Mini-Biography

Since 2000 Uwe Radetzki is research staff member of the Department of Computer Science III at the University of Bonn. He joined the laboratory of scientific information systems (GIS, Life Science) and software technology. His topics of research are interoperable and adaptable information systems in science applications. In the area of software technology Uwe Radetzki is interested in middleware and service-oriented architectures, especially in Web Services. (www.cs.uni-bonn.de/~ur)

Prof. Dr. Ulf Leser is head of the Knowledge Management Group in Bioinformatics at the Humboldt-Universität Berlin since 2002. Previously he was project manager at PSI AG. He received his PhD in computer science in 2000 at the Technical University Berlin for his work on query optimization in information integration. His research topics include management of biomedical databases, information integration, and text mining for the life sciences. (www.informatik.hu-berlin.de/~leser)

Since 1990 Prof. Dr. Armin B. Cremers is Director of the Department of Computer Science III at the University of Bonn. His fields of research encompass software technology, information systems, and artificial intelligence. He leads research groups working on middleware, P2P, Grid, and component technology. A. B. Cremers is Dean of the Faculty for Mathematics and Natural Sciences at the University of Bonn. (www.cs.uni-bonn.de/~abc)

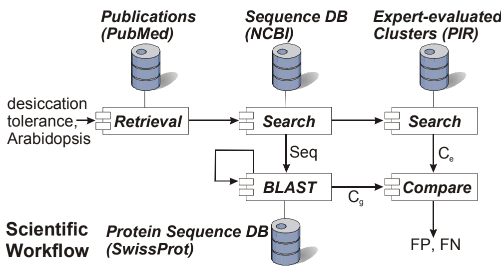
IRIS: A MEDIATOR-BASED APPROACH ACHIEVING INTEROPERABILITY OF WEB SERVICES IN LIFE SCIENCE APPLICATIONS

Uwe Radetzki[1], Ulf Leser[2], Armin B. Cremers[1]

[1]*Department of Computer Science III, University of Bonn, D-53117 Bonn, Germany, {ur,abc}@iai.uni-bonn.de*
[2]*Department of Computer Science, Humboldt-Universität Berlin, D-10099 Berlin, Germany, leser@informatik.hu-berlin.de*

Web Services are loosely coupled software components that are published, located, and invoked across the web. The growing number of web services raises new challenges, namely the discovery, composition, and integration of services required for a given task. Especially in life science applications the scenery is changing rapidly, with new services emerging and others being no longer maintained. For instance, the Molecular Biology Database Collection (MBDC) has doubled the number of registered data and method providers in less than four years [2], which makes the development of stable applications using external web services a difficult task.

On the other hand, biologists need several information services and analyzing tools in their daily work. Due to heterogeneity, especially in semantic aspects, the current automatic interoperation of services still requires high user efforts and constant, error-prone adaptation to changes. Most distributed query processing systems assume static data sources and a centralized mediated schema and cannot cope with such a dynamic environment [8].

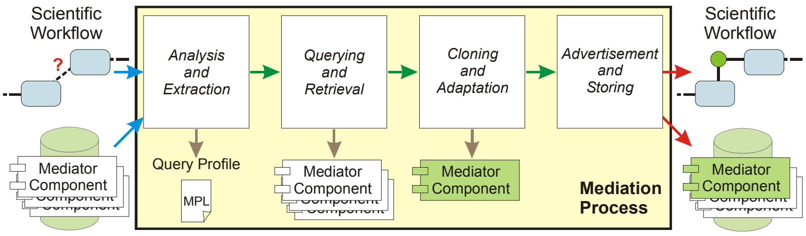


In IRIS we address the issue of service interoperability by a semiautomatic mediation process, where transformation units are identified and inserted between web services to form a desired application process. These transformation units are called mediators and they are realized as fully-fledged software components [7] that implement the necessary adaptation facilities.

	service	mediator
granularity	high	low
usage	explicit	implicit
context	scientific workflow	mediation process
solve	scientific problems in scientific processes	mediation problem (service interoperability)
example query	retrieve all homologous proteins related with desiccation tolerance in Arabidopsis	retrieve transformation units that convert AGAVE information of service A into BSML information of serivce B
example	BLAST, Text-Retrieval	AGAVE-to-BSML

