CSE392 - Computers playing Jeopardy! Course Information

CSE 392, Computers Playing Jeopardy!, Fall 2011

Stony Brook University

http://www.cs.stonybrook.edu/~cse392

Course Description

- "IBM Watson is a computer system capable of answering rich natural language questions and estimating its confidence in those answers at a level of the best humans at the task. On Feb 14-16, in an televised event, Watson triumphed over the best human players of all time on the American quiz show, Jeopardy!. In this course we will discuss the main principles of natural language processing, computer representation of knowledge and discuss how Watson solved some of its answers (right and wrong)."
- Prerequisites: some experience in a programming language.
  - No background in NLP or KR is necessary

Course Focus

- Unstructured Information Managing Architecture UIMA (in Java)
- Natural Language Processing (NLP)
- Knowledge Representation (KR)

Instructor Information

- Dr. Paul Fodor
  1437 Computer Science Building
- Office hours: We 12:00-2:00PM and Th 1:00-2:00PM
  - I am also available by appointment
- Email: pfodor (at) cs (dot) stonybrook (dot) edu
- TAs: TBD
- Please include "CSE 392" in the email subject and your name in your email correspondence

General Information

- Meeting Information:
  - Lectures: TuTh 9:50-11:10am, CS 2114
- Course Web page: http://www.cs.stonybrook.edu/~cse392
  - Blackboard will be used for assignments, grades and course material
  - I also use Blackboard to send email to the class, so make sure that your email address in Blackboard is up-to-date.

Textbook

- No textbook is required
  - Jurafsky, D. and Martin, J. H. Speech and Language Processing. Prentice Hall: 2000. ISBN: 0130950696.
  - Manning, C. D. and H. Schütze: Foundations of Statistical Natural Language Processing. The MIT Press. 1999. ISBN 0-262-13360-1
- Necessary Software:
  - Java 1.5 or higher: download from http://www.oracle.com/technetwork/java/javase/downloads
  - Eclipse IDE: http://www.eclipse.org
- Important Dates
  - Final Exam: Wed. Dec 14, 11:15am-1:45pm, class
    - http://www.stonybrook.edu/registrar/finals.shtml

Coursework

- **Grading Schema**
  - Homework = 50%
    - Programming homework assignments
  - Midterm exams (2) = 30% (15% each)
  - Final exam = 20%

Assignment Submission

- All assignments should be submitted electronically
  - Blackboard

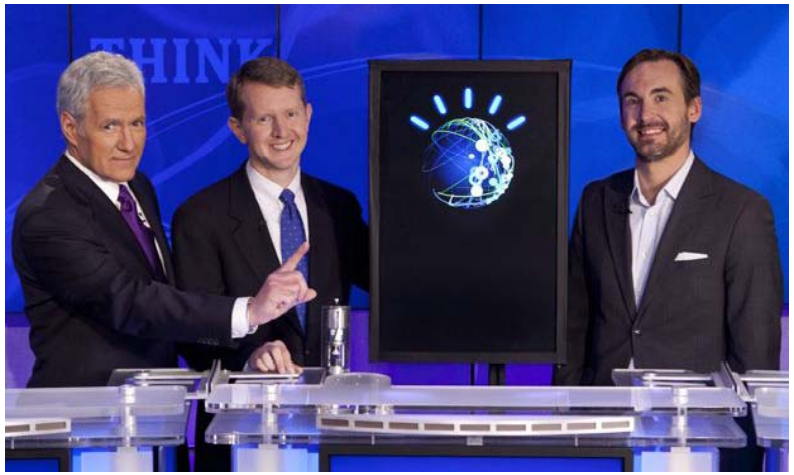
Academic Integrity

- You can discuss general assignment concepts with other students
- You MAY NOT share assignments, source code or other answers
  - Assignments are subject to manual and automated similarity checking
- If you cheat, you MAY be brought up on academic dishonesty charges without warning - we follow the university policy:
  - http://www.stonybrook.edu/uaa/academicjudiciary

Please

- Please be on time
- Please show respect for your classmates
- Please turn off (or use vibrate for) your cellphones
- On-topic questions are welcome


