Answering Questions from the Semantic Web

Chris Welty

Semantic Analysis and Integration Group
IBM Research
IBM Semantic Web Rep.

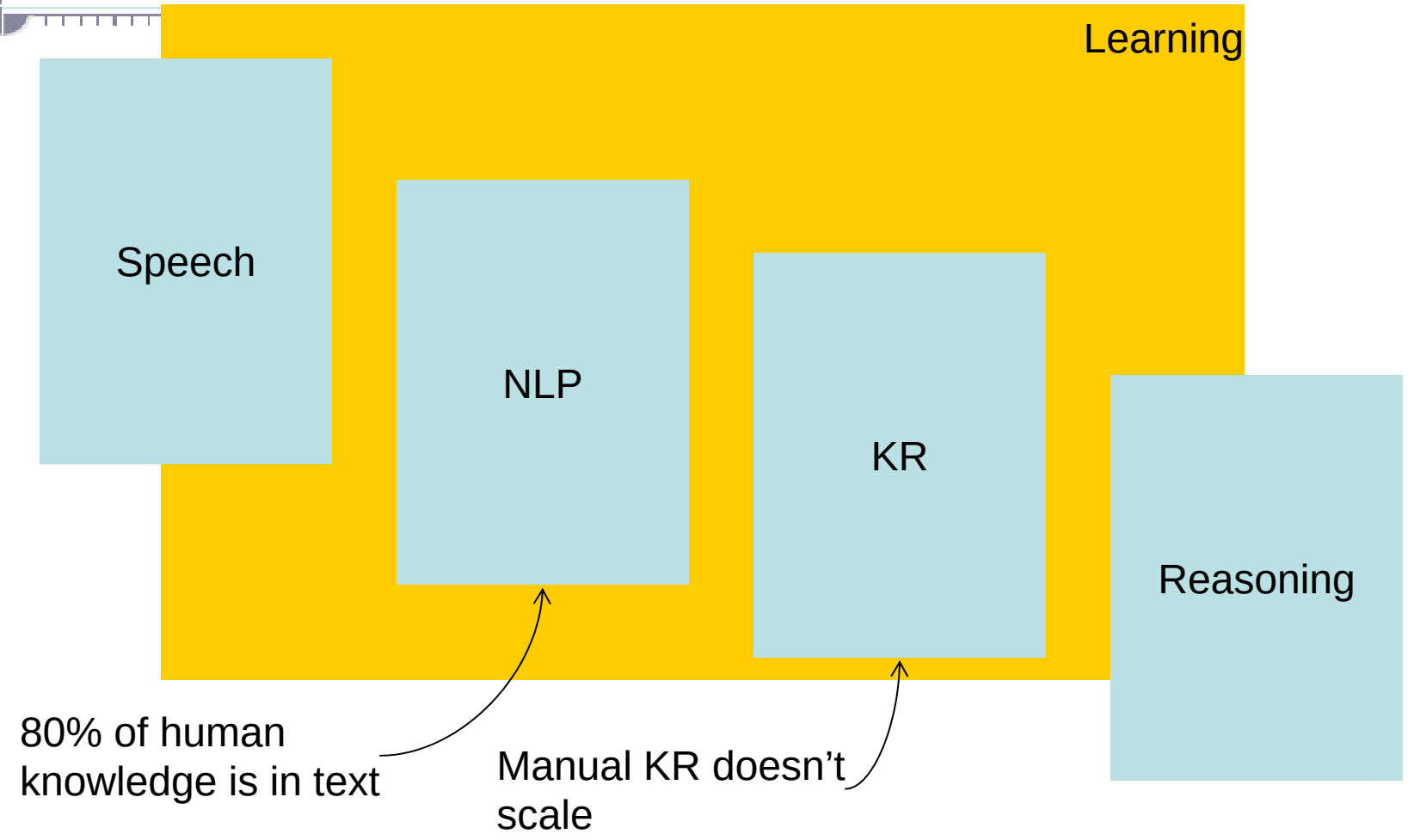
Overview

- Brief History
- Current Problems in NLP
- Current Problems in KR
- The Gap
- Bridging the Gap

AI in Computer Science

- 60s
  - Mostly computer science
  - Driven by simple intuitive ideas
- 70s
  - Some rigor added
  - Some actual computation
- 80s
  - KR and NLP split
  - KR becomes more a pure CS discipline
  - NLP driven mainly by syntax
- 90s
  - KR association with logic becomes entrenched
  - NLP and linguistics merge
  - The NLP/KR gap widens
- 00s
  - KR standards emerge
  - IR-based NLP dominates
  - No one in NLP or KR knows what's going on in the other

The AI Pipeline



Current Problems in NLP

- Machine Translation (MT)
  - Translation between natural languages
  - Mostly shallow and semantic free
  - Driven mainly by statistical correlation
- Search (IR)
  - Mostly shallow and semantic free
  - Bag of words, cosine measure, etc
- Semantic Search (IE)
  - Add simple typing to search
  - Very slight increment over IR
- Question Answering (QA)
  - Most advanced techniques use “structured sources”
    - Wordnet, IMDB

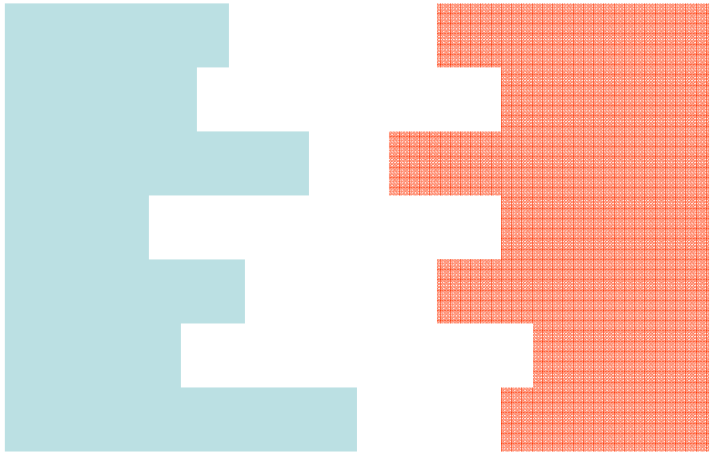
Modern NLP tends to be grounded in practical problems and driven by what works

Current Problems in KR

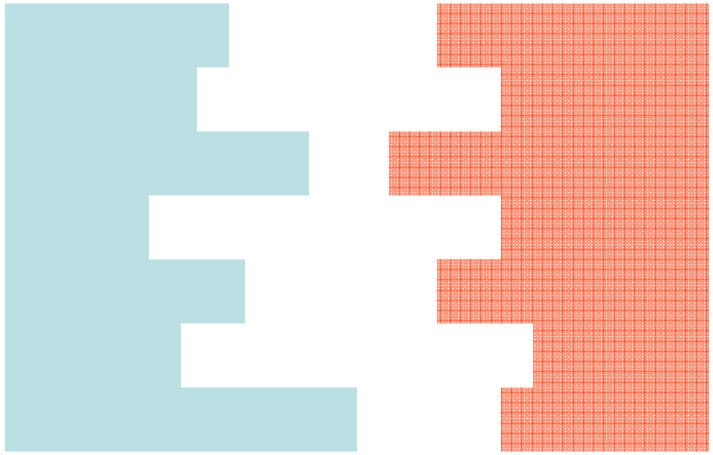
- New FOL fragments
  - Mapping the computational space
- Deepening understanding of modal variants
- Contexts
  - Time, space, belief, …
- Knowledge Acquisition

