PowerLoom Overview, Features and Examples

Hans Chalupsky
Project Leader, USC/ISI Loom KR&R Group
Overview

- Logic-based KR&R
  - What is it, advantages, disadvantages
- PowerLoom
  - Quick overview
  - Basic features
  - Tutorial
  - Advanced features relevant for scientific reasoning
  - Debugging failed inferences
- Conclusion
Logic-Based Knowledge Representation & Reasoning
Logic-Based KR&R

- Knowledge Representation & Reasoning
  - Studies a wide variety of paradigms & algorithms for
    - Modeling salient aspects of a world of interest
    - Reasoning with such models

- There are many different modeling paradigms
  - Logic-based, frame-based, graph-based, probabilistic, etc.

- PowerLoom is a logic-based KR&R system

- How do logic-based models represent the world?
Logical Models 101

Terms represent entities:
- **Joe, car002**

Predicates represent relations:
- **owns**

Sentences represent what is true in the world (facts):
- (Person Joe)
- (= (age Joe) 17)
- (Car car002)
- (owns Joe car002)
- (model car002 Ford)
- (not (rich Joe))

Rules define terms and represent domain regularities:
- (<=) (and (> (age ?x) 12) (< (age ?x) 20))
- (Teenager ?x))
- (=> (and (Teenager ?x) (car ?y) (owns ?x ?y))
- (happy ?x))

Facts + rules + inference derive concluded facts:
- (Teenager Joe)
- (happy Joe)
How Does Logic Model the World?

- Terms correspond to entities in the (some) world
- Predicates model properties and relations between entities
- **Domain rules** define and constrain relations, for example, “If Joe is a teenager who owns a car then Joe is happy”
- **Logical inference** rules define the propagation of truth between logical sentences, for example: from X and X => Y it must be true that Y
- The more rules and sentences we add, the higher constrained their “interpretation” (what they could mean) becomes
- However, every consistent theory always has infinitely many (formal) interpretations
Advantages of Logic-based Models

- Tradition
  - Well-understood syntax and semantics
  - Very large amount of relevant research (> 2000 yrs.)
  - Many available logic-based tools
    - Provers, constraint reasoners, learners, planners, KR&R systems, etc.

- Representational power
  - Negation
  - Disjunction
  - Equality (object identity)
  - Logical connectives
  - Quantification
  - Rules, constraints
  - Abstraction
  - Definitions
  - Extendable vocabulary, ontologies
  - “If you can’t say it in logic, you probably don’t want to say it” 😊
Advantages of Logic-based Models

- General purpose, well-understood inference mechanisms
  - Deduction
  - Abduction
  - Induction
  - Constraint satisfaction
  - Automated reasoners
Advantages of Logic-based Models

- **Formalizes reasoning and gives justification**
  - Proofs provide justifications for derived facts
  - If one accepts the premises one must/should accept the conclusions

- **Explanation and understandability**
  - Proofs are a good starting point to provide explanations
  - Logical models are “easy” to understand and interpret (e.g., rules learned by an ILP method)
  - Logical models are easier to debug than other approaches

- **Translatability**
  - Different logical representations are (often) easily translatable into each other (e.g., this diffuses the attribute-vs.-collection distinction)
Disadvantages?

- Disadvantages
  - Difficult to handle uncertainty and probabilistic reasoning
    - But, various efforts to combine logical and probabilistic models (e.g., PRM’s)
  - Complexity of reasoning algorithms
  - Sometimes too expressive, too many different ways of saying the same thing
  - Hard to handle grey areas, but the world is grey
    - Have to make hard decisions (true, false)
    - Hard to say “many”, “few”, “nearly”, etc. (frustrates NLP people)
PowerLoom
PowerLoom KR&R System

- Successor to the successful Loom KR&R system with
  - More expressive representation language
  - Less arcane syntax
  - Better scalability
  - Better portability & extensibility (available in Lisp, C++, Java)
- Based on first-order predicate logic
- Not a description logic but has description logic features
  - E.g., classifier, type and cardinality reasoning, subsumption
- Focus on expressivity + scalability
- Pragmatic stance
  - Usability is more important than theoretical “neatness”
  - Expressivity is more important than inferential completeness
Expressivity vs. Inferential Completeness

- Qualitative comparison of KR-system philosophies
  - Description logics: restricted expressivity, sound, complete, tractable
  - PowerLoom: representationally promiscuous, sound, 80/20 complete & tractable (handle most expected inferences in reasonable time)
Inference Capabilities

- Query engine to retrieve asserted and inferable statements
- Prolog-style backward inference enhanced by
  - Recursive subgoal detection
  - Proper handling of negation
  - Hypothetical reasoning
  - Resource-bounded depth-first or iterative deepening search
  - Proof tree (justification) recording
- Forward inference and simple constraint propagation
- Equality and inequality reasoning
- Subsumption reasoning + relation & instance classification
- Partial match for reasoning with incomplete information
- “WhyNot” abductive inference for query diagnosis
- Extensible reasoning specialists architecture
Knowledge Base Management

- Incremental monotonic and non-monotonic updates
  - Interleave definitions, assertions, retractions with retrieval and inference
  - Truth maintenance via inference caches
- Context mechanism
  - Separate name and assertion spaces with inheritance
  - Provides powerful structuring mechanism for KBs
  - Facilitates scenarios and hypothetical reasoning
- Simple load and save KB mechanism via files
  - Experimental RDBMS persistence in the works
Built-in Ontologies

- **PL-KERNEL-KB**
  - Minimal upper level that defines the core representational vocabulary
    - Relation, concept, function, set, holds, proposition, range-cardinality,…
  - Emphasis on minimality
    - Represent what’s absolutely necessary to make PowerLoom work
    - The less it contains, the less opportunities for modeling conflicts there are when preexisting ontologies get imported
    - Not yet minimal enough, various things still need to be extracted into their own loadable ontology

- **Other available ontologies**
  - Small time ontology: supports time points, durations, temporal arithmetic
  - Units ontology: units support loadable on demand
  - Various translations of upper models (e.g., Cyc, SENSUS), application ontologies (e.g., EELD, Seismology, …)
Tools and APIs

- Ontosaurus KB browser
  - Web-based, dynamic generation of HTML pages viewable in standard browser

- PowerLoom GUI
  - Java-based browse/edit/query environment
  - Client/server based, deployable via Java WebStart in standard browser

- Interactive command-line interface

- Programmatic PowerLoom Interface (PLI)
  - Lisp, C++ and Java bindings

- Lisp-based Loom API
  - Facilitates import of legacy Loom KBs

- OntoMorph translation system
  - Facilitates import of KBs in other languages (e.g. Flogic)

- Initial Semantic Web support
  - Import translator for RDF/RDFS
PowerLoom GUI
Ontosaurus Browser

Relation BASIN-DEPTH-2.5

Textual Definition

Depth to the 2.5 km/s Vs boundary in a basin.

