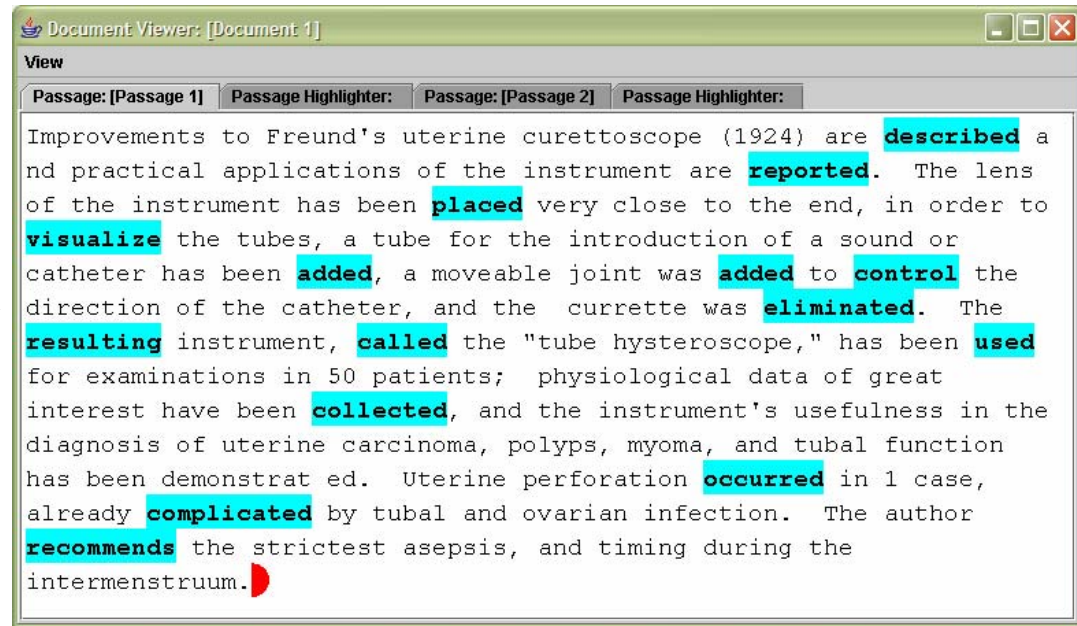
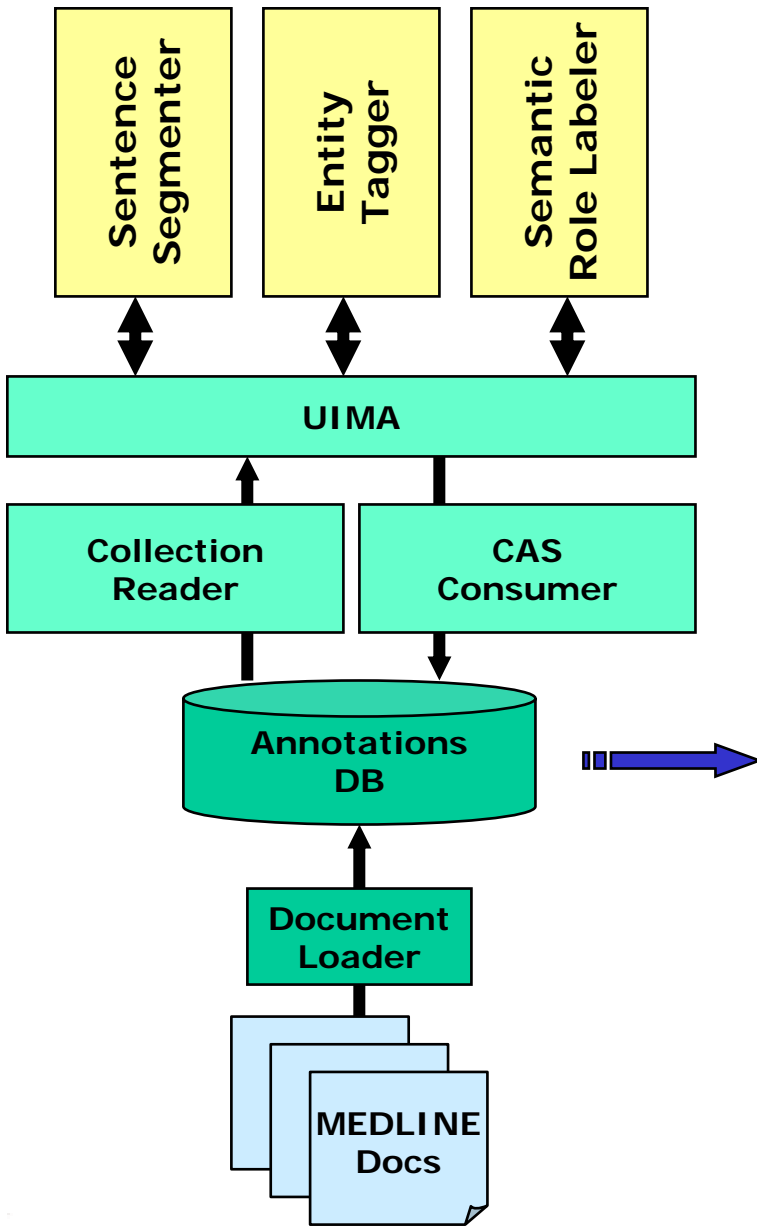


What's Ready Now?

- Segmenter: MX Terminator
- Tagger: BBN Identifinder
- Role Labeler: ASSERT



Annotations Viewer



Semantic Role Labeling

- We've been using ASSERT to label predicate targets and arguments:
<http://oak.colorado.edu/assert/>
- Also see Dan Roth's SRL:
<http://l2r.cs.uiuc.edu/~cogcomp/demo.php?dkey=SRL>
(top system at CoNLL)



Semantic Indexing

- ADB documents that have been annotated can be indexed and searched using Lemur/INDRI
- Semantic information (e.g. predicate targets, arguments) is indexed and queried in addition to keywords
- Example:
http://durazno.lti.cs.cmu.edu/javelin_public/predicate_demo/
- See Matt Bilotti for more details



ADB Pointers

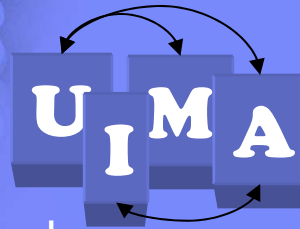
- ADB Documentation Page:
http://guadalajara.lti.cs.cmu.edu/annotations_db/
 - Data model description
 - Java libraries for ADB API, ADB Viewer
 - Javadoc
- Sample ADB server on seit1.lti.cs.cmu.edu
 - mySQL (download Query Browser from mysql.com, or use ADB Viewer)
 - Username: seituser
 - Password: tagit
 - Databases of interest: medline_small, phosphorylate



UIMA Pointers

- UIMA SDK / tutorial:
<http://www.alphaworks.ibm.com/tech/uima>
- UIMA source code (for experts):
<http://uima-framework.sourceforge.net/>
- UIMA Component Repository (new):
<http://uima.lti.cs.cmu.edu/>