- Modern KR tends to be grounded in logical problems and driven by perfection

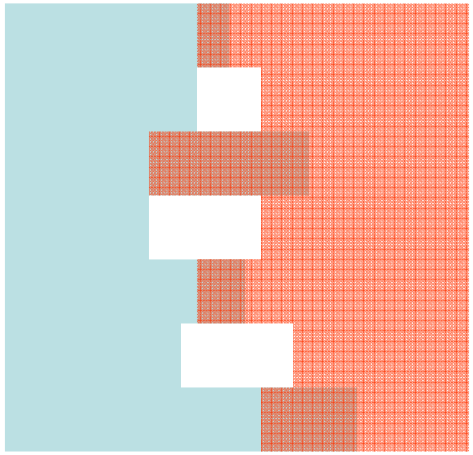
IE ↔ KR



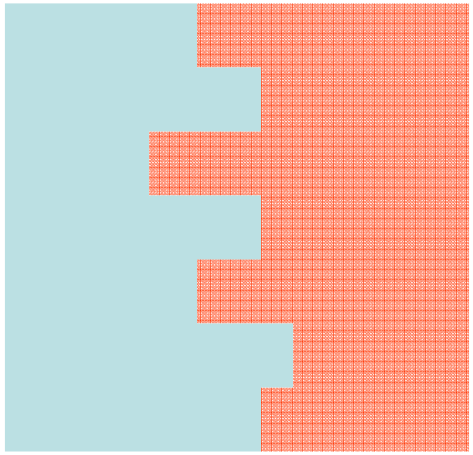
IE ↔ KR



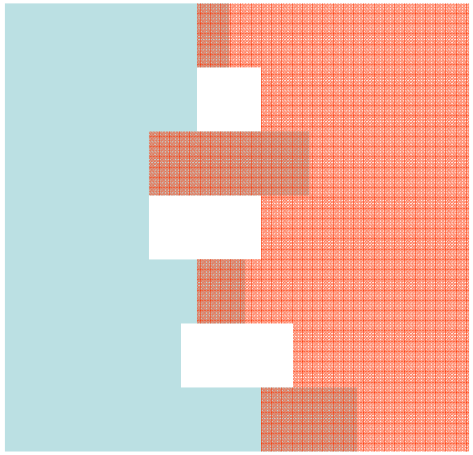
IE ↔ KR



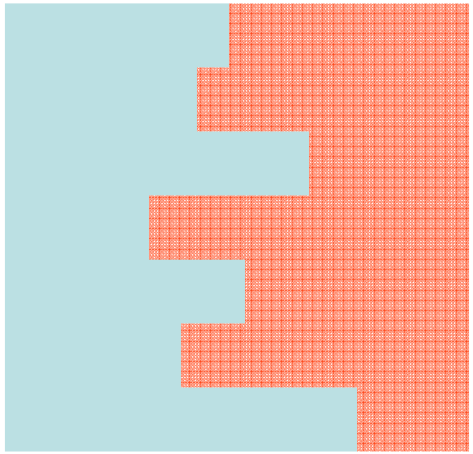
IE ↔ KR



IE ↔ KR

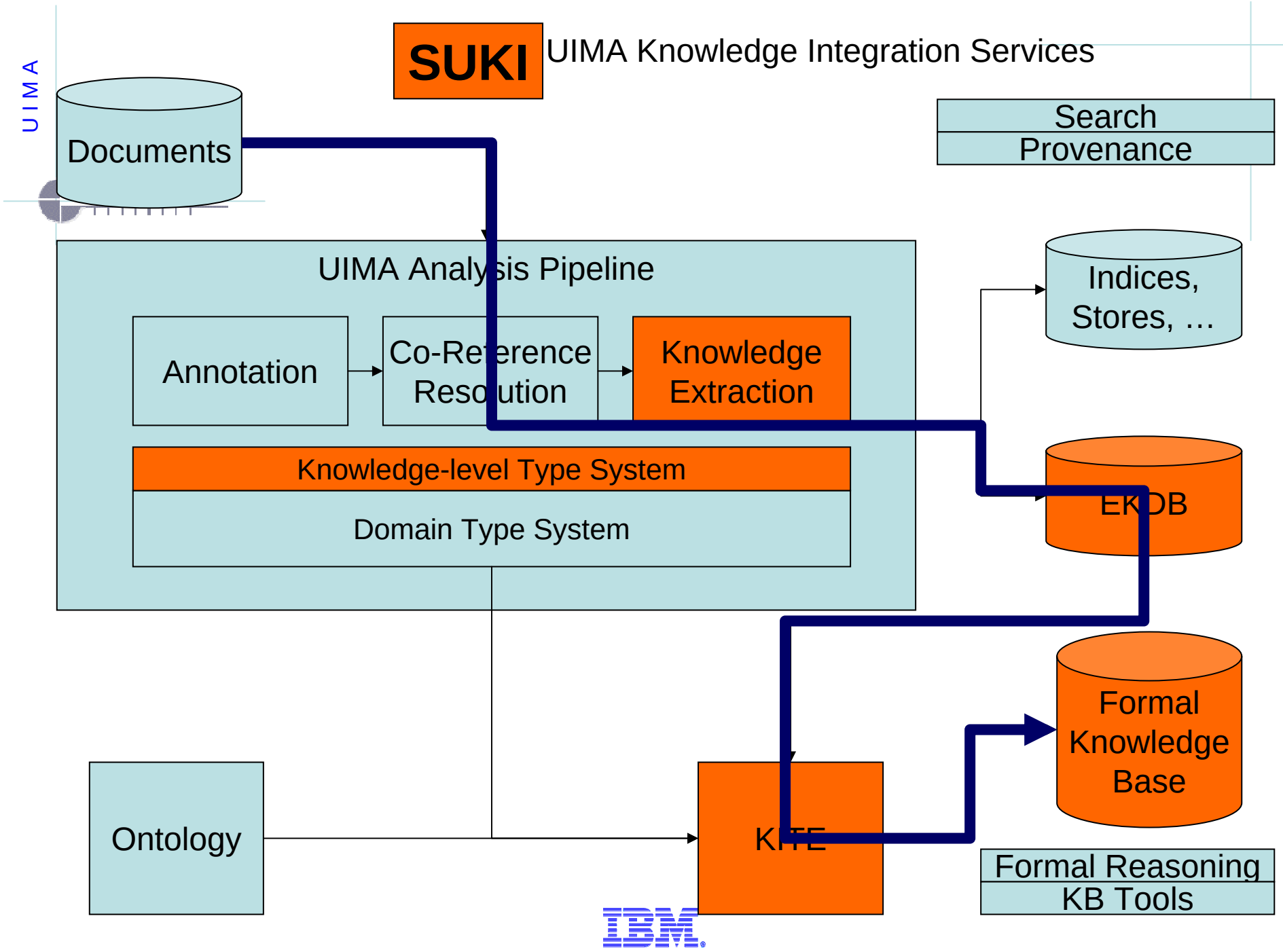


IE ↔ KR

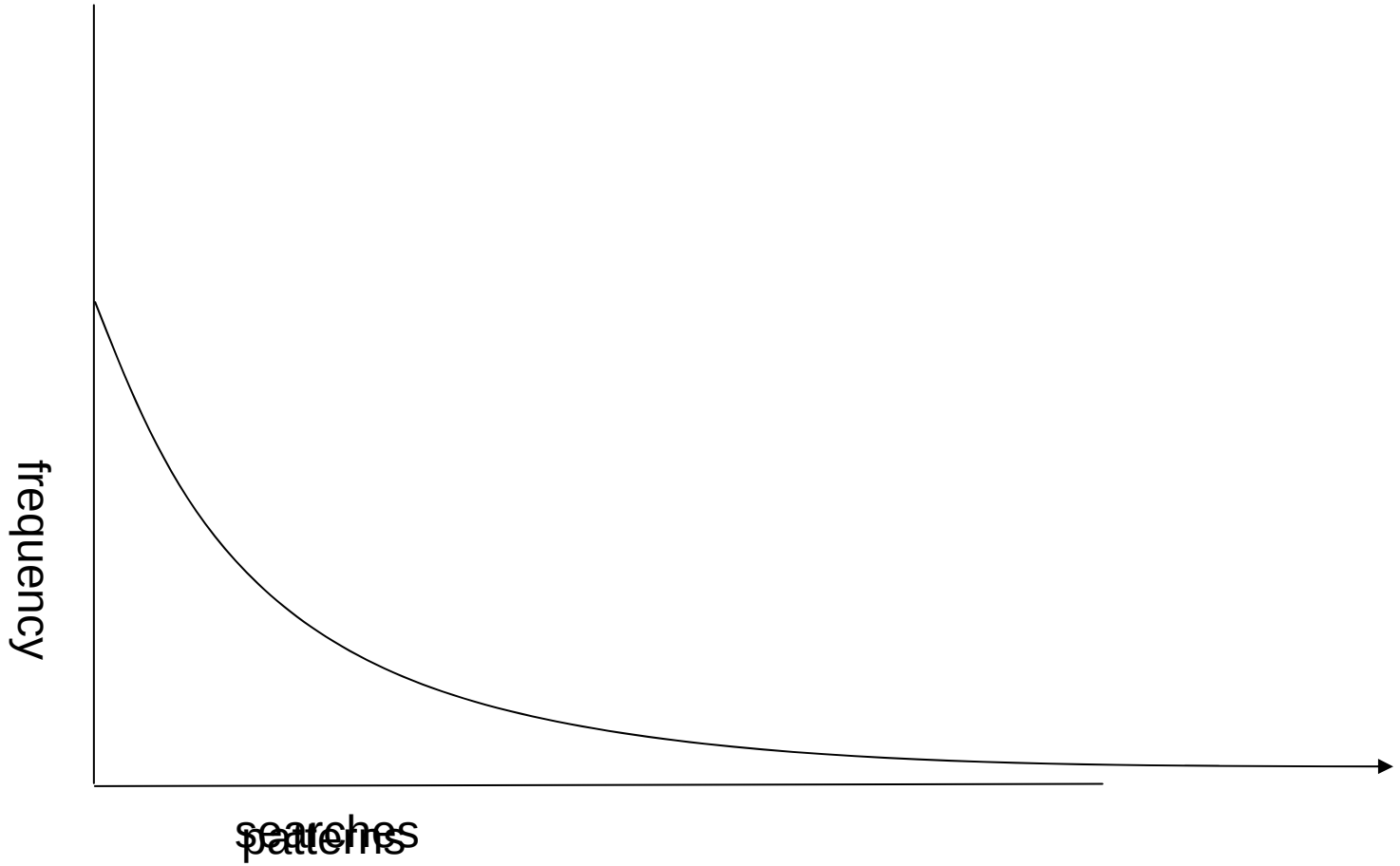


Putting them together

- Precision
  - NLP techniques make mistakes
  - KR&R techniques require perfection
- Recall
  - NLP techniques miss data
  - Reasoning problems often require complete data
- Scale
  - NLP techniques are scalable
  - KR&R techniques are not scalable
- Coreference
  - One of the hardest NLP Problems
  - Required by KR&R
- Relations
  - NLP techniques w/very long tail
  - KR&R useless w/o relations
- Time
  - NLP extremely difficult
  - KR well understood
- Logical Form
  - NLP: Syntactic
  - KR: Semantic
- Soundness
  - NLP: who cares
  - KR: a major focus
- Completeness
  - NLP: who cares
  - KR: a major focus
- Decidability
  - NLP: who cares
  - KR: a major focus
- Explanation
  - NLP not really explainable
  - KR is supposed to be…
- Standards
  - NLP has none (UIMA?)
  - KR now has many
- Ambiguity
  - Language is ambiguous
  - KR is not


