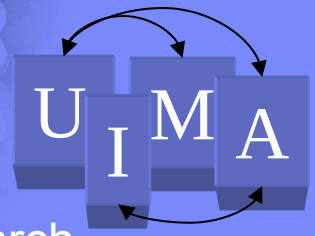


Putting the Semantic In the Semantic Web

An overview of UIMA and its role in Accelerating the Semantic Revolution

Ontolog Forum, May 11, 2006

David A. Ferrucci
Senior Manager, Semantic Analysis and Integration
Chief Architect, UIMA
IBM T.J. Watson Research Center

© 2006 IBM Corporation – All Rights Reserved

Opening Observations

- **Semantics are key to Knowledge Discovery in Unstructured Sources**
  - Unstructured sources is where the action is
  - Keywords and Known-Item Search don't cut it
  - Need Richer Semantics
- **Massive manual semantic annotation is not likely**
  - A challenge for the Semantic Web vision
  - User View vs. Authors view
  - Automated Annotation is Essential
  - Many Component Annotators must emerge and combine
- **The right enabler and the right expectations**
  - Must enable rapid annotator development and integration
  - Applications can add value while tolerating less that perfect precision
  - Automation is necessary to tag and help humans reduce/focus content

© 2006 IBM Corporation – All Rights Reserved –

Overview

- **Semantics Matters – A Motivating Example**
- **UIMA: Facilitating Automatic Semantic Discovery**
- **Using Semantics to Raise the Search *Bar***
  - *Finding the Sweet-Spot*
  - *Semantic Search and the SAW*
  - *Knowledge Extraction*

© 2006 IBM Corporation – All Rights Reserved –

Semantic Match vs. Key Word Match

SEARCH: Going rate for leasing a billboard near Triborough Bridge

Top hits from popular search engines miss the mark…

Keywords may Match

BUT **WRONG** *Content returned*

And right content ***MISSED***

© 2006 IBM Corporation – All Rights Reserved –

SEARCH: Going rate for leasing a billboard near Triborough Bridge

Wired New York Forum - Long Island City Development
"it's Right over The 59th Street **Bridge**." **...** it's **going** to become our SoHo. The Powerhouse is most likely **going** to become a performing Arts facility for **...**
www.wirednewyork.com/forum/ printthread.php?t=5160&pp=50 - 150k - Supplemental Result - Cached - Similar pages

Long Island City Development [Archive] - Wired New York Forum
... Creek to the **Triborough Bridge** - would become an attractive destination. **...** Several Long Island City sites are now in play - Junger himself's **going ...**
www.wirednewyork.com/forum/ archive/index.php/t-5160.html - 255k - Cached - Similar pages
[More results from www.wirednewyork.com]

[PDF] Austin's Wireless Future
File Format: PDF/Adobe Acrobat - View as HTML
In Section 2 we explore the **near**-term future of wireless-- ubiquitous mobile **...** applications such as the **Triborough**. **Bridge** installation, wireless sensors **...**
www.ic2.org/publications/AustinsWirelessFuture.pdf - Similar pages

[PDF] Funding of the High-Speed Rail System
File Format: PDF/Adobe Acrobat - View as HTML
Triborough Bridge. and Tunnel. Authority**. Oklahoma. Turnpike Authority **...** management fees, reserves / on **going** TI's, and **leasing** commissions. **...**
www.cahighspeedrail.ca.gov/plan/pdf/Financial_Plan.pdf - Similar pages

The rail magazine for The computer age ...
Surprisingly, thanks to 9600 baud modems, lower **rates**, and automated access **...** his previous positions included President, **Triborough Bridge** and tunnel **...**
www.railfan.net/railpix/text/ro_1_6.txt - Supplemental Result - Similar pages

Checklist 1989
Triborough Bridge and Tunnel Authority, selected procurement and contract **...** Distribution of high school graduates and college-**going rate**, New York State. **...**
www.nysl.nysed.gov/edocs/education/check89.htm - 513k - Cached - Similar pages

[PDF] www.icc.utexas.edu/publications/AustinsWirelessFut...
File Format: PDF/Adobe Acrobat - View as HTML
Supplemental Result - Similar pages

"Going", "Bridge", "Rate", "Leasing", etc.

But All Misses

© 2006 IBM Corporation – All Rights Reserved –

SEARCH: Going rate for leasing a billboard near Triborough Bridge

* Were you looking for ***Going rate for leaving a billboard near Triborough Bridge***

Web Results

Page 1 of 4 results containing **Going rate for leasing a billboard near Triborough Bridge** (0.70 seconds)

Wired New York Forum - Long Island City Development

… on Vernon Boulevard **near** the Queensboro **Bridge**, the founders of Silvercup … Newtown Cr… the **Triborough Bridge** - would become an attractive … Pearson St. - where he's **going** to c… 20-story …

www.wirednewyork.com/forum/printthread.php?t=5160&pp=50 Cached page

FINANCIAL PLAN

Triborough Bridge and Tunnel Authority** Oklahoma Turnpike Authority … A…rage Annual Gr… **Rate**) 1979-1982 10.9% 1980-1990 5 … Leveraged **Leasing** of Rail Rolling Stock When rail rolling …

www.cahighspeedrail.ca.gov/plan/pdf/Financial_Plan.pdf Cached page PDF file

NEW YORK: Stadiums/Arenas (NETS, JETS, YANKS, and others) [Archive …

… backward from the adjacent **Triborough Bridge** elevated roadway that leads to … I was **going** to… the name of one of my companies on it to get … The area **near** the Continental Arena has become… construction …

forum.skyscraperpage.com/archive/index.php/t-65433.html Cached page

http://odur.let.rug.nl/ftp/pub/prolog-app/DutchBrillTagging/TaggerSoftware/Bin_and_Data/LEXICON.WSJ

Billboard NNP Billboarding NNP Billed VBN Billerica NNP Billheimer NNP Billiards NNP NN Billie NNP Billing NN Billings NNS NNP Billion CD Billionaire NN Billions NNS Billmeyer NNP Billock NNP Bills NNS …

odur.let.rug.nl/ftp/pub/prolog-app/DutchBrillTagging/TaggerSoftware/Bin_and_Data/LEXICON.W… Cached page

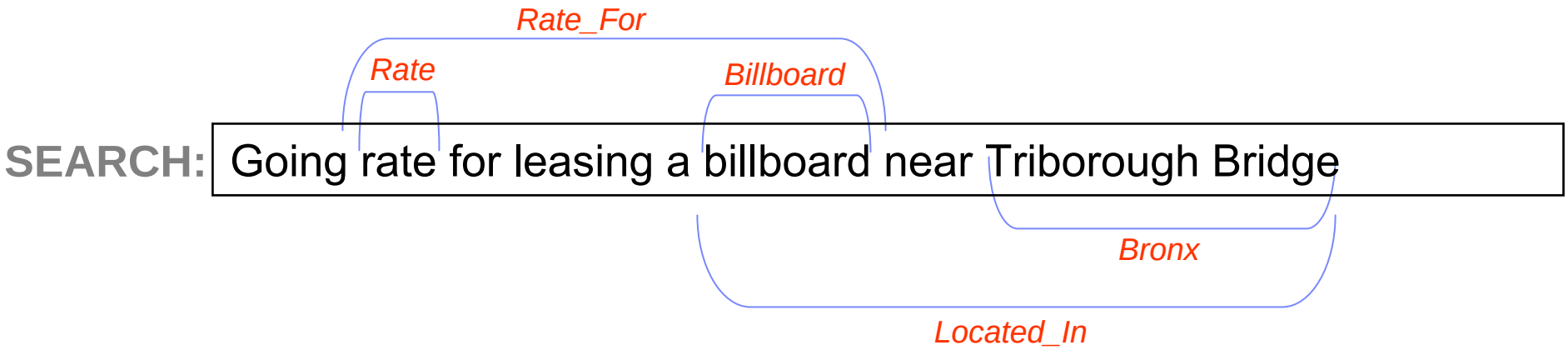
Another Search Engine
Some Different **Hits**
but
Still All Misses

Makes you wonder…

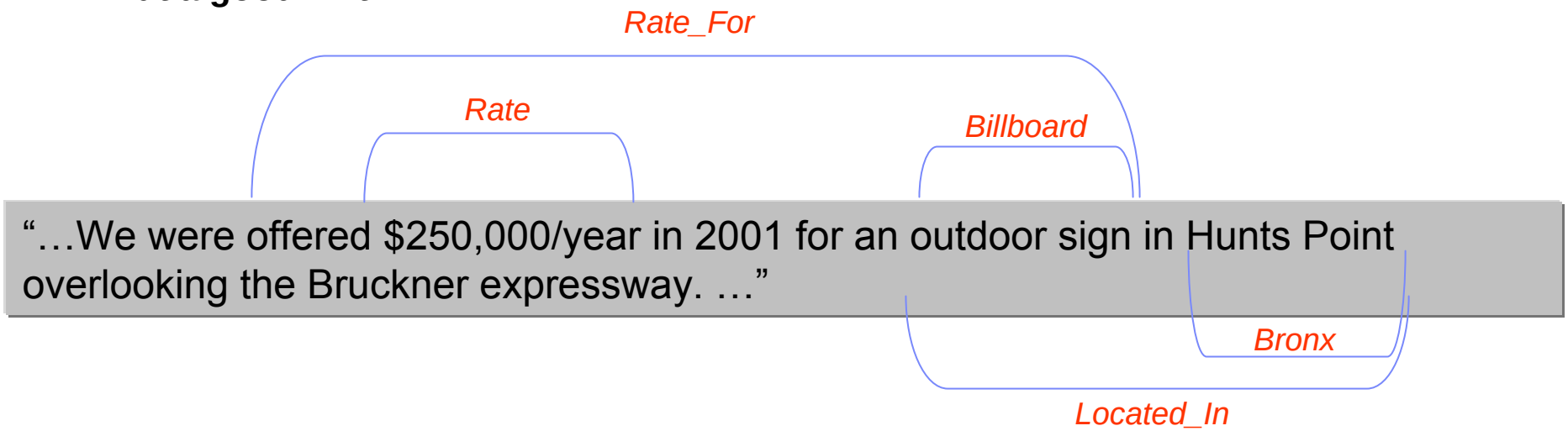
Could be out there?

© 2006 IBM Corporation – All Rights Reserved –

Semantic Match Improves Recall

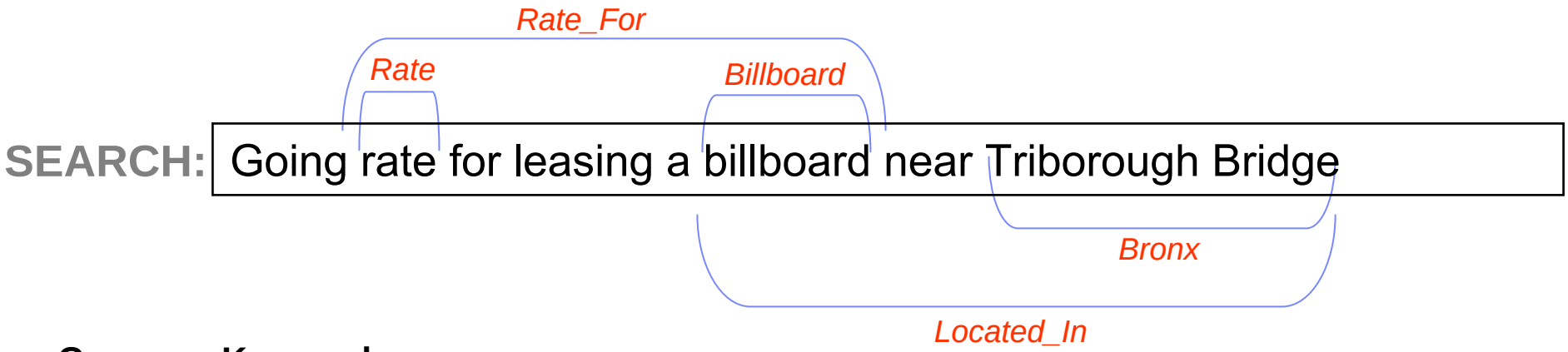


No Keywords in Common
But a good "hit"



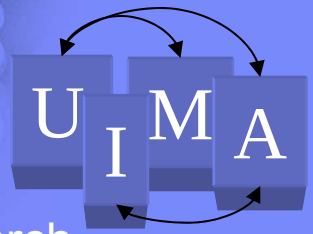
© 2006 IBM Corporation – All Rights Reserved –

Semantic Match Improves Precision



Common Keywords
Bad Semantic Match

© 2006 IBM Corporation – All Rights Reserved –



UIMA Overview

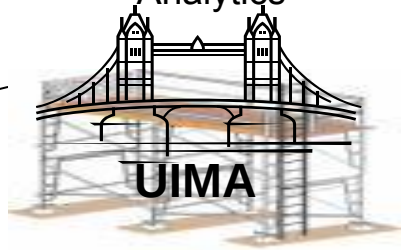
Unstructured Information Management Architecture

A open-source, pluggable framework for accelerating the science and business of semantic analysis

© 2006 IBM Corporation – All Rights Reserved

Analytics Bridge the Unstructured & Structured Worlds

Text and Multi-Modal Analytics

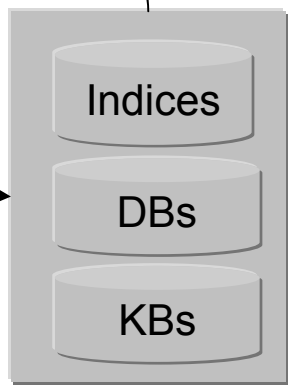


Unstructured Information

Discover Relevant Semantics → Build into Structure

- Docs, Emails, Phone Calls, Reports
- Topics, Entities, Relationships
- People, Places, Org, Times, Events
- Customer Opinions, Products, Problems
- Threats, Chemicals, Drugs, Drug Interactions....

Structured Information



- ***High-Value***
- ***Most Current***
- ***Fastest Growing (80% of Corporate Data)***

...BUT ...

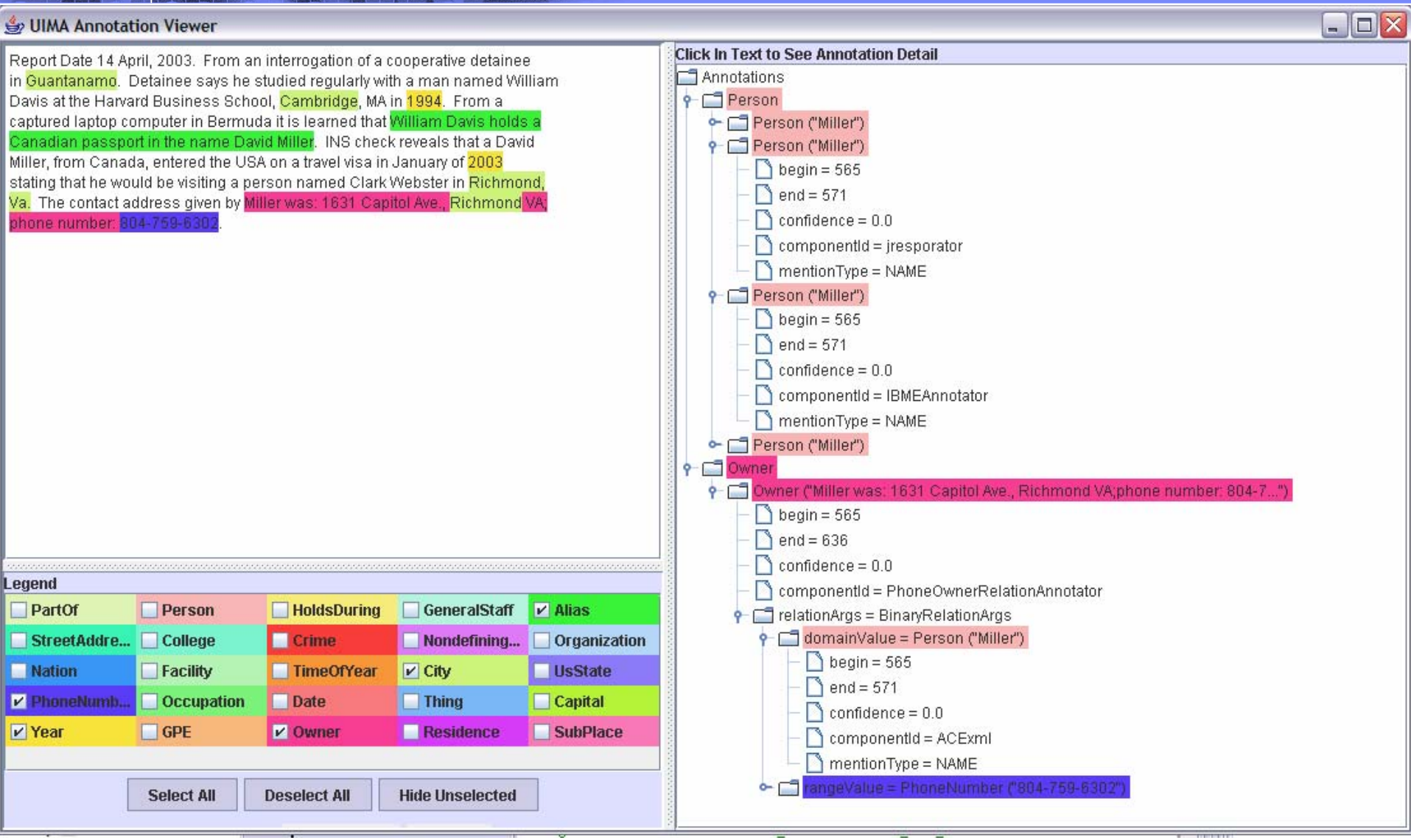
- ***Buried in Huge Volumes (Noise)***
- ***Implicit Semantics***
- ***Inefficient Search***

- ***Explicit Semantics***
- ***Efficient Search***
- ***Focused Content***

...BUT...

- ***Slow Growing***
- ***Narrow Coverage***
- ***Less Current/Relevant***

© 2006 IBM Corporation – All Rights Reserved –



© 2006 IBM Corporation – All Rights Reserved –

Semantic Analytics: The Promise and the Challenge

- Independently developed
- From an increasing # of sources

- Different technologies & interfaces
- Highly specialized & fine grained

Analysis Capabilities

- **Language, Speaker Identifiers**
- **Tokenizers**
- **Part of Speech Detectors**
- **Relationship Detectors**
- **Classifiers ...**

Capability Specializations

- **Modality**
- **Human Language**
- **Domain of Interest**
- **Performance/Precision Tradeoffs...**

The right analysis for the job will likely be a **best-of-breed combination** integrating across many dimensions.