Structured Description

Formal logical encoding of one constraint implied by the textual definition

(Basin-depth = 0m) ⇔ (Vs30 > 2.5 km/s)
Status and Distribution

- Written in STELLA
  - Available in Lisp, C++ and Java

- Current release: PowerLoom 3.0.2.beta
  - Basic KR&R system only
  - Distributed as Lisp, C++ and Java source
  - ~500 downloads world-wide
  - ~400 subscribers to the mailing lists

- Licensing Terms
  - Open Source, user choice of 3 standard licences
    1. GPL
    2. LGPL
    3. Mozilla
PowerLoom Features

- Full-function, robust and stable KR&R system
  - Representation, reasoning, query language, storage, extensive API
  - Available in Java (useful for integration with Protégé)
- Expressivity
  - KR failures often due to “we could not express X”
- Meta-representation & reasoning
  - Concepts, relations, contexts, rules, queries, etc. are all first-class citizens which can be represented and reasoned about
- Explanation support
  - Built-in recording and rendering of proof trees
  - Explanation of failed inferences (“WhyNot”)
- Sophisticated context & module system
  - Encapsulation and organization of knowledge
  - Efficient inference
  - Representation of assumptions: e.g., “all reactions modeled here assume 20°C ambient temperature”
- Sophisticated support for units & measures
PowerLoom Features /2

- Extensible architecture
  - Easy to add new specialized reasoning procedures
- Scalability
  - Caveat: PowerLoom inference is worst-case exponential complexity
  - But: many design features to deal with performance
    - Common inferences (e.g. subsumption) supported by specialists
    - Expensive inference (e.g., classifier) available on demand
    - Various search control directives, e.g., forward/backward-only rules, resource bounded inference
    - Different inference levels
    - Modules to focus reasoning
    - Database interface to offload data-intensive operations onto RDBMS
- Successfully handled very large KBs
  - Reasoned with full Cyc KB (~1,000,000 facts, 35,000 rules)
  - Large EELD ontologies and datasets (not loadable into XSB deductive database)
    - O(1000) ontology & rules, O(10,000) instances, O(100,000) assertions (see example later)
PowerLoom Language Concepts
PowerLoom Representation Language

- PowerLoom language is based on KIF
  - The Knowledge Interchange Format (Genesereth 91)
  - Developed as part of DARPA’s knowledge sharing effort
  - Proposed ANSI standard, now one of the accepted syntaxes of the Common-Logic effort
  - Syntax and declarative semantics for First-Order Predicate Logic
  - Lisp-based, uniform prefix syntax, similar to CycL

Example:

Facts:
- (person fred)
- (citizen-of fred germany)
- (national-language-of germany german)

Rules:
- (forall (?p ?c ?l)
  - (=> (and (person ?p)
    (citizen-of ?p ?c)
    (national-language-of ?c ?l))
  (speaks-language ?p ?l)))
Many extensions to standard FOL:

- Type, set & cardinality relations, e.g., `subset-of`, `instance-of`, `range-cardinality`, etc.
- Second-order definitions via `holds`
- Selective closed-world assumption (OWA is default)
- Classical negation and negation-by-failure
- Defaults (still need work)

Frame-style definition language as syntactic sugar

- `defconcept`, `defrelation`, `deffunction`, `definstance`, `defrule`
- Allows concise definitions but expands internally into standard (more verbose) logic assertions
PowerLoom Knowledge Bases

- Terminology Definitions
  - Concepts (classes), functions, and relations define the vocabulary of a domain, e.g., person, citizen-of, age, etc.

- Assertions
  - Describe what is true in a domain
  - Facts, e.g. (person Fred)
  - Rules, e.g., (forall ?x (=> (rich ?x) (happy ?x)))

- Contexts & Modules
  - Knowledge is organized into modules
  - Facts & rules are not asserted globally but relative to modules, can have different truth values in different modules
  - Hierarchical module structure, assertions from higher modules are inherited to lower modules
Terms, Relations & Propositions

- A KB captures a useful representation of a physical or virtual world
- Entities in the world are modeled in the KB by terms
  - “Georgia”, “Ben Franklin”, 3, “abc”, concept “Person”
- Terms are categorized and related via relations
  - “has age”, “greater than”, “is married to”, “plus”, “Person”
  - Concepts such as “Person” are considered unary relations
- Propositions are sentences with an associated truth value
  - “Ben Franklin is a person”, “Bill is married to Hillary”, “two plus three equals six” (which is false)
- PowerLoom uses KIF terms and sentences to represent propositions
  - (person Ben-Franklin) (married-to Bill Hillary)
  - (= (+ 2 3) 6)
Logical Connectives & Rules

- Predicate logic uses **logical connectives** to construct complex sentences from simpler ones:
  - and, or, not, <=, =>, <=>, quantifiers exists and forall

- Examples:
  - “Richard is not a crook”:
    \[(\text{not (crook Richard)})\]
  - “Every person has a mother”:
    \[(\text{forall } ?p
      \quad (\Rightarrow (\text{person } ?p)
        \quad (\exists ! ?m
          \quad (\text{has-mother } ?p ?m))))\]
Definitions

- Terminology (relations, concepts) need to be defined before they are used via `defconcept`, `deffunction` & `defrelation`.

- Examples:

  ```lisp
  (defconcept person)
  (defrelation married-to ((?p1 person) (?p2 person))
  (deffunction + ((?n1 number) (?n2 number))
      -> (?sum number))
  ```

- Advantage & Disadvantage
  
  - Allows certain amount of error checking (e.g., misspelled relations, argument type violations)
  
  - A bit more tedious and can sometime generate ordering problems
Definition Ordering

- Circular references are only allowed within definitions
- Evaluation of rules within definitions is deferred until query time
- Example:

  \[
  \text{(defconcept parent (?p)}
  \Rightarrow \text{(and (person ?p)}
  \Rightarrow \text{(exists ?c (parent-of ?p ?c))))}
  \]
  \[
  \text{(defrelation parent-of ((?p parent) (?c person)))}
  \]

- Equivalent definition but illegal circular reference:

  \[
  \text{(defconcept parent)}
  \]
  \[
  \text{(assert)}
  \]
  \[
  \text{(forall (?p)}
  \Rightarrow \text{(=> (parent ?p)}
  \Rightarrow \text{(and (person ?p) (exists ?c (parent-of ?p ?c))))})
  \]
  \[
  \text{(defrelation parent-of ((?p parent) (?c person)))}
  \]
Redefinitions

- Definition constructs primarily serve two roles
  1. Convenience; more compact syntax for often used idioms
  2. Linking sets of related axioms to a name to facilitate redefinition
- Redefinition is useful during interactive ontology and KB development

- Example Definition:
  (defrelation parent-of ((?p person) (?c person))
   :=> (relative-of ?p ?c))

- Example Redefinition:
  (defrelation parent-of ((?p parent) (?c person)))

- Result:
  - Redefines parent-of with a different domain
Truth Values

- Each PowerLoom proposition (sentence) is associated with a truth value (relative to a context or module)

- Five possible truth values:
  - true, false, default-true, default-false, unknown

- Standard assertion assigns truth value true
  
  (assert (person Bill))

- Negation asserts truth value false
  
  (assert (not (crook Richard)))

- Presume command asserts default truth values
  
  (presume (=> (bird ?x) (flies ?x)))

- Propositions that are assigned true and false generate a clash (or contradiction)

  - Useful to detect certain constraint violations or errors
  - Used by proof-by-contradiction specialist
  - Contradictory propositions do not bring down the system and are treated as unknown
Changing Truth Values

