Towards Digital Library Mediation for Web Services

Claudia Niederée[1], Ulrike Steffens[2], Matthias Hemmje[1]

[1]*Fraunhofer-IPSI,*
Darmstadt, Germany
{niederee,hemmje}@ipsi.fhg.de
ipsi.fhg.de

[2]*Software Systems Institute*
TU Hamburg-Harburg,
Hamburg, Germany
ul.steffens@tuhh.de
www.sts.tu-harburg.de

Abstract

Web services and related technologies are a ubiquitous topic within the IT community. Due to their flexible system composition characteristics these technologies can be considered a promising building block for the dynamic networking of people and organizations in our increasingly globalized and quickly changing economies.

With the growing number of Web services we are in need of elaborate mechanisms or service location, selection, and access as well as for assuring service quality going far beyond the functionalities of service registries. Fortunately, digital library research and practice provides advanced concepts for these tasks.

The paper discusses requirements imposed on content description, structuring, enrichment, and retrieval when considering services as a new type of library content and presents a blueprint for a Web service digital library aiming to support the match-making between service providers and potential service users.

1. Introduction

Globalization of markets and the rapid development and high competition of economies pushes a trend towards dynamic networking. Stable organizational structures and relationships are replaced by dynamic networks of organizations and people as well as systems. The IT and service infrastructure used in such a context has to be adaptable to the changing requirements and working processes implied by dynamic networking.

Achieving a high degree of platform and programming language independence and hence facilitating the interaction and cooperation of independent system components as well as their replacement by new improved components, the currently emerging Web service technology provides a promising starting point for a dynamic networking infrastructure. Web service technology provides an important building block for the development of run-time adaptable systems dynamically composed from reusable service components [26][25]:

Web as programming interface: Web services provide a network accessible interface to application functionality. Web sites, up to now interpreted merely by humans, are extended by programming interface capabilities.

Flexible system composition: By employing standards like XML for component interaction the integration of heterogeneous components is enabled.

Just-in-time integration: Due to the use of service registries the choice and binding of the service to be employed can even happen at run time.

These properties of the Web service technology can be exploited to achieve the IT infrastructure flexibility needed in dynamic networking. The autonomous development of a growing number of competing as well as complementing Web services and the necessity of frequent service selection and reconfiguration in dynamic networking also impose some special challenges: How to identify and locate an adequate service for a task? How to choose between several similar services? How to assure quality and long-term availability of a chosen service?

The value adding mediator role of a digital library can be applied to Web services considering them as library content, hence facilitating service selection and access as well as assuring service quality. However, additional library functionality and metadata for content description and enrichment become necessary as discussed in section 2. Starting from existing Web service technologies, section 3 presents a first design of required service components and metadata for the effective management and demand-oriented mediation of high quality Web services in a digital library. The paper concludes with a discussion of related work and future research directions.

2. Services as Library Content

Development environments like *.net* of *Microsoft* and *Apache's Axis* project simplify the development of new Web services as well as the provision of existing compo-

nents in the form of Web services. This will lead to a rapidly growing number of available Web services for a wide variety of purposes provided by a heterogeneous community of developers and organizations.

Most of the components that a system integrator needs for the system he is about to construct are available somewhere as ready-to-use Web services. But how to identify and locate the adequate high-quality services that he is ready to build his application upon? Do the discovered services behave as required?

Selecting, collecting, structuring, and annotating Web services, a digital library can act as mediator between the available service content and the user community, that consists of both human service users and software agents.
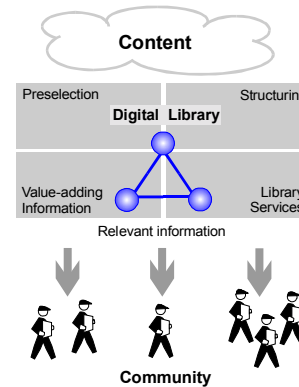
2.1. Content-To-Community Mediators

A digital library to mediates between the information needs of the user community and the available content [24]. This is achieved by (see figure 1):

Content preselection: The library selects high-quality content potentially relevant for its user community;

Content structuring: The library structures the content according to the predominant domain understanding of the user community;

Content enrichment: Domain experts and community members enrich content objects with descriptive and value-adding metadata;

Library services: Services for content retrieval, access, annotation, etc. support the identification of relevant material and facilitate content access;

The benefit of the four contributions above also applies to new forms of library content like Web services.