JEOPARDY!
The IBM Challenge

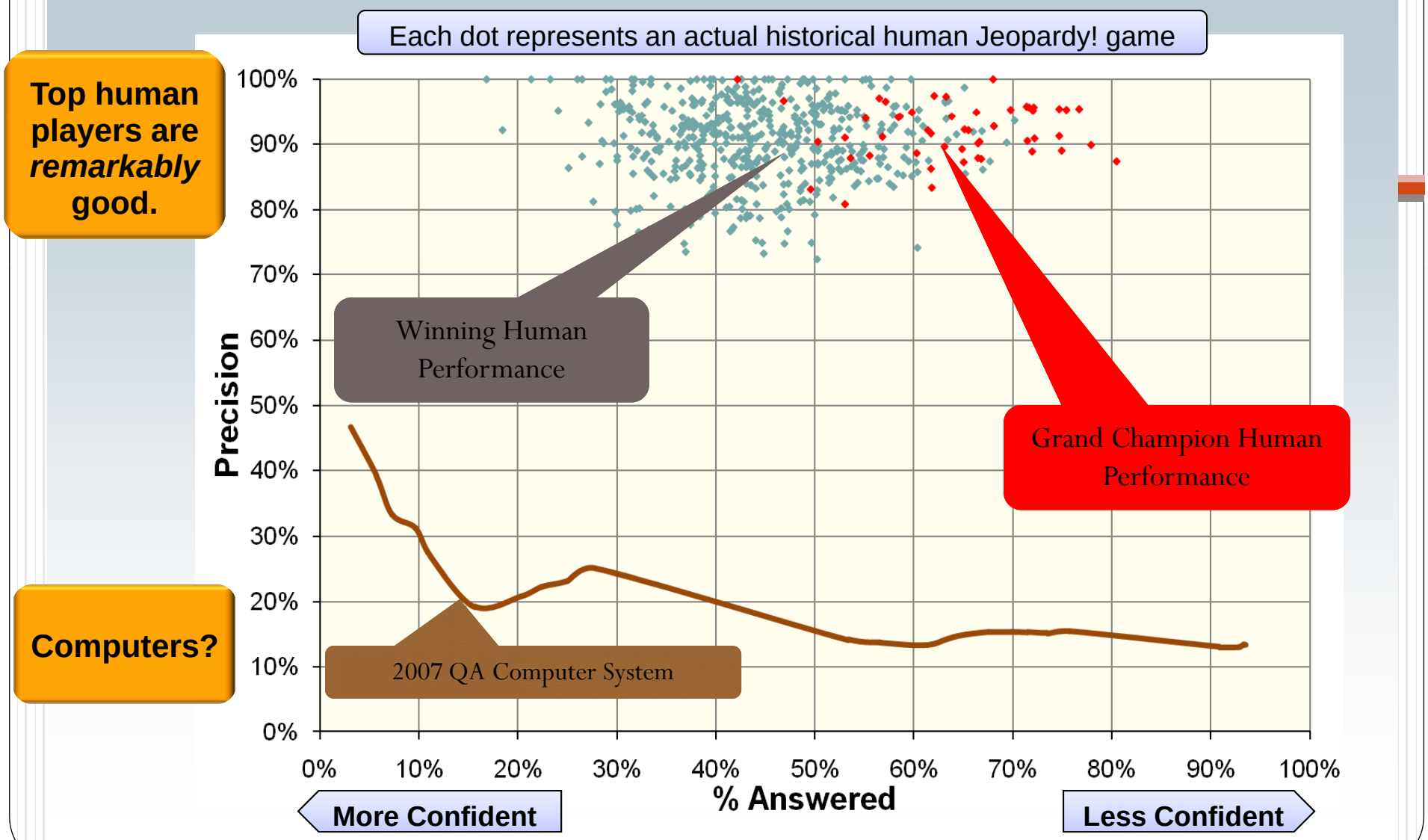

THINK

Real Language is Real Hard

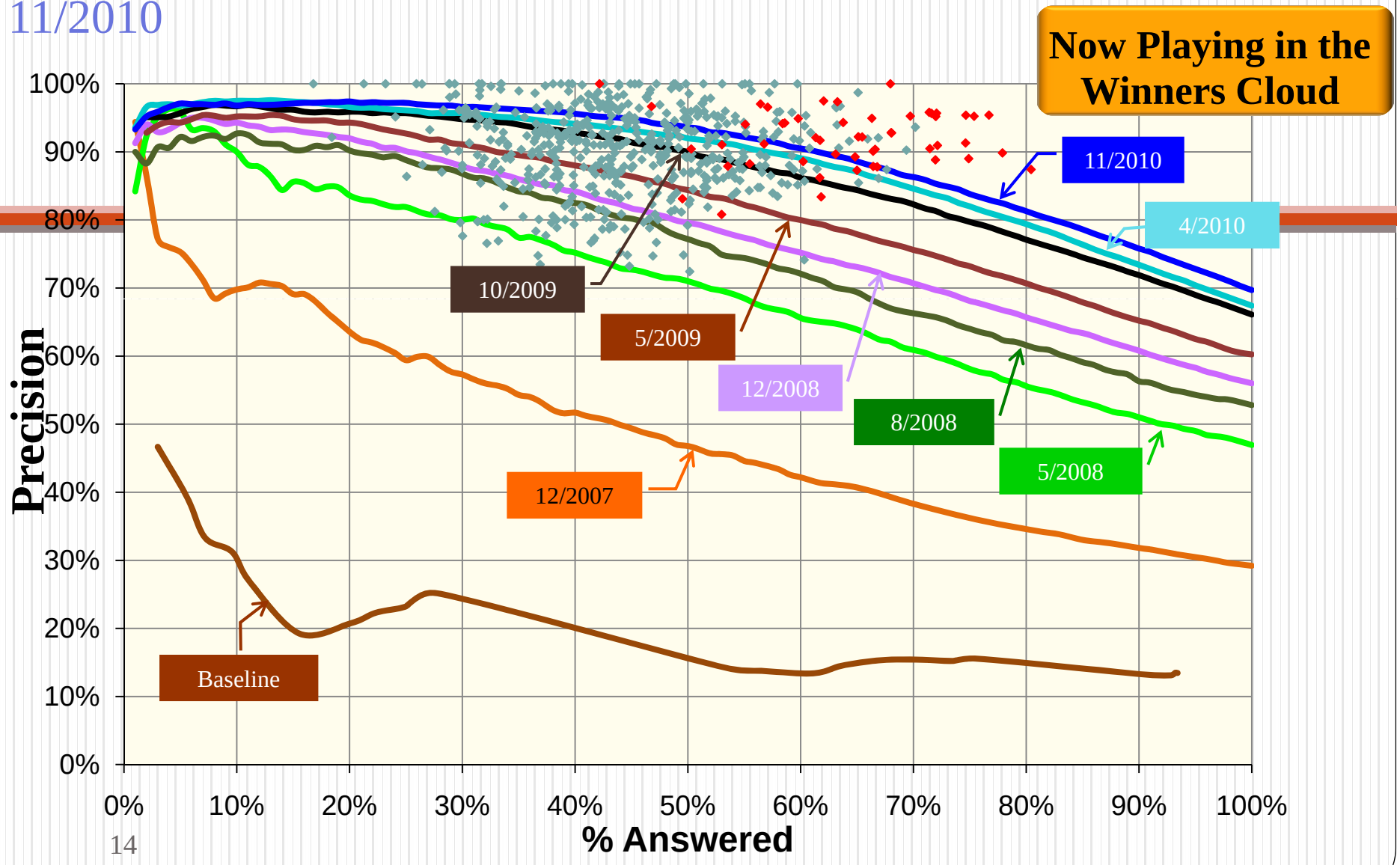
- Chess
  - A finite, mathematically well-defined search space
  - Limited number of moves and states
  - Grounded in **explicit**, **unambiguous** mathematical rules

- Human Language
  - Ambiguous, contextual and implicit
  - Grounded only in **human cognition**
  - Seemingly infinite number of ways to express the same meaning