© 2006 IBM Corporation – All Rights Reserved –

Over 200 IBM researchers across 6 WW labs with UIM as a Strategic Focus

© 2006 IBM Corporation – All Rights Reserved –

UIMA: Unstructured Information Management Architecture

The Architecture

- **Architecture for composing analytics that extract knowledge from unstructured sources & integrate results with structured information**
  - Interfaces, Data Representation Schemes, Design Patterns
- **Principal Architectural Commitments**
  - Common representation scheme
  - Common component engine interfaces (task and domain-independent)
  - Common component metadata
  - Pluggable Workflow and Transports
  - Embeddable (can be layered on top of system middleware)
- **Independent of but interoperable with application specific**
  - Data models
  - Algorithms
  - Language-level or domain-level concepts or tools
  - Workflows or workflow engines
  - Transports (e.g., SOAP, Vinci)
  - Back-end Systems (DB, Search Engine, KB Interfaces)

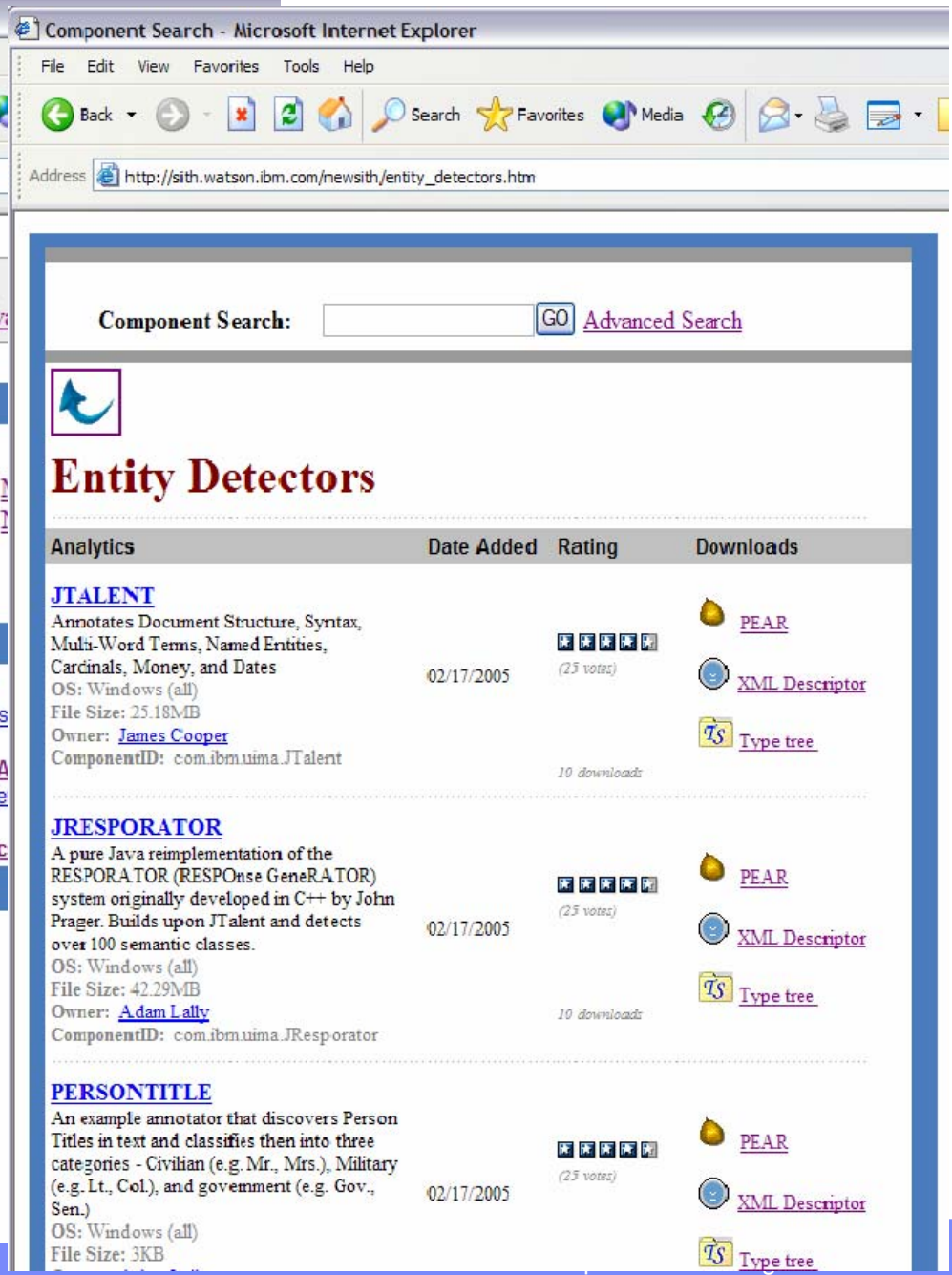
© 2006 IBM Corporation – All Rights Reserved –

UIMA: The Software

- **Supports UIMA-compliant development, composition & deployment**
- **Java Framework Implementation**
  - Components in other languages (e.g., C++, PERL, PYTHON)
  - Support for co-located and service-oriented deployments
    - Just change XML Descriptor
    - Balances performance and flexibility
    - Support for C++ co-Located Aggregates for high-performance communication
  - High-Performances APIs to common data representation (CAS)
- **Where to Get It**
  - **Software Development Kit (SDK) Freely Available from IBM alphaWorks**
    - http://www.ibm.com/research/uima
    - http://www.alphaworks.ibm.com/tech/uima
    - Stand-alone Java Install
    - C++ Enablement Layer
    - Tutorial and Development-Level Utilities and Tooling
    - “Semantic Search” Engine and Indexer
  - Core Software Framework **Open-Source on Source Forge**
    - http://uima-framework.sourceforge.net/
  - Emerging Component Community at CMU
    - http://uima.lti.cs.cmu.edu/index.html

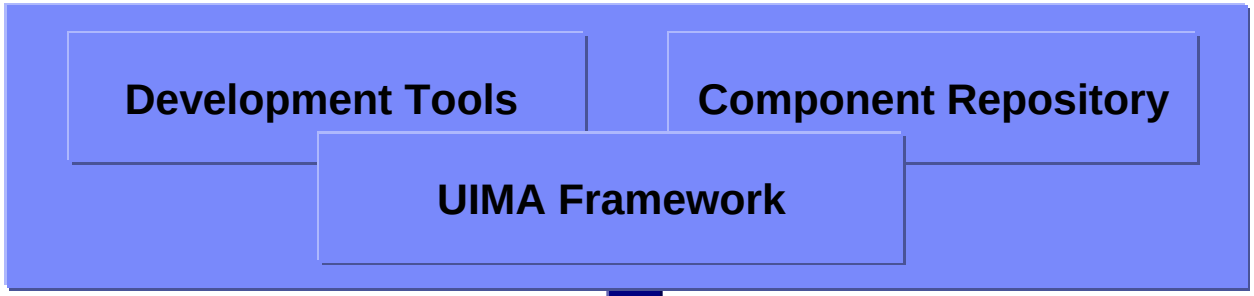
© 2006 IBM Corporation – All Rights Reserved –

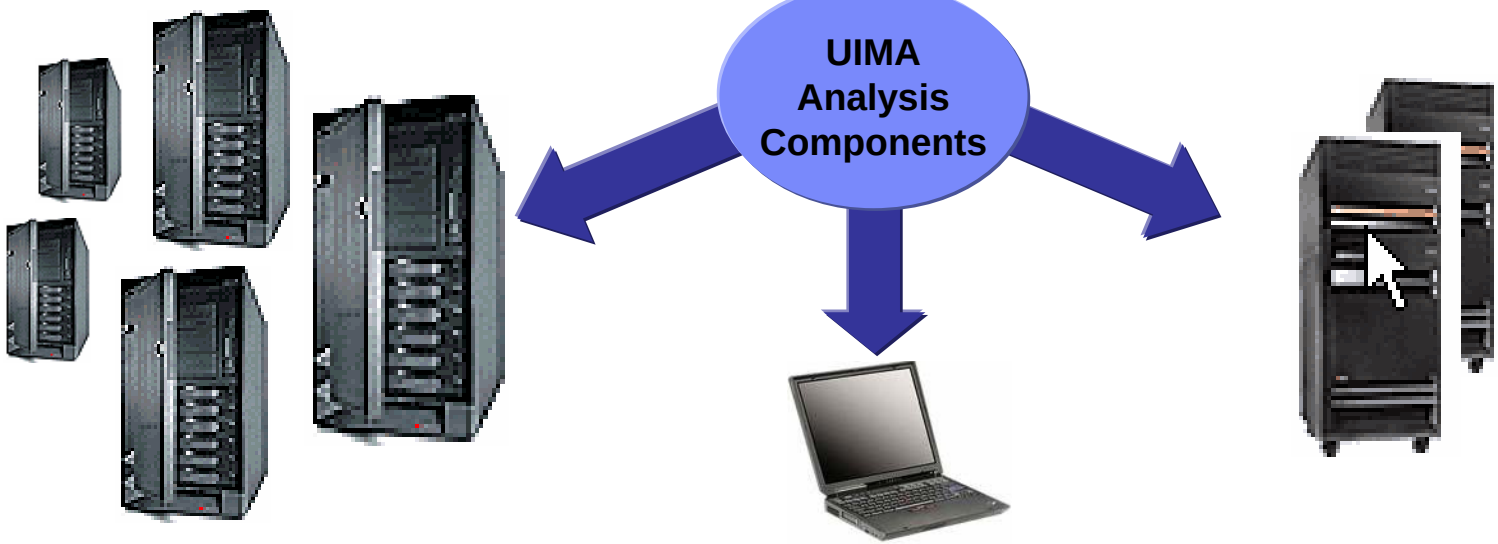
Search - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://sith.watson.ibm.com/newsith/default.htm
Component Search: GO Adv
UIMA Component Library
The available components include analysis engines, CAS processors (analysis engines, collection readers and CAS initializers) and CAS consumers (indexers, printers, etc.). They can be search by using search field above based on description or can be browsed based on the categories.
CAS Processors
Entity Detectors
Relation Detectors
Tokenizers, Parsers
CAS Consumers
Indexers, Filewriters
Domain Specific A
Life Science, Marke
OpenNLP Analytic
Collection Readers
XML Collection Reader
DB2 Collection Reader


Component Search - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://sith.watson.ibm.com/newsith/entity_detectors.htm
Component Search: GO Advanced Search
Entity Detectors
Analytics Date Added Rating Downloads
JTALENT
Annotates Document Structure, Syntax, Multi-Word Terms, Named Entities, Cardinals, Money, and Dates
OS: Windows (all)
File Size: 25.18MB
Owner: James Cooper
ComponentID: com.ibm.uima.JTalent
02/17/2005
(25 votes)
10 downloads
PEAR
XML Descriptor
Type tree
JRESPORATOR
A pure Java reimplementation of the RESPORATOR (RESPOnse GeneRATOR) system originally developed in C++ by John Prager. Builds upon JTalent and detects over 100 semantic classes.
OS: Windows (all)
File Size: 42.29MB
Owner: Adam Lally
ComponentID: com.ibm.uima.JResporator
02/17/2005
(25 votes)
10 downloads
PEAR
XML Descriptor
Type tree
PERSONTITLE
An example annotator that discovers Person Titles in text and classifies then into three categories - Civilian (e.g. Mr., Mrs.), Military (e.g. Lt., Col.), and government (e.g. Gov., Sen.)
OS: Windows (all)
File Size: 3KB
02/17/2005
(25 votes)
PEAR
XML Descriptor
Type tree

Common Infrastructure

Development, Composition, and Deployment





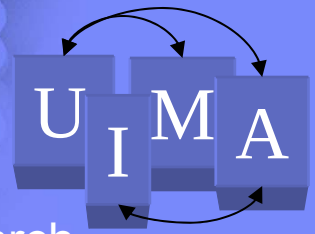
Large-Scale Distributed Deployment

Standalone

Product Embeddings

Across various types of deployments

© 2006 IBM Corporation – All Rights Reserved –



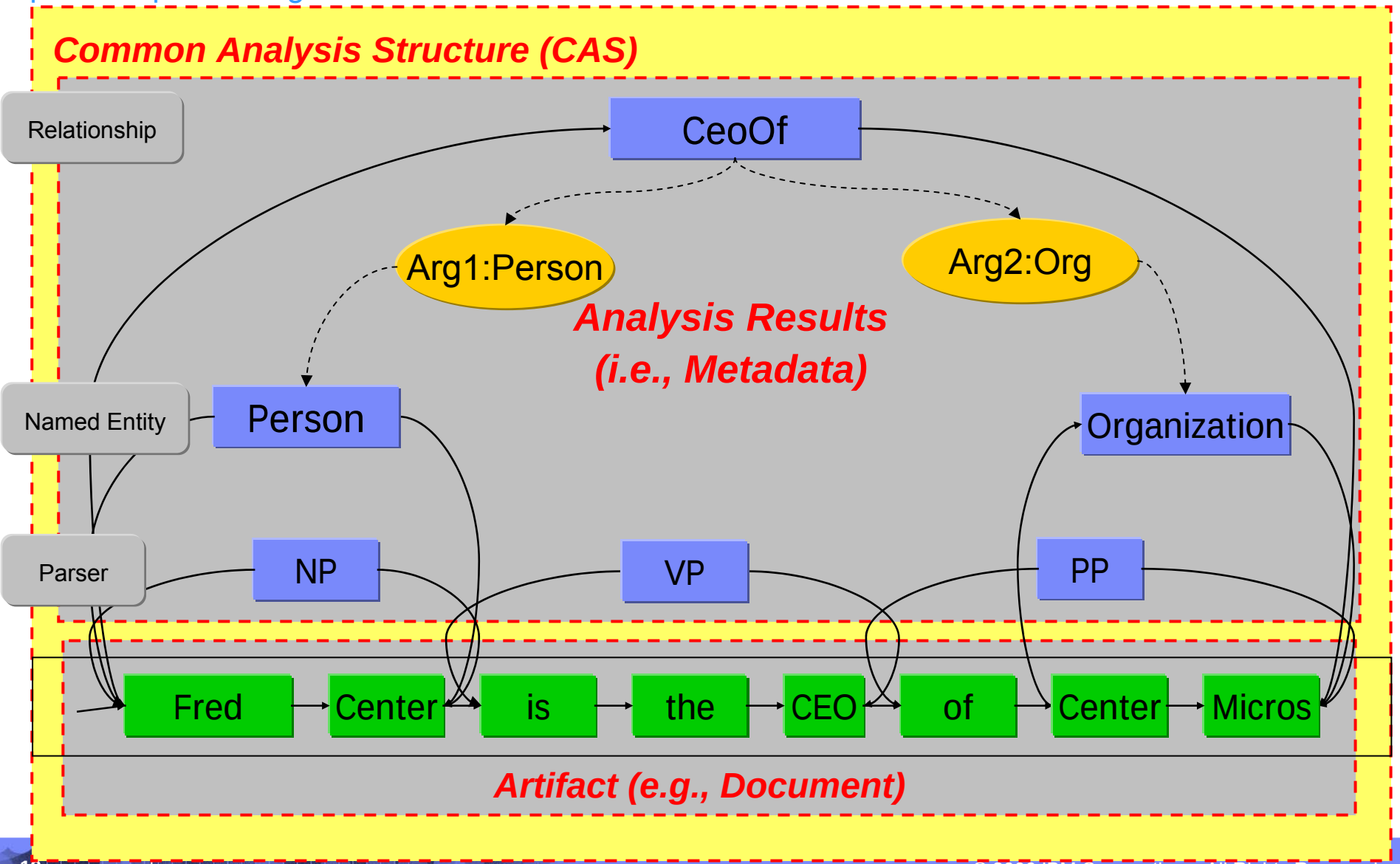
Architecture Highlights

Common Representations, Interfaces and Design Patterns

How it works

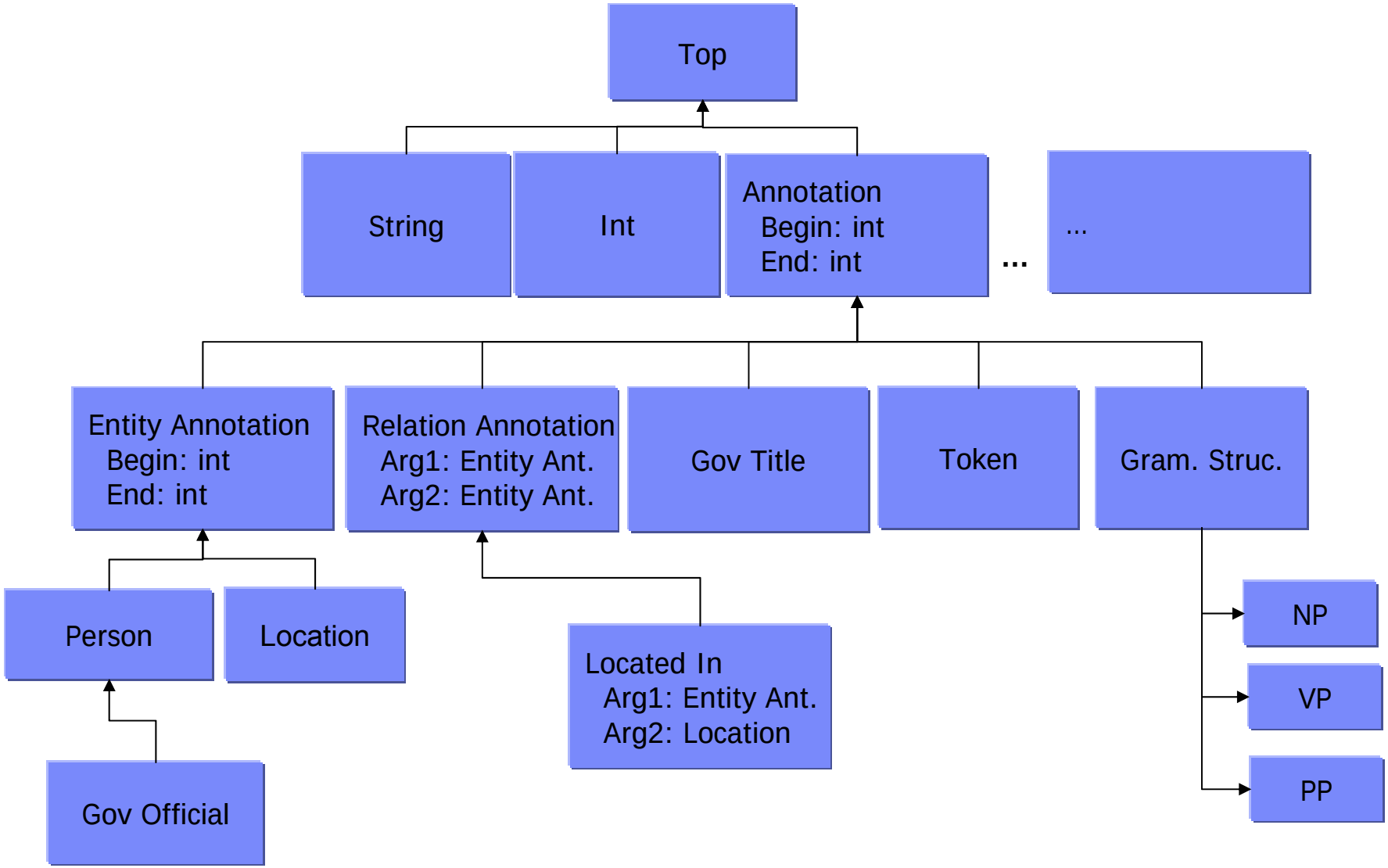
© 2006 IBM Corporation – All Rights Reserved

UIMA's Basic Building Blocks are **Annotators.** They iterate over an artifact to discover new types based on existing ones and update the **Common Analysis Structure (CAS)** for upstream processing.



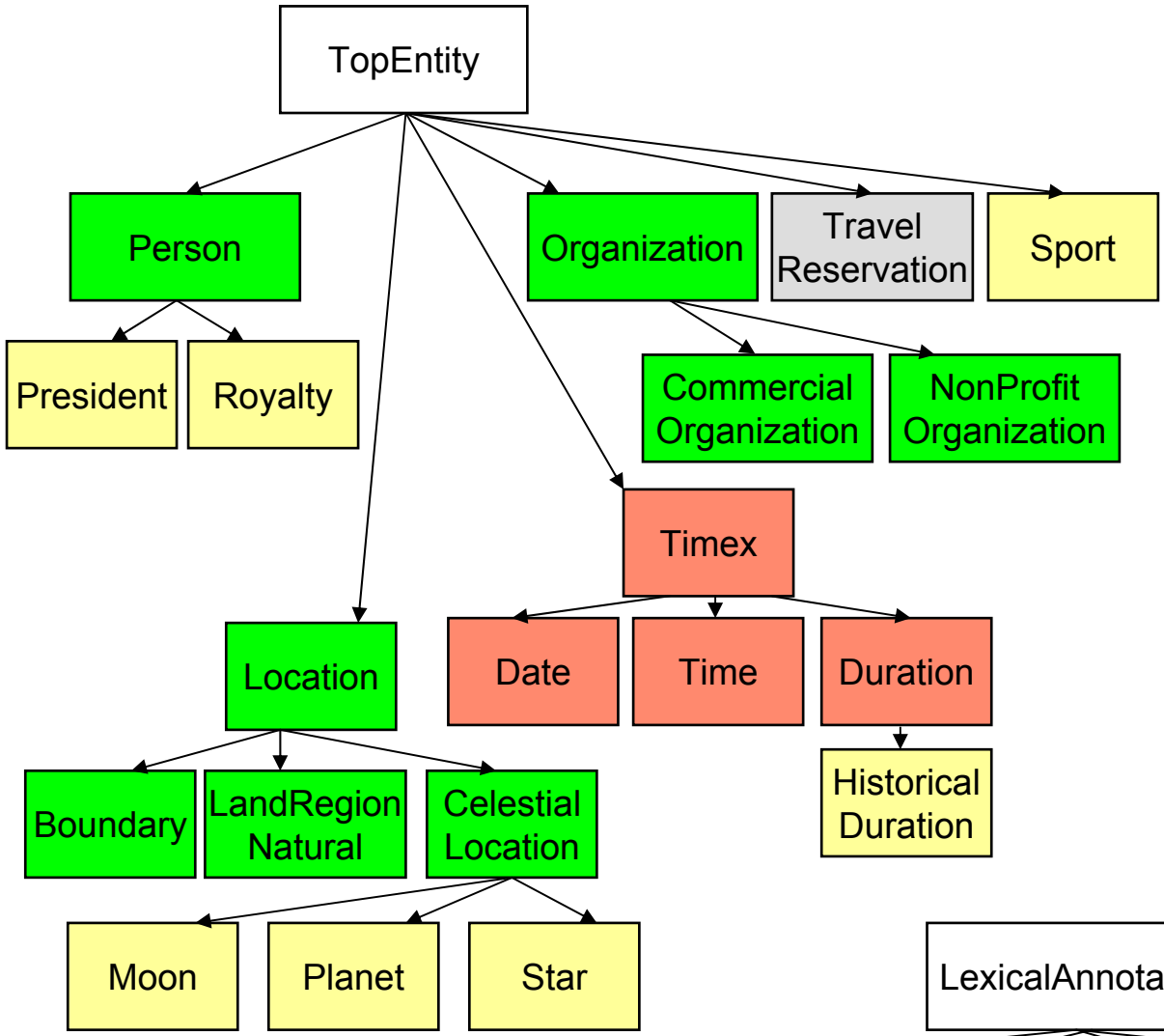
© 2006 IBM Corporation – All Rights Reserved –