- The truth value of assertions can be changed
  - Implicitly, by strengthening the truth value, e.g., from default-true to true
  - By explicit retraction of the old truth value and new assertion, e.g.,
    ```lisp
    (assert (not (crook Richard)))
    (retract (not (crook Richard)))
    (assert (crook Richard))
    ```

- Truth values of inferred propositions cannot be retracted

```lisp
(defconcept employee (?e) ::= (person ?e))
(assert (employee Mary))
(ask (person Mary)) ⇒ true
(retract (person Mary))
(ask (person Mary)) ⇒ true
```
Contexts & Modules

- Knowledge is organized into contexts
  - Modules define name-spaces + assertion spaces
  - Worlds define assertion spaces only
- Facts & rules are not asserted globally but relative to modules, can have different truth values in different modules
- Hierarchical module structure, assertions from higher modules are inherited to lower modules
  - Non-monotonic inheritance is possible (e.g., override some inherited assertions for scenario reasoning)
- Contexts are first-class objects that can be asserted to and queried about in the KB
  - Allows attachment of meta-information, e.g., source, assumptions, etc.
- Very efficient, light-weight implementation derived from OPLAN
  - Support built in at a very low level (STELLA)
An Annotated Example
Using Modules

- We define a separate `BUSINESS` module for our example
  - Inherits built-in PowerLoom definitions from `PL-KERNEL/PL-USER`
  - Sets up a separate name and assertion space to avoid unwanted interference with/from other loaded knowledge bases
  - Allows easy experimentation (clearing/changing/editing/saving)
  - All PowerLoom commands are interpreted relative to current module

```
(defmodule "BUSINESS"
    :documentation "Module for the Business demo example."
    :includes ("PL-USER"))

(in-module "BUSINESS")

(clear-module "BUSINESS")
```
Concepts

- Concepts define classes of entities
  - Defined via the `defconcept` command
  - Can have zero or more parent concepts (they all inherit `THING`)
  - Used to introduce typed instances

```
(defconcept company)
(defconcept corporation (?c company))

(assert (company ACME-cleaners))
(assert (corporation megasoft))

(retrieve all ?x (company ?x))
There are 2 solutions:
  #1: ?X=ACME-CLEANERS
  #2: ?X=MEGASOFT

(retrieve all ?x (corporation ?x))
There is 1 solution:
  #1: ?X=MEGASOFT
```
Relations define sets of relationships between entities
- Defined via the defrelation command (& deffunction see later)
- Can have one or more arguments (unary to n-ary)
- Can be fixed or variable arity
- Can be single or multi-valued
- Usually specify types for each argument
- Used to specify relationships between entities

Simple binary relation
- Argument/role variable
- Argument type = domain
- Argument type = range

(defrelation company-name ((?c company) (?name STRING)))
(assert (company-name ACME-cleaners "ACME Cleaners, LTD"))
(assert (company-name megasoftware "MegaSoft, Inc."))
Relations /2

- Retrieve all relations asserted in the BUSINESS module:

(retrieve all (company-name ?x ?y))

There are 2 solutions:
  #2: ?X=ACME-CLEANERS, ?Y="ACME Cleaners, LTD"

(retrieve all (?y ?x) (company-name ?x ?y))

There are 2 solutions:
  #1: ?Y="MegaSoft, Inc.", ?X=MEGASOFT
  #2: ?Y="ACME Cleaners, LTD", ?X=ACME-CLEANERS
Relation Hierarchies

- Hierarchies for concepts as well as relations are supported
  - PowerLoom represents a subconcept/subrelation relationship by asserting an “implication” relation (or an “implies” link)
  - Link is equivalent to a logic rule but allows more efficient inference
  - Various syntactic shortcuts are available to support often-used implication relations

(defrelation fictitious-business-name ((?c company) (?name STRING))) => (company-name ?c ?name))

(forall (?c ?name)
  (=) (fictitious-business-name ?c ?name)
  (company-name ?c ?name))

(subset-of fictitious-business-name company-name)
Relation Hierarchies /2

- Retrieve all names of MegaSoft, fictitious or not
  - Illustrates that company-name is a multi-valued relation

(assert (fictitious-business-name megasoft "MegaSoft"))

(retrieve all ?x (company-name megasoft ?x))
There are 2 solutions:
  #1: ?X="MegaSoft, Inc."
  #2: ?X="MegaSoft"

Directly asserted

Inferred via the subrelation rule/link
Functions are term-producing, single-valued relations

- Defined via the `deffunction` command
- Very similar to relations defined via `defrelation` but:
  - Term producing: a function applied to its first n-1 input arguments specifies a unique, intensional term, e.g., “Fred’s age”
  - Single-valued: each set of input arguments has at most one output argument (the last argument), e.g., “Fred’s age is 42”
- By default, functions are assumed to be partial, i.e., could be undefined for some legal input values (e.g., 1/0)

```
(deffunction number-of-employees ((?c company)) :-> (?n INTEGER))

(assert (= (number-of-employees ACME-cleaners) 8))
(assert (= (number-of-employees megasoft) 10000))
```
Functions /2

Functions syntax often results in shorter expressions than using similar relation syntax:

```
(retrieve all (and (company ?x)
                  (< (number-of-employees ?x) 50)))
```

There is 1 solution:

#1: ?X=ACME-CLEANERS

Compare to:

```
(retrieve all (and (company ?x)
                  (exists ?n
                   (and (number-of-employees ?x ?n)
                        (< ?n 50)))))
```

There is 1 solution:

#1: ?X=ACME-CLEANERS

Multiple function terms:

```
(retrieve all (> (number-of-employees ?x) (number-of-employees ?y)))
```

There is 1 solution:

#1: ?X=MEGASOFT, ?Y=ACME-CLEANERS

Defined Concepts

- Concepts (and functions/relations) can be defined completely in terms of rules.
- Useful to name often-used queries or subexpressions and build up powerful vocabulary.

(\texttt{defconcept small-company} (\texttt{?c company})
  \texttt{:=>} (\texttt{and} (\texttt{company} ?c)
              (\texttt{(< (number-of-employees} ?c) 50))))

(\texttt{forall} ?c (\texttt{=>} (\texttt{and} (\texttt{company} ?c)
               (\texttt{(< (number-of-employees} ?c) 50)))
               (\texttt{small-company} ?c)))

(\texttt{forall} ?c (\texttt{=>} (\texttt{small-company} ?c)
               (\texttt{and} (\texttt{company} ?c)
               (\texttt{(< (number-of-employees} ?c) 50)))))
Defined Concepts /2

Retrieve small companies even if we don’t know exactly how many employees they have

(assert (and (company zz-productions) (< (number-of-employees zz-productions) 20)))

(retrieve all (small-company ?x))

There are 2 solutions:

#1: ?X=ZZ-PRODUCTIONS
#2: ?X=ACME-CLEANERS

All we know is that ZZ Productions has less than 20 employees

Rule-based inference + transitivity of ‘<’
Negation & Open/Closed-World Semantics

- PowerLoom uses classical negation and an open-world assumption (OWA) by default
  - KB is not assumed to be a complete model of the world: if something can’t be derived the answer is UNKNOWN, not FALSE
  - Can distinguish between failure and falsity!
  - Inference engine uses asymmetric effort to derive the truth or falsity of a query
    - Focuses effort on deriving truth, picks up falsity only via quick, shallow disproofs
    - Full effort for falsity available by asking for the negated query
    - Possible extension: 3-valued ask (similar to Loom)

```lisp
(defconcept s-corporation ((?c corporation)))

(ask (s-corporation zz-productions)) ⇒ UNKNOWN
(ask (not (s-corporation zz-productions))) ⇒ UNKNOWN

(assert (not (s-corporation zz-productions)))

(ask (s-corporation zz-productions)) ⇒ FALSE
(ask (not (s-corporation zz-productions))) ⇒ TRUE
```

Due to open-world assumption
Quick disproof from assertion
Negation & Open/Closed-World Semantics /2

- Falsity can also come from sources other than explicit assertion
  - Single-valued functions and relations
  - Inequalities
  - Disjoint types
  - Negated rule heads, etc.