The Best Human Performance: *Our Analysis Reveals the **Winner's Cloud***



DeepQA: Incremental Progress in Precision and Confidence 6/2007-11/2010

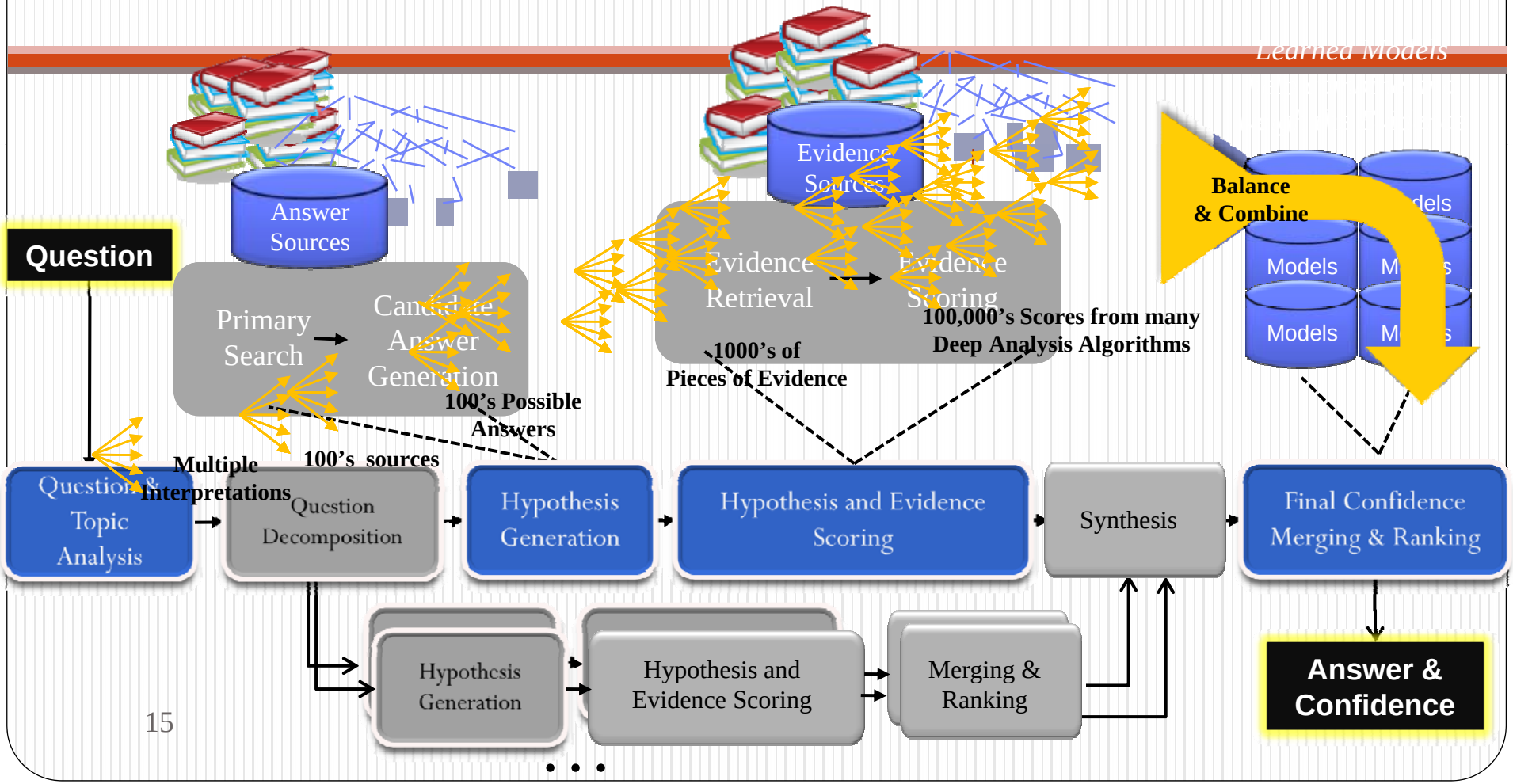


DeepQA: The Technology Behind *Watson*

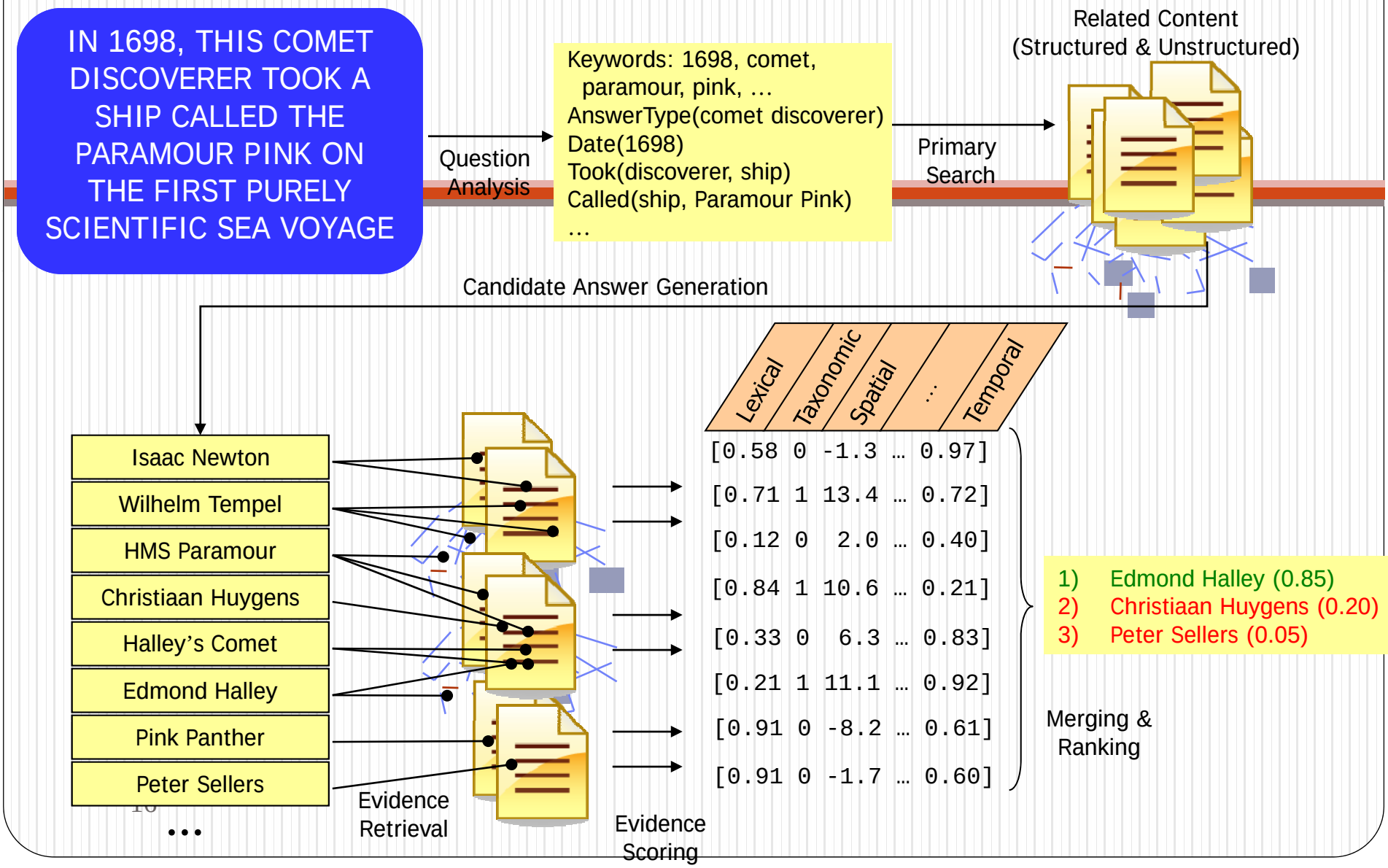
Massively Parallel Probabilistic Evidence-Based Architecture