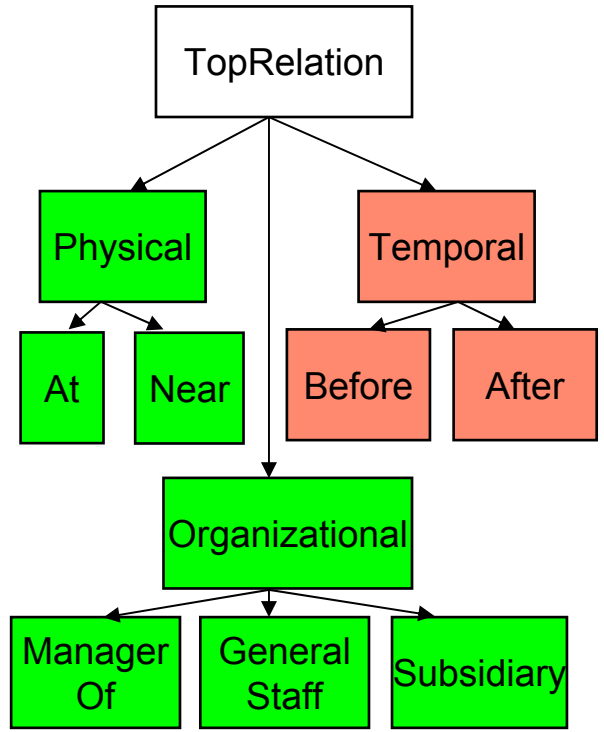
Sample Type System

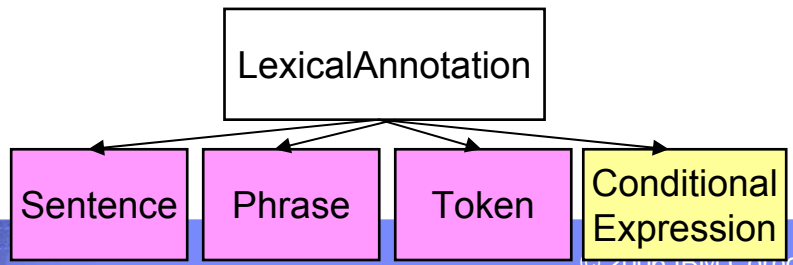


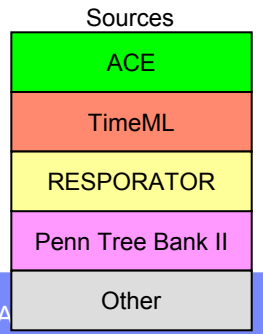
© 2006 IBM Corporation – All Rights Reserved –

Partial *HUTT* Type System (254 concepts in total)

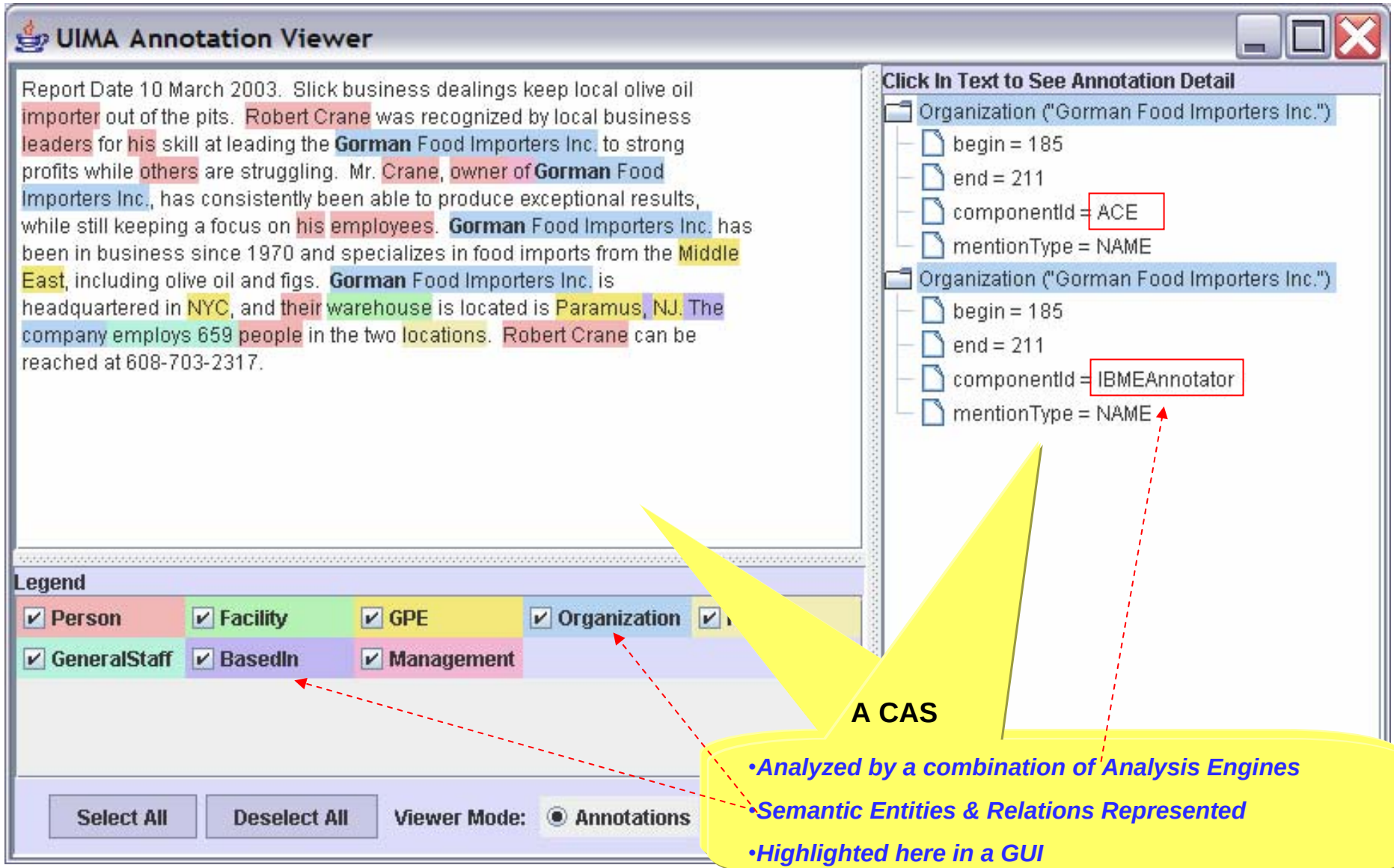






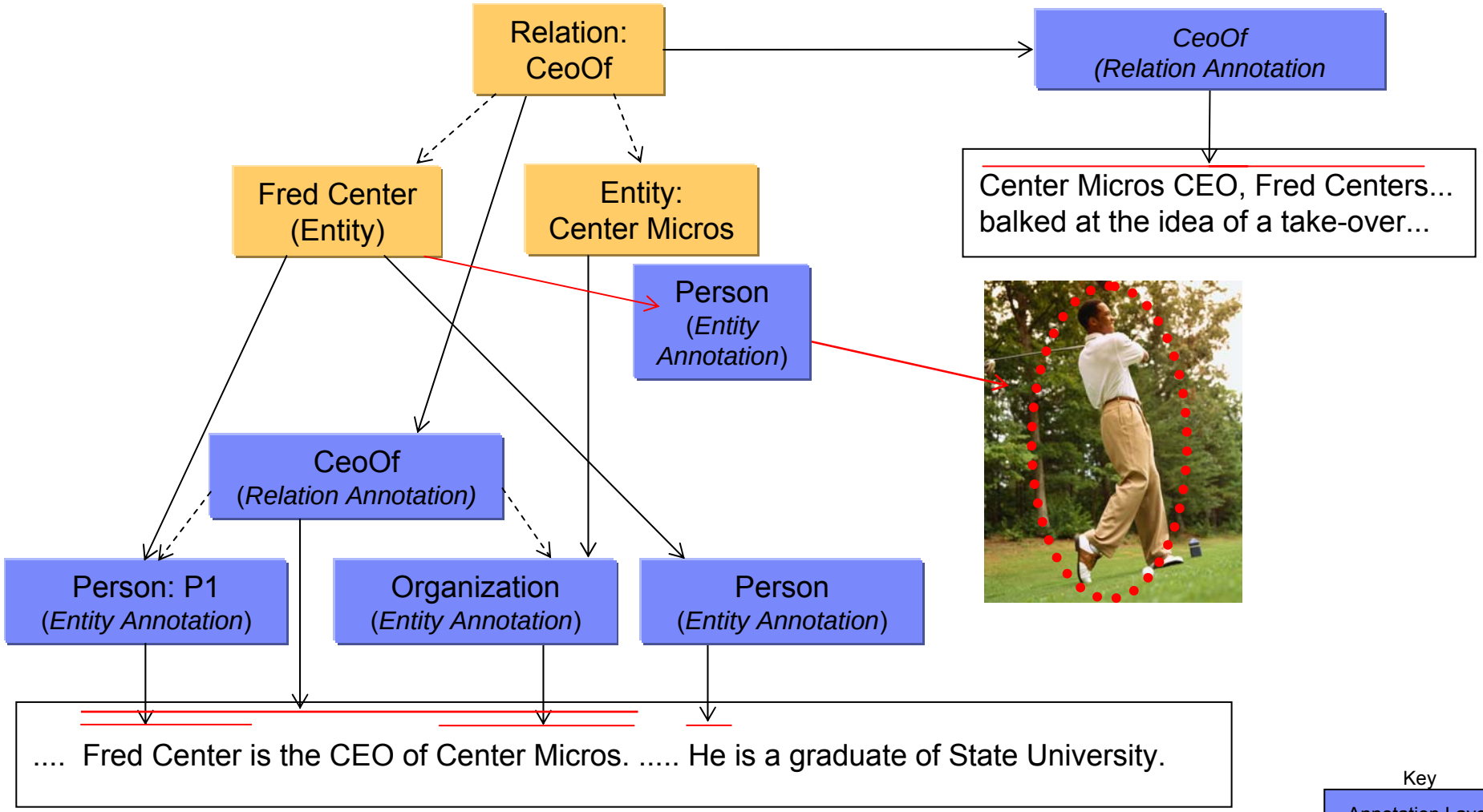


© 2006 IBM Corporation – A



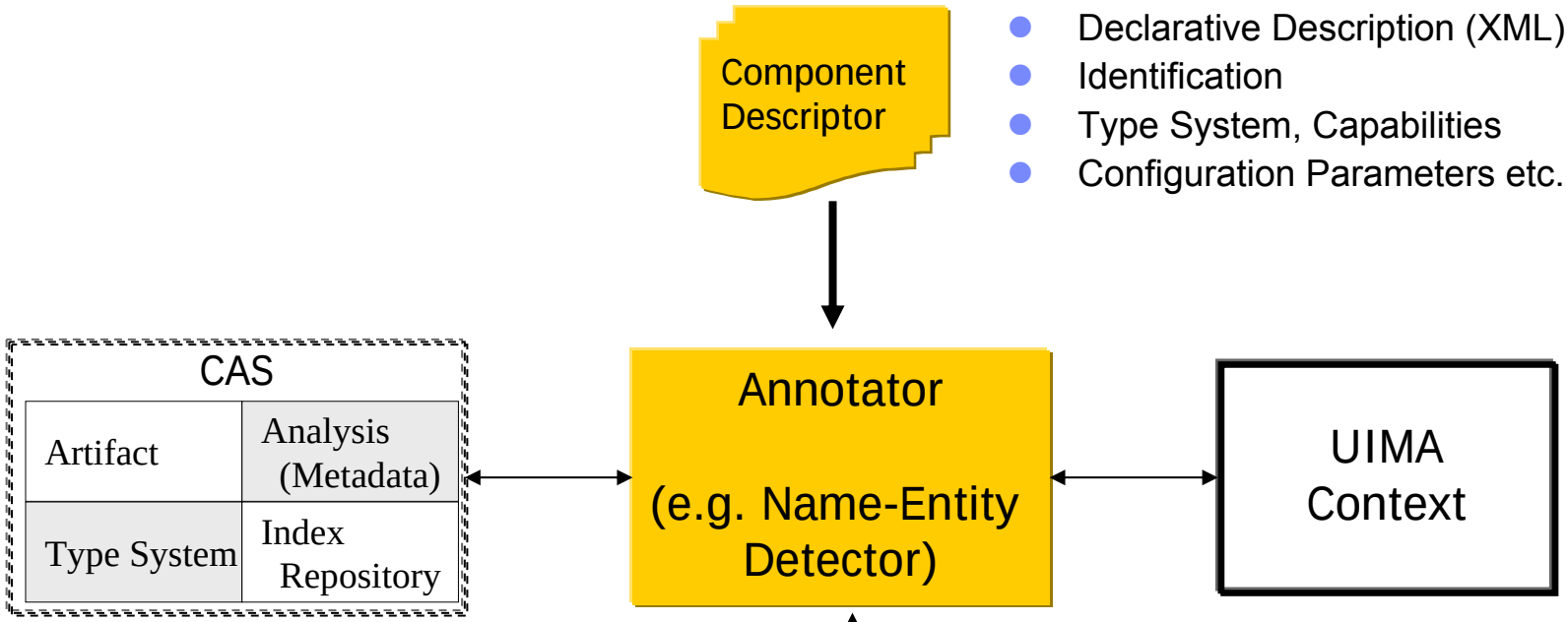
© 2006 IBM Corporation – All Rights Reserved –

Annotations and Referents



© 2006 IBM Corporation – All Rights Reserved –

The Basic UIMA Component Interfaces

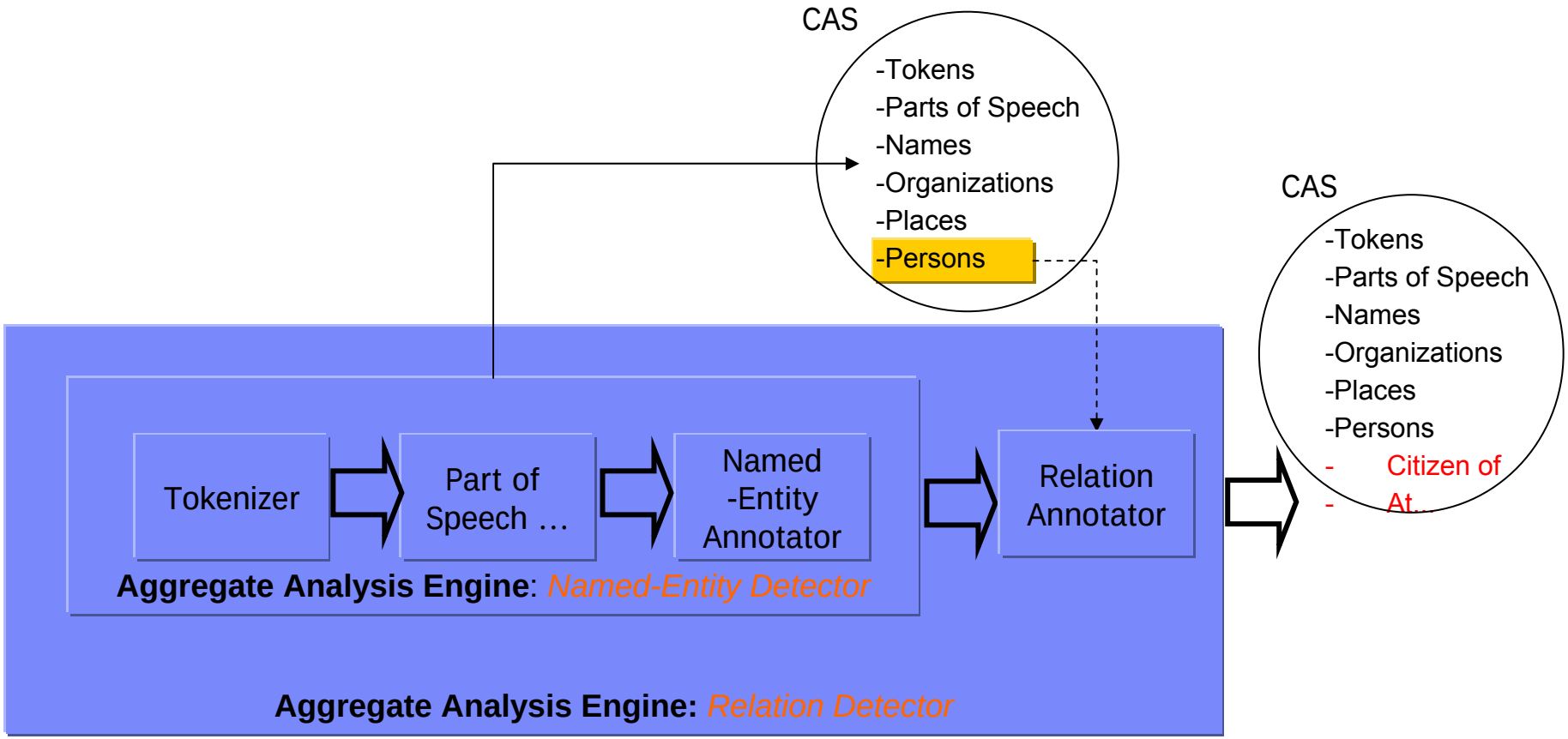


- Provides access to Artifact and the current analysis (metadata)
- JCAS Interface presents native Java model of CAS

- Manages access to shared external resources
- Allowing local APIs to access resources shared by cooperating engines

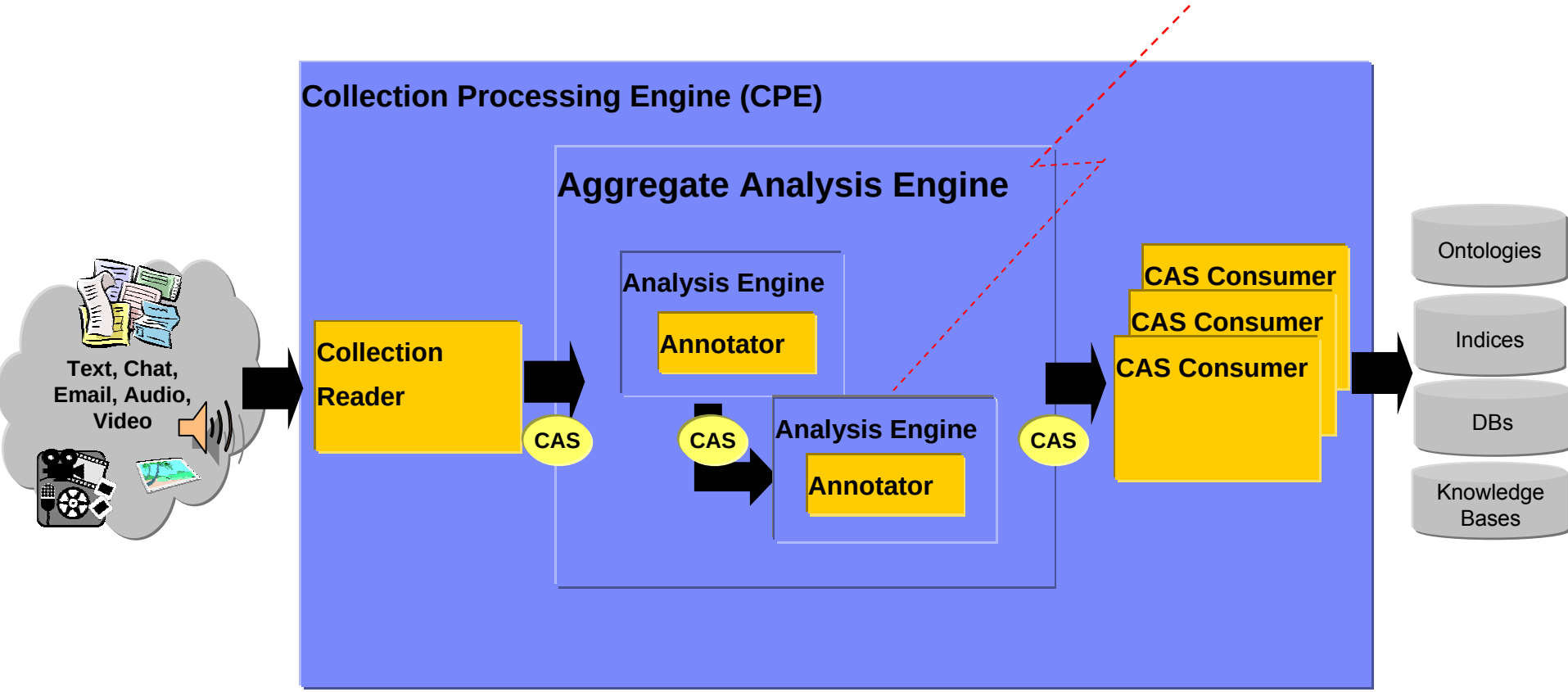
© 2006 IBM Corporation – All Rights Reserved –

The UIMA framework allows the developer to compose & encapsulate a set of engines as a single **aggregate** component, insulating applications or users from its implementation details/complexities – Promoting Reuse.