(ask (= (number-of-employees ACME-cleaners) 8)) ⇒ TRUE
(ask (= (number-of-employees ACME-cleaners) 10)) ⇒ FALSE
(ask (not (= (number-of-employees ACME-cleaners) 10))) ⇒ TRUE
(ask (= (number-of-employees zz-productions) 100)) ⇒ FALSE
(ask (= (number-of-employees zz-productions) 10)) ⇒ UNKNOWN

Quick disproof since functions are single-valued
Quick disproof via inequality constraints
Truly unknown since there is not enough information
Selective closed-world semantics and negation-by-failure are also available (as used by Prolog, deductive databases, F-Logic, etc.)

- Useful in cases where we do have complete knowledge
- If something can’t be derived, it is assumed to be false
- Closed-world semantics specified by marking relations as closed
- Negation-by-failure via `fail` instead of `not`

```
(defrelation works-for (?p (?c company)))

(assert (works-for shirly ACME-cleaners))
(assert (works-for jerome zz-productions))

(ask (not (works-for jerome megasoft))) ⇒ UNKNOWN

(assert (closed works-for))
(ask (not (works-for jerome megasoft))) ⇒ TRUE

(retract (closed works-for))
(ask (not (works-for jerome megasoft))) ⇒ UNKNOWN
(ask (fail (works-for jerome megasoft))) ⇒ TRUE
```

Due to open world
Mark relation as closed
Via selective closed-world semantics
Via explicit negation-by-failure
Retraction

Retraction allows the erasure or change of a previously asserted truth-value of a proposition

- Useful for error correction or iterative “change of mind” during development
- Useful to change certain aspects of a scenario without having to reload the whole knowledge base
- Allows efficient, fine-grained change
  - Some cached information is lost and needs to be regenerated
  - Loss can be minimized by careful structuring of module hierarchy (put more stable knowledge higher up in the hierarchy)
- Allows the exploration of hypothetical conjectures
  - What would change if F were true or false?
  - Module system allows us to consider both possibilities at the same time
Retraction /2

Some geographic terminology and information

(defconcept geographic-location)
(defconcept country (\l geographic-location))
(defconcept state (\l geographic-location))
(defconcept city (\l geographic-location))
(defrelation contains (\l1 geographic-location)
    (\l2 geographic-location))

(assert (and
    (country united-states)
    (geographic-location eastern-us)
    (contains united-states eastern-us)
    (state georgia) (contains eastern-us georgia)
    (city atlanta) (contains georgia atlanta)
    (geographic-location southern-us)
    (contains united-states southern-us)
    (state texas) (contains eastern-us texas)
    (city dallas) (contains texas dallas)
    (city austin) (contains texas austin)))
Retraction /3

- Retraction to fix an incorrect assertion

(ask (contains eastern-us texas)) ⇒ TRUE

(retract (contains eastern-us texas))
(assert (contains southern-us texas))

(ask (contains eastern-us texas)) ⇒ UNKNOWN
Value Clipping

- Functions allow implicit retraction via value clipping
  - Assertion of a function value automatically retracts a preexisting value
  - Justified, since functions are single-valued

(deffunction headquarters ((?c company)) :-> (?city city))

(assert (= (headquarters zz-productions) atlanta))
(retrieve all (= ?x (headquarters zz-productions)))
The there is 1 solution:
  #1: ?X=ATLANTA

(assert (= (headquarters zz-productions) dallas))
(retrieve all (= ?x (headquarters zz-productions)))
The there is 1 solution:
  #1: ?X=DALLAS

Assertion automatically clips previous value
DALLAS value replaced ATLANTA
Value Clipping /2

- Clipping also works for single-valued relations

(defrelation headquartered-in ((?c company) (?city city))
  :axioms (single-valued headquartered-in))

(assert (headquartered-in megasoft atlanta))
(retrieve all (headquartered-in megasoft ?x))
There is 1 solution:
  #1: ?X=ATLANTA

(assert (headquartered-in megasoft dallas))
(retrieve all (headquartered-in megasoft ?x))
There is 1 solution:
  #1: ?X=DALLAS
Contradictions

- Propositions that are both **true** and **false** are contradictory
  - Contradictions can result from explicit assertions, during forward-chaining, or as the result of a refutation proof
  - Contradictory propositions are treated as **unknown** to allow the system to continue to function

```lisp
(assert (not (state texas)))
```

Derived both **true** and **false** for the proposition `|P#|(STATE TEXAS)'.
Clash occurred in module `|MDL|/PL-KERNEL-KB/BUSINESS'.

```lisp
(ask (state texas)) ⇒ UNKNOWN
(ask (not (state texas))) ⇒ UNKNOWN
```
Rule-Based Inference

- Logic rules can be used to model complex relationships
- Rules can be unnamed or named via `defrule`
- Most definition commands expand into one or more rules
- Inference engines apply rules to derive conclusions

```
(retrieve all (contains southern-us ?x))
There is 1 solution:
  #1: ?x=TEXAS
```

```
(defrule transitive-contains
  (forall (?l1 ?l2 ?l3)
    (=> (and (contains ?l1 ?l2)
      (contains ?l2 ?l3))
    (contains ?l1 ?l3))))

(retrieve all (contains southern-us ?x))
There are 3 solutions:
  #1: ?x=TEXAS
  #2: ?x=AUSTIN
  #3: ?x=DALLAS
```

Finds only directly asserted values
Defines `contains` to be transitive
Same rule via implicit quantification
Named Rules & Axiom Schemata

- Logic rules can be defined and named via `defrule`
- Rules are propositions which are in the domain of discourse
  - Allows meta-annotations and reasoning
- Naming rules (or any proposition) provides extra level of convenience
- Axiom schemata allow simple definition of commonly used rule patterns

```
(retract transitive-contains)

(retrieve all (contains southern-us ?x))
There is 1 solution:
  #1: ?X=TEXAS

(assert (transitive contains))

(retrieve all (contains southern-us ?x))
There are 3 solutions:
  #1: ?X=TEXAS
  #2: ?X=AUSTIN
  #3: ?X=DALLAS
```

Retract rule by name

Reassert transitivity via meta-relation + axiom schema

```
(defrelation transitive ( (?r RELATION))
  :=>> (and (binary-relation ?r)
             (not (function ?r)))
  :=>> (=> (and (?r ?x ?y)
                 (?r ?y ?z))
           (?r ?x ?z)))
```

Transitivity relation and axiom schema from PL-KERNEL KB
Justifications and Explanation