2.2. Service Content Characteristics

A digital library is open for a wide variety of new content types [18]. However, new content types may require adapted metadata, structuring, and service support as well as new methods for the preselection of high-quality content objects. This also holds true for service content. Services are not static information objects, but operational components implying the following characteristics:

1. Services expose dynamic operational characteristics like efficiency, availability, etc.
2. Service semantics is coded into software and therefore not subject to "understanding by reading", but subject to "understanding by testing or using";
3. Service changes are under control of an independent service provider;
4. The users of a digital service library can be humans as well as software agents;
5. Services might be used to build up more complex systems;
6. Content access is a two-step-process: service location is followed by accessing the service functionality;

Figure 1: The Digital Library as Mediator

7. Only service *descriptions* may be stored by the library not the content itself, the services;

2.3. Service Content Requirements

For information objects, like text documents, effective content classification as well as content description and enrichment by metadata are well-understood issues. Agreed upon sets of metadata fields are summarized in standards like Dublin Core [15] comprising information for content retrieval, selection, and interpretation. It is also agreed that other kinds of information objects like images require a different or extended set of metadata fields for their effective description (see e.g. [29]).

Clearly, service content exhibits relevant characteristics different from those of text documents. Document subject as a central semantic property is, for example, replaced by functional service semantics. The service provider (as equivalent to a document creator) and service functionality contribute to a changed basis for content classification. This leads to the first of three central requirements for the support of service content:

Additional types of metadata are needed for functional and operational service characteristics relevant for service selection and classification; metadata models must also be suited for the interpretation by software agents.

Key aspects for content preselection in a library are relevance for the user community and quality. Decisions are based on a good understanding of the community as well as of the underlying domain. Service preselection may be based on similar approaches: Expertise of qualified personnel, community feedback, and usage statistics.

Content quality is a more severe issue. When composing a complex system from Web services the operation of the composed system heavily relies on its components' quality. The quality of service content should be assessed by applying objective forms of automatic testing.

New methods of quality assessment are necessary to ensure the quality of the offered service content. Service quality, however, may change over time, so that continuous re-assessment becomes necessary.
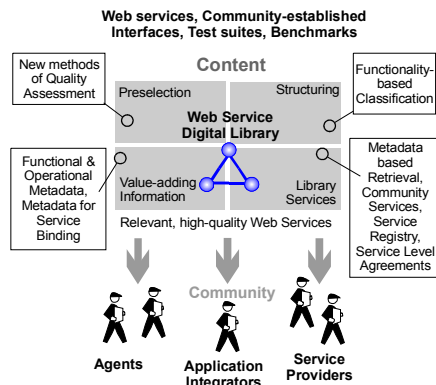
Figure 2: Additional Components of a Web Service Digital Library

Querying a Web service digital library, most users come up with clear expectations of service features like the required interface or behavior. Traditional retrieval methods may not extract this type of information from a service without previous efforts on service description.

Metadata-based retrieval taking into account the description of service interfaces and semantics is of central importance in service discovery.

3. A Digital Library of Web Services

A digital library with Web service content has to provide solutions for the identified requirements. This section presents a blueprint for such a library (see figure 2).

3.1. Overview

The basic Web service architecture mainly consists of three interacting components (see e.g. [27]): The service provider, the service client, and a service registry. In terms of this architecture the digital library takes the role of an advanced service registry. Service providers as well as agents and human application integrators looking for adequate services are members of the library's user community. Service clients directly interact with the service provider after choosing a service from the library.

Figure 2 summarizes the additional components and information structures necessary for the management of Web services in a digital library:

- New automatic, semi-automatic, and community-driven methods of quality assessment influence the decisions in content preselection and maintenance.
- Metadata describing functional and operational service properties contribute to content enrichment.
- Content is either structured by service functionality or by service provider.
- Additional library services are envisioned. This includes community services for quality control, retrieval based on service metadata and registration of services.