Generates and scores many hypotheses using a combination of 1000's ***Natural Language Processing***, ***Information Retrieval***, ***Machine Learning*** *and* ***Reasoning Algorithms.***

These gather, evaluate, weigh and balance different types of ***evidence*** *to deliver the answer with the best support it can find.*

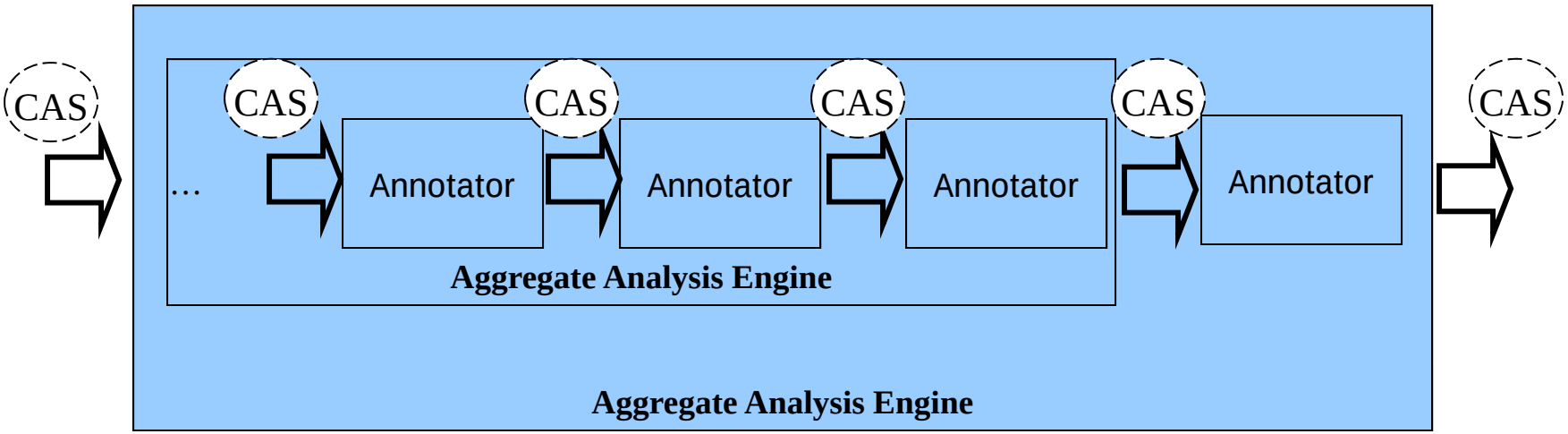


How Watson Processes a Question

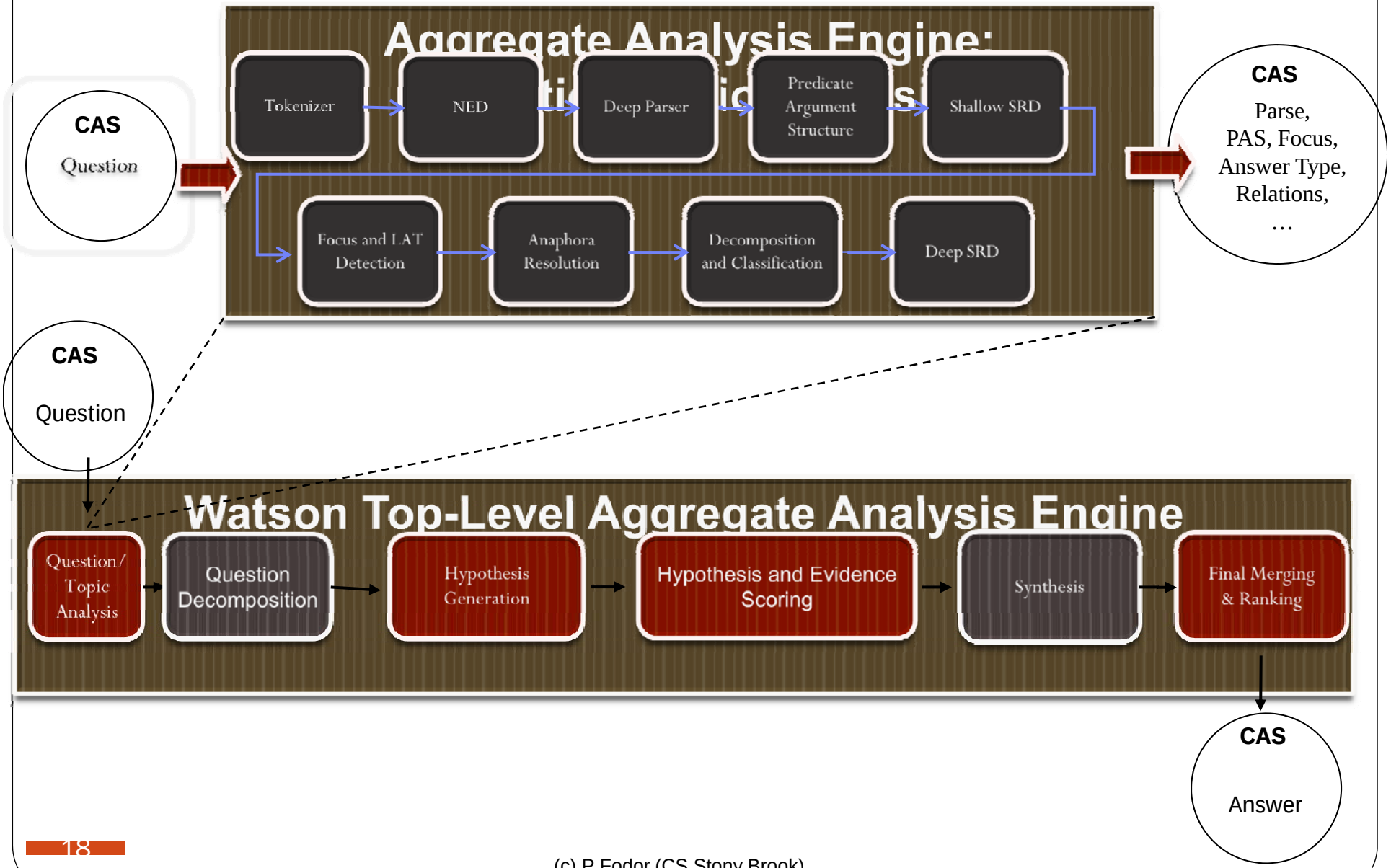


Apache UIMA

- Open-source framework and tools for building NLP applications
- Key Concepts