- Explanation of true/false queries
- Backward inference can store proof trees that can be rendered into explanations
- Simple built-in explanation mechanism
  - Various rendering possibilities, ASCII, HTML, XML
  - Eliminates explanation of duplicate and low-level goals
  - Explanation strings for different audiences (technical, lay)

(ask (contains southern-us dallas)) ⇒ TRUE

(why)
1 (CONTAINS SOUTHERN-US DALLAS)
   follows by Modus Ponens
   since 1.1 ! (FORALL (?l1 ?l3)
      (<= (CONTAINS ?l1 ?l3)
          (EXISTS (?l2)
              (AND (CONTAINS ?l1 ?l2)
                  (CONTAINS ?l2 ?l3)))))
and 1.2 ! (CONTAINS SOUTHERN-US TEXAS)
and 1.3 ! (CONTAINS TEXAS DALLAS)
Explanation /2

- Explanation of retrieved results
  - Separate explanation for each derived solution
  - **why** explains most recently retrieved solution

(retrieve 3 (contains southern-us ?x))

There are 3 solutions so far:

#1: ?X=DALLAS
#2: ?X=TEXAS
#3: ?X=AUSTIN

(why)

1 (CONTAINS SOUTHERN-US AUSTIN)
follows by Modus Ponens
since 1.1 ! (FORALL (?l1 ?l3)
 (<= (CONTAINS ?l1 ?l3)
   (EXISTS (?l2)
    (AND (CONTAINS ?l1 ?l2)
      (CONTAINS ?l2 ?l3)))))

and 1.2 ! (CONTAINS SOUTHERN-US TEXAS)
and 1.3 ! (CONTAINS TEXAS AUSTIN)
Hypothetical or scenario reasoning can be achieved by
creating a new context which inherits existing set of facts and
allows the exploration of "assumptions".
In this example, we show how certain inherited assertions can be
retracted and changed.

(defmodule "ALTERNATE-BUSINESS"
 :includes "BUSINESS")

(in-module "ALTERNATE-BUSINESS")

(assert (and (company web-phantoms)
             (company-name web-phantoms "Web Phantoms, Inc.")))

(retract (company-name megasoft "MegaSoft, Inc.")
(assert (company-name megasoft "MegaZorch, Inc.")))
The ALTERNATE-BUSINESS module

- inherits all of the information of its parent module
- is subject to the specific changes made in the local module.

(in-module "BUSINESS")

(retrieve all (company-name ?x ?y))
There are 3 solutions:
   #2: ?X=ACME-CLEANERS, ?Y="ACME Cleaners, LTD"
   #3: ?X=MEGASOFT, ?Y="MegaSoft"

(in-module "ALTERNATE-BUSINESS")

(retrieve all (company-name ?x ?y))
There are 4 solutions:
   #3: ?X=ACME-CLEANERS, ?Y="ACME Cleaners, LTD"
   #4: ?X=MEGASOFT, ?Y="MegaSoft"

Changed local assertion

New local assertion

From “fictitious business name” assertion
Cross-Contextual Reasoning

- Normally queries operate in the current module.
- The IST (IS-TRUE) relation (J. McCarthy) allows us to query about the state of knowledge in other modules.
- This also allows cross-module inference by binding variables across forms
- Example: “find all companies whose names differ in the two modules”

```
(in-module "BUSINESS")

(retrieve all (ist alternate-business (company-name ?x ?y)))
There are 4 solutions:
  #3: ?X=ACME-CLEANERS, ?Y="ACME Cleaners, LTD"
  #4: ?X=MEGASOFT, ?Y="MegaSoft"

(retrieve all (and (ist business (company-name ?x ?y))
  (fail (ist alternate-business (company-name ?x ?y)))))
There is 1 solution:
RDBMS to PowerLoom Mapping

Defining a PowerLoom database instance edb1:

```
(DEFDB edb1
   :dsn "EELD_EDB17JUN03_COMPLIANT_PUBLIC_PL"
   :user "scott" :host "blackcat.isi.edu")
```

Defining a PowerLoom relation EDB-Person that maps onto the EDB table Person:

```
(DEFTABLE EDB-Person edb1 "PERSON"
  (?ENTITYID ?REPORTID ?SOURCEID
```

Defining a PowerLoom lifting axiom that maps the ontology concept Person onto EDB-Person:

```
(ASSERT
   (=> (QUERY
   ;MATCH-MODE :EELD :HOW-MANY :ALL)
   (Person ?p)))
```
Advanced Topics
Concept Definitions

Define new concept term `event`

```
(defconcept event
   :documentation "The class of events."
)
```

Documentation string

Define `movement-event` as subconcept of `event`

```
(defconcept movement-event (?ev event)
   :documentation "The class of movement events."
   :=> (= (* (time ?ev) (speed ?ev))
        (distance ?ev)))
```

Constraint rule in KIF syntax
Definitions are Syntactic Sugar

- Constructs such as `defconcept` facilitate concise expression of commonly needed definition tasks
- Example: Previous definitions expand into the following more verbose set of assertions:

```lisp
(defconcept event)
(defconcept movement-event)
(assert (documentation movement-event "The class of movement events."

(assert (forall ?ev
  (=> (movement-event ?ev)
      (event ?ev))))

(assert (forall ?ev
  (=> (movement-event ?ev)
      (= (* (time ?ev) (speed ?ev))
          (distance ?ev))))
```

Meta assertion about the concept `movement-event`

Represents the subconcept relationship

Represents the constraint
Relation Definitions

Define binary relation sub-event with domain and range event:

(defrelation sub-event ((?sub event) (?super event))
 :documentation “Links a sub-event to its super-event.”)

Relations can have arbitrary as well as variable arity.

Functions are single-valued, term-generating relations:

(deffunction time ((?ev movement-event) ?time)
 :documentation “The duration of a movement event.”)

(deffunction speed ((?ev movement-event) ?speed)
 :documentation “The speed of a movement event.”)

(deffunction distance ((?ev movement-event) ?distance)
 :documentation “The distance covered by a movement.”)
Define event instance \texttt{ev1} with various properties:

\begin{verbatim}
(definstance ev1
  :movement-event true :speed 10 :time 20)
\end{verbatim}

The above concise, frame-style definition expands into the following individual assertions:

\begin{verbatim}
(assert (movement-event ev1))
(assert (= (speed ev1) 10))
(assert (= (time ev1) 20))
\end{verbatim}

Function term “time of event \texttt{ev1}”
Rules are also in the domain of discourse and can be named for easy attachment of other information

For example, documentation strings or explanation templates:

```lisp
(defun speed-rule
  (forall ?ev
    (= (movement-event ?ev)
      (= (* (time ?ev) (speed ?ev))
        (distance ?ev)))
  :documentation "Distance = time * speed of a movement."
)

(assert
  (explanation-template speed-rule
    "The distance ?(distance ?ev) of the movement event ?ev equals its duration ?(time ?ev) times its speed ?(speed ?ev).")
)
Arithmetic Constraint Reasoning

- Arithmetic reasoning is important for scientific, engineering and everyday reasoning
- PoweLoom’s built-in arithmetic specialists can compute a result from any two bound arguments
  - Allows us to model this formula via a single “speed” rule (instead of three – was an issue with Cyc in Phase-1)
  - Example: \((+ \ 5 \ ?x \ 2) \Rightarrow ?x = -3\)

