

D-Spin Meeting April 2009 Tübingen - Organization -

Additional information can be found on: http://www.sfs.uni-tuebingen.de/dspin/meetingApril09.shtml

Participants

Tübingen:	
Erhard Hinrichs	
Marie Hinrichs	
Kathrin Beck	
Iris Vogel	
Yana Panchenko	
Alexander Kislev	
Thomas Zastrow	
Stuttgart:	
Helmut Schmid	We
Fabienne Fritzinger	We
Max Kisselew	We
Leipzig:	
Uwe Quasthoff	Tu
Marco Büchler	Tu
Mannheim:	
Helge Krause (beide Teile)	We
Jens Rinne (nur Teil 2)	We
(101.10.10.10.10.10.10.10.10.10.10.10.10.	
Berlin:	
Jörg Didakowski	We
Lothar Lemnitzer	We
Clarin:	
Peter Wittenburg	Tu
Milan Agatonovic	Tu
Marc Kemp-Snijders	Tu
Radu Ion	Tu
Santi	Tu
Victor	Tu
Shef	Tu

Agenda

Monday, 4/27

(For those arriving early, lunch at 12:00)

1:00

Start of first part - Framework for WS: metadata, provenance, registries, wrappers,

3:00 - 3:30 Coffee Break

until 6:00

Continuation part one

7:00

Dinner at the Flavours of India

Tuesday, 28/4

9:15

Continuation of framework discussion

12:00 - 1:00

End of part one, lunch break

3:00 - 3:30

Coffee Break

until 6:00

Continuation part two

Mittwoch, 29/4

9:15

Continuation part two

12:00

End of the meeting

Abstract First Part

Peter Wittenburg

Web Services combined in processing chains exchange more or less complex messages which they do based on formal interface descriptions and they create enrichments in form of new resources or annotations on existing resources. To be able to do this web services need to be embedded in a framework which we would like to describe with a few statements:

- project files contain detailed descriptions about workflow and information flow
- an instance needs to keep control about the state of processing in particular in case of asynchronous processes
- resources and services are described by metadata in many of the required characteristics which are contained in registries, new resources need to be described by metadata and be registered
- provenance files memorize all relevant information to allow restarts and replications
- wrappers need to be provided so that developers can focus on the algorithmic solutions; an alternative solution can be used by defining a CLARIN Service Bus
- user start such processes via applications that will carry out an authentication step and mechanisms need to be in place that pass through the user credentials - etc

In the previous CLARIN WG2.6 discussions all these issues were raised resulting in the draft requirements document which is based on the experience and knowledge of a number of European experts in this area. Also in the collaboration between L-T-S within D-SPIN similar issues were discussed. This first day wants to present the ideas for a framework as described in the CLARIN WG2.6 document and wants to discuss them in more detail. So this part of the workshop does not want to discuss the complexity of the I/E problem, but it will raise the issue of what kind of data structures we will need to describe the complexity in MD elements for example.

All issues are still under debate and the discussion will keep us busy until the end of the preparatory phase at least. The workshop will be organized without prepared talks with the exception of the introduction, but will discuss ideas, comments, architectural aspects as they will come up.

Themensammlung zweiter Teil

Volker Boehlke, Marie Hinrichs, Helmut Schmid, Thomas Zastrow

Architektur

- Gibt es Inkompatibilitäten zur CLARIN-Architektur?
- Ideen zur Verbesserung der Architektur
- Diskussion einzelner Aspekte der Architektur
 - o Ist ein einzelnes XML-Dokument als Webservice-Interface generell genug?
 - Sind die Dokumentattribute in Verbindung mit IO-Spezifikationen der Webservices ein ausreichender Mechanismus für die Prüfung der Korrektheit von Webservice-Verkettungen?
 - Wie werden die Dokumentattribute technisch definiert?
- Genaues festschreiben der Prototyp Architektur:
 - Wie sieht ein zentrales Repository aus? Welche Daten gehören in welcher Form hinein? Was kann hier an bereits vorhandener Software evtl. wiederverwendet werden?
 - Welche Aufgabenteilung besteht zwischen "Toolwrapper-Service" und dem allgemeinen Webservice? Welche dieser Probleme wollen/sollen wir im Prototypen addressieren? Die Authentifizierung ist hier kein sehr lohnendes Thema, da es hier in Clarin schon konkrete Ideen und Experten gibt. Lastverteilung/Ticketsystem sowie die Frage, ob auch auf dieser Ebene Funktionalität rund um die Metadaten vorhanden sein sollte, sind hier interessante und wichtige Punkte.

Datenformat(e)

- Was sollte/muss noch am Toolwrapper-Format (eher allgemein; konkret am Beispiel des TextCorpus Formates) geändert werden?!
- Wie kann der "Header" formaler ausgedrückt werden; postagging="STTS" ist hier vielleicht nicht der beste Weg. Hier wäre es sinnvoll, für diese Attribute und ihre gültigen Werte (eine saubere, standardkonforme Spezifikation) entsprechende Kategorien anzulegen etc. Hier stellt isoCAT einen interessanten Ansatz dar.
- Welche Schemasprache(n) soll(en) verwendet werden (W3C, RelaxNG, DTD, Schematron) und für welche Zwecke?
- Welche externen Standards (ISOcat, LAF, UIMA, ...) können wie konkret genutzt werden?
- Wie wird die Definition der XML-Formate organisiert? Wer definiert die Formate?
- Wäre es sinnvoll, die XML-Formate für Metadaten und Inhaltsdaten unabhängig voneinander zu entwickeln?

Clients

- Welche Bedeutung hat die Entwicklung von Clients in D-Spin/CLARIN?
- Welche Softwaretechniken eignen sich für die Implementierung solcher Clients?
- Wer entwickelt Clients?
- Viele Tools koennen parametrisiert werden, und Benutzer koennen die Parameter selbst bestimmen wollen, oder aber Default-Konfigurationen verwenden. Wie wird Parametrisierbarkeit sichergestellt und in evtl. Clients, Formaten und Prozessketten abgebildet?

Evaluierung

• Es ist wichtig, sich langfristig Gedanken zu einer möglichen Evaluierung zu machen. Welche Aspekte wollen wir testen (Performance, Stabilität, Kosten etc.)? Wie wollen wir diese testen (Beispiele für Prozessketten; konkrete Probleme...)?

Integration weiterer Webservices

- Welche Webservices existieren ausser den bereits integrierten in D-Spin?
- Wie können diese in die bereits existierende Tool-Chain eingebunden werden?
- Sollte eine Integration nicht möglich sein: Welche weiteren Tool-Chain(s) liessen sich auf Basis bestehender Webservices entwickeln?
- Welche weiteren Webservices liessen sich erstellen bzw. sind bereits angedacht?