SUKI
UIMA Knowledge Integration Services
UIMA
Documents
Search
Provenance
UIMA Analysis Pipeline
Annotation
Co-Reference Resolution
Knowledge Extraction
Knowledge-level Type System
Domain Type System
Indices, Stores, …
EKDB
Formal Knowledge Base
Ontology
KITE
Formal Reasoning
KB Tools
IBM

The very long tail



Towards *semantic* relations

Chris visited Rome

- NLP
  - Verb: visit
    - Tense: past
  - Object: Rome
  - Subject: Chris
- KR
  - Visit(Chris, Rome)

Towards *semantic* relations

Chris went to Rome

- NLP
  - Verb: go
    - Tense: past
  - Object: Rome
  - Subject: Chris
- KR
  - Visit(Chris, Rome)

Towards *semantic* relations

Chris travelled to Rome

- NLP
  - Verb: travel
    - Tense: past
  - Object: Rome
  - Subject: Chris
- KR
  - Visit(Chris, Rome)

Towards *semantic* relations

To Rome Chris did go

- NLP
  - Verb: do
    - Tense: past
  - Object: direct: go; indirect: Rome
  - Subject: Chris
- KR
  - Visit(Chris, Rome)

Towards *semantic* relations

Chris took a trip to Europe, his first stop was Rome

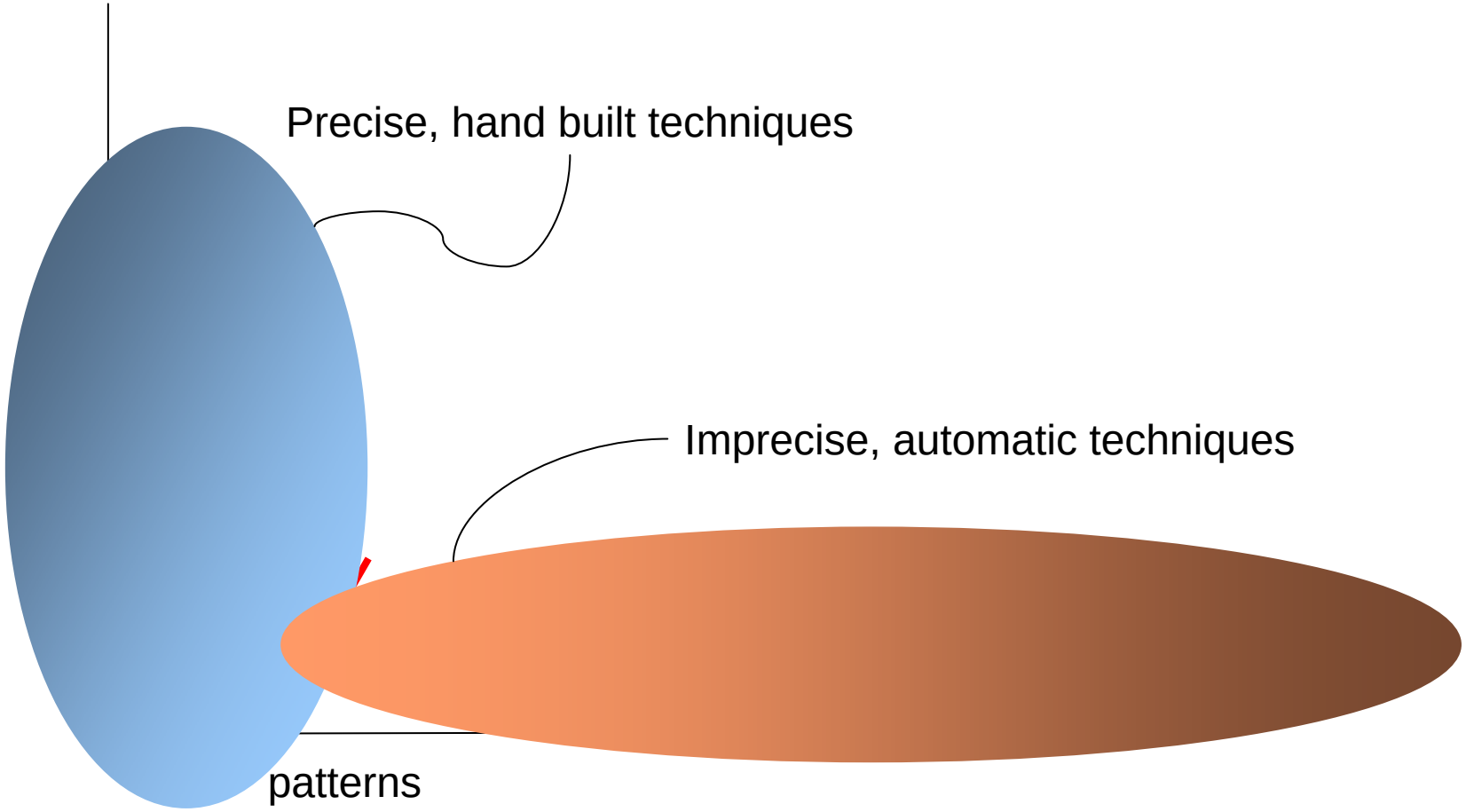
- NLP
  - Verb: be
    - Tense: past
  - Object: Rome
  - Subject: stop
- KR
  - Visit(Chris, Rome)

Towards *semantic* relations

In homage to his name-sake, St. Christopher, the itinerant traveler breezed through the *Citta de Buon Consiglio* like a politcal candidate, stopping only for a few glasses of Trentino wine and some speeches to well-wishing onlookers…

- NLP
  - ???
- KR
  - Visit(Chris, Rome)

The very long tail



Practical considerations

- *Be realistic:* Proper combination of NLP and KR (or any technology pairing) requires looking at the state of the art in *both* fields to see how they can be practically combined.
- *Don't proceed blindly:* Being realistic requires understanding capabilities of all parts of a system.
- *Improvements over perfection:* Being better is all that matters, don't let possible failure prevent a technique from being used
- *Go with what works:* Be prepared to shift directions as opposed to pushing one thing endlessly.

Question Answering

- Q: What *weapon* was featured in the ballet “Fall River Legend?”
- Ans from search: American Ballet Theatre
- Ans from Semantic Web [RDF]: none
- Ans from Semantic Web [OWL-DL]:

  …still waiting for result

- Shift directions: can we use SW to check for “stupid answers”?
  - E.g: Is “American Ballet Theatre” a weapon?

UIMA

Question Answering

- Q: What *gum's* motto was "Double your pleasure, double your fun"?
- A: personal lubricant

IBM

Precision

- Range/Domain Constraints
  - "He arrived at Bush International Airport at 10:00"
    - Person("Bush International Airport"), Person(He)
    - Org("Bush International Airport")
    - At(He, "Bush International Airport")
- Lack of Background Knowledge
  - "Kisumu is the AIDS capital of Kenya."
  - capitalOf(Kisumu,Kenya)
- Logical Consistency
  - "The governor of Alabama visited Georgia, Oct 1."
  - At(gov, Alabama, t1), At(gov, Georgia, t1)

Knowledge for Improving *Precision*

- Range/Domain Constraints
  - At ☒ Person × Place
- Background Knowledge
  - "Kisumu is the AIDS capital of Kenya."
  - Capital of Kenya is Nairobi
- Logical Consistency
  - Vehicle in two places at the same time
- Temporal Consistency
  - A time point both *before* and *after* the same event

Recall

- (Linguistic) Coreference is largely syntactic, misses “Joe, *the nose to his friends, …*”
- Basic search/NLP miss answers for “people in Mass.” given “Joe in Boston”, same for time
- Search/NLP cannot typically connect entities across documents
  - NYT article by Joe
  - Other article: Joe in Boston
  - Looking for Journalists in Boston

Knowledge for Improving *Recall*

- Semantic Relations
  - Has-alias
- Containment axioms
  - Boston in Massachusetts in US
  - April 23, 2003 in April, 2003, in 2003
- Classification axioms
  - Author of newspaper article a journalist
- Spatial axioms
  - Passenger on vehicle located where vehicle is located
- Temporal ordering axioms
  - Transitivity of *before, after*
  - Full Allen calculus
- Requires different notion of evaluation