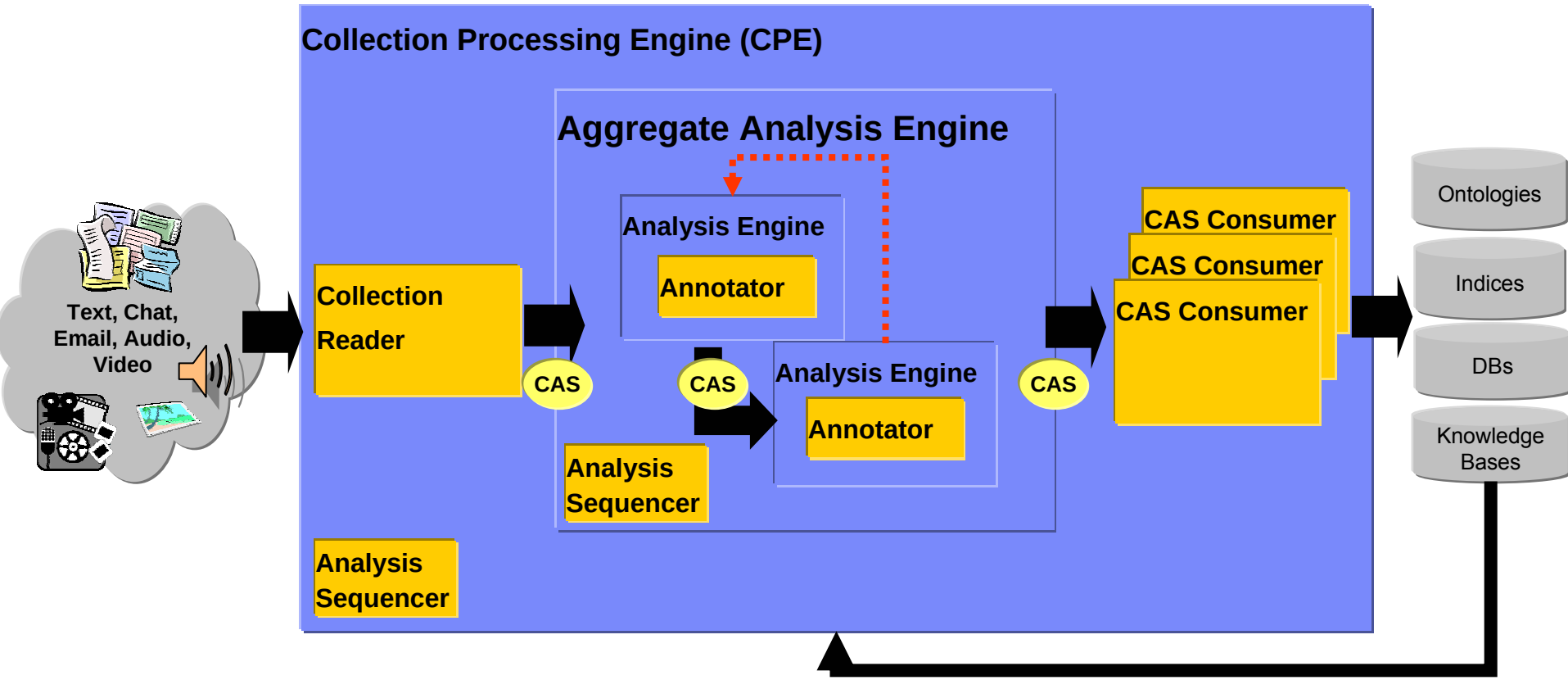
© 2006 IBM Corporation – All Rights Reserved –

Collection Processing Engines: Aggregate Analytic Components From "Source to Sink"



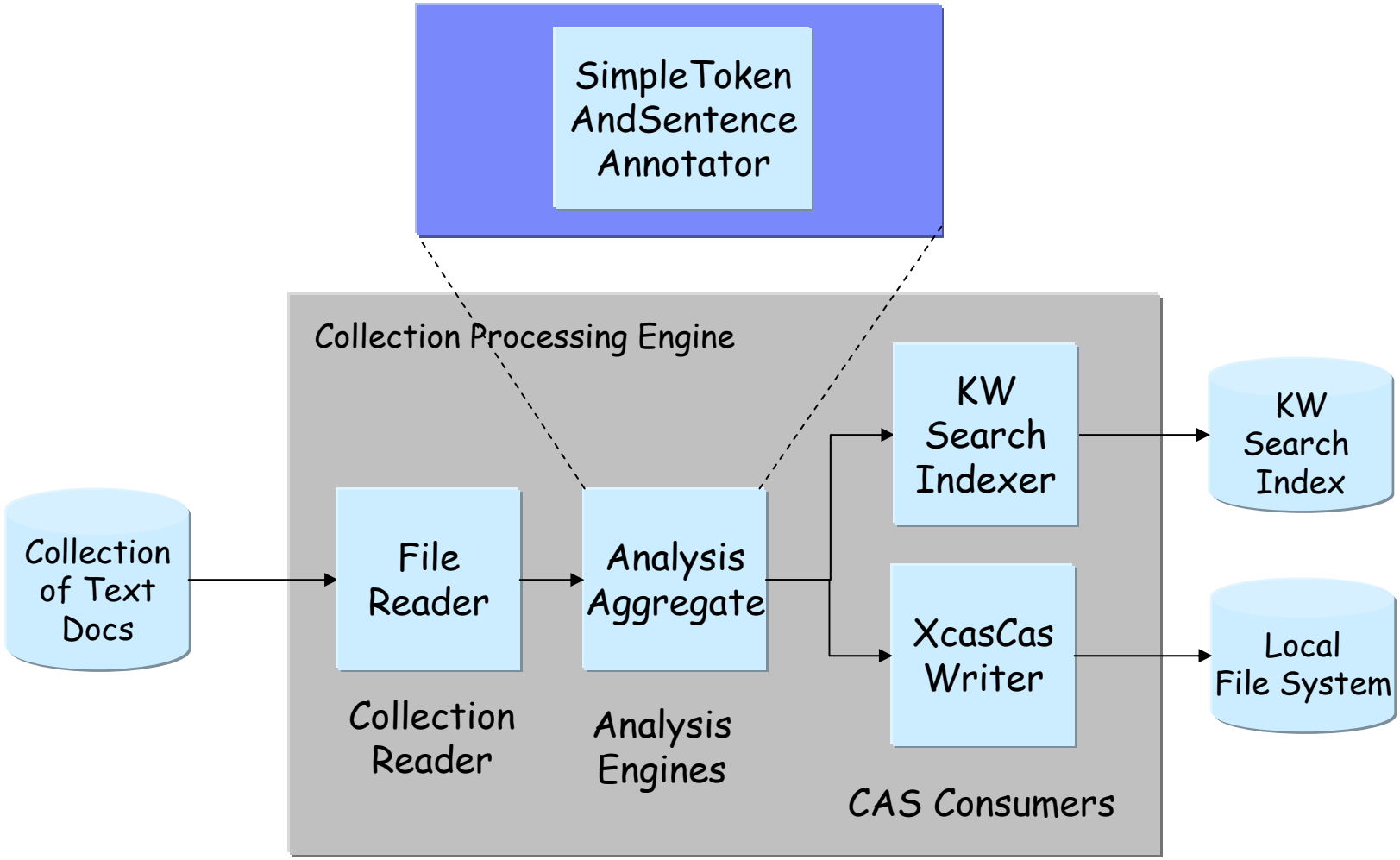
© 2006 IBM Corporation – All Rights Reserved –

User Definable Workflow (3Q 2006)



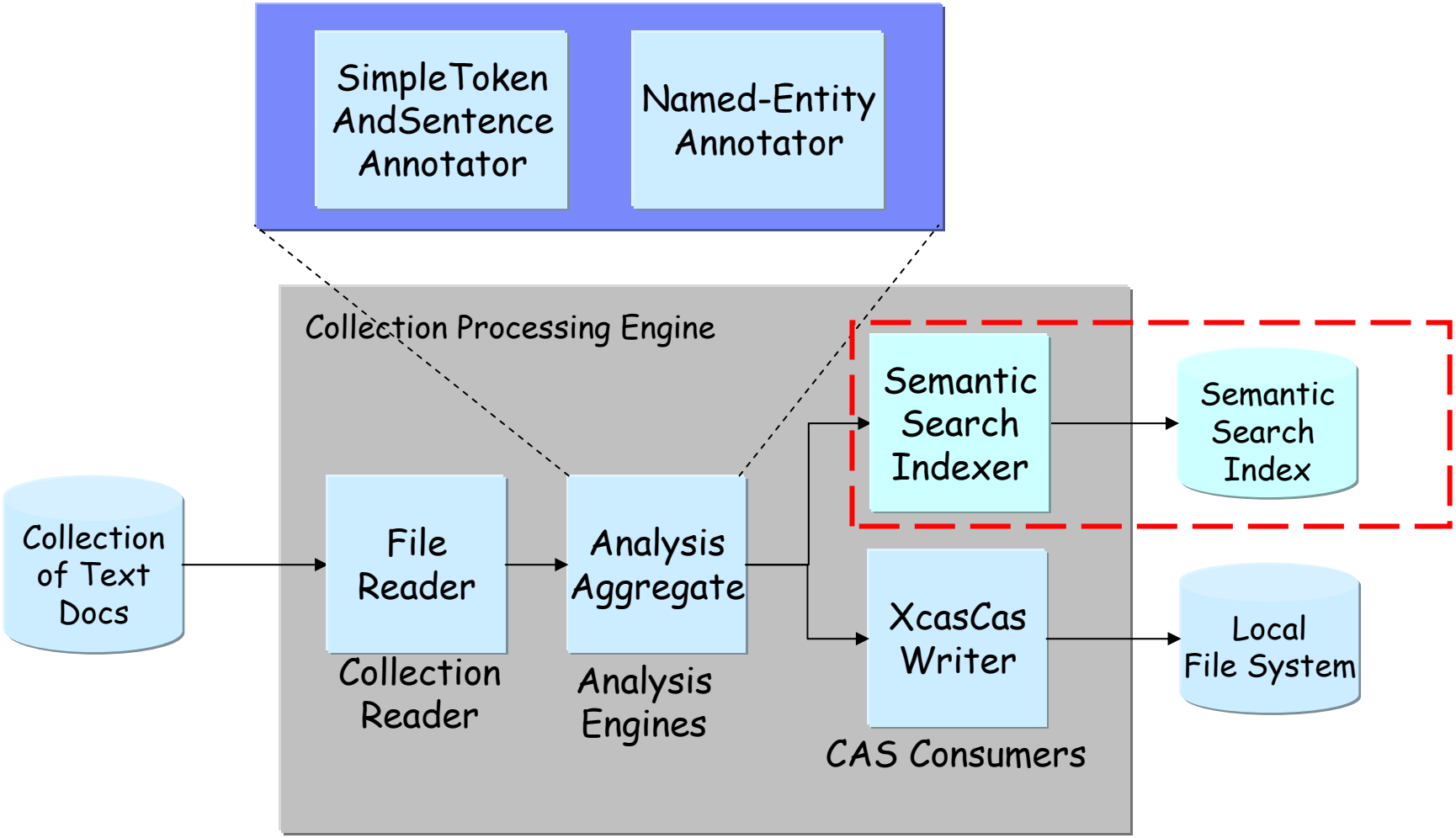
© 2006 IBM Corporation – All Rights Reserved –

CPE for Keyword Search



© 2006 IBM Corporation – All Rights Reserved –

CPE for Semantic Search



© 2006 IBM Corporation – All Rights Reserved –

We index and search over tokens AND the semantics annotations from UIMA

"first" is an ambiguous term so is "center"

We are looking for these terms with particular semantics possibley detected by the UIMA analytics

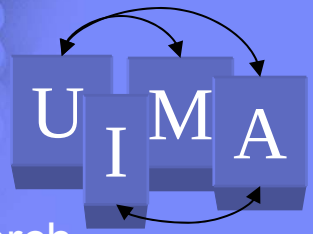
"first" as it appears in the name of an organization

The JuruXML Query Language Exploits the results of Analysis

KeyWord Query:	"first"
Semantic Search Query:	<organization> first </organization>
Semantic Search Query:	<ceo_of> <person> Center </person> </ceo_of>

"Center" appearing in a name of a person who is the ceo of something

© 2006 IBM Corporation – All Rights Reserved –



Raising the Search *Bar*

From matching keywords to gathering knowledge

© 2006 IBM Corporation – All Rights Reserved

From Finding Documents to Gathering Knowledge

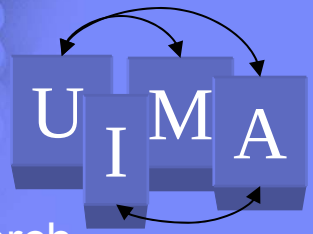
- **Known-Item Search**
  - Problem
    - Find the items (e.g., documents) that individually contain some combination of keywords
  - Operating Assumptions
    - User assumed that exactly what he/she wants is there in a document
    - There is one specific item out there that best solves information need
    - Native artifact boundaries are rigid
    - Contents = Used Words
    - User provides contained key words or phrases
  - Magic: Ranking
- **Knowledge Gathering and Synthesis**
  - Problem
    - Produce an *efficient* collection of knowledge that is relevant to information need
  - Operating Assumptions
    - Multiple items at varying grains together will solve information need
    - Artifact boundaries are permeable
    - Contents ~ Knowledge
    - User provides a conceptual space of interest (possibly using words or phrases)
  - Magic: Semantic Matching

© 2006 IBM Corporation – All Rights Reserved –

Raising the Search Bar

- **Exploit Semantic Analysis**
  - Tolerate less-than-perfect analytics
  - Use to Eliminate Noise and Focus Content
  - Degrade gracefully toward key-word search
- **The more in the query the better the results should be**
  - Web-Search conditions users to ~ 2 key words
  - Recall drops off dramatically
  - The more users type the less they know about the desired artifact & we punish them for it
  - Encourage the user to fully describe their information need in their terms
- **Predictive Query Interpretation and Refinement**
  - Should be clear on how to modify a query to direct results
  - Throttle Precision and Recall
- **Cross Artifact Boundaries**
  - The best solution is not always the single-based document
- **Identify the best Grain (relates to Density)**
  - Document, Passage, Entity, Fact, Set of…

© 2006 IBM Corporation – All Rights Reserved –



Finding the Sweet-Spot

Using an Ontology and an interactive dialog to elicit priorities and select or focus the content

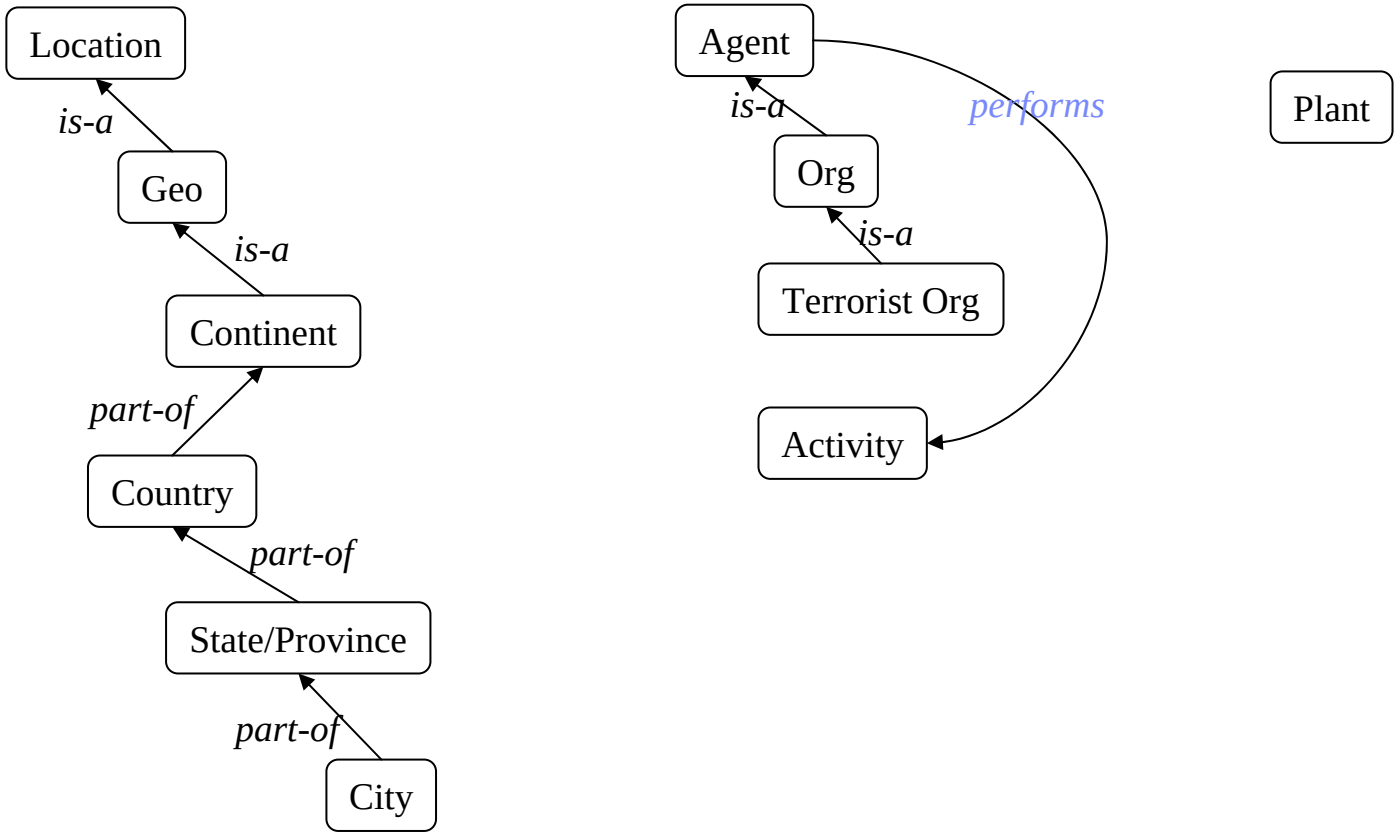
© 2006 IBM Corporation – All Rights Reserved

The QA-Cube

- **Produces conceptual dimensions present in a scenario**
  - Each represented by a concept taxonomy
- **Interacts with user to elicit relative priories**
  - Where the user needs to be most specific and
  - Where he is willing to relax the requirements
- **Evaluate or Generate Focused Corpora**
  - Less Noise, More Relevant
  - Worthy of Deeper Analysis/Exploration

© 2006 IBM Corporation – All Rights Reserved –

Example Base Ontology

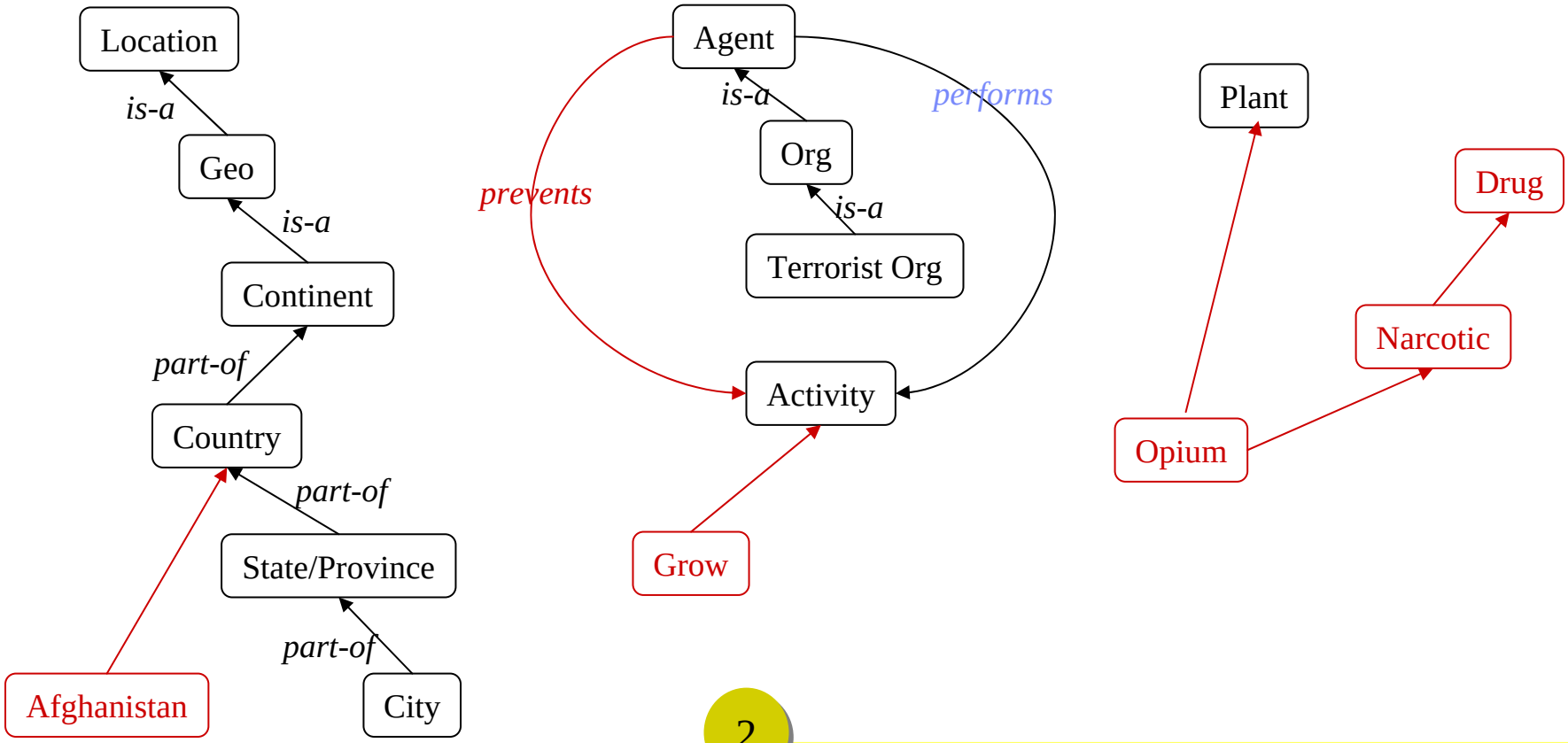


© 2006 IBM Corporation – All Rights Reserved –

NLP on scenario text (e.g., Topic, NE Detection)

Scenario Ontology Expansion

"…explore the prevention of opium growing in Afghanistan…"