  - *Common Analysis Structure (CAS):* Container for Data Structures in user-defined data model (which can be defined in UML)
  - *Annotator:* Pluggable component (Java or C++, among others) that reads and writes a CAS
  - *Aggregate Analysis Engine:* Collection of Annotators



Watson in UIMA

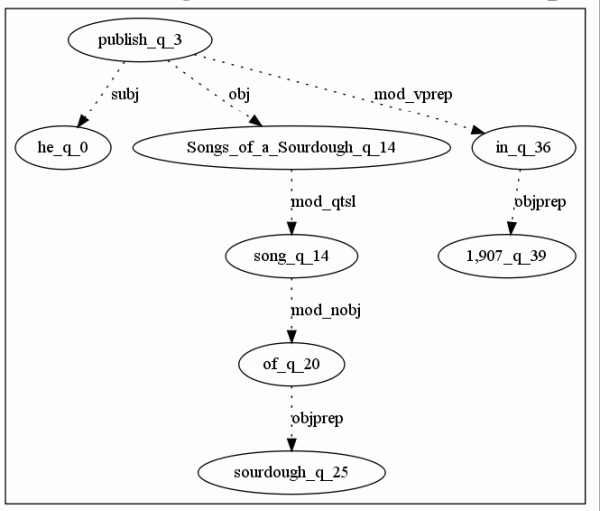


Natural Language Processing In Watson

Text
(Question or
Evidence)

Tokenization
Deep Parsing
Predicate Argument Structure
Named Entity Recognition

Predicate Argument Structure Graph



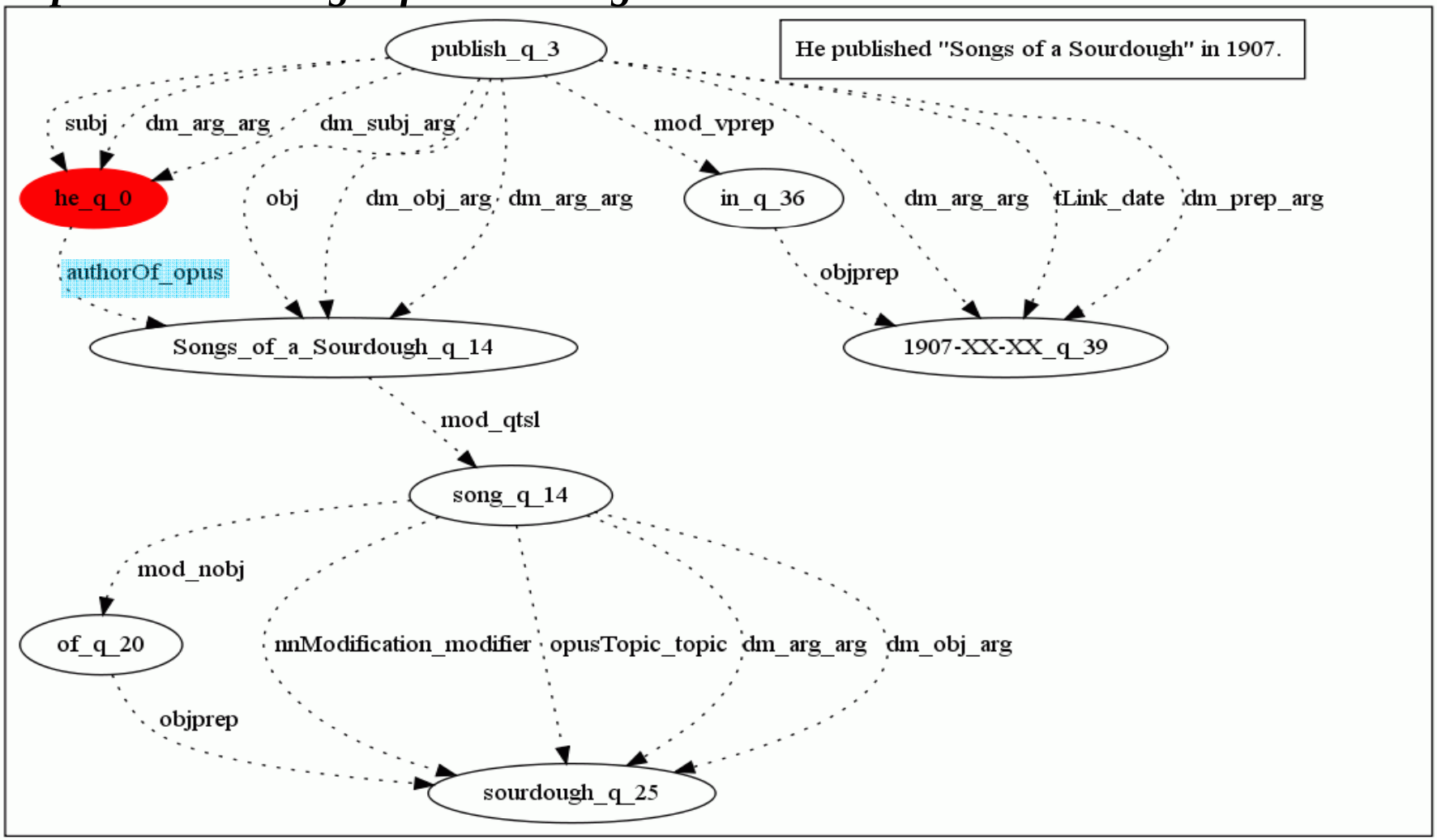
Rule-Based and
Statistical
Pattern Matching

