tr>
<td>A supramolecular structure is an assemblage of several molecules.</td>
<td>Ψ(h6)(supramolecular(structure(h6))⇒Ψ(h6)(assemblage(h6)(Ψ(h6)(molecule(h6))))))</td>
</tr>
<tr>
<td>Cellulose is made by enzymes that are located within the plasma membrane of a plant cell.</td>
<td>Ψ(h6)(cellulose(h6)⇒Ψ(h6)(enzyme(h6))⇒Ψ(h6)(be(Ψ(h6)(locate(h6)(Ψ(h6)(within(h6)Ψ(h6)(plasmamembrane(h6))))))))</td>
</tr>
<tr>
<td>Proteins are embedded in the phospholipid bilayer.</td>
<td>Ψ(h6)(protein(h6)⇒Ψ(h6)(be(Ψ(h6)(embedded(h6)(Ψ(h6)(phospholipid(h6))))))))</td>
</tr>
<tr>
<td>Membranes must be fluid to function properly.</td>
<td>Ψ(h6)(membrane(h6)⇒Ψ(h6)(be(Ψ(h6)(fluid(h6))))⇒Ψ(h6)(function(h6))))</td>
</tr>
<tr>
<td>Phagocytosis is engulfment and digestion.</td>
<td>Ψ(h6)(phagocyte(h6)⇒Ψ(h6)(engulfment(h6))⇒Ψ(h6)(消化(h6))))</td>
</tr>
<tr>
<td>Enzymes in plasma membranes that make cellulose deposit it on the outer surface of them.</td>
<td>Ψ(h6)(enzyme(h6)⇒Ψ(h6)(membrane(h6))⇒Ψ(h6)(deposit(h6)(Ψ(h6)(cellulose(h6))))⇒Ψ(h6)(on(h6)(Ψ(h6)(outer(h6))))))</td>
</tr>
<tr>
<td>Ribosomes carry out the synthesis of protein.</td>
<td>Ψ(h6)(ribosome(h6)⇒Ψ(h6)(synthesis(h6)(Ψ(h6)(protein(h6))))))</td>
</tr>
<tr>
<td>Transportation across a cell's membrane of some compound is a key component of the regulation of transport in a cell.</td>
<td>Ψ(h8)(transport(h8)(of(h8)⇒Ψ(h8)(membrane(h8))⇒Ψ(h8)(component(h8)⇒Ψ(h8)(key(h8)⇒Ψ(h8)(component(h8)⇒Ψ(h8)(transport(h8))))))))</td>
</tr>
<tr>
<td>Internal membranes compartmentalize the functions of a eukaryotic cell.</td>
<td>Ψ(h8)(internal(membrane(h8))⇒Ψ(h8)(compartment(h8)(Ψ(h8)(function(h8)(Ψ(h8)(eukaryotic(cell(h8))))))))</td>
</tr>
<tr>
<td>Carrier proteins are transport proteins.</td>
<td>Ψ(h6)(carrier(h6)⇒Ψ(h6)(protein(h6))⇒Ψ(h6)(transport(h6))))</td>
</tr>
<tr>
<td>A cell membrane consists of a lipid bilayer with embedded proteins.</td>
<td>Ψ(h6)(cell(h6)⇒Ψ(h6)(membrane(h6)⇒Ψ(h6)(consists(h6)(Ψ(h6)(lipid(h6))⇒Ψ(h6)(bilayer(h6))))⇒Ψ(h6)(contain(h6)(Ψ(h6)(protein(h6))))))</td>
</tr>
<tr>
<td>A bilayer is a double layer of molecules that are closely packed together.</td>
<td>Ψ(h6)(bilayer(h6)⇒Ψ(h6)(molecule(h6)⇒Ψ(h6)(be(Ψ(h6)(close(h6)(Ψ(h6)(back(h6))))⇒Ψ(h6)(together(h6))))⇒Ψ(h6)(double(h6)))))</td>
</tr>
</tbody>
</table>
Knowledge Acquisition

Note: the “parse” ranked first by machine learning techniques is usually not the correct interpretation.
**Query Formulation**

- Are the passage ways provided by channel proteins hydrophilic or hydrophobic?

