

January 10th UIMA OASIS Teleconference
Meeting Minutes

Roster Check

Present:

=====

- Eric Nyberg
- Kevin Bair
- Dave Ferrucci
- Thilo Goetz
- Thomas Hampp-Bahnmueller
- Adam Lally
- Karin Verspoor
- Antonio San Fillipo
- Jonathan Michel
- Ian Roberts
- Pascal Coupet

Absent:

=====

- Clifford Thompson
- Christopher Chute
- Adrian Miley
- James Clark
- Robin Cover

Other Participants (not on roster)

=====

- Hamish Cunningham (Sheffield) - waiting to be approved
- Alex Rankov - waiting to be approved
- Yoshino Kano & Nguyen Ngan (University of Tokyo) - waiting to be approved

Put minutes into proper place on the web site. Want the minutes to be publicly available, so we can let the Apache community know about what's going on? There were no objections.

Start discussing what's in the white paper. The points in the paper are illustrated with examples from the Apache framework implementation; these are treated separately.

Reviewed Last Calls Minutes

Reviewed Agenda for Today

Discuss Design Goals (Section 3)

1. Data representation. Reaffirm that the original artifact is immutable. But note that you can use additional sofas to handle variations on the original document. What are the integrity constraints on operations on the data? The spec should state relevant assumptions clearly for example, the immutability of the artifact data.

Consider rephrasing the definition of an artifact in 2.1.1 to express the idea that "An artifact is a collection of its representations".

How will we characterize the "representational power" of the CAS?

For example, can we support cross-CAS references?

Will CASes always be accessed as a stream? What about sets of CASes exchanged between interoperable components? How can we refer to CASes? etc.

2. Other comments on design goals

Current text doesn't say much about deployment environment (beyond discovery and composition in point 3). Error handling and monitoring must be handled in a consistent way across all annotators, in order for composed services to work. How can we extend the responsibilities of each annotator to handle this?

The APIs for components may vary depending on whether they are being run in batch mode on a cluster, or over the web for interactive composition by remote users.

What service protocols should the spec consider? What overheads do they add to high-volume processing? In-process API bindings part of the spec? (e.g. we can offer different compliance points for developers to adhere to or not as they wish).

For example an in-process API binding for Java would not be difficult to derive from the existing SOAP interfaces, using an XML binding technology such as Sun's JAXB. APACHE UIMA already submits Java bindings. Do we adopt that as the strawman? Should this be included as part of the specification?

3. Are we going to exchange metadata models, schemata, etc.? (assume the answer is 'yes', but should be clearer in the design goals). "A schema that can represent what a component can do with the documents". How can we improve the current descriptor languages to facilitate discovery and composition?

4. How does the question of schema interoperation relate to integrating data together from multiple components? (type system mapping, etc.) (see 7.1 of the spec for proposed positioning on this).

Perhaps we can call out a type of process/component that does mapping, which can be included in a workflow. We need not address into the details of a particular mapping language or implementation.

- * Discuss Use Cases (Section 4)

(not started)

- * Discuss Structure of the Spec (Section 5)

(not started)