UIMA | IBM Research

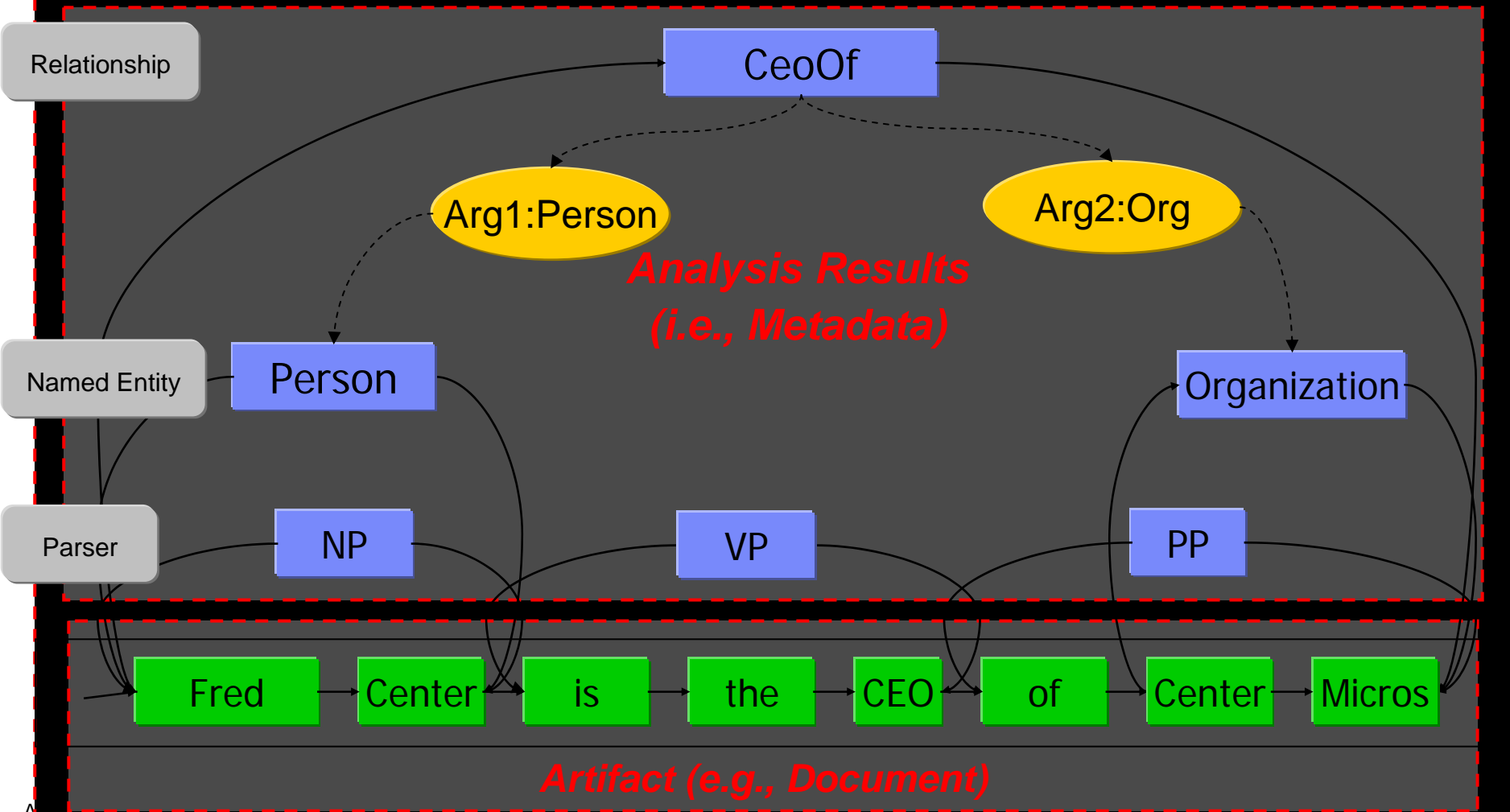
Architecture Highlights

Common Representations, Interfaces and Design Patterns

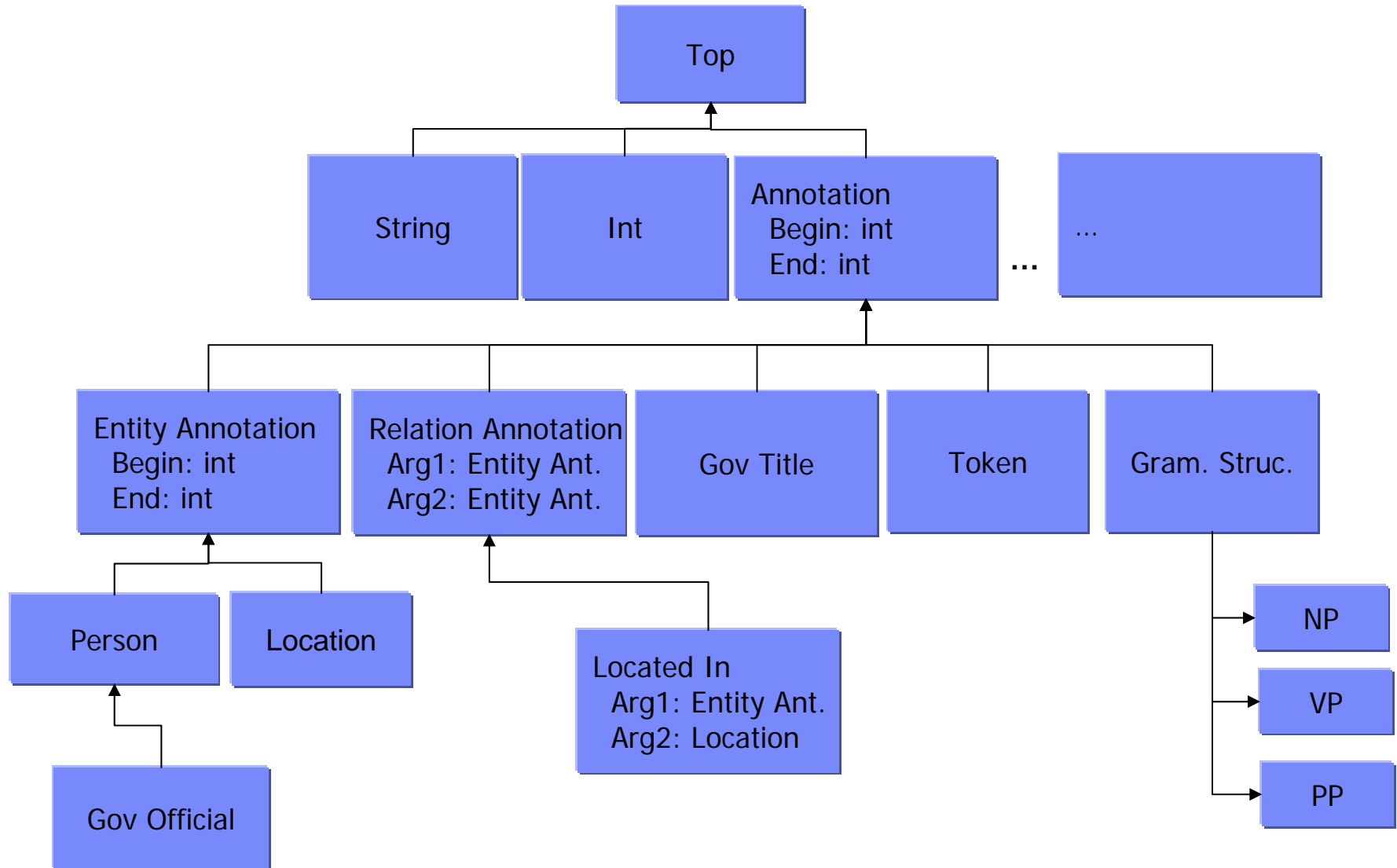
How it works

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.

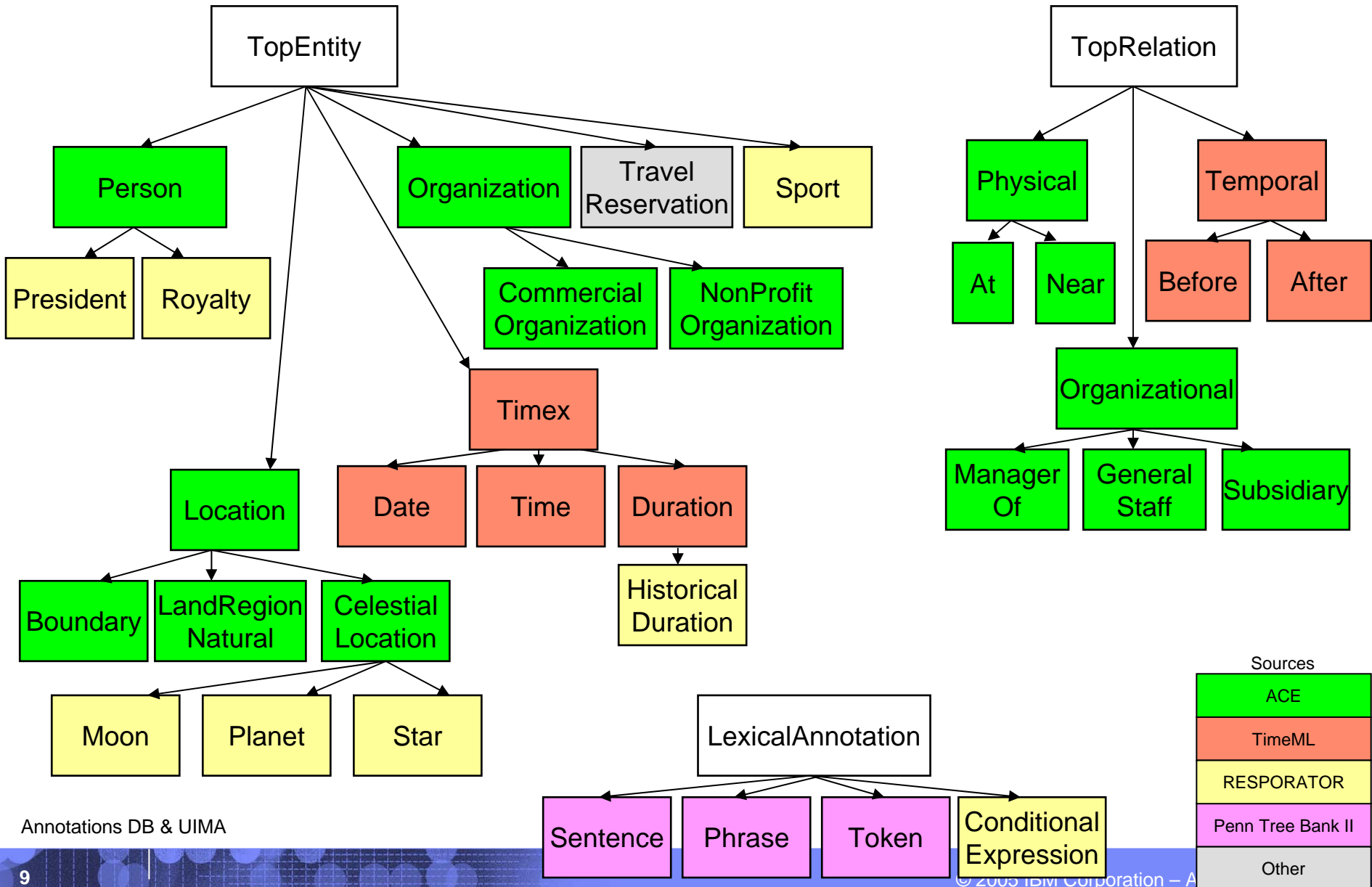
Common Analysis Structure (CAS)



Sample Type System



Partial HUTT Type System (254 concepts in total)



UIMA Annotation Viewer

Report Date 10 March 2003. Slick business dealings keep local olive oil importer out of the pits. Robert Crane was recognized by local business leaders for his skill at leading the Gorman Food Importers Inc. to strong profits while others are struggling. Mr. Crane, owner of Gorman Food Importers Inc., has consistently been able to produce exceptional results, while still keeping a focus on his employees. Gorman Food Importers Inc. has been in business since 1970 and specializes in food imports from the Middle East, including olive oil and figs. Gorman Food Importers Inc. is headquartered in NYC, and their warehouse is located in Paramus, NJ. The company employs 659 people in the two locations. Robert Crane can be reached at 608-703-2317.

Click In Text to See Annotation Detail

- Organization ("Gorman Food Importers Inc.")
 - begin = 185
 - end = 211
 - componentId = ACE
 - mentionType = NAME
- Organization ("Gorman Food Importers Inc.")
 - begin = 185
 - end = 211
 - componentId = IBMEAnnotator
 - mentionType = NAME

Legend

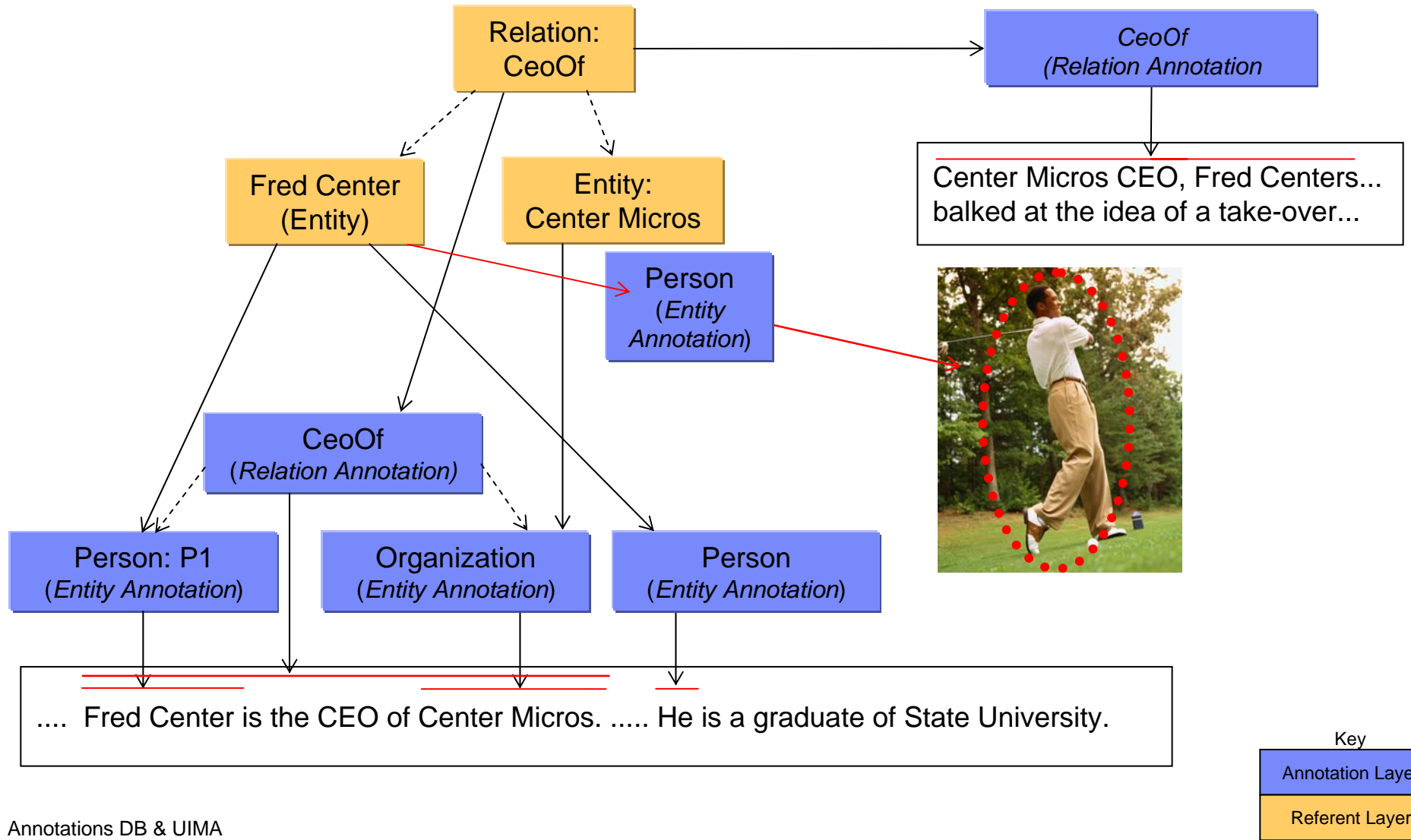
<input checked="" type="checkbox"/> Person	<input checked="" type="checkbox"/> Facility	<input checked="" type="checkbox"/> GPE	<input checked="" type="checkbox"/> Organization	<input checked="" type="checkbox"/> ...
<input checked="" type="checkbox"/> GeneralStaff	<input checked="" type="checkbox"/> BasedIn	<input checked="" type="checkbox"/> Management		