Relations
Co-Reference Resolution
Question Focus
Lexical Answer Types (LATs)
Question Classification

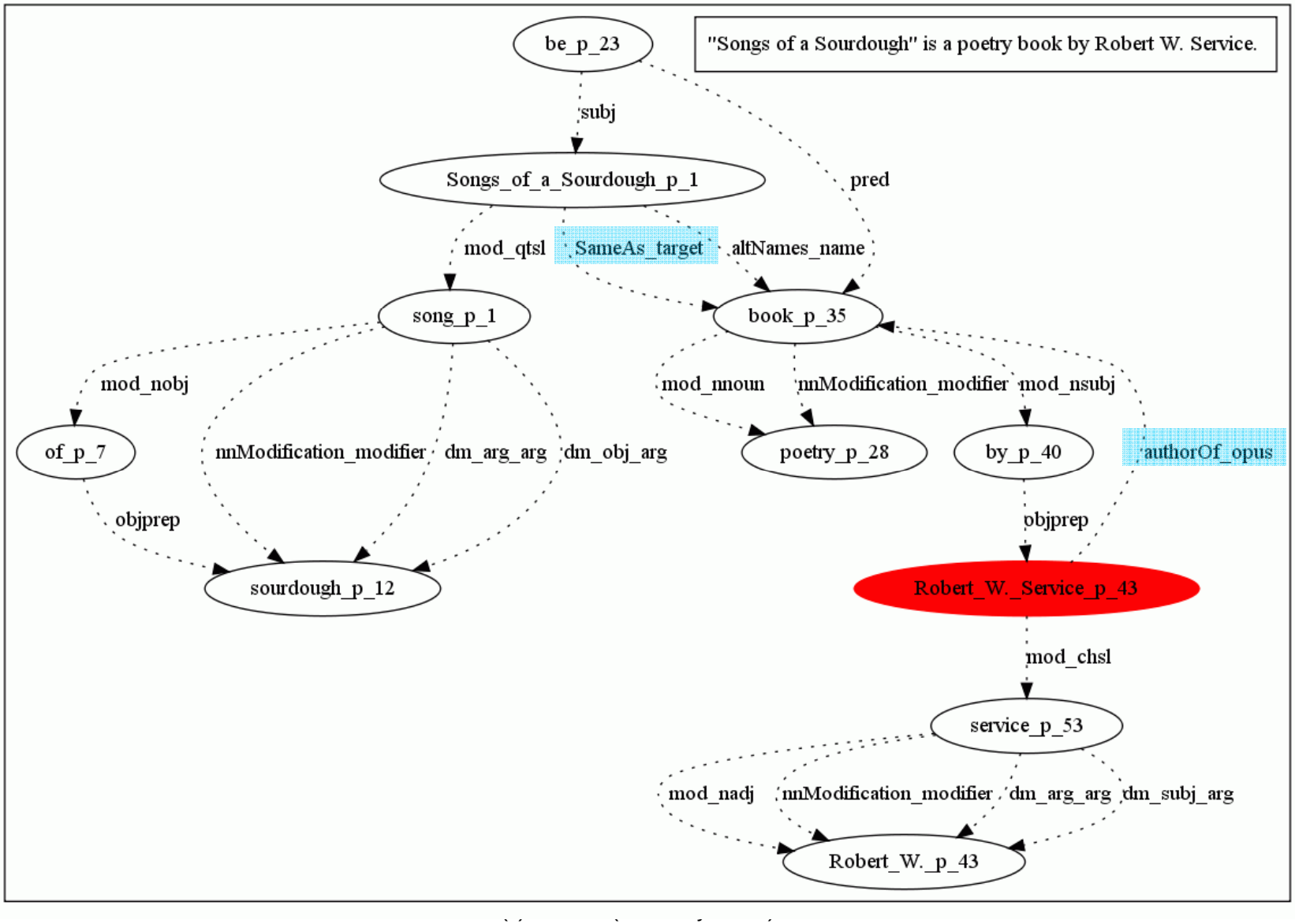
Predicate Argument Structure and Relations in Question

POETS & POETRY: He was a bank clerk in the Yukon before

he published "Songs of a Sourdough" in 1907

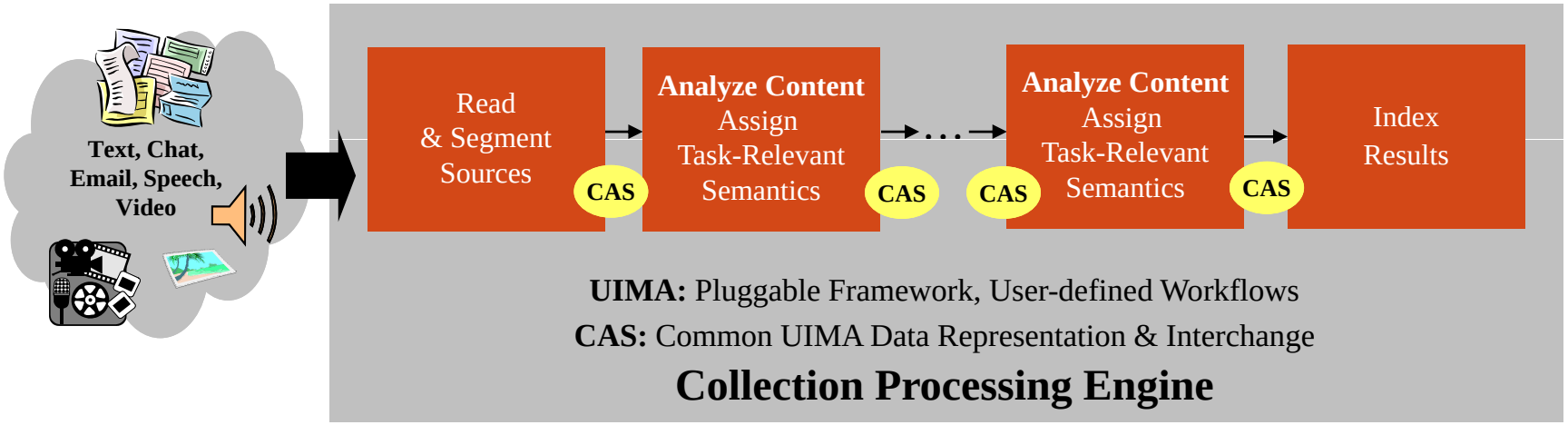


PAS and Relations in a Supporting Passage



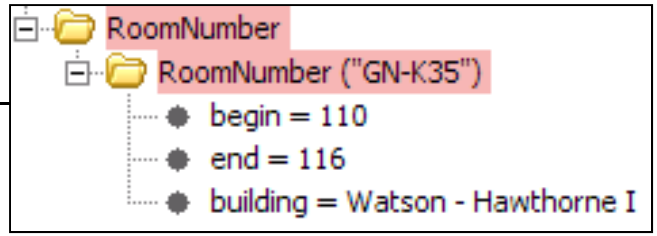
Unstructured Information Management Architecture (UIMA)