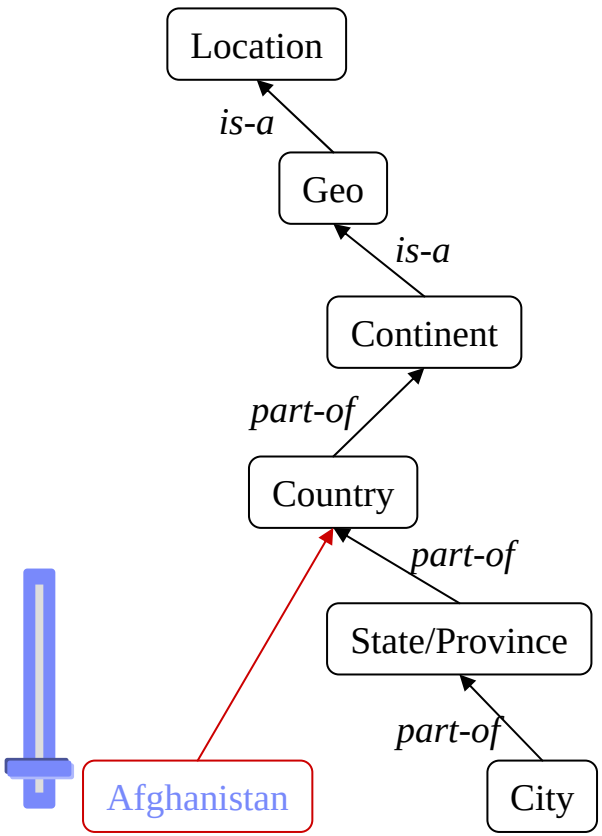


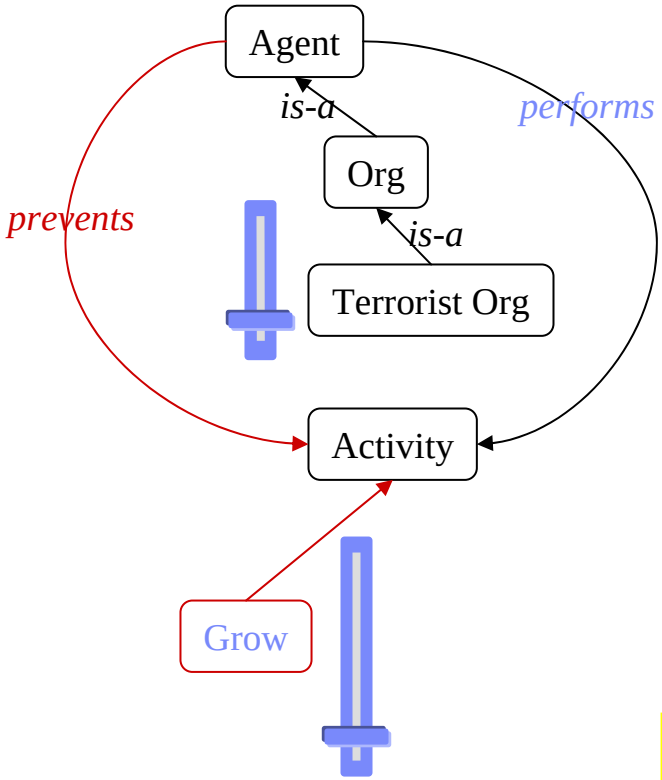
2

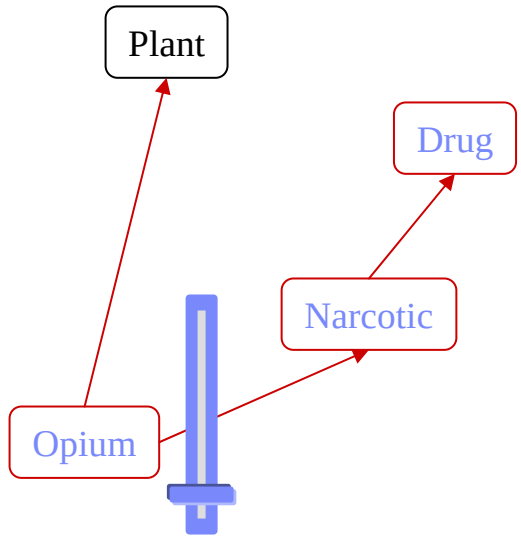
- Identify and query ontologies & KBs
- Find relevant concepts/relations
- Integrate with Scenario Ontology accordingly

© 2006 IBM Corporation – All Rights Reserved –

Selects key dimensions in QA Cube





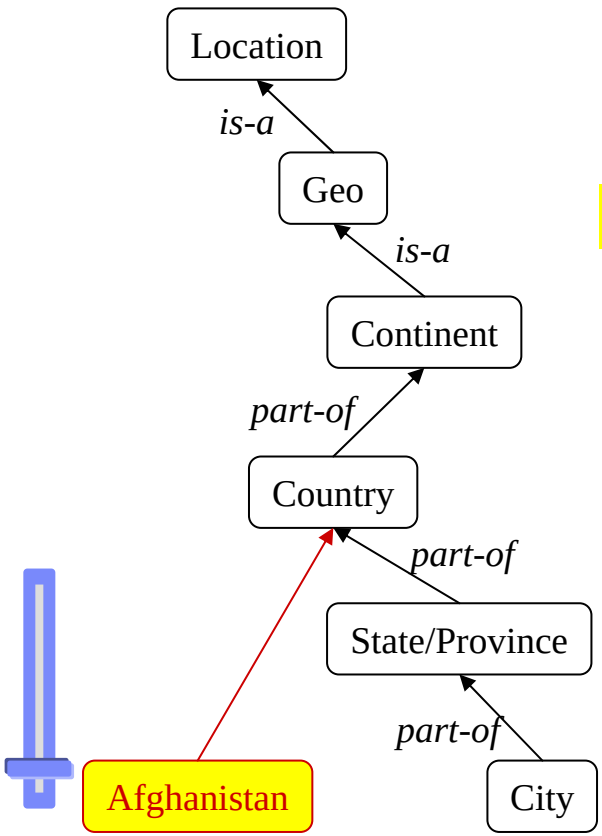


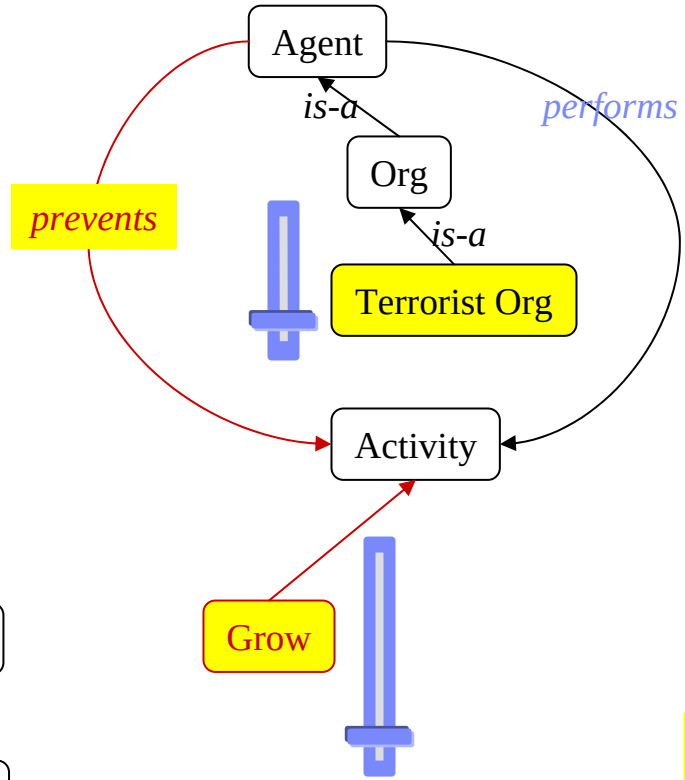
Hits: 0
Clusters: 0

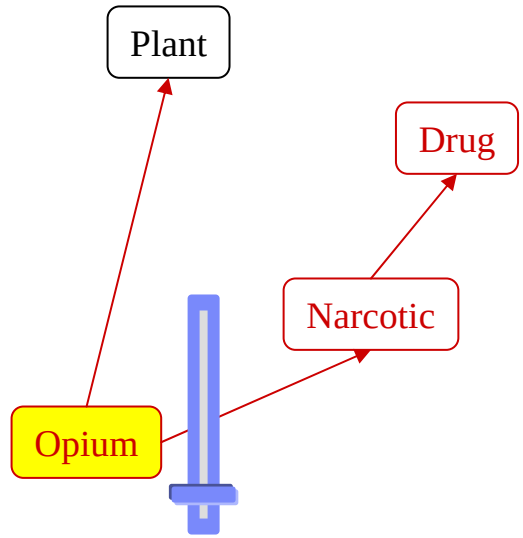
Immediate feedback on effectiveness of conceptual query

© 2006 IBM Corporation – All Rights Reserved –

Find the Sweet Spot



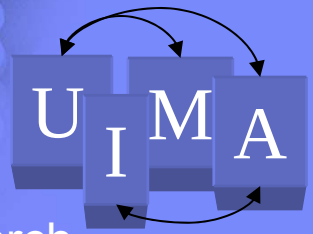




Hits: 300
Clusters: 10

Immediate feedback on effectiveness of conceptual query

© 2006 IBM Corporation – All Rights Reserved –



IBM

Semantic Search and Analysis Workbench

Exploiting Task Ontologies & Automatically Discovered Semantics to Improve Search Quality

© 2006 IBM Corporation – All Rights Reserved

Semantic Search

Broadly Defined…

An extension of search that exploits *semantic analysis* to improve *precision*, *recall*, and *density*.

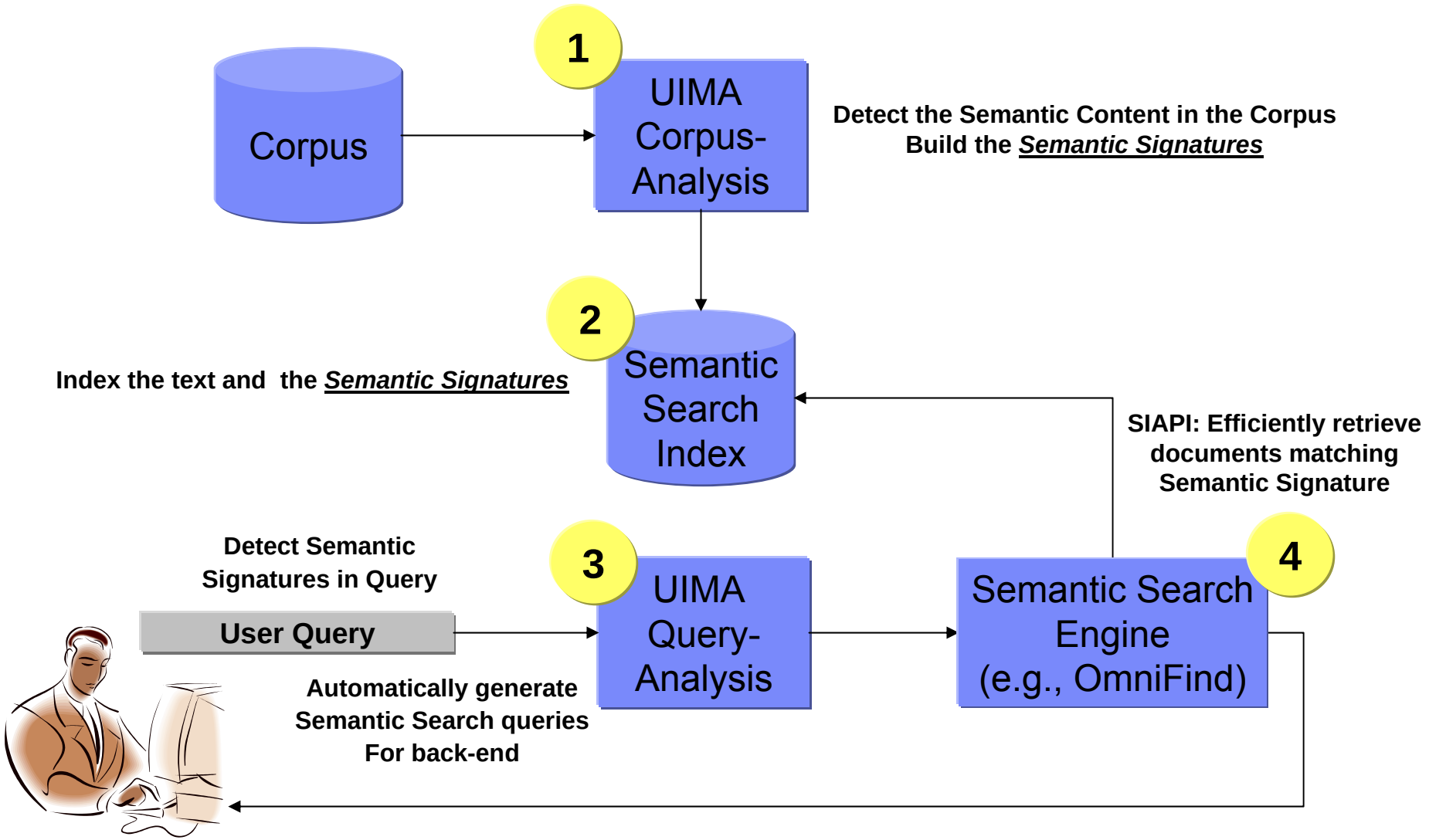
© 2006 IBM Corporation – All Rights Reserved –

The SAW: A Tool for Knowledge Gathering and Synthesis

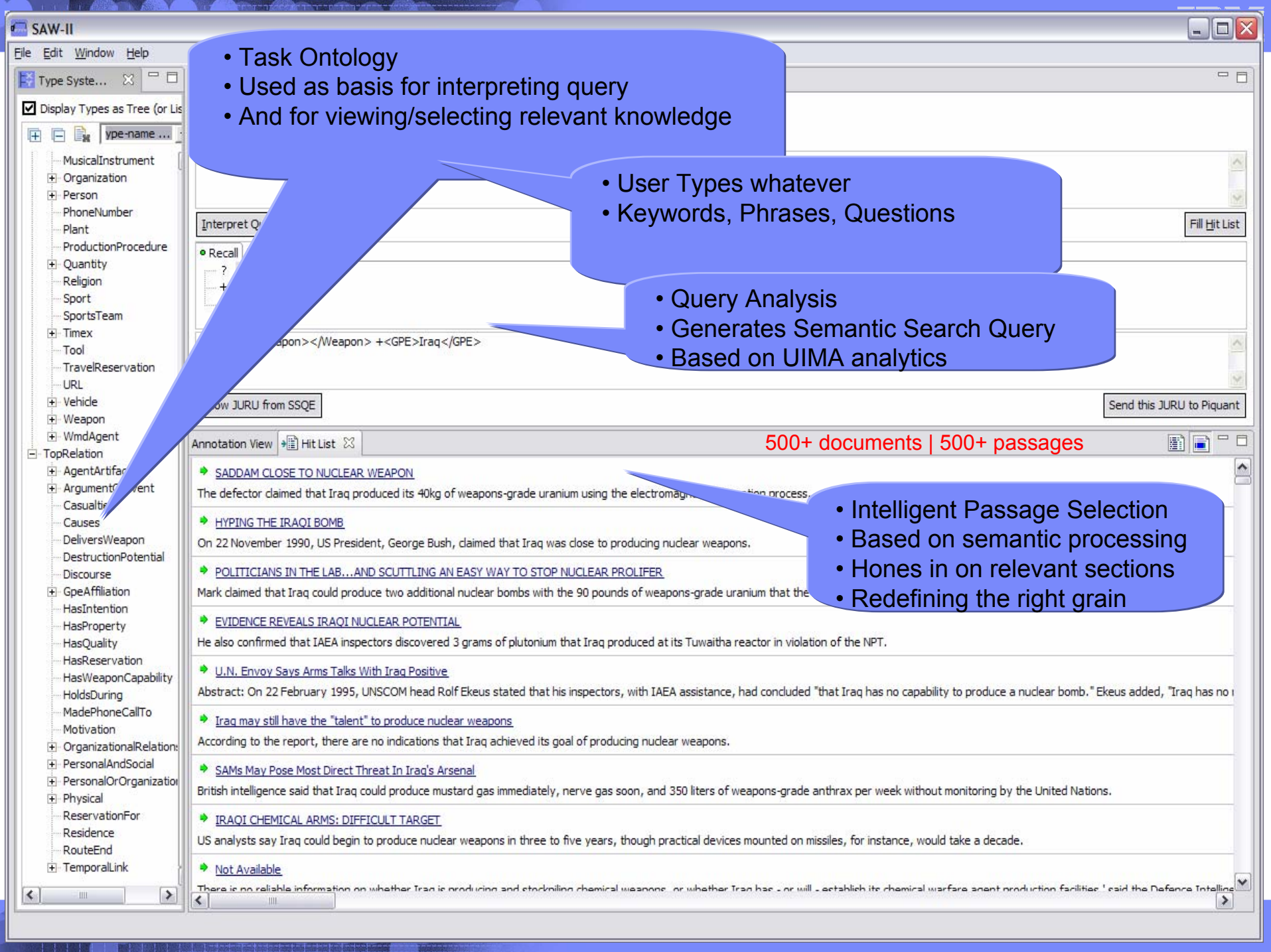
- **Beyond Keyword Search**
- **Ontologies**
- **Automatically Discovered Semantics**
- **Throttle Precision and Recall**
- **Toward Knowledge Gathering and Synthesis**

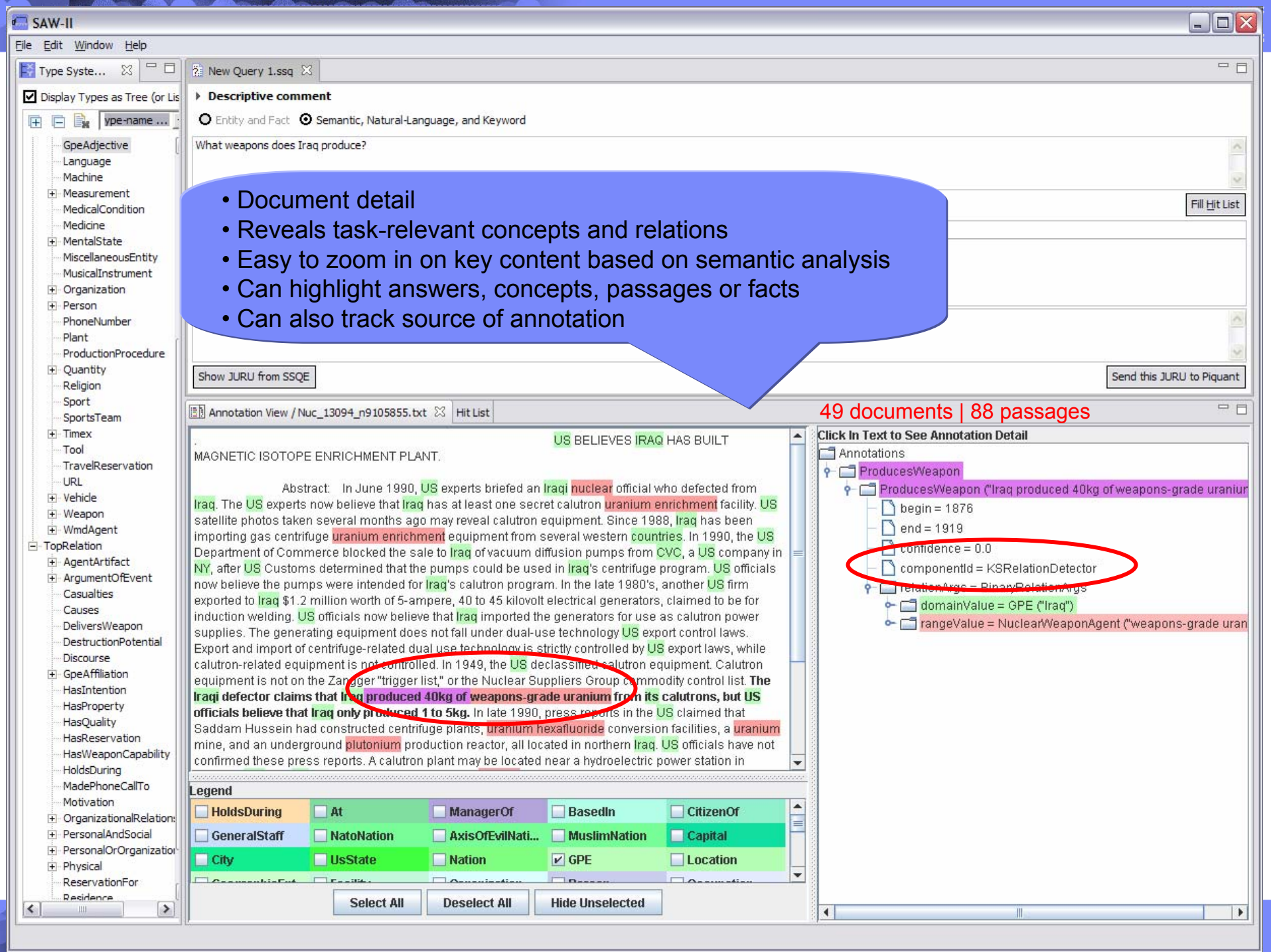
© 2006 IBM Corporation – All Rights Reserved –