There is an important difference between mediators and services. Technically there are identical, as both are realized and described by Web Services techniques, mainly WSDL. The differences mainly appear in granularity and usage. Nevertheless, this is not a strict separation, because sometimes a service can be used explicitly within a service composition, whereas sometimes it is used implicitly in a mediation process. For instance, assume a service which calculates a protein sequence based on a RNA sequence. This service can be explicitly used in a user-defined scientific workflow by including the appropriate WSDL description in the specification. In another scientific workflow a user might want to combine a retrieval service returning gene sequences with a BLASTP service requiring protein sequences. In this scenario, IRIS will automatically discover and integrate this protein calculation service. Thus the usage of this service is implicit and not directly part of the scientific workflow. However, in general the granularity of services and mediators is different. Services have a higher granularity solving scientific problems in scientific processes, while

mediators have a finer granularity solving interoperability problems between services.
The whole mediation process consists of four independent subtasks: query generation, discovery, adaptation and storage of mediators. For discovery and description of mediator capabilities we have developed the Mediator Profile Language (MPL). MPL is developed as a top-level ontology using the Web Ontology Language (OWL). Every mediator has to be associated to a MPL-compliant mediator profile. Thus it is possible to inspect the capabilities of a previously unknown mediator even at runtime. Queries for mediators are specified as a MPL profile, too. Such query profiles are generated automatically by analyzing the WSDL documents of services within a service composition. The user, can enhance the profile, if desired.



Discovering suitable mediators, services, or agents (called matchmaking) is a complex problem. Several approaches have been discussed using clustering algorithms, information retrieval techniques, linguistic analysis, signature matching, or ontologies [1,3,6].
The discovery component of IRIS provides advanced signature matching and concept matching, in such a way that they not only support subsumption of data types and concepts, but also identify new, previously nonexistent mediator compositions [4]. The IRIS matchmaking uses linguistic methods to allow for fuzzy matching of similar mediators. Because of the complexity of these algorithms, we implemented several filters to reduce the search space before the matchmaking starts.
The adaptation task within the mediation process has several facets: selection of suitable mediator components, defining thresholds for ranking lists, changing the behavior of a mediator by modifying mediator's properties, modifying newly created or already existing mediator compositions, and creating new mediator components. The creation of new mediators can sometimes be done automatically (e.g. by schema matching approaches [5]).
Finally we will take a closer look on future directions of IRIS. First of all, the development of benchmarks for evaluating precision and recall of the implemented discovery algorithms and to compare it with other approaches is an important goal.
With each iteration of the mediation process, the knowledgebase of suitable mediators grows, because newly created or adapted mediators are stored in the system. Hence, we plan to develop tools for managing this evolution of the knowledgebase, i.e. supporting a process manager in the reduction and aggregation of mediators and mediator compositions, respectively.
Finally, we plan to adopt the core platform of IRIS to Grid and P2P technologies. That will allow the creation of virtual organizations and groups, thus enabling an increased level of privacy in and between organizations.

Keywords: service interoperability, mediation process, web services, software components, mediator

[1] X. Dong, A. Halevy, J. Madhavan, E. Nemes, J. Zhang: *Similarity Search for Web Services*. Proceedings of the 30th VLDB Conference, Toronto, Canada, 2004.
[2] M. Y. Galperin: *The Molecular Biology Database Collection: 2004 update.* Nucleic Acids Research, 32: 3-22, 2004.
[3] M. Paolucci, T. Kawamura, T. R. Payne, K. Sycara: *Semantic Matching of Web Services Capabilities*. First International Semantic Web Conference (ISWC), LNCS 2342, pp. 333-347. Sardinia, Italy, Juni 2002.
[4] U. Radetzki, A. B. Cremers: *IRIS: A Framework for Mediator-Based Composition of Service-Oriented Software.* 2004 IEEE International Conference on Web Services (ICWS 2004), July 6-9, 2004, San Diego, CA, USA.
[5] E. Rahm, P. A. Bernstein: *A survey of approaches to automatic schema matching.* VLDB Journal, 10(4): 334-350, 2001.
[6] K. Sycara, S. Widoff, M. Klusch, J. Lu: *LARKS: Dynamic Matchmaking Among Heterogeneous Software Agents in Cyberspace.* Autonomous Agents and Multi-Agent Systems, Kluwer Academic Publishers, 5:173–203, 2002.
[7] C. Szyperski, D. Gruntz, S. Murer: *Component Software: Beyond Object-Oriented Programming. Component Software Series.* ACM Press, Addison-Wesley, 2nd Edition, 2002.
[8] I. Tatarinov, Z. Ives, J. Madhavan, A. Halevy, D. Suciu, N. Dalvi, X. Dong, Y. Kadiyska, G. Miklau, P. Mork: *The Piazza peer data management project.* SIGMOD Record 32(3): 2003.