Select All Deselect All Viewer Mode: Annotations

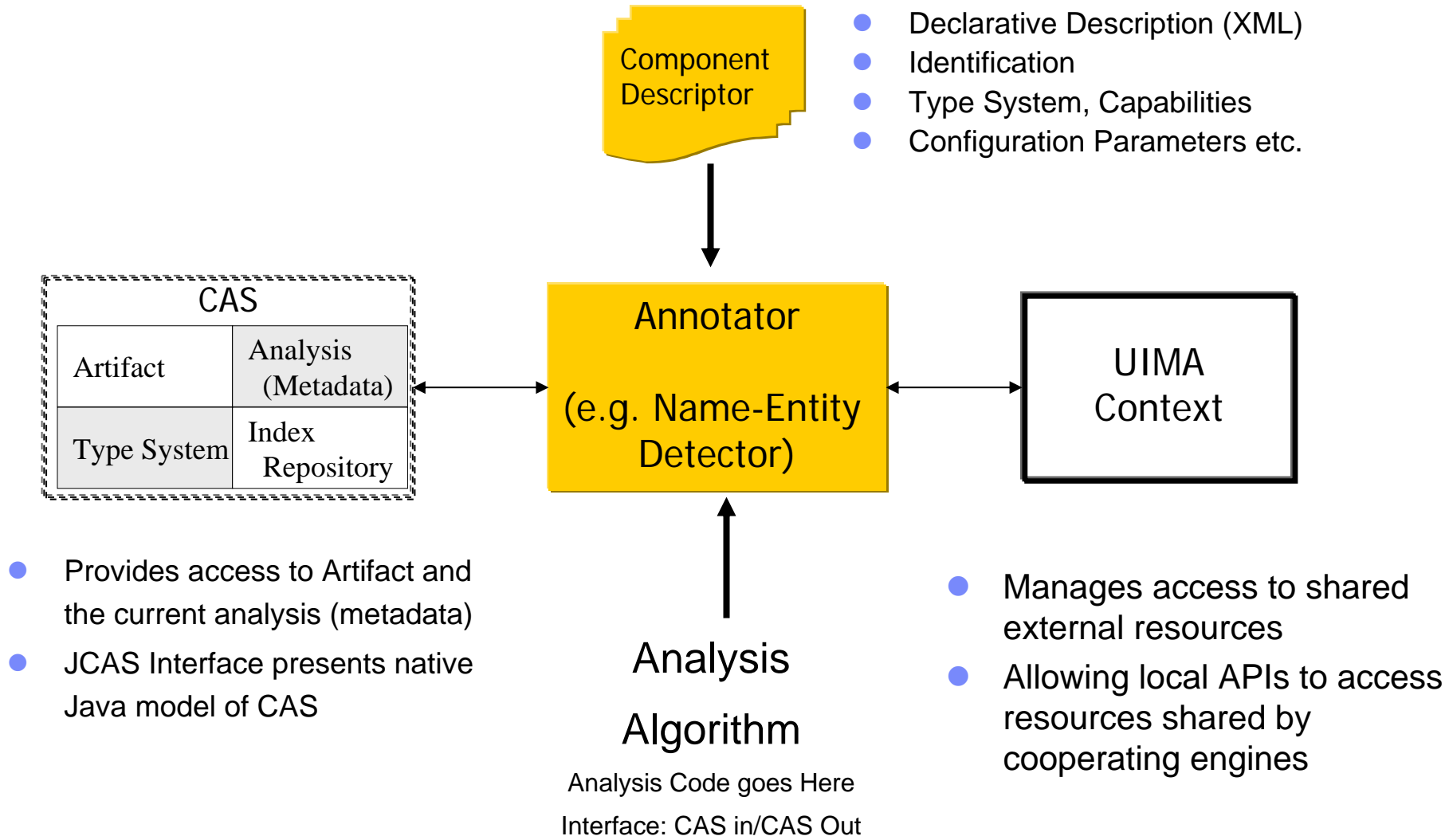
A CAS

- Analyzed by a combination of Analysis Engines
- Semantic Entities & Relations Represented
- Highlighted here in a GUI

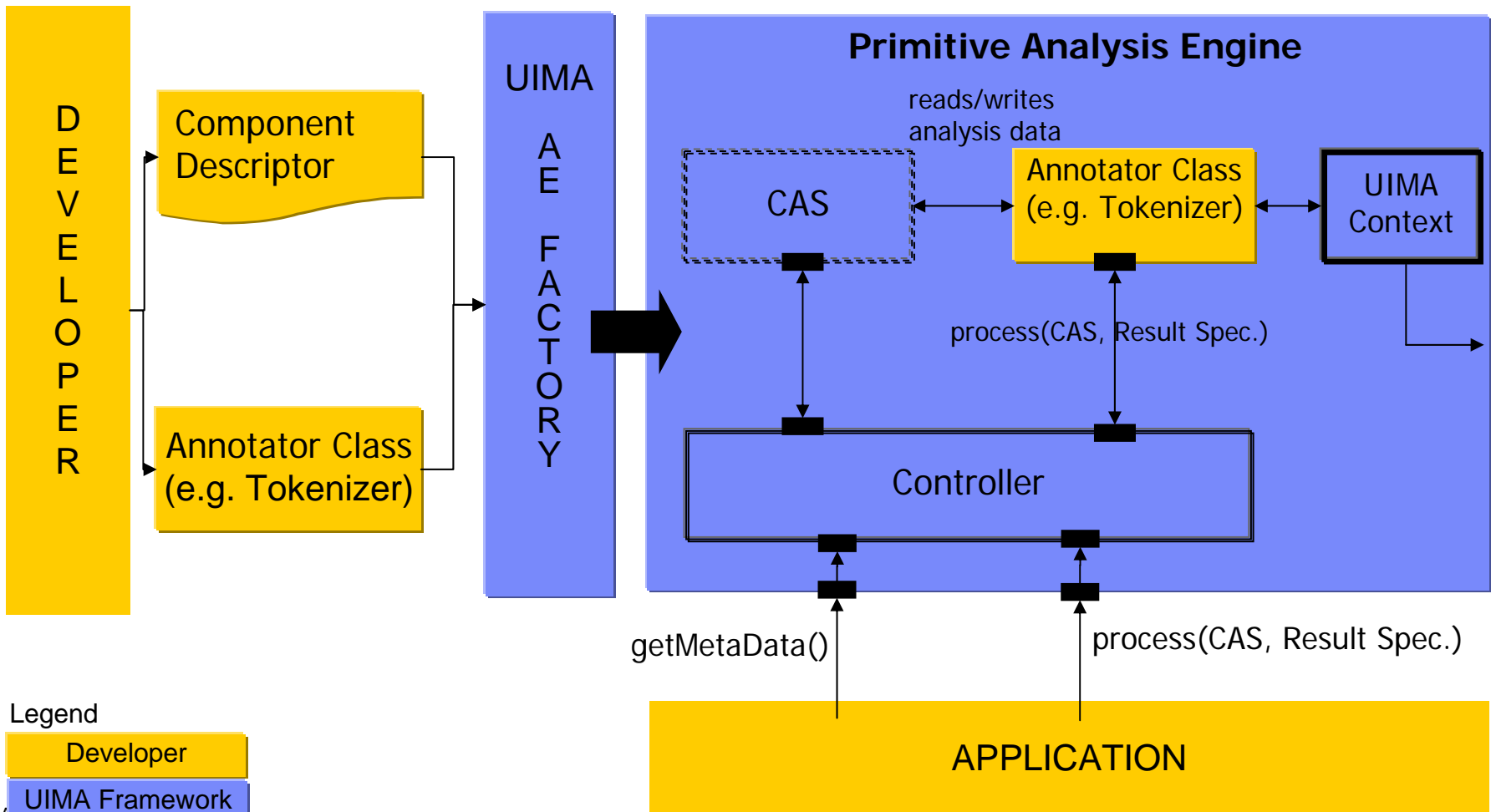
Annotations and Referents



The Basic UIMA Component Interfaces

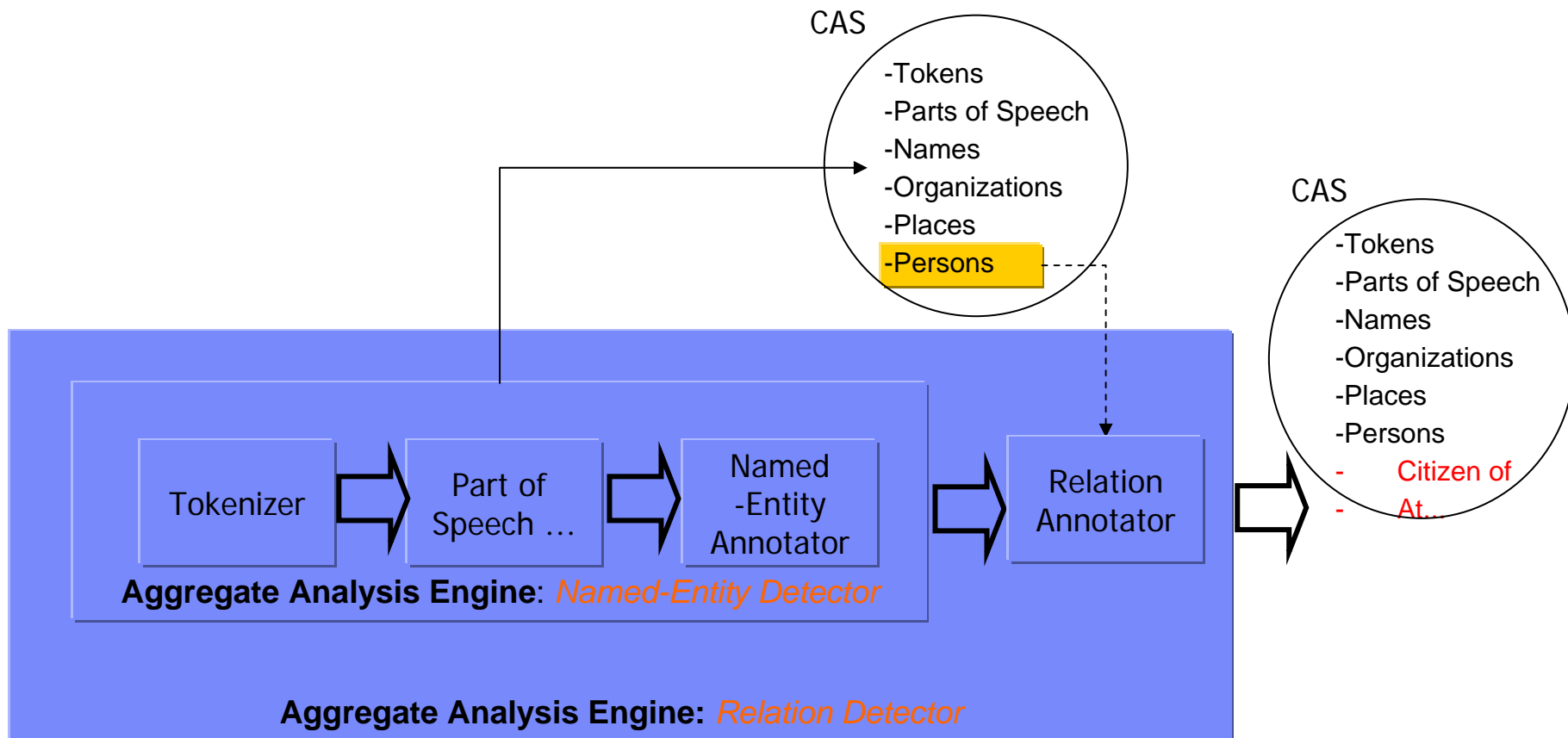


UIMA Frameworks has factories that take developers code and build pluggable engines