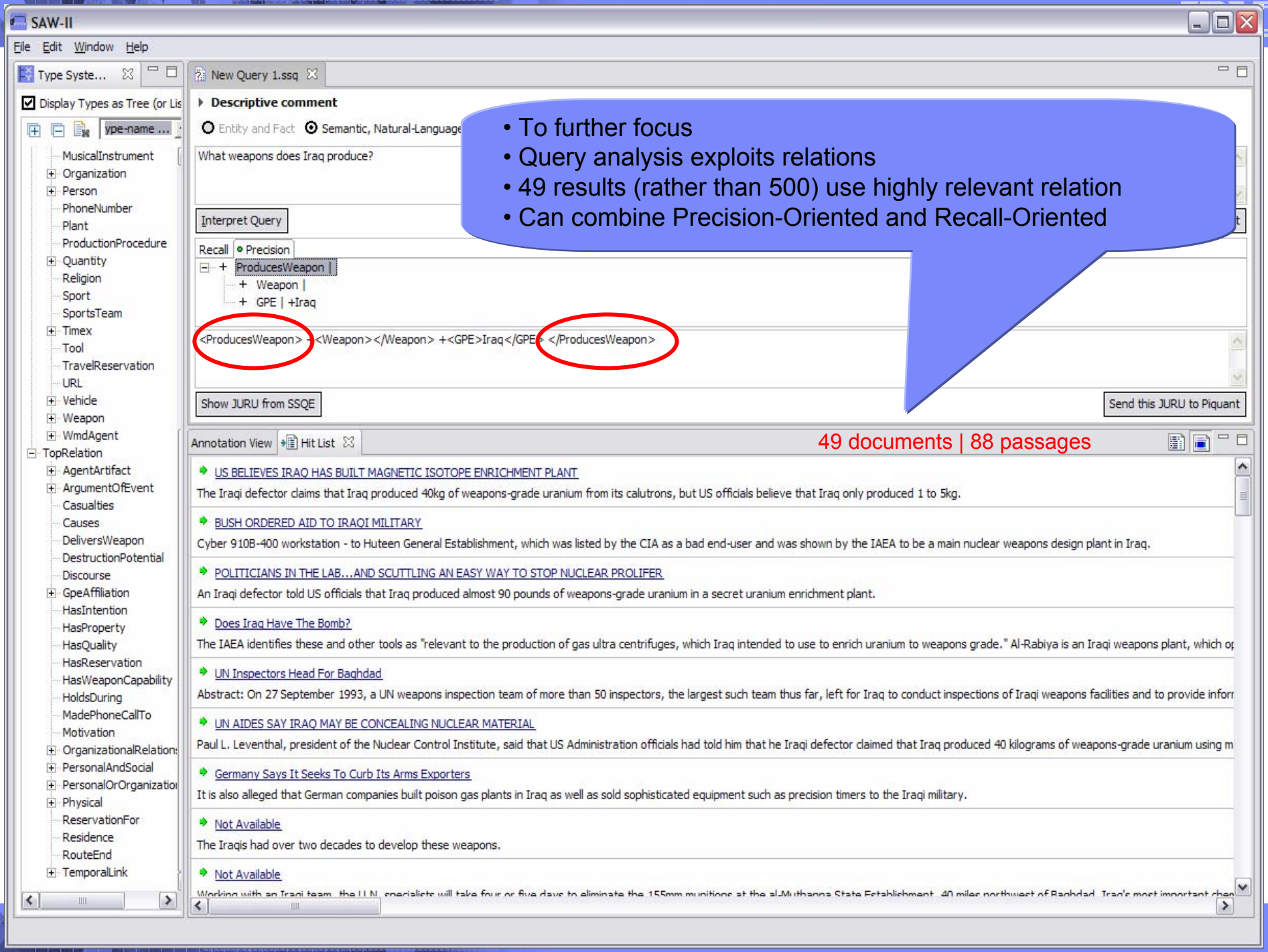
The Semantic Search 4-Step Program

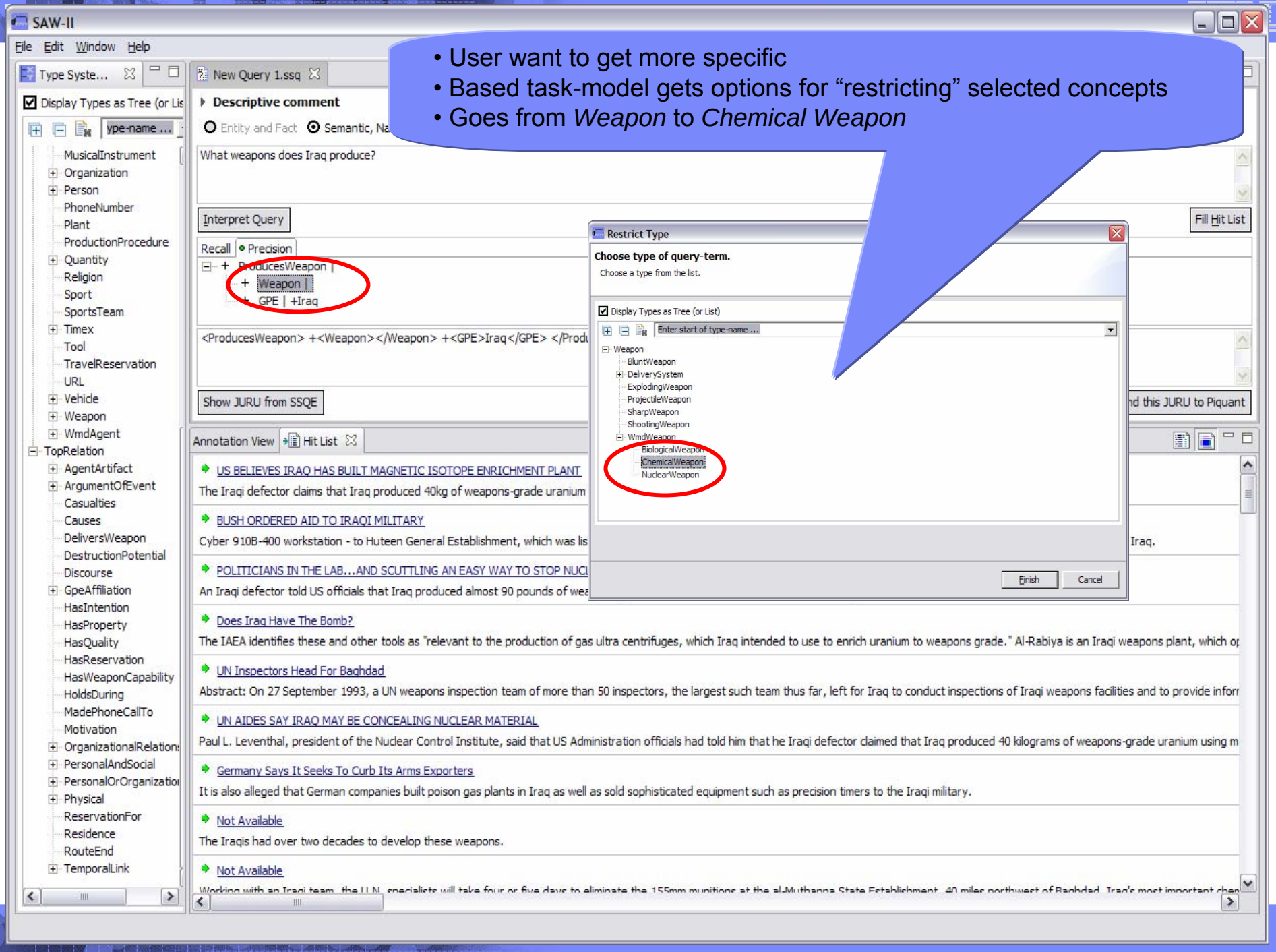


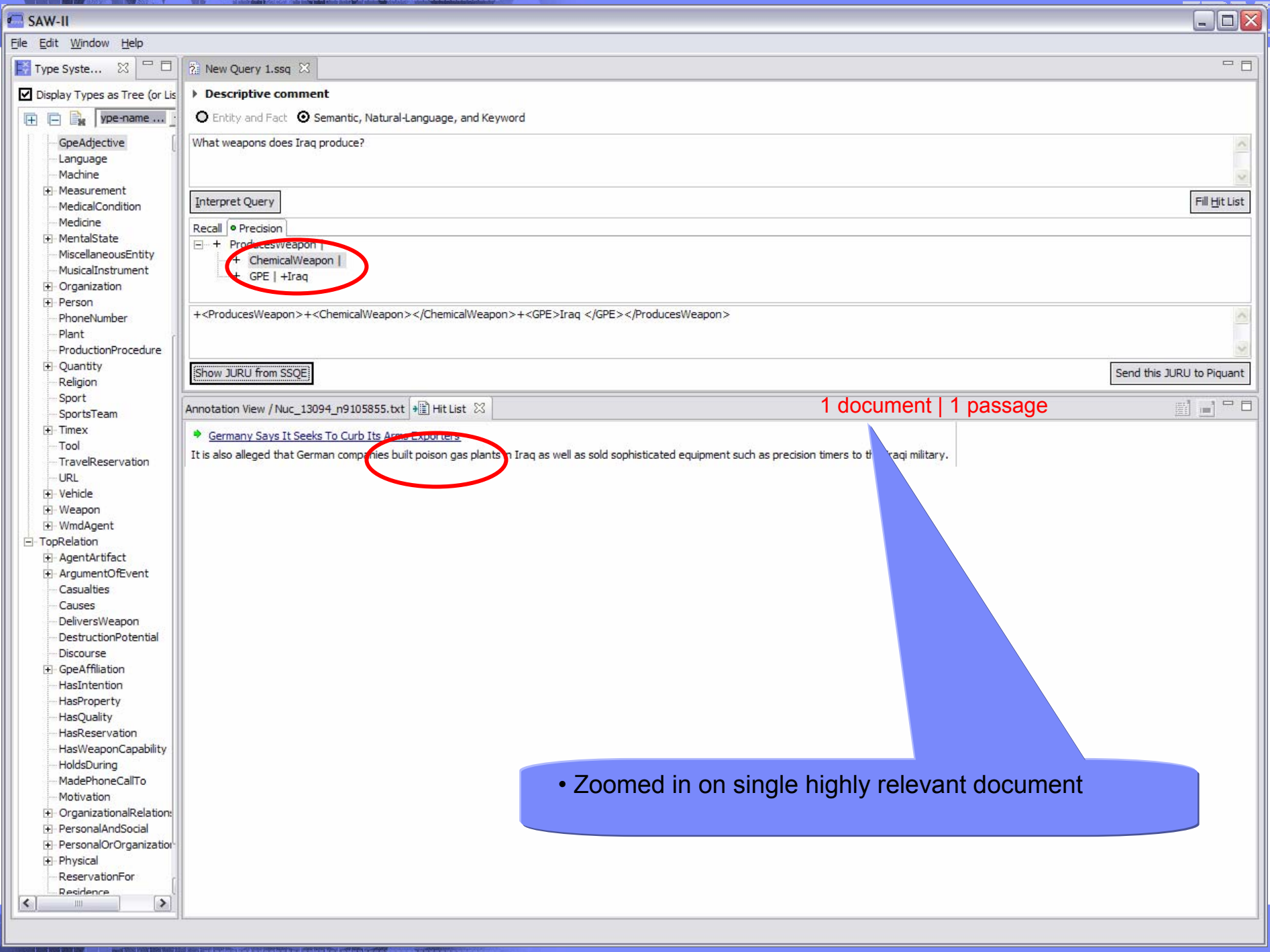
© 2006 IBM Corporation – All Rights Reserved –

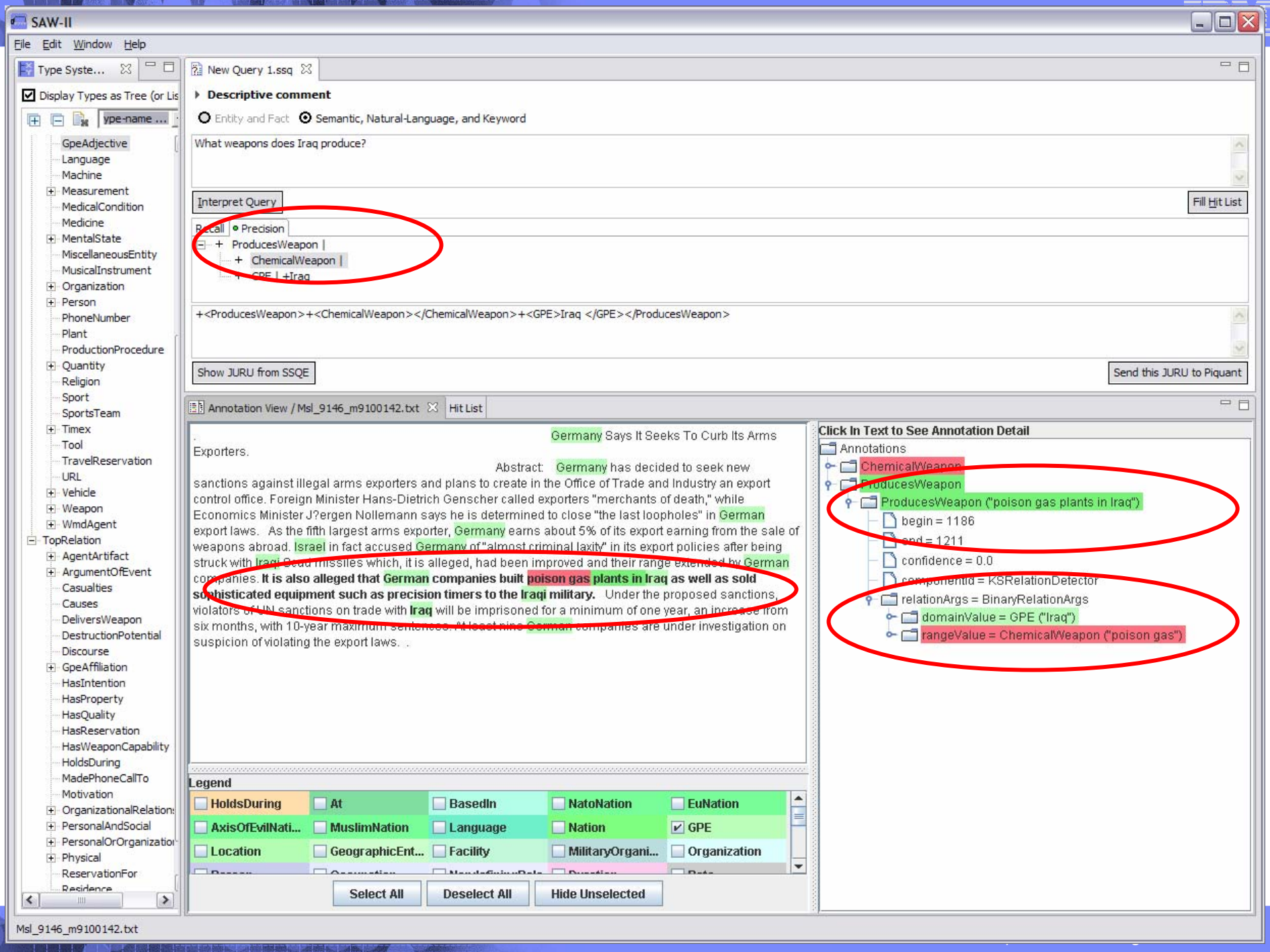

SAW-II
File Edit Window Help
Type Syste...
Display Types as Tree (or Lis
MusicalInstrument
Organization
Person
PhoneNumber
Plant
ProductionProcedure
Quantity
Religion
Sport
SportsTeam
Timex
Tool
TravelReservation
URL
Vehicle
Weapon
WmdAgent
TopRelation
Causes
DeliversWeapon
DestructionPotential
Discourse
GpeAffiliation
HasIntention
HasProperty
HasQuality
HasReservation
HasWeaponCapability
HoldsDuring
MadePhoneCallTo
Motivation
PersonalAndSocial
Physical
ReservationFor
Residence
RouteEnd
TemporalLink
Interpret Q
Fill Hit List
Recall
apon></Weapon> +<GPE>Iraq</GPE>
ow JURU from SSQE
Send this JURU to Piquant
Annotation View
Hit List
500+ documents | 500+ passages
SADDAM CLOSE TO NUCLEAR WEAPON
The defector claimed that Iraq produced its 40kg of weapons-grade uranium using the electromagn
HYPING THE IRAQI BOMB
On 22 November 1990, US President, George Bush, claimed that Iraq was close to producing nuclear weapons.
POLITICIANS IN THE LAB...AND SCUTTLING AN EASY WAY TO STOP NUCLEAR PROLIFER
Mark claimed that Iraq could produce two additional nuclear bombs with the 90 pounds of weapons-grade uranium that the
EVIDENCE REVEALS IRAQI NUCLEAR POTENTIAL
He also confirmed that IAEA inspectors discovered 3 grams of plutonium that Iraq produced at its Tuwaitha reactor in violation of the NPT.
U.N. Envoy Says Arms Talks With Iraq Positive
Abstract: On 22 February 1995, UNSCOM head Rolf Ekeus stated that his inspectors, with IAEA assistance, had concluded "that Iraq has no capability to produce a nuclear bomb." Ekeus added, "Iraq has no
Iraq may still have the "talent" to produce nuclear weapons
According to the report, there are no indications that Iraq achieved its goal of producing nuclear weapons.
SAMs May Pose Most Direct Threat In Iraq's Arsenal
British intelligence said that Iraq could produce mustard gas immediately, nerve gas soon, and 350 liters of weapons-grade anthrax per week without monitoring by the United Nations.
IRAQI CHEMICAL ARMS: DIFFICULT TARGET
US analysts say Iraq could begin to produce nuclear weapons in three to five years, though practical devices mounted on missiles, for instance, would take a decade.
Not Available
• Task Ontology
• Used as basis for interpreting query
• And for viewing/selecting relevant knowledge
• User Types whatever
• Keywords, Phrases, Questions
• Query Analysis
• Generates Semantic Search Query
• Based on UIMA analytics
• Intelligent Passage Selection
• Based on semantic processing
• Hones in on relevant sections
• Redefining the right grain

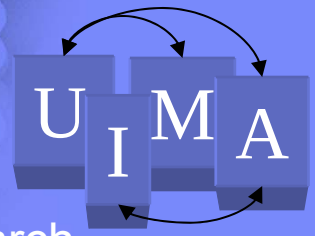

SAW-II
File Edit Window Help
Type Syste...
Display Types as Tree (or Lis
GpeAdjective
Language
Machine
Measurement
MedicalCondition
Medicine
MentalState
MiscellaneousEntity
MusicalInstrument
Organization
Person
PhoneNumber
Plant
ProductionProcedure
Quantity
Religion
Sport
SportsTeam
Timex
Tool
TravelReservation
URL
Vehicle
Weapon
WmdAgent
TopRelation
AgentArtifact
ArgumentOfEvent
Casualties
Causes
DeliversWeapon
DestructionPotential
Discourse
GpeAffiliation
HasIntention
HasProperty
HasQuality
HasReservation
HasWeaponCapability
HoldsDuring
MadePhoneCallTo
Motivation
OrganizationalRelation
PersonalAndSocial
PersonalOrOrganizatio
Physical
ReservationFor
New Query 1.ssq
Descriptive comment
Entity and Fact
Semantic, Natural-Language, and Keyword
What weapons does Iraq produce?
Fill Hit List
Show JURU from SSQE
Send this JURU to Piquant
• Document detail
• Reveals task-relevant concepts and relations
• Easy to zoom in on key content based on semantic analysis
• Can highlight answers, concepts, passages or facts
• Can also track source of annotation
Annotation View / Nuc_13094_n9105855.txt
Hit List
49 documents | 88 passages
US BELIEVES IRAQ HAS BUILT
MAGNETIC ISOTOPE ENRICHMENT PLANT.
Abstract: In June 1990, US experts briefed an Iraqi nuclear official who defected from Iraq. The US experts now believe that Iraq has at least one secret calutron uranium enrichment facility. US satellite photos taken several months ago may reveal calutron equipment. Since 1988, Iraq has been importing gas centrifuge uranium enrichment equipment from several western countries. In 1990, the US Department of Commerce blocked the sale to Iraq of vacuum diffusion pumps from CVC, a US company in NY, after US Customs determined that the pumps could be used in Iraq's centrifuge program. US officials now believe the pumps were intended for Iraq's calutron program. In the late 1980's, another US firm exported to Iraq $1.2 million worth of 5-ampere, 40 to 45 kilovolt electrical generators, claimed to be for induction welding. US officials now believe that Iraq imported the generators for use as calutron power supplies. The generating equipment does not fall under dual-use technology US export control laws. Export and import of centrifuge-related dual use technology is strictly controlled by US export laws, while calutron-related equipment is not controlled. In 1949, the US declassified calutron equipment. Calutron equipment is not on the Zangger "trigger list," or the Nuclear Suppliers Group commodity control list. The Iraqi defector claims that Iraq produced 40kg of weapons-grade uranium from its calutrons, but US officials believe that Iraq only produced 1 to 5kg. In late 1990, press reports in the US claimed that Saddam Hussein had constructed centrifuge plants, uranium hexafluoride conversion facilities, a uranium mine, and an underground plutonium production reactor, all located in northern Iraq. US officials have not confirmed these press reports. A calutron plant may be located near a hydroelectric power station in
Legend
HoldsDuring
At
ManagerOf
BasedIn
CitizenOf
GeneralStaff
NatoNation
AxisOfEvilNati...
MuslimNation
Capital
City
UsState
Nation
GPE
Location
Select All
Deselect All
Hide Unselected
Click In Text to See Annotation Detail
Annotations
ProducesWeapon
ProducesWeapon ("Iraq produced 40kg of weapons-grade uraniur
begin = 1876
end = 1919
confidence = 0.0
componentId = KSRelationDetector
domainValue = GPE ("Iraq")
rangeValue = NuclearWeaponAgent ("weapons-grade uran