In Web service technology there are already some basic developments in the areas of service description and discovery, related to the aims of a Web service digital library. The descriptive capabilities of the Web service description language (WSDL) are addressed in section 3.2. The Universal Description, Discovery, and Integration (UDDI) specification describes basic support for service registration and discovery [22]. A UDDI registry manages Web services and information about them providing business entities. By introducing the concept of tModels, UDDI also supports service classification. tModels are registered like services and subsequently used as reference points for description and classification.

3.2. Functional Metadata

A full description of a service's functionality consists of a specification of the service interface and of the service semantics. Both aspects have to be covered by the metadata managed in the Web service digital library.

Service Interfaces: For the RPC style Web services considered in this paper a service offers a set of methods that can be accessed via SOAP envelopes [13]. This set of method signatures constitutes the service interface.

The *Web Service Description Language* (*WSDL*, [6]), defines a format for describing service interfaces in terms of employed data types, method names, and exchanged messages that are used to model input and output parameters as well as faults that may occur during method execution. A WSDL description also provides the technical prerequisites for service access by service clients.

The WSDL specification provides a basis for the service interface description in a digital library. For human users an XML description conforming to traditional APIs can be automatically generated from an existing WSDL description using, e.g., XSLT.

Service Semantics: The interface has to be complemented by a description of the service semantics. In traditional APIs targeting human readers service semantics is given by a textual description. In the Web service digital library such descriptions enable the use of traditional retrieval methods for service retrieval. However, textual semantics description is vague and imposes problems for correct service selection, especially for agents. Formal specification of software functionality, as e.g., described in [23], facilitates service selection, but requires skilled experts and a considerable amount of work.

Alternative approaches for the description of service semantics are the combination of process models with process ontologies [3] and description based on Semantic Web technology [17],[5],[9], combining RDF descriptions with service ontologies enabling more effective and efficient automatic service discovery methods [3][21].

A more pragmatic approach is the community-driven establishment of interfaces that can be implemented by different Web services. Agreement on such interfaces is achieved through discussion within the user community.

Established interfaces represent a kind of standard for the respective service functionality. Referring to traditional APIs, the establishment and use of agreed upon interfaces is common practice for widely used functionality like XML parsing (see, e.g., SAX and DOM interfaces). The UDDI specification [22] enables the registration of *tModels* which describe compliance with a specification, a concept, or a shared design. They can therefore also represent community-accepted interfaces.

Web services implementing the same interface become comparable, which facilitates service selection, and are exchangeable, which is one of the aims of Web service technology. Interface establishment combined with semantic service specification reduces the effort for semantic description, as this is done only once for the interface. All implementing services will then refer to this specification, only describing semantic differences.

3.3. Operational Metadata

A Web service is exposing dynamic operational characteristics which are further important criteria in service selection:

Operation Time: Typical operation times for Web-based applications are 7x24, i.e. all days, day and night. Web services may differ in their operation schedule.

Availability: System problems can make a service unavailable. Typical parameters for the description of availability are percentage of system availability, maximum time to fix, meantime to repair, etc.

Privacy: The way a Web service handles sensitive personal data may e.g. be described by a privacy policy according to P3P [8] which may be automatically matched with the user's privacy preferences.

Security: The security level of a Web service may be described in terms of its commitment to available security standards like [1][14].

Efficiency: The definition and use of common benchmarks increases the comparability of data describing the efficiency of a service. Some properties like response times are network dependent and thus unpredictable.

Accessibility: The Web service may be restricted to certain user groups or require registration and payment.

Validity/Reliability: Applying quality assessment methods to a Web service (see next section) results in a review on its validity, e.g. with respect to a claimed functionality, and on its reliability. Service validity may also refer to the contexts in which the service is applicable.

The availability of operational metadata in the library enables the selection of adequate services. Combined with community-established interfaces, operational metadata facilitates service comparison.

3.4. Methods for Service Quality Assessment

The preselection of high-quality content has been identified as one of the main contributions of a digital library. With traditional content a library ensures a certain quality level by relying on the reviewing procedures of established organizations and by carefully selecting and acquiring relevant content. For service quality a thorough quality assessment, however, requires service testing or relying on reports of former service users.

