Integrating Natural Language, Knowledge Representation and Reasoning, and Analogical Processing to Learn by Reading

Kenneth D. Forbus, Christopher Riesbeck, Lawrence Birnbaum, Kevin Livingston, Abhishek Sharma, Leo Ureel

NORTHWESTERN UNIVERSITY

Overview

- Systems that learn by reading would
  - radically change the economics of building large knowledge bases
  - provide a platform for cognitive simulations of larger-scale phenomena
- Learning Reader learns by reading simplified language texts
  - Manages syntactic complexity
  - Unconstrained vocabulary, unlike controlled languages
- Learning Reader combines
  - Natural language processing
  - A large-scale knowledge base
  - Deductive reasoning
  - Analogical processing
- Novel aspects
  - Large scale integration
  - Memory-based parsing over a large scale knowledge base
  - Rumination over new knowledge
  - Automatic axiom extraction for reasoning
- Experiments indicate that Learning Reader successfully accumulates and applies new knowledge from simplified texts

Portion of KB Accessible by DMAP

	Find	Create
Collections (27,649)	99%	57%
Predicates (8,892)	15%	15%
Assertions (1.2 M)	43%	

Reading Speed

- 77% of sentences in under 1 second each
- 93% in under 4 seconds each
- 99% when allowed to take up to 1.1 minutes

DMAP Rule Pattern Matching Example

Input: "An attack occurred in Madrid"

Attack-TheWord, Occur-TheWord, In-TheWord, CityOfMadridSpain

(isa ?x AttackOnObject)

Rule:

Pattern: ?Event Occur-TheWord In-TheWord ?GeographicLocation

Produces: (eventOccursAt ?Event ?GeographicLocation)

Constraints: (isa ?Event AttackOnObject)

Bindings: ((?GeographicLocation . CityOfMadridSpain))

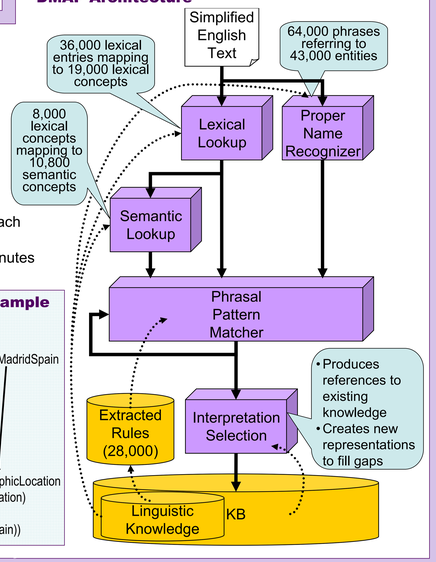
Direct Memory Access Parsing (DMAP)

- Maps textual patterns directly to knowledge structures
  - Uses existing semantic and episodic memory
  - Creates new representation to fill gaps
- Relevant KB knowledge is applied to generate expectations during sentence understanding
  - "George Bush announced…" generates very different expectations than "Motorola announced…"