Experiment

- Question
  - Will dropping inconsistent triples lead to an improvement in precision?
- Approach
  - During knowledge integration, add each mapped triple to KB and check consistency
  - If inconsistent, remove triple from KB (drop)
    - Creates an order dependency
  - If consistent, leave it

UIMA

Experimental Setup

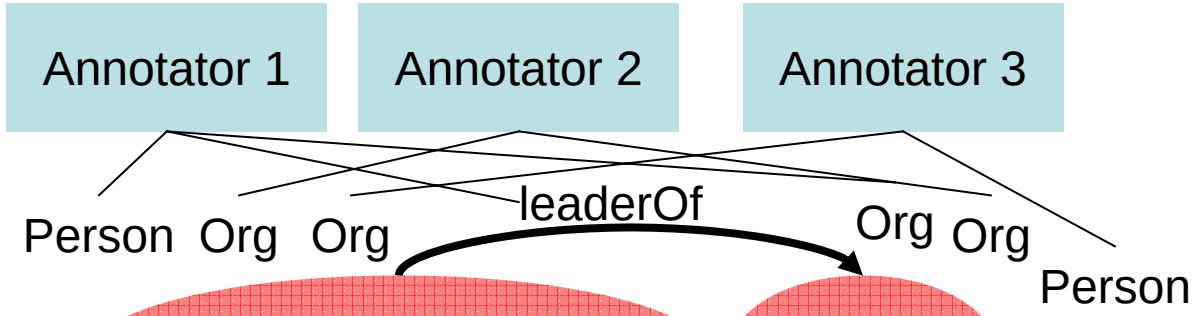
- OWL Ontology: 56 classes, 62 Object properties
  - global domain/range axioms
  - disjointness
- Corpus: 378 of 30K news articles
- Analytics
  - 42 OTS UIMA components producing 205 entity types and 79 relation types
  - 322 entity annotations/document
  - 21 relation annotations/document
  - 15 entity annotations/entity (coreference)
  - 1.8 relation annotations/relation (coreference)
- Knowledge Integration (RDF)
  - 6281 individuals
  - 834 object property triples

IBM

Results

- Experiment 1
  - Dropped 67 of 834 triples
  - 2 were correct (should not have been dropped)
    - Based on order dependency
  - Precision improves by ~8.7%
- Experiment 2
  - Dropped 1173 of 8898 triples
  - 400 evaluated, all incorrect
  - Precision improves by ~15%

Improving Precision by Dropping Triples



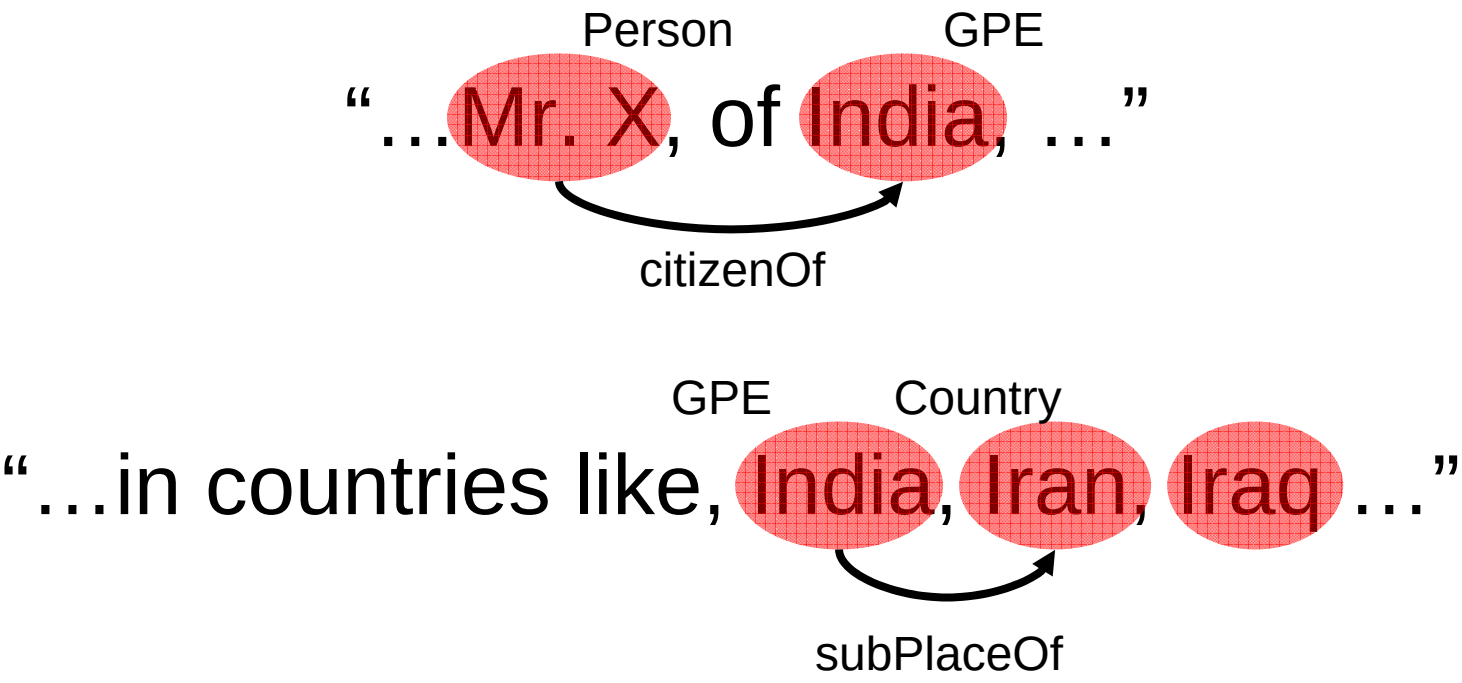
"…*Khmer Rouge forces allegedly drove off the personnel sent to repair the road…*"

Ontology

leaderOf: Person x Org

Person disjoint Org

Dropping triples after coreference



Ontology: Country subClass GPE

Country subClass (all subPlaceOf (GPE and (not (Country)))

citizenOf: Person x Country

Axiomatic Constraints

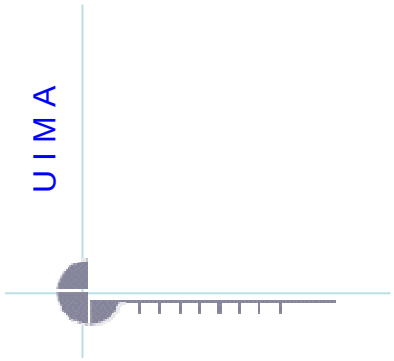
- Global
  - *basedIn*: (Facility | ORG) x GPE
  - *employeeOf*: Person x (GPE | ORG)
    - *managerOf* subPropertyOf *employeeOf*
  - *ownerOf*: (GPE | ORG | Person) x **not**(Person)
  - *memberOf*: (GPE | ORG | Person) x ORG
  - *citizenOf*: Person x Country
  - *from*: ChangePos x GeoPlace
  - *arrester*: Arrest x (GPE | ORG | Person)
- Local Constraints:
  - State: *subPlaceOf* . Country
  - Country: *subPlaceOf* . **not**(Country | State | City)
  - GeoPlace: *near* . (GeoPlace | SpatialObject)
  - SpatialObject: *near* . (GeoPlace | SpatialObject)
  - Person: *phoneRegistry* . PhoneNumber