- Apache UIMA: http://incubator.apache.org/uima/
- Platform independent standard for interoperable text and multi-modal analytics

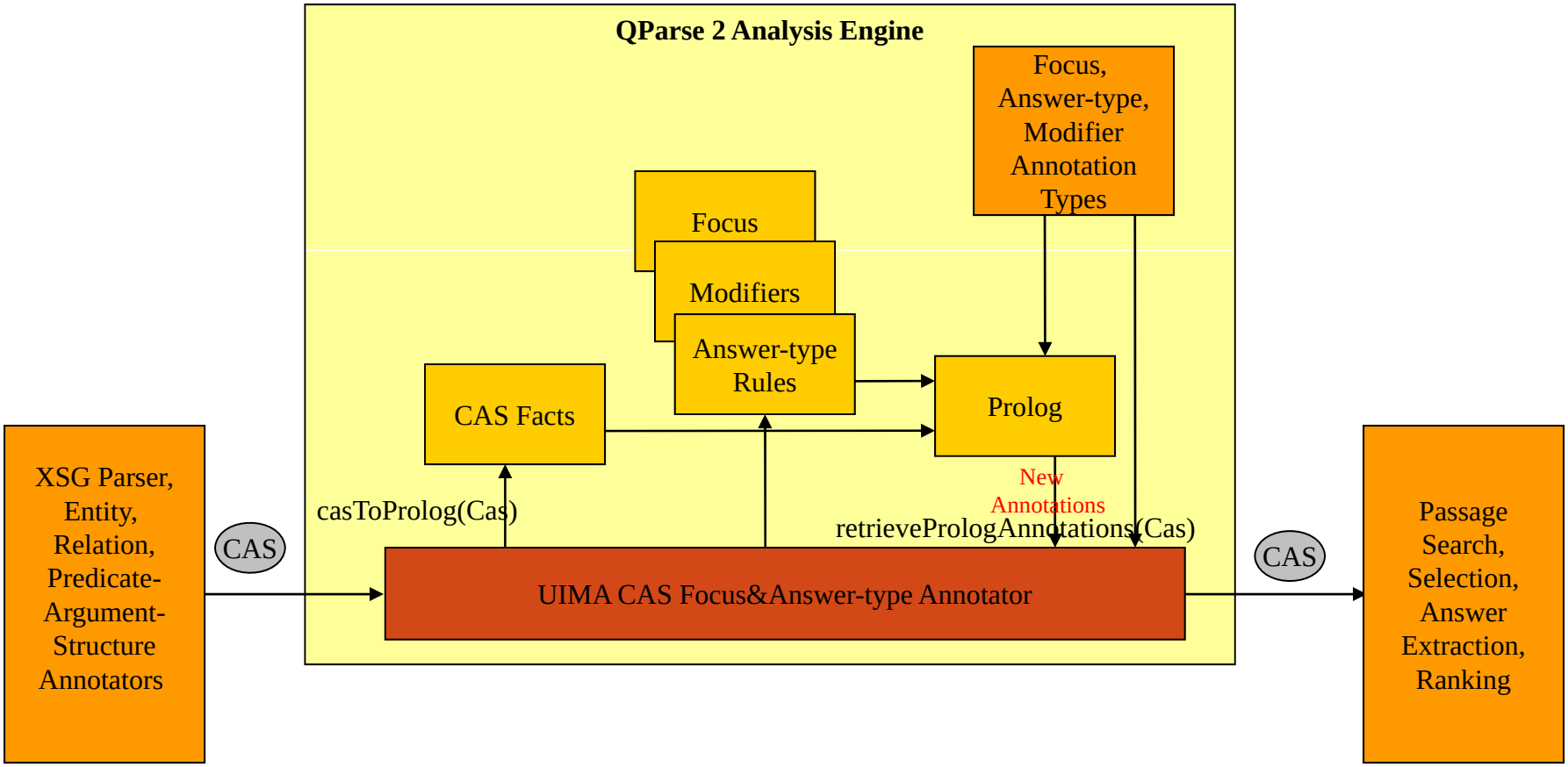


UIMA Annotation

Document text:
"…seminar in GN-K35 on October 24, 2007"



My work in IBM Watson - UIMA CAS Prolog Interface Architecture



Focus Computation rules

- The focus is the "node" that refers to the unspecified answer
  - "What is the name of the airport in Dallas?"
    - Focus = "airport"
  - "What is the population of Iceland?"
    - Focus = "population"
- The focus abstracts different syntactical constructs:

  1) What X …
  2) What is the X that…
  3) Which of the X …
  4) What is the name of the X that…
  5) Name the X that…

  …
- Applications:
  - Answer-type detection
  - Logical form answer-selection

Example QParse2 Focus Detection Rules

- "How much/many" rule:
  - Pattern: HOW_MANY/MUCH XVERB …?
  - Examples:

    "How many hexagons are on a soccer ball?"

    "How much does the capitol dome weigh?"