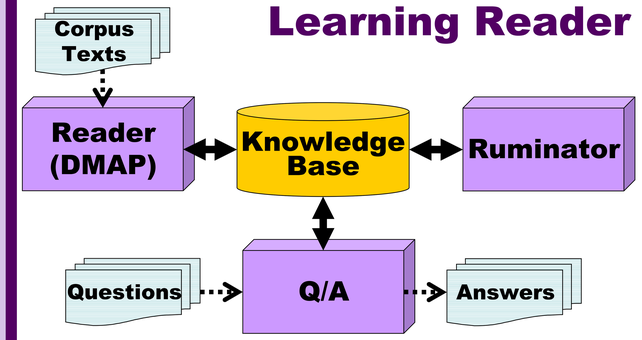
Example

Input	An attack occurred in Madrid. The bombing killed 190 people. The bombing was perpetrated by Al-Qaida.
Interpretation	
Consistent Assertions	
The perpetrator of the terrorist attack in Madrid on September 8, 2003 is al-Qaida.	(perpetrator TerroristAttack-September-8-2003-Madrid AlQaida)
The number of Person who died in the terrorist attack in Madrid on September 8, 2003 is 190.	(deathToll TerroristAttack-September-8-2003-Madrid Person 190)
the terrorist attack in Madrid on September 8, 2003 is a bombing.	(isa TerroristAttack-September-8-2003-Madrid Bombing)
Confirmed Assertions	
the terrorist attack in Madrid on September 8, 2003 occurred at Madrid, Spain.	(eventOccursAt TerroristAttack-September-8-2003-Madrid CityOfMadridSpain)

DMAP Architecture



Learning Reader



Question Answering (Q/A)

- Used to test the learning and reasoning capabilities of the system.
- Answers parameterized questions by backchaining.
- Utilizes methods for identifying useful and efficient axioms.

Parameterized Questions

1. Who is <Person>?
2. Where did <Event> occur?
3. Where might <Person> be?
4. What are the goals of <Person>?
5. What are the consequences of <Event>?
6. When did <Event> occur?
7. Who is involved in <Event>?
8. Who is acquainted with (or knows) <IntelligentAgent>?
9. Why did <Event> occur?
10. Where is <SpatialThing>?

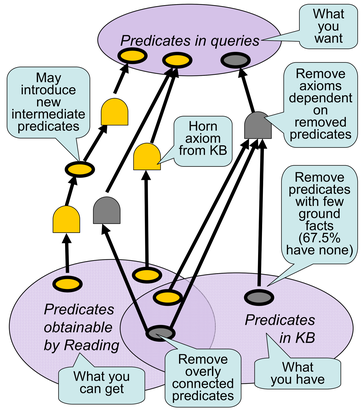
Example Question and Answers

Question: What are the (possible) consequences of **IraqOccupiesKuwait** ?

Answer(s):

```
1.
(eventResults IraqOccupiesKuwait
              BlockadeOfIraq)
2.
(increasesCausally
 IraqOccupiesKuwait
 (thereExists ?RETALIATION
  (and
   (isa ?RETALIATION
        AttackOnTangible)
   (maleficiary ?RETALIATION
                (ArmyFn Iraq))))
 likelihood)
```

Axiom Extraction and Optimization



Speedup versus Completeness

	Speedup Factor Over Baseline	Completeness Over Baseline
Average	129	32.3%
Worst	4	-8.5%

Aggregate results from three experiments

Results

- Automatic static analysis can remove reasoning bottlenecks
- Averaged two orders of magnitude improvement in speed
- Averaged an order of magnitude improvement in completeness
- Worst case small penalty in completeness with small speedup

Ruminator

- Improves the system's understanding by off-line computation
- Asks itself questions about what has been read
- Uses analogical reasoning for generation of questions and filling knowledge gaps
- Detects errors

Deductive Question Answering Example

Question	Answer
Who is involved in or affected by IraqOccupiesKuwait ?	(affectedAgent IraqOccupiesKuwait ThirdUSArmy)
What are the consequences of IraqOccupiesKuwait ?	(increasesCausally IraqOccupiesKuwait (thereExists ?RETALIATION (and (isa ?RETALIATION AttackOnTangible) (performedBy ?RETALIATION Kuwait) (startsAfterStartingOf ?RETALIATION IraqOccupiesKuwait) (maleficiary ?RETALIATION (ArmyFn Iraq)))) likelihood)
What are the consequences of IraqOccupiesKuwait ?	(eventResults IraqOccupiesKuwait BlockadeOfIraq)
What are the consequences of IraqOccupiesKuwait ?	(eventResults IraqOccupiesKuwait IraqDeclaresKuwaitIts19thProvince)