Test-based Quality Assessment: Since Web services are software components in operation, quality assessment means basically testing. Such tests may target functional and operational properties. Setting up own quality assessment procedures the digital library must take into account that service quality may change over time requiring regular re-assessment which imposes a considerable workload on the digital library.

Instead of pre-checking the service quality, the library may also enable the user to check service quality on demand. Based on the idea of managing community-established interfaces (see section 3.2), the digital library may also provide test suites for these interfaces. This ensures up-to-date test results and reduces workload. The library may also outsource quality assessment.

Community-driven Quality Assessment: Typically, the user community of a Web service digital library consists of technical experts, who can provide valuable feedback on services quality and reliability. The library can collect opinions from community members and include them in its quality assessment procedures applying content rating and collaborative filtering methods (e.g. [4]).

Service Level Agreements: In professional relationships between service providers and consumers, service level agreements (SLAs) play an important role in assuring service quality. A SLA is a contract between provider and consumer on operational characteristics of the service provision [20]. Adopting this approach, the digital library may manage SLA templates, that represent a SLA offer by the service provider. Support for automatic negotiation of SLAs is desirable in the Web service context.

4. Related Work

Popular metadata formats like Dublin Core [15] are targeted towards traditional library content. Beyond, there are also content types for which the extension of these standard formats (see [28][29]) or the development of content type-specific metadata formats is required (e.g. geospatial information objects [12]).

The Semantic Web activities [17][9][7] are strongly related to the Web service activities in general (also see section 3.2). For the *orchestration* of Web services, i.e., the composition of complex from simpler Web services, different approaches are under development. This includes the definition of composition languages like WSFL [19] and XLANG [26] and service semantics description enabling semi-automatic composition [21]. The operational properties provided by the digital library might be used

during orchestration for service selection as well as for deducing the properties of the composed service.

Our approach is related to brokering approaches that aim at automating the mediation between available services and service clients. [2] e.g. presents an approach for flexible brokering of problem solving methods for building knowledge based systems.

4. Conclusions and Further Research

Services as a new kind of library content imply extensions and adaptations in the areas of digital library structures, functionality and metadata. Based on these extra requirements we presented a Web service digital library blueprint. The discussion of the blueprint focussed on operational and functional metadata for Web services and on test-based, contract-based and community-driven methods for content quality assessment and assurance.

The envisioned service digital library can be used for selecting adequate service components in IT environments. For full support of dynamic IT environments, this functionality has to be extended by a Web service orchestration language (see section 4) and by flexible tools for on-the-fly system customization [10], [11].

Starting from this first blueprint for a Web service digital library, one challenging research issue is the development and evaluation of semantic service descriptions combining expressiveness, feasibility with respect to the required definition effort, and effective support of intelligent agents in service discovery (see section 3.2). A further important research issue is the adequate representation and management of the envisioned operational metadata. The ABC model and ontology [16] can be used as a guideline for setting up a service metadata model.

6 References

[1]   M. Bartel, J. Boyer, B. Fox, B. LaMacchia, and E. Simon. XML-Signature Syntax and Processing - W3C Rec., 12 Feb. 2002.

[2]   V. Richard Benjamins, B. Wielinga, J. Wielemaker and D. Fensel. "Towards Brokering Problem-Solving Knowledge on the Internet". In *European Knowledge Acquisition Workshop (EKAW-99),* May 1999

[3]   A. Bernstein and M. Klein. High Precision Service Retrieval. In *Proc. of the 1st International Semantic Web Conference* (ISWC'02), Sardinia, Italy (to appear), 2002.

[4]   V. Bouthors and O. Dedieu. "Pharos, a Collaborative Infrastructure for Web Knowledge Sharing". In *Research and Advanced Technology for Digital Libraries, Proc. of the 3rd European Conference, ECDL'99*, Paris, France, 1999.

[5]   D. Brickley and R.V. Guha. "Resource Description Framework (RDF) Schema Spec. 1.0, W3C Cand. Rec." March 2000.

[6]   E. Christensen, F. Curbera, G. Meredith, and S. Weerawarana. Web Services Description Language (WSDL) 1.1 - W3C Note, 15 March 2001.

[7]   D. Connolly, F.van Harmelen, I. Horrocks, D. L. McGuinness, P. F.Patel-Schneider, and L.A. Stein. DAML+OIL (March

DAML+OIL (March 2001) Reference Description-W3C Note, 18 December 2001.