UIMA

Conclusions

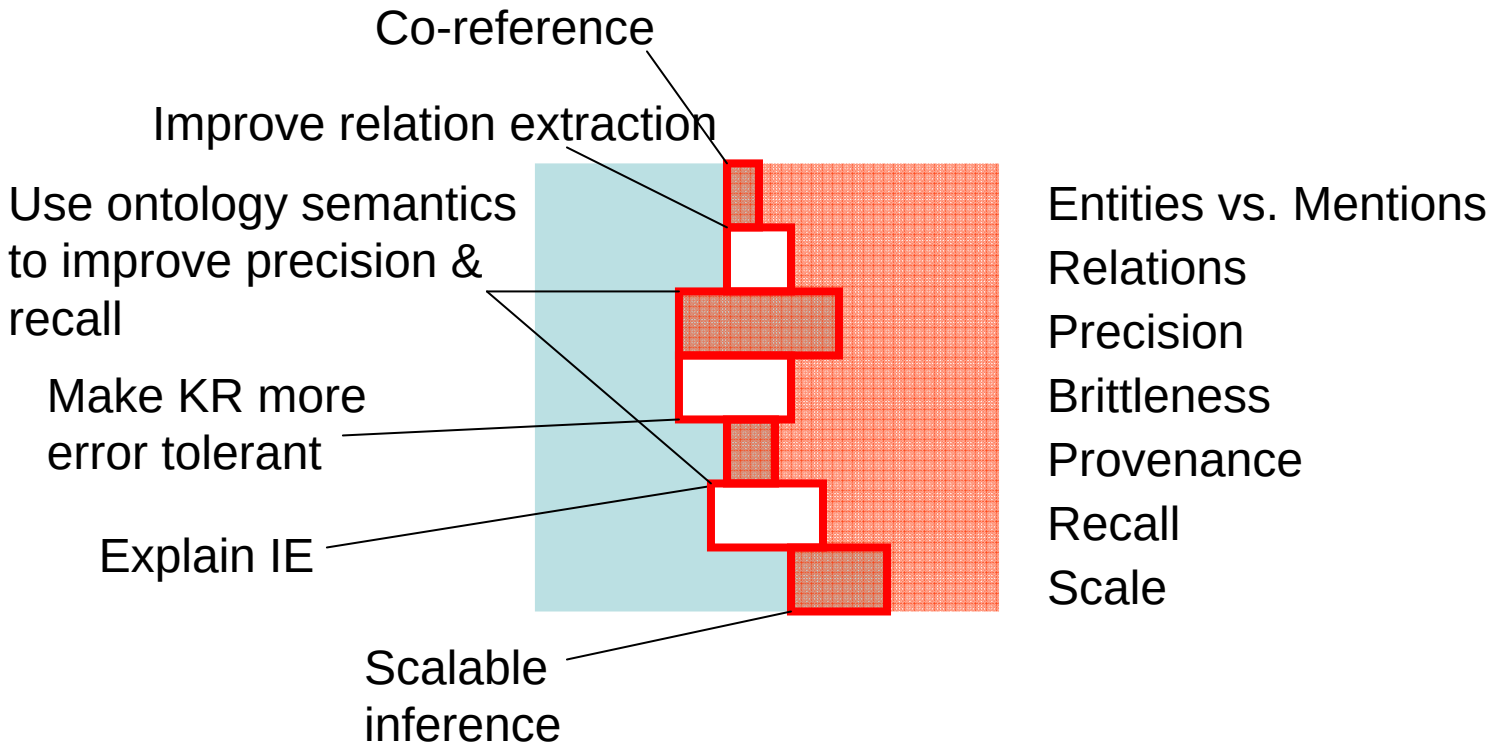
- Combining KR & IE faces many challenges
  - Different precision & recall assumptions
  - KR&R requires relations
  - Referents vs. annotations
  - Scalability
  - Explainability/Provenance
- Addressing *each* in a variety of ways
- SemantiClean
  - Improving P&R with semantic constraints
  - Could be used for any kind of *noisy* data
    - Sensors, databases, etc.
- Interested in *improvements* not perfection

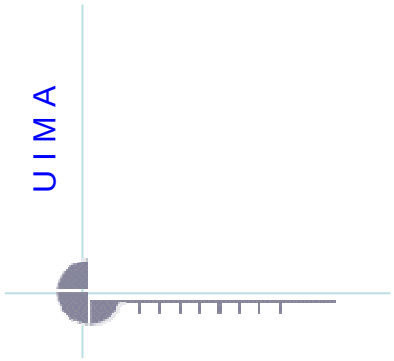
Use SemantiClean for all your data cleaning needs!


UIMA


IBM

IE ↔ KR


Co-reference
Improve relation extraction
Use ontology semantics to improve precision & recall
Make KR more error tolerant
Explain IE
Scalable inference
Entities vs. Mentions
Relations
Precision
Brittleness
Provenance
Recall
Scale



Recall

Same word, different meaning

- Logical form
  - NLP: vp(
- Semantic

Current Problems in NLP

- Improve precision & recall of search
  - Precision: constraints
  - Recall: deduction
- Note: not about being *correct*
  - *Just better*
  - Show that KR&R technologies *add value* to the problem

UIMA

Real World Reasoning?

- The acquisition bottleneck
- Reasoning is *hard*
  - Complexity
    NP < Exp < Non-Elementary < Undecidable
  - More than 1,000,000 RDF triples
  - Can't hold results of 1000 documents in memory
- Reasoning is *sound*
  - Not tolerant of errors
    <Person>Bush International</Person>
  - Not just precision
    Iraqi press agency says <event>the war is ended</event>
- Reasoning can be *inscrutable*
  - If I have no friends then all my friends are doctors

Text Analysis

How much reasoning when

Bounded reasoning

Explanation

What is needed?

- Reasoning
  - Temporal containment
  - Spatial containment
- Knowledge
  - Geography
  - Basic Ontology
- Coreference
  - Multiple mentions of same *entities, relations*
  - Find time that something happened
- Extraction
  - Recognize people, places
  - Recognize times

Former Iraqi Ambassador to Canada
Hisham Al Shawi, who defected to the UK in August 1993, was a key figure in launching Iraq's nuclear program. From 1972 to
1974, Al Shawi was chairman of Iraq's Atomic Energy Commission and was Iraq's representative on the board of the IAEA. Al Shawi helped coordinate Iraq's efforts to train Iraqi students and send them abroad to universities and international research facilities. However, there is no evidence that Al Shawi knew of efforts to use these trainees in a military nuclear programme.
During the week of Feb. 26, 2003, Al Shawi travelled to France and proceeded up the Seine by boat. He spent three days in the latin quarter, and returned on Mar. 1.
The purpose of his trip was not known.
Al Shawi and a fellow defector, former ambassador to Tunisia Hamid al Jabbouri, denied having any detailed knowledge of Iraq's nuclear program or its procurement network.

UIMA

General Problem

- *Given*
  - an ontology in OWL
  - A background knowledge base in RDF
  - Inference procedures
  - A collection of existing IE components
- Use the results of IE to populate the KB
  - Map IE semantics to KB semantics
  - Map extracted entities to possibly existing KB instances

Semantic Integration Goals

- Process results of IE into a form *suitable for reasoning*
  - i.e. by devil's advocate
- *Map* linguistic structures into knowledge-base
  - Different ontologies
  - Different semantics

  - Moving targets
  - Declarative
- *Clean up* the results using ontology semantics
  - Improve precision and recall
  - Propose candidate contexts
- *Scale* the results along some dimension of "massive"
- *Evaluate* the quality of the results

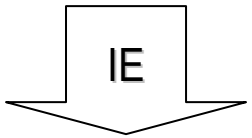
Broader Contributions

- Reuse&Adapt existing work
  - Chimaera, OntoMerge, Prompt
- Develop catalog of integration inferences
  - Extension to Prompt
  - Moving towards semi-automation for linguistic ontologies
- Mapping TimeML ↔ DAML-Time
- Deeper semantics of IE
- Improving P&R of IE

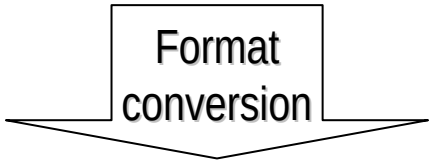
UIMA

Semantic Integration

“13 delegates from Turkey arrived today.”



“13 delegates from <country>Turkey</country> arrived today.”



<OWL:country rdf:ID=“Turkey” />

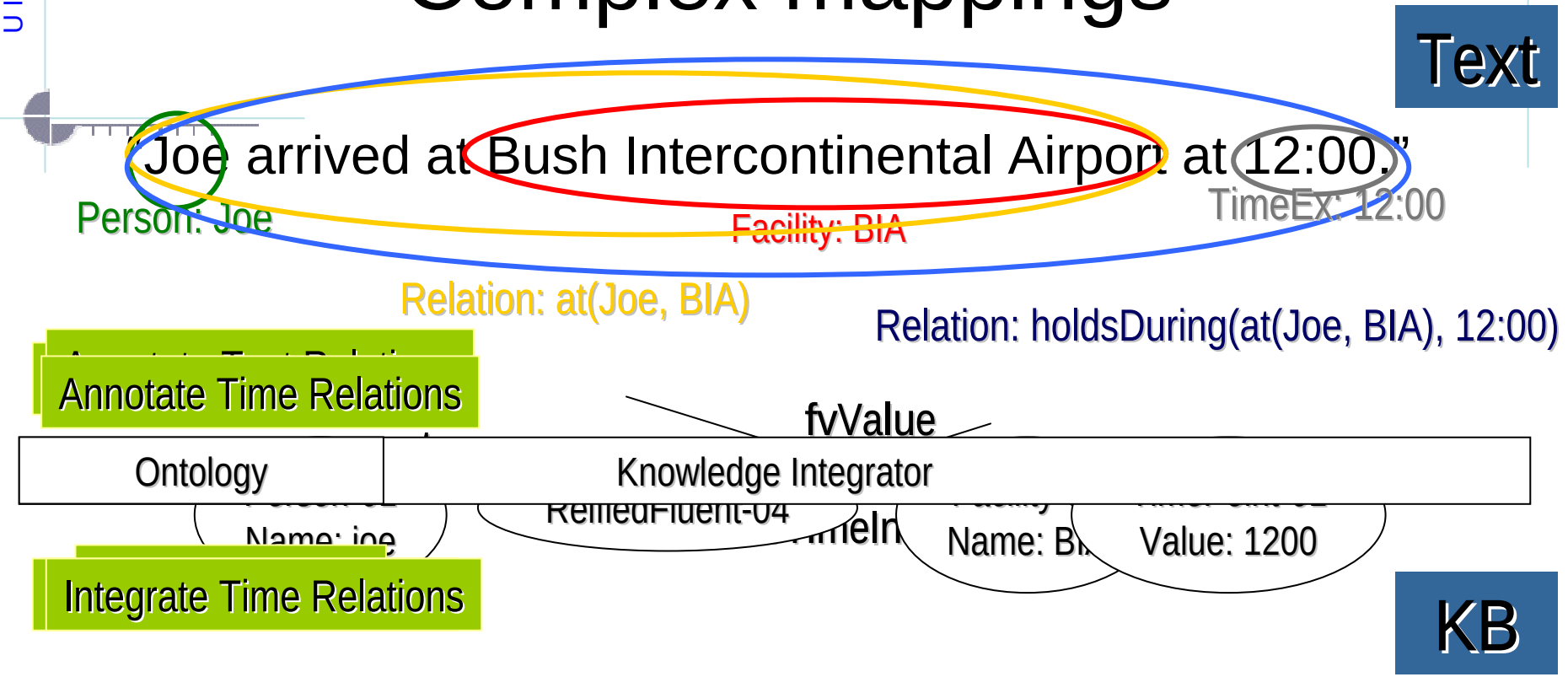
Easy!!!