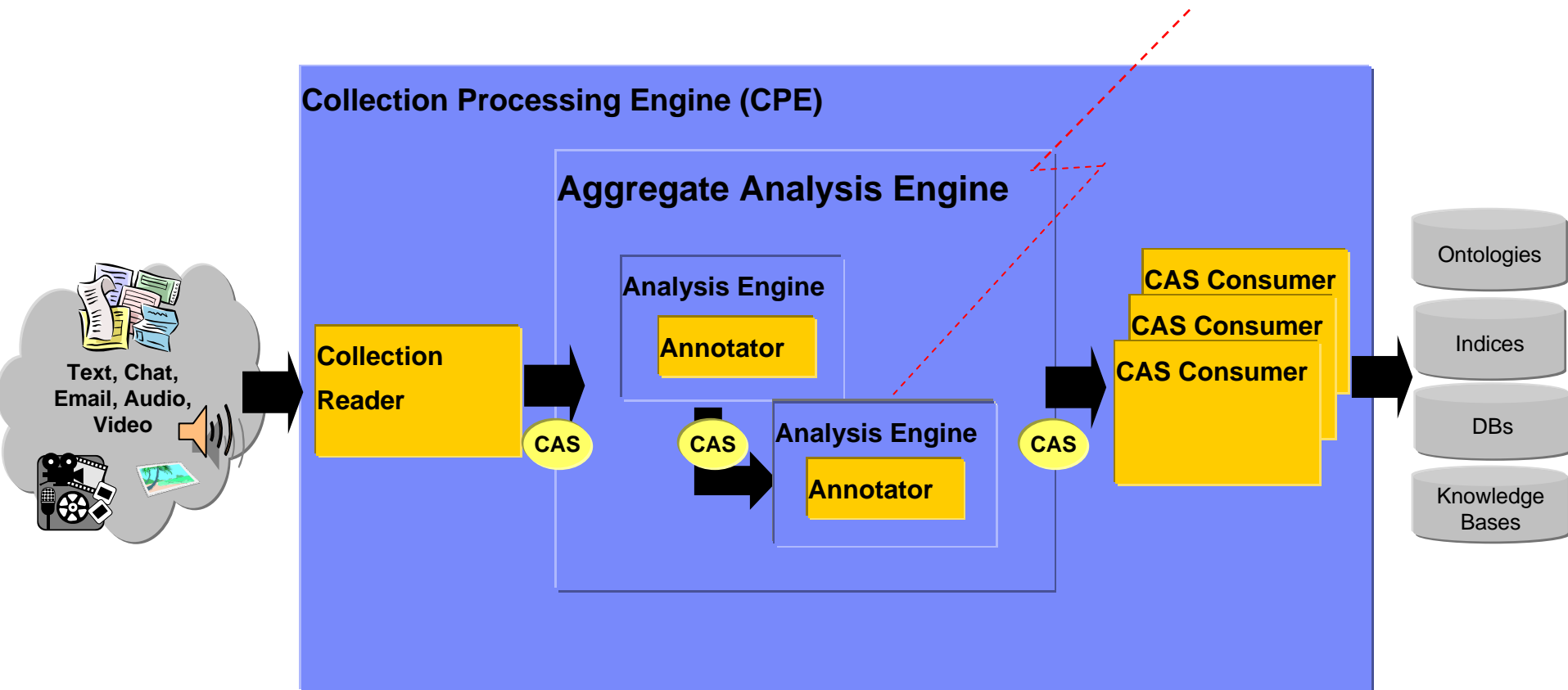


- Legend
- Developer
 - UIMA Framework

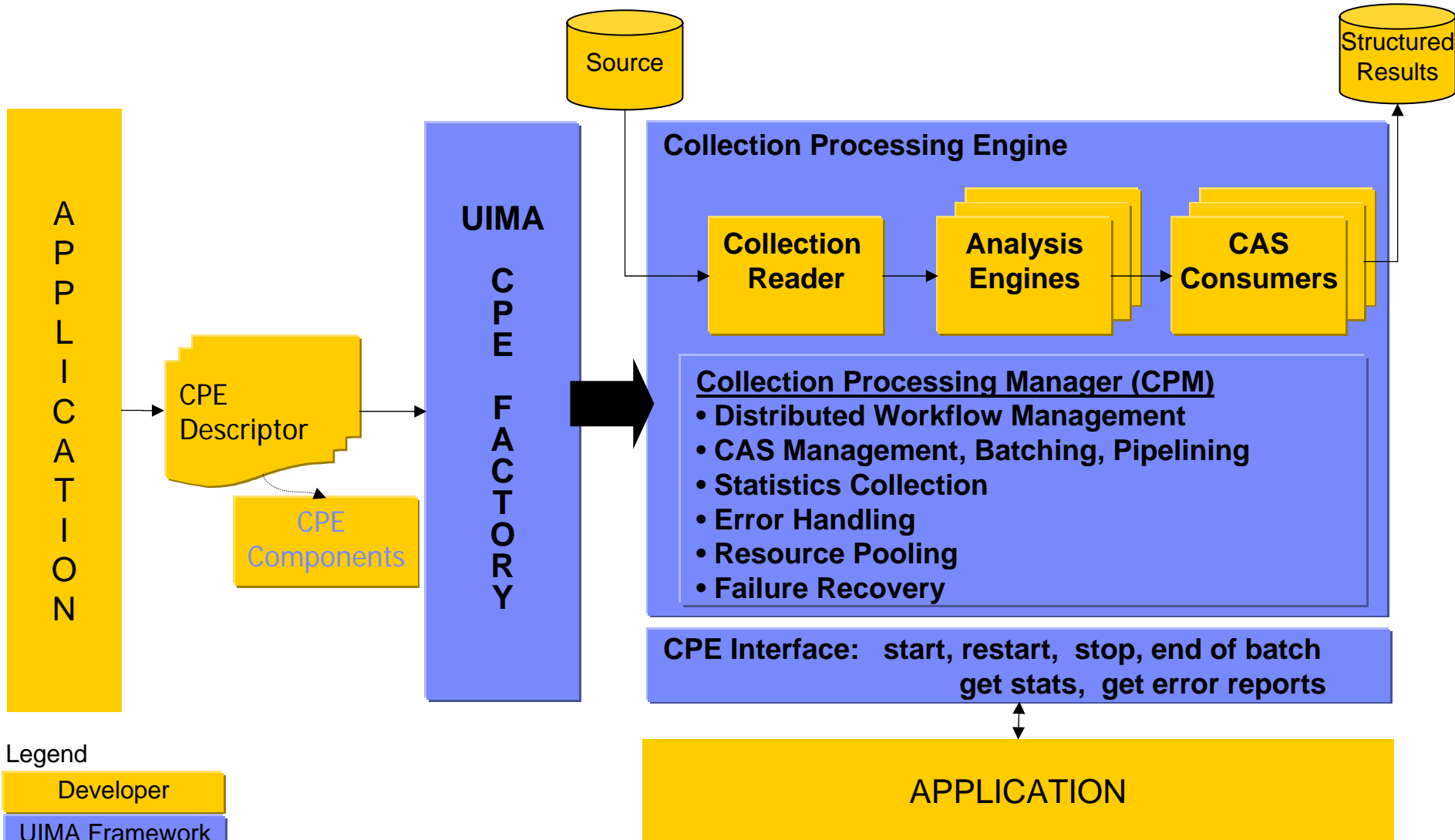
The UIMA framework allows the developer to compose & encapsulate a set of engines as a single **aggregate** component, insulating applications or users from its implementation details/complexities – Promoting Reuse.



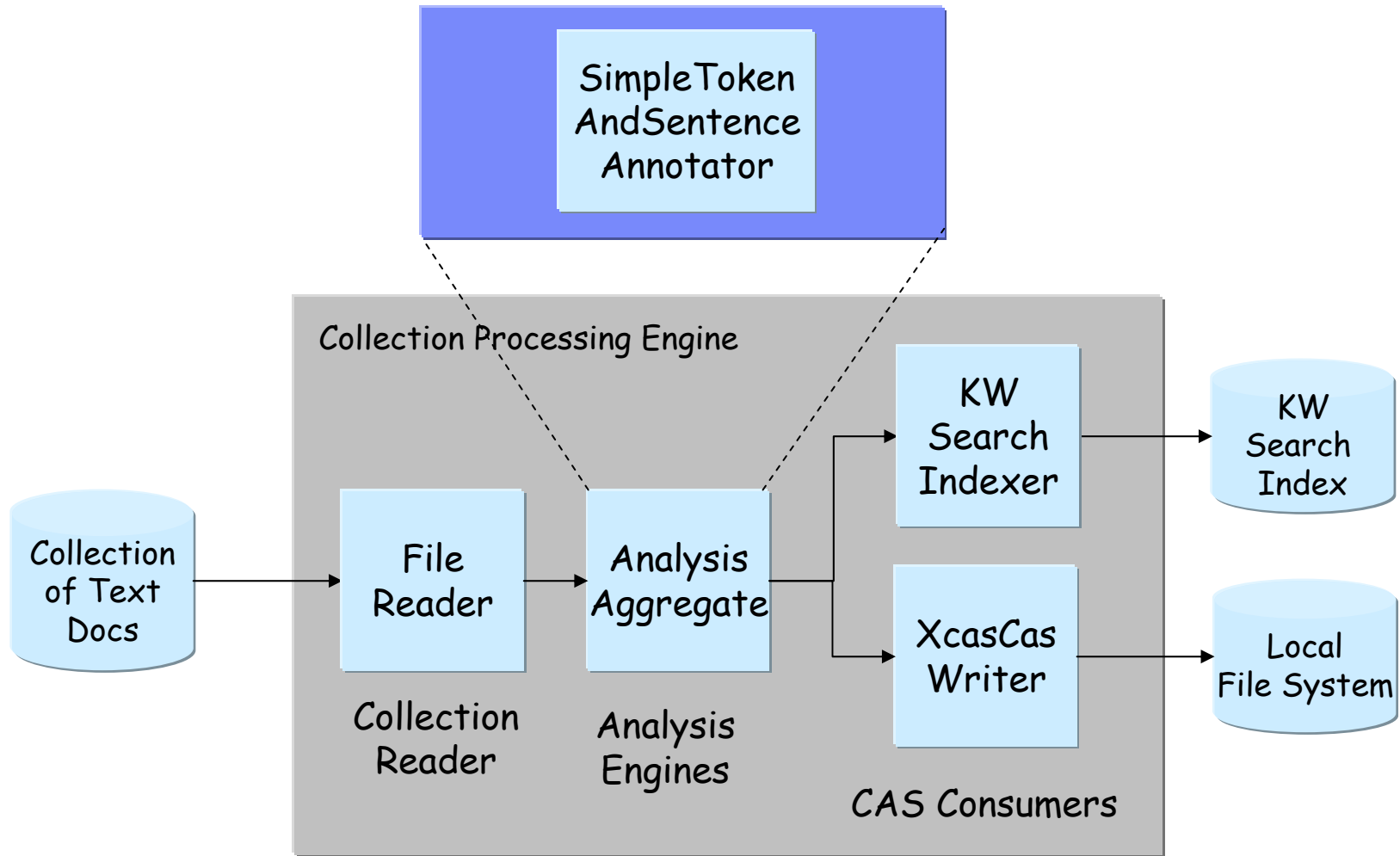
Collection Processing Engines: Aggregate Analytic Components From “Source to Sink”