(definistance ev1 :movement-event true :speed 10 :time 20)
(definistance ev2 :movement-event true :speed 10 :distance 50)

(retrieve all (distance ev1 ?x))
There is 1 solution:
  #1: ?X=200

(retrieve all (time ev2 ?x))
There is 1 solution:
  #1: ?X=5
General Queries

- Many other systems have problems with general queries that ask about classes of things:
  - “Is it true that the ionization of diluted solutions is higher than those of concentrated solutions?”
  - This is often worked around by introducing a specific solution individual and asking the question about the instance

- In PowerLoom we can ask the universal question directly:

```lisp
(defconcept solution :documentation "The class of chemical solutions.")
(deffunction concentration-level ((?s solution) ?level) :documentation "Concentration ?level of some particular solution ?s.")
(deffunction ionization ((?s solution) ?level) :documentation "Ionization ?level of some particular solution ?s.")
(defrelation greater-than (?x ?y) :documentation "Qualitative `>' relation.")
```
Two (mock) rules describing relationships between concentration levels of solutions and their ionization level

Note that these rules operate at the instance level: given a specific solution instance and its concentration, we can infer the solution’s ionization level

(defrule ionization-rule1
  (=> (and (solution ?x)
            (concentration-level ?x diluted))
       (= (ionization ?x) high))
  :documentation "Diluted solutions have high ionization."
)

(defrule ionization-rule2
  (=> (and (solution ?x)
            (concentration-level ?x concentrated))
       (= (ionization ?x) low))
  :documentation "Concentrated solutions have low ionization."
)

(assert (greater-than high low))

Qualitative ordering
General Queries /3

- Can phrase the general query directly as a universally quantified statement
  - PowerLoom’s Universal Introduction reasoner is used to prove it
  - Automatically introduces hypothetical solution individuals with the necessary properties in a hypothetical world

\[(\text{ask } (\forall x \forall y) (\Rightarrow (\text{and } (\text{solution } x) (\text{solution } y) (\text{concentration-level } x \text{ diluted}) (\text{concentration-level } y \text{ concentrated}) (\text{greater-than } (\text{ionization } x) (\text{ionization } y)))))) \Rightarrow \text{TRUE}\]

- Contrast this with the “hand-reification” approach:

\[(\text{assert } (\text{and } (\text{solution sol1}) (\text{concentration-level sol1 diluted})))
(\text{assert } (\text{and } (\text{solution sol2}) (\text{concentration-level sol2 concentrated})))

(\text{ask } (\text{greater-than } (\text{ionization sol1}) (\text{ionization sol2}))) \Rightarrow \text{TRUE}\]
Representing Queries

- Question answering applications
  - Reasoning about the queries itself is often important (e.g., answering a multiple-choice question by identifying the incorrect answers)

- PowerLoom can represent queries as terms to facilitate such query-level reasoning
  - Example: Wh-query can be represented via a KAPPA term and then evaluated via the query engine to generate the answers

```
(deffunction wh-query (?q) :-> (?kappa SET))

(assert (wh-query q1
  (kappa ?x
    (and (solution ?x) (concentration-level ?x diluted)))))

(assert (and (solution s1) (concentration-level s1 diluted)))
(assert (and (solution s2) (concentration-level s2 concentrated)))
(assert (and (solution s3) (concentration-level s3 diluted)))

(retrieve all ?x (member-of ?x (wh-query q1)))
There are 2 solutions:
#1: ?X=S3
#2: ?X=S1
```
Units and Dimensions

- Scientific reasoning uses various units and dimensions
- PowerLoom has full support for units
  - Large number of predefined units
    - SI and other measurement systems
    - Fundamental quantities: mass, distance, time, angle, solid angle, amount of substance, electric current, luminous intensity, data
- Arithmetic operations on units
- Arbitrary combinations of units introduced by formulae
  - not limited to predefined combinations
- Extensible via STELLA code

Integration with Ontology
- Datatype introduced via the `units` function
- All logical operations and inferences work with units
Reasoning with Units
Assertions and Comparisons

- The `units` function introduces the data types
  - Assertions
  - Comparisons

```lisp
(assert (= (age Fred) (units 35 "yr")))
(assert (= (age Pebbles) (units 18 "month")))
(assert (= (answer problem) (units 42 "m2kg/s^3")))
```

```lisp
(ask (< (units 10 "mm") (units 10 "ft")))
TRUE
(ask (< (units 10 "ft") (units 11 "m")))
TRUE
(ask (< (units 11 "m") (units 10 "ft")))
FALSE
```

```lisp
(ask (< (units 11 "kg") (units 10 "ft")))
UNKNOWN
```

- Magnitude of expression
- Units (as string)
- Arbitrary unit combinations
- Comparisons normally give true or false answers
- Incompatible units, so no meaningful answer
Reasoning with Units
Conversions

- Conversions
- All units are stored internally in canonical form (SI mks)
- Conversions are performed on input or output

(retrieve (= (units ?x "mile") (units 100 "km")))
There is 1 solution so far:
  #1: ?X=62.13711922373341
(retrieve (= (units ?x ?y) (units 100 "km")))
There is 1 solution so far:
  #1: ?X=100000.0, ?Y="m"
(retrieve all (= (age Fred) (units ?x "yr")))
There is 1 solution:
  #1: ?X=35.0
(retrieve (= (units 1000 ?y) (units 1 "km")))
No solutions.

Miles to kilometers
Both magnitude and unit (in canonical units)
More useful example
Too open-ended
Reasoning with Units
Arithmetic

- Arithmetic
  - Units combine appropriately
  - Arbitrary units combinations

(retrieve (= (units ?x ?y) (u-div (units 20 "km") (units 1 "h"))))
There is 1 solution so far:
  #1: ?X=5.555555555555555, ?Y="m/s"

(retrieve (= (units ?x "km/h") (u-div (units 20 "km") (units 1 "h"))))
There is 1 solution so far:
  #1: ?X=20.0

(retrieve (= (units ?x "km") (u* (units 20 "km/h") (units 1.5 "hr"))))
There is 1 solution so far:
  #1: ?X=30.0

? (retrieve (= (units ?x ?y) (u-div (units 1 "h") (units 20 "km"))))
There is 1 solution so far:
  #1: ?X=0.18, ?Y="s/m"

Creates 20 km/h unit, a common unit
Canonical internal representation.
Converted back to km/h
Synonyms
Creates 20h/km, a quite uncommon unit
Time Points and Durations

- Time is an important aspect of the world
- PowerLoom has support for exact time points and durations
  - Time point specification uses flexible strings and `timepoint-of` function
    - ISO-8601 extended format for dates
    - Time zones are specified numerically as offset from UTC (i.e., what you add to UTC to get local time)
      Common time zone strings are also supported: UTC, Z, PST, EDT
  - Duration uses simple strings of days and milliseconds and `duration-of` function
    - "plus 5 days; 85000 ms", "minus 3 days; 0 ms"
    - Integrated and interchangeable with `units` function
- Arithmetic operations on time points and durations
- Comparisons of time points or durations
- Integration with Ontology
  - Datatypes introduced via the `timepoint-of` and `duration-of` functions
  - All logical operations and inferences work with time points and durations
  - Durations interoperate with the `units` function
Reasoning with Time