SAW-II
File Edit Window Help
Type Syste...
Display Types as Tree (or Lis
MusicalInstrument
Organization
Person
PhoneNumber
Plant
ProductionProcedure
Quantity
Religion
Sport
SportsTeam
Timex
Tool
TravelReservation
URL
Vehicle
Weapon
WmdAgent
TopRelation
AgentArtifact
ArgumentOfEvent
Casualties
Causes
DeliversWeapon
DestructionPotential
Discourse
GpeAffiliation
HasIntention
HasProperty
HasQuality
HasReservation
HasWeaponCapability
HoldsDuring
MadePhoneCallTo
Motivation
OrganizationalRelation
PersonalAndSocial
PersonalOrOrganizatio
Physical
ReservationFor
Residence
RouteEnd
TemporalLink
New Query 1.ssq
Descriptive comment
Entity and Fact
Semantic, Natural-Language
What weapons does Iraq produce?
Interpret Query
Recall
Precision
ProducesWeapon |
Weapon |
GPE | +Iraq
<ProducesWeapon> +<Weapon></Weapon> +<GPE>Iraq</GPE> </ProducesWeapon>
Show JURU from SSQE
Send this JURU to Piquant
• To further focus
• Query analysis exploits relations
• 49 results (rather than 500) use highly relevant relation
• Can combine Precision-Oriented and Recall-Oriented
Annotation View
Hit List
49 documents | 88 passages
US BELIEVES IRAQ HAS BUILT MAGNETIC ISOTOPE ENRICHMENT PLANT
The Iraqi defector claims that Iraq produced 40kg of weapons-grade uranium from its calutrons, but US officials believe that Iraq only produced 1 to 5kg.
BUSH ORDERED AID TO IRAQI MILITARY
Cyber 910B-400 workstation - to Huteen General Establishment, which was listed by the CIA as a bad end-user and was shown by the IAEA to be a main nuclear weapons design plant in Iraq.
POLITICIANS IN THE LAB...AND SCUTTLING AN EASY WAY TO STOP NUCLEAR PROLIFER
An Iraqi defector told US officials that Iraq produced almost 90 pounds of weapons-grade uranium in a secret uranium enrichment plant.
Does Iraq Have The Bomb?
The IAEA identifies these and other tools as "relevant to the production of gas ultra centrifuges, which Iraq intended to use to enrich uranium to weapons grade." Al-Rabiya is an Iraqi weapons plant, which o
UN Inspectors Head For Baghdad
Abstract: On 27 September 1993, a UN weapons inspection team of more than 50 inspectors, the largest such team thus far, left for Iraq to conduct inspections of Iraqi weapons facilities and to provide infor
UN AIDES SAY IRAQ MAY BE CONCEALING NUCLEAR MATERIAL
Paul L. Leventhal, president of the Nuclear Control Institute, said that US Administration officials had told him that he Iraqi defector claimed that Iraq produced 40 kilograms of weapons-grade uranium using m
Germany Says It Seeks To Curb Its Arms Exporters
It is also alleged that German companies built poison gas plants in Iraq as well as sold sophisticated equipment such as precision timers to the Iraqi military.
Not Available
The Iraqis had over two decades to develop these weapons.
Not Available


SAW-II
File Edit Window Help
Type Syste...
Display Types as Tree (or Lis
MusicalInstrument
Organization
Person
PhoneNumber
Plant
ProductionProcedure
Quantity
Religion
Sport
SportsTeam
Timex
Tool
TravelReservation
URL
Vehicle
Weapon
WmdAgent
TopRelation
AgentArtifact
ArgumentOfEvent
Casualties
Causes
DeliversWeapon
DestructionPotential
Discourse
GpeAffiliation
HasIntention
HasProperty
HasQuality
HasReservation
HasWeaponCapability
HoldsDuring
MadePhoneCallTo
Motivation
OrganizationalRelation
PersonalAndSocial
PersonalOrOrganizatio
Physical
ReservationFor
Residence
RouteEnd
TemporalLink
New Query 1.ssq
Descriptive comment
Entity and Fact
Semantic, Na
What weapons does Iraq produce?
Interpret Query
Fill Hit List
Recall
Precision
ProducesWeapon
Weapon
GPE | +Iraq
<ProducesWeapon> +<Weapon></Weapon> +<GPE>Iraq</GPE> </Prod
Show JURU from SSQE
nd this JURU to Piquant
Annotation View
Hit List
US BELIEVES IRAQ HAS BUILT MAGNETIC ISOTOPE ENRICHMENT PLANT
The Iraqi defector claims that Iraq produced 40kg of weapons-grade uranium
BUSH ORDERED AID TO IRAQI MILITARY
Cyber 910B-400 workstation - to Huteen General Establishment, which was lis
Iraq.
POLITICIANS IN THE LAB...AND SCUTTLING AN EASY WAY TO STOP NUCL
An Iraqi defector told US officials that Iraq produced almost 90 pounds of wea
Does Iraq Have The Bomb?
The IAEA identifies these and other tools as "relevant to the production of gas ultra centrifuges, which Iraq intended to use to enrich uranium to weapons grade." Al-Rabiya is an Iraqi weapons plant, which o
UN Inspectors Head For Baghdad
Abstract: On 27 September 1993, a UN weapons inspection team of more than 50 inspectors, the largest such team thus far, left for Iraq to conduct inspections of Iraqi weapons facilities and to provide infor
UN AIDES SAY IRAQ MAY BE CONCEALING NUCLEAR MATERIAL
Paul L. Leventhal, president of the Nuclear Control Institute, said that US Administration officials had told him that he Iraqi defector claimed that Iraq produced 40 kilograms of weapons-grade uranium using m
Germany Says It Seeks To Curb Its Arms Exporters
It is also alleged that German companies built poison gas plants in Iraq as well as sold sophisticated equipment such as precision timers to the Iraqi military.
Not Available
The Iraqis had over two decades to develop these weapons.
Not Available
Restrict Type
Choose type of query-term.
Choose a type from the list.
Display Types as Tree (or List)
Enter start of type-name ...
Weapon
BluntWeapon
DeliverySystem
ExplodingWeapon
ProjectileWeapon
SharpWeapon
ShootingWeapon
WmdWeapon
BiologicalWeapon
ChemicalWeapon
NuclearWeapon
Finish
Cancel
• User want to get more specific
• Based task-model gets options for “restricting” selected concepts
• Goes from *Weapon* to *Chemical Weapon*


SAW-II
File Edit Window Help
Type Syste...
Display Types as Tree (or Lis
GpeAdjective
Language
Machine
Measurement
MedicalCondition
Medicine
MentalState
MiscellaneousEntity
MusicalInstrument
Organization
Person
PhoneNumber
Plant
ProductionProcedure
Quantity
Religion
Sport
SportsTeam
Timex
Tool
TravelReservation
URL
Vehicle
Weapon
WmdAgent
TopRelation
AgentArtifact
ArgumentOfEvent
Casualties
Causes
DeliversWeapon
DestructionPotential
Discourse
GpeAffiliation
HasIntention
HasProperty
HasQuality
HasReservation
HasWeaponCapability
HoldsDuring
MadePhoneCallTo
Motivation
OrganizationalRelation
PersonalAndSocial
PersonalOrOrganizatio
Physical
ReservationFor
New Query 1.ssq
Descriptive comment
Entity and Fact
Semantic, Natural-Language, and Keyword
What weapons does Iraq produce?
Interpret Query
Fill Hit List
Recall
Precision
ChemicalWeapon |
GPE | +Iraq
+<ProducesWeapon>+<ChemicalWeapon></ChemicalWeapon>+<GPE>Iraq </GPE></ProducesWeapon>
Show JURU from SSQE
Send this JURU to Piquant
Annotation View / Nuc_13094_n9105855.txt
Hit List
1 document | 1 passage
Germany Says It Seeks To Curb Its Arms Exporters
It is also alleged that German companies built poison gas plants in Iraq as well as sold sophisticated equipment such as precision timers to the Iraqi military.
• Zoomed in on single highly relevant document


SAW-II
File Edit Window Help
Type Syste...
Display Types as Tree (or Lis
GpeAdjective
Language
Machine
Measurement
MedicalCondition
Medicine
MentalState
MiscellaneousEntity
MusicalInstrument
Organization
Person
PhoneNumber
Plant
ProductionProcedure
Quantity
Religion
Sport
SportsTeam
Timex
Tool
TravelReservation
URL
Vehicle
Weapon
WmdAgent
TopRelation
AgentArtifact
ArgumentOfEvent
Casualties
Causes
DeliversWeapon
DestructionPotential
Discourse
GpeAffiliation
HasIntention
HasProperty
HasQuality
HasReservation
HasWeaponCapability
HoldsDuring
MadePhoneCallTo
Motivation
OrganizationalRelation
PersonalAndSocial
PersonalOrOrganizatio
Physical
ReservationFor
Residence
New Query 1.ssq
Descriptive comment
Entity and Fact
Semantic, Natural-Language, and Keyword
What weapons does Iraq produce?
Interpret Query
Fill Hit List
Recall
Precision
ProducesWeapon |
ChemicalWeapon |
GPE | +Iraq
+<ProducesWeapon>+<ChemicalWeapon></ChemicalWeapon>+<GPE>Iraq </GPE></ProducesWeapon>
Show JURU from SSQE
Send this JURU to Piquant
Annotation View / Msl_9146_m9100142.txt
Hit List
Germany Says It Seeks To Curb Its Arms Exporters.
Abstract: Germany has decided to seek new sanctions against illegal arms exporters and plans to create in the Office of Trade and Industry an export control office. Foreign Minister Hans-Dietrich Genscher called exporters "merchants of death," while Economics Minister J?ergen Nollemann says he is determined to close "the last loopholes" in German export laws. As the fifth largest arms exporter, Germany earns about 5% of its export earning from the sale of weapons abroad. Israel in fact accused Germany of "almost criminal laxity" in its export policies after being struck with Iraqi Scud missiles which, it is alleged, had been improved and their range extended by German companies. It is also alleged that German companies built poison gas plants in Iraq as well as sold sophisticated equipment such as precision timers to the Iraqi military. Under the proposed sanctions, violators of UN sanctions on trade with Iraq will be imprisoned for a minimum of one year, an increase from six months, with 10-year maximum sentences. At least nine German companies are under investigation on suspicion of violating the export laws. .
Legend
HoldsDuring
At
BasedIn
NatoNation
EuNation
AxisOfEvilNati...
MuslimNation
Language
Nation
GPE
Location
GeographicEnt...
Facility
MilitaryOrgani...
Organization
Select All
Deselect All
Hide Unselected
Click In Text to See Annotation Detail
Annotations
ChemicalWeapon
ProducesWeapon
ProducesWeapon ("poison gas plants in Iraq")
begin = 1186
end = 1211
confidence = 0.0
componentId = KSRelationDetector
relationArgs = BinaryRelationArgs
domainValue = GPE ("Iraq")
rangeValue = ChemicalWeapon ("poison gas")
Msl_9146_m9100142.txt