IBM

Mapping from Extraction to Knowledge

- Type-Class Mapping
  - Person → Person
  - Country → Location & Political Entity
- Entity Mapping
  - Person(Abduhl Ramazi) → kani:person-101
- Relations
  - IE:At(Person,Place) → KANI:At
  - IE:At(Place,Place) → KANI:subPlace
- Complex Mappings
  - HoldsDuring(At(Person, Place), TimeInterval) →
    At(Person, *fv*) & fvValue(*fv*, Place) & fvTimeInterval(*fv*, TimeInterval)
  - Uses(BioTerrorism, Diseases) → …

Complex mappings



UIMA

Catalog of Ontology Merging Operations

- Simple mappings
  - $Class_1 \rightarrow$ Existing $Class_2$
  - $Class_1 \rightarrow$ New $Class_2$
  - $Class_1$+$Class_2$ Subclass new $Class_2$
  - …
- Complex mappings
  - $Class_1 \rightarrow Set_2$ of Classes
  - Set_1 of Classes $\rightarrow Class_2$
  - …
- Language Expressivity
  - DL-expressible [Halevy] [Lenzerini]
  - Function-free FOL expressible [McDermott & Smith]
  - FOL expressible [Chalupsky & MacGregor]
  - Non-FOL

IBM

Mapping TimeML to DAML-Time

- TimeML
  - Markup language with linguistic-based semantics
  - time expressions (Timex)
  - Events
  - Links (before, after, …)
- DAML-Time
  - Ontology-based specification of time points and intervals
  - Based on Allen calculus
  - No events
- High level correspondence [Pustejovsky&Hobbs]
- Generate complete DAML-Time RDF for TimeBank 1.1 corpus

Deeper Semantics for IE

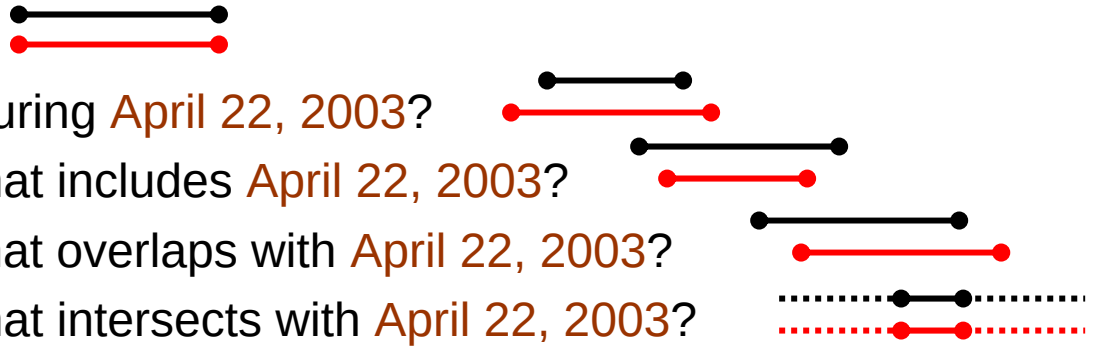
- IE focused on surface semantics
- Surface semantics appear obvious
  - Person(Abduhl Ramazi)
  - onBoard(ramazi, train)
- Requirement for reasoning exposes problems

"Ramazi was in his office on April 22, 2003."

Holds(in(Ramazi, office), *t1*)

What is *t1:*

- April 22, 2003?
- A time interval during April 22, 2003?
- A time interval that includes April 22, 2003?
- A time interval that overlaps with April 22, 2003?
- A time interval that intersects with April 22, 2003?



UIMA

Using Semantics to boost precision/recall

- "A man was arrested, his name was given as Chris"

Co-reference chain

Relation: nameOf(his, Chris)

Cannot find a link from "his" to "Chris" – the relation is not lexical, it's semantic

During integration, the semantics of *name* relation are processed and Chris assigned as name

- "He arrived at Bush Intercontinental at 12:00"

Relation: at(he, Bush Intercontinental)

Type: Person, Facility

Entity extraction tags "Bush Intercontinental" with Person and Facility

Relation extraction finds at(person, Bush Intercontinental)

Semantics of *at* relation requires range be a facility or place

Semantics of Person and Facility are of disjoint classes

Person annotation thrown away

IBM

Knowledge for Improving Precision

- Range/Domain Constraints
  - At ☒ Person × Place
- Background Knowledge
  - "Kisumu is the AIDS capital of Kenya."
  - Capital of Kenya is Nairobi
- Logical Consistency
  - Vehicle in two places at the same time
- Temporal Consistency
  - A time point both *before* and *after* the same event

Improving Recall

- Semantic Relations
  - Has-alias
- Containment axioms
  - Boston in Massachusetts in US
  - April 23, 2003 in April, 2003, in 2003
- Classification axioms
  - Author of newspaper article a journalist
- Spatial axioms
  - Passenger on vehicle located where vehicle is located
- Temporal ordering axioms
  - Transitivity of *before, after*
  - Full Allen calculus

Trials and Tribulations

- Reasoning helps information processing
  - Increase in recall through deductive inference
  - Increase in precision through constraint processing
- General-purpose reasoning algorithms are complex
  - OWL-lite is Co-NP
  - OWL-DL is EXPTIME
  - OWL, General First-order reasoning is undecidable
  - Allen Calculus undecidable
- In *practice* this means reasoning must be bounded
  - A *tradeoff* between scale and effectiveness of reasoning
- Research agenda – explore this tradeoff
  - Techniques for bounding data (e.g. partitioning)
  - Techniques for "hiding" data
  - Incrementally apply more sophisticated reasoning (staged reasoning)
  - Measure KB and Ontology *complexity* (what is massive?)

Hiding Data - Motivation

- IE generates a lot of data
- Need to reduce the burden on reasoning-based processes
  - RDF-based reasoners limited to ~500K nodes
- Can we reduce the amount of data?
- What if it is needed later?

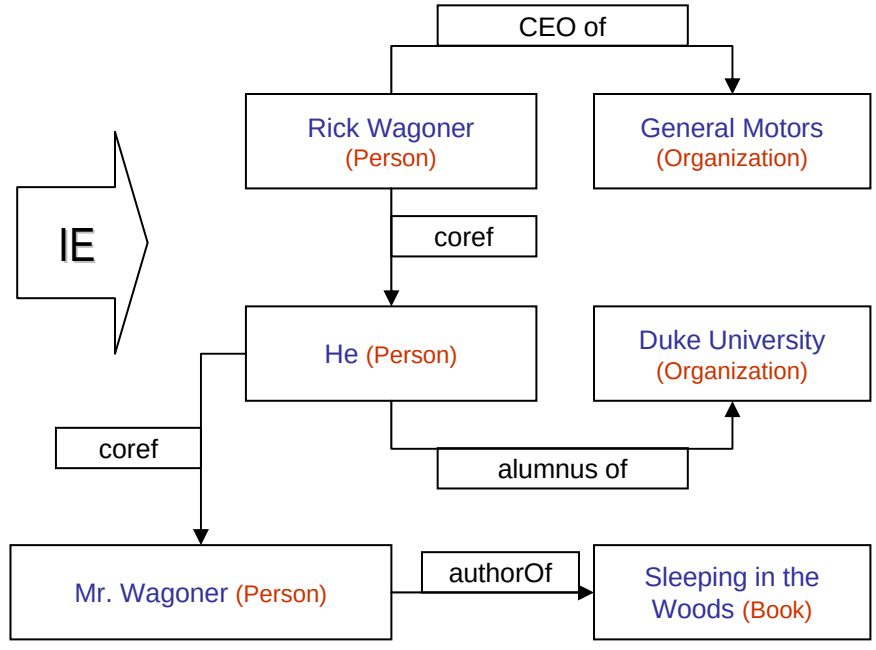
UIMA

Data Volume Analysis

… Rick Wagoner is CEO of General Motors. He is an alumnus of Duke University…

… Mr. Wagoner, the author of *Sleeping in the Woods*…

IE



SI

P1 type Person
P2 type Person
O1 type Org
O2 type Org
P1 ceoOf O1
P2 alumnusOf O2
P1 sameAs P2
P3 type Person
P3 sameAs P2
P3 authorOf O3
O3 type Book

