UIMA Workshop

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
Program (ctd.)

- 14:00-14:25  Heyer & Buechler: *Kookurenberechnungen 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*

Program (ctd.)


- 16:25-16:50 Netter & Meyer: ICE - Intelligent Content Engineering

- 17:00-17:45 Discussion
Unstructured Information Management Architecture

Executive Overview
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
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)
The Basic UIMA Component Interfaces

- Declarative Description (XML)
- Identification
- Type System, Capabilities
- Configuration Parameters etc.

CAS

<table>
<thead>
<tr>
<th>Artifact</th>
<th>Analysis (Metadata)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type System</td>
<td>Index Repository</td>
</tr>
</tbody>
</table>

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

Component Descriptor

- Component Descriptor

Annotator

(e.g. Name-Entity Detector)

- Analysis Algorithm
- Analysis Code goes Here

UIMA Context

- Manages access to shared external resources
- Allowing local APIs to access resources shared by cooperating engines
**UIMA:** Pluggable Framework, User-defined Workflows
**CAS:** Common UIMA Data Representation & Interchange
Aligned with OMG & W3C standards (i.e., XMI, SOAP, RDF)

- Read & Segment Sources
- Analyze Content: Assign Task-Relevant Semantics
- Index Results

- End-User Application Interfaces
- Query Services
- Relevant Knowledge
- Query Interface(s)
- Text, Chat, Email, Speech, Video

- Web Crawler
- File System Reader
- Streaming Speech Segmenter
- Video Object Detector
- Entity & Relation Detector(s)
- Deep Parser
- Arabic-English Translator (Web Service)
- Index Tokens & Annotations in IR Engine
- Index Entities & Relations in RDB or OWL KB

- Any UIMA-Compliant Collection Reader(s)
- Any UIMA-Compliant Analysis Engine(s)
- Any UIMA-Compliant Transcription Engine
- Any UIMA-Compliant CAS Consumer(s)
- Relational Database
- OWL Knowledge-Base
- Text IR Engine Index
- Video Search Index

- CAS
Apache UIMA

- 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/
UIMA TC at OASIS

- 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
While you’re here…