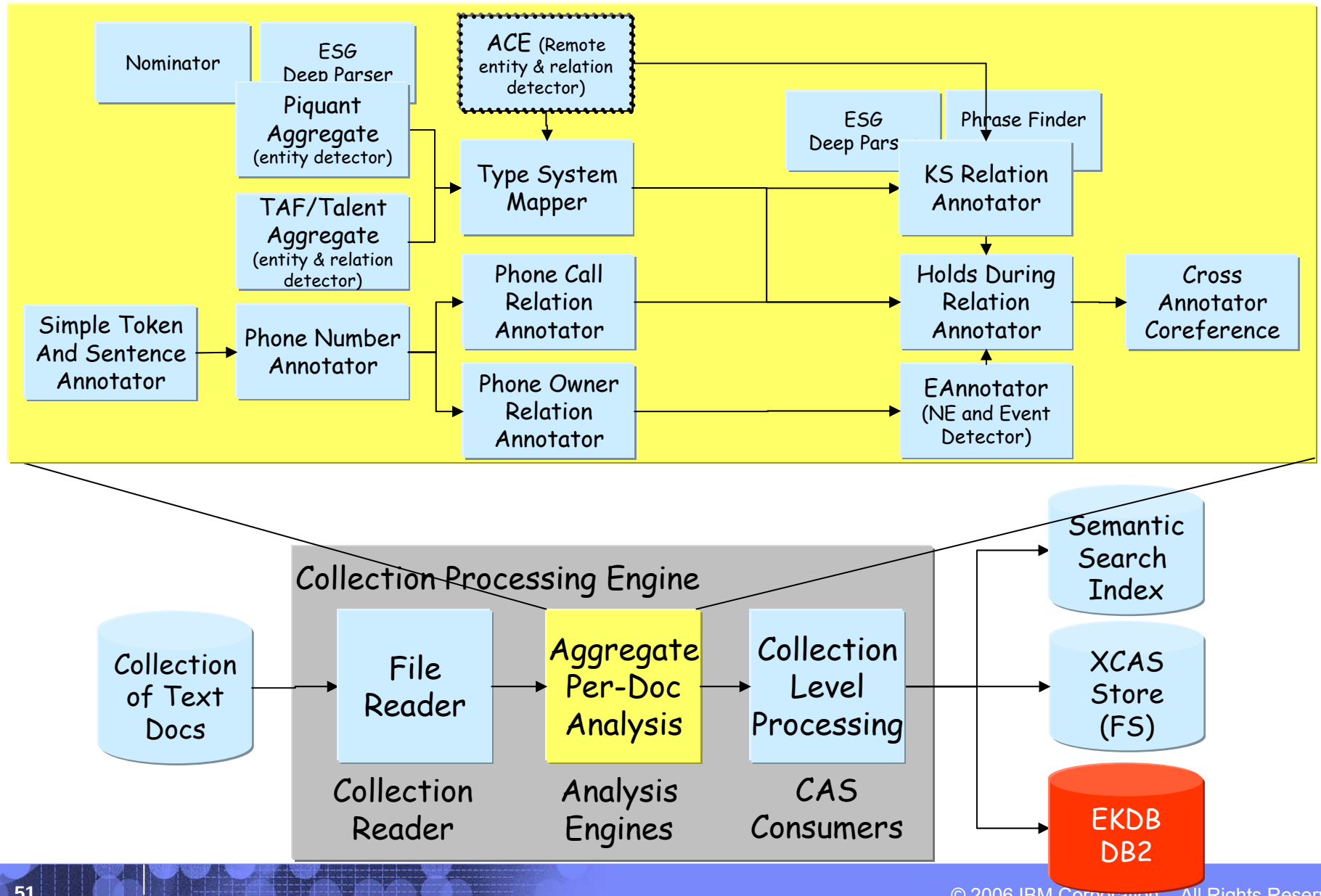


IBM

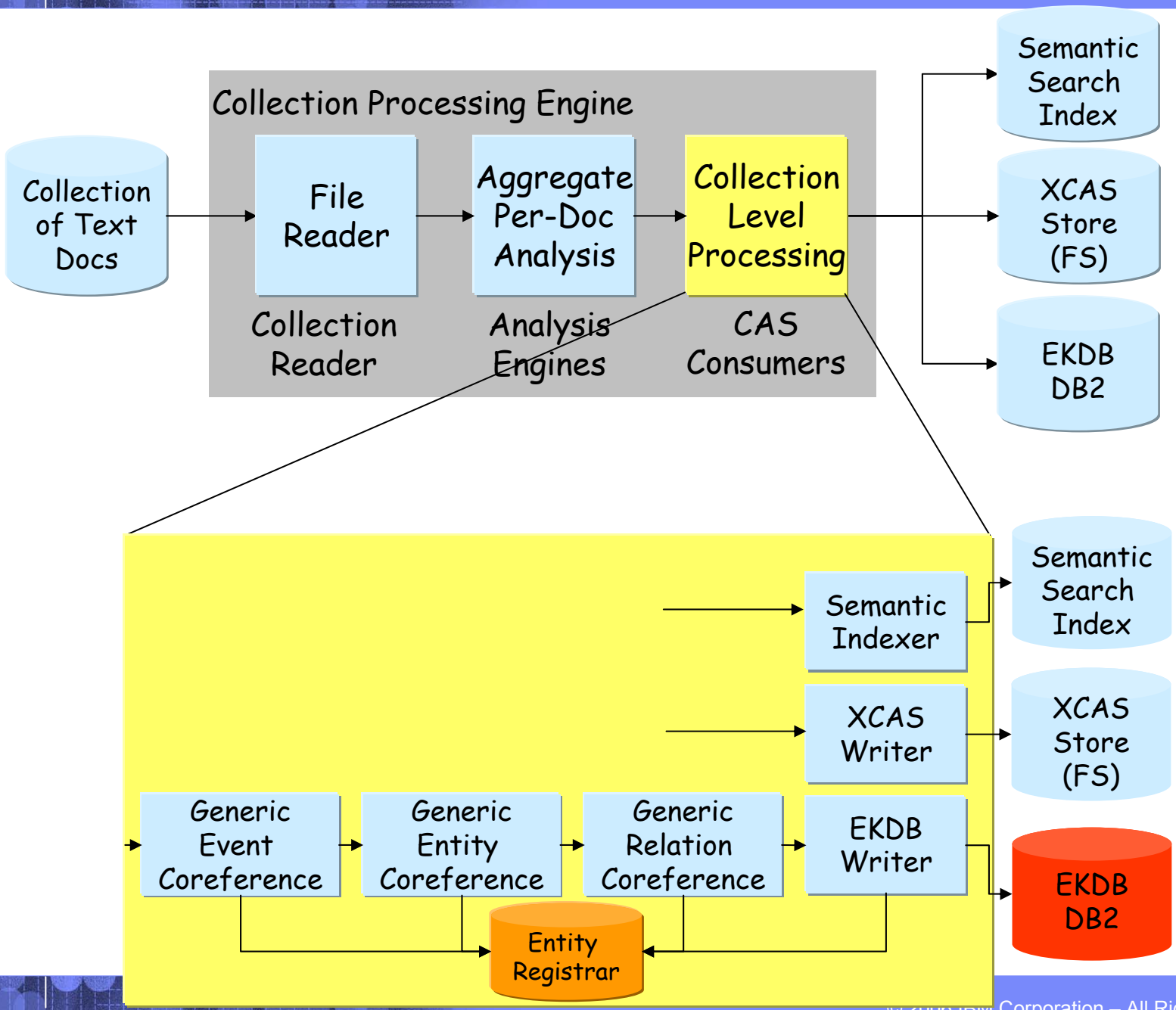
Knowledge Extraction and Integration

Assisting the Transformation from Text to Formal Knowledge

© 2006 IBM Corporation – All Rights Reserved



© 2006 IBM Corporation – All Rights Reserved –

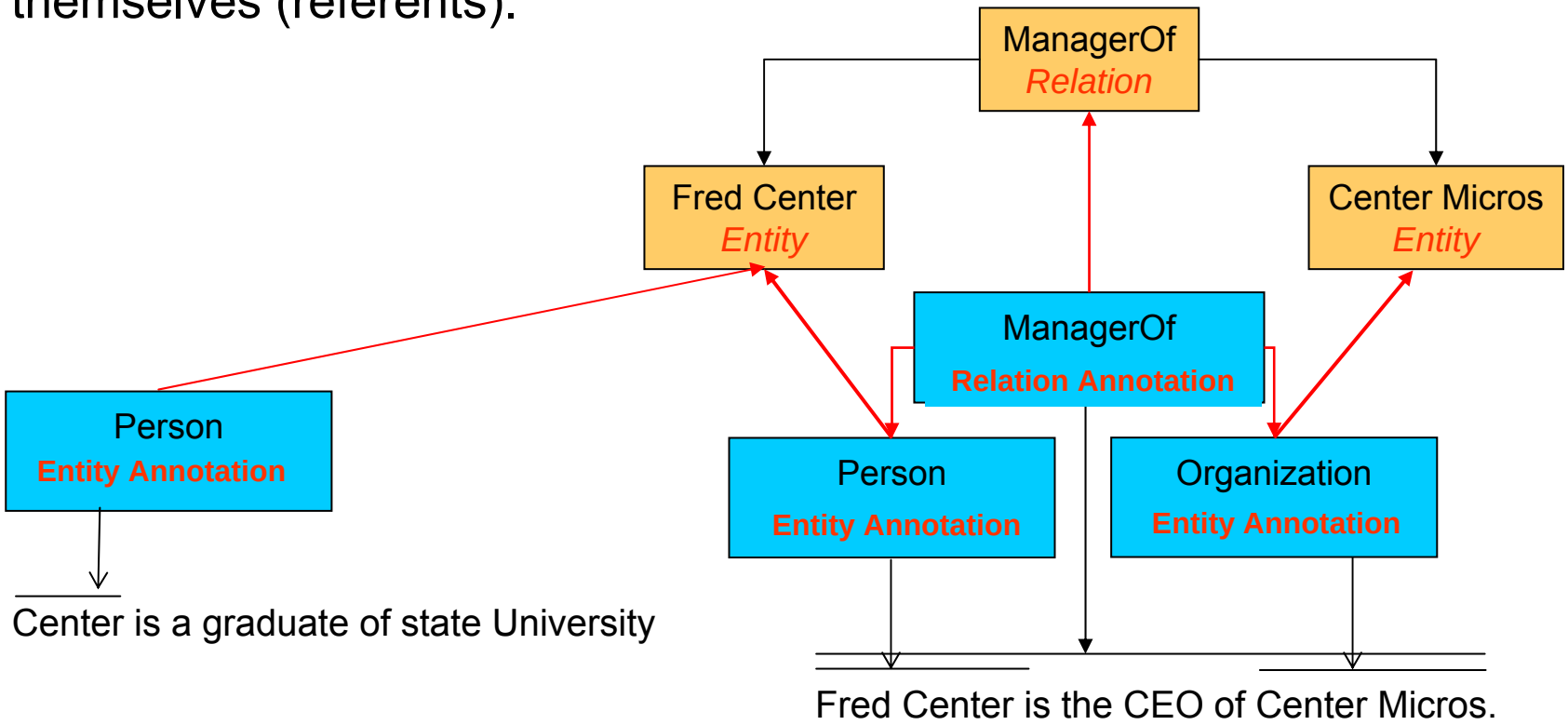


© 2006 IBM Corporation – All Rights Reserved –

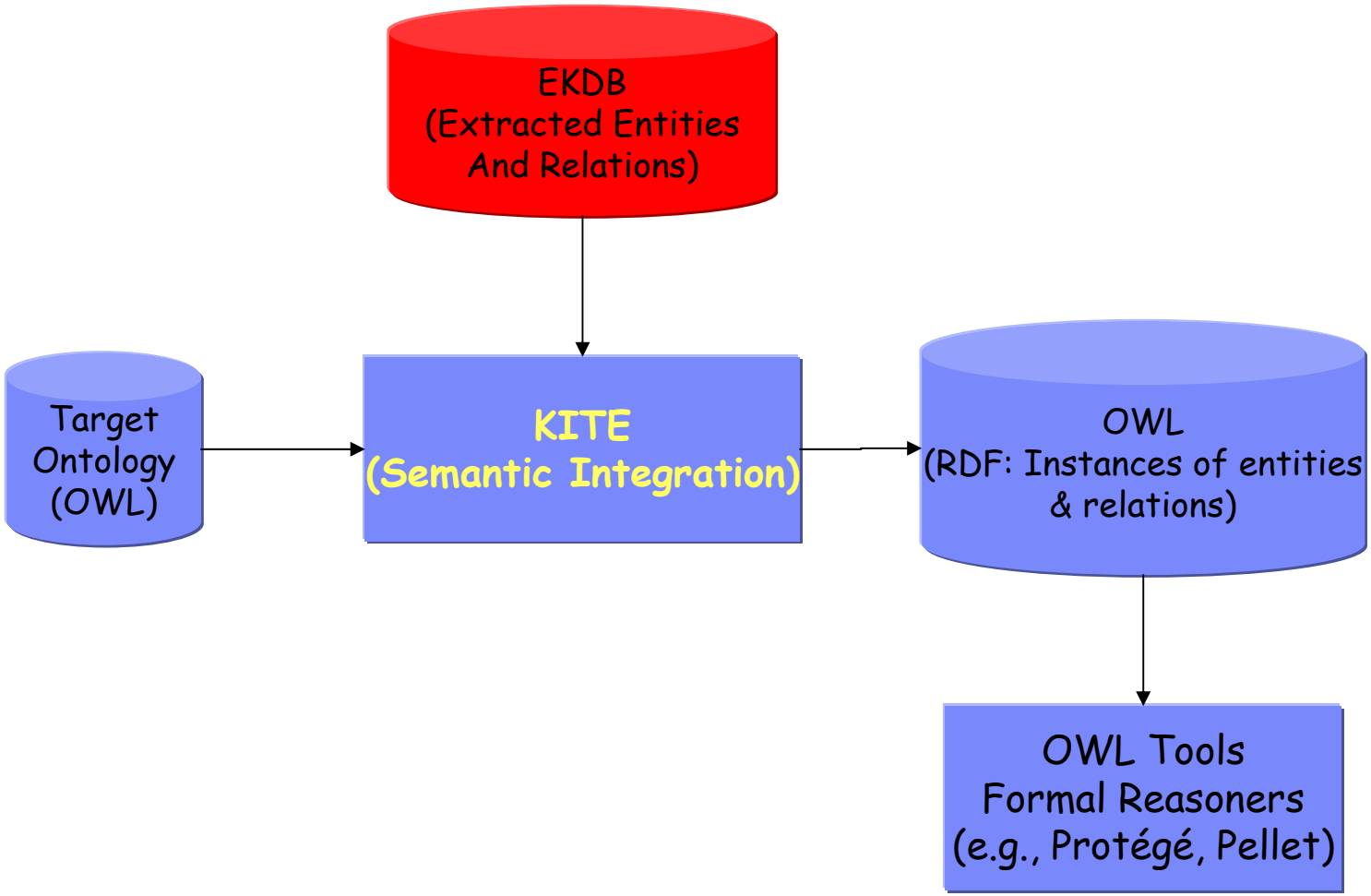
The Knowledge-Level Type (KLT) System imposes essential ontological distinctions on linguistic analysis results

The transformation of analysis results into formal knowledge…

Requires analysis engines to distinguish between entity annotations, relation annotations, mentions of individuals and the individuals themselves (referents).



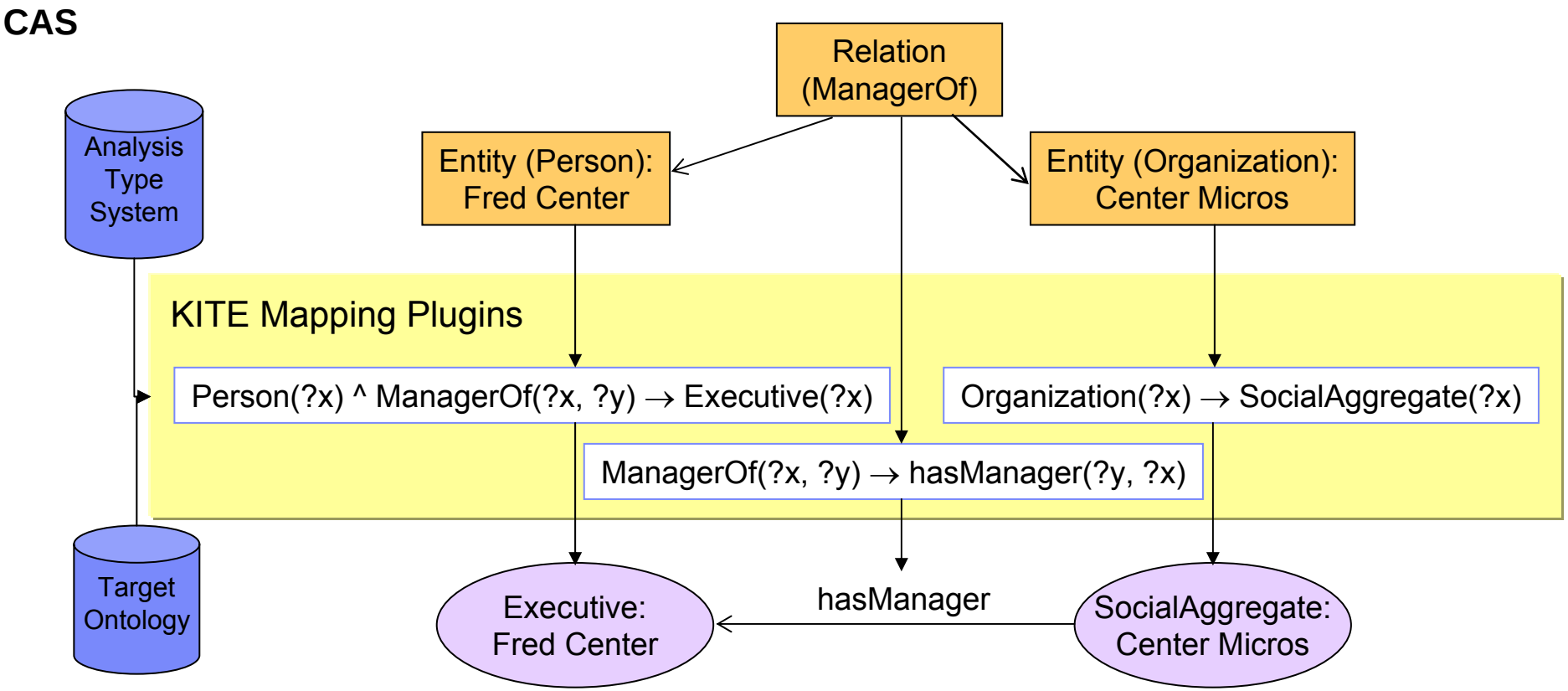
© 2006 IBM Corporation – All Rights Reserved –



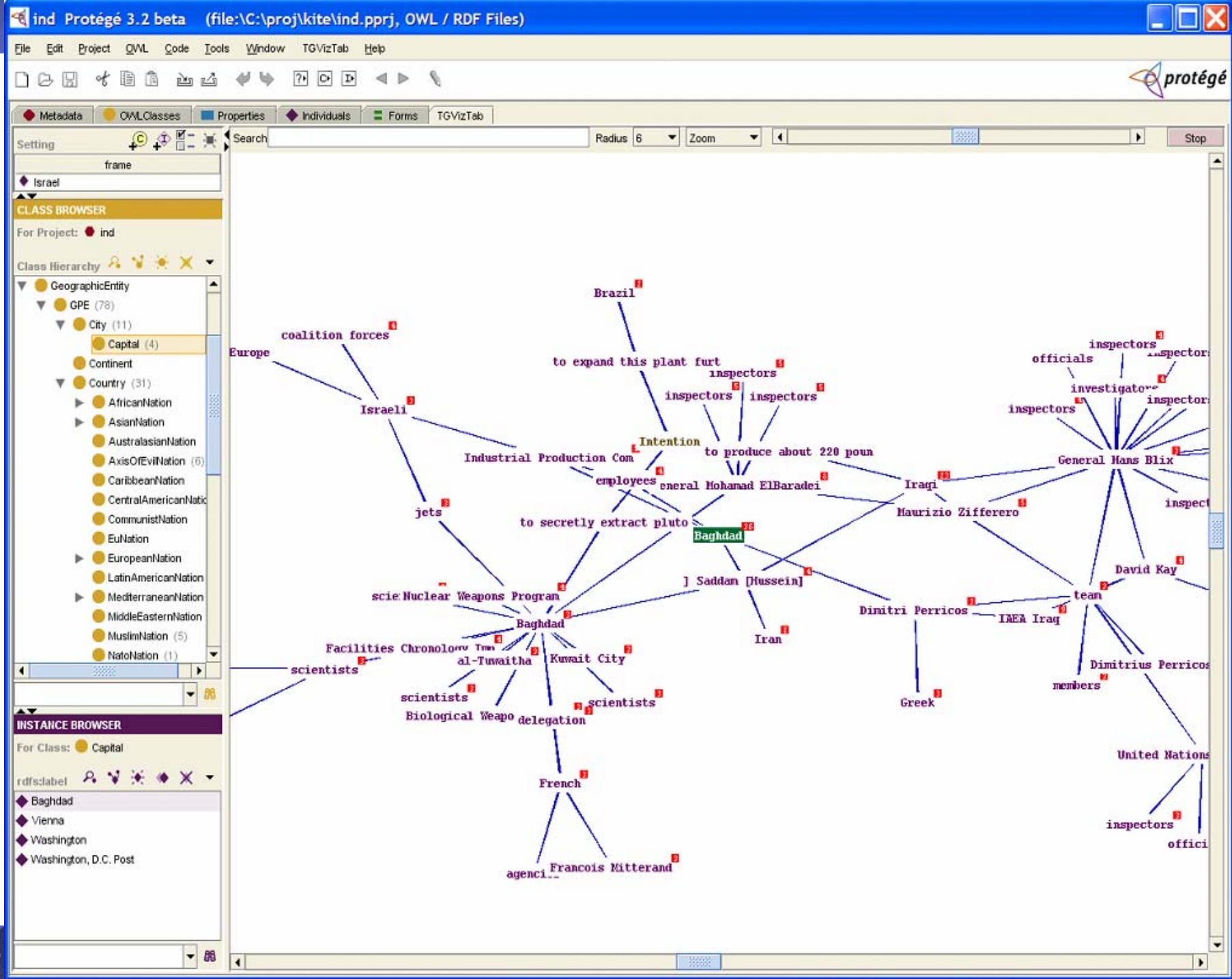
© 2006 IBM Corporation – All Rights Reserved –

Knowledge Integration and Transformation Engine

The transformation of analysis results into formal knowledge requires the explicit mapping of types in the type system to classes and properties in the ontology.

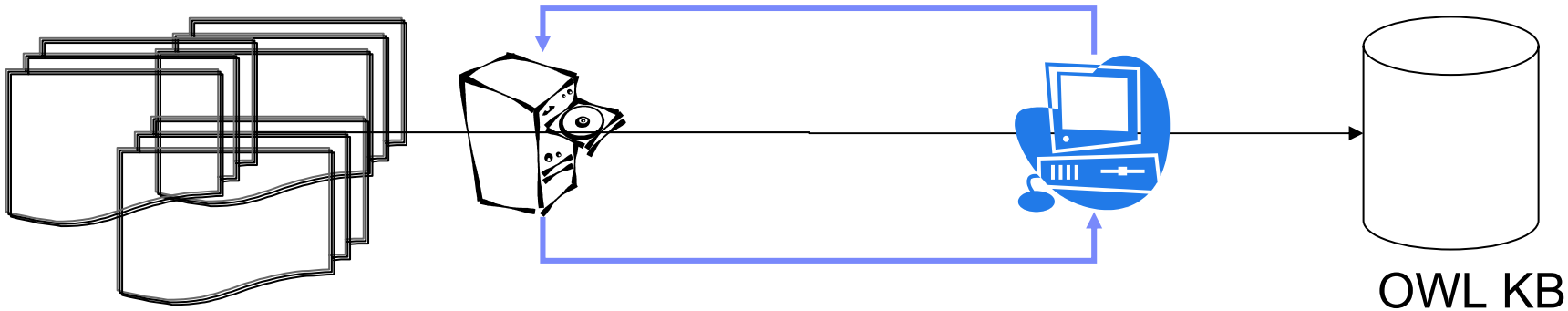


© 2006 IBM Corporation – All Rights Reserved –


ind Protégé 3.2 beta (file:\C:\proj\kite\ind.pprj, OWL / RDF Files)
File Edit Project OWL Code Tools Window TGVizTab Help
protégé
Metadata OWLClasses Properties Individuals Forms TGVizTab
Setting
frame
Israel
CLASS BROWSER
For Project: ind
Class Hierarchy
GeographicEntity
GPE (78)
City (11)
Capital (4)
Continent
Country (31)
AfricanNation
AsianNation
AustralasianNation
AxisOfEvilNation (6)
CaribbeanNation
CentralAmericanNatic
CommunistNation
EuNation
EuropeanNation
LatinAmericanNation
MediterraneanNation
MiddleEasternNation
MuslimNation (5)
NatoNation (1)
INSTANCE BROWSER
For Class: Capital
rdfs:label
Baghdad
Vienna
Washington
Washington, D.C. Post
Search
Radius 6
Zoom
Stop
Brazil
coalition forces
Europe
to expand this plant furt
inspectors
inspectors
inspectors
officials
inspectors
Israeli
Intention
Industrial Production Com
to produce about 220 poun
General Hans Blix
employees
eneral Mohamad ElBaradei
Iraqi
jets
to secretly extract pluto
Maurizio Zifferero
Baghdad
David Kay
] Saddam [Hussein]
scie Nuclear Weapons Program
team
Dimitri Perricos
IAEA Iraq
Baghdad
Iran
Facilities Chronolo
al-Tuwaitha
Kuwait City
scientists
Dimitrius Perricos
members
scientists
scientists
Biological Weapo
delegation
Greek
United Nations
French
inspectors
Francois Mitterand

Automated Knowledge Acquisition

…Not Gonna Happen…



Documents or other media

- Unstructured
- Implicit Semantics
- Ambiguous
- Imprecise
- Incomplete
- Inconsistent
- Partially relevant

Automation + User Interaction

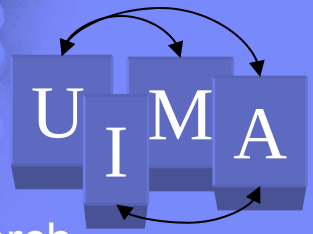
Tools and process tolerates less than prefect precision to produce candidate task-relevant knowledge

OWL KB

- Perfect Knowledge
- Precise & Complete Rendering
- Ready to Reason

Task-Relevant Knowledge Assessable to Humans & Automated Reasoners

© 2006 IBM Corporation – All Rights Reserved –



Thank You

© 2006 IBM Corporation – All Rights Reserved

Query Analysis: Simple *Answer Type Insertion*

Want User to Type these….

- **What is Webster's Phone Number**
  - Webster's Phone Number
  - …
- **Where is Smith?**
  - Smith's Place
  - …

The System to Generate These….

- **+Webster phone number**
- **+<PhoneNumber> </PhoneNumber>**

- **+Smith**
- **<Location> </Location>**

- **+Smith place**
- **<Location> </Location>**

© 2006 IBM Corporation – All Rights Reserved –

Query Analysis: *Exploiting Relation Detection*

Want User to Type these….

- **Who owns Select Gourmet Foods?**
- **Select Gourmet Foods owner?**

Still targeted but *Greater Recall*

The System to Generate These….

Greater Precision

- <Owner>
  - +<> <Person> </Person> <Organization> </Organization><>
  - +<Organization> Select Gourmet Foods </Organization>
- </Owner>

- owner
- <Organization> +Select +Gourmet +Foods </Organization>
- +<>
  - <Organization> </Organization>
  - <Person> </Person>
- **</>**

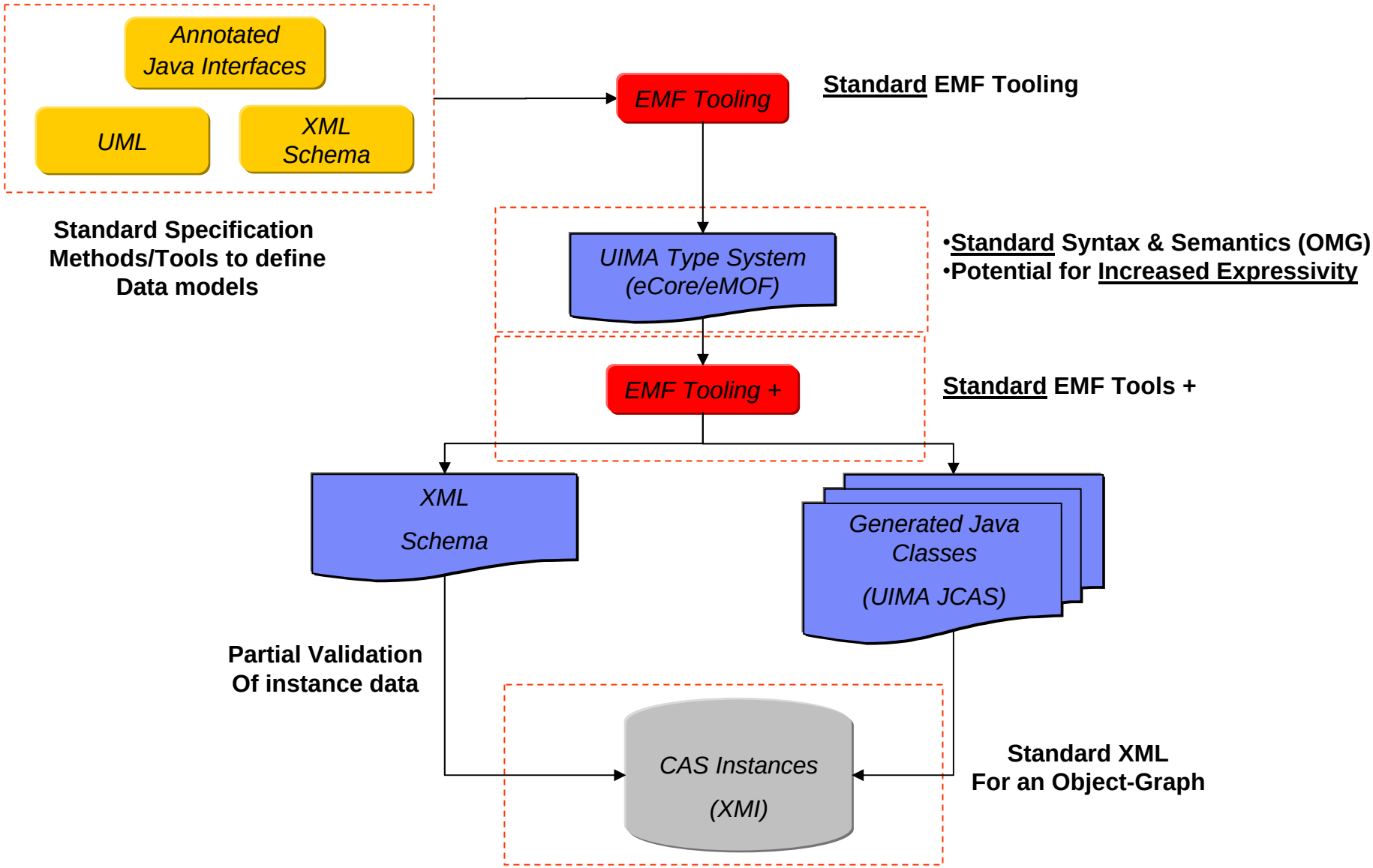
© 2006 IBM Corporation – All Rights Reserved –

UIMA Data Representation Alignment with Standards

- **CAS (Instances)**
  - Semantics: Object Graph
    - A *stand-off* representation leaves original artifact untouched
    - Expressivity allows multiple interpretations of artifacts, overlapping annotations
  - XML Syntax: XMI (inherit open-source tooling: http://www.eclipse.org/emf/)
  - Inline XML may be generated
  - OWL RDF may be generated
- **CAS Type System (Schema)**
  - Semantics: Ecore/EMOF
  - XML Syntax: XMI
  - XML Schema for partial instance validation can be generated
  - Java Classes for CAS Types can be generated
- **Component Metadata**
  - For Discovery, Reuse and Application-Level Plug-in-Play
  - XML document declaratively capture identification, configuration, behavior and deployment information about UIMA components.
  - Tools to create standard **OSGi** Bundles for painless application integration from servers to cell phones http://www.aqute.biz/osgi.html

© 2006 IBM Corporation – All Rights Reserved –

Leverage open-source Eclipse/EMF Tooling



© 2006 IBM Corporation – All Rights Reserved –