Two triples for every co-reference

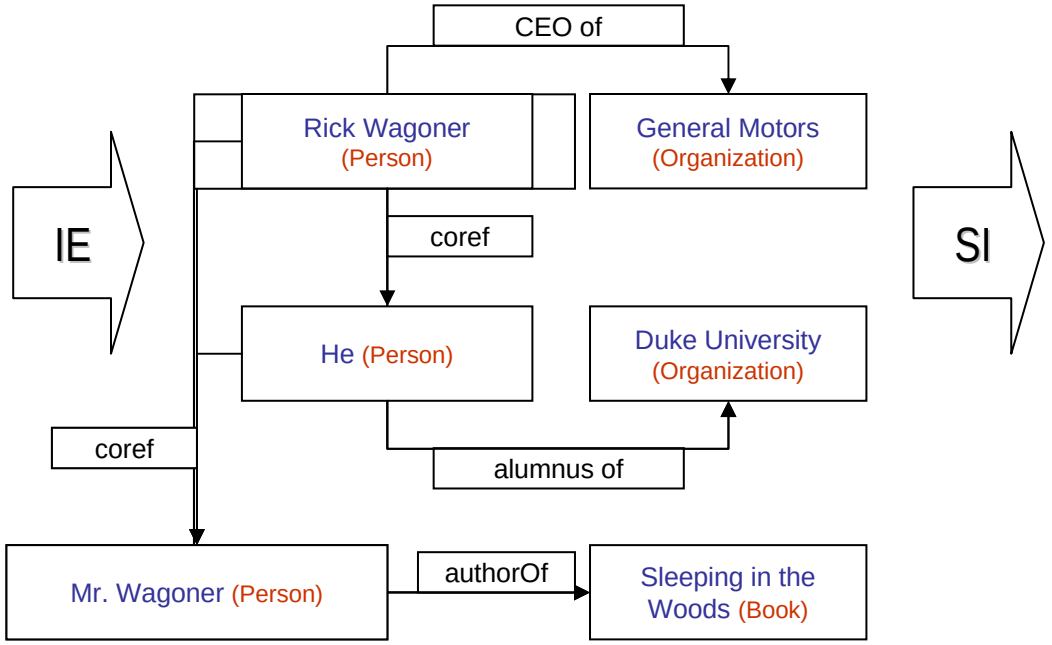
- 2.6 mentions per instance
- 20 instances per document
- 136 triples per document
- *64 triples per document for representing coreference.*

IBM

Hiding Data

… Rick Wagoner is CEO of General Motors. He is an alumnus of Duke University…

… Mr. Wagoner, the author of *Sleeping in the Woods*…



SI

P1 type Person

O1 type Org
O2 type Org
P1 ceoOf O1
P1 alumnusOf O2

P1 authorOf O3
O3 type Book

- 47% reduction in the number of RDF triples
- 62% reduction in the number of RDF nodes
- But what if it's wrong…

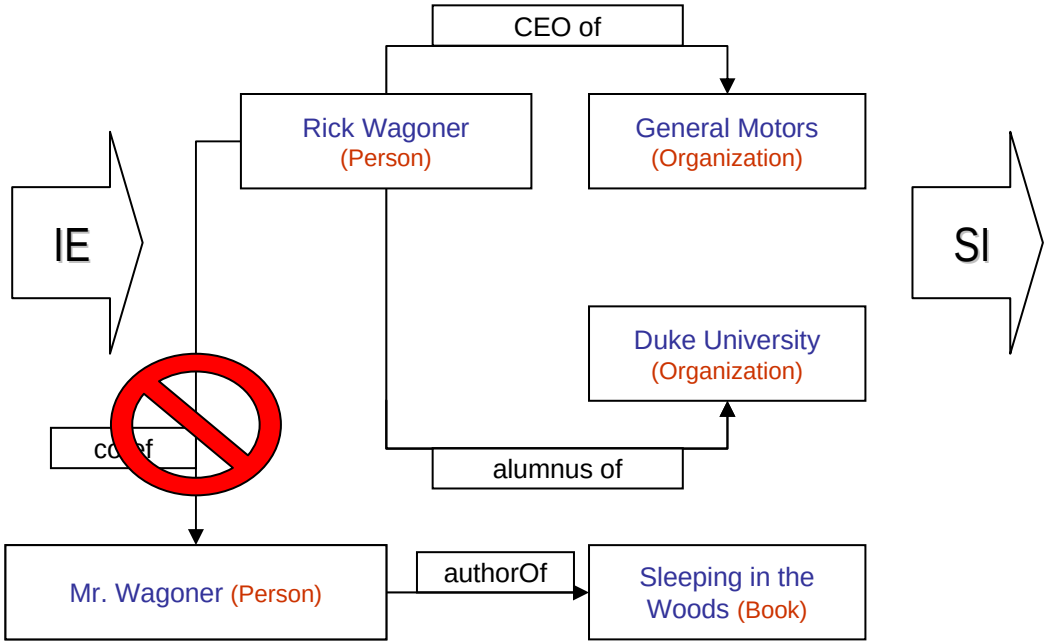
IBM

UIMA

Hiding Data

… Rick Wagoner is CEO of General Motors. He is an alumnus of Duke University…

… Mr. Wagoner, the author of *Sleeping in the Woods*…



SI

P1 type Person

O1 type Org
O2 type Org
P1 ceoOf O1
P1 alumnusOf O2

P3 type Person
P3 sameAs P1
P3 authorOf O3
O3 type Book

- 47% reduction in the number of RDF triples
- 62% reduction in the number of RDF nodes
- But what if it's wrong…
- "Hide" IE inferences in provenance
- Expose when problem arises