    "How much folic acid should an expectant mother get daily?"

```
focus(QuestionRoot, [Determiner]):-
        getDescendantNodes(QuestionRoot,Determiner),
        lemmaForm(Determiner,DeterminerString),
        howMuchMany(DeterminerString),!. % "how much/many", "this much",…
```

Example QParse2 Focus Detection Rules

- "What is X …" rule:
  - Pattern: WHAT IS X …?
  - Example:
    "What is the democratic party symbol?"
    "What is the longest river in the world?"

```
focus(QuestionRoot, [Pred]):-
  getDescendantNodes(QuestionRoot,Verb),
  lemmaForm(Verb,"be"),
  subj(Verb,Subj),
  lemmaForm(Subj,SubjString),
  whatWord(SubjString), % e.g., "what","which“ ("this","these“)
  pred(Verb,Pred),!.
```

Answer-type Computation Rules

- Computes the type of the answer - heuristics

Focus lexicalization (R2 annotations and lexical chains using Prolog WordNet followed by a mapping to our taxonomy):

Question	QParse 2 AnswerType
What American revolutionary general turned over West Point to the British?	[com.ibm.hutt.MilitaryLeader]

Table lookup for the verb:

Question	QParse 2 AnswerType
How did Jimi Hendrix die?	[com.ibm.hutt.Disease com.ibm.hutt.MannerOfKilling com.ibm.hutt.TypeOfInjury]

Table lookup for the focus:

Question	QParse 2 AnswerType
How far is it from the pitcher's mound to home plate?	[com.ibm.hutt.Length]
When was Lyndon B Johnson president?	[com.ibm.hutt.Year]

Table lookup for the focus (noun) + the verb:

Question	QParse 2 AnswerType
What instrument measures radioactivity?	[com.ibm.hutt.Tool]
What instrument did Louis Armstrong play?	[com.ibm.hutt.MusicalInstrument]

Answer-type Computation Rules

- Heuristics
  - cascading rules in order of generality
    - first rule that fires returns the most specific answer-type for the question

Look at the focus + verb:

Question	QParse 2 AnswerType
How much did Marilyn Monroe weigh?	[com.ibm.hutt.Weight]
How much did the first Barbie cost?	[com.ibm.hutt.Money]

Look at the focus + noun:

Question	QParse 2 AnswerType
How many Earth days does it take for Mars to orbit the sun?	[com.ibm.hutt.Duration]
How many people visited Disneyland in 1999?	[com.ibm.hutt.Population]

Look at the focus:

Question	QParse 2 AnswerType
How many moons does Venus have?	[com.ibm.hutt.WholeNumber]
How much calcium is in broccoli?	[com.ibm.hutt.Number]

Example QParse 2 Answer-type Detection Rules

- Time rule (e.g. when, then):

  Pattern: WHEN VERB OBJ; OBJ VERB THEN

  Example: When was the US capitol built?

  answerType => ["com.ibm.hutt.Year"]

```
answerType(_QuestionRoot,FocusList,timeAnswerType,ATList):-
  member(Mod,FocusList),
  lemmaForm(Mod,ModString),
  wh_time(ModString), % "when", "then"
  whadv(Verb,Mod),
  lemmaForm(Verb,VerbString),
  timeTableLookup(VerbString,ATList),!.
```

Example QParse Answer-type Detection Rules

- "How … VERB" rule:

Pattern: How … VERB?

Example: "How did Virginia Woolf die?"

answerType => ["com.ibm.hutt.Disease", "com.ibm.hutt.MannerOfKilling", "com.ibm.hutt.TypeOfInjury"]

```
answerType(_QuestionRoot,FocusList,howVerb1,ATList):-
  member(Mod,FocusList),
  lemmaForm(Mod,"how"),
  whadv(Verb,Mod),
  lemmaForm(Verb,VerbString),
  howVerbTableLookup(VerbString,ATList), !.
```

QParse2 Evaluation

- 370 correct matches with the standard (89.5%)

343 *exact* answer-type (83%):

Question	QParse 2 AnswerType
Who created the literary character Phineas Fogg?	[com.ibm.hutt.ContentCreator]
What is the name of the airport in Dallas Ft Worth?	[com.ibm.hutt.Facility]
What city is Disneyland in?	[com.ibm.hutt.City]
What color belt is first in karate?	[com.ibm.hutt.Color]

27 of the correct matches were NounPhrase (6.5%):

- one cannot determine the type (unless he already knows the answer of the question)

Question
What did Peter Minuit buy for the equivalent of 2400?
What is the gift for the 20th anniversary?
What did Ozzy Osbourne bite the head off of?

- no type in our taxonomy

Question
What is the word which means one hiring his relatives?
What is a word spelled the same backward and forward called?

QParse2 Evaluation

- 3 results had a subset of the manually annotated answer types

Question	Standard Answer Type	QParse 2 AnswerType
What flavor filling did the original Twinkies have?	[com.ibm.hutt.Food com.ibm.hutt.Material]	[com.ibm.hutt.Material]

- 17 results had extra types than the manually annotated answer types
  - 11 results had a superset of the manually annotated answer types

Question	Standard Answer Type	QParse 2 AnswerType
How big is a keg?	[com.ibm.hutt.Volume com.ibm.hutt.Weight]	[com.ibm.hutt.Length, com.ibm.hutt.Area, com.ibm.hutt.Volume, com.ibm.hutt.Weight, com.ibm.hutt.Number]
How long before bankruptcy is removed from a credit report?	[com.ibm.hutt.Duration]	[com.ibm.hutt.Duration, com.ibm.hutt.Length]
How long is a quarter in an NBA game?	[com.ibm.hutt.Duration]	[com.ibm.hutt.Duration, com.ibm.hutt.Length]

  - 6 results had a super-type of the manually annotated answer types

Question	Standard Answer Type	QParse 2 AnswerType
What are the measurements for a king-size bed?	[com.ibm.hutt.Area com.ibm.hutt.Length com.ibm.hutt.Volume]	[com.ibm.hutt.Measurement]
When did International Volunteers Day begin?	[com.ibm.hutt.Year]	[com.ibm.hutt.DateTime]

QParse2 Evaluation

- 23 results different than the standard:
  - Need for more answer-type detection rules

Question	Standard Answer Type	QParse 2 AnswerType
What does an English stone equal?	[com.ibm.hutt.Weight]	[com.ibm.hutt.NounPhrase]
How do you say “cat” in the French language?	[com.ibm.hutt.NounPhrase com.ibm.hutt.Translation com.ibm.hutt.VerbPhrase]	[com.ibm.hutt.Method]
What did Caesar say before he died?	[com.ibm.hutt.Quotation]	[com.ibm.hutt.NounPhrase]

  - WordNet word sense disambiguation algorithm

Question	Standard Answer Type	QParse 2 AnswerType
What Liverpool club spawned the Beatles?	[com.ibm.hutt.Facility]	[com.ibm.hutt.SportsTeam]

  - Wrong Parse

Question	Standard Answer Type	QParse 2 AnswerType
What 20th century American president died at Warm Springs, Georgia?	[com.ibm.hutt.President]	[com.ibm.hutt.Date]

Results!