[8]   L. Cranor, M. Langheinrich, M. Marchiori, and J. Reagle. The Platform for Privacy Preferences 1.0 (P3P1.0) Spec., W3C Rec. 16 April 2002.

[9]   D. Fensel, F. van Harmelen, I. Horrocks, D.L. McGuiness, and P.F. Patel-Schneider. OIL: An Ontology Infrastructure for the Semantic Web. *IEEE Intelligent Systems*, 16(2):38 -45, April 2001.

[10] G. Fisher. Social Creativity, Symmetry of Ignorance, and Metadesign. *Knowledge-based Systems Journal*, 13(7-8):527-537, December 2000.

[11] M. Fuchs, C. Muscogiuri, C. Niederée, and M. Hemmje. An Open Framework for Integrated Qualification Management Portals. In *Proc. of the 3rd Int. Workshop on Management of Information on the Web (MIW'2002),* Sept. 2002 (to appear).

[12] Metadata Ad Hoc Working Group. Content Standard for Digital Geospatial Metadata. Standard Document FGDC-STD-001-1998, Federal Geographic Data Committee, Reston, Virginia, June 1998. http://www.fgdc.gov/metadata/csdgm/.

[13] M. Gudgin, M. Hadley, J.-J. Moreau, and H. Frystyk Nielsen. SOAP Version 1.2 Part 1: Messaging Framework - W3C Working Draft, 17 December 2001.

[14] T. Imamura, B. Dillaway, and E. Simon. XML EncryptionSyntax and Processing - W3C Cand. Rec., 04 March 2002.

[15] Dublin Core Metadata Inititative. Dublin Core Metadata Element Set, Version 1.1: Reference Description. July 1999.

[16] C. Lagoze and J. Hunter. The ABC Ontology and Model. *Journal of Digital Information*, 2(2), November 2001.

[17] O. Lassila and R. R. Swick. Resource Description Framework (RDF) Model and Syntax, W3C Rec. February 1999.

[18] D. M. Levy and C. C. Marshall. "What Color was George Washington's White Horse? A Look at the Assumptions Underlying Digital Libraries". In *1st Ann. Conf. on the Theory and Practice of Digital Libraries*, College Station, USA, 1994.

[19] F. Leymann. Web Services Flow Language (WSFL 1.0). Whitepaper, IBM Software Group, May 2001.

[20] I. Macfarlane and C. Rudd. "IT Service Management".*The IT Infrastructure Library*. itSMF Ltd., Reading, UK, 2001.

[21] S. A. McIlraith, T. C. Son, and H. Zeng. Semantic Web Services. *IEEE Intelligent Systems*, 16(2):46 - 53, April 2001.

[22] B. McKee, D. Ehnebuske, and D. Rogers. UDDI Version 2.0 API Spec. UDDI - Open Draft Spec., 8 June 2001.

[23] A. Mili, R. Mili, and R. T. Mittermeir. Storing and Retrieving Software Components: A Renement Based System. *IEEE Transactions on Software Engineering*, 23(7), July 1997.

[24] J.W. Schmidt, G. Schröder, C. Niederee, and F. Matthes. "Linguistic and Architectural Requirements for Personalized Digital Libraries". *Int. Journal on Digital Libraries*, 1(1), 1997.

[25] J. Snell, D. Tidwell, and P. Kulchenko. Programming Web Services with SOAP. Building Distributed Applications. O'Reilly, CA, 2002.

[26] S. Thatte. "XLANG: Web Services for Business Pro-cess Design". Whitepaper, Microsoft Corporation, 2001. http://www.gotdotnet.com/team/xml wsspecs/xlang-c/.

[27] S. J. Vaughan-Nichols. "Web Services: Beyond the Hype". *Computer*, 35(2):18-21, February 2002.

[28] S. Weibel and R. Iannella. "The 4th Dublin Core Metadata Workshop Report". *D-Lib Magazine*, June 1997.

[29] S. Weibel and E. Miller. "Image Description on the Internet - A Summary of the CNI/OCLC Image Metadata Workshop". *D-Lib Magazine*, January 1997