Assertions and Comparisons

- The `timepoint-of` function introduces time points and the `duration-of` function introduces durations

- Assertions
  
  ```lisp
  (assert (= (birthday Fred) (timepoint-of "2001-Jan-8 7:00Z")))
  (assert (= (duration Project85) (duration-of "180 days")))
  (assert (= (duration Concert75) (duration-of "0 days; 7200000 ms")))
  ```

- Comparisons
  
  ```lisp
  (ask (< (timepoint-of "2005-Jul-3")
          (timepoint-of "2005-Jul-4")))
  TRUE
  
  (ask (< (timepoint-of "2005-07-03T12:30Z")
          (timepoint-of "2005-07-03T18:30+8:00")))
  FALSE
  
  (ask (< (timepoint-of "2006-10-May")
          (duration-of "2 days")))
  UNKNOWN
  ```
Reasoning with Time Durations
Conversions

- **Conversions**
  - Durations and time units can be converted between each other.
  - Time points can be destructured using the `timepoint-of*` function

\[
\text{(retrieve } (\text{= (duration-of } ?x) \text{ (units 10 "day")))}\]

There is 1 solution so far:
- #1: \(?X=\text{"plus 10 days; 0 ms"}\)

\[
\text{(retrieve } (\text{= (duration-of } ?x) \text{ (units 2 "h")))}\]

There is 1 solution so far:
- #1: \(?X=\text{"plus 0 days; 7200000 ms"}\)

\[
\text{(retrieve all } (\text{= (duration Concert75) (units } ?x \text{ "h")))}\]

There is 1 solution:
- #1: \(?X=2.0\)

\[
\text{(retrieve all } (\text{= (duration Concert75) (units } ?x \text{ ?y)))}\]

There is 1 solution:
- #1: \(?X=7200.0, \ ?Y=\text{"s"}\)

\[
\text{(retrieve } (\text{= (timepoint-of* } ?y \text{ ?m ?d ?hh ?mm ?ss "PST"}) \text{ (timepoint-of "Feb/5/2002 00:25:30 EST")))}\]

There is 1 solution so far:
- #1: \(?Y=2002, \ ?M=2, \ ?D=4, \ ?HH=21, \ ?MM=25, \ ?SS=30.0\)
Reasoning with Time Arithmetic

- Arithmetic
  - Addition and Subtraction of points and durations
  - Time types combine appropriately

(retrieve all (time- (timepoint-of "2006-10-20")
               (timepoint-of "2006-10-15")
               (duration-of ?interval)))

There is 1 solution:
  #1: ?INTERVAL="plus 5 days; 0 ms"

(retrieve all (time+ (timepoint-of "2006-12-25")
                   (duration-of "12 days")
                   (timepoint-of ?date)))

There is 1 solution:
  #1: ?DATE="2007-JAN-06 7:00:00.000 UTC"

(retrieve all (time+ (duration-of "12 days")
                   (units ?n "h")
                   (duration-of "14 days")))

There is 1 solution:
  #1: ?N=48.0

Amount of time between two time points
Adding a duration to a time points to get a new time point
Works together with units
Output variable can be in any position
Forward Inference (FIX)

- Simple propositional reasoning, e.g.,
  \((\text{or } p \ q), \neg p \models q\)
- Simple equality reasoning
- Forward skolemization, e.g.,
  \((\forall x \text{ Person}) (\exists y \text{ Person}) (\text{mother-of } x y))\)
WhyNot Query Debugging
Debugging Queries in Large KBs

- Logic-based knowledge representation & reasoning system
  - Use language of some logic L to represent knowledge (e.g., KIF)
  - Use implementation of proof procedure for L as reasoning engine

- Some (partially) developed knowledge base:

Facts:
- (person fred)
- (citizen-of fred germany)
- (national-language-of germany german)

Rules:
- (forall (?p ?c ?l)
  => (and (person ?p)
           (citizen-of ?p ?c)
           (national-language-of ?c ?l))
  (speaks-language ?p ?l))

Queries:
- (speaks-language fred german)? \implies TRUE
- (speaks-language fred french)? \implies UNKNOWN
The Problem

Failed query:  \((\text{speaks-language} \ \text{fred} \ \text{french})?)\quad \rightarrow \text{UNKNOWN}

Diagnosis is **simple**: the query failed because:

- Not asserted as a fact
- Not enough facts to infer it via the known rule
- Open world assumption!

Ask similar query in Cyc (1,000,000 facts, 35,000 rules)

- “Does Doug Lenat speak German?”
- Answer: **UNKNOWN**

Diagnosis is **very difficult**: the query failed because

- Not asserted as a fact
- All attempts to infer the answer failed
  - Search tree explored by Cyc is **very large**, timeout at 30 CPU seconds
- Hard to debug for knowledge engineers, impossible for non-experts
Solution: Explaining Query Failures via Plausible Partial Proofs

- Standard explanation technique for logic-based reasoners:
  - Find and record a proof for the query
  - Present the proof to the user in an understandable way
  - Problem: No proof ⇒ no explanation

- Solution: Need to create a proof even though none could be found

- Generate plausible partial proofs for a query
  - Partial proofs can be explained
  - Proof "holes" identify potential knowledge or inference gaps
  - Multiple plausible partial proofs to explain different failure modes
  - Top-ranked partial proofs focus on most plausible failures

- Challenges:
  - What is a plausible partial proof?
  - Scaling, find plausible proofs without looking at too many options
    ⇒ there are infinitely many possible proofs!
Example Explanation of Query Failure

Explanation #1 score=0.708:

1 \(\text{(speaks-language fred german)}\)
   is true to some part because an if-then rule applies
   with substitution \{?p/fred, ?l/german, ?l/phil\}
   since 1.1 Rule:  \(\forall p \exists f \text{(speaks-language } p f)\)
   \(\exists f \text{(parent-of } p f)\)
   \(\text{(native-language-of } f l)\))
   and 1.2 Fact: \(\text{(parent-of fred phil)}\)
   and 1.3 Inferred: \(\text{(native-language-of phil german)}\)

1.3 \(\text{(native-language-of phil german)}\)
   is true to some part because an if-then rule applies
   with substitution \{?p/phil, ?l/german, ?c/germany\}
   since 1.3.1 Rule:  \(\forall p \exists c \text{(native-language-of } p c)\)
   \(\forall c \text{(person } p)\)
   \(\text{(birth-place-of } p c)\)
   \(\text{(national-language-of } c l)\))
   and 1.3.2 Fact: \(\text{(person phil)}\)
   and 1.3.3 Unknown: \(\text{(birth-place-of phil germany)}\)
   and 1.3.4 Fact: \(\text{(national-language-of germany german)}\)
PowerLoom’s “WhyNot” Query Diagnosis Tool

- **PowerLoom KR&R system**
  - First-order-logic-based KR&R system
  - Representation language is KIF (variant of FOL)
  - Natural deduction reasoner combining forward, backward reasoning plus variety of reasoning specialists
  - Type & cardinality reasoning, relation subsumption, classifier
  - Selective closed-world reasoning
  - Modules and light-weight worlds for hypothetical reasoning

- **“WhyNot”** built into inference engine of PowerLoom
  - Partial inference mode to generate plausible partial proofs
  - Score propagation instead of truth-values
  - Various plausibility heuristics

- PowerLoom explanation component used to explain partial proofs

- Only diagnosis of missing facts at the moment
“WhyNot” Plug-in to Debug Queries in Large Cyc Knowledge Bases

- Cyc-based KRAKEN KA Tool
  - input and output in natural language
  - very large amount of background knowledge (over 1,000,000 facts, O(10,000) rules)
  - query diagnosis is very difficult

- PowerLoom “WhyNot”
  - external knowledge source integrated via blackboard
  - dynamically fetches and translates Cyc knowledge
  - performs partial inference against very large KB
  - pinpoints potential knowledge gaps
  - ships explanations to KRAKEN UIA display
ISI Why Not

Query Critique from ISI's WhyNot Module

For the question

Does Doug Lenat speak the German language?

the Why-Not module has determined that if any of the following facts were known, the query would succeed.