<table>
<thead>
<tr>
<th>formula</th>
<th>logic</th>
<th>co-reference</th>
<th>within</th>
<th>inequality</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \leq \text{proteins} (?x16) )</td>
<td>( \forall )</td>
<td>( \forall )</td>
<td>( \leq \text{proteins} (?x16) )</td>
<td></td>
</tr>
<tr>
<td>the('ways'(?x4))</td>
<td>( \forall )</td>
<td>( \forall )</td>
<td>the('ways'(?x4))</td>
<td></td>
</tr>
<tr>
<td>('or')</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \leq \text{channel} (?x23) )</td>
<td>( \exists )</td>
<td></td>
<td>( \leq \text{proteins} (?x16) )</td>
<td></td>
</tr>
<tr>
<td>( \leq \text{passage} (?x9) )</td>
<td>( \exists )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Readings (1)

\[
\forall (?x16) \text{channel(protein) (?x16)} \Rightarrow \forall (?x4) \text{provide} (?x16, ?x4) \land \text{passage(way) (?x4)} \Rightarrow \text{hydrophilic (?x4) } \lor \text{hydrophobic (?x4)}
\]
The Answer is “Hydrophilic”

- Hypothetical query uses “presumption” below
- Presumption yields tuples with skolems
- The answer is on the last line below
Logic to text (not focal in KA experiment)
BLOOMS TAXONOMY

**KNOWLEDGE**
- Recall of information;
- Discovery; Observation;
- Listing; Locating; Naming

**COMPREHENSION**
- Understanding; Translating;
- Summarising; Demonstrating;
- Discussing

**APPLICATION**
- Using and applying knowledge;
- Using problem solving methods;
- Manipulating; Designing; Experimenting

**ANALYSIS**
- Identifying and analyzing patterns;
- Organisation of ideas;
- Recognizing trends

**SYNTHESIS**
- Using old concepts to create new ideas;
- Design and Invention; Composing; Imagining;
- Inferring; Modifying; Predicting; Combining

**EVALUATION**
- Assessing theories; Comparison of ideas;
- Evaluating outcomes; Solving; Judging;
- Recommending; Rating
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))
∧∀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?
KA Advantages of Approach

• Approach = Textual Logic + Rulelog

• Interactive disambiguation: relatively rapidly produces rich K
  • With logical and semantic precision
  • Starting from effectively unconstrained text

• 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”

• Rulelog as rich target logic
  • Can handle exceptions and change, and is tractable

• 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

• Research breakthrough on two aspects:
  
  • 1. rapid acquisition of rich logical knowledge
  
  • 2. reasoning with rich logical knowledge
  
• 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 to aid disambiguation
  
  • More experiments, e.g.: push on QA; scale up
June 2013 News

• Company created to commercialize approach

  Coherent Knowledge Systems

  http://coherentknowledge.com

• Announced at SemTechBiz SF 2013 on June 3, 2013

• Target markets: policy-centric, NL QA and HCI
recap: Scalable Rich KA – Requirements

Text-based
KA & QA
Disambig.

Defeasible & Tractable
Logic KRR
recap: Scalable Rich K – Approach

Textual Logic
Text-based KA & QA

Rulelog
Logic KRR

Logic-based map of text ↔ logic
Textual terminology
Interactive disambiguation

Defeasible + Tractable
Argumentation theories
Restraint
Omniform
Textual Logic

Text-based & KA QA

Logic-based map of text $\leftrightarrow$ logic

Textual terminology

Interactive disambiguation

Defeasible + Tractable Argumentation theories

Restraint

Omniform

Rulelog

Logic KRR

Databases Services

Specialized UI Service interfaces

Usage Context for Approach
Acknowledgements

• 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.

• Thanks to:
  • The HalAR/SILK team
  • The Project Sherlock team at Automata
  • The Project Halo team at Vulcan
  • RuleML and W3C, for their cooperation on Rulelog
Thank You