IBM

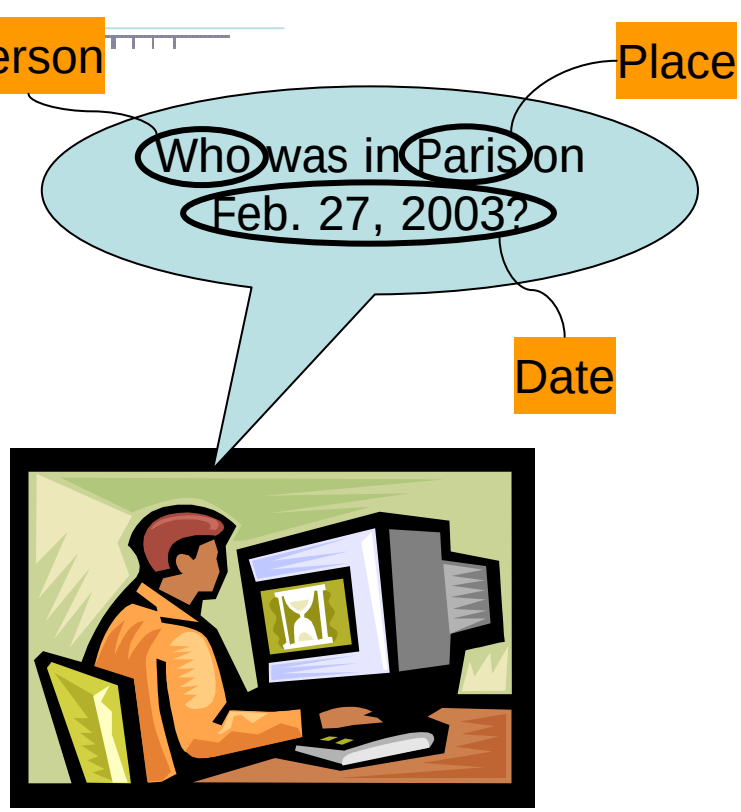
Key Challenges

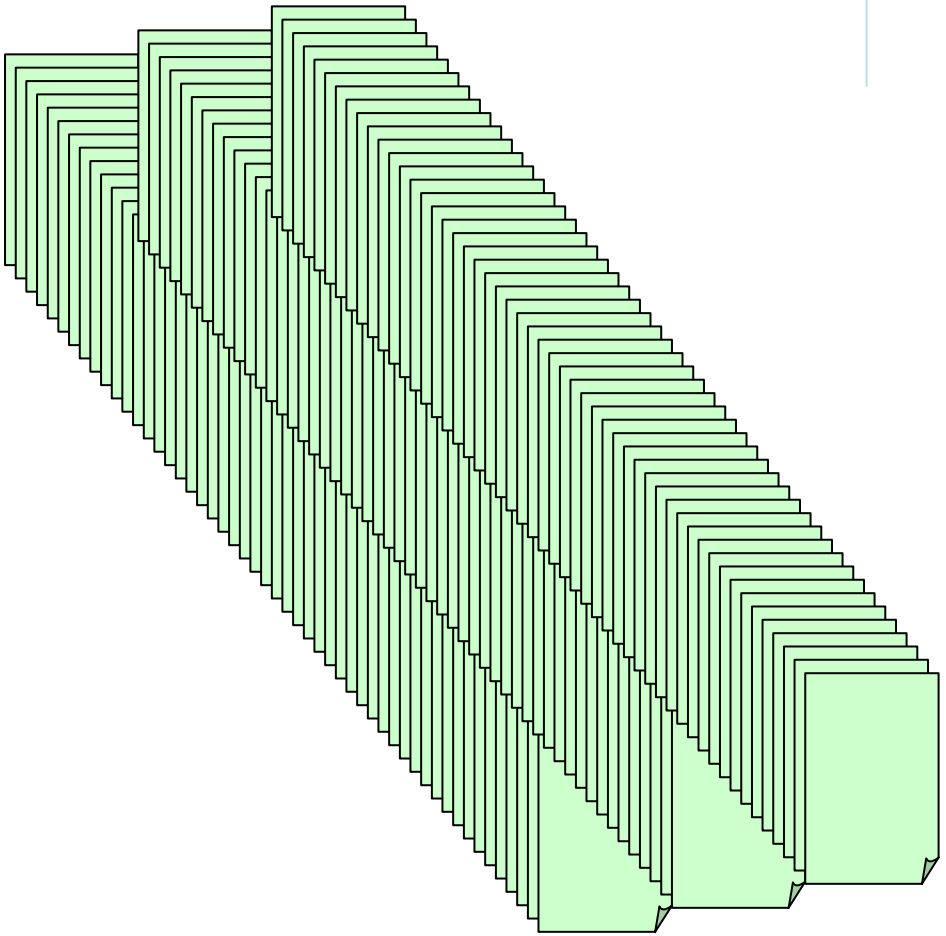
- How much reasoning when
- Artifacts of established IE
  - Imprecision
  - Co-reference
  - Multiple annotations
- Confidence, Trustworthiness
- Maintaining provenance through mapping
- Limits of Automated Reasoning for integration
  - Limits of description logic reasoning (OWL/DL) [Lenzerini, 2002], [Halevy, 2001]
  - Limits of first order reasoning (FOL) [McDermott, et al, 2002]
- Axiomatic semantics of information extraction
  - Further clarify semantics [ACE, KDD]
  - Boost precision
- Adaptability
- Evaluation

How much reasoning when?

- Incremental value of increased processing
- Many dimensions of complexity
  - Representational
    - Worst-case complexity
    - Special-purpose reasoning
    - Optimizations
  - Instance level
    - Number of instances, relations
    - Connectedness
    - Precision
  - Ontology level
    - Number of classes, properties
    - Number of axioms, constraints

Why extraction?





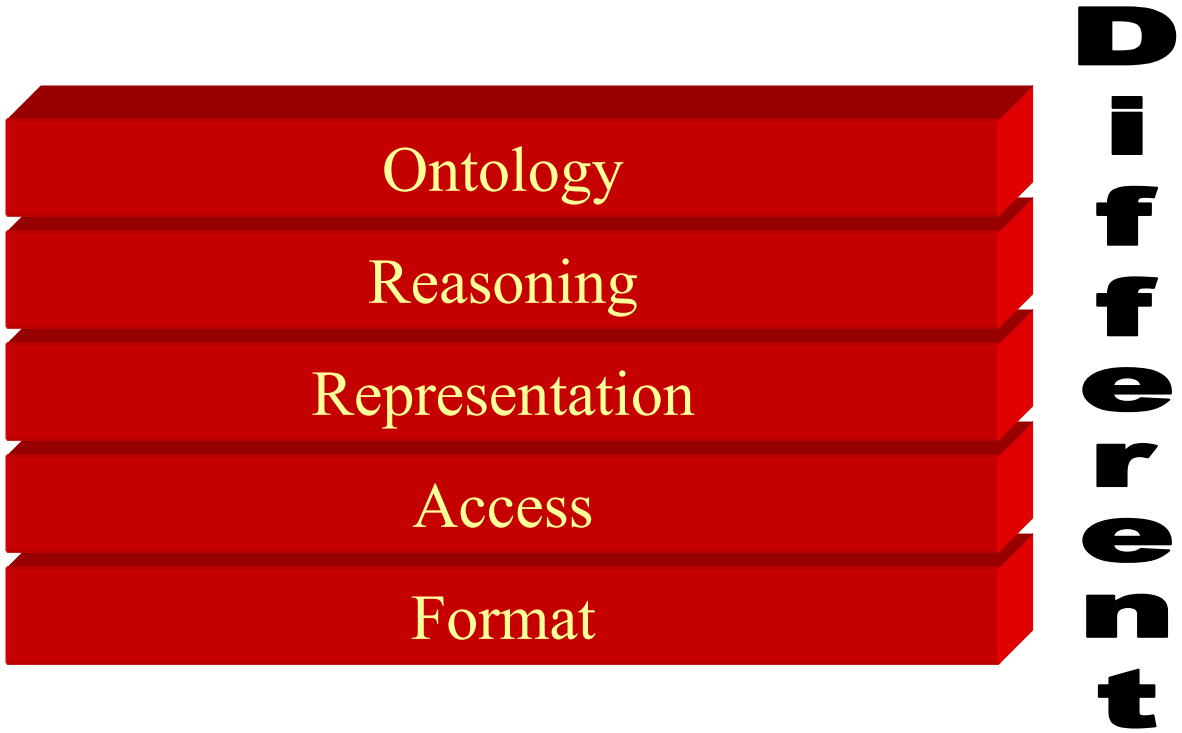
Why reasoning?

Former Iraqi Ambassador to Canada
Hisham Al Shawi, who defected to the UK in
August 1993, was a key figure in launching Iraq's nuclear program. From 1972 to
1974, Al Shawi was chairman of Iraq's Atomic Energy Commission and was Iraq's representative on the board of the IAEA. Al Shawi helped coordinate Iraq's
efforts to train Iraqi students and send them abroad to universities and international research facilities. However, there is no evidence that Al Shawi knew of efforts to use these trainees in a military nuclear programme.
During the week of Feb. 26, 2003, Al Shawi travelled to France and proceeded up the Seine by boat. He spent three days in the latin quarter, and returned on Mar. 1.
The purpose of his trip was not known.
Al Shawi and a fellow defector, former ambassador to Tunisia Hamid al Jabbouri, denied having any detailed knowledge of Iraq's nuclear program or its procurement network.

Why reasoning?

Former Iraqi Ambassador to Canada
Hisham Al Shawi, who defected to the UK in August 1993, was a key figure in launching Iraq's nuclear program. From 1972 to
1974, Al Shawi was chairman of Iraq's Atomic Energy Commission and was Iraq's representative on the board of the IAEA. Al Shawi helped coordinate Iraq's efforts to train Iraqi students and send them abroad to universities and international research facilities. However, there is no evidence that Al Shawi knew of efforts to use these trainees in a military nuclear programme.
During the week of Feb. 26, 2003, Al Shawi travelled to France and proceeded up the Seine by boat. He spent three days in the latin quarter, and returned on Mar. 1.
The purpose of his trip was not known.
Al Shawi and a fellow defector, former ambassador to Tunisia Hamid al Jabbouri, denied having any detailed knowledge of Iraq's nuclear program or its procurement network.

Obstacles


Ontology
Reasoning
Representation
Access
Format
Different

UIMA

Starting Point

Simple Token and Sentence Annotator

Annotates tokens and sentences.

Ace Annotator

A statistical entity and relation annotator that performs within-document coreference resolution.

KS Relation Detector Aggregate

Aggregate that includes KS Relation Detector and TAE's that provide its inputs

XsgParsing Annotator

Performs deep parse using slot grammar parsers

PhraseFinder Annotator

Creates annotations for phrases to be passed to ESG. Phrase sources are WordNet and pre-annotated Resporator phrases.

EAnnotator

IBM EAnnotator - A statistical entity, relation, and event annotator.

Cross Annotator Coreference

Resolves correference disputes across annotators

KS Relation Annotator

Knowledge Structures Relation Annotator (pattern-based relation detector)

Simple Phone Call Relation Annotator

Finds mentions of phone calls in text and annotates them as MadePhoneCallTo relations.

Simple Phone Number Annotator

Finds phone numbers in text

UIMA

Starting Point

KS Relation Detector Aggregate

Aggregate that includes KS Relation Detector and TAE's that provide

Ace Annotator

tistical entity and on annotator that rms within- ment corefer

GlossOnt

Finds taxonomic and other definitional relations from

Simple Token a Sentence Anno

A se

HoldsDuring

Extracts Relations between TimeEx3 entities and relations.

CrossDocumentCoreference

Merges coreference chains across documents.

Nominator

Finds prope and other clues

tations for passed to ESG.

ordNet

E

IBM EAnnotator - A statistical entity, relation, and event ann

KDDAnnotator

A statistical entity relation annotator within- coreference .

TFSTOnBoard

Finds onBoard relations between people and vehicles.

TFSTVehicle

Finds vehicles.

entities.

TFSTAddr

Finds addresses and extracts the subplace relation.

Simp Anno

Finds mentions of phone calls in text and annotates them as MadePhoneCallT relations.

reference

TFSTTime

Extracts TimeE entities.

notator

e Structures nnotator (pattern-

JResporator

Extracts over 80 classes of entities.

TFST

Finds a number of standard named entities.

Finds phone numbers in text

Issues in the combination

- Components have *overlapping* semantics
  - Common type system, but…
  - Different meanings, precision, recall
- Multiple annotations on a single span
  - Often agree
  - Occasionally do not
- Multiple overlapping co-reference chains

Knowledge Integration