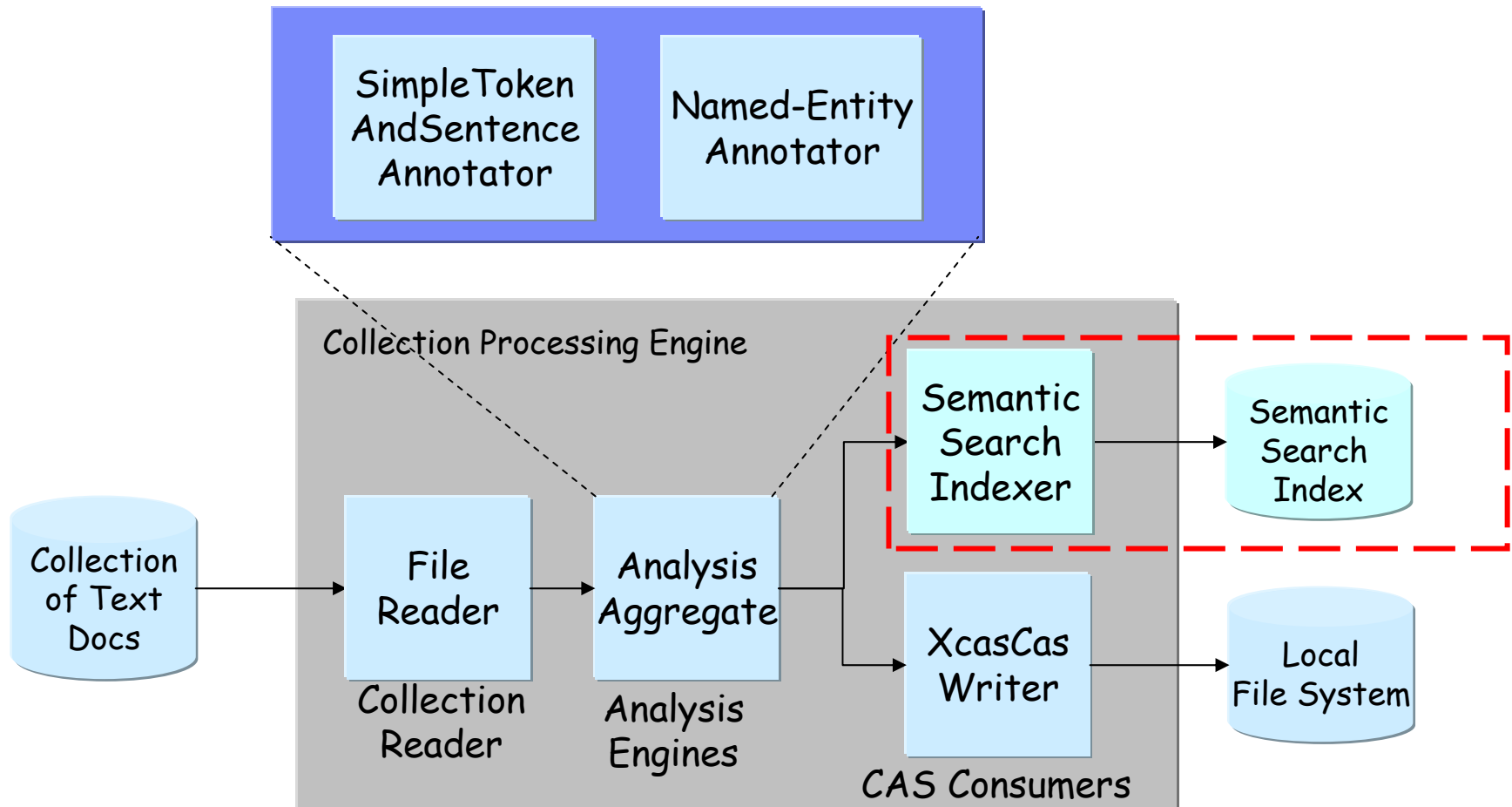
The UIMA Framework provides exploits scaleable infrastructure for running CPEs over large collections in a variety of deployment scenarios.



UIMA Pipeline for Keyword Search



UIMA Pipe Line for Semantic Search



Looking for a person with “first” in his/her name

KeyWord Query:

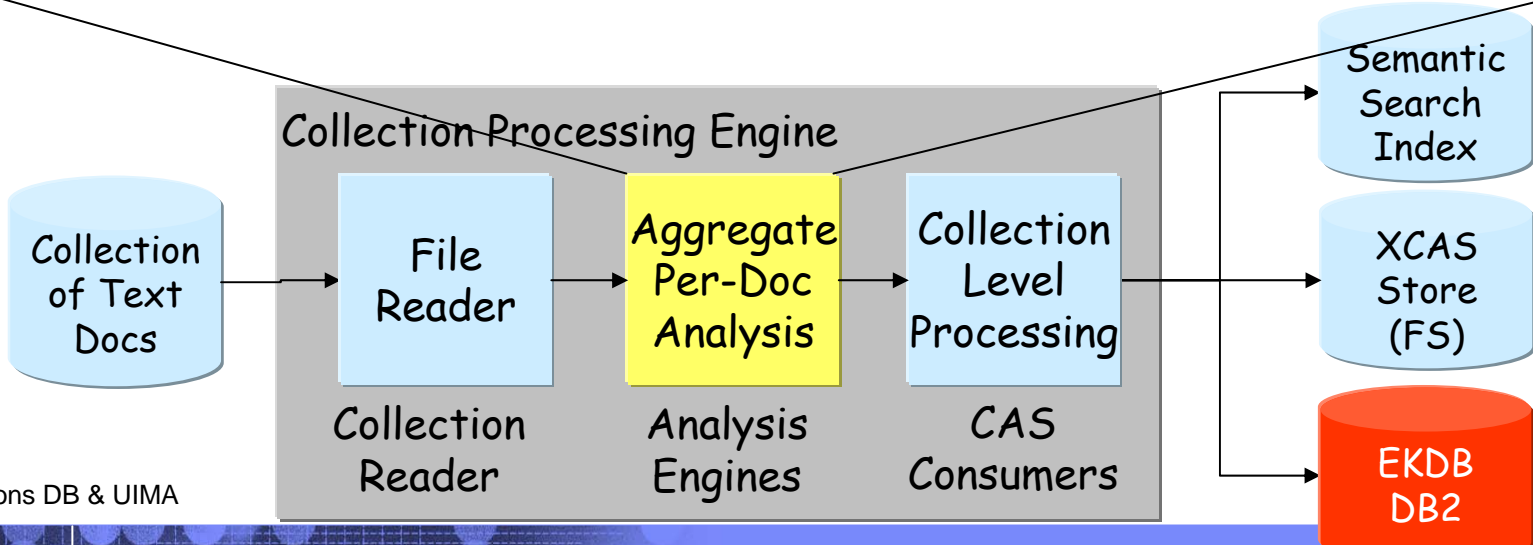
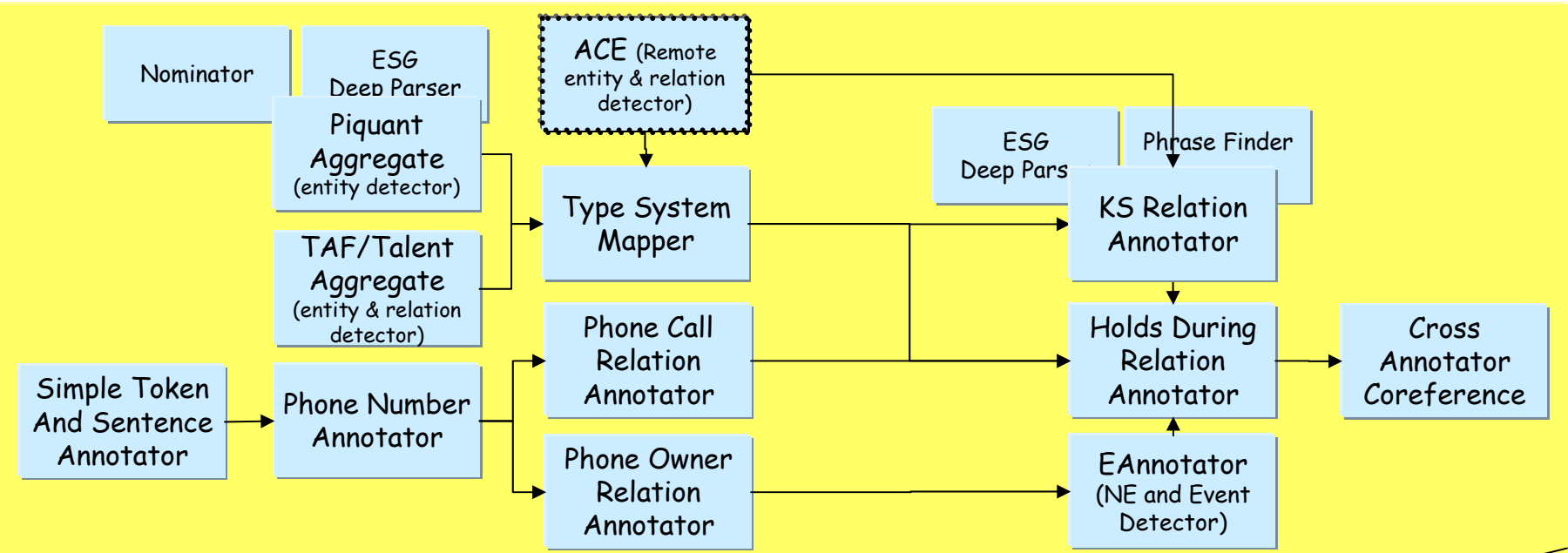
“first”

Semantic Search Query:

<organization> first </organization>

Semantic Search Query:

<ceo_of> <person>Center</person> </ceo_of>



Annotations DB & UIMA

UIMA Pipe Video segmenter

Keyframe
And Icon
Extractor



MARVEL MPEG-7 Video Search Engine 2005 Edition

File Edit View Favorites Tools Help

Back Search Favorites Media

Zoom: 0.5x, 0.75x, 1x, 1.5x, 2x Hits: 12, 24, 48, 100, 200, 500 Icons: Thumbnails, Frames, Shots