- [F] ?X speaks the German language.
- [F] Doug Lenat is a child of ?X.

Explanation Detail:

Explanation 1, score=0.50:

1 Doug Lenat speaks the German language.
   is true to some part because an f-fthen rule applies
   with substitution {??Doug Lenat, Z?High German, X?one of Kurt Godel, Georg Cantor,
   Johann Sebastian Bach, etc}

   since 1.1 Rule: If Y is a child of X and X speaks Z then Y speaks Z.
   and 1.2 Unknown: Doug Lenat is a child of X.
   and 1.3 Fact: ?X speaks the German language.

Legend:
- The color green marks rules and facts that are known either because they were asserted or they follow by some logical inference.
- The color yellow marks logically inferred facts that are only partly true, since their inferences depend on some unknowns.
- The color red marks completely unknown facts that might be missing from the knowledge base (if they actually do make sense). Asserting all of them would allow Cyc to answer the
Does Doug Lenat Speak German?

Query Critique from ISI’s WhyNot Module

For the question

Is it true that Douglas Lenat speaks the German language?

the Why-Not module has determined that if any of the following facts were known, the query would succeed.

- [+] Douglas Lenat is a child of Kurt Godel.

Explanation 1, score=0.50:

1. Douglas Lenat speaks the German language.
   is true to some part because an if-then rule applies
   with substitution \{Y/Douglas Lenat, Z/the German language, X/Kurt Godel\}
   since 1.1 If some intelligent agent Y is a child of some intelligent agent X and X speaks some natural language, then Y speaks the natural language.
   and 1.2 Douglas Lenat is a child of Kurt Godel.
   and 1.3 Kurt Godel speaks the German language.

Many similar explanations
Need to generalize
Improved Explanation by Generalizing Similar Proofs

For the question

Does Doug Lenat speak the German language?

the Why-Not module has determined that if the following facts were known, the query would succeed.

- [+]? X speaks the German language.
- [+]? Doug Lenat is a child of ?X.

[Continue]

Explanation Detail:

Explanation 1, score=0.50:

1 Doug Lenat speaks the German language.
   is true to some part because an if-then rule applies
   with substitution {Y/Doug Lenat, Z/High German, X/one of Kurt Godel, Georg Cantor,
   Johann Sebastian Bach, etc.}
   since 1.1 Rule: If Y is a child of X and X speaks Z, then Y speaks Z.
   and 1.2 Unknown: Doug Lenat is a child of X.
   and 1.3 Fact: X speaks the German language.
Alternative Lower-Score Explanation

- Explanation 2 (score 0.38):

For the question

Does Doug Lenat speak the German language?

the Why-Not module has determined that if the following facts were known, the query would succeed.

- [+] ?X speaks the German language.
- [+] Julius Caesar is a child of ?x.
- [+] Augustus and Doug Lenat are siblings.
Example Explanation 2 Detail (score 0.38)

Explanation 2, score=0.38:

1. **Doug Lenat speaks the German language.**
   is true to some part because an if-then rule applies
   with substitution \{Y/Doug Lenat, Z/High German, X/Julius Caesar\}
   since 1.1 Rule: \( Y \) is a child of \( X \) and \( X \) speaks \( Z \), then \( Y \) speaks \( Z \).
   and 1.2 Inferred: Doug Lenat is a child of Julius Caesar.
   and 1.3 Inferred: Julius Caesar speaks the German language.

2. **Doug Lenat is a child of Julius Caesar.**
   is true to some part because an if-then rule applies
   with substitution \{Y/Julius Caesar, Z/Doug Lenat, X/Augustus\}
   since 1.2.1 Rule: \( X \) is a child of \( Y \) and \( X \) and \( Z \) are siblings, then \( Z \) is a child of \( Y \).
   and 1.2.2 Unknown: Augustus and Doug Lenat are siblings.
   and 1.2.3 Fact: Augustus is a child of Julius Caesar.

3. **Julius Caesar speaks the German language.**
   is true to some part because an if-then rule applies
   with substitution \{Y/Julius Caesar, Z/High German, X/\{one of Kurt Godel, Georg Cantor, Johann Sebastian Bach, etc.\}\}
   since 1.1 Rule: \( Y \) is a child of \( X \) and \( X \) speaks \( Z \), then \( Y \) speaks \( Z \).
   and 1.3.1 Unknown: Julius Caesar is a child of \( X \).
   and 1.3.2 Fact: \( X \) speaks the German language.
Partial Inference Application #2: Pattern Matching for Link Discovery (EELD)

- Link Discovery Problem
  - **Given**: large amounts of evidence
    - people, organizations, places, events, relationships, accounts, transactions, etc.
  - **Discover**: high-level activities of interest
    - Contract murders, gang wars, industry takeovers, terrorist activities, etc.

- KOJAK approach:
  - Represent evidence as large-scale PowerLoom evidence KBs
  - Represent domain knowledge via logic rules
  - Represent patterns via logic rules and queries
  - Use **partial inference** to detect patterns of interest

- Challenges:
  - Scale, incompleteness, noise, corruption
Example: Using “WhyNot” Partial Inference in EELD Evaluation Domain

- Example domain (small/medium size):
  - 150 concepts, 200 relations, 6000 individuals
  - 10,000 asserted facts
  - 125 rules

- Example query:
  
  \[(\text{contract-murder UID517 UID3})?\]

- Strict proof is fairly large:
  - 121 facts (or leaves)
  - 80 rule applications
  - Chaining depth up to 9
  - **Impossible to debug manually if it fails**

- Great domain for WhyNot partial match to show its utility:
  - Example: explain query failure caused by 1 missing assertion
Strict proof: ~1 second
Partial proof: ~2 minutes
Large space of potential partial proofs explored
Conclusion

- PowerLoom is well-suited for the representation & reasoning tasks:
  - Full-function, robust and stable KR&R system
    - Expressive representation, reasoning, query language, storage, extensive API
      - Available in Java (useful for integration with Protégé)
  - Meta-representation & reasoning
  - Concepts, relations, contexts, rules, queries, etc. are all first-class citizens which can be represented and reasoned about
  - Explanation support for successful and failed reasoning
  - Sophisticated context & module system
    - Encapsulation, efficient inference, representation of assumptions
  - Sophisticated support for units & measures
  - Support for simple timepoint reasoning.