UIMA Workshop

Thilo Götz

4/11/2007

Program

- **11:00 – 11:15** *Workshop kick-off*
- **11:15 – 11:40 Gurevych et al.:** *Darmstadt Knowledge Processing Repository Based on UIMA*
- **11:40-12:05 Gnjatovic et al.:** *Processing dialogue-based data in the UIMA framework*
- **12:20-12:45 Tomanek et al.:** *An UIMA-Based Tool Suite for Semantic Text Processing*
- **12:45-13:10 Hahn et al.:** *An UIMA Annotation Type System for a Generic Text Mining Architecture*
- **13:10-14:00 Lunch**

© 2006 IBM Corporation

Program (ctd.)

- **14:00-14:25 Heyer & Buechler:** *Kookurenzberechnungen mit UIMA und Medusa*
- **14:25-14:50 Blohm et al.:** *Iterative Learning of Relation Patterns for Market Analysis with UIMA*
- **15:00-15:25 Wilcock:** *GATE and UIMA in Language Technology Teaching*
- **15:25-15:50 Gurevych et al.:** *Teaching "Unstructured Information Management: Theory and Applications" to Computational Linguistics Students*

© 2006 IBM Corporation

Program (ctd.)

- **16:00-16:25** **Kaiser & Zavrel:** *A new UIMA Annotator: the Textkernel Information Extraction Engine*
- **16:25-16:50** **Netter & Meyer:** *ICE - Intelligent Content Engineering*
- **17:00-17:45** *Discussion*

© 2006 IBM Corporation

Unstructured Information Management Architecture

Executive Overview

ON DEMAND BUSINESS™

UIMA – Architecture and Framework Implementation

- The Open Software Architecture
  - Common Interface Definitions, Data Representation Schemes, Design Patterns
  - For defining analysis engines and applications that
  - Capture and integrate knowledge from unstructured sources
- The Open-Source Software Framework Implementation
  - Tools, Utilities and Runtime
  - Supports Developer in
    - Implementing architected interfaces,
    - Defining and manipulating Data Models,
    - Conforming to Design Patterns and
    - Integrating and deploying results

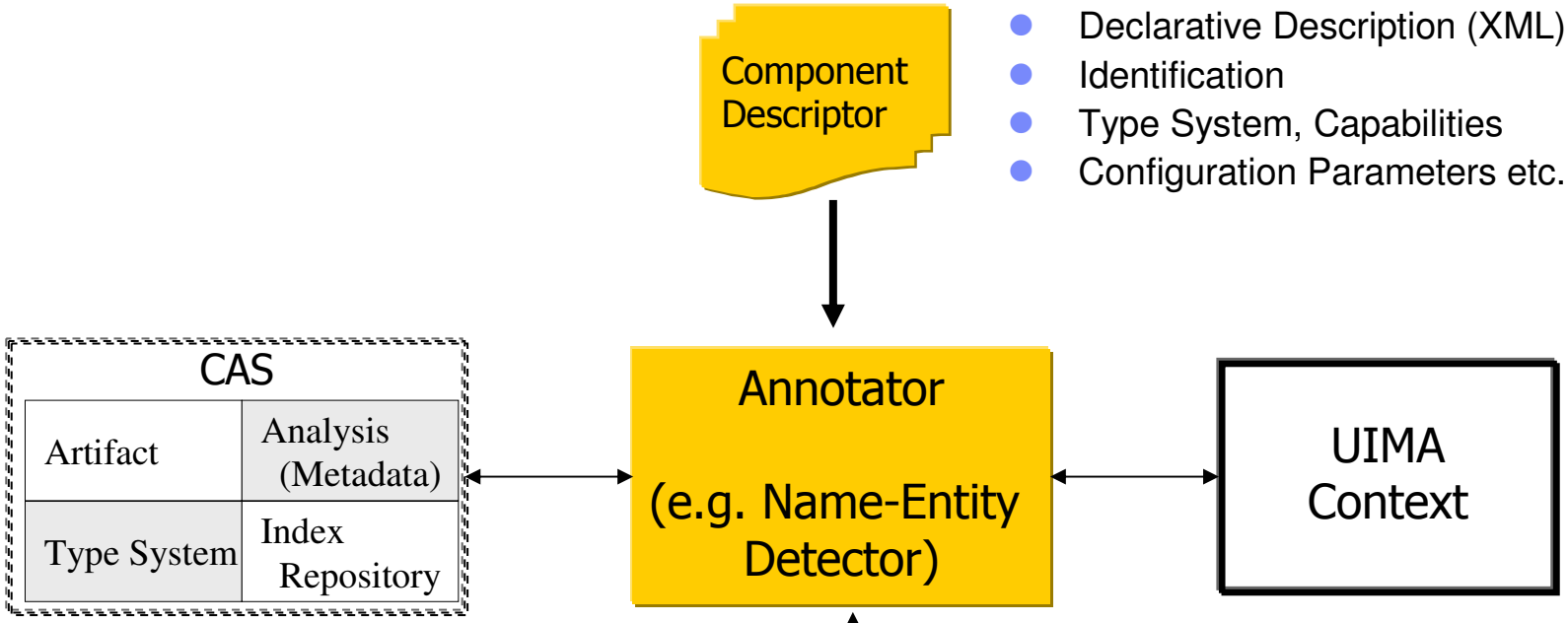
© 2006 IBM Corporation

Focus on the Architecture

- Principal Architectural Commitments
  - Common representation scheme
  - Common component engine interfaces (task and domain-independent)
  - Common component metadata
  - Embeddable (can be layered on top of system middleware)
- Independent of but interoperable with application specific
  - Data models
  - Modality (e.g., Text, Speech, Chat, Image, Video)
  - Algorithms (e.g., Rule-based, Statistical)
  - Language-level or domain-level tooling
  - Systems Level Middleware: Workflow engines, Transports (e.g., SOAP, Vinci)
  - Back-end Systems (DBs, Search Engines, Knowledgebases)

© 2006 IBM Corporation

The Basic UIMA Component Interfaces



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

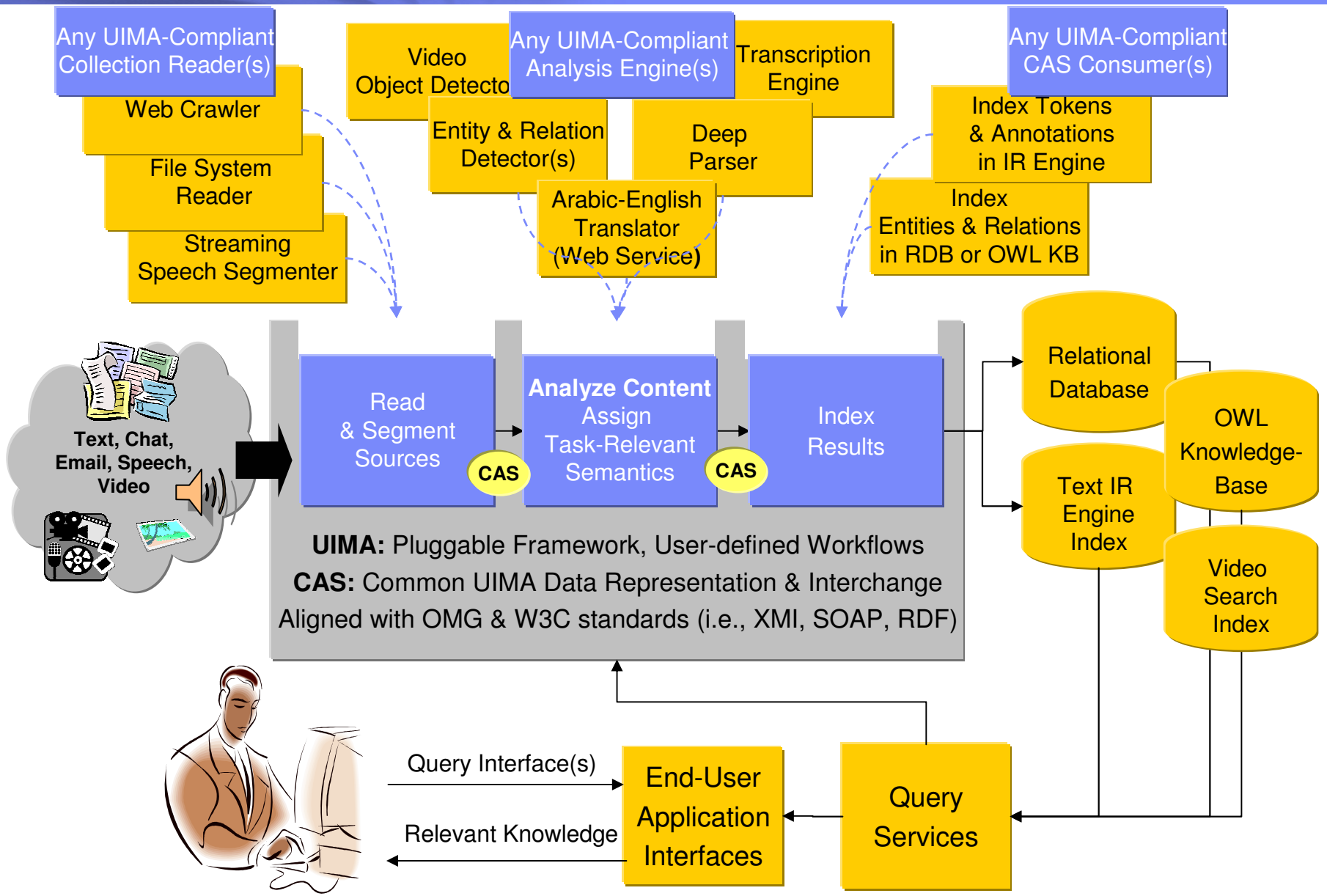
Analysis Algorithm

Analysis Code goes Here

Interface: CAS in/CAS Out

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

© 2006 IBM Corporation



© 2006 IBM Corporation

Apache UIMA

The Apache Software Foundation
http://www.apache.org/

- UIMA accepted as Apache Incubator project 10/06
- Future development of UIMA will happen on Apache as open source development
- IBM will continue to support UIMA development, but cedes sole control
- http://incubator.apache.org/uima/

© 2006 IBM Corporation

UIMA TC at OASIS

Advancing E-Business Standards Since 1993

- OASIS -- Organization for the Advancement of Structured Information Standards
- OASIS has established a technical committee for the standardization of UIMA
- The UIMA TC is open to participation to all OASIS members
- http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=uima

© 2006 IBM Corporation

While you're here…

© 2006 IBM Corporation