

# At-a-Glance: IBM's UIMA



Ballardvale Research  
Andover, Massachusetts

September 8, 2005

# Summary

---

- UIMA Background and Benefits
- Two Levels of Analysis
  - Document Level
  - Collection Level
- Components
  - Common Analysis System (CAS)
  - Analysis Engine (AE), Annotator
  - Collection Processing Engine (CPE)
  - CAS Consumers
- Recommended Developer Roles

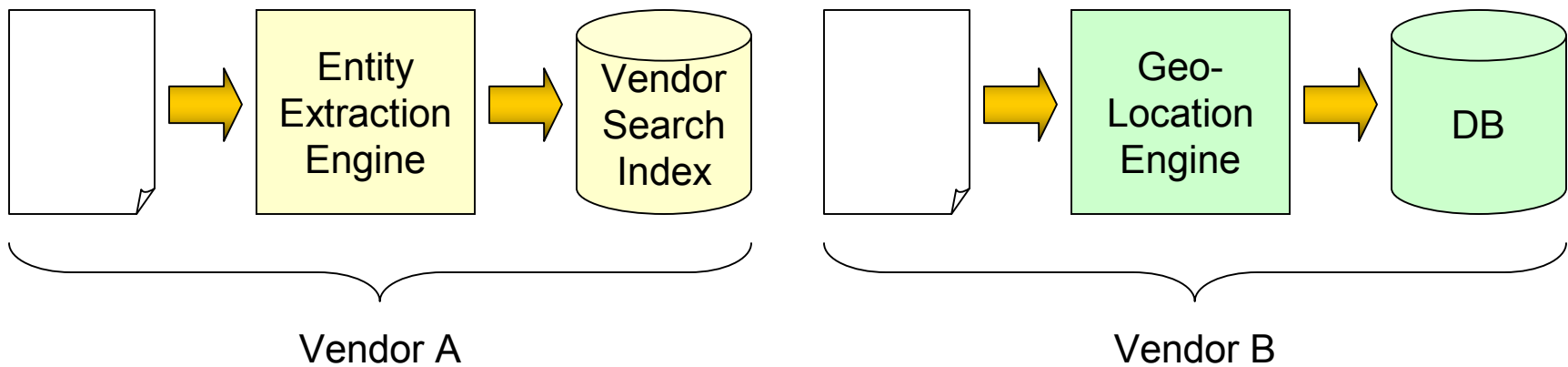
# Background

---

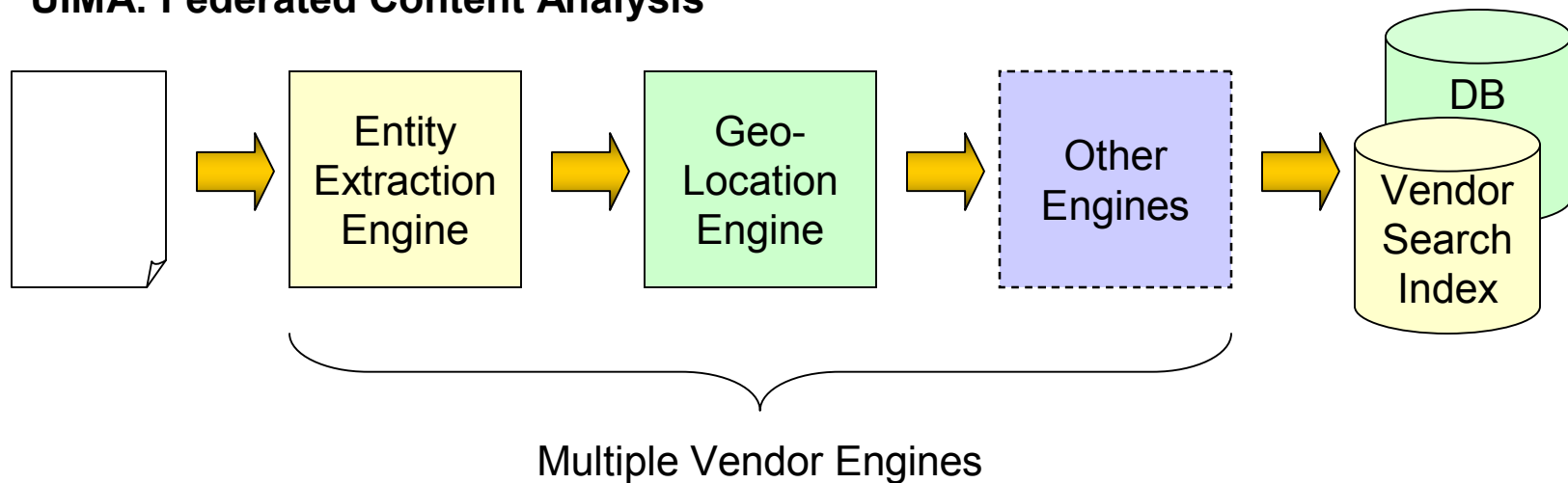
- Short for Unstructured Information Management Architecture
- Is a framework, not a product
- A Java-based SDK is available from IBM
- Will be available to open source community
- Developed over 4 years
  - Funded by DARPA
  - Companies: BBN, MITRE, SAIC
  - Universities: Carnegie Mellon, Columbia, UMass/Amherst

# UIMA: Federated Content Analysis

Today: Separate Content Analysis Applications, Difficult to Integrate



UIMA: Federated Content Analysis



# Benefits by Role

---

- *Developers:* Can string together multiple analysis engines and send the results to search engines or databases
- *Integrators:* Can avoid having to learn different API sets when working on an integration project
- *Companies:* Can leverage analysis technologies from multiple vendors in a single solution

# Two Levels of Analysis

---

- *Document Level* – Analyzes an individual document. Results can include entities (e.g., name, location) and document summaries.
- *Collection Level* – Analyzes an entire collection of documents. Results can include databases and search indexes.

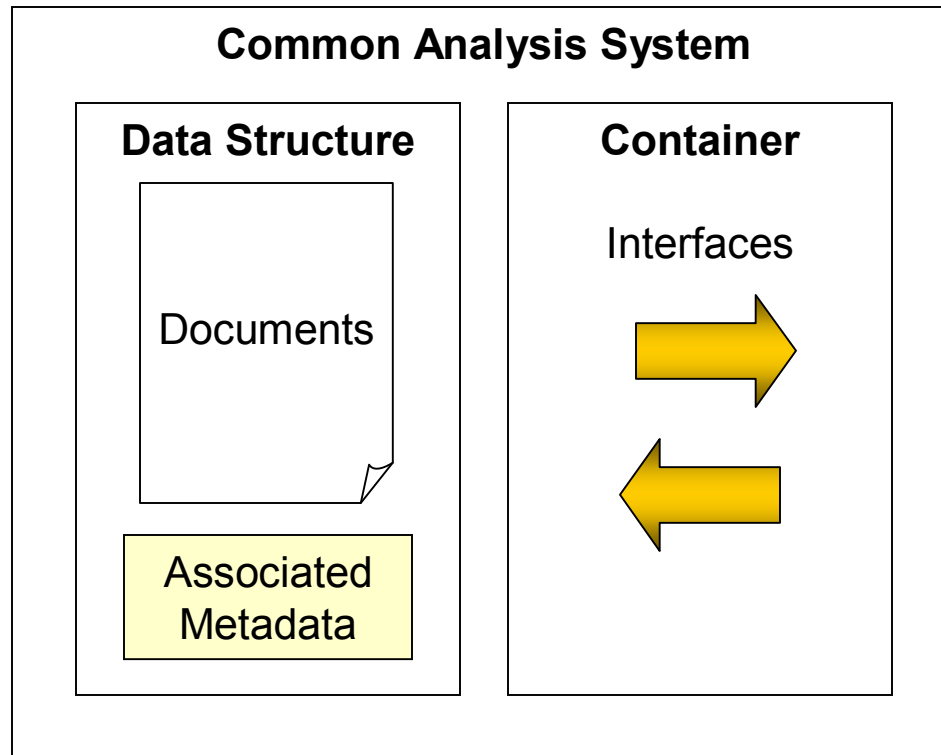
# UIMA Components

---

- Common Analysis System (CAS)
- Analysis Engine (AE)
- Annotator
- Collection Processing Engine (CPE)
- CAS Consumers

# Common Analysis System (CAS)

---

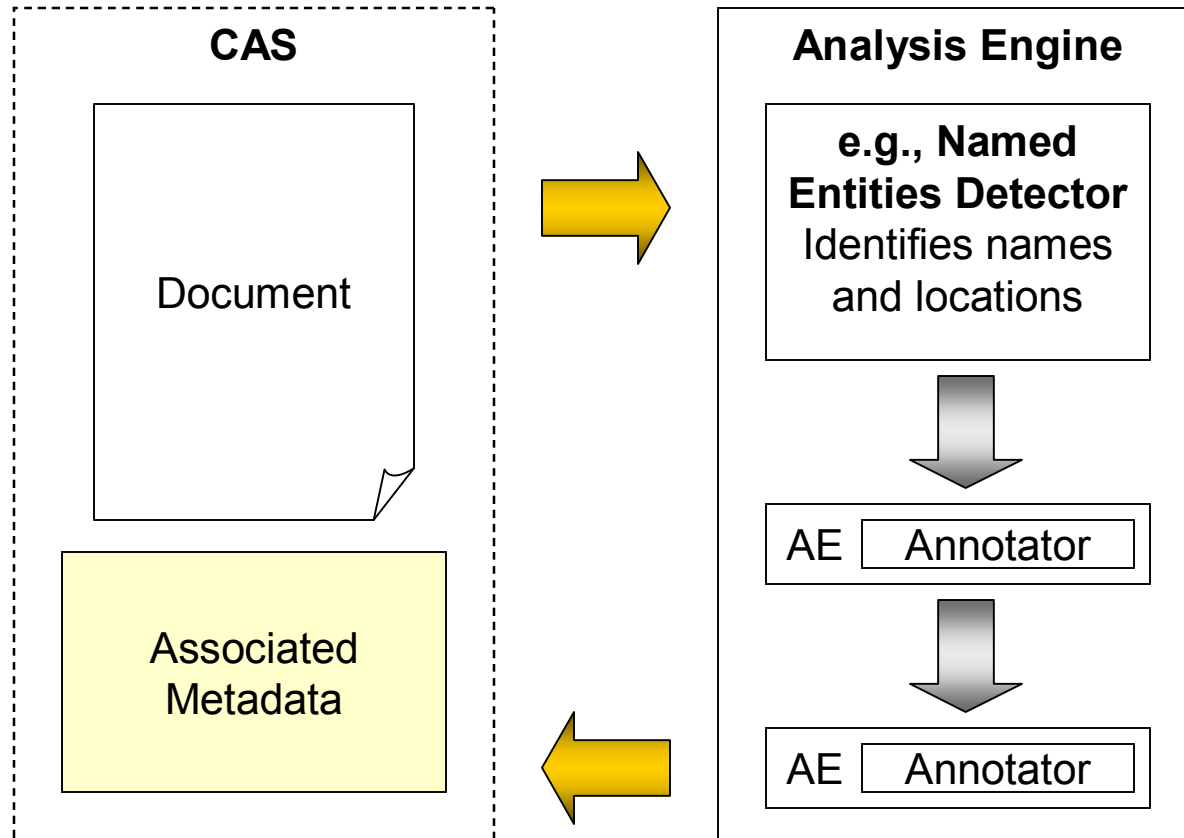


The Common Analysis System is made up of:

- Data Structure -- contains documents and associated metadata
- Container Object -- interfaces that allow indexing, reading, and updating of the data structure



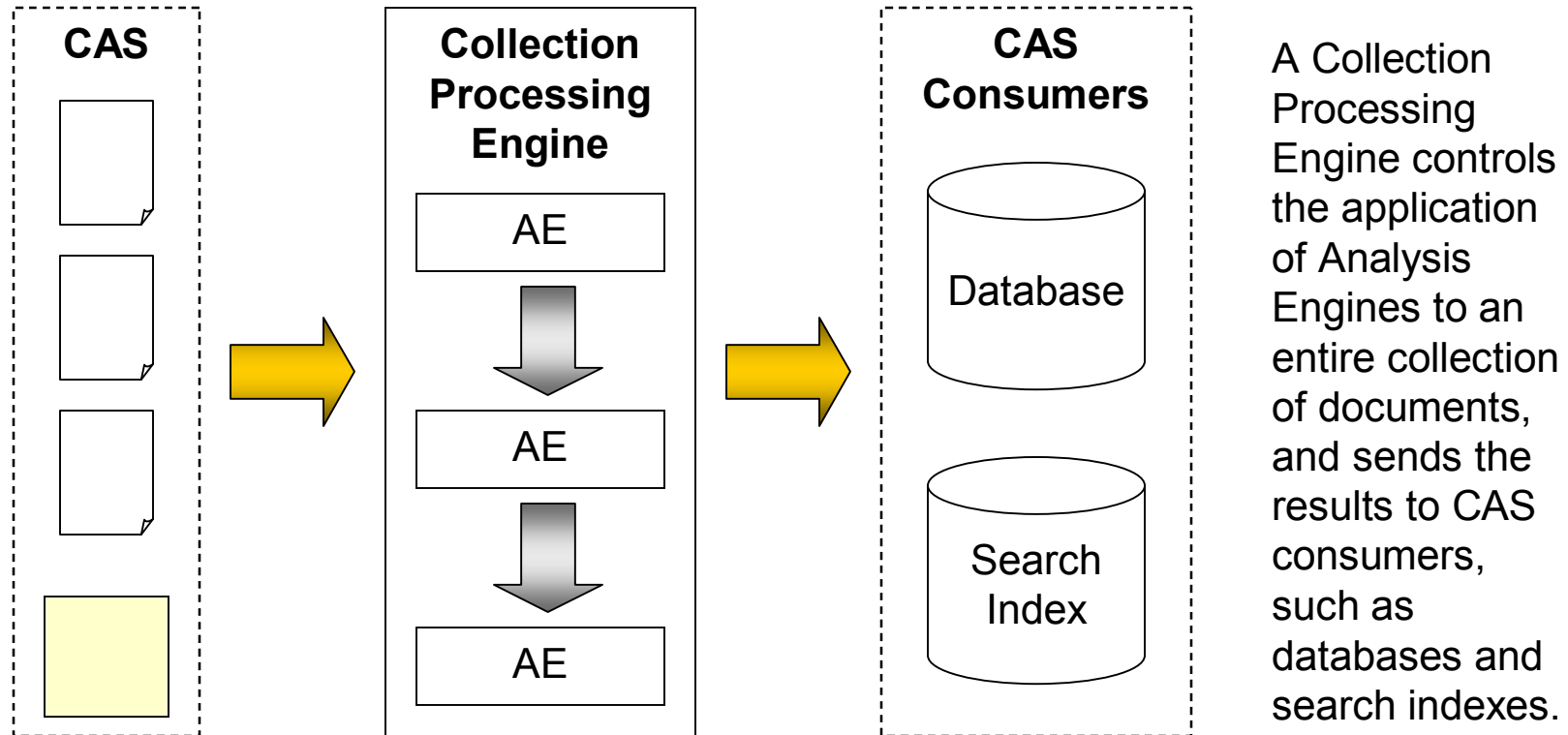
# Analysis Engine (AE)



An Analysis Engine analyzes content via the CAS and generates metadata which is added to the CAS.

The algorithm component of an AE is an annotator; AEs can contain other AEs.

# Collection Processing Engine (CPE)



# Five Development Roles

---

Due to the complexity of the work and the different skill sets required, IBM recommends five development roles in implementing UIMA solutions:

- Annotator Developer
- Analysis Engine Assembler
- Collection Processing Component Developer
- CPE Assembler
- Component Deployer

# This Report & Its Author

---

An [At-a-Glance Deck](#) is designed for today's busy executive: it distills issues into concise points and crisp visuals that lead to quick understanding. You may reuse/copy any portion of the Deck that you find useful, as long as you cite Ballardvale Research as the source. To aid in that reuse, if you would like to request a PowerPoint version, please send an e-mail to [deck@ballardvale.com](mailto:deck@ballardvale.com).

Guy Creese founded Ballardvale Research in 2004 and has been an industry analyst for over 7 years. You may contact him at:

- E-mail: [guy.creese@ballardvale.com](mailto:guy.creese@ballardvale.com)
- Blog: [www.ballardvale.typepad.com](http://www.ballardvale.typepad.com)