Operation: None Apply

Local System
Semantic Search Index
MARVEL Index

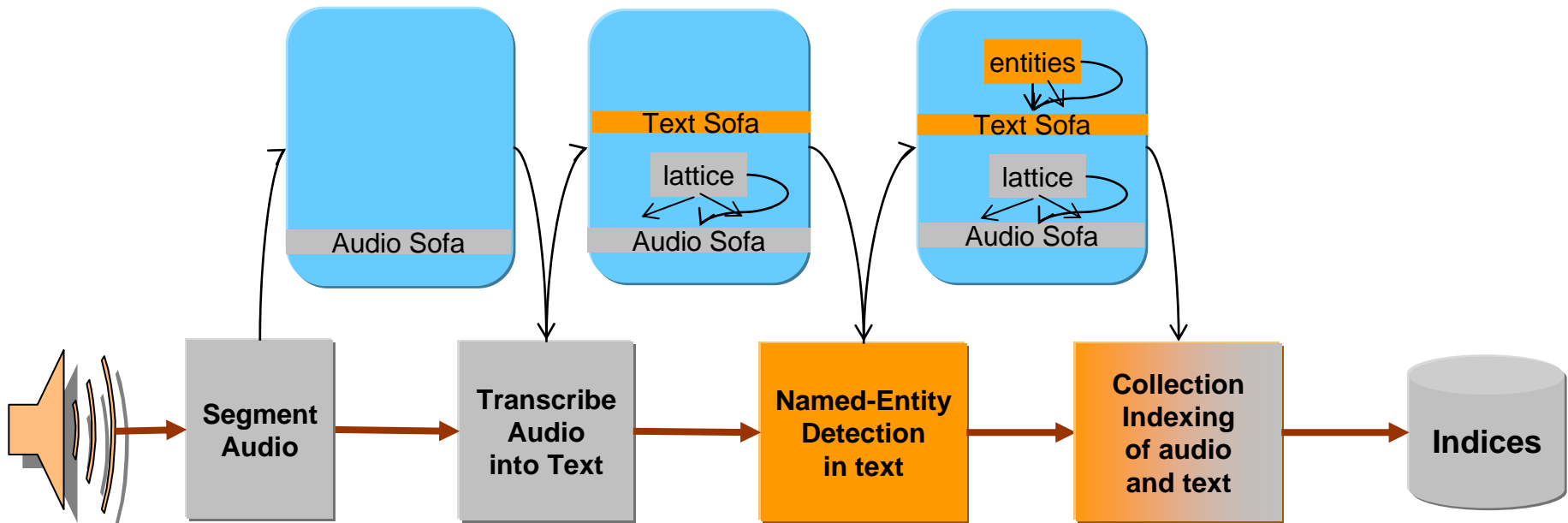
Annotations DB & UIMA

http://localhost/cgi-marvel-bin/WebCBR03.exe?8.COLS=28TD=08L=12548.COMB=18AIGCA=38CP=238DM=2

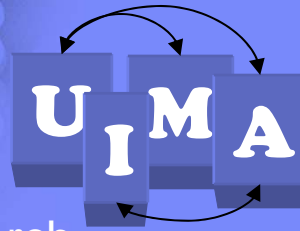


Supports Multi-Modal and Multi-Lingual Analysis

Multiple Subject of Analysis (Sofa) within a CAS



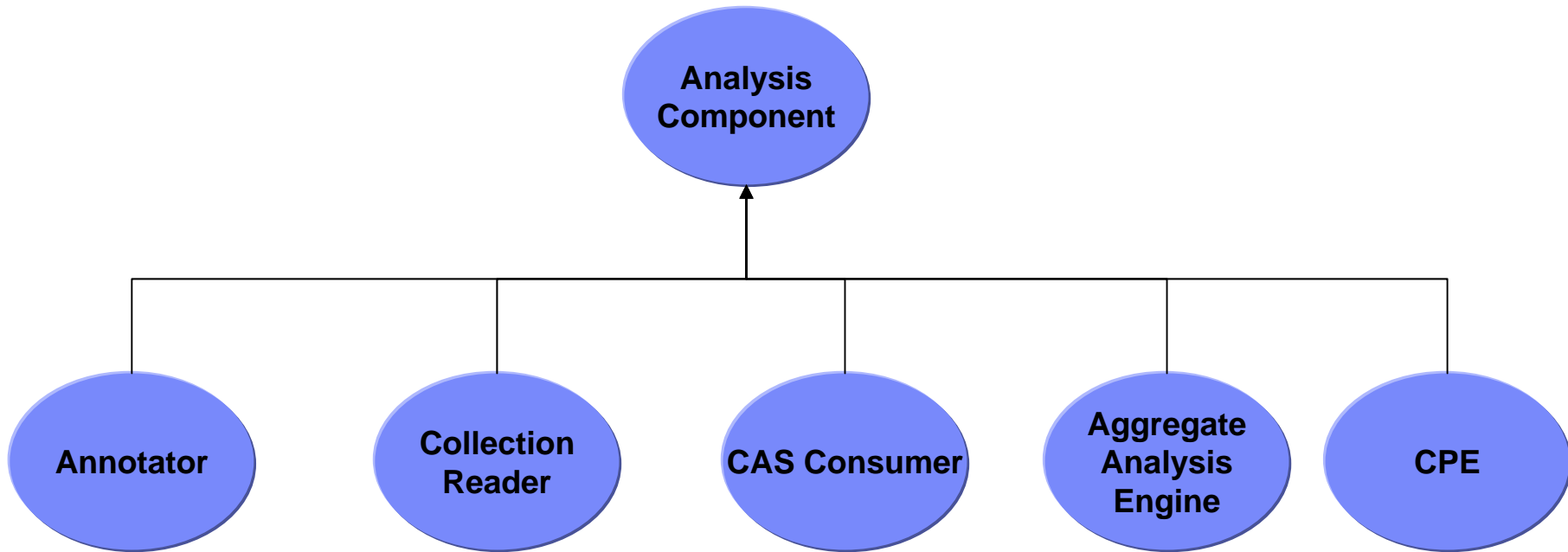
- **Multiple views of an artifact can each support independent sets of attributes**
- **Focus can change from 1 to the other, (e.g., audio to text) to both**
- **Attributes directed to one or more Sofas**
- **Maintain compatibility with existing UIMA text analytic components**

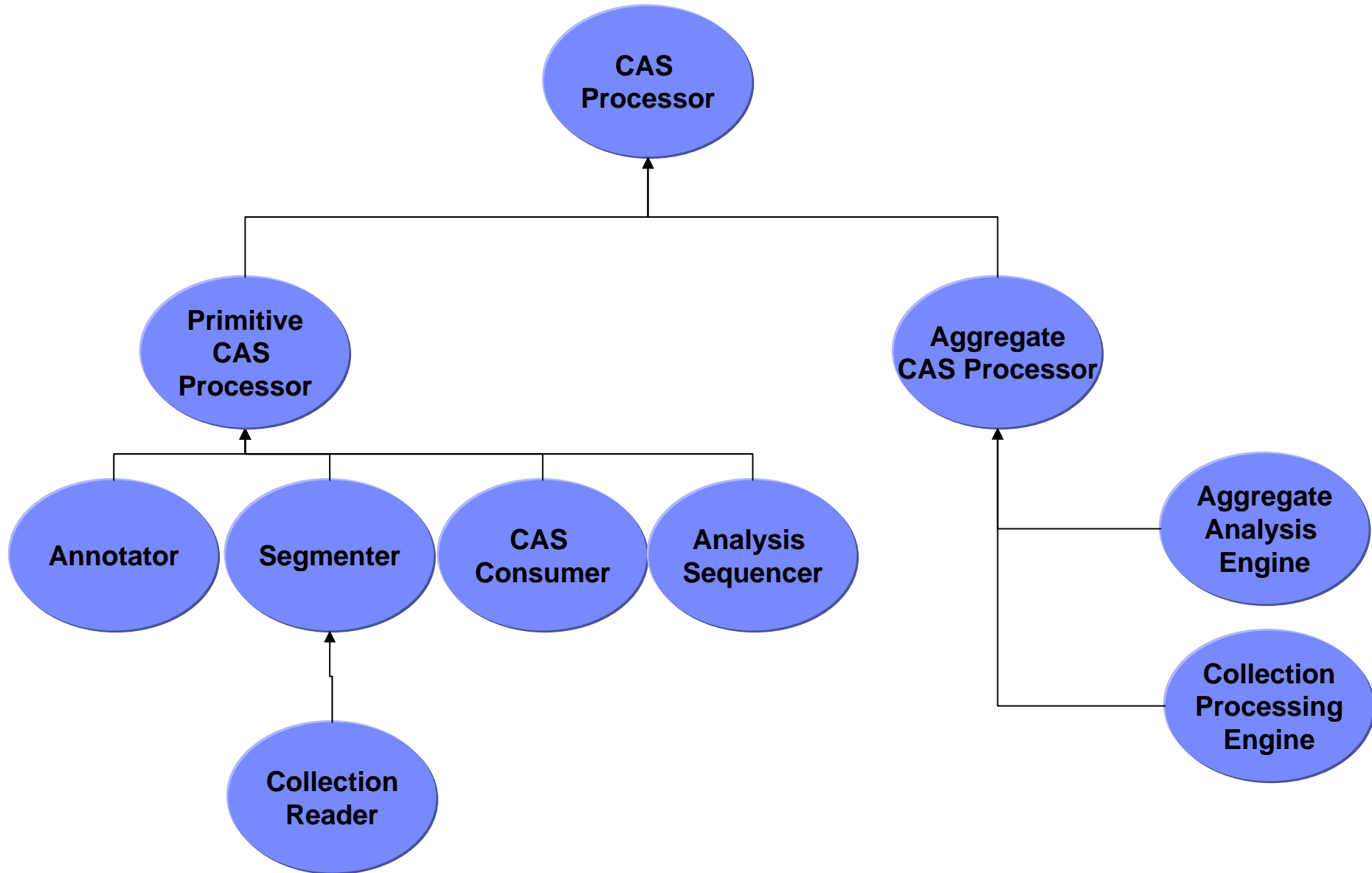


UIMA | IBM Research

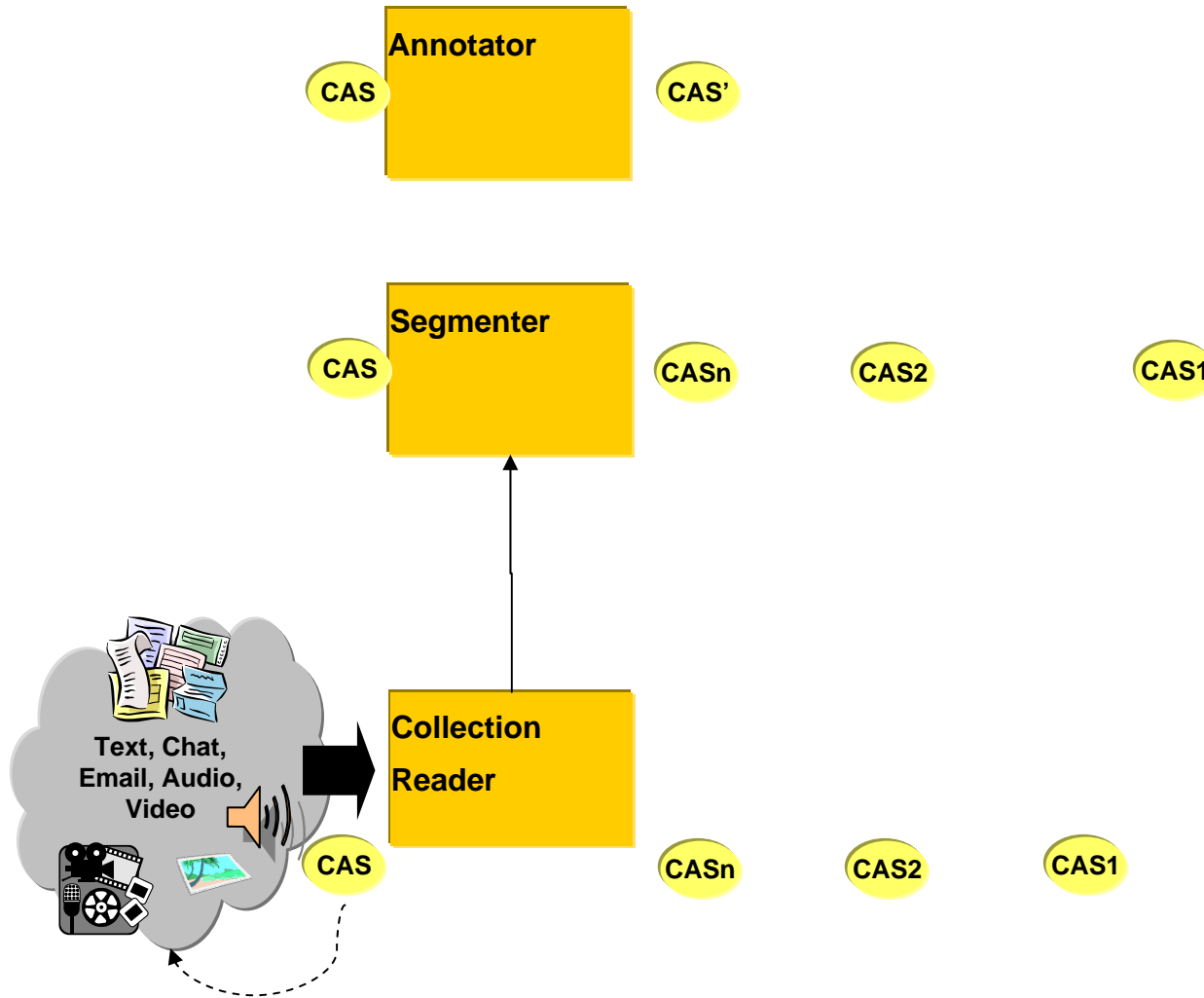
Advanced Architecture Preview

CAS Processor Generalization, Analysis Sequencing, Multiple Sofas, Views and Segmenters, XML in UIMA

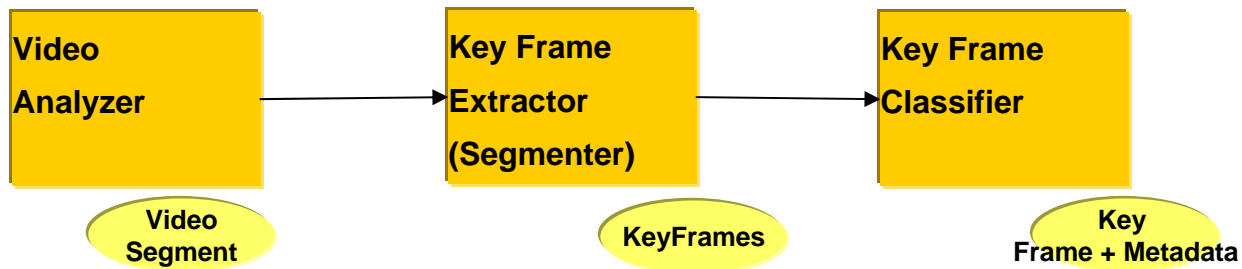




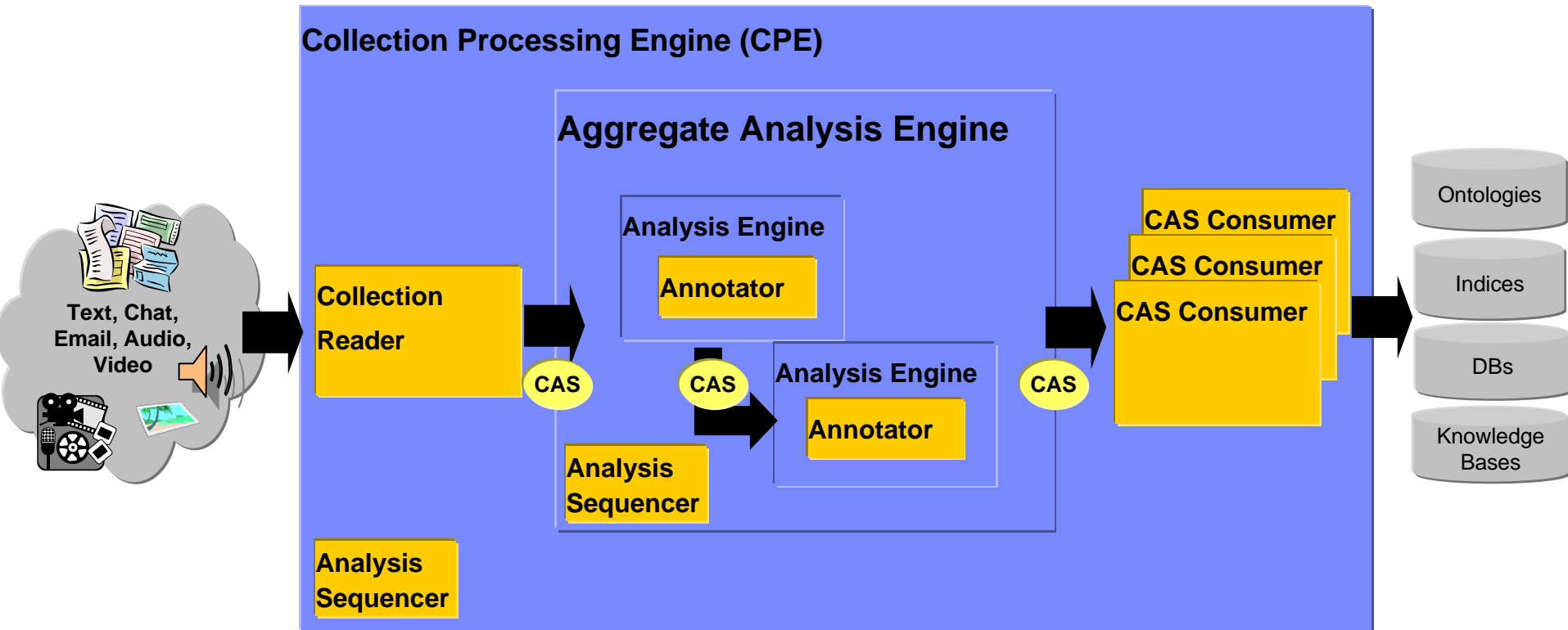
Annotators and Segmenters



Example Inline Segmenter



Flexible, User Definable Workflow

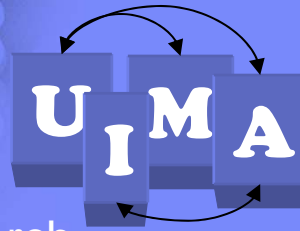


XML in UIMA

- **Inline XML**
 - CAS is a stand-off representation
 - Allows multiple interpretations of artifact
 - Overlapping annotations
 - Immutability of artifact
 - Utilities produce Inline XML to represent text annotations in CAS making assumptions about overlapping annotations
- **XCAS (and soon XMI)**
 - The full CAS can be viewed as an object graph where annotations have refernces into artifact
 - XMI is used to represent full CAS contents as an XML document
- **Component Metadata**
 - XML Document declaratively capture identification, configuration, behavior and deployment information about UIMA components.

Helps to think and plan projects in terms of Development Roles

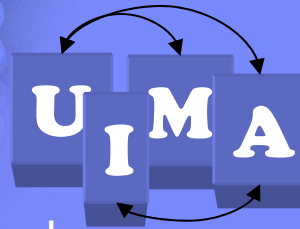
- **Data Modeler**
 - Defines Type Systems
- **Component Developer Roles**
 - Develop Algorithms
 - Implement Architected Component Interfaces
 - e.g., Annotator Interfaces
- **Engine Assembler**
 - Selects Component Implementations
 - Assembles into workflow to create aggregate capability
- **Engine Deployer**
 - Selects deployment model (Collocated, Distributed (e.g., SOAP))
 - Configures Engine Descriptor



UIMA | IBM Research

UIMA Development Tooling and Component Repository

How it works



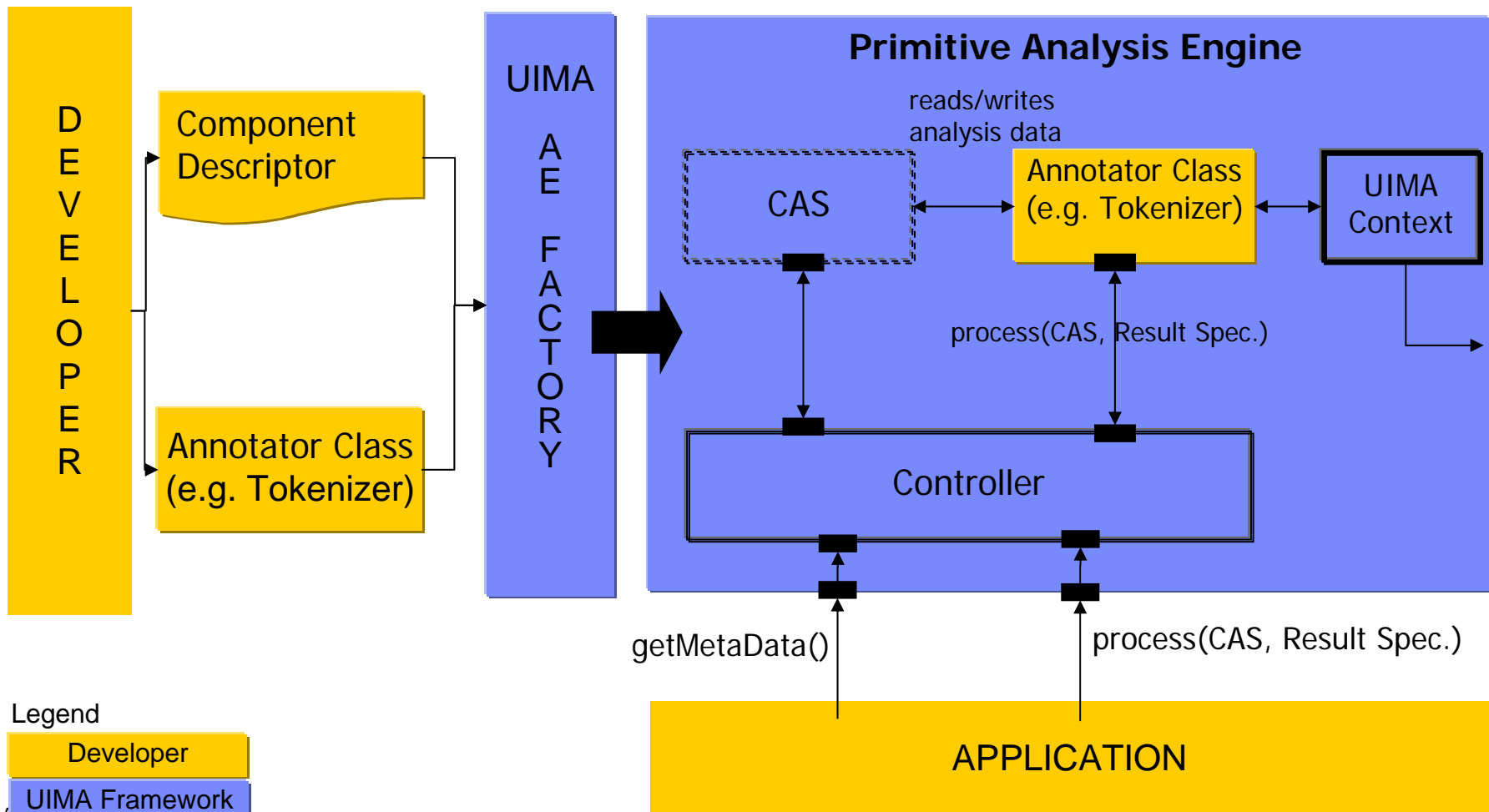
UIMA | IBM Research

UIMA Framework Functions

The plumbing you don't want to build
(over and over and over again)

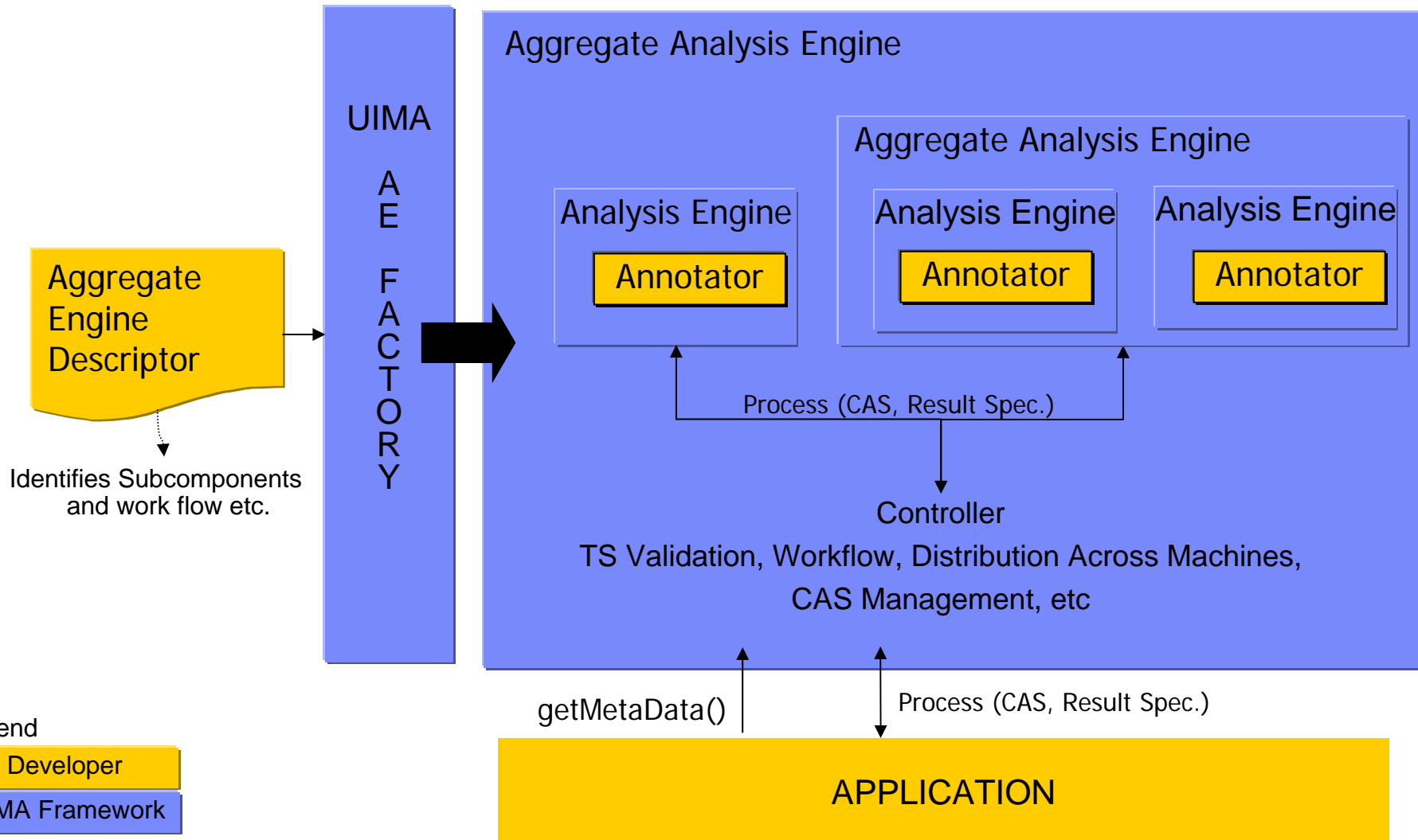
How it works

To create an analysis engine, the developer needs an **Annotator Class** and a **Component Descriptor**. The Framework does the rest.

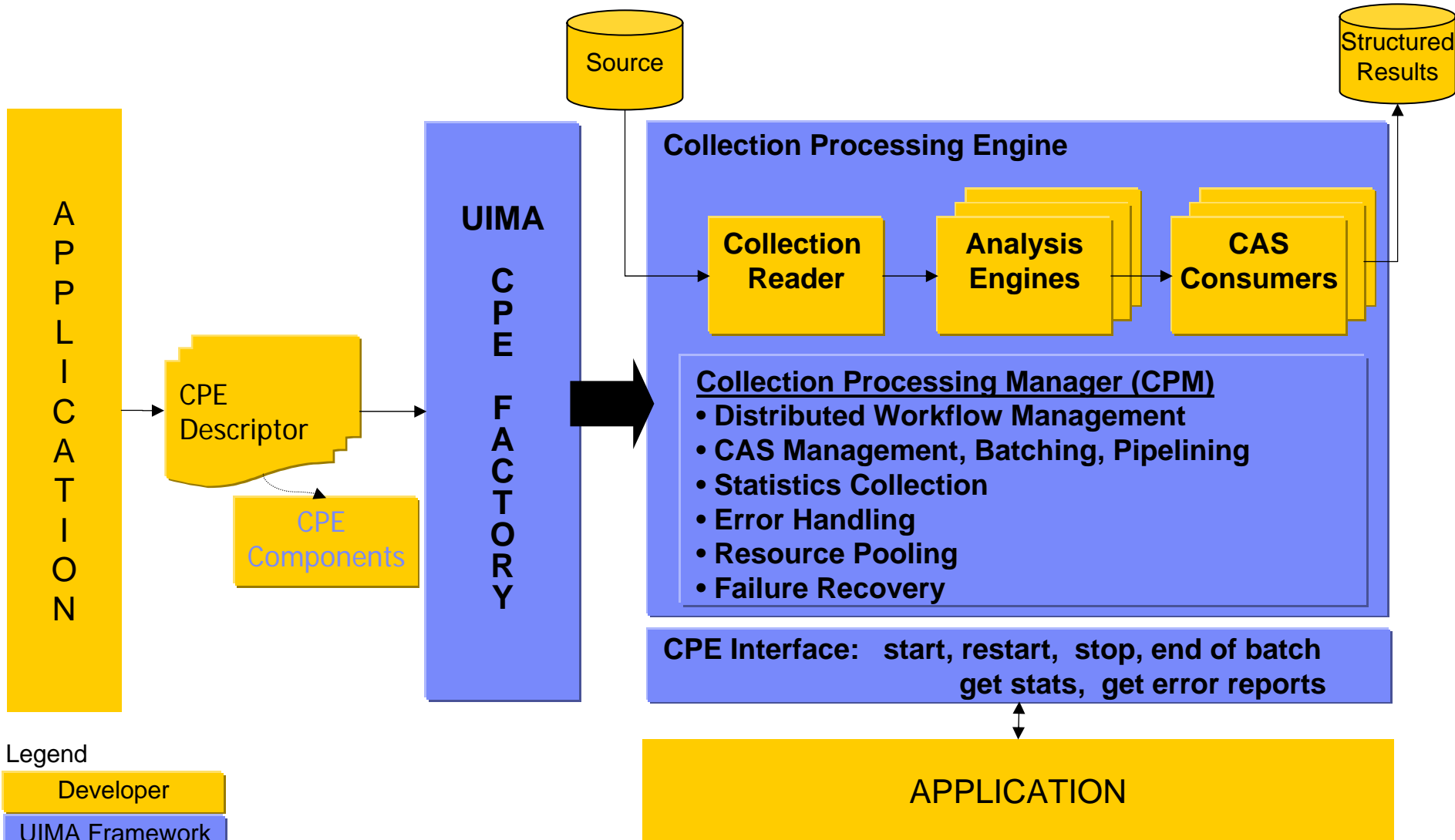


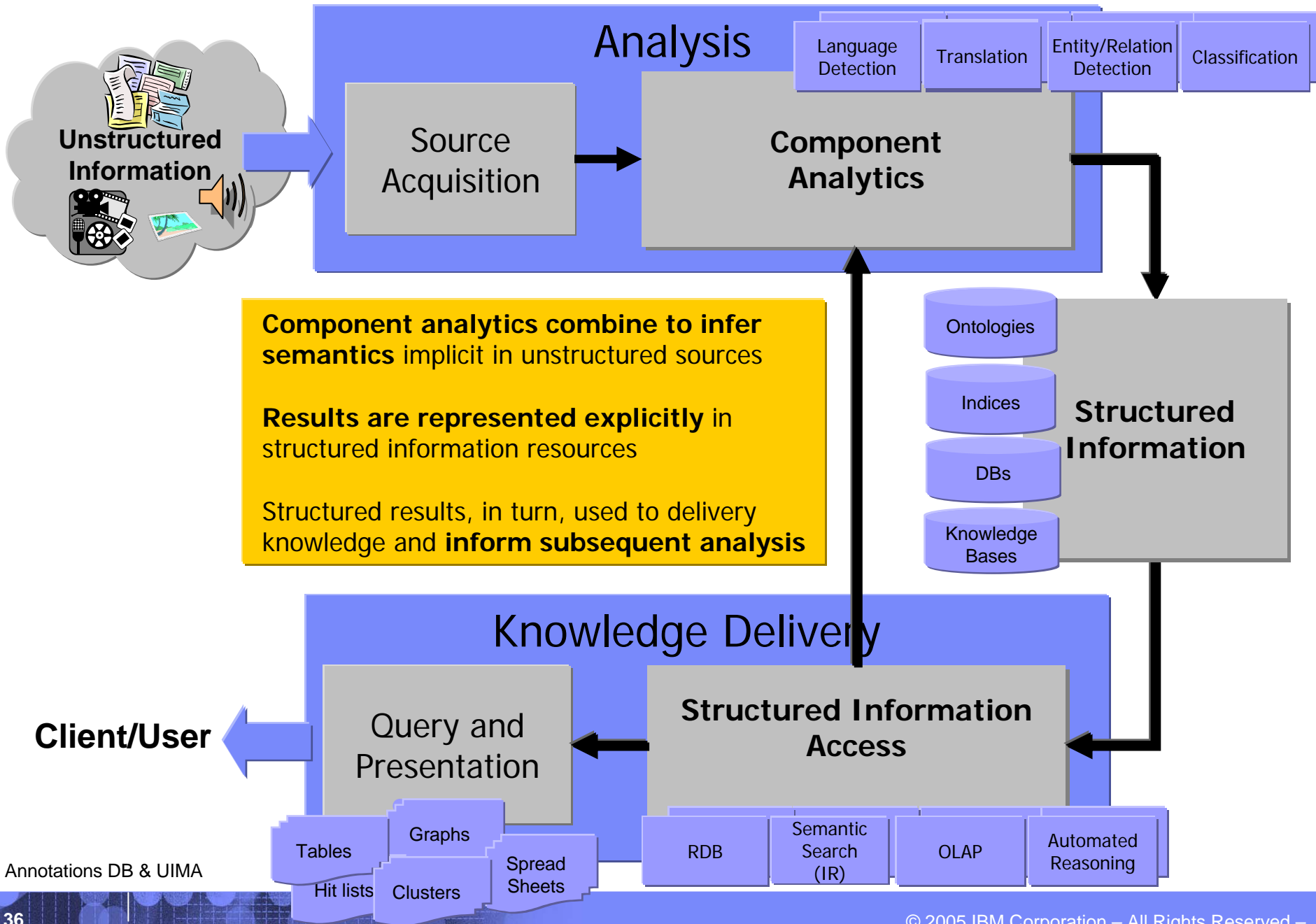
- Legend
- Developer
 - UIMA Framework

Aggregate Analysis Engine



The UIMA Framework provides a CPE Factory and implements scaleable infrastructure for running CPEs in a variety of deployment scenarios.





The **Common Analysis Structure (CAS)** enables Analysis Writers access to the artifact being analyzed and the current analysis results.