Analogical Processing Example

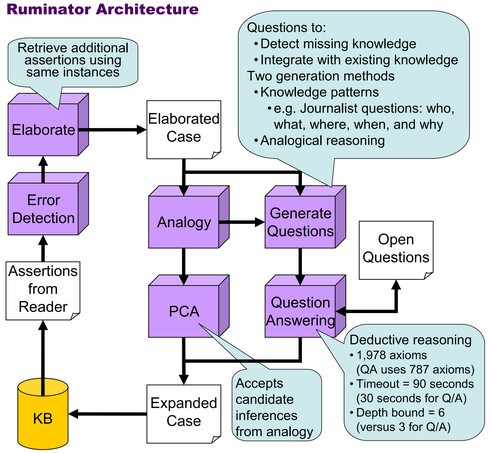
Most Similar case: (MinimalCaseFn IraqInvadingIran-1980)

Candidate Inference	Accepted /Rejected	Justification
(mostNotableIsa IraqOccupiesKuwait (AnalogySkolemFn MilitaryInvasion))	Rejected	Contains Skolem Variables
(startingDate IraqOccupiesKuwait (MonthFn February (YearFn 1991)))	Rejected	Inconsistent Assertion
(maleficiary IraqOccupiesKuwait Kuwait)	Accepted	No inconsistency
(isa IraqOccupiesKuwait MilitaryInvasion)	Accepted	No inconsistency
(directingAgent IraqOccupiesKuwait Iraq)	Accepted	No inconsistency

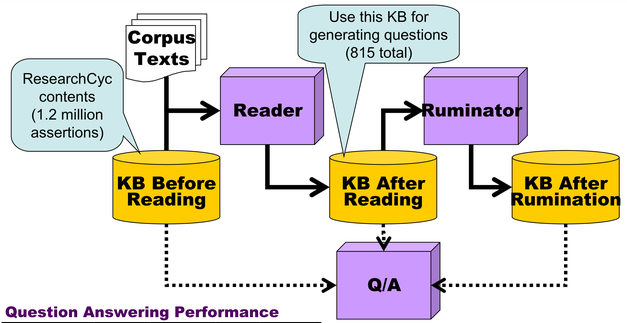
Promiscuous Conjecture Acceptance (PCA)

- Accepts (stores into KB) candidate inferences from analogy
  - Rejects inferences (shown in red) that contain skolem variables or are known to be inconsistent with the KB

Ruminator Architecture



Full System Experiment



Question Answering Performance

Condition	Answers	Questions	Wrong	Accuracy
Before Reading	87	10%	0	100%
Reading only	320	37%	1	99.7%
Reading + Deductive Rumination	434	50%	3	99.3%
Reading + Deductive + PCA	525	60%	48	90.8%

Corpus

- 62 stories (956 sentences).
- Divided into 186 text snippets
- Each marked with topic (e.g. MilitaryAction) topic only used in creating case libraries and SEQL operations in Ruminator

Results and Observations

- Learning Reader can learn from simplified English
  - Can answer questions after reading that it could not before
  - Can answer even more questions after rumination
  - Can answer still more questions after non-deductive rumination
- DMAP provides an effective way of moving from language to predicate calculus
- Analogical reasoning in Rumination is useful for integrating and extending knowledge acquired from reading.
- Large scale learning systems should be robust to errors generated at the various stages of processing.

Future Work

- Expand corpus
  - Increased number of stories over all topics
  - One (or more) sets of stories about a particular event or set of events (e.g. the Israel Lebanon conflict of 2006)
    - Learning complex causal relations
    - Use to explore the impact of noise
- Use DMAP for question-parsing instead of parameterized questions.
  - expand coverage
  - provide the basis for interactive dialogue
    - allow trainers to ask follow-up questions
    - allow the Ruminator to ask its trainers a limited number of questions, with answers being interpreted also via DMAP
- Extend Ruminator processes which scrutinizes newly learned knowledge for inconsistencies to also implement repairs
  - Use techniques of model-based diagnosis, applied to a model of the system's own processing
- Expand the role of evidential reasoning in the Ruminator
  - Exploit the probabilities generated in analogical processing to help address misunderstandings, when diagnosed.

Acknowledgements

This project was funded through DARPA, grant HR0011-04-1-0051. Thanks to Cycorp